VOLUME

1

# ENCYCLOPEDIA OF MATHEMATICAL PHYSICS

EDITED BY

JEAN-PIERRE FRANÇOISE

GREGORY L. NABER

TSOU SHEUNG TSUN

$$\delta \int \mathcal{L} = 0$$

AP

# EDITORS

**Jean-Pierre Françoise**
Université P.-M. Curie, Paris VI
Paris, France

**Gregory L. Naber**
Drexel University
Philadelphia, PA, USA

**Tsou Sheung Tsun**
University of Oxford
Oxford, UK

# EDITORIAL ADVISORY BOARD

# FOREWORD

In bygone centuries, our physical world appeared to be filled to the brim with mysteries. Divine powers could provide for genuine miracles; water and sunlight could turn arid land into fertile pastures, but the same powers could lead to miseries and disasters. The force of life, the *vis vitalis*, was assumed to be the special agent responsible for all living things. The heavens, whatever they were for, contained stars and other heavenly bodies that were the exclusive domain of the Gods.

Mathematics did exist, of course. Indeed, there was one aspect of our physical world that was recognised to be controlled by precise, mathematical logic: the geometric structure of space, elaborated to become a genuine form of art by the ancient Greeks. From my perspective, the Greeks were the first practitioners of 'mathematical physics', when they discovered that all geometric features of space could be reduced to a small number of axioms. Today, these would be called 'fundamental laws of physics'. The fact that the flow of *time* could be addressed with similar exactitude, and that it could be handled geometrically together with space, was only recognised much later. And, yes, there were a few crazy people who were interested in the magic of numbers, but the real world around us seemed to contain so much more that was way beyond our capacities of analysis.

Gradually, all this changed. The Moon and the planets appeared to follow geometrical laws. Galilei and Newton managed to identify their logical rules of motion, and by noting that the concept of mass could be applied to things in the sky just like apples and cannon balls on Earth, they made the sky a little bit more accessible to us. Electricity, magnetism, light and sound were also found to behave in complete accordance with mathematical equations.

Yet all of this was just a beginning. The real changes came with the twentieth century. A completely new way of thinking, by emphasizing mathematical, logical analysis rather than empirical evidence, was pioneered by Albert Einstein. Applying advanced mathematical concepts, only known to a few pure mathematicians, to notions as mundane as space and time, was new to the physicists of his time. Einstein himself had a hard time struggling through the logic of connections and curvatures, notions that were totally new to him, but are only too familiar to students of mathematical physics today. Indeed, there is no better testimony of Einstein's deep insights at that time, than the fact that we now teach these things regularly in our university classrooms.

Special and general relativity are only small corners of the realm of modern physics that is presently being studied using advanced mathematical methods. We have notoriously complex subjects such as phase transitions in condensed matter physics, superconductivity, Bose–Einstein condensation, the quantum Hall effect, particularly the fractional quantum Hall effect, and numerous topics from elementary particle physics, ranging from fibre bundles and renormalization groups to supergravity, algebraic topology, superstring theory, Calabi–Yau spaces and what not, all of which require the utmost of our mental skills to comprehend them.

The most bewildering observation that we make today is that it seems that our *entire* physical world appears to be controlled by mathematical equations, and these are not just sloppy and debatable models, but precisely documented properties of materials, of systems, and of phenomena in all echelons of our universe.

Does this really apply to our entire world, or only to parts of it? Do features, notions, entities exist that are emphatically *not* mathematical? What about intuition, or dreams, and what about consciousness? What about religion? Here, most of us would say, one should not even try to apply mathematical analysis, although even here, some brave social scientists are making attempts at coordinating rational approaches.

No, there are clear and important differences between the physical world and the mathematical world. Where the physical world stands out is the fact that it refers to 'reality', whatever 'reality' is. Mathematics is the world of pure logic and pure reasoning. In physics, it is the experimental evidence that ultimately decides whether a theory is acceptable or not. Also, the methodology in physics is different.

A beautiful example is the serendipitous discovery of superconductivity. In 1911, the Dutch physicist Heike Kamerlingh Onnes was the first to achieve the liquefaction of helium, for which a temperature below 4.25 K had to be realized. Heike decided to measure the specific conductivity of mercury, a metal that is frozen solid at such low temperatures. But something appeared to go wrong during the measurements, since the volt meter did not show any voltage at all. All experienced physicists in the team assumed that they were dealing with a malfunction. It would not have been the first time for a short circuit to occur in the electrical equipment, but, this time, in spite of several efforts, they failed to locate it. One of the assistants was responsible for keeping the temperature of the sample well within that of liquid helium, a dull job, requiring nothing else than continuously watching some dials. During one of the many tests, however, he dozed off. The temperature rose, and suddenly the measurements showed the normal values again. It then occurred to the investigators that the effect and its temperature dependence were completely reproducible. Below 4.19 degrees Kelvin the conductivity of mercury appeared to be strictly infinite. Above that temperature, it is finite, and the transition is a very sudden one. Superconductivity was discovered (D. van Delft, "Heike Kamerling Onnes", Uitgeverij Bert Bakker, Amsterdam, 2005 (in Dutch)).

This is not the way mathematical discoveries are made. Theorems are not produced by assistants falling asleep, even if examples do exist of incidents involving some miraculous fortune.

The hybrid science of mathematical physics is a very curious one. Some of the topics in this Encyclopedia are undoubtedly physical. High $T_c$ superconductivity, breaking water waves, and magneto-hydrodynamics, are definitely topics of physics where experimental data are considered more decisive than any high-brow theory. Cohomology theory, Donaldson–Witten theory, and AdS/CFT correspondence, however, are examples of purely mathematical exercises, even if these subjects, like all of the others in this compilation, are strongly inspired by, and related to, questions posed in physics.

It is inevitable, in a compilation of a large number of short articles with many different authors, to see quite a bit of variation in style and level. In this Encyclopedia, theoretical physicists as well as mathematicians together made a huge effort to present in a concise and understandable manner their vision on numerous important issues in advanced mathematical physics. All include references for further reading. We hope and expect that these efforts will serve a good purpose.

Gerard 't Hooft,
Spinoza Institute,
Utrecht University,
The Netherlands.

# PREFACE

Mathematical Physics as a distinct discipline is relatively new. The International Association of Mathematical Physics was founded only in 1976. The interaction between physics and mathematics has, of course, existed since ancient times, but the recent decades, perhaps partly because we are living through them, appear to have witnessed tremendous progress, yielding new results and insights at a dizzying pace, so much so that an encyclopedia seems now needed to collate the gathered knowledge.

Mathematical Physics brings together the two great disciplines of Mathematics and Physics to the benefit of both, the relationship between them being symbiotic. On the one hand, it uses mathematics as a tool to organize physical ideas of increasing precision and complexity, and on the other it draws on the questions that physicists pose as a source of inspiration to mathematicians. A classical example of this relationship exists in Einstein's theory of relativity, where differential geometry played an essential role in the formulation of the physical theory while the problems raised by the ensuing physics have in turn boosted the development of differential geometry. It is indeed a happy coincidence that we are writing now a preface to an encyclopedia of mathematical physics in the centenary of Einstein's annus mirabilis.

The project of putting together an encyclopedia of mathematical physics looked, and still looks, to us a formidable enterprise. We would never have had the courage to undertake such a task if we did not believe, first, that it is worthwhile and of benefit to the community, and second, that we would get the much-needed support from our colleagues. And this support we did get, in the form of advice, encouragement, and practical help too, from members of our Editorial Advisory Board, from our authors, and from others as well, who have given unstintingly so much of their time to help us shape this Encyclopedia.

Mathematical Physics being a relatively new subject, it is not yet clearly delineated and could mean different things to different people. In our choice of topics, we were guided in part by the programs of recent International Congresses on Mathematical Physics, but mainly by the advice from our Editorial Advisory Board and from our authors. The limitations of space and time, as well as our own limitations, necessitated the omission of certain topics, but we have tried to include all that we believe to be core subjects and to cover as much as possible the most active areas.

Our subject being interdisciplinary, we think it appropriate that the Encyclopedia should have certain special features. Applications of the same mathematical theory, for instance, to different problems in physics will have different emphasis and treatment. By the same token, the same problem in physics can draw upon resources from different mathematical fields. This is why we divide the Encyclopedia into two broad sections: physics subjects and related mathematical subjects. Articles in either section are deliberately allowed a fair amount of overlap with one another and many articles will appear under more than one heading, but all are linked together by elaborate cross referencing. We think this gives a better picture of the subject as a whole and will serve better a community of researchers from widely scattered yet related fields.

The Encyclopedia is intended primarily for experienced researchers but should be of use also to beginning graduate students. For the latter category of readers, we have included eight elementary introductory articles for easy reference, with those on mathematics aimed at physics graduates and those on physics aimed at mathematics graduates, so that these articles can serve as their first port of call to enable them to embark on any of the main articles without the need to consult other material beforehand. In fact, we think these articles may even form the

foundation of advanced undergraduate courses, as we know that some authors have already made such use of them.

In addition to the printed version, an on-line version of the Encyclopedia is planned, which will allow both the contents and the articles themselves to be updated if and when the occasion arises. This is probably a necessary provision in such a rapidly advancing field.

This project was some four years in the making. Our foremost thanks at its completion go to the members of our Editorial Advisory Board, who have advised, helped and encouraged us all along, and to all our authors who have so generously devoted so much of their time to writing these articles and given us much useful advice as well. We ourselves have learnt a lot from these colleagues, and made some wonderful contacts with some among them. Special thanks are due also to Arthur Greenspoon whose technical expertise was indispensable.

The project was started with Academic Press, which was later taken over by Elsevier. We thank warmly members of their staff who have made this transition admirably seamless and gone on to assist us greatly in our task: both Carey Chapman and Anne Guillaume, who were in charge of the whole project and have been with us since the beginning, and Edward Taylor responsible for the copy-editing. And Martin Ruck, who manages to keep an overwhelming amount of details constantly at his fingertips, and who is never known to have lost a single email, deserves a very special mention.

As a postscript, we would like to express our gratitude to the very large number of authors who generously agreed to donate their honorariums to support the Committee for Developing Countries of the European Mathematical Society in their work to help our less fortunate colleagues in the developing world.

Jean-Pierre Françoise  
Gregory L. Naber  
Tsou Sheung Tsun

# PERMISSION ACKNOWLEDGMENTS

The following material is reproduced with kind permission of Nature Publishing Group
Figures 11 and 12 of “Point-vortex Dynamics”
http://www.nature.com/nature
The following material is reproduced with kind permission of Oxford University Press
Figure 1 of “Random Walks in Random Environments”
http://www.oup.co.uk

# GUIDE TO USE OF THE ENCYCLOPEDIA

## Structure of the Encyclopedia

The material in this Encyclopedia is organised into two sections. At the start of Volume 1 are eight **Introductory Articles**. The introductory articles on mathematics are aimed at physics graduates; those on physics are aimed at mathematics graduates. It is intended that these articles should serve as the first port of call for graduate students, to enable them to embark on any of the main entries without the need to consult other material beforehand.

Following the Introductory Articles, the main body of the Encyclopedia is arranged as a series of entries in alphabetical order. These entries fill the remainder of Volume 1 and all of the subsequent volumes (2–5).

To help you realize the full potential of the material in the Encyclopedia we have provided four features to help you find the topic of your choice: a contents list by subject, an alphabetical contents list, cross-references, and a full subject index.

## 1. Contents List by Subject

Your first point of reference will probably be the contents list by subject. This list appears at the front of each volume, and groups the entries under subject headings describing the broad themes of mathematical physics. This will enable the reader to make quick connections between entries and to locate the entry of interest. The contents list by subject is divided into two main sections: *Physics Subjects* and *Related Mathematics Subjects*. Under each main section heading, you will find several subject areas (such as GENERAL RELATIVITY in Physics Subjects or NONCOMMUTATIVE GEOMETRY in Related Mathematics Subjects). Under each subject area is a list of those entries that cover aspects of that subject, together with the volume and page numbers on which these entries may be found.

Because mathematical physics is so highly interconnected, individual entries may appear under more than one subject area. For example, the entry GAUGE THEORY: MATHEMATICAL APPLICATIONS is listed under the Physics Subject GAUGE THEORY as well as in a broad range of Related Mathematics Subjects.

## 2. Alphabetical Contents List

The alphabetical contents list, which also appears at the front of each volume, lists the entries in the order in which they appear in the Encyclopedia. This list provides both the volume number and the page number of the entry.

You will find "dummy entries" where obvious synonyms exist for entries or where we have grouped together related topics. Dummy entries appear in both the contents list and the body of the text.

*Example*
If you were attempting to locate material on path integral methods via the alphabetical contents list:

PATH INTEGRAL METHODS *see* Functional Integration in Quantum Physics; Feynman Path Integrals

The dummy entry directs you to two other entries in which path integral methods are covered. At the appropriate locations in the contents list, the volume and page numbers for these entries are given.

If you were trying to locate the material by browsing through the text and you had looked up Path Integral Methods, then the following information would be provided in the dummy entry:

| **Path Integral Methods** *see* Functional Integration in Quantum Physics; Feynman Path Integrals |
|---|

## 3. Cross-References

All of the articles in the Encyclopedia have been extensively cross-referenced. The cross-references, which appear at the end of an entry, serve three different functions:

i. To indicate if a topic is discussed in greater detail elsewhere.

ii. To draw the reader's attention to parallel discussions in other entries.

iii. To indicate material that broadens the discussion.

*Example*
The following list of cross-references appears at the end of the entry STOCHASTIC HYDRODYNAMICS

> *See also*: Cauchy Problem for Burgers-Type Equations; Hamiltonian Fluid Dynamics; Incompressible Euler Equations: Mathematical Theory; Malliavin Calculus; Non-Newtonian Fluids; Partial Differential Equations: Some Examples; Stochastic Differential Equations; Turbulence Theories; Viscous Incompressible Fluids: Mathematical Theory; Vortex Dynamics

Here you will find examples of all three functions of the cross-reference list: a topic discussed in greater detail elsewhere (e.g. Incompressible Euler Equations: Mathematical Theory), parallel discussion in other entries (e.g. Stochastic Differential Equations) and reference to entries that broaden the discussion (e.g. Turbulence Theories).

The eight Introductory Articles are not cross-referenced from any of the main entries, as it is expected that introductory articles will be of general interest. As mentioned above, the Introductory Articles may be found at the start of Volume 1.

## 4. Index

The index will provide you with the volume and page number where the material is located. The index entries differentiate between material that is a whole entry, is part of an entry, or is data presented in a figure or table. Detailed notes are provided on the opening page of the index.

## 5. Contributors

A full list of contributors appears at the beginning of each volume.

# CONTRIBUTORS

**A Abbondandolo**
Università di Pisa
Pisa, Italy

**M J Ablowitz**
University of Colorado
Boulder, CO, USA

**S L Adler**
Institute for Advanced Study
Princeton, NJ, USA

**H Airault**
Université de Picardie
Amiens, France

**G Alberti**
Università di Pisa
Pisa, Italy

**S Albeverio**
Rheinische Friedrich–Wilhelms-Universität Bonn
Bonn, Germany

**S T Ali**
Concordia University
Montreal, QC, Canada

**R Alicki**
University of Gdańsk
Gdańsk, Poland

**G Altarelli**
CERN
Geneva, Switzerland

**C Amrouche**
Université de Pau et des Pays de l'Adour
Pau, France

**M Anderson**
State University of New York at Stony Brook
Stony Brook, NY, USA

**L Andersson**
University of Miami
Coral Gables, FL, USA and Albert Einstein Institute
Potsdam, Germany

**B Andreas**
Humboldt-Universität zu Berlin
Berlin, Germany

**V Araújo**
Universidade do Porto
Porto, Portugal

**A Ashtekar**
Pennsylvania State University
University Park, PA, USA

**W Van Assche**
Katholieke Universiteit Leuven
Leuven, Belgium

**G Aubert**
Université de Nice Sophia Antipolis
Nice, France

**H Au-Yang**
Oklahoma State University
Stillwater, OK, USA

**M A Aziz-Alaoui**
Université du Havre
Le Havre, France

**V Bach**
Johannes Gutenberg-Universität
Mainz, Germany

**C Bachas**
Ecole Normale Supérieure
Paris, France

**V Baladi**
Institut Mathématique de Jussieu
Paris, France

**D Bambusi**
Università di Milano
Milan, Italy

**C Bardos**
Université de Paris 7
Paris, France

**D Bar-Natan**
University of Toronto
Toronto, ON, Canada

**E L Basor**
California Polytechnic State University
San Luis Obispo, CA, USA

**M T Batchelor**
Australian National University
Canberra, ACT, Australia

**S Bauer**
Universität Bielefeld
Bielefeld, Germany

**V Beffara**
Ecole Nomale Supérieure de Lyon
Lyon, France

**R Beig**
Universität Wien
Vienna, Austria

**M I Belishev**
Petersburg Department of Steklov Institute of Mathematics
St. Petersburg, Russia

**P Bernard**
Université de Paris Dauphine
Paris, France

**D Birmingham**
University of the Pacific
Stockton, CA, USA

**Jiří Bičák**
Charles University, Prague, Czech Republic and Albert Einstein Institute
Potsdam, Germany

**C Blanchet**
Université de Bretagne-Sud
Vannes, France

**M Blasone**
Università degli Studi di Salerno
Baronissi (SA), Italy

**M Blau**
Université de Neuchâtel
Neuchâtel, Switzerland

**S Boatto**
IMPA
Rio de Janeiro, Brazil

**L V Bogachev**
University of Leeds
Leeds, UK

**L Boi**
EHESS and LUTH
Paris, France

**M Bojowald**
The Pennsylvania State University
University Park, PA, USA

**C Bonatti**
Université de Bourgogne
Dijon, France

**P Bonckaert**
Universiteit Hasselt
Diepenbeek, Belgium

**F Bonetto**
Georgia Institute of Technology
Atlanta, GA, USA

**G Bouchitté**
Université de Toulon et du Var
La Garde, France

**A Bovier**
Weierstrass Institute for Applied Analysis and Stochastics
Berlin, Germany

**H W Braden**
University of Edinburgh
Edinburgh, UK

**H Bray**
Duke University
Durham, NC, USA

**Y Brenier**
Université de Nice Sophia Antipolis
Nice, France

**J Bros**
CEA/DSM/SPhT, CEA/Saclay
Gif-sur-Yvette, France

**R Brunetti**
Universität Hamburg
Hamburg, Germany

**M Bruschi**
Università di Roma “La Sapienza”
Rome, Italy

**T Brzeziński**
University of Wales Swansea
Swansea, UK

**D Buchholz**
Universität Göttingen
Göttingen, Germany

**N Burq**
Université Paris-Sud
Orsay, France

**F H Busse**
Universität Bayreuth
Bayreuth, Germany

**G Buttazzo**
Università di Pisa
Pisa, Italy

**P Buttà**
Università di Roma “La Sapienza”
Rome, Italy

**S L Cacciatori**
Università di Milano
Milan, Italy

**P T Callaghan**
Victoria University of Wellington
Wellington, New Zealand

**Francesco Calogero**
University of Rome, Rome, Italy and Institute
Nazionale di Fisica Nucleare
Rome, Italy

**A Carati**
Università di Milano
Milan, Italy

**J Cardy**
Rudolf Peierls Centre for Theoretical Physics
Oxford, UK

**R Caseiro**
Universidade de Coimbra
Coimbra, Portugal

**A S Cattaneo**
Universität Zürich
Zürich, Switzerland

**A Celletti**
Università di Roma “Tor Vergata”
Rome, Italy

**D Chae**
Sungkyunkwan University
Suwon, South Korea

**G-Q Chen**
Northwestern University
Evanston, IL, USA

**L Chierchia**
Università degli Studi “Roma Tre”
Rome, Italy

**S Chmutov**
Petersburg Department of Steklov
Institute of Mathematics
St. Petersburg, Russia

**M W Choptuik**
University of British Columbia
Vancouver, Canada

**Y Choquet-Bruhat**
Université P.-M. Curie, Paris VI
Paris, France

**P T Chruściel**
Université de Tours
Tours, France

**Chong-Sun Chu**
University of Durham
Durham, UK

**F Cipriani**
Politecnico di Milano
Milan, Italy

**R L Cohen**
Stanford University
Stanford, CA, USA

**T H Colding**
University of New York
New York, NY, USA

**J C Collins**
Penn State University
University Park, PA, USA

**G Comte**
Université de Nice Sophia Antipolis
Nice, France

**A Constantin**
Trinity College
Dublin, Republic of Ireland

**D Crowdy**
Imperial College
London, UK

**A B Cruzeiro**
University of Lisbon
Lisbon, Portugal

**G Dal Maso**
SISSA
Trieste, Italy

**F Dalfovo**
Università di Trento
Povo, Italy

**A S Dancer**
University of Oxford
Oxford, UK

**P D'Ancona**
Università di Roma "La Sapienza"
Rome, Italy

**S R Das**
University of Kentucky
Lexington, KY, USA

**E Date**
Osaka University
Osaka, Japan

**N Datta**
University of Cambridge
Cambridge, UK

**G W Delius**
University of York
York, UK

**G F dell'Antonio**
Università di Roma "La Sapienza"
Rome, Italy

**C DeWitt-Morette**
The University of Texas at Austin
Austin, TX, USA

**L Diósi**
Research Institute for Particle and Nuclear Physics
Budapest, Hungary

**A Doliwa**
University of Warmia and Mazury in Olsztyn
Olsztyn, Poland

**G Dolzmann**
University of Maryland
College Park, MD, USA

**S K Donaldson**
Imperial College
London, UK

**T C Dorlas**
Dublin Institute for Advanced Studies
Dublin, Republic of Ireland

**M R Douglas**
Rutgers, The State University of New Jersey
Piscataway, NJ, USA

**M Dütsch**
Universität Zürich
Zürich, Switzerland

**B Dubrovin**
SISSA-ISAS
Trieste, Italy

**J J Duistermaat**
Universiteit Utrecht
Utrecht, The Netherlands

**S Duzhin**
Petersburg Department of Steklov Institute of Mathematics
St. Petersburg, Russia

**G Ecker**
Universität Wien
Vienna, Austria

**M Efendiev**
Universität Stuttgart
Stuttgart, Germany

**T Eguchi**
University of Tokyo
Tokyo, Japan

**J Ehlers**
Max Planck Institut für Gravitationsphysik
(Albert-Einstein Institut)
Golm, Germany

**P E Ehrlich**
University of Florida
Gainesville, FL, USA

**D Einzel**
Bayerische Akademie der Wissenschaften
Garching, Germany

**G A Elliott**
University of Toronto
Toronto, Canada

**G F R Ellis**
University of Cape Town
Cape Town, South Africa

**C L Epstein**
University of Pennsylvania
Philadelphia, PA, USA

**J Escher**
Universität Hannover
Hannover, Germany

**J B Etnyre**
University of Pennsylvania
Philadelphia, PA, USA

**G Falkovich**
Weizmann Institute of Science
Rehovot, Israel

**M Farge**
Ecole Normale Supérieure
Paris, France

**B Ferrario**
Università di Pavia
Pavia, Italy

**R Finn**
Stanford University
Stanford, CA, USA

**D Fiorenza**
Università di Roma “La Sapienza”
Rome, Italy

**A E Fischer**
University of California
Santa Cruz, CA, USA

**A S Fokas**
University of Cambridge
Cambridge, UK

**J-P Françoise**
Université P.-M. Curie, Paris VI
Paris, France

**S Franz**
The Abdus Salam ICTP
Trieste, Italy

**L Frappat**
Université de Savoie
Chambery-Annecy, France

**J Frauendiener**
Universität Tübingen
Tübingen, Germany

**K Fredenhagen**
Universität Hamburg
Hamburg, Germany

**S Friedlander**
University of Illinois-Chicago
Chicago, IL, USA

**M R Gaberdiel**
ETH Zürich
Zürich, Switzerland

**G Gaeta**
Università di Milano
Milan, Italy

**L Galgani**
Università di Milano
Milan, Italy

**G Gallavotti**
Università di Roma “La Sapienza”
Rome, Italy

**R Gambini**
Universidad de la República
Montevideo, Uruguay

**G Gentile**
Università degli Studi “Roma Tre”
Rome, Italy

**A Di Giacomo**
Università di Pisa
Pisa, Italy

**P B Gilkey**
University of Oregon
Eugene, OR, USA

**R Gilmore**
Drexel University
Philadelphia, PA, USA

**S Gindikin**
Rutgers University
Piscataway, NJ, USA

**A Giorgilli**
Università di Milano
Milan, Italy

**G A Goldin**
Rutgers University
Piscataway, NJ, USA

**G González**
Louisiana State University
Baton Rouge, LA, USA

**R Gopakumar**
Harish-Chandra Research Institute
Allahabad, India

**D Gottesman**
Perimeter Institute
Waterloo, ON, Canada

**H Gottschalk**
Rheinische Friedrich-Wilhelms-Universität Bonn
Bonn, Germany

**O Goubet**
Université de Picardie Jules Verne
Amiens, France

**T R Govindarajan**
The Institute of Mathematical Sciences
Chennai, India

**A Grassi**
University of Pennsylvania
Philadelphia, PA, USA

**P G Grinevich**
L D Landau Institute for
Theoretical Physics
Moscow, Russia

**Ch Gruber**
Ecole Polytechnique Fédérale de Lausanne
Lausanne, Switzerland

**J-L Guermond**
Université de Paris Sud
Orsay, France

**F Guerra**
Università di Roma “La Sapienza”
Rome, Italy

**T Guhr**
Lunds Universitet
Lund, Sweden

**C Guillopé**
Université Paris XII – Val de Marne
Créteil, France

**C Gundlach**
University of Southampton
Southampton, UK

**S Gutt**
Université Libre de Bruxelles
Brussels, Belgium

**K Hannabuss**
University of Oxford
Oxford, UK

**M Haragus**
Université de Franche-Comté
Besançon, France

**S G Harris**
St. Louis University
St. Louis, MO, USA

**B Hasselblatt**
Tufts University
Medford, MA, USA

**P Hayden**
McGill University
Montreal, QC, Canada

**D C Heggie**
The University of Edinburgh
Edinburgh, UK

**B Helffer**
Université Paris-Sud
Orsay, France

**G M Henkin**
Université P.-M. Curie, Paris VI
Paris, France

**M Henneaux**
Université Libre de Bruxelles
Bruxelles, Belgium

**S Herrmann**
Université Henri Poincaré, Nancy 1
Vandoeuvre-lès-Nancy, France

**C P Herzog**
University of California at Santa Barbara
Santa Barbara, CA, USA

**J G Heywood**
University of British Columbia
Vancouver, BC, Canada

**A C Hirshfeld**
Universität Dortmund
Dortmund, Germany

**A S Holevo**
Steklov Mathematical Institute
Moscow, Russia

**T J Hollowood**
University of Wales Swansea
Swansea, UK

**D D Holm**
Imperial College
London, UK

**J-W van Holten**
NIKHEF
Amsterdam, The Netherlands

**A Huckleberry**
Ruhr-Universität Bochum
Bochum, Germany

**K Hulek**
Universität Hannover
Hannover, Germany

**D Iagolnitzer**
CEA/DSM/SPhT, CEA/Saclay
Gif-sur-Yvette, France

**R Illge**
Friedrich-Schiller-Universität Jena
Jena, Germany

**P Imkeller**
Humboldt Universität zu Berlin
Berlin, Germany

**G Iooss**
Institut Non Linéaire de Nice
Valbonne, France

**M Irigoyen**
Université P.-M. Curie, Paris VI
Paris, France

**J Isenberg**
University of Oregon
Eugene, OR, USA

**R Ivanova**
University of Hawaii Hilo
Hilo, HI, USA

**E M Izhikevich**
The Neurosciences Institute
San Diego, CA, USA

**R W Jackiw**
Massachusetts Institute of Technology
Cambridge, MA, USA

**J K Jain**
The Pennsylvania State University
University Park, PA, USA

**M Jardim**
IMECC–UNICAMP
Campinas, Brazil

**L C Jeffrey**
University of Toronto
Toronto, ON, Canada

**J Jiménez**
Universidad Politecnica de Madrid
Madrid, Spain

**S Jitomirskaya**
University of California at Irvine
Irvine, CA, USA

**P Jizba**
Czech Technical University
Prague, Czech Republic

**A Joets**
Université Paris-Sud
Orsay, France

**K Johansson**
Kungl Tekniska Högskolan
Stockholm, Sweden

**G Jona-Lasinio**
Università di Roma “La Sapienza”
Rome, Italy

**V F R Jones**
University of California at Berkeley
Berkeley, CA, USA

**N Joshi**
University of Sydney
Sydney, NSW, Australia

**D D Joyce**
University of Oxford
Oxford, UK

**C D Jäkel**
Ludwig-Maximilians-Universität München
München, Germany

**G Kasperski**
Université Paris-Sud XI
Orsay, France

**L H Kauffman**
University of Illinois at Chicago
Chicago, IL, USA

**R K Kaul**
The Institute of Mathematical Sciences
Chennai, India

**Y Kawahigashi**
University of Tokyo
Tokyo, Japan

**B S Kay**
University of York
York, UK

**R Kenyon**
University of British Columbia
Vancouver, BC, Canada

**M Keyl**
Università di Pavia
Pavia, Italy

**T W B Kibble**
Imperial College
London, UK

**S Kichenassamy**
Université de Reims Champagne-Ardenne
Reims, France

**J Kim**
University of California at Irvine
Irvine, USA

**S B Kim**
Chonnam National University
Gwangju, South Korea

**A Kirillov**
University of Pennsylvania
Philadelphia, PA, USA

**A Kirillov, Jr.**
Stony Brook University
Stony Brook, NY, USA

**K Kirsten**
Baylor University
Waco, TX, USA

**F Kirwan**
University of Oxford
Oxford, UK

**S Klainerman**
Princeton University
Princeton, NJ, USA

**I R Klebanov**
Princeton University
Princeton, NJ, USA

**Y Kondratiev**
Universität Bielefeld
Bielefeld, Germany

**A Konechny**
Rutgers, The State University of New Jersey
Piscataway, NJ, USA

**K Konishi**
Università di Pisa
Pisa, Italy

**T H Koornwinder**
University of Amsterdam
Amsterdam, The Netherlands

**P Kornprobst**
INRIA
Sophia Antipolis, France

**V P Kostov**
Université de Nice Sophia Antipolis
Nice, France

**R Kotecký**
Charles University
Prague, Czech Republic and the
University of Warwick, UK

**Y Kozitsky**
Uniwersytet Marii Curie-Sklodowskiej
Lublin, Poland

**P Kramer**
Universität Tübingen
Tübingen, Germany

**C Krattenthaler**
Universität Wien
Vienna, Austria

**M Krbec**
Academy of Sciences
Prague, Czech Republic

**D Kreimer**
IHES
Bures-sur-Yvette, France

**A Kresch**
University of Warwick
Coventry, UK

**D Kretschmann**
Technische Universität Braunschweig
Braunschweig, Germany

**P B Kronheimer**
Harvard University
Cambridge, MA, USA

**B Kuckert**
Universität Hamburg
Hamburg, Germany

**Y Kuramoto**
Hokkaido University
Sapporo, Japan

**J M F Labastida**
CSIC
Madrid, Spain

**G Labrosse**
Université Paris-Sud XI
Orsay, France

**C Landim**
IMPA, Rio de Janeiro, Brazil and UMR 6085
and Université de Rouen
France

**E Langmann**
KTH Physics
Stockholm, Sweden

**S Laporta**
Università di Parma
Parma, Italy

**O D Lavrentovich**
Kent State University
Kent, OH, USA

**G F Lawler**
Cornell University
Ithaca, NY, USA

**C Le Bris**
CERMICS – ENPC
Champs Sur Marne, France

**A Lesne**
Université P.-M. Curie, Paris VI
Paris, France

**D Levi**
Università "Roma Tre"
Rome, Italy

**J Lewandowski**
Uniwersyte Warszawski
Warsaw, Poland

**R G Littlejohn**
University of California at Berkeley
Berkeley, CA, USA

**R Livi**
Università di Firenze
Sesto Fiorentino, Italy

**R Longoni**
Università di Roma "La Sapienza"
Rome, Italy

**J Lowengrub**
University of California at Irvine
Irvine, USA

**C Lozano**
INTA
Torrejón de Ardoz, Spain

**T T Q Lê**
Georgia Institute of Technology
Atlanta, GA, USA

**B Lucquin-Desreux**
Université P.-M. Curie, Paris VI
Paris, France

**V Lyubashenko**
Institute of Mathematics
Kyiv, Ukraine

**M Lyubich**
University of Toronto
Toronto, ON, Canada and Stony Brook University
NY, USA

**R Léandre**
Université de Bourgogne
Dijon, France

**P Lévay**
Budapest University of Technology and Economics
Budapest, Hungary

**R Maartens**
Portsmouth University
Portsmouth, UK

**N MacKay**
University of York
York, UK

**J Magnen**
Ecole Polytechnique
France

**F Magri**
Università di Milano Bicocca
Milan, Italy

**J Maharana**
Institute of Physics
Bhubaneswar, India

**S Majid**
Queen Mary, University of London
London, UK

**C Marchioro**
Università di Roma "La Sapienza"
Rome, Italy

**K Marciniak**
Linköping University
Norrköping, Sweden

**M Marcolli**
Max-Planck-Institut für Mathematik
Bonn, Germany

**M Mariño**
CERN
Geneva, Switzerland

**J Marklof**
University of Bristol
Bristol, UK

**C-M Marle**
Université P.-M. Curie, Paris VI
Paris, France

**L Mason**
University of Oxford
Oxford, UK

**V Mastropietro**
Università di Roma "Tor Vergata"
Rome, Italy

**V Mathai**
University of Adelaide
Adelaide, SA, Australia

**J Mawhin**
Université Catholique de Louvain
Louvain-la-Neuve, Belgium

**S Mazzucchi**
Università di Trento
Povo, Italy

**B M McCoy**
State University of New York at Stony Brook
Stony Brook, NY, USA

**E Meinrenken**
University of Toronto
Toronto, ON, Canada

**I Melbourne**
University of Surrey
Guildford, UK

**J Mickelsson**
KTH Physics
Stockholm, Sweden

**W P Minicozzi II**
University of New York
New York, NY, USA

**S Miracle-Solé**
Centre de Physique Théorique, CNRS
Marseille, France

**A Miranville**
Université de Poitiers
Chasseneuil, France

**P K Mitter**
Université de Montpellier 2
Montpellier, France

**V Moncrief**
Yale University
New Haven, CT, USA

**Y Morita**
Ryukoku University
Otsu, Japan

**P J Morrison**
University of Texas at Austin
Austin, TX, USA

**J Mund**
Universidade de São Paulo
São Paulo, Brazil

**F Musso**
Università "Roma Tre"
Rome, Italy

**G L Naber**
Drexel University
Philadelphia, PA, USA

**B Nachtergaele**
University of California at Davis
Davis, CA, USA

**C Nash**
National University of Ireland
Maynooth, Ireland

**Š Nečasová**
Academy of Sciences
Prague, Czech Republic

**A I Neishtadt**
Russian Academy of Sciences
Moscow, Russia

**N Neumaier**
Albert-Ludwigs-University in Freiburg
Freiburg, Germany

**S E Newhouse**
Michigan State University
E. Lansing, MI, USA

**C M Newman**
New York University
New York, NY, USA

**S Nikčević**
SANU
Belgrade, Serbia and Montenegro

**M Nitsche**
University of New Mexico
Albuquerque, NM, USA

**R G Novikov**
Université de Nantes
Nantes, France

**J M Nunes da Costa**
Universidade de Coimbra
Coimbra, Portugal

**S O'Brien**
Tyndall National Institute
Cork, Republic of Ireland

**A Okounkov**
Princeton University
Princeton, NJ, USA

**A Onuki**
Kyoto University
Kyoto, Japan

**J-P Ortega**
Université de Franche-Comté
Besançon, France

**H Osborn**
University of Cambridge
Cambridge, UK

**Maciej P Wojtkowski**
University of Arizona
Tucson, AZ, USA and Institute of Mathematics PAN
Warsaw, Poland

**J Palmer**
University of Arizona
Tucson, AZ, USA

**J H Park**
Sungkyunkwan University
Suwon, South Korea

**P E Parker**
Wichita State University
Wichita KS, USA

**S Paycha**
Université Blaise Pascal
Aubière, France

**P A Pearce**
University of Melbourne
Parkville VIC, Australia

**P Pearle**
Hamilton College
Clinton, NY, USA

**M Pedroni**
Università di Bergamo
Dalmine (BG), Italy

**B Pelloni**
University of Reading
UK

**R Penrose**
University of Oxford
Oxford, UK

**A Perez**
Penn State University,
University Park, PA, USA

**J H H Perk**
Oklahoma State University
Stillwater, OK, USA

**T Peternell**
Universität Bayreuth
Bayreuth, Germany

**D Petz**
Budapest University of Technology and Economics
Budapest, Hungary

**M J Pflaum**
Johann Wolfgang Goethe-Universität
Frankfurt, Germany

**B Piccoli**
Istituto per le Applicazioni del Calcolo
Rome, Italy

**C Piquet**
Université P.-M. Curie, Paris VI
Paris, France

**L P Pitaevskii**
Università di Trento
Povo, Italy

**S Pokorski**
Warsaw University
Warsaw, Poland

**E Presutti**
Università di Roma “Tor Vergata”
Rome, Italy

**E Previato**
Boston University
Boston, MA, USA

**B Prinari**
Università degli Studi di Lecce
Lecce, Italy

**J Pullin**
Louisiana State University
Baton Rouge, LA, USA

**M Pulvirenti**
Università di Roma “La Sapienza”
Rome, Italy

**O Ragnisco**
Università “Roma Tre”
Rome, Italy

**P Ramadevi**
Indian Institute of Technology Bombay
Mumbai, India

**S A Ramakrishna**
Indian Institute of Technology
Kanpur, India

**J Rasmussen**
Princeton University
Princeton, NJ, USA

**L Rastelli**
Princeton University
Princeton, NJ, USA

**T S Ratiu**
Ecole Polytechnique Fédérale de Lausanne
Lausanne, Switzerland

**S Rauch-Wojciechowski**
Linköping University
Linköping, Sweden

**K-H Rehren**
Universität Göttingen
Göttingen, Germany

**E Remiddi**
Università di Bologna
Bologna, Italy

**J E Roberts**
Università di Roma “Tor Vergata”
Rome, Italy

**L Rey-Bellet**
University of Massachusetts
Amherst, MA, USA

**R Robert**
Université Joseph Fourier
Saint Martin D’Hères, France

**F A Rogers**
King’s College London
London, UK

**R M S Rosa**
Universidade Federal do Rio de Janeiro
Rio de Janeiro, Brazil

**C Rovelli**
Université de la Méditerranée et Centre
de Physique Théorique
Marseilles, France

**S N M Ruijsenaars**
Centre for Mathematics and Computer Science
Amsterdam, The Netherlands

**F Russo**
Université Paris 13
Villetaneuse, France

**L H Ryder**
University of Kent
Canterbury, UK

**S Sachdev**
Yale University
New Haven, CT, USA

**H Sahlmann**
Universiteit Utrecht
Utrecht, The Netherlands

**M Salmhofer**
Universität Leipzig
Leipzig, Germany

**P M Santini**
Università di Roma “La Sapienza”
Rome, Italy

**A Sarmiento**
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil

**R Sasaki**
Kyoto University
Kyoto, Japan

**A Savage**
University of Toronto
Toronto, ON, Canada

**M Schechter**
University of California at Irvine
Irvine, CA, USA

**D-M Schlingemann**
Technical University of Braunschweig
Braunschweig, Germany

**R Schmid**
Emory University
Atlanta, GA, USA

**G Schneider**
Universität Karlsruhe
Karlsruhe, Germany

**K Schneider**
Université de Provence
Marseille, France

**B Schroer**
Freie Universität Berlin
Berlin, Germany

**T Schücker**
Université de Marseille
Marseille, France

**S Scott**
King’s College London
London, UK

**P Selick**
University of Toronto
Toronto, ON, Canada

**M A Semenov-Tian-Shansky**
Steklov Institute of Mathematics
St. Petersburg, Russia and and Université de Bourgogne
Dijon, France

**A N Sengupta**
Louisiana State University
Baton Rouge LA, USA

**S Serfaty**
New York University
New York, NY, USA

**E R Sharpe**
University of Utah
Salt Lake City, UT, USA

**D Shepelsky**
Institute for Low Temperature Physics and Engineering
Kharkov, Ukraine

**S Shlosman**
Université de Marseille
Marseille, France

**A Siconolfi**
Università di Roma “La Sapienza”
Rome, Italy

**V Sidoravicius**
IMPA
Rio de Janeiro, Brazil

**J A Smoller**
University of Michigan
Ann Arbor MI, USA

**M Socolovsky**
Universidad Nacional Autónoma de México
México DF, México

**J P Solovej**
University of Copenhagen
Copenhagen, Denmark

**A Soshnikov**
University of California at Davis
Davis, CA, USA

**J M Speight**
University of Leeds
Leeds, UK

**H Spohn**
Technische Universität München
Garching, Germany

**J Stasheff**
Lansdale, PA, USA

**D L Stein**
University of Arizona
Tucson, AZ, USA

**K S Stelle**
Imperial College
London, UK

**G Sterman**
Stony Brook University
Stony Brook, NY, USA

**S Stringari**
Università di Trento
Povo, Italy

**S J Summers**
University of Florida
Gainesville, FL, USA

**V S Sunder**
The Institute of Mathematical Sciences
Chennai, India

**Y B Suris**
Technische Universität München
München, Germany

**R J Szabo**
Heriot-Watt University
Edinburgh, UK

**S Tabachnikov**
Pennsylvania State University
University Park, PA, USA

**H Tasaki**
Gakushuin University
Tokyo, Japan

**M E Taylor**
University of North Carolina
Chapel Hill, NC, USA

**R Temam**
Indiana University
Bloomington, IN, USA

**B Temple**
University of California at Davis
Davis, CA, USA

**R P Thomas**
Imperial College
London, UK

**U Tillmann**
University of Oxford
Oxford, UK

**K P Tod**
University of Oxford
Oxford, UK

**J A Toth**
McGill University
Montreal, QC, Canada

**C A Tracy**
University of California at Davis
Davis, CA, USA

**A Trautman**
Warsaw University
Warsaw, Poland

**D Treschev**
Moscow State University
Moscow, Russia

**L Triolo**
Università di Roma “Tor Vergata”
Rome, Italy

**J Troost**
Ecole Normale Supérieure
Paris, France

**Tsou Sheung Tsun**
University of Oxford
Oxford, UK

**V Turaev**
IRMA
Strasbourg, France

**D Ueltschi**
University of Arizona
Tucson, AZ, USA

**A M Uranga**
Consejo Superior de Investigaciones Cientificas
Madrid, Spain

**A Valentini**
Perimeter Institute for Theoretical Physics
Waterloo, ON, Canada

**M Vaugon**
Université P.-M. Curie, Paris VI
Paris, France

**P Di Vecchia**
Nordita
Copenhagen, Denmark

**A F Verbeure**
Institute for Theoretical Physics
KU Leuven, Belgium

**Y Colin de Verdière**
Université de Grenoble 1
Saint-Martin d'Hères, France

**M Viana**
IMPA
Rio de Janeiro, Brazil

**G Vitiello**
Università degli Studi di Salerno
Baronissi (SA), Italy

**D-V Voiculescu**
University of California at Berkeley
Berkeley, CA, USA

**S Waldmann**
Albert-Ludwigs-Universität Freiburg
Freiburg, Germany

**J Wambsganss**
Universität Heidelberg
Heidelberg, Germany

**R S Ward**
University of Durham
Durham, UK

**E Wayne**
Boston University
Boston, MA, USA

**F W Wehrli**
University of Pennsylvania
Philadelphia, PA, USA

**R F Werner**
Technische Universität Braunschweig
Braunschweig, Germany

**H Widom**
University of California at Santa Cruz
Santa Cruz, CA, USA

**C M Will**
Washington University
St. Louis, MO, USA

**N M J Woodhouse**
University of Oxford
Oxford, UK

**Siye Wu**
University of Colorado
Boulder, CO, USA

**V Wünsch**
Friedrich-Schiller-Universität Jena
Jena, Germany

**D R Yafaev**
Université de Rennes
Rennes, France

**M Yamada**
Kyoto University
Kyoto, Japan

**M Yuri**
Hokkaido University
Sapporo, Japan

**D Žubrinić**
University of Zagreb
Zagreb, Croatia

**V Županović**
University of Zagreb
Zagreb, Croatia

**R Zecchina**
International Centre for Theoretical Physics (ICTP)
Trieste, Italy

**S Zelditch**
Johns Hopkins University
Baltimore, MD, USA

**S Zelik**
Universität Stuttgart
Stuttgart, Germany

**S-C Zhang**
Stanford University
Stanford, CA, USA

**M B Ziane**
University of Southern California
Los Angeles, CA, USA

**M R Zirnbauer**
Universität Köln
Köln, Germany

**A Zumpano**
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil

# CONTENTS LIST BY SUBJECT

Location references refer to the volume number and page number (separated by a colon).

## INTRODUCTORY ARTICLES



## PHYSICS SUBJECTS

### Classical Mechanics



### Classical, Conformal and Topological Field Theory



### Condensed Matter and Optics

## Disordered Systems



## Dynamical Systems



## Equilibrium Statistical Mechanics



## Fluid Dynamics

## Gauge Theory



## General Relativity



## Integrable Systems

## M-Theory *see* String Theory and M-Theory

## Nonequilibrium Statistical Mechanics



## Quantum Field Theory

## Quantum Gravity



## Quantum Information and Computation



## Quantum Mechanics



## String Theory and *M*-Theory



## Yang–Mills Theory *see* Gauge Theory

# RELATED MATHEMATICS SUBJECTS

## Algebraic Techniques

## Algebraic Topology



## Complex Geometry



## Differential Geometry



## Discrete Mathematics



## Functional Analysis and Operator Algebras



## Lie Groups and Lie Algebras



## Low Dimensional Geometry

## Noncommutative Geometry



## Ordinary and Partial Differential Equations



## Quantization Methods and Path Integration



## Quantum Groups



## Stochastic Methods



## Symplectic Geometry and Topology

## Variational Techniques

# CONTENTS

## VOLUME 1

## C

## VOLUME 2

### D



### E

## F



## G

## H



# VOLUME 3

## I

## N



## O

# VOLUME 4

## P



## Q

## VOLUME 5

## T



## U



## V



## W



## Y

# Introductory Articles

## Introductory Article: Classical Mechanics

**G Gallavotti**, Università di Roma “La Sapienza,” Rome, Italy



### General Principles

Classical mechanics is a theory of motions of point particles. If $X=(x_1,\ldots,x_n)$ are the particle positions in a Cartesian inertial system of coordinates, the equations of motion are determined by their masses $(m_1,\ldots,m_n)$, $m_j>0$, and by the potential energy of interaction, $V(x_1,\ldots,x_n)$, as

$$m_i\ddot{x}_i=-\partial_{x_i}V(x_1,\ldots,x_n),\quad i=1,\ldots,n \qquad [1]$$

here $x_i=(x_{i1},\ldots,x_{id})$ are coordinates of the $i$th particle and $\partial_{x_i}$ is the gradient $(\partial_{x_{i1}},\ldots,\partial_{x_{id}})$; $d$ is the space dimension (i.e., $d=3$, usually). The potential energy function will be supposed “smooth,” that is, analytic except, possibly, when two positions coincide. The latter exception is necessary to include the important cases of gravitational attraction or, when dealing with electrically charged particles, of Coulomb interaction. A basic result is that if $V$ is bounded below, eqn [1] admits, given initial data $X_0=X(0),\dot{X}_0=\dot{X}(0)$, a unique global solution $t\to X(t)$, $t\in(-\infty,\infty)$; otherwise a solution can fail to be global if and only if, in a finite time, it reaches infinity or a singularity point (i.e., a configuration in which two or more particles occupy the same point: an event called a collision).

In eqn [1], $-\partial_{x_i}V(x_1,\ldots,x_n)$ is the force acting on the points. More general forces are often admitted. For instance, velocity-dependent friction forces: they are not considered here because of their phenomenological nature as models for microscopic phenomena which should also, in principle, be explained in terms of conservative forces (furthermore, even from a macroscopic viewpoint, they are rather incomplete models, as they should be considered together with the important heat generation phenomena that accompany them). Another interesting example of forces not corresponding to a potential are certain velocity-dependent forces like the Coriolis force (which, however, appears only in noninertial frames of reference) and the closely related Lorentz force (in electromagnetism): they could be easily accommodated in the Hamiltonian formulation of mechanics; see Appendix 2.

The action principle states that an equivalent formulation of the eqns [1] is that a motion $t\to X_0(t)$ satisfying [1] during a time interval $[t_1,t_2]$ and leading from $X^1=X_0(t_1)$ to $X^2=X_0(t_2)$, renders stationary the action

$$\mathcal{A}(\{X\})=\int_{t_1}^{t_2}\left(\sum_{i=1}^{n}\frac{1}{2}m_i\dot{X}_i(t)^2-V(X(t))\right)\mathrm{d}t \qquad [2]$$

within the class $\mathcal{M}_{t_1,t_2}(X^1,X^2)$ of smooth (i.e., analytic) “motions” $t\to X(t)$ defined for $t\in[t_1,t_2]$ and leading from $X^1$ to $X^2$.

The function

$$\mathcal{L}(Y,X)=\frac{1}{2}\sum_{i=1}^{n}m_i y_i^2-V(X)\overset{\text{def}}{=}K(Y)-V(X),\quad Y=(y_1,\ldots,y_n)$$

is called the Lagrangian function and the action can be written as

$$\int_{t_1}^{t_2}\mathcal{L}(\dot{X}(t),X(t))\,\mathrm{d}t$$

The quantity $K(\dot{X}(t))$ is called kinetic energy and motions satisfying [1] conserve energy as time $t$ varies, that is,

$$K(\dot{X}(t))+V(X(t))=E=\text{const.} \qquad [3]$$

Hence the action principle can be intuitively thought of as saying that motions proceed by keeping constant the energy, sum of the kinetic and potential energies, while trying to share as evenly as possible their (average over time) contribution to the energy.

In the special case in which $V$ is translation invariant, motions conserve linear momentum $Q\overset{\text{def}}{=}\sum_i m_i\dot{x}_i$; if $V$

is rotation invariant around the origin $O$, motions conserve angular momentum $\boldsymbol{M} \stackrel{\text{def}}{=} \sum_i m_i \boldsymbol{x}_i \wedge \dot{\boldsymbol{x}}_i$, where $\wedge$ denotes the vector product in $\mathbb{R}^d$, that is, it is the tensor $(\boldsymbol{a} \wedge \boldsymbol{b})_{ij} = a_i b_j - b_i a_j, i, j = 1, \ldots, d$: if the dimension $d = 3$ the $\boldsymbol{a} \wedge \boldsymbol{b}$ will be naturally regarded as a vector. More generally, to any continuous symmetry group of the Lagrangian correspond conserved quantities: this is formalized in the *Noether theorem*.

It is convenient to think that the scalar product in $\mathbb{R}^{dn}$ is defined in terms of the ordinary scalar product in $\mathbb{R}^d$, $\boldsymbol{a} \cdot \boldsymbol{b} = \sum_{j=1}^d a_j b_j$, by $(\boldsymbol{v}, \boldsymbol{w}) = \sum_{i=1}^n m_i \boldsymbol{v}_i \cdot \boldsymbol{w}_i$: so that kinetic energy and line element $\mathrm{d}s$ can be written as $K(\dot{\boldsymbol{X}}) = \frac{1}{2}(\dot{\boldsymbol{X}}, \dot{\boldsymbol{X}})$ and $\mathrm{d}s^2 = \sum_{i=1}^n m_i \, \mathrm{d}\boldsymbol{x}_i^2$, respectively. Therefore, the metric generated by the latter scalar product can be called *kinetic energy metric*.

The interest of the kinetic metric appears from the *Maupertuis' principle* (equivalent to [1]): the principle allows us to identify the trajectory traced in $\mathbb{R}^d$ by a motion that leads from $\boldsymbol{X}^1$ to $\boldsymbol{X}^2$ moving with energy $E$. Parametrizing such trajectories as $\tau \to \boldsymbol{X}(\tau)$ by a parameter $\tau$ varying in [0, 1] so that the line element is $\mathrm{d}s^2 = (\partial_\tau \boldsymbol{X}, \partial_\tau \boldsymbol{X}) \, \mathrm{d}\tau^2$, the principle states that the trajectory of a motion with energy $E$ which leads from $\boldsymbol{X}^1$ to $\boldsymbol{X}^2$ makes stationary, among the analytic curves $\boldsymbol{\xi} \in \mathcal{M}_{0,1}(\boldsymbol{X}^1, \boldsymbol{X}^2)$, the function

$$L(\boldsymbol{\xi}) = \int_{\boldsymbol{\xi}} \sqrt{E - V(\boldsymbol{\xi}(s))} \, \mathrm{d}s \qquad [4]$$

so that the possible trajectories traced by the solutions of [1] in $\mathbb{R}^{nd}$ and with energy $E$ can be identified with the geodesics of the metric $\mathrm{d}m^2 \stackrel{\text{def}}{=} (E - V(\boldsymbol{X})) \cdot \mathrm{d}s^2$.

For more details, the reader is referred to Landau and Lifshitz (1976) and Gallavotti (1983).

## Constraints

Often particles are subject to constraints which force the motion to take place on a surface $M \subset \mathbb{R}^{nd}$, i.e., $\boldsymbol{X}(t)$ is forced to be a point on the manifold $M$. A typical example is provided by rigid systems in which motions are subject to forces which keep the mutual distances of the particles constant: $|\boldsymbol{x}_i - \boldsymbol{x}_j| = \rho_{ij}$, with $\rho_{ij}$ time-independent positive quantities. In essentially all cases, the forces that imply constraints, called constraint reactions, are velocity dependent and, therefore, are not in the class of conservative forces considered here, cf. [1]. Hence, from a fundamental viewpoint admitting only conservative forces, constrained systems should be regarded as idealizations of systems subject to conservative forces which approximately imply the constraints.

In general, the $\ell$-dimensional manifold $M$ will not admit a global system of coordinates: however, it will be possible to describe points in the vicinity of any $\boldsymbol{X}^0 \in M$ by using $N = nd$ coordinates $\boldsymbol{q} = (q_1, \ldots, q_\ell, q_{\ell+1}, \ldots, q_N)$ varying in an open ball $B_{\boldsymbol{X}^0}$: $\boldsymbol{X} = \boldsymbol{X}(q_1, \ldots, q_\ell, q_{\ell+1}, \ldots, q_N)$.

The $q$-coordinates can be chosen well adapted to the surface $M$ and to the kinetic metric, i.e., so that the points of $M$ are identified by $q_{\ell+1} = \cdots = q_N = 0$ (which is the meaning of "adapted"); furthermore, infinitesimal displacements $(0, \ldots, 0, \mathrm{d}\varepsilon_{\ell+1}, \ldots, \mathrm{d}\varepsilon_N)$ out of a point $\boldsymbol{X}^0 \in M$ are orthogonal to $M$ (in the kinetic metric) and have a length independent of the position of $\boldsymbol{X}^0$ on $M$ (which is the meaning of "well adapted" to the kinetic metric).

Motions constrained on $M$ arise when the potential $V$ has the form

$$V(\boldsymbol{X}) = V_a(\boldsymbol{X}) + \lambda W(\boldsymbol{X}) \qquad [5]$$

where $W$ is a smooth function which reaches its minimum value, say equal to 0, precisely on the manifold $M$ while $V_a$ is another smooth potential. The factor $\lambda > 0$ is a parameter called the rigidity of the constraint.

A particularly interesting case arises when the level surfaces of $W$ also have the geometric property of being "parallel" to the surface $M$: in the precise sense that the matrix $\partial^2_{q_i q_j} W(\boldsymbol{X})$, $i, j > \ell$ is positive definite and $\boldsymbol{X}$-independent, for all $\boldsymbol{X} \in M$, in a system of coordinates well adapted to the kinetic metric.

A potential $W$ with the latter properties can be called an approximately ideal constraint reaction. In fact, it can be proved that, given an initial datum $\boldsymbol{X}^0 \in M$ with velocity $\dot{\boldsymbol{X}}^0$ tangent to $M$, i.e., given an initial datum whose coordinates in a local system of coordinates are $(\boldsymbol{q}_0, \boldsymbol{0})$ and $(\dot{\boldsymbol{q}}_0, \boldsymbol{0})$ with $\boldsymbol{q}_0 = (q_{01}, \ldots, q_{0\ell})$ and $\dot{\boldsymbol{q}}_0 = (\dot{q}_{01}, \ldots, \dot{q}_{0\ell})$, the motion generated by [1] with $V$ given by [5] is a motion $t \to \boldsymbol{X}_\lambda(t)$ which

1. as $\lambda \to \infty$ tends to a motion $t \to \boldsymbol{X}_\infty(t)$;
2. as long as $\boldsymbol{X}_\infty(t)$ stays in the vicinity of the initial data, say for $0 \le t \le t_1$, so that it can be described in the above local adapted coordinates, its coordinates have the form $t \to (\boldsymbol{q}(t), \boldsymbol{0}) = (q_1(t), \ldots, q_\ell(t), 0, \ldots, 0)$: that is, it is a motion developing on the constraint surface $M$; and
3. the curve $t \to \boldsymbol{X}_\infty(t)$, $t \in [0, t_1]$, as an element of the space $\mathcal{M}_{0,t_1}(\boldsymbol{X}^0, \boldsymbol{X}_\infty(t_1))$ of analytic curves on $M$ connecting $\boldsymbol{X}^0$ to $\boldsymbol{X}_\infty(t_1)$, renders the action

$$A(\boldsymbol{X}) = \int_0^{t_1} \big( K(\dot{\boldsymbol{X}}(t)) - V_a(\boldsymbol{X}(t)) \big) \mathrm{d}t \qquad [6]$$

stationary.

The latter property can be formulated "intrinsically," that is, referring only to $M$ as a surface, via the restriction of the metric $\mathrm{d}s^2$ to line elements $\mathrm{d}\boldsymbol{s} = (\mathrm{d}q_1, \ldots, \mathrm{d}q_\ell, 0, \ldots, 0)$ tangent to $M$ at the point $\boldsymbol{X} = (q_0, 0, \ldots, 0) \in M$; we write $\mathrm{d}s^2 = \sum_{i,j}^{1,\ell} g_{ij}(q) \times \mathrm{d}q_i\,\mathrm{d}q_j$. The $\ell \times \ell$ symmetric positive-definite matrix $g$ can be called the *metric* on $M$ induced by the kinetic energy. Then the action in [6] can be written as

$$\mathcal{A}(q) = \int_0^{t_1} \left( \frac{1}{2} \sum_{i,j}^{1,\ell} g_{ij}(q(t))\dot{q}_i(t)\dot{q}_j(t) - \overline{V}_a(q(t)) \right) \mathrm{d}t \qquad [7]$$

where $\overline{V}_a(q) \stackrel{\text{def}}{=} V_a(\boldsymbol{X}(q_1, \ldots, q_\ell, 0, \ldots, 0))$: the function

$$\mathcal{L}(\boldsymbol{\eta}, q) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{i,j}^{1,\ell} g_{ij}(q)\eta_i\eta_j - \overline{V}_a(q) \equiv \frac{1}{2} g(q)\boldsymbol{\eta} \cdot \boldsymbol{\eta} - \overline{V}_a(q) \qquad [8]$$

is called the constrained Lagrangian of the system.

An important property is that the constrained motions conserve the energy defined as $E = \frac{1}{2}(g(q)\dot{q}, \dot{q}) + \overline{V}_a(q)$; see next section.

The constrained motion $\boldsymbol{X}_\infty(t)$ of energy $E$ satisfies the Maupertuis' principle in the sense that the curve on $M$ on which the motion develops renders

$$L(\boldsymbol{\xi}) = \int_{\boldsymbol{\xi}} \sqrt{E - V_a(\boldsymbol{\xi}(s))}\, \mathrm{d}s \qquad [9]$$

stationary among the (smooth) curves that develop on $M$ connecting two fixed values $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. In the particular case in which $\ell = n$ this is again Maupertuis' principle for unconstrained motions under the potential $V(\boldsymbol{X})$. In general, $\ell$ is called the number of degrees of freedom because a complete description of the initial data requires $2\ell$ coordinates $q(0), \dot{q}(0)$.

If $W$ is minimal on $M$ but the condition on $W$ of having level surfaces parallel to $M$ is not satisfied, i.e., if $W$ is not an approximate ideal constraint reaction, it still remains true that the limit motion $\boldsymbol{X}_\infty(t)$ takes place on $M$. However, in general, it will not satisfy the above variational principles. For this reason, motions arising as limits (as $\lambda \to \infty$) of motions developing under the potential [5] with $W$ having minimum on $M$ and level curves parallel (in the above sense) to $M$ are called ideally constrained motions or motions subject by ideal constraints to the surface $M$.

As an example, suppose that $W$ has the form $W(\boldsymbol{X}) = \sum_{i,j \in \mathcal{P}} w_{ij}(|\boldsymbol{x}_i - \boldsymbol{x}_j|)$ with $w_{ij}(|\boldsymbol{\xi}|) \geq 0$ an analytic function vanishing only when $|\boldsymbol{\xi}| = \rho_{ij}$ for $i$, $j$ in some set of pairs $\mathcal{P}$ and for some given distances $\rho_{ij}$ (e.g., $w_{ij}(\boldsymbol{\xi}) = (\boldsymbol{\xi}^2 - \rho_{ij}^2)^2\gamma, \gamma > 0$). Then $W$ can be shown to satisfy the mentioned conditions and therefore, the so constrained motions $\boldsymbol{X}_\infty(t)$ of the body satisfy the variational principles mentioned in connection with [7] and [9]: in other words, the above natural way of realizing a rather general rigidity constraint is ideal.

The modern viewpoint on the physical meaning of the constraint reactions is as follows: looking at motions in an inertial Cartesian system, it will appear that the system is subject to the applied forces with potential $V_a(\boldsymbol{X})$ and to constraint forces which are defined as the differences $\boldsymbol{R}_i = m_i\ddot{\boldsymbol{x}}_i + \boldsymbol{\partial}_{\boldsymbol{x}_i} V_a(\boldsymbol{X})$. The latter reflect the action of the forces with potential $\lambda W(\boldsymbol{X})$ in the limit of infinite rigidity ($\lambda \to \infty$).

In applications, sometimes the action of a constraint can be regarded as ideal: the motion will then verify the variational principles mentioned and $\boldsymbol{R}$ can be computed as the differences between the $m_i\ddot{\boldsymbol{x}}_i$ and the active forces $-\boldsymbol{\partial}_{\boldsymbol{x}_i} V_a(\boldsymbol{X})$. In dynamics problems it is, however, a very difficult and important matter, particularly in engineering, to judge whether a system of particles can be considered as subject to ideal constraints: this leads to important decisions in the construction of machines. It simplifies the calculations of the reactions and fatigue of the materials but a misjudgment can have serious consequences about stability and safety. For statics problems, the difficulty is of lower order: usually assuming that the constraint reaction is ideal leads to an overestimate of the requirements for stability of equilibria. Hence, employing the action principle to statics problems, where it constitutes the principle of *virtual work*, generally leads to economic problems rather than to safety issues. Its discovery even predates Newtonian mechanics.

We refer the reader to Arnol'd (1989) and Gallavotti (1983) for more details.

## Lagrange and Hamilton Forms of the Equations of Motion

The stationarity condition for the action $\mathcal{A}(q)$, cf. [7], [8], is formulated in terms of the Lagrangian $\mathcal{L}(\boldsymbol{\eta}, \boldsymbol{\xi})$, see [8], by

$$\frac{\mathrm{d}}{\mathrm{d}t} \partial_{\eta_i} \mathcal{L}(\dot{q}(t), q(t)) = \partial_{\xi_i} \mathcal{L}(\dot{q}(t), q(t)), \quad i = 1, \ldots, \ell \qquad [10]$$

which is a second-order differential equation called the Lagrangian equation of motion. It can be cast in "normal form": for this purpose, adopting the convention of "summation over repeated indices," introduce the "generalized momenta"

$$p_i \stackrel{\text{def}}{=} g(q)_{ij}\dot{q}_j, \quad i = 1, \ldots, \ell \qquad [11]$$

Since $g(q) > 0$, the motions $t \to q(t)$ and the corresponding velocities $t \to \dot{q}(t)$ can be described equivalently by $t \to (q(t), p(t))$: and the equations of motion [10] become the first-order equations

$$\dot{q}_i = \partial_{p_i} \mathcal{H}(p, q), \qquad \dot{p}_i = -\partial_{q_i} H(p, q) \qquad [12]$$

where the function $\mathcal{H}$, called the Hamiltonian of the system, is defined by

$$\mathcal{H}(p, q) \stackrel{\text{def}}{=} \tfrac{1}{2}(g(q)^{-1} p, p) + \overline{V}_a(q) \qquad [13]$$

Equations [12], regarded as equations of motion for phase space points $(p, q)$, are called Hamilton equations. In general, $q$ are local coordinates on $M$ and motions are specified by giving $q, \dot{q}$ or $p, q$.

Looking for a coordinate-free representation of motions consider the pairs $X$, $Y$ with $X \in M$ and $Y$ a vector $Y \in T_X$ tangent to $M$ at the point $X$. The collection of pairs $(Y, X)$ is denoted $T(M) = \cup_{X \in M} (T_X \times \{X\})$ and a motion $t \to (\dot{X}(t), X(t)) \in T(M)$ in local coordinates is represented by $(\dot{q}(t), q(t))$. The space $T(M)$ can be called the space of initial data for Lagrange's equations of motion: it has $2\ell$ dimensions (also known as the "tangent bundle" of $M$).

Likewise, the space of initial data for the Hamilton equations will be denoted $T^*(M)$ and it consists of pairs $X, P$ with $X \in M$ and $P = g(X)Y$ with $Y$ a vector tangent to $M$ at $X$. The space $T^*(M)$ is called the phase space of the system: it has $2\ell$ dimensions (and it is occasionally called the "cotangent bundle" of $M$).

Immediate consequence of [12] is

$$\frac{d}{dt} \mathcal{H}(p(t), q(t)) \equiv 0$$

and it means that $\mathcal{H}(p(t), q(t))$ is constant along the solutions of [12]. Noting that $\mathcal{H}(p, q) = (1/2)(g(q) \;\; \dot{q}, \dot{q}) + \overline{V}_a(q)$ is the sum of the kinetic and potential energies, it follows that the conservation of $\mathcal{H}$ along solutions means energy conservation in presence of ideal constraints.

Let $S_t$ be the flow generated on the phase space variables $(p, q)$ by the solutions of the equations of motion [12], that is, let $t \to S_t(p, q) \equiv (p(t), q(t))$ denote a solution of [12] with initial data $(p, q)$. Then a (measurable) set $\Delta$ in phase space evolves in time $t$ into a new set $S_t \Delta$ with the same volume: this is obvious because the Hamilton equations [12] have manifestly zero divergence ("Liouville's theorem").

The Hamilton equations also satisfy a variational principle, called the Hamilton action principle: that is, if $\mathcal{M}_{t_1, t_2}((p_1, q_1), (p_2, q_2); M)$ denotes the space of the analytic functions $\varphi : t \to (\pi(t), \kappa(t))$ which in the time interval $[t_1, t_2]$ lead from $(p_1, q_1)$ to $(p_2, q_2)$, then the condition that $\varphi_0(t) = (p(t), q(t))$ satisfies [12] can be equivalently formulated by requiring that the function

$$\mathcal{A}_{\mathcal{H}}(\varphi) \stackrel{\text{def}}{=} \int_{t_1}^{t_2} \Big( \pi(t) \cdot \dot{\kappa}(t) - \mathcal{H}(\pi(t), \kappa(t)) \Big) \, dt \qquad [14]$$

be stationary for $\varphi = \varphi_0$: in fact, eqns [12] are the stationarity conditions for the Hamilton action [14] on $\mathcal{M}_{t_0, t_1}((p_1, q_1), (p_2, q_2); M)$. And, since the derivatives of $\pi(t)$ do not appear in [14], stationarity is even achieved in the larger space $\mathcal{M}_{t_1, t_2}(q_1, q_2; M)$ of the motions $\varphi : t \to (\pi(t), \kappa(t))$ leading from $q_1$ to $q_2$ without any restriction on the initial and final momenta $p_1, p_2$ (which, therefore, cannot be prescribed *a priori* independently of $q_1, q_2$). If the prescribed data $p_1, q_1, p_2, q_2$ are not compatible with the equations of motion (e.g., $H(p_1, q_2) \neq H(p_2, q_2)$), then the action functional has no stationary trajectory in $\mathcal{M}_{t_1, t_2}((p_1, q_1), (p_2 q_2); M)$.

For more details, the reader is referred to Landau and Lifshitz (1976), Arnol'd (1989), and Gallavotti (1983).

## Canonical Transformations of Phase Space Coordinates

The Hamiltonian form, [13], of the equations of motion turns out to be quite useful in several problems. It is, therefore, important to remark that it is invariant under a special class of transformations of coordinates, called canonical transformations.

Consider a local change of coordinates on phase space, i.e., a smooth, smoothly invertible map $\mathcal{C}(\pi, \kappa) = (\pi', \kappa')$ between an open set $U$ in the phase space of a Hamiltonian system with $\ell$ degrees of freedom, into an open set $U'$ in a $2\ell$-dimensional space. The change of coordinates is said to be *canonical* if for any solution $t \to (\pi(t), \kappa(t))$ of equations like [12], for any Hamiltonian $\mathcal{H}(\pi, \kappa)$ defined on $U$, the $\mathcal{C}$–image $t \to (\pi'(t), \kappa'(t)) = \mathcal{C}(\pi(t), \kappa(t))$ is a solution of [12] with the "same" Hamiltonian, that is, with Hamiltonian $\mathcal{H}'(\pi', \kappa') \stackrel{\text{def}}{=} \mathcal{H}(\mathcal{C}^{-1}(\pi', \kappa'))$.

The condition that a transformation of coordinates is canonical is obtained by using the arbitrariness of the function $\mathcal{H}$ and is simply expressed as a necessary and sufficient property of the Jacobian $L$,

$$\begin{gathered} L = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \\ A_{ij} = \partial_{\pi_j} \pi'_i, \quad B_{ij} = \partial_{\kappa_j} \pi'_i, \\ C_{ij} = \partial_{\pi_j} \kappa'_i, \quad D_{ij} = \partial_{\kappa_j} \kappa'_i \end{gathered} \qquad [15]$$

where $i, j = 1, \ldots, \ell$. Let

$$E = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

denote the $2\ell \times 2\ell$ matrix formed by four $\ell \times \ell$ blocks, equal to the 0 matrix or, as indicated, to the $\pm$(identity matrix); then, if a superscript T denotes matrix transposition, the condition that the map be canonical is that

$$L^{-1} = EL^{\mathrm{T}}E^{\mathrm{T}} \text{ or } L^{-1} = \begin{pmatrix} D^{\mathrm{T}} & -B^{\mathrm{T}} \\ -C^{\mathrm{T}} & A^{\mathrm{T}} \end{pmatrix} \qquad [16]$$

which immediately implies that $\det L = \pm 1$. In fact, it is possible to show that [16] implies $\det L = 1$. Equation [16] is equivalent to the four relations $AD^{\mathrm{T}} - BC^{\mathrm{T}} = 1$, $-AB^{\mathrm{T}} + BA^{\mathrm{T}} = 0$, $CD^{\mathrm{T}} - DC^{\mathrm{T}} = 0$, and $-CB^{\mathrm{T}} + DA^{\mathrm{T}} = 1$. More explicitly, since the first and the fourth relations coincide, these can be expressed as

$$\{\pi'_i, \kappa'_j\} = \delta_{ij}, \ \{\pi'_i, \pi'_j\} = 0, \ \{\kappa'_i, \kappa'_j\} = 0 \qquad [17]$$

where, for any two functions $F(\boldsymbol{\pi}, \boldsymbol{\kappa})$, $G(\boldsymbol{\pi}, \boldsymbol{\kappa})$, the *Poisson bracket* is

$$\{F, G\}(\boldsymbol{\pi}, \boldsymbol{\kappa}) \stackrel{\text{def}}{=} \sum_{k=1}^{\ell} \big(\partial_{\pi_k} F(\boldsymbol{\pi}, \boldsymbol{\kappa}) \partial_{\kappa_k} G(\boldsymbol{\pi}, \boldsymbol{\kappa}) - \partial_{\kappa_k} F(\boldsymbol{\pi}, \boldsymbol{\kappa}) \, \partial_{\pi_k} G(\boldsymbol{\pi}, \boldsymbol{\kappa})\big) \qquad [18]$$

The latter satisfies *Jacobi's identity*: $\{\{F, G\}, Q\} + \{\{G, Q\}, F\} + \{\{Q, F\}, G\} = 0$, for any three functions $F, G, Q$ on the phase space. It is quite useful to remark that if $t \to (\boldsymbol{p}(t), \ \boldsymbol{q}(t)) = S_t(\boldsymbol{p}, \boldsymbol{q})$ is a solution to Hamilton equations with Hamiltonian $\mathcal{H}$ then, given any observable $F(\boldsymbol{p}, \boldsymbol{q})$, it "evolves" as $F(t) \stackrel{\text{def}}{=} F(\boldsymbol{p}(t), \boldsymbol{q}(t))$ satisfying

$$\partial_t F(\boldsymbol{p}(t), \ \boldsymbol{q}(t)) = \{\mathcal{H}, \ F\}(\boldsymbol{p}(t), \ \boldsymbol{q}(t))$$

Requiring the latter identity to hold for all observables $F$ is equivalent to requiring that the $t \to (\boldsymbol{p}(t), \boldsymbol{q}(t))$ be a solution of Hamilton's equations for $\mathcal{H}$.

Let $\mathcal{C} : U \longleftrightarrow U'$ be a smooth, smoothly invertible transformation between two open $2\ell$-dimensional sets: $\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\kappa}) = (\boldsymbol{\pi}', \boldsymbol{\kappa}')$. Suppose that there is a function $\Phi(\boldsymbol{\pi}', \boldsymbol{\kappa})$ defined on a suitable domain $W$ such that

$$\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\kappa}) = (\boldsymbol{\pi}', \boldsymbol{\kappa}') \ \Rightarrow \ \begin{cases} \boldsymbol{\pi} = \partial_{\boldsymbol{\kappa}} \Phi(\boldsymbol{\pi}', \boldsymbol{\kappa}) \\ \boldsymbol{\kappa}' = \partial_{\boldsymbol{\pi}'} \Phi(\boldsymbol{\pi}', \boldsymbol{\kappa}) \end{cases} \qquad [19]$$

then $\mathcal{C}$ is canonical. This is because [19] implies that if $\boldsymbol{\kappa}, \boldsymbol{\pi}'$ are varied and if $\boldsymbol{\pi}, \boldsymbol{\kappa}', \boldsymbol{\pi}', \boldsymbol{\kappa}$ are related by $\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\kappa}) = (\boldsymbol{\pi}', \boldsymbol{\kappa}')$, then $\boldsymbol{\pi} \cdot \mathrm{d}\boldsymbol{\kappa} + \boldsymbol{\kappa}' \cdot \mathrm{d}\boldsymbol{\pi}' = \mathrm{d}\Phi(\boldsymbol{\pi}', \boldsymbol{\kappa})$, which implies that

$$\boldsymbol{\pi} \cdot \mathrm{d}\boldsymbol{\kappa} - \mathcal{H}(\boldsymbol{\pi}, \boldsymbol{\kappa})\mathrm{d}t \equiv \boldsymbol{\pi}' \cdot \mathrm{d}\boldsymbol{\kappa}' - \mathcal{H}(\mathcal{C}^{-1}(\boldsymbol{\pi}', \boldsymbol{\kappa}'))\mathrm{d}t + \mathrm{d}\Phi(\boldsymbol{\pi}', \boldsymbol{\kappa}) - \mathrm{d}(\boldsymbol{\pi}' \cdot \boldsymbol{\kappa}') \qquad [20]$$

It means that the Hamiltonians $\mathcal{H}(\boldsymbol{p}, \boldsymbol{q})$ and $\mathcal{H}'(\boldsymbol{p}', \boldsymbol{q}')) \stackrel{\text{def}}{=} \mathcal{H}(\mathcal{C}^{-1}(\boldsymbol{p}', \boldsymbol{q}'))$ have Hamilton actions $\mathcal{A}_{\mathcal{H}}$ and $\mathcal{A}_{\mathcal{H}'}$ differing by a constant, if evaluated on corresponding motions $(\boldsymbol{p}(t), \boldsymbol{q}(t))$ and $(\boldsymbol{p}'(t), \boldsymbol{q}'(t)) = \mathcal{C}(\boldsymbol{p}(t), \boldsymbol{q}(t))$.

The constant depends only on the initial and final values $(\boldsymbol{p}(t_1), \boldsymbol{q}(t_1))$ and $(\boldsymbol{p}(t_2), \boldsymbol{q}(t_2))$ and, respectively, $(\boldsymbol{p}'(t_1), \boldsymbol{q}'(t_1))$ and $(\boldsymbol{p}'(t_2), \boldsymbol{q}'(t_2))$ so that if $(\boldsymbol{p}(t), \boldsymbol{q}(t))$ makes $\mathcal{A}_{\mathcal{H}}$ extreme, then $(\boldsymbol{p}'(t), \boldsymbol{q}'(t)) = \mathcal{C}(\boldsymbol{p}(t), \boldsymbol{q}(t))$ also makes $\mathcal{A}_{\mathcal{H}'}$ extreme.

Hence, if $t \to (\boldsymbol{p}(t), \boldsymbol{q}(t))$ solves the Hamilton equations with Hamiltonian $\mathcal{H}(\boldsymbol{p}, \boldsymbol{q})$ then the motion $t \to (\boldsymbol{p}'(t), \boldsymbol{q}'(t)) = \mathcal{C}(\boldsymbol{p}(t), \boldsymbol{q}(t))$ solves the Hamilton equations with Hamiltonian $\mathcal{H}'(\boldsymbol{p}', \boldsymbol{q}') = \mathcal{H}(\mathcal{C}^{-1}(\boldsymbol{p}', \boldsymbol{q}'))$ no matter which it is: therefore, the transformation is canonical. The function $\Phi$ is called its generating function.

Equation [19] provides a way to construct canonical maps. Suppose that a function $\Phi(\boldsymbol{\pi}', \boldsymbol{\kappa})$ is given and defined on some domain $W$; then setting

$$\begin{cases} \boldsymbol{\pi} = \partial_{\boldsymbol{\kappa}} \Phi(\boldsymbol{\pi}', \boldsymbol{\kappa}) \\ \boldsymbol{\kappa}' = \partial_{\boldsymbol{\pi}'} \Phi(\boldsymbol{\pi}', \boldsymbol{\kappa}) \end{cases}$$

and inverting the first equation in the form $\boldsymbol{\pi}' = \Xi(\boldsymbol{\pi}, \boldsymbol{\kappa})$ and substituting the value for $\boldsymbol{\pi}'$ thus obtained, in the second equation, a map $\mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\kappa}) = (\boldsymbol{\pi}', \boldsymbol{\kappa}')$ is defined on some domain (where the mentioned operations can be performed) and if such domain is open and not empty then $\mathcal{C}$ is a canonical map.

For similar reasons, if $\Gamma(\boldsymbol{\kappa}, \boldsymbol{\kappa}')$ is a function defined on some domain then setting $\boldsymbol{\pi} = \partial_{\boldsymbol{\kappa}} \Gamma(\boldsymbol{\kappa}, \boldsymbol{\kappa}')$, $\boldsymbol{\pi}' = -\partial_{\boldsymbol{\kappa}'} \Gamma(\boldsymbol{\kappa}, \boldsymbol{\kappa}')$ and solving the first relation to express $\boldsymbol{\kappa}' = \boldsymbol{\Delta}(\boldsymbol{\pi}, \boldsymbol{\kappa})$ and substituting in the second relation a map $(\boldsymbol{\pi}', \boldsymbol{\kappa}') = \mathcal{C}(\boldsymbol{\pi}, \boldsymbol{\kappa})$ is defined on some domain (where the mentioned operations can be performed) and if such domain is open and not empty then $\mathcal{C}$ is a canonical map.

Likewise, canonical transformations can be constructed starting from *a priori* given functions $F(\boldsymbol{\pi}, \boldsymbol{\kappa}')$ or $G(\boldsymbol{\pi}, \boldsymbol{\pi}')$. And the most general canonical map can be generated locally (i.e., near a given point in phase space) by a single one of the above four ways, possibly composed with a few "trivial" canonical maps in which one pair of coordinates $(\pi_i, \kappa_i)$ is transformed into $(-\kappa_i, \pi_i)$. The necessity of also including the trivial maps can be traced to the existence of homogeneous canonical maps, that is, maps such that $\boldsymbol{\pi} \cdot \mathrm{d}\boldsymbol{\kappa} = \boldsymbol{\pi}' \cdot \mathrm{d}\boldsymbol{\kappa}'$ (e.g., the identity map, see below or [49] for nontrivial examples) which are action preserving hence canonical, but which evidently cannot be generated by a function $\Phi(\boldsymbol{\kappa}, \boldsymbol{\kappa}')$ although they can be generated by a function depending on $\boldsymbol{\pi}', \boldsymbol{\kappa}$.

Simple examples of homogeneous canonical maps are maps in which the coordinates $\boldsymbol{q}$ are changed into $\boldsymbol{q}' = \boldsymbol{R}(\boldsymbol{q})$ and, correspondingly, the $\boldsymbol{p}$'s are transformed as $\boldsymbol{p}' = (\partial_q \boldsymbol{R}(\boldsymbol{q}))^{-1\,\mathrm{T}}\boldsymbol{p}$, linearly: indeed, this map is generated by the function $F(\boldsymbol{p}', \boldsymbol{q}) \stackrel{\text{def}}{=} \boldsymbol{p}' \cdot \boldsymbol{R}(\boldsymbol{q})$.

For instance, consider the map "Cartesian–polar" coordinates $(q_1, q_2) \longleftrightarrow (\rho, \theta)$ with $(\rho, \theta)$ the polar coordinates of $\boldsymbol{q}$ (namely $\rho = \sqrt{q_1^2 + q_2^2}, \theta =$ arctan $(q_2/q_1)$) and let $\boldsymbol{n} \stackrel{\text{def}}{=} \boldsymbol{q}/|\boldsymbol{q}| = (n_1, n_2)$ and $\boldsymbol{t} = (-n_2, n_1)$. Setting $p_\rho \stackrel{\text{def}}{=} \boldsymbol{p}\cdot\boldsymbol{n}, p_\theta \stackrel{\text{def}}{=} \rho\boldsymbol{p}\cdot\boldsymbol{t}$, the map $(p_1, p_2, q_1, q_2) \longleftrightarrow (p_\rho, p_\theta, \rho, \theta)$ is homogeneous canonical (because $\boldsymbol{p}\cdot \mathrm{d}\boldsymbol{q} = \boldsymbol{p}\cdot\boldsymbol{n}\,\mathrm{d}\rho + \boldsymbol{p}\cdot\boldsymbol{t}\rho\,\mathrm{d}\theta = p_\rho\,\mathrm{d}\rho + p_\theta\,\mathrm{d}\theta$).

As a further example, any area-preserving map $(p, q) \longleftrightarrow (p', q')$ defined on an open region of the plane $\mathbb{R}^2$ is canonical: because in this case the matrices $A, B, C, D$ are just numbers, which satisfy $AD - BC = 1$ and, therefore, [16] holds.

For more details, the reader is referred to Landau and Lifshitz (1976) and Gallavotti (1983).

## Quadratures

The simplest mechanical systems are integrable by quadratures. For instance, the Hamiltonian on $\mathbb{R}^2$,

$$\mathcal{H}(p, q) = \frac{1}{2m}p^2 + V(q) \qquad [21]$$

generates a motion $t \to q(t)$ with initial data $q_0, \dot{q}_0$ such that $\mathcal{H}(p_0, q_0) = E$, i.e., $\frac{1}{2}m\dot{q}_0^2 + V(q_0) = E$, satisfying

$$\dot{q}(t) = \pm\sqrt{\frac{2}{m}(E - V(q(t)))}$$

If the equation $E = V(q)$ has only two solutions $q_-(E) < q_+(E)$ and $|\partial_q V(q_\pm(E))| > 0$, the motion is periodic with period

$$T(E) = 2\int_{q_-(E)}^{q_+(E)} \frac{\mathrm{d}x}{\sqrt{(2/m)(E - V(x))}} \qquad [22]$$

The special solution with initial data $q_0 = q_-(E), \dot{q}_0 = 0$ will be denoted $Q(t)$, and it is an analytic function (by the general regularity theorem on ordinary differential equations). For $0 \le t \le T/2$ or for $T/2 \le t \le T$ it is given, respectively, by

$$t = \int_{q_-(E)}^{Q(t)} \frac{\mathrm{d}x}{\sqrt{(2/m)(E - V(x))}} \qquad [23a]$$

or

$$t = \frac{T}{2} - \int_{q_+(E)}^{Q(t)} \frac{\mathrm{d}x}{\sqrt{(2/m)(E - V(x))}} \qquad [23b]$$

The most general solution with energy $E$ has the form $q(t) = Q(t_0 + t)$, where $t_0$ is defined by $q_0 = Q(t_0)$, $\dot{q}_0 = \dot{Q}(t_0)$, i.e., it is the time needed for the "standard solution" $Q(t)$ to reach the initial data for the new motion.

If the derivative of $V$ vanishes in one of the extremes or if at least one of the two solutions $q_\pm(E)$ does not exist, the motion is not periodic and it may be unbounded: nevertheless, it is still expressible via integrals of the type [22]. If the potential $V$ is periodic in $q$ and the variable $q$ is considered to be varying on a circle then essentially all solutions are periodic: exceptions can occur if the energy $E$ has a value such that $V(q) = E$ admits a solution where $V$ has zero derivative.

Typical examples are the harmonic oscillator, the pendulum, and the Kepler oscillator: whose Hamiltonians, if $m, \omega, g, h, G, k$ are positive constants, are, respectively,

$$\begin{aligned} &\frac{p^2}{2m} + \frac{1}{2}m\omega^2 q^2 \\ &\frac{p^2}{2m} + mg\left(1 - \cos\frac{q}{h}\right) \\ &\frac{p^2}{2m} - mk\frac{1}{|q|} + m\frac{G^2}{2q^2} \end{aligned} \qquad [24]$$

the Kepler oscillator Hamiltonian has a potential which is singular at $q = 0$ but if $G \ne 0$ the energy conservation forbids too close an approach to $q = 0$ and the singularity becomes irrelevant.

The integral in [23] is called a *quadrature* and the systems in [21] are therefore *integrable by quadratures*. Such systems, at least when the motion is periodic, are best described in new coordinates in which periodicity is more manifest. Namely when $V(q) = E$ has only two roots $q_\pm(E)$ and $\mp V'(q_\pm(E)) > 0$ the *energy–time coordinates* can be used by replacing $q, \dot{q}$ or $p, q$ by $E, \tau$, where $\tau$ is the time needed for the standard solution $t \to Q(t)$ to reach the given data, that is, $Q(\tau) = q, \dot{Q}(\tau) = \dot{q}$. In such coordinates, the motion is simply $(E, \tau) \to (E, \tau + t)$ and, of course, the variable $\tau$ has to be regarded as varying on a circle of radius $T/2\pi$. The $E, \tau$ variables are a kind of polar coordinates, as can be checked by drawing the curves of constant $E$, "energy levels," in the plane $p, q$ in the cases in [24]; see **Figure 1**.

In the harmonic oscillator case, all trajectories are periodic. In the pendulum case, all motions are periodic except the ones which separate the oscillatory motions (the closed curves in the second drawing) from the rotatory motions (the apparently open curves) which, in fact, are on closed curves as well if the $q$ coordinate, that is, the vertical

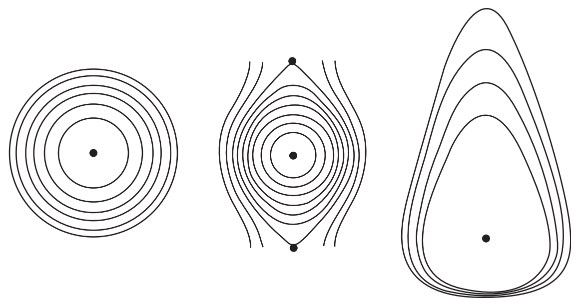


**Figure 1** The energy levels of the harmonic oscillator, the pendulum, and the Kepler motion.

coordinate in Figure 1, is regarded as "periodic" with period $2\pi h$. In the Kepler case, only the negative-energy trajectories are periodic and a few of them are drawn in Figure 1. The single dots represent the equilibrium points in phase space.

The region of phase space where motions are periodic is a set of points $(p, q)$ with the topological structure of $\cup_{u \in U}(\{u\} \times C_u)$, where $u$ is a coordinate varying in an open interval $U$ (e.g., the set of values of the energy), and $C_u$ is a closed curve whose points $(p, q)$ are identified by a coordinate (e.g., by the time necessary for an arbitrarily fixed datum with the same energy to evolve into $(p, q)$).

In the above cases, [24], if the "radial" coordinate is chosen to be the energy the set $U$ is the interval $(0, +\infty)$ for the harmonic oscillator, $(0, 2mg)$ or $(2mg, +\infty)$ for the pendulum, and $(-\frac{1}{2}mk^2/G^2, 0)$ in the Kepler case. The fixed datum for the reference motion can be taken, in all cases, to be of the form $(0, q_0)$ with the time coordinate $t_0$ given by [23].

It is remarkable that the energy–time coordinates are canonical coordinates: for instance, in the vicinity of $(p_0, q_0)$ and if $p_0 > 0$, this can be seen by setting

$$S(q, E) = \int_{q_0}^{q} \sqrt{2m(E - V(x))}\mathrm{d}x \qquad [25]$$

and checking that $p = \partial_q S(q, E)$, $t = \partial_E S(q, E)$ are identities if $(p, q)$ and $(E, t)$ are coordinates for the same point so that the criterion expressed by [20] applies.

It is convenient to standardize the coordinates by replacing the time variable by an angle $\alpha = (2\pi/T(E))t$; and instead of the energy any invertible function of it can be used.

It is natural to look for a coordinate $A = A(E)$ such that the map $(p, q) \longleftrightarrow (A, \alpha)$ is a canonical map: this is easily done as the function

$$\hat{S}(q, A) = \int_{q_0}^{q} \sqrt{2m(E(A) - V(x))}\,\mathrm{d}x \qquad [26]$$

generates (locally) the correspondence between $p = \sqrt{2m(E(A) - V(q))}$ and

$$\alpha = E'(A) \int_0^q \frac{\mathrm{d}x}{\sqrt{2m^{-1}(E(A) - V(x))}}$$

Therefore, by the criterion [20], if

$$E'(A) = \frac{2\pi}{T(E(A))}$$

i.e., if $A'(E) = T(E)/2\pi$, the coordinates $(A, \alpha)$ will be canonical coordinates. Hence, by [22], $A(E)$ can be taken as

$$A = \frac{1}{2\pi} 2 \int_{q_-(E)}^{q_+(E)} \sqrt{2m(E - V(q))}\mathrm{d}q \equiv \frac{1}{2\pi} \oint p\,\mathrm{d}q \qquad [27]$$

where the last integral is extended to the closed curve of energy $E$; see Figure 1. The *action–angle coordinates* $(A, \alpha)$ are defined in open regions of phase space covered by periodic motions: in action–angle coordinates such regions have the form $W = J \times \mathbb{T}$ of a product of an open interval $J$ and a one-dimensional "torus" $\mathbb{T} = [0, 2\pi]$ (i.e., a unit circle).

For details, the reader is again referred to Landau and Lifshitz (1976), Arnol'd (1989), and Gallavotti (1983).

## Quasiperiodicity and Integrability

A Hamiltonian is called *integrable* in an open region $W \subset T^*(M)$ of phase space if

1. there is an analytic and nonsingular (i.e., with nonzero Jacobian) change of coordinates $(\boldsymbol{p}, \boldsymbol{q}) \longleftrightarrow (\boldsymbol{I}, \boldsymbol{\varphi})$ mapping $W$ into a set of the form $\mathcal{I} \times \mathbb{T}^\ell$ with $\mathcal{I} \subset \mathbb{R}^\ell$ (open); and furthermore
2. the flow $t \to S_t(\boldsymbol{p}, \boldsymbol{q})$ on phase space is transformed into $(\boldsymbol{I}, \boldsymbol{\varphi}) \to (\boldsymbol{I}, \boldsymbol{\varphi} + \boldsymbol{\omega}(\boldsymbol{I})t)$ where $\boldsymbol{\omega}(\boldsymbol{I})$ is a smooth function on $\mathcal{I}$.

This means that, in suitable coordinates, which can be called "integrating coordinates," the system appears as a set of $\ell$ points with coordinates $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_\ell)$ moving on a unit circle at angular velocities $\boldsymbol{\omega}(\boldsymbol{I}) = (\omega_1(\boldsymbol{I}), \ldots, \omega_\ell(\boldsymbol{I}))$ depending on the actions of the initial data.

A system integrable in a region $W$ which, in integrating coordinates $\boldsymbol{I}, \boldsymbol{\varphi}$, has the form $\mathcal{I} \times \mathbb{T}^\ell$ is said to be *anisochronous* if $\det \partial_{\boldsymbol{I}} \boldsymbol{\omega}(\boldsymbol{I}) \neq 0$. It is said to be isochronous if $\boldsymbol{\omega}(\boldsymbol{I}) \equiv \boldsymbol{\omega}$ is independent of $\boldsymbol{I}$. The motions of integrable systems are called *quasiperiodic* with frequency spectrum $\boldsymbol{\omega}(\boldsymbol{I})$, or with *frequencies* $\boldsymbol{\omega}(\boldsymbol{I})/2\pi$, in the coordinates $(\boldsymbol{I}, \boldsymbol{\varphi})$.

Clearly, an integrable system admits $\ell$ independent constants of motion, the $\boldsymbol{I} = (I_1, \ldots, I_\ell)$, and, for each

choice of $\boldsymbol{I}$, the other coordinates vary on a "standard" $\ell$-dimensional torus $\mathbb{T}^\ell$: hence, it is possible to say that a phase space region of integrability is *foliated* into $\ell$-dimensional invariant tori $\mathcal{T}(\boldsymbol{I})$ parametrized by the values of the constants of motion $\boldsymbol{I} \in \mathcal{I}$.

If an integrable system is anisochronous then it is *canonically integrable*: that is, it is possible to define on $W$ a canonical change of coordinates $(\boldsymbol{p}, \boldsymbol{q}) = \mathcal{C}(\boldsymbol{A}, \boldsymbol{\alpha})$ mapping $W$ onto $J \times \mathbb{T}^\ell$ and such that $\mathcal{H}(\mathcal{C}(\boldsymbol{A}, \boldsymbol{\alpha})) = h(\boldsymbol{A})$ for a suitable $h$. Then, if $\boldsymbol{\omega}(\boldsymbol{A}) \stackrel{\text{def}}{=} \partial_{\boldsymbol{A}} h(\boldsymbol{A})$, the equations of motion become

$$\dot{\boldsymbol{A}} = \boldsymbol{0}, \qquad \dot{\boldsymbol{\alpha}} = \boldsymbol{\omega}(\boldsymbol{A}) \qquad [28]$$

Given a system $(\boldsymbol{I}, \boldsymbol{\varphi})$ of coordinates integrating an anisochronous system the construction of action–angle coordinates can be performed, in principle, via a classical procedure (under a few extra assumptions).

Let $\gamma_1, \ldots, \gamma_\ell$ be $\ell$ topologically independent circles on $\mathbb{T}^\ell$, for definiteness let $\gamma_i(\boldsymbol{I}) = \{\boldsymbol{\varphi} \,|\, \varphi_1 = \varphi_2 = \cdots = \varphi_{i-1} = \varphi_{i+1} = \cdots = 0,\ \varphi_i \in [0, 2\pi]\}$, and set

$$A_i(\boldsymbol{I}) = \frac{1}{2\pi} \oint_{\gamma_i(\boldsymbol{I})} \boldsymbol{p} \cdot \mathrm{d}\boldsymbol{q} \qquad [29]$$

If the map $\boldsymbol{I} \longleftrightarrow \boldsymbol{A}(\boldsymbol{I})$ is analytically invertible as $\boldsymbol{I} = \boldsymbol{I}(\boldsymbol{A})$, the function

$$S(\boldsymbol{A}, \boldsymbol{\varphi}) = (\lambda) \int_{\boldsymbol{0}}^{\boldsymbol{\varphi}} \boldsymbol{p} \cdot \mathrm{d}\boldsymbol{q} \qquad [30]$$

is well defined if the integral is over any path $\lambda$ joining the points $(p(\boldsymbol{I}(\boldsymbol{A}), \boldsymbol{0}), q(\boldsymbol{I}(\boldsymbol{A}), \boldsymbol{0}))$ and $(p(\boldsymbol{I}(\boldsymbol{A}), \boldsymbol{\varphi})), q(\boldsymbol{I}(\boldsymbol{A}), \boldsymbol{\varphi})$ and lying on the torus parametrized by $\boldsymbol{I}(\boldsymbol{A})$.

The key remark in the proof that [30] really defines a function of the only variables $\boldsymbol{A}, \boldsymbol{\varphi}$ is that anisochrony implies the vanishing of the Poisson brackets (cf. [18]): $\{I_i, I_j\} = 0$ (hence also $\{A_i, A_j\} \equiv \sum_{h,k} \partial_{I_k} A_i\, \partial_{I_h} A_j \{I_k, I_h\} = 0$). And the property $\{I_i, I_j\} = 0$ can be checked to be precisely the integrability condition for the differential form $\boldsymbol{p} \cdot \mathrm{d}\boldsymbol{q}$ restricted to the surface obtained by varying $\boldsymbol{q}$ while $\boldsymbol{p}$ is constrained so that $(\boldsymbol{p}, \boldsymbol{q})$ stays on the surface $\boldsymbol{I} =$ constant, i.e., on the invariant torus of the points with fixed $\boldsymbol{I}$.

The latter property is necessary and sufficient in order that the function $S(\boldsymbol{A}, \boldsymbol{\varphi})$ be well defined (i.e., be independent on the integration path $\lambda$) up to an additive quantity of the form $\sum_i 2\pi n_i A_i$ with $\boldsymbol{n} = (n_1, \ldots, n_\ell)$ integers.

Then the action–angle variables are defined by the canonical change of coordinates with $S(\boldsymbol{A}, \boldsymbol{\varphi})$ as generating function, i.e., by setting

$$\alpha_i = \partial_{A_i} S(\boldsymbol{A}, \boldsymbol{\varphi}), \qquad I_i = \partial_{\boldsymbol{\varphi}_i} S(\boldsymbol{A}, \boldsymbol{\varphi}) \qquad [31]$$

and, since the computation of $S(\boldsymbol{A}, \boldsymbol{\varphi})$ is "reduced to integrations" which can be regarded as a natural extension of the quadratures discussed in the one-dimensional cases, such systems are also called *integrable by quadratures*. The just-described construction is a version of the more general *Arnol'd–Liouville theorem*.

In practice, however, the actual evaluation of the integrals in [29], [30] can be difficult: its analysis in various cases (even as "elementary" as the pendulum) has in fact led to key progress in various domains, for example, in the theory of special functions and in group theory.

In general, any surface on phase space on which the restriction of the differential form $\boldsymbol{p} \cdot \mathrm{d}\boldsymbol{q}$ is locally integrable is called a *Lagrangian manifold*: hence the invariant tori of an anisochronous integrable system are Lagrangian manifolds.

If an integrable system is anisochronous, it cannot admit more than $\ell$ independent constants of motion; furthermore, it does not admit invariant tori of dimension $> \ell$. Hence $\ell$-dimensional invariant tori are called maximal.

Of course, invariant tori of dimension $< \ell$ can also exist: this happens when the variables $\boldsymbol{I}$ are such that the frequencies $\boldsymbol{\omega}(\boldsymbol{I})$ admit nontrivial rational relations; i.e., there is an integer components vector $\boldsymbol{\nu} \in \mathbb{Z}^\ell$, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_\ell) \neq 0$ such that

$$\boldsymbol{\omega}(\boldsymbol{I}) \cdot \boldsymbol{\nu} = \sum_i \omega_i(\boldsymbol{I}) \nu_i = 0 \qquad [32]$$

in this case, the invariant torus $\mathcal{T}(\boldsymbol{I})$ is called *resonant*. If the system is anisochronous then $\det \partial_{\boldsymbol{I}} \boldsymbol{\omega}(\boldsymbol{I}) \neq 0$ and, therefore, the resonant tori are associated with values of the constants of motion $\boldsymbol{I}$ which form a set of measure zero in the space $\mathcal{I}$ but which is not empty and dense.

Examples of isochronous systems are the systems of harmonic oscillators, i.e., systems with Hamiltonian

$$\sum_{i=1}^{\ell} \frac{1}{2m_i} p_i^2 + \frac{1}{2} \sum_{i,j}^{1,\ell} c_{ij} q_i q_j$$

where the matrix $v$ is a positive-definite matrix. This is an isochronous system with frequencies $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_\ell)$ whose squares are the eigenvalues of the matrix $m_i^{-1/2} c_{ij} m_j^{-1/2}$. It is integrable in the region $W$ of the data $\boldsymbol{x} = (\boldsymbol{p}, \boldsymbol{q}) \in \mathbb{R}^{2\ell}$ such that, setting

$$A_\beta = \frac{1}{2\omega_\beta} \left( \left( \sum_{i=1}^{\ell} \frac{v_{\beta,i} p_i}{\sqrt{m_i}} \right)^2 + \omega_\beta^2 \left( \sum_{i=1}^{\ell} \frac{v_{\beta,i} q_i}{\sqrt{m_i^{-1}}} \right)^2 \right)$$

for all eigenvectors $\boldsymbol{v}_\beta$, $\beta = 1, \ldots, \ell$, of the above matrix, the vectors $\boldsymbol{A}$ have all components $>0$.

Even though this system is isochronous, it nevertheless admits a system of canonical action–angle coordinates in which the Hamiltonian takes the simplest form

$$h(\boldsymbol{A}) = \sum_{\beta=1}^{\ell} \omega_\beta A_\beta \equiv \boldsymbol{\omega} \cdot \boldsymbol{A} \qquad [33]$$

with

$$\alpha_\beta = -\arctan\left(\frac{\sum_{i=1}^{\ell} \frac{v_{\beta,i} p_i}{\sqrt{m_i}}}{\sum_{i=1}^{\ell} \sqrt{m_i}\omega_\beta v_{\beta,i} q_i}\right)$$

as conjugate angles.

An example of anisochronous system is the *free rotators* or *free wheels*: i.e., $\ell$ noninteracting points on a circle of radius $R$ or $\ell$ noninteracting homogeneous coaxial wheels of radius $R$. If $J_i = m_i R^2$ or, respectively, $J_i = (1/2) m_i R^2$ are the inertia moments and if the positions are determined by $\ell$ angles $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_\ell)$, the angular velocities are constants related to the angular momenta $\boldsymbol{A} = (A_1, \ldots, A_\ell)$ by $\omega_i = A_i / J_i$. The Hamiltonian and the spectrum are

$$h(\boldsymbol{A}) = \sum_{i=1}^{\ell} \frac{1}{2J_i} A_i^2, \qquad \boldsymbol{\omega}(\boldsymbol{A}) = \left(\frac{1}{J_i} A_i\right)_{i=1,\ldots,\ell} \qquad [34]$$

For further details see Landau and Lifshitz (1976), Gallavotti (1983), Arnol'd (1989), and Fassò (1998).

## Multidimensional Quadratures: Central Motion

Several important mechanical systems with more than one degree of freedom are integrable by canonical quadratures in vast regions of phase space. This is checked by showing that there is a foliation into invariant tori $\mathcal{T}(\boldsymbol{I})$ of dimension equal to the number of degrees of freedom ($\ell$) parametrized by $\ell$ constants of motion $\boldsymbol{I}$ in involution, i.e., such that $\{I_i, I_j\} = 0$. One then performs, if possible, the construction of the action–angle variables by the quadratures discussed in the previous section.

The above procedure is well illustrated by the theory of the planar motion of a unit mass attracted by a coplanar center of force: the Lagrangian is, in polar coordinates $(\rho, \theta)$,

$$\mathcal{L} = \frac{m}{2}(\dot\rho^2 + \rho^2 \dot\theta^2) - V(\rho)$$

The planarity of the motion is not a strong restriction as central motion always takes place on a plane.

Hence, the equations of motion are

$$\frac{\mathrm{d}}{\mathrm{d}t} m\rho^2\dot\theta = 0$$

i.e., $m\rho^2\dot\theta = G$ is a constant of motion (it is the angular momentum), and

$$\begin{aligned} \ddot\rho &= -\partial_\rho V(\rho) + \partial_\rho \frac{m}{2}\rho^2\dot\theta^2 \\ &= -\partial_\rho V(\rho) + \frac{G^2}{m\rho^3} \\ &\stackrel{\text{def}}{=} -\partial_\rho V_G(\rho) \end{aligned}$$

Then the energy conservation yields a second constant of motion $E$,

$$\begin{aligned} \frac{m}{2}\dot\rho^2 + \frac{1}{2}\frac{G^2}{m\rho^2} + V(\rho) &= E \\ = \frac{1}{2m}p_\rho^2 + \frac{1}{2m}\frac{p_\theta^2}{\rho^2} + V(\rho) & \end{aligned} \qquad [35]$$

The right-hand side (rhs) is the Hamiltonian for the system, derived from $\mathcal{L}$, if $p_\rho$, $p_\theta$ denote conjugate momenta of $\rho$, $\theta$: $p_\rho = m\dot\rho$ and $p_\theta = m\rho^2\dot\theta$ (note that $p_\theta = G$).

Suppose $\rho^2 V(\rho) \xrightarrow[\rho\to 0]{} 0$: then the singularity at the origin cannot be reached by any motion starting with $\rho > 0$ if $G > 0$. Assume also that the function

$$V_G(\rho) \stackrel{\text{def}}{=} \frac{1}{2}\frac{G^2}{m\rho^2} + V(\rho)$$

has only one minimum $E_0(G)$, no maximum and no horizontal inflection, and tends to a limit $E_\infty(G) \le \infty$ when $\rho \to \infty$. Then the system is integrable in the domain $W = \{(\boldsymbol{p}, \boldsymbol{q}) \mid E_0(G) < E < E_\infty(G), G \neq 0\}$.

This is checked by introducing a "standard" periodic solution $t \to R(t)$ of $m\ddot\rho = -\partial_\rho V_G(\rho)$ with energy $E_0(G) < E < E_\infty(G)$ and initial data $\rho = \rho_{E,-}(G)$, $\dot\rho = 0$ at time $t = 0$, where $\rho_{E,\pm}(G)$ are the two solutions of $V_G(\rho) = E$, see the section "Quadratures": this is a periodic analytic function of $t$ with period

$$T(E,G) = 2\int_{\rho_{E,-}(G)}^{\rho_{E,+}(G)} \frac{\mathrm{d}x}{\sqrt{(2/m)(E - V_G(x))}}$$

The function $R(t)$ is given, for $0 \le t \le \frac{1}{2}T(E,G)$ or for $\frac{1}{2}T(E,G) \le t \le T(E,G)$, by the quadratures

$$t = \int_{\rho_{E,-}(G)}^{R(t)} \frac{\mathrm{d}x}{\sqrt{(2/m)(E - V_G(x))}} \qquad [36a]$$

or

$$t = \frac{T(E,G)}{2} - \int_{\rho_{E,+}(G)}^{R(t)} \frac{\mathrm{d}x}{\sqrt{(2/m)(E - V_G(x))}} \qquad [36b]$$

respectively. The analytic regularity of $R(t)$ follows from the general existence, uniqueness, and regularity theorems applied to the differential equation for $\ddot{\rho}$.

Given an initial datum $\dot{\rho}_0, \rho_0, \dot{\theta}_0, \theta_0$ with energy $E$ and angular momentum $G$, define $t_0$ to be the time such that $R(t_0) = \rho_0, \dot{R}(t_0) = \dot{\rho}_0$: then $\rho(t) \equiv R(t + t_0)$ and $\theta(t)$ can be computed as

$$\theta(t) = \theta_0 + \int_0^t \frac{G}{mR(t' + t_0)^2} \mathrm{d}t'$$

a second quadrature. Therefore, we can use as coordinates for the motion $E, G, t_0$, which determine $\dot{\rho}_0, \rho_0, \dot{\theta}_0$ and a fourth coordinate that determines $\theta_0$ which could be $\theta_0$ itself but which is conveniently determined, via the second quadrature, as follows.

The function $Gm^{-1}R(t)^{-2}$ is periodic with period $T(E, G)$; hence it can be expressed in a Fourier series

$$\chi_0(E, G) + \sum_{k \neq 0} \chi_k(E, G) \exp\left(\frac{2\pi}{T(E, G)} \mathrm{i}tk\right)$$

the quadrature for $\theta(t)$ can be performed by integrating the series terms. Setting

$$\bar{\theta}(t_0) \stackrel{\text{def}}{=} \frac{T(E, G)}{2\pi} \sum_{k \neq 0} \frac{\chi_k(E, G)}{k} \exp\left(\frac{2\pi}{T(E, G)} \mathrm{i}t_0 k\right)$$

and $\varphi_1(0) = \theta_0 - \bar{\theta}(t_0)$, the expression

$$\theta(t) = \theta_0 + \int_0^t \frac{G}{mR(t' + t_0)^2} \mathrm{d}t'$$

becomes

$$\varphi_1(t) = \varphi_1(0) + \chi_0(E, G)\, t \qquad [37]$$

Hence the system is integrable and the spectrum is $\boldsymbol{\omega}(E, G) = (\omega_0(E, G), \omega_1(E, G)) \equiv (\omega_0, \omega_1)$ with

$$\omega_0 \stackrel{\text{def}}{=} \frac{2\pi}{T(E, G)} \quad \text{and} \quad \omega_1 \stackrel{\text{def}}{=} \chi_0(E, G)$$

while $\boldsymbol{I} = (E, G)$ are constants of motion and the angles $\boldsymbol{\varphi} = (\varphi_0, \varphi_1)$ can be taken as

$$\varphi_0 \stackrel{\text{def}}{=} \omega_0 t_0, \qquad \varphi_1 \stackrel{\text{def}}{=} \theta_0 - \bar{\theta}(t_0)$$

At $E$, $G$ fixed, the motion takes place on a two-dimensional torus $\mathcal{T}(E, G)$ with $\varphi_0, \varphi_1$ as angles.

In the anisochronous cases, i.e., when $\det \partial_{E,G} \boldsymbol{\omega}(E, G) \neq 0$, canonical action–angle variables conjugated to $(p_\rho, \rho, p_\theta, \theta)$ can be constructed via [29], [30] by using two cycles $\gamma_1, \gamma_2$ on the torus $\mathcal{T}(E, G)$. It is convenient to choose

1. $\gamma_1$ as the cycle consisting of the points $\rho = x, \theta = 0$ whose first half (where $p_\rho \geq 0$) consists in the set $\rho_{E,-}(G) \leq x \leq \rho_{E,+}(G), p_\rho = \sqrt{2m(E - V_G(x))}$ and $\mathrm{d}\theta = 0$; and
2. $\gamma_2$ as the cycle $\rho = \text{const}, \theta \in [0, 2\pi]$ on which $\mathrm{d}\rho = 0$ and $p_\theta = G$ obtaining

$$A_1 = \frac{2}{2\pi} \int_{\rho_{E,-}(G)}^{\rho_{E,+}(G)} \sqrt{2m(E - V_G(x))} \mathrm{d}x, \qquad A_2 = G \qquad [38]$$

According to the general theory (cf. the previous section) a generating function for the canonical change of coordinates from $(p_\rho, \rho, p_\theta, \theta)$ to action–angle variables is (if, to fix ideas, $p_\rho > 0$)

$$S(A_1, A_2, \rho, \theta) = G\theta + \int_{\rho_{E,-}}^{\rho} \sqrt{2m(E - V_G(x))} \mathrm{d}x \qquad [39]$$

In terms of the above $\omega_0, \chi_0$ the Jacobian matrix $\partial(E, G)/\partial(A_1, A_2)$ is computed from [38], [39] to be $\begin{pmatrix} \omega_0 & \chi_0 \\ 0 & 1 \end{pmatrix}$. It follows that $\partial_E S = t, \partial_G S = \theta - \bar{\theta}(t) - \chi_0{}^t$ so that, see [31],

$$\alpha_1 \stackrel{\text{def}}{=} \partial_{A_1} S = \omega_0 t, \qquad \alpha_2 \stackrel{\text{def}}{=} \partial_{A_2} S = \theta - \bar{\theta}(t) \qquad [40]$$

and $(A_1, \alpha_1), (A_2, \alpha_2)$ are the action–angle pairs.

For more details, see Landau and Lifshitz (1976) and Gallavotti (1983).

## Newtonian Potential and Kepler's Laws

The anisochrony property, that is, $\det \partial(\omega_0, \chi_0)/\partial(A_1, A_2) \neq 0$ or, equivalently, $\det \partial(\omega_0, \chi_0)/\partial(E, G) \neq 0$, is not satisfied in the important cases of the harmonic potential and the Newtonian potential. Anisochrony being only a sufficient condition for canonical integrability it is still possible (and true) that, nevertheless, in both cases the canonical transformation generated by [39] integrates the system. This is expected since the two potentials are limiting cases of anisochronous ones (e.g., $|\boldsymbol{q}|^{2+\varepsilon}$ and $|\boldsymbol{q}|^{-1-\varepsilon}$ with $\varepsilon \to 0$).

The Newtonian potential

$$\mathcal{H}(\boldsymbol{p}, \boldsymbol{q}) = \frac{1}{2m} \boldsymbol{p}^2 - \frac{km}{|\boldsymbol{q}|}$$

is integrable in the region $G \neq 0, E_0(G) = -k^2 m^3 / 2G^2 < E < 0, |G| < \sqrt{k^2 m^3 / (-2E)}$. Proceeding as in the last section, one finds integrating coordinates and that the integrable motions develop on ellipses with one focus on the center of attraction $S$ so that motions are periodic, hence not anisochronous: nevertheless, the construction of the canonical coordinates via [29]–[31] (hence [39]) works and leads to canonical coordinates $(L', \lambda', G', \gamma')$. To obtain action–angle variables with a simple

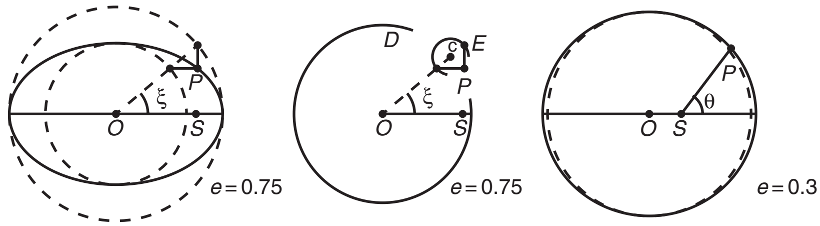



**Figure 2** Eccentric and true anomalies of $P$, which moves on a small circle $E$ centered at a point $c$ moving on the circle $D$ located half-way between the two concentric circles containing the Keplerian ellipse: the anomaly of $c$ with respect to the axis $OS$ is $\xi$. The circle $D$ is *eccentric* with respect to $S$ and therefore $\xi$ is, even today, called eccentric anomaly, whereas the circle $D$ is, in ancient terminology, the *deferent* circle (eccentric circles were introduced in astronomy by Ptolemy). The small circle $E$ on which the point $P$ moves is, in ancient terminology, an *epicycle*. The deferent and the epicyclical motions are synchronous (i.e., they have the same period); Kepler discovered that his key *a priori* hypothesis of inverse proportionality between angular velocity on the deferent and distance between $P$ and $S$ (i.e., $\rho\dot{\xi} = \text{constant}$) implied both synchrony and elliptical shape of the orbit, with focus in $S$. The latter law is equivalent to $\rho^2\dot{\theta} = \text{constant}$ (because of the identity $a\dot{\xi} = \rho\dot{\theta}$). Small eccentricity ellipses can hardly be distinguished from circles.

interpretation, it is convenient to perform on the variables $(L', \lambda', G', \gamma')$ (constructed by following the procedure just indicated) a further trivial canonical transformation by setting $L = L' + G'$, $G = G'$, $\lambda = \lambda'$, $\gamma = \gamma' - \lambda'$; then

1. $\lambda$ (average anomaly) is the time necessary for the point $P$ to move from the pericenter to its actual position, in units of the period, times $2\pi$;
2. $L$ (action) is essentially the energy $E = -k^2m^3/2L^2$;
3. $G$ (angular momentum);
4. $\gamma$ (axis longitude), is the angle between a fixed axis and the major axis of the ellipse oriented from the center of the ellipse $O$ to the center of attraction $S$.

The eccentricity of the ellipse is $e$ such that $G = \pm L\sqrt{1-e^2}$. The ellipse equation is $\rho = a(1 - e\cos\xi)$, where $\xi$ is the eccentric anomaly (see Figure 2), $a = L^2/km^2$ is the major semiaxis, and $\rho$ is the distance to the center of attraction S.

Finally, the relations between eccentric anomaly $\xi$, average anomaly $\lambda$, true anomaly $\theta$ (the latter is the polar angle), and SP distance $\rho$ are given by the Kepler equations

$$\begin{aligned} &\lambda = \xi - e\sin\xi \\ &(1 - e\cos\xi)(1 + e\cos\theta) = 1 - e^2 \\ &\lambda = (1-e^2)^{3/2}\int_0^\theta \frac{\mathrm{d}\theta'}{(1 + e\cos\theta')^2} \\ &\frac{\rho}{a} = \frac{1-e^2}{1+e\cos\theta} \end{aligned} \qquad [41]$$

and the relation between true anomaly and average anomaly can be inverted in the form

$$\begin{aligned} \xi &= \lambda + g_\lambda \\ \theta &= \lambda + f_\lambda \;\Rightarrow\; \frac{\rho}{a} = \frac{1-e^2}{1 + e\cos(\lambda + f_\lambda)} \end{aligned} \qquad [42]$$

where $g_\lambda = g(e\sin\lambda, e\cos\lambda)$, $f_\lambda = f(e\sin\lambda, e\cos\lambda)$, and $g(x,y), f(x,y)$ are suitable functions analytic for $|x|, |y| < 1$. Furthermore, $g(x,y) = x(1 + y + \cdots)$, $f(x,y) = 2x(1 + \frac{5}{4}y + \cdots)$ and the ellipses denote terms of degree 2 or higher in $x, y$, containing only even powers of $x$.

For more details, the reader is referred to Landau and Lifshitz (1976) and Gallavotti (1983).

## Rigid Body

Another fundamental integrable system is the rigid body in the absence of gravity and with a fixed point O. It can be naturally described in terms of the Euler angles $\theta_0, \varphi_0, \psi_0$ (see Figure 3) and their derivatives $\dot{\theta}_0$, $\dot{\varphi}_0$, $\dot{\psi}_0$.

Let $I_1, I_2, I_3$ be the three *principal inertia moments* of the body along the three principal axes with unit vectors $\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{i}_3$. The inertia moments and the principal axes are the eigenvalues and the associated unit eigenvectors of the $3\times 3$ inertia matrix $\mathcal{I}$, which is defined by $\mathcal{I}_{hk} = \sum_{i=1}^n m_i(\boldsymbol{x}_i)_h(\boldsymbol{x}_i)_k$, where $h, k = 1, 2, 3$ and $\boldsymbol{x}_i$ is the position of the $i$th particle in a reference frame with origin at O and in which

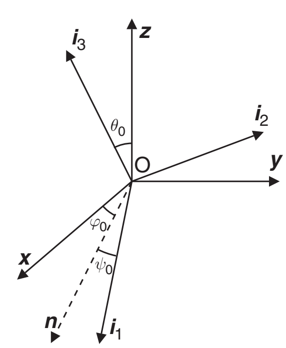



**Figure 3** The *Euler angles* of the comoving frame $\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{i}_3$ with respect to a fixed frame $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$. The direction $\boldsymbol{n}$ is the "node line, intersection between the planes $\boldsymbol{x}, \boldsymbol{y}$ and $\boldsymbol{i}_1, \boldsymbol{i}_2$.

all particles are at rest: this *comoving frame* exists as a consequence of the rigidity constraint. The principal axes form a coordinate system which is comoving as well: that is, in the frame $(O; \boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{i}_3)$ as well, the particles are at rest.

The Lagrangian is simply the kinetic energy: we imagine the rigidity constraint to be ideal (e.g., as realized by internal central forces in the limit of infinite rigidity, as mentioned in the section "Lagrange and Hamilton forms of equations of motion"). The *angular velocity* of the rigid motion is defined by

$$\boldsymbol{\omega} = \dot{\theta}_0 \boldsymbol{n} + \dot{\varphi}_0 \boldsymbol{z} + \dot{\psi}_0 \boldsymbol{i}_3 \qquad [43]$$

expressing that a generic infinitesimal motion must consist of a variation of the three Euler angles and, therefore, it has to be a rotation of speeds $\dot{\theta}_0, \dot{\varphi}_0, \dot{\psi}_0$ around the axes $\boldsymbol{n}, \boldsymbol{z}, \boldsymbol{i}_3$ as shown in Figure 3.

Let $(\omega_1, \omega_2, \omega_3)$ be the components of $\boldsymbol{\omega}$ along the principal axes $\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{i}_3$: for brevity, the latter axes will often be called $\mathbf{1}, \mathbf{2}, \mathbf{3}$. Then the angular momentum $\boldsymbol{M}$, with respect to the pivot point $O$, and the kinetic energy $K$ can be checked to be

$$\begin{aligned} \boldsymbol{M} &= I_1\omega_1 \boldsymbol{i}_1 + I_2\omega_2 \boldsymbol{i}_2 + I_3\omega_3 \boldsymbol{i}_3 \\ K &= \frac{1}{2}(I_1\omega_1^2 + I_2\omega_2^2 + I_2\omega_3^2) \end{aligned} \qquad [44]$$

and are constants of motion. From Figure 3 it follows that $\omega_1 = \dot{\theta}_0 \cos\psi_0 + \dot{\varphi}_0 \sin\theta_0 \sin\psi_0$, $\omega_2 = -\dot{\theta}_0 \sin\psi_0 + \dot{\varphi}_0 \sin\theta_0 \cos\psi_0$ and $\omega_3 = \dot{\varphi}_0 \cos\theta_0 + \dot{\psi}_0$, so that the Lagrangian, uninspiring at first, is

$$\begin{aligned} \mathcal{L} \stackrel{\text{def}}{=} & \frac{1}{2} I_1 (\dot{\theta}_0 \cos\psi_0 + \dot{\varphi}_0 \sin\theta_0 \sin\psi_0)^2 \\ & + \frac{1}{2} I_2 (-\dot{\theta}_0 \sin\psi_0 + \dot{\varphi}_0 \sin\theta_0 \cos\psi_0)^2 \\ & + \frac{1}{2} I_3 (\dot{\varphi}_0 \cos\theta_0 + \dot{\psi}_0)^2 \end{aligned} \qquad [45]$$

Angular momentum conservation does not imply that the components $\omega_j$ are constants because $\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{i}_3$ also change with time according to

$$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{i}_j = \boldsymbol{\omega} \wedge \boldsymbol{i}_j, \quad j = 1, 2, 3$$

Hence, $\dot{\boldsymbol{M}} = \mathbf{0}$ becomes, by the first of [44] and denoting $I\boldsymbol{\omega} = (I_1\omega_1, I_2\omega_2, I_3\omega_3)$, the *Euler equations* $I\dot{\boldsymbol{\omega}} + \boldsymbol{\omega} \wedge I\boldsymbol{\omega} = \mathbf{0}$, or

$$\begin{aligned} I_1\dot{\omega}_1 &= (I_2 - I_3)\omega_2\omega_3 \\ I_2\dot{\omega}_2 &= (I_3 - I_1)\omega_3\omega_1 \\ I_3\dot{\omega}_3 &= (I_1 - I_2)\omega_1\omega_2 \end{aligned} \qquad [46]$$

which can be considered together with the conserved quantities [44].

Since angular momentum is conserved, it is convenient to introduce the *laboratory frame* $(O; \boldsymbol{x}_0, \boldsymbol{y}_0, \boldsymbol{z}_0)$ with fixed axes $\boldsymbol{x}_0, \boldsymbol{y}_0, \boldsymbol{z}_0$ and (see Figure 4):

1. $(O; \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$, the *momentum frame* with fixed axes, but with $\boldsymbol{z}$-axis oriented as $\boldsymbol{M}$, and $\boldsymbol{x}$-axis coinciding with the node (i.e., the intersection) of the $\boldsymbol{x}_0$–$\boldsymbol{y}_0$ plane and the $\boldsymbol{x}$–$\boldsymbol{y}$ plane (orthogonal to $\boldsymbol{M}$). Therefore, $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ is determined by the two Euler angles $\zeta, \gamma$ of $(O; \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ in $(O; \boldsymbol{x}_0, \boldsymbol{y}_0, \boldsymbol{z}_0)$;
2. $(O; \mathbf{1}, \mathbf{2}, \mathbf{3})$, the comoving frame, that is, the frame fixed with the body, and with unit vectors $\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{i}_3$ parallel to the principal axes of the body. The frame is determined by three Euler angles $\theta_0, \varphi_0, \psi_0$;
3. the Euler angles of $(O; \mathbf{1}, \mathbf{2}, \mathbf{3})$ with respect to $(O; \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$, which are denoted $\theta, \varphi, \psi$;
4. $G$, the total angular momentum: $G^2 = \sum_j I_j^2 \omega_j^2$;
5. $M_3$, the angular momentum along the $z_0$ axis; $M_3 = G\cos\zeta$; and
6. $L$, the projection of $\boldsymbol{M}$ on the axis $\mathbf{3}$, $L = G\cos\theta$.

The quantities $G, M_3, L, \varphi, \gamma, \psi$ determine $\theta_0, \varphi_0, \psi_0$ and $\dot{\theta}_0, \dot{\varphi}_0, \dot{\psi}_0$, or the $p_{\theta_0}, p_{\varphi_0}, p_{\psi_0}$ variables conjugated to $\theta_0, \varphi_0, \psi_0$ as shown by the following comment.

Considering Figure 4, the angles $\zeta, \gamma$ determine location, in the fixed frame $(O; \boldsymbol{x}_0, \boldsymbol{y}_0, \boldsymbol{z}_0)$ of the direction of $\boldsymbol{M}$ and the node line $\boldsymbol{m}$, which are, respectively, the $\boldsymbol{z}$-axis and the $\boldsymbol{x}$-axis of the fixed frame associated with the angular momentum; the angles $\theta, \varphi, \psi$ then determine the position of the comoving frame with respect to the fixed frame $(O; \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$, hence its position with respect to $(O; \boldsymbol{x}_0, \boldsymbol{y}_0, \boldsymbol{z}_0)$, that is, $(\theta_0, \varphi_0, \psi_0)$. From this and $G$, it is possible to determine $\boldsymbol{\omega}$ because

$$\begin{aligned} \cos\theta = \frac{I_3\omega_3}{G}, \qquad \tan\psi = \frac{I_2\omega_2}{I_1\omega_1} \\ \omega_2^2 = I_2^{-2}(G^2 - I_1^2\omega_1^2 - I_3^2\omega_3^2) \end{aligned} \qquad [47]$$

and, from [43], $\dot{\theta}_0, \dot{\varphi}_0, \dot{\psi}_0$ are determined.

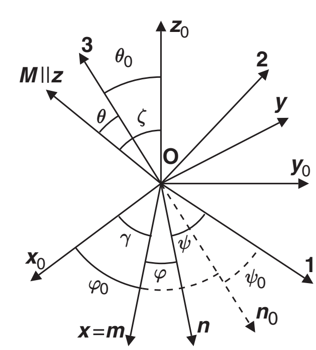



**Figure 4** The laboratory frame, the angular momentum frame, and the comoving frame (and the Deprit angles).

The Lagrangian [45] gives immediately (after expressing $\boldsymbol{\omega}$, i.e., $\boldsymbol{n}, \boldsymbol{z}, \boldsymbol{i}_3$, in terms of the Euler angles $\theta_0, \varphi_0, \psi_0$) an expression for the variables $p_{\theta_0}, p_{\varphi_0}, p_{\psi_0}$ conjugated to $\theta_0, \varphi_0, \psi_0$:

$$p_{\theta_0} = \boldsymbol{M}\cdot\boldsymbol{n}_0, \quad p_{\varphi_0} = \boldsymbol{M}\cdot\boldsymbol{z}_0, \quad p_{\psi_0} = \boldsymbol{M}\cdot\boldsymbol{i}_3 \qquad [48]$$

and, in principle, we could proceed to compute the Hamiltonian.

However, the computation can be avoided because of the very remarkable property (Deprit), which can be checked with some patience, making use of [48] and of elementary spherical trigonometry identities,

$$\begin{aligned} &M_3\,\mathrm{d}\gamma + G\,\mathrm{d}\varphi + L\,\mathrm{d}\psi \\ &\quad = p_{\varphi_0}\,\mathrm{d}\varphi_0 + p_{\psi_0}\,\mathrm{d}\psi_0 + p_{\theta_0}\,\mathrm{d}\theta_0 \end{aligned} \qquad [49]$$

which means that the map $((M_3,\gamma),(L,\psi),(G,\varphi)) \longleftrightarrow ((p_{\theta_0},\theta_0),(p_{\varphi_0},\varphi_0),(p_{\psi_0},\psi_0))$ is a canonical map. And in the new coordinates, the kinetic energy, hence the Hamiltonian, takes the form

$$K = \frac{1}{2}\left[\frac{L^2}{I_3} + (G^2 - L^2)\left(\frac{\sin^2\psi}{I_1} + \frac{\cos^2\psi}{I_2}\right)\right] \qquad [50]$$

This again shows that $G, M_3$ are constants of motion, and the $L, \psi$ variables are determined by a quadrature, because the Hamilton equation for $\psi$ combined with the energy conservation yields

$$\dot\psi = \pm\left(\frac{1}{I_3} - \frac{\sin^2\psi}{I_1} - \frac{\cos^2\psi}{I_2}\right) \times \sqrt{\frac{2E - G^2\left(\frac{\sin^2\psi}{I_1} + \frac{\cos^2\psi}{I_2}\right)}{\frac{1}{I_3} - \frac{\sin^2\psi}{I_1} - \frac{\cos^2\psi}{I_2}}} \qquad [51]$$

In the integrability region, this motion is periodic with some period $T_L(E,G)$. Once $\psi(t)$ is determined, the Hamilton equation for $\varphi$ leads to the further quadrature

$$\dot\varphi = \left(\frac{\sin^2\psi(t)}{I_1} + \frac{\cos^2\psi(t)}{I_2}\right)G \qquad [52]$$

which determines a second periodic motion with period $T_G(E,G)$. The $\gamma, M_3$ are constants and, therefore, the motion takes place on three-dimensional invariant tori $\mathcal{T}_{E,G,M_3}$ in phase space, each of which is "always" foliated into two-dimensional invariant tori parametrized by the angle $\gamma$ which is constant (by [50], because $K$ is $M_3$-independent): the latter are in turn foliated by one-dimensional invariant tori, that is, by periodic orbits, with $E, G$ such that the value of $T_L(E,G)/T_G(E,G)$ is rational.

Note that if $I_1 = I_2 = I$, the above analysis is extremely simplified. Furthermore, if gravity $g$ acts on the system the Hamiltonian will simply change by the addition of a potential $-mgz$ if $z$ is the height of the center of mass. Then (see **Figure 4**), if the center of mass of the body is on the axis $\boldsymbol{i}_3$ and $z = h\cos\theta_0$, and $h$ is the distance of the center of mass from $O$, since $\cos\theta_0 = \cos\theta\cos\zeta - \sin\theta\sin\zeta\cos\varphi$, the Hamiltonian will become $\mathcal{H} = K - mgh\cos\theta_0$ or

$$\mathcal{H} = \frac{G^2}{2I_3} + \frac{G^2 - L^2}{2I} - mgh\left(\frac{M_3 L}{G^2} - \left(1 - \frac{M_3^2}{G^2}\right)^{1/2} \times \left(1 - \frac{L^2}{G^2}\right)^{1/2}\cos\varphi\right) \qquad [53]$$

so that, again, the system is integrable by quadratures (with the roles of $\psi$ and $\varphi$ "interchanged" with respect to the previous case) in suitable regions of phase space. This is called the *Lagrange's gyroscope*.

A less elementary integrable case is when the inertia moments are related as $I_1 = I_2 = 2I_3$ and the center of mass is in the $\boldsymbol{i}_1$–$\boldsymbol{i}_2$ plane (rather than on the $\boldsymbol{i}_3$-axis) and only gravity acts, besides the constraint force on the pivot point $O$; this is called *Kowalevskaia's gyroscope*.

For more details, see Gallavotti (1983).

## Other Quadratures

An interesting classical integrable motion is that of a point mass attracted by two equal-mass centers of gravitational attraction, or a point ideally constrained to move on the surface of a general ellipsoid.

New integrable systems have been discovered quite recently and have generated a wealth of new developments ranging from group theory (as integrable systems are closely related to symmetries) to partial differential equations.

It is convenient to extend the notion of integrability by stating that a system is integrable in a region $W$ of phase space if

1. there is a change of coordinates $(\boldsymbol{p},\boldsymbol{q}) \in W \longleftrightarrow \{\boldsymbol{A},\boldsymbol{\alpha},\boldsymbol{Y},\boldsymbol{y}\} \in (U \times \mathbb{T}^\ell) \times (V \times \mathbb{R}^m)$ where $U \subset \mathbb{R}^\ell$, $V \subset \mathbb{R}^m$, with $\ell + m \geq 1$, are open sets; and
2. the $\boldsymbol{A}, \boldsymbol{Y}$ are constants of motion while the other coordinates vary "linearly":

$$(\boldsymbol{\alpha},\boldsymbol{y}) \rightarrow (\boldsymbol{\alpha} + \boldsymbol{\omega}(\boldsymbol{A},\boldsymbol{Y})t,\ \boldsymbol{y} + \boldsymbol{v}(\boldsymbol{A},\boldsymbol{Y})t) \qquad [54]$$

where $\boldsymbol{\omega}(\boldsymbol{A},\boldsymbol{Y}), \boldsymbol{v}(\boldsymbol{A},\boldsymbol{Y})$ are smooth functions.

In the new sense, the systems studied in the previous sections are integrable in much wider regions (essentially on the entire phase space with the exception of a set of data which lie on lower-dimensional surfaces

forming sets of zero volume). The notion is convenient also because it allows us to say that even the systems of free particles are integrable.

Two very remarkable systems integrable in the new sense are the Hamiltonian systems, respectively called *Toda lattice* (KRUSKAL, ZABUSKY), and *Calogero lattice* (CALOGERO, MOSER); if $(p_i, q_i) \in \mathbb{R}^2$, they are

$$\begin{aligned} \mathcal{H}_T(\boldsymbol{p},\boldsymbol{q}) &= \frac{1}{2m}\sum_{i=1}^{n} p_i^2 + \sum_{i=1}^{n-1} g\,\mathrm{e}^{-\kappa(q_{i+1}-q_i)} \\ \mathcal{H}_C(\boldsymbol{p},\boldsymbol{q}) &= \frac{1}{2m}\sum_{i=1}^{n} p_i^2 + \sum_{i<j}^{n} \frac{g}{(q_i-q_j)^2} \\ &\quad + \frac{1}{2}\sum_{i=1}^{n} m\omega^2 q_i^2 \end{aligned} \qquad [55]$$

where $m > 0$ and $\kappa, \omega, g \geq 0$. They describe the motion of $n$ interacting particles on a line.

The integration method for the above systems is again to find first the constants of motion and later to look for quadratures, when appropriate. The constants of motion can be found with the method of the *Lax pairs*. One shows that there is a pair of self-adjoint $n \times n$ matrices $M(\boldsymbol{p},\boldsymbol{q}), N(\boldsymbol{p},\boldsymbol{q})$ such that the equations of motion become

$$\frac{\mathrm{d}}{\mathrm{d}t} M(\boldsymbol{p},\boldsymbol{q}) = \mathrm{i}[M(\boldsymbol{p},\boldsymbol{q}), N(\boldsymbol{p},\boldsymbol{q})], \quad \mathrm{i} = \sqrt{-1} \qquad [56]$$

which imply that $M(t) = U(t)M(0)U(t)^{-1}$, with $U(t)$ a unitary matrix. When the equations can be written in the above form, it is clear that the $n$ eigenvalues of the matrix $M(0) = M(\boldsymbol{p}_0, \boldsymbol{q}_0)$ are constants of motion. When appropriate (e.g., in the Calogero lattice case with $\omega > 0$), it is possible to proceed to find canonical action–angle coordinates: a task that is quite difficult due to the arbitrariness of $n$, but which is possible.

The Lax pairs for the Calogero lattice (with $\omega = 0$, $g = m = 1$) are

$$\begin{aligned} M_{hh} &= p_h, \qquad N_{hh} = 0 \\ M_{hk} &= \frac{\mathrm{i}}{(q_h - q_k)}, \qquad N_{hk} = \frac{1}{(q_h - q_k)^2}\, h \neq k \end{aligned} \qquad [57]$$

while for the Toda lattice (with $m = g = \frac{1}{2}\kappa = 1$) the nonzero matrix elements of $M$, $N$ are

$$\begin{aligned} M_{hh} &= p_h, \qquad M_{h,h+1} = M_{h+1,h} = \mathrm{e}^{-(q_h - q_{h+1})} \\ N_{h,h+1} &= -N_{h+1,h} = \mathrm{i}\,\mathrm{e}^{-(q_h - q_{h+1})} \end{aligned} \qquad [58]$$

which are checked by first trying the case $n = 2$.

Another integrable system (SUTHERLAND) is

$$\mathcal{H}_S(\boldsymbol{p},\boldsymbol{q}) = \frac{1}{2m}\sum_{i=k}^{n} p_k^2 + \sum_{h<k}^{n} \frac{g}{\sinh^2(q_h - q_k)} \qquad [59]$$

whose Lax pair is related to that of the Calogero lattice.

By taking suitable limits as $n \to \infty$ and as the other parameters tend to 0 or $\infty$ at suitable rates, integrability of a few differential equations, among which the *Korteweg–deVries equation* or the nonlinear *Schrödinger equation*, can be derived.

As mentioned in the introductory section, symmetry properties under continuous groups imply existence of constants of motion. Hence, it is natural to think that integrability of a mechanical system reflects enough symmetry to imply the existence of as many constants of motion, independent and in involution, as the number of degrees of freedom, $n$.

This is in fact always true, and in some respects it is a tautological statement in the anisochronous cases. Integrability in a region $W$ implies existence of canonical action–angle coordinates $(\boldsymbol{A}, \boldsymbol{\alpha})$ (see the section "Quasiperiodicity and integrability") and the Hamiltonian depends solely on the $A$'s: therefore, its restriction to $W$ is invariant with respect to the action of the continuous commutative group $\mathcal{T}^n$ of the translations of the angle variables. The actions can be seen as constants of motion whose existence follows from Noether's theorem, at least in the anisochronous cases in which the Hamiltonian formulation is equivalent to a Lagrangian one.

What is nontrivial is to recognize, prior to realizing integrability, that a system admits this kind of symmetry: in most of the interesting cases, the systems either do not exhibit obvious symmetries or they exhibit symmetries apparently unrelated to the group $\mathcal{T}^n$, which nevertheless imply existence of sufficiently many independent constants of motion as required for integrability. Hence, nontrivial integrable systems possess a "hidden" symmetry under $\mathcal{T}^n$: the rigid body is an example.

However, very often the symmetries of a Hamiltonian $H$ which imply integrability also imply partial isochrony, that is, they imply that the number of independent frequencies is smaller than $n$ (see the section "Quasiperiodicity and integrability"). Even in such cases, often a map exists from the original coordinates $(\boldsymbol{p}, \boldsymbol{q})$ to the integrating variables $(\boldsymbol{A}, \boldsymbol{\alpha})$ in which $\boldsymbol{A}$ are constants of motion and the $\boldsymbol{\alpha}$ are uniformly rotating angles (some of which are also constant) with spectrum $\boldsymbol{\omega}(\boldsymbol{A})$, which is the gradient $\partial_{\boldsymbol{A}} h(\boldsymbol{A})$ for some function $h(\boldsymbol{A})$ depending only on a few of the $\boldsymbol{A}$ coordinates. However, the map might fail to be canonical. The system is then said to be bi-Hamiltonian: in the sense that one can represent motions in two systems of canonical coordinates, not related by a canonical transformation, and by two Hamiltonian functions $H$ and $H' \equiv h$ which generate the same motions in the respective

coordinates (the latter changes of variables are sometimes called "canonical with respect to the pair $H, H'$" while the transformations considered in the section "Canonical transformations of phase space coordination" are called completely canonical).

For more details, we refer the reader to Calogero and Degasperis (1982).

## Generic Nonintegrability

It is natural to try to prove that a system "close" to an integrable one has motions with properties very close to quasiperiodic. This is indeed the case, but in a rather subtle way. That there is a problem is easily seen in the case of a perturbation of an anisochronous integrable system.

Assume that a system is integrable in a region $W$ of phase space which, in the integrating action–angle variables $(\boldsymbol{A}, \boldsymbol{\alpha})$, has the standard form $U \times \mathbb{T}^\ell$ with a Hamiltonian $h(\boldsymbol{A})$ with gradient $\boldsymbol{\omega}(\boldsymbol{A}) = \partial_{\boldsymbol{A}} h(\boldsymbol{A})$. If the forces are perturbed by a potential which is smooth then the new system will be described, in the same coordinates, by a Hamiltonian like

$$\mathcal{H}_\varepsilon(\boldsymbol{A}, \boldsymbol{\alpha}) = h(\boldsymbol{A}) + \varepsilon f(\boldsymbol{A}, \boldsymbol{\alpha}) \qquad [60]$$

with $h, f$ analytic in the variables $\boldsymbol{A}, \boldsymbol{\alpha}$.

If the system really behaved like the unperturbed one, it ought to have $\ell$ constants of motion of the form $F_\varepsilon(\boldsymbol{A}, \boldsymbol{\alpha})$ analytic in $\varepsilon$ near $\varepsilon = 0$ and uniform, that is, single valued (which is the same as periodic) in the variables $\boldsymbol{\alpha}$. However, the following theorem (POINCARÉ) shows that this is a somewhat unlikely possibility.

**Theorem 1** *If the matrix* $\partial^2_{\boldsymbol{AA}} h(\boldsymbol{A})$ *has rank* $\geq 2$, *the Hamiltonian* [60] *"generically"* (an intuitive notion precised below) *cannot be integrated by a canonical transformation* $\mathcal{C}_\varepsilon(\boldsymbol{A}, \boldsymbol{\alpha})$ *which*

(i) *reduces to the identity as* $\varepsilon \to 0$; *and*
(ii) *is analytic in* $\varepsilon$ *near* $\varepsilon = 0$ *and in* $(\boldsymbol{A}, \boldsymbol{\alpha}) \in U' \times \mathbb{T}^\ell$, *with* $U' \subset U$ *open.*

*Furthermore, no uniform constants of motion* $F_\varepsilon(\boldsymbol{A}, \boldsymbol{\alpha})$, *defined for* $\varepsilon$ *near 0 and* $(\boldsymbol{A}, \boldsymbol{\alpha})$ *in an open domain* $U' \times \mathbb{T}^\ell$, *exist other than the functions of* $\mathcal{H}_\varepsilon$ *itself.*

Integrability in the sense (i), (ii) can be called analytic integrability and it is the strongest (and most naive) sense that can be given to the attribute.

The first part of the theorem, that is, (i), (ii), holds simply because, if integrability was assumed, a generating function of the integrating map would have the form $\boldsymbol{A}' \cdot \boldsymbol{\alpha} + \Phi_\varepsilon(\boldsymbol{A}', \boldsymbol{\alpha})$ with $\Phi$ admitting a power series expansion in $\varepsilon$ as $\Phi_\varepsilon = \varepsilon \Phi^1 + \varepsilon^2 \Phi^2 + \cdots$. Hence, $\Phi^1$ would have to satisfy

$$\boldsymbol{\omega}(\boldsymbol{A}') \cdot \partial_{\boldsymbol{\alpha}} \Phi^1(\boldsymbol{A}', \boldsymbol{\alpha}) + f(\boldsymbol{A}', \boldsymbol{\alpha}) = \overline{f}(\boldsymbol{A}') \qquad [61]$$

where $\overline{f}(\boldsymbol{A}')$ depends only on $\boldsymbol{A}'$ (hence integrating both sides with respect to $\boldsymbol{\alpha}$, it appears that $\overline{f}(\boldsymbol{A}')$ must coincide with the average of $f(\boldsymbol{A}', \boldsymbol{\alpha})$ over $\boldsymbol{\alpha}$).

This implies that the Fourier transform $f_{\boldsymbol{\nu}}(\boldsymbol{A})$, $\boldsymbol{\nu} \in \mathbb{Z}^\ell$, should satisfy

$$f_{\boldsymbol{\nu}}(\boldsymbol{A}') = 0 \qquad \text{if } \boldsymbol{\omega}(\boldsymbol{A}') \cdot \boldsymbol{\nu} = 0, \quad \boldsymbol{\nu} \neq \boldsymbol{0} \qquad [62]$$

which is equivalent to the existence of $\widetilde{f}_{\boldsymbol{\nu}}(\boldsymbol{A}')$ such that $f_{\boldsymbol{\nu}}(\boldsymbol{A}) = \boldsymbol{\omega}(\boldsymbol{A}') \cdot \boldsymbol{\nu} \widetilde{f}_{\boldsymbol{\nu}}(\boldsymbol{A})$ for $\boldsymbol{\nu} \neq \boldsymbol{0}$. But since there is no relation between $\boldsymbol{\omega}(\boldsymbol{A})$ and $f(\boldsymbol{A}, \boldsymbol{\alpha})$, this property "generically" will not hold in the sense that as close as wished to an $f$ which satisfies the property [62] there will be another $f$ which does not satisfy it essentially no matter how "closeness" is defined, (e.g., with respect to the metric $||f - g|| = \sum_{\boldsymbol{\nu}} |f_{\boldsymbol{\nu}}(\boldsymbol{A}) - g_{\boldsymbol{\nu}}(\boldsymbol{A})||$). This is so because the rank of $\partial^2_{\boldsymbol{AA}} h(\boldsymbol{A})$ is higher than 1 and $\boldsymbol{\omega}(\boldsymbol{A})$ varies at least on a two-dimensional surface, so that $\boldsymbol{\omega} \cdot \boldsymbol{\nu} = 0$ becomes certainly possible for some $\boldsymbol{\nu} \neq \boldsymbol{0}$ while $f_{\boldsymbol{\nu}}(\boldsymbol{A})$ in general will not vanish, so that $\Phi^1$, hence $\Phi_\varepsilon$, does not exist.

This means that close to a function $f$ there is a function $f'$ which violates [62] for some $\boldsymbol{\nu}$. Of course, this depends on what is meant by "close": however, here essentially any topology introduced on the space of the functions $f$ will make the statement correct. For instance, if the distance between two functions is defined by $\sum_{\boldsymbol{\nu}} \sup_{\boldsymbol{A} \in U} |f_{\boldsymbol{\nu}}(\boldsymbol{A}) - g_{\boldsymbol{\nu}}(\boldsymbol{A})|$ or by $\sup_{\boldsymbol{A}, \boldsymbol{\alpha}} |f(\boldsymbol{A}, \boldsymbol{\alpha}) - g(\boldsymbol{A}, \boldsymbol{\alpha})|$.

The idea behind the last statement of the theorem is in essence the same: consider, for simplicity, the anisochronous case in which the matrix $\partial^2_{\boldsymbol{AA}} h(\boldsymbol{A})$ has maximal rank $\ell$, that is, the determinant $\det \partial^2_{\boldsymbol{AA}} h(\boldsymbol{A})$ does not vanish. Anisochrony implies that $\boldsymbol{\omega}(\boldsymbol{A}) \cdot \boldsymbol{\nu} \neq 0$ for all $\boldsymbol{\nu} \neq \boldsymbol{0}$ and $\boldsymbol{A}$ on a dense set, and this property will be used repeatedly in the following analysis.

Let $B(\varepsilon, \boldsymbol{A}, \boldsymbol{\alpha})$ be a "uniform" constant of motion, meaning that it is single valued and analytic in the non-simply-connected region $U \times \mathbb{T}^\ell$ and, for $\varepsilon$ small,

$$B(\varepsilon, \boldsymbol{A}, \boldsymbol{\alpha}) = B_0(\boldsymbol{A}, \boldsymbol{\alpha}) + \varepsilon B_1(\boldsymbol{A}, \boldsymbol{\alpha}) + \varepsilon^2 B_2(\boldsymbol{A}, \boldsymbol{\alpha}) + \cdots \qquad [63]$$

The condition that $B$ is a constant of motion can be written order by order in its expansion in $\varepsilon$: the first two orders are

$$\begin{aligned} &\boldsymbol{\omega}(\boldsymbol{A}) \cdot \partial_{\boldsymbol{\alpha}} B_0(\boldsymbol{A}, \boldsymbol{\alpha}) = 0 \\ &\partial_{\boldsymbol{A}} f(\boldsymbol{A}, \boldsymbol{\alpha}) \cdot \partial_{\boldsymbol{\alpha}} B_0(\boldsymbol{A}, \boldsymbol{\alpha}) - \partial_{\boldsymbol{\alpha}} f(\boldsymbol{A}, \boldsymbol{\alpha}) \cdot \partial_{\boldsymbol{A}} B_0(\boldsymbol{A}, \boldsymbol{\alpha}) \\ &\quad + \boldsymbol{\omega}(\boldsymbol{A}) \cdot \partial_{\boldsymbol{\alpha}} B_1(\boldsymbol{A}, \boldsymbol{\alpha}) = 0 \end{aligned} \qquad [64]$$

Then the above two relations and anisochrony imply (1) that $B_0$ must be a function of $\boldsymbol{A}$ only and (2) that $\boldsymbol{\omega}(\boldsymbol{A})\cdot\boldsymbol{\nu}$ and $\partial_{\boldsymbol{A}}B_0(\boldsymbol{A})\cdot\boldsymbol{\nu}$ vanish simultaneously for all $\boldsymbol{\nu}$. Hence, the gradient of $B_0$ must be proportional to $\boldsymbol{\omega}(\boldsymbol{A})$, that is, to the gradient of $h(\boldsymbol{A}):\boldsymbol{\partial}_{\boldsymbol{A}}B_0(\boldsymbol{A})=\lambda(\boldsymbol{A})\boldsymbol{\partial}_{\boldsymbol{A}}h(\boldsymbol{A})$. Therefore, generically (because of the anisochrony) it must be that $B_0$ depends on $\boldsymbol{A}$ through $h(\boldsymbol{A}):B_0(\boldsymbol{A})=F(h(\boldsymbol{A}))$ for some $F$.

Looking again, with the new information, at the second of [64] it follows that at fixed $\boldsymbol{A}$ the $\boldsymbol{\alpha}$-derivative in the direction $\boldsymbol{\omega}(\boldsymbol{A})$ of $B_1$ equals $F'(h(\boldsymbol{A}))$ times the $\boldsymbol{\alpha}$-derivative of $f$, that is, $B_1(\boldsymbol{A},\boldsymbol{\alpha})=f(\boldsymbol{A},\boldsymbol{\alpha})F'(h(\boldsymbol{A}))+C_1(\boldsymbol{A})$.

Summarizing: the constant of motion $B$ has been written as $B(\boldsymbol{A},\boldsymbol{\alpha})=F(h(\boldsymbol{A}))+\varepsilon F'(h(\boldsymbol{A}))f(\boldsymbol{A},\boldsymbol{\alpha})+\varepsilon C_1(\boldsymbol{A})+\varepsilon^2B_2+\cdots$ which is equivalent to $B(\boldsymbol{A},\boldsymbol{\alpha})=F(\mathcal{H}_\varepsilon)+\varepsilon(B_0'+\varepsilon B_1'+\cdots)$ and therefore $B_0'+\varepsilon B_1'+\cdots$ is another analytic constant of motion. Repeating the argument also $B_0'+\varepsilon B_1'+\cdots$ must have the form $F_1(\mathcal{H}_\varepsilon)+\varepsilon(B_0''+\varepsilon B_1''+\cdots)$; conclusion

$$B=F(\mathcal{H}_\varepsilon)+\varepsilon F_1(\mathcal{H}_\varepsilon)+\varepsilon^2F_2(\mathcal{H}_\varepsilon)+\cdots+\varepsilon^nF_n(\mathcal{H}_\varepsilon)+O(\varepsilon^{n+1}) \qquad [65]$$

By analyticity, $B=F_\varepsilon(\mathcal{H}_\varepsilon(\boldsymbol{A},\boldsymbol{\alpha}))$ for some $F_\varepsilon$: hence generically all constants of motion are trivial.

Therefore, a system close to integrable cannot behave as it would naively be expected. The problem, however, was not manifest until Poincaré's proof of the above results: because in most applications the function $f$ has only finitely many Fourier components, or at least is replaced by an approximation with this property, so that at least [62] and even a few of the higher-order constraints like [64] become possible in open regions of action space. In fact, it may happen that the values of $\boldsymbol{A}$ of interest are restricted so that $\boldsymbol{\omega}(\boldsymbol{A})\cdot\boldsymbol{\nu}=0$ only for "large" values of $\boldsymbol{\nu}$ for which $f_{\boldsymbol{\nu}}=0$. Nevertheless, the property that $f_{\boldsymbol{\nu}}(\boldsymbol{A})=(\boldsymbol{\omega}(\boldsymbol{A})\cdot\boldsymbol{\nu})\widetilde{f}_{\boldsymbol{\nu}}(\boldsymbol{A})$ (or the analogous higher-order conditions, e.g., [64]), which we have seen to be necessary for analytic integrability of the perturbed system, can be checked to fail in important problems, if no approximation is made on $f$. Hence a conceptual problem arises.

For more details see Poincaré (1987).

## Perturbing Functions

To check, in a given problem, the nonexistence of nontrivial constants of motion along the lines indicated in the previous section, it is necessary to express the potential, usually given in Cartesian coordinates as $\varepsilon V(\boldsymbol{x})$, in terms of the action–angle variables of the unperturbed, integrable, system.

In particular, the problem arises when trying to check nonexistence of nontrivial constants of motion when the anisochrony assumption (cf. the previous section) is not satisfied. Usually it becomes satisfied "to second order" (or higher): but to show this, a more detailed information on the structure of the perturbing function expressed in action–angle variables is needed. For instance, this is often necessary even when the perturbation is approximated by a trigonometric polynomial, as it is essentially always the case in celestial mechanics.

Finding explicit expressions for the action–angle variables is in itself a rather nontrivial task which leads to many problems of intrinsic interest even in seemingly simple cases. For instance, in the case of the planar gravitational central motion, the Kepler equation $\lambda=\xi-\varepsilon\sin\xi$ (see the first of [41]) must be solved expressing $\xi$ in terms of $\lambda$ (see the first of [42]). It is obvious that for small $\varepsilon$, the variable $\xi$ can be expressed as an analytic function of $\varepsilon$: nevertheless, the actual construction of this expression leads to several problems. For small $\varepsilon$, an interesting algorithm is the following.

Let $h(\lambda)=\xi-\lambda$, so that the equation to solve (i.e., the first of [41]) is

$$h(\lambda)=\varepsilon\sin(\lambda+h(\lambda))\equiv-\varepsilon\frac{\partial c}{\partial\lambda}(\lambda+h(\lambda)) \qquad [66]$$

where $c(\lambda)=\cos\lambda$; the function $\lambda\to h(\lambda)$ should be periodic in $\lambda$, with period $2\pi$, and analytic in $\varepsilon,\lambda$ for $\varepsilon$ small and $\lambda$ real. If $h(\lambda)=\varepsilon h^{(1)}+\varepsilon^2h^{(2)}+\cdots$, the Fourier transform of $h^{(k)}(\lambda)$ satisfies the recursion relation

$$h_\nu^{(k)}=-\sum_{p=1}^{\infty}\frac{1}{p!}\sum_{\substack{k_1+\cdots+k_p=k-1\\ \nu_0+\nu_1+\cdots+\nu_p=\nu}}(\mathrm{i}\nu_0)c_{\nu_0}(\mathrm{i}\nu_0)^p\times\prod h_{\nu_j}^{(k_j)},\quad k>1 \qquad [67]$$

with $c_\nu$ the Fourier transform of the cosine ($c_{\pm1}=\frac{1}{2}$, $c_\nu=0$ if $\nu\neq\pm1$), and (of course) $h_\nu^{(1)}=-\mathrm{i}\nu c_\nu$. Equation [67] is obtained by expanding the RHS of [66] in powers of $h$ and then taking the Fourier transform of both sides retaining only terms of order $k$ in $\varepsilon$.

Iterating the above relation, imagine drawing all trees $\theta$ with $k$ "branches," or "lines," distinguished by a label taking $k$ values, and $k$ nodes and attach to each node $v$ a harmonic label $\nu_v=\pm1$ as in Figure 5. The trees will be assumed to start with a root line $vr$ linking a point $r$ and the "first node" $v$ (see **Figure 5**)

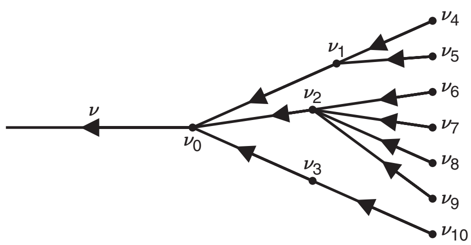



**Figure 5** An example of a tree graph and its labels. It contains only one simple node (3). Harmonics are indicated next to their nodes. Labels distinguishing lines are not marked.

and then bifurcate arbitrarily (such trees are sometimes called "rooted trees").

Imagine the tree oriented from the endpoints towards the root $r$ (not to be considered a node) and given a node $v$ call $v'$ the node immediately following it. If $v$ is the first node before the root $r$, let $v' = r$ and $\nu_{v'} = 1$. For each such decorated tree define its numerical value

$$\mathrm{Val}(\theta) = \frac{-\mathrm{i}}{k!} \prod_{\text{lines } l=v'v} (\nu_{v'}\nu_v) \prod_{\text{nodes}} c_{\nu_v} \qquad [68]$$

and define a current $\nu(l)$ on a line $l = v'v$ to be the sum of the harmonics of the nodes preceding $v'$: $\nu(l) = \sum_{w\le v} \nu_v$. Call $\nu(\theta)$ the current flowing in the root branch and call order of $\theta$ the number of nodes (or branches). Then

$$h_\nu^{(k)} = \sum_{\substack{\theta,\nu(\theta)=\nu \\ \mathrm{order}(\theta)=k}} \mathrm{Val}(\theta) \qquad [69]$$

provided trees are considered identical if they can be overlapped (labels included) after suitably scaling the lengths of their branches and pivoting them around the nodes out of which they emerge (the root is always imagined to be fixed at the origin).

If the trees are stripped of the harmonic labels, their number is finite and it can be estimated to be $\le k!4^k$ (because the labels which distinguish the lines can be attached to an unlabeled tree in many ways). The harmonic labels (i.e., $\nu_v = \pm 1$) can be laid down in $2^k$ ways, and the value of each tree can be bounded by $(1/k!)2^{-k}$ (because $c_{\pm1} = \frac{1}{2}$).

Hence $\sum_\nu |h_\nu^{(k)}| \le 4^k$, which gives a (rough) estimate of the radius of convergence of the expansion of $h$ in powers of $\varepsilon$: namely 0.25 (easily improvable to 0.3678 if $4^k k!$ is replaced by $k^{k-1}$ using Cayley's formula for the enumeration of rooted trees). A simple expression for $h^{(k)}(\psi)$ (Lagrange) is

$$h^{(k)}(\psi) = \frac{1}{k!}\partial_\psi^{k-1} \sin^k \psi$$

(also readable from the tree representation): the actual radius of convergence, first determined by Laplace, of the series for $h$ can also be determined from the latter expression for $h$ (Rouché) or directly from the tree representation: it is $\sim 0.6627$.

One can find better estimates or at least more efficient methods for evaluating the sums in [69]: in fact, in performing the sum in [69] important cancellations occur. For instance, the harmonic labels can be subject to the further strong constraint that no line carries zero current because the sum of the values of the trees of fixed order and with at least one line carrying zero current vanishes.

The above expansion can also be simplified by *partial resummations*. For the purpose of an example, let the nodes with one entering and one exiting line (see Figure 5) be called as "simple" nodes. Then all tree graphs which, on any line between two nonsimple nodes, contain any number of simple nodes can be eliminated. This is done by replacing, in evaluating the (remaining) tree values, the factors $\nu_{v'}\nu_v$ in [68] by $\nu_{v'}\nu_v/(1-\varepsilon\cos\psi)$: then the value of $\theta$ (denoted $\mathrm{Val}(\theta)_\psi$) for a tree becomes a function of $\psi$ and $\varepsilon$ and [69] is replaced by

$$h(\psi) = \sum_{k=1}^{\infty} \sum_{\substack{\theta,\,\nu(\theta)=\nu \\ \mathrm{order}(\theta)=k}}^{*} \varepsilon^k \mathrm{e}^{\mathrm{i}\nu\psi} \mathrm{Val}(\theta)_\psi \qquad [70]$$

where the $*$ means that the trees are subject to the further restriction of not containing any simple node. It should be noted that the above graphical representation of the solution of the Kepler equation is strongly reminiscent of the representations of quantities in terms of graphs that occur often in quantum field theory. Here the trees correspond to *Feynman graphs*, the factors associated with the nodes are the *couplings*, the factors associated with the lines are the *propagators*, and the resummations are analogous to the *self-energy resummations*, while the cancellations mentioned above can be related to the class of identities called *Ward identities*. Not only the analogy can be shown not to be superficial, but it also turns out to be very helpful in key mechanical problems: see Appendix 1.

The existence of a vast number of identities relating the tree values is shown already by the *simple* form of the Lagrange series and by the even more remarkable resummation (Levi-Civita) leading to

$$h(\psi) = \sum_{k=1}^{\infty} \frac{(\varepsilon\sin\psi)^k}{k!} \left(\frac{1}{1-\varepsilon\cos\psi}\partial_\psi\right)^k \psi \qquad [71]$$

It is even possible to further collect the series terms to express it as a series with much better convergence properties; for instance, its terms can be reorganized and collected (resummed) so that $h$ is expressed as a power series in the parameter

$$\eta = \frac{\varepsilon\, e^{\sqrt{1-\varepsilon^2}}}{1+\sqrt{1-\varepsilon^2}} \qquad [72]$$

with radius of convergence 1, which corresponds to $\varepsilon = 1$ (via a simple argument by Levi-Civita). The analyticity domain for the Lagrange series is $|\eta| < 1$. This also determines the value of Laplace radius, which is the point closest to the origin of the complex curve $|\eta(\varepsilon)| = 1$: it is imaginary so that it is the root of the equation

$$\varepsilon e^{\sqrt{1+\varepsilon^2}}/(1+\sqrt{1+\varepsilon^2}) = 1$$

The analysis provides an example, in a simple case of great interest in applications, of the kind of computations actually necessary to represent the perturbing function in terms of action–angle variables. The property that the function $c(\lambda)$ in [66] is the cosine has been used only to limit the range of the label $\nu$ to be $\pm 1$; hence the same method, with similar results, can be applied to study the inversion of the relation between the average anomaly $\lambda$ and the true anomaly $\theta$ and to efficiently obtain, for instance, the properties of $f$, $g$ in [42].

For more details, the reader is referred to Levi-Civita (1956).

## Lindstedt and Birkhoff Series: Divergences

Nonexistence of constants of motion, rather than being the end of the attempts to study motions close to integrable ones by perturbation methods, marks the beginning of renewed efforts to understand their nature.

Let $(\boldsymbol{A}, \boldsymbol{\alpha}) \in U \times \mathbb{T}^\ell$ be action–angle variables defined in the integrability region for an analytic Hamiltonian and let $h(\boldsymbol{A})$ be its value in the action–angle coordinates. Suppose that $h(\boldsymbol{A}_0)$ is anisochronous and let $f(\boldsymbol{A}, \boldsymbol{\alpha})$ be an analytic perturbing function. Consider, for $\varepsilon$ small, the Hamiltonian $\mathcal{H}_\varepsilon(\boldsymbol{A}, \boldsymbol{\alpha}) = \mathcal{H}_0(\boldsymbol{A}) + \varepsilon f(\boldsymbol{A}, \boldsymbol{\alpha})$.

Let $\boldsymbol{\omega}_0 = \boldsymbol{\omega}(\boldsymbol{A}_0) \equiv \boldsymbol{\partial}_A \mathcal{H}_0(\boldsymbol{A})$ be the frequency spectrum (see the section "Quasiperiodicity and integrability") of one of the invariant tori of the unperturbed system corresponding to an action $\boldsymbol{A}_0$. Short of integrability, the question to ask at this point is whether the perturbed system admits an analytic invariant torus on which the motion is quasiperiodic and

1. has the same spectrum $\boldsymbol{\omega}_0$,
2. depends analytically on $\varepsilon$ at least for $\varepsilon$ small,
3. reduces to the "unperturbed torus" $\{\boldsymbol{A}_0\} \times \mathbb{T}^\ell$ as $\varepsilon \to 0$.

More concretely, the question is:

> Are there functions $\boldsymbol{H}_\varepsilon(\boldsymbol{\psi}), \boldsymbol{h}_\varepsilon(\boldsymbol{\psi})$ analytic in $\boldsymbol{\psi} \in \mathbb{T}^\ell$ and in $\varepsilon$ near 0, vanishing as $\varepsilon \to 0$ and such that the torus with parametric equations
>
> $$\boldsymbol{A} = \boldsymbol{A}_0 + \boldsymbol{H}_\varepsilon(\boldsymbol{\psi}), \ \boldsymbol{\alpha} = \boldsymbol{\psi} + \boldsymbol{h}_\varepsilon(\boldsymbol{\psi}), \quad \boldsymbol{\psi} \in \mathbb{T}^\ell \quad [73]$$
>
> is invariant and, if $\boldsymbol{\omega}_0 \stackrel{\text{def}}{=} \boldsymbol{\omega}(\boldsymbol{A}_0)$, the motion on it is simply $\boldsymbol{\psi} \to \boldsymbol{\psi} + \boldsymbol{\omega}_0 t$, i.e., it is quasiperiodic with spectrum $\boldsymbol{\omega}_0$?

In this context, Poincaré's theorem (in the section "Generic nonintegrability") had followed another key result, earlier developed in particular cases and completed by him, which provides a partial answer to the question.

Suppose that $\boldsymbol{\omega}_0 = \boldsymbol{\omega}(\boldsymbol{A}_0) \in \mathbb{R}^\ell$ satisfies a Diophantine property, namely suppose that there exist constants $C, \tau > 0$ such that

$$|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}| \geq \frac{1}{C|\boldsymbol{\nu}|^\tau}, \quad \text{for all } \boldsymbol{0} \neq \boldsymbol{\nu} \in \mathbb{Z}^\ell \qquad [74]$$

which, for each $\tau > \ell - 1$ fixed, is a property enjoyed by all $\boldsymbol{\omega} \in \mathbb{R}^\ell$ but for a set of zero measure. Then the motions on the unperturbed torus run over trajectories that fill the torus densely because of the "irrationality" of $\boldsymbol{\omega}_0$ implied by [74]. Writing Hamilton's equations,

$$\dot{\boldsymbol{\alpha}} = \partial_A \mathcal{H}_0(\boldsymbol{A}) + \varepsilon \boldsymbol{\partial}_A f(\boldsymbol{A}, \boldsymbol{\alpha}), \quad \dot{\boldsymbol{A}} = -\varepsilon \boldsymbol{\partial}_\alpha f(\boldsymbol{A}, \boldsymbol{\alpha})$$

with $\boldsymbol{A}, \boldsymbol{\alpha}$ given by [73] with $\boldsymbol{\psi}$ replaced by $\boldsymbol{\psi} + \boldsymbol{\omega} t$, and using the density of the unperturbed trajectories implied by [74], the condition that [73] are equations for an invariant torus on which the motion is $\boldsymbol{\psi} \to \boldsymbol{\psi} + \boldsymbol{\omega}_0 t$ are

$$\begin{aligned} \boldsymbol{\omega}_0 + (\boldsymbol{\omega}_0 \cdot \boldsymbol{\partial}_\psi)\boldsymbol{h}_\varepsilon(\boldsymbol{\psi}) &= \boldsymbol{\partial}_A \mathcal{H}_0(\boldsymbol{A}_0 + \boldsymbol{H}_\varepsilon(\boldsymbol{\psi})) \\ &\quad + \varepsilon \boldsymbol{\partial}_A f(\boldsymbol{A}_0 + \boldsymbol{H}_\varepsilon(\boldsymbol{\psi}), \boldsymbol{\psi} + \boldsymbol{h}_\varepsilon(\boldsymbol{\psi}))(\boldsymbol{\omega}_0 \cdot \boldsymbol{\partial}_\psi)\boldsymbol{H}_\varepsilon(\boldsymbol{\psi}) \\ &= -\varepsilon \boldsymbol{\partial}_\alpha f(\boldsymbol{A}_0 + \boldsymbol{H}_\varepsilon(\boldsymbol{\psi}), \boldsymbol{\psi} + \boldsymbol{h}_\varepsilon(\boldsymbol{\psi})) \end{aligned} \qquad [75]$$

The theorem referred to above (Poincaré) is that

**Theorem 2** *If the unperturbed system is anisochronous and $\boldsymbol{\omega}_0 = \boldsymbol{\omega}(\boldsymbol{A}_0)$ satisfies [74] for some $C, \tau > 0$ there exist two well defined power series $\boldsymbol{h}_\varepsilon(\boldsymbol{\psi}) = \sum_{k=1}^\infty \varepsilon^k \boldsymbol{h}^{(k)}(\boldsymbol{\psi})$ and $\boldsymbol{H}_\varepsilon(\boldsymbol{\psi}) = \sum_{k=1}^\infty \varepsilon^k \boldsymbol{H}^{(k)}(\boldsymbol{\psi})$ which*

*solve [75] to all orders in $\varepsilon$. The series for $\boldsymbol{H}_\varepsilon$ is uniquely determined, and such is also the series for $\boldsymbol{h}_\varepsilon$ up to the addition of an arbitrary constant at each order, so that it is unique if $\boldsymbol{h}_\varepsilon$ is required, as henceforth done with no loss of generality, to have zero average over $\boldsymbol{\psi}$.*

The algorithm for the construction is illustrated in a simple case in the next section (see eqns [83], [84]). Convergence of the above series, called *Lindstedt series*, even for small $\varepsilon$ has been a problem for rather a long time. Poincaré proved the existence of the formal solution; but his other result, discussed in the section "Generic nonintegrability," casts doubts on convergence although it does not exclude it, as was immediately stressed by several authors (including Poincaré himself). The result in that section shows the impossibility of solving [75] for all $\boldsymbol{\omega}_0$'s near a given spectrum, analytically and uniformly, but it does not exclude the possibility of solving it for a single $\boldsymbol{\omega}_0$.

The theorem admits several extensions or analogs: an interesting one is to the case of isochronous unperturbed systems:

> Given the Hamiltonian $\mathcal{H}_\varepsilon(\boldsymbol{A},\boldsymbol{\alpha}) = \boldsymbol{\omega}_0 \cdot \boldsymbol{A} + \varepsilon f(\boldsymbol{A},\boldsymbol{\alpha})$, with $\boldsymbol{\omega}_0$ satisfying [74] and $f$ analytic, there exist power series $\mathcal{C}_\varepsilon(\boldsymbol{A}',\boldsymbol{\alpha}')$, $u_\varepsilon(\boldsymbol{A}')$ such that $\mathcal{H}_\varepsilon(\mathcal{C}_\varepsilon(\boldsymbol{A}',\boldsymbol{\alpha}')) = \boldsymbol{\omega}_0 \cdot \boldsymbol{A}' + u_\varepsilon(\boldsymbol{A}')$ holds as an equality between formal power series (i.e., order by order in $\varepsilon$) and at the same time the $\mathcal{C}_\varepsilon$, regarded as a map, satisfies order by order the condition (i.e., (4.3)) that it is a canonical map.

This means that there is a generating function $\boldsymbol{A}' \cdot \boldsymbol{\alpha} + \boldsymbol{\Phi}_\varepsilon(\boldsymbol{A}',\boldsymbol{\alpha})$ also defined by a formal power series $\boldsymbol{\Phi}_\varepsilon(\boldsymbol{A}',\boldsymbol{\alpha}) = \sum_{k=1}^{\infty} \varepsilon^k \boldsymbol{\Phi}^{(k)}(\boldsymbol{A}',\boldsymbol{\alpha})$, that is, such that if $\mathcal{C}_\varepsilon(\boldsymbol{A}',\boldsymbol{\alpha}') = (\boldsymbol{A},\boldsymbol{\alpha})$ then it is true, order by order in powers of $\varepsilon$, that $\boldsymbol{A} = \boldsymbol{A}' + \boldsymbol{\partial}_{\boldsymbol{\alpha}} \boldsymbol{\Phi}_\varepsilon(\boldsymbol{A}',\boldsymbol{\alpha})$ and $\boldsymbol{\alpha}' = \boldsymbol{\alpha} + \boldsymbol{\partial}_{\boldsymbol{A}'} \boldsymbol{\Phi}_\varepsilon(\boldsymbol{A}',\boldsymbol{\alpha})$. The series for $\boldsymbol{\Phi}_\varepsilon, u_\varepsilon$ are called Birkhoff series.

In this isochronous case, if Birkhoff series were convergent for small $\varepsilon$ and $(\boldsymbol{A}',\boldsymbol{\alpha})$ in a region of the form $U \times \mathbb{T}^\ell$, with $U \subset \mathbb{R}^\ell$ open and bounded, it would follow that, for small $\varepsilon, \mathcal{H}_\varepsilon$ would be integrable in a large region of phase space (i.e., where the generating function can be used to build a canonical map: this would essentially be $U \times \mathbb{T}^\ell$ deprived of a small layer of points near the boundary of $U$). However, convergence for small $\varepsilon$ is false (in general), as shown by the simple two-dimensional example

$$\mathcal{H}_\varepsilon(\boldsymbol{A},\boldsymbol{\alpha}) = \boldsymbol{\omega}_0 \cdot \boldsymbol{A} + \varepsilon\,(A_2 + f(\boldsymbol{\alpha})) \qquad (\boldsymbol{A},\boldsymbol{\alpha}) \in \mathbb{R}^2 \times \mathbb{T}^2 \qquad [76]$$

with $f(\boldsymbol{\alpha})$ an arbitrary analytic function with all Fourier coefficients $f_{\boldsymbol{\nu}}$ positive for $\boldsymbol{\nu} \neq 0$ and $f_{\mathbf{o}} = 0$. In the latter case, the solution is

$$u_\varepsilon(\boldsymbol{A}') = \varepsilon A_2$$

$$\boldsymbol{\Phi}_\varepsilon(\boldsymbol{A}',\boldsymbol{\alpha}) = \sum_{k=1}^{\infty} \varepsilon^k \sum_{0 \neq \boldsymbol{\nu} \in \mathbb{Z}^2} f_{\boldsymbol{\nu}}\, \mathrm{e}^{\mathrm{i}\boldsymbol{\alpha}\cdot\boldsymbol{\nu}} \frac{(\mathrm{i}\,\nu_2)^k}{(\mathrm{i}(\omega_{01}\nu_1 + \omega_{02}\nu_2))^{k+1}} \qquad [77]$$

The series does not converge: in fact, its convergence would imply integrability and, consequently, bounded trajectories in phase space: however, the equations of motion for [76] can be easily solved explicitly and in any open region near given initial data there are other data which have unbounded trajectories if $\omega_{01}/(\omega_{02} + \varepsilon)$ is rational.

Nevertheless, even in this elementary case a formal sum of the series yields

$$u(\boldsymbol{A}') = \varepsilon A_2'$$

$$\boldsymbol{\Phi}_\varepsilon(\boldsymbol{A}',\boldsymbol{\alpha}) = \varepsilon \sum_{0 \neq \boldsymbol{\nu} \in \mathbb{Z}^2} \frac{f_{\boldsymbol{\nu}}\, \mathrm{e}^{\mathrm{i}\boldsymbol{\alpha}\cdot\boldsymbol{\nu}}}{\mathrm{i}(\omega_{01}\nu_1 + (\omega_{20} + \varepsilon)\nu_2)} \qquad [78]$$

and the series in [78] (no longer a power series in $\varepsilon$) is really convergent if $\boldsymbol{\omega} = (\omega_{01}, \omega_{02} + \varepsilon)$ is a Diophantine vector (by [74], because analyticity implies exponential decay of $|f_{\boldsymbol{\nu}}|$). Remarkably, for such values of $\varepsilon$ the Hamiltonian $\mathcal{H}_\varepsilon$ is integrable and it is integrated by the canonical map generated by [78], in spite of the fact that [78] is obtained, from [77], via the nonrigorous sum rule

$$\sum_{k=0}^{\infty} z^k = \frac{1}{1-z} \quad \text{for } z \neq 1 \qquad [79]$$

(applied to cases with $|z| \geq 1$, which are certainly realized for a dense set of $\varepsilon$'s even if $\boldsymbol{\omega}$ is Diophantine because the $z$'s have values $z = \nu_2/\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}$). In other words, the integration of the equations is elementary and once performed it becomes apparent that, if $\boldsymbol{\omega}$ is diophantine, the solutions can be rigorously found from [78]. Note that, for instance, this means that relations like $\sum_{k=0}^{\infty} 2^k = -1$ are really used to obtain [78] from [77].

Another extension of Lindstedt series arises in a perturbation of an anisochronous system when asking the question as to what happens to the unperturbed invariant tori $\mathcal{T}_{\boldsymbol{\omega}_0}$ on which the spectrum is resonant, that is, $\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu} = 0$ for some $\boldsymbol{\nu} \neq 0$, $\boldsymbol{\nu} \in \mathbb{Z}^\ell$. The result is that even in such a case there is a formal power series solution showing that at least a few of the (infinitely many) invariant tori into which $\mathcal{T}_{\boldsymbol{\omega}_0}$ is in turn foliated in the unperturbed case can be formally continued at $\varepsilon \neq 0$ (see the section "Resonances and their stability").

For more details, we refer the reader to Poincaré (1987).

## Quasiperiodicity and KAM Stability

To discuss more advanced results, it is convenient to restrict attention to a special (nontrivial) paradigmatic case

$$\mathcal{H}_\varepsilon(A,\boldsymbol{\alpha}) = \tfrac{1}{2}A^2 + \varepsilon f(\boldsymbol{\alpha}) \qquad [80]$$

In this simple case (called *Thirring model*: representing $\ell$ particles on a circle interacting via a potential $\varepsilon f(\boldsymbol{\alpha})$) the equations for the maximal tori [75] reduce to equations for the only functions $\boldsymbol{h}_\varepsilon$:

$$(\boldsymbol{\omega}\cdot\boldsymbol{\partial}_{\boldsymbol{\psi}})^2\boldsymbol{h}_\varepsilon(\boldsymbol{\psi}) = -\varepsilon\boldsymbol{\partial}_{\boldsymbol{\alpha}}f(\boldsymbol{\psi}+\boldsymbol{h}_\varepsilon(\boldsymbol{\psi})), \quad \boldsymbol{\psi}\in\mathbb{T}^\ell \qquad [81]$$

as the second of [75] simply becomes the definition of $\boldsymbol{H}_\varepsilon$ because the RHS does not involve $\boldsymbol{H}_\varepsilon$.

The real problem is therefore whether the formal series considered in the last section converge at least for small $\varepsilon$: and the example [76] on the Birkhoff series shows that sometimes sum rules might be needed in order to give a meaning to the series. In fact, whenever a problem (of physical interest) admits a formal power series solution which is not convergent, or which is such that it is not known whether it is convergent, then one should look for sum rules for it.

The modern theory of perturbations starts with the proof of the convergence for $\varepsilon$ small enough of the Lindstedt series (Kolmogorov). The general "KAM" result is:

**Theorem 3** (KAM) *Consider the Hamiltonian* $\mathcal{H}_\varepsilon(\boldsymbol{A},\boldsymbol{\alpha}) = h(\boldsymbol{A}) + \varepsilon f(\boldsymbol{A},\boldsymbol{\alpha})$, *defined in* $U = V\times\mathbb{T}^\ell$ *with* $V\subset\mathbb{R}^\ell$ *open and bounded and with* $f(\boldsymbol{A},\boldsymbol{\alpha})$, $h(\boldsymbol{A})$ *analytic in the closure* $\overline{V}\times\mathbb{T}^\ell$ *where* $h(\boldsymbol{A})$ *is also anisochronous; let* $\boldsymbol{\omega}_0 \stackrel{\text{def}}{=} \boldsymbol{\omega}(\boldsymbol{A}_0) = \partial_{\boldsymbol{A}}h(\boldsymbol{A}_0)$ *and assume that* $\boldsymbol{\omega}_0$ *satisfies* [74]. *Then*

- (i) *there is* $\varepsilon_{C,\tau} > 0$ *such that the Lindstedt series converges for* $|\varepsilon| < \varepsilon_{C,\tau}$;
- (ii) *its sum yields two function* $\boldsymbol{H}_\varepsilon(\boldsymbol{\psi}), \boldsymbol{h}_\varepsilon(\boldsymbol{\psi})$ *on* $\mathbb{T}^\ell$ *which parametrize an invariant torus* $\mathcal{T}_{C,\tau}(\boldsymbol{A}_0,\varepsilon)$;
- (iii) *on* $\mathcal{T}_{C,\tau}(\boldsymbol{A}_0,\varepsilon)$ *the motion is* $\boldsymbol{\psi}\to\boldsymbol{\psi}+\boldsymbol{\omega}_0 t$, *see* [73]; *and*
- (iv) *the set of data in* $U$ *which belong to invariant tori* $\mathcal{T}_{C,\tau}(\boldsymbol{A}_0,\varepsilon)$ *with* $\boldsymbol{\omega}(\boldsymbol{A}_0)$ *satisfying* [74] *with prefixed* $C,\tau$ *has complement with volume* $<$const $C^{-a}$ *for a suitable* $a>0$ *and with area also* $<$const $C^{-a}$ *on each nontrivial surface of constant energy* $\mathcal{H}_\varepsilon = E$.

In other words, for small $\varepsilon$ the spectra of most unperturbed quasiperiodic motions can still be found as spectra of perturbed quasiperiodic motions developing on tori which are close to the corresponding unperturbed ones (i.e., with the same spectrum).

This is a stability result: for instance, in systems with two degrees of freedom the invariant tori of dimension two which lie on a given three-dimensional energy surface, will separate the points on the energy surface into the set which is "inside" the torus and the set which is "outside." Hence, an initial datum starting (say) inside cannot reach the outside. Likewise, a point starting between two tori has to stay in between forever. Further, if the two tori are close, this means that motion will stay very localized in action space, with a trajectory accessing only points close to the tori and coming close to all such points, within a distance of the order of the distance between the confining tori. The case of three or more degrees of freedom is quite different (see sections "Diffusion in phase space" and "The three-body problem").

In the simple case of the rotators system [80] the equations for the parametric representation of the tori are given by [81]. The latter bear some analogy with the easier problem in [66]: but [81] are $\ell$ equations instead of one and they are differential equations rather than ordinary equations. Furthermore, the function $f(\boldsymbol{\alpha})$ which plays here the role of $c(\lambda)$ in [66] has Fourier coefficient $f_{\boldsymbol{\nu}}$ with no restrictions on $\boldsymbol{\nu}$, while the Fourier coefficients $c_\nu$ for $c$ in [66] do not vanish only for $\nu = \pm 1$.

The above differences are, to some extent, "minor" and the power series solution to [81] can be constructed by the same algorithm as used in the case of [66]: namely one forms trees as in Figure 5 with the harmonic labels $\nu_v\in\mathbb{Z}$ replaced by $\boldsymbol{\nu}_v\in\mathbb{Z}^\ell$ (still to be thought of as possible harmonic indices in the Fourier expansion of the perturbing function $f$). All other labels affixed to the trees in the section "Generic nonintegrability" will be the same. In particular, the current flowing on a branch $l = v'v$ will be defined as the sum of the harmonics of the nodes $w\le v$ preceding $v$:

$$\boldsymbol{\nu}(l) \stackrel{\text{def}}{=} \sum_{w\le v}\boldsymbol{\nu}_w \qquad [82]$$

and we call $\boldsymbol{\nu}(\theta)$ the current flowing in the root branch.

Here the value $\mathrm{Val}(\theta)$ of a tree has to be defined differently because the equation to be solved ([81]) contains the differential operator $(\boldsymbol{\omega}_0\cdot\boldsymbol{\partial}_{\boldsymbol{\psi}})^2$ which, when Fourier transformed, becomes multiplication of the Fourier component with harmonic $\boldsymbol{\nu}$ by $(\mathrm{i}\boldsymbol{\omega}\cdot\boldsymbol{\nu})^2$.

The variation due to the presence of the operator $(\boldsymbol{\omega}_0\cdot\boldsymbol{\partial}_{\boldsymbol{\psi}})^2$ and the necessity of its inversion in the evaluation of $\boldsymbol{u}\cdot\boldsymbol{h}^{(k)}_{\boldsymbol{\nu}}$, that is, of the component of $\boldsymbol{h}^{(k)}_{\boldsymbol{\nu}}$ along an arbitrary unit vector $\boldsymbol{u}$, is nevertheless quite simple: the value of a tree graph $\theta$ of order $k$

(i.e., with $k$ nodes and $k$ branches) has to be defined by (cf. [68])

$$\mathrm{Val}(\theta) \stackrel{\text{def}}{=} \frac{-\mathrm{i}(-1)^k}{k!}\left(\prod_{\text{lines } l=v'v} \frac{\boldsymbol{\nu}_{v'}\cdot\boldsymbol{\nu}_v}{(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}(l))^2}\right) \times \left(\prod_{\text{nodes } v} f_{\boldsymbol{\nu}_v}\right) \qquad [83]$$

where the $\boldsymbol{\nu}_{v'}$ appearing in the factor relative to the root line $rv$ from the first node $v$ to the root $r$ (see Figure 5) is interpreted as a unit vector $\boldsymbol{u}$ (it was interpreted as 1 in the one-dimensional case [66]). Equation [83] makes sense only for trees in which no line carries zero current. Then the component along $\boldsymbol{u}$ (the harmonic label attached to the root of a tree) of $\boldsymbol{h}^{(k)}$ is given (see also [69]) by

$$\boldsymbol{u}\cdot\boldsymbol{h}_{\boldsymbol{\nu}}^{(k)} = \sum_{\substack{\theta,\boldsymbol{\nu}(\theta)=\boldsymbol{\nu}\\ \mathrm{order}(\theta)=k}}^{*} \mathrm{Val}(\theta) \qquad [84]$$

where the $*$ means that the sum is only over trees in which a nonzero current $\boldsymbol{\nu}(l)$ flows on the lines $l\in\theta$. The quantity $\boldsymbol{u}\cdot\boldsymbol{h}_{\mathbf{0}}^{(k)}$ will be defined to be $\mathbf{0}$ (see the previous section).

In the case of [66] zero-current lines could appear: but the contributions from tree graphs containing at least one zero current line would cancel. In the present case, the statement that the above algorithm actually gives $\boldsymbol{h}_{\boldsymbol{\nu}}^{(k)}$ by simply ignoring trees with lines with zero current is nontrivial. It was Poincaré's contribution to the theory of Lindstedt series to show that even in the general case (cf. [75]) the equations for the invariant tori can be solved by a formal power series. Equation [84] is proved by induction on $k$ after checking it for the first few orders.

The algorithm just described leading to [83] can be extended to the case of the general Hamiltonian considered in the KAM theorem.

The convergence proof is more delicate than the (elementary) one for eqn [66]. In fact, the values of trees of order $k$ can give large contributions to $\boldsymbol{h}_{\boldsymbol{\nu}}^{(k)}$: because the "new" factors $(\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}(l))^2$, although not zero, can be quite small and their small size can overwhelm the smallness of the factors $f_{\boldsymbol{\nu}}$ and $\varepsilon$. In fact, even if $f$ is a trigonometric polynomial (so that $f_{\boldsymbol{\nu}}$ vanishes identically for $|\boldsymbol{\nu}|$ large enough) the currents flowing in the branches can be very large, of the order of the number $k$ of nodes in the tree; see [82].

This is called the *small-divisors problem*. The key to its solution goes back to a related work (Siegel) which shows that

**Theorem 4** *Consider the contribution to the sum in [82] from graphs $\theta$ in which no pairs of lines which lie on the same path to the root carry the same current and, furthermore, the node harmonics are bounded by $|\boldsymbol{\nu}|\le N$ for some $N$. Then the number of lines $\ell$ in $\theta$ with divisor $\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_\ell$ satisfying $2^{-n} < C|\boldsymbol{\omega}_0\cdot\boldsymbol{\nu}_\ell| \le 2^{-n+1}$ does not exceed $4Nk2^{-n/\tau}$.*

Hence, setting

$$F \stackrel{\text{def}}{=} C^2\max_{|\boldsymbol{\nu}|\le N}|f_{\boldsymbol{\nu}}|$$

the corresponding $\mathrm{Val}(\theta)$ can be bounded by

$$\frac{1}{k!}F^kN^{2k}\prod_{n=0}^{\infty}2^{2n(4Nk2^{-n/\tau})} \stackrel{\text{def}}{=} \frac{1}{k!}B^k \qquad B = FN^2 2\sum_n 8n2^{-n/\tau} \qquad [85]$$

since the product is convergent. In the case in which $f$ is a trigonometric polynomial of degree $N$, the above restricted contributions to $\boldsymbol{u}\cdot\boldsymbol{h}_{\boldsymbol{\nu}}^{(k)}$ would generate a convergent series for $\varepsilon$ small enough. In fact, the number of trees is bounded (as in the section "Perturbing functions") by $k!4^k(2N+1)^{\ell k}$ so that the series $\sum_{\boldsymbol{\nu}}|\varepsilon|^k|\boldsymbol{u}\cdot\boldsymbol{h}_{\boldsymbol{\nu}}^{(k)}|$ would converge for small $\varepsilon$ (i.e., $|\varepsilon| < (B\cdot 4(2N+1)^\ell)^{-1}$).

Given this comment, the analysis of the "remaining contributions" becomes the real problem, and it requires new ideas because among the excluded trees there are some simple $k$th order trees whose value alone, if considered separately from the other contributions, would generate a factorially divergent power series in $\varepsilon$.

However, the contributions of all large-valued trees of order $k$ can be shown to cancel: although not exactly (unlike the case of the elementary problem in the section "Perturbing functions," where the cancellation is not necessary for the proof, in spite of its exact occurrence), but enough so that in spite of the existence of exceedingly large values of individual tree graphs their total sum can still be bounded by a constant to the power $k$ so that the power series actually converges for $\varepsilon$ small enough. The idea is discussed in Appendix 1.

For more details, the reader is referred to Poincaré (1987), Kolmogorov (1954), Moser (1962), and Arnol'd (1989).

## Resonances and their Stability

A quasiperiodic motion with $r$ rationally independent frequencies is called resonant if $r$ is strictly less than the number of degrees of freedom, $\ell$. The difference $s=\ell-r$ is the degree of the resonance.

Of particular interest are the cases of a perturbation of an integrable system in which resonant motions take place.

A typical example is the $n$-body problem which studies the mutual perturbations of the motions of $n-1$ particles gravitating around a more massive particle. If the particle masses can be considered to be negligible, the system will consist of $n-1$ central Keplerian motions: it will therefore have $\ell = 3(n-1)$ degrees of freedom. In general, only one frequency per body occurs in the absence of the perturbations (the period of the Keplerian orbit). Hence, $r \leq n-1$ and $s \geq 2(n-1)$ (or in the planar case $s \geq (n-1)$) with equality holding when the periods are rationally independent.

Another example is the rigid body with a fixed point perturbed by a conservative force: in this case, the unperturbed system has three degrees of freedom but, in general, only two frequencies (see the discussion following [52]).

Furthermore, in the above examples there is the possibility that the independent frequencies assume, for special initial data, values which are rationally related, giving rise to resonances of even higher order (i.e., with smaller values of $r$).

In an integrable anisochronous system, resonant motions will be dense in phase space because the frequencies $\boldsymbol{\omega}(\boldsymbol{A})$ will vary as much as the actions and therefore resonances of any order (i.e., any $r < \ell$) will be dense in phase space: in particular, the periodic motions (i.e., the highest-order resonances) will be dense.

Resonances, in integrable systems, can arise in *a priori* stable integrable systems and in *a priori* unstable systems: the former are systems whose Hamiltonian admits canonical action–angle coordinates $(\boldsymbol{A}, \boldsymbol{\alpha}) \in U \times \mathbb{T}^\ell$ with $U \subset \mathbb{R}^\ell$ open, while the latter are systems whose Hamiltonian has, in suitable local canonical coordinates, the form

$$\mathcal{H}_0(\boldsymbol{A}) + \sum_{i=1}^{s_1} \frac{1}{2}(p_i^2 - \lambda_i^2 q_i^2) + \sum_{j=1}^{s_2} \frac{1}{2}(\pi_j^2 + \mu_j^2 \kappa_j^2), \quad \lambda_i, \mu_j > 0 \qquad [86]$$

where $(\boldsymbol{A}, \boldsymbol{\alpha}) \in U \times \mathbb{T}^r$, $U \in \mathbb{R}^r$, $(\boldsymbol{p}, \boldsymbol{q}) \in V \subset \mathbb{R}^{2s_1}$, $(\boldsymbol{\pi}, \boldsymbol{\kappa}) \in V' \subset \mathbb{R}^{2s_2}$ with $V, V'$ neighborhoods of the origin and $\ell = r + s_1 + s_2$, $s_i \geq 0$, $s_1 + s_2 > 0$ and $\pm\sqrt{\lambda_j}$, $\pm\sqrt{\mu_j}$ are called Lyapunov coefficients of the resonance. The perturbations considered are supposed to have the form $\varepsilon f(\boldsymbol{A}, \boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{q}, \boldsymbol{\pi}, \boldsymbol{\kappa})$. The denomination of *a priori* stable or unstable refers to the properties of the "*a priori* given unperturbed Hamiltonian." The label "*a priori* unstable" is certainly appropriate if $s_1 > 0$: here also $s_1 = 0$ is allowed for notational convenience implying that the Lyapunov coefficients in *a priori* unstable cases are all of order 1 (whether real $\lambda_j$ or imaginary $\mathrm{i}\sqrt{\mu_j}$). In other words, the *a priori* stable case, $s_1 = s_2 = 0$ in [86], is the only excluded case. Of course, the stability properties of the motions when a perturbation acts will depend on the perturbation in both cases.

The *a priori* stable systems usually have a great variety of resonances (e.g., in the anisochronous case, resonances of any dimension are dense). The *a priori* unstable systems have (among possible other resonances) some very special $r$-dimensional resonances occurring when the unstable coordinates $(\boldsymbol{p}, \boldsymbol{q})$ and $(\boldsymbol{\pi}, \boldsymbol{\kappa})$ are zero and the frequencies of the $r$ action–angle coordinates are rationally independent.

In the first case (*a priori* stable), the general question is whether the resonant motions, which form invariant tori of dimension $r$ arranged into families that fill $\ell$-dimensional invariant tori, continue to exist, in presence of small enough perturbations $\varepsilon f(\boldsymbol{A}, \boldsymbol{\alpha})$, on slightly deformed invariant tori. Similar questions can be asked in the *a priori* unstable cases. To examine the matter more closely consider the formulation of the simplest problems.

*A priori* stable resonances: more precisely, suppose $\mathcal{H}_0 = \frac{1}{2}\boldsymbol{A}^2$ and let $\{\boldsymbol{A}_0\} \times \mathbb{T}^\ell$ be the unperturbed invariant torus $\mathcal{T}_{\boldsymbol{A}_0}$ with spectrum $\boldsymbol{\omega}_0 = \boldsymbol{\omega}(\boldsymbol{A}_0) = \partial_{\boldsymbol{A}}\mathcal{H}_0(\boldsymbol{A}_0)$ with only $r$ rationally independent components. For simplicity, suppose that $\boldsymbol{\omega}_0 = (\omega_1, \ldots, \omega_r, 0, \ldots, 0) \stackrel{\text{def}}{=} (\boldsymbol{\omega}, \boldsymbol{0})$ with $\boldsymbol{\omega} \in \mathbb{R}^r$. The more general case in which $\boldsymbol{\omega}$ has only $r$ rationally independent components can be reduced to the special case above by a canonical linear change of coordinates at the price of changing the $\mathcal{H}_0$ to a new one, still quadratic in the actions but containing mixed products $A_i B_j$: the proofs of the results that are discussed here would not be really affected by such more general form of $\mathcal{H}$.

It is convenient to distinguish between the "fast" angles $\alpha_1, \ldots, \alpha_r$ and the "resonant" angles $\alpha_{r+1}, \ldots, \alpha_\ell$ (also called "slow" or "secular") and call $\boldsymbol{\alpha} = (\boldsymbol{\alpha}', \boldsymbol{\beta})$ with $\boldsymbol{\alpha}' \in \mathbb{T}^r$ and $\boldsymbol{\beta} \in \mathbb{T}^s$. Likewise, we distinguish the fast actions $\boldsymbol{A}' = (A_1, \ldots, A_r)$ and the resonant ones $A_{r+1}, \ldots, A_\ell$ and set $\boldsymbol{A} = (\boldsymbol{A}', \boldsymbol{B})$ with $\boldsymbol{A}' \in \mathbb{R}^r$ and $\boldsymbol{B} \in \mathbb{R}^s$.

Therefore, the torus $\mathcal{T}_{\boldsymbol{A}_0}, \boldsymbol{A}_0 = (\boldsymbol{A}'_0, \boldsymbol{B}_0)$, is in turn a continuum of invariant tori $\mathcal{T}_{\boldsymbol{A}_0, \boldsymbol{\beta}}$ with trivial parametric equations: $\boldsymbol{\beta}$ fixed, $\boldsymbol{\alpha}' = \boldsymbol{\psi}, \boldsymbol{\psi} \in \mathbb{T}^r$, and $\boldsymbol{A}' = \boldsymbol{A}'_0, \boldsymbol{B} = \boldsymbol{B}_0$. On each of them the motion is: $\boldsymbol{A}', \boldsymbol{B}, \boldsymbol{\beta}$ constant and $\boldsymbol{\alpha}' \to \boldsymbol{\alpha}' + \boldsymbol{\omega} t$, with rationally independent $\boldsymbol{\omega} \in \mathbb{R}^r$.

Then the natural question is whether there exist functions $\boldsymbol{h}_\varepsilon, \boldsymbol{k}_\varepsilon, \boldsymbol{H}_\varepsilon, \boldsymbol{K}_\varepsilon$ smooth in $\varepsilon$ near $\varepsilon = 0$ and in $\boldsymbol{\psi} \in \mathbb{T}^r$, vanishing for $\varepsilon = 0$, and such that the torus $\mathcal{T}_{\boldsymbol{A}_0, \boldsymbol{\beta}_0, \varepsilon}$ with parametric equations

$$\begin{aligned} \boldsymbol{A}' &= \boldsymbol{A}'_0 + \boldsymbol{H}_\varepsilon(\boldsymbol{\psi}), & \boldsymbol{\alpha}' &= \boldsymbol{\psi} + \boldsymbol{h}_\varepsilon(\boldsymbol{\psi}), \\ \boldsymbol{B} &= \boldsymbol{B}_0 + \boldsymbol{K}_\varepsilon(\boldsymbol{\psi}), & \boldsymbol{\beta} &= \boldsymbol{\beta}_0 + \boldsymbol{k}_\varepsilon(\boldsymbol{\psi}) \end{aligned} \qquad \boldsymbol{\psi} \in \mathbb{T}^r \qquad [87]$$

is invariant for the motions with Hamiltonian

$$\mathcal{H}_\varepsilon(\boldsymbol{A},\boldsymbol{\alpha}) = \tfrac{1}{2}\boldsymbol{A}'^2 + \tfrac{1}{2}\boldsymbol{B}^2 + \varepsilon f(\boldsymbol{\alpha}', \boldsymbol{\beta})$$

and the motions on it are $\boldsymbol{\psi} \to \boldsymbol{\psi} + \boldsymbol{\omega} t$. The above property, when satisfied, is summarized by saying that the unperturbed resonant motions $\boldsymbol{A} = (\boldsymbol{A}'_0, \boldsymbol{B}_0)$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}'_0 + \boldsymbol{\omega}' t, \boldsymbol{\beta}_0)$ can be continued in presence of perturbation $\varepsilon f$, for small $\varepsilon$, to quasiperiodic motions with the same spectrum and on a slightly deformed torus $\mathcal{T}_{\boldsymbol{A}'_0,\boldsymbol{\beta}_0,\varepsilon}$.

*A priori* unstable resonances: here the question is whether the special invariant tori continue to exist in presence of small enough perturbations, of course slightly deformed. This means asking whether, given $\boldsymbol{A}_0$ such that $\boldsymbol{\omega}(\boldsymbol{A}_0) = \partial_{\boldsymbol{A}}\mathcal{H}_0(\boldsymbol{A}_0)$ has rationally independent components, there are functions $(\boldsymbol{H}_\varepsilon(\boldsymbol{\psi}), \boldsymbol{h}_\varepsilon(\boldsymbol{\psi})), (\boldsymbol{P}_\varepsilon(\boldsymbol{\psi}), \boldsymbol{Q}_\varepsilon(\boldsymbol{\psi}))$ and $(\boldsymbol{\Pi}_\varepsilon(\boldsymbol{\psi}), \boldsymbol{K}_\varepsilon(\boldsymbol{\psi}))$ smooth in $\varepsilon$ near $\varepsilon = 0$, vanishing for $\varepsilon = 0$, analytic in $\boldsymbol{\psi} \in \mathbb{T}^r$ and such that the $r$-dimensional surface

$$\begin{aligned} \boldsymbol{A} &= \boldsymbol{A}_0 + \boldsymbol{H}_\varepsilon(\boldsymbol{\psi}), & \boldsymbol{\alpha} &= \boldsymbol{\psi} + \boldsymbol{h}_\varepsilon(\boldsymbol{\psi}) \\ \boldsymbol{p} &= \boldsymbol{P}_\varepsilon(\boldsymbol{\psi}), & \boldsymbol{q} &= \boldsymbol{Q}_\varepsilon(\boldsymbol{\psi}) \qquad \boldsymbol{\psi} \in \mathbb{T}^r \\ \boldsymbol{\pi} &= \boldsymbol{\Pi}_\varepsilon(\boldsymbol{\psi}), & \boldsymbol{\kappa} &= \boldsymbol{K}_\varepsilon(\boldsymbol{\psi}) \end{aligned} \qquad [88]$$

is an invariant torus $\mathcal{T}_{\boldsymbol{A}_0,\varepsilon}$ on which the motion is $\boldsymbol{\psi} \to \boldsymbol{\psi} + \boldsymbol{\omega}(\boldsymbol{A}_0)t$. Again, the above property is summarized by saying that the unperturbed special resonant motions can be continued in presence of perturbation $\varepsilon f$ for small $\varepsilon$ to quasiperiodic motions with the same spectrum and on a slightly deformed torus $\mathcal{T}_{\boldsymbol{A}_0,\varepsilon}$.

Some answers to the above questions are presented in the following section. For more details, the reader is referred to Gallavotti *et al.* (2004).

## Resonances and Lindstedt Series

We discuss eqns [87] in the paradigmatic case in which the Hamiltonian $\mathcal{H}_0(\boldsymbol{A})$ is $\frac{1}{2}\boldsymbol{A}^2$ (cf. [80]). It will be $\boldsymbol{\omega}(\boldsymbol{A}') \equiv \boldsymbol{A}'$ so that $\boldsymbol{A}_0 = \boldsymbol{\omega}$, $\boldsymbol{B}_0 = \boldsymbol{0}$ and the perturbation $f(\boldsymbol{\alpha})$ can be considered as a function of $\boldsymbol{\alpha} = (\boldsymbol{\alpha}', \boldsymbol{\beta})$: let $\overline{f}(\boldsymbol{\beta})$ be defined as its average over $\boldsymbol{\alpha}'$. The determination of the invariant torus of dimension $r$ which can be continued in the sense discussed in the last section is easily understood in this case.

A resonant invariant torus which, among the tori $\mathcal{T}_{\boldsymbol{A}_0,\boldsymbol{\beta}}$, has parametric equations that can be continued as a formal power series in $\varepsilon$ is the torus $\mathcal{T}_{\boldsymbol{A}_0,\boldsymbol{\beta}_0}$ with $\boldsymbol{\beta}_0$ a stationarity point for $\overline{f}(\boldsymbol{\beta})$, that is, an equilibrium point for the average perturbation: $\partial_{\boldsymbol{\beta}}\overline{f}(\boldsymbol{\beta}_0) = 0$. In fact, the following theorem holds:

**Theorem 5** *If $\boldsymbol{\omega} \in \mathbb{R}^r$ satisfies a Diophantine property and if $\boldsymbol{\beta}_0$ is a nondegenerate stationarity point for the "fast angle average" $\overline{f}(\boldsymbol{\beta})$* (i.e., *such that* $\det \partial^2_{\boldsymbol{\beta}\boldsymbol{\beta}}\overline{f}(\boldsymbol{\beta}_0) \neq 0$), *then the following equations for the functions $\boldsymbol{h}_\varepsilon, \boldsymbol{k}_\varepsilon$,*

$$\begin{aligned} (\boldsymbol{\omega} \cdot \partial_{\boldsymbol{\psi}})^2 \boldsymbol{h}_\varepsilon(\boldsymbol{\psi}) &= -\varepsilon \partial_{\boldsymbol{\alpha}'} f(\boldsymbol{\psi} + \boldsymbol{h}_\varepsilon(\boldsymbol{\psi}), \boldsymbol{\beta}_0 + \boldsymbol{k}_\varepsilon(\boldsymbol{\psi})) \\ (\boldsymbol{\omega} \cdot \partial_{\boldsymbol{\psi}})^2 \boldsymbol{k}_\varepsilon(\boldsymbol{\psi}) &= -\varepsilon \partial_{\boldsymbol{\beta}} f(\boldsymbol{\psi} + \boldsymbol{h}_\varepsilon(\boldsymbol{\psi}) + \boldsymbol{k}_\varepsilon(\boldsymbol{\psi})) \end{aligned} \qquad [89]$$

*can be formally solved in powers of $\varepsilon$.*

Given the simplicity of the Hamiltonian [80] that we are considering, it is not necessary to discuss the functions $\boldsymbol{H}_\varepsilon, \boldsymbol{K}_\varepsilon$ because the equations that they should obey reduce to their definitions as in the section "Quasiperiodicity and KAM stability," and for the same reason.

In other words, also the resonant tori admit a Lindstedt series representation. It is however very unlikely that the series are, in general, convergent.

Physically, this new aspect is due to the fact that the linearization of the motion near the torus $\mathcal{T}_{\boldsymbol{A}_0,\boldsymbol{\beta}_0}$ introduces oscillatory motions around $\mathcal{T}_{\boldsymbol{A}'_0,\boldsymbol{\beta}_0}$ with frequencies proportional to the square roots of the positive eigenvalues of the matrix $\varepsilon\partial^2_{\boldsymbol{\beta}\boldsymbol{\beta}}\overline{f}(\boldsymbol{\beta}_0)$: therefore, it is naively expected that it has to be necessary that a Diophantine property be required on the vector $(\boldsymbol{\omega}, \sqrt{\varepsilon\mu_1}, \ldots)$, where $\varepsilon\mu_j$ are the positive eigenvalues. Hence, some values of $\varepsilon$, namely those for which $(\boldsymbol{\omega}, \sqrt{\varepsilon\mu_1}, \ldots)$ is not a Diophantine vector or is too close to a non-Diophantine vector, should be excluded or at least should be expected to generate difficulties. Note that the problem arises irrespective of the assumptions about the nondegenerate matrix $\partial^2_{\boldsymbol{\beta}\boldsymbol{\beta}}\overline{f}(\boldsymbol{\beta}_0)$ (since $\varepsilon$ can have either sign), and no matter how small $|\varepsilon|$ is supposed to be. But we can expect that if the matrix $\partial^2_{\boldsymbol{\beta}\boldsymbol{\beta}}\overline{f}(\boldsymbol{\beta}_0)$ is (say) positive definite (i.e., $\boldsymbol{\beta}_0$ is a minimum point for $\overline{f}(\boldsymbol{\beta})$) then the problem should be easier for $\varepsilon < 0$ and vice versa, if $\boldsymbol{\beta}_0$ is a maximum, it should be easier for $\varepsilon > 0$ (i.e., in the cases in which the eigenvalues of $\varepsilon\partial^2_{\boldsymbol{\beta}\boldsymbol{\beta}}\overline{f}(\boldsymbol{\beta}_0)$ are negative and their roots do not have the interpretation of frequencies).

Technically, the sums of the formal series can be given (so far) a meaning only via summation rules involving divergent series: typically, one has to identify in the formal expressions (denumerably many) geometric series which, although divergent, can be given a meaning by applying the rule [79]. Since the rule can only be applied if $z \neq 1$, this leads to conditions on the parameter $\varepsilon$, in order to exclude that the various $z$ that have to be considered are very close to 1. Hence, this stability result turns out to be rather different from the KAM result for the maximal tori. Namely the series can be given a

meaning via summation rules provided $f$ and $\boldsymbol{\beta}_0$ satisfy certain additional conditions and provided certain values of $\varepsilon$ are excluded. An example of a theorem is the following:

**Theorem 6** *Given the Hamiltonian [80] and a resonant torus $\mathcal{T}_{A_0',\boldsymbol{\beta}_0}$ with $\boldsymbol{\omega}=A_0' \in \mathbb{R}^r$ satisfying a Diophantine property let $\boldsymbol{\beta}_0$ be a nondegenerate maximum point for the average potential $\overline{f}(\boldsymbol{\beta}) \stackrel{\text{def}}{=} (2\pi)^{-r}\int_{\mathbb{T}^r} f(\boldsymbol{\alpha}',\boldsymbol{\beta})d^r\boldsymbol{\alpha}'$. Consider the Lindstedt series solution for eqns [89] of the perturbed resonant torus with spectrum $(\boldsymbol{\omega},0)$. It is possible to express the single nth-order term of the series as a sum of many terms and then rearrange the series thus obtained so that the resummed series converges for $\varepsilon$ in a domain $\mathcal{E}$ which contains a segment $[0,\varepsilon_0]$ and also a subset of $[-\varepsilon_0,0]$ which, although with open dense complement, is so large that it has 0 as a Lebesgue density point. Furthermore, the resummed series for $\boldsymbol{h}_\varepsilon,\boldsymbol{k}_\varepsilon$ define an invariant r-dimensional analytic torus with spectrum $\boldsymbol{\omega}$.*

More generally, if $\boldsymbol{\beta}_0$ is only a nondegenerate stationarity point for $\overline{f}(\boldsymbol{\beta})$, the domain of definition of the resummed series is a set $\mathcal{E} \subset [-\varepsilon_0,\varepsilon_0]$ which on both sides of the origin has an open dense complement although it has 0 as a Lebesgue density point.

Theorem 6 can be naturally extended to the general case in which the Hamiltonian is the most general perturbation of an anisochronous integrable system $\mathcal{H}_\varepsilon(\boldsymbol{A},\boldsymbol{\alpha})=h(\boldsymbol{A})+\varepsilon f(\boldsymbol{A},\boldsymbol{\alpha})$ if $\partial^2_{\boldsymbol{AA}}h$ is a nonsingular matrix and the resonance arises from a spectrum $\boldsymbol{\omega}(A_0)$ which has $r$ independent components (while the remaining are not necessarily zero).

We see that the convergence is a delicate problem for the Lindstedt series for nearly integrable resonant motions. They might even be divergent (mathematically, a proof of divergence is an open problem but it is a very reasonable conjecture in view of the above physical interpretation); nevertheless, Theorem 6 shows that sum rules can be given that sometimes (i.e., for $\varepsilon$ in a large set near $\varepsilon=0$) yield a true solution to the problem.

This is reminiscent of the phenomenon met in discussing perturbations of isochronous systems in [76], but it is a much more complex situation. It leaves many open problems: foremost among them is the question of uniqueness. The sum rules of divergent series always contain some arbitrary choices, which lead to doubts about the uniqueness of the functions parametrizing the invariant tori constructed in this way. It might even be that the convergence set $\mathcal{E}$ may depend upon the arbitrary choices, and that considering several of them no $\varepsilon$ with $|\varepsilon|<\varepsilon_0$ is left out.

The case of *a priori* unstable systems has also been widely studied. In this case too resonances with Diophantine $r$-dimensional spectrum $\boldsymbol{\omega}$ are considered. However, in the case $s_2=0$ (called *a priori* unstable hyperbolic resonance) the Lindstedt series can be shown to be convergent, while in the case $s_1=0$ (called *a priori* unstable elliptic resonance) or in the mixed cases $s_1,s_2>0$ extra conditions are needed. They involve $\boldsymbol{\omega}$ and $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_{s_2})$ (cf. [86]) and properties of the perturbations as well. It is also possible to study a slightly different problem: namely to look for conditions on $\boldsymbol{\omega},\boldsymbol{\mu},f$ which imply that, for small $\varepsilon$, invariant tori with spectrum $\varepsilon$-dependent but close, in a suitable sense, to $\boldsymbol{\omega}$ exist.

The literature is vast, but it seems fair to say that, given the above comments, particularly those concerning uniqueness and analyticity, the situation is still quite unsatisfactory. We refer the reader to Gallavotti *et al.* (2004) for more details.

## Diffusion in Phase Space

The KAM theorem implies that a perturbation of an analytic anisochronous integrable system, i.e., with an analytic Hamiltonian $\mathcal{H}_\varepsilon(\boldsymbol{A},\boldsymbol{\alpha})=\mathcal{H}_0(\boldsymbol{A})+\varepsilon f(\boldsymbol{A},\boldsymbol{\alpha})$ and nondegenerate Hessian matrix $\partial^2_{\boldsymbol{AA}}h(\boldsymbol{A})$, generates large families of maximal invariant tori. Such tori lie on the energy surfaces but do not have codimension 1 on them, i.e., they do not split the $(2\ell-1)$–dimensional energy surfaces into disconnected regions except, of course, in the case of systems with two degrees of freedom (see the section "Quasiperiodicity and KAM stability").

Therefore, there might exist trajectories with initial data close to $\boldsymbol{A}^{\mathrm{i}}$ in action space which reach phase space points close to $\boldsymbol{A}^{\mathrm{f}} \neq \boldsymbol{A}^{\mathrm{i}}$ in action space for $\varepsilon \neq 0$, *no matter how small*. Such *diffusion* phenomenon would occur in spite of the fact that the corresponding trajectory has to move in a space in which very close to each $\{\boldsymbol{A}\}\times\mathbb{T}^\ell$ there is an invariant surface on which points move keeping $\boldsymbol{A}$ constant within $O(\varepsilon)$, which for $\varepsilon$ small can be $\ll|\boldsymbol{A}^{\mathrm{f}}-\boldsymbol{A}^{\mathrm{i}}|$.

In *a priori* unstable systems (cf. the section "Resonances and their stability") with $s_1=1$, $s_2=0$, it is not difficult to see that the corresponding phenomenon can actually occur: the paradigmatic example (Arnol'd) is the *a priori* unstable system

$$\mathcal{H}_\varepsilon = \frac{A_1^2}{2}+A_2+\frac{p^2}{2}+g(\cos q-1) + \varepsilon(\cos\alpha_1+\sin\alpha_2)(\cos q-1) \qquad [90]$$

This is a system describing a motion of a "pendulum" ($(p, q)$ coordinates) interacting with a "rotating wheel" ($(A_1, \alpha_1)$ coordinates) and a "clock" ($(A_2, \alpha_2)$ coordinates) *a priori* unstable near the points $p = 0$, $q = 0$, $2\pi$ ($s_1 = 1$, $s_2 = 0$, $\lambda_1 = \sqrt{g}$, cf. [86]). It can be proved that on the energy surface of energy $E$ and for each $\varepsilon \neq 0$ small enough (no matter how small) there are initial data with action coordinates close to $\boldsymbol{A}^{\mathrm{i}} = (A_1^{\mathrm{i}}, A_2^{\mathrm{i}})$ with $(1/2)A_1^{\mathrm{i}2} + A_2^{\mathrm{i}}$ close to $E$ eventually evolving to a datum $\boldsymbol{A}' = (A_1', A_2')$ with $A_1'$ at a distance from $A_1^{\mathrm{f}}$ smaller than an arbitrarily prefixed distance (of course with energy $E$). Furthermore, during the whole process the pendulum energy stays close to zero within $o(\varepsilon)$ (i.e., the pendulum swings following closely the unperturbed separatrices).

In other words, [90] describes a machine (the pendulum) which, working approximately in a cycle, extracts energy from a reservoir (the clock) to transfer it to a mechanical device (the wheel). The statement that diffusion is possible means that the machine can work as soon as $\varepsilon \neq 0$, if the initial actions and the initial phases (i.e., $\alpha_1, \alpha_2, p, q$) are suitably tuned (as functions of $\varepsilon$).

The peculiarity of the system [90] is that the fixed points $P_\pm$ of the unperturbed pendulum (i.e., the equilibria $p = 0, q = 0, 2\pi$) *remain* unstable equilibria even when $\varepsilon \neq 0$ and this is an important simplifying feature.

It is a peculiarity that permits bypassing the obstacle, arising in the analysis of more general cases, represented by the resonance surfaces consisting of the $\boldsymbol{A}$'s with $A_1\nu_1 + \nu_2 = 0$: the latter correspond to harmonics $(\nu_1, \nu_2)$ present in the perturbing function, i.e., the harmonics which would lead to division by zero in an attempt to construct (as necessary in studying [90] by Arnol'd's method) the parametric equations of the perturbed invariant tori with action close to such $\boldsymbol{A}$'s. In the case of [90] the problem arises only on the resonance marked in **Figure 6** by a heavy line, i.e., $A_1 = 0$, corresponding to $\cos\alpha_1$ in [90].

If $\varepsilon = 0$, the points $P_-$ with $p = 0$, $q = 0$ and the point $P_+$ with $p = 0, q = 2\pi$ are both unstable equilibria (and they are, of course, the same point, if $q$ is an angular variable). The unstable manifold (it is a curve) of $P_+$ coincides with the stable manifold of $P_-$ and vice versa. So that the unperturbed system admits nontrivial motions leading from $P_+$ to $P_-$ and from $P_-$ to $P_+$, both in a bi-infinite time interval $(-\infty, \infty)$: the $p, q$ variables describe a pendulum and $P_\pm$ are its unstable equilibria which are connected by the separatrices (which constitute the zero-energy surfaces for the pendulum).

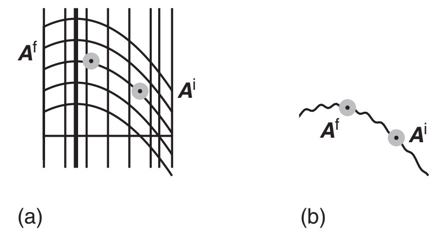



**Figure 6** (a) The $\varepsilon = 0$ geometry: the "partial energy" lines are parabolas, $(1/2)A_1^2 + A_2 = \text{const}$. The vertical lines are the resonances $A_1 = \text{rational}$ (i.e., $\nu_1 A_1 + \nu_2 = 0$). The disks are neighborhoods of the points $\boldsymbol{A}^{\mathrm{i}}$ and $\boldsymbol{A}^{\mathrm{f}}$ (the dots at their centers). (b) $\varepsilon \neq 0$; an artist's rendering of a trajectory in $\boldsymbol{A}$ space, driven by the pendulum swings to accelerate the wheel from $A_1^i$ to $A_1^f$ at the expenses of the clock energy, sneaking through invariant tori not represented and (approximately) located "away" from the intersections between resonances and partial energy lines (a dense set, however). The pendulum coordinates are not shown: its energy stays close to zero, within a power of $\varepsilon$. Hence the pendulum swings, staying close to the separatrix. The oscillations symbolize the wiggly behavior of the partial energy $(1/2)A_1^2 + A_2$ in the process of sneaking between invariant tori which, because of their invariance, would be impossible without the pendulum. The energy $(1/2)A_1^2$ of the wheel increases slightly at each pendulum swing: accurate estimates yield an increase of the wheel speed $A_1$ of the order of $\varepsilon/(\log \varepsilon^{-1})$ at each swing of the pendulum implying a transition time of the order of $g^{-1/2}\varepsilon^{-1} \log \varepsilon^{-1}$.

The latter property remains true for more general *a priori* unstable Hamiltonians

$$\mathcal{H}_\varepsilon = \mathcal{H}_0(\boldsymbol{A}) + \mathcal{H}_{\mathrm{u}}(p, q) + \varepsilon f(\boldsymbol{A}, \boldsymbol{\alpha}, p, q) \quad \text{in } (U \times \mathbb{T}^\ell) \times (\mathbb{R}^2) \qquad [91]$$

where $\mathcal{H}_{\mathrm{u}}$ is a one-dimensional Hamiltonian which has two unstable equilibrium points $P_+$ and $P_-$ linearly repulsive in one direction and linearly attractive in another which are connected by two heteroclinic trajectories which, as time tends to $\pm\infty$, approach $P_-$ and $P_+$ and vice versa.

Actually, the points need not be different but, if coinciding, the trajectories linking them must be nontrivial: in the case [90] the variable $q$ can be considered an angle and then $P_+$ and $P_-$ would coincide (but are connected by nontrivial trajectories, i.e., by trajectories that also visit points different from $P_\pm$). Such trajectories are called heteroclinic if $P_+ \neq P_-$ and homoclinic if $P_+ = P_-$.

In the general case, besides the homoclinicity (or heteroclinicity) condition, certain weak genericity conditions, automatically satisfied in the example [90], have to be imposed in order to show that, given $\boldsymbol{A}^{\mathrm{i}}$ and $\boldsymbol{A}^{\mathrm{f}}$ with the same unperturbed energy $E$, one can find, for all $\varepsilon$ small enough but not equal to zero, initial data ($\varepsilon$-dependent) with actions arbitrarily close to $\boldsymbol{A}^{\mathrm{i}}$ which evolve to data with actions arbitrarily close to $\boldsymbol{A}^{\mathrm{f}}$. This is a phenomenon

called the *Arnol'd diffusion*. Simple sufficient conditions for a transition from near $\boldsymbol{A}^{\mathrm{i}}$ to near $\boldsymbol{A}^{\mathrm{f}}$ are expressed by the following result:

**Theorem 7** *Given the Hamiltonian [91] with $\mathcal{H}_{\mathrm{u}}$ admitting two hyperbolic fixed points $P_{\pm}$ with heteroclinic connections, $t \to (p_a(t), q_a(t))$, $a = 1, 2$, suppose that:*

(i) *On the unperturbed energy surface of energy $E = \mathcal{H}(\boldsymbol{A}^{\mathrm{i}}) + \mathcal{H}_{\mathrm{u}}(P_{\pm})$ there is a regular curve $\gamma : s \to \boldsymbol{A}(s)$ joining $\boldsymbol{A}^{\mathrm{i}}$ to $\boldsymbol{A}^{\mathrm{f}}$ such that the unperturbed tori $\{\boldsymbol{A}(s)\} \times \mathbb{T}^{\ell}$ can be continued at $\varepsilon \neq 0$ into invariant tori $\mathcal{T}_{\boldsymbol{A}(s),\varepsilon}$ for a set of values of $s$ which fills the curve $\gamma$ leaving only gaps of size of order $o(\varepsilon)$.*

(ii) *The $\ell \times \ell$ matrix $D_{ij}$ of the second derivatives of the integral of $f$ over the heteroclinic motions is not degenerate, that is,*

$$|\det D| = \left| \det \left( \int_{-\infty}^{\infty} \mathrm{d}t \; \partial_{\alpha_i \alpha_j} f(\boldsymbol{A}, \boldsymbol{\alpha} + \boldsymbol{\omega}(\boldsymbol{A})t, p_a(t), q_a(t)) \right) \right| > c > 0 \qquad [92]$$

*for all $\boldsymbol{A}$'s on the curve $\gamma$ and all $\boldsymbol{\alpha} \in \mathbb{T}^2$.*

*Given arbitrary $\rho > 0$, for $\varepsilon \neq 0$ small enough there are initial data with action and energy closer than $\rho$ to $\boldsymbol{A}^{\mathrm{i}}$ and $E$, respectively, which after a long enough time acquire an action closer than $\rho$ to $\boldsymbol{A}^{\mathrm{f}}$ (keeping the initial energy).*

The above two conditions can be shown to hold generically for many pairs $\boldsymbol{A}^{\mathrm{i}} \neq \boldsymbol{A}^{\mathrm{f}}$ (and many choices of the curves $\gamma$ connecting them) if the number of degrees of freedom is $\geq 3$. Thus, the result, obtained by a simple extension of the argument originally outlined by Arnol'd to discuss the paradigmatic example [90], proves the existence of diffusion in *a priori* unstable systems. The integral in [92] is called *Melnikov integral*.

The real difficulty is to estimate the time needed for the transition: it is a time that obviously has to diverge as $\varepsilon \to 0$. Assuming $g$ fixed (i.e., $\varepsilon$ independent) a naive approach easily leads to estimates which can even be worse than $O(\exp(a\varepsilon^{-b}))$ with some $a, b > 0$. It has finally been shown that in such cases the minimum time can be, for rather general perturbations $\varepsilon f(\boldsymbol{\alpha}, q)$, estimated above by $O(\varepsilon^{-1} \log \varepsilon^{-1})$, which is the best that can be hoped for under generic assumptions.

The reader is referred to Arnol'd (1989) and Chierchia and Valdinoci (2000) for more details.

## Long-Time Stability of Quasiperiodic Motions

A more difficult problem is whether the same phenomenon of migration in action space occurs in *a priori* stable systems. The root of the difficulty is a remarkable stability property of quasiperiodic motions. Consider Hamiltonians $\mathcal{H}_{\varepsilon}(\boldsymbol{A}, \boldsymbol{\alpha}) = h(\boldsymbol{A}) + \varepsilon f(\boldsymbol{A}, \boldsymbol{\alpha})$ with $\mathcal{H}_0(\boldsymbol{A}) = h(\boldsymbol{A})$ strictly convex, analytic, and anisochronous on the closure $\overline{U}$ of an open bounded region $U \subset \mathbb{R}^{\ell}$, and a perturbation $\varepsilon f(\boldsymbol{A}, \boldsymbol{\alpha})$ analytic in $\overline{U} \times \mathbb{T}^{\ell}$.

Then *a priori* bounds are available on how long it can possibly take to migrate from an action close to $\boldsymbol{A}_1$ to one close to $\boldsymbol{A}_2$: and the bound is of "exponential type" as $\varepsilon \to 0$ (i.e., it admits a lower bound which behaves as the exponential of an inverse power of $\varepsilon$). The simplest theorem is (Nekhorossev):

**Theorem 7** *There are constants $0 < a, b, d, g, \tau$ such that any initial datum $(\boldsymbol{A}, \boldsymbol{\alpha})$ evolves so that $\boldsymbol{A}$ will not change by more than $a\varepsilon^g$ before a long time bounded below by $\tau \exp(b\varepsilon^{-d})$.*

Thus, this puts an exponential bound, i.e., a bound exponential in an inverse power of $\varepsilon$, to the diffusion time: before a time $\tau \exp(b\varepsilon^{-d})$ actions can only change by $O(\varepsilon^g)$ so that their variation cannot be large no matter how small $\varepsilon \neq 0$ is chosen. This places a (long) lower bound to the time of diffusion in *a priori* stable systems.

The proof of the theorem provides, actually, an interesting and detailed picture of the variations in actions showing that some actions may vary more slowly than others.

The theorem is constructive, i.e., all constants $0 < a, b, d, \tau$ can be explicitly chosen and depend on $\ell, \mathcal{H}_0, f$ although some of them can be fixed to depend only on $\ell$ and on the minimum curvature of the convex graph of $\mathcal{H}_0$. Its proof can be adapted to cover many cases which do not fall in the class of systems with strictly convex unperturbed Hamiltonian, and even to cases with a resonant unperturbed Hamiltonian.

However, in important problems (e.g., in the three-body problems met in celestial mechanics) there is empirical evidence that diffusion takes place at a fast pace (i.e., not exponentially slow in the above sense) while the above results would forbid a rapid migration in phase space if they applied: however, in such problems the assumptions of the theorem are not satisfied, because the unperturbed system is strongly resonant (as in the celestial mechanics problems, where the number of independent frequencies is a fraction of the number

of degrees of freedom and $h(\boldsymbol{A})$ is far from strictly convex), leaving wide open the possibility of observing rapid diffusion.

Further, changing the assumptions can dramatically change the results. For instance, rapid diffusion can sometimes be proved even though it might be feared that it should require exponentially long times: an example that has been proposed is the case of a three-timescales system, with Hamiltonian

$$\omega_1 A_1 + \omega_2 A_2 + \frac{p^2}{2} + g(1+\cos q) + \varepsilon f(\alpha_1, \alpha_2, p, q) \qquad [93]$$

with $\boldsymbol{\omega}_\varepsilon \stackrel{\text{def}}{=} (\omega_1, \omega_2)$, where $\omega_1 = \varepsilon^{-1/2}\overline{\omega}$, $\omega_2 = \varepsilon^{1/2}\widetilde{\omega}$ and $\overline{\omega}, \widetilde{\omega} > 0$ constants. The three scales are $\omega_1^{-1}$, $\sqrt{g^{-1}}$, $\omega_2^{-1}$. In this case, there are many (although by no means all) pairs $\boldsymbol{A}_1, \boldsymbol{A}_2$ which can be connected within a time that can be estimated to be of order $O(\varepsilon^{-1}\log\varepsilon^{-1})$.

This is a rapid-diffusion case in an *a priori* unstable system in which condition [92] is not satisfied: because the $\varepsilon$-dependence of $\boldsymbol{\omega}(\boldsymbol{A})$ implies that the lower bound $c$ in [92] must depend on $\varepsilon$ (and be exponentially small with an inverse power of $\varepsilon$ as $\varepsilon \to 0$).

The unperturbed system in [93] is nonresonant in the $\mathcal{H}_0$ part for $\varepsilon > 0$ outside a set of zero measure (i.e., where the vector $\boldsymbol{\omega}_\varepsilon$ satisfies a suitable Diophantine property) and, furthermore, it is *a priori* unstable: cases met in applications can be *a priori* stable and resonant (and often not anisochronous) in the $\mathcal{H}_0$ part. In such a system, not only the speed of diffusion is not understood but proposals to prove its existence, if present (as expected), have so far not given really satisfactory results.

For more details, the reader in referred to Nekhorossev (1977).

## The Three-Body Problem

Mechanics and the three-body problem can be almost identified with each other, in the sense that the motion of three gravitating masses has long been a key astronomical problem and at the same time the source of inspiration for many techniques: foremost among them the theory of perturbations.

As an introduction, consider a special case. Let three masses $m_S = m_0$, $m_J = m_1$, $m_M = m_2$ interact via gravity, that is, with interaction potential $-km_i m_j |\boldsymbol{x}_i - \boldsymbol{x}_j|^{-1}$: the simplest problem arises when the third body has a neglegible mass compared to the two others and the latter are supposed to be on a circular orbit; furthermore, the mass $m_J$ is $\varepsilon m_S$ with $\varepsilon$ small and the mass $m_M$ moves in the plane of the circular orbit. This will be called the "circular restricted three-body problem."

In a reference system with center $S$ and rotating at the angular speed of $J$ around $S$ inertial forces (centrifugal and Coriolis) act. Supposing that the body $J$ is located on the axis with unit vector $\boldsymbol{i}$ at distance $R$ from the origin $S$, the acceleration of the point $M$ is

$$\ddot{\varrho} = \boldsymbol{F} + \omega_0^2\left(\varrho - \frac{\varepsilon R}{1+\varepsilon}\boldsymbol{i}\right) - 2\boldsymbol{\omega}_0 \wedge \dot{\varrho}$$

if $\boldsymbol{F}$ is the force of attraction and $\boldsymbol{\omega}_0 \wedge \dot{\varrho} \equiv \omega_0 \dot{\varrho}^\perp$ where $\boldsymbol{\omega}_0$ is a vector with $|\boldsymbol{\omega}_0| = \omega_0$ and perpendicular to the orbital plane and $\varrho^\perp \stackrel{\text{def}}{=} (-\rho_2, \rho_1)$ if $\varrho = (\rho_1, \rho_2)$. Here, taking into account that the origin $S$ rotates around the fixed center of mass, $\omega_0^2(\varrho - \varepsilon R/(1+\varepsilon)\boldsymbol{i})$ is the centrifugal force while $-2\boldsymbol{\omega}_0 \wedge \dot{\varrho}$ is the Coriolis force. The equations of motion can therefore be derived from a Lagrangian

$$\mathcal{L} = \frac{1}{2}\dot{\varrho}^2 - W + \omega_0 \varrho^\perp \cdot \dot{\varrho} + \frac{1}{2}\omega_0^2 \varrho^2 - \omega_0^2 \frac{\varepsilon R}{1+\varepsilon}\varrho \cdot \boldsymbol{i} \qquad [94]$$

with

$$\omega_0^2 R^3 = km_S(1+\varepsilon) \stackrel{\text{def}}{=} g_0$$
$$W = -\frac{km_S}{|\varrho|} - \frac{km_S\varepsilon}{|\varrho - R\boldsymbol{i}|}$$

where $k$ is the gravitational constant, $R$ the distance between $S$ and $J$, and finally the last three terms in [94] come from the Coriolis force (the first) and from the centripetal force (the other two, taking into account that the origin $S$ rotates around the fixed center of mass).

Setting $g = g_0/(1+\varepsilon) \equiv km_S$, the Hamiltonian of the system is

$$\mathcal{H} = \frac{1}{2}(\boldsymbol{p} - \omega_0 \varrho^\perp)^2 - \frac{g}{|\varrho|} - \frac{1}{2}\omega_0^2\varrho^2 - \varepsilon\frac{g}{R}\left(\left|\frac{\varrho}{R} - \boldsymbol{i}\right|^{-1} - \frac{\varrho}{R}\cdot\boldsymbol{i}\right) \qquad [95]$$

The first part can be expressed immediately in the action–angle coordinates for the two-body problem (cf. the section "Newtonian potential and Kepler's laws"). Calling such coordinates $(L_0, \lambda_0, G_0, \gamma_0)$ and $\theta_0$ the polar angle of $M$ with respect to the major axis of the ellipse and $\lambda_0$ the mean anomaly of $M$ on its ellipse, the Hamiltonian becomes, taking into account that for $\varepsilon = 0$ the ellipse axis rotates at speed $-\omega_0$,

$$\mathcal{H} = -\frac{g^2}{2L_0^2} - \omega_0 G_0 - \varepsilon\frac{g}{R}\left(\left|\frac{\varrho}{R} - \boldsymbol{i}\right|^{-1} - \frac{\varrho}{R}\cdot\boldsymbol{i}\right) \qquad [96]$$

which is convenient if we study the interior problem, i.e., $|\varrho| < R$. This can be expressed in the action–angle coordinates via [41], [42]:

$$\theta_0 = \lambda_0 + f_{\lambda_0}, \qquad \theta_0 + \gamma_0 = \lambda_0 + \gamma_0 + f_{\lambda_0}$$
$$e = \left(1 - \frac{G_0^2}{L_0^2}\right)^{1/2}, \quad \frac{|\varrho|}{R} = \frac{G_0^2}{gR} \frac{1}{1 + e\cos(\lambda_0 + f_{\lambda_0})} \qquad [97]$$

where (see [42]), $f_\lambda = f(e\sin\lambda, e\cos\lambda)$ and

$$f(x, y) = 2x\left(1 + \frac{5}{4}y + \cdots\right)$$

with the ellipsis denoting higher orders in $x, y$ even in $x$. The Hamiltonian takes the form, if $\omega^2 = gR^{-3}$,

$$\mathcal{H}_\varepsilon = -\frac{g^2}{2L_0^2} - \omega G_0 + \varepsilon \frac{g}{R} F(G_0, L_0, \lambda_0, \lambda_0 + \gamma_0) \qquad [98]$$

where the only important feature (for our purposes) is that $F(L, G, \alpha, \beta)$ is an analytic function of $L, G, \alpha, \beta$ near a datum with $|G| < L$ (i.e., $e > 0$) and $|\varrho| < R$. However, the domain of analyticity in $G$ is rather small as it is constrained by $|G| < L$ excluding in particular the circular orbit case $G = \pm L$.

Note that apparently the KAM theorem fails to be applicable to [98] because the matrix of the second derivatives of $\mathcal{H}_0(L, G)$ has vanishing determinant. Nevertheless, the proof of the theorem also goes through in this case, with minor changes. This can be checked by studying the proof or, following a remark by Poincaré, by simply noting that the "squared" Hamiltonian $\mathcal{H}'_\varepsilon \stackrel{\text{def}}{=} (\mathcal{H}_\varepsilon)^2$ has the form

$$\mathcal{H}'_\varepsilon = \left(-\frac{g^2}{2L_0^2} - \omega G_0\right)^2 + \varepsilon F'(G_0, L_0, \lambda_0, \lambda_0 + \gamma_0) \qquad [99]$$

with $F'$ still analytic. *But* this time

$$\det \frac{\partial^2 \mathcal{H}'_0}{\partial(G_0, L_0)} = -6g^2 L_0^{-4} \omega_0^2 h \neq 0$$
$$\text{if } h = -g^2 L_0^{-2} - 2\omega G_0 \neq 0$$

Therefore, the KAM theorem applies to $\mathcal{H}'_\varepsilon$ and the key observation is that the orbits generated by the Hamiltonian $(\mathcal{H}_\varepsilon)^2$ are geometrically the same as those generated by the Hamiltonian $\mathcal{H}_\varepsilon$: they are only run at a different speed because of the need of a time rescaling by the constant factor $2\mathcal{H}_\varepsilon$.

This shows that, given an unperturbed ellipse of parameters $(L_0, G_0)$ such that $\boldsymbol{\omega} = (g^2/L_0^3, -\omega)$, $G_0 > 0$, with $\omega_1/\omega_2$ Diophantine, then the perturbed system admits a motion which is quasiperiodic with spectrum proportional to $\boldsymbol{\omega}$ and takes place on an orbit which wraps around a torus remaining forever close to the unperturbed torus (which can be visualized as described by a point moving, according to the area law on an ellipse rotating at a rate $-\omega_0$) with actions $(L_0, G_0)$, provided $\varepsilon$ is small enough. Hence,

> The KAM theorem answers, at least conceptually, the classical question: can a solution of the three-body problem remain close to an unperturbed one forever? That is, is it possible that a solar system is stable forever?

Assuming $e, |\varrho|/R \ll 1$ and retaining only the lowest orders in $e$ and $|\varrho|/R \ll 1$ the Hamiltonian [98] simplifies into

$$\mathcal{H} = -\frac{g^2}{2L_0^2} - \omega G_0 + \delta_\varepsilon(G_0) - \frac{\varepsilon g}{2R} \frac{G_0^4}{g^2 R^2}\Big(3\cos 2(\lambda_0 + \gamma_0)$$
$$- e\cos\lambda_0 - \frac{9}{2} e\cos(\lambda_0 + 2\gamma_0)$$
$$+ \frac{3}{2} e\cos(3\lambda_0 + 2\gamma_0)\Big) \qquad [100]$$

where

$$\delta_\varepsilon(G_0) = -((1+\varepsilon)^{1/2} - 1)\omega G_0 - \frac{\varepsilon g}{2R} \frac{G_0^4}{g^2 R^2}$$
$$e = \left(1 - \frac{G_0^2}{L_0^2}\right)^{1/2}$$

It is an interesting exercise to estimate, assuming as model the Hamiltonian [100] and following the proof of the KAM theorem, how small has $\varepsilon$ to be if a planet with the data of Mercury can be stable forever on a (slowly precessing) orbit with actions close to the present-day values under the influence of a mass $\varepsilon$ times the solar mass orbiting on a circle, at a distance from the Sun equal to that of Jupiter. It is possible to follow either the above reduction to the ordinary KAM theorem or to apply directly to [100] the Lindstedt series expansion, proceeding along the lines of the section "Quasiperiodicity and KAM stability." The first approach is easy but the second is more efficient: in both cases, unless the estimates are done in a particularly careful manner, the value found for $\varepsilon m_S$ is not interesting from the viewpoint of astronomy.

The reader is refered to Arnol'd (1989) for more details.

## Rationalization and Regularization of Singularities

Often integrable systems have interesting data which lie on the boundary of the integrability domain. For instance, the central motion when $L = G$ (circular orbits) or the rigid body in a rotation around one of the principal axes or the two-body problem when $G = 0$ (collisional data). In such cases, perturbation

theory cannot be applied as discussed above. Typically, the perturbation depends on quantities like $\sqrt{L-G}$ and is *not analytic* at $L=G$. Nevertheless, it is sometimes possible to enlarge phase space and introduce new coordinates in the vicinity of the data which in the initial phase space are singular.

A notable example is the failure of the analysis of the circular restricted three-body problem: it apparently fails when the orbit that we want to perturb is circular.

It is convenient to introduce the canonical coordinates $L, \lambda$ and $G, \gamma$:

$$\begin{aligned} L &= L_0, \qquad G = L_0 - G_0 \\ \lambda &= \lambda_0 + \gamma_0, \qquad \gamma = -\gamma_0 \end{aligned} \qquad [101]$$

so that $e = \sqrt{2GL^{-1}}\sqrt{1-G(2L)^{-1}}$ and $\lambda_0 = \lambda + \gamma$ and $\theta_0 = \lambda_0 + f_{\lambda_0}$, where $f_{\lambda_0}$ is defined in [42] (see also [97]). Hence,

$$\begin{aligned} &\theta_0 = \lambda + \gamma + f_{\lambda+\gamma}, \qquad \theta_0 + \gamma_0 = \lambda + f_{\lambda+\gamma} \\ &e = \sqrt{2G}\sqrt{\frac{1}{L}\left(1 - \frac{G}{2L}\right)} \\ &\frac{|\varrho|}{R} = \frac{L^2(1-e^2)}{gR}\frac{1}{1+e\cos(\lambda+\gamma+f_{\lambda+\gamma})} \end{aligned} \qquad [102]$$

and the Hamiltonian [100] takes the form

$$\begin{aligned} \mathcal{H}_\varepsilon = &-\frac{g^2}{2L^2} - \omega L + \omega G \\ &+ \varepsilon\frac{g}{R}F(L-G, L, \lambda+\gamma, \lambda) \end{aligned} \qquad [103]$$

In the coordinates $L$,$G$ of [101] the unperturbed circular case corresponds to $G=0$ and [96], once expressed in the action–angle variables $G, L, \gamma, \lambda$, is analytic in a domain whose size is controlled by $\sqrt{G}$. Nevertheless, very often problems of perturbation theory can be "regularized."

This is done by "enlarging the integrability" domain by adding to it points (one or more) around the singularity (a boundary point of the domain of the coordinates) and introducing new coordinates to describe simultaneously the data close to the singularity and the newly added points: in many interesting cases, the equations of motion are no longer singular (i.e., become analytic) in the new coordinates and are therefore apt to describe the motions that reach the singularity in a finite time. One can say that the singularity was only apparent.

Perhaps this is best illustrated precisely in the above circular restricted three-body problem, with the singularity occurring where $G=0$, that is, at a circular unperturbed orbit. If we describe the points with $G$ small in a new system of coordinates obtained from the one in [101] by letting alone $L, \lambda$ and setting

$$p = \sqrt{2G}\cos\gamma, \qquad q = \sqrt{2G}\sin\gamma \qquad [104]$$

then $p$, $q$ vary in a neighborhood of the origin with the origin itself excluded.

Adding the origin of the $p$–$q$ plane then in a full neighborhood of the origin, the Hamiltonian [96] is analytic in $L, \lambda, p, q$. This is because it is analytic (cf. [96], [97]) as a function of $L, \lambda$ and $e\cos\theta_0$ and of $\cos(\lambda_0+\theta_0)$. Since $\theta_0 = \lambda+\gamma+f_{\lambda+\gamma}$ and $\theta_0+\lambda_0 = \lambda+f_{\lambda+\gamma}$ by [97], the Hamiltonian [96] is analytic in $L, \lambda, e\cos(\lambda+\gamma+f_{\lambda+\gamma})$, $\cos(\lambda+f_{\lambda+\gamma})$ for $e$ small (i.e., for $G$ small) and, by [42], $f_{\lambda+\gamma}$ is analytic in $e\sin(\lambda+\gamma)$ and $e\cos(\lambda+\gamma)$. Hence the trigonometric identities

$$\begin{aligned} e\sin(\lambda+\gamma) &= \frac{p\sin\lambda + q\cos\lambda}{\sqrt{L}}\sqrt{1-\frac{G}{2L}} \\ e\cos(\lambda+\gamma) &= \frac{p\cos\lambda - q\sin\lambda}{\sqrt{L}}\sqrt{1-\frac{G}{2L}} \end{aligned} \qquad [105]$$

together with $G=(1/2)(p^2+q^2)$ imply that [103] is analytic near $p=q=0$ and $L>0, \lambda\in[0,2\pi]$. The Hamiltonian becomes analytic and the new coordinates are suitable to describe motions crossing the origin: for example, by setting

$$C \stackrel{\text{def}}{=} \frac{1}{2}\left(1 - \frac{p^2+q^2}{4L}\right)L^{-1/2}$$

[100] becomes

$$\begin{aligned} \mathcal{H} = &-\frac{g^2}{2L^2} - \omega L + \omega\tfrac{1}{2}(p^2+q^2) \\ &+ \delta_\varepsilon(\tfrac{1}{2}(p^2+q^2)) - \frac{\varepsilon g}{2R}\frac{(L-\frac{1}{2}(p^2+q^2))^4}{g^2R^2} \\ &\times (3\cos 2\lambda - ((-11\cos\lambda + 3\cos 3\lambda)p \\ &- (7\sin\lambda + 3\sin 3\lambda)q)C) \end{aligned} \qquad [106]$$

The KAM theorem does not apply in the form discussed above to "Cartesian coordinates," that is, when, as in [106], the unperturbed system is not assigned in action–angle variables. However, there are versions of the theorem (actually its corollaries) which do apply and therefore it becomes possible to obtain some results even for the perturbations of circular motions by the techniques that have been illustrated here.

Likewise, the Hamiltonian of the rigid body with a fixed point $O$ and subject to analytic external forces becomes singular, if expressed in the action–angle coordinates of Deprit, when the body motion nears a rotation around a principal axis or, more generally, nears a configuration in which any two of

the axes $\boldsymbol{i}_3, \boldsymbol{z}$, or $\boldsymbol{z}_0$ coincide (i.e., any two among the principal axis, the angular momentum axis and the inertial $z$-axis coincide; see the section "Rigid body"). Nevertheless, by imitating the procedure just described in the simpler cases of the circular three-body problem, it is possible to enlarge the phase space so that in the new coordinates the Hamiltonian is analytic near the singular configurations.

A regularization also arises when considering collisional orbits in the unrestricted planar three-body problem. In this respect, a very remarkable result is the regularization of collisional orbits in the planar three-body problem. After proving that if the total angular momentum does not vanish, simultaneous collisions of the three masses cannot occur within any finite time interval, the question is reduced to the regularization of two-body collisions, under the assumption that the total angular momentum does not vanish.

The local change of coordinates, which changes the relative position coordinates $(x, y)$ of two colliding bodies as $(x, y) \to (\xi, \eta)$, with $x + iy = (\xi + i\eta)^2$, is not one to one, hence it has to be regarded as an enlargement of the positions space, if points with different $(\xi, \eta)$ are considered different. However, the equations of motion written in the variables $\xi, \eta$ have no singularity at $\xi, \eta = 0$ (Levi-Civita).

Another celebrated regularization is the regularization of the Schwartzschild metric, i.e., of the general relativity version of the two-body problem: it is, however, somewhat out of the scope of this review (Synge, Kruskal).

For more details, the reader is refered to Levi-Civita (1956).

## Appendix 1: KAM Resummation Scheme

The idea to control the "remaining contributions" is to reduce the problem to the case in which there are no pairs of lines that follow each other in the tree order and which have the same current. Mark by a scale label "0" the lines, see [74], [83], of a tree whose divisors $C/\boldsymbol{\omega}_0.\boldsymbol{\nu}(l)$ are >1: these are lines which give no problems in the estimates. Then mark by a scale label "≥1" the lines with current $\boldsymbol{\nu}(l)$ such that $|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(l)| \le 2^{-n+1}$ for $n = 1$ (i.e., the remaining lines).

The lines labeled 0 are said to be on scale 0, while those labeled ≥1 are said to be on scale ≥1. A cluster of scale 0 will be a maximal collection of lines of scale 0 forming a connected subgraph of a tree $\theta$.

Consider only trees $\theta_0 \in \Theta_0$ of the family $\Theta_0$ of trees containing no clusters of lines with scale label 0 which have only one line entering the cluster and one exiting it with equal current.

It is useful to introduce the notion of a line $\ell_1$ situated "between" two lines $\ell, \ell'$ with $\ell' > \ell$: this will mean that $\ell_1$ precedes $\ell'$ but not $\ell$.

All trees $\theta$ in which there are some pairs $l' > l$ of consecutive lines of scale label ≥1 which have equal current and such that all lines between them bear scale label 0 are obtained by "inserting" on the lines of trees in $\Theta_0$ with label ≥1 any number of clusters of lines and nodes, with lines of scale 0 and with the property that the sum of the harmonics of the nodes inserted vanishes.

Consider a line $l_0 \in \theta_0 \in \Theta_0$ linking nodes $v_1 < v_2$ and labeled ≥1 and imagine inserting on it a cluster $\gamma$ of lines of scale 0 with sum of the node harmonics vanishing and out of which emerges one line connecting a node $v_{\rm out}$ in $\gamma$ to $v_2$ and into which enters one line linking $v_1$ to a node $v_{\rm in} \in \gamma$. The insertion of a $k$–lines, $|\gamma| = (k+1)$-nodes, cluster changes the tree value by replacing the line factor, that will be briefly called "value of the cluster $\gamma$", as

$$\frac{\boldsymbol{\nu}_{v_1} \cdot \boldsymbol{\nu}_{v_2}}{\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(l_0)^2} \to \frac{(\boldsymbol{\nu}_{v_1} \cdot M(\gamma; \boldsymbol{\nu}(l_0))\boldsymbol{\nu}_{v_2})}{\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(l_0)^2} \frac{1}{\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(l_0)^2} \quad [107]$$

where $M$ is an $\ell \times \ell$ matrix

$$M_{rs}(\gamma, \boldsymbol{\nu}(l_0)) = \frac{\varepsilon^{|\gamma|}}{k!} \nu_{out,r} \nu_{in,s} \prod_{v \in \gamma} (-f_{\boldsymbol{\nu}_v}) \prod_{l \in \gamma} \frac{\boldsymbol{\nu}_v \cdot \boldsymbol{\nu}_{v'}}{\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(l)^2}$$

if $\ell = v'v$ denotes a line linking $v'$ and $v$. Therefore, if all possible connected clusters are inserted and the resulting values are added up, the result can be taken into account by attributing to the original line $l_0$ a factor like [107] with $M^{(0)}(\boldsymbol{\nu}(l_0)) \stackrel{\rm def}{=} \sum_\gamma M(\gamma; \boldsymbol{\nu}(l_0))$ replacing $M(\gamma; \boldsymbol{\nu}(l_0))$.

If several connected clusters $\gamma$ are inserted on the same line and their values are summed, the result is a modification of the factor associated with the line $l_0$ into

$$\sum_{k=0}^{\infty} \boldsymbol{\nu}_{v_1} \cdot \left( \frac{M^{(0)}(\boldsymbol{\nu}(l_0))}{\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(l_0)^2} \right)^k \boldsymbol{\nu}_{v_2} \frac{1}{\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(l_0)^2}$$
$$= \left( \boldsymbol{\nu}_{v_1} \cdot \frac{1}{\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(l_0)^2 - M^{(0)}(\boldsymbol{\nu}(l_0))} \boldsymbol{\nu}_{v_2} \right) \quad [108]$$

The series defining $M^{(0)}$ involves, by construction, only trees with lines of scale 0, hence with large divisors, so that it converges to a matrix of small size of order $\varepsilon$ (actually $\varepsilon^2$, more precisely) if $\varepsilon$ is small enough.

Convergence can be established by simply remarking that the series defining $M^{(1)}$ is built with lines with values >(1/2) of the propagator, so that it certainly converges for $\varepsilon$ small enough (by the estimates in the section "Perturbing functions," where the propagators were identically 1) and the

sum is of order $\varepsilon$ (actually $\varepsilon^2$), hence <1. However, such an argument cannot be repeated when dealing with lines with smaller propagators (which still have to be discussed). Therefore, a method not relying on so trivial a remark on the size of the propagators has eventually to be used when considering lines of scale higher than 1, as it will soon become necessary.

The advantage of the collection of terms achieved with [108] is that we can represent $\boldsymbol{h}$ as a sum of values of trees which are simpler because they contain no pair of lines of scale $\geq 1$ with in between lines of scale 0 with total sum of the node harmonics vanishing. The price is that the divisors are now more involved and we even have a problem due to the fact that we have not proved that the series in [108] converges. In fact, it is a geometric series whose value is the RHS of [108] obtained by the sum rule [79] unless we can prove that the ratio of the geometric series is <1. This is trivial in this case by the previous remark: but it is better to note that there is another reason for convergence, whose use is not really necessary here but will become essential later.

The property that the ratio of the geometric series is <1 can be regarded as due to the consequence of the cancellation mentioned in the section "Quasi-periodicity and KAM stability" which can be shown to imply that the ratio is <1 because $M^{(0)}(\boldsymbol{\nu}) = \varepsilon^2(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu})^2 m^{(0)}(\boldsymbol{\nu})$ with $C\,|m^{(0)}(\boldsymbol{\nu})| < D_0$ for some $D_0 > 0$ and for all $|\varepsilon| < \varepsilon_0$ for some $\varepsilon_0$. Then for small $\varepsilon$ the divisor in [108] is essentially still what it was before starting the resummation.

At this point, an induction can be started. Consider trees evaluated with the new rule and place a scale level "$\geq 2$" on the lines with $C\,|\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(l)| \leq 2^{-n+1}$ for $n = 2$: leave the label "0" on the lines already marked so and label by "1" the other lines. The lines of scale "1" will satisfy $2^{-n} < |\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(l)| \leq 2^{-n+1}$ for $n = 1$. The graphs will now possibly contain lines of scale 0, 1 or $\geq 2$ while lines with label "$\geq 1$" no longer can appear, by construction.

A cluster of scale 1 will be a maximal collection of lines of scales 0, 1 forming a connected subgraph of a tree $\theta$ and containing at least one line of scale 1.

The construction carried out by considering clusters of scale 0 can be repeated by considering trees $\theta_1 \in \Theta_1$, with $\Theta_1$ the collection of trees with lines marked 0, 1, or $\geq 2$ and in which no pairs of lines with equal momentum appear to follow each other if between them there are only lines marked 0 or 1.

Insertion of connected clusters $\gamma$ of such lines on a line $l_0$ of $\theta_1$ leads to define a matrix $M^{(1)}$ formed by summing tree values of clusters $\gamma$ with lines of scales 0 or 1 evaluated with the line factors defined in [107] and with the restriction that in $\gamma$ there are no pairs of lines $\ell < \ell'$ with the same current and which follow each other while any line between them has lower scale (i.e., 0), here "between" means "preceding $l'$ but not preceding $l$," as above.

Therefore, a scale-independent method has to be devised to check the convergence for $M^{(1)}$ and for the matrices to be introduced later to deal with even smaller propagators. This is achieved by the following extension of Siegel's theorem mentioned in the section "Quasiperiodicity and KAM stability":

**Theorem 8** *Let $\boldsymbol{\omega}_0$ satisfy [74] and set $\boldsymbol{\omega} = C\boldsymbol{\omega}_0$. Consider the contribution to the sum in [82] from graphs $\theta$ in which*

(i) *no pairs $\ell' > \ell$ of lines which lie on the same path to the root carry the same current $\boldsymbol{\nu}$ if all lines $\ell_1$ between them have current $\boldsymbol{\nu}(\ell_1)$ such that $|\boldsymbol{\omega} \cdot \boldsymbol{\nu}(\ell_1)| > 2|\boldsymbol{\omega} \cdot \boldsymbol{\nu}|$;*
(ii) *the node harmonics are bounded by $|\boldsymbol{\nu}| \leq N$ for some $N$.*

*Then the number of lines $\ell$ in $\theta$ with divisor $\boldsymbol{\omega} \cdot \boldsymbol{\nu}_\ell$ satisfying $2^{-n} < |\boldsymbol{\omega} \cdot \boldsymbol{\nu}_\ell| \leq 2^{-n+1}$ does not exceed $4\,Nk2^{-n/\tau}$, $n = 1, 2, \ldots$.*

This implies, by the same estimates in [85], that the series defining $M^{(1)}$ converges. Again, it must be checked that there are cancellations implying that $M^{(1)}(\boldsymbol{\nu}) = \varepsilon^2(\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu})^2 m^{(1)}(\boldsymbol{\nu})$ with $|m^{(1)}(\boldsymbol{\nu})| < D_0$ for the same $D_0 > 0$ and the same $\varepsilon_0$.

At this point, one deals with trees containing only lines carrying labels $0, 1, \geq 2$, and the line factors for the lines $\ell = v'v$ of scale 0 are $\boldsymbol{\nu}_{v'} \cdot \boldsymbol{\nu}_v / (\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(\ell))^2$, those of the lines $\ell = v'v$ of scale 1 have line factors $\boldsymbol{\nu}_{v'} \cdot (\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(\ell)^2 - M^{(0)}(\boldsymbol{\nu}(\ell)))^{-1}\boldsymbol{\nu}_v$, and those of the lines $\ell = v'v$ of scale $\geq 2$ have line factors

$$\boldsymbol{\nu}_{v'} \cdot (\boldsymbol{\omega}_0 \cdot \boldsymbol{\nu}(\ell)^2 - M^{(1)}(\boldsymbol{\nu}(\ell)))^{-1}\boldsymbol{\nu}_v$$

*Furthermore*, no pair of lines of scale "1" or of scale "$\geq 2$" with the same momentum and with only lines of lower scale (i.e., of scale "0" in the first case or of scale "0", "1" in the second) between them can follow each other.

This procedure can be iterated until, after infinitely many steps, the problem is reduced to the evaluation of tree values in which each line carries a scale label $n$ and there are no pairs of lines which follow each other and which have only lines of lower scale in between. Then the Siegel argument applies once more and the series so resumed is an absolutely convergent series of functions analytic in $\varepsilon$: hence the original series is convergent.

Although at each step there is a lower bound on the denominators, it would not be possible to avoid using Siegel's theorem. In fact, the lower bound would become worse and worse as the scale increases. In order to check

the estimates of the constants $D_0, \varepsilon_0$ which control the scale independence of the convergence of the various series, it is necessary to take advantage of the theorem, and of the absence (at each step) of the necessity of considering trees with pairs of consecutive lines with equal momentum and intermediate lines of higher scale.

One could also perform the analysis by bounding $h^{(k)}$ order by order with no resummations (i.e., without changing the line factors) and exhibiting the necessary cancellations. Alternatively, the paths that Kolmogorov, Arnol'd and Moser used to prove the first three (somewhat different) versions of the theorem, by successive approximations of the equations for the tori, can be followed.

The invariant tori are Lagrangian manifolds just as the unperturbed ones (cf. comments after [31]) and, in the case of the Hamiltonian [80], the generating function $\boldsymbol{A} \cdot \boldsymbol{\psi} + \Phi(\boldsymbol{A}, \boldsymbol{\psi})$ can be expressed in terms of their parametric equations

$$\begin{aligned}
&\Phi(\boldsymbol{A}, \boldsymbol{\psi}) = G(\boldsymbol{\psi}) + \boldsymbol{a} \cdot \boldsymbol{\psi} + \boldsymbol{h}(\boldsymbol{\psi}) \cdot (\boldsymbol{A} - \boldsymbol{\omega} - \Delta \boldsymbol{h}(\boldsymbol{\psi})) \\
&\partial_{\boldsymbol{\psi}} G(\boldsymbol{\psi}) \stackrel{\text{def}}{=} -\Delta \boldsymbol{h}(\boldsymbol{\psi}) + \underset{\sim}{h}(\boldsymbol{\psi}) \partial_{\boldsymbol{\psi}} \Delta \underset{\sim}{h}(\boldsymbol{\psi}) - \boldsymbol{a} \\
&\boldsymbol{a} \stackrel{\text{def}}{=} \int (-\Delta \boldsymbol{h}(\boldsymbol{\psi}) + \underset{\sim}{h}(\boldsymbol{\psi}) \partial_{\boldsymbol{\psi}} \Delta \underset{\sim}{h}(\boldsymbol{\psi})) \frac{\mathrm{d}\boldsymbol{\psi}}{(2\pi)^{\ell}} \\
&\quad = \int \underset{\sim}{h}(\boldsymbol{\psi}) \partial_{\boldsymbol{\psi}} \Delta \underset{\sim}{h}(\boldsymbol{\psi}) \frac{\mathrm{d}\boldsymbol{\psi}}{(2\pi)^{\ell}}
\end{aligned} \qquad [109]$$

where $\Delta = (\boldsymbol{\omega} \cdot \boldsymbol{\partial}_{\boldsymbol{\psi}})$ and the invariant torus corresponds to $\boldsymbol{A}' = \boldsymbol{\omega}$ in the map $\boldsymbol{\alpha} = \boldsymbol{\psi} + \boldsymbol{\partial}_{\boldsymbol{A}} \boldsymbol{\Phi}(\boldsymbol{A}, \boldsymbol{\psi})$ and $\boldsymbol{A}' = \boldsymbol{A} + \boldsymbol{\partial}_{\boldsymbol{\psi}} \Phi(\boldsymbol{A}, \boldsymbol{\psi})$. In fact, by [109] the latter becomes $\boldsymbol{A}' = \boldsymbol{A} - \Delta \boldsymbol{h}$ and, from the second of [75] written for $f$ depending only on the angles $\boldsymbol{\alpha}$, it is $\boldsymbol{A} = \boldsymbol{\omega} + \Delta \boldsymbol{h}$ when $\boldsymbol{A}, \boldsymbol{\alpha}$ are on the invariant torus.

Note that if $\boldsymbol{a}$ exists it is necessarily determined by the third relation in [109] but the check that the second equation in [109] is soluble (i.e., that the RHS is an exact gradient up to a constant) is nontrivial. The canonical map generated by $\boldsymbol{A} \cdot \boldsymbol{\psi} + \boldsymbol{\Phi}(\boldsymbol{A}, \boldsymbol{\psi})$ is also defined for $\boldsymbol{A}'$ close to $\boldsymbol{\omega}$ and foliates the neighborhood of the invariant torus with other tori: of course, for $\boldsymbol{A}' \neq \boldsymbol{\omega}$ the tori defined in this way are, in general, not invariant.

The reader is referred to Gallavotti *et al.* (2004) for more details.

## Appendix 2: Coriolis and Lorentz Forces – Larmor Precession

Larmor precession refers to the motion of an electrically charged particle in a magnetic field $H$ (in an inertial frame of reference). It is due to the Lorentz force which, on a unit mass with unit charge, produces an acceleration $\ddot{\varrho} = \boldsymbol{v} \wedge \boldsymbol{H}$ if the speed of light is $c = 1$.

Therefore, if $\boldsymbol{H} = H\boldsymbol{k}$ is directed along the $\boldsymbol{k}$-axis, the acceleration it produces is the same that the Coriolis force would impress on a unit mass located in a reference frame which rotates with angular velocity $\omega_0 \boldsymbol{k}$ around the $\boldsymbol{k}$-axis if $\boldsymbol{H} = 2\omega_0 \boldsymbol{k}$.

The above remarks imply that a homogeneous sphere electrically charged uniformly with a unit charge and freely pivoting about its center in a constant magnetic field $H$ directed along the $\boldsymbol{k}$-axis undergoes the same motion as it would follow if not subject to the magnetic field but seen in a noninertial reference frame rotating at constant angular velocity $\omega_0$ around the $\boldsymbol{k}$-axis if $H$ and $\omega_0$ are related by $H = 2\omega_0$: in this frame, the Coriolis force is interpreted as a magnetic field.

This holds, however, only if the centrifugal force has zero moment with respect to the center: true in the spherical symmetry case only. In spherically nonsymmetric cases, the centrifugal forces have in general nonzero moment, so the equivalence between Coriolis force and the Lorentz force is only approximate.

The Larmor theorem makes this more precise. It gives a quantitative estimate of the difference between the motion of a general system of particles of mass $m$ in a magnetic field and the motion of the same particles in a rotating frame of reference but in the absence of a magnetic field. The approximation is estimated in terms of the size of the Larmor frequency $eH/2mc$, which should be small compared to the other characteristic frequencies of the motion of the system: the physical meaning is that the centrifugal force should be small compared to the other forces.

The vector potential $\boldsymbol{A}$ for a constant magnetic field in the $\boldsymbol{k}$-direction, $\boldsymbol{H} = 2\omega_0 \boldsymbol{k}$, is $\boldsymbol{A} = 2\omega_0 \boldsymbol{k} \wedge \varrho \equiv 2\omega_0 \varrho^{\perp}$. Therefore, from the treatment of the Coriolis force in the section "Three-body problem" (see [95]), the motion of a charge $e$ with mass $m$ in a magnetic field $\boldsymbol{H}$ with vector potential $\boldsymbol{A}$ and subject to other forces with potential $W$ can be described, in an inertial frame and in generic units, in which the speed of light is $c$, by a Hamiltonian

$$\mathcal{H} = \frac{1}{2m}\left(\boldsymbol{p} - \frac{e}{c}\boldsymbol{A}\right)^2 + W(\varrho) \qquad [110]$$

where $\boldsymbol{p} = m\dot{\varrho} + (e/c)\boldsymbol{A}$ and $\varrho$ are canonically conjugate variables.

## Further Reading


Arnol'd VI (1989) *Mathematical Methods of Classical Mechanics*. Berlin: Springer.

Calogero F and Degasperis A (1982) *Spectral Transform and Solitons*. Amsterdam: North-Holland.

Chierchia L and Valdinoci E (2000) A note on the construction of Hamiltonian trajectories along heteroclinic chains. *Forum Mathematicum* 12: 247–255.
Fassò F (1998) Quasi-periodicity of motions and complete integrability of Hamiltonian systems. *Ergodic Theory and Dynamical Systems* 18: 1349–1362.
Gallavotti G (1983) *The Elements of Mechanics*. New York: Springer.
Gallavotti G, Bonetto F, and Gentile G (2004) *Aspects of the Ergodic, Qualitative and Statistical Properties of Motion*. Berlin: Springer.
Kolmogorov N (1954) On the preservation of conditionally periodic motions. *Doklady Akademia Nauk SSSR* 96: 527–530.
Landau LD and Lifshitz EM (1976) *Mechanics*. New York: Pergamon Press.
Levi-Civita T (1956) Opere Matematiche. *Accademia Nazionale dei Lincei*. Bologna: Zanichelli.
Moser J (1962) On invariant curves of an area preserving mapping of the annulus. *Nachricten Akademie Wissenschaften Göttingen* 11: 1–20.
Nekhorossev V (1977) An exponential estimate of the time of stability of nearly integrable Hamiltonian systems. *Russian Mathematical Surveys* 32(6): 1–65.
Poincaré H (1987) *Méthodes nouvelles de la mécanique celèste* vol. I. Paris: Gauthier-Villars. (reprinted by Gabay, Paris, 1987).

# Introductory Article: Differential Geometry

**S Paycha**, Université Blaise Pascal, Aubière, France



Differential geometry is the study of differential properties of geometric objects such as curves, surfaces and higher-dimensional manifolds endowed with additional structures such as metrics and connections. One of the main ideas of differential geometry is to apply the tools of analysis to investigate geometric problems; in particular, it studies their “infinitesimal parts,” thereby linearizing the problem. However, historically, geometric concepts often anticipated the analytic tools required to define them from a differential geometric point of view; the notion of tangent to a curve, for example, arose well before the notion of derivative.

In its barely more than two centuries of existence, differential geometry has always had strong (often two-way) interactions with physics. Just to name a few examples, the theory of curves is used in kinematics, symplectic manifolds arise in Hamiltonian mechanics, pseudo-Riemannian manifolds in general relativity, spinors in quantum mechanics, Lie groups and principal bundles in gauge theory, and infinite-dimensional manifolds in the path-integral approach to quantum field theory.

## Curves and Surfaces

The study of differential properties of curves and surfaces resulted from a combination of the coordinate method (or analytic geometry) developed by Descartes and Fermat during the first half of the seventeenth century and infinitesimal calculus developed by Leibniz and Newton during the second half of the seventeenth and beginning of the eighteenth century.

Differential geometry appeared later in the eighteenth century with the works of Euler *Recherches sur la courbure des surfaces* (1760) (Investigations on the curvature of surfaces) and Monge *Une application de l'analyse à la géométrie* (1795) (An application of analysis to geometry). Until Gauss' fundamental article *Disquisitiones generales circa superficies curvas* (General investigations of curved surfaces) published in Latin in 1827 (of which one can find a partial translation to English in Spivak (1979)), surfaces embedded in $\mathbb{R}^3$ were either described by an equation, $W(x, y, z) = 0$, or by expressing one variable in terms of the others. Although Euler had already noticed that the coordinates of a point on a surface could be expressed as functions of two independent variables, it was Gauss who first made a systematic use of such a parametric representation, thereby initiating the concept of “local chart” which underlies differential geometry.

## Differentiable Manifolds

The actual notion of $n$-manifold independent of a particular embedding in a Euclidean space goes back to a lecture *Über die Hypothesen, welche der Geometrie zu Grunde liegen* (On the hypotheses which lie at the foundations of geometry) (of which one can find a translation to English and comments in Spivak (1979)) delivered by Riemann at Göttingen University in 1854, in which he makes clear the fact that $n$-manifolds are locally like $n$-dimensional Euclidean space. In his work, Riemann mentions the existence of infinite-dimensional manifolds, such as function spaces, which today play an important role since they naturally arise as configuration spaces in quantum field theories.

In modern language a differentiable manifold modeled on a topological space $V$ (which can be

finite dimensional, Fréchet, Banach, or Hilbert for example) is a topological space $M$ equipped with a family of local coordinate charts $(U_i, \phi_i)_{i \in I}$ such that the open subsets $U_i \subset M$ cover $M$ and where $\phi_i : U_i \to V$, $i \in I$, are homeomorphisms which give rise to smooth transition maps $\phi_i \circ \phi_j^{-1} : \phi_j(U_i \cap U_j) \to \phi_i(U_i \cap U_j)$. An $n$-dimensional differentiable manifold is a differentiable manifold modeled on $\mathbb{R}^n$. The sphere $S^{n-1} := \{(x_1, \ldots, x_n) \in \mathbb{R}^n, \sum_{i=1}^n x_i^2 = 1\}$ is a differentiable manifold of dimension $n - 1$.

Simple differentiable curves in $\mathbb{R}^n$ are one-dimensional differentiable manifolds locally specified by coordinates $x(t) = (x_1(t), \ldots, x_n(t)) \in \mathbb{R}^n$, where $t \mapsto x_j(t)$ is of class $C^k$. The tangent at point $x(t_0)$ to such a curve, which is a straight line passing through this point with direction given by the vector $x'(t_0)$, generalizes to the concept of tangent space $T_mM$ at point $m \in M$ of a smooth manifold $M$ modeled on $V$ which is a vector space isomorphic to $V$ spanned by tangent vectors at point $m$ to curves $\gamma(t)$ of class $C^1$ on $M$ such that $\gamma(t_0) = m$.

In order to make this more precise, one needs the notion of differentiable mapping. Given two differentiable manifolds $M$ and $N$, a mapping $f : M \to N$ is differentiable at point $m$ if, for every chart $(U, \phi)$ of $M$ containing $m$ and every chart $(V, \psi)$ of $N$ such that $f(U) \subset V$, the mapping $\psi \circ f \circ \phi^{-1} : \phi(U) \to \psi(V)$ is differentiable at point $\phi(m)$. In particular, differentiable mappings $f : M \to \mathbb{R}$ form the algebra $C^\infty(M, \mathbb{R})$ of smooth real-valued functions on $M$. Differentiable mappings $\gamma : [a, b] \to M$ from an interval $[a, b] \subset \mathbb{R}$ to a differentiable manifold $M$ are called "differentiable curves" on $M$. A differentiable mapping $f : M \to N$ which is invertible and with differentiable inverse $f^{-1} : N \to M$ is called a diffeomorphism.

The derivative of a function $f \in C^\infty(M, \mathbb{R})$ along a curve $\gamma : [a, b] \to M$ at point $\gamma(t_0) \in M$ with $t_0 \in [a, b]$ is given by

$$Xf := \frac{\mathrm{d}}{\mathrm{d}t_{|_{t=t_0}}} f \circ \gamma(t)$$

and the map $f \mapsto Xf$ is called the tangent vector to the curve $\gamma$ at point $\gamma(t_0)$. Tangent vectors to some curve $\gamma : [a, b] \to M$ at a given point $m \in \gamma([a, b])$ form a vector space $T_mM$ called the "tangent space" to $M$ at point $m$.

A (smooth) map which, to a point $m \in M$, assigns a tangent vector $X \in T_mM$ is called a (smooth) vector field. It can also be seen as a derivation $\tilde{X} : f \mapsto Xf$ on $C^\infty(M, \mathbb{R})$ defined by $(\tilde{X}f)(m) := X(m)f$ for any $m \in M$ and the bracket of vector fields is thereby defined from the operator bracket $\widetilde{[X, Y]} := \tilde{X} \circ \tilde{Y} - \tilde{Y} \circ \tilde{X}$. The linear operations on tangent vectors carry out to vector fields $(X + Y)(m) := X(m) + Y(m), (\lambda X)(m) := \lambda X(m)$ for any $m \in M$ and for any $X, Y \in T_mM, \lambda \in \mathbb{R}$ so that vector fields on $M$ build a linear space.

One can generate tangent vectors to $M$ via local one-parameter groups of differentiable transformations of $M$, that is, mappings $(t, m) \mapsto \phi_t(m)$ from $]-\epsilon, \epsilon[ \times U$ to $U$ (with $\epsilon > 0$ and $U \subset M$ an open subset of $M$) such that $\phi_0 = \mathrm{Id}, \phi_{t+s} = \phi_t \circ \phi_s$ $\forall s, t \in ]-\epsilon, \epsilon[$ with $t + s \in ]-\epsilon, \epsilon[$ and $m \mapsto \phi_t(m)$ is a diffeomorphism of $U$ onto an open subset $\phi_t(U)$. The tangent vector at $t = 0$ to the curve $\gamma(t) = \phi_t(m)$ yields a tangent vector to $M$ at point $m = \gamma(0)$. Conversely, when $M$ is finite dimensional, the fundamental theorem for systems of ordinary equations yields, for any vector field $X$ on $M$, the existence (around any point $m \in M$) of a local one-parameter group of local transformations $\phi : ]-\epsilon, \epsilon[ \times U \to M$ (with $U$ an open subset containing $m$) which induces the tangent vector $X(m) \in T_mM$.

A differentiable mapping $\phi : M \to N$ induces a map $\phi_*(m) : T_mM \to T_{\phi(m)}M$ defined by $\phi_* Xf = X(f \circ \phi)$. An "immersion" of a manifold $M$ in a manifold $N$ is a differentiable mapping $\phi : M \to N$ such that the maps $\phi_*(m)$ are injective at any point $m \in M$. Such a map is an embedding if it is moreover injective in which case $\phi(M) \subset N$ is a submanifold of $N$. The unit sphere $S^n$ is a submanifold of $\mathbb{R}^{n+1}$. Whitney showed that every smooth real $n$-dimensional manifold can be embedded in $\mathbb{R}^{2n+1}$.

A differentiable manifold whose coordinate charts take values in a complex vector space $V$ and whose transition maps are holomorphic is called a complex manifold, which is complex $n$-dimensional if $V = \mathbb{C}^n$. The complex projective space $\mathbb{C}P^n$, the union of complex straight lines through 0 in $\mathbb{C}^{n+1}$, is a compact complex manifold of dimension $n$. Similarly to the notion of differentiable mapping between differentiable manifolds, we have the notion of holomorphic mapping between complex manifolds.

A smooth family $m \mapsto J_m$ of endomorphisms of the tangent spaces $T_mM$ to a differentiable manifold $M$ such that $J_m^2 = -\mathrm{Id}$ gives rise to an almost-complex manifold. The prototype is the almost-complex structure on $\mathbb{C}^n$ defined by $J(\partial_{x_i}) = \partial_{y_i}; J(\partial_{y_i}) = -\partial_{x_i}$ with $z = (x_1 + \mathrm{i}y_1, \ldots, x_n + \mathrm{i}y_n) \in \mathbb{C}^n$ which can be transferred to a complex manifold $M$ by means of local charts. An almost-complex structure $J$ on a manifold $M$ is called complex if $M$ is the underlying differentiable manifold of a complex manifold which induces $J$ in this way.

Studying smooth functions on a differentiable manifold can provide information on the topology of the manifold: for example, the behavior of a smooth function on a compact manifold as its critical points strongly restricted by the topological properties of the manifold. This leads to the Morse

critical point theory which extends to infinite-dimensional manifolds and, among other consequences, leads to conclusions on extremals or closed extremals of variational problems. Rather than privileging points on a manifold, one can study instead the geometry of manifolds from the point of view of spaces of functions, which leads to an algebraic approach to differential geometry. The initial concept there is a commutative ring (which becomes a possibly noncommutative algebra in the framework of noncommutative geometry), namely the ring of smooth functions on the manifold, while the manifold itself is defined in terms of the ring as the space of maximal ideals. In particular, this point of view proves to be fruitful to understand supermanifolds, a generalization of manifolds which is important for supersymmetric field theories.

One can further consider the sheaf of smooth functions on an open subset of the manifold; this point of view leads to sheaf theory which provides a unified approach to establishing connections between local and global properties of topological spaces.

## Metric Properties

Riemann focused on the metric properties of manifolds but the first clear formulation of the concept of a manifold equipped with a metric was given by Weyl in *Die Idee der Riemannsche Fläche*. A Riemannian metric on a differentiable manifold $M$ is a positive-definite scalar product $g_m$ on $T_mM$ for every point $m \in M$ depending smoothly on the point $m$. A manifold equipped with a Riemannian metric is called a Riemannian manifold. A Weyl transformation, which is multiplying the metric by a smooth positive function, yields a new Riemannian metric with the same angle measurement as the original one, and hence leaves the "conformal" structure on $M$ unchanged.

Riemann also suggested considering metrics on the tangent spaces that are not induced from scalar products; metrics on the manifold built this way were first systematically investigated by Finsler and are therefore called Finsler metrics. Geodesics on a Riemannian manifold $M$ which correspond to smooth curves $\gamma : [a, b] \to M$ that minimize the length functional

$$L(\gamma) := \frac{1}{2}\int_a^b \sqrt{g_{\gamma(t)}\left(\frac{\mathrm{d}\gamma}{\mathrm{d}t}, \frac{\mathrm{d}\gamma}{\mathrm{d}t}\right)}\,\mathrm{d}t$$

then generalize to curves which realize the shortest distance between two points chosen sufficiently close.

Euclid's axioms which naturally lead to Riemannian geometry are also satisfied up to the axiom of parallelism by a geometry developed by Lobatchevsky in 1829 and Bolyai in 1832. Non-Euclidean geometries actually played a major role in the development of differential geometry and Lobachevsky's work inspired Riemann and later Klein.

Dropping the positivity assumption for the bilinear forms $g_m$ on $T_mM$ leads to Lorentzian manifolds which are $(n+1)$-dimensional smooth manifolds equipped with bilinear forms on the tangent spaces with signature $(1, n)$. These occur in general relativity and tangent vectors with negative, positive, or vanishing squared length are called timelike, spacelike, and lightlike, respectively.

Just as complex vector spaces can be equipped with positive-definite Hermitian products, a complex manifold $M$ can come equipped with a Hermitian metric, namely a positive-definite Hermitian product $h_m$ on $T_mM$ for every point $m \in M$ depending smoothly on the point $m$; every Hermitian metric induces a Riemannian one given by its real part. The complex projective space $\mathbb{C}P^n$ comes naturally equipped with the Fubini–Study Hermitian metric.

## Transformation Groups

Metric properties can be seen from the point of view of transformation groups. Poncelet in his *Traité projectif des figures* (1822) had investigated classical Euclidean geometry from a projective geometric point of view, but it was not until Cayley (1858) that metric properties were interpreted as those stable under any "projective" transformation which leaves "cyclic points" (points at infinity on the imaginary axis of the complex plane) invariant. Transformation groups were further investigated by Lie, leading to the modern concept of Lie group, a smooth manifold endowed with a group structure such that the group operations are smooth.

A vector field $X$ on a Lie group $G$ is called left- (resp. right-) invariant if it is invariant under left translations $L_g : h \mapsto gh$ (resp. right translations $R_g : h \mapsto hg$) for every $g \in G$, that is, if $(L_g)_* X(h) = X(gh)\ \forall (g, h) \in G^2$ (resp. $(R_g)_* X(h) = X(gh)\ \forall (g, h) \in G^2$). The set of all left-invariant vector fields equipped with the sum, scalar multiplication, and the bracket operation on vector fields form an algebra called the Lie algebra of $G$.

The group $\mathrm{Gl}_n(\mathbb{R})$ (resp. $\mathrm{Gl}_n(\mathbb{C})$) of all real (resp. complex) invertible $n \times n$ matrices is a Lie group with Lie algebra, the algebra $\mathrm{gl}_n(\mathbb{R})$ (resp. $\mathrm{gl}_n(\mathbb{C})$) of all real (resp. complex) $n \times n$ matrices and the bracket operation reads $[A, B] = AB - BA$.

The orthogonal (resp. unitary) group $\mathrm{O}_n(\mathbb{R}) := \{A \in \mathrm{Gl}_n(\mathbb{R}), A^{\mathrm{t}}A = 1\}$, where $A^{\mathrm{t}}$ denotes the transposed matrix (resp. $U_n(\mathbb{C}) := \{A \in \mathrm{Gl}_n(\mathbb{C}), A^*A = 1\}$, where $A^* = \bar{A}^{\mathrm{t}}$), is a compact Lie group with Lie

algebra $o_n(\mathbb{R}) := \{A \in \mathrm{Gl}_n(\mathbb{R}), A^t = -A\}$ (resp. $u_n(\mathbb{C}) := \{A \in \mathrm{Gl}_n(\mathbb{C}), A^* = -A\}$).

A left-invariant vector field $X$ on a finite-dimensional Lie group $G$ (or equivalently an element $X$ of the Lie algebra of $G$) generates a global one-parameter group of transformations $\phi_X(t), t \in \mathbb{R}$. The mapping from the Lie algebra of $G$ into $G$ defined by $\exp(X) := \phi_X(1)$ is called the exponential mapping. The exponential mapping on $\mathrm{Gl}_n(\mathbb{R})$ (resp. $\mathrm{Gl}_n(\mathbb{C})$) is given by the series $\exp(A) = \sum_{i=0}^{\infty} A^i/i!$.

As symmetry groups of physical systems, Lie groups play an important role in physics, in particular in quantum mechanics and Yang–Mills theory. Infinite-dimensional Lie groups arise as symmetry groups, such as the group of diffeomorphisms of a manifold in general relativity, the group of gauge transformations in Yang–Mills theory, and the group of Weyl transformations of metrics on a surface in string theory. The principle "the physics should not depend on how it is described" translates to an invariance under the action of the (possibly infinite-dimensional group) of symmetries of the theory. Anomalies arise when such an invariance holds for the classical action of a physical theory but "breaks" at the quantized level.

In his Erlangen program (1872), Klein puts the concept of transformation group in the foreground introducing a novel idea by which one should consider a space endowed with some properties as a set of objects invariant under a given group of transformations. One thereby reaches a classification of geometric results according to which group is relevent in a particular problem as, for example, the projective linear group for projective geometry, the orthogonal group for Riemannian geometry, or the symplectic group for "symplectic" geometry.

## Fiber Bundles

Transformation groups give rise to principal fiber bundles which play a major role in Yang–Mills theory. The notion of fiber bundle first arose out of questions posed in the 1930s on the topology and the geometry of manifolds, and by 1950 the definition of fiber bundle had been clearly formulated by Steenrod.

A smooth fiber bundle with typical fiber a manifold $F$ is a triple $(E, \pi, B)$, where $E$ and $B$ are smooth manifolds called the total space and the base space, and $\pi : E \to B$ is a smooth surjective map called the projection of the bundle such that the preimage $\pi^{-1}(b)$ of a point $b \in B$ called the fiber of the bundle over $b$ is isomorphic to $F$ and any base point $b$ has a neighborhood $U \subset B$ with preimage $\pi^{-1}(U)$ diffeomorphic to $U \times F$, where the diffeomophisms commute with the projection on the base space. Smooth sections of $E$ are maps $\sigma : B \to E$ such that $\pi \circ \sigma = I_B$.

When $F$ is a vector space and when, given open subsets $U_i \subset B$ that cover $B$ with corresponding coordinate charts $(U_i, \phi_i)_{i\in I}$, the local diffeomorphisms $\tau_i : \pi^{-1}(U_i) \simeq \phi_i(U_i) \times F$ give rise to transition maps $\tau_i \circ \tau_j^{-1} : \phi_j(U_i \cap U_j) \times F \to \phi_i(U_i \cap U_j) \times F$ that are linear in the fiber, the bundle is called a "vector bundle." The tangent bundle $TM = \bigcup_{m\in M} T_m M$ to a differentiable manifold $M$ modeled on a vector space $V$ is a vector bundle with typical fiber $V$ and transition maps $\tau_{ij} = (\phi_i \circ \phi_j^{-1}, \mathrm{d}(\phi_i \circ \phi_j^{-1}))$ expressed in terms of the differentials of the transition maps on the manifold $M$. So are the cotangent bundle, the dual of the tangent bundle, and tensor products of the tangent and cotangent vector bundles with typical fiber the dual $V^*$ and tensor products of $V$ and $V^*$. Vector fields defined previously are sections of the tangent bundle, 1-forms on $M$ are sections of the cotangent bundle, and contravariant tensors, resp. covariant tensors are sections of tensor products of the tangent, resp. cotangent bundles. A differentiable mapping $\phi : M \to N$ takes covariant $p$-tensor fields on $N$ to their pullbacks by $\phi$, covariant $p$-tensors on $M$ given by

$$(\phi^* T)(X_1, \ldots, X_p) := T(\phi_* X_1, \ldots, \phi_* X_p)$$

for any vector fields $X_1, \ldots, X_p$ on $M$.

Differentiating a smooth function $f$ on $M$ gives rise to a 1-form $\mathrm{d}f$ on $M$. More generally, exterior $p$-forms are antisymmetric smooth covariant $p$-tensors so that $\omega(X_{\sigma(1)}, \ldots, X_{\sigma(p)}) = \epsilon(\sigma)\omega(X_1, \ldots, X_p)$ for any vector fields $X_1, \ldots, X_p$ on $M$ and any permutation $\sigma \in \Sigma_p$ with signature $\epsilon(\sigma)$.

Riemannian metrics are covariant 2-tensors and the space of Riemannian metrics on a manifold $M$ is an infinite-dimensional manifold which arises as a configuration space in string theory and general relativity.

A principal bundle is a fiber bundle $(P, \pi, B)$ with typical fiber a Lie group $G$ acting freely and properly on the total space $P$ via a right action $(p, g) \in P \times G \mapsto pg = R_g(p) \in P$ and such that the local diffeomorphisms $\pi^{-1}(U) \simeq U \times G$ are $G$-equivariant. Given a principal fiber bundle $(P, \pi, B)$ with structure group a finite-dimensional Lie group $G$, the action of $G$ on $P$ induces a homomorphism which to an element $X$ of the Lie algebra of $G$ assigns a vector field $X^*$ on $P$ called the "fundamental vector field" generated by $X$. It is defined at $p \in P$ by

$$X^*(p) := \frac{\mathrm{d}}{\mathrm{d}t_{|_{t=0}}} R_{\exp(tX)}(p)$$

where exp is the exponential map on $G$.

Given an action of $G$ on a vector space $V$, one builds from a principal bundle with typical fiber $G$ an associated vector bundle with typical fiber $V$. Principal bundles are essential in gauge theory; $U(1)$-principal bundles arise in electro-magnetism and nonabelian structure groups arise in Yang–Mills theory. There the fields are connections on the principal bundle, and the action of gauge transformations on (irreducible) connections gives rise to an infinite-dimensional principal bundle over the moduli space with structure group given by gauge transformations. Infinite-dimensional bundles arise in other field theories such as string theory where the moduli space corresponds to inequivalent complex structures on a Riemann surface and the infinite-dimensional structure group is built up from Weyl transformations of the metric and diffeomorphisms of the surface.

## Connections

On a manifold there is no canonical method to identify tangent spaces at different points. Such an identification, which is needed in order to differentiate vector fields, can be achieved on a Riemannian manifold via "parallel transport" of the vector fields. The basic concepts of the theory of covariant differentiation on a Riemannian manifold were given at the end of the nineteenth century by Ricci and, in a more complete form, in 1901 in collaboration with Levi-Civita in *Méthodes de calcul différentiel absolu et leurs applications*; on a Riemannian manifold, it is possible to define in a canonical manner a parallel displacement of tangent vectors and thereby to differentiate vector field covariantly using the since then called Levi-Civita connection.

More generally, a (linear) connection (or equivalently a covariant derivation) on a vector bundle $E$ over a manifold $M$ provides a way to identify fibers of the vector bundle at different points; it is a map $\nabla$ taking sections $\sigma$ of $E$ to $E$-valued 1-forms on $M$ which satisfies a Leibniz rule, $\nabla(f\sigma) = \mathrm{d}f\sigma + f\nabla\sigma$, for any smooth function $f$ on $M$. When $E$ is the tangent bundle over $M$, curves $\gamma$ on the manifold with covariantly constant velocity $\nabla\dot{\gamma}(t) = 0$ give rise to geodesics. Given an initial velocity $\dot{\gamma}(0) = X \in T_mM$ and provided $X$ has small enough norm, $\gamma_X(1)$ defines a point on the corresponding geodesic and the map $exp : X \mapsto \gamma_X(1)$ a diffeomorphism from a neighborhood of 0 in $T_mM$ to a neighborhood of $m \in M$ called the "exponential map" of $\nabla$.

The concept of connection extends to principal bundles where it was developed by Ehresmann building on the work of Cartan. A connection on a principal bundle $(P, \pi, B)$ with structure group $G$, which is a smooth equivariant (under the action of the group $G$) decomposition of the tangent space $T_pP = H_pP \oplus V_pP$ at each point $p$ into a horizontal space $H_pP$ and the vertical space $V_pP = \mathrm{Ker}\, \mathrm{d}\pi_p$, gives rise to a linear connection on the associated vector bundle.

A connection on $P$ gives rise to a 1-form $\omega$ on $P$ with values in the Lie algebra of the structure group $G$ called the connection 1-form and defined as follows. For each $X \in T_pP, \omega(X)$ is the unique element $U$ of the Lie algebra of $G$ such that the corresponding fundamental vector field $U^*(p)$ at point $p$ coincides with the vertical component of $X$. In particular, $\omega(U^*) = U$ for any element $U$ of the Lie algebra of $G$.

The space of connections which is an infinite-dimensional manifold arises as a configuration space in Yang–Mills theory and also comes into play in the Seiberg–Witten theory.

## Geometric Differential Operators

From connections one defines a number of differential operators on a Riemannian manifold, among them second-order Laplacians. In particular, the Laplace–Beltrami operator $f \mapsto -\mathrm{tr}(\nabla^{T*M}\,\mathrm{d}f)$ on smooth functions, where $\nabla^{T^*M}$ is the connection on the cotangent bundle induced by the Levi-Civita connection on $M$, generalizes the ordinary Laplace operator on Euclidean space. This in turn generalizes to second-order operators $\Delta^E := -\mathrm{tr}(\nabla^{T^*M\otimes E}\nabla^E)$ acting on smooth sections of a vector bundle $E$ over a Riemannian manifold $M$, where $\nabla^E$ is a connection on $E$ and $\nabla^{T^*M\otimes E}$ the connection on $T^*M \otimes E$ induced by $\nabla^E$ and the Levi-Civita connection on $M$.

The Dirac operator on a spin Riemannian manifold, a first-order differential operator whose square coincides with the Laplace–Beltrami operator up to zeroth-order terms, can be best understood going back to the initial idea of Dirac. A first-order differential operator with constant matrix coefficients $\sum_{i=1}^n \gamma_i(\partial/\partial x_i)$ has square given by the Laplace operator $-\sum_{i=1}^n \partial^2/\partial x_i^2$ on $\mathbb{R}^n$ if and only if its coefficients satisfy the the Clifford relations

$$\gamma_i^2 = -1 \quad \forall\, i = 1, \ldots, n$$
$$\gamma_i\gamma_j + \gamma_j\gamma_i = 0 \quad \forall\, i \neq j$$

The resulting Clifford algebra, once complexified, is isomorphic in even dimensions $n = 2k$ to the space $\mathrm{End}(S_n)$ (and $\mathrm{End}(S_n) \oplus \mathrm{End}(S_n)$ in odd dimensions $n = 2k + 1$) of endomorphisms of the space $S_n = \mathbb{C}^{2^k}$ of complex $n$-spinors. When instead of the canonical metric on $\mathbb{R}^n$ one starts from the the metric on the tangent bundle $TM$ induced by the Riemannian

metric on $M$ and provided the corresponding spinor spaces patch up to a "spinor bundle" over $M$, $M$ is called a spin manifold. The Dirac operator on a spin Riemannian manifold $M$ is a first-order differential operator acting on spinors given by $D_g = \sum_{i=1}^{n} \gamma_i \nabla_{e_i}$, where $\nabla$ is the connection on spinors (sections of the spinor bundle $S$) induced by the Levi-Civita connection and $e_1, \ldots, e_n$ is an orthonormal frame of the tangent bundle $TM$. This is a particular case of more general twisted Dirac operators $D_g^W$ on a twisted spinor bundle $S \otimes W$ equipped with the connection $\nabla^{S\otimes W}$ which combines the connection $\nabla$ with a connection $\nabla^W$ on an auxilliary vector bundle $W$. Their square $(D_g^W)^2$ relates to the Laplacian $\Delta^{S\otimes W}$ built from this twisted connection via the Lichnerowicz formula which is useful for estimates on the spectrum of the Dirac operator in terms of the underling geometric data.

When there is no spin structure on $M$, one can still hope for a Spin$^c$ structure and a Dirac $D^c$ operator associated with a connection compatible with that structure. In particular, every compact orientable 4-manifold can be equipped with a Spin$^c$ structure and one can build invariants of the differentiable manifold called Seiberg–Witten invariants from solutions of a system of two partial differential equations, one of which is the Dirac equation $D^c\Phi = 0$ associated with a connection compatible with the Spin$^c$ structure and the other a nonlinear equation involving the curvature.

## Curvature

The concept of "curvature," which is now understood in terms of connections (the curvature of a connection $\nabla$ is defined by $\Omega = \nabla^2$), historically arose prior to that of connection. In its modern form, the concept of curvature dates back to Gauss. Using a spherical representation of surfaces – the Gauss map $\nu$, which sends a point $m$ of an oriented surface $\Sigma \subset \mathbb{R}^3$ to the outward pointing unit normal vector $\nu_m$ – Gauss defined what is since then called the Gaussian curvature $K_m$ at point $m \in U \subset \Sigma$ as the limit when the area of $U$ tends to zero of the ratio $\text{area}(\nu(U))/\text{area}(U)$. It measures the obstruction to finding a distance-preserving map from a piece of the surface around $m$ to a region in the standard plane. Gauss' *Teorema Egregium* says that the Gaussian curvature of a smooth surface in $\mathbb{R}^3$ is defined in terms of the metric on the surface so that it agrees for two isometric surfaces.

From the curvature $\Omega$ of a connection on a Riemannian manifold $(M, g)$, one builds the Riemannian curvature tensor, a 4-tensor which in local coordinates reads

$$R_{ijkl} := g\left(\Omega\left(\frac{\partial}{\partial_i}, \frac{\partial}{\partial_j}\right)\frac{\partial}{\partial_k}, \frac{\partial}{\partial_l}\right)$$

further taking a partial trace leads to the Ricci curvature given by the 2-tensor $\text{Ric}_{ij} = \sum_k R_{ikjk}$, the trace of which gives in turn the scalar curvature $R = \sum_i \text{Ric}_{ii}$. Sectional curvature at a point $m$ in the direction of a two-dimensional plane spanned by two vectors $U$ and $V$ corresponds to $K(U, V) = g(\Omega(U, V)V, U)$. A manifold has constant sectional curvature whenever $K(U, V)/\|U \wedge V\|^2$ is a constant $K$ for all linearly independent vectors $U, V$. A Riemannian manifold with constant sectional curvature is said to be spherical, flat, or hyperbolic type depending on whether $K > 0, K = 0$, or $K < 0$, respectively. One owes to Cartan the discovery of an important class of Riemannian manifolds, symmetric spaces, which contains the spheres, the Euclidean spaces, the hyperbolic spaces, and compact Lie groups. A connected Riemannian manifold $M$ equipped at every point $m$ with an isometry $\sigma_m$ such that $\sigma_m(m) = m$ and the tangent map $T_m\sigma_m$ equals –Id on the tangent space (it therefore reverses the geodesics through $m$) is called symmetric. $\mathbb{C}P^n$ equipped with the Fubini–Study metric is a symmetric space with the isometry given by the reflection with respect to a line in $\mathbb{C}^{n+1}$. A compact symmetric space has non-negative sectional curvature $K$.

Constraints on the curvature can have topological consequences. Spheres are the only simply connected manifolds with constant positive sectional curvature; if a simply connected complete Riemannian manifold of dimension $>1$ has non-positive sectional curvature along every plane, then it is homeomorphic to the Euclidean space.

A manifold with Ricci curvature tensor proportional to the metric tensor is called an Einstein manifold. Since Einstein, curvature is a cornerstone of general relativity with gravitational force being interpreted in terms of curvature. For example, the vacuum Einstein equation reads $\text{Ric}_g = (1/2)R_g\, g$ with $\text{Ric}_g$ the Ricci curvature of a metric $g$ and $R_g$ its scalar curvature. In addition, Kaluza–Klein supergravity is a unified theory modeled on a direct product of the Mikowski four-dimensional space and an Einstein manifold with positive scalar curvature.

The Ricci flow $\text{d}g(t)/\text{d}t = -2\text{Ric}_{g(t)}$, which is related with the Einstein equation in general relativity, was only fairly recently introduced in the mathematical literature. Hopes are strong to get a classification of closed 3-manifolds using the Ricci flow as an essential ingredient.

## Cohomology

Differentiation of functions $f \mapsto \mathrm{d}f$ on a differentiable manifold $M$ generalizes to exterior differentiation $\alpha \mapsto d\alpha$ of differential forms. A form $\alpha$ is closed whenever it is in the kernel of $d$ and it is exact whenever it lies in the range of $d$. Since $d^2 = 0$, exact forms are closed.

Cartan's structure equations $d\omega = -(1/2)[\omega, \omega] + \Omega$ relate the exterior differential of the connection 1-form $\omega$ on a principal bundle to its curvature $\Omega$ given by the exterior covariant derivative $D\omega := d\omega \circ h$, where $h : T_pP \to H_pP$ is the projection onto the horizontal space.

On a complex manifold, forms split into sums of $(p, q)$-forms, those with $p$-holomorphic and $q$-antiholomorphic components, and exterior differentiation splits as $d = \partial + \bar{\partial}$ into holomorphic and antiholomorphic derivatives, with $\partial^2 = \bar{\partial}^2 = 0$.

Geometric data are often expressed in terms of closedness conditions on certain differential forms. For example, a "symplectic manifold" is a manifold $M$ equipped with a closed nondegenerate differential 2-form called the "symplectic form." The theory of $J$-holomorphic curves on a manifold equipped with an almost-complex structure $J$ has proved fruitful in building invariants on symplectic manifolds. A Kähler manifold is a complex manifold equipped with a Hermitian metric $h$ whose imaginary part $\mathrm{Im}\, h$ yields a closed $(1, 1)$-form. The complex projective space $\mathbb{C}P^n$ is Kähler.

The exterior differentation $d$ gives rise to de Rham cohomology as $\mathrm{Ker}\, d/\mathrm{Im}\, d$, and de Rham's theorem establishes an isomorphism between de Rham cohomology and the real singular cohomology of a manifold. Chern (or characteristic) classes are topological invariants associated to fiber bundles and play a crucial role in index theory. Chern–Weil theory builds representatives of these de Rham cohomology classes from a connection $\nabla$ of the form $\mathrm{tr}(f(\nabla^2))$, where $f$ is some analytic function.

When the manifold is Riemannian, the Laplace–Beltrami operator on functions generalizes to differential forms in two different ways, namely to the Bochner Laplacian $\Delta^{\Lambda T^* M}$ on forms (i.e., sections of $\Lambda T^* M$), where the contangent bundle $T^* M$ is equipped with a connection induced by the Levi-Civita connection and to the Laplace–Beltrami operator on forms $(d + d^*)^2 = d^* d + d\, d^*$, where $d^*$ is the (formal) adjoint of the exterior differential $d$. These are related via Weitzenböck's formula which in the particular case of 1-forms states that the difference of those two operators is measured by the Ricci curvature.

When the manifold is compact, Hodge's theorem asserts that the de Rham cohomology groups are isomorphic to the space of harmonic (i.e., annihilated by the Laplace–Beltrami operator) differential forms. Thus, the dimension of the set of harmonic $k$-forms equals the $k$th Betti numbers from which one can define the Euler characteristic $\chi(M)$ of the manifold $M$ taking their alternate sum. Hodge theory plays an important role in mirror symmetry which posits a duality between different manifolds on the geometric side and between different field theories via their correlation functions on the physics side. Calabi–Yau manifolds, which are Ricci-flat Kähler manifolds, are studied extensively in the context of duality.

## Index Theory

While the Gaussian curvature is the solution to a local problem, it has strong influence on the global topology of a surface. The Gauss–Bonnet formula (1850) relates the Euler characteristic on a closed surface to the Gaussian curvature by

$$\chi(M) = \frac{1}{2\pi} \int_M K_m \, \mathrm{d}A_m$$

where $\mathrm{d}A_m$ is the volume element on $M$. This is the first result relating curvature to global properties and can be seen as one of the starting points for index theory. It generalizes to the Chern–Gauss–Bonnet theorem (1944) on an even-dimensional closed manifold and can be interpreted as an example of the Atiyah–Singer index theorem (1963)

$$\mathrm{ind}(D_g^W) = \int_M \hat{A}(\Omega_g) \, \mathrm{e}^{-\mathrm{tr}(\Omega^W)}$$

where $g$ denotes a Riemannian metric on a spin manifold $M$, $D_g^W$ a Dirac operator acting on sections of some twisted bundle $S \otimes W$ with $S$ the spinor bundle on $M$ and $W$ an auxiliary vector bundle over $M$, $\mathrm{ind}(D_g^W)$ the "index" of the Dirac operator, and $\Omega_g$, $\Omega^W$ respectively the curvatures of the Levi-Civita connection and a connection on $W$, and $\hat{A}(\Omega_g)$ a particular Chern form called the $\hat{A}$-genus. Index theorems are useful to compute anomalies in gauge theories arising from functional quantisation of classical actions.

Given an even-dimensional closed spin manifold $(M, g)$ and a Hermitian vector bundle $W$ over $M$, the index of the associated Dirac operator $D_g^W$ yields the so-called Atiyah map $K^0(M) \mapsto \mathbb{Z}$ defined by $W \mapsto \mathrm{ind}(D_g^W)$, where $K^0(M)$ is the group of formal differences of stable homotopy classes of smooth vector bundles over $M$. This is the starting point for the noncommutative geometry approach to index theory, in which the space of smooth functions on a

manifold which arises here in a disguised from since $K^0(M) \simeq K_0(C^\infty(M))$ (which consists of formal differences of smooth homotopy classes of idempotents in the inductive limit of spaces of matrices $\mathrm{gl}_n(C^\infty(M))$) is generalized to any noncommutative smooth algebra.

## Further Reading


Bishop R and Crittenden R (2001) *Geometry of Manifolds*. Providence, RI: AMS Chelsea Publishing.

Chern SS, Chen WH, and Lam KS (2000) Lectures on Differential Geometry, *Series on University Mathematics*. Singapore: World Scientific.

Choquet-Bruhat Y, de Witt-Morette C, and Dillard-Bleick M (1982) *Analysis, Manifolds and Physics*, 2nd edn. Amsterdam–New York: North Holland.

Gallot S, Hulin D, and Lafontaine J (1993) Riemannian Geometry, *Universitext*. Berlin: Springer.

Helgason S (2001) *Differential, Lie Groups and Symmetric Spaces*. Graduate Studies in Mathematics 36. AMS, Providence, RI.

Husemoller D (1994) *Fibre Bundles*, 3rd edn. Graduate Texts in Mathematics 20. New York: Springer Verlag.

Jost J (1998) Riemannian Geometry and Geometric Analysis, *Universitext*. Berlin: Springer.

Klingenberg W (1995) *Riemannian Geometry*, 2nd edn. Berlin: de Gruyter.

Kobayashi S and Nomizou K (1996) *Foundations of Differential Geometry I, II*. Wiley Classics Library, a Wiley-Interscience Publication. New York: Wiley.

Lang S (1995) *Differential and Riemannian Manifolds*, 3rd edn. Graduate Texts in Mathematics, 160. New York: Springer Verlag.

Milnor J (1997) *Topology from the Differentiate Viewpoint*. Princeton Landmarks in Mathematics. Princeton, NJ: Princeton University Press.

Nakahara M (2003) *Geometry, Topology and Physics*, 2nd edn. Bristol: Institute of Physics.

Spivak M (1979) *A Comprehensive Introduction to Differential Geometry*, vols. 1, 2 and 3. Publish or Perish Inc., Wilmington, Delaware.

Sternberg S (1983) *Lectures on Differential Geometry*, 2nd edn. New York: Chelsea Publishing Co.


# Introductory Article: Electromagnetism

**N M J Woodhouse**, University of Oxford, Oxford, UK



This article is adapted from Chapters 2 and 3 of *Special Relativity*, N M J Woodhouse, Springer-Verlag, 2002, by kind permission of the publisher.

## Introduction

The modern theory of electromagnetism is built on the foundations of Maxwell's equations:

$$\operatorname{div} \boldsymbol{E} = \frac{\rho}{\epsilon_0} \quad [1]$$

$$\operatorname{div} \boldsymbol{B} = 0 \quad [2]$$

$$\operatorname{curl} \boldsymbol{B} - \frac{1}{c^2}\frac{\partial \boldsymbol{E}}{\partial t} = \mu_0 \boldsymbol{J} \quad [3]$$

$$\operatorname{curl} \boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} = 0 \quad [4]$$

On the left-hand side are the electric and magnetic fields, $\boldsymbol{E}$ and $\boldsymbol{B}$, which are vector-valued functions of position and time. On the right are the sources, the charge density $\rho$, which is a scalar function of position and time, and the current density $\boldsymbol{J}$. The source terms encode the distribution and velocities of charges, and the equations, together with boundary conditions at infinity, determine the fields that they generate. From these equations, one can derive the familiar predictions of electrostatics and magnetostatics, as well as the dynamical behavior of fields and charges, in particular, the generation and propagation of electromagnetic waves – light waves.

Maxwell would not have recognized the equations in this compact vector notation – still less in the tensorial form that they take in special relativity. It is notable that although his contribution is universally acknowledged in the naming of the equations, it is rare to see references to "Maxwell's theory." This is for a good reason. In his early studies of electromagnetism, Maxwell worked with elaborate mechanical models, which he saw as analogies rather than as literal descriptions of the underlying physical reality. In his later work, the mechanical models, in particular the mechanical properties of the "lumiferous ether" through which light waves propagate, were put forward more literally as the foundations of his electromagnetic theory. The equations survive in the modern theory, but the mechanical models with which Maxwell, Faraday, and others wrestled live on only in the survival of archaic terminology, such as "lines of force" and "magnetic flux." The luminiferous ether evaporated with the advent of special relativity.

Maxwell's legacy is not his "theory," but his equations: a consistent system of partial differential equations that describe the whole range of known interactions of electric and magnetic fields with

moving charges. They unify the treatment of electricity and magnetism by revealing for the first time the full duality between the electric and magnetic fields. They have been verified over an almost unimaginable variety of physical processes, from the propagation of light over cosmological distances, through the behavior of the magnetic fields of stars and the everyday applications in electrical engineering and laboratory experiments, down – in their quantum version – to the exchange of photons between individual electrons.

The history of Maxwell's equations is convoluted, with many false turns. Maxwell himself wrote down an inconsistent form of the equations, with a different sign for $\rho$ in the first equation, in his 1865 work "A dynamical theory of the electromagnetic field." The consistent form appeared later in his *Treatise on Electricity and Magnetism* (1873); see Chalmers (1975).

In this article, we shall not follow the historical route to the equations. Some of the complex story of the development hinted at in the remarks above can be found in the articles by Chalmers (1975), Siegel (1985), and Roche (1998). Neither shall we follow the traditional pedagogic route of many textbooks in building up to the full dynamical equations through the study of basic electrical and magnetic phenomena. Instead, we shall follow a path to Maxwell's equations that is informed by knowledge of their most critical feature, invariance under Lorentz transformations. Maxwell, of course, knew nothing of this.

We shall start with a summary of basic facts about the behavior of charges in electric and magnetic fields, and then establish the full dynamical framework by considering this behavior as seen from moving frames of reference. It is impossible, of course, to do this consistently within the framework of classical ideas of space and time since Maxwell's equations are inconsistent with Galilean relativity. But it is at least possible to understand some of the key features of the equations, in particular the need for the term involving the time derivative of $\boldsymbol{E}$, the so-called "displacement current," in the third of Maxwell's equations.

We shall begin with some remarks concerning the role of relativity in classical dynamics.

## Relativity in Newtonian Dynamics

Newton's laws hold in all inertial frames. The formalism of classical mechanics is invariant under Galilean transformations and it is impossible to tell by observing the dynamical behavior of particles and other bodies whether a frame of reference is at rest or in uniform motion. In the world of classical mechanics, therefore:

**Principle of Relativity** There is no absolute standard of rest; only relative motion is observable.

In his "Dialogue concerning the two chief world systems," Galileo illustrated the principle by arguing that the uniform motion of a ship on a calm sea does not affect the behavior of fish, butterflies, and other moving objects, as observed in a cabin below deck.

Relativity theory takes the principle as fundamental, as a statement about the nature of space and time as much as about the properties of the Newtonian equations of motion. But if it is to be given such universal significance, then it must apply to all of physics, and not just to Newtonian dynamics. At first this seems unproblematic – it is hard to imagine that it holds at such a basic level, but not for more complex physical interactions. Nonetheless, deep problems emerge when we try to extend it to electromagnetism since Galilean invariance conflicts with Maxwell's equations.

All appears straightforward for systems involving slow-moving charges and slowly varying electric and magnetic fields. These are governed by laws that appear to be invariant under transformations between uniformly moving frames of reference. One can imagine a modern version of Galileo's ship also carrying some magnets, batteries, semiconductors, and other electrical components. Salviati's argument for relativity would seem just as compelling.

The problem arises when we include rapidly varying fields – in particular, when we consider the propagation of light. As Einstein (1905) put it, "Maxwell's electrodynamics . . . , when applied to moving bodies, leads to asymmetries which do not appear to be inherent in the phenomena." The central difficulty is that Maxwell's equations give light, along with other electromagnetic waves, a definite velocity: in empty space, it travels with the same speed in every direction, independently of the motion of the source – a fact that is incompatible with Galilean invariance. Light traveling with speed $c$ in one frame should have speed $c + u$ in a frame moving towards the source of the light with speed $u$. Thus, it should be possible for light to travel with any speed. Light that travels with speed $c$ in a frame in which its source is at rest should have some other speed in a moving frame; so Galilean invariance would imply dependence of the velocity of light on the motion of the source.

A full resolution of the conflict can only be achieved within the special theory of relativity: here, remarkably, Maxwell's equations retain exactly

their classical form, but the transformations between the space and time coordinates of frames of reference in relative motion do not. The difference appears when the velocities involved are not insignificant when compared with the velocity of light. So long as one can ignore terms of order $u^2/c^2$, Maxwell's equations are compatible with the Galilean principle of relativity.

## Charges, Fields, and the Lorentz-Force Law

The basic objects in the modern form of electromagnetic theory are

- charged particles; and
- the electric and magnetic fields $\boldsymbol{E}$ and $\boldsymbol{B}$, which are vector quantities that depend on position and time.

The charge $e$ of a particle, which can be positive or negative, is an intrinsic quantity analogous to gravitational mass. It determines the strength of the particle's interaction with the electric and magnetic fields – as its mass determines the strength of its interaction with gravitational fields.

The interaction is in two directions. First, electric and magnetic fields exert a force on a charged particle which depends on the value of the charge, the particle's velocity, and the values of $\boldsymbol{E}$ and $\boldsymbol{B}$ at the location of the particle. The force is given by the Lorentz-force law

$$\boldsymbol{f} = e(\boldsymbol{E} + \boldsymbol{u} \wedge \boldsymbol{B}) \qquad [5]$$

in which $e$ is the charge and $\boldsymbol{u}$ is the velocity. It is analogous to the gravitational force

$$\boldsymbol{f} = m\boldsymbol{g} \qquad [6]$$

on a particle of mass $m$ in a gravitational field $\boldsymbol{g}$. It is through the force law that an observer can, in principle, measure the electric and magnetic fields at a point, by measuring the force on a standard charge moving with known velocity.

Second, moving charges generate electric and magnetic fields. We shall not yet consider in detail the way in which they do this, beyond stating the following basic principles.

**EM1.** The fields depend linearly on the charges.

This means that if we superimpose two distributions of charge, then the resultant $\boldsymbol{E}$ and $\boldsymbol{B}$ fields are the sums of the respective fields that the two distributions generate separately.

**EM2.** A stationary point charge $e$ generates an electric field, but no magnetic field. The electric field is given by

$$\boldsymbol{E} = \frac{ke\boldsymbol{r}}{r^3} \qquad [7]$$

where $\boldsymbol{r}$ is the position vector from the charge, $r = |\boldsymbol{r}|$, and k is a positive constant, analogous to the gravitational constant.

By combining [7] and [5], we obtain an inverse-square law electrostatic force

$$\frac{kee'}{r^2} \qquad [8]$$

between two stationary charges; unlike gravity, it is repulsive when the charges have the same sign.

**EM3.** A point charge moving with velocity $\boldsymbol{v}$ generates a magnetic field

$$\boldsymbol{B} = \frac{k' e\boldsymbol{v} \wedge \boldsymbol{r}}{r^3} \qquad [9]$$

where $k'$ is a second positive constant.

This is extrapolated from measurements of the magnetic field generated by currents flowing in electrical circuits.

The constants $k$ and $k'$ in EM2 and EM3 determine the strengths of electric and magnetic interactions. They are usually denoted by

$$k = \frac{1}{4\pi\epsilon_0}, \qquad k' = \frac{\mu_0}{4\pi} \qquad [10]$$

Charge $e$ is measured in coulombs, $|\boldsymbol{B}|$ in teslas, and $|\boldsymbol{E}|$ in volts per meter. With other quantities in SI units,

$$\epsilon_0 = 8.9 \times 10^{-12}, \qquad \mu_0 = 1.3 \times 10^{-6} \qquad [11]$$

The charge of an electron is $-1.6 \times 10^{-19}$ C; the current through an electric fire is a flow of 5–10 C s$^{-1}$. The earth's magnetic field is about $4 \times 10^{-5}$ T; a bar magnet's is about 1 T; there is a field of about 50 T on the second floor of the Clarendon Laboratory in Oxford; and the magnetic field on the surface of a neutron star is about $10^8$ T.

Although we are more aware of gravity in everyday life, it is very much weaker than the electrostatic force – the electrostatic repulsion between two protons is a factor of $1.2 \times 10^{36}$ greater than their gravitational attraction (at any separation, both forces obey the inverse-square law).

Our aim is to pass from EM1–EM3 to Maxwell's equations, by replacing [7] and [9] by partial differential equations that relate the field strengths to the charge and current densities $\rho$ and $\boldsymbol{J}$ of a

continuous distribution of charge. The densities are defined as the limits

$$\rho = \lim_{V\to 0}\left(\frac{\sum e}{V}\right), \qquad \boldsymbol{J} = \lim_{V\to 0}\left(\frac{\sum e\boldsymbol{v}}{V}\right) \qquad [12]$$

where $V$ is a small volume containing the point, $e$ is a charge within the volume, and $\boldsymbol{v}$ is its velocity; the sums are over the charges in $V$ and the limits are taken as the volume is shrunk (although we shall not worry too much about the precise details of the limiting process).

## Stationary Distributions of Charge

We begin the task of converting the basic principles into partial differential equations by looking at the electric field of a stationary distribution of charge, where the passage to the continuous limit is made by using the Gauss theorem to restate the inverse-square law.

The Gauss theorem relates the integral of the electric field over a closed surface to the total charge contained within it. For a point charge, the electric field is given by EM2:

$$\boldsymbol{E} = \frac{e\boldsymbol{r}}{4\pi\epsilon_0 r^3}$$

Since $\operatorname{div}\boldsymbol{r} = 3$ and $\operatorname{grad} r = \boldsymbol{r}/r$, we have

$$\operatorname{div}(\boldsymbol{E}) = \operatorname{div}\left(\frac{e\boldsymbol{r}}{\pi\epsilon_0 r^3}\right) = \frac{e}{4\pi\epsilon_0}\left(\frac{3}{r^3} - \frac{3\boldsymbol{r}\cdot\boldsymbol{r}}{r^5}\right) = 0$$

everywhere except at $r = 0$. Therefore, by the divergence theorem,

$$\int_{\partial V} \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{S} = 0 \qquad [13]$$

for any closed surface $\partial V$ bounding a volume $V$ that does not contain the charge.

What if the volume does contain the charge? Consider the region bounded by the sphere $S_R$ of radius $R$ centered on the charge; $S_R$ has outward unit normal $\boldsymbol{r}/r$. Therefore,

$$\int_{S_R} \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{S} = \frac{e}{4\pi R^2\epsilon_0}\int_{S_R} \mathrm{d}S = \frac{e}{\epsilon_0}$$

In particular, the value of the surface integral on the left-hand side does not depend on $R$.

Now consider arbitrary finite volume bounded by a closed surface $S$. If the charge is not inside the volume, then the integral of $\boldsymbol{E}$ over $S$ vanishes by [13]. If it is, then we can apply [13] to the volume $V$ between $S$ and a small sphere $S_R$ to deduce that

$$\int_S \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{S} - \int_{S_R} \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{S} = \int_{\partial V} \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{S} = 0$$

and that the integrals of $\boldsymbol{E}$ over $S$ and $S_R$ are the same. Therefore,

$$\int_S \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{S} = \begin{cases} e/\epsilon_0 & \text{if the charge is in the volume bounded by } S \\ 0 & \text{otherwise} \end{cases}$$

When we sum over a distribution of charges, the integral on the left picks out the total charge within $S$. Therefore, we have the Gauss theorem.

**The Gauss theorem.** For any closed surface $\partial V$ bounding a volume $V$,

$$\int_{\partial V} \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{S} = Q/\epsilon_0$$

where $\boldsymbol{E}$ is the total electric field and $Q$ is the total charge within $V$.

Now we can pass to the continuous limit. Suppose that $\boldsymbol{E}$ is generated by a distribution of charges with density $\rho$ (charge per unit volume). Then by the Gauss theorem,

$$\int_{\partial V} \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{S} = \frac{1}{\epsilon_0}\int_V \rho\, \mathrm{d}V$$

for any volume $V$. But then, by the divergence theorem,

$$\int_V (\operatorname{div}\boldsymbol{E} - \rho/\epsilon_0)\, \mathrm{d}V = 0$$

Since this holds for any volume $V$, it follows that

$$\operatorname{div}\boldsymbol{E} = \rho/\epsilon_0 \qquad [14]$$

By an argument in a similar spirit, we can also show that the electric field of a stationary distribution of charge is conservative in the sense that the total work done by the field when a charge is moved around a closed loop vanishes; that is,

$$\oint \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{s} = 0$$

for any closed path. This is equivalent to

$$\operatorname{curl}\boldsymbol{E} = 0 \qquad [15]$$

since, by Stokes' theorem,

$$\oint \boldsymbol{E}\cdot \mathrm{d}\boldsymbol{s} = \int_S \operatorname{curl}\boldsymbol{E}\cdot \mathrm{d}\boldsymbol{S}$$

where $S$ is any surface spanning the path. This vanishes for every path and for every $S$ if and only if [15] holds.

The field of a single stationary charge is conservative since

$$\boldsymbol{E} = -\mathrm{grad}\ \phi, \quad \phi = \frac{e}{4\pi\epsilon_0 r}$$

and therefore $\mathrm{curl}\,\boldsymbol{E} = 0$ since the curl of a gradient vanishes identically. For a continuous distribution, $\boldsymbol{E} = -\mathrm{grad}\ \phi$, where

$$\phi(\boldsymbol{r}) = \frac{1}{4\pi\epsilon_0} \int_{\boldsymbol{r}' \in V} \frac{\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|}\, \mathrm{d}V' \qquad [16]$$

In the integral, $\boldsymbol{r}$ (the position of the point at which $\phi$ is evaluated) is fixed, and the integration is over the positions $\boldsymbol{r}'$ of the individual charges. In spite of the singularity at $\boldsymbol{r} = \boldsymbol{r}'$, the integral is well defined. So, [15] also holds for a continuous distribution of stationary charge.

## The Divergence of the Magnetic Field

We can apply the same argument that established the Gauss theorem to the magnetic field of a slow-moving charge. Here,

$$\boldsymbol{B} = \frac{\mu_0 e \boldsymbol{v} \wedge \boldsymbol{r}}{4\pi r^3}$$

where $\boldsymbol{r}$ is the vector from the charge to the point at which the field is measured. Since $\boldsymbol{r}/r^3 = -\mathrm{grad}(1/r)$, we have

$$\mathrm{div}\left(\boldsymbol{v} \wedge \frac{\boldsymbol{r}}{r^3}\right) = \boldsymbol{v} \wedge \mathrm{curl}\left(\mathrm{grad}\frac{1}{r}\right) = 0$$

Therefore, $\mathrm{div}\,\boldsymbol{B} = 0$ except at $r = 0$, as in the case of the electric field. However, in the magnetic case, the integral of the field over a surface surrounding the charge also vanishes, since if $S_R$ is a sphere of radius $R$ centered on the charge, then

$$\int_{S_R} \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{S} = \frac{\mu_0 e}{4\pi} \int_{S_R} \frac{\boldsymbol{v} \wedge \boldsymbol{r}}{r^3} \cdot \frac{\boldsymbol{r}}{r}\, \mathrm{d}S = 0$$

By the divergence theorem, the same is true for any surface surrounding the charge. We deduce that if magnetic fields are generated only by moving charges, then

$$\int_{\partial V} \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{S} = 0$$

for any volume $V$, and hence that

$$\mathrm{div}\,\boldsymbol{B} = 0 \qquad [17]$$

Of course, if there were free "magnetic poles" generating magnetic fields in the same way that charges generate electric fields, then this would not hold; there would be a "magnetic pole density" on the right-hand side, by analogy with the charge density in [14].

## Inconsistency with Galilean Relativity

Our central concern is the compatibility of the laws of electromagnetism with the principle of relativity. As Einstein observed, simple electromagnetic interactions do indeed depend only on relative motion; the current induced in a conductor moving through the field of a magnet is the same as that generated in a stationary conductor when a magnet is moved past it with the same relative velocity (Einstein 1905). Unfortunately, this symmetry is not reflected in our basic principles. We very quickly come up against contradictions if we assume that they hold in every inertial frame of reference.

One emerges as follows. An observer $O$ can measure the values of $\boldsymbol{B}$ and $\boldsymbol{E}$ at a point by measuring the force on a particle of standard charge, which is related to the velocity $\boldsymbol{v}$ of the charge by the Lorentz-force law,

$$\boldsymbol{f} = e(\boldsymbol{E} + \boldsymbol{v} \wedge \boldsymbol{B})$$

A second observer $O'$ moving relative to the first with velocity $\boldsymbol{v}$ will see the same force, but now acting on a particle at rest. He will therefore measure the electric field to be $\boldsymbol{E}' = \boldsymbol{f}/e$. We conclude that an observer moving with velocity $\boldsymbol{v}$ through a magnetic field $\boldsymbol{B}$ and an electric field $\boldsymbol{E}$ should see an electric field

$$\boldsymbol{E}' = \boldsymbol{E} + \boldsymbol{v} \wedge \boldsymbol{B} \qquad [18]$$

By interchanging the roles of the two observers, we should also have

$$\boldsymbol{E} = \boldsymbol{E}' - \boldsymbol{v} \wedge \boldsymbol{B}' \qquad [19]$$

where $\boldsymbol{B}'$ is the magnetic field measured by the second observer. If both are to hold, then $\boldsymbol{B} - \boldsymbol{B}'$ must be a scalar multiple of $\boldsymbol{v}$.

But this is incompatible with EM3; if the fields are those of a point charge at rest relative to the first observer, then $\boldsymbol{E}$ is given by [7], and

$$\boldsymbol{B} = 0$$

On the other hand, the second observer sees the field of a point charge moving with velocity $-\boldsymbol{v}$. Therefore,

$$\boldsymbol{B}' = -\frac{\mu_0 e \boldsymbol{v} \wedge \boldsymbol{r}}{4\pi r^3}$$

So $\boldsymbol{B} - \boldsymbol{B}'$ is orthogonal to $\boldsymbol{v}$, not parallel to it.

This conspicuous paradox is resolved, in part, by the realization that EM3 is not exact; it holds only when the velocities are small enough for the magnetic force between two particles to be negligible in comparison with the electrostatic force. If $v$ is a typical velocity, then the condition is that $v^2\mu_0$

should be much less than $1/\epsilon_0$. That is, the velocities involved should be much less than

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} = 3 \times 10^8 \ \mathrm{m\,s^{-1}}$$

This, of course, is the velocity of light.

## The Limits of Galilean Invariance

Our basic principles EM1–EM3 must now be seen to be approximations – they describe the interactions of particles and fields when the particles are moving relative to each other at speeds much less than that of light. To emphasize that we cannot expect, in particular, EM3 to hold for particles moving at speeds comparable with $c$, we must replace it by

**EM3′**. A charge moving with velocity $\boldsymbol{v}$, where $v \ll c$, generates a magnetic field

$$\boldsymbol{B} = \frac{\mu_0 e \boldsymbol{v} \wedge \boldsymbol{r}}{4\pi r^3} + \mathrm{O}(v^2/c^2) \qquad [20]$$

The magnetic field of a system of charges in general motion satisfies

$$\operatorname{div} \boldsymbol{B} = 0 \qquad [21]$$

In the second part, we have retained [21] as a differential form of the statement that there are no free magnetic poles; the magnetic field is generated only by the motion of the charges. With this change, the theory is consistent with the principle of relativity, provided that we ignore terms of order $v^2/c^2$. The substitution of EM3′ for EM3 resolves the conspicuous paradox; the symmetry noted by Einstein between the current generated by the motion of the conductor in a magnetic field and by the motion of a magnet past a conductor is explained, provided that the velocities are much less than that of light.

The central problem remains however; the equations of electromagnetism are not invariant under a Galilean transformation with velocity comparable to $c$. The paradox is still there, but it is more subtle than it appeared to be at first. There are three possible ways out: (1) the noninvariance is real and has observable effects (necessarily of order $v^2/c^2$ or smaller); (2) Maxwell's theory is wrong; or (3) the Galilean transformation is wrong. Disconcertingly, it is the last path that physics has taken. But that is to jump ahead in the story. Our task is to complete the derivation of Maxwell's equations.

## Faraday's Law of Induction

The magnetic field of a slow-moving charge will always be small in relation to its electric field (even when we replace $\boldsymbol{B}$ by $c\boldsymbol{B}$ to put it into the same units as $\boldsymbol{E}$). The magnetic fields generated by currents in electrical circuits are not, however, dominated by large electric fields. This is because the currents are created by the flow, at slow velocity, of electrons, while overall the matter in the wire is roughly electrically neutral, with the electric fields of the positively charged nuclei and negatively charged electrons canceling.

This is the physical context to keep in mind in the following deduction of Faraday's law of induction from Galilean invariance for velocities much less than $c$. The law relates the electromotive force or "voltage" around an electrical circuit to the rate of change of the magnetic field $\boldsymbol{B}$ over a surface spanning the circuit. In its differential form, the law becomes one of Maxwell's equations.

Suppose first that the fields are generated by charges all moving relative to a given inertial frame of reference $R$ with the same velocity $\boldsymbol{v}$. Then in a second frame $R'$ moving relative to $R$ with velocity $\boldsymbol{v}$, there is a stationary distribution of charge. If the velocity is much less than that of light, then the electric field $\boldsymbol{E}'$ measured in $R'$ is related to the electric and magnetic $\boldsymbol{E}$ and $\boldsymbol{B}$ measured in $R$ by

$$\boldsymbol{E}' = \boldsymbol{E} + \boldsymbol{v} \wedge \boldsymbol{B}$$

Since the field measured in $R'$ is that of a stationary distribution of charge, we have

$$\operatorname{curl} \boldsymbol{E}' = 0$$

In $R$, the charges are all moving with velocity $\boldsymbol{v}$, so their configuration looks exactly the same from the point $\boldsymbol{r}$ at time $t$ as it does from the point $\boldsymbol{r} + \boldsymbol{v}\tau$ at time $t + \tau$. Therefore,

$$\boldsymbol{B}(\boldsymbol{r} + \boldsymbol{v}\tau, t + \tau) = \boldsymbol{B}(\boldsymbol{r}, t)$$
$$\boldsymbol{E}(\boldsymbol{r} + \boldsymbol{v}\tau, t + \tau) = \boldsymbol{E}(\boldsymbol{r}, t)$$

and hence by taking derivatives with respect to $\tau$ at $\tau = 0$,

$$\begin{aligned} \boldsymbol{v} \cdot \operatorname{grad} \boldsymbol{B} + \frac{\partial \boldsymbol{B}}{\partial t} &= 0 \\ \boldsymbol{v} \cdot \operatorname{grad} \boldsymbol{E} + \frac{\partial \boldsymbol{E}}{\partial t} &= 0 \end{aligned} \qquad [22]$$

So we must have

$$\begin{aligned} 0 &= \operatorname{curl} \boldsymbol{E}' \\ &= \operatorname{curl} \boldsymbol{E} + \operatorname{curl}(\boldsymbol{v} \wedge \boldsymbol{B}) \\ &= \operatorname{curl} \boldsymbol{E} + \boldsymbol{v} \operatorname{div} \boldsymbol{B} - \boldsymbol{v} \cdot \operatorname{grad} \boldsymbol{B} \\ &= \operatorname{curl} \boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} \end{aligned} \qquad [23]$$

since $\operatorname{div} \boldsymbol{B} = 0$. It follows that

$$\operatorname{curl} \boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} = 0 \qquad [24]$$

Equation [24] is linear in $\boldsymbol{B}$ and $\boldsymbol{E}$; so by adding the magnetic and electric fields of different streams of charges moving relative to $R$ with different velocities, we deduce that it holds generally for the electric and magnetic fields generated by moving charges.

Equation [24] encodes Faraday's law of electromagnetic induction, which describes how changing magnetic fields can generate currents. In the static case

$$\frac{\partial \boldsymbol{B}}{\partial t} = 0$$

and the equation reduces to $\operatorname{curl} \boldsymbol{E} = 0$ – the condition that the electrostatic field should be conservative; that is, it should do no net work when a charge is moved around a closed loop.

More generally, consider a wire loop in the shape of a closed curve $\gamma$. Let $S$ be a fixed surface spanning $\gamma$. Then we can deduce from eqn [24] that

$$\begin{aligned} \oint_\gamma \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{s} &= \int_S \operatorname{curl} \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{S} \\ &= -\int_S \frac{\partial \boldsymbol{B}}{\partial t} \cdot \mathrm{d}\boldsymbol{S} \\ &= -\frac{\mathrm{d}}{\mathrm{d}t} \int_S (\boldsymbol{B} \cdot \mathrm{d}\boldsymbol{S}) \end{aligned} \qquad [25]$$

If the magnetic field is varying, so that the integral of $\boldsymbol{B}$ over $S$ is not constant, then the integral of $\boldsymbol{E}$ around the loop will not be zero. There will be a nonzero electric field along the wire, which will exert a force on the electrons in the wire and cause a current to flow.

The quantity

$$\oint \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{s}$$

which is measured in volts, is the work done by the electric field when a unit charge makes one circuit of the wire. It is called the electromotive force around the circuit. The integral is the magnetic flux linking the circuit. The relationship [25] between electromotive force and rate of change of magnetic flux is Faraday's law.

## The Field of Charges in Uniform Motion

We can extract another of Maxwell's equations from this argument. By EM3′, a single charge $e$ with velocity $\boldsymbol{v}$ generates an electric field $\boldsymbol{E}$ and a magnetic field

$$\boldsymbol{B} = \frac{\mu_0 e \boldsymbol{v} \wedge \boldsymbol{r}}{4\pi r^3} + \mathrm{O}(v^2/c^2)$$

where $\boldsymbol{r}$ is the vector from the charge to the point at which the field is measured. In the frame of reference $R'$ in which the charge is at rest, its electric field is

$$\boldsymbol{E}' = \frac{e\boldsymbol{r}}{4\pi\epsilon_0 r^3}$$

In the frame in which it is moving with velocity $\boldsymbol{v}$, $\boldsymbol{E} = \boldsymbol{E}' + \mathrm{O}(v/c)$. Therefore,

$$c\boldsymbol{B} = \frac{\boldsymbol{v} \wedge \boldsymbol{E}'}{c} = \frac{\boldsymbol{v} \wedge \boldsymbol{E}}{c} + \mathrm{O}\left(\frac{v^2}{c^2}\right)$$

By taking the curl of both sides, and dropping terms of order $v^2/c^2$,

$$\begin{aligned} \operatorname{curl}(c\boldsymbol{B}) &= \operatorname{curl}\left(\frac{\boldsymbol{v} \wedge \boldsymbol{E}}{c}\right) \\ &= \frac{1}{c}(\boldsymbol{v} \operatorname{div} \boldsymbol{E} - \boldsymbol{v} \cdot \operatorname{grad} \boldsymbol{E}) \end{aligned}$$

But

$$\operatorname{div} \boldsymbol{E} = \rho/\epsilon_0, \qquad \boldsymbol{v} \cdot \operatorname{grad} \boldsymbol{E} = -\frac{\partial \boldsymbol{E}}{\partial t}$$

by [22]. Therefore,

$$\operatorname{curl}(c\boldsymbol{B}) - \frac{1}{c}\frac{\partial \boldsymbol{E}}{\partial t} = \frac{1}{c\epsilon_0}\boldsymbol{J} = c\mu_0 \boldsymbol{J}$$

where $\boldsymbol{J} = \rho\boldsymbol{v}$. By summing over the separate particle velocities, we conclude that

$$\operatorname{curl} \boldsymbol{B} - \frac{1}{c^2}\frac{\partial \boldsymbol{E}}{\partial t} = \mu_0 \boldsymbol{J}$$

holds for an arbitrary distribution of charges, provided that their velocities are much less than that of light.

## Maxwell's Equations

The basic principles, together with the assumption of Galilean invariance for velocities much less than that of light, have allowed us to deduce that the electric and magnetic fields generated by a continuous distribution of moving charges in otherwise empty space satisfy

$$\operatorname{div} \boldsymbol{E} = \frac{\rho}{\epsilon_0} \qquad [26]$$

$$\operatorname{div} \boldsymbol{B} = 0 \qquad [27]$$

$$\operatorname{curl} \boldsymbol{B} - \frac{1}{c^2}\frac{\partial \boldsymbol{E}}{\partial t} = \mu_0 \boldsymbol{J} \qquad [28]$$

$$\operatorname{curl} \boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} = 0 \qquad [29]$$

where $\rho$ is the charge density, $\boldsymbol{J}$ is the current density, and $c^2 = 1/\epsilon_0\mu_0$. These are Maxwell's equations, the basis of modern electrodynamics. Together with the Lorentz-force law, they describe the dynamics of charges and electromagnetic fields.

We have arrived at them by considering how basic electromagnetic processes appear in moving frames of reference – an unsatisfactory route because we have seen on the way that the principles on which we based the derivation are incompatible with Galilean invariance for velocities comparable with that of light. Maxwell derived them by analyzing an elaborate mechanical model of electric and magnetic fields – as displacements in the luminiferous ether. That is also unsatisfactory because the model has long been abandoned. The reason that they are accepted today as the basis of theoretical and practical applications of electromagnetism has little to do with either argument. It is first that they are self-consistent, and second that they describe the behavior of real fields with unreasonable accuracy.

## The Continuity Equation

It is not immediately obvious that the equations are self-consistent. Given $\rho$ and $\boldsymbol{J}$ as functions of the coordinates and time, Maxwell's equations are two scalar and two vector equations in the unknown components of $\boldsymbol{E}$ and $\boldsymbol{B}$. That is, a total of eight equations for six unknowns – more equations than unknowns. Therefore, it is possible that they are in fact inconsistent.

If we take the divergence of eqn [29], then we obtain

$$\frac{\partial}{\partial t}(\mathrm{div}\,\boldsymbol{B}) = 0$$

which is consistent with eqn [27]; so no problem arises here. However, by taking the divergence of eqn [28] and substituting from eqn [26], we get

$$\begin{aligned} 0 &= \mathrm{div\,curl}\,\boldsymbol{B} \\ &= \frac{1}{c^2}\frac{\partial}{\partial t}(\mathrm{div}\,\boldsymbol{E}) + \mu_0 \mathrm{div}\boldsymbol{J} \\ &= \mu_0\left(\frac{\partial \rho}{\partial t} + \mathrm{div}\boldsymbol{J}\right) \end{aligned}$$

This gives a contradiction unless

$$\frac{\partial \rho}{\partial t} + \mathrm{div}\boldsymbol{J} = 0 \qquad [30]$$

So the choice of $\rho$ and $\boldsymbol{J}$ is not unconstrained; they must be related by the continuity equation [30]. This holds for physically reasonable distributions of charge; it is a differential form of the statement that charges are neither created nor destroyed.

## Conservation of Charge

To see the connection between the continuity equation and charge conservation, let us look at the total charge within a fixed $V$ bounded by a surface $S$. If charge is conserved, then any increase or decrease in a short period of time must be exactly balanced by an inflow or outflow of charge across $S$.

Consider a small element $\mathrm{d}\boldsymbol{S}$ of $S$ with outward unit normal and consider all the particles that have a particular charge $e$ and a particular velocity $\boldsymbol{v}$ at time $t$. Suppose that there are $\sigma$ of these per unit volume ($\sigma$ is a function of position). Those that cross the surface element between $t$ and $t + \delta t$ are those that at time $t$ lie in the region of volume

$$|\boldsymbol{v}\cdot\boldsymbol{n}\,\mathrm{d}S\,\delta t|$$

shown in Figure 1. They contribute $e\sigma\boldsymbol{v}\cdot\mathrm{d}\boldsymbol{S}\delta t$ to the outflow of charge through the surface element. But the value of $\boldsymbol{J}$ at the surface element is the sum of $e\sigma\boldsymbol{v}$ over all possible values of $\boldsymbol{v}$ and $e$. By summing over $\boldsymbol{v}$, $e$, and the elements of the surface, therefore, and by passing to the limit of a continuous distribution, the total rate of outflow is

$$\int_S \boldsymbol{J}\cdot\mathrm{d}\boldsymbol{S}$$

Charge conservation implies that the rate of outflow should be equal to the rate of decrease in the total charge within $V$. That is,

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_V \rho\,\mathrm{d}V + \int_S \boldsymbol{J}\cdot\mathrm{d}\boldsymbol{S} = 0 \qquad [31]$$

By differentiating the first term under the integral sign and by applying the divergence theorem to the second integral,

$$\int_V \left(\frac{\partial \rho}{\partial t} + \mathrm{div}\boldsymbol{J}\right)\mathrm{d}V = 0 \qquad [32]$$

If this is to hold for any choice of $V$, then $\rho$ and $\boldsymbol{J}$ must satisfy the continuity equation. Conversely, the continuity equation implies charge conservation.

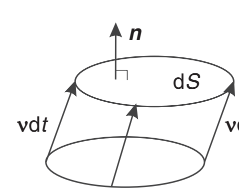



**Figure 1** The outflow through a surface element.

## The Displacement Current

The third of Maxwell's equations can be written as

$$\operatorname{curl} \boldsymbol{B} = \mu_0 \left( \boldsymbol{J} + \epsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} \right) \quad [33]$$

in which form it can be read as an equation for an unknown magnetic field $\boldsymbol{B}$ in terms of a known current distribution $\boldsymbol{J}$ and electric field $\boldsymbol{E}$. When $\boldsymbol{E}$ and $\boldsymbol{J}$ are independent of $t$, it reduces to

$$\operatorname{curl} \boldsymbol{B} = \mu_0 \boldsymbol{J}$$

which determines the magnetic field of a steady current, in a way that was already familiar to Maxwell's contemporaries. But his second term on the right-hand side of [33] was new; it adds to $\boldsymbol{J}$ the so-called vacuum displacement current

$$\epsilon_0 \frac{\partial \boldsymbol{E}}{\partial t}$$

The name comes from an analogy with the behavior of charges in an insulating material. Here no steady current can flow, but the distribution of charges within the material is distorted by an external electric field. When the field changes, the distortion also changes, and the result appears as a current – the displacement current – which flows during the period of change. Maxwell's central insight was that the same term should be present even in empty space. The consequence was profound; it allowed him to explain the propagation of light as an electromagnetic phenomenon.

## The Source-Free Equations

In a region of empty space, away from the charges generating the electric and magnetic fields, we have $\rho = 0 = \boldsymbol{J}$, and Maxwell's equations reduce to

$$\operatorname{div} \boldsymbol{E} = 0 \quad [34]$$

$$\operatorname{div} \boldsymbol{B} = 0 \quad [35]$$

$$\operatorname{curl} \boldsymbol{B} - \frac{1}{c^2} \frac{\partial \boldsymbol{E}}{\partial t} = 0 \quad [36]$$

$$\operatorname{curl} \boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} = 0 \quad [37]$$

where $c = 1/\sqrt{\epsilon_0 \mu_0}$. By taking the curl of eqn [36] and by substituting from eqns [35] and [37], we obtain

$$\begin{aligned} 0 &= \operatorname{grad}(\operatorname{div} \boldsymbol{B}) - \nabla^2 \boldsymbol{B} - \frac{1}{c^2} \operatorname{curl}\left( \frac{\partial \boldsymbol{E}}{\partial t} \right) \\ &= -\nabla^2 \boldsymbol{B} - \frac{1}{c^2} \frac{\partial}{\partial t} (\operatorname{curl} \boldsymbol{E}) \\ &= -\nabla^2 \boldsymbol{B} + \frac{1}{c^2} \frac{\partial^2 \boldsymbol{B}}{\partial t^2} \end{aligned} \quad [38]$$

Therefore, the three components of $\boldsymbol{B}$ in empty space satisfy the (scalar) wave equation

$$\Box u = 0$$

Here $\Box$ is the d'Alembertian operator, defined by

$$\Box = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2 = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}$$

By taking the curl of eqn [37], we also obtain $\Box \boldsymbol{E} = 0$.

## Monochromatic Plane Waves

The fact that $\boldsymbol{E}$ and $\boldsymbol{B}$ are vector-valued solutions of the wave equation in empty space suggests that we look for "plane wave" solutions of Maxwell's equations in which

$$\boldsymbol{E} = \boldsymbol{\alpha} \cos \Omega + \boldsymbol{\beta} \sin \Omega \quad [39]$$

where $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are constant vectors and

$$\Omega = \frac{\omega}{c}(ct - \boldsymbol{r} \cdot \boldsymbol{e}), \quad \boldsymbol{e} \cdot \boldsymbol{e} = 1 \quad [40]$$

with $\omega > 0$, $\alpha, \beta$, and $\boldsymbol{e}$ constant; $\omega$ is the frequency and $\boldsymbol{e}$ is a unit vector that gives the direction of propagation (adding $\tau$ to $t$ and $c\tau\boldsymbol{e}$ to $\boldsymbol{r}$ leaves $u$ unchanged). This satisfies the wave equation, but for a general choice of the constants, it will not be possible to find $\boldsymbol{B}$ such that eqns [34]–[37] also hold.

By taking the divergence of eqn [39], we obtain

$$\operatorname{div} \boldsymbol{E} = \frac{\omega}{c} (\boldsymbol{e} \cdot \boldsymbol{\alpha} \sin \Omega - \boldsymbol{e} \cdot \boldsymbol{\beta} \cos \Omega) \quad [41]$$

For eqn [34] to hold, therefore, we must choose $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ orthogonal to $\boldsymbol{e}$. For eqn [37] to hold, we must find $\boldsymbol{B}$ such that

$$\operatorname{curl} \boldsymbol{E} = \frac{\omega}{c} (\boldsymbol{e} \wedge \boldsymbol{\alpha} \sin \Omega - \boldsymbol{e} \wedge \boldsymbol{\beta} \cos \Omega) = -\frac{\partial \boldsymbol{B}}{\partial t} \quad [42]$$

A possible choice is

$$\boldsymbol{B} = \frac{\boldsymbol{e} \wedge \boldsymbol{E}}{c} = \frac{1}{c} (\boldsymbol{e} \wedge \boldsymbol{\alpha} \cos \Omega + \boldsymbol{e} \wedge \boldsymbol{\beta} \sin \Omega) \quad [43]$$

and it is not hard to see that $\boldsymbol{E}$ and $\boldsymbol{B}$ then satisfy [35] and [36] as well.

The solutions obtained in this way are called "monochromatic electromagnetic plane waves."

Note that such waves are transverse in the sense that $\boldsymbol{E}$ and $\boldsymbol{B}$ are orthogonal to the direction of propagation. The definition $\boldsymbol{E}$ can be written more concisely in the form

$$\boldsymbol{E} = \mathrm{Re}\left[(\boldsymbol{\alpha} + \mathrm{i}\boldsymbol{\beta})\mathrm{e}^{-\mathrm{i}\Omega}\right] \qquad [44]$$

It is an exercise in Fourier analysis to show every solution in empty space is a combination of monochromatic plane waves. A plane wave has "plane" or "linear" polarization if $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are proportional. It has "circular" polarization if $\boldsymbol{\alpha} \cdot \boldsymbol{\alpha} = \boldsymbol{\beta} \cdot \boldsymbol{\beta}, \boldsymbol{\alpha} \cdot \boldsymbol{\beta} = 0$.

At the heart of Maxwell's theory was the idea that a light wave with definite frequency or color is represented by a monochromatic plane solution of his equations.

## Potentials

For every solution of Maxwell's equations *in vacuo*, the components of $\boldsymbol{E}$ and $\boldsymbol{B}$ satisfy the three-dimensional wave equation; but the converse is not true. That is, it is not true in general that if

$$\Box \boldsymbol{B} = 0, \qquad \Box \boldsymbol{E} = 0$$

then $\boldsymbol{E}$ and $\boldsymbol{B}$ satisfy Maxwell's equations. For this to happen, the divergence of both fields must vanish, and they must be related by [36] and [37]. These additional constraints are somewhat simpler to handle if we work not with the fields themselves, but with auxiliary quantities called "potentials."

The definition of the potentials depends on standard integrability conditions from vector calculus. Suppose that $\boldsymbol{v}$ is a vector field, which may depend on time. If $\mathrm{curl}\,\boldsymbol{v} = 0$, then there exists a function $\phi$ such that

$$\boldsymbol{v} = \mathrm{grad}\,\phi \qquad [45]$$

If $\mathrm{div}\,\boldsymbol{v} = 0$, then there exists a second vector field $\boldsymbol{a}$ such that

$$\boldsymbol{v} = \mathrm{curl}\,\boldsymbol{a} \qquad [46]$$

Neither $\phi$ nor $\boldsymbol{a}$ is uniquely determined by $\boldsymbol{v}$. In the first case, if [45] holds, then it also holds when $\phi$ is replaced by $\phi' = \phi + f$, where $f$ is a function of time alone; in the second, if [46] holds, then it also holds when $\boldsymbol{a}$ is replaced by

$$\boldsymbol{a}' = \boldsymbol{a} + \mathrm{grad}\,u$$

for any scalar function $u$ of position and time. It should be kept in mind that the existence statements are local. If $\boldsymbol{v}$ is defined on a region $U$ with nontrivial topology, then it may not be possible to find a suitable $\phi$ or $\boldsymbol{a}$ throughout the whole of $U$.

Suppose now that we are given fields $\boldsymbol{E}$ and $\boldsymbol{B}$ satisfying Maxwell's equations [26]–[29] with sources represented by the charge density $\rho$ and the current density $\boldsymbol{J}$. Since $\mathrm{div}\,\boldsymbol{B} = 0$, there exists a time-dependent vector field $\boldsymbol{A}(t, x, y, z)$ such that

$$\boldsymbol{B} = \mathrm{curl}\,\boldsymbol{A}$$

If we substitute $\boldsymbol{B} = \mathrm{curl}\,\boldsymbol{A}$ into [29] and interchange curl with the time derivative, then we obtain

$$\mathrm{curl}\left(\boldsymbol{E} + \frac{\partial \boldsymbol{A}}{\partial t}\right) = 0$$

It follows that there exists a scalar $\phi(t, x, y, z)$ such that

$$\boldsymbol{E} = -\mathrm{grad}\,\phi - \frac{\partial \boldsymbol{A}}{\partial t} \qquad [47]$$

Such a vector field $\boldsymbol{A}$ is called a "magnetic vector potential"; a function $\phi$ such that eqn [47] holds is called an "electric scalar potential."

Conversely, given scalar and vector functions $\phi$ and $\boldsymbol{A}$ of $t$, $x$, $y$, $z$, we can define $\boldsymbol{B}$ and $\boldsymbol{E}$ by

$$\boldsymbol{B} = \mathrm{curl}\,\boldsymbol{A}, \qquad \boldsymbol{E} = -\mathrm{grad}\,\phi - \frac{\partial \boldsymbol{A}}{\partial t} \qquad [48]$$

Then two of Maxwell's equations hold automatically, since

$$\mathrm{div}\,\boldsymbol{B} = 0, \qquad \mathrm{curl}\,\boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} = 0$$

The remaining pair translate into conditions on $\boldsymbol{A}$ and $\phi$. Equation [26] becomes

$$\mathrm{div}\,\boldsymbol{E} = -\nabla^2\phi - \frac{\partial}{\partial t}(\mathrm{div}\,\boldsymbol{A}) = \frac{\rho}{\epsilon_0}$$

and eqn [28] becomes

$$\begin{aligned}\mathrm{curl}\,\boldsymbol{B} - \frac{1}{c^2}\frac{\partial \boldsymbol{E}}{\partial t} &= -\nabla^2\boldsymbol{A} + \mathrm{grad}\ \mathrm{div}\,\boldsymbol{A} \\ &\quad + \frac{1}{c^2}\frac{\partial}{\partial t}\left(\mathrm{grad}\,\phi + \frac{\partial \boldsymbol{A}}{\partial t}\right) \\ &= \mu_0 \boldsymbol{J}\end{aligned}$$

If we put

$$\alpha = \frac{1}{c^2}\frac{\partial \phi}{\partial t} + \mathrm{div}\,(\boldsymbol{A})$$

then we can rewrite the equations for $\boldsymbol{A}$ and $\phi$ more simply as

$$\Box\phi - \frac{\partial \alpha}{\partial t} = \frac{\rho}{\epsilon_0}$$

$$\Box\boldsymbol{A} + \mathrm{grad}\ \alpha = \mu_0 \boldsymbol{J}$$

Here we have four equations (one scalar, one vector) in four unknowns ($\phi$ and the components of $\boldsymbol{A}$). Any set of solutions $\phi, \boldsymbol{A}$ determines a solution of Maxwell's equations via [48].

## Gauge Transformations

Given solutions $\boldsymbol{E}$ and $\boldsymbol{B}$ of Maxwell's equations, what freedom is there in the choice of $\boldsymbol{A}$ and $\phi$? First, $\boldsymbol{A}$ is determined by $\operatorname{curl} \boldsymbol{A} = \boldsymbol{B}$ up to the replacement of $\boldsymbol{A}$ by

$$\boldsymbol{A}' = \boldsymbol{A} + \operatorname{grad} u$$

for some function $u$ of position and time. The scalar potential $\phi'$ corresponding to $\boldsymbol{A}'$ must be chosen so that

$$\begin{aligned} -\operatorname{grad} \phi' &= \boldsymbol{E} + \frac{\partial \boldsymbol{A}'}{\partial t} \\ &= \boldsymbol{E} + \frac{\partial \boldsymbol{A}}{\partial t} + \operatorname{grad}\left(\frac{\partial u}{\partial t}\right) \\ &= -\operatorname{grad}\left(\phi - \frac{\partial u}{\partial t}\right) \end{aligned}$$

That is, $\phi' = \phi - \partial u/\partial t + f(t)$, where $f$ is a function of $t$ alone. We can absorb $f$ into $u$ by subtracting

$$\int f \, \mathrm{d}t$$

(this does not alter $\boldsymbol{A}'$). So the freedom in the choice of $\boldsymbol{A}$ and $\phi$ is to make the transformation

$$\boldsymbol{A} \mapsto \boldsymbol{A}' = \boldsymbol{A} + \operatorname{grad} u, \qquad \phi \mapsto \phi' = \phi - \frac{\partial u}{\partial t} \qquad [49]$$

for any $u = u(t, x, y, z)$. The transformation [49] is called a "gauge transformation."

Under [49],

$$\alpha \mapsto \alpha' = \frac{1}{c^2}\frac{\partial \phi'}{\partial t} + \operatorname{div}(\boldsymbol{A}') = \alpha - \Box u$$

It is possible to show, under certain very mild conditions on $\alpha$, that the inhomogeneous wave equation

$$\Box u = \alpha \qquad [50]$$

has a solution $u = u(t, x, y, z)$. If we choose $u$ so that [50] holds, then the transformed potentials $\boldsymbol{A}'$ and $\phi'$ satisfy

$$\operatorname{div}(\boldsymbol{A}') + \frac{1}{c^2}\frac{\partial \phi'}{\partial t} = 0$$

This is the "Lorenz gauge condition," named after L Lorenz (not the H A Lorentz of the "Lorentz contraction").

If we impose the Lorenz condition, then the only remaining freedom in the choice of $\boldsymbol{A}$ and $\phi$ is to make gauge transformations [49] in which $u$ is a solution of the wave equation $\Box u = 0$. Under the Lorenz condition, Maxwell's equations take the form

$$\Box \phi = \rho/\epsilon_0, \qquad \Box \boldsymbol{A} = \mu_0 \boldsymbol{J} \qquad [51]$$

Consistency with the Lorenz condition follows from the continuity equation on $\phi$ and $\boldsymbol{J}$.

In the absence of sources, therefore, Maxwell's equations for the potential in the Lorenz gauge reduce to

$$\Box \phi = 0, \qquad \Box \boldsymbol{A} = 0 \qquad [52]$$

together with the constraint

$$\operatorname{div} \boldsymbol{A} + \frac{1}{c^2}\frac{\partial \phi}{\partial t} = 0$$

We can, for example, choose three arbitrary solutions of the scalar wave equation for the components of the vector potential, and then define $\phi$ by

$$\phi = c^2 \int \operatorname{div} \boldsymbol{A} \mathrm{d}t$$

Whatever choice we make, we shall get a solution of Maxwell's equations, and every solution of Maxwell's equations (without sources) will arise from some such choice.

## Historical Note

At the end of the eighteenth century, four types of electromagnetic phenomena were known, but not the connections between them.

- *Magnetism*, the word derives from the Greek for "stone from Magnesia."
- *Static electricity*, produced by rubbing amber with fur; the word "electricity" derives from the Greek for "amber."
- *Light*.
- *Galvanism* or "animal electricity" – the electricity produced by batteries, discovered by Luigi Galvani.

The construction of a unified theory was a slow and painful business. It was hindered by attempts, which seem bizarre in retrospect, to understand electromagnetism in terms of underlying mechanical models involving such inventions as "electric fluids" and "magnetic vortices." We can see the legacy of this period, which ended with Einstein's work in 1905, in the misleading and archaic terms that still survive in modern terminology: "magnetic flux," "lines of force," "electric displacement," and so on.

Maxwell's contribution was decisive, although much of what we now call "Maxwell's theory" is due to his successors (Lorentz, Hertz, Einstein, and so on); and, as we shall see, a key element in Maxwell's own description of electromagnetism – the "electromagnetic ether," an all-pervasive medium which was supposed to transmit electromagnetic waves – was thrown out by Einstein.

A rough chronology is as follows.

- 1800 Volta demonstrated the connection between galvanism and static electricity.
- 1820 Oersted showed that the current from a battery generates a force on a magnet.
- 1822 Ampère suggested that light was a wave motion in a "luminiferous ether" made up of two types of electric fluid. In the same year, Galileo's "Dialogue concerning the two chief world systems" was removed from the index of prohibited books.
- 1831 Faraday showed that moving magnets can induce currents.
- 1846 Faraday suggested that light is a vibration in magnetic lines of force.
- 1863 Maxwell published the equations that describe the dynamics of electric and magnetic fields.
- 1905 Einstein's paper "On the electrodynamics of moving bodies."

## Further Reading

Chalmers AF (1975) Maxwell and the displacement current. *Physics Education* January 1975: 45–49.

Einstein A (1905) *On the Electrodynamics of Moving Bodies*. A translation of the paper can be found in *The Principle of Relativity* by Lorentz HA, Einstein A, Minkowski H, and Weyl H, with notes by Sommerfeld A. New York: Dover, 1952.

Roche J (1998) The present status of Maxwell's displacement current. *European Journal of Physics* 19: 155–166.

Siegel DM (1985) Mechanical image and reality in Maxwell's electromagnetic theory. In: Harman PM (ed.) *Wranglers and Physicists*. Manchester: Manchester University Press.

# Introductory Article: Equilibrium Statistical Mechanics

**G Gallavotti**, Università di Roma "La Sapienza," Rome, Italy



## Foundations: Atoms and Molecules

Classical statistical mechanics studies properties of macroscopic aggregates of particles, atoms, and molecules, based on the assumption that they are point masses subject to the laws of classical mechanics. Distinction between macroscopic and microscopic systems is evanescent and in fact the foundations of statistical mechanics have been laid on properties, proved or assumed, of few-particle systems.

Macroscopic systems are often considered in stationary states, which means that their microscopic configurations follow each other as time evolves while looking the same macroscopically. Observing time evolution is the same as sampling ("not too closely" time-wise) independent copies of the system prepared in the same way.

A basic distinction is necessary: a stationary state may or may not be in equilibrium. The first case arises when the particles are enclosed in a container $\Omega$ and are subject only to their mutual conservative interactions and, possibly, to external conservative forces: a typical example is a gas in a container subject to forces due to the walls of $\Omega$ and gravity, besides the internal interactions. This is a very restricted class of systems and states.

A more general case is when the system is in a stationary state but it is also subject to nonconservative forces: a typical example is a gas or fluid in which a wheel rotates, as in the Joule experiment, with some device acting to keep the temperature constant. The device is called a thermostat and in statistical mechanics it has to be modeled by forces, including nonconservative ones, which prevent an indefinite energy transfer from the external forcing to the system: such a transfer would impede the occurrence of stationary states. For instance, the thermostat could simply be a constant friction force (as in stirred incompressible liquids or as in electric wires in which current circulates because of an electromotive force).

A more fundamental approach would be to imagine that the thermostat device is not a phenomenologically introduced nonconservative force (e.g., a friction force) but is due to the interaction with an external infinite system which is in "equilibrium at infinity."

In any event nonequilibrium stationary states are intrinsically more complex than equilibrium states. Here attention will be confined to equilibrium

statistical mechanics of systems of $N$ identical point particles $\boldsymbol{Q}=(\boldsymbol{q}_1,\ldots,\boldsymbol{q}_N)$ enclosed in a cubic box $\Omega$, with volume $V$ and side $L$, normally assumed to have perfectly reflecting walls.

Particles of mass $m$ located at $\boldsymbol{q},\boldsymbol{q}'$ will be supposed to interact via a pair potential $\varphi(\boldsymbol{q}-\boldsymbol{q}')$. The microscopic motion follows the equations

$$m\ddot{\boldsymbol{q}}_i = -\sum_{j=1}^{N}\partial_{\boldsymbol{q}_i}\varphi(\boldsymbol{q}_i-\boldsymbol{q}_j)+\sum_i W_{\text{wall}}(\boldsymbol{q}_i) \stackrel{\text{def}}{=} -\partial_{\boldsymbol{q}_i}\Phi(\boldsymbol{Q}) \qquad [1]$$

where the potential $\varphi$ is assumed to be smooth except, possibly, for $|\boldsymbol{q}-\boldsymbol{q}'|\le r_0$ where it could be $+\infty$, that is, the particles cannot come closer than $r_0$, and at $r_0$ [1] is interpreted by imagining that they undergo elastic collisions; the potential $W_{\text{wall}}$ models the container and it will be replaced, unless explicitly stated, by an elastic collision rule.

The time evolution $(\boldsymbol{Q},\dot{\boldsymbol{Q}})\to S_t(\boldsymbol{Q},\dot{\boldsymbol{Q}})$ will, therefore, be described on the position – velocity space, $\widehat{\mathcal{F}}(N)$, of the $N$ particles or, more conveniently, on the phase space, i.e., by a time evolution $S_t$ on the momentum – position ($\boldsymbol{P},\boldsymbol{Q}$, with $\boldsymbol{P}=m\dot{\boldsymbol{Q}}$) space, $\mathcal{F}(N)$. The motion being conservative, the energy

$$U \stackrel{\text{def}}{=} \sum_i \frac{1}{2m}\boldsymbol{p}_i^2+\sum_{i<j}\varphi(\boldsymbol{q}_i-\boldsymbol{q}_j)+\sum_i W_{\text{wall}}(\boldsymbol{q}_i) \stackrel{\text{def}}{=} K(\boldsymbol{P})+\Phi(\boldsymbol{Q})$$

will be a constant of motion; the last term in $\Phi$ is missing if walls are perfect. This makes it convenient to regard the dynamics as associated with two dynamical systems $(\mathcal{F}(N),S_t)$ on the $6N$-dimensional phase space, and $(\mathcal{F}_U(N),S_t)$ on the $(6N-1)$-dimensional surface of energy $U$. Since the dynamics [1] is Hamiltonian on phase space, with Hamiltonian

$$H(\boldsymbol{P},\boldsymbol{Q}) \stackrel{\text{def}}{=} \sum_i \frac{1}{2m}\boldsymbol{p}_i^2+\Phi(\boldsymbol{Q}) \stackrel{\text{def}}{=} K+\Phi$$

it follows that the volume $\mathrm{d}^{3N}\boldsymbol{P}\mathrm{d}^{3N}\boldsymbol{Q}$ is conserved (i.e., a region $E$ has the same volume as $S_tE$) and also the area $\delta(H(\boldsymbol{P},\boldsymbol{Q})-U)\mathrm{d}^{3N}\boldsymbol{P}\mathrm{d}^{3N}\boldsymbol{Q}$ is conserved.

The above dynamical systems are well defined, i.e., $S_t$ is a map on phase space globally defined for all $t\in(-\infty,\infty)$, when the interaction potential is bounded below: this is implied by the *a priori* bounds due to energy conservation. For gravitational or Coulomb interactions, much more has to be said, assumed, and done in order to even define the key quantities needed for a statistical theory of motion.

Although our world is three dimensional (or at least was so believed to be until recent revolutionary theories), it will be useful to consider also systems of particles in dimension $d\neq 3$: in this case the above $6N$ and $3N$ become, respectively, $2dN$ and $dN$. Systems with dimension $d=1,2$ are in fact sometimes very good models for thin filaments or thin films. For the same reason, it is often useful to imagine that space is discrete and particles can only be located on a lattice, for example, on $\mathbb{Z}^d$ (see the section "Lattice models").

The reader is referred to Gallavotti (1999) for more details.

## Pressure, Temperature, and Kinetic Energy

The beginning was BERNOULLI's derivation of the perfect gas law via the identification of the *pressure* at *numerical density* $\rho$ with the average momentum transferred per unit time to a surface element of area $\mathrm{d}S$ on the walls: that is, the average of the observable $2mv\rho v\,\mathrm{d}S$, with $v$ the normal component of the velocity of the particles that undergo collisions with $\mathrm{d}S$. If $f(v)\mathrm{d}v$ is the distribution of the normal component of velocity and $f(\boldsymbol{v})\mathrm{d}^3\boldsymbol{v}\equiv\prod_i f(v_i)\mathrm{d}^3\boldsymbol{v}$, $\boldsymbol{v}=(v_1,v_2,v_3)$, is the total velocity distribution, the average of the momentum transferred is $p\mathrm{d}S$ given by

$$\mathrm{d}S\int_{v>0}2mv^2\rho f(v)\mathrm{d}v=\mathrm{d}S\int mv^2\rho f(v)\mathrm{d}v = \rho\frac{2}{3}\mathrm{d}S\int\frac{m}{2}\boldsymbol{v}^2 f(\boldsymbol{v})\mathrm{d}^3\boldsymbol{v}=\rho\frac{2}{3}\left\langle\frac{K}{N}\right\rangle\mathrm{d}S \qquad [2]$$

Furthermore $(2/3)\langle K/N\rangle$ was identified as proportional to the absolute temperature $\langle K/N\rangle \stackrel{\text{def}}{=}$ const $(3/2)T$ which, with present-day notations, is written as $(2/3)\langle K/N\rangle=k_{\text{B}}T$. The constant $k_{\text{B}}$ was (later) called Boltzmann's constant and it is the same for at least all perfect gases. Its independence on the particular nature of the gas is a consequence of *Avogadro's law* stating that equal volumes of gases at the same conditions of temperature and pressure contain equal number of molecules.

Proportionality between average kinetic energy and temperature via the universal constant $k_{\text{B}}$ became in fact a fundamental assumption extending to all aggregates of particles gaseous or not, never challenged in all later works (until quantum mechanics, where this is no longer true, see the section "Quantum statistics".

For more details, we refer the reader to Gallavotti (1999).

## Heat and Entropy

After Clausius' discovery of entropy, Boltzmann, in order to explain it mechanically, introduced the *heat theorem*, which he developed to full generality between 1866 and 1884. Together with the mentioned identification of absolute temperature with average kinetic energy, the heat theorem can also be considered a founding element of statistical mechanics.

The theorem makes precise the notion of time average and then states in great generality that given any mechanical system one can associate with its dynamics four quantities $U, V, p, T$, defined as time averages of suitable mechanical observables (i.e., functions on phase space), so that when the external conditions are infinitesimally varied and the quantities $U, V$ change by $\mathrm{d}U, \mathrm{d}V$, respectively, the ratio $(\mathrm{d}U + p\mathrm{d}V)/T$ is exact, i.e., there is a function $S(U, V)$ whose corresponding variation equals the ratio. It will be better, for the purpose of considering very large boxes ($V \to \infty$) to write this relation in terms of intensive quantities $u \overset{\text{def}}{=} U/N$ and $v = V/N$ as

$$\frac{\mathrm{d}u + p\mathrm{d}v}{T} \text{ is exact} \qquad [3]$$

i.e., the ratio equals the variation $\mathrm{d}s$ of $s(U/N, V/N) \equiv (1/N)S(U, V)$.

The proof originally dealt with *monocyclic* systems, i.e., systems in which all motions are periodic. The assumption is clearly much too restrictive and justification for it developed from the early "nonperiodic motions can be regarded as periodic with infinite period" (1866), to the later *ergodic hypothesis* and finally to the realization that, after all, the heat theorem does not really depend on the ergodic hypothesis (1884).

Although for a one-dimensional system the proof of the heat theorem is a simple check, it was a real breakthrough because it led to an answer to the general question as to under which conditions one could define mechanical quantities whose variations were constrained to satisfy [3] and therefore could be interpreted as a mechanical model of Clausius' macroscopic thermodynamics. It is reproduced in the following.

Consider a one-dimensional system subject to forces with a confining potential $\varphi(x)$ such that $|\varphi'(x)| > 0$ for $|x| > 0, \varphi''(0) > 0$ and $\varphi(x) \xrightarrow[x\to\infty]{} +\infty$. All motions are periodic, so that the system is monocyclic. Suppose that the potential $\varphi(x)$ depends on a parameter $V$ and define a *state* to be a motion with given energy $U$ and given $V$; let

$$\begin{aligned} &U = \text{total energy of the system} \equiv K + \Phi \\ &T = \text{time average of the kinetic energy } K = \langle K \rangle \\ &V = \text{the parameter on which } \varphi \text{ is supposed to depend} \\ &p = -\text{time average of } \partial_{V\varphi}, -\langle \partial_{V\varphi} \rangle \end{aligned} \qquad [4]$$

A state is thus parametrized by $U$, $V$. If such parameters change by $\mathrm{d}U, \mathrm{d}V$, respectively, and if $\mathrm{d}L \overset{\text{def}}{=} -p\mathrm{d}V, \mathrm{d}Q \overset{\text{def}}{=} \mathrm{d}U + p\mathrm{d}V$, then [3] holds. In fact, let $x_\pm(U, V)$ be the extremes of the oscillations of the motion with given $U$, $V$ and define $S$ as

$$S = 2 \log \int_{x_-(U,V)}^{x_+(U,V)} \sqrt{(U - \varphi(x))}\mathrm{d}x$$
$$\Rightarrow \mathrm{d}S = \frac{\int (\mathrm{d}U - \partial_{V\varphi}(x)\mathrm{d}V)(\mathrm{d}x/\sqrt{K})}{\int (\mathrm{d}x/\sqrt{K})K} \qquad [5]$$

Noting that $\mathrm{d}x/\sqrt{K} = \sqrt{2/m}\,\mathrm{d}t$, [3] follows because time averages are given by integrating with respect to $\mathrm{d}x/\sqrt{K}$ and dividing by the integral of $1/\sqrt{K}$.

For more details, the reader is referred to Boltzmann (1968b) and Gallavotti (1999).

## Heat Theorem and Ergodic Hypothesis

Boltzmann tried to extend the result beyond the one-dimensional systems (e.g., to Keplerian motions, which are not monocyclic unless only motions with a fixed eccentricity are considered). However, the early statement that "aperiodic motions can be regarded as periodic with infinite period" is really the heart of the application of the heat theorem for monocyclic systems to the far more complex gas in a box.

Imagine that the gas container $\Omega$ is closed by a piston of section $A$ located to the right of the origin at distance $L$ and acting as a lid, so that the volume is $V = AL$. The microscopic model for the piston will be a potential $\overline{\varphi}(L - \xi)$ if $x = (\xi, \eta, \zeta)$ are the coordinates of a particle. The function $\overline{\varphi}(r)$ will vanish for $r > r_0$, for some $r_0 \ll L$, and diverge to $+\infty$ at $r = 0$. Thus, $r_0$ is the width of the layer near the piston where the force of the wall is felt by the particles that happen to be roaming there.

The contribution to the total potential energy $\Phi$ due to the walls is $W_{\text{wall}} = \sum_j \overline{\varphi}(L - \xi_j)$ and $\partial_V \overline{\varphi} = A^{-1} \partial_L \overline{\varphi}$; assuming monocyclicity, it is necessary to evaluate the time average of $\partial_L \Phi(x) = \partial_L W_{\text{wall}} \equiv -\sum_j \overline{\varphi}'(L - \xi_j)$. As time evolves, the particles $x_j$ with $\xi_j$ in the layer within $r_0$ of the wall will feel the force exercised by the wall and

bounce back. One particle in the layer will contribute to the average of $\partial_L \Phi(x)$ the amount

$$\frac{1}{\text{total time}} 2 \int_{t_0}^{t_1} -\overline{\varphi}'(L - \xi_j) \mathrm{d}t \qquad [6]$$

if $t_0$ is the first instant when the point $j$ enters the layer and $t_1$ is the instant when the $\xi$-component of the velocity vanishes "against the wall." Since $-\overline{\varphi}'(L - \xi_j)$ is the $\xi$-component of the force, the integral is $2m|\dot{\xi}_j|$ (by Newton's law), provided, of course, $\dot{\xi}_j > 0$.

Suppose that no collisions between particles occur while the particles travel within the range of the potential of the wall, i.e., the mean free path is much greater than the range of the potential $\overline{\varphi}$ defining the wall. The contribution of collisions to the average momentum transfer to the wall per unit time is therefore given by, see [2],

$$\int_{v>0} 2mv\, f(v) \rho_{\text{wall}} A v \, \mathrm{d}v$$

if $\rho_{\text{wall}}, f(v)$ are the average density near the wall and, respectively, the average fraction of particles with a velocity component normal to the wall between $v$ and $v + \mathrm{d}v$. Here $p$, $f$ are supposed to be independent of the point on the wall: this should be true up to corrections of size $o(A)$.

Thus, writing the average kinetic energy per particle and per velocity component, $\int (m/2) v^2 f(v) \mathrm{d}v$, as $(1/2)\beta^{-1}$ (cf. [2]) it follows that

$$p \stackrel{\text{def}}{=} - \langle \partial_V \Phi \rangle = \rho_{\text{wall}} \beta^{-1} \qquad [7]$$

has the physical interpretation of pressure. $(1/2)\beta^{-1}$ is the average kinetic energy per degree of freedom: hence, it is proportional to the absolute temperature $T$ (cf. see the section "Pressure, temperature, and kinetic energy").

On the other hand, if motion on the energy surface takes place on a single periodic orbit, the quantity $p$ in [7] is the right quantity that would make the heat theorem work; see [4]. Hence, regarding the trajectory on each energy surface as periodic (i.e., the system as monocyclic) leads to the heat theorem with $p, U, V, T$ having the right physical interpretation corresponding to their appellations. This shows that monocyclic systems provide natural models of thermodynamic behavior.

Assuming that a chaotic system like a gas in a container of volume $V$ will satisfy, for practical purposes, the above property, a quantity $p$ can be defined such that $\mathrm{d}U + p\mathrm{d}V$ admits the inverse of the average kinetic energy $\langle K \rangle$ as an integrating factor and, furthermore, $p, U, V, \langle K \rangle$ have the physical interpretations of pressure, energy, volume, and (up to a proportionality factor) absolute temperature, respectively.

Boltzmann's conception of space (and time) as discrete allowed him to conceive the property that the energy surface is constituted by "points" all of which belong to a single trajectory: a property that would be impossible if the phase space was really a continuum. Regarding phase space as consisting of a finite number of "cells" of finite volume $h^{dN}$, for some $h > 0$ (rather than of a continuum of points), allowed him to think, without logical contradiction, that the energy surface consisted of a single trajectory and, hence, that motion was a cyclic permutation of its points (actually cells).

Furthermore, it implied that the time average of an observable $F(\boldsymbol{P}, \boldsymbol{Q})$ had to be identified with its average on the energy surface computed via the Liouville distribution

$$C^{-1} \int F(\boldsymbol{P}, \boldsymbol{Q}) \delta(H(\boldsymbol{P}, \boldsymbol{Q}) - U) \mathrm{d}\boldsymbol{P}\, \mathrm{d}\boldsymbol{Q}$$

with

$$C = \int \delta(H(\boldsymbol{P}, \boldsymbol{Q}) - U) \mathrm{d}\boldsymbol{P}\, \mathrm{d}\boldsymbol{Q}$$

(the appropriate normalization factor): a property that was written symbolically

$$\frac{\mathrm{d}t}{T} = \frac{\mathrm{d}\boldsymbol{P}\, \mathrm{d}\boldsymbol{Q}}{\int \mathrm{d}\boldsymbol{P}\, \mathrm{d}\boldsymbol{Q}}$$

or

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T F(S_t(\boldsymbol{P}, \boldsymbol{Q})) \mathrm{d}t = \frac{\int F(\boldsymbol{P}', \boldsymbol{Q}') \delta(H(\boldsymbol{P}', \boldsymbol{Q}') - U)\, \mathrm{d}\boldsymbol{P}' \mathrm{d}\boldsymbol{Q}'}{\int \delta(H(\boldsymbol{P}', \boldsymbol{Q}') - U)\, \mathrm{d}\boldsymbol{P}' \mathrm{d}\boldsymbol{Q}'} \qquad [8]$$

The validity of [8] for all (piecewise smooth) observables $F$ and for all points of the energy surface, with the exception of a set of zero area, is called the ergodic hypothesis.

For more details, the reader is referred to Boltzmann (1968) and Gallavotti (1999).

## Ensembles

Eventually Boltzmann in 1884 realized that the validity of the heat theorem for averages computed via the right-hand side (rhs) of [8] held independently of the ergodic hypothesis, that is, [8] was not necessary because the heat theorem (i.e., [3]) could also be derived under the only assumption that the averages involved in its formulation were computed

as averages over phase space with respect to the probability distribution on the rhs of [8].

Furthermore, if $T$ was identified with the average kinetic energy, $U$ with the average energy, and $p$ with the average force per unit surface on the walls of the container $\Omega$ with volume $V$, the relation [3] held for a variety of families of probability distributions on phase space, besides [8]. Among these are:

1. The "microcanonical ensemble," which is the collection of probability distributions on the rhs of [8] parametrized by $u = U/N, v = V/N$ (energy and volume per particle),

$$\mu^{\mathrm{mc}}_{u,v}(\mathrm{d}\boldsymbol{P}\,\mathrm{d}\boldsymbol{Q}) = \frac{1}{Z_{\mathrm{mc}}(U,N,V)}\delta(H(\boldsymbol{P},\boldsymbol{Q}) - U)\frac{\mathrm{d}\boldsymbol{P}\,\mathrm{d}\boldsymbol{Q}}{N!h^{dN}} \qquad [9]$$

   where $h$ is a constant with the dimensions of an action which, in the discrete representation of phase space mentioned in the previous section, can be taken such that $h^{dN}$ equals the volume of the cells and, therefore, the integrals with respect to [9] can be interpreted as an (approximate) sum over the cells conceived as microscopic configurations of $N$ indistinguishable particles (whence the $N!$).

2. The "canonical ensemble," which is the collection of probability distributions parametrized by $\beta, v = V/N$,

$$\mu^{\mathrm{c}}_{\beta,v}(\mathrm{d}\boldsymbol{P}\mathrm{d}\boldsymbol{Q}) = \frac{1}{Z_{\mathrm{c}}(\beta,N,V)}\mathrm{e}^{-\beta H(\boldsymbol{P},\boldsymbol{Q})}\frac{\mathrm{d}\boldsymbol{P}\mathrm{d}\boldsymbol{Q}}{N!h^{dN}} \qquad [10]$$

   to which more ensembles can be added, such as the grand canonical ensemble (Gibbs).

3. The "grand canonical ensemble" which is the collection of probability distributions parameterized by $\beta, \lambda$ and defined over the space $\mathcal{F}_{\mathrm{gc}} = \cup_{N=0}^{\infty}\mathcal{F}(N)$,

$$\mu^{\mathrm{gc}}_{\beta,\lambda}(\mathrm{d}\boldsymbol{P}\mathrm{d}\boldsymbol{Q}) = \frac{1}{Z_{\mathrm{gc}}(\beta,\lambda,V)}\mathrm{e}^{\beta\lambda N - \beta H(\boldsymbol{P},\boldsymbol{Q})}\frac{\mathrm{d}\boldsymbol{P}\mathrm{d}\boldsymbol{Q}}{N!h^{dN}} \qquad [11]$$

Hence, there are several different models of thermodynamics. The key tests for accepting them as real microscopic descriptions of macroscopic thermodynamics are as follows.

1. A correspondence between the macroscopic states of thermodynamic equilibrium and the elements of a collection of probability distributions on phase space can be established by identifying, on the one hand, macroscopic thermodynamic states with given values of the thermodynamic functions and, on the other, probability distributions attributing the same average values to the corresponding microscopic observables (i.e., whose averages have the interpretation of thermodynamic functions).
2. Once the correct correspondence between the elements of the different ensembles is established, that is, once the pairs $(u,v), (\beta,v), (\beta,\lambda)$ are so related to produce the same values for the averages $U, V, k_{\mathrm{B}}T \stackrel{\mathrm{def}}{=} \beta^{-1}, p|\partial\Omega|$ of

$$H(\boldsymbol{P},\boldsymbol{Q}), V, \frac{2K(\boldsymbol{P})}{3N}, \int \delta_{\partial\Omega}(\boldsymbol{q}_1)2m(\boldsymbol{v}_1\cdot\boldsymbol{n})^2\,\mathrm{d}\boldsymbol{q}_1 \qquad [12]$$

where ($\delta_{\partial\Omega}(\boldsymbol{q}_1)$ is a delta-function pinning $\boldsymbol{q}_1$ to the surface $\partial\Omega$), then the averages of all physically interesting observables *should coincide* at least in the thermodynamic limit, $\Omega \to \infty$. In this way, the elements $\mu$ of the considered collection of probability distributions can be identified with the states of macroscopic equilibrium of the system. The $\mu$'s depend on parameters and therefore they form an ensemble: each of them corresponds to a macroscopic equilibrium state whose thermodynamic functions are appropriate averages of microscopic observables and therefore are functions of the parameters identifying $\mu$.

**Remark** The word "ensemble" is often used to indicate the individual probability distributions of what has been called here an ensemble. The meaning used here seems closer to the original sense in the 1884 paper of Boltzmann (in other words, often by "ensemble" one means that collection of the phase space points on which a given probability distribution is considered, and this does not seem to be the original sense).

For instance, in the case of the microcanonical distributions this means interpreting energy, volume, temperature, and pressure of the equilibrium state with specific energy $u$ and specific volume $v$ as proportional, through appropriate universal proportionality constants, to the integrals with respect to $\mu^{\mathrm{mc}}_{u,v}(\mathrm{d}\boldsymbol{P}\,\mathrm{d}\boldsymbol{Q})$ of the mechanical quantities in [12]. The averages of other thermodynamic observables in the state with specific energy $u$ and specific volume $v$ should be given by their integrals with respect to $\mu^{\mathrm{mc}}_{u,v}$.

Likewise, one can interpret energy, volume, temperature, and pressure of the equilibrium state with specific energy $u$ and specific volume $v$ as the averages of the mechanical quantities [12] with respect to the canonical distribution $\mu^{\mathrm{c}}_{\beta,v}(\mathrm{d}\boldsymbol{P}\,\mathrm{d}\boldsymbol{Q})$ which has average specific energy precisely $u$. The averages of other thermodynamic observables in the state with specific energy and volume $u$ and $v$ are

given by their integrals with respect to $\mu^{c}_{\beta,v}$. A similar definition can be given for the description of thermodynamic equilibria via the grand canonical distributions.

For more details, see Gibbs (1981) and Gallavotti (1999).

## Equivalence of Ensembles

BOLTZMANN proved that, computing averages via the microcanonical or canonical distributions, the essential property [3] was satisfied when changes in their parameters (i.e., $u, v$ or $\beta, v$, respectively) induced changes $\mathrm{d}u$ and $\mathrm{d}v$ on energy and volume, respectively. He also proved that the function $s$, whose existence is implied by [3], was the same function once expressed as a function of $u, v$ (or of any pair of thermodynamic parameters, e.g., of $T, v$ or $p, u$). A close examination of Boltzmann's proof shows that the [3] holds exactly in the canonical ensemble and up to corrections tending to 0 as $\Omega \to \infty$ in the microcanonical ensemble. Identity of thermodynamic functions evaluated in the two ensembles holds, as a consequence, up to corrections of this order. In addition, Gibbs added that the same held for the grand canonical ensemble.

Of course, not every collection of stationary probability distributions on phase space would provide a model for thermodynamics: Boltzmann called "orthodic" the collections of stationary distributions which generated models of thermodynamics through the above-mentioned identification of its elements with macroscopic equilibrium states. The microcanonical, canonical, and the later grand canonical ensembles are the chief examples of orthodic ensembles. Boltzmann and Gibbs proved these ensembles to be not only orthodic but to generate the same thermodynamic functions, that is to generate the same thermodynamics.

This meant freedom from the analysis of the truth of the doubtful ergodic hypothesis (still unproved in any generality) or of the monocyclicity (manifestly false if understood literally rather than regarding the phase space as consisting of finitely many small, discrete cells), and allowed Gibbs to formulate the problem of statistical mechanics of equilibrium as follows.

**Problem** Study the properties of the collection of probability distributions constituting (any) one of the above ensembles.

However, by no means the three ensembles just introduced exhaust the class of orthodic ensembles producing the same models of thermodynamics in the limit of infinitely large systems. The wealth of ensembles with the orthodicity property, hence leading to equivalent mechanical models of thermodynamics, can be naturally interpreted in connection with the phenomenon of phase transition (see the section "Phase transitions and boundary conditions").

Clearly, the quoted results do not "prove" that thermodynamic equilibria "are" described by the microcanonical, canonical, or grand canonical ensembles. However, they certainly show that, for most systems, independently of the number of degrees of freedom, one can define quite unambiguously a mechanical model of thermodynamics establishing parameter-free, system-independent, physically important relations between thermodynamic quantities (e.g., $\partial_u(p(u,v)/T(u,v)) \equiv \partial_v(1/T(u,v))$, from [3]).

The ergodic hypothesis which was at the root of the mechanical theorems on heat and entropy cannot be taken as a justification of their validity. Naively one would expect that the time scale necessary to see an equilibrium attained, called recurrence time scale, would have to be at least the time that a phase space point takes to visit all possible microscopic states of given energy: hence, an explanation of why the necessarily enormous size of the recurrence time is not a problem becomes necessary.

In fact, the recurrence time can be estimated once the phase space is regarded as discrete: for the purpose of countering mounting criticism, Boltzmann assumed that momentum was discretized in units of $(2mk_{\mathrm{B}}T)^{1/2}$ (i.e., the average momentum size) and space was discretized in units of $\rho^{-1/3}$ (i.e., the average spacing), implying a volume of cells $h^{3N}$ with $h \stackrel{\mathrm{def}}{=} \rho^{-1/3}(2mk_{\mathrm{B}}T)^{1/2}$; then he calculated that, even with such a gross discretization, a cell representing a microscopic state of $1\,\mathrm{cm}^3$ of hydrogen at normal condition would require a time (called "recurrence time") of the order of $\sim 10^{10^{19}}$ times the age of the Universe (!) to visit the entire energy surface. In fact, the phase space volume is $\Gamma = (\rho^{-3}N(2mk_{\mathrm{B}}T)^{3/2})^N \equiv h^{3N}$ and the number of cells of volume $h^{3N}$ is $\Gamma/(N!h^{3N}) \simeq \mathrm{e}^{3N}$; and the time to visit all will be $\mathrm{e}^{3N}\tau_0$, with $\tau_0$ a typical atomic unit, e.g., $10^{-12}\,\mathrm{s}$ – but $N = 10^{19}$. In this sense, the statement boldly made by young Boltzmann that "aperiodic motions can be regarded as periodic with infinite period" was even made quantitative.

The recurrence time is clearly so long to be irrelevant for all purposes: nevertheless, the correctness of the microscopic theory of thermodynamics can still rely on the microscopic dynamics once it is understood (as stressed by Boltzmann) that the reason why we observe approach to equilibrium, and equilibrium itself, over "human" timescales

(which are far shorter than the recurrence times) is due to the property that on most of the energy surface the (very few) observables whose averages yield macroscopic thermodynamic functions (namely pressure, temperature, energy, ...) *assume the same value* even if $N$ is only very moderately large (of the order of $10^3$ rather than $10^{19}$). This implies that this value coincides with the average and therefore satisfies the heat theorem without any contradiction with the length of the recurrence time. The latter rather concerns the time needed to the *generic observable* to thermalize, that is, to reach its time average: the generic observable will indeed take a very long time to "thermalize" but no one will ever notice, because the generic observable (e.g., the position of a pre-identified particle) is not relevant for thermodynamics.

The word "proof" is not used in the mathematical sense so far in this article: the relevance of a mathematically rigorous analysis was widely realized only around the 1960s at the same time when the first numerical studies of the thermodynamic functions became possible and rigorous results were needed to check the correctness of various numerical simulations.

For more details, the reader is referred to Boltzmann (1968a, b) and Gallavotti (1999).

## Thermodynamic Limit

Adopting Gibbs axiomatic point of view, it is interesting to see the path to be followed to achieve an equivalence proof of three ensembles introduced in the section "Heat theorem and ergodic hypothesis."

A preliminary step is to consider, given a cubic box $\Omega$ of volume $V = L^d$, the normalization factors $Z^{gc}(\beta, \lambda, V), Z^{c}(\beta, N, V)$, and $Z^{mc}(U, N, V)$ in [9], [10], and [11], respectively, and to check that the following thermodynamic limits exist:

$$\begin{aligned} \beta p_{gc}(\beta,\lambda) &\stackrel{\text{def}}{=} \lim_{V\to\infty} \frac{1}{V} \log Z^{gc}(\beta,\lambda,V) \\ -\beta f_c(\beta,\rho) &\stackrel{\text{def}}{=} \lim_{V\to\infty, \frac{N}{V}=\rho} \frac{1}{N} \log Z^{c}(\beta,N,V) \\ k_B^{-1} s_{mc}(u,\rho) & \\ &\stackrel{\text{def}}{=} \lim_{V\to\infty, N/V=\rho,\, U/N=u} \frac{1}{N} \log Z^{mc}(U,N,V) \end{aligned} \qquad [13]$$

where the density $\rho \stackrel{\text{def}}{=} v^{-1} \equiv N/V$ is used, instead of $v$, for later reference. The normalization factors play an important role because they have simple thermodynamic interpretation (see the next section): they are called grand canonical, canonical, and microcanonical partition functions, respectively.

Not surprisingly, assumptions on the interparticle potential $\varphi(q - q')$ are necessary to achieve an existence proof of the limits in [13]. The assumptions on $\varphi$ are not only quite general but also have a clear physical meaning. They are

1. *stability*: that is, existence of a constant $B \ge 0$ such that $\sum_{i<j}^{N} \varphi(q_i - q_j) \ge -BN$ for all $N \ge 0$, $q_1, \dots, q_N \in \mathbb{R}^d$, and
2. *temperedness*: that is, existence of constants $\varepsilon_0$, $R > 0$ such that $|\varphi(q - q')| < B|q - q'|^{-d-\varepsilon_0}$ for $|q - q'| > R$.

The assumptions are satisfied by essentially all microscopic interactions with the notable exceptions of the gravitational and Coulombic interactions, which require a separate treatment (and lead to somewhat different results on the thermodynamic behavior).

For instance, assumptions (1), (2) are satisfied if $\varphi(q)$ is $+\infty$ for $|q| < r_0$ and smooth for $|q| > r_0$, for some $r_0 \ge 0$, and furthermore $\varphi(q) > B_0|q|^{-(d+\varepsilon_0)}$ if $r_0 < |q| \le R$, while for $|q| > R$ it is $|\varphi(q)| < B_1|q|^{-(d+\varepsilon_0)}$, for some $B_0, B_1, \varepsilon_0 > 0, R > r_0$. Briefly, $\varphi$ is fast diverging at contact and fast approaching 0 at large distance. This is called a (generalized) Lennard–Jones potential. If $r_0 > 0$, $\varphi$ is called a hard-core potential. If $B_1 = 0$, the potential is said to have finite range. (See Appendix 1 for physical implications of violations of the above stability and temperedness properties.) However, in the following, it will be necessary, both for simplicity and to contain the length of the exposition, to restrict consideration to the case $B_1 = 0$, i.e., to

$$\begin{aligned} \varphi(q) &> B_0|q|^{-(d+\varepsilon_0)}, \quad r_0 < |q| \le R, \\ |\varphi(q)| &\equiv 0, \quad |q| > R \end{aligned} \qquad [14]$$

unless explicitly stated.

Assuming stability and temperedness, the existence of the limits in [13] can be mathematically proved: in Appendix 2, the proof of the first is analyzed to provide the simplest example of the technique. A remarkable property of the functions $\beta p_{gc}(\beta, \lambda)$, $-\beta\rho f_c(\beta, \rho)$, and $\rho s_{mc}(u, \rho)$ is that they are convex functions: hence, they are continuous in the interior of their domains of definition and, at one variable fixed, are differentiable with respect to the other with at most countably many exceptions.

In the case of a potential without hard core ($\rho_{max} = \infty$), $-\rho f_c(\beta, \rho)$ can be checked to tend to 0 slower than $\rho$ as $\rho \to 0$, and to $-\infty$ faster than $-\rho$ as $\rho \to \infty$ (essentially proportionally to $-\rho \log \rho$ in both cases). Likewise, in the same case, $s_{mc}(u, \rho)$ can be shown to tend to 0 slower than $u - u_{min}$ as $u \to u_{min}$, and to $-\infty$ faster than $-u$ as $u \to \infty$. The latter

asymptotic properties can be exploited to derive, from the relations between the partition functions in [13],

$$Z^{gc}(\beta,\lambda,V)=\sum_{N=0}^{\infty} e^{\beta\lambda N}Z^{c}(\beta,N,V)$$
$$Z^{c}(\beta,N,V)=\int_{-B}^{\infty} e^{-\beta U}Z^{mc}(U,N,V)\,dU \qquad [15]$$

and, from the above-mentioned convexity, the consequences

$$\beta p_{mc}(\beta,\lambda)=\max_{v}(\beta\lambda v^{-1}-\beta v^{-1}f_c(\beta,v^{-1}))$$
$$-\beta f_c(\beta,v^{-1})=\max_{u}(-\beta u+k_B^{-1}s_{mc}(u,v^{-1})) \qquad [16]$$

and that the maxima are attained in points, or intervals, internal to the intervals of definition. Let $v_{gc}, u_c$ be points where the maxima are, respectively, attained in [16].

Note that the quantity $e^{\beta\lambda N}Z^c(\beta,N,V)/Z^{gc}(\beta,\lambda,V)$ has the interpretation of probability of a density $v^{-1}=N/V$ evaluated in the grand canonical distribution. It follows that, if the maximum in the first of [16] is strict, that is, it is reached at a single point, the values of $v^{-1}$ in closed intervals not containing the maximum point $v_{gc}^{-1}$ have a probability behaving as $<e^{-cV}$, $c>0$, as $V\to\infty$, compared to the probability of $v^{-1}$'s in any interval containing $v_{gc}^{-1}$. Hence, $v_{gc}$ has the interpretation of average value of $v$ in the grand canonical distribution, in the limit $V\to\infty$.

Likewise, the interpretation of

$$e^{-\beta uN}Z^{mc}(uN,\ N,\ V)/Z^c(\beta,\ N,\ V)$$

as probability in the canonical distribution of an energy density $u$ shows that, if the maximum in the second of [16] is strict, the values of $u$ in closed intervals not containing the maximum point $u_c$ have a probability behaving as $<e^{-cV}, c>0$, as $V\to\infty$, compared to the probability of $u$'s in any interval containing $u_c$. Hence, in the limit $\Omega\to\infty$, the average value of $u$ in the canonical distribution is $u_c$.

If the maxima are strict, [16] also establishes a relation between the grand canonical density, the canonical free energy and the grand canonical parameter $\lambda$, or between the canonical energy, the microcanonical entropy, and the canonical parameter $\beta$:

$$\lambda=\partial_{v^{-1}}(v_{gc}^{-1}f_c(\beta,v_{gc}^{-1})),\qquad k_B\beta=\partial_u s_{mc}(u_c,v^{-1}) \qquad [17]$$

where convexity and strictness of the maxima imply the derivatives existence.

**Remark** Therefore, in the equivalence between canonical and microcanonical ensembles, the canonical distribution with parameters $(\beta,v)$ should correspond with the microcanonical with parameters $(u_c,v)$. The grand canonical distribution with parameters $(\beta,\lambda)$ should correspond with the canonical with parameters $(\beta,v_{gc})$.

For more details, the reader is referred to Ruelle (1969) and Gallavotti (1999).

## Physical Interpretation of Thermodynamic Functions

The existence of the limits [13] implies several properties of interest. The first is the possibility of finding the physical meaning of the functions $p_{gc}, f_c, s_{mc}$ and of the parameters $\lambda,\beta$

Note first that, for all $V$ the grand canonical average $\langle K\rangle_{\beta,\lambda}$ is $(d/2)\beta^{-1}\langle N\rangle_{\beta,\lambda}$ so that $\beta^{-1}$ is proportional to the temperature $T_{gc}=T(\beta,\lambda)$ in the grand canonical distribution: $\beta^{-1}=k_BT(\beta,\lambda)$. Proceeding heuristically, the physical meaning of $p(\beta,\lambda)$ and $\lambda$ can be found through the following remarks.

Consider the microcanonical distribution $\mu_{u,v}^{mc}$ and denote by $\int^*$ the integral over $(\boldsymbol{P},\boldsymbol{Q})$ extended to the domain of the $(\boldsymbol{P},\boldsymbol{Q})$ such that $H(\boldsymbol{P},\boldsymbol{Q})=U$ and, at the same time, $q_1\in dV$, where $dV$ is an infinitesimal volume surrounding the region $\Omega$. Then, by the microscopic definition of the pressure $p$ (see the introductory section), it is

$$p\,dV=\frac{N}{Z(U,N,V)}\int^*\delta\frac{2}{3}\frac{p_1^2}{2m}\frac{d\boldsymbol{P}\,d\boldsymbol{Q}}{N!h^{dN}}$$
$$\equiv\frac{2}{3Z(U,N,V)}\int^*\delta K(\boldsymbol{P})\frac{d\boldsymbol{P}\,d\boldsymbol{Q}}{N!h^{dN}} \qquad [18]$$

where $\delta\equiv\delta(H(\boldsymbol{P},\boldsymbol{Q})-U)$. The RHS of [18] can be compared with

$$\frac{\partial_V Z(U,N,V)dV}{Z(U,N,V)}=\frac{N}{Z(U,N,V)}\int^*\frac{d\boldsymbol{P}\,d\boldsymbol{Q}}{N!h^{dN}}$$

to give

$$\frac{\partial_V Z\,dV}{Z}=N\frac{p\,dV}{(2/3)\langle K\rangle^*}=\beta p\,dV$$

because $\langle K\rangle^*$, which denotes the average $\int^*K/\int^*1$, should be essentially the same as the microcanonical average $\langle K\rangle_{mc}$ (i.e., insensitive to the fact that one particle is constrained to the volume $dV$) if $N$ is large. In the limit $V\to\infty, V/N=v$, the latter remark together with the second of [17] yields

$$k_B^{-1}\partial_v s_{mc}(u,v^{-1})=\beta p(u,v),$$
$$k_B^{-1}\partial_u s_{mc}(u,v)=\beta \qquad [19]$$

respectively. Note that $p\geq 0$ and it is not increasing in $v$ because $s_{mc}(\rho)$ is concave as a function of $v=\rho^{-1}$ (in fact, by the remark following [14] $\rho s_{mc}(u,\rho)$ is convex in $\rho$ and, in general, if $\rho g(\rho)$ is convex in $\rho$ then $g(v^{-1})$ is always concave in $v=\rho^{-1}$).

Hence, $ds_{mc}(u,v) = (du + p dv)/T$, so that taking into account the physical meaning of $p$, $T$ (as pressure and temperature, see the section "Pressure, temperature, and kinetic energy"), $s_{mc}$ is, in thermodynamics, the entropy. Therefore (see the second of [16]), $-\beta f_c(\beta,\rho) = -\beta u_c + k_B^{-1} s_{mc}(u_c,\rho)$ becomes

$$f_c(\beta,\rho) = u_c - T_c s_{mc}(u_c,\rho), \qquad df_c = -p\,dv - s_{mc}\,dT \tag{20}$$

and since $u_c$ has the interpretation (as mentioned in the last section) of average energy in the canonical distribution $\mu^c_{\beta,v}$ it follows that $f_c$ has the thermodynamic interpretation of free energy (once compared with the definition of free energy, $F = U - TS$, in thermodynamics).

By [17] and [20],

$$\lambda = \partial_{v^{-1}}(v_{gc}^{-1} f_c(\beta, v_{gc}^{-1})) \equiv u_c - T_c s_{mc} + p v_{gc}$$

and $v_{gc}$ has the meaning of specific volume $v$. Hence, after comparison with the definition of chemical potential, $\lambda V = U - TS + pV$, in thermodynamics, it follows that the thermodynamic interpretation of $\lambda$ is the chemical potential and (see [16], [17]), the grand canonical relation

$$\beta p_{gc}(\beta,\lambda) = \beta\lambda v_{gc}^{-1} - \beta v_{gc}^{-1}(-\beta u_c + k_B^{-1} s_{mc}(u_c, v^{-1}))$$

shows that $p_{gc}(\beta,\lambda) \equiv p$, implying that $p_{gc}(\beta,\lambda)$ is the pressure expressed, however, as a function of temperature and chemical potential.

To go beyond the heuristic derivations above, it should be remarked that convexity and the property that the maxima in [16], [17] are reached in the interior of the intervals of variability of $v$ or $u$ are sufficient to turn the above arguments into rigorous mathematical deductions: this means that given [19] as definitions of $p(u,v), \beta(u,v)$, the second of [20] follows as well as $p_{gc}(\beta,\lambda) \equiv p(u_v, v_{gc}^{-1})$. But the values $v_{gc}$ and $u_c$ in [16] are not necessarily unique: convex functions can contain horizontal segments and therefore the general conclusion is that the maxima may possibly be attained in intervals. Hence, instead of a single $v_{gc}$, there might be a whole interval $[v_-, v_+]$, where the rhs of [16] reaches the maximum and, instead of a single $u_c$, there might be a whole interval $[u_-, u_+]$ where the rhs of [17] reaches the maximum.

Convexity implies that the values of $\lambda$ or $\beta$ for which the maxima in [16] or [17] are attained in intervals rather than in single points are rare (i.e., at most denumerably many): the interpretation is, in such cases, that the thermodynamic functions show discontinuities, and the corresponding phenomena are called phase transitions (see the next section).

For more details the reader is referred to Ruelle (1969) and Gallavotti (1999).

## Phase Transitions and Boundary Conditions

The analysis in the last two sections of the relations between elements of ensembles of distributions describing macroscopic equilibrium states not only allows us to obtain mechanical models of thermodynamics but also shows that the models, for a given system, coincide at least as $\Omega \to \infty$. Furthermore, the equivalence between the thermodynamic functions computed via corresponding distributions in different ensembles can be extended to a full equivalence of the distributions.

If the maxima in [16] are attained at single points $v_{gc}$ or $u_c$ the equivalence should take place in the sense that a correspondence between $\mu^{gc}_{\beta,\lambda}, \mu^c_{\beta,v}, \mu^{mc}_{u,v}$ can be established so that, given any local observable $F(\boldsymbol{P},\boldsymbol{Q})$, defined as an observable depending on $(\boldsymbol{P},\boldsymbol{Q})$ only through the $\boldsymbol{p}_i, \boldsymbol{q}_i$ with $\boldsymbol{q}_i \in \Lambda$, where $\Lambda \subset \Omega$ is a finite region, has the same average with respect to corresponding distributions in the limit $\Omega \to \infty$.

The correspondence is established by considering $(\lambda,\beta) \leftrightarrow (\beta, v_{gc}) \leftrightarrow (u_{mc}, v)$, where $v_{gc}$ is where the maximum in [16] is attained, $u_{mc} \equiv u_c$ is where the maximum in [17] is attained and $v_{gc} \equiv v$, (cf. also [19], [20]). This means that the limits

$$\lim_{V\to\infty} \int F(\boldsymbol{P},\boldsymbol{Q})\mu^a(d\boldsymbol{P}\,d\boldsymbol{Q}) \stackrel{\text{def}}{=} \langle F\rangle_a \qquad (a - \text{independent}),\ a = \text{gc}, \text{c}, \text{mc} \tag{21}$$

coincide if the averages are evaluated by the distributions $\mu^{gc}_{\beta,\lambda}, \mu^c_{\beta,v_c}, \mu^{mc}_{u_{mc},v_{mc}}$.

Exceptions to [21] are possible: and are certainly likely to occur at values of $u$, $v$ where the maxima in [16] or [17] are attained in intervals rather than in isolated points; but this does not exhaust, in general, the cases in which [21] may not hold.

However, no case in which [21] fails has to be regarded as an exception. It rather signals that an interesting and important phenomenon occurs. To understand it properly, it is necessary to realize that the grand canonical, canonical, and microcanonical families of probability distributions are by far not the only ensembles of probability distributions whose elements can be considered to generate models of thermodynamics, that is, which are orthodic in the sense of the discussion in the section "Equivalence of ensembles." More general families of orthodic statistical ensembles of probability

distributions can be very easily conceived. In particular:

**Definition** Consider the grand canonical, canonical, and microcanonical distributions associated with an energy function in which the potential energy contains, besides the interaction $\Phi$ between particles located inside the container, also the interaction energy $\Phi_{\text{in, out}}$ between particles inside the container and external particles, identical to the ones in the container but not allowed to move and fixed in positions such that in every unit cube $\Delta$ external to $\Omega$ there is a finite number of them bounded independently of $\Delta$. Such configurations of external particles will be called "boundary conditions of fixed external particles."

The thermodynamic limit with such boundary conditions is obtained by considering the grand canonical, canonical, and microcanonical distributions constructed with potential energy function $\Phi + \Phi_{\text{in, out}}$ in containers $\Omega$ of increasing size taking care that, while the size increases, the fixed particles that would become internal to $\Omega$ are eliminated. The argument used in the section "Thermodynamic limit" to show that the three models of thermodynamics, considered there, did define the same thermodynamic functions can be repeated to reach the conclusion that also the (infinitely many) "new" models of thermodynamics in fact give rise to the same thermodynamic functions and averages of local observables. Furthermore, the values of the limits corresponding to [13] can be computed using the new partition functions and coincide with the ones in [13] (i.e., they are independent of the boundary conditions).

However, it may happen, and in general it is the case, for many models and for particular values of the state parameters, that the limits in [21] do not coincide with the analogous limits computed in the new ensembles, that is, the averages of some local observables are unstable with respect to changes of boundary conditions with fixed particles.

There is a very natural interpretation of such apparent ambiguity of the various models of thermodynamics: namely, at the values of the parameters that are selected to describe the macroscopic states under consideration, there may correspond different equilibrium states with the same parameters. When the maximum in [16] is reached on an interval of densities, one should not think of any failure of the microscopic models for thermodynamics: rather one has to think that there are several states possible with the same $\beta, \lambda$ and that they can be identified with the probability distributions obtained by forming the grand canonical, canonical, or microcanonical distributions with different kinds of boundary conditions.

For instance, a boundary condition with high density may produce an equilibrium state with parameters $\beta, \lambda$ which also has high density, i.e., the density $v_+^{-1}$ at the right extreme of the interval in which the maximum in [16] is attained, while using a low-density boundary condition the limit in [21] may describe the averages taken in a state with density $v_-^{-1}$ at the left extreme of the interval or, perhaps, with a density intermediate between the two extremes. Therefore, the following definition emerges.

**Definition** If the grand canonical distributions with parameters $(\beta, \lambda)$ and different choices of fixed external particles boundary conditions generate for some local observable $F$ average values which are different by more than a quantity $\delta > 0$ for all large enough volumes $\Omega$ then one says that the system has a phase transition at $(\beta, \lambda)$. This implies that the limits in [21], when existing, will depend on the boundary condition and their values will represent averages of the observables in "different phases." A corresponding definition is given in the case of the canonical and microcanonical distributions when, given $(\beta, v)$ or $(u, v)$, the limit in [21] depends on the boundary conditions for some $F$.

**Remarks**

1. The idea is that by fixing one of the thermodynamic ensembles and by varying the boundary conditions one can realize all possible states of equilibrium of the system that can exist with the given values of the parameters determining the state in the chosen ensemble (i.e., $(\beta, \lambda)$, $(\beta, v)$, or $(u, v)$ in the grand canonical, canonical, or microcanonical cases, respectively).
2. The impression that in order to define a phase transition the thermodynamic limit is necessary is incorrect: the definition does not require considering the limit $\Omega \to \infty$. The phenomenon that occurs is that by changing boundary conditions the average of a local observable can change at least by amounts independent of the system size. Hence, occurrence of a phase transition is perfectly observable in finite volume: it suffices to check that by changing boundary conditions the average of some observable changes by an amount whose minimal size is volume independent. It is a manifestation of an instability of the averages with respect to changes in boundary conditions: an instability which does not fade away when the boundary recedes to infinity, i.e., boundary perturbations produce

bulk effects and at a phase transition the averages of the local observable, if existing at all, will exhibit a nontrivial dependence on the boundary conditions. This is also called "long range order."

3. It is possible to show that when this happens then some thermodynamic function whose value is independent of the boundary condition (e.g., the free energy in the canonical distributions) has discontinuous derivatives in terms of the parameters of the ensemble. This is in fact one of the frequently-used alternative definitions of phase transitions: the latter two natural definitions of first-order phase transition are equivalent. However, it is very difficult to prove that a given system shows a phase transition. For instance, existence of a liquid–gas phase transition is still an open problem in systems of the type considered until the section "Lattice models" below.
4. A remarkable unification of the theory of the equilibrium ensembles emerges: all distributions of any ensemble describe equilibrium states. If a boundary condition is fixed once and for all, then some equilibrium states might fail to be described by an element of an ensemble. However, if all boundary conditions are allowed then all equilibrium states should be realizable in a given ensemble by varying the boundary conditions.
5. The analysis leads us to consider as completely equivalent without exceptions grand canonical, canonical, or microcanonical ensembles enlarged by adding to them the distributions with potential energy augmented by the interaction with fixed external particles.
6. The above picture is really proved only for special classes of models (typically in models in which particles are constrained to occupy points of a lattice and in systems with hard core interactions, $r_0 > 0$ in [14]) but it is believed to be correct in general. At least it is consistent with all that is known so far in classical statistical mechanics. The difficulty is that, conceivably, one might even need boundary conditions more complicated than the fixed particles boundary conditions (e.g., putting different particles outside, interacting with the system with an arbitrary potential, rather than via $\varphi$).

The discussion of the equivalence of the ensembles and the question of the importance of boundary conditions has already imposed the consideration of several limits as $\Omega \to \infty$. Occasionally, it will again come up. For conciseness, it is useful to set up a formal definition of equilibrium states of an infinite-volume system: although infinite volume is an idealization void of physical reality, it is nevertheless useful to define such states because certain notions (e.g., that of pure state) can be sharply defined, with few words and avoiding wide circumvolutions, in terms of them. Therefore, let:

**Definition** An infinite-volume state with parameters $(\beta, v), (u, v)$ or $(\beta, \lambda)$ is a collection of average values $F \to \langle F \rangle$ obtained, respectively, as limits of finite-volume averages $\langle F \rangle_{\Omega_n}$ defined from canonical, microcanonical, or grand canonical distributions in $\Omega_n$ with fixed parameters $(\beta, v), (u, v)$ or $(\beta, \lambda)$ and with general boundary condition of fixed external particles, on sequences $\Omega_n \to \infty$ for which such limits exist simultaneously for all local observables $F$.

Having set the definition of infinite-volume state consider a local observable $G(X)$ and let $\tau_\xi G(X) = G(X + \xi), \xi \in \mathbb{R}^d$, with $X + \xi$ denoting the configuration $X$ in which all particles are translated by $\xi$: then an infinite-volume state is called a pure state if for any pair of local observables $F, G$ it is

$$\langle F \tau_\xi G \rangle - \langle F \rangle \langle \tau_\xi G \rangle \underset{\xi \to \infty}{\longrightarrow} 0 \qquad [22]$$

which is called a cluster property of the pair $F, G$.

The result alluded to in remark (6) is that at least in the case of hard-core systems (or of the simple lattice systems discussed in the section "Lattice models") the infinite-volume equilibrium states in the above sense exhaust at least the totality of the infinite-volume pure states. Furthermore, the other states that can be obtained in the same way are convex combinations of the pure states, i.e., they are "statistical mixtures" of pure phases. Note that $\langle \tau_\xi G \rangle$ cannot be replaced, in general, by $\langle G \rangle$ because not all infinite-volume states are necessarily translation invariant and in simple cases (e.g., crystals) it is even possible that no translation-invariant state is a pure state.

**Remarks**

1. This means that, in the latter models, generalizing the boundary conditions, for example considering external particles to be not identical to the ones inside the system, using periodic or partially periodic boundary conditions, or the widely used alternative of introducing a small auxiliary potential and first taking the infinite-volume states in presence of it and then letting the potential vanish, does not enlarge further the set of states (but may sometimes be useful: an example of a study of a phase transition by using the latter method of small fields will be given in the section "Continuous symmetries: 'no $d = 2$ crystal' theorem").

2. If $\chi$ is the indicator function of a local event, it will make sense to consider the probability of occurrence of the event in an infinite-volume state defining it as $\langle\chi\rangle$. In particular, the probability density for finding $p$ particles at $x_1, x_2, \ldots, x_p$, called the $p$-point correlation function, will thus be defined in an infinite-volume state. For instance, if the state is obtained as a limit of canonical states $\langle\cdot\rangle_{\Omega_n}$ with parameters $\beta, \rho, \rho = N_n/V_n$, in a sequence of containers $\Omega_n$, then

$$\rho(\boldsymbol{x}) = \lim_n \left\langle \sum_{j=1}^{N_n} \delta(\boldsymbol{x} - \boldsymbol{q}_j) \right\rangle_{\Omega_n}$$

$$\rho(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_p) = \lim_n \left\langle \sum_{i_1,\ldots,i_p}^{N_n} \prod_{j=1}^{p} \delta(\boldsymbol{x}_j - \boldsymbol{q}_{i_j}) \right\rangle_{\Omega_n}$$

where the sum is over the ordered $p$-ples $(j_1, \ldots, j_p)$. Thus, the *pair correlation* $\rho(\boldsymbol{q}, \boldsymbol{q}')$ and its possible *cluster property* are

$$\rho(\boldsymbol{q}, \boldsymbol{q}') \stackrel{\text{def}}{=} \lim_n \frac{\int_{\Omega_n} \exp(-\beta U(\boldsymbol{q}, \boldsymbol{q}', \boldsymbol{q}_1, \ldots, \boldsymbol{q}_{N_n-2}))\, d\boldsymbol{q}_1 \cdots d\boldsymbol{q}_{N_n-2}}{(N_n-2)!\, Z_0^c(\beta, \rho, V_n)}$$

$$\rho(\boldsymbol{q}, (\boldsymbol{q}' + \boldsymbol{\xi})) - \rho(\boldsymbol{q})\rho(\boldsymbol{q}' + \boldsymbol{\xi}) \underset{\xi \to \infty}{\longrightarrow} 0 \qquad [23]$$

where

$$Z_0^c \stackrel{\text{def}}{=} \int e^{-\beta U(Q)} dQ$$

is the "configurational" partition function.

The reader is referred to Ruelle (1969), Dobrushin (1968), Lanford and Ruelle (1969), and Gallavotti (1999).

## Virial Theorem and Atomic Dimensions

For a long time it has been doubted that "just changing boundary conditions" could produce such dramatic changes as macroscopically different states (i.e., phase transitions in the sense of the definition in the last section). The first evidence that by taking the thermodynamic limit very regular analytic functions like $N^{-1} \log Z^c(\beta, N, V)$ (as a function of $\beta, v = V/N$) could develop, in the limit $\Omega \to \infty$, singularities like discontinuous derivatives (corresponding to the maximum in [16] being reached on a plateau and to a consequent existence of several pure phases) arose in the van der Waals' theory of liquid–gas transition.

Consider a real gas with $N$ identical particles with mass $m$ in a container $\Omega$ with volume $V$. Let the force acting on the $i$th particle be $\boldsymbol{f}_i$; multiplying both sides of the equations of motion, $m\ddot{\boldsymbol{q}}_i = \boldsymbol{f}_i$, by $-(1/2)\boldsymbol{q}_i$ and summing over $i$, it follows that

$$-\frac{1}{2}\sum_{i=1}^{N} m\boldsymbol{q}_i \cdot \ddot{\boldsymbol{q}}_i = -\frac{1}{2}\sum_{i=1}^{N} \boldsymbol{q}_i \cdot \boldsymbol{f}_i \stackrel{\text{def}}{=} \frac{1}{2} C(\boldsymbol{q})$$

and the quantity $C(\boldsymbol{q})$ defines the *virial of the forces* in the configuration $\boldsymbol{q}$. Note that $C(\boldsymbol{q})$ is not translation invariant because of the presence of the forces due to the walls.

Writing the force $\boldsymbol{f}_i$ as a sum of the internal and the external forces (due to the walls) the virial $C$ can be expressed naturally as sum of the virial $C_{\text{int}}$ of the internal forces (translation invariant) and of the virial $C_{\text{ext}}$ of the external forces.

By dividing both sides of the definition of the virial by $\tau$ and integrating over the time interval $[0, \tau]$, one finds in the limit $\tau \to +\infty$, that is, up to quantities relatively infinitesimal as $\tau \to \infty$, that

$$\langle K \rangle = \tfrac{1}{2}\langle C \rangle \quad \text{and} \quad \langle C_{\text{ext}} \rangle = 3pV$$

where $p$ is the pressure and $V$ the volume. Hence

$$\langle K \rangle = \tfrac{3}{2} pV + \tfrac{1}{2}\langle C_{\text{int}} \rangle$$

or

$$\frac{1}{\beta} = pv + \frac{\langle C_{\text{int}} \rangle}{3N} \qquad [24]$$

Equation [24] is Clausius' *virial theorem*: in the case of no internal forces, it yields $\beta pv = 1$, the ideal-gas equation.

The internal virial $C_{\text{int}}$ can be written, if $\boldsymbol{f}_{j\to i} = -\partial_{\boldsymbol{q}_i}\varphi(\boldsymbol{q}_i - \boldsymbol{q}_j)$, as

$$C_{\text{int}} = -\sum_{i=1}^{N}\sum_{i \neq j} \boldsymbol{f}_{j\to i} \cdot \boldsymbol{q}_i \equiv -\sum_{i<j} \partial_{\boldsymbol{q}_i}\varphi(\boldsymbol{q}_i - \boldsymbol{q}_j) \cdot (\boldsymbol{q}_i - \boldsymbol{q}_j)$$

which shows that the contribution to the virial by the internal repulsive forces is negative while that of the attractive forces is positive. The average of $C_{\text{int}}$ can be computed by the canonical distribution, which is convenient for the purpose. van der Waals first used the virial theorem to perform an actual computation of the corrections to the perfect-gas laws. Simply neglect the third-order term in the density and use the approximation $\rho(\boldsymbol{q}_1, \boldsymbol{q}_2) = \rho^2 e^{-\beta\varphi(\boldsymbol{q}_1 - \boldsymbol{q}_2)}$ for the pair correlation function, [23], then

$$\frac{1}{2}\langle C_{\text{int}} \rangle = V \frac{3}{2\beta} \rho^2 I(\beta) + VO(\rho^3) \qquad [25]$$

where

$$I(\beta) = \frac{1}{2}\int (e^{-\beta\varphi(q)} - 1)d^3 q$$

and the equation of state [24] becomes

$$pv + \frac{I(\beta)}{\beta v} + O(v^{-2}) = \beta^{-1}$$

For the purpose of illustration, the calculation of $I$ can be performed approximately at "high temperature" ($\beta$ small) in the case

$$\varphi(r) = 4\varepsilon\left(\left(\frac{r_0}{r}\right)^{12} - \left(\frac{r_0}{r}\right)^6\right)$$

(the classical Lennard–Jones potential), $\varepsilon, r_0 > 0$. The result is

$$I \cong -(b - \beta a)$$

$$b = 4v_0,\ a = \frac{32}{3}\varepsilon v_0,\ v_0 = \frac{4\pi}{3}\left(\frac{r_0}{2}\right)^3$$

Hence,

$$pv + \frac{a}{v} - \frac{b}{\beta v} = \frac{1}{\beta} + O\left(\frac{1}{\beta v^2}\right)$$

$$\left(p + \frac{a}{v^2}\right)v = \left(1 + \frac{b}{v}\right)\frac{1}{\beta} = \frac{1}{1 - b/v}\frac{1}{\beta} + O\left(\frac{1}{\beta v^2}\right)$$

or

$$\left(p + \frac{a}{v^2}\right)(v - b)\beta = 1 + O(v^{-2}) \qquad [26]$$

which gives the equation of state for $\beta\varepsilon \ll 1$. Equation [26] can be compared with the well-known empirical *van der Waals equation* of state:

$$\beta\left(p + \frac{a}{v^2}\right)(v - b) = 1$$

or

$$(p + An^2/V^2)(V - nB) = nRT \qquad [27]$$

where, if $N_A$ is Avogadro's number, $A = aN_A^2$, $B = bN_A, R = k_B N_A, n = N/N_A$. It shows the possibility of accessing the microscopic parameters $\varepsilon$ and $r_0$ of the potential $\varphi$ via measurements detecting deviations from the *Boyle–Mariotte law*, $\beta pv = 1$, of the rarefied gases: $\varepsilon = 3a/8b = 3A/8BN_A$ $r_0 = (3b/2\pi)^{1/3} = (3B/2\pi N_A)^{1/3}$.

As a final comment, it is worth stressing that the virial theorem gives in principle the exact corrections to the equation of state, in a rather direct and simple form, as time averages of the virial of the internal forces. Since the virial of the internal forces is easy to calculate from the positions of the particles as a function of time, the theorem provides a method for computing the equation of state in numerical simulations. In fact, this idea has been exploited in many numerical experiments, in which [24] plays a key role.

For more details, the reader is referred to Gallavotti (1999).

## van der Waals Theory

Equation [27] is empirically used beyond its validity region (small density and small $\beta$) by regarding $A, B$ as phenomenological parameters to be experimentally determined by measuring them near generic values of $p, V, T$. The measured values of $A, B$ do not "usually vary too much" as functions of $v, T$ and, apart from this small variability, the predictions of [27] have reasonably agreed with experience until, as experimental precision increased over the years, serious inadequacies eventually emerged.

Certain consequences of [27] are appealing: for example, Figure 1 shows that it does not give a $p$ monotonic nonincreasing in $v$ if the temperature is small enough. A critical temperature can be defined as the largest value, $T_c$, of the temperature below which the graph of $p$ as a function of $v$ is not monotonic decreasing; the critical volume $V_c$ is the value of $v$ at the horizontal inflection point occurring for $T = T_c$.

For $T < T_c$ the van der Waals interpretation of the equation of state is that the function $p(v)$ may describe metastable states while the actual equilibrium states would follow an equation with a monotonic dependence on $v$ and $p(v)$ becoming horizontal in the coexistence region of specific volumes. The precise value of $p$ where to draw the plateau (see Figure 1) would then be fixed by experiment or theoretically predicted via the simple rule that the plateau associated with the represented isotherm is drawn at a height such that the area of the two cycles in the resulting loop are equal.

This is *Maxwell's rule*: obtained by assuming that the isotherm curve joining the extreme points of the plateau and the plateau itself define a cycle

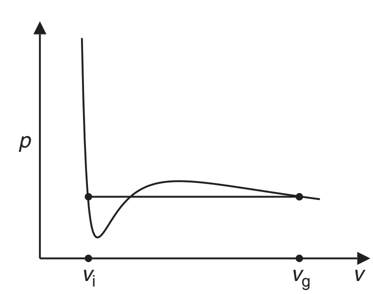



**Figure 1** The van der Waals equation of state at a temperature $T < T_c$ where the pressure is not monotonic. The horizontal line illustrates the "Maxwell rule."

(see **Figure 1**) representing a sequence of possible macroscopic equilibrium states (the ones corresponding to the plateau) or states with extremely long time of stability ("metastable") represented by the curved part. This would be an isothermal Carnot cycle which, therefore, could not produce work: since the work produced in the cycle (i.e., $\oint p\mathrm{d}v$) is the signed area enclosed by the cycle the rule just means that the area is zero. The argument is doubtful at least because it is not clear that the intermediate states with $p$ increasing with $v$ could be realized experimentally or could even be theoretically possible.

A striking prediction of [27], taken literally, is that the gas undergoes a gas–liquid phase transition with a critical point at a temperature $T_c$, volume $v_c$, and pressure $p_c$ that can be computed via [27] and are given by $RT_c = 8A/27B$, $V_c = 3B$ $(n = 1)$.

At the same time, the above prediction is interesting as it shows that there are simple relations between the critical parameters and the microscopic interaction constants, i.e., $\varepsilon \simeq k_B T_c$ and $r_0 \simeq (V_c/N_A))^{1/3}$: or more precisely $\varepsilon = 81 k_B T_c/64$, $r_0 = (V_c/2\pi N_A)^{1/3}$ if a classical Lennard–Jones potential (i.e., $\varphi = 4\varepsilon((r_0/|q|)^{12} - (r_0/|q|)^6)$; see the last section) is used for the interaction potential $\varphi$.

However, [27] cannot be accepted acritically not only because of the approximations (essentially the neglecting of $O(v^{-1})$ in the equation of state), but mainly because, as remarked above, for $T < T_c$ the function $p$ is no longer monotonic in $v$ as it must be; see comment following [19].

The van der Waals equation, refined and complemented by Maxwell's rule, predicts the following behavior:

$$(p - p_c) \propto (v - v_c)^{\delta}, \quad \delta = 3, \ T = T_c$$
$$(v_g - v_l) \propto (T_c - T)^{\beta}, \quad \beta = 1/2, \quad \text{for } T \to T_c^- \qquad [28]$$

which are in sharp contrast with the experimental data gathered in the twentieth century. For the simplest substances, one finds instead $\delta \cong 5$, $\beta \cong 1/3$.

Finally, blind faith in the equation of state [27] is untenable, last but not least, also because nothing in the analysis would change if the space dimension was $d = 2$ or $d = 1$: but for $d = 1$, it is easily proved that the system, if the interaction decays rapidly at infinity, does not undergo phase transitions (see next section).

In fact, it is now understood that van der Waals' equation represents rigorously only a limiting situation, in which particles have a hard-core interaction (or a strongly repulsive one at close distance) and a further smooth interaction $\varphi$ with very long range. More precisely, suppose that the part of the potential outside a hard-core radius $r_0 > 0$ is attractive (i.e., non-negative) and has the form $\gamma^d \varphi_1(\gamma^{-1}|q|) \leq 0$ and call $P_0(v)$ the ($\beta$-independent) product of $\beta$ times the pressure of the hard-core system without any attractive tail ($P_0(v)$ is not explicitly known except if $d = 1$, in which case it is $P_0(v)(v - b) = 1, b = r_0$), and let

$$a = -\frac{1}{2}\int_{|q|>r_0} |\varphi_1(q)| \mathrm{d}q$$

If $p(\beta, v; \gamma)$ is the pressure when $\gamma > 0$ then it can be proved that

$$\beta p(\beta, v) \overset{\text{def}}{=} \lim_{\gamma \to 0} \beta p(\beta, v; \gamma) = \left[-\frac{\beta a}{v^2} + P_0(v)\right]_{\text{Maxwell's rule}} \qquad [29]$$

where the subscript means that the graph of $p(\beta, v)$ as a function of $v$ is obtained from the function in square bracket by applying to it Maxwell's rule, described above in the case of the van der Waals equation. Equation [29] reduces exactly to the van der Waals equation for $d = 1$, and for $d > 1$ it leads to an equation with identical critical behavior (even though $P_0(v)$ cannot be explicitly computed).

The reader is referred to Lebowitz and Penrose (1979) and Gallavotti (1999) for more details.

## Absence of Phase Transitions: $d = 1$

One of the most quoted no-go theorems in statistical mechanics is that one-dimensional systems of particles interacting via short-range forces do not exhibit phase transitions (cf. the next section) unless the somewhat unphysical situation of having zero absolute temperature is considered. This is particularly easy to check in the case of "nearest-neighbor hard-core interactions." Let the hard-core size be $r_0$, so that the interaction potential $\varphi(r) = +\infty$ if $r \leq r_0$, and suppose also that $\varphi(r) \equiv 0$ if $f \geq 2r_0$. In this case, the thermodynamic functions can be exactly computed and checked to be analytic: hence the equation of state cannot have any phase transition plateau. This is a special case of *van Hove's theorem* establishing smoothness of the equation of state for interactions extending beyond the nearest neighbor and rapidly decreasing at infinity.

If the definition of phase transition based on the sensitivity of the thermodynamic limit to variations of boundary conditions is adopted then a more general, conceptually simple, argument can be given to show that in one-dimensional systems there cannot be any phase transition if the potential energy of mutual interaction between a

configuration $Q$ of particles to the left of a reference particle (located at the origin O, say) and a configuration $Q'$ to the right of the particle (with $Q \cup O \cup Q'$ compatible with the hard cores) is uniformly bounded below. Then a mathematical proof can be devised showing that the influence of boundary conditions disappears as the boundaries recede to infinity. One also says that no long-range order can be established in a one-dimensional case, in the sense that one loses any trace of the boundary conditions imposed.

The analysis fails if the space dimension is $\geq 2$: in this case, even if the interaction is short-ranged, the energy of interaction between two regions of space separated by a boundary is of the order of the boundary area. Hence, one cannot bound above and below the probability of any two configurations in two half-spaces by the product of the probabilities of the two configurations, each computed as if the other was not there. This is because such a bound would be proportional to the exponential of the surface of separation, which tends to $\infty$ when the surface grows large. This means that we cannot consider, at least not in general, the configurations in the two half-spaces as independently distributed.

Analytically, a condition on the potential sufficient to imply that the energy between a configuration to the left and one to the right of the origin is bounded below, if $d = 1$, is simply expressed by

$$\int_{r'}^{\infty} r|\varphi(r)|\mathrm{d}r < +\infty \quad \text{for } r' > r_0$$

Therefore, in order to have phase transitions in $d = 1$, a potential is needed that is "so long range" that it has a divergent first moment. It can be shown by counterexamples that if the latter condition fails there can be phase transitions even in $d = 1$ systems.

The results just quoted also apply to discrete models like lattice gases or lattice spin models that will be considered later in the article.

For more details, we refer the reader to Landau and Lifschitz (1967), Dyson (1969), Gallavotti (1999), and Gallavotti *et al.* (2004).

## Continuous Symmetries: "No $d = 2$ Crystal" Theorem

A second case in which it is possible to rule out existence of phase transitions or at least of certain kinds of transitions arises when the system under analysis enjoys large symmetry. By symmetry is meant a group of transformations acting on the configurations and transforming each of them into a configuration which, at least for one boundary condition (e.g., periodic or open), has the same energy.

A symmetry is said to be "continuous" if the group of transformations is a continuous group. For instance, continuous systems have translational symmetry if considered in a container $\Omega$ with periodic boundary conditions. Systems with "too much symmetry" sometimes cannot show phase transitions. For instance, the continuous translation symmetry of a gas in a container $\Omega$ with periodic boundary conditions is sufficient to exclude the possibility of crystallization in dimension $d = 2$.

To discuss this, which is a prototype of a proof which can be used to infer absence of many transitions in systems with continuous symmetries, consider the translational symmetry and a potential satisfying, besides the usual [14] and with the symbols used in [14], the further property that $|\boldsymbol{q}|^2|\partial^2_{ij}\varphi(\boldsymbol{q})| < \overline{B}|\boldsymbol{q}|^{-(d+\varepsilon_0)}$, with $\varepsilon_0 > 0$, for some $\overline{B}$ holds for $r_0 < |\boldsymbol{q}| \leq R$. This is a very mild extra requirement (and it allows for a hard-core interaction).

Consider an "ideal crystal" on a square lattice (for simplicity) of spacing a, exactly fitting in its container $\Omega$ of side $L$ assumed with periodic boundary conditions: so that $N = (L/a)^d$ is the number of particles and $a^{-d}$ is the density, which is supposed to be smaller than the close packing density if the interaction $\varphi$ has a hard core. The probability distribution of the particles is rather trivial:

$$\overline{\mu} = \sum_p \prod_{\boldsymbol{n}} \delta(\boldsymbol{q}_{p(\boldsymbol{n})} - a\,\boldsymbol{n}) \frac{\mathrm{d}Q}{N!}$$

the sum running over the permutations $\boldsymbol{m} \to p(\boldsymbol{m})$ of the sites $\boldsymbol{m} \in \Omega, \boldsymbol{m} \in \mathbb{Z}^d, 0 < m_i \leq La^{-1}$. The density at $\boldsymbol{q}$ is

$$\widehat{\rho}(\boldsymbol{q}) = \sum_{\boldsymbol{n}} \delta(\boldsymbol{q} - a\,\boldsymbol{n}) \equiv \left\langle \sum_{j=1}^{N} \delta(\boldsymbol{q} - \boldsymbol{q}_j) \right\rangle$$

and its Fourier transform is proportional to

$$\rho(\boldsymbol{k}) \stackrel{\mathrm{def}}{=} \frac{1}{N} \left\langle \sum_j \mathrm{e}^{-i\boldsymbol{k}\cdot\boldsymbol{q}_j} \right\rangle, \quad \boldsymbol{k} = \frac{2\pi}{L}\boldsymbol{n},\ \boldsymbol{n} \in \mathbb{Z}^d$$

$\rho(\boldsymbol{k})$ has value 1 for all $\boldsymbol{k}$ of the form $\boldsymbol{K} = (2\pi/a)\boldsymbol{n}$ and $(1/N)O(\max_{c=1,2} |\mathrm{e}^{\mathrm{i}k_c a} - 1|^{-2})$ otherwise. In presence of interaction, it has to be expected that, in a crystal state, $\rho(\boldsymbol{k})$ has peaks near the values $\boldsymbol{K}$: but the value of $\rho(\boldsymbol{k})$ can depend on the boundary conditions.

Since the system is translation invariant a crystal state defined as a state with a distribution "close" to $\overline{\mu}$,

i.e., with $\hat{\rho}(\boldsymbol{q})$ with peaks at the ideal lattice points $\boldsymbol{q} = \boldsymbol{n}a$, cannot be realized under periodic boundary conditions, even when the system state is crystalline. To realize such a state, a symmetry-breaking term is needed in the interaction.

This can be done in several ways, for example, by changing the boundary condition. Such a choice implies a discussion of how much the boundary conditions influence the positions of the peaks of $\rho(\boldsymbol{k})$: for instance, it is not obvious that a boundary condition will not generate a state with a period different from the one that *a priori* has been selected for disproval (a possibility which would imply a reciprocal lattice of $K$'s different from the one considered to begin with). Therefore, here the choice will be to imagine that an external weak force with potential $\varepsilon W(\boldsymbol{q})$ acts forcing a symmetry breaking that favors the occupation of regions around the points of the ideal lattice (which would mark the average positions of the particles in the crystal state that is being sought). The proof (*Mermin's theorem*) that no equilibrium state with particles distribution "close" to $\overline{\mu}$, i.e., with peaks in place of the delta functions (see below), is essentially reproduced below.

Take $W(\boldsymbol{q}) = \sum_{\boldsymbol{n}a\in\Omega} \chi(\boldsymbol{q} - \boldsymbol{n}a)$, where $\chi(\boldsymbol{q}) \le 0$ is smooth and zero everywhere except in a small vicinity of the lattice points around which it decreases to some negative minimum keeping a rotation symmetry around them. The potential $W$ is invariant under translations by the lattice steps. By the choice of the boundary condition and $\varepsilon W$, the density $\widetilde{\rho}_\varepsilon(\boldsymbol{q})$ will be periodic with period $a$ so that $\rho_\varepsilon(\boldsymbol{k})$ will, possibly, not have a vanishing limit as $N \to \infty$ only if $\boldsymbol{k}$ is a reciprocal vector $\boldsymbol{K} = (2\pi/a)\boldsymbol{n}$. If the potential is $\varphi + \varepsilon W$ and if there exists a crystal state in which particles have higher probability of being near the lattice points $\boldsymbol{n}a$, it should be expected that for small $\varepsilon > 0$ the system will be found in a state with Fourier transform of the density, $\rho_\varepsilon(\boldsymbol{k})$, satisfying, for some vector $\boldsymbol{K} \ne \boldsymbol{0}$ in the reciprocal lattice,

$$\lim_{\varepsilon\to 0}\lim_{N\to\infty} |\rho_\varepsilon(\boldsymbol{K})| = r > 0 \qquad [30]$$

that is, the requirement is that uniformly in $\varepsilon \to 0$ the Fourier transform of the density has a peak at some $\boldsymbol{K} \ne \boldsymbol{0}$. Note that if $\boldsymbol{k}$ is not in the reciprocal lattice $\rho_\varepsilon(\boldsymbol{k}) \xrightarrow[N\to\infty]{} 0$, being bounded above by

$$\frac{1}{N} O\left(\max_{j=1,2} |e^{ik_j a} - 1|^{-2}\right)$$

because $(1/N)\widetilde{\rho}_\varepsilon$ is periodic and its integral over $\boldsymbol{q}$ is equal to 1. Hence, excluding the existence of a crystal will be identified with the impossibility of the [30]. Other criteria can be imagined, for example, considering crystals with a lattice different from simple cubic, which lead to the same result by following the same technique. Nevertheless, it is not mathematically excluded (but unlikely) that, with some weaker existence definition, a crystal state could be possible even in two dimensions.

The following inequalities hold under the present assumptions on the potential and in the canonical distribution with periodic boundary conditions and parameters $(\beta, \rho), \rho = a^{-3}$ in a box $\Omega$ with side multiple of $a$ (so that $N = (La^{-1})^d$) and potential of interaction $\varphi + \varepsilon W$. The further assumption that the lattice $\boldsymbol{n}a$ is not a close-packed lattice is (of course) necessary when the interaction potential has a hard core. Then, for suitable $B_0, B, B_1, B_2 > 0$, independent of $N$, and $\varepsilon$ and for $|\boldsymbol{\kappa}| < \pi/a$ and for all $\Omega$ (*if* $\boldsymbol{K} \ne \boldsymbol{0}$)

$$\frac{1}{N}\left\langle \left| \sum_{j=1}^{N} e^{-i(\boldsymbol{\kappa}+\boldsymbol{K})\cdot \boldsymbol{q}_j} \right|^2 \right\rangle \ge B\frac{(\rho_\varepsilon(\boldsymbol{K}) + \rho_\varepsilon(\boldsymbol{K} + 2\boldsymbol{\kappa}))^2}{B_1\boldsymbol{\kappa}^2 + \varepsilon B_2}$$

$$\frac{1}{N}\sum_{\boldsymbol{\kappa}} \gamma(\boldsymbol{\kappa})\, \frac{d\boldsymbol{\kappa}}{N} \left\langle \left| \sum_{j=1}^{N} e^{-i(\boldsymbol{\kappa}+\boldsymbol{K})\cdot \boldsymbol{q}_j} \right|^2 \right\rangle \le B_0 < \infty \qquad [31]$$

where the averages are in the canonical distribution $(\beta, \rho)$ with periodic boundary conditions and a symmetry-breaking potential $\varepsilon W(\boldsymbol{q}); \gamma(\boldsymbol{k}) \ge 0$ is an (arbitrary) smooth function vanishing for $2|\boldsymbol{\kappa}| \ge \delta$ with $\delta < 2\pi/a$ and $B_0$ depends on $\gamma$. See Appendix 3 for a derivation of [31].

Multiplying both sides of the first equation in [31] by $N^{-1}\gamma(\boldsymbol{\kappa})$ and summing over $\boldsymbol{\kappa}$, the crystallinity condition in the form [30] implies

$$B_0 \ge Br^2 a^d \int_{|\boldsymbol{\kappa}|<\delta} \frac{\gamma(\boldsymbol{\kappa})\, d\boldsymbol{\kappa}}{\boldsymbol{\kappa}^2 B_1 + \varepsilon B_2}$$

For $d = 1, 2$ the integral diverges, as $\varepsilon^{-1/2}$ or $\log \varepsilon^{-1}$, respectively, implying $|\rho_\varepsilon(\boldsymbol{K})| \xrightarrow[\varepsilon\to 0]{} r = 0$: the criterion of crystallinity, [30] cannot be satisfied if $d = 1, 2$.

The above inequality is an example of a general class of inequalities called *infrared inequalities* stemming from another inequality called *Bogoliubov's inequality* (see Appendix 3), which lead to the proof that certain kinds of ordered phases cannot exist if the dimension of the ambient space is $d = 2$ when a finite volume, under suitable boundary conditions (e.g., periodic), shows a continuous symmetry. The excluded phenomenon is, more precisely, the nonexistence of equilibrium states exhibiting, in the thermodynamic limit, a symmetry lower than the continuous symmetry holding in a finite volume.

In general, existence of thermodynamic equilibrium states with symmetry lower than the

symmetry enjoyed by the system in finite volume and under suitable boundary conditions is called a "spontaneous symmetry breaking." It is yet another manifestation of instability with respect to changes in boundary conditions, hence its occurrence reveals a phase transition. There is a large class of systems for which an infrared inequality implies absence of spontaneous symmetry breaking: in most of the one- or two-dimensional systems a continuous symmetry cannot be spontaneously broken.

The limitation to dimension $d \le 2$ is a strong limitation to the generality of the applicability of infrared theorems to exclude phase transitions. More precisely, systems can be divided into classes each of which has a "critical dimension" below which too much symmetry implies absence of phase transitions (or of certain kinds of phase transitions).

It should be stressed that, at the critical dimension, the symmetry breaking is usually so weakly forbidden that one might need astronomically large containers to destroy small effects (due to boundary conditions or to very small fields) which break the symmetry. For example, in the crystallization just discussed, the Fourier transform peaks are only bounded by $O(1/\sqrt{\log \varepsilon^{-1}})$. Hence, from a practical point of view, it might still be possible to have some kind of order even in large containers.

The reader is referred to Mermin (1968), Hohenberg (1969), and Ruelle (1969).

## High Temperature and Small Density

There is another class of systems in which no phase transitions take place. These are the systems with stable and tempered interactions $\varphi$ (e.g., those satisfying [14]) in the high-temperature and low-density region. The property is obtained by showing that the equation of state is analytic in the variables $(\beta, \rho)$ near the origin $(0, 0)$.

A simple algorithm (*Mayer's series*) yields the coefficients of the virial series

$$\beta p(\beta, \rho) = \rho + \sum_{k=2}^{\infty} c_k(\beta)\rho^k$$

It has the drawback that the $k$th order coefficient $c_k(\beta)$ is expressed as a sum of many terms (a number growing more than exponentially fast in the order $k$) and it is not so easy (but possible) to show combinatorially that their sum is bounded exponentially in $k$ if $\beta$ is small enough. A more efficient approach leads quickly to the desired solution. Denoting $\boldsymbol{\Phi}(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n) \stackrel{\text{def}}{=} \sum_{i<j} \varphi(\boldsymbol{q}_i - \boldsymbol{q}_j)$, consider the ("spatial or configurational") correlation functions defined, in the grand canonical distribution with parameters $\beta, \lambda$ (and empty boundary conditions), by

$$\rho_\Omega(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n) \stackrel{\text{def}}{=} \frac{1}{Z^{gc}(\beta, \lambda, V)} \sum_{m=0}^{\infty} z^{n+m} \times \int_\Omega \mathrm{e}^{-\beta\Phi(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_m)} \frac{\mathrm{d}\boldsymbol{y}_1 \cdots \mathrm{d}\boldsymbol{y}_m}{m!} \quad [32]$$

This is the probability density for finding particles with any momentum in the volume element $\mathrm{d}\boldsymbol{q}_1 \cdots \mathrm{d}\boldsymbol{q}_n$ (irrespective of where other particles are), and $z = \mathrm{e}^{\beta\lambda}(\sqrt{2\pi m \beta^{-1} h^{-2}})^d$ accounts for the integration over the momenta variables and is called the activity: it has the dimension of a density (cf. [23]).

Assuming that the potential has a hard core (for simplicity) of radius $R$, the interaction energy $\Phi_{\boldsymbol{q}_1}(\boldsymbol{q}_2, \ldots, \boldsymbol{q}_n)$ of a particle at $\boldsymbol{q}_1$ with any number of other particles at $\boldsymbol{q}_2, \ldots, \boldsymbol{q}_m$ with $|\boldsymbol{q}_i - \boldsymbol{q}_j| > R$ is bounded below by $-B$ for some $B \ge 0$ (related but not equal to the $B$ in [14]). The functions $\rho_\Omega$ will be regarded as a sequence of functions "of one, two, ... particle positions": $\rho_\Omega = \{\rho_\Omega(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n)\}_{n=1}^{\infty}$ vanishing for $\boldsymbol{q}_j \notin \Omega$. Then, one checks that

$$\rho_\Omega(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n) = z\delta_{n,1}\chi_\Omega(\boldsymbol{q}_1) + K\rho_\Omega(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n) \quad [33a]$$

with

$$K\rho_\Omega(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n) \stackrel{\text{def}}{=} \mathrm{e}^{-\beta\Phi_{\boldsymbol{q}_1}(\boldsymbol{q}_2, \ldots, \boldsymbol{q}_n)} \Big(\rho_\Omega(\boldsymbol{q}_2, \ldots, \boldsymbol{q}_n)\delta_{n>1} + \sum_{s=1}^{\infty} \int_\Omega \frac{\mathrm{d}\boldsymbol{y}_1 \cdots \mathrm{d}\boldsymbol{y}_s}{s!} \prod_{k=1}^{s} (\mathrm{e}^{-\beta\varphi(\boldsymbol{q}_1 - \boldsymbol{y}_k)} - 1) \times \rho_\Omega(\boldsymbol{q}_2, \ldots, \boldsymbol{q}_n, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_s)\Big) \quad [33b]$$

where $\delta_{n,1}, \delta_{n>1}$ are Kronecker deltas and $\chi_\Omega(\boldsymbol{q})$ is the indicator function of $\Omega$. Equation [33] is called the *Kirkwood–Salzburg equation* for the family of correlation functions in $\Omega$. The kernel $K$ of the equations is independent of $\Omega$, but the domain of integration is $\Omega$.

Calling $\alpha_\Omega$ the sequence of functions $\alpha_\Omega(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n) \equiv 0$ if $n \neq 1$ and $\alpha_\Omega(\boldsymbol{q}) = z\chi_\Omega(\boldsymbol{q})$, a recursive expansion arises, namely

$$\rho_\Omega = z\alpha_\Omega + z^2 K\alpha_\Omega + z^3 K^2\alpha_\Omega + z^4 K^3 \alpha_\Omega + \cdots \quad [34]$$

It gives the correlation functions, provided the series converges. The inequality

$$|K^p \alpha_\Omega(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n)| \le \mathrm{e}^{(2\beta B+1)p} \left( \int |\mathrm{e}^{-\beta\varphi(\boldsymbol{q})} - 1| \mathrm{d}\boldsymbol{q} \right)^p \stackrel{\text{def}}{=} \mathrm{e}^{(2\beta B+1)p} r(\beta)^{3p} \quad [35]$$

shows that the series [34], called Mayer's series, converges if $|z| < \mathrm{e}^{-(2\beta B+1)} r(\beta)^{-3}$. Convergence is uniform (as $\Omega \to \infty$) and $(K^p)\alpha_\Omega(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n)$ tends to a limit as $V \to \infty$ at fixed $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ and the limit is simply $(K^p\alpha)(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n)$, if $\alpha(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n) \equiv 0$ for $n \neq 1$, and $\alpha(\boldsymbol{q}_1) \equiv 1$. This is because the kernel $K$ contains

the factors $(e^{-\beta\varphi(q_1-y)}-1)$ which decay rapidly or, if $\varphi$ has finite range, will eventually even vanish. It is also clear that $(K^p\alpha)(q_1,\ldots,q_n)$ is translation invariant.

Hence, if $|z|e^{2\beta B+1}r(\beta)^3<1$, the limits, as $\Omega\to\infty$, of the correlation functions exist and can be computed by a convergent power series in $z$; the correlation functions will be translation invariant (in the thermodynamic limit).

In particular, the one-point correlation function $\rho=\rho(q)$ is $\rho=z(1+O(zr(\beta)^3))$, which, to lowest order in $z$, just shows that activity and density essentially coincide when they are small enough. Furthermore, $\beta p_\Omega=(1/V)\log Z^{gc}(\beta,\lambda,V)$ is such that

$$z\partial_z\,\beta p_\Omega=\frac{1}{V}\int\rho_\Omega(q)\,dq$$

(from the definition of $\rho_\Omega$ in [32]). Therefore,

$$\begin{aligned}\beta p(\beta,z)&=\lim_{V\to\infty}\frac{1}{V}\log Z^{gc}(\beta,\lambda,V)\\&=\int_0^z\frac{dz'}{z'}\rho(\beta,z')\end{aligned}\qquad[36]$$

and, since the density $\rho$ is analytic in $z$ as well and $\rho\simeq z$ for $z$ small, the grand canonical pressure is analytic in the density and $\beta p=\rho(1+O(\rho^2))$, at small density. In other words, the equation of state is, to lowest order, essentially the equation of a perfect gas. All quantities that are conceivably of some interest turn out to be analytic functions of temperature and density. The system is essentially a free gas and it has no phase transitions in the sense of a discontinuity or of a singularity in the dependence of a thermodynamic function in terms of others. Furthermore, the system cannot show phase transitions in the sense of sensitive dependence on boundary conditions of fixed external particles. This also follows, with some extra work, from the Kirkwood–Salzburg equations.

The reader is referred to Ruelle (1969) and Gallavotti (1969) for more details.

## Lattice Models

The problem of proving the existence of phase transitions in models of homogeneous gases with pair interactions is still open. Therefore, it makes sense to study the problem of phase transitions in simpler models, tractable to some extent but nontrivial, and which are of practical interest in their own right.

The simplest models are the so-called lattice models in which particles are constrained to points of a lattice: they cannot move in the ordinary sense of the word (but, of course, they could jump) and therefore their configurations do not contain momentum variables.

The interaction energy is just the potential energy, and ensembles are defined as collections of probability distributions on the position coordinates of the particle configurations. Usually, the potential is a pair potential decaying fast at $\infty$ and, often, with a hard-core forbidding double or higher occupancy of the same lattice site. For instance, the *lattice gas* with potential $\varphi$, in a cubic box $\Omega$ with $|\Omega|=V=L^d$ sites of a square lattice with mesh $a>0$, is defined by the potential energy attributed to the configuration $X$ of occupied distinct sites, i.e., subsets $X\subset\Omega$:

$$H(X)=-\sum_{(x,y)\in X}\varphi(x-y)\qquad[37]$$

where the sum is over pairs of distinct points in $X$. The canonical ensemble and the grand canonical ensemble are the collections of distributions, parametrized by $(\beta,\rho),(\rho=N/V)$, or, respectively, by $(\beta,\lambda)$, attributing to $X$ the probability

$$p_{\beta,\rho}(X)=\frac{e^{-\beta H(X)}}{Z_p^c(\beta,N,\Omega)}\delta_{|X|,N}\qquad[38a]$$

or

$$p_{\beta,\lambda}(X)=\frac{e^{\beta\lambda|X|}e^{-\beta H(X)}}{Z_p^{gc}(\beta,\lambda,\Omega)}\qquad[38b]$$

where the denominators are normalization factors that can, respectively, be called, in analogy with the theory of continuous systems, canonical and grand canonical partition functions; the subscript $p$ stands for particles.

A lattice gas in which in each site there can be at most one particle can be regarded as a model for the distribution of a family of spins on a lattice. Such models are quite common and useful (e.g., they arise in studying systems with magnetic properties). Simply identify an "occupied" site with a "spin up" or + and an "empty" site with a "spin down" or − (say). If $\boldsymbol{\sigma}=\{\sigma_x\}_{x\in\Omega}$ is a spin configuration, the energy of the configuration "for potential $\varphi$ and magnetic field $h$" will be

$$H(\boldsymbol{\sigma})=-\sum_{(x,y)\in\Omega}\varphi(x-y)\sigma_x\sigma_y-h\sum_x\sigma_x\qquad[39]$$

with the sum running over pairs $(x,y)\in\Omega$ of distinct sites. If $\varphi(x-y)\equiv J_{xy}\geq 0$, the model is called a *ferromagnetic Ising model*. As in the case of continuous systems, it will be assumed to have a finite range for $\varphi$: that is, $\varphi(x)=0$ for $|x|>R$, for some $R$, unless explicitly stated otherwise.

The canonical and grand canonical ensembles in the box $\Omega$ with respective parameters $(\beta, m)$ or $(\beta, h)$ will be defined as the probability distributions on the spin configurations $\boldsymbol{\sigma} = \{\sigma_x\}_{x\in\Omega}$ with $\sum_{x\in\Omega}\sigma_x = M = mV$ or without constraint on $M$, respectively; hence,

$$p_{\beta,m}(\boldsymbol{\sigma}) = \frac{\exp\left(-\beta\sum_{(x,y)}\varphi(x-y)\sigma_x\sigma_y\right)}{Z_s^{\rm c}(\beta, M, \Omega)}$$
$$p_{\beta,h}(\boldsymbol{\sigma}) = \frac{\exp\left(-\beta h\sum\sigma_x - \beta\sum_{(x,y)}\varphi(x-y)\sigma_x\sigma_y\right)}{Z_s^{\rm gc}(\beta, h, \Omega)} \qquad [40]$$

where the denominators are normalization factors again called, respectively, the canonical and grand canonical partition functions. As in the study of the previous continuous systems, canonical and grand canonical ensembles with "external fixed particle configurations" can be defined together with the corresponding ensembles with "external fixed spin configurations"; the subscript $s$ stands for spins.

For each configuration $X \subset \Omega$ of a lattice gas, let $\{n_x\}$ be $n_x = 1$ if $x \in X$ and $n_x = 0$ if $x \notin X$. Then the transformation $\sigma_x = 2n_x - 1$ establishes a correspondence between lattice gas and spin distributions. In the correspondence, the potential $\varphi(x-y)$ of the lattice gas generates a potential $(1/4)\varphi(x-y)$ for the corresponding spin system and the chemical potential $\lambda$ for the lattice gas is associated with a magnetic field $h$ for the spin system with $h = (1/2)(\lambda + \sum_{x\neq 0}\varphi(x))$.

The correspondence between boundary conditions is natural: for instance, a boundary condition for the lattice gas in which all external sites are occupied becomes a boundary condition in which external sites contain a spin +. The close relation between lattice gas and spin systems permits switching from one to the other with little discussion.

In the case of spin systems, empty boundary conditions are often considered (no spins outside $\Omega$). In lattice gases and spin systems (as well as in continuum systems), often periodic and semiperiodic boundary conditions are considered (i.e., periodic in one or more directions and with empty or fixed external particles or spins in the others).

Thermodynamic limits for the partition functions

$$\begin{aligned}
-\beta f(\beta, v) &= \lim_{\substack{\Omega\to\infty \\ V/N=v}} \frac{1}{N}\log Z_p^{\rm c}(\beta, N, \Omega) \\
\beta p(\beta, \lambda) &= \lim_{\Omega\to\infty} \frac{1}{V}\log Z_p^{\rm gc}(\beta, \lambda, \Omega) \\
-\beta g(\beta, m) &= \lim_{\substack{\Omega\to\infty, \\ M/V\to m}} \frac{1}{V}\log Z_s^{\rm c}(\beta, M, \Omega) \\
\beta f(\beta, h) &= \lim_{\Omega\to\infty} \frac{1}{V}\log Z_s^{\rm gc}(\beta, \lambda, \Omega)
\end{aligned} \qquad [41]$$

can be shown to exist by a method similar to the one discussed in Appendix 2. They have convexity and continuity properties as in the cases of the continuum systems. In the case of a lattice gas, the $f, p$ functions are still interpreted as free energy and pressure, respectively. In the case of spin, $f(\beta, h)$ has the interpretation of magnetic free energy, while $g(\beta, m)$ does not have a special name in the thermodynamics of magnetic systems. As in the continuum systems, it is occasionally useful to define infinite-volume equilibrium states:

**Definition** An infinite-volume state with parameters $(\beta, h)$ or $(\beta, m)$ is a collection of average values $F \to \langle F\rangle$ obtained, respectively, as limits of finite-volume averages $\langle F\rangle_{\Omega_n}$ defined from canonical or grand canonical distributions in $\Omega_n$ with fixed parameters $(\beta, h)$ or $(\beta, m)$, or $(u, v)$ and with general boundary condition of fixed external spins or empty sites, on sequences $\Omega_n \to \infty$ for which such limits exist simultaneously for all local observables $F$.

This is taken verbatim from the definition in the section "Phase transitions and boundary conditions." In this way, it makes sense to define the spin correlation functions for $X = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n)$ as $\langle\sigma_X\rangle$ if $\sigma_X = \prod_j \sigma_{\xi_j}$. For instance, we shall call $\rho(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \stackrel{\rm def}{=} \langle\sigma_{\xi_1}\sigma_{\xi_2}\rangle$ and a pure phase can be defined as an infinite-volume state such that

$$\langle\sigma_X\sigma_{Y+\xi}\rangle - \langle\sigma_X\rangle\langle\sigma_{Y+\xi}\rangle \underset{\xi\to\infty}{\longrightarrow} 0 \qquad [42]$$

Again, for more details, we refer the reader to Ruelle (1969) and Gallavotti (1969).

## Thermodynamic Limits and Inequalities

An interesting property of lattice systems is that it is possible to study delicate questions like the existence of infinite-volume states in some (moderate) generality. A typical tool is the use of inequalities. As the simplest example of a vast class of inequalities, consider the ferromagnetic Ising model with some finite (but arbitrary) range interaction $J_{xy} \geq 0$ in a field $h_x \geq 0$: $J, h$ may even be not translationally invariant. Then the average of $\sigma_X \stackrel{\rm def}{=} \sigma_{x_1}\sigma_{x_2}\cdots\sigma_{x_n}$, $X = (x_1, \ldots, x_n)$, in a state with "empty boundary conditions" (i.e., no external spins) satisfies the inequalities

$$\langle\sigma_X\rangle,\ \partial_{h_x}\langle\sigma_X\rangle,\ \partial J_{xy}\langle\sigma_X\rangle \geq 0 \quad X = (x_1, \ldots, x_n)$$

More generally, let $H(\boldsymbol{\sigma})$ in [39] be replaced by $H(\boldsymbol{\sigma}) = -\sum_X J_X\sigma_X$ with $J_X \geq 0$ and $X$ can be any finite set; then, if $Y = (y_1, \ldots, y_n)$, $X = (x_1, \ldots, x_n)$, the following *Griffiths inequalities* hold:

$$\langle\sigma_X\rangle \geq 0, \quad \partial_{J_Y}\langle\sigma_X\rangle \equiv \langle\sigma_X\sigma_Y\rangle - \langle\sigma_X\rangle\langle\sigma_Y\rangle \geq 0 \quad [43]$$

The inequalities can be used to check, in ferromagnetic Ising models, [39], existence of infinite-volume states (cf. the sections "Phase transitions and boundary conditions" and "Lattice models") obtained by fixing the boundary condition $\mathcal{B}$ to be either "all external spins +" or "all external sites empty." If $\langle F\rangle_{\mathcal{B},\Omega}$ denotes the grand canonical average with boundary condition $\mathcal{B}$ and any fixed $\beta, h > 0$, this means that for all local observables $F(\boldsymbol{\sigma}_\Lambda)$ (i.e., for all $F$ depending on the spin configuration in any fixed region $\Lambda$) all the following limits exist:

$$\lim_{\Omega\to\infty} \langle F\rangle_{\mathcal{B},\Omega} = \langle F\rangle_{\mathcal{B}} \qquad [44]$$

The reason is that the inequalities [43] imply that all averages $\langle \sigma_X\rangle_{\mathcal{B},\Omega}$ are monotonic in $\Omega$ for all fixed $X \subset \Omega$: so the limit [44] exists for $F(\boldsymbol{\sigma}) = \sigma_X$. Hence, it exists for all $F$'s depending only on finitely many spins, because any local function $F$ "measurable in $\Lambda$" can be expressed (uniquely) as a linear combination of functions $\sigma_X$ with $X \subseteq \Lambda$.

Monotonicity with empty boundary conditions is seen by considering the sites outside $\Omega$ and in a region $\Omega'$ with side one unit larger than that of $\Omega$ and imagining that the couplings $J_X$ with $X \subset \Omega'$ but $X \not\subset \Omega$ vanish. Then, $\langle\sigma_X\rangle_{\Omega'} \geq \langle\sigma_X\rangle_{\Omega}$, because $\langle\sigma_X\rangle_{\Omega'}$ is an average computed with a distribution corresponding to an energy with the couplings $J_X$ with $X \not\subset \Omega$, but $X \subset \Omega'$, changed from 0 to $J_X \geq 0$.

Likewise, if the boundary condition is +, then enlarging the box from $\Omega$ to $\Omega'$ corresponds to decreasing an external field $h$ acting on the external spins from $+\infty$ (which would force all external spins to be +) to a finite value $h \geq 0$: so, increasing the box $\Omega$ causes $\langle\sigma_X\rangle_{+,\Omega}$ to decrease. Therefore, as $\Omega$ increases, Ising ferromagnets spin correlations increase if the boundary condition is empty and decrease if it is +.

The inequalities can be used in similar ways to prove that the infinite-volume states obtained from + or empty boundary conditions are translation invariant; and that in zero external field, $h = 0$, the + and − boundary conditions generate pure states if the interaction potential is only a pair ferromagnetic interaction.

There are many other important inequalities which can be used to prove several existence theorems along very simple paths. Unfortunately, their use is mostly restricted to lattice systems and requires very special assumptions on the energy (e.g., ferromagnetic interactions in the above example). The quoted examples were among the first discovered and provide a way to exhibit nontrivial thermodynamic limits and pure states.

For more details, see Ruelle (1969), Lebowitz (1974), Gallavotti (1999), Lieb and Thirring (2001), and Lieb (2002).

## Symmetry-Breaking Phase Transitions

The simplest phase transitions (see the section "Phase transitions and boundary conditions") are symmetry-breaking transitions in lattice systems: they take place when the energy of the system in a container $\Omega$ and with some special boundary condition (e.g., periodic, antiperiodic, or empty) is invariant with respect to the action of a group $\mathcal{G}$ on phase space. This means that on the points $x$ of phase space acts a group of transformations $\mathcal{G}$ so that with each $\gamma \in \mathcal{G}$ is associated a map $x \to x\gamma$ which transforms $x$ into $x\gamma$ respecting the composition law in $\mathcal{G}$, that is, $(x\gamma)\gamma' \equiv x(\gamma\gamma')$. If $F$ is an observable, the action of the group on phase space induces an action on the observable $F$ changing $F(x)$ into $F_\gamma(x) \stackrel{\text{def}}{=} F(x\gamma^{-1})$.

A symmetry-breaking transition occurs when, by fixing suitable boundary conditions and taking the thermodynamic limit, a state $F \to \langle F\rangle$ is obtained in which some local observable shows a nonsymmetric average $\langle F\rangle \neq \langle F_\gamma\rangle$ for some $\gamma$.

An example is provided by the "nearest-neighbor ferromagnetic Ising model" on a d-dimensional lattice with energy function given by [39] with $h = 0$ and $\varphi(x - y) \equiv 0$ unless $|x - y| = 1$, i.e., unless $x, y$ are nearest neighbors, in which case $\varphi(x - y) = J > 0$. With periodic or empty boundary conditions, it exhibits a discrete "up–down" symmetry $\boldsymbol{\sigma} \to -\boldsymbol{\sigma}$.

Instability with respect to boundary conditions can be revealed by considering the two boundary conditions, denoted + or −, in which the lattice sites outside the container $\Omega$ are either occupied by spins + or by spins −. Consider also, for later reference, (1) the boundary conditions in which the boundary spins in the upper half of the boundary are + and the ones in the lower part are −: call this the ±-boundary condition (see Figure 2); or (2) the boundary conditions in

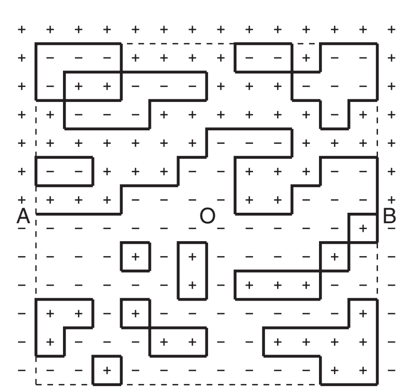



**Figure 2** The dashed line is the boundary of $\Omega$; the outer spins correspond to the $\pm$ boundary condition. The points $A$, $B$ are points where an open "line" $\lambda$ ends.

which some of the opposite sides of $\Omega$ are identified while + or − conditions are assigned on the remaining sides: call these "cylindrical or semiperiodic boundary conditions."

A new description of the spin configurations is useful: given $\boldsymbol{\sigma}$, draw a unit segment perpendicular to the center of each bond $b$ having opposite spins at its extremes. An example of this construction is provided by Figure 2 for the boundary condition $\pm$.

The set of segments can be grouped into lines separating regions where the spins are positive from regions where they are negative. If the boundary condition is + or −, the lines form "closed polygons", whereas, if the condition is $\pm$, there is also a single polygon $\lambda_1$ which is not closed (as in Figure 2). If the boundary condition is periodic or cylindrical, all polygons are closed but some may "go around" $\Omega$. The polygons are also called "contours" and the length of a polygon $\gamma$ will be denoted $|\gamma|$.

The correspondence $(\gamma_1, \gamma_2, \ldots, \gamma_n, \lambda_1) \longleftrightarrow \boldsymbol{\sigma}$, for the boundary condition $\pm$ or, for the boundary condition + (or −), $\boldsymbol{\sigma} \longleftrightarrow (\gamma_1, \ldots, \gamma_n)$ is one-to-one and, if $h = 0$, the energy $H_\Omega(\boldsymbol{\sigma})$ of a configuration is higher than $-J \times$(number of bonds in $\Omega$) by an amount $2J(|\lambda_1| + \sum_i |\gamma_i|)$ or, respectively, $2J\sum_i |\gamma_i|$. The grand canonical probability of each spin configuration is therefore proportional, if $h = 0$, respectively, to

$$e^{-2\beta J(|\lambda_1| + \sum_i |\gamma_i|)} \quad \text{or} \quad e^{-2\beta J \sum_i |\gamma_i|} \qquad [45]$$

and the "up–down" symmetry is clearly reflected by [45].

The average $\langle \sigma_x \rangle_{\Omega,+}$ of $\sigma_+$ with + boundary conditions is given by $\langle \sigma_x \rangle_{\Omega,+} = 1 - 2P_{\Omega,+}(-)$, where $P_{\Omega,+}(-)$ is the probability that the spin $\sigma_x$ is –1. If the site $x$ is occupied by a negative spin then the point $x$ is inside some contour $\gamma$ associated with the spin configuration $\boldsymbol{\sigma}$ under consideration. Hence, if $\rho(\gamma)$ is the probability that a given contour belongs to the set of contours describing a configuration $\boldsymbol{\sigma}$, it is $P_{\Omega,+}(-) \le \sum_{\gamma o x} \rho(\gamma)$ where $\gamma o x$ means that $\gamma$ "surrounds" $x$.

If $\Gamma = (\gamma_1, \ldots, \gamma_n)$ is a spin configuration and if the symbol $\Gamma\,\text{comp}\gamma$ means that the contour $\gamma$ is "disjoint" from $\gamma_1, \ldots, \gamma_n$ (i.e., $\{\gamma \cup \Gamma\}$ is a new spin configuration), then

$$\begin{aligned} \rho(\gamma) &= \frac{\sum_{\Gamma \ni \gamma} e^{-2\beta J \sum_{\gamma' \in \Gamma} |\gamma'|}}{\sum_{\Gamma} e^{-2\beta J \sum_{\gamma' \in \Gamma} |\gamma'|}} \\ &\equiv e^{-2\beta J |\gamma|} \frac{\sum_{\Gamma \text{comp} \gamma} e^{-2\beta J \sum_{\gamma' \in \Gamma} |\gamma'|}}{\sum_{\Gamma} e^{-2\beta J \sum_{\gamma' \in \Gamma} |\gamma'|}} \\ &\le e^{-2\beta J |\gamma|} \end{aligned} \qquad [46]$$

because the last ratio in [46] does not exceed 1. Note that there are $> 3^p$ different shapes of $\gamma$ with perimeter $p$ and at most $p^2$ congruent $\gamma$'s containing $x$; therefore, the probability that the spin at $x$ is − when the boundary condition is + satisfies the inequality

$$P_{\Omega,+}(-) \le \sum_{p=4}^{\infty} p^2 3^p e^{-2\beta J p} \underset{\beta \to \infty}{\longrightarrow} 0$$

This probability can be made arbitrarily small so that $\langle \sigma_x \rangle_{\Omega,+}$ is estimated by a quantity which is as close to 1 as desired provided $\beta$ is large enough and the closeness of $\langle \sigma_x \rangle_{\Omega,+}$ to 1 is estimated by a quantity which is both $x$ and $\Omega$ independent.

A similar argument for the (−)-boundary condition, or the remark that for $h = 0$ it is $\langle \sigma_x \rangle_{\Omega,-} = -\langle \sigma_x \rangle_{\Omega,+}$, leads to conclude that, at large $\beta$, $\langle \sigma_x \rangle_{\Omega,-} \neq \langle \sigma_x \rangle_{\Omega,+}$ and the difference between the two quantities is positive uniformly in $\Omega$. This is the proof (*Peierls' theorem*) of the fact that there is, if $\beta$ is large, a strong instability, of the magnetization with respect to the boundary conditions, i.e., the nearest-neighbor Ising model in dimension 2 (or greater, by an identical argument) has a phase transition. If the dimension is 1, the argument clearly fails and no phase transition occurs (see the section "Absence of phase transitions: $d = 1$").

For more details, see Gallavotti (1999).

## Finite-Volume Effects

The description in the last section of the phase transition in the nearest-neighbor Ising model can be made more precise both from physical and mathematical points of view giving insights into the nature of the phase transitions. Assume that the boundary condition is the (+)-boundary condition and describe a spin configuration $\boldsymbol{\sigma}$ by means of the associated closed disjoint polygons $(\gamma_1, \ldots, \gamma_n)$. Attribute to $\boldsymbol{\sigma} = (\gamma_1, \ldots, \gamma_n)$ a probability proportional to [45]. Then the following *Minlos–Sinai's theorem* holds:

**Theorem** *If $\beta$ is large enough there exist $C > 0$, $\rho(\gamma) > 0$ with $\rho(\gamma) \le e^{-2\beta J|\gamma|}$ and such that a spin configuration $\boldsymbol{\sigma}$ randomly chosen out of the grand canonical distribution with + boundary conditions and $h = 0$ will contain, with probability approaching 1 as $\Omega \to \infty$, a number $K_{(\gamma)}(\boldsymbol{\sigma})$ of contours congruent to $\gamma$ such that*

$$|K_{(\gamma)}(\boldsymbol{\sigma}) - \rho(\gamma)|\Omega|| \le C\sqrt{|\Omega|}\, e^{-\beta J|\gamma|} \qquad [47]$$

*and this relation holds simultaneously for all $\gamma$'s.*

Thus, there are very few contours (and the larger they are the smaller is, in absolute and relative value, their number): a typical spin configuration in the grand canonical ensemble with (+)-boundary conditions is such that the large majority of the spins is "positive" and, in the "sea" of positive spins, there are a few negative spins distributed in small and rare regions (their number, however, is still of order of $|\Omega|$).

Another consequence of the analysis in the last section concerns the the approximate equation of state near the phase transition region at low temperatures and finite $\Omega$. If $\Omega$ is finite, the graph of $h$ versus $m_\Omega(\beta, h)$ will have a rather different behavior depending on the possible boundary conditions. For example, if the boundary condition is $(+)$ or $(-)$, one gets, respectively, the results depicted in Figure 3a and 3b, where $m^*(\beta)$ denotes the spontaneous magnetization (i.e., $m^*(\beta) \stackrel{\text{def}}{=} \lim_{h\to 0^+} \lim_{\Omega\to\infty} m_\Omega(\beta, h)$).

With periodic or empty boundary conditions, the diagram changes as in Figure 4. The thermodynamic limit $m(\beta, h) = \lim_{\Omega\to\infty} m_\Omega(\beta, h)$ exists for all $h \neq 0$ and the resulting graph is in Figure 4b, which shows that at $h = 0$ the limit is discontinuous. It can be proved, if $\beta$ is large enough, that $\infty > \lim_{h\to 0^+} \partial_h m(\beta, h) = \chi(\beta) > 0$ (i.e., the angle between the vertical part of the graph and the rest is sharp).

Furthermore, it can be proved that $m(\beta, h)$ is analytic in $h$ for $h \neq 0$. If $\beta$ is small enough, analyticity holds at all $h$. For $\beta$ large, the function $f(\beta, h)$ has an essential singularity at $h = 0$: a result that can be interpreted as excluding a naive theory of metastability as a description of states governed by an equation of state obtained from an analytic continuation to negative values of $h$ of $f(\beta, h)$.

The above considerations and results further clarify the meaning of a phase transition for a finite system. For more details, we refer the reader to Gallavotti (1999) and Friedli and Pfister (2004).

## Beyond Low Temperatures (Ferromagnetic Ising Model)

A limitation of the results discussed above is the condition of low temperature ("$\beta$ large enough"). A natural problem is to go beyond the low-temperature region and to describe fully the phenomena in the region where boundary condition instability takes place and first develops. A number of interesting partial results are known, which considerably improve the picture emerging from the previous analysis. A striking list, but far from exhaustive, of such results follows and focuses on the properties of ferromagnetic Ising spin systems. The reason for restricting to such cases is that they are simple enough to allow a rather fine analysis, which sheds considerable light on the structure of statistical mechanics suggesting precise formulation

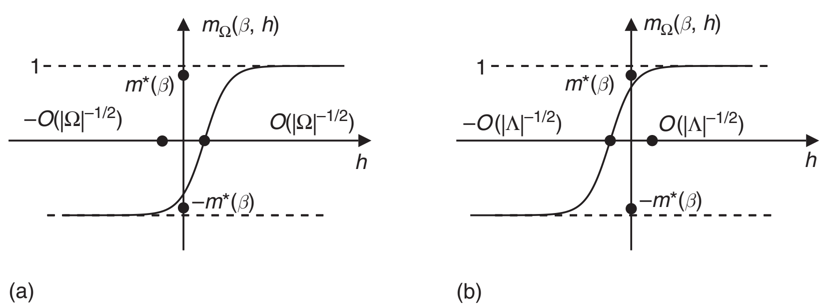



**Figure 3** The $h$ vs $m_\Omega(\beta, h)$ graphs for $\Omega$ finite and (a) + and (b) − conditions.

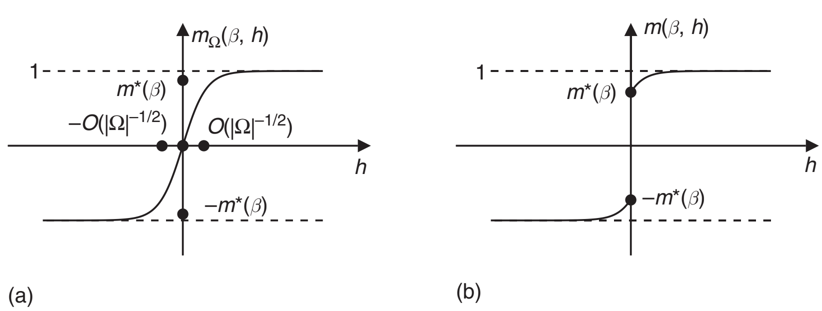



**Figure 4** (a) The $h$ vs $m_\Omega(\beta, h)$ graph for periodic or empty boundary conditions. (b) The discontinuity (at $h = 0$) of the thermodynamic limit.

of the problems that it would be desirable to understand in more general systems.

1. Let $z \stackrel{\text{def}}{=} e^{\beta h}$ and consider that the product of $z^V$ ($V$ is the number of sites $|\Omega|$ of $\Omega$) times the partition function with periodic or perfect-wall boundary conditions and with finite-range ferromagnetic interaction, not necessarily nearest-neighbor; a polynomial in $z$ (of degree $2V$) is thus obtained. Its zeros lie on the unit circle $|z| = 1$: this is *Lee–Yang's theorem*. It implies that the only singularities of $f(\beta, h)$ in the region $0 < \beta < \infty, -\infty < h < +\infty$ can be found at $h = 0$.

   A singularity can appear only if the point $z = 1$ is an accumulation point of the limiting distribution (as $\Omega \to \infty$) of the zeros on the unit circle: if the zeros are $z_1, \ldots, z_{2V}$ then

$$\frac{1}{V}\log z^V Z(\beta, h, \Omega, \text{periodic}) = 2\beta J + \beta h + \frac{1}{V}\sum_{i=1}^{2V}\log(z - z_i)$$

   and if

$$V^{-1} \times (\text{number of zeros of the form } z_j = e^{i\theta_j}, \theta \le \theta_j \le \theta + d\theta) \underset{\Omega\to\infty}{\longrightarrow} \frac{d\rho_\beta(\theta)}{2\pi}$$

   it is

$$\beta f(\beta, h) = 2\beta J + \frac{1}{2\pi}\int_{-\pi}^{\pi}\log(z - e^{i\theta})\, d\rho_\beta(\theta) \qquad [48]$$

   The existence of the measure $d\rho_\beta(\theta)$ follows from the existence of the thermodynamic limit: but $d\rho_\beta(\theta)$ is not necessarily $d\theta$-continuous, i.e., not necessarily proportional to $d\theta$.
2. It can be shown that, with not necessarily a nearest-neighbor interaction, the zeros of the partition function do not move too much under small perturbations of the potential even if one perturbs the energy (at perfect-wall or periodic boundary conditions) into

$$\begin{aligned} H'_\Omega(\boldsymbol{\sigma}) &= H_\Omega(\boldsymbol{\sigma}) + (\delta H_\Omega)(\boldsymbol{\sigma}) \\ (\delta H_\Omega)(\boldsymbol{\sigma}) &= \sum_{X \subset \Omega} J'(X)\,\sigma_X \end{aligned} \qquad [49]$$

   where $J'(X)$ is very general and defined on subsets $X = (x_1, \ldots, x_k) \subset \Omega$ such that the quantity $||J'|| = \sup_{y\in Z^d}\sum_{y\in X}|J'(X)|$ is small enough. More precisely, with a ferromagnetic pair potential $J$ fixed, suppose that one knows that, when $J' = 0$, the partition function zeros in the variable $z = e^{\beta h}$ lie in a certain closed set $N$ (of the unit circle) in the $z$-plane. Then, if $J' \neq 0$, they lie in a closed set $N^1$, $\Omega$-independent and contained in a neighborhood of $N$ of width shrinking to 0 when $||J'|| \to 0$. This allows to establish various relations between analyticity properties and boundary condition instability as described in (3) below.
3. In the ferromagnetic Ising model, with not necessarily a nearest-neighbor interaction, one says that there is a gap around 0 if $d\rho_\beta(\theta) = 0$ near $\theta = 0$. It can be shown that if $\beta$ is small enough there is a gap for all $h$ of width *uniform* in $h$.
4. Another question is whether the boundary condition instability is always revealed by the one-spin correlation function (i.e., by the magnetization) or whether it might be shown only by some correlation functions of higher order. It can be proved that no boundary condition instability occurs for $h \neq 0$; at $h = 0$ it is possible only if

$$\lim_{h\to 0^-} m(\beta, h) \neq \lim_{h\to 0^+} m(\beta, h) \qquad [50]$$

5. A consequence of the Griffiths' inequalities (cf. the section "Thermodynamic limits and inequalities") is that if [50] is true for a given $\beta_0$ then it is true for all $\beta > \beta_0$. Therefore, item (4) leads to a natural definition of the critical temperature $T_c$ as the least upper bound of the $T$'s such that [50] holds ($k_B T = \beta^{-1}$).
6. If $d = 2$ the free energy of the nearest-neighbor ferromagnetic Ising model has a singularity at $\beta_c$ and the value of $\beta_c$ is known exactly from the exact solutions of the model: $m(\beta, 0^+) \stackrel{\text{def}}{=} m^*(\beta) \equiv (1 - \sinh^4 2\beta J)^{1/8}$. The location and nature of the singularities of $f(\beta, 0)$ as a function of $\beta$ remains an open question for $d = 3$. In particular, the question whether there is a singularity of $f(\beta, 0)$ at $\beta = \beta_c$ is open.
7. For $\beta < \beta_c$ there is instability with respect to boundary conditions (see (6) above) and a natural question is: how many "pure" phases can exist in the ferromagnetic Ising model? (cf. the section "Phase transitions and boundary conditions," eqn [22]). Intuition suggests that there should be only two phases: the positively magnetized and the negatively magnetized ones.

   One has to distinguish between translation-invariant pure phases and non-translation-invariant ones. It can be proved that, in the case of the two-dimensional nearest-neighbor ferromagnetic Ising models, all infinite-volume states (cf. the section "Lattice models") are translationally invariant. Furthermore, they can be obtained by

considering just the two boundary conditions + and −: the latter states are also pure states for models with non-nearest-neighbor ferromagnetic interaction. The solution of this problem has led to the introduction of many new ideas and techniques in statistical mechanics and probability theory.

8. In any dimension $d \geq 2$, for $\beta$ large enough, it can be proved that the nearest-neighbor Ising model has only two translation-invariant phases. If the dimension is $\geq 3$ and $\beta$ is large, the + and − phases exhaust the set of translation-invariant pure phases but there exist non-translation-invariant phases. For $\beta$ close to $\beta_c$, however, the question is much more difficult.

For more details, see Onsager (1944), Lee and Yang (1952), Ruelle (1971), Sinai (1991), Gallavotti (1999), Aizenman (1980), Higuchi (1981), and Friedli and Pfister (2004).

## Geometry of Phase Coexistence

Intuition about the phenomena connected with the classical phase transitions is usually based on the properties of the liquid–gas phase transition; this transition is usually experimentally investigated in situations in which the total number of particles is fixed (canonical ensemble) and in presence of an external field (gravity).

The importance of such experimental conditions is obvious; the external field produces a nontranslationally invariant situation and the corresponding separation of the two phases. The fact that the number of particles is fixed determines, on the other hand, the fraction of volume occupied by each of the two phases.

Once more, consider the nearest-neighbor ferromagnetic Ising model: the results available for it can be used to obtain a clear picture of the solution to problems that one would like to solve but which in most other models are intractable with present-day techniques.

It will be convenient to discuss phase coexistence in the canonical ensemble distributions on configurations of fixed total magnetization $M = mV$ (see the section "Lattice models"; [40]). Let $\beta$ be large enough to be in the two-phase region and, for a fixed $\alpha \in (0, 1)$, let

$$m = \alpha\, m^*(\beta) + (1-\alpha)\,(-m^*(\beta)) \\ = (1-2\alpha)\, m^*(\beta) \qquad [51]$$

that is, $m$ is in the vertical part of the diagram $m = m(\beta, h)$ at $\beta$ fixed (see Figure 4).

Fixing $m$ as in [51] does not yet determine the separation of the phases in two different regions; for this effect, it will be necessary to introduce some external cause favoring the occupation of a part of the volume by a single phase. Such an asymmetry can be obtained in at least two ways: through a weak uniform external field (in complete analogy with the gravitational field in the liquid–vapor transition) or through an asymmetric field acting only on boundary spins. The latter should have the same qualitative effect as the former, because in a phase transition region a boundary perturbation produces volume effects (see sections "Phase transitions and inequalities" and "Symmetry-breaking phase transitions"). From a mathematical point of view, it is simpler to use a boundary asymmetry to produce phase separations and the simplest geometry is obtained by considering $\pm$-cylindrical or $++$-cylindrical boundary conditions: this means $++$ or $\pm$ boundary conditions periodic in one direction (e.g., in Figure 2 imagine the right and left boundary identified after removing the boundary spins on them).

Spins adjacent to the bases of $\Omega$ act as symmetry-breaking external fields. The $++$-cylindrical boundary condition should favor the formation inside $\Omega$ of the positively magnetized phase; therefore, it will be natural to consider, in the canonical distribution, this boundary condition only when the total magnetization is fixed to be the spontaneous magnetization $m^*(\beta)$.

On the other hand, the $\pm$-boundary condition favors the separation of phases (positively magnetized phase near the top of $\Omega$ and negatively magnetized phase near the bottom). Therefore, it will be natural to consider the latter boundary condition in the case of a canonical distribution with magnetization $m = (1-2\alpha)m^*(\beta)$ with $0 < \alpha < 1$ ([51]). In the latter case, the positive phase can be expected to adhere to the top of $\Omega$ and to extend, in some sense to be discussed, up to a distance $O(L)$ from it; and then to change into the negatively magnetized pure phase.

To make the phenomenological description precise, consider the spin configurations $\boldsymbol{\sigma}$ through the associated sets of disjoint polygons (cf. the section "Symmetry-breaking phase transitions"). Fix the boundary conditions to be $++$ or $\pm$-cylindrical boundary conditions and note that polygons associated with a spin configuration $\boldsymbol{\sigma}$ are all closed and of two types: the ones of the first type, denoted $\gamma_1, \ldots, \gamma_n$, are polygons which do not encircle $\Omega$; the second type of polygons, denoted by the symbols $\lambda_\alpha$, are the ones which wind up, at least once, around $\Omega$.

So, a spin configuration $\boldsymbol{\sigma}$ will be described by a set of polygons; the statistical weight of a configuration $\boldsymbol{\sigma} = (\gamma_1, \ldots, \gamma_n, \lambda_1, \ldots, \lambda_h)$ is (cf. [45]):

$$e^{-2\beta J\left(\sum_i |\gamma_i| + \sum_j |\lambda_j|\right)} \qquad [52]$$

The reason why the contours $\lambda$ that go around the cylinder $\Omega$ are denoted by $\lambda$ (rather than by $\gamma$) is that they "look like" open contours (see the section "Symmetry-breaking phase transitions") if one forgets that the opposite sides of $\Omega$ have to be identified. In the case of the $\pm$-boundary conditions then the number of polygons of $\lambda$-type must be odd (hence $\neq 0$), while for the $++$-boundary condition the number of $\lambda$-type polygons must be even (hence it could be 0).

For more details, the reader is referred to Sinai (1991) and Gallavotti (1999).

## Separation and Coexistence of Phases

In the context of the geometric description of the spin configuration in the last section, consider the canonical distributions with $++$-cylindrical or the $\pm$-cylindrical boundary conditions and zero field: they will be denoted briefly as $\mu_{\beta,++}$, $\mu_{\beta,\pm}$, respectively. The following theorem (*Minlos–Sinai's theorem*) provided the foundations of the microscopic theory of coexistence: it is formulated in dimension $d=2$ but, modulo obvious changes, it holds for $d \geq 2$.

**Theorem** *For* $0 < \alpha < 1$ *fixed, let* $m = (1-2\alpha)\, m^*(\beta)$; *then for* $\beta$ *large enough a spin configuration* $\boldsymbol{\sigma} = (\gamma_1, \ldots, \gamma_n, \lambda_1, \ldots, \lambda_{2h+1})$ *randomly chosen with the distribution* $\mu_{\beta,\pm}$ *enjoys the properties (i)–(iv) below with a* $\mu_{\beta,\pm}$*-probability approaching* 1 *as* $\Omega \to \infty$:

(i) $\boldsymbol{\sigma}$ *contains only one contour of* $\lambda$*-type and*

$$||\lambda| - (1+\varepsilon(\beta))L| < o(L) \qquad [53]$$

*where* $\varepsilon(\beta) > 0$ *is a suitable (*$\alpha$*-independent) function of* $\beta$ *tending to zero exponentially fast as* $\beta \to \infty$.

(ii) *If* $\Omega_\lambda^+, \Omega_\lambda^-$ *denote respectively, the regions above and below* $\lambda$, *and* $|\Omega| \equiv V, |\Omega^+|, |\Omega^-|$ *are, respectively, the volumes of* $\Omega$, $\Omega^+$, $\Omega^-$ *then*

$$||\Omega_\lambda^+| - \alpha V| < \kappa(\beta)\, V^{3/4}$$
$$|\Omega_\lambda^-| - (1-\alpha)V| < \kappa(\beta)\, V^{3/4} \qquad [54]$$

*where* $\kappa(\beta) \xrightarrow[\beta\to\infty]{}$ *exponentially fast; the exponent* 3/4, *here and below, is not optimal.*

(iii) *If* $M_\lambda^+ = \sum_{x\in\Omega_\lambda^+} \sigma_x$ and $M_\lambda^- = \sum_{x\in\Omega_\lambda^-} \sigma_x$, *then*

$$|M_\lambda^+ - \alpha\, m^*(\beta)\, V| < \kappa(\beta)\, V^{3/4}$$
$$M_\lambda^- - (1-\alpha)\, m^*(\beta)\, V| < \kappa(\beta) V^{3/4} \qquad [55]$$

(iv) *If* $K_\gamma^\lambda(\boldsymbol{\sigma})$ *denotes the number of contours congruent to a given* $\gamma$ *and lying in* $\Omega_\lambda^+$ *then, simultaneously for all the shapes of* $\gamma$:

$$|K_\gamma^\lambda(\boldsymbol{\sigma}) - \rho(\gamma)\,\alpha\, V| \leq Ce^{-\beta J|\gamma|}V^{1/2}, \quad C>0 \qquad [56]$$

*where* $\rho(\gamma) \leq e^{-2\beta J|\gamma|}$ *is the same quantity as already mentioned in the text of the theorem of "Finite-volume effects". A similar result holds for the contours below* $\lambda$ (cf. *the comments on* [47]).

The above theorem not only provides a detailed and rather satisfactory description of the phase separation phenomenon, but it also furnishes a precise microscopic definition of the line of separation between the two phases, which should be naturally identified with the (random) line $\lambda$.

A similar result holds in the canonical distribution $\mu_{\beta,++,m^*(\beta)}$ where (i) is replaced by: no $\lambda$-type polygon is present, while (ii), (iii) become superfluous, and (iv) is modified in the obvious way. In other words, a typical configuration for the distribution the $\mu_{\beta,++,m^*(\beta)}$ has the same appearance as a typical configuration of the corresponding grand canonical ensemble with $(+)$-boundary condition (whose properties are described by the theorem given in the section "Beyond low temperatures (ferromagnetic Ising model")).

For more details, see Sinai (1991) and Gallavotti (1999).

## Phase Separation Line and Surface Tension

Continuing to refer to the nearest-neighbor Ising ferromagnet, the theorem of the last section means that, if $\beta$ is large enough, then the microscopic line $\lambda$, separating the two phases, is almost straight (since $\varepsilon(\beta)$ is small). The deviations of $\lambda$ from a straight line are more conveniently studied in the grand canonical distributions $\mu_\pm^0$ with boundary condition set to +1 in the upper half of $\partial\Omega$, vertical sites included, and to –1 in the lower half: this is illustrated in Figure 2 (see the section "Symmetry-breaking phase transitions"). The results can be converted into very similar results for grand canonical distributions with $\pm$-cylindrical boundary conditions of the last section.

Define $\lambda$ to be *rigid* if the probability that $\lambda$ passes through the center of the box $\Omega$ (i.e., 0) does not tend to 0 as $\Omega \to \infty$; otherwise, it is *not rigid*.

The notion of rigidity distinguishes between the possibilities for the line $\lambda$ to be "straight." The "excess" length $\varepsilon(\beta)L$ (see [53]) can be obtained in two ways: either the line $\lambda$ is essentially straight (in the geometric sense) with a few "bumps" distributed with a density of order $\varepsilon(\beta)$ or, otherwise, it is only locally straight and with an important part of the excess length being gained through a small bending on a large length scale. In three dimensions a similar phenomenon is possible. Rigidity of $\lambda$, or its failure, can in principle be investigated by optical means;

there can be interference of coherent light scattered by macroscopically separated surface elements of $\lambda$ only if $\lambda$ is rigid in the above sense.

It has been rigorously proved that, the line $\lambda$ is *not rigid* in dimension 2. And, at least at low temperature, the fluctuation of the middle point is of the order $O(\sqrt{L})$. In dimension 3 however, it has been shown that the surface $\lambda$ is *rigid* at low enough temperature.

A deeper analysis is needed to study the shape of the separation surface under other conditions, for example, with + boundary conditions in a canonical distribution with magnetization intermediate between $\pm m^*(\beta)$. It involves, as a prerequisite, the definition and many properties of the surface tension between the two phases. Here only the definition of surface tension in the case of $\pm$-boundary conditions in the two-dimensional case will be mentioned. If $Z^{++}(\Omega, m^*(\beta))$ and $Z^{+-}(\Omega, m)$ are, respectively, the canonical partition functions for the ++- and $\pm$-cylindrical boundary conditions the tension $\tau(\beta)$ is defined as

$$\beta\tau(\beta) = -\lim_{\Omega\to\infty} \frac{1}{L} \log \frac{Z^{+-}(\Omega, m)}{Z^{++}(\Omega, m^*(\beta))}$$

The limit can be shown to be $\alpha$-independent for $\beta$ large enough: the definition and its justification is based on the microscopic geometric description in the section "Geometry of phase co-existence." The definition can be naturally extended to higher dimension (and to more general non-nearest-neighbor models). If $d = 2$, the tension $\tau$ can be exactly computed at all temperatures below criticality and is $\beta\tau(\beta) = 2\beta J + \log \tanh \beta J$.

More remarkably, the definition can be extended to define the surface tension $\tau(\beta, \boldsymbol{n})$ in the "direction $\boldsymbol{n}$," that is, when the boundary conditions are such that the line of separation is in the average orthogonal to the unit vector $\boldsymbol{n}$. In this way, if $d = 2$ and $\alpha \in (0, 1)$ is fixed, it can be proved that at low enough temperature the canonical distribution with + boundary conditions and intermediate magnetization $m = (1 - 2\alpha)m^*(\beta)$ has typical configurations containing a spin − region of area $\sim\alpha V$; furthermore, if the container is rescaled to size $L = 1$, the region will have a limiting shape filling an area $\alpha$ bounded by a smooth curve whose form is determined by the classical macroscopic *Wulff's theory* of the shape of crystals in terms of the surface tension $\tau(\boldsymbol{n})$.

An interesting question remains open in the three-dimensional case: it is conceivable that the surface, although rigid at low temperature, might become "loose" at a temperature $\widetilde{T}_c$ smaller than the critical temperature $T_c$ (the latter being defined as the highest temperature below which there are at least two pure phases). The temperature $\widetilde{T}_c$, whose existence is rather well established in numerical experiments, would be called the "roughening transition" temperature. The rigidity of $\lambda$ is connected with the existence of translationally non-invariant equilibrium states. The latter exist in dimension $d = 3$, but not in dimension $d = 2$, where the discussed nonrigidity of $\lambda$, established all the way to $T_c$, provides the intuitive reason for the absence of non-translation-invariant states. It has been shown that in $d = 3$ the roughening temperature $\widetilde{T}_c(\beta)$ necessarily cannot be smaller than the critical temperature of the two-dimensional Ising model with the same coupling.

Note that existence of translationally noninvariant equilibrium states is not necessary for the description of coexistence phenomena. The theory of the nearest-neighbor two-dimensional Ising model is a clear proof of this statement.

The reader is referred to Onsager (1944), van Beyeren (1975), Sinai (1991), Miracle-Solé (1995), Pfister and Velenik (1999), and Gallavotti (1999) for more details.

## Critical Points

Correlation functions for a system with short-range interactions and in an equilibrium state (which is a pure phase) have cluster properties (see [22]): their physical meaning is that in a pure phase there is independence between fluctuations occurring in widely separated regions. The simplest cluster property concerns the "pair correlation function," that is, the probability density $\rho(\boldsymbol{q}_1, \boldsymbol{q}_2)$ of finding particles at points $\boldsymbol{q}_1, \boldsymbol{q}_2$ independently of where the other particles may happen to be (see [23]). In the case of spin systems, the pair correlation $\rho(\boldsymbol{q}_1, \boldsymbol{q}_2) = \langle \sigma_{q_1} \sigma_{q_2} \rangle$ will be considered. The pair correlation of a translation-invariant equilibrium state has a cluster property ([22], [42]), if

$$|\rho(\boldsymbol{q}_1, \boldsymbol{q}_2) - \rho^2| \xrightarrow[|q_1 - q_2| \to \infty]{} 0 \qquad [57]$$

where $\rho$ is the probability density for finding a particle at $\boldsymbol{q}$ (i.e., the physical density of the state) or $\rho = \langle \sigma_q \rangle$ is the average of the value of the spin at $\boldsymbol{q}$ (i.e., the magnetization of the state).

A general definition of critical point is a point $c$ in the space of the parameters characterizing equilibrium states, for example, $\beta, \lambda$ in grand canonical distributions, $\beta$, $v$ in canonical distributions, or $\beta$, $h$ in the case of lattice spin systems in a grand canonical

distribution. In systems with short-range interaction (i.e., with $\varphi(\boldsymbol{r})$ vanishing for $|\boldsymbol{r}|$ large enough) the point $c$ is a critical point if the pair correlation tends to 0 (see [57]), slower than exponential (e.g., as a power of the distance $|\boldsymbol{r}| = |\boldsymbol{q}_1 - \boldsymbol{q}_2|$).

A typical example is the two-dimensional Ising model on a square lattice and with nearest-neighbor ferromagnetic interaction of size $J$. It has a single critical point at $\beta = \beta_c$, $h = 0$ with $\sinh 2\beta_c J = 1$. The cluster property is that $\langle \sigma_x \sigma_y \rangle - \langle \sigma_x \rangle \langle \sigma_y \rangle \xrightarrow[|x-y|\to\infty]{} 0$ as

$$A_+(\beta)\frac{e^{-\kappa(\beta)|x-y|}}{\sqrt{|x-y|}}, \qquad A_-(\beta)\frac{e^{-\kappa(\beta)|x-y|}}{|x-y|^2}$$
$$A_c\frac{1}{|x-y|^{1/4}}, \qquad [58]$$

for $\beta < \beta_c$, $\beta > \beta_c$, or $\beta = \beta_c$, respectively, where $A_\pm(\beta)$, $A_c$, $\kappa(\beta) > 0$. The properties [58] stem from the exact solution of the model.

At the critical point, several interesting phenomena occur: the lack of exponential decay indicates lack of a length scale over which really distinct phenomena can take place, and properties of the system observed at different length scales are likely to be simply related by suitable scaling transformations. Many efforts have been dedicated at finding ways of understanding quantitatively the scaling properties pertaining to different observables. The result has been the development of the *renormalization group* approach to critical phenomena (cf. the section "Renormalization group"). The picture that emerges is that the closer the critical point is the larger becomes the maximal scale of length below which scaling properties are observed. For instance, in a lattice spin system in zero field the magnetization $M|\Lambda|^{-a}$ in a box $\Lambda \subset \Omega$ should have essentially the same distribution for all $\Lambda$'s with side $< l_0(\beta)$ and $l_0(\beta) \to \infty$ as $\beta \to \beta_c$, provided $a$ is suitably chosen. The number $a$ is called a *critical exponent*.

There are several other "critical exponents" that can be defined near a critical point. They can be associated with singularities of the thermodynamic function or with the behavior of the correlation functions involving joint densities at two or more than two points. As an example, consider a lattice spin system: then the "$2n$–spins correlation" $\langle \sigma_0 \sigma_{\xi_1} \ldots \sigma_{\xi_{2n-1}} \rangle_c$ could behave proportionally to $\chi_{2n}(0, \xi_1, \ldots, \xi_{2n-1}), n = 1, 2, 3, \ldots$, for a suitable family of homogeneous functions $\chi_n$, of some degree $\omega_{2n}$, of the coordinates $(\xi_1, \ldots, \xi_{2n-1})$ at east when the reciprocal distances are large but $< l_0(\beta)$ and

$$l_0(\beta) = \text{const.}(\beta - \beta_c)^{-\nu} \xrightarrow[\beta\to\beta_0]{} \infty$$

This means that if $\xi_i$ are regarded as points in $\mathbb{R}^d$ there are functions $\chi_{2n}$ such that

$$\chi_{2n}\left(0, \frac{\xi_1}{\lambda}, \ldots, \frac{\xi_{2n-1}}{\lambda}\right) = \lambda^{\omega_{2n}} \chi_{2n}(0, \xi_1, \ldots, \xi_{2n-1})$$
$$0 < \lambda \in \mathbb{R} \qquad [59]$$

and $\langle \sigma_0 \sigma_{\xi_1} \ldots \sigma_{\xi_{2n-1}} \rangle \propto \chi_{2n}(0, \xi_1, \ldots, \xi_{2n-1})$ if $1 \ll |x_i - x_j| \ll l_0(\beta)$. The numbers $\omega_{2n}$ define a sequence of critical exponents.

Other critical exponents can be associated with approaching the critical point along other directions (e.g., along $h \to 0$ at $\beta = \beta_c$). In this case, the length up to which there are scaling phenomena is $l_0(h) = \ell_o h^{-\bar{\nu}}$. Further, the magnetization $m(h)$ tends to 0 as $h \to 0$ at fixed $\beta = \beta_c$ as $m(h) = m_0 h^{1/\delta}$ for $\delta > 0$.

None of the feautres of critical exponents is known rigorously, including their existence. An exception is the case of the two–dimensional nearest-neighbor Ising ferromagnet where some exponents are known exactly (e.g., $\omega_2 = 1/4$, $\omega_{2n} = n\omega_2$, or $\nu = 1$, while $\delta$, $\bar{\nu}$ are not rigorously known). Nevertheless, for Ising ferromagnets (not even nearest-neighbor but, as always here, finite-range) in all dimensions, all of the exponents mentioned are conjectured to be the same as those of the nearest-neighbor Ising ferromagnet. A further exception is the derivation of rigorous relations between critical exponents and, in some cases, even their values under the assumption that they exist.

**Remark** Naively it could be expected that in a pure state in zero field with $\langle \sigma_x \rangle = 0$ the quantity $s = |\Lambda|^{-1/2} \sum_{x\in\Lambda} \sigma_x$, if $\Lambda$ is a cubic box of side $\ell$, should have a probability distribution which is Gaussian, with dispersion $\lim_{\Lambda\to\infty}\langle s^2 \rangle$. This is "usually true," but not always. Properties [58] show that in the $d = 2$ ferromagnetic nearest-neighbor Ising model, $\langle s^2 \rangle$ diverges proportionally to $\ell^{2-\frac{1}{4}}$ so that the variable $s$ cannot have the above Gaussian distribution. The variable $S = |\Lambda|^{-7/8} \sum_{x\in\Lambda} \sigma_x$ will have a finite dispersion: however, there is no reason that it should be Gaussian. This makes clear the great interest of a fluctuation theory and its relevance for the critical point studies (see the next two sections).

For more details, the reader is referred to Onsager (1944), Domb and Green (1972), McCoy and Wu (1973), and Aizenman (1982).

## Fluctuations

As it appears from the discussion in the last section, fluctuations of observables around their averages have interesting properties particularly at critical points. Of particular interest are observables that

are averages, over large volumes $\Lambda$, of local functions $F(x)$ on phase space: this is so because macroscopic observables often have this form. For instance, given a region $\Lambda$ inside the system container $\Omega$, $\Lambda \subset \Omega$, consider a configuration $x = (\boldsymbol{P}, \boldsymbol{Q})$ and the number of particles $N_\Lambda = \sum_{q \in \Lambda} 1$ in $\Lambda$, or the potential energy $\boldsymbol{\Phi}_\Lambda = \sum_{(q,q') \in \Lambda} \varphi(\boldsymbol{q} - \boldsymbol{q}')$ or the kinetic energy $K_\Lambda = \sum_{q \in \Lambda} (1/2m)\boldsymbol{p}^2$. In the case of lattice spin systems, consider a configuration $\boldsymbol{\sigma}$ and, for instance, the magnetization $M_\Lambda = \sum_{i \in \Lambda} \sigma_i$ in $\Lambda$. Label the above four examples by $\alpha = 1, \ldots, 4$.

Let $\mu_\alpha$ be the probability distribution describing the equilibrium state in which the quantities $X_\Lambda$ are considered; let $x_\Lambda = \langle X_\Lambda / |\Lambda| \rangle_{\mu_\alpha}$ and $p \stackrel{\text{def}}{=} (X_\Lambda - x_\Lambda)/|\Lambda|$. Then typical properties of fluctuations that should be investigated are ($\alpha = 1, \ldots, 4$):

1. for all $\delta > 0$ it is $\lim_{\Lambda \to \infty} \mu_\alpha(|p| > \delta) = 0$ (*law of large numbers*);
2. there is $D_\alpha > 0$ such that

$$\mu(p\sqrt{|\Lambda|} \in [a,b]) \underset{\Lambda \to \infty}{\longrightarrow} \int_a^b \frac{\mathrm{d}z}{\sqrt{2\pi D_\alpha}} \mathrm{e}^{-z^2/2D_\alpha}$$

   (*central limit law*); and
3. there is an interval $I_\alpha = (p^*_{\alpha,-}, p^*_{\alpha,+})$ and a concave function $F_\alpha(p)$, $p \in I$, such that if $[a,b] \subset I$ then

$$\frac{1}{|\Lambda|} \log \mu(p \in [a,b]) \underset{\Lambda \to \infty}{\longrightarrow} \max_{p \in [a,b]} F_\alpha(p)$$

   (*large deviations law*).

The law of large numbers provides the certainty of the macroscopic values; the central limit law controls the small fluctuations (of order $\sqrt{|\Lambda|}$) of $X_\Lambda$ around its average; and the large deviations law concerns the fluctuations of order $|\Lambda|$.

The relations (1)–(3) above are not always true: they can be proved under further general assumptions if the potential $\varphi$ satisfies [14] in the case of particle systems or if $\sum_q |\varphi(q)| < \infty$ in the case of lattice spin systems. The function $F_\alpha(p)$ is defined in terms of the thermodynamic limits of suitable thermodynamic functions associated with the equilibrium state $\mu_\alpha$. The further assumption is, essentially in all cases, that a suitable thermodynamic function in terms of which $F_\alpha(p)$ will be expressed is smooth and has a nonvanishing second derivative.

For the purpose of a simple concrete example, consider a lattice spin system of Ising type with energy $-\sum_{x,y \in \Omega} \varphi(x-y)\sigma_x\sigma_y - \sum_x h\sigma_x$ and the fluctuations of the magnetization $M_\Lambda = \sum_{x \in \Lambda} \sigma_x$, $\Lambda \subset \Omega$, in the grand canonical equilibrium states $\mu_{h,\beta}$.

Let the free energy be $\beta f(\beta, h)$ (see [41]), let $m = m(h) \stackrel{\text{def}}{=} \langle M_\Lambda / |\Lambda| \rangle$ and let $h(m)$ be the inverse function of $m(h)$. If $p = M_\Lambda / |\Lambda|$ the function $F(p)$ is given by

$$F(p) = \beta(f(\beta, h(p)) - f(\beta, h) - \partial_h f(\beta, h)(h(p) - h)) \quad [60]$$

then a quite general result is:

**Theorem** *The relations* (1)–(3) *hold if the potential satisfies* $\sum_x |\varphi(x)| < \infty$ *and if* $F(p)$ [60] *is smooth and* $F''(p) \neq 0$ *in open intervals around those in which* $p$ *is considered, that is, around* $p = 0$ *for the law of large numbers and for the central limit law or in an open interval containing* $a, b$ *for the case of the large deviations law.*

In the cases envisaged, the theory of equivalence of ensembles implies that the function $F$ can also be computed via thermodynamic functions naturally associated with other equilibrium ensembles. For instance, instead of the grand canonical $f(\beta, h)$, one could consider the canonical $\beta g(\beta, m)$ (see [41]), then

$$F(p) = -\beta(g(\beta, p) - g(\beta, m) - \partial_m g(\beta, m)(p - m)) \quad [61]$$

It has to be remarked that there should be a strong relation between the central limit law and the law of large deviations. Setting aside stating the conditions for a precise mathematical theorem, the statement can be efficiently illustrated in the case of a ferromagnetic lattice spin system and with $\Lambda \equiv \Omega$, by showing that the law of large deviations in small intervals, around the average $m(h_0)$, at a value $h_0$ of the external field, is implied by the validity of the central limit law for all values of $h$ near $h_0$ and vice versa (here $\beta$ is fixed). Taking $h_0 = 0$ (for simplicity), the heuristic reasons are the following. Let $\mu_{h,\Omega}$ be the grand canonical distribution in external field $h$. Then:

1. The probability $\mu_{h,\Omega}(p \in dp)$ is proportional, by definition, to $\mu_{0,\Omega}(p \in \mathrm{d}p)\mathrm{e}^{-\beta h p |\Omega|}$. Hence, if the central limit law holds for all $h$ near $h_0 = 0$, there will exist two functions $m(h)$ and $D(h) > 0$, defined for $h$ near $h_0 = 0$, with $m(0) = 0$ and

$$\mu_0(p \in \mathrm{d}p)\mathrm{e}^{-\beta h p^{|\Omega|}} = \text{const.} \exp\left(-|\Omega| \frac{(p - m(h))^2}{2D(h)} + o(\Omega)\right) \mathrm{d}p \quad [62]$$

2. There is a function $\zeta(m)$ such that $\partial_m \zeta(m(h)) = \beta h$ and $\partial_m^2 \zeta(m(h)) = D(h)^{-1}$. (This is obtained by noting that, given $D(h)$, the differential equation $\partial_m \beta h = D(h)^{-1}$ with the initial value $h(0) = 0$ determines the function $h(m)$; therefore, $\zeta(m)$ is determined by a second integration, from $\partial_m \zeta(m) = \beta h(m)$.

It then follows, heuristically, that the probability of $p$ in zero field has the form const. $e^{\zeta(p)|\Omega|}\,dp$ so that the probability that $p \in [a,b]$ will be const $\exp(|\Omega| \max_{p\in[a,b]} \zeta(p))$.

Conversely, the large deviations law for $p$ at $h=0$ implies the validity of the central limit law for the fluctuations of $p$ in all small enough fields $h$: this simply arises from the function $F(p)$ having a negative second derivative.

This means that there is a "duality" between central limit law and large deviation law or that the law of large deviations is a "global version" of the central limit law, in the sense that:

1. if the central limit law holds for h in an interval around $h_0$ then the fluctuations of the magnetization at field $h_0$ satisfy a large deviation law in a small enough interval $J$ around $m(h_0)$; and
2. if a large deviation law is satisfied in an interval around $h_0$ then the central limit law holds for the fluctuations of magnetization around its average in all fields h with $h - h_0$ small enough.

Going beyond the heuristic level in establishing the duality amounts to giving a precise meaning to "small enough" and to discuss which properties of $m(h)$ and $D(h)$, or $F(p)$ are needed to derive properties (1), (2).

For purposes of illustration consider the Ising model with ferromagnetic short range interaction $\varphi$: then the central limit law holds for all $h$ if $\beta$ is small enough and, under the same condition on $\beta$, the large deviations law holds for all $h$ and all intervals $[a,b] \subset (-1,1)$. If $\beta$ is not small then the condition $h \neq 0$ has to be added. Hence, the conditions are fairly weak and the apparent exceptions concern the value $h=0$ and $\beta$ not small where the statements may become invalid because of possible phase transitions.

In presence of phase transitions, the law of large numbers, the central limit law, and law of large deviations should be reformulated. Basically, one has to add the requirement that fluctuations are considered in pure phases and change, in a natural way, the formulation of the laws. For instance, the large fluctuations of magnetization in a pure phase of the Ising model in zero field and large $\beta$ (i.e., in a state obtained as limit of finite-volume states with + or − boundary conditions) in intervals $[a,b]$ which do not contain the average magnetization $m^*$ are not necessarily exponentially small with the size of $|\Lambda|$: if $[a,b] \subset [-m^*, m^*]$ they are exponentially small but only with the size of the surface of $\Lambda$ (i.e., with $|\Lambda|^{(d-1)/d}$) while they are exponentially small with the volume if $[a,b] \cap [-m^*, m^*] = \emptyset$.

The discussion of the last section shows that at the critical point the nature of the large fluctuations is also expected to change: no central limit law is expected to hold in general because of the example of [58] with the divergence of the average of the normal second moment of the magnetization in a box as the side tends to $\infty$.

For more details the reader is referred to Olla (1987).

## Renormalization Group

The theory of fluctuations just discussed concerns only fluctuations of a single quantity. The problem of joint fluctuations of several quantities is also interesting and in fact led to really new developments in the 1970s. It is necessary to restrict attention to rather special cases in order to illustrate some ideas and the philosophy behind the approach. Consider, therefore, the equilibrium distribution $\mu_0$ associated with one of the classical equilibrium ensembles. To fix the ideas we consider the equilibrium distribution of an Ising energy function $\beta H_0$, having included the temperature factor in the energy: the inclusion is done because the discussion will deal with the properties of $\mu_0$ as a function of $\beta$. It will also be assumed that the average of each spin is zero ("no magnetic field," see [39] with $h=0$). Keeping in mind a concrete case, imagine that $\beta H_0$ is the energy function of the nearest-neighbor Ising ferromagnet in zero field.

Imagine that the volume $\Omega$ of the container has periodic boundary conditions and is very large, ideally infinite. Define the family of blocks $k\boldsymbol{\xi}$, parametrized by $\boldsymbol{\xi} \in \mathbb{Z}^d$ and with $k$ an integer, consisting of the lattice sites $\boldsymbol{x} = \{k\xi_i \le x_i < (k+1)\xi_i\}$. This is a lattice of cubic blocks with side size $k$ that will be called the "$k$-rescaled lattice."

Given $\alpha$, the quantities $m_{\boldsymbol{\xi}} = k^{-\alpha d} \sum_{\boldsymbol{x}\in k\boldsymbol{\xi}} \sigma_{\boldsymbol{x}}$ are called the block spins and define the map $R^*_{\alpha,k}\mu_0 = \mu_k$ transforming the initial distribution on the original spins into the distribution of the block spins. Note that if the initial spins have only two values $\sigma_{\boldsymbol{x}} = \pm 1$, the block spins take values between $-k^d/k^{\alpha d}$ and $k^d/k^{\alpha d}$ at steps of size $2/k^{\alpha d}$. Furthermore, the map $R^*_{\alpha,k}$ makes sense independently of how many values the initial spins can assume, and even if they assume a continuum of values $S_{\boldsymbol{x}} \in \mathbb{R}$.

Taking $\alpha = 1$ means, for $k$ large, looking at the probability distribution of the joint large fluctuations in the blocks $k\boldsymbol{\xi}$. Taking $\alpha = 1/2$ corresponds to studying a joint central limit property for the block variables.

Considering a one-parameter family of initial distributions $\mu_0$ parametrized by a parameter $\beta$

(that will be identified with the inverse temperature), typically there will be a unique value $\alpha(\beta)$ of $\alpha$ such that the joint fluctuations of the block variables admit a limiting distribution,

$$\text{prob}_k(m_\xi \in [a_\xi, b_\xi], \boldsymbol{\sigma} \in \Lambda) \underset{k\to\infty}{\longrightarrow} \int_{\{a_\xi\}}^{\{b_\xi\}} g_\Lambda((S_\xi)_{\xi\in\Lambda}) \prod_{\xi\in\Lambda} \mathrm{d}S_\xi \qquad [63]$$

for some distribution $g_\Lambda(z)$ on $\mathbb{R}^\Lambda$.

If $\alpha > \alpha(\beta)$, the limit will then be $\prod_{\xi\in\Lambda}\delta(S_\xi)\,\mathrm{d}S_\xi$, or if $\alpha < \alpha(\beta)$ the limit will not exist (because the block variables will be too large, with a dispersion diverging as $k \to \infty$).

It is convenient to choose as sequence of $k \to \infty$ the sequence $k = 2^n$ with $n = 0, 1, \ldots$ because in this way it is $R^*_{\alpha,k} \equiv R^{*n}_{\alpha,1}$ and the limits $k \to \infty$ along the sequence $k = 2^n$ can be regarded as limits on a sequence of iterations of a map $R^*_{\alpha,1}$ acting on the probability distributions of generic spins $S_x$ on the lattice $\mathbb{Z}^d$ (the sequence $3^n$ would be equally suited).

It is even more convenient to consider probability distributions that are expressed in terms of energy functions $H$ which generate, in the thermodynamic limit, a distribution $\mu$: then $R^*_{\alpha,1}$ defines an action $R_\alpha$ on the energy functions so that $R_\alpha H = H'$ if $H$ generates $\mu$, $H'$ generates $\mu'$ and $R^*_{\alpha,1}\mu = \mu'$. Of course, the energy function will be more general than [39] and at least a form like $\delta U$ in [49] has to be admitted.

In other words, $R_\alpha$ gives the result of the action of $R^*_{\alpha,1}$ expressed as a map acting on the energy functions. Its iterates also define a semigroup which is called the block spin renormalization group.

While the map $R^*_{\alpha,1}$ is certainly well defined as a map of probability distributions into probability distributions, it is by no means clear that $R_\alpha$ is well defined as a map on the energy functions. Because, if $\mu$ is given by an energy function, it is not clear that $R^*_{\alpha,1}\mu$ is such.

A remarkable theorem can be (easily) proved when $R^*_{\alpha,1}$ and its iterates act on initial $\mu_0$'s which are equilibrium states of a spin system with short-range interactions and at high temperature ($\beta$ small). In this case, if $\alpha = 1/2$, the sequence of distributions $R^{*n}_{1/2,1}\mu_0(\beta)$ admits a limit which is given by a product of independent Gaussians:

$$\text{prob}_k(m_\xi \in [a_\xi, b_\xi], \boldsymbol{\sigma} \in \Lambda) \underset{k\to\infty}{\longrightarrow} \int_{\{a_\xi\}}^{\{b_\xi\}} \prod_{\xi\in\Lambda} \exp\left(-\frac{1}{2D(\beta)} S_\xi^2\right) \prod_{\xi\in\Lambda} \frac{\mathrm{d}S_\xi}{\sqrt{2\pi D(\beta)}} \qquad [64]$$

Note that this theorem is stated without even mentioning the renormalization maps $R^n_{1/2}$: it can nevertheless be interpreted as stating that

$$R^n_{1/2}\beta H_0 \underset{n\to\infty}{\longrightarrow} \sum_{\xi\in\mathbb{Z}^d} \frac{1}{2D(\beta)} S_\xi^2 \qquad [65]$$

but the interpretation is not rigorous because [64] does not state require that $R^n_{1/2}H_0(\beta)$ makes sense for $n \geq 1$. It states that at high temperature block spins have normal independent fluctuations: it is therefore an extension of the central limit law.

There are a few cases in which the map $R_\alpha$ can be rigorously shown to be well defined at least when acting on special equilibrium states like the high-temperature lattice spin systems: but these are exceptional cases of relatively little interest.

Nevertheless, there is a vast literature dealing with approximate representations of the map $R_\alpha$. The reason is that, assuming not only its existence but also that it has the properties that one would normally expect to hold for a map acting on a finite dimensional space, it follows that a number of consequences can be drawn; quite nontrivial ones as they led to the first theory of the critical point that goes beyond the van der Waals theory discribed in the section "van der Waals theory."

The argument proceeds essentially as follows. At the critical point, the fluctuations are expected to be anomalous (cf. the last remark in the section "Critical points") in the sense that $\langle(\sum_{x\in\Lambda}\sigma_\xi/\sqrt{|\Lambda|})^2\rangle$ will tend to $\infty$, because $\alpha = 1/2$ does not correspond to the right fluctuation scale of $\sum_{\xi\in\Lambda}\sigma_\xi$, signaling that $R^{*n}_{1/2,1}\mu_0(\beta_c)$ will not have a limit but, possibly, there is $\alpha_c > 1/2$ such that $R^{*n}_{\alpha_c,1}\mu_0(\beta_c)$ converges to a limit in the sense of [63]. In the case of the critical nearest-neighbor Ising ferromagnetic $\alpha_c = 7/8$ (see ending remark in the section "Critical points"). Therefore, if the map $R^*_{\alpha_c,1}$ is considered as acting on $\mu_0(\beta)$, it will happen that for all $\beta < \beta_c$, $R^{*n}_{\alpha_c,1}\mu_0(\beta_c)$ will converge to a trivial limit $\prod_{\xi\in\Lambda}\delta(S_\xi)\,\mathrm{d}S_\xi$ because the value $\alpha_c$ is greater than 1/2 while normal fluctuations are expected.

If the map $R_{\alpha_c}$ can be considered as a map on the energy functions, this says that $\prod_{\xi\in\Lambda}\delta(S_\xi)\,\mathrm{d}S_\xi$ is a "(trivial) fixed point of the renormalization group" which "attracts" the energy functions $\beta H_0$ corresponding to the high-temperature phases.

The existence of the critical $\beta_c$ can be associated with the existence of a *nontrivial fixed point* $H^*$ for $R_{\alpha_c}$ which is hyperbolic with just one Lyapunov exponent $\lambda > 1$; hence, it has a stable manifold of codimension 1. Call $\mu^*$ the probability distribution corresponding to $H^*$.

The migration towards the trivial fixed point for $\beta < \beta_c$ can be explained simply by the fact that for

such values of $\beta$ the initial energy function $\beta H_0$ is outside the stable manifold of the nontrivial fixed point and under application of the renormalization transformation $R^n_{\alpha_c}$, $\beta H_0$ migrates toward the trivial fixed point, which is attractive in all directions.

By increasing $\beta$, it may happen that, for $\beta = \beta_c$, $\beta H_0$ crosses the stable manifold of the nontrivial fixed point $H^*$ for $R_{\alpha_c}$. Then $R^n_{\alpha_c}\beta_c H_0$ will no longer tend to the trivial fixed point but it will tend to $H^*$: this means that the block spin variables will exhibit a completely different fluctuation behavior. If $\beta$ is close to $\beta_c$, the iterations of $R_{\alpha_c}$ will bring $R^n_{\alpha_c}\beta H_0$ close to $H^*$, only to be eventually repelled along the unstable direction reaching a distance from it increasing as $\lambda^n|\beta - \beta_c|$.

This means that up to a scale length $O(2^{n(\beta)})$ lattice units with $\lambda^{n(\beta)}|\beta - \beta_c| = 1$ (i.e., up to a scale $O(|\beta - \beta_c|^{-\log_2 \lambda})$), the fluctuations will be close to those of the fixed point distribution $\mu^*$, but beyond that scale they will come close to those of the trivial fixed point: to see them the block spins would have to be normalized with index $\alpha = 1/2$ and they would appear as uncorrelated Gaussian fluctuations (cf. [64], [65]).

The next question concerns finding the nontrivial fixed points, which means finding the energy functions $H^*$ and the corresponding $\alpha_c$ which are fixed points of $R_{\alpha_c}$. If the above picture is correct, the distributions $\mu^*$ corresponding to the $H^*$ would describe the critical fluctuations and, if there was only one choice, or a limited number of choices, of $\alpha_c$ and $H^*$ this would open the way to a universality theory of the critical point hinted already by the "primitive" results of van der Waals' theory.

The initial hope was, perhaps, that there would be a very small number of critical values $\alpha_c$ and $H^*$ possible: but it rapidly faded away leaving, however, the possibility that the critical fluctuations could be classified into universality classes. Each class would contain many energy functions which, upon iterated actions of $R_{\alpha_c}$, would evolve under the control of the trivial fixed point (always existing) for $\beta$ small while, for $\beta = \beta_c$, they would be controlled, instead, by a nontrivial fixed point $H^*$ for $R_{\alpha_c}$ with the same $\alpha_c$ and the same $H^*$. For $\beta < \beta_c$, a "resolution" of the approach to the trivial fixed point would be seen by considering the map $R_{1/2}$ rather than $R_{\alpha_c}$ whose iterates would, however, lead to a Gaussian distribution like [64] (and to a limit energy function like [65]).

The picture is highly hypothetical: but it is the first suggestion of a mechanism leading to critical points with the character of universality and with exponents different from those of the van der Waals theory or, for ferromagnets on a lattice, from those of its lattice version (the *Curie–Weiss theory*). Furthermore, accepting the approximations (e.g., the Wilson–Fisher $\varepsilon$-expansion) that allow one to pass from the well-defined $R^*_{\alpha,1}$ to the action of $R_\alpha$ on the energy functions, it is possible to obtain quite unambiguously values for $\alpha_c$ and expressions for $H^*$ which are associated with the action of $R_{\alpha_c}$ on various classes of models.

For instance, it can lead to conclude that the critical behavior of all ferromagnetic finite-range lattice spin systems (with energy functions given by [39]) have critical points controlled by the same $\alpha_c$ and the same nontrivial fixed point: this property is far from being mathematically proved, but it is considered a major success of the theory. One has to compare it with van der Waals' critical point theory: *for the first time*, an approximation scheme has led, even though under approximations not fully controllable, to computable critical exponents which are not equal to those of the van der Waals theory.

The renormalization group approach to critical phenomena has many variants, depending on which kind of fluctuations are considered and on the models to which it is applied. In statistical mechanics, there are a few mathematically complete applications: certain results in higher dimensions, theory of dipole gas in $d = 2$, hierarchical models, some problems in condensed matter and in statistical mechanics of lattice spins, and a few others. Its main mathematical successes have occured in various related fields where not only the philosophy described above can be applied but it leads to renormalization transformations that can be defined precisely and studied in detail: for example, constructive field theory, KAM theory of quasiperiodic motions, and various problems in dynamical systems.

However, the applications always concern special cases and in each of them the general picture of the trivial–nontrivial fixed point dichotomy appears realized but without being accompanied, except in rare cases (like the hierarchical models or the universality theory of maps of the interval), by the full description of stable manifold, unstable direction, and action of the renormalization transformation on objects other than the one of immediate interest (a generality which looks often an intractable problem, but which also turns out not to be necessary).

In the renormalization group context, mathematical physics has played an important role also by providing clear evidence that universality classes could not be too few: this was shown by the numerous exact solutions after Onsager's solution of the nearest-neighbor Ising ferromagnet: there are in fact several lattice models in $d = 2$ that exhibit critical points with some critical exponents exactly computable and that depend continuously on the models parameters.

For more details, we refer the reader to McCoy and Wu (1973), Baxter (1982), Bleher and Sinai (1975), Wilson and Fisher (1972), Gawedzky and Kupiainen (1983, 1985), Benfatto and Gallavotti (1995), and Mastropietro (2004).

## Quantum Statistics

Statistical mechanics is extended to assemblies of quantum particles rather straightforwardly. In the case of $N$ identical particles, the observables are operators $O$ on the Hilbert space

$$\mathcal{H}_N = L_2(\Omega)^N_\alpha \quad \text{or} \quad \mathcal{H}_N = (L_2(\Omega) \otimes \mathcal{C}^2)^N_\alpha$$

where $\alpha = +, -$, of the symmetric ($\alpha = +$, bosonic particles) or antisymmetric ($\alpha = -$, fermionic particles) functions $\psi(\boldsymbol{Q})$, $\boldsymbol{Q} = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N)$, of the position coordinates of the particles or of the position and spin coordinates $\psi(\boldsymbol{Q}, \boldsymbol{\sigma})$, $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_N)$, normalized so that

$$\int |\psi(\boldsymbol{Q})|^2 \, \mathrm{d}\boldsymbol{Q} = 1 \quad \text{or} \quad \sum_{\boldsymbol{\sigma}} \int |\psi(\boldsymbol{Q}, \boldsymbol{\sigma})|^2 \, \mathrm{d}\boldsymbol{Q} = 1$$

here only $\sigma_j = \pm 1$ is considered. As in classical mechanics, a state is defined by the average values $\langle O \rangle$ that it attributes to the observables.

Microcanonical, canonical, and grand canonical ensembles can be defined quite easily. For instance, consider a system described by the Hamiltonian ($\hbar$ = Planck's constant)

$$H_N = -\frac{\hbar^2}{2m} \sum_{j=1}^{N} \Delta_{q_j} + \sum_{j<j'} \varphi(\boldsymbol{q}_j - \boldsymbol{q}_{j'}) + \sum_j w(\boldsymbol{q}_j) \stackrel{\text{def}}{=} K + \boldsymbol{\Phi} \qquad [66]$$

where periodic boundary conditions are imagined on $\Omega$ and $w(\boldsymbol{q})$ is periodic, smooth potential (the side of $\Omega$ is supposed to be a multiple of the periodic potential period if $w \neq 0$). Then a canonical equilibrium state with inverse temperature $\beta$ and specific volume $v = V/N$ attributes to the observable $O$ the average value

$$\langle O \rangle \stackrel{\text{def}}{=} \frac{\operatorname{tr} \mathrm{e}^{-\beta H_N} O}{\operatorname{tr} \mathrm{e}^{-\beta H_N}} \qquad [67]$$

Similar definitions can be given for the grand canonical equilibrium states.

Remarkably, the ensembles are orthodic and a "heat theorem" (see the section "Heat theorem and ergodic hypothesis") can be proved. However, "equipartition" does not hold: that is, $\langle K \rangle \neq (d/2)N\beta^{-1}$, although $\beta^{-1}$ is still the integrating factor of $\mathrm{d}U + p\,\mathrm{d}V$ in the heat theorem; hence, $\beta^{-1}$ continues to be proportional to temperature.

Lack of equipartition is important, as it solves paradoxes that arise in classical statistical mechanics applied to systems with infinitely many degrees of freedom, like crystals (modeled by lattices of coupled oscillators) or fields (e.g., the electromagnetic field important in the study of black body radiation). However, although this has been the first surprise of quantum statistics (and in fact responsible for the very discovery of quanta), it is by no means the last.

At low temperatures, new unexpected (i.e., with no analogs in classical statistical mechanics) phenomena occur: Bose–Einstein condensation (superfluidity), Fermi surface instability (superconductivity), and appearance of off-diagonal long-range order (ODLRO) will be selected to illustrate the deeply different kinds of problems of quantum statistical mechanics. Largely not yet understood, such phenomena pose very interesting problems not only from the physical point of view but also from the mathematical point of view and may pose challenges even at the level of a definition. However, it should be kept in mind that in the interesting cases (i.e., three-dimensional systems and even most two- and one-dimensional systems) there is no proof that the objects defined below really exist for the systems like [66] (see, however, the final comment for an important exception).

### Bose–Einstein Condensation

In a canonical state with parameters $\beta, v$, a definition of the occurrence of Bose condensation is in terms of the eigenvalues $\nu_j(\Omega, N)$ of the kernel $\rho(\boldsymbol{q}, \boldsymbol{q}')$ on $L_2(\Omega)$, called the *one-particle reduced density matrix*, defined by

$$N \sum_{n=1}^{\infty} \frac{\mathrm{e}^{-\beta E_n(\Omega,N)}}{\operatorname{tr} \mathrm{e}^{-\beta H_N}} \int \overline{\psi}_n(\boldsymbol{q}, \boldsymbol{q}_1, \ldots, \boldsymbol{q}_{N-1}) \times \psi_n(\boldsymbol{q}', \boldsymbol{q}_1, \ldots, \boldsymbol{q}_{N-1}) \, \mathrm{d}\boldsymbol{q}_1 \ldots \mathrm{d}\boldsymbol{q}_{N-1} \qquad [68]$$

where $E_n(\Omega, N)$ are the eigenvalues of $H_N$ and $\psi_n(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N)$ are the corresponding eigenfunctions. If $\nu_j$ are ordered by increasing value, the state with parameters $\beta, v$ is said to contain a *Bose–Einstein condensate* if $\nu_1(\Omega, N) \geq bN > 0$ for all large $\Omega$ at $v = V/N, \beta$ fixed. This receives the interpretation that there are more than $bN$ particles with equal momentum. The free Bose gas exhibits a Bose condensation phenomenon at fixed density and small temperature.

### Fermi Surface

The wave functions $\psi_n(\boldsymbol{q}_1, \sigma_1, \ldots, \boldsymbol{q}_N, \sigma_N) \equiv \psi_n(\boldsymbol{Q}, \boldsymbol{\sigma})$ are now antisymmetric in the permutations of the pairs $(\boldsymbol{q}_i, \sigma_i)$. Let $\psi(\boldsymbol{Q}, \boldsymbol{\sigma}; N, n)$ denote the $n$th

eigenfunction of the $N$-particle energy $H_N$ in [66] with eigenvalue $E(N,n)$ (labeled by $n=0,1,\ldots$ and nondecreasingly ordered). Setting $\boldsymbol{Q}''=(\boldsymbol{q}''_1,\ldots,\boldsymbol{q}''_{N-p})$, $\boldsymbol{\sigma}''=(\sigma''_1,\ldots,\sigma''_{N-p})$, introduce the kernels $\rho_p^{H_N}(\boldsymbol{Q},\boldsymbol{\sigma};\boldsymbol{Q}',\boldsymbol{\sigma}')$ by

$$\rho_p(\boldsymbol{Q},\boldsymbol{\sigma};\boldsymbol{Q}',\boldsymbol{\sigma}') \stackrel{\text{def}}{=} p!\binom{N}{p}\int\sum_{\boldsymbol{\sigma}''}\mathrm{d}^{N-p}\boldsymbol{Q}''\sum_{n=0}^{\infty}\frac{\mathrm{e}^{-\beta E(N,n)}}{\operatorname{tr}\mathrm{e}^{-\beta H_N}} \times\overline{\psi}(\boldsymbol{Q},\boldsymbol{\sigma};\boldsymbol{Q}'',\boldsymbol{\sigma}'';N,n)\psi(\boldsymbol{Q}',\boldsymbol{\sigma}';\boldsymbol{Q}'',\boldsymbol{\sigma}'';N,n) \quad [69]$$

which are called *p-particle reduced density matrices* (extending the corresponding one-particle reduced density matrix [68]). Denote $\rho(\boldsymbol{q}_1-\boldsymbol{q}_2)\stackrel{\text{def}}{=}\sum_\sigma\rho_1(\boldsymbol{q}_1,\sigma,\boldsymbol{q}_2,\sigma)$. It is also useful to consider spinless fermionic systems: the corresponding definitions are obtained simply by suppressing the spin labels and will not be repeated.

Let $r_1(\boldsymbol{k})$ be the Fourier transform of $\rho_1(\boldsymbol{q}-\boldsymbol{q}')$: the Fermi surface can be defined as the locus of the $\boldsymbol{k}$'s in the neighborhood of which $\partial_{\boldsymbol{k}} r_1(\boldsymbol{k})$ is unbounded as $\Omega\to\infty$, $\beta\to\infty$. The limit as $\beta\to\infty$ is important because the notion of a Fermi surface is, possibly, precise only at zero temperature, that is at $\beta=\infty$.

So far, existence of Fermi surface (i.e., the smoothness of $r_1(\boldsymbol{k})$ except on a smooth surface in $\boldsymbol{k}$-space) has been proved in free Fermi systems ($\varphi=0$) and

1. certain exactly soluble one-dimensional spinless systems and
2. in rather general one-dimensional spinless systems or systems with spin and repulsive pair interaction, possibly in an external periodic potential.

The spinning case in a periodic potential and dimension $d\ge 2$ is the most interesting case to study for its relevance in the theory of conduction in crystals. Essentially no mathematical results are available as the above-mentioned ones do not concern any case in dimension >1: this is a rather deceiving aspect of the theory and a challenge.

In dimension 2 or higher, for fermionic systems with Hamiltonian [66], not only there are no results available, even without spin, but it is not even clear that a Fermi surface can exist in presence of interesting interactions.

### Cooper Pairs

The superconductivity theory has been phenomenologically related to the existence of Cooper pairs. Consider the Hamiltonian [66] and define (cf. [69])

$$\rho(\boldsymbol{x}-\boldsymbol{y},\sigma;\boldsymbol{x}'-\boldsymbol{y}',\sigma';\boldsymbol{x}-\boldsymbol{x}') \stackrel{\text{def}}{=} \rho_2(\boldsymbol{x},\sigma,\boldsymbol{y},-\sigma;\boldsymbol{x}',\sigma',\boldsymbol{y}',-\sigma')$$

The system is said to contain *Cooper pairs* with spins $\sigma,-\sigma$ ($\sigma=+$ or $\sigma=-$) if there exist functions $g^\alpha(\boldsymbol{q},\sigma)\neq 0$ with

$$\int\overline{g^\alpha}(\boldsymbol{q},\sigma)g^{\alpha'}(\boldsymbol{q},\sigma)\,\mathrm{d}\boldsymbol{q}=0\quad\text{if }\alpha\neq\alpha'$$

such that

$$\lim_{V\to\infty}\rho(\boldsymbol{x}-\boldsymbol{y},\sigma,\boldsymbol{x}'-\boldsymbol{y}',\sigma',\boldsymbol{x}-\boldsymbol{x}') \xrightarrow[\boldsymbol{x}-\boldsymbol{x}'\to\infty]{} \sum_\alpha g^\alpha(\boldsymbol{x}-\boldsymbol{y},\sigma)\overline{g^\alpha}(\boldsymbol{x}'-\boldsymbol{y}',\sigma') \quad [70]$$

In this case, $g^\alpha(\boldsymbol{x}-\boldsymbol{y},\sigma)$ with largest $L_2$ norm can be called, after normalize, the wave function of the paired state of lowest energy: this is the analog of the plane wave for a free particle (and, like it, it is manifestly not normalizable, i.e., it is not square integrable as a function of $\boldsymbol{x},\boldsymbol{y}$). If the system contains Cooper pairs and the nonleading terms in the limit [70] vanish quickly enough the two-particle reduced density matrix [70] regarded as a kernel operator has an eigenvalue of order $V$ as $V\to\infty$: that is, the state of lowest energy is "macroscopically occupied," quite like the free Bose condensation in the ground state.

Cooper pairs instability might destroy the Fermi surface in the sense that $r_1(\boldsymbol{k})$ becomes analytic in $\boldsymbol{k}$; but it is also possible that, even in the presence of them, there remains a surface which is the locus of the singularities of the function $r_1(\boldsymbol{k})$. In the first case, there should remain a trace of it as a very steep gradient of $r_1(\boldsymbol{k})$ of the order of an exponential in the inverse of the coupling strength; this is what happens in the *BCS model* for superconductivity. The model is, however, a mean-field model and this particular regularity aspect might be one of its peculiarities. In any event, a smooth singularity surface is very likely to exist for some interesting density matrix (e.g., in the BCS model with "gap parameter $\gamma$" the wave function

$$g(\boldsymbol{x}-\boldsymbol{y},\sigma)\equiv\frac{1}{(2\pi)^d}\int_{\varepsilon(\boldsymbol{k})>0}\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot(\boldsymbol{x}-\boldsymbol{y})}\frac{\gamma}{\sqrt{\varepsilon(\boldsymbol{k})^2+\gamma^2}}\mathrm{d}\boldsymbol{k}$$

of the lowest energy level of the Cooper pairs is singular on a surface coinciding with the Fermi surface of the free system).

### ODLRO

Consider the $k$-fermion reduced density matrix $\rho_k(\boldsymbol{Q},\boldsymbol{\sigma};\boldsymbol{Q}',\boldsymbol{\sigma}')$ as kernel operators $O_k$ on $L_2((\Omega\times C^2)^k)_-$. Suppose $k$ is even, then if $O_k$ has a (generalized) eigenvalue of order $N^{k/2}$ as $N\to\infty, N/V=\rho$, the system is said to exhibit *off-diagonal long-range* order of order $k$. For $k$ odd, ODLRO is defined to exist if $O_k$ has an eigenvalue of order $N^{(k-1)/2}$ and $k\ge 3$ (if $k=1$ the largest eigenvalue of $O_1$ is necessarily $\le 1$).

For bosons, consider the reduced density matrix $\rho_k(Q;Q')$ regarding it as a kernel operator $O_k$ on $L_2(\Omega)_+^k$ and define ODLRO of order $k$ to be present if $O(k)$ has a (generalized) eigenvalue of order $N^k$ as $N \to \infty, N/V = \rho$.

ODLRO can be regarded as a unification of the notions of Bose condensation and of the existence of Cooper pairs, because Bose condensation could be said to correspond to the kernel operator $\rho_1(q_1 - q_2)$ in [68] having a (generalized) eigenvalue of order $N$, and to be a case of ODLRO of order 1. If the state is pure in the sense that it has a cluster property (see the sections "Phase transitions and boundary conditions" and "Lattice models"), then the existence of ODLRO, Bose condensation, and Cooper pairs implies that the system shows a spontaneously broken symmetry: conservation of particle number and clustering imply that the off-diagonal elements of (all) reduced density matrices vanish at infinite separation in states obtained as limits of states with periodic boundary conditions and Hamiltonian [66], and this is incompatible with ODLRO.

The free Fermi gas has no ODLRO, the BCS model of superconductivity has Cooper pairs and ODLRO with $k = 2$, but no Fermi surface in the above sense (possibly too strict). Fermionic systems cannot have ODLRO of order 1 (because the reduced density matrix of order 1 is bounded by 1).

The contribution of mathematical physics has been particularly effective in providing exactly soluble models: however, the soluble models deal with one-dimensional systems and it can be shown that in dimensions 1, 2 no ODLRO can take place. A major advance is the recent proof of ODLRO and Bose condensation in the case of a lattice version of [66] at a special density value (and $d \geq 3$).

In no case, for the Hamiltonian [66] with $\varphi \neq 0$, existence of Cooper pairs has been proved nor existence of a Fermi surface for $d > 1$. Nevertheless, both Bose condensation and Cooper pairs formation can be proved to occur rigorously in certain limiting situations. There are also a variety of phenomena (e.g., simple spectral properties of the Hamiltonians) which are believed to occur once some of the above-mentioned ones do occur and several of them can be proved to exist in concrete models.

If $d = 1, 2$, ODLRO can be proved to be impossible at $T > 0$ through the use of Bogoliubov's inequality (used in the "no $d = 2$ crystal theorem," see the section "Continuous symmetries: 'no $d = 2$ crystal' theorem").

For more details, the reader is referred to Penrose and Onsager (1956), Yang (1962), Ruelle (1969), Hohenberg (1967), Gallavotti (1999), and Aizenman *et al.* (2004).

## Appendix 1: The Physical Meaning of the Stability Conditions

It is useful to see what would happen if the conditions of stability and temperedness (see [14]) are violated. The analysis also illustrates some of the typical methods of statistical mechanics.

### Coalescence Catastrophe due to Short-Distance Attraction

The simplest violation of the first condition in [14] occurs when the potential $\varphi$ is smooth and negative at the origin.

Let $\delta > 0$ be so small that the potential at distances $\leq 2\delta$ is $\leq -b < 0$. Consider the canonical distribution with parameters $\beta, N$ in a (cubic) box $\Omega$ of volume $V$. The probability $P_{\text{collapse}}$ that all the $N$ particles are located in a little sphere of radius $\delta$ around the center of the box (or around any prefixed point of the box) is estimated from below by remarking that

$$\Phi \leq -b\binom{N}{2} \sim -\frac{b}{2}N^2$$

so that

$$P_{\text{collapse}} = \frac{\int_c \frac{\mathrm{d}p\,\mathrm{d}q}{h^{3N}N!}\mathrm{e}^{-\beta(K(p)+\Phi(q))}}{\int \frac{\mathrm{d}p\,\mathrm{d}q}{h^{3N}N!}\mathrm{e}^{-\beta(K(p)+\Phi(q))}} \geq \frac{\left(\frac{4\pi\sqrt{2m\beta^{-1}}^3}{3h^3}\right)^N \frac{\delta^{3N}}{N!}\mathrm{e}^{\beta b(1/2)N(N-1)}}{\int \frac{\mathrm{d}q}{h^{3N}N!}\mathrm{e}^{-\beta\Phi(q)}} \qquad [71]$$

The phase space is extremely small: nevertheless, such configurations are far more probable than the configurations which "look macroscopically correct," that is, configurations with particles more or less spaced by the average particle distance expected in a macroscopically homogeneous configuration, namely $(N/V)^{-1/3} = \rho^{-1/3}$. Their energy $\Phi(q)$ is of the order of $uN$ for some $u$, so that their probability will be bounded above by

$$P_{\text{regular}} \leq \frac{\int \frac{\mathrm{d}p\,\mathrm{d}q}{h^{3N}N!}\mathrm{e}^{-\beta(K(p)+uN)}}{\int \frac{\mathrm{d}p\,\mathrm{d}q}{h^{3N}N!}\mathrm{e}^{-\beta(K(p)+\Phi(q))}} = \frac{\frac{V^N\sqrt{2m\beta^{-1}}^3}{h^{3N}N!}\mathrm{e}^{-\beta uN}}{\int \frac{\mathrm{d}q}{h^{3N}N!}\mathrm{e}^{-\beta\Phi(q)}} \qquad [72]$$

However, no matter how small $\delta$ is, the ratio $P_{\text{regular}}/P_{\text{collapse}}$ will approach 0 as $V \to \infty$, $N/V \to v^{-1}$; this occurs extremely rapidly because $e^{\beta b N^2/2}$ eventually dominates over $V^N \sim e^{N \log N}$.

Thus, it is far more probable to find the system in a microscopic volume of size $\delta$ rather than in a configuration in which the energy has some macroscopic value proportional to $N$. This catastrophe can be called an ultraviolet catastrophe (as it is due to the behavior at very short distances) and it causes the collapse of the particles into configurations concentrated in regions as small as we please as $V \to \infty$.

#### Coalescence Catastrophe due to Long-Range Attraction

It occurs when the potential is too attractive near $\infty$. For simplicity, suppose that the potential has a hard core, i.e., it is $+\infty$ for $r < r_0$, so that the above-discussed coalescence cannot occur and the system density bounded above by a certain quantity $\rho_{\text{cp}} < \infty$ (*close-packing density*).

The catastrophe occurs if $\varphi(\boldsymbol{q}) \sim -g|\boldsymbol{q}|^{-3+\varepsilon}$, $g, \varepsilon > 0$, for $|\boldsymbol{q}|$ large. For instance, this is the case for matter interacting gravitationally; if $k$ is the gravitational constant, $m$ is the particle mass, then $g = km^2$ and $\varepsilon = 2$.

The probability $P_{\text{regular}}$ of "regular configurations," where particles are at distances of order $\rho^{-1/3}$ from their close neighbors, is compared with the probability $P_{\text{collapse}}$ of "catastrophic configurations," with the particles at distances $r_0$ from their close neighbors to form a configuration of density $\rho_{\text{cp}}/(1+\delta)^3$ almost in close packing (so that $r_0$ is equal to the hard-core radius times $1+\delta$). In the latter case, the system does not fill the available volume and leaves empty a region whose volume is a fraction $\sim((\rho_{\text{cp}} - \rho)/\rho_{\text{cp}})V$ of $V$. Further, it can be checked that the ratio $P_{\text{regular}}/P_{\text{collapse}}$ tends to 0 at a rate $O(\exp(g\frac{1}{2}N(\rho_{\text{cp}}(1+\delta)^{-3} - \rho)))$ if $\delta$ is small enough (and $\rho < \rho_{\text{cp}}$).

A system which is too attractive at infinity will not occupy the available volume but will stay confined in a close-packed configuration even in empty space.

This is important in the theory of stars: stars cannot be expected to obey "regular thermodynamics" and in particular will not "evaporate" because their particles interact via the gravitational force at large distances. Stars do not occupy the whole volume given to them (i.e., the universe); they do not collapse to a point only because the interaction has a strongly repulsive core (even when they are burnt out and the radiation pressure is no longer able to keep them at a reasonable size).

#### Evaporation Catastrophe

This is another *infrared catastrophe*, that is, a catastrophe due to the long-range structure of the interactions in the above subsection; it occurs when the potential is too repulsive at $\infty$, that is,

$$\varphi(\boldsymbol{q}) \sim +g|\boldsymbol{q}|^{-3+\varepsilon} \quad \text{as} \quad \boldsymbol{q} \to \infty$$

so that the temperedness condition is again violated.

In addition, in this case, the system does not occupy the whole volume: it will generate a layer of particles sticking, in close-packed configuration, to the walls of the container. Therefore, if the density is lower than the close-packing density, $\rho < \rho_{\text{cp}}$, the system will leave a region around the center of the container $\Omega$ empty; and the volume of the empty region will still be of the order of the total volume of the box (i.e., its diameter will be a fraction of the box side $L$). The proof is completely analogous to the one of the previous case; except that now the configuration with lowest energy will be the one sticking to the wall and close packed there, rather than the one close packed at the center.

Also this catastrophe is important as it is realized in systems of charged particles bearing the same charge: the charges adhere to the boundary in close-packing configuration, and dispose themselves so that the electrostatic potential energy is minimal. Therefore, charges deposited on a metal will not occupy the whole volume: they will rather form a surface layer minimizing the potential energy (i.e., so that the Coulomb potential in the interior is constant). In general, charges in excess of neutrality do not behave thermodynamically: for instance, besides not occupying the whole volume given to them, they will not contribute normally to the specific heat.

Neutral systems of charges behave thermodynamically if they have hard cores, so that the ultraviolet catastrophe cannot occur or if they obey quantum-mechanical laws and consist of fermionic particles (plus possibly bosonic particles with charges of only one sign).

For more details, we refer the reader to Lieb and Lebowitz (1972) and Lieb and Thirring (2001).

## Appendix 2: The Subadditivity Method

A simple consequence of the assumptions is that the exponential in (5.2) can be bounded above by $e^{\beta BN} \exp(-\frac{\beta}{2m}\sum_{i=1}^{N} \boldsymbol{P}_i^2)$ so that

$$1 \le Z_{\text{gc}}(\beta, \lambda, V) \le \exp\left(V e^{\beta\lambda} e^{\beta B} \sqrt{2m\beta^{-1}}^{\,d}\right)$$
$$\Rightarrow 0 \le \frac{1}{V} \log Z_{\text{gc}}(\beta, \lambda, V) \le e^{\beta\lambda} e^{\beta B} \sqrt{2m\beta^{-1}}^{\,d} \quad [73]$$

Consider, for simplicity, the case of a hard-core interaction with finite range (cf. [14]). Consider a

sequence of boxes $\Omega_n$ with sides $2^n L_0$, where $L_0 > 0$ is arbitrarily fixed to be $>2R$. The partition function $Z_{gc}(\beta, z)$ relative to the volume $\Omega_n$ is

$$Z_n = \sum_{N=0}^{\infty} \frac{z^N}{N!} \int_{\Omega_n} \mathrm{d}Q \mathrm{e}^{-\beta\Phi(Q)}$$

because the integral over the $P$ variables can be explicitly performed and included in $z^N$ if $z$ is defined as $z = \mathrm{e}^{\beta\lambda}(2m\beta^{-1})^{d/2}$.

Then the box $\Omega_n$ contains $2^d$ boxes $\Omega_{n-1}$ for $n \geq 1$ and

$$1 \leq Z_n \leq Z_{n-1}^{2^d} \exp\left(\beta B 2d (L_{n-1}/R)^{d-1} 2^{2d}\right) \qquad [74]$$

because the corridor of width $2R$ around the boundaries of the $2^d$ cubes $\Omega_{n-1}$ filling $\Omega_n$ has volume $2RL_{n-1}2^d$ and contains at most $(L_{n-1}/R)^{d-1}2^d$ particles, each of which interacts with at most $2^d$ other particles. Therefore,

$$\begin{aligned}\beta p_n &\stackrel{\text{def}}{=} L_n^d \log Z_n \\ &\leq L_{n-1}^d \log Z_{n-1} + \beta B \gamma_d 2^{-n}(L_0/R)^{d-1}\end{aligned}$$

for some $\gamma_d > 0$. Hence, $0 \leq \beta p_n \leq \beta p_{n-1} + \Gamma_d 2^{-n}$ for some $\Gamma_d > 0$ and $p_n$ is bounded above and below uniformly in $n$. So, the limit [13] exists on the sequence $L_n = L_0 2^n$ and defines a function $\beta p_\infty(\beta, \lambda)$.

A box of arbitrary size $L$ can be filled with about $(L/L_{\bar{n}})^d$ boxes of side $L_{\bar{n}}$ with $\bar{n}$ so large that, prefixed $\delta > 0$, $|p_\infty - p_n| < \delta$ for all $n \geq \bar{n}$. Likewise, a box of size $L_n$ can be filled by about $(L_n/L)^d$ boxes of size $L$ if $n$ is large. The latter remarks lead us to conclude, by standard inequalities, that the limit in [13] exists and coincides with $p_\infty$.

The subadditivity method just demonstrated for finite-range potentials with hard core can be extended to the potentials satisfying just stability and temperedness (cf. the section "Thermodynamic limit").

For more details, the reader is referred to Ruelle (1969) and Gallavotti (1999).

## Appendix 3: An Infrared Inequality

The infrared inequalities stem from *Bogoliubov's inequality*. Consider as an example the problem of crystallization discussed in the section "Continuous symmetries: 'no $d=2$ crystal' theorem". Let $\langle \cdot \rangle$ denote average over a canonical equilibrium state with Hamiltonian

$$H = \sum_{j=1}^{N} \frac{p_j^2}{2} + U(Q) + \varepsilon W(Q)$$

with given temperature and density parameters $\beta, \rho, \rho = a^{-3}$. Let $\{X, Y\} = \sum_j (\partial_{p_j} X \, \partial_{q_j} Y - \partial_{q_j} X \, \partial_{p_j} Y)$ be the Poisson bracket. Integration by parts, with periodic boundary conditions, yields

$$\begin{aligned}\langle A^* \{C, H\}\rangle &\equiv -\frac{\int A^* \{C, \mathrm{e}^{-\beta H}\} \mathrm{d}P \mathrm{d}Q}{\beta Z_c(\beta, \rho, N)} \\ &\equiv -\beta^{-1} \langle \{A^*, C\}\rangle \end{aligned} \qquad [75]$$

as a general identity. The latter identity implies, for $A = \{C, H\}$, that

$$\langle \{H, C\}^* \{H, C\}\rangle = -\beta^{-1} \langle \{C, \{H, C^*\}\}\rangle \qquad [76]$$

Hence, the Schwartz inequality $\langle A^* A\rangle \langle \{H, C\}^* \{H, C\}\rangle \geq |\langle \{A^*, C\}\rangle|^2$ combined with the two relations in [75], [76] yields Bogoliubov's inequality:

$$\langle A^* A\rangle \geq \beta^{-1} \frac{|\langle \{A^*, C\}\rangle|^2}{\langle \{C, \{C^*, H\}\}\rangle} \qquad [77]$$

Let $g$, $h$ be arbitrary complex (differentiable) functions and $\partial_j = \partial_{q_j}$

$$A(Q) \stackrel{\text{def}}{=} \sum_{j=1}^{N} g(q_j), \qquad C(P, Q) \stackrel{\text{def}}{=} \sum_{j=1}^{N} p_j h(q_j) \qquad [78]$$

Then $H = \sum \frac{1}{2} p_j^2 + \Phi(q_1, \ldots, q_N)$, if

$$\Phi(q_1, \ldots, q_N) = \frac{1}{2} \sum_{j \neq j'} \varphi(|q_j - q_{j'}|) + \varepsilon \sum_j W(q_j)$$

so that, via algebra,

$$\{C, H\} \equiv \sum_j (h_j \partial_j \Phi - p_j (p_j \cdot \partial_j) h_j)$$

with $h_j \stackrel{\text{def}}{=} h(q_j)$. If $h$ is real valued, $\langle \{C, \{C^*, H\}\}\rangle$ becomes, again via algebra,

$$\left\langle \sum_{jj'} h_j h_{j'} \partial_j \cdot \partial_{j'} \Phi(Q) \right\rangle + \left\langle \varepsilon \sum_j h_j^2 \Delta W(q_j) + \frac{4}{\beta} \sum_j (\partial_j h_j)^2 \right\rangle$$

(integrals on $p_j$ just replace $p_j^2$ by $2\beta^{-1}$ and $\langle (p_j)_i (p_j)_{i'}\rangle = \beta^{-1} \delta_{i,i'}$). Therefore, the average $\langle \{C, \{C^*, H\}\}\rangle$ becomes

$$\left\langle \frac{1}{2} \sum_{jj'} (h_j - h_{j'})^2 \Delta\varphi(|q_j - q_{j'}|) + \varepsilon \sum_j h_j^2 \Delta W(q_j) + 4\beta^{-1} \sum_j (\partial_j h_j)^2 \right\rangle \qquad [79]$$

Choose $g(q) \equiv e^{-i(\kappa + K)\cdot q}$, $h(q) = \cos q \cdot \kappa$ and bound $(h_j - h_{j'})^2$ by $\kappa^2 (q_j - q_{j'})^2$, $(\partial_j h_j)^2$ by $\kappa^2$ and

$h_j^2$ by 1. Hence [79] is bounded above by $ND(\boldsymbol{\kappa})$ with

$$D(\boldsymbol{\kappa}) \stackrel{\text{def}}{=} \left\langle \boldsymbol{\kappa}^2 \left( 4\beta^{-1} + \frac{1}{2N} \sum_{j \neq j'} (\boldsymbol{q}_j - \boldsymbol{q}_{j'})^2 |\Delta\varphi(\boldsymbol{q}_j - \boldsymbol{q}_{j'})| \right) + \varepsilon \frac{1}{N} \sum_j |\Delta W(\boldsymbol{q}_j)| \right\rangle \qquad [80]$$

This can be used to estimate the denominator in [77]. For the LHS remark that

$$\langle A^*, A \rangle = |\sum_{j=1}^{N} \mathrm{e}^{-\mathrm{i}\boldsymbol{q}\cdot(\boldsymbol{\kappa}+\boldsymbol{K})}|^2$$

and

$$|\langle\{A^*, \mathbf{C}\}\rangle|^2 = \left|\left\langle \sum_j h_j \partial g_j \right\rangle\right|^2 = |\boldsymbol{K} + \boldsymbol{\kappa}|^2 N^2 (\rho_\varepsilon(\boldsymbol{K}) + \rho_\varepsilon(\boldsymbol{K} + 2\boldsymbol{\kappa}))^2$$

hence [77] becomes, after multiplying both sides by the auxiliary function $\gamma(\boldsymbol{\kappa})$ (assumed even and vanishing for $|\boldsymbol{\kappa}| > \pi/a$) and summing over $\boldsymbol{\kappa}$,

$$D_1 \stackrel{\text{def}}{=} \frac{1}{N} \sum_{\boldsymbol{\kappa}} \gamma(\boldsymbol{\kappa}) \left\langle \frac{1}{N} |\sum_{j=1}^{N} \mathrm{e}^{-\mathrm{i}(\boldsymbol{K}+\boldsymbol{\kappa})\cdot \boldsymbol{q}_j}|^2 \right\rangle \geq \frac{1}{N} \sum_{\boldsymbol{\kappa}} \gamma(\boldsymbol{\kappa}) \times \frac{|\boldsymbol{K}|^2}{4\beta} \frac{(\rho_\varepsilon(\boldsymbol{K}) + \rho_\varepsilon(\boldsymbol{K} + 2\boldsymbol{\kappa}))^2}{D(\boldsymbol{\kappa})} \qquad [81]$$

To apply [77] the averages in [80], [81] have to be bounded above: this is a technical point that is discussed here, as it illustrates a general method of using the results on the thermodynamic limits and their convexity properties to obtain estimates.

Note that $\langle (1/N) \sum_{\boldsymbol{k}} \gamma(\boldsymbol{k}) \mathrm{d}^d \boldsymbol{k} | \sum_{j=1}^{N} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot \boldsymbol{q}_j}|^2 \rangle$ is identically $\widetilde{\varphi}(0) + (2/N) \langle \sum_{j<j'} \widetilde{\varphi}(\boldsymbol{q}_j - \boldsymbol{q}_{j'}) \rangle$ with $\widetilde{\varphi}(\boldsymbol{q}) \stackrel{\text{def}}{=} (1/N) \sum_{\boldsymbol{\kappa}} \gamma(\boldsymbol{\kappa}) \mathrm{e}^{\mathrm{i}\boldsymbol{\kappa}\cdot \boldsymbol{q}}$.

Let $\varphi_{\lambda,\zeta}(\boldsymbol{q}) \stackrel{\text{def}}{=} \varphi(\boldsymbol{q}) + \lambda \boldsymbol{q}^2 |\Delta\varphi(\boldsymbol{q})| + \eta \widetilde{\varphi}(\boldsymbol{q})$ and let $F_V(\lambda, \eta, \zeta) \stackrel{\text{def}}{=} (1/N) \log Z^{\mathrm{c}}(\lambda, \eta, \zeta)$ with $Z^{\mathrm{c}}$ the partition function in the volume $\Omega$ computed with energy $U' = \sum_{jj'} \varphi_{\lambda,\zeta}(\boldsymbol{q}_j - \boldsymbol{q}_{j'}) + \varepsilon \sum_j W(\boldsymbol{q}_j) + \eta\varepsilon \sum |\Delta W(\boldsymbol{q}_j)|$. Then $F_V(\lambda, \eta, \zeta)$ is convex in $\lambda, \eta$ and it is uniformly bounded above and below if $|\eta|, |\varepsilon|, |\zeta| \leq 1$ (say) and $|\lambda| \leq \lambda_0$: here $\lambda_0 > 0$ exists if $\boldsymbol{r}^2 |\Delta\varphi(\boldsymbol{r})|$ satisfies the assumption set at the beginning of the section "Continuous symmetries: 'no $d = 2$ crystal' theorem" and the density is smaller than a close packing (this is because the potential $U'$ will still satisfy conditions similar to [14] uniformly in $|\varepsilon|, |\eta| < 1$ and $|\lambda|$ small enough).

Convexity and boundedness above and below in an interval imply bounds on the derivatives in the interior points, in this case on the derivatives of $F_V$ with respect to $\lambda, \eta, \zeta$ at 0. The latter are identical to the averages in [80], [81]. In this way, the constants $B_1, B_2, B_0$ such that $D(\boldsymbol{\kappa}) \leq \boldsymbol{\kappa}^2 B_1 + \varepsilon B_2$ and $B_0 > D_1$ are found.

For more details, the reader is referred to Mermin (1968).

## Further Reading


Aizenman M (1980) Translation invariance and instability of phase coexistence in the two dimensional Ising system. *Communications in Mathematical Physics* 73: 83–94.

Aizenman M (1982) *Geometric analysis of $\varphi^4$ fields and Ising models*. 86: 1–48.

Aizenman M, Lieb EH, Seiringer R, Solovej JP, and Yngvason J (2004) Bose–Einstein condensation as a quantum phase transition in a optical lattice, *Physical Review A* 70: 023612.

Baxter R (1982) *Exactly Solved Models*. London: Academic Press.

Benfatto G and Gallavotti G (1995) *Renormalization group*. Princeton: Princeton University Press.

Bleher P and Sinai Y (1975) Critical indices for Dyson's asymptotically hierarchical models. *Communications in Mathematical Physics* 45: 247–278.

Boltzmann L (1968a) Über die mechanische Bedeutung des zweiten Haupsatzes der Wärmetheorie. In: Hasenöhrl F (ed.) *Wissenschaftliche Abhandlungen*, vol. I, pp. 9–33. New York: Chelsea.

Boltzmann L (1968b) Über die Eigenshaften monzyklischer und anderer damit verwandter Systeme. In: Hasenöhrl FP (ed.) *Wissenshafltliche Abhandlungen*, vol. III, pp. 122–152. New York: Chelsea.

Dobrushin RL (1968) Gibbsian random fields for lattice systems with pairwise interactions. *Functional Analysis and Applications* 2: 31–43.

Domb C and Green MS (1972) *Phase Transitions and Critical Points*. New York: Wiley.

Dyson F (1969) Existence of a phase transition in a one–dimensional Ising ferromagnet. *Communications in Mathematical Physics* 12: 91–107.

Dyson F and Lenard A (1967, 1968) Stability of matter. *Journal of Mathematical Physics* 8: 423–434, 9: 698–711.

Friedli S and Pfister C (2004) On the singularity of the free energy at a first order phase transition. *Communications in Mathematical Physics* 245: 69–103.

Gallavotti G (1999) *Statistical Mechanics*. Berlin: Springer.

Gallavotti G, Bonetto F and Gentile G (2004) *Aspects of the Ergodic, Qualitative and Statistical Properties of Motion*. Berlin: Springer.

Gawedzky K and Kupiainen A (1983) Block spin renormalization group for dipole gas and $(\partial\phi)^4$. *Annals of Physics* 147: 198–243.

Gawedzky K and Kupiainen A (1985) Massless lattice $\phi_4^4$ theory: rigorous control of a renormalizable asymptotically free model. *Communications in Mathematical Physics* 99: 197–252.

Gibbs JW (1981) *Elementary Principles in Statistical Mechanics*. Woodbridge (Connecticut): Ox Bow Press (reprint of the 1902 edition).

Higuchi Y (1981) On the absence of non translationally invariant Gibbs states for the two dimensional Ising system. In: Fritz J, Lebowitz JL, and Szaz D (eds.) *Random Folds*. Amsterdam: North-Holland.

Hohenberg PC (1967) Existence of long range order in one and two dimensions. *Physical Review* 158: 383–386.

Landau L and Lifschitz LE (1967) *Physique Statistique*. Moscow: MIR.
Lanford O and Ruelle D (1969) Observables at infinity and states with short range correlations in statistical mechanics. *Communications in Mathematical Physics* 13: 194–215.
Lebowitz JL (1974) GHS and other inequalities. *Communications in Mathematical Physics* 28: 313–321.
Lebowitz JL and Penrose O (1979) Towards a rigorous molecular theory of metastability. In: Montroll EW and Lebowitz JL (eds.) *Fluctuation Phenomena*. Amsterdam: North-Holland.
Lee TD and Yang CN (1952) Statistical theory of equations of state and phase transitions, II. Lattice gas and Ising model. *Physical Review* 87: 410–419.
Lieb EH (2002) *Inequalities*. Berlin: Springer.
Lieb EH and Lebowitz JL (1972) Lectures on the Thermodynamic Limit for Coulomb Systems, In: Lenard A (ed.) *Springer Lecture Notes in Physics*, vol. 20, pp. 135–161. Berlin: Springer.
Lieb EH and Thirring WE (2001) *Stability of Matter from Atoms to Stars*. Berlin: Springer.
Mastropietro V (2004) Ising models with four spin interaction at criticality. *Communications in Mathematical Physics* 244: 595–642.
McCoy BM and Wu TT (1973) *The two Dimensional Ising Model*. Cambridge: Harvard University Press.
Mermin ND (1968) Crystalline order in two dimensions. *Physical Review* 176: 250–254.
Miracle-Solé S (1995) Surface tension, step free energy and facets in the equilibrium crystal shape. *Journal Statistical Physics* 79: 183–214.
Olla S (1987) Large deviations for Gibbs random fields. *Probability Theory and Related Fields* 77: 343–357.
Onsager L (1944) Crystal statistics. I. A two dimensional Ising model with an order–disorder transition. *Physical Review* 65: 117–149.
Penrose O and Onsager L (1956) Bose–Einstein condensation and liquid helium. *Physical Review* 104: 576–584.
Pfister C and Velenik Y (1999) Interface, surface tension and Reentrant pinning transition in the 2D Ising model. *Communications in Mathematical Physics* 204: 269–312.
Ruelle D (1969) *Statistical Mechanics*. New York: Benjamin.
Ruelle D (1971) Extension of the Lee–Yang circle theorem. *Physical Review Letters* 26: 303–304.
Sinai Ya G (1991) *Mathematical Problems of Statistical Mechanics*. Singapore: World Scientific.
van Beyeren H (1975) Interphase sharpness in the Ising model. *Communications in Mathematical Physics* 40: 1–6.
Wilson KG and Fisher ME (1972) Critical exponents in 3.99 dimensions. *Physical Review Letters* 28: 240–243.
Yang CN (1962) Concept of off-diagonal long-range order and the quantum phases of liquid He and of superconductors. *Reviews of Modern Physics* 34: 694–704.

# Introductory Article: Functional Analysis

**S Paycha**, Université Blaise Pascal, Aubière, France



## Introduction

Functional analysis is concerned with the study of functions and function spaces, combining techniques borrowed from classical analysis with algebraic techniques. Modern functional analysis developed around the problem of solving equations with solutions given by functions. After the differential and partial differential equations, which were studied in the eighteenth century, came the integral equations and other types of functional equations investigated in the nineteenth century, at the end of which arose the need to develop a new analysis, with functions of an infinite number of variables instead of the usual functions. In 1887, Volterra, inspired by the calculus of variations, suggested a new infinitesimal calculus where usual functions are replaced by functionals, that is, by maps from a function space to $\mathbb{R}$ or $\mathbb{C}$, but he and his followers were still missing some algebraic and topological tools to be developed later. Modern analysis was born with the development of an "algebra of the infinite" closely related to classical linear algebra which by 1890 had (up to the concept of duality, which was developed later) settled on firm ground. Strongly inspired by algebraic methods, Fredholm's work at the turn of the nineteenth century, in which emerged the concept of kernel of an operator, became a founding stone for the modern theory of integral equations. Hilbert developed further Fredholm's methods for symmetric kernels, exploiting analogies with the theory of real quadratic forms and thereby making clear the importance of the notion of square-integrable functions. With Hilbert's *Grundzüge einer allgemeinen Theorie der Integralgleichung*, a further step was made from the "algebra of the infinite" to the "geometry of the infinite." The contribution of Fréchet, who introduced the abstract notion of a space endowed with a distance, made it possible to transfer Euclidean geometry to the framework of what have since then been called Hilbert spaces, a basic concept in mathematics and quantum physics.

The usefulness of functional analysis in the study of quantum systems became clear in the 1950s when Kato proved the self-adjointness of atomic Hamiltonians, and Garding and Wightman formulated axioms for quantum field theory. Ever since functional analysis lies at the very heart of many approaches to quantum field theory. Applications of functional analysis stretch out to many branches of mathematics, among which are numerical

analysis, global analysis, the theory of pseudodifferential operators, differential geometry, operator algebras, noncommutative geometry, etc.

## Topological Vector Spaces

Most topological spaces one comes across in practice are metric spaces. A metric on a topological space $E$ is a map $d: E \times E \to [0, +\infty[$ which is symmetric, such that $d(u,v) = 0 \Leftrightarrow u = v$ and which verifies the triangle inequality $d(u,w) \leq d(u,v) + d(v,w)$ for all vectors $u, v, w$. A topological space $E$ is metrizable if there is a metric $d$ on $E$ compatible with the topology on $E$, in which case the balls with radius $1/n$ centered at any point $x \in E$ form a local base at $x$ – that is, a collection of neighborhoods of $x$ such that every neighborhood of $x$ contains a member of this collection. A sequence $(u_n)$ in $E$ then converges to $u \in E$ if and only if $d(u_n, u)$ converges to 0.

The Banach fixed-point theorem on a complete metric space $(E, d)$ is a useful tool in nonlinear functional analysis: it states that *a (strict) contraction on* $E$*, that is, a map* $T: E \to E$ *such that* $d(Tu, Tv) \leq k(u,v)$ *for all* $u \neq v \in E$ *and fixed* $0 < k < 1$*, has a unique fixed point* $T u_0 = u_0$. In particular, it provides local existence and uniqueness of solutions of differential equations $\mathrm{d}y/\mathrm{d}t = F(y,t)$ with initial condition $y(0) = y_0$, where $F$ is Lipschitz continuous.

Linear functional analysis starts from topological vector spaces, that is, vector spaces equipped with a topology for which the operations are continuous. A topological vector space equipped with a local base whose members are convex is said to be locally convex. Examples of locally convex spaces are normed linear spaces, namely vector spaces equipped with a norm, a concept that first arose in the work of Fréchet. A seminorm on a vector space $V$ is a map $\rho: V \to [0, \infty[$ which obeys the triangle identity $\rho(u+v) \leq \rho(u) + \rho(v)$ for any vectors $u, v$ and such that $\rho(\lambda u) = |\lambda| \rho(u)$ for any scalar $\lambda$ and any vector $u$; if $\rho(u) = 0 \Rightarrow u = 0$, it is a norm, often denoted by $\|\cdot\|$. A norm on a vector space $E$ gives rise to a translation-invariant distance function $d(u,v) = \|u - v\|$ making it a metric space.

Historically, one of the first examples of normed spaces is the space $C([0,1])$ investigated by Riesz of (real- or complex-valued) continuous functions on the interval $[0,1]$ equipped with the supremium norm $\|f\|_\infty := \sup_{x \in [0,1]} |f(x)|$. In the 1920s, the general definition of Banach space arose in connection with the works of Hahn and Banach. A normed linear space is a Banach space if it is complete as a metric space for the induced metric, $C([0,1])$ being a prototype of a Banach space. More generally, for any non-negative integer $k$, the space $C^k([0,1])$ of functions on $[0,1]$ of class $C^k$ equipped with the norm $\|f\|_k = \sum_{i=0}^k \|f^{(i)}\|_\infty$ expressed in terms of a finite number of seminorms $\|f^{(i)}\|_\infty = \sup_{x \in [0,1]} |f^{(i)}(x)|, i = 0, \ldots, k$, is also a Banach space.

The space $C^\infty([0,1])$ of smooth functions on the interval $[0,1]$ is not anymore a Banach space since its topology is described by a countable family of seminorms $\|f\|_k$ with $k$ varying in the positive integers. The metric

$$d(f,g) = \sum_{k=1}^{\infty} 2^{-k} \frac{\|f - g\|_k}{1 + \|f - g\|_k}$$

turns it into a Fréchet space, that is, a locally convex complete metric space. The space $\mathcal{S}(\mathbb{R}^n)$ of rapidly decreasing functions, which are smooth functions $f$ on $\mathbb{R}^n$ for which

$$\|f\|_{\alpha,\beta} := \sup_{x \in \mathbb{R}^n} |x^\alpha D_x^\beta f(x)|$$

is finite for any multiindices $\alpha$ and $\beta$, is also a Fréchet space with the topology given by the seminorms $\|\cdot\|_{\alpha,\beta}$. Further examples of Fréchet spaces are the space $C_0^\infty(K)$ of smooth functions with support in a fixed compact subset $K \subset \mathbb{R}^n$ equipped with the countable family of seminorms

$$\|D^\alpha f\|_{\infty, K} = \sup_{x \in K} |D_x^\alpha f(x)|, \quad \alpha \in \mathbb{N}_0^n$$

and the space $C^\infty(M, E)$ of smooth sections of a vector bundle $E$ over a closed manifold $M$ equipped with a similar countable family of seminorms. Given an open subset $\Omega = \cup_{p \in \mathbb{N}} K_p$ with $K_p, p \in \mathbb{N}$ compact subsets of $\mathbb{R}^n$, the space $\mathcal{D}(\Omega) = \cup_{p \in \mathbb{N}} C_0^\infty(K_p)$ equipped with the inductive limit topology – for which a sequence $(f_n)$ in $\mathcal{D}(\Omega)$ converges to $f \in \mathcal{D}(\Omega)$ if each $f_n$ has support in some fixed compact subset $K$ and $(D^\alpha f_n)$ converges uniformly to $D^\alpha f$ on $K$ for each mutilindex $\alpha$ – is a locally convex space.

Among Banach spaces are Hilbert spaces which have properties very similar to those of finite-dimensional spaces and are historically the first type of infinite-dimensional space to appear with the works of Hilbert at the beginning of the twentieth century. A Hilbert space is a Banach space equipped with a norm $\|\cdot\|$ that derives from an inner product, that is, $\|u\|^2 = \langle u, u \rangle$ with $\langle \cdot, \cdot \rangle$ a positive-definite bilinear (or sesquilinear according to whether the base space is real or complex) form. Hilbert spaces are fundamental building blocks in quantum mechanics; using (closed) tensor products, from a Hilbert space $H$ one builds the Fock space $\mathcal{F}(H) = \sum_{k=0}^\infty \otimes^k H$ and from there the bosonic Fock space $\mathcal{F}(H) = \sum_{k=0}^\infty \otimes_s^k H$ (where $\otimes_s$ stands for the (closed) symmetrized tensor product) as well

as the fermionic Fock space $\mathcal{F}(H)=\sum_{k=0}^{\infty}\Lambda^k H$ (where $\Lambda^k$ stands for the antisymmetrized (closed) tensor product).

A prototype of Hilbert space is the space $l_2(\mathbb{Z})$ of complex-valued sequences $(u_n)_{n\in\mathbb{Z}}$ such that $\sum_{n\in\mathbb{Z}}|u_n|^2$ is finite, which is already implicit in Hilbert's Grundzügen. Shortly afterwords, Riesz and Fischer, with the help of the integration tool introduced by Lebesgue, showed that the space $L^2(]0,1[)$ (first introduced by Riesz) of square-summable functions on the interval $]0,1[$, that is, functions $f$ such that

$$\|f\|_{L^2}=\left(\int_0^1|f(x)|^2\,\mathrm{d}x\right)^{1/2}$$

is finite, provides an example of Hilbert space. These were then further generalized to spaces $L^p(]0,1[)$ of $p$-summable $(1\le p<\infty)$ functionals on $]0,1[$ (i.e., functions $f$ such that

$$\|f\|_{L^p}=\left(\int_0^1|f(x)|^p\,\mathrm{d}x\right)^{1/p}$$

is finite), which are not Hilbert unless $p=2$ but which provide further examples of Banach spaces, the space $L^\infty(]0,1[)$ of functions on $]0,1[$ bounded almost everywhere with respect to the Lebesgue measure, offering yet another example of Banach space.

In 1936, Sobolev gave a generalization of the notion of function and their derivatives through integration by parts, which led to the so-called Sobolev spaces $W^{k,p}(]0,1[)$ of functions $f\in L^p(]0,1[)$ with derivatives up to order $k$ lying in $L^p(]0,1[)$, obtained as the closure of $C^\infty(]0,1[)$ for the norm

$$f\mapsto\|f\|_{W^{k,p}}=\left(\sum_{j=1}^{k}\|\partial^j f\|_{L^p}^p\right)^{1/p}$$

(for $p=2, W^{k,p}(]0,1[)$ is a Hilbert space often denoted by $H^k(]0,1[)$. They differ from the Sobolev spaces $W_0^{k,p}(]0,1[)$, which correspond to the closure of the set $\mathcal{D}(]0,1[)$ for the norm $f\mapsto\|f\|_{W^{k,p}}$; for example, an element $u\in W^{1,p}(]0,1[)$ lies in $W_0^{1,p}(]0,1[)$ if and only if it vanishes at 0 and 1, that is, if and only if it satisfies Dirichlet-type boundary conditions on the boundary of the interval. Similarly, one defines Sobolev spaces $W_0^{k,p}(\mathbb{R})=W^{k,p}(\mathbb{R})$ on $\mathbb{R}$, Sobolev spaces $W^{k,p}(\Omega)$ and $W_0^{k,p}(\Omega)$ on open subsets $\Omega\subset\mathbb{R}^n$ and using a partition of unity on a closed manifold $M$, Sobolev spaces $H^k(M,E)=W^{k,2}(M,E)$ of sections of vector bundles $E$ over $M$. Using the Fourier transform (discussed later), one can drop the assumption that $k$ be an integer and extend the notion of Sobolev space to define $W^{s,p}(\Omega)$ and $H^s(M,E)$ with $s$ any real number.

Sobolev spaces arise in many areas of mathematics; one central example in probability theory is the Cameron–Martin space $H^1([0,t])$ embedded in the Wiener space $C([0,t])$. This embedding is a particular case of more general Sobolev embedding theorems, which embed (possibly continuously, sometimes even compactly (the notion of compact operator is discussed in a later section)) $W^{k,p}$-Sobolev spaces in $L^q$-spaces with $q>p$ such as the continuous inclusion $W^{k,p}(\mathbb{R}^n)\subset L^q(\mathbb{R}^n)$ with $1/q=1/p-k/n$, or in $C^l$-spaces with $l\le k$ such as, for a bounded open and regular enough subset $\Omega$ of $\mathbb{R}^n$ and for any $s\ge l+n/p$ with $p>n$, the continuous inclusion $W^{s,p}(\Omega)\subset C^l(\bar\Omega)$ (the set of functions in $C^l(\Omega)$ such that $D^\alpha u$ can be continuously extended to the closure $\bar\Omega$ for all $|\alpha|\le l$). Sobolev embeddings have important applications for the regularity of solutions of partial differential equations, when showing that weak solutions one constructs are in fact smooth. In particular, on an $n$-dimensional closed manifold $M$ for $s>l+n/2$, the Sobolev space $H^s(M,E)$ can be continuously embedded in the space $C^l(M,E)$ of sections of $E$ of class $C^l$, which in particular implies that the solutions of a hypoelliptic partial differential equation $Au=v$ with $v\in L^2(M,E)$ are smooth, as for example in the case of solutions of the Seiberg–Witten equations.

## Duality

The concept of duality (in a topological sense) was initiated at the beginning of the twentieth century by Hadamard, who was looking for continuous linear functionals on the Banach space $C(I)$ of continuous functions on a compact interval $I$ equipped with a uniform topology. It is implicit in Hilbert's theory and plays a central part in Riesz' work, who managed to express such continuous functionals as Stieltjes integrals, one of the starting points for the modern theory of integration.

The topological dual of a topological vector space $E$ is the space $E^*$ of continuous linear forms on $E$ which, when $E$ is a normed space, can be equipped with the dual norm $\|L\|_{E^*}=\sup_{u\in E,\,\|u\|\le 1}|L(u)|$.

Dual spaces often provide a receptacle for singular objects; any of the functions $f\in L^p(\mathbb{R}^n)(p\ge 1)$ and the delta-function at point $x\in\mathbb{R}^n$, $\delta_x:f\mapsto f(x)$, all lie in the space $\mathcal{S}'(\mathbb{R}^n)$ dual to $\mathcal{S}(\mathbb{R}^n)$ of tempered distributions on $\mathbb{R}^n$, which is itself contained in the space $\mathcal{D}'(\mathbb{R}^n)$ of distributions dual to $\mathcal{D}(\mathbb{R}^n)$. Furthermore, the topological dual $E^*$ of a nuclear space $E$ contains the support of a probability

measure with characteristic function (see the next section) given by a continuous positive-definite function on $E$. Among nuclear spaces are projective limits $E = \cap_{p\in\mathbb{N}} H_p$ (a sequence $(u_n) \in E$ converges to $u \in E$ whenever it converges to $u$ in each $H_p$) of countably many nested Hilbert spaces $\cdots \subset H_p \subset H_{p-1} \subset \cdots \subset H_0$ such that the embedding $H_p \subset H_{p-1}$ is a trace-class operator (see the section "Operator algebras"). If $H_p$ is the closure of $E$ for the norm $\|\cdot\|_p$, the topological dual $E'$ of $E$ for the norm $\|\cdot\|_0$ is an inductive limit $E' = \cup_{p\in\mathbb{N}_0} H_{-p}$, where $H_{-p}$ are the dual (with respect to $\|\cdot\|_0$) Hilbert spaces with norm $\|\cdot\|_{-p}$ (a sequence $(u_n) \in E'$ converges to $u \in E'$ whenever it lies in some $H_{-p}$ and converges to $u$ for the topology of $H_{-p}$) and we have

$$\begin{aligned} E &\subset \cdots \subset H_p \subset H_{p-1} \subset \cdots \subset H_0 \\ &= H_0' \subset H_{-1} \subset \cdots \subset H_{-p} \subset \cdots \subset E' \end{aligned}$$

As a result of the theory of elliptic operators on a closed manifold, the Fréchet space $C^\infty(M, E)$ of smooth sections of a vector bundle over a closed manifold $M$ is nuclear as the inductive limit of countably many Sobolev spaces $H^p(M, E)$ with $L^2$-dual given by the projective limit of countably many Sobolev spaces $H^{-p}(M, E)$.

The existence of nontrivial continuous linear forms on a normed linear space $E$ is ensured by the Hahn–Banach theorem, which asserts that *for any closed linear subspace $F$ of $E$, there is a nonvanishing continuous linear form that vanishes on $F$*. When the space is a Hilbert space $(H, \langle\cdot,\cdot\rangle_H)$, it follows from the Riesz–Fréchet theorem that any continuous linear form $L$ on $H$ is represented in a unique way by a vector $v \in H$ such that $L(u) = \langle v, u\rangle_H$ for all $u \in H$, thus relating the dual pairing on the left with the Hilbert inner product on the right and identifying the topological dual $H^*$ with $H$.

The strong topology induced by the norm $\|\cdot\|$ on a normed vector space $E$ – that is, the topology in which a sequence $(u_n)$ converges to $u$ whenever $\|u_n - u\| \to 0$ – is too refined to have compact sets when $E$ is infinite dimensional since the compactness of the unit ball in $E$ for the strong topology characterizes finite-dimensional spaces. Since compact sets are useful for existence theorems, one is inclined to weaken the topology: the weak topology on $E$ – which coincides with the strong topology when $E$ is finite dimensional and for which a sequence $(u_n)$ converges to $u$ if and only if $L(u_n) \to L(u)\ \forall L \in E^*$ – has compact unit ball if and only if $E$ is reflexive or, in other words, if $E$ can be canonically identified with its double dual $(E^*)^*$. For $1 < p < \infty$, given an open subset $\Omega \subset \mathbb{R}^n$, the topological dual of $L^p(\Omega)$ can be identified via the Riesz representation with $L^{p^*}(\Omega)$ with $p^*$ conjugate to $p$, that is, $1/p + 1/p^* = 1$ and $L^p(\Omega)$ is reflexive, whereas the topological duals of $W^{s,p}(\Omega)$ and $W_0^{s,p}(\Omega)$ both coincide with $W_0^{-s,p^*}(\Omega)$ so that only $W_0^{s,p}(\Omega)$ is reflexive. Neither $L^1(\Omega)$ nor its topological dual $L^\infty(\Omega)$ is reflexive since $L^1(\Omega)$ is strictly contained in the topological dual of $L^\infty(\Omega)$ for there are continuous linear forms $L$ on $L^\infty(\Omega)$ that are not of the form

$$L(u) = \int_\Omega uv\ \forall u \in L^\infty(\Omega) \quad \text{with } v \in L^1(\Omega)$$

Similarly, the topological dual $E^*$ of a normed linear space $E$ can be equipped with the topology induced by the dual norm $\|\cdot\|_{E^*}$ and the the weak $*$-topology, namely the weakest one for which the maps $L \mapsto L(u), u \in E$, are continuous, and the unit ball in $E^*$ is indeed compact for this topology (Banach–Alaoglu theorem).

Duality does not always preserve separability – a topological vector space is separable if it has a countable dense subspace – since $L^\infty(\Omega)$, which is not separable, is the topological dual of $L^1(\Omega)$, which is separable. However, as a consequence of the Hahn–Banach theorem, if the topological dual of a Banach space is separable then so is the original space and one has equivalence when adding the reflexivity assumption; a Banach space is reflexive and separable whenever its topological dual is. For $1 \le p < \infty, L^p(\Omega)$ and $W_0^{s,p}(\Omega)$ are separable and moreover reflexive if $p \neq 1$.

## Fourier Transform

In the middle of the eighteenth century, oscillations of a vibrating string were interpreted by Bernouilli as a limit case for the oscillation of $n$-point masses when $n$ tends the infinity, and Bernouilli introduced the novel idea of the superposition principle by which the general oscillation of the string should decompose in a superposition of "proper oscillations." This point of view triggered off a discussion as to whether or not an arbitrary function can be expanded as a trigonometric series. Other examples of expansions in "orthogonal functions" (this terminology actually only appears with Hilbert) had been found in the mean time in relation to oscillation problems and investigations on heat theory, but it was only in the nineteenth century, with the works of Fourier and Dirichlet, that the superposition problem was solved.

Separable Hilbert spaces can be equipped with a countable orthonormal system $\{e_n\}_{n\in\mathbb{Z}}$ ($\langle e_n, e_m\rangle_H = \delta_{mn}$ with $\langle\cdot,\cdot\rangle_H$ the scalar product on $H$) which is

complete, that is, any vector $u \in H$ can be expanded in this system in a unique way $u = \sum_{n\in\mathbb{Z}} \hat{u}_n e_n$ with Fourier coefficients $\hat{u}_n = \langle u, e_n\rangle$. The latter obey Parseval's relation $\sum_{n\in\mathbb{Z}} |\hat{u}_n|^2 = \|u\|^2$ (where $\|\cdot\|$ is the norm associated with $\langle\cdot,\cdot\rangle$), and the Fourier transform $u \mapsto (\hat{u}(n))_{n\in\mathbb{Z}}$ gives rise to an isometric isomorphism between the separable Hilbert space $H$ and the Hilbert space $l^2(\mathbb{Z})$ of square-summable sequences of complex numbers. In particular, the space $L^2(S^1)$ of $L^2$-functions on the unit circle $S^1 = \mathbb{R}/\mathbb{Z}$ with its usual Haar measure $\mathrm{d}t$ is separable with complete orthonormal system $t \mapsto e_n(t) = \mathrm{e}^{2\mathrm{i}\pi nt}, n \in \mathbb{Z}$ and the Fourier transform

$$u \mapsto \left( t \mapsto \hat{u}(n) = \int_0^1 \mathrm{e}^{-2\mathrm{i}\pi nt} u(t)\,\mathrm{d}t \right)_{n\in\mathbb{Z}}$$

identifies it with the space $l^2(\mathbb{Z})$. Under this identification, the Hilbert subspace $l^2(\mathbb{N})$ obtained as the range in $l^2(\mathbb{Z})$ of the projection $p_+ : (u)_{n\in\mathbb{Z}} \mapsto (u_n)_{n\in\mathbb{N}}$ corresponds to the Hardy space $\mathcal{H}^2(S^1)$.

The Fourier transform extends to the space $\mathcal{S}(\mathbb{R}^n)$, sending a function $f \in \mathcal{S}(\mathbb{R}^n)$ to the map

$$\xi \mapsto \hat{f}(\xi) = \frac{1}{\sqrt{(2\pi)^n}} \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}\xi\cdot x} f(x)\,\mathrm{d}x$$

and maps $\mathcal{S}(\mathbb{R}^n)$ onto itself linearly and continuously with continuous inverse $f \mapsto \hat{f}(-\xi)$. When $n = 1$, the Poisson formula relates $f \in \mathcal{S}(\mathbb{R})$ with its Fourier transform $\hat{f}$ by $\sum_{n=-\infty}^{\infty} f(2\pi n) = \sum_{n=-\infty}^{\infty} \hat{f}(n)$.

Since Fourier transformation turns (up to a constant multiplicative factor) differentiation $D_\xi^\alpha$ for a multiindex $\alpha = (\alpha_1, \ldots, \alpha_n)$ into multiplication by $\xi^\alpha = \xi_1^{\alpha_1} \cdots \xi_n^{\alpha_n}$, it can be used to define $W^{s,p}$-Sobolev spaces with $s$ a real number as the space of $L^p$-functions with finite Sobolev norms $\|u\|_{W^{s,p}} = (\int |(1+|\xi|)^s \hat{u}(\xi)|^p)^{1/p}$ (which coincide with the ones defined previously when $s = k$ is a non-negative integer).

Fourier transforms are also used to describe a linear pseudodifferential operator $A$ (see next two sections where the notions of bounded and unbounded linear operator are discussed) of order $a$ acting on smooth functions on an open subset $U$ of $\mathbb{R}^n$ in terms of its symbol $\sigma_A$ – a smooth map $\sigma$ on $U \times \mathbb{R}^n$ with compact support in $x$ such that for any multi-indices $\alpha, \beta \in \mathbb{N}_0^n$, there is a constant $C_{\alpha,\beta}$ with

$$|D_x^\alpha D_\xi^\beta \sigma(x,\xi) \le C_{\alpha,\beta}(1+|\xi|)^{a-|\beta|}$$

for any $\xi \in \mathbb{R}^n$ – by

$$(Af)(x) = \frac{1}{\sqrt{(2\pi)^n}} \int_{\mathbb{R}^n} \mathrm{e}^{-\mathrm{i}x\cdot\xi} \sigma_A(x,\xi) \hat{f}(\xi)\,\mathrm{d}\xi$$

Fourier transform maps a Gaussian function $x \mapsto \mathrm{e}^{-(1/2)\lambda|x|^2}$ on $\mathbb{R}^n$, where $\lambda$ is a nonzero scalar, to another Gaussian function $\xi \mapsto \mathrm{e}^{-(1/2)\lambda^{-1}|\xi|^2}$ (up to a nonzero multiplicative factor), a starting point for $T$-duality in string theory. More generally, the characteristic function

$$\hat{\mu}(\xi) := \int_H \mathrm{e}^{\mathrm{i}\langle x,\xi\rangle_H} \mu(\mathrm{d}x)$$

of a Gaussian probability measure $\mu$ with covariance $C$ on a Hilbert space $H$ is the function $\xi \mapsto \mathrm{e}^{-(1/2)\langle \xi, C\xi\rangle_H}$. Such probability measures typically arise in Euclidean quantum field theory; in axiomatic quantum field theory, the analyticity properties of $n$-point functions can be derived from the Wightman axioms using Fourier transforms. Thus, Fourier transformation underlies many different aspects of quantum field theory.

## Fredholm operators

A complex-valued continuous function $K$ on $[0,1] \times [0,1]$ gives rise to an integral operator

$$A : f \to \int_0^1 K(x,y) f(y)\,\mathrm{d}y$$

on complex-valued continuous functions on $[0,1]$ (equipped with the supremum norm $\|\cdot\|_\infty$) with the following upper bound property:

$$\|Af\|_\infty \le \mathrm{Sup}_{[0,1]\times[0,1]} |K(x,y)|\, \|f\|_\infty$$

In other words, $A$ is a bounded linear operator with norm bounded from above by $\sup_{[0,1]\times[0,1]} |K(x,y)|$; a linear operator $A : E \to F$ from a normed linear space $(E, \|\cdot\|_E)$ to a normed linear space $(F, \|\cdot\|_F)$ is bounded (or continuous) if and only if its (operator) norm $|||A||| := \sup_{\|u\|_E \le 1} \|Au\|_F$ is bounded.

An integral operator

$$A : f \to \int_0^1 K(x,y) f(y)\,\mathrm{d}y$$

defined by a continuous kernel $K$ is, moreover, compact; a compact operator is a bounded operator of normed spaces that maps bounded sets to a precompact sets, that is, to sets whose closure is compact. Other examples of compact operators on normed spaces are finite-rank operators, operators with finite-dimensional range. In fact, any compact operator on a separable Hilbert space can be approximated in the topology induced by the operator norm $|||\cdot|||$ by a sequence of finite-rank operators.

Inspired by the work of Volterra, who, in the case of the integral operator defined above, produced

continuous solutions $\phi = (I - A)^{-1} f$ of the equation $f = (I - A)\phi$ for $f \in C([0,1])$, Fredholm in 1900 (*Sur une classe d'*équations *fonctionnelles*) studied the equation $f = (I - \lambda A)\phi$, introducing a complex parameter $\lambda$. He proved what is since then called the Fredholm alternative, which states that either the equation $f = (I - \lambda A)\phi$ has a unique solution for every $f \in C([0,1])$ or the corresponding homogeneous equation $(I - \lambda A)\phi = 0$ has nontrivial solutions. In modern language, it means that the resolvent $R(A,\mu) = (A - \mu I)^{-1}$ of a compact linear operator $A$ is surjective if and only if it is injective. The Fredholm alternative is a powerful tool to solve partial differential equations among which the Dirichlet problem, the solutions of which are harmonic functions $u$ (i.e., $\Delta u = 0$, where $\Delta = -\sum_{i=1}^{n} \partial^2 u/\partial x_i^2$) on some domain $\Omega \in \mathbb{R}^n$ with Dirichlet boundary conditions $u_{|\partial\Omega} = f$, where $f$ is a continuous function on the boundary $\partial\Omega$. The Dirichlet problem has geometric applications, in particular to the nonlinear Plateau problem, which minimizes the area of a surface in $\mathbb{R}^d$ with given boundary curves and which reduces to a (linear) Dirichlet problem.

The operator $B = I - A$ built from the compact operator $A$ is a particular Fredholm operator, namely a bounded linear operator $B : E \to F$ which is invertible "up to compact operators," that is, such that there is a bounded linear operator $C : F \to E$ with both $BC - I_F$ and $CB - I_E$ compact. A Fredholm operator $B$ has a finite-dimensional kernel Ker $B$ and when $(E, \langle\cdot,\cdot\rangle_E)$ and $(F, \langle\cdot,\cdot\rangle_F)$ are Hilbert spaces its cokernel Ker $B^*$, where $B^*$ is the adjoint of $B$ defined by

$$\langle B u, v\rangle_F = \langle u, B^* v\rangle_E \quad \forall u \in E,\ \forall v \in F$$

is also finite dimensional, so that it has a well-defined index $\mathrm{ind}(B) = \dim(\mathrm{Ker}\, B) - \dim(\mathrm{Ker}\, B^*)$, a starting point for index theory. Töplitz operators $T_\phi$, where $\phi$ is a continuous function on the unit circle $S^1$, provide first examples of Fredholm operators; they act on the Hardy space $\mathcal{H}^2(S^1)$ by

$$T_{e_{-n}}\left(\sum_{m\geq 0} a_m e_m\right) = \sum_{m \geq 0} a_{m+n} e_m$$

under the identification $\mathcal{H}^2(S^1) \simeq l^2(\mathbb{N}) \subset l^2(\mathbb{Z})$, with $l^2(\mathbb{Z})$ equipped with the canonical complete orthonormal basis $(e_n, n \in \mathbb{Z})$. The Fredholm index $\mathrm{ind}(T_{e_{-n}})$ is exactly the integer $n$ so that the index of its adjoint is $-n$, as a consequence of which the index map from Fredholm operators to integers is onto.

## One-Parameter (Semi) groups

Unlike in the finite-dimensional situation, a linear operator $A : E \to F$ between two normed linear spaces $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ is not expected to be bounded. Unbounded operators arise in partial differential equations that involve differential operators such as the Laplacian $\Delta$ on an open subset $\Omega \subset \mathbb{R}^n$. The following equations provide fundamental examples of partial differential equations which arose over time from the study of various problems in mathematical physics with the works of Poisson, Fourier, and Cauchy:

$$\begin{aligned} \Delta u &= 0 \quad \text{Laplace equation} \\ \frac{\partial^2 t}{\partial t^2} + \Delta u &= 0 \quad \text{wave equation} \\ \frac{\partial u}{\partial t} + \Delta u &= 0 \quad \text{heat equation} \end{aligned}$$

and later the Schrödinger equation in quantum mechanics:

$$\mathrm{i}\frac{\partial u}{\partial t} = \Delta u$$

where $t$ is a time parameter.

An unbounded linear operator on an infinite-dimensional normed space is usually defined on a domain $D(A)$ which is strictly contained in $E$. The Laplacian $\Delta$ is defined on the dense domain $D(A) = H^2(\mathbb{R}^n)$ in $L^2(\mathbb{R}^n)$; it defines a bounded operator from $H^2(\mathbb{R}^n)$ to $L^2(\mathbb{R}^n)$ but does not extend to a bounded operator on $L^2(\mathbb{R}^n)$. Like this operator, most unbounded operators $A : E \to F$ one comes across have dense domain $D(A)$ in $E$ and are closed, that is, their graph $\{(u, Au), u \in D(A)\}$ is closed as a subset of the normed linear space $E \times F$. When not actually closed, they can be closable, that is, they can have a closed extension called the closure of the operator. By the closed-graph theorem, when $E$ and $F$ are Banach spaces, a linear operator $A : E \to F$ is continuous whenever its graph is closed, as a consequence of which a closed linear operator $A : E \to F$ defined on a dense domain is bounded provided its domain coincides with the whole space.

For a closed operator $A : E \to F$ with dense domain $D(A)$, when $E$ and $F$ are Hilbert spaces equipped with inner products $\langle\cdot,\cdot\rangle_E$ and $\langle\cdot,\cdot\rangle_F$, the adjoint $A^*$ of $A$ is defined on its domain $D(A^*)$ by

$$\langle Au, v\rangle_F = \langle u, A^* v\rangle_E \quad \forall (u,v) \in D(A) \times D(A^*)$$

A self-adjoint operator $A$ with domain $D(A)$ is one for which $D(A) = D(A^*)$ and $A = A^*$; the Laplacian $\Delta$ on $\mathbb{R}^n$ is self-adjoint on the Sobolev space $H^2(\mathbb{R}^n)$ but it is only essentially self-adjoint on the dense domain $\mathcal{D}(\mathbb{R}^n)$, the latter meaning that its closure is self-adjoint.

Unbounded self-adjoint operators can arise as generators of one-parameter semigroups of bounded

operators. A one-parameter family of bounded operators $T_t, t \geq 0$ ($T_t, t \in \mathbb{R}$) on a Hilbert space $H$ is a semigroup (resp. group) if $T_s T_t = T_{t+s}$ $\forall t, s \geq 0$ (resp. $\forall t, s \in \mathbb{R}$) and it is strongly continuous (or simply continuous) if $\lim_{t \to t_0} T_t u = T_{t_0} u$ at any $t_0 \geq 0$ (resp. $t_0 \in \mathbb{R}$) and for any $u \in H$.

Stones' theorem sets up a one-to-one correspondence between continuous one-parameter unitary ($U_t^* U_t = U_t U_t^* = I$) groups $U_t, t \in \mathbb{R}$ on a Hilbert space such that $U_0 = \mathrm{Id}$ and self-adjoint operators $A$ obtained as infinitesimal generators, that is, as the strong limit

$$Au = \lim_{t \to 0} \frac{U_t u - u}{t}, \quad u \in H$$

of $U_t, t \in \mathbb{R}$, which in a compact form reads $U_t = \mathrm{e}^{\mathrm{i}tA}$. An important example in quantum mechanics is $U_t = \mathrm{e}^{\mathrm{i}tH} U_0, t \in \mathbb{R}$ with $H$ a self-adjoint Hamiltonian, which solves the Schrödinger equation $\mathrm{d}/\mathrm{d}tu = \mathrm{i}Hu$. The Lie–Trotter formula, which has important applications for Feynman path integrals, expresses the unitary semigroup generated by $A + B$, where $A$, $B$, and $A + B$ are self-adjoint on their respective domains as a strong limit

$$\mathrm{e}^{\mathrm{i}t(A+B)} = \lim_{t \to \infty} \left( \mathrm{e}^{\frac{\mathrm{i}tA}{n}} \mathrm{e}^{\frac{\mathrm{i}tB}{n}} \right)^n$$

On the other hand, positive operators on a Hilbert space $(H, \langle \cdot, \cdot \rangle_H)$ – that is, $A$ self-adjoint and such that $\langle Au, u \rangle_H \geq 0 \quad \forall u \in D(A)$ – generate one-parameter semigroups $T_t = \mathrm{e}^{-tA}, t \geq 0$. Hille and Yosida proved that on a Hilbert space, strongly continuous contraction (i.e., $|||T_t||| \leq 1 \quad \forall t > 0$) semigroups such that $T_0 = \mathrm{Id}$ are in one-to-one correspondence with densely defined positive operators $A : D(A) \subset H \to H$ that are maximal (i.e., $I + A$ is onto), obtained as (minus the) infinitesimal generators

$$-Au = \lim_{t \to 0} \frac{T_t u - u}{t}, \quad u \in H$$

of the corresponding semigroups. Similarly, a positive densely defined self-adjoint operator $A$ on a Hilbert space $H$ gives rise to a densely defined closed symmetric sesquilinear form $(u, v) \mapsto \langle \sqrt{A} u, \sqrt{A} v \rangle_H$ (see next section for a definition of $\sqrt{A}$; $\langle \cdot, \cdot \rangle_H$ is the scalar product on $H$) and this map yields a one-to-one correspondence between operators and sesquilinear forms on $H$ with the aforementioned properties, one of the starting points for the theory of Dirichlet forms. To a probability measure $\mu$ on a separable Banach space $E$, one can associate a densely defined closed symmetric sesquilinear form (it is in fact a Dirichlet form) on a Hilbert space $H$ such that $E^* \subset H^* = H \subset E$, which in the particular case of the standard Wiener measure $\mu$ on the Wiener space $E = C([0, t])$ and with Hilbert space given by the Cameron–Martin space $H = H^1([0, t])$, is the bilinear form

$$(u, v) \mapsto \int \langle \bar{\nabla} u, \bar{\nabla} v \rangle_H$$

with $\bar{\nabla}$ the (closed) gradient of Malliavin calculus.

The operator $-\Delta$, where $\Delta$ is the Laplacian on $\mathbb{R}^n$, generates the heat-operator semigroup $\mathrm{e}^{-\Delta t}, t \geq 0$. It has a smooth kernel $K_t \in C^\infty(\mathbb{R}^n \times \mathbb{R}^n)$ defined by

$$(\mathrm{e}^{-\Delta t} f)(x) = \int_{\mathbb{R}^n} K_t(x, y) f(y) \mathrm{d}y \quad \forall f \in C_0^\infty(\mathbb{R}^n)$$

and defines a smoothing operator, an operator that maps Sobolev function to smooth function. In general, a pseudodifferential operators $A$ on an open subset $U$ of $\mathbb{R}^n$ with symbol $\sigma_A$ only has a distribution kernel

$$K_A(x, y) = \int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}\langle x - y, \xi \rangle} \sigma(\xi) \mathrm{d}\xi$$

The kernel of the inverse Laplacian $(\Delta + m^2)^{-1}$ on $\mathbb{R}^n$ (the non-negative real number $m^2$ stands for the mass) called Green's function on $\mathbb{R}^n$, plays an essential role in the theory of Feynman graphs.

## Spectral Theory

Spectral theory is the study of the distribution of the values of the complex parameter $\lambda$ for which, given a linear operator $A$ on a normed space $E$, the operator $A - \lambda I$ has an inverse and of the properties of this inverse when it exists, the resolvent $R(A, \lambda) = (A - \lambda I)^{-1}$ of $A$. The resolvent $\rho(A)$ of $A$ is the set of complex numbers $\lambda$ for which $A - \lambda I$ is invertible with densely defined bounded inverse. The spectrum $\mathrm{Sp}(A)$ of $A$ is the complement in $\mathbb{C}$ of the resolvent; it consists of a union of three disjoint sets: the set of all complex numbers $\lambda$ for which $A - \lambda I$ is not injective, called the point spectrum – such a $\lambda$ is an eigenvalue of $A$ with associated eigenfunction any $u \in D(A)$ such that $Au = \lambda u$; the set of points $\lambda$ for which $A - \lambda I$ has a densely defined unbounded inverse $R(A, \lambda)$ called the continuous spectrum; and the set of points $\lambda$ for which $A - \lambda I$ has a well-defined unbounded but not densely defined inverse $R(A, \lambda)$ called the residual spectrum.

A bounded operator has bounded spectrum and a self-adjoint operator $A$ acting on a Hilbert space has real spectrum and no residual spectrum since the range of $A - \lambda I$ is dense. As a consequence of the

Fredholm alternative, the spectrum of a compact operator consists only of point spectrum; it is countable with accumulation point at 0. A Hamiltonian of a quantum mechanichal system can have both point and continuous spectra, but its point spectrum is of special interest because the corresponding eigenfunctions are stationary states of the system. As was first pointed out by Kac ("Can you hear the shape of a drum?"), the spectrum of an operator acting on functions can reflect the geometry of the space these functions are defined on, a starting point for many interesting and far-reaching questions in differential geometry.

A self-adjoint linear operator on a Hilbert space can be described in terms of a family of projections $E_\lambda$, $\lambda \in \mathbb{R}$ via the spectral representation

$$A = \int_{\mathrm{Sp}(A)} \lambda \mathrm{d}E_\lambda$$

Given a Borel real-valued function $f$ on $\mathbb{R}$, the operator

$$f(A) = \int_{\mathrm{Sp}(A)} f(\lambda) \mathrm{d}E_\lambda$$

yields another self-adjoint operator. A positive operator $A$ on a dense domain $D(A)$ of some Hilbert space $(H, \langle \cdot, \cdot \rangle_H)$ has non-negative spectrum and for any positive real number $t$, the map $\lambda \mapsto \mathrm{e}^{-t\lambda}$ gives the associated bounded heat-operator

$$\mathrm{e}^{-tA} = \int_{\mathrm{Sp}(A)} \mathrm{e}^{-t\lambda} \mathrm{d}E_\lambda$$

while the map $\lambda \mapsto \sqrt{\lambda}$ gives rise to a positive operator $\sqrt{A}$ such that $\sqrt{A}^2 = A$.

The resolvent can also be used to define new operators

$$f(A) = \frac{1}{2\mathrm{i}\pi} \int_C f(\lambda) R(A, \lambda) \mathrm{d}\lambda$$

from a linear operator via a Cauchy-type integral along a countour $C$ around the spectrum; this way one defines complex powers $A^{-z}$ of (essentially self-adjoint) positive elliptic pseudodiffferential operators which enter the definition of the zeta-function, $z \mapsto \zeta(A, z)$, of the operator $A$. The $\zeta$-function is a useful tool to extend the ordinary determinant to $\zeta$-determinants of self-adjoint elliptic operators, thereby providing an ansatz to give a meaning to partition functions in the path integral approach to quantum field theory.

## Operator Algebras

Bounded linear operators on a Hilbert space $H$ form an algebra $\mathcal{L}(H)$ closed for the operator norm with involution given by the adjoint operation $A \mapsto A^*$; it is a $C^*$-algebra, that is, an algebra over $\mathbb{C}$ with a norm $\|\cdot\|$ and an involution $*$ such that $A$ is closed for this norm and such that $\|ab\| \leq \|a\|\|b\|$ and $\|a^* a\| = \|a\|^2$ for all $a, b \in A$ and by the Gelfand–Naimark theorem, every $C^*$-algebra is isomorphic to a sub-$C^*$-algebra of some $\mathcal{L}(H)$. The notion of spectrum extends from bounded operators to $C^*$-algebras; the spectrum $\mathrm{sp}(a)$ of an element $a$ in a $C^*$-algebra $A$ is a (compact) set of complex numbers such that $a - \lambda \cdot 1$ is not invertible. The notion of self-adjointness also extends $(a = a^*)$, and just as a self-adjoint operator $B \in \mathcal{L}(H)$ is non-negative (in which case its spectrum lies in $\mathbb{R}^+$) if and only if $B = A^* A$ for some bounded operator $A$, an element $b \in A$ is said to be non-negative if and only if $b = a^* a$ for some $a \in A$, in which case $\mathrm{sp}(a) \subset \mathbb{R}_0^+$.

The algebra $C(X)$ of continuous functions $f : X \to \mathbb{C}$ vanishing at infinity on some locally compact Hausdorff space $X$ equipped with the supremum norm and the conjugation $f \mapsto \bar{f}$ is also a $C^*$-algebra and a prototype for abelian $C^*$-algebras, since Gelfand showed that every abelian $C^*$-algebra is isometrically isomorphic to $C(X)$, with $X$ compact if the algebra is unital. To a $C^*$-algebra $A$, one can associate an abelian group $K_0(A)$ which is dual to the Grothendieck group $K^0(X)$ of isomorphism classes of vector bundles over a compact Hausdorff space $X$.

Compact operators on a Hilbert space $H$ form the only proper two-sided ideal $\mathcal{K}(H)$ of the $C^*$-algebra $\mathcal{L}(H)$ which is closed for the operator norm topology on $\mathcal{L}(H)$. The quotient $\mathcal{L}(H)/\mathcal{K}(H)$ is called the Calkin space, after Calkin, who classified all two-sided ideals in $\mathcal{L}(H)$ for a separable Hilbert space $H$; one can set up a one-to-one correspondence between such ideals and certain sequence spaces. Corresponding to the Banach space $l^1(\mathbb{Z})$ of complex-valued sequences $(u_n)$ such that $\sum_{n \in \mathbb{N}} |u_n| < \infty$, is the $*$-ideal $\mathcal{I}_1(H)$ of trace-class operators. The trace $\mathrm{tr}(A) = \sum_{n \in \mathbb{Z}} \langle A\, e_n, e_n \rangle_H$ of a negative operator $A \in \mathcal{L}(H)$ lies in $[0, +\infty]$ and is independent of the choice of the complete orthonormal basis $\{e_n, n \in \mathbb{Z}\}$ of $H$ equipped with the inner product $\langle \cdot, \cdot \rangle_H$. $\mathcal{I}_1(H)$ is the Banach space of bounded linear operators on $H$ such that $\|A\|_1 = \mathrm{tr}(|A|)$ is bounded. Given an (esssentially self-adjoint) positive differential operator $D$ of order $d$ acting on smooth functions on a closed $n$-dimensional Riemannian manifold $M$, its complex power $D^{-z}$ is a trace class on the space of $L^2$-functions on $M$ provided $\mathrm{Re}(z) > n/d$ and the corresponding trace $\mathrm{tr}(D^{-z})$ extends to a meromorphic function on the whole plane, the $\zeta$-function $\zeta(D, z)$ which is holomorphic at 0.

More generally, Banach spaces $l^p(\mathbb{Z})$, $1 \leq p < \infty$, of complex-valued sequences $(u_n)_{n\in\mathbb{Z}}$ such that $\sum_{n\in\mathbb{Z}} |u_n|^p < \infty$ relate to Schatten ideals $\mathcal{I}_p(H)$, $1 \leq p < \infty$, where $\mathcal{I}_p(H)$ is the Banach space of bounded linear operators on $H$ such that $\|A\|_p = (\mathrm{tr}(|A|^p))^{1/p}$ is bounded. Just as all $l^p$-sequences converge to 0, the Schatten ideals $\mathcal{I}_p(H)$ all lie in $\mathcal{K}(H)$ and we have $\cdots \subset \mathcal{I}_{p+1}(H) \subset \mathcal{I}_p(H) \subset \cdots \subset \mathcal{K}(H)$.

Compact operators and Schatten ideals are useful to extend index theory to a noncommutative context; a Fredholm module $(H, F)$ over an involutive algebra $A$ is given by an involutive representation $\pi$ of $A$ in a Hilbert space $H$ and a self-adjoint bounded linear operator $F$ on $H$ such that $F^2 = \mathrm{Id}_H$ and the operator brackets $[F, \pi(a)]$ are compact for all $a \in A$. To a $p$-summable Fredholm module $(H, F)$, that is, $[F, \pi(a)] \in \mathcal{I}_p(H)$ for all $a \in A$, one associates a representative $\tau$ of the Chern character $\mathrm{ch}^*(H, F)$ given by a cyclic cocycle on $A$, which pairs up with $K$-theory to build an integer-valued index map $\tau$ on $K$-theory.

Schatten ideals are also useful to investigate the geometry of infinite-dimensional spaces such as loop groups, for which the Hilbert–Schmidt operators (operators in $\mathcal{I}_2(H)$ are also called Hilbert–Schmidt operators) are particularly useful. A Hölder-type inequality shows that the product of two Hilbert–Schmidt operators is trace-class. Moreover, for any two Hilbert–Schmidt operators $A$ and $B$, the "cyclicity property" that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ holds, and the sesquilinear form $(A, B) \mapsto \mathrm{tr}(AB^*)$ makes $\mathcal{L}_2(H)$ a Hilbert space.

## Further Reading


Adams R (1975) *Sobolev Spaces*. London: Academic Press.

Dunford N and Schwartz J (1971) *Linear Operators*. Part I. General Theory. Part II. Spectral Theory. Part III. Spectral Operators. New York: Wiley.

Hille E (1972) *Methods in Classical and Functional Analysis*. London: Academic Press and Addison-Wesley.

Kato T (1982) *A Short Introduction to Perturbation Theory for Linear Operators*. New York–Berlin: Springer.

Reed M and Simon B (1980) *Methods of Modern Mathematical Physics* vols. I–IV, 2nd edn. New York: Academic Press.

Riesz F and SZ-Nagy B (1968) *Leçons d'analyse fonctionnelle*. Paris: Gauthier–Villars: Budapest Akademiai Kiado.

Rudin W (1994) *Functional Analysis*, 2nd edn. New York: International Series in Pure and Applied Mathematics.

Yosida K (1980) *Functional Analysis*, 6th edn. Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen Band vol. 132. Berlin–New York: Springer.


# Introductory Article: Minkowski Spacetime and Special Relativity

**G L Naber**, Drexel University, Philadelphia, PA, USA



## Introduction

Minkowski spacetime is generally regarded as the appropriate mathematical context within which to formulate those laws of physics that do not refer specifically to gravitational phenomena. Here we shall describe this context in rigorous terms, postulate what experience has shown to be its correct physical interpretation, and illustrate by means of examples its appropriateness for the formulation of physical laws.

## Minkowski Spacetime and the Lorentz Group

Minkowski spacetime $\mathcal{M}$ is a four-dimensional real vector space on which is defined a bilinear form $\boldsymbol{g} : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ that is symmetric ($\boldsymbol{g}(v, w) = \boldsymbol{g}(w, v)$ for all $v, w \in \mathcal{M}$) and nondegenerate ($\boldsymbol{g}(v, w) = 0$ for all $w \in \mathcal{M}$ implies $v = 0$). Further, $\boldsymbol{g}$ has index 1, that is, there exists a basis $\{e_1, e_2, e_3, e_4\}$ for $\mathcal{M}$ with

$$\boldsymbol{g}(e_a, e_b) = \eta_{ab} = \begin{cases} 1 & \text{if } a = b = 1, 2, 3 \\ -1 & \text{if } a = b = 4 \\ 0 & \text{if } a \neq b \end{cases}$$

$\boldsymbol{g}$ is called a Lorentz inner product for $\mathcal{M}$ and any basis of the type just described is an orthonormal basis for $\mathcal{M}$. We shall often write $v \cdot w$ for the value $\boldsymbol{g}(v, w)$ of $\boldsymbol{g}$ on $(v, w) \in \mathcal{M} \times \mathcal{M}$. A vector $v \in \mathcal{M}$ is said to be spacelike, timelike, or null if $v \cdot v$ is positive, negative, or zero, respectively, and the set $C_N$ of all null vectors is called the null cone in $\mathcal{M}$. If $\{e_1, e_2, e_3, e_4\}$ is an orthonormal basis and if we write $v = v^1 e_1 + v^2 e_2 + v^3 e_3 + v^4 e_4 = v^a e_a$ (using the Einstein summation convention, according to which a repeated index, one subscript and one superscript, is summed over its possible values) and $w = w^b e_b$, then

$$\begin{aligned} v \cdot w &= v^1 w^1 + v^2 w^2 + v^3 w^3 - v^4 w^4 \\ &= \eta_{ab} v^a w^b \end{aligned}$$

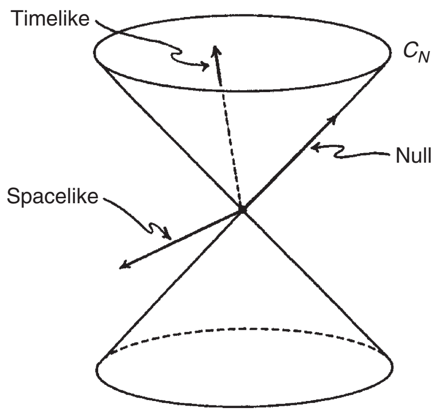



**Figure 1** Spacelike, timelike and null vectors.

In particular, $v$ is null if and only if

$$(v^4)^2 = (v^1)^2 + (v^2)^2 + (v^3)^2$$

(hence the name null "cone" for $C_N$). Timelike vectors are "inside" the null cone and spacelike vectors are "outside" (see Figure 1).

We select some orientation for the vector space $\mathcal{M}$ and will henceforth consider only oriented, orthonormal bases for $\mathcal{M}$. From the Schwartz inequality for $\mathbb{R}^3$, one can show (Naber 1992, theorem 1.3.1) that, if $v$ is timelike and $w$ is either timelike or null and nonzero, then $v \cdot w < 0$ if and only if $v^4 w^4 > 0$ in any orthonormal basis. In particular, one can define an equivalence relation on the set of all timelike vectors by decreeing that two such, $v$ and $w$, are equivalent if and only if $v \cdot w < 0$. For reasons that will emerge shortly we then say that $v$ and $w$ have the same time orientation. There are precisely two equivalence classes, one of which we select and designate future directed. Timelike vectors in the other class are then called past directed. One can show (Naber 1992, section 1.3 and corollary 1.4.5) that this classification can be extended to nonzero null vectors as well (but not to spacelike vectors). We will call an oriented, orthonormal basis time oriented if its timelike vector $e_4$ is future directed and will consider only these in what follows. An oriented, time-oriented, orthonormal basis for $\mathcal{M}$ will be called an admissible basis. If $\{e_1, e_2, e_3, e_4\}$ and $\{\hat{e}_1, \hat{e}_2, \hat{e}_3, \hat{e}_4\}$ are two such bases and if we write

$$\begin{aligned} e_b &= \Lambda^1{}_b \hat{e}_1 + \Lambda^2{}_b \hat{e}_2 + \Lambda^3{}_b \hat{e}_3 + \Lambda^4{}_b \hat{e}_4 \\ &= \Lambda^a{}_b \hat{e}_a, \quad b = 1, 2, 3, 4 \end{aligned} \qquad [1]$$

then the matrix $\Lambda = (\Lambda^a{}_b)$ ($a$ = row index, $b$ = column index) can be shown to satisfy the following three conditions (Naber 1992, section 1.3):

1. (orthogonality) $\Lambda^{\mathrm{T}} \eta \Lambda = \eta$,
   where T means transpose and

$$\eta = (\eta_{ab}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

2. (orientability) $\det \Lambda = 1$, and
3. (time orientability) $\Lambda^4{}_4 \geq 1$.

We shall refer to any $4 \times 4$ matrix $\Lambda = (\Lambda^a{}_b)$ satisfying these three conditions as a Lorentz transformation (although one often sees the adjectives "proper" and "orthochronous" appended to emphasize conditions (2) and (3), respectively). The set $\mathcal{L}$ of all such matrices forms a group under matrix multiplication that we call simply the Lorentz group. It is a simple matter to show (Naber 1992, lemma 1.3.4) from the orthogonality condition (1) that, if $\Lambda^4{}_4 = 1$, then $\Lambda$ must be of the form

$$\begin{pmatrix} & & & 0 \\ & (R^i{}_j) & & 0 \\ & & & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

where $(R^i{}_j)$ is an element of SO(3), that is, a $3 \times 3$ orthogonal matrix with determinant 1. The set $\mathcal{R}$ of all matrices of this form is a subgroup of $\mathcal{L}$ called the rotation subgroup. Although it will play no role in what we do here, it should be pointed out that in many applications (e.g., in particle physics) it is necessary to consider the larger group of transformations of $\mathcal{M}$ generated by the Lorentz group and spacetime translations ($x^a \to x^a + \Lambda^a$, for some constants $\Lambda^a, a = 1, 2, 3, 4$). This is called the inhomogeneous Lorentz group, or Poincaré group.

## Physical Interpretation

For the purpose of describing how one is to think of Minkowski spacetime and the Lorentz group physically it will be convenient to distinguish (intuitively and terminologically, if not mathematically) between a "vector" in $\mathcal{M}$ and a "point" in $\mathcal{M}$ (the "tip" of a vector). The points in $\mathcal{M}$ are called events and are to be thought of as actual physical occurrences, albeit idealized as "point events" which have no spatial extension and no duration. One might picture, for example, an instantaneous collision, or explosion, or an "instant" in the history of some point material particle or photon ("particle of light").

Events are observed and identified by the assignment of coordinates. We will be interested in coordinates assigned in a very particular way by a

very particular type of observer. Specifically, our admissible observers preside over three-dimensional, right-handed, Cartesian spatial coordinate systems, relative to which photons always move along straight lines in any direction. With a single clock located at the origin, such an observer can determine the speed, $c$, of light *in vacuo* by the so-called Fizeau procedure (emit a photon from the origin when the clock there reads $t_1$, bounce it back from a mirror located at $(x^1, x^2, x^3)$, receive the photon at the origin again when the clock there reads $t_2$ and set $c = 2\sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2/(t_2 - t_1)})$. Now place an identical clock at each spatial point and synchronize them by emitting from the origin a spherical electromagnetic wave (photons in all directions) and setting the clock whose location is $(x^1, x^2, x^3)$ to read $\sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2}/c$ at the instant the wave arrives. An observer now assigns to an event the three spatial coordinates of the location at which it occurred in his coordinate system as well as the time reading on the clock at that location at the instant the event occurred. We shall assume also that our admissible observers are inertial in the sense of Newtonian mechanics (the trajectory of a particle on which no forces act, when described in terms of the coordinates just introduced, is a point or a straight line traversed at constant speed). It is an experimental fact (and quite a remarkable one) that all of these admissible observers (whether or not they are in relative motion) agree on the numerical value of the speed of light *in vacuo* ($c \approx 3.00 \times 10^{10}\,\mathrm{cm\,s^{-1}}$). We shall exploit this fact at the outset to have all of our admissible observers measure time in units of distance by simply multiplying their time coordinates $t$ by $c$. The resulting time coordinate is denoted $x^4 = ct$. In these units all speeds are dimensionless and the speed of light *in vacuo* is 1.

In our mathematical model $\mathcal{M}$ of the world of events, this very subtle and complex notion of an admissible observer is fully identified with the conceptually very simple notion of an admissible basis $\{e_1, e_2, e_3, e_4\}$. If $x \in \mathcal{M}$ is an event and if we write $x = x^a e_a$, then $(x^1, x^2, x^3)$ are the spatial and $x^4$ is the time coordinate supplied for $x$ by the corresponding observer. If $\{\hat{e}_1, \hat{e}_2, \hat{e}_3, \hat{e}_4\}$ is another basis/observer related to $\{e_1, e_2, e_3, e_4\}$ by [1] and if we write $x = \hat{x}^a \hat{e}_a$, then

$$\hat{x}^a = \Lambda^a{}_b x^b, \quad a = 1, 2, 3, 4 \qquad [2]$$

Thus, Lorentz transformations relate the space and time coordinates supplied for any given event by two admissible observers. If $(\Lambda^a{}_b) \in \mathcal{R}$, then the two observers differ only in the orientation of their spatial coordinate axes. On the other hand, for any real number $\theta$ one can define an element $L(\theta)$ of $\mathcal{L}$ by

$$L(\theta) = \begin{pmatrix} \cosh\theta & 0 & 0 & -\sinh\theta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh\theta & 0 & 0 & \cosh\theta \end{pmatrix} \qquad [3]$$

and, if two admissible bases are related by this Lorentz transformation, then the coordinate transformation [2] becomes

$$\begin{aligned} \hat{x}^1 &= (\cosh\theta)\, x^1 - (\sinh\theta)\, x^4 \\ \hat{x}^2 &= x^2 \\ \hat{x}^3 &= x^3 \\ \hat{x}^4 &= -(\sinh\theta)\, x^1 + (\cosh\theta)\, x^4 \end{aligned} \qquad [4]$$

Letting $\beta = \tanh\theta$ (so that $-1 < \beta < 1$) and suppressing $\hat{x}^2 = x^2$ and $\hat{x}^3 = x^3$, one obtains

$$\begin{aligned} \hat{x}^1 &= \frac{1}{\sqrt{1-\beta^2}} x^1 - \frac{\beta}{\sqrt{1-\beta^2}} x^4 \\ \hat{x}^4 &= -\frac{\beta}{\sqrt{1-\beta^2}} x^1 + \frac{1}{\sqrt{1-\beta^2}} x^4 \end{aligned} \qquad [5]$$

This corresponds to two observers whose spatial axes are oriented as shown in Figure 2 with the hatted coordinate system moving along the common $x^1$-, $\hat{x}^1$-axis with speed $|\beta|$, to the right if $\beta > 0$ and to the left if $\beta < 0$.

We remark that, reverting to traditional time units, $\beta = v/c$, where $|v|$ is the relative speed of the two coordinate systems, and [5] becomes what is generally referred to as a "Lorentz transformation" in elementary expositions of special relativity, that is,

$$\begin{aligned} \hat{x}^1 &= \frac{x^1 - vt}{\sqrt{1 - v^2/c^2}} \\ \hat{t} &= \frac{t - (v/c^2)x^1}{\sqrt{1 - v^2/c^2}} \end{aligned} \qquad [6]$$

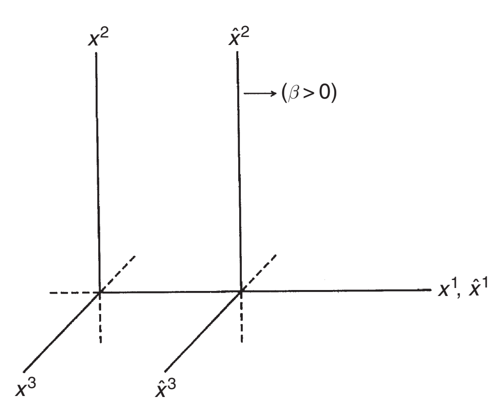



**Figure 2** Observers in standard configuration.

There is a sense in which, to understand the kinematic effects of special relativity, it is enough to restrict one's attention to the so-called special Lorentz transformations $L(\theta)$. Specifically, one can show (Naber 1992, theorem 1.3.5) that if $\Lambda \in \mathcal{L}$ is any Lorentz transformation, then there exists a real number $\theta$ and two rotations $R_1, R_2 \in \mathcal{R}$ such that $\Lambda = R_1 L(\theta) R_2$. Since $R_1$ and $R_2$ involve no relative motion, all of the kinematics is contained in $L(\theta)$. We shall explore these kinematic effects in more detail shortly.

Now suppose that $x$ and $x_0$ are two distinct events in $\mathcal{M}$ and consider the displacement vector $x - x_0$ from $x_0$ to $x$. If $\{e_1, e_2, e_3, e_4\}$ is an admissible basis and if we write $x = x^a e_a$ and $x_0 = x_0^a e_a$, then $x - x_0 = (x^a - x_0^a)e_a = \Delta x^a e_a$. If $x - x_0$ is null, then

$$(\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 = (\Delta x^4)^2$$

so the spatial separation of the two events is equal to the distance light would travel during the time lapse between the events. The same must be true in any other admissible basis since Lorentz transformations are the matrices of linear maps that preserve the Lorentz inner product. Consequently, all admissible observers agree that $x_0$ and $x$ are "connectible by a photon." They even agree as to which of the two events is to be regarded as the "emission" of the photon and which is to be regarded as its "reception" since one can show (Naber 1992, theorem 1.3.3) that, when a vector is either timelike or null and nonzero, the sign of its fourth coordinate is the same in every admissible basis (because $\Lambda^4{}_4 \geq 1$). Thus, $x^4 - x_0^4$ is either positive for all admissible observers ($x_0$ occurred before $x$) or negative for all admissible observers ($x_0$ occurred after $x$). Since photons move along straight lines in admissible coordinate systems we adopt the following terminology. If $x_0, x \in \mathcal{M}$ are such that $x - x_0$ is null, then the straight line in $\mathcal{M}$ containing $x_0$ and $x$ is called the world line of a photon in $\mathcal{M}$ and is to be thought of as the set of all events in the history of some particle of light that "experiences" both $x_0$ and $x$.

Let us now suppose instead that $x - x_0$ is timelike. Then, in any admissible basis,

$$(\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 < (\Delta x^4)^2$$

so the spatial separation of $x_0$ and $x$ is less than the distance light would travel during the time lapse between the events. In this case, one can prove (Naber 1992, section 1.4) that there exists an admissible basis $\{\hat{e}_1, \hat{e}_2, \hat{e}_3, \hat{e}_4\}$ in which $\Delta\hat{x}^1 = \Delta\hat{x}^2 = \Delta\hat{x}^3 = 0$, that is, there is an admissible observer for whom the two events occur at the same spatial location, one after the other. Thinking of this location as occupied by some material object (e.g., the observer's clock situated at that point) we find that the events $x_0$ and $x$ are both "experienced" by this material particle and that, moreover, $\sqrt{|g(x - x_0, x - x_0)|}$ is just the time lapse between the events recorded by a clock carried along by this material particle. To any other admissible observer this material particle appears "free" (not subject to forces) because it moves on a straight line with constant speed. This leads us to the following definitions. If $x_0, x \in \mathcal{M}$ are such that $x - x_0$ is timelike, then the straight line in $\mathcal{M}$ containing $x_0$ and $x$ is called the world line of a free material particle in $\mathcal{M}$ and $\sqrt{|g(x - x_0, x - x_0)|}$, usually written $\tau(x - x_0)$, or simply $\Delta\tau$, is the proper time separation of $x_0$ and $x$. One can think of $\tau(x - x_0)$ as a sort of "length" for $x - x_0$ measured, however, by a clock carried along by a free material particle that experiences both $x_0$ and $x$. It is an odd sort of length, however, since it satisfies not the usual triangle inequality, but the following "reversed" version.

**Reversed triangle inequality** (Naber 1992, theorem 1.4.2) *Let $x_0$, $x$ and $y$ be events in $\mathcal{M}$ for which $y - x$ and $x - x_0$ are timelike with the same time orientation. Then $y - x_0 = (y - x) + (x - x_0)$ is timelike and*

$$\tau(y - x_0) \geq \tau(y - x) + \tau(x - x_0) \qquad [7]$$

*with equality holding if and only if $y - x$ and $x - x_0$ are linearly dependent.*

The sense of the inequality in [7] has interesting consequences about which we will have more to say shortly.

Finally, let us suppose that $x - x_0$ is spacelike. Then, in any admissible basis

$$(\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 > (\Delta x^4)^2$$

so the spatial separation of $x_0$ and $x$ is greater than the distance light could travel during the time lapse that separates them. There is clearly no admissible observer for whom the events occur at the same location. No free material particle (or even photon) can experience both $x_0$ and $x$. However, one can show (Naber 1992, section 1.5) that, given any real number $T$ (positive, negative, or zero), one can find an admissible basis $\{\hat{e}_1, \hat{e}_2, \hat{e}_3, \hat{e}_4\}$ in which $\Delta\hat{x}^4 = T$. Some admissible observers will judge the events simultaneous, some will assert that $x_0$ occurred before $x$, and others will reverse the order. Temporal order, cause and effect, have no meaning for such pairs of events. For those admissible observers for whom the events are simultaneous ($\Delta\hat{x}^4 = 0$), the quantity $\sqrt{g(x - x_0, x - x_0)}$ is the distance between them and for this reason this quantity is called the proper spatial separation of $x_0$ and $x$ (whenever $x - x_0$ is spacelike).

For any two events $x_0, x \in \mathcal{M}, g(x - x_0, x - x_0)$ is given in any admissible basis by $(\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 - (\Delta x^4)^2$ and is called the interval separating $x_0$ and $x$. It is the closest analog in Minkowskian geometry to the (squared) length in Euclidean geometry. It can, however, assume any real value depending on the physical relationship between the events $x_0$ and $x$. Historically, of course, it was the various physical interpretations of this interval that we have just described which led Minkowski (Einstein *et al.* 1958) to the introduction of the structure that bears his name.

## Kinematic Effects

All of the well-known kinematic effects of special relativity (the addition of velocities formula, the relativity of simultaneity, time dilation, and length contraction) follow easily from what we have done. Because it eases visualization and because, as we mentioned earlier, it suffices to do so, we will limit our discussion to the special Lorentz transformations.

Let $\theta_1$ and $\theta_2$ be two real numbers and consider the corresponding elements $L(\theta_1)$ and $L(\theta_2)$ of $\mathcal{L}$ defined by [3]. Sum formulas for $\sinh\theta$ and $\cosh\theta$ imply that $L(\theta_1)L(\theta_2) = L(\theta_1 + \theta_2)$. Defining $\beta_i = \tanh\theta_i, i = 1, 2$, and $\beta = \tanh(\theta_1 + \theta_2)$, the sum formula for $\tanh\theta$ then gives

$$\beta = \frac{\beta_1 + \beta_2}{1 + \beta_1\beta_2} \qquad [8]$$

The physical interpretation is simple. One has three admissible observers whose spatial axes are related in the manner shown in **Figure 2**. If the speed of the second relative to the first is $\beta_1$ and the speed of the third relative to the second is $\beta_2$, then the speed of the third relative to the first is not $\beta_1 + \beta_2$ as a Newtonian predisposition would lead one to expect, but rather $\beta$, given by [8]. This is the relativistic addition of velocities formula.

We have seen already that, when the interval between $x_0$ and $x$ is spacelike, the events will be judged simultaneous by some admissible observers, but not by others. Indeed, if $\Delta x^4 = 0$ and the observers are related by [5], then $\Delta\hat{x}^4 = -(\beta/\sqrt{1-\beta^2})\Delta x^1 = -\beta\Delta\hat{x}^1$, which will not be zero unless $\beta = 0$ and so there is no relative motion ($\Delta\hat{x}^1$ cannot be zero since then $\Delta\hat{x}^a = 0$ for $a = 1, 2, 3, 4$ and $x = x_0$). This phenomenon is called the relativity of simultaneity and we now construct a simple geometrical representation of it.

Select two perpendicular lines in the plane to represent the $x^1$- and $x^4$-axes (the Euclidean orthogonality of the lines has no physical significance and is unnecessary, but makes the pictures easier to draw). The $\hat{x}^1$-axis will be represented by the straight line $\hat{x}^4 = 0$ which, from [5], is given by $x^4 = \beta x^1$ (in **Figure 3** we have assumed that $\beta > 0$). Similarly, the $\hat{x}^4$-axis is identified with the line $x^4 = (1/\beta)x^1$. Since Lorentz transformations leave the Lorentz inner product invariant, the hyperbolas $(x^1)^2 - (x^4)^2 = k$ coincide with $(\hat{x}^1)^2 - (\hat{x}^4)^2 = k$ and we calibrate the axes accordingly, for example, the branch of $(x^1)^2 - (x^4)^2 = 1$ with $x^1 > 0$ intersects the $x^1$-axis at the point $(x^1, x^4) = (1, 0)$ and intersects the $\hat{x}^1$-axis at the point $(\hat{x}^1, \hat{x}^4) = (1, 0)$. This necessitates a different scale on the hatted and unhatted axes, but one can show (Naber 1992, section 1.3) that, with this calibration, all coordinates can be obtained geometrically by projecting parallel to the opposite axis (e.g., the $x^4$- and $\hat{x}^4$-coordinates of an event result from projecting parallel to the $x^1$- and $\hat{x}^1$-axes, respectively).

Thus, a line of simultaneity in the hatted (respectively, unhatted) coordinates is parallel to the $\hat{x}^1$- (respectively, $x^1$-) axis so that, in general, a pair of events lying on one will not lie on the other (note, however, that these lines are "really" three-dimensional hyperplanes so what appears to be a point of intersection is actually a two-dimensional "plane of agreement", any two events in which are judged simultaneous by both observers).

For any two events whatsoever the relationship between the time lapse $\Delta\hat{x}^4$ in the hatted coordinates and the time lapse $\Delta x^4$ in the unhatted coordinates is, from [5],

$$\Delta\hat{x}^4 = -\frac{\beta}{\sqrt{1-\beta^2}}\Delta x^1 + \frac{1}{\sqrt{1-\beta^2}}\Delta x^4$$

so the two are generally not equal. Consider, in particular, two events on the world line of a point at rest in the unhatted coordinate system, for

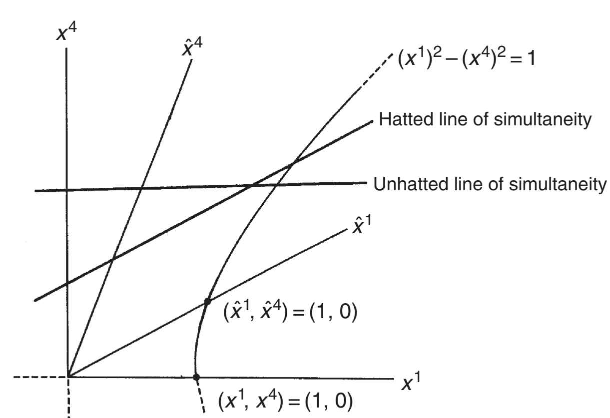



**Figure 3** Relativity of simultaneity.

example, two readings on the clock at rest at the origin in this system. Then $\Delta x^1 = 0$ so

$$\Delta \hat{x}^4 = \frac{1}{\sqrt{1-\beta^2}} \Delta x^4 > \Delta x^4$$

This effect is entirely symmetrical since, if $\Delta \hat{x}^1 = 0$, then [5] implies

$$\Delta x^4 = \frac{1}{\sqrt{1-\beta^2}} \Delta \hat{x}^4 > \Delta \hat{x}^4$$

Each observer judges the other's clocks to be running slow. This phenomenon is called time dilation and is clearly visible in the spacetime diagram in Figure 4 (e.g., both observers agree on the time reading "0" for the clock at the origin of the unhatted system, but the line $\hat{x}^4 = 1$ intersects the world line of the clock, i.e., the $x^4$-axis, at a point below $(x^1, x^4) = (0, 1)$).

We should emphasize that this phenomenon is quite "real" in the physical sense. For example, certain types of elementary particles (mesons) found in cosmic radiation are so short-lived (at rest) that, even if they could travel at the speed of light, the time required to traverse our atmosphere would be some ten times their normal life span. They should not be able to reach the earth, but they do. Time dilation "keeps them young" in the sense that what seems a normal life time to the meson appears much longer to us.

Finally, since admissible observers generally disagree on which events are simultaneous and since the only way to measure the "length" of a moving object (say, a measuring rod) is to locate its end points "simultaneously," it should come as no surprise that length, like simultaneity, and time, depends on the admissible observer measuring it. Specifically, let us consider a measuring rod lying at rest along the $\hat{x}^1$-axis of the hatted coordinate

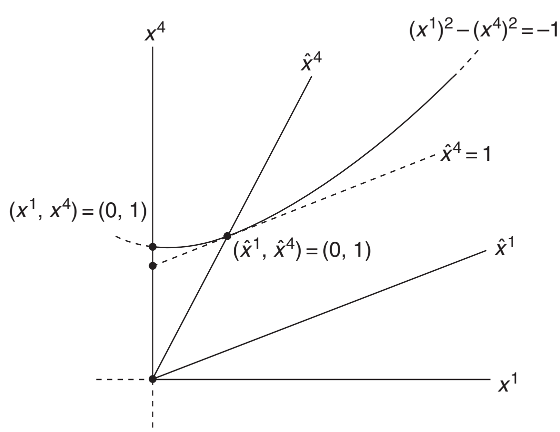



**Figure 4** Time dilation.

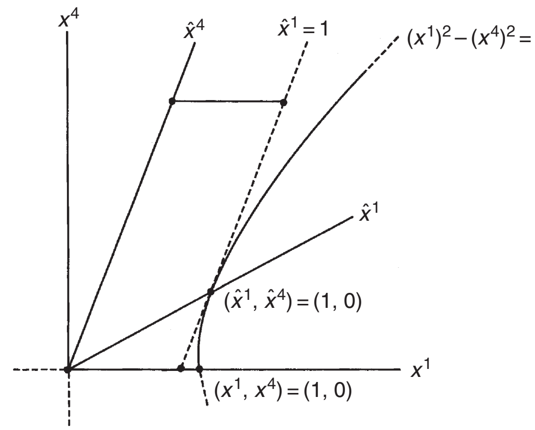



**Figure 5** Length contraction.

system. Its "length" in this coordinate system is $\Delta \hat{x}^1$. The world lines of its end points are two straight lines parallel to the $\hat{x}^4$-axis. If the unhatted observer locates two events on these world lines "simultaneously" their coordinates will satisfy $\Delta x^4 = 0$ and, by [5] $\Delta \hat{x}^1 = (1/\sqrt{1-\beta^2})\Delta x^1$ so

$$\Delta x^1 = \sqrt{1-\beta^2}\, \Delta \hat{x}^1 < \Delta \hat{x}^1$$

and the moving measuring rod appears contracted in its direction of motion by a factor of $\sqrt{1-\beta^2}$. As for time dilation, this phenomenon, known as length contraction, is entirely symmetrical, quite real, and clearly visible in a spacetime diagram (Figure 5).

## The Relativity Principle

We have found that admissible observers can disagree about some rather startling things (whether or not two events are simultaneous, the time lapse between two events even when no one thinks they are simultaneous, and the length of a measuring rod). This would be a matter of no concern at all, of course, if one could determine, in any given situation, who was really right. Surely, two events are either simultaneous or they are not and we need only sort out which admissible observer has the correct view of the situation? Unfortunately (or fortunately, depending on one's point of view) this distinction between the judgments made by different admissible observers is precisely what physics forbids.

**The relativity principle** (Einstein *et al.* 1958). *All admissible observers are completely equivalent for the formulation of the laws of physics.*

We must be clear that this is not a mathematical statement. It is rather a statement about the physical world around us and how it should be described, gleaned from observations, some of which are

complex and subtle and some of which are commonplace (a passenger in a smooth, quiet airplane traveling at constant groundspeed cannot "feel" his motion relative to the earth). It is a powerful guide for constructing the laws of relativistic physics, but even more fundamentally it prohibits us from regarding any particular admissible observer as having a privileged view of the universe. In particular, we are forbidden from attaching any objective significance to such questions as, "were the two supernovae simultaneous?", "How long did the meson survive?", and "What is the distance between the Crab Nebula and Alpha Centauri?" This is severe, but one must deal with it.

## Particles and 4-Momentum

If $I \subseteq \mathbb{R}$ is an interval, then a map $\alpha : I \to \mathcal{M}$ is a curve in $\mathcal{M}$. Relative to any admissible basis we can write

$$\alpha(\xi) = x^a(\xi)\, e_a$$

for each $\xi \in I$. We shall assume that $\alpha$ is smooth in the sense that each $x^a(\xi), a = 1, 2, 3, 4$, is infinitely differentiable ($C^\infty$) on $I$ and the velocity vector

$$\alpha'(\xi) = \frac{dx^a}{d\xi} e_a$$

is nonzero for every $\xi \in I$ (we adopt the usual custom, in a vector space, of identifying the tangent space at each point with the vector space itself). This definition of smoothness clearly does not depend on the choice of admissible basis for $\mathcal{M}$. The curve $\alpha$ is said to be spacelike, timelike, or null if

$$\alpha'(\xi) \cdot \alpha'(\xi) = \eta_{ab} \frac{dx^a}{d\xi} \frac{dx^b}{d\xi}$$

is positive, negative, or zero, respectively, for each $\xi \in I$. A timelike curve $\alpha$ for which $\alpha'(\xi)$ is future directed for each $\xi \in I$ is called a timelike world line and its image is identified with the set of all events in the history of some (not necessarily free) point material particle. If $I = [\xi_0, \xi_1]$ and $\alpha : [\xi_0, \xi_1] \to \mathcal{M}$ is a timelike world line, then the proper time length of $\alpha$ is defined by

$$\begin{aligned} L(\alpha) &= \int_{\xi_0}^{\xi_1} \sqrt{|\boldsymbol{g}(\alpha'(\xi), \alpha'(\xi))|}\, d\xi \\ &= \int_{\xi_0}^{\xi_1} \sqrt{-\eta_{ab} \frac{dx^a}{d\xi} \frac{dx^b}{d\xi}}\, d\xi \end{aligned}$$

and interpreted as the time lapse between the events $\alpha(\xi_0)$ and $\alpha(\xi_1)$ as recorded by a clock carried along by the particle whose world line is $\alpha$. This interpretation is easily motivated by writing out a Riemann sum approximation to the integral and appealing to our interpretation of the proper time separation $\Delta\tau = \sqrt{-\eta_{ab}\, \Delta x^a\, \Delta x^b}$. There are subtleties, however, both mathematical and physical (Naber 1992, section 1.4). The mathematical ones are addressed by the following result (which combines theorems 1.4.6 and 1.4.8 of Naber (1992)).

**Theorem** *Let $x_0$ and $x$ be two events in $\mathcal{M}$. Then $x - x_0$ is timelike and future directed if and only if there exists a timelike world line $\alpha : [\xi_0, \xi_1] \to \mathcal{M}$ in $\mathcal{M}$ with $\alpha(\xi_0) = x_0$ and $\alpha(\xi_1) = x$ and, in this case,*

$$L(\alpha) \leq \tau(x - x_0) \qquad [9]$$

*with equality holding if and only if $\alpha$ is a parametrization of a timelike straight line.*

The inequality [9] asserts that if two material particles experience both $x_0$ and $x$, then the one that is free (and so can be regarded as at rest in some admissible coordinate system) has longer to wait for the occurrence of the second event (moving clocks run slow). For many years this basically obvious fact was christened "The Twin Paradox."

Just as a smooth curve in Euclidean space has an arc length parametrization, so a timelike world line has a proper time parametrization defined as follows. For each $\xi$ in $[\xi_0, \xi_1]$ let

$$\tau = \tau(\xi) = \int_{\xi_0}^{\xi} \sqrt{|\boldsymbol{g}(\alpha'(\zeta), \alpha'(\zeta))|}\, d\zeta$$

(the proper time length of $\alpha$ from $\alpha(\xi_0)$ to $\alpha(\xi)$). Then $\tau = \tau(\xi)$ has a smooth inverse $\xi = \xi(\tau)$ so $\alpha$ can be reparametrized by $\tau$. We will abuse our notation slightly and write

$$\alpha(\tau) = x^a(\tau) e_a$$

The velocity vector with this parametrization is denoted

$$U = U(\tau) = \frac{dx^a}{d\tau} e_a$$

called the 4-velocity of the world line and is the unit tangent vector field to $\alpha$, that is,

$$U(\tau) \cdot U(\tau) = -1 \qquad [10]$$

for each $\tau$. An admissible observer is, of course, more likely to parametrize a world line by his own time coordinate $x^4$. Then

$$\alpha'(x^4) = \frac{dx^1}{dx^4} e_1 + \frac{dx^2}{dx^4} e_2 + \frac{dx^3}{dx^4} e_3 + e_4$$

so

$$\left|\boldsymbol{g}\left(\alpha'(x^4),\, \alpha'(x^4)\right)\right| = 1 - \|\boldsymbol{V}\|^2$$

where

$$\|V\| = \sqrt{\left(\frac{dx^1}{dx^4}\right)^2 + \left(\frac{dx^2}{dx^4}\right)^2 + \left(\frac{dx^3}{dx^4}\right)^2}$$

is the usual magnitude of the particle's velocity vector

$$\begin{aligned} V &= V(x^4) \\ &= \frac{dx^1}{dx^4} e_1 + \frac{dx^2}{dx^4} e_2 + \frac{dx^3}{dx^4} e_3 \\ &= V^i e_i \end{aligned}$$

in the given admissible coordinate system. One finds then that

$$U = \left(1 - \|V\|^2\right)^{-1/2} (V + e_4) \qquad [11]$$

We shall identify a material particle in $\mathcal{M}$ with a pair $(\alpha, m)$, where $\alpha$ is a timelike world line and $m$ is a positive constant called the particle's proper mass (or rest mass). If each $dx^a/d\xi$, $a = 1, 2, 3, 4$, is constant, then $(\alpha, m)$ is a free material particle with proper mass $m$. The 4-momentum of $(\alpha, m)$ is defined by $P = mU$. Thus,

$$P \cdot P = -m^2 \qquad [12]$$

In any admissible basis we write

$$\begin{aligned} P &= P^a e_a = mU^a e_a = m\frac{dx^a}{d\tau} e_a \\ &= m\left(1 - \|V\|^2\right)^{-1/2} (V + e_4) \qquad [13] \end{aligned}$$

The "spatial part" of $P$ in these coordinates is

$$\boldsymbol{P} = \frac{m}{\sqrt{1 - \|V\|^2}} V$$

which, for $\|V\| \ll 1$, is approximately $mV$. Identifying $m$ with the inertial mass of Newtonian mechanics (measured by an observer for whom the particle's speed is small), this is simply the classical momentum of the particle. Somewhat more explicitly, if one expands $1/\sqrt{1 - \|V\|^2}$ by the Binomial Theorem one finds that

$$\begin{aligned} P^i &= \frac{m}{\sqrt{1 - \|V\|^2}} V^i \\ &= mV^i + \frac{1}{2} mV^i \|V\|^2 + \cdots, \quad i = 1, 2, 3 \qquad [14] \end{aligned}$$

which gives the components of the classical momentum plus "relativistic corrections." In order to preserve a formal similarity with Newtonian mechanics one often sees $m/\sqrt{1 - \|V\|^2}$ referred to as the "relativistic mass" of the particle, but we shall avoid this terminology. The fourth component of $P$ is given by

$$\begin{aligned} P^4 &= -P \cdot e_4 \\ &= \frac{m}{\sqrt{1 - \|V\|^2}} = m + \frac{1}{2} m \|V\|^2 + \cdots \qquad [15] \end{aligned}$$

The appearance of the term $(1/2)m\|V\|^2$ corresponding to the Newtonian kinetic energy suggests that $P^4$ be denoted $E$ and called the total relativistic energy measured by the given admissible observer for the particle:

$$E = -P \cdot e_4 \qquad [16]$$

Now, one must understand that the concept of "energy" in physics is a subtle one and simply giving $-P \cdot e_4$ this name does not ensure that there is any physical content. Whether or not the name is appropriate can only be determined experimentally. In particular, one should ask if the appearance of the term $m$ in [15] is consistent with the view that $P^4$ represents the "energy" of the particle. Observe that if $\|V\| = 0$ (i.e., if the particle is at rest relative to the given observer), then [15] gives

$$E = m \; (= mc^2, \text{ in standard units}) \qquad [17]$$

which we interpret as saying that, even when the particle is at rest, it still has energy. If this is really "energy" in the physical sense, then it should be possible to liberate and use it. That this is, indeed, possible has, of course, been rather convincingly demonstrated.

Next we observe that not only material particles, but also photons possess "momentum" and "energy" and therefore should have 4-momentum (witness, e.g., the photoelectric effect in which photons collide with and eject electrons from their orbits in an atom). Unlike a material particle, however, a photon's characteristic feature is not proper mass, but frequency $\nu$, or wavelength $\lambda = 1/\nu$, related to its energy $\mathcal{E}$ by $\mathcal{E} = h\nu$ ($h$ being Planck's constant) and these are highly observer dependent (Doppler effect). There is, moreover, no "proper frequency" analogous to "proper mass" since there is no admissible observer for whom the photon is at rest. In an attempt to model these features we consider a point $x_0 \in \mathcal{M}$, a future directed null vector $N$ and an interval $I \subseteq \mathbb{R}$. The curve $\alpha : I \to \mathcal{M}$ defined by

$$\alpha(\xi) = x_0 + \xi N \qquad [18]$$

is a parametrization of the world line of a photon through $x_0$. Being null, $N$ can be written in any admissible basis as

$$N = (-N \cdot e_4)(\boldsymbol{d} + e_4) \qquad [19]$$

where

$$\boldsymbol{d} = \left[(N \cdot e_1)^2 + (N \cdot e_2)^2 + (N \cdot e_3)^2\right]^{-1/2} \left[(N \cdot e_1)e_1 + (N \cdot e_2)e_2 + (N \cdot e_3)e_3\right] \qquad [20]$$

is the direction vector of the world line in the corresponding spatial coordinate system. Now, by analogy with [16], we define a photon in $\mathcal{M}$ to be a curve in $\mathcal{M}$ of the form [18], take $N$ to be its 4-momentum and define the energy $\mathcal{E}$ of the photon in the admissible basis $\{e_1, e_2, e_3, e_4\}$ by

$$\mathcal{E} = -N \cdot e_4 \qquad [21]$$

Then, by [19],

$$N = \mathcal{E}(\boldsymbol{d} + e_4) \qquad [22]$$

The corresponding frequency $\nu$ and wavelength $\lambda$ are then defined by $\nu = \mathcal{E}/h$ and $\lambda = 1/\nu$. In another admissible basis, one has $N = \hat{\mathcal{E}}(\hat{\boldsymbol{d}} + \hat{e}_4)$, where $\hat{\boldsymbol{d}}$ and $\hat{\mathcal{E}}$ are defined by the hatted versions of [20] and [21]. One can then show (Naber 1992, section 1.8) that

$$\frac{\hat{\mathcal{E}}}{\mathcal{E}} = \frac{\hat{\nu}}{\nu} = \frac{1 - \beta \cos\theta}{\sqrt{1 - \beta^2}} = (1 - \beta\cos\theta) + \frac{1}{2}\beta^2(1 - \beta\cos\theta) + \cdots \qquad [23]$$

where $\beta$ is the relative speed of the two spatial coordinate systems and $\theta$ is the angle (in the unhatted spatial coordinate system) between the direction $\boldsymbol{d}$ of the photon and the direction of motion of the hatted spatial coordinate system. Equation [23] is the formula for the relativistic Doppler effect with the first term in the series being the classical formula.

We conclude this section by examining a few simple interactions between particles of the sort modeled by our definitions, assuming only that 4-momentum is conserved in the interaction. For convenience, we will use the term free particle to refer to either a free material particle or a photon. If $\mathcal{A}$ is a finite set of free particles, then each element of $\mathcal{A}$ has a unique 4-momentum which is a future-directed timelike or null vector. The sum of any such collection of vectors is timelike and future directed, except when all of the vectors are null and parallel, in which case the sum is null and future directed (Naber 1992, lemma 1.4.3). We call this sum the total 4-momentum of $\mathcal{A}$. Now we formulate a definition which is intended to model a finite set of free particles colliding at some event with a (perhaps new) set of free particles emerging from the collision (e.g., an electron and proton collide, with a neutron and neutrino emerging from the collision). A contact interaction in $\mathcal{M}$ is a triple $(\mathcal{A}, x, \tilde{\mathcal{A}})$, where $\mathcal{A}$ and $\tilde{\mathcal{A}}$ are two finite sets of free particles, neither of which contains a pair of particles with linearly dependent 4-momenta (which would presumably be physically indistinguishable) and $x \in \mathcal{M}$ is an event such that

1. $x$ is the terminal point of all of the particles in $\mathcal{A}$ (i.e., for each world line $\alpha : [\xi_0, \xi_1] \to \mathcal{M}$ of a particle in $\mathcal{A}$, $\alpha(\xi_1) = x$);
2. $x$ is the initial point of all the particles in $\tilde{\mathcal{A}}$, and
3. the total 4-momentum of $\mathcal{A}$ equals the total 4-momentum of $\tilde{\mathcal{A}}$.

Properly (3) is called the conservation of 4-momentum. If $\mathcal{A}$ consists of a single free particle, then $(\mathcal{A}, x, \tilde{\mathcal{A}})$ is called a decay (e.g., a neutron decays into a proton, an electron and an antineutrino).

Consider, for example, an interaction $(\mathcal{A}, x, \tilde{\mathcal{A}})$ for which $\tilde{\mathcal{A}}$ consists of a single photon. The total 4-momentum of $\tilde{\mathcal{A}}$ is null so the same must be true of $\mathcal{A}$. Since the 4-momenta of the individual particles in $\mathcal{A}$ are timelike or null and future directed their sum can be null only if they are, in fact, all null and parallel. Since $\mathcal{A}$ cannot contain distinct photons with parallel 4-momenta, it must consist of a single photon which, by (3), must have the same 4-momentum as the photon in $\tilde{\mathcal{A}}$. In essence, "nothing happened at $x$." We conclude that *no nontrivial interaction of the type modeled by our definition can result in a single photon and nothing else*. Reversing the roles of $\mathcal{A}$ and $\tilde{\mathcal{A}}$ shows that, if 4-momentum is to be conserved, *a photon cannot decay*.

Next let us consider the decay of a single material particle into two material particles, for example, the spontaneous disintegration of an atom through $\alpha$-emission. Thus, we consider a contact interaction $(\mathcal{A}, x, \tilde{\mathcal{A}})$ in which $\mathcal{A}$ consists of a single free material particle of proper mass $m_0$ and $\tilde{\mathcal{A}}$ consists of two free material particles with proper masses $m_1$ and $m_2$. Let $P_0, P_1$, and $P_2$ be the 4-momenta of the particles of proper mass $m_0$, $m_1$, and $m_2$, respectively. Then $P_0 = P_1 + P_2$. Appealing to the "reversed triangle inequality," the fact that $P_1$ and $P_2$ are linearly independent and future directed, and [12] we conclude that

$$m_0 > m_1 + m_2 \qquad [23]$$

The excess mass $m_0 - (m_1 + m_2)$ of the initial particle is regarded, via [17], as a measure of the amount of energy required to split $m_0$ into two pieces. Stated somewhat differently, when the two particles in $\tilde{\mathcal{A}}$ were held together to form the single particle in $\mathcal{A}$, the "binding energy" contributed to the mass of this latter particle.

Reversing the roles of $\mathcal{A}$ and $\tilde{\mathcal{A}}$ in the last example gives a contact interaction modelling an inelastic collision (two free material particles with masses $m_1$ and $m_2$ collide and coalesce to form a third of mass $m_0$). The inequality [23] remains true, of course, and a somewhat more detailed analysis (Naber 1992, section 1.8) yields an approximate formula for $m_0 - (m_1 + m_2)$ which can be compared (favorably) with the Newtonian formula for the loss in kinetic energy that results from the collision (energy which, classically, is viewed as taking the form of heat in the combined particle). An analysis of the interaction in which both $\mathcal{A}$ and $\tilde{\mathcal{A}}$ consist of an electron and a photon yields (Naber 1992, section 1.8) a formula for the so-called Compton effect. Many more such examples of this sort are treated in great detail in Synge (1972, chapter VI, § 14).

## Charged Particles and Electromagnetic Fields

A charged particle in $\mathcal{M}$ is a triple $(\alpha, m, q)$, where $(\alpha, m)$ is a material particle and $q$ is a nonzero real number called the charge of the particle. Charged particles do two things of interest to us. By their very presence they create electromagnetic fields and they also respond to the electromagnetic fields created by other charges.

Charged particles "respond" to an electromagnetic field by experiencing changes in 4-momentum. The quantitative nature of this response, that is, the equation of motion, is generally taken to be the so-called Lorentz 4-force law which expresses the proper time rate of change of the particle's 4-momentum at each point of the world line as a linear function of the 4-velocity. Thus, at each point $\alpha(\tau)$ of the world line

$$\frac{\mathrm{d}P(\tau)}{\mathrm{d}\tau} = q\tilde{F}_{\alpha(\tau)}(U(\tau)) \qquad [24]$$

where $\tilde{F}_{\alpha(\tau)}:\mathcal{M} \to \mathcal{M}$ is a linear transformation determined, in each admissible coordinate system, by the classical electric $\boldsymbol{E}$ and magnetic $\boldsymbol{B}$ fields (here we are assuming that the contribution of $q$ to the ambient electromagnetic field is negligible, that is, $(\alpha, m, q)$ is a "test charge"). Let us write [24] more simply as

$$\tilde{F}(U) = \frac{m}{q}\frac{\mathrm{d}U}{\mathrm{d}\tau} \qquad [25]$$

Dotting both sides of [25] with $U$ gives

$$\tilde{F}(U) \cdot U = \frac{m}{q}\frac{\mathrm{d}U}{\mathrm{d}\tau} \cdot U = \frac{m}{2q}\frac{\mathrm{d}}{\mathrm{d}\tau}(U \cdot U) = \frac{m}{2q}\frac{\mathrm{d}}{\mathrm{d}\tau}(-1) = 0$$

Since any future-directed timelike unit vector $u$ is the 4-velocity of some charged particle, we find that $\tilde{F}(u) \cdot u = 0$ for any such vector. Linearity then implies $\tilde{F}(v) \cdot v = 0$ for any timelike vector. Now, if $u$ and $v$ are timelike and future directed, then $u + v$ is timelike so $0 = \tilde{F}(u + v) \cdot (u + v) = \tilde{F}(u) \cdot v + u \cdot \tilde{F}(v)$ and therefore $\tilde{F}(u) \cdot v = -u \cdot \tilde{F}(v)$. But $\mathcal{M}$ has a basis of future-directed timelike vectors so

$$\tilde{F}(x) \cdot y = -x \cdot \tilde{F}(y) \qquad [26]$$

for all $x, y \in \mathcal{M}$. Thus, at each point, the linear transformation $\tilde{F}$ must be skew-symmetric with respect to the Lorentz inner product. One could therefore model an electromagnetic field on $\mathcal{M}$ by an assignment to each point of a skew-symmetric linear transformation whose job it is to assign to the 4-velocity of a charged particle whose world line passes through that point the change in 4-momentum that the particle should expect to experience because of the presence of the field. However, a slightly different perspective has proved more convenient. Notice that a skew-symmetric linear transformation $\tilde{F}:\mathcal{M} \to \mathcal{M}$ and the Lorentz inner product together determine a bilinear form $F:\mathcal{M} \times \mathcal{M} \to \mathbb{R}$ given by

$$F(x, y) = \tilde{F}(x) \cdot y$$

which is also skew-symmetric $(F(y, x) = \tilde{F}(y) \cdot x = -F(x, y))$ and that, conversely, a skew-symmetric bilinear form uniquely determines a skew-symmetric linear transformation. Now, an assignment of a skew-symmetric bilinear form to each point of $\mathcal{M}$ is nothing other than a 2-form on $\mathcal{M}$ and it is in the language of forms that we choose to phrase classical electromagnetic theory (a concise introduction to this language is available, for example, in Spivak (1965, chapter 4).

Nature imposes a certain restriction on which 2-forms can reasonably represent an electromagnetic field on $\mathcal{M}$ ("Maxwell's equations"). To formulate these we introduce a source 1-form $J$ as follows: If

$x^1, x^2, x^3, x^4$ is any admissible coordinate system on $\mathcal{M}$, then

$$J = J_1 dx^1 + J_2 dx^2 + J_3 dx^3 - \rho dx^4 \qquad [27]$$

where $\rho : \mathcal{M} \to \mathbb{R}$ is a charge density function and $\boldsymbol{J} = J_1 e_1 + J_2 e_2 + J_3 e_3$ is a current density vector field (these are to be regarded as the usual "smoothed out," pointwise versions of "charge per unit volume" and "charge flow per unit area per unit time" as measured by the corresponding admissible observer). Now, our formal definition is as follows: The electromagnetic field on $\mathcal{M}$ determined by the source 1-form $J$ on $\mathcal{M}$ is a 2-form $F$ on $\mathcal{M}$ that satisfies Maxwell's equation

$$dF = 0 \qquad [28]$$

and

$$^*d^*F = J \qquad [29]$$

A few comments are in order here. We have chosen units in which not only the speed of light, but also various other constants that one often finds in Maxwell's equations (the dielectric constant $\epsilon_0$ and magnetic permeability $\mu_0$) are 1 and a factor of $4\pi$ in [29] is "normalized out." The $^*$ in [29] is the Hodge star operator determined by the Lorentz inner product and the chosen orientation of $\mathcal{M}$. This is a natural isomorphism

$$^* : \Omega^p(\mathcal{M}) \to \Omega^{4-p}(\mathcal{M}), \quad p = 0, 1, 2, 3, 4$$

of the $p$-forms on $\mathcal{M}$ to the $(4-p)$-forms on $\mathcal{M}$ and is most simply defined as follows: let $x^1, x^2, x^3, x^4$ be any admissible coordinate system on $\mathcal{M}$. If $1 \in \Omega^0(\mathcal{M})$ is the constant function (0-form) on $\mathcal{M}$ whose value is $1 \in \mathbb{R}$, then

$$^*1 = dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4$$

is the volume form on $\mathcal{M}$. If $1 \le i_1 < \cdots < i_k \le 4$, then $^*(dx^{i_1} \wedge \cdots \wedge dx^{i_k})$ is uniquely determined by

$$\left(dx^{i_1} \wedge \cdots \wedge dx^{i_k}\right) \wedge^* \left(dx^{i_1} \wedge \cdots \wedge dx^{i_k}\right) = -dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4$$

Thus, for example, $^*dx^2 = dx^1 \wedge dx^3 \wedge dx^4$, $^*(dx^1 \wedge dx^2) = -dx^3 \wedge dx^4$, $^*(dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4) = -1$, etc. It follows that, if $\mu$ is a $p$-form on $\mathcal{M}$, then

$$^{**}\mu = (-1)^{p+1}\mu \qquad [30]$$

(a more thorough discussion is available in Choquet-Bruhat *et al.* (1977, chapter V A3)). In particular, [29] is equivalent to

$$d^*F = {}^*J \qquad [31]$$

On regions in which there are no charges, so that $J = 0$, [28] and [31] become the source free Maxwell equations

$$dF = 0 \qquad [32]$$

and

$$d^*F = 0 \qquad [33]$$

that is, both $F$ and $^*F$ are closed 2-forms.

Any 2-form $F$ on $\mathcal{M}$ can be written in any admissible coordinate system as $F = (1/2)F_{ab}dx^a \wedge dx^b$ (summation convention!), where $(F_{ab})$ is the skew-symmetric matrix of components of $F$. In order to make contact with the notation generally employed in physics, we introduce the following names for these components:

$$(F_{ab}) = \begin{pmatrix} 0 & B^3 & -B^2 & E^1 \\ -B^3 & 0 & B^1 & E^2 \\ B^2 & -B^1 & 0 & E^3 \\ -E^1 & -E^2 & -E^3 & 0 \end{pmatrix} \qquad [34]$$

Thus,

$$\begin{aligned} F = {} & E^1 dx^1 \wedge dx^4 + E^2 dx^2 \wedge dx^4 \\ & + E^3 dx^3 \wedge dx^4 + B^3 dx^1 \wedge dx^2 \\ & + B^2 dx^3 \wedge dx^1 + B^1 dx^2 \wedge dx^3 \end{aligned} \qquad [35]$$

Computing $^*F, dF, d^*F$ and $^*d^*F$ and writing $\boldsymbol{E} = E^1 e_1 + E^2 e_2 + E^3 e_3$ and $\boldsymbol{B} = B^1 e_1 + B^2 e_2 + B^3 e_3$ one finds that $dF = 0$ is equivalent to

$$\operatorname{div} \boldsymbol{B} = 0 \qquad [36]$$

and

$$\operatorname{curl} \boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} = 0 \qquad [37]$$

while $^*d^*F = J$ is equivalent to

$$\operatorname{div} \boldsymbol{E} = \rho \qquad [38]$$

and

$$\operatorname{curl} \boldsymbol{B} - \frac{\partial \boldsymbol{E}}{\partial t} = \boldsymbol{J} \qquad [39]$$

Equations [36]–[39] are the more traditional renderings of Maxwell's equations.

In another admissible coordinate system $\hat{x}^1, \hat{x}^2, \hat{x}^3, \hat{x}^4$ on $\mathcal{M}$ (related to the first by [2]) the 2-form $F$ would be written $F = (1/2)\hat{F}_{ab} d\hat{x}^a \wedge d\hat{x}^b$. Setting $\hat{x}^a = \Lambda^a{}_\alpha x^\alpha$ and $\hat{x}^b = \Lambda^b{}_\beta x^\beta$ gives $F = (1/2)(\Lambda^a{}_\alpha \Lambda^b{}_\beta \hat{F}_{ab}) dx^\alpha \wedge dx^\beta$, so

$$F_{\alpha\beta} = \Lambda^a{}_\alpha \Lambda^b{}_\beta \hat{F}_{ab}, \quad \alpha, \beta = 1, 2, 3, 4 \qquad [40]$$

Now, suppose that we wish to describe the electromagnetic field of a uniformly moving charge. According to the relativity principle, it does not matter at all whether we view the charge as moving

relative to a "fixed" admissible observer, or the observer as moving relative to a "stationary" charge. Thus, we shall write out the field due to a charge fixed at the origin of the hatted coordinate system ("Coulomb's law") and transform, by [40], to an unhatted coordinate system moving relative to it. Relative to $\hat{x}^1, \hat{x}^2, \hat{x}^3, \hat{x}^4$, the familiar inverse square law for a fixed point charge $q$ located at the spatial origin gives $\hat{\boldsymbol{B}} = 0$ and $\hat{\boldsymbol{E}} = (q/\hat{r}^3)\hat{\boldsymbol{r}}$, where $\hat{\boldsymbol{r}} = \hat{x}^1\hat{e}_1 + \hat{x}^2\hat{e}_2 + \hat{x}^3\hat{e}_3$ and $\hat{r} = ((\hat{x}^1)^2 + (\hat{x}^2)^2 + (\hat{x}^3)^2)^{1/2}$ (note that $\hat{\boldsymbol{E}}$ is defined only on $\mathcal{M} - \text{Span}\{\hat{e}_4\}$). Thus,

$$(\hat{F}_{ab}) = \frac{q}{\hat{r}^3}\begin{pmatrix} 0 & 0 & 0 & \hat{x}^1 \\ 0 & 0 & 0 & \hat{x}^2 \\ 0 & 0 & 0 & \hat{x}^3 \\ -\hat{x}^1 & -\hat{x}^2 & -\hat{x}^3 & 0 \end{pmatrix} \qquad [41]$$

It is a simple matter to verify that, on its domain, $(\hat{F}_{ab})$ satisfies the source free Maxwell equations. Taking $\Lambda$ to be the special Lorentz transformation corresponding to [5] and writing out [40] with $(\hat{F}_{ab})$ given by [41] yields

$$\begin{aligned} E^1 &= q\left(\frac{\hat{x}^1}{\hat{r}^3}\right) \\ E^2 &= \frac{q}{\sqrt{1-\beta^2}}\left(\frac{\hat{x}^2}{\hat{r}^3}\right) \\ E^3 &= \frac{q}{\sqrt{1-\beta^2}}\left(\frac{\hat{x}^3}{\hat{r}^3}\right) \\ B^1 &= 0 \\ B^2 &= \frac{-q\beta}{\sqrt{1-\beta^2}}\left(\frac{\hat{x}^3}{\hat{r}^3}\right) \\ B^3 &= \frac{q\beta}{\sqrt{1-\beta^2}}\left(\frac{\hat{x}^2}{\hat{r}^3}\right) \end{aligned} \qquad [42]$$

We wish to express these in terms of measurements made by the unhatted observer at the instant the charge passes through his spatial origin. Setting $x^4 = 0$ in [5] gives

$$\hat{x}^1 = \frac{1}{\sqrt{1-\beta^2}}x^1, \quad \hat{x}^2 = x^2, \quad \hat{x}^3 = x^3$$

and so

$$\hat{r}^2 = \frac{1}{1-\beta^2}(x^1)^2 + (x^2)^2 + (x^3)^2$$

which, for convenience, we write $r_\beta^2$. Making these substitutions in [42] gives

$$\begin{aligned} \boldsymbol{E} &= \frac{q}{\sqrt{1-\beta^2}}\left(\frac{1}{r_\beta^3}\right)\left(x^1e_1 + x^2e_2 + x^3e_3\right) \\ &= \frac{q}{\sqrt{1-\beta^2}}\left(\frac{1}{r_\beta^3}\right)\boldsymbol{r} \end{aligned} \qquad [43]$$

and

$$\begin{aligned} \boldsymbol{B} &= \frac{q}{\sqrt{1-\beta^2}}\left(\frac{1}{r_\beta^3}\right)\left(0e_1 - \beta x^3e_2 + \beta x^2 e_3\right) \\ &= \frac{q}{\sqrt{1-\beta^2}}\left(\frac{1}{r_\beta^3}\right)((\beta e_1)\times \boldsymbol{r}) \end{aligned} \qquad [44]$$

for the field of a charge moving uniformly with velocity $\beta e_1$ at the instant the charge passes through the origin. Observe that when $\beta \ll 1, r_\beta \approx r$, so [43] says that the electric field of a slowly moving charge is approximately the Coulomb field. When $\beta \ll 1$, [44] reduces to the Biot–Savart law.

Let us consider one other simple application, that is, the response of a charged particle $(\alpha, m, q)$ to an electromagnetic field which, for some admissible observer, is constant and purely magnetic. For simplicity, we assume that, for this observer $\boldsymbol{E} = \boldsymbol{0}$ and $\boldsymbol{B} = be_3$, where $b$ is a nonzero constant. The corresponding 2-form $F$ has components

$$(F_{ab}) = \begin{pmatrix} 0 & b & 0 & 0 \\ -b & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

(from [34]). The corresponding linear transformation $\tilde{F}$ has the same matrix relative to this basis so, with $\alpha(\tau) = x^a(\tau)e_a$ and $U(\tau) = U^a(\tau)e_a$, the Lorentz 4-force law [25] reduces to the system of linear differential equations

$$\frac{\mathrm{d}U^1}{\mathrm{d}\tau} = \frac{bq}{m}U^2, \qquad \frac{\mathrm{d}U^2}{\mathrm{d}\tau} = -\frac{bq}{m}U^1$$
$$\frac{\mathrm{d}U^3}{\mathrm{d}\tau} = 0, \qquad \frac{\mathrm{d}U^4}{\mathrm{d}\tau} = 0$$

The system is easily solved and the results easily integrated to give

$$\begin{aligned} \alpha(\tau) &= x_0 + a\sin\left(\frac{bq\tau}{m} + \phi\right)e_1 \\ &\quad + a\cos\left(\frac{bq\tau}{m} + \phi\right)e_2 \\ &\quad + c\tau e_3 + \left(1 + \frac{a^2b^2q^2}{m^2} + c^2\right)\tau e_4 \end{aligned} \qquad [45]$$

where $x_0 = x_0^a e_a \in \mathcal{M}$ is constant and $a, \phi$, and $c$ are real constants with $a > 0$ (we have used $U \cdot U = -1$ to eliminate one other arbitrary real constant). Note that, at each point on $\alpha$, $(x^1 - x_0^1)^2 + (x^2 - x_0^2)^2 = a^2$. Thus, if $c \neq 0$ the spatial trajectory in this coordinate system is a helix along the $e_3$-direction (i.e., along the magnetic field lines). If $c = 0$, the trajectory is a circle in the $x^1$–$x^2$ plane. This case is of some practical significance since one can

introduce constant magnetic fields in a bubble chamber so as to induce a particle of interest to follow a circular path. We show now how to measure the charge-to-mass ratio for such a particle. Taking $c = 0$ in [45] and computing $U(\tau)$, then using [11] to solve for the coordinate velocity vector $\boldsymbol{V}$ of the particle gives

$$\boldsymbol{V} = \frac{abq/m}{\sqrt{1 - \|\boldsymbol{V}\|^2}}\left(\cos\left(\frac{bq\tau}{m} + \phi\right)e_1 + \sin\left(\frac{bq\tau}{m} + \phi\right)e_2\right)$$

From this one computes

$$\|\boldsymbol{V}\|^2 = \left(1 + \frac{m^2}{a^2b^2q^2}\right)^{-1}$$

(note that this is a constant). Solving this last equation for $q/m$ (and assuming $q > 0$ for convenience) one arrives at

$$\frac{q}{m} = \frac{1}{a|b|}\frac{\|\boldsymbol{V}\|}{\sqrt{1 - \|\boldsymbol{V}\|^2}}$$

Since $a$, $b$, and $\|\boldsymbol{V}\|$ are measurable, one obtains the desired charge-to-mass ratio.

To conclude we wish to briefly consider the existence and use of "potentials" for electromagnetic fields. Suppose $F$ is an electromagnetic field defined on some connected, open region $X$ in $\mathcal{M}$. Then $F$ is a 2-form on $X$ which, by [28], is closed. Suppose also that the second de Rham cohomology $H^2(X;\mathbb{R})$ of $X$ is trivial (since $\mathcal{M}$ is topologically $\mathbb{R}^4$ this will be the case, for example, when $X$ is all of $\mathcal{M}$, or an open ball in $\mathcal{M}$, or, more generally, an open "star-shaped" region in $\mathcal{M}$). Then, by definition, every closed 2-form on $X$ is exact so, in particular, there exists a 1-form $A$ on $X$ satisfying

$$F = \mathrm{d}A \qquad [46]$$

In particular, such a 1-form $A$ always exists locally on a neighborhood of any point in $X$ for any $F$. Such an $A$ is not uniquely determined, however, because, if $A$ satisfies [46], then so does $A + \mathrm{d}f$ for any smooth real-valued function (0-form) $f$ on $X$ ($\mathrm{d}^2 = 0$ implies $\mathrm{d}(A + \mathrm{d}f) = \mathrm{d}A + \mathrm{d}^2 f = \mathrm{d}A = F$). Any 1-form $A$ satisfying [46] is called a (gauge) potential for $F$. The replacement $A \to A + \mathrm{d}f$ for some $f$ is called a gauge transformation of the potential and the freedom to make such a replacement without altering [46] is called gauge freedom.

One can show that, given $F$, it is always possible to locally solve $\mathrm{d}A = F$ for $A$ subject to an arbitrary specification of the 0-form ${}^*\mathrm{d}^*A$. More precisely, if $F$ is any 2-form satisfying $\mathrm{d}F = 0$ and $g$ is an arbitrary 0-form, then locally, on a neighborhood of any point, there exists a 1-form $A$ satisfying

$$\mathrm{d}A = F \quad \text{and} \quad {}^*\mathrm{d}^*A = g \qquad [47]$$

(a more general result is proved in Parrott (1987, appendix 2) and a still more general one in section 2.9 of this same source). The usefulness of the second condition in [47] can be illustrated as follows. Suppose we are given some (physical) configuration of charges and currents (i.e., some source 1-form $J$) and we wish to find the corresponding electromagnetic field $F$. We must solve Maxwell's equations $\mathrm{d}F = 0$ and ${}^*\mathrm{d}^*F = J$ (subject to whatever boundary conditions are appropriate). Locally, at least, we may seek instead a corresponding potential $A$ (so that $F = \mathrm{d}A$). Then the first of Maxwell's equations is automatically satisfied ($\mathrm{d}F = \mathrm{d}(\mathrm{d}A) = 0$) and we need only solve ${}^*\mathrm{d}^*(\mathrm{d}A) = J$. To simplify the notation let us temporarily write $\delta = {}^*\mathrm{d}^*$ and consider the operator $\Delta = \mathrm{d} \circ \delta + \delta \circ \mathrm{d}$ on forms (variously called the Laplace–Beltrami operator, Laplace–de Rham operator, or Hodge Laplacian on Minkowski spacetime). Then

$$\Delta A = \mathrm{d}(\delta A) + \delta(\mathrm{d}A) = \mathrm{d}({}^*\mathrm{d}^*A) + {}^*\mathrm{d}^*(\mathrm{d}A) \qquad [48]$$

According to the result quoted above, we may narrow down our search by imposing the condition ${}^*\mathrm{d}^*A = 0$, that is

$$\delta A = 0 \qquad [49]$$

(this is generally referred to as imposing the Lorentz gauge). With this, [48] becomes $\Delta A = {}^*\mathrm{d}^*(\mathrm{d}A)$ and to satisfy the second Maxwell equation we must solve

$$\Delta A = J \qquad [50]$$

Thus, we see that the problem of (locally) solving Maxwell's equations for a given source $J$ reduces to that of solving [49] and [50] for the potential $A$. To understand how this simplifies the problem, we note that a calculation in admissible coordinates shows that the operator $\Delta$ reduces to the componentwise d'Alembertian $\Box$, defined on real-valued functions by

$$\Box = \frac{\partial^2}{\partial(x^1)^2} + \frac{\partial^2}{\partial(x^2)^2} + \frac{\partial^2}{\partial(x^3)^2} - \frac{\partial^2}{\partial(x^4)^2}$$

Thus, eqn [50] decouples into four scalar equations

$$\Box A_a = J_a, \quad a = 1, 2, 3, 4 \qquad [51]$$

each of which is the well-studied inhomogeneous wave equation.

## Further Reading

Choquet-Bruhat Y, De Witt-Morette C, and Dillard-Bleick M (1977) *Analysis, Manifolds and Physics*. Amsterdam: North-Holland.
Einstein A *et al.* (1958) *The Principle of Relativity*. New York: Dover.
Naber GL (1992) *The Geometry of Minkowski Spacetime*. Berlin: Springer.
Parrott S (1987) *Relativistic Electrodynamics and Differential Geometry*. Berlin: Springer.
Spivak M (1965) *Calculus on Manifolds*. New York: W A Benjamin.
Synge JL (1972) *Relativity: The Special Theory*. Amsterdam: North-Holland.

# Introductory Article: Quantum Mechanics

**G F dell'Antonio**, Università di Roma "La Sapienza," Rome, Italy



## Historical Background

In this section we shall briefly recall the basic empirical facts and the first theoretical attempts from which the theory and the formalism of present-day quantum mechanics (QM) has grown. In the next sections we shall give the mathematical and computational structure of QM, mention the physical problems that QM has solved with much success, and describe the serious conceptual consistency problems which are posed by QM (and which remain unsolved up to now).

Empirical rules of discretization were observed already, starting from the 1850s, in the absorption and in the emission of light. Fraunhofer noticed that the dark lines in the absorption spectrum of the light of the sun coincide with the bright lines in the emission lines of all elements. G Kirchhoff and R Bunsen reached the conclusion that the relative intensities of the emission and absorption of light implied that the ratio between energy emitted and absorbed is independent of the atom considered. This was the starting point of the analysis by Planck.

On the other hand, by the end of the eighteenth century, the spatial structure of the atom had been investigated; the most successful model was that of Rutherford, in which the atom appeared as a small nucleus of charge $Z$ surrounded by $Z$ electrons attracted by the nucleus according to Coulomb's law. This model represents, for distances of the order of the size of an atom, a complete departure from Newton's laws combined with the laws of classical electrodynamics; indeed, according to these laws, the atom would be unstable against collapse, and would certainly not exhibit a discrete energy spectrum. We must conclude that the classical laws are inadequate for the description of emission and absorption of light, in which the internal structure of the atom plays a major role.

The birth of the old quantum theory is placed traditionally at the date of M Planck's discussion of the blackbody radiation in 1900.

Planck put forward the postulate that light is emitted and absorbed by matter in discrete energy quanta through "resonators" that have an energy proportional to their frequency. This assumption led, through the use of Gibb's rules of Statistical Mechanics applied to a gas of resonators, to a law (Planck's law) which reproduces the empirical findings on the radiation from a blackbody. It led Einstein to ascribe to light (which had, since the times of Maxwell, a successful description in terms of waves) a discrete, particle-like nature. Nine years later A Einstein gave further support to Planck's postulate by showing that it can reproduce correctly the energy fluctuations in blackbody radiation and even clarifies the properties of specific heat. Soon afterwards, Einstein (1924, 1925) proved that the putative particle of light satisfied the relativistic laws (relation between energy and momentum) of a particle with zero mass.

This dual nature of light received further support from the experiments on the Compton effect and from description, by Einstein, of the photoelectric effect (Einstein 1905). It should be emphasized that while Planck considered with light in interaction with matter $\nu$ as composed of bits of energy $h\nu$ ($h \simeq 6,6 \times 10^{-27}$ erg s), Einstein's analysis went much further in assigning to the quantum of light properties of a particle-like (localized) object. This marks a complete departure from the laws of classical electromagnetism. Therefore, quoting Einstein,

> It is conceivable that the wave theory of light, which retains its effectiveness for the representation of purely optical phenomena and is based on continuous functions over space, will lead to contradiction with the experiments when applied to phenomena in which there is creation or conversion of light; indeed these phenomena can be better

> described on the assumption that light is distributed discontinuously in space and described by a finite number of quanta which move without being divided and which must be absorbed or emitted as a whole.

Notice that, for wavelength of $8 \times 10^3$Å, a 30 W lamp emits roughly $10^{20}$ photons $s^{-1}$; for macroscopic objects the discrete nature of light has no appreciable consequence.

Planck's postulate and energy conservation imply that in emitting and absorbing light the atoms of the various elements can lose or gain energy only by discrete amounts. Therefore, atoms as producers or absorbers of radiation are better described by a theory that assigns to each atom a (possible infinite) discrete set of states which have a definite energy.

The old quantum theory of matter addresses precisely this question. Its main proponent is N Bohr (Bohr 1913, 1918). The new theory is entirely phenomenological (as is Planck's theory) and based on Rutherford's model and on three more postulates (Born 1924):

(i) The states of the atom are stable periodic orbits, as given by Newton's laws, of energy $E_n, n \in Z^+$, given by $E_n = h\nu_n f(n)$, where $h$ is Plank's constant, $\nu_n$ is the frequency of the electron on that orbit, and $f(n)$ is for each atom a function approximately linear in $Z$ at least for small values of $Z$.
(ii) When radiation is emitted or absorbed, the atom makes a transition to a different state. The frequency of the radiation emitted or absorbed when making a transition is $\nu_{n,m} = h^{-1}|E_n - E_m|$.
(iii) For large values of $n$ and $m$ and small values of $(n-m)/(n+m)$ the prediction of the theory should agree with those of the classical theory of the interaction of matter with radiation.

Later, A Sommerfeld gave a different version of the first postulate, by requiring that the allowed orbits be those for which the classical action is an integer multiple of Planck's constant.

The old quantum theory met success when applied to simple systems (atoms with $Z < 5$) but it soon appeared evident that a new, radically different point of view was needed and a fresh start; the new theory was to contain few free parameters, and the role of postulate (iii) was now to fix the value of these parameters.

There were two (successful) attempts to construct a consistent theory; both required a more sharply defined mathematical formalism. The first one was sparked by W Heisenberg, and further important ideas and mathematical support came from M Born, P Jordan, W Pauli, P Dirac and, on the mathematical side, also by J von Neumann and A Weyl. This formulation maintains that one should only consider relations between observable quantities, described by elements that depend only on the initial and final states of the system; each state has an internal energy. By energy conservation, the difference between the energies must be proportional (with a universal constant) to the frequency of the radiation absorbed or emitted. This is enough to define the energy of the state of a single atom modulo an additive constant. The theory must also take into account the probability of transitions under the influence of an external electromagnetic field.

We shall give some details later on, which will help to follow the basis of this approach.

The other attempt was originated by L de Broglie following early remarks by HW Bragg and M Brillouin. Instead of emphasizing the discrete nature of light, he stressed the possible wave nature of particles, using as a guide the Hamilton–Jacobi formulation of classical mechanics. This attempt was soon supported by the experiments of Davisson and Germer (1927) of scattering of a beam of ions from a crystal. These experiments showed that, while electrons are recorded as "point particles," their distribution follows the law of the intensity for the diffraction of a (dispersive) wave. Moreover, the relation between momentum and frequency was, within experimental errors, the same as that obtained by Einstein for photons.

The theory started by de Broglie was soon placed in almost definitive form by E Schrödinger. In this approach one is naturally led to formulate and solve partial differential equations and the full development of the theory requires regularity results from the theory of functions.

Schrödinger soon realized that the relations which were found in the approach of Heisenberg could be easily (modulo technical details which we shall discuss later) obtained within the formalism he was advocating and indeed he gave a proof that the two formalisms were equivalent. This proof was later refined, from the mathematical point of view, by J von Neumann and G Mackey.

In fact, Schrödinger's approach has proved much more useful in the solution of most physical problems in the nonrelativistic domain, because it can rely on the developments and practical use of the theory of functions and of partial differential equations. Heisenberg's "algebraic" approach has therefore a lesser role in solving concrete problems in (nonrelativistic) QM.

If one considers processes in which the number of particles may change in time, one is forced to

introduce a Hilbert space that accommodates states with an arbitrarily large number of particles, as is the case of the theory of relativistic quantized field or in quantum statistical mechanics; it is then more difficult to follow the line of Schrödinger, due to difficulties in handling spaces of functions of infinitely many variables. The approach of Heisenberg, based on the algebra of matrices, has a rather natural extension to suitable algebras of operators; the approach of Schrödinger, based on the description of a state as a (wave) function, encounters more difficulties since one must introduce functionals over spaces of functions and the description of dynamics does not have a simple form.

From this point of view, the generalization of Heisenberg's approach has led to much progress in the understanding of the structure of the resulting theory. Still some relevant results have been obtained in a Schrödinger representation. We shall not elaborate further on this point.

We shall end this introductory section with a short description of the emergence of the structure of QM in Heisenberg's and Schrödinger's approaches; this will provide a motivation for the axiom of QM which we shall introduce in the following section. For an extended analysis, see, for example, Jammer (1979).

The specific form that was postulated by de Broglie (1923) for the wave nature of a particle relies on the relation of geometrical optics with wave propagation and on the formulation of Hamiltonian mechanics as a sort of "wave front propagation" through the solution of the Hamilton–Jacobi equation and the introduction of group velocity.

By the analogy with electromagnetic wave, it is natural to associate with a free nonrelativistic particle of momentum $p$ and mass $m$ the plane wave

$$\phi_p(x,t) = e^{i(px-Et)/\hbar}, \quad \hbar = \frac{h}{2\pi}, \ E = \frac{p^2}{2m}$$

Schrödinger obtained the equation for a quantum particle in a field of conservative forces with potential $V(x)$ by considering an analogy with the propagation of an electromagnetic wave in a medium with refraction index $n(x,\omega)$ that varies slowly on the scale of the wavelength. Indeed, in this case the "wave" follows the laws of geometrical optics, and has therefore a "particle-like" behavior. If one denotes by $\hat{u}(x,\omega)$ the Fourier transom (with respect to time) of a generic component of the electric field and one assumes that the field be essentially monochromatic (so that the support of $\hat{u}(x,\omega)$ as a function of $\omega$ is in a very small neighborhood of $\omega_0$), one finds that $\hat{u}(x,\omega)$ is an approximate solution of the equation

$$-\Delta\hat{u}(x,\omega) = \frac{\omega_0^2}{c^2} n^2(x,\omega)\hat{u}(x,\omega) \qquad [1]$$

Writing $u(x,\omega) = A(x,\omega)\, e^{i(\omega/c)W(x,\omega)}$ the phase $W(x,\omega)$ satisfies, in the high-frequency limit, the eikonal equation $|\nabla W(x,\omega)|^2 = n^2(x,\omega)$. One can define for the solution a *phase velocity* $v_f$ and it turns out that $v_f = c/|\nabla W(x,\omega)|$.

On the other hand, classical mechanics can also be described by propagation of surfaces of constant value for the solution $W(x, t)$ of the Hamilton–Jacobi equation $H(x,\nabla W) = E$, with $H = p^2/2m + V(x)$. Recall that high-frequency (the realm of geometric optics) corresponds to small distances. This analogy led Schrödinger (1926) to postulate that the dynamics satisfied by the waves associated with the particles was given by the (Schrödinger) equation

$$i\hbar\frac{\partial\psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m}\Delta_x\psi(x,t) + V(x)\psi(x,t) \qquad [2]$$

This wave was to describe the particle and its motion, but, being complex valued, it could not represent any measurable property. It is a mathematical property of the solutions of [2] that the quantity $\int|\psi(x,t)|^2\, d^3x$ is preserved in time. Furthermore, if one sets

$$\rho(x,t) \equiv |\psi(x,t)|^2$$

$$j(x,t) \equiv -i\frac{\hbar}{2m}[\bar{\psi}(x,t)\nabla\psi(x,t) - \psi(x,t)\nabla\bar{\psi}(x,t)] \qquad [3]$$

one easily verifies the local conservation law

$$\frac{\partial\rho}{\partial t} + \mathrm{div}\, j(x,t) = 0 \qquad [4]$$

These mathematical properties led to the statistical interpretation given by Max Born: in those experiments in which the position of the particles is measured, the integral of $|\psi(x,t)|^2$ over a region $\Omega$ of space gives the probability that at time $t$ the particle is localized in the region $\Omega$. Moreover, the current associated with a charged particle is given locally by $j(x,t)$ defined above.

Let us now briefly review Heisenberg's approach. At the heart of this approach are: empirical formulas for the intensities of emission and absorption of radiation (dispersion relations), Sommerfeld's quantum condition for the action and the vague statement "the analogue of the derivative for the discrete action variable is the corresponding finite difference quotient." And, most important, the remark that the correct description of atomic physics was through quantities associated with pairs of states, that is, (infinite) matrices and the

empirical fact that the frequency (or rather the wave number) $\omega_{k,j}$ of the radiation (emitted or absorbed) in the transition between the atomic levels $k$ and $j$ ($k \neq j$) satisfies the Ritz combination principle $\omega_{m,j} + \omega_{j,k} = \omega_{m,k}$. It easy to see that any doubly indexed family satisfying this relation must have the form $\omega_{m,k} = E_m - E_k$ for suitable constant $E_j$.

It was empirically verified by Kramers that the dipole moment of an atom in an external monochromatic external field with frequency $\nu$ was proportional to the field with a coefficient (of polarization)

$$P - \frac{e^2}{4\pi m} \sum_i \left[ \frac{f_i}{\nu_i^2 - \nu^2} - \frac{F_i}{\nu_i^2 - \nu^2} \right] \qquad [5]$$

where $e$, $m$ are the charge and the mass of the electron and $f_i, F_i$ are the probabilities that the frequency $\nu$ is emitted or absorbed.

A detailed analysis of the phenomenon of polarization in classical mechanics, with the clearly stated aim "of presenting the results in a way that may give hints for the construction of a New Mechanics" was made by Max Born (1924). He makes use of action-angle variables $\{J_i, \theta_i\}$ assuming that the atom can be considered as a collection of harmonic oscillators with frequency $\nu_i$ coupled linearly to the electric field of frequency $\mu$.

In the dipole approximation one obtains the following result for the polarization $P$ (linear response in energy to the electric field):

$$P = - \sum_{(\nu \cdot m) > 0} 2(m \cdot \nabla_J) \frac{|A(J)|^2 (\nu \cdot m)}{((m \cdot \nu)^2 - \mu} \qquad [6]$$

where $\nu_k = \partial H / \partial J_k$, $H$ is the interaction Hamiltonian), and $A(J)$ is a suitable matrix. In order to derive the new dynamics, having as a guide the correspondence principle, one has to compare this result with the Kramers dispersion relation, which we write (to make the comparison easier) in the form

$$P = \frac{e^2}{4\pi m} \sum_{n,m} \frac{f_{m,n}}{\nu_{n,m}^2 - \mu^2} - \frac{f_{n,m}}{\nu_{n,m}^2 - \mu^2} \qquad E_m > E_n \qquad [7]$$

Bohr's rule implies that $\nu(n + \tau, n) = (E(n + \tau - E(n))/\hbar$.

Born and Heisenberg noticed that, for $n$ sufficiently large and $k$ small, one can approximate the differential operator in [6] with the corresponding difference operator, with an error of the order of $k/n$. Therefore, [6] could be substituted by

$$P = - \hbar^{-1} \sum_{m_k > 0} \left[ \frac{|A_{n+m,n}|^2}{\nu(n+m)^2 - \mu^2} - \frac{|A_{n-m,n}|^2}{\nu(n-m)^2 - \mu^2} \right] \qquad [8]$$

The conclusion Born and Heisenberg drew is that the matrix $A$ that takes the place of the momentum in the classical theory must be such that $|A_{n+m,n}|^2 = e^2 \hbar m^{-1} f(n+m, n)$. In the same vein, considering the polarization in a static electric field, it is possible to find an expression for the matrix that takes the place of the coordinate $x$ in classical Hamiltonian theory.

In general, the new approach (matrix mechanics) associates matrices with some relevant classical observables (such as functions of position or momentum) with a time dependence that is derived from the empirical dispersion relations of Kramers, the correspondence principle, Bohr's rule, Sommerfeld action principle and first- (and second-) order perturbation theory for the interaction of an atom with an external electromagnetic field. It was soon clear to Born and Jordan (1925) that this dynamics took the form $i\hbar \dot{A} = AH - HA$ for a matrix $H$ that for the case of the hydrogen atom is obtained for the classical Hamiltonian with the prescription given for the coordinates $x$ and $p$. It was also seen as plausible the relation $[\hat{x}_h, \hat{p}_k] = \mathrm{i}I$ among the matrices $\hat{x}_k$ and $\hat{p}_k$ corresponding to position and momentum. One year later P Dirac (1926) pointed out the structural identity of this relation with the Poisson bracket of Hamiltonian dynamics, developed a "quantum algebra" and a "quantum differentiation" and proved that any $^*$-derivation $\delta$ (derivation which preserves the adjoint) of the algebra $\mathcal{B}_N$ of $N \times N$ matrices is inner, that is, is given by $\delta(a) = \mathrm{i}[a, h]$ for a Hermitian matrix $h$. Much later this theorem was extended (with some assumptions) to the algebra of all bounded operators on a separable Hilbert space. Since the derivations are generators of a one-parameter continuous group of automorphisms, that is, of a dynamics, this result led further strength to the ideas of Born and Heisenberg.

The algebraic structure introduced by Born, Jordan, and Heisenberg (1926) was used by Pauli (1927) to give a purely group-theoretical derivation of the spectrum of the hydrogen atom, following the lines of the derivation in symplectic mechanics of the SO(4) symmetry of the Coulomb system. This remarkable success gave much strength to the Heisenberg formulation of QM, which was soon recognized as an efficient instrument in the study of the atomic world.

The algebraic formulation was also instrumental in the description given by Pauli (1928) of the "spin" (a property of electrons empirically postulated by Goudsmidt and Uhlenbeck to account for a hyperfine splitting of some emission lines) as "internal" degree of freedom without reference to spatial coordinates and still connected with the

properties of the the system under the group of spatial rotations. This description through matrices has a major role also in the formulation by Pauli of the exclusion principle (and its relation with Fermi–Dirac statistics), which gave further credit to the Heisenberg's theory by helping in reproducing correctly the classification of the atoms.

These features may explain why the "standard" formulation of the axioms of QM given in the next section shows the influence of Heisenberg's approach. On the other hand, comparison with experiments is usually set in the framework in Schrödinger's approach. Posing the problems in terms of properties of the solution of the Schrödinger equation, one is led to a pragmatic use of the formalism, leaving aside difficulties of interpretation. This separation of "the axioms" from the "practical use" may be one of the reasons why a serious analysis of the axioms and of the problems that arise from them is apparently not a concern for most of the research in QM, even from the point of view of mathematical physics.

One should stress that both the approach of Born and Heisenberg and that of de Broglie and Schrödinger are rooted in a mixture of attention to the experimental data, deep understanding of the previous theory, bold analogies and approximations, and deep concern for the consistency of the "new mechanics."

There is an essential difference between the starting points of the two approaches. In Heisenberg's approach, the atom has *a priori* no spatial structure; the description is entirely in terms of its properties under emission and absorption of light, and therefore its observable quantities are represented by matrices. Dynamics enters through the study of the interaction with the electromagnetic field, and some analogies with the classical theory of electrodynamics in an asymptotic regime (correspondence principle). In this way, as we have briefly indicated, the special role of some matrices, which have a mutual relation similar to the relation of position and momentum in Hamiltonian theory. Following this analogy, it is possible to extend the theory beyond its original scope and consider phenomena in which the electrons are not bound to an atom.

In the approach of Schrödinger, on the other hand, particles and collections of particles are represented by spatial structures (waves). Spatial coordinates are therefore introduced *a priori*, and the position of a particle is related to the intensity of the corresponding wave (this was stressed by Born). Position and momentum are both basic measurable quantities as in classical mechanics. Physical interpretation forces the particle wave to be square integrable, and mathematics provides a limitation on the simultaneous localization in momentum and position leading to Heisenberg's uncertainty principle. Dynamics is obtained from a particle–wave duality and an analogy with the relativistic wave equation in the low-energy regime. The presence of bound states with quantized energies is seen as a consequence of the well-known fact that waves confined to a bounded spatial region have their wave number (and therefore energy) quantized.

## Formal Structure

In this section we describe the formal mathematical structure that is commonly associated with QM. It constitutes a coherent mathematical theory, but the interpretation axiom it contains leads to conceptual difficulties.

We state the axioms in the form in which they were codified by J von Neumann (1966); they constitute a mathematically precise rendering of the formalism of Born, Heisenberg, and Jordan. The formalism of Schrödinger *per se* does not require general statements about the category of observables.

**Axiom I**

(i) Observables are represented by self-adjoint operators in a complex separable Hilbert space $\mathcal{H}$.
(ii) Every such operator represents an observable.

**Remark** Axiom I (ii) is introduced only for mathematical simplicity. There is no physical justification for part (ii). In principle, an observable must be connected to a procedure of measurement (observation) and for most of the self-adjoint operators on $\mathcal{H}$ (e.g., in the Schrödinger representation for $\mathrm{i}x_k(\partial/\partial x_h)x_k$) such procedure has not yet been given).

**Axiom II**

(i) Pure states of the systems are represented by normalized vectors in $\mathcal{H}$.
(ii) If a measurement of the observable $A$ is made on a system in the state represented by the element $\phi \in \mathcal{H}$, the average of the numerical values one obtains is $<\phi, A\phi>$, a real number because $A$ is self-adjoint (we have denoted by $<\phi, \psi>$ the scalar product in $\mathcal{H}$).

**Remark** Notice that Axiom II makes no statement about the outcome of a single measurement.

Using the natural complex structure of $\mathcal{B}(\mathcal{H})$, pure states can be extended as linear real functionals on $\mathcal{B}(\mathcal{H})$.

One defines a state as any linear real positive functional on $\mathcal{B}(\mathcal{H})$ (all bounded operators on the separable Hilbert space $\mathcal{H}$) and says that a state is normal if it is continuous in the strong topology. It can be proved that a normal state can be decomposed into a convex combination of at most a denumerable set of pure states. With these definitions a state is pure iff it has no nontrivial decomposition. It is worth stressing that this statement is true only if the operators that correspond to observable quantities generate all of $\mathcal{B}(\mathcal{H})$; one refers to this condition by stating that there are no superselection rules.

By general results in the theory of the algebra $\mathcal{B}(\mathcal{H})$, a normal state $\rho$ is represented by a positive operator of trace class $\sigma$ through the formula $\rho(A) = Tr(\sigma A)$. Since a positive trace-class operator (usually referred to as density matrix in analogy with its classical counterpart) has eigenvalues $\lambda_k$ that are positive and sum up to 1, the decomposition of the normal state $\rho$ takes the form $\sigma = \sum_k \lambda_k \Pi_k$, where $\Pi_k$ is the projection operator onto the $k$th eigenstate (counting multiplicity).

It is also convenient to know that if a sequence of normal states $\sigma_k$ on $\mathcal{B}(\mathcal{H})$ converges weakly (i.e., for each $A \in \mathcal{B}(\mathcal{H})$ the sequence $\sigma_k(A)$ converges) then the limit state is normal. This useful result is false in general for closed subalgebras of $\mathcal{B}(\mathcal{H})$, for example, for algebras that contain no minimal projections.

Note that no pure state is dispersion free with respect to all the observables (contrary to what happens in classical mechanics). Recall that the dispersion of the state $\rho_\sigma$ with respect to the observable $A$ is defined as $\Delta_\sigma(A) \equiv \sigma(A^2) - (\sigma(A))^2$.

The connection of the state with the outcome of a single measurement of an observable associated with an operator $A$ is given by the following axiom, which we shall formulate only for the case when the self-adjoint operator $A$ has only discrete spectrum. The generalization to the other case is straightforward but requires the use of the spectral projections of $A$.

**Axiom III**

(i) If $A$ has only discrete spectrum, the possible outcomes of a measurement of $A$ are its eigenvalues $\{a_k\}$.

(ii) If the state of the system immediately before the measurement is represented by the vector $\phi \in \mathcal{H}$, the probability that the outcome be $a_k$ is $\sum_h |<\psi, \phi_h^{A;k}>|$, where $\phi_h^{A;k}$ are a complete orthonormal set in the Hilbert space spanned by the eigenvectors of $A$ to the eigenvalue $a_k$.

(iii) If a system is in the pure state $\phi$ and one performs a measurement of the observable $A$ with outcome $a_j \in (b-\delta, b+\delta)$ for some $b, \delta \in R$ then immediately after the measurement the system can be in any (not necessarily pure) state which lies in the convex hull of the pure states which are in the spectral subspace of the operator $A$ in the interval $\Delta_{b;\delta} \equiv (b-\delta, b+\delta)$.

*Note* Statements (ii) and (iii) can be extended without modification to the case in which the initial state is not a pure state, and is represented by a density matrix $\sigma$.

**Remark 1** Axiom III makes sure that if one performs, immediately after the first, a further measurement of the same observable $A$ the outcome will still lie in the interval $\Delta_{b;\delta}$. This is needed to give some objectivity to the statement made about the outcome; notice that one must place the condition "immediately after" because the evolution may not leave invariant the spectral subspaces of $A$. If the operator $A$ has, in the interval $\Delta_{b;\delta}$, only discrete (pure point) spectrum, one can express Axiom III in the following way: the outcome can be any state that can be represented by a convex affine superposition of the eigenstates of $A$ with eigenvalues contained in $\Delta_{b;\delta}$.

In the very special case when $A$ has only one eigenvalue in $\Delta_{b;\delta}$ and this eigenvalue is not degenerate, one can state Axiom III in the following form (commonly referred to as "reduction of the wave packet"): the system after the measurement is pure and is represented by an eigenstate of the operator $A$.

**Remark 2** Notice that the third axiom makes a statement about the state of the system after the measurement is completed.

It follows from Axiom III that one can measure "simultaneously" only observables which are represented by self-adjoint operators that commute with each other (i.e., their spectral projections mutually commute). It follows from the spectral representation of the self-adjoint operators that a family $\{A_k\}$ of commuting operators can be considered (i.e., there is a representation in which they are) functions over a common measure space.

Axioms I–III give a mathematically consistent formulation of QM and allow a statistical description (and statistical prediction) of the outcome of the measurement of any observable. It is worth remarking that while the predictions will have only a statistical nature, the dynamical evolution of the observables (and by duality of the states) will be described by deterministic laws. The intrinsically statistical aspect of the predictions comes only from

the third postulate, which connects the mathematical content of the theory with the measurement process.

The third axiom, while crucial for the connection of the mathematical formalism with the experimental data, contains the seed of the conceptual difficulties which plague QM and have not been cured so far.

Indeed, the third axiom indicates that the process of measurement is described by laws that are intrinsically different from the laws that rule the evolution without measurement. This privileged role of the changing by effect of a measurement leads to serious conceptual difficulties since the changing is independent of whether or not the result is recorded by some observer; one should therefore have a way to distinguish between measurements and generic interactions with the environment.

A related problem that is originated by Axiom III is that the formulation of this axiom refers implicitly to the presence of a classical observer that certifies the outcomes of measurements and is allowed to make use of classical probability theory. This observer is not subjected therefore to the laws of QM.

These two aspects of the conceptual difficulties have their common origin in the separation of the measuring device and of the measured systems into disjoint entities satisfying different laws. The difficulties in the theory of measurement have not yet received a satisfactory answer, but various attempts have been made, with various degree of success, and some of them are described briefly in the section "Interpretation problems." It appears therefore that QM in its present formulation is a refined and successful instrument for the description of the nonrelativistic phenomena at the Planck scale, but its internal consistency is still standing on shaky ground.

Returning to the axioms, it is worth remarking explicitly that according to Axiom II a state is a linear functional over the observables, but it is represented by a sesquilinear function on the complex Hilbert space $\mathcal{H}$. Since Axiom II states that any normalized element of $\mathcal{H}$ represents a state (and elements that differ only by a phase represent the same state) together with $\phi, \gamma$ also $\xi \equiv a\phi + b\psi$, $|a|^2 + |b|^2 = 1$ represent a state superposition of $\phi$ and $\psi$ (superposition principle).

But for an observable $A$, one has in general $\rho_\xi(A) \neq |a|^2 \rho_\phi(A) + |b|^2 \rho_\psi(A)$, due to the cross-terms in the scalar product. The superposition principle is one of the characteristic features of QM. The superposition of the two pure states $\phi$ and $\psi$ has properties completely different from those of a statistical mixture of the same two states, defined by the density matrix $\sigma = |a|^2 \Pi_\phi + |b|^2 \Pi_\psi$, where we have denoted by $\Pi_\phi$ the orthogonal projection onto the normalized vector $\phi$. Therefore, the search for these interference terms is one of the means to verify the predictions of QM, and their smallness under given conditions is a sign of quasiclassical behavior of the system under study.

Strictly connected to superposition are entanglement and the partial trace operation. Suppose that one has two systems which when considered separately are described by vectors in two Hilbert spaces $\mathcal{H}_i, i = 1, 2$, and which have observables $A_i \in \mathcal{B}(\mathcal{H}_i)$. When we want to study their mutual interaction, it is natural to describe both of them in the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ and to consider the observables $A_1 \otimes I$ and $I \otimes A_2$.

When the systems interact, the interaction will not in general commute with the projection operator $\Pi_1$ onto $\mathcal{H}_1$. Therefore, even if the initial state is of the form $\phi_1 \otimes \phi_2, \phi_i \in \mathcal{H}_i$, the final state (after the interaction) is a vector $\xi \in \mathcal{H}_1 \otimes \mathcal{H}_2$ which cannot be written as $\xi = \zeta_1 \otimes \zeta_2$ with $\zeta_i \in \mathcal{H}_i$. It can be shown, however, that there always exist two orthonormal family vectors $\phi_n \in \mathcal{H}_1$ and $\psi_n \in \mathcal{H}_2$ such that $\xi = \sum c_n \phi_n \otimes \psi_n$ for suitable $c_n \in C$, $\sum |c_n|^2 = 1$ (this decomposition is not unique in general).

Recalling that $\rho_{\phi\otimes\psi}(A_1 \otimes I) = \rho_\phi(A_1)$, one can write

$$\rho_\xi(A_1 \otimes I) = \sum |c_n|^2 \rho_{\zeta_n}(A_1) = \rho_\sigma(A_1)$$
$$\sigma \equiv \sum_n |c_n|^2 \Pi_{\phi_n}$$

The map $\Gamma_2 : \rho_\xi \to \rho_{\sigma_1}$ is called reduction or also conditioning) with respect to $\mathcal{H}_2$; it is also called "partial trace" with respect to $\mathcal{H}_2$. The first notation reflects the analogy with conditioning in classical probability theory.

The map $\Gamma_2$ can be extended by linearity to a map from normal states (density matrices) on $\mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ to normal states on $\mathcal{B}(\mathcal{H}_1)$ and gives rise to a positivity-preserving and trace-preserving map.

One can in fact prove (Takesaki 1971) that any conditioning for normal states of a von Neumann algebra $\mathcal{M}$ is completely positive in the sense that it remains positive after tensorization of $\mathcal{M}$ with $\mathcal{B}(\mathcal{K})$, where $\mathcal{K}$ is an arbitrary Hilbert space.

It can also be proved that a partial converse is true, that is, that every completely positive trace-preserving map $\Phi$ on normal states of a von Neumann algebra $\mathcal{A} \subset \mathcal{B}(\mathcal{H})$ can be written, for a suitable choice of a larger Hilbert space $\mathcal{K}$ and partial isometries $V_k$, in the form (Kraus form) $\Phi(a) = \sum_k V_k^* a V_k$.

But it must be remarked that, if $U(t)$ is a one-parameter group of unitary operators on $\mathcal{H}_1 \otimes \mathcal{H}_2$ and $\sigma$ is a density matrix, the one-parameter family of maps $\Gamma(t) \equiv \sigma \to \Gamma_2(U(t)\sigma U^*(t))$ does not, in general, have the semigroup property $\Gamma(t+s) = \Gamma(t) \cdot \Gamma(s)$ $s, t > 0$ and therefore there is in general no generator (of a reduced dynamics) associated with it. Only in special cases and under very strong hypothesis and approximations is there a reduced dynamics given by a semigroup (Markov property).

Since entanglement and (nontrivial) conditioning are marks of QM, and on the other side the Markov property described above is typical of conditioning in classical mechanics, it is natural to search for conditions and approximations under which the Markov property is recovered, and more generally under which the coherence properties characteristic of QM are suppressed (decoherence). We shall discuss briefly this problem in the section "Interpretation problems," devoted to the attempts to overcome the serious conceptual difficulties that descend from Axiom III.

It is seen from the remarks and definitions above that normal states (density matrices) play the role that in classical mechanics is attributed to measures over phase space, with the exception that pure states in QM do not correspond to Dirac measures (later on we shall discuss the possibility of describing a quantum-mechanical states with a function (Wigner function) on phase space).

In this correspondence, evaluation of an observable (a measurable function over phase space) over a state (a normalized, positive measure) is related to finding the (Hilbert space) trace of the product of an operator in $\mathcal{B}(\mathcal{H})$ with a density matrix. Notice that the trace operation shares some of the properties of the integral, in particular $\operatorname{tr} AB = \operatorname{tr} BA$ if $A$ is in trace class and $B \in \mathcal{B}(\mathcal{H})$ (cf. $g \in L^1$ and $f \in L^\infty$) and $\operatorname{tr} AB > 0$ if $A$ is a density matrix and $B$ is a positive operator. This suggests to define functions over the density matrices that correspond to quantities which are important in the theory of dynamical systems, in particular the entropy.

This is readily done if the Hilbert space is finite dimensional, and in the infinite-dimensional case if one takes as observables all Hermitian bounded operators. In quantum statistical mechanics one is led to consider an infinite collection of subsystems, each one described with a Hilbert space (finite or infinite dimensional) $\mathcal{H}_i, i = 1, 2, \ldots$, the space of representation is a subspace $\mathcal{K}$ of $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots$, and the observables are a (weakly closed) subalgebra $\mathcal{A}$ of $\mathcal{B}(\mathcal{K})$ (typically constructed as an inductive limit of elements of the form $I \otimes I \cdots \otimes A_k \otimes I \cdots$). In this context one also considers normal states on $\mathcal{A}$ and defines a trace operation, with the properties described above for a trace. Most of the definitions (e.g., of entropy) can be given in this enlarged context, but differences may occur, since in general $\mathcal{A}$ does not contain finite-dimensional projections, and therefore the trace function is not the trace commonly defined in a Hilbert space. We shall not describe further this very interesting and much developed theory, of major relevance in quantum statistical mechanics. For a thorough presentation see Ohya and Petz (1993).

The simplest and most-studied example is the case when each Hilbert space $\mathcal{H}_i$ is a complex two-dimensional space. The resulting system is constructed in analogy with the Ising model of classical statistical mechanics, but in contrast to that system it possesses, for each value of the index $i$, infinitely many pure states. The corresponding algebra of observables is a closed subalgebra of $(C^2 \times C^2)^{\otimes Z}$ and generically does not contain any finite-dimensional projection.

This model, restricted to the case $(C^2 \times C^2)^K$, $K$ a finite integer, has become popular in the study of quantum information and quantum computation, in which case a normalized element of $\mathcal{H}_i$ is called a q-bit (in analogy with the bits of information in classical information theory). It is clear that the unit sphere in $(C^2 \times C^2)$ contains many more than four points, and this gives much more freedom for operations on the system. This is the basis of quantum computation and quantum information, a very interesting field which has received much attention in recent years.

## Quantization and Dynamics

The evolution in nonrelativistic QM is described by the Schrödinger equation in the representation in which for an $N$-particle system the Hilbert space is $L^2(R^{3N} \otimes C^k$, where $C^k$ is a finite-dimensional space which accounts for the fact that some of the particles may have a spin content.

Apart from (often) inessential parameters, the Schrödinger equation for spin-0 particles can be written typically as

$$
\begin{aligned}
&\mathrm{i}\hbar \frac{\partial \phi}{\partial t} = H\phi \\
&\quad H \equiv \sum_{k=1}^{N} m_k (\mathrm{i}\hbar \nabla_k + A_k)^2 \\
&\qquad + \sum_{k=1}^{N} V_k(x_k) + \sum_{i \neq k, 1}^{N} V_{i,k}(x_i - x_k)
\end{aligned} \qquad [9]
$$

where $\hbar$ is Planck's constant, $A_k$ are vector-valued functions (vector potentials), and $V_k$ and $V_{i,k}$ are scalar-valued function (scalar potentials) on $R^3$.

If some particles have of spin 1/2, the corresponding kinetic energy term should read $-(\mathrm{i}\hbar\sigma\cdot\nabla)^2$, where $\sigma_k$, $k=1, 2, 3$, are the Pauli matrices and one must add a term $W(x)$ which is a matrix field with values in $C^k\otimes C^k$ and takes into account the coupling between the spin degrees of freedom. Notice that the local operator $\mathrm{i}\sigma\cdot\nabla$ is a "square root" of the Laplacian.

A relativistic extension of the Schrödinger equation for a free particle of mass $m\geq 0$ in dimension 3 was obtained by Dirac in a space of spinor-valued functions $\psi_k(x,t), k=0,1,2,3$, which carries an irreducible representation of the Lorentz group. In analogy with the electromagnetic field, for which a linear partial differential equation (PDE) can be written using a four-dimensional representation of the Lorentz group, the relativistic Dirac equation is the linear PDE

$$\mathrm{i}\sum_{k=0}^{3}\gamma_k\frac{\partial}{\partial x_k}\psi=m\psi,\quad x_0\equiv ct$$

where the $\gamma_k$ generate the algebra of a representation of the Lorentz group. The operator $\sum(\partial/\partial x_k)\gamma_k$ is a local square root of the relativistically invariant d'Alembert operator $-\partial^2/\partial x_0^2+\Delta-m\cdot I$.

When one tries to introduce (relativistically invariant) local interactions, one faces the same problem as in the classical mechanics, namely one must introduce relativistically covariant fields (e.g., the electromagnetic field), that is, systems with an infinite number of degrees of freedom. If this field is considered as external, one faces technical problems, which can be overcome in favorable cases. But if one tries to obtain a fully quantized theory (by also quantizing the field) the obstacles become unsurmountable, due also to the nonuniqueness of the representation of the canonical commutation relations if these are taken as the basis of quantization, as in the finite-dimensional case.

In a favorable case (e.g., the interaction of a quantum particle with the quantized electromagnetic field) one can set up a perturbation scheme in a parameter $\alpha$ (the physical value of $\alpha$ in natural units is roughly 1/137). We shall come back later to perturbation schemes in the context of the Schrödinger operator; in the present case one has been able to find procedures (renormalization) by which the series in $\alpha$ that describe relevant physical quantities are well defined term by term. But even in this favorable case, where the sum of the first few terms of the series is in excellent agreement with the experimental data, one has reasons to believe that the series is not convergent, and one does not even know whether the series is asymptotic.

One is led to wonder whether the structure of fields (operator-valued elements in the dual of compactly supported smooth functions on classical spacetime), taken over in a simple way from the field structure of classical electromagnetism, is a valid instrument in the description of phenomena that take place at a scale incomparably smaller than the scale (atomic scale) at which we have reasons to believe that the formalisms of Schrödinger and Heisenberg provide a suitable model for the description of natural phenomena.

The phenomena which are related to the interaction of a quantum nonrelativistic particle interacting with the quantized electromagnetic field take place at the atomic scale. These phenomena have been the subject of very intense research in theoretical physics, mostly within perturbation theory, and the analysis to the first few orders has led to very spectacular results (although there is at present no proof that the perturbation series are at least asymptotic).

In this field rigorous results are scarce, but recently some progress has been made, establishing, among other things, the existence of the ground state (a nontrivial result, because there is no gap separating the ground-state energy from the continuous part of the spectrum) and paving the way for the description of scattering phenomena; the latter result is again nontrivial because the photon field may lead to an anomalous infrared (long-range) behavior, much in the same way that the long-range Coulomb interaction requires a special treatment in nonrelativistic scattering theory.

This contribution to the Encyclopedia is meant to be an introduction to QM and therefore we shall limit ourselves to the basic structure of nonrelativistic theory, which deals with systems of a finite number of particles interacting among themselves and with external (classical) potential fields, leaving for more specialized contributions a discussion of more advanced items in QM and of the successes and failures of a relativistically invariant theory of interaction between quantum particles and quantized fields.

We shall return therefore to basics.

One may begin a section on dynamics in QM by discussing some properties of the solutions of the Schrödinger equation, in particular dispersive effects and the related scattering theory, the problem of bound states and resonances, the case of time-dependent perturbation and the ionization effect, the binding of atoms and molecules, the Rayleigh scattering, the Hall effect and other effects in nanophysics, the various multiscale and adiabatic limits, and in general all the physical problems that

have been successfully solved by Schrödinger's QM (as well as the very many interesting and unsolved problems).

We will consider briefly these issues and the approximation schemes that have been developed in order to derive explicit estimates for quantities of physical interest. Since there are very many excellent reviews of present-day research in QM (e.g., Araki and Ezawa (2004), Blanchard and Dell'Antonio (2004), Cycon *et al.* (1986), Islop and Sigal (1996), Lieb (1990), Le Bris (2005), Simon (2002), and Schlag (2004)) we refer the reader to the more specialized contributions to this Encyclopedia for a detailed analysis and precise statements about the results.

We prefer to come back first to the foundations of the theory; we shall take the point of view of Heisenberg and start discussing the mapping properties of the algebra of observables and of the states. Since transition probabilities play an important role, we consider only transformations $\alpha$ which are such that, for any pair of pure states $\phi_1$ and $\phi_2$, one has $<\alpha(\phi_1), \alpha(\phi_2)> = <\phi_1, \phi_2>$. We call these maps Wigner automorphisms.

A result of Wigner (see Weyl (1931)) states that if $\alpha$ is a Wigner automorphism then there exists a unique operator $U_\alpha$, either unitary or antiunitary, such that $\alpha(P) = U^*_\alpha P U_\alpha$ for all projection operators. If there is a one-parameter group of such automorphisms, the corresponding operators are all unitary (but they need not form a group).

A generalization of this result is due to Kadison. Denoting by $I_{1,+}$ the set of density matrices, a Kadison automorphism $\beta$ is, by definition, such that for all $\sigma_1, \sigma_2 \in I_{1,+}$ and all $0 < s < 1$ one has $\beta(s\sigma_1 + (1-s)\sigma_2) = s\beta(\sigma_1) + (1-s)\beta(\sigma_2)$. For Kadison automorphisms the same result holds as for Wigner's.

A similar result holds for automorphisms of the observables. Notice that the product of two Hermitian operators is not Hermitian in general, but Hermiticity is preserved under Jordan's product defined as $A \times B \equiv (1/2)[AB + BA]$.

A Segal automorphism is, by definition, an automorphism of the Hermitian operators that preserves the Jordan product structure. A theorem of Segal states that $\gamma$ is a Segal automorphism if and only if there exist an orthogonal projector $E$, a unitary operator $U$ in $E\mathcal{H}$, and an antiunitary operator $V$ in $(I-E)\mathcal{H}$ such that $\gamma(A) = WAW^*$, where $W \equiv U \oplus V$.

We can study now in more detail the description of the dynamics in terms of automorphism of Wigner or Kadison type when it refers to states and of Segal type when it refers to observables. We require that the evolution be continuous in suitable topologies. The strongest result refers to Wigner's case. One can prove that if a one-parameter group of Wigner automorphism $\alpha_t$ is measurable in the weak topology (i.e., $\alpha_t\sigma(A)$ is measurable in $t$ for every choice of $A$ and $\sigma$) then it is possible to choose the $U(t)$ provided by Wigner's theorem in such a way that they form a group which is continuous in the strong topology. Similar results are obtained for the cases of Kadison and Segal automorphism, but in both cases one has to assume continuity of $\alpha_t$ in a stronger topology (the strong operator topology in the Segal case, the norm topology in Kadison's). Weak continuity is sufficient if the operator product is preserved (in this case one speaks of automorphisms of the algebra of bounded operators). The existence of the continuous group $U(t)$ defines a Hamiltonian evolution. One has indeed:

**Theorem 1** (Stone). *The map $t \to U(t), t \in R$ is a weakly continuous representation of $R$ in the set of unitary operators in a Hilbert space $\mathcal{H}$ if and only if there exists a self-adjoint operator $H$ on (a dense set of) $\mathcal{H}$ such that $U(t) = \mathrm{e}^{itH}$ and therefore*

$$\phi \in D(H) \to \mathrm{i}\frac{dU(t)}{dt}\phi = HU(t)\phi \qquad [10]$$

*The operator $H$ is called generator of the dynamics described by $U(t)$.*

*Note* In Schrödinger's approach the operator described in Stone's theorem is called Hamiltonian, in analogy with the classical case. In the case of one particle of mass $m$ in $R^3$ subject to a conservative force with potential energy $V(x)$ it has the following form, in units in which $\hbar = 1$:

$$H = -\frac{1}{2m}\Delta + V(x), \quad \Delta = \sum_k \frac{\partial^2_{x_k}}{\partial x_k^2} \qquad [11]$$

If the potential $V$ depends on time, Stone's theorem is not directly applicable but still the spectral properties of the self-adjoint operators $H_t$ and of the Kernel of the group $\tau \to \mathrm{e}^{\mathrm{i}H_t\tau}$ are essential to solve the (time-dependent) Schrödinger equation.

The semigroup $t \to \mathrm{e}^{-tH_0}$ is usually a positivity-preserving semigroup of contractions and defines a Markov process; in favorable cases, the same is true of $t \to \mathrm{e}^{-tH}$ (Feynmann–Kac formula).

There is an analogous situation in the general theory of dynamical systems on a von Neumann algebra; in analogy with the case of elliptic operators, one defines as "dissipation" a map $\Delta$ on a von Neumann algebra $\mathcal{M}$ which satisfies $\Delta(a^*a) \geq a^*\Delta(a) + \Delta(a^*)a$ for all $a \in \mathcal{M}$. The positive dissipation $\Delta$ is called completely positive if it remains positive after tensorization with $\mathcal{B}(\mathcal{K})$ for any

Hilbert space $\mathcal{K}$. Notice that according to this definition every $^*$-derivation is a completely positive dissipation. For dissipations there is an analog of the theorem of Stinespring, and often bounded dissipation can be written as

$$\Delta(a) = \mathrm{i}[h, a] + \sum V_k^* a V_k - \left(\frac{1}{2}\right) \sum \{V_k^* V_k, a\} \quad \text{for } a \in \mathcal{M}$$

(the symbols $\{.,.\}$ denote the anticommutator).

In general terms, by quantization is meant the construction of a theory by deforming a commutative algebra of functions on a classical phase $X$ in such a way that the dynamics of the quantum system can be derived from the prescription of deformation, usually by deforming the Poisson brackets if $X$ is a cotangent bundle $T^*\mathcal{M}$ (Halbut 2002, Landsman 2002). We shall discuss only the Weyl quantization (Weyl 1931) that has its roots in Heisenberg's formulation of QM and refers to the case in which the configuration space is $R^N$, or, with some variant (Floquet–Zak) the $N$-dimensional torus. We shall add a few remarks on the Wick (anti-Weyl) quantization. More general formulations are needed when one tries to quantize a classical system defined on the cotangent bundle of a generic variety and even more so if it defined on a generic symplectic manifold.

The Weyl quantization is a mathematically accurate rendering of the essential content of the procedure adopted by Born and Heisenberg to construct dynamics by finding operators which play the role of symplectic coordinates.

Consider a system with one degree of freedom. The first naive attempt would be to find operators $\hat{q}, \hat{p}$ that satisfy the relation

$$[\hat{q}, \hat{p}] \subset \mathrm{i}I \qquad [12]$$

and to construct the Hamiltonian in analogy with the classical case. To play a similar role, the operators $\hat{q}$ and $\hat{p}$ must be self-adjoint and satisfy [12] at least in a weak sense. If both are bounded, [12] implies $\mathrm{e}^{-\mathrm{i}b\hat{p}}\hat{q}\mathrm{e}^{-\mathrm{i}b\hat{p}} = \hat{q} + bI$ (the exponential is defined through a convergent series) and therefore the spectrum of $\hat{q}$ is the entire real line, a contradiction. Therefore, that inclusion sign in [12] is strict and we face domain problems, and as a consequence [12] has many inequivalent solutions ("equivalence" here means "unitary equivalence").

Apart from "pathological" ones, defined on $L^2$-spaces over multiple coverings of $R$, there are inequivalent solutions of [12] which are effectively used in QM.

The most common solution is on the Hilbert space $L^2(R)$ (with Lebesgue measure), with $\hat{x}$ defined as the essentially self-adjoint operator that acts on the smooth functions with compact support as multiplication by the coordinate $x$ and $\hat{p}$ is defined similarly in Fourier space. This representation can be trivially generalized to construct operators $\hat{q}_k$ and $\hat{p}_k$ in $L^2(R^N)$.

Another frequently used representation of [12] is on $L^2(S^1)$ (and when generalized to $N$ degrees of freedom, on $T^N$). In this representation, the operator $\hat{p}$ is defined by $c_k \to kc_k$ on functions $f(\theta) = \sum_{k=-M}^{N} c_k \mathrm{e}^{\mathrm{i}k\theta/2\pi}, 0 \geq M, N < \infty$. In this case the operator $\hat{q}$ is defined as multiplication by the angle coordinate $\theta$. It is easy to check that this representation is inequivalent to the previous one and that [12] is satisfied (as an identity) on the (dense) set of vectors which are in the domain both of $\hat{p}\hat{q}$ and of $\hat{q}\hat{p}$. But notice that the domain of essential self-adjointness of $\hat{p}$ is not left invariant by the action of $\hat{q}$ ($\theta f(\theta)$ is a function on $S^1$ only if $f(2\pi) = 0$).

We shall denote $\hat{p}$ in this representation by the symbol $\partial/\partial\theta_{\mathrm{per}}$ and refer to it as the Bloch representation. It can be modified by setting the action of $\hat{p}$ as $c_n \to nc_n + \alpha, 0 < \alpha < 2\pi$, and this gives rise to the various Bloch–Zak and magnetic representations.

The Bloch representation can be extended to periodic functions on $R^1$ noticing that $L^2(R) = L^2(S^1) \otimes l^2(N)$; similarly, the Bloch–Zak and the magnetic representation can be extended to $L^2(R^N)$.

The difference between the representations can be seen more clearly if one considers the one-parameter groups of unitary operators generated by the "canonical operators" $\hat{q}$ and $\hat{p}$. In the Schrödinger representation on $L^2(R)$, these groups satisfy

$$U(a)V(b) = \mathrm{e}^{\mathrm{i}ab}V(b)U(a)$$
$$U(a) = \mathrm{e}^{\mathrm{i}a\hat{q}}, \qquad V(b) = \mathrm{e}^{\mathrm{i}b\hat{p}}$$

and therefore, setting $z = a + \mathrm{i}b$ and $W(z) \equiv \mathrm{e}^{-\mathrm{i}ab/2}V(b)U(a)$ one has

$$W(z)W(z') = \mathrm{e}^{-\mathrm{i}\omega(z,z')/2}W(z+z') \qquad z \in C, \quad \omega(z,z') = \mathrm{Im}(\bar{z}, z') \qquad [13]$$

The unitary operators $W(z)$ are therefore projective representations of the additive group $C$. This generalizes immediately to the case of $N$ degrees of freedom; the representation is now of the additive group $C^N$ and $\omega$ is the standard symplectic form on $C^N$.

In the Bloch representation, the unitaries $U(a)V(b)U^*(a)V^*(b)$ are not multiples of the identity, and have no particularly simple form. The map $C^N \ni z \to W(z)$ with the structure [13] is called Weyl system; it plays a major role in QM. The following

theorem has therefore a major importance in the mathematical theory of QM.

**Theorem 2** (von Neumann 1965). *There exists only one, modulo unitary equivalence, irreducible representation of the Weil system.*

The proof of this theorem follows a general pattern in the theory of group representations. One introduces an algebra $\mathcal{W}^{(N)}$ of operators

$$W_f \equiv \int f(z) W(z) \mathrm{d}z, \quad f \in L^1(C^N)$$

called Weyl algebra.

It easy to see that $|W_f| = |f|_1$ and that $f \to W_f$ is a linear isomorphism of algebras if one considers $\mathcal{W}^{(N)}$ with its natural product structure and $L^1$ as a noncommutative algebra with product structure

$$f * g \equiv \int \mathrm{d}z' f(z - z') g(z') \exp \frac{\mathrm{i}}{2} \omega(z, z') \qquad [14]$$

So far the algebra $\mathcal{W}^{(N)}$ is a concrete algebra of bounded operators on $L^2(R^2)$. But it can also be considered an abstract $C^*$-algebra which we still denote by $\mathcal{W}^{(N)}$.

It is easy to see that, according to [14], if $f_0$ is chosen to be a suitable Gaussian, then $W_{f_0}$ is a projection operator which commutes with all the $W_f$'s. Moreover, $W_f W_g = \phi_{f,g} W_{f*g}$ for a suitable phase factor $\phi$. Considering the Gelfand–Neumark–Segal construction for the $C^*$-algebra $\mathcal{W}^{(N)}$, one finds that these properties lead to a decomposition of any representation in cyclic irreducible equivalent ones, completing the proof of the theorem.

The Weyl system has a representation (equivalent to the Schrödinger one) in the space $L^2(R^N, g)$, where $g$ is Gauss's measure. This allows an extension in which $C^N$ is replaced by an infinite-dimensional Banach space equipped with a Gauss measure (weak distribution (Segal 1965, Gross 1972, Wiener 1938)). Uniqueness fails in this more general setting (uniqueness is strictly connected with the compactness of the unit ball in $C^N$). Notice that in the Schrödinger representation (and, therefore, in any other representation) the Hamiltonian for the harmonic oscillator defines a positive self-adjoint operator

$$N = \sum_1^N N_k, \quad N_k = -\frac{\partial^2}{\partial x_{k^2}} + x_k^2 - 1$$

The spectrum of each of the commuting operators $N_k$ consists of the positive integers (including 0) and is therefore called number operator for the $k$th degree of freedom. The operator $N_k$ can be written as $N_k = a_k^* a_k$, where $a_k = (1/\sqrt{2})(x_k + \partial/\partial x_k)$ and $a_k^*$ is the formal adjoint of $a_k$ in $L^2(R)$. One has $|a_k(N_k + 1)^{-1/2}| < 1$. In the domain of $N$ these operators satisfy the following relations (canonical commutation relations)

$$\begin{aligned} [a_k, a_h^*] &= \delta_{k,h}, \qquad [a_h, a_k] = 0 \\ [N_k, a_h] &= -a_h \delta_{h,k}, \qquad [N_h, a_k^*] = a_k^* \delta_{h,k} \end{aligned} \qquad [15]$$

In view of the last two relations, the operator $a_k$ is called the annihilation operator (relative to the $k$th degree of freedom) and its formal adjoint is called the creation operator. The operators $a_k$ have as spectrum the entire complex plane, the operators $a_k^*$ have empty spectrum; the eigenvectors of $N_k$ are the Hermite polynomials in the variable $x_k$. The eigenvectors of $a_k$ (i.e., the solutions in $L^2(R)$ of the equation $a_k \phi_\lambda = \lambda \phi_\lambda$, $\lambda \in C$) are called coherent states; they have a major role in the Bargmann–Fock–Segal quantization and in general in the semiclassical limit.

The operators $\{N_k\}$ generate a maximal abelian system and therefore the space $L^2(R^N)$ has a natural representation as the symmetrized subspace of $\oplus_k (C^N)^k$ (Fock representation). In this representation, a natural basis is given by the common eigenvectors $\phi_{\{n_k\}}, k = 1, \ldots, N$, of the operators $N_k$. A generic vector can be written as

$$\psi = \sum_{\{n_k\}} c_{\{n_k\}} \phi_{\{n_k\}}, \quad \sum_{\{n_k\}} |c_{\{n_k\}}|^2 < \infty$$

and therefore can be represented by the sequence $c_{\{n_k\}}$.

Notice that the creation operators do not create particles in $R^N$ but rather act as a shift in the basis of the Hermite polynomials.

It is traditional to denote by $\gamma(L^2(R^N))$ the Fock representation (also called second quantization because for each degree of freedom the wave function is written in the quantized basis of the harmonic oscillator) and to denote by $\Gamma(A)$ the lift of a matrix $A \in B(C^N)$. These notations are especially used if $C^N$ is substituted with a Banach space $X$. This terminology was introduced by Segal in his work on quantization of the wave equation; it is used ever since, mostly in a perturbative context.

In the theory of quantized fields, the space $C^N$ is substituted with a Banach space, $X$, of functions. In this setting, "second quantization" (Segal 1965, Nelson 1974) considers the state $\phi_{\{n_k\}}$ as representing a configuration of the system in which there are precisely $n_k$ particles in the $k$th physical state (this presupposes having chosen a basis in the space of distribution on $R^3$). There is no problem in doing this (Gross 1972) and one can choose for $X$ a suitable Sobolev space (which one depends on the Gaussian measure given in $X$) if one wants that the

generalization of the commutation relations [15] be of the form $[a^*(f), a(g)] = <f, g>$ with a suitable scalar product $<\cdot,\cdot>$ in $X$. The problem with quantization of relativistic fields is that, in order to ensure locality, one is forced to use a Sobolev space of negative index (depending on the dimension of physical space), and this gives rise to difficulties in the definition of the dynamics for nonlinear vector fields.

One should notice that in the work of Segal (1965), and then in Constructive field theory (Nelson 1974), the Fock representation is placed in a Schrödinger context exhibiting the relevant operators as acting on a space $L^2(X, g)$, where $X$ is a subspace of the space of Schwartz distributions on the physical space of the particles one wants to describe and $g$ is a suitably defined Gauss measure on $X$.

The Fock representation is related to the Bargmann–Fock–Segal representation (Bargmann 1967), a representation in a space of holomorhic functions on $C^N$ square integrable with respect to a Gaussian measure. For its development, this representation relies on the properties of Toeplitz operators and on Tauberian estimates. It is much used in the study of the semiclassical limit and in the formulation of QM in systems for which the classical version has, for phase space, a manifold which is not a cotangent bundle (e.g., the 2-sphere).

**Remark** The Fock representation associated with the Weyl system in the infinite-dimensional context can describe only particles obeying Bose–Einstein statistics; indeed, the states are qualified by their particle content for each element of the basis chosen and there is no possibility of identifying each particle in an $N$-particle state. This is obvious in the finite-dimensional case: the Hermite polynomial of order 2 cannot be seen as "composed" of two polynomials of order 1.

In the infinite-dimensional context, if one wants to treat particles which obey Fermi–Dirac statistics, one must rely on the Pauli exclusion principle (Pauli 1928), which states that two such particles cannot be in the same configuration; to ensure this, the wave function must be antisymmetric under permutation of the particle symbols. It is a matter of fact (and a theorem in relativistic quantum field theory which follows in that theory from covariance, locality and positivity of the energy (Streater and Wightman 1964) that particles with half-integer spin obey the Fermi–Dirac statistics. Therefore, to quantize such systems, one must introduce (commutation) relations different from those of Weyl. Since it must now be that $(a^*)^2 = 0$, due to antisymmetry, it is reasonable to introduce the following relations (canonical anticommutation relations:

$$\begin{aligned} &\{a_k, a_h^*\} = \delta_{k,h}, \qquad \{a_h, a_k\} = 0 \\ &[N_k, a_h] = -a_h \delta_{h,k}, \quad \{A, B\} \equiv AB - BA \end{aligned} \qquad [16]$$

The Hilbert space is now $\otimes^N \mathcal{H}_2$, where $\mathcal{H}_2$ is a two-dimensional complex Hilbert space. Notice that $\mathcal{H}_2$ carries an irreducible two-dimensional representation of $sU(2) \equiv o(3)$ (spin representation) so that this quantization associates spin 1/2 and antisymmetry.

The operators in [16] are all bounded (in fact bounded by 1 in norm). The Fock representation is constructed as in the case of Weyl (see Araki (1988)), with $n_k$ equal 0 or 1 for each index $k$. The infinite-dimensional case is defined in the same way, and leads to inequivalent irreducible representations (Araki 1988); only in one of them is the number operator defined and bounded below. Some of these representations can be given a Schrödinger-like form, with the introduction of a gauge and an integration formalism based on a trace (Gross 1972). This system is much used in quantum statistical mechanics because it deals with bounded operators and can take advantage of strong results in the theory of $C^*$-algebras. In the finite-dimensional case (and occasionally also in the general case) it is used in quantum information (the space $\mathcal{H}_2$ is the space of a quantum bit).

Returning to the Weyl system, we now introduce the strictly related Wigner function which plays an important role in the analysis of the semiclassical limit and in the discussion of some scaling limits, in particular the hydrodynamical limit and the Bose–Einstein condensation when $N \to \infty$.

The Wigner function $W_\phi$ for a pure state $\phi$ is a real-valued function on the phase space of the classical system which represents the state faithfully. It is defined as

$$W_\psi(x, \xi) = (2\pi)^{-n} \int_{R^n} \mathrm{e}^{-\mathrm{i}(\xi, x)} \psi\left(x + \frac{y}{2}\right) \bar{\psi}\left(x - \frac{y}{2}\right) \mathrm{d}y$$

The Wigner function is not positive in general (the only exceptions are those Gaussian states that satisfy $\Delta(x) \cdot \Delta(p) \geq \hbar$). But is has the interesting property that its marginals reproduce correctly the Born rule. In fact, one has $\int W_\phi(x, \xi)\, \mathrm{d}x = |\hat{\phi}(\xi)|^2$. If the function $\phi(t, x)$ $x \in R^n$ is a solution of the free Schrödinger equation $\mathrm{i}\hbar \partial\phi/\partial t = -\hbar^2 \Delta$ then its Wigner function satisfies the Liouville (transport) equation $\partial W_\phi / \partial t + \xi \cdot \nabla W = 0$.

The Wigner function is strictly linked with the Weyl quantization. This quantization associates with every function $\sigma(p, x)$ in a given regularity

class an operator $\sigma(D,x)$ (the Weyl symbol of the function $\sigma$) defined by

$$(\sigma(D,x)f,g) \equiv \int \sigma(\xi,x)W(f,g)(\xi,x)\,\mathrm{d}\xi\,\mathrm{d}x$$
$$W(f,g)(\xi,x) \equiv \int \mathrm{e}^{-\mathrm{i}(\xi,p)}f\left(x+\frac{p}{2},x-\frac{p}{2}\right)\mathrm{d}p$$

It can be verified that the action of $F$ preserves the Schwartz classes $S$ and $S'$ and is unitary in $L^2(R^{2N})$. Moreover, one has $\sigma(D,x)^* = \bar{\sigma}(D,x)$.

The relation between Weyl's quantization and Wigner functions can be readily seen from the natural duality between bounded operators and pure states:

$$\mathrm{tr}(\hat{A}\,\hat{\rho}) \equiv \int a(p,q)\rho(p,q)\,\mathrm{d}p\,\mathrm{d}q$$
$$\rho(p,q) = \int \mathrm{e}^{\mathrm{i}(p,q')}\rho(q',q)\,\mathrm{d}q'$$

We give now a brief discussion of the general structure of a quantization, and apply it to the Weyl quantization. By quantization of a Hamiltonian system we mean a correspondence, parametrized by a small parameter $\hbar$, between classical observables (real functions on a phase space $\mathcal{F}$) and quantum observables (self-adjoint operators on a Hilbert space $\mathcal{H}$) with the property that the corresponding structures coincide in the limit $\hbar \to 0$ and the difference for $\hbar \neq 0$ can be estimated in a suitable topology.

This last requirement is important for the applications and, from this point of view, Weyl's quantization gives stronger results than the other formalisms of quantization.

We limit our analysis to the case $\mathcal{F} \equiv T^*X$, with $X \equiv R^N$, and we make use of the realization of $\mathcal{H}$ as $L^2(R^N)$.

Let $\{x_i\}$ be Cartesian coordinates in $R^N$ and consider a correspondence $A \to \hat{A}$ that satisfies the following requirements:

1. $A \leftrightarrow \hat{A}$ is linear;
2. $x_k \leftrightarrow \hat{x}_k$ where $\hat{x}_k$ is multiplication by $x_k$;
3. $p_k \leftrightarrow -\mathrm{i}\hbar\partial/\partial x_k$;
4. if $f$ is a continuous function in $R^N$, one has $f(x) \leftrightarrow f(\hat{x})$ and $\hat{f}(p) = (Ff)(\hat{x})$, where $F$ denotes a Fourier transform;
5. $L_\zeta \leftrightarrow \hat{L}_\zeta, \zeta \equiv (\alpha,\beta), \alpha,\beta \in R^N$, where $L_\zeta$ is the generator of the translations in phase space in the direction $\zeta$ and $\hat{L}_\zeta$ is the generator of the one-parameter group $t \to W(t\zeta)$ associated with $\zeta$ by the Weyl system.

Note that (1) and (4) imply (2) and (3) through a limit procedure.

Under the correspondence $A \leftrightarrow \hat{A}$, linear symplectic maps correspond to unitary transformations. This is not in general the case for nonlinear maps.

One can prove that conditions (1)–(5) give a complete characterization of the map $A \leftrightarrow \hat{A}$. Moreover, the correspondence cannot be extended to other functions in phase space. Indeed, one has:

**Theorem 3** (van Hove). *Let $G$ be the class of functions $C^\infty$ on $R^{2N}$ which are generators of global symplectic flows. For $g \in G$ let $\Phi_g(t)$ be the corresponding group. There cannot exist for every $g$ a correspondence $g \leftrightarrow \hat{g}$, with $\hat{g}$ self-adjoint, such that $\hat{g}(x,p) = g(\hat{x},\hat{p})$.*

We described the Weyl quantization as a correspondence between functions in the Schwartz class $S$ and a class of bounded operators. Weyl's quantization can be extended to a much wider class of functions. Operators that can be so constructed are called Fourier integral operators. One uses the notation $\hat{\sigma} \equiv \sigma(D,x)$.

We have the following useful theorems (Robert 1987):

**Theorem 4** *Let $l_1,\ldots,l_K$ be linear functions on $R^N$ such that $\{l_i l_k\} = 0$. Let $P$ be a polynomial and let $\sigma(\xi,x) \equiv P[l_1(\xi,x), l_K(\xi,x)]$. Then*

(i) *$\sigma(D,x)$ maps $S$ in $L^2(R^N)$ and self-adjoint;*
(ii) *if $g$ is continuous, then $(g(\sigma)(D,x) = g(\sigma(D,x))$.*

One proves that $\sigma(D,x)$ extends to a continuous map $S'(X) \to S'(X)$ and, moreover,

**Theorem 5** (Calderon–Vaillancourt). *If $\sigma_0 \equiv \sum_{|\alpha|+|\beta|\leq 2N+1} |D_\xi^\alpha D_x^\beta \sigma| < \infty$ the norm of the operator $\sigma(D,x)$ is bounded by $\sigma_0$.*

Any operator obtained from a suitable class of functions through Weyl's quantization is called a pseudodifferential operator. If $\sigma(q,p) = P(p)$, where $P$ is a polynomial, $\hat{\sigma}(p,q)$ is a differential operator.

Moreover, if $\sigma(p,x) \in L^2$ then $\sigma(D,x)$ is a Hilbert–Schmidt operator and

$$|\sigma(D,x)|_{\mathrm{HS}} = (2\pi\hbar)^{-n/2}\left[\int |A(z)|^2\,\mathrm{d}z\right]^{1/2}$$

Pseudodifferential operators turn out to be very important in particular in the quantum theory of molecules (Le Bris 2003), where adiabatic analysis and Peierls substitution rules force the use of pseudodifferential operators.

The next important problem in the theory of quantization is related to dynamics.

Let $\beta$ be a quantization procedure and let $H(p,q)$ be a classical Hamiltonian on phase space. Let $A_t$ be

the evolution of a classical observable $A$ under the flow defined by $H$ and assume that $\beta(A_t)$ is well defined or all $t$.

Is there a self-adjoint operator $\hat{H}$ such that $\beta(A_t) = e^{it\hat{H}} \beta(A) e^{-it\hat{H}}$? If so, can one estimate $|\hat{H} - \beta(H)|$? Conversely, if the generator of the quantized flow is, by definition, $\hat{H}$ (as is usually assumed), is it possible to give an estimate of the difference $|\beta(A_t) - (\beta(A))_t|\phi$ for a dense set of $\phi \in \mathcal{H}$, where $A_t \equiv e^{it\hat{H}} A e^{-it\hat{H}}$, or to estimate $|\tilde{A}_t - A_t|_\infty$, where $\tilde{A}_t$ is defined by $\beta(\tilde{A}_t) = (\beta(A))_t$. Is it possible to write an asymptotic series in $\hbar$ for the differences?

For the Weyl quantization some quantitative results have been obtained if one makes use of the semiclassical observables (Robert 1987). We shall not elaborate further on this point.

For completeness, we briefly mention another quantization procedure which is often used in mathematical physics.

### Wick Quantization

This quantization assigns positive operators to positive functions, but does not preserve polynomial relations. It is strictly related to the Bargmann–Fock–Segal representation.

Call coherent state centered in the point $(y, \eta)$ of phase space the normalized solution of $(i\hat{p} + \hat{x} - i\eta + x)\phi_{y,\eta}(x) = 0$.

Wick's quantization of the classical observable $A$ is by definition the map $A \to \mathrm{Op}^W(A)$, where

$$\mathrm{Op}^W(A)\psi \equiv (2\pi\hbar)^{-n} \int A(y, \eta)(\psi, \bar{\phi}_{y,\eta})\phi_{y,\eta}\, dy\, d\eta$$

One can prove, either directly or going through Weyl's representation, that

1. if $A \geq 0$ then $\mathrm{Op}^W_\hbar(A) \geq 0$;
2. the Weyl symbol of the operator $\mathrm{Op}^W_\hbar(A)$ is
$$(\pi\hbar)^{-n} \int\int A(y, \eta) e^{-\frac{1}{\hbar}[(x-y)^2 + (\xi-\eta)^2]}\, dy\, d\eta$$
3. for every $A \in O(0)$ one has $\|\mathrm{Op}^W_\hbar(A) - \hat{A}\| = O(\hbar)$.

Wick's quantization associates with every vector $\phi \in \mathcal{H}$ a positive Radon measure $\mu_\phi$ in phase space, called Husimi measure. It is defined by $\int A\, d\mu_\psi = (\mathrm{Op}^W_\hbar(A)\psi \cdot \psi)$, $A \in S(z)$. Wick's quantization is less adapted to the treatment of nonrelativistic particles, in particular Eherenfest's rule does not apply, and the semiclassical propagation theorem has a more complicated formulation. It is very much used for the analysis in Fock space in the theory of quantized relativistic fields, where a special role is assigned to Wick ordering, according to which the polynomials in $\hat{x}_\hbar$ and $\hat{p}_\hbar$ are reordered in terms of creation and annihilation operators by placing all creation operators to the left.

We now come back to Schrödinger's equation and notice that it can be derived within Heisenberg's formalism and Weyl's quantization scheme from the Hamiltonian of an $N$-particle system in Hamiltonian mechanics (at least if one neglects spin, which has no classical analog).

Apart from (often) inessential parameters, the Schrödinger equation for $N$ scalar particles in $R^3$ can be written as

$$i\hbar \frac{\partial\phi}{\partial t} = \sum_{k=1}^{N} (i\hbar\nabla_k + A_k)^2 \phi + V\phi \equiv H\phi \qquad \phi \in L^2(R^{3N}) \tag{17}$$

where $A_k$ are vector-valued functions (vector potentials) and $V = V_k(x_k) + V_{i,k}(x_i - x_k)$ are scalar-valued function (scalar potentials) on $R^3$.

Typical problems in Schrödinger's quantum mechanics are:

1. Self-adjointness of $H$, existence of bound states (discrete spectrum of the operator), their number and distribution, and, in general, the properties of the spectrum.
2. Existence, completeness, and continuity properties of the wave operators
$$W_\pm \equiv s - \lim_{\mp\infty} e^{itH_0} e^{-itH} \tag{18}$$
and the ensuing existence and properties of the $S$-matrix and of the scattering cross sections. In [18] $H_0$ is a suitable reference operator, usually $-\Delta$ (with periodic boundary conditions if the potentials are periodic in space), for which Schrödinger's equation can be somewhat analytically controlled.
3. Existence and property of a semiclassical limit.

In [17] and [18] we have implicitly assumed that $H$ is time independent; very interesting problems arise when $H$ depends on time, in particular if it is periodic or quasiperiodic in time, giving rise to ionization phenomena. In the periodic case, one is helped by Floquet's theory, but even in this case many interesting problems are still unsolved.

If the potentials are sufficiently regular, the spectrum of $H$ consists of an absolutely continuous part (made up of several bands in the space-periodic case) and a discrete part, with few accumulation points.

On the contrary, if $V(x, \omega)$ is a measurable function on some probability space $\Omega$, with a suitable distribution (e.g., Gaussian), the spectrum may have totally different properties almost surely.

For example, in the case $N=1$ (so that the terms $V_{i,j}$ are absent) in one and two spatial dimensions the spectrum is pure point and dense, with eigenfunctions which decrease at infinity exponentially fast (although not uniformly); as a consequence, the evolution group does not give rise to a dispersive motion. The same is true in three dimensions if the potential is sufficiently strong and the kinetic energy content of the initial state is sufficiently limited. This very interesting behavior is due roughly to the randomness of the "barriers" generated by the potential and is also present, to a large extent, for potentials quasiperiodic in space (Pastur and Figotin 1992).

In these as well as in most problems related to Schrödinger's equation, a crucial role is taken by the resolvent operator $(H - \lambda I)^{-1}$, where $\lambda$ is any complex number outside the spectrum of $H$; many of the results are obtained when the difference $(H - \lambda I)^{-1} - (H_0 - \lambda I)^{-1}$ is a compact operator.

Problems of type (1) and (2) are of great physical interest, and are of course common with theoretical physics and quantum chemistry (Le Bris 2003), although the instruments of investigation are somewhat different in mathematical physics. The semiclassical limit is often more of theoretical interest, but its analysis has relevance in quantum chemistry and its methods are very useful whenever it is convenient to use multiscale methods, as in the study of molecular spectra.

We start with a brief description of point (3); it provides a valid instrument in the description of quantum-mechanical systems at a scale where it is convenient to use units in which the physical constant $\hbar$ has a very small value ($\hbar \simeq 10^{-27}$ in CGS units). From Heisenberg's commutation relations, $[\hat{x}, \hat{p}] \subset \hbar I$, it follows that the product of the dispersion (uncertainty) of the position and momentum variables is proportional to $\hbar$ and therefore at least one of these two quantities must have very large values (compared to $\hbar$). One considers usually the case in which these dispersions have comparable values, which is therefore very small, of the order of magnitude $\hbar^{1/2}$ (but very large as compared with $\hbar$). In order to make connection with the Hamilton–Jacobi formalism of classical mechanics one can also consider the case in which the dispersion in momentum is of the order $\hbar$ (the WKB method).

The semiclassical limit takes advantage mathematically from the fact that the parameter $\hbar$ is very small in natural units, and performs an asymptotic analysis, in which the terms of "lowest order" are exactly described and the difference is estimated. The problem one faces is that the Schrödinger equation becomes, in the "mathematical limit" $\hbar \to 0$, a very singular PDE (the coefficients of the differential terms go to zero in this limit).

Dividing each term of the equation by $\hbar$ (because we do not want to change the scale of time) leads, in the case of one quantum particle in $R^3$ in potential field $V(x)$ (we treat, for simplicity, only this case), to the equation

$$\mathrm{i}\frac{\partial \phi(x,t)}{\partial t} = -\hbar \Delta \phi(x,t) + \hbar^{-1} V(x) \phi(x,t) \qquad [19]$$

It is convenient therefore to "rescale" the spatial variables by a factor $\hbar^{1/2}$ (i.e., choose different units) setting $x = \sqrt{\hbar} X$ and look for solutions of [19] which remain regular in the limit $\hbar \to 0$ as functions of the rescaled variable $X$. One searches therefore for solutions that on the "physical scale" have support that becomes "vanishingly small" in the limit. It is therefore not surprising that, in the limit, these solutions may describe point particles; the main result of semiclassical analysis is that he coordinates of these particles obey Hamilton's laws of classical mechanics.

This can be roughly seen as follows (accurate estimates are needed to make this empirical analysis precise). Using multiscale analysis, one may write the solution in the form $\phi(X, x, t)$ and seek solutions which are smooth in $X$ and $x$. Both terms on the right-hand side of [19] contain contributions of order $-2$ and $-1$ in $\sqrt{\hbar}$ and in order to have regular solutions one must have cancellations between equally singular contributions. For this, one must perform an expansion to the second order of the potential (assumed at least twice differentiable) around a suitable trajectory $q(t)$, $q \in R^3$, and choose this trajectory in such a way that the cancellations take place.

A formal analysis shows that this is achieved only if the trajectory chosen is precisely a solution of the classical Lagrange equations. Of course, a more refined analysis and good estimates are needed to make this argument precise, and to estimate the error that is made when one neglects in the resulting equation terms of order $\sqrt{\hbar}$; in favorable cases, for each chosen $T$ the error in the solution for most initial conditions of the type described is of order $\sqrt{\hbar}$ for $|t| < T$.

This semiclassical result is most easily visualized using the formalism of Wigner functions (the technical details, needed to to make into a proof the formal arguments, take advantage of regularity estimates in the theory of functions).

In natural units, one defines

$$W_{\hbar,\rho}(x,\xi,t) = \left(\frac{\mathrm{i}}{2\pi}\right)^N W_\rho\left(x, \frac{\xi}{\hbar}, t\right)$$

In terms of the Wigner function $W_{\hbar,\rho}$ the Schrödinger equation [19] takes the form

$$\frac{\partial f^\hbar}{\partial t} + \xi \cdot \nabla_x f^\hbar + K_\hbar * f^\hbar = 0 \qquad [20]$$
$$\rho^\hbar(t=0) = \rho_0(\hbar)$$

where

$$K_\hbar = \frac{\mathrm{i}}{(2\pi)^N} \mathrm{e}^{-\mathrm{i}\xi,y}\hbar^{-1}\left[V\left(x+\frac{\hbar y}{2}\right) - V\left(x-\frac{\hbar y}{2}\right)\right]$$

It can be proved (Robert 1987) that if the potential is sufficiently regular and if the initial datum converges in a suitable topology to a positive measure $f_0$, then, for all times, $W_{\hbar,\rho}(x,t)$ converges to a (weak) solution of the Liouville equation

$$\frac{\partial f}{\partial t} + \xi \cdot \nabla_x f - \nabla V(x) \cdot \nabla_\xi f = 0$$

This leads to the semiclassical limit if, for example, one considers a sequence of initial data $\rho_{\phi_n}$ where $\phi_n$ is a sequence of functions centered at $x_0$ with Fourier transform centered at $p_0$ and dispersion of order $\hbar^{1/2}$ both in position and in momentum. In this case, the limit measure is a Dirac measure centered on the classical paths.

In the course of the proof of the semiclassical limit theorem, one becomes aware of the special status of the Hamiltonians that are at most quadratic in $\hat{x}$ and $\hat{p}$. Indeed, it is easy to verify that for these Hamiltonians the expectation values of $\hat{x}$ and $\hat{p}$ obey the classical equation of motion (P Ehrenfest rule).

From the point of view of Heisenberg, this can be understood as a consequence of the fact that operators at most bilinear in $a$ and $a^*$ form an algebra $\mathcal{D}$ under commutation and, moreover, the homogeneous part of order 2 is a closed subalgebra such that its action on $\mathcal{D}$ (by commutation) has the same structure as the algebra of generators of the Hamiltonian flow and its tangent flow. Apart from (important) technicalities, the proof of the semiclassical limit theorem reduces to the proof that one can estimate the contribution of the terms of order higher than 2 in the expansion of the quantum Hamiltonian at the classical trajectory as being of order $\hbar^{1/2}$ in a suitable topology (Hepp 1974).

We end this overview by giving a brief analysis of problems (1) and (2), which refer to the description of phenomena that are directly accessible to comparison with experimental data, and therefore have been extensively studied in theoretical physics and quantum chemistry (Mc Weeny 1992); some of them have been analyzed with the instruments of mathematical physics, often with considerable success. We give here a very naive introduction to these problems and refer the reader to the more specialized contributions to this Encyclopedia for a rigorous analysis and exact statements.

Of course, most of the problems of physical interest are not "exactly solvable," in the sense that rarely the final result is given explicitly in terms of simple functions. As a consequence, exact numerical results, to be compared with experimental data, are rarely obtained in physically relevant problems, and most often one has to rely on approximation schemes with (in favorable cases) precise estimates on the error.

Formal perturbation theory is the easiest of such schemes, but it seldom gives reliable results to physically interesting problems. One writes

$$H_\epsilon \equiv H + \epsilon V \qquad [21]$$

where $\epsilon$ is a small real parameter, and sets a formal scheme in case (1) by writing

$$H_\epsilon\phi_\epsilon \equiv E_\epsilon\phi_\epsilon,\ E_\epsilon \equiv \sum_0^\infty \epsilon^k E_k,\ \phi_\epsilon \equiv \sum_0^\infty \epsilon^k \phi_k$$

and, in case (2), iterating Duhamel's formula

$$\mathrm{e}^{-\mathrm{i}tH_\epsilon} = \mathrm{e}^{-\mathrm{i}tH_0} + \mathrm{i}\epsilon\int_0^t \mathrm{e}^{-\mathrm{i}(t-s)H_\epsilon} V \mathrm{e}^{-\mathrm{i}sH_0}\mathrm{d}s \qquad [22]$$

Very seldom the perturbation series converges, and one has to resort to more refined procedures.

In some cases, it turns out to be convenient to consider the formal primitive $\tilde{E}_\epsilon$ of $E_\epsilon$ (as a differentiable function of $\epsilon$) and prove that it is differentiable in $\epsilon$ for $0 < \epsilon < \epsilon_0$ (but not for $\epsilon = 0$). In favorable cases, this procedure may lead to

$$E_\epsilon = \sum_0^N \epsilon^k E_k + R_N(\epsilon), \quad \lim_{N\to\infty} |R_N|(\epsilon) = +\infty$$

with explicit estimates of $|R_N(\epsilon)|$ for $0 \le \epsilon < \epsilon_0$.

Re-summation techniques of the formal power series may be of help in some cases.

The estimate of the lowest eigenvalues of an operator bounded below is often done by variational analysis, making use of min–max techniques applied to the quadratic form $Q(\phi) \equiv (\phi, H\phi)$.

Semiclassical analysis can be useful to search for the distribution of eigenvalues and in the study of the dynamics of states whose dispersions both in position and in momentum are very large in units in which $\hbar = 1$.

A case of particular interest in molecular and atomic physics occurs when the physical parameters which appear in $H_\epsilon$ (typically the masses of the particles involved in the process) are such that one

can *a priori* guess the presence of coordinates which have a rapid dependence on time (fast variables) and a complementary set of coordinates whose dependence on time is slow. This suggests that one can try an asymptotic analysis, often in connection with adiabatic techniques. Seldom one deals with cases in which the hypotheses of elementary adiabatic theorems are satisfied, and one has to refine the analysis, mostly through subtle estimates which ensure the existence of quasi invariant subspaces.

Asymptotic techniques and refined estimates are also needed to study the effective description of a system of $N$ interacting identical particles when $N$ becomes very large; for example, in statistical mechanics, one searches for results which are valid when $N \to \infty$.

The most spectacular results in this direction are the proof of stability of matter by E Lieb and collaborators, and the study of the phenomenon of Bose–Einstein condensation and the related Gross–Pitaevskii (nonlinear Schrödinger) equation. The experimental discovery of the state of matter corresponding to a Bose–Einstein condensate is a clear evidence of the nonclassical behavior of matter even at a comparatively macroscopic size. From the point of view of mathematical physics, the ongoing research in this direction is very challenging.

One should also recognize the increasing role that research in QM is taking in applications, also in connection with the increasing success of nanotechnology. In this respect, from the point of view of mathematical physics, the study of nanostructure (quantum-mechanical systems constrained to very small regions of space or to lower-dimensional manifolds, such as sheets or graphs) is still in its infancy and will require refined mathematical techniques and most likely entirely new ideas.

Finally, one should stress the important role played by numerical analysis (Le Bris 2003) and especially computer simulations. In problems involving very many particles, present-day analytical techniques provide at most qualitative estimates and in favorable cases bounds on the value of the quantities of interest. Approximation schemes are not always applicable and often are not reliable.

Hints for a progress in the mathematical treatment of some relevant physical phenomena of interest in QM (mostly in condensed matter physics) may come from the *ab initio* analysis made by simulations on large computers; this may provide a qualitative and, to a certain extent, quantitative behavior of the solutions of Schrödinger's equation corresponding to "typical" initial conditions. In recent times the availability of more efficient computing tools has made computer simulation more reliable and more apt to concur with mathematical investigation to a fuller comprehension of QM.

## Interpretation Problems

In this section we describe some of the conceptual problems that plague present-day QM and some of the attempts that have been made to cure these problems, either within its formalism or with an altogether different approach.

### Approaches within the QM Formalism

We begin with the approaches "from within." We have pointed out that the main obstacle in the measurement problem is the description of what occurs during an act of measurement. Axiom III claims that it must be seen as a "destruction" act, and the outcome is to some extent random. The final state of the system is one of the eigenstates of the observable, and the dependence on the initial state is only through an *a priori* probability assignment; the act of measurement is therefore not a causal one, contrary to the (continuous) causal reversible description of the interaction with the environment. One should be able to distinguish *a priori* the acts of measurement from a generic interaction.

There is a further difficulty. Due to the superposition principle, if a system $\mathcal{S}$ on which we want to make a measurement of the property associated with the operator $A$ "interacts" with an instrument $\mathcal{I}$ described by the operator $S$, the final state $\xi$ of the combined system will be a coherent superposition of tensor product of (normalized) eigenstates of the two systems

$$\xi = \sum_{n,m} c_{n,m} \phi_n^A \otimes \psi_m^S, \quad \sum_{n,m} |c_{n,m}|^2 = 1 \qquad [23]$$

Measurement as described by Axiom III of QM claims that once the measurement is over, the measured system is, with probability $\sum_m |c_{n,m}|^2$, in the state $\phi_n^A$ and the instrument is in a state which carries the information about the final state of the system (after all, what one reads at the end is an indicator of the final state of the instrument).

It is therefore convenient to write $\xi$ in the form

$$\xi = \sum d_n \phi_n^A \otimes \zeta_n, \quad \sum_n |d_n|^2 = 1 \qquad [24]$$

(this defines $\zeta_n$ if the spectrum of $A$ is pure point and nondegenerate). It is seen from [24] that, due to the reduction postulate, we know that the the measured system is in the state $\phi_{n_0}^A$ if a measurement of an observable $T$ with nondegenerate spectrum,

eigenvectors $\{\zeta_n\}$, and eigenvalues $\{z_n\}$ gives the results $z_{n_0}$.

Along these lines, one does not solve the measurement problem (the outcome is still probabilistic) but at least one can find the reason why the measuring apparatus may be considered "classical."

It is more convenient to go back to [23] and to assume that one is able to construct the measuring apparatus in such a way that one divides (roughly) its pure (microscopic) states in sets $\Phi_n$ (each corresponding to a "macroscopic" state) which are (roughly) in one-to-one correspondence to the eigenstates of $A$. The sets $\Phi_n$ contain a very large number, $N_{\Phi_n}$, of elements, so that the sets $\Phi_n$ need not be given with extreme precision. And the sets $\Phi_n$ must be in a sense "stable" under small external perturbations.

It is clear from this rough description that the apparatus should contain a large number of small components and still its interaction with the "small" system $A$ should lead to a more or less sudden change of the sets $\Phi_n$.

A concrete model of this mechanism has been proposed by K Hepp (1972) for the case when $A$ is a $2 \times 2$ matrix, and the measuring apparatus is made of a chain of $N$ spins, $N \to \infty$; the analysis was recently completed by Sewell (2005) with an estimate on the error which is made if $N$ is finite but large. This is a dynamical model, in which the observable $A$ (a spin) interacts with a chain of spins ("moves over the spins") leaving the trace of its passage. It is this trace (final macroscopic state of the apparatus) which is measured and associated with the final state of $A$. The interaction is not "instantaneous" but may require a very short time, depending on the parameters used to describe the apparatus and the interaction.

We call "decoherence" the weakening of the superposition principle due to the interaction with the environment.

Two different models of decoherence have been analyzed in some detail; we shall denote them thermal-bath model and scattering model; both are dynamical models and both point to a solution, to various extents, of the problem of the reduction to a final density matrix which commutes with the operator $A$ (and therefore to the suppression of the interference terms).

The thermal-bath model makes use of the Heisenberg representation and relies on results of the theory of $C^*$-algebras. This approach is closely linked with (quantum) statistical mechanics; its aim is to prove, after conditioning with respect to the degrees of freedom of the bath, that a special role emerges for a commuting set of operators of the measured system, and these are the observables that specify the outcome of the measurement in probabilistic terms.

The scattering approach relies on the Schrödinger approach to QM, and on results from the theory of scattering. This approach describes the interaction of the system $\mathcal{S}$ (typically a heavy particle) with an environment made of a large number of light particles and seeks to describe the state of $\mathcal{S}$ after the interaction when one does not have any information on the final state of the light particle. One seeks to prove that the reduced density matrix is (almost) diagonal in a given representation (typically the one given by the spatial coordinates). This defines the observable (typically, position) that can be measured and the probability of each outcome.

Both approaches rely on the loss of information in the process to cancel the effect of the superposition principle and to bring the measurement problem within the realm of classical probability theory. None of them provides a causal dependence of the result of the measurement on the initial state of the system.

We describe only very briefly these attempts.

In its more basic form, the "scattering approach" has as starting point the Schrödinger equation for a system of two particles, one of which has mass very much smaller than the other one. The heavy particle may be seen as representing the system on which a measurement is being made. The outline of the method of analysis (which in favorable cases can be made rigorous) (Joos and Zeh 1985, Tegmark 1993) is the following. One chooses units in which the mass of the heavy particle is 1, and one denotes by $\epsilon$ the mass of the light particle. If $x$ is the coordinate of the heavy particle and $y$ that of the light one, and if the initial state of the system is denoted by $\Phi_0(x, y)$, the solution of the equation for the system is (apart from inessential factors)

$$\Phi_t = \exp\{\mathrm{i}(-\Delta_x - \epsilon^{-1}\Delta_y + W(x) + V(x-y))t\}\Phi_0$$

Making use of center-of-mass and relative coordinates, one sees that when $\epsilon$ is very small one should be able to describe the system on two timescales, one fast (for the light particle) and one slow (for the heavy one) and, therefore, place oneself in a setting which may allow the use of adiabatic techniques. In this setting, for the measure of the heavy particle (e.g., its position) one may be allowed to consider the light particle in a scattering regime, and use the wave operator corresponding to a potential $V_x(y) \equiv V(y - x)$.

Taking the partial trace with respect to the degrees of freedom of the light particle (this

corresponds to no information of its final state) one finds, at least heuristically, that the state of the heavy particle is now described (due to the trace operation) by a density matrix $\sigma$ for which in the coordinate representation the off-diagonal terms $\sigma_{x,x'}$ are slightly suppressed by a factor $\xi_{x,x'} = 1 - (W_x^+\psi, W_{x'}^+\psi)$ where $\psi$ represents the initial state of the light particle and $W_x^+$ is the wave operator for the motion of the light particle in the potential $\epsilon V_x$. One must assume that function $\phi$ which represents the initial state of the heavy particle is sufficiently localized so that $\xi_{x,x'} < 1$ for every $x' \neq x$ in its support.

If the environment is made of very many particles (their number $N(\epsilon)$ must be such that $\lim_{\epsilon \to 0} \epsilon N(\epsilon) = \infty$) and the heavy particle can be supposed to have separate interactions with all of them, the off-diagonal elements of the density matrix tend to 0 as $\epsilon \to 0$ and the resulting density matrix tends to have the form $\Phi(x, x') = \delta(x - x') \rho(x), \rho(x) \geq 0, \int \rho(x)\, dx = 1$. If it can be supposed that all interactions take place within a time $T(\epsilon) \leq \epsilon^\alpha$, $\alpha > 0$ one has $\rho(x) = |\psi(x)|^2$.

If the interactions are not independent, the analysis becomes much more involved since it has to be treated by many-body scattering theory; this suggests that the scattering approach can be hardly used in the context of the "thermal-bath model." In any case, the selection of a "preferred basis" (the coordinate representation) depends on the fact that one is dealing with a scattering phenomenon. A few steps have been made for a rigorous analysis (Teta 2004) but we are very far from a mathematically satisfactory answer.

The thermal-bath approach has been studied within the algebraic formulation of QM and stands on good mathematical ground (Alicki 2002, Blanchard *et al.* 2003, Sewell 2005). Its drawback is that it is difficult to associate the formal scheme with actual physical situations and it is difficult to give a realistic estimate on the decoherence time.

The thermal-bath approach attributes the decoherence effect to the practical impossibility of distinguishing between a vast majority of the pure states of the systems and the corresponding statistical mixtures. In this approach, the observables are represented by self-adjoint elements of a weakly closed subalgebra $\mathcal{M}$ of all bounded operators $\mathcal{B}(\mathcal{H})$ on a Hilbert space $\mathcal{H}$. This subalgebra may depend on the measuring apparatus (i.e, not all the apparatuses are fit to measure a set of observables). A "classical" observable by definition commutes with all other observables and therefore must belong to the center of $\mathcal{A}$ which is isomorphic to a collection of functions on a probability space $\mathcal{M}$. So the appearance of classical properties of a quantum system corresponds to the "emergence" of an algebra with nontrivial center. Since automorphic evolutions of an algebra preserve its center, this program can be achieved only if we admit the loss of quantum coherence, and this requires that the quantum systems we describe are open and interact with the environment, and moreover that the commutative algebra which emerges be stable for time evolution.

It may be shown that one must consider quantum environment in the thermodynamic limit, that is, consider the interaction of the system to be measured with a thermal bath. A discussion of the possible emergence of classical observables and of the corresponding dynamics is given by Gell-Mann (1993). In all these approaches, the commutative subalgebra is selected by the specific form of the interaction; therefore, the measuring apparatus determines the algebra of classical observables.

On the experimental side, a number of very interesting results have been obtained, using very refined techniques; these experiments usually also determine the "decoherence time." The experimental results, both for the collision model (Hornberger *et al.* 2003) and for the thermal-bath model (Hackermueller *et al.* 2004), are done mostly with fullerene (a molecule which is heavy enough and is not deflected too much after a collision with a particle of the gas). They show a reasonable accordance with the (rough) theoretical conclusions.

The most refined experiments about decoherence are those connected with quantum optics (circularly polarized atoms in superconducting cavities). These are not related to the wave nature of the particles but in a sense to the "wave nature" of a photon as a single unit. The electromagnetic field is now regarded as an incoherent superposition of states with an arbitrarily large number of photons. Polarized photons can be produced one by one, and they retain their individuality and their polarization until each of them interacts with "the environment" (e.g., the boundary of the cavity or a particle of the gas). In a sense, these experimental results refer to a "decoherence by collision" theory.

The experiments by Haroche (2003) prove that coherence may persist for a measurable interval of time and are the most controlled experiments on coherence so far.

### Other Approaches

We end this section with a brief discussion of the problem of "hidden variables" and a presentation of an entirely different approach to QM, originated by

D Bohm (1952) and put recently on firm mathematical grounds by Duerr *et al.* (1999). The approach is radically different from the traditional one and it is not clear at present whether it can give a solution to the measurement problem and a description of all the phenomena which traditional QM accounts for. But it is very interesting from the point of view of the mathematics involved.

We have remarked that the formulation of QM that is summarized in the three axioms given earlier has many unsatisfactory aspects, mainly connected with the superposition principle (described in its extremal form by the Schrödinger's cat "paradox") and with the problem of measurement which reveals, for example, through the Einstein–Rosen–Podolski "paradox," an intrinsic nonlocality if one maintains that their "objective" properties can be attributed to systems which are far apart. From the very beginning of QM, attempts have been made to attribute these features to the presence of "hidden variables"; the statistical nature of the predictions of QM is, from this point of view, due to the incompleteness of the parameters used to describe the systems. The impossibility of matching the statistical prediction of QM (confirmed by experimental findings) with a local theory based on hidden variables and classical probability theory has been known for sometime (Kochen and Specker 1967), also through the use of "Bell inequalities" (Bell 1964) among correlations of outcomes of separate measurements performed on entangled system (mainly two photons or two spin-1/2 particles created in a suitable entangled state).

A proof of the intrinsic nonlocality of QM (in the above sense) was given by L Hardy (see Haroche (2003)).

While experimental results prove that one cannot substitute QM with a "naive" theory of hidden variables, more refined attempts may have success. We shall only discuss the approach of Bohm (following a previous attempt by de Broglie) as presented in Duerr *et al.* (1999). It is a dynamical theory in which representative points follow "classical paths" and their motion is governed by a time-dependent vector "velocity" field (in this sense, it is not Newtonian). In a sense, Bohmian mechanics is a minimal completion of QM if one wants to keep the position as primitive observable. To these primitive objects, Bohm's theory adds a complex-valued function $\phi$ (the "guiding wave" in Bohm's terminology) defined on the configuration space $\mathcal{Q}$ of the particles. In the case of particles with spin, the function $\phi$ is spinor-valued. Dynamics is given by two equations: one for the coordinates of the particles and one for the guiding wave. If $x \equiv x_1, \ldots, x_N$ describes the configuration of the points, the dynamics in a potential field $V(x)$ is described in the following way: for the wave $\phi$ by a nonrelativistic Schrödinger equation with potential $V$ and for the coordinates by the ordinary differential equation (ODE)

$$\dot{x}_k = (\hbar/m_k)\mathrm{Im}\left[\frac{\phi^* \nabla_k \phi}{\phi^* \phi}\right](x), \quad x_k \in R^3$$

where $m_k$ is the mass of the $m$th particle.

Notice that the vector field is singular at the zeros of the wave function, therefore global existence and uniqueness must be proved. To see why Bohmian mechanics is empirically equivalent to QM, at least for measurement of position, notice that the equation for the points coincides with the continuity equation in QM. It follows that if one has at time zero a collection of points distributed with density $|\phi_0|^2$, the density at time $t$ will be $|\phi(t)|^2$ where $\phi(t)$ is the solution of the Schrödinger equation with initial datum $\phi_0$.

Bohm (1952) formulated the theory as a modification of Newton's laws (and in this form it has been widely used) through the introduction of a "quantum potential" $V_{\mathrm{Q}}$. This was achieved by writing the wave function in its polar form $\phi = R\mathrm{e}^{\mathrm{i}S/\hbar}$ and writing the continuity equation as a modified Hamilton–Jacobi equation. The version of Bohm's theory discussed in Duerr *et al.* (1999) introduces only the guiding wave function and the coordinates of the points, and puts the theory on firm mathematical grounds. Through an impressive series of mathematical results, these authors and their collaborators deal with the completeness of the velocity vector field, the asymptotic behavior of the points trajectories (both for the scattering regime and for the trapped trajectories, which are shown to correspond to bound states in QM), with a rigorous analysis of the theorem on the flux across a surface (a cornerstone in scattering theory) and the detailed analysis of the "two-slit" experiment through a study of the interaction with the measuring apparatus. The theory is completely causal, both for the trajectories of the points and for the time development of the pilot wave, and can also accommodate points with spin. It leads to a mathematically precise formulation of the semiclassical limit, and it may also resolve the measurement problem by relating the pilot wave of the entire system to its approximate decomposition in incoherent superposition of pilot wave associated with the particle and to the measuring apparatus (this would be the way to see the "collapse of the wave function" in QM). A weak point of this approach is the relation of the representative points with observable quantities.

## Further Reading


Alicki R (2004) Pure decoherence in quantum systems. *Open Syst. Inf. Dyn.* 11: 53–61.

Araki H (1988) In: Jorgensen P and Muhly P (eds.) *Operator Algebras and Mathematical Physics*, Contemporary Mathematics 62. Providence, RI: American Mathematical Society.

Araki H and Ezawa H (eds.) (2004) *Topics in the theory of Schroedinger Operators*. River Edge, NJ: World Scientific.

Bach V, Froelich J, and Sigal IM (1998) Quantum electrodynamics of constrained non relativistic particles. *Advanced Mathematics* 137: 299–395.

Bargmann V (1967) On a Hilbert space of analytic functions and an associated integral transform. *Communications of Pure and Applied Mathematics* 20: 1–101.

Bell J (1966) On the problem of hidden variables in quantum mechanics. *Reviews of Modem Physics* 38: 4247–4280.

Blanchard P and Dell'Antonio GF (eds.) (2004) Multiscale methods in quantum mechanics, theory and experiments. *Trends in Mathematics*. Boston: Birkhauser.

Blanchard P and Olkiewiz R (2003) Decoherence in the Heisenberg representation. *International Journal of Physics B* 18: 501–507.

Bohm D (1952) A suggested interpretation of quantum theory in terms of "hidden" variables I, II. *Physical Reviews* 85: 161–179, 180–193.

Bohr N (1913) On the constitution of atoms and molecules. *Philosophical Magazine* 26: 1–25, 476–502, 857–875.

Bohr N (1918) On the quantum theory of line spectra. *Kongelige Danske Videnskabernes Selskabs Skrifter Series* 8, IV, 1, 1–118.

Born M (1924) Uber quantenmechanik. *Zeitschrift fur Physik* 32: 379–395.

Born M and Jordan P (1925) Zur quantenmechanik. *Zeitschrift fur Physik* 34: 858–888.

Born M, Jordan P, and Heisenberg W (1926) Zur quantenmechanik II. *Zeitschrift fur Physik* 35: 587–615.

Cycon HL, Frese RG, Kirsh W, and Simon B (1987) Schroedinger operators with application to quantum mechanics and geometry. *Texts and Monogrphs in Physics.* Berlin: Springer Verlag.

de Broglie L (1923) Ondes et quanta. *Comptes Rendue* 177: 507–510.

Dell'Antonio GF (2004) On decoherence. *Journal of Mathematical Physics* 44: 4939–4955.

Dirac PAM (1925) The fundamental equations of quantum mechanics. *Proceedings of the Royal Society of London A* 109: 642–653.

Dirac PAM (1926) The quantum algebra. *Proceedings of the Cambridge Philosophical Society* 23: 412–428.

Dirac PAM (1928) The quantum theory of the electron. *Proc. Royal Soc. London A* 117: 610–624, 118: 351–361.

Duerr D, Golstein S, and Zanghì N (1996) Bohmian mechanics as the foundation of quantum mechanics. *Boston Studies Philosophical Society* 184: 21–44. Dordrecht: Kluwer Academic.

Einstein A (1905) *The Collected Papers of Albert Einstein*, vol. 2, pp. 347–377, 564–585. Princeton, NJ: Princeton University Press.

Einstein A (1924–1925) Quantentheorie des einatomigen idealen gases. *Berliner Berichte* (1924) 261–267, (1925) 3–14.

Gell-Mann M and Hartle JB (1997) Strong decoherence. *Quantum-Classical Correspondence*, pp. 3–35. Cambridge, MA: International Press.

Gross L (1972) Existence and uniqueness of physical ground state. *Journal of Functional Analysis* 19: 52–109.

Haroche S (2003) Quantum Information in cavity quantum electrodynamics. *Royal Society of London Philosophical Transactions Serial A Mathematical and Physics Engineering Science* 361: 1339–1347.

Heisenberg W (1925) Uber Quanteteoretische Umdeutung Kinematisches und Mechanischer Beziehungen. *Zeitschrift fur Physik* 33: 879–893.

Heisenberg W (1926) Uber quantentheoretische Kinematik und Mechanik. *Matematishes Annalen* 95: 694–705.

Hepp K (1974) The classical limit of quantum correlation functions. *Communications in Mathematical Physics* 35: 265–277.

Hepp K (1975) Results and problem in the irreversible statistical mechanics of open systems. *Lecture Notes in Physics 39*. Berlin: Spriger Verlag.

Hornberger K and Sype EJ (2003) Collisional decoherence reexamined. *Physical Reviews A* 68: 012105, 1–16.

Islop P and Sigal S (1996) Introduction to spectral theory with application to Schroedinger operators. *Applied Mathematical Sciences 113*. New York: Springer Verlag.

Jammer M (1989) *The Conceptual Development of Quantum Mechanics*, 2nd edn. Tomash Publishers, American Institute of Physics.

Joos E *et al*. (eds.) (2003) *Quantum Theory and the Appearance of a Classical World*, second edition. Berlin: Springer Verlag.

Kochen S and Speker EP (1967) The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics* 17: 59–87.

Le Bris C (2002) Problematiques numeriques pour la simulation moleculaire. *ESAIM Proceedings of the 11th Society on Mathematics and Applied Industries*, pp. 127–190. Paris.

Le Bris C and Lions PL (2005) From atoms to crystals: a mathematical journey. *Bulletin of the American Mathematical Society (NS)* 42: 291–363.

Lieb E (1990) From atoms to stars. *Bulletin of the American Mathematical Society* 22: 1–49.

Mackey GW (1963) *Mathematical Foundations of Quantum Mechanics.* New York–Amsterdam: Benjamin.

Mc Weeny E (1992) An overview of molecular quantum mechanics. *Methods of Computational Molecular Physics.* New York: Plenum Press.

Nelson E (1973) Construction of quantum fields from Markoff fields. *Journal of Functional Analysis* 12: 97–112.

von Neumann J (1996) Mathematical foundation of quantum mechanics. *Princeton Landmarks in Mathematics.* Princeton NJ: Princeton University Press.

Nielsen M and Chuang I (2000) *Quantum Computation and Quantum Information.* Cambridge, MA: Cambridge University Press.

Ohya M and Petz D (1993) *Quantum Entropy and Its Use*. Text and Monographs in Physics. Berlin: Springel Verlag.

Pauli W (1927) Zur Quantenmechanik des magnetische Elektron. *Zeitschrift fur Physik* 43: 661–623.

Pauli W (1928) *Collected Scientific Papers*, vol. 2. 151–160, 198–213, 1073–1096.

Robert D (1987) *Aoutur de l'approximation semi-classique*. Progress in Mathematics 68. Boston: Birkhauser.

Schroedinger E (1926) Quantizierung als Eigenwert probleme. *Annalen der Physik* 79: 361–376, 489–527, 80: 437–490, 81: 109–139.

Segal I (1996) *Quantization, Nonlinear PDE and Operator Algebra*, pp. 175–202. Proceedings of the Symposium on Pure Mathematics 59. Providence, RI: American Mathematical Society.

Sewell J (2004) Interplay between classical and quantum structure in algebraic quantum theory. *Rend, Circ. Mat. Palermo Suppl.* 73: 127–136.

Simon B (2000) Schrodinger operators in the twentieth century. *Journal of Mathematical Physics* 41: 3523–3555.

Streater RF and Wightman AS (1964) *PCT, Spin and Statistics and All That*. New York–Amsterdam: Benjamin.

Takesaki M (1971) One parameter autmorphism groups and states of operator algebras. *Actes du Congrés International des Mathamaticiens Nice*, 1970, Tome 2, pp. 427–432. Paris: Gauthier Villars.

Teta A (2004) On a rigorous proof of the Joos–Zeh formula for decoherence in a two-body problem. *Multiscale Methods in Quantum Mechanics*, pp. 197–205. Trends in Mathematics. Boston: Birkhauser.

Weyl A (1931) *The Theory of Groups and Quantum Mechanics*. New York: Dover.

Wiener N (1938) The homogeneous chaos. *American Journal of Mathematics* 60: 897–936.

Wigner EP (1952) Die Messung quantenmechanischer operatoren. *Zeitschrift fur Physik* 133: 101–108.

Yafaev DR (1992) Mathematical scattering theory. *Transactions of Mathematical Monographs*. Providence, RI: American Mathematical Society.

Zee HI (1970) On the interpretation of measurement in quantum theory. *Foundations of Physics* 1: 69–76.

Zurek WH (1982) Environment induced superselection rules. *Physical Reviews D* 26(3): 1862–1880.

# Introductory Article: Topology

**Tsou Sheung Tsun**, University of Oxford, Oxford, UK



## Introduction

This will be an elementary introduction to general topology. We shall not even touch upon algebraic topology, which will be dealt with in Cohomology Theories, although in some mathematics departments it is introduced in an advanced undergraduate course.

We believe such an elementary article is useful for the encyclopaedia, purely for quick reference. Most of the concepts will be familiar to physicists, but usually in a general rather vague sense. This article will provide the rigorous definitions and results whenever they are needed when consulting other articles in the work. To make sure that this is the case, we have in fact experimentally tested the article on physicists for usefulness.

Topology is very often described as "rubber-sheet geometry," that is, one is allowed to deform objects without actually breaking them. This is the all-important concept of continuity, which underlies most of what we shall study here.

We shall give full definitions, state theorems rigorously, but shall not give any detailed proofs. On the other hand, we shall cite many examples, with a view to applications to mathematical physics, taking for granted that familiar more advanced concepts there need not be defined. By the same token, the choice of topics will also be so dictated.

## ",1,5,1,0,0pc,0pc,0pc,0pc>Essential Concepts

**Definition 1** Let $X$ be a set. A collection $\mathcal{T}$ of subsets of $X$ is called a *topology* if the following are satisfied:

(i) $\emptyset, X \in \mathcal{T}$.

(ii) Let $\mathcal{I}$ be an index set. then

$$A_\alpha \in \mathcal{T}, \alpha \in \mathcal{I} \implies \bigcup_{\alpha \in \mathcal{I}} A_\alpha \in \mathcal{T}$$

(iii) $A_i \in \mathcal{T}, i = 1, \ldots, n \implies \bigcap_{i=1}^{n} A_i \in \mathcal{T}$.

**Definition 2** A member of the topology $\mathcal{T}$ is called an *open set* (of $X$ with topology $\mathcal{T}$).

**Remark** The last two properties are more easily put as *arbitrary* unions of open sets are open, and *finite* intersections of open sets are open. One can easily see the significance of this: if we take the "usual topology" (which will be defined in due course) of the real line, then the intersection of all open intervals $(-1/n, 1/n), n$ a positive integer, is just the single point {0}, which is manifestly not open in the usual sense.

**Example** If we postulate that $\emptyset$, and the entire set $X$, are the only open subsets, we get what is called the *indiscrete* or *coarsest* topology. At the other extreme, if we postulate that all subsets are open, then we get the *discrete* or *finest* topology. Both seem quite unnatural if we think in terms of the real line or plane, but in fact it would be more unnatural to explicitly exclude them from the definition. They prove to be quite useful in certain respects.

**Definition 3** A subset of $X$ is *closed* if its complement in $X$ is open.

**Remarks**

(i) One could easily build a topology using closed sets instead of open sets, because of the simple relation that the complement of a union is the intersection of the complements.

(ii) From the definitions, there is nothing to prevent a set being both open and closed, or neither

**Definition 4** A set equipped with a topology is called a *topological space* (with respect to the given topology). Elements of a topological space are sometimes called *points*.

**Definition 5** Let $x \in X$. A *neighborhood* of $x$ is a subset of $X$ containing an open set which contains $x$.

**Remark** This seems a clumsy definition, but turns out to be more useful in the general case than restricting to open neighborhoods, which is often done.

**Definition 6** A subcollection of open sets $\mathcal{B} \subseteq \mathcal{T}$ is called a *basis* for the topology $\mathcal{T}$ if every open set is a union of sets of $\mathcal{B}$.

**Definition 7** A subcollection of open sets $\mathcal{S} \subseteq \mathcal{T}$ is called a *sub-basis* for the topology $\mathcal{T}$ if every open set is a union of finite intersections of sets of $\mathcal{S}$.

**Definition 8** The *closure* $\bar{A}$ of a subset $A$ of $X$ is the smallest closed set containing $A$.

**Definition 9** The *interior* $\mathring{A}$ of a subset $A$ of $X$ is the largest open set contained in $A$.

**Remark** It is sometimes useful to define the *boundary* of $A$ as the set $\bar{A} \backslash \mathring{A} = \{x \in \bar{A}, x \notin \mathring{A}\}$.

**Definition 10** Let $A$ be a subset of a topological space $X$. A point $x \in X$ is called a *limit point* of $A$ if every open set containing $x$ contains some point of $A$ other than $x$.

**Definition 11** A subset $A$ of $X$ is said to be *dense* in $X$ if $\bar{A} = X$.

**Definition 12** A topological space $X$ is called a *Hausdorff space* if for any two distinct points $x, y \in X$, there exist an open neighborhood of $A$ of $x$ and an open neighborhood $B$ of $y$ such that $A$ and $B$ are disjoint (that is, $A \cap B = \emptyset$).

**Remark and Examples**

(i) This is looking more like what we expect. However, certain mildly non-Hausdorff spaces turn out to be quite useful, for example, in twistor theory. A "pocket" furnishes such an example. Explicitly, consider $X$ to be the subset of the real plane consisting of the interval $[-1, 1]$ on the $x$-axis, together with the interval $[0, 1]$ on the line $y = 1$, where the following pairs of points are identified: $(x, 0) \cong (x, 1), 0 < x \leq 1$. Then the two points (0, 0) and (0, 1) do not have any disjoint neighborhoods. Strictly speaking, one needs the notion of a quotient topology, introduced below.

(ii) For a more "truly" non-Hausdorff topology, consider the space of positive integers $\mathbb{N} = \{1, 2, 3, \dots\}$, and take as open sets the following: $\emptyset$, $\mathbb{N}$, and the sets $\{1, 2, \dots, n\}$ for each $n \in \mathbb{N}$. This space is neither Hausdorff nor compact (see later for definition of compactness).

**Definition 13** Let $X$ and $Y$ be two topological spaces and let $f : X \to Y$ be a map from $X$ to $Y$. We say that $f$ is *continuous* if $f^{-1}(A)$ is open (in $X$) whenever $A$ is open (in $Y$).

**Remark** Continuity is the single most important concept here. In this general setting, it looks a little different from the "$\epsilon$–$\delta$" definition, but this latter works only for metric spaces, which we shall come to shortly.

**Definition 14** A map $f : X \to Y$ is a *homeomorphism* if it is a continuous bijective map such that its inverse $f^{-1}$ is also continuous.

**Remark** Homeomorphisms are the natural maps for topological spaces, in the sense that two homeomorphic spaces are "indistinguishable" from the point of view of topology. Topological invariants are properties of topological spaces which are preserved under homeomorphisms.

**Definition 15** Let $B \subseteq A$. Then one can define the *relative topology* of $B$ by saying that a subset $C \subseteq B$ is open if and only if there exists an open set $D$ of $A$ such that $C = D \cap B$.

**Definition 16** A subset $B \subseteq A$ equipped with the relative topology is called a *subspace* of the topological space $A$.

**Remark** Thus, if for subsets of the real line, we consider $A = [0, 3], B = [0, 2]$, then $C = (1, 2]$ is open in $B$, in the relative topology induced by the usual topology of $\mathbb{R}$.

**Definition 17** Given two topological spaces $X$ and $Y$, we can define a *product topological space* $Z = X \times Y$, where the set is the Cartesian product of the two sets $X$ and $Y$, and sets of the form $A \times B$, where $A$ is open in $X$ and $B$ is open in $Y$, form a basis for the topology.

**Remark** Note that the open sets of $X \times Y$ are not always of this product form $(A \times B)$.

**Definition 18** Suppose there is a partition of $X$ into disjoint subsets $A_\alpha, \alpha \in \mathcal{I}$, for some index set $\mathcal{I}$, or equivalently, there is defined on $X$ an equivalence relation $\sim$. Then one can define the *quotient topology* on the set of equivalence classes $\{A_\alpha, \alpha \in \mathcal{I}\}$, usually denoted as the quotient space $X/\sim\, = Y$, as follows. Consider the map $\pi : X \to Y$, called the canonical projection, which maps the element $x \in X$ to its equivalence class $[x]$. Then a subset $U \subseteq Y$ is open if and only if $\pi^{-1}(U)$ is open.

**Proposition 1** *Let $\mathcal{T}$ be the quotient topology on the quotient space $Y$. Suppose $\mathcal{T}'$ is another*

*topology on Y such that the canonical projection is continuous, then* $\mathcal{T}' \subseteq \mathcal{T}$.

**Definition 19** An (*open*) *cover* $\{U_\alpha : \alpha \in \mathcal{I}\}$ for $X$ is a collection of open sets $U_\alpha \subseteq X$ such that their union equals $X$. A subcover of this cover is then a subset of the collection which is itself a cover for $X$.

**Definition 20** A topological space $X$ is said to be *compact* if every cover contains a finite subcover.

**Remark** So for a compact space, however one chooses to cover it, it is always sufficient to use a finite number of open subsets. This is one of the essential differences between an open interval (not compact) and a closed interval (compact). The former is in fact homeomorphic to the entire real line.

**Definition 21** A topological space $X$ is said to be *connected* if it cannot be written as the union of two nonempty disjoint open sets.

**Remark** A useful equivalent definition is that any continuous map from $X$ to the two-point set {0, 1}, equipped with the discrete topology, cannot be surjective.

**Definition 22** Given two points $x$, $y$ in a topological space $X$, a *path* from $x$ to $y$ is a continuous map $f : [0,1] \to X$ such that $f(0) = x, f(1) = y$. We also say that such a path joins $x$ and $y$.

**Definition 23** A topological space $X$ is *path-connected* if every two points in $X$ can be joined by a path lying entirely in $X$.

**Proposition 2** *A path-connected space is connected.*

**Proposition 3** *A connected open subspace of* $\mathbb{R}^n$ *is path-connected.*

**Definition 24** Given a topological space $X$, define an equivalence relation by saying that $x \sim y$ if and only if $x$ and $y$ belong to the same connected subspace of $X$. Then the equivalence classes are called (*connected*) *components* of $X$.

**Examples**

(i) The Lie group O(3) of $3 \times 3$ orthogonal matrices has two connected components. The identity connected component is SO(3) and is a subgroup.
(ii) The proper orthochronous Lorentz transformations of Minkowski space form the identity component of the group of Lorentz transformations.

## Metric Spaces

A special class of topological spaces plays an important role: metric spaces.

**Definition 25** A *metric space* is a set $X$ together with a function $d : X \times X \to \mathbb{R}$ satisfying

(i) $d(x,y) \geq 0$,
(ii) $d(x,y) = 0 \Leftrightarrow x = y$,
(iii) $d(x,z) \leq d(x,y) + d(y,z)$ ("triangle inequality").

**Remarks**

(i) The function $d$ is called the metric, or distance function, between the two points.
(ii) This concept of metric is what is generally known as "Euclidean" metric in mathematical physics. The distinguishing feature is the positive definiteness (and the triangle inequality). One can, and does, introduce indefinite metrics (for example, the Minkowski metric) with various signatures. But these metrics are not usually used to induce topologies in the spaces concerned.

**Definition 26** Given a metric space $X$ and a point $x \in X$, we define the *open ball* centred at $x$ with radius $r$ (a positive real number) as

$$B_r(x) = \{y \in X : d(x,y) < r\}$$

Given a metric space $X$, we can immediately define a topology on it by taking all the open balls in $X$ as a basis. We say that this is the topology induced by the given metric. Then we can recover our usual "$\epsilon$–$\delta$" definition of continuity.

**Proposition 4** *Let* $f : X \to Y$ *be a map from the metric space X to the metric space Y. Then f is continuous* (*with respect to the corresponding induced topologies*) *at* $x \in X$ *if and only if given any* $\epsilon > 0, \exists \delta > 0$ *such that* $d(x,x') < \delta$ *implies* $d(f(x,), f(x')) < \epsilon$.

Note that we do not bother to give two different symbols to the two metrics, as it is clear which spaces are involved. The proof is easily seen by taking the relevant balls as neighborhoods. Equally easy is the following:

**Proposition 5** *A metric space is Hausdorff.*

**Definition 27** A map $f : X \to Y$ of metric spaces is *uniformly continuous* if given any $\epsilon > 0$ there exists $\delta > 0$ such that for any $x_1, x_2 \in X, d(x_1, x_2) < \delta$ implies $d(f(x_1), f(x_2)) < \epsilon$.

**Remark** Note the difference between continuity and uniform continuity: the latter is stronger and requires the same $\delta$ for the whole space.

**Definition 28** Two metrics $d_1$ and $d_2$ defined on $X$ are *equivalent* if there exist positive constants $a$ and $b$ such that for any two points $x, y \in X$ we have

$$ad_1(x,y) \leq d_2(x,y) \leq bd_1(x,y)$$

**Remark** This is clearly an equivalence relation. Two equivalent metrics induce the same topology.

**Examples**

(i) Given a set $X$, we can define the discrete metric as follows: $d_0(x, y) = 1$ whenever $x \neq y$. This induces the discrete topology on $X$. This is quite a convenient way of describing the discrete topology.

(ii) In $\mathbb{R}$, the usual metric is $d(x, y) = |x - y|$, and the usual topology is the one induced by this.

(iii) More generally, in $\mathbb{R}^n$, we can define a metric for every $p \geq 1$ by

$$d_p(x, y) = \left\{ \sum_{k=1}^{n} |x_k - y_k|^p \right\}^{1/p}$$

where $x = (x_1, x_2, \ldots, x_n), y = (y_1, y_2, \ldots, y_n)$. In particular, for $p = 2$ we have the usual Euclidean metric, but the other cases are also useful. To continue the series, one can define

$$d_\infty = \max_{1<k<n} \{|x_k - y_k|\}$$

All these metrics induce the same topology on $\mathbb{R}^n$.

(iv) In a vector space $V$, say over the real or the complex field, a function $\|\cdot\| : V \to \mathbb{R}^+$ is called a *norm* if it satisfies the following axioms:
(a) $\|x\| = 0$ if and only if $x = 0$,
(b) $\|\alpha x\| = |\alpha| \|x\|$, and
(c) $\|x + y\| \leq \|x\| + \|y\|$.

Then it is easy to see that a metric can be defined using the norm

$$d(x, y) = \|x - y\|$$

In many cases, for example, the metrics defined in example (iii) above, one can define the norm of a vector as just the distance of it from the origin. One obvious exception is the discrete metric.

A slightly more general concept is found to be useful for spaces of functions and operators: that of seminorms. A *seminorm* is one which satisfies the last two of the conditions, but not necessarily the first, for a norm, as listed above.

**Definition 29** Given a metric space $X$, a sequence of points $\{x_1, x_2, \ldots\}$ is called a *Cauchy sequence* if, given any $\epsilon > 0$, there exists a positive integer $N$ such that for any $k, \ell > N$ we have $d(x_k, x_\ell) < \epsilon$.

**Definition 30** Given a sequence of points $\{x_1, x_2, \ldots\}$ in a metric space $X$, a point $x \in X$ is called a *limit* of the sequence if given any $\epsilon > 0$, there exists a positive integer $N$ such that for any $n > N$ we have $d(x, x_n) < \epsilon$. We say that the sequence *converges* to $x$.

**Definition 31** A metric space $X$ is *complete* if every Cauchy sequence in $X$ converges to a limit in it.

**Examples**

(i) The closed interval $[0, 1]$ on the real line is complete, whereas the open interval $(0, 1)$ is not. For example, the Cauchy sequence $\{1/n, n = 2, 3, \ldots\}$ has no limit in this open interval. (Considered as a sequence on the real line, it has of course the limit point 0.)

(ii) The spaces $\mathbb{R}^n$ are complete.

(iii) The Hilbert space $\ell^2$ consisting of all sequences of real numbers $\{x_1, x_2, \ldots\}$ such that $\sum_1^\infty x_k^2$ converges is complete with respect to the obvious metric which is a generalization to infinite dimension of $d_2$ above. For arbitrary $p \geq 1$, one can similarly define $\ell^p$, which are also complete and are hence Banach spaces.

**Remarks** Completeness is not a topological invariant. For example, the open interval $(-1, 1)$ and the whole real line are homeomorphic (with respect to the usual topologies) but the former is not complete while the latter is. The homeomorphism can conveniently be given in terms of the trigonometric function tangent.

**Definition 32** A subset $B$ of the metric space $X$ is *bounded* if there exists a ball of radius $R$ $(R > 0)$ which contains it entirely.

**Theorem 1** (Heine–Borel) *Any closed bounded subset of $\mathbb{R}^n$ is compact.*

**Remark** The converse is also true. We have thus a nice characterization of compact subsets of $\mathbb{R}^n$ as being closed and bounded.

**Proposition 6** *Any bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.*

**Definition 33** Consider a sequence $\{f_n\}$ of real-valued functions on a subset A (usually an interval) of $\mathbb{R}$. We say that $\{f_n\}$ *converges pointwise* in $A$ if the sequence of real numbers $\{f_n(x)\}$ converges for every $x \in A$. We can then define a function $f : A \to \mathbb{R}$ by $f(x) = \lim_{n\to\infty} f_n(x)$, and write $f_n \to f$.

**Definition 34** A sequence of functions $f_n : A \to \mathbb{R}, A \subseteq \mathbb{R}$ is said to *converge uniformly* to a function $f : A \to \mathbb{R}$ if given any $\epsilon > 0$, there exists a positive integer $N$ such that, for all $x$, $|f_n(x) - f(x)| < \epsilon$ whenever $n > N$.

**Theorem 2** *Let $f_n : (a, b) \to \mathbb{R}$ be a sequence of functions continuous at the point $c \in (a, b)$, and suppose $f_n$ converges uniformly to $f$ on $(a, b)$. Then $f$ is continuous at $c$.*

**Remark and Example** The pointwise limit of continuous functions need not be continuous, as can be shown by the following example: $f_n(x) = x^n, x \in [0, 1]$. We see that the limit function $f$ is not continuous:

$$f(x) = \begin{cases} 0 & x \neq 1 \\ 1 & x = 1 \end{cases}$$

**Definition 35** Let $X$ be a metric space. A map $f: X \to X$ is a *contraction* if there exists $c < 1$ such that $d(f(x), f(y)) \leq cd(x, y)$ for all $x, y \in X$.

**Theorem 3** (Banach) *If $X$ is a complete metric space and $f$ is a contraction in $X$, then $f$ has a unique fixed point $x \in X$, that is, $f(x) = x$.*

## Some Function and Operator Spaces

The spaces of functions and operators can be equipped with different topologies, given by various concepts of convergence and of norms (or sometimes seminorms), very often with different such concepts for the same space. As we saw earlier, a norm in a vector space gives rise to a metric, and hence to a topology. Similarly with the concept of convergence for sequences of functions and operators, as one then knows what the limit points, and hence closed sets, are.

But before we do that, let us introduce, in a slightly different context, a topology which is in some sense the natural one for the space of continuous maps from one space to another.

**Definition 36** Consider a family $F$ of maps from a topological space $X$ to a topological space $Y$, and define $W(K, U) = \{f : f \in F, f(K) \subseteq U\}$. Then the family of all sets of the form $W(K, U)$ with $K$ compact (in $X$) and $U$ open (in $Y$) form a sub-basis for the *compact open topology* for $F$.

Consider a topological space $X$ and sequences of functions $(f_n)$ on it. Let $D \subseteq X$. We can then define pointwise convergence and uniform convergence exactly as for functions on subsets of the real line.

**Definition 37** Let $X$, $D$ and $(f_n)$ as above.

(i) The functions $f_n$ *converge pointwise* on $D$ to a function $f$ if the sequence of numbers $f_n(x) \to f(x), \forall x \in D$.
(ii) The functions $f_n$ *converge uniformly* on $D$ to a function $f$ if given $\epsilon > 0$, there exists $N$ such that for all $n > N$ we have $|f_n(x) - f(x)| < \epsilon, \forall x \in D$.

Next we consider the Lebesgue spaces $L^p$, that is, functions $f$ defined on subsets of $\mathbb{R}^n$, such that $|f(x)|^p$ is Lebesgue integrable, for real numbers $p \geq 1$. To define these spaces, we tacitly take equivalence classes of functions which are equal almost everywhere (that is, up to a null set), but very often we can take representatives of these classes and just deal with genuine functions instead. Note that of all $L^p$, only $L^2$ is a Hilbert space.

**Definition 38** In the space $L^p$, we define its norm by

$$\|f\| = \left( \int |f(x)|^p \, dx \right)^{1/p}$$

Now we turn to general normed spaces, and operators on them.

**Definition 39** Convergence in the norm is also called strong convergence. In other words, a sequence $(x_n)$ in a normed space $X$ is said to *converge strongly* to $x$ if

$$\lim_{n \to \infty} \|x_n - x\| = 0$$

**Definition 40** A sequence $(x_n)$ in a normed space $X$ is said to *converge weakly* to $x$ if

$$\lim_{n \to \infty} f(x_n) = f(x)$$

for all bounded linear functionals $f$.

Consider the space $B(X, Y)$ of bounded linear operators $T$ from $X$ to $Y$. We can make this into a normed space by defining the following norm:

$$\|T\| = \sup_{x \in X, \|x\| = 1} \|Tx\|$$

Then we can define three different concepts of convergence on $B(X, Y)$. There are in fact more in current use in functional analysis.

**Definition 41** Let $X$ and $Y$ be normed spaces and let $(T_n)$ be a sequence of operators $T_n \in B(X, Y)$.

(i) $(T_n)$ is *uniformly convergent* if it converges in the norm.
(ii) $(T_n)$ is *strongly convergent* if $(T_n x)$ converges strongly for every $x \in X$.
(iii) $(T_n)$ is *weakly convergent* if $(T_n x)$ converges weakly for every $x \in X$.

**Remark** Clearly we have: uniform convergence $\Longrightarrow$ strong convergence $\Longrightarrow$ weak convergence, and the limits are the same in all three cases. However, the converses are in general not true.

## Homotopy Groups

The most elementary and obvious property of a topological space $X$ is the number of connected components it has. The next such property, in a certain sense, is the number of holes $X$ has. There

are higher analogues of these, called the *homotopy groups*, which are topological invariants, that is, they are invariant under homeomorphisms. They play important roles in many topological considerations in field theory and other topics of mathematical physics. The articles Topological Defects and Their Homotopy Classification and Electric-Magnetic Duality contain some examples.

**Definition 42** Given a topological space $X$, the zeroth homotopy set, denoted $\pi_0(X)$, is the set of connected components of $X$. One sometimes writes $\pi_0(X) = 0$ if $X$ is connected.

To define the fundamental group of $X$, or $\pi_1(X)$, we shall need the concept of closed loops, which we shall find useful in other ways too. For simplicity, we shall consider based loops (that is, loops passing through a fixed point in $X$). It seems that in most applications, these are the relevant ones. One could consider loops of various smoothness (when $X$ is a manifold), but in view of applications to quantum field theory, we shall consider continuous loops, which are also the ones relevant for topology.

**Definition 43** Given a topological space $X$ and a point $x_0 \in X$, a (*closed*) (*based*) *loop* is a continuous function of the parametrized circle to $X$:

$$\xi : [0, 2\pi] \to X$$

satisfying $\xi(0) = \xi(2\pi) = x_0$.

**Definition 44** Given a connected topological space $X$ and a point $x_0 \in X$, the space of all closed based loops is called the (parametrized based) *loop space* of $X$, denoted $\Omega X$.

**Remarks**

(i) The loop space $\Omega X$ inherits the relative compact–open topology from the space of continuous maps from the closed interval $[0, 2\pi]$ to $X$. It also has a natural base point: the constant function mapping all of $[0, 2\pi]$ to $x_0$. Hence it is easy to iterate the construction and define $\Omega^k X, k \geq 1$.

(ii) Here we have chosen to parametrize the circle by $[0, 2\pi]$, as is more natural if we think in terms of the phase angle. We could easily have chosen the unit interval [0,1] instead. This would perhaps harmonize better with our previous definition of paths and the definitions of homotopies below.

**Proposition 7** *The fundamental group of a topological space $X$, denoted $\pi_1(X)$, consists of classes of closed loops in $X$ which cannot be continuously deformed into one another while preserving the base point.*

**Definition 45** A space $X$ is called *simply connected* if $\pi_1(X)$ is trivial.

To define the higher homotopy groups, let us go into a little detail about homotopy.

**Definition 46** Given two topological spaces $X$ and $Y$, and maps

$$p, q : X \to Y$$

we say that $h$ is a *homotopy* between the maps $p, q$ if

$$h : X \times I \to Y$$

is a continuous map such that $h(x, 0) = p(x)$, $h(x, 1) = q(x)$, where $I$ is the unit interval $[0, 1]$. In this case, we write $p \simeq q$.

**Definition 47** A map $f : X \to Y$ is a *homotopy equivalence* if there exists a map $g : Y \to X$ such that $g \circ f \simeq \mathrm{id}_X$ and $f \circ g \simeq \mathrm{id}_Y$.

**Remark** This is an equivalence relation.

**Definition 48** For a topological space $X$ with base point $x_0$, we define $\pi_n(X), n \geq 0$ as the set of homotopy equivalence classes of based maps from the $n$-sphere $S^n$ to $X$.

**Remark** This coincides with the previous definitions for $\pi_0$ and $\pi_1$.

There is a very nice relation between homotopy classes and loop spaces.

**Proposition 8** $\pi_n(X) = \pi_{n-1}(\Omega X) = \cdots = \pi_0(\Omega^n X)$.

**Remarks**

(i) When we consider the gauge group $G$ in a Yang–Mills theory, its fundamental group classifies the monopoles that can occur in the theory.

(ii) For $n \geq 1, \pi_n(X)$ is a group, the group action coming from the joining of two loops together to form a new loop. On the other hand, $\pi_0(X)$ in general is not a group. However, when $X$ is a Lie group, then $\pi_0(X)$ inherits a group structure from $X$, because it can be identified with the quotient group of $X$ by its identity-connected component. For example, the two components of O(3) can be identified with the two elements of the group $\mathbb{Z}_2$, the component where the determinant equals 1 corresponding to 0 in $\mathbb{Z}_2$ and the component where the determinant equals $-1$ corresponding to 1 in $\mathbb{Z}_2$.

(iii) For $n \geq 2$, the group $\pi_n(X)$ is always abelian.

(iv) Examples of nonabelian $\pi_1$ are the fundamental groups of some Riemann surfaces.

(v) Since $\pi_1$ is not necessarily abelian, much of the direct-sum notation we use for the homotopy

groups should more correctly be written multiplicatively. However, in most literature in mathematical physics, the additive notation seems to be preferred.

**Examples**

(i) $\pi_n(X \times Y) = \pi_n(X) + \pi_n(Y)$, $n \geq 1$.
(ii) For the spheres, we have the following results:

$$\pi_i(S^n) = \begin{cases} 0 & \text{if } i > n \\ \mathbb{Z} & \text{if } i = n \end{cases}$$
$$\pi_i(S^1) = 0 \qquad \text{if } i > 1$$
$$\pi_{n+1}(S^n) = \mathbb{Z}_2 \qquad \text{if } n \geq 3$$
$$\pi_{n+2}(S^n) = \mathbb{Z}_2 \qquad \text{if } n \geq 2$$
$$\pi_6(S^3) = \mathbb{Z}_{12}$$

(iii) From the theory of sphere bundles, we can deduce:

$$\pi_i(S^2) = \pi_{i-1}(S^1) + \pi_i(S^3) \quad \text{if } i \geq 2$$
$$\pi_i(S^4) = \pi_{i-1}(S^3) + \pi_i(S^7) \quad \text{if } i \geq 2$$
$$\pi_i(S^8) = \pi_{i-1}(S^7) + \pi_i(S^{15}) \quad \text{if } i \geq 2$$

and the first of these relations give the following more succinct result:

$$\pi_i(S^3) = \pi_i(S^2) \quad \text{if } i \geq 3$$

(iv) A result of Serre says that all the homotopy groups of spheres are in fact finite except $\pi_n(S^n)$ and $\pi_{4n-1}(S^{2n}), n \geq 1$.

**Definition 49** Given a connected space $X$, a map $\pi : B \to X$ is called a *covering* if (i) $\pi(B) = X$, and (ii) for each $x \in X$, there exists an open connected neighborhood $V$ of $x$ such that each component of $\pi^{-1}(V)$ is open in $B$, and $\pi$ restricted to each component is a homeomorphism. The space $B$ is called a *covering space*.

**Examples**

(i) The real line $\mathbb{R}$ is a covering of the group U(1).
(ii) The group SU(2) is a double cover of the group SO(3).
(iii) The group SL(2, $\mathbb{C}$) is a double cover of the Lorentz group SO(1, 3).
(iv) The group SU(2, 2) is a 4-fold cover of the conformal group in four dimensions. This local isomorphism is of great importance in twistor theory.

**Remarks**

(i) By considering closed loops in $X$ and their coverings in $B$ it is easily seen that the fundamental group $\pi_1(X)$ acts on the coverings of $X$. If we further assume that the action is transitive, then we have the following nice result: coverings of $X$ are in 1–1 correspondence with normal subgroups of $\pi_1(X)$.
(ii) Given a connected space $X$, there always exists a unique connected simply connected covering space $\widetilde{X}$, called the universal covering space. Furthermore, $\widetilde{X}$ covers all the other covering spaces of $X$. For the higher homotopy groups, one has

$$\pi_n(X) = \pi_n(\widetilde{X}), \quad n \geq 2$$

One very important class of homotopy groups are those of Lie groups. To simplify matters, we shall consider only connected groups, that is, $\pi_0(G) = 0$. Also we shall deal mainly with the classical groups, and in particular, the orthogonal and unitary groups.

**Proposition 9** *Suppose that G is a connected Lie group.*

(i) *If G is compact and semi-simple, then* $\pi_1(G)$ *is finite. This implies that* $\widetilde{G}$ *is still compact.*
(ii) $\pi_2(G) = 0$.
(iii) *For G compact, simple, and nonabelian,* $\pi_3(G) = \mathbb{Z}$.
(iv) *For G compact, simply connected, and simple,* $\pi_4(G) = 0$ or $\mathbb{Z}_2$.

**Examples**

(i) $\pi_1(\mathrm{SU}(n)) = 0$.
(ii) $\pi_1(\mathrm{SO}(n)) = \mathbb{Z}_2$.
(iii) Since the unitary groups U($n$) are topologically the product of SU($n$) with a circle $S^1$, their homotopy groups are easily computed using the product formula. We remind ourselves that U(1) is topologically a circle and SU(2) topologically $S^3$.
(iv) For $i \geq 2$, we have:

$$\pi_i(\mathrm{SO}(3)) = \pi_i(\mathrm{SU}(2))$$
$$\pi_i(\mathrm{SO}(5)) = \pi_i(\mathrm{Sp}(2))$$
$$\pi_i(\mathrm{SO}(6)) = \pi_i(\mathrm{SU}(4))$$

Just for interest, and to show the richness of the subject, some isomorphisms for homotopy groups are shown in Table 1 and some homotopy groups for low SU($n$) and SO($n$) are listed in Table 2.

**Table 1** Some isomorphisms for homotopy groups

| *Isomorphism* | *Range* |
|---|---|
| $\pi_i(\mathrm{SO}(n)) \cong \pi_i(\mathrm{SO}(m))$ | $n, m \geq i+2$ |
| $\pi_i(\mathrm{SU}(n)) \cong \pi_i(\mathrm{SU}(m))$ | $n, m \geq \frac{1}{2}(i+1)$ |
| $\pi_i(\mathrm{Sp}(n)) \cong \pi_i(\mathrm{Sp}(m))$ | $n, m \geq \frac{1}{4}(i-1)$ |
| $\pi_i(G_2) \cong \pi_i(\mathrm{SO}(7))$ | $2 \leq i \leq 5$ |
| $\pi_i(F_4) \cong \pi_i(\mathrm{SO}(9))$ | $2 \leq i \leq 6$ |
| $\pi_i(\mathrm{SO}(9)) \cong \pi_i(\mathrm{SO}(7))$ | $i \leq 13$ |

**Table 2** Some homotopy groups for low SU($n$) and SO($n$)

| | $\pi_4$ | $\pi_5$ | $\pi_6$ | $\pi_7$ | $\pi_8$ | $\pi_9$ | $\pi_{10}$ |
|---|---|---|---|---|---|---|---|
| SU(2) | $\mathbb{Z}_2$ | $\mathbb{Z}_2$ | $\mathbb{Z}_{12}$ | $\mathbb{Z}_2$ | $\mathbb{Z}_2$ | $\mathbb{Z}_3$ | $\mathbb{Z}_{15}$ |
| SU(3) | 0 | $\mathbb{Z}$ | $\mathbb{Z}_6$ | 0 | $\mathbb{Z}_{12}$ | $\mathbb{Z}_3$ | $\mathbb{Z}_{30}$ |
| SU(4) | 0 | $\mathbb{Z}$ | 0 | $\mathbb{Z}$ | $\mathbb{Z}_{24}$ | $\mathbb{Z}_2$ | $\mathbb{Z}_{120}+\mathbb{Z}_2$ |
| SU(5) | 0 | $\mathbb{Z}$ | 0 | $\mathbb{Z}$ | 0 | $\mathbb{Z}$ | $\mathbb{Z}_{120}$ |
| SU(6) | 0 | $\mathbb{Z}$ | 0 | $\mathbb{Z}$ | 0 | $\mathbb{Z}$ | $\mathbb{Z}_3$ |
| SO(5) | $\mathbb{Z}_2$ | $\mathbb{Z}_2$ | 0 | $\mathbb{Z}$ | 0 | 0 | $\mathbb{Z}_{120}$ |
| SO(6) | 0 | $\mathbb{Z}$ | 0 | $\mathbb{Z}$ | $\mathbb{Z}_{24}$ | $\mathbb{Z}_2$ | $\mathbb{Z}_{120}+\mathbb{Z}_2$ |
| SO(7) | 0 | 0 | 0 | $\mathbb{Z}$ | $\mathbb{Z}_2+\mathbb{Z}_2$ | $\mathbb{Z}_2+\mathbb{Z}_2$ | $\mathbb{Z}_{24}$ |
| SO(8) | 0 | 0 | 0 | $\mathbb{Z}+\mathbb{Z}$ | $\mathbb{Z}_2+\mathbb{Z}_2+\mathbb{Z}_2$ | $\mathbb{Z}_2+\mathbb{Z}_2+\mathbb{Z}_2$ | $\mathbb{Z}_{24}+\mathbb{Z}_{24}$ |
| SO(9) | 0 | 0 | 0 | $\mathbb{Z}$ | $\mathbb{Z}_2+\mathbb{Z}_2$ | $\mathbb{Z}_2+\mathbb{Z}_2$ | $\mathbb{Z}_{24}$ |
| SO(10) | 0 | 0 | 0 | $\mathbb{Z}$ | $\mathbb{Z}_2$ | $\mathbb{Z}+\mathbb{Z}_2$ | $\mathbb{Z}_{12}$ |

## Appendix: A Mathematician's Basic Toolkit

The following is a drastically condensed list, most of which is what a mathematics undergraduate learns in the first few weeks. The rest is included for easy reference. These notations and concepts are used universally in mathematical writing. We have not endeavored to arrange the material in a logical order. Furthermore, given structures such as sets, groups, etc., one can usually define "substructures" such as subsets, subgroups, etc., in a straightforward manner. We shall therefore not spell this out.

### Sets

| | |
|---|---|
| $A \cup B = \{x : x \in A \text{ or } x \in B\}$ | union |
| $A \cap B = \{x : x \in A \text{ and } x \in B\}$ | intersection |
| $A \backslash B = \{x : x \in A \text{ and } x \notin B\}$ | complement |
| $A \times B = \{(x, y) : x \in A, y \in B\}$ | Cartesian product |

### Maps

1. A map or mapping $f : A \to B$ is an assignment of an element $f(x)$ of $B$ for every $x \in A$.
2. A map $f : A \to B$ is injective if $f(x) = f(y) \Longrightarrow x = y$. This is sometimes called a 1–1 map, a term to be avoided.
3. A map $f : A \to B$ is surjective if for every $y \in B$ there exists an $x \in A$ such that $y = f(x)$. This is sometimes called an "onto" map.
4. A map $f : A \to B$ is bijective if it is both surjective and injective. This is also sometimes called a 1–1 map, a term to be equally avoided.
5. For any map $f : A \to B$ and any subset $C \subseteq B$, the inverse image $f^{-1}(C) = \{x : f(x) \in C\} \subseteq A$ is always defined, although, of course, it can be empty. On the other hand, the map $f^{-1}$ is defined if and only if $f$ is bijective.
6. A map from a set to either the real or complex numbers is usually called a function.
7. A map between vector spaces, and more particularly normed spaces (including Hilbert spaces), is called an operator. Most often, one considers linear operators.
8. An operator from a vector space to its field of scalars is called a functional. Again, one considers almost exclusively linear functionals.

### Relations

1. A relation $\sim$ on a set $A$ is a subset $R \subseteq A \times A$. We say that $x \sim y$ if $(x, y) \in R$.
2. We shall only be interested in equivalence relations. An equivalence relation $\sim$ is one satisfying, for all $x, y, z \in A$:
   (a) $x \sim x$ ("reflexive"),
   (b) $x \sim y \Longrightarrow y \sim x$ ("symmetric"),
   (c) $x \sim y, y \sim z \Longrightarrow x \sim z$ ("transitive").
3. If $\sim$ is an equivalence relation in $A$, then for each $x \in A$, we can define its equivalence class:

   $$[x] = \{y \in A : y \sim x\}$$

   It can be shown that equivalence classes are nonempty, any two equivalence classes are either equal or disjoint, and they together partition the set $A$. Subgroup equivalence classes are called cosets.
4. An element of an equivalence class is called a representative.

### Groups

A group is a set G with a map, called multiplication or group law

$$\begin{aligned} G \times G &\longrightarrow G \\ (x, y) &\longmapsto xy \end{aligned}$$

satisfying

1. $(xy)z = x(yz), \forall x, y, z \in G$ ("associative");
2. there exists a neutral element (or identity) 1 such that $1x = x1 = x, \forall x \in G$; and
3. every element $x \in G$ has an inverse $x^{-1}$, that is, $xx^{-1} = x^{-1}x = 1$.

A map such as the multiplication in the definition is an example of a binary operation. Note that we have denoted the group law as multiplication here. It is usual to denote it additively if the group is abelian, that is, if $xy = yx, \forall x, y \in G$. In this case, we may write the condition as $x + y = y + x$, and call the identity element 0.

### Rings

A ring is a set $R$ equipped with two binary operations, $x + y$ called addition, and $xy$ called multiplication, such that

1. $R$ is an abelian group under addition;
2. the multiplication is associative; and
3. $(x + y)z = xz + yz, x(y + z) = xy + xz, \forall x, y, z \in R$ ("distributive").

If the multiplication is commutative ($xy = yx$) then the ring is said to be commutative. A ring may contain a multiplicative identity, in which case it is called a ring with unit element.

An ideal $I$ of $R$ is a subring of $R$, satisfying in addition

$$r \in R, a \in I \implies ra \in I, ar \in I$$

One can define in an obvious fashion a left-ideal and a right- ideal. The above definition will then be for a two-sided ideal.

### Modules

Given a ring $R$, an $R$-module is an abelian group $M$, together with an operation, $M \times R \to M$, denoted multiplicatively, satisfying, for $x, y \in M, r, s \in R$,

1. $(x + y)r = xr + yr$,
2. $x(r + s) = xr + xs$,
3. $x(rs) = (xr)s$, and
4. $x1 = x$

The term right $R$-module is sometimes used, to distinguish it from obviously defined left $R$-modules.

### Fields

A field $F$ is a commutative ring in which every nonzero element is invertible.

The additive identity 0 is never invertible, unless $0 = 1$, so it is usual to assume that a field has at least two elements, 0 and 1.

The most common fields we come across are, of course, the number fields: the rationals, the reals, and the complex numbers.

### Vector Spaces

A vector space, or sometimes linear space, $V$, over a field $F$, is an abelian group, written additively, with a map $F \times V \to V$ such that, for $x, y \in V, \alpha, \beta \in F$,

1. $\alpha(x + y) = \alpha x + \alpha y$ ("linearity"),
2. $(\alpha + \beta)x = \alpha x + \beta x$,
3. $(\alpha\beta)x = \alpha(\beta x)$, and
4. $1x = x$.

A vector space is then a right (or left) $F$-module. The elements of $V$ are called vectors, and those of $F$ scalars.

### Algebras

An algebra $A$ over a field $F$ is a ring which is a vector space over $F$, such that

$$\alpha(ab) = (\alpha a)b = a(\alpha b), \quad \alpha \in F, \ a, b \in A$$

Note that in some older literature, particularly the Russian school, an algebra of operators is called a ring of operators.

## Further Reading


Borel A (1955) Topology of Lie groups and characteristic classes. *Bulletin American Mathematical Society* 61: 397–432.

Kelly JL (1955) *General Topology*. New York: Van Nostrand Reinhold.

Kreyszig E (1978) *Introductory Functional Analysis with Applications*. New York: Wiley.

Mc Carty G (1967) *Topology: An Introduction with Application to Topological Groups*. New York: McGraw-Hill.

Simmons GF (1963) *Introduction to Topology and Modern Analysis*. New York: McGraw-Hill.

# A

## Abelian and Nonabelian Gauge Theories Using Differential Forms

**A C Hirshfeld**, Universität Dortmund, Dortmund, Germany



### Introduction

Quantum electrodynamics is the theory of the electromagnetic interactions of photons and electrons. When attempting to generalize this theory to other interactions it turns out to be necessary to identify its essential components. The essential properties of electrodynamics are contained in its formulation as an "abelian gauge theory." The generalization to include other interactions is then reduced to incorporating the structure of nonabelian groups. This becomes particularly clear when we formulate the theory in the language of differential forms.

Here we first present the formulation of electrodynamics using differential forms. The electromagnetic fields are introduced via the Lorentz force equation. They are recognized as the components of a differential 2-form. This form fulfills two differential conditions, which are equivalent to Maxwell's equations. These are expressed with the help of a differential operator and its Hermitian conjugate, the codifferential operator. We consider the effects of charge conservation and introduce electromagnetic potentials, which are defined up to gauge transformations. We finally consider Weyl's argument for the existence of the electromagnetic interaction as a consequence of the local phase invariance of the electron wave function.

We then go on to present the nonabelian generalization. The gauge bosons appear in a theory with fermions by requiring invariance of the theory with respect to local gauge transformations. When the fermions group into symmetry multiplets this gives rise to a gauge group SU($N$) involving $N^2-1$ gauge bosons mediating the interaction, where $N$ is the dimension of the Lie algebra. The interaction arises through the necessity of replacing the usual derivatives by covariant derivatives, which transform in a natural way in order to preserve the gauge invariance. The covariant derivatives involve the gauge potentials, whose transformation properties are dictated by those of the covariant derivative. Whereas for an abelian gauge theory such as electromagnetism scalar-valued $p$-forms are sufficient (actually only $p=1,2$), a nonabelian gauge theory involves the use of Lie-algebra-valued $p$-forms. These are introduced and used to construct the Yang–Mills action, which involves the field strength tensor which is determined from the gauge potentials. This action leads to the Yang–Mills equations for the gauge potentials, which are the nonabelian generalizations of the Maxwell equations.

### Relativistic Kinematics

The trajectory of a mass point is described as $x^\mu(\tau)$, where $\tau$ is the invariant proper time interval:

$$d\tau^2 = dt^2 - d\boldsymbol{x}\cdot d\boldsymbol{x} = dt^2(1-\boldsymbol{v}^2) \qquad [1]$$

with $\boldsymbol{v} = d\boldsymbol{x}/dt$. With the abbreviation $\gamma = (1-\boldsymbol{v}^2)^{-1/2}$ this yields $d\tau = (1/\gamma)dt$.

The 4-velocity of a point is defined as $u^\mu = dx^\mu/d\tau = \gamma(dx^\mu/dt)$. The quantity

$$u^2 = g_{\mu\nu}u^\mu u^\nu = \frac{dx^\mu dx_\mu}{d\tau^2} = 1 \qquad [2]$$

is a relativistic invariant. Here

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \qquad [3]$$

is the metric of Minkowski space.

The 4-momentum of a particle is $p^\mu = m_0 u^\mu = (m_0\gamma, m_0\gamma\boldsymbol{v})$, and $p^\mu p_\mu = m_0^2$. The 4-force is

$$f^\mu = \frac{dp^\mu}{d\tau} = \gamma\frac{dp^\mu}{dt} = \gamma\left(\frac{dp^0}{dt}, \boldsymbol{f}\right) \qquad [4]$$

with the 3-force

$$\boldsymbol{f} = \frac{d(m_0\gamma\boldsymbol{v})}{dt} \qquad [5]$$

Differentiate $p^2 = m_0^2$ with respect to $\tau$, this yields

$$2p^{\mu}f_{\mu} = 2m_0\gamma^2\left(\frac{dp^0}{dt} - \boldsymbol{f}\cdot\boldsymbol{v}\right) = 0 \qquad [6]$$

or

$$\frac{dp^0}{dt} = \boldsymbol{f}\cdot\boldsymbol{v} = \boldsymbol{f}\cdot\frac{d\boldsymbol{x}}{dt} \qquad [7]$$

This says that

$$dp^0 = \boldsymbol{f}\cdot d\boldsymbol{x} = dW \qquad [8]$$

where $W$ is the work done and $p^0$ is the energy.

For a charged particle, the Lorentz force is

$$\boldsymbol{f} = q(\boldsymbol{E} + \boldsymbol{v}\times\boldsymbol{B}) \qquad [9]$$

where $q$ is the charge of the particle, $\boldsymbol{E}$ is the electric, and $\boldsymbol{B}$ the magnetic field strength. Since $\boldsymbol{f}\cdot\boldsymbol{v} = q\boldsymbol{E}\cdot\boldsymbol{v}$, we have the four-dimensional form of the Lorentz force:

$$f^{\mu} = q\gamma(\boldsymbol{E}\cdot\boldsymbol{v}, \boldsymbol{E} + \boldsymbol{v}\times\boldsymbol{B}) \qquad [10]$$

## The Lorentz Force Equation with Differential Forms

We write the Lorentz force equation as an equation for a differential form $f = f_{\mu}dx^{\mu}$, with $f_{\mu} = g_{\mu\nu}f^{\nu}$. The velocity-dependent Lorentz force is

$$f = -qi_uF \qquad [11]$$

with

$$u = \gamma\left(\frac{\partial}{\partial t} + v^x\frac{\partial}{\partial x} + v^y\frac{\partial}{\partial y} + v^z\frac{\partial}{\partial z}\right) \qquad [12]$$

the 4-velocity and $F$ the electromagnetic field strength:

$$F = \mathcal{E}\wedge dt + \mathcal{B} \qquad [13]$$

where $\mathcal{E}$ is a 1-form in three dimensions,

$$\mathcal{E} = E_x dx + E_y dy + E_z dz \qquad [14]$$

and $\mathcal{B}$ is a 2-form in three dimensions,

$$\mathcal{B} = B_x dy\wedge dz + B_y dz\wedge dx + B_z dx\wedge dy \qquad [15]$$

The symbol $i_u$ indicates a contraction of a 2-form with a vector, which is defined as

$$i_uF(v) = F(u, v) \qquad [16]$$

for an arbitrary vector $v$. The contraction of a 2-form with a vector yields a 1-form.

It is easily seen that a 2-form can be expressed in terms of a polar vector and an axial vector: if it is to be invariant with respect to parity transformations with

$$t\to t,\quad x\to -x,\quad y\to -y,\quad z\to -z \qquad [17]$$

the fields in eqn [13] must transform as

$$\boldsymbol{E}\to -\boldsymbol{E},\quad \boldsymbol{B}\to\boldsymbol{B} \qquad [18]$$

Now we check the validity of eqn [11]. We have

$$\begin{aligned} f &= -qi_uF \\ &= q\gamma(\boldsymbol{v}\cdot\boldsymbol{E})dt - q\gamma[(E^x + (\boldsymbol{v}\times\boldsymbol{B})^x)dx \\ &\quad + (E^y + (\boldsymbol{v}\times\boldsymbol{B})^y)dy + (E^z + (\boldsymbol{v}\times\boldsymbol{B})^z)dz] \end{aligned} \qquad [19]$$

in agreement with eqn [10]. We remember to change the signs in $E_x = -E^x, B_x = -B^x$, etc.

## The Codifferential Operator

The space of $p$-forms on an $n$-dimensional manifold is an

$$\binom{n}{p} = \binom{n}{n-p} = \frac{n!}{(n-p)!p!} \qquad [20]$$

dimensional vector space. The space of $p$-forms is thus isomorphic to the space of $(n-p)$-forms. The Hodge dual operator maps the $p$-forms into the $(n-p)$-forms, and is defined by

$$\alpha\wedge *\beta = \langle\alpha,\beta\rangle dx^1\wedge\cdots\wedge dx^n \qquad [21]$$

Here $\langle\alpha,\beta\rangle$ is the scalar product of two $p$-forms:

$$\langle\alpha,\beta\rangle = \alpha_{i_1\cdots i_p}\beta^{i_1\cdots si_p} \qquad [22]$$

where $\alpha_{i_1\cdots si_p}$ are the coefficients of the form $\alpha$,

$$\alpha = \alpha_{i_1\cdots i_p}dx^{i_1}\wedge\cdots\wedge dx^{i_p} \qquad [23]$$

$\beta_{j_1\cdots sj_p}$ are the coefficients of the form $\beta$,

$$\beta = \beta_{j_1\cdots j_p}dx^{j_1}\wedge\cdots\wedge dx^{j_p} \qquad [24]$$

and

$$\beta^{i_1\cdots i_p} = g^{i_1j_1}\cdots g^{i_pj_p}\beta_{j_1\cdots j_p} \qquad [25]$$

The indices satisfy $i_1 < \cdots < i_p$ and $j_1 < \cdots < j_p$.

The basis elements are orthogonal with respect to this scalar product, and

$$\begin{aligned} &\langle dx^{i_1}\wedge\cdots\wedge dx^{i_p}, dx^{i_1}\wedge\cdots\wedge dx^{i_p}\rangle \\ &= g_{i_1i_1}\cdots g_{i_pi_p} \end{aligned} \qquad [26]$$

The Hodge dual has the property that

$$* \left( dx^{\sigma(1)} \wedge \cdots \wedge dx^{\sigma(p)} \right) = g_{\sigma(1)\sigma(1)} \cdots g_{\sigma(p)\sigma(p)} (\text{sign}\, \sigma) \times \left( dx^{\sigma(p+1)} \wedge \cdots \wedge dx^{\sigma(n)} \right) \quad [27]$$

where $\sigma$ is a permutation of the indices $(1, \ldots, n)$, $\sigma(1) < \cdots < \sigma(p)$, and $\sigma(p+1) < \cdots < \sigma(n)$. We also have

$$* \left( dx^{\sigma(p+1)} \wedge \cdots \wedge dx^{\sigma(n)} \right) = g_{\sigma(p+1)\sigma(p+1)} \cdots g_{\sigma(n)\sigma(n)} (-1)^{p(n-p)} (\text{sign}\, \sigma) \times \left( dx^{\sigma(1)} \wedge \cdots \wedge dx^{\sigma(p)} \right) \quad [28]$$

We therefore find that the application of the Hodge dual to a $p$-form twice yields

$$\begin{aligned} **\left( dx^{\sigma(1)} \wedge \cdots \wedge dx^{\sigma(p)} \right) &= g_{\sigma(1)\sigma(1)} \cdots g_{\sigma(p)\sigma(p)} (\text{sign}\, \sigma) * \left( dx^{\sigma(p+1)} \wedge \cdots \wedge dx^{\sigma(n)} \right) \\ &= g_{\sigma(1)\sigma(1)} \cdots g_{\sigma(n)\sigma(n)} (-1)^{p(n-p)} dx^{\sigma(1)} \wedge \cdots \wedge dx^{\sigma(p)} \end{aligned} \quad [29]$$

or

$$** = (-1)^{p(n-p)} (-1)^{\text{Ind}\, g} \text{Id} \quad [30]$$

where $\text{Ind}\, g$ is the number of times $(-1)$ occurs along the diagonal of $g$.

Now let $\alpha$ be a $(p-1)$-form, and $\beta$ a $p$-form. Then $d * \beta$ is an $(n-p+1)$-form, and

$$\begin{aligned} d(\alpha \wedge * \beta) &= d\alpha \wedge * \beta + (-1)^{p-1} \alpha \wedge d * \beta \\ &= d\alpha \wedge * \beta + (-1)^{(p-1)} (-1)^{(n-p+1)(p-1)} \times (-1)^{\text{Ind} g} \alpha \wedge (**) d * \beta \\ &= d\alpha \wedge * \beta + (-1)^{n(p-1)} (-1)^{\text{Ind} g} \times \alpha \wedge * (* d * \beta) \end{aligned} \quad [31]$$

We then have

$$(d\alpha, \beta) - (\alpha, d^* \beta) = \int_M d(\alpha \wedge * \beta) \quad [32]$$

with

$$d^* = -(-1)^{n(p-1)} (-1)^{\text{Ind}\, g} * d * \quad [33]$$

We are here using the scalar product of two $p$-forms

$$(\alpha, \beta) := \int_M (\alpha \wedge * \beta) \quad [34]$$

With the help of Stokes' theorem the last integral in eqn [32] may be turned into a surface term at infinity, which vanishes for $\alpha$ and $\beta$ with compact support. $d^*$ is the adjoint operator to $d$ with respect to the scalar product $( \, , \, )$. Whereas the differential operator $d$ maps $p$-forms into $(p+1)$-forms, the codifferential operator $d^*$ maps $p$-forms into $(p-1)$-forms.

The relation $d^2 = 0$ leads to

$$(d^*)^2 \propto (*d*)(*d*) \propto *d^2* = 0 \quad [35]$$

This fact plays an essential role in connection with the conservation laws.

Finally, we want to obtain a coordinate expression for $d^* \beta$. Indeed $d^* \beta = -\text{Div}\, \beta$ for

$$(\text{Div} \beta)_K = \frac{\partial \beta^j_K}{\partial x^j} \quad [36]$$

where $K$ is the multi-index of the coeffecients in $\beta = \beta_{\underline{K}} dx^K$, and $\underline{K}$ indicates that $K = (k_1, \ldots, k_p)$ is in the order $k_1 < \cdots < k_p$. We will show that $(\alpha, d^* \beta) = (\alpha, -\text{Div} \beta)$ for an arbitrary $(p-1)$-form $\alpha$. It is a fact that

$$(\alpha, d^* \beta) = (d\alpha, \beta) = \int (d\alpha)_{\underline{I}} \beta^I * 1 \quad [37]$$

Now we have the coordinate expressions

$$d\alpha = (d\alpha_{\underline{L}}) \wedge dx^L \quad [38]$$

and $(dx^L)_K = \delta^L_K$. It follows that

$$(d\alpha)_{\underline{I}} = (d\alpha_{\underline{L}} \wedge dx^L)_{\underline{I}} = \delta^{jK}_{\underline{I}} \frac{\partial \alpha_{\underline{L}}}{\partial x^j} \delta^L_K \quad [39]$$

or

$$(d\alpha)_{\underline{I}} = \delta^{j\underline{K}}_{\underline{I}} \frac{\partial \alpha_K}{\partial x^j} \quad [40]$$

Here we use

$$(\alpha \wedge \beta)_I = \delta^{KL}_I \alpha_K \beta_L \quad [41]$$

where

$$\delta^{KL}_I = \begin{cases} 1 & \text{if (KL) is an even permutation of I} \\ -1 & \text{if (KL) is an odd permutation of I} \\ 0 & \text{otherwise} \end{cases} \quad [42]$$

Use of the Leibnitz rule yields

$$\begin{aligned} \int (d\alpha)_{\underline{I}} \beta^I * 1 &= \int \delta^{j\underline{K}}_{\underline{I}} \frac{\partial \alpha_K}{\partial x^j} \beta^I * 1 \\ &= \int \frac{\partial (\delta^{j\underline{K}}_{\underline{I}} \alpha_K \beta^I)}{\partial x^j} * 1 \\ &\quad - \int \alpha_{\underline{K}} \delta^{jK}_{\underline{I}} \frac{\partial \beta^I}{\partial x^j} * 1 \end{aligned} \quad [43]$$

The first term corresponds to a surface integration and we can neglect it. We then have $\delta^{jK}_{\underline{I}}\beta^I = \beta^{jK}$ from the antisymmetry of $\beta$, so that

$$(\alpha, d^*\beta) = -\int \alpha_{\underline{K}} \frac{\partial \beta^{jK}}{\partial x^j} * 1 = (\alpha, -\mathrm{Div}\beta) \qquad [44]$$

## The Maxwell Equations

The Maxwell equations become remarkably concise when expressed in terms of differential forms, namely

$$dF = 0, \qquad d^*F = -j \qquad [45]$$

where $F$ is the field strength and $j$ is the current density. We wish to demonstrate this. We use a $(3+1)$-separation of the exterior derivative into a timelike and a spacelike part:

$$d = \boldsymbol{d} + dt \wedge \frac{\partial}{\partial t} \qquad [46]$$

We then get

$$dF = \left(\boldsymbol{d}\mathcal{E} + \frac{\partial \mathcal{B}}{\partial t}\right) \wedge dt + \boldsymbol{d}\mathcal{B} = 0 \qquad [47]$$

By comparing coefficients, we arrive at

$$\boldsymbol{d}\mathcal{E} = -\frac{\partial \mathcal{B}}{dt}, \qquad \boldsymbol{d}\mathcal{B} = 0 \qquad [48]$$

In vector notation

$$\mathrm{curl}\,\boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t}, \qquad \mathrm{div}\,\boldsymbol{B} = 0 \qquad [49]$$

the usual form of the homogeneous Maxwell equations.

By direct application of the formula [27], one finds

$$*F = -\star\mathcal{B} \wedge dt + \star\mathcal{E} \qquad [50]$$

where $\star$ means the Hodge dual in three space dimensions. One finds

$$d * F = \boldsymbol{d} \star \mathcal{E} - \left(\boldsymbol{d} \star \mathcal{B} - \frac{\partial \star \mathcal{E}}{\partial t}\right) \wedge dt \qquad [51]$$

Therefore,

$$\begin{aligned} d * F = &-(\mathrm{div}\,\boldsymbol{E}) dx \wedge dy \wedge dz \\ &+ \left((\mathrm{curl}\,\boldsymbol{B})^x - \frac{\partial E^x}{dt}\right) dy \wedge dz \wedge dt \\ &+ \left((\mathrm{curl}\,\boldsymbol{B})^y - \frac{\partial E^y}{dt}\right) dz \wedge dx \wedge dt \\ &+ \left((\mathrm{curl}\,\boldsymbol{B})^z - \frac{\partial E^z}{dt}\right) dx \wedge dy \wedge dt \end{aligned} \qquad [52]$$

We apply again the Hodge dual:

$$\begin{aligned} *d * F = &-(\mathrm{div}\,\boldsymbol{E}) dt + \left((\mathrm{curl}\,\boldsymbol{B})^x - \frac{\partial E^x}{\partial t}\right) dx \\ &+ \left((\mathrm{curl}\,\boldsymbol{B})^y - \frac{\partial E^y}{\partial t}\right) dy \\ &+ \left((\mathrm{curl}\,\boldsymbol{B})^z - \frac{\partial E^z}{\partial t}\right) dz \end{aligned} \qquad [53]$$

In Minkowski space the expression $*d*$ equals the codifferential. Therefore, the equation $d^*F = *d*F = -j$ holds, with $j$ given by $j^\mu = (\rho, \boldsymbol{J})$, which is equivalent to

$$\mathrm{div}\,\boldsymbol{E} = \rho, \qquad \mathrm{curl}\,\boldsymbol{B} - \frac{\partial \boldsymbol{E}}{\partial \boldsymbol{t}} = \boldsymbol{J} \qquad [54]$$

the inhomogeneous Maxwell equations.

## Current Conservation

The electromagnetic 4-current is

$$j^\mu = \rho_0 \boldsymbol{u}^\mu = (\rho_0\gamma, \rho_0\gamma\boldsymbol{v}) = (\rho, \boldsymbol{J}) \qquad [55]$$

where $\rho$ is the charge density and $\boldsymbol{J}$ the current density. This corresponds to a 1-form

$$j = \rho dt - J^x dx - J^y dy - J^z dz \qquad [56]$$

The Hodge dual is $*j = \sigma^3 - j^2 \wedge dt$, with the 3-form $\sigma^3 = \rho dx \wedge dy \wedge dz$, and the 2-form

$$j^2 = -J^x dy \wedge dz - J^y dz \wedge dx - J^z dx \wedge dy \qquad [57]$$

From the Maxwell equation $d^*F = -j$, it follows that

$$(d^*)^2 F = -d^* j = 0 \qquad [58]$$

that is

$$\begin{aligned} *d(*j) &= *d(\sigma^3 - j^2 \wedge dt) = *(\boldsymbol{d}\sigma^3 - \boldsymbol{d}j^2 \wedge dt) \\ &= *\left(\frac{\partial \rho}{\partial t} + \mathrm{div}\boldsymbol{J}\right) dt \wedge dx \wedge dy \wedge dz \\ &= \frac{\partial \rho}{\partial t} + \mathrm{div}\boldsymbol{J} = 0 \end{aligned} \qquad [59]$$

This is the “continuity equation.”

The total charge inside a volume $V$ is $Q = \int_V \rho dV$, therefore

$$-\frac{dQ}{dt} = -\frac{d}{dt}\int_V \rho\, dV = \int_{\partial V} \boldsymbol{J} \cdot \boldsymbol{n}\, dS \qquad [60]$$

where $\partial V$ is the surface which encloses the volume $V$, $dS$ is the surface element, and $\boldsymbol{n}$ is the normal vector to this surface. This is current conservation.

## The Gauge Potential

The "Poincaré lemma" tells us that $dF=0$ implies $F=dA$, with the 4-potential $A$:

$$A = \phi dt + \boldsymbol{A} \quad [61]$$

and the vector potential $\boldsymbol{A}=A_x dx + A_y dy + A_z dz$. From

$$\begin{aligned} F = \mathcal{E} \wedge dt + \mathcal{B} &= \left(\boldsymbol{d} + dt \wedge \frac{\partial}{\partial t}\right) A \\ &= \boldsymbol{d}\phi \wedge dt + \boldsymbol{dA} + dt \wedge \frac{\partial \boldsymbol{A}}{\partial t} \end{aligned} \quad [62]$$

it follows by comparing coefficients that

$$\mathcal{E} = \boldsymbol{d}\phi - \frac{\partial \boldsymbol{A}}{\partial t}, \qquad \mathcal{B} = \boldsymbol{dA} \quad [63]$$

In vector notation this is

$$\boldsymbol{E} = \mathrm{grad}\phi - \frac{\partial \boldsymbol{A}}{\partial t}, \qquad \boldsymbol{B} = \mathrm{curl}\,\boldsymbol{A} \quad [64]$$

The 4-potential is determined up to a gauge function $\Lambda$:

$$A' = A + d\Lambda \quad [65]$$

This gauge freedom has no influence on the observable quantities $\boldsymbol{E}$ and $\boldsymbol{B}$:

$$F' = dA' = dA + d^2\Lambda = dA = F \quad [66]$$

The Laplace operator is $\triangle = (d^* + d)^2 = dd^* + d^*d$, so when the 4-potential $A$ fulfills the condition $d^*A = 0$, we have

$$\triangle A = d^* dA = d^* F = -j \quad [67]$$

the "classical wave equation." The condition $d^*A=0$ is called the "Lorentz gauge condition." This condition can always be fulfilled by using the gauge freedom: $d^*(A + d\Lambda) = 0$ is fulfilled when $d^*d\Lambda = \triangle\Lambda = -d^*A$, where we have used the fact that $d^*\Lambda = 0$ for functions. That is to say, $d^*A = 0$ is fulfilled when $\Lambda$ is a solution of the inhomogeneous wave equation.

## Gauge Invariance

In quantum mechanics, the electron is described by a wave function which is determined up to a free phase. Indeed, at every point in space this phase can be chosen arbitrarily:

$$\begin{aligned} \psi(x) &\rightarrow \psi'(x) = \exp\{\mathrm{i}\alpha(x)\}\psi(x) \\ \bar{\psi}(x) &\rightarrow \bar{\psi}'(x) = \bar{\psi}(x)\exp\{-\mathrm{i}\alpha(x)\} \end{aligned} \quad [68]$$

with the only condition being that $\alpha(x)$ is a continuous function. The gauge transformation is of the form $g = \exp\{\mathrm{i}\alpha(x)\}$, with $g$ an element of the abelian gauge group $G = \mathrm{U}(1)$. The free action is

$$S_0 = \int \mathcal{L}_0 \, d^4x \quad [69]$$

with

$$\mathcal{L}_0 = \bar{\psi}(\mathrm{i}\gamma^\mu \partial_\mu - m)\psi \quad [70]$$

the "Lagrange density." This action is not invariant under gauge transformations:

$$\mathcal{L}_0 \rightarrow \mathcal{L}_0' = \bar{\psi}(\mathrm{i}\gamma^\mu\partial_\mu - m)\psi - (\partial_\mu\alpha)\bar{\psi}\gamma^\mu\psi \quad [71]$$

The undesired term can be compensated by the introduction of a gauge potential $\omega$ in a covariant derivative of $\psi$,

$$D\psi = (d + \omega)\psi \quad [72]$$

which has the desired transformation property $D\psi \rightarrow \exp\{\mathrm{i}\alpha\}D\psi$ when besides the transformation $\psi(x) \rightarrow \exp\{\mathrm{i}\alpha(x)\}\psi(x)$ of the matter field the gauge potential simultaneously transforms according to the gauge transformation $\omega \rightarrow \omega - \mathrm{i}d\alpha$. The new Lagrange density is

$$\mathcal{L} = \bar{\psi}(\mathrm{i}\gamma^\mu D_\mu - m)\psi = \mathcal{L}_0 + \mathrm{i}\omega_\mu\bar{\psi}(x)\gamma^\mu\psi(x) \quad [73]$$

The substitution $\partial_\mu \rightarrow D_\mu$ is known to physicists; with $\omega = -\mathrm{i}qA$ it is the ansatz of minimal coupling for taking into account electromagnetic effects: $\partial_\mu \rightarrow \partial_\mu - \mathrm{i}qA_\mu$. The Lagrange density becomes in this notation $\mathcal{L} = \mathcal{L}_0 - A_\mu J^\mu$, where $J^\mu = -q\bar{\psi}\gamma^\mu\psi$.

The Lagrange density must now be completed by a kinetic term for the gauge potential and we get the complete electromagnetic Lagrange density

$$\mathcal{L} = \mathcal{L}_0 - A_\mu J^\mu - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad [74]$$

with $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. In the action this corresponds to

$$S = S_0 - \int_M A_\mu J^\mu \mathrm{vol}^4 - \frac{1}{4}\int_M F_{\mu\nu}F^{\mu\nu}\mathrm{vol}^4 \quad [75]$$

We get the field equations for the potential $A$ by demanding that the variation of the action vanishes:

$$\delta S[A] = -\int_M \delta A_\mu J^\mu \mathrm{vol}^4 - \frac{1}{4}\delta\int_M F_{\mu\nu}F^{\mu\nu}\mathrm{vol}^4 \quad [76]$$

We write now

$$\int_M \delta A_\mu J^\mu \mathrm{vol}^4 = (\delta A, j) \quad [77]$$

and

$$\begin{aligned}\frac{1}{4}\delta\int_M F_{\mu\nu}F^{\mu\nu}\mathrm{vol}^4 \\ =\frac{1}{2}\delta\int_M F\wedge *F=\frac{1}{2}\delta(F,F) \\ =(\delta dA,F)=(d\delta A,F)=(\delta A,d^*F)\end{aligned} \quad [78]$$

where we have exchanged the action of $\delta$ and $d$. Since this holds for arbitrary variations $\delta A$ we find

$$d^*F=-j \quad [79]$$

the inhomogeneous Maxwell equation.

## Nonabelian Gauge Theories

In SU($N$) gauge theory the elementary particles are taken to be members of symmetry multiplets. For example, in electroweak theory the left-handed electron and the neutrino are members of an SU(2) doublet:

$$\psi=\begin{pmatrix}e^-\\ \nu\end{pmatrix} \quad [80]$$

A gauge transformation is

$$\psi'(x)=g^{-1}(x)\psi(x),\qquad \bar{\psi}'(x)=\bar{\psi}(x)g(x) \quad [81]$$

with

$$g(x)=\exp\{\Lambda(x)\} \quad [82]$$

where $g(x)$ is an element of the Lie group SU(2) and $\Lambda$ is an element of the Lie algebra su(2). The Lie algebra is a vector space, and its elements may be expanded in terms of a basis:

$$\Lambda(x)=\Lambda^a(x)T_a \quad [83]$$

For su(2) the basis elements are traceless and anti-Hermitian (see below), they are conventionally expressed in terms of the Pauli matrices,

$$T_a=\frac{\sigma_a}{2\mathrm{i}} \quad [84]$$

with

$$\sigma_1=\begin{pmatrix}0&1\\1&0\end{pmatrix},\qquad \sigma_2=\begin{pmatrix}0&-\mathrm{i}\\ \mathrm{i}&0\end{pmatrix} \\ \sigma_3=\begin{pmatrix}1&0\\0&-1\end{pmatrix} \quad [85]$$

They are conventionally normalized according to

$$\mathrm{tr}(T_aT_b)=-\tfrac{1}{2}\delta_{ab} \quad [86]$$

The Dirac Lagrangian is not invariant with respect to local gauge transformations:

$$\begin{aligned}\mathcal{L}_0=\bar{\psi}(\mathrm{i}\gamma^\mu\partial_\mu-m)\psi\to\mathcal{L}_0' \\ =\mathcal{L}_0+\mathrm{i}\bar{\psi}\gamma^\mu(g\partial_\mu g^{-1})\psi\end{aligned} \quad [87]$$

We introduce the gauge potential

$$\omega_\mu(x)=\omega_\mu^a(x)T_a \quad [88]$$

with a gauge transformation

$$\omega_\mu\to\omega_\mu'=g^{-1}\omega_\mu g+g^{-1}\partial_\mu g \quad [89]$$

The Lagrange density is modified through a covariant derivative:

$$\partial_\mu\to D_\mu=\partial_\mu+\omega_\mu \quad [90]$$

The covariant derivative $D_\mu$ transforms according to

$$D_\mu\to D_\mu{}'=g^{-1}D_\mu g \quad [91]$$

and thus the modified Lagrange density

$$\mathcal{L}=\bar{\psi}(\mathrm{i}\gamma^\mu D_\mu-m)\psi=\mathcal{L}_0+\mathrm{i}\bar{\psi}\gamma^\mu\omega_\mu\psi \quad [92]$$

is invariant with respect to local gauge transformations.

The extra term in the Langrange density is conventionally written

$$-J_a^\mu A_\mu^a \quad [93]$$

with

$$A_\mu^a=-\mathrm{i}q\omega_\mu^a \quad [94]$$

and

$$J_a^\mu=\bar{\psi}\gamma^\mu T_a\psi \quad [95]$$

In mathematical terminology $\omega$ is called a connection. The quantity $A$ is the physicist's gauge potential. The connection is anti-Hermitian and the gauge potential Hermitian. The gauge potential also includes the coupling constant $q$. We will refer to both $\omega$ and $A$ as the gauge potential, where the relation between them is given by eqn [94].

We can write the gauge potential as $A=A_\mu^a dx^\mu T_a$ or, in the SU(2) case, as

$$A_\mu=A_\mu^1T_1+A_\mu^2T_2+A_\mu^3T_3 \quad [96]$$

where we see explicitly that it involves three vector fields, which couple to the electroweak currents [95] with the single coupling constant $q$, and which will become after symmetry breaking the three vector bosons $W_+, W_-, Z_0$ of the electroweak gauge theory. Actually, a mix of the neutral gauge boson and the photon will combine to yield the $Z_0$ boson, while the orthogonal mixture gives rise to the electromagnetic interaction, in an $\mathrm{SU}(2)\times\mathrm{U}(1)$ theory. At this stage,

the gauge bosons are all massless, their masses are generated by the "Higgs' mechanism."

## Lie-Algebra-Valued $p$-Forms

To describe nonabelian fields, we need Lie-algebra-valued $p$-forms:

$$\phi = T_a \phi^a \qquad [97]$$

where $T_a$ is a generator of the Lie algebra, the index $a$ runs over the number of generators of the Lie algebra, and the $\phi^a$ are the usual scalar-valued $p$-forms. The composition in a Lie algebra is a Lie bracket, which is defined for two Lie-algebra-valued $p$-forms by

$$[\phi, \psi] := [T_a, T_b]\phi^a \wedge \psi^b \qquad [98]$$

The Lie bracket in the algebra is

$$[T_a, T_b] = f^c_{ab} T_c \qquad [99]$$

where $f^a_{bc}$ are the structure constants. It follows from this that

$$[\psi, \phi] = [T_a, T_b]\psi^a \wedge \phi^b = -[T_b, T_a]\psi^a \wedge \phi^b \qquad [100]$$

or

$$[\psi, \phi] = (-1)^{pq+1}[\phi, \psi] \qquad [101]$$

when $\phi$ is a $p$-form and $\psi$ is a $q$-form. In the special case that $T_a$ is a matrix, also the product $T_a T_b$ is defined, and from this the product of two Lie-algebra-valued $p$-forms

$$\phi \wedge \psi = T_a \phi^a \wedge T_b \psi^b = T_a T_b \phi^a \wedge \psi^b \qquad [102]$$

Now the Lie bracket is a commutator:

$$[T_a, T_b] = T_a T_b - T_b T_a \qquad [103]$$

and

$$\begin{aligned}[\phi, \psi] &= [T_a, T_b]\phi^a \wedge \psi^b \\ &= T_a\phi^a \wedge T_b\psi^b - (-1)^{pq} T_b\psi^b \wedge T_a\phi^a \\ &= \phi \wedge \psi - (-1)^{pq}\psi \wedge \phi \end{aligned} \qquad [104]$$

From this relation it follows that for $\phi$ and $\psi$ odd $p$-forms

$$[\phi, \psi] = \phi \wedge \psi + \psi \wedge \phi \qquad [105]$$

For $\phi$ an odd $p$-form

$$[\phi, \phi] = \phi \wedge \phi + \phi \wedge \phi = 2(\phi \wedge \phi) \qquad [106]$$

## The Gauge Potential and the Field Strength

The generalization of the abelian relationship between the gauge potential and the field strength, $F = dA$, is

$$\theta = d\omega + \tfrac{1}{2}[\omega, \omega] = d\theta + \omega \wedge \omega \qquad [107]$$

where because $\omega$ is a 1-form we can use eqn [106]. The mathematician refers to $\theta$ as the curvature. The physicist writes, in analogy to eqn [94],

$$F = -\mathrm{i}\, q\theta = \tfrac{1}{2} F^a_{\mu\nu} dx^\mu \wedge dx^\nu T_a \qquad [108]$$

One obtains for the components

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - \mathrm{i} q f^a_{bc} A^b_\mu A^c_\nu \qquad [109]$$

A generalization of the gauge transformation of $A$, that is, $A' = A + d\Lambda$, is eqn [89]:

$$\omega' = g^{-1}\omega g + g^{-1} dg \qquad [110]$$

A quantity $\phi$ with the transformation property

$$\phi' = g^{-1}\phi g \qquad [111]$$

is called a "tensorial" quantity. The gauge potential $\omega$ is according to this definition nontensorial. Nevertheless the field strength is tensorial. Indeed

$$\begin{aligned}\theta' = {} & d(g^{-1}\omega g) + (dg^{-1}) \wedge dg \\ & + \tfrac{1}{2}[g^{-1}\omega g + g^{-1}dg, g^{-1}\omega g + g^{-1}dg] \\ = {} & (dg^{-1}) \wedge \omega g + g^{-1} d\omega g - g^{-1}\omega \wedge dg + (dg^{-1}) \wedge dg \\ & + \tfrac{1}{2} g^{-1}[\omega, \omega] g + \tfrac{1}{2}[g^{-1}\omega g, g^{-1}dg] \\ & + \tfrac{1}{2}[g^{-1}dg, g^{-1}\omega g] + \tfrac{1}{2}[g^{-1}dg, g^{-1}dg] \\ = {} & g^{-1}\theta g + (dg^{-1}) \wedge \omega g - g^{-1}\omega \wedge dg + (dg^{-1}) \wedge dg \\ & + g^{-1}\omega \wedge dg + g^{-1}dg \wedge g^{-1}\omega g + g^{-1}dg \wedge g^{-1}dg \\ = {} & g^{-1}\theta g \end{aligned} \qquad [112]$$

where we have used the derivation of the relation $g^{-1}g = \mathrm{Id}$ to get

$$dg^{-1} = -g^{-1} dg\, g^{-1} \qquad [113]$$

In the abelian case, we had $dF = 0$. The nonabelian analog is

$$\begin{aligned} d\theta &= d\omega \wedge \omega - \omega \wedge d\omega \\ &= (\theta - \omega \wedge \omega) \wedge \omega - \omega \wedge (\theta - \omega \wedge \omega) \\ &= \theta \wedge \omega - \omega \wedge \theta \end{aligned} \qquad [114]$$

or

$$d\theta + \omega \wedge \theta - \theta \wedge \omega = 0 \qquad [115]$$

the Bianchi identity. It can also be written as

$$d\theta + \omega \wedge \theta - \theta \wedge \omega = d\theta + [\omega, \theta] = 0 \qquad [116]$$

because from eqn [104]

$$\omega \wedge \theta + (-1)^{2 \cdot 1} \theta \wedge \omega = [\omega, \theta] \qquad [117]$$

The covariant derivative $D$ is defined as

$$D\phi := d\phi + [\omega, \phi] \qquad [118]$$

for $\phi$ a tensorial quantity. The covariant derivative takes tensorial $p$-forms into tensorial $(p+1)$-forms:

$$\begin{aligned} D'\phi' &= d(g^{-1}\phi g) + [g^{-1}\omega g + g^{-1}dg, g^{-1}\phi g] \\ &= dg^{-1} \wedge \phi g + g^{-1} d\phi g + (-1)^p g^{-1} \phi \wedge dg \\ &\quad + [g^{-1}\omega g, g^{-1}\phi g] + [g^{-1}dg, g^{-1}\phi g] \\ &= g^{-1} D\phi g + dg^{-1} \wedge \phi g + (-1)^p g^{-1} \phi \wedge dg \\ &\quad + g^{-1} dg g^{-1} \wedge \phi g - (-1)^p g^{-1} \phi \wedge dg \\ &= g^{-1} D\phi g \end{aligned} \qquad [119]$$

We have thereby verified the transformation property of eqn [91].

## The Gauge Group

From the gauge transformation $\psi' = g\psi$ the requirement $|\psi'|^2 = |\psi|^2$ leads to $g^\dagger g = 1$. That means that $g$ belongs to the unitary Lie group $G = \mathrm{U}(n)$, whose elements fulfill $g^\dagger = \bar{g}^{\mathrm{T}} = g^{-1}$. For elements of the Lie algebra $\mathcal{G} = \mathrm{u}(n)$ this implies

$$\left(\mathrm{e}^X\right)^\dagger = \mathrm{e}^{\bar{X}^{\mathrm{T}}} = \mathrm{e}^{-X} \qquad [120]$$

or

$$X^\dagger = \bar{X}^{\mathrm{T}} = -X \qquad [121]$$

where $\bar{X}$ is complex conjugation and $X^{\mathrm{T}}$ means transposition.

For elements of the Lie algebra we can define a scalar product (the Killing metric)

$$\langle X, Y \rangle := -\mathrm{tr}\,(XY) = -X^\alpha{}_\beta X^\beta{}_\alpha \qquad [122]$$

The scalar product is real:

$$\langle \bar{X}, \bar{Y} \rangle = -\bar{X}^\beta{}_\alpha \bar{Y}^\alpha{}_\beta = -X^\alpha{}_\beta X^\beta{}_\alpha = \langle X, Y \rangle \qquad [123]$$

symmetric:

$$\langle X, Y \rangle = -\mathrm{tr}(X, Y) = -\mathrm{tr}(Y, X) = \langle Y, X \rangle \qquad [124]$$

and positive definite:

$$\langle X, X \rangle = -X^\alpha{}_\beta X^\beta{}_\alpha = X^\alpha{}_\beta \bar{X}^\alpha{}_\beta = |X^\alpha{}_\beta|^2 \qquad [125]$$

The scalar product is invariant under the action of $G$ on $\mathcal{G}$: for $g \in G$

$$\begin{aligned} \langle gXg^{-1}, gYg^{-1} \rangle &= -\mathrm{tr}\,(gXYg^{-1}) \\ &= -\mathrm{tr}(X, Y) = \langle X, Y \rangle \end{aligned} \qquad [126]$$

or for $X, Y, Z \in \mathcal{G}$

$$\langle \mathrm{e}^{tX} Y \mathrm{e}^{-tX}, \mathrm{e}^{tX} Z \mathrm{e}^{-tX} \rangle = \langle Y, Z \rangle \qquad [127]$$

We take the derivative of this equation with respect to $t$ at the value $t = 0$ and get:

$$\langle [X, Y], Z \rangle + \langle Y, [X, Z] \rangle = 0 \qquad [128]$$

We define an action of the algebra $\mathcal{G}$ on itself: $ad(X): \mathcal{G} \to \mathcal{G}$

$$ad(X)Y = [X, Y] \qquad [129]$$

We can then formulate our conclusion as follows: the action of $\mathcal{G}$ on itself is anti-Hermitian:

$$\langle ad(X)Y, Z \rangle = -\langle Y, ad(X)Z \rangle \qquad [130]$$

or

$$[ad(X)]^\dagger = -ad(X) \qquad [131]$$

From $g^\dagger g = 1$ we have $|\det(g)|^2 = 1$. For the gauge group $G = \mathrm{SU}(N)$ we require in addition $\det(g) = 1$. Since

$$\det(g) = \det(\exp(X)) = \exp(\mathrm{tr}(X)) \qquad [132]$$

the elements $X \in \mathrm{su}(N)$ must be traceless. A basis of the vector space of traceless, anti-Hermitian $(2 \times 2)$ matrices is given by the Pauli matrices, eqn [85].

## The Yang–Mills Action

The SU(2) Yang–Mills action is, in analogy to the abelian case,

$$\begin{aligned} S &= -\frac{1}{4q^2} \int_M F^a_{\mu\nu} F^{a\mu\nu} \mathrm{vol}^4 = \frac{1}{2q^2} \int_M \mathrm{tr}(F_{\mu\nu} F^{\mu\nu}) \mathrm{vol}^4 \\ &= \frac{1}{2q^2} \int_M \mathrm{tr}(F \wedge * F) \end{aligned} \qquad [133]$$

We have included the trace in our definition of the scalar product:

$$(\phi, \psi) := -\int_M \mathrm{tr} < \phi_I \psi^I > \mathrm{vol}^n = -\int_M \mathrm{tr}(\phi \wedge * \psi) \qquad [134]$$

We then write eqn [133] as

$$S[\omega] = \tfrac{1}{2} (\theta, \theta) \qquad [135]$$

taking into account the relation between $\theta$ and the field strength $F$, and indicating the dependence on

the gauge potential. Since $\theta$ is tensorial the action is invariant.

Now we calculate the variation von $S[\omega]$ with respect to a variation of the gauge potential:

$$\begin{aligned}\delta S[\omega] &= \frac{d}{dt}S[\omega(t)]|_{t=0} = \frac{1}{2}\delta(\theta,\theta) \\ &= \frac{1}{2}((\delta\theta,\theta)+(\theta,\delta\theta)) \\ &= (\delta\theta,\theta) = \left(\delta\left(d\omega+\frac{1}{2}[\omega,\omega]\right),\theta\right) \\ &= \left(\delta d\omega+\frac{1}{2}[\delta\omega,\omega]+\frac{1}{2}[\omega,\delta\omega],\theta\right) \\ &= (d\delta\omega+[\omega,\delta\omega],\theta)\end{aligned} \quad [136]$$

where we have exchanged the order of $\delta$ and $d$. We remark that although $\omega$ is not a tensorial section, $\delta\omega$ is: for $\omega'_1 = g^{-1}\omega_1 g + g^{-1}dg$ and $\omega'_2 = g^{-1}\omega_2 g + g^{-1}dg$ is

$$\delta\omega = \omega'_1 - \omega'_2 = g^{-1}(\omega_1-\omega_2)g \quad [137]$$

The quantity $\theta$ is in any case tensorial. Therefore, the covariant derivative is defined, and we have

$$D\delta\omega = d\delta\omega + [\omega,\delta\omega] \quad [138]$$

and

$$D\theta = d\theta + [\omega,\theta] \quad [139]$$

In general, the action of the covariant derivative on tensorial quantities can be written as $D = d + ad(\omega)$, where $ad(X)$ is the representation of the Lie algebra on itself introduced in the previous section. We now have

$$\delta S[\omega] = (D\delta\omega,\theta) = (\delta\omega, D^*\theta) = 0 \quad [140]$$

for an arbitrary variation $\delta\omega$. Therefore, $D^*\theta = 0$.

We have obtained

$$D^*\theta = 0 \quad [141]$$

the "Yang–Mills equations," and

$$D\theta = 0 \quad [142]$$

the "Bianchi identites." These are the generalizations of the Maxwell equations $d^*F = 0$ and $dF = 0$ in the absence of external sources. For the general case of interacting fermions, we write out the full action, in analogy to eqn [74], and obtain, in analogy to eqns [79] and [58],

$$D^*\theta = -J, \qquad D^*J = 0 \quad [143]$$

We shall now derive, again for the pure gauge sector, coordinate expressions for the Yang–Mills equations. Consider the expression

$$\begin{aligned}\delta S[\omega] &= (D\delta\omega,\theta) = (\delta\omega, D^*\theta) \\ &= (d\delta\omega + [\omega,\delta\omega],\theta)\end{aligned} \quad [144]$$

The first term in the last expression is

$$(d\delta\omega,\theta) = (\delta\omega, d^*\theta) = -\mathrm{tr}\int_M \delta\omega_\nu\{d^*\theta\}^\nu \mathrm{vol}^4 \quad [145]$$

The second term can be computed using

$$\begin{aligned}[][\omega,\delta\omega]_{\mu\nu} &= \{\omega\wedge\delta\omega + \delta\omega\wedge\omega\}(\partial_\mu,\partial_\nu) \\ &= \omega_\mu\delta\omega_\nu - \omega_\nu\delta\omega_\mu + \delta\omega_\mu\omega_\nu - \delta\omega_\nu\omega_\mu\end{aligned} \quad [146]$$

and hence

$$[\omega,\delta\omega]_{\mu\nu}\theta^{\mu\nu} = 2[\omega_\mu,\delta\omega_\nu]\theta^{\mu\nu} \quad [147]$$

because $\theta$ is antisymmetric, $\theta^{\mu\nu} = -\theta^{\nu\mu}$. Thus,

$$\begin{aligned}([\omega,\delta\omega],\theta) &= -\int_M \mathrm{tr}([\omega,\delta\omega]\wedge *\theta) \\ &= -\frac{1}{2}\int_M \mathrm{tr}([\omega,\delta\omega]_{\mu\nu}\theta^{\mu\nu})\mathrm{vol}^4 \\ &= -\int_M \mathrm{tr}([\omega_\mu,\delta\omega_\nu]\theta^{\mu\nu})\mathrm{vol}^4 \\ &= \int_M \langle[\omega_\mu,\delta\omega_\nu],\theta^{\mu\nu}\rangle\mathrm{vol}^4\end{aligned} \quad [148]$$

where $\langle\,,\rangle$ is the scalar product in $\mathcal{G}$. From eqn [128] this equals

$$\begin{aligned}&-\int_M \langle\delta\omega_\nu,[\omega_\mu,\theta^{\mu\nu}]\rangle\mathrm{vol}^4 \\ &\quad = \int_M \mathrm{tr}(\delta\omega_\nu[\omega_\mu,\theta^{\mu\nu}])\mathrm{vol}^4\end{aligned} \quad [149]$$

Combining this with eqn [144] gives

$$\begin{aligned}(\delta\omega, D^*\theta) &= -\int_M \mathrm{tr}(\delta\omega_\nu\{(d^*\theta)^\nu - [\omega_\mu,\theta^{\mu\nu}]\})\mathrm{vol}^4 \\ &= (\delta\omega, \{(d^*\theta)^\nu - [\omega_\mu,\theta^{\mu\nu}]\})\end{aligned} \quad [150]$$

We can now insert the coordinate expression for

$$(d\theta)^\nu = -\partial_\mu\theta^{\mu\nu} \quad [151]$$

Finally, the coordinate expressions of the Yang–Mills equations $D^*\theta = 0$ are

$$(D^*\theta)^\nu = -\{\partial_\mu\theta^{\mu\nu} + [\omega_\mu,\theta^{\mu\nu}]\} = 0 \quad [152]$$

## The Analogy with Electromagnetism

The Yang–Mills equation and the Bianchi identity in the absence of external sources are

$$\partial_\nu F^{\mu\nu} - \mathrm{i}q[A_\nu, F^{\mu\nu}] = 0 \quad [153]$$

and

$$\begin{aligned}&\partial_\mu F_{\nu\tau} + \partial_\tau F_{\mu\nu} + \partial_\nu F_{\tau\mu} - \mathrm{i}q\{[A_\mu, F_{\nu\tau}] \\ &\quad + [A_\tau, F_{\mu\nu}] + [A_\nu, F_{\tau\mu}]\} = 0\end{aligned} \quad [154]$$

We shall write these equations in terms of the fields

$$F^{i0} = E^i, \quad i = 1,2,3 \qquad [155]$$

$$F^{12} = B^3, \qquad F^{31} = B^2, \qquad F^{12} = B^3 \qquad [156]$$

where the **E** and **B** vectors may be thought of as "electric" and "magnetic" fields, even though they have Lie-algebra indices, $F^{i0} = (F^a)^{i0} T_a$, etc. In the context of the SU(3) theory, they are referred to as the "chromoelectric" and "chromomagnetic" fields, respectively.

The Yang–Mills equations with $\mu = 0$ are

$$\partial_i F^{i0} - \mathrm{i}q[A_i, F^{i0}] = 0 \qquad [157]$$

with $i = 1,2,3$ a spatial index. In vector notation this is

$$\operatorname{div} \boldsymbol{E} = \mathrm{i}q(\boldsymbol{A} \cdot \boldsymbol{E} - \boldsymbol{E} \cdot \boldsymbol{A}) \qquad [158]$$

This is the analog of Gauss's equation. Even though we started out without external sources, $\mathrm{i}q(\boldsymbol{A} \cdot \boldsymbol{E} - \boldsymbol{E} \cdot \boldsymbol{A})$ plays the role of a "charge density." The Yang–Mills field $\boldsymbol{E}$ and the potential $\boldsymbol{A}$ combine to act as a source for the Yang–Mills field. This is an essential feature of nonabelian gauge theories in which they differ from the abelian case, due to the fact that the commutator $[\boldsymbol{A}, \boldsymbol{E}]$ is nonvanishing.

Now consider the Yang–Mills equations with a spatial index $\mu = i$:

$$\partial_0 F^{i0} + \partial_j F^{ij} - \mathrm{i}q[A_0, F^{i0}] - \mathrm{i}q[A_j F^{ij}] = 0 \qquad [159]$$

In vector notation this is

$$\operatorname{curl} \boldsymbol{B} = \frac{\partial \boldsymbol{E}}{\partial t} - \mathrm{i}q(A_0 \boldsymbol{E} - \boldsymbol{E} A_0) + \mathrm{i}q(\boldsymbol{A} \times \boldsymbol{B} + \boldsymbol{B} \times \boldsymbol{A}) \qquad [160]$$

replacing the Ampere–Maxwell law. Note that there are two extra contributions to the "current" other than the displacement current.

The analogs of the laws of Faraday and of the absence of magnetic monopoles are derived similarly from the Bianchi identities. The results are

$$\operatorname{curl} \boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} = \mathrm{i}q\{(\boldsymbol{A} \times \boldsymbol{E} + \boldsymbol{E} \times \boldsymbol{A}) + (A_0 \boldsymbol{B} - \boldsymbol{B} A_0)\} \qquad [161]$$

and

$$\operatorname{div} \boldsymbol{B} = \mathrm{i}q(\boldsymbol{A} \cdot \boldsymbol{B} - \boldsymbol{B} \cdot \boldsymbol{A}) \qquad [162]$$

## Further Remarks

The foundations of the mathematics of differential forms were laid down by Poincaré (1953). They were applied to the description of electrodynamics already by Cartan (1923). A modern presentation of differential forms and the manifolds on which they are defined is given in Abraham *et al.* (1983). A recent treatment of electrodynamics in this approach is Hehl and Obukhov (2003). Weyl's argument is in his paper of 1929.

Nonabelian gauge theories today explain the electromagnetic, the strong and weak nuclear interactions. The original paper is that of Yang and Mills (1954). Glashow, Salam, and Weinberg (1980) saw the way to apply it to the weak interactions by using spontaneous symmetry breaking to generate the masses through the use of the Higgs' (1964) mechanism. t'Hooft and Veltman (1972) showed that the resulting quantum field theory was renormalizable. The strong interactions were recognized as the nonabelian gauge theory with gauge group SU(3) by Gell-Mann (1972). For a modern treatment which puts nonabelian gauge theories in the context of differential geometry, see Frankel (1987).

*See also:* Dirac Fields in Gravitation and Nonabelian Gauge Theory; Electroweak Theory; Measure on Loop Spaces; Nonperturbative and Topological Aspects of Gauge Theory; Quantum Electrodynamics and its Precision Tests.

## Further Reading

Abraham A, Marsden J, and Ratiu T (1983) *Manifolds, Tensor Analysis, and Applications*. MA: Addison-Wesley.

Cartan É (1923) *On manifolds with an Affine Connection and the Theory of General Relativity*. English translation of the French original 1923/1924 (Bibliopolis, Napoli 1986).

Frankel T (1987) *The Geometry of Physics, An Introduction*. Cambridge University Press.

Gell-Mann M (1972) Quarks: developments in the quark theory of hadrons. *Acta Physica Austriaca Suppl*. IV: 733.

Glashow SL (1980) Towards a unified theory: threads in a tapestry. *Reviews of Modern Physics* 52: 539.

Hehl FW and Obukhov YN (2003) *Foundations of Classical Electrodynamics*. Boston: Birkhäuser.

Higgs PW (1964) Broken symmetries and the masses of gauge bosons. *Physical Review Letters* 13: 508.

t'Hooft G and Veltman M (1972) Regularization and renormalization of gauge fields. *Nuclear Physics B* 44: 189.

Poincaré H (1953) *Oeuvre*. Paris: Gauthier-Villars.

Salam A (1980) Gauge unification of fundamental forces. *Reviews of Modern Physics* 52: 525.

Weinberg SM (1980) Conceptual foundations of the unified theory of weak and electromagnetic interactions. *Reviews of Modern Physics* 52: 515.

Weyl H (1929) Elektron und gravitation. *Zeitschrift fuer Physik* 56: 330.

Yang CN and Mills RL (1954) Construction of isotopic spin and isotopic gauge invariance. *Physical Review* 96: 191.

# Abelian Higgs Vortices

**J M Speight**, University of Leeds, Leeds, UK



## Introduction

For the purpose of this article, vortices are topological solitons arising in field theories in $(2+1)$-dimensional spacetime when a complex-valued field $\phi$ is allowed to acquire winding at infinity, meaning that the phase of $\phi(t, \boldsymbol{x})$, as $\boldsymbol{x}$ traverses a large circle in the spatial plane, changes by $2\pi n$, where $n$ is a nonzero integer. Such winding cannot be removed by any continuous deformation of $\phi$ (hence "topological") and traps a considerable amount of energy which tends to coalesce into smooth, stable lumps with highly particle-like characteristics (hence "solitons"). Clearly, the universe is $(3+1)$ dimensional. Nonetheless, planar field theories are of physical interest for two main reasons. First, the theory may arise by dimensional reduction of a $(3+1)$-dimensional model under the assumption of translation invariance in one direction. Vortices are then transverse slices through straight tube-like objects variously interpreted as magnetic flux tubes in a superconductor or cosmic strings. Second, a crucial ingredient of the standard model of particle physics is spontaneous breaking of gauge symmetry by a Higgs field. As well as endowing the fundamental gauge bosons and chiral fermions with mass, this mechanism can potentially generate various types of topological solitons (monopoles, strings, and domain walls) whose structure and interactions one would like to understand. Vortices in $(2+1)$ dimensions are interesting in this regard because they arise in the simplest field theory exhibiting the Higgs mechanism, the abelian Higgs model (AHM). They are thus a useful theoretical laboratory in which to test ideas which may ultimately find application in more realistic theories. This article describes the properties of abelian Higgs vortices and explains how, using a mixture of numerical and analytical techniques, a good understanding of their dynamical interactions has been obtained.

## The Abelian Higgs Model

Throughout this article spacetime will be $\mathbb{R}^{2+1}$ endowed with the Minkowski metric with signature $(+,-,-)$, and Cartesian coordinates $x^\mu, \mu = 0,1,2$, with $x^0 = t$ (the speed of light $c = 1$). A spacetime point will be denoted $x$, its spatial part by $\boldsymbol{x} = (x^1, x^2)$. Latin indices $j, k, \ldots$ range over 1, 2, and repeated indices (Latin or Greek) are summed over. We sometimes use polar coordinates in the spatial plane, $\boldsymbol{x} = r(\cos\theta, \sin\theta)$, and sometimes a complex coordinate $z = x^1 + \mathrm{i}x^2 = r\mathrm{e}^{\mathrm{i}\theta}$. Occasionally, it is convenient to think of $\mathbb{R}^{2+1}$ as a subspace of $\mathbb{R}^{3+1}$ and denote by $\boldsymbol{k}$ the unit vector in the (fictitious) third spatial direction. The complex scalar Higgs field is denoted $\phi$, and the electromagnetic gauge potential $A_\mu$, best thought of as the components of a 1-form $A = A_\mu \,\mathrm{d}x^\mu$. $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the field strength tensor which, in $\mathbb{R}^{2+1}$, has only three independent components, identified with the magnetic field $B = F_{12}$ and electric field $(E_1, E_2) = (F_{01}, F_{02})$. The gauge-covariant derivative is $D_\mu\phi = \partial_\mu\phi - \mathrm{i}eA_\mu\phi$, $e$ being the electric charge of the Higgs. Under a U(1) gauge transformation,

$$\phi \mapsto \mathrm{e}^{\mathrm{i}\Lambda}\phi, \qquad A_\mu \mapsto A_\mu + \mathrm{e}^{-1}\partial_\mu\Lambda \qquad [1]$$

$\Lambda : \mathbb{R}^{2+1} \to \mathbb{R}$ being any smooth function, $F_{\mu\nu}$ and $|\phi|$ remain invariant, while $D_\mu\phi \mapsto \mathrm{e}^{\mathrm{i}\Lambda}D_\mu\phi$. Only gauge-invariant quantities are physically observable (classically).

With these conventions, the AHM has Lagrangian density

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{\alpha}{2}D_\mu\phi\overline{D^\mu\phi} - \frac{\lambda}{8}(\nu^2 - |\phi|^2)^2 \qquad [2]$$

which is manifestly gauge invariant. By rescaling $\phi, A_\mu, x$ and the unit of action, we can (and henceforth will) assume that $e = \nu = \alpha = 1$. The only parameter which cannot be scaled away is $\lambda > 0$. Its value greatly influences the model's behavior.

The field equations, obtained by demanding that $\phi(x), A_\mu(x)$ be a local extremal of the action $S = \int \mathcal{L} d^3x$, are

$$\begin{aligned} D_\mu D^\mu\phi + \frac{\lambda}{2}(1 - |\phi|^2)\phi = 0 \\ \partial^\mu F_{\mu\nu} + \frac{\mathrm{i}}{2}(\phi\overline{D_\nu\phi} - \bar{\phi}D_\nu\phi) = 0 \end{aligned} \qquad [3]$$

This is a coupled set of nonlinear second-order PDEs. Of particular interest are solutions which have finite total energy. Energy is not a Lorentz-invariant quantity. To define it we must choose an inertial frame and, having broken Lorentz invariance, it is convenient to work in a temporal gauge, for which $A_0 \equiv 0$ (which may be obtained by a gauge transformation with $\Lambda(t, \boldsymbol{x}) = \int_0^t A_0(t', \boldsymbol{x})\,\mathrm{d}t'$, after which only time-independent gauge transformations are permitted). The potential energy of a field is then

$$\begin{aligned} E &= \frac{1}{2}\int \left(B^2 + D_i\phi\overline{D_i\phi} + \frac{\lambda}{4}(1 - |\phi|^2)^2\right)\mathrm{d}x^1\mathrm{d}x^2 \\ &= E_{\mathrm{mag}} + E_{\mathrm{grad}} + E_{\mathrm{self}} \end{aligned} \qquad [4]$$

while its kinetic energy is

$$E_{\text{kin}} = \frac{1}{2}\int\left(|\partial_0 \boldsymbol{A}|^2 + \partial_0\phi\overline{\partial_0\phi}\right)\mathrm{d}x^1\mathrm{d}x^2 \qquad [5]$$

If $\phi, \boldsymbol{A}$ satisfy the field equations then the total energy $E_{\text{tot}} = E_{\text{kin}} + E$ is independent of $t$. By Derrick's theorem, static solutions have $E_{\text{mag}} \equiv E_{\text{self}}$ (Manton and Sutcliffe 2004, pp. 82–87).

Configurations with finite energy have quantized total magnetic flux. To see this, note that $E$ finite implies $|\phi| \to 1$ as $r \to \infty$, so $\phi \sim e^{i\chi(r,\theta)}$ at large $r$ for some real (in general, multivalued) function $\chi$. The winding number of $\phi$ is its winding around a circle of large radius $R$, that is, the integer $n = (\chi(R, 2\pi) - \chi(R, 0))/2\pi$. Although the phase of $\phi$ is clearly gauge dependent, $n$ is not, because to change this, a gauge transformation $e^{i\Lambda} : \mathbb{R}^2 \to U(1)$ would itself need nonzero winding around the circle, contradicting smoothness of $e^{i\Lambda}$. The model is invariant under spatial reflexions, under which $n \mapsto -n$, so we will assume (unless noted otherwise) that $n \geq 0$. Finiteness of $E$ also implies that $D\phi = \mathrm{d}\phi - iA\phi \to 0$, so $A \sim -i\mathrm{d}\phi/\phi \sim \mathrm{d}\chi$ as $r \to \infty$ (note $\phi \neq 0$ for large $r$). Hence, the total magnetic flux is

$$\int_{\mathbb{R}^2} B\,\mathrm{d}^2\boldsymbol{x} = \lim_{R\to\infty}\oint_{S_R} A = \lim_{R\to\infty}\int_0^{2\pi}\partial_\theta\chi\,\mathrm{d}\theta = 2\pi n \qquad [6]$$

where $S_R = \{\boldsymbol{x} : |\boldsymbol{x}| = R\}$ and we have used Stokes's theorem. The above argument uses only generic properties of $E$, namely that finite $E_{\text{self}}$ requires $|\phi|$ to assume a nonzero constant value as $r \to \infty$. So flux quantization is a robust feature of this type of model. As presented, the argument is somewhat formal, but it can be made mathematically rigorous at the cost of gauge-fixing technicalities (Manton and Sutcliffe 2004, pp. 164–166). Note that if $n \neq 0$ then, by continuity, $\phi(\boldsymbol{x})$ must vanish at some $\boldsymbol{x} \in \mathbb{R}^2$, and one expects a lump of energy density to be associated with each such $\boldsymbol{x}$ since $\phi = 0$ maximizes the integrand of $E_{\text{self}}$.

## Radially Symmetric Vortices

The model supports static solutions within the radially symmetric ansatz $\phi = \sigma(r)e^{in\theta}, A = a(r)\,\mathrm{d}\theta$, which reduces the field equations to a coupled pair of nonlinear ODEs:

$$\begin{aligned}\frac{\mathrm{d}^2\sigma}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}\sigma}{\mathrm{d}r} - \frac{1}{r^2}(n-a)^2\sigma + \frac{\lambda}{2}(1-\sigma^2)\sigma = 0\\ \frac{\mathrm{d}^2 a}{\mathrm{d}r^2} - \frac{1}{r}\frac{\mathrm{d}a}{\mathrm{d}r} + (n-a)\sigma^2 = 0\end{aligned} \qquad [7]$$

Finite energy requires $\lim_{r\to\infty}\sigma(r) = 1$, $\lim_{r\to\infty}a(r) = n$ while smoothness requires $\sigma(r) \sim \text{const}_1 r^n, a(r) \sim \text{const}_2 r^2$ as $r \to 0$. It is known that solutions to this system, which we shall call $n$-vortices, exist for all $n, \lambda$, though no explicit formulas for them are known. They may be found numerically, and are depicted in Figure 1. Note that $\sigma$ and $a$ always rise monotonically to their vacuum values, and $B$ always falls monotonically to 0, as $r$ increases. These solutions have their magnetic flux concentrated in a single, symmetric lump, a flux tube in the $\mathbb{R}^{3+1}$ picture. In contrast, the total energy density (integrand of $E$ in [4]) is nonmonotonic for $n \geq 2$, being peaked on a ring whose radius grows with $n$. This is a common feature of planar solitons.

The large $r$ asymptotics of $n$-vortices are well understood. For $\lambda \leq 4$ one may linearize [7] about $\sigma = 1, a = n$, yielding

$$\sigma(r) \sim 1 + \frac{q_n}{2\pi}K_0(\sqrt{\lambda}r) \qquad [8]$$

$$a(r) \sim n + \frac{m_n}{2\pi}rK_1(r) \qquad [9]$$

where $q_n, m_n$ are unknown constants and $K_\alpha$ denotes the modified Bessel's function. For $\lambda > 4$ linearization is no longer well justified, and the asymptotic behaviour of $\sigma$ (though not $a$) is quite different (Manton and Sutcliffe 2004, pp. 174–175). We shall not consider this rather extreme regime further. Note that

$$K_\alpha(r) \sim \sqrt{\frac{\pi}{2r}}e^{-r} \quad \text{as } r \to \infty \qquad [10]$$

for all $\alpha$, so both $\sigma$ and $a$ approach their vacuum values exponentially fast, but with different decay lengths: $1/\sqrt{\lambda}$ for $\sigma$, 1 for $a$. This can be seen in **Figure 1a**. The constants $q_n$ and $m_n$ depend on $\lambda$ and must be inferred by comparing the numerical solutions with [8], [9]; $q = q_1$ and $m = m_1$ will receive a physical interpretation shortly.

The 1-vortex (henceforth just "vortex") is stable for all $\lambda$, but $n$-vortices with $n \geq 2$ are unstable to break up into $n$ separate vortices if $\lambda > 1$. We shall say that the AHM is type I if $\lambda < 1$, type II if $\lambda > 1$, and critically coupled if $\lambda = 1$, based on this distinction. Let $E_n$ denote the energy of an $n$-vortex. **Figure 2** shows the energy per vortex $E_n/n$ plotted against $n$ for $\lambda = 0.5, 1$, and 2. It decreases with $n$ for $\lambda = 0.5$, indicating that it is energetically favorable for isolated vortices to coalesce into higher winding lumps. For $\lambda = 2$, by contrast, $E_n/n$ increases with $n$ indicating that it is energetically favorable for $n$-vortices to fission into their constituent vortex parts. The case $\lambda = 1$ balances between these behaviors: $E_n/n$ is independent of $n$. In fact, the energy of a collection of vortices is independent of their positions in this case.

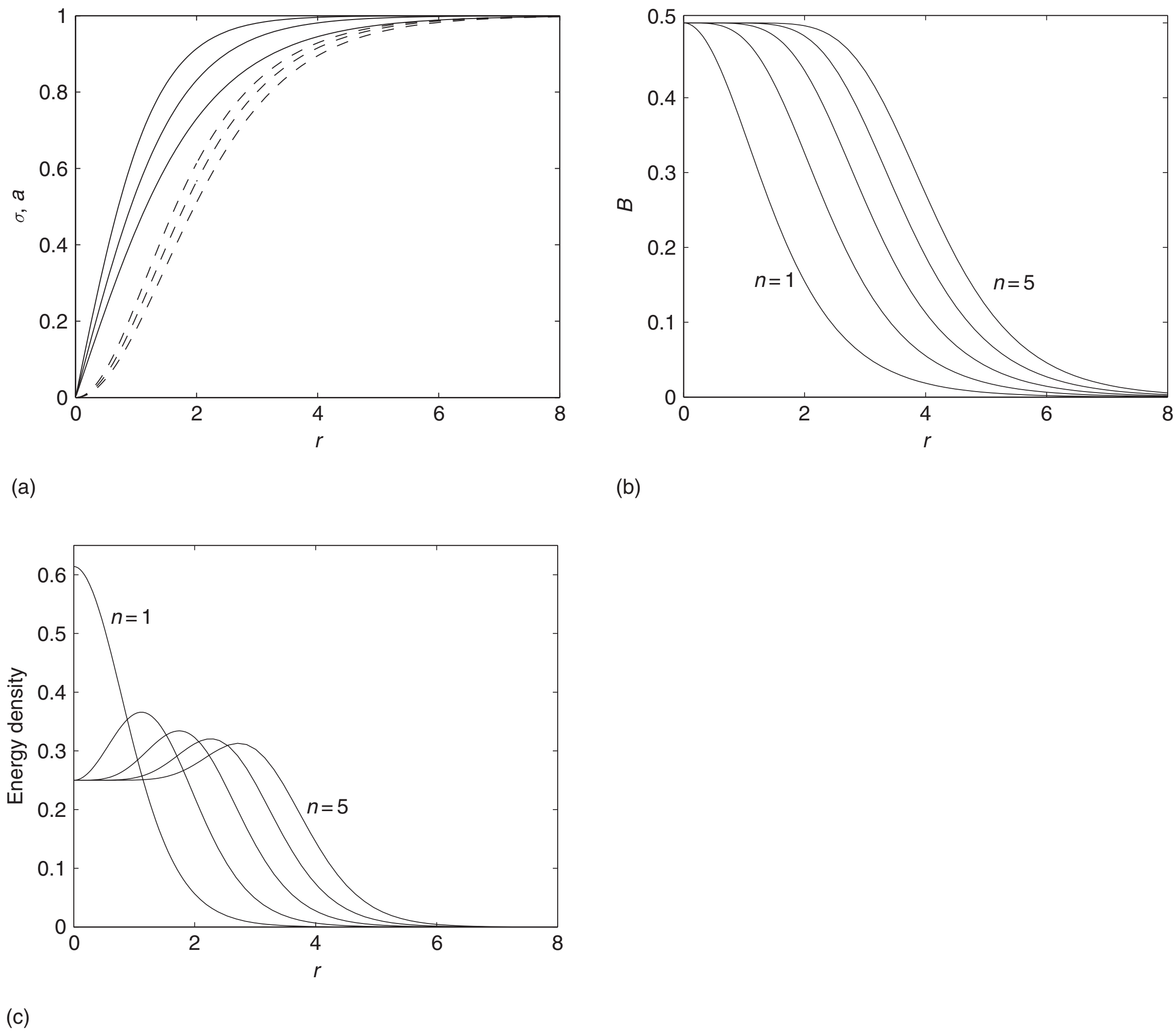



**Figure 1** Static, radially symmetric *n*-vortices: (a) the 1-vortex profile functions $\sigma(r)$ (solid curve) and $a(r)$ (dashed curve) for $\lambda = 2, 1$, and 1/2, left to right; (b) the magnetic field $B$; and (c) the energy density of *n*-vortices, $n = 1$ to 5, left to right, for $\lambda = 1$.

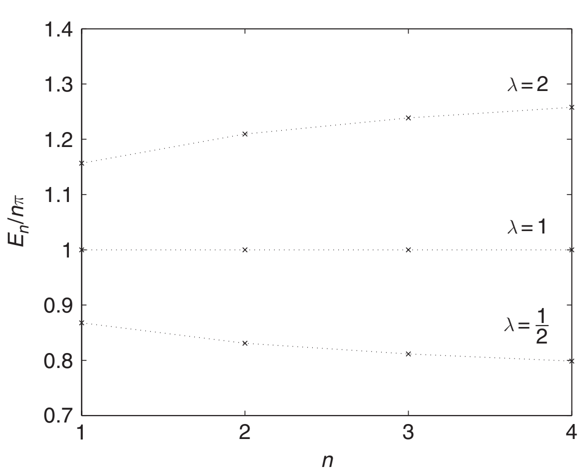



**Figure 2** The energy per unit winding $E_n/n$ of radially symmetric *n*-vortices for $\lambda = 1/2, 1$, and 2.

## Interaction Energy

A precise understanding of the type I/II dichotomy can be obtained using the 2-vortex interaction energy $E_{\text{int}}(s)$ introduced by Jacobs and Rebbi. This is defined to be the minimum of $E$ over all $n = 2$ configurations for which $\phi(\boldsymbol{x}) = 0$ at some pair of points $\boldsymbol{x}_1, \boldsymbol{x}_2$ distance $s$ apart. One interprets $\boldsymbol{x}_1, \boldsymbol{x}_2$ as the vortex positions. $E_{\text{int}}$ can only depend on their separation $s = |\boldsymbol{x}_1 - \boldsymbol{x}_2|$, by translation and rotation invariance. Figure 3 presents graphs of $E_{\text{int}}(s)$ generated by a lattice minimization algorithm. For $\lambda < 1$, vortices uniformly attract one another, so a vortex pair has least energy when coincident. For $\lambda > 1$, vortices uniformly repel, always lowering their energy by moving further apart. The graph for $\lambda = 1$ would be a horizontal line, $E_{\text{int}}(s) = 2\pi$.

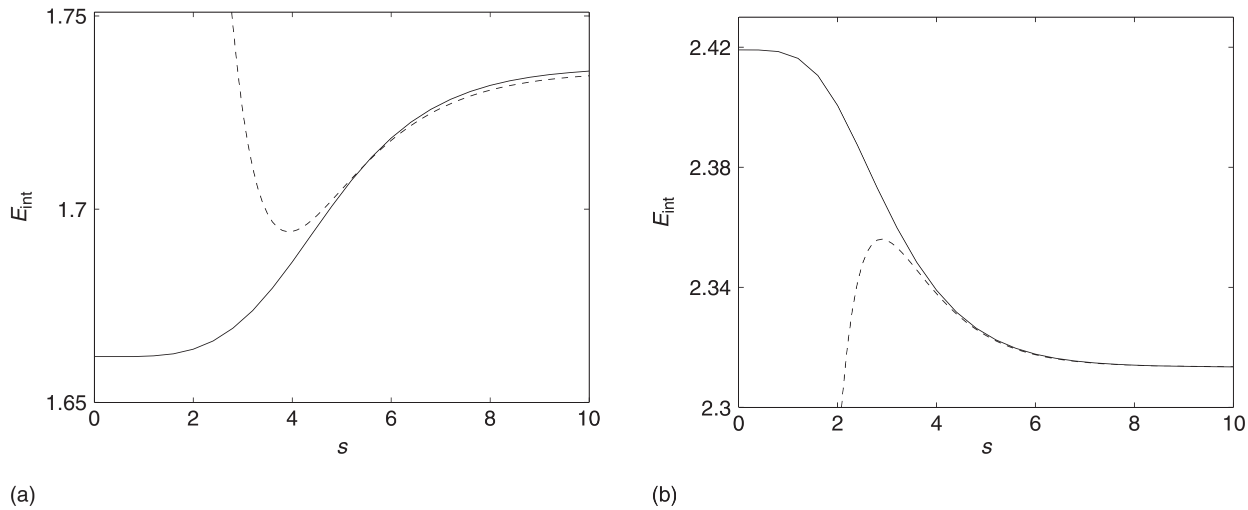



**Figure 3** The 2-vortex interaction energy $E_{\text{int}}(s)$ as a function of vortex separation (solid curve), in comparison with its asymptotic form $E_{\text{int}}^{\infty}(s)$ (dashed curve) for (a) $\lambda = 1/2$ and (b) $\lambda = 2$.

The large $s$ behavior of $E_{\text{int}}(s)$ is known, and can be understood in two ways (Manton and Sutcliffe 2004, pp. 177–181). Speight, adapting ideas of Manton on asymptotic monopole interactions, observed that, in the real $\phi$ gauge ($\phi \mapsto e^{-i\theta}\phi$, $A \mapsto A - d\theta$), the difference between the vortex and the vacuum $\phi = 1, A = 0$ at large $r$,

$$\psi = \phi - 1 \sim \frac{q}{2\pi} K_0(\sqrt{\lambda}\, r) \qquad [11]$$

$$(A_0, \boldsymbol{A}) \sim \frac{m}{2\pi}(0, \boldsymbol{k} \times \nabla K_0(r)) \qquad [12]$$

is identical to the solution of a linear Klein–Gordon–Proca theory,

$$(\partial_\mu \partial^\mu + \lambda)\psi = \kappa, \qquad (\partial_\mu \partial^\mu + 1)A_\nu = j_\nu \qquad [13]$$

in the presence of a composite point source,

$$\kappa = q\delta(\boldsymbol{x}), \qquad (j_0, \boldsymbol{j}) = m(0, \boldsymbol{k} \times \nabla \delta(\boldsymbol{x})) \qquad [14]$$

located at the vortex position. Viewed from afar, therefore, a vortex looks like a point particle carrying both a scalar monopole charge $q$ and a magnetic dipole moment $m$, a "point vortex," inducing a real scalar field of mass $\sqrt{\lambda}$ (the Higgs particle) and a vector boson field of mass 1 (the "photon"). If physics is to be model independent, therefore, the interaction energy of a pair of well-separated vortices should approach that of the corresponding pair of point vortices as the separation grows. Computing the latter is an easy exercise in classical linear field theory, yielding

$$E_{\text{int}}(s) \sim E_{\text{int}}^{\infty}(s) = 2E_1 - \frac{q^2}{2\pi} K_0(\sqrt{\lambda} s) + \frac{m^2}{2\pi} K_0(s) \qquad [15]$$

Bettencourt and Rivers obtained the same formula by a more direct superposition ansatz approach, though they did not give the constants $q$, $m$ a physical interpretation.

The force between a well-separated vortex pair, $-E_{\text{int}}{}'(s)$, consists of the mutual attraction of identical scalar monopoles, of range $1/\sqrt{\lambda}$, and the mutual repulsion of identical magnetic dipoles, of range 1. If $\lambda < 1$, scalar attraction dominates at large $s$ so vortices attract. If $\lambda > 1$, magnetic repulsion dominates and they repel. If $\lambda = 1$ then $q \equiv m$, as we shall see, so the forces cancel exactly. Figure 3 shows both $E_{\text{int}}$ and $E_{\text{int}}^{\infty}$ for $\lambda = 0.5, 2$. The agreement is good for $s$ large, but breaks down for $s < 4$, as one expects. Vortices are not point particles, as in the linear model, and when they lie close together the overlap of their cores produces significant effects.

The same method predicts the interaction energy between an $n_1$-vortex and an $n_2$-vortex at large separation. We just replace $2E_1$ by $E_{n_1} + E_{n_2}$, $q^2$ by $q_{n_1} q_{n_2}$, and $m^2$ by $m_{n_1} m_{n_2}$. In particular, an antivortex ((−1)-vortex) has $E_{-1} = E_1, q_{-1} = q_1 = q$, and $m_{-1} = -m_1 = -m$, so the interaction energy for a vortex–antivortex pair is

$$E_{\text{int}}^{v\bar{v}}(s) \sim 2E_1 - \frac{q^2}{2\pi} K_0(\sqrt{\lambda} r) - \frac{m^2}{2\pi} K_0(r) \qquad [16]$$

which is uniformly attractive. It would be pleasing if $q_n, m_n$ could be deduced easily from $q$, $m$. One might guess $q_n = |n|q, m_n = nm$, in analogy with monopoles. Unfortunately, this is false: $q_n, m_n$ grow approximately exponentially with $|n|$.

## Vortex Scattering

The AHM being Lorentz invariant, one can obtain time-dependent solutions wherein a single $n$-vortex travels at constant velocity, with speed $0 < v < 1$ and $E_{\text{tot}} = (1 - v^2)^{-1/2} E_n$, by Lorentz boosting the static solutions described above. Of more dynamical interest are solutions in which two or more vortices undergo relative motion. The simplest problem is vortex scattering. Two vortices, initially well separated, are propelled towards one another. In the center-of-mass (COM) frame they have, as $t \to -\infty$, equal speed $v$, and approach one another along parallel lines distance $b$ (the impact parameter) apart, see Figure 4. If $b = 0$, they approach head-on. Assuming they do not capture one another, they interact and, as $t \to \infty$, recede along parallel straight lines having been deflected through an angle $\Theta$ (the scattering angle). If scattering is elastic, the exit lines also lie $b$ apart and each vortex travels at speed $v$ as $t \to \infty$. The dependence of $\Theta$ on $v$, $b$, and $\lambda$ has been studied through lattice simulations by several authors, perhaps most comprehensively by Myers, Rebbi, and Strilka (1992). We shall now describe their results.

Note first that vortex scattering is actually inelastic: vortices recede with speed $< v$ because some of their initial kinetic energy is dispersed by the collision as small-amplitude traveling waves ("radiation"). This energy loss can be as high as 80% in very fast collisions at small $b$. At small $v$ the energy loss is tiny, but can still have important consequences for type I vortices: if $v$ is very small, they start with only just enough energy to escape their mutual attraction. In undergoing a small $b$ collision they can lose enough of this energy to become trapped in an oscillating bound state. In this case they do not truly scatter and $\Theta$ is ill-defined. Myers *et al.* find that $v \geq 0.2$ suffices to avoid

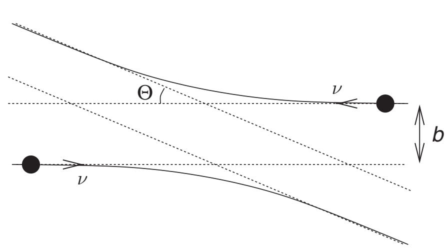



**Figure 4** The geometry of vortex scattering.

capture when $\lambda = 1/2$. Since type I vortices attract, one might expect $\Theta$ to be always negative, indicating that the vortices deflect towards one another. In fact, as Figure 5a shows, this happens only for small $v$ and large $b$. Another naive expectation is that $\Theta = 0$ or $\Theta = 180°$ when $b = 0$ (either vortices pass through one another or ricochet backwards in a head-on collision). In fact $\Theta = 90°$, the only other possibility allowed by reflexion symmetry of the initial data. Figure 6 depicts snapshots of such a scattering process at modest $v$. The vortices deform each other as they get close until, at the moment of coincidence, they are close to the static 2-vortex ring. They then break apart along a line perpendicular to their line of approach. One may consider them to have exchanged half-vortices, so that each emergent vortex is a mixture of the incoming vortices. This rather surprising phenomenon was actually predicted by Ruback in advance of any numerical simulations and turns out to be a generic feature of planar topological solitons.

Consider now the type II case ($\lambda = 2$, Figure 5b). Here, $\Theta > 0$ for all $v$, $b$ as one expects of particles that repel each other. Head-on scattering is more interesting now since two regimes emerge: for $v > v_{\text{crit}} \approx 0.3$, one has the surprising 90° scattering already described, while for $v < v_{\text{crit}}$ the vortices bounce backwards, $\Theta = 180°$. This is easily explained. In order to undergo 90° head-on scattering, the vortices must become coincident (otherwise reflexion symmetry is violated), hence must have initial energy at least $E_2$. For $v < v_{\text{crit}}$, where

$$\frac{2E_1}{\sqrt{1 - {v_{\text{crit}}}^2}} = E_2 \qquad [17]$$

they have too little energy, so come to a halt before coincidence, then recede from one another. The solution $v_{\text{crit}}$ of [17] depends on $\lambda$ and is plotted in Figure 7. For $v$ slightly above $v_{\text{crit}}$, we see that, in contrast to the type I case, $\Theta(b)$ is not monotonic: maximum deflection occurs at nonzero $b$.

The point vortex formalism yields a simple model of type II vortex scattering which is remarkably successful at small $v$. One writes down the Lagrangian for two identical (nonrelativistic) point particles of mass $E_1$ moving along trajectories $\boldsymbol{x}_1(t), \boldsymbol{x}_2(t)$ under the influence of the repulsive potential $E_{\text{int}}^{\infty}$,

$$L = \tfrac{1}{2} E_1 (|\dot{\boldsymbol{x}}_1|^2 + |\dot{\boldsymbol{x}}_2|^2) - E_{\text{int}}^{\infty}(|\boldsymbol{x}_1 - \boldsymbol{x}_2|) \qquad [18]$$

Energy and angular momentum conservation reduce $\Theta(v, b)$ to an integral over one variable ($s = |\boldsymbol{x}_1 - \boldsymbol{x}_2|$) which is easily computed numerically. To illustrate, Figure 5b shows the result for $\lambda = 2, v = 0.1$ in comparison with the lattice simulations of

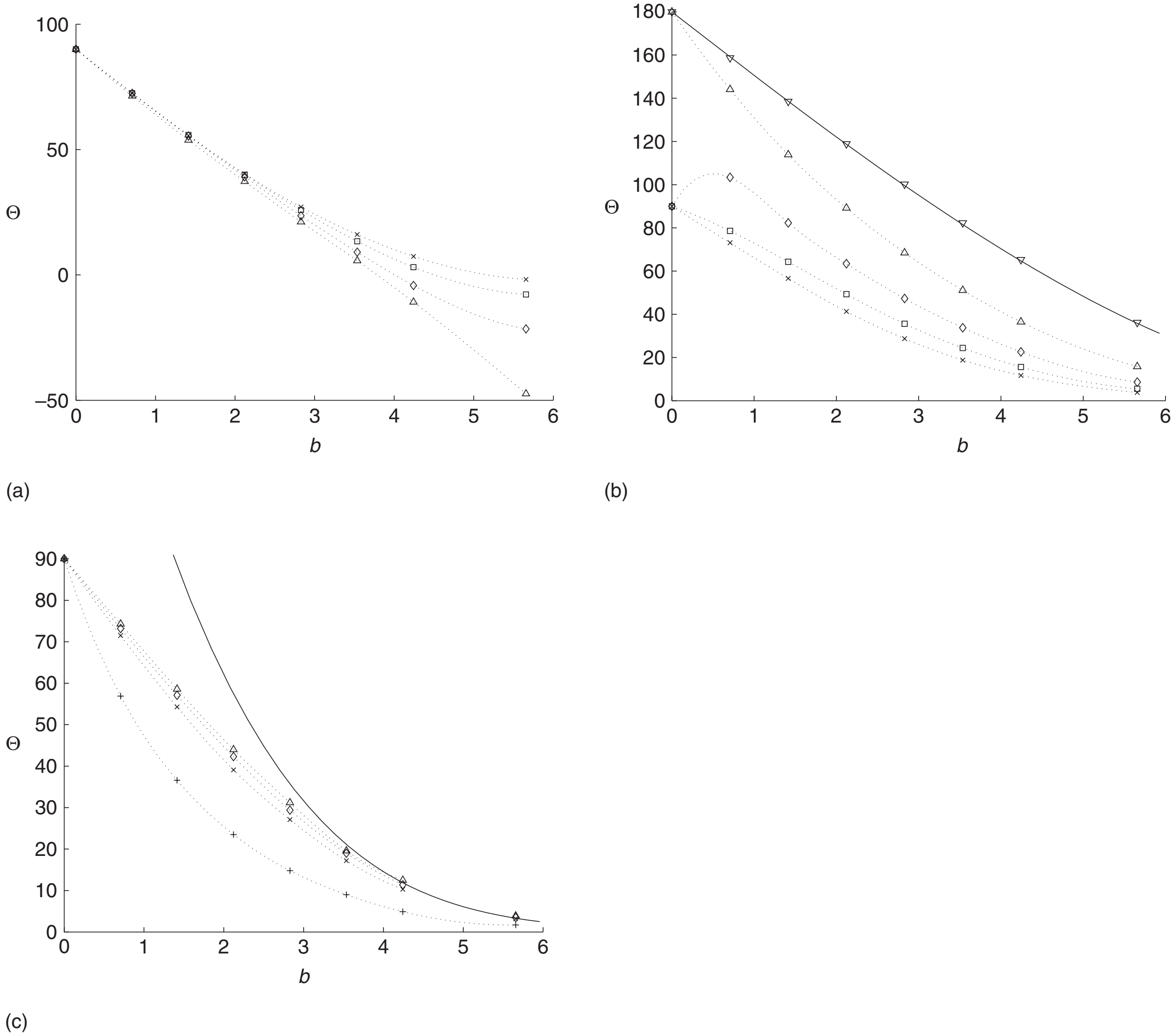



**Figure 5** The 2-vortex scattering angle $\Theta$ as a function of impact parameter $b$ for $v = 0.1$ ($\bigtriangledown$), $v = 0.2$ ($\bigtriangleup$), $v = 0.3$ ($\Diamond$), $v = 0.4$ ($\approx$), $v = 0.5$ ($\times$), and $v = 0.9$ ($+$), as computed by Myers *et al.* (1992): (a) $\lambda = 1/2$; (b) $\lambda = 2$; (c) $\lambda = 1$. The dotted curves are merely guides to the eye. The solid curves in (b), (c) were computed using the point vortex model. Note that Myers *et al.* use different normalizations, so $b = \sqrt{2} b_{MRS}$ and $\lambda = \lambda_{MRS}/2$.

Myers *et al.* The agreement is almost perfect. For large $v$ the approximation breaks down not only because relativistic corrections become significant, but also because small $b$ collisions then probe the small $|\boldsymbol{x}_1 - \boldsymbol{x}_2|$ region where vortex core overlap effects become important. For the same reason, the point vortex model is less useful for type I scattering. Here there is no repulsion to keep the vortices well separated, so its validity is restricted to the small $v$, large $b$ regime.

Critical coupling is theoretically the most interesting regime, where most analytic progress has been made. Since $E_{\text{int}} \equiv E_{\text{int}}^{\infty} \equiv 0$, one might expect vortex scattering to be trivial ($\Theta(v, b) \equiv 0$), but this is quite wrong, as shown in **Figure 5c**. In particular, $\Theta(v, 0) = 90°$ for all $v$, just as in the large $v$ type I and type II cases. The point is that scalar attraction and magnetic repulsion of vortices are mediated by fields with different Lorentz transformation properties. While they cancel for static vortices, there is no reason to expect them to cancel for vortices in relative motion.

## Critical Coupling

The AHM with $\lambda = 1$ has many remarkable properties, at which we have so far only hinted. These all stem from Bogomol'nyi's crucial observation (Manton and Sutcliffe 2004, pp. 197–202) that the potential energy in this case can be rewritten as

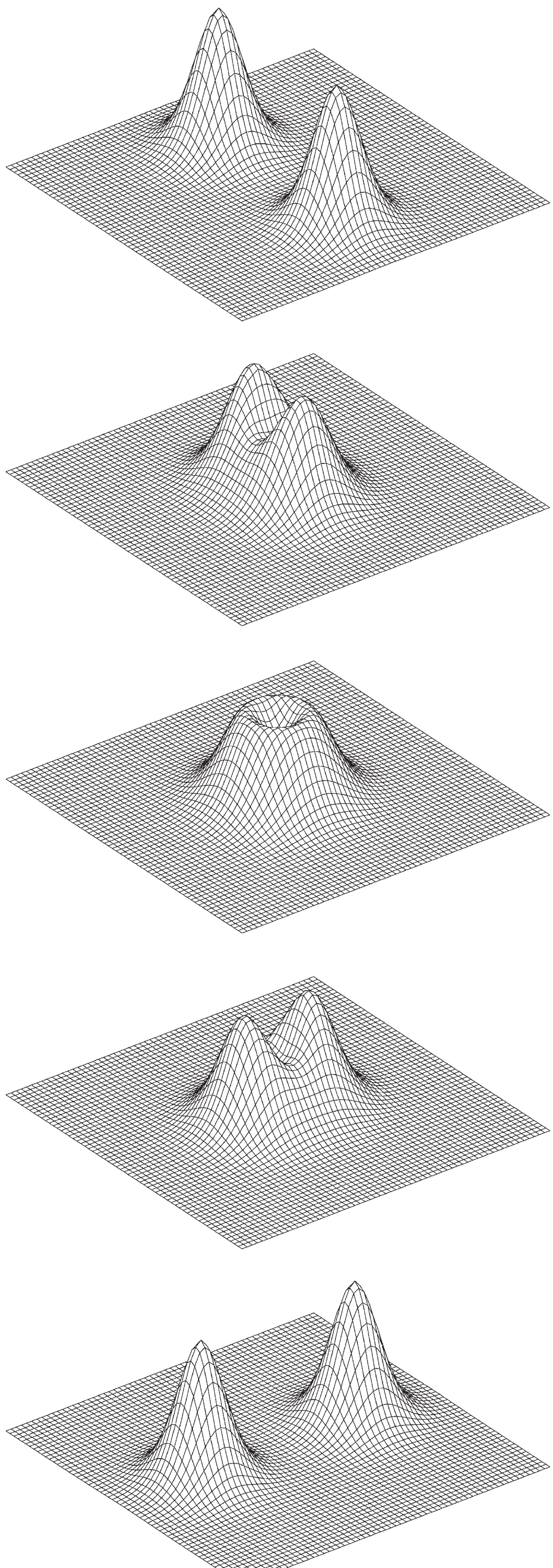


**Figure 6** Snapshots of the energy density during a head-on collision of vortices. This 90° scattering phenomenon is a generic feature of planar topological soliton dynamics.

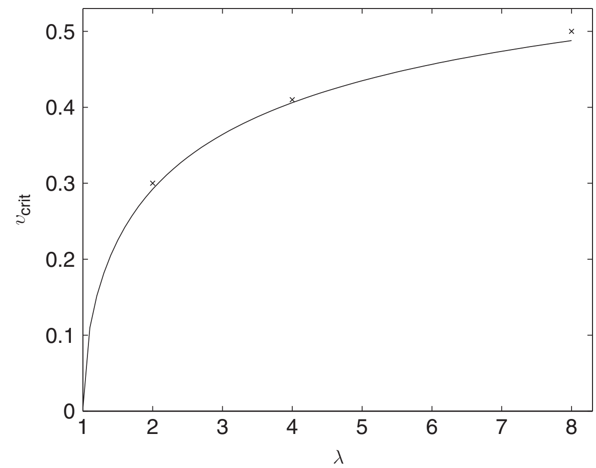



**Figure 7** The critical velocity for 90° head-on scattering of type II vortices $v_{\mathrm{crit}}$ as a function of $\lambda$, as predicted by equation [17] (solid curve), in comparison with the results of Myers *et al.* (1992), (crosses).

$$E = \frac{1}{2}\int\left\{\left(B - \frac{1}{2}(1-|\phi|^2)\right)^2 + |D_1\phi + \mathrm{i}D_2\phi|^2 + B\right\}\mathrm{d}^2\boldsymbol{x} - \mathrm{i}\int_{\mathbb{R}^2}\mathrm{d}(\bar{\phi}D\phi) \qquad [19]$$

The last integral vanishes by Stokes's theorem, so $E \geq \pi n$ by flux quantization [6], and $E = \pi n$ if and only if

$$(D_1 + \mathrm{i}D_2)\phi = 0 \qquad [20]$$

$$\tfrac{1}{2}(1 - |\phi|^2) = B \qquad [21]$$

Note that system [20], [21] is first order, in contrast to the second-order field equations [3]. No explicit solutions of [20], [21] are known. However, Taubes has proved that for each unordered list $[z_1, z_2, \ldots, z_n]$ of $n$ points in $\mathbb{C}$, not necessarily distinct, there exists a solution of [20], [21], unique up to gauge transformations, with $\phi(z_1) = \phi(z_2) = \cdots = \phi(z_n) = 0$ and $\phi$ nonvanishing elsewhere, the zero at $z_r$ having the same multiplicity as $z_r$ has in the list. Note that the list is unordered: a solution is uniquely determined by the positions and multiplicities of the zeroes of $\phi$, but the order in which we label these is irrelevant. The solution minimizes $E$ within the class $\mathcal{C}_n$ of winding $n$ configurations, so is automatically a stable static solution of the model.

Equation [20] applied to the symmetric $n$-vortex, $\phi = \sigma(r)\,\mathrm{e}^{\mathrm{i}n\theta}, A = a(r)\,\mathrm{d}\theta$ implies $a(r) = n - r\sigma'(r)/\sigma(r)$. Comparing with [8], [9], it follows that $q_n = m_n$ when $\lambda = 1$ as previously claimed, since $K_1 = -K_0'$. Tong has conjectured, based on a string duality argument, that $q_1 = -2\pi 8^{1/4}$. This is consistent with current numerics but has no direct derivation so far.

Taubes's theorem shows that this $n$-vortex is just one point, corresponding to the list $[0, 0, \ldots, 0]$, in a $2n$-dimensional space of static multivortex solutions called the moduli space $\mathsf{M}_n$. This space may be visualized as the flat, finite-dimensional valley bottom in $\mathcal{C}_n$ on which $E$ attains its minimum value, $\pi n$. Points in $\mathsf{M}_n$ are in one-to-one correspondence with distinct unordered lists $[z_1, z_2, \ldots, z_n]$, which are themselves in one-to-one correspondence with points in $\mathbb{C}^n$, as follows. To each list, we assign the unique monic polynomial whose roots are $z_r$,

$$\begin{aligned} p(z) &= (z - z_1)(z - z_2) \cdots (z - z_n) \\ &= a_0 + a_1 z + \cdots + a_{n-1} z^{n-1} + z^n \end{aligned} \qquad [22]$$

This polynomial is uniquely determined by its coefficients $(a_0, a_1, \ldots, a_{n-1}) \in \mathbb{C}^n$, which give good global coordinates on $\mathsf{M}_n \cong \mathbb{C}^n$. The zeros $z_r$ of $\phi$ may be used as local coordinates on $\mathsf{M}_n$, away from $\Delta$, the subset of $\mathsf{M}_n$ on which two or more of the zeros $z_r$ coincide, but are not good global coordinates.

Let $(\phi, \boldsymbol{A})_{\boldsymbol{a}}$ denote the static solution corresponding to $\boldsymbol{a} \in \mathbb{C}^n$. If the zeros $z_r$ are all at least $s$ apart, Taubes showed the solution is just a linear superposition of 1-vortices located at $z_r$, up to corrections exponentially small in $s$. Imagine these constituent vortices are pushed with small initial velocities. Then $(\phi(t), \boldsymbol{A}(t))$ must remain close to the valley bottom $\mathsf{M}_n$, since departing from it costs kinetic energy, of which there is little. Manton has suggested, therefore, that the dynamics is well approximated by the constrained variational problem wherein $(\phi(t), \boldsymbol{A}(t)) = (\phi, \boldsymbol{A})_{\boldsymbol{a}(t)} \in \mathsf{M}_n$ for all $t$. Since the action $S = \int \mathcal{L}\, d^3x = \int (E_{\text{kin}} - E)\, dt$, and $E = \pi n$, constant, on $\mathsf{M}_n$, this constrained problem amounts to Lagrangian mechanics on configuration space $\mathsf{M}_n$ with Lagrangian $L = E_{\text{kin}}|_{\mathsf{M}_n}$. Now $E_{\text{kin}}$ is real, positive, and quadratic in time derivatives of $\phi$, $\boldsymbol{A}$, so

$$L = \tfrac{1}{2} \gamma_{rs}(\boldsymbol{a}) \dot{a}_r \dot{\bar{a}}_s \qquad [23]$$

$\gamma_{rs}$ forming the entries of a positive-definite $n \times n$ Hermitian matrix ($\gamma_{sr} \equiv \overline{\gamma_{rs}}$). Since $(\phi, \boldsymbol{A})_{\boldsymbol{a}}$ is not known explicitly, neither are $\gamma_{rs}(\boldsymbol{a})$. Observe, however, that $L$ is the Lagrangian for geodesic motion in $\mathsf{M}_n$ with respect to the Riemannian metric

$$\gamma = \gamma_{rs}(\boldsymbol{a})\, da_r\, d\bar{a}_s \qquad [24]$$

Manton originally proposed this geodesic approximation for monopoles, but it is now standard for all topological solitons of Bogomol'nyi type (where one has a moduli space of static multisolitons saturating a topological lower bound on $E$). Note that geodesics are independent of initial speed, which agrees with Myers *et al*: **Figure 5c** shows that $\Theta(v, b)$ is approximately independent of $v$ for $v \leq 0.5$. Further, Stuart (1994) has proved that, for initial speeds of order $\epsilon$, small, the fields stay (pointwise) $\epsilon^2$ close to their geodesic approximant for times of order $\epsilon^{-1}$.

On symmetry grounds, two vortex dynamics in the COM frame reduces to geodesic motion in $\mathsf{M}_2^0 \cong \mathbb{C}$, the subspace of centered 2-vortices ($a_1 = 0$, so $z_1 = -z_2$), with induced metric

$$\gamma^0 = G(|a_0|)\, da_0\, d\bar{a}_0 \qquad [25]$$

$G$ being some positive function. Note that $a_0 = z_1 z_2$, so the intervortex distance $|z_1 - z_2| = 2|z_1| = 2|a_0|^{1/2}$. The line $a_0 = \beta \in \mathbb{R}$, traversed with $\beta$ increasing, say, is geodesic in $\mathsf{M}_2^0$. The vortex positions (roots of $z^2 + a_0$) are $\pm\sqrt{|\beta|}$ for $\beta \leq 0$ and $\pm i\sqrt{\beta}$ for $\beta > 0$. This describes perfectly the 90° scattering phenomenon: two vortices approach head-on along the $x^1$ axis, coincide to form a 2-vortex ring, then break apart along the $x^2$ axis, as in **Figure 6**. This behavior occurs because $a_0 = z_1 z_2$, rather than $z_1 - z_2$, is the correct global coordinate on $\mathsf{M}_2^0$, since vortices are classically indistinguishable.

Samols found a useful formula (Manton and Sutcliffe 2004, pp. 205–215) for $\gamma$ in terms of the behavior of $|\phi_{\boldsymbol{a}}|$ close to its zeros, using which he devised an efficient numerical scheme to evaluate $G(|a_0|)$, and computed $\Theta(b)$ in detail, finding excellent agreement with lattice simulations at low speeds. He also studied the quantum scattering of vortices, approximating the quantum state by a wave function $\Psi$ on $\mathsf{M}_n$ evolving according to the natural Schrödinger equation for quantum geodesic motion,

$$i\hbar \frac{\partial \Psi}{\partial t} = -\tfrac{1}{2} \hbar^2 \Delta_\gamma \Psi \qquad [26]$$

where $\Delta_\gamma$ is the Laplace–Beltrami operator on $(\mathsf{M}_n, \gamma)$. This technique, introduced for monopoles by Gibbons and Manton, is now standard for solitons of Bogomol'nyi type.

By analyzing the forces between moving point vortices at $\lambda = 1$, Manton and Speight (2003) showed that, as the vortex separations become uniformly large, the metric on $\mathsf{M}_n$ approaches

$$\begin{aligned} \gamma^\infty = \pi \sum_r \Bigg[ & dz_r\, d\bar{z}_r - \frac{q^2}{4\pi} \sum_{s \neq r} K_0(|z_r - z_s|) \\ & \times (dz_r - dz_s)(d\bar{z}_r - d\bar{z}_s) \Bigg] \end{aligned} \qquad [27]$$

This formula can also be obtained by a method of matched asymptotic expansions. We can use [27] to study 2-vortex scattering for large $b$, when the

vortices remain well separated. (Note that $\gamma^\infty$ is not positive definite if any $|z_r - z_s|$ becomes too small.) The results are good, provided $v \le 0.5$ and $b \ge 3$ (see Figure 5c).

## Other Developments

The (critically coupled) AHM on a compact physical space $\Sigma$ is of considerable theoretical and physical interest. Bradlow showed that $\mathsf{M}_n(\Sigma)$ is empty unless $V = \text{Area}(\Sigma) \ge 4\pi n$, so there is a limit to how many vortices a space of finite area can accommodate (Manton and Sutcliffe 2004, pp. 227–230). Manton has analyzed the thermodynamics of a gas of vortices by studying the statistical mechanics of geodesic flow on $\mathsf{M}_n(\Sigma)$. In this context, spatial compactness is a technical device to allow nonzero vortex density $n/V$ for finite $n$, without confining the fields to a finite box, which would destroy the Bogomol'nyi properties. In the limit of interest, $n, V \to \infty$ with $n/V$ fixed, the thermodynamical properties turn out to depend on $\Sigma$ only through $V$, so $\Sigma = S^2$ and $\Sigma = T^2$ give equivalent results, for example. The equation of state of the gas is ($P$ = pressure, $T$ = temperature)

$$P = \frac{nT}{V - 4\pi n} \qquad [28]$$

which is similar, at low density $n/V$, to that of a gas of hard disks of area $2\pi$. The crucial step in deriving [28] is to find the volume of $\mathsf{M}_n(\Sigma)$ which, despite there being no formula for $\gamma$, may be computed exactly by remarkable indirect arguments (Manton and Sutcliffe 2004, pp. 231–234).

The static AHM coincides with the Ginzburg–Landau model of superconductivity, which has precisely the same type I/II classification. Here the "Higgs" field represents the wave function of a condensate of Cooper pairs, usually (but not always) electrons. There has been a parallel development of the static model by condensed matter theorists, therefore; see Fossheim and Sudbo (2004), for example. In fact the vortex was actually first discovered by Abrikosov in the condensed matter context. One important difference is that type I superconductors do not support vortex solutions in an external magnetic field $B_{\text{ext}}$ because the critical $|B_{\text{ext}}|$ required to create a single vortex is greater than the critical $|B_{\text{ext}}|$ required to destroy the condensate completely ($\phi \equiv 0$). Type II superconductors do support vortices, and there are such superconductors with $\lambda \approx 1$, but the vortex dynamics we have described is not relevant to these systems. In this context there is an obvious preferred reference frame (the rest frame of the superconductor) so it is unsurprising that the Lorentz-invariant AHM is inappropriate. Insofar as vortices move at all, they seem to obey a first-order (in time) dynamical system, in contrast to the second-order AHM. Manton has devised a first-order system which may have relevance to superconductivity, by replacing $E_{\text{kin}}$ with a Chern–Simons–Schrödinger functional (Manton and Sutcliffe 2004, pp. 193–197). Rather than attracting or repelling, vortices now tend to orbit one another at constant separation. There is again a moduli space approximation to slow vortex dynamics for $\lambda \approx 1$, but it has a Hamiltonian-mechanical rather than Riemannian-geometric flavor.

Finally, an interesting simplification of the AHM, which arises, for example, as a phenomenological model of liquid helium-4, is obtained if we discard the gauge field $A_\mu$, or equivalently set the electric charge of $\phi$ to $e = 0$. There is now no type I/II classification, since $\lambda$ may be absorbed by rescaling. The resulting model, which has only global U(1) phase symmetry, supports $n$-vortices $\phi = \sigma(r)e^{in\theta}$ for all $n$, but these are not exponentially spatially localized,

$$\sigma(r) = 1 - \frac{n^2}{\lambda r^2} - \frac{n^2(8 + n^2)}{2\lambda^2 r^4} + O(r^{-6}) \qquad [29]$$

and cannot have finite $E$ by Derrick's theorem. They are unstable for $|n| > 1$, and 1-vortices uniformly repel one another. They can be given an interesting first-order dynamics (the Gross–Pitaevski equation).

## Abbreviations

| | |
|---|---|
| $A_\mu$ | electromagnetic gauge potential |
| $b$ | impact parameter |
| $D_\mu$ | gauge-covariant derivative |
| $E$ | potential energy |
| $E_{\text{kin}}$ | kinetic energy |
| $F_{\mu\nu}$ | electromagnetic field strength tensor |
| $L$ | Lagrangian |
| $\mathcal{L}$ | Lagrangian density |
| $S$ | action |
| $\phi$ | Higgs field |
| $\Theta$ | scattering angle |

*See also:* Fractional Quantum Hall Effect; Ginzburg–Landau Equation; High $T_c$ Superconductor Theory; Integrable Systems: Overview; Nonperturbative and Topological Aspects of Gauge Theory; Quantum Fields with Topological Defects; Solitons and Other Extended Field Configurations; Symmetry Breaking in Field Theory; Topological Defects and Their Homotopy Classification; Variational Techniques for Ginzburg–Landau Energies.

## Further Reading


Atiyah M and Hitchin N (1988) *The Geometry and Dynamics of Magnetic Monopoles*. Princeton: Princeton University Press.
Fossheim K and Sudbo A (2004) *Superconductivity: Physics and Applications*. Hoboken NJ: Wiley.
Jaffe A and Taubes C (1980) *Vortices and Monopoles: Structure of Static Gauge Theories*. Boston: Birkhäuser.
Nakahara M (1990) *Geometry, Topology and Physics*. Bristol: Adam-Hilger.
Manton NS and Speight JM (2003) Asymptotic interactions of critically coupled vortices. *Communications in Mathematical Physics* 236: 535–555.
Manton NS and Sutcliffe PM (2004) *Topological Solitons*. Cambridge: Cambridge University Press.
Myers E, Rebbi C, and Strilka R (1992) Study of the interaction and scattering of vortices in the abelian Higgs (or Ginzburg-Landau) model. *Physical Review* 45: 1355–1364.
Rajaraman R (1989) *Solitons and Instantons*. Amsterdam: North-Holland.
Stuart D (1994) Dynamics of abelian Higgs vortices in the near Bogomolny regime. *Communications in Mathematical Physics* 159: 51–91.
Vilenkin A and Shellard EPS (1994) *Cosmic Strings and Other Topological Defects*. Cambridge: Cambridge University Press.


# Adiabatic Piston


**Ch Gruber**, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
**A Lesne**, Université P.-M. Curie, Paris VI, Paris, France




## Introduction

### Macroscopic Problem

The "adiabatic piston" is an old problem of thermodynamics which has had a long and controversial history. It is the simplest example concerning the time evolution of an adiabatic wall, that is, a wall which does not conduct heat. The system consists of a gas in a cylinder divided by an adiabatic wall (the piston). Initially, the piston is held fixed by a clamp and the two gases are in thermal equilibrium characterized by $(p^\pm, T^\pm, N^\pm)$, where the index $-/+$ refers to the gas on the left/right side of the piston and $(p, T, N)$ denote the pressure, the temperature, and the number of particles (Figure 1). Since the piston is adiabatic, the whole system remains in equilibrium even if $T^- \neq T^+$. At time $t = 0$, the clamp is removed and the piston is let free to move without any friction in the cylinder. The question is to find the final state, that is, the final position $X_\mathrm{f}$ of the piston and the parameters $(p_\mathrm{f}^\pm, T_\mathrm{f}^\pm)$ of the gases.

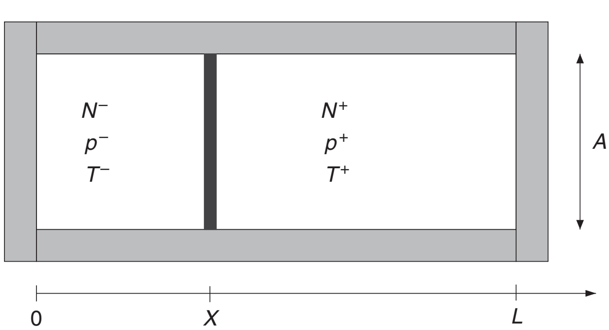



**Figure 1** The adiabatic piston problem.

In the late 1950s, using the two laws of equilibrium thermodynamics (i.e., thermostatics), Landau and Lifshitz concluded that the adiabatic piston will evolve toward a final state where $p^-/T^- = p^+/T^+$. Later, Callen (1963) and others realized that the maximum entropy condition implies that the system will reach mechanical equilibrium where the pressures are equal $p_\mathrm{f}^- = p_\mathrm{f}^+$; however, nothing could be said concerning the final position $X_\mathrm{f}$ or the final temperatures $T_\mathrm{f}^\pm$ which should depend explicitly on the viscosity of the fluids. It thus became a controversial problem since one was forced to accept that the two laws of thermostatics are not sufficient to predict the final state as soon as adiabatic movable walls are involved (see early references in Gruber (1999)).

Experimentally, the adiabatic piston was used already before 1924 to measure the ratio $c_p/c_v$ of the specific heats of gases. In 2000, new measurements have shown that one has to distinguish between two regimes, corresponding to weak damping or strong damping, with very different properties, for example, for weak damping the frequency of oscillations corresponds to adiabatic oscillations, whereas for strong damping it corresponds to isothermal oscillations.

### Microscopic Problem

The "adiabatic piston" was first considered from a microscopic point of view by Lebowitz who introduced in 1959 a simple model to study heat conduction. In this model, the gas consists of point particles of mass $m$ making purely elastic collisions on the wall of the cylinder and on the piston. Furthermore, the gas is very dilute so that the

equation of state $p = nk_BT$ is satisfied at equilibrium, where $n$ is the density of particles in the gas and $k_B$ the Boltzmann constant. The adiabatic piston is taken as a heavy particle of mass $M \gg m$ without any internal degree of freedom. Using this same model Feynman (1965) gave a qualitative analysis in *Lectures in Physics*. He argued intuitively but correctly that the system should converge first toward a state of mechanical equilibrium where $p^- = p^+$ and then very slowly toward thermal equilibrium. This approach toward thermal equilibrium is associated with the "wiggles" of the piston induced by the random collisions with the atoms of the gas. Of course, this stochastic behavior is not part of thermodynamics and the evolution beyond the mechanical equilibrium cannot appear in the macroscopical framework assuming that the piston does not conduct heat.

From a microscopical point of view, one is confronted with two different problems: the approach toward mechanical equilibrium in the absence of any *a priori* friction (where the entropy of both gases should increase) and, on a different timescale, the approach toward thermal equilibrium (where the entropy of one gas should decrease but the total entropy increase).

The conceptual difficulties of the problem beyond mechanical equilibrium come from the following intuitive reasoning. When the piston moves toward the hotter gas, the atoms of the hotter gas gain energy, whereas those of the cooler gas lose energy. When the piston moves toward the cooler side, it is the opposite. Since on an average the hotter side should cool down and the cold side should warm up, we are led to conclude that on an average the piston should move toward the colder side. On the other hand, from $p = nk_BT$, the piston should move toward the warmer side to maintain pressure balance.

In 1996, Crosignani, Di Porto, and Segev introduced a kinetic model to obtain equations describing the adiabatic approach toward mechanical equilibrium. Starting with the microscopical model introduced by Lebowitz, Gruber, Piasecki, and Frachebourg, later joined by Lesne and Pache, initiated in 1998 a systematic investigation of the adiabatic piston within the framework of statistical mechanics, together with a large number of numerical simulations. This analysis was based on the fact that $m/M$ is a very small parameter to investigate expansions in powers of $m/M$ (see Gruber and Piasecki (1999) and Gruber *et al.* (2003) and reference therein). An approach using dynamical system methods was then developed by Lebowitz *et al.* (2000) and Chernov *et al.* (2002). An extension to hard-disk particles was analyzed at the same time by Kestemont *et al.* (2000). Recently, several other authors have contributed to this subject.

The general picture which emerges from all the investigations is the following. For an infinite cylinder, starting with mechanical equilibrium $p^- = p^+ = p$, the piston evolves to a stationary stochastic state with nonzero velocity toward the warmer side

$$\langle V \rangle = \frac{m}{M}\sqrt{\frac{\pi k_B}{8m}}\left(\sqrt{T^+} - \sqrt{T^-}\right) + o\left(\frac{m}{M}\right) \qquad [1]$$

with relaxation time

$$\tau = \frac{M}{A}\sqrt{\frac{\pi k_B}{8m}}\frac{1}{p}\left(\frac{1}{\sqrt{T^-}} + \frac{1}{\sqrt{T^+}}\right)^{-1} \qquad [2]$$

where $M/A$ is the mass per unit area of the piston. In this state the piston has a temperature $T_P = \sqrt{T^+T^-}$ and there is a heat flux

$$j_Q = \left(\sqrt{T^-} - \sqrt{T^+}\right)\frac{m}{M}\sqrt{\frac{8k_B}{m\pi}}p + o\left(\frac{m}{M}\right) \quad (p^- = p^+ = p) \qquad [3]$$

For a finite cylinder and $p^+ \neq p^-$, the evolution proceeds in four different stages. The first two are deterministic and adiabatic. They correspond to the thermodynamic evolution of the (macroscopic) adiabatic piston. The last two stages, which go beyond thermodynamics, are stochastic with heat transfer across the piston. More precisely:

1. In the first stage whose duration is the time needed for the shock wave to bounce back on the piston, the evolution corresponds to the case of the infinite cylinder (with $p^- \neq p^+$). If $R = Nm/M > 10$, the piston will be able to reach and maintain a constant velocity

$$V = (p^- - p^+)\sqrt{\frac{\pi k_B}{8m}}\frac{\sqrt{T^-T^+}}{p^+\sqrt{T^-} + p^-\sqrt{T^+}} + \mathcal{O}\left(\frac{m}{M}\right) \quad \text{for } |p^- - p^+| \ll 1 \qquad [4]$$

2. In the second stage the evolution toward mechanical equilibrium is either weakly or strongly damped depending on $R$. If $R < 1$, the evolution is very weakly damped, the dynamics takes place on a timescale $t' = \sqrt{R}t$, and the effect of the collisions on the piston is to introduce an external potential $\phi(X) = c_1/X^2 + c_2/(L - X)^2$. On the other hand, if $R > 4$, the evolution is strongly damped (with two oscillations only) and depends neither on $M$ nor on $R$.

3. After mechanical equilibrium has been reached, the third stage is a stochastic approach toward thermal equilibrium associated with heat transfer across the piston. This evolution is very slow and exhibits a scaling property with respect to $t' = mt/M$.
4. After thermal equilibrium has been reached ($T^- = T^+$, $p^- = p^+$), in a fourth stage the gas will evolve very slowly toward a state with Maxwellian distribution of velocities, induced by the collision with the stochastic piston.

The general conclusion is thus that a wall which is adiabatic when fixed will become a heat conductor under a stochastic motion. However, it should be stressed that the time required to reach thermal equilibrium will be several orders of magnitude larger than the age of the universe for a macroscopical piston and such a wall could not reasonably be called a heat conductor. However, for mesoscopic systems, the effect of stochasticity may lead to very interesting properties, as shown by Van den Broeck *et al.* (2004) in their investigations of Brownian (or biological) motors.

## Microscopical Model

The system consists of two fluids separated by an "adiabatic" piston inside a cylinder with $x$-axis, length $L$, and area $A$. The fluids are made of $N^{\pm}$ identical light particles of mass $m$. The piston is a heavy flat disk, without any internal degree of freedom, of mass $M \gg m$, orthogonal to the $x$-axis, and velocity parallel to this $x$-axis. If the piston is fixed at some position $X_0$, and if the two fluids are in thermal equilibrium characterized by $(p_0^{\pm}, T_0^{\pm}, N^{\pm})$, then they will remain in equilibrium forever even if $T_0^+ \neq T_0^-$: it is thus an "adiabatic piston" in the sense of thermodynamics. At a certain time $t = 0$, the piston is let free to move and the problem is to study the time evolution. To define the dynamics, we consider that the system is purely Hamiltonian, that is, the particles and the piston move without any friction according to the laws of mechanics. In particular, the collisions between the particles and the walls of the cylinder, or the piston, are purely elastic and the total energy of the system is conserved. In most studies, one considers that the particles are point particles making purely elastic collisions. Since the piston is bound to move only in the $x$-direction, the velocity components of the particles in the transverse directions play no role in this problem. Moreover, since there is no coupling between the components in the $x$- and transverse directions, one can simplify the model further by assuming that all probability distributions are independent of the transverse coordinates. We are thus led to a formally one-dimensional problem (except for normalizations). Therefore, in this review, we consider that the particles are noninteracting and all velocities are parallel to the $x$-axis. From the collision law, if $v$ and $V$ denote the velocities of a particle and the piston before a collision, then under the collision on the piston:

$$\begin{aligned} v \to v' &= 2V - v + \alpha(v - V) \\ V \to V' &= V + \alpha(v - V) \end{aligned} \qquad [5]$$

where

$$\alpha = \frac{2m}{M + m} \qquad [6]$$

Similarly, under a collision of a particle with the boundary at $x = 0$ or $x = L$:

$$v \to v' = -v \qquad [7]$$

Let us mention that more general models have also been considered, for example, the case where the two fluids are made of point particles with different masses $m^{\pm}$, or two-dimensional models where the particles are hard disks. However, no significant differences appear in these more general models and we restrict this article to the simplest case.

One can study different situations: $L = \infty$, $L$ finite, and $L \to \infty$. Furthermore, taking first $M$ and $A$ finite, one can investigate several limits.

1. Thermodynamic limit for the piston only. In this limit, $L$ is fixed (finite or infinite) and $A \to \infty, M \to \infty$, keeping constant the initial densities $n^{\pm}$ of the fluid and the parameter

$$\gamma = \frac{2mA}{M + m} = \alpha A \sim 2m\,\frac{A}{M} \qquad [8]$$

   If $L$ is finite, this means that $N^{\pm} \to \infty$ while keeping constant the parameters

$$R^{\pm} = \frac{mN^{\pm}}{M} = \frac{M_{\text{gas}}^{\pm}}{M} \qquad [9]$$

2. Thermodynamic limit for the whole system, where $L \to \infty$ and $A \sim L^2, N^{\pm} \sim L^3$. In this limit, space and time variables are rescaled according to $x' = x/L$ and $t' = t/L$. This limit can be considered as a limiting case of (1) where $R^{\pm} \sim \sqrt{A} \to \infty$ (and time is scaled).
3. Continuum limit where $L$ and $M$ are fixed and $N^{\pm} \to \infty, m \to 0$ keeping $M_{\text{gas}}^{\pm}$ constant, that is, $R^{\pm} = cte$.

The case $L$ infinite and the limit (1) have been investigated using statistical mechanics (Liouville or

Boltzmann's equations). On the other hand, the limit (2) has been studied using dynamical system methods, reducing first the system to a billiard in an $(N^+ + N^- + 1)$-dimensional polyhedron. The limit (3) has been introduced to derive hydrodynamical equations for the fluids.

In this article, we present the approach based on statistical mechanics. Although not as rigorous as (2) on a mathematical level, it yields more informations on the approach toward mechanical and thermal equilibrium. Moreover, it indicates what are the open problems which should be mathematically solved. In all investigations, advantage is taken of the fact that $m/M$ is very small and one introduces the small parameter

$$\epsilon = \sqrt{m/M} \ll 1 \qquad [10]$$

Let us note that $\epsilon$ measures the ratio of thermal velocities for the piston and a fluid particle, whereas $\alpha \sim \epsilon^2$ measures the ratio of velocity changes during a collision.

## Starting Point: Exact Equations

Using the statistical point of view, the time evolution is given by Liouville's equation for the probability distribution on the whole phase space for $(N^+ + N^- + 1)$ particles, with $L, A, N^{\pm}$, and $M$ finite. Initially $(t \le 0)$, the piston is fixed at $(X_0, V_0 = 0)$ and the fluids are in thermal equilibrium with homogeneous densities $n_0^{\pm}$, velocity distributions $\varphi_0^{\pm}(v) = \varphi_0^{\pm}(-v)$, and temperatures

$$T_0^{\pm} = m \int_{-\infty}^{\infty} \mathrm{d}v \, n_0^{\pm} \varphi_0^{\pm}(v) v^2 \qquad [11]$$

Integrating out the irrelevant degrees of freedom, the Liouville's equation yields the equations for the distribution $\rho^{\pm}(x, v; t)$ of the right and left particles:

$$\partial_t \rho^{\pm}(x, v; t) + v \partial_x \rho^{\pm}(x, v; t) = I^{\pm}(x, v; t) \qquad [12]$$

The collision term $I^{\pm}(x, v; t)$ is a functional of $\rho_{\pm,P}(X, v; X, V; t)$, the two-point correlation function for a right (resp. left) particle at $(x = X, v)$ and the piston at $(X, V)$. Similarly, one obtains for the velocity distribution of the piston:

$$\begin{aligned} \partial_t \Phi(V; t) = & A \int_{-\infty}^{\infty} (V - v)\big[\theta(V - v)\rho^-_{\mathrm{surf}}(v'; V'; t) \\ & + \theta(v - V)\rho^-_{\mathrm{surf}}(v; V; t)\big]\mathrm{d}v \\ & - A \int_{-\infty}^{\infty} (V - v)\big[\theta(v - V)\rho^+_{\mathrm{surf}}(v'; V'; t) \\ & + \theta(V - v)\rho^+_{\mathrm{surf}}(v; V; t)\big]\mathrm{d}v \end{aligned} \qquad [13]$$

where $(v', V')$ are given by eqn [5] and

$$\rho^{\pm}_{\mathrm{surf}}(v; V; t) = \int_{-\infty}^{\infty} \mathrm{d}X \rho_{\pm,P}(X, v; X, V; t) \qquad [14]$$

We thus have to solve eqns [12]–[13] with initial conditions

$$\begin{aligned} \rho^-(x, v; t = 0) &= n_0^- \varphi_0^-(v)\theta(x)\theta(X_0 - x) \\ \rho^+(x, v; t = 0) &= n_0^+ \varphi_0^+(v)\theta(L - x)\theta(x - X_0) \\ \Phi(V; t = 0) &= \delta(V) \end{aligned} \qquad [15]$$

Using the fact that $\alpha = 2m/(M + m) \ll 1$, we can rewrite eqn [13] as a formal series in powers of $\alpha$:

$$\partial_t \Phi(V; t) = \gamma \sum_{k=1}^{\infty} \frac{(-1)^k \alpha^{k-1}}{k!} \left(\frac{\partial}{\partial V}\right)^k \tilde{F}_{k+1}(V; t) \qquad [16]$$

$$\begin{aligned} \tilde{F}_k(V; t) = & \int_V^{\infty} (v - V)^k \rho^-_{\mathrm{surf}}(v; V; t)\mathrm{d}v \\ & - \int_{-\infty}^{V} (v - V)^k \rho^+_{\mathrm{surf}}(v; V; t)\mathrm{d}v \end{aligned} \qquad [17]$$

from which one obtains the equations for the moments of the piston velocity:

$$\frac{1}{\gamma}\frac{\mathrm{d}\langle V^n \rangle}{\mathrm{d}t} = \sum_{k=1}^{n} \alpha^{k-1} \frac{n!}{k!(n-k)!} \int_{-\infty}^{\infty} \mathrm{d}V \, V^{n-k} \tilde{F}_{k+1}(V; t) \qquad [18]$$

However, we do not know the two-point correlation functions.

If the length of the cylinder is infinite, the condition $M \gg m$ implies that the probability for a particle to make more than one collision on the piston is negligible. Alternatively, one could choose initial distributions $\varphi_0^{\pm}(v)$ which are zero for $|v| < v_{\min}$, where $v_{\min}$ is taken such that the probability of a recollision is strictly zero. Therefore, if $L = \infty$, one can consider that before a collision on the piston the particles are distributed with $\varphi_0^{\pm}(v)$ for all $t$, and the two-point correlation functions factorize, that is,

$$\begin{aligned} \rho^-_{\mathrm{surf}}(v; V; t) &= \rho^-_{\mathrm{surf}}(v; t)\Phi(V; t), \quad \text{if } v > V \\ \rho^+_{\mathrm{surf}}(v; V; t) &= \rho^+_{\mathrm{surf}}(v; t)\Phi(V; t), \quad \text{if } v < V \end{aligned} \qquad [19]$$

where for $L = \infty$, $\rho^{\pm}_{\mathrm{surf}}(v; t) = n_0^{\pm}\varphi_0^{\pm}(v)$ and thus the conditions to obtain eqn [18] are satisfied.

If $L$ is finite, one can show that the factorization property (eqn [19]) is an exact relation in the thermodynamic limit for the piston $(A \to \infty, M/A = cte)$. For finite $L$ and finite $A$, we introduce

**Assumption 1** (Factorization condition). Before a collision the two-point correlation functions have the factorization property (eqn [19]) to first order in $\alpha$.

Under the factorization condition, we have

$$\tilde{F}_k(V;t) = F_k(V;t)\Phi(V;t) \qquad [20]$$

with

$$\begin{aligned} F_k(V;t) &= \int_V^{\infty} \mathrm{d}v(v-V)^k \rho^{-}_{\mathrm{surf}}(v;t) \\ &\quad - \int_{-\infty}^{V} \mathrm{d}v(v-V)^k \rho^{+}_{\mathrm{surf}}(v;t) \\ &= F_k^{-}(V;t) - F_k^{+}(V;t) \end{aligned} \qquad [21]$$

and from eqn [18]

$$\left(\frac{M}{A}\right)\frac{\mathrm{d}}{\mathrm{d}t}\langle V\rangle = M\alpha\langle F_2(V;t)\rangle_\Phi \qquad [22]$$

$$\left(\frac{M}{A}\right)\frac{\mathrm{d}}{\mathrm{d}t}\langle V^2\rangle = M\alpha[\langle VF_2(V;t)\rangle_\Phi + \alpha\langle F_3(V;t)\rangle_\Phi] \qquad [23]$$

Introducing $\bar{V} = \langle V\rangle_\Phi$ then from eqns [12] and [20], it follows that the (kinetic) energies satisfy

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\langle E^{\pm}\rangle}{A}\right) &= \pm M\alpha\Big[\langle F_2^{\pm}(V;t)\rangle_\Phi \bar{V} \\ &\quad + \langle (V-\bar{V})F_2^{\pm}(V;t)\rangle_\Phi \\ &\quad + \frac{\alpha}{2}\langle F_3^{\pm}(V;t)\rangle_\Phi\Big] \end{aligned} \qquad [24]$$

which implies conservation of energy.

From the first law of thermodynamics,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\langle E^{\pm}\rangle}{A}\right) = \frac{1}{A}\left[P_W^{P\to\pm} + P_Q^{P\to\pm}\right] \qquad [25]$$

where $P_W^{P\to\pm}$ and $P_Q^{P\to\pm}$ denote the work- and heat-power transmitted by the piston to the fluid, we conclude from eqns [22] and [25] that the heat flux is

$$\begin{aligned} \frac{1}{A}P_Q^{P\to\pm} &= \pm M\alpha\Big[\langle (V-\bar{V})F_2^{\pm}(V;t)\rangle_\Phi \\ &\quad + \frac{\alpha}{2}\langle F_3^{\pm}(V;t)\rangle_\Phi\Big] \end{aligned} \qquad [26]$$

Since $\alpha \ll 1$, it is interesting to introduce the irreducible moments

$$\Delta_r = \langle (V-\bar{V})^r\rangle_\Phi \qquad [27]$$

and the expansion around $\bar{V} = \langle V\rangle_t$,

$$F_n^{\pm}(V;t) = \sum_{r=0}^{\infty}\frac{1}{r!}F_n^{(r,\pm)}(\bar{V})(V-\bar{V})^r \qquad [28]$$

from which one obtains equations for $\mathrm{d}\Delta_r/\mathrm{d}t$. In particular, using the identities

$$F_3^{(r+1,\pm)} = -3F_2^{(r,\pm)}, \qquad F_2^{(r+2,\pm)} = 2F_0^{(r,\pm)} \qquad [29]$$

in [22] and [24], we have

$$\begin{aligned} \langle F_2^{\pm}(V;t)\rangle_\Phi &= F_2^{\pm}(\bar{V};t) \\ &\quad + \sum_{r\geq 0}\frac{2}{(2+r)!}F_0^{(r,\pm)}\Delta_{2+r} \end{aligned} \qquad [30]$$

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\langle E^{\pm}\rangle}{A}\right) &= \pm M\alpha\Big[\langle F_2^{\pm}(V;t)\rangle_\Phi \bar{V} \\ &\quad + \frac{\alpha}{2}F_3^{\pm}(\bar{V};t) + \frac{1}{2}\sum_{r\geq 2}\frac{1}{r!}(2r-3\alpha) \\ &\quad \times F_2^{(r-1,\pm)}(\bar{V};t)\Delta_r\Big] \end{aligned} \qquad [31]$$

Depending on the questions or approximations one wants to study, either the distribution $\Phi(V;t)$ or the moments $\langle V^n\rangle_t$ will be the interesting objects. Finally, with the condition [19], one can take eqn [12] for $x \neq X_t$ and impose the boundary conditions at $x = X_t$:

$$\begin{aligned} \rho^{-}(X_t,v;t) &= \rho^{-}(X_t,v';t), \quad \text{if } v < V_t \\ \rho^{+}(X_t,v;t) &= \rho^{+}(X_t,v';t), \quad \text{if } v > V_t \end{aligned} \qquad [32]$$

and similarly for $x = 0$ and $x = L$ with $v' = -v$.

Let us note that this factorization condition is of the same nature as the molecular chaos assumption introduced in kinetic theory, and with this condition eqn [13] yields the Boltzmann equation for this model.

In the following, to obtain explicit results as a function of the initial temperatures $T_0^{\pm}$, we take Maxwellian distributions $\varphi_0^{\pm}(v)$ and initial conditions $(p_0^{\pm}, T_0^{\pm}, n_0^{\pm})$ such that the velocity of the piston remains small (i.e., $|\langle V\rangle_t| \ll |\langle v^{\pm}\rangle_0|$).

## Distribution $\Phi(V;t)$ for the Infinite Cylinder ($L = \infty$)

To lowest order in $\epsilon = \sqrt{m/M}$, and assuming $|1 - p^{+}/p^{-}|$ is of order $\epsilon$, one obtains from eqn [16] the usual Fokker–Planck equation whose solution gives

$$\Phi_0(V;t) = \frac{1}{\sqrt{2\pi}}\frac{1}{\Delta(t)}\exp-\left(\frac{(V-\bar{V}(t))^2}{2\Delta^2(t)}\right) \qquad [33]$$

with

$$\bar{V}(t) = (p^- - p^+)\sqrt{\frac{\pi k_B}{8m}}\left[\frac{p^+}{\sqrt{T^+}} + \frac{p^-}{\sqrt{T^-}}\right]^{-1}(1 - e^{-\lambda t})$$
$$\lambda = \frac{A}{M}\sqrt{\frac{8m}{\pi k_B}}\left[\frac{p^+}{\sqrt{T^+}} + \frac{p^-}{\sqrt{T^-}}\right] \quad [34]$$
$$\Delta^2(t) = \frac{k_B}{M}\sqrt{T^- T^+}\,\frac{p^+\sqrt{T^+} + p^-\sqrt{T^-}}{p^+\sqrt{T^-} + p^-\sqrt{T^+}}(1 - e^{-2\lambda t})$$

where we have dropped the index "zero" on the variable $T^\pm, n^\pm$ and used the equation of state $p^\pm = n^\pm k_B T^\pm$.

In conclusion, in the thermodynamic limit for the piston ($M \to \infty, M/A$ fixed), eqn [33] shows that the evolution is deterministic, that is, $\Phi(V;t) = \delta(V - \bar{V}(t))$, where the velocity $\bar{V}(t)$ of the piston tends exponentially fast toward stationary value $V_{stat} = \bar{V}(\infty)$ with relaxation time $\tau = \lambda^{-1}$.

Let us note that for $p^+ = p^-$, we have $\bar{V}(t) \equiv 0$ and the evolution [33] is identical to the Ornstein–Uhlenbeck process of thermalization of the Brownian particle starting with zero velocity and friction coefficient $\lambda$. The analysis of [16] to first order in $\epsilon$ yields then

$$\Phi(V;t) = \left[1 + \epsilon\sum_{k=0}^{3} a_k(t)(V - \bar{V}(t))^k\right]\Phi_0(V;t) \quad [35]$$

where $a_k(t)$ can be explicitly calculated and $a_0(t) = -\Delta^2(t)a_2(t)$ because of the normalization condition. Moreover, $a_2(t) \sim (p^- - p^+)$, that is, $a_2(t) = 0$ if $p^- = p^+$. From [35], one obtains

$$\begin{aligned}\langle V\rangle_t = &\sqrt{\frac{\pi k_B}{8m}}\frac{\sqrt{T^- T^+}}{p^+\sqrt{T^-} + p^-\sqrt{T^+}} \\ &\times \Big\{(p^- - p^+)(1 - e^{-\lambda t}) \\ &+ (p^- - p^+)^2\frac{\pi}{8}\frac{(p^- T^+ - p^+ T^-)}{(p^+\sqrt{T^-} + p^-\sqrt{T^+})^2} \\ &\times (1 - 2\lambda t e^{-\lambda t} - e^{-2\lambda t}) \\ &+ \frac{m}{M}\frac{1}{\sqrt{T^- T^+}}(p^- T^+ - p^+ T^-) \\ &\times \left(\frac{p^+\sqrt{T^+} + p^-\sqrt{T^-}}{p^+\sqrt{T^-} + p^-\sqrt{T^+}}\right)(1 - e^{-\lambda t})^2\Big\}\end{aligned} \quad [36]$$

and

$$\langle V^2\rangle_t - \langle V\rangle_t^2 = \Delta^2(t)\left[1 + \sqrt{\frac{m}{M}}2\Delta^2(t)a_2(t)\right] \quad [37]$$

From eqn [36], we now conclude that for equal pressures $p^- = p^+$, the piston will evolve stochastically to a stationary state with nonzero velocity toward the warmer side

$$\left.\begin{aligned}\langle V\rangle_{stat} &= \frac{m}{M}\sqrt{\frac{\pi k_B}{8m}}(\sqrt{T^+} - \sqrt{T^-}) \\ \langle V^2\rangle_{stat} - \langle V\rangle_{stat}^2 &= \frac{k_B}{M}\sqrt{T^- T^+}\end{aligned}\right\} \text{ if } p^- = p^+ \quad [38]$$

Let us remark that we have established eqn [35] under the condition that $|1 - p^+/p^-| = \mathcal{O}(\epsilon)$, but as we see in the next section, the stationary value $V_{stat}$ obtained from eqn [36] remains valid whenever $|(1 - p^+/p^-)(1 - \sqrt{T^+/T^-})| \ll 1$.

## Moments $\langle V^n\rangle_t$: Thermodynamic Limit for the Piston

### General Equations: Adiabatic Evolution

In the thermodynamic limit $M \to \infty$, $\alpha \to 0$, $\gamma = \alpha A$ is fixed and eqn [16] reduces to

$$\partial_t\Phi(V;t) = -\gamma\frac{\partial}{\partial V}\tilde{F}_2(V;t) \quad [39]$$

Integrating [39] with initial condition $\Phi(V;t=0) = \delta(V)$ yields

$$\Phi(V,t) = \delta(V - \bar{V}(t)), \text{ that is, } \langle V^n\rangle_t = \langle V\rangle_t^n \quad [40]$$

where

$$\frac{d}{dt}V(t) = \gamma F_2(V(t);t), \quad V(t=0) = 0 \quad [41]$$

Moreover,

$$\tilde{F}_2(V;t) = F_2(V;t)\Phi(V;t) \quad [42]$$

and

$$\begin{aligned}\rho_{\pm,P}(X, v; X, V; t) = \rho^\pm(x, v; t)\delta(X - X(t)) \\ \times \delta(V - V(t))\end{aligned} \quad [43]$$

where $dX(t)/dt = V(t), X(t=0) = X_0$.

In conclusion, as already mentioned, in this limit the factorization condition (eqn [19]) is an exact relation. Let us note that $\rho^\pm_{surf}(v;t) = \rho^\pm_{surf}(2V - v;t)$ if $v > V(t)$ (on the right) or $v < V(t)$ (on the left). Let us also remark that $2mF_2^\pm(V(t);t)$ represents the effective pressure from the right/left exerted on the piston. Moreover, since for any distribution $\rho^\pm_{surf}(v;t)$, the functions $F_2^-(V;t)$ and $-F_2^+(V;t)$ are monotonically decreasing, we can introduce the decomposition

$$p^\pm_{surf} = 2mF_2^\pm(V;t) = \hat{p}^\pm \pm \left(\frac{M}{A}\right)\lambda^\pm(V;t)V \quad [44]$$

where the static pressure at the surface is $\hat{p}^\pm(t) = p^\pm_{surf}(V=0;t)$ and the friction coefficients

$\lambda^{\pm}(V;t)$ are strictly positive. The evolution [41] is thus of the form

$$\frac{d}{dt}V(t) = \frac{A}{M}(\hat{p}^- - \hat{p}^+) - \lambda(V)V \qquad [45]$$

It involves the difference of static pressure and the friction coefficient $\lambda(V) = \lambda^-(V) + \lambda^+(V)$. Finally, from eqn [12], we obtain the evolution of the (kinetic) energy per unit area for the fluids in the left and right compartments:

$$\frac{d}{dt}\left(\frac{<E^{\pm}>}{A}\right) = \pm 2mF_2^{\pm}(V;t)V \qquad [46]$$

Therefore, from [40] and [46], and the first law of thermodynamics, we recover the conclusions obtained in the previous section, that is, in the thermodynamic limit for the piston, the evolution (eqns [41], [12], and [35]) is deterministic and adiabatic (i.e., in [46] only work and no heat is involved).

### Infinite Cylinder ($L = \infty, M = \infty$)

As already discussed, for $L = \infty$ we can neglect the recollisions. Therefore, in $F_2^{\pm}$ the distribution $\rho^{\pm}(v;t)$ can be replaced by $n_0^{\pm}\varphi_0^{\pm}(v)$ and $F_2^{\pm}(V)$ is independent of $t$. In this case, the evolution of the piston is simply given by the ordinary differential equation

$$\frac{d}{dt}V(t) = \frac{A}{M}2mF_2(V), \quad V(t=0) = 0 \qquad [47]$$

where $F_2(V)$ is a strictly decreasing function of $V$. If $p_0^+ = p_0^-$, then $V(t) = 0$, that is, the piston remains at rest and the two fluids remain in their original thermal equilibrium. If $p_0^+ \neq p_0^-$, that is, $n_0^+ k_B T_0^+ \neq n_0^- k_B T_0^-$, the piston will evolve monotonically to a stationary state with constant velocity $V_{stat}$ solution of $F_2(V_{stat}) = 0$. From [34], it follows that $V_{stat}$ is a function of $n_0^+/n_0^-, T_0^-, T_0^+$ but does not depend on the value $M/A$. Moreover, the approach to this stationary state is exponentially fast with relaxation time $\tau_0 = 1/\lambda(V=0)$. For Maxwellian distributions $\varphi_0^{\pm}(v)$, $V_{stat}$ is a solution of

$$k_B\left(n_0^- T_0^- - n_0^+ T_0^+\right) - V_{stat}\sqrt{\frac{8k_B m}{\pi}}\left(n_0^-\sqrt{T_0^-} - n_0^+\sqrt{T_0^+}\right) + V_{stat}^2 m\left(n_0^- - n_0^+\right) + \mathcal{O}\left(V_{stat}^3\right) = 0 \qquad [48]$$

Moreover,

$$\tau_0^{-1} = \frac{A}{M}\sqrt{\frac{8k_B m}{\pi}}\left(n_0^-\sqrt{T_0^-} + n_0^+\sqrt{T_0^+}\right) \qquad [49]$$

which implies that the relaxation time will be very small either if $M/A \ll 1$, or if $n_0^{\pm} = \xi\,\tilde{n}_0^{\pm}$ with $\xi \gg 1$. In this case, the piston acquires almost immediately its final velocity $V_{stat}$ and one can solve eqn [12] to obtain the evolution of the fluids.

### Finite Cylinder ($L < \infty$, $M = \infty$)

For finite $L$, introducing the average temperature in the fluids

$$T_{av}^{\pm} = \frac{2\langle E^{\pm}\rangle_t}{k_B N^{\pm}} \qquad [50]$$

we have to solve [41] and [46], that is,

$$\begin{aligned} \frac{d}{dt}V(t) &= \frac{A}{M}2m\left[F_2^-(V;t) - F_2^+(V;t)\right] \\ k_B\frac{d}{dt}T_{av}^{\pm} &= \pm 4m\frac{A}{N^{\pm}}F_2^{\pm}(V;t)V \end{aligned} \qquad [51]$$

where $F_2^{\pm}(V;t)$ is a functional of $\rho_{surf}^{\pm}(v;t)$ which we decompose as

$$F_2^{\pm}(V;t) = \hat{n}^{\pm}(t)k_B\hat{T}^{\pm}(t) \pm \left(\frac{M}{A}\right)\lambda^{\pm}(V;t)V \qquad [52]$$

with

$$\begin{aligned} \hat{n}^-(t) &= \int_0^{\infty} dv\rho_{surf}^-(v;t) \\ \hat{n}^+(t) &= \int_{-\infty}^{0} dv\rho_{surf}^+(v;t) \end{aligned} \qquad [53]$$

and

$$\hat{n}^{\pm}k_B\hat{T}^{\pm} = \hat{p}^{\pm} \qquad [54]$$

For a time interval $\tau_1 = L\sqrt{m/k_B T}$ which is the time for the shock wave to bounce back, the piston will evolve as already discussed. In particular, if $R^{\pm}$ is sufficiently large, then after a time $\tau_0 = \mathcal{O}((R^{\pm})^{-1})$ the piston will reach the velocity $\bar{V}$ given by $F_2(\bar{V}, t) = 0$ (eqn [47]). For $t > \tau_1$, $F_2^{\pm}(V;t)$ depends explicitly on time. For $R^{\pm}$ sufficiently large, we can expect that for all $t$ the velocity $V(t)$ will be a functional of $\rho_{surf}^{\pm}(v;t)$ given by $F_2[V(t); \rho_{surf}^{\pm}(.;t)] = 0$, and thus the problem is to solve eqn [12] with the boundary condition (eqn [32]). Since $V(t)$ so defined is independent of $M/A$, the evolution will be independent of $M/A$ if $R^{\pm}$ is sufficiently large. This conclusion, which we cannot prove rigorously, will be confirmed by numerical simulations.

To give a qualitative discussion of the evolution for arbitrary values of $R^{\pm}$, we shall use the following assumption already introduced in the experimental measurement of $c_p/c_v$.

**Assumption 2** (Average assumption). The surface coefficients $\hat{n}^{\pm}(t)$ and $\hat{T}^{\pm}(t)$ (eqns [52]–[53]) coincide to order 1 in $\alpha$ with the average value of the density and temperature in the fluids, that is,

$$\hat{n}^- = \frac{N^-}{AX(t)}, \quad \hat{n}^+ = \frac{N^+}{A(L - X(t))}$$
$$\hat{T}^{\pm} = T^{\pm}_{\mathrm{av}}(t) \qquad [55]$$

We still need an expression for the friction coefficients. From

$$F_2^{\pm}(V;t) = \hat{p}^{\pm}(t) - 4mVF_1^{\pm}(V=0;t) + mV^2\hat{n}^{\pm}(t) + \mathcal{O}(V^3) \qquad [56]$$

then, assuming that to first order in $\alpha$, $F_1^{\pm}(V=0;t)$ is the same function of $\hat{T}^{\pm}(t)$ as for Maxwellian distributions, we have

$$\lambda^{\pm}(V) = \left(\frac{A}{M}\right)m\hat{n}^{\pm}\left[\sqrt{\frac{8k_{\mathrm{B}}\hat{T}^{\pm}}{m\pi}} \pm V\right] + \mathcal{O}(V^2) \qquad [57]$$

Therefore, choosing initial condition such that $V(t)$ is small for all time, eqn [51] yields

$$\sqrt{\hat{T}^-}X - \sqrt{\hat{T}^+}(L-X) = C = \sqrt{\hat{T}_0^-}X_0 - \sqrt{\hat{T}_0^+}(L-X_0) \qquad [58]$$

We thus obtain the equilibrium point for the adiabatic evolution ($M=\infty$):

$$\left(\frac{N^-}{A}\right)T_{\mathrm{f}}^- = \frac{2E_0}{Ak_{\mathrm{B}}}\frac{X_{\mathrm{f}}}{L} \qquad [59]$$

$$\left(\frac{N^+}{A}\right)T_{\mathrm{f}}^+ = \frac{2E_0}{Ak_{\mathrm{B}}}\left(1 - \frac{X_{\mathrm{f}}}{L}\right) \qquad [60]$$

where

$$\frac{2E_0}{Ak_{\mathrm{B}}} = \left(\frac{N^-}{A}\right)T_0^- + \left(\frac{N^+}{A}\right)T_0^+ \qquad [61]$$

and

$$\sqrt{\left(\frac{A}{N^-}\right)X_{\mathrm{f}}^3} - \sqrt{\left(\frac{A}{N^+}\right)(L-X_{\mathrm{f}})^3} = \sqrt{\frac{AL}{2E_0k_{\mathrm{B}}}}C \qquad [62]$$

Solving [58]–[62] gives the equilibrium state $(X_{\mathrm{f}}, T_{\mathrm{f}}^{\pm})$, which is a state of mechanical equilibrium $p_{\mathrm{f}}^- = p_{\mathrm{f}}^+$, but not thermal equilibrium $T_{\mathrm{f}}^- \neq T_{\mathrm{f}}^+$. Moreover, this equilibrium state does not depend on $M$. Having obtained the equilibrium point, we can then investigate the evolution close to the equilibrium point. Linearizing eqn [51] around $(X_{\mathrm{f}}, T_{\mathrm{f}}^{\pm})$ yields

$$\frac{\mathrm{d}}{\mathrm{d}t}V = k_{\mathrm{B}}\left[\left(\frac{N^-}{M}\right)\frac{T_{\mathrm{f}}^-X_{\mathrm{f}}^2}{X^3} - \left(\frac{N^+}{M}\right)\frac{T_{\mathrm{f}}^+(L-X_{\mathrm{f}})^2}{(L-X)^3}\right] - \lambda(V=0)V \qquad [63]$$

In other words, the effect of collisions on the piston is to induce an external potential of the form $[c_1|X|^{-2} + c_2(L-X)^{-2}]$ and a friction force. It is a damped harmonic oscillator with

$$\omega_0^2 = 6\left(\frac{E_0}{M}\right)\frac{1}{X_{\mathrm{f}}(L-X_{\mathrm{f}})}$$
$$\lambda = 4\sqrt{\frac{1}{\pi}}\sqrt{\frac{E_0}{ML}}\left[\sqrt{\frac{R^-}{X_{\mathrm{f}}}} + \sqrt{\frac{R^+}{(L-X_{\mathrm{f}})}}\right] \qquad [64]$$

(recall that $R^{\pm} = mN^{\pm}/M$). For the case $N^- = N^+$ to be considered in the simulations, eqn [64] implies that the motion is weakly damped if

$$R < R_{\max} = \frac{3\pi}{2}\left[\sqrt{\frac{X_{\mathrm{f}}}{L}} + \sqrt{1 - \frac{X_{\mathrm{f}}}{L}}\right]^{-2} \qquad [65]$$

with period

$$\tau = \frac{2\pi}{\omega_0}\frac{1}{\sqrt{R - R_{\max}}} \qquad [66]$$

and strongly damped if $R > R_{\max}$, in agreement with experimental observations.

## Moments $\langle V^n \rangle_t$: Piston with Finite Mass

### Equation to First Order in $\alpha = 2m/(M+m)$

If the mass of the piston is finite with $M \gg m$, then the irreducible moments $\Delta_r$ are of the order $\alpha^{[(r+1)/2]}$ where $[(r+1)/2]$ is the integral part of $(r+1)/2$. If the factorization condition [19] is satisfied, to first order in $\alpha$ we have

$$\langle V^n \rangle_t = V^n(t) + \frac{n(n-1)}{2}V^{n-2}(t)\Delta_2(t) \qquad [67]$$

where $V(t) = \langle V \rangle_t$ and $\Delta_2(t) = \langle V^2 \rangle_t - \langle V \rangle_t^2$ are solutions of

$$\frac{1}{\gamma}\frac{\mathrm{d}}{\mathrm{d}t}V(t) = F_2 + \Delta_2 F_0$$
$$\frac{1}{\gamma}\frac{\mathrm{d}}{\mathrm{d}t}\Delta_2(t) = -4\Delta_2 F_1 + \alpha F_3 \qquad [68]$$
$$\frac{1}{\gamma}\frac{\mathrm{d}}{\mathrm{d}t}\langle E^{\pm}\rangle_t = \pm\left\{M[F_2^{\pm} + \Delta_2 F_0^{\pm}]V + (M/2)[4\Delta_2 F_1^{\pm} - \alpha F_3^{\pm}]\right\}$$

and $\Delta_2 \doteq k_{\mathrm{B}}T_{\mathrm{P}}/M$ defines the temperature of the piston.

### Infinite Cylinder: Heat Transfer

For the infinite cylinder, the factorization assumption is an exact relation and in this case the functions $F_k(V;t)$ are independent of $t$. The solution

of the autonomous system [68] with $F_k = F_k(V)$ shows that the piston evolves to a stationary state with velocity $\bar{V}$ given by

$$F_2(\bar{V}) + \frac{\alpha}{4}\frac{F_3(\bar{V})F_0(\bar{V})}{F_1(\bar{V})} = 0 \qquad [69]$$

The temperature of the piston is

$$\bar{\Delta}_2 = \frac{k_B T_P}{M} = \frac{\alpha}{4}\frac{F_3(\bar{V})}{F_1(\bar{V})} \qquad [70]$$

and the heat flux from the piston to the fluid is

$$\frac{1}{A}P_Q^{P\to -} = \frac{m^2}{2M}\left[\frac{F_3^+F_1^- - F_3^-F_1^+}{F_1^- - F_1^+}\right] \qquad [71]$$

If we choose initial conditions such that $|V(t)| \ll 1$ for all $t$, and Maxwellian distributions $\varphi^{\pm}(v)$, the solutions $V(t)$, $\Delta_2(t)$ coincide with the solutions previously obtained (eqns [36] and [37]) and

$$\frac{1}{A}P_Q^{P\to -} = (T^+ - T^-) \times \frac{m}{M}\sqrt{\frac{8k_B}{m\pi}} \times \frac{p^-p^+}{(p^+\sqrt{T^-} + p^-\sqrt{T^+})} \qquad [72]$$

In conclusion, to first order in $m/M$, there is a heat flux from the warm side to the cold one proportional to $(T^+ - T^-)$, induced by the stochastic motion of the piston.

### Finite Cylinder ($L < \infty, M < \infty$)

**Singular character of the perturbation approach** Whereas the leading order is actually the "thermodynamic behavior" $M = \infty$ in the first two stages of the evolution (fast relaxation toward mechanical equilibrium), the fluctuations of order $\mathcal{O}(\alpha)$ rule the slow relaxation toward thermal equilibrium. It is thus obvious that a naive perturbation approach cannot give access to "both" regimes. This difficulty is reminiscent of the boundary-layer problems encountered in hydrodynamics, and the perturbation method to be used here is the exact temporal analog of the matched perturbative expansion method developed for these boundary layers. The idea is to implement two different perturbation approaches:

1. one at short times, with time variable $t$ describing the fast dynamics ruling the fast relaxation toward mechanical equilibrium; and
2. one for longer times, with a rescaled time variable $\tau = \alpha t$.

The second perturbation approach above is supplemented with a "slaving principle," expressing that at each time of the slow evolution, that is, at fixed $\tau$, the still present fast dynamics has reached a local asymptotic state, slaved to the values of the slow observables. The initial conditions are set on the first-stage solution. The initial conditions of the second regime match the asymptotic behavior of the first-stage solution ("matching condition").

The slaving principle is implemented by interpreting an evolution equation of the form

$$\frac{da}{dt} \equiv \alpha\frac{da}{d\tau} = A(\tau, a), \quad A = \mathcal{O}(1) \qquad [73]$$

as follows: it indicates that $a$ is in fact a fast quantity relaxing at short times ($\ll\tau$) toward a stationary state $a_{eq}(\tau)$ slaved to the slow evolution and determined by the condition

$$A[\tau, a_{eq}(\tau)] = 0 \qquad [74]$$

(at lowest order in $\alpha$, actually $A[\tau, a_{eq}(\tau)] = \mathcal{O}(\alpha)$ which prescribes the leading order of $a_{eq}(\tau)$); the following-order terms can be arbitrarily fixed as long as only the first order of perturbation is implemented. Physically, such a condition arises to express that an instantaneous mechanical equilibrium takes place at each time $\tau$ of the slow relaxation to thermal equilibrium.

**Equations for the fluctuation-induced evolution of the system** Following this procedure, we arrive at explicit expressions for the rescaled quantities (of order $\mathcal{O}(1)$) $\widetilde{V} = V/\alpha$, $\widetilde{\Delta}_2 = \Delta_2/\alpha$, and $\widetilde{\Pi} = (p^- - p^+)/\alpha$:

$$\begin{aligned}
\widetilde{V} &= \frac{m}{3}\left(\frac{AL}{E_0}\right)\left(\frac{F_3^-F_1^+ - F_3^+F_1^-}{F_1}\right) + \mathcal{O}(\alpha) \\
\frac{\widetilde{\Pi}}{2m} &= \frac{2m}{3}\left(\frac{AL}{E_0}\right)(F_3^-F_1^+ - F_3^+F_1^-) \\
&\quad - \frac{F_3F_1}{4F_1} + \mathcal{O}(\alpha) \\
\widetilde{\Delta}_2 &= \frac{F_3}{4F_1} + \mathcal{O}(\alpha)
\end{aligned} \qquad [75]$$

We then introduce a (dimensionless) rescaled position for the piston

$$\xi = \frac{1}{2} - \frac{X}{L} \in \left[-\frac{1}{2}, \frac{1}{2}\right] \qquad [76]$$

which satisfies

$$\frac{d\xi}{d\tau} = -k_B(T^- - T^+)\left(\frac{2A}{3E_0}\right)\frac{F_1^-F_1^+}{F_1} \qquad [77]$$

To discuss eqn [77], a third assumption has to be introduced.

**Assumption 3 (Maxwellian Identities).** In the regime when $V = \mathcal{O}(\alpha)$, the relations between the functionals $F_1, F_2$, and $F_3$ are the same at lowest order in $\alpha$ as if the distributions $\rho_{surf}^{\pm}(v; V; t)$ were Maxwellian in $v$:

$$F_1^{\pm}(V) \approx \mp\rho^{\pm}\sqrt{\frac{k_B T^{\pm}}{2m\pi}}$$
$$F_3^{\pm}(V) \approx \left(\frac{2k_B T^{\pm}}{m}\right)F_1^{\pm}(V) - VF_2^{\pm}(V) \quad [78]$$

Using these identities and the (dimensionless) rescaled time

$$s = \tau\frac{2}{3L}\sqrt{\frac{k_B}{m\pi}}\sqrt{\frac{2(N^- T_0^- + N^+ T_0^+)}{N}} \quad [79]$$

where $N = N^+ + N^-$, we obtain a deterministic equation describing the piston motion (Gruber *et al.* 2003):

$$\frac{d\xi}{ds} = -\left[\sqrt{\frac{N}{2N^+}(1+2\xi)} - \sqrt{\frac{N}{2N^-}(1-2\xi)}\right]$$
$$\xi(0) = \frac{1}{2} - \frac{X_{ad}}{L} \quad [80]$$

where $X_{ad}$ is the piston position at the end of the adiabatic regime (i.e., $X_f$, eqn [62]). The meaningful observables straightforwardly follow from the solution $\xi(s)$:

$$X(s) = L\left(\frac{1}{2} - \xi(s)\right)$$
$$T^{\pm}(s) = [1 \pm 2\xi(s)]\left(\frac{N^- T_0^- + N^+ T_0^+}{2N^{\pm}}\right) \quad [81]$$

The first-order perturbation analysis using a single rescaled time $t_1 = \alpha t_0$ is valid in the regime when $V = \mathcal{O}(\alpha)$ and it gives access to the relaxation toward thermal equilibrium up to a temperature difference $T^+ - T^- = \mathcal{O}(\alpha)$. For the sake of technical completeness (rather that physical relevance, since the above first-order analysis is enough to get the observable, meaningful behavior), let us mention that the perturbation analysis can be carried over at higher orders; using further rescaled times $t_2 = \alpha^2 t_0, \ldots, t_n = \alpha^n t_0$, it would allow us to control the evolution up to a temperature difference $|T^+ - T^-| = \mathcal{O}(\alpha^n)$; however, one could expect that the factorization condition does not hold at higher orders.

## Numerical Simulations

As we have seen, the results were established under the condition that $m/M$ is a small parameter. Moreover for finite systems $(L < \infty, M < \infty)$, it was assumed that before collisions and to first order in $m/M$, the factorization and the average assumptions are satisfied. The numerical simulations are thus essential to check the validity of these assumptions, to determine the range of acceptable values $m/M$ for the perturbation expansion, to investigate the thermodynamic limit, and to guide the intuition.

In all simulation, we have taken $k_B = 1, m = 1$, $T^- = 1$ and usually $T^+ = 10$. For $L$ finite, we have taken $L = 60, X_0 = 10, A = 10^5$, and $N^+ = N^- = N/2$, that is, $p^- = R(M/A)(1/10)$ and $p^+ = 2p^-$. The number of particles $N$ was varied from a few hundreds to one or several millions; the mass $M$ of the piston from 1 to $10^5$. We give below some of the results which have been obtained for $L = \infty$ (**Figures 2 and 3**)

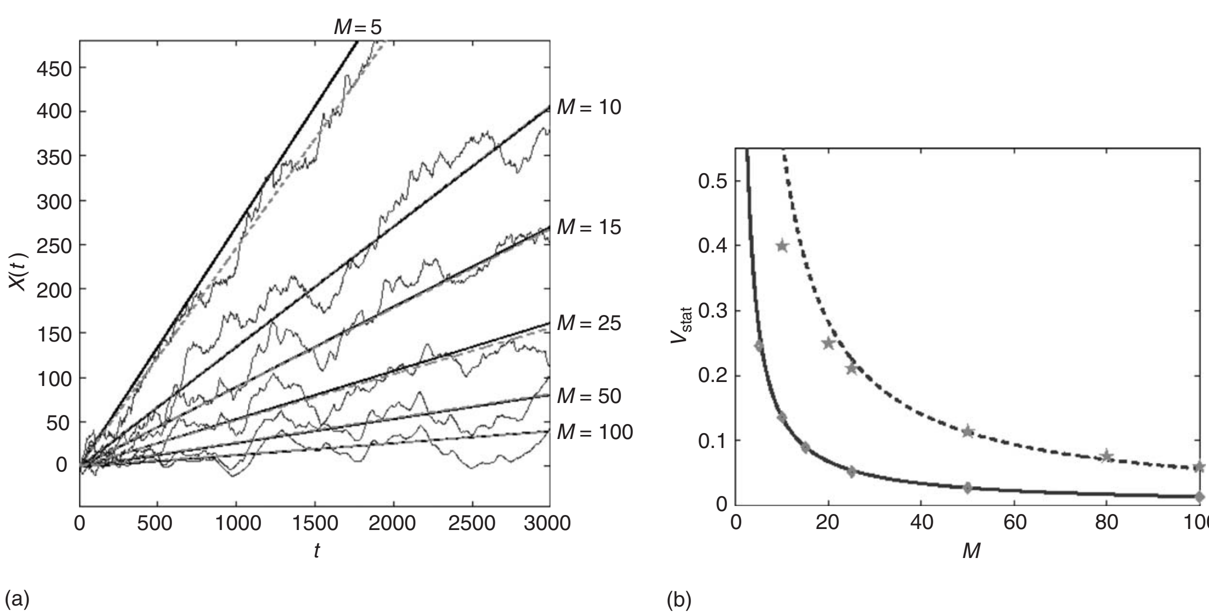



**Figure 2** Evolution of the piston for $L = \infty$, and $p^- = p^+ = 1$ as observed in simulations (stochastic line in (a), dots in (b)) compared with prediction: (a) position $X(t)$ for $T^+ = 10$; and (b) stationary velocity for $T^+ = 10$ (continuous line) and $T^+ = 100$ (dotted line), as a function of $M$.

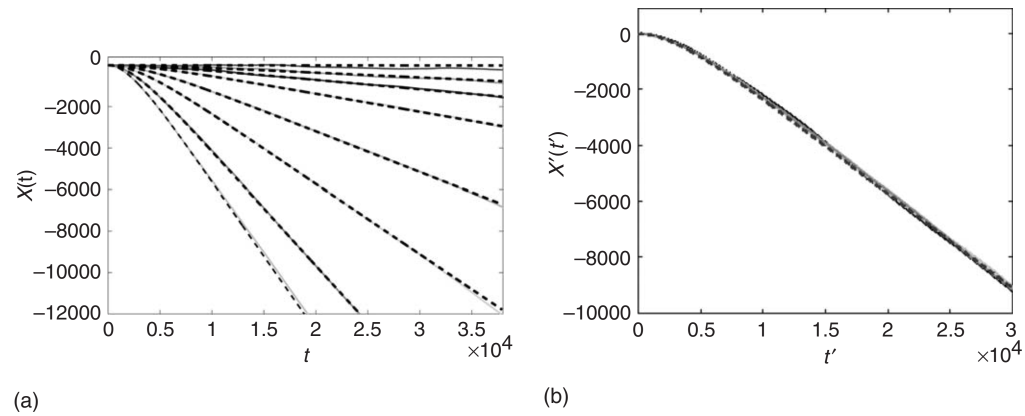



**Figure 3** Evolution of the piston for $L=\infty, M=10^4$, and $p^+ \neq p^-$ as observed in simulations (continuous line) compared with predictions (dotted line): (a) $p^-=1, p^+=p^-+\Delta p$, from top to bottom $\Delta p/p^-=0.05, 0.1, 0.2, 1, 2, 3$; and (b) $p^-=\zeta,\ p^+=2\zeta$, $\Delta p/p^-=1; X'=\zeta X, t'=\zeta t,\ \zeta=10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3, 10^4$.

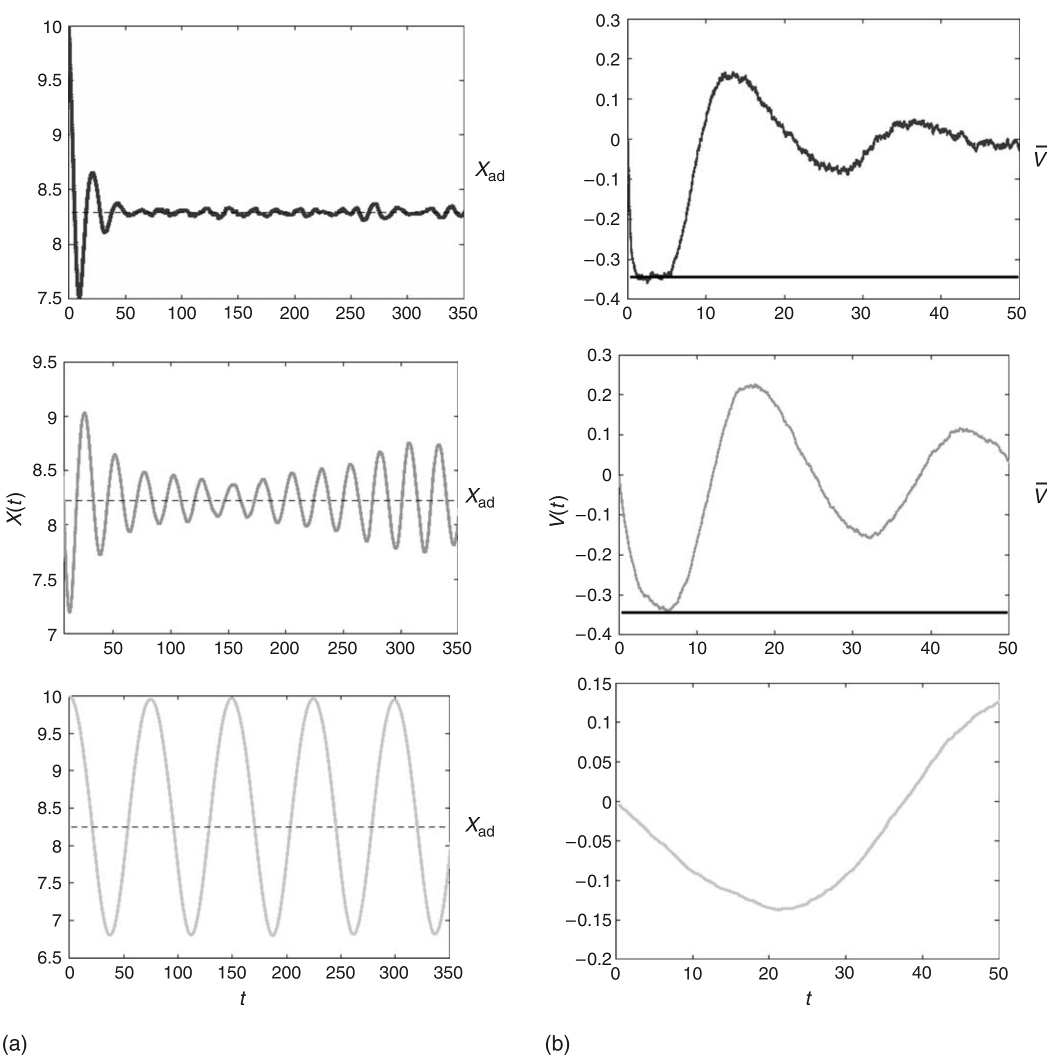



**Figure 4** "Deterministic" evolution toward mechanical equilibrium for $L<\infty, M=10^5$: (a) position $X(t)$; one finds $X_{\rm ad}^{\rm sim}=8.3$ whereas $X_{\rm ad}^{\rm th}=8.42$ and (b) velocity $V(t)$; one finds $\bar{V}^{\rm sim}=-0.343$ whereas $\bar{V}^{\rm th}=-0.3433$. From top to bottom: $R=12$: strong damping, independent of $R$ and $M$ for $R>4$ and $M>10^3$. $R=2$: critical damping. $R=0.1$: weak damping; damping coefficient increases with $R$ and $\omega_0 \sim \sqrt{R}$ for $R<1$ but is independent of $M$ for $M>10^3$.

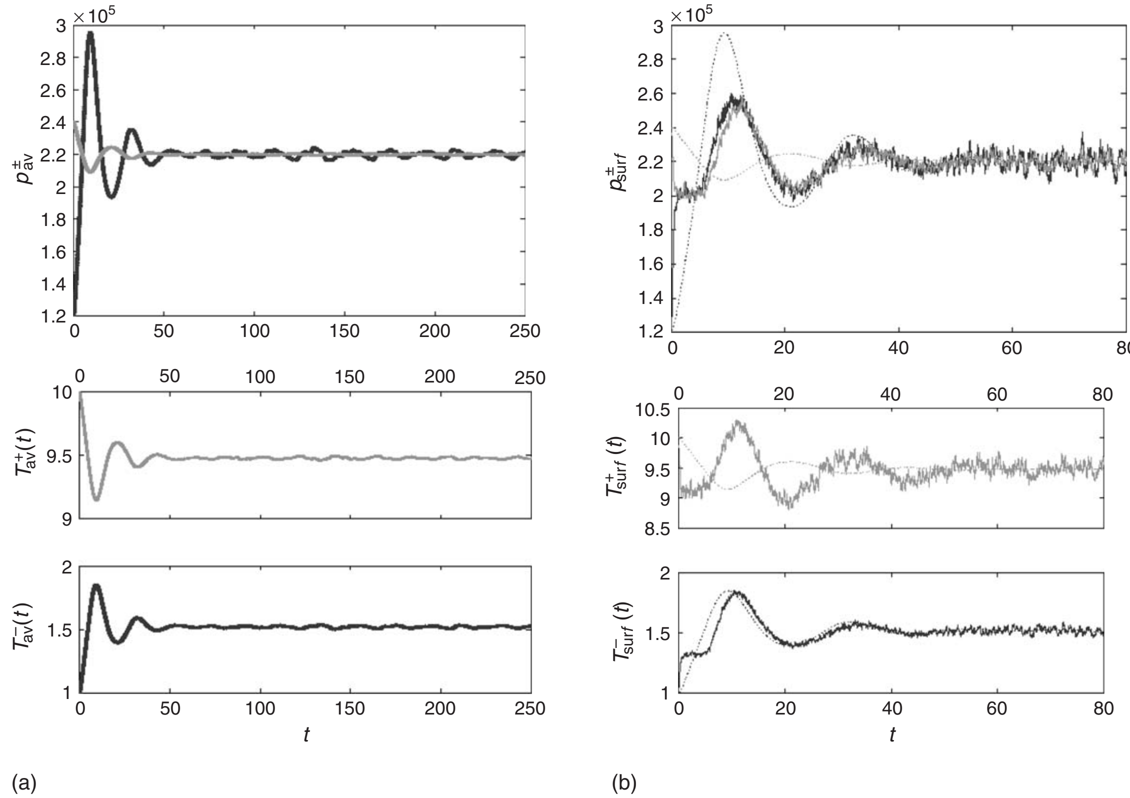



**Figure 5** Same conditions as **Figure 4**, $R = 12$: (a) average pressure and temperature in the fluid: $p^{\pm}_{\text{av}}(t) = 2E^{\pm}n^{\pm}/N^{\pm}$, $T^{\pm}_{\text{av}} = E^{\pm}/N^{\pm}k_B$ and (b) pressure and temperature at the surface of the piston. Prediction: $T^{-}_{\text{ad}} = 1.54$, $T^{+}_{\text{ad}} = 9.46$, $p^{-}_{\text{ad}} = p^{+}_{\text{ad}} = 2.2$. Simulations: $T^{-}_{\text{ad}} = 1.52$, $T^{+}_{\text{ad}} = 9.48$, $p^{-}_{\text{ad}} = p^{+}_{\text{ad}} = 2.2$.

and for $L < \infty$ approach to mechanical equilibrium (Figures 4–6) and to thermal equilibrium (Figures 7 and 8).

## Conclusions and Open Problems

In this article, the adiabatic piston has been investigated to first order in the small parameter $m/M$, but no attempt has been made to control the remainder terms. For an infinite cylinder, no other assumptions were necessary and the numerical simulations (Figures 2 and 3) are in perfect agreement with the theoretical prediction in particular for the stationary velocity $V_{\text{stat}}$, the friction coefficient $\lambda(V)$, and the relaxation time $\tau$.

For a finite cylinder ($L < \infty$) and in the thermodynamic limit ($M = \infty$), we were forced to introduce the average assumption to obtain a set of autonomous equations. As we have seen when initially $p^- \neq p^+$, this limiting case also describes the evolution to lowest order during the first two stages characterized by a time of the order $t_1 = L\sqrt{m/k_B T}$, where the evolution is adiabatic and deterministic. In the first stage, that is, before the shock wave bounces back on the piston, the simulations confirm the theoretical predictions. In particular, they show that if $R > 4$, the piston will be able to reach and maintain for some time the velocity $V_{\text{stat}}$, whereas this will not be the case for $R < 1$ (Figure 4b). In the second stage of the evolution, the simulations (Figure 4) exhibit damped oscillations toward mechanical equilibrium which are in very good agreement with the predictions for the final state $(X_{\text{ad}}, T^{\pm}_{\text{ad}})$, the frequency of oscillations and the existence of weak and strong damping depending on $R < 1$ or $R > 4$. Moreover, the general behavior of the evolution observed in the simulations as a function of the parameters was as predicted. However, the damping coefficient of these oscillations is wrong by one or several orders of magnitude. To understand this discrepancy, we note that using the average assumption we have related the damping to the friction coefficient. However, the simulations clearly show that those two dissipative effects have totally different origins. Indeed, as one can see with $L = \infty$, friction is associated with the fact that the density of the gas in front and in the back of the piston is not the same as in the bulk, and this generates a shock wave that propagates in the fluid. For finite $L$, when $R > 4$, the stationary velocity $V_{\text{stat}}$ is reached and the effect of friction is

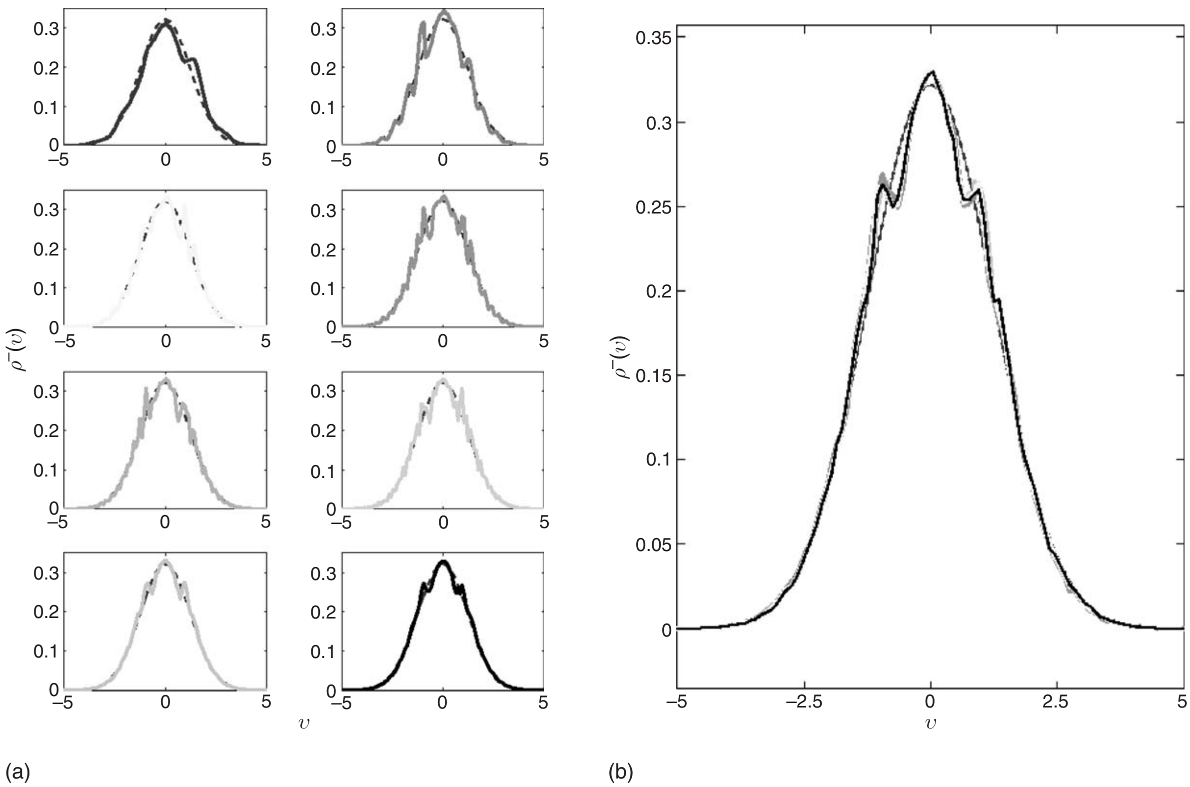



**Figure 6** Velocity distribution in the left compartment. Same conditions as **Figure 4**, $R = 12$. Dotted line corresponds to Maxwellian with $T^- = 1.52$: (a) $t = 12, 24, 36, 48, 60, 92, 144, 240$ from top to bottom and (b) $t = 276 - 460$.

to transfer in this first stage more and more energy to the fluid on one side and vice versa on the other side. However, to stop the piston and reverse its motion, only a certain amount of the transferred energy is necessary and the rest remains as dissipated energy in the fluid leading to a strong damping. On the other hand, for $R < 1$, the value $V_{\text{stat}}$ is never reached and all the energy transferred is necessary to revert the motion. In this case very little dissipation is involved and the damping will be very small. This indicates that the mechanism responsible for damping is associated with shock waves bouncing back and forth and the average assumption, which corresponds to a homogeneity condition throughout the gas, cannot describe the situation. In fact, the simulations (**Figure 5b**) indicate that the average assumption does

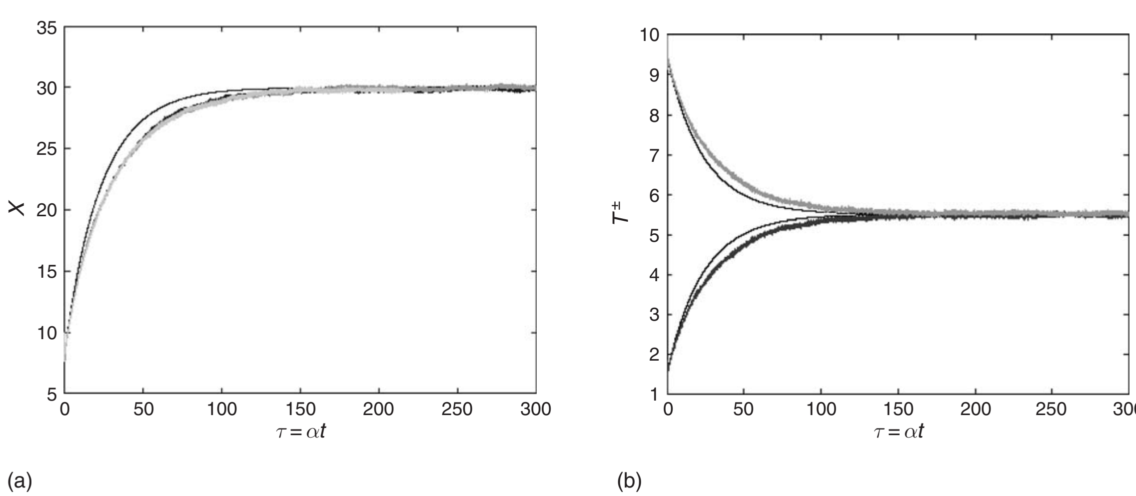
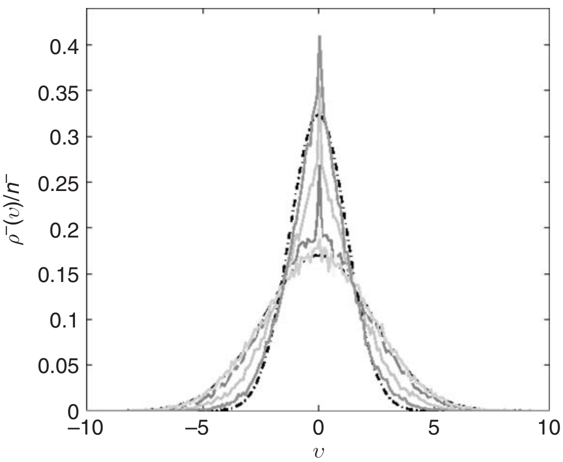
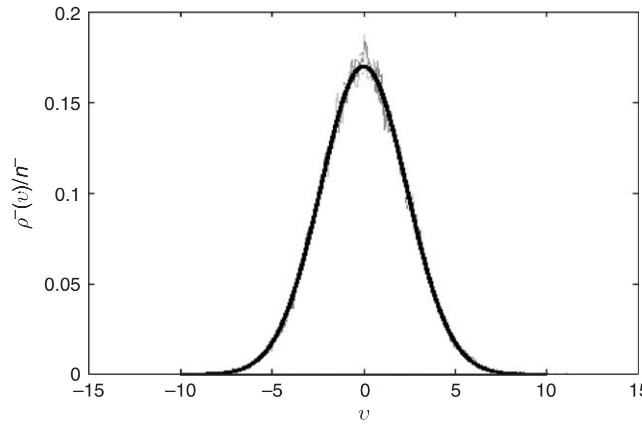



**Figure 7** Approach to thermal equilibrium, $N^{\pm} = 3 \times 10^4$. The smooth curves correspond to the predictions, the stochastic curves to simulations: (a) position $X(\tau)$, $\tau = \alpha t$, no visible difference for $M = 100, 200, 1000$ and (b) average temperatures $T^{\pm}(\tau)$, $\tau = \alpha t$, $M = 200$.




(a)




(b)

**Figure 8** Approach to thermal equilibrium from $T_{ad}^- = 1.54$ (dotted line in(a)) to $T_f^- = 5.5$ (heavy line in (b)). Velocity distribution function on the left for $M = 200$, $N^{\pm} = 5 \times 10^4$. (a) $\tau = \alpha\ t = 2, 4, 14, 48, 92, 144$ and (b) approach to Maxwellian distribution for $\tau > 445$.

not hold in this second stage. In conclusion, one is forced to admit that to describe correctly the adiabatic evolution, it is necessary to study the coupling between the motion of the piston and the hydrodynamic equations of the gas. Preliminary investigations have been initiated, but this is still one of the major open problems. Another problem would be to study the evolution in the case of interacting particles. However, investigations with hard disks suggest that no new effects should appear. To investigate adiabatic evolution, a simpler version of the adiabatic piston problem, without any controversy, has been introduced: this is the model of a standard piston with a constant force acting on it.

In the third stage, that is, the very slow approach to thermal equilibrium, another assumption was necessary, namely the factorization condition. The simulations (Figure 7) show a very good agreement with the prediction, and in particular the scaling property with $t' = t/M$ is perfectly verified. It appears that the small discrepancy between simulations and theoretical predictions could be due to the fact that, to compute explicitly the coefficients in the equations of motion, we have taken Maxwellian relations for the velocities of the gas particles, which is clearly not the case (Figure 8a).

The fourth stage of the evolution, that is, the approach to Maxwellian distributions (Figure 8b), is still another major open problem. Some preliminary studies have been conducted, where one investigates the stability and the evolution of the system when initially the two gases are in the same equilibrium state, but characterized by a distribution function which is not Maxwellian.

Finally, let us mention that the relation between the piston problem and the second law of thermodynamics is one more major problem. The question of entropy production out of equilibrium, and the validity of the second law, are still highly controversial. Again, preliminary results can be found in the literature. Among other things, this question has led to a model of heat conductivity gases, which reproduces the correct behavior (Gruber and Lesne 2005).

*See also:* Billiards in Bounded Convex Domains; Boltzmann Equation (Classical and Quantum); Hamiltonian Fluid Dynamics; Multiscale Approaches; Nonequilibrium Statistical Mechanics (Stationary): Overview; Nonequilibrium Statistical Mechanics: Dynamical Systems Approach.

## Further Reading


Callen HB (1963) *Thermodynamics*. New York: Wiley. (Appendix C. See also Callen HB (1985) *Thermodynamics and Thermostatics*, 2nd edn., pp. 51 and 53. New York: Wiley.)

Chernov N, Sinai YaG, and Lebowitz JL (2002) Scaling dynamic of a massive piston in a cube filled with ideal gas: exact results. *Journal of Statistical Physics* 109: 529–548.

Feynman RP (1965) *Lectures in Physics I*. New York: Addison-Wesley.

Gruber Ch (1999) Thermodynamics of systems with internal adiabatic constraints: time evolution of the adiabatic piston. *European Journal of Physics* 20: 259–266.

Gruber Ch and Lesne A (2005) Hamiltonian model of heat conductivity and Fourier law. *Physica A* 351: 358.

Gruber Ch, Pache S, and Lesne A (2003) Two-time-scale relaxation towards thermal equilibrium of the enigmatic piston. *Journal of Statistical Physics* 112: 1199–1228.

Gruber Ch and Piasecki J (1999) Stationary motion of the adiabatic piston. *Physica A* 268: 412–442.
Kestemont E, Van den Broeck C, and MalekMM (2000) The "adiabatic" piston: and yet it moves. *Europhysics Letters* 49: 143.
Lebowitz JL, Piasecki J, and Sinai YaG (2000) Scaling dynamics of a massive piston in an ideal gas. In: Szász D (ed.) *Hard Ball Systems and the Lorentz Gas*, Encyclopedia of Mathematical Sciences Series, vol. 101, pp. 217–227. Berlin: Springer.
Van den Broeck C, Meurs P, and Kawai R (2004) From Maxwell demon to Brownian motor. *New Journal of Physics* 7: 10.

# AdS/CFT Correspondence

**C P Herzog**, University of California at Santa Barbara, Santa Barbara, CA, USA
**I R Klebanov**, Princeton University, Princeton, NJ, USA



## Introduction

The anti-de Sitter/conformal field theory (AdS/CFT) correspondence is a conjectured equivalence between a quantum field theory in $d$ spacetime dimensions with conformal scaling symmetry and a quantum theory of gravity in $(d+1)$-dimensional anti-de Sitter space. The most promising approaches to quantizing gravity involve superstring theories, which are most easily defined in 10 spacetime dimensions, or $M$-theory which is defined in 11 spacetime dimensions. Hence, the AdS/CFT correspondences based on superstrings typically involve backgrounds of the form $AdS_{d+1} \times Y_{9-d}$ while those based on $M$-theory involve backgrounds of the form $AdS_{d+1} \times Y_{10-d}$, where $Y$ are compact spaces.

The examples of the AdS/CFT correspondence discussed in this article are dualities between (super)conformal nonabelian gauge theories and superstrings on $AdS_5 \times Y_5$, where $Y_5$ is a five-dimensional Einstein space (i.e., a space whose Ricci tensor is proportional to the metric, $R_{ij} = 4g_{ij}$). In particular, the most basic (and maximally supersymmetric) such duality relates $\mathcal{N} = 4$ SU($N$) super Yang–Mills (SYM) and type IIB superstring in the curved background $AdS_5 \times S^5$.

There exist special limits where this duality is more tractable than in the general case. If we take the large-$N$ limit while keeping the 't Hooft coupling $\lambda = g_{YM}^2 N$ fixed ($g_{YM}$ is the Yang–Mills coupling strength), then each Feynman graph of the gauge theory carries a topological factor $N^\chi$, where $\chi$ is the Euler characteristic of the graph. The graphs of spherical topology (often called "planar"), to be identified with string tree diagrams, are weighted by $N^2$; the graphs of toroidal topology, to be identified with string one-loop diagrams, by $N^0$, etc. This counting corresponds to the closed-string coupling constant of order $N^{-1}$. Thus, in the large-$N$ limit the gauge theory becomes "planar," and the dual string theory becomes classical. For small $g_{YM}^2 N$, the gauge theory can be studied perturbatively; in this regime the dual string theory has not been very useful because the background becomes highly curved. The real power of the AdS/CFT duality, which already has made it a very useful tool, lies in the fact that, when the gauge theory becomes strongly coupled, the curvature in the dual description becomes small; therefore, classical supergravity provides a systematic starting point for approximating the string theory.

There is a strong motivation for an improved understanding of dualities of this type. In one direction, generalizations of this duality provide the tantalizing hope of a better understanding of quantum chromodynamics (QCD); QCD is a nonabelian gauge theory that describes the strong interactions of mesons, baryons, and glueballs, and has a conformal symmetry which is broken by quantum effects. In the other direction, AdS/CFT suggests that quantum gravity may be understandable as a gauge theory. Understanding the confinement of quarks and gluons that takes place in low-energy QCD and quantizing gravity are well acknowledged to be two of the most important outstanding problems of theoretical physics.

## Some Geometrical Preliminaries

The $d$-dimensional sphere of radius $L$, $S^d$, may be defined by a constraint

$$\sum_{i=1}^{d+1} (X^i)^2 = L^2 \qquad [1]$$

on $d+1$ real coordinates $X^i$. It is a positively curved maximally symmetric space with symmetry group SO$(d+1)$. We will denote the round metric on $S^d$ of unit radius by $d\Omega_d^2$.

The $d$-dimensional anti-de Sitter space, $AdS_d$, may be defined by a constraint

$$(X^0)^2 + (X^d)^2 - \sum_{i=1}^{d-1}(X^i)^2 = L^2 \qquad [2]$$

This constraint shows that the symmetry group of $AdS_d$ is $SO(2, d-1)$. $AdS_d$ is a negatively curved maximally symmetric space, that is, its curvature tensor is related to the metric by

$$R_{abcd} = -\frac{1}{L^2}[g_{ac}g_{bd} - g_{ad}g_{bc}] \qquad [3]$$

Its metric may be written as

$$ds^2_{AdS} = L^2\left(-(y^2+1)dt^2 + \frac{dy^2}{y^2+1} + y^2\, d\Omega^2_{d-2}\right) \qquad [4]$$

where the radial coordinate $y \in [0, \infty)$, and $t$ is defined on a circle of length $2\pi$. This space has closed timelike curves; to eliminate them, we will work with the universal covering space where $t \in (-\infty, \infty)$. The boundary of $AdS_d$, which plays an important role in the AdS/CFT correspondence, is located at infinite $y$. There exists a subspace of $AdS_d$ called the Poincaré wedge, with the metric

$$ds^2 = \frac{L^2}{z^2}\left(dz^2 - (dx^0)^2 + \sum_{i=1}^{d-2}(dx^i)^2\right) \qquad [5]$$

where $z \in [0, \infty)$.

A Euclidean continuation of $AdS_d$ is the Lobachevsky space (hyperboloid), $\boldsymbol{L}_d$. It is obtained by reversing the sign of $(X^d)^2, dt^2$, and $(dx^0)^2$ in [2], [4], and [5], respectively. After this Euclidean continuation, the metrics [4] and [5] become equivalent; both of them cover the entire $\boldsymbol{L}_d$. Another equivalent way of writing the metric is

$$ds^2_L = L^2\left(d\rho^2 + \sinh^2\rho\, d\Omega^2_{d-1}\right) \qquad [6]$$

which shows that the boundary at infinite $\rho$ has the topology of $S^{d-1}$. In terms of the Euclideanized metric [5], the boundary consists of the $\boldsymbol{R}^{d-1}$ at $z = 0$, and a single point at $z = \infty$.

## The Geometry of Dirichlet Branes

Our path toward formulating the $AdS_5/CFT_4$ correspondence requires introduction of Dirichlet branes, or D-branes for short. They are soliton-like "membranes" of various internal dimensionalities contained in type II superstring theories. A Dirichlet $p$-brane (or D$p$ brane) is a $(p+1)$-dimensional hyperplane in $(9+1)$-dimensional spacetime where strings are allowed to end. A D-brane is much like a topological defect: upon touching a D-brane, a closed string can open up and turn into an open string whose ends are free to move along the D-brane. For the endpoints of such a string the $p+1$ longitudinal coordinates satisfy the conventional free (Neumann) boundary conditions, while the $9-p$ coordinates transverse to the D$p$ brane have the fixed (Dirichlet) boundary conditions, hence the origin of the term "Dirichlet brane." The D$p$ brane preserves half of the bulk supersymmetries and carries an elementary unit of charge with respect to the $(p+1)$-form gauge potential from the Ramond–Ramond (RR) sector of type II superstring.

For this article, the most important property of D-branes is that they realize gauge theories on their world volume. The massless spectrum of open strings living on a D$p$ brane is that of a maximally supersymmetric U(1) gauge theory in $p+1$ dimensions. The $9-p$ massless scalar fields present in this supermultiplet are the expected Goldstone modes associated with the transverse oscillations of the D$p$ brane, while the photons and fermions provide the unique supersymmetric completion. If we consider $N$ parallel D-branes, then there are $N^2$ different species of open strings because they can begin and end on any of the D-branes. $N^2$ is the dimension of the adjoint representation of U($N$), and indeed we find the maximally supersymmetric U($N$) gauge theory in this setting.

The relative separations of the D$p$ branes in the $9-p$ transverse dimensions are determined by the expectation values of the scalar fields. We will be interested in the case where all scalar expectation values vanish, so that the $N$ D$p$ branes are stacked on top of each other. If $N$ is large, then this stack is a heavy object embedded into a theory of closed strings which contains gravity. Naturally, this macroscopic object will curve space: it may be described by some classical metric and other background fields including the RR $(p+2)$-form field strength. Thus, we have two very different descriptions of the stack of D$p$ branes: one in terms of the U($N$) supersymmetric gauge theory on its world volume, and the other in terms of the classical RR charged $p$-brane background of the type II closed superstring theory. The relation between these two descriptions is at the heart of the connections between gauge fields and strings that are the subject of this article.

### Coincident D3 Branes

Gauge theories in $3+1$ dimensions play an important role in physics, and as explained above, parallel D3 branes realize a $(3+1)$-dimensional U($N$) SYM

theory. Let us compare a stack of D3 branes with the RR-charged black 3-brane classical solution where the metric assumes the form

$$\begin{aligned} ds^2 =& H^{-1/2}(r)\left[-f(r)(dx^0)^2 + (dx^i)^2\right] \\ &+ H^{1/2}(r)\left[f^{-1}(r)dr^2 + r^2\, d\Omega_5{}^2\right] \end{aligned} \qquad [7]$$

where $i = 1, 2, 3$ and

$$H(r) = 1 + \frac{L^4}{r^4}, \qquad f(r) = 1 - \frac{r_0{}^4}{r^4}$$

The solution also contains an RR self-dual 5-form field strength

$$\begin{aligned} F =& dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3 \wedge d(H^{-1}) \\ &+ 4L^4 \operatorname{vol}(S^5) \end{aligned} \qquad [8]$$

so that the Einstein equation of type IIB supergravity, $R_{\mu\nu} = F_{\mu\alpha\beta\gamma\delta}F_\nu{}^{\alpha\beta\gamma\delta}/96$, is satisfied.

In the extremal limit $r_0 \to 0$, the 3-brane metric becomes

$$\begin{aligned} ds^2 =& \left(1 + \frac{L^4}{r^4}\right)^{-1/2}\left(-(dx^0)^2 + (dx^i)^2\right) \\ &+ \left(1 + \frac{L^4}{r^4}\right)^{1/2}(dr^2 + r^2\, d\Omega_5^2) \end{aligned} \qquad [9]$$

Just like the stack of parallel, ground-state D3 branes, the extremal solution preserves 16 of the 32 supersymmetries present in the type IIB theory. Introducing $z = L^2/r$, one notes that the limiting form of [9] as $r \to 0$ factorizes into the direct product of two smooth spaces, the Poincaré wedge [5] of $AdS_5$, and $S^5$, with equal radii of curvature $L$. The 3-brane geometry may thus be viewed as a semi-infinite throat of radius $L$ which, for $r \gg L$, opens up into flat $(9+1)$-dimensional space. Thus, for $L$ much larger than the string length scale, $\sqrt{\alpha'}$, the entire 3-brane geometry has small curvatures everywhere and is appropriately described by the supergravity approximation to type IIB string theory.

The relation between $L$ and $\sqrt{\alpha'}$ may be found by equating the gravitational tension of the extremal 3-brane classical solution to $N$ times the tension of a single D3 brane:

$$\frac{2}{\kappa^2} L^4 \operatorname{vol}(S^5) = N\frac{\sqrt{\pi}}{\kappa} \qquad [10]$$

where $\operatorname{vol}(S^5) = \pi^3$ is the volume of a unit 5-sphere, and $\kappa = \sqrt{8\pi G}$ is the ten-dimensional gravitational constant. It follows that

$$L^4 = \frac{\kappa}{2\pi^{5/2}} N = g_{YM}^2 N\alpha'^2 \qquad [11]$$

where we used the standard relations $\kappa = 8\pi^{7/2} g_{st}\alpha'^2$ and $g_{YM}^2 = 4\pi g_{st}$ [10]. Thus, the size of the throat in string units is $\lambda^{1/4}$. This remarkable emergence of the 't Hooft coupling from gravitational considerations is at the heart of the success of the AdS/CFT correspondence. Moreover, the requirement $L \gg \sqrt{\alpha'}$ translates into $\lambda \gg 1$: the gravitational approach is valid when the 't Hooft coupling is very strong and the perturbative field-theoretic methods are not applicable.

### Example: Thermal Gauge Theory from Near-Extremal D3 Branes

An important black hole observable is the Bekenstein–Hawking (BH) entropy, which is proportional to the area of the event horizon. For the 3-brane solution [7], the horizon is located at $r = r_0$. For $r_0 > 0$ the 3-brane carries some excess energy $E$ above its extremal value, and the BH entropy is also nonvanishing. The Hawking temperature is then defined by $T^{-1} = \partial S_{BH}/\partial E$.

Setting $r_0 \ll L$ in [9], we obtain a near-extremal 3-brane geometry, whose Hawking temperature is found to be $T = r_0/(\pi L^2)$. The eight-dimensional "area" of the horizon is

$$A_h = (r_0/L)^3 V_3 L^5 \operatorname{vol}(S^5) = \pi^6 L^8 T^3 V_3 \qquad [12]$$

where $V_3$ is the spatial volume of the D3 brane (i.e., the volume of the $x^1$, $x^2$, $x^3$ coordinates). Therefore, the BH entropy is

$$S_{BH} = \frac{2\pi A_h}{\kappa^2} = \frac{\pi^2}{2} N^2 V_3 T^3 \qquad [13]$$

This gravitational entropy of a near-extremal 3-brane of Hawking temperature $T$ is to be identified with the entropy of $\mathcal{N} = 4$ supersymmetric U($N$) gauge theory (which lives on $N$ coincident D3 branes) heated up to the same temperature.

The entropy of a free U($N$) $\mathcal{N} = 4$ supermultiplet – which consists of the gauge field, $6N^2$ massless scalars, and $4N^2$ Weyl fermions – can be calculated using the standard statistical mechanics of a massless gas (the blackbody problem), and the answer is

$$S_0 = \frac{2\pi^2}{3} N^2 V_3 T^3 \qquad [14]$$

It is remarkable that the 3-brane geometry captures the $T^3$ scaling characteristic of a conformal field theory (CFT) (in a CFT this scaling is guaranteed by the extensivity of the entropy and the absence of dimensionful parameters). Also, the $N^2$ scaling indicates the presence of $O(N^2)$ unconfined degrees

of freedom, which is exactly what we expect in the $\mathcal{N}=4$ supersymmetric U($N$) gauge theory. But what is the explanation of the relative factor of 3/4 between $S_{\text{BH}}$ and $S_0$? In fact, this factor is not a contradiction but rather a prediction about the strongly coupled $\mathcal{N}=4$ SYM theory at finite temperature. As we argued above, the supergravity calculation of the BH entropy, [13], is relevant to the $\lambda \to \infty$ limit of the $\mathcal{N}=4$ SU($N$) gauge theory, while the free-field calculation, [14], applies to the $\lambda \to 0$ limit. Thus, the relative factor of 3/4 is not a discrepancy: it relates two different limits of the theory. Indeed, on general field-theoretic grounds, we expect that in the 't Hooft large-$N$ limit, the entropy is given by

$$S = \frac{2\pi^2}{3} N^2 f(\lambda) V_3 T^3 \qquad [15]$$

The function $f$ is certainly not constant: perturbative calculations valid for small $\lambda = g_{\text{YM}}^2 N$ give

$$f(\lambda) = 1 - \frac{3}{2\pi^2}\lambda + \frac{3+\sqrt{2}}{\pi^3}\lambda^{3/2} + \cdots \qquad [16]$$

Thus, the BH entropy in supergravity, [13], is translated into the prediction that

$$\lim_{\lambda\to\infty} f(\lambda) = \frac{3}{4} \qquad [17]$$

## The Essentials of the AdS/CFT Correspondence

The AdS/CFT correspondence asserts a detailed map between the physics of type IIB string theory in the throat of the classical 3-brane geometry, that is, the region $r \ll L$, and the gauge theory living on a stack of D3 branes. As already noted, in this limit $r \ll L$, the extremal D3 brane geometry factors into a direct product of $\text{AdS}_5 \times S^5$. Moreover, the gauge theory on this stack of D3 branes is the maximally supersymmetric $\mathcal{N}=4$ SYM.

Since the horizon of the near-extremal 3-brane lies in the region $r \ll L$, the entropy calculation could have been carried out directly in the throat limit, where $H(r)$ is replaced by $L^4/r^4$. Another way to motivate the identification of the gauge theory with the throat is to think about the absorption of massless particles. In the D-brane description, a particle incident from asymptotic infinity is converted into an excitation of the stack of D-branes, that is, into an excitation of the gauge theory on the world volume. In the supergravity description, a particle incident from the asymptotic (large $r$) region tunnels into the $r \ll L$ region and produces an excitation of the throat. The fact that the two different descriptions of the absorption process give identical cross sections supports the identification of excitations of $\text{AdS}_5 \times S^5$ with the excited states of the $\mathcal{N}=4$ SYM theory.

Maldacena (1998) motivated this correspondence by thinking about the low-energy ($\alpha' \to 0$) limit of the string theory. On the D3 brane side, in this low-energy limit, the interaction between the D3 branes and the closed strings propagating in the bulk vanishes, leaving a pure $\mathcal{N}=4$ SYM theory on the D3 branes decoupled from type IIB superstrings in flat space. Around the classical 3-brane solutions, there are two types of low-energy excitations. The first type propagate in the bulk region, $r \gg L$, and have a cross section for absorption by the throat which vanishes as the cube of their energy. The second type are localized in the throat, $r \leq L$, and find it harder to tunnel into the asymptotically flat region as their energy is taken smaller. Thus, both the D3 branes and the classical 3-brane solution have two decoupled components in the low-energy limit, and in both cases, one of these components is type IIB superstrings in flat space. Maldacena conjectured an equivalence between the other two components.

Immediate support for this identification comes from symmetry considerations. The isometry group of $\text{AdS}_5$ is SO(2, 4), and this is also the conformal group in $3+1$ dimensions. In addition, we have the isometries of $S^5$ which form $\text{SU}(4) \sim \text{SO}(6)$. This group is identical to the R-symmetry of the $\mathcal{N}=4$ SYM theory. After including the fermionic generators required by supersymmetry, the full isometry supergroup of the $\text{AdS}_5 \times S^5$ background is SU(2, 2|4), which is identical to the $\mathcal{N}=4$ superconformal symmetry. We will see that, in theories with reduced supersymmetry, the $S^5$ factor is replaced by other compact Einstein spaces $Y_5$, but $\text{AdS}_5$ is the "universal" factor present in the dual description of any large-$N$ CFT and makes the SO(2, 4) conformal symmetry a geometric one.

The correspondence extends beyond the supergravity limit, and we must think of $\text{AdS}_5 \times Y_5$ as a background of string theory. Indeed, type IIB strings are dual to the electric flux lines in the gauge theory, providing a string-theoretic setup for calculating correlation functions of Wilson loops. Furthermore, if $N \to \infty$ while $g_{\text{YM}}^2 N$ is held fixed and finite, then there are string scale corrections to the supergravity limit (Maldacena 1998, Gubser *et al.* 1998, Witten 1998) which proceed in powers of $\alpha'/L^2 = (g_{\text{YM}}^2 N)^{-1/2}$. For finite $N$, there are also

string loop corrections in powers of $\kappa^2/L^8 \sim N^{-2}$. As expected, with $N \to \infty$ we can take the classical limit of the string theory on $AdS_5 \times Y_5$. However, in order to understand the large-$N$ gauge theory with finite 't Hooft coupling, we should think of $AdS_5 \times Y_5$ as the target space of a two-dimensional sigma model describing the classical string physics.

### Correlation Functions and the Bulk/Boundary Correspondence

A basic premise of the AdS/CFT correspondence is the existence of a one-to-one map between gauge-invariant operators in the CFT and fields (or extended objects) in AdS. Gubser *et al.* (1998) and Witten (1998) formulated precise methods for calculating correlation functions of various operators in a CFT using its dual formulation. A physical motivation for these methods comes from earlier calculations of absorption by 3-branes. When a wave is absorbed, it tunnels from asymptotic infinity into the throat region, and then continues to propagate toward smaller $r$. Let us separate the 3-brane geometry into two regions: $r \gtrsim L$ and $r \lesssim L$. For $r \lesssim L$ the metric is approximately that of $AdS_5 \times S^5$, while for $r \gtrsim L$ it becomes very different and eventually approaches the flat metric. Signals coming in from large $r$ (small $z = L^2/r$) may be considered as disturbing the "boundary" of $AdS_5$ at $r \sim L$, and then propagating into the bulk of $AdS_5$. Discarding the $r \gtrsim L$ part of the 3-brane metric, the gauge theory correlation functions are related to the response of the string theory to boundary conditions at $r \sim L$. It is therefore natural to identify the generating functional of correlation functions in the gauge theory with the string theory path integral subject to the boundary conditions that $\phi(\boldsymbol{x}, z) = \phi_0(\boldsymbol{x})$ at $z = L$ (at $z = \infty$ all fluctuations are required to vanish). In calculating correlation functions in a CFT, we will carry out the standard Euclidean continuation; then on the string theory side, we will work with $L_5$, which is the Euclidean version of $AdS_5$.

More explicitly, we identify a gauge theory quantity $W$ with a string-theory quantity $Z_{\text{string}}$:

$$W[\phi_0(\boldsymbol{x})] = Z_{\text{string}}[\phi_0(\boldsymbol{x})] \qquad [18]$$

$W$ generates the connected Euclidean Green's functions of a gauge-theory operator $\mathcal{O}$,

$$W[\phi_0(\boldsymbol{x})] = \left\langle \exp \int d^4x \phi_0 \mathcal{O} \right\rangle \qquad [19]$$

$Z_{\text{string}}$ is the string theory path integral calculated as a functional of $\phi_0$, the boundary condition on the field $\phi$ related to $\mathcal{O}$ by the AdS/CFT duality. In the large-$N$ limit, the string theory becomes classical which implies

$$Z_{\text{string}} \sim e^{-I[\phi_0(\boldsymbol{x})]} \qquad [20]$$

where $I[\phi_0(\boldsymbol{x})]$ is the extremum of the classical string action calculated as a functional of $\phi_0$. If we are further interested in correlation functions at very large 't Hooft coupling, then the problem of extremizing the classical string action reduces to solving the equations of motion in type IIB supergravity whose form is known explicitly. A simple example of such a calculation is presented in the next subsection.

Our reasoning suggests that from the point of view of the metric [5], the boundary conditions are imposed not quite at $z = 0$, which is the true boundary of $\boldsymbol{L}_5$, but at some finite value $z = \epsilon$. It does not matter which value it is since the metric [5] is unchanged by an overall rescaling of the coordinates $(z, \boldsymbol{x})$; thus, such a rescaling can take $z = L$ into $z = \epsilon$ for any $\epsilon$. The physical meaning of this cutoff is that it acts as a UV regulator in the gauge theory. Indeed, the radial coordinate $z$ is to be considered as the effective energy scale of the gauge theory, and decreasing $z$ corresponds to increasing the energy. A safe method for performing calculations of correlation functions, therefore, is to keep the cutoff on the $z$-coordinate at intermediate stages and remove it only at the end.

### Two-Point Functions and Operator Dimensions

In the following, we present a brief discussion of two-point functions of scalar operators in $CFT_d$. The corresponding field in $\boldsymbol{L}_{d+1}$ is a scalar field of mass $m$ whose Euclidean action is proportional to

$$\frac{1}{2}\int d^d x \, dz \, z^{-d+1}\left[(\partial_z\phi)^2 + \sum_{a=1}^{d}(\partial_a\phi)^2 + \frac{m^2L^2}{z^2}\phi^2\right] \qquad [21]$$

In calculating correlation functions of vertex operators from the AdS/CFT correspondence, the first problem is to reconstruct an on-shell field in $\boldsymbol{L}_{d+1}$ from its boundary behavior. The near-boundary, that is, small $z$, behavior of the classical solution is

$$\begin{aligned}\phi(z, \boldsymbol{x}) \to {} & z^{d-\Delta}\left[\phi_0(\boldsymbol{x}) + O(z^2)\right] \\ & + z^{\Delta}\left[A(\boldsymbol{x}) + O(z^2)\right]\end{aligned} \qquad [22]$$

where $\Delta$ is one of the roots of

$$\Delta(\Delta - d) = m^2L^2 \qquad [23]$$

$\phi_0(\boldsymbol{x})$ is regarded as a "source" in [19] that couples to the dual gauge-invariant operator $\mathcal{O}$ of dimension $\Delta$, while $A(\boldsymbol{x})$ is related to the expectation value,

$$A(\boldsymbol{x}) = \frac{1}{2\Delta - d}\langle \mathcal{O}(\boldsymbol{x})\rangle \qquad [24]$$

It is possible to regularize the Euclidean action to obtain the following value as a functional of the source:

$$I[\phi_0(\boldsymbol{x})] = -(\Delta - (d/2))\pi^{-d/2}\frac{\Gamma(\Delta)}{\Gamma(\Delta - (d/2))} \times \int \mathrm{d}^d\boldsymbol{x} \int \mathrm{d}^d\boldsymbol{x}' \frac{\phi_0(\boldsymbol{x})\phi_0(\boldsymbol{x}')}{|\boldsymbol{x} - \boldsymbol{x}'|^{2\Delta}} \qquad [25]$$

Varying twice with respect to $\phi_0$, we find that the two-point function of the corresponding operator is

$$\langle \mathcal{O}(\boldsymbol{x})\mathcal{O}(\boldsymbol{x}')\rangle = \frac{(2\Delta - d)\Gamma(\Delta)}{\pi^{d/2}\Gamma(\Delta - (d/2))}\frac{1}{|\boldsymbol{x} - \boldsymbol{x}'|^{2\Delta}} \qquad [26]$$

Which of the two roots, $\Delta_+$ or $\Delta_-$, of [23]

$$\Delta_\pm = \frac{d}{2} \pm \sqrt{\frac{d^2}{4} + m^2L^2} \qquad [27]$$

should we choose for the operator dimension? For positive $m^2$, $\Delta_+$ is certainly the right choice: here the other root, $\Delta_-$, is negative. However, it turns out that for

$$-\frac{d^2}{4} < m^2L^2 < -\frac{d^2}{4} + 1 \qquad [28]$$

both roots of [23] may be chosen. Thus, there are two possible CFTs corresponding to the same classical AdS action: in one of them the corresponding operator has dimension $\Delta_+$, while in the other the dimension is $\Delta_-$. We note that $\Delta_-$ is bounded from below by $(d-2)/2$, which is precisely the unitarity bound on dimensions of scalar operators in $d$-dimensional field theory! Thus, the ability to choose dimension $\Delta_-$ is crucial for consistency of the AdS/CFT duality.

Whether string theory on $\mathrm{AdS}_5 \times Y_5$ contains fields with $m^2$ in the range [28] depends on $Y_5$. The example discussed in the next section, $Y_5 = T^{1,1}$, turns out to contain such fields, and the possibility of having dimension $\Delta_-$, [27], is crucial for consistency of the AdS/CFT duality in that case. However, for $Y_5 = \boldsymbol{S}^5$, which is dual to the $\mathcal{N}=4$ large-$N$ SYM theory, there are no such fields and all scalar dimensions are given by [27].

The operators in the $\mathcal{N}=4$ large-$N$ SYM theory naturally break up into two classes: those that correspond to the Kaluza–Klein states of supergravity and those that correspond to massive string states. Since the radius of the $S^5$ is $L$, the masses of the Kaluza–Klein states are proportional to $1/L$. Thus, the dimensions of the corresponding operators are independent of $L$ and therefore also of $\lambda$. On the gauge-theory side, this independence is explained by the fact that the supersymmetry protects the dimensions of certain operators from being renormalized: they are completely determined by the representation under the superconformal symmetry. All families of the Kaluza–Klein states, which correspond to such protected operators, were classified long ago. Correlation functions of such operators in the strong 't Hooft coupling limit may be obtained from the dependence of the supergravity action on the boundary values of corresponding Kaluza–Klein fields, as in [19]. A variety of explicit calculations have been performed for two-, three-, and even four-point functions. The four-point functions are particularly interesting because their dependence on operator positions is not determined by the conformal invariance.

On the other hand, the masses of string excitations are $m^2 = 4n/\alpha'$, where $n$ is an integer. For the corresponding operators the formula [27] predicts that the dimensions do depend on the 't Hooft coupling and, in fact, blow up for large $\lambda = g_{\mathrm{YM}}^2 N$ as $2\lambda^{1/4}\sqrt{n}$.

### Calculation of Wilson Loops

The Wilson loop operator of a nonabelian gauge theory

$$W(\mathcal{C}) = \mathrm{tr}\left[P\exp\left(\mathrm{i}\oint_{\mathcal{C}} A\right)\right] \qquad [29]$$

involves the path-ordered integral of the gauge connection $A$ along a contour $\mathcal{C}$. For $\mathcal{N}=4$ SYM, one typically uses a generalization of this loop operator which incorporates other fields in the $\mathcal{N}=4$ multiplet, the adjoint scalars and fermions. Using a rectangular contour, we can calculate the quark–antiquark potential from the expectation value $\langle W(\mathcal{C})\rangle$. One thinks of the quarks located a distance $L$ apart for a time $T$, yielding

$$\langle W\rangle \sim \mathrm{e}^{-TV(L)} \qquad [30]$$

where $V(L)$ is the potential.

According to Maldacena, and Rey and Yee, the AdS/CFT correspondence relates the Wilson loop expectation value to a sum over string world sheets ending on the boundary of $L_5(z=0)$ along the contour $\mathcal{C}$:

$$\langle W\rangle \sim \int \mathrm{e}^{-S} \qquad [31]$$

where $S$ is the action functional of the string world sheet. In the large 't Hooft coupling limit $\lambda \to \infty$, this path integral may be evaluated using a saddle-point approximation. The leading answer is $\sim e^{-S_0}$, where $S_0$ is the action for the classical solution, which is proportional to the minimal area of the string world sheet in $L_5$ subject to the boundary conditions. The area as currently defined is actually divergent, and to regularize it one must position the contour at $z = \epsilon$ (this is the same type of regulator as used in the definition of correlation functions).

Consider a circular Wilson loop of radius $a$. The action of the corresponding classical string world sheet is

$$S_0 = \sqrt{\lambda}\left(\frac{a}{\epsilon} - 1\right) \quad [32]$$

Subtracting the linearly divergent term, which is proportional to the length of the contour, one finds

$$\ln\langle W \rangle = \sqrt{\lambda} + O(\ln \lambda) \quad [33]$$

a result which has been duplicated in field theory by summing certain classes of rainbow Feynman diagrams in $\mathcal{N} = 4$ SYM. From these sums, one finds

$$\langle W \rangle_{\text{rainbow}} = \frac{2}{\sqrt{\lambda}} I_1\left(\sqrt{\lambda}\right) \quad [34]$$

where $I_1$ is a Bessel function. This formula is one of the few available proposals for extrapolation of an observable from small to large coupling. At large $\lambda$,

$$\langle W \rangle_{\text{rainbow}} \sim \sqrt{\frac{2}{\pi}} \frac{e^{\sqrt{\lambda}}}{\lambda^{3/4}} \quad [35]$$

in agreement with the geometric prediction.

The quark–antiquark potential is extracted from a rectangular Wilson loop of width $L$ and length $T$. After regularizing the divergent contribution to the energy, one finds the attractive potential

$$V(L) = -\frac{4\pi^2\sqrt{\lambda}}{\Gamma(1/4)^4 L} \quad [36]$$

The Coulombic $1/L$ dependence is required by the conformal invariance of the theory. The fact that the potential scales as the square root of the 't Hooft coupling indicates some screening of the charges at large coupling.

## Conformal Field Theories and Einstein Manifolds

Interesting generalizations of the duality between $AdS_5 \times S^5$ and $\mathcal{N} = 4$ SYM with less supersymmetry and more complicated gauge groups can be produced by placing D3 branes at the tip of a Ricci-flat six-dimensional cone $X$ (see Figure 1). The cone metric may be cast in the form

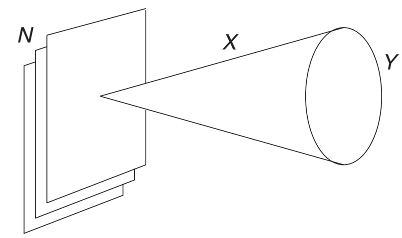



**Figure 1** D3 branes placed at the tip of a Ricci-flat cone $X$.

$$ds_X{}^2 = dr^2 + r^2\, ds_Y{}^2 \quad [37]$$

where $Y$ is the level surface of $X$. In particular, $Y$ is a positively curved Einstein manifold, that is, one for which $R_{ij} = 4g_{ij}$. In order to preserve the $\mathcal{N} = 1$ supersymmetry, $X$ must be a Calabi–Yau space; then $Y$ is defined to be Sasaki–Einstein.

The D3 branes appear as a point in $X$ and span the transverse Minkowski space $\mathbb{R}^{3,1}$. The ten-dimensional metric they produce assumes the form [9], but with the sphere metric $d\Omega_5{}^2$ replaced by the metric on $Y$, $ds_Y^2$. The equality of tensions [10] now requires that

$$L^4 = \frac{\sqrt{\pi}\kappa N}{2\,\text{vol}(Y)} = 4\pi g_s N \alpha'^2 \frac{\pi^3}{\text{vol}(Y)} \quad [38]$$

In the near-horizon limit, $r \to 0$, the geometry factors into $AdS_5 \times Y$. Because the D3 branes are located at a singularity, the gauge theory becomes much more complicated, typically involving a product of several SU($N$) factors coupled to matter in bifundamental representations, often described using a quiver diagram (see Figure 2 for an example).

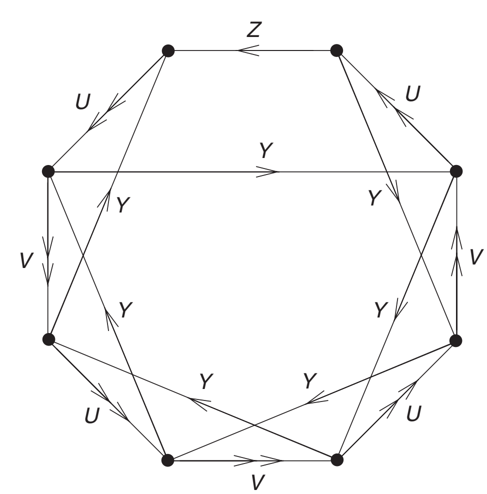



**Figure 2** The quiver for $Y^{4,3}$. Each node corresponds to an SU($N$) gauge group and each arrow to a bifundamental chiral superfield.

The simplest examples of $X$ are orbifolds $\mathbb{C}^3/\Gamma$, where $\Gamma$ is a discrete subgroup of SO(6). Indeed, if $\Gamma \subset \mathrm{SU}(3)$, then $\mathcal{N}=1$ supersymmetry is preserved. The level surface of such an $X$ is $Y = S^5/\Gamma$. In this case, the product structure of the gauge theory can be motivated by thinking about image stacks of D3 branes from the action of $\Gamma$.

The next simplest example of a Calabi–Yau cone $X$ is the conifold which may be described by the following equation in four complex variables:

$$\sum_{a=1}^{4} {z_a}^2 = 0 \qquad [39]$$

Since this equation is symmetric under an overall rescaling of the coordinates, this space is a cone. The level surface $Y$ of the conifold is a coset manifold $T^{1,1} = (\mathrm{SU}(2) \times \mathrm{SU}(2))/\mathrm{U}(1)$. This space has the $\mathrm{SO}(4) \sim \mathrm{SU}(2) \times \mathrm{SU}(2)$ symmetry which rotates the $z$'s, and also the U(1) R-symmetry under $z_a \to e^{i\theta} z_a$. The metric on $T^{1,1}$ is known explicitly; it assumes the form of an $S^1$ bundle over $S^2 \times S^2$.

The supersymmetric field theory on the D3 branes probing the conifold singularity is $\mathrm{SU}(N) \times \mathrm{SU}(N)$ gauge theory coupled to two chiral superfields, $A_i$, in the $(\boldsymbol{N}, \overline{\boldsymbol{N}})$ representation and two chiral superfields, $B_j$, in the $(\overline{\boldsymbol{N}}, \boldsymbol{N})$ representation. The $A$'s transform as a doublet under one of the global SU(2)'s, while the $B$'s transform as a doublet under the other SU(2). Cancelation of the anomaly in the U(1) R-symmetry requires that the $A$'s and the $B$'s each have R-charge 1/2. For consistency of the duality, it is necessary that we add an exactly marginal superpotential which preserves the $\mathrm{SU}(2) \times \mathrm{SU}(2) \times \mathrm{U}(1)_\mathrm{R}$ symmetry of the theory. Since a marginal superpotential has R-charge equal to 2 it must be quartic, and the symmetries fix it uniquely up to overall normalization:

$$W = \epsilon^{ij}\epsilon^{kl} \operatorname{tr} A_i B_k A_j B_l \qquad [40]$$

There are in fact infinite families of Calabi–Yau cones $X$, but there are two problems one faces in studying these generalized AdS/CFT correspondences. The first is geometric: the cones $X$ are not all well understood and only for relatively few do we have explicit metrics. However, it is often possible to calculate important quantities such as the vol($Y$) without knowing the metric. The second problem is gauge theoretic: although many techniques exist, there is no completely general procedure for constructing the gauge theory on a stack of D-branes at an arbitrary singularity.

Let us mention two important classes of Calabi–Yau cones $X$. The first class consists of cones over the so-called $Y^{p,q}$ Sasaki–Einstein spaces. Here, $p$ and $q$ are integers with $p \geq q$. Gauntlett *et al.* (2004) discovered metrics on all the $Y^{p,q}$, and the quiver gauge theories that live on the D-branes probing the singularity are now known. Making contact with the simpler examples discussed above, the $Y^{p,0}$ are orbifolds of $T^{1,1}$ while the $Y^{p,p}$ are orbifolds of $S^5$.

In the second class of cones $X$, a del Pezzo surface shrinks to zero size at the tip of the cone. A del Pezzo surface is an algebraic surface of complex dimension 2 with positive first Chern class. One simple del Pezzo surface is a complex projective space of dimension 2, $\mathbb{P}^2$, which gives rise to the $\mathcal{N}=1$ preserving $S^5/\mathbb{Z}_3$ orbifold. Another simple case is $\mathbb{P}^1 \times \mathbb{P}^1$, which leads to $T^{1,1}/\mathbb{Z}_2$. The remaining del Pezzos surfaces $B_k$ are $\mathbb{P}^2$ blown up at $k$ points, $1 \leq k \leq 8$. The cone where $B_1$ shrinks to zero size has level surface $Y^{2,1}$. Gauge theories for all the del Pezzos have been constructed. Except for the three del Pezzos just discussed, and possibly also for $B_6$, metrics on the cones over these del Pezzos are not known. Nevertheless, it is known that for $3 \leq k \leq 8$, the volume of the Sasaki–Einstein manifold $Y$ associated with $B_k$ is $\pi^3(9-k)/27$.

### The Central Charge

The central charge provides one of the most amazing ways to check the generalized AdS/CFT correspondences. The central charge $c$ and conformal anomaly $a$ can be defined as coefficients of certain curvature invariants in the trace of the stress energy tensor of the conformal gauge theory:

$$\langle T^\alpha_\alpha \rangle = -aE_4 - cI_4 \qquad [41]$$

(The curvature invariants $E_4$ and $I_4$ are quadratic in the Riemann tensor and vanish for Minkowski space.) As discussed above, correlators such as $\langle T_{\mu\nu} \rangle$ can be calculated from supergravity, and one finds

$$a = c = \frac{\pi^3 N^2}{4\,\mathrm{vol}(Y)} \qquad [42]$$

On the gauge-theory side of the correspondence, anomalies completely determine $a$ and $c$:

$$a = \tfrac{3}{32}(3 \operatorname{tr} R^3 - \operatorname{tr} R)$$
$$c = \tfrac{1}{32}(9 \operatorname{tr} R^3 - 5 \operatorname{tr} R) \qquad [43]$$

The trace notation implies a sum over the R-charges of all of the fermions in the gauge theory. (From the geometric knowledge that $a = c$, we can conclude that $\operatorname{tr} R = 0$.)

The R-charges can be determined using the principle of $a$-maximization. For a superconformal gauge theory, the R-charges of the fermions maximize $a$ subject to the constraints that the

Novikov–Shifman–Vainshtein–Zakharov (NSVZ) beta function of each gauge group vanishes and the R-charge of each superpotential term is 2.

For the $Y^{p,q}$ spaces mentioned above, one finds that

$$\mathrm{vol}(Y^{p,q}) = \frac{q^2\left(2p+\sqrt{4p^2-3q^2}\right)}{3p^2\left(3q^2-2p^2+p\sqrt{4p^2-3q^2}\right)}\pi^3 \qquad [44]$$

The gauge theory consists of $p-q$ fields $Z$, $p+q$ fields $Y$, $2p$ fields $U$, and $2q$ fields $V$. These fields all transform in the bifundamental representation of a pair of SU($N$) gauge groups (the quiver diagram for $Y^{4,3}$ is given in **Figure 2**). The NSVZ beta function and superpotential constraints determine the R-charges up to two free parameters $x$ and $y$. Let $x$ be the R-charge of $Z$ and $y$ the R-charge of $Y$. Then the $U$ have R-charge $1-(1/2)(x+y)$ and the $V$ have R-charge $1+(1/2)(x-y)$.

The technique of $a$ maximization leads to the result

$$x = \frac{1}{3q^2}\left(-4p^2+2pq+3q^2+(2p-q)\sqrt{4p^2-3q^2}\right)$$

$$y = \frac{1}{3q^2}\left(-4p^2-2pq+3q^2+(2p+q)\sqrt{4p^2-3q^2}\right)$$

Thus, as calculated by Benvenuti *et al.* (2004) and Bertolini *et al.* (2004)

$$a(Y^{p,q}) = \frac{\pi^3 N^2}{4\,\mathrm{vol}(Y^{p,q})} \qquad [45]$$

in remarkable agreement with the prediction [42] of the AdS/CFT duality.

### A Path to a Confining Theory

There exists an interesting way of breaking the conformal invariance for spaces $Y$ whose topology includes an $S^2$ factor (examples of such spaces include $T^{1,1}$ and $Y^{p,q}$, which are topologically $S^2 \times S^3$). At the tip of the cone over $Y$, one may add $M$ wrapped D5 branes to the $N$ D3 branes. The gauge theory on such a combined stack is no longer conformal; it exhibits a novel pattern of quasiperiodic renormalization group flow, called a duality cascade.

To date, the most extensive study of a theory of this type has been carried out for the conifold, where one finds an $\mathcal{N}=1$ supersymmetric SU($N$) × SU($N+M$) theory coupled to chiral superfields $A_1$, $A_2$ in the $(\mathbf{N}, \overline{\mathbf{N}+\mathbf{M}})$ representation, and $B_1$, $B_2$ in the $(\overline{\mathbf{N}}, \mathbf{N}+\mathbf{M})$ representation. D5 branes source RR 3-form flux; hence, the supergravity dual of this theory has to include $M$ units of this flux. Klebanov and Strassler (2000) found an exact nonsingular supergravity solution incorporating the 3-form and the 5-form RR field strengths, and their back-reaction on the geometry. This back-reaction creates a "geometric transition" to the deformed conifold

$$\sum_{a=1}^{4} z_a^2 = \epsilon^2 \qquad [46]$$

and introduces a "warp factor" so that the full ten-dimensional geometry has the form

$$\mathrm{d}s_{10}{}^2 = h^{-1/2}(\tau)(-(\mathrm{d}x^0)^2 + (\mathrm{d}x^i)^2) + h^{1/2}(\tau)\,\mathrm{d}\tilde{s}_6{}^2 \qquad [47]$$

where $\mathrm{d}\tilde{s}_6{}^2$ is the Calabi–Yau metric of the deformed conifold, which is known explicitly.

The field-theoretic interpretation of this solution is unconventional. After a finite amount of RG flow, the SU($N+M$) group undergoes a Seiberg duality transformation. After this transformation, and an interchange of the two gauge groups, the new gauge theory is SU($\tilde{N}$) × SU($\tilde{N}+M$) with the same matter and superpotential, and with $\tilde{N}=N-M$. The self-similar structure of the gauge theory under the Seiberg duality is the crucial fact that allows this pattern to repeat many times. If $N=(k+1)M$, where $k$ is an integer, then the duality cascade stops after $k$ steps, and we find SU($M$) × SU($2M$) gauge theory. This IR gauge theory exhibits a multitude of interesting effects visible in the dual supergravity background. One of them is confinement, which follows from the fact that the warp factor $h$ is finite and nonvanishing at the smallest radial coordinate, $\tau=0$. The methods presented in the section "Calculation of Wilson loops," then imply that the quark–antiquark potential grows linearly at large distances. Other notable IR effects are chiral symmetry breaking and the Goldstone mechanism. Particularly interesting is the appearance of an entire "baryonic branch" of the moduli space in the gauge theory, whose existence has been demonstrated also in the dual supergravity language.

## Conclusions

This article tries to present a logical path from studying gravitational properties of D-branes to the formulation of an exact duality between conformal field theories and string theory in anti-de Sitter backgrounds, and also sketches some methods for breaking the conformal symmetry. Due to space limitations, many aspects and applications of the AdS/CFT correspondence have been omitted. At the moment, practical applications of this duality are limited mainly to very strongly coupled, large-$N$ gauge theories, where the dual string description is well approximated by classical supergravity. To understand the implications of the duality for more general parameters, it is necessary to find better

methods for attacking the world sheet approach to string theories in anti-de Sitter backgrounds with RR background fields turned on. When such methods are found, it is likely that the material presented here will have turned out to be just a tiny tip of a monumental iceberg of dualities between fields and strings.

## Acknowledgments

The authors are very grateful to all their collaborators on gauge/string duality for their valuable input over many years. The research of I R Klebanov is supported in part by the National Science Foundation (NSF) grant no. PHY-0243680. The research of C P Herzog is supported in part by the NSF under grant no. PHY99-07949. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

*See also:* Brane Construction of Gauge Theories; Branes and Black Hole Statistical Mechanics; Einstein Equations: Exact Solutions; Gauge Theories from Strings; Large-*N* and Topological Strings; Large-*N* Dualities; Mirror Symmetry: A Geometric Survey; Quantum Chromodynamics; Quantum Field Theory in Curved Spacetime; Superstring Theories.

## Further Reading

Aharony O, Gubser SS, Maldacena JM, Ooguri H, and Oz Y (2000) Large N field theories, string theory and gravity. *Physics Reports* 323: 183 (arXiv:hep-th/9905111).

Benvenuti S, Franco S, Hanany A, Martelli D, and Sparks J (2005) An infinite family of superconformal quiver gauge theories with Sasaki–Einstein duals. *JHEP* 0506: 064 (arXiv:hep-th/0411264).

Bertolini M, Bigazzi F, and Cotrone AL (2004) New checks and subtleties for AdS/CFT and a-maximization. *JHEP* 0412: 024 (arXiv:hep-th/0411249).

Bigazzi F, Cotrone AL, Petrini M, and Zaffaroni A (2002) Supergravity duals of supersymmetric four dimensional gauge theories. *Rivista del Nuovo Cimento* 25N12: 1 (arXiv:hep-th/0303191).

D'Hoker E and Freedman DZ (2002) Supersymmetric gauge theories and the AdS/CFT correspondence, arXiv:hep-th/0201253.

Gauntlett J, Martelli D, Sparks J, and Waldram D (2004) Sasaki–Einstein metrics on $S^2 \times S^3$. *Advances in Theoretical Mathematics in Physics* 8: 711 (arXiv:hep-th/0403002).

Gubser SS, Klebanov IR, and Polyakov AM (1998) Gauge theory correlators from noncritical string theory. *Physics Letters B* 428: 105 (hep-th/9802109).

Herzog CP, Klebanov IR, and Ouyang P (2002) D-branes on the conifold and $N=1$ gauge/gravity dualities, arXiv:hep-th/0205100.

Klebanov IR (2000) TASI lectures: introduction to the AdS/CFT correspondence, arXiv:hep-th/0009139.

Klebanov IR and Strassler MJ (2000) Supergravity and a confining gauge theory: Duality cascades and $\chi$-resolution of naked singularities. *JHEP* 0008: 052 (arXiv:hep-th/0007191).

Maldacena J (1998) The large N limit of superconformal field theories and supergravity. *Advances in Theoretical and Mathematical Physics* 2: 231 (hep-th/9711200).

Maldacena JM (1998) Wilson loops in large N field theories. *Physics Review Letters* 80: 4859 (arXiv:hep-th/9803002).

Polchinski J (1998) *String Theory*. Cambridge: Cambridge University Press.

Polyakov AM (1999) The wall of the cave. *International Journal of Modern Physics A* 14: 645.

Rey SJ and Yee JT (2001) Macroscopic strings as heavy quarks in large N gauge theory and anti-de Sitter supergravity. *European Physics Journal C* 22: 379 (arXiv:hep-th/9803001).

Semenoff GW and Zarembo K (2002) Wilson loops in SYM theory: from weak to strong coupling. *Nuclear Physics Proceeding Supplements* 108: 106 (arXiv:hep-th/0202156).

Strassler MJ The duality cascade, TASI 2003 lectures, arXiv:hep-th/0505153.

Witten E (1998) Anti-de Sitter space and holography. *Advances in Theoretical and Mathematical Physics* 2: 253 (hep-th/9802150).

# Affine Quantum Groups

**G W Delius and N MacKay**, University of York, York, UK



Affine quantum groups are certain pseudoquasitriangular Hopf algebras that arise in mathematical physics in the context of integrable quantum field theory, integrable quantum spin chains, and solvable lattice models. They provide the algebraic framework behind the spectral parameter dependent Yang–Baxter equation

$$R_{12}(u)R_{13}(u+v)R_{23}(v) = R_{23}(v)R_{13}(u+v)R_{12}(u) \qquad [1]$$

One can distinguish three classes of affine quantum groups, each leading to a different dependence of the $R$-matrices on the spectral parameter $u$: Yangians lead to rational $R$-matrices, quantum affine algebras lead to trigonometric R-matrices, and elliptic quantum groups lead to elliptic $R$-matrices. We will mostly concentrate on the quantum affine algebras but many results hold similarly for the other classes.

After giving mathematical details about quantum affine algebras and Yangians in the first two sections, we describe how these algebras arise in different areas of mathematical physics in the three following sections. We end with a description of boundary quantum groups which extend the formalism to the boundary Yang–Baxter (reflection) equation.

## Quantum Affine Algebras

### Definition

A quantum affine algebra $U_q(\hat{\mathfrak{g}})$ is a quantization of the enveloping algebra $U(\hat{\mathfrak{g}})$ of an affine Lie algebra (Kac–Moody algebra) $\hat{\mathfrak{g}}$. So we start by introducing affine Lie algebras and their enveloping algebras before proceeding to give their quantizations.

Let $\mathfrak{g}$ be a semisimple finite-dimensional Lie algebra over $\mathbb{C}$ of rank $r$ with Cartan matrix $(a_{ij})_{i,j=1,\ldots,r}$, symmetrizable via positive integers $d_i$, so that $d_i a_{ij}$ is symmetric. In terms of the simple roots $\alpha_i$, we have

$$a_{ij} = 2\frac{\alpha_i \cdot \alpha_j}{|\alpha_i|^2} \qquad \text{and} \qquad d_i = \frac{|\alpha_i|^2}{2}.$$

We can introduce an $\alpha_0 = \sum_{i=1}^r n_i \alpha_i$ in such a way that the extended Cartan matrix $(a_{ij})_{i,j=0,\ldots,r}$ is of affine type – that is, it is positive semidefinite of rank $r$. The integers $n_i$ are referred to as Kac indices. Choosing $\alpha_0$ to be the highest root of $\mathfrak{g}$ leads to an untwisted affine Kac–Moody algebra while choosing $\alpha_0$ to be the highest short root of $\mathfrak{g}$ leads to a twisted affine Kac–Moody algebra.

One defines the affine Lie algebra $\hat{\mathfrak{g}}$ corresponding to this affine Cartan matrix as the Lie algebra (over $\mathbb{C}$) with generators $H_i, E_i^{\pm}$ for $i = 0, 1, \ldots, r$ and $D$ with relations

$$\begin{gathered}\left[H_i, E_j^{\pm}\right] = \pm a_{ij} E_i^{\pm}; \qquad [H_i, H_j] = 0 \\ \left[E_i^{+}, E_j^{-}\right] = \delta_{ij} H_i \\ [D, H_i] = 0, \qquad \left[D, E_i^{\pm}\right] = \pm \delta_{i,0} E_i^{\pm} \\ \sum_{k=0}^{1-a_{ij}} (-1)^k \binom{1-a_{ij}}{k} \left(E_i^{\pm}\right)^k E_j^{\pm} \left(E_i^{\pm}\right)^{1-a_{ij}-k} = 0, \quad i \neq j\end{gathered} \qquad [2]$$

The $E_i^{\pm}$ are referred to as Chevalley generators and the last set of relations are known as Serre relations. The generator $D$ is known as the canonical derivation. We will denote the algebra obtained by dropping the generator $D$ by $\hat{\mathfrak{g}}'$.

In applications to physics, the affine Lie algebra $\hat{\mathfrak{g}}$ often occurs in an isomorphic form as the loop Lie algebra $\mathfrak{g}[z, z^{-1}] \oplus \mathbb{C} \cdot c$ with Lie product (for untwisted $\hat{\mathfrak{g}}$)

$$\begin{aligned}&[Xz^k, Yz^l] = [X, Y]z^{k+l} + \delta_{k,-l}(X, Y)c, \\ &\quad \text{for } X, Y \in \mathfrak{g}, \; k, l \in \mathbb{Z}\end{aligned} \qquad [3]$$

and $c$ being the central element.

The universal enveloping algebra $U(\hat{\mathfrak{g}})$ of $\hat{\mathfrak{g}}$ is the unital algebra over $\mathbb{C}$ with generators $H_i, E_i^{\pm}$ for $i = 0, 1, \ldots, r$ and $D$ and with relations given by [2] where now $[\,,\,]$ stands for the commutator instead of the Lie product.

To define the quantization of $U(\hat{\mathfrak{g}})$, one can either define $U_h(\hat{\mathfrak{g}})$ (Drinfeld 1985) as an algebra over the ring $\mathbb{C}[[h]]$ of formal power series over an indeterminate $h$ or one can define $U_q(\hat{\mathfrak{g}})$ (Jimbo 1985) as an algebra over the field $\mathbb{Q}(q)$ of rational functions of $q$ with coefficients in $\mathbb{Q}$. We will present $U_h(\hat{\mathfrak{g}})$ first.

The quantum affine algebra $U_h(\hat{\mathfrak{g}})$ is the unital algebra over $\mathbb{C}[[h]]$ topologically generated by $H_i, E_i^{\pm}$ for $i = 0, 1, \ldots, r$ and $D$ with relations

$$\begin{gathered}\left[H_i, E_j^{\pm}\right] = \pm a_{ij} E_i^{\pm}; \qquad [H_i, H_j] = 0 \\ \left[E_i^{+}, E_j^{-}\right] = \delta_{ij} \frac{q_i^{H_i} - q_i^{-H_i}}{q_i - q_i^{-1}} \\ [D, H_i] = 0, \qquad \left[D, E_i^{\pm}\right] = \pm \delta_{i,0} E_i^{\pm} \\ \sum_{k=0}^{1-a_{ij}} (-1)^k \begin{bmatrix} 1-a_{ij} \\ k \end{bmatrix}_{q_i} \left(E_i^{\pm}\right)^k E_j^{\pm} \left(E_i^{\pm}\right)^{1-a_{ij}-k} = 0, \quad i \neq j\end{gathered} \qquad [4]$$

where $q_i = q^{d_i}$ and $q = e^h$. The $q$-binomial coefficients are defined by

$$[n]_q = \frac{q^n - q^{-n}}{q - q^{-1}} \qquad [5]$$

$$[n]_q! = [n]_q \cdot [n-1]_q \ldots [2]_q [1]_q \qquad [6]$$

$$\begin{bmatrix} m \\ n \end{bmatrix}_q = \frac{[m]_q!}{[n]_q! [m-n]_q!} \qquad [7]$$

The quantum affine algebra $U_h(\hat{\mathfrak{g}})$ is a Hopf algebra with coproduct

$$\begin{aligned}\Delta(D) &= D \otimes 1 + 1 \otimes D \\ \Delta(H_i) &= H_i \otimes 1 + 1 \otimes H_i \\ \Delta\left(E_i^{\pm}\right) &= E_i^{\pm} \otimes q_i^{-H_i/2} + q_i^{H_i/2} \otimes E_i^{\pm}\end{aligned} \qquad [8]$$

antipode

$$\begin{gathered}S(D) = -D, \qquad S(H_i) = -H_i \\ S\left(E_i^{\pm}\right) = -q_i^{\mp 1} E_i^{\pm}\end{gathered} \qquad [9]$$

and co-unit

$$\epsilon(D) = \epsilon(H_i) = \epsilon\left(E_i^{\pm}\right) = 0 \qquad [10]$$

It is easy to see that the classical enveloping algebra $U(\hat{\mathfrak{g}})$ can be obtained from the above by setting $h = 0$, or more formally,

$$U_h(\hat{\mathfrak{g}}) / h U_h(\hat{\mathfrak{g}}) = U(\hat{\mathfrak{g}})$$

We can also define the quantum affine algebra $U_q(\hat{\mathfrak{g}})$ as the algebra over $\mathbb{Q}(q)$ with generators $K_i, E_i^{\pm}, D$ for $i = 0, 1, \ldots, r$ and relations that are

obtained from the ones given above for $U_h(\hat{\mathfrak{g}})$ by setting

$$q_i^{H_i/2} = K_i, \quad i = 0, \ldots, r \qquad [11]$$

One can go further to an algebraic formulation over $\mathbb{C}$ in which $q$ is a complex number (with some points including $q = 0$ not allowed). This has the advantage that it becomes possible to specialize, for example, to $q$ a root of unity, where special phenomena occur.

### Representations

For applications in physics, the finite-dimensional representations of $U_h(\hat{\mathfrak{g}}')$ are the most interesting. As will be explained in later sections, these occur, for example, as particle multiplets in 2D quantum field theory or as spin Hilbert spaces in quantum spin chains. In the next subsection, we will use them to derive matrix solutions to the Yang–Baxter equation.

While for a nonaffine quantum algebra $U_h(\mathfrak{g})$ the ring of representations is isomorphic to that of the classical enveloping algebra $U(\mathfrak{g})$ (because in fact the algebras are isomorphic, as Drinfeld has pointed out), the corresponding fact is no longer true for affine quantum groups, except in the case $\hat{\mathfrak{g}} = a_n^{(1)} = \widehat{\mathfrak{sl}_{n+1}}$.

For the classical enveloping algebras $U(\hat{\mathfrak{g}}')$, any finite-dimensional representation of $U(\mathfrak{g})$ also carries a finite-dimensional representation of $U(\hat{\mathfrak{g}}')$. In the quantum case, however, in general, an irreducible representation of $U_h(\hat{\mathfrak{g}}')$ reduces to a sum of representations of $U_h(\mathfrak{g})$.

To classify the finite-dimensional representations of $U_h(\hat{\mathfrak{g}}')$, it is necessary to use a different realization of $U_h(\hat{\mathfrak{g}}')$ that looks more like a quantization of the loop algebra realization [3] than the realization in terms of Chevalley generators. In terms of the generators in this alternative realization, which we do not give here because of its complexity, the finite-dimensional representations can be viewed as pseudo-highest-weight representations. There is a set of $r$ "fundamental" representations $V^a$, $a = 1, \ldots r$, each containing the corresponding $U_h(\mathfrak{g})$ fundamental representation as a component, from the tensor products of which all the other finite-dimensional representations may be constructed. The details can be found in Chari and Pressley (1994).

Given some representation $\rho : U_h(\hat{\mathfrak{g}}') \to \mathrm{End}(V)$, we can introduce a parameter $\lambda$ with the help of the automorphism $\tau_\lambda$ of $U_h(\hat{\mathfrak{g}}')$ generated by $D$ and given by

$$\begin{aligned} \tau_\lambda(E_i^\pm) &= \lambda^{\pm s_i} E_i^\pm \\ \tau_\lambda(H_i) &= H_i \end{aligned} \qquad i = 0, \ldots, r \qquad [12]$$

Different choices for the $s_i$ correspond to different gradations. Commonly used are the "homogeneous gradation," $s_0 = 1, s_1 = \cdots = s_r = 0$, and the "principal gradation," $s_0 = s_1 = \cdots = s_r = 1$. We shall also need the "spin gradation" $s_i = d_i^{-1}$. The representations

$$\rho_\lambda = \rho \circ \tau_\lambda$$

play an important role in applications to integrable models where $\lambda$ is referred to as the (multiplicative) spectral parameter. In applications to particle scattering introduced in a later section, it is related to the rapidity of the particle. The generator $D$ can be realized as an infinitesimal scaling operator on $\lambda$ and thus plays the role of the Lorentz boost generator.

The tensor product representations $\rho_\lambda^a \otimes \rho_\mu^b$ are irreducible generically but become reducible for certain values of $\lambda/\mu$, a fact which again is important in applications (fusion procedure, particle-bound states).

### *R*-Matrices

A Hopf algebra $A$ is said to be "almost cocommutative" if there exists an invertible element $\mathcal{R} \in A \otimes A$ such that

$$\mathcal{R}\Delta(x) = (\sigma \circ \Delta(x))\mathcal{R}, \quad \text{for all } x \in A \qquad [13]$$

where $\sigma : x \otimes y \mapsto y \otimes x$ exchanges the two factors in the coproduct. In a quasitriangular Hopf algebra, this element $\mathcal{R}$ satisfies

$$\begin{aligned} (\Delta \otimes \mathrm{id})(\mathcal{R}) &= \mathcal{R}_{13}\mathcal{R}_{23} \\ (\mathrm{id} \otimes \Delta)(\mathcal{R}) &= \mathcal{R}_{13}\mathcal{R}_{12} \end{aligned} \qquad [14]$$

and is known as the "universal $R$-matrix" (*see* Hopf Algebras and $q$-Deformation Quantum Groups). As a consequence of [13] and [14], it automatically satisfies the Yang–Baxter equation

$$\mathcal{R}_{12}\mathcal{R}_{13}\mathcal{R}_{23} = \mathcal{R}_{23}\mathcal{R}_{13}\mathcal{R}_{12} \qquad [15]$$

For technical reasons, to do with the infinite number of root vectors of $\hat{\mathfrak{g}}$, the quantum affine algebra $U_h(\hat{\mathfrak{g}})$ does not possess a universal $R$-matrix that is an element of $U_h(\hat{\mathfrak{g}}) \otimes U_h(\hat{\mathfrak{g}})$. However, as pointed out by Drinfeld (1985), it possesses a pseudouniversal $R$-matrix $\mathcal{R}(\lambda) \in (U_h(\hat{\mathfrak{g}}') \otimes U_h(\hat{\mathfrak{g}}'))((\lambda))$. The $\lambda$ is related to the automorphism $\tau_\lambda$ defined in [12]. When using the homogeneous gradation, $\mathcal{R}(\lambda)$ is a formal power series in $\lambda$.

When the pseudouniversal $R$-matrix is evaluated in the tensor product of any two indecomposable finite-dimensional representations $\rho_1$ and $\rho_2$, one obtains a numerical $R$-matrix

$$R^{12}(\lambda) = (\rho^1 \otimes \rho^2)\mathcal{R}(\lambda) \qquad [16]$$

The entries of these numerical $R$-matrices are rational functions of the multiplicative spectral parameter $\lambda$ but when written in terms of the additive spectral parameter $u = \log(\lambda)$ they are trigonometric functions of $u$ and satisfy the Yang–Baxter equation in the form given in [1]. The matrix

$$\check{R}^{12}(\lambda) = \sigma \circ R^{12}(\lambda)$$

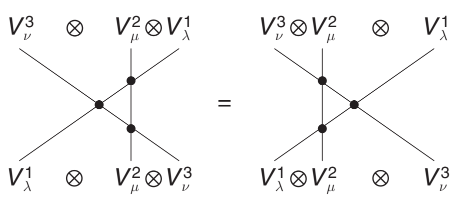
satisfies the intertwining relation

$$\begin{aligned} &\check{R}^{12}(\lambda/\mu) \cdot \left(\rho_\lambda^1 \otimes \rho_\mu^2\right)(\Delta(x)) \\ &\quad = \left(\rho_\mu^2 \otimes \rho_\lambda^1\right)(\Delta(x)) \cdot \check{R}^{12}(\lambda/\mu) \end{aligned} \qquad [17]$$

for any $x \in U_h(\hat{\mathfrak{g}}')$. It follows from the irreducibility of the tensor product representations that these $R$-matrices satisfy the Yang–Baxter equations

$$\begin{aligned} &(\mathrm{id} \otimes \check{R}^{23}(\mu/\nu))(\check{R}^{13}(\lambda/\nu) \otimes \mathrm{id})(\mathrm{id} \otimes \check{R}^{12}(\lambda/\mu)) \\ &\quad = (\check{R}^{12}(\lambda/\mu) \otimes \mathrm{id})(\mathrm{id} \otimes \check{R}^{13}(\lambda/\nu)) \\ &\qquad \times (\check{R}^{23}(\mu/\nu) \otimes \mathrm{id}) \end{aligned} \qquad [18]$$

or, graphically,




Explicit formulas for the pseudouniversal $R$-matrices were found by Khoroshkin and Tolstoy. However, these are difficult to evaluate explicitly in specific representations so that in practice it is easiest to find the numerical $R$-matrices $\check{R}^{ab}(\lambda)$ by solving the intertwining relation [17]. It should be stressed that solving the intertwining relation, which is a linear equation for the $R$-matrix, is much easier than directly solving the Yang–Baxter equation, a cubic equation.

## Yangians

As remarked by Drinfeld (1986), for untwisted $\hat{\mathfrak{g}}$ the quantum affine algebra $U_h(\hat{\mathfrak{g}}')$ degenerates as $h \to 0$ into another quasipseudotriangular Hopf algebra, the "Yangian" $Y(\mathfrak{g})$ (Drinfeld 1985). It is associated with $R$-matrices which are rational functions of the additive spectral parameter $u$. Its representation ring coincides with that of $U_h(\hat{\mathfrak{g}}')$.

Consider a general presentation of a Lie algebra $\mathfrak{g}$, with generators $I_a$ and structure constants $f_{abc}$, so that

$$[I_a, I_b] = f_{abc} I_c, \quad \Delta(I_a) = I_a \otimes 1 + 1 \otimes I_a$$

(with summation over repeated indices). The Yangian $Y(\mathfrak{g})$ is the algebra generated by these and a second set of generators $J_a$ satisfying

$$\begin{aligned} &[I_a, J_b] = f_{abc} J_c \\ &\Delta(J_a) = J_a \otimes 1 + 1 \otimes J_a + \tfrac{1}{2} f_{abc} I_c \otimes I_b \end{aligned}$$

The requirement that $\Delta$ be a homomorphism imposes further relations:

$$[J_a, [J_b, I_c]] - [I_a, [J_b, J_c]] = \alpha_{abcdeg} \{I_d, I_e, I_g\}$$

and

$$\begin{aligned} &[[J_a, J_b], [I_l, J_m]] + [[J_l, J_m], [I_a, J_b]] \\ &\quad = \left(\alpha_{abcdeg} f_{lmc} + \alpha_{lmcdeg} f_{abc}\right)\{I_d, I_e, J_g\} \end{aligned}$$

where

$$\alpha_{abcdeg} = \frac{1}{24} f_{adi} f_{bej} f_{cgk} f_{ijk}, \quad \{x_1, x_2, x_3\} = \sum_{i \neq j \neq k} x_i x_j x_k$$

When $\mathfrak{g} = \mathfrak{sl}_2$ the first of these is trivial, while for $\mathfrak{g} \neq \mathfrak{sl}_2$ the first implies the second. The co-unit is $\epsilon(I_a) = \epsilon(J_a) = 0$; the antipode is $s(I_a) = -I_a$, $s(J_a) = -J_a + (1/2) f_{abc} I_c I_b$. The Yangian may be obtained from $U_h(\hat{\mathfrak{g}}')$ by expanding in powers of $h$. For the precise relationship, see Drinfeld (1985) and MacKay (2005). In the spin gradation, the automorphism [12] generated by $D$ descends to $Y(\mathfrak{g})$ as $I_a \mapsto I_a$, $J_a \mapsto J_a + u I_a$.

There are two other realizations of $Y(\mathfrak{g})$. The first (see, for example, Molev 2003) defines $Y(\mathfrak{gl}_n)$ directly from

$$R(u - v) T_1(u) T_2(v) = T_2(v) T_1(u) R(u - v)$$

where $T_1(u) = T(u) \otimes \mathrm{id}$, $T_2(v) = \mathrm{id} \otimes T(v)$, and

$$\begin{aligned} &T(u) = \sum_{i,j=1}^{n} t_{ij}(u) \otimes e_{ij} \\ &t_{ij}(u) = \delta_{ij} + I_{ij} u^{-1} + J_{ij} u^{-2} + \cdots \end{aligned}$$

where $e_{ij}$ are the standard matrix units for $\mathfrak{gl}_n$. The rational $R$-matrix for the $n$-dimensional representation of $\mathfrak{gl}_n$ is

$$R(u - v) = 1 - \frac{P}{u - v}, \quad \text{where } P = \sum_{i,j=1}^{n} e_{ij} \otimes e_{ji}$$

is the transposition operator. $Y(\mathfrak{gl}_n)$ is then defined to be the algebra generated by $I_{ij}, J_{ij}$, and must be quotiented by the "quantum determinant" at its center to define $Y(\mathfrak{sl}_n)$. The coproduct takes a particularly simple form,

$$\Delta(t_{ij}(u)) = \sum_{k=1}^{n} t_{ik}(u) \otimes t_{kj}(u)$$

Here we do not give explicitly the third realization, namely Drinfeld's "new" realization of $Y(\mathfrak{g})$ (Drinfeld 1988), but we remark that it was in this presentation that Drinfeld found a correspondence between certain sets of polynomials and finite-dimensional irreducible representations of $Y(\mathfrak{g})$, thus classifying these (although not thereby deducing their dimension or constructing the action of $Y(\mathfrak{g})$). As remarked earlier, the structure is as in the earlier section: $Y(\mathfrak{g})$ representations are in general $\mathfrak{g}$-reducible, and there is a set of $r$ fundamental $Y(\mathfrak{g})$-representations, containing the fundamental $\mathfrak{g}$-representations as components, from which all other representations can be constructed.

## Origins in the Quantum Inverse-Scattering Method

Quantum affine algebras for general $\hat{\mathfrak{g}}$ first appear in Drinfeld (1985, 1986) and Jimbo (1985, 1986), but they have their origin in the "quantum inverse-scattering method" (QISM) of the St. Petersburg school, and the essential features of $U_h(\widehat{\mathfrak{sl}_2})$ first appear in Kulish and Reshetikhin (1983). In this section, we explain how the quantization of the Lax-pair description of affine Toda theory led to the discovery of the $U_h(\hat{\mathfrak{g}})$ coproduct, commutation relations, and $R$-matrix. We use the normalizations of Jimbo (1986), in which the $H_i$ are rescaled so that the Cartan matrix $a_{ij} = \alpha_i \cdot \alpha_j$ is symmetric.

We begin with the affine Toda field equations

$$\partial^\mu \partial_\mu \phi_i = -\frac{m^2}{\beta} \sum_{j=1}^{r} \left( e^{\beta a_{ij} \phi_j} - n_i e^{\beta \alpha_0 . \alpha_j \phi_j} \right)$$

an integrable model in $\mathbb{R}^{1+1}$ of $r$ real scalar fields $\phi_i(x,t)$ with a mass parameter $m$ and coupling constant $\beta$. Equivalently, we may write $[\partial_x + L_x, \partial_t + L_t] = 0$ for the Lax pair

$$L_x(x,t) = \frac{\beta}{2} \sum_{i=1}^{r} H_i \partial_t \phi_i + \frac{m}{2} \sum_{i,j=1}^{r} e^{(\beta/2) a_{ij} \phi_j} \left( E_i^+ + E_i^- \right) + \frac{m}{2} \sum_{j=1}^{r} e^{(\beta/2) a_{0j} \phi_j} \left( \lambda E_0^+ + \frac{1}{\lambda} E_0^- \right)$$

$$L_t(x,t) = \frac{\beta}{2} \sum_{i=1}^{r} H_i \partial_x \phi_i + \frac{m}{2} \sum_{i,j=1}^{r} e^{(\beta/2) a_{ij} \phi_j} \left( E_i^+ - E_i^- \right) + \frac{m}{2} \sum_{j=1}^{r} e^{(\beta/2) a_{0j} \phi_j} \left( \lambda E_0^+ - \frac{1}{\lambda} E_0^- \right)$$

with arbitrary $\lambda \in \mathbb{C}$. The classical integrability of the system is seen in the existence of $r(\lambda, \lambda')$ such that

$$\{T(\lambda) \otimes T(\lambda')\} = [r(\lambda, \lambda'), T(\lambda) \otimes T(\lambda')]$$

where $T(\lambda) = T(-\infty, \infty; \lambda)$ and $T(x,y;\lambda) = \boldsymbol{P} \exp(\int_x^y L(\xi;\lambda)\, d\xi)$. Taking the trace of this relation gives an infinity of charges in involution.

Quantization is problematic, owing to divergences in $T$. The QISM regularizes these by putting the model on a lattice of spacing $\Delta$, defining the lattice Lax operator to be

$$L_n(\lambda) = T((n-1/2)\Delta, (n+1/2)\Delta; \lambda) = \boldsymbol{P} \exp\left( \int_{(n-(1/2))\Delta}^{(n+(1/2))\Delta} L(\xi;\lambda)\, d\xi \right)$$

The lattice monodromy matrix is then $T(\lambda) = \lim_{l \to -\infty, m \to \infty} T_l^m$ where $T_l^m = L_m L_{m-1} \ldots L_{l+1}$, and its trace again yields an infinity of commuting charges, provided that there exists a quantum $R$-matrix $R(\lambda_1, \lambda_2)$ such that

$$R(\lambda_1,\lambda_2) L_n^1(\lambda_1) L_n^2(\lambda_2) = L_n^2(\lambda_2) L_n^1(\lambda_1) R(\lambda_1,\lambda_2) \qquad [19]$$

where $L_n^1(\lambda_1) = L_n(\lambda_1) \otimes \mathrm{id}$, $L_n^2(\lambda_2) = \mathrm{id} \otimes L_n(\lambda_2)$. That $R$ solves the Yang–Baxter equation follows from the equivalence of the two ways of intertwining $L_n(\lambda_1) \otimes L_n(\lambda_2) \otimes L_n(\lambda_3)$ with $L_n(\lambda_3) \otimes L_n(\lambda_2) \otimes L_n(\lambda_1)$.

To compute $L_n(\lambda)$, one uses the canonical, equal-time commutation relations for the $\phi_i$ and $\dot{\phi}_i$. In terms of the lattice fields

$$p_{i,n} = \int_{(n-(1/2))\Delta}^{(n+(1/2))\Delta} \dot{\phi}_i(x)\, dx$$

$$q_{i,n} = \int_{(n-(1/2))\Delta}^{(n+(1/2))\Delta} \sum_j e^{(\beta/2) a_{ij} \phi_j(x)}\, dx$$

the only nontrivial relation is $[p_{i,n}, q_{j,n}] = (i\hbar\beta/2)\delta_{ij} q_{j,n}$, and one finds

$$L_n(\lambda) = \exp\left( \frac{\beta}{2} \sum_i H_i p_{i,n} \right) + \exp\left( \frac{\beta}{4} \sum_j H_j p_{j,n} \right) \times \frac{m}{2} \left[ \sum_i q_{i,n} (E_i^+ + E_i^-) + \prod_i q_{i,n}^{-n_i} \left( \lambda E_0^+ + \frac{1}{\lambda} E_0^- \right) \right] \times \exp\left( \frac{\beta}{4} \sum_j H_j p_{j,n} \right) + O(\Delta^2)$$

the expression used by the St Petersburg school and by Jimbo. We now make the replacement $E_i^\pm \mapsto q^{-H_i/4} E_i^\pm q^{H_i/4}$, where $q = \exp(i\hbar\beta^2/2)$, and compute the $O(\Delta)$ terms in [19], which reduce to

$$
\begin{aligned}
&R(z)(H_i \otimes 1 + 1 \otimes H_i) \\
&\quad = (H_i \otimes 1 + 1 \otimes H_i)R(z) \\
&R(z)\left(E_i^{\pm} \otimes q^{-H_i/2} + q^{H_i/2} \otimes E_i^{\pm}\right) \\
&\quad = \left(q^{-H_i/2} \otimes E_i^{\pm} + E_i^{\pm} \otimes q^{H_i/2}\right)R(z) \\
&R(z)\left(z^{\pm 1}E_0^{\pm} \otimes q^{-H_0/2} + q^{H_0/2} \otimes E_0^{\pm}\right) \\
&\quad = \left(q^{-H_0/2} \otimes E_0^{\pm} + z^{\pm 1}E_0^{\pm} \otimes q^{H_0/2}\right)R(z)
\end{aligned}
$$

where $z = \lambda_1/\lambda_2$. We recognize in these the $U_h(\hat{\mathfrak{g}})$ coproduct and thus the intertwining relations, in the homogeneous gradation. These equations were solved for $R$ in defining representations of nonexceptional $\mathfrak{g}$ by Jimbo (1986).

For $\hat{\mathfrak{g}} = \widehat{\mathfrak{sl}_2}$, it was Kulish and Reshetikhin (1983) who first discovered that the requirement that the coproduct must be an algebra homomorphism forces the replacement of the commutation relations of $U(\widehat{\mathfrak{sl}_2})$ by those of $U_h(\widehat{\mathfrak{sl}_2})$; more generally it requires the replacement of $U(\hat{\mathfrak{g}})$ by $U_h(\hat{\mathfrak{g}})$.

## Affine Quantum Group Symmetry and the Exact *S*-Matrix

In the last section, we saw the origins of $U_h(\hat{\mathfrak{g}})$ in the "auxiliary" algebra introduced in the Lax pair. However, the quantum affine algebras also play a second role, as a symmetry algebra. An imaginary-coupled affine Toda field theory based on the affine algebra $\hat{\mathfrak{g}}^{\vee}$ possesses the quantum affine algebra $U_h(\hat{\mathfrak{g}})$ as a symmetry algebra, where $\hat{\mathfrak{g}}^{\vee}$ is the Langland dual to $\hat{\mathfrak{g}}$ (the algebra obtained by replacing roots by coroots).

The solitonic particle states in affine Toda theories form multiplets which transform in the fundamental representations of the quantum affine algebra. Multiparticle states transform in tensor product representations $V^a \otimes V^b$. The scattering of two solitons of type $a$ and $b$ with relative rapidity $\theta$ is described by the $S$-matrix $S^{ab}(\theta): V^a \otimes V^b \to V^b \otimes V^a$, graphically represented in Figure 1a. It then follows from the symmetry that the two-particle scattering matrix

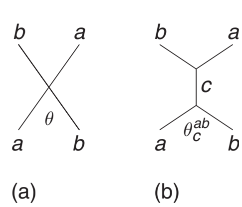


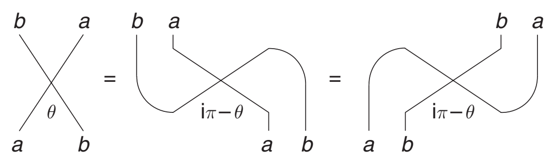

**Figure 1** (a) Graphical representation of a two-particle scattering process described by the $S$-matrix $S_{ab}(\theta)$. (b) At special values $\theta_{ab}^{c}$ of the relative spectral parameter, the two particles of types $a$ and $b$ form a bound state of type $c$.

($S$-matrix) for solitons must be proportional to the intertwiner for these tensor product representations, the $R$ matrix:

$$
S^{ab}(\theta) = f^{ab}(\theta)\check{R}^{ab}(\theta)
$$

with $\theta$ proportional to $u$, the additive spectral parameter. The scalar prefactor $f^{ab}(\theta)$ is not determined by the symmetry but is fixed by other requirements like unitarity, crossing symmetry, and the bootstrap principle.

It turns out that the axiomatic properties of the $R$-matrices are in perfect agreement with the axiomatic properties of the analytic $S$-matrix. For example, crossing symmetry of the $S$-matrix, graphically represented by

 [20]

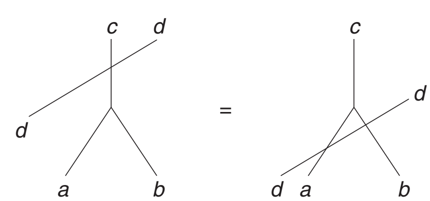

is a consequence of the property of the universal $R$-matrix with respect to the action of the antipode $S$,

$$
(S \otimes 1)\mathcal{R} = \mathcal{R}^{-1}
$$

An $S$-matrix will have poles at certain imaginary rapidities $\theta_c^{ab}$ corresponding to the formation of virtual bound states. This is graphically represented in Figure 1b. The location of the pole is determined by the masses of the three particles involved,

$$
m_c^2 = m_a^2 + m_b^2 + 2m_a m_b \cos(\mathrm{i}\theta_c^{ab})
$$

At the bound state pole, the $S$-matrix will project onto the multiplet $V^c$. Thus, the $\check{R}$ matrix has to have this projection property as well and indeed, this turns out to be the case. The bootstrap principle, whereby the $S$-matrix for a bound state is obtained from the $S$-matrices of the constituent particles,

 [21]


is a consequence of the property [14] of the universal $R$-matrix with respect to the coproduct.

There is a famous no-go theorem due to Coleman and Mandula which states the "impossibility of combining space-time and internal symmetries in any but a trivial way." Affine quantum group symmetry circumvents this no-go theorem. In fact, the derivation $D$ is the infinitesimal two-dimensional Lorentz boost generator and the other symmetry

charges transform nontrivially under these Lorentz transformations, see [2].

The noncocommutative coproduct [8] means that a $U_h(\hat{\mathfrak{g}})$ symmetry generator, when acting on a 2-soliton state, acts differently on the left soliton than on the right soliton. This is only possible because the generator is a nonlocal symmetry charge – that is, a charge which is obtained as the space integral of the time component of a current which itself is a nonlocal expression in terms of the fields of the theory.

Similarly, many nonlinear sigma models possess nonlocal charges which form $Y(\mathfrak{g})$, and the construction proceeds similarly, now utilizing rational $R$-matrices, and with particle multiplets forming fundamental representations of $Y(\mathfrak{g})$. In each case, the three-point couplings corresponding to the formation of bound states, and thus the analogs for $U_h(\hat{\mathfrak{g}})$ and $Y(\mathfrak{g})$ of the Clebsch–Gordan couplings, obey a rather beautiful geometric rule originally deduced in simpler, purely elastic scattering models (Chari and Pressley 1996).

More details about this topic can be found in Delius (1995) and MacKay (2005).

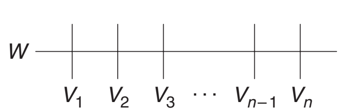
## Integrable Quantum Spin Chains

Affine quantum groups provide an unlimited supply of integrable quantum spin chains. From any $R$-matrix $R(\theta)$ for any tensor product of finite-dimensional representations $W \otimes V$, one can produce an integrable quantum system on the Hilbert space $V^{\otimes n}$. This Hilbert space can then be interpreted as the space of $n$ interacting spins. The space $W$ is an auxiliary space required in the construction but not playing a role in the physics.

Given an arbitrary $R$-matrix $R(\theta)$, one defines the monodromy matrix $T(\theta) \in \mathrm{End}(W \otimes V^{\otimes n})$ by

$$T(\theta) = R_{01}(\theta - \theta_1)R_{02}(\theta - \theta_2) \cdots R_{0n}(\theta - \theta_n)$$

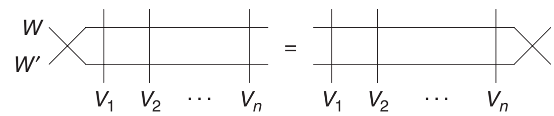
where, as usual, $R_{ij}$ is the $R$-matrix acting on the $i$th and $j$th component of the tensor product space. The $\theta_i$ can be chosen arbitrarily for convenience. Graphically the monodromy matrix can be represented as




As a consequence of the Yang–Baxter equation satisfied by the $R$-matrices the monodromy matrix satisfies

$$RTT = TTR \qquad [22]$$

or, graphically,




One defines the transfer matrix

$$\tau(\theta) = \mathrm{tr}_W T(\theta)$$

which is now an operator on $V^{\otimes n}$, the Hilbert space of the quantum spin chain. Due to [22], two transfer matrices commute,

$$[\tau(\theta), \tau(\theta')] = 0$$

and thus the $\tau(\theta)$ can be seen as a generating function of an infinite number of commuting charges, one of which will be chosen as the Hamiltonian. This Hamiltonian can then be diagonalized using the algebraic Bethe ansatz.

One is usually interested in the thermodynamic limit where the number of spins goes to infinity. In this limit, it has been conjectured, the Hilbert space of the spin chain carries a certain infinite-dimensional representation of the quantum affine algebra and this has been used to solve the model algebraically, using vertex operators (Jimbo and Miwa 1995).

## Boundary Quantum Groups

In applications to physical systems that have a boundary, the Yang–Baxter equation [1] appears in conjunction with the boundary Yang–Baxter equation, also known as the reflection equation,

$$\begin{aligned} R_{12}(u-v)K_1(u)R_{21}(u+v)K_2(v) \\ = K_2(v)R_{12}(u+v)K_1(u)R_{21}(u-v) \end{aligned} \qquad [23]$$

The matrices $K$ are known as reflection matrices. This equation was originally introduced by Cherednik to describe the reflection of particles from a boundary in an integrable scattering theory and was used by Sklyanin to construct integrable spin chains and quantum field theories with boundaries.

Boundary quantum groups are certain co-ideal subalgebras of affine quantum groups. They provide the algebraic structures underlying the solutions of the boundary Yang–Baxter equation in the same way in which affine quantum groups underlie the solutions of the ordinary Yang–Baxter equation. Both allow one to find solutions of the respective Yang–Baxter equation by solving a linear intertwining relation. In the case without spectral parameters these algebras appear in the theory of braided groups (*see* Hopf Algebras and $q$-Deformation Quantum Groups and Braided and Modular Tensor Categories).

For example, the subalgebra $B_\epsilon(\hat{\mathfrak{g}})$ of $U_h(\hat{\mathfrak{g}}')$ generated by

$$Q_i = q_i^{H_i/2}(E_i^+ + E_i^-) + \epsilon_i(q_i^{H_i} - 1), \quad i = 0, \ldots, r \qquad [24]$$

is a boundary quantum group for certain choices of the parameters $\epsilon_i \in \mathbb{C}[[h]]$. It is a left co-ideal subalgebra of $U_h(\hat{\mathfrak{g}}')$ because

$$\Delta(Q_i) = Q_i \otimes 1 + q_i^{H_i} \otimes Q_i \in U_h(\hat{\mathfrak{g}}') \otimes B_\epsilon(\hat{\mathfrak{g}}) \qquad [25]$$

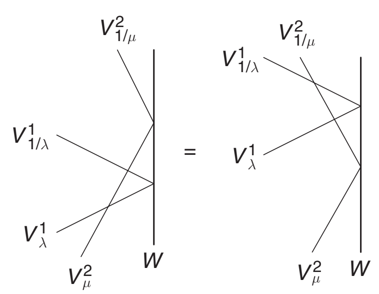
Intertwiners $K(\lambda) : V_{\eta\lambda} \to V_{\eta/\lambda}$ for some constant $\eta$ satisfying

$$K(\lambda)\rho_{\eta\lambda}(Q) = \rho_{\eta/\lambda}(Q)K(\lambda), \quad \text{for all } Q \in B_\epsilon(\hat{\mathfrak{g}}) \qquad [26]$$

provide solutions of the reflection equation in the form

$$(\mathrm{id} \otimes K^2(\mu))\check{R}^{12}(\lambda\mu)(\mathrm{id} \otimes K^1(\lambda))\check{R}^{21}(\lambda/\mu) = \check{R}^{12}(\lambda/\mu)(\mathrm{id} \otimes K^1(\lambda)) \times \check{R}^{21}(\lambda\mu)(\mathrm{id} \otimes K^2(\mu)) \qquad [27]$$

This can be extended to the case where the boundary itself carries a representation $W$ of $B_\epsilon(\hat{\mathfrak{g}})$. The boundary Yang–Baxter equation can be represented graphically as




Another example is provided by twisted Yangians where, when the $I_a$ and $J_a$ are constructed as nonlocal charges in sigma models, it is found that a boundary condition which preserves integrability leaves only the subset

$$I_i \quad \text{and} \quad \tilde{J}_p = J_p + \tfrac{1}{4} f_{piq}(I_i I_q + I_q I_i)$$

conserved, where $i$ labels the $\mathfrak{h}$-indices and $p, q$ the $\mathfrak{k}$-indices of a symmetric splitting $\mathfrak{g} = \mathfrak{h} + \mathfrak{k}$. The algebra $Y(\mathfrak{g}, \mathfrak{h})$ generated by the $I_i, \tilde{J}_p$ is, like $B_\epsilon(\hat{\mathfrak{g}})$, a co-ideal subalgebra, $\Delta(Y(\mathfrak{g}, \mathfrak{h})) \subset Y(\mathfrak{g}) \otimes Y(\mathfrak{g}, \mathfrak{h})$, and again yields an intertwining relation for $K$-matrices. For $\mathfrak{g} = \mathfrak{sl}_n$ and $\mathfrak{h} = \mathfrak{so}_n$ or $\mathfrak{sp}_{2n}$, $Y(\mathfrak{g}, \mathfrak{h})$ is the "twisted Yangian" described in Molev (2003).

All the constructions in earlier sections of this review have analogs in the boundary setting. For more details see Delius and MacKay (2003) and MacKay (2005).

*See also:* Bethe Ansatz; Boundary Conformal Field Theory; Classical *r*-Matrices, Lie Bialgebras, and Poisson Lie Groups; Hopf Algebras and *q*-Deformation Quantum Groups; Riemann–Hilbert Problem; Solitons and Kac–Moody Lie Algebras; Yang–Baxter Equations.

## Further Reading

Chari V and Pressley AN (1994) *Quantum Groups*. Cambridge: Cambridge University Press.

Chari V and Pressley AN (1996) Yangians, integrable quantum systems and Dorey's rule. *Communications in Mathematical Physics* 181: 265–302.

Delius GW (1995) Exact S-matrices with affine quantum group symmetry. *Nuclear Physics B* 451: 445–465.

Delius GW and MacKay NJ (2003) Quantum group symmetry in sine-Gordon and affine Toda field Theories on the Half-Line, *Communications in Mathematical Physics* 233: 173–190.

Drinfeld V (1985) Hopf algebras and the quantum Yang–Baxter equation. *Soviet Mathematics Doklady* 32: 254–258.

Drinfeld V (1986) *Quantum Groups*, Proc. Int. Cong. Math. (Berkeley), pp. 798–820.

Drinfeld V (1988) A new realization of Yangians and quantized affine algebras. *Soviet Mathematics Doklady* 36: 212–216.

Jimbo M (1985) A $q$-difference analogue of $U(\mathfrak{g})$ and the Yang–Baxter equation. *Letters in Mathematical Physics* 10: 63–69.

Jimbo M (1986) Quantum R-matrix for the generalized Toda system. *Communications in Mathematical Physics* 102: 537–547.

Jimbo M and Miwa T (1995) *Algebraic Analysis of Solvable Lattice Models*. Providence, RI: American Mathematical Society.

Kulish PP and Reshetikhin NY (1983) Quantum linear problem for the sine-Gordon equation and higher representations. *Journal of Soviet Mathematics* 23: 2435.

MacKay NJ (2005) Introduction to Yangian symmetry in integrable field theory. *International Journal of Modern Physics* (to appear).

Molev A (2003) Yangians and their applications. In: Hazewinkel M (ed.) *Handbook of Algebra*, vol. 3, pp. 907–959. Elsevier.

# Aharonov–Bohm Effect

**M Socolovsky**, Universidad Nacional Autónoma de México, México DF, México



## Introduction

In classical electrodynamics, the interaction of charged particles with the electromagnetic field is local, through the pointlike coupling of the electric charge of the particles with the electric and magnetic fields, $\boldsymbol{E}$ and $\boldsymbol{B}$, respectively. This is mathematically expressed by the Lorentz-force law. The scalar and vector potentials, $\varphi$ and $\boldsymbol{A}$, which are the time and space components of the relativistic 4-potential $A_\mu$, are considered auxiliary quantities in terms of which the field strengths $\boldsymbol{E}$ and $\boldsymbol{B}$, the observables, are expressed in a gauge-invariant manner. The homogeneous or first pair of Maxwell equations are a direct consequence of the definition of the field strengths in terms of $A_\mu$. The inhomogeneous or second pair of Maxwell equations, which involve the charges and currents present in the problem, are also usually written in terms of $\boldsymbol{E}$ and $\boldsymbol{B}$; however when writing them in terms of $A_\mu$, the number of degrees of freedom of the electromagnetic field is explicitly reduced from six to four; and finally, with two additional gauge transformations, one ends with the two physical degrees of freedom of the electromagnetic field.

In quantum mechanics, however, both the Schrödinger equation and the path-integral approaches for scalar and unpolarized charged particles in the presence of electromagnetic fields, are written in terms of the potential and not of the field strengths. Even in the case of the Schrödinger–Pauli equation for spin 1/2 electrons with magnetic moment $\boldsymbol{\mu}$ interacting with a magnetic field $\boldsymbol{B}$, one knows that the coupling $-\boldsymbol{\mu}\cdot\boldsymbol{B}$ is the nonrelativistic limit of the Dirac equation, which depends on $A_\mu$ but not on $\boldsymbol{E}$ and $\boldsymbol{B}$. Since gauge invariance also holds in the quantum domain, it was thought that $\boldsymbol{A}$ and $\varphi$ were mere auxiliary quantities, like in the classical case.

Aharonov and Bohm, in 1959, predicted a quantum interference effect due to the motion of charged particles in regions where $\boldsymbol{B}(\boldsymbol{E})$ vanishes, but not $\boldsymbol{A}(\varphi)$, leading to a nonlocal gauge-invariant effect depending on the flux of the magnetic field in the inaccessible region, in the magnetic case, and on the difference of the integrals over time of time-varying potentials, in the electric case. (The magnetic effect was already noticed 10 years before by Ehrenberg and Siday in a paper on the refractive index of electrons.)

In the context of the Schrödinger equation, one can show that due to gauge invariance, if $\psi_0$ is a solution to the equation in the absence of an electromagnetic potential, then the product of $\psi_0(\boldsymbol{x})$ times the integral of $A_\mu$ over a path joining an arbitrary reference point $\boldsymbol{x}_0$ to $\boldsymbol{x}$ is also a solution, if the integral is path independent. However, it is the path integral of Feynman which in the formulas for propagators of charged particles in the presence of electromagnetic fields clearly shows that *the action of these fields on charged particles is nonlocal, and it is given by the celebrated nonintegrable (path-dependent) phase factor of* Wu and Yang (1975). Moreover, this fact provides an additional proof of the nonlocal character of quantum mechanics: to surround fluxes, or to develop a potential difference, the particle has to travel simultaneously at least through two paths.

Thus, the fact that the Aharonov–Bohm (A–B) effect was verified experimentally, by Chambers and others, demonstrates the necessity of introducing the (gauge-dependent) potential $A_\mu$ in describing the electromagnetic interactions of the quantum particle. This is widely regarded as the single most important piece of evidence for electromagnetism being a gauge theory. Moreover, it shows, to paraphrase Yang, that the field underdescribes the physical theory, while the potential overdescribes it, and it is the phase factor which describes it exactly.

The content of this article is essentially twofold. The first four sections are mainly physical, where we describe the magnetic A–B effect using the Schrödinger equation and the Feynman path integral. The fifth section is geometrical and is the longest of the article. We describe the effect in the context of fiber bundles and connections, namely as a result of the coupling of the wave function (section of an associated bundle) to a nontrivial flat connection (non-pure gauge vector potential with zero magnetic field) in a trivial bundle (the A–B bundle) with topologically nontrivial (non-simply-connected) base space. We discuss the moduli space of flat connections and the holonomy groups giving the phase shifts of the interference patterns. Finally, in the last section, we briefly comment on the nonabelian A–B effect.

## Electromagnetic Fields in Classical Physics

In classical physics, the motion of charged particles in the presence of electromagnetic fields is governed by the equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{p} = q\left(\boldsymbol{E} + \frac{\boldsymbol{v}}{c} \times \boldsymbol{B}\right) \qquad [1]$$

where

$$\boldsymbol{p} = \frac{m\boldsymbol{v}}{\sqrt{1-(v^2/c^2)}}$$

is the mechanical momentum of the particle with electric charge $q$, mass $m$, and velocity $\boldsymbol{v} = \dot{\boldsymbol{x}}$ ($c$ is the velocity of light in vacuum, and for $|\boldsymbol{v}| \ll c$ the left-hand side (LHS) of [1] is approximately $m\boldsymbol{v}$); the right-hand side (RHS) is the Lorentz force, where $\boldsymbol{E}$ and $\boldsymbol{B}$ are, respectively, the electric and magnetic fields at the spacetime point $(t, \boldsymbol{x})$ where the particle is located. Equation [1] is easily derived from the Euler–Lagrange equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \boldsymbol{v}}\right) - \frac{\partial L}{\partial \boldsymbol{x}} = 0 \qquad [2]$$

with the Lagrangian $L$ given by the sum of the free Lagrangian for the particle,

$$L_0 = -mc^2\sqrt{1-\frac{v^2}{c^2}} \qquad [3]$$

and the Lagrangian describing the particle–field interaction,

$$L_{\mathrm{int}} = \frac{q}{c}\boldsymbol{A}\cdot\boldsymbol{v} - q\varphi \qquad [4]$$

In [4], $\boldsymbol{A}$ and $\varphi$ are, respectively, the vector potential and the scalar potential, which together form the 4-potential $A_\mu = (A_0, -\boldsymbol{A}) = (\varphi, -A^i)$, $i = 1, 2, 3$, in terms of which the electric and magnetic field strengths are given by

$$\boldsymbol{E} = -\frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{A} - \nabla\varphi \qquad [5a]$$

$$\boldsymbol{B} = \nabla \times \boldsymbol{A} \qquad [5b]$$

The classical action corresponding to a given path of the particle is

$$S = \int_{t_1}^{t_2} \mathrm{d}t\, L = \int_{t_1}^{t_2} \mathrm{d}t(L_0 + L_{\mathrm{int}}) = \int_{t_1}^{t_2} \mathrm{d}t\, L_0 + \int_{t_1}^{t_2} \mathrm{d}t\, L_{\mathrm{int}} \equiv S_0 + S_{\mathrm{int}} \qquad [6]$$

$\boldsymbol{E}$, $\boldsymbol{B}$, and $S$ are invariant under the gauge transformation

$$\boldsymbol{A} \to \boldsymbol{A}' = \boldsymbol{A} - \nabla\Lambda \qquad [7a]$$

$$\varphi \to \varphi' = \varphi + \frac{1}{c}\frac{\partial}{\partial t}\Lambda \qquad [7b]$$

where $\Lambda$ is a real-valued differentiable scalar function (at least of class $C^2$) on spacetime. That is, if $\boldsymbol{E}'$, $\boldsymbol{B}'$, and $S'_{\mathrm{int}}$ are defined in terms of $\boldsymbol{A}'$ and $\varphi'$ as $\boldsymbol{E}$, $\boldsymbol{B}$, and $S_{\mathrm{int}}$ are defined in terms of $\boldsymbol{A}$ and $\varphi$, then $\boldsymbol{E}' = \boldsymbol{E}$, $\boldsymbol{B}' = \boldsymbol{B}$, and $S'_{\mathrm{int}} = S_{\mathrm{int}}$. This fact leads to the concept that, classically, the observables $\boldsymbol{E}$ and $\boldsymbol{B}$ are the physical quantities, while $A_\mu$ is only an auxiliary quantity. Also, and most important in the present context, eqn [1] states that the motion of the particles is determined by the values or state of the field strengths in an infinitesimal neighborhood of the particles, that is, classically, $\boldsymbol{E}$ and $\boldsymbol{B}$ act locally. If one defines the differential 1-form $A \equiv A_\mu \mathrm{d}x^\mu$ (with $\mathrm{d}x^0 = c\,\mathrm{d}t$), then the components of the differential 2-form $F = \mathrm{d}A = (1/2)(\partial_\mu A_\nu - \partial_\nu A_\mu)\mathrm{d}x^\mu \wedge \mathrm{d}x^\nu \equiv (1/2)F_{\mu\nu}$ $\mathrm{d}x^\mu \wedge \mathrm{d}x^\nu$ are precisely the electric and magnetic fields:

$$F_{\mu\nu} = \begin{pmatrix} 0 & E^1 & E^2 & E^3 \\ -E^1 & 0 & -B^3 & B^2 \\ -E^2 & B^3 & 0 & -B^1 \\ -E^3 & -B^2 & B^1 & 0 \end{pmatrix} \qquad [8]$$

At the level of $A$,

$$\mathrm{d}F = \mathrm{d}^2 A = 0 \qquad [9]$$

is an identity, but at the level of $\boldsymbol{E}$ and $\boldsymbol{B}$, [9] amounts to the homogeneous (or first pair of) Maxwell equations obeyed by the field strengths:

$$\nabla \cdot \boldsymbol{B} = 0 \qquad [10a]$$

$$\nabla \times \boldsymbol{E} + \frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{B} = 0 \qquad [10b]$$

Therefore, these equations have a geometrical origin. The second pair of Maxwell equations is dynamical, and is obtained from the field action (in the Heaviside system of units)

$$S_{\mathrm{field}} = -\frac{1}{4c}\int \mathrm{d}^4x F_{\mu\nu}F^{\mu\nu} \qquad [11]$$

which leads to $\qquad \nabla \cdot \boldsymbol{E} = 4\pi\rho \qquad [12a]$

$$\nabla \times \boldsymbol{B} - \frac{1}{c}\frac{\partial}{\partial t}\boldsymbol{E} = \frac{4\pi \boldsymbol{j}}{c} \qquad [12b]$$

where $(\rho, -\boldsymbol{j}) = (j^0, -\boldsymbol{j})$ is the 4-current satisfying, as a consequence of [12a] and [12b], the conservation law

$$\partial_\mu j^\mu = 0 \qquad [13]$$

For a pointlike particle, $\rho(t, \boldsymbol{x}) = q\delta^3(\boldsymbol{x} - \boldsymbol{x}(t))$ and $\boldsymbol{j} = \rho\boldsymbol{v}$.

## Electromagnetic Fields in Quantum Physics

In quantum physics, the motion of charged particles in external electromagnetic fields is governed by the Schrödinger equation or, equivalently, by the Feynman path integral. In both cases, however, it is the 4-potential $A_\mu$ which appears in the equations, and not the field strengths. For simplicity, we consider here scalar (spinless) charged particles or unpolarized electrons (spin-(1/2)particles), both of which, in the nonrelativistic approximation, can be described quantum mechanically by a complex wave function $\psi(t, \boldsymbol{x})$.

To derive the Schrödinger equation, one starts from the classical Hamiltonian

$$H = \boldsymbol{P}\cdot\boldsymbol{v} - L - mc^2 = \frac{1}{2}\left(\boldsymbol{P} - \frac{q}{c}\boldsymbol{A}\right)^2 + q\varphi \qquad [14]$$

where

$$\boldsymbol{P} = \frac{\partial}{\partial \boldsymbol{v}}L = \boldsymbol{p} + \frac{q}{c}\boldsymbol{A}$$

is the canonical momentum of the particle, and we have subtracted its rest energy. The replacements $\boldsymbol{P} \to -\mathrm{i}\hbar\nabla$ and $H \to \mathrm{i}\hbar\partial/\partial t$ lead to

$$\begin{aligned}\mathrm{i}\hbar\frac{\partial}{\partial t}\psi &= \left(\frac{1}{2m}\left(\mathrm{i}\hbar\nabla + \frac{q}{c}\boldsymbol{A}\right)^2 + q\varphi\right)\psi \\ &= \left(-\frac{\hbar^2}{2m}\nabla^2 + \frac{q^2}{2mc^2}A^2 \right.\\ &\quad \left. + \frac{i\hbar q}{2mc}\nabla\cdot\boldsymbol{A} + \frac{i\hbar q}{mc}\boldsymbol{A}\cdot\nabla + q\varphi\right)\psi \qquad [15]\end{aligned}$$

The gauge transformation [7a] and [7b] is a symmetry of this equation, if simultaneously to the change of the 4-potential, the wave function transforms as follows:

$$\psi(t, \boldsymbol{x}) \to \psi'(t, \boldsymbol{x}) = \mathrm{e}^{-(\mathrm{i}q/\hbar c)\Lambda}\,\psi(t, \boldsymbol{x}) \qquad [7c]$$

So, $\boldsymbol{A}'$ and $\psi'$ obey [15]. At each $(t, \boldsymbol{x})$, $\mathrm{e}^{-(\mathrm{i}q/\hbar c)\Lambda}$ belongs to U(1), the unit circle in the complex plane.

In the path-integral approach, the kernel $K(t', \boldsymbol{x}'; t, \boldsymbol{x})$, which gives the probability amplitude for the propagation of the particle from the spacetime point $(t, \boldsymbol{x})$ to the spacetime point $(t', \boldsymbol{x}')$ $(t < t')$, is given by

$$\begin{aligned}&K(t', \boldsymbol{x}'; t, \boldsymbol{x}) \\ &= \int_{\boldsymbol{x}(t)=\boldsymbol{x}}^{\boldsymbol{x}(t')=\boldsymbol{x}'} D\boldsymbol{x}(\tau)\exp\left(\frac{\mathrm{i}}{\hbar}(S_0 + S_{\text{int}})\right) \\ &= \int_{\boldsymbol{x}(t)=\boldsymbol{x}}^{\boldsymbol{x}(t')=\boldsymbol{x}'} D\boldsymbol{x}(\tau)\exp\left(\frac{\mathrm{i}}{\hbar}\int_t^{t'} \mathrm{d}\tau\left(\frac{1}{2}m\dot{x}^2 \right.\right.\\ &\quad \left.\left. + \frac{q}{c}\boldsymbol{A}\cdot\boldsymbol{v} - q\varphi\right)\right) \\ &= \int_{\boldsymbol{x}(t)=\boldsymbol{x}}^{\boldsymbol{x}(t')=\boldsymbol{x}'} D\boldsymbol{x}(\tau)\exp\left(\frac{\mathrm{i}}{\hbar}\int_t^{t'} \mathrm{d}\tau\frac{1}{2}m\dot{x}^2\right) \\ &\quad \times \exp\left(\frac{\mathrm{i}q}{\hbar c}\int_t^{t'}(\boldsymbol{A}\cdot\mathrm{d}\boldsymbol{x} - \varphi\,\mathrm{d}x^0)\right) \\ &= \int_{\boldsymbol{x}(t)=\boldsymbol{x}}^{\boldsymbol{x}(t')=\boldsymbol{x}'} D\boldsymbol{x}(\tau)\exp\left(\frac{\mathrm{i}}{\hbar}\int_t^{t'} \mathrm{d}\tau\frac{1}{2}m\dot{x}^2\right) \\ &\quad \times \exp\left(\frac{\mathrm{i}q}{\hbar c}\int_t^{t'}\mathrm{d}x^\mu A_\mu\right) \qquad [16]\end{aligned}$$

where the integral $\int D\boldsymbol{x}(\tau)\ldots$ is over all continuous spacetime paths $(\tau, \boldsymbol{x}(\tau))$ which join $(t, \boldsymbol{x})$ with $(t', \boldsymbol{x}')$. If one knows the wave function at $(t, \boldsymbol{x})$, then the wave function at $(t', \boldsymbol{x}')$ is given by

$$\psi(t', \boldsymbol{x}') = \int \mathrm{d}^3\boldsymbol{x}\; K(t', \boldsymbol{x}'; t, \boldsymbol{x})\psi(t, \boldsymbol{x}) \qquad [17]$$

An important point is the natural appearance in the integrand of the functional integral of the factor

$$\mathrm{e}^{(\mathrm{i}q/\hbar c)\int_\gamma A}$$

for each path $\gamma$ joining $(t, \boldsymbol{x})$ with $(t', \boldsymbol{x}')$.

## A Solution to the Schrödinger Equation

In what follows, we shall restrict ourselves to static magnetic fields; then in the previous formulas, we set $\varphi = 0$ and $\boldsymbol{A}(t, \boldsymbol{x}) = \boldsymbol{A}(\boldsymbol{x})$. It is then easy to show that if $\boldsymbol{x}_0$ is an arbitrary reference point and the integral $\int_{\boldsymbol{x}_0}^{\boldsymbol{x}} \boldsymbol{A}(\boldsymbol{x}')\cdot\mathrm{d}\boldsymbol{x}'$ is independent of the integration path from $\boldsymbol{x}_0$ to $\boldsymbol{x}$, that is, it is a well-defined function $f$ of $\boldsymbol{x}$, and if $\psi_0$ is a solution of the free Schrödinger equation, that is,

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\psi_0 = -\frac{\hbar^2}{2m}\nabla^2\psi_0 \qquad [18]$$

then

$$\psi(t, \boldsymbol{x}) = \exp\left(\frac{\mathrm{i}q}{\hbar c}\int_{\boldsymbol{x}_0}^{\boldsymbol{x}}\boldsymbol{A}(\boldsymbol{x}')\cdot\mathrm{d}\boldsymbol{x}'\right)\psi_0(t, \boldsymbol{x}) \qquad [19]$$

is a solution of [15]. In fact, replacing [19] in [15], the LHS gives

$$\exp\left(\frac{\mathrm{i}q}{\hbar c}f(\boldsymbol{x})\right)\mathrm{i}\hbar\frac{\partial}{\partial t}\psi_0$$

while for the RHS one has

$$\exp\left(\frac{\mathrm{i}q}{\hbar c}f(\boldsymbol{x})\right)\left(-\frac{\hbar^2}{2m}\right)\nabla^2\psi_0$$

The cancelation of the exponential factors shows that, under the condition of path independence, there is no effect of the potential on the charged particles. Another way to see this is by making a gauge transformation [7a]–[7c] with $\Lambda(\boldsymbol{x}) = f(\boldsymbol{x})$, which changes $\psi \to \psi_0$ and $\boldsymbol{A} \to \boldsymbol{A}' = \boldsymbol{A} - \nabla \int_{x_0}^{x} \boldsymbol{A}(\boldsymbol{x}') \cdot \mathrm{d}\boldsymbol{x}' = \boldsymbol{A} - \boldsymbol{A} = 0$.

The condition of path independence amounts, however, to the condition that no magnetic field is present since, if $\int_\gamma A$ depends on $\gamma$, then for some pair of paths $\gamma$ and $\gamma'$ from $(t, \boldsymbol{x})$ to $(t', \boldsymbol{x}')$, $0 \neq \int_\gamma A - \int_{\gamma'} A = \int_\gamma A + \int_{-\gamma'} A = \oint_{\gamma\cup(-\gamma')} A = \int_\Sigma \mathrm{d}\boldsymbol{\sigma} \cdot (\nabla \times \boldsymbol{A})$, where in the last equality we applied Stokes theorem ($\Sigma$ is any surface with boundary $\gamma \cup (-\gamma')$), which shows that $\boldsymbol{B} = \nabla \times \boldsymbol{A}$ must not vanish everywhere and has a nonzero flux $\Phi$ through $\Sigma$ given by

$$\Phi = \int_\Sigma \mathrm{d}\boldsymbol{\sigma} \cdot \boldsymbol{B} \qquad [20]$$

The conclusion of this section is that the ansatz [19] for solving [15] can only be applied in simply connected regions with no magnetic field strength present.

## Aharonov–Bohm Proposal

In 1959, Aharonov and Bohm proposed an experiment to test, in quantum mechanics, the coupling of electric charges to electromagnetic field strengths through a local interaction with the electromagnetic potential $A_\mu$, but not with the field strengths themselves. However, as we saw before, no physical effect exists, that is, $A_\mu$ can be gauged away, unless magnetic and/or electric fields exist somewhere, although not necessarily overlapping the wave function of the particles.

Consider the usual two-slit experiment as depicted in **Figure 1**, with the additional presence, behind the slits, of a long and narrow solenoid enclosing a nonvanishing magnetic flux $\Phi$ due to a constant and homogeneous magnetic field $\boldsymbol{B}$ normal to the plane of the figure (in direction $z$); outside of the solenoid, the magnetic field is zero. If the radius of the solenoid is $R$, a vector potential $\boldsymbol{A}$ that produces such field strength is given by

$$\boldsymbol{A}(\boldsymbol{x}) = \begin{cases} (|\boldsymbol{B}|r/2)\hat{\varphi}, & r \leq R \\ (\Phi/2\pi r)\hat{\varphi}, & r > R \end{cases} \qquad [21]$$

where $\Phi = \pi R^2 |\boldsymbol{B}|$ and $\hat{\varphi}$ is a unit vector in the azimuthal direction. In fact,

$$\boldsymbol{B} = \nabla \times \boldsymbol{A}(\boldsymbol{x}) = \begin{cases} |\boldsymbol{B}|\hat{z}, & r \leq R \\ 0, & r > R \end{cases} \qquad [22]$$

Notice that at $r = R$, $\boldsymbol{A}$ is continuous but not continuously differentiable. Also, the ideal limit of an infinitely long solenoid makes the problem two-dimensional, that is, in the $x$–$y$ plane.

The probability amplitude for an electron emitted at the source S to arrive at the point P on the screen $\Pi$, is given by the sum of two probability amplitudes, namely those corresponding to passing through the slits 1 and 2. The solenoid is assumed to be impenetrable to the electrons; mathematically, this corresponds to a motion in a non-simply-connected region. In the approximation for the path integral [16], in which one considers the contribution of only two classes of paths, that is, the class $\{\gamma\}$ represented by path I, and the class $\{\gamma'\}$ represented by path II, if the wave function at the source is $\psi_\mathrm{S}$, then the wave function at P is given by

$$\begin{aligned} \psi_\mathrm{P} &= \left( \int_{\{\gamma\}} \mathrm{e}^{(\mathrm{i}/\hbar)S_0(\gamma)} \mathrm{e}^{-(\mathrm{i}|e|/\hbar c)\int_\gamma A} \right. \\ &\quad \left. + \int_{\{\gamma'\}} \mathrm{e}^{(\mathrm{i}/\hbar)S_0(\gamma')} \mathrm{e}^{-(\mathrm{i}|e|/\hbar c)\int_{\gamma'} A} \right) \psi_\mathrm{S} \\ &= \mathrm{e}^{-(\mathrm{i}|e|/\hbar c)\int_\mathrm{I} A} \int_{\{\gamma\}} \mathrm{e}^{(\mathrm{i}/\hbar)S_0(\gamma)} \psi_\mathrm{S} \\ &\quad + \mathrm{e}^{-(\mathrm{i}|e|/\hbar c)\int_\mathrm{II} A} \int_{\{\gamma'\}} \mathrm{e}^{(\mathrm{i}/\hbar)S_0(\gamma')} \psi_\mathrm{S} \\ &= \mathrm{e}^{-(\mathrm{i}|e|/\hbar c)\int_\mathrm{I} A} \left( \psi_\mathrm{P}^0(\mathrm{I}) \right. \\ &\quad \left. + \mathrm{e}^{-(\mathrm{i}|e|/\hbar c)\left(\int_{\mathrm{II}\cup(-\mathrm{I})} A\right)} \psi_\mathrm{P}^0(\mathrm{II}) \right) \\ &= \mathrm{e}^{-(\mathrm{i}|e|/\hbar c)\int_\mathrm{I} A} \left( \psi_\mathrm{P}^0(\mathrm{I}) + \mathrm{e}^{-2\pi\mathrm{i}(\Phi/\Phi_0)} \psi_\mathrm{P}^0(\mathrm{II}) \right) \end{aligned} \qquad [23]$$

where, in the second line, we used the path independence of the integral of $A$ within each class of paths;

$$\psi_\mathrm{P}^0(\mathrm{I}) = \int_{\{\gamma\}} \mathrm{e}^{(\mathrm{i}/\hbar)\int_{\{\gamma\}} S_0(\gamma)} \psi_\mathrm{S}$$

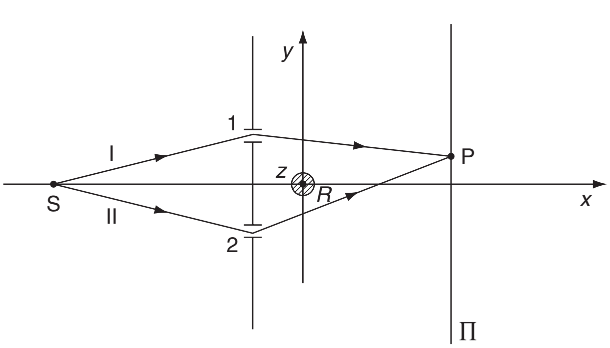



**Figure 1** Magnetic Aharonov–Bohm effect.

and

$$\psi_P^0(\mathrm{II}) = \int_{\{\gamma'\}} e^{(i/\hbar)S_0(\gamma')}\psi_S$$

and, in the last equality, we applied the extended version of Stokes theorem (by Craven), to allow for noncontinuously differentiable vector potentials; and the quantum of magnetic flux associated with the charge $|e|$ is defined by

$$\Phi_0 = 2\pi\frac{\hbar c}{|e|} \cong 4.135 \times 10^{-7}\,\mathrm{G\,cm^2} \qquad [24]$$

($=2\pi/|e| = \sqrt{\pi/\alpha} \cong \sqrt{137\pi}$ in the natural system of units (n.s.u.) $\hbar = c = 1$; $\alpha$ is the fine structure constant). Then the probability of finding the electron at P is proportional to

$$|\psi_P|^2 = |\psi_P^0(\mathrm{I})|^2 + |\psi_P^0(\mathrm{II})|^2 + 2\mathrm{Re}\left(e^{2\pi i(\Phi/\Phi_0)}\psi_P^0(\mathrm{I})\psi_P^0(\mathrm{II})^*\right) \qquad [25]$$

which exhibits an interference pattern shifted with respect to that without the magnetic field: as $\boldsymbol{B}$ and therefore $\Phi$ change, dark and bright interference fringes alternate periodically at the screen $\Pi$, with period $\Phi_0$. This is the magnetic A–B effect, which has been quantitatively verified in many experiments, the first one in 1960 by Chambers. The effect is:

1. *gauge invariant*, since $\boldsymbol{B}$ and therefore $\Phi$ are gauge invariant;
2. *nonlocal*, since it depends on the magnetic field inside the solenoid, where the electrons never enter;
3. *quantum mechanical*, since classically the charges do not feel any force and therefore no effect would be expected in this limit; and
4. *topological*, since the electrons necessarily move in a non-simply-connected space.

But perhaps the most important implication of the A–B effect is a dramatic additional confirmation of the nonlocal character of quantum mechanics: the electron has to "travel" along the two paths (I and II) simultaneously; on the contrary, no flux would be surrounded and then no shift of the (then nonexistent) interference fringes would be observed at the screen $\Pi$.

Calculations in the path-integral approach including the whole set of homotopy classes of paths around the solenoid, indexed by an integer $m$, have been performed by several authors, leading to a formula of the type

$$\psi_P = \sum_{m=-\infty}^{\infty} e^{-im\delta}\psi_P^0(m) \qquad [26]$$

with

$$\delta = 2\pi\frac{\Phi}{\Phi_0} \qquad [27]$$

(Schulman 1971, Kobe 1979). As in [23],

$$\psi_P(\Phi + k\Phi_0) = \psi_P(\Phi), \quad k \in \mathbb{Z} \qquad [28]$$

There is a close relation between the A–B effect and the Dirac quantization condition (DQC) in the presence of electric and magnetic charges: according to [25] (or [26]) the A–B effect disappears when the flux $\Phi$ equals $n\Phi_0 = 2\pi n(\hbar c/|e|)$, $n \in \mathbb{Z}$, that is, when the condition

$$|e|\Phi = nhc \qquad [29]$$

holds. But this is the DQC (Dirac 1931) when $\Phi$ is the flux associated with a magnetic charge $g$: $\Phi(g) = (g/4\pi r^2) \times 4\pi r^2 = g$, leading to $|e|g = nhc$ ($2\pi n$ in the n.s.u.). This is precisely the condition for the Dirac string to be unobservable in quantum mechanics: to give no A–B effect.

## Geometry of the A–B Effect

In this section we study the space of gauge classes of flat potentials outside the solenoid, which determine the A–B effect; the topological structure of the A–B bundle; and the holonomy groups of the connections, which precisely give the phase shifts of the wave functions. We use the n.s.u. system; in particular, if [L] is the unit of length, then $[A_\mu] = [\mathrm{L}]^{-1}$, $[|e|] = [\mathrm{L}]^0$, and $\Phi_0 = 2\pi/|e| = \sqrt{\pi/\alpha} \cong \sqrt{137\pi}$, where $\alpha$ is the fine structure constant.

To synthesize, one can say that *the abelian A–B effect is a nonlocal gauge-invariant quantum effect due to the coupling of the wave function (section of an associated bundle) to a nontrivial (non-exact) flat (closed) connection in a trivial principal bundle with a non-simply-connected base space*. In the following subsections, we will give a detailed explanation of these statements.

### The A–B Bundle

The gauge group of electromagnetism is the abelian Lie group U(1) with Lie algebra (the tangent space at the identity) $u(1) = i\mathbb{R}$. In the limit of an infinitely long and infinitesimally thin solenoid carrying the magnetic flux $\Phi$, the space available to the electrons is the plane minus a point, that is, $\mathbb{R}^{2*}$, which is of the same homotopy type as the circle $S^1$. Then the set of isomorphism classes of U(1) bundles over $\mathbb{R}^{2*}$ is in one-to-one correspondence with the set of homotopy classes of maps from $S^0$ to $S^1$ (Steenrod 1951), which consists of only one point: if $f, g$:

$S^0 \to S^1$ are given by $f(1) = e^{i\varphi_1}$, $f(-1) = e^{i\varphi_2}$, $g(1) = e^{i\theta_1}$, and $g(-1) = e^{i\theta_2}$, then $H : S^0 \times [0,1] \to S^1$ given by $H(1,t) = e^{i((1-t)\varphi_1 + t\theta_1)}$ and $H(-1,t) = e^{i((1-t)\varphi_2 + t\theta_2)}$ is a homotopy between $f$ and $g$. Then, up to equivalence, the relevant bundle for the A–B effect is the product bundle

$$\xi_{A-B} : U(1) \to \mathbb{R}^{2*} \times U(1) \to \mathbb{R}^{2*} \qquad [30a]$$

Since $\mathbb{R}^{2*}$ is homeomorphic to an open disk minus a point $(D_0^2)^*$, then the total space of the bundle is homeomorphic to an open solid 2-torus minus a circle, since $(T_0^2)^* = (D_0^2)^* \times S^1$. Then the A–B bundle has the topological structure

$$\xi_{A-B} : S^1 \to (T_0^2)^* \to (D_0^2)^* \qquad [30b]$$

### The Gauge Group and the Moduli Space of Flat Connections

The gauge group of the bundle $\xi_{A-B}$ is the set of smooth functions from the base space to the structure group, that is, $\mathcal{G} = C^\infty(\mathbb{R}^{2*}, U(1))$. Since $\mathcal{G} \subset C^0(\mathbb{R}^{2*}, U(1)) = \{\text{continuous functions } \mathbb{R}^{2*} \to U(1)\}$ and $[\mathbb{R}^{2*}, U(1)] = \{\text{homotopy classes of continuous functions } \mathbb{R}^{2*} \to U(1)\} \cong [S^1, S^1] \cong \pi_1(S^1) \cong \mathbb{Z}$, given $f \in \mathcal{G}$ there exists a unique $n \in \mathbb{Z}$ such that $f$ is homotopic to $f_n (f \sim f_n)$, where $f_n : \mathbb{R}^{2*} \to U(1)$ is given by $f_n(re^{i\varphi}) = e^{in\varphi}$, $\varphi \in [0, 2\pi)$.

$\mathcal{G}$ acts on the space of flat connections on $\xi_{A-B}$ given by the closed $u(1)$-valued differential 1-forms on $\mathbb{R}^{2*}$:

$$\mathcal{C}_0 = \{\mathcal{A} \in \Omega^1(\mathbb{R}^{2*}; u(1)), d\mathcal{A} = 0\} \qquad [31]$$

through

$$\mathcal{C}_0 \times \mathcal{G} \to \mathcal{C}_0, \quad (\mathcal{A}, f) \to \mathcal{A} + f^{-1}df \qquad [32]$$

where $f^{-1}(x,y) = (f(x,y))^{-1}$. The moduli space

$$\mathcal{M}_0 = \frac{\mathcal{C}_0}{\mathcal{G}} = \{\text{gauge equivalence classes of flat connections on } \xi_{A-B}\} = \{[\mathcal{A}] = \{\mathcal{A} + f^{-1}df,\ f \in \mathcal{G}\}, \mathcal{A} \in \mathcal{C}_0\} \qquad [33]$$

is isomorphic to the circle $S^1$ with length 1. This can be seen as follows: the de Rham cohomology of $\mathbb{R}^{2*}$ with coefficients in $i\mathbb{R}$ in dimension 1 is

$$H^1_{DR}(\mathbb{R}^{2*}; i\mathbb{R}) = \{\lambda[\mathcal{A}_0]_{DR}, \lambda \in \mathbb{R}\} \cong H^1_{DR}(S^1; i\mathbb{R}) \cong \mathbb{R} \qquad [34]$$

where

$$\mathcal{A}_0 = i\frac{x\,dy - y\,dx}{x^2 + y^2} \in \mathcal{C}_0 \qquad [35]$$

is the connection that, once multiplied by $-|e|^{-1}$ (see below) generates the flux $-\Phi_0$ and therefore no A–B effect: $\mathcal{A}_0$ is closed ($d\mathcal{A}_0 = 0$) but not exact ($(x\,dy - y\,dx)/(x^2+y^2) = d\varphi$ only for $\varphi \in (0, 2\pi)$, $\varphi = 0$ is excluded); $[\mathcal{A}_0]_{DR} = \mathcal{A}_0 + d\beta$ with $\beta \in \Omega^0(R^{2*}; i\mathbb{R})$. $\beta$ gives an element of $\mathcal{G}$ through the composite $\exp \circ \beta : \mathbb{R}^{2*} \to U(1), (x,y) \mapsto e^{i\beta(x,y)}$. The A–B effect with flux $\Phi = -\lambda\Phi_0$ is produced by the connection $\mathcal{A} = \lambda\mathcal{A}_0$. To determine $\mathcal{M}_0$, one finds the smallest $\sigma \in \mathbb{R}$ such that $(\lambda + \sigma)\mathcal{A}_0 \sim \lambda\mathcal{A}_0$, that is, $(\lambda+\sigma)\mathcal{A}_0 \in [\lambda\mathcal{A}_0]$, which means, from [33], that $(\lambda+\sigma)\mathcal{A}_0 = \lambda\mathcal{A}_0 + f^{-1}df$ or $\sigma\mathcal{A}_0 = f^{-1}df$. For $\varphi \neq 0$, $\mathcal{A}_0 = i d\varphi$ and $f_1^{-1}df_1 = i d\varphi$, then $\sigma = 1$, and therefore $(\lambda+1)\mathcal{A}_0 \sim \lambda\mathcal{A}_0$, in particular $\mathcal{A}_0 \sim 0$.

A remark concerning the gauge group $\mathcal{G}$ is the following. In classical electrodynamics, according to [7a] and [7b], the symmetry group could be taken to be the additive group $(\mathbb{R}, +)$ instead of the multiplicative group U(1). Since $\mathbb{R}$ is contractible, then the gauge group would be $\mathcal{G}_{cl} = C^\infty(\mathbb{R}^{2*}, \mathbb{R})$ with $[\mathbb{R}^{2*}, \mathbb{R}] \cong 0$, so that the homomorphism $\Psi : \mathcal{G}_{cl} \to \mathcal{G}$, $\Psi(f)(x) = e^{if(x)}$ would not exhaust $\mathcal{G}$ since $\Psi(f) \in [1]$ for any $f \in \mathcal{G}_{cl}$: in fact, $H : \mathbb{R}^{2*} \times [0,1] \to U(1)$ given by $H(x,t) = e^{i(1-t)f(x)}$ is a homotopy between $\Psi(f)$ and 1. However, the quantization of electric charges implies that in fact the gauge group is U(1) and not $\mathbb{R}$. This is equivalent mathematically to the possible existence of magnetic monopoles which require nontrivial bundles for their description.

### Covariant Derivative, Parallel Transport, and Holonomy

Let $G$ be a matrix Lie group with Lie algebra $\boldsymbol{g}$, $B$ a differentiable manifold, $\xi : G \to P \xrightarrow{\pi} B$ a principal bundle, $V$ a vector space, $G \times V \to V$ an action, and $\xi_V : V \to P \times_G V \xrightarrow{\pi_V} B$ the corresponding associated vector bundle ($\xi_V$ is trivial if $\xi$ is trivial). Call $\Gamma(\xi_V)$ the sections of $\xi_V$, $\Gamma(TB)(\Gamma(TP))$ the sections of the tangent bundle of $B(P)$, and $\Gamma_{eq}(P, V)$ the set of functions $\gamma : P \to V$ satisfying $\gamma(pg) = g^{-1}\gamma(p)$ (equivariant functions from $P$ to $V$). $s \in \Gamma(\xi_V)$ induces $\gamma_s \in \Gamma_{eq}(P,V)$ with $\gamma_s(p) = v$, where $s(\pi(p)) = [p, v]$ and $\gamma \in \Gamma_{eq}(P,V)$ induces $s_\gamma \in \Gamma(\xi_V)$ with $s_\gamma(b) = [p, \gamma(p)]$, where $p \in \pi^{-1}(\{b\})$. If $H$ is a connection on $\xi$, that is, a smooth assignment of a (horizontal) vector subspace $H_p$ of $T_pP$ at each $p$ of $P$, algebraically determined by a smooth $\boldsymbol{g}$-valued 1-form $\omega$ on $P$ through $H_p = \ker(\omega_p)$, $s \in \Gamma(\xi_V)$, $X \in \Gamma(TB)$, and $X^\uparrow \in \Gamma(TP)$ the horizontal lifting of $X$ by $\omega$, then $X^\uparrow(\gamma_s) \in \Gamma_{eq}(P,V)$, and covariant

derivative of $s$ with respect to $\omega$ in the direction of $X$ is defined by

$$\nabla_X^\omega s := s_{X\uparrow(\gamma_s)} \quad [36a]$$

If $\phi:\pi^{-1}(U) \to U \times G$ is a local trivialization of $\xi$, $x^\mu$, $\mu = 1, \ldots, \dim B$ are local coordinates on $U$, and $e_i$, $i = 1, \ldots, \dim V$ is a basis of the local sections in $\pi_V^{-1}(U)$, then the local expression of [36a] is

$$\nabla^{\omega_U}_{X^\mu\partial/\partial x^\mu}(s^i e_i) = X^\mu\left(\delta_i^j \frac{\partial}{\partial x^\mu} + \mathcal{A}^j_{\mu i}\right)s^i e_j \quad [36b]$$

where

$$\mathcal{A}_{U_i}^{\ j} = \mathcal{A}^j_{\mu i}\mathrm{d}x^\mu = (\sigma^*\omega_U)^j_i \quad [36c]$$

is the geometrical gauge potential in $U$, given by the pullback of $\omega_U$, the restriction of $\omega$ to $\pi^{-1}(U)$, by the local section $\sigma: U \to \pi^{-1}(U)$, $\sigma(b) = \phi^{-1}(b,1)$. ($\mathcal{A}^j_{\mu i}$ is defined through $\nabla^{\omega_U}_{\partial/\partial x^\mu} e_i = \mathcal{A}^j_{\mu i} e_j$.) The operator

$$D^j_{\mu i} = \delta^j_i \frac{\partial}{\partial x^\mu} + \mathcal{A}^j_{\mu i} \quad [36d]$$

is the usual local covariant derivative. In an overlapping trivialization, [36b] is replaced by

$$\nabla^{\omega_{U'}}_{X^\mu\partial/\partial x^\mu}(s'^i e'_i) = X^\mu\left(\delta_i^j \frac{\partial}{\partial x^\mu} + \mathcal{A}'^j_{\mu i}\right)s'^i e'_j$$

with $e'_j = g^k_j e_k$ and $s'^i = g^{-1}{}^i_l s^l$ on $U \cap U'$, then the local potential transforms as

$$\mathcal{A}^j_{\mu l} = g^j_k \mathcal{A}'^k_{\mu i} g^{-1}{}^i_l + (\partial_\mu g^j_k) g^{-1}{}^k_l \quad [36e]$$

which for $G$ abelian has the form [32].

For each smooth path $c:[0,1] \to B$ joining the points $b$ and $b'$, and each $p \in P_b = \pi^{-1}(\{b\})$, there exists a unique path $c^\uparrow$ in $P$ through $p$ with $\dot{c}^\uparrow(t) \in H_{c(t)}$ for all $t \in [0,1]$. $c^\uparrow$ is the horizontal lifting of $c$ by $\omega$ through $p$. Thus, for each connection and path there exists a diffeomorphism $P^\omega_c: P_b \to P_{b'}$ called parallel transport. If $c$ is a loop at $b$, then $P^\omega_c \in \mathrm{Diff}(P_b)$ is called the holonomy of $\omega$ at $b$ along $c$. To the loop space of $B$ at $b$, $\Omega(B;b)$, corresponds a subgroup $\mathrm{Hol}^\omega_b$ of $\mathrm{Diff}(P_b)$ called the holonomy of $\omega$ at $b$. If $c \in \Omega(B;b)$ and $\beta$ is a lifting of $c$ through $q \in P_b$, then there exists a unique path $g:[0,1] \to G$ such that $c^\uparrow(t) = \beta(t)g(t)$ with $c^\uparrow(0) = qg(0) = p$; $g$ satisfies the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t} g(t) + \omega_{\beta(t)}(\dot{\beta}(t)) = 0 \quad [37]$$

whose solution is the time-ordered exponential

$$\begin{aligned} g(t)g(0)^{-1} &= T\exp\left(\int_0^t \mathrm{d}\tau\, \omega_{\beta(\tau)}(\dot{\beta}(\tau))\right) \\ &= 1 + \sum_{m=1}^\infty (-1)^m \int_0^t \mathrm{d}\tau_1 \omega_{\beta(\tau_1)}(\dot{\beta}(\tau_1)) \\ &\quad \times \int_0^{\tau_1} \mathrm{d}\tau_2 \omega_{\beta(\tau_2)}(\dot{\beta}(\tau_2)) \cdots \\ &\quad \times \int_0^{\tau_{m-1}} \mathrm{d}\tau_m \omega_{\beta(\tau_m)}(\dot{\beta}(\tau_m)) \end{aligned} \quad [38]$$

If $q = p$ then $g(0) = 1$. For each $p \in P$, the set of elements $g' \in G$ such that $c^\uparrow(1) = pg'$ for $c \in \Omega(B;\pi(p))$ is a subgroup of $G$, $\mathrm{Hol}^\omega_p$, called the holonomy *of* $\omega$ at $p$. (For each $p$, there exists a group isomorphism $\mathrm{Hol}^\omega_p \to \mathrm{Hol}^\omega_{\pi(p)}$, and if $p$ and $p'$ are connected by a horizontal curve, then $\mathrm{Hol}^\omega_p = \mathrm{Hol}^\omega_{p'}$; if all $p'$s in $P$ are horizontally connected, then $\mathrm{Hol}^\omega_p = G$ for all $p \in P$.) If $(U,\phi)$ is a local trivialization of $\xi$, $c \subset U$, and $\beta(t) = \sigma(c(t))$, then one has the local formula

$$c^\uparrow(t) = \phi^{-1}(c(t),1)\left(T\exp\left(-\int_{c(0)}^{c(t)} \mathcal{A}_U\right)\right)g(0) \quad [39]$$

In particular, if $\xi$ is a product bundle, then $\phi$ is the identity, and choosing $g(0)=1$ gives

$$c^\uparrow(t) = \left(c(t), T\exp\left(-\int_{c(0)}^{c(t)} \mathcal{A}_U\right)\right) \quad [40]$$

In our case, $V = \mathbb{C}$, $\xi$ is a product bundle, $s = \psi$, the wave function, is a global section of the associated bundle

$$\xi_\mathbb{C}: \mathbb{C} \to \mathbb{R}^{2*} \times \mathbb{C} \xrightarrow{\pi_\mathbb{C}} \mathrm{R}^{2*} \quad [41]$$

$G = \mathrm{U}(1)$ with $\boldsymbol{g} = \mathrm{i}\mathbb{R}$ and an action $\mathrm{U}(1) \times \mathbb{C} \to \mathbb{C}$, $(\mathrm{e}^{\mathrm{i}\varphi}, z) \mapsto \mathrm{e}^{\mathrm{i}\varphi} z$; therefore, $\mathcal{A}_\mu = \mathcal{A}_{0\mu} = \mathrm{i}a_\mu$ with $a_\mu$ real valued, and the covariant derivative is

$$D_\mu\psi = \left(\frac{\partial}{\partial x_\mu} + \mathrm{i}a_\mu\right)\psi \quad [36f]$$

If $\psi$ carries the electric charge $q$, we define the physical gauge potential $A_\mu$ through

$$a_\mu = qA_\mu \quad [42]$$

and, for the covariant derivative, after multiplying by i, we obtain the operator appearing in eqn [15], $\mathrm{i}D_\mu\psi = (\mathrm{i}(\partial/\partial x^\mu) - qA_\mu)\psi$: in fact, for the spatial part the coupling is $(\mathrm{i}\nabla + q\boldsymbol{A})\psi$, and for the temporal part one has $(\mathrm{i}\partial/\partial t - q\varphi)\psi$. For the electron, $q = -|e|$ and $a_\mu = -|e|A_\mu = -(2\pi/\Phi_0)A_\mu$.

For $c \in \Omega(\mathbb{R}^{2*};(x_0, y_0))$, which turns $n$ times around the solenoid at $(0, 0)$, eqn [40] gives

$$c^{\uparrow} = ((x_0, y_0), e^{-n \oint_c A}) = ((x_0, y_0), e^{-in \oint_c a}) \\ = ((x_0, y_0), e^{-i|e|n \oint_c A \cdot dx}) = ((x_0, y_0), e^{-2\pi i n \Phi/\Phi_0})$$

and therefore, for $\Phi/\Phi_0 = \lambda \in [0, 1)$ we have the holonomy groups

$$\mathrm{Hol}^{\omega(\Phi)}_{((x_0,y_0),1)} = \{e^{-2\pi i n(\Phi/\Phi_0)}\}_{n\in\mathbb{Z}} = \begin{cases} \mathbb{Z}_q, & \lambda = p/q,\ p, q \in \mathbb{Z},\ (p, q) = 1 \\ \mathbb{Z}, & \lambda \notin \mathbb{Q} \end{cases} \quad [43]$$

In the second case, $\mathrm{Hol}^{\omega(\Phi)}_{((x_0,y_0),1)}$ is dense in U(1): in fact, suppose that for $n_1, n_2 \in \mathbb{Z}$, $n_1 \neq n_2$, $e^{2\pi i n_1 \lambda} = e^{2\pi i n_2 \lambda}$, then $e^{2\pi i (n_1 - n_2)\lambda} = 1$ and so $(n_1 - n_2)\lambda = m$ for some $m \in \mathbb{Z}$; therefore, $\lambda \in \mathbb{Q}$, which is a contradiction.

Finally, we should mention that the A–B effect can be understood as a geometric phase *à la* Berry, though not necessarily through an adiabatic change of the parameters on which the Hamiltonian depends. The Berry potential $a_B$ turns out to be proportional to the real magnetic vector potential $\boldsymbol{A}$: in the n.s.u., and for electrons,

$$a_B = -|e|\boldsymbol{A} \quad [44]$$

### Nonabelian and Gravitational A–B Effects

Since the fundamental group $\Pi_1(\mathbb{R}^{2*}, (x_0, y_0)) \cong \mathbb{Z}$, eqn [43] shows that there is a homomorphism $\varphi(\omega)$: $\Pi_1(\mathbb{R}^{2*}, (x_0, y_0)) \to U(1)$, $\varphi(\omega)(n) = e^{-2\pi i n \lambda}$, with $\varphi(\omega)\ (\Pi_1(\mathbb{R}^{2*})) = \mathrm{Hol}^{\omega(\Phi)}_{((x_0,y_0),1)}$, which characterizes the A–B effect in that case. In general, an A–B effect in a $G$-bundle with a connection $\omega$ is characterized by a group homomorphism from the fundamental group of the base space $B$ onto the holonomy group of the connection, which is a subgroup of the structure group. The A–B effect is nonabelian if the holonomy group is nonabelian, which requires both $G$ and $\Pi_1(B, \boldsymbol{x})$ to be nonabelian. Examples with Yang–Mills and gravitational fields are considered in the literature.

### Acknowledgment

The author thanks the University of Valencia, Spain, where part of this work was done.

*See also:* Deformation Quantization and Representation Theory; Fractional Quantum Hall Effect; Geometric Phases; Moduli Spaces: An Introduction; Quantum Chromodynamics; Variational Techniques for Ginzburg–Landau Energies.

### Further Reading

Aguilar MA and Socolovsky M (2002) Aharonov–Bohm effect, flat connections, and Green's theorem. *International Journal of Theoretical Physics* 41: 839–860.
Aharonov Y and Bohm D (1959) Significance of electromagnetic potentials in the quantum theory. *Physical Review* 15: 485–491.
Berry MV (1984) Quantal phase factors accompanying adiabatic changes. *Proceedings of the Royal Society of London A* 392: 45–57.
Chambers RG (1960) Shift of an electron interference pattern by enclosed magnetic flux. *Physical Review Letters* 5: 3–5.
Corichi A and Pierri M (1995) Gravity and geometric phases. *Physical Review D* 51: 5870–5875.
Dirac PMA (1931) Quantised singularities in the electromagnetic field. *Proceedings of the Royal Society of London A* 133: 60–72.
Kobe DH (1979) Aharonov–Bohm effect revisited. *Annals of Physics* 123: 381–410.
Peshkin M and Tonomura A (1989) *The Aharonov–Bohm Effect.* Berlin: Springer.
Schulman LS (1971) Approximate topologies. *Journal of Mathematical Physics* 12: 304–308.
Steenrod N (1951) *The Topology of Fibre Bundles.* Princeton, NJ: Princeton University Press.
Sundrum R and Tassie LJ (1986) Non-abelian Aharonov–Bohm effects, Feynman paths, and topology. *Journal of Mathematical Physics* 27: 1566–1570.
Wu TT and Yang CN (1975) Concept of nonintegrable phase factors and global formulation of gauge fields. *Physical Review D* 12: 3845–3857.

# Algebraic Approach to Quantum Field Theory

**R Brunetti and K Fredenhagen**, Universität Hamburg, Hamburg, Germany



### Introduction

Quantum field theory may be understood as the incorporation of the principle of locality, which is at the basis of classical field theory, into quantum physics. There are, however, severe obstacles against a straightforward translation of concepts of classical field theory into quantum theory, among them the notorious divergences of quantum field theory and the intrinsic nonlocality of quantum physics. Therefore, the concept of locality is somewhat obscured in the formalism of quantum field theory as it is typically exposed in textbooks. Nonlocal concepts such as the vacuum, the notion of particles or the $S$-matrix play a fundamental role, and neither the

relation to classical field theory nor the influence of background fields can be properly treated.

Algebraic quantum field theory (AQFT; synonymously, local quantum physics), on the contrary, aims at emphasizing the concept of locality at every instance. As the nonlocal features of quantum physics occur at the level of states ("entanglement"), not at the level of observables, it is better not to base the theory on the Hilbert space of states but on the algebra of observables. Subsystems of a given system then simply correspond to subalgebras of a given algebra. The locality concept is abstractly encoded in a notion of independence of subsystems; two subsystems are independent if the algebra of observables which they generate is isomorphic to the tensor product of the algebras of the subsystems.

Spacetime can then – in the spirit of Leibniz – be considered as an ordering device for systems. So, one associates with regions of spacetime the algebras of observables which can be measured in the pertinent region, with the condition that the algebras of subregions of a given region can be identified with subalgebras of the algebra of the region.

Problems arise if one aims at a generally covariant approach in the spirit of general relativity. Then, in order to avoid pitfalls like in the "hole problem," systems corresponding to isometric regions must be isomorphic. Since isomorphic regions may be embedded into different spacetimes, this amounts to a simultaneous treatment of all spacetimes of a suitable class. We will see that category theory furnishes such a description, where the objects are the systems and the morphisms the embeddings of a system as a subsystem of other systems.

States arise as secondary objects via Hilbert space representations, or directly as linear functionals on the algebras of observables which can be interpreted as expectation values and are, therefore, positive and normalized. It is crucial that inequivalent representations ("sectors") can occur, and the analysis of the structure of the sectors is one of the big successes of AQFT. One can also study the particle interpretation of certain states as well as (equilibrium and nonequilibrium) thermodynamical properties.

The mathematical methods in AQFT are mainly taken from the theory of operator algebras, a field of mathematics which developed in close contact to mathematical physics, in particular to AQFT. Unfortunately, the most important field theories, from the point of view of elementary particle physics, as quantum electrodynamics or the standard model could not yet be constructed beyond formal perturbation theory with the annoying consequence that it seemed that the concepts of AQFT could not be applied to them. However, it has recently been shown that formal perturbation theory can be reshaped in the spirit of AQFT such that the algebras of observables of these models can be constructed as algebras of formal power series of Hilbert space operators. The price to pay is that the deep mathematics of operator algebras cannot be applied, but the crucial features of the algebraic approach can be used.

AQFT was originally proposed by Haag as a concept by which scattering of particles can be understood as a consequence of the principle of locality. It was then put into a mathematically precise form by Araki, Haag, and Kastler. After the analysis of particle scattering by Haag and Ruelle and the clarification of the relation to the Lehmann–Symanzik–Zimmermann (LSZ) formalism by Hepp, the structure of superselection sectors was studied first by Borchers and then in a fundamental series of papers by Doplicher, Haag, and Roberts (DHR) (see, e.g., Doplicher *et al.* (1971, 1974)) (soon after Buchholz and Fredenhagen established the relation to particles), and finally Doplicher and Roberts uncovered the structure of superselection sectors as the dual of a compact group thereby generalizing the Tannaka–Krein theorem of characterization of group duals.

With the advent of two-dimensional conformal field theory, new models were constructed and it was shown that the DHR analysis can be generalized to these models. Directly related to conformal theories is the algebraic approach to holography in anti-de Sitter (AdS) spacetime by Rehren.

The general framework of AQFT may be described as a covariant functor between two categories. The first one contains the information on local relations and is crucial for the interpretation. Its objects are topological spaces with additional structures (typically globally hyperbolic Lorentzian spaces, possibly spin bundles with connections, etc.), its morphisms being the structure-preserving embeddings. In the case of globally hyperbolic Lorentzian spacetimes, one requires that the embeddings are isometric and preserve the causal structure. The second category describes the algebraic structure of observables. In quantum physics the standard assumption is that one deals with the category of $C^*$-algebras where the morphisms are unital embeddings. In classical physics, one looks instead at Poisson algebras, and in perturbative quantum field theory one admits algebras which possess nontrivial representations as formal power series of Hilbert space operators. It is the leading principle of AQFT that the functor $\mathscr{A}$ contains all physical information. In particular, two theories are equivalent if the corresponding functors are naturally equivalent.

In the analysis of the functor $\mathscr{A}$, a crucial role is played by natural transformations from other functors on the locality category. For instance, a field $A$ may be defined as a natural transformation from the category of test function spaces to the category of observable algebras via their functors related to the locality category.

## Quantum Field Theories as Covariant Functors

The rigorous implementation of the generally covariant locality principle uses the language of category theory.

The following two categories are used:

**Loc**: The class of objects obj(**Loc**) is formed by all (smooth) $d$-dimensional ($d \geq 2$ is held fixed), globally hyperbolic Lorentzian spacetimes $M$ which are oriented and time oriented. Given any two such objects $M_1$ and $M_2$, the morphisms $\psi \in \hom_{\mathbf{Loc}}(M_1, M_2)$ are taken to be the isometric embeddings $\psi: M_1 \to M_2$ of $M_1$ into $M_2$ but with the following constraints:

(i) if $\gamma: [a, b] \to M_2$ is any causal curve and $\gamma(a), \gamma(b) \in \psi(M_1)$ then the whole curve must be in the image $\psi(M_1)$, that is, $\gamma(t) \in \psi(M_1)$ for all $t \in [a, b]$;

(ii) any morphism preserves orientation and time orientation of the embedded spacetime. The composition is defined as the composition of maps, the unit element in $\hom_{\mathbf{Loc}}(M, M)$ is given by the identical embedding $\mathrm{id}_M : M \mapsto M$ for any $M \in \mathrm{obj}(\mathbf{Loc})$.

**Obs**: The class of objects obj(**Obs**) is formed by all $C^*$-algebras possessing unit elements, and the morphisms are faithful (injective) unit-preserving $*$-homomorphisms. The composition is again defined as the composition of maps, the unit element in $\hom_{\mathbf{Obs}}(\mathcal{A}, \mathcal{A})$ is for any $\mathcal{A} \in \mathrm{obj}(\mathbf{Obs})$ given by the identical map $\mathrm{id}_{\mathcal{A}} : A \mapsto A, A \in \mathcal{A}$.

The categories are chosen for definitiveness. One may envisage changes according to particular needs, as, for instance, in perturbation theory where instead of $C^*$-algebras general topological $*$-algebras are better suited. Or one may use von Neumann algebras, in case particular states are selected. On the other hand, one might consider for **Loc** bundles over spacetimes, or (in conformally invariant theories) admit conformal embeddings as morphisms. In case one is interested in spacetimes which are not globally hyperbolic, one could look at the globally hyperbolic subregions (where one needs to be careful about the causal convexity condition (i) above).

The concept of locally covariant quantum field theory is defined as follows.

**Definition 1**

(i) A locally covariant quantum field theory is a covariant functor $\mathscr{A}$ from **Loc** to **Obs** and (writing $\alpha_\psi$ for $\mathscr{A}(\psi)$) with the covariance properties

$$\alpha_{\psi'} \circ \alpha_\psi = \alpha_{\psi' \circ \psi}, \qquad \alpha_{\mathrm{id}_M} = \mathrm{id}_{\mathscr{A}(M)}$$

for all morphisms $\psi \in \hom_{\mathbf{Loc}}(M_1, M_2)$, all morphisms $\psi' \in \hom_{\mathbf{Loc}}(M_2, M_3)$, and all $M \in \mathrm{obj}(\mathbf{Loc})$.

(ii) A locally covariant quantum field theory described by a covariant functor $\mathscr{A}$ is called "causal" if the following holds: whenever there are morphisms $\psi_j \in \hom_{\mathbf{Loc}}(M_j, M), j = 1, 2$, so that the sets $\psi_1(M_1)$ and $\psi_2(M_2)$ are causally separated in $M$, then one has

$$\left[\alpha_{\psi_1}(\mathscr{A}(M_1)), \alpha_{\psi_2}(\mathscr{A}(M_2))\right] = \{0\}$$

where the element-wise commutation makes sense in $\mathscr{A}(M)$.

(iii) One says that a locally covariant quantum field theory given by the functor $\mathscr{A}$ obeys the "time-slice axiom" if

$$\alpha_\psi(\mathscr{A}(M)) = \mathscr{A}(M')$$

holds for all $\psi \in \hom_{\mathbf{Loc}}(M, M')$ such that $\psi(M)$ contains a Cauchy surface for $M'$.

Thus, a quantum field theory is an assignment of $C^*$-algebras to (all) globally hyperbolic spacetimes so that the algebras are identifiable when the spacetimes are isometric, in the indicated way. This is a precise description of the generally covariant locality principle.

## The Traditional Approach

The traditional framework of AQFT, in the Araki–Haag–Kastler sense, on a fixed globally hyperbolic spacetime can be recovered from a locally covariant quantum field theory, that is, from a covariant functor $\mathscr{A}$ with the properties listed above.

Indeed, let $M$ be an object in obj(**Loc**). $\mathcal{K}(M)$ denotes the set of all open subsets in $M$ which are relatively compact and also contain, with each pair of points $x$ and $y$, all $\boldsymbol{g}$-causal curves in $M$ connecting $x$ and $y$ (cf. condition (i) in the definition of **Loc**). $O \in \mathcal{K}(M)$, endowed with the metric of $M$ restricted to $O$ and with the induced orientation and time orientation, is a member of obj(**Loc**), and the injection map $\iota_{M,O} : O \to M$, that is, the identical map restricted to $O$, is an element in $\hom_{\mathbf{Loc}}(O, M)$.

With this notation, it is easy to prove the following assertion:

**Theorem 1** *Let $\mathscr{A}$ be a covariant functor with the above-stated properties, and define a map $\mathcal{K}(M) \ni O \mapsto \mathcal{A}(O) \subset \mathscr{A}(M)$ by setting*

$$\mathcal{A}(O) := \alpha_{\iota_{M,O}}(\mathscr{A}(O))$$

*Then the following statements hold*:

(i) *The map fulfills isotony, that is,*

$$O_1 \subset O_2 \Rightarrow \mathcal{A}(O_1) \subset \mathcal{A}(O_2) \\ \text{for all } O_1, O_2 \in \mathcal{K}(M)$$

(ii) *If there exists a group G of isometric diffeomorphisms $\kappa : M \to M$ (so that $\kappa * \mathbf{g} = \mathbf{g}$) preserving orientation and time orientation, then there is a representation $G \ni \kappa \mapsto \tilde{\alpha}_\kappa$ of G by $C^*$-algebra automorphisms $\tilde{\alpha}_\kappa : \mathscr{A}(M) \to \mathscr{A}(M)$ such that*

$$\tilde{\alpha}_\kappa(\mathcal{A}(O)) = \mathcal{A}(\kappa(O)), \quad O \in \mathcal{K}(M)$$

(iii) *If the theory given by $\mathscr{A}$ is additionally causal, then it holds that*

$$[\mathcal{A}(O_1), \mathcal{A}(O_2)] = \{0\}$$

*for all $O_1, O_2 \in \mathcal{K}(M)$ with $O_1$ causally separated from $O_2$.*

These properties are just the basic assumptions of the Araki–Haag–Kastler framework.

## The Achievements of the Traditional Approach

In the Araki–Haag–Kastler approach in Minkowski spacetime $\mathbb{M}$, many results have been obtained in the last 40 years, some of them also becoming a source of inspiration to mathematics. A description of the achievements can be organized in terms of a length-scale basis, from the small to the large. We assume in this section that the algebra $\mathscr{A}(\mathbb{M})$ is faithfully and irreducibly represented on a Hilbert space $\mathcal{H}$, that the Poincaré transformations are unitarily implemented with positive energy, and that the subspace of Poincaré invariant vectors is one dimensional (uniqueness of the vacuum). Moreover, algebras correponding to regions which are spacelike to a nonempty open region are assumed to be weakly closed (i.e., von Neumann algebras on $\mathcal{H}$), and the condition of weak additivity is fulfilled, that is, for all $O \in \mathcal{K}(\mathbb{M})$ the algebra generated from the algebras $\mathcal{A}(O + x), x \in \mathbb{M}$ is weakly dense in $\mathscr{A}(\mathbb{M})$.

### Ultraviolet Structure and Idealized Localizations

This section deals with the problem of inspecting the theory at very small scales. In the limiting case, one is interested in idealized localizations, eventually the points of spacetimes. But the observable algebras are trivial at any point $x \in \mathbb{M}$, namely

$$\bigcap_{O \ni x} \mathcal{A}(O) = \mathbb{C}\mathbf{1}, \quad O \in \mathcal{K}(\mathbb{M})$$

Hence, pointlike localized observables are necessarily singular. Actually, the Wightman formulation of quantum field theory is based on the use of distributions on spacetime with values in the algebra of observables (as a topological $*$-algebra). In spite of technical complications whose physical significance is unclear, this formalism is well suited for a discussion of the connection with the Euclidean theory, which allows, in fortunate cases, a treatment by path integrals; it is more directly related to models and admits, via the operator-product expansion, a study of the short-distance behavior. It is, therefore, an important question how the algebraic approach is related to the Wightman formalism. The reader is referred to the literature for exploring the results on this relation.

Whereas these results point to an essential equivalence of both formalisms, one needs in addition a criterion for the existence of sufficiently many Wightman fields associated with a given local net. Such a criterion can be given in terms of a compactness condition to be discussed in the next subsection. As a benefit, one derives an operator-product expansion which has to be assumed in the Wightman approach.

In the purely algebraic approach, the ultraviolet structure has been investigated by Buchholz and Verch. Small-scale properties of theories are studied with the help of the so-called scaling algebras whose elements can be described as orbits of observables under all possible renormalization group motions. There results a classification of theories in the scaling limit which can be grouped into three broad classes: theories for which the scaling limit is purely classical (commutative algebras), those for which the limit is essentially unique (stable ultraviolet fixed point) and not classical, and those for which this is not the case (unstable ultraviolet fixed point). This classification does not rely on perturbation expansions. It allows an intrinsic definition of confinement in terms of the so-called ultraparticles, that is, particles which are visible only in the scaling limit.

### Phase-Space Analysis

As far as finite distances are concerned, there are two apparently competing principles, those of

nuclearity and modularity. The first one suggests that locally, after a cutoff in energy, one has a situation similar to that of old quantum mechanics, namely a finite number of states in a finite volume of phase space. Aiming at a precise formulation, Haag and Swieca introduced their notion of compactness, which Buchholz and Wichmann sharpened into that of nuclearity. The latter authors proposed that the set generated from the vacuum vector $\Omega$,

$$\{e^{-\beta H} A\Omega \mid A \in \mathcal{A}(O), \|A\| < 1\}$$

$H$ denoting the generator of time translations (Hamiltonian), is nuclear for any $\beta > 0$, roughly stating that it is contained in the image of the unit ball under a trace class operator. The nuclear size $Z(\beta, O)$ of the set plays the role of the partition function of the model and has to satisfy certain bounds in the parameter $\beta$. The consequence of this constraint is the existence of product states, namely those normal states for which observables localized in two given spacelike separated regions are uncorrelated. A further consequence is the existence of thermal equilibrium states (KMS states) for all $\beta > 0$.

The second principle concerns the fact that, even locally, quantum field theory has infinitely many degrees of freedom. This becomes visible in the Reeh–Schlieder theorem, which states that every vector $\Phi$ which is in the range of $e^{-\beta H}$ for some $\beta > 0$ (in particular, the vacuum $\Omega$) is cyclic and separating for the algebras $\mathcal{A}(O), O \in \mathcal{K}(\mathbb{M})$, that is, $\mathcal{A}(O)\Phi$ is dense in $\mathcal{H}$ ($\Phi$ is cyclic) and $A\Phi = 0, A \in \mathcal{A}(O)$ implies $A = 0$ ($\Phi$ is separating). The pair $(\mathcal{A}(O), \Omega)$ is then a von Neumann algebra in the so-called standard form. On such a pair, the Tomita–Takesaki theory can be applied, namely the densely defined operator

$$SA\Omega = A^*\Omega, \quad A \in \mathcal{A}(O)$$

is closable, and the polar decomposition of its closure $\bar{S} = J\Delta^{1/2}$ delivers an antiunitary involution $J$ (the modular conjugation) and a positive self-adjoint operator $\Delta$ (the modular operator) associated with the standard pair $(\mathcal{A}(O), \Omega)$. These operators have the properties

$$J\mathcal{A}(O)J = \mathcal{A}(O)'$$

where the prime denotes the commutant, and

$$\Delta^{it}\mathcal{A}(O)\Delta^{-it} = \mathcal{A}(O), \quad t \in \mathbb{R}$$

The importance of this structure is based on the fact disclosed by Bisognano and Wichmann using Poincaré-covariant Wightman fields and local algebras generated by them, that for specific regions in Minkowski spacetime the modular operators have a geometrical meaning. Indeed, these authors showed for the pair $(\mathcal{A}(W), \Omega)$, where $W$ denotes the wedge region $W = \{x \in \mathbb{M} \mid |x^0| < x^1\}$, that the associated modular unitary $\Delta^{it}$ is the Lorentz boost with velocity $\tanh(2\pi t)$ in the direction 1 and that the modular conjugation $J$ is the $CP_1T$ symmetry operator with parity $P_1$ the reflection with respect to the $x^1 = 0$ plane. Later, Borchers discovered that already on the purely algebraic level a corresponding structure exists. He proved that, given any standard pair $(\mathcal{A}, \Phi)$ and a one-parameter group of unitaries $\tau \to U(\tau)$ acting on the Hilbert space $\mathcal{H}$ with a positive generator and such that $\Phi$ is invariant and $U(\tau)\mathcal{A}U(\tau)^* \subset \mathcal{A}, \tau > 0$, then the associated modular operators $\Delta$ and $J$ fulfill the commutation relations

$$\Delta^{it}U(\tau)\Delta^{-it} = U(e^{-2\pi t}\tau)$$
$$JU(\tau)J = U(-\tau)$$

which are just the commutation relations between boosts and lightlike translations.

Surprisingly, there is a direct connection between the two concepts of nuclearity and modularity. Indeed, in the nuclearity condition, it is possible to replace the Hamiltonian operator by a specific function of the modular operator associated with a slightly larger region. Furthermore, under mild conditions, nuclearity and modularity together determine the structure of local algebras completely; they are isomorphic to the unique hyperfinite type $III_1$ von Neumann algebra.

### Sectors, Symmetries, Statistics, and Particles

Large scales are appropriate for discussing global issues like superselection sectors, statistics and symmetries as far as large spacelike distances are concerned, and scattering theory, with the resulting notions of particles and infraparticles, as far as large timelike distances are concerned.

In purely massive theories, where the vacuum sector has a mass gap and the mass shell of the particles are isolated, a very satisfactory description of the multiparticle structure at large times can be given. Using the concept of almost local particle generators,

$$\Psi = A(t)\Omega$$

where $\Psi$ is a single-particle state (i.e., an eigenstate of the mass operator), $A(t)$ is a family of almost local operators essentially localized in the kinematical region accessible from a given point by a motion with the velocities contained in the spectrum of $\Psi$, one obtains the multiparticle states as limits of products $A_1(t)\cdots A_n(t)\Omega$ for disjoint velocity supports. The corresponding closed subspaces are

invariant under Poincaré transformations and are unitarily equivalent to the Fock spaces of noninteracting particles.

For massless particles, no almost-local particle generators can be expected to exist. In even dimensions, however, one can exploit Huygens principle to construct asymptotic particle generators which are in the commutant of the algebra of the forward or backward lightcone, respectively. Again, their products can be determined and multiparticle states obtained.

Much less well understood is the case of massive particles in a theory which also possesses massless particles. Here, in general, the corresponding states are not eigenstates of the mass operator. Since quantum electrodynamics (QED) as well as the standard model of elementary particles have this problem, the correct treatment of scattering in these models is still under discussion. One attempt to a correct treatment is based on the concept of the so-called particle weights, that is, unbounded positive functionals on a suitable algebra. This algebra is generated by positive almost-local operators annihilating the vacuum and interpreted as counters.

The structure at large spacelike scales may be analyzed by the theory of superselection sectors. The best-understood case is that of locally generated sectors which are the objects of the DHR theory. Starting from a distinguished representation $\pi_0$ (vacuum representation) which is assumed to fulfill the Haag duality,

$$\pi_0\big(\mathcal{A}(O)\big) = \pi_0\big(\mathcal{A}(O')\big)'$$

for all double cones $O$, one may look at all representations which are equivalent to the vacuum representation if restricted to the observables localized in double cones in the spacelike complement of a given double cone. Such representations give rise to endomorphisms of the algebra of observables, and the product of endomorphisms can be interpreted as a product of sectors ("fusion"). In general, these representations violate the Haag duality, but there is a subclass of the so-called finite statistics sectors where the violation of Haag duality is small, in the sense that the nontrivial inclusion

$$\pi\big(\mathcal{A}(O)\big) \subset \pi\big(\mathcal{A}(O')\big)'$$

has a finite Jones index. These sectors form (in at least three spacetime dimensions) a symmetric tensor category with some further properties which can be identified, in a generalization of the Tannaka–Krein theorem, as the dual of a unique compact group. This group plays the role of a global gauge group. The symmetry of the category is expressed in terms of a representation of the symmetric group. One may then enlarge the algebra of observables and obtain an algebra of operators which transform covariantly under the global gauge group and satisfy Bose or Fermi commutation relations for spacelike separation.

In two spacetime dimensions, one obtains instead braided tensor categories. They have been classified under additional conditions (conformal symmetry, central charge $c < 1$) in a remarkable work by Kawahigashi and Longo. Moreover, in their paper, one finds that by using completely new methods ($Q$-systems) a new model is unveiled, apparently inaccessible by methods used by others. To some extent, these categories can be interpreted as duals of generalized quantum groups.

The question arises whether all representations describing elementary particles are, in the massive case, DHR representations. One can show that in the case of a representation with an isolated mass shell there is an associated vacuum representation which becomes equivalent to the particle representation after restriction to observables localized spacelike to a given infinitely extended spacelike cone. This property is weaker than the DHR condition but allows, in four spacetime dimensions, the same construction of a global gauge group and of covariant fields with Bose and Fermi commutation relations, respectively, as the DHR condition. In three space dimensions, however, one finds a braided tensor category, which has similar properties as those known from topological field theories in three dimensions.

The sector structure in massless theories is not well understood, due to the infrared problem. This is in particular true for QED.

## Fields as Natural Transformations

In order to be able to interpret the theory in terms of measurements, one has to be able to compare observables associated with different regions of spacetime, or, even different spacetimes. In the absence of nontrivial isometries, such a comparison can be made in terms of locally covariant fields. By definition, these are natural transformations from the functor of quantum field theory to another functor on the category of spacetimes **Loc**.

The standard case is the functor which associates with every spacetime $M$ its space $\mathcal{D}(M)$ of smooth compactly supported test functions. There, the morphisms are the pushforwards $\mathcal{D}\psi \equiv \psi_*$.

**Definition 2** A locally covariant quantum field $\Phi$ is a natural transformation between the functors $\mathscr{D}$ and $\mathscr{A}$, that is, for any object $M$ in obj(**Loc**) there exists a morphism $\Phi_M : \mathcal{D}(M) \to \mathscr{A}(M)$ such that for

any pair of objects $M_1$ and $M_2$ and any morphism $\psi$ between them, the following diagram commutes:

$$\begin{array}{ccc} \mathcal{D}(M_1) & \xrightarrow{\Phi_{M_1}} & \mathcal{A}(M_1) \\ \psi_* \downarrow & & \downarrow \alpha_\psi \\ \mathcal{D}(M_2) & \underset{\Phi_{M_2}}{\longrightarrow} & \mathcal{A}(M_2) \end{array}$$

The commutativity of the diagram means, explicitly, that

$$\alpha_\psi \circ \Phi_{M_1} = \Phi_{M_2} \circ \psi_*$$

which is the requirement sought for the covariance of fields. It contains, in particular, the standard covariance condition for spacetime isometries.

Fields in the above sense are not necessarily linear. Examples for fields which are also linear are the scalar massive free Klein–Gordon fields on all globally hyperbolic spacetimes and its locally covariant Wick polynomials. In particular, the energy–momentum tensors can be constructed as locally covariant fields, and they provide a crucial tool for discussing the back-reaction problem for matter fields.

An example for the more general notion of a field are the local $S$-matrices in the Stückelberg–Bogolubov–Epstein–Glaser sense. These are unitaries $S_M(\lambda)$ with $M \in \mathrm{obj}(\mathbf{Loc})$ and $\lambda \in \mathcal{D}(M)$ which satisfy the conditions

$$S_M(0) = 1$$
$$S_M(\lambda + \mu + \nu) = S_M(\lambda + \mu) S_M(\mu)^{-1} S_M(\mu + \nu)$$

for $\lambda, \mu, \nu \in \mathcal{D}(M)$ such that the supports of $\lambda$ and $\nu$ can be separated by a Cauchy surface of $M$ with $\mathrm{supp}\,\lambda$ in the future of the surface.

The importance of these $S$-matrices relies on the fact that they can be used to define a new quantum field theory. The new theory is locally covariant if the original theory is and if the local $S$-matrices satisfy the condition of the locally covariant field above. A perturbative construction of interacting quantum field theory on globally hyperbolic spacetimes was completed in this way by Hollands and Wald, based on previous work by Brunetti and Fredenhagen.

*See also:* Axiomatic Quantum Field Theory; Constructive Quantum Field Theory; Current Algebra; Deformation Quantization and Representation Theory; Dispersion Relations; Indefinite Metric; Integrability and Quantum Field Theory; Operads; Perturbative Renormalization Theory and BRST; Quantum Central Limit Theorems; Quantum Field Theory: A Brief Introduction; Quantum Field Theory in Curved Spacetime; Quantum Fields with Indefinite Metric: Non-Trivial Models; Quantum Fields with Topological Defects; Quantum Geometry and its Applications; Scattering in Relativistic Quantum Field Theory: Fundamental Concepts and Tools; Scattering in Relativistic Quantum Field Theory: The Analytic Program; Spin Foams; Symmetries in Quantum Field Theory: Algebraic Aspects; Symmetries in Quantum Field Theory of Lower Spacetime Dimensions; Tomita–Takesaki Modular Theory; Two-Dimensional Models; von Neumann Algebras: Introduction, Modular Theory and Classification Theory; von Neumann Algebras: Subfactor Theory.

## Further Reading

Araki H (1999) *Mathematical Theory of Quantum Fields.* Oxford: Oxford University Press.

Baumgærtel H and Wollemberg M (1992) *Causal Nets of Operator Algebras.* Berlin: Akademie Verlag.

Borchers HJ (1996) *Translation Group and Particle Representation in Quantum Field Theory*, Lecture Notes in Physics. New Series m: Monographs, 40. Berlin: Springer.

Borchers HJ (2000) On revolutionizing quantum field theory with Tomita's modular theory. *Journal of Mathematical Physics* 41: 3604–3673.

Bratteli O and Robinson DW (1987) *Operator Algebras and Quantum Statistical Mechanics*, vol. 1 Berlin: Springer.

Brunetti R and Fredenhagen K (2000) Microlocal analysis and interacting quantum field theories: renormalization on physical backgrounds. *Communications in Mathematical Physics* 208: 623–661.

Brunetti R, Fredenhagen K, and Verch R (2003) The generally covariant locality principle – a new paradigm for local quantum field theory. *Communications in Mathematical Physics* 237: 31–68.

Buchholz D and Haag R (2000) The quest for understanding in relativistic quantum physics. *Journal of Mathematical Physics* 41: 3674–3697.

Dixmier J (1964) *Les C*-algèbres et leurs représentations.* Paris: Gauthier-Villars.

Doplicher S, Haag R, and Roberts JE (1971) Local observables and particle statistics I. *Communications in Mathematical Physics* 23: 199–230.

Doplicher S, Haag R, and Roberts JE (1974) Local observables and particle statistics II. *Communications in Mathematical Physics* 35: 49–85.

Evans DE and Kawahigashi Y (1998) *Quantum Symmetries on Operator Algebras.* New York: Clarendon Press.

Haag R (1996) *Local Quantum Physics*, 2nd edn. Berlin: Springer.

Haag R and Kastler D (1964) An algebraic approach to quantum field theory. *Journal of Mathematical Physics* 5: 848–861.

Hollands S and Wald RM (2001) Local Wick polynomials and time ordered products of quantum fields in curved spacetime. *Communications in Mathematical Physics* 223: 289–326.

Hollands S and Wald RM (2002) Existence of local covariant time ordered products of quantum field in curved spacetime. *Communications in Mathematical Physics* 231: 309–345.

Kastler D (ed.) (1990) *The Algebraic Theory of Superselection Sectors. Introductions and Recent Results.* Singapore: World Scientific.

Kawahigashi Y and Longo R (2004) Classification of local conformal nets. Case $c < 1$. *Annals of Mathematics* 160: 1–30.

Takesaki M (2003) *Theory of Operator Algebras I, II, III*, Encyclopedia of Mathematical Sciences, vols. 124, 125, 127. Berlin: Springer.

Wald RM (1994) *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics.* Chicago: University of Chicago Press.

**Anderson Localization** *see* Localization for Quasi-Periodic Potentials

# Anomalies

**S L Adler**, Institute for Advanced Study, Princeton, NJ, USA



## Synopsis

*Anomalies* are the breaking of classical symmetries by quantum mechanical radiative corrections, which arise when the regularizations needed to evaluate small fermion loop Feynman diagrams conflict with a classical symmetry of the theory. They have important implications for a wide range of issues in quantum field theory, mathematical physics, and string theory.

## Chiral Anomalies, Abelian and Nonabelian

Consider quantum electrodynamics, with the fermionic Lagrangian density

$$\mathcal{L} = \bar{\psi}(\mathrm{i}\gamma^{\mu}\partial_{\mu} - e_0\gamma^{\mu}B_{\mu} - m_0)\psi \qquad [1a]$$

where $\bar{\psi} = \psi^{\dagger}\gamma^0$, $e_0$ and $m_0$ are the bare charge and mass, and $B_{\mu}$ is the electromagnetic gauge potential. (We reserve the notation $A$ for axial-vector quantities.) Under a chiral transformation

$$\psi \to \mathrm{e}^{\mathrm{i}\lambda\gamma_5}\psi \qquad [1b]$$

with constant $\lambda$, the kinetic term in eqn [1a] is invariant (because $\gamma_5$ commutes with $\gamma^0\gamma^{\mu}$), whereas the mass term is not invariant. Therefore, naive application of Noether's theorem would lead one to expect that the axial-vector current

$$j^5_{\mu} = \bar{\psi}\gamma_{\mu}\gamma_5\psi \qquad [1c]$$

obtained from the Lagrangian density by applying a chiral transformation with spatially varying $\lambda$, should have a divergence given by the change under chiral transformation of the mass term in eqn [1a]. Up to tree approximation, this is indeed true, but when one computes the $AVV$ Feynman diagram with one axial-vector and two vector vertices (see Figure 1), and insists on conservation of the vector current $j_{\mu} = \bar{\psi}\gamma_{\mu}\psi$, one finds that to order $e_0^2$, the classical Noether theorem is modified to read

$$\partial^{\mu}j^5_{\mu}(x) = 2\mathrm{i}m_0 j^5(x) + \frac{e_0^2}{16\pi^2}F^{\xi\sigma}(x)F^{\tau\rho}(x)\epsilon_{\xi\sigma\tau\rho} \qquad [2]$$

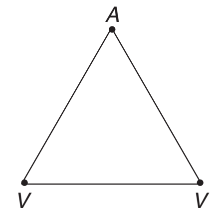



**Figure 1** The *AVV* triangle diagram responsible for the abelian chiral anomaly.

with $F^{\xi\sigma}(x) = \partial^{\sigma}B^{\xi}(x) - \partial^{\xi}B^{\sigma}(x)$ the electromagnetic field strength tensor. The second term in eqn [2], which would be unexpected from the application of the classical Noether theorem, is the abelian axial-vector anomaly (often called the Adler–Bell–Jackiw (or ABJ) anomaly after the seminal papers on the subject). Since vector current conservation, together with the axial-vector current anomaly, implies that the left- and right-handed chiral currents $j_{\mu} \pm j^5_{\mu}$ are also anomalous, the axial-vector anomaly is frequently called the "chiral anomaly," and we shall use the terms interchangeably in this article.

There are a number of different ways to understand why the extra term in eqn [2] appears. (1) Working through the formal Feynman diagrammatic Ward identity proof of the Noether theorem, one finds that there is a step where the closed fermion loop contributions are eliminated by a shift of the loop-integration variable. For Feynman diagrams that are convergent, this is not a problem, but the $AVV$ diagram is linearly divergent. The linear divergence vanishes under symmetric integration, but the shift then produces a finite residue, which gives the anomaly. (2) If one defines the $AVV$ diagram by Pauli–Villars regularization with regulator mass $M_0$ that is allowed to approach infinity at the end of the calculation, one finds a classical Noether theorem in the regulated theory,

$$\partial^{\mu}j^5_{\mu}|_{m_0} - \partial^{\mu}j^5_{\mu}|_{M_0} = 2\mathrm{i}m_0 j^5|_{m_0} - 2\mathrm{i}M_0 j^5|_{M_0} \qquad [3a]$$

with the subscripts $m_0$ and $M_0$ indicating that fermion loops are to be calculated with fermion mass $m_0$ and $M_0$, respectively. Taking the vacuum to two-photon matrix element of eqn [3a], one finds that the matrix element $\langle 0|j^5|_{M_0}|\gamma\gamma\rangle$, which is unambiguously computable after imposing vector-current conservation, falls off only as $M_0^{-1}$ as the regulator mass approaches infinity. Thus, the product of $2\mathrm{i}M_0$ with this matrix element has a finite limit, which gives the anomaly. (3) If the

gauge-invariant axial-vector current is defined by point-splitting

$$j^5_\mu(x) = \bar{\psi}(x+\epsilon/2)\gamma_\mu\gamma_5\psi(x-\epsilon/2)\mathrm{e}^{-\mathrm{i}e_0\epsilon^\sigma B_\sigma(x)} \quad [3b]$$

with $\epsilon \to 0$ at the end of the calculation, one observes that the divergence of eqn [3b] contains an extra term with a factor of $\epsilon$. On careful evaluation, one finds that the coefficient of this factor is an expression that behaves as $\epsilon^{-1}$, which gives the anomaly in the limit of vanishing $\epsilon$. (4) Finally, if the field theory is defined by a functional integral over the classical action, the standard Noether analysis shows that the classical action is invariant under the chiral transformation of eqn [1b], apart from the contribution of the mass term, which gives the naive axial-vector divergence. However, as pointed out by Fujikawa, the chiral transformation must also be applied to the functional integration measure, and since the measure is an infinite product, it must be regularized to be well defined. Careful calculation shows that the regularized measure is not chiral invariant, but contributes an extra term to the axial-vector Ward identity that is precisely the chiral anomaly.

A key feature of the anomaly is that it is irreducible: a local polynomial counter term cannot be added to the $AVV$ diagram that preserves vector-current conservation and eliminates the anomaly. More generally, one can show that there is no way of modifying quantum electrodynamics so as to eliminate the chiral anomaly, without spoiling either vector-current conservation (i.e., electromagnetic gauge invariance), renormalizability, or unitarity. Thus, the chiral anomaly is a new physical effect in renormalizable quantum field theory, which is not present in the prequantization classical theory.

The abelian chiral anomaly is the simplest case of the anomaly phenomenon. It was extended to nonabelian gauge theories by Bardeen using a point-splitting method to compute the divergence, followed by adding polynomial counter terms to remove as many of the residual terms as possible. The resulting irreducible divergence is the nonabelian chiral anomaly, which in terms of Yang–Mills field strengths for vector and axial-vector gauge potentials $V^\mu$ and $A^\mu$,

$$\begin{aligned} F_V^{\mu\nu}(x) &= \partial^\mu V^\nu(x) - \partial^\nu V^\mu(x) - \mathrm{i}[V^\mu(x), V^\nu(x)] \\ &\quad - \mathrm{i}[A^\mu(x), A^\nu(x)] \\ F_A^{\mu\nu}(x) &= \partial^\mu A^\nu(x) - \partial^\nu A^\mu(x) - \mathrm{i}[V^\mu(x), A^\nu(x)] \\ &\quad - \mathrm{i}[A^\mu(x), V^\nu(x)] \end{aligned} \quad [4a]$$

is given by

$$\begin{aligned} \partial^\mu j^a_{5\mu}(x) &= \text{normal divergence term} \\ &\quad + (1/4\pi^2)\epsilon_{\mu\nu\sigma\tau}\mathrm{tr}\lambda^a_A[(1/4)F_V^{\mu\nu}(x)F_V^{\sigma\tau}(x) \\ &\quad + (1/12)F_A^{\mu\nu}(x)F_A^{\sigma\tau}(x) \\ &\quad + (2/3)\mathrm{i}A^\mu(x)A^\nu(x)F_V^{\sigma\tau}(x) \\ &\quad + (2/3)\mathrm{i}F_V^{\mu\nu}(x)A^\sigma(x)A^\tau(x) \\ &\quad + (2/3)\mathrm{i}A^\mu(x)F_V^{\nu\sigma}(x)A^\tau(x) \\ &\quad - (8/3)A^\mu(x)A^\nu(x)A^\sigma(x)A^\tau(x)] \end{aligned} \quad [4b]$$

In eqn [4b], "tr" denotes a trace over internal degrees of freedom, and $\lambda^a_A$ is the internal symmetry matrix associated with the axial-vector external field. In the abelian case, where there is no internal symmetry structure, the terms involving two or four factors of $A^\mu, A^\nu, \ldots$ vanish by antisymmetry of $\epsilon_{\mu\nu\sigma\tau}$, and one recovers the $AVV$ triangle anomaly, as well as a kinematically related anomaly in the $AAA$ triangle diagram. In the nonabelian case, with nontrivial internal symmetry structure, there are also box- and pentagon-diagram anomalies.

In addition to coupling to spin-1 gauge fields, fermions can also couple to spin-2 gauge fields, associated with the graviton. When the coupling of fermions to gravitation is taken into account, the axial-vector current $\bar{\psi}T\gamma_\mu\gamma_5\psi$, with $T$ an internal symmetry matrix, has an additional anomalous contribution to its divergence proportional to

$$\mathrm{tr}\, T\epsilon_{\xi\sigma\tau\rho}R^{\xi\sigma\alpha\beta}R^{\tau\rho}{}_{\alpha\beta} \quad [4c]$$

where $R_{\xi\sigma\tau\rho}$ is the Riemann curvature tensor of the gravitational field.

## Chiral Anomaly Nonrenormalization

A salient feature of the chiral anomaly is the fact that it is not renormalized by higher-order radiative corrections. In other words, the one-loop expressions of eqns [2] and [4b] give the exact anomaly coefficient without modification in higher orders of perturbation theory. In gauge theories such as quantum electrodynamics and quantum chromodynamics, this result (the Adler–Bardeen theorem) can be understood heuristically as follows. Write down a modified Lagrangian, in which regulators are included for all gauge-boson fields. Since the gauge-boson regulators do not influence the chiral-symmetry properties of the theory, the divergences of the chiral currents are not affected by their inclusion, and so the only sources of anomalies in the regularized theory are small single-fermion loops, giving the anomaly expressions of eqns [2] and [4b]. Since the renormalized theory is obtained as the limit of

the regularized theory as the regulator masses approach infinity, this result applies to the renormalized theory as well.

The above argument can be made precise, and extends to nongauge theories such as the $\sigma$-model as well. For both gauge theories and the $\sigma$-model, cancellation of radiative corrections to the anomaly coefficient has been explicitly demonstrated in fourth-order calculations. Nonperturbative demonstrations of anomaly renormalization have also been given using the Callan–Symanzik equations. For example, in quantum electrodynamics, Zee, and Lowenstein and Schroer, showed that a factor $f$ that gives the ratio of the true anomaly to its one-loop value obeys the differential equation

$$\left(m\frac{\partial}{\partial m}+\alpha\beta(\alpha)\frac{\partial}{\partial\alpha}\right)f=0 \qquad [5]$$

Since $f$ is dimensionless, it can have no dependence on the mass $m$, and since $\beta(\alpha)$ is nonzero this implies $\partial f/\partial\alpha=0$. Thus, $f$ has no dependence on $\alpha$, and so $f=1$.

## Applications of Chiral Anomalies

Chiral anomalies have numerous applications in the standard model of particle physics and its extensions, and we describe here a few of the most important ones.

### Neutral Pion Decay $\pi^0\to\gamma\gamma$

As a result of the abelian chiral anomaly, the partially conserved axial-vector current (PCAC) equation relevant to neutral pion decay is modified to read

$$\partial^\mu\mathcal{F}^5_{3\mu}(x) = \left(f_\pi\mu_\pi^2/\sqrt{2}\right)\phi_\pi(x)+S\frac{\alpha_0}{4\pi}F^{\xi\sigma}(x)F^{\tau\rho}(x)\epsilon_{\xi\sigma\tau\rho} \qquad [6a]$$

with $\mu_\pi$ the pion mass, $f_\pi\simeq 131\,\text{MeV}$ the charged-pion decay constant, and $S$ a constant determined by the constituent fermion charges and axial-vector couplings. Taking the matrix element of eqn [6a] between the vacuum state and a two-photon state, and using the fact that the left-hand side has a kinematic zero (the Sutherland–Veltman theorem), one sees that the $\pi^0\to\gamma\gamma$ amplitude $F$ is completely determined by the anomaly term, giving the formula

$$F=-(\alpha/\pi)2S\sqrt{2}/f_\pi \qquad [6b]$$

For a single set of fractionally charged quarks, the amplitude $F$ is a factor of three too small to agree with experiment; for three fractionally charged quarks (or an equivalent Han–Nambu triplet), eqn [6b] gives the correct neutral pion decay rate. This calculation was one of the first pieces of evidence for the color degree of freedom of quarks.

### Anomaly Cancellation in Gauge Theories

In quantum electrodynamics, the gauge particle (the photon) couples to the vector current, and so the anomalous conservation properties of the axial-vector current have no effect. The same statement holds for the gauge gluons in quantum chromodynamics, when treated in isolation from the other interactions. However, in the electroweak theory that embeds quantum electrodynamics in a theory of the weak force, the gauge particles (the $W^\pm$ and $Z$ intermediate bosons) couple to chiral currents, which are left- or right-handed linear combinations of the vector and axial-vector currents. In this case, the chiral anomaly leads to problems with the renormalizability of the theory, unless the anomalies cancel between different fermion species. Writing all fermions as left-handed, the condition for anomaly cancellation is

$$\begin{aligned}&\mathrm{tr}\{T_\alpha,T_\beta\}T_\gamma=\mathrm{tr}(T_\alpha T_\beta+T_\beta T_\alpha)T_\gamma=0\\&\quad\text{for all }\alpha,\beta,\gamma\end{aligned} \qquad [7]$$

with $T_\alpha$ the coupling matrices of gauge bosons to left-handed fermions. These conditions are obeyed in the standard model, by virtue of three nontrivial sum rules on the fermion gauge couplings being satisfied (four sum rules, if one includes the gravitational contribution to the chiral anomaly given in eqn [4c], which also cancels in the standard model). Note that anomaly cancellation in the locally gauged currents of the standard model does not imply anomaly cancellation in global-flavor currents. Thus, the flavor axial-vector current anomaly that gives the $\pi^0\to\gamma\gamma$ matrix element remains anomalous in the full electroweak theory. Anomaly cancellation imposes important constraints on the construction of grand unified models that combine the electroweak theory with quantum chromodynamics. For instance, in SU(5) the fermions are put into a $\bar{5}$ and 10 representation, which together, but not individually, are anomaly free. The larger unification groups SO(10) and $E_6$ satisfy eqn [7] for all representations, and so are automatically anomaly free.

### Instanton Physics and the Theta Vacuum

The theory of anomalies is intimately tied to the physics associated with instanton classical Yang–Mills theory solutions. Since the instanton field

strength is self-dual, the nonvanishing instanton Euclidean action

$$S_E = \int d^4x \frac{1}{4} F_{\mu\nu}F^{\mu\nu} = 8\pi^2 \qquad [8a]$$

implies that the integral of the pseudoscalar density $F_{\mu\nu}F_{\lambda\sigma}\epsilon^{\mu\nu\lambda\sigma}$ over the instanton is also nonzero,

$$\int d^4x F_{\mu\nu}F_{\lambda\sigma}\epsilon^{\mu\nu\lambda\sigma} = 64\pi^2 \qquad [8b]$$

Referring back to eqn [4b], this means that the integral of the nonabelian chiral anomaly for fermions in the background field of an instanton is an integer, which in the Minkowski space continuation has the interpretation of a topological winding number change produced by the instanton tunneling solution. This fact has a number of profound consequences. Since a vacuum with a definite winding number $|\nu\rangle$ is unstable under instanton tunneling, careful analysis shows that the nonabelian vacuum that has correct clustering properties is a Fourier superposition

$$|\theta\rangle = \sum_\nu e^{i\theta\nu}|\nu\rangle \qquad [8c]$$

giving rise to the $\theta$-vacuum of quantum chromodynamics, and a host of issues associated with (the lack of) strong *CP* violation, the Peccei–Quinn mechanism, and axion physics. Also, the fact that the integral of eqn [8b] is nonzero means that the U(1) chiral symmetry of quantum chromodynamics is broken by instantons, which as shown by 't Hooft resolves the longstanding "U(1) problem" of strong interactions, that of explaining why the flavor singlet pseudoscalar meson $\eta'$ is not light, unlike its flavor octet partners.

### Anomaly Matching Conditions

The anomaly structure of a theory, as shown by 't Hooft, leads to important constraints on the formation of massless composite bound states. Consider a theory with a set of left-handed fermions $\psi^{if}$, with *i* a "color" index acted on by a nonabelian gauge force, and *f* an ungauged family or "flavor" index. Suppose that the family multiplet structure is such that the global chiral symmetries associated with the flavor index have nonvanishing anomalies $\mathrm{tr}\{T_\alpha, T_\beta\}T_\gamma$. Then the 't Hooft condition asserts that if the color forces result in the formation of composite massless bound states of the original completely confined fermions, and if there is no spontaneous breaking of the original global flavor symmetries, then these bound states must contain left-handed spin-1/2 composites with a representation structure *S* that has the same anomaly coefficient as that in the underlying theory. In other words, we must have

$$\mathrm{tr}\{S_\alpha, S_\beta\}S_\gamma = \mathrm{tr}\{T_\alpha, T_\beta\}T_\gamma \qquad [9]$$

To prove this, one adjoins to the theory a set of right-handed spectator fermions $\psi^f$ with the same flavor structure as the original set, but which are not acted on by the color force. These right-handed fermions cancel the original anomaly, making the underlying theory anomaly free at zero color coupling; since dynamics cannot spontaneously generate anomalies, the theory, when the color dynamics is turned on, must also have no global chiral anomalies. This implies that the bound-state spectrum must conspire to cancel the anomalies associated with the right-handed spectators; in other words, the bound-state anomaly structure must match that of the original fermions. This anomaly matching condition has found applications in the study of the possible compositeness of quarks and leptons. It has also been applied to the derivation of nonperturbative dynamical results in whole classes of supersymmetric theories, where the combined tools of holomorphicity, instanton physics, and anomaly matching have given incisive results.

## Global Structure of Anomalies

We noted earlier that chiral anomalies are irreducible, in that they cannot be eliminated by adding a local polynomial counter-term to the action. However, anomalies can be described by a nonlocal effective action, obtained by integrating out the fermion field dynamics, and this point of view proves very useful in the nonabelian case. Starting with the abelian case for orientation, we note that if $A^\mu$ is an external axial-vector field, and we write an effective action $\Gamma[A]$, then the axial-vector current $j^5_\mu$ associated with $A^\mu$ is given (up to an overall constant) by the variational derivative expression

$$j^5_\mu(x) = \frac{\delta\Gamma[A]}{\delta A^\mu(x)} \qquad [10a]$$

and the abelian anomaly appears as the fact that the expression

$$\partial^\mu j^5_\mu = X\Gamma[A] = G \neq 0, \quad X = \partial^\mu \frac{\delta}{\delta A^\mu(x)} \qquad [10b]$$

is nonvanishing even when the theory is classically chiral invariant. Turning now to the nonabelian case, the variational derivative appearing in eqns [10a] and [10b] must be replaced by an appropriate

covariant derivative. In terms of the internal-symmetry component fields $A_\mu^a$ and $V_\mu^a$ of the Yang–Mills potentials of eqn [4a], one introduces operators

$$-X^a(x) = \partial^\mu \frac{\delta}{\delta A_\mu^a(x)} + f_{abc} V_\mu^b \frac{\delta}{\delta A_\mu^c(x)} + f_{abc} A_\mu^b \frac{\delta}{\delta V_\mu^c(x)}$$
$$-Y^a(x) = \partial^\mu \frac{\delta}{\delta V_\mu^a(x)} + f_{abc} V_\mu^b \frac{\delta}{\delta V_\mu^c(x)} + f_{abc} A_\mu^b \frac{\delta}{\delta A_\mu^c(x)} \qquad [11a]$$

with $f_{abc}$ the antisymmetric nonabelian group structure constants. The operators $X^a$ and $Y^a$ are easily seen to obey the commutation relations

$$[X^a(x), X^b(y)] = f_{abc}\delta(x-y)Y_c(x)$$
$$[X^a(x), Y^b(y)] = f_{abc}\delta(x-y)X_c(x) \qquad [11b]$$
$$[Y^a(x), Y^b(y)] = f_{abc}\delta(x-y)Y_c(x)$$

Let $\Gamma[V, A]$ be the effective action as a functional of the fields $V^\mu, A^\mu$, constructed so that the vector currents are covariantly conserved, as expressed formally by

$$Y^a \Gamma[V, A] = 0 \qquad [12a]$$

Then the nonabelian axial-vector current anomaly is given by

$$X^a \Gamma[V, A] = G^a \qquad [12b]$$

From eqns [12a] and [12b] and the first line of eqn [11b], we have

$$X^b G^a - X^a G^b = (X^b X^a - X^a X^b)\Gamma[V, A] \propto f_{abc} Y^c \Gamma[V, A] = 0 \qquad [12c]$$

which is the Wess–Zumino consistency condition on the structure of the anomaly $G^a$. It can be shown that this condition uniquely fixes the form of the nonabelian anomaly to be that of eqn [4b], up to an overall constant, which can be determined by comparison with the simplest anomalous $AVV$ triangle graph. A physical consequence of the consistency condition is that the $\pi^0 \to \gamma\gamma$ decay amplitude determines uniquely certain other anomalous amplitudes, such as $2\gamma \to 3\pi, \gamma \to 3\pi$, and a five pseudoscalar vertex.

Although the action $\Gamma[V, A]$ is necessarily nonlocal, Wess and Zumino were able to write down a local action, involving an auxiliary pseudoscalar field, that obeys the anomalous Ward identities and the consistency conditions. Subsequently, Witten gave a new construction of this local action, in terms of the integral of a fifth-rank antisymmetric tensor over a five-dimensional disk which has a four-dimensional space as its boundary. He also showed that requiring $e^{i\Gamma}$ to be independent of the choice of the spanning disk requires, in analogy with Dirac's quantization condition for monopole charge, the condition that the overall coefficient in the nonabelian anomaly be quantized in integer multiples. Comparison with the lowest-order triangle diagram shows that in the case of $SU(N_c)$ gauge theory, this integer is just the number of colors $N_c$. Thus, global considerations tightly constrain the nonabelian chiral anomaly structure, and dictate that up to an integer-proportionality constant, it must have the form given in eqns [4a] and [4b].

## Trace Anomalies

The discovery of chiral anomalies inspired the search for other examples of anomalous behavior. First indications of a perturbative trace anomaly obtained in a study of broken scale invariance by Coleman and Jackiw were shown by Crewther, and by Chanowitz and Ellis, to correspond to an anomaly in the three-point function $\theta_\sigma^\sigma V_\mu V_\nu$, where $\theta_\mu^\mu$ is the energy–momentum tensor. Letting $\Delta_{\mu\nu}(p)$ be the momentum space expression for this three-point function, and $\Pi_{\mu\nu}$ the corresponding $V_\mu V_\nu$ two-point function, the trace anomaly equation in quantum electrodynamics reads

$$\Delta_{\mu\nu}(p) = \left(2 - p_\sigma \frac{\partial}{\partial p_\sigma}\right)\Pi_{\mu\nu}(p) - \frac{R}{6\pi^2}(p_\mu p_\nu - \eta_{\mu\nu}p^2) \qquad [13a]$$

with the first term on the right-hand side the naive divergence, and the second term the trace anomaly, with anomaly coefficient $R$ given by

$$R = \sum_{i,\mathrm{spin}\,\frac{1}{2}} Q_i^2 + \frac{1}{4}\sum_{i,\mathrm{spin}\,0} Q_i^2 \qquad [13b]$$

The fact that there should be a trace anomaly can readily be inferred from a trace analog of the Pauli–Villars regulator argument for the chiral anomaly given in eqn [3a]. Letting $j = \bar{\psi}\psi$ be the scalar current in abelian electrodynamics, one has

$$\theta_\mu^\mu|_{m_0} - \theta_\mu^\mu|_{M_0} = m_0 j|_{m_0} - M_0 j|_{M_0} \qquad [13c]$$

Taking the vacuum to two-photon matrix element of this equation, and imposing vector-current conservation, one finds that the matrix element $\langle 0|j|_{M_0}|\gamma\gamma\rangle$ is proportional to $M_0^{-1}\langle 0|F_{\lambda\sigma}F^{\lambda\sigma}|\gamma\gamma\rangle_{M_0}$ for a large regulator mass, and so makes a

nonvanishing contribution to the right-hand side of eqn [13c], giving the lowest-order trace anomaly.

Unlike the chiral anomaly, the trace anomaly is renormalized in higher orders of perturbation theory; heuristically, the reason is that whereas boson field regulators do not affect the chiral symmetry properties of a gauge theory (which are determined just by the fermionic terms in the Lagrangian), they do alter the energy–momentum tensor, since gravitation couples to all fields, including regulator fields. An analysis using the Callan–Symanzik equations shows, however, that the trace anomaly is computable to all orders in terms of various renormalization group functions of the coupling. For example, in abelian electrodynamics, defining $\beta(\alpha)$ and $\delta(\alpha)$ by $\beta(\alpha) = (m/\alpha)\partial\alpha/\partial m$ and $1 + \delta(\alpha) = (m/m_0)\partial m_0/\partial m$, the trace of the energy–momentum tensor is given to all orders by

$$\theta^{\mu}_{\mu} = [1 + \delta(\alpha)]m_0\bar{\psi}\psi + \tfrac{1}{4}\beta(\alpha)N[F_{\lambda\sigma}F^{\lambda\sigma}] + \cdots \quad [14]$$

with $N[\ ]$ specifying conditions that make the division into two terms in eqn [14] unique, and with the ellipsis $\cdots$ indicating terms that vanish by the equations of motion. A similar relation holds in the nonabelian case, again with the $\beta$ function appearing as the coefficient of the anomalous $\mathrm{tr}\,N[F_{\lambda\sigma}F^{\lambda\sigma}]$ term.

Just as in the chiral anomaly case, when spin-0, spin-1/2, or spin-1 fields propagate on a background spacetime, there are curvature-dependent contributions to the trace anomaly, in other words, gravitational anomalies. These typically take the form of complicated linear combinations of terms of the form $R^2, R_{\mu\nu}R^{\mu\nu}, R_{\mu\nu\lambda\sigma}R^{\mu\nu\lambda\sigma}, R_{,\mu}{}^{;\mu}$, with coefficients depending on the matter fields involved.

In supersymmetric theories, the axial-vector current and the energy–momentum tensor are both components of the supercurrent, and so their anomalies imply the existence of corresponding supercurrent anomalies. The issue of how the nonrenormalization of chiral anomalies (which have a supercurrent generalization given by the Konishi anomaly), and the renormalization of trace anomalies, can coexist in supersymmetric theories originally engendered considerable confusion. This apparent puzzle is now understood in the context of a perturbatively exact expression for the $\beta$ function in supersymmetric field theories (the so-called NSVZ, for Novikov, Shifman, Vainshtein, and Zakharov, $\beta$ function). Supersymmetry anomalies can be used to infer the structure of effective actions in supersymmetric theories, and these in turn have important implications for possibilities for dynamical supersymmetry breaking. Anomalies may also play a role, through anomaly mediation, in communicating supersymmetry breaking in "hidden sectors" of a theory, which do not contain the physical fields that we directly observe, to the "physical sector" containing the observed fields.

## Further Anomaly Topics

The above discussion has focused on some of the principal features and applications of anomalies. There are further topics of interest in the physics and mathematics of anomalies that are discussed in detail in the references cited in the "Further reading" section. We briefly describe a few of them here.

### Anomalies in Other Spacetime Dimensions and in String Theory

The focus above has been on anomalies in four-dimensional spacetime, but anomalies of various types occur both in lower-dimensional quantum field theories (such as theories in two- and three-dimensional spacetimes) and in quantum field theories in higher-dimensional spacetimes (such as $N = 1$ supergravity in ten-dimensional spacetime). Anomalies also play an important role in the formulation and consistency of string theory. The bosonic string is consistent only in 26-dimensional spacetime, and the analogous supersymmetric string only in ten-dimensional spacetime, because in other dimensions both these theories violate Lorentz invariance after quantization. In the Polyakov path-integral formulation of these string theories, these special dimensions are associated with the cancellation of the Weyl anomaly, which is the relevant form of the trace anomaly discussed above. Yang–Mills, gravitational, and mixed Yang–Mills gravitational anomalies make an appearance both in $N = 1$ ten-dimensional supergravity and in superstring theory, and again special dimensions play a role. In these theories, only when the associated internal symmetry groups are either SO(32) or $E_8 \times E_8$ is elimination of all anomalies possible, by cancellation of hexagon-diagram anomalies with anomalous tree diagrams involving exchange of a massless antisymmetric two-form field. This mechanism, due to Green and Schwarz, requires the factorization of a sixth-order trace invariant that appears in the hexagon anomaly in terms of lower-order invariants, as well as two numerical conditions on the adjoint representation generator structure, restricting the allowed gauge groups to the two noted above.

### Covariant versus Consistent Anomalies; Descent Equations

The nonabelian anomaly of eqns [4a] and [4b] is called the "consistent anomaly," because it obeys the

Wess–Zumino consistency conditions of eqn [12c]. This anomaly, however, is not gauge covariant, as can be seen from the fact that it involves not only the Yang–Mills field strengths $F^{\mu\nu}_{V,A}$, but the potentials $V^{\mu}, A^{\mu}$ as well. It turns out to be possible, by adding appropriate polynomials to the currents, to transform the consistent anomaly to a form, called the "covariant anomaly," which is gauge covariant under gauge transformations of the potentials $V^{\mu}, A^{\mu}$. This anomaly, however, does not obey the Wess–Zumino consistency conditions, and cannot be obtained from variation of an effective action functional.

The consistent anomalies (but not the covariant anomalies) obey a remarkable set of relations, called the Stora–Zumino descent equations, which relate the abelian anomaly in $2n+2$ spacetime dimensions to the nonabelian anomaly in $2n$ spacetime dimensions. This set of equations has been interpreted physically by Callan and Harvey as reflecting the fact that the Dirac equation has chiral zero modes in the presence of strings in $2n+2$ dimensions and of domain walls in $2n+1$ dimensions.

### Anomalies and Fermion Doubling in Lattice Gauge Theories

A longstanding problem in lattice formulations of gauge field theories is that when fermions are introduced on the lattice, the process of discretization introduces an undesirable doubling of the fermion particle modes. In particular, when an attempt is made to put chiral gauge theories, such as the electroweak theory, on the lattice, one finds that the doublers eliminate the chiral anomalies, by cancellation between modes with positive and negative axial-vector charge. Thus, for a long time, it appeared doubtful whether chiral gauge theories could be simulated on the lattice. However, recent work has led to formulations of lattice fermions that use a mathematical analog of a domain wall to successfully incorporate chiral fermions and the chiral anomaly into lattice gauge theory calculations.

### Relation of Anomalies to the Atiyah–Singer Index Theorem

The singlet ($\lambda_A^a = 1$) anomaly of eqn [4b] is closely related to the Atiyah–Singer index theorem. Specifically, the Euclidean spacetime integral of the singlet anomaly constructed from a gauge field can be shown to give the index of the related Dirac operator for a fermion moving in that background gauge field, where the index is defined as the difference between the numbers of right- and left-handed zero-eigenvalue normalizable solutions of the Dirac equation. Since the index is a topological invariant, this again implies that the Euclidean spacetime integral of the anomaly is a topological invariant, as noted above in our discussion of instanton-related applications of anomalies.

## Retrospect

The wide range of implications of anomalies has surprised – even astonished – the founders of the subject. New anomaly applications have appeared within the last few years, and very likely the future will see continued growth of the area of quantum field theory concerned with the physics and mathematics of anomalies.

## Acknowledgment

This work is supported, in part, by the Department of Energy under grant #DE-FG02-90ER40542.

*See also:* Bosons and Fermions in External Fields; BRST Quantization; Effective Field Theories; Gauge Theories from Strings; Gerbes in Quantum Field Theory; Index Theorems; Lagrangian Dispersion (Passive Scalar); Lattice Gauge Theory; Nonperturbative and Topological Aspects of Gauge Theory; Quantum Electrodynamics and Its Precision Tests; Quillen Determinant; Renormalization: General Theory; Seiberg–Witten Theory.

## Further Reading


Adler SL (1969) Axial-vector vertex in spinor electrodynamics. *Physical Review* 177: 2426–2438.

Adler SL (1970) Perturbation theory anomalies. In: Deser S, Grisaru M, and Pendleton H (eds.) *Lectures on Elementary Particles and Quantum Field Theory*, vol. 1, pp. 3–164. Cambridge, MA: MIT Press.

Adler SL (2005) Anomalies to all orders. In: 't Hooft G (ed.) *Fifty Years of Yang–Mills Theory*, pp. 187–228. Singapore: World Scientific.

Adler SL and Bardeen WA (1969) Absence of higher order corrections in the anomalous axial-vector divergence equation. *Physical Review* 182: 1517–1536.

Bardeen W (1969) Anomalous ward identities in spinor field theories. *Physical Review* 184: 1848–1859.

Bell JS and Jackiw R (1969) A PCAC puzzle: $\pi^0 \to \gamma\gamma$ in the $\sigma$-model. *Nuovo Cimento A* 60: 47–61.

Bertlmann RA (1996) *Anomalies in Quantum Field Theory*. Oxford: Clarendon.

De Azcárraga JA and Izquierdo JM (1995) *Lie Groups, Lie Algebras, Cohomology and Some Applications in Physics*, ch. 10. Cambridge: Cambridge University Press.

Fujikawa K and Suzuki H (2004) *Path Integrals and Quantum Anomalies*. Oxford: Oxford University Press.

Golterman M (2001) Lattice chiral gauge theories. *Nuclear Physics Proceeding Supplements* 94: 189–203.

Green MB, Schwarz JH, and Witten E (1987) *Superstring Theory*. vol. 2, sects. 13.3–13.5. Cambridge: Cambridge University Press.

Hasenfratz P (2005) Chiral symmetry on the lattice. In: 't Hooft G (ed.) *Fifty Years of Yang–Mills Theory*, pp. 377–398. Singapore: World Scientific.

Jackiw R (1985) Field theoretic investigations in current algebra and topological investigations in quantum gauge theories. In: Treiman S, Jackiw R, Zumino B, and Witten E (eds.) *Current Algebra and Anomalies*. Singapore: World Scientific and Princeton: Princeton University Press.

Jackiw R (2005) Fifty years of Yang–Mills theory and our moments of triumph. In: 't Hooft G (ed.) *Fifty Years of Yang–Mills Theory*, pp. 229–251. Singapore: World Scientific.

Makeenko Y (2002) *Methods of Contemporary Gauge Theory*, ch. 3. Cambridge: Cambridge University Press.

Neuberger H (2000) Chiral fermions on the lattice. *Nuclear Physics Proceeding Supplements* 83: 67–76.

Polchinski J (1999) *String Theory*, vol. 1, sect. 3.4; vol. 2, sect. 12.2. Cambridge: Cambridge University Press.

Shifman M (1997) Non-perturbative dynamics in supersymmetric gauge theories. *Progress in Particle and Nuclear Physics* 39: 1–116.

van Nieuwenhuizen P (1988) *Anomalies in Quantum Field Theory: Cancellation of Anomalies in d = 10 Supergravity*. Leuven: Leuven University Press.

Volovik GE (2003) *The Universe in a Helium Droplet*, ch. 18. Oxford: Clarendon.

Weinberg S (1996) *The Quantum Theory of Fields, Vol. II Modern Applications*, ch. 22. Cambridge: Cambridge University Press.

Zee A (2003) *Quantum Field Theory in a Nutshell*, sect. IV.7. Princeton: Princeton University Press.


# Arithmetic Quantum Chaos

**J Marklof**, University of Bristol, Bristol, UK



## Introduction

The central objective in the study of quantum chaos is to characterize universal properties of quantum systems that reflect the regular or chaotic features of the underlying classical dynamics. Most developments of the past 25 years have been influenced by the pioneering models on statistical properties of eigenstates (Berry 1977) and energy levels (Berry and Tabor 1977, Bohigas *et al.* 1984). Arithmetic quantum chaos (AQC) refers to the investigation of quantum systems with additional arithmetic structures that allow a significantly more extensive analysis than is generally possible. On the other hand, the special number-theoretic features also render these systems nongeneric, and thus some of the expected universal phenomena fail to emerge. Important examples of such systems include the modular surface and linear automorphisms of tori ("cat maps") which will be described below.

The geodesic motion of a point particle on a compact Riemannian surface $\mathcal{M}$ of constant negative curvature is the prime example of an Anosov flow, one of the strongest characterizations of dynamical chaos. The corresponding quantum eigenstates $\varphi_j$ and energy levels $\lambda_j$ are given by the solution of the eigenvalue problem for the Laplace–Beltrami operator $\Delta$ (or Laplacian for short)

$$(\Delta + \lambda)\varphi = 0, \qquad \|\varphi\|_{L^2(\mathcal{M})} = 1 \qquad [1]$$

where the eigenvalues

$$\lambda_0 = 0 < \lambda_1 \le \lambda_2 \le \cdots \to \infty \qquad [2]$$

form a discrete spectrum with an asymptotic density governed by Weyl's law

$$\#\{j : \lambda_j \le \lambda\} \sim \frac{\text{Area}(\Gamma\backslash\mathbb{H})}{4\pi}\lambda, \quad \lambda \to \infty \qquad [3]$$

We rescale the sequence by setting

$$X_j = \frac{\text{Area}(\Gamma\backslash\mathbb{H})}{4\pi}\lambda_j \qquad [4]$$

which yields a sequence of asymptotic density 1. One of the central conjectures in AQC says that, if $\mathcal{M}$ is an arithmetic hyperbolic surface (see the next section for examples of this very special class of surfaces of constant negative curvature), the eigenvalues of the Laplacian have the same local statistical properties as independent random variables from a Poisson process (see, e.g., the surveys by Sarnak (1995) and Bogomolny *et al.* (1997)). This means that the probability of finding $k$ eigenvalues $X_j$ in randomly shifted interval $[X, X+L]$ of fixed length $L$ is distributed according to the Poisson law $L^k e^{-L}/k!$. The gaps between eigenvalues have an exponential distribution,

$$\frac{1}{N}\#\{j \le N : X_{j+1} - X_j \in [a,b]\} \to \int_a^b e^{-s}\,ds \qquad [5]$$

as $N \to \infty$, and thus eigenvalues are likely to appear in clusters. This is in contrast to the general expectation that the energy level statistics of generic chaotic systems follow the distributions of random matrix ensembles; Poisson statistics are usually associated with quantized integrable systems. Although we are at present far from a proof of [5], the deviation from random matrix theory is well understood (see the section "Eigenvalue statistics and Selberg trace formula").

Highly excited quantum eigenstates $\varphi_j (j \to \infty)$ (cf. Figure 1) of chaotic systems are conjectured to behave locally like random wave solutions of [1],

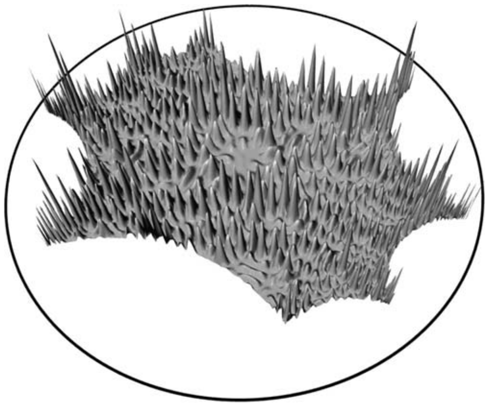


**Figure 1** Image of the absolute-value-squared of an eigenfunction $\varphi_j(z)$ for a nonarithmetic surface of genus 2. The surface is obtained by identifying opposite sides of the fundamental region. Reproduced from Aurich and Steiner (1993) Statistical properties of highly excited quantum eigenstates of a strongly chaotic system. *Physica D* 64(1–3): 185–214, with permission from R Aurich.

where boundary conditions are ignored. This hypothesis was put forward by Berry in 1977 and tested numerically, for example, in the case of certain arithmetic and nonarithmetic surfaces of constant negative curvature (Hejhal and Rackner 1992, Aurich and Steiner 1993). One of the implications is that eigenstates should have uniform mass on the surface $\mathcal{M}$, that is, for any bounded continuous function $g:\mathcal{M}\to\mathbb{R}$

$$\int_{\mathcal{M}} |\varphi_j|^2 g\, \mathrm{d}A \to \int_{\mathcal{M}} g\, \mathrm{d}A, \quad j\to\infty \qquad [6]$$

where $\mathrm{d}A$ is the Riemannian area element on $\mathcal{M}$. This phenomenon, referred to as quantum unique ergodicity (QUE), is expected to hold for general surfaces of negative curvature, according to a conjecture by Rudnick and Sarnak (1994). In the case of arithmetic hyperbolic surfaces, there has been substantial progress on this conjecture in the works of Lindenstrauss, Watson, and Luo–Sarnak (discussed later in this article; see also the review by Sarnak (2003)). For general manifolds with ergodic geodesic flow, the convergence in [6] is so far established only for subsequences of eigenfunctions of density 1 (Schnirelman–Zelditch–Colin de Verdière theorem, *see* Quantum Ergodicity and Mixing of Eigenfunctions), and it cannot be ruled out that exceptional subsequences of eigenfunctions have singular limit, for example, localized on closed geodesics. Such "scarring" of eigenfunctions, at least in some weak form, has been suggested by numerical experiments in Euclidean domains, and the existence of singular quantum limits is a matter of controversy in the current physics and mathematics literature. A first rigorous proof of the existence of scarred eigenstates has recently been established in the case of quantized toral automorphisms. Remarkably, these quantum cat maps may also exhibit QUE. A more detailed account of results for these maps is given in the section "Quantum eigenstates of cat maps"; see also Rudnick (2001) and De Bièvre (to appear).

There have been a number of other fruitful interactions between quantum chaos and number theory, in particular the connections of spectral statistics of integrable quantum systems with the value distribution properties of quadratic forms, and analogies in the statistical behavior of energy levels of chaotic systems and the zeros of the Riemann zeta function. We refer the reader to Marklof (2006) and Berry and Keating (1999), respectively, for information on these topics.

## Hyperbolic Surfaces

Let us begin with some basic notions of hyperbolic geometry. The hyperbolic plane $\mathbb{H}$ may be abstractly defined as the simply connected two-dimensional Riemannian manifold with Gaussian curvature –1. A convenient parametrization of $\mathbb{H}$ is provided by the complex upper-half plane, $\mathfrak{H}=\{x+\mathrm{i}y: x\in\mathbb{R}, y>0\}$, with Riemannian line and volume elements

$$\mathrm{d}s^2=\frac{\mathrm{d}x^2+\mathrm{d}y^2}{y^2}, \qquad \mathrm{d}A=\frac{\mathrm{d}x\,\mathrm{d}y}{y^2} \qquad [7]$$

respectively. The group of orientation-preserving isometries of $\mathbb{H}$ is given by fractional linear transformations

$$\mathfrak{H}\to\mathfrak{H}, \qquad z\mapsto\frac{az+b}{cz+d}$$
$$\begin{pmatrix} a & b\\ c & d\end{pmatrix}\in \mathrm{SL}(2,\mathbb{R}) \qquad [8]$$

where $\mathrm{SL}(2,\mathbb{R})$ is the group of $2\times 2$ matrices with unit determinant. Since the matrices 1 and –1 represent the same transformation, the group of orientation-preserving isometries can be identified with $\mathrm{PSL}(2,\mathbb{R}):=\mathrm{SL}(2,\mathbb{R})/\{\pm 1\}$. A finite-volume hyperbolic surface may now be represented as the quotient $\Gamma\backslash\mathbb{H}$, where $\Gamma\subset\mathrm{PSL}(2,\mathbb{R})$ is a Fuchsian group of the first kind. An arithmetic hyperbolic surface (such as the modular surface) is obtained, if $\Gamma$ has, loosely speaking, some representation in $n\times n$ matrices with integer coefficients, for some suitable $n$.

This is evident in the case of the modular surface, where the fundamental group is the modular group

$$\Gamma = \mathrm{PSL}(2,\mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}(2,\mathbb{R}) \colon a,b,c,d \in \mathbb{Z} \right\} / \{\pm 1\}$$

A fundamental domain for the action of the modular group $\mathrm{PSL}(2,\mathbb{Z})$ on $\mathfrak{H}$ is the set

$$\mathcal{F}_{\mathrm{PSL}(2,\mathbb{Z})} = \left\{ z \in \mathfrak{H} : |z| > 1, -\tfrac{1}{2} < \mathrm{Re}\, z < \tfrac{1}{2} \right\} \qquad [9]$$

(see **Figure 2**). The modular group is generated by the translation

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} : z \mapsto z+1$$

and the inversion

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} : z \mapsto -1/z$$

These generators identify sections of the boundary of $\mathcal{F}_{\mathrm{PSL}(2,\mathbb{Z})}$. By gluing the fundamental domain along identified edges, we obtain a realization of the modular surface, a noncompact surface with one cusp at $z \to \infty$, and two conic singularities at $z = \mathrm{i}$ and $z = 1/2 + \mathrm{i}\sqrt{3}/2$.

An interesting example of a compact arithmetic surface is the "regular octagon," a hyperbolic surface of genus 2. Its fundamental domain is shown in **Figure 3** as a subset of the Poincaré disc $\mathfrak{D} = \{z \in \mathbb{C} \colon |z| < 1\}$, which yields an alternative parametrization of the hyperbolic plane $\mathbb{H}$. In these coordinates, the Riemannian line and volume element read

$$\mathrm{d}s^2 = \frac{4(\mathrm{d}x^2 + \mathrm{d}y^2)}{(1 - x^2 - y^2)^2}, \quad \mathrm{d}A = \frac{4\mathrm{d}x\,\mathrm{d}y}{(1 - x^2 - y^2)^2} \qquad [10]$$

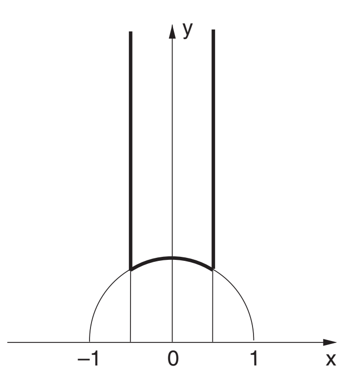



**Figure 2** Fundamental domain of the modular group PSL(2, *Z*) in the complex upper-half plane.

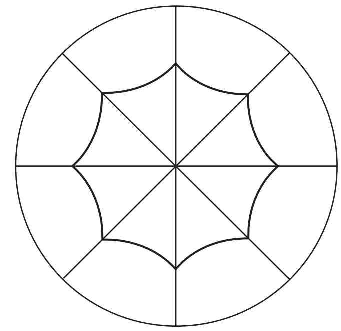


**Figure 3** Fundamental domain of the regular octagon in the Poincaré disk.

The group of orientation-preserving isometries is now represented by $\mathrm{PSU}(1,1) = \mathrm{SU}(1,1)/\{\pm 1\}$, where

$$\mathrm{SU}(1,1) = \left\{ \begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{pmatrix} \colon \alpha, \beta \in \mathbb{C}, |\alpha|^2 - |\beta|^2 = 1 \right\} \qquad [11]$$

acting on $\mathfrak{D}$ as above via fractional linear transformations. The fundamental group of the regular octagon surface is the subgroup of all elements in $\mathrm{PSU}(1,1)$ with coefficients of the form

$$\alpha = k + l\sqrt{2}, \quad \beta = (m + n\sqrt{2})\sqrt{1 + \sqrt{2}} \qquad [12]$$

where $k, l, m, n \in \mathbb{Z}[\mathrm{i}]$, that is, Gaussian integers of the form $k_1 + \mathrm{i}k_2, k_1, k_2 \in \mathbb{Z}$. Note that not all choices of $k, l, m, n \in \mathbb{Z}[\mathrm{i}]$ satisfy the condition $|\alpha|^2 - |\beta|^2 = 1$. Since all elements $\gamma \neq 1$ of $\Gamma$ act fix-point free on $\mathbb{H}$, the surface $\Gamma \backslash \mathbb{H}$ is smooth without conic singularities.

In the following, we will restrict our attention to a representative case, the modular surface with $\Gamma = \mathrm{PSL}(2,\mathbb{Z})$.

## Eigenvalue Statistics and Selberg Trace Formula

The statistical properties of the rescaled eigenvalues $X_j$ (cf. [4]) of the Laplacian can be characterized by their distribution in small intervals

$$\mathcal{N}(x, L) := \#\{j : x \le X_j \le x + L\} \qquad [13]$$

where $x$ is uniformly distributed, say, in the interval $[X, 2X]$, $X$ large. Numerical experiments by Bogomolny, Georgeot, Giannoni, and Schmit, as well as Bolte, Steil, and Steiner (see references in

Bogomolny (1997)) suggest that the $X_j$ are asymptotically Poisson distributed:

**Conjecture 1** For any bounded function $g:\mathbb{Z}_{\geq 0}\to\mathbb{C}$ we have

$$\frac{1}{X}\int_X^{2X} g(\mathcal{N}(x,L))\,\mathrm{d}x \to \sum_{k=0}^{\infty} g(k)\frac{L^k\mathrm{e}^{-L}}{k!} \qquad [14]$$

as $T\to\infty$.

One may also consider larger intervals, where $L\to\infty$ as $X\to\infty$. In this case, the assumption on the independence of the $X_j$ predicts a central-limit theorem. Weyl's law [3] implies that the expectation value is asymptotically, for $T\to\infty$,

$$\frac{1}{X}\int_X^{2X} \mathcal{N}(x,L)\,\mathrm{d}x \sim L \qquad [15]$$

This asymptotics holds for any sequence of $L$ bounded away from zero (e.g., $L$ constant, or $L\to\infty$).

Define the variance by

$$\Sigma^2(X,L) = \frac{1}{X}\int_X^{2X}(\mathcal{N}(x,L)-L)^2\,\mathrm{d}x \qquad [16]$$

In view of the above conjecture, one expects $\Sigma^2(X,L)\sim L$ in the limit $X\to\infty, L/\sqrt{X}\to 0$ (the variance exhibits a less universal behavior in the range $L\gg\sqrt{X}$ (the notation $A\ll B$ means there is a constant $c>0$ such that $A\leq cB$), cf. Sarnak (1995), and a central-limit theorem for the fluctuations around the mean:

**Conjecture 2** For any bounded function $g:\mathbb{R}\to\mathbb{C}$ we have

$$\frac{1}{X}\int_X^{2X} g\left(\frac{\mathcal{N}(x,L)-L}{\sqrt{\Sigma^2(x,L)}}\right)\mathrm{d}x \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g(t)\,\mathrm{e}^{-(1/2)t^2}\,\mathrm{d}t \qquad [17]$$

as $X,L\to\infty, L\ll X$.

The main tool in the attempts to prove the above conjectures has been the Selberg trace formula. It relates sums over eigenvalues of the Laplacians to sums over lengths of closed geodesics on the hyperbolic surface. The trace formula is in its simplest form in the case of compact hyperbolic surfaces; we have

$$\sum_{j=0}^{\infty} h(\rho_j) = \frac{\mathrm{Area}(\mathcal{M})}{4\pi}\int_{-\infty}^{\infty} h(\rho)\tanh(\pi\rho)\rho\,\mathrm{d}\rho + \sum_{\gamma\in H_*}\sum_{n=1}^{\infty}\frac{\ell_\gamma\, g(n\ell_\gamma)}{2\sinh(n\ell_\gamma/2)} \qquad [18]$$

where $H_*$ is the set of all primitive oriented closed geodesics $\gamma$, and $\ell_\gamma$ their lengths. The quantity $\rho_j$ is related to the eigenvalue $\lambda_j$ by the equation $\lambda_j=\rho_j^2+1/4$. The trace formula [18] holds for a large class of even test functions $h$. For example, it is sufficient to assume that $h$ is infinitely differentiable, and that the Fourier transform of $h$,

$$g(t) = \frac{1}{2\pi}\int_{\mathbb{R}} h(\rho)\,\mathrm{e}^{-\mathrm{i}\rho t}\,\mathrm{d}\rho \qquad [19]$$

has compact support. The trace formula for noncompact surfaces has additional terms from the parabolic elements in the corresponding group, and includes also sums over the resonances of the continuous part of the spectrum. The noncompact modular surface behaves in many ways like a compact surface. In particular, Selberg showed that the number of eigenvalues embedded in the continuous spectrum satisfies the same Weyl law as in the compact case (Sarnak 2003).

Setting

$$h(\rho) = \chi_{[X,X+L]}\left(\frac{\mathrm{Area}(\mathcal{M})}{4\pi}\left(\rho^2+\frac{1}{4}\right)\right) \qquad [20]$$

where $\chi_{[X,X+L]}$ is the characteristic function of the interval $[X,X+L]$, we may thus view $\mathcal{N}(X,L)$ as the left-hand side of the trace formula. The above test function $h$ is, however, not admissible, and requires appropriate smoothing. Luo and Sarnak (cf. Sarnak (2003)) developed an argument of this type to obtain a lower bound on the average number variance,

$$\frac{1}{L}\int_0^L \Sigma^2(X,L')\,\mathrm{d}L' \gg \frac{\sqrt{X}}{(\log X)^2} \qquad [21]$$

in the regime $\sqrt{X}/\log X\ll L\ll\sqrt{X}$, which is consistent with the Poisson conjecture $\Sigma^2(X,L)\sim L$. Bogomolny, Levyraz, and Schmit suggested a remarkable limiting formula for the two-point correlation function for the modular surface (cf. Bogomolny *et al.* (1997) and Bogomolny (2006)), based on an analysis of the correlations between multiplicities of lengths of closed geodesics. A rigorous analysis of the fluctuations of multiplicities is given by Peter (cf. Bogomolny (2006)). Rudnick (2005) has recently established a smoothed version of Conjecture 2 in the regime

$$\frac{\sqrt{X}}{L}\to\infty, \qquad \frac{\sqrt{X}}{L\log X}\to 0 \qquad [22]$$

where the characteristic function in [20] is replaced by a certain class of smooth test functions.

All of the above approaches use the Selberg trace formula, exploiting the particular properties of the

distribution of lengths of closed geodesics in arithmetic hyperbolic surfaces. These will be discussed in more detail in the next section, following the work of Bogomolny, Georgeot, Giannoni and Schmit, Bolte, and Luo and Sarnak (see Bogomolny *et al.* (1997) and Sarnak (1995) for references).

## Distribution of Lengths of Closed Geodesics

The classical prime geodesic theorem asserts that the number $N(\ell)$ of primitive closed geodesics of length less than $\ell$ is asymptotically

$$N(\ell) \sim \frac{e^{\ell}}{\ell} \qquad [23]$$

One of the significant geometrical characteristics of arithmetic hyperbolic surfaces is that the number of closed geodesics with the same length $\ell$ grows exponentially with $\ell$. This phenomenon is most easily explained in the case of the modular surface, where the set of lengths $\ell$ appearing in the lengths spectrum is characterized by the condition

$$2\cosh(\ell/2) = |\mathrm{tr}\ \gamma| \qquad [24]$$

where $\gamma$ runs over all elements in $\mathrm{SL}(2, \mathbb{Z})$ with $|\mathrm{tr}\gamma| > 2$. It is not hard to see that any integer $n > 2$ appears in the set $\{|\mathrm{tr}\ \gamma| : \gamma \in \mathrm{SL}(2, \mathbb{Z})\}$, and hence the set of distinct lengths of closed geodesics is

$$\mathcal{L} = \{2\ \mathrm{arcosh}(n/2) : n = 3, 4, 5, \ldots\} \qquad [25]$$

Therefore, the number of distinct lengths less than $\ell$ is asymptotically (for large $\ell$)

$$N'(\ell) = \#(\mathcal{L} \cap [0, \ell]) \sim e^{\ell/2} \qquad [26]$$

Equations [26] and [23] say that on average the number of geodesics with the same lengths is at least $\asymp e^{\ell/2}/\ell$.

The prime geodesic theorem [23] holds equally for all hyperbolic surfaces with finite area, while [26] is specific to the modular surface. For general arithmetic surfaces, we have the upper bound

$$N'(\ell) \leq c e^{\ell/2} \qquad [27]$$

for some constant $c > 0$ that may depend on the surface. Although one expects $N'(\ell)$ to be asymptotic to $(1/2)N(\ell)$ for generic surfaces (since most geodesics have a time-reversal partner which thus has the same length, and otherwise all lengths are distinct), there are examples of nonarithmetic Hecke triangles where numerical and heuristic arguments suggest $N'(\ell) \sim c_1 e^{c_2 \ell}/\ell$ for suitable constants $c_1 > 0$ and $0 < c_2 < 1/2$ (cf. Bogomolny (2006)). Hence exponential degeneracy in the length spectrum seems to occur in a weaker form also for nonarithmetic surfaces.

A further useful property of the length spectrum of arithmetic surfaces is the bounded clustering property: there is a constant $C$ (again surface dependent) such that

$$\#(\mathcal{L} \cap [\ell, \ell + 1]) \leq C \qquad [28]$$

for all $\ell$. This fact is evident in the case of the modular surface; the general case is proved by Luo and Sarnak (cf. Sarnak (1995)).

## Quantum Unique Ergodicity

The unit tangent bundle of a hyperbolic surface $\Gamma\backslash\mathbb{H}$ describes the physical phase space on which the classical dynamics takes place. A convenient parametrization of the unit tangent bundle is given by the quotient $\Gamma\backslash\mathrm{PSL}(2, \mathbb{R}$ – this may be seen be means of the Iwasawa decomposition for an element $g \in \mathrm{PSL}(2, \mathbb{R})$,

$$g = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix} \times \begin{pmatrix} \cos\theta/2 & \sin\theta/2 \\ -\sin\theta/2 & \cos\theta/2 \end{pmatrix} \qquad [29]$$

where $x + iy \in \mathfrak{H}$ represents the position of the particle in $\Gamma\backslash\mathbb{H}$ in half-plane coordinates, and $\theta \in [0, 2\pi)$ the direction of its velocity. Multiplying the matrix [29] from the left by $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and writing the result again in the Iwasawa form [29], one obtains the action

$$(z, \phi) \mapsto \left( \frac{az + b}{cz + d}, \theta - 2\arg(cz + d) \right) \qquad [30]$$

which represents precisely the geometric action of isometries on the unit tangent bundle.

The geodesic flow $\Phi^t$ on $\Gamma\backslash\mathrm{PSL}(2, \mathbb{R})$ is represented by the right translation

$$\Phi^t : \Gamma g \mapsto \Gamma g \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} \qquad [31]$$

The Haar measure $\mu$ on $\mathrm{PSL}(2, \mathbb{R})$ is thus trivially invariant under the geodesic flow. It is well known that $\mu$ is not the only invariant measure, that is, $\Phi^t$ is not uniquely ergodic, and that there is in fact an abundance of invariant measures. The simplest examples are those with uniform mass on one, or a countable collection of, closed geodesics.

To test the distribution of an eigenfunction $\varphi_j$ in phase space, one associates with a function

$a \in C^\infty(\Gamma\backslash PSL(2, \mathbb{R}))$ the quantum observable $Op(a)$, a zeroth order pseudodifferential operator with principal symbol $a$. Using semiclassical techniques based on Friedrich's symmetrization, one can show that the matrix element

$$\nu_j(a) = \langle Op(a)\varphi_j, \varphi_j \rangle \qquad [32]$$

is asymptotic (as $j \to \infty$) to a positive functional that defines a probability measure on $\Gamma\backslash PSL(2, \mathbb{R})$. Therefore, if $\mathcal{M}$ is compact, any weak limit of $\nu_j$ represents a probability measure on $\Gamma\backslash PSL(2, \mathbb{R})$. Egorov's theorem (*see* Quantum Ergodicity and Mixing of Eigenfunctions) in turn implies that any such limit must be invariant under the geodesic flow, and the main challenge in proving QUE is to rule out all invariant measures apart from Haar.

**Conjecture 3** (Rudnick and Sarnak (1994); see Sarnak (1995, 2003)). For every compact hyperbolic surface $\Gamma\backslash\mathbb{H}$, the sequence $\nu_j$ converges weakly to $\mu$.

Lindenstrauss has proved this conjecture for compact arithmetic hyperbolic surfaces of congruence type (such as the second example in the section "Hyperbolic surfaces") for special bases of eigenfunctions, using ergodic-theoretic methods. These will be discussed in more detail in the next section. His results extend to the noncompact case, that is, to the modular surface where $\Gamma = PSL(2, \mathbb{Z})$. Here he shows that any weak limit of subsequences of $\nu_j$ is of the form $c\mu$, where $c$ is a constant with values in $[0, 1]$. One believes that $c = 1$, but with present techniques it cannot be ruled out that a proportion of the mass of the eigenfunction escapes into the noncompact cusp of the surface. For the modular surface, $c = 1$ can be proved under the assumption of the generalized Riemann hypothesis (see the section "Eigenfunctions and $L$-functions" and Sarnak (2003)). QUE also holds for the continuous part of the spectrum, which is furnished by the Eisenstein series $E(z, s)$, where $s = 1/2 + ir$ is the spectral parameter. Note that the measures associated with the matrix elements

$$\nu_r(a) = \langle Op(a)E(\cdot, 1/2 + ir), E(\cdot, 1/2 + ir)\rangle \qquad [33]$$

are not probability measures but only Radon measures, since $E(z, s)$ is not square-integrable. Luo and Sarnak, and Jakobson have shown that

$$\lim_{r\to\infty} \frac{\nu_r(a)}{\nu_r(b)} = \frac{\mu(a)}{\mu(b)} \qquad [34]$$

for suitable test functions $a, b \in C^\infty(\Gamma\backslash PSL(2, \mathbb{R}))$ (cf. Sarnak (2003)).

## Hecke Operators, Entropy and Measure Rigidity

For compact surfaces, the sequence of probability measures approaching the matrix elements $\nu_j$ is relatively compact. That is, every infinite sequence contains a convergent subsequence. Lindenstrauss' central idea in the proof of QUE is to exploit the presence of Hecke operators to understand the invariance properties of possible quantum limits. We will sketch his argument in the case of the modular surface (ignoring issues related to the noncompactness of the surface), where it is most transparent.

For every positive integer $n$, the Hecke operator $T_n$ acting on continuous functions on $\Gamma\backslash\mathbb{H}$ with $\Gamma = SL(2, \mathbb{Z})$ is defined by

$$T_n f(z) = \frac{1}{\sqrt{n}} \sum_{\substack{a,d=1 \\ ad=n}}^{n} \sum_{b=0}^{d-1} f\left(\frac{az+b}{d}\right) \qquad [35]$$

The set $M_n$ of matrices with integer coefficients and determinant $n$ can be expressed as the disjoint union

$$M_n = \bigcup_{\substack{a,d=1 \\ ad=n}}^{n} \bigcup_{b=0}^{d-1} \Gamma \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \qquad [36]$$

and hence the sum in [35] can be viewed as a sum over the cosets in this decomposition. We note the product formula

$$T_m T_n = \sum_{d|\gcd(m,n)} T_{mn/d^2} \qquad [37]$$

The Hecke operators are normal, form a commuting family, and in addition they commute with the Laplacian $\Delta$. In the following, we consider an orthonormal basis of eigenfunctions $\varphi_j$ of $\Delta$ that are simultaneously eigenfunctions of all Hecke operators. We will refer to such eigenfunctions as Hecke eigenfunctions. The above assumption is automatically satisfied, if the spectrum of $\Delta$ is simple (i.e., no eigenvalues coincide), a property conjectured by Cartier and supported by numerical computations. Lindenstrauss' work is based on the following two observations. Firstly, all quantum limits of Hecke eigenfunctions are geodesic-flow invariant measures of positive entropy, and secondly, the only such measure of positive entropy that is recurrent under Hecke correspondences is the Lebesgue measure.

The first property is proved by Bourgain and Lindenstrauss (2003) and refines arguments of Rudnick and Sarnak (1994) and Wolpert (2001) on the distribution of Hecke points (see Sarnak (2003) for

references to these papers). For a given point $z \in \mathbb{H}$ the set of Hecke points is defined as

$$T_n(z) := M_n z \quad [38]$$

For most primes, the set $T_{p^k}(z)$ comprises $(p+1)p^{k-1}$ distinct points on $\Gamma\backslash\mathbb{H}$. For each $z$, the Hecke operator $T_n$ may now be interpreted as the adjacency matrix for a finite graph embedded in $\Gamma\backslash\mathbb{H}$, whose vertices are the Hecke points $T_n(z)$. Hecke eigenfunctions $\varphi_j$ with

$$T_n\varphi_j = \lambda_j(n)\varphi_j \quad [39]$$

give rise to eigenfunctions of the adjacency matrix. Exploiting this fact, Bourgain and Lindenstrauss show that for a large set of integers $n$

$$|\varphi_j(z)|^2 \ll \sum_{w\in T_n(z)} |\varphi_j(w)|^2 \quad [40]$$

that is, pointwise values of $|\varphi_j|^2$ cannot be substantially larger than its sum over Hecke points. This, and the observation that Hecke points for a large set of integers $n$ are sufficiently uniformly distributed on $\Gamma\backslash\mathbb{H}$ as $n\to\infty$, yields the estimate of positive entropy with a quantitative lower bound.

Lindenstrauss' proof of the second property, which shows that Lebesgue measure is the only quantum limit of Hecke eigenfunctions, is a result of a currently very active branch of ergodic theory: measure rigidity. Invariance under the geodesic flow alone is not sufficient to rule out other possible limit measures. In fact, there are uncountably many measures with this property. As limits of Hecke eigenfunctions, all quantum limits possess an additional property, namely recurrence under Hecke correspondences. Since the explanation of these is rather involved, let us recall an analogous result in a simpler setup. The map $\times 2 : x \mapsto 2x \bmod 1$ defines a hyperbolic dynamical system on the unit circle with a wealth of invariant measures, similar to the case of the geodesic flow on a surface of negative curvature. Furstenberg conjectured that, up to trivial invariant measures that are localized on finitely many rational points, Lebesgue measure is the only $\times 2$-invariant measure that is also invariant under action of $\times 3 : x \mapsto 3x \bmod 1$. This fundamental problem is still unsolved and one of the central conjectures in measure rigidity. Rudolph, however, showed that Furstenberg's conjecture is true if one restricts the statement to $\times 2$-invariant measures of positive entropy (cf. Lindenstrauss (to appear)). In Lindenstrauss' work, $\times 2$ plays the role of the geodesic flow, and $\times 3$ the role of the Hecke correspondences. Although here it might also be interesting to ask whether an analog of Furstenberg's conjecture holds, it is inessential for the proof of QUE due to the positive entropy of quantum limits discussed in the previous paragraph.

## Eigenfunctions and *L*-Functions

An even eigenfunction $\varphi_j(z)$ for $\Gamma = \mathrm{SL}(2,\mathbb{Z})$ has the Fourier expansion

$$\varphi_j(z) = \sum_{n=1}^{\infty} a_j(n) y^{1/2} K_{\mathrm{i}\rho_j}(2\pi n y)\cos(2\pi n x) \quad [41]$$

We associate with $\varphi_j(z)$ the Dirichlet series

$$L(s,\varphi_j) = \sum_{n=1}^{\infty} a_j(n) n^{-s} \quad [42]$$

which converges for Re $s$ large enough. These series have an analytic continuation to the entire complex plane $\mathbb{C}$ and satisfy a functional equation,

$$\Lambda(s,\varphi_j) = \Lambda(1-s,\varphi_j) \quad [43]$$

where

$$\Lambda(s,\varphi_j) = \pi^{-s}\Gamma\left(\frac{s+\mathrm{i}\rho_j}{2}\right)\Gamma\left(\frac{s-\mathrm{i}\rho_j}{2}\right)L(s,\varphi_j) \quad [44]$$

If $\varphi_j(z)$ is in addition an eigenfunction of all Hecke operators, then the Fourier coefficients in fact coincide (up to a normalization constant) with the eigenvalues of the Hecke operators

$$a_j(m) = \lambda_j(m) a_j(1) \quad [45]$$

If we normalize $a_j(1) = 1$, the Hecke relations [37] result in an Euler product formula for the $L$-function,

$$L(s,\varphi_j) = \prod_{p \text{ prime}} \left(1 - a_j(p)p^{-s} + p^{-1-2s}\right)^{-1} \quad [46]$$

These $L$-functions behave in many other ways like the Riemann zeta or classical Dirichlet $L$-functions. In particular, they are expected to satisfy a Riemann hypothesis, that is, all nontrivial zeros are constrained to the critical line $\mathrm{Im}\, s = 1/2$.

Questions on the distribution of Hecke eigenfunctions, such as QUE or value distribution properties, can now be translated to analytic properties of $L$-functions. We will discuss two examples.

The asymptotics in [6] can be established by proving [6] for the choices $g = \varphi_k, k = 1, 2, \ldots,$ that is,

$$\int_{\mathcal{M}} |\varphi_j|^2 \varphi_k \, \mathrm{d}A \to 0 \quad [47]$$

Watson discovered the remarkable relation (Sarnak 2003)

$$\left|\int_{\mathcal{M}} \varphi_{j_1}\varphi_{j_2}\varphi_{j_3}\,\mathrm{d}A\right|^2 = \frac{\pi^4 \Lambda(\frac{1}{2}, \varphi_{j_1}\times\varphi_{j_2}\times\varphi_{j_3})}{\Lambda(1,\mathrm{sym}^2\varphi_{j_1})\Lambda(1,\mathrm{sym}^2\varphi_{j_2})\Lambda(1,\mathrm{sym}^2\varphi_{j_3})} \quad [48]$$

The $L$-functions $\Lambda(s, g)$ in Watson's formula are more advanced cousins of those introduced earlier (see Sarnak (2003) for details). The Riemann hypothesis for such $L$-functions then implies, via [48], a precise rate of convergence to QUE for the modular surface,

$$\int_{\mathcal{M}} |\varphi_j|^2 g\,\mathrm{d}A = \int_{\mathcal{M}} g\,\mathrm{d}A + O(\lambda_j^{-1/4+\epsilon}) \quad [49]$$

for any $\epsilon > 0$, where the implied constant depends on $\epsilon$ and $g$.

A second example on the connection between statistical properties of the matrix elements $\nu_j(a) = \langle \mathrm{Op}(a)\varphi_j, \varphi_j\rangle$ (for fixed $a$ and random $j$) and values $L$-functions has appeared in the work of Luo and Sarnak (cf. Sarnak (2003)). Define the variance

$$V_\lambda(a) = \frac{1}{N(\lambda)}\sum_{\lambda_j\le\lambda}\left|\nu_j(a) - \mu(a)\right|^2 \quad [50]$$

with $N(\lambda) = \#\{j: \lambda_j \le \lambda\}$; cf. [3]. Following a conjecture by Feingold–Peres and Eckhardt *et al.* (see Sarnak (2003) for references) for "generic" quantum chaotic systems, one expects a central-limit theorem for the statistical fluctuations of the $\nu_j(a)$, where the normalized variance $N(\lambda)^{1/2}V_\lambda(a)$ is asymptotic to the classical autocorrelation function $C(a)$, see eqn [54].

**Conjecture 4** For any bounded function $g: \mathbb{R}\to\mathbb{C}$ we have

$$\frac{1}{N(\lambda)}\sum_{\lambda_j\le\lambda} g\left(\frac{\nu_j(a)-\mu(a)}{\sqrt{V_\lambda(a)}}\right) \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g(t)\mathrm{e}^{-(1/2)t^2}\mathrm{d}t \quad [51]$$

as $\lambda\to\infty$.

Luo and Sarnak prove that in the case of the modular surface the variance has the asymptotics

$$\lim_{\lambda\to\infty} N(\lambda)^{1/2}V_\lambda(a) = \langle Ba, a\rangle \quad [52]$$

where $B$ is a non-negative self-adjoint operator which commutes with the Laplacian $\Delta$ and all Hecke operators $T_n$. In particular, we have

$$B\varphi_j = \tfrac{1}{2}L(\tfrac{1}{2},\varphi_j)C(\varphi_j)\varphi_j \quad [53]$$

where

$$C(a) := \int_{\mathbb{R}}\int_{\Gamma\backslash\mathrm{PSL}(2,\mathbb{R})} a(\Phi^t(g))\overline{a(g)}\,\mathrm{d}\mu(g)\,\mathrm{d}t \quad [54]$$

is the classical autocorrelation function for the geodesic flow with respect to the observable $a$ (Sarnak 2003). Up to the arithmetic factor $(1/2)L(1/2, \varphi_j)$, eqn [53] is consistent with the Feingold–Peres prediction for the variance of generic chaotic systems. Furthermore, recent estimates of moments by Rudnick and Soundararajan (2005) indicate that Conjecture 4 is not valid in the case of the modular surface.

## Quantum Eigenstates of Cat Maps

Cat maps are probably the simplest area-preserving maps on a compact surface that are highly chaotic. They are defined as linear automorphisms on the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$,

$$\Phi_A : \mathbb{T}^2 \to \mathbb{T}^2 \quad [55]$$

where a point $\xi\in\mathbb{R}^2(\mathrm{mod}\,\mathbb{Z}^2)$ is mapped to $A\xi(\mathrm{mod}\,\mathbb{Z}^2)$; $A$ is a fixed matrix in $\mathrm{GL}(2,\mathbb{Z})$ with eigenvalues off the unit circle (this guarantees hyperbolicity). We view the torus $\mathbb{T}^2$ as a symplectic manifold, the phase space of the dynamical system. Since $\mathbb{T}^2$ is compact, the Hilbert space of quantum states is an $N$-dimensional vector space $\mathcal{H}_N$, $N$ integer. The semiclassical limit, or limit of small wavelengths, corresponds here to $N\to\infty$.

It is convenient to identify $\mathcal{H}_N$ with $\mathrm{L}^2(\mathbb{Z}/N\mathbb{Z})$, with the inner product

$$\langle\psi_1,\psi_2\rangle = \frac{1}{N}\sum_{Q \bmod N}\psi_1(Q)\overline{\psi_2(Q)} \quad [56]$$

For any smooth function $f\in\mathrm{C}^\infty(\mathbb{T}^2)$, define a quantum observable

$$\mathrm{Op}_N(f) = \sum_{n\in\mathbb{Z}^2}\widehat{f}(n)T_N(n)$$

where $\widehat{f}(n)$ are the Fourier coefficients of $f$, and $T_N(n)$ are translation operators

$$T_N(n) = \mathrm{e}^{\pi\mathrm{i}n_1n_2/N}t_2^{n_2}t_1^{n_1} \quad [57]$$

$$[t_1\psi](Q) = \psi(Q+1)$$
$$[t_2\psi](Q) = \mathrm{e}^{2\pi\mathrm{i}Q/N}\psi(Q) \quad [58]$$

The operators $\mathrm{Op}_N(a)$ are the analogs of the pseudodifferential operators discussed in the section "Quantum unique ergodicity."

A quantization of $\Phi_A$ is a unitary operator $U_N(A)$ on $L^2(\mathbb{Z}/N\mathbb{Z})$ satisfying the equation

$$U_N(A)^{-1}\mathrm{Op}_N(f)U_N(A) = \mathrm{Op}_N(f \circ \Phi_A) \qquad [59]$$

for all $f \in C^\infty(\mathbb{T}^2)$. There are explicit formulas for $U_N(A)$ when $A$ is in the group

$$\Gamma = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}): ab \equiv cd \equiv 0 \bmod 2 \right\} \qquad [60]$$

These may be viewed as analogs of the Shale–Weil or metaplectic representation for SL(2). for example, the quantization of

$$A = \begin{pmatrix} 2 & 1 \\ 3 & 2 \end{pmatrix} \qquad [61]$$

yields

$$U_N(A)\psi(Q) = N^{-1/2} \sum_{Q' \bmod N} \exp\left[\frac{2\pi i}{N}(Q^2 - QQ' + Q'^2)\right]\psi(Q') \qquad [62]$$

In analogy with [1], we are interested in the statistical features of the eigenvalues and eigenfunctions of $U_N(A)$, that is, the solutions to

$$U_N(A)\varphi = \lambda\varphi, \quad \|\varphi\|_{L^2(\mathbb{Z}/N\mathbb{Z})} = 1 \qquad [63]$$

Unlike typical quantum-chaotic maps, the statistics of the $N$ eigenvalues

$$\lambda_{N1}, \lambda_{N2}, \ldots, \lambda_{NN} \in S^1 \qquad [64]$$

do not follow the distributions of unitary random matrices in the limit $N \to \infty$, but are rather singular (Keating 1991). In analogy with the Selberg trace formula for hyperbolic surfaces [18], there is an exact trace formula relating sums over eigenvalues of $U_N(A)$ with sums over fixed points of the classical map (Keating 1991).

As in the case of arithmetic surfaces, the eigenfunctions of cat maps appear to behave more generically. The analog of the Schnirelman–Zelditch–Colin de Verdière theorem states that, for any orthonormal basis of eigenfunctions $\{\varphi_{Nj}\}_{j=1}^N$ we have, for all $f \in C^\infty(\mathbb{T}^2)$,

$$\langle \mathrm{Op}(f)\varphi_{Nj}, \varphi_{Nj}\rangle \to \int_{\mathbb{T}^2} f(\xi)\mathrm{d}\xi \qquad [65]$$

as $N \to \infty$, for all $j$ in an index set $J_N$ of full density, that is, $\#J_N \sim N$. Kurlberg and Rudnick (see Rudnick (2001)) have characterized special bases of eigenfunctions $\{\varphi_{Nj}\}_{j=1}^N$ (termed Hecke eigenbases, in analogy with arithmetic surfaces) for which QUE holds, generalizing earlier work of Degli Esposti, Graffi, and Isola (1995). That is, [65] holds for all $j = 1, \ldots, N$. Rudnick and Kurlberg, and more recently Gurevich and Hadani, have established results on the rate of convergence analogous to [49]. These results are unconditional. Gurevich and Hadani use methods from algebraic geometry based on those developed by Deligne in his proof of the Weil conjectures (an analog of the Riemann hypothesis for finite fields).

In the case of quantum-cat maps, there are values of $N$ for which the number of coinciding eigenvalues can be large, a major difference to what is expected for the modular surface. Linear combinations of eigenstates with the same eigenvalue are as well eigenstates, and may lead to different quantum limits. Indeed, Faure, Nonnenmacher, and De Bièvre (see De Bièvre (to appear)) have shown that there are subsequences of values of $N$, so that, for all $f \in C^\infty(\mathbb{T}^2)$,

$$\langle \mathrm{Op}(f)\varphi_{Nj}, \varphi_{Nj}\rangle \to \frac{1}{2}\int_{\mathbb{T}^2} f(\xi)\mathrm{d}\xi + \frac{1}{2}f(0) \qquad [66]$$

that is, half of the mass of the quantum limit localizes on the hyperbolic fixed point of the map. This is the first, and to date the only, rigorous result concerning the existence of scarred eigenfunctions in systems with chaotic classical limit.

## Acknowledgment

The author is supported by an EPSRC Advanced Research Fellowship.

*See also:* Quantum Ergodicity and Mixing of Eigenfunctions; Random Matrix Theory in Physics.

## Further Reading

Aurich R and Steiner F (1993) Statistical properties of highly excited quantum eigenstates of a strongly chaotic system. *Physica D* 64(1–3): 185–214.

Berry MV and Keating JP (1999) The Riemann zeros and eigenvalue asymptotics. *SIAM Review* 41(2): 236–266.

Bogomolny EB (2006) Quantum and arithmetical chaos. In: Cartier PE, Julia B, Moussa P, and Vanhove P (eds.) *Frontiers in Number Theory, Physics and Geometry on Random Matrices, Zeta Functions, and Dynamical Systems*, Springer Lecture Notes. Les Houches.

Bogomolny EB, Georgeot B, Giannoni M-J, and Schmit C (1997) Arithmetical chaos. *Physics Reports* 291(5–6): 219–324.

De Bièvre S *Recent Results on Quantum Map Eigenstates*, Proceedings of QMATH9, Giens 2004 (to appear).

Hejhal DA and Rackner BN (1992) On the topography of Maass waveforms for PSL(2, Z). *Experiment. Math.* 1(4): 275–305.

Keating JP (1991) The cat maps: quantum mechanics and classical motion. *Nonlinearity* 4(2): 309–341.

Lindenstrauss E Rigidity of multi-parameter actions. *Israel Journal of Mathematics* (Furstenberg Special Volume) (to appear).

Marklof J (2006) Energy level statistics, lattice point problems and almost modular functions. In: Cartier PE, Julia B, Moussa P, and Vanhove P (eds.) *Frontiers in Number Theory, Physics and Geometry on Random Matrices, Zeta Functions, and Dynamical Systems*, Springer Lecture Notes. Les Houches.

Rudnick Z (2001) On quantum unique ergodicity for linear maps of the torus. In: *European Congress of Mathematics*, (Barcelona, 2000), Progr. Math., vol. 202, pp. 429–437. Basel: Birkhäuser.

Rudnick Z (2005) A central limit theorem for the spectrum of the modular group, Park city lectures. *Annales Henri Poincaré* 6: 863–883.

Sarnak P Arithmetic quantum chaos. The Schur lectures (1992) (Tel Aviv), Israel Math. Conf. Proc., 8, pp. 183–236. Bar-Ilan Univ., Ramat Gan, 1995.

Sarnak P (2003) Spectra of hyperbolic surfaces. *Bulletin of the American Mathematical Society (N.S.)* 40(4): 441–478.

# Asymptotic Structure and Conformal Infinity

**J Frauendiener**, Universität Tübingen, Tübingen, Germany



## Introduction

A major motivation for studying the asymptotic structure of spacetimes has been the need for a rigorous description of what should be understood by an "isolated system" in Einstein's theory of gravity. As an example, consider a gravitating system somewhere in our universe (e.g., a galaxy, a cluster of galaxies, a binary system, or a star) evolving according to its own gravitational interaction, and possibly reacting to gravitational radiation impinging on it from the outside. Thereby it will emit gravitational radiation. We are interested in describing these waves because they provide us with important information about the physics governing the system.

To adequately describe this situation, we need to idealize the real situation in an appropriate way, since it is hopeless to try to analyze the behavior of the system in its interaction with the rest of the universe. We are mainly interested in the behavior of the system, and not so much in other processes taking place at large distances from the system. Since we would like to ignore those regions, we need a way to isolate the system from their influence.

The notion of an isolated system allows us to select individual subsystems of the universe and describe their properties regardless of the rest of the universe so that we can assign to each subsystem such physical attributes as its energy–momentum, angular momentum, or its emitted radiation field. Without this notion, we would always have to take into account the interaction of the system with its environment in full detail.

In general relativity (GR) it turns out to be a rather difficult task to describe an isolated system and the reason is – as always in Einstein's theory – the fact that the metric acts both as the physical field and as the background. In other theories, like electrodynamics, the physical field, such as the Maxwell field, is very different from the background field, the flat metric of Minkowski space. The fact that the metric in GR plays a dual role makes it difficult to extract physical meaning from the metric because there is no nondynamical reference point.

Imagine a system alone in the universe. As we recede from the system we would expect its influence to decrease. So we expect that the spacetime which models this situation mathematically will resemble the flat Minkowski spacetime and it will approximate it even better the farther away we go. This implies that one needs to impose fall-off conditions for the curvature and that the manifold will be asymptotically flat in an appropriate sense. However, there is the problem that fall-off conditions necessarily imply the use of coordinates and it is awkward to decide which coordinates should be "good ones." Thus, it is not clear whether the notion of an asymptotically flat spacetime is an invariant concept.

What is needed, therefore, is an invariant definition of asymptotically flat spacetimes. The key observation in this context is that "infinity" is far away with respect to the spacetime metric. This means that geodesics heading away from the system should be able to "run forever," that is, be defined for arbitrary values of their affine parameter $s$. "Infinity" will be reached for $s \to \infty$. However, suppose we do not use the spacetime metric $g$ but a metric $\hat{g}$ which is scaled down with respect to $g$, that is, in such a way that $\hat{g} = \Omega^2 g$ for some function $\Omega$. Then it might be possible to arrange $\Omega$ in such a way that geodesics for the metric $\hat{g}$ cover the same events (strictly speaking, this holds only for null geodesics, but this is irrelevant for the present plausibility argument) as those for the metric $g$ yet that their affine parameter $\hat{s}$ (which is also scaled down with respect to $s$) approaches a finite value $\hat{s}_0$ for $s \to \infty$. Then we could attach a boundary to the spacetime manifold consisting of all the limit points corresponding to the events with $\hat{s} = \hat{s}_0$ on the $\hat{g}$-geodesics.

This boundary would have to be interpreted as "infinity" for the spacetime because it takes infinitely long for the $g$-geodesics to get there.

We arrived at this idea of attaching a boundary by considering the metric structure only "up to arbitrary scaling," that is, by looking at metrics which differ only by a factor. This is the conformal structure of the spacetime manifold in question. By considering the spacetime only from the point of view of its conformal structure we obtain a picture of the spacetime which is essentially finite but which leaves its causal properties unchanged, and hence in particular the properties of wave propagation. This is exactly what is needed for a rigorous treatment of radiation emitted by the system.

## Infinity for Minkowski Spacetime

The above discussion suggests that we should consider the spacetime metric only up to scale, that is, to focus on the conformal structure of the spacetime in question. Since we are interested in systems which approach Minkowski spacetime at large distances from the source, it is illuminating to study Minkowski spacetime as a preliminary example. So consider the manifold $\mathbb{M} = \mathbb{R}^4$ equipped with the flat metric

$$g = \mathrm{d}t^2 - \mathrm{d}r^2 - r^2 \mathrm{d}\sigma^2 \qquad [1]$$

where $r$ is the standard radial coordinate defined by $r^2 = x^2 + y^2 + z^2$ and

$$\mathrm{d}\sigma^2 = \mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2$$

is the standard metric on the unit sphere $S^2$. We now introduce retarded and advanced time coordinates, which are adapted to the null cone and hence to the conformal structure of $g$ by the definition

$$u = t - r, \qquad v = t + r$$

and obtain the metric in the form

$$g = \mathrm{d}u\,\mathrm{d}v - \tfrac{1}{4}(v - u)^2 \mathrm{d}\sigma^2$$

The coordinates $u$ and $v$ both take arbitrary real values but they are restricted by the relation $v - u = 2r \geq 0$. In order to see what happens "at infinity," we introduce the coordinates $U$ and $V$ by the relations

$$u = \tan\, U, \qquad v = \tan\, V$$

Then $U$ and $V$ both take values in the open interval $(-\pi/2, \pi/2)$ with $V \geq U$ and the metric is transformed to

$$g = \frac{1}{4\cos^2 U \cos^2 V}\left[4\mathrm{d}U\,\mathrm{d}V - \sin^2(V - U)\,\mathrm{d}\sigma^2\right] \qquad [2]$$

Clearly, the metric is undefined at events with $\cos U = 0$ or $\cos V = 0$. These would correspond to events with $u = \pm\infty$ or $v = \pm\infty$ which do not lie in $\mathbb{M}$. However, by defining the function

$$\Omega = 2\cos U \cos V$$

we find that the metric $\hat{g} = \Omega^2 g$ with

$$\hat{g} = 4\mathrm{d}U\,\mathrm{d}V - \sin^2(V - U)\,\mathrm{d}\sigma^2 \qquad [3]$$

is conformally equivalent to $g$ and is regular for all values of $U$ and $V$ (keeping $V \geq U$). In fact, by defining the coordinates

$$T = V + U, \qquad R = V - U$$

this metric takes the form

$$\hat{g} = \mathrm{d}T^2 - \mathrm{d}R^2 - \sin^2 R\,\mathrm{d}\sigma^2 \qquad [4]$$

the metric of the static Einstein universe $\mathbb{E}$. Thus, we may regard the Minkowski spacetime as the part of the Einstein cylinder defined by restricting the coordinates $T$ and $R$ to the region $|T| + R < \pi$ as illustrated in Figure 1. Although $\mathbb{M}$ can be considered as being diffeomorphic to the shaded part in Figure 1, these two manifolds are not isometric. This is obvious from considering the properties of the events lying on

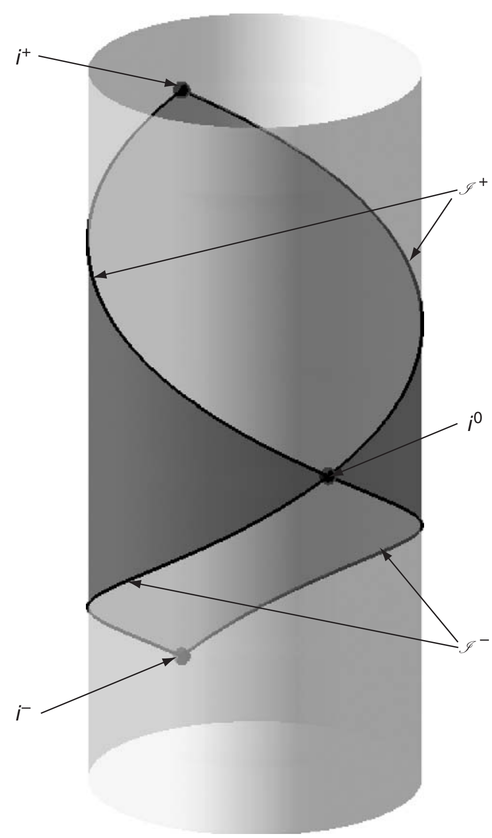



**Figure 1** The embedding of Minkowski spacetime into the Einstein cylinder.

the boundary $\partial\mathbb{M}$ of $\mathbb{M}$ in $\mathbb{E}$. Fix a point $P$ inside $\mathbb{M}$ and follow a null geodesic with respect to the metric $\hat{g}$ from $P$ toward the future. It will intersect $\partial\mathbb{M}$ after a finite amount of its affine parameter has elapsed. When we follow a null geodesic with respect to $g$ from $P$ in the same direction, we find that it does not reach $\partial\mathbb{M}$ for any value of its affine parameter. Thus, the boundary is at infinity for the metric $g$ but at a finite location with respect to the metric $\hat{g}$. When we consider all possible kinds of geodesics for the metric $g$ we find that $\partial\mathbb{M}$ consists of five qualitatively different pieces. The future pointing timelike geodesics all approach the point $i^+$ given by $(T, R) = (\pi, 0)$, while the past-pointing geodesics approach $i^-$ with coordinates $(-\pi, 0)$. All spacelike geodesics come arbitrarily close to a point $i^0$ with coordinates $(0, \pi)$ (located on the front of the cylinder in Figure 1). Null geodesics, however, are different. For any point $(T, \pi - |T|)$ with $T \neq 0, \pm\pi$ on $\partial\mathbb{M}$ there are $g$-null-geodesics which come arbitrarily close.

In this sense, we may regard $\partial\mathbb{M}$ as consisting of limit points obtained by tracing-geodesics for infinite values of their affine parameters. According to the causal character of the geodesics the set of their respective limit points is called future/past timelike infinity $i^\pm$, spacelike infinity $i^0$ or future/past null-infinity, denoted by $\mathscr{I}^\pm$. These two parts of null-infinity are three-dimensional regular submanifolds of the embedding manifold $\mathbb{E}$, while the points $i^\pm, i^0$ are regular points in $\mathbb{E}$ in the sense that the metric $\hat{g}$ is regular there. This is not automatic, considering the fact that infinitely many geodesics converge to a single point. However, the flatness of Minkowski spacetime guarantees that the geodesics approach at just the appropriate rate for the limit points to be regular.

This example shows that the structure of the boundary is determined entirely by the metric $g$ of Minkowski spacetime. If we had chosen a different function $\Omega' = \omega\Omega$ with $\omega > 0$ then we would not have obtained the Einstein cylinder but some different Lorentzian manifold $(\mathcal{M}', g')$. Yet, the boundary of $\mathbb{M}$ in $\mathcal{M}'$ would have had the same properties.

## Asymptotically Flat Spacetimes

The physical idea of an isolated system is captured mathematically by an asymptotically flat spacetime. Since such a spacetime $\mathcal{M}$ is expected to approach Minkowski spacetime asymptotically, the asymptotic structure of $\mathcal{M}$ is also expected to be similar to that of $\mathbb{M}$. This expectation is expressed in

**Definition 1** A spacetime $(\mathcal{M}, g_{ab})$ is called "asymptotically simple" if there exists a manifold-with-boundary $\widehat{\mathcal{M}}$ with metric $\hat{g}_{ab}$ and scalar field $\Omega$ on $\widehat{\mathcal{M}}$ and boundary $\mathcal{I} = \partial\mathcal{M}$ such that the following conditions hold:

1. $\mathcal{M}$ is the interior of $\widehat{\mathcal{M}}$: $\mathcal{M} = \text{int}\ \widehat{\mathcal{M}}$;
2. $\hat{g}_{ab} = \Omega^2 g_{ab}$ on $\mathcal{M}$;
3. $\Omega$ and $\hat{g}_{ab}$ are smooth on all of $\bar{\mathcal{M}}$;
4. $\Omega > 0$ on $\mathcal{M}$; $\Omega = 0, \nabla_a\Omega \neq 0$ on $\mathscr{I}$; and
5. each null geodesic acquires both future and past endpoints on $\mathscr{I}$.

This definition formalizes the construction which was explicitly performed above, by which one attaches a regular (nonempty) boundary to a spacetime after suitably rescaling its metric. Asymptotically simple spacetimes are exactly those for which this process of conformal compactification is possible. The purpose of condition 5 is to exclude pathological cases. There are spacetimes which do not satisfy this condition (e.g., the Schwarzschild spacetime, where some of the null geodesics enter the event horizon and cannot escape to infinity). Yet, one would like to include them as being asymptotically simple in a sense, because they clearly describe isolated systems. For these cases, there exists the notion of weakly asymptotically simple spacetimes.

In order to arrive at asymptotically flat spacetimes, one needs to make certain assumptions about the behavior of the curvature near the boundary, thus:

**Definition 2** An asymptotically simple spacetime is called "asymptotically flat" if its Ricci tensor Ric[$g$] vanishes in a neighborhood of $\mathscr{I}$.

Note that this definition imposes a rather strong restriction on the Ricci curvature; less restrictive assumptions are possible. This condition applies only near $\mathscr{I}$. Thus, it is possible to consider spacetimes which contain matter fields as long as these fields do not extend to infinity.

Other asymptotically simple spacetimes which are not asymptotically flat are the de Sitter and anti-de Sitter spacetimes which are solutions of the Einstein equations with nonvanishing cosmological constant $\lambda$. It is a simple consequence of the definition that the boundary $\mathscr{I}$ is a regular three-dimensional hypersurface of the embedding spacetime $\widehat{\mathcal{M}}$ which is timelike, spacelike, or null depending on the sign of $\lambda$. In particular, for the Minkowski spacetime $(\lambda = 0)$ the boundary is necessarily a null hypersurface, as noted above.

The requirement that the vacuum Einstein equations hold near $\mathscr{I}$ has several important

consequences. First, $\mathscr{I}$ is a null hypersurface with the special property of being shear-free. This means that any cross section of a bundle of its null generators does not suffer any distortions when moved along the generators. Only expansion or contraction can occur. The global structure of $\mathscr{I}$ is the same as the one from the example above. Null infinity consists of two connected components, $\mathscr{I}^{\pm}$, each of which is diffeomorphic to $S^2 \times \mathbb{R}$. Thus, topologically, $\mathscr{I}^{\pm}$ are cylinders. The cone-like appearance as seen in Figure 1 is artificial. It depends on the particular conformal factor $\Omega$ chosen for the conformal compactification. Furthermore, it is only in very exceptional cases that the metric $\hat{g}$ is regular at $i^0$ or $i^{\pm}$.

The most important consequence, however, concerns the conformal Weyl tensor $C^a{}_{bcd}$. This is the part of the full Riemann curvature tensor $R^a{}_{bcd}$ which is trace-free. It is invariant under conformal rescalings of the metric. Thus, on $\mathcal{M}$, $C^a{}_{bcd} = \hat{C}^a{}_{bcd}$. When the vanishing of the Ricci tensor near $\mathscr{I}$ is assumed then it turns out that the Weyl tensor necessarily vanishes on $\mathscr{I}$. This is the ultimate justification for calling such manifolds asymptotically flat because the entire curvature vanishes on $\mathscr{I}$.

## Some Consequences

There are several consequences of the existence of the conformal boundary $\mathscr{I}$. They all can be traced back to the fact that this boundary can be used to separate the geometric fields into a universal background field and dynamical fields which propagate on it. The background is given by the boundary points attached to an asymptotically flat spacetime which always form a three-dimensional null hypersurface $\mathscr{I}$ with two connected components (in the sequel, we restrict our attention to $\mathscr{I}^+$ only; $\mathscr{I}^-$ is treated similarly), each with the topology of a cylinder. And in each case, $\mathscr{I}$ is shear-free.

### The BMS Group

Since the structure of null-infinity is universal over all asymptotically flat spacetimes, it is obvious that its symmetry group should also possess a universal meaning. This group, the so-called Bondi–Metzner–Sachs (BMS) group is in many respects similar to the Poincaré group, the symmetry group of $\mathbb{M}$. It is the semidirect product of the Lorentz group with an abelian group which, however, is not the four-dimensional translation group but an infinite-dimensional group of supertranslations. This group is a normal subgroup, so the factor group is isomorphic to the Lorentz group.

In physical terms, the supertranslations arise because there are infinitely many directions from which observers at infinity (whose world lines coincide with the null generators of $\mathscr{I}$ in a certain limit) can observe the system and because each observer is free to choose its own origin of proper time $u$. The observers surrounding the system are not synchronized, because under the assumptions made there is no natural way to fix a unique common origin. Hence, a supertranslation is a shift of the parameter along each null generator of $\mathscr{I}^+$ corresponding to a change of origin for each individual observer. It can be given as a map $S^2 \to \mathbb{R}$. A choice of origin on each null generator of $\mathscr{I}^+$ is referred to as a "cut" of $\mathscr{I}^+$. It is a two-dimensional surface of spherical topology which intersects each null generator exactly once. It is an open question whether one can always synchronize the observers by imposing canonical conditions at $i^0$ or $i^{\pm}$, thereby reducing the BMS group to the smaller Poincaré group.

The supertranslations contain a unique four-dimensional normal subgroup. In $\mathbb{M}$ these special supertranslations are the ones which are induced by the translations of Minkowski spacetime in the following way. Take the future light cone of some event $P$ and follow it out to $\mathscr{I}^+$, where its intersection defines an origin for each observer located there. Now consider the light cone of another event $Q$ obtained from $P$ by a translation in a spatial direction. Then the light emitted from $Q$ will arrive at $\mathscr{I}^+$ earlier than that from $P$ for observers in the direction of the translation, while it will be delayed for observers in the opposite direction. This change in arrival time defines a specific supertranslation. Similarly, for a translation in a temporal direction, the light from $Q$ will arrive later than that from $P$ for all observers. Thus, every translation in $\mathbb{M}$ defines a particular supertranslation on $\mathscr{I}^+$. These can be characterized in a different way, which is intrinsic to $\mathscr{I}^+$ and which can be used in the general case even though there will be no Killing vectors present in a general asymptotically flat spacetime. In an appropriate coordinate system, the asymptotic translations are given as linear combinations of the first four spherical harmonics $Y_{00}, Y_{10}, Y_{1\pm1}$. The space of asymptotic translations $\mathbb{T}$ is in a natural way isometric to $\mathbb{M}$.

### The Peeling Property

Now consider the Weyl tensor $C^a{}_{bcd}$ on $\widehat{\mathcal{M}}$. Since it vanishes on $\mathscr{I}$ where $\Omega = 0$ we may form the quotient

$$K^a{}_{bcd} = \Omega^{-1} C^a{}_{bcd}$$

which can be shown to be smooth on $\mathscr{I}^+$. The physical interpretation of this tensor field is based on the following properties. In source-free regions the field satisfies the spin-2 zero-rest-mass equation

$$\widehat{\nabla}_a K^a{}_{bcd} = 0$$

which is very similar to the Maxwell equations for the electromagnetic (spin-1) Faraday tensor. Thus, $K^a{}_{bcd}$ is interpreted as the gravitational field, which describes the gravitational waves contained inside the system. The zero-rest-mass equation for $K^a{}_{bcd}$ and the fact that the field is smooth on $\mathscr{I}$ implies that the Weyl tensor satisfies the "peeling" property. This is a characteristic conspiracy between the fall-off behavior of certain components of the Weyl tensor along outgoing $g$-null-geodesics approaching $\mathscr{I}^+$ in $\mathcal{M}$ with respect to an affine parameter $s$ for $s \to \infty$ and their algebraic type. Symbolically, the Weyl tensor has the following behavior as $s \to \infty$ along the null geodesic:

$$C = \frac{[4]}{s} + \frac{[31]}{s^2} + \frac{[211]}{s^3} + \frac{[1111]}{s^4} + O(s^{-5}) \qquad [5]$$

where the numerator of each component indicates its Petrov type. The repeated principal null direction (PND) in the first three components and one of the PNDs in the fourth component are aligned with the tangent vector of the geodesic. This implies that the farthest reaching component of the Weyl tensor, which is $O(1/s)$, has the Petrov type of a radiation field. It is customary to combine the components which are $O(1/s^i)$ into one complex function and denote it by $\psi_{5-i}$. When expressed in terms of the field $K^a{}_{bcd}$ on $\widehat{\mathcal{M}}$, this fall-off behavior implies that of all components of $K^a{}_{bcd}$ only $\psi_4$ does not necessarily vanish on $\mathscr{I}^+$.

In special cases like the Minkowski, Schwarzschild, Kerr, and more generally in all asymptotically flat stationary spacetimes, even $\psi_4$ vanishes on $\mathscr{I}^+$. For these reasons, $\psi_4$ is called the radiation field of the system, that is, that part of the gravitational field which can be registered by the observers at infinity. It describes the outgoing radiation which is being emitted by the system during its evolution.

### The Bondi–Sachs Mass-Loss Formula

Gravitational waves carry away energy from the system. This is a consequence of the Bondi–Sachs mass-loss formula. The Bondi–Sachs energy–momentum is related to a weighted integral over a cut $\mathcal{C}$,

$$P_{\mathcal{C}}[W] = -\frac{1}{4\pi G}\int_{\mathcal{C}} W[\psi_2 + \sigma\dot{\bar{\sigma}}]\,\mathrm{d}^2S \qquad [6]$$

The quantity in brackets, the mass aspect, is a combination of the scalar $\psi_2$ which in a sense measures the strength of the Coulomb-like part of the gravitational field on $\mathscr{I}^+$ and the complex quantity $\sigma$. In a so-called Bondi coordinate system, this quantity is related to the radiation field $\psi_4$ by the relation

$$\psi_4 = -\ddot{\bar{\sigma}}$$

the dot indicating differentiation with respect to the affine parameter along the null generators. Thus, $\sigma$ is essentially the second time integral of the radiation field. The mass aspect is integrated against a function $W$ which is an asymptotic translation, that is, a linear combination of the first four spherical harmonics. Thus, one can view the expression [6] as defining a linear map $\mathbb{T} \to \mathbb{R}$. Since $\mathbb{T}$ and $\mathbb{M}$ are isometric this defines a covector $P_a$ on $\mathbb{M}$, which can always be shown to be timelike, $P_aP^a \geq 0$. This positivity property together with the fact that in the special cases of Schwarzschild and Kerr spacetimes the integral yields the mass parameters when evaluated for a time translation ($W = 1$) motivates the interpretation of $P_{\mathcal{C}}$ as the energy–momentum 4-vector of the spacetime at the instant defined by the cut $\mathcal{C}$. In particular, for $W = 1$ the integral gives the time component of $P_{\mathcal{C}}$, the Bondi–Sachs energy $E$.

The interpretation of [6] as energy–momentum is strengthened by the fact that $P_{\mathcal{C}}$ arises as dual to the translations which is familiar from Lagrangian field theories where energy and momentum appear as generators for time and space translations. In fact, one can set up a Hamiltonian framework where the role of the Bondi–Sachs energy–momentum as generator of asymptotic translations is made explicit.

This point of view suggests that one should also be able to define a notion of angular momentum for asymptotically flat spacetimes because angular momentum arises as the generator of rotations, which can also be defined asymptotically. However, while there is a unique notion of translation on $\mathscr{I}^+$, this is not the case for rotations (and boosts). The reason is hidden in the structure of the BMS group where the Lorentz group appears naturally as a factor group but not as a unique subgroup. In physical terms, the angular momentum depends on an origin but there is no natural way to choose an origin on $\mathscr{I}^+$. This ambiguity in the choice of origin leads to several nonequivalent expressions for angular momentum in the literature.

Consider now two cuts $\mathcal{C}$ and $\mathcal{C}'$, with $\mathcal{C}'$ later than $\mathcal{C}$. Then we may compute the difference $\Delta E = E - E'$ of the Bondi–Sachs energies with respect to the two

cuts. It turns out that this difference can be expressed as an integral over the (three-dimensional) piece $\Sigma$ of $\mathscr{I}^+$ which is bounded by the two cuts (i.e., $\partial\Sigma = \mathcal{C}' - \mathcal{C}$):

$$E' - E = -\frac{1}{4\pi G}\int_\Sigma \dot{\sigma}\dot{\bar{\sigma}}\, \mathrm{d}^3 V \qquad [7]$$

This result means that the Bondi–Sachs energy of the system decreases, since $E' < E$ and the rate of decrease is given by the (positive-definite) amount of gravitational radiation which leaves the system during the period defined by the two cuts.

It is necessary to point out that in this article the structure of null infinity has been postulated based on physical reasonings. The Einstein equations have been used only in a very weak sense, namely only in a neighborhood of $\mathscr{I}$. It is an entirely different question whether the field equations are compatible with this postulated structure. To answer it, one needs to show that there are global solutions of the Einstein equations which exhibit the postulated behavior in the asymptotic region. This question has been settled recently in the affirmative: there are many global spacetimes which are asymptotically flat in the sense described here.

This article discussed has the notion of null infinity, that is, of spacetimes which are asymptotically flat in lightlike directions. Spacetimes which are asymptotically flat in spacelike directions have not been covered. The latter is a notion which has been developed largely independently of null infinity since it is essentially a property of an initial data set and not of the entire four-dimensional spacetime. Ultimately, these two notions should coincide, in the sense that if one has an initial data set which is asymptotically flat in spatial directions in an appropriate sense then its Cauchy development will be an asymptotically flat spacetime. However, as of yet, it is not clear what the appropriate conditions should be because the structure of the gravitational field in the neighborhood of spacelike infinity $i^0$ is not sufficiently well understood so far.

*See also:* Black Hole Mechanics; Boundaries for Spacetimes; Canonical General Relativity; Einstein Equations: Exact Solutions; Einstein Equations: Initial Value Formulation; General Relativity: Overview; Gravitational Waves; Quantum Entropy; Spacetime Topology, Causal Structure and Singularities; Stability of Minkowski Space; Stationary Black Holes.

## Further Reading

Ashtekar A (1987) *Asymptotic Quantization.* Naples: Bibliopolis.

Bondi H, van der Burg MGJ, and Metzner AWK (1962) Gravitational waves in general relativity VII. Waves from axi-symmetric isolated systems. *Proceedings of the Royal Society of London, Series A* 269: 21–52.

Frauendiener J (2004) Conformal infinity. *Living Reviews in Relativity*, vol. 3. http://relativity.livingreviews.org/Articles/lrr-2004-1/index.html.

Friedrich H (1992) Asymptotic structure of space-time. In: Janis AI and Porter JR (eds.) *Recent Advances in General Relativity*. Boston: Birkhäuser.

Friedrich H (1998a) Einstein's equation and conformal structure. In: Huggett SA, Mason LJ, Tod KP, Tsou SS, and Woodhouse NMJ (eds.) *The Geometric Universe: Science, Geometry and the Work of Roger Penrose*. Oxford: Oxford University Press.

Friedrich H (1998b) Gravitational fields near space-like and null infinity. *Journal of Geometry and Physics* 24: 83–163.

Geroch R (1977) Asymptotic structure of space-time. In: Esposito FP and Witten L (eds.) *Asymptotic Structure of Space-Time*. New York: Plenum.

Hawking S and Ellis GFR (1973) *The Large Scale Structure of Space-Time*. Cambridge: Cambridge University Press.

Penrose R (1965) Zero rest-mass fields including gravitation: asymptotic behaviour. *Proceedings of the Royal Society of London, Series A* 284: 159–203.

Penrose R (1968) Structure of space-time. In: DeWitt CM and Wheeler JA (eds.) *Battelle Rencontres*. New York: W. A. Benjamin.

Penrose R and Rindler W (1984, 1986) *Spinors and Space-Time*, Cambridge: Cambridge University Press.

Sachs RK (1962) Gravitational waves in general relativity VIII. Waves in asymptotically flat space-time. *Proceedings of the Royal Society of London, Series A* 270: 103–127.

# Averaging Methods

**A I Neishtadt**, Russian Academy of Sciences, Moscow, Russia



## Introduction

Averaging methods are the methods of perturbation theory that are based on the averaging principle and the idea of dividing the dynamics into slow drift and fast oscillations. The most common field of applications of averaging methods is the analysis of the behavior of dynamical systems that differ from integrable systems by small perturbations.

## Averaging Principle

Equations of motion of a system that differ from an integrable system by small perturbations often can be written in the form

$$\dot{I} = \varepsilon g(I, \varphi, \varepsilon), \quad \dot{\varphi} = \omega(I) + \varepsilon f(I, \varphi, \varepsilon)$$
$$I = (I_1, \ldots, I_n) \in \mathbb{R}^n \qquad [1]$$
$$\varphi = (\varphi_1, \ldots, \varphi_m) \in \mathbb{T}^m \bmod 2\pi, 0 < \varepsilon \ll 1$$

The small parameter $\varepsilon$ characterizes the amplitude of the perturbation. For $\varepsilon = 0$ one gets the unperturbed system. The equation $I = \text{const.}$ singles out an invariant $m$-dimensional torus of the unperturbed system. The motion on this torus is quasiperiodic with frequency vector $\omega(I)$; components of vector $I$ are called "slow variables" whereas components of vector $\varphi$ are called "fast variables" or "phases." The right-hand sides of system [1] are $2\pi$-periodic with respect to all $\varphi_j$. It is assumed that they are smooth enough functions of all arguments. It is also assumed that components of the frequency vector are not linearly dependent over the ring of integer numbers identically with respect to $I$. System [1] is called a "system with rotating phases."

In applications, one is often interested mainly in the behavior of slow variables. The "averaging principle" (or method) consists in replacing the system of perturbed equations [1] by the "averaged system"

$$\dot{J} = \varepsilon G(J), \quad G(J) = (2\pi)^{-m} \oint_{\mathbb{T}^m} g(J, \varphi, 0)\, d\varphi \qquad [2]$$

for the purpose of providing an approximate description of the evolution of the slow variables over time intervals of order $1/\varepsilon$ or longer. Here, $d\varphi = d\varphi_1 \cdots d\varphi_m$. System [2] contains only slow variables and, therefore, is much simpler for investigation than system [1]. When passing from system [1] to system [2], one ignores the terms $g(I, \varphi, 0) - G(I)$ on the right-hand side of [1]. The averaging principle is based on the idea that these terms oscillate and lead only to small oscillations which are superimposed on the drift described by the averaged system. To justify the averaging principle, one should establish a relation between the behavior of the solutions of systems [1] and [2]. This problem is still far from being completely solved.

Another version of the averaging principles is used in the case when frequencies are approximately in resonance. This means that one or several relations of the form $(k, \omega) = 0$ approximately are valid with irreducible integer coefficient vectors $k \neq 0$; here, $(k, \omega)$ is the standard scalar product in $\mathbb{R}^m$. Let $\Gamma$ be a sublattice of the integer lattice $\mathbb{Z}^m$ generated by these vectors. Let $r = \operatorname{rank} \Gamma$ and $k^{(1)}, k^{(2)}, \ldots, k^{(m)}$ be a basis in $\mathbb{Z}^m$, the first $r$ vectors of which belong to $\Gamma$. Instead of $\varphi$, one can introduce new variables:

$$\vartheta = (\vartheta_1, \ldots, \vartheta_r) \in \mathbb{T}^r \bmod 2\pi$$
$$\chi = (\chi_1, \ldots, \chi_{m-r}) \in \mathbb{T}^{m-r} \bmod 2\pi$$
$$\vartheta_i = (k^{(i)}, \varphi), \qquad \chi_j = (k^{(r+j)}, \varphi)$$

Let $R$ be an $r \times m$ matrix whose rows are vectors $k^{(i)}, 1 \leq i \leq r$. For an approximate description of the behavior of variables $I, \vartheta$, the averaging principle prescribes replacing system [1] by the system

$$\dot{J} = \varepsilon G_\Gamma(J, \gamma), \quad \dot{\gamma} = R\omega(J) + \varepsilon R F_\Gamma(J, \gamma)$$
$$G_\Gamma(J, \vartheta) = (2\pi)^{-(m-r)} \oint_{\mathbb{T}^{m-r}} g(J, \varphi, 0)\, d\chi \qquad [3]$$
$$F_\Gamma(J, \vartheta) = (2\pi)^{-(m-r)} \oint_{\mathbb{T}^{m-r}} f(J, \varphi, 0)\, d\chi$$

(one should express $g, f$ through $\vartheta, \chi$ and then integrate over $\chi, d\chi = d\chi_1 \cdots d\chi_{m-r}$). System [3] is called "partially averaged system" for resonances in $\Gamma$. Functions $G_\Gamma, F_\Gamma$ can be obtained from Fourier series expansions of functions $g$, $f$ for $\varepsilon = 0$ by throwing away harmonics $\exp(i(k, \varphi)), k \notin \Gamma$ (nonresonant harmonics). Passing from system [1] to system [3] is based on the idea that the ignored nonresonant harmonics oscillate fast and do not affect essentially the evolution of the slow variables.

Now let system [1] be a Hamiltonian system close to an integrable one. The Hamiltonian function has the form

$$H = H_0(p) + \varepsilon H_1(p, \varphi, y, x, \varepsilon)$$

where $\varphi, x$ are coordinates and $p, y$ are conjugated to them. The equations of motion have the same form as [1], with $I = (p, y, x)$:

$$\dot{p} = -\varepsilon \frac{\partial H_1}{\partial \varphi}, \quad \dot{y} = -\varepsilon \frac{\partial H_1}{\partial x}$$
$$\dot{x} = \varepsilon \frac{\partial H_1}{\partial y}, \qquad \dot{\varphi} = \frac{\partial H_0}{\partial I} + \varepsilon \frac{\partial H_1}{\partial I} \qquad [4]$$

The averaging principle in the case when there are no resonant relations leads to the system

$$\dot{p} = 0, \quad \dot{y} = -\varepsilon \frac{\partial \mathcal{H}_1}{\partial x}, \quad \dot{x} = \varepsilon \frac{\partial \mathcal{H}_1}{\partial y}$$
$$\mathcal{H}_1 = (2\pi)^{-m} \oint_{\mathbb{T}^m} H_1(p, \varphi, y, x, 0)\, d\varphi \qquad [5]$$

Therefore, in this case there is no drift in $p$, and the behavior of $y, x$ is described by the Hamiltonian system, which contains $p$ as a parameter. Equations of motion of planets around the Sun can be reduced to the form [4]. The issue of the absence of the evolution of momenta $p$ is known in this problem as

the Lagrange–Laplace theorem, about the absence of the evolution of semimajor axes of planetary orbits.

## Elimination of Fast Variables, Decoupling of Slow and Fast Motions

The basic role in the averaging method is played by the idea that the exact system can be in the principal approximation transformed into the averaged system by means of a transformation of variables close to the identical one. The extension of this idea is the idea that similar transformation of variables allows one to eliminate, up to an arbitrary degree of accuracy, the fast phases from the right-hand sides of the equations of perturbed motion and in this way decouple the slow motion from the fast one. For system [1], provided there are no resonant relations between frequencies, the elimination of fast variables is performed as follows. The desirable transformation of variables $(I, \varphi) \mapsto (J, \psi)$ is sought as a formal series

$$\begin{aligned} I &= J + \varepsilon u_1(J, \psi) + \varepsilon^2 u_2(J, \psi) + \cdots \\ \varphi &= \psi + \varepsilon v_1(J, \psi) + \varepsilon^2 v_2(J, \psi) + \cdots \end{aligned} \qquad [6]$$

where functions $u_j, v_j$ are $2\pi$-periodic in $\psi$. The transformation [6] should be chosen in such a way that in the new variables the right-hand sides of equations of motion do not contain fast variables, that is, the equations of motion should have the form

$$\begin{aligned} \dot{J} &= \varepsilon G_0(J) + \varepsilon^2 G_1(J) + \cdots \\ \dot{\psi} &= \omega(J) + \varepsilon F_0(J) + \varepsilon^2 F_1(J) + \cdots \end{aligned} \qquad [7]$$

Substituting [6] into [7], taking into account [1], and equating the terms of the same order in $\varepsilon$, we obtain the following set of relations:

$$\begin{aligned} G_0(J) &= g(J, \psi, 0) - \frac{\partial u_1}{\partial \psi}\omega \\ F_0(J) &= f(J, \psi, 0) + \frac{\partial \omega}{\partial J}u_1 - \frac{\partial v_1}{\partial \psi}\omega \\ G_i(J) &= X_i(J, \psi) - \frac{\partial u_{i+1}}{\partial \psi}\omega \\ F_i(J) &= Y_i(J, \psi) + \frac{\partial \omega}{\partial J}u_{i+1} - \frac{\partial v_{i+1}}{\partial \psi}\omega, \quad i \geq 1 \end{aligned} \qquad [8]$$

The functions $X_i, Y_i$ are uniquely determined by the terms $u_1, v_1, \ldots, u_i, v_i$ in expansion [6]. The first equation in [8] implies that

$$\begin{aligned} G_0(J) &= g_0(J) = G(J) \\ u_1(J, \psi) &= \sum_{k \neq 0} \frac{g_k}{\mathrm{i}(k,\ \omega)} \exp(\mathrm{i}(k, \psi)) + u_1^0(J) \end{aligned} \qquad [9]$$

where $g_k, k \in \mathbb{Z}^m$, are Fourier coefficients of function $g$ at $\varepsilon = 0$, and $u_1^0$ is an arbitrary function of $J$. It is assumed that the denominators in [9] do not vanish, and that the series in [9] converges and determines a smooth function. In the same way, from the other equations in [8] one can sequentially determine $F_0, v_1, \ldots, G_i, u_{i+1}, F_i, v_{i+1}, i \geq 1$.

On truncating the series in [6] and [7] at the terms of order $\varepsilon^l$, we obtain a truncated system of the $l$th approximation. The equation for $J$ is decoupled from the other equations and can be solved separately. Then the behavior of $\psi$ is determined by means of quadrature. The behavior of original variable $I$ in this approximation is a slow drift (described by the equation for $J$), on which small oscillations (described by transformation of variables) are superimposed. The behavior of $\varphi$ can be represented as a rotation with slowly varying frequency, on which oscillations are also superimposed. For $l = 1$, the truncated system coincides with the averaged system [2].

If the sublattice $\Gamma \subset \mathbb{Z}^m$ specifying possible resonant relations is given, then in an analogous manner one can construct a formal transformation of variables $(I, \varphi) \mapsto (J, \psi)$ such that, in the new variables, the fast phase $\psi$ will appear on the right-hand sides of the differential equations for the new variables only in combinations $(k, \psi)$, with $k \in \Gamma$ (see, e.g., Arnol'd *et al.* (1988)). Again, on truncating the series on the right-hand sides of the differential equations for the new variables at the terms of order $\varepsilon^l$, we obtain a truncated system of the $l$th approximation. At $l = 1$, this truncated system coincides with the partially averaged system [3] (for some special choice of arbitrary functions that are contained in the formulas for transformation of variables). If the original system is a Hamiltonian system of the form [4], then the transformation of variables eliminating the fast phases from the right-hand sides of the differential equations can be chosen to be symplectic. The corresponding procedures are called "Lindstedt method" and "Newcomb method" (nonresonant case for $n = m$), "Delaunay method" (resonant case for $n = m$), and "von Zeipel method" (resonant case for $n \geq m$) (see Poincaré (1957) and Arnol'd *et al.* (1988)).

The calculation of high-order terms in the procedures of elimination of fast variables is rather cumbersome. There are versions of these procedures which are convenient for symbolic processors (especially for Hamiltonian systems, e.g., the Deprit–Hori method; Giacaglia 1972).

The averaging method consists in using the averaged system for the description of motion in the first approximation and the truncated systems

obtained by means of the procedures of elimination of fast variables in the higher approximations, together with the corresponding transformations of variables.

## Justification of the Averaging Method

To justify the averaging method, one should establish conditions under which the deviation of the slow variables along the solutions of the exact system from the solutions of the averaged system with appropriate initial data on time intervals of order $1/\varepsilon$ or longer tends to 0 as $\varepsilon \to 0$. It is desirable to have estimates from the above for these deviations. The estimates of deviations of the solutions of the exact system from the solutions of the truncated systems obtained by means of the procedure of elimination of fast phases are important as well. It can happen that there are "bad" initial data for which the slow component of the solution of the exact system deviates from the solution of the averaged system by a value of order 1 over time of order $1/\varepsilon$. In this case, one should have estimates from above for the measure of the set of such "bad" initial data; on the complementary set of initial data, one should have estimates from above for the deviation of slow variables along the solutions of the exact system from the solution of the averaged system. These problems are currently far from being completely solved. Some general results are described in the following.

Let functions $\omega, f, g$ on the right-hand side of system [1] be defined and bounded together with a sufficient number of derivatives in the domain $D\{I\} \times \mathbb{T}^m\{\varphi\} \times [0, \varepsilon_0]$. Let $J(t)$ be the solution of the averaged system [2] with initial condition $I_0 \in D$. Let $(I(t), \varphi(t))$ be the solution of the exact system [1] with initial conditions $(I_0, \varphi_0)$. So, $I(0) = J(0)$. It is assumed that the solution $J(t)$ is defined and stays at a positive distance from the boundary of $D$ on the time interval $0 \le t \le K/\varepsilon, K = \text{const} > 0$.

If system [1] is a one-frequency system ($m = 1$), and the frequency $\omega$ does not vanish in $D$, then for $0 \le t \le K/\varepsilon$ the solution $(I(t), \varphi(t))$ is well defined, and $|I(t) - J(t)| < C\varepsilon, C = \text{const.} > 0$. For $\omega = 1$, this assertion was proved by P Fatou (1928) and, by a different method, by L I Mandel'shtam and L D Papaleksi (1934). This was historically the first result on the justification of the averaging method (Mintropol'skii 1971). There is a proof based on the elimination of fast variables (see, e.g., Arnol'd (1983)). For a one-frequency system, higher approximations of the procedure of elimination of fast variables allow the description of the dynamics with an accuracy of the order of any power in $\varepsilon$ on time intervals of order $1/\varepsilon$ (Bogolyubov and Mitropol'skii 1961).

If system [1] is a multifrequency system ($m \ge 2$), but the vector of frequencies is constant and nonresonant, then for any $\rho > 0$ and small enough $\varepsilon < \varepsilon_0(\rho)$ it holds that $|I(t) - J(t)| < \rho$ for $0 \le t \le K/\varepsilon$ (Bogolyubov 1945, Bogolyubov and Mitropol'skii 1961). If, in addition, the frequencies satisfy the Diophantine condition $|(k, \omega)| > \text{const}\,|k|^{-\nu}$ for all $k \in \mathbb{Z}^m \backslash \{0\}$ and some $\nu > 0$, then one can choose $\rho = O(\varepsilon)$. In this case, higher approximations of the procedure of elimination of fast variables allow one to describe the dynamics with an accuracy of the order of any power in $\varepsilon$ on time intervals of order $1/\varepsilon$ (see, e.g., Arnol'd *et al.* (1988)).

If the system is a multifrequency system, and frequencies are not constant (but depend on the slow variables $I$), then due to the evolution of slow variables the frequencies themselves are evolving slowly. At certain time moments, they can satisfy certain resonant relations. One of the phenomena that can take place here is a capture into a resonance; this capture leads to a large deviation of the solutions of the exact and averaged systems. However, the general Anosov averaging theorem (Anosov 1960) implies that if the frequencies $\omega$ are nonresonant for almost all $I$, then for any $\rho > 0$, the inequality $|I(t) - J(t)| < \rho$ is satisfied for $0 \le t \le K/\varepsilon$ for all initial data outside a set $E(\rho, \varepsilon)$ whose measure tends to 0 as $\varepsilon \to 0$. In many cases, it turns out that $\text{mes}\, E(\rho, \varepsilon) = O(\sqrt{\varepsilon}/\rho)$ (in particular, the sufficient condition for the last estimate is that $\text{rank}(\partial\omega/\partial I) = m$) (Arnol'd *et al.* (1988)).

The knowledge about averaging in two-frequency systems ($m = 2$) on time intervals, of order of $1/\varepsilon$, is relatively more complete (see Arnol'd (1983), Arnol'd *et al.* (1988), and Lochak and Meunier (1988)). For Hamiltonian and reversible systems, the justification of the averaging method is a by-product of Kolmogorov–Arnold–Moser (KAM) theory. The KAM theory provides estimates of the difference between the solutions of the exact and averaged systems for majority of initial data on infinite time interval $-\infty < t < +\infty$. For remaining data this difference can grow because of Arnol'd diffusion, but, in general, very slowly. According to the Nekhoroshev theorem, this difference is small on time intervals whose length grows exponentially when the perturbation decays linearly (for an analytic Hamiltonian if the unperturbed Hamiltonian is a generic function, the so-called steep function).

Another aspect of justification of the averaging method is establishing relations between invariant manifolds of the exact and averaged systems. Consider, in particular, the case of a one-frequency

system and a multifrequency system with constant Diophantine frequencies. Suppose that the averaged system has an equilibrium such that real parts of all its eigenvalues are different from 0, or a limit cycle such that the absolute values of all but one of its multipliers are different from 1. Then the exact system has an invariant torus, respectively, $m$- or $(m+1)$-dimensional, whose projection onto the space of the slow variables is $O(\varepsilon)$-close to the equilibrium (cycle) of the averaged system. This torus is stable or unstable together with the equilibrium (cycle) of the averaged system. For Hamiltonian and reversible systems, the problem of invariant manifolds is considered in the framework of the KAM theory.

## Averaging in Bogolyubov's Systems

Systems in the standard form of Bogolyubov (1945) are of the form

$$\dot{x} = \varepsilon X(t, x, \varepsilon), \quad x \in \mathbb{R}^p,\ 0 < \varepsilon \ll 1 \qquad [10]$$

It is assumed that the function $X$, besides the usual smoothness conditions, satisfies the condition of uniform average: the limit (time average)

$$X_0(x) = \lim_{T\to\infty} \frac{1}{T} \int_0^T X(t, x, 0)\,\mathrm{d}t \qquad [11]$$

exists uniformly in $x$. The averaging principle of Bogolyubov consists of the replacement of the original system in standard form by the averaged system

$$\dot{\xi} = \varepsilon X_0(\xi) \qquad [12]$$

with a goal to provide an approximate description of the behavior of $x$. This approach generalizes the approach of the section "Averaging principle" for the case of constant frequencies ($\omega = \text{const}$). Upon introducing in the given system with constant frequencies the deviation from uniform rotation $\alpha = \varphi - \omega t$ and denoting $x = (I, \alpha)$, we obtain a system in the standard form [10]. Here the condition of uniform average is fulfilled because $X(t, x, 0)$ is a quasiperiodic function of time $t$. The averaged system [12] for nonresonant frequencies coincides with the averaged system [2]; for resonant frequencies, it coincides with the partially averaged system [3] (one should only supply systems [2] and [3] with equations for some components of the vector $\varphi - \omega t$ that do not enter into the right-hand side of the averaged system).

The averaging principle of Bogolyubov is justified by three Bogolyubov theorems. According to the first theorem, if $\xi(t), 0 \le t \le K/\varepsilon$, is a solution of the averaged system, and $x(t)$ is a solution of the exact system with initial condition $x(0) = \xi(0)$, then for any $\rho > 0$ there exists $\varepsilon_0(\rho) > 0$ such that $|x(t) - \xi(t)| < \rho$ for $0 \le t \le K/\varepsilon$ and $0 < \varepsilon < \varepsilon_0(\rho)$. The second and the third Bogolyubov theorems describe the motion in the neighborhoods of equilibria and the limit cycles of the averaged system. In particular, if for an equilibrium real parts of all its eigenvalues are different from 0, or, for a limit cycle, the absolute values of all but one multipliers are different from 1, then the exact system has a solution which eternally stays near this equilibrium (cycle). The stability properties of this solution are the same as the stability properties of the corresponding equilibrium (cycle) of the averaged system.

For systems of the form [10] a procedure exists that, similarly to the procedure in the section "Elimination of fast variables, decoupling of slow and fast motions," allows us to eliminate time $t$ from the right-hand side of the system with an accuracy of the order of any power in $\varepsilon$ by means of a transformation of variables. (To perform this procedure, one should assume that the conditions of uniform average are satisfied for functions that arise in the process of constructing higher approximations in this procedure (Bogolyubuv and Mitropol'skii 1961).) In the first approximation, such a transformation of variables transforms the original system into the averaged one.

The condition of uniform average is very important for theory. If the limit in [11] exists, but convergence is nonuniform in $x$, then the time average $X_0$ could be, for example, a discontinuous function of $x$, and the averaged system would not be well defined.

## Averaging in Slow–Fast Systems

Systems of the form [1] are particular cases of the systems of the form

$$\dot{x} = f(x, y, \varepsilon), \qquad \dot{y} = \varepsilon g(x, y, \varepsilon) \qquad [13]$$

which are called "slow–fast systems" (or systems with slow and fast motions, with slow and fast variables). The generalization of the approach of the section "Averaging principle" for these systems is the following averaging principle of Anosov (1960). In the system [6], let $x \in M, y \in \mathbb{R}^n$, where $M$ is a smooth compact $m$-dimensional manifold. At $\varepsilon = 0$, the system for fast variables $x$ contains slow variables $y$ as parameters. Assume that this system (which is called "fast system") has a finite smooth

invariant measure $\mu_y$ and is ergodic for almost all values of $y$. Introduce the averaged system

$$\dot{Y} = \varepsilon G(Y), \quad G(Y) = \frac{1}{\mu_Y(M)} \int_M g(x, Y, 0) \mathrm{d}\mu_Y$$

According to the averaging principle, one should use the solution $Y(t)$ of the averaged system with initial condition $Y(0) = y(0)$ for approximate description of slow motion $y(t)$ in the original system. This averaging principle is justified by the following Anosov theorem [1]: *for any positive $\rho$ the measure of the set $E(\rho, \varepsilon)$ of initial data (from a compact in the phase space) such that*

$$\max_{0 \le t \le 1/\varepsilon} |y(t) - Y(t)| > \rho$$

*tends to* $0$ *as* $\varepsilon \to 0$.

The particular case when the original system is a Hamiltonian system depending on slowly varying parameter $\lambda = \varepsilon t$, and for almost all values of $\lambda$ the motion of the system with $\lambda = \text{const}$ is ergodic on almost all energy levels, is considered in Kasuga (1961).

For the case when the has strong mixing properties, see Bakhtin (2004) and Kifer (2004).

For slow–fast systems, there is also a generalization of approach of the previous section that uses time averaging and the condition of uniform average (Volosov 1962).

## Applications of the Averaging Method

The averaging method is one of the most productive methods of perturbation theory, and its applications are immense. It is widely used in celestial mechanics and space flight dynamics for the description of the evolution of motions of celestial bodies, in plasma physics and theory of accelerators for description of motion of charged particles, and in radio engineering for the description of nonlinear oscillatory regimes. There are also applications in hydrodynamics, physics of lasers, optics, acoustics, etc. (see Arnol'd *et al.* (1988), Bogolyubov and Mitropol'skii (1961), Lochak and Meunier (1988), Mitropol'skii (1971), and Volosov (1962)).

*See also:* Central Manifolds, Normal Forms; Diagrammatic Techniques in Perturbation Theory; Hamiltonian Systems: Stability and Instability Theory; KAM Theory and Celestial Mechanics; Multiscale Approaches; Random Walks in Random Environments; Separatrix Splitting; Stability Problems in Celestial Mechanics; Stability Theory and KAM.

## Further Reading

Anosov DV (1960) Averaging in systems of ordinary differential equations with rapidly oscillating solutions. *Izvestiya Akademii Nauk SSSR, Ser. Mat.* 24(5): 721–742 (Russian).

Arnol'd VI (1983) *Geometrical Methods in the Theory of Ordinary Differential Equations.* New York–Berlin: Springer.

Arnol'd VI, Kozlov VV, and Neishtadt AI (1988) *Mathematical Aspects of Classical and Celestial Mechanics*, Encyclopaedia of Mathematical Sciences, vol. 3. Berlin: Springer.

Bakhtin VI (2004) Cramér asymptotics in the averaging method for systems with fast hyperbolic motions. *Proceedings of the Steklov Institute of Mathematics* 244(1): 79.

Bogolyubov NN (1945) On some statistical methods in mathematical physics. *Akad. Nauk USSR*. L'vov (Russian).

Bogolyubov NN and Mitropol'skii YuA (1961) *Asymptotic Methods in the Theory of Nonlinear Oscillations*. New York: Gordon and Breach.

Giacaglia GEO (1972) *Perturbation Methods in Nonlinear Systems*, Applied Mathematical Science, vol. 8. Berlin: Springer.

Kasuga T (1961) On the adiabatic theorem for the Hamiltonian system of differential equations in the classical mechanics I, II, III. *Proceedings of the Japan Academy* 37(7): 366–382.

Kevorkian J and Cole JD (1996) *Multiple Scale and Singular Perturbations Methods*, Applied Mathematical Sciences, vol. 114. New York: Springer.

Kifer Y (2004) Some recent advances in averaging. In: *Modern Dynamical Systems and Applications*, 403. Cambridge: Cambridge University Press.

Lochak P and Meunier P (1988) *Multiphase Averaging for Classical Systems*, Applied Mathematical Sciences, vol. 72. New York: Springer.

Mitropol'skii YuA (1971) *Averaging Method in Nonlinear Mechanics*. Kiev: Naukova Dumka (Russian).

Poincaré H (1957) *Les Méthodes Nouvelles de la Mécanique Céleste*, vols. 1–3. New York: Dover.

Sanders JA and Verhulst F (1985) *Averaging Methods in Nonlinear Dynamical Systems*, Applied Mathematical Sciences, vol. 59. New York: Springer.

Volosov VM (1962) Averaging in systems of ordinary differential equations. *Russian Mathematical Surveys* 17(6): 1–126.

# Axiomatic Approach to Topological Quantum Field Theory

**C Blanchet**, Université de Bretagne-Sud, Vannes, France
**V Turaev**, IRMA, Strasbourg, France



## Introduction

The idea of topological invariants defined via path integrals was introduced by A S Schwartz (1977) in a special case and by E Witten (1988) in its full power. To formalize this idea, Witten (1988) introduced a notion of a topological quantum field theory (TQFT). Such theories, independent of Riemannian metrics, are rather rare in quantum physics. On the other hand, they admit a simple axiomatic description first suggested by M Atiyah (1989). This description was inspired by G Segal's (1988) axioms for a two-dimensional conformal field theory. The axiomatic formulation of TQFTs makes them suitable for a purely mathematical research combining methods of topology, algebra, and mathematical physics. Several authors explored axiomatic foundations of TQFTs (see Quinn (1995) and Turaev (1994).

## Axioms of a TQFT

An $(n+1)$-dimensional TQFT $(\boldsymbol{V}, \tau)$ over a scalar field $\boldsymbol{k}$ assigns to every closed oriented $n$-dimensional manifold $X$ a finite-dimensional vector space $\boldsymbol{V}(X)$ over $\boldsymbol{k}$ and assigns to every cobordism $(M, X, Y)$ a $\boldsymbol{k}$-linear map

$$\tau(M) = \tau(M, X, Y) : \boldsymbol{V}(X) \to \boldsymbol{V}(Y)$$

Here a cobordism $(M, X, Y)$ between $X$ and $Y$ is a compact oriented $(n+1)$-dimensional manifold $M$ endowed with a diffeomorphism $\partial M \approx \overline{X} \amalg Y$ (the overline indicates the orientation reversal). All manifolds and cobordisms are supposed to be smooth. A TQFT must satisfy the following axioms.

1. *Naturality* Any orientation-preserving diffeomorphism of closed oriented $n$-dimensional manifolds $f : X \to X'$ induces an isomorphism $f_\sharp : \boldsymbol{V}(X) \to \boldsymbol{V}(X')$. For a diffeomorphism $g$ between the cobordisms $(M, X, Y)$ and $(M', X', Y')$, the following diagram is commutative:

$$\begin{array}{ccc} \boldsymbol{V}(X) & \xrightarrow{(g|_X)_\sharp} & \boldsymbol{V}(X') \\ {\scriptstyle \tau(M)}\downarrow & & \downarrow{\scriptstyle \tau(M')} \\ \boldsymbol{V}(Y) & \xrightarrow{(g|_Y)_\sharp} & \boldsymbol{V}(Y') \end{array}$$

2. *Functoriality* If a cobordism $(W, X, Z)$ is obtained by gluing two cobordisms $(M, X, Y)$ and $(M', Y', Z)$ along a diffeomorphism $f : Y \to Y'$, then the following diagram is commutative:

$$\begin{array}{ccc} \boldsymbol{V}(X) & \xrightarrow{\tau(W)} & \boldsymbol{V}(Z) \\ {\scriptstyle \tau(M)}\downarrow & & \downarrow{\scriptstyle \tau(M')} \\ \boldsymbol{V}(Y) & \xrightarrow{f_\sharp} & \boldsymbol{V}(Y') \end{array}$$

3. *Normalization* For any $n$-dimensional manifold $X$, the linear map

$$\tau([0,1] \times X) : \boldsymbol{V}(X) \to \boldsymbol{V}(X)$$

is identity.

4. *Multiplicativity* There are functorial isomorphisms

$$\boldsymbol{V}(X \amalg Y) \approx \boldsymbol{V}(X) \otimes \boldsymbol{V}(Y)$$
$$\boldsymbol{V}(\emptyset) \approx \boldsymbol{k}$$

such that the following diagrams are commutative:

$$\begin{array}{ccc} \boldsymbol{V}((X \amalg Y) \amalg Z) & \approx & (\boldsymbol{V}(X) \otimes \boldsymbol{V}(Y)) \otimes \boldsymbol{V}(Z) \\ \downarrow & & \downarrow \\ \boldsymbol{V}(X \amalg (Y \amalg Z)) & \approx & \boldsymbol{V}(X) \otimes (\boldsymbol{V}(Y) \otimes \boldsymbol{V}(Z)) \end{array}$$

$$\begin{array}{ccc} \boldsymbol{V}(X \amalg \emptyset) & \approx & \boldsymbol{V}(X) \otimes \boldsymbol{k} \\ \downarrow & & \downarrow \\ \boldsymbol{V}(X) & = & \boldsymbol{V}(X) \end{array}$$

Here $\otimes = \otimes_{\boldsymbol{k}}$ is the tensor product over $\boldsymbol{k}$. The vertical maps are respectively the ones induced by the obvious diffeomorphisms, and the standard isomorphisms of vector spaces.

5. *Symmetry* The isomorphism

$$\boldsymbol{V}(X \amalg Y) \approx \boldsymbol{V}(Y \amalg X)$$

induced by the obvious diffeomorphism corresponds to the standard isomorphism of vector spaces

$$\boldsymbol{V}(X) \otimes \boldsymbol{V}(Y) \approx \boldsymbol{V}(Y) \otimes \boldsymbol{V}(X)$$

Given a TQFT $(\boldsymbol{V}, \tau)$, we obtain an action of the group of diffeomorphisms of a closed oriented $n$-dimensional manifold $X$ on the vector space $\boldsymbol{V}(X)$. This action can be used to study this group.

An important feature of a TQFT $(\boldsymbol{V}, \tau)$ is that it provides numerical invariants of compact oriented $(n+1)$-dimensional manifolds without boundary. Indeed, such a manifold $M$ can be considered as a cobordism between two copies of $\emptyset$ so that $\tau(M) \in Hom_{\boldsymbol{k}}(\boldsymbol{k}, \boldsymbol{k}) = \boldsymbol{k}$. Any compact oriented $(n+1)$-dimensional manifold $M$ can be considered as a

cobordism between $\emptyset$ and $\partial M$; the TQFT assigns to this cobordism a vector $\tau(M)$ in $\mathrm{Hom}_{\boldsymbol{k}}(\boldsymbol{k}, \boldsymbol{V}(\partial M)) = \boldsymbol{V}(\partial M)$ called the vacuum vector.

The manifold $[0,1] \times X$, considered as a cobordism from $\overline{X} \amalg X$ to $\emptyset$ induces a nonsingular pairing

$$\boldsymbol{V}(\overline{X}) \otimes \boldsymbol{V}(X) \to \boldsymbol{k}$$

We obtain a functorial isomorphism $\boldsymbol{V}(\overline{X}) = \boldsymbol{V}(X)^* = \mathrm{Hom}_{\boldsymbol{k}}(\boldsymbol{V}(X), \boldsymbol{k})$.

We now outline definitions of several important classes of TQFTs.

If the scalar field $\boldsymbol{k}$ has a conjugation and all the vector spaces $\boldsymbol{V}(X)$ are equipped with natural nondegenerate Hermitian forms, then the TQFT $(\boldsymbol{V}, \tau)$ is Hermitian. If $\boldsymbol{k} = \boldsymbol{C}$ is the field of complex numbers and the Hermitian forms are positive definite, then the TQFT is unitary.

A TQFT $(\boldsymbol{V}, \tau)$ is nondegenerate or cobordism generated if for any closed oriented $n$-dimensional manifold $X$, the vector space $\boldsymbol{V}(X)$ is generated by the vacuum vectors derived as above from the manifolds bounded by $X$.

Fix a Dedekind domain $D \subset \boldsymbol{C}$. A TQFT $(\boldsymbol{V}, \tau)$ over $\boldsymbol{C}$ is almost $D$-integral if it is nondegenerate and there is $d \in \boldsymbol{C}$ such that $d\tau(M) \in D$ for all $M$ with $\partial M = \emptyset$. Given an almost integral TQFT $(\boldsymbol{V}, \tau)$ and a closed oriented $n$-dimensional manifold $X$, we define $S(X)$ to be the $D$-submodule of $\boldsymbol{V}(X)$ generated by all the vacuum vectors. This module is preserved under the action of self-diffeomorphisms of $X$ and yields a finer "arithmetic" version of $\boldsymbol{V}(X)$.

The notion of an $(n+1)$-dimensional TQFT over $\boldsymbol{k}$ can be reformulated in the categorical language as a symmetric monoidal functor from the category of $n$-manifolds and $(n+1)$-cobordisms to the category of finite-dimensional vector spaces over $\boldsymbol{k}$. The source category is called the $(n+1)$-dimensional cobordism category. Its objects are closed oriented $n$-dimensional manifolds. Its morphisms are cobordisms considered up to the following equivalence: cobordisms $(M, X, Y)$ and $(M', X, Y)$ are equivalent if there is a diffeomorphism $M \to M'$ compatible with the diffeomorphisms $\partial M \approx \overline{X} \amalg Y \approx \partial M'$.

## TQFTs in Low Dimensions

TQFTs in dimension $0 + 1 = 1$ are in one-to-one correspondence with finite-dimensional vector spaces. The correspondence goes by associating with a one-dimensional TQFT $(\boldsymbol{V}, \tau)$ the vector space $\boldsymbol{V}(pt)$ where $pt$ is a point with positive orientation.

Let $(\boldsymbol{V}, \tau)$ be a two-dimensional TQFT. The linear map $\tau$ associated with a pair of pants (a 2-disk with two holes considered as a cobordism between two circles $S^1 \amalg S^1$ and one circle $S^1$) defines a commutative multiplication on the vector space $\mathcal{A} = \boldsymbol{V}(S^1)$. The 2-disk, considered as a cobordism between $S^1$ and $\emptyset$, induces a nondegenerate trace on the algebra $\mathcal{A}$. This makes $\mathcal{A}$ into a commutative Frobenius algebra (also called a symmetric algebra). This algebra completely determines the TQFT $(\boldsymbol{V}, \tau)$. Moreover, this construction defines a one-to-one correspondence between equivalence classes of two-dimensional TQFTs and isomorphism classes of finite dimensional commutative Frobenius algebras (Kock 2003).

The formalism of TQFTs was to a great extent motivated by the three-dimensional case, specifically, Witten's Chern–Simons TQFTs. A mathematical definition of these TQFTs was first given by Reshetikhin and Turaev using the theory of quantum groups. The Witten–Reshetikhin–Turaev three-dimensional TQFTs do not satisfy exactly the definition above: the naturality and the functoriality axioms only hold up to invertible scalar factors called framing anomalies. Such TQFTs are said to be projective. In order to get rid of the framing anomalies, one has to add extra structures on the three-dimensional cobordism category. Usually one endows surfaces $X$ with Lagrangians (maximal isotropic subspaces in $H_1(X; \boldsymbol{R})$). For 3-cobordisms, several competing – but essentially equivalent – additional structures are considered in the literature: 2-framings (Atiyah 1989), $p_1$-structures (Blanchet *et al.* 1995), numerical weights (K Walker, V Turaev).

Large families of three-dimensional TQFTs are obtained from the so-called modular categories. The latter are constructed from quantum groups at roots of unity or from the skein theory of links. *See* Quantum 3-Manifold Invariants.

## Additional Structures

The axiomatic definition of a TQFT extends in various directions. In dimension 2 it is interesting to consider the so-called open–closed theories involving 1-manifolds formed by circles and intervals and two-dimensional cobordisms with boundary (G Moore, G Segal). In dimension 3 one often considers cobordisms including framed links and graphs whose components (resp. edges) are labeled with objects of a certain fixed category $\mathcal{C}$. In such a theory, surfaces are endowed with finite sets of points labeled with objects of $\mathcal{C}$ and enriched with tangent directions. In all dimensions one can study manifolds and cobordisms endowed with homotopy classes of mappings to a fixed space (homotopy quantum field theory, in the sense of Turaev). Additional structures on the tangent bundles – spin

structures, framings, etc. – may be also considered provided the gluing is well defined.

*See also:* Braided and Modular Tensor Categories; Hopf Algebras and *q*-Deformation Quantum Groups; Indefinite Metric; Quantum 3-Manifold Invariants; Topological Gravity, Two-Dimensional; Topological Quantum Field Theory: Overview.

## Further Reading


Atiyah M (1989) Topological Quantum Field Theories. *Publications Mathématiques de l'Ihés* 68: 175–186.

Bakalov B and Kirillov A Jr. (2001) Lectures on Tensor Categories and Modular Functors. *University Lecture Series* vol. 21. Providence, RI: American Mathematical Society.

Blanchet C, Habegger N, Masbaum G, and Vogel P (1995) Topological quantum field theories derived from the Kauffman bracket. *Topology* 34: 883–927.

Kock J (2003) Frobenius Algebras and 2D Topological Quantum Field Theories. *LMS Student Texts,* vol. 59. Cambridge: Cambridge University Press.

Quinn F (1995) Lectures on axiomatic topological quantum field Freed DS and Uhlenbeck KK (eds.) *Geometry and Quantum Field Theory*, pp. 325–453. IAS/Park City Mathematical Series, University of Texas, Austin: American Mathematical Society.

Segal G (1988) Two-dimensional conformal field theories and modular functors. In: Simon B, Truman A, and Davies IM (eds.) *IXth International Congress on Mathematical Physics*, pp. 22–37. Bristol: Adam Hilger Ltd.

Turaev V (1994) Quantum Invariants of Knots and 3-Manifolds. *de Gruyter Studies in Mathematics,* vol. 18. Berlin: Walter de Gruyter.

Witten E (1988) Topological quantum field theory. *Communication in Mathematical Physics* 117(3): 353–386.


# Axiomatic Quantum Field Theory

**B Kuckert**, Universität Hamburg, Hamburg, Germany



## Introduction

The term "axiomatic quantum field theory" subsumes a collection of research branches of quantum field theory analyzing the general principles of relativistic quantum physics. The content of the results typically is structural and retrospective rather than quantitative and predictive.

The first axiomatic activities in quantum field theory date back to the 1950s, when several groups started investigating the notion of scattering and *S*-matrix in detail (Lehmann, Symanzik, and Zimmermann 1955 (LSZ-approach), Bogoliubov and Parasiuk 1957, Hepp and Zimmermann (BPHZ-approach), Haag 1957–59 and Ruelle 1962 (Haag–Ruelle theory) (*see* Scattering, Asymptotic Completeness and Bound States and Scattering in Relativistic Quantum Field Theory: Fundamental Concepts and Tools).

Wightman (1956) analyzed the properties of the vacuum expectation values used in these approaches and formulated a system of axioms that the vacuum expectation values ought to satisfy in general. Together with Gårding (1965), he later formulated a system of axioms in order to characterize general quantum fields in terms of operator-valued functionals, and the two systems have been found to be equivalent.

A couple of spectacular theorems such as the PCT theorem and the spin–statistics theorem have been obtained in this setting, but no interacting quantum fields satisfying the axioms have been found so far (in $1+3$ spacetime dimensions). So, the development of alternatives and modifications of the setting got into the focus of the theory, and the axioms themselves became the objects of research. Their role as axioms – understood in the common sense – turned into the role of mere properties of quantum fields. Today, the term "axiomatic quantum field theory" is widely avoided for this reason.

In a long list of publications spread over the 1960s, Araki, Borchers, Haag, Kastler, and others worked out an algebraic approach to quantum field theory in the spirit of Segal's "postulates for general quantum Mechanics" (1947) (*see* Algebraic Approach to Quantum Field Theory).

The Wightman setting was the basis of a framework into which the causal construction of the *S*-matrix developed by Stückelberg (1951) and Bogoliubov and Shirkov (1959) has been fitted by Epstein and Glaser (1973). The causality principle fixes the time-ordered products up to a finite number of parameters at each order, which are to be put in as the renormalization constants.

Already in 1949, Dyson had seen that problems in the formulation of quantum electrodynamics (QED) could be avoided by "just" multiplying the time variable and, correspondingly, the energy variable by the imaginary unit constant ("Wick rotation"). Schwinger then investigated time-ordered Green functions of QED in this Euclidean setting. This approach was formulated in terms of axioms by Osterwalder and Schrader (1973, 1975) (*see* Euclidean Field Theory).

Other extensions of the aforementioned settings are objects of current research (*see* Indefinite Metric,

Quantum Field Theory in Curved Spacetime, Symmetries in Quantum Field Theory of Lower Spacetime Dimensions, and Thermal Quantum Field Theory).

## Quantum Fields

Gårding and Wightman characterized operator-valued quantum fields on the Minkowski spacetime $\mathbb{R}^{1+3}$ by a couple of axioms. Given additional assumptions concerning the high-energy behavior, the Gårding–Wightman fields are in one–one correspondence with algebraic field theories.

Without specifying or presupposing these additional assumptions, the axioms will now be formulated and discussed in detail and compared to the corresponding conditions in the algebraic setting. Adjoint operators are marked by an asterisk, and Einstein's summation convention is used.

**Operator-valued functionals** *The components of a field* $\boldsymbol{F}$ *are an n-tuple* $F_1 \cdots F_n$ *of linear maps that assign to each test function* $\varphi \in C_0^\infty(\mathbb{R}^{1+3})$ *linear operators* $F_1(\varphi) \cdots F_n(\varphi)$ *in a Hilbert space* $\mathcal{H}$ *with domains of definition* $D(F_1(\varphi)) \cdots D(F_n(\varphi))$. *There exists a dense subspace* $\mathcal{D}$ *of* $\mathcal{H}$ *with* $\mathcal{D} \subset D(F_\nu(\varphi)) \cap D(F_\nu(\varphi)^*)$ *and* $F_\nu(\varphi)\mathcal{D} \cup F_\nu(\varphi)^*\mathcal{D} \subset \mathcal{D}$ *for all indices* $\nu$. *Consider* $m$ *such fields* $\boldsymbol{F}^1 \cdots \boldsymbol{F}^m$ *with components* $F_\nu^a, 1 \le a \le m, \; 1 \le \nu \le n_a$. *Assume there to be an involution* $* : (1 \cdots m) \to (1 \cdots m)$ *such that* $F_\nu^{a^*}(\varphi) = F_\nu^a(\overline{\varphi})^*$, *where* $\overline{\varphi}(x) := \overline{\varphi(x)}$.

Quantum fields cannot be operator-valued functions on $\mathbb{R}^{1+3}$ if one wants them to exhibit (part of) the properties to follow. But point fields can be quadratic forms; typically this is the case for fields in a Fock space.

For each component $F_\nu^a$ and each open region $\mathcal{O} \subset \mathbb{R}^{1+3}$, the field operators $F_\nu^a(\varphi)$ with $\operatorname{supp} \varphi \subset \mathcal{O}$ generate a $^*$-algebra $\mathcal{F}_\nu^a(\mathcal{O})$ of operators defined on $\mathcal{D}$. These operators typically are unbounded, which is one of the differences with the traditional setting of the algebraic approach. There a $C^*$-algebra $\mathfrak{A}(\mathcal{O})$ is assigned to each open region $\mathcal{O}$ in such a way that $\mathcal{O} \subset \mathcal{P}$ implies $\mathfrak{A}(\mathcal{O}) \subset \mathfrak{A}(\mathcal{P})$. Each $C^*$-algebra is a $^*$-algebra, but in contrast to a $C^*$-algebra, a$^*$-algebra does not need to be endowed with a norm. The fundamental observables in quantum theory are bounded positive operators (typically, but not always, projections), and these generate a $C^*$-algebra.

There is no fundamental physical motivation for confining the setting to fields with a finite number of components, except that it includes most of the fields known from "daily life."

**Continuity as a distribution** *For all* $\Phi, \Psi \in \mathcal{D}$, *the linear functionals* $T_{\Phi,\Psi,\nu}$ *on* $C_0^\infty(\mathbb{R}^{1+3})$ *defined by*

$$T^a_{\nu,\Phi,\Psi}(\varphi) := \langle \Phi, F^a_\nu(\varphi)\Psi \rangle$$

*are distributions. They can be extended to tempered distributions.*

The Fourier transform of a tempered distribution is well defined as a tempered distribution. It is mainly due to the importance of Fourier transformations that the preceding assumption is convenient. Bogoliubov *et al.* (1975) remark that the assumption is not a mere technicality, since it rules out nonrenormalizable quantum fields.

**Microcausality (Bose–Fermi alternative)** *If* $\varphi$ *and* $\psi$ *are test functions with spacelike separated support, then*

$$F^a_\nu(\varphi)F^b_\mu(\psi)|_{\mathcal{D}} = \pm F^b_\mu(\psi)F^a_\nu(\varphi)|_{\mathcal{D}}.$$

The sign depends on the statistics of the fields, it is "−" if and only if both $\boldsymbol{F}^a$ and $\boldsymbol{F}^b$ are fermion fields.

Microcausality is closely related to Einstein causality. Einstein causality requires that any two observables located in spacelike separated regions commute in the strong sense, that is, their spectral measures commute. But fields with Fermi–Dirac statistics are not observables, and not even for Bose–Einstein fields with self-adjoint field operators does the above condition imply that the spectral projections commute, which is the criterion for commensurability. The sign on the right-hand side does, however, specify the statistics of the field.

This is a crucial difference with the algebraic approach. If $\mathcal{O}$ and $P$ are spacelike separated open regions and if $A \in \mathfrak{A}(\mathcal{O})$ and $B \in \mathfrak{A}(P)$, then one assumes, like in the above case, that $AB = BA$ (locality). But being elements of $C^*$-algebras, $A$ and $B$ are bounded operators (or can be represented accordingly), so if $A$ and $B$ are self-adjoint, they are, indeed, commensurable.

Doplicher, Haag, and Roberts (1974) and Buchholz and Fredenhagen (1984) have derived from this input of observables a field structure of *localized particle states*, and they showed that the statistics of these fields is Bose–Einstein, Fermi–Dirac, or some corresponding parastatistics (which is, *a priori*, forbidden if one assumes microcausality).

Recall that the unimodular group $SL(2,\mathbb{C})$ is isomorphic to the universal covering group of the restricted Lorentz group $L_+^\uparrow$ (the connected component containing the unit element). Denote by $\Lambda : SL(2,\mathbb{C}) \to L_+^\uparrow$ a covering map.

**Covariance** *There exist strongly continuous unitary representations U and T of* $\mathrm{SL}(2,\mathbb{C})$ *and* $(\mathbb{R}^{1+3},+)$, *respectively, and representations* $\boldsymbol{D}^1\cdots\boldsymbol{D}^m$ *of* $\mathrm{SL}(2,\mathbb{C})$ *in* $\mathbb{C}^{n_1}\cdots\mathbb{C}^{n_m}$, *respectively, such that*

$$U(g)F^a_\nu(\varphi)U(g)^* = D^a(g^{-1})^\mu_\nu F^a_\mu(\varphi(\Lambda(g)^{-1}\cdot))$$

*and*

$$T(y)F^a_\nu(\varphi)T(y)^* = F^a_\mu(\varphi(\cdot - y)),$$

*where* $D^a(g^{-1})^\mu_\nu$ *are the elements of the matrix* $\boldsymbol{D}^a(g^{-1})$. *Dropping coordinate indices, this reads*

$$U(g)\boldsymbol{F}^a(\varphi)U(g)^* = \boldsymbol{D}^a(g^{-1})\boldsymbol{F}^a(\varphi(\Lambda(g)^{-1}\cdot))$$

*and*

$$T(y)\boldsymbol{F}^a(\varphi)T(y)^* = \boldsymbol{F}^a(\varphi(\cdot - y)).$$

*The representations U and T generate a representation of the universal covering of the restricted Poincaré group.*

As it stands, this assumption is a very strong one, since it manifestly fixes the action of the representation on the field operators. In the algebraic approach, the covariance assumption is more modestly formulated. Namely, it is assumed that $U(g)\mathfrak{A}(\mathcal{O})U(g)^* = \mathfrak{A}(\Lambda(g)\mathcal{O})$ and $T(y)\mathfrak{A}(\mathcal{O})T(y)^* = \mathfrak{A}(\mathcal{O}+y)$, leaving open how the representation acts on the single local observables.

**Vacuum vector** *There exists a unique (up to a multiple) vector* $\Omega\in\mathcal{D}$ *that is invariant under the representations U and T and cyclic with respect to the algebra* $\mathcal{F}(\mathbb{R}^{1+3})$ *generated by all field operators* $F^a_\nu(\varphi)$, *that is,* $\overline{\mathcal{F}(\mathbb{R}^{1+3})\Omega} = \mathcal{H}$.

**Spectrum condition** *The joint spectrum of the components of the 4-momentum, i.e., of the generators of the spacetime translations, has support in the closed forward light cone* $\overline{V}_+$, *that is, the set* $\{k^2 \geq 0, k_0 \geq 0\}$.

The existence of an invariant ground state called the vacuum is standard in algebraic quantum field theory as well.

## *N*-Point Functions

Consider the above fields $\boldsymbol{F}^1\cdots\boldsymbol{F}^m$. For each $N\in\mathbb{N}$ and each $N$-tuple $(a_1\cdots a_N)$ of natural numbers $\leq m$ (labeling fields), define families $(\boldsymbol{F}^{a_1\cdots a_N}) := (F^{a_1\cdots a_N}_{\nu_1\cdots\nu_N})_{\nu_i\leq n_{a_i}}$ and $\boldsymbol{w}^{a_1\cdots a_N} := (w^{a_1\cdots a_N}_{\nu_1\cdots\nu_N})_{\nu_i\leq n_{a_i}}$ of distributions on $(\mathbb{R}^{1+3})^N$ by

$$F^{a_1\cdots a_N}_{\nu_1\cdots\nu_N}(\varphi_1\otimes\cdots\otimes\varphi_N) := F^{a_1}_{\nu_1}(\varphi_1)\cdots F^{a_N}_{\nu_N}(\varphi_N)$$

(using the nuclear theorem) and

$$w^{a_1\cdots a_N}_{\nu_1\cdots\nu_N}(\psi) := \langle\Omega, F^{a_1\cdots a_N}_{\nu_1\cdots\nu_N}(\psi)\Omega\rangle. \qquad [1]$$

These distributions are called the "$N$-point functions" of the fields $\boldsymbol{F}^1\cdots\boldsymbol{F}^m$ and yield the vacuum expectation values of the theory. It is straightforward to deduce the following properties from the Gårding–Wightman axioms.

**Microcausality (Bose–Fermi alternative)** *If* $\varphi_i$ *and* $\varphi_{i+1}$ *have spacelike separated supports, then*

$$\begin{aligned}&w^{a_1\cdots a_i a_{i+1}\cdots a_N}_{\nu_1\cdots\nu_i\nu_{i+1}\cdots\nu_N}(\varphi_1\otimes\cdots\otimes\varphi_i\otimes\varphi_{i+1}\otimes\cdots\otimes\varphi_N)\\&\quad = \pm w^{a_1\cdots a_{i+1}a_i\cdots a_N}_{\nu_1\cdots\nu_{i+1}\nu_i\cdots\nu_N}(\varphi_1\otimes\cdots\otimes\varphi_{i+1}\otimes\varphi_i\otimes\cdots\otimes\varphi_N).\end{aligned}$$

*or dropping coordinate indices,*

$$\begin{aligned}&\boldsymbol{w}^{a_1\cdots a_i a_{i+1}\cdots a_N}(\varphi_1\otimes\cdots\otimes\varphi_i\otimes\varphi_{i+1}\otimes\cdots\otimes\varphi_N)\\&\quad = \pm \boldsymbol{w}^{a_1\cdots a_{i+1}a_i\cdots a_N}(\varphi_1\otimes\cdots\otimes\varphi_{i+1}\otimes\varphi_i\otimes\cdots\otimes\varphi_N).\end{aligned}$$

**Invariance** *For all* $g\in\mathrm{SL}(2,\mathbb{C})$ *and* $y\in\mathbb{R}^{1+3}$, *one has*

$$\begin{aligned}&w^{a_1\cdots a_N}_{\mu_1\cdots\mu_N}(\varphi_1\otimes\cdots\otimes\varphi_N)\\&\quad = D^{a_1}(g^{-1})^{\nu_1}_{\mu_1}\cdots D^{a_N}(g^{-1})^{\nu_N}_{\mu_N}\\&\qquad \times w^{a_1\cdots a_N}_{\nu_1\cdots\nu_N}(\Lambda(g)\varphi_1\otimes\cdots\otimes\Lambda(g)\varphi_N)\\&\quad = w^{a_1\cdots a_N}_{\mu_1\cdots\mu_N}(\varphi_1(\cdot - y)\otimes\cdots\otimes\varphi_N(\cdot - y))\end{aligned}$$

*or dropping coordinate indices,*

$$\begin{aligned}&\boldsymbol{w}^{a_1\cdots a_N}(\varphi_1\otimes\cdots\otimes\varphi_N)\\&\quad = \left(\boldsymbol{D}^{a_1}(g^{-1})\otimes\cdots\otimes\boldsymbol{D}^{a_N}(g^{-1})\right)\\&\qquad \times \boldsymbol{w}^{a_1\cdots a_N}(\Lambda(g)\varphi_1\otimes\cdots\otimes\Lambda(g)\varphi_N)\\&\quad = \boldsymbol{w}^{a_1\cdots a_N}(\varphi_1(\cdot - y)\otimes\cdots\otimes\varphi_N(\cdot - y)).\end{aligned}$$

By translation invariance, the $N$-point functions $w^{a_1\cdots a_N}_{\nu_1\cdots\nu_N}(x_1\cdots x_N)$ only depend on the $N-1$ *relative*-position vectors $\xi_1 := x_1 - x_2$, $\xi_2 := x_2 - x_3, \ldots,$ $\xi_{N-1} := x_{N_1} - x_N$. This means that there are distributions $W^{a_1\cdots\nu_N}_{\nu_1\cdots\nu_N}$ on $(\mathbb{R}^{1+3})^{N-1}$ related to the $N$-point functions by the symbolic condition

$$w^{a_1\cdots a_N}_{\nu_1\cdots\nu_N}(x_1\cdots x_N) = W^{a_1\cdots a_N}_{\nu_1\cdots\nu_N}(\xi_1\cdots\xi_{N-1}).$$

In precise notation, this reads

$$w^{a_1\cdots a_N}_{\nu_1\cdots\nu_N}(\varphi) = \int_{1+3} W^{a_1\cdots a_N}_{\nu_1\cdots\nu_N}(\varphi_x)\,dx,$$

where

$$\varphi_x(\xi_1\cdots\xi_{N-1}) := \varphi(x, x-\xi_1, x-\xi_1-\xi_2, \ldots, x-\xi_1 - \cdots - \xi_{N-1}).$$

The functions $W^{a_1\cdots a_N}_{\nu_1\cdots\nu_N}$ are called the *Wightman functions*, and they have the following property because of the spectrum condition of the field.

**Spectrum condition** *The support of the Fourier transform of each* $W^{a_1 \cdots a_N}_{\nu_1 \cdots \nu_N}$ *is contained in* $(\overline{V}_+)^{N-1}$.

The uniqueness of the vacuum vector (up to a phase) is equivalent to the following condition.

**Cluster property** *For* $N \geq 2$, *let* $x$ *be a spacelike vector in* $\mathbb{R}^{1+3}$, *let* $L$ *be a natural number* $< N$, *and let* $\varphi$ *and* $\psi$ *be tempered test functions on* $(\mathbb{R}^{1+3})^L$ *and* $(\mathbb{R}^{1+3})^{N-L}$, *respectively. then*

$$\lim_{0<\lambda\to\infty} w^{a_1 \cdots a_N}_{\nu_1 \ldots \nu_N}(\varphi \otimes \psi(\cdot - \lambda x)) = w^{a_1 \cdots a_L}_{\nu_1 \cdots \nu_L}(\varphi) w^{a_{L+1} \cdots a_N}_{\nu_{L+1} \cdots \nu_N}(\psi).$$

On the one hand, these properties have been deduced from the Gårding–Wightman axioms via eqn [1]. Conversely, a family of distributions labeled in the above fashion and satisfying the above properties may be used to construct a Gårding–Wightman field theory provided that two more conditions – which hold for all systems of $N$-point functions – are satisfied. This requires some elementary notation.

Define the index sets

$$\mathcal{I}_N := \left\{ \begin{pmatrix} a_1 \cdots a_N \\ \nu_1 \cdots \nu_N \end{pmatrix} : 1 \leq a_i \leq m, 1 \leq \nu_i \leq n_{a_i} \text{ for all } 1 \leq i \leq N \right\}, \quad N \in \mathbb{N}$$

$\mathcal{I}_0 := \{\emptyset\}$, and $\mathcal{I} := \bigcup_{N \in \mathbb{N}_0} \mathcal{I}_N$. On $\mathcal{I}$ a concatenation $\circ$ is defined by

$$\begin{pmatrix} a_1 \cdots a_N \\ \nu_1 \cdots \nu_N \end{pmatrix} \circ \begin{pmatrix} b_1 \cdots b_M \\ \mu_1 \cdots \mu_M \end{pmatrix} := \begin{pmatrix} a_1 \cdots a_N \, b_1 \cdots b_M \\ \nu_1 \cdots \nu_N \, \mu_1 \cdots \mu_M \end{pmatrix}$$

and

$$\emptyset \circ \kappa := \kappa \circ \emptyset := \kappa$$

and an involution $*$ by

$$\begin{pmatrix} a_1 \cdots a_N \\ \nu_1 \cdots \nu_N \end{pmatrix}^* := \begin{pmatrix} a_N^* \cdots a_1^* \\ \nu_N \cdots \nu_1 \end{pmatrix} \quad \text{and} \quad \emptyset^* := \emptyset.$$

Define an antilinear involution $*$ on $\mathcal{S}^N := \mathcal{S}((\mathbb{R}^{1+3})^N)$ by

$$\psi(x_1 \cdots x_N) := \overline{\psi(x_N \cdots x_1)}$$

for each $N \in \mathbb{N}$. Put $\mathcal{S}^0 := \mathbb{C}$ and $z^* := \overline{z}$ for all $z \in \mathbb{C}$.

Define $\mathcal{S}^{\mathcal{I}_N} := \mathcal{S}^N \times \mathcal{I}_N$, and $\mathcal{S}^{\mathcal{I}} := \bigcup_N \mathcal{S}^{\mathcal{I}_N}$. For each $\kappa \in \mathcal{I}_N$, the set $\mathcal{S}^\kappa := \mathcal{S}((\mathbb{R}^{1+3})^N) \times \{\kappa\}$ is a linear space. On the direct sum $\mathcal{B}^{\mathcal{I}} := \bigoplus_{\kappa \in \mathcal{I}} \mathcal{S}^\kappa$ define an associative product by

$$(\psi, \kappa)(\chi, \lambda) := (\psi \otimes \chi, \kappa \circ \lambda)$$

and an antilinear involution $*$ by $(\psi, \kappa)^* := (\psi^*, \kappa^*)$. This endows $\mathcal{B}^{\mathcal{I}}$ with the structure of a nonabelian $*$-algebra with unit element $1 = (1, \emptyset)$ (Borchers algebra).

If one defines $F_\emptyset(z) := z\mathbf{1}$, then $w_\emptyset(z) = z$, and the Wightman functions induce a $\mathbb{C}$-linear functional $\omega$ on $\mathcal{B}^{\mathcal{I}}$ by

$$\omega(\psi, \kappa) := w_\kappa(\psi) \qquad [2]$$

$\omega$ exhibits the following two properties, which are the announced additional conditions required for reconstructing the fields from the $N$-point functions.

**Hermiticity** $\omega(\xi^*) = \overline{\omega(\xi)}$.

**Positivity** $\omega(\xi^* \xi) \geq 0$.

To see Hermiticity, compute

$$\begin{aligned} \omega(\psi^*, \kappa^*) &= \langle \Omega, F_{\kappa^*}(\psi^*)\Omega \rangle \\ &= \langle F_\kappa(\psi)\Omega, \Omega \rangle = \overline{\omega(\psi, \kappa)} \end{aligned}$$

and use $\mathbb{C}$-linearity to prove the statement for arbitrary $\xi \in \mathcal{B}$. For positivity, write any $\xi$ as a finite sum $\xi = (\psi_1, \kappa_1) + \cdots + (\psi_M, \kappa_M)$, and compute

$$\begin{aligned} \omega(\xi^* \xi) &= \omega\left( \sum_{i,j=1}^M (\psi_i, \kappa_i)^* (\psi_j, \kappa_j) \right) \\ &= \omega\left( \sum_{ij} (\psi_i^* \otimes \psi_j, \kappa_i^* \circ \kappa_j) \right) \\ &= \sum_{ij} w_{\kappa_i^* \circ \kappa_j}(\psi_i^* \otimes \psi_j) \\ &= \sum_{ij} \langle \Omega, F_{\kappa_i^* \circ \kappa_j}(\psi_i^* \otimes \psi_j)\Omega \rangle \\ &= \sum_{ij} \langle \Omega, F_{\kappa_i^*}(\psi_i^*) F_{\kappa_j}(\psi_j)\Omega \rangle \\ &= \sum_{ij} \langle F_{\kappa_i}(\psi_i)\Omega, F_{\kappa_j}(\psi_j)\Omega \rangle \\ &= \left\| \sum_i F_{\kappa_i}(\psi_i)\Omega \right\|^2 \geq 0. \end{aligned}$$

**Theorem 1** (Wightman's reconstruction theorem). *Let* $m$ *and* $n_1 \cdots n_m$ *be natural numbers, let* $\mathcal{I}_0, \mathcal{I}_1, \mathcal{I}_2, \ldots,$ *and* $\mathcal{I}$ *be the above index sets, and let* $\mathcal{B}^{\mathcal{I}}$ *be the above Borchers algebra. Let* $D_1 \cdots D_m$ *be matrix representations of* SL(2, $\mathbb{C}$) *in* $\mathbb{C}^{n_1} \cdots \mathbb{C}^{n_m}$, *respectively.*

*For each natural number* $N$, *let* $(w_\kappa)_{\kappa \in \mathcal{I}_N}$ *be a family of distributions on* $(\mathbb{R}^{1+3})^N$. *Suppose the family* $(w_\kappa)_{\kappa \in \mathcal{I}}$ *defined this way satisfies microcausality, covariance, spectrum condition, and the cluster property. If the linear functional* $\omega$ *defined on* $\mathcal{B}^{\mathcal{I}}$ *by eqn [2] is Hermitian and positive, then*

*there is (up to unitary equivalence) a unique family* $F^1 \cdots F^m$ *of Gårding–Wightman fields with* $n_1 \cdots n_m$ *components such that eqn [1] holds.*

The proof uses the GNS construction known from the theory of operator algebras. The Borchers algebra plays several roles. On the one hand, it is a linear space with an inner product. The Hilbert space $\mathcal{H}$ and the invariant space $\mathcal{D}$ of the field theory are constructed from this structure. On the other hand, the Borchers algebra acts on itself as an algebra of linear operators by its own algebra multiplication. This is the structure the $*$-algebra of field operators is constructed from.

## Results

The mathematical and structural analysis of quantum fields has improved the understanding of scattering theory in the different approaches mentioned above; see Bogoliubov *et al.* (1975) and the relevant articles in this encyclopedia. Apart from this, the following results deserve to be mentioned. Evidently, many others have to be omitted for practical reasons.

### PCT Symmetry

An early famous result was Lüders's proof (1957) that all fields in the above setting exhibit PCT symmetry, that is, the symmetry under reflections in all space and time variables combined with a charge conjugation. This symmetry is exhibited by all particle reactions observed so far. The proof, like several of the main results, made extensive use of the fact that the $N$-point functions are boundary values of analytic functions due to the spectrum condition, and that a fundamental theorem by Bargmann, Hall, and Wightman (1957) yields invariant analytic extensions.

### Reeh–Schlieder Theorem

For each field $F^a_\nu$ and each bounded open region $\mathcal{O} \subset \mathbb{R}^{1+3}$, the vacuum vector is cyclic with respect to $\mathcal{F}^a_\nu(\mathcal{O})$ (Reeh and Schlieder 1961). So excitations of the vacuum vector by field operators located in $\mathcal{O}$ are not to be considered as state vectors of a particle localized in $\mathcal{O}$, since they are not perpendicular to the excitations by field operators located outside $\mathcal{O}$.

### Unruh Effect and Modular $P_1CT$ Symmetry

In the 1970s, Bisognano and Wichmann (1975, 1976) discovered a surprising link of symmetries to the intrinsic algebraic structure of quantum fields, which is established by the Tomita–Takesaki modular theory (*see* Tomita–Takesaki Modular Theory). Namely, the unitary operators implementing the Lorentz boosts on the fields are elements of modular groups. This means that a uniformly accelerated observer perceives the vacuum as a thermal state with a temperature proportional to its acceleration, corresponding to the famous Unruh effect.

In addition, it was shown that $P_1CT$ symmetries (i.e., PCT combined with rotations by the angle $\pi$) are implemented by modular conjugations (modular $P_1CT$ symmetry). Modular $P_1CT$ symmetry is a consequence of the Unruh effect (Guido and Longo 1995).

### Spin and Statistics

Immediately following Lüders's PCT theorem, the spin–statistics theorem was proved for the $N$-point functions of the Wightman setting (Lüders and Zumino 1958, Burgoyne 1958, Dell'Antonio 1961). This was a remarkable and widely acknowledged progress. But as remarked earlier, the confinement to finite-component fields, which is used in the proof, cannot be motivated by physical first principles (i.e., in a truly axiomatic fashion). The representation $D$ of $SL(2, \mathbb{C})$ acting on the components, however, is forced to be finite dimensional by this assumption, and since the representations $D^a$ are objects of investigation, a considerable part of the result is assumed this way from the outset. Even more so, there are examples of fields with a "wrong" spin–statistics connection and infinitely many components.

This was one reason to continue working on the subject. At the beginning of the 1990s, it was found that the spin–statistics theorem can be derived from the symmetries discovered by Bisognano and Wichmann, and Unruh. Two approaches not referring to the number of internal degrees of freedom have been worked out: one assumes the Unruh effect (Guido and Longo 1995), the other modular $P_1CT$ symmetry (Kuckert 1995, 2005, Kuckert and Lorenzen 2005). The first approach has been generalized to conformal fields, the second to the case that the symmetry group's homogeneous part is not $SL(2, \mathbb{C})$, but only $SU(2)$.

Both approaches can be applied to infinite-component fields. They yield existence theorems; a distinguished representation is constructed from the modular symmetries, and this representation exhibits Pauli's spin–statistics connection. As mentioned before, nothing more can be expected at this level of generality. The line of argument works in both the algebraic and the Wightman setting.

### A Dynamical Property of the Vacuum

One can derive the spectrum condition, the Bisognano–Wichmann symmetries/the Unruh effect, and

covariance from the condition that no (inertial or) uniformly accelerated observer can extract mechanical energy from the field *in vacuo* by means of a cyclic process (Kuckert 2002).

#### Interacting Fields

The examples of interacting quantum fields that fit into the above settings live in one or two spatial dimensions only, and their relevance for physics mainly consists in being such examples. This has contributed to some frustration and to doubts on whether one is not, in fact, proving theorems on pretty empty sets, or in other words, working on "the most sophisticated theory of the free field."

The computations in quantum field theory are, like most of the computations in physics, perturbative. In order to be successful, they need to yield good agreement with experiment with reasonable computational efforts, that is, by evolution up to the second or third order. This asymptotic convergence is more important than convergence of the series as a whole. There are low-dimensional examples of interacting Wightman fields (e.g., $(\varphi^4)_2$; cf. the monograph by Glimm and Jaffe (1987)), and time will tell whether four-dimensional interacting Wightman fields exist. But there is no reason to expect convergence for general interacting fields; for example, QED does not fit into the Wightman framework.

The appropriate extension of the Wightman setting has been formulated by Epstein and Glaser (1973). It defines the *S-matrix* rather than the field itself as a (in general divergent) formal power series of operator-valued distributions.

The above results apply to this somewhat more modest setting as well, so the "axiomatic" approaches do help in understanding the known high-energy physics interactions. This even includes gauge theories (*see* Perturbative Renormalization Theory and BRST). The high-precision results of QED can be reproduced within this setting, and there occur no UV singularities: renormalization amounts to the need to extend distributions by fixing some parameters, that is, the renormalization constants. The infrared problem is circumvented by considering the *S*-matrix as a (position-dependent) distribution taking values in the unitary formal power series of distributions rather than as a single (global) unitary operator (or unitary power series).

#### Quantum Energy Inequalities

Energy densities of Wightman fields admit negative expectation values (Epstein, Glaser, and Jaffe 1965). This is in contrast to the positivity conditions that the energy–momentum tensors of classical general (and, hence, also special) relativity have to satisfy to ensure causality. But the conflict can be solved by smearing the densities out in space or time, as has first been realized by Ford (1991). The extent to which the energy density can become negative depends on the extent to which it is smeared out: "more smearing means less violation of positivity," so the classical positivity conditions are restored at medium and large scales. There are many ways to make this principle concrete. Quantum energy inequalities hold for thermodynamically well-behaved quantum fields on causally well-behaved classical spacetime backgrounds.

## Bibliographic Notes

Important monographs on axiomatic quantum field theory are those by Streater and Wightman (1964), Jost (1965), Bogoliubov *et al.* (1975), and Bogoliubov *et al.* (1990). Note that the books of Bogoliubov *et al.* differ in setup fundamentally and that neither replaces the other. For a lecture notes volume, see also Völkel (1977), and for a review article, see Streater (1975). A valuable discussion of the Wightman axioms can also be found in the second volume of the series by Reed and Simon (1970).

The first monograph on the algebraic approach to quantum field theory is due to Haag (1992), a more recent one has been written by Araki (1999). Concerning the sufficient conditions for "switching" between the Gårding–Wightman and the algebraic approach, see Wollenberg (1988) and the Ph.D. thesis of Bostelmann (2000) and references given there. Dynamical and thermodynamical foundation of standard axioms, the Bisognano–Wichmann symmetries (Unruh effect), and the spin–statistics theorem, have been investigated by Kuckert (2002, 2005), see also the references given there for related work.

In different formulations and at differing degrees of mathematical sophistication, the causal approach to perturbation theory can be found in the monographs by Bogoliubov and Shirkov (1959), Scharf (1989, 2001), and Steinmann (2000). Two modern review articles have been written by Brunetti and Fredenhagen (2000) and by Dütsch and Fredenhagen (2004).

The reference original articles on the Euclidean axioms are those of Osterwalder and Schrader (1973, 1975). Note that the first one contains an error. (cf. also Zinoviev (1995)). A monograph on Euclidean field theory and its relations to the other axiomatic settings of quantum field theory and to statistical mechanics is that by Glimm and Jaffe (1987).

A recent review on quantum energy inequalities is due to Fewster (2003).

## Acknowledgments

The author is a fellow of the Emmy-Noether Programme (DFG). Thanks for discussions are due to Professor D Arlt.

*See also:* Algebraic Approach to Quantum Field Theory; $C^*$-Algebras and Their Classification; Constructive Quantum Field Theory; Dispersion Relations; Euclidean Field Theory; Indefinite Metric; Perturbative Renormalization Theory and BRST; Quantum Field Theory: A Brief Introduction; Quantum Field Theory in Curved Spacetime; Scattering, Asymptotic Completeness and Bound States; Scattering in Relativistic Quantum Field Theory: Fundamental Concepts and Tools; Scattering in Relativistic Quantum Field Theory: The Analytic Program; Symmetries in Quantum Field Theory: Algebraic Aspects; Symmetries in Quantum Field Theory of Lower Spacetime Dimensions; Thermal Quantum Field Theory; Tomita–Takesaki Modular Theory; Two-Dimensional Models.

## Further Reading

Araki H (1999) *Mathematical Theory of Quantum Fields*. Oxford: Oxford University Press.

Bogoliubov NN, Logunov AA, and Todorov IT (1975) *Introduction to Axiomatic Quantum Field Theory*, (Russian original edition: Nauka (Moskow) 1969). New York: Benjamin.

Bogoliubov NN, Logunov AA, Oksak AI, and Todorov IT (1990) *General Principles of Quantum Field Theory*, (Russian original edition Nauka (Moskow) 1987). Dordrecht–Boston–London: Kluwer.

Bostelmann H (2000) *Lokale Algebren und Operatorprodukte am Punkt* (in German). Ph.D. thesis, Göttingen.

Brunetti R and Fredenhagen K (2004) Microlocal Analysis and Interacting Quantum Field Theories: Renormalization on Physical Backgrounds. *Communications in Mathematical Physics* 208: 623.

Dütsch and Fredenhagen K (2004) Causal Perturbation Theory in terms of retarded products, and a proof of the Action Ward Identity, to appear in. *Rev. Math. Phys*.

Fewster CJ (2003) *Energy Inequalities in Quantum Field Theory*. Proceedings of the International Conference on Mathematical Physics (revised version under math-ph/0501073).

Glimm J and Jaffe A (1987) *Quantum Physics: A Functional Integral Point of View*, 2nd edn. Berlin–Heidelberg–New York: Springer.

Guido D and Longo R (1995) An algebraic spin and statistics Theorem. *Communications in Mathematical Physics* 172: 517.

Haag R (1992) *Local Quantum Physics*. Berlin–Heidelberg–New York: Springer.

Jost R (1965) *The General Theory of Quantized Fields*. American Mathematical Society.

Kuckert B (2002) Covariant thermodynamics of quantum systems: passivity, semipassivity, and the Unruh effect. *Annals of Physics* 295: 216.

Kuckert B (2005) Spin, statistics, and reflections, I. *Annales Henri Poincaré* 6: 849.

Kuckert B and Lorenzen R (2005) Spin, Statistics, and Reflections, II. Preprint (math-ph/0512068).

Osterwalder K and Schrader R (1973) Axioms for Euclidean Green's functions. *Communications in Mathematical Physics* 31: 83.

Osterwalder K and Schrader R (1975) Axioms for Euclidean Green's functions. 2. *Communications in Mathematical Physics* 42: 281.

Reed M and Simon B (1970) *Methods of Modern Mathematical Physics*, (4 volumes). London: Academic Press.

Scharf G (1989) *Finite Quantum Electrodynamics*. Berlin–Heidelberg–New York: Springer.

Scharf G (2001) *Quantum Gauge Theories A True Ghost Story*. Weinheim: Wiley.

Streater RF (1975) Outline of axiomatic quantum field theory. *Reports on Progress in Physics* 38: 771.

Streater RF and Wightman AS (1964) *PCT, Spin & Statistics, and All That*. New York: Benjamin.

Völkel AH (1977) Fields, Particles, and Currents. *Lecture Notes in Physics*, vol. 66. Berlin–Heidelberg–New York: Springer.

Wollenberg M (1988) The existence of quantum fields for local nets of observables. *Journal of Mathematical Physics* 29: 2106.

Zinoviev YM (1995) Equivalence of Euclidean and Wightman field theories. *Communications in Mathematical Physics* 174: 1.

# B

## Bäcklund Transformations

**D Levi**, Università "Roma Tre", Rome, Italy



### Introduction

Bäcklund transformations appeared for the first time in the work of the geometers of the end of the nineteenth century, for instance, Bianchi, Lie, Bäcklund, and Darboux, when studying surfaces of constant curvature. If on a surface in three-dimensional Euclidean space, the asymptotic directions are taken as coordinate directions, then the surface metric may be written as

$$\mathrm{d}s^2 = \mathrm{d}x^2 + 2\cos(w)\,\mathrm{d}x\,\mathrm{d}y + \mathrm{d}y^2 \qquad [1]$$

where $w(x,y)$ is a function of the surface coordinates $x, y$. A necessary and sufficient condition for the surface to be of constant curvature is that $w$ satisfies the nonlinear partial differential equation

$$w_{,xy} = \sin(w) \qquad [2]$$

where the subscript denotes partial derivative. Equation [2] is nowadays called the sine Gordon (sG) equation. Bianchi (1879), Lie (1888, 1890, 1893), and Bäcklund (1874) introduced a transformation which allows one to pass from a solution of eqn [2] to a new solution, that is, from a surface of constant curvature to a new one. Starting from the work of Clarin (1903), this transformation has been referred to as Bäcklund transformation (BT). The BT for eqn [2] reads

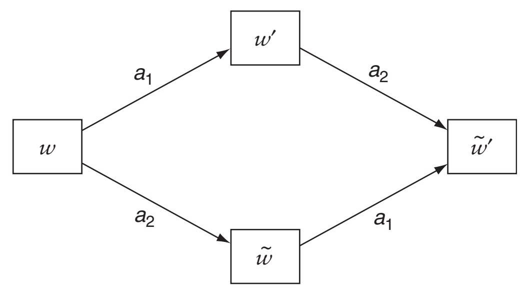
$$\tilde{w}_{,x} = w_{,x} + 2a\sin\left(\frac{\tilde{w}+w}{2}\right) \qquad [3a]$$

$$\tilde{w}_{,y} = -w_{,y} + \frac{2}{a}\sin\left(\frac{\tilde{w}-w}{2}\right) \qquad [3b]$$

where $a$ is a nonzero constant parameter and $\tilde{w}$ is a different solution of eqn [2]. It is immediate to prove by appropriate differentiation of eqns [3] with respect to $y$ and $x$ that both $w$ and $\tilde{w}$ must satisfy eqn [2]. The BT [3] provides a denumerable set of exact solutions once a solution $w$ is known. Bianchi showed that four such solutions can be related in an algebraic way:

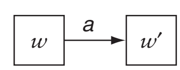
$$\tan\left(\frac{\tilde{w}'-w}{4}\right) = \frac{a_1+a_2}{a_1-a_2}\tan\left(\frac{w'-\tilde{w}}{4}\right) \qquad [4]$$

Equation [4] is derived using the permutability theorem proved by Bianchi in his Ph.D. thesis in 1879:




[5]

whereby the diagram




we mean a BT from $w$ to $w'$ with parameter $a$.

For sG equation [2] a trivial solution is given, for example, by $w(x,y) = \pi$. Then, from eqn [3a] we get

$$\tilde{w}(x,y) = 2\arcsin\left(\frac{1-\mathrm{e}^{-2[ax+\alpha(y)]}}{1+\mathrm{e}^{-2[ax+\alpha(y)]}}\right)$$

Introducing this result in eqn [3b], we get $\alpha_{,y} = -1/a$. So, the application of the BT [3] to sG equation gives the nontrivial solution

$$\boxed{w=\pi} \longrightarrow \boxed{\tilde{w} = 4\arctan\left(\frac{1-\mathrm{e}^{-[ax-y/a]}}{1+\mathrm{e}^{-[ax-y/a]}}\right)} \qquad [6]$$

Clarin (1903) extended the results of Bäcklund to the case of a generic partial differential equation of second order,

$$F(x, y, w, w_{,x}, w_{,y}, w_{,xx}, w_{,xy}, w_{,yy}) = 0 \qquad [7]$$

by assuming that

$$\begin{aligned} w_{,x} &= f(w, \tilde{w}, \tilde{w}_{,x}, \tilde{w}_{,y}) \\ w_{,y} &= g(w, \tilde{w}, \tilde{w}_{,x}, \tilde{w}_{,y}) \end{aligned} \qquad [8]$$

If the compatibility of eqns [8]

$$f_{,y} - g_{,x} = 0 \qquad [9]$$

is identically satisfied by eqn [7] for the variable $\tilde{w}(x, y)$, then we say that eqns [8] are an auto-Bäcklund transformation for eqn [7]. In this case, eqns [8] transform a solution of eqn [7] into a new solution of the same equation. Thus, eqns [8] simplify the problem of finding solutions of eqn [7]. Given one solution $w(x, y)$ of eqn [7], the existence of a BT reduces the problem of integrating eqn [7] into that of solving two first-order ordinary differential equations. From this point of view, the Cauchy–Riemann relations

$$w_{,x} = \tilde{w}_{,y}, \qquad w_{,y} = -\tilde{w}_{,x} \qquad [10]$$

for the Laplace equation

$$w_{,xx} + w_{,yy} = 0 \qquad [11]$$

are a BT *ante litteram* (however, without a free parameter).

Consider the case when $\tilde{w}(x, y)$ satisfies a different partial differential equation,

$$G(x, y, \tilde{w}, \tilde{w}_{,x}, \tilde{w}_{,y}, \tilde{w}_{,xx}, \tilde{w}_{,xy}, \tilde{w}_{,yy}) = 0 \qquad [12]$$

In this case, one still has a BT, but not an auto-BT. The best-known cases are when $F_1 = w_{,y} + w_{,xxx} + ww_{,x}$ and $G_1 = \tilde{w}_{,y} + \tilde{w}_{,xxx} + \tilde{w}^2\tilde{w}_{,x}$, and $F_2 = w_{,xy} - e^w$ and $G_2 = \tilde{w}_{,xy}$ (Lamb 1976). In the first case, the BT relates the Korteweg–de Vries (KdV) equation to the modified KdV equation and this transformation paved the way to the discovery of the complete integrability of the KdV equation by Gardner *et al.* (1967). In the second case, the BT relates the Liouville equation to the wave equation, and can be used to solve it completely. Due to the first example, often a non-auto-BT is denoted as Miura transformation.

One can now state an operative definition of BT, extending the results of Bäcklund and Clarin to more general equations.

**Definition 1** Consider two partial differential equations of order $m_1$ and $m_2$:

$$F_1(\boldsymbol{x}, \boldsymbol{u}, \underset{(1)}{\boldsymbol{u}}, \underset{(2)}{\boldsymbol{u}}, \ldots, \underset{(m_1)}{\boldsymbol{u}}) = 0 \qquad [13a]$$

$$F_2(\boldsymbol{x}, \tilde{\boldsymbol{u}}, \underset{(1)}{\tilde{\boldsymbol{u}}}, \underset{(2)}{\tilde{\boldsymbol{u}}}, \ldots, \underset{(m_2)}{\tilde{\boldsymbol{u}}}) = 0 \qquad [13b]$$

where $\boldsymbol{x} \in \mathbb{R}^n$ and $(\boldsymbol{u}, \tilde{\boldsymbol{u}}) \in \mathbb{C}^p$, and $\underset{(k)}{\boldsymbol{u}}$ is the set of $k$-order derivative of $\boldsymbol{u}$. The set of $n$ equations

$$G_j(\boldsymbol{x}, \boldsymbol{u}, \underset{(1)}{\boldsymbol{u}}, \ldots, \underset{(s_1)}{\boldsymbol{u}}; \tilde{\boldsymbol{u}}, \underset{(1)}{\tilde{\boldsymbol{u}}}, \ldots, \underset{(s_2)}{\tilde{\boldsymbol{u}}}) = 0$$
$$j = 1, 2, \ldots, n \qquad [14]$$

with $s_1 < m_1$ and $s_2 < m_2$, represents the BT of eqns [13] iff the compatibility of eqns [14] is identically satisfied on the solutions of eqns [13] and $G_j$ depends on a set of essential arbitrary constant parameters.

The Clarin formulation [8] and the classical BT for the sG [3] are clearly special subcases of this definition. When a solution of $F_1 = 0$ is known, a solution of $F_2 = 0$ is obtained by solving a set of lower-order partial differential equations. By a proper choice of the BT parameters, once a new solution is obtained by solving the BT [14], one can use the obtained solution as a starting point to construct another one, and so on. In this way, one can construct a whole ladder of solutions, *a priori* a denumerable set of solutions. This same construction has been applied also to the case of functional equations. In particular, it has been considered for the case of differential–difference and difference–difference equations both for finite (dynamical systems (Wojciechowski 1982)) and infinite lattices (Toda 1989).

In the case when $F_1$ and $F_2$ represent the same equation, $s_1 = s_2 = 1$ and the BTs $G_j = 0$ are linear in $\underset{(1)}{\boldsymbol{u}}$, then Definition 1 is strictly related to the notion of nonclassical symmetry or conditional symmetry (Levi and Winternitz 1989, Olver 1993), an extension of the concept of Lie symmetry used to reduce and integrate a differential equation. In the case of the nonclassical symmetries, the known solution $\tilde{\boldsymbol{u}}$ is included in the arbitrary $\boldsymbol{x}$-dependent coefficients of the transformation. In this case, the BT is just a way to construct an explicit solution of the differential equation [7].

Definition 1 is often too general to be able to get explicit results. It is constructive for any partial differential equation, linear or nonlinear, but if one is not able to get a nontrivial BT this does not mean that a BT does not exist. As noted later, the existence of an auto-BT is associated to the existence of an infinity of symmetries, and this is a condition for the exact integrability of eqn [13] (Fokas 1980, Ibragimov and Shabat 1980). So, the existence of a BT is closely related to the integrability of eqn [13].

## Bäcklund via Integrability

One can derive the BT from the integrability properties of eqn [13a]. Equation [13a] is said to be integrable if it can be written as the compatibility condition of an overdetermined system of linear partial differential equations for an auxiliary function depending on a free parameter belonging to the

complex $C$ plane. The prototype of such a situation is given by the Lax pair for the KdV equation

$$u_{,t} + u_{,xxx} - 6uu_{,x} = 0 \quad [15]$$

introduced by Lax (1968):

$$L\psi = k^2\psi, \quad L = -\partial_x^2 + u(x,t) \quad [16a]$$

$$\psi_{,t} = -M\psi, \quad M = 4\partial_{xxx} - 3(u\partial_x + \partial_x u) \quad [16b]$$

where $k$ is a free parameter and $\psi = \psi(x,t;k)$. As eqn [16a] is nothing else but the stationary Schrödinger equation, the function $\psi$ can be interpreted as a wave function, and $k^2$ is the spectral parameter corresponding to the potential $u(x,t)$. The condition for the existence of a solution $\psi$ of the overdetermined system of eqns [16] is given by the operator equation

$$L_{,t} = [L, M] \quad [17]$$

the so-called Lax equation. In the case of asymptotically bounded potentials, eqn [16a] defines the spectrum unique. Introducing the following asymptotic boundary conditions for the wave function $\psi$,

$$\begin{aligned} \psi(x,t;k) &\underset{x\to-\infty}{\longrightarrow} T(k,t)e^{-ikx} \\ \psi(x,t;k) &\underset{x\to+\infty}{\longrightarrow} e^{-ikx} + R(k,t)e^{ikx} \end{aligned} \quad [18]$$

where $R(k,t)$ and $T(k,t)$ are, respectively, the reflection and the transmission coefficient, the spectrum is defined in the complex plane of the variable $k$ by

$$S[u] \equiv \{R(k,t),\ -\infty < k < \infty;\ p_n, c_n(t),\ j = 1,2,\ldots,N\} \quad [19]$$

where $p_n$ are the bound state parameters corresponding to isolated singularities of the reflection coefficients on the imaginary positive $k$-axis corresponding to a solution $\phi_n(x,t;p_n)$ of the spectral problem vanishing for $x \to -\infty$ and such that

$$\lim_{x\to+\infty} [e^{p_n x}\phi_n(x,t;p_n)] = 1 \quad [20]$$

and $c_n$ are some functions of $t$ related to the residues of $R(k,t)$ at the poles $p_n$. There is a one-to-one correspondence between the evolution of the potential $u(x,t)$ in eqn [15] and that of the spectrum $S[u]$ of the Schrödinger spectral problem [16a]. In particular, for the KdV, taking into account eqn [16b], the evolution of the reflection coefficient $R(k,t)$ is given by

$$\frac{dR(k,t)}{dt} = 8ik^3 R(k,t) \quad [21]$$

In eqn [21] and henceforth, $d/dt$ denotes the total derivative with respect to $t$.

In the following, for the sake of the simplicity of exposition and for the concreteness of the presentation, all the results presented on the BT will be derived for the KdV equation. Similar results can be obtained and have been obtained in the literature for many classes of integrable partial differential equations in two and three dimensions and for differential–difference and difference–difference equations. For a partial review of the available recent literature on the subject, see Rogers and Shadwick (1982) and Coley *et al.* (2001)

A more general form of introducing the nonlinear partial differential equation as a compatibility of an overdetermined system of linear equations has been provided by Zaharov and Shabat (1979) with the dressing method (DM). In the DM, the differential equations [16] are substituted by a matrix system of linear equations

$$\Psi_{,x} = U(u(x,t),k)\Psi \quad [22a]$$

$$\Psi_{,t} = V(u(x,t),k)\Psi \quad [22b]$$

where $\Psi = \Psi(x,t;k)$ and $U$ and $V$ are matrix functions. The existence of a nonsingular solution of the system of linear equations [22] requires that the matrix functions $U$ and $V$ satisfy the equation

$$U_{,t} - V_{,x} + [U,V] = 0 \quad [23]$$

often called zero-curvature condition. The KdV equation [15] in the DM is obtained by choosing

$$\begin{aligned} U(u(x,t),k) &= \begin{pmatrix} ik & u(x,t) \\ 1 & -ik \end{pmatrix} \\ V(u(x,t),k) & \\ &= \begin{pmatrix} 2u + 4k^2 & -u_x - 2iku - 4ik^3 \\ u_x + 2iku + 4ik^3 & 2u(u+2k^2) - 2iku_x - u_x x \end{pmatrix} \end{aligned} \quad [24]$$

The existence of an auto-BT implies the existence of a differential equation (see Definition 1) which relates two solutions of the same nonlinear equation. The new solution $\tilde{u}(x,t)$ of eqn [15] will be associated to a different Lax operator and a different spectral problem (but of the same operational form)

$$\tilde{L} = -\partial_{xx} + \tilde{u}(x,t) \quad [25a]$$

$$\tilde{L}\tilde{\psi} = k^2\tilde{\psi} \quad [25b]$$

The existence of a relation between the potentials $u(x,t)$ and $\tilde{u}(x,t)$ thus implies that there must be a $(u,\tilde{u};k)$-dependent operator $D$ such that

$$\tilde{\psi} = D\psi \qquad [26]$$

The compatibility of eqns [16a], [25b], and [26] implies that $\tilde{L}D\psi = Dk^2\psi$, that is,

$$\tilde{L}D = DL \qquad [27]$$

Equation [27] is the auto-BT in the Lax formalism. If $\tilde{L}$ and $L$ are two different spectral problems related to two different nonlinear partial differential equations, then eqn [27] will provide a Miura transformation. In the DM, the requirement of the existence of a BT is given again by eqn [26] with $\psi$ and $\tilde{\psi}$ substituted by $\Psi$ and $\tilde{\Psi}$ and the operator $D$ substituted by a matrix function $\mathcal{D}$. The BT in the DM is given by

$$\mathcal{D}_{,x} = U(\tilde{u}(x,t),k)\mathcal{D} - \mathcal{D}U(u(x,t),k) \qquad [28a]$$

$$\mathcal{D}_{,t} = V(\tilde{u}(x,t),k)\mathcal{D} - \mathcal{D}V(u(x,t),k) \qquad [28b]$$

In the particular case of the Hilbert–Riemann problem with zeros, providing the soliton solutions, the matrix $\mathcal{D}$ can be expressed as a function of $\Psi$. In this way, one derives the Moutard or Darboux transformation (DT) (Moutard 1878, Levi *et al.* 1984), the most efficient way to get soliton solutions of the nonlinear partial differential equation.

Given a linear ordinary differential equation for the unknown $\psi$, depending on a set of arbitrary functions $u(x)$ and parameters $k$, the DT provides a discrete transformation which leaves the equation invariant. In the particular case of the KdV equation associated with the stationary Schrödinger spectral problem [16a], we have

$$\tilde{u}(x,t) = u(x,t) - 2(\log F(x,t))_{,xx} \qquad [29a]$$

$$\tilde{\psi}(x,t;k) = -\frac{\mathrm{i}}{k+\mathrm{i}p}\psi_{,x}(x,t;k) - \frac{F_x(x,t)}{F(x,t)}\psi(x,t;k) \qquad [29b]$$

where the intermediate wave function

$$F(x,t) = \psi(x,t;k=\mathrm{i}p) + a\psi(x,t;k=-\mathrm{i}p)$$

is a linear combination of the Jost solution of the Schrödinger spectral problem with $p$ a real parameter and $a$ an arbitrary constant. If one looks for an equation involving only the potentials $u$ and $\tilde{u}$, from eqns [29], one gets the BT for the KdV equation. Given a trivial solution of the KdV equation, together with the corresponding solution of the spectral problem, eqn [29a] provides a new solution of the KdV, while eqn [29b] gives a new solution of the spectral problem. This procedure can be carried out recursively and gives a ladder of explicit solutions for the KdV equation.

The DM is a particularly simple setting in which one can derive DTs. In fact, expressing the matrix $\mathcal{D}$ in terms of $\Psi$, eqn [28a] gives a relation between the potentials of the type given by eqn [29a], while eqn [26] gives eqn [29b]. Depending on the form of the matrix $\mathcal{D}$ in terms of $k$, one can introduce more parameters in the DT. The classical DT [29] depends on just one parameter; however, in the case of the Schrödinger spectral problem [16a], one can also have DTs depending on two parameters, a TDT.

A more general DT, which can provide solutions even when the initial solution is not bounded asymptotically, can be obtained for many equations and, in particular, also for the KdV equation. This is obtained in a particular limit of the TDT when the parameters coincide (Levi 1988) and it is often referred to as binary DT (Matveev and Salle 1991). The binary DT for the KdV is given by

$$\tilde{u}(x,t) = u(x,t) - 2(\log F(x,t))_{,xx} \qquad [30a]$$

$$\tilde{\psi}(x,t;k) = \frac{1}{k^2-\mu^2}\left\{\left(k^2-\mu^2-\frac{F(x,t)_{,xx}}{2F(x,t)}\right)\psi_{,x}(x,t;k) - \frac{F_x(x,t)}{F(x,t)}\psi(x,t;k)\right\} \qquad [30b]$$

where $\mu$ is a value of $k$ for which the function $\psi(x,t;k)$ is asymptotically bounded at $+\infty$ and the function $F(x,t)$ is given by

$$F(x,t) = 1 + \rho\int_x^{+\infty}\psi(y,t;\mu)^2\,\mathrm{d}y \qquad [31]$$

with $\rho$ an arbitrary constant. The corresponding BT obtained eliminating the function $F$ from eqns [30] reads

$$\tilde{q}_{,xx} - q_{,xx} = -\frac{1}{8}(\tilde{q}-q)^3 - [\tilde{q}_x + q_x - 2g(x) + 2\mu](\tilde{q}-q) + \frac{1}{2}\frac{(\tilde{q}_x - q_x)^2}{\tilde{q}-q} \qquad [32]$$

where $q = \int_x^\infty u_0(y,t)\,\mathrm{d}y$ with $u_0(x,t) = u(x,t) - g(x)$, the asymptotically bounded part of $u(x,t)$, and $g(x)$ its asymptotic behavior, and $\tilde{q} = \int_x^\infty \tilde{u}_0(y,t)\,\mathrm{d}y$ with $\tilde{u}_0(x,t) = \tilde{u}(x,t) - g(x)$.

Once the Lax operator $L$ is given, we can obtain in a constructive way the operators $M$ which give the admissible nonlinear partial differential

equations and the operators $D$ which give the admissible BT. A technique to do so is provided by the so-called Lax technique introduced by Bruschi and Ragnisco (1980a–c). Using the Lax technique, we can easily obtain the nonlinear partial differential equations and BT associated with the Lax operator [16a] both in the isospectral and nonisospectral case (when $k_{,t}=0$ and when $k_{,t}\neq 0$) and the corresponding evolution of the spectrum. We have

$$u_{,t} = f(\mathcal{L},t)u_x + g(\mathcal{L},t)[xu_x + 2u] \qquad [33a]$$

$$k_{,t} = kg(-4k^2,t) \qquad \frac{dR(k,t)}{dt} = 2ikf(-4k^2,t)R(k,t) \qquad [33b]$$

$$F(\Lambda)(\tilde{u}-u) + G(\Lambda)\Gamma\ 1 = 0 \qquad [33c]$$

$$\tilde{R}(k,t) = \frac{F(-4k^2) - 2ikG(-4k^2)}{F(-4k^2)+2ikG(-4k^2)}R(k,t) \qquad [33d]$$

where the functions $f$, $g$, $F$, and $G$ are entire functions of their first argument and the recursive operators $\mathcal{L}$ and $\Lambda$ are given by

$$\mathcal{L}f(x) = f_{,xx}(x) - 4u(x,t)f(x) + 2u_{,x}(x,t)\int_x^{+\infty} f(y)\,dy \qquad [34a]$$

$$\Lambda f(x) = f_{,xx}(x) - 2[\tilde{u}(x,t)+u(x,t)]f(x) + \Gamma\int_x^{+\infty} f(y)\,dy \qquad [34b]$$

$$\Gamma f(x) = [\tilde{u}_{,x}(x,t)+u_{,x}(x,t)]f(x) + [\tilde{u}(x,t)-u(x,t)] \times \int_x^{+\infty}[\tilde{u}(y,t)-u(y,t)]f(y)\,dy \qquad [34c]$$

In the limit when $\tilde{u}\to u$ the operator $\Lambda\to\mathcal{L}$. A BT is obtained by choosing the functions $F$ and $G$ in eqn [33c]. The simplest BT is obtained by setting $F=\sigma$ and $G=1$:

$$\tilde{v}_{,x} + v_{,x} + (\tilde{v}-v)\left[\sigma - \tfrac{1}{2}(\tilde{v}-v)\right] = 0 \qquad [35]$$

with $u(x,t) = -v_{,x}(x,t)$ and $\sigma$ is the Bäcklund parameter. By combining together BT of the form [35] with different parameters as in eqn [5], we get the permutability theorem for the KdV BTs:

$$\tilde{v}' = v - \frac{(\sigma_1+\sigma_2)[v'-\tilde{v}]}{\sigma_1-\sigma_2+(1/2)(v'-\tilde{v})} \qquad [36]$$

Its proof is immediate from the point of view of the spectrum.

## Bäcklund and Symmetries

A symmetry of the nonlinear equation [15] is given by a flow commuting with it, that is, by an equation

$$u_{,\epsilon} = f(u,u_x,u_t,\ldots) \qquad [37]$$

where $\epsilon$ is the group parameter, $u=u(x,t;\epsilon)$, and the $\epsilon$ derivative of [15] is zero on its set of solutions. A group transformation is obtained by integrating it. Usually this is possible only when eqn [37] is a quasilinear partial differential equation of the first order. Taking into account the evolution of the spectrum of the KdV equation [15], it is easy to prove that its symmetries are given by

$$u_{,\epsilon} = \left\{\sum_{n=0}^{+\infty}\alpha_n\mathcal{L}^n - 3\sum_{n=0}^{+\infty}\beta_n t\mathcal{L}^n\right\}u_{,x} + \left\{\sum_{n=0}^{+\infty}\beta_n\mathcal{L}^n\right\}[xu_{,x}+2u] \qquad [38]$$

where $\alpha_n$ and $\beta_n$ are a set of constant parameters. For each choice of the parameters $\alpha_n$ and $\beta_n$, one gets a symmetry of the KdV equation [15]. With eqn [38] one can associate the following evolution of the reflection coefficient $R(k,t;\epsilon)$:

$$\frac{dR}{d\epsilon} = 2ik\left\{\sum_{n=0}^{+\infty}\alpha_n(-4k^2)^n - 3\sum_{n=0}^{+\infty}\beta_n t(-4k^2)^{n+1}\right\}R \qquad [39]$$

and of the spectral parameter $k$

$$k_{,\epsilon} = \sum_{n=0}^{+\infty}\beta_n(-4k^2)^n k \qquad [40]$$

As $-(1/2)\mathcal{L}\ 1 = xu_{,x}+2u$, one can add to the symmetries [38] the exceptional one (which has no spectral counterpart as $u$ is not bounded asymptotically):

$$u_{,\epsilon} = 1 + 6tu_{,x} \qquad [41]$$

By a proper natural choice of the constant parameters $\alpha_n$ and $\beta_n$, one can define two infinite series of symmetries. The first one is obtained by choosing $\beta_n=0$ and $\alpha_n=\delta_{n,m}$ with $m=1,2,\ldots,\infty$ and can be denoted as the isospectral series as $k_{,\epsilon}=0$. This is formed by commuting symmetries. The second one is given by $\alpha_n=0$ and $\beta_n=\delta_{n,m}$ with $m=1,2,\ldots,\infty$ and can be denoted as the nonisospectral series as $k_{,\epsilon}\neq 0$. The nonisospectral symmetries have a nonzero commutation relation among themselves and with the isospectral ones.

Except for a few Lie point symmetries (given by eqn [41] and by choosing inside the series [38] those with different from zero only $\beta_0$ or $\alpha_0$ or $\alpha_1$) they are all generalized symmetries (Olver 1993). By analyzing their spectrum, it is easy to prove that the choice [38] is such that they are all independent. For the isospectral class, the evolution of the spectrum is simple and can be integrated to provide the group transformation of the spectrum

$$R(k,t;\epsilon) = R(k,t) \times \exp\left[2\mathrm{i}k\left\{\sum_{n=0}^{+\infty}\alpha_n(-4k^2)^n\right\}\epsilon\right] \qquad [42]$$

Let us now consider the simplest BT obtained by choosing, in eqn [33c], $F(\Lambda)=\sigma$ and $G(\Lambda)=1$, where $\sigma$ is an arbitrary parameter. In the spectral space, this corresponds to the following change of the spectrum:

$$\tilde{R}(k,t) = \frac{\sigma - 2\mathrm{i}k}{\sigma + 2\mathrm{i}k}R(k,t) \qquad [43]$$

Defining $\tilde{R}(k,t)=R(k,t;\epsilon)$, eqn [42] is equal to eqn [43] iff

$$\alpha_n = -\frac{2}{\epsilon\sigma^{2n+1}(2n+1)}, \quad n = 0,1,\ldots,\infty \qquad [44]$$

So we need an infinite number of symmetries to be able to reconstruct the change of the spectrum given by the BT. This shows that the existence of a BT is strictly connected to the existence of an infinity of symmetries which is a condition for the exact integrability of the nonlinear partial differential equation (Fokas 1980, Ibragimov and Shabat 1980).

## Discretization via Bäcklund

BTs, apart from providing classes of exact solutions to nonlinear equations, play a very important role in the discretization of partial differential equations. As noted earlier, an auto-BT is a differential relation between two different solutions of the same nonlinear partial differential equation. If it is assumed that the new solution $\tilde{u}$ is just the old solution $u$ computed in a different point of a lattice, then the BT becomes just a differential–difference equation (Chiu and Ladik 1977, Levi and Benguria 1980). This can be carried out also at the level of the associated compatibility condition and in such a way one is able to also obtain its Lax pair. This demonstrates the integrability of the differential–difference equation

$$v(n+1,t)_{,t} + v(n,t)_{,t} + [v(n+1,t) - v(n,t)] \times \left\{\sigma - \tfrac{1}{2}[v(n+1,t) - v(n,t)]\right\} = 0 \qquad [45]$$

which is an integrable differential–difference approximation to the KdV equation or

$$w(n+1,t)_{,t} = w(n,t)_{,t} + 2a\sin\left[\frac{w(n+1,t)+w(n,t)}{2}\right] \qquad [46]$$

a discrete integrable differential–difference approximation to the sG equation (Hirota 1977, Orfanidis 1978).

As the nonlinear superposition formulas are purely algebraic relations involving potentials associated with integrable nonlinear partial differential equations, one can interpret them as difference–difference equations. In the case of the sG equation from eqn [7], we have

$$w_{n+1,m+1} - w_{n,m} = 4\arctan^{-1}\left(\frac{a_1+a_2}{a_1-a_2}\tan\frac{w_{n,m+1}-w_{n+1,m}}{4}\right) \qquad [47]$$

where $w(x,t)=w_{n,m}$, $\tilde{w}(x,t)=w_{n+1,m}$, $w'(x,t)=w_{n,m+1}$, and $\tilde{w}'(x,t)=w_{n+1,m+1}$. In a similar manner, from [36], one gets

$$v_{n+1,m+1} = v_{n,m} - \frac{(\sigma_1+\sigma_2)[v_{n+1,m} - v_{n,m+1}]}{\sigma_1 - \sigma_2 + \frac{1}{2}[v_{n+1,m} - v_{n,m+1}]} \qquad [48]$$

The continuous limit of eqn [47], obtained by setting $x=\epsilon_1 n$ and $y=\epsilon_2 m$ and choosing

$$\frac{a_1}{a_2} = \frac{\epsilon_1\epsilon_2}{4}$$

gives back eqn [2] (Rogers and Schief 1997). It is worth mentioning that one can also use known nonlinear lattice equations to construct BT for nonlinear partial differential equations (Levi 1981).

*See also:* Integrable Systems and Discrete Geometry; Integrable Systems: Overview; Painlevé Equations; Solitons and Kac–Moody Lie Algebras; Toda Lattices.

## Further Reading

Bäcklund AV (1874) Einiges über Curven und Flächentransformationen. *Lund Universitëts Arsskrift* 10: 1–12.

Bianchi L (1879) Ricerche sulle superficie a curvatura costante e sulle elicoidi. *Annali della R. Scuola normale superiore di Pisa* 2: 285.

Bruschi M and Ragnisco O (1980a) Existence of a Lax pair for any member of the class of nonlinear evolution equations associated to the matrix Schrödinger spectral problem. *Lettere al Nuovo Cimento* 29: 321–326.

Bruschi M and Ragnisco O (1980b) Extension of the Lax method to solve a class of nonlinear evolution equations with $x$-dependent coeffcients associated to the matrix Schrödinger spectral problem. *Lettere al Nuovo Cimento* 29: 327–330.

Bruschi M and Ragnisco O (1980c) Bäcklund transformations and Lax technique. *Lettere al Nuovo Cimento* 29: 331–334.

Chiu S-C and Ladik JF (1977) Generating exactly soluble nonlinear discrete evolution equations by a generalized

Wronskian technique. *Journal of Mathematical Physics* 18: 690–700.
Clarin J (1903) Sur quelques équations aux dérivées partielles du second ordre. *Annales de la Facult des Sciences de Toulouse pour les Sciences Mathmatiques et les Sciences Physiques. Serie 2* 5: 437–458.
Coley A, Levi D, Milson R, Rogers C, and Winternitz P (eds.) (2001) Bäcklund and Darboux transformations. The Geometry of solitons. *Proceedings of the AARMS-CRM Workshop, Halifax, NS, June 4–9, 1999*. CRM Proceedings and Lecture Notes, vol. 29. Providence, RI: American Mathematical Society.
Faddeev LD and Takhtajan LA (1987) *Hamiltonian Methods in the Theory of Solitons*. Berlin: Springer.
Fokas AS (1980) A symmetry approach to exactly solvable evolution equations. *Journal of Mathematical Physics* 21: 1318–1325.
Gardner CS, Greene JM, Kruskal MD, and Miura RM (1967) Method for solving the Korteweg–de Vries equation. *Physical Review Letters* 19: 1095–1097.
Hirota R (1978) Nonlinear partial difference equations. III. Discrete sine-Gordon equation. *Journal of the Physical Society of Japan* 43: 2079–2086.
Ibragimov NH and Shabat AB (1980) Infinite Lie–Bcklund algebras (in Russian). *Funktsional. Anal. i Prilozhen* 14: 79–80.
Lamb GL (1976) Bäcklund transformations at the turn of the century. In: Miura RM (ed.) *Bäcklund Transformations*, pp. 69–79. Berlin: Springer.
Lax PD (1968) Integrals of nonlinear equations of evolution and solitary waves. *Communications in Pure and Applied Mathematics* 21: 647–690.
Levi D (1981) Nonlinear differential difference equations as Backlund transformations. *Journal of Physics A: Mathematical and General* 14: 1083–1098.
Levi D (1988) On a new Darboux transformation for the construction of exact solutions of the Schroedinger equation. *Inverse Problems* 4: 165–172.
Levi D and Benguria R (1980) Bäcklund transformations and nonlinear differential difference equations. *Proceedings of the National Academy of Science USA* 77: 5025–5027.
Levi D and Winternitz P (1989) Non-classical symmetry reduction: example of the Boussinesq equation. *Journal of Physics A: Mathematical and General* 22: 2915–2924.
Levi D, Ragnisco O, and Sym A (1984) Dressing method vs. classical Darboux transformation. *Il Nuovo Cimento* 83B: 34–42.
Lie S (1888, 1890, 1893) *Theorie der Transformationgruppen*. Leipzig: B.G. Teubner.
Matveev VB and Salle LA (1991) *Darboux Transformations and Solitons*. Berlin: Springer.
Moutard Th-F (1878) Sur la construction des équations de la forme $(1/z)(\mathrm{d}^2z/\mathrm{d}x\mathrm{d}y)=\lambda(x,y)$, qui admettent une integrale général explicite. *Journal de l'Ecole Polytechnique, Paris* 28: 1–11.
Olver PJ (1993) *Applications of Lie Groups to Differential Equations*. New York: Springer.
Orfanidis SJ (1978) Discrete sine-Gordon equations. *Physical Review D* 18: 3822–3827.
Rogers C and Schief WK (1997) The classical Bäcklund transformation and integrable discretization of characteristic equations. *Physics Letters A* 232: 217–223.
Rogers C and Shadwick WF (1982) *Bäcklund Transformations and Their Applications*. New York: Academic Press.
Toda M (1989) *Theory of Nonlinear Lattices*. Berlin: Springer.
Wojciechowski S (1982) The analogue of the Bäcklund transformation for integrable many-body systems. *Journal of Physics A: Mathematical and General* 15: L653–L657.
Zaharov VE and Shabat AB (1979) Integration of the nonlinear equations of mathematical physics by the method of the inverse scattering problem. II (Russian). *Funktsional Analiz i ego Prilozheniya* 13: 13–22. (English translation: *Functional Analysis and Applications* 13: 166–173 (1980)).

# Batalin–Vilkovisky Quantization

**A C Hirshfeld**, Universität Dortmund, Dortmund, Germany



## Introduction

The Batalin–Vilkovisky formalism for quantizing gauge theories has a long history of development. It begins with the Faddeev–Popov procedure for quantizing Yang–Mills theory, involving the Faddeev–Popov ghost fields (Faddeev and Popov 1967). It continued with the discovery of BRST symmetry by Becchi *et al.* (1976). Then Zinn-Justin (1975) introduced sources for these transformations, and a symmetric structure in the space of fields and sources in his study of renormalizability of these theories. Finally, Batalin and Vilkovisky (1981) systematized and generalized these developments. A more detailed account of this history can be found in Gomis *et al.* (1994), where many worked examples of the Batalin–Vilkovisky formalism are given. At the present time, it is the most general treatment available. Alexandrov, Kontsevich, Schwarz, and Zabaronsky (AKSZ 1997) have presented a geometric interpretation for the case in which the action is topologically invariant.

## Structure of the Set of Gauge Transformations

Consider a system whose dynamics is governed by a classical action $S[\phi^i]$ which depends on the fields $\phi^i(x), i=1,\ldots,n$. We employ a compact notation in which the multi-index $i$ may denote the various fields involved, the discrete indices on which they depend, and the dependence on the spacetime variables as well. The generalized summation convention then means that a repeated index may denote not only a sum over discrete variables, but also integration over the spacetime variables. $\epsilon_i=\epsilon(\phi^i)$ denotes the

Grassmann parity of the fields. Fields with $\epsilon_i = 0$ are called bosonic, with $\epsilon_i = 1$ fermionic. The graded commutation rule is

$$\phi^i(x)\phi^j(y) = (-1)^{\epsilon_i\epsilon_j}\phi^j(y)\phi^i(x) \qquad [1]$$

For a gauge theory the action is invariant under a set of gauge transformations with infinitesimal form

$$\delta\phi^i = R^i_\alpha \varepsilon^\alpha, \quad \alpha = 1 \text{ or } 2 \text{ or} \ldots m \qquad [2]$$

The $\varepsilon^\alpha$ are the infinitesimal gauge parameters and $R^i_\alpha$ the generators of the gauge transformations. When $\epsilon_\alpha = \epsilon(\varepsilon^\alpha) = 0$ we have an ordinary symmetry, when $\epsilon_\alpha = 1$ the equation is characteristic of a supersymmetry. The Grassmann parity of $R^i_\alpha$ is $\epsilon(R^i_\alpha) = \epsilon_i + \epsilon_\alpha \pmod 2$.

A subscript after a comma denotes the right derivative with respect to the corresponding field, that is, the field is to be commutated to the far right and then dropped. The field equations may then be written as

$$S_{0,i} = 0 \qquad [3]$$

where $S_0$ is the classical action. Let $\Sigma$ denote the surface in the space of solutions where the field equations are satisfied:

$$S_{0,i}|_\Sigma = 0 \qquad [4]$$

If the gauge transformations are "independent" on-shell, that is,

$$\operatorname{rank} R^i_\alpha|_\Sigma = m \qquad [5]$$

the gauge theory is said to be "irreducible." We assume here that this is the case. When it is not, the theory is "reducible." For details of the treatment in that case, see Gomis, Paris, and Samuel. The classical solutions are $\phi_0 \in \Sigma$.

The Noether identities are

$$S_{0,i}R^i_\alpha = 0 \qquad [6]$$

The general solution to the Noether identity is

$$\lambda^i = R^i_\alpha T^\alpha + S_{0,j}E^{ji} \qquad [7]$$

The commutator of two gauge transformations is

$$[\delta_1, \delta_2]\phi^i = \left(R^i_{\alpha,j}R^j_\beta - (-1)^{\epsilon_\alpha\epsilon_\beta}R^i_{\beta,j}R^j_\alpha\right)\varepsilon^\beta_1\varepsilon^\alpha_2 \qquad [8]$$

Since this commutator is a symmetry of the action, it satisfies the Noether identity

$$S_{0,i}\left(R^i_{\alpha,j}R^j_\beta - (-1)^{\epsilon_\alpha\epsilon_\beta}R^i_{\beta,j}R^j_\alpha\right) = 0 \qquad [9]$$

which by eqn [7] implies that

$$R^i_{\alpha,j}R^j_\beta - (-1)^{\epsilon_\alpha\epsilon_\beta}R^i_{\beta,j}R^j_\alpha = R^i_\gamma T^\gamma_{\alpha\beta} + S_{0,j}E^{ji}_{\alpha\beta} \qquad [10]$$

Equations [8] and [10] lead to the following condition:

$$[\delta_1, \delta_2]\phi^i = \left(R^i_\gamma T^\gamma_{\alpha\beta} - S_{0,j}E^{ji}_{\alpha\beta}\right)\varepsilon^\beta_1\varepsilon^\alpha_2 \qquad [11]$$

The tensors $T^\gamma_{\alpha\beta}$ are called the structure constants of the gauge algebra, although they depend, in general, on the fields of the theory. When $E^{ij}_{\alpha\beta} = 0$, the gauge algebra is said to be "closed," otherwise it is "open." Equation [11] defines a Lie algebra if the algebra is closed and the $T^\gamma_{\alpha\beta}$ are independent of the fields.

The gauge tensors have the following graded symmetry properties:

$$\begin{aligned} T^\gamma_{\alpha\beta} &= -(-1)^{\epsilon_\alpha\epsilon_\beta}T^\gamma_{\beta\alpha} \\ E^{ij}_{\alpha\beta} &= -(-1)^{\epsilon_\alpha\epsilon_\beta}E^{ji}_{\alpha\beta} = -(-1)^{\epsilon_\alpha\epsilon_\beta}E^{ij}_{\beta\alpha} \end{aligned} \qquad [12]$$

The Grassmann parities are

$$\epsilon(T^\gamma_{\alpha\beta}) = \epsilon_\alpha + \epsilon_\beta + \epsilon_\gamma \pmod 2 \qquad [13]$$

and

$$\epsilon(E^{ij}_{\alpha\beta}) = \epsilon_i + \epsilon_j + \epsilon_\alpha + \epsilon_\beta \pmod 2 \qquad [14]$$

Various restrictions are imposed by the Jacobi identity

$$\sum_{\text{cyclic}(123)} [\delta_1, [\delta_2, \delta_3]] = 0 \qquad [15]$$

These restrictions are

$$\sum_{\text{cyclic}(123)} \left(R^i_\delta A^\delta_{\alpha\beta\gamma} - S_{0,j}B^{ji}_{\alpha\beta\gamma}\right)\varepsilon^\gamma\varepsilon^\beta\varepsilon^\alpha = 0 \qquad [16]$$

where

$$\begin{aligned} 3A^\delta_{\alpha\beta\gamma} \equiv & \left(T^\delta_{\alpha\beta k}R^k_\gamma - T^\delta_{\alpha\eta}T^\eta_{\beta\gamma}\right) + (-1)^{\epsilon_\alpha(\epsilon_\beta+\epsilon_\gamma)} \\ & \times\left(T^\delta_{\beta\gamma k}R^k_\gamma - T^\delta_{\beta\eta}T^\eta_{\gamma\alpha}\right) \\ & + (-1)^{\epsilon_\gamma(\epsilon_\alpha+\epsilon_\beta)}\left(T^\delta_{\gamma\alpha k}R^k_\beta - T^\delta_{\gamma\eta}T^\eta_{\alpha\beta}\right) \end{aligned}$$

and

$$\begin{aligned} 3B^{ji}_{\alpha\beta\gamma} \equiv & \Big(E^{ji}_{\alpha\beta k}R^k_\alpha - E^{ji}_{\alpha\delta}T^\delta_{\beta\gamma} - (-1)^{\epsilon_i\epsilon_\alpha} \\ & \times R^j_{\alpha,k}E^{ki}_{\beta\gamma} + (-1)^{\epsilon_j(\epsilon_i+\epsilon_\alpha)}R^i_{\alpha,k}E^{kj}_{\beta\gamma}\Big) \\ & + (-1)^{\epsilon_\alpha(\epsilon_\beta+\epsilon_\gamma)}(\alpha \to \beta \to \gamma) + (-1)^{\epsilon_\gamma(\epsilon_\alpha+\epsilon_\beta)} \\ & \times (\alpha \to \gamma \to \beta) \end{aligned}$$

As in the familiar Faddeev–Popov procedure, it is useful to introduce ghost fields $C^\alpha$ with opposite Grassmann parities to the gauge parameters $\varepsilon^\alpha$:

$$\epsilon(C^\alpha) = \epsilon_\alpha + 1 \pmod 2 \qquad [17]$$

and to replace the gauge parameters by ghost fields. One must then modify the graded symmetry properties of the gauge structure tensors according to

$$T_{\alpha_1\alpha_2\alpha_3\alpha_4\ldots} \to (-1)^{\epsilon_{\alpha_2+\alpha_4+\cdots}} T_{\alpha_1\alpha_2\alpha_3\alpha_4\ldots} \qquad [18]$$

The Noether identities then take the form

$$S_{0,i} R^i_\alpha C^\alpha = 0 \qquad [19]$$

and the structure relations [10] become

$$(2R^i_{\alpha,j} R^j_\beta - R^i_\gamma T^\gamma_{\alpha\beta} + S_{0,j} E^{ji}_{\alpha\beta}) C^\beta C^\alpha = 0 \qquad [20]$$

## Introducing the Antifields

We incorporate the ghost fields into the field set $\Phi^A = \{\phi^i, C^\alpha\}$, where $i = 1, \ldots, n$ and $\alpha = 1, \ldots, m$. Clearly $A = 1, \ldots, N$, where $N = n + m$. One then further increases the set by introducing an antifield $\Phi^*_A$ for each field $\Phi^A$. The Grassmann parity of the antifields is

$$\epsilon\left(\Phi^*_A\right) = \epsilon(\Phi^A) + 1 \ (\mathrm{mod}\, 2) \qquad [21]$$

Each field is assigned a ghost number, with

$$\begin{aligned} \mathrm{gh}[\phi^i] &= 0 \\ \mathrm{gh}[C^\alpha] &= 1 \\ \mathrm{gh}\left[\Phi^*_A\right] &= -\mathrm{gh}[\Phi^A] - 1 \end{aligned} \qquad [22]$$

In the space of fields and antifields, the antibracket is defined by

$$(X, Y) = \frac{\partial_r X}{\partial \Phi^A} \frac{\partial_l Y}{\partial \Phi^*_A} - \frac{\partial_r X}{\partial \Phi^*_A} \frac{\partial_l Y}{\partial \Phi^A} \qquad [23]$$

where $\partial_r$ denotes the right, $\partial_l$ the left derivative. The antibracket is graded antisymmetric:

$$(X, Y) = -(-1)^{(\epsilon_X+1)(\epsilon_Y+1)} (Y, X) \qquad [24]$$

It satisfies a graded Jacobi identity

$$((X, Y), Z) + (-1)^{(\epsilon_X+1)(\epsilon_Y+1)} \times ((Y, Z), X) + (-1)^{(\epsilon_Z+1)(\epsilon_X+\epsilon_Y)} ((Z, X), Y) = 0 \qquad [25]$$

It is a graded derivation

$$\begin{aligned} (X, YZ) &= (X, Y)Z + (-1)^{\epsilon_X \epsilon_Y} (X, Z) Y \\ (XY, Z) &= X(Y, Z) + (-1)^{\epsilon_X \epsilon_Y} Y(X, Z) \end{aligned} \qquad [26]$$

It has ghost number

$$\mathrm{gh}[(X, Y)] = \mathrm{gh}[X] + \mathrm{gh}[Y] + 1 \qquad [27]$$

and Grassmann parity

$$\epsilon((X, Y)) = \epsilon(X) + \epsilon(Y) + 1 \ (\mathrm{mod}\, 2) \qquad [28]$$

For bosonic fields

$$(B, B) = 2 \frac{\partial B}{\partial \Phi^A} \frac{\partial B}{\partial \Phi^*_A} \qquad [29]$$

for fermionic fields

$$(F, F) = 0 \qquad [30]$$

and for any $X$

$$((X, X), X) = 0 \qquad [31]$$

If one groups the fields and the antifields together into the set

$$z^a = \{\Phi^A, \Phi^*_A\}, \quad a = 1, \ldots, 2N \qquad [32]$$

then the antibracket is seen to define a symplectic structure on the space of fields and antifields

$$(X, Y) = \frac{\partial_r X}{\partial z^a} \omega^{ab} \frac{\partial_l Y}{\partial z^b} \qquad [33]$$

with

$$\omega^{ab} = \begin{pmatrix} 0 & \delta^A_B \\ -\delta^A_B & 0 \end{pmatrix} \qquad [34]$$

The antifields can be thought of as conjugate variables to the fields, since

$$\left(\Phi^A, \Phi^*_B\right) = \delta^A_B \qquad [35]$$

## The Classical Master Equation

Let $S[\Phi^A, \Phi^*_A]$ be a functional of the fields and antifields with the dimension of an action, vanishing ghost number and even Grassmann parity. The equation

$$(S, S) = 2 \frac{\partial S}{\partial \Phi^A} \frac{\partial S}{\partial \Phi^*_A} = 0 \qquad [36]$$

is the classical master equation. Solutions of the classical master equation with suitable boundary conditions turn out to be generating functionals for the gauge structure of the theory. $S$ is also the starting point for the quantization. One denotes by $\Sigma$ the subspace of stationary points of the action in the space of fields and antifields:

$$\Sigma = \left\{ z^a \left| \frac{\partial S}{\partial z^a} = 0 \right. \right\} \qquad [37]$$

Given a classical solution $\phi_0$ of $S_0$ one stationary point is

$$\phi^i = \phi^i_0, \quad C^a = 0, \quad \Phi^*_A = 0 \qquad [38]$$

An action which satisfies the classical master equation has its own set of invariances:

$$\frac{\partial S}{\partial z^a} R^a_b = 0 \qquad [39]$$

with

$$R^a_b = \omega^{ac} \frac{\partial_l \partial_r S}{\partial z^c \partial z^b} \qquad [40]$$

This equation implies

$$R^a_c R^a_b \big|_\Sigma = 0 \qquad [41]$$

One says that $R^a_b$ is invariant on-shell. A nilpotent $2N \times 2N$ matrix has rank $\leq N$. Let $r$ be the rank of the hessian of $S$ at the stationary point:

$$r = \operatorname{rank} \frac{\partial_l \partial_r S}{\partial z^a \partial z^b}\bigg|_\Sigma \qquad [42]$$

We then have $r \leq N$. The relevant solutions of the classical master equation are those for which $r = N$. In this case the number of independent gauge invariances of the type in eqn [39] equals the number of antifields. When at a later stage the gauge is fixed, the nonphysical antifields are eliminated.

To ensure the correct classical limit, the proper solution must contain the classical action $S_0$ in the sense that

$$S\left[\Phi^A, \Phi^*_A\right]\big|_{\Phi^*_A = 0} = S_0[\phi^i] \qquad [43]$$

The action $S[\Phi^A, \Phi^*_A]$ can be expanded in a series in the antifields, while maintaining vanishing ghost number and even Grassmann parity:

$$\begin{aligned} S[\Phi, \Phi^*] = {} & S_0 + \phi^*_i R^i_\alpha C^\alpha + C^*_a \tfrac{1}{2} T^\alpha_{\beta\gamma} (-1)^{\epsilon_\beta} C^\gamma C^\beta \\ & + \phi^*_i \phi^*_j (-1)^{\epsilon_i} \tfrac{1}{4} E^{ji}_{\alpha\beta} (-1)^{\epsilon_\alpha} C^\beta C^\alpha + \cdots \end{aligned} \qquad [44]$$

When this is inserted into the classical master equation, one finds that this equation implies the gauge structure of the classical theory.

## Gauge Fixing and Quantization

Equation [39] shows that the action $S$ still possesses gauge invariances, and hence is not yet suitable for quantization via the path integral approach: a gauge-fixing procedure is necessary. In the Batalin–Vilkovisky approach the gauge is fixed, and the antifields eliminated, by use of a gauge-fixing fermion $\Psi$ which has Grassmann parity $\epsilon(\Psi) = 1$ and $\text{gh}[\Psi] = -1$. It is a functional of the fields $\Phi^A$ only; its relation to the antifields is

$$\Phi^*_A = \frac{\partial \Psi}{\partial \Phi^A} \qquad [45]$$

We define a surface in functional space

$$\Sigma_\Psi = \left\{ \left(\Psi^A, \Psi^*_A\right) | \Psi^*_A = \frac{\partial \Psi}{\partial \Phi^A} \right\} \qquad [46]$$

so that for any functional $X[\Phi, \Phi^*]$

$$X|_{\Sigma_\Psi} = X\left[\Psi, \frac{\partial \Psi}{\partial \Phi}\right] \qquad [47]$$

To construct a gauge-fixing fermion $\Psi$ of ghost number –1, one must again introduce additional fields. The simplest choice utilizes a trivial pair $\bar{C}_\alpha, \bar{\pi}_\alpha$ with

$$\begin{aligned} \epsilon(\bar{C}_\alpha) &= \epsilon_\alpha + \mathbf{1}, & \epsilon(\bar{\pi}_\alpha) &= \epsilon_\alpha \\ \text{gh}[\bar{C}_\alpha] &= -1, & \text{gh}[\bar{\pi}_\alpha] &= 0 \end{aligned} \qquad [48]$$

The fields $\bar{C}_\alpha$ are the Faddeev–Popov antighosts. Along with these fields we include the corresponding antifields $\bar{C}^{*\alpha}, \bar{\pi}^{*\alpha}$. Adding the term $\bar{\pi}_\alpha \bar{C}^{*\alpha}$ to the action $S$ does not spoil its properties as a proper solution to the classical master equation, and one gets the nonminimal action

$$S^{\text{non}} = S + \bar{\pi}_\alpha \bar{C}^{*\alpha} \qquad [49]$$

The simplest possibility for $\Psi$ is

$$\Psi = \bar{C}_\alpha \chi^\alpha(\phi) \qquad [50]$$

where $\chi^\alpha$ are the gauge-fixing conditions for the fields $\psi$. The gauge-fixed action is denoted by

$$S_\Psi = S^{\text{non}}|_{\Sigma_\psi} \qquad [51]$$

Quantization is performed using the path integral to calculate a correlation function $X$, with the constraint [45] implemented by a $\delta$-function:

$$\begin{aligned} I_\Psi(X) = {} & \int D\Phi D\Phi^* \delta\left(\Phi^*_A - \frac{\partial \Psi}{\partial \Phi^A}\right) \\ & \times \exp\left(\frac{\mathrm{i}}{\hbar} W[\Phi, \Phi^*]\right) X[\Phi, \Phi^*] \end{aligned} \qquad [52]$$

Here $W$ is the quantum action, which reduces to $S$ in the limit $\hbar \to 0$. An admissible $\Psi$ leads to well-defined propagators when the path integral is expressed as a perturbation series expansion.

The results of a calculation should be independent of the gauge fixing. Consider the integrand in eqn [52],

$$I[\Phi, \Phi^*] = \exp\left(\frac{\mathrm{i}}{\hbar} W[\Phi, \Phi^*]\right) X[\Phi, \Phi^*] \qquad [53]$$

Under an infinitesimal change in $\Psi$

$$I_{\Psi + \delta\Psi}(X) - I_\Psi(X) \approx \int D\Phi \, \Delta I \delta\Psi \qquad [54]$$

where the Laplacian $\Delta$ is

$$\Delta = (-1)^{\epsilon_A+1} \frac{\partial}{\partial \Phi^A} \frac{\partial}{\partial \Phi^*_A} \quad [55]$$

Obviously, the integral $I_\Psi(X)$ is independent of $\Psi$ if $\Delta I = 0$. For $X = 1$ one gets the requirement

$$\Delta \exp\left(\frac{i}{\hbar} W\right) = \exp\left(\frac{i}{\hbar} W\right) \times \left(\frac{i}{\hbar}\Delta W - \frac{1}{2\hbar^2}(W, W)\right) = 0 \quad [56]$$

The formula

$$\tfrac{1}{2}(W, W) = i\hbar \Delta W \quad [57]$$

is the quantum master equation. A gauge-invariant correlation function satisfies

$$(X, W) = i\hbar \Delta X \quad [58]$$

The terms of higher order in $\hbar$ by which the quantum action $W$ may differ from the solution of the classical master equation $S$ correspond to the counter-terms of the renormalizable gauge theory if

$$\Delta S = 0 \quad [59]$$

One must, of course, use a regularization scheme which respects the symmetries of the theory. For $W = S + O(\hbar)$ the quantum master equation [57] reduces in this case to the classical master equation

$$(S, S) = 0 \quad [60]$$

Hence, up to possible counter-terms, one may simply choose $W = S$.

To implement the gauge fixing, one uses for the action $W = S^{\text{non}}$. For the path integral $Z = I_\Psi(X = 1)$, the integration over the antifields in eqn [52] is performed by using the $\delta$-function. The result is

$$Z = \int D\Phi \exp\left(\frac{i}{\hbar} S_\Psi\right) \quad [61]$$

## Geometrical Interpretation of Topological Field Theories

The Batalin–Vilkovisky formalism for topological field theories has been given a geometrical interpretation by AKSZ (1997).

A supermanifold equipped with an odd vector field satisfying $Q^2 = 0$ is called a $Q$-manifold. A $Q$-manifold provided with an odd symplectic structure $\omega$ (P-structure) is called a QP-manifold if the odd symplectic structure is $Q$-invariant, that is, $L_Q \omega = 0$. Every solution to the classical master equation determines a QP-structure on $M$ and vice versa. The geometric object corresponding to a classical mechanical system in the Batalin–Vilkovisky formalism is a QP-manifold.

The nondegenerate closed 2-form $\omega$ is written as

$$\omega = dz^a \zeta_{ab} dz^b \quad [62]$$

where $z^a$ are local coordinates in the supermanifold $M$. For functions on $M$, an (odd) Poisson bracket is defined as in eqn [33], where $\omega^{ab}$ stands for the inverse matrix of $\omega_{ab}$. An even function $S$ on $M$ satisfies the classical master equation if $(S, S) = 0$. The correspondence between vector fields and functions on $M$ is given by $K_F G = (G, F)$, where $K_F$ is the vector field, $F$ the given function, and $G$ an arbitrary function. The function $F$ is called the Hamiltonian of the vector field $K_F$.

Geometrically, equivalent QP-manifolds describe the same physics. In particular, one can consider an even Hamiltonian vector field $K_F$ corresponding to an odd function $F$. This vector field determines an infinitesimal transformation preserving P-structure. It transforms a solution $S$ to the classical master equation into the physically equivalent solution $S + \epsilon(S, F)$, where $\epsilon$ is an infinitesimally small parameter.

A submanifold $L$ of a P-manifold $M$ is called a Lagrangian submanifold if the restriction of the form $\omega$ to $L$ vanishes. In the particular case when $M = \Pi T^*N$ (the cotangent bundle to $N$ with reversed parity of fibres) with standard P-structure, one can construct many examples of Lagrangian submanifolds in the following way. Fix an odd function $\Psi$ on $N$, the gauge fermion. The submanifold $L_\Psi \in M$ determined by the equation

$$\xi^a = \frac{\partial \Psi}{\partial x^a} \quad [63]$$

where $\{x^a, \xi_a\}$ are coordinates corresponding to the identification of $M$, will be a Lagrangian submanifold of $M$.

The P-manifold M in the neighborhood of $L$ can be identified with $\Pi T^*L$. In other words, one can find such a neighborhood $U$ of $L$ in $M$ and a neighborhood $V$ of $L$ in $\Pi T^*L$ that there exists an isomorphism of P-manifolds $U$ and $V$ leaving $L$ intact. Using this isomorphism a function $\Psi$ defined on a Lagrangian submanifold $L \subset M$ determines another Lagrangian submanifold $L_\Psi \subset M$.

Consider a solution $S$ to the classical master equation on $M$. In the Batalin–Vilkovisky formalism we have to restrict $S$ to a Lagrangian submanifold $L \in M$, then the quantization of $S$ can be performed by integration of $\exp(iS/\hbar)$ over $L$. One may construct an odd vector field $Q$ on $L$ in such a

way that the functional $S$ restricted to $L$ is $Q$-invariant. This invariance is BRST invariance.

AKSZ apply these geometric constructions to obtain in a natural way the action functionals of two-dimensional sigma-models (Witten 1998) and to show that the Chern–Simons theory (Axelrod and Singer 1991) in Batalin–Vilkovisky formalism arises as a sigma-model with target space $\Pi\mathcal{G}$, where $\mathcal{G}$ stands for a Lie algebra and $\Pi$ denotes parity inversion.

## The Poisson-Sigma Model

The quantization of the Poisson-sigma model was performed by Hirshfeld and Schwarzweller (2000) and by Cattaneo and Felder (2001). The Poisson-sigma model is the simplest topological field theory in two dimensions. It is a field theory on a two-dimensional world sheet without boundary (Schaller and Strobl 1994). It involves a set of bosonic scalar fields, which can be seen as a set of maps $X^i : M \to N$, where $N$ is a Poisson manifold. In addition, one has a 1-form $A$ on the world sheet $M$ which takes values in $T^*(N)$, for $x$ coordinates on $M$ we have $A = A_{\mu i}\mathrm{d}x^i \wedge \mathrm{d}X^i$. Its action is

$$S_0[X,A] = \int_M \mu\big(\epsilon^{\mu\nu}(A_{\mu i}\partial_\nu X^i + P^{ij}(X)A_{\mu i}A_{\nu j}\big) \quad [64]$$

where $\epsilon^{\mu\nu}$ is the antisymmetric tensor and $\mu$ is the volume form on $M$. The gauge transformations of the model are

$$\delta X^i = P^{ij}(X)\varepsilon_j, \qquad \delta A_{\mu i} = D^j_{\mu i}\varepsilon_j \quad [65]$$

where $D^j_{\mu i} = \partial_\mu \delta^j_i + P^{kj}{}_{,i}A_{\mu k}$. The equations of motion are

$$\epsilon^{\mu\nu} D^j_{\mu i} A_{\nu j} = 0 \quad [66]$$

and

$$\epsilon^{\mu\nu}(\partial_\nu X^i + P^{ij}A_{\nu j}) = \epsilon^{\mu\nu}D_\nu X^i = 0 \quad [67]$$

The gauge algebra is given by

$$\begin{aligned}
[\delta(\varepsilon_1), \delta(\varepsilon_2)]X^i &= P^{ji}(P^{mn}{}_{,j}\varepsilon_{1n}\varepsilon_{2m}) \\
[\delta(\varepsilon_1), \delta(\varepsilon_2)]A_{\mu i} &= D^j_{\mu i}(P^{mn}{}_{,j}\varepsilon_{1n}\varepsilon_{2m}) \\
&\quad - (\epsilon^{\nu\rho}D_\rho X^j)\epsilon_{\nu\mu}P^{mn}{}_{,ji}\varepsilon_{1n}\varepsilon_{2m}
\end{aligned} \quad [68]$$

In our general notation the generators of the gauge transformations $R$ are here $P^{ij}$ and $D^j_{\mu i}$. The gauge tensors $T$ and $E$ are $P^{ij}{}_{,k}$ and $\epsilon_{\nu\mu}P^{mn}{}_{,ji}$. The higher-order gauge tensors $A$ and $B$ vanish.

The ghost fields are again denoted by $C^i$. The Noether identities are then

$$\int_M \mu\Big(\epsilon^{\mu\nu}D^j_{\mu i}A_{\nu j}P^{ki} + (\epsilon^{\mu\nu}D_\nu X^i)D^k_{\mu i}\Big)C_k = 0 \quad [69]$$

Considering the commutator of two gauge transformations leads to (see eqns [8]–[11])

$$\begin{aligned}
&\int_M \mu\big(2P^{mi}{}_{,j}P^{nj} - P^{ji}P^{mn}{}_{,j}\big)C_mC_n = 0 \\
&\int_M \mu\Big(2(P^{jk}{}_{,i}D^l_{\mu j} + P^{mk}{}_{,ij}A_{\mu m}P^{jl}) \\
&\quad - D^m_{\mu i}P^{kl}{}_{,m} + (\epsilon^{\rho\nu}D_\rho X^i)\epsilon_{\mu\nu}P^{kl}{}_{,ji}\Big)C_lC_k = 0
\end{aligned} \quad [70]$$

The Jacobi identity is

$$P^{ij}{}_{,m}P^{mk}C_iC_jC_k = 0 \quad [71]$$

The fields and antifields of the model are

$$\Phi^A = \{A^{\mu i}, X^i, C_i\} \quad \text{and} \quad \Phi^*_A = \{A^{\mu i *}, X^*_i, C^{i*}\} \quad [72]$$

The extended action is

$$\begin{aligned}
\mathcal{S} = \int_M \mu\Big(&\epsilon^{\mu\nu}(A_{\mu i}\partial_\nu X^i + P^{ij}(X)A_{\mu i}A_{\nu j}) \\
&+ A^{\mu i*}D^j_{\mu i}C_j + X^*_i P^{ji}(X)C_j + \frac{1}{2}C^{i*}P^{jk}{}_{,i}(X)C_jC_k \\
&+ \frac{1}{4}A^{\mu i*}A^{\nu j*}\epsilon_{\mu\nu}P^{kl}{}_{,ij}(X)C_kC_l\Big)
\end{aligned} \quad [73]$$

The gauge-fixing conditions are taken to be of the form $\chi_i(A,X)$, so that the gauge fermion [50] becomes $\Psi = \bar{C}^i\chi_i(A,X)$. The antifields are then fixed to be

$$\begin{aligned}
A^*_{\mu i} &= \bar{C}_j \frac{\partial \chi_j(A,X)}{\partial A_{\mu i}} \\
X^*_i &= \bar{C}_j \frac{\partial \chi_j(A,X)}{\partial X^i} \\
C^*_i &= 0 \\
\bar{C}^*_i &= \chi_i(A,X)
\end{aligned} \quad [74]$$

The gauge-fixed action is

$$\begin{aligned}
S_\Psi = \int_M \mu\Big(&\epsilon^{\mu\nu}(A_{\mu i}\partial_\nu X^i + P^{ij}(X)A_{\mu i}A_{\nu j}) \\
&+ \bar{C}^k \frac{\partial \chi_k(A,X)}{\partial A_{\mu i}}D^j_{\mu i}C_j + \bar{C}^k\frac{\partial\chi_k(A,X)}{\partial X^i}P^{ij}C_j \\
&+ \frac{1}{4}\bar{C}^m\frac{\partial\chi_m(A,X)}{\partial A_{\mu i}}\bar{C}^n\frac{\partial\chi_n(A,X)}{\partial A_{\nu j}}\epsilon_{\mu\nu}P^{kl}{}_{,ij}(X) \\
&\times C_kC_l + \bar{\pi}^i\chi_i(A,X)\Big)
\end{aligned} \quad [75]$$

Now consider different gauge conditions:

1. First, the Landau gauge for the gauge potential $\chi_i = \partial^\mu A_{\mu i}$, so that the gauge fermion becomes $\Psi = \bar{C}^i\partial^\mu A_{\mu i}$. The antifields are fixed to be

$$\begin{aligned}
A^{*\mu i} &= \partial^\mu \bar{C}^i \\
X^*_i &= C^{*i} = 0 \\
\bar{C}^*_i &= \partial^\mu A_{\mu i}
\end{aligned} \quad [76]$$

for this gauge choice the gauge-fixed action is

$$S_\Psi = \int_M \mu\Big(\epsilon^{\mu\nu}(A_{\mu i}\partial_\nu X^i + P^{ij}(X)A_{\mu i}A_{\nu j}) + \bar{C}^i\partial^\mu D^j_{\mu i}C_j + \frac{1}{4}(\partial^\mu\bar{C}^i)(\partial^\nu\bar{C}^j)\epsilon_{\mu\nu}P^{kl}{}_{,ij}(X) \times C_k C_l - \bar{\pi}^i(\partial^\mu A_{\mu i})\Big) \quad [77]$$

Translating this action into the notation of Cattaneo and Felder, one sees that it is exactly the expression they use to derive the perturbation series.

2. Now consider the temporal gauge $\chi_i = A_{0i}$. The gauge fermion is given by $\Psi = \bar{C}^i A_{0i}$. The antifields are fixed to

$$A^{*0i} = \bar{C}^i \quad A^{*1i} = 0 \quad X^*_i = C^{*i} = 0 \quad \bar{C}^*_i = A_{0i} \quad [78]$$

The gauge-fixed action is

$$S_\Psi = \int_M \mu\Big(\epsilon^{\mu\nu}(A_{\mu i}\partial_\nu X^i + P^{ij}(X)A_{\mu i}A_{\nu j}) + \bar{C}^i D^j_{0i}C_j - \bar{\pi}^i(A_{0i})\Big) \quad [79]$$

3. Finally consider the Schwinger–Fock gauge $\chi_i = x^\mu A_{\mu i}$. Then the antifields are fixed to be

$$A^{*\mu i} = x^\mu\bar{C}^i \quad X^*_i = C^{*i} = 0 \quad \bar{C}^*_i = x^\mu A_{\mu i} \quad [80]$$

for this gauge choice the gauge-fixed action is

$$S_\Psi = \int_M \mu\Big(\epsilon^{\mu\nu}(A_{\mu i}\partial_\nu X^i + P^{ij}(X)A_{\mu i}A_{\nu j}) + \bar{C}^i x^\mu D^j_{\mu i}C_j - \bar{\pi}^i(\partial^\mu A_{\mu i})\Big) \quad [81]$$

Notice that in the noncovariant gauges 2 and 3 the action simplifies, in that the term which arose because of the nonclosed nature of the gauge algebra vanishes.

*See also:* BF Theories; BRST Quantization; Constrained Systems; Graded Poisson Algebras; Operads; Perturbative Renormalization Theory and BRST; Supermanifolds; Topological Sigma Models.

## Further Reading


Alexandrov M, Kontsevich M, Schwarz A, and Zaboronsky O (1997) Geometry of the Master Equation. *International Journal of Modern Physics* A12: 1405–1430.

Axelrod S and Singer IM (1991) *Chern–Simons Perturbation Theory*, Proceedings of the XXth Conference on Differential Geometric Methods in Physics, Baruch College/CUNY, NY. (hep-th/9110056).

Batalin IA and Vilkovisky GA (1977) Gauge algebra and quantization. *Physics Letters* 69B: 309–312.

Becchi C, Rouet A, and Stora R (1976) Renormalization of gauge theories. *Annals of Physics (NY)* 98: 287–321.

Cattaneo AS and Felder G (2001) On the AKSZ formulation of the Poisson–Sigma model. *Letters of Mathematical Physics* 56: 163–179.

Faddeev LD and Popov VN (1967) Feynman diagrams for the Yang–Mills field. *Physics Letters* 25B: 29–30.

Gomis J, Paris J, and Samuel S (1994) Antibracket Antifields and gauge-theory quantization. *Physics Reports* 269: 1–145.

Hirshfeld AC and Schwarzweller T (2000) Path integral quantization of the Poisson–Sigma model. *Annals of Physics (Leipzig)* 9: 83–101.

Schaller P and Strobl T (1994) Poisson structure induced (topological) field theories. *Modern Physics Letters* A9: 3129–3136.

Witten E (1988) Topological sigma models. *Communications in Mathematical Physics* 118: 411–449.

Zinn-Justin J (1975) Renormalization of gauge theories. In: Rollnik H and Dietz K (eds.) *Trends in Elementary Particle Physics*, Lecture Notes in Physics, vol. 37. Berlin: Springer.


# Bethe Ansatz

**M T Batchelor**, Australian National University, Canberra, ACT, Australia



## Introduction

The Bethe ansatz is a particular form of wave function introduced in the diagonalization of the Heisenberg spin chain. It underpins the majority of exactly solved models in statistical mechanics and quantum field theory. At the heart of the Bethe ansatz is the way in which multibody interactions factor into two-body interactions. The Bethe ansatz is thus intimately entwined with the theory of integrability.

The way in which the Bethe ansatz works is best understood by working through an explicit hands-on example. The canonical example is the isotropic antiferromagnetic Heisenberg Hamiltonian

$$H = \sum_{i=1}^{L-1} h_{i,i+1} + h_{L,1}, \quad h_{ij} = \tfrac{1}{2}(\boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_j + 1) \quad [1]$$

where $\boldsymbol{\sigma} = (\sigma^x, \sigma^y, \sigma^z)$ are Pauli matrices and $L$ is the length of the chain. Periodic boundary conditions are imposed. However, open boundary conditions may also be treated, along with the addition of magnetic bulk and boundary fields. The $z$-components of each of the spins are either up or down. Since the $z$-component of the total spin commutes with the Hamiltonian, the total number $n$ of up spins serves as a good quantum number. A state of the system can therefore be conveniently described in terms of the coordinates of all the up spins. Denote these coordinates by $x_i$, with $1 \leq x_i \leq L$. The quantum number $n$ ensures that the Hamiltonian decomposes into $L+1$ sectors, each of size $L$ choose $n$. The antiferromagnetic ground state occurs in the largest sector.

The normalization of the Hamiltonian [1] is such that its action is that of the permutation operator:

$$\begin{aligned} h|--\rangle &= |--\rangle, \qquad h|++\rangle = |++\rangle \\ h|+-\rangle &= |-+\rangle, \qquad h|-+\rangle = |+-\rangle \end{aligned} \qquad [2]$$

## Diagonalization of Sectors

One can address the diagonalization of the sectors for various cases.

### Case 1: $n=0$

Consider the case with all spins down. The eigenstate is $\Psi = |- \cdots -\rangle$, with $H\Psi = L\Psi$ and, thus, $E = L$ is the trivial solution.

### Case 2: $n=1$

There are $L$ states, with

$$\Psi = \sum_{x=1}^{L} a(x)|\psi(x)\rangle \qquad [3]$$

where $|\psi(x)\rangle$ is the state with an up spin at site $x$. The aim is to find the amplitudes $a(x)$. It is clear that

$$\begin{aligned} H|\psi(x)\rangle = (L-2)|\psi(x)\rangle + |\psi(x-1)\rangle \\ + |\psi(x+1)\rangle \end{aligned} \qquad [4]$$

in the bulk (away from either boundary). Insertion of [3] into $H\Psi = E\Psi$ gives

$$Ea(x) = (L-2)a(x) + a(x-1) + a(x+1) \qquad [5]$$

Substitution of spin waves $a(x) = e^{ikx}$ gives

$$E = L - 2 + 2\cos k \qquad [6]$$

The boundary conditions are such that $a(0) = a(L)$ and $a(L+1) = a(1)$; either gives $e^{ikL} = 1$, from which the $L$ values of $k$ follow.

### Case 3: $n=2$

Here the wave function can be written in terms of the two flipped spins as

$$\Psi = \sum_{x<y} a(x,y)|\psi(x,y)\rangle \qquad [7]$$

It is to be emphasized that one is working in the region with $x < y$. There are two cases to consider: (1) $y > x+1$ and (2) $y = x+1$. Consider the interactions in the bulk. For (1) the action of the Hamiltonian implies

$$\begin{aligned} Ea(x,y) = (L-4)a(x,y) + a(x-1,y) + a(x+1,y) \\ + a(x,y-1) + a(x,y+1) \end{aligned} \qquad [8]$$

and for (2)

$$\begin{aligned} Ea(x,x+1) = (L-2)a(x,x+1) \\ + a(x-1,x+1) + a(x,x+2) \end{aligned} \qquad [9]$$

The compatibility of these two equations requires that

$$2a(x,x+1) = a(x,x) + a(x+1,x+1) \qquad [10]$$

which is known as the "collision" or "meeting" condition.

Some adjustments need to be made for spins which get flipped at the boundaries. Looking at [8] and [9] with $x = 1$ and $x = L$, it is evident that one can take

$$a(y, x+L) = a(x,y) \qquad [11]$$

to restore the original ordering. The terms which arise involve up spins at sites 0 and $L+1$. This illustrates the periodic boundary condition.

We now assume (the Bethe ansatz) that

$$a(x,y) = A_{12}e^{ik_1x}e^{ik_2y} + A_{21}e^{ik_2x}e^{ik_1y} \qquad [12]$$

Substitution of the ansatz [12] into [8] gives

$$E = L - 4 + 2\cos k_1 + 2\cos k_2 \qquad [13]$$

Substitution of [12] into [10] gives

$$\frac{A_{12}}{A_{21}} = -\frac{1 - 2\,e^{ik_1} + e^{i(k_1+k_2)}}{1 - 2\,e^{ik_2} + e^{i(k_1+k_2)}} \qquad [14]$$

The three relations [11], [12], and [14] give the Bethe equations

$$e^{ik_1L} = \frac{A_{12}}{A_{21}} \quad \text{and} \quad e^{ik_2L} = \frac{A_{21}}{A_{12}} \qquad [15]$$

which are to be solved for $k_1$ and $k_2$. Note that $e^{i(k_1+k_2)L} = 1$.

**Case 4:** $n = 3$

The full power of the Bethe ansatz method becomes evident for three particles. Here

$$\Psi = \sum_{x<y<z} a(x,y,z)|\psi(x,y,z)\rangle \qquad [16]$$

There are several cases to consider:

1. $y > x+1$ and $z > y+1$, where

$$Ea(x,y,z) = (L-6)a(x,y,z) + a(x \pm 1, y, z) + a(x, y \pm 1, z) + a(x, y, z \pm 1) \qquad [17]$$

By $a(x \pm 1, y, z)$, we mean $a(x+1,y,z) + a(x-1,y,z)$, etc.

2. $y = x+1$ and $z > y+1$, with

$$Ea(x,x+1,z) = (L-4)a(x,x+1,z) + a(x-1,x+1,z) + a(x,x+2,z) + a(x,x+1,z \pm 1) \qquad [18]$$

3. $y > x+1$ and $z = y+1$, where

$$Ea(x,y,y+1) = (L-4)a(x,y,y+1) + a(x \pm 1, y, y+1) + a(x,y-1,y+1) + a(x,y,y+2) \qquad [19]$$

4. $y = x+1$ and $z = y+1$, for which

$$Ea(x,x+1,x+2) = (L-2)a(x-1,x+1,x+2) + a(x,x+1,x+3) \qquad [20]$$

Again, we must ensure that these equations are compatible. This involves comparison of the last three equations with [17]. The three equations to be satisfied are

$$2a(x,x+1,z) = a(x,x,z) + a(x+1,x+1,z) \qquad [21]$$

$$2a(x,y,y+1) = a(x,y,y) + a(x,y+1,y+1) \qquad [22]$$

$$4a(x,x+1,x+2) = a(x,x,x+2) + a(x,x+1,x+1) + a(x,x+2,x+2) + a(x+1,x+1,x+2) \qquad [23]$$

But note that setting $z = x+2$ in [21] and $y = x+1$ in [22] leads to [23] being automatically satisfied. We are thus left with only two equations [21] and [22]. Note the similarity between these two equations and the meeting condition [10] for the $n = 2$ case.

In this case the Bethe ansatz is

$$a(x,y,z) = A_{123} z_1^x z_2^y z_3^z + A_{132} z_1^x z_3^y z_2^z + A_{213} z_2^x z_1^y z_3^z + A_{231} z_2^x z_3^y z_1^z + A_{321} z_3^x z_2^y z_1^z + A_{312} z_3^x z_1^y z_2^z \qquad [24]$$

in which $z_j = e^{ik_j}$. This is a sum over the 3! permutations of the integers 1, 2, 3. Inserting this ansatz into [17] gives

$$E = L - 6 + 2(\cos k_1 + \cos k_2 + \cos k_3) \qquad [25]$$

To determine the $k_j$, it is convenient to define

$$s_{ij} = 1 - 2z_j + z_i z_j \qquad [26]$$

Substitution of [24] into the meeting conditions [21] and [22] then gives

$$s_{12}A_{123} + s_{21}A_{213} + s_{13}A_{132} + s_{31}A_{312} + s_{23}A_{231} + s_{32}A_{321} = 0 \qquad [27]$$

$$s_{23}A_{123} + s_{32}A_{132} + s_{13}A_{213} + s_{31}A_{231} + s_{21}A_{321} + s_{12}A_{312} = 0 \qquad [28]$$

These equations are assumed to be satisfied in permutation pairs, that is,

$$s_{12}A_{123} + s_{21}A_{213} = 0$$
$$s_{23}A_{123} + s_{32}A_{132} = 0, \text{ etc.} \qquad [29]$$

Up to an overall constant, the relations [27] and [28] are satisfied by

$$A_{123} = s_{21}s_{31}s_{32}, \quad A_{132} = -s_{31}s_{21}s_{23}$$
$$A_{312} = s_{13}s_{23}s_{21}, \quad A_{321} = -s_{23}s_{13}s_{12} \qquad [30]$$
$$A_{231} = s_{32}s_{12}s_{13}, \quad A_{213} = -s_{12}s_{32}s_{31}$$

The boundary condition, $a(y,z,x+L) = a(x,y,z)$, gives

$$(z_1^L A_{321} - A_{132}) z_1^x z_3^y z_2^x + (z_2^L A_{312} - A_{231}) z_2^x z_3^y z_1^x + (z_1^L A_{231} - A_{123}) z_1^x z_2^y z_3^x + (z_3^L A_{213} - A_{321}) z_3^x z_2^y z_1^x + (z_2^L A_{132} - A_{213}) z_2^x z_1^y z_3^x + (z_3^L A_{123} - A_{312}) z_3^x z_1^y z_2^x = 0 \qquad [31]$$

This leads to the equations

$$z_1^L = \frac{A_{123}}{A_{231}} = \frac{A_{132}}{A_{321}} = \frac{s_{21}s_{31}}{s_{12}s_{13}}$$
$$z_2^L = \frac{A_{213}}{A_{132}} = \frac{A_{231}}{A_{312}} = \frac{s_{12}s_{32}}{s_{21}s_{23}} \qquad [32]$$
$$z_3^L = \frac{A_{321}}{A_{213}} = \frac{A_{312}}{A_{123}} = \frac{s_{13}s_{23}}{s_{31}s_{32}}$$

which can be solved for the Bethe roots $k_j$.

### General $n$

The general Bethe ansatz is

$$a(x_1, \ldots, x_n) = \sum_P A_{p_1,\ldots,p_n} z_{p_1}^{x_1} \ldots z_{p_n}^{x_n} \qquad [33]$$

where the sum is over all $n!$ permutations $P = \{p_1, \ldots, p_n\}$ of the integers $1, \ldots, n$. The boundary condition is

$$a(x_2, x_3, \ldots, x_n, x_1 + L) = a(x_1, x_2, \ldots, x_n) \qquad [34]$$

leading to the Bethe equations

$$z_{p_1}^L = \frac{A_{p_1,\ldots,p_n}}{A_{p_2,\ldots,p_n,p_1}} \qquad [35]$$

for all permutations, with

$$A_{p_1,\ldots,p_n} = \epsilon_P \prod_{1 \le i < j \le n} s_{p_j,p_i} \qquad [36]$$

where $\epsilon_P$ is the signature of the permutation. Finally,

$$z_{p_1}^L = (-)^{n-1} \prod_{\ell=2}^{n} \frac{s_{p_\ell,p_1}}{s_{p_1,p_\ell}} \quad \text{or} \quad z_j^L = (-)^{n-1} \prod_{\substack{\ell=1 \\ \neq j}}^{n} \frac{s_{\ell,j}}{s_{j,\ell}} \qquad [37]$$

for $j = 1, \ldots, n$. The eigenvalues are given by

$$E = L + \sum_{j=1}^{n} (2 \cos k_j - 2) \qquad [38]$$

Another form of the Bethe equations is obtained by defining

$$e^{ik_j} = \frac{u_j - (1/2)i}{u_j + (1/2)i} \qquad [39]$$

which gives

$$E = L - \sum_{j=1}^{n} \frac{1}{u_j^2 + 1/4} \qquad [40]$$

with $u_j$ satisfying

$$\left( \frac{u_j - (1/2)i}{u_j + (1/2)i} \right)^L = - \prod_{\ell=1}^{n} \frac{u_j - u_\ell - i}{u_j - u_\ell + i} \qquad [41]$$

for $j = 1, \ldots, n$.

All eigenvalues of the Heisenberg spin chain may be obtained in terms of the Bethe ansatz solution. For example, the distribution of roots $u_j$ for the ground state are real and symmetric about the origin. Excitations may involve complex roots. Although obtained exactly in terms of the Bethe roots, the Bethe ansatz wave function is cumbersome.

We have thus seen how the Bethe ansatz works for the Heisenberg spin chain. The underlying mechanism is the way in which the collision or meeting conditions can be handled in terms of two-body interactions. To see this more clearly, the six permutation pair equations [29] can be written in the general form $A_{abc} = Y_{ab} A_{bac}$ and $A_{abc} = Y_{bc} A_{acb}$, where $Y_{ab} = -s_{ba}/s_{ab}$. Now there are two possible paths to get from $A_{abc}$ to $A_{cba}$, namely

$$\begin{aligned} A_{cba} &= Y_{ab} Y_{ac} Y_{bc} A_{abc} \\ A_{cba} &= Y_{bc} Y_{ac} Y_{ab} A_{abc} \end{aligned} \qquad [42]$$

Both paths must be equivalent, with

$$Y_{ab} Y_{ba} = 1 \quad \text{and} \quad Y_{ab} Y_{ac} Y_{bc} = Y_{bc} Y_{ac} Y_{ab} \qquad [43]$$

The latter is a condition of nondiffraction or equivalently a manifestation of the Yang–Baxter equation.

Historically, the next model to be exactly solved in terms of the Bethe ansatz was the one-dimensional model of $N$ interacting bosons on a line of length $L$ defined by the Hamiltonian

$$H = - \sum_{i=1}^{N} \frac{\partial^2}{\partial x_i^2} + 2c \sum_{1 \le i < j \le N} \delta(x_i - x_j) \qquad [44]$$

where $c$ is a measure of the interaction strength. For this model the Bethe ansatz wave function is of the same form as [33] with the two-body interaction term given by

$$s_{ab} = k_a - k_b + ic \qquad [45]$$

The Bethe equations are given by

$$\exp(ik_j L) = - \prod_{\ell=1}^{N} \frac{k_j - k_\ell + ic}{k_j - k_\ell - ic} \quad \text{for} \quad j = 1, \ldots, N \qquad [46]$$

The energy eigenvalue is

$$E = \sum_{j=1}^{N} k_j^2 \qquad [47]$$

For repulsive ($c > 0$) interactions, one can prove that all Bethe roots are real.

The Bethe ansatz has been applied to a number of other and more general models, both for discrete spins and in the continuum. These include the anisotropic Heisenberg ($XXZ$) spin chain, for which the above working readily generalizes to trigonometric functions. The underlying ansatz [33] remains the same. One key generalization is the nested Bethe ansatz, which arises, for example, in the solution of the general $N$-state permutator model, the Hubbard model, and the Gaudin–Yang model of interacting fermions. For such models the nested Bethe ansatz involves an additional level of work to determine the amplitudes appearing in the

wave function [33] due to higher symmetries. This results in Bethe equations involving different types or colors of roots.

The exactly solved one-dimensional quantum spin chains may also be obtained from their two-dimensional classical counterparts – the vertex models. For example, the six-vertex model shares the same Bethe ansatz wave function and Bethe equations as the *XXZ* spin chain. The more general permutator Hamiltonians are related to multistate vertex models. One may also consider other spin-*S* models.

The discussion in this article has centered on what is known as the coordinate Bethe ansatz. Another formulation is the algebraic Bethe ansatz, which was developed for the systematic treatment of the higher-spin models. In this formulation, operators create the Bethe states by acting on a vacuum. The algebraic Bethe ansatz goes hand-in-hand with the quantum inverse-scattering method. In all of the exactly solved Bethe ansatz models, it is possible to derive quantities like the ground-state energy per site via the root density method, which assumes that the Bethe roots form a uniform distribution in the infinite-size limit. The thermodynamics of the Bethe ansatz solvable models may also be calculated in a systematic fashion.

Despite Bethe's early optimism, the Bethe ansatz has not been extended to higher-dimensional systems.

*See also:* Affine Quantum Groups; Eight Vertex and Hard Hexagon Models; Integrability and Quantum Field Theory; Integrable Systems: Overview; Quantum Spin Systems; Yang–Baxter Equations.

## Further Reading

Baxter RJ (1983) *Exactly Solved Models in Statistical Mechanics*. London: Academic Press.
Baxter RJ (2003) Completeness of the Bethe ansatz for the six- and eight-vertex models. *Journal of Statistical Physics* 108: 1–48.
Bethe HA (1931) Zur Theorie der Metalle I. Eigenwerte und Eigenfunktionen der linearen Atomkette. *Zeitschrift für Physik* 71: 205–226.
Gaudin M (1967) Un Système à Une Dimension de Fermions en Interaction. *Physics Letters A* 24: 55–56.
Gaudin M (1983) *la Fonction d'onde de Bethe*. Paris: Masson.
Korepin VE, Izergin AG, and Bogoliubov NM (1993) *Quantum Inverse Scattering Method and Correlation Functions*. Cambridge: Cambridge University Press.
Lieb EH and Liniger W (1963) Exact analysis of an interacting Bose gas I. The general solution and the ground state. *Physical Review* 130: 1605–1616.
Mattis DC (1993) *The Many-Body Problem: An Encyclopaedia of Exactly Solved Models in One-Dimension*. Singapore: World Scientific.
McGuire JB (1964) Study of exactly soluble one-dimensional *N*-body problems. *Journal of Mathematical Physics* 5: 622–636.
Sutherland B (2004) *Beautiful Models: 70 Years of Exactly Solved Quantum Many–Body Problems*. Singapore: World Scientific.
Takahashi M (1999) *Thermodynamics of One-Dimensional Solvable Models*. Cambridge: Cambridge University Press.
Yang CN (1967) Some exact results for the many-body problem in one-dimension with repulsive Delta-function interaction. *Physical Review Letters* 19: 1312–1315.

# BF Theories

**M Blau**, Université de Neuchâtel, Neuchâtel, Switzerland



## Introduction

BF theories are a class of gauge theories with a nontrivial metric-independent classical action. As such these theories are candidate topological field theories akin to the Chern–Simons theory in three dimensions, but in contrast to the Chern–Simons theory these exist and are well defined in arbitrary dimensions.

The name "BF theories" derives from the fact that, roughly (see [1] below and the subsequent discussion for a more precise description), the action of the BF theory takes the form $\int B \wedge F_A$ with $F_A$ the curvature of a connection $A$ and $B$ a Lagrange multiplier. The classical equations of motion imply that $A$ is flat, $F_A = 0$, and thus BF theories are topological gauge theories of flat connections.

Abelian BF theories and their relation to topological invariants (the Ray–Singer torsion) were originally discussed by Schwarz (1978, 1979). In the context of the topological field theory, nonabelian BF theories were introduced in Horowitz (1989) and Blau and Thompson (1989, 1991).

Since then, BF theories have attracted a lot of attention as simple toy-models of (topological) gauge theories, and also because of their relationships with the Chern–Simons theory, the Yang–Mills theory, and gauge-theory formulations of gravity, as well as because of the rather rich and intricate structure of their quantum theories.

The purpose of this article is to provide an overview of these various features of BF theories. The standard reference for the basic classical and quantum properties of BF theories is Birmingham *et al.* (1991).

## Basic Classical Properties of BF Theories

### Nonabelian BF Theories

**The classical action and equations of motion** Typically, the classical action of the BF theory takes the form

$$S_{\mathrm{BF}}(A,B) = \int_M \mathrm{tr}_G\, B \wedge F_A \qquad [1]$$

where $F_A$ is the curvature of a connection $A$ on a principal $G$-bundle $P \to M$ over an $n$-dimensional manifold $M$, $B$ is an ad-equivariant horizontal $(n-2)$-form on $P$, and $\mathrm{tr}_G$ (a trace) denotes an ad-invariant nondegenerate scalar product on the Lie algebra $\mathfrak{g}$ of the Lie group $G$. Generalizations of this are possible, in particular, for $G$ abelian or for $n=3$ and are mentioned below.

We consider $F_A$ and $B$ as forms on $M$ taking values in the bundle of Lie algebras $\mathrm{ad}\, P = P \times_{\mathrm{ad}} \mathfrak{g}$ and refer to such objects as elements of $\Omega^*(M,\mathfrak{g})$. Then $\mathrm{tr}\, B \wedge F_A \in \Omega^n(M,\mathbb{R})$ is a volume form on $M$. In order to simplify the exposition, in the following we will mostly assume that $G$ is compact semisimple and that $M$ is compact without a boundary (even though relaxing any one of these conditions is possible and also of interest in its own right).

Varying the action [1] with respect to $A$ and $B$, one obtains the classical equations of motion

$$F_A = 0, \qquad d_A B = 0 \qquad [2]$$

where

$$d_A B = \mathrm{d}B + [A,B] \qquad [3]$$

is the covariant exterior derivative. In particular, therefore, the equations of motion imply that the connection $A$ is flat.

**Gauge invariance** For any $n$, the action [1] is invariant under $G$ gauge transformations (vertical automorphisms of $P$) acting on $A$ and $B$ as

$$A \to g^{-1}Ag + g^{-1}\mathrm{d}g, \qquad B \to g^{-1}Bg \qquad [4]$$

(the latter is what is meant by the fact that $B$ takes values in ad $P$), because $F_A$ is also ad-equivariant, $F_A \to g^{-1}F_{Ag}$, and $\mathrm{tr}_G$ is ad-invariant. The infinitesimal version of this statement is that the action is invariant under the variations

$$\delta A = d_A\lambda, \qquad \delta B = [B,\lambda] \qquad [5]$$

where $\lambda \in \Omega^0(M,\mathfrak{g})$ can (formally) be thought of as an element of the Lie algebra of the group of gauge transformations.

Gauge-fixing this symmetry can proceed in the usual way (via the Faddeev–Popov or Becchi–Rouet–Stora–Tyupkin procedure), a typical gauge choice being $d_{A_0} \star (A - A_0) = 0$ where $A_0$ is a reference connection, and $\star$ is the Hodge duality operator corresponding to a choice of metric on $M$.

**Local $p$-form symmetries** For $n=2$, the only local symmetries of the BF action are the above $G$ gauge transformations. For $n>2$, however, there are other local symmetries associated with shifts of $B_p \in \Omega^p(M,\mathfrak{g})$ with $p = n-2 > 0$. Indeed, integration by parts using Stokes' theorem and $\partial M = 0$ shows that [1] is invariant under

$$A \to A, \quad B_p \to B_p + d_A\lambda_{p-1}, \quad \lambda_{p-1} \in \Omega^{p-1}(M,\mathfrak{g}) \qquad [6]$$

For $p=1$, $\lambda$ is a 0-form and the invariance follows. For $p>1$, however, the gauge parameter has, in some sense, its own gauge invariance. Namely, under the shift

$$\lambda_{p-1} \to \lambda_{p-1} + d_A\lambda_{p-2} \qquad [7]$$

one has

$$d_A\lambda_{p-1} \to d_A\lambda_{p-1} + [F_A, \lambda_{p-2}] \qquad [8]$$

Thus for $F_A = 0$, the shift [7] has no effect on the local symmetry [6]. Likewise, for $p>2$ the parameter $\lambda_{p-2}$ itself has a similar invariance, etc. Since $F_A=0$ is one of the classical equations of motion, the shift symmetry [6] is what is called an "on-shell reducible symmetry." Gauge-fixing such symmetries is not straightforward, and one generally appeals to the Batalin–Vilkovisky formalism to accomplish this.

**Diffeomorphisms and local symmetries** One manifestation of the general covariance of the BF action [1] is the on-shell equivalence of (infinitesimal) diffeomorphisms and (infinitesimal) local symmetries. Diffeomorphisms are generated by the Lie derivative $L_X$ along a vector field $X$. The action of $L_X$ on differential forms is given by the Cartan formula $L_X = \mathrm{d}i_X + i_X\,\mathrm{d}$, where $i_{(.)}$ is the operation of contraction. The action of the Lie derivative on $A$ and $B$ can be written in gauge covariant form as

$$\begin{aligned} L_X A &= i_X F_A + d_A\lambda(X), \\ L_X B &= i_X\, d_A B + [B,\lambda(X)] + d_A\lambda'(X) \end{aligned} \qquad [9]$$

where $\lambda(X) = i_X A$ and $\lambda'(X) = i_X B$. This shows that on-shell diffeomorphisms are equivalent to field-dependent gauge and $p$-form symmetries of the BF action.

**The classical moduli space** The classical moduli space $\mathcal{C} = \mathcal{C}(P,M,G)$ is the space of solutions to the classical equations of motion modulo the local symmetries of the action. Since the field content

and the nature of the local symmetries of the BF theory depend strongly on the dimension $n$ of $M$, the structure and interpretation of the classical moduli space also depend on $n$.

For $n=2$, by [5] the equation of motion [2] for $B \in \Omega^0(M, \mathfrak{g})$ says that $A$ is invariant under the infinitesimal gauge transformation generated by $B$. Thus if $A$ is "irreducible," there are no nontrivial solutions for $B$ and, away from reducible flat connections, the classical moduli space is just the moduli space of flat connections on $P \to M$ over the surface $M$:

$$\mathcal{C}_{n=2} = \mathcal{M}_{\text{flat}}(P, G) \quad [10]$$

This space may or may not be empty, depending on whether $P$ admits flat connections or not.

For $n=3$, the equation of motion [2] for $B \in \Omega^1(M, \mathfrak{g})$ says that $B$ is a tangent vector to the space of flat connections at the flat connection $A$, in the sense that under the variation $\delta A = B$, one has

$$\delta F_A = d_A B = 0 \quad [11]$$

The local $G$ gauge symmetry and the 1-form symmetry [6] now imply that the moduli space of classical solutions can be identified with the (co-)tangent bundle of the moduli space of flat connections on $P \to M$ over the 3-manifold $M$:

$$\mathcal{C}_{n=3} = T\mathcal{M}_{\text{flat}}(P, G) \quad [12]$$

In higher dimensions there appears to be less geometrical structure associated with BF theories, and all that can be said in general is that the tangent space to $\mathcal{C}_n$ at a solution $(A, B)$ of the equations of motion [2] is the vector space:

$$T_{(A,B)}\mathcal{C}_n = H^1_A(M, \mathfrak{g}) \oplus H^{n-2}_A(M, \mathfrak{g}) \quad [13]$$

where $H^k_A(M, \mathfrak{g})$ are the cohomology groups of the deformation complex

$$d_A : \Omega^*(M, \mathfrak{g}) \to \Omega^{*+1}(M, \mathfrak{g}) \quad [14]$$

associated with the flat connection $A$, $F_A = (d_A)^2 = 0$.

When $M$ is topologically of the form $M = \Sigma \times \mathbb{R}$ (where one can think of $\mathbb{R}$ as time), one has

$$T_{(A,B)}\mathcal{C}_n = H^1_A(\Sigma, \mathfrak{g}) \oplus H^{n-2}_A(\Sigma, \mathfrak{g}) \quad [15]$$

This is naturally a symplectic vector space (necessary for a phase space), the nondegenerate antisymmetric pairing being given by Poincaré duality:

$$\omega([a_1], [b_1]; [a_2], [b_2]) = \int_\Sigma \operatorname{tr}_G(a_1 \wedge b_2 - a_2 \wedge b_1) \quad [16]$$

**Metric independence** Perhaps the most important property of the action [1] is that, in contrast to, for example, the usual Yang–Mills action for nonabelian gauge fields

$$S_{\text{YM}} = \frac{1}{4g^2} \int_M \operatorname{tr}_G F_A \wedge \star F_A \quad [17]$$

it does not require a metric (or the corresponding Hodge duality operator $\star$) for its formulation. This makes it a candidate action for a "topological field theory," this term loosely referring to field theories which, in a suitable sense, do not depend on additional structures imposed on the underlying space(-time) manifold $M$, in this case a Riemannian structure.

To establish that BF theories are "topological quantum field theories," one needs to show that the partition function (and correlation functions) of the quantized BF theory are also metric independent. This is not completely automatic as typically the metric enters in the gauge fixing of the local symmetries of the action which is required to make the quantum theory well defined. The usual lore is that since the metric only enters through the gauge fixing and since the quantum theory should be independent of the choice of gauge, it should also be metric independent. In the case of nonabelian BF theories, the complexity of their local symmetries complicates the analysis somewhat, but it can nevertheless be shown that BF theories indeed define topological field theories also at the quantum level.

### Special Features of Abelian BF Theories

All the features of nonabelian BF theories discussed above are, of course, also valid when $G$ is abelian (with some obvious modifications and simplifications). However, when $G$ is abelian, a more general action than [1] is possible. Indeed, although there is no obvious higher $p$-form analog of nonabelian gauge fields, in the abelian case $G = \mathrm{U}(1)$ or $G = \mathbb{R}$, and the condition $F_A \in \Omega^2(M, \mathbb{R})$ can be relaxed. In particular, one can consider the actions

$$S(n,p) \equiv S(B_p, C_{n-p-1}) = \int_M B_p \wedge \mathrm{d}C_{n-p-1} \quad [18]$$

with $B_p \in \Omega^p(M, \mathbb{R}), C_{n-p-1} \in \Omega^{n-p-1}(M, \mathbb{R})$, and $F_C = \mathrm{d}C$, its $(n-p)$-form field strength. More generally, one can also consider the hybrid action

$$S_A(n,p) = \int_M B_p \wedge d_A C_{n-p-1} \quad [19]$$

where $A$ is a fixed (nondynamical) flat $G$-connection, $d_A^2 = 0$, and $B$ and $C$ take values in the corresponding adjoint bundle. This action can be considered as the linearization of the nonabelian BF action [1] around

the flat connection $A$, and it reduces to the abelian BF action [18] for $\mathfrak{g} = \mathbb{R}$.

The action is invariant under the (reducible) local symmetries

$$\begin{aligned} B_p &\to B_p + d_A \lambda_{p-1} \\ C_{n-p-1} &\to C_{n-p-1} + d_A \lambda'_{n-p-2} \end{aligned} \qquad [20]$$

The space of solutions to the equations of motion $d_A C = d_A B = 0$ modulo gauge symmetries is (cf. [13]) the finite-dimensional vector space

$$\mathcal{C}_{n,p} = H_A^p(M, \mathfrak{g}) \oplus H_A^{n-p-1}(M, \mathfrak{g}) \qquad [21]$$

which is naturally symplectic for $M = \Sigma \times \mathbb{R}$.

## Uses and Applications of Quantum Abelian BF Theories

### Quantization of Abelian BF Theories and the Ray–Singer Torsion

We will now show that the partition function of the abelian BF theory (actually more generally that of the linearized nonabelian BF action [19]) is related to the Ray–Singer torsion of $M$. This requires some preparatory material on Gaussian path integrals, determinants, and gauge fixing that we present first.

In order to simplify the exposition, we assume that there are no harmonic modes, either because they have been gauged away or because the cohomology groups of $d_A$ are trivial, $H_A^k(M, \mathfrak{g}) = 0$, that is, the deformation complex [14] is "acyclic."

**Laplacians, determinants, and the Ray–Singer torsion** Choosing a Riemannian metric $g$ (and Hodge duality operator $\star$) on $M$, the twisted Laplacian on $p$-forms is

$$\Delta_A^{(p)} = (d_A + d_A^\star)^2 = d_A d_A^\star + d_A^\star d_A \qquad [22]$$

where $d_A^\star = \pm \star d_A \star$ is the adjoint of d with respect to the scalar product on $p$-forms defined by $\star$. This is an elliptic operator whose determinant can be defined, for example, by a $\zeta$-function regularization. Denoting the (nonzero) eigenvalues of $\Delta_A^{(p)}$ by $\lambda_k^{(p)}$, its $\zeta$-function is

$$\zeta^{(p)}(s) = \sum_k \left(\lambda_k^{(p)}\right)^{-s} \qquad [23]$$

This converges for Re($s$) sufficiently large and can be analytically continued to a meromorphic function of $s$ analytic at $s = 0$, so that

$$\det \Delta_A^{(p)} := \mathrm{e}^{-\zeta^{(p)\prime}(0)} \qquad [24]$$

is well defined. The Ray–Singer torsion of $(M, \mathfrak{g})$ (with respect to the flat connection $A$) is then defined by

$$T_A(M) = \prod_{p=0}^{n} \left(\det \Delta_A^{(p)}\right)^{(-1)^p p/2} \qquad [25]$$

Even though this definition depends strongly on the metric $g$ on $M$, the Ray–Singer torsion has the remarkable property of being independent of $g$. The Ray–Singer torsion can be shown to be trivial (essentially =1 modulo zero-mode contributions) in even dimensions, but is a nontrivial topological invariant in odd dimensions. Henceforth, we will suppress the dependence on $M$ and denote the $n$-dimensional Ray–Singer torsion by $T_A(n)$.

**Gaussian path integrals and determinants** The path integral for abelian BF theories is modeled on the usual formula for a $\delta$-function

$$\delta^n(x) = \frac{1}{(\sqrt{2\pi})^n} \int_{\mathbb{R}^n} \mathrm{d}^n \pi \, \mathrm{e}^{\mathrm{i}\pi x} \qquad [26]$$

from which one deduces the Gaussian integral formula

$$\begin{aligned} &\frac{1}{(\sqrt{2\pi})^n} \int_{\mathbb{R}^n \times \mathbb{R}^n} \mathrm{d}^n \pi \, \mathrm{d}^n x \, \mathrm{e}^{\mathrm{i}\pi D x + \mathrm{i}K x + \mathrm{i}\pi J} \\ &= \int_{\mathbb{R}^n} \mathrm{d}^n x \delta^n(Dx + J) \, \mathrm{e}^{\mathrm{i}Kx} \\ &= \frac{1}{\det D} \, \mathrm{e}^{-\mathrm{i}K.D^{-1}J} \end{aligned} \qquad [27]$$

Here, we have assumed that the operator (matrix) $D$ is invertible. The model that one uses in the path integral is that

$$\int \mathrm{d}[\phi] \, \mathrm{d}[\chi] \, \mathrm{e}^{\mathrm{i} \int_M \phi \star D \chi} = (\det D)^{\mp 1} \qquad [28]$$

where $\phi$ is a set of fields and the $\chi$ are a set of dual fields with $D$ again a nondegenerate operator. The inverse determinant arises for Grassmann even fields (as in [27]), while it is the determinant that appears for Grassmann odd fields.

**Gauge fixing – the Faddeev–Popov trick** If the action [19], $S_A(n, p) = \int B_p d_A C_{n-p-1}$, were nondegenerate, its partition function could be defined directly by [28]. However, because of gauge invariance of the action, the kinetic term is degenerate and one needs to eliminate the gauge freedom to obtain an (at least formally) well-defined expression for the partition function. Concretely, this degeneracy can be seen by

recalling that, when there are no harmonic forms (as we have assumed), there is a unique orthogonal Hodge decomposition of a $p$-form $B_p \in \Omega^p(M, \mathfrak{g})$ into a sum of a $d_A$-exact and a $d_A$-coexact form:

$$B_p = d_A \lambda_{p-1} + d_A^\star \tau_{p+1} \qquad [29]$$

(and likewise for $C$). Evidently, the exact (longitudinal) parts $d_A\lambda$ of $B$ and $C$ do not appear in the action, and these are precisely the gauge-dependent parts of $B$ and $C$ under the gauge transformation [20]. Gauge fixing amounts to imposing a condition $\mathcal{F}(B_p) = 0$ on $B_p$ that determines the longitudinal part uniquely in terms of the transversal part $d_A^\star \tau$. A natural condition is

$$d_A \lambda_{p-1} = 0 \Leftrightarrow \mathcal{F}(B_p) = d_A^\star B_p = 0 \qquad [30]$$

A gauge-fixing condition independent of the partition function results from inserting “1” in the form of

$$1 = \int_{\mathcal{G}} \mathrm{d}[g] \delta(\mathcal{F}(B^g)) \Delta_{\mathcal{F}}(B) \qquad [31]$$

into the functional integral (the Faddeev–Popov trick), where $\mathcal{G}$ is the gauge group. This defines the Faddeev–Popov determinant $\Delta_{\mathcal{F}}$, and the functional properties of the delta functional imply that $\Delta_{\mathcal{F}}$ is the determinant of the operator that one obtains upon gauge variation of $\mathcal{F}(B)$.

In the general case of reducible gauge symmetries, the nature of the gauge group is complicated and requires some more thought. In the irreducible case, however, that is, for $p = 1$, the Lie algebra of the gauge group can be identified with $\Omega^0(M, \mathfrak{g})$, and $\Delta_{\mathcal{F}}$ is the determinant of the operator:

$$\frac{\delta \mathcal{F}}{\delta B} d_A : \Omega^0(M, \mathfrak{g}) \to \Omega^0(M, \mathfrak{g}) \qquad [32]$$

For [30], this is simply the Laplacian on 0-forms, and thus

$$\Delta_{\mathcal{F}} = \det \Delta_A^{(0)} \qquad [33]$$

**The partition function** Following the finite-dimensional model, both the $\delta$-function implementing the gauge-fixing condition and the Faddeev–Popov determinant can be lifted into the exponential, the former by a Lagrange multiplier $\pi$ [26], a Grassmann even 0-form, and the latter by a pair of Grassmann odd 0-forms $c$ and $\bar{c}$ [28], the ghost and antighost fields, respectively. The sum of the classical action and these gauge-fixing and ghost terms defines the (BRST-invariant) “quantum action” $S_A^q(n, p)$, and the partition function is

$$Z_A(n, p) = \int \mathrm{d}[\phi] \mathrm{e}^{\mathrm{i} S_A^q(n,p)(\phi)} \qquad [34]$$

where $\phi$ denotes collectively all the fields. Concretely, when $n = 2$ and $p = 0$ (or, equivalently, $p = 1$), the quantum action is

$$S_A^q(2, 0) = \int B_0 d_A C_1 + \pi d_A \star C_1 + \bar{c} \star \Delta_A^{(0)} c \qquad [35]$$

Likewise, for $n = 3$ and $p = 1$ (the only other case when the gauge symmetry is indeed irreducible), both $B_1$ and $C_1$ require separate gauge fixing, and the quantum action is

$$S_A^q(3, 1) = \int B_1 d_A C_1 + \pi d_A \star C_1 + \bar{c} \star \Delta_A^{(0)} c + \pi' d_A \star B_1 + \bar{c}' \star \Delta_A^{(0)} c' \qquad [36]$$

Formally, therefore, the two-dimensional partition function is

$$Z_A(2, 0) = \frac{\det \Delta^{(0)}}{\det D_A} \qquad [37]$$

where $D_A$ is the operator:

$$D_A = \begin{pmatrix} \star d_A \\ \star d_A \star \end{pmatrix} : \Omega^1(M, \mathfrak{g}) \to \Omega^0(M, \mathfrak{g}) \oplus \Omega^0(M, \mathfrak{g}) \qquad [38]$$

One can define the determinant of this operator as the square root of the determinant of the operator $D_A^\star D_A = \Delta_A^{(1)}$, and therefore the partition function

$$Z_A(2, 0) = \det \Delta^{(0)} (\det \Delta^{(1)})^{-1/2} = T_A(2) \qquad [39]$$

is equal to the two-dimensional Ray–Singer torsion [25]. In this case, it is easy to see directly that the even-dimensional Ray–Singer torsion is trivial, as one could have equally well defined the determinant of $D_A$ as the square root of the operator $D_A D_A^\star = \Delta_A^{(0)} \oplus \Delta_A^{(0)}$, which implies $Z_A(2, 0) = 1$.

In three dimensions, the two pairs of ghosts each contribute a $\det \Delta_A^{(0)}$, and thus

$$Z_A(3, 1) = \frac{(\det \Delta^{(0)})^2}{\det D_A} \qquad [40]$$

where

$$D_A = \begin{pmatrix} \star d_A & d_A \\ d_A \star & 0 \end{pmatrix} : \Omega^0(M, \mathfrak{g}) \oplus \Omega^1(M, \mathfrak{g}) \to \Omega^0(M, \mathfrak{g}) \oplus \Omega^1(M, \mathfrak{g}) \qquad [41]$$

is the operator acting on the fields $(B_1, C_1, \pi, \pi')$. As before, this operator can be diagonalized by squaring it, $D_A^* D_A = \Delta^{(0)} \oplus \Delta^{(1)}$, and thus

$$Z_A(3, 1) = (\det \Delta_A^{(0)})^{3/2} (\det \Delta_A^{(1)})^{-1/2} = T_A(3)^{-1} \qquad [42]$$

is again related to the (this time genuinely nontrivial) Ray–Singer torsion.

In spite of the complications caused by reducible gauge symmetries, it can be shown that all of the above generalizes to arbitrary $n$ and $p$, with the result that (for $n$ odd)

$$Z_A(n,p) = T_A(n)^{(-1)^p} \qquad [43]$$

confirming the topological nature of BF theories.

In the nonabelian case, the situation is significantly more complicated because of the complexity of the classical moduli space, the (higher cohomology) zero modes, and the on-shell reducibility of the gauge symmetries. Nevertheless, ignoring all the zero modes except those of $A$, that is, except the moduli $m$ of flat connections $A(m)$, the result is similar to that in the abelian case, in that the partition function reduces to an integral over the moduli space of flat connections, with measure determined by the Ray–Singer torsion $T_{A(m)}$.

#### Linking Numbers as Observables of Abelian BF Theories

With the exception of $p=0$, there are no interesting "local" observables (gauge-invariant functionals of the fields $C$ and $B$) in the abelian BF theory, since the gauge-invariant field strengths d$C$ and d$B$ vanish by the equations of motion. (For $p=0$, $B$ is a gauge-invariant 0-form and hence $B(x)$ is a good local observable.) However, as in the Chern–Simons and Yang–Mills theories, certain (weakly) nonlocal observables such as Wilson loops are also of interest. In the case at hand (eqn [18]), we have abelian Wilson surface operators

$$W_S[B] = \int_S B, \qquad W_{S'}[C] = \int_{S'} C \qquad [44]$$

associated with $p$- and $(n-p-1)$-dimensional submanifolds $S$ and $S'$ of $M$, respectively. These operators are gauge invariant, that is, invariant under the local symmetries [20] provided that $\partial S = \partial S' = 0$, so that $S$ and $S'$ represent homology cycles of $M$.

For $M = \mathbb{R}^n$, correlation functions of these operators are related to the topological linking number of $S$ and $S'$. We choose $S = \partial\Sigma$ and $S' = \partial\Sigma'$ to be disjoint compact-oriented boundaries of oriented submanifolds $\Sigma$ and $\Sigma'$ of $\mathbb{R}^n$. We also introduce de Rham currents $\Delta_\Sigma$ and $\Delta_S$ (essentially distributional differential forms with $\delta$-function support on $\Sigma$ or $S$, respectively), characterized by the properties

$$\begin{aligned} \int_S \omega_p &= \int_M \Delta_S \wedge \omega_p \\ \int_\Sigma \omega_{p+1} &= \int_M \Delta_\Sigma \wedge \omega_{p+1} \end{aligned} \qquad [45]$$

for all $\omega_k \in \Omega^k(M, \mathbb{R})$ (and likewise for $S'$ and $\Sigma'$).

Since the dimension of $\Sigma$ is equal to the codimension of $S' = \partial\Sigma'$, $\Sigma$ and $S'$ will generically intersect transversally at isolated points, and we define the "linking number" of $S$ and $S'$ to be the intersection number of $\Sigma$ and $S'$, expressed in terms of de Rham currents as

$$L(S,S') = \int_\Sigma \Delta_{S'} = \int_M \Delta_\Sigma \Delta_{S'} \qquad [46]$$

In terms of de Rham currents, the Wilson surface operators can be written as $W_S[B] = \int_M \Delta_S \wedge B$, etc. Thus, the generating functional for correlation functions of Wilson surface operators

$$\langle e^{i\beta W_S[B]} e^{i\alpha W_{S'}[C]} \rangle = \int D[C]D[B] e^{i\int_M (B\,dC + \alpha\Delta_{S'}C + \beta\Delta_S B)} \qquad [47]$$

is simply a Gaussian path integral. Using the defining properties of de Rham currents, this can be formally evaluated (using [27]) to give

$$\langle e^{i\beta W_S[B]} e^{i\alpha W_{S'}[C]} \rangle = e^{\pm i\alpha\beta L(S,S')} \qquad [48]$$

As expected, correlation functions of these topological field theories encode topological information.

## Uses and Applications of Classical Nonabelian BF Theories

Low-dimensional BF theories are closely related to other theories of interest, for example, the Yang–Mills theory, the Chern–Simons theory, and gravity. Here, we briefly review some of these relationships. In order to avoid the complexities of quantum nonabelian BF theories, we focus on their classical features. Brief suggestions for further reading are provided at the end of each subsection.

#### Relation with Yang–Mills Theory

In any dimension, the nonabelian BF action can be regarded as the zero-coupling limit $g^2 \to 0$ of the Yang–Mills theory since the Yang–Mills action [17] can be written in first-order form as

$$\frac{1}{4g^2}\int_M \mathrm{tr}_G\, F_A \wedge \star F_A \equiv \int_M \mathrm{tr}_G[iB_{n-2} \wedge F_A + g^2 B_{n-2} \wedge \star B_{n-2}] \qquad [49]$$

However, whereas for $n \geq 3$ the $B^2$-term breaks the $p$-form gauge invariance of the BF action (and thus liberates the physical Yang–Mills degrees of freedom), this limit is nonsingular in two dimensions where this $p$-form symmetry is absent and, indeed, both theories have zero physical degrees of freedom.

A nonsingular BF-like zero coupling limit of the Yang–Mills theory for $n \geq 3$ can be obtained by introducing an auxiliary (Stückelberg) field $\eta \in \Omega^{n-3}(M, \mathfrak{g})$ which restores the $p$-form gauge invariance. The resulting BF Yang–Mills action is

$$S_{\mathrm{BFYM}} = \int_M \mathrm{tr}_G \Big[ \mathrm{i}B_{n-2} \wedge F_A + g^2 \left( B_{n-2} - \frac{1}{\sqrt{2}g} d_A \eta \right) \wedge * \left( B_{n-2} - \frac{1}{\sqrt{2}g} d_A \eta \right) \Big] \quad [50]$$

This action is not only invariant under ordinary $G$ gauge transformations, but also under the $p$-form gauge symmetry $B \to B + d_A \lambda$ [6] provided that $\eta$ transforms as $\eta \to \eta + \sqrt{2}g\lambda$. Thus, this shift can be used to set $\eta$ to zero, upon which one recovers the first-order form of the Yang–Mills action. Moreover, in the zero-coupling limit all that survives is a standard (and nontopological) minimal coupling of $\eta$ to the BF action:

$$\lim_{g^2 \to 0} S_{\mathrm{BFYM}} = \int_M \mathrm{tr}_G \left[ \mathrm{i}B_{n-2} \wedge F_A + \tfrac{1}{2} d_A \eta \wedge * d_A \eta \right] \quad [51]$$

accounting for the correct number of degrees of freedom of the Yang–Mills theory (the $(n-3)$-form $\eta$ being absent for $n=2$).

Two-dimensional quantum BF and Yang–Mills theories have a variety of interesting topological properties. An account of some of them can be found in Blau and Thompson (1994) and Witten (1991). For a detailed discussion of the gauge symmetries and gauge fixing of the BFYM action, see Cattaneo *et al.* (1998).

### Chern–Simons Theory, Gravity, and (Deformed) BF Theory

The Chern–Simons theory is a three-dimensional gauge theory. The Chern–Simons action for an $H$-connection $C$, $H$ the gauge group, is

$$S_{\mathrm{CS}}(C) = \int_M \mathrm{tr}_H \left( C \wedge \mathrm{d}C + \tfrac{2}{3} C \wedge C \wedge C \right) \quad [52]$$

It is invariant under the infinitesimal gauge transformations $\delta C = d_C \lambda$, $\lambda \in \Omega^0(M, \mathfrak{h})$, and the gauge-invariant equation of motion is the flatness condition $F_C = 0$. Now let $H = TG$ be the tangent bundle group $TG \sim G \times_s \mathfrak{g}$. This is a semidirect product group with $G$ acting on $\mathfrak{g}$ via the adjoint and $\mathfrak{g}$ regarded as an abelian Lie algebra of translations. Thus, in terms of generators $(J_a, P_a)$, where the $J_a$ are generators of $G$, the commutation relations are $[J_a, J_b] = f_{ab}^c J_c$, $[J_a, P_b] = f_{ab}^c P_c$ and $[P_a, P_b] = 0$, and the curvature of the $TG$-connection $C = J_a A^a + P_a B^a$ is

$$F_C = J_a F_A^a + P_a d_A B^a \quad [53]$$

Thus, the equations of motion of the $TG$ Chern–Simons theory are equivalent to the equations of motion [2] of the BF theory with gauge group $G$. This equivalence also holds at the level of the action:

$$\tfrac{1}{2} S_{\mathrm{CS}}(C) = S_{\mathrm{BF}}(A, B) \quad [54]$$

provided that one chooses the nondegenerate invariant scalar product to be

$$\begin{aligned} \mathrm{tr}_{TG}(J_a P_b) &= \mathrm{tr}_G(J_a J_b) \\ \mathrm{tr}_{TG}(J_a J_b) &= \mathrm{tr}_{TG}(P_a P_b) = 0 \end{aligned} \quad [55]$$

For $G = \mathrm{SO}(3)$, $TG$ is the Euclidean group of isometries of $\mathbb{R}^3$ and for $G = \mathrm{SO}(2,1)$, $TG$ is the Poincaré group of isometries of the three-dimensional Minkowski space $\mathbb{R}^{2,1}$. For these gauge groups, the BF action takes the form of the three-dimensional (Euclidean or Lorentzian) Einstein–Hilbert action, with the interpretation of $B = e$ as the dreibein and $A = \omega$ as the spin connection. The equations of motion for $e$ and $\omega$ express the vanishing of the torsion and the Riemann tensor (equivalent to the vanishing of the Ricci tensor for $n=3$), respectively. This Chern–Simons interpretation of three-dimensional gravity extends to gravity with a cosmological constant, with $H$ the appropriate de Sitter or anti-de Sitter isometry group (SO(4), SO(3, 1), or SO(2, 2), depending on the signature and the sign of the cosmological constant). In terms of the BF interpretation, this corresponds to the simple topological deformation

$$S_{\mu\mathrm{BF}}(A, B) = \int_M \mathrm{tr}_G \left( B \wedge F_A + \tfrac{1}{3} \mu B \wedge B \wedge B \right) \quad [56]$$

of the BF action, which has the deformed local symmetries (cf. [5] and [6])

$$\delta A = d_A \lambda + \mu [B, \lambda'], \qquad \delta B = [B, \lambda] + d_A \lambda' \quad [57]$$

A simple way to understand these symmetries is to note that the action can be written as the difference of two Chern–Simons actions:

$$\begin{aligned} &S_{\mathrm{CS}}(A + \sqrt{\mu} B) - S_{\mathrm{CS}}(A - \sqrt{\mu} B) \\ &\quad = 4\sqrt{\mu} S_{\mu\mathrm{BF}}(A, B) \end{aligned} \quad [58]$$

whose evident standard local gauge symmetries $\delta(A \pm \sqrt{\mu} B) = d_{A \pm \sqrt{\mu} B} \lambda^{\pm}$ are equivalent to [57] for $\lambda^{\pm} = \lambda \pm \sqrt{\mu} \lambda'$.

A detailed account of three-dimensional classical and quantum gravity can be found in Carlip (1998).

### Relation with Gravity

Theories of two-dimensional gravity and topological gravity also have a BF formulation (Blau and Thompson 1991, Birmingham *et al.* 1991) which resembles the Chern–Simons BF formulation of three-dimensional gravity described above, the natural gauge group now being SO(2, 1) or SO(3) or one of its contractions.

In the first-order (Palatini) formulation, the Einstein–Hilbert action for four-dimensional gravity can be written as

$$S_{\mathrm{EH}} = \int \mathrm{tr}(e \wedge e \wedge F_\omega) \qquad [59]$$

where $e$ is the vierbein and $\omega$ is the spin connection. This action has the general form of a BF action with a constraint that $B = e \wedge e$ be a simple bi(co-)vector. Thus, four-dimensional general relativity can be regarded as a constrained BF theory. Although this constraint drastically changes the number of physical degrees of freedom (BF theory has zero degrees of freedom, while four-dimensional gravity has two), this is nevertheless a fruitful analogy which also lies at the heart of the spin-foam quantization approach to quantum gravity. This constrained BF description of gravity is also available for higher-dimensional gravity theories.

For further details, and references, see Freidel *et al.* (1999) and the review article (Baez 2000).

### Knot and Generalized Knot Invariants

The known relationship between Wilson loop observables of the Chern–Simons theory with a compact gauge group and knot invariants (Witten 1989), and the interpretation of the three-dimensional BF theory as a Chern–Simons theory with a noncompact gauge group raise the question of the relation of observables of an $n = 3$ BF theory to knot invariants, and suggest the possibility of using an $n \geq 4$ BF theory to define higher-dimensional analogs of knot invariants. It turns out that an appropriate observable of $n = 3$ BF theory for $G = \mathrm{SU}(2)$ is related to the Alexander–Conway polynomial. The analysis of higher-dimensional BF theories requires the full power of the Batalin–Vilkovisky (BV) formalism. BV observables generalizing Wilson loops have been shown to give rise to cohomology classes on the space of imbedded curves.

For a detailed discussion of these issues, see Cattaneo and Rossi (2001) and references therein. A relation between the algebra of generalized Wilson loops and string topology has been investigated in Cattaneo *et al.* (2003).

*See also:* Batalin–Vilkovisky Quantization; BRST Quantization; Chern–Simons Models: Rigorous Results; Gauge Theories From Strings; Knot Invariants and Quantum Gravity; Loop Quantum Gravity; Moduli Spaces: An Introduction; Nonperturbative and Topological Aspects of Gauge Theory; Schwarz-Type Topological Quantum Field Theory; Spin Foams; Topological Quantum Field Theory: Overview.

## Further Reading


Baez J (2000) An introduction to spin foam models of quantum gravity and BF theory. *Lecture Notes in Physics* 543: 25–94.

Birmingham D, Blau M, Rakowski M, and Thompson G (1991) Topological field theory. *Physics Reports* 209: 129–340.

Blau M and Thompson G (1989) A New Class of Topological Field Theories and the Ray–Singer Torsion. *Physics Letters B* 228: 64–68.

Blau M and Thompson G (1991) Topological gauge theories of antisymmetric tensor fields. *Annals of Physics* 205: 130–172.

Blau M and Thompson G (1994) Lectures on 2d gauge theories: topological aspects and path integral techniques. In: Gava E, Masiero A, Narain KS, Randjbar–Daemi S, and Shafi Q (eds.) *Proceedings of the 1993 Trieste Summer School on High Energy Physics and Cosmology*, pp. 175–244. Singapore: World Scientific.

Carlip S (1998) *Quantum Gravity in* 2 + 1 *Diemensions*. Cambridge: Cambridge University Press.

Cattaneo A and Rossi C (2001) Higher-dimensional BF theories in the Batalin–Vilkovisky formalism: the BV action and generalized Wilson loops. *Communications in Mathematical Physics* 221: 591–657.

Cattaneo A, Cotta-Ramusino P, Fucito F, Martellini M, and Rinaldi M, *et al.* (1998) Four-dimensional Yang–Mills theory as a deformation of topological BF theory. *Communications in Mathematical Physics* 197: 571–621.

Cattaneo A, Pedrini P, and Fröhlich J (2003) Topological field theory interpretation of string topology. *Communications in Mathematical Physics* 240: 397–421.

Freidel L, Krasnov K, and Puzio R (1999) BF description of higher-dimensional gravity theories. *Advances in Theoretical and Mathematical Physics* 3: 1289–1324.

Horowitz GT (1989) Exactly soluble diffeomorphism invariant theories. *Communications in Mathematical Physics* 125: 417–437.

Schwarz AS (1978) The partition function of a degenerate quadratic functional and Ray–Singer Invariants. *Letters in Mathematical Physics* 2: 247–252.

Schwarz AS (1979) The partition function of a degenerate functional. *Communications in Mathematical Physics* 67: 1–16.

Witten E (1989) Quantum field theory and the Jones polynomial. *Communications in Mathematical Physics* 127: 351–399.

Witten E (1991) On quantum gauge theories in two dimensions. *Communications in Mathematical Physics* 141: 153–209.

# Bicrossproduct Hopf Algebras and Noncommutative Spacetime

**S Majid**, Queen Mary, University of London,
London, UK



## Introduction

One of the sources of quantum groups is a bicrossproduct construction coming in the case of Lie groups from considerations of Planck-scale physics in the 1980s. This article describes these objects and their currently known applications. See also the overview of Hopf algebras which provides the algebraic context (*see* Hopf Algebras and $q$-Deformation Quantum Groups).

The construction of quantum groups here is viewed as a microcosm of the problem of quantization in a manner compatible with geometry. Here quantization enters in the noncommutativity of the algebra of observables and "curvature" enters as a quantum nonabelian group structure on phase space. Among the main features of the resulting bicrossproduct models (Majid 1988) are

1. Compatibility takes the form of nonlinear "matched pair equations" generically leading to singular accumulation regions (event horizons or a maximum value of momentum depending on context).
2. The equations are solved in an "equal and opposite" form from local factorization of a larger object.
3. Different classical limits are related by observer–observed symmetry and Hopf algebra duality.
4. Nonabelian Born reciprocity re-emerges and is linked to $T$-duality.

It has also been argued that noncommutative geometry should emerge as an effective theory of the first corrections to geometry coming from any unknown theory of quantum gravity. Concrete models of noncommutative spacetime currently provide the first framework for the experimental verification of such effects. The most basic of these possible effects is curvature in momentum space or "cogravity." We start with this.

## Cogravity

We recall that curvature in space or spacetime means by definition noncommutativity among the covariant derivatives $D_i$. Here the natural momenta are $p_i = -i\hbar D_i$ and the situation is typified by the top line in Figure 1. There are also mixed relations between the $D_i$ and position functions as indicated

| | Position | Momentum |
|---|---|---|
| Gravity | Curved $\Sigma_\mu x_\mu^2 = \frac{1}{\gamma^2}$ | Noncommutative $[p_i, p_j] = i\hbar\gamma\epsilon_{ijk}p_k$ |
| Cogravity | Noncommutative $[x_i, x_j] = 2i\lambda\epsilon_{ijk}x_k$ | Curved $\Sigma_\mu P_\mu^2 = \frac{1}{\lambda^2}$ |
| Quantum mechanics | $[x_i, p_j] = i\hbar\delta_{ij}$ | |

**Figure 1** Noncommutative spacetime means curvature in momentum space. The equations are for illustration.

for flat space in the bottom line, which is quantum mechanics (there is a similar story for quantum mechanics on a curved space). We see however a third and dual possibility – noncommutativity in position space which should be interpreted as curvature in momentum space, that is, the dual of gravity. This is an independent physical effect and comes therefore with its own length scale which we denote $\lambda$. These ideas were made precise in the mid 1990s using the quantum group Fourier transform; see Majid (2000). Here we show what is involved on three illustrative examples.

1. We consider the "spin space" algebra

$$\mathbb{R}^3_\lambda : [x_i, x_j] = i2\lambda\epsilon_{ij}{}^k x_k$$

where ${\epsilon_{12}}^3 = 1$ and where it is convenient to insert a factor 2. This is the enveloping algebra $U(su_2)$, that is, just angular momentum space but now regarded "upside down" as a coordinate algebra (*see* Hopf Algebras and $q$-Deformation Quantum Groups). Then a plane wave is of the form

$$\psi_p = e^{ip\cdot x}, \quad p \in \mathbb{R}^3$$

where we set $\hbar = 1$ for this discussion. The momenta $p_i$ are nothing but local coordinates for the corresponding point $e^{i\lambda p\cdot\sigma} \in SU_2$ where $\lambda\sigma$ is the representation by Pauli matrices. It is really elements of this curved space $SU_2$ where momenta live. Here $\mathbb{R}^3_\lambda = U(su_2)$ has dual $\mathbb{C}[SU_2]$ and Hopf algebra Fourier transform (after suitable completion) takes one between these spaces. Thus, in one direction

$$\mathcal{F}(f) = \int_{SU_2} du f(u) u \approx \int d^3p J(p) f(p)\, e^{ip\cdot x}$$

for $f$ a function on $SU_2$. We use the Haar measure on $SU_2$. The local result on the right has $J$ the Jacobian for the change to the local $p$ coordinates and $f$ is written in terms of these. Note that the coproduct in

$\mathbb{C}[SU_2]$ in terms of the $p^i$ generators is an infinite series given by the Campbell–Baker–Hausdorff series, and not the usual linear one (this is why the measure is not the Lebesgue one). The physical content here is in the plane waves themselves, one can use any other momentum coordinates to parametrize them with the corresponding measure and coproduct. Differential operators on $\mathbb{R}^3_\lambda$ are given by the action of elements of $\mathbb{C}[SU_2]$ and are diagonal on these plane waves,

$$f.\psi_{\boldsymbol{p}} = f(\boldsymbol{p})\psi_{\boldsymbol{p}}$$

which corresponds under Fourier transform simply to pointwise multiplication in $\mathbb{C}[SU_2]$. For example, the function $\lambda^{-2}(\mathrm{tr}-2)$ as a function on $SU_2$ will give a rotationally invariant wave operator which is also invariant under inversion in the group. Its value on plane waves is

$$\frac{1}{\lambda^2}\mathrm{tr}(\mathrm{e}^{\mathrm{i}\lambda \boldsymbol{p}\cdot\sigma}-1) = \frac{2}{\lambda^2}(\cos(\lambda|\boldsymbol{p}|)-1)$$

In the limit $\lambda \to 0$ this gives the usual wave operator on $\mathbb{R}^3$.

It is also possible to put a differential graded algebra (DGA) structure of differential forms on this algebra, the natural one being

$$\mathrm{d}x_i = \lambda\sigma_i, \qquad x_i\theta - \theta x_i = \mathrm{i}\frac{\lambda^2}{\mu}\,\mathrm{d}x_i$$

$$(\mathrm{d}x_i)x_j - x_j\mathrm{d}x_i = \mathrm{i}\lambda\epsilon_{ij}{}^k\mathrm{d}x_k + \mathrm{i}\mu\delta_{ij}\theta$$

where $\theta$ is the $2\times 2$ identity matrix which, together with the Pauli matrices $\sigma_i$, completes the basis of left-invariant 1-forms. The 1-form $\theta$ provides a natural time direction, even though there is no time coordinate, and the new parameter $\mu \neq 0$ appears as the freedom to change its normalization. The partial derivatives $\partial^i$ are defined by

$$\mathrm{d}\psi(x) = (\partial^i\psi)\mathrm{d}x_i + (\partial^0\psi)\theta$$

and act diagonally on plane waves as

$$\partial^i = \frac{\mathrm{i}}{2\lambda}\mathrm{tr}(\sigma_i(\ )) = \mathrm{i}\frac{p^i}{\lambda|\boldsymbol{p}|}\sin(\lambda|\boldsymbol{p}|)$$

while $\partial^0 = \mathrm{i}\mu(\mathrm{tr}-2)/2\lambda^2$ is computed as above.

Note that $\mu$ cannot be taken to be zero due to an anomaly for translation invariance of the DGA. It is in fact a typical feature of noncommutative differential geometry that there is a 1-form $\theta$ generating d by commutator which can be required as an extra cotangent direction with its associated partial derivative an induced Hamiltonian. In the present model we have

$$\partial^0\psi = \mathrm{i}\frac{\mu}{2}\sum_i(\partial^i)^2\psi + O(\lambda^2)$$

which is of the form of Schrödinger's equation with respect to an auxiliary time variable and for a particle with mass $1/\mu$.

The reader may ask what happens to the Euclidean group of translations and rotations in this context. From the above we find that $U_\lambda(\mathrm{poinc}_3) = \mathbb{C}[SU_2] \bowtie U(su_2)$, the semidirect product generated by translations $\partial^i$ and usual rotations. This in turn is the quantum double $D(U(su_2))$ of the classical enveloping algebra, and as such a quantum group with braiding etc. (*see* Hopf Algebras and $q$-Deformation Quantum Groups). This quantum double has been identified as part of an effective theory in $2+1$ quantum gravity in a Euclidean version based on Chern–Simons theory with Lie algebra $\mathrm{poinc}_3$ and the spin space algebra proposed as an effective theory for this. The quotient of $\mathbb{R}^3_\lambda$ by an allowed value of the quadratic Casimir $\boldsymbol{x}^2$ (which then makes it a matrix algebra) is called a "fuzzy sphere" and appears as a "world-volume algebra" in certain string theories and reduced matrix models. The noncommutative differential geometry that we have described is due to Batista and the author.

2. We take the same type of construction to obtain the "bicrossproduct model" spacetime algebra

$$\mathbb{R}^{1,3}_\lambda: \qquad [t, x_i] = \mathrm{i}\lambda x_i, \quad [x_i, x_j] = 0$$

These are the relations of a Lie algebra $b_+$ (say) but again regarded as coordinates on a noncommutative spacetime. Here $\lambda$ is a timescale which can be written as a mass scale $\kappa = 1/\lambda$ instead. We parametrize the plane waves as

$$\psi_{\boldsymbol{p},p^0} = \mathrm{e}^{\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{x}}\mathrm{e}^{\mathrm{i}p^0 t}, \qquad \psi_{\boldsymbol{p},p^0}\psi_{\boldsymbol{p}',p^{0\prime}} = \psi_{\boldsymbol{p}+\mathrm{e}^{-\lambda p^0}\boldsymbol{p}',p^0+p^{0\prime}}$$

which identifies the $p^\mu$ as the coordinates of the nonabelian group $B_+ = \mathbb{R} \rhd\!\!\!< _\lambda \mathbb{R}^3$ with Lie algebra $b_+$. The group law in these coordinates is read off as usual from the product of plane waves, which also gives the coproduct of $\mathbb{C}[B_+]$ on the $p^\mu$. We have parametrized plane waves in this way (rather than the canonical way by the Lie algebra as before) in order to have a more manageable form for this. We do pay a price that in these coordinates group inversion is not simply $-p^\mu$, but

$$(\boldsymbol{p}, p^0)^{-1} = (-\mathrm{e}^{\lambda p^0}\boldsymbol{p}, -p^0)$$

which is also the action of the antipode $S$ on the abstract $p^\mu$ generators.

In particular, the right-invariant Haar measure on $B_+$ in these coordinates is the usual $\mathrm{d}^4 p$ so the

quantum group Fourier transform reduces to the usual one but normal ordered,

$$\mathcal{F}(f) = \int_{\mathbb{R}^4} \mathrm{d}^4 p\, f(p) \mathrm{e}^{\mathrm{i}p\cdot x} \mathrm{e}^{\mathrm{i}p^0 t}$$

(one can also Fourier transform with respect to the left-invariant measure $\mathrm{d}^4 p\, \mathrm{e}^{3\lambda p^0}$ on $B_+$). The inverse is again given in terms of the usual inverse transform if we specify general fields $\psi$ in $\mathbb{R}^{1,3}_\lambda$ by normal ordering of usual functions, which we shall do. As before, the action of elements of $\mathbb{C}[B_+]$ defines differential operators on $\mathbb{R}^{1,3}_\lambda$ and these act diagonally on plane waves.

We also have a natural DGA with

$$(\mathrm{d}x_j)x_\mu = x_\mu \mathrm{d}x_j, \qquad (\mathrm{d}t)x_\mu - x_\mu \mathrm{d}t = \mathrm{i}\lambda \mathrm{d}x_\mu$$

which leads to the partial derivatives

$$\partial^i \psi =: \frac{\partial}{\partial x_i}\psi(x,t) := \mathrm{i}p^i.\psi$$

$$\partial^0 \psi =: \frac{\psi(x, t+\mathrm{i}\lambda) - \psi(x,t)}{\mathrm{i}\lambda} := \frac{\mathrm{i}}{\lambda}(1 - \mathrm{e}^{-\lambda p^0}).\psi$$

for normal-ordered polynomial functions $\psi$ or in terms of the action of the coordinates $p^\mu$ in $\mathbb{C}[B_+]$. These $\partial^\mu$ do respect our implicit $*$-structure (unitarity) on $\mathbb{R}^{1,3}_\lambda$ but in a Hopf algebra sense which is not the usual sense, since the action of the antipode $S$ is not just $-p^\mu$. This can be remedied by using adjusted derivatives $L^{-(1/2)}\partial^\mu$ where

$$L\psi =: \psi(x, t+\mathrm{i}\lambda) := \mathrm{e}^{-\lambda p^0}.\psi$$

In this case the natural 4D Laplacian is $L^{-1}((\partial^0)^2 - \sum_i (\partial^i)^2)$, which acts on plane waves as

$$-\frac{2}{\lambda^2}(\cosh(\lambda p^0) - 1) + \boldsymbol{p}^2 \mathrm{e}^{\lambda p^0}$$

where

$$\boldsymbol{p}^2 = \sum_{i=1}^{3} p_{i^2}$$

This deforms the usual Laplacian in such a way as to remain invariant under the Lorentz group (which now acts nonlinearly on $B_+$ in this model) and under group inversion.

This model may provide the first experimental test for noncommutative spacetime and cogravity. For the analysis of an experiment, we assume the identification of noncommutative waves in the above normal-ordered form with classical ones that a detector might register. In that case one may argue (Amelino-Camelia and Majid 2000) that the dispersion relation for such waves has the classical derivation as $\partial p^0/\partial p^i$ which now computes as propagation speed for a massless particle:

$$\left|\frac{\partial p^0}{\partial \boldsymbol{p}}\right| = \mathrm{e}^{\lambda p^0}$$

in units where 1 is the usual speed of light. So the prediction is that the speed of light depends on energy. What is remarkable is that even if $\lambda \sim 10^{-44}$ s (the Planck timescale), this prediction could in principle be tested, for example using $\gamma$-ray bursts. These are known in some cases to travel cosmological distances before arriving on Earth, and have a spread of energies from 0.1–100 MeV. According to the above, the relative time delay $\Delta_t$ on traveling distance $L$ for frequencies corresponding to $p^0, p^0 + \Delta_{p^0}$ is

$$\Delta_t \sim \lambda \Delta_{p^0} \frac{L}{c} \sim 10^{-44}\mathrm{s} \times 100\,\mathrm{MeV} \times 10^{10}\mathrm{y} \sim 1\,\mathrm{ms}$$

which is in principle observable by statistical analysis of a large number of bursts correlated with distance (determined, e.g., by using the Hubble telescope to lock in on the host galaxy of each burst). Although the above is only one of a class of predictions, it is striking that even Planck-scale effects are now in principle within experimental reach.

We now explain what happens to the full Poincaré symmetry here. The nonlinear action of the Lorentz group on $B_+$ Fourier transforms to an action on the generators of $\mathbb{R}^{1,3}_\lambda$, which combines with the above action of the $p^\mu$ to generate an entire Poincaré quantum group $U(so_{1,3}) \triangleright\!\!\blacktriangleleft \mathbb{C}[B_+]$. We will say more about its "bicrossproduct" structure in a later section. The above wave operator in momentum space is the natural Casimir in these momentum coordinates. A common mistake in the literature for this model is to suppose that the Casimir relation alone amounts to a physical prediction, whereas in fact the momentum coordinates are arbitrary and have meaning only in conjunction with the plane waves that they parametrize. The deformed Poincaré as an algebra alone is actually isomorphic to the undeformed one by a different choice of generators, so by itself has no physical content; one needs rather the noncommutative spacetime as well. Prior work on the relevant deformed Poincaré algebra either did not consider it acting on spacetime or took it acting on classical (commutative) Minkowski spacetime with inconsistent results (there is no such action as a quantum group).

The above model was introduced by Majid and Ruegg (1994) and later tied up with a dual approach of Woronowicz. There is also a previous "$\kappa$-Poincaré" version of the Hopf algebra alone obtained (Lukierski *et al.* 1991) in another context (by contraction of $U_q(so_{2,3})$) but with fundamentally different generators and relations and hence different physical content (e.g., the Lorentz

generators there do not close among themselves but mix with momentum).

3. The usual Heisenberg algebra of quantum mechanics is another possible noncommutative (phase) space; one may also take the same algebra and view it as a noncommutative spacetime, so:

$$\mathbb{R}^{1,3}_\theta: \quad [x_\mu, x_\nu] = \mathrm{i}\theta_{\mu\nu}$$

for any antisymmetric tensor $\theta_{\mu\nu}$. This is not a Hopf algebra but it turns out that this model can also be completely solved by Hopf algebra methods, namely the theory of covariant twists. Twist models also include versions of the noncommutative torus studied by Connes, and related $\theta$-spaces, which are nontrivial at the level of $C^*$-algebras. However, at an algebraic level, all covariant structures are automatically provided by applying the twisting functor $\mathcal{T}$ to the desired classical construction (*see* Hopf Algebras and $q$-Deformation Quantum Groups). This is not usually appreciated in the physics literature on such models, but see Oeckl (2000).

Thus, consider $H = U(\mathbb{R}^{1,3})$ with generators $p^\mu = -\mathrm{i}\partial^\mu$ acting as usual on functions on Minkowski space. It has a cocycle

$$F = \mathrm{e}^{(\mathrm{i}/2)p^\mu \otimes p^\nu \theta_{\mu\nu}}$$

which induces a new product $\bullet$ on functions by $\phi \bullet \psi = \cdot(F^{-1}(\phi \otimes \psi))$. This is just the standard Moyal product, in the present case on $\mathbb{R}^{1,3}$, viewed as a covariant twist using Hopf algebra methods. The Hopf algebra $U(\mathbb{R}^{1,3})$ in principle has a twisted coproduct given by $\Delta_F = F(\Delta(\ ))F^{-1}$ but this does not change as the algebra is commutative.

Next, $H$ also acts covariantly on $\Omega(\mathbb{R}^{1,3})$, the usual algebra of differential forms, and twisting this in the same way gives

$$\psi(x) \bullet \mathrm{d}x_\mu = \psi \mathrm{d}x_\mu = (\mathrm{d}x_\mu)\psi = (\mathrm{d}x_\mu) \bullet \psi$$

unchanged. This is because no terms higher than $p^\mu \otimes p^\nu \theta_{\mu\nu}$ contribute and then $\mathrm{d}(1) = 0$. The associated partial derivatives defined by d are likewise unchanged and act in the usual way as derivations with respect to both the $\bullet$ product and the undeformed product. The result may look different when the same $\psi(x)$ is expressed as a function of the variables with the $\bullet$ product. In other words, the only deformation comes from the Moyal product itself, with the rest being automatic. Moreover, the plane waves themselves are unchanged because $(x \cdot k)^{\bullet n} = (x \cdot k)^n$ due to $\theta$ being antisymmetric. Hence,

$$\psi_k(x) = \mathrm{e}_\bullet^{\mathrm{i}x\cdot k} = \mathrm{e}^{\mathrm{i}x.k}, \qquad p^\mu \psi_k(x) = k^\mu \psi_k(x)$$

where $p^\mu = -\mathrm{i}\partial^\mu$. The wave operator $-\partial_\mu\partial^\mu$ is therefore given by the action of $p_\mu p^\mu$ and has value $k_\mu k^\mu$ as usual on plane waves. On the other hand,

$$\psi_k \bullet \psi_{k'} = \mathrm{e}^{(\mathrm{i}/2)k^\mu k'^\nu \theta_{\mu\nu}} \psi_{k+k'}$$

or in algebraic terms the twist functor $\mathcal{T}$ applied to the Fourier transform implies also a twisted coproduct or coaddition law for the abstract $k^\mu$ generators, now different from the linear one for the covariance momentum operators $p^\mu$. This leads to some of the more interesting features of the model.

One immediately also has a Poincaré quantum group here, $U_\theta(\mathrm{poinc}_{1,3})$, obtained by similarly twisting the classical $U(\mathrm{poinc}_{1,3})$. We just view $F$ as living here rather than in the original $H$. The translation sector is unchanged as before but if $M^{\alpha\beta}$ are the usual Lorentz generators, then

$$\begin{aligned}\Delta_F M^{\alpha\beta} = {}& M^{\alpha\beta} \otimes 1 + 1 \otimes M^{\alpha\beta} \\ & + \tfrac{1}{2}(p^\alpha \otimes \theta^\beta{}_\mu p^\mu - \theta^\beta{}_\mu p^\mu \otimes p^\alpha) \\ & - \tfrac{1}{2}(p^\beta \otimes \theta^\alpha{}_\mu p^\mu - \theta^\alpha{}_\mu p^\mu \otimes p^\beta)\end{aligned}$$

using the metric $\eta_{\mu\nu}$ to raise or lower indices. The antipode is also modified according to the theory in Majid (1995). The relations in the Poincaré algebra are not modified (so, e.g., $p_\mu p^\mu$ will remain central). Any construction originally Poincaré covariant becomes covariant under this twisted one after application of the twisting functor. As with the differentials above, the action on $\mathbb{R}^{1,3}_\theta$ is not actually modified but may appear so when functions are expressed in terms of the $\bullet$ product.

The above model is popular at the time of writing in connection with string theory. Here, an effective description of the endpoints of open strings landing on a fixed 4-brane has been modeled conveniently in terms of the $\bullet$ product above (Seiberg and Witten 1999). It should be borne in mind, however, that this fixed 4-brane lives in some of the higher dimensions of the string spacetime, so this is not necessarily a prediction of noncommutative spacetime $\mathbb{R}^{1,3}$.

In fact, a proposal superficially similar to $\mathbb{R}^{1,3}_\theta$ above was already proposed in Snyder (1947). Here

$$[x^\mu, x^\nu] = \mathrm{i}\lambda^2 M^{\mu\nu}$$

where $\lambda$ is our length scale and the $M^{\mu\nu}$ are now operators with the usual commutation rules for the Lorentz algebra with themselves and with $x^\mu$ and the momenta $p^\mu$. The latter obey

$$[p^\mu, x^\nu] = \mathrm{i}(\eta^{\mu\nu} - \lambda^2 p^\mu p^\nu), \quad [p^\mu, p^\nu] = 0$$

so the entire Poincaré algebra is undeformed but the phase-space relations are deformed. Snyder also constructed the orbital angular momentum realization $M^{\mu\nu}=x^\mu p^\nu - x^\nu p^\mu$. This model is not a proposal for a noncommutative spacetime because the algebra does not even close among the $x^\mu$. Rather it is a proposal for "mixing" of position and Lorentz generators. On the other hand (which was the point of view in Snyder (1947)), in any representation of the Poincaré algebra, the $M^{\mu\nu}$ become operators and in some sense numerical. The rotational sector has discrete eigenvalues as usual, so to this extent the spacetime has been discretized. Although not fitting into the methods in this article, it is also of interest that the relations above were motivated by considering $p^\mu$ as coordinates projected from a 5D flat space to de Sitter space and $x^\mu$ as the 5-component of orbital angular momentum in the flat space.

To conclude this section, let us note that there are further models that we have not included for lack of space. One of them is a much-studied $\mathbb{R}_q^{1,3}$ in which $t$ is central but the $x_i$ enjoy complicated $q$-relations best understood as $q$-deformed Hermitian matrices. One of the motivations in the theory was the result in Majid (1990) that $q$-deformation could be used to regularize infinities in quantum field theory as poles at $q=1$. Another entire class is to use noncommutative geometry and quantum group methods on finite or discrete spaces. Unlike lattice theory where a finite lattice is viewed as approximation, these models are not approximations but exact noncommutative geometries valid even on a few points. The noncommutativity enters into the fact that finite differences are bilocal and hence naturally have different left and right multiplications by functions. Both aspects are mentioned briefly in the overview article (*see* Hopf Algebras and $q$-Deformation Quantum Groups). Also, on the experimental front, another large area that we have not had room to cover is the prediction of modified uncertainty relations both in spacetime and phase space (Kempf *et al.* 1995).

Moreover, for all of the models above, once one has a noncommutative differential calculus one may proceed to gauge theory etc., on noncommutative spacetimes, at least at the level where a connection is a noncommutative (anti-Hermitian) 1-form $\alpha$. Gauge transformations are invertible (unitary) elements $u$ of the noncommutative "coordinate algebra" and the connection and curvature transform as

$$\alpha \to u^{-1}\alpha u + u^{-1}\,\mathrm{d}u$$
$$F(\alpha) = \mathrm{d}\alpha + \alpha\wedge\alpha \to u^{-1}F(\alpha)u$$

The full extent of quantum bundles and gravity (*see* Quantum Group Differentials, Bundles and Gauge Theory) and quantum field theory is not always possible, although both have been done for covariant twist examples (for functorial reasons) and for small finite sets. For the first two models above, for example, it is not clear at the time of writing how to interpret scattering when the addition of momenta is nonabelian.

## Matched Pair Equations

Although we have presented noncommutative spacetime first, the first actual application of quantum group methods to Planck-scale physics was the Planck-scale Hopf algebra obtained by a theory of bicrossproducts. Like the Snyder model, the intention here was to deform phase space itself, but since then bicrossproducts have had many further applications. The main ingredient here is the notion of a pair of groups $(G, M)$, say, acting on each other as we explain now. The mathematics here goes back to the early 1910s in group theory, but also arose in mathematical physics as a toy version of Einstein's equation in the sense of compatibility between quantization and curvature (see the next section).

By definition, $(G, M)$ are a matched pair of groups if there are left and right actions

$$M \overset{\triangleleft}{\longleftarrow} M\times G \overset{\triangleright}{\rightarrow} G$$

of each group on the set of the other, such that

$$s\triangleleft e = s,\quad e\triangleright u = u,\quad s\triangleright e = e,\quad e\triangleleft u = e$$
$$(s\triangleleft u)\triangleleft v = s\triangleleft(uv),\ s\triangleright(t\triangleright u) = (st)\triangleright u$$
$$s\triangleright(uv) = (s\triangleright u)((s\triangleleft u)\triangleright v)$$
$$(st)\triangleleft u = (s\triangleleft(t\triangleright u))(t\triangleleft u)$$

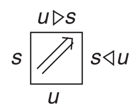
for all $u, v \in G, s, t \in M$. Here $e$ denotes the relevant group unit element. As a first application of such data, one may make a "double cross product group" $G\bowtie M$ with product

$$(u,s).(v,t) = (u(s\triangleright v), (s\triangleleft v)t)$$

and with $G$, $M$ as subgroups. Since it is built on the direct product space, the bigger group factorizes into these subgroups. Conversely, if $X$ is a group factorization such that the product $G\times M\to X$ is bijective, each group acts on the other by actions $\triangleright, \triangleleft$ defined by $su=(s\triangleright u)(s\triangleleft u)$ for $u\in G$ and $s\in M$, where $s$, $u$ are multiplied in $X$ and the product is factorized as something in $G$ and something in $M$. So finite group matched pairs are equivalent to group factorizations. In the Lie group context, the

corresponding system of differential equations is equivalent to a local factorization.

There is a nice graphical representation of the matched pair conditions which relates to "surface integration." Thus, consider squares



labeled by elements of $M$ on the left edge and elements of $G$ on the bottom edge. We can fill in the other two edges by thinking of an edge transformed by the other edge as it goes through the square either horizontally or vertically, the two together is the surface transport $\Rightarrow$ across the square. The matched pair equations have the meaning that a square can be subdivided either vertically or horizontally as shown in Figure 2, where the labeling on vertical edges is to be read from top down. The transport operation here is nothing other than normal ordering in the factorizing group. In the Lie setting, it means that the equations can be solved from infinitesimal solutions (a matched pair of Lie algebras) by a simultaneous double integration over the group (i.e., building up a large box from many small ones). If one considers solving the quantum Yang–Baxter equations on groups, they appear in this notation as an equality of surface transport going two ways around a cube, and the classical Yang–Baxter equations as curvature of the underlying higher-order connection.

Also in this notation there is a bicrossproduct quantum group defined in Figure 3, at least when $M$ is finite. The expressions are considered zero unless the juxtaposed edges have the same group labels. In that case, the product is a semidirect product algebra $\mathbb{C}(M){>\!\!\triangleleft}\mathbb{C}G$ of functions on $M$ by the group algebra of $G$. The coproduct is the adjoint of

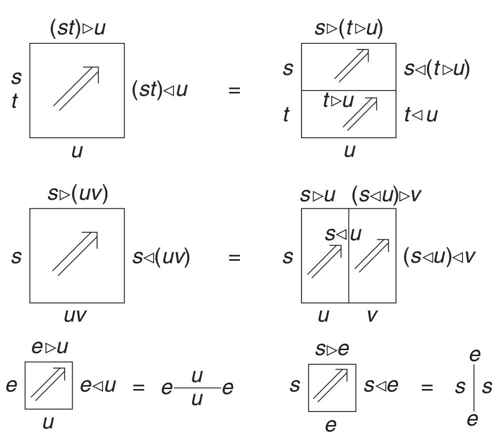


**Figure 2** Matched pair condition as a subdivision property.

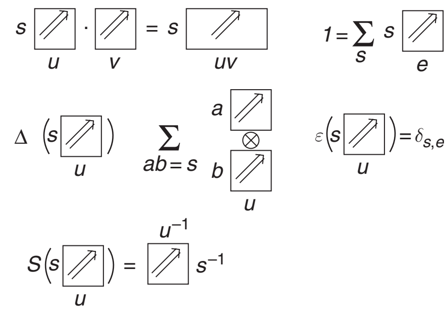


**Figure 3** Bicrossproduct Hopf algebra showing horizontal product and vertical coproduct as an "unproduct."

this, so is a semidirect coalgebra $\mathbb{C}(M){\blacktriangleright\!\!<}\mathbb{C}G$. Hence the two together are denoted $\mathbb{C}(M){\blacktriangleright\!\!\triangleleft}\mathbb{C}G$. The dual needs $G$ finite and has the same form but with vertical and horizontal compositions interchanged, that is, a bicrossproduct $\mathbb{C}M{\triangleright\!\!\blacktriangleleft}\mathbb{C}(G)$. Both Hopf algebras have the above labeled squares as basis.

It is possible to generalize both bicrossproducts and double cross products associated to matched pairs to general Hopf algebras $H_1{\blacktriangleright\!\!\triangleleft}H_2$ and $H_1{\bowtie}H_2$, respectively, where $H_1, H_2$ are Hopf algebras (*see* Majid 1990) and to relate the two in general by dualization of one factor. Another general result (Majid 1995) is that $H_1{\blacktriangleright\!\!\triangleleft}H_2$ acts covariantly on the algebra $H_1^*$ from the right, or $H_1{\triangleright\!\!\blacktriangleleft}H_2$ acts covariantly on $H_2^*$ from the left. A third general result is that bicrossproducts solve the extension problem

$$H_1 \to H \to H_2$$

meaning that such a Hopf algebra $H$ subject to some technical requirements (such as an algebra splitting map $H_2 \to H$) is of the form $H \cong H_1{\blacktriangleright\!\!\triangleleft}H_2$. The theory was also extended to include cocycle bicrossproducts at the end of the 1980s (by the author). The finite group case, however, was first found by Kac and Paljutkin (1966) in the Russian literature and later rediscovered independently in Takeuchi (1981) and in the course of Majid (1988).

## The Planck-Scale Hopf Algebra

We consider a quantum algebra of observables $H$ and ask when it is a Hopf algebra extending some classical position coordinate algebra $\mathbb{C}[M]$ and some possibly noncommutative momentum coordinate algebra $U(\mathfrak{g})$ in the form of a strict extension

$$\mathbb{C}[M] \to H \to U(\mathfrak{g})$$

From the theory above this problem is governed by local solutions of the matched pair equations on $(G, M)$. It requires that $H \cong \mathbb{C}[M]{>\!\!\triangleleft}U(\mathfrak{g})$ as an algebra, that is,

the quantization of a particle moving on orbits in $M$ under some action of $G$ (in an algebraic setting, or one can use von Neumann or $C^*$-algebras etc.). And it requires the classical phase space to be a nonabelian or "curved" group $M \bowtie \mathfrak{g}^*$. This extends to a coproduct on $H$ which becomes the bicrossproduct Hopf algebra $\mathbb{C}[M] \blacktriangleright\!\!\triangleleft U(\mathfrak{g})$. In this way, the problem which was open at the start of the 1980s of finding true examples of Hopf algebras was given a physical interpretation as being equivalent to finding quantum-mechanical systems reconciled with curvature, and the equations that governed this were the matched pair ones (Majid 1988).

We still have to solve these equations. In the Lie case, they mean a pair of cross-coupled first-order equations on $G \times M$. These can be solved locally as a double-holonomy construction in line with the surface transport point of view, but are nonlinear typically with singularities in the noncompact case. The equations are also symmetric under interchange of $G, M$ so Born reciprocity between position and momentum is extended to the quantum system with generally "curved" position and momentum spaces. Moreover, in so far as Einstein's equation $G_{\mu\nu} = 8\pi T_{\mu\nu}$ is also a compatibility between a quantity in position space and a quantity originating (ultimately) in momentum space, the matched pair equations can be viewed as a toy version of these.

Let us note that the reason to look for $H$ a Hopf algebra in the first place, aside from the reasons already given, is for observer–observed symmetry (this was put forward as a postulate for Planck-scale physics). Thus, $H^*$ is also an algebra of observables of some dual system, in our case $U(\mathfrak{m}) \triangleright\!\!\blacktriangleleft \mathbb{C}[G]$ or particles in $G$ moving on orbits under $M$. Thus, Born reciprocity is truly implemented in the quantum/curved system by Hopf algebra duality. Put another way, Hopf algebras are the simplest objects after abelian groups that admit Fourier transform (*see* Hopf Algebras and $q$-Deformation Quantum Groups) and we require this on phase space if Born reciprocity is to be extended to the quantum/curved system.

The Planck-scale Hopf algebra is the simplest example of these ideas (Majid 1988). Here $G = M = \mathbb{R}$ and the matched pair equations can be solved completely. The general solution is

$$\hat{p} = i\hbar(1 - e^{-\gamma x})\frac{\partial}{\partial x}, \qquad \hat{x} = \frac{i}{\hbar}(1 - e^{-\hbar\gamma p})\frac{\partial}{\partial p}$$

for the action of one group with generator $p$ on functions of $x$ in the other group and vice-versa. It has two parameters which we have denoted as $\hbar$ and a background curvature scale $\gamma$, and the corresponding bicrossproduct $\mathbb{C}[p] \triangleright\!\!\blacktriangleleft \mathbb{C}[x]$ is

$$\begin{aligned} &[p, x] = i\hbar(1 - e^{-\gamma x}), \qquad \Delta x = x \otimes 1 + 1 \otimes x \\ &\Delta p = p \otimes e^{-\gamma x} + 1 \otimes p, \quad \epsilon x = \epsilon p = 0 \\ &Sx = -x, \qquad Sp = -pe^{\gamma x} \end{aligned}$$

where we should allow power series or take $e^{\gamma x}$ as an invertible generator.

It is important to note that the matched pair equations here have only this solution and it is necessarily singular at $p = 0$ or $x = 0$. The interpretation in position space is as follows. Consider an infalling particle of mass $m$ with fixed momentum $p = mv_\infty$ (in terms of the velocity at infinity). By definition, $p$ is the free-particle momentum and acts on $\mathbb{R}$ as above. This corresponds to a free-particle Hamiltonian $\hat{p}^2/2m$ and induces

$$\begin{aligned} &\dot{p} = 0 \\ &\dot{x} = \frac{p}{m}(1 - e^{-\gamma x}) = v_\infty\left(1 - \frac{1}{1 + \gamma x + \cdots}\right) \end{aligned}$$

at the classical level. We see that the particle takes an infinite time to reach the origin, which is an accumulation point. This can be compared with the formula in standard radial infalling coordinates

$$\dot{x} = v_\infty\left(1 - \frac{1}{1 + \frac{c^2 x}{2GM}}\right)$$

for distance $x$ from the event horizon of a black hole of mass $M$ (here $G$ is Newton's constant and $c$ the speed of light). So $\gamma \sim c^2/GM$ and for the sake of further discussion we will use this value. With a little more work, one can then see that

$$\begin{array}{l} mM \ll m_{\rm P}^2 \\ \mathbb{C}[x] \blacktriangleright\!\!\triangleleft \mathbb{C}[p] \begin{array}{l} \nearrow \mathbb{C}[x]\,\mathbb{C}[p] \text{usual qu. mech.} \\ \searrow \mathbb{C}(X) \text{usual curved geometry} \end{array} \\ mM \gg m_{\rm P}^2 \end{array}$$

where $m_{\rm P}$ is the Planck mass of the order of $10^{-5}$ g and $X = \mathbb{R} \triangleright\!\!< \mathbb{R}$ is a nonabelian group. In the first limit, the particle motion is not detectably different from usual flat space quantum mechanics outside the Compton wavelength from the origin. In the second limit, the estimate is such that noncommutativity would not show up for length scales much larger than the background curvature scale.

This Hopf algebra is also the simplest way to extend classical position $\mathbb{C}[x]$ and momentum $\mathbb{C}[p]$ in the sense above. In other words, requiring to maintain observer–observed symmetry or Born reciprocity throws up both quantum mechanics (in the form of $\hbar$) and something with the flavor of

gravity (in the form of $\gamma$) and both are required for a nontrivial Hopf algebra. Moreover, the construction necessarily has a self-dual form and indeed the dually paired Hopf algebra is $\mathbb{C}[p]\triangleright\!\!\blacktriangleleft\mathbb{C}[x]$ with new parameters $\hbar' = 1/\hbar$ and $\gamma' = \hbar\gamma$ if we take the standard pairing $x, p$ across the two algebras. Hopf algebra duality realized by the quantum group Fourier transform $\mathcal{F}$ takes one between the two models.

## Bicrossproduct Poincaré Quantum Groups

Another example from the 1980s in the same family as the Planck-scale Hopf algebra is $G = SU_2$ and $M = B_+$, a nonabelian version of $\mathbb{R}^3$ with Lie algebra $b_+$ of the form

$$[x_3, x_i] = \mathrm{i}\lambda x_i, \qquad [x_i, x_j] = 0$$

for $i = 1, 2$. The required solution of the matched pair equations was found in Majid (1990) and has a nonlinear action of rotations on $B_+$. The interpretation of $\mathbb{C}[B_+]\blacktriangleright\!\!\triangleleft U(su_2)$ is of particles moving along orbits which are deformed spheres in $B_+$, and there is a dual model where particles move instead on orbits in $SU_2$ under the action of $b_+$. Moreover, from the general theory of bicrossproducts, we automatically have a covariant action of $\mathbb{C}[B_+]\blacktriangleright\!\!\triangleleft U(su_2)$ on the auxiliary noncommutative space $\mathbb{R}^3_\lambda = U(b_+)$ with relations as above.

The quantum group here was actually obtained as a Hopf–von Neumann algebra but we limit ourselves to the underlying algebraic version. Also, there is of course nothing stopping one considering this Hopf algebra equally well as $U_\lambda(\mathrm{poinc}_3)$, that is, a deformation of the group of motions on $\mathbb{R}^3$, rather than as an algebra of observables. The only difference is to denote the generators of $\mathbb{C}[B_+]$ by the symbols $p^i$, reserving $x_i$ instead for the auxiliary noncommutative space. We lower $i, j, k$ indices using the Euclidean metric. Then the bicrossproduct has the form

$$[p_i, p_j] = 0, \qquad [M_i, M_j] = \mathrm{i}\epsilon_{ij}{}^k M_k$$
$$[M_3, p_j] = \mathrm{i}\epsilon_{3j}{}^k p_k, \qquad [M_i, p_3] = \mathrm{i}\epsilon_{i3}{}^k p_k$$

as usual, for $i, j = 1, 2, 3$, and the modified relations

$$[M_i, p_j] = \frac{\mathrm{i}}{2}\epsilon_{ij}{}^3\left(\frac{1 - \mathrm{e}^{-2\lambda p_3}}{\lambda} - \lambda \boldsymbol{p}^2\right) + \mathrm{i}\lambda\epsilon_i{}^{k3} p_j p_k$$

for $i, j = 1, 2$ and $\boldsymbol{p}^2 = p_1^2 + p_2^2$. The coproducts are

$$\Delta M_i = M_i \otimes \mathrm{e}^{-\lambda p_3} + \lambda M_3 \otimes p_i + 1 \otimes M_i$$
$$\Delta p_i = p_i \otimes \mathrm{e}^{-\lambda p_3} + 1 \otimes p_i$$

for $i = 1, 2$ and the usual additive ones for $p_3, M_3$. There is also an appropriate counit and antipode. The deformed spheres under the nonlinear rotation in Majid (1990) are constant values of the Casimir for the above algebra. This is

$$\frac{2}{\lambda^2}(\cosh(\lambda p_3) - 1) + \boldsymbol{p}^2 \mathrm{e}^{\lambda p_3}$$

which from the group of motions point of view generates the noncommutative Laplacian when acting on $\mathbb{R}^3_\lambda$. The model here is a Euclidean inhomogeneous one.

The four-dimensional (4D) version $U(so_{1,3})\triangleright\!\!\blacktriangleleft\mathbb{C}[B_+]$ of this construction (Majid and Ruegg 1994) is again linked to Planck-scale predictions, this time as a generalized symmetry. In terms of translation generators $p^\mu$, rotations $M_i$ and boosts $N_i$ we have

$$[p^\mu, p^\nu] = 0, \qquad [M_i, M_j] = \mathrm{i}\epsilon_{ij}{}^k M_k$$
$$[N_i, N_j] = -\mathrm{i}\epsilon_{ij}{}^k M_k, \qquad [M_i, N_j] = \mathrm{i}\epsilon_{ij}{}^k N_k$$
$$[p^0, M_i] = 0, \qquad [p^i, M_j] = \mathrm{i}\epsilon^i{}_{jk} p^k, \qquad [p^0, N_i] = -\mathrm{i}p_i$$

as usual, and the modified relations and coproduct

$$[p^i, N_j] = -\frac{\mathrm{i}}{2}\delta^i_j\left(\frac{1 - \mathrm{e}^{-2\lambda p^0}}{\lambda} + \lambda\boldsymbol{p}^2\right) + \mathrm{i}\lambda p^i p_j$$
$$\Delta N_i = N_i \otimes 1 + \mathrm{e}^{-\lambda p^0} \otimes N_i + \lambda\epsilon_{ijk} p^j \otimes M_k$$
$$\Delta p^i = p^i \otimes 1 + \mathrm{e}^{-\lambda p^0} \otimes p^i$$

and the usual additive coproducts on $p^0, M_i$. This time the Lorentz group orbits in $B_+$ are deformed hyperboloids rather than deformed spheres, and the Casimir that controls this has the same form as above but with $-$ in the cosh term, that is, the model is a Lorentzian one. We know from the general theory of bicrossproducts that this Hopf algebra acts on $U(b_+) = \mathbb{R}^{1,3}_\lambda$ the spacetime in the section "Cogravity," and the Casimir induces the wave operator as we have seen there.

Let us look a bit more closely at the deformed hyperboloids. Because neither group here is compact, one expects from the general theory of bicrossproducts to have limiting accumulation regions. This is visible in the contour plot of $p^0$ against $|\boldsymbol{p}|$ in **Figure 4**, where the $p^0 > 0$ mass shells are now cups with almost vertical walls, compressed into the vertical tube

$$|\boldsymbol{p}| < \lambda^{-1}$$

In other words, the 3-momentum is bounded above by the Planck momentum scale (if $\lambda$ is the Planck time). Indeed, the light-cone equation (setting the Casimir to zero) reads $\lambda|\boldsymbol{p}| = 1 - \mathrm{e}^{-\lambda p_3}$ so this is

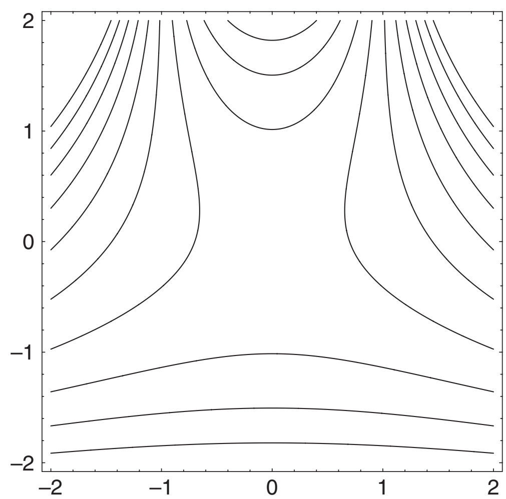



**Figure 4** Deformed mass-shell orbits in the bicrossproduct curved momentum space for $\lambda = 1$.

immediate. Nevertheless, this observation is so striking that the bicrossproduct model has been dubbed "doubly special" and spawned the search for other such models. Such accumulation regions are a main discovery of the noncompact bicrossproduct theory visible already in the Planck-scale Hopf algebra. The model further confirms the role of the matched pair equations as a toy version of Einstein's.

## Poisson–Lie *T*-Duality

We have explained in Section 3 that the matched pair equations are equivalent to a local factorization of Lie groups, with the action and back-reaction created "equally and oppositely" from this. For the two models in the last section, these are $SL_2(\mathbb{C})$ factorizing as $SU_2$ and a 3D $B_+$, and $SO_{2,3}$ locally as $SO_{1,3}$ and a 4D $B_+$. The first of these examples is in fact one of a general family based on the Iwasawa decomposition $G_{\mathbb{C}} = G \bowtie G^*$ where $G$ is a compact Lie group with complexification $G_{\mathbb{C}}$ and $G^*$ a certain solvable group. From this, one may construct a solution $(G, G^*)$ of the matched pair equations and bicrossproduct quantum group

$$\mathbb{C}[G^*] \blacktriangleright\!\!\triangleleft U(\mathfrak{g})$$

associated to all complex simple Lie algebras. This is again part of the bicrossproduct theory from the 1980s. On the other hand, the Lie algebra $\mathfrak{g}^*$ here can be identified with the dual of $\mathfrak{g}$ in which case its Lie algebra corresponds to a Lie coproduct $\delta : \mathfrak{g} \to \mathfrak{g} \otimes \mathfrak{g}$ and makes $(\mathfrak{g}, \delta)$ into a Lie bialgebra in the sense of Drinfeld. This $\delta$ exponentiates to a Poisson bracket on $G$ making it a "Poisson–Lie group" and the quantization of this is provided by the quantum group coordinate algebras $\mathbb{C}_q[G]$ (*see* Hopf Algebras and $q$-Deformation Quantum Groups and Classical $r$-matrices, Lie Bialgebras, and Poisson Lie Groups). The bicrossproduct quantum groups are nevertheless unrelated to the latter even though they spring form related classical data.

As already discussed, one interpretation here is of quantized particles in $G^*$ moving on orbits under $G$ and in vice versa in the dual model. The dual model is equivalent in the sense that the states of one (in the sense of positive-linear functionals) lie in the algebra of observables of the other and we also saw in the Planck-scale example inversion of structure constants reminiscent of $T$-duality in string theory. Motivated in part by this duality Klimcik (1996) along with Severa in the mid 1990s showed that indeed a $\sigma$-model on $G$ could be constructed in such a way that there was a matching dual $\sigma$-model on $G^*$ in some sense equivalent in terms of solutions to the equations of motion. The Lagrangians here have the usual form

$$\mathcal{L} = E_u(u^{-1}\partial_+ u, u^{-1}\partial_- u),$$
$$\hat{\mathcal{L}} = \hat{E}_s(s^{-1}\partial_+ s, s^{-1}\partial_- s)$$

where $u : \mathbb{R}^{1,1} \to G$ and $s : \mathbb{R}^{1,1} \to G^*$ are the dynamical fields, except that the inner products $E, \hat{E}$ are not constant. Rather they are obtained by solving nonlinear differential equations on the groups defined through the structure constants of $\mathfrak{g}, \mathfrak{g}^*$ and the Drinfeld double $D(\mathfrak{g})$. At the time, $T$-duality here was well understood in the case of abelian groups while these Poisson–Lie $T$-duality models provided the first convincing nonabelian models.

This construction was extended by Beggs and Majid (2001) to a general matched pair $(G, M)$, that is, a $\sigma$-model on $G$ dual to one on $M$. The Poisson–Lie case is the special case where the actions are coadjoint actions and the Lie algebra of $G \bowtie M$ is $D(\mathfrak{g})$. The solutions of the equations of motion for the two systems are created "equally and oppositely" from one on the factorizing group. It could be expected that $T$-duality ideas again play a role in Planck-scale physics.

## Other Bicrossproducts

There are also infinite-dimensional factorizations such as the Riemann–Hilbert problem (*see* Riemann–Hilbert Problem) in the theory of integrable systems and hence infinite-dimensional matched pairs and bicrossproducts linked to

them. Here we mention just one partly infinite example of current interest.

Thus, the diffeomorphisms on the line $\mathbb{R}$ may be factorized into transformations of the form $ax + b$ and diffeomorphisms that fix the origin and have unit differential there. After a (logarithmic) change of generators to arrive at an algebraic picture, one has a bicrossproduct

$$H(1) = U(b_+) \triangleright\!\!\blacktriangleleft H_\infty$$

where $b_+$ is now the two-dimensional (2D) Lie algebra with relations $[x, y] = x$ and $H_\infty$ is the algebra of polynomials in generators $\delta_n$ and a certain coalgebra as a model of the coordinate algebra of the group of diffeomorphisms that fix the origin with unit differential. The Hopf algebra $H(1)$ was introduced by Connes and Moscovici (1998) although not actually as a bicrossproduct (but motivated by the bicrossproduct theory) as part of a family $H(n)$ useful in cyclic cohomology computations. It has cross relations and coproduct determined by

$$[\delta_n, x] = \delta_{n+1}, \qquad [\delta_n, y] = n\delta_n,$$
$$\Delta\delta_1 = \delta_1 \otimes 1 + 1 \otimes \delta_1$$
$$\Delta x = x \otimes 1 + 1 \otimes x + \delta_1 \otimes y,$$
$$\Delta y = y \otimes 1 + 1 \otimes y$$

which we see has a semidirect product form where $\delta_n \triangleleft x = \delta_{n+1}, \delta_n \triangleleft y = n\delta_n$. The coalgebra is also a semidirect coproduct by means of a back-reaction of $H_\infty$ in $B_+$ (expressed as a coaction). From the bicrossproduct theory, we also have a dual model

$$\mathbb{C}[B_+] \blacktriangleright\!\!\triangleleft U(\mathrm{diff}_0)$$

where $\mathrm{diff}_0$ is the Lie algebra of the group of diffeomorphisms fixing the origin. As such it could be viewed as in the family of examples in the section "Bicrossproduct Poincaré quantum groups" but now with a 2D $B_+$. We also conclude from the bicrossproduct theory that this acts covariantly on $\mathbb{R}^2_\lambda = U(b_+)$ after introducing the scaling parameter $\lambda$.

Finally, the Hopf algebra $H(1)$ is also part of a family of bicrossproduct Hopf algebras built on rooted trees and related to bookkeeping of overlapping divergences in renormalizable quantum field theories (*see* Hopf Algebra Structure of Renormalizable Quantum Field Theory). While we have not had room to cover all bicrossproduct quantum groups of interest, it would appear that bicrossproducts are indeed intimately tied up with actual quantum physics.

*See also:* Classical *r*-Matrices, Lie Bialgebras, and Poisson Lie Groups; Hopf Algebra Structure of Renormalizable Quantum Field Theory; Hopf Algebras and *q*-Deformation Quantum Groups; Quantum Group Differentials, Bundles and Gauge Theory; Riemann–Hilbert Problem; von Neumann Algebras: Introduction, Modular Theory, and Classification Theory.

## Further Reading


Amelino-Camelia G and Majid S (2000) Waves on noncommutative spacetime and gamma-ray bursts. *International Journal of Modern Physics A* 15: 4301–4323.

Beggs E and Majid S (2001) Poisson–Lie *T*-duality for quasitriangular Lie bialgebras. *Communications in Mathematical Physics* 220: 455–488.

Connes A and Moscovici H (1998) Hopf algebras, cyclic cohomology and the transverse index theory. *Communications in Mathematical Physics* 198: 199–246.

Kac GI and Paljutkin VG (1966) Finite ring groups. *Transactions of the American Mathematical Society* 15: 251–294.

Kempf A, Mangano G, and Mann RB (1995) Hilbert space representation of the minimal length uncertainty relation. *Physical Review D* 52: 1108–1118.

Klimcik C (1996) Poisson–Lie *T*-duality. *Nuclear Physics B (Proc. Suppl.)* 46: 116–121.

Lukierski J, Nowicki A, Ruegg H, and Tolstoy VN (1991) *q*-Deformation of Poincaré algebra. *Physics Letters B* 268: 331–338.

Majid S (1988) Hopf algebras for physics at the Planck scale. *Journal of Classical and Quantum Gravity* 5: 1587–1606.

Majid S (1990) Physics for algebraists: non-commutative and non-cocommutative Hopf algebras by a bicrossproduct construction. *Journal of Algebra* 130: 17–64.

Majid S (1990) Matched pairs of Lie groups associated to solutions of the Yang–Baxter equations. *Pacific Journal of Mathematics* 141: 311–332.

Majid S (1990) On *q*-regularization. *International Journal of Modern Physics A* 5: 4689–4696.

Majid S (1995) *Foundations of Quantum Group Theory*. Cambridge: Cambridge University Press.

Majid S (2000) Meaning of noncommutative geometry and the Planck-scale quantum group. *Springer Lecture Notes in Physics* 541: 227–276.

Majid S and Ruegg H (1994) Bicrossproduct structure of the $\kappa$-Poincaré group and non-commutative geometry. *Physics Letters B* 334: 348–354.

Oeckl R (2000) Untwisting noncommutative $\mathbb{R}^d$ and the equivalence of quantum field theories. *Nuclear Physics B* 581: 559–574.

Seiberg N and Witten E (1999) String theory and noncommutative geometry. *Journal of High Energy Physics* 9909: 032.

Snyder HS (1947) Quantized space-time *Physical Review D* 67: 38–41.

Takeuchi M (1981) Matched pairs of groups and bismash products of Hopf algebras. *Communications in Algebra* 9: 841.

# Bifurcation Theory

**M Haragus**, Université de Franche-Comté, Besançon, France
**G Iooss**, Institut Non Linéaire de Nice, Valbonne, France



## Introduction

Consider the following equation:

$$F(X, \mu) = 0 \qquad [1]$$

where $X$ is the variable, $\mu$ is a parameter, and $X, \mu, F$ belong to appropriate (finite- or infinite-dimensional) spaces. The problem of *bifurcation theory* is to *describe the singularities of the set of solutions*

$$S_\mu = \{X; (X, \mu) \text{ satisfies } F(X, \mu) = 0\}$$

The word "bifurcation" was introduced by H Poincaré (1885) in his study of equilibria of rotating liquid masses.

The simplest example is the study of the real roots $x$ of a quadratic polynomial

$$x^2 + bx + c = 0 \qquad [2]$$

where $\mu$ is represented by the pair of parameters $(b, c) \in \mathbb{R}^2$. As it is well known, real roots are determined by the sign of

$$\Delta \overset{\text{def}}{=} b^2 - 4c$$

For $\Delta < 0$, there is no real solution of [2], while there are two solutions $x_\pm$ in the region $\Delta > 0$, which merge when the distance between the point $(b, c)$ and the parabola $\Delta = 0$ tends towards 0. It is then clear that a *singularity occurs in the structure of the set of solutions of* [2] *at the crossing of the parabola* $\Delta = 0$ or, in other words, a *bifurcation occurs in the parameter space* $(b, c)$ *on the parabola* $\Delta = 0$. A point $(\mu_0, x_0) \in \mathbb{R}^3$ is then called a *bifurcation point* if $\mu_0 = (b, c)$ satisfies $\Delta = 0$, and $x_0 = -b/2$.

In the theory of differential equations, $F(X, \mu)$ often represents a vector field. This study is then concerned with the existence of equilibrium solutions to the differential equation

$$\frac{\mathrm{d}X}{\mathrm{d}t} = F(X, \mu) \qquad [3]$$

and is therefore referred to as *static bifurcation theory*. In addition, *dynamic bifurcation theory* is concerned here with "changes" in the dynamic properties of the solutions of the differential equation as $\mu$ varies. A widely used way to characterize these "changes" is to say that the vector field $F(\cdot, \mu_0)$ is *structurally stable* if the sets of orbits of the differential equation are homeomorphic for $\mu$ close to $\mu_0$, with homeomorphisms which preserve the orientation of the orbits in time $t$. Then a bifurcation occurs at $\mu = \mu_0$ if $F(\cdot, \mu_0)$ is not structurally stable. It turns out that there is a close link between the stability properties of equilibrium solutions of the differential equation and the type of the bifurcation in static theory.

The tools developed in bifurcation theory are extensively used to solve concrete problems arising in physics and natural sciences. These problems may be modeled by ordinary or partial differential equations, integral equations, but also delay equations or iteration maps, and in all these cases the presence of parameters naturally leads to bifurcation phenomena. They can be regarded as problems of the form [1] or [3], in suitable function spaces, and bifurcation theory allows to detect solutions and to describe their qualitative properties. During the last decades, a class of problems in which the use of bifurcation theory led to significant progress is concerned with nonlinear waves in partial differential equations, including hydrodynamic problems, nonlinear water waves, elasticity, but also pattern formation, front propagation, or spiral waves in reaction–diffusion type systems.

## Examples in One and Two Dimensions

The most complete results in bifurcation theory are available in one and two dimensions. The study of static bifurcations in one dimension is concerned with scalar equations

$$f(x, \mu) = 0 \qquad [4]$$

where $x \in \mathbb{R}, \mu \in \mathbb{R}$, and the function $f$ is supposed to be regular enough with respect to $(x, \mu)$. When $f(x_0, \mu_0) = 0$ and the derivative of $f$ with respect to $x$ satisfies $\partial_x f(x_0, \mu_0) \neq 0$, the implicit function theorem gives a unique branch of solutions $x(\mu)$ for $\mu$ close to $\mu_0$, and shows the absence of bifurcation points near $(\mu_0, x_0)$. Bifurcation theory intervenes when

$$\partial_x f(x_0, \mu_0) = 0 \qquad [5]$$

and one cannot apply the implicit function theorem for solving with respect to $x$ near $x_0$. A complete description of the set of solutions near $(x_0, \mu_0)$ can be obtained by looking at the partial derivatives of $f$ with respect to $x$ and $\mu$.

For example, if

$$\partial_\mu f(x_0, \mu_0) \neq 0,$$

it is possible to solve with respect to $\mu$ and obtain a regular solution $\mu(x)$ such that $\mu(x_0) = \mu_0$ and $f(x, \mu(x)) \equiv 0$. In addition, if the second order derivative

$$\partial_x^2 f(x_0, \mu_0) \neq 0$$

the picture of the solution set in the plane $(\mu, x)$, also called *bifurcation diagram*, shows a turning point with a fold opened to the left or to the right depending upon the sign of the product $\partial_\mu f(x_0, \mu_0) \cdot \partial_x^2 f(x_0, \mu_0)$; see **Figure 1**. Notice that here the bifurcation point $(\mu_0, x_0) \in \mathbb{R}^2$ corresponds to the appearance of a pair of solutions of [4] "from nowhere". This is the simplest example of a *one-sided bifurcation* in which the bifurcating solutions exist for either $\mu > \mu_0$ or $\mu < \mu_0$.

A particularly interesting situation arises when the equation possesses a symmetry. For example, assume that in [4] the function $f$ is odd with respect to $x$. This implies that we always have the solution $x = 0$, for any value of the parameter $\mu$. Assume now that $f$ satisfies

$$\partial_x f(0, \mu_0) = 0 \qquad [6]$$

and that

$$\partial_{x\mu}^2 f(0, \mu_0) \neq 0, \qquad \partial_x^3 f(0, \mu_0) \neq 0 \qquad [7]$$

Then the point $(\mu_0, 0)$ is a *pitchfork bifurcation point*, this denomination being related with the bifurcation diagram in the plane $(\mu, x)$; see **Figure 2**. Notice that here, the bifurcation point $(\mu_0, x_0) \in \mathbb{R}^2$ corresponds to the bifurcation from the origin of a pair of solutions exchanged by the symmetry $x \to -x$, in addition to the persistent "trivial" solution $x = 0$ which is invariant under the above symmetry. Such a bifurcation is also referred to as a *symmetry-breaking bifurcation*. Similar bifurcation diagrams are found when the equation [4] has a "known" branch of

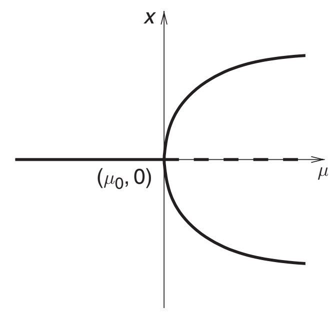



**Figure 2** Supercritical pitchfork bifurcation in the case $\partial_{x\mu}^2 f(0, \mu_0) > 0$ and $\partial_x^3 f(0, \mu_0) < 0$.. The solid (dashed) lines indicate the branch of stable (unstable) solutions in the differential equation.

solutions $x(\mu)$ for $\mu$ close to $\mu_0$. This situation arises often in applications where usually this branch consists of trivial solutions $x(\mu) = 0$. Then at a bifurcation point $(\mu_0, x_0)$ a second branch of solutions appears forming either a one-sided bifurcation, or a two-sided bifurcation; see **Figure 3.**

We can now view $f$ as a vector field in the ordinary differential equation

$$\frac{dx}{dt} = f(x, \mu) \qquad [8]$$

and the study above corresponds to looking for equilibrium solutions of [8]. The stability of such a solution is determined by the sign of the derivative $\partial_x f(x, \mu)$ of $f$ at this equilibrium, and it is closely related to the type of the static bifurcation.

In the case of a *turning point bifurcation*, when $\partial_x^2 f(x_0, \mu_0) \neq 0$, the sign of $\partial_x f(x, \mu)$ is different for the two bifurcating solutions. This means that one solution is attracting (i.e., stable), the other one being repelling (i.e., unstable); see **Figure 1**. In the case of a *pitchfork bifurcation* as above, the stability of the trivial solution $x = 0$ changes when $\mu$ crosses $\mu_0$, and the stability of both bifurcating nonzero solutions is the opposite from the stability of the origin on the side of the bifurcation. The bifurcation

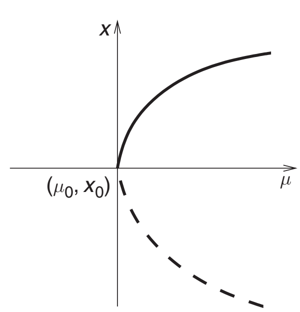



**Figure 1** Turning point bifurcation in the case $\partial_\mu f(x_0, \mu_0) > 0$ and $\partial_x^2 f(x_0, \mu_0) < 0$. The solid (dashed) line indicates the branch of stable (unstable) solutions in the differential equation.

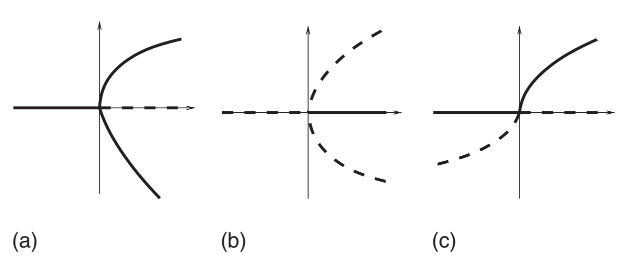



**Figure 3** Typical bifurcation diagrams in the case of a branch of trivial solutions. One-sided bifurcations: (a) supercritical, (b) subcritical; two-sided bifurcation: (c) transcritical. The solid (dashed) lines indicate the branch of stable (unstable) solutions in the differential equation.

is called *supercritical* if the bifurcating solutions lie on the side of the bifurcation point where the basic solution $x=0$ is unstable and *subcritical* otherwise; see Figure 2. The situation is the same in the case of one-sided bifurcations for an equation which has a "known" branch of solutions. In the case of a two-sided bifurcation, there is an *exchange of stability* at the bifurcation point $(\mu_0, x_0)$, solutions on the two branches having opposite stability for $\mu > \mu_0$ and $\mu < \mu_0$, which changes at $(\mu_0, x_0)$. Such a bifurcation is also referred to as *transcritical*; see Figure 3.

Notice that the study of fixed points or periodic points for maps enter in the above frame. Specifically, the period-doubling process occurring in successive bifurcations of one-dimensional maps is a common phenomenon in physics.

The analysis of bifurcations in two dimensions leads to more complicated scenarios. Consider the differential equation [8] in which now $x \in \mathbb{R}^2$ and $f(x,\mu) \in \mathbb{R}^2$, and assume that $f(x_0,\mu_0)=0$. The behavior of solutions near $(x_0,\mu_0)$ is determined by the differential $D_x f(x_0,\mu_0) =: L$ of $f$ with respect to $x$, which can be identified with a $2 \times 2$ matrix. For steady solutions, the implicit function theorem insures the existence of a unique branch of solutions $x(\mu)$ provided $L$ is invertible or, in other words, zero does not belong to the spectrum of $L$. Consequently, the study of bifurcations of steady solutions is concerned with the case when zero belongs to the spectrum of $L$, and can be performed following the strategy described for one dimension, provided that the zero eigenvalue of $L$ is simple. For example, assuming that the second eigenvalue is negative leads in general to a *saddle–node bifurcation*, where an additional dimension is added to the previous picture of a turning point bifurcation, in which one of the two bifurcating steady solutions is a stable node, while the other one is a saddle. If, in addition, there is a symmetry $S$ commuting with $f$, that is, such that $f(Sx,\mu)=Sf(x,\mu)$, and if, for example, $x_0$ is invariant under $S, Sx_0=x_0$, and the eigenvector $\zeta_0$ associated to the zero eigenvalue of $L$ is antisymmetric, $L\zeta_0 = -\zeta_0$, then there is again a *pitchfork bifurcation*. The equation possesses a branch of symmetric steady solutions the stability of which changes when crossing the value $\mu_0$ of the parameter, node on one side and saddle on the other, and a pair of solutions is created in a one-sided bifurcation which are exchanged by the symmetry $S$ and have stability opposite to the one of the symmetric solution, just as in the one-dimensional pitchfork bifurcation above.

A new type of bifurcation that arises for vector fields in two dimensions is the so-called *Hopf bifurcation*. This bifurcation was first understood by Poincaré, and then proved in two dimensions by Andronov (1937) using a Poincaré map, and later in $n$ dimensions by Hopf (1948) by means of a Liapunov–Schmidt-type method. For the differential equation, the absence of the zero eigenvalue in the spectrum of $L$ is not enough to ensure that the vector field $f(\cdot,\mu_0)$ is structurally stable in a neighborhood of $x_0$. This only holds when the spectrum of $L$ does not contain purely imaginary eigenvalues, as asserted by the Hartman–Grobman theorem. We are then left with the case when $L$ has a pair of purely imaginary eigenvalues $\pm i\omega, \omega \in \mathbb{R}^*$. Static bifurcation theory gives that the system has a unique branch of equilibria $(x(\mu),\mu)$ for $\mu$ close to $\mu_0$, and typically their stability changes as $\mu$ crosses $\mu_0$. For the differential equation a *Hopf bifurcation* occurs in which a branch of periodic orbits bifurcates on one side of $\mu_0$, and their stability is opposite to that of the steady solution on this side; see Figure 4. A convenient way to study this bifurcation is through "normal form theory," which is briefly described below.

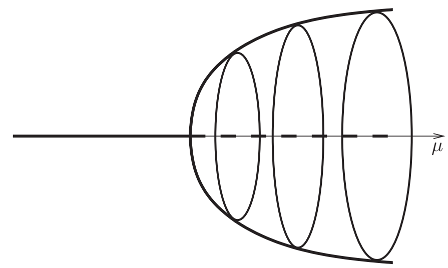



**Figure 4** Supercritical Hopf bifurcation.

## Local Bifurcation Theory

There are two aspects of bifurcation theory, *local* and *global* theory. As this designation suggests, local theory is concerned with (local) properties of the set of solutions in a neighborhood of a "known" solution, while global theory investigates solutions in the entire space.

An important class of tools in *local* bifurcation theory consists of *reduction methods*, among which the *Liapunov–Schmidt reduction* and the *center manifold reduction* are often used to investigate static and dynamic bifurcations, respectively. The basic idea is to replace the bifurcation problem by an equivalent problem in lower dimensions, for example, a one- or a two-dimensional problem as the ones above.

Consider again the equation [1] in which $F: \mathcal{X} \times \mathcal{M} \to \mathcal{Y}$ is sufficiently regular, and $\mathcal{X}, \mathcal{Y}$, and $\mathcal{M}$ are Banach spaces. Assume, without loss of generality,

that $F(0,0)=0$, or, in other words, that one solution is known. The equation can be then written as

$$LX + G(X,\mu) = 0$$

in which $L = D_X F(0,0)$ represents the differential of $F$ with respect to $X$ at (0, 0), and is assumed to have a closed range. The implicit function theorem shows absence of bifurcation if $L$ has a bounded inverse, so that bifurcations are related to the existence of a nontrivial kernel of $L$. The *Liapunov–Schmidt reduction* then goes as follows.

Let $N(L)$ and $R(L)$ denote the kernel and the range of $L$, respectively, and consider continuous projections $P:\mathcal{X}\to N(L)$ and $Q:\mathcal{Y}\to R(L)$. Then there exists a bounded linear operator $B:R(L)\to(\mathrm{id}-P)\mathcal{X}$, the right inverse of $L$, satisfying $LB=\mathrm{id}$ on $R(L)$ and $BL=\mathrm{id}-P$ on $\mathcal{X}$. For $X\in\mathcal{X}$ one may write

$$X = X_0 + X_1, \quad X_0 = PX, X_1 = (\mathrm{id}-P)X$$

and then by projecting with $\mathrm{id}-Q$ and $Q$ the equation becomes

$$(\mathrm{id}-Q)G(X_0+X_1,\mu) = 0$$
$$X_1 + BQG(X_0+X_1,\mu) = 0$$

The implicit function theorem allows to solve the second equation for $X_1=\psi(X_0,\mu)$ in a neighborhood of the origin. Substitution into the first equation leads to the equation in $(\mathrm{id}-Q)\mathcal{Y}$ for $X_0$ in $P\mathcal{X}$,

$$(\mathrm{id}-Q)G(X_0+\psi(X_0,\mu),\mu) = 0$$

also called *bifurcation equation*. This equation completely describes the set of solutions to [1] in a neighborhood of (0, 0), and this problem is then posed in a space of dimension much smaller than the dimension of $\mathcal{X}$.

The basic principle of the Liapunov–Schmidt method has been discovered and used independently by different authors. E Schmidt (1908) used this method for integral equations, while Liapunov used it to study the stability of the zero solution of nonlinear partial differential equations when the linear part has zero eigenvalues (1947), and later in 1960 for the bifurcation problem studied by Poincaré (1885). In working in a Banach space of $t$-periodic functions, the Liapunov–Schmidt method may be used to solve the Hopf bifurcation problem, as did Hopf himself in 1948.

The analog of this reduction procedure for the differential equation [3] is the *center manifold reduction*. Assuming that $F(0,0)=0$, we obtain the differential equation

$$\frac{\mathrm{d}X}{\mathrm{d}t} = LX + G(X,\mu)$$

Since dynamic bifurcations are related to the existence of purely imaginary spectral values of $L$, the kernel of $L$ alone is not enough to describe this situation. One has to consider the spectral space $\mathcal{Y}_c$ of $L$ associated to the purely imaginary spectrum of $L$. A spectral gap is needed between this part of the spectrum and the rest (always true in finite dimensions), so that the spectral projection $P$ onto $\mathcal{Y}_c$ is well defined. One writes

$$X = X_c + X_h, \quad X_c = PX, \ X_h = (\mathrm{id}-P)X$$

and obtains the decomposed system

$$\frac{\mathrm{d}X_c}{\mathrm{d}t} = LX_c + PG(X_c+X_h,\mu)$$
$$\frac{\mathrm{d}X_h}{\mathrm{d}t} = LX_h + (\mathrm{id}-P)G(X_c+X_h,\mu)$$

The reduction procedure works provided the nonhomogeneous linear equation

$$\frac{\mathrm{d}X_h}{\mathrm{d}t} = LX_h + f(t)$$

possesses a unique solution in suitably chosen function spaces with weak exponential growth, such that one can then solve the second equation for $X_h=\Psi(X_c)$ in a neighborhood of the origin in these function spaces. This property is always true in finite dimensions, but it has to be checked in infinite dimensions. Different results showing the solvability of this equation are available in both Banach and Hilbert spaces, relying upon additional conditions on the spectrum of $L$, decaying properties of the resolvent of $L$ on the imaginary axis, and regularity properties of the nonlinearity $G$. The map $\Psi$ is then used to construct a map $\psi:P\mathcal{X}\times\mathcal{M}\to(\mathrm{id}-P)\mathcal{X}$, defined in a neighborhood of the origin, which parametrizes a *local center manifold* invariant under the flow of the equation. The flow on this center manifold is governed by the *reduced equation* in $\mathcal{Y}_c$,

$$\frac{\mathrm{d}X_c}{\mathrm{d}t} = LX_c + PG(X_c+\psi(X_c,\mu),\mu)$$

which completely describes the bifurcation problem.

The first proofs of this result were given in finite dimensions by Pliss (1964) and Kelley (1967). Center manifolds in infinite dimensions have been studied in different settings determined by assumptions on the linear part $L$ and the nonlinear part $G$. One typical assumption in infinite dimensions is that the spectrum of $L$ contains only a finite number of purely imaginary eigenvalues, so that the reduced equation above is a differential equation in a finite-dimensional space.

These reduction methods work for a large class of problems and the advantage of such an approach is that one is left with a bifurcation problem in a lower-dimensional space. The methods involved in

solving this reduced bifurcation problem can be very different from one problem to another, and often make use of some additional structure in the problem, such as a gradient-like structure, Hamiltonian structure, or the presence of symmetries, which are preserved by the reduction procedure.

A powerful tool for the analysis of these reduced differential equations is provided by the *normal form theory*, which goes back to works of Poincaré (1885) and Birkhoff (1927). The idea is to use coordinate transformations to make the expression of the vector field as simple as possible. The transformed vector field is called *normal form*. There is an extensive literature on normal forms for vector fields in many different contexts, in both finite- and infinite-dimensional cases. Typically the classes of normal forms are characterized in terms of the linear part of the differential equation.

For differential equations of the form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = Lx + g(x, \mu) \qquad [9]$$

in which $L$ is a matrix and $g$ a sufficiently regular map such that $g(0,0)=0, D_x g(0,0)=0$, as encountered in bifurcation theory, one possible characterization of normal forms makes use of the adjoint matrix $L^*$. Fixing any order $k \geq 2$, there exist polynomials $\Phi$ and $N$ of degree $k$ in $x$ with coefficients which are regular functions of $\mu$, and $\Phi(0,0)=N(0,0)=0, D_x\Phi(0,0)=D_xN(0,0)=0$, such that by the change of variables

$$x = y + \Phi(y, \mu)$$

the equation [9] is transformed into the normal form

$$\frac{\mathrm{d}y}{\mathrm{d}t} = Ly + N(y, \mu) + o(\|y\|^k) \qquad [10]$$

in which the polynomial $N$ is characterized through

$$N(\mathrm{e}^{tL^*}y, \mu) = \mathrm{e}^{tL^*}N(y, \mu)$$

for all $y, \mu$, and $t$, or, equivalently,

$$D_y N(y, \mu)L^* y = L^* N(y, \mu)$$

for all $y$ and $\mu$. This characterization allows to determine the classes of possible normal forms for a given matrix $L$, and also provides an efficient way to compute the normal form for a given vector field $g$. As for the reduction methods, normal form transformations can be made to preserve the additional structure of the problem, such as Hamiltonian structure or symmetries.

As an example, consider a differential equation of the form [9] with $x \in \mathbb{R}^n$ and $\mu \in \mathbb{R}$, which supports a Hopf bifurcation so that $L$ has simple eigenvalues $\pm\mathrm{i}\omega, \omega > 0$, and no other eigenvalues with zero real part. The center manifold reduction provides a two-dimensional reduced system with linear part having the simple eigenvalues $\pm\mathrm{i}\omega$, for which it is convenient to write the normal form in complex variables

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \mathrm{i}\omega A + AQ\left(|A|^2, \mu\right) + o\left(|A|^{2k+2}\right)$$

for $A(t) \in \mathbb{C}$, where $Q$ is a complex polynomial of degree $k$ in $|A|^2$ with $Q(0,0)=0$, or, equivalently, in polar coordinates $A = r\mathrm{e}^{\mathrm{i}\phi}$,

$$\frac{\mathrm{d}r}{\mathrm{d}t} = rQ_r\left(r^2, \mu\right) + o\left(r^{2k+2}\right)$$

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \omega + Q_\phi\left(r^2, \mu\right) + o\left(r^{2k+1}\right)$$

$Q_r$ and $Q_\phi$ being the real and imaginary part of $Q$, respectively. The radial equation for $r$ truncated at order $2k+1$ decouples and admits a pitchfork bifurcation. The bifurcating steady solutions of this equation then lead first to periodic solutions for the truncated system, which are then shown to persist for the full equation by a standard perturbation analysis.

A situation that occurs in a large class of problems is when the problem possesses a reversibility symmetry, which often comes from some reflection invariance in the physical space, that is, when the vector field $F(\,\cdot\,, \mu)$ anticommutes with a symmetry operator $S$. One of the simplest examples is the case of a differential equation [9] when the matrix $L$ has a double eigenvalue in 0, no other eigenvalues with zero real part, and a one-dimensional kernel which is invariant by $S$. In this case, the center manifold reduction provides a two-dimensional reduced reversible system, which can be put in the normal form

$$\frac{\mathrm{d}a}{dt} = b$$

$$\frac{\mathrm{d}b}{dt} = \mu - a^2 + o((|a| + |b|)^3)$$

which anticommutes with the symmetry $(a, b) \mapsto (a, -b)$. The above system undergoes a *reversible Takens–Bogdanov bifurcation* and has for $\mu > 0$ a phase portrait as in **Figure 5**. There are two equilibria, one a saddle, the other a center, and a family of periodic orbits with the zero-amplitude limit at the center equilibrium, and the infinite-period limit a homoclinic orbit, originating at the saddle point. In concrete problems the bounded orbits of such a reduced system determine the shape of physically interesting solutions of the full system of equations, such as, for example, in water-wave theory where to homoclinic and periodic orbits correspond solitary and periodic waves, respectively.

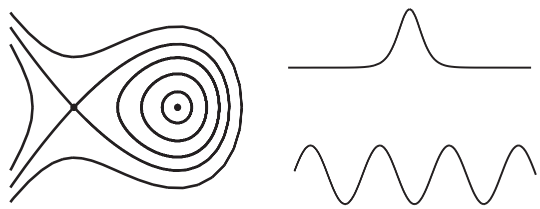


**Figure 5** Phase portrait of the reduced system in a reversible Takens–Bogdanov bifurcation (left) and sketch of the *a*-component of solutions corresponding to homoclinic and periodic orbits (right).

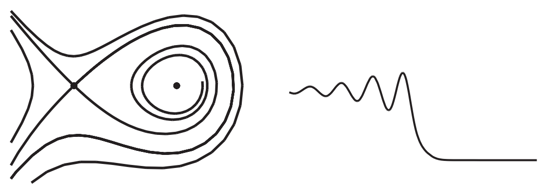


**Figure 6** Phase portrait of the reduced system in absence of reversibility (left) and sketch of the *a*-component of the solution corresponding to the bounded orbit (right).

Notice that in the absence of the reversibility symmetry, the same type of bifurcation may lead to a completely different phase portrait for the reduced system as, for example, the one in Figure 6 in which the homoclinic and the periodic orbits disappear. This situation often occurs in the presence of a small dissipation in nearly reversible systems.

## Global Bifurcation Theory

Most of the existing results in global bifurcation theory concern the static problem [1]. The analysis of *global sets* of solutions often relies upon topological methods, degree theory, but also variational methods, or analytic function theory. Significant progress in understanding global branches of solutions has been made in the 1970s, in particular, for nonlinear eigenvalue problems and the Hopf bifurcation problem (see, e.g., works by Rabinowitz, Crandall, Dancer, and Alexander, Yorke, Ize, respectively).

A now-classical result in the topological theory of global bifurcations is the following theorem by Rabinowitz (1970), which gives a characterization of global sets of solutions for *eigenvalue problems* of the form

$$X = F(X, \mu) = \mu L X + H(X, \mu)$$

$H(X, \mu) = o(\|X\|)$, posed for $(X, \mu) \in \mathcal{X} \times \mathbb{R}, \mathcal{X}$ being a Banach space. In contrast to local theory where the function $F$ is usually $k$-times differentiable (with a suitable $k$), in the global theory a typical assumption is that $F: \mathcal{X} \times \mathbb{R} \to \mathcal{X}$ is *compact*. The equation above possesses a "trivial" branch of solutions $(0, \mu)$ for any $\mu$. The bifurcation result asserts that if for some real parameter value $\mu_0$ zero is an eigenvalue of odd multiplicity of the operator $\mathrm{id} - \mu_0 L$, then the set $S$ of nontrivial solutions $(X, \mu)$ possesses a maximal subcontinuum which contains $(0, \mu_0)$ and meets either infinity in $\mathcal{X} \times \mathbb{R}$ or another trivial solution $(0, \mu_1), \mu_1 \neq \mu_0$. In particular, $(\mu_0, 0)$ is a bifurcation point. A local version of this result is often referred to as Krasnoselski's theorem.

Different versions and extensions of these theorems can be found in the literature, as, for example, in the case of a simple eigenvalue, or if the field $F$ is real-analytic when the set of solutions is path-connected. More recent works address the question of lack of compactness, and a number of results are now available for problems with additional structure (gradient-like or Hamiltonian structure), but also for concrete problems, such as the water-wave problem.

*See also:* Bifurcations in Fluid Dynamics; Bifurcations of Periodic Orbits; Central Manifolds, Normal Forms; Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Ginzburg–Landau Equation; Integrable Systems: Overview; Leray–Schauder Theory and Mapping Degree; Singularity and Bifurcation Theory; Stability Theory and KAM; Symmetry and Symmetry Breaking in Dynamical Systems.

## Further Reading

Arnold VI (1988) Geometrical Methods in the Theory of Ordinary Differential Equations. *Grundlehren der Mathematischen Wissenschaften*, vol. 250. New York: Springer.

Buffoni B and Toland J (2003) *Analytic Theory of Global Bifurcation*. Princeton: Princeton University Press.

Chossat P and Lauterbach R (2000) Methods in Equivariant Bifurcations and Dynamical Systems. *Advanced Series in Nonlinear Dynamics*, vol. 15. River Edge, NJ: World Scientific.

Chow S-N and Hale JK (1982) Methods of Bifurcation Theory. *Grundlehren der Mathematischen Wissenschaften*, vol. 251. New York: Springer.

Golubitsky M and Schaeffer DG (1985) Singularities and Groups in Bifurcation Theory, Vol. I. *Applied Mathematical Sciences*, vol. 51. New York: Springer.

Golubitsky M, Stewart I, and Schaeffer DG (1988) Singularities and Groups in Bifurcation Theory, Vol. II. *Applied Mathematical Sciences*, vol. 69. New York: Springer.

Guckenheimer J and Holmes P (1990) Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields. *Applied Mathematical Sciences*, vol. 42. New York: Springer.

Iooss G and Adelmeyer M (1998) Topics in Bifurcation Theory and Applications, *Advances Series in Nonlinear Dynamics*, 2nd edn., vol. 3, Singapore: World Scientific.

Iooss G, Helleman RHG, and Stora R (eds.) (1983) *Chaotic behavior of deterministic systems*. Session XXXVI of the Summer School in Theoretical Physics held at Les Houches June 29–July 31, 1981. Amsterdam: North-Holland.

Ize J and Vignoli A (2003) Equivariant Degree Theory. *de Gruyter Series in Nonlinear Analysis and Applications*, vol. 8. Berlin: de Gruyter and Co.

Kielhöfer H (2004) *Bifurcation Theory. An Introduction with Applications to PDEs*, Applied Mathematical Sciences, vol. 156. New York: Springer.

Kuznetsov YA (2004) Elements of Applied Bifurcation Theory, 3rd edn. *Applied Mathematical Sciences*, vol. 112. New York: Springer.

Ruelle D (1989) *Elements of Differentiable Dynamics and Bifurcation Theory*. Boston MA: Academic Press.

Vanderbauwhede A (1989) Centre Manifolds, Normal Forms and Elementary Bifurcations. *Dynamics Reported, Dynam. Report. Ser. Dynam. Systems Appl.*, vol. 2, pp. 89–169. Chichester: Wiley.

Vanderbauwhede A and Iooss G (1992) Center Manifold Theory in Infinite Dimensions. *Dynamics Reported: Expositions in Dynamical Systems*, vol. 1, pp. 125–163. Berlin: Springer.

# Bifurcations in Fluid Dynamics

**G Schneider**, Universität Karlsruhe, Karlsruhe, Germany



## Introduction

Almost all classical hydrodynamical stability problems are experiments or gedankenexperiment which have been designed to understand and to extract special phenomena in more complicated situations. Examples are the Taylor–Couette problem, Bénard's problem, Poiseuille flow, or Kolmogorov flow.

The Taylor–Couette problem consists in finding the flow of a viscous incompressible fluid contained in between two coaxial co- or counterrotating cylinders, cf. Figure 1. If the rotational velocity of the inner cylinder is below a certain threshold, the trivial solution, called the Couette flow, is asymptotically stable. At the threshold, this spatially homogenous solution becomes unstable and bifurcates via a pitchfork bifurcation or a Hopf bifurcation into different spatially periodic patterns, that is, depending on the rotational velocity of the outer cylinder the basic patterns are stationary (called the Taylor vortices) or time-periodic. If the rotational velocity of the inner cylinder is increased further, more complicated patterns occur. The bifurcation scenario is well understood from experiments and analytic investigations.

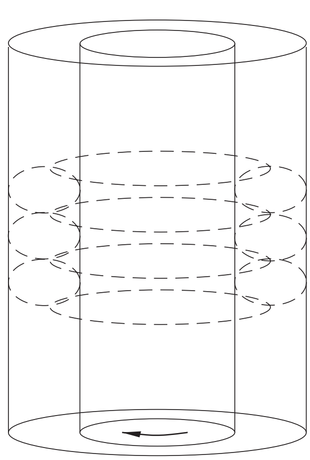


Figure 1 The Taylor–Couette problem with the Taylor vortices.

Bénard's problem consists in finding the flow of a viscous incompressible fluid contained in between two plates, where the lower plate is heated and the upper plate is kept at a constant temperature, cf. Figure 2. If the temperature difference between the two plates is below a certain threshold, the transport of energy from below to above is made by pure conduction. At this threshold, this spatially homogenous solution becomes unstable, convection sets in, and spatially periodic patterns as rolls or hexagons occur. Convection problems play a big role in geophysical applications, that is, in spherical domains, as the earth. The paradigm for an anisotropic pattern-forming system is electroconvection in nematic crystals.

Poiseuille flow consists in finding the flow of a viscous incompressible fluid flowing through a pipe driven by some pressure gradient, cf. Figure 3. In noncircular pipes, the trivial laminar flow becomes unstable at a critical pressure gradient. Experimentally, a direct transition to turbulent flow with large amplitudes is observed, according to the fact that in general at the instability point of the trivial solution a subcritical bifurcation occurs.

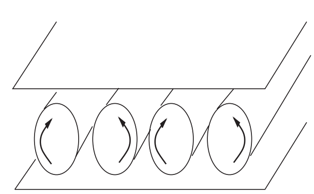


Figure 2 Bénard's problem with rolls.

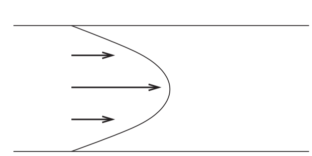


Figure 3 Poiseuille flow with the trivial solution.

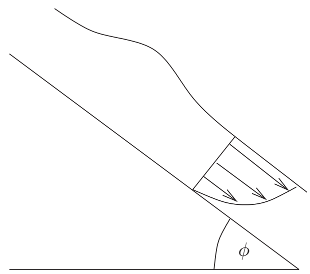



**Figure 4** The inclined-plane problem. The trivial Nusselt solution possesses a flat top surface and a parabolic flow profile.

Kolmogorov flow consists in finding the flow of a viscous incompressible fluid under the action of an external force parallel to the flow direction $x$ and varying periodically in the perpendicular $y$-direction. This gedankenexperiment has been designed by Kolmogorov in 1958 as a simplified model for the Poiseuille flow problem in order to study the nature of turbulence. The trivial solution which is called Kolmogorov flow can become unstable via a long-wave instability along the flow direction.

The inclined-plane problem consists in finding the flow of a viscous liquid running down an inclined plane, cf. Figure 4. The trivial solution, the so-called Nusselt solution, becomes sideband-unstable if the inclination angle $\phi$ is increased. Then the dynamics is dominated by traveling pulse trains, although the individual pulses are unstable due to the long-wave instability of the flat surface. Time series taken from the motion of the individual pulses indicates the occurrence of chaos directly at the onset of instability.

There are other famous hydrodynamical stability problems, with arbitrarily complicated bifurcation scenarios.

## Spectral Analysis of the Trivial Solution

All classical hydrodynamical stability problems are described by the Navier–Stokes equations

$$\begin{aligned} \partial_t U &= \frac{1}{\nu}\Delta U - \nabla p - (U\cdot\nabla)U + f \\ 0 &= \nabla\cdot U \end{aligned} \qquad [1]$$

where $U = U(x,t) \in \mathbb{R}^d$ with $d = 2,3$ is the velocity field, $p = p(x,t) \in \mathbb{R}$ the pressure field, $f$ some external forcing, and $\nu$ the dynamic viscosity. These equations are completed with boundary conditions. In case of Bénard's problem, the Navier–Stokes equations are coupled to a nonlinear heat equation.

By projecting $U$ onto the space of divergence-free vector fields and by taking the trivial solution as new origin all problems from the previous section can be written as evolutionary system

$$\partial_t U = \Lambda U + N(U)$$

where $U = 0$ corresponds to the trivial solution, where $\Lambda$ is a linear and $N(U) = \mathcal{O}(U^2)$ for $U \to 0$ a nonlinear operator. Most of the examples from the previous section are semilinear, that is, from a functional analytic point of view, the nonlinear operator $N$ can be controlled in terms of the linear operator $\Lambda$.

Since the form of the bifurcating pattern is only slightly influenced by far away boundaries, that is, for instance, the upper and lower end of the rotating cylinders in the Taylor–Couette problem, the problems are considered from a theoretical point of view in unbounded domains, $\Omega = \mathbb{R}^d \times \Sigma$, with $\Sigma \subseteq \mathbb{R}^m$ the bounded cross section that is, for instance, that the Taylor–Couette problem is considered with two cylinders of infinite length. Then the eigenfunctions of the linear operator $\Lambda$ are given by Fourier modes, that is,

$$\Lambda(e^{ik\cdot x}\varphi_{k,n}(z)) = \lambda_n(k)e^{ik\cdot x}\varphi_{k,n}(z)$$

with $x \in \mathbb{R}^d, k \in \mathbb{R}^d, k\cdot x = \sum_{j=1}^d k_j x_j, z \in \Sigma, n \in \mathbb{N}$. If an external control parameter is changed, independent of the underlying physical problem, the trivial solution becomes unstable, then the surface $k \mapsto \mathrm{Re}\lambda_1(k)$ intersects the plane $\{\mathrm{Re}\lambda_1(k) = 0\}$. Generically, this happens first at a nonzero wave vector $k_c \neq 0$ (cf. Figure 5).

Examples for such an instability are the Taylor–Couette problem, Bénard's problem, or Poiseuille flow. Very often, due to some conserved quantity in the problem we have $\mathrm{Re}\lambda_1(0) = 0$ for all values of the bifurcation parameter. Then, a so-called sideband instability can occur, cf. Figure 6.

Examples for such an instability are the Kolmogorov flow problem or the inclined plane problem.

According to some symmetries in the problem, for instance, reflection along the cylinders in the Taylor–Couette problem or rotational symmetry in Bénard's problem, the curves in Figure 5 are double or rotational symmetric.

In case of $\Omega$ being spherical symmetric, we have

$$\Lambda(f_l(r)\varphi_{l,n}(z)) = \lambda_l f_l(r)\varphi_{l,n}(z)$$

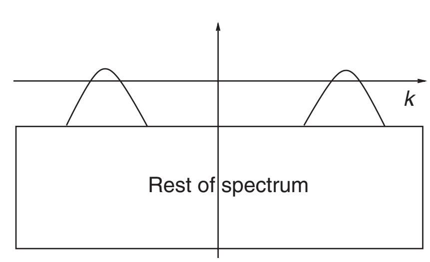



**Figure 5** Real part of the spectrum in case of an instability at a wave number $k_c \neq 0$. Definition of the small bifurcation parameter $\varepsilon$.

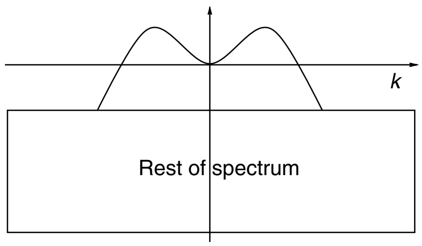



**Figure 6** Real part of the spectrum in case of a sideband instability. Definition of the small bifurcation parameter $\varepsilon$.

with $r \geq 0, z \in S^d, \varphi_{l,n}$ for $l \in \mathbb{N}_0$ and $m = -l, l-1, \ldots, l+1, l$ being a spherical harmonic, that is, if $\lambda_{l_0}$ is the eigenvalue having first positive real part, then by symmetry, simultaneously $2l_0 + 1$ eigenvalues cross the imaginary axis.

## Reduction of the Dimension

In order to understand the occurrence of the spatially periodic Taylor vortices in the Taylor–Couette problem and of the roll solutions and hexagons in Bénard's problem, the problems are considered with periodic boundary conditions along the unbounded directions. Then the instability of the trivial solution occurs when at least one eigenvalue crosses the imaginary axis. Generically, this happens by a simple real eigenvalue or a pair of complex-conjugate eigenvalues crossing the imaginary axis (Figure 7). Center manifold theory and the Lyapunov–Schmidt reduction allow to reduce the *a priori* infinite-dimensional bifurcation problem to a finite-dimensional one.

In case of a real eigenvalue $\lambda_1$ crossing the imaginary axis, the solution $u$ can be written as a sum of the weakly unstable mode and the stable modes, that is, $u = c_1\varphi_1 + u_r$, $(c_1 \in \mathbb{R})$, where $u_r$ lives in the closure of the span of the stable eigenfunctions $\{\varphi_2, \varphi_3, \ldots\}$. For the linearized system all solutions are attracted by the one-dimensional set $E_c = \{u \mid u_r = 0\}$, in which all solutions diverge to infinity.

For the nonlinear system and small bifurcation parameter this attracting structure survives, no longer as a linear space, but as a manifold

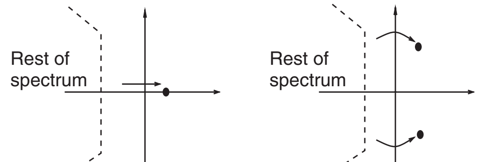



**Figure 7** Generically, a simple real eigenvalue or a pair of complex-conjugate eigenvalues cross the imaginary axis.

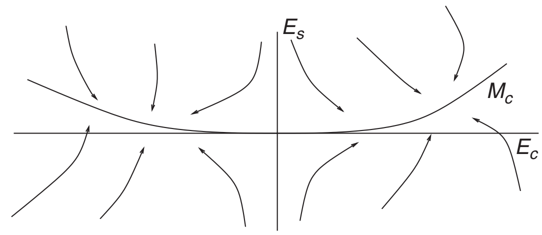



**Figure 8** The center manifold is invariant under the flow, is tangential to the central subspace $E_c$, and attracts nearby solutions with some exponential rate.

$$M_c = \{u = c_1\varphi_1 + h(c_1)| \\ h(c_1) \in \overline{\text{span}\{\varphi_2, \varphi_3, \ldots\}}\}$$

the so-called center manifold which is tangential to $E_c$, that is, $\|h(c_1)\| \leq C\|c_1\|^2$ (Figure 8). The dynamics on $M_c$ is no longer trivial due to the nonlinear terms.

Due to the fact that real problems are considered $\text{Re}\lambda_1(k_c) = 0$ implies $\text{Re}\lambda_1(-k_c) = 0$, that is, in case of $2\pi/k_c$-periodic boundary conditions always two eigenvalues cross the imaginary axis simultaneously. For Bénards's problem in a strip or for the Taylor–Couette problem in case of a bifurcation of fixed points, the reduced system on the center manifold is derived with the ansatz

$$U = \varepsilon A(\varepsilon^2 t)\mathrm{e}^{\mathrm{i}k_c x} + \text{c.c.} + \mathcal{O}(\varepsilon^2)$$

where $0 < \varepsilon \ll 1$ is the small bifurcation parameter, cf. Figure 5. Then due to $\mathrm{e}^{\mathrm{i}k_c x}\mathrm{e}^{\mathrm{i}k_c x}\mathrm{e}^{-\mathrm{i}k_c x} = \mathrm{e}^{\mathrm{i}k_c x}$ the complex-valued amplitude $A$ satisfies the so-called Landau equation

$$\partial_T A = A - \gamma A|A|^2 + \mathcal{O}(\varepsilon^2)$$

where the Landau coefficient $\gamma \in \mathbb{R}$ is obtained by classical perturbation analysis (Figure 9). The reduced system is symmetric under the $S^1$-symmetry

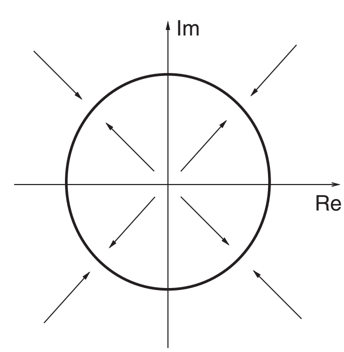



**Figure 9** The dynamics of the Landau equation. Except of the origin which corresponds to the Couette flow, all solutions converge towards the circle of fixed points, which corresponds to the family of Taylor vortices. The translation invariance of the Taylor–Couette problem is reflected by the rotational symmetry of the reduced system.

$A \mapsto Ae^{i\phi}$ with $\phi \in \mathbb{R}$ which corresponds to the translation invariance of the original systems.

This so-called equivariant bifurcation theory has been applied successfully to convection problems in the plane and on the sphere.

The stability of time-periodic flows can be analyzed with Floquet multipliers. Bifurcations from a time-periodic solution can lead to quasiperiodic motion in time. Ruelle and Takens (1971) showed that already the next bifurcation leads to chaotic dynamics. Since this time many classical hydrodynamical stability problems have been analyzed with bifurcation theory up to turbulent flows.

It was observed that center manifold theory can also be applied successfully to elliptic PDE problems posed in spatially unbounded cylindrical domains. A famous example is the construction of capillary-gravity solitary waves for the so-called water-wave problem.

## Modulation Equations

The analysis of the last section is of no use in case of a sideband instability occurring at the wave number $k_c = 0$, as it happens in the inclined-plane problem or in the Kolmogorov flow problem. Moreover, in case of an instability at a wave vector $k_c \neq 0$, based on the above analysis, front solutions cannot be described. In such situations, the method of modulation equations generalizes the role of the finite-dimensional amplitude equations from the last section.

The complex cubic Ginzburg–Landau equation in normal form is given by

$$\partial_T A = (1 + i\alpha)\partial_X^2 A + A - (1 + i\beta)A|A|^2$$

where the coefficients $\alpha, \beta \in \mathbb{R}$ are real, and we have $X \in \mathbb{R}, T \geq 0$, and $A(X,T) \in \mathbb{C}$. The Ginzburg–Landau equation is a universal amplitude equation that describes slowly varying modulations, in space and time, of the amplitude of bifurcating spatially periodic solutions in pattern-forming systems close to the threshold of the first instability. Whenever the instability drawn in Figure 5 occurs, that is, for the Taylor–Couette problem and Bénard's problem in a strip, that is, $d = 1$, it can be derived by a multiple scaling ansatz

$$u(x,t) \sim \varepsilon A(\varepsilon(x - c_g t), \varepsilon^2 t)e^{i(k_c x - \omega_0 t)} + \text{c.c.}$$

For instance, in case of $\alpha = \beta = 0$, the Ginzburg–Landau equation possesses front solutions connecting the stable fixed point $A = 1$ with the unstable fixed point $A = 0$. Such solutions correspond in the Taylor–Couette problem to modulating fronts connecting the stable Taylor vortices with the unstable Couette flow, cf. Figure 10.

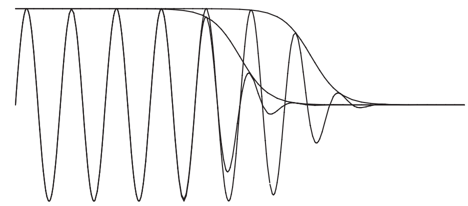


**Figure 10** The front solution of the Ginzburg–Landau equation modulates the underlying pattern in the original system.

The diffusion operator in the Ginzburg–Landau equation reflects the parabolic shape of $\text{Re}\lambda_1$ close to $k = k_c$ in Figure 5. In case of the long-wave instability, as drawn in Figure 6, the second-order differential operator changes in a fourth-order differential operator.

For Kolmogorov flow with $T = \varepsilon^4 t$ and $X = \varepsilon x$ and the amplitude scaled with $\varepsilon$, we obtain that in lowest order $A$ has to satisfy a Cahn–Hilliard equation

$$\partial_T A = -\sqrt{2}\partial_X^2 A - 3\partial_X^4 A + \gamma\partial_X^2(A^3)$$

where $A(X,T) \in \mathbb{R}$ and $\gamma \in \mathbb{R}$ a constant (cf. Figure 6).

The Kuramoto–Shivashinsky (KS)-perturbed KdV equation

$$\partial_T A = -\partial_X^3 u - \partial_X(A^2)/2 - \varepsilon(\partial_x^2 + \partial_x^4)u$$

with $A = A(X,T) \in \mathbb{R}, X \in \mathbb{R}, T \geq 0$, where $0 < \varepsilon \ll 1$ is still a small parameter, can be derived for the inclined problem with $T = \varepsilon^3 t$ and $X = \varepsilon x$ and the amplitude scaled with $\varepsilon^2$.

The theory of modulation equations is nowadays a well-established mathematical tool which allows us to construct special solutions, global existence results for the solutions of pattern-forming systems, or allows to characterize the attractors in such systems. The method is based on approximation results, showing that solutions of the original systems can be approximated by the modulation equation and attractivity results showing that every solution of the original system develops in such a way that it can be described by the modulation equation.

This method can also be applied to secondary bifurcations describing instabilities of spatially periodic wave trains. Then the so-called phase-diffusion equations, conservation laws, Burgers equations, and again the KS equations occur.

However, this method cannot be applied successfully in all situations. There are counterexamples showing that not every formally derived modulation equation describes the original system in a correct way. Moreover, very often according to some symmetries in the original problem no consistent

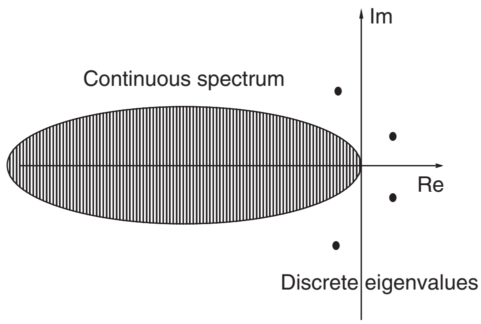



**Figure 11** Spectrum for the flow around an obstacle.

multiple scaling analysis is possible, that is, that the modulation equations still depend on $\varepsilon$.

## Discussion

There is no satisfactory bifurcation analysis for situations where boundary layers play a role. The most simple problem is the flow around some obstacle. The difficulties are according to the fact that due to the unbounded flow region there is always continuous spectrum up to the imaginary axis. From the localized obstacle discrete eigenvalues are created, (cf. Figure 11).

In such a situation, so far there is no mathematical bifurcation theory available.

*See also:* Bifurcation Theory; Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Leray–Schauder Theory and Mapping Degree; Multiscale Approaches; Newtonian Fluids and Thermohydraulics; Symmetry and Symmetry Breaking in Dynamical Systems; Turbulence Theories; Variational Methods in Turbulence.

## Further Reading

Chandrasekhar S (1961) *Hydrodynamic and Hydromagnetic Stability*. Oxford: Clarendon.

Chang H-C and Demekhin EA (2002) *Complex Wave Dynamics on Thin Films*, Studies in Interface Science, vol. 14. Amsterdam: Elsevier.

Chossat P and Iooss G (1994) *The Taylor–Couette Problem*, Applied Mathematical Sciences, vol. 102. Springer.

Chow S-N and Hale J (1982) *Methods of Bifurcation Theory*, Grundlehren der Mathematischen Wissenschaften, vol. 251. Berlin: Springer.

Golubitsky M and Schaeffer DG (1985) *Singularities and Groups in Bifurcation Theory I*, Applied Mathematical Sciences, vol. 51. Berlin: Springer.

Golubitsky M, Stewart I, and Schaeffer DG (1988) *Singularities and Groups in Bifurcation Theory II*, Applied Mathematical Sciences, vol. 69. Berlin: Springer.

Haken H (1987) *Advanced Synergetics*. Berlin: Springer.

Henry D (1981) *Geometric Theory of Semilinear Parabolic Equations*, Lecture Notes in Mathematics, vol. 840. Berlin: Springer.

Mielke A (2002) The Ginzburg–Landau equation in its role as a modulation equation. In: Fiedler B (ed.) *Handbook of Dynamical Systems II*, pp. 759–834. Amsterdam: North-Holland.

Ruelle D and Takens F (1971) On the nature of turbulence. *Communications in Mathematical Physics* 20: 167–192.

Temam R (1988) *Infinite-Dimensional Systems in Mechanics and Physics*. Berlin: Springer.

# Bifurcations of Periodic Orbits

**J-P Françoise**, Université P.-M. Curie, Paris VI, Paris, France



## Introduction

Bifurcation theory of periodic orbits relates to modeling of quite diverse subjects. It appeared classically in the field of celestial mechanics with the contributions of H Poincaré. Van der Pol (1926, 1927, 1928, 1931) observed the frequency-locking phenomenon in electrical circuits. More recently, Malkin's theory (Malkin 1952, 1956, Roseau 1966) was used to justify synchronization of weakly coupled oscillators modeling the electrical activity of the cells of the sinusal node in the heart. This article provides the essential mathematical background necessary for existence of frequency locking. Applications can be found, for instance, in Weakly Coupled Oscillators.

## The Asymptotic Phase of a Stable Periodic Orbit

Let $\Gamma$ be a periodic orbit of a vector field and let $S(\Gamma)$ denote the stable manifold of $\Gamma$ (resp. $U(\Gamma)$ denotes the unstable manifold of $\Gamma$). The following theorem can be found, for instance, in Hartman (1964).

**Theorem** *There exist $\alpha$ and $K$ such that $\mathrm{Re}(\lambda_j) < -\alpha$, $j = 1, \ldots, k$ and $\mathrm{Re}(\lambda_j) > \alpha$, $j = k+1, \ldots,$ and for all $x \in S(\Gamma)$, there is an asymptotic phase $t_0$ such that for all $t \geq 0$*

$$| \phi_t(x) - \gamma(t - t_0) | < K\,\mathrm{e}^{-\alpha(t/T)}$$

*Similarly, for any $x \in U(\Gamma)$, there is a $t_0$ such that $t \leq 0$,*

$$| \phi_t(x) - \gamma(t - t_0) | < K\,\mathrm{e}^{\alpha(t/T)}$$

If the periodic orbit is stable, the local stable manifold coincides with an open neighborhood of $\Gamma$. In such a case, there is a foliation of this open set

whose leaves are the points with a given asymptotic phase. The asymptotic phase can be considered as a coordinate function $\phi$ defined on the neighborhood $S(\Gamma)$.

If we consider now the particular case of a plane system, this function can be completed with the square of the distance function to the orbit into a coordinate system called the "amplitude–phase" system and denoted as $(\rho, \phi)$.

## Frequency Locking and Phase Locking

The term "oscillator" has two meanings. A conservative "oscillator" is a plane vector field which displays an open set of periodic orbits. It is said to be isochronous if all orbits have same period. A dissipative "oscillator" is a planar vector field which displays an attractive limit cycle (attractive periodic orbit).

We consider $N$ dissipative oscillators:

$$\begin{aligned}\frac{dx_i}{dt} &= f(x_i, y_i)\\ \frac{dy_i}{dt} &= g(x_i, y_i)\end{aligned} \qquad [1]$$

where $i = 1, \ldots, m$.

The dynamical system obtained by considering the space of all the variables $(x_i, y_i)$, $i = 1, \ldots, m$, displays an invariant torus full of periodic orbits that we denote by $T^m(0)$.

Assume now that the $N$ oscillators are weakly coupled:

$$\begin{aligned}\frac{dx_i}{dt} &= f(x_i, y_i) + \epsilon F_i(x, y, \epsilon)\\ \frac{dy_i}{dt} &= g(x_i, y_i) + \epsilon G_i(x, y, \epsilon)\end{aligned} \qquad [2]$$

where $\epsilon$ can be considered as small as we wish.

**Definition** The system [2] has a frequency locking if it displays a family of stable periodic orbits $\Gamma_\epsilon$ for all values of $\epsilon$ small enough which tends to (in the sense of Hausdorff's topology) a periodic orbit of [1] contained in the periodic torus $T^m(0)$.

Assume now that [2] has a frequency locking associated with the periodic orbit $\Gamma(t)$. Consider the projections $\Gamma_i(t)$ of $\Gamma(t)$ on the coordinates plane $(x_i, y_i)$, $i = 1, \ldots, m$. Assume that $\epsilon$ is small enough so that the projection belongs to the open set $S_i$ on which are defined the "amplitude–phase" coordinates of the system [1]. We can write the system [2], restricted to the open set $S = \Pi_{i=1}^m S_i$, as

$$\begin{aligned}\frac{d\rho_i}{dt} &= f_i(\rho, \alpha, \epsilon)\\ \frac{d\alpha_i}{dt} &= \mathbf{\Phi}_i(\rho, \alpha, \epsilon), \quad i = 1, \ldots, m\end{aligned} \qquad [3]$$

**Definition** The system [2] has a phase locking if the system induced by [3] on $\Gamma(t)$

$$\frac{d\alpha_i}{dt} = \mathbf{\Phi}_i(0, \alpha, \epsilon) \qquad [4]$$

has an attractive singular point.

As the attractive singular points are structurally stable, this is enough to assume that the system

$$\frac{d\alpha_i}{dt} = \mathbf{\Phi}_i(0, \alpha, 0) \qquad [5]$$

displays an attractive singular point.

## Periodic Orbits of Linear Systems

Consider the linear system

$$\frac{dx}{dt} = P(t) \cdot x + q(t) \qquad [6]$$

where $P$ is a continuous $T$-periodic matrix function and $q$ is a vector $T$-periodic continuous function, $x = (x_1, \ldots, x_n)$. Consider also the two associated homogeneous equations:

$$\frac{dx}{dt} = P(t) \cdot x \qquad [7a]$$

$$\frac{dx}{dt} = -P^*(t) \cdot x \qquad [7b]$$

where $P^*$ denotes the transposed of $P$.

The set of $T$-periodic solutions of [7b] is a vector space. $m$ denotes its dimension. Let $U^j(t)$, $j = 1, \ldots, m$, be a basis of this vector space. This basis is completed by adding $n - m$ solutions $U^j(t)$, $j = m+1, \ldots, n$, to obtain a basis of $R^n$. Let $U(t)$ be the matrix whose columns are these vectors; denote $U_{ij}(t)$ the elements of this matrix.

With the change of variable $x = U^*(0)^{-1}y$, system [6] gets transformed into

$$\frac{dy}{dt} = Q(t)y + r(t) \qquad [8]$$

with $Q(t) = U^*(0)P(t)U^*(0)^{-1}$ and $r(t) = U^*(0)q(t)$.

Matrix $V(t) = U^{-1}(0)U(t)$ is such that

$$\frac{dV}{dt} + Q^*(t)V = 0, \quad V(0) = I$$

and the $k$ first column vectors $V(t)$, denoted as $V^j(t)$, $j = 1, \ldots, m$, are $T$-periodic.

Let $X(t)$ be the fundamental solution defined by

$$\frac{\mathrm{d}X}{\mathrm{d}t} = Q(t) \cdot X, \quad X(0) = I$$

then,

$$X^{-1}(t) = V^*(t)$$

The solution of [8] can be written as

$$y(t) = X(t) \cdot y(0) + X(t) \cdot \int_0^t X^{-1}(u) r(u)\,\mathrm{d}u \qquad [9]$$

This yields that $T$-periodic solutions of [8] have initial data $y(0)$ given by

$$(V^*(T) - I) \cdot y(0) = \int_0^T V^*(s) r(s)\,\mathrm{d}s \qquad [10]$$

Conversely, given a solution $y(0)$ of [10], $T$-periodicity of $P$ and $q$ and uniqueness of solutions of a differential equation imply that $y(0)$ represents the initial data of a $T$-periodic solution of [8]. Hence, the $T$-periodic solutions of [8] are in one-to-one correspondence with the affine space defined by the solutions of [10]. The $m$ first rows of $V^*(T) - I$ are zero and its rank is exactly $n - m$. In the following, assume that the determinant $\Delta$ formed by the $(n - m)$ last rows and last columns of $(V^*(T) - I)$ is not zero.

A necessary and sufficient condition so that [8] displays a $T$-periodic solution is

$$\int_0^T \sum_{j=1}^n V_{jk}(u) r_j(u)\,\mathrm{d}u = 0, \quad k = 1, \ldots, m \qquad [11a]$$

$$\sum_{j=m+1}^n (V_{jk}(T) - \delta_{jk}) y_j(0) = \sum_{j=1}^n \int_0^T V_{jk}(s) r_j(s)\,\mathrm{d}s, \quad m+1 \le s \le n \qquad [11b]$$

This yields the Fredholm alternative, if the $m$ conditions,

$$\sum_{j=1}^n \int_0^T U_{jk}(s) q_j(s)\,\mathrm{d}s = 0, \quad k = 1, \ldots, m \qquad [12]$$

are satisfied, then [6] displays a family $x_\alpha(t)$ of $T$-periodic solutions depending of $m$ parameters $(\alpha_1, \ldots, \alpha_m)$:

$$x_\alpha(t) = \alpha_1 \phi_1(t) + \cdots + \alpha_m \phi_m(t) + \bar{x}(t) \qquad [13]$$

where $\bar{x}(t)$ is a particular $T$-periodic solution and $\phi_j(t)$ denote $T$-periodic independent solutions of [7a]. To be more specific, one can choose $\bar{x}(t)$ to be the unique solution of [6] such that $y(0)_k = 0, k = m+1, \ldots, n$, and $\phi_j(t)$ solutions of [7a], such that $y(0)_k = \delta_{jk}$. With these notations, $x_\alpha(t)$ is such that

$$y(0)_k = \alpha_k, \quad k = 1, \ldots, m$$

and its other initial conditions $y(0)_k = \beta_k$, $k = m + 1, \ldots, n$, are fixed:

$$\beta_k = \beta_k^0$$

## Malkin's Theorem for Quasilinear Systems

Consider now nonlinear systems with the perturbation:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = P(t) \cdot x + q(t) + \epsilon f(x, t, \epsilon) \qquad [14]$$

where $f$ is $C^1$ and $T$-periodic in $t$.

Assume that the solutions $y(t, y(0), \epsilon)$ of [14] exist for all values of $t$, $0 \le t \le T$. The solutions define a differential function of their initial data $y(0)$. This is, for instance, true for perturbations of linear systems if $\epsilon$ is small enough.

Assume that $q$ satisfies la condition [12] and that there is a solution

$$(\alpha_1^0, \ldots, \alpha_m^0)$$

to the equations

$$\psi_k(\alpha) = \sum_{j=1}^n \int_0^T U_{jk}(u) f_j(x_\alpha(u), u, 0)\,\mathrm{d}u = 0, \quad k = 1, \ldots, m \qquad [15a]$$

so that

$$\frac{\partial \psi_k(\alpha)}{\partial \alpha_j}\Big|_{\alpha=\alpha^0}; \quad k = 1, \ldots m,\ j = 1, \ldots, m \qquad [15b]$$

is invertible.

Proceed as in previous section with the coordinate change $x = U^*(0)^{-1} y$. Equation [14] gets transformed into

$$\frac{\mathrm{d}y}{\mathrm{d}t} = Q(t) y + r(t) + \epsilon F(y, t, \epsilon) \qquad [16]$$

with $F = U^*(0) f(U^*(0)^{-1} \cdot y, t, \epsilon)$.

Solutions of [16] are uniquely determined by their initial data. We can understand the parameters $(\alpha, \beta)$ as coordinates on the space of solutions. With this viewpoint, for instance, the set of $T$-periodic solutions of [6] is an affine space of dimension $m$

given by the equations $\beta = \beta^0$ and is parametrized by the coordinates $\alpha$. In this space, we pick up a point (which corresponds to a particular $T$-periodic solution of [6]): $(\alpha = \alpha^0)$. $T$-periodic solutions of [16] are in one-to-one correspondence with the solutions of

$$C_k(\alpha, \beta, \epsilon) = \sum_{j=1}^{n} \int_0^T V_{jk}(s) F_j(y(s, \epsilon, \alpha, \beta), s, \epsilon) \mathrm{d}s = 0, \quad k = 1, \ldots, m \qquad [17a]$$

$$\begin{aligned} C_k(\alpha, \beta, \epsilon) = &\sum_{j=m+1,\ldots,n} (V_{jk}(T) - I)\beta_j \\ &- \sum_{j=1}^{n} \int_0^T V_{jk}(s) r_j(s)\, \mathrm{d}s \\ &- \epsilon \sum_{j=1}^{n} \int_0^T V_{jk}(s) F_j(y(s, \epsilon, \alpha, \beta), s, \epsilon) \mathrm{d}s = 0, \\ &k = m+1, \ldots, n \end{aligned} \qquad [17b]$$

where $\alpha_k, k = 1, \ldots, m$ and $\beta_k = y_k(0)$, $k = m + 1, \ldots, n$ parametrize the solutions $y(t, \epsilon, \alpha, \beta)$ of [14] in this way:

$$y(0) = U^*(0) \cdot x(0), \quad x(0) = \sum_{j=1}^{m} \alpha_j \phi_j(0) + \bar{x}(0) \qquad [18]$$

Consider the determinant of the Jacobian matrix of the mapping

$$(\alpha, \beta) \mapsto C(\alpha, \beta, \epsilon) \qquad [19]$$

for $\alpha = \alpha^0$, $\beta_k = \beta_k^0$, $k = m+1, \ldots, n$, $\epsilon = 0$. This is equal to the product of $\Delta$ and the determinant of

$$\frac{\partial \psi_k(\alpha)}{\partial \alpha_j}\Big|_{\alpha=\alpha^0} \qquad [20]$$

which is nonzero.

The implicit-function theorem shows that the differential equation [14] (and thus [16] as well) has, for $\epsilon$ small enough, a unique $T$-periodic solution which tends to $x_{\alpha^0}$ when $\epsilon$ tends to 0.

## Generalization of Malkin's Theorem

Finally, we consider the most general situation of the perturbation of a general system (not necessarily linear):

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x, t) + \epsilon g(x, t, \epsilon) \qquad [21]$$

where we assume that

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x, t) \qquad [22]$$

displays an $m$-parameter family $x_\alpha(t)$ of $T$-periodic orbits.

Assume that the solutions $y(t, y(0), \epsilon)$ exist for all $0 \le t \le T$ and define a differentiable mapping of the initial data $y(0)$. This is, for instance, the case if we assume that the nonperturbed equation defines a flow and if $\epsilon$ is small enough.

Assume also that the different solutions $x_\alpha(t)$ are independent in the sense that the mapping

$$\alpha \mapsto x_\alpha(t)$$

is an immersion for any $t$. In other words, the $m$ vectors $\mathrm{d}x_\alpha(t)/\mathrm{d}\alpha_j$ are independent.

We linearize the solution along the family of periodic orbits:

$$x = x_\alpha(t) + \epsilon\xi \qquad [23]$$

Equation [21] gets transformed into

$$\frac{\mathrm{d}\xi}{\mathrm{d}t} = Df_x(x_\alpha(t), t) \cdot \xi + g(x_\alpha(t), t, 0) + \epsilon F(\xi, t, \epsilon) \qquad [24]$$

Set, furthermore,

$$P(t) = Df_x(x_\alpha(t), t), \quad r(t) = g(x_\alpha(t), t, 0)$$

and denote $U(t)$ the fundamental solution of [7b] described earlier.

**Theorem** *Assume that there is a solution*

$$(\alpha_1^0, \ldots, \alpha_m^0)$$

*of the m equations*:

$$\sigma_k(\alpha) = \sum_{j=1}^{n} \int_0^T U_{jk}(u) g_j(x_\alpha(u), u, 0)\, \mathrm{d}u = 0, \quad k = 1, \ldots, m \qquad [25a]$$

*such that*

$$\frac{\partial \sigma_k(\alpha)}{\partial \alpha_j}\Big|_{\alpha=\alpha^0}; \quad k = 1, \ldots m, \ j = 1, \ldots, m \qquad [25b]$$

*is invertible. Then, for all $\epsilon$ sufficiently small*, eqn [21] *has a unique T-periodic solution which tends to $x_{\alpha^0}$ when $\epsilon$ tends to* 0.

We show that under the hypothesis of the theorem, we can apply the results proved in the preceding section. Note that one can prove the theorem for eqn [24] because it reduces to [21] with the change of variables [23].

Note first that the $m$ conditions [25a] imply that the $m$ equations,

$$\frac{d\xi}{dt} = Df_x(x_{\alpha^0}(t), t) \cdot \xi + g(x_{\alpha^0}(t), t, 0)$$

display a family of $T$-periodic solutions which depend on $m$ parameters $\gamma = (\gamma_1, \ldots, \gamma_m)$. From (13), one can write

$$\xi_\gamma(t) = \gamma_1\phi_1(t) + \cdots + \gamma_m\phi_m(t) + \bar{\xi}(t) \qquad [26]$$

where $\bar{\xi}(t)$ is a particular $T$-periodic solution and the $\phi_j(t)$ are independent $T$-periodic solutions of (22a).

**Lemma 1** *A possible choice for the solutions* $\phi_j(t)$ is $\partial x_\alpha(t)/\partial\alpha_j\,|_{\alpha=\alpha^0}$.

We have already assumed that these vectors are independent. They are obviously $T$-periodic solutions to (22a).

In the following, we will assume that all other periodic solutions of (22a) are linear combinations of these.

As a consequence of what was proved in the section on periodic orbits of linear systems, system [24] displays a periodic solution (for $\epsilon$ small enough) if there exists a solution

$$(\gamma_1^0, \ldots, \gamma_m^0)$$

to equations

$$\nu_k(\gamma) = \sum_{j=1}^{n} \int_0^T U_{jk}(s) F_j(\xi_\gamma(s), s, 0)\, ds = 0,$$
$$k = 1, \ldots, m$$

such that

$$\frac{\partial\nu_k(\gamma)}{\partial\gamma_j}\Big|_{\gamma=\gamma^0}; \quad k = 1, \ldots m,\ j = 1, \ldots, m$$

is invertible.

**Lemma 2** *The quantities* $\nu_k(\gamma)$ *depend linearly in* $\gamma$.

**Proof** Observe first that the quantities $F_j(\xi, s, 0)$ depend quadratically of $\xi$:

$$F_j(\xi, s, 0) = \frac{1}{2}\sum_{k,l} \frac{\partial^2 f_j}{\partial z_k \partial z_l}(x_{\alpha^0}(s), s)\xi_k\xi_l + \sum_k \frac{\partial g_j}{\partial z_k}(x_{\alpha^0}(s), s, 0) + \frac{\partial g_j}{\partial\epsilon}(x_{\alpha^0}(s), s, 0) \qquad [27]$$

Then, the solutions $\xi(t)$ depend linearly on $\gamma$. We thus obtain that *a priori* $\nu_p(\gamma)$ are quadratic functions of $\gamma$:

$$\nu_p(\gamma_1, \ldots, \gamma_m) = \frac{1}{2}\sum_{qrkl} \gamma_q\gamma_r \int_0^T U_{jp} \frac{\partial^2 f_j}{\partial z_k \partial z_l} \cdot \frac{\partial z_k}{\partial\gamma_q} \cdot \frac{\partial z_l}{\partial\gamma_r}\, ds + \sum_{qkl} \gamma_q \int_0^T U_{jp} \left[ \frac{1}{2}\frac{\partial^2 f_j}{\partial z_k \partial z_l} \left( \frac{\partial z_k}{\partial\gamma_q} \cdot \bar{\xi}_l + \frac{\partial z_l}{\partial\gamma_q}\bar{\xi}_k \right) + \frac{\partial g_j}{\partial z_k} \cdot \frac{\partial z_k}{\partial\gamma_q} \right] ds + \cdots \qquad [28]$$

where the dots represent quantities independent of $\gamma$. We use then the expression

$$\frac{d}{dt}\left(\frac{\partial^2 z_j}{\partial\gamma_q \partial\partial_r}\right) = \sum_{kl} \frac{\partial^2 f_j}{\partial z_k \partial z_l} \cdot \frac{\partial z_k}{\partial\gamma_q} \cdot \frac{\partial z_l}{\partial\gamma_r} + \sum_k \frac{\partial f_j}{\partial z_k} \frac{\partial^2 z_k}{\partial\gamma_q \partial\partial_r}$$

This allows one to find the homogeneous quadratic part as

$$\sum_{jkl} \int_0^T U_{jp} \frac{\partial^2 f_j}{\partial z_k \partial z_l} \cdot \frac{\partial z_k}{\partial\gamma_q} \cdot \frac{\partial z_l}{\partial\gamma_r}\, ds = \sum_j \int_0^T U_{jp}(s) \frac{d}{ds}\left(\frac{\partial^2 z_j}{\partial\gamma_q \partial\partial_r}\right) ds - \sum_{jk} \int_0^T U_{jp}(s) \frac{\partial f_j}{\partial z_k} \frac{\partial^2 z_k}{\partial\gamma_q \partial\gamma_r}\, ds$$

Integration by parts yields

$$\sum_{jkl} \int_0^T U_{jp} \frac{\partial^2 f_j}{\partial z_k \partial z_l} \cdot \frac{\partial z_k}{\partial\gamma_q} \cdot \frac{\partial z_l}{\partial\gamma_r}\, ds = -\sum_j \int_0^T \left( \frac{dU_{jp}}{ds} + U_{jp}(s)\frac{\partial f_j}{\partial z_k} \right) \frac{\partial^2 z_k}{\partial\gamma_q \partial\gamma_r}\, ds = 0$$

because $U^*$ is solution to [7a]. This shows that [28] is linear in $\gamma$. Suffices to show that the determinant of this system does not vanish to have existence and uniqueness of the solution such that

$$\frac{\partial\nu_1, \ldots, \nu_m}{\partial\gamma_1, \ldots, \gamma_m} \neq 0$$

Consider now the coefficient of the linear part:

$$\sum_{kl} \int_0^T U_{jp} \left[ \frac{\partial^2 f_j}{\partial z_k \partial z_l} \cdot \bar{\xi}_l + \frac{\partial g_j}{\partial z_k} \right] \cdot \frac{\partial z_k}{\partial\gamma_q}\, ds$$

and the coefficient

$$\sigma_p(\alpha) = \sum_{j=1}^{n} \int_0^T U_{jp}(u) g_j(x_\alpha(u), u, 0)\, \mathrm{d}u$$

We can write

$$\frac{\mathrm{d}\sigma_p}{\mathrm{d}\alpha_q} = \int_0^T \left( \frac{\partial U_{jp}}{\partial \alpha_q} \cdot g_j + U_{jp} \frac{\partial g_j}{\partial z_k} \cdot \frac{\partial z_k}{\partial \alpha_q} \right) \mathrm{d}s$$

Note that

$$\frac{\mathrm{d}\bar{\xi}_j}{\mathrm{d}t} = \sum_r \frac{\partial f_j}{\partial z_r} \bar{\xi}_r + g_j(z(t), \alpha^0, 0)$$

and we obtain

$$\frac{\mathrm{d}\sigma_p}{\mathrm{d}\alpha_q} = \int_0^T \left( \frac{\partial U_{jp}}{\partial \alpha_q} \cdot \left( \frac{\mathrm{d}\bar{\xi}_j}{\mathrm{d}s} - \sum_r \frac{\partial f_j}{\partial z_r} \bar{\xi}_r \right) + U_{jp} \frac{\partial g_j}{\partial z_k} \cdot \frac{\partial z_k}{\partial \alpha_q} \right) \mathrm{d}s$$

Integration by parts yields

$$\left. \frac{\mathrm{d}\sigma_p}{\mathrm{d}\alpha_q} \right|_{\alpha=\alpha^0} = -\int_0^T \left( \frac{\mathrm{d}}{\mathrm{d}s} \left( \frac{\partial U_{jp}}{\partial \alpha_q} \right) \cdot \bar{\xi}_j + \sum_r \frac{\partial f_j}{\partial z_r} \bar{\xi}_r \right) + \int_0^T U_{jp} \left( \frac{\partial g_j}{\partial z_k} \cdot \frac{\partial z_k}{\partial \alpha_q} \right) \mathrm{d}s$$

From the equation

$$\frac{\mathrm{d}U_{jp}}{\mathrm{d}t} + \sum_k \frac{\partial f_k}{\partial z_j} U_{kp} = 0$$

we deduce that

$$\frac{\mathrm{d}}{\mathrm{d}t} \left( \frac{\partial U_{jp}}{\partial \alpha_q} \right) = -\sum_k \frac{\partial f_k}{\partial z_j} \frac{\partial U_{jp}}{\partial \alpha_q} + \sum_k \frac{\partial^2 f_k}{\partial z_j \partial z_r} U_{kp} \frac{\partial z_r}{\partial \alpha_q}$$

and thus this shows that

$$\left. \frac{\mathrm{d}\sigma_p}{\mathrm{d}\alpha_q} \right|_{\alpha=\alpha^0} = \sum_{kl} \int_0^T U_{jp} \left[ \frac{\partial^2 f_j}{\partial z_k \partial z_l} \cdot \bar{\xi}_l + \frac{\partial g_j}{\partial z_k} \right] \cdot \frac{\partial z_k}{\partial \gamma_q} \mathrm{d}s$$

This achieves the proof of the theorem. In the special case of Hamiltonian systems, in the case of the peturbations of an isochronous system, the method explained is equivalent to Moser's averaging theory.

The reader is referred to other articles in this encyclopedia for a discussion of other aspects of synchronization, frequency locking, and phase locking.

*See also:* Bifurcation Theory; Fractal Dimensions in Dynamics; Integrable Systems: Overview; Isochronous Systems; Leray–Schauder Theory and Mapping Degree; Ljusternik–Schnirelman Theory; Singularity and Bifurcation Theory; Symmetry and Symmetry Breaking in Dynamical Systems; Synchronization of Chaos; Weakly Coupled Oscillators.

## Further Reading

Hartman P (1964) *Ordinary Differential Equations*. New York: Wiley.
Malkin I (1952) *Stability Theory of the Motion*. Moscou–Leningrad: Izdat. Gos.
Malkin I (1956) *Some Problems in the Theory of Nonlinear Oscillations*. Gostekhisdat.
Moser J (1970) Regularization of Kepler's problem and the averaging method on a manifold. *Communication of Pure and Applied Mathematics* 23: 609–636.
Roseau M (1966) *Vibrations non linéaires et théorie de la stabilité*, Springer Tracts in Natural Philosophy, vol. 8. Berlin: Springer.
Van der Pol B (1926) On relaxation-oscillations. *Philosophical Magazine* 3(7): 978–992.
Van der Pol B (1931) Oscillations sinusoidales et de relaxation. *L'onde électrique* 245–256.
Van der Pol B and Van der Mark J (1927) Frequency demultiplication. *Nature* 120: 363–364.
Van der Pol B and Van der Mark J (1928) The heart beat considered as a relaxation oscillation, and an electrical model of the heart. *Philosophical Magazine* 6(7): 763–775.

# Bi-Hamiltonian Methods in Soliton Theory

**M Pedroni**, Università di Bergamo, Dalmine (BG), Italy



## Introduction

At the end of the 1960s, the theory of integrable systems received a great boost by the discovery (made by Gardner, Green, Kruskal, and Miura) of the inverse-scattering method (*see* Integrable Systems: Overview). It allows one to reduce the solution of the (nonlinear) Korteweg–de Vries equation (henceforth simply the KdV equation)

$$u_t = \tfrac{1}{4}(u_{xxx} - 6uu_x) \qquad [1]$$

to the solution of linear equations. After the KdV equation, a lot of other nonlinear partial differential equations, solvable by means of the inverse-scattering method, were found out. A common feature of such equations is the existence of soliton solutions, that is, solutions in the shape of a solitary wave (with additional interaction properties). For this reason they are called "soliton equations."

It was soon observed that the KdV equation can be seen as an infinite-dimensional Hamiltonian system with an infinite sequence of constants of motion in involution; the corresponding (commuting) vector fields are symmetries for the KdV equation, and form the so-called KdV hierarchy. In particular, Zakharov and Faddeev constructed action-angle variables for the KdV equation. These facts pointed out that the KdV equation is an infinite-dimensional analog of a classical integrable Hamiltonian system (Dubrovin *et al.* 2001), whose theory has been developed during the nineteenth century by Liouville, Jacobi, and many others. Moreover, the infinite-dimensional case suggested methods (such as the existence of a Lax pair) which were applied successfully also to finite-dimensional cases such as the Toda lattices and the Calogero systems. More recently, after the discovery by Witten and Kontsevich of remarkable relations between the KdV hierarchy and matrix models of two-dimensional (2D) quantum gravity, there has been a renewed interest in the study of soliton equations in the community of theoretical physicists. We also mention that the classical versions of the extended $\mathcal{W}_n$-algebras of 2D conformal field theory are the (second) Poisson structures of the Gelfand–Dickey hierarchies.

In this article we describe the so-called bi-Hamiltonian formulation of soliton equations. This approach to integrable systems springs from the observation, made by Magri at the end of the 1970s, that the KdV equation can be seen as a Hamiltonian system in two different ways. In the same circle of ideas, there were important works by Adler, Dorfman, Gelfand, Kupershmidt, Wilson, and many others. Thus, the concept of bi-Hamiltonian manifold, which constitutes the geometric setting for the study of bi-Hamiltonian systems, emerged. This notion and its applications to the theory of finite-dimensional integrable systems is discussed in Multi-Hamiltonian Systems.

In the first section of this article, we discuss the Hamiltonian form of soliton equations and, more generally, we present an important class of infinite-dimensional Poisson (also called Hamiltonian) structures, namely those of hydrodynamic type. Then we show how to use the bi-Hamiltonian properties of the KdV equation in order to construct its conserved quantities. We also recall that the KdV equation can be seen as an Euler equation on the dual of the Virasoro algebra. In the third section, we deal with other examples of integrable evolution equations admitting a bi-Hamiltonian representation, that is, the Boussinesq and the Camassa–Holm equations, and we consider the bi-Hamiltonian structures of hydrodynamic type.

## Hamiltonian Methods in Soliton Theory

The most famous example of soliton equation is the KdV equation [1], where $u$ is usually a periodic or rapidly decreasing real function. The choice of the coefficients in the equation has no special meaning, since they can be changed arbitrarily by rescaling $x$, $t$, and $u$. Right after the discovery of the inverse-scattering method for solving the Cauchy problem for the KdV equation, it was realized that this equation can be seen as an infinite-dimensional Hamiltonian system. Indeed, from a geometrical point of view, eqn [1] defines a vector field $X(u)=(1/4)(u_{xxx}-6uu_x)$ on $\mathcal{M}$, the infinite-dimensional vector space of $C^\infty$ functions from the unit circle $S^1$ to $\mathbb{R}$. (For the sake of simplicity, we consider only the periodic case; the integrals in this article are therefore understood to be taken on $S^1$.) The vector field $X$ associated with the KdV equation is Hamiltonian, that is, it can be factorized as

$$X(u) = [-2\partial_x]\left[\tfrac{1}{8}(-u_{xx}+3u^2)\right]$$

where $\mathrm{d}H=(1/8)(-u_{xx}+3u^2)$ is the differential of the functional

$$H(u)=\frac{1}{8}\int\left(u^3+\frac{1}{2}u_x^2\right)\mathrm{d}x$$

that is, the variational derivative $\delta h/\delta u$ of the density $h=(1/8)(u^3+(1/2)u_x^2)$, and $P=-2\partial_x$ is a Poisson (or Hamiltonian) operator. This means that the corresponding composition law

$$\{F,G\}=\int \mathrm{d}F\,P(\mathrm{d}G)\,\mathrm{d}x=-2\int \mathrm{d}F\,(\mathrm{d}G)_x\,\mathrm{d}x \qquad [2]$$

between functionals of $u$ has the usual properties of the Poisson bracket, that is, it is $\mathbb{R}$-bilinear and skew-symmetric, and it fulfills the Leibniz rule and the Jacobi identity. In other words, $(\mathcal{M},P)$ is an infinite-dimensional Poisson manifold. Using the Poisson bracket [2], eqn [1] can be written as

$$u_t=\{u,H\} \qquad [3]$$

corresponding to the usual Hamilton equation in $\mathbb{R}^{2n}$

$$\dot{z}^i=\{z^i,H\},\quad i=1,\ldots,2n \qquad [4]$$

up to the replacement of $z$ with $u$, and of the discrete index $i$ with the continuous index $x$. More precisely, in the expression $u_t=\{u,H\}$ the symbol $u$ should be replaced by $u^x$ (in analogy with $z^i$), the functional assigning to the generic function $v\in\mathcal{M}$ its value at a fixed point $x$, that is, $u^x: v\mapsto v(x)$. In

these notations, the Poisson bracket [2] takes the form

$$\{u^x, u^y\} = -2\delta'(x-y)$$

where the $\delta$-function is as usual defined as

$$\int f(y)\delta(x-y)\,\mathrm{d}x = f(x)$$

so that its derivatives are given by

$$\int f(y)\delta^{(k)}(x-y)\,\mathrm{d}x = f^{(k)}(x)$$

Another important example is given by the Boussinesq equation

$$u_{tt} = \tfrac{1}{3}\left(-u_{xxxx} + 4u_x^2 + 4uu_{xx}\right) \qquad [5]$$

describing, like KdV, shallow water (soliton) waves in a nonlinear approximation. It can be obtained by the first-order (in time) system

$$u^1{}_t = \tfrac{2}{3}u^2u_x^2 + u_{xx}^1 - \tfrac{2}{3}u_{xxx}^2, \quad u^2{}_t = 2u_x^1 - u_{xx}^2 \qquad [6]$$

by taking the derivative of its second equation with respect to $t$, plugging the result in the first one, and setting $u = u^2$. The system [6] is Hamiltonian, since it can be written as

$$u^1{}_t = \left(\frac{\delta h}{\delta u^2}\right)_x, \qquad u^2{}_t = \left(\frac{\delta h}{\delta u^1}\right)_x$$

with $h = (u^1)^2 + (1/9)(u^2)^3 - u^1u_x^2 + (1/3)(u_x^2)^2$, and

$$\begin{pmatrix} 0 & \partial_x \\ \partial_x & 0 \end{pmatrix} \qquad [7]$$

is easily seen to be a Poisson operator. Thus, the Poisson manifold associated with the Boussinesq equation is the space of periodic $C^\infty$ functions with values in $\mathbb{R}^2$. More generally, one can consider the space $\mathcal{M}^n$ of $C^\infty$ functions from the unit circle $S^1$ to $\mathbb{R}^n$. If $P^{ij}$, for $i, j = 1, \ldots, n$, are the entries of a constant skew-symmetric matrix and $u^{i,x}$ assigns to the generic function $v \in \mathcal{M}^n$ the value of its $i$th components at a fixed point $x$, then

$$\{u^{i,x}, u^{j,y}\} = P^{ij}\delta(x-y)$$

defines a Poisson bracket on $\mathcal{M}^n$. One can also let the $P^{ij}$ depend on the $u^k$ in such a way that they form the components of a Poisson tensor on $\mathbb{R}^n$. If $H = \int h\,\mathrm{d}x$ is a functional on $\mathcal{M}^n$ with density $h$, the associated Hamiltonian vector field gives rise to the following system of partial differential equations:

$$u_t^i = \sum_{j=1}^n P^{ij}\frac{\delta h}{\delta u^j}, \quad i = 1, \ldots, n$$

In particular, if $n = 2N$ and

$$[P^{ij}] = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

then we have the Hamiltonian formulation of the field equations,

$$q_t^i = \frac{\delta h}{\delta p^i}, \qquad p_t^i = -\frac{\delta h}{\delta q^i}, \quad i = 1, \ldots, N$$

Another important example of Poisson bracket on $\mathcal{M}^n$ is given by

$$\{u^{i,x}, u^{j,y}\} = g^{ij}\delta'(x-y) \qquad [8]$$

where $g^{ij}$ are the entries of a constant symmetric matrix. In this case, the Hamiltonian vector field associated with $H = \int h\,\mathrm{d}x$ is given by

$$u_t^i = \sum_{j=1}^n g^{ij}\partial_x\left(\frac{\delta h}{\delta u^j}\right), \quad i = 1, \ldots, n \qquad [9]$$

Notice that this vector field is zero if $H = \int u^k\,\mathrm{d}x$, with $k = 1, \ldots, n$. This amounts to saying that such an $H$ is a Casimir function of the Poisson bracket [8], that is, that $\{H, F\} = 0$ for all functionals $F$. A simple example of this class (with $n = 2$) is given by the Poisson structure of the Boussinesq equation, corresponding to the choice $g^{11} = g^{22} = 0$ and $g^{12} = g^{21} = 1$. Suppose now that the matrix with entries $g^{ij}$ is invertible. Then they can be interpreted as the contravariant components of a flat pseudo-Riemannian metric in $\mathbb{R}^n$. A change of coordinates $(u^1, \ldots, u^n) \mapsto (\bar{u}^1, \ldots, \bar{u}^n)$ in $\mathbb{R}^n$ transforms the Poisson bracket [9] in

$$\{\bar{u}^{i,x}, \bar{u}^{j,y}\} = g^{ij}(\bar{u})\delta'(x-y) + \Gamma_k^{ij}(\bar{u})\bar{u}_x^k\delta(x-y) \qquad [10]$$

where $g^{ij}(\bar{u})$ are the components of the metric in the new coordinates and the $\Gamma_k^{ij}$ are the contravariant Christoffel symbols related to the usual Christoffel symbols by

$$\Gamma_k^{ij} = -g^{il}\Gamma_{lk}^j \qquad [11]$$

Conversely, the expression [10] gives a Poisson bracket if the metric defined by $g^{ij}$ is flat and its Christoffel symbols are related to the $\Gamma_k^{ij}$ by [11]. These are the Poisson structures of hydrodynamic type introduced by Dubrovin and Novikov. We will consider them again later.

## Bi-Hamiltonian Formulation of the KdV Equation

The KdV equation [1] has a lot of remarkable properties, such as the Lax representation and the existence of a $\tau$-function. In this section, we recall a geometrical feature of KdV, namely, the fact that it

has a second Hamiltonian structure, and we show that the integrability of KdV can be seen as a natural consequence of its double Hamiltonian representation. We have already seen that the KdV vector field $X(u) = (1/4)(u_{xxx} - 6uu_x)$ can be written as

$$X(u) = P_0\, \mathrm{d}H_2$$

where $P_0 = -2\partial_x$ and

$$H_2 = \frac{1}{8}\int \left(u^3 + \frac{1}{2}u_x^2\right) \mathrm{d}x$$

But $X$ admits another Hamiltonian representation:

$$X(u) = P_1\, \mathrm{d}H_1$$

where $P_1 = -(1/2)\partial_{xxx} + 2u\partial_x + u_x$ and

$$H_1 = -\frac{1}{4}\int u^2\, \mathrm{d}x$$

The important point is that $P_1$ is also a Poisson operator. Moreover, it is compatible with $P_0$, that is, any linear combination of $P_0$ and $P_1$ is still a Poisson operator. Thus, the KdV equation is a bi-Hamiltonian system, that is, it can be seen in two different (but compatible) ways as a Hamiltonian system. Next, we will show how this property can be used to construct an infinite sequence of conserved quantities for the KdV equation, which are in involution with respect to the Poisson brackets $\{\cdot\,,\cdot\}_0$ and $\{\cdot\,,\cdot\}_1$ associated with $P_0$ and $P_1$. In particular, the phase space $\mathcal{M}$ of KdV is a bi-Hamiltonian manifold, that is, it has two different (but compatible) Poisson structures. Let us rename $X_1 = X$ the KdV vector field. Since $X = P_0\, \mathrm{d}H_2 = P_1\, \mathrm{d}H_1$, one is naturally led to consider the vector fields

$$X_0 = P_0\, \mathrm{d}H_1, \quad X_2 = P_1\, \mathrm{d}H_2$$

Explicitly, $X_0(u) = u_x$ and $X_2(u) = (1/16)(u_{xxxxx} - 10uu_{xxx} - 20u_xu_{xx} + 30u^2u_x)$. One can check that these vector fields are also bi-Hamiltonian. Indeed, $X_0(u) = P_1\, \mathrm{d}H_0$, with $H_0 = \int u\, \mathrm{d}x$, and

$$X_2 = P_0\, \mathrm{d}H_3 \quad \text{with}$$
$$H_3 = -\frac{1}{64}\int \left(u_{xx}^2 + 5uu_x^2 + \frac{5}{2}u^4\right) \mathrm{d}x$$

The functional $H_0$ is a Casimir of $P_0$, that is, $P_0\, \mathrm{d}H_0 = 0$, so that the iteration ends on this side, but it can be continued indefinitely from the other side, as shown below. For the time being, let us take for granted that there exists an infinite sequence $\{H_k\}_{k\geq 0}$ of functionals such that $P_1\, \mathrm{d}H_k = P_0\, \mathrm{d}H_{k+1}$; in other words,

$$\{\cdot, H_k\}_1 = \{\cdot, H_{k+1}\}_0 \qquad [12]$$

Such relations are often called Lenard–Magri relations. Then the functionals $H_k$ are in involution with respect to both Poisson brackets. Indeed, for $k > j$, one has

$$\{H_j, H_k\}_0 = \{H_j, H_{k-1}\}_1 = \{H_{j+1}, H_{k-1}\}_0$$
$$= \cdots = \{H_k, H_j\}_0$$

so that $\{H_j, H_k\}_0 = 0$ for all $j, k \geq 0$, and therefore $\{H_j, H_k\}_1 = 0$ for all $j, k \geq 0$. Hence, these functionals are constants of motion (in involution) for the KdV equation. The Hamiltonian vector fields associated with them are symmetries for the KdV equation; the corresponding evolution equations are called higher-order KdV equations. The set of such equations is the well-known KdV hierarchy. We remark that the existence of a sequence of functionals $\{H_k\}_{k\geq 0}$, fulfilling the Lenard–Magri relations [12] and starting from a Casimir of $P_0$, is equivalent to the existence of a Casimir function $H(\lambda) = \sum_{k\geq 0} H_k\lambda^{-k}$ for the Poisson pencil $P_\lambda = P_1 - \lambda P_0$, where $\lambda$ is a real parameter. A straightforward way (due essentially to Miura, Gardner, and Kruskal) to determine such a Casimir function is to consider the (generalized) Miura map $h \mapsto u = h_x + h^2 - \lambda$. As shown by Kupershmidt and Wilson, it transforms the Poisson structure $(1/2)\partial_x$ (in the variable $h$) into the Poisson pencil $P_\lambda = -(1/2)\partial_{xxx} + 2(u + \lambda)\partial_x + u_x$. Given $u$, the Riccati equation

$$h_x + h^2 = u + \lambda \qquad [13]$$

admits a unique solution with the asymptotic expansion $h = z + \sum_{k\geq 1} h_k z^{-k}$, where $z^2 = \lambda$. Moreover, the coefficients $h_k$ are differential polynomials in $u$ (i.e., polynomials in $u$ and its $x$-derivatives) that can be computed by recurrence. Thus, the generalized Miura map can be seen as an invertible transformation. Since the functional $h \mapsto \int h\, \mathrm{d}x$ is a Casimir of the Poisson structure $(1/2)\partial_x$, it follows that if $h(u)$ is the solution of the Riccati equation [13], then $u \mapsto \int h(u)\, \mathrm{d}x$ is a Casimir of the Poisson pencil $P_\lambda$. More precisely, one has to introduce the functional $H(\lambda) = z\int h(u)\, \mathrm{d}x$, that turns out to be a Laurent series in $\lambda$, because the even coefficients of $h(u)$ are $x$-derivatives. This is the Casimir function we were looking for. Explicitly, one finds that the first terms of $h(u)$ are

$$h_1 = \tfrac{1}{2}u, \qquad h_2 = -\tfrac{1}{4}u_x, \qquad h_3 = \tfrac{1}{8}(u_{xx} - u^2)$$
$$h_4 = -\tfrac{1}{16}(u_{xxx} - 4uu_x)$$
$$h_5 = \tfrac{1}{32}(u_{xxxx} - 6uu_{xx} - 5u_x^2 + 2u^3)$$

Obviously, $h_1$ is the density of a Casimir function of $P_0$, while $h_3$ and $h_5$ are (one-half of) the densities of the

two Hamiltonians $H_1$ and $H_2$ of the KdV equation. We conclude this section showing that, as observed by Khesin and Ovsienko (Arnol'd and Khesin 1998), the bi-Hamiltonian structures of KdV have a clear Lie-algebraic origin. Indeed, the second Hamiltonian structure is the Lie–Poisson structure on the dual of the Virasoro algebra, while the first one can be obtained by "freezing" the second one at a suitable point. Let $\mathcal{X}(S^1)$ be the Lie algebra of vector fields on $S^1$. The Virasoro algebra is the vector space $\mathfrak{g}=\mathcal{X}(S^1)\oplus\mathbb{R}$ endowed with the Lie-algebra structure

$$\left[\left(f(x)\frac{\partial}{\partial x},a\right),\left(g(x)\frac{\partial}{\partial x},b\right)\right] = \left((f'(x)g(x)-g'(x)f(x))\frac{\partial}{\partial x},\ \int f'(x)g''(x)\,dx\right) \quad [14]$$

It is called a central extension of $\mathcal{X}(S^1)$ since it is obtained by considering the usual commutator between vector fields (up to a sign) and by adding a copy of $\mathbb{R}$, which turns out to be the center of the Virasoro algebra. Equation [14] gives rise indeed to a Lie-algebra structure because the expression $\int f'(x)g''(x)\,dx$ defines a 2-cocycle of $\mathcal{X}(S^1)$. The dual space $\mathfrak{g}^*$ of $\mathfrak{g}$ can be considered as the space of the pairs $(u\,dx\otimes\,dx,c)$, where $u\in C^\infty(S^1)$ and $c\in\mathbb{R}$. The pairing is obviously given by

$$\left\langle (u\,dx\otimes\,dx,c),\left(f\frac{\partial}{\partial x},a\right)\right\rangle \int u(x)f(x)\,dx+ac$$

The Lie–Poisson structure on the dual $\mathfrak{g}^*$ of a Lie algebra $\mathfrak{g}$ is defined as

$$\{F,G\}(X)=\langle X,[dF(X),dG(X)]\rangle \quad [15]$$

where $F,G\in C^\infty(\mathfrak{g})^*$ and their differentials at $X\in\mathfrak{g}^*$ are seen as elements of $\mathfrak{g}$. When $\mathfrak{g}$ is the Virasoro algebra and $F(u,c)=\int f(u,c)\,dx$, $G(u,c)=\int g(u,c)\,dx$ are two functionals on $\mathfrak{g}^*$ whose densities $f$ and $g$ are differential polynomials in $u$, one has

$$\begin{aligned}\{F,G\}(u,c) &= \left\langle (u\,dx\otimes\,dx,c),\left(\left(\frac{\delta f}{\delta u}\right)'\left(\frac{\delta g}{\delta u}\right) - \left(\frac{\delta g}{\delta u}\right)'\left(\frac{\delta f}{\delta u}\right)\right)\frac{\partial}{\partial x},\int\left(\frac{\delta f}{\delta u}\right)'\left(\frac{\delta g}{\delta u}\right)''dx\right\rangle \\ &= \int u\left(\left(\frac{\delta f}{\delta u}\right)'\left(\frac{\delta g}{\delta u}\right)-\left(\frac{\delta g}{\delta u}\right)'\left(\frac{\delta f}{\delta u}\right)\right)dx \\ &\quad + \int c\left(\frac{\delta f}{\delta u}\right)'\left(\frac{\delta g}{\delta u}\right)''dx\end{aligned} \quad [16]$$

This is (up to rescaling) the second Poisson bracket of KdV. The KdV equation is therefore an Euler equation, that is, it can be obtained from the Euler equations for the rigid body by replacing the Lie algebra of the rotation group with the Virasoro algebra. To be more precise, the Hamiltonian vector field associated with $H_1(u,c)=-(1/2)(\int u^2\,dx+c)$ is

$$u_t+3uu_x+cu_{xxx}=0,\quad c_t=0$$

If $c\neq 0$, this is (up to rescaling) the KdV equation [1]. For $c=0$, we have the Burgers equation (also called dispersionless KdV equation), to be discussed again later on. The first Poisson bracket for the KdV hierarchy can be obtained by "freezing" the Lie–Poisson bracket at the point $((1/2)dx\otimes dx,0)$ of the dual of the Virasoro algebra. This means that instead of [16] one has to consider

$$\begin{aligned}\{F,G\}_0(u,c) &= \left\langle\left(\frac{1}{2}dx\otimes dx,0\right),\left(\left(\frac{\delta f}{\delta u}\right)'\left(\frac{\delta g}{\delta u}\right) - \left(\frac{\delta g}{\delta u}\right)'\left(\frac{\delta f}{\delta u}\right)\right)\frac{\partial}{\partial x},\int\left(\frac{\delta f}{\delta u}\right)'\left(\frac{\delta g}{\delta u}\right)''dx\right\rangle \\ &= \frac{1}{2}\int\left(\left(\frac{\delta f}{\delta u}\right)'\left(\frac{\delta g}{\delta u}\right)-\left(\frac{\delta g}{\delta u}\right)'\left(\frac{\delta f}{\delta u}\right)\right)dx\end{aligned} \quad [17]$$

The corresponding Hamiltonian is $H_2=(1/2)\int(-u^3+cu_x^2)\,dx$. From this (Lie algebraic) point of view, the compatibility between the two Poisson brackets follows from the fact that the pencil $\{\cdot,\cdot\}_\lambda=\{\cdot,\cdot\}-\lambda\{\cdot,\cdot\}_0$ is obtained from the Lie–Poisson bracket $\{\cdot,\cdot\}$ by applying the translation

$$(u\,dx\otimes\,dx,c)\mapsto\left(\left(u+\frac{\lambda}{2}\right)dx\otimes\,dx,c\right)$$

## Other Examples

In the previous section, we have presented the bi-Hamiltonian structure of the KdV equation and some of its properties. Now we give two more examples of equations – the Boussinesq equation and the Camassa–Holm equation – admitting a bi-Hamiltonian formulation. We have seen in an earlier section that the system [6] associated with the Boussinesq equation [5] is Hamiltonian with respect to the Poisson structure [7] and the Hamiltonian

$$H_1(u^1,u^2)=\int\left((u^1)^2+\tfrac{1}{9}(u^2)^3-u^1u_x^2+\tfrac{1}{3}(u_x^2)^2\right)dx$$

A more complicated Poisson structure for this system is

$$P = \begin{pmatrix} A & -3\partial_x^4 + 3u^2\partial_x^2 + 9u^1\partial_x + 3u_x^1 \\ B & -6\partial_x^3 + 6u^2\partial_x + 3u_x^2 \end{pmatrix} \quad [18]$$

with

$$A = 2\partial_x^5 - 4u^2\partial_x^3 - 6u_x^2\partial_x^2 + (2(u^2)^2 + 6u_x^1 - 6u_{xx}^2)\partial_x + (3u_{xx}^1 - 2u_{xxx}^2 + 2u^2u_x^2)$$

and

$$B = 3\partial_x^4 - 3u^2\partial_x^2 + (9u^1 - 6u_x^2)\partial_x + (6u_x^1 - 3u_{xx}^2)$$

It can be obtained by means of the Drinfeld–Sokolov reduction (or also by means of a bi-Hamiltonian reduction) from the Lie–Poisson structure (modified with the cocycle $\partial_x$) on the space of $C^\infty$ maps from $S^1$ to the Lie algebra of $3 \times 3$ traceless matrices. This is the reason why it is a Poisson structure, compatible with [7]. The system [6] can be written as

$$\begin{pmatrix} u_t^1 \\ u_t^2 \end{pmatrix} = P\begin{pmatrix} (\delta h_2/\delta u^1) \\ (\delta h_2/\delta u^2) \end{pmatrix}$$

where $h_2 = (1/3)u_1$ is the density of a Casimir of the Poisson structure [7]. Thus, the Boussinesq equation is a bi-Hamiltonian system and can be shown to possess, like KdV, an infinite sequence of conserved quantities and symmetries, forming the Boussinesq hierarchy. The KdV and the Boussinesq hierarchy are indeed particular examples of Gelfand–Dickey hierarchies (Dickey 2003). They are hierarchies of systems of $n$ equations with $n$ unknown functions and they are related, via the Drinfeld–Sokolov approach, to the Lie algebra $\mathfrak{sl}(n+1)$. As shown by Adler, Dickey, and Gelfand, these hierarchies have a bi-Hamiltonian formulation. Also the generalized KdV equations, associated by Drinfeld and Sokolov with an arbitrary affine Kac–Moody Lie algebra, are bi-Hamiltonian (or are obtained as suitable reductions of bi-Hamiltonian systems). Let us consider now the (dispersionless) Camassa–Holm equation

$$u_t - u_{txx} = -3uu_x + 2u_xu_{xx} + uu_{xxx} \quad [19]$$

which also describes shallow water waves, and possesses remarkable solutions called peakons, since they represent traveling waves with discontinuous first derivative. In order to supply this equation with a (bi-)Hamiltonian structure, one has to perform the change of variable $m = u - u_{xx}$, whose inverse, in the space of period-1 functions, turns out to be given by

$$u(x) = \int_0^x m(y)\,\sinh(y-x)\,dy + \frac{1}{2\,\sinh(1/2)}\int_0^1 m(y)\cosh\left(y - x - \frac{1}{2}\right)dy$$

The Camassa–Holm equation is then bi-Hamiltonian with respect to the Poisson pair

$$P_1 = \partial_{xxx} - \partial_x, \qquad P_2 = 2m\partial_x + m_x$$

Indeed, it can be written as $m_t = P_1\,dH_2 = P_1\,dH_2$, where

$$H_1 = -\frac{1}{2}\int (u^2 + u_x^2)\,dx$$
$$H_2 = \frac{1}{2}\int (u^3 + uu_x^2)\,dx$$

Notice that the Poisson pair of the Camassa–Holm equation can be obtained from that of KdV by moving the cocycle $\partial_{xxx}$ from the second Poisson structure to the first one. Indeed,

$$P_{(a,b,c)} = a\partial_{xxx} + b\partial_x + c(2m\partial_x + m_x) \qquad a,b,c \in \mathbb{R} \quad [20]$$

is a family of pairwise compatible Poisson operators. Moreover, we mention that Misiołek has shown that also the Camassa–Holm equation is an Euler equation on the dual of the Virasoro algebra. We conclude this article with a brief discussion concerning the so-called bi-Hamiltonian structures of hydrodynamic type. They play a relevant role in the theory of Frobenius manifolds, that, in turn, have deep relations with many important topics in contemporary mathematics and physics, such as Gromov–Witten invariants and isomonodromic deformations. As we have seen in the earlier section, a Poisson structure of hydrodynamic type is given, on the space of $C^\infty$ maps from $S^1$ to (an open set of) $\mathbb{R}^n$, by

$$\{u^{i,x}, u^{j,y}\} = g^{ij}(u)\delta'(x-y) + \Gamma_k^{ij}(u)u_x^k\delta(x-y) \quad [21]$$

where $g^{ij}(u)$ are the contravariant components of a (pseudo-)Riemannian flat metric and the $\Gamma_k^{ij}$ are the (contravariant) Christoffel symbols of the metric. If two Poisson structures of hydrodynamic type are given, it can be shown that they are compatible if and only if the two corresponding metrics form a flat pencil. This means that their linear combinations (with constant coefficients) are still flat (pseudo-)Riemannian metrics, and that the contravariant Christoffel symbols of the linear combinations are the linear combinations of the contravariant Christoffel symbols of the

two metrics. The simplest example is given by the bi-Hamiltonian formulation of the Burgers (or dispersionless KdV) equation,

$$u_t + 3uu_x = 0$$

that we have already encountered. We know that this equation is Hamiltonian with respect to the (Lie–)Poisson operator $2u\partial_x + u_x$, with Hamiltonian function $H_1 = -(1/2)\int u^2\,dx$, and with respect to the Poisson operator $\partial_x$, with Hamiltonian function $H_2 = -(1/2)\int u^3\,\mathrm{d}x$. This also means that the bi-Hamiltonian structure of the Burgers equation comes from the family [20]. The first Hamiltonian structure corresponds to the standard metric on $\mathbb{R}$, that is, $\mathrm{d}u \otimes \mathrm{d}u$, whereas the second one is given by the metric $(2u)^{-1}\,\mathrm{d}u \otimes \mathrm{d}u$.

*See also:* Classical *r*-Matrices, Lie Bialgebras, and Poisson Lie Groups; Hamiltonian Fluid Dynamics; Infinite-Dimensional Hamiltonian Systems; Integrable Systems and Recursion Operators on Symplectic and Jacobi Manifolds; Integrable Systems: Overview; Korteweg–de Vries Equation and Other Modulation Equations; Multi-Hamiltonian Systems; Recursion Operators in Classical Mechanics; Solitons and Kac–Moody Lie Algebras; Toda Lattices; WDVV Equations and Frobenius Manifolds.

## Further Reading

Arnol'd VI and Khesin BA (1998) *Topological Methods in Hydrodynamics*. New York: Springer.

Błaszak M (1998) *Multi-Hamiltonian Theory of Dynamical Systems*. Berlin: Springer.

Dickey LA (2003) *Soliton Equations and Hamiltonian Systems*, 2nd edn. River Edge: World Scientific.

Dorfman I (1993) *Dirac Structures and Integrability of Nonlinear Evolution Equations*. Chichester: Wiley.

Drinfeld VG and Sokolov VV (1985) Lie algebras and equations of Korteweg–de Vries type. *Journal of Soviet Mathematics* 30: 1975–2036.

Dubrovin BA (1996) Geometry of 2D topological field theories. In: Donagi R *et al.* (ed.) *Integrable Systems and Quantum Groups (Montecatini Terme, 1993)*, Lecture Notes in Mathematics, vol. 1620, pp. 120–348. Berlin: Springer.

Dubrovin BA, Krichever IM, and Novikov SP (2001) Integrable systems. I. In: Arnol'd VI (ed.) *Encyclopaedia of Mathematical Sciences. Dynamical Systems IV*, pp. 177–332. Berlin: Springer.

Faddeev LD and Takhtajan LA (1987) *Hamiltonian Methods in the Theory of Solitons*. Berlin: Springer.

Magri F, Falqui G, and Pedroni M (2003) The method of Poisson pairs in the theory of nonlinear PDEs. In: Conte R *et al.* (ed.) *Direct and Inverse Methods in Nonlinear Evolution Equations*, Lecture Notes in Physics, vol. 632, pp. 85–136. Berlin: Springer.

Marsden JE and Ratiu TS (1999) *Introduction to Mechanics and Symmetry*, 2nd edn. New York: Springer.

Olver PJ (1993) *Applications of Lie Groups to Differential Equations*, 2nd edn. New York: Springer.

# Billiards in Bounded Convex Domains

**S Tabachnikov**, Pennsylvania State University, University Park, PA, USA



## Billiard Flow and Billiard Ball Map

The billiard system describes the motion of a free particle inside a domain with elastic reflection off the boundary. More precisely, a billiard table is a Riemannian manifold $M$ with a piecewise smooth boundary, for example, a domain in the plane. The point moves along a geodesic line with a constant speed until it hits the boundary. At a smooth boundary point, the billiard ball reflects so that the tangential component of its velocity remains the same, while the normal component changes its sign. This means that both energy and momentum are conserved. In dimension 2, this collision is described by a well-known law of geometrical optics: the angle of incidence equals the angle of reflection. Thus, the theory of billiards has much in common with geometrical optics. If the billiard ball hits a corner, its further motion is not defined.

The billiard reflection law satisfies a variational principle. Let $A$ and $B$ be fixed points in the billiard table and let $AXB$ be a billiard trajectory from $A$ to $B$ with reflection at a boundary point $X$. Then, the position of a variable point $X$ extremizes the length $AXB$. This is the Fermat principle of geometrical optics.

In this article, we discuss billiards in bounded convex domains with smooth boundary, also called Birkhoff billiards. A related article treats billiards in polygons (*see* Polygonal Billiards).

The billiard flow is defined as a continuous-time dynamical system. The time-$t$ billiard transformation acts on unit tangent vectors to $M$ which constitute the phase space of the billiard flow, and the manifold $M$ is its configuration space. Thus, the billiard flow is the geodesic flow on a manifold with boundary.

It is useful to reduce the dimensions by one and to replace continuous time by discrete one, that is, to replace the billiard flow by a mapping, called the billiard ball map and denoted by $T$. The phase space of the billiard ball map consists of unit tangent vectors $(x, v)$ with the foot point $x$ on the boundary of $M$ and the inward direction $v$. A vector $(x, v)$ moves along the geodesic through $x$ in the direction of $v$ to the next point of its intersection $x_1$ with the boundary $\partial M$, and then $v$ reflects in $\partial M$ to the new

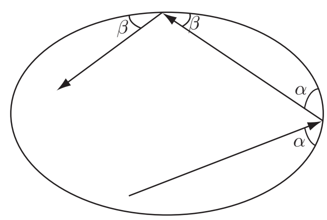



**Figure 1** Billiard ball map.

inward vector $v_1$. Then, one has: $T(x, v) = (x_1, v_1)$. For a convex $M$, the map $T$ is continuous. If $M$ is $n$-dimensional, then the dimension of the phase space of the billiard ball map is $2n - 2$.

Equivalently, and more in the spirit of geometrical optics, one considers $\mathcal{L}$, the space of oriented geodesics (rays of light) that intersect the billiard table. This space of lines is in one-to-one correspondence with the phase space of the billiard ball map: to an inward unit vector $(x, v)$ there corresponds the oriented line through $x$ in the direction $v$ (Figure 1).

The space of rays $\mathcal{L}$ carries a canonical symplectic structure, that is, a closed nondegenerate differential 2-form. In the Euclidean case, this symplectic structure $\omega$ is defined as follows. Given an oriented line $\ell$ in $\boldsymbol{R}^n$, let $q$ be the unit vector along $\ell$ and $p$ be the vector obtained by dropping the perpendicular from the origin to $\ell$. Then, $\omega = \mathrm{d}p \wedge \mathrm{d}q = \sum \mathrm{d}p_i \wedge \mathrm{d}q_i$. This construction identifies $\mathcal{L}$ with the cotangent bundle of the unit sphere: $q$ is a unit vector and $p$ is a (co)tangent vector at $q$, and $\omega$ identifies with the canonical symplectic structure of $T^*S^{n-1}$. In the general case of a Riemannian manifold $M$, the symplectic structure on the space of oriented geodesics is obtained from that on $T^*M$ by symplectic reduction.

One has an important result: the billiard ball map preserves the symplectic structure $T^*(\omega) = \omega$. As a consequence, $T$ is also measure preserving. In the planar case, one has the following explicit formula for this measure. Let $t$ be an arc length parameter along the boundary of the billiard table and let $\alpha \in [0, \pi]$ be the angle made by the unit vector with this boundary. Then, $(\alpha, t)$ are coordinates in the phase space, identified with the cylinder, and the invariant measure is $\sin\alpha \, \mathrm{d}\alpha \, \mathrm{d}t$.

As a consequence, the total area of the phase space equals $2L$ where $L$ is the perimeter length of the boundary of the billiard table, and the mean free path equals $\pi A/L$, where $A$ is the area of the billiard table. In the general $n$-dimensional case, the mean free path equals

$$\frac{\mathrm{vol}(S^{n-1})}{\mathrm{vol}(B^{n-1})} \frac{\mathrm{vol}(M)}{\mathrm{vol}(\partial M)}$$

where $S^{n-1}$ and $B^{n-1}$ are the unit sphere and the unit disk in Euclidean spaces.

## Existence and Nonexistence of Caustics

Given a plane billiard table, a caustic is a curve inside the table such that if a segment of a billiard trajectory is tangent to this curve then so is each reflected segment. Caustics correspond to invariant circles of the billiard ball map (i.e., invariant curves that go around the phase cylinder): such an invariant circle is a one-parameter family of oriented lines, and the respective caustic is their envelop. An envelop may have cusp-like singularities but if the boundary of the billiard table is a smooth curve with positive curvature then a caustic, sufficiently close to the boundary, is smooth and convex.

One can recover the table from a caustic by the following string construction. Let $\gamma$ be a caustic. Wrap a closed nonstretchable string around $\gamma$, pull it tight at a point and move this point around $\gamma$ to obtain a new curve $\Gamma$. Then, $\gamma$ is a caustic for the billiard inside $\Gamma$. Note that this construction has one parameter, the length of the string.

The following useful "mirror equation" relates various quantities depicted in Figure 2:

$$\frac{1}{a} + \frac{1}{b} = \frac{2k}{\sin\alpha}$$

where $k$ is the curvature of the boundary at the impact point.

Do caustics exist for every convex billiard table? This is important to know, in particular, because the existence of a caustic implies that the billiard ball map is not ergodic. The answer is given by a theorem of Lazutkin: *if the boundary of the billiard table is sufficiently smooth and its curvature never vanishes, then there exists a collection of smooth caustics in the vicinity of the billiard curve whose union has a positive area.* Originally this theorem asked for 553 continuous derivatives; later this was reduced to six. This result uses the techniques of the KAM (Kolmogorov–Arnol'd–Moser) theory. The

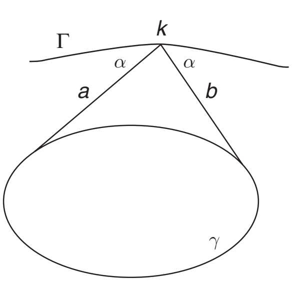



**Figure 2** String construction and mirror equation.

crucial fact is that, in appropriate coordinates, the billiard ball map is approximated, near the boundary of the phase cylinder, by the integrable map $(x, y) \mapsto (x + y, y)$.

On the other hand, by a theorem of Mather, if the curvature of a convex smooth billiard curve vanishes at some point, then this billiard ball map has no invariant circles. This result belongs to the well-developed theory of area-preserving twist maps of the cylinder, of which the billiard ball map is an example.

## Integrable Billiards

Let a plane billiard table be an ellipse with foci $F_1$ and $F_2$. It is known since antiquity that a billiard ball shot from $F_1$ reflects to $F_2$. A generalization of this optical property of the ellipse is the following theorem: *a billiard trajectory inside an ellipse forever remains tangent to a fixed confocal conic*. More precisely, if a segment of a billiard trajectory does not intersect the segment $F_1F_2$, then all the segments of this trajectory do not intersect $F_1F_2$ and are all tangent to the same ellipse with foci $F_1$ and $F_2$; and if a segment of a trajectory intersects $F_1F_2$, then all the segments of this trajectory intersect $F_1F_2$ and are all tangent to the same hyperbola with foci $F_1$ and $F_2$.

It follows that confocal ellipses are the caustics of the billiard inside an ellipse. In particular, a neighborhood of the boundary of such a billiard table is foliated by caustics. A long-standing conjecture, attributed to Birkhoff, asserts that if a neighborhood of a strictly convex smooth boundary of a billiard table is foliated by caustics, then this table is an ellipse. This conjecture remains open. The best result in this direction is a theorem of Bialy: *if almost every phase point of the billiard ball map in a strictly convex billiard table belongs to an invariant circle, then the billiard table is a disk*.

The multidimensional analogs of the optical properties of an ellipse are as follows. Consider an ellipsoid $M$ in $\boldsymbol{R}^n$ given by the equation

$$\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} + \cdots + \frac{x_n^2}{a_n^2} = 1 \qquad [1]$$

and define the confocal family of quadrics $M_\lambda$ by the equation

$$\frac{x_1^2}{a_1^2 + \lambda} + \frac{x_2^2}{a_2^2 + \lambda} + \cdots + \frac{x_n^2}{a_n^2 + \lambda} = 1$$

where $\lambda$ is a real parameter. The topological type of $M_\lambda$ changes as $\lambda$ passes the values $-a_i^2$.

One has the following theorem: *a billiard trajectory inside $M$ remains tangent to fixed $(n-1)$ confocal quadrics*. A similar and closely related result holds for the geodesic curves on $M$: the tangent lines to a fixed geodesic on $M$ are tangent to $(n-2)$ other fixed quadrics, confocal with $M$. For a triaxial ellipsoid, this theorem goes back to Jacobi.

Explicit formulas for the integrals of the billiard in an $n$-dimensional ellipsoid [1] are as follows. Let $(x, v)$ be a phase point, a unit inward tangent vector whose foot point $x$ lies on the boundary. The following functions are invariant under the billiard ball map:

$$F_i(x, v) = v_i^2 + \sum_{j \neq i} \frac{(v_i x_j - v_j x_i)^2}{a_j^2 - a_i^2}, \quad i = 1, \ldots, n$$

these functions are not independent: $F_1 + \cdots + F_n = 1$.

In fact, the integrals $F_i$ Poisson-commute (with respect to the Poisson bracket associated with the symplectic structure in the phase space of the billiard ball map that was described above). According to the Arnol'd–Liouville theorem, this complete integrability of the billiard inside an ellipsoid implies that the phase space is foliated by invariant tori and, in appropriate coordinates, the map on each torus is a parallel translation.

Similar results on complete integrability hold for billiards inside quadrics in spaces of constant positive or negative curvature. The former is the intersection of a quadratic cone with the unit sphere, and the latter with the unit pseudosphere.

## Periodic Orbits

Periodic billiard trajectories inside a planar billiard table correspond to inscribed polygons of extremal perimeter length. When counting periodic trajectories, one does not distinguish between polygons obtained from each other by cyclic permutation or reversing the order of the vertices. In other words, one counts the orbits of the dihedral group $D_n$ acting on $n$-periodic billiard polygons.

An additional topological characteristic of a periodic billiard trajectory is the rotation number defined as follows. Assume that the boundary $\gamma$ of a billiard table is parametrized by the unit circle and consider a polygon $(x_1, x_2, \ldots, x_n)$ inscribed in $\gamma$. For all $i$, one has $x_{i+1} = x_i + t_i$ with $t_i \in (0, 1)$. Since the polygon is closed, $t_1 + \cdots + t_n \in \boldsymbol{Z}$. This integer, that takes values from 1 to $n-1$, is called the rotation number of the polygon and denoted by $\rho$. Changing the orientation of a polygon replaces the

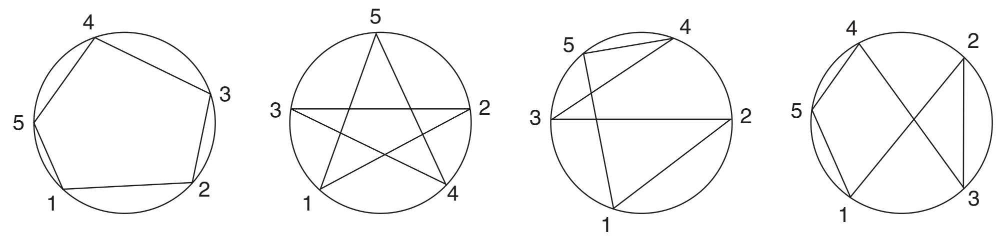



**Figure 3** Rotation numbers of periodic trajectories.

rotation number $\rho$ by $n - \rho$. The leftmost 5-periodic trajectory in Figure 3 has $\rho = 1$ and the other three $\rho = 2$.

The following theorem is due to Birkhoff: *for every* $n \geq 2$ *and* $\rho \leq \lfloor (n-1)/2 \rfloor$, *coprime with* $n$, *there exist two geometrically distinct* $n$*-periodic billiard trajectories with the rotation number* $\rho$. For example, there are at least two 2-periodic billiard trajectories inside every smooth oval: one is the diameter, the longest chord, and another one is of minimax type, similar to the minor axis of an ellipse.

In higher dimensions, lower bounds on the number of periodic billiard trajectories inside strictly convex domains with smooth boundaries were obtained only recently by Farber and the present author. Here is one of the results: for a generic billiard table in $\boldsymbol{R}^m$, the number of $n$-periodic trajectories is not less than $(n-1)(m-1)$. The proof consists in using the Morse theory to estimate below the number of critical points of the perimeter length function on the space of inscribed $n$-gons and its quotient space by the dihedral group $D_n$, and the main difficulty is in describing the topology of these spaces.

Returning to convex smooth planar billiards, the following conjecture remains open for a long time: the set of $n$-periodic points of the billiard ball map has zero measure. This is easy for $n = 2$; for $n = 3$ this is a theorem by M Rychlik. The motivation for this question comes from spectral geometry. In particular, according to a theorem of Ivrii, the above conjecture implies the Weyl conjecture on the second term for the spectral asymptotics of the Laplacian in a bounded domain with the Dirichlet or Neumann boundary conditions.

## Length Spectrum

The set of lengths of the closed trajectories in a convex billiard $M$ is called the length spectrum of $M$. There is a remarkable relation between the length spectrum and the spectrum of the Laplace operator in $M$ with the Dirichlet boundary condition: $\Delta f = \lambda f$, $f|_{\partial M} = 0$. From the physical point of view, the eigenvalues $\lambda$ are the eigenfrequencies of the membrane $M$ with a fixed boundary. Roughly speaking, one can recover the length spectrum from that of the Laplacian. More precisely, the following theorem of K Anderson and R Melrose holds:

$$\sum_{\lambda_i \in \text{spec } \Delta} \cos\left(t\sqrt{-\lambda_i}\right)$$

*is a well-defined generalized function (distribution) of* $t$, *smooth away from the length spectrum*. That is, if $l > 0$ belongs to the singular support of this distribution, then there exists either a closed billiard trajectory of length $l$, or a closed geodesic of length $l$ in the boundary of the billiard table.

This relation between the Laplacian and the length spectrum is due to the fact that geometric optics is not a very accurate description of light. In wave optics, light is considered as electromagnetic waves, and geometric optics gives a realistic approximation only when the wave length is small. This small-wave approximation is based on the assumption that the waves are locally almost harmonic, while their amplitudes change slowly from point to point. The substitution of such a function into the corresponding PDEs gives, in the first approximation, the equations of wave fronts, that is, of geometric optics.

Here is another spectral result concerning a smooth strictly convex plane domain, due to S Marvizi and R Melrose. Let $L_n$ be the supremum and $l_n$ the infimum of the perimeters of simple billiard $n$-gons. Then,

$$\lim_{n \to \infty} n^k (L_n - l_n) = 0$$

for any positive $k$. Furthermore, $L_n$ has an asymptotic expansion, as $n \to \infty$,

$$L_n \sim l + \sum_{i=1}^{\infty} \frac{c_i}{n^{2i}}$$

where $l$ is the length of the boundary of billiard table and $c_i$ are constants, depending on the curvature of the boundary.

## Acknowledgments

This work was partially supported by NSF.

*See also:* Adiabatic Piston; Hamiltonian Systems: Obstructions to Integrability; Hyperbolic Billiards; Integrable Discrete Systems; Integrable Systems and Algebraic Geometry; Optical Caustics; Integrable Systems: Overview; Polygonal Billiards; Semiclassical Spectra and Closed Orbits; Separatrix Splitting; Stability Theory and KAM.

## Further Reading

Chernov N and Markarian R, *Theory of Chaotic Billiards* (to appear).

Farber M and Tabachnikov S (2002) Topology of cyclic configuration spaces and periodic orbits of multi-dimensional billiards. *Topology* 41: 553–589.

Gutkin E (2003) Billiard dynamics: a survey with the emphasis on open problems. *Regular and Chaotic Dynamics* 8: 1–13.

Katok A and Hasselblatt B (1995) *Introduction to the Modern Theory of Dynamical Systems*. Cambridge: Cambridge University Press.

Kozlov V and Treshchev D (1991) *Billiards. A Genetic Introduction to the Dynamics of Systems with Impacts*. Providence: American Mathematical Society.

Lazutkin V (1993) *KAM Theory and Semiclassical Approximations to Eigenfunctions*. Berlin: Springer.

Moser J (1980) *Various Aspects of Integrable Hamiltonian Systems*. Progress in Mathematics, vol. 8, pp. 233–289. Basel: Birhauser.

Siburg KF (2004) *The Principle of Last Action in Geometry and Dynamics*, Lecture Notes in Mathematics, vol. 1844. Berlin: Springer.

Sinai Ya (1976) *Introduction to Ergodic Theory*. Princeton: Princeton University Press.

Tabachnikov S (1995) *Billiards*, Société Math. de France, Panoramas et Syntheses, No 1.

Tabachnikov S (2005) *Geometry and Billiards*. American Mathematic Society (to appear).

# Black Hole Mechanics

**A Ashtekar**, Pennsylvania State University, University Park, PA, USA



## Introduction

Over the last 30 years, black holes have been shown to have a number of surprising properties. These discoveries have revealed unforeseen relations between the otherwise distinct areas of general relativity, quantum physics, and statistical mechanics. This interplay, in turn, led to a number of deep puzzles at the very foundations of physics. Some have been resolved while others continue to baffle physicists. The starting point of these fascinating developments was the discovery of laws of black hole mechanics by Bardeen, Bekenstein, Carter, and Hawking. They dictate the behavior of black holes in equilibrium, under small perturbations away from equilibrium, and in fully dynamical situations. While they are consequences of classical general relativity alone, they have a close similarity with the laws of thermodynamics. The origin of this seemingly strange coincidence lies in quantum physics. For further discussion, *see* Asymptotic Structure and Conformal Infinity; Loop Quantum Gravity; Quantum Geometry and Its Applications; Quantum Field Theory in Curved Spacetime; Stationary Black Holes.

The focus of this article is just on black hole mechanics. The discussion is divided into three parts. In the first, we will introduce the notions of event horizons and black hole regions and discuss properties of globally stationary black holes. In the second, we will consider black holes which are themselves in equilibrium but in surroundings which may be time dependent. Finally, in the third part, we summarize what is known in the fully dynamical situations. For simplicity, all manifolds and fields are assumed to be smooth and, unless otherwise stated, spacetime is assumed to be four dimensional, with a metric of signature $-, +, +, +$, and the cosmological constant is assumed to be zero. An arrow under a spacetime index denotes the pullback of that index to the horizon.

## Global Equilibrium

To capture the intuitive notion that black hole is a region from which signals cannot escape to the asymptotic part of spacetime, one needs a precise definition of future infinity. The standard strategy is to use Penrose's conformal boundary $\mathcal{J}^+$. A black hole region $\mathcal{B}$ of a spacetime $(\mathbb{M}, g_{ab})$ is defined as $\mathcal{B} = \mathbb{M} \setminus I^-(\mathcal{J}^+)$, where $I^-$ denotes "chronological past." The boundary $\partial\mathcal{B}$ of the black hole region is called the "event horizon" and denoted by $\mathcal{E}$. Thus, $\mathcal{E}$ is the boundary of the past of $\mathcal{J}^+$. It therefore follows that $\mathcal{E}$ is a null 3-surface, ruled by future inextendible null geodesics without caustics. If the spacetime is globally hyperbolic, an "instant of time" is represented by a Cauchy surface $M$. The intersection of $\mathcal{B}$ with $M$ may have several disjoint components, each representing a black hole at that instant of time. If $M'$ is a Cauchy surface to the future of $M$, the number of disjoint components of $M' \cup \mathcal{B}$ in the causal future of $M \cup \mathcal{B}$ must be less than or equal to those of $M \cup \mathcal{B}$

(see Hawking and Ellis (1973)). Thus, black holes can merge but can not bifurcate. (By a time reversal, i.e., by replacing $\mathcal{J}^+$ with $\mathcal{J}^-$ and $I^-$ with $I^+$, one can define a white hole region $\mathcal{W}$. However, here we will focus only on black holes.)

A spacetime $(\mathbb{M}, g_{ab})$ is said to be stationary (i.e., time independent) if $g_{ab}$ admits a Killing field $t^a$ that represents an asymptotic time translation. By convention, $t^a$ is assumed to be unit at infinity. $(\mathbb{M}, g_{ab})$ is said to be axisymmetric if $g_{ab}$ admits a Killing field $\phi^a$ generating an SO(2) isometry. By convention $\phi^a$ is normalized such that the affine length of its integral curves is $2\pi$. Stationary spacetimes with nontrivial $\mathbb{M}\setminus I^-(\mathcal{J}^+)$ represent black holes which are in global equilibrium. In the Einstein–Maxwell theory in four dimensions, there exists a unique three-parameter family of stationary black hole solutions, generally parametrized by mass $m$, angular momentum $J$, and electric charge $Q$. This is the celebrated Kerr–Newman family. Therefore, in general relativity a great deal of work on black holes has focused on these solutions and perturbations thereof. The Kerr–Newman family is axisymmetric and furthermore, its metric has the property that the 2-flats spanned by the Killing fields $t^a$ and $\phi^a$ are orthogonal to a family of 2-surfaces. This property is called "$t$–$\phi$ orthogonality." These features of Kerr–Newman space-times are widely used in black hole physics. Note however that uniqueness fails in higher dimensions, and also in the presence of nonabelian gauge fields or rings of perfect fluids around black holes in four dimensions. In mathematical physics, there is significant literature on the new stationary black hole solutions in Einstein–Yang–Mills–Higgs theories. These are called "hairy black holes." Research on stationary black hole solutions with rings received a boost by a recent discovery that these black holes can violate the Kerr inequality $J \leq Gm^2$ between angular momentum $J$ and mass $m$.

A null 3-manifold $\mathcal{K}$ in $\mathbb{M}$ is said to be a "Killing horizon" if $g_{ab}$ admits a Killing field $K^a$ which is everywhere normal to $\mathcal{K}$. On a Killing horizon, one can show that the acceleration of $K^a$ is proportional to $K^a$ itself:

$$K^a \nabla_a K^b = \kappa K^b \qquad [1]$$

The proportionality function $\kappa$ is called "surface gravity." We will show in the next section that if a mild energy condition holds on $\mathcal{K}$, then $\kappa$ must be constant. Note that if we rescale $K^a$ via $K^a \to cK^a$, where $c$ is a constant, surface gravity also rescales as $\kappa \to c\kappa$.

In the Kerr–Newman family, the event horizon is a Killing horizon. More generally, if an axisymmetric, stationary black hole spacetime $(\mathbb{M}, g_{ab})$ satisfies the $t$–$\phi$ orthogonality property, its event horizon $\mathcal{E}$ is a Killing horizon. (Although one can envisage stationary black holes in which these additional symmetry conditions are not met, this possibility has been ignored in black hole mechanics on stationary spacetimes. Quasilocal horizons, discussed below, do not require any spacetime symmetries.) In these cases, the normalization freedom in $K^a$ is fixed by requiring that $K^a$ have the form

$$K^a = t^a + \Omega \phi^a \qquad [2]$$

on the horizon, where $\Omega$ is a constant, called the "angular velocity of the horizon." The resulting $\kappa$ is called the surface gravity of the black hole. It is remarkable that $\kappa$ is constant for all such black holes, even when their horizon is highly distorted (i.e., far from being spherically symmetric) either due to rotation or due to external matter fields. This is analogous to the fact that the temperature of a thermodynamical system in equilibrium is constant, independently of the details of the system. In analogy with thermodynamics, constancy of $\kappa$ is referred to as the "zeroth law of black hole mechanics."

Next, let us consider an infinitesimal perturbation $\delta$ within the three-parameter Kerr–Newman family. A simple calculation shows that the changes in the Arnowitt–Deser–Misner (ADM) mass $m$, angular momentum $J$, and the total charge $Q$ of the spacetime and in the area $a$ of the horizon are constrained via

$$\delta m = \frac{\kappa}{8\pi G}\,\delta a + \Omega\,\delta J + \Phi \delta Q \qquad [3]$$

where the coefficients $\kappa, \Omega, \Phi$ are black hole parameters, $\Phi = A_a K^a$ being the electrostatic potential at the horizon. The last two terms, $\Omega\delta J$ and $\Phi\delta Q$, have the interpretation of "work" required to spin the black hole up by an amount $\delta J$ or to increase its charge by $\delta Q$. Therefore, [3] has a striking resemblance to the first law, $\delta E = T\delta S + \delta W$, of thermodynamics if (as the zeroth law suggests) $\kappa$ is made proportional to the temperature $T$, and the horizon area $a$ to the entropy $S$. Therefore, [3] and its generalizations discussed below are referred to as the "first law of black hole mechanics."

In Kerr–Newman spacetimes, the only contribution to the stress–energy tensor comes from the Maxwell field. Bardeen *et al.* (1973) consider stationary black holes with matter such as perfect fluids in the exterior region and stationary perturbations $\delta$ thereof. Using Einstein's equations, they show that the form [3] of the first law does not change; the only modification is addition of certain matter terms on the right-hand side which can be

interpreted as the work $\delta W$ done on the total system. A generalization in another direction was made by Iyer and Wald (1994) using Noether currents. They allow nonstationary perturbations and, more importantly, drop the restriction to general relativity. Instead, they consider a wide class of diffeomorphism-invariant Lagrangian densities $L(g_{ab}, R_{abcd}, \nabla_a R_{bcde}, \ldots, \Phi^{\cdot\cdot}{}_{..}, \nabla_a \Phi^{\cdot\cdot}{}_{..}, \ldots)$ which depend on the metric $g_{ab}$, matter fields $\Phi^{\cdot\cdot}{}_{..}$, and a finite number of derivatives of the Riemann tensor and matter fields. Finally, they restrict themselves to $\kappa \neq 0$. In this case, on the maximal analytic extension of the spacetime, the Killing field $K^a$ vanishes on a 2-sphere $S_o$ called the bifurcate horizon. Then, [3] is generalized to

$$\delta m = \frac{\kappa}{2\pi} \delta S_{\text{hor}} + \delta W \qquad [4]$$

Here $\delta W$ again represents "work terms" and $S_{\text{hor}}$ is given by

$$S_{\text{hor}} = -2\pi \oint_{S_o} \frac{\delta L}{\delta R_{abcd}} n_{ab} n_{cd} \qquad [5]$$

where $n_{ab}$ is the binormal to $S_o$ (with $n_{ab}n^{ab} = -2$), and the functional derivative inside the integral is evaluated by formally viewing the Riemann tensor as a field independent of the metric. For the Einstein–Hilbert action, this yields $S_{\text{hor}} = a/4G$ and one recovers [3].

These results are striking. However, the underlying assumptions have certain unsatisfactory aspects. First, although the laws are meant to refer just to black holes, one assumes that the entire spacetime is stationary. In thermodynamics, by contrast, one only assumes that the system under consideration is in equilibrium, not the whole universe. Second, in the first law, quantities $a, \Omega, \Phi$ are evaluated at the horizon while $M$, $J$ are evaluated at infinity and include contributions from possible matter fields outside the black hole. A more satisfactory law of black hole mechanics would involve attributes of the black hole alone. Finally, the notion of the event horizon is extremely global and teleological since it explicitly refers to $\mathcal{J}^+$. An event horizon may well be developing in the very room you are sitting today in anticipation of a gravitational collapse in the center of our galaxy which may occur a billion years hence. This feature makes it impossible to generalize the first law to fully dynamical situations and relate the change in the event horizon area to the flux of energy and angular momentum falling across it. Indeed, one can construct explicit examples of dynamical black holes in which an event horizon $\mathcal{E}$ forms and grows in the flat part of a spacetime where nothing happens physically. These considerations call for a replacement of $\mathcal{E}$ by a quasilocal horizon which leads to a first law involving only horizon attributes, and which can grow only in response to the influx of energy. Such horizons are discussed in the next two sections.

## Local Equilibrium

The key idea here is drop the requirement that spacetime should admit a stationary Killing field and ask only that the intrinsic horizon geometry be time independent. Consider a null 3-surface $\Delta$ in a spacetime $(\mathbb{M}, g_{ab})$ with a future-pointing normal field $\ell^a$. The pullback $q_{ab} := g_{\underleftarrow{ab}}$ of the spacetime metric to $\Delta$ is the intrinsic, degenerate "metric" of $\Delta$ with signature 0, +, +. The first condition is that it be "time independent," that is, $\mathcal{L}_\ell q_{ab} = 0$ on $\Delta$. Then by restriction, the spacetime derivative operator $\nabla$ induces a natural derivative operator $D$ on $\Delta$. While $D$ is compatible with $q_{ab}$, that is, $D_a q_{bc} = 0$, it is not uniquely determined by this property because $q_{ab}$ is degenerate. Thus, $D$ has extra information, not contained in $q_{ab}$. The pair $(q_{ab}, D)$ is said to determine the intrinsic geometry of the null surface $\Delta$. This notion leads to a natural definition of a horizon in local equilibrium. Let $\Delta$ be a null, three-dimensional submanifold of $(\mathbb{M}, g_{ab})$ with topology $\mathbb{S} \times \mathbb{R}$, where $\mathbb{S}$ is compact and without boundary.

**Definition 1** $\Delta$ is said to be "isolated horizon" if it admits a null normal $\ell^a$ such that:

(i) $\mathcal{L}_\ell \, q_{ab} = 0$ and $[\mathcal{L}_\ell, D] = 0$ on $\Delta$ and
(ii) $-T^a{}_b \ell^b$ is a future pointing causal vector on $\Delta$.

On can show that, generically, this null normal field $\ell^a$ is unique up to rescalings by positive constants.

Both conditions are local to $\Delta$. In particular, $(\mathbb{M}, g_{ab})$ is not required to be asymptotically flat and there is no longer any teleological feature. Since $\Delta$ is null and $\mathcal{L}_\ell q_{ab} = 0$, the area of any of its cross sections is the same, denoted by $a_\Delta$. As one would expect, one can show that there is no flux of gravitational radiation or matter across $\Delta$. This captures the idea that the black hole itself is in equilibrium. Condition (ii) is a rather weak "energy condition" which is satisfied by all matter fields normally considered in classical general relativity. The nontrivial condition is (i). It extracts from the notion of a Killing horizon just a "tiny part" that refers only to the intrinsic geometry of $\Delta$. As a result, every Killing horizon $\mathcal{K}$ is, in particular, an isolated horizon. However, a spacetime with an isolated horizon $\Delta$ can admit gravitational radiation and dynamical matter fields away from $\Delta$. In fact, as a family of Robinson–Trautman spacetimes illustrates,

gravitational radiation could even be present arbitrarily close to $\Delta$. Because of these possibilities, there are many nontrivial examples and the transition from event horizons of stationary spacetimes to isolated horizons represents a significant generalization of black hole mechanics. (In fact, the derivation of the zeroth and the first law requires slightly weaker assumptions, encoded in the notion of a "weakly isolated horizon" (Ashtekar *et al.* 2000, 2001).)

An immediate consequence of the requirement $\mathcal{L}_\ell q_{ab} = 0$ is that there exists a 1-form $\omega_a$ on $\Delta$ such that $D_a \ell^b = \omega_a \ell^b$. Following the definition of $\kappa$ on a Killing horizon, the surface gravity $\kappa_{(\ell)}$ of $(\Delta, \ell)$ is defined as $\kappa_{(\ell)} = \omega_a \ell^a$. Again, under $\ell^a \to c\ell^a$, we have $\kappa_{(c\ell)} = c\kappa_\ell$. Together with Einstein's equations, the two conditions of Definition 1 imply $\mathcal{L}_\ell \omega_a = 0$ and $\ell^a D_{[a}\omega_{b]} = 0$. The Cartan identity relating the Lie and exterior derivative now yields

$$D_a(\omega_b \ell^b) \equiv D_a \kappa_{(\ell)} = 0 \qquad [6]$$

Thus, surface gravity is constant on every isolated horizon. This is the zeroth law, extended to horizons representing local equilibrium. In the presence of an electromagnetic field, Definition 1 and the field equations imply $\mathcal{L}_\ell \underline{F_{ab}} = 0$ and $\ell^a \underline{F_{ab}} = 0$. The first of these equations implies that one can always choose a gauge in which $\mathcal{L}_\ell \underline{A_a} = 0$. By Cartan identity it then follows that the electrostatic potential $\Phi_{(\ell)} := A_a \ell^a$ is constant on the horizon. This is the Maxwell analog of the zeroth law.

In this setting, the first law is derived using a Hamiltonian framework (Ashtekar *et al.* 2000, 2001). For concreteness, let us assume that we are in the asymptotically flat situation and the only gauge field present is electromagnetic. One begins by restricting oneself to horizon geometries such that $\Delta$ admits a rotational vector field $\varphi^a$ satisfying $\mathcal{L}_\varphi q_{ab} = 0$. (In fact for black hole mechanics, it suffices to assume only that $\mathcal{L}_\varphi \epsilon_{ab} = 0$, where $\epsilon_{ab}$ is the intrinsic area 2-form on $\Delta$. The same is true on dynamical horizons discussed in the next section.) One then constructs a phase space $\boldsymbol{\Gamma}$ of gravitational and matter fields such that (1) $\mathbb{M}$ admits an internal boundary $\Delta$ which is an isolated horizon; and (2) all fields satisfy asymptotically flat boundary conditions at infinity. Note that the horizon geometry is allowed to vary from one phase-space point to another; the pair $(q_{ab}, D)$ induced on $\Delta$ by the spacetime metric only has to satisfy Definition 1 and the condition $\mathcal{L}_\varphi q_{ab} = 0$.

Let us begin with angular momentum. Fix a vector field $\phi^a$ on $\mathbb{M}$ which coincides with the fixed $\varphi^a$ on $\Delta$ and is an asymptotic rotational symmetry at infinity. (Note that $\phi^a$ is not restricted in any way in the bulk.) Lie derivatives of gravitational and matter fields along $\phi^a$ define a vector field $X(\phi)$ on $\boldsymbol{\Gamma}$. One shows that it is an infinitesimal canonical transformation, that is, satisfies $\mathcal{L}_{X(\phi)}\boldsymbol{\Omega} = 0$, where $\boldsymbol{\Omega}$ is the symplectic structure on $\boldsymbol{\Gamma}$. The Hamiltonian $H(\phi)$ generating this canonical transformation is given by

$$\begin{aligned} H(\phi) &= J_\Delta^{(\phi)} - J_\infty^{(\phi)} \\ J_\Delta^{(\phi)} &= -\frac{1}{8\pi G}\oint_{\mathbb{S}} (\omega_a \phi^a)\epsilon - \frac{1}{4\pi}\oint_{\mathbb{S}} (A_a \phi^a)^\star F \end{aligned} \qquad [7]$$

where $J_\infty^{(\phi)}$ is the ADM angular momentum at infinity, $\mathbb{S}$ is any cross section of $\Delta$, and $\epsilon$ the area element thereon. The term $J_\Delta^{(\phi)}$ is independent of the choice of $\mathbb{S}$ made in its evaluation and interpreted as the "horizon angular momentum." It has numerous properties that support this interpretation. In particular, it yields the standard angular momentum expression in Kerr–Newman spacetimes.

To define horizon energy, one has to introduce a "time-translation" vector field $t^a$. At infinity, $t^a$ must tend to a unit time translation. On $\Delta$, it must be a symmetry of $q_{ab}$. Since $\ell^a$ and $\varphi^a$ are both horizon symmetries, $t^a = c\ell^a + \Omega\varphi^a$ on $\Delta$, for some constants $c$ and $\Omega$. However, unlike $\phi^a$, the restriction of $t^a$ to $\Delta$ cannot be fixed once and for all but must be allowed to vary from one phase-space point to another. In particular, on physical grounds, one expects $\Omega$ to be zero at a phase-space point representing a nonrotating black hole but nonzero at a point representing a rotating black hole. This freedom in the boundary value of $t^a$ introduces a qualitatively new element. The vector field $X(t)$ on $\boldsymbol{\Gamma}$ defined by the Lie derivatives of gravitational and matter fields does not, in general, satisfy $\mathcal{L}_{X(t)}\boldsymbol{\Omega} = 0$; it need not be an infinitesimal canonical transformation. The necessary and sufficient condition is that $(\kappa_{(c\ell)}/8\pi G)\delta a_\Delta + \Omega\delta J_\Delta + \Phi_{(c\ell)}\delta Q_\Delta$ be an exact variation. That is, $X(t)$ generates a Hamiltonian flow if and only if there exists a function $E_\Delta^{(t)}$ on $\boldsymbol{\Gamma}$ such that

$$\delta E_\Delta^{(t)} = \frac{\kappa_{(c\ell)}}{8\pi G}\delta a_\Delta + \Omega\delta J_\Delta + \Phi_{(c\ell)}\delta Q_\Delta \qquad [8]$$

This is precisely the first law. Thus, the framework provides a deeper insight into the origin of the first law: it is the necessary and sufficient condition for the evolution generated by $t^a$ to be Hamiltonian. Equation [8] is a genuine restriction on the choice of phase-space functions $c$ and $\Omega$, that is, of restrictions to $\Delta$ of evolution fields $t^a$. It is easy to verify that $\mathbb{M}$ admits many such vector fields. Given one, the Hamiltonian $H(t)$ generating the time evolution along $t^a$ takes the form

$$H(t) = E_\infty^{(t)} - E_\Delta^{(t)} \qquad [9]$$

re-enforcing the interpretation of $E_{\Delta}^{(t)}$ as the horizon energy.

In general, there is a multitude of first laws, one for each vector field $t^a$, the evolution along which preserves the symplectic structure. In the Einstein–Maxwell theory, given any phase-space point, one can choose a canonical boundary value $t_o^a$ exploiting the uniqueness theorem. $E_{\Delta}^{(t_o)}$ is then called the horizon mass and denoted simply by $m_{\Delta}$. In the Kerr–Newman family, $\boldsymbol{H}(t_o)$ vanishes and $m_{\Delta}$ coincides with the ADM mass $m_{\infty}$. Similarly, if $\phi^a$ is chosen to be a global rotational Killing field, $J_{\Delta}^{(\phi)}$ equals $J_{\infty}^{(\phi)}$. However, in more general spacetimes where there is matter field or gravitational radiation outside $\Delta$, these equalities do not hold; $m_{\Delta}$ and $J_{\Delta}$ represent quantities associated with the horizon alone while the ADM quantities represent the total mass and angular momentum in the spacetime, including contributions from matter fields and gravitational radiation in the exterior region. In the first law [8], only the contributions associated with the horizon appear.

When the uniqueness theorem fails, as, for example, in the Einstein–Yang–Mills–Higgs theory, first laws continue to hold but the horizon mass $m_{\Delta}$ becomes ambiguous. Interestingly, these ambiguities can be exploited to relate properties of hairy black holes with those of the corresponding solitons. (For a summary, see Ashtekar and Krishnan (2004).)

## Dynamical Situations

A natural question now is whether there is an analog of the second law of thermodynamics. Using event horizons, Hawking showed that the answer is in the affirmative (see Hawking and Ellis (1973)). Let $(\mathbb{M}, g_{ab})$ admit an event horizon $\mathcal{E}$. Denote by $\ell^a$ a geodesic null normal to $\mathcal{E}$. Its expansion is defined as $\theta_{(\ell)} := q^{ab}\nabla_a \ell_b$, where $q^{ab}$ is any inverse of the degenerate intrinsic metric $q_{ab}$ on $\mathcal{E}$, and determines the rate of change of the area element of $\mathcal{E}$ along $\ell^a$. Assuming that the null energy condition and Einstein's equations hold, the Raychaudhuri equation immediately implies that if $\theta_{(\ell)}$ were to become negative somewhere it would become infinite within a finite affine parameter. Hawking showed that, if there is a globally hyperbolic region containing $I^-(\mathcal{J}^+) \cup \mathcal{E}$ – that is, if there are no naked singularities – this can not happen, whence $\theta(\ell) \geq 0$ on $\mathcal{E}$. Hence, if a cross section $S_2$ of $\mathcal{E}$ is to the future of a cross section $S_1$, we must have $a_{S_2} \geq a_{S_1}$. Thus, in any (i.e., not necessarily infinitesimal) dynamical process, the change $\Delta a$ in the horizon area is always non-negative. This result is known as the "second law of black hole mechanics." As in the first law, the analog of entropy is the horizon area.

It is tempting to ask if there is a local physical process directly responsible for the growth of area. For event horizons, the answer is in the negative since they can grow in a flat portion of spacetime. However, one can introduce quasilocal horizons also in the dynamical situations and obtain the desired result (Ashtekar and Krishnan 2003). These constructions are strongly motivated by earlier ideas introduced by Hayward (1994).

**Definition 2** A three-dimensional spacelike submanifold $\mathcal{H}$ of $(\mathbb{M}, g_{ab})$ is said to be a "dynamical horizon" if it admits a foliation by compact 2-manifolds $\mathbb{S}$ (without boundary) such that:

(i) the expansion $\theta_{(\ell)}$ of one (future directed) null normal field $\ell^a$ to $\mathbb{S}$ vanishes and the expansion of the other (future directed) null normal field, $n^a$ is negative; and
(ii) $-T^a{}_b \ell^b$ is a future pointing causal vector on $\mathcal{H}$.

One can show that this foliation of $\mathcal{H}$ is unique and that $\mathbb{S}$ is either a 2-sphere or, under degenerate and physically over-restrictive conditions, a 2-torus. Each leaf $\mathbb{S}$ is a marginally trapped surface and referred to as a "cut" of $\mathcal{H}$. Unlike event horizons $\mathcal{E}$, dynamical horizons $\mathcal{H}$ are locally defined and do not display any teleological feature. In particular, they cannot lie in a flat portion of spacetime. Dynamical horizons commonly arise in numerical simulations of evolving black holes as world tubes of apparent horizons. As the black hole settles down, $\mathcal{H}$ asymptotes to an isolated horizon $\Delta$, which tightly hugs the asymptotic future portion of the event horizon. However, during the dynamical phase, $\mathcal{H}$ typically lies well inside $\mathcal{E}$.

The two conditions in Definition 2 immediately imply that the area of cuts of $\mathcal{H}$ increases monotonically along the "outward direction" defined by the projection of $\ell^a$ on $\mathcal{H}$. Furthermore, this change turns out to be directly related to the flux of energy falling across $\mathcal{H}$. Let $R$ denote the "radius function" on $\mathcal{H}$ so that the area of any cut $\mathbb{S}$ is given by $a_{\mathbb{S}} = 4\pi R^2$. Let $N$ denote the norm of $\partial_a R$ and $\Delta\mathcal{H}$, the portion of $\mathcal{H}$ bounded by two cross sections $\mathbb{S}_1$ and $\mathbb{S}_2$. The appropriate energy turns out to be associated with the vector field $N\ell^a$, where $\ell^a$ is normalized such that its projection on $\mathcal{H}$ is the unit normal $\hat{r}^a$ to the cuts $\mathbb{S}$. In the generic and physically interesting case when $\mathbb{S}$ is a 2-sphere, the Gauss and the Codazzi (i.e., constraint) equations imply

$$\frac{1}{2G}(R_2 - R_1) = \int_{\Delta\mathcal{H}} T_{ab} N \ell^a \hat{r}^b \, \mathrm{d}^3V + \frac{1}{16\pi G} \times \int_{\Delta\mathcal{H}} N\left(\sigma_{ab}\sigma^{ab} + 2\zeta_a\zeta^a\right) \mathrm{d}^3V \qquad [10]$$

Here $\hat{r}^a$ is the unit normal to $\mathcal{H}$, $\sigma^{ab}$ the shear of $\ell^a$ (i.e., the tracefree part of $q^{am}q^{bm}\nabla_m\ell_n$), and $\zeta^a = q^{ab}\hat{r}^c\nabla_c\ell_b$, where $q^{ab}$ is the projector onto the tangent space of the cuts $\mathbb{S}$. The first integral on the right-hand side can be directly interpreted as the flux across $\Delta\mathcal{H}$ of matter–energy (relative to the vector field $N\ell^a$). The second term is purely geometric and is interpreted as the flux of energy carried by gravitational waves across $\Delta\mathcal{H}$. It has several properties which support this interpretation. Thus, not only does the second law of black hole mechanics hold for a dynamical horizon $\mathcal{H}$, but the "cause" of the increase in the area can be directly traced to physical processes happening near $\mathcal{H}$.

Another natural question is whether the first law [8] can be generalized to fully dynamical situations, where $\delta$ is replaced by a finite transition. Again, the answer is in the affirmative. We will outline the idea for the case when there are no gauge fields on $\mathcal{H}$. As with isolated horizons, to have a well-defined notion of angular momentum, let us suppose that the intrinsic 3-metric on $\mathcal{H}$ admits a rotational Killing field $\varphi$. Then, the angular momentum associated with any cut $\mathbb{S}$ is given by

$$J_{\mathbb{S}}^{(\varphi)} = -\frac{1}{8\pi G}\oint_{\mathbb{S}} K_{ab}\varphi^a\hat{r}^b\, \mathrm{d}^2V \equiv \frac{1}{8\pi G}\oint_{\mathbb{S}} j^{(\varphi)}\, \mathrm{d}^2V \qquad [11]$$

where $K_{ab}$ is the extrinsic curvature of $\mathcal{H}$ in $(\mathbb{M}, g_{ab})$ and $j^{(\varphi)}$ is interpreted as "the angular momentum density." Now, in the Kerr family, the mass, surface gravity, and the angular velocity can be unambiguously expressed as well-defined functions $\bar{m}(a,J)$, $\bar{\kappa}(a,J)$, and $\bar{\Omega}(a,J)$ of the horizon area $a$ and angular momentum $J$. The idea is to use these expressions to associate mass, surface gravity, and angular velocity with each cut of $\mathcal{H}$. Then, a surprising result is that the difference between the horizon masses associated with cuts $\mathbb{S}_1$ and $\mathbb{S}_2$ can be expressed as the integral of a locally defined flux across the portion $\Delta\mathcal{H}$ of $\mathcal{H}$ bounded by $\mathcal{H}_1$ and $\mathcal{H}_2$:

$$\bar{m}_2 - \bar{m}_1 = \frac{1}{8\pi G}\int_{\Delta\mathcal{H}} \bar{\kappa}\, \mathrm{d}a + \frac{1}{8\pi G}\left\{ \oint_{\mathbb{S}_2} \bar{\Omega} j^{\varphi}\, \mathrm{d}^2V - \oint_{\mathbb{S}_1} \bar{\Omega} j^{\varphi}\, \mathrm{d}^2V - \int_{\bar{\Omega}_1}^{\bar{\Omega}_2} \mathrm{d}\bar{\Omega} \oint_{\mathbb{S}} j^{\varphi}\, \mathrm{d}^2V \right\} \qquad [12]$$

If the cuts $\mathbb{S}_2$ and $\mathbb{S}_1$ are only infinitesimally separated, this expression reduces precisely to the standard first law involving infinitesimal variations. Therefore, [12] is an integral generalization of the first law.

Let us conclude with a general perspective. On the whole, in the passage from event horizons in stationary spacetimes to isolated horizons and then to dynamical horizons, one considers increasingly more realistic situations. In all the three cases, the analysis has been extended to allow the presence of a cosmological constant $\Lambda$. (The only significant change is that the topology of cuts $\mathbb{S}$ of dynamical horizons is restricted to be $\mathbb{S}^2$ if $\Lambda > 0$ and is completely unrestricted if $\Lambda < 0$.) In the first two frameworks, results have also been extended to higher dimensions. Since the notions of isolated and dynamical horizons make no reference to infinity, these frameworks can be used also in spatially compact spacetimes. The notion of an event horizon, by contrast, does not naturally extend to these spacetimes. On the other hand, the generalization [4] of the first law [3] is applicable to event horizons of stationary spacetimes in a wide class of theories while so far the isolated and dynamical horizon frameworks are tied to general relativity (coupled to matter satisfying rather weak energy conditions). From a mathematical physics perspective, extension to more general theories is an important open problem.

*See also:* Asymptotic Structure and Conformal Infinity; Branes and Black Hole Statistical Mechanics; Dirac Fields in Gravitation and Nonabelian Gauge Theory; Geometric Flows and the Penrose Inequality; Loop Quantum Gravity; Minimal Submanifolds; Quantum Field Theory in Curved Spacetime; Quantum Geometry and its Applications; Random Algebraic Geometry, Attractors and Flux Vacua; Shock Wave Refinement of the Friedman–Robertson–Walker Metric; Stationary Black Holes.

## Further Reading

Ashtekar A, Beetle C, and Lewandowski J (2001) Mechanics of rotating black holes. *Physical Review* 64: 044016 (gr-qc/0103026).

Ashtekar A, Fairhurst S, and Krishnan B (2000) Isolated horizons: Hamiltonian evolution and the first law. *Physical Review D* 62: 104025 (gr-qc/0005083).

Ashtekar A and Krishnan B (2003) Dynamical horizons and their properties. *Physical Review D* 68: 104030 (gr-qc/0308033).

Ashtekar A and Krishnan B (2004) Isolated and dynamical horizons and their applications. *Living Reviews in Relativity* 10: 1–78 (gr-qc/0407042).

Bardeen JW, Carter B, and Hawking SW (1973) The four laws of black hole mechanics. *Communications in Mathematical Physics* 31: 161.

DeWitt BS and DeWitt CM (eds.) (1972) *Black Holes*. Amsterdam: North-Holland.

Frolov VP and Novikov ID (1998) *Black Hole Physics*. Dordrecht: Kluwer.

Hawking SW and Ellis GFR (1973) *Large Scale Structure of Space-Time*. Cambridge: Cambridge University Press.

Hayward S (1994) General laws of black hole dynamics. *Physical Review D* 49: 6467–6474.

Iyer V and Wald RM (1994) Some properties of noether charge and a proposal for dynamical black hole entropy. *Physical Review D* 50: 846–864.

Wald RM (1994) *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*. Chicago: University of Chicago Press.

# Boltzmann Equation (Classical and Quantum)

**M Pulvirenti**, Università di Roma “La Sapienza,” Rome, Italy



## Introduction

Ludwig Boltzmann (1872) established an evolution equation to describe the behavior of a rarefied gas, starting from the mathematical model of elastic balls and using mechanical and statistical considerations. The importance of this equation is twofold. First, it provides a reduced description (as well as the hydrodynamical equations) of the microscopic world. Second, it is also an important tool for the applications, especially for dilute fluids when the hydrodynamical equations fail to hold.

The starting point of the Boltzmann analysis is to abandon the study of the gas in terms of the detailed motion of molecules which constitute it because of their large number. Instead, it is better to investigate a function $f(x, v)$, which is the probability density of a given particle, where $x$ and $v$ denote its position and velocity. Actually, $f(x, v)\mathrm{d}x\,\mathrm{d}v$ is often confused with the fraction of molecules falling in the cell of the phase space of size $\mathrm{d}x\,\mathrm{d}v$ around $x$, $v$. The two concepts are not exactly the same, but they are asymptotically equivalent (when the number of particles is diverging) if a law of large numbers holds.

The Boltzmann equation is the following:

$$(\partial_t + v \cdot \nabla_x) f = Q(f, f) \qquad [1]$$

where $Q$, the collision operator, is defined by eqn [2]:

$$Q(f, f) = \int_{\mathbb{R}^3} \mathrm{d}v_1 \int_{S^2_+} \mathrm{d}n (v - v_1) \cdot n \times [f(x, v') f(x, v'_1) - f(x, v) f(x, v_1)] \qquad [2]$$

and

$$\begin{aligned} v' &= v - n[n \cdot (v - v_1)] \\ v'_1 &= v_1 + n[n \cdot (v - v_1)] \end{aligned} \qquad [3]$$

Moreover, $n$ (the impact parameter) is a unitary vector and $S^2_+ = \{n | n \cdot (v - v_1) \geq 0\}$.

Note that $v', v'_1$ are the outgoing velocities after a collision of two elastic balls with incoming velocities $v$ and $v_1$ and centers $x$ and $x + rn$, $r$ being the diameter of the spheres. Obviously, the collision takes place if $n \cdot (v - v_1) \geq 0$. Equations [3] are a consequence of the conservation of total energy, momentum, and angular momentum. Note also that $r$ does not enter in eqn [1] as a parameter.

As fundamental features of eqn [1], we have the conservation in time of the following five quantities

$$\int \mathrm{d}x \int \mathrm{d}v\, f(x, v; t) v^\alpha \qquad [4]$$

with $\alpha = 0, 1, 2$, expressing conservation of the probability, momentum, and energy.

From now on we shall set $\int = \int_{\mathbb{R}^3}$ for notational simplicity.

Moreover, Boltzmann introduced the (kinetic) entropy defined as

$$H(f) = \int \mathrm{d}x \int \mathrm{d}v\, f \, \log f(x, v) \qquad [5]$$

and proved the famous $H$-theorem asserting the decreasing of $H(f(t))$ along the solutions to eqn [1].

Finally, in the case of bounded domains or homogeneous solutions ($f = f(v; t)$ is independent of $x$), the distribution defined for some $\beta > 0, \rho > 0$, and $u \in \mathbb{R}^3$ by

$$M(x, v) = \frac{\rho}{(2\pi/\beta)^{3/2}} \mathrm{e}^{-(\beta/2)|v-u|^2} \qquad [6]$$

called Maxwellian distribution, is stationary for the evolution given by eqn [1]. In addition, $M$ minimizes $H$ among all distributions with given total mass $\rho$, given mean velocity $u$, and mean energy. The parameter $\beta$ is interpreted as the inverse temperature.

In conclusion, Boltzmann was able to introduce not only an evolutionary equation with the remarkable properties expressing mass, momentum, and energy conservation, but also the trend to the thermal equilibrium. In other words, he tried to conciliate the Newton’s laws with the second principle of thermodynamics.

## The Boltzmann Heuristic Argument

Thus, we want to find an evolution equation for the quantity $f(x, v; t)$. The molecular system we are considering consists of $N$ identical particles of diameter $r$ in the whole space $\mathbb{R}^3$. We denote by $x_1, v_1, \ldots, x_N, v_N$ a state of the system, where $x_i$ and $v_i$ indicate the position and the velocity of the particle $i$. The particles cannot overlap (i.e., the centers of two particles cannot be at a distance smaller than the particle diameter $r$).

The particles are moving freely up to the first instance of contact, that is, the first time when two particles (say particles $i$ and $j$) arrive at a distance $r$. Then the pair interacts when an elastic collision occurs. This means that they change instantaneously

their velocities, according to the conservation of the energy and linear and angular momentum. More precisely, the velocities after a collision with incoming velocities $v$ and $v_1$ are those given by formula [3]. After the first collision, the system evolves by iterating the procedure. Here we neglect triple collisions because they are unlikely. The evolution equation for a tagged particle is then of the form

$$(\partial_t + v \cdot \nabla_x)f = \text{Coll} \qquad [7]$$

where Coll denotes the variation of $f$ due to the collisions.

We have

$$\text{Coll} = G - L \qquad [8]$$

where $L$ and $G$ (the loss and gain terms, respectively) are the negative and positive contributions to the variation of $f$ due to the collisions. More precisely, $L\,\mathrm{d}x\,\mathrm{d}v\,\mathrm{d}t$ is the probability of the test particle to disappear from the cell $\mathrm{d}x\,\mathrm{d}v$ of the phase space because of a collision in the time interval $(t, t + \mathrm{d}t)$ and $G\mathrm{d}x\,\mathrm{d}v\,\mathrm{d}t$ is the probability to appear in the same time interval for the same reason. Let us consider the sphere of center $x$ with radius $r$ and a point $x + rn$ over the surface, where $n$ denotes the generic unit vector. Consider also the cylinder with base area $\mathrm{d}S = r^2\,\mathrm{d}n$ and height $|V|\mathrm{d}t$ along the direction of $V = v_2 - v$.

Then a given particle (say particle 2) with velocity $v_2$ can contribute to $L$ because it can collide with the test particle in the time $\mathrm{d}t$, provided it is localized in the cylinder and if $V \cdot n \le 0$. Therefore, the contribution to $L$ due to the particle 2 is the probability of finding such a particle in the cylinder (conditioned to the presence of the first particle in $x$). This quantity is $f_2(x, v, x + nr, v_2)\,|\,(v_2 - v) \cdot n|r^2\,\mathrm{d}n\,\mathrm{d}v_2\,\mathrm{d}t$, where $f_2$ is the joint distribution of two particles. Integrating in $\mathrm{d}n$ and $\mathrm{d}v_2$, we obtain that the total contribution to $L$ due to any predetermined particle is

$$r^2 \int \mathrm{d}v_2 \int_{S^2_-} \mathrm{d}n\, f_2(x, v, x + nr, v_2)|(v_2 - v) \cdot n| \qquad [9]$$

where $S^2_-$ is the unit hemisphere $(v_2 - v) \cdot n < 0$. Finally, we obtain the total contribution multiplying by the total number of particles:

$$L = (N-1)r^2 \int \mathrm{d}v_2 \times \int_{S_-} \mathrm{d}n\, f_2(x, v, x + nr, v_2)|(v_2 - v) \cdot n| \qquad [10]$$

The gain term can be derived analogously by considering that we are looking at particles which have velocities $v$ and $v_2$ after the collisions so that we have to integrate over the hemisphere $S^2_+ = \{(v_2 - v) \cdot n > 0\}$:

$$G = (N-1)r^2 \int \mathrm{d}v_2 \times \int_{S_+} \mathrm{d}n\, f_2(x, v, x + nr, v_2)|(v_2 - v) \cdot n| \qquad [11]$$

Summing $G$ and $-L$, we get

$$\text{Coll} = (N-1)r^2 \int \mathrm{d}v_2 \times \int \mathrm{d}n\, f_2(x, v, x + nr, v_2)(v_2 - v) \cdot n \qquad [12]$$

which, however, is not a very useful expression because the time derivative of $f$ is expressed in terms of another object, namely $f_2$. An evolution equation for $f_2$ will imply $f_3$, the joint distribution of three particles, and so on, up to we include the total particle number $N$. Here the basic main assumption of Boltzmann enters, namely that two given particles are uncorrelated if the gas is rarefied, namely

$$f(x, v, x_2, v_2) = f(x, v)f(x_2, v_2) \qquad [13]$$

Condition [13], referred to as the propagation of chaos, seems contradictory at first sight: if two particles collide, correlations are created. Even though we could assume eqn [13] at some time, if the test particle collides with particle 2, such an equation cannot be satisfied anymore after the collision.

Before discussing the propagation of chaos hypothesis, we first analyze the size of the collision operator. We remark that, in practical situations for a rarefied gas, the combination $Nr^3 \approx 10^{-4}\,\mathrm{cm}^3$ (i.e., the volume occupied by the particles) is very small, while $Nr^2 = O(1)$. This implies that $G = O(1)$. Therefore, since we are dealing with a very large number of particles, we are tempted to perform the limit $N \to \infty$ and $r \to 0$ in such a way that $r^2 = O(N^{-1})$. As a consequence, the probability that two tagged particles collide (which is of the order of the surface of a ball, i.e., $O(r^2)$) is negligible. However, the probability that a given particle performs a collision with any one of the remaining $N - 1$ particles (which is $O(Nr^2) = O(1)$) is not negligible. Therefore, condition [13] is referring to two preselected particles (say particles 1 and 2), so that it is not unreasonable to conceive that it holds in the limiting situation in which we are working.

However, we cannot insert [13] in [12] because this latter equation refers to instants before and after the collision and, if we know that a collision took place, we certainly cannot invoke eqn [13]. Hence, it is more convenient to assume eqn [13] in the loss term and work over the gain term to keep advantage

of the factorization property which will be assumed only before the collision.

Coming back to eqn [11] for the outgoing pair velocities $v, v_2$ (satisfying the condition $(v_2 - v) \cdot n > 0$), we make use of the continuity property

$$f_2(x, v, x + nr, v_2) = f_2(x, v', x + nr, v'_2) \quad [14]$$

where the pair $v', v'_2$ is pre-collisional. On $f_2$ expressed before the collision, we can reasonably apply condition [13] and obtain

$$G - L = (N-1)r^2 \int dv_2 \int_{S^2_+} dn (v - v_2) \cdot n \times [f(x, v') f(x - nr, v'_2) - f(x, v) f(x + nr, v_2)] \quad [15]$$

after a change $n \to -n$ in the gain term, using the notation $S^2_+$ for the hemisphere $\{n| = (v_2 - v) \cdot n \geq 0\}$. This transforms the pair $v', v'_2$ from a pre-collisional to a post-collisional pair.

Finally, in the limit $N \to \infty$, $r \to 0$, $Nr^2 = \lambda^{-1}$, we find

$$(\partial_t + v \cdot \nabla_x) f = \lambda^{-1} \int dv_2 \int_{S_+} dn (v - v_2) \cdot n \times [f(x, v') f(x, v'_2) - f(x, v) f(x, v_2)] \quad [16]$$

The parameter $\lambda$, called mean free path, represents, roughly speaking, the typical length a particle can cover without undergoing any collision. In eqns [1] and [2], we just chose $\lambda = 1$.

Equation [16] (or, equivalently, eqns [1] and [2]) is the Boltzmann equation for hard spheres. Such an equation has a statistical nature, and it is not equivalent to the Hamiltonian dynamics from which it has been derived. Indeed, the $H$-theorem shows that such an equation is not reversible in time as expected of any law of mechanics.

This concludes the heuristic preliminary analysis of the Boltzmann equation. We certainly know that the above arguments are delicate and require a more rigorous and deeper analysis. If we want the Boltzmann equation not to be a phenomenological model, derived by *ad hoc* assumptions and justified only by its practical relevance, but rather that it is a consequence of a mechanical model, we must derive it rigorously. In particular, the propagation of chaos should be not a hypothesis but the statement of a theorem.

## Beyond the Hard Spheres

The heuristic arguments we have developed so far can be extended to different potentials than that of the hard-sphere systems. If the particles interact via a two-body interaction $V = V(r)$, the resulting Boltzmann equation is eqn [1], with

$$Q(f, f) = \int dv_1 \int_{S^2_+} dn\, B(v - v_1; n) [f' f'_1 - f f_1] \quad [17]$$

where we are using the usual shorthand notation:

$$f' = f(x, v'), \quad f'_1 = f(x, v'_1), \quad f = f(x, v), \quad f_1 = f(x, v_1) \quad [18]$$

and $B = B(v - v_1; n)$ is a suitable function of the relative velocity $v - v_1$ and the impact parameter $n$, which is proportional to the cross section relative to the potential $V$. Another equivalent, sometimes more convenient, way, to express eqn [17] is

$$Q(f, f) = \int dv_1 \int dv' \int dv'_1\, W(v, v_1 | v', v'_1) [f' f'_1 - f f_1] \quad [19]$$

with

$$W(v, v_1 | v', v'_1) = w(v, v_1 | v', v'_1) \times \delta(v + v_1 - v' - v'_1) \times \delta\left(\tfrac{1}{2}\left(v^2 + v_1^2 - (v')^2 - (v'_1)^2\right)\right) \quad [20]$$

where $w$ is a suitable kernel. All the qualitative properties, such as the conservation laws and the $H$-theorem, are obviously still valid.

## Consequences

The Boltzmann equation provoked a debate involving Loschmidt, Zermelo, and Poincaré, who outlined inconsistencies between the irreversibility of the equation and the reversible character of the Hamiltonian dynamics. Boltzmann argued the statistical nature of his equation and his answer to the irreversibility paradox was that "most" of the configurations behave as expected by the thermodynamical laws. However, he did not have the probabilistic tools for formulating in a precise way the statements of which he had a precise intuition.

Grad (1949) stated clearly the limit $N \to \infty$, $r \to 0$, $Nr^2 \to$ const., where $N$ is the number of particles and $r$ is the diameter of the molecules, in which the Boltzmann equation is expected to hold. This limit is usually called the Boltzmann–Grad limit (B–G limit in the sequel).

The problem of a rigorous derivation of the Boltzmann equation was an open and challenging problem for a long time. Lanford (1975) showed that, although for a very short time, the Boltzmann equation can be derived starting from the mechanical model of the hard-sphere system. The proof has a deep content but is relatively simple from a technical viewpoint.

## Existence

The mathematical study of the Boltzmann equation starts with the problem of proving the existence of the solutions. One would like to be able to show that, for all (or at least for a physically significant family of) initial distributions (which are positive and summable functions) with finite momentum, energy, and entropy, there exists a unique solution to eqn [1] with the same mass, momentum, and energy as of the initial distribution. Moreover, the entropy should decrease and the solution should approach the right Maxwellian as $t \to \infty$. The problem, in such a generality, is still unsolved, but several results in this direction have been achieved since the pioneering works due to Carleman (1933) for the homogeneous equation. Actually, there are satisfactory results for some special situations, such as the homogeneous solutions (independent of $x$) close to the equilibrium, to the vacuum, or to homogeneous data. The most general result we have up to now is, unfortunately, not constructive. This is due to Di Perna and Lions (1989), who showed the existence of suitable weak solutions to eqn [1]. However, we still do not know whether such solutions, which preserve mass and momentum, and satisfy the $H$-theorem, are unique and also preserve the energy.

## Hydrodynamics

The derivation of hydrodynamical equations from the Boltzmann equation is a problem as old as the equation itself and, in fact, it goes back to Maxwell and Hilbert. Preliminary to the discussion of the hydrodynamic limit, we establish a few properties of the collision kernel.

It is a well-known fact that the only solution to the equation

$$Q(f,f) = 0 \qquad [21]$$

is a local Maxwellian, namely

$$f(x,v) := M(x,v) = \frac{\rho(x)}{(2\pi T(x))^{3/2}} \mathrm{e}^{-|v-u(x)|^2/2T(x)} \qquad [22]$$

where the local parameters $\rho, \rho u$, and $T$ satisfy the relations

$$\int M \, \mathrm{d}v = \rho \qquad [23]$$

$$\int vM = \rho u \qquad [24]$$

$$\frac{1}{2}\int v^2 M \, \mathrm{d}v = \frac{3}{2}\rho T + \frac{1}{2}\rho u^2 \qquad [25]$$

Moreover, the only solution to the equation

$$\int h(v) Q(f,f) \, \mathrm{d}v = 0 \qquad [26]$$

is any linear combination of the quantities $(1, v, v^2)$, called collision invariants. The last property obviously corresponds to the mass, momentum, and energy conservation.

With this in mind, consider a change of variables in the Boltzmann equation [1], passing from microscopic to macroscopic variables, $x \to \varepsilon x$, $t \to \varepsilon t$. Here $\varepsilon$ is a small scale parameter expressing the ratio between the typical inter-particle distances and the typical distances over which the macroscopic equations are varying. Such a change yields

$$(\partial_t + v \cdot \nabla_x) f_\varepsilon = \frac{1}{\varepsilon} Q(f_\varepsilon, f_\varepsilon) \qquad [27]$$

We need to allow the small parameter $\varepsilon$ (mean free path or the Knudsen number) to tend to zero. In order to eliminate the singularity on the right-hand side of [27], we multiply both sides by the collision invariants $v^\alpha$ with $\alpha = 0, 1, 2$, and obtain the five equations:

$$\int \mathrm{d}v \, v^\alpha (\partial_t + v \cdot \nabla_x) f_\varepsilon = 0 \qquad [28]$$

On the other hand, if $f_\varepsilon$ converges to $f$, as $\varepsilon \to 0$, necessarily $Q(f,f) = 0$ and hence $f = M$. Therefore, we expect that in the limit $\varepsilon \to 0$,

$$\int \mathrm{d}v \, v^\alpha (\partial_t + v \cdot \nabla_x) M = 0 \qquad [29]$$

Equation [29] fixes a relation among the fields $\rho, u, T$ as functions of $x$ and $t$. A standard computation gives us the Euler equations for compressible gas

$$\partial_t \rho + \mathrm{div}(\rho u) = 0 \qquad [30]$$

$$\partial_t u + (u \cdot \nabla) u + \frac{1}{\rho} \nabla p = 0 \qquad [31]$$

$$\partial_t T + (u \cdot \nabla) T + \tfrac{2}{3} T \nabla u = 0 \qquad [32]$$

where the pressure $p$ is related to the density $\rho$ and the temperature $T$ by the perfect gas law

$$p = \rho T \qquad [33]$$

In order to make the above arguments rigorous, Hilbert (1916) developed a useful tool, called the Hilbert expansion, to control the limiting procedure.

Namely, he expressed a formal solution to eqn [27] in the form of a power series expansion:

$$f_\varepsilon = \sum_{j\geq 0} f_j\, \varepsilon^j \qquad [34]$$

where $f_0$ is the local Maxwellian, with the parameters $\rho, u, T$ satisfying the Euler equations. All the other coefficients $f_j$ of the developments can be determined by recurrence, inverting suitable operators. However, the series is not expected to be convergent, so that the way to show the validity of the hydrodynamical limit rigorously is to truncate the expansion and to control the remainder. The first result in this direction was obtained by Caflisch (1980). However, this approach is based on the regularity of the solutions to the Euler equations, which is known to hold only for short times since shocks can be formed. How to approximate the shocks in terms of a kinetic description is still a difficult and open problem.

Note that the hydrodynamical picture of the Boltzmann equation just means that we are looking at the solutions of this equation at a suitable macroscopic scale. The rarefaction hypothesis underlying the Boltzmann description is reflected in the law of perfect gas, which states that the particles, in the local thermal equilibrium, are free.

## Stationary Problems

Stationary non-Maxwellian solutions to the Boltzmann equation should describe stationary nonequilibrium states exhibiting nontrivial flows. In spite of the physical relevance of these problems, not many complete mathematical results are, at the moment, available. Among them, there is the traveling-wave problem, which can be formulated in the following way. We look for a solution $f = f(x - ct, v), f: \mathbb{R}\times\mathbb{R}^3 \to \mathbb{R}^+$, constant in form but traveling with a constant velocity $c > 0$, to

$$(v_1 - c)f' = Q(f, f) \qquad [35]$$

where $v_1$ is the first component of $v$ and $f'$ denotes the spatial derivative of $f$. Equation [35] must be complemented by the boundary conditions which are $f \to M_\pm$, as $x \to \infty$, where $M_\pm$ are the right and left Maxwellians, namely two prescribed equilibrium situations at infinity. The parameters (density, mean velocity, and temperature) of the Maxwellians, however, cannot be chosen arbitrarily. Indeed, the conservations of the mass, momentum, and energy (which are properties of $Q$) imply the conservations (in $x$) of the fluxes of these quantities. Hence, we have to impose five equations that relate the upstream and the downstream values of the densities, mean velocities, and temperatures. Such relations are known in gas dynamics as the Rankine–Hugoniot conditions. A solution of this problem has been found by Caflisch and Nikolaenko (1983) in case of a weak shock (namely, when $M_+$ and $M_-$ are close) by using Hilbert expansion techniques. More recently, Liu and Yu (2004) established also stability and positivity of this solution.

## Quantum Kinetic Theory

Uehling and Uhlembeck (1933) introduced the following kinetic equation for describing a large system of weakly interacting bosons or fermions:

$$\begin{aligned}(\partial_t + v\cdot\nabla_x)f = \int \mathrm{d}v_1 \int \mathrm{d}v' \int \mathrm{d}v_1'\, W(v, v_1|v', v_1') \\ \times \{(1\pm f)(1\pm f_1)f'f_1' \\ - (1\pm f')(1\pm f_1')ff_1\}\end{aligned} \qquad [36]$$

Here the +/− sign, stand for bosons/fermions, respectively, and

$$\begin{aligned}W(v, v_1|v', v_1') \\ = (\hat{V}(v' - v) - \hat{V}(v' - v_1))^2\delta(v + v_1 - v' - v_1') \\ \times \delta\left(\tfrac{1}{2}\left(v^2 + v_1^2 - (v')^2 - (v_1')^2\right)\right)\end{aligned} \qquad [37]$$

Moreover,

$$\hat{V}(p) = 4\pi \int \mathrm{d}x\, \mathrm{e}^{\mathrm{i}p\cdot x} \qquad [38]$$

where $V$ is the interaction potential. Note that eqn [37] is the expression of the cross section of a quantum scattering in the Born approximation.

The unknown $f = f(x, v; t)$ in eqn [37] is the expected number of molecules falling in the unit (quantum) cell of the phase space. This function is proportional to the one-particle Wigner function, introduced by Wigner (1932) to handle kinetic problems in quantum mechanics, and defined as (setting $\hbar = 1$):

$$\frac{1}{(2\pi)^3}\int \mathrm{d}y\, \mathrm{e}^{\mathrm{i}y\cdot v}\rho\left(x + \tfrac{1}{2}y; x - \tfrac{1}{2}y\right)$$

where $\rho(x; z)$ is the kernel of a one-particle density matrix. Basically, the Wigner function is an equivalent way to describe a state of a quantum system. For instance, eqn [40] below expresses the equilibrium distributions for bosons and fermions in terms of Wigner functions. In general, the Wigner functions, due to the uncertainty principle, are real but not necessarily positive; however, the integral with respect to $x$ and $v$ gives the probability

distributions of the velocity and the position, respectively. In the kinetic regime, in which we are interested, the scales are mesoscopic, namely the typical quantum oscillations are on a scale much smaller than the characteristic scales of the problem, so that we expect that $f$ should be a genuine probability distribution, since the Heisenberg principle does not play an essential role. However, the interaction occurs on a microscopic scale, so that we expect that the statistics play a role in addition to the quantum rules for the scattering.

In this framework, the entropy functional is

$$H(f) = \int \mathrm{d}x \int \mathrm{d}v\,[f(x,v)\log f(x,v) \mp (1 \pm f(x,v))\log(1 \pm f(x,v))] \qquad [39]$$

It is decreasing along the solutions to eqn [35] and it is also minimized (among the distributions with given mass, momentum, and energy) by the equilibria

$$M(v) = \frac{z}{\mathrm{e}^{(\beta/2)|v-u|^2} \mp z} \qquad [40]$$

namely the Bose–Einstein and the Fermi–Dirac distributions, respectively. Here $\beta > 1$ and $z > 0$ are the inverse temperature and the activity, respectively. Note that, for the Bose–Einstein distribution, $z < 1$. This creates, in a sense, an inconsistency with eqn [36]. Indeed, assuming $u = 0$ and an initial distribution $f = f_0(v)$ with the density larger than the maximal density allowed by eqn [40], namely

$$\rho_c := \int \mathrm{d}v \frac{1}{\mathrm{e}^{(\beta/2)v^2} - 1} \qquad [41]$$

it cannot converge to any equilibrium. In order to overcome this difficulty related to the Bose condensation, one can enlarge the definition of the equilibria family by setting

$$M(v) = \frac{1}{\mathrm{e}^{(\beta/2)v^2} - 1} + \mu\delta(v) \qquad [42]$$

to take care of excess of mass by means of a condensate component. However, it is not clear whether eqn [36] can actually describe the Bose condensation since its derivation from the Schrödinger equation requires, just from the very beginning, the existence of bosonic quasifree states which can be constructed only if the density is moderate. Further analyses are certainly needed to clarify the situation. A rigorous derivation of the Uehling and Uhlembeck equation is, up to now, far from being obtained even for short times; nevertheless, such an equation is extensively used in the applications. Equation [36] concerns a weakly interacting gas of quantum particles. From a mathematical viewpoint, it is expected to be valid in the so-called weak-coupling limit, which consists in scaling space and time and the interaction potential $\phi$ as

$$x \to \varepsilon x, \quad t \to \varepsilon t, \quad \phi \to \sqrt{\varepsilon}\phi \qquad [43]$$

where $\varepsilon^{-1} = N^{1/3}$ is a parameter diverging when the number of particles $N$ tends to infinity.

We mention, incidentally, that under such a scaling, a classical system is described by a transport equation, called Fokker–Planck–Landau equation, with a diffusion operator in the velocity space.

The B–G limit considered for classical particle systems is different from that considered here for weakly interacting quantum systems. It is actually equivalent to rescaling space and time according to

$$x \to \varepsilon x, \qquad t \to \varepsilon t \qquad [44]$$

leaving the interaction unscaled but, in order to control the total interaction, we make the density diverging gently as $\varepsilon^{-1} = N^{1/2}$.

A quantum system under such a scaling is expected to be described by a Boltzmann equation [1] with the collision operator $Q$ computed with the full quantum cross section. Now we do not have any effect of the statistics because in this rarefaction limit these corrections disappear. On the other hand, the cross section is that arising from the analysis of the quantum scattering. Since we do not rescale the interaction, all the other terms in the Born expansion of the cross section play a role. This kind of Boltzmann equation is a good description of a rarefied gas in which quantum effects are not negligible.

*See also:* Adiabatic Piston; Evolution Equations: Linear and Nonlinear; Gravitational *N*-Body Problem (Classical); Interacting Particle Systems and Hydrodynamic Equations; Kinetic Equations; Multiscale Approaches; Nonequilibrium Statistical Mechanics: Dynamical Systems Approach; Quantum Dynamical Semigroups.

## Further Reading

Balesku R (1978) *Equilibrium and Nonequilibrium Statistical Mechanics*. Moscow: Mir (distributed by Imported Publications, Chicago, Ill).

Caflisch RE (1980) The fluid dynamical limit of the nonlinear Boltzmann equation. *Communications of Pure and Applied Mathematics* 33: 651–666.

Caflisch RE and Nicolaenko B (1983) Shock waves and the Boltzmann equation. Nonlinear partial differential equations. *Contemporary Mathematics* 17: 35–44.

Carleman T (1933) Sur la théorie de l'équation intégro-differentielle de Boltzmann. *Acta Mathematica* 60: 91–146.

Cercignani C (1998) *Ludwig Boltzmann. The Man Who Trusted Atoms*. Oxford: Oxford University Press.

Cercignani C, Illner R, and Pulvirenti M (1994) The Mathematical Theory of Dilute Gases. *Springer Series in Applied Mathematics*, vol. 106. New York: Springer.

Di Perna RJ and Lions P-L (1989) On the Cauchy problem for the Boltzmann equations: Global existence and weak stability. *Annals of Mathematics* 130: 321–366.

Grad H (1949) On the kinetic theory of rarified gases. *Communications in Pure and Applied Mathematics* 2: 331–407.

Hilbert D (1916) Begründung der Kinetischen Gastheorie. *Mathematische Annalen* 72: 331–407.

Lanford OE III (1975) Time evolution of large classical systems. In: Ehlers J, Hepp K, and Weidenmüller HA (eds.) *Lecture Notes in Physics*, vol. 38, pp. 1–111. Berlin: Springer.

Liu T-P and Yu S-H (2004) Boltzmann equation: micro–macro decompositions and positivity of shock profiles. *Communications in Mathematical Physics* 246(1): 133–179.

Spohn H (1994) Quantum kinetic equations. In: Fannes M, Maes C, and Verbeure A (eds.) *On Three Levels: Micro, Meso and Macro Approaches in Physics*. New York: Plenum.

Uehling EA and Uhlembeck GE (1933) Transport phenomena in Einstein–Bose and Fermi–Dirac gases. I. *Physical Reviews* 43: 552–561.

Wigner EP (1932) On the quantum correction for thermodynamic equilibrium. *Physical Reviews* 40: 749–759.

# Bose–Einstein Condensates

**F Dalfovo, L P Pitaevskii, and S Stringari**,
Università di Trento, Povo, Italy



## Introduction

In 1924 the Indian physicist S N Bose introduced a new statistical method to derive the blackbody radiation law in terms of a gas of light quanta (photons). His work, together with the contemporary de Broglie's idea of matter–wave duality, led A Einstein to apply the same statistical approach to a gas of $N$ indistinguishable particles of mass $m$. An amazing result of his theory was the prediction that below some critical temperature a finite fraction of all the particles condense into the lowest-energy single-particle state. This phenomenon, named Bose–Einstein condensation (BEC), is a consequence of purely statistical effects. For several years, such a prediction received little attention, until 1938, when F London argued that BEC could be at the basis of the superfluid properties observed in liquid $^4$He below 2.17 K. A strong boost to the investigation of Bose–Einstein condensates was given in 1995 by the observation of BEC in dilute gases confined in magnetic traps and cooled down to temperatures of the order of a few nK. Differently from superfluid helium, these gases allow one to tune the relevant parameters (confining potential, particle density, interactions, etc.), so to make them an ideal test-ground for concepts and theories on BEC.

## What Is BEC?

In nature, particles have either integer or half-integer spin. Those having half-integer spin, like electrons, are called fermions and obey the Fermi–Dirac statistics; those having integer spin are called bosons and obey the Bose–Einstein statistics. Let us consider a system of $N$ bosons. In order to introduce the concept of BEC on a general ground, one can start with the definition of the one-body density matrix

$$n^{(1)}(\boldsymbol{r},\boldsymbol{r}') = \left\langle \hat{\Psi}^\dagger(\boldsymbol{r})\hat{\Psi}(\boldsymbol{r}')\right\rangle \qquad [1]$$

The quantities $\hat{\Psi}^\dagger(\boldsymbol{r})$ and $\hat{\Psi}(\boldsymbol{r})$ are the field operators which create and annihilate a particle at point $\boldsymbol{r}$, respectively; they satisfy the bosonic commutation relations

$$[\hat{\Psi}(\boldsymbol{r}),\hat{\Psi}^\dagger(\boldsymbol{r}')] = \delta(\boldsymbol{r}-\boldsymbol{r}'), \quad [\hat{\Psi}(\boldsymbol{r}),\hat{\Psi}(\boldsymbol{r}')] = 0 \qquad [2]$$

If the system is in a pure state described by the $N$-body wave function $\Psi(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_N)$, then the average [1] is taken following the standard rules of quantum mechanics and the one-body density matrix can be written as

$$n^{(1)}(\boldsymbol{r},\boldsymbol{r}') = N\int \mathrm{d}\boldsymbol{r}_2\cdots\mathrm{d}\boldsymbol{r}_N\Psi^*(\boldsymbol{r},\boldsymbol{r}_2,\ldots,\boldsymbol{r}_N)\Psi(\boldsymbol{r}',\boldsymbol{r}_2,\ldots,\boldsymbol{r}_N) \qquad [3]$$

involving the integration over the $N-1$ variables $\boldsymbol{r}_2,\ldots,\boldsymbol{r}_N$. In the more general case of a statistical mixture of pure states, expression [3] must be averaged according to the probability for a system to occupy the different states.

Since $n^{(1)}(\boldsymbol{r},\boldsymbol{r}') = (n^{(1)}(\boldsymbol{r}',\boldsymbol{r}))^*$ the quantity $n^{(1)}$, when regarded as a matrix function of its indices $\boldsymbol{r}$ and $\boldsymbol{r}'$, is Hermitian. It is therefore always possible to find a complete orthonormal basis of single-particle eigenfunctions, $\varphi_i(\boldsymbol{r})$, in terms of which the density matrix takes the diagonal form

$$n^{(1)}(\boldsymbol{r},\boldsymbol{r}') = \sum_i n_i\varphi_i^*(\boldsymbol{r})\varphi_i(\boldsymbol{r}') \qquad [4]$$

The real eigenvalues $n_i$ are subject to the normalization condition $\sum_i n_i = N$ and have the meaning of occupation numbers of the single-particle states $\varphi_i$. BEC occurs when one of these numbers (say, $n_0$) becomes macroscopic, that is, when $n_0 \equiv N_0$ is a number of order $N$, all the others remaining of order 1.

In this case eqn [4] can be conveniently rewritten in the form

$$n^{(1)}(\boldsymbol{r},\boldsymbol{r}') = N_0\varphi_0^*(\boldsymbol{r})\varphi_0(\boldsymbol{r}') + \sum_{i\neq 0} n_i\varphi_i^*(\boldsymbol{r})\varphi_i(\boldsymbol{r}') \quad [5]$$

and the state represented by $\varphi_0(\boldsymbol{r})$ is called Bose–Einstein condensate. This definition is rather general, since it applies to any macroscopic ($N \gg 1$) system of indistinguishable bosons independently of mutual interactions and external fields.

The one-body density matrix [1] contains information on important physical observables. By setting $\boldsymbol{r} = \boldsymbol{r}'$ one finds the diagonal density of the system

$$n(\boldsymbol{r}) \equiv n^{(1)}(\boldsymbol{r},\boldsymbol{r}) = \langle\hat{\Psi}^\dagger(\boldsymbol{r})\hat{\Psi}(\boldsymbol{r})\rangle \quad [6]$$

with $N = \int \mathrm{d}\boldsymbol{r}\, n(\boldsymbol{r})$. The off-diagonal components can instead be used to calculate the momentum distribution

$$n(\boldsymbol{p}) = \langle\hat{\Psi}^\dagger(\boldsymbol{p})\hat{\Psi}(\boldsymbol{p})\rangle \quad [7]$$

where $\hat{\Psi}(\boldsymbol{p}) = (2\pi\hbar)^{-3/2}\int \mathrm{d}\boldsymbol{r}\hat{\Psi}(\boldsymbol{r})\exp[-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{r}/\hbar]$ is the field operator in momentum representation. By inserting this expression for $\hat{\Psi}(\boldsymbol{p})$ into eqn [7] one finds

$$n(\boldsymbol{p}) = \frac{1}{(2\pi\hbar)^3}\int \mathrm{d}\boldsymbol{R}\,\mathrm{d}\boldsymbol{s}\; n^{(1)}\left(\boldsymbol{R}+\frac{\boldsymbol{s}}{2}, \boldsymbol{R}-\frac{\boldsymbol{s}}{2}\right)\mathrm{e}^{-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{s}/\hbar} \quad [8]$$

where $\boldsymbol{s} = \boldsymbol{r} - \boldsymbol{r}'$ and $\boldsymbol{R} = (\boldsymbol{r}+\boldsymbol{r}')/2$.

Let us consider a uniform system of $N$ particles in a volume $V$ and take the thermodynamic limit $N, V \to \infty$ with density $N/V$ kept fixed. The eigenfunctions of the density matrix are plane waves and the lowest-energy state has zero momentum, $\boldsymbol{p} = 0$, and constant wave function $\varphi_0(\boldsymbol{r}) = V^{-1/2}$. BEC in this state implies a macroscopic number of particles having zero momentum and constant density $N_0/V$. The density matrix only depends on $\boldsymbol{s} = \boldsymbol{r} - \boldsymbol{r}'$ and can be written as

$$n^{(1)}(\boldsymbol{s}) = \frac{N_0}{V} + \frac{1}{V}\sum_{\boldsymbol{p}\neq 0} n_{\boldsymbol{p}}\,\mathrm{e}^{-\mathrm{i}\boldsymbol{p}\cdot\boldsymbol{s}/\hbar} \quad [9]$$

In the $s \to \infty$ limit, the sum on the right vanishes due to destructive interference between different plane waves, but the first term survives. One thus finds that, in the presence of BEC, the one-body density matrix tends to a constant finite value at large distances. This behavior is named *off-diagonal long-range order*, since it involves the off-diagonal components of the density matrix. Its counterpart in momentum space is the appearance of a singular term at $\boldsymbol{p} = 0$:

$$n(\boldsymbol{p}) = N_0\delta(\boldsymbol{p}) + \sum_{\boldsymbol{p}'\neq 0} n_{\boldsymbol{p}'}\delta(\boldsymbol{p}-\boldsymbol{p}') \quad [10]$$

The sum on the right is the number of noncondensed particles ($N - N_0$), and the quantity $N_0/N$ is called condensate fraction.

If the system is not uniform, the eigenfunctions of the density matrix are no longer plane waves but, provided $N$ is sufficiently large, the concept of BEC is still well defined, being associated with the occurrence of a macroscopic occupation of a single-particle eigenfunction $\varphi_0(\boldsymbol{r})$ of the density matrix. Thus, the condensed bosons can be described by means of the function $\Psi(\boldsymbol{r}) = \sqrt{N_0}\varphi_0(\boldsymbol{r})$, which is a classical complex field playing the role of an *order parameter*. This is the analog of the classical limit of quantum electrodynamics, where the electromagnetic field replaces the microscopic description of photons. The function $\Psi$ may also depend on time and can be written as

$$\Psi(\boldsymbol{r},t) = |\Psi(\boldsymbol{r},t)|\,\mathrm{e}^{\mathrm{i}S(\boldsymbol{r},t)} \quad [11]$$

Its modulus determines the contribution of the condensate to the diagonal density [6], while the phase $S$ is crucial in characterizing the coherence and superfluid properties of the system. The order parameter [11], also named *macroscopic wave function* or *condensate wave function*, is defined only up to a constant phase factor. One can always multiply this function by the numerical factor $\mathrm{e}^{\mathrm{i}\alpha}$ without changing any physical property. This reflects the gauge symmetry exhibited by all the physical equations of the problem. Making an explicit choice for the value of the order parameter, and hence for the phase, corresponds to a formal breaking of gauge symmetry.

## BEC in Ideal Gases

Once we have defined what is a Bose–Einstein condensate, the next question is when such a condensation occurs in a given system. The ideal Bose gas provides the simplest example. So, let us consider a gas of noninteracting bosons described by the Hamiltonian $\hat{H} = \sum_i \hat{H}_i^{(1)}$, where the Schrödinger equation $\hat{H}^{(1)}\varphi_i(\boldsymbol{r}) = \epsilon_i\varphi_i(\boldsymbol{r})$ gives the spectrum of single-particle wave functions and energies. One can define an occupation number $n_i$ as the number of particles in the state with energy $\epsilon_i$. Thus, any given state of the many-body system is specified by a set $\{n_i\}$. The mean occupation numbers, $\bar{n}_i$, can be calculated by using the standard rules of statistical mechanics. For instance, by considering a grand canonical ensemble at temperature $T$, one finds

$$\bar{n}_i = \{\exp[\beta(\epsilon_i - \mu)] - 1\}^{-1} \quad [12]$$

with $\beta = 1/(k_B T)$. The chemical potential $\mu$ is fixed by the normalization condition $\sum_i \bar{n}_i = N$, where $N$ is the average number of particles in the gas. For $T \to \infty$ the chemical potential is negative and large. It increases monotonically when $T$ is lowered. Let us call $\epsilon_0$ the lowest single-particle level in the spectrum. If at some critical temperature $T_c$ the normalization condition can be satisfied with $\mu \to \epsilon_0^-$, then the occupation of the lowest state, $\bar{n}_0 = N_0$, becomes of order $N$ and BEC is realized. Below $T_c$ the normalization condition must be replaced with $N = N_0 + N_T$, where $N_T = \sum_{i \neq 0} \bar{n}_i$ is the number of particles out of the condensate, that is, the *thermal* component of the gas. Whether BEC occurs or not, and what is the value of $T_c$ depends on the dimensionality of the system and the type of single-particle spectrum.

The simplest case is that of a gas confined in a cubic box of volume $V = L^3$ with periodic boundary conditions, where $\hat{H}^{(1)} = -(\hbar^2/2m)\nabla^2$. The eigenfunctions are plane waves $\varphi_p(\boldsymbol{r}) = V^{-1/2} \exp[-i\boldsymbol{p} \cdot \boldsymbol{r}/\hbar]$, with energy $\epsilon_p = p^2/2m$ and momentum $\boldsymbol{p} = 2\pi\hbar\boldsymbol{n}/L$. Here $\boldsymbol{n}$ is a vector whose components $n_x, n_y, n_z$ are 0 or $\pm$ integers. The lowest eigenvalue has zero energy ($\epsilon_0 = 0$) and zero momentum. The mean occupation numbers are given by $\bar{n}_p = \{\exp[\beta(p^2/2m - \mu)] - 1\}^{-1}$. In the thermodynamic limit ($N, V \to \infty$ with $N/V$ kept constant), one can replace the sum $\sum_p$ with the integral $\int \mathrm{d}\epsilon \rho(\epsilon)$, where $\rho(\epsilon) = (2\pi)^{-2} V (2m/\hbar^2)^{3/2} \sqrt{\epsilon}$ is the density of states. In this way, one can calculate the thermal component of the gas as a function of $T$, finding the critical temperature

$$k_B T_c = \frac{2\pi\hbar^2}{m} \left( \frac{N}{V\zeta(3/2)} \right)^{2/3} \qquad [13]$$

where $\zeta$ is the Riemann zeta function and $\zeta(3/2) \simeq 2.612$. For $T > T_c$, one has $\mu < 0$ and $N_T = N$. For $T < T_c$ one instead has $\mu = 0, N_T = N - N_0$ and

$$N_0(T) = N[1 - (T/T_c)^{3/2}] \qquad [14]$$

The critical temperature turns out to be fully determined by the density $N/V$ and by the mass of the constituents. These results were first obtained by A Einstein in his seminal paper and used by F London in the context of superfluid helium. We notice that the replacement of the sum with an integral in the above derivation is justified only if the thermal energy $k_B T$ is much larger than the energy spacing between single-particle levels, that is, if $k_B T \gg \hbar^2/2mV^{2/3}$. Is is also worth noticing that the above expression for $T_c$ can be written as $\lambda_T^3 N/V \simeq 2.612$, where $\lambda_T = [2\pi\hbar^2/(mk_B T)]^{1/2}$ is the thermal de Broglie wavelength. This is equivalent to saying that BEC occurs when the mean distance between bosons is of the order of their de Broglie wavelength.

Another interesting case, which is relevant for the recent experiments with BEC in dilute gases confined in magnetic and/or optical traps, is that of an ideal gas subject to harmonic potentials. Let us consider, for simplicity, an isotropic external potential $V_{ext}(r) = (1/2)m\omega_{ho}^2 r^2$. The single-particle Hamiltonian is $\hat{H}^{(1)} = -(\hbar^2/2m)\nabla^2 + V_{ext}(r)$ and its eigenvalues are $\epsilon_{n_x,n_y,n_z} = (n_x + n_y + n_z + 3/2)\hbar\omega_{ho}$. The corresponding density of states is $\rho(\epsilon) = (1/2)(\hbar\omega_{ho})^{-3}\epsilon^2$. A natural thermodynamic limit for this system is obtained by letting $N \to \infty$ and $\omega_{ho} \to 0$, while keeping the product $N\omega_{ho}^3$ constant. The condition for BEC to occur is that $\mu$ approaches the value $\epsilon_{000} = (3/2)\hbar\omega_{ho}$ from below by cooling the gas down to $T_c$. Following the same procedure as for the uniform gas, one finds

$$k_B T_c = \hbar\omega_{ho}[N/\zeta(3)]^{1/3} = 0.94\hbar\omega_{ho} N^{1/3} \qquad [15]$$

and

$$N_0(T) = N[1 - (T/T_c)^3] \qquad [16]$$

Notice that the condensate is not uniform in this case, since it corresponds to the lowest eigenfunction of the harmonic oscillator, which is a Gaussian of width $a_{ho} = [\hbar/(m\omega_{ho})]^{1/2}$. Correspondingly, the condensate in the momentum space is also a Gaussian, of width $a_{ho}^{-1}$. This implies that, differently from the gas in a box, here the condensate can be seen both in coordinate and momentum space in the form of a narrow distribution emerging from a wider thermal component. Finally, results [15] and [16] remain valid even for anisotropic harmonic potentials, with trapping frequencies $\omega_x, \omega_y$, and $\omega_z$, provided the frequency $\omega_{ho}$ is replaced by the geometric average $(\omega_x\omega_y\omega_z)^{1/3}$.

## BEC in Interacting Gases

Actual condensates are made of interacting particles. The full many-body Hamiltonian is

$$\hat{H} = \int \mathrm{d}\boldsymbol{r} \hat{\Psi}^\dagger(\boldsymbol{r}) \hat{H}_0 \hat{\Psi}(\boldsymbol{r}) + \frac{1}{2} \int \mathrm{d}\boldsymbol{r}' \mathrm{d}\boldsymbol{r}\, \hat{\Psi}^\dagger(\boldsymbol{r}) \hat{\Psi}^\dagger(\boldsymbol{r}') V(\boldsymbol{r} - \boldsymbol{r}') \hat{\Psi}(\boldsymbol{r}') \hat{\Psi}(\boldsymbol{r}) \qquad [17]$$

where $V(\boldsymbol{r} - \boldsymbol{r}')$ is the particle–particle interaction and $\hat{H}_0 = -(\hbar^2/2m)\nabla^2 + V_{ext}(\boldsymbol{r})$. Differently from the case of ideal gases, $\hat{H}$ is no longer a sum of single-particle Hamiltonians. However, the general definitions given in the section "What is BEC?" are still valid. In particular, the one-body density matrix, in the presence of BEC, can be separated as in eqn [5]. One

can write $n^{(1)}(\boldsymbol{r},\boldsymbol{r}')=\Psi^*(\boldsymbol{r})\Psi(\boldsymbol{r}')+\tilde{n}^{(1)}(\boldsymbol{r},\boldsymbol{r}')$, where $\Psi$ is the order parameter of the condensate ($\Psi^*(\boldsymbol{r})\Psi(\boldsymbol{r}')$ being of order $N$), while $\tilde{n}^{(1)}(\boldsymbol{r},\boldsymbol{r}')$ vanishes for large $|\boldsymbol{r}-\boldsymbol{r}'|$. This is equivalent to say that the bosonic field operator splits in two parts,

$$\hat{\Psi}(\boldsymbol{r}) = \Psi(\boldsymbol{r}) + \delta\hat{\Psi}(\boldsymbol{r}) \quad [18]$$

where the first term is a complex function and the second one is the field operator associated with the noncondensed particles. This decomposition is particularly useful when the depletion of the condensate, that is, the fraction of noncondensed particles, is small. This happens when the interaction is weak, but also for particles with arbitrary interaction, provided the gas is dilute. In this case, one can expand the many-body Hamiltonian by treating the operator $\delta\hat{\Psi}$ as a small quantity.

A suitable strategy consists in writing the Heisenberg equation for the evolution of the field operators, $\mathrm{i}\hbar\partial_t\hat{\Psi}=[\hat{\Psi},\hat{H}]$, using the many-body Hamiltonian [17]:

$$\mathrm{i}\hbar\partial_t\hat{\Psi}(\boldsymbol{r},t) = \left(\hat{H}_0 + \int \mathrm{d}\boldsymbol{r}'\hat{\Psi}^\dagger(\boldsymbol{r}',t)V(\boldsymbol{r}-\boldsymbol{r}')\hat{\Psi}(\boldsymbol{r}',t)\right) \times \hat{\Psi}(\boldsymbol{r},t) \quad [19]$$

The zeroth-order is thus obtained by replacing the operator $\hat{\Psi}$ with the classical field $\Psi$. In the integral containing the interaction $V(\boldsymbol{r}-\boldsymbol{r}')$, this replacement is, in general, a poor approximation when short distances $(\boldsymbol{r}-\boldsymbol{r}')$ are involved. In a dilute and cold gas, one can nevertheless obtain a proper expression for the interaction term by observing that, in this case, only binary collisions at low energy are relevant and these collisions are characterized by a single parameter, the *s*-wave scattering length, $a$, independently of the details of the two-body potential. This allows one to replace $V(\boldsymbol{r}-\boldsymbol{r}')$ in $\hat{H}$ with an effective interaction $V(\boldsymbol{r}-\boldsymbol{r}')=g\delta(\boldsymbol{r}-\boldsymbol{r}')$, where the coupling constant $g$ is given by $g=4\pi\hbar^2a/m$. The scattering length can be measured with several experimental techniques or calculated from the exact two-body potential. Using this pseudopotential and replacing the operator $\hat{\Psi}$ with the complex function $\Psi$ in the Heisenberg equation of motion, one gets

$$\mathrm{i}\hbar\partial_t\Psi(\boldsymbol{r},t) = \left(-\frac{\hbar^2\nabla^2}{2m} + V_{\mathrm{ext}}(\boldsymbol{r}) + g|\Psi(\boldsymbol{r},t)|^2\right)\Psi(\boldsymbol{r},t) \quad [20]$$

This is known as Gross–Pitaevskii (GP) equation and it was first introduced in 1961. It has the form of a *nonlinear Schrödinger equation*, the nonlinearity coming from the mean-field term, proportional to $|\Psi|^2$. It has been derived assuming that $N$ is large while the fraction of noncondensed atoms is negligible. On the one hand, this means that quantum fluctuations of the field operator have to be small, which is true when $n|a|^3 \ll 1$, where $n$ is the particle density. In fact, one can show that, at $T=0$ the quantum depletion of the condensate is proportional to $(n|a|^3)^{1/2}$. On the other hand, thermal fluctuations have also to be negligible and this means that the theory is limited to temperatures much lower than $T_{\mathrm{c}}$. Within these limits, one can identify the total density with the condensate density.

The stationary solution of eqn [20] corresponds to the condensate wave function in the ground state. One can write $\Psi(\boldsymbol{r},t)=\Psi_0(\boldsymbol{r})\exp(-\mathrm{i}\mu t/\hbar)$, where $\mu$ is the chemical potential. Then the GP equation [20] becomes

$$\left(-\frac{\hbar^2\nabla^2}{2m} + V_{\mathrm{ext}}(\boldsymbol{r}) + g|\Psi_0(\boldsymbol{r})|^2\right)\Psi_0(\boldsymbol{r}) = \mu\Psi_0(\boldsymbol{r}) \quad [21]$$

where $n(\boldsymbol{r})=|\Psi_0(\boldsymbol{r})|^2$ is the particle density. The same equation can be obtained by minimizing the energy of the system written as a functional of the density:

$$E[n] = \int \mathrm{d}\boldsymbol{r}\left[\frac{\hbar^2}{2m}|\boldsymbol{\nabla}\sqrt{n}|^2 + nV_{\mathrm{ext}}(\boldsymbol{r}) + \frac{gn^2}{2}\right] \quad [22]$$

The first term on the right corresponds to the quantum kinetic energy coming from the uncertainty principle; it is usually named "quantum pressure" and vanishes for uniform systems.

The next order in $\delta\hat{\Psi}$ gives the excited states of the condensate. In a uniform gas the ground-state order parameter, $\Psi_0$, is a constant and the first-order expansion of $\hat{H}$ was introduced by N Bogoliubov in 1947. In particular, he found an elegant way to diagonalize the Hamiltonian by using simple linear combinations of particle creation and annihilation operators. These are known as Bogoliubov's transformations and stay at the basis of the concept of *quasiparticle*, one of the most important concepts in quantum many-body theory.

A generalization of Bogoliubov's approach to the case of nonuniform condensates is obtained by considering small deviations around the ground state in the form

$$\Psi(\boldsymbol{r},t) = \mathrm{e}^{-\mathrm{i}\mu t/\hbar}\left[\Psi_0(\boldsymbol{r}) + u(\boldsymbol{r})\mathrm{e}^{-\mathrm{i}\omega t} + v^*(\boldsymbol{r})\mathrm{e}^{\mathrm{i}\omega t}\right] \quad [23]$$

Inserting this expression into eqn [20] and keeping terms linear in the complex functions $u$ and $v$, one gets

$$\hbar\omega u(\boldsymbol{r}) = [\hat{H}_0 - \mu + 2g\Psi_0^2(\boldsymbol{r})]u(\boldsymbol{r}) + g\Psi_0^2(\boldsymbol{r})v(\boldsymbol{r}) \quad [24]$$

$$-\hbar\omega v(\boldsymbol{r}) = [\hat{H}_0 - \mu + 2g\Psi_0^2(\boldsymbol{r})]v(\boldsymbol{r}) + g\Psi_0^2(\boldsymbol{r})u(\boldsymbol{r}) \quad [25]$$

These coupled equations allow one to calculate the energies $\varepsilon=\hbar\omega$ of the excitations. They also give the so-called quasiparticle amplitudes $u$ and $v$, which obey the normalization condition

$$\int \mathrm{d}\boldsymbol{r}[u_i^*(\boldsymbol{r})u_j(\boldsymbol{r}) - v_i^*(\boldsymbol{r})v_j(\boldsymbol{r})] = \delta_{ij}$$

In a uniform gas, $u$ and $v$ are plane waves and one recovers the famous Bogoliubov's spectrum

$$\hbar\omega = \left[\frac{\hbar^2 q^2}{2m}\left(\frac{\hbar^2 q^2}{2m} + 2gn\right)\right]^{1/2} \qquad [26]$$

where $\boldsymbol{q}$ is the wave vector of the excitations. For large momenta the spectrum coincides with the free-particle energy $\hbar^2q^2/2m$. At low momenta, it instead gives the phonon dispersion $\omega=cq$, where $c=[gn/m]^{1/2}$ is the Bogoliubov sound velocity. The transition between the two regimes occurs when the excitation wavelength is of the order of the *healing length*,

$$\xi = [8\pi na]^{-1/2} = \hbar/(mc\sqrt{2}) \qquad [27]$$

which is an important length scale for superfluidity. When the order parameter is forced to vanish at some point (by an impurity, a wall, etc.), the healing length provides the typical distance over which it recovers its bulk value. In a nonuniform condensate the excitations are no longer plane waves but, at low energy, they have still a phonon-like character, in the sense that they involve a collective motion of the condensate.

The GP equation [20] is the starting point for an accurate mean-field description of BEC in dilute cold gases, which is rigorous at $T=0$ and for $n|a|^3 \ll 1$. Static and dynamics properties of condensates in different geometries can be calculated by solving the GP equation numerically or using suitable approximated methods. The inclusion of effects beyond mean field is a highly nontrivial and interesting problem. A rather extreme case is represented by liquid $^4$He, which is a dense system where the interaction between atoms causes a large depletion of the condensate even at $T=0$ ($N_0/N$ being less than 10%) and thus a full many-body treatment is required for its rigorous description. Nevertheless, even in this case, the general definitions of the section "What is BEC?" are still useful.

## Superfluidity and Coherence

With the word superfluidity, one summarizes a complex of macroscopic phenomena occurring in quantum fluids under particular conditions: persistent currents, equilibrium states at rest in rotating vessels, viscousless motion, quantized vorticity, and others. These features can also be observed in BEC. The link between BEC and superfluidity is given by the phase of the order parameter [11]. To understand this point, let us consider a uniform system. If $\hat{\Psi}(\boldsymbol{r},t)$ is a solution of the Heisenberg equation [19] with $V_{\mathrm{ext}}=0$, then

$$\hat{\Psi}'(\boldsymbol{r},t) = \hat{\Psi}(\boldsymbol{r}-\boldsymbol{v}t,t)\exp\left[\frac{\mathrm{i}}{\hbar}\left(m\boldsymbol{v}\cdot\boldsymbol{r} - \frac{1}{2}mv^2t\right)\right] \qquad [28]$$

where $\boldsymbol{v}$ is a constant vector, is also a solution. This equation gives the Galilean transformation of the field operator and also applies to its condensate component $\Psi$. At equilibrium, the ground-state order parameter is given by $\Psi_0=\sqrt{n}\exp(-\mathrm{i}\mu t/\hbar)$, where $n$ is a constant independent of $\boldsymbol{r}$. In a frame where the condensate moves with velocity $\boldsymbol{v}$, the order parameter instead takes the form $\Psi_0=\sqrt{n}\exp(\mathrm{i}S)$, with $S(\boldsymbol{r},t)=\hbar^{-1}[m\boldsymbol{v}\cdot\boldsymbol{r}-(mv^2/2+\mu)t]$. The velocity of the condensate can thus be identified with the gradient of the phase $S$:

$$\boldsymbol{v}(\boldsymbol{r},t) = \frac{\hbar}{m}\nabla S(\boldsymbol{r},t) \qquad [29]$$

This definition is also valid for $\boldsymbol{v}$ varying slowly in space and time. The modulus of the order parameter plays a minor role in this definition and it is not necessary to assume the gas to be dilute and close to $T=0$. Indeed, the relation [29] between the velocity field and the phase of the order parameter also applies in the presence of large quantum depletion, as in superfluid $^4$He, and at $T\neq 0$. In this case, $n$ should not be identified with the condensate density. Conversely, in dilute gases at $T=0$, $n$ is the condensate density and the velocity [29] can be simply obtained by applying the usual definition of current density operator, $\hat{\boldsymbol{j}}$, to the order parameter [11].

The velocity [29] describes a potential flow and corresponds to a collective motion of many particles occupying a single quantum state. Being equal to the gradient of a scalar function, it is irrotational ($\nabla\times\boldsymbol{v}_{\mathrm{s}}=0$) and satisfies the Onsager–Feynman quantization condition $\oint \boldsymbol{v}_{\mathrm{s}}\cdot\mathrm{d}\boldsymbol{l}=\kappa h/m$, with $\kappa$ non-negative integer. These conditions are not satisfied by a classical fluid, where the hydrodynamic velocity field, $\boldsymbol{v}(\boldsymbol{r},t)=\boldsymbol{j}(\boldsymbol{r},t)/n(\boldsymbol{r},t)$, is the average over many different states and does not correspond to a potential flow.

By using the definition of the phase $S$ and velocity $\boldsymbol{v}$, together with particle conservation, one can show that the dynamics of a condensate, as far as macroscopic motions are concerned, is governed by the hydrodynamic equations of an irrotational

nonviscous fluid. Within the mean-field theory, this can be easily seen by rewriting the GP equation [20] in terms of the density $n = |\Psi|^2$ and the velocity [29]. Neglecting the quantum pressure term $\nabla^2\sqrt{n}$ (hence limiting the description to length scales larger than the healing length $\xi$), one gets

$$\frac{\partial}{\partial t} n + \boldsymbol{\nabla} \cdot (\boldsymbol{v} n) = 0 \qquad [30]$$

and

$$m \frac{\partial}{\partial t} \boldsymbol{v} + \boldsymbol{\nabla}\left( V_{\text{ext}} + \mu(n) + \frac{m v^2}{2} \right) = 0 \qquad [31]$$

with the local chemical potential $\mu(n) = gn$. These equations have the typical structure of the dynamic equations of superfluids at zero temperature and can be viewed as the $T = 0$ case of the more general Landau's two-fluid theory.

One of the most striking evidences of superfluidity is the observation of quantized vortices, that is, vortices obeying the Onsager–Feynman quantization condition. A vast literature is devoted to vortices in superfluid helium and, more recently, vortices have also been produced and studied in condensates of ultracold gases, including nice configurations of many vortices in regular triangular lattices, similar to the Abrikosov lattices in superconductors. Other phenomena, such as the reduction of the moment of inertia, the occurrence of Josephson tunneling through barriers, the existence of thresholds for dissipative processes (Landau criterion), and others, are typical subjects of intense investigation.

Another important consequence of the fact that BEC is described by an order parameter with a well-defined phase is the occurrence of coherence effects which, in different words, mean that condensates behave like *matter waves*. For instance, one can measure the phase difference between two condensates by means of interference. This can be done in coordinate space by confining two condensates in two potential minima, $a$ and $b$, at a distance $d$. Let us take $d$ along $z$ and assume that, at $t = 0$, the order parameter is given by the linear combination $\Psi(\boldsymbol{r}) = \Psi_a(\boldsymbol{r}) + \exp(\mathrm{i}\phi)\Psi_b(\boldsymbol{r})$ with $\Psi_a$ and $\Psi_b$ real and without overlap. Then let us switch off the confining potentials so that the condensates expand and overlap. If the overlap occurs when the density is small enough to neglect interactions, the motion is ballistic and the phase of each condensate evolves as $S(\boldsymbol{r}, t) \simeq mr^2/(2\hbar t)$, so that $\boldsymbol{v} = \boldsymbol{r}/t$. This implies a relative phase $\phi + S(x, y, z + d/2) - S(x, y, z - d/2) = \phi + mdz/\hbar t$. The total density $n = |\Psi|^2$ thus exhibits periodic modulations along $z$ with wavelength $\hbar t/md$. This interference pattern has indeed been observed in condensates of ultracold atoms. In these systems it was also possible to measure the coherence length, that is, the distance $|\boldsymbol{r} - \boldsymbol{r}'|$ at which the one-body density vanishes and the phase of the order parameter is no more well defined. In most situations, the coherence length turns out to be of the order of, or larger than the size of the condensates. However, interesting situations exist when the coherence length is shorter but the system still preserves some features of BEC (quasicondensates).

## Final Remarks

Bose–Einstein condensates of ultracold atoms are easily manipulated by changing and tuning the external potentials. This means, for instance, that one can prepare condensates in different geometries, including very elongated (quasi-1D) or disk-shaped (quasi-2D) condensates. This is conceptually important, since BEC in lower dimensions is not as simple as in three dimensions: thermal and quantum fluctuations play a crucial role, superfluidity must be properly re-defined, and very interesting limiting cases can be explored (Tonks–Girardeau regime, Luttinger liquid, etc.). Another possibility is to use laser beams to produce standing waves acting as an external periodic potential (optical lattice). Condensates in optical lattices behave as a sort of perfect crystal, whose properties are the analog of the dynamic and transport properties in solid-state physics, but with controllable spacing between sites, no defects and tunable lattice geometry. One can investigate the role of phase coherence in the lattice, looking, for instance, at Josephson effects as in a chain of junctions. By tuning the lattice depth one can explore the transition from a superfluid phase and a Mott-insulator phase, which is a nice example of quantum phase transition. Controlling cold atoms in optical lattice can be a good starting point for application in quantum engineering, interferometry, and quantum information.

Another interesting aspect of BECs is that the key equation for their description in mean-field theory, namely the GP equation [20], is a nonlinear Schrödinger equation very similar to the ones commonly used, for instance, in nonlinear quantum optics. This opens interesting perspectives in exploiting the analogies between the two fields, such as the occurrence of dynamical and parametric instabilities, the possibility to create different types of solitons, the occurrence of nonlinear processes like, for example, higher harmonic generation and mode mixing.

A relevant part of the current research also involves systems made of mixtures of different gases, Bose–Bose or Fermi–Bose, and many activities with ultracold atoms now involve fermionic gases, where BEC can

also be realized by condensing molecules of fermionic pairs. An extremely active research now concerns the BCS–BEC crossover, which can be obtained in Fermi gases by tuning the scattering length (and hence the interaction) by means of Feshbach resonances.

Ten years after the first observation of BEC in ultracold gases, it is almost impossible to summarize all the researches done in this field. A large amount of work has already been devoted to characterize the condensates and several new lines have been opened. Rather detailed review articles and books are already available for the interested readers.

*See also:* Interacting Particle Systems and Hydrodynamic Equations; Quantum Phase Transitions; Quantum Statistical Mechanics: Overview; Renormalization: Statistical Mechanics and Condensed Matter; Superfluids; Variational Techniques for Ginzburg–Landau Energies.

## Further Reading

Cornell EA and Wieman CE (2002) Nobel lecture: Bose–Einstein condensation in a dilute gas, the first 70 years and some recent experiments. *Reviews of Modern Physics* 74: 875.

Dalfovo F, Giorgini S, Pitaevskii LP, and Stringari S (1999) Theory of Bose–Einstein condensation in trapped gases. *Reviews of Modern Physics* 71: 463.

Griffin A, Snoke DW, and Stringari S (1995) *Bose–Einstein Condensation*. Cambridge: Cambridge University Press.

Huang K (1987) *Statistical Mechanics*, 2nd edn. New York: Wiley.

Inguscio M, Stringari S, and Wieman CE (1999) *Bose–Einstein Condensation in Atomic Gases*, Proceedings of the International School of Physics "Enrico Fermi," Course CXL. Amsterdam: IOS Press.

Ketterle W (2002) Nobel lecture: when atoms behave as waves: Bose–Einstein condensation and the atom laser. *Reviews of Modern Physics* 74: 1131.

Landau LD and Lifshitz EM (1980) *Statistical Physics*, Part 1. Oxford: Pergamon Press.

Leggett AJ (2001) Bose–Einstein condensation in the alkali gases: some fundamental concepts. *Reviews of Modern Physics* 73: 307.

Lifshitz EM and Pitaevskii LP (1980) *Statistical Physics*, Part 2. Oxford: Pergamon Press.

Pethick CJ and Smith H (2002) *Bose–Einstein Condensation in Dilute Gases*. Cambridge: Cambridge University Press.

Pitaevskii LP and Stringari S (2003) *Bose–Einstein Condensation*. Oxford: Clarendon Press.

# Bosons and Fermions in External Fields

**E Langmann**, KTH Physics, Stockholm, Sweden



## Introduction

In this article we discuss quantum theories which describe systems of nondistinguishable particles interacting with external fields. Such models are of interest also in the nonrelativistic case (in quantum statistical mechanics, nuclear physics, etc.), but the relativistic case has additional, interesting complications: relativistic models are genuine quantum field theories, that is, quantum theories with an infinite number of degrees of freedom, with nontrivial features like divergences and anomalies. Since interparticle interactions are ignored, such models can be regarded as a first approximation to more complicated theories, and they can be studied by mathematically precise methods.

Models of relativistic particles in external electromagnetic fields have received considerable attention in the physics literature, and interesting phenomena like the Klein paradox or particle–antiparticle pair creation in overcritical fields have been studied; see Rafelski *et al.* (1978) for an extensive review. We will not discuss these physics questions but only describe some prototype examples and a general Hamiltonian framework which has been used in mathematically precise work on such models. The general framework for this latter work is the mathematical theory of Hilbert space operators (see, e.g., Reed and Simon (1975)), but in our discussion we try to avoid presupposing knowledge of that theory. As mentioned briefly in the end, this work has had close relations to various topics of recent interest in mathematical physics, including anomalies, infinite-dimensional geometry and group theory, conformal field theory, and noncommutative geometry.

We restrict our discussion to spin-0 bosons and spin-1/2 fermions, and we will not discuss models of particles in external gravitational fields but only refer the interested reader to DeWitt (2003). We also only mention in passing that external field problems have also been studied using functional integral approaches, and mathematically precise work on this can be found in the extensive literature on determinants of differential operators.

## Examples

Consider the Schrödinger equation describing a nonrelativistic particle of mass $m$ and charge $e$

moving in three-dimensional space and interacting with an external vector and scalar potentials $\boldsymbol{A}$ and $\phi$, respectively,

$$\mathrm{i}\partial_t\psi = H\psi, \quad H = \frac{1}{2m}(-\mathrm{i}\nabla + e\boldsymbol{A})^2 - e\phi \qquad [1]$$

(we set $\hbar = c = 1$, $\partial_t = \partial/\partial t$, and $\psi$,$\phi$, and $\boldsymbol{A}$ can depend on the space and time variables $\boldsymbol{x} \in \mathbb{R}^3$ and $t \in \mathbb{R}$). This is a standard quantum-mechanical model, with $\psi$ the one-particle wave function allowing for the usual probabilistic interpretation.

One interesting generalization to the relativistic regime is the Klein–Gordon equation

$$\left[(\mathrm{i}\partial_t + e\phi)^2 - (-\mathrm{i}\nabla + e\boldsymbol{A})^2 - m^2\right]\psi = 0 \qquad [2]$$

with a $\mathbb{C}$-valued function $\psi$. There is another important relativistic generalization, the Dirac equation

$$[(\mathrm{i}\partial_t + e\phi) - (-\mathrm{i}\nabla + e\boldsymbol{A})\cdot\boldsymbol{\alpha} + m\beta]\psi = 0 \qquad [3]$$

with $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ and $\beta$ Hermitian $4 \times 4$ matrices satisfying the relations

$$\alpha_i\alpha_j + \alpha_j\alpha_i = \delta_{ij}, \quad \alpha_i\beta = -\beta\alpha_i, \quad \beta^2 = 1 \qquad [4]$$

and a $\mathbb{C}^4$-valued function $\psi$ (we also write 1 for the identity). These two relativistic equations differ by the transformation properties of $\psi$ under Lorentz transformations: in [2] it transforms like a scalar and thus describes spin-0 particles, and it transforms like a spinor describing spin-1/2 particles in [3]. While these equations are natural relativistic generalizations of the Schrödinger equation, they no longer allow to consistently interpret $\psi$ as one-particle wave functions. The physical reason is that, in a relativistic theory, high-energy processes can create particle–antiparticle pairs, and this makes the restriction to a fixed particle number inconsistent. This problem can be remedied by constructing a many-body model allowing for an arbitrary number of particles and antiparticles. The requirement that this many-body model should have a ground state is an important ingredient in this construction.

It is obviously of interest to formulate and study many-body models of nondistinguishable particles already in the nonrelativistic case. An important empirical fact is that such particles come in two kinds, bosons and fermions, distinguished by their exchange statistics (we ignore the interesting possibility of exotic statistics). For example, the fermion many-particle version of [1] for suitable $\phi$ and $\boldsymbol{A}$ is a useful model for electrons in a metal. An elegant method to go from the one- to the many-particle description is the formalism of second quantization: one promotes $\psi$ to a quantum field operator with certain (anti-) commutator relations, and this is a convenient way to construct the appropriate many-particle Hilbert space, Hamiltonian, etc. In the nonrelativistic case, this formalism can be regarded as an elegant reformulation of a pedestrian construction of a many-body quantum-mechanical model, which is useful since it provides convenient computational tools. However, this formalism naturally generalizes to the relativistic case where the one-particle model no longer has an acceptable physical interpretation, and one finds that one can nevertheless give a consistent physical interpretation to [2] and [3] provided that $\psi$ are interpreted as quantum field operators describing bosons and fermions. This particular exchange statistics of the relativistic particles is a special case of the spin-statistics theorem: *integer-spin particles are bosons and half-integer spin particles are fermions*. While many structural features of this formalism are present already in the simpler nonrelativistic models, the relativistic models add some nontrivial features typical for quantum field theories.

In the following, we discuss a precise mathematical formulation of the quantum field theory models described above. We emphasize the functorial nature of this construction, which makes manifest that it also applies to other situations, for example, where the bosons and fermions are also coupled to a gravitational background, are considered in other spacetime dimensions than $3 + 1$, etc.

## Second Quantization: Nonrelativistic Case

Consider a quantum system of nondistinguishable particles where the quantum-mechanical description of one such particle is known. In general, this one-particle description is given by a Hilbert space $h$ and one-particle observables and transformations which are self-adjoint and unitary operators on $h$, respectively. The most important observable is the Hamiltonian $H$. We will describe a general construction of the corresponding many-body system.

**Example** As a motivating example we take the Hilbert space $h = L^2(\mathbb{R}^3)$ of square-integrable functions $f(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^3$, and the Hamiltonian $H$ in [1]. A specific example for a unitary operator on $h$ is the gauge transformation $(Uf)(\boldsymbol{x}) = \exp(\mathrm{i}\chi(\boldsymbol{x}))f(\boldsymbol{x})$ with $\chi$ a smooth, real-valued function on $\mathbb{R}^3$.

In this example, the corresponding wave functions for $N$ identical such particles are the $L^2$-functions $f_N(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$, $\boldsymbol{x}_j \in \mathbb{R}^3$. It is obvious how to extend

one-particle observables and transformations to such $N$-particle states: for example, the $N$-particle Hamiltonian corresponding to $H$ in [1] is

$$H_N = \sum_{j=1}^{N} \frac{1}{2m}(-\mathrm{i}\nabla_{x_j} + e\boldsymbol{A}(t,\boldsymbol{x}_j))^2 - e\phi(t,\boldsymbol{x}_j) \quad [5]$$

and the $N$-particle gauge transformation $U_N$ is defined through multiplication with $\prod_{j=1}^{N} \exp(\mathrm{i}\chi(\boldsymbol{x}_j))$.

For systems of indistinguishable particles it is enough to restrict to wave functions which are even or odd under particle exchanges,

$$\begin{aligned} &f_N(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_j,\ldots,\boldsymbol{x}_k,\ldots,\boldsymbol{x}_N) \\ &\quad = \pm f_N(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k,\ldots,\boldsymbol{x}_j,\ldots,\boldsymbol{x}_N) \end{aligned} \quad [6]$$

for all $1 \le j < k \le N$, with the upper and lower signs corresponding to bosons and fermions, respectively (this empirical fact is usually taken as a postulate in nonrelativistic many-body quantum physics). It is convenient to define the zero-particle Hilbert space as $\mathbb{C}$ (complex numbers) and to introduce a Hilbert space containing states with all possible particle numbers: this so-called Fock space contains all states

$$\begin{pmatrix} f_0 \\ f_1(\boldsymbol{x}_1) \\ f_2(\boldsymbol{x}_1,\boldsymbol{x}_2) \\ f_3(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_3) \\ \vdots \end{pmatrix} \quad [7]$$

with $f_0 \in \mathbb{C}$. The definition of $H_N$ and $U_N$ then naturally extends to this Fock space; see below.

### General Construction

The construction of Fock spaces and many-particle observables and transformations just outlined in a specific example is conceptually simple. An alternative, more efficient construction method is to use "quantum fields," which we denote as $\psi(\boldsymbol{x})$ and $\psi^\dagger(\boldsymbol{x})$, $\boldsymbol{x} \in \mathbb{R}^3$. They can be fully characterized by the following (anti-) commutator relations:

$$[\psi(\boldsymbol{x}),\psi^\dagger(\boldsymbol{y})]_\mp = \delta^3(\boldsymbol{x}-\boldsymbol{y}), \quad [\psi(\boldsymbol{x}),\psi(\boldsymbol{y})]_\mp = 0 \quad [8]$$

where $[a,b]_\mp \equiv ab \mp ba$, with the commutator and anticommutators (upper and lower signs, respectively) corresponding to the boson and fermion case, respectively. It is convenient to "smear" these fields with one-particle wave functions and define

$$\begin{aligned} \psi(f) &= \int_{\mathbb{R}^3} \mathrm{d}^3x\overline{f(\boldsymbol{x})}\psi(\boldsymbol{x}) \\ \psi^\dagger(f) &= \int_{\mathbb{R}^3} \mathrm{d}^3x\psi^\dagger(\boldsymbol{x})f(\boldsymbol{x}) \end{aligned} \quad [9]$$

for all $f \in h$. Then the relations characterizing the field operators can be written as

$$\begin{aligned} &[\psi(f),\psi^\dagger(g)]_\mp = (f,g) \\ &[\psi(f),\psi(g)]_\mp = 0 \\ &\quad \forall f,g \in h \end{aligned} \quad [10]$$

where

$$(f,g) = \int_{\mathbb{R}^3} \mathrm{d}^3x\overline{f(\boldsymbol{x})}g(\boldsymbol{x})$$

is the inner product in $h$. The Fock space $\mathcal{F}_\mp(h)$ can then be defined by postulating that it contains a normalized vector $\Omega$ called "vacuum" such that

$$\psi(f)\Omega = 0 \quad \forall f \in h \quad [11]$$

and that all $\psi^{(\dagger)}(f)$ are operators on $\mathcal{F}_\mp(h)$ such that $\psi^\dagger(f) = \psi(f)^*$, where $*$ is the Hilbert space adjoint. Indeed, from this we conclude that $\mathcal{F}_\mp(h)$, as vector space, is generated by

$$f_1 \wedge f_2 \wedge \cdots \wedge f_N \equiv \psi^\dagger(f_1)\psi^\dagger(f_2)\cdots\psi^\dagger(f_N)\Omega \quad [12]$$

with $f_j \in h$ and $N = 0,1,2,\ldots$, and that the Hilbert space inner product of such vectors is

$$\begin{aligned} &\langle f_1 \wedge f_2 \wedge \cdots \wedge f_N, g_1 \wedge g_2 \wedge \cdots \wedge g_M\rangle \\ &\quad = \delta_{N,M} \sum_{P\in S_N} (\pm 1)^{|P|} \prod_{j=1}^{N} (f_j, g_{Pj}) \end{aligned} \quad [13]$$

with $S_N$ the permutation group, with $(+1)^{|P|} = 1$ always, and $(-1)^{|P|} = +1$ and $-1$ for even and odd permutations, respectively. The many-body Hamiltonian $q(H)$ corresponding to the one-particle Hamiltonian $H$ can now be defined by the following relations:

$$q(H)\Omega = 0, \quad [q(H),\psi^\dagger(f)] = \psi^\dagger(Hf) \quad [14]$$

for all $f \in h$ such that $Hf$ is defined. Indeed, this implies that

$$\begin{aligned} &q(H) f_1 \wedge f_2 \wedge \cdots \wedge f_N \\ &\quad = \sum_{j=1}^{N} f_1 \wedge f_2 \wedge \cdots \wedge (Hf_j) \wedge \cdots \wedge f_N \end{aligned} \quad [15]$$

which defines a self-adjoint operator on $\mathcal{F}_\mp(h)$, and it is easy to check that this coincides with our down-to-earth definition of $H_N$ above. Similarly, the many-body transformation $Q(U)$ corresponding to a one-particle transformation $U$ can be defined as

$$Q(U)\Omega = \Omega, \quad Q(U)\psi^\dagger(f) = \psi^\dagger(Uf)Q(U) \quad [16]$$

for all $f \in h$, which implies that

$$\begin{aligned} &Q(U) f_1 \wedge f_2 \wedge \cdots \wedge f_N \\ &\quad = (Uf_1) \wedge (Uf_2) \wedge \cdots \wedge (Uf_N) \end{aligned} \quad [17]$$

and thus coincides with our previous definition of $U_N$.

While we presented the construction above for a particular example, it is important to note that it actually does not make reference to what the one-particle formalism actually is. For example, if we had a model of particles on a space $\mathcal{M}$ given by some "nice" manifold of any dimension and with $M$ internal degrees of freedom, we would take $h = L^2(\mathcal{M}) \otimes \mathbb{C}^M$ and replace [9] by

$$\psi(f) = \int_{\mathcal{M}} \mathrm{d}\mu(\boldsymbol{x}) \sum_{j=1}^{M} \overline{f_j(\boldsymbol{x})}\psi_j(\boldsymbol{x}) \qquad [18]$$

and its Hermitian conjugate, with the measure $\mu$ on $\mathcal{M}$ defining the inner product in $h$,

$$(f, g) = \int \mathrm{d}\mu(\boldsymbol{x}) \sum_j \overline{f_j(\boldsymbol{x})} g_j(\boldsymbol{x})$$

With that, all formulas after [9] hold true as they stand. *Given any one-particle Hilbert space $h$ with inner product $(\cdot\,,\cdot)$, observable $H$, and transformation $U$, the formulas above define the corresponding Fock spaces $\mathcal{F}_\mp(h)$ and many-body observable $q(H)$ and transformation $Q(U)$.* It is also interesting to note that this construction has various beautiful general (functorial) properties: the set of one-particle observables has a natural Lie algebra structure with the Lie bracket given by the commutator (strictly speaking: i times the commutator, but we drop the common factor i for simplicity). The definitions above imply that

$$[q(A), q(B)] = q([A, B]) \qquad [19]$$

for one-particle observables $A$, $B$, that is, the above-mentioned Lie algebra structure is preserved under this map $q$. In a similar manner, the set of one-particle transformations has a natural group structure preserved by the map $Q$,

$$Q(U)Q(V) = Q(UV), \quad Q(U)^{-1} = Q(U^{-1}) \qquad [20]$$

Moreover, if $A$ is self-adjoint, then $\exp(\mathrm{i}A)$ is unitary, and one can show that

$$Q(\exp(\mathrm{i}A)) = \exp(\mathrm{i}q(A)) \qquad [21]$$

For later use, we note that, if $\{f_n\}_{n\in\mathbb{Z}}$ is some complete, orthonormal basis in $h$, then operators $A$ on $h$ can be represented by infinite matrices $(A_{mn})_{m,n\in\mathbb{Z}}$ with $A_{mn} = (f_m, Af_n)$, and

$$q(A) = \sum_{m,n} A_{mn}\psi_m^\dagger\psi_n \qquad [22]$$

where $\psi_n^{(\dagger)} = \psi^{(\dagger)}(f_n)$ obey

$$[\psi_m, \psi_n^\dagger]_\mp = \delta_{m,n}, \qquad [\psi_m, \psi_n^\dagger]_\mp = 0 \qquad [23]$$

for all $m$, $n$. We also note that, in our definition of $q(A)$, we made a convenient choice of normalization, but there is no physical reason to not choose a different normalization and define

$$q'(A) = q(A) - b(A) \qquad [24]$$

where $b$ is some linear function mapping self-adjoint operators $A$ to real numbers. For example, one may wish to use another reference vector $\tilde{\Omega}$ instead of $\Omega$ in the Fock space, and then would choose $b(A) = \langle\tilde{\Omega}, q(A)\tilde{\Omega}\rangle$. Then the relations in [19] are changed to

$$[q'(A), q'(B)] = q'([A, B]) + S_0(A, B) \qquad [25]$$

where $S_0(A, B) = b([A, B])$. However, the $\mathbb{C}$-number term $S_0(A, B)$ in the relations [25] is trivial, since it can be removed by going back to $q(A)$.

### Physical Interpretation

The Fock space $\mathcal{F}_\mp(h)$ is the direct sum of subspaces of states with different particle numbers $N$,

$$\mathcal{F}_\mp(h) = \bigoplus_{N=0}^{\infty} h_\mp^{(N)} \qquad [26]$$

where the zero-particle subspace $h_\mp^{(0)} = \mathbb{C}$ is generated by the vacuum $\Omega$, and $h_\mp^{(N)}$ is the $N$-particle subspace generated by the states $f_1 \wedge f_2 \wedge \cdots \wedge f_N$, $f_j \in h$. We note that

$$\mathcal{N} \equiv q(1) \qquad [27]$$

is the "particle-number operator," $\mathcal{N}F_N = NF_N$ for all $F_N \in h_\mp^{(N)}$. The field operators obviously change the particle number: $\psi^\dagger(f)$ increases the particle number by one (maps $h_\mp^{(N)}$ to $h_\mp^{(N+1)}$), and $\psi(f)$ decreases it by one. Since every $f \in h$ can be interpreted as one-particle state, it is natural to interpret $\psi^\dagger(f)$ and $\psi(f)$ as "creation" and "annihilation" operators, respectively: they create and annihilate one particle in the state $f \in h$. It is important to note that, in the fermion case, [10] implies that $\psi^\dagger(f)^2 = 0$, which is a mathematical formulation of the Pauli exclusion principle: *it is not possible to have two fermions in the same one-particle state*. In the boson case, there is no such restriction. Thus, even though the formalisms used to describe boson and fermion systems look very similar, they describe dramatically different physics.

### Applications

In our example, the many-body Hamiltonian $\mathcal{H}_0 \equiv q(H)$ can also be written in the following suggestive form:

$$\mathcal{H}_0 = \int \mathrm{d}^3x\, \psi^\dagger(\boldsymbol{x})(H\psi)(\boldsymbol{x}) \qquad [28]$$

and similar formulas hold true for other observables and other Hilbert spaces $h = L^2(\mathcal{M}) \otimes \mathbb{C}^n$. It is rather easy to solve the model defined by such Hamiltonian: all necessary computations can be reduced to one-particle computations. For example, in the static case, where $\boldsymbol{A}$ and $\phi$ are time independent, a main quantity of interest in statistical physics is the free energy

$$\mathcal{E} \equiv -\beta^{-1} \log(\mathrm{tr}(\exp(-\beta[\mathcal{H}_0 - \mu\mathcal{N}]))) \qquad [29]$$

where $\beta > 0$ is the inverse temperature, $\mu$ the chemical potential, and the trace over the Fock space $\mathcal{F}_\mp(h)$. One can show that

$$\mathcal{E} = \pm\mathrm{tr}\big(\beta^{-1}\log(1 \mp \exp(-\beta[H - \mu]))\big) \qquad [30]$$

where the trace is over the one-particle Hilbert space $h$. Thus, to compute $\mathcal{E}$, one only needs to find the eigenvalues of $H$.

It is important to mention that the framework discussed here is not only for external field problems but can be equally well used to formulate and study more complicated models with interparticle interactions. For example, while the model with the Hamiltonian $\mathcal{H}_0$ above is often too simple to describe systems in nature, it is easy to write down more realistic models, for example, the Hamiltonian

$$\mathcal{H} = \mathcal{H}_0 + (e^2/2)\int \mathrm{d}^3\boldsymbol{x}\int \mathrm{d}^3\boldsymbol{y}\,\psi^\dagger(\boldsymbol{x})\psi^\dagger(\boldsymbol{y}) \times |\boldsymbol{x}-\boldsymbol{y}|^{-1}\psi(\boldsymbol{y})\psi(\boldsymbol{x}) \qquad [31]$$

describes electrons in an external electromagnetic field interacting through Coulomb interactions. This illustrates an important point which we would like to stress: the task in quantum theory is twofold, namely to formulate and to solve (exact of otherwise) models. Obviously, in the nonrelativistic case, it is equally simple to formulate many-body models with and without interparticle interactions, and only the latter are simpler because they are easier to solve: the two tasks of formulating and solving models can be clearly separated. As we will see, in the relativistic case, even the formulation of an external field problem is nontrivial, and one finds that one cannot formulate the model without at least partially solving it. This is a common feature of quantum field theories making them challenging and interesting.

## Relativistic Fermion and Boson Systems

We now generalize the formalism developed in the previous section to the relativistic case.

### Field Algebras and Quasifree Representations

In the previous section, we identified the field operators $\psi^{(\dagger)}(f)$ with particular Fock space operators. This is analogous to identifying the operators $p_j = -\mathrm{i}\partial_{x_j}$ and $q_j = x_j$ on $L^2(\mathbb{R}^M)$ with the generators of the Heisenberg algebra, as usually done. (We recall: the Heisenberg algebra is the star algebra generated by $P_j$ and $Q_j$, $j = 1, 2, \ldots, M < \infty$, with the well-known relations

$$[P_j, P_k] = -\mathrm{i}\delta_{jk}, \qquad [P_j, P_k] = [P_j, Q_k] = 0$$
$$P_j^\dagger = P_j, \qquad Q_j^\dagger = Q_j \qquad [32]$$

for all $j$, $k$.) Identifying the Heisenberg algebra with a particular representation is legitimate since, as is well known, all its irreducible representations are (essentially) the same (this statement is made precise by a celebrated theorem due to von Neumann).

However, in case of the algebra generated by the field operators $\psi^{(\dagger)}(f)$, there exist representations which are truly different from the ones discussed in the last section, and such representations are needed to construct relativistic external field problems. It is therefore important to distinguish the fields as generators of an algebra from the operators representing them. We thus define the (boson or fermion) field algebra $\mathcal{A}_\mp(h)$ over a Hilbert space $h$ as the star algebra generated by $\Psi^\dagger(f), f \in h$, such that the map $f \to \Psi(f)$ is linear and the relations

$$[\Psi(f), \Psi^\dagger(g)]_\mp = (f, g)$$
$$[\Psi(f), \Psi(g)]_\mp = 0 \qquad [33]$$
$$\Psi^\dagger(f)^\dagger = \Psi(f)$$

are fulfilled for all $f, g \in h$, with $\dagger$ the star operation in $\mathcal{A}_\mp(h)$. The particular representation of this algebra discussed in the last section will be denoted by $\pi_0$, $\pi_0(\Psi^{(\dagger)}(f)) = \psi^{(\dagger)}(f)$. Other representations $\pi_{P_-}$ can be constructed from any projection operators $P_-$ on $h$, that is, any operator $P_-$ on $h$ satisfying $P_-^* = P_-^2 = P_-$. Writing $\hat{\psi}^{(\dagger)}(f)$ short for $\pi_{P_-}(\Psi^{(\dagger)}(f))$, this so-called quasifree representation is defined by

$$\hat{\psi}^\dagger(f) = \psi^\dagger(P_+f) + \psi(\overline{P_-f})$$
$$\hat{\psi}(f) = \psi(P_+f) \mp \psi^\dagger(\overline{P_-f}) \qquad [34]$$

where the bar means complex conjugation. It is important to note that, while the star operation is identical with the Hilbert space adjoint $*$ in the fermion case, we have

$$\hat{\psi}(f)^\dagger = \psi(Ff)^* \quad \text{with}$$
$$F = P_+ - P_- \quad \text{for bosons} \qquad [35]$$

where $F$ is a grading operator, that is, $F^* = F$ and $F^2 = 1$. We stress that the "physical" star operation always is $*$, that is, physical observables $A$ obey $A = A^*$.

The present framework suggests to regard quantization as the procedure which amounts to going from a one-particle Hilbert space $h$ to the corresponding field algebra $\mathcal{A}_+(h)$. Indeed, the Heisenberg algebra is identical with the boson field algebra $\mathcal{A}_-(\mathbb{C}^M)$ (since the latter is obviously identical with the algebra of $M$ harmonic oscillators), and thus conventional quantum mechanics can be regarded as boson quantization in the special case where the one-particle Hilbert space is finite dimensional. It is interesting to note that "fermion quantum mechanics" $\mathcal{A}_-(\mathbb{C}^M)$ is the natural framework for formulating and studying lattice fermion and spin systems which play an important role in condensed matter physics.

In the following, we elaborate the naive interpretations of the relativistic equations in [2] and [3] as a quantum theory of one particle, and we discuss why they are unphysical. For simplicity, we assume that the electromagnetic fields $\phi$, $\boldsymbol{A}$ are time independent. We then show that quasifree representations as discussed above can provide physically acceptable many-particle theories. We first consider the Dirac case, which is somewhat simpler.

### Fermions

**One-particle formalism** Recalling that $\mathrm{i}\partial_t$ is the energy operator, we define the Dirac Hamiltonian $D$ by rewriting [3] in the following form:

$$\mathrm{i}\partial_t\psi = D\psi, \quad D = (-\mathrm{i}\nabla + e\boldsymbol{A}) \cdot \boldsymbol{\alpha} + m\beta - e\phi \qquad [36]$$

This Dirac Hamiltonian is obviously a self-adjoint operator on the one-particle Hilbert space $h = L^2(\mathbb{R}^4) \otimes \mathbb{C}^4$, but, different from the Schrödinger Hamiltonian in [1], it is not bounded from below: for any $E_0 > -\infty$, one can find a state $f$ such that the energy expectation value $(f, Df)$ is less than $E_0$. This can be easily seen for the simplest case where the external potential vanishes, $\boldsymbol{A} = \phi = 0$. Then the eigenvalues of $D$ can be computed by Fourier transformation, and one finds

$$E = \pm\sqrt{\boldsymbol{p}^2 + m^2}, \quad \boldsymbol{p} \in \mathbb{R}^3 \qquad [37]$$

Due to the negative energy eigenvalues we conclude that there is no ground state, and the Dirac Hamiltonian thus describes an unstable system, which is physically meaningless.

To summarize: a (unphysical) one-particle description of relativistic fermions is given by a Hilbert space $h$ together with a self-adjoint Hamiltonian $D$ unbounded from below. Other observables and transformations are given by self-adjoint and unitary operators on $h$, respectively.

**Many-body formalism** We now explain how to construct a physical many-body description from these data. To simplify notation, we first assume that $D$ has a purely discrete spectrum (which can be achieved by using a compact space). We can then label the eigenfunctions $f_n$ by integers $n$ such that the corresponding eigenvalues $E_n \geq 0$ for $n \geq 0$ and $E_n < 0$ for $n < 0$. Using the naive representation of the fermion field algebra discussed in the last section, we get (we use the notation introduced in [22])

$$q(D) = \sum_{n\geq 0} |E_n|\psi_n^\dagger\psi_n - \sum_{n<0} |E_n|\psi_n^\dagger\psi_n \qquad [38]$$

which is obviously not bounded from below and thus not physically meaningful. However, $\psi_n^\dagger\psi_n = 1 - \psi_n\psi_n^\dagger$, which suggests that we can remedy this problem by interchanging the creation and annihilation operators for $n < 0$. This is possible: it is easy to see that

$$\hat{\psi}_n \equiv \psi_n \quad \forall n \geq 0 \quad \text{and} \quad \hat{\psi}_n \equiv \psi_n^\dagger \quad \forall n < 0 \qquad [39]$$

provides a representation of the algebra in [23]. We thus define

$$\hat{q}(D) \equiv \sum_{n\in\mathbb{Z}} E_n : \hat{\psi}_n^\dagger\hat{\psi}_n : \qquad [40]$$

with the so-called normal ordering prescription

$$:\psi_m^\dagger\psi_n :\equiv \psi_m^\dagger\psi_n - \langle\Omega, \psi_m^\dagger\psi_n\Omega\rangle \qquad [41]$$

where we made use of the freedom of normalization explained after [23] to eliminate unwanted additive constants. We get $q(D) = \sum_{n\in\mathbb{Z}} |E_n|\psi_n^\dagger\psi_n$, which is manifestly a non-negative self-adjoint operator with $\Omega$ as ground state. We thus found a physical many-body description for our model. We can now define for other one-particle observables,

$$\hat{q}(A) \equiv \sum_{n\in\mathbb{Z}} A_{mn} : \hat{\psi}_m^\dagger\hat{\psi}_n : \qquad [42]$$

and, by straightforward computations, we obtain

$$[\hat{q}(A), \hat{q}(B)] = \hat{q}([A, B]) + S(A, B) \qquad [43]$$

where $S(A, B) = \sum_{m<0}\sum_{n\geq 0}(A_{mn}B_{nm} - B_{mn}A_{nm})$, that is,

$$S(A, B) = \mathrm{tr}(P_-AP_+BP_- - P_-BP_+AP_-) \qquad [44]$$

with $P_- = \sum_{n<0} f_n(f_n, \cdot)$ the projection onto the subspace spanned by the negative energy eigenvectors of $D$ and $P_+ = 1 - P_-$. One can show that $\hat{q}(A)$ is no longer defined for all operators but only if

$$P_-AP_+ \text{ and } P_+AP_- \text{ are Hilbert–Schmidt operators} \qquad [45]$$

(we recall that $a$ is a Hilbert–Schmidt operator if $\mathrm{tr}(a^*a) < \infty$). The $\mathbb{C}$-number term $S(A,B)$ in [43] is

often called Schwinger term and, different from the similar term in [25], it is now nontrivial, that is, it is no longer possible to remove it by a redefinition $\hat{q}'(A) = \hat{q}(A) - b(A)$. This Schwinger term is an example of an anomaly, and it has various interesting implications.

In a similar manner, one can construct the many-body transformations $\hat{Q}(U)$ of unitary operators $U$ on $h$ satisfying the very Hilbert–Schmidt condition in [45], and one obtains

$$\hat{Q}(U)\hat{Q}(V) = \chi(U, V)\hat{Q}(UV) \qquad [46]$$

with interesting phase-valued functions $\chi$.

More generally, for any one-particle Hilbert space $h$ and Dirac Hamiltonian $D$, the physical representation is given by the quasifree representation $\pi_{P_-}$ in [34] with $P_-$ the projection onto the negative energy subspace of $D$. The results about $\hat{q}$ and $\hat{Q}$ mentioned hold true in any such representation.

Thus the one-particle Hamiltonian $D$ determines which representation one has to use, and one therefore cannot construct the "physical" representation without specific information about $D$. However, not all these representations are truly different: if there is a unitary operator $\mathcal{U}$ on the Fock space $\mathcal{F}_+(h)$ such that

$$\mathcal{U}^* \pi_{P_-^{(1)}}(\psi^{(\dagger)}(f))\mathcal{U} = \pi_{P_-^{(2)}}(\psi^{(\dagger)}(f)) \qquad [47]$$

for all $f \in h$, then the quasifree representations associated with the different projections $P_-^{(1)}$ and $P_-^{(2)}$ are physically equivalent: one could equally well formulate the second model using the representation of the first. Two such quasifree representations are called unitarily equivalent, and a fundamental theorem due to Shale and Stinespring states that *two quasifree representations* $\pi_{P^{(1,2)}_-}$ *are unitarily equivalent if and only if* $P_-^{(1)} - P_-^{(2)}$ *is a Hilbert–Schmidt operator* (a similar result holds true in the boson case).

## Bosons

**One-particle formalism** Similarly as for the Dirac case, the solutions of the Klein–Gordon equation in [2] also do not define a physically acceptable one-particle quantum theory with a ground state: the energy eigenvalues in [37] for $A = \phi = 0$ are a consequence the relativistic invariance and thus equally true for the Klein–Gordon case. However, in this case there is a further problem. To find the one-particle Hamiltonian, one can rewrite the second-order equation in [2] as a system of first-order equations,

$$\mathrm{i}\partial_t \Phi = K\Phi$$
$$\Phi = \begin{pmatrix} \psi \\ \pi^\dagger \end{pmatrix}, \quad K = \begin{pmatrix} C & \mathrm{i} \\ -\mathrm{i}B^2 & C \end{pmatrix} \qquad [48]$$

with

$$B^2 \equiv (-\mathrm{i}\nabla + e\mathbf{A})^2 + m^2, \qquad C \equiv -e\phi \qquad [49]$$

Thus, one sees that the natural one-particle Hilbert space for the Klein–Gordon equation is $h = L^2(\mathbb{R}^3) \otimes \mathbb{C}^2$; here, and in the following, we identify $h$ with $h_0 \oplus h_0, h_0 = L^2(\mathbb{R}^3)$, and use a convenient $2 \times 2$ matrix notation naturally associated with that splitting. However, the one-particle Hamiltonian is not self-adjoint but rather obeys

$$K^* = JKJ, \quad J \equiv \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \qquad [50]$$

with $*$ the Hilbert space adjoint. It is important to note that $J$ is a grading operator. Thus, we can define a sesquilinear form

$$(f, g)_J \equiv (f, Jg) \quad \forall f, g \in h \qquad [51]$$

with $(\cdot\,, \cdot)$ the standard inner product, and [50] is equivalent to $K$ being self-adjoint with respect to this sesquilinear form; in this case, we say that $K$ is $J$-self-adjoint. Thus, in the Klein–Gordon case, this sesquilinear form takes the role of the Hilbert space inner product and, in particular, not $(\Phi,\Phi)$ but $(\Phi,\Phi)_J$ is preserved under time evolution. However, different from $\Phi^\dagger\Phi$, $\Phi^\dagger J\Phi$ is not positive definite, and it is therefore not possible to interpret it as probability density as in conventional quantum mechanics. For consistency, one has to require that one-particle transformations $U$ are unitary with respect to $(\Phi,\Phi)_J$, that is, $U^{-1} = JUJ$. We call such operators $J$-unitary.

To summarize: a (unphysical) one-particle description of relativistic bosons is given by a Hilbert space of the form $h = h_0 \oplus h_0$, the grading operator $J$ in [50], and a $J$-self-adjoint Hamiltonian $K$ of the form as in eqn [48], where $B \geq 0$ and $C$ are self-adjoint operators on $h_0$. Other observables and transformations are given by $J$-self-adjoint and $J$-unitary operators on $h$, respectively.

**Many-body formalism** We first consider the quasifree representation $\pi_{P_-^{(0)}}$ of the boson field algebra $\mathcal{A}_-(h)$ so that the grading operator in [35] is equal to $J$, that is, $P_-^{(0)} = (1 - J)/2$. Writing $\pi_{P_-^{(0)}}(\Psi^{(\dagger)}(f)) = \psi^{(\dagger)}(f)$, one finds that

$$q(A)^* = q(JAJ), \qquad Q(U)^* = Q(JU^*J) \qquad [52]$$

and thus $J$-self-adjoint operators and $J$-unitary operators are mapped to proper observables and transformations. In particular, $q(K)$ is a self-adjoint

operator, which resolves one problem of the one-particle theory. However, $q(K)$ is not bounded from below, and thus $\pi_{P_{\underline{-}}(0)}$ is not yet the physical representation.

The physical representation can be constructed using the operators

$$T = \frac{1}{\sqrt{2}}\begin{pmatrix} B^{1/2} & \mathrm{i}B^{-1/2} \\ B^{1/2} & \mathrm{i}B^{-1/2} \end{pmatrix}, \quad F = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad [53]$$

(for simplicity, we restrict ourselves to the case $C=0$ and $B > 0$; we use the calculus of self-adjoint operators here) with the following remarkable properties:

$$T^{-1} = JT^{*}F$$
$$TKT^{-1} = \begin{pmatrix} B & 0 \\ 0 & -B \end{pmatrix} \equiv \hat{K} \qquad [54]$$

One can check that

$$\hat{\psi}^{\dagger}(f) \equiv \psi^{\dagger}(Tf), \qquad \hat{\psi}(f) \equiv \psi(T^{-1}f) \qquad [55]$$

is a quasifree representation $\pi_{P_{-}}$ of $\mathcal{A}_{-}(h)$ with $P_{-}=(1-F)/2$. With that the construction of $\hat{q}$ and $\hat{Q}$ is very similar to the fermion case described above (the crucial simplification is that $\hat{K}$ and $F$ now are diagonal). In particular, $\hat{q}(K)$ is a non-negative operator with the ground state $\Omega$, and $\hat{q}(A)$ and $\hat{Q}(U)$ are self-adjoint and unitary for every one-particle observable $A$ and transformation $U$, respectively. One also gets relations as in [43] and [46].

## Related Topics of Recent Interest

The impossibility to construct relativistic quantum-mechanical models played an important role in the early history of quantum field theory, as beautifully discussed in chapter 1 of Weinberg (1995).

The abstract formalism of quasifree representations of fermion and boson field algebras was developed in many papers (see, e.g., Ruijsenaars (1977), Grosse and Langmann (1992), and Langmann (1994) for explicit results on $\hat{Q}$ and $\chi$). A nice textbook presentation with many references can be found in chapter 13 of Gracia-Bondía *et al.* (2001) (this chapter is rather self-contained but mainly restricted to the fermion case).

Based on the Shale–Stinespring theorem, there has been considerable amount of work to investigate whether the quasifree representations associated with different external electromagnetic fields $\psi_1, \boldsymbol{A}_1$ and $\psi_2, \boldsymbol{A}_2$ are unitarily equivalent, if and which time-dependent many-body Hamiltonians exist, etc. (see chapter 13 of Gracia-Bondía *et al.* (2001), and references therein).

The infinite-dimensional Lie algebra $g_2$ of Hilbert space operators satisfying the condition in [45] is an interesting infinite-dimensional Lie algebra with a beautiful representation theory. This subject is closely related to conformal field theory (see, e.g., Kac and Raina (1987) for a textbook presentation and Carey and Ruijsenaars (1987) for a detailed mathematical account within the framework described by us).

It turns out that the mathematical framework discussed in the previous section is sufficient for constructing fully interacting quantum field theories, in particular Yang–Mills gauge theories, in $1+1$ but not in higher dimensions. The reason is that, in $3+1$ dimensions, the one-particle observables $A$ of interest do not obey the Hilbert–Schmidt condition in [45] but only the weaker condition

$$\mathrm{tr}(a^{*}a)^{n} < \infty, \quad a = P_{\mp}AP_{\pm} \qquad [56]$$

with $n=2$, and the natural analog of $g_2$ in $3+1$ dimensions thus seems to be the Lie algebra $g_{2n}$ of operators satisfying this condition with $n=2$. Various results on the representation theory of such Lie algebras $g_{2n>2}$ have been developed (see Mickelsson (1989), where various interesting relations to infinite-dimensional geometry are also discussed).

As mentioned, the Schwinger term $S(A,B)$ in [44] is an example of an anomaly. Mathematically, it is a nontrivial 2-cocycle of the Lie algebra $g_2$, and analogs for the groups $g_{2n>2}$ have been found. These cocycles provide a natural generalization of anomalies (in the meaning of particle physics) to operator algebras. They not only shed some interesting light on the latter, but also provide a link to notions and results from noncommutative geometry (see, e.g., Gracia-Bondía *et al.* (2001)). We believe that this link can provide a fruitful driving force and inspiration to find ways to deepen our understanding of quantum Yang–Mills theories in $3+1$ dimensions (Langmann 1996).

*See also:* Anomalies; $C^{*}$-Algebras and Their Classification; Dirac Fields in Gravitation and Nonabelian Gauge Theory; Dirac Operator and Dirac Field; Gerbes in Quantum Field Theory; Quantum Field Theory in Curved Spacetime; Quantum *n*-Body Problem; Superfluids; Two-Dimensional Models.

## Further Reading


Carey AL and Ruijsenaars SNM (1987) On fermion gauge groups, current algebras and Kac–Moody algebras. *Acta Applicandae Mathematicae* 10: 1–86.

DeWitt B (2003) *The Global Approach to Quantum Field Theory*, International Series of Monographs on Physics, vols. 1 and 2, p. 114. New York: Oxford University Press.

Gracia-Bondía JM, Várilly JC, and Figueroa H (2001) *Elements of Noncommutative Geometry*, Birkhäuser Advanced Texts: Basel Textbooks. Boston: Birkhäuser.

Grosse H and Langmann E (1992) A superversion of quasifree second quantization. *Journal of Mathematical Physics* 33: 1032–1046.

Kac VG and Raina AK (1987) *Bombay Lectures on Highest Weight Representations of Infinite-Dimensional Lie Algebras*,

Advanced Series in Mathematical Physics, vol. 2. Teaneck: World Scientific Publishing.
Langmann E (1994) Cocycles for boson and fermion Bogoliubov transformations. *Journal of Mathematical Physics* 96–112.
Langmann E (1996) Quantum gauge theories and noncommutative geometry. *Acta Physica Polonica B* 27: 2477–2496.
Mickelsson J (1989) *Current Algebras and Groups*, Plenum Monographs in Nonlinear Physics. New York: Plenum Press.
Rafelski J, Fulcher LP, and Klein A (1978) Fermions and bosons interacting with arbitrary strong external fields. *Physics Reports* 38: 227–361.
Reed M and Simon B (1975) *Methods of Modern Mathematical Physics. II. Fourier Analysis, Self-Adjointness*. New York: Academic Press.
Ruijsenaars SNM (1977) On Bogoliubov transformations for systems of relativistic charged particles. *Journal of Mathematical Physics* 18: 517–526.
Weinberg S (1995) *The Quantum Theory of Fields*, vol. I (English summary) Foundations. Cambridge: Cambridge University Press.

# Boundaries for Spacetimes

**S G Harris**, St. Louis University, St. Louis, MO, USA



## Introduction

There is a common practice in mathematics of placing a boundary on an object which may not appear to come naturally equipped with one; this is often thought of as adding ideal points to the object. Perhaps the most famous example is the addition of a single "point at infinity" to the complex plane, resulting in the Riemann sphere: this is a boundary point in the sense of providing an ideal endpoint for lines and other endless curves in the plane. Often, there is more than one reasonable way to construct a boundary for a given object, depending on the intent; for instance, the plane is sometimes equipped, not with a single point at infinity, but with a circle at infinity, resulting in a space homeomorphic to a closed disk. Both these boundaries on the plane have useful but different things to tell us about the nature of the plane; the common feature is that, by bringing the infinite reach of the plane within the confines of a more finite object, we are better able to grasp the behavior of the original object.

The general usefulness of the construction of boundaries for an object is to allow behavior of structures in the "completed" object to aid in visualization of behavior in the original object, such as by providing a degree of measurement or other classification of processes at infinity. This utility has not been overlooked for spacetimes. A variety of purposes may be served by various boundary construction methods: providing a locale for singularities (as the spacetime itself is modeled by a smooth manifold with a smooth metric, free of singular points); providing a platform from which to measure global properties such as total energy or angular momentum; displaying in finite form the causal structure at infinity; or providing a compact (or quasicompact) topological envelope for the spacetime while preserving the causal structure.

This article will consider several of the methods that have been used or proposed for constructing boundaries for spacetimes, ranging from the *ad hoc* (but practical) to the universal. Perhaps the simplest way to classify these methods is into those which employ or analyze embeddings of the spacetime in question and those that do not.

## Boundaries from Embeddings

### General

The simplest and most common method of constructing a boundary for a spacetime $M$ is to find a suitable manifold $N$ (of the same dimension) and an appropriate map $\phi : M \to N$ which is a topological embedding, that is, a homeomorphism onto its image $\phi(M)$. We can consider $\bar{M}_\phi$, the closure of $\phi(M)$ in $N$, as the $\phi$-completion of $M$, and $\partial_\phi(M) = \bar{M}_\phi - \phi(M)$ as the $\phi$-boundary. Typically, this embedding is chosen in such a way that curves of interest in $M$ – such as timelike or null geodesics or causal curves of bounded acceleration – which have no endpoints in $M$, do have endpoints in $\partial_\phi(M)$; in other words, if $c : [0, \infty) \to M$ is such a curve of interest, then $\lim_{t \to \infty} \phi(c(t))$ exists in $N$.

The common practice, initiated by Penrose in 1967, is to choose $N$ to be another spacetime – often called the unphysical spacetime, while $M$ is considered the spacetime of physical interest – and to require the embedding $\phi$ to be a conformal mapping, that is, $\phi$ carries the spacetime metric in $M$ to a scalar multiple of the spacetime metric in $N$. As conformal maps preserve the local causal structure, leaving unchanged the notions of timelike curve or null curve, this means that $\bar{M}_\phi$ inherits from $N$ a causal structure which, locally, is an extension of that of $M$. This allows us to speak of causal relationships within $\bar{M}_\phi$, closely related to those in $M$.

### Minkowski Space

The prototypical example is the conformal embedding of Minkowski space into the Einstein static spacetime.

Let $\mathbb{R}^n$ denote Euclidean $n$-space, $\mathbb{S}^n$ the unit $n$-sphere, and $\mathbb{L}^n$ Minkowski $n$-space, that is, $\mathbb{R}^n$ with metric $ds^2 = dx_1^2 + \cdots + dx_{n-1}^2 - dt^2$ (so $\mathbb{L}^n = \mathbb{R}^{n-1} \times \mathbb{L}^1$). The $n$-dimensional Einstein static spacetime is the product spacetime $\mathbb{E}^n = \mathbb{S}^{n-1} \times \mathbb{L}^1$. Consider $\mathbb{S}^{n-1}$ as embedded in $\mathbb{R}^n = \mathbb{R}^{n-1} \times \mathbb{R}^1$. Then the conformal embedding is $\phi : \mathbb{L}^n \to \mathbb{E}^n$, expressed as $\phi : \mathbb{R}^{n-1} \times \mathbb{L}^1 \to \mathbb{S}^{n-1} \times \mathbb{L}^1 \subset \mathbb{R}^{n-1} \times \ \mathbb{R}^1 \times \mathbb{L}^1$ given by $\phi(x,t) = ((x/|x|) \sin\theta, \cos\theta, \tau)$, where $\theta = \tan^{-1}(t+|x|) - \tan^{-1}(t-|x|)$ and $\tau = \tan^{-1}(t+|x|) + \tan^{-1}(t-|x|)$. The boundary $\partial_\phi(\mathbb{L}^n)$ consists of the following: the points $\{\theta + \tau = \pi; 0 < \tau \le \pi\}$, composed of an $\mathbb{S}^{n-2}$ of null lines coming together at the point $i^+ = (0, 1, \pi)$; a similar cone of null lines $\{\theta - \tau = \pi; -\pi \le \tau < 0\}$ with vertex at $i^- = (0, 1, -\pi)$; and a single limit-point for both cones at $i^0 = (0, -1, 0)$. The $\tau > 0$ null cone is called $\mathfrak{I}^+$ (the letter is read "scri" for "script-I"), its counterpart $\mathfrak{I}^-$ (Figures 1 and 2). As all future-directed timelike geodesics in $\mathbb{L}^n$ have $i^+$ as an endpoint in $\mathbb{E}^n$, $i^+$ is called future-timelike infinity; similarly, $i^-$ is past-timelike infinity. Every future-directed null geodesic ends up on $\mathfrak{I}^+$, which is thus termed future-null infinity, and $\mathfrak{I}^-$ is past-null infinity. All spacelike geodesics come to $i^0$, spacelike infinity.

For $n = 2$, this picture produces the familiar diamond representation of $\mathbb{L}^2$ (Figure 3): as $\mathbb{E}^2$ is easily unrolled into another copy of $\mathbb{L}^2$ (metric

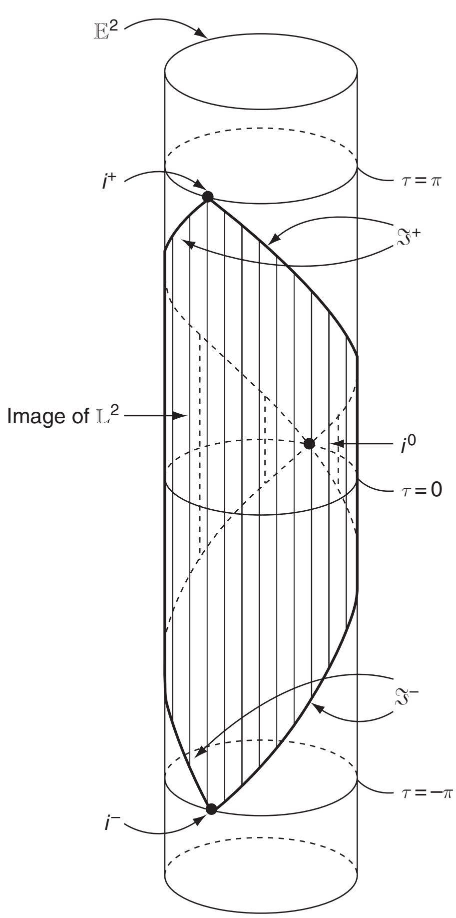



**Figure 1** $\mathbb{L}^2$ conformally embedded in $\mathbb{E}^2 = \mathbb{S}^1 \times \mathbb{L}^1$.

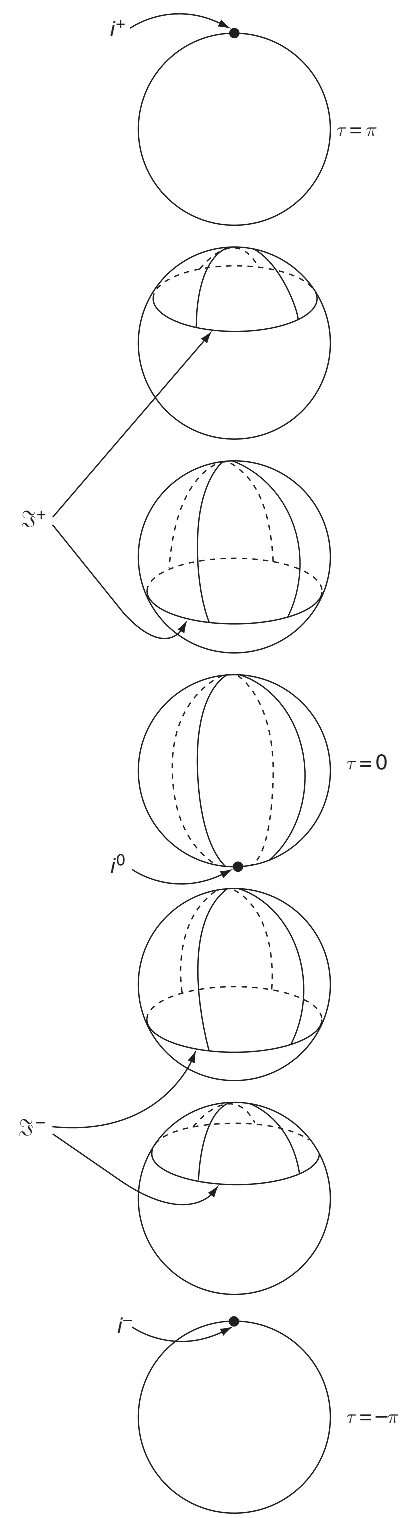



**Figure 2** $\mathbb{L}^3$ conformally embedded in $\mathbb{E}^3 = \mathbb{S}^2 \times \mathbb{L}^1$.

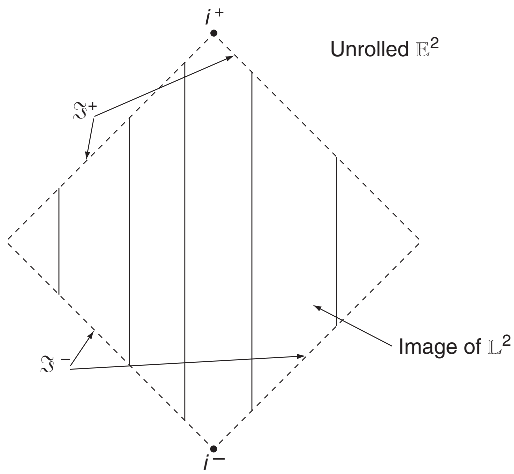



**Figure 3** $\mathbb{L}^2$ conformally embedded in unrolled $\mathbb{E}^2$, i.e., $\mathbb{R}^1 \times \mathbb{L}^1 = \mathbb{L}^2$.

$\mathrm{d}\theta^2 - \mathrm{d}\tau^2$), this means that $\phi(\mathbb{L}^2)$ is the region $|\theta| + |\tau| < \pi$ in $\mathbb{L}^2$; timelike curves and null geodesics in the original $\mathbb{L}^2$ are the same as in $\phi(\mathbb{L}^2)$, and their endpoints in the boundary of the diamond are evident. For higher dimensions, the picture is not as visually obvious, since $\mathbb{E}^n$ cannot be unrolled; but the principle of reading the causal structure at infinity of $\mathbb{L}^n$ via its boundary points in $\mathbb{E}^n$ remains the same.

### Conformal Embeddings

There have been various formulations designed to emulate the conformal mapping of $\mathbb{L}^n$ with respect to spacetimes, which are, in some sense, asymptotically like Minkowski space being conformally mapped into larger spacetimes. A spacetime $M$ with metric $g$ is called asymptotically simple or (alternatively) asymptotically flat if there is a spacetime $N$ with metric $h$, an embedding $\phi: M \to N$, and a scalar function $\Omega$ defined on $N$ with $\phi^* h = (\Omega \circ \phi)^2 g$ (i.e., $\phi$ is conformal with $\Omega^2$ the conformal factor) and $\Omega = 0$ on $\partial_\phi(M), \mathrm{d}\Omega \neq 0$ on $\partial_\phi(M)$, and various other restrictions on $\Omega$, depending on the intent. One can define asymptotic symmetries of $M$ by means of motions within $\partial_\phi(M)$, leading to notions of global energy and angular momentum (see Hawking and Ellis (1973) and Wald (1984) for details).

### Classifications of Embeddings

As a general rule, there is no uniqueness in the choice of an embedding $\phi$ for a spacetime $M$ to construct a boundary, nor in the topology of the resulting boundary $\partial_\phi(M)$, or even of which curves of interest end up having endpoints in the boundary. In an attempt to categorize which embeddings yield equivalent results and what sort of results there are in terms of endpoints of curves, Scott and Szekeres (1994) formulated what they called the abstract boundary of a spacetime. This depends on a choice of class of "interesting" curves, each characterizable as having either infinite or finite parameter length; typical choices for this class would be timelike geodesics or causal geodesics or timelike curves of bounded acceleration. For instance, a boundary point may be said to represent a singularity with respect to the chosen class of curves if it is the endpoint of one such curve with finite parameter length; nonsingular points are points at infinity. These classifications do not require conformal embeddings, nor even that the target of the embeddings be spacetimes; they accommodate boundaries of a far more general type than Penrose's notion stemming from conformal embeddings.

A somewhat different study of boundaries from embeddings has been formulated by García-Parrado and Senovilla (2003), classifying points at infinity and singularities in $\partial_\phi(M)$ for embeddings $\phi: M \to N$ in which $N$ is a spacetime, $\phi$ preserves the chronology relation $\ll$, and there is also a diffeomorphism $\psi: \phi(M) \to N$ which again preserves $\ll$ (the chronology relation in a spacetime is defined thus: $x \ll y$ if and only if there is a future-directed timelike curve from $x$ to $y$). This scheme applies more generally than to conformal embeddings, but the requirement for chronology-preserving maps in both directions guarantees a strong sensitivity to causality; it amounts to a mild extension of Penrose's notion that is often much easier to construct.

## Universal Constructions

### B-Boundary

Attempts have been made to formulate boundary concepts specifically for defining singularities as ideal endpoints for finite-length geodesics. The most complete venture in this direction is the b-boundary ("b" for "bundle") of Schmidt (Hawking and Ellis 1973, pp. 276–284). This is a formulation that takes note only of the connection in the linear frames bundle $L(M)$ of a spacetime $M$ (or of any manifold with a linear connection, metric or otherwise); in other words, it takes no particular note of the spacetime metric or even of the causal structure of the spacetime, but only of the notion of parallel translation of tangent vectors along curves. Parallel translation of a frame (a basis for the tangent space) along a curve is used to obtain an *ad hoc* length for the curve by treating the translated frame as positive-definite orthonormal at each point; whether this length is finite or infinite is independent of the choice of the original frame. The Schmidt construction

defines a boundary on $M$ which gives an endpoint for each curve, endless in $M$, which is finite in that sense: Select a positive-definite metric on $L(M)$, give it a boundary by means of Cauchy completion, and then take the appropriate quotient by the bundle group. This has an appealing universality of application, but the problems of putting it into practice are quite formidable. Also, the fact that it takes no special note of the spacetime character of $M$ suggests that it may not be of particular utility for physical insights.

### Causal Boundary: Basics

In 1972 Geroch, Kronheimer, and Penrose (GKP) formulated a notion of boundary – the causal boundary – that is specifically adapted to the causal character of a spacetime $M$; indeed, it is defined in such a way that one need know only the chronology relation $\ll$ on $M$ without any further reference to the metric (another way of saying this is that the causal boundary is conformally invariant). Like Schmidt's b-boundary, the causal boundary is a universal construction, not depending on any extraneous choices; however, although it has an obvious clarity in its causal structure, there are subtleties in the choice of an appropriate topology which are perhaps not yet fully resolved. As this boundary construction appears to embody the best hopes for a practical universal construction, it is detailed here in some depth.

The causal boundary construction applies only to strongly causal spacetimes; essentially, this means that the local causal structure at each point is exactly reflective of the global causal structure.

The basic construction of the causal boundary of a spacetime $M$ starts with two separate parts: the future and past (pre-)boundaries of $M$, intended as yielding endpoints for, respectively, future- and past-endless causal curves. Part of the difficulty of the causal boundary is knowing how best to meld these two into one; currently, there are several answers to this conundrum.

The elements of the future causal boundary of $M$ are defined in terms of the past-set operator $I^-$. For a point $x \in M$, the past of $x$ is $I^-(x) = \{y \mid y \ll x\}$; for a set $A \subset M, I^-[A] = \bigcup_{x \in A} I^-(x)$. A set $P \subset M$ is called a past set if $I^-[P] = P$; anything of the form $P = I^-[A]$ is a past set, and all past sets have this form. A past set $P$ is an indecomposable past set (IP) if $P$ cannot be written as $P_1 \cup P_2$ for past sets which are proper subsets $P_i \subsetneq P$. IPs come in exactly two varieties: pointlike IPs (PIPs), of the form $I^-(x)$ (Figure 4), and terminal IPs (TIPs), of the form $I^-[c]$ for $c$ a future-endless causal curve (Figure 5). (Of course, any $I^-(x)$ can also be expressed as $I^-[c]$ for $c$ a causal curve ending at $x$.) The future causal boundary of $M$, $\hat{\partial}(M)$, consists of all the TIPs of $M$; the future causal completion of $M$ is $\hat{M} = \hat{\partial}(M) \cup M$. But that is just a set; the causal structure of $M$ needs to be extended to $\hat{M}$.

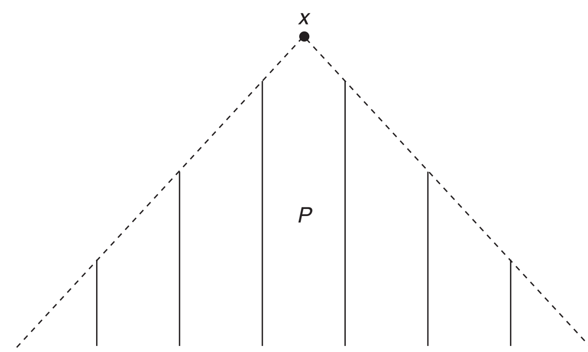



**Figure 4** PIP $P = I^-(x)$.

For any $x \in M$ and $P \in \hat{\partial}(M)$, set $x \ll P$ if and only if $x \in P$; set $P \ll x$ if and only if $P \subset I^-(y)$ for some $y \ll x$ $(y \in M)$; and for $P$ and $Q$ in $\hat{\partial}(M)$, set $P \ll Q$ if and only if $P \subset I^-(y)$ for some $y \in Q$. If we consider this an extension of the $\ll$ relation on $M$, then we end up with a relation which, like that on $M$, is transitive and antireflexive. Furthermore, it has the property that for all $\alpha, \beta \in \hat{M}, \alpha \ll \beta$ if and only if for some $x \in M, \alpha \ll x \ll \beta$. (One can also amend the chronology relation within $M$ to be more like the definition in the extension; that is not of major import.)

We can also extend the causality relation $\prec$ on $M$ to one on $\hat{M}$ (in $M, x \prec y$ if there is a future-directed

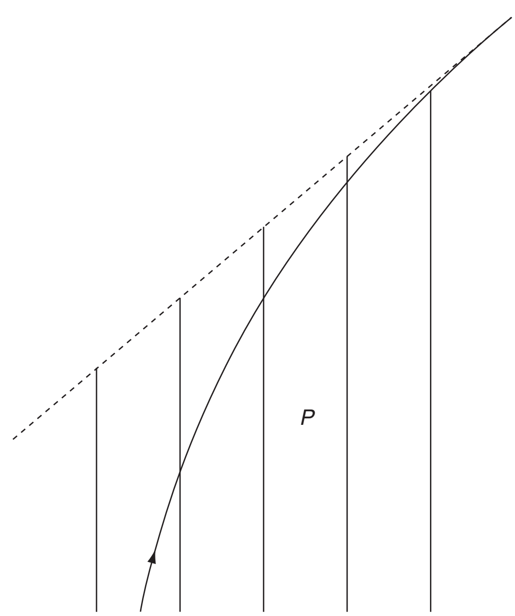



**Figure 5** TIP $P = I^- c$.

causal curve from $x$ to $y$): for $x \in M$ and $P, Q \in \hat{\partial}(M)$, $x \prec P$ for $I^-(x) \subset P$, $P \prec x$ for $P \subset I^-(x)$, and $P \prec Q$ for $P \subset Q$.

The intent is to have the elements of $\hat{\partial}(M)$ provide future endpoints for future-endless causal curves in $M$; in particular, we want two such curves, $c_1$ and $c_2$, to be assigned the same future endpoint precisely when $I^-[c_1] = I^-[c_2]$. This is accomplished by the simple expedient of defining the future endpoint of a future-endless causal curve $c$ to be $P = I^-[c]$. We do not have a topology on $\hat{M}$ as yet, but it is worth noting that if $P$ is the assigned future endpoint of $c$, then $I^-(P) = I^-[c]$; this is at least the correct causal behavior for a putative future endpoint of $c$.

We can perform all the operations above in the time-dual manner, obtaining the past causal boundary $\check{\partial}(M)$, consisting of terminal indecomposable future sets (TIFs), and the past causal completion $\check{M} = \check{\partial}(M) \cup M$. The full causal boundary of $M$ consists of the union of $\hat{\partial}(M)$ with $\check{\partial}(M)$ with some sort of identifications to be made.

As an example of the need for identifications, consider $M$ to be $\mathbb{L}^2$ with a closed timelike line segment deleted, say $M = \mathbb{L}^2 - \{(0, t) \mid 0 \le t \le 1\}$. For $\hat{\partial}(M)$, we have first the boundary elements at infinity: the TIP $i^+ = M$ (the past of the positive time axis) and the set of TIPs making up $\Im^+$ (the pasts of null lines going out to infinity in $\mathbb{L}^2$); and then, the boundary elements coming from the deleted points: for each $t$ with $0 < t \le 1$, two IPs emanating from $(0, t)$, that is, $P_t^+$, the past of the null line going pastwards from $(0, t)$ toward $x > 0$, and $P_t^-$, the past of the null line going pastwards from $(0, t)$ toward $x < 0$; and $P_0$, emanating from $(0, 0)$, that is, the past of the negative time axis. Similarly, $\check{\partial}(M)$ consists of $i^-, \Im^-$, TIFs $F_t^+$ and $F_t^-$ emanating from $(0, t)$ for $0 \le t < 1$, and the TIF $F_1$ emanating from $(0, 1)$. We probably want to make at least the following identifications for each $t$ with $0 < t < 1$, $P_t^+ \equiv F_t^+$ and $P_t^- \equiv F_t^-; P_1^+ \equiv F_1 \equiv P_1^-$; and $F_0^+ \equiv P_0 \equiv F_0^-$. This results in a two-sided replacement for the deleted segment; for some purposes, it might be deemed desirable to identify the two sides as one, but a universal boundary is probably a good idea, leaving further identifications as optional quotients of the universal object.

How best to define the appropriate identifications in general is a matter of some controversy. GKP defined a somewhat complicated topology on $\bar{M} = \hat{\partial}(M) \cup \check{\partial}(M) \cup M$, then used an identification intended to result in a Hausdorff space. There are significant problems with this approach in some *outré* spacetimes, as pointed out by Budic and Sachs (1974) and Szabados (1989), both of whom recommended a different set of identifications. But what is of more concern is that the topology prescribed by GKP is not what might be expected in even the simplest of cases, for example, Minkowski space: $\overline{\mathbb{L}^n}$ needs no identifications among boundary points (no matter whose identification procedure is followed). The GKP topology on $\overline{\mathbb{L}^n}$, restricted to $\hat{\partial}(\mathbb{L}^n)$, is not that of a cone ($\mathbb{S}^{n-2} \times \mathbb{R}^1$ with a point added), as is the case for $\Im^+$ in the conformal embedding into $\mathbb{E}^n$; but, instead, each null line in $\hat{\partial}(\mathbb{L}^n)$ (not including $i^+$) is an open set, and $i^+$ has no neighborhood in $\hat{\partial}(\mathbb{L}^n)$ save for the entire boundary. This is a topology bearing no relation at all to that of any embedding.

### Future Causal Boundary

**Construction** An alternative approach, initiated by Harris (1998), is to forego the full causal boundary and concentrate only on $\hat{M}$ and $\check{M}$ separately. There is an advantage to this in that the process of future causal completion – that is to say, forming $\hat{M}$ from $M$ – can be made functorial in an appropriate category of "chronological sets": a set $X$ with a relation $\ll$ which is transitive and antireflexive such that it possesses a countable subset $S$ which is "chronologically dense," that is, for any $x, y \in X$, there is some $s \in S$ with $x \ll s \ll y$. Any strongly causal spacetime $M$ is a chronological set, as is $\hat{M}$. The entire construction of the future causal boundary works just as well for a chronological set. The role of a timelike curve in a chronological set is taken by a future chain: a sequence $c = \{x_n\}$ with $x_n \ll x_{n+1}$ for all $n$. For any future chain $c$, $I^-[c]$ is an IP, and any IP can be so expressed; but unlike in spacetimes, $I^-(x)$ may or may not be an IP for $x \in X$. Then, $\hat{X}$ is always future complete in the sense that for any future chain $c$ in $\hat{X}$, there is an element $\alpha \in \hat{X}$ with $I^-(\alpha) = I^-[c]$: for instance, if the chain $c$ lies in $X$ but there is no $x \in X$ with $I^-(x) = I^-[c]$, just let $\alpha = I^-[c]$, which is an element of $\hat{\partial}(X)$. This yields a functor of future completion from the category of chronological sets to the category of future-complete chronological sets, and the embedding $X \to \hat{X}$ is a universal object in the sense of the category theory; this implies that it is categorically unique and is the minimal future-completion process.

However, it is crucial to have more than the chronology relation operating in what is to be a boundary; topology of some sort is needed. This is accomplished by defining what might be called the future-chronological topology for any chronological set – including for $\hat{M}$ when $M$ is a strongly causal spacetime. This topology is defined by means of a limit-operator $\hat{L}$ on sequences: if $X$ is the chronological set, then for any sequence of points $\sigma = \{x_n\}$ in $X, \hat{L}(\sigma)$ denotes a subset of $X$ which is the set of

limits of $\sigma$. It is explicitly recognized that there may be more than one limit of a sequence, as the space may not be Hausdorff; no attempt is made to remove any non-Hausdorffness, as this is viewed as giving important information on how, possibly, two points in the future causal boundary represent very similar and yet not identical pieces of information about the causal structure at infinity. Once the limit operator is in place, the actual topology on $X$ is defined thus: a subset $A \subset X$ is said to be closed if and only if for any sequence $\sigma \subset A, \hat{L}(\sigma) \subset A$ (and open sets are complements of closed sets). This yields the elements of $\hat{L}(\sigma)$ as topological limits of $\sigma$.

The definition of $\hat{L}$ is simplest when $X$ has the property that $I^-(x)$ is an IP for any $x \in X$; as this is true for $X$ being either a spacetime $M$ or the future causal completion $\hat{M}$ of a spacetime, the discussion here is restricted to this situation. Let us also make the common assumption that $X$ is past-distinguishing, that is, $I^-(x) = I^-(y)$ implies $x = y$.

Let $\sigma = \{x_n\}$ be a sequence of points in a past-distinguishing chronological set $X$ in which the past of any point is an IP. Then $\hat{L}(\sigma)$ consists of those points $x$ for which (see Figures 6 and 7)

1. for all $y \in I^-(x)$, for $n$ sufficiently large, $y \ll x_n$, and
2. for any IP $P \supsetneq I^-(x)$, there is some $z \in P$ such that for $n$ sufficiently large, $z \not\ll x_n$.

Then the future-chronological topology on $X$ has these features:

1. It is a $T_1$ topology, that is, points are closed.
2. If $I^-(x) = I^-[c]$ for a future chain $c = \{x_n\}$, then $x$ is a topological limit of the sequence $\{x_n\}$.
3. If $X = M$, a strongly causal spacetime, then the future-chronological topology is precisely the manifold topology.
4. If $X = \hat{M}$, the future causal completion of a strongly causal spacetime $M$, then the induced topology on $M$ is the manifold topology, $\hat{\partial}(M)$ is a closed subset of $\hat{M}$, and $M$ is dense in $\hat{M}$. As per property (2), for any future-endless causal curve $c$ in $M$, the point $I^-[c]$ in $\hat{\partial}(M)$ is the topological endpoint of $c$ in $\hat{M}$.
5. If $X = \widehat{\mathbb{L}^n}$, then $X$ is homeomorphic to the conformal image of $\mathbb{L}^n$ in $\mathbb{E}^n$ together with $\Im^+$ and $i^+$; in particular, $\hat{\partial}(\mathbb{L}_n)$ has the topology of a cone.

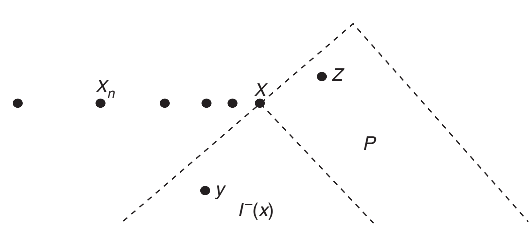



**Figure 6** $x \in \hat{L}(\{x_n\})$: for all $y \in I^-(x)$, eventually $y \ll x_n$, and for all IP $P \supsetneq I^-(x)$, there is some $z \in P$ such that eventually $z \not\ll x_n$.

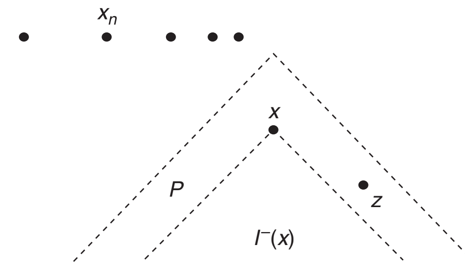



**Figure 7** $x \notin \hat{L}(\{x_n\})$: there is some IP $P \supsetneq I^-(x)$ such that for all $z \in P$, $z \ll x_n$ for infinitely many $n$.

**Examples** The future causal boundary with the future-chronological topology can be calculated with a fair degree of success. For instance, if $M$ is conformal to a simple product spacetime $Q \times \mathbb{L}^1$ ($Q$ a Riemannian manifold), then $\hat{\partial}(M)$ is much like $\hat{\partial}(\mathbb{L}^n)$ in that it consists of null or timelike lines factored over a particular boundary construction $\partial(Q)$ on $Q$, coming together at a single point $i^+$ (the IP which is all of $M$); if $Q$ is complete, then these are all null lines, and together they may be called $\Im^+$.

The elements of $\partial(Q)$ are defined in terms of the Lipschitz-1 functions on $Q$ known as Busemann functions: if $c : [\alpha, \omega) \to Q$ is any endless unit-speed curve (typically, $\omega = \infty$), then the Busemann function $b_c : Q \to \mathbb{R}$ is defined by $b_c(q) = \lim_{s \to \omega} (s - d(c(s), q))$, where $d$ is the distance function in $Q$; this function is either finite for all $q$ or infinite for all $q$. The set $B(Q)$ of finite Busemann functions has an $\mathbb{R}$-action defined by $a \cdot b_c = b_{a \cdot c}$, where $(a \cdot c)(s) = c(s + a)$. Then $\partial(Q) = B(Q)/\mathbb{R}$. For any $P \in \hat{\partial}(M)$, the boundary of $P$, as a subset of $Q \times \mathbb{L}^1 \cong Q \times \mathbb{R}$, is the graph of a Busemann function (the function is $b_c$ for $P$ generated by a null curve projecting to $c$); and a point $x = (q, t)$ in $M$ can be represented by $\partial(I^-(x))$, which is the graph of the function $t - d(-, q)$. Thus, one could use the function-space topology on $B(Q)$ to topologize $\hat{M}$; in that function-space topology $\hat{\partial}(M)$ is a cone on $\partial(Q)$, and $\hat{M}$, apart from $i^+$, is the topological product of $\mathbb{R}$ with $Q \cup \partial(Q)$. The future-chronological topology is sometimes different from the function-space topology, allowing more convergent sequences than the function-space topology does. When this happens, the result is non-Hausdorff, revealing pairs of points in $\hat{\partial}(M)$ which are more closely related to one another than the function-space

topology reveals; but it is still the case that $\hat{\partial}(M)$, apart from $i^+$, is fibered by $\mathbb{R}$ over $\partial(Q)$.

If $Q$ is a warped product $Q = (a, b) \times K$ for a compact manifold $K$ with metric $\mathrm{d}r^2 + \mathrm{e}^{\phi(r)}h$ with $h$ a metric on $K$, then one can calculate more precisely: if, for instance, $\phi$ has a minimum in the interior of $(a, b)$ and has suitable growth on either end, then $\partial(Q)$ represents two copies of $K$ (one for each end of $(a, b) \times K$), the future-chronological topology is the same as the function-space topology, and $\hat{M}$ (apart from $i^+$) is a simple product of $\mathbb{R}$ with $Q \cup \partial(Q)$: $\hat{\partial}(M)$ is precisely a null cone over two copies of $K$. This applies, for instance, to exterior Schwarzschild, where $K = \mathbb{S}^2$; the boundary at one end of exterior Schwarzschild is the usual $\Im^+$, and the boundary at the other end is the null cone $\{r = 2m\}$, where exterior attaches to interior Schwarzschild.

Calculations for the future-chronological topology become much easier when $\hat{\partial}(M)$ is purely spacelike, that is, no $P \in \hat{\partial}(M)$ is contained in the past of any other element of $\hat{M}$. For instance, if $M$ is conformal to a multiwarped product, $Q_1 \times \cdots \times Q_m \times (a, b)$ with metric $f_1(t)^2 h_1 + \cdots + f_m(t)^2 h_m - \mathrm{d}t^2$, where $h_i$ is a Riemannian metric on $Q_i$, then $\hat{\partial}(M)$ will be purely spacelike if all the Riemannian factors are complete and for each $i$, $\int_{b^-}^{b} 1/f_i(t)\,\mathrm{d}t < \infty$; in that case, $\hat{\partial}(M) \cong Q$, where $Q = Q_1 \times \cdots \times Q_m$ and $\hat{M} \cong Q \times (a, b)$. This applies, for instance, to interior Schwarzschild, where $Q_1 = \mathbb{R}^1$ and $Q_2 = \mathbb{S}^2$, yielding the topology of $\mathbb{R}^1 \times \mathbb{S}^2$ for the Schwarzschild singularity.

There is a categorical universality for spacelike boundaries and the future-chronological topology. This means that any other reasonable way of future-completing interior Schwarzschild must yield $\mathbb{R}^1 \times \mathbb{S}^2$ or a topological quotient of that for the singularity; and if the result is to be past-distinguishing, $\mathbb{R}^1 \times \mathbb{S}^2$ is the only possibility.

Of course, all this can be done in the time-dual fashion, using the past-chronological topology on $\check{M}$. It would be desirable to combine the future and past causal boundaries with a suitable topology as well as appropriate identifications. There has been some work in that direction.

### Causal Boundary: Revisited

Marolf and Ross (2003) have proposed an identification of TIPs and TIFs that relies on the equivalence relation defined by Szabados. For an IP $P$ and IF $F$, call $(P, F)$ a Szabados pair if $P \subset I^-(x)$ for all $x \in F$, $P$ is maximal among IPs for that property, and dually for $F$ with respect to $P$. For instance, for any $x \in M$, $(I^-(x), I^+(x))$ is a Szabados pair. The Marolf–Ross version of the causal boundary, $\bar{\partial}(M)$, consists of all Szabados pairs formed of TIPs and TIFs, plus any TIP or TIF that cannot be paired; this produces an appropriate set of identifications within $\hat{\partial}(M) \cup \check{\partial}(M)$. The chronology relation on $M$ is extended to $\bar{M} = \bar{\partial}(M) \cup M$ by treating each point $x$ in $M$ as the Szabados pair $(I^-(x), I^+(x))$ and each unpaired IP $P$ as $(P, \emptyset)$ and unpaired IF $F$ as $(\emptyset, F)$, and then defining $(P, F) \ll (P', F')$ whenever $F \cap P' \neq \emptyset$.

The resulting chronological set is not necessarily either past- or future-distinguishing, but it is (past and future)-distinguishing. The topology they propose places endpoints in $\bar{\partial}(M)$ for all causal curves which are endless in $M$, but there may be multiple future endpoints for a single future-endless curve. The topology need not be $\mathrm{T}_1$: points can fail to be closed. For a product spacetime $M = Q \times \mathbb{L}^1$, the Marolf–Ross topology on $\bar{M}$ is always the function-space topology.

As of this writing, there is active research by J L Flores to institute a Marolf–Ross type of identification of $\hat{\partial}(M)$ with $\check{\partial}(M)$ using a topology that partakes more of the future- and past-chronological topologies.

*See also:* Asymptotic Structure and Conformal Infinity; Spacetime Topology, Causal Structure and Singularities.

## Further Reading

Budic R and Sachs RK (1974) Causal boundaries for general relativistic space-times. *Journal of Mathematical Physics* 15: 1302–1309.

García-Parrado A and Senovilla JMM (2003) Causal relationship: a new tool for the causal characterization of Lorentzian manifolds. *Classical and Quantum Gravity* 20: 625–664.

Geroch RP, Kronheimer EH, and Penrose R (1972) Ideal points in space-time. *Proceedings of the Royal Society of London, Series A* 327: 545–567.

Harris SG (1998) Universality of the future chronological boundary. *Journal of Mathematical Physics* 39: 5427–5445.

Harris SG (2000) Topology of the future chronological boundary: universality for spacelike boundaries. *Classical and Quantum Gravity* 17: 551–603.

Harris SG (2001) Causal boundary for standard static spacetimes. *Nonlinear Analysis* 47: 2971–2981 (Special Edition: Proceedings of the Third World Congress in Nonlinear Analysis).

Harris SG (2004a) Boundaries on spacetimes: an outline. *Classical and Quantum Gravity* 359: 65–85.

Harris SG (2004b) Discrete group actions on spacetimes: causality conditions and the causal boundary. *Classical and Quantum Gravity* 21: 1209–1236.

Harris SG and Dray T (1990) The causal boundary of the trousers space. *Classical and Quantum Gravity* 7: 149–161.

Hawking SW and Ellis GFR (1973) *The Large Scale Structure of Space-Time*. Cambridge: Cambridge University Press.

Marolf D and Ross SF (2003) A new recipe for causal completions. *Classical and Quantum Gravity* 20: 4085–4118.

Schmidt BG (1972) Local completeness of the *b*-boundary. *Communications in Mathematical Physics* 29: 49–54.

Scott SM and Szekeres P (1994) The abstract boundary – a new approach to singularities of manifolds. *Journal of Geometry and Physics* 13: 223–253.

Szabados LB (1988) Causal boundary for strongly causal spacetimes. *Classical and Quantum Gravity* 5: 121–134.
Szabados LB (1989) Causal boundary for strongly causal spacetimes: II. *Classical and Quantum Gravity* 6: 77–91.
Wald RM (1984) *General Relativity*. Chicago: University of Chicago Press.

# Boundary Conformal Field Theory

**J Cardy**, Rudolf Peierls Centre for Theoretical Physics, Oxford, UK



Boundary conformal field theory (BCFT) is simply the study of conformal field theory (CFT) in domains with a boundary. It gains its significance [1] because, in some ways, it is mathematically simpler: the algebraic and geometric structures of CFT appear in a more straightforward manner; and [2] because it has important applications: in string theory in the physics of open strings and D-branes, and in condensed matter physics in boundary critical behavior and quantum impurity models.

This article, however, describes the basic ideas from the point of view of quantum field theory, without regard to particular applications or to any deeper mathematical formulations.

## Review of CFT

### Stress Tensor and Ward Identities

Two-dimensional CFTs are massless, local, relativistic renormalized quantum field theories. Usually they are considered in imaginary time, that is, on two-dimensional manifolds with Euclidean signature. In this article, the metric is also taken to be Euclidean, although the formulation of CFTs on general Riemann surfaces is also of great interest, especially for string theory. For the time being, the domain is the entire complex plane.

Heuristically, the correlation functions of such a field theory may be thought of as being given by the Euclidean path integral, that is, as expectation values of products of local densities with respect to a Gibbs measure $Z^{-1}\,e^{-S_E(\{\psi\})}[d\psi]$, where the $\{\psi(x)\}$ are some set of fundamental local fields, $S_E$ is the Euclidean action, and the normalization factor $Z$ is the partition function. Of course, such an object is not in general well defined, and this picture should be seen only as a guide to formulating the basic principles of CFT which can then be developed into a mathematically consistent theory.

In two dimensions, it is useful to use the so-called complex coordinates $z = x^1 + ix^2, \bar{z} = x^1 - ix^2$. In CFT, there are local densities $\phi_j(z, \bar{z})$, called primary fields, whose correlation functions transform covariantly under conformal mappings $z \to z' = f(z)$:

$$\langle \phi_1(z_1, \bar{z}_1)\phi_2(z_2, \bar{z}_2)\cdots\rangle = \prod_i f'(z_j)^{h_j}\bar{f}'(z_j)^{\bar{h}_j}\langle \phi_1(z'_1, \bar{z}'_1)\phi_2(z'_2, \bar{z}'_2)\cdots\rangle \qquad [1]$$

where $(h_j, \bar{h}_j)$ (usually real numbers, not complex conjugates of each other) are called the conformal weights of $\phi_j$. These local fields can in general be normalized so that their two-point functions have the form

$$\langle \phi_j(z_j, \bar{z}_j)\phi_k(z_k, \bar{z}_k)\rangle = \delta_{jk}/(z_j - z_k)^{2h_j}(\bar{z}_j - \bar{z}_k)^{2\bar{h}_j} \qquad [2]$$

They satisfy an algebra known as the operator product expansion (OPE)

$$\phi_i(z_1, \bar{z}_1)\cdot\phi_j(z_2, \bar{z}_2) = \sum_k c_{ijk}(z_1 - z_2)^{-h_i - h_j + h_k} \times (\bar{z}_1 - \bar{z}_2)^{-\bar{h}_i - \bar{h}_j + \bar{h}_k}\phi_k(z_1, \bar{z}_1) + \cdots \qquad [3]$$

which is supposed to be valid when inserted into higher-order correlation functions in the limit when $|z_1 - z_2|$ is much less than the separations of all the other points. The ellipses denote the contributions of other nonprimary scaling fields to be described below. The structure constants $c_{ijk}$, along with the conformal weights, characterize the particular CFT.

An essential role is played by the energy–momentum tensor, or, in Euclidean field theory language, the stress tensor $T^{\mu\nu}$. Heuristically, it is defined as the response of the partition function to a local change in the metric:

$$T^{\mu\nu}(x) = -(2\pi)\,\delta \ln Z/\delta g_{\mu\nu}(x) \qquad [4]$$

(the factor of $2\pi$ is included so that similar factors disappear in later equations).

The symmetry of the theory under translations and rotations implies that $T^{\mu\nu}$ is conserved, $\partial_\mu T^{\mu\nu} = 0$, and symmetric. Scale invariance implies that it is also traceless $\Theta \equiv T^\mu_\mu = 0$. It should be noted that the vanishing of the trace of the stress tensor for a scale invariant classical field theory does

not usually survive when quantum corrections are taken into account: indeed, $\Theta \propto \beta(g)$, the renormalization group (RG) beta-function. A quantum field theory is thus only a CFT when this vanishes, that is, at an RG fixed point. In complex coordinates, the components $T_{z\bar{z}} = T_{\bar{z}z} = 4\Theta$ vanish, while the conservation equations read

$$\partial_{\bar{z}} T_{zz} = \partial_z T_{\bar{z}\bar{z}} = 0 \qquad [5]$$

Thus, correlators of $T(z) \equiv T_{zz}$ are locally analytic (in fact, globally meromorphic) functions of $z$, while those of $\bar{T}(\bar{z}) \equiv T_{\bar{z}\bar{z}}$ are antianalytic. It is this property of analyticity which makes CFTs tractable in two dimensions.

Since an infinitesimal conformal transformation $z \to z + \alpha(z)$ induces a change in the metric, its effect on a correlation function of primary fields, given by [1], may also be expressed through an appropriate integral involving an insertion of the stress tensor. This leads to the conformal Ward identity:

$$\int_C \langle T(z) \prod_j \phi_j(z_j, \bar{z}_j) \rangle \, \alpha(z) \, dz = \sum_j \left(h_j \alpha'(z_j) + \alpha(z_j)(\partial/\partial z_j)\right) \left\langle \prod_j \phi_j(z_j, \bar{z}_j) \right\rangle \qquad [6]$$

where $C$ is a contour encircling all the points $\{z_j\}$. (A similar equation holds for the insertion of $\bar{T}$.) Using Cauchy's theorem, this determines the first few terms in the OPE of $T$ with any primary density:

$$T(z) \cdot \phi_j(z_j, \bar{z}_j) = \frac{h_j}{(z - z_j)^2} \phi(z_j, \bar{z}_j) + \frac{1}{z - z_j} \partial_{z_j} \phi(z_j, \bar{z}_j) + O(1) \qquad [7]$$

The other, regular, terms in the OPE generate new scaling fields, which are not in general primary, called descendants. One way of defining a density to be primary is by the condition that the most singular term in its OPE with $T$ is a double pole.

The OPE of $T$ with itself has the form

$$T(z) \cdot T(z_1) = \frac{c/2}{(z - z_1)^4} + \frac{2}{(z - z_1)^2} T(z_1) + \cdots \qquad [8]$$

The first term is present because $\langle T(z) T(z_1) \rangle$ is nonvanishing, and must take the form shown, with $c$ being some number (which cannot be scaled to unity, since the normalization of $T$ is fixed by its definition) which is a property of the CFT. It is known as the conformal anomaly number or the central charge. This term implies that $T$ is not itself primary. In fact, under a finite conformal transformation $z \to z' = f(z)$,

$$T(z) \to f'(z)^2 T(z') + \frac{c}{12} \{z', z\} \qquad [9]$$

where $\{z', z\} = (f'''f' - \frac{3}{2} f''2)/f'^2$ is the Schwartzian derivative.

### Virasoro Algebra

As with any quantum field theory, the local fields can be realized as linear operators acting on a Hilbert space. In ordinary QFT, it is customary to quantize on a constant-time hypersurface. The generator of infinitesimal time translations is the Hamiltonian $\hat{H}$, which itself is independent of which time slice is chosen, because of time translational symmetry. It is also given by the integral over the hypersurface of the time–time component of the stress tensor. In CFT, because of scale invariance, one may instead quantize on fixed circle of a given radius. The analog of the Hamiltonian is the dilatation operator $\hat{D}$, which generates scale transformations. Unlike $\hat{H}$, the spectrum of $\hat{D}$ is usually discrete, even in an infinite system. It may also be expressed as an integral over the radial component of the stress tensor:

$$\begin{aligned} \hat{D} &= \frac{1}{2\pi} \int_0^{2\pi} r \, \hat{T}_{rr} \, r d\theta \\ &= \frac{1}{2\pi i} \int_C z \hat{T}(z) dz - \frac{1}{2\pi i} \int_C \bar{z} \hat{\bar{T}}(\bar{z}) d\bar{z} \\ &\equiv \hat{L}_0 + \hat{\bar{L}}_0 \end{aligned} \qquad [10]$$

where, because of analyticity, $C$ can be any contour encircling the origin.

This suggests that one define other operators

$$\hat{L}_n \equiv \frac{1}{2\pi} \int_C z^{n+1} \hat{T}(z) dz \qquad [11]$$

and similarly the $\hat{\bar{L}}_n$. From the OPE [8] then follows the Virasoro algebra $\mathcal{V}$:

$$[\hat{L}_n, \hat{L}_m] = (n - m) \hat{L}_{n+m} + \frac{c}{12} n(n^2 - 1) \delta_{n+m,0} \qquad [12]$$

with an isomorphic algebra $\bar{\mathcal{V}}$ generated by the $\hat{\bar{L}}_n$.

In radial quantization, there is a vacuum state $|0\rangle$. Acting on this with the operator corresponding to a scaling field gives a state $|\phi_j\rangle \equiv \hat{\phi}_j(0,0)|0\rangle$ which is an eigenstate of $\hat{D}$: in fact,

$$\hat{L}_0 |\phi_j\rangle = h_j |\phi_j\rangle, \quad \hat{\bar{L}}_0 |\phi_j\rangle = \bar{h}_j |\phi_j\rangle \qquad [13]$$

From the OPE [7], one sees that $|L_n \phi_j\rangle \propto \hat{L}_n |\phi_j\rangle$, and, if $\phi_j$ is primary, $\hat{L}_n |\phi_j\rangle = 0$ for all $n \geq 1$.

The states corresponding to a given primary field, and those generated by acting on these with all the $\hat{L}_n$ with $n < 0$ an arbitrary number of times, form a

highest-weight representation of $\mathcal{V}$. However, this is not necessarily irreducible. There may be null vectors, which are linear combinations of states at a given level which are themselves annihilated by all the $\hat{L}_n$ with $n > 0$. They exist whenever $h$ takes a value from the Kac table:

$$h = h_{r,s} = \frac{(r(m+1) - sm)^2 - 1}{4m(m+1)} \qquad [14]$$

with the central charge parametrized as $c = 1 - 6/(m(m+1))$, and $r$, $s$ are non-negative integers. These null states should be projected out, giving an irreducible representation $\mathcal{V}_h$.

The full Hilbert space of the CFT is then

$$\mathcal{H} = \bigoplus_{h,\bar{h}} n_{h,\bar{h}} \mathcal{V}_h \otimes \bar{\mathcal{V}}_{\bar{h}} \qquad [15]$$

where the non-negative integers $n_{h,\bar{h}}$ specify how many distinct primary fields of weights $(h, \bar{h})$ there are in the CFT.

The consistency of the OPE [3] with the existence of null vectors leads to the fusion algebra of the CFT. This applies separately to the holomorphic and antiholomorphic sectors, and determines how many copies of $\mathcal{V}_c$ occur in the fusion of $\mathcal{V}_a$ and $\mathcal{V}_b$:

$$\mathcal{V}_a \odot \mathcal{V}_b = \sum_c N_{ab}^c \mathcal{V}_c \qquad [16]$$

where the $N_{ab}^c$ are non-negative integers.

A particularly important subset of all CFTs consists of the minimal models. These have rational central charge $c = 1 - 6(p-q)^2/pq$, in which case the fusion algebra closes with a finite number of possible values $1 \leq r \leq q, 1 \leq s \leq p$ in the Kac formula [14]. For these models, the fusion algebra takes the form

$$\mathcal{V}_{r_1,s_1} \odot \mathcal{V}_{r_2,s_2} = \sum_{r=|r_1-r_2|}^{r_1+r_2-1\,\prime} \sum_{s=|s_1-s_2|}^{s_1+s_2-1\,\prime} \mathcal{V}_{r,s} \qquad [17]$$

where the prime on the sums indicates that they are to be restricted to the allowed intervals of $r$ and $s$.

There is an important theorem which states that the only unitary CFTs with $c < 1$ are the minimal models with $p/q = (m+1)/m$, where $m$ is an integer $\geq 3$.

## Modular Invariance

The fusion algebra limits which values of $(h, \bar{h})$ might appear in a consistent CFT, but not which ones actually occur, that is, the values of the $n_{h,\bar{h}}$. This is answered by the requirement of modular invariance on the torus. First consider the theory on an infinitely long cylinder, of unit circumference. This is related to the (punctured) plane by the conformal mapping $z \to (1/2\pi) \ln z \equiv t + ix$. The result is a QFT on the circle $0 \leq x < 1$, in imaginary time $t$. The generator of infinitesimal time translations is related to that for dilatations in the plane:

$$\hat{H} = 2\pi\hat{D} - \frac{\pi c}{6} = 2\pi(\hat{L}_0 + \hat{\bar{L}}_0) - \frac{\pi c}{6} \qquad [18]$$

where the last term comes from the Schwartzian derivative in [9]. Similarly, the generator of translations in $x$, the total momentum operator, is $\hat{P} = 2\pi(\hat{L}_0 - \hat{\bar{L}}_0)$.

A general torus is, up to a scale transformation, a parallelogram with vertices $(0, 1, \tau, 1+\tau)$ in the complex plane, with the opposite edges identified. We can make this by taking a cylinder of unit circumference and length Im, $\tau$, twisting the ends by a relative amount Re $\tau$, and sewing them together. This means that the partition function of the CFT on the torus can be written as

$$Z(\tau, \bar{\tau}) = \mathrm{tr}\, e^{-(\mathrm{Im}\,\tau)\hat{H} + i(\mathrm{Im}\,\tau)\hat{P}} = \mathrm{tr}\, q^{\hat{L}_0 - c/24} \bar{q}^{\hat{\bar{L}}_0 - c/24} \qquad [19]$$

using the above expressions for $\hat{H}$ and $\hat{P}$ and introducing $q \equiv e^{2\pi i \tau}$.

Through the decomposition [15] of $\mathcal{H}$, the trace sum can be written as

$$Z(\tau, \bar{\tau}) = \sum_{h,\bar{h}} n_{h,\bar{h}} \chi_h(q)\, \chi_{\bar{h}}(q) \qquad [20]$$

where

$$\chi_h(q) \equiv \mathrm{tr}_{\mathcal{V}_h} q^{\hat{L}_0 - c/24} = \sum_N d_h(N) q^{h-(c/24)+N} \qquad [21]$$

is the character of the representation of highest weight $h$, which counts the degeneracy $d_h(N)$ at level $N$. It is purely an algebraic property of the Virasoro algebra, and its explicit form is known in many cases.

All of this would be less interesting were it not for the observation that the parametrization of the torus through $\tau$ is not unique. In fact, the transformations $S : \tau \to -1/\tau$ and $T : \tau \to \tau + 1$ give the same torus (see Figure 1). Together, these

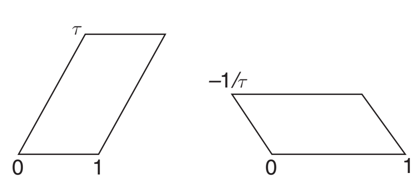



**Figure 1** Two equivalent parametrizations of the same torus.

operations generate the modular group SL(2, Z), and the partition function $Z(\tau, \bar{\tau})$ should be invariant under them. $T$-invariance is simply implemented by requiring that $h - \bar{h}$ is an integer, but the $S$-invariance of the right-hand side of [20] places highly nontrivial constraints on the $n_{h,\bar{h}}$. That this can be satisfied at all relies on the remarkable property of the characters that they transform linearly under $S$:

$$\chi_h(\mathrm{e}^{-2\pi\mathrm{i}/\tau}) = \sum_{h'} S_h^{h'} \chi_{h'}(\mathrm{e}^{2\pi\mathrm{i}\tau}) \qquad [22]$$

This follows from applying the Poisson sum formula to the explicit expressions for the characters, which are related to Jacobi theta-functions. In many cases (e.g., the minimal models) this representation is finite dimensional, and the matrix $S$ is symmetric and orthogonal. This means that one can immediately obtain a modular invariant partition function by forming the diagonal sum

$$Z = \sum_h \chi_h(q)\chi_h(\bar{q}) \qquad [23]$$

so that $n_{h,\bar{h}} = \delta_{h\bar{h}}$. However, because of various symmetries of the characters, other modular invariants are possible: for the minimal models (and some others) these have been classified. Because of an analogy of the results with the classification of semisimple Lie algebras, the diagonal invariants are called the A-series.

## Boundary CFT

In any field theory in a domain with a boundary, one needs to consider how to impose a set of consistent boundary conditions. Since CFT is formulated independently of a particular set of fundamental fields and a Lagrangian, this must be done in a more general manner. A natural requirement is that the off-diagonal component $T_{\parallel\perp}$ of the stress tensor parallel/perpendicular to the boundary should vanish. This is called the conformal boundary condition. If the boundary is parallel to the time axis, it implies that there is no momentum flow across the boundary. Moreover, it can be argued that, under the RG, any uniform boundary condition will flow into a conformally invariant one. For a given bulk CFT, however, there may be many possible distinct such boundary conditions, and it is one task of BCFT to classify these.

To begin with, take the domain to be the upper-half plane, so that the boundary is the real axis. The conformal boundary condition then implies that $T(z) = \bar{T}(\bar{z})$ when $z$ is on the real axis. This has the immediate consequence that correlators of $\bar{T}$ are those of $T$, analytically continued into the lower-half plane. The conformal Ward identity, cf. [7], now reads

$$\left\langle T(z) \prod_j \phi_j(z_j, \bar{z}_j) \right\rangle = \sum_j \left( \frac{h_j}{(z - z_j)^2} + \frac{1}{z - z_j}\partial_{z_j} + \frac{\bar{h}_j}{(\bar{z} - \bar{z}_j)^2} + \frac{1}{\bar{z} - \bar{z}_j}\partial_{\bar{z}_j} \right) \left\langle \prod_j \phi_j(z_j, \bar{z}_j) \right\rangle \qquad [24]$$

In radial quantization, in order that the Hilbert spaces defined on different hypersurfaces be equivalent, one must choose semicircles centered on some point on the boundary, conventionally the origin. The dilatation operator is now

$$\hat{D} = \frac{1}{2\pi\mathrm{i}} \int_S z\hat{T}(z)\mathrm{d}z - \frac{1}{2\pi\mathrm{i}} \int_S \bar{z}\hat{\bar{T}}(\bar{z})\,\mathrm{d}\bar{z} \qquad [25]$$

where $S$ is a semicircle. Using the conformal boundary condition, this can also be written as

$$\hat{D} = \hat{L}_0 = \frac{1}{2\pi\mathrm{i}} \int_C z\hat{T}(z)\,\mathrm{d}z \qquad [26]$$

where $C$ is a complete circle around the origin. As before, one may similarly define the $\hat{L}_n$, and they satisfy a Virasoro algebra.

Note that there is now only one Virasoro algebra. This is related to the fact that conformal mappings which preserve the real axis correspond to real analytic functions. The eigenstates of $\hat{L}_0$ correspond to boundary operators $\hat{\phi}_j(0)$ acting on the vacuum state $|0\rangle$. It is well known that in a renormalizable QFT operators at the boundary require a different renormalization from those in the bulk, and this will in general lead to a different set of conformal weights. It is one of the tasks of BCFT to determine these, for a given allowed boundary condition.

However, there is one feature unique to boundary CFT in two dimensions. Radial quantization also makes sense, leading to the same form [26] for the dilation operator, if the boundary conditions on the negative and positive real axes are different. As far as the structure of BCFT goes, correlation functions with this mixed boundary condition behave as though a local scaling field were inserted at the origin. This has led to the term "boundary condition changing (bcc) operator," but it must be stressed that these are not local operators in the conventional sense.

## The Annulus Partition Function

Just as consideration of the partition function on the torus illuminates the bulk operator content $n_{h,\bar{h}}$, it

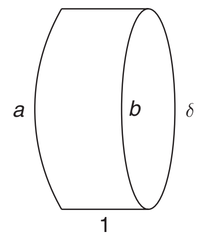



**Figure 2** The annulus, with boundary conditions $a$ and $b$ on either boundary.

turns out that consistency on the annulus helps classify both the allowed boundary conditions, and the boundary operator content. To this end, consider a CFT in an annulus formed of a rectangle of unit width and height $\delta$, with the top and bottom edges identified (see Figure 2). The boundary conditions on the left and right edges, labeled by $a, b, \ldots$, may be different. The partition function with boundary conditions $a$ and $b$ on either edge is denoted by $Z_{ab}(\delta)$.

One way to compute this is by first considering the CFT on an infinitely long strip of unit width. This is conformally related to the upper-half plane (with an insertion of bcc operators at 0 and $\infty$ if $a \neq b$) by the mapping $z \to (1/\pi)\ln z$. The generator of infinitesimal translations along the strip is

$$\hat{H}_{ab} = \pi\hat{D} - \pi c/24 = \pi\hat{L}_0 - \pi c/24 \qquad [27]$$

Thus, for the annulus,

$$Z_{ab}(\delta) = \operatorname{tr} e^{-\delta \hat{H}_{ab}} = \operatorname{tr} q^{\hat{L}_0 - \pi c/24} \qquad [28]$$

with $q \equiv e^{-\pi\delta}$. As before, this can be decomposed into characters:

$$Z_{ab}(\delta) = \sum_h n_h^{ab} \chi_h(q) \qquad [29]$$

but note that now the expression is linear. The nonnegative integers $n_{ab}^h$ give the operator content with the boundary conditions $(ab)$: the lowest value of $h$ with $n_{ab}^h > 0$ gives the conformal weight of the bcc operator, and the others give conformal weights of the other allowed primary fields which may also sit at this point.

On the other hand, the annulus partition function may be viewed, up to an overall rescaling, as the path integral for a CFT on a circle of unit circumference, being propagated for (imaginary) time $\delta^{-1}$. From this point of view, the partition function is no longer a trace, but rather the matrix element of $e^{-\hat{H}/\delta}$ between boundary states:

$$Z_{ab}(\delta) = \langle a | e^{-\hat{H}/\delta} | b \rangle \qquad [30]$$

Note that $\hat{H}$ is the same Hamiltonian that appears in [18], and the boundary states lie in $\mathcal{H}$, [15].

How are these boundary states to be characterized? Using the transformation law [9] the conformal boundary condition applied to the circle implies that $L_n = \bar{L}_{-n}$. This means that any boundary state $|B\rangle$ lies in the subspace satisfying

$$\hat{L}_n |B\rangle = \hat{\bar{L}}_{-n} |B\rangle \qquad [31]$$

Moreover, because of the decomposition [15] of $\mathcal{H}$, $|B\rangle$ is also some linear superposition of states from $\mathcal{V}_h \otimes \bar{\mathcal{V}}_{\bar{h}}$. This condition can therefore be applied in each subspace. Taking $n = 0$ in [31] constrains $\bar{h} = h$. For simplicity, consider only the diagonal CFTs with $n_{h,\bar{h}} = \delta_{h,\bar{h}}$. It can then be shown that the solution of [31] is unique and has the following form. The subspace at level $N$ of $\mathcal{V}_h$ has dimension $d_h(N)$. Denote an orthonormal basis by $|h, N; j\rangle$, with $1 \le j \le d_h(N)$, and the same basis for $\bar{\mathcal{V}}_h$ by $\overline{|h, N; j\rangle}$. The solution to [31] in this subspace is then

$$|h\rangle\rangle \equiv \sum_{N=0}^{\infty} \sum_{j=1}^{d_h(N)} |h, N; j\rangle \otimes \overline{|h, N; j\rangle} \qquad [32]$$

These are called Ishibashi states. Matrix elements of the translation operator along the cylinder between them are simple:

$$\begin{aligned} &\langle\langle h' | e^{-\hat{H}/\delta} | h \rangle\rangle \\ &= \sum_{N'=0}^{\infty} \sum_{j'=1}^{d_{h'}(N')} \sum_{N=0}^{\infty} \sum_{j=1}^{d_h(N)} \langle h', N'; j'| \\ &\quad \otimes \overline{\langle h', N'; j'|} e^{-(2\pi/\delta)(\hat{L}_0 + \hat{\bar{L}}_0 - c/12)} \end{aligned} \qquad [33]$$

$$\begin{aligned} &|h, N; j\rangle \otimes \overline{|h, N; j\rangle} \\ &= \delta_{h'h} \sum_{N=0}^{\infty} \sum_{j=1}^{d_h(N)} e^{-(4\pi/\delta)(h + N - (c/24))} \end{aligned} \qquad [34]$$

$$= \delta_{h'h} \chi_h(e^{-4\pi/\delta}) \qquad [35]$$

Note that the characters which appear are related to those in [29] by the modular transformation $S$.

The physical boundary states satisfying [29], sometimes called the Cardy states, are linear combinations of the Ishibashi states:

$$|a\rangle = \sum_h \langle\langle h | a \rangle |h\rangle\rangle \qquad [36]$$

Equating the two different expressions [29] and [30] for $Z_{ab}$, and using the modular transformation law

[22] and the linear independence of the characters gives the (equivalent) conditions:

$$n_{ab}^h = \sum_{h'} S_{h'}^h \langle a|h'\rangle\rangle\langle\langle h'|b\rangle \qquad [37]$$

$$\langle a|h'\rangle\rangle\langle\langle h'|b\rangle = \sum_h S_h^{h'} n_{ab}^h \qquad [38]$$

These are called the Cardy conditions. The requirements that the right-hand side of [37] should give a non-negative integer, and that the right-hand side of [38] should factorize in $a$ and $b$, give highly nontrivial constraints on the allowed boundary states and their operator content.

For the diagonal CFTs considered here (and for the nondiagonal minimal models) a complete solution is possible. It can be shown that the elements $S_0^h$ of $S$ are all non-negative, so one may choose $\langle\langle h|\tilde{0}\rangle = (S_0^h)^{1/2}$. This defines a boundary state

$$|\tilde{0}\rangle \equiv \sum_h \left(S_0^h\right)^{1/2} |h\rangle\rangle \qquad [39]$$

and a corresponding boundary condition such that $n_{00}^h = \delta_{h0}$. Then, for each $h' \neq 0$, one may define a boundary state

$$\langle\langle h|\tilde{h'}\rangle \equiv S_{h'}^h / (S_0^h)^{1/2} \qquad [40]$$

From [37], this gives $n_{h'0}^h = \delta_{h'h}$. For each allowed $h'$ in the torus partition function, there is therefore a boundary state $|\tilde{h'}\rangle$ satisfying the Cardy conditions. However, there is a further requirement:

$$n_{h'h''}^h = \frac{S_{h'}^h S_{h''}^h}{S_0^h} \qquad [41]$$

should be a non-negative integer. Remarkably, this combination of elements of $S$ occurs in the Verlinde formula, which follows from considering consistency of the CFT on the torus. This states that the right-hand side of [41] is equal to the fusion algebra coefficient $N_{h'h''}^h$. Since these are non-negative integers, the consistency of the above ansatz for the boundary states is consistent.

We conclude that, at least for the diagonal models, there is a bijection between the allowed primary fields in the bulk CFT and the allowed conformally invariant boundary conditions. For the minimal models, with a finite number of such primary fields, this correspondence has been followed through explicitly.

**Example** The simplest example is the diagonal $c = \frac{1}{2}$ unitary CFT corresponding to $m = 3$. The allowed values of the conformal weights are $h = 0, \frac{1}{2}, \frac{1}{16}$, and

$$S = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix} \qquad [42]$$

from which one finds the allowed boundary states

$$|\tilde{0}\rangle = \frac{1}{\sqrt{2}}|0\rangle\rangle + \frac{1}{\sqrt{2}}\left|\frac{1}{2}\right\rangle\right\rangle + \frac{1}{2^{1/4}}\left|\frac{1}{16}\right\rangle\right\rangle \qquad [43]$$

$$\left|\tilde{\frac{1}{2}}\right\rangle = \frac{1}{\sqrt{2}}|0\rangle\rangle + \frac{1}{\sqrt{2}}\left|\frac{1}{2}\right\rangle\right\rangle - \frac{1}{2^{1/4}}\left|\frac{1}{16}\right\rangle\right\rangle \qquad [44]$$

$$\left|\tilde{\frac{1}{16}}\right\rangle = |0\rangle\rangle - \left|\frac{1}{2}\right\rangle\right\rangle \qquad [45]$$

The nontrivial part of the fusion algebra of this CFT is

$$\mathcal{V}_{\frac{1}{16}} \odot \mathcal{V}_{\frac{1}{16}} = \mathcal{V}_0 + \mathcal{V}_{\frac{1}{2}} \qquad [46]$$

$$\mathcal{V}_{\frac{1}{16}} \odot \mathcal{V}_{\frac{1}{2}} = \mathcal{V}_{\frac{1}{16}} \qquad [47]$$

$$\mathcal{V}_{\frac{1}{2}} \odot \mathcal{V}_{\frac{1}{2}} = \mathcal{V}_0 \qquad [48]$$

from which can be read off the boundary operator content

$$n_{\tilde{h}}^h = 1 \quad n_{\tilde{\frac{1}{16}}\tilde{\frac{1}{16}}}^0 = n_{\tilde{\frac{1}{16}}\tilde{\frac{1}{16}}}^{\frac{1}{2}} = n_{\tilde{\frac{1}{16}}\tilde{\frac{1}{16}}}^{\frac{1}{2}} = n_{\tilde{\frac{1}{2}}\tilde{\frac{1}{16}}}^{\frac{1}{16}} = 1 \qquad [49]$$

The $c = \frac{1}{2}$ CFT is known to describe the continuum limit of the critical Ising model, in which spins $s = \pm 1$ are localized on the sites of a regular lattice. The above boundary conditions may be interpreted as the continuum limit of the lattice boundary conditions $s = 1$, free and $s = -1$, respectively. Note there is a symmetry of the fusion rules which means that one could equally well have inverted the ordering of this correspondence.

## Other Topics

### Boundary Entropy

The partition function on annulus of length $L$ and circumference $\beta$ can be thought of as the quantum statistical mechanics partition function for a one-dimensional QFT in an interval of length $L$, at temperature $\beta^{-1}$. It is interesting to consider this in the thermodynamic limit when $\delta = L/\beta$ is large. In that case, only the ground state of $\hat{H}$ contributes in [30], giving

$$Z_{ab}(L, \beta) \sim \langle a|0\rangle\langle 0|b\rangle e^{\pi cL/6\beta} \qquad [50]$$

from which the free energy $F_{ab} = -\beta^{-1} \ln Z_{ab}$ and the entropy $\mathcal{S}_{ab} = -\beta^2(\partial F_{ab}/\partial\beta)$ can be obtained. The result is

$$\mathcal{S}_{ab} = (\pi c/3\beta)L + s_a + s_b + o(1) \qquad [51]$$

where the first term is the usual extensive contribution. The other two pieces $s_a \equiv \ln(\langle a|0\rangle)$ and $s_b \equiv \ln(\langle b|0\rangle)$ may be identified as the boundary entropy associated with the corresponding boundary states. A similar definition may be made in massive QFTs. It is an unproven but well-verified conjecture that the boundary entropy is a nonincreasing function along boundary RG flows, and is stationary only for conformal boundary states.

### Bulk–Boundary OPE

The boundary Ward identity [24] has the implication that, from the point of view of the dependence of its correlators on $z_j$ and $\bar{z}_j$, a primary field $\phi_j(z_j, \bar{z}_j)$ may be thought of as the product of two local fields which are holomorphic functions of $z_j$ and $\bar{z}_j$, respectively. These will satisfy OPEs as $|z_j - \bar{z}_j| \to 0$, with the appearance of primary fields on the right-hand side being governed by the fusion rules. These fields are localized on the real axis: they are the boundary operators. There is therefore a kind of bulk–boundary OPE:

$$\phi_j(z_j, \bar{z}_j) = \sum_k d_{jk}(\mathrm{Im}\, z_j)^{-h_j - \bar{h}_j + h_k} \phi_k^b(\mathrm{Re}\, z_j) \qquad [52]$$

where the sum on the right-hand side is, in principle, over all the boundary fields consistent with the boundary condition, and the coefficients $d_{jk}$ are analogous to the OPE coefficients in the bulk. As before, they are nonvanishing only if allowed by the fusion algebra: a boundary field of conformal weight $h_k$ is allowed only if $N^{h_k}_{h_j \bar{h}_j} > 0$.

For example, in the $c = \frac{1}{2}$ CFT, the bulk operator with $h = \bar{h} = \frac{1}{16}$ goes over into the boundary operator with $h = 0$, or that with $h = \frac{1}{2}$, depending on the boundary condition. The bulk operator with $h = \bar{h} = \frac{1}{2}$, however, can only go over into the identity boundary operator with $h = 0$ (or a descendent thereof.)

The fusion rules also apply to the boundary operators themselves. The consistency of these with bulk–boundary and bulk–bulk fusion rules, as well as the modular properties of partition functions, was examined by Lewellen.

### Extended Algebras

CFTs may contain other conserved currents apart from the stress tensor, which generate algebras (Kac–Moody, superconformal, W-algebras) which extend the Virasoro algebra. In BCFT, in addition to the conformal boundary condition, it is possible (but not necessary) to impose further boundary conditions relating the holomorphic and antiholomorphic parts of the other currents on the boundary. It is believed that all rational CFTs can be obtained from Kac–Moody algebras via the coset construction. The classification of boundary conditions from this point of view is fruitful and also important for applications, but is beyond the scope of this article.

### Stochastic Loewner Evolution

In recent years, there has emerged a deep connection between BCFT and conformally invariant measures on curves in the plane which start at a boundary of a domain. These arise naturally in the continuum limit of certain statistical mechanics models. The measure is constructed dynamically as the curve is extended, using a sequence of random conformal mappings called stochastic Loewner evolution (SLE). In CFT, the point where the curve begins can be viewed as the insertion of a boundary operator. The requirement that certain quantities should be conserved in mean under the stochastic process is then equivalent to this operator having a null state at level two. Many of the standard results of CFT correspond to an equivalent property of SLE.

## Acknowledgments

This article was written while the author was a member of the Institute for Advanced Study. He thanks the School of Mathematics and the School of Natural Sciences for their hospitality. The work was supported by the Ellentuck Fund.

*See also:* Affine Quantum Groups; Eight Vertex and Hard Hexagon Models; Indefinite Metric; Operator Product Expansion in Quantum Field Theory; Quantum Phase Transitions; Stochastic Loewner Evolutions; String Field Theory; Superstring Theories; Symmetries in Quantum Field Theory: Algebraic Aspects; Two-Dimensional Conformal Field Theory and Vertex Operator Algebras.

## Further Reading

Affleck I (1997) Boundary condition changing operators in conformal field theory and condensed matter physics. *Nuclear Physics B Proceedings Supplement* 58: 35.

Cardy J (1984) Conformal invariance and surface critical behavior. *Nuclear Physics B* 240: 514–532.

Cardy J (1989) Boundary conditions, fusion rules and the Verlinde formula. *Nuclear Physics B* 324: 581.

di Francesco P, Mathieu P, and Senechal D (1999) *Conformal Field Theory*. New York: Springer.

Kager W and Nienhuis B (2004) A guide to stochastic Loewner evolution and its applications. *Journal of Statistical Physics* 115: 1149.

Lawler G (2005) *Conformally Invariant Processes in the Plane*. American Mathematical Society.

Lewellen DC (1992) Sewing constraints for conformal field theories on surfaces with boundaries. *Nuclear Physics B* 372: 654.

Petkova V and Zuber JB *Conformal Boundary Conditions and What They Teach Us*, Lectures given at the Summer School and Conference on Nonperturbative Quantum Field Theoretic

Methods and their Applications, August 2000, Budapest, Hungary, hep-th/0103007.
Verlinde E (1988) Fusion rules and modular transformations in 2D conformal field theory. *Nuclear Physics B* 300: 360.
Werner W *Random Planar Curves and Schramm–Loewner Evolutions*, Springer Lecture Notes (to appear), math.PR/0303354.

# Boundary Control Method and Inverse Problems of Wave Propagation

**M I Belishev**, Petersburg Department of Steklov Institute of Mathematics, St. Petersburg, Russia



## Introduction

Inverse problems are generally positioned as the problems of determination of a system (its structure, parameters, etc.) from its "input → output" correspondence.

The boundary-value inverse problems deal with systems which describe processes (wave, heat, electromagnetic ones, etc.) occurring in media occupying a spatial domain. The process is initiated by a boundary source (input) and is described by a solution of a certain partial differential equation in the domain. Certain additional information about the solution, which can be extracted from measurements on the boundary, plays the role of the output. The objective is to determine the parameters of the medium – in particular, the coefficients in the equation – from this information.

The boundary control (BC) method (Belishev 1986) is an approach to the boundary-value inverse problems based on their links with the control theory and system theory. The present article is a version of the BC method which solves the problem of reconstruction of a Riemannian manifold from its boundary spectral or dynamical data.

## Forward Problems

### Manifold

Let $(\Omega, d)$ be a smooth compact Riemannian manifold with the boundary $\Gamma$, $\dim \Omega \geq 2$; $d$ is the distance determined by the metric tensor $g$. For $A \subset \Omega$ denote

$$\langle A \rangle^r := \{x \in \Omega \,|\, d(x, A) \leq r\}, \quad r \geq 0$$

the hypersurfaces $\Gamma^T := \{x \in \Omega \,|\, d(x, \Gamma) = T\}$, $T > 0$ are equidistant to $\Gamma$. In terms of the dynamics of the system, the value

$$T_* := \min\{T > 0 \,|\, \langle \Gamma \rangle^T = \Omega\} = \max_{\Omega} d(\cdot\,, \Gamma)$$

means the time needed for waves, moving from $\Gamma$ with the unit speed, to fill $\Omega$.

A point $x \in \Omega$ is said to belong to the set $c_0 \subset \Omega$ if $x$ is connected with $\Gamma$ via more than one shortest geodesic. The set $c := \bar{c}_0$ is called the separation set (cut locus) of $\Omega$ with respect to $\Gamma$. It is a closed set of zero volume. Let $\tau_*(\gamma)$ be the length of the geodesic emanating from $\gamma \in \Gamma$ orthogonally to $\Gamma$ and connecting $\gamma$ with $c$. The function $\tau_*(\cdot)$ is continuous on $\Gamma$.

For $x \in \Omega \setminus c$ the pair $(\gamma, \tau)$, such that $\tau = d(x, \Gamma) = d(x, \gamma)$, constitutes the semigeodesic coordinates of $x$. The set of these coordinates

$$\Theta := \{(\gamma, \tau) \,|\, \gamma \in \Gamma, 0 \leq \tau < \tau_*(\gamma)\} \subset \Gamma \times [0, T_*]$$

is called the pattern of $\Omega$. Pictorially, to get the pattern, one needs to slit $\Omega$ along $c$ and then pull it on the cylinder $\Gamma \times [0, T_*]$. The part $\Theta^T := \Theta \cap (\Gamma \times [0, T])$ of the pattern consists of the semigeodesic coordinates of the points $x \in \langle \Gamma \rangle^T \setminus c$ (**Figure 1**).

### Dynamical System

Propagation of waves in the manifold is described by a dynamical system $\alpha^T$ of the form

$$u_{tt} - \Delta_g u = h \qquad \text{in } \Omega \times (0, T) \qquad [1]$$

$$u\,|_{t=0} = u_t\,|_{t=0} = 0 \quad \text{in } \Omega \qquad [2]$$

$$u = f \qquad \text{on } \Gamma \times [0, T] \qquad [3]$$

where $\Delta_g$ is the Beltrami–Laplace operator, $0 < T \leq \infty$, $f$ and $h$ are the boundary and volume sources (controls), $u = u^{f,h}(x,t)$ is the solution (wave).

Set $\mathcal{H} := L_2(\Omega)$; the spaces of the controls are

$$\mathcal{F}^T := L_2(\Gamma \times [0, T]), \quad \mathcal{G}^T := L_2([0, T]; \mathcal{H})$$

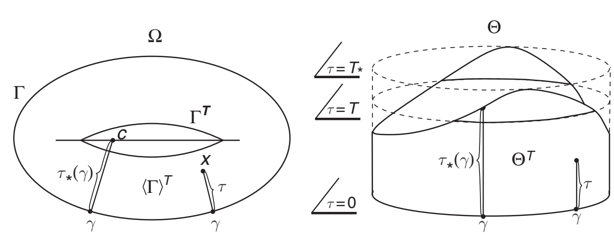



**Figure 1** Manifold and pattern. (Data from Belishev (1997).)

The "input $\mapsto$ state" map of the system $\alpha^T$ is realized by the control operator $W^T$:

$$\mathcal{F}^T \times \mathcal{G}^T \to \mathcal{H}, \quad W^T\{f, h\} := u^{f,h}(\cdot, T)$$

and its parts

$$W^T_{\mathrm{bd}} : \mathcal{F}^T \to \mathcal{H}, \quad W^T_{\mathrm{vol}} : \mathcal{G}^T \to \mathcal{H}$$
$$W^T_{\mathrm{bd}} f := u^{f,0}(\cdot, T), W^T_{\mathrm{vol}} h := u^{0,h}(\cdot, T)$$

In the case $f = 0$ the evolution of the system is governed by the operator $L := -\Delta_g$ defined on the Sobolev class $H^2(\Omega) \cap H^1_0(\Omega)$ of functions vanishing on $\Gamma$, and the semigroup representation

$$u^{0,h}(\cdot, r) = W^r_{\mathrm{vol}} h = \int_0^r L^{-1/2} \sin\left[(r-t)L^{1/2}\right] h(\cdot, t)\mathrm{d}t \quad [4]$$

holds for all $r \geq 0$.

The "input $\mapsto$ output" map is implemented by the response operator $R^T: \mathcal{F}^T \to \mathcal{F}^T$,

$$R^T f := \partial_\nu u^{f,0} \quad \text{on } \Gamma \times [0, T]$$

defined on controls $f \in H^1(\Gamma \times [0, T])$ vanishing on $\Gamma \times \{t = 0\}$; here $\nu = \nu(\gamma)$ is the outward normal to $\Gamma$. The normal derivative $\partial_\nu u^{f,0}$ describes the forces appearing on $\Gamma$ as a result of interaction of the wave with the boundary.

The map $C^T : \mathcal{F}^T \to \mathcal{F}^T, C^T := (W^T_{\mathrm{bd}})^* W^T_{\mathrm{bd}}$, which is called the connecting operator, can be represented via the response operator of the system $\alpha^{2T}$:

$$C^T = \tfrac{1}{2}(S^T)^* R^{2T} J^{2T} S^T \quad [5]$$

$S^T: \mathcal{F}^T \to \mathcal{F}^{2T}$ being the extension of controls from $\Gamma \times [0, T]$ onto $\Gamma \times [0, 2T]$ as odd functions of $t$ with respect to $t = T$, and $J^{2T}: \mathcal{F}^{2T} \to \mathcal{F}^{2T}$ being the integration

$$(J^{2T} f)(\cdot, t) = \int_0^t f(\cdot, s)\mathrm{d}s$$

### Controllability

Open subsets $\sigma \subset \Gamma$ and $\omega \subset \Omega$ determine the subspaces

$$\mathcal{F}^T_\sigma := \{f \in \mathcal{F}^T \mid \mathrm{supp}\, f \subset \overline{\sigma} \times [0, T]\}$$
$$\mathcal{G}^T_\omega := \{h \in \mathcal{G}^T \mid \mathrm{supp}\, h \subset \overline{\omega} \times [0, T]\}$$

of controls acting from $\sigma$ and $\omega$, respectively. In view of hyperbolicity of the problem [1]–[3], the relation

$$\mathrm{supp}\, u^{f,h}(\cdot, t) \subset \langle\overline{\sigma}\rangle^t \cup \langle\overline{\omega}\rangle^t, \quad t \geq 0 \quad [6]$$

holds for $f \in \mathcal{F}^T_\sigma$ and $h \in \mathcal{G}^T_\omega$. This means that the waves propagate in $\Omega$ with the speed $= 1$.

The sets of waves

$$\mathcal{U}^T_\sigma := W^T_{\mathrm{bd}} \mathcal{F}^T_\sigma, \qquad \mathcal{U}^T_\omega := W^T_{\mathrm{vol}} \mathcal{G}^T_\omega$$

are said to be reachable at time $t = T$ from $\sigma$ and $\omega$, respectively. Denoting

$$\mathcal{H}A := \{y \in \mathcal{H} \mid \mathrm{supp}\ y \subset \overline{A}\}$$

by virtue of [6] one has the embeddings $\mathcal{U}^T_\sigma \subset \mathcal{H}\langle\overline{\sigma}\rangle^T$ and $\mathcal{U}^T_\omega \subset \mathcal{H}\langle\overline{\omega}\rangle^T$. The property of the system $\alpha^T$ that plays the key role in inverse problems is that these embeddings are dense:

$$\mathrm{cl}\,\mathcal{U}^T_\sigma = \mathcal{H}\langle\overline{\sigma}\rangle^T, \qquad \mathrm{cl}\,\mathcal{U}^T_\omega = \mathcal{H}\langle\overline{\omega}\rangle^T \quad [7]$$

for any $T > 0$ (cl denotes the closure in $\mathcal{H}$).

In control theory, relations [7] are interpreted as an approximate controllability of the system in subdomains filled with waves; the name "BC method" is derived from the first one (boundary controllability). This property means that the sets of waves are rich enough: any function supported in the subdomain $\langle\overline{\sigma}\rangle^T$ reachable for waves excited on $\sigma$ can be approximated with any precision in $\mathcal{H}$-norm by the wave $u^{f,0}(\cdot, T)$ due to appropriate choice of the control $f$ acting from $\sigma$. The proof of [7] relies on the fundamental Holmgren–John–Tataru unique continuation theorem for the wave equation (Tataru 1993).

### Laplacian on Waves

If $h = 0$, so that the system is governed only by boundary controls, its trajectory $\{u^{f,0}(\cdot, t) | 0 \leq t \leq T\}$ does not leave the reachable set $\mathcal{U}^T_\Gamma$. In this case, the system possesses one more intrinsic operator $L^T$ which acts in the subspace $\mathrm{cl}\,\mathcal{U}^T_\Gamma$ and is introduced through its graph

$$\mathrm{gr}\, L^T := \mathrm{cl}\left\{\{W^T_{\mathrm{bd}} f, -W^T_{\mathrm{bd}} f_{tt}\} | f \in C^\infty_0(\Gamma \times (0, T))\right\} \quad [8]$$

(closure in $\mathcal{H} \times \mathcal{H}$). By virtue of the relation $L^T W^T_{\mathrm{bd}} f = -\Delta_g W^T_{\mathrm{bd}} f$ following from the wave equation [1] and [6], the operator $L^T$ is interpreted as Laplacian on waves filling the subdomain $\langle\Gamma\rangle^T$.

In the case $T > T_*$, one has $\langle\Gamma\rangle^T = \Omega$, $\mathrm{cl}\,\mathcal{U}^T_\Gamma = \mathcal{H}$, and $L^T$ is a densely defined operator in $\mathcal{H}$, satisfying $L^T \subset L$. Using [7], one proves the equality $L^T = L$. This equality and representation [4] imply that

$$W^r_{\mathrm{vol}} h = \int_0^r (L^T)^{-1/2} \sin\left[(r-t)(L^T)^{1/2}\right] h(\cdot, t)\mathrm{d}t \quad [9]$$

for all $r \geq 0$ and any fixed $T > T^*$.

### Spectral Problem

The Dirichlet homogeneous boundary-value problem is to find nontrivial solutions of the system

$$-\Delta_g \varphi = \lambda \varphi \quad \text{in } \Omega \qquad [10]$$

$$\varphi = 0 \quad \text{on } \Gamma \qquad [11]$$

This problem is equivalent to the spectral analysis of the operator $L$; it has the discrete spectrum $\{\lambda_k\}_{k=1}^{\infty}, 0 < \lambda_1 < \lambda_2 \leq \cdots, \lambda_k \to \infty$; the eigenfunctions $\{\varphi_k\}_{k=1}^{\infty}, L\varphi_k = \lambda_k \varphi_k$, form an orthonormal basis in $\mathcal{H}$.

Expanding the solutions of the problem (1)–(3) over the eigenfunctions of the problem [10], [11] one derives the spectral representation of waves:

$$u^{f,0}(\cdot, T) = W_{\text{bd}}^T f = \sum_{k=1}^{\infty} (f, s_k^T)_{\mathcal{F}^T} \varphi_k(\cdot) \qquad [12]$$

where

$$s_k^T(\gamma, t) := \lambda_k^{-1/2} \sin\left[(T-t)\lambda_k^{1/2}\right] \partial_\nu \varphi_k(\gamma)$$

Thus, for a given control $f$, the Fourier coefficients of the wave $u^{f,0}$ are determined by the spectrum $\{\lambda_k\}_{k=1}^{\infty}$ and the derivatives $\{\partial_\nu \varphi_k\}_{k=1}^{\infty}$.

## Inverse problems

### General Setup

The set of pairs $\Sigma := \{\lambda_k; \partial_\nu \varphi_k\}_{k=1}^{\infty}$ associated with the problem [10], [11] is said to be the Dirichlet spectral data of the manifold $(\Omega, d)$. The spectral (frequency domain) inverse problem is to recover the manifold from its spectral data.

Since the speed of wave propagation is unity, the response operator $R^T$ contains the information not about the entire manifold but only about its part $\langle \Gamma \rangle^{T/2}$. This fact is taken into account in the dynamical (time domain) inverse problem which aims to recover the manifold from the operator $R^{2T}$ given for a fixed $T > T_*$.

If the manifolds $(\Omega', d')$ and $(\Omega'', d'')$ are isometric via an isometry $i : \Omega' \to \Omega''$, then, identifying the boundaries by $i(\gamma) \equiv \gamma$, one gets two manifolds with the common boundary $\Gamma = \partial\Omega' = \partial\Omega''$ which possess identical inverse data: $\Sigma' = \Sigma'', R'^{2T} = R''^{2T}$. Such manifolds are called equivalent: they are indistinguishable for the external observer extracting $\Sigma$ or $R^{2T}$ from the boundary measurements. Therefore, these data do not determine the manifold uniquely and both of the inverse problems need to be clarified. The precise formulation is given in the form of two questions:

1. Does the coincidence of the inverse data imply the equivalence of the manifolds?
2. Given the inverse data of an unknown manifold, how to construct a manifold possessing these data?

The BC method gives an affirmative answer to the first question and provides a procedure producing a representative of the class of equivalent manifolds from its inverse data. The method is based on the concepts of model and "coordinatization."

### Model

A pair consisting of an auxiliary Hilbert space $\tilde{\mathcal{H}}$ and an operator $\tilde{W}_{\text{bd}}^T : \mathcal{F}^T \to \tilde{\mathcal{H}}$ is said to be a model of the system $\alpha^T$, if $\tilde{W}_{\text{bd}}^T$ is determined by inverse data, and the map $U : W_{\text{bd}}^T f \mapsto \tilde{W}_{\text{bd}}^T f$ is an isometry from Ran $W_{\text{bd}}^T \subset \mathcal{H}$ onto Ran $\tilde{W}_{\text{bd}}^T \subset \tilde{\mathcal{H}}$. The model is an intermediate object in solving inverse problems. It plays the role of an auxiliary copy of the original dynamical system which an external observer can build from measurements on the boundary. While the genuine wave process inside $\Omega$, initiated by a boundary control, remains unaccessible for direct measurements, its $\tilde{\mathcal{H}}$-representation can be visualized by means of the model control operator $\tilde{W}_{\text{bd}}^T$. This is illustrated by the diagram on Figure 2, where the upper part is invisible for an external observer, whereas the lower part can be extracted from inverse data.

Each type of data determines a corresponding model. The spectral model is the pair

$$\tilde{\mathcal{H}} := l_2, \qquad \tilde{W}_{\text{bd}}^T := \{(\cdot, s_k^T)_{\mathcal{F}^T}\}_{k=1}^{\infty} \qquad [13]$$

(see [12]); the role of isometry $U$ is played by the Fourier transform $F : \mathcal{H} \to \tilde{\mathcal{H}}, Fy := \{(y, \varphi)_{\mathcal{H}}\}_{k=1}^{\infty}$. By virtue of [4], the data $\Sigma$ also determine the operator $\tilde{W}^r{}_{\text{vol}} : L_2([0, r]; \tilde{\mathcal{H}}) \to \tilde{\mathcal{H}}$,

$$\tilde{W}^r{}_{\text{vol}} = \int_0^r \tilde{L}^{-1/2} \sin\left[(r-t)(\tilde{L})^{1/2}\right](\cdot)(t)\, dt, \quad r \geq 0 \qquad [14]$$

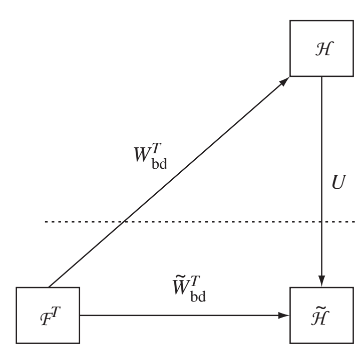



**Figure 2** Model of a system. (Data from Belishev (1997).)

where $\tilde{L} := ULU^* = \mathrm{diag}\{\lambda_k\}_{k=1}^{\infty}$. Thus, the spectral model allows one to see the Fourier images of invisible waves.

According to [5], the response operator $R^{2T}$ determines the modulus of the control operator

$$|W_{\mathrm{bd}}^T| = [(W_{\mathrm{bd}}^T)^* W_{\mathrm{bd}}^T]^{1/2} = (C^T)^{1/2}$$

which enters in the polar decomposition $W_{\mathrm{bd}}^T = \Phi |W_{\mathrm{bd}}^T|$. Along with it, the response operator determines the dynamical model

$$\tilde{\mathcal{H}} := \mathrm{cl}\ \mathrm{Ran}(C^T)^{1/2}, \qquad \tilde{W}_{\mathrm{bd}}^T := (C^T)^{1/2} \qquad [15]$$

The correspondence "system → model" is realized by the isometry $U = \Phi^* : W_{\mathrm{bd}}^T f \mapsto |W_{\mathrm{bd}}^T| f$. The operator $\tilde{L}^T := UL^T U^*$ dual to the Laplacian on waves, is determined by its graph

$$\mathrm{gr}\,\tilde{L}^T := \mathrm{cl}\left\{\{\tilde{W}_{\mathrm{bd}}^T f, -\tilde{W}_{\mathrm{bd}}^T f_{tt}\} | f \in C_0^{\infty}(\Gamma \times (0,T))\right\} \qquad [16]$$

(see [8]) and, therefore, $\tilde{L}^T$ is also determined by $R^{2T}$. In the case $T > T_*$, the operator $\tilde{W}_{\mathrm{vol}}^r : L_2([0,r]; \tilde{\mathcal{H}}) \to \tilde{\mathcal{H}}$ dual to $W_{\mathrm{vol}}^r$, is represented in the form

$$\tilde{W}_{\mathrm{vol}}^r = \int_0^r (\tilde{L}^T)^{-1/2} \sin\left[(r-t)(\tilde{L}^T)^{1/2}\right](\cdot)(t)\,\mathrm{d}t, \quad r \geq 0 \qquad [17]$$

in accordance with [9]. Thus, the dynamical model visualizes the $\Phi^*$-images of the waves propagating inside $\Omega$.

### Wave Coordinatization

In a general sense, a coordinatization is a correspondence between points $x$ of the studied set $\mathcal{A}$ and elements $\tilde{x}$ of another set $\tilde{\mathcal{A}}$ such that: (i) the elements of $\tilde{\mathcal{A}}$ are accessible and distinguishable; (ii) the map $x \mapsto \tilde{x}$ is a bijection; and (iii) relations between elements of $\tilde{\mathcal{A}}$ determine those between points of $\mathcal{A}$ which are studied (H Weyl). Coordinatization enables one to study $\mathcal{A}$ via operations with coordinates $\tilde{x} \in \tilde{\mathcal{A}}$.

The external observer investigating the manifold probes $\Omega$ with waves initiated by sources on $\Gamma$. The relevant coordinatization of $\Omega$ described below uses such waves and is implemented in three steps.

*Step 1* (*subdomains*) Let $x(\gamma, \tau)$ be the end point of the geodesic of the length $\tau > 0$ emanating from $\gamma \in \Gamma$ in the direction $-\nu(\gamma)$, and let $\sigma_\gamma^\varepsilon \subset \Gamma$ be a small neighborhood shrinking to $\gamma$ as $\varepsilon \to 0$. If $\tau \leq \tau_*(\gamma)$, then the family of subdomains

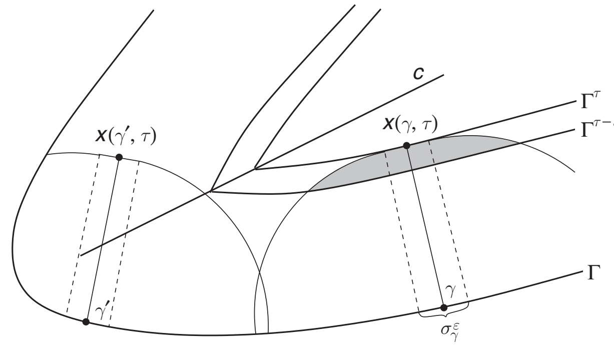



**Figure 3** The subdomains.

$$\omega^\varepsilon(\gamma, \tau) := [\langle\Gamma\rangle^\tau \setminus \langle\Gamma\rangle^{\tau-\varepsilon}] \cap \langle\sigma_\gamma^\varepsilon\rangle^\tau$$

(shaded domain on Figure 3) shrinks to $x(\gamma, \tau)$; if $\tau > \tau_*(\gamma)$, then the family terminates: $\omega^\varepsilon(\gamma, \tau) = \emptyset$ as $\varepsilon < \varepsilon_0(\gamma)$ (the case $\gamma = \gamma'$ in Figure 3). Such behavior of subdomains implies that

$$\lim_{\varepsilon \to 0} \left\langle [\langle\Gamma\rangle^\tau \setminus \langle\Gamma\rangle^{\tau-\varepsilon}] \cap \langle\sigma_\gamma^\varepsilon\rangle^\tau \right\rangle^r = \begin{cases} \langle x(\gamma,\tau)\rangle^r, & \tau \leq \tau_*(\gamma) \\ \emptyset, & \tau > \tau_*(\gamma) \end{cases} \qquad [18]$$

*Step 2* (*wave subspaces*) Pass from the subdomains to the corresponding subspaces $\mathcal{H}\langle\Gamma\rangle^\tau, \mathcal{H}\langle\sigma_\gamma^\varepsilon\rangle^\tau$, $\mathcal{H}\langle\omega^\varepsilon(\gamma,\tau)\rangle^r$, and represent them via reachable sets by [7]:

$$\mathcal{H}\langle\Gamma\rangle^\tau = \mathrm{cl}\ W_{\mathrm{bd}}^\tau \mathcal{F}^\tau, \qquad \mathcal{H}\langle\sigma_\gamma^\varepsilon\rangle^\tau = \mathrm{cl}\ W_{\mathrm{bd}}^\tau \mathcal{F}_{\sigma_\gamma^\varepsilon}^\tau$$

$$\begin{aligned}\mathcal{H}\langle\omega^\varepsilon(\gamma,\tau)\rangle^r &= \mathrm{cl}\ W_{\mathrm{vol}}^r L_2([0,r]; \mathcal{H}\omega^\varepsilon(\gamma,\tau)) \\ &= \mathrm{cl}\ W_{\mathrm{vol}}^r L_2\Big([0,r]; [\mathcal{H}\langle\Gamma\rangle^\tau \ominus \mathcal{H}\langle\Gamma\rangle^{\tau-\varepsilon}] \cap \mathcal{H}\langle\sigma_\gamma^\varepsilon\rangle^\tau\Big) \\ &= \mathrm{cl}\ W_{\mathrm{vol}}^r L_2\Big([0,r]; [\mathrm{cl}\ W_{\mathrm{bd}}^\tau \mathcal{F}^\tau \ominus \mathrm{cl}\ W_{\mathrm{bd}}^{\tau-\varepsilon} \mathcal{F}^{\tau-\varepsilon}] \cap \mathrm{cl}\ W_{\mathrm{bd}}^\tau \mathcal{F}_{\sigma_\gamma^\varepsilon}^\tau\Big)\end{aligned}$$

Define

$$\mathcal{W}_{(\gamma,\tau)}^r := \lim_{\varepsilon \to 0} \mathrm{cl}\ W_{\mathrm{vol}}^r L_2\Big([0,r]; [\mathrm{cl}\ W_{\mathrm{bd}}^\tau \mathcal{F}^\tau \ominus \mathrm{cl}\ W_{\mathrm{bd}}^{\tau-\varepsilon} \mathcal{F}^{\tau-\varepsilon}] \cap \mathrm{cl}\ W_{\mathrm{bd}}^\tau \mathcal{F}_{\sigma_\gamma^\varepsilon}^\tau\Big) \qquad [19]$$

$\mathcal{W}_{(\gamma,0)}^r := \mathcal{W}_{(\gamma,+0)}^r$, $r \geq 0$ (the limits in the sense of the strong operator convergence of the projections in $\mathcal{H}$ on the corresponding subspaces). By the definitions, one has $\mathcal{W}_{(\gamma,\tau)}^r = \lim_{\varepsilon \to 0} \mathcal{H}\langle\omega^\varepsilon(\gamma,\tau)\rangle^r$, whereas [18] leads to the equality

$$\mathcal{W}_{(\gamma,\tau)}^r = \begin{cases} \mathcal{H}\langle x(\gamma,\tau)\rangle^r, & \tau \leq \tau_*(\gamma) \\ \{0\}, & \tau > \tau_*(\gamma) \end{cases} \qquad [20]$$

for all $\gamma \in \Gamma$, $\tau \geq 0$, $r \geq 0$. As a result, since any $x \in \Omega$ can be represented as $x = x(\gamma, \tau)$, one attaches to every point of the manifold a family of expanding subspaces $\{\mathcal{W}^r_{(\gamma,\tau)} | r \geq 0\}$ built out of waves. As is seen from [20], the family is determined by the point $x$ (not dependent on the representation $x = x(\gamma,\tau)$); the subspaces which it consists of coincide with $\mathcal{H}\langle x\rangle^r$.

Expressing the distance as

$$d(x', x'') = 2\ \inf\{r > 0\,|\,\mathcal{H}\langle x'\rangle^r \cap \mathcal{H}\langle x''\rangle^r \neq \{0\}\}$$

in accordance with [20], one can represent

$$d(x', x'') = 2\ \inf\{r > 0\,|\,\mathcal{W}^r_{(\gamma',\tau')} \cap \mathcal{W}^r_{(\gamma'',\tau'')} \neq \{0\}\} \qquad [21]$$

where $x' = x(\gamma', \tau')$, $x'' = x(\gamma'', \tau'')$, and hence find the distance via the above families.

*Step 3* (*wave copy*) By varying $\gamma \in \Gamma, \tau \geq 0$, gather all nonzero families $\{\mathcal{W}^r_{(\gamma,\tau)} | r \geq 0\} =: \tilde{x}$ in the set $\tilde{\Omega} = \{\tilde{x}\}$. Redenoting $\mathcal{W}^r_{\tilde{x}} := \mathcal{W}^r_{(\gamma,\tau)} \in \tilde{x}$, endow the set with the distance

$$\tilde{d}(\tilde{x}', \tilde{x}'') := 2\ \inf\{r > 0\,|\,\mathcal{W}^r_{\tilde{x}'} \cap \mathcal{W}^r_{\tilde{x}''} \neq \{0\}\} \qquad [22]$$

In view of [21], one has $d(x', x'') = \tilde{d}(\tilde{x}', \tilde{x}'')$, so that the metric space $(\tilde{\Omega}, \tilde{d})$ is an isometric copy of $(\Omega, d)$ by construction. Thus, the correspondence $x \mapsto \tilde{x}$ ("point $\mapsto$ family") is an isometry and satisfies the general principles (i)–(iii) of coordinatization.

The manifold $(\tilde{\Omega}, \tilde{d})$ is the end product of the wave coordinatization. It represents the original manifold as a collection of infinitesimal sources interacting with each other via the waves which they produce.

### Solving Inverse Problems

The motivation for the above coordinatization is that the wave copy can be reproduced via any model. Namely, the external observer with the knowledge of $\Sigma$ or $R^{2T}(T > T_*)$ can recover $(\tilde{\Omega}, \tilde{d})$ up to isometry by the following procedure:

1. Construct the model corresponding to the given inverse data and determine the operators $\tilde{W}^\tau_{\text{bd}}$, $0 \leq \tau \leq T$ by [13], [15]; then determine $\tilde{L}, \tilde{L}^T$, and $\tilde{W}^r_{\text{vol}}$ by [14] or [16], [17].
2. Replace on the right-hand side of [19] all operators $W$ without tildes by the ones with tildes, and get the subspaces $\tilde{\mathcal{W}}^r_{(\gamma,\tau)} = U\mathcal{W}^r_{(\gamma,\tau)}$, $\gamma \in \Gamma$, $\tau \geq 0$, $r \geq 0$.
3. Gather all nonzero families $\{\tilde{\mathcal{W}}^r_{(\gamma,\tau)} | r \geq 0\} =: \hat{x}$ in the set $\hat{\Omega} = \{\hat{x}\}$ and redenote the subspaces as $\tilde{\mathcal{W}}^r_{\hat{x}} := \tilde{\mathcal{W}}^r_{(\gamma,\tau)} \in \hat{x}$; endow the set with the metric $\hat{d}(\hat{x}', \hat{x}'') := 2\ \inf\{r > 0\,|\,\tilde{\mathcal{W}}^r_{\hat{x}'} \cap \tilde{\mathcal{W}}^r_{\hat{x}''} \neq \{0\}\}$ (see [22]), and get a sample $(\hat{\Omega}, \hat{d})$ of the wave copy $(\tilde{\Omega}, \tilde{d})$.

This sample is isometric to the original $(\Omega, d)$ by construction. Identifying properly the boundaries $\partial\hat{\Omega}$ and $\Gamma$, one turns $(\hat{\Omega}, \hat{d})$ into a canonical representative of the class of equivalent manifolds possessing the given inverse data.

If the response operator $R^{2T}$ is given for a fixed $T < T_*$, the above procedure produces the wave copy of the submanifold $(\langle\Gamma\rangle^T, d)$. This locality in time is an intrinsic feature and advantage of the BC method: longer time of observation on $\Gamma$ increases the depth of penetration into $\Omega$.

### Amplitude Formula

Another variant of the BC method is based on geometrical optics formulas describing the propagation of singularities of the waves.

Let $y \in \mathcal{H}$, and let $\beta$ be the density of the volume in semigeodesic coordinates: $dx = \beta\, d\Gamma\, d\tau$; the function

$$\tilde{y}(\gamma, \tau) := \begin{cases} \beta^{1/2}(\gamma, \tau)\, y(x(\gamma, \tau)), & (\gamma, \tau) \in \Theta \\ 0, & \text{otherwise} \end{cases}$$

defined on $\Gamma \times [0, T_*]$ is called the image of $y$. The amplitude formula represents the images of waves initiated by boundary controls in the form

$$\widetilde{u^{f,0}(\cdot, T)}(\gamma, \tau) = \lim_{t \to T-\tau-0} [(W^T_{\text{bd}})^*(I - P^\tau)W^T_{\text{bd}} f](\gamma, t)$$
$$0 < \tau < T$$

where $I$ is the identity operator and $P^\tau$ is the projection in $\mathcal{H}$ onto $\text{cl}\, W^\tau_{\text{bd}}\mathcal{F}^\tau$. The formula is derived by the ray method going back to J Hadamard, the derivation uses the controllability [7].

Any model determines the right-hand side of the last relation by the isometry: $(W^T_{\text{bd}})^*(I - P^\tau) W^T_{\text{bd}} = (\tilde{W}^T_{\text{bd}})^*(\tilde{I} - \tilde{P}^\tau)\tilde{W}^T_{\text{bd}}$, where $\tilde{W}^T_{\text{bd}} = UW^T_{\text{bd}}, \tilde{I}$ is the identity operator, and $\tilde{P}^\tau = UP^\tau U^*$ is the projection in $\tilde{\mathcal{H}}$ onto $\text{cl}\, \tilde{W}^\tau_{\text{bd}}\mathcal{F}^\tau$. This leads to the representation

$$\widetilde{u^{f,0}(\cdot, T)}(\gamma, \tau) = \lim_{t \to T-\tau-0} [(\tilde{W}^T_{\text{bd}})^*(\tilde{I} - \tilde{P}^\tau)\tilde{W}^T_{\text{bd}} f](\gamma, t)$$
$$0 < \tau < T \qquad [23]$$

and makes the amplitude formula a useful tool for solving the inverse problems. The external observer can construct a model via inverse data and then visualize by [23] the wave images on the part $\Theta^T$ of the pattern (see **Figure 1**). The collection of images $\widetilde{u^{f,0}}$ corresponding to all possible controls $f$ is rich enough for recovering the tensor $g$ on $\Theta^T$ (i.e., the metric tensor in semigeodesic coordinates) and turning the pattern into an isometric copy of the submanifold $(\langle\Gamma\rangle^T, d)$. This variant of the method is

more appropriate if one needs to recover unknown coefficients of the wave equation in $\Omega$ – it can be realized in terms of numerical algorithms.

## Extensions of the Method

Electromagnetic waves are also well suited for coordinatization and for constructing the wave copy $(\tilde{\Omega}, \tilde{d})$. An appropriate version of the amplitude formula also exists for the system governed by the Maxwell equations (see Further Reading). At present (2004), the applicability of the BC method to three-dimensional inverse problems of elasticity theory is still an open question. The following hypothesis concerns the Lamé system: the wave coordinatization procedure (steps 1–3) using the elastic waves instead of the above $u^{f,0}$, gives rise to the copy of $\Omega \subset \mathbf{R}^3$ endowed with the metric $|dx|^2/c_p^2$ where $c_p = \sqrt{(\lambda + 2\mu)/\rho}$ is the speed of the pressure waves.

The concept of model is used for solving inverse problems for the heat and Schrödinger equations (Avdonin and Belishev, 1995–2004), as well as for the problem of boundary data continuation (Belishev 2001, Kurylev and Lassas 2002). A variant of the BC method allows one to recover not only the manifold but also the Schrödinger type operators on it and/or the dissipative term in the scalar wave equation (Kurylev and Lassas 1993–2003).

An appropriate version of the amplitude formula solves the inverse problem for one-dimensional two-velocity dynamical system which describes the waves consisting of two modes propagating with different speeds and interacting with each other (Belishev, Blagoveschenskii, Ivanov, 1997–2000).

One more variant of coordinatization going back to the first paper on the BC method, associates with points $x \in \Omega$ the Dirac measures $\delta_x$; then, their images $\tilde{\delta}_x$ are identified via suitable models. This variant solves inverse problems on graphs and the two-dimensional elliptic Calderon problem. The reader is referred to articles by the present author listed in Further Reading.

Within the scope of the method, one derives some natural analogs of the classical Gelfand–Levitan–Krein–Marchenko equations (Belishev, 1987–2001). Also, an appropriate analog solves the kinematic inverse problem for a class of two-dimensional manifolds (Pestov 2004).

There exists an abstract version of the approach, embedding the BC method into the framework of linear system theory (Belishev 2001). The method is also related to the problem of triangular factorization of operators (Belishev and Pushnitski 1996).

Numerical algorithms for solving two-dimensional spectral and dynamical inverse problems for the wave equation $\rho u_{tt} - \Delta u = 0$ which recover the variable density $\rho$ have been developed and tested (Filippov, Gotlib, Ivanov, 1994–1999).

*See also:* Dynamical Systems and Thermodynamics; Geophysical Dynamics; Inverse Problem in Classical Mechanics.

## Further Reading

Belishev MI (1988) On an approach to multidimensional inverse problems for the wave equation. *Soviet Mathematics. Doklady* 36(3): 481–484.

Belishev MI (1996) Canonical model of a dynamical system with boundary control in the inverse problem of heat conductivity. *St. Petersburg Mathematical Journal* 7(6): 869–890.

Belishev MI (1997) Boundary control in reconstruction of manifolds and metrics. *Inverse Problems* 13(5): R1–R45.

Belishev MI (2001) Dynamical systems with boundary control: models and characterization of inverse data. *Inverse Problems* 17: 659–682.

Belishev MI (2002) How to see waves under the Earth surface (the BC-method for geophysicists). In: Kabanikhin SI and Romanov VG (eds.) *Ill-Posed and Inverse Problems*, pp. 67–84. Utrecht/Boston: VSP.

Belishev MI (2003) The Calderon problem for two-dimensional manifolds by the BC-method. *SIAM Journal of Mathematical Analysis* 35(1): 172–182.

Belishev MI (2004) Boundary spectral inverse problem on a class of graphs (trees) by the BC-method. *Inverse Problems* 20(3): 647–672.

Belishev MI and Glasman AK (2001) Dynamical inverse problem for the Maxwell system: recovering the velocity in the regular zone (the BC-method). *St. Petersburg Mathematical Journal* 12(2): 279–319.

Belishev MI and Gotlib VYu (1999) Dynamical variant of the BC-method: theory and numerical testing. *Journal of Inverse and Ill-Posed Problems* 7(3): 221–240.

Belishev MI, Isakov VM, Pestov LN, and Sharafutdinov VA (2000) On reconstruction of metrics from external electromagnetic measurements. *Russian Academy of Sciences. Doklady. Mathematics* 61(3): 353–356.

Belishev MI and Ivanov SA (2002) Characterization of data of dynamical inverse problem for two-velocity system. *Journal of Mathematical Sciences* 109(5): 1814–1834.

Belishev MI and Lasiecka I (2002) The dynamical Lamé system: regularity of solutions, boundary controllability and boundary data continuation. *ESAIM COCV* 8: 143–167.

Katchalov A, Kurylev Y, and Lassas M (2001) Inverse Boundary Spectral Problems. *Chapman and Hall/CRC Monographs and Surveys in Pure and Applied Mathematics*, vol. 123. Boca Raton, FL: Chapman and Hall/CRC.

# Boundary-Value Problems for Integrable Equations

**B Pelloni**, University of Reading, UK



## Introduction

Integrable equations are a special class of nonlinear equations arising in the modeling of a wide variety of physical phenomena. It has been argued that integrable PDEs are in a certain, specific sense "universal" models for physical phenomena involving weak nonlinearity. Indeed, integrable equations are obtained by a procedure involving rescaling and an asymptotic expansion from very large classes of nonlinear evolution equations, which preserves integrability while retaining in the limit weakly nonlinear effects. For this reason, integrable equations are a very important class of PDEs. Important examples are the nonlinear Schrödinger (NLS) equation

$$\mathrm{i}q_t + q_{xx} - 2\lambda|q|^2 q = 0, \quad \lambda = \pm 1 \qquad [1]$$

the Korteweg–deVries (KdV) equation

$$q_t + q_x \pm q_{xxx} + 6qq_x = 0 \qquad [2]$$

the modified KdV (mKdV) equation

$$q_t \pm q_{xxx} \mp 6\lambda q^2 q_x = 0, \quad \lambda = \pm 1 \qquad [3]$$

and the sine-Gordon (SG) equation in light-cone or laboratory coordinates

$$q_{xt} + \sin q = 0 \quad \text{or} \quad q_{tt} - q_{xx} + \sin q = 0 \qquad [4]$$

A general method for solving the initial-value problem for integrable equations in one space dimension was discovered in 1967, when in a pioneering and much celebrated work (Gardner *et al.* 1967), the initial-value problems for KdV with decaying initial condition was completely solved. Soon afterwards, it was understood that this method, now known as the "inverse scattering transform," is of more general applicability. Indeed, it can be applied to those nonlinear equations that can be written as the compatibility condition of a pair of linear eigenvalue equations. The method of solution for the Cauchy problem essentially relies on the possibility of expressing the equation through this pair, now called a Lax pair after the work of Lax (1968), who first clarified the connection. Zakharov and Shabat (1972) constructed such a pair for the NLS equation, and in subsequent years the Lax pairs associated with all important integrable equations in one and two spatial variables were constructed. These include the NLS, sG, mKdV, Davey–Stewartson I and II, and Kamdotsev–Petviashvili I and II equations.

There is no universally accepted definition of an integrable PDE, but on account of the above results, the existence of a Lax pair can be taken as the defining property of such equations. In the course of the 1970s, the inverse scattering transform was applied to solve the initial-value (Cauchy) problem for many integrable equations. In principle, there is no obstruction to solving analytically the initial-value problem by the inverse scattering transform as soon as a Lax pair is constructed for the equation, and appropriate decaying initial conditions are prescribed. The solution is then characterized in terms of a certain integral equation. This approach is equivalent to associating with the initial-value problem a classical problem in complex analysis, namely a matrix Riemann–Hilbert problem, defined in the complex spectral space. This point of view is currently taken by many authors as it provides a unifying and very flexible framework for the analysis.

After the success of the inverse scattering transform in solving the Cauchy problem, it was natural to attempt to generalize the approach to boundary-value problems. To describe the difficulties involved in this generalization, consider the case of evolution equations in one space and one time dimensions. The independent variables can be denoted by $(x, t)$, with $t > 0$ representing time. While the initial-value problem is posed on the full real line, hence for $x \in (-\infty, \infty)$, the simplest boundary-value problem is posed on a half-line, for $x \in (0, \infty)$. In addition to initial conditions for initial time $t = 0$, it is necessary to prescribe conditions at the boundary $x = 0$. The number of conditions that must be prescribed to obtain a problem which admits a unique solution depends on the particular equation, but for evolution equation it is roughly equal to half the number of $x$-derivatives involved in the equation. For example, for the NLS equation, a well-posed problem is defined as soon as one boundary condition at $x = 0$ is prescribed; hence a typical boundary-value problem for this equation is obtained, for example, when $q(x, 0) = q_0(x)$ and $q(0, t) = g_0(t)$ are prescribed and compatible, so that $q_0(0) = g_0(0)$. It follows that, while $q_{xx}(0, t)$ can be computed from the equation, $q_x(0, t)$ is not immediately known. An even more difficult situation arises for the KdV equation [2] (with the + sign), for which a well-posed problem is again defined as soon as one boundary condition is prescribed, so that there are two unknown boundary values.

Because of this simple fact, a straightforward application of the ideas of the inverse scattering transform immediately encounters one crucial difficulty. This transform method yields an integral representation of the solution which involves not only the given boundary conditions $f(t)$, but also the other "unknown" boundary values – in our example for the NLS equation, the function $q_x(0,t)$. The problem of characterizing these unknown boundary values has impeded progress in this direction for over thirty years.

On account of their physical significance, various boundary-value problems for the KdV equation have been considered, and classical PDE techniques (not specific to integrable models) have been used to establish existence and uniqueness results (Bona *et al.* 2001, Colin and Ghidaglia 2001, Colliander and Kenig 2001). These approaches, and in particular the approach of Colliander and Kenig, are quite general and possibly of wide applicability, and give global existence results in wide functional classes. However, they do not rely on integrability properties. Indeed, none of these results use the integrable structure of the equation in any fundamental or systematic way. However, the fact that these equations are integrable on the full line implies very special properties that should be exploited in the analysis and it is natural to try to generalize the inverse scattering transform approach.

Such a generalization is sometimes directly possible. For example, it has been used for studying the problem on the half-line for the hyperbolic version of the sG equation [4a] which does not involve unknown boundary values (Fokas 2000, Pelloni). It has also been used to study some specific boundary-value problems for the NLS equation, for example, for homogeneous Dirichlet or Neumann conditions, when it is possible to use even or odd extensions of the problem to the full line (Ablowitz and Segur 1974), or more recently in Degasperis *et al.* (2001). In the latter case, however, the unknown boundary values are characterized through an integral Fredholm equation, which does not admit a unique solution. Some special cases of boundary-value problems for the KdV equation (Adler *et al.* 1997, Habibullin 1999) and elliptic sG (Sklyanin 1987) have also been studied via the inverse scattering transform. However all the examples considered are nongeneric, and it has recently been shown (Fokas, in press) that the boundary conditions chosen fall in the special class of the so-called "linearizable" boundary conditions, for which the problem can be solved as if it were posed on the full line. One cannot hope to use similar methods to solve the problem with generic boundary conditions.

Recently, Fokas (2000) introduced a general methodology to extend the ideas of the inverse scattering transform to boundary-value problems. This methodology provides the tools to analyze boundary-value problems for integrable equations to a considerable degree of generality. We note as a side remark that linear PDEs are trivially integrable, in the sense of admitting a Lax pair (in this case the Lax pair can be found algorithmically, while the construction of the Lax pair associated with a nonlinear equation is by no means trivial). As a consequence of this remark, the extension of the inverse scattering transform also provides a method for solving boundary-value problems for a large variety of linear PDEs of mathematical physics.

What follows is a general description of the approach of Fokas, considering, for the sake of concreteness, the case of an integrable PDE in the two variables $(x,t)$ which vary in the domain $D$ (typically, for an evolution problem $D=(0,\infty)\times(0,T)$). We assume that $q(x,t)$ denotes the unique solution of a boundary-value problem posed for such an equation.

The method consists of the following steps.

1. Write the PDE as the compatibility condition of a Lax pair. This is a pair of linear ODEs for the function $\mu=\mu(x,t,k)$ involving the solution $q(x,t)$ of the PDE, the derivatives of this solution, and a complex parameter $k$, called the spectral parameter. This can be done algorithmically for linear PDEs, and in this case $\mu(x,t,k)$ is a scalar function. For nonlinear integrable PDEs, $\mu(x,t,k)$ is in general a matrix-valued function.

   The equivalence of the PDE with a Lax pair can be reformulated in the language of differential forms, and in this language it is easier to describe the methodology in general. Assume then that $\Omega(x,t,k)$ is a differential 1-form expressed in terms of a function $q(x,t)$ and its derivatives, and of a complex variable $k$, and one which is characterized by the property that $d\Omega=0$ if and only if $q(x,t)$ satisfies the given PDE. The closure of the form $\Omega$ yields the two important consequences 2(a) and 2(b) below.
2. (a) Since the domain $D$ under consideration is simply connected, the closed form $\Omega$ is also exact; hence, it is possible to find the particular, 0-form $\mu(x,t,k)$, solving $d\mu=\Omega$. In particular, $\mu(x,t,k)$ can be chosen to be sectionally bounded with respect to $k$ by solving either a Riemann–Hilbert problem or a $d$-bar problem in the complex spectral $k$ plane, and the solution $\mu(x,t,k)$ is then expressed in terms of certain "spectral functions" depending on all the boundary values

of the solution $q(x,t)$ of the PDE. The function $q(x,t)$ can then be expressed in terms of $\mu(x,t,k)$. (b) The integral of $\Omega$ along the boundary of the domain $D$ vanishes. This yields an integral constraint between all boundary values of the solution of the PDE, which becomes an algebraic constraint for the spectral functions. The resulting algebraic identity is called the "global relation."

3. The last step is the analysis the $k$-invariance properties of the global relation. This analysis yields the characterization of the spectral functions in terms only of the given boundary conditions.

The crucial and most difficult step in the solution process is the characterization described above. The analysis required depends on the type of problem under consideration. For nonlinear integrable evolution PDEs posed on the half-line $x > 0$, in general the characterization mentioned in step (3) involves solving a system of nonlinear Volterra integral equations. This is an important difference from the case of the Cauchy problem, where the solution is given by a single integral equation where all the terms are explicitly known.

The method outlined above has been applied successfully to solve a variety of boundary-value problems for linear and integrable nonlinear PDEs. For concreteness, here the focus is on the important case of integrable evolution PDEs in one space, which illustrates clearly the generalities of this method.

## Integrable Evolution Equations in One Space Dimension

The crucial property of integrable PDEs which is used in the inverse scattering transform approach to solve the initial-value problem is the fact that they can be written as the compatibility of a Lax pair. Many integrable evolution equations of physical significance (such as NLS, KdV, sG, and mKdV) admit a Lax pair of the form

$$\begin{aligned} \mu_x + \mathrm{i}f_1(k)\sigma_3\mu &= Q(x,t,k)\mu \\ \mu_t + \mathrm{i}f_2(k)\sigma_3\mu &= \tilde{Q}(x,t,k)\mu \end{aligned} \qquad [5]$$

where $\mu(x,t,k)$ is a $2 \times 2$ matrix, $\sigma_3 = \mathrm{diag}(1, -1)$, $f_i(k), i = 1, 2$, are analytic functions of the complex parameter $k$, and $Q$, $\tilde{Q}$ are analytic functions of $k$, of the function $q(x,t)$ (and of its complex conjugate $\overline{q(x,t)}$ for complex-valued problems) and of its derivatives. For example, the NLS equation [1] is equivalent to the compatibility condition of the pair

$$\begin{aligned} \mu_x + \mathrm{i}k\sigma_3\mu &= Q\mu, \quad Q = \begin{pmatrix} 0 & q \\ \lambda\bar{q} & 0 \end{pmatrix} \\ \mu_t + 2\mathrm{i}k^2\sigma_3\mu &= (2kQ - \mathrm{i}Q_x\sigma_3 - \mathrm{i}\lambda|q|^2\sigma_3)\mu \end{aligned} \qquad [6]$$

The first step towards a systematic new approach to solving boundary-value problem was the work of Fokas and Its, who associated the boundary-value problem for NLS on the half-line to a single Riemann–Hilbert problem determined by both equations in the Lax pair. The jump determining this Riemann–Hilbert problem has an explicit exponential dependence on both $x$ and $t$. This differs from the classical inverse scattering approach, in which the $x$-part of the Lax pair is used to determine an $x$-transform with $t$-dependent scattering data, and the $t$-part of the Lax pair is then exploited to find the time evolution of these data. The work of Fokas and Its led to the understanding that both equations in the Lax pair [6] must be considered in order to construct a spectral transform appropriate to solve boundary-value problems. Fokas (2000) reviews his systematic way to solve these problems by performing the simultaneous spectral analysis of both equations in the Lax pair. The transform thus obtained, which is a nonlinearization of the Fourier transform, precisely generalizes the inverse scattering transform.

This simultaneous analysis also leads naturally to the identification of the "global relation" which holds between initial and boundary data, and which plays an essential role in deriving an expression for the solution of the problem which does not involve unknown boundary values.

The Riemann–Hilbert problem with explicit $(x,t)$ dependence, the global relation, and the invariance properties of the latter with respect to the spectral parameter are the fundamental ingredients of this systematic approach to solve boundary-value problems for integrable equations.

The steps involved in this method are summarized in the introduction. While steps (1) and (2) can be described generally, and, once the Lax pair is identified, can be performed algorithmically (at least under the assumption that the solution of the PDE exists), the last step is the most difficult part of the analysis, and it needs to be considered separately for each given problem. However, it is this step that yields the effective characterization of the solution.

The results obtained for the particular case of eqn [1] are reviewed in detail in the next section, as they provide an important example, which can be generalized without any conceptual difficulty to eqns [2]–[4].

### The NLS Equation

As already mentioned, the initial-value problem for NLS was solved, for decaying initial condition, by Zakharov and Shabat, and studied in depth by many others. However, by the mid-1990s only a handful of papers had been written on the solution of the boundary-value problem posed on the half-line, all on a specific example or aspect of the problem, or attempts at solving the problem using general PDE techniques.

For this equation, the approach of Fokas yields the following results. Let the complex-valued function $q(x,t)$ satisfy the NLS equation [1], for $x>0$ and $t>0$, for prescribed one initial and one boundary conditions. For the sake of concreteness, we select the specific initial and boundary conditions

$$\begin{aligned} q(x,0) &= q_0(x) \in \mathcal{S}(\mathbb{R}^+) \\ q(0,t) &= g_0(t) \in \mathcal{S}(\mathbb{R}^+) \\ q_0(0) &= g_0(0) \end{aligned} \qquad [7]$$

where $\mathcal{S}$ denotes the space of Schwartz functions (similar results hold for different choices of boundary conditions, and less restrictive function classes).

The solution of this initial boundary-value (IBV) problem can be constructed as follows (Fokas 2000, 2002; in press):

- Given $q_0(x)$ construct the spectral functions $\{a(k), b(k)\}$. These functions are defined by

  $$a(k) = \phi_2(0,k), \quad b(k) = \phi_1(0,k)$$

  where the vector $\phi(x,k)$ with components $\phi_1(x,k)$ and $\phi_2(x,k)$ is the following solution of the $x$-problem of the associated Lax pair evaluated at $t=0$:

  $$\begin{aligned} \phi_x + \mathrm{i}k\sigma_3\phi &= Q(x,0,k)\phi, \quad 0<x<\infty,\ \mathrm{Im}\,k\geq 0 \\ \phi(x,k) &= \mathrm{e}^{\mathrm{i}kx}\left(\begin{pmatrix}0\\1\end{pmatrix}+o(1)\right) \text{ as } x\to\infty \\ Q(x,0,k) &= \begin{pmatrix} 0 & q_0(x) \\ \lambda\bar{q}_0(x) & 0\end{pmatrix}\end{aligned}$$

  ($\sigma_3$ and $Q(x,t,k)$ are defined after eqns [5] and [6], respectively).
- Given $q_0(x)$ and $g_0(t)$ characterize $g_1(t)$ by the requirement that the spectral functions $\{A(t,k), B(t,k)\}$ satisfy the global relation

  $$\begin{aligned} B(t,k) - R(k)A(t,k) &= \mathrm{e}^{4\mathrm{i}k^2t}\frac{c(t,k)}{a(k)} \\ R(k) &= \frac{b(k)}{a(k)},\ t\in[0,T],\ k\in\bar{D}\end{aligned} \qquad [8]$$

  where $D$ denotes the first quadrant of the complex $k$-plane:

  $$D = \{k|\mathrm{Re}\,k>0, \mathrm{Im}\,k>0\}$$

  $\bar{D}$ denotes the closure of $D$, and $c(t,k)$ is a function of $k$ analytic in $D$ and of order $\mathrm{O}(1/k)$ as $k\to\infty$. The spectral functions are defined by

  $$\begin{aligned} A(t,k) &= \mathrm{e}^{2\mathrm{i}k^2t}\overline{\Phi_2(t,\bar{k})}, \\ B(t,k) &= -\mathrm{e}^{2\mathrm{i}k^2t}\Phi_1(t,k)\end{aligned} \qquad [9]$$

  where the vector $\Phi(t,k)$ with components $\Phi_1$ and $\Phi_2$ is the following solution of the $t$-problem of the associated Lax pair evaluated at $x=0$:

  $$\begin{aligned} &\Phi_t + 2\mathrm{i}k^2\sigma_3\Phi = \tilde{Q}(0,t,k)\Phi \\ &\quad 0<t<T, \quad k\in\mathbb{C} \\ &\quad \Phi(0,k) = \begin{pmatrix}0\\1\end{pmatrix} \\ &\quad \tilde{Q}(0,t,k) = \\ &\begin{pmatrix} -|g_0(t)|^2 & 2kg_0(t) + \mathrm{i}\lambda\lambda g_1(t) \\ 2k\bar{g}_0(t) - \mathrm{i}\lambda\bar{g}_1(t) & |g_0(t)|^2\end{pmatrix}\end{aligned} \qquad [10]$$

- Given $a(k), b(k)$ and $A(k), B(k)$, define a $2\times 2$ matrix Riemann–Hilbert problem. This problem has the distinctive feature that its jump has explicit $(x,t)$ dependence in the exponential form of $\exp\{\mathrm{i}kx+2\mathrm{i}k^2t\}$. Determine $q(x,t)$ in terms of the solution of this Riemann–Hilbert problem by using the fact that these functions are related by the Lax pair. Then the function $q(x,t)$ solves the IBV problem [1]–[7] with $q(x,0)=q_0(x), q(0,t)=g_0(t)$, and $q'_x(0,t)=g_1(t)$.

The above construction can be summarized in the following theorem (Fokas 2002):

**Theorem 1** *Consider the boundary-value problem for the NLS equation [1] determined by the conditions [7]. Let $a(k), b(k)$ be given by [8], and suppose that there exists a function $g_1(t)$ such that if $A(k), B(k)$ are defined by [9], then the global relation [8] holds.*

*Let $M(x,t,k)$ be the solution of the $2\times 2$ Riemann–Hilbert problem with jump on the real and imaginary axes given by*

- *$M_-(x,t,k)=M_+(x,t,k)J(x,t,k)$ with $M=M_-$ in the second and fourth quadrants of $\mathbb{C}$, $M=M_+$ in the first and third quadrants of $\mathbb{C}$, and $J(x,t,k)$ is defined in terms of a, b, A, B and the exponential $\mathrm{e}^{\mathrm{i}kx-2\mathrm{i}k^2t}$.*
- *$M=I+\mathrm{O}(1/k)$ as $k\to\infty$ and has appropriate residue conditions if there are poles*

  *Then M(x,t,k) exists and is unique, and*

  $$q(x,t) = 2\mathrm{i}\lim_{k\to\infty}(kM(x,t,k))_{12}$$

The result above relies on characterizing the unknown boundary value $g_1(t)$ *a priori* by requiring that the global relation hold. Recently, substantial progress has been made in this direction in the case of integrable nonlinear evolution equations, in particular of NLS. Namely Fokas (in press) contains an effective description of the map assigning to each given $q(x,0)=q_0(x)$ and $g_0(t)=q(0,t)$ a unique value for $q_x(0,t)$ (called the Dirichlet to Neumann map) for the NLS, as well as for a version of the Korteweg–deVries and sG equations. We state below the relevant theorem for the case of the NLS equation.

**Theorem 2** *Let $q(x,t)$ satisfy the NLS equation on the half-line $0<x<\infty, t>0$ with the initial and boundary conditions [7]. Then $g_1(t) := q_x(0,t)$ is given by*

$$\begin{aligned} g_1(t) = {} & \frac{g_0(t)}{\pi}\int_{\partial D} \mathrm{e}^{-2ik^2t}(\Phi_2(t,k)-\Phi_2(t,-k))\mathrm{d}k \\ & + \frac{4i}{\pi}\int_{\partial D} \mathrm{e}^{-2ik^2t} kR(k)\overline{\Phi_2(t,\bar{k})}\mathrm{d}k \\ & + \frac{2i}{\pi}\int_{\partial D} \mathrm{e}^{-2ik^2t}(k[\Phi_1(t,k)-\Phi_1(t,-k)]+\mathrm{i}g_0(t))\mathrm{d}k \end{aligned}$$

*with $\Phi=(\Phi_1,\Phi_2)^\tau$ given by the solution of [10]. The Neumann datum $g_1(t)$ is unique and exists globally in $t$.*

This result yields a rigorous proof of the global existence of the solution of boundary-value problems on the half-line for the NLS equation. Therefore, the assumption in Theorem 1 that a suitable function $g_1(t)$ exists can be dropped.

## Generalizations and Summary of Results

Results analogous to the ones presented in the previous section can be phrased exclusively in terms of integral equations rather than in terms of Riemann–Hilbert problems, as done for example in Khruslov and Kotlyarov (2003). This is the point of view of the school of Gelfand and Marchenko, and in this setting the functions $\Phi$ are given in the so-called Gelfand–Levitan–Marchenko representation. Results on boundary-value problems for the NLS equation using this representation have been obtained only under additional assumptions on the unknown part of the boundary values. It was only after the idea that the $x$- and $t$-parts of the spectral equations should be treated simultaneously that this approach yielded complete results. However, the Gelfand–Levitan–Marchenko representation yields a crucial simplification for deriving the explicit form of the Dirichlet to Neumann map and proving Theorem 2. This representation has now been derived for all equations [1]–[3], see Fokas (in press).

The analysis of the invariance properties of the global relation with respect to $k$ also yields the characterization of all the boundary conditions for which the transform obtained to represent the solution linearizes. For these boundary conditions, called linearizable, the solution can be represented as effectively as for the Cauchy problem. For example, the linearizable boundary conditions for the NLS equation are given by any boundary values that satisfy

$$g_0(t)\overline{g_1(t)} - \overline{g_0(t)}g_1(t) = 0$$

An example of boundary condition satisfying this constraint, encompassing also Dirichlet and Neumann homogeneous conditions, is $q(0,t) - \chi q_x(0,t) = 0$, with $\chi$ a non-negative constant.

As mentioned at the beginning of the previous section, the approach described in general can be used to obtain results similar to those given for the NLS equation for many other integrable evolution equations, in particular, mKdV (Boutet de Monvel *et al.* 2004), sG, and KdV (Fokas 2002). The results obtained are essentially the same as for NLS, starting from the general form [5] of the Lax pair, and include the derivation of the solution representation, the complete characterization of linearizable boundary conditions, and the analysis of the Dirichlet to Neumann map.

The approach above can also be used for studying boundary-value problems posed on finite domains, for $x \in [0,1]$. This has been done for a model for transient simulated Raman scattering (Fokas and Menyuk 1999), for the sG equation in light-cone coordinates (Pelloni, in press), and for the NLS equation (Fokas and Its 2004). In this case also the method yields a representation of the solution which is suitable for asymptotic analysis. In this respect, the question of soliton generation from boundary data is of some importance, and has been recently considered by various authors (Fokas and Menyuk 1999, Boutet de Monvel and Kotlyarov 2003, Pelloni in press, Boutet de Monvel *et al.* 2004). The results are however still considered case by case, and there is no general framework for this problem identified yet. For problem on the half-line, solitons may be generated but not necessarily in correspondence to the singularities that generate soliton for the full line problem, even when the same singularities are present. For problems posed on finite domains, in some specific cases at least for the simulated Raman scattering, and the sG equations, it appears that the dominant asymptotic behavior is given by a similarity solution.

In conclusion, the extension of the inverse scattering transform given by Fokas provides the tool for analyzing boundary-value problems specific to nonlinear integrable equations. This tool relies, in an essential way, on the integrability structure of the problem, and yields a full characterization of the solution as well as uniqueness and existence results. The solution representation thus obtained is not always fully explicit, but it is always suitable for asymptotic analysis using standard techniques such as the recent nonlinearization of the classical steepest descent method.

*See also:* $\bar{\partial}$ Approach to Integrable Systems; Integrable Discrete Systems; Integrable Systems and the Inverse Scattering Method; Integrable Systems: Overview; Nonlinear Schrödinger Equations; Riemann–Hilbert Methods in Integrable Systems; Separation of Variables for Differential Equations; Sine-Gordon Equation.

## Further Reading


Ablowitz MJ and Segur HJ (1974) The inverse scattering transform: semi-infinite interval. *Journal of Mathematical Physics* 16: 1054.

Adler VE, Gurel B, Gurses M, and Habibullin IT (1997) *Journal of Physics A* 30: 3505.

Bona J, Sun S, and Zhang BY (2001) A non-homogeneous boundary value problem for the Korteweg–deVries equation. *Transactions of the American Mathematical Society* 354: 427–490.

Boutet de Monvel A, Fokas AS, and Shepelsky D (2004) The modified KdV equation on the half-line. *Journal of the Institute of Mathematics of Jussieu* 3: 139–164.

Boutet de Monvel A and Kotlyarov VP (2003) Generation of asymptotic solitons of the nonlinear Schrödinger equation by boundary data. *Journal of Mathematical Physics* 44: 3185–3215.

Colin T and Ghidaglia J-M (2001) An initial-boundary value problem for the Korteweg–deVries equation posed on a finite interval. *Advanced Differential Equations* 6(12): 1463–1492.

Colliander JE and Kenig CE (2001) The generalized Korteweg–deVries equation on the half line (http://arxiv.org/abs/math.AP/0111294).

Degasperis A, Manakov S, and Santini PM (2001) The nonlinear Schrödinger equation on the half line. *JETP Letters* 74(10): 481–485.

Fokas AS (2000) On the integrability of linear and nonlinear PDEs. *Journal of Mathematical Physics* 41: 4188.

Fokas AS (2002) Integrable nonlinear evolution equations on the half line. *Communications in Mathematical Physics* 230: 1–39.

Fokas AS (2005) A generalised Dirichlet to Neumann map for certain nonlinear evolution PDEs. *Communications on Pure and Applied Mathematics* 58: 639–670.

Fokas AS and Its AR (2004) The nonlinear Schrödinger equation on the interval. *Journal of Physics A: Mathematical and General* 37: 6091–6114.

Fokas AS and Menyuk CR (1999) Integrability and self-similarity in transient stimulated Raman scattering. *Journal of Nonlinear Science* 9: 1–31.

Gardner GS, Greene JM, Kruskal MD, and Miura RM (1967) Method for solving the Korteweg–de Vries equation. *Physical Review Letters* 19: 1095.

Habibullin IT (1999) KdV equation on a half-line with the zero boundary condition. *Theoretical and Mathematical Fizika* 119: 397.

Khruslov E and Kotlyarov VP (2003) Generation of asymptotic solitons in an integrable model of stimulated Raman scattering by periodic boundary data. *Mat. Fiz. Anal. Geom.* 10(3): 366–384.

Lax PD (1968) Integrals of nonlinear equations of evolution and solitary waves. *Communications in Pure and Applied Mathematics* 21: 467–490.

Pelloni B (2005) The asymptotic behaviour of the solution of boundary value problems for the Sine–Gordon equation on a finite interval. *Journal of Nonlinear Mathematical Physics* 12: 518–529.

Sklyanin EK (1987) Boundary conditions for integrable equations. *Functional Analysis and its Applications* 21: 86–87.

Zakharov VE and Shabat AB (1972) An exact theory of two-dimensional self-focusing and one-dimensional automodulation of waves in a nonlinear medium. *Soviet Physics – JEPT* 34: 62–78.


# Braided and Modular Tensor Categories

**V Lyubashenko**, Institute of Mathematics, Kyiv, Ukraine



## Introduction

Tensor or monoidal categories are encountered in various branches of modern mathematical physics. First examples came without mentioning the name of a monoidal category as categories of modules over a group or a Lie algebra. The operation of a monoidal product in this case is the usual tensor product $X \otimes_{\mathbb{C}} Y$ of modules (representations) $X$ and $Y$. These categories are symmetric: the modules $X \otimes Y$ and $Y \otimes X$ are isomorphic; moreover, the permutation isomorphism (the twist) $\boldsymbol{c}: X \otimes Y \mapsto Y \otimes X$, $x \otimes y \to y \otimes x$, is involutive, $\boldsymbol{c}^2 = \mathrm{id}_{X \otimes Y}$. Next examples of monoidal categories were given by categories of representations of supergroups or Lie superalgebras. They are also symmetric: now the symmetry (Koszul's rule) $\boldsymbol{c}: X \otimes Y \to Y \otimes X$, $x \otimes y \mapsto (-1)^{\deg x \cdot \deg y} y \otimes x$, is the twist with a sign, which depends on the degree (or parity) $\deg x$ of elements $x \in X$.

The development of the theory of exactly solvable models in statistical mechanics led Drinfeld (1987) to the notion of quantum groups – Hopf algebras $H$ with additional structures (quasitriangular Hopf algebras). $H$-Modules also form a monoidal category; however, it is not symmetric, but only braided.

It means that a canonical braiding isomorphism $\boldsymbol{c}: X \otimes Y \to Y \otimes X$ still exists, but it is not involutive any more, $\boldsymbol{c}^2 \neq \mathrm{id}$. The braiding $\boldsymbol{c}$ satisfies the Yang–Baxter equation

$$(\boldsymbol{c} \otimes 1)(1 \otimes \boldsymbol{c})(\boldsymbol{c} \otimes 1) = (1 \otimes \boldsymbol{c})(\boldsymbol{c} \otimes 1)(1 \otimes \boldsymbol{c}) : X \otimes Y \otimes Z \to Z \otimes Y \otimes X$$

for any three $H$-modules $X$, $Y$, $Z$.

In the above examples, we also have an obvious isomorphism of associativity $\boldsymbol{a}: X \otimes (Y \otimes Z) \to (X \otimes Y) \otimes Z$ of the iterated tensor product. There are, however, monoidal categories of modules, where such an isomorphism is nontrivial, namely, modules over quasi-Hopf algebras. These were introduced by Drinfeld (1989a, b) in connection with the Knizhnik–Zamolodchikov equations. These nontrivial associativity isomorphisms $\boldsymbol{a}: X \otimes (Y \otimes Z) \to (X \otimes Y) \otimes Z$ are required to satisfy the pentagon equation of Mac Lane and Stasheff.

Braided monoidal categories also arise in rational conformal field theories (RCFTs), integrable models of statistical mechanics and topological quantum field theories (TQFTs). The common feature of these categories is that they are semisimple abelian with finite number of simple modules. In other words, such a category $\mathcal{C}$ is equivalent to the category of finite-dimensional $\mathbb{C}^n = \mathbb{C} \times \cdots \times \mathbb{C}$-modules for some $n$. However, not monoidally equivalent, the monoidal structure can be rather involved. For instance, from the Ising model one can obtain the monoidal category with two simple objects $I$ and $X$, which obey the monoidal law $\mathbb{1} \otimes \mathbb{1} = \mathbb{1}, \mathbb{1} \otimes X = X \otimes \mathbb{1} = X, X \otimes X = \mathbb{1} \oplus X$. Clearly, such relations cannot be satisfied by finite-dimensional $\mathbb{C}$-vector spaces $\mathbb{1}$ and $X$, if $\otimes$ would mean the usual tensor product $\otimes_{\mathbb{C}}$ of $\mathbb{C}$-vector spaces. However, here $\otimes$ means simply a functor $\otimes: \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ with certain properties. Categories which come from RCFT, integrable models or TQFT often enjoy additional properties. They are *rigid* – for each object $X$, there exists a dual object $X^\vee$. They are *ribbon* (balanced) – there is a canonical endomorphism $\upsilon_X : X \to X$ for each object $X$, which is related to the braiding. They are *modular*, which is defined as nondegeneracy of a certain matrix. The meaning of modularity is that the ribbon category is suitable for producing a TQFT out of it.

For categories equivalent to the category of $\mathbb{C} \times \cdots \times \mathbb{C}$-modules, the ribbon (braided) monoidal structure can be specified by a finite number of complex matrices. For instance, $6j$-symbols or $q$-$6j$-symbols encode the associativity isomorphism. In this form, modular categories appeared in the work of Moore and Seiberg (1989) on RCFTs. Such categories can be realized as categories of modules over *weak Hopf algebras*, but we stress again that the monoidal product for such modules does not coincide with the tensor product of vector spaces. So, general features are better seen at the level of category theory, and we now start with precise definitions.

## Rigid Monoidal Categories

We recall here the basic definitions of monoidal categories, monoidal functors, and dual objects.

**Definition 1** A monoidal category $(\mathcal{C}, \otimes, \boldsymbol{a}, \mathbb{1}, \boldsymbol{l}, \boldsymbol{r})$ is a category $\mathcal{C}$, a functor $\otimes: \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ (called the tensor product), a functorial isomorphism $\boldsymbol{a}: X \otimes (Y \otimes Z) \to (X \otimes Y) \otimes Z$, the associativity isomorphism, a unit object $\mathbb{1}$, and two functorial isomorphisms $\boldsymbol{l}: \mathbb{1} \otimes X \to X, \boldsymbol{r}: X \otimes \mathbb{1} \to X$ such that

$$\begin{array}{ccc} X \otimes (Y \otimes (Z \otimes W)) \xrightarrow{a} (X \otimes Y) \otimes (Z \otimes W) & \xrightarrow{a} & ((X \otimes Y) \otimes Z) \otimes W \\ {\scriptstyle X \otimes a}\downarrow & & \uparrow{\scriptstyle a \otimes W} \\ X \otimes ((Y \otimes Z) \otimes W) & \xrightarrow{\quad a \quad} & (X \otimes (Y \otimes Z)) \otimes W \end{array}$$

commutes (the pentagon equation) and

$$\boldsymbol{a}_{X,\mathbb{1},Y} = \left( X \otimes (\mathbb{1} \otimes Y) \xrightarrow{X \otimes l_Y} X \otimes Y \xrightarrow{r_X^{-1} \otimes Y} (X \otimes \mathbb{1}) \otimes Y \right)$$

**Definition 2** A monoidal functor $(F, \phi, \boldsymbol{f}): (\mathcal{C}, \otimes) \to (\mathcal{D}, \otimes)$ is a functor $F: \mathcal{C} \to \mathcal{D}$, a functorial isomorphism $\phi = \phi_{X,Y} : F(X) \otimes F(Y) \to F(X \otimes Y) \in \mathcal{D}$, and an isomorphism $\boldsymbol{f}: \mathbb{1} \to F\mathbb{1} \in \mathcal{D}$ such that

$$\begin{array}{ccc} FX \otimes (FY \otimes FZ) \xrightarrow{1 \otimes \phi} FX \otimes F(Y \otimes Z) & \xrightarrow{\phi} & F(X \otimes (Y \otimes Z)) \\ \boldsymbol{a}\downarrow & & \downarrow F\boldsymbol{a} \\ (FX \otimes FY) \otimes FZ \xrightarrow{\phi \otimes 1} F(X \otimes Y) \otimes Z & \xrightarrow{\phi} & F((X \otimes Y) \otimes Z) \end{array}$$

$$\begin{array}{ccc} F\mathbb{1} \otimes FX \xrightarrow{\phi} F(\mathbb{1} \otimes X) & & FX \otimes F\mathbb{1} \xrightarrow{\phi} F(X \otimes \mathbb{1}) \\ {\scriptstyle f \otimes 1}\uparrow \qquad \downarrow{\scriptstyle Fl,} & & {\scriptstyle 1 \otimes f}\uparrow \qquad \downarrow{\scriptstyle Fr} \\ \mathbb{1} \otimes FX \xrightarrow{\ l\ } FX & & FX \otimes \mathbb{1} \xrightarrow{\ r\ } FX \end{array}$$

commute. A morphism of monoidal functors $\lambda: (F, \phi, \boldsymbol{f}) \to (G, \psi, \boldsymbol{g})$ is a functorial morphism $\lambda: F \to G$ such that

$$\begin{array}{ccc} FX \otimes FY & \xrightarrow{\phi} & F(X \otimes Y) \\ {\scriptstyle \lambda \otimes \lambda}\downarrow & & \downarrow{\scriptstyle \lambda} \\ GX \otimes GY & \xrightarrow{\psi} & G(X \otimes Y) \end{array}$$

$$g = (\mathbb{1} \xrightarrow{f} F\mathbb{1} \xrightarrow{\lambda} G\mathbb{1})$$

The $f$ datum of a monoidal functor $(F, \phi, f)$ is uniquely determined by the $(F, \phi)$ data, so we can denote a monoidal functor as $(F, \phi)$ or even $F$.

The coherence theorem of Mac Lane (1963) states that any monoidal category $\mathcal{C}$ is equivalent to a strictly monoidal category, in which $X \otimes (Y \otimes Z) = (X \otimes Y) \otimes Z, \mathbb{1} \otimes X = X = X \otimes \mathbb{1}$, and the isomorphisms $\boldsymbol{a}, \boldsymbol{l}, \boldsymbol{r}$ are identity isomorphisms. Thus, in theoretical constructions, one may ignore the associativity isomorphism. It is not always so in practice. For instance, working with quasi-Hopf algebras related with the Knizhnik–Zamolodchikov equation one prefers to keep the original category, which is (a deformation of) the category of modules over a Lie algebra, rather than to replace it with a strict monoidal category, that is not a category of modules any more.

**Definition 3** A rigid category $\mathcal{C}$ is a monoidal category in which, to every object $X \in \mathcal{C}$, dual objects $X^\vee$ and ${}^\vee X \in \mathcal{C}$ are assigned together with morphisms of evaluation and coevaluation

$$\mathrm{ev}_X : X \otimes X^\vee \to \mathbb{1} = X \bigcup X^\vee$$

$$\mathrm{ev}'_X : {}^\vee X \otimes X \to \mathbb{1} = {}^\vee X \bigcup X$$

$$\mathrm{coev}_X : \mathbb{1} \to X^\vee \otimes X = X^\vee \bigcap X$$

$$\mathrm{coev}'_X : \mathbb{1} \to X \otimes {}^\vee X = X \bigcap {}^\vee X$$

The evaluations and coevaluations are chosen such that the compositions

$$X \xrightarrow{r^{-1}} X \otimes \mathbb{1} \xrightarrow{1\otimes\mathrm{coev}} X \otimes (X^\vee \otimes X) \xrightarrow{a} (X \otimes X^\vee) \otimes X \xrightarrow{\mathrm{ev}\otimes 1} \mathbb{1} \otimes X \xrightarrow{1} X$$

$$X \xrightarrow{1^{-1}} \mathbb{1} \otimes X \xrightarrow{\mathrm{coev}'\otimes 1} (X \otimes {}^\vee X) \otimes X \xrightarrow{a^{-1}} X \otimes ({}^\vee X \otimes X) \xrightarrow{1\otimes\mathrm{ev}'} X \otimes \mathbb{1} \xrightarrow{r} X$$

$$X^\vee \xrightarrow{1^{-1}} \mathbb{1} \otimes X^\vee \xrightarrow{\mathrm{coev}\otimes 1} (X^\vee \otimes X) \otimes X^\vee \xrightarrow{a^{-1}} X^\vee \otimes (X \otimes X^\vee) \xrightarrow{1\otimes\mathrm{ev}} X^\vee \otimes \mathbb{1} \xrightarrow{r} X^\vee$$

$${}^\vee X \xrightarrow{r^{-1}} {}^\vee X \otimes \mathbb{1} \xrightarrow{1\otimes\mathrm{coev}'} {}^\vee X \otimes (X \otimes {}^\vee X) \xrightarrow{a} ({}^\vee X \otimes X) \otimes {}^\vee X \xrightarrow{\mathrm{ev}'\otimes 1} \mathbb{1} \otimes {}^\vee X \xrightarrow{1} {}^\vee X$$

are all identity morphisms.

In a rigid monoidal category $\mathcal{C}$, there is a pairing

$$(X \otimes Y) \otimes (Y^\vee \otimes X^\vee) \xrightarrow{\sim} (X \otimes (Y \otimes Y^\vee)) \otimes X^\vee \xrightarrow{X\otimes\mathrm{ev}\otimes X^\vee} (X \otimes \mathbb{1}) \otimes X^\vee \xrightarrow{r\otimes X^\vee} X \otimes X^\vee \xrightarrow{\mathrm{ev}} \mathbb{1}$$

which induces an isomorphism $j_{+X,Y} : Y^\vee \otimes X^\vee \to (X \otimes Y)^\vee$, such that the above pairing coincides with

$$(X \otimes Y) \otimes (Y^\vee \otimes X^\vee) \xrightarrow{1\otimes j_+} (X \otimes Y) \otimes (X \otimes Y)^\vee \xrightarrow{\mathrm{ev}} \mathbb{1}$$

The equation

$$\mathrm{coev}_{X\otimes Y} = \Big(\mathbb{1} \xrightarrow{\mathrm{coev}_Y} Y^\vee \otimes Y \simeq Y^\vee \otimes \mathbb{1} \otimes Y \xrightarrow{1\otimes\mathrm{coev}_X\otimes 1} Y^\vee \otimes X^\vee \otimes X \otimes Y \xrightarrow{j_+\otimes 1} (X \otimes Y)^\vee \otimes (X \otimes Y)\Big)$$

also holds. Similarly, there is an isomorphism $j_{-X,Y} : {}^\vee Y \otimes {}^\vee X \to {}^\vee(X \otimes Y)$.

Morphisms constructed from braidings and (co)-evaluations are often described by tangles. The conventions are listed in Figure 1. The suggested assignment of morphisms in $\mathcal{C}$ to elementary pictures extends to a unique functor $\Phi$ from the category of $\mathcal{C}$-colored tangles to the category $\mathcal{C}$ itself. With the above interpretation, these tangles need not be oriented. We shall use the same notation for framed tangles, and the framing will be within the plane.

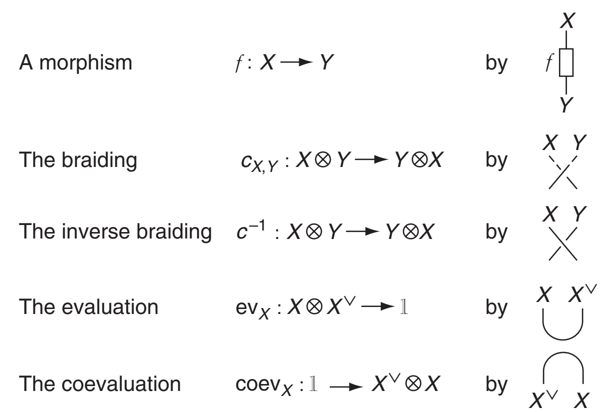
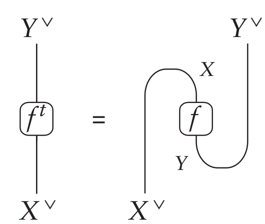
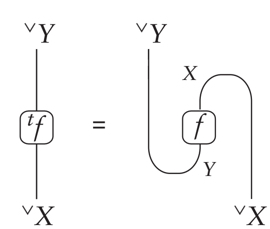



**Figure 1** Conventions for notation of morphisms from tangles.

The maps $\mathrm{Ob}\,\mathcal{C} \to \mathrm{Ob}\,\mathcal{C}, X \mapsto X^\vee$, and $X \mapsto {}^\vee X$ extend to contravariant self-equivalences $\mathcal{C} \to \mathcal{C}$, $f \mapsto f^{\mathrm{t}}$, and $f \mapsto {}^{\mathrm{t}}f$. For given $f$, the morphisms $f^{\mathrm{t}}$ and ${}^{\mathrm{t}}f$ can be defined, respectively, by the following pictures using the assignment from Figure 1:




We have a monoidal self-equivalence of $\mathcal{C}$,

$$(-^{\vee\vee}, j_2) : (\mathcal{C}, \otimes, \mathbb{1}) \to (\mathcal{C}, \otimes, \mathbb{1}),\ X \mapsto X^{\vee\vee},\ f \mapsto f^{\mathrm{tt}}$$

$$j_{2X,Y} = \left(X^{\vee\vee} \otimes Y^{\vee\vee} \xrightarrow{j_+} (Y^\vee \otimes X^\vee)^\vee \xrightarrow{j_+^{-1\mathrm{t}}} (X \otimes Y)^{\vee\vee}\right) \qquad [1]$$

It is not always true that the two duals $X^\vee$ and ${}^\vee X$ are isomorphic. However, there are canonical isomorphisms

$$X \to {}^\vee(X^\vee), \qquad X \to ({}^\vee X)^\vee$$

We may replace the category $\mathcal{C}$ with an equivalent one, such that the above isomorphisms become identity morphisms, and the functors $-^\vee$ and ${}^\vee-$ are inverse to each other. We shall assume this to simplify notations. Finally, we denote the iterated duals by $X^{(n\vee)} = X^{\vee\cdots\vee}$ ($n$ times) and $X^{(-n\vee)} = {}^{\vee\cdots\vee}X$ ($n$ times) for $n \geq 0$.

## Braided Categories

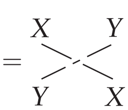
Here we review the definitions of the braiding isomorphism and further derived isomorphisms. Several basic relations between them are listed. Two important classes of examples of braided categories are given by the categories of modules over quasitriangular Hopf algebras and the categories of tangles.

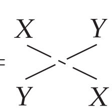
**Definition 4** A braided category $(\mathcal{C}, c)$ is a monoidal category $\mathcal{C}$ equipped with a functorial isomorphism $c = c_{X,Y} : X \otimes Y \to Y \otimes X$ – the braiding, or the commutativity isomorphism – such that the two hexagons commute,

$$\begin{array}{ccccc} X\otimes(Y\otimes Z) & \xrightarrow{1\otimes c^{\pm1}} & X\otimes(Z\otimes Y) & \xrightarrow{a} & (X\otimes Z)\otimes Y \\ a\downarrow & & & & \downarrow c^{\pm1}\otimes 1 \\ (X\otimes Y)\otimes Z & \xrightarrow{c^{\pm1}} & Z\otimes(X\otimes Y) & \xrightarrow{a} & (Z\otimes X)\otimes Y \end{array}$$

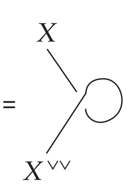
(one for $c$ and one for $c^{-1}$).

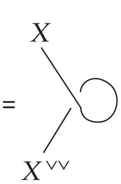
The graphical notation for the braiding and its inverse is

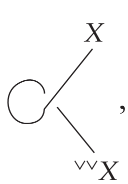
$c = (c_{X,Y} : X \otimes Y \to Y \otimes X) =$ 

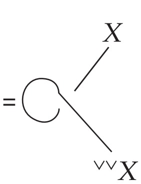

$c =$ 


In a rigid braided category, we can define functorial isomorphisms using again the conventions from Figure 1:

$u_1^2 =$ 

, $u_{-1}^2 =$ 


$u_{-1}^{-2} =$ 

, $u_{-1}^{-2} =$ 


These are isomorphisms of monoidal functors (see [1])

$$u_1^2 : (\mathrm{Id}, c^{-2}) \longrightarrow (-^{\vee\vee}, j_2)$$
$$u_{-1}^2 : (\mathrm{Id}, c^2) \longrightarrow (-^{\vee\vee}, j_2)$$

In particular, this implies the commutativity of the diagram

$$\begin{array}{ccc} X\otimes Y & \xrightarrow{c^{-2}} & X\otimes Y \\ u_1^2\otimes u_1^2\downarrow & & \downarrow u_1^2 \\ X^{\vee\vee}\otimes Y^{\vee\vee} & \xrightarrow{j_2} & (X\otimes Y)^{\vee\vee} \end{array}$$

The square of the monoidal functor $(-^{\vee\vee}, j_2)$ is

$$(-^{\vee\vee\vee\vee}, j_4) : (\mathcal{C}, \otimes, \mathbb{1}) \longrightarrow (\mathcal{C}, \otimes, \mathbb{1}),$$
$$X \mapsto X^{\vee\vee\vee\vee}, \quad f \mapsto f^{\mathrm{tttt}}$$

where

$$j_{4X,Y} = \left( X^{\vee\vee\vee\vee} \otimes Y^{\vee\vee\vee\vee} \xrightarrow{j_2} (X^{\vee\vee} \otimes Y^{\vee\vee})^{\vee\vee} \xrightarrow{j_2^{\mathrm{tt}}} (X \otimes Y)^{\vee\vee\vee\vee} \right)$$

The natural isomorphism $u_0^4 = u_{-1}^2 \circ u_1^2$ is, in fact, an isomorphism of monoidal functors $u_0^4 : (\mathrm{Id}, \mathrm{id}) \to (-^{\vee\vee\vee\vee}, j_4)$.

## Ribbon Categories

Now we define balancing and recall some properties of balanced (ribbon) categories.

**Definition 5** Let $\mathcal{C}$ be a rigid braided category. A balancing $\beta_X : X \to X^{\vee\vee}$ is an isomorphism of monoidal functors $\beta : (\mathrm{Id}, \mathrm{id}, \mathrm{id}) \to (-^{\vee\vee}, j_2, d_2)$ such that $\beta^2 = u_0^4$ and $\beta_X^{\mathrm{t}} = \beta_{X^\vee}^{-1} : X^{\vee\vee\vee} \to X^\vee$. The category $\mathcal{C}$ equipped with a balancing is called balanced.

We also use the notation $u_0^2 = \beta$. In any balanced category, there exists a canonical ribbon twist $v$. A *ribbon twist* $v = v_X : X \to X, v : \mathrm{Id} \to \mathrm{Id}$ is a self-adjoint $(v_{X^\vee} = v_X^{\mathrm{t}})$ automorphism of the identity functor such that $c^2 = (v_X^{-1} \otimes v_Y^{-1}) \circ v_{X\otimes Y}$. It can be determined from the equations

$$u_0^2 = u_1^2 \circ v^{-1} = u_{-1}^2 \circ v : X \to X^{\vee\vee}$$
$$\beta^{-1} = u_0^{-2} = u_1^{-2} \circ v^{-1} = u_{-1}^{-2} \circ v : X \to {}^{\vee\vee}X$$

In particular, its square is given by the canonical isomorphism $v^2 = u_1^{-2} \circ u_1^2$. Conversely, in any rigid braided category with a ribbon twist (called *ribbon category*) there exists a canonical balancing $u_0^2$ given by the above formulas. Thus, ribbon categories and balanced categories are synonyms.

In the case of $X = \mathbb{1}$, we have $v_{\mathbb{1}} = \mathrm{id}_{\mathbb{1}}$.

The following result can be used to simplify notations:

**Proposition 1** *For any ribbon category $\mathcal{C}$ there exists a ribbon category $\mathcal{D}$ equivalent to $\mathcal{C}$ such that in it*

(i) $\mathbb{1}^\vee = \mathbb{1}$;
(ii) *for any object $X$ we have* ${}^\vee X = X^\vee, X^{\vee\vee} = X$, *and* $\beta_X = \mathrm{id}_X : X \to X^{\vee\vee} = X$.
(iii) *for any object $X$ we have* $\mathrm{ev}_X = \mathrm{ev}'_{X^\vee} : X \otimes X^\vee \to \mathbb{1}$, *and* $\mathrm{coev}_X = \mathrm{coev}'_{X^\vee} : \mathbb{1} \to X^\vee \otimes X$.

In the category $\mathcal{C} = H$-mod, where $H$ is a ribbon Hopf algebra, the equation $X^\vee = {}^\vee X$ is not necessarily satisfied. Nevertheless, $X^\vee$ is canonically isomorphic to ${}^\vee X$. The same holds in any ribbon category. We identify these objects via $\beta = \boldsymbol{u}_0^2$: ${}^\vee X \to X^\vee$. This allows us to use the right dual objects in place of the left ones. In that role, the right duals are equipped with the left evaluation and coevaluation, called flipped evaluation and coevaluation, respectively:

$$\widetilde{\mathrm{ev}} : X^\vee \otimes X \xrightarrow{X^\vee \otimes \beta} X^\vee \otimes X^{\vee\vee} \xrightarrow{\mathrm{ev}} \mathbb{1}$$

$$\widetilde{\mathrm{coev}} : \mathbb{1} \xrightarrow{\mathrm{coev}} X^{\vee\vee} \otimes X^\vee \xrightarrow{\beta^{-1} \otimes X^\vee} X \otimes X^\vee$$

They are often denoted simply ev and coev and should be replaced by $\widetilde{\mathrm{ev}}$ and $\widetilde{\mathrm{coev}}$ in applications. In the context of Hopf algebra, $\beta$ is given by the action of a group-like element introduced by Drinfeld.

## Hopf Algebras in Braided Categories

Let $\mathcal{C}$ be a braided monoidal category. A Hopf algebra $H$ in $\mathcal{C}$ is an object $H \in \mathrm{Ob}\,\mathcal{C}$ together with an associative multiplication $m : H \otimes H \to H$ and an associative comultiplication $\Delta : H \to H \otimes H$, obeying the bialgebra axiom

$$\left(H \otimes H \xrightarrow{m} H \xrightarrow{\Delta} H \otimes H\right)$$
$$= \left(H \otimes H \xrightarrow{\Delta \otimes \Delta} H \otimes H \otimes H \otimes H \xrightarrow{H \otimes c \otimes H} H \otimes H \otimes H \otimes H \xrightarrow{m \otimes m} H \otimes H\right)$$

Moreover, $H$ has a unit $\eta : \mathbb{1} \to H$, a counit $\varepsilon : H \to \mathbb{1}$, an antipode $\gamma : H \to H$, and the inverse antipode $\gamma^{-1} : H \to H$. The defining relations for these are the same as in the classical case. Notice, in particular, that the unit is also a morphism. Associativity of multiplication, as well as coassociativity of comultiplication, is formulated with the use of associativity isomorphism (in the nonstrict case).

Hopf algebras in braided categories have also been called braided groups. Their basic properties are very similar to those of usual Hopf algebras, for example, the antipode is antimultiplicative with respect to the braiding (see, e.g., Majid (1993)). For Hopf algebras in rigid braided categories, there exist integrals in a sense very much similar to the case of ordinary finite-dimensional Hopf algebras, as shown by Bespalov *et al.* (2000).

## Modular Categories

Assume that a braided rigid monoidal category $\mathcal{C}$ is equivalent as a category (with monoidal structure ignored) to the category of finite-dimensional modules over a finite-dimensional algebra. In particular, $\mathcal{C}$ is abelian. Then there exists an object $F$ in $\mathcal{C}$, equipped with a morphism $i_X : X \otimes X^\vee \to F$ for each $X \in \mathrm{Ob}\,\mathcal{C}$, such that the diagram

$$\begin{array}{ccc} X \otimes Y^\vee & \xrightarrow{f \otimes Y^\vee} & Y \otimes Y^\vee \\ {\scriptstyle X \otimes f^t}\downarrow & & \downarrow{\scriptstyle i_Y} \\ X \otimes X^\vee & \xrightarrow{i_X} & F \end{array}$$

is commutative for all morphisms $f : X \to Y$ of $\mathcal{C}$, and, moreover, $F$ is universal between objects with such properties. Here $f^t : Y^\vee \to X^\vee$ is the transpose of a morphism $f : X \to Y$. In other words, $F$ is a direct limit, called the coend and denoted as $F = \int^{Z \in \mathcal{C}} Z \otimes Z^\vee$. It can also be defined via an exact sequence

$$\bigoplus_{f : X \to Y \in \mathcal{C}} X \otimes Y^\vee \xrightarrow{f \otimes Y^\vee - X \otimes f^t} \bigoplus_{Z \in \mathcal{C}} Z \otimes Z^\vee \xrightarrow{\oplus i_Z} F \to 0$$

It turns out that the coend $F$ is a Hopf algebra in the braided category $\mathcal{C}$, when it is equipped with the following operations. The comultiplication in $F$ is uniquely determined by the equation

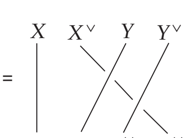
$$\left(X \otimes X^\vee \xrightarrow{i_X} F \xrightarrow{\Delta} F \otimes F\right)$$
$$= \left(X \otimes X^\vee = X \otimes \mathbb{1} \otimes X^\vee \xrightarrow{X \otimes \mathrm{coev} \otimes X^\vee} X \otimes X^\vee \otimes X \otimes X^\vee \xrightarrow{i_X \otimes i_X} F \otimes F\right)$$

The counit in $F$ is determined by the equation

$$\left(X \otimes X^\vee \xrightarrow{i_X} F \xrightarrow{\varepsilon} \mathbb{1}\right) = \left(X \otimes X^\vee \xrightarrow{\mathrm{ev}} \mathbb{1}\right)$$

The multiplication $m : F \otimes F \to F$ is defined by the following diagram:

$m$ =




and

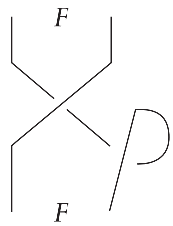
$$\begin{array}{ccc} X \otimes X^\vee \otimes (Y \otimes Y^\vee) & \xrightarrow{i_X \otimes i_Y} & F \otimes F \\ {\scriptstyle X \otimes c}\downarrow & & \exists\downarrow{\scriptstyle m} \\ X \otimes Y \otimes (X \otimes Y)^\vee & \xrightarrow{i_{X \otimes Y}} & F \end{array}$$

The unit is given by the morphism

$$\eta : \mathbb{1} = \mathbb{1} \otimes \mathbb{1}^{\vee} \xrightarrow{i_{\mathbb{1}}} F$$

The diagram corresponding to the antipode $\gamma_F : F \to F$ is given by

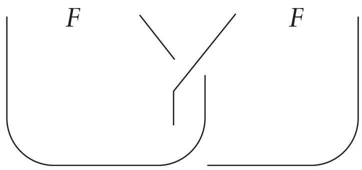
$\gamma_F =$




The structure of the coend $F$ as a Hopf algebra can also be found directly from its universal property, as in Majid (1993).

There is a pairing of Hopf algebras $\omega : F \otimes F \to \mathbb{1}$ in $\mathcal{C}$:

$\omega =$




It induces a homomorphism of Hopf algebras $F \to F^{\vee}$.

**Definition 6** A ribbon category $\mathcal{C}$, equivalent as a category to the category of finite-dimensional modules over a finite-dimensional algebra, is called modular if the pairing $\omega$ is nondegenerate, that is, the induced morphism $F \to F^{\vee}$ is invertible.

Examples of nonsemisimple modular categories include $\mathcal{C} = H$-mod, where $H = u_q(\mathfrak{g})$ is a finite-dimensional algebra, quotient of the quantum universal enveloping algebra $U_q(\mathfrak{g})$, and $q$ is a root of unity of odd degree. In these examples, the coalgebra $F$ identifies with the dual Hopf algebra $H^*$, but the multiplication in $F$ differs from that of $H^*$. Explicit formula for the multiplication in $F$ uses the $R$-matrix for $H$ (see, e.g., Majid (1993)). A definition of modularity for another type of categories (not necessarily abelian) was given by Turaev (1994).

When the category $\mathcal{C}$ is modular, the integrals for the Hopf algebra $F$ have especially simple properties. The integral element in $F$ is two sided. It is a morphism $\mu : \mathbb{1} \to F$ such that

$$\left(F = F \otimes \mathbb{1} \xrightarrow{1 \otimes \mu} F \otimes F \xrightarrow{m} F\right)$$
$$= \left(F \xrightarrow{\varepsilon} \mathbb{1} \xrightarrow{\mu} F\right)$$
$$= \left(F = \mathbb{1} \otimes F \xrightarrow{\mu \otimes 1} F \otimes F \xrightarrow{m} F\right)$$

and $\mu$ is universal between morphisms with such property. By duality, the integral functional $\lambda : F \to \mathbb{1}$ is also two sided. It satisfies

$$\left(F \xrightarrow{\Delta} F \otimes F \xrightarrow{1 \otimes \lambda} F \otimes \mathbb{1} = F\right)$$
$$= \left(F \xrightarrow{\lambda} \mathbb{1} \xrightarrow{\eta} F\right)$$
$$= \left(F \xrightarrow{\Delta} F \otimes F \xrightarrow{\lambda \otimes 1} \mathbb{1} \otimes F = F\right)$$

and is universal between morphisms with such property. The integral element and the integral functional are unique up to a multiplication by an element of $\mathrm{Aut}_{\mathcal{C}}\, \mathbb{1}$.

## Semisimple Abelian Modular Categories

Reshetikhin and Turaev proposed to construct invariants of 3-manifolds via quantum groups. More precisely, they use certain abelian semisimple ribbon categories obtained from quantum groups at roots of unity as trace quotients. One can forget about the origin of these categories and work simply with semisimple modular categories. We shall describe them as input data for the modular functor construction.

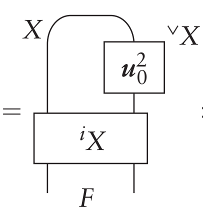
Let $\mathcal{C}$ be a $\mathbb{C}$-linear abelian semisimple modular ribbon category. Assume that the number of isomorphism classes of simple objects is finite. Assume also that $\mathbb{1}$ is simple and for each simple object $X$ the endomorphism algebra End $X = \mathbb{C}$. We denote by $\mathcal{S} = \{X_i\}_i$ the list of (representatives of isomorphism classes of) all simple objects.

Under these assumptions, many formulas simplify. The coend $F \in \mathcal{C}$ takes the form

$$F = \bigoplus_{X \in \mathcal{S}} X \otimes X^{\vee} \in \mathcal{C}$$

Any morphism $\mathbb{1} \to F$ is a $\mathbb{C}$-linear combination of the standard morphisms for $X \in \mathcal{S}$,

$\phi_X =$




$$: \mathbb{1} \xrightarrow{\mathrm{coev}} X \otimes {}^{\vee}X \xrightarrow{1 \otimes u_0^2} X \otimes X^{\vee} \xrightarrow{i} F$$

The morphisms $\phi_X$ form a basis of the commutative algebra $\mathrm{Inv}\, F = \mathrm{Hom}_{\mathcal{C}}(\mathbb{1}, F)$. The Grothendieck ring of the category $\mathcal{C}$ determines the multiplication law in $\mathrm{Inv}\, F$ via the algebra isomorphism $\mathbb{C} \otimes_{\mathbb{Z}} K_0(\mathcal{C}) \to \mathrm{Inv}\, F, [X] \mapsto \phi_X$.

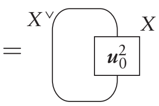
Any morphism $F \to \mathbb{1}$ can be represented as a linear combination of the morphisms

$$\psi_X : F \xrightarrow{\mathrm{pr}_X} X \otimes X^{\vee} \xrightarrow{\mathrm{ev}_X} \mathbb{1}$$

where $X \in \mathcal{S}$. The functional $\psi_1 : F \to \mathbb{1}$ satisfies the properties of a two-sided integral $\lambda$ of the braided Hopf algebra $F$.

### The Verlinde Formula

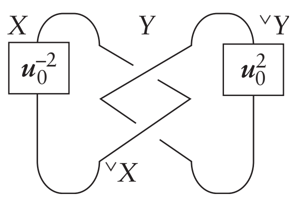
The number

$\dim_q(X) =$ 


$$: \mathbb{1} \xrightarrow{\text{coev}} X^\vee \otimes X \xrightarrow{1\otimes u_0^2} X^\vee \otimes X^{\vee\vee} \xrightarrow{\text{ev}} \mathbb{1}$$

is called the dimension of an object $X \in \mathrm{Ob}\,\mathcal{C}$. (The index $q$ reminds us that this number coincides with the $q$-dimension in the case $\mathcal{C} = U_q(\mathfrak{g})$-mod.) We have $\dim_q(X^\vee) = \dim_q(X)$.

**Definition 7** Introduce a biadditive function of two variables $\mathsf{s} : \mathrm{Ob}\,\mathcal{C} \times \mathrm{Ob}\,\mathcal{C} \to \mathbb{C}$ on the class of objects of $\mathcal{C}$:

$\mathsf{s}_{XY} =$ 


In particular, its restriction to $\mathcal{S}$ is a matrix $\mathsf{s}|_{\mathcal{S}} : \mathcal{S} \times \mathcal{S} \to \mathbb{C}$, denoted again by $\mathsf{s} = (\mathsf{s}_{XY})_{X,Y\in\mathcal{S}}$ by abuse of notation; here $X$ and $Y$ run over simple objects.

Notice that $\mathsf{s}_{XY} = \mathsf{s}_{YX}$, so the matrix $\mathsf{s}$ is symmetric. Let us consider the $\mathbb{C}$-algebra $\mathrm{Inv}\,F = \mathrm{Hom}_{\mathcal{C}}(\mathbb{1}, F)$. It has the basis $\phi_X, X \in \mathcal{S}$; hence, it is $n$-dimensional, where $n = \mathrm{Card}\,\mathcal{S}$. The form $\omega$ on $F$ induces a bilinear form

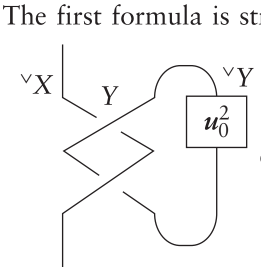
$$\omega' : \mathrm{Inv}\,F \times \mathrm{Inv}\,F \xrightarrow{\otimes} \mathrm{Hom}(\mathbb{1}, F \otimes F) \xrightarrow{\mathrm{Hom}(1,\omega)} \mathbb{1}$$

The matrix $(\mathsf{s}_{XY})$ is the matrix of the form $\omega'$ in the basis $(\phi_X)$.

**Lemma 1** (The Verlinde formula) *For any simple $X \in \mathcal{S}$ and any objects $Y$ and $Z$ of $\mathcal{C}$, we have*

$$\mathsf{s}_{X1} = \dim_q(X), \qquad \mathsf{s}_{X1}\mathsf{s}_{X,Y\otimes Z} = \mathsf{s}_{XY}\mathsf{s}_{XZ} \qquad [2]$$

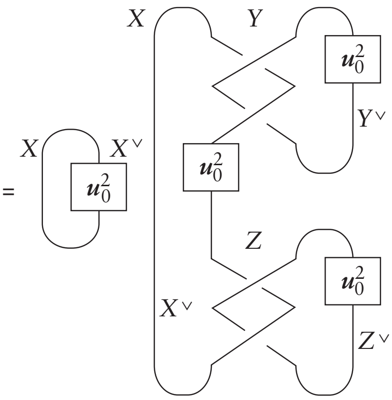
**Proof** The first formula is straightforward. Since

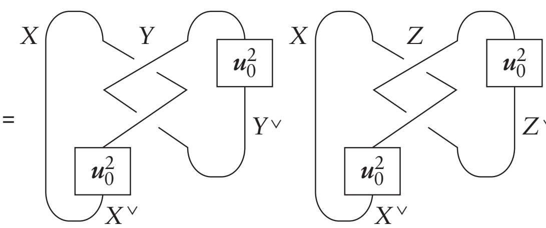


$\in \mathrm{End}\,^\vee X \simeq \mathbb{C}$

is a number, we can move it from the second factor to the first in the following computation:

$\mathsf{s}_{X1}\mathsf{s}_{X,Y\otimes Z}$

$=$ 


$=$ 


$= \mathsf{s}_{XY}\mathsf{s}_{XZ}$

This proves the second formula. □

**Proposition 2** (Criterion of modularity) *In the above assumption of semisimplicity, the following conditions are equivalent:*

(i) *$\mathcal{C}$ is modular ($\omega$ is nondegenerate);*
(ii) *the matrix $(\mathsf{s}_{XY})_{X,Y\in\mathcal{S}}$ is nondegenerate;*
(iii) *for any $X \in \mathcal{S}$ its dimension $\dim_q X$ does not vanish, and there exist numbers $\mu'_Y, Y \in \mathcal{S}$, such that for all $X \in \mathcal{S}$ we have $\sum_{Y\in\mathcal{S}} \mathsf{s}_{XY}\mu'_Y = \delta_{X1}$; and*
(iv) *for each simple $X \not\simeq \mathbb{1}$ we have $\sum_{Y\in\mathcal{S}} \mathsf{s}_{XY} \dim_q Y = 0$ and $\dim_q X \neq 0$.*

The easy implication (ii) $\Longrightarrow$ (iii) can be deduced from the Verlinde formula. If the dimension $\dim_q(X) = \mathsf{s}_{X1}$ of a simple object $X$ vanishes, then $\mathsf{s}_{XY}^2 = 0$ for all $Y \in \mathrm{Ob}\,\mathcal{C}$. This contradicts to the assumption of nondegeneracy of $(\mathsf{s}_{XY})$.

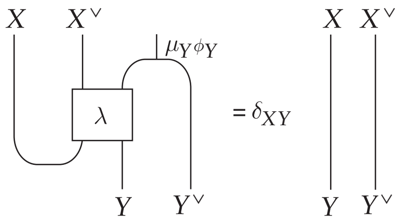
Let us determine the coefficients $\mu_Y$ of the integral element

$$\mu = \sum_{Y\in\mathcal{S}} \mu_Y \phi_Y : \mathbb{1} \to F$$

of the Hopf algebra $F$. It also has a two-sided integral-functional $\lambda : F \to \mathbb{1}$. The corresponding endomorphism is

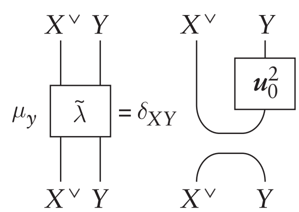
$$\tilde{\lambda}_Z = \left(Z \xrightarrow{\delta_Z} F \otimes Z \xrightarrow{\lambda\otimes Z} \mathbb{1} \otimes Z =\!= Z\right)$$

for an arbitrary object $Z$ of $\mathcal{C}$, where $\delta_Z$ is the natural coaction. The equation




[3]

follows from the properties of the two-sided integral $\lambda$ of the Hopf algebra $F$. Due to uniqueness of integrals, $\lambda$ is proportional to $\psi_1$. In eqn [3], $X$ and $Y$ vary over $\mathcal{S}$. The right-hand side is the identity morphism if $X = Y$, and vanishes otherwise. Substituting the definition of $\phi_Y$, we rewrite the equation as follows:




[4]

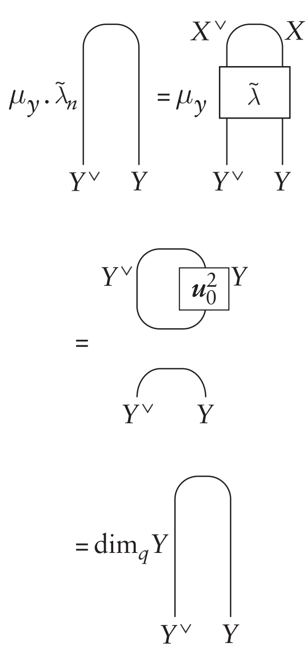
For $X = 1$, we get

$$\mu_Y \cdot \tilde{\lambda}_Y = \delta_{1Y} \cdot \mathrm{id}_Y : Y \to Y \qquad [5]$$

If $Y \not\cong \mathbb{1}$, then $\tilde{\lambda}_Y = 0$. So [5] tells essentially that

$$\mu_1 \cdot \tilde{\lambda}_1 = \mathrm{id}_1 : \mathbb{1} \to \mathbb{1} \qquad [6]$$

Now return to [4] with $X = Y$. If we compose that equation with $\mathrm{coev} : \mathbb{1} \to Y^\vee \otimes Y$, we obtain

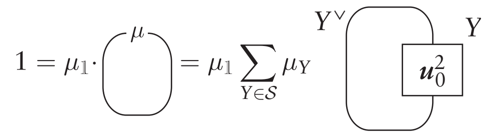



[7]

Multiplying both sides of [7] with $\mu_1$, we find

$$\mu_Y = \mu_1 \cdot \dim_q(Y)$$

The normalization is fixed by eqn [6], which we can write as




$$= \mu_1^2 \sum_{Y \in \mathcal{S}} (\dim_q(Y))^2$$

Hence,

$$(\mu_1)^2 = \left( \sum_{Y \in \mathcal{S}} (\dim_q(Y))^2 \right)^{-1} \qquad [8]$$

So, we find $\mu_1$, unique up to a sign.

## Conjugation Properties

From the Verlinde formula [2], we conclude that the commutative $\mathbb{C}$-algebra $\mathrm{Inv}\, F$ possesses homomorphisms

$$\chi_X : \mathrm{Inv}\, F \to \mathbb{C}$$
$$\phi_Y \mapsto (\dim_q(X))^{-1} s_{XY} = s_{XY}/s_{X1}$$

The matrix $s$ is invertible, so that its columns cannot be proportional. Hence, all $\chi_X$ are different characters. Their number is $n = \mathrm{Card}\, \mathcal{S} = \dim_{\mathbb{C}} F$; hence, there is an isomorphism of $\mathbb{C}$-algebras

$$\chi : \mathrm{Inv}\ F \to \mathbb{C} \times \cdots \times \mathbb{C} = \mathbb{C}^n$$
$$\phi \mapsto (\chi_1(\phi), \ldots, \chi_n(\phi))$$

Now we show that the dimensions $\dim_q(Y)$ are real numbers, so that $\mu_1$ is also a real number. One can introduce in $\mathrm{Inv}\ F$ an antilinear involution,

$$-^* : \mathrm{Inv}\, F \to \mathrm{Inv}\, F, \quad (\phi_X)^* = \phi_{X^\vee}$$

and a scalar (Hermitian) product

$$(\phi_X | \phi_Y) = \delta_{XY}, \quad X, Y \in \mathcal{S}$$

Then $\mathrm{Inv}\ F$ becomes a finite-dimensional commutative Hilbert algebra. Indeed,

$$\begin{aligned}(\phi_X \phi_Y | \phi_Z) &= \dim \mathrm{Hom}(X \otimes Y, Z)\\ &= \dim \mathrm{Hom}(X, Y^\vee \otimes Z) = (\phi_X | \phi_Y^* \phi_Z)\end{aligned}$$

From the theory of finite-dimensional commutative Hilbert algebras, we know that idempotents in the algebra $\mathrm{Inv}\ F$ are self-adjoint (only in that case the scalar product can be positive definite). Hence, $\chi$ is a $*$-morphism, that is, $\chi_X(\phi^*) = \overline{\chi_X(\phi)}$. Therefore,

$s_{XY^\vee}/s_{X\mathbb{1}} = \overline{s_{XY}}/\overline{s_{X\mathbb{1}}}$. In the particular case of $X = \mathbb{1}$, we obtain

$$\dim_q(Y) = \dim_q(Y^\vee) = s_{\mathbb{1}Y^\vee} = \overline{s_{\mathbb{1}Y}} = \overline{\dim_q(Y)}$$

since $s_{\mathbb{1}\mathbb{1}} = 1$. This proves that for any $Y \in \mathcal{C}$ its dimension $\dim_q(Y)$ is a real number.

It is natural to take for $\mu_1$ the positive root of the right-hand side of [8]. Positiveness fixes $\mu_1$ uniquely.

### Examples of Semisimple Modular Categories

In their original paper, Reshetikhin and Turaev (1991) use as algebraic input data the representation theory of the quantum deformation $U = U_q(\mathrm{sl}_2)$ of the Lie algebra $\mathrm{sl}(2, \mathbb{C})$, where $q$ is a root of unity. They construct the invariant as a trace over $U$-equivariant morphisms, and prove the necessary modularity condition concerning the nondegeneracy of the braided pairing.

The general picture is drawn by Turaev (1994), where 3-manifold invariants and TQFTs are constructed from semisimple modular categories. He shows how to obtain the latter as quotients of certain subcategories of representations of a modular Hopf algebra by the ideal of trace-negligible morphisms.

Finkelberg (1996), based on results of Gelfand and Kazhdan, establishes (via the theory of Kazhdan and Lusztig) an equivalence between two modular categories. The first is the semisimple category $\mathcal{C}$ of integrable modules over an affine Lie algebra $\hat{\mathfrak{g}}$ of positive integer level $k$. The second is a certain subquotient of the category of $U_q(\mathfrak{g})$-modules for $q = \exp(\pi i m^{-1}/(k + h^\vee))$, where $m \in \{1, 2, 3\}$ and $h^\vee$ is the dual Coxeter number of $\mathfrak{g}$. Huang and Lepowsky (1999) describe the rigid braided structure of $\mathcal{C}$ using vertex operators. Bakalov and Kirillov (2001) use geometrical constructions to make $\mathcal{C}$ into a modular category, associated with the Wess–Zumino–Witten (WZW) model. They construct the corresponding WZW modular functor.

## Modular Functor and TQFT

Modular categories give rise to a modular functor and a TQFT. The meanings of those differ from author to author, but the common features are the following. Such a TQFT is a functor from the category whose objects are smooth surfaces with additional structures and morphisms are three-dimensional manifolds with additional structures to the category of vector spaces. A modular functor is the restriction of such TQFT to the subcategory whose morphisms are homeomorphisms of surfaces. One of the constructions due to Kerler and Lyubashenko (2001) takes a nonsemisimple modular category as an input and assigns to it a double TQFT functor, that is, a functor between double categories. The target is the 2-category of abelian categories.

*See also:* Axiomatic Approach to Topological Quantum Field Theory; Hopf Algebras and $q$-Deformation Quantum Groups; The Jones Polynomial; Knot Invariants and Quantum Gravity; Quantum 3-Manifold Invariants; Symmetries in Quantum Field Theory of Lower Spacetime Dimensions; Topological Quantum Field Theory: Overview; von Neumann Algebras: Introduction, Modular Theory, and Classification Theory; von Neumann Algebras: Subfactor Theory.

## Further Reading


Bakalov B and Kirillov A Jr. (2001) *Lectures on Tensor Categories and Modular Functors,* University Lecture Series, vol. 21. Providence, RI: American Mathematical Society.

Bespalov Y, Kerler T, Lyubashenko VV, and Turaev VG (2000) Integrals for braided Hopf algebras. *Journal of Pure and Applied Algebra* 148(2): 113–164 (arXiv:math.QA/9709020).

Drinfeld VG (1987) Quantum groups. In: Gleason A (ed.) *Proceedings of the International Congress of Mathematicians (Berkeley, 1986)*, vol. 1, pp. 798–820. Providence, RI: American Mathematical Society.

Drinfeld VG (1989a) Quasi-Hopf algebras. *Algebra i Analiz* 1(6): 114–148.

Drinfeld VG (1989b) Quasi-Hopf algebras and Knizhnik–Zamolodchikov equations. In: *Problems of Modern Quantum Field Theory*, pp. 1–13. Berlin–New York: Springer.

Finkelberg M (1996) An equivalence of fusion categories. *Geometric and Functional Analysis* 6(2): 249–267.

Huang Y-Z and Lepowsky J (1999) Intertwining operator algebras and vertex tensor categories for affine Lie algebras. *Duke Mathematical Journal* 99(1): 113–134 (arXiv:q-alg/9706028) (arXiv:q-alg/9706028).

Joyal A and Street RH (1991) Tortile Yang–Baxter operators in tensor categories. *Journal of Pure and Applied Algebra* 71: 43–51.

Kerler T and Lyubashenko VV (2001) *Non-Semisimple Topological Quantum Field Theories for 3-Manifolds with Corners*, Lecture Notes in Mathematics, vol. 1765, vi + 379 pp. Heidelberg: Springer.

Mac Lane S (1971) *Categories for the Working Mathematician*, GTM, vol. 5. New York: Springer.

Majid S (1993) Braided groups. *Journal of Pure and Applied Algebra* 86(2): 187–221.

Majid S (1995) *Foundations of Quantum Group Theory*. Cambridge: Cambridge University Press.

Moore G and Seiberg N (1989) Classical and quantum conformal field theory. *Communications in Mathematical Physics* 123: 177–254.

Reshetikhin NY and Turaev VG (1991) Invariants of 3-manifolds via link polynomials and quantum groups. *Inventiones Mathematicae* 103(3): 547–597.

Turaev VG (1994) Quantum Invariants of Knots and 3-Manifolds, *de Gruyter Stud. Math*, vol. 18. Berlin–New York: Walter de Gruyter.

# Brane Construction of Gauge Theories

**S L Cacciatori**, Università di Milano, Milan, Italy



## Introduction

Branes appear in string theories and M-theory as extended objects which contain some nonperturbative information about the theory, and, apart from gravity, they can couple with gauge fields.

At low energies, M-theory can be approximated with an 11-dimensional $N=1$ supergravity, which in fact is unique and contains a graviton field (the metric $g_{\mu\nu}$), a spin 3/2 field $\psi$ (the gravitino) and a gauge field consisting of a 3-form potential field $c$. The gauge field, whose field strength is a 4-form $G=\mathrm{d}c$, can then couple electrically with two-dimensional extended objects, called M2 membranes. Moving in spacetime, an M2 membrane describes a three-dimensional world volume $W3$ so that its coupling to the gauge field is

$$S_2 = k\int_{W3} c \qquad [1]$$

$k$ representing the charge.

With $c$ we can associate a dual field $\tilde{c}$ such that $\mathrm{d}\tilde{c} = {}^*G$. It is a 6-form and can then electrically couple with a five-dimensional object, the M5 membrane. However, as $c$ is the true field, we say that M5 couples magnetically with $c$.

In superstring theories, which however are related to M-theory by a dualities web, there are many more objects to be considered. In particular, we will consider type II strings, which at low energies are described by ten-dimensional $N=2$ supergravity theories. They contain a Neveu–Schwarz sector consisting of a graviton $g_{\mu\nu}$, a 2-form potential $B_{\mu\nu}$, and a scalar field $\phi$, the dilaton. The content of the Ramond–Ramond fields depends on the chirality of the supercharges.

Type IIA strings are nonchiral (their left and right supercharges having opposite chiralities) and contain only odd-dimensional $p$-form potentials $A^{(p)}$, with $p=1,3,5,7,9$.

Type IIB strings are chiral and contain only even-dimensional $p$-form potentials $A^{(p)}$, with $p=0,2,4,6,8$.

Proceeding as before, we see that a $(p+1)$-form potential can couple electrically with a $p$-dimensional object and magnetically with a $(6-p)$-dimensional object. Such objects in fact exist in type II strings: the D$p$ branes are $p$-dimensional extended objects, with $p=0,2,4,6,8$ for IIA strings and $p=-1,1,3,5,7,9$ for IIB strings. In particular, D0 and D1 branes are called D-particles and D-strings respectively, whereas D$(-1)$ branes are instantons, that is, points in spacetime. Concretely, D-branes are extended regions in spacetime where the endpoints of open strings are constrained to live. Mathematically, they are defined imposing Dirichlet conditions (whence the "D" of D-brane) on the ends of the string, along certain spatial directions. Excitation of these string states gives rise to the dynamic of the brane. They correspond to a ten-dimensional U(1) gauge field, whose components, which are tangent to the brane world volume, give rise to a gauge field in $p+1$ dimensions, whereas the orthogonal components generate deformations of the brane shape. Moreover, if $n$ parallel $p$-branes overlap, the gauge theory on the world volume is enhanced to a U($n$) gauge theory. Closed strings can generate gravitational interactions responsible for wrappings of the brane. However, in the cases when gravitational interaction is negligible, we can use this mechanism to construct $(p+1)$-dimensional gauge theories, as we will see.

Before explaining how the construction works let us remember that there are two other interesting objects which often appear. In fact, we have not yet considered the Neveu–Schwarz $B$-field: this field can couple electrically with a one-dimensional object and magnetically with a five-dimensional object. These are the usual string (also called a fundamental or F-string) and a five-dimensional membrane called NS5 brane.

We will see how supersymmetric gauge theory configurations can be realized geometrically, considering more or less simple configurations of branes. We will also show that quantum corrections, be they exact or perturbative, can be described in this geometrical fashion. To be explicit, we will work with four-dimensional gauge theories, but it is clear that similar constructions can be done in different dimensions.

## Gauge Groups on the Branes

A deeper understanding of how D-branes and related world-volume gauge theories work requires the introduction of dualities, but a quite simple heuristic argument can be given, giving up some rigor in favor of intuition.

To set our ideas, let us think of an open string moving in a nearly flat (but ten-dimensional) spacetime. Its trajectory will describe a two-dimensional surface having a boundary traced by the ends of the string (Figure 1). The string can then be described by a map from a two-dimensional surface $\Sigma$, having a

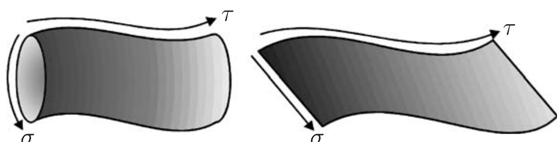



**Figure 1** Strings moving in spacetime.

boundary $\gamma = \partial\Sigma$, to spacetime, say $X^\mu(\sigma, \tau)$ with $\mu = 0, 1, \ldots, 9$. Here we chose on $\Sigma$ local coordinates $\sigma^\alpha = (\sigma, \tau)$, where $\sigma \in [0, \pi]$ is a spacelike coordinate and $\tau$ is a timelike one. Then $\sigma = 0, \pi$ individuate the ends of the string and are identified for the closed string. Now, on a given background, the string evolution is usually described as a two-dimensional (supersymmetric) conformal field theory for the fields $X^\mu(\sigma, \tau)$. The action for the bosonic part is the same for both type IIA and IIB strings, and reads

$$S[X] = \frac{1}{4\pi\alpha'} \int_\Sigma \mathrm{d}^\sigma \sqrt{h} h^{\alpha\beta} g_{\mu\nu}(X) \frac{\partial X^\mu}{\partial \sigma^\alpha} \frac{\partial X^\nu}{\partial \sigma^\beta} + \frac{1}{4\pi\alpha'} \int_\Sigma B_{\mu\nu}(X) \frac{\partial X^\mu}{\partial \sigma^\alpha} \frac{\partial X^\nu}{\partial \sigma^\beta} \mathrm{d}\sigma^\alpha \wedge \mathrm{d}\sigma^\beta \qquad [2]$$

where $g_{\mu\nu}$ and $B$ are the metric and a 2-form potential field for the given spacetime background, and $h_{\alpha\beta}$ is a metric for $\Sigma$. In general, we must also add a scalar field $\phi(X)$, but it will not play any role here. Using conformal invariance, we can reduce $h_{\alpha\beta}$ to the flat metric. Also consider a flat background $g_{\mu\nu}(X) = \eta_{\mu\nu}$ and concentrate for a moment on the $B$-field.

Conceived as a 2-form field over the spacetime, the potential field $B$ is a gauge field: its field strength 3-form $H = \mathrm{d}B$ is unchanged under a shift

$$B \longrightarrow B + \mathrm{d}A \qquad [3]$$

generated by the 1-form field $A(X)$. Here $A$ should be a totally unphysical field. However, note that if one considers open strings, the action for the $B$-field, and then the full action is shifted by a boundary term

$$S[X] \longrightarrow S[X] + \frac{1}{4\pi\alpha'} \int_\gamma A_\mu(X) \frac{\partial X^\mu}{\partial \sigma^\alpha} \mathrm{d}\sigma^\alpha \qquad [4]$$

The boundary $\gamma$ just describes the timelike world lines of the ends of the string. Thus, the ends of the string carry a U(1) charge and, even though the $B$-field vanishes, we can have the open-string action

$$S[X] = \frac{1}{4\pi\alpha'} \int_\Sigma \partial_\alpha X^\mu \partial^\alpha X_\mu \mathrm{d}^2\sigma + \int_\gamma A_\mu(X) \partial_\alpha X^\mu \mathrm{d}\sigma^\alpha \qquad [5]$$

Here we conventionally rescaled the $A$ field to normalize the action. To define the equation of motion, however, we must also specify boundary conditions for $X^\mu(\sigma, \tau)$ on $\gamma$. Let us choose Neumann conditions for $\mu = 0, 1, \ldots, p$ and Dirichlet conditions for the remaining directions

$$\partial_\sigma X^a(\gamma) = 0, \quad a = 0, \ldots, p \qquad [6]$$

$$\partial_\sigma X^i(\gamma) = 0, \quad i = p + 1, \ldots, 9 \qquad [7]$$

This means that the extrema of the string are bound on a $(p + 1)$-dimensional region (including time): the D$p$ brane. If for $\Sigma$ we consider the full strip $(\sigma, \tau) = [0, \pi] \times \mathbb{R}$ then the U(1) action reduces to

$$S_A[X] = \int_{-\infty}^{\infty} A_a \partial_\tau X^a(\pi, \tau) - \int_{-\infty}^{\infty} A_a \partial_\tau X^a(0, \tau) \qquad [8]$$

Thus, only the components of $A_a$ tangent to the brane interact with the ends of the strings. What about the normal components $A_i$?

To understand its meaning, let us proceed to compute the mean momentum transferred by the string, as it would be rigid. Imitating the Hamilton–Jacobi procedures for particles, let us consider the action up to a fixed time, say $\tau = 0$, so that $\Sigma = [0, \pi] \times [-\infty, 0]$. It is then a function of the position $X^\mu(\sigma, 0)$ of the string at the instant $\tau = 0$. To compute the momentum, we must vary the action by changing the position by a constant shift $\delta X^\mu(\sigma) = \Delta_0^\mu$. The variation will then contain some boundary terms which, for reasons of consistency, we must make vanish.

Before doing such a computation, let us make some further comments. It is plausible to assume that the two ends of the string could be charged for different U(1) fields. To the states of the open string we can in fact add two discrete labels $I, J = 1, \ldots, n$, for some integer $n$, called Chan–Paton factors, and referring, respectively, to the two ends of the string. We will indicate the ends of the string as $X^\mu(0, \tau; I)$ and $X^\mu(\pi, \tau; J)$ when we need to specify the states. If the string is in the excited state $(I, J)$, then $X(0, \tau; I)$ can couple with the field $A^I$ and $X(\pi, \tau; J)$ with $A^{(J)}$. For simplicity, we will now assume that these fields are constant. Note however that $A^{(I)}$ must be intended as a function of $X(0, \tau)$ only, and similarly for $A^{(J)}$. Also to realize the variation we can vary $X^\mu(\sigma, \tau)$ by a function $\delta X^\mu(\sigma, \tau) = \Delta^\mu(\tau)$ strictly picked to $\Delta_0^\mu$ at $\tau = 0$ so that essentially

$$\partial_\tau \Delta^\mu(\tau) = \Delta_0^\mu \delta(\tau) \qquad [9]$$

where $\delta(\tau)$ is the Dirac delta function.

Using the chosen boundary conditions, the variation of the full action contains the boundary terms

$$\begin{aligned}\delta S_{\text{bound}} &= \left(A_i^{(J)} - A_i^{(I)}\right)\int_{-\infty}^{0} \partial_\tau \Delta^i(\tau)\mathrm{d}\tau \\ &\quad + \frac{1}{2\pi\alpha'}\int_0^\pi \Delta^i \partial_\sigma X_i(\sigma,0)\mathrm{d}\sigma \\ &= \frac{\Delta^i}{2\pi\alpha'}\Big[X_i(\pi,0) - X_i(0,0) \\ &\quad + 2\pi\alpha'\left(A_i^{(J)} - A_i^{(I)}\right)\Big]\end{aligned} \quad [10]$$

Imposing the condition of its vanishing gives the physical interpretation for the normal components of the U(1) fields

$$X_i(\pi,0) - X_i(0,0) = -2\pi\alpha'\left(A_i^{(J)} - A_i^{(I)}\right) \quad [11]$$

This means that, up to a constant shift, the fields $A_i^{(K)}$ measure the positions of the ends of the strings in the transverse directions! (Figure 2). Equivalently, we can say that the string ends on two different D$p$ branes, parallel but displaced in the transverse directions by a quantity $-2\pi\alpha'\left(A_i^{(J)} - A_i^{(I)}\right)$. We are thus also able to interpret the Chan–Paton factors. They mean that the string is living in a background of $n$ parallel branes, stretched between the $I$th and the $J$th brane. On every brane, a U(1) gauge group lives so that the full gauge group is $\mathrm{U}(1)^n$. However, when $k$ of the branes overlap, the corresponding set of states become indistinguishable, so that the gauge group can be enhanced to a U($k$) group. In conclusion, $n$ overlapping parallel D$p$ branes carry a $(p+1)$-dimensional U($n$) gauge theory which breaks in U($k_i$) block factors if the branes separate in stacks of $k_i$ overlapping branes.

We can say a little bit more about this. If the string excited states represent gauge degree of freedom, they must become massive to break gauge symmetry when the branes separate. To see this, let us conclude by computing the mean momentum carried by the string. After elimination of the boundary terms, the total variation of the action due to the shift $\delta X^\mu(\sigma,0) = \Delta^\mu$ becomes

$$\begin{aligned}\delta S &= \frac{1}{2\pi\alpha'}\int_\Sigma \partial_\tau \Delta^\mu \partial_\tau X_\mu \mathrm{d}\sigma^2 \\ &= \frac{\Delta^\mu}{2\pi\alpha'}\int_0^\pi \partial_\tau X_\mu(\sigma,0)\mathrm{d}\sigma\end{aligned} \quad [12]$$

The resulting momentum is

$$P_\mu = \frac{1}{2\pi\alpha'}\int_0^\pi \partial_\tau X_\mu(\sigma,0)\mathrm{d}\sigma$$

On the bulk, the fields $X^\mu$ satisfy the standard wave equation in two dimensions, so that the general solution is the sum of a left-moving and a right-moving part, $X^\mu(\sigma,\tau) = X_\mathrm{L}^\mu(\tau+\sigma) + X_\mathrm{R}^\mu(\tau-\sigma)$. Imposing the boundary conditions, one finds

$$\begin{aligned}X^a(\sigma,\tau) &= X_\mathrm{L}^a(\tau+\sigma) + X_\mathrm{L}^a(\tau-\sigma) \\ &\quad + 2\pi\alpha' p^a \tau + X_0^a\end{aligned} \quad [13]$$

$$\begin{aligned}X^i(\sigma,\tau) &= X_\mathrm{L}^i(\tau+\sigma) - X_\mathrm{L}^i(\tau-\sigma) \\ &\quad + 2\alpha'\left(A^{(J)i} - A^{(I)i}\right)\sigma + X_0^i\end{aligned} \quad [14]$$

Here $X_0^\mu$ and $p^a$ are integration constants and $X_\mathrm{L}^i(\tau+\pi) - X_\mathrm{L}^i(\tau-\pi) = 0$. A direct computation then shows that $P^a = p^a$ and $P^i = 0$, which is also what intuition suggests: the string can freely move along the branes but is fixed between them in the orthogonal directions. However, if it is stretched between two separated branes (i.e., if $I \neq J$), there is another contribution to the energy. In fact the factor $T := 1/(2\pi\alpha')$ represents the string tension, so that if $\Delta$ is its minimal length, its minimal contribution to the energy will be $\delta E = T\Delta$. This energy must equally contribute to the spectrum of the excited modes, the gauge field bosons. Here in fact, is where $T$-duality comes into play, but we will not discuss it.

The conclusion is that the spectrum corresponding to the stretched string must satisfy the condition $E \geq T\Delta$, which is as if the string states acquired a mass $T\Delta$, that is,

$$m^2 = \sum_{i=p+1}^{9}\left(A^{(J)i} - A^{(I)i}\right)^2 \quad [15]$$

This gives us a geometric tool to construct $(p+1)$-dimensional gauge theories: on $n$ coincident D$p$ branes there exists a U($n$) gauge theory which can be broken separating the branes and thus giving a mass to the gauge bosons. Such a mass is proportional to the distance between the branes (Figure 3).

Before continuing with some examples, let us make two comments. First, the theory obtained in this way is a supersymmetric one, because the

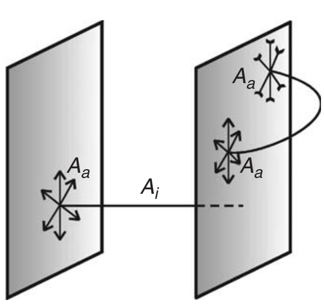



**Figure 2** Tangential components of $A_a$ appear as gauge modes. Normal components $A_i$ appear as shift modes.

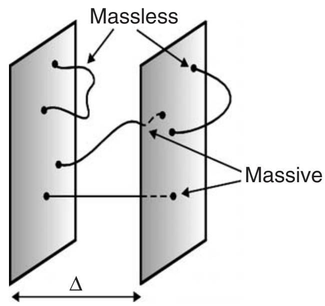



**Figure 3** Stretched strings acquire a mass.

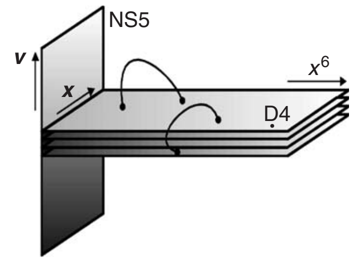



**Figure 4** D4 branes ending on an NS5 brane. Gauge degrees of freedom are frozen in four dimensions.

Dirichlet conditions allow the action of supersymmetric transformations of the form $\epsilon_L Q_L + \epsilon_R Q_R$, where $Q_L$ and $Q_R$ are the fermionic left and right supercharge operators and $\epsilon_L$, $\epsilon_R$ are spinors satisfying the brane projection condition $\epsilon_L = \pm\Gamma^0\Gamma^1 \cdot \ldots \cdot \Gamma^p \epsilon_R$. Here $\Gamma^\mu$ are the ten-dimensional Dirac matrices and one refers to "antibranes" for the negative sign.

Second, the gauge group can be converted into an SO($n$) or an Sp($n/2$) (for even $n$), adding an orientifold plane parallel to the branes. The orientifold plane acts on the orthogonal spacetime directions with a $\mathbb{Z}_2$-action

$$X^i \sim -X^i \qquad [16]$$

if $X^i = 0$ is the position of the orientifold. It further acts on the string world sheet as $\sigma \sim \pi - \sigma$ making it an unoriented string. The effect is to project out some states from the spectra, thus reducing the gauge group.

## Geometric Engineering of Gauge Theories from Branes

To illustrate how brane construction of gauge theories works, we will consider a particular configuration of branes (Witten 1997).

We would like to obtain a four-dimensional U($n$) gauge theory. A possibility could be to take $n$ D3 branes in a type IIB string background. However, such a model would contain too many supersymmetries: in ten dimensions, supersymmetries are generated by two 16-dimensional chiral spinors $\epsilon_L$, $\epsilon_R$ ($\Gamma^0 \cdot \ldots \cdot \Gamma^9 \epsilon_{L,R} = \epsilon_{L,R}$). From the four-dimensional point of view, each of them represents four four-dimensional spinors giving an $N = 8$ supersymmetric theory. The projection condition, due to the branes, reduces the number of supersymmetries to four. Supersymmetry not being manifest in nature, it is desirable to have fewer supersymmetric gauge theories at hand. Because different brane projection conditions can further reduce supersymmetry, we can try to consider the coexistence of more kinds of branes.

One way to do this is to consider $n$ parallel 4-branes ending on an NS5 brane in type IIA string theory (Figure 4), and then analyze the gauge theory restricted to the four-dimensional intersection (here the theory is nonchiral as $\Gamma^0 \cdot \ldots \cdot \Gamma^9 \epsilon_{L/R} = \pm\epsilon_{L/R}$). What kind of branes can end on other kind of branes can be established, starting from the fact that strings can end on a brane, and using the dualities tool (Giveon and Kutasov 1999).

Let us fix some conventions. We will indicate with $\boldsymbol{x} = (x^0, x^1, x^2, x^3) \in \mathbb{R}^4$ the coordinates on the intersection, so that $(\boldsymbol{x}; \boldsymbol{v}) = (\boldsymbol{x}; x^4, x^5) \in \mathbb{R}^6$ define the NS5 brane, and $(\boldsymbol{x}, x^6)$, with $x^6 \in [0, \infty)$, the 4-branes. Also $\boldsymbol{v}_I$ will indicate the position of the $I$th 4-brane on the 5-brane, and $\boldsymbol{y} = (x^7, x^8, x^9)$ will collect the remaining coordinates. Finally, we will indicate the product of Γ-matrices, corresponding to given directions, indicizing a simple Γ with the respective coordinates. For example $\Gamma^{\boldsymbol{v}} = \Gamma^4\Gamma^5$. With these conventions, the brane projection conditions for D4 and NS5 branes, respectively, read

$$\epsilon_L = \Gamma^{\boldsymbol{x}}\Gamma^6 \epsilon_R \qquad [17]$$

$$\epsilon_L = \Gamma^{\boldsymbol{x}}\Gamma^{\boldsymbol{v}} \epsilon_L, \qquad \epsilon_R = \Gamma^{\boldsymbol{x}}\Gamma^{\boldsymbol{v}} \epsilon_R \qquad [18]$$

These projections reduce supersymmetry to $N = 2$. After a short manipulation and using for example antichirality of $\epsilon_R$, it is easy to see that the first condition can be substituted by

$$\epsilon_L = \Gamma^{\boldsymbol{x}}\Gamma^{\boldsymbol{y}} \epsilon_R \qquad [19]$$

In other words, we could add a number of 6-branes in the $(\boldsymbol{x}, \boldsymbol{y})$ directions, without further reducing supersymmetry. We will consider this possibility later.

On the D4 branes there is an eventually broken U($n$) gauge theory. Here the vector fields $A_\mu$, $\mu = 0, 1, 2, 3, 6$, and the scalar fields $\boldsymbol{v}_I$ and $\boldsymbol{y}$ live. The last ones are set to zero by the Dirichlet conditions, whereas $\boldsymbol{v}_I$ measure the fluctuations of the D3 brane positions over NS5. The O(2) group

of rotations of the $(x^4, x^5)$ coordinates acts on them, which can be broken by an expectation value $\langle v_I \rangle \neq 0$. The SO(3) rotations of $(x^6, x^7, x^8)$ (under which $v_I$ are singulets) do not influence the projection conditions and can then be identified with the R-symmetry group $SU(2)_R$. It could be broken by a nonvanishing expectation value $\langle y \rangle \neq 0$, but as we said it cannot happen in the actual configuration. This highlights an unbroken supersymmetric Coulomb branch.

What is the physics as seen by an observer living on the four-dimensional spacetime $x$? The components $A_\alpha$, $\alpha = 0, 1, 2, 3$, of the vector fields transform as vectors with respect to the four-dimensional Lorentz group SO(1, 3). They satisfy Neumann boundary conditions on $x^6 = 0$ and then survive as $U(n)$ gauge vector fields. The $A_6$ component behaves as a scalar with respect to SO(1, 3) but is eliminated by a Dirichlet condition in $x^6 = 0$. The $v$ scalar field will be responsible for the eventual breaking of the gauge group.

This seems to be quite a good scenario but actually the situation is unsatisfactory. If a 4-brane extends to the interval $[0, L]$ in the $x^6$ direction, the effective action for the gauge fields goes like this:

$$\frac{1}{g_{D_4}^2}\int_0^L \mathrm{d}x^6 \int_{\mathbb{R}^4} \mathrm{d}^4 x \mathrm{tr} F_{\mu\nu}F^{\mu\nu} \approx \frac{L}{g_{D_4}^2}\int_{\mathbb{R}^4} \mathrm{d}^4 x \mathrm{tr} F_{\alpha\beta}F^{\alpha\beta} \qquad [20]$$

where $\alpha, \beta = 0, 1, 2, 3$. Thus, the gauge coupling in four dimensions appears to be $g_4 = (g_{D_4})/\sqrt{L}$. In our case, where $L$ goes to infinity, the gauge coupling vanishes and the gauge degrees of freedom are frozen. Moreover, an argument similar to the one made for the stretched strings shows that the energy of the D4 brane is very high and makes the mechanism of gauge group breaking difficult. The same is true for the NS5 brane, which also turns out to be extremely massive and does not participate in the dynamics. But this is what we want.

To solve the problem and restore gauge dynamics in four dimensions, one must consider a stack of 4-branes of finite length in the $x^6$ direction. This can be achieved placing in $x^6 = L$ a second NS5 brane parallel to the first one and in the same point in $y$ (Figure 5). In this way, the D4 branes can stretch between the NS5 branes. If $L$ is little enough, the gauge dynamics is restored also requiring a small value for $g_{D_4}$, to ensure the gravitational coupling (and the couplings with the Kaluza–Klein and NS5 modes) to be negligible. However, $L$ must be bigger then the $\delta X^6$ fluctuations in order to avoid quantum corrections.

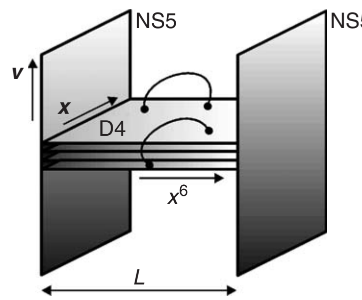



**Figure 5** $N = 2$ four-dimensional super Yang–Mills theory, with $U(n)$ gauge group.

What we just obtained is an $N = 2$ supersymmetric classical $U(n)$ gauge theory in four dimensions, without matter, and in the Coulomb branch. Before considering quantization, let us briefly discuss some possible generalizations. For example, matter can be realized attaching to the left-hand side NS5 brane, new D4 branes parallel to the previous ones, but extended in the $x^6$ direction from $-\infty$ to 0 (Figure 6). Considering strings stretched between long and short branes, we obtain states whose half-gauge action, associated with the end connected to the long brane, is frozen. The corresponding states thus appear in the fundamental representation and can be interpreted as matter states.

To consider the Higgs branch, one should be able to break supersymmetry giving an expectation value to $y$. As mentioned above, in the actual configuration this cannot happen because $y$ is set to 0 by Dirichlet conditions. Fortunately, as we said, one can add 6-branes in the $(x, y)$ directions. If we insert such branes to stop the long D4 branes in a large but finite value of $x^6$, say $x^6 = -M$ with $M \gg L$, then long branes have Neumann conditions in the $y$ directions. Thus, fluctuations of the long branes can give an expectation value to $y$, breaking supersymmetry and subsequently the Higgs branch can be tuned, shifting 4-branes stretched between 6-branes (Figure 7).

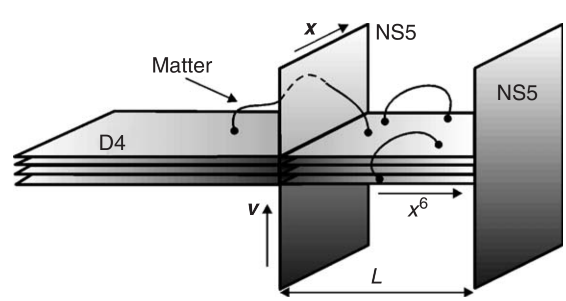



**Figure 6** Adding matter.

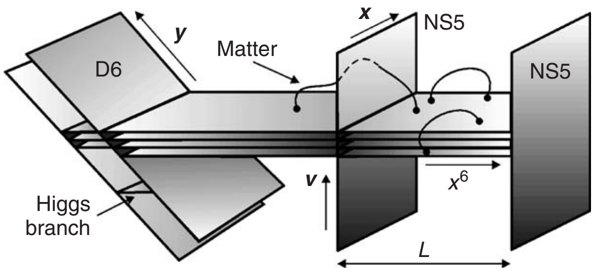



**Figure 7** Permitting Higgs phases.

The details require some careful inspection, but we shall stop our analysis here (Giveon and Kutasov 1999).

More general gauge configurations can be realized by adding more parallel NS5 branes, and thus obtaining product groups. Adding orientifold planes, one can change gauge groups as explained in the previous section (Figure 8).

Finally, we can take a further step towards more physical models, constructing $N = 1$ gauge theories. For example, this can be achieved from the previous $N = 2$ model, rotating the second NS5 brane from the $(\boldsymbol{x}, \boldsymbol{v})$ position, to the $(\boldsymbol{x}, \boldsymbol{w})$ position, where $\boldsymbol{w} = (x^8, x^9)$ (Figure 9). Then a new brane projection condition appears ($\epsilon_{\mathrm{L}} = \Gamma^x \Gamma^w \epsilon_{\mathrm{R}}$), breaking supersymmetry down to $N = 1$.

In this case, one could also obtain chiral matter, adding, for example, orientifold planes.

## Quantum Corrections from M-Theory

Up to this point we have considered classical gauge configurations. Quantum corrections could be computed switching on brane fluctuations. However, it is an amusing fact that working with $M$-theory one can obtain exact quantum results. As an example, let us sketch how the exact Seiberg–Witten solution can be obtained for the $N = 2$ model described in the previous section, in the simplest case without matter.

The full web of dualities suggests the existence of a unique unifying theory called M-theory. At low energies, M-theory appears as the strong-coupling limit of type IIA strings. In such a limit, D0 branes become the dominant objects and the corresponding states can be interpreted as Kaluza–Klein modes coming from an eleventh dimension $x^{10}$ compactified on a circle $S^1$ (Figure 10).

Thus, M-theory manifests itself as an 11-dimensional supergravity. In particular, it can be shown that there can be only a unique 11-dimensional supergravity. As said, here the nonperturbative objects are two- or five-dimensional membranes.

From the M-theory point of view, the D4 branes considered in our model appear as M5 membranes wrapped on the eleventh direction $S^1$ (Figure 11). Because quantum corrections are no longer negligible, we can no longer think of these branes as stretched in the $x^6$ direction, but $\boldsymbol{v}$ must also be considered. Thus, the M5 membranes will describe, in $\mathbb{R}^{10} \times S^1$, a region $\mathbb{R}^4 \times S$, where $\mathbb{R}^4$ are the $\boldsymbol{x}$ coordinates, and $S$ is a Riemann surface immersed in $Q \times S^1, Q$ being spanned by the $(\boldsymbol{v}, x^6)$ coordinates. In fact, supersymmetry constrains the surface to be a holomorphic curve, so that to describe it, it is convenient to collect $\boldsymbol{v} = (x^4, x^5)$ and $(x^6, x^{10})$ into complex coordinates $v = x^4 + \mathrm{i}x^5$ and $s = x^6 + \mathrm{i}x^{10}$.

To compute quantum fluctuations, let us note that the end of a D4 brane over an NS5 brane is free to move along the $v$ directions. A fully free end of a brane would satisfy a free wave equation. However, as $x^6$ is constrained in all directions but the $v$ ones, it will simply satisfy a Laplace equation in two dimensions: $\Delta_v X^6 = 0$. Let us solve it, for a fixed NS5 brane. It will be (at least for large values of $v$)

$$x^6(v) = k \sum_{i=1}^{n_{\mathrm{L}}} \log |v - v_{\mathrm{L}i}^{(\alpha)}| - k \sum_{i=1}^{n_{\mathrm{R}}} \log |v - v_{\mathrm{R}i}^{(\alpha)}| \quad [21]$$

where $n_{\mathrm{L}}$ is the number of D4 branes ending on the left-hand side of the NS5 brane, in the positions $v_{\mathrm{L}i}^{(\alpha)}$, and similar for the R index, which refers to

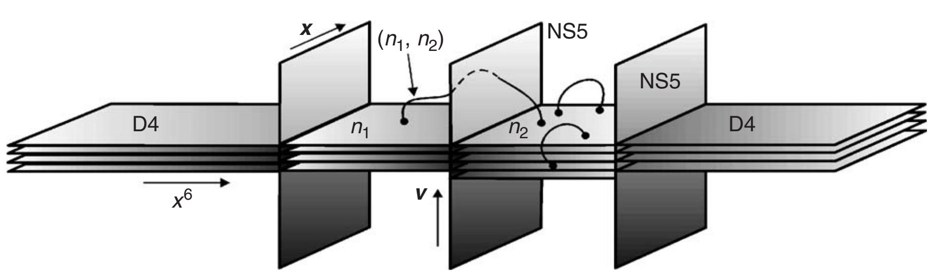



**Figure 8** $N = 2$ four-dimensional super Yang–Mills theory with $\mathrm{U}(n_1) \times \mathrm{U}(n_2)$ gauge group and matter. Strings crossing the central NS5 brane give matter in the $(n_1, n_2)$ representation.

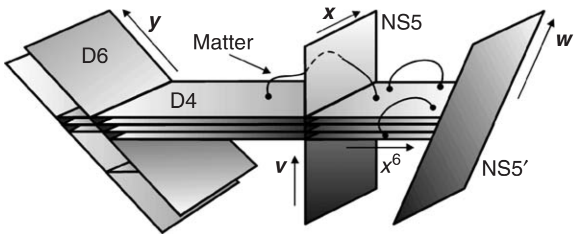



**Figure 9** Going down to $N=1$ supersymmetry.

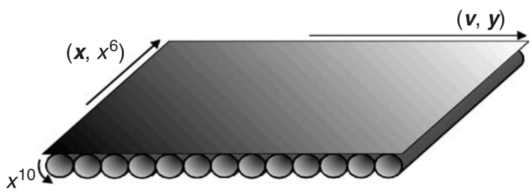



**Figure 10** In M-theory one can think as if at any ten-dimensional spacetime point, there is attached an $S^1$ circle of ray $R_{10}$.

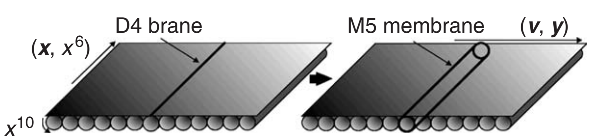



**Figure 11** D4 branes become M5 membranes in M-theory.

the right-hand side. Here $(\alpha)$ refers to the $\alpha$th NS5 brane, and $k$ is an integration constant.

Because $x^6$ is the real part of a holomorphic field, whose imaginary part is compactified on a circle of ray $R_{10}$, we then find

$$s(v) = R_{10} \sum_{i=1}^{n_{\mathrm{L}}} \log\left(v - v_{\mathrm{L}i}^{(\alpha)}\right) - R_{10} \sum_{i=1}^{n_{\mathrm{R}}} \log\left(v - v_{\mathrm{R}i}^{(\alpha)}\right) \qquad [22]$$

This describes the quantum fluctuations of the NS5 brane as seen in M-theory. In particular, because of the imaginary part of $s$, the ends of the D4 branes appear as vortices on the NS5 brane. In place of $s$, it is now convenient to introduce a new field $t := \exp(-s/R_{10})$ so that

$$t(v) = \frac{\prod_{i=1}^{n_{\mathrm{R}}} \left(v - v_{\mathrm{R}i}^{(\alpha)}\right)}{\prod_{i=1}^{n_{\mathrm{L}}} \left(v - v_{\mathrm{L}i}^{(\alpha)}\right)} \qquad [23]$$

Before continuing, let us look a bit again at the classical limit. In this case, a fixed value of $v$ will correspond to the position of a D4 brane, whereas a fixed value of $s$ will correspond to the fixed position of an NS5 brane. The classical configuration is then

$$\left(s - s^{(1)}\right)\left(s - s^{(2)}\right) \prod_{i=1}^{n} (v - v_i) = 0 \qquad [24]$$

Here $s^{(\alpha)}$ are the positions of the NS5 branes, and the positions $v_i$ of the D4 branes coincide for both the NS5 branes. Also, for large values of $v$, one has $t^{(1)} \approx v^n$ and $t^{(2)} \approx v^{-n}$.

Quantum mechanically, the configuration is determined in terms of $v$ and $t$ by the holomorphic curve $S$, which can be described as an algebraic curve $F(v,t) = 0$, generalizing the classical configuration. As there are two NS5 branes and $n$ D4 branes, $F$ must be a polynomial of degree 2 in $t$,

$$F(v,t) = A_2(v)t^2 + A_1(v)t + A_0(v) \qquad [25]$$

where $A_a$, $a = 1,2,3$, are all polynomials of degree $n$. Note that values of $v$ such that $A_1$ vanishes give the solution $t = 0$, which corresponds to sending the right-hand side NS5 brane to $\infty$. Similarly, $A_2 = 0$ sends the other NS5 brane to $-\infty$. To avoid these undesirable configurations, we can set $A_0 = A_2 = 1$. For $A_1$, we can take the most general choice, up to an eventual shift in $v$, giving the quantum configuration

$$t^2 + \left[v^n + a_{n-2}v^{n-2} + \cdots + a_1 v + a_0\right]t + 1 = 0 \qquad [26]$$

This realizes a quantum-mechanical correspondence between the M5 membrane configurations described by the given polynomials, and the $N=2$ super Yang–Mills vacua. But this is also the claimed Seiberg–Witten curve. In particular, M-theory gives a concrete physical meaning for the support Riemann surfaces of the Seiberg–Witten solutions.

To conclude, let us make some further comments. It is clear how the construction can be extended for involving more configurations, for example, with more NS5 branes, or adding matter.

Also, we have seen that the geometrical picture which branes give of gauge theories extends at the quantum level.

A similar construction can be made for the $N=1$ model, which also permits a full geometrical proof of the Seiberg duality at both classical and quantum levels.

Finally, we should note that there are also other methods, which work in spacetimes where extra dimensions are compactified. There, the branes wrap around certain singular loci which contain information about gauge symmetries (Lerche 1997).

*See also:* AdS/CFT Correspondence; Compactification of Superstring Theory; Gauge Theories from Strings; Noncommutative Geometry from Strings; Seiberg–Witten Theory; Supergravity; Superstring Theories; Supersymmetric Particle Models.

## Further Reading


Giveon A and Kutasov D (1999) Brane dynamics and gauge theory. *Reviews of Modern Physics* 71: 983.

Johnson CV (2003) *D-Branes*.Cambridge: Cambridge University Press.

Lerche W (1997) Introduction to Seiberg–Witten Theory and Its Stringy Origin. *Nucl. Phys. Proc. Suppl. B* 55: 83.

Polchinski J (1998) *String Theory. Vol. 1: An Introduction to the Bosonic String*. Cambridge: Cambridge University Press.

Polchinski J (2004) *String Theory. Vol. 2: Superstring Theory and Beyond*. Cambridge: Cambridge University Press.

Witten E (1997) Solutions of four-dimentional field theories via M-theory. *Nuclear Physics B* 500: 3.

Zwiebach B (2004) *A First Course in String Theory*. Cambridge: Cambridge University Press.


# Brane Worlds

**R Maartens**, Portsmouth University, Portsmouth, UK



## Introduction

At high enough energies, Einstein's classical theory of general relativity breaks down, and will be superseded by a quantum gravity theory. The singularities predicted by general relativity in gravitational collapse and in the hot big bang origin of the universe are thought to be artifacts of the classical nature of Einstein's theory, which will be removed by a quantum theory of gravity. Developing a quantum theory of gravity and a unified theory of all the forces and particles of nature are the two main goals of current work in fundamental physics. The problem is that general relativity and quantum field theory cannot simply be molded together. There is as yet no generally accepted (pre-)quantum gravity theory.

The quest for a quantum gravity theory has a long and thus far not very successful history. Many different lines of attack have been developed, each having a different way of dealing with the classical singularities that arise from point particles and smooth spacetime geometry. String theory does away with zero-dimensional point particles, and particles are modeled as different states of new fundamental objects, the one-dimensional strings. It turns out, however, that there is a price to pay – the number of spacetime dimensions must be greater than four for a consistent theory. When fermions are included, which leads to superstring theory, the required number of dimensions is ten – one time and nine space dimensions.

There are in fact five distinct (1+9)-dimensional superstring theories. In the mid-1990s, duality transformations were discovered that relate these superstring theories to each other and to the (1+10)-dimensional supergravity theory. This led to the conjecture that all of these theories arise as different limits of a single theory, which has come to be known as M theory. It was also discovered that extended objects of higher dimension than strings play a fundamental role in the theory. These objects are known as "branes" (from membranes), and the relation between them and strings leads to a new picture of how gravity and matter may be connected in the universe. Roughly speaking, open strings describe the particles of the nongravitational sector, and their ends are attached to branes, while closed strings, which describe the graviton and associated particles of the gravitational sector, can move freely in all dimensions.

Thus, the observable universe could be a $(1+3)$-surface – a "brane," embedded in a $(1+3+d)$-dimensional spacetime – the "bulk," with standard-model particles and fields trapped on the brane, while gravity is free to access the bulk. Brane-world models offer a phenomenological way to test some of the novel predictions and corrections to general relativity that are implied by M theory.

## Higher-Dimensional Gravity

Brane worlds can be seen as reviving the original higher-dimensional ideas of Kaluza and Klein in the 1920s, but in a new context of quantum gravity. An important consequence of extra dimensions is that the four-dimensional Planck scale $M_{\rm p} \equiv M_{(4)} = 1.2 \times 10^{19}$ GeV is no longer the fundamental energy scale of gravity. The fundamental scale is instead $M_{(4+d)}$. This can be seen from the modification of the gravitational potential. For an Einstein–Hilbert gravitational action,

$$S_{\rm gravity} = \frac{1}{2\kappa^2_{(4+d)}} \int {\rm d}^4x\, {\rm d}^dy \sqrt{-{}^{(4+d)}g} \times \left[ {}^{(4+d)}R - 2\Lambda_{(4+d)} \right] \quad [1]$$

we have the higher-dimensional Einstein field equations,

$$\begin{aligned} {}^{(4+d)}G_{AB} &\equiv {}^{(4+d)}R_{AB} - \tfrac{1}{2}{}^{(4+d)}R\,{}^{(4+d)}g_{AB} \\ &= -\Lambda_{(4+d)}{}^{(4+d)}g_{AB} + \kappa^2_{(4+d)}{}^{(4+d)}T_{AB} \quad [2] \end{aligned}$$

where $x^A = (x^a, y^1, \ldots, y^d)$ and $\kappa^2_{(4+d)}$ is the gravitational coupling constant given by

$$\kappa^2_{(4+d)} = 8\pi G_{(4+d)} = \frac{8\pi}{M^{2+d}_{(4+d)}} \qquad [3]$$

The static weak field limit of the field equations leads to the $(4+d)$-dimensional Poisson equation, whose solution is the gravitational potential

$$V(r) \propto \frac{\kappa^2_{(4+d)}}{r^{1+d}} \qquad [4]$$

In the simplest scenario, we can assume a toroidal configuration for the $d$ extra dimensions, with each compactified on the same length scale $L$. Then on scales $r \lesssim L$, the potential is $(4+d)$-dimensional, $V \sim r^{-(1+d)}$. By contrast, on scales large relative to $L$, where the extra dimensions do not contribute to variations in the potential, $V$ behaves like a four-dimensional potential, $V \sim L^{-d} r^{-1}$. This means that the usual Planck scale becomes an effective coupling constant, describing gravity on scales much larger than the extra dimensions, and related to the fundamental scale via the volume of the extra dimensions:

$$M_{\rm p}^2 \sim M^{2+d}_{(4+d)} L^d \qquad [5]$$

## Large Extra Dimensions

If the extra-dimensional volume is significantly above the Planck scale, then the true fundamental scale $M_{(4+d)}$ can be much less than the effective scale $M_{\rm p}$,

$$L^d \gg M_{\rm p}^{-d} \;\Rightarrow\; M_{(4+d)} \ll M_{\rm p} \qquad [6]$$

In this case, we understand the weakness of gravity as due to the fact that it "spreads" into extra dimensions, and only a part of it is felt in four dimensions.

A lower limit on $M_{(4+d)}$ is given by null results in table-top experiments to test for deviations from Newton's law in four dimensions, $V \propto r^{-1}$. These experiments currently probe submillimeter scales, and find no detectable deviation, so that

$$L \lesssim 10^{-1}\,\text{mm} \sim (10^{-15}\,\text{TeV})^{-1} \\ \Rightarrow\; M_{(4+d)} \gtrsim 10^{(32-15d)/(d+2)}\,\text{TeV} \qquad [7]$$

Stronger bounds can be derived from null results in particle accelerators in some brane-world models, or from constraints imposed by observations of supernovae or of light-element abundance.

Brane worlds, arising in the framework of string theory, thus incorporate the possibility that the fundamental scale is much less than the Planck scale felt in four dimensions. This emerges by virtue of the large size of the extra dimensions. It is not necessary for all extra dimensions to be of equal size for this mechanism to operate. There are string theory solutions (Horava–Witten solutions) with two $(1+9)$-branes located at the boundaries of the bulk, at the endpoints of an $S^1/Z_2$ orbifold, that is, a circle folded on itself across a diameter. The orbifold extra dimension is the large one, whereas the other six extra dimensions on the branes are compactified on a very small scale, close to the fundamental scale, and their effect on the dynamics is felt through "moduli" fields, that is, five-dimensional scalar fields.

These solutions can be thought of as effectively five dimensional, with an extra dimension that can be large relative to the fundamental scale. They provide the basis for the Randall–Sundrum 1 (RS1) phenomenological models of five-dimensional gravity. The single-brane Randall–Sundrum 2 (RS2) models with infinite extra dimension arise when the orbifold radius tends to infinity. The RS models are not the only phenomenological realizations of M theory ideas. They were preceded by the brane-world models of Arkani-Hamed, Dimopoulos, and Dvali (ADD), which put forward the idea that a large volume for the compact extra dimensions would lower the effective Planck scale $M_{(4+d)}$. If $M_{(4+d)}$ is close to the electroweak scale, $M_{\rm ew}$, then this would address the long-standing "hierarchy" problem, that is, why there is such a large gap between $M_{\rm ew} \sim 1\,\text{TeV}$ and $M_{\rm p} \sim 10^{16}\,\text{TeV}$.

In the ADD models, more than one extra dimension is required for agreement with experiments, and there is "democracy" among the equivalent extra dimensions, which, in addition, are flat. By contrast, the RS models have a "preferred" extra dimension, with other extra dimensions treated as ignorable (i.e., stabilized except at energies near the fundamental scale). Furthermore, this extra dimension is curved or "warped" rather than flat: the bulk is a portion of anti-de Sitter ($\text{AdS}_5$) spacetime. The RS branes are $Z_2$-symmetric (mirror symmetry), and have a tension, which serves to counter the influence on the brane of the negative bulk cosmological constant. This also means that the self-gravity of the branes is incorporated in the RS models. The novel feature of the RS models compared to previous higher-dimensional models is that the observable three dimensions are protected from the large extra dimension (at low energies) by curvature (warping), rather than straightforward compactification.

The RS brane worlds provide phenomenological models that reflect at least some of the features of

M theory, and that bring exciting new geometric and particle physics ideas into play. The RS2 models also provide a framework for exploring holographic ideas that have emerged in M theory. Roughly speaking, holography suggests that higher-dimensional dynamics may be determined from a knowledge of the fields on a lower-dimensional boundary. The AdS/CFT correspondence is an example in which the classical dynamics of the higher-dimensional AdS gravitational field are equivalent to the quantum dynamics of a conformal field theory (CFT) on the boundary.

## Kaluza–Klein Modes

The dilution of gravity via extra dimensions not only weakens gravity, it also broadens the range of graviton modes felt on the brane. The graviton is more than just the four-dimensional massless mode of four-dimensional gravity – other modes, with an effective mass on the brane, arise from the fact that the graviton is a (4+d)-dimensional massless particle. These extra modes on the brane are known as Kaluza–Klein (KK) modes of the graviton.

For simplicity, consider a flat brane with one flat extra dimension, compactified through the identification $y \leftrightarrow y + 2\pi nL$, where $n = 0, 1, 2, \ldots$ . The perturbative five-dimensional graviton is defined via

$$^{(5)}\eta_{AB} \rightarrow {}^{(5)}\eta_{AB} + h_{AB} \qquad [8]$$

where $^{(5)}\eta_{AB}$ is the five-dimensional Minkowski metric and $h_{AB}$ is a small transverse traceless perturbation. Its amplitude can be Fourier expanded as

$$h(x^a, y) = \sum_n \mathrm{e}^{iny/L}\, h_n(x^a) \qquad [9]$$

where $h_n$ are the amplitudes of the KK modes, that is, the effective four-dimensional modes of the five-dimensional graviton. To see that these KK modes are massive from the brane viewpoint, we start from the five-dimensional wave equation that the massless five-dimensional field $h$ satisfies (in a suitable gauge):

$$^{(5)}\Box h = 0 \;\Rightarrow\; \Box h + \partial_y^2 h = 0 \qquad [10]$$

It follows that the KK modes satisfy a four-dimensional Klein–Gordon equation with an effective four-dimensional mass, $m_n$:

$$\Box h_n = m_n^2 h_n, \quad m_n = \frac{n}{L} \qquad [11]$$

The massless mode, $h_0$, is the usual four-dimensional graviton mode. But there is a tower of massive modes, $L^{-1}, 2L^{-1}, \ldots,$ which imprint the effect of the five-dimensional gravitational field on the four-dimensional brane. Compactness of the extra dimension leads to discreteness of the spectrum. For an infinite extra dimension, $L \to \infty$, the separation between the modes disappears and the tower forms a continuous spectrum.

## Randall–Sundrum Brane Worlds

RS brane worlds do not rely on compactification to localize gravity at the brane, but on the curvature of the bulk. What prevents gravity from "leaking" into the extra dimension at low energies is a negative bulk cosmological constant,

$$\Lambda_{(5)} = -\frac{6}{\ell^2} = -6\mu^2 \qquad [12]$$

where $\ell$ is the curvature radius of $AdS_5$ and $\mu$ is the corresponding energy scale. The bulk cosmological constant with its repulsive gravity effect acts to "squeeze" the gravitational potential closer to the brane. We can see this clearly in Gaussian normal coordinates $x^A = (x^\mu, y)$ based on the brane at $y = 0$, for which the metric takes the form

$$^{(5)}\mathrm{d}s^2 = \mathrm{d}y^2 + \mathrm{e}^{-2|y|/\ell}\eta_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu \qquad [13]$$

with $\eta_{\mu\nu}$ the Minkowski metric. The exponential warp factor reflects the confining role of the bulk cosmological constant. The $Z_2$-symmetry about the brane at $y = 0$ is incorporated via the $|y|$ term. In the bulk, this metric is a solution of the five-dimensional Einstein equations,

$$^{(5)}G_{AB} = -\Lambda_{(5)}{}^{(5)}g_{AB} \qquad [14]$$

that is, $^{(5)}T_{AB} = 0$ in eqn [2]. The brane is a flat Minkowski spacetime, $g_{AB}(x^\mu, 0) = \eta_{\mu\nu}\delta^\mu{}_A\delta^\nu{}_B$, with self-gravity in the form of brane tension.

The two RS models are distinguished as follows:

**RS1** There are two branes in RS1, at $y = 0$ and $y = L$, with $Z_2$-symmetry identifications

$$y \leftrightarrow -y, \qquad y + L \leftrightarrow L - y \qquad [15]$$

The branes have equal and opposite tensions, $\pm\lambda$, where

$$\lambda = \frac{3}{4\pi}\,\frac{M_{\mathrm{p}}^2}{\ell^2} \qquad [16]$$

The positive-tension "TeV" brane has fundamental scale $M_{(5)} \sim 1$ TeV. Because of the exponential

warping factor, the effective scale on the negative tension "Planck" brane at $y = L$ is $M_p$. On the positive tension brane,

$$M_p^2 = M_{(5)}^3 \ell \left[1 - e^{-2L/\ell}\right] \quad [17]$$

So RS1 gives a new approach to the hierarchy problem. Because of the finite separation between the branes, the KK spectrum is discrete.

**RS2** In RS2, there is only one, positive-tension, brane. This may be thought of as arising from sending the negative tension brane off to infinity, $L \to \infty$. Then the energy scales are related via

$$M_{(5)}^3 = \frac{M_p^2}{\ell} \quad [18]$$

On the RS2 brane, the negative $\Lambda_{(5)}$ is offset by the positive brane tension $\lambda$. The fine-tuning in eqn [16] ensures that there is zero effective cosmological constant on the brane, so that the brane has the induced geometry of Minkowski spacetime. To see how gravity is localized at low energies, we consider the five-dimensional graviton perturbations of the metric:

$$\begin{gathered} {}^{(5)}g_{AB} \to {}^{(5)}g_{AB} + h_{AB} \\ h_{Ay} = 0 = h^\mu{}_\mu = \partial_\nu h^{\mu\nu} \end{gathered} \quad [19]$$

We split the amplitude $h$ into three-dimensional Fourier modes, and the linearized five-dimensional Einstein equations lead to the wave equation ($y > 0$)

$$e^{2y/\ell}\left[\ddot{h} + k^2 h\right] = h'' - \frac{4}{\ell} h' \quad [20]$$

Separability means we can write

$$h(t, y) = \sum_m \varphi_m(t)\, h_m(y) \quad [21]$$

and the wave equation reduces to

$$\ddot{\varphi}_m + (m^2 + k^2)\varphi_m = 0 \quad [22]$$

$$h''_m - \frac{4}{\ell} h'_m + e^{2y/\ell} h_m = 0 \quad [23]$$

The zero-mode solution is

$$\varphi_0(t) = A_{0+}\, e^{+ikt} + A_{0-}\, e^{-ikt} \quad [24]$$

$$h_0(y) = B_0 + C_0 e^{4y/\ell} \quad [25]$$

and the massive KK mode ($m > 0$) solutions are

$$\begin{aligned} \varphi_m(t) = {} & A_{m+} \exp\left(+i\sqrt{m^2 + k^2}\, t\right) \\ & + A_{m-} \exp\left(-i\sqrt{m^2 + k^2} t\right) \end{aligned} \quad [26]$$

$$\begin{aligned} h_m(y) = {} & e^{2y/\ell}\Big[B_m J_2\left(ml e^{y/\ell}\right) \\ & + C_m Y_2\left(ml e^{y/\ell}\right)\Big] \end{aligned} \quad [27]$$

where $J_2, Y_2$ are Bessel functions.

The boundary condition for the perturbations is $h'(t, 0) = 0$, which implies

$$C_0 = 0, \qquad C_m = -\frac{J_1(m\ell)}{Y_1(m\ell)} B_m \quad [28]$$

In the RS1 model, we have a further boundary condition, $h'(t, L) = 0$, which leads to a discrete eigenspectrum, namely the masses $m$ that satisfy

$$J_1\left(m\ell e^{L/\ell}\right) Y_1(m\ell) - Y_1\left(m\ell e^{L/\ell}\right) J_1(m\ell) = 0 \quad [29]$$

The zero mode is normalizable, since

$$\left| \int_0^\infty B_0 e^{-2y/\ell} dy \right| < \infty \quad [30]$$

Its contribution to the gravitational potential $V = (1/2)h_{00}$ gives the four-dimensional result, $V \propto r^{-1}$. The contribution of the massive KK modes sums to a correction of the four-dimensional potential. For $r \ll \ell$, one obtains

$$V(r) \approx \frac{GM}{r}\left(1 + \frac{\ell}{r}\right) \approx \frac{GM\ell}{r^2} \quad [31]$$

which simply reflects the fact that the potential becomes truly five dimensional on small scales. For $r \gg \ell$,

$$V(r) \approx \frac{GM}{r}\left(1 + \frac{2\ell^2}{3r^2}\right) \quad [32]$$

which gives the small correction to four-dimensional gravity at low energies from extra-dimensional effects.

## Cosmological Brane Worlds

The RS models contain vacuum (Minkowski) branes. In order to pursue brane-world ideas in cosmology, we need to generalize the RS models to incorporate cosmological branes with matter and radiation on them. The effective field equations on the brane are the vehicle for brane-bound observers to interpret cosmological dynamics. They arise from projecting the five-dimensional field equations onto the brane, via the Gauss–Codazzi equations. These equations involve also the extrinsic curvature $K_{\mu\nu}$ of the brane, which determines how the brane is imbedded in the bulk.

The stress-energy on the brane (tension, matter, radiation) means that there is a jump in $K_{\mu\nu}$ across

the brane. More precisely, the junction conditions across the brane are

$$g^{+}_{\mu\nu} - g^{-}_{\mu\nu} = 0 \quad [33]$$

$$K^{+}_{\mu\nu} - K^{-}_{\mu\nu} = -\kappa^2_{(5)}\left[T^{\text{brane}}_{\mu\nu} - \tfrac{1}{3}T^{\text{brane}}\, g_{\mu\nu}\right] \quad [34]$$

where

$$T^{\text{brane}}_{\mu\nu} = T_{\mu\nu} - \lambda g_{\mu\nu} \quad [35]$$

is the total energy–momentum tensor on the brane and $T^{\text{brane}} = g^{\mu\nu}T^{\text{brane}}_{\mu\nu}$. The $Z_2$-symmetry means that when approaching the brane from one side and going through it, one emerges into a bulk that looks the same, but with the normal reversed. This implies that

$$K^{-}_{\mu\nu} = -K^{+}_{\mu\nu} \quad [36]$$

so that we can use the junction condition (eqn [34]) to determine the extrinsic curvature:

$$K_{\mu\nu} = -\tfrac{1}{2}\kappa^2_{(5)}\left[T_{\mu\nu} + \tfrac{1}{3}(\lambda - T)g_{\mu\nu}\right] \quad [37]$$

where $T = T^{\mu}{}_{\mu}$, we have dropped the (+) and we evaluate quantities on the brane by taking the limit $y \to +0$.

Together with the Gauss–Codazzi equations, eqn [37] leads to the induced field equations on the brane:

$$G_{\mu\nu} = -\Lambda g_{\mu\nu} + \kappa^2 T_{\mu\nu} + 6\frac{\kappa^2}{z\lambda}\mathcal{S}_{\mu\nu} - \mathcal{E}_{\mu\nu} \quad [38]$$

where

$$\kappa^2 \equiv \kappa^2_{(4)} = \tfrac{1}{6}\lambda\kappa^4_{(5)} \quad [39]$$

$$\Lambda \equiv \Lambda_{(4)} = \tfrac{1}{2}\left[\Lambda_{(5)} + \kappa^2\lambda\right] \quad [40]$$

$$\mathcal{S}_{\mu\nu} = \tfrac{1}{12}TT_{\mu\nu} - \tfrac{1}{4}T_{\mu\alpha}T^{\alpha}{}_{\nu} + \tfrac{1}{24}g_{\mu\nu}\left[3T_{\alpha\beta}T^{\alpha\beta} - T^2\right] \quad [41]$$

and

$$\mathcal{E}_{\mu\nu} = {}^{(5)}C_{ACBD}n^C n^D g_{\mu}{}^{A} g_{\nu}{}^{B} \quad [42]$$

where $n^A$ is the unit normal to the brane and ${}^{(5)}C_{ACBD}$ is the Weyl tensor in the bulk.

The induced field equations [38] show two key modifications to the standard four-dimensional Einstein field equations arising from extra-dimensional effects.

- $\mathcal{S}_{\mu\nu} \sim (T_{\mu\nu})^2$ is the high-energy correction term, which is negligible for $\rho \ll \lambda$, but dominant for $\rho \gg \lambda$ (where $\rho$ is the energy density):

$$\frac{|\kappa^2\mathcal{S}_{\mu\nu}/\lambda|}{|\kappa^2 T_{\mu\nu}|} \sim \frac{|T_{\mu\nu}|}{\lambda} \sim \frac{\rho}{\lambda} \quad [43]$$

- $\mathcal{E}_{\mu\nu}$, the projection of the bulk Weyl tensor on the brane, encodes corrections from KK or five-dimensional graviton effects. From the brane-observer viewpoint, the energy–momentum corrections in $\mathcal{S}_{\mu\nu}$ are local, whereas the KK corrections in $\mathcal{E}_{\mu\nu}$ are nonlocal, since they incorporate five-dimensional gravity wave modes. These nonlocal corrections cannot be determined purely from data on the brane. In the perturbative analysis of RS2 which leads to the corrections in the gravitational potential, eqn [32], the KK modes that generate this correction are responsible for a nonzero $\mathcal{E}_{\mu\nu}$; this term is what carries the modification to the weak-field field equations.

The effective field equations are not a closed system. One needs to supplement them by five-dimensional equations governing $\mathcal{E}_{\mu\nu}$, which are obtained from the five-dimensional Einstein equations.

## Cosmological Dynamics

A (1+4)-dimensional spacetime with spatial 4-isotropy (four-dimensional spherical/ plane/ hyperbolic symmetry) has a natural splitting into hypersurfaces of symmetry, which are (1+3)-dimensional surfaces with 3-isotropy and 3-homogeneity, that is, Friedmann–Robertson–Walker (FRW) surfaces. In particular, the $AdS_5$ bulk of the RS2 brane world, which admits a foliation into Minkowski surfaces, also admits an FRW foliation since it is 4-isotropic. The generalization of $AdS_5$ that preserves 4-isotropy and solves the five-dimensional Einstein equation is Schwarzschild $AdS_5$, and this bulk therefore admits an FRW foliation. It follows that an FRW cosmological brane world can be embedded in Schwarzschild $AdS_5$ spacetime.

The black hole in the bulk is felt on the brane via the $\mathcal{E}_{\mu\nu}$ term. The bulk black hole gives rise to "dark radiation" on the brane via its Coulomb effect. The FRW brane can be thought of as moving radially along the fifth dimension, with the junction conditions determining the velocity via the Friedmann equation. Thus, one can interpret the expansion of the universe as motion of the brane through the static bulk. In the special case of no black hole and no brane motion, the brane is empty and has Minkowski geometry, that is, the original RS2 brane world is recovered, in different coordinates.

An intriguing aspect of the cosmological metric is that five-dimensional gravitational wave signals can take "shortcuts" through the bulk in traveling

between points A and B on the brane. The travel time for such a graviton signal is less than the time taken for a photon signal (which is stuck to the brane) from A to B.

Cosmological dynamics on the brane are governed by the modified Friedmann equation:

$$H^2 = \frac{\kappa^2}{3}\rho\left(1+\frac{\rho}{2\lambda}\right)+\frac{m}{a^4}+\frac{1}{3}\Lambda-\frac{K}{a^2} \qquad [44]$$

where $H = \dot{a}/a$ is the Hubble expansion rate, $a(t)$ is the scale factor, $K$ is the curvature index, and $m$ is the mass of the bulk black hole.

The $\rho^2/\lambda$ term is the high-energy term. When $\rho \gg \lambda$, in the early universe, then $H^2 \propto \rho^2$. This means that a given energy density produces a greater rate of expansion that it would in standard four-dimensional gravity. As a consequence, inflation in the early universe is modified in interesting ways, some of which may leave a signature in cosmological observations.

The $m/a^4$ term in eqn [44] is the "dark radiation," so called because it redshifts with expansion like ordinary radiation. But, unlike ordinary radiation, it is not a form of detectable matter, but the imprint on the brane of the gravitational field in the bulk (the Coulomb effect of the bulk black hole). This additional effective relativistic degree of freedom is constrained by nucleosynthesis in the early universe. Any extra radiative energy not thermally coupled to radiation affects the rate of production of light elements, and observed abundances place tight constraints on such extra energy. The dark radiation can be no more than ~3% of the radiation energy density at nucleosynthesis:

$$\frac{3m}{\kappa^2\rho_{\rm nuc}} \lesssim 0.03 \qquad [45]$$

The other modification to the Hubble rate is via the high-energy correction $\rho/\lambda$. In order to recover the observational successes of general relativity, the high-energy regime where significant deviations occur must take place before nucleosynthesis, that is, cosmological observations impose the lower limit

$$\lambda > (1\,\mathrm{MeV})^4 \;\Rightarrow\; M_{(5)} > 10^4\,\mathrm{GeV} \qquad [46]$$

This is much weaker than the limit imposed by table-top experiments, which limit the curvature radius to $\ell \lesssim 0.2\,\mathrm{mm}$, leading to

$$\lambda > (100\,\mathrm{GeV})^4 \Rightarrow M_{(5)} > 10^8\,\mathrm{GeV} \qquad [47]$$

The high-energy regime during radiation domination is short-lived. Since $\rho^2/\lambda$ decays as $a^{-8}$ during the radiation era, it will rapidly drop below one, and the universe will enter the low-energy four-dimensional regime. However, traces of the high-energy era may be left in the perturbation spectra that leave an imprint in the cosmic microwave background radiation.

In conclusion, simple brane-world models of RS2 type provide a rich phenomenology for exploring some of the ideas that are emerging from M theory. The higher-dimensional degrees of freedom for the gravitational field, and the confinement of standard model fields to the visible brane, lead to a complex but fascinating interplay between gravity, particle physics, and geometry, which enlarges and enriches general relativity in the direction of a quantum gravity theory. High-precision astronomical data mean that cosmology is a potential laboratory for testing and constraining these brane worlds. The models predict extra-dimensional signatures in the cosmic microwave background and other observations, and these predictions can in principle be tested against data.

*See also:* String Theory: Phenomenology; Supergravity; Superstring Theories.

## Further Reading


Brax P and van de Bruck C (2003) Cosmology and brane worlds: a review. *Classical and Quantum Gravity* 20: R201 (arXiv: hep-th/0303095) (arXiv: hep-th/0303095).

Cavaglia M (2003) Black hole and brane production in TeV gravity: a review. *International Journal of Modern Physics* A18: 1843 (arXiv:hep-ph/0210296).

Langlois D (2003) Cosmology in a brane-universe. *Astrophysics and Space Science* 283: 469 (arXiv:astro-ph/0301022).

Maartens R (2004) Brane-world gravity. *Living Reviews in Relativity* 7: 7 (arXiv:gr-qc/0312059).

Quevedo F (2002) Lectures on string/brane cosmology. *Classical and Quantum Gravity* 19: 5721 (arXiv:hep-th/0210292).

Rubakov V (2001) Large and infinite extra dimensions. *Physics-Uspekhi* 44: 871 (arXiv:hep-ph/0104152).

Wands D (2002) String-inspired cosmology. *Classical and Quantum Gravity* 19: 3403 (arXiv:hep-th/0203107).

# Branes and Black Hole Statistical Mechanics

**S R Das**, University of Kentucky, Lexington, KY, USA



## Introduction

In classical general relativity, a black hole is a solution of Einstein's equations with a region of spacetime which is causally disconnected from the asymptotic region at infinity. The boundary of such a region is called the "event horizon." The spacetime around the simplest black hole in three space dimensions is described by the Schwarzschild metric

$$ds^2 = -\left(1 - \frac{2GM}{rc^2}\right) dt^2 + \left(1 - \frac{2GM}{rc^2}\right)^{-1} dr^2 + r^2\, d\Omega^2 \qquad [1]$$

where $G$ is Newton's gravitational constant, $c$ is the velocity of light, and we have used spherical coordinates with $d\Omega$ the line element on an $S^2$. A nonrotating, uncharged star which is too massive to form a neutron star will eventually collapse, and at late times the metric will be given by [1]. The horizon is a null surface $S^2 \times t$ and the radius of the $S^2$ is $r_{\text{horizon}} = 2GM/c^2$. The Schwarzschild solution has generalizations to black holes with charge and angular momentum and no-hair theorems guarantee that a black hole has no other characteristic property. All these solutions can be generalized to other theories like supergravity in various dimensions.

In 1974, Hawking showed that due to pair production of particles near the horizon, black holes radiate thermally. Hawking's calculation is valid for black holes whose masses are much larger than the Planck mass: for such black holes, the curvature at the horizon is weak and normal semiclassical quantization is valid. Remarkably, the properties of Hawking radiation are quite universal. A black hole can be characterized by an entropy called the Bekenstein–Hawking entropy. The leading result for the entropy $S_{\text{BH}}$ for all black holes in any theory with the standard Einstein–Hilbert action is given by

$$S_{\text{BH}} = \frac{A_{\text{H}}}{4G} \qquad [2]$$

where $A_{\text{H}}$ denotes the area of the horizon. The temperature $T_{\text{H}}$ is given by

$$T_{\text{H}} = \frac{\kappa}{2\pi} \qquad [3]$$

where $\kappa$ is the surface gravity at the horizon. The principle of detailed balance further ensures that the radiation rate of some species of particle $i$, $\Gamma_i(k)$, in some given momentum range $(\boldsymbol{k}, \boldsymbol{k} + d\boldsymbol{k})$ is related to the corresponding absorption cross section $\sigma_i(k)$ by

$$\Gamma(k) = \frac{\sigma_i(k)}{e^{\omega/T_{\text{H}}} \pm 1} \frac{d^d k}{(2\pi)^d} \qquad [4]$$

where $\omega$ is the energy and $d$ denotes the number of spatial dimensions. The $\pm$ sign refers to fermions (bosons), respectively. A nontrivial $k$ dependence of $\sigma_i$ signifies a departure from black-body behavior. Consequently, $\sigma_i(k)$ is often called a grey-body factor. Equations [2] and [3] may be derived by combining Hawking's calculation of the radiation with standard thermodynamic relations. Alternatively, they follow from the leading semiclassical approximations of path-integral formulations of Euclidean gravity based on the standard Einstein–Hilbert action. For an account of black-hole thermodynamics, see Wald (1994).

Unlike usual thermodynamic systems, black holes appear to pose a deep puzzle. In usual systems, thermodynamics is a coarse-grained description of a system which is in a highly degenerate state. Typically, such systems are described in terms of a few macroscopic parameters such as the total energy, the total volume, the total charge. For each set of values of these macroscopic parameters, there are a large number of microscopic states which can be described in terms of the constituents such as atoms or molecules. This degeneracy manifests itself as an entropy $S$ which is related to the number of microscopic states for a given set of values of the macroscopic parameters, $\Omega$ by Boltzmann's relation

$$S = \log(\Omega) \qquad [5]$$

where units have been chosen such that the Boltzmann constant is unity. For a black hole, the macrostates are specified by its mass, charge, and angular momentum. No-hair theorems, however, seem to suggest that there are no other properties and hence no obvious candidate for microstates. In the absence of such a statistical basis, one would be inevitably led to the conclusion that there is loss of information in processes involving black holes.

In a consistent quantum theory of gravity, there would be such a statistical basis since quantum mechanics is unitary. String theory is a strong candidate for a unified theory which contains gravity. Indeed, string theory provides a microscopic description for a class of black holes.

## Black Hole Solutions in String Theory

Perturbatively, the basic excitations of string theory are fundamental closed and open strings characterized by a string tension $T_s$ and hence a length scale, the string length $l_s = 1/\sqrt{2\pi T_s}$. Consistency requires that the string should be able to propagate in ten spacetime dimensions and should be supersymmetric at the fundamental level. Formulated in this fashion, there are several consistent string theories: type IIA, type IIB, and heterotic string theory (which contain only closed strings perturbatively) and type I theory (which contains both open and closed strings).

At energies much smaller than $1/l_s$, only the massless modes of the string can be excited. For all these string theories, the massless spectrum of closed strings contains the graviton and the low-energy dynamics is given by the appropriate supersymmetric generalization of general relativity, supergravity. In addition, the closed-string spectrum contains a neutral scalar field, the dilaton $\phi$, whose expectation value gives rise to a dimensionless parameter governing interactions, called the string coupling $g_s$:

$$g_s = e^{<\phi>} \qquad [6]$$

The ten-dimensional gravitational constant is given by

$$G_{10} = 8\pi^6 g_s^2 l_s^8 \qquad [7]$$

Ten-dimensional supergravity has a wide variety of black hole solutions, the simplest of which is the straightforward generalization of the Schwarzschild solution.

### Black $p$-Brane Solutions

More significantly, there are solutions which are charged with respect to the various gauge fields that appear in the supergravity spectrum. Generically, these charged solutions represent extended objects. For accounts of such solutions, see Maldacena (1996).

Consider, for example, the supergravity which follows from type IIB string theory. This theory has a pair of 2-form gauge fields $B_{MN}$ and $B'_{MN}$ and a 4-form gauge field $A_{MNPQ}$ with a self-dual field strength. Just as an ordinary point electric charge produces a 1-form gauge field, a $(p+1)$-form gauge field may be sourced by an electrically charged $p$-dimensional extended object. The corresponding field strength is a $(p+2)$-form, whose Hodge dual in $d$ spacetime dimensions is a $(d-p-2)$ form. This shows that there should be magnetically charged $(d-p-4)$-dimensional extended objects as well. These extended objects are called "branes."

In the type IIB example, there should be two kinds of one-dimensional extended objects which carry electric charge under $B_{MN}, B'_{MN}$, called the F-string and the D-string, respectively. There are also two kinds of five-dimensional branes which carry magnetic charges under $B_{MN}, B'_{MN}$, called the NS 5-brane and D5 brane, respectively. Finally, there should be a 3-brane, since the corresponding 5-form field strength is self-dual as well as a D7 brane. A similar catalog can be prepared for other string theories, as well as for 11-dimensional supergravity, which is the low-energy limit of M-theory.

The classical solutions for a set of $p$-branes of the same kind generally have inner and outer horizons which have the topology $t \times S^{8-p} \times R^p$. The outer horizon is then associated with a Hawking temperature and a Bekenstein–Hawking entropy. Of particular interest are extremal limits. In this limit, the inner and outer horizons coincide and the mass density is simply proportional to the charge. Given some charge, the extremal solution has the lowest energy. Extremal limits are interesting because in supergravity these correspond to solutions in which some of the supersymmetries (in this case, half of the supersymmetries) are retained – such solutions are called Bogomolny–Prasad–Sommerfeld (BPS) saturated solutions. The charge in question appears as a central charge in the extended supersymmetry algebra. This fact may be used to show that such BPS solutions are absolutely stable. Indeed, for the particular solution considered here, the Hawking temperature $T_H \to 0$, so that there is no Hawking radiation, as required by stability. Furthermore, the entropy $S_{BH} \to 0$. The horizon shrinks to a point which appears as a naked null singularity.

All the ten dimensions of string theory need not be noncompact. In fact, to describe the real world, one must have a solution of string theory in which six of the dimensions are wrapped up and form a compact space. In principle, however, one can compactify any number of dimensions. In the above example of a $p$-brane, it is trivial to compactify the directions along which the brane is extended to a $p$-dimensional torus, $T^p$, which can be chosen to be a product of $p$ circles each of radius $R$. At length scales much smaller than $R$, the theory then becomes a $(10-p)$-dimensional theory. The $p$-brane appears as a black hole with a spherical horizon and, since the original $p$-form gauge field now behaves as an ordinary 1-form gauge field with a nonzero time component, this is an electrically charged black hole.

### D1–D5–*N* System and Five-Dimensional Black Holes

For reasons which will become clear in the next section, it is useful to get extremal black holes with large horizon areas, so that Hawking's semiclassical formulas are valid. It turns out that such solutions involve branes of various types which intersect each other and are suitably wrapped on compact internal spaces. Such black holes then have necessarily different kinds of charges. It turns out that the simplest case is a five-dimensional black hole with three kinds of charges, which is obtained by brane systems wrapped on a compact five-dimensional space. An example is a type IIB solution which has D5 branes which are wrapped on either $T^4 \times S^1$ or $K3 \times S^1$, together with D1 branes wrapped on the $S^1$ as well as some momentum along the $S^1$. From the noncompact five-dimensional point of view, this is a black hole with three kinds of gauge charges: the D5 charge $Q_5$, the D1 charge $Q_1$, and a Kaluza–Klein charge $N$ coming from the momentum $P = N/R$ along the circle of radius $R$.

When the internal space is $T^4 \times S^1$ the five-dimensional Einstein frame metric is given by

$$ds^2 = -[f(r)]^{-2/3}\left(1 - \frac{r_0^2}{r^2}\right)dt^2 + [f(r)]^{1/3}\left[\frac{dr^2}{(1 - r_0^2/r^2)} + r^2\, d\Omega_3^2\right] \qquad [8]$$

where

$$f(r) = \left(1 + \frac{r_0^2 \sinh^2 \alpha_1}{r^2}\right)\left(1 + \frac{r_0^2 \sinh^2 \alpha_5}{r^2}\right) \times \left(1 + \frac{r_0^2 \sinh^2 \sigma}{r^2}\right) \qquad [9]$$

and the three charges are

$$Q_1 = \frac{V r_0^2 \sinh 2\alpha_1}{32\pi^4 g_s l_s^6}, \qquad Q_5 = \frac{r_0^2 \sinh 2\alpha_5}{2 g_s l_s^2}$$
$$N = \frac{V R^2}{32\pi^4 l_s^8 g_s^2} r_0^2 \sinh 2\sigma \qquad [10]$$

where $V$ is the volume of the $T^4$ and $R$ is the radius of the circle $S^1$.

The ADM mass of the black hole is

$$M_{\text{ADM}} = \frac{R V r_0^2}{32\pi^4 g_s^2 l_s^8} \times [\cosh 2\alpha_1 + \cosh 2\alpha_5 + \cosh 2\alpha] \qquad [11]$$

The Bekenstein–Hawking entropy is given by

$$S_{\text{BH}} = \frac{R V r_0^3}{8\pi^3 l_s^8 g_s^2} \cosh \alpha_1 \cosh \alpha_5 \cosh \sigma \qquad [12]$$

while the Hawking temperature is

$$T_{\text{H}} = \frac{1}{2\pi r_0 \cosh \alpha_1 \cosh \alpha_5 \cosh \sigma} \qquad [13]$$

The extremal limit of this solution is given by

$$r_0 \to 0, \quad \alpha_1, \alpha_5, \sigma \to \infty$$
$$Q_1,\, Q_5,\, N = \text{fixed} \qquad [14]$$

The extremal solution is a BPS saturated state and retains four of the original supersymmetries. In this limit, the inner and outer horizons coincide. However, the horizon is now a smooth $S^3$ with a finite area in the Einstein frame metric. Consequently, the extremal Bekenstein–Hawking entropy is also finite and may be seen to be

$$S_{\text{BH}}^{\text{3-charge extremal}} = 2\pi\sqrt{Q_1 Q_5 N} \qquad [15]$$

The temperature, however, is zero in this limit, which is consistent with the stability of a BPS saturated state.

The above five-dimensional black hole is in fact a generalization of the Reissner–Nordtsrom black hole. Similar solutions with large horizon areas in the extremal limit can be constructed in four dimensions. One such construction is in the IIB theory wrapped on $T^6$ in which there are four sets of D3 branes which wrap four different $T^3$'s contained in the $T^6$. Black holes with lower supersymmetry may be obtained by replacing the $T^6$ by a Calabi–Yau space.

### Duality and Branes

String theory has a rich set of symmetries called duality symmetries which relate different kinds of string theories that are suitably compactified. These symmetries relate different classical solutions. For example, application of these symmetries relate the five-dimensional black holes above with other five-dimensional black holes with different kinds of charges. Furthermore, at the level of supergravity, these various theories may be derived from a yet unknown 11-dimensional theory called the M-theory whose low-energy limit is 11-dimensional supergravity.

## Branes in String Theory

For a given string theory, the perturbative spectrum consists of strings. However, at the nonperturbative

level, there are, in addition, extended objects of other dimensionalities. Duality symmetries imply that these extended objects are as "fundamental" as the strings themselves. Such extended objects are also called branes. For an exhaustive account of branes in string theory, see Johnson (2003).

Like their counterparts in supergravity, branes in string theory are typically charged with respect to some gauge fields. While supergravity solutions are possible with any value of the charge, in string theory the brane charges have to be quantized. Multiple units of the minimum quantum of charge can appear as collections of branes each with unit charge or, alternatively, branes which wrap around compact cycles in space a multiple number of times.

### D-Branes

The extended objects in string theory are described in terms of their collective excitations. These are best understood for the class of branes called D-branes in the type II theory, discovered by Polchinski. These are D1, D3, D5, and D7 branes in type IIB and D0, D2, D4, and D6 branes in type IIA theory. D$p$ branes are characterized by the fact that they couple to, and act as sources for, $(p+1)$-form gauge fields which belong to the Ramond–Ramond sector of the theory. Collective excitations of a $p$-dimensional extended object in field theory are expected to be described by waves on its $(p+1)$-dimensional world volume. The collective coordinate action would be a quantum field theory which has vectors, corresponding to longitudinal oscillations of the brane, and scalars which correspond to transverse oscillations. For D-branes in string theory, the theory of collective excitations is a string field theory of open strings whose endpoints lie on the brane. (This is the origin of the nomenclature D-brane: an open string whose ends are constrained to lie on the brane has a world-sheet description in which the bosonic fields corresponding to transverse target space coordinates have Dirichlet boundary conditions.) The lowest-energy states of open superstrings are ordinary massless gauge fields and their supersymmetric partners so that the low-energy limit of the string field theory is a supersymmetric gauge theory.

The fact that the underlying theory is a string theory has an important consequence. For a system of $N$ parallel D-branes of the same type, one would have open strings which join different branes as well as the same brane. The low-energy theory then becomes a supersymmetric nonabelian gauge theory with gauge group U($N$). In a suitable gauge, the off-diagonal gauge fields and their supersymmetric partners (which include scalar fields in the adjoint representation) are the low-energy degrees of freedom of open strings which connect different branes.

The mass density or tension $T_p$ of a single D$p$ brane is given by

$$T_p = \frac{1}{g_s (2\pi)^p l_s^{p+1}} \qquad [16]$$

This couples to the $(p+1)$-form gauge field with a charge

$$\mu_p = g_s T_p \qquad [17]$$

and the Yang–Mills coupling constant for the collective theory on the brane world volume is given by

$$g^2_{\mathrm{YM-D}p} = (2\pi)^{p-2} g_s l_s^{p-3} \qquad [18]$$

The ground state of a single D$p$ brane is a BPS state which preserves 16 of the 32 supersymmetries of the original theory. One consequence of this is that two or more parallel D$p$ branes of the same type form a threshold bound state preserving the same supersymmetries, with no net force between them. As a result, the tension of $N$ parallel D$p$ branes is simply $NT_p$.

Branes of different dimensionalities can also form bound states. Of particular interest are configurations which can form threshold bound states which preserve some supersymmetries. For example, a set of $N_1$ parallel D$p$ branes can form a threshold bound state with a set of $N_2$ parallel D$(4+p)$ branes with all the $p$ branes lying entirely along the $(4+p)$-branes. This configuration is also a BPS saturated state preserving eight of the original supersymmetries and would have charges under both $(p+1)$-form and $(p+5)$-form gauge potentials. The BPS nature ensures that the total mass density is the sum of the individual mass densities.

### NS Branes

The other extended objects in string theory are called NS branes since they couple to $p$-form gauge fields which arise from the Neveu–Schwarz/Neveu–Schwarz sector of the world-sheet theory. These are present in all the five string theories and appear in two types. The first is a macroscopic fundamental string which may be wound around a compact direction. The second is called a solitonic 5-brane. While the collective dynamics of a fundamental string is the standard world-sheet description of string theory, the description for the NS 5-brane is rather complicated and not known in full detail. The rest of this article deals exclusively with D-branes.

## D-Branes and Black Branes

The idea that black holes correspond to highly degenerate states in string theory is quite old and dates back to 't Hooft (1990) and Susskind (1993). In the following two sections we discuss such black holes which are described by D-branes. For reviews see Maldacena (1996), Das and Mathur (2001), and David *et al.* (2002).

We have so far discussed the string-theoretic branes in two different ways. In the first description, branes are solutions of the low-energy equations of motion – this is the setting in which branes provide conventional descriptions of black holes. In the second description, branes are certain states in the quantum theory of superstrings. More specifically, D-branes are described in terms of states of the open-string field theory which lives on the branes. The first description is necessarily approximate. On the other hand, the second description is exact in principle, although in practice one might not know how to write down and analyze the string-field theory in an exact fashion.

The description in terms of open-string field theory should reduce to the description in terms of a classical solution when the charges and masses become large. If black-hole thermodynamics has a microscopic origin, D-branes should be highly degenerate states in this limit and the entropy should be given by the Boltzmann formula. Furthermore, Hawking radiation should be understood as an ordinary decay process.

For a system of $Q_p$ parallel D$p$ branes, the mass is $Q_p/g_s$, while Newton's gravitational constant $G \sim g_s^2$. Gravitational effects are controlled by $GM \sim g_s Q_p$. A semiclassical limit in closed-string theory requires $g_s \to 0$, while a nontrivial gravitational effect in this limit requires $g_s Q_p$ finite, which implies one must have $Q_p \gg 1$. Furthermore, when $g_s Q_p \gg 1$ the typical curvatures are small compared to the string scale and the semiclassical string theory reduces to classical supergravity. This is the limit in which branes are well described as classical solutions.

Similar considerations apply for brane systems with multiple charges. For example, in the D1–D5–$N$ system the classical solution becomes a good description when all the quantities $g_s Q_1$, $g_s Q_5$, and $g_s^2 N$ become large. (The relevant quantity which comes with the momentum has $g_s^2$ rather than $g_s$ because the mass contribution from the momentum is simply $N/R$ without any inverse power of $g_s$.) However, $g_s$ is the square of the coupling constant of the open-string theory living on the brane – in fact, eqn [18] shows this relation in the low-energy limit. It is well known that in a U($Q_p$) gauge theory the real coupling constant is $g_{\mathrm{YM}}\sqrt{Q_P} \sim \sqrt{g_s Q_p}$. This means that the semiclassical limit corresponds to a strongly coupled string-field theory which reduces to strongly coupled gauge theory in the low-energy limit and the picture of D-branes as a collection of open strings is not very useful. In fact, known calculational methods in gauge theory or open-string theory are not valid in this regime.

### Microscopic Entropy for Two-Charge Systems

The prospects are much better for extremal black holes, which appear as BPS states in string theory. This is because the spectra of BPS states do not depend on the coupling. The degeneracy of such states may therefore be calculated at weak coupling, where techniques are well known and the result can be extrapolated to strong coupling without change.

The simplest BPS state is the ground state of a set of parallel D-branes of the same type. This state is indeed 128-fold degenerate, which would imply a microscopic entropy. This entropy, however, is small and therefore invisible in the corresponding classical solution. Indeed, the classical solution shows that in the extremal limit the horizon area is zero, leading to a vanishing Bekenstein–Hawking entropy.

The next interesting class of states consists of threshold bound states with two kinds of charges. Consider, for example, the D1–D5 system on $T^4 \times S^1$ considered above with no momentum along the D1's. By known duality transformations, this is equivalent to a fundamental IIB string which is wound $Q_5$ times around the $S^1$ and with a net momentum $P = Q_1/2\pi Q_5 R$ (where $R$ is the radius of the $S^1$), with four of the transverse directions compactified on a $T^4$. For this system, it is easy to count the number of states for given values of $Q_1$ and $Q_5$ at weak string coupling by simply enumerating the perturbative oscillator states of the string. For large values of $Q_1$ and $Q_5$, we can alternatively calculate this entropy by using a canonical ensemble of eight massless bosons corresponding to the eight transverse polarizations and their supersymmetric partners – eight massless fermions – moving on the string with some temperature $T$ and a chemical potential $\alpha$ for the total momentum.

Consider a noninteracting gas of $f$ massless bosons and $f$ fermions living on a circle with circumference $L$. The average number of left- and right-moving particles with some energy $e$, denoted by $\rho_L$, $\rho_R$, respectively, are

$$\rho_i(e) = \frac{1}{e^{e/T_i} \pm 1}, \quad i = L, R \qquad [19]$$

where the ± sign refers to fermions and bosons, respectively, and we have introduced left- and right-moving temperatures $T_L, T_R$. The physical temperature is

$$\frac{1}{T} = \frac{1}{2}\left(\frac{1}{T_L} + \frac{1}{T_R}\right) \qquad [20]$$

The extensive quantities, such as the energy $E$, momentum $P$, and entropy $S$, then become the sum of left- and right-moving pieces:

$$E = E_L + E_R, \quad P = P_L + P_R, \quad S = S_L + S_R \qquad [21]$$

and the distribution function [19] leads to the following thermodynamic relations:

$$T_i = \sqrt{\frac{3E_i}{L\pi f}} = \frac{4S_i}{\pi f L}, \quad i = \mathrm{L, R} \qquad [22]$$

Since the total momentum $P = P_R + P_L = E_R - E_L$ is nonzero, the lowest-energy state is clearly the one in which all the particles move in the same direction, for example, right moving. This is a BPS state and corresponds to the extremal solution in supergravity. Then $E = E_R = P = P_R$. This approach to the black hole entropy was initiated by Das and Mathur (1996) and Callan and Maldacena (1996).

For our two-charge system, $f = 8, P = 2\pi Q_1/L$, and $L = 2\pi R Q_1 Q_5$. Using [22] we get

$$S_{\text{micro}}^{\text{2-charge-II}} = 2\pi\sqrt{2Q_1Q_5} \qquad [23]$$

This is the microscopic entropy for the fundamental string with momentum in the type II theory. By duality, this is also the microscopic entropy of the D1–D5 system. This is a large number which should agree with the macroscopic entropy calculated from the corresponding classical solution.

The discussion is almost identical for the fundamental heterotic string, except that now we have 24 right-moving bosons, eight left-moving bosons, and eight left-moving fermions, and the BPS state consists only of right movers. If $n_w$ denotes the winding number and $n_p$ the quantized momentum the extremal heterotic string entropy is

$$S_{\text{micro}}^{\text{2-charge heterotic}} = 4\pi\sqrt{n_p n_w} \qquad [24]$$

The supergravity solution for the D1–D5 system may be obtained by substituting $\sigma = 0$ in eqns [8]–[13]. In the extremal limit, the classical Bekenstein–Hawking entropy vanishes as is clear from the expression [15], in which $N = 0$. This appears to be in contradiction with the fact that the state has a large microscopic entropy.

The key point, however, is that the two-charge solution has a singular horizon where the string frame curvature is large. Consequently, low-energy tree-level supergravity breaks down near the horizon and higher-derivative terms (e.g., higher powers of curvature) become important. This issue has been best studied for the fundamental heterotic string compactified on $T^6$. This is dual to the D1–D5 system in type IIB theory compactified on $K3 \times T^2$. The classical supergravity solution is then a singular black hole in four spacetime dimensions. In one of the first papers on the string-theoretic understanding of black hole thermodynamics, Sen (1995) showed that, for large $n_p$, $n_w$, string-loop effects are small near the horizon so that the only relevant corrections are higher-derivative terms coming from integrating out the massive modes of the string at tree level. Furthermore, a robust scaling argument shows that regardless of the detailed nature of the derivative corrections, the macroscopic entropy defined through the horizon area must be of the form $a\sqrt{n_p n_w}$, where $a$ is a pure number. Finally, one can define a "stretched horizon" as the surface where the curvature becomes of the order of the string scale and the area of the stretched horizon is indeed proportional to $\sqrt{n_p n_w}$. This result gives a strong indication that string theory provides a microscopic basis for black hole thermodynamics, although the coefficient $a$ cannot be determined without more detailed knowledge of higher-derivative terms.

## Microscopic Entropy of Extremal Three-Charge System

Brane bound states with three kinds of charge provide examples of black holes whose extremal limits have large horizons with curvatures much smaller than the string scale. In this case, a microscopic count of states in string theory should exactly account for the Bekenstein–Hawking formula, without corrections coming from higher derivatives. This is indeed true, as first found by Strominger and Vafa (1996). In the following, we will outline how this calculation can be done in the D1–D5–$N$ system on $K3 \times S^1$ or $T^4 \times S^1$ following the treatment of Dijkgraaf *et al.* (1996).

D1 branes can be considered as "instanton strings" in the six-dimensional supersymmetric $U(Q_5)$ gauge theory of D5 branes (actually, these should be called solitonic strings rather than instantons, since the configurations are time independent). The total instanton number is the D1-brane charge $Q_1$. The moduli space of these instantons is then a blown-up version of the

orbifold $(T^4)^{Q_1Q_5}/S(Q_1Q_5)$ or $(K3)^{Q_1Q_5}/S(Q_1Q_5)$ and is $4Q_1Q_5$ dimensional. Since any instanton configuration is independent of time $x^0$ and the $S^1$ direction $x^5$, the collective coordinate dynamics is a $(1+1)$-dimensional field theory which lives in the $(x^0, x^5)$ space. At low energies, this flows to a conformal field theory with a central charge $c = 6Q_1Q_5$ since there are $4Q_1Q_5$ bosons each contributing 1 to the central charge and an equal number of fermions each contributing 1/2. The BPS state with momentum $N/R$ is a purely right- or left-moving state in this conformal field theory which has a conformal weight $N$. From general principles of conformal invariance, the degeneracy of such states for large $N$ is given by Cardy's formula

$$d(N) \sim e^{2\pi\sqrt{cN/6}} \qquad [25]$$

so that the microscopic entropy is

$$S^{\text{micro}}_{\text{3-charge}} = \log d(n) = 2\pi\sqrt{cN/6} \qquad [26]$$

Substituting the value of $c = 6Q_1Q_5$, this is in exact agreement with the Bekenstein–Hawking entropy of the classical solution given in [15].

## Nonextremal Black Holes and Hawking Radiation

The BPS property of ground states of D-brane systems enables us to compute the degeneracy of microstates exactly in the regime of parameters where the state can be reliably described as a black hole solution in the low-energy theory. However, extremal black holes have vanishing temperature and do not radiate. To understand the microscopic origins of Hawking radiation, one has to go away from extremality. Such states are not supersymmetric and an extrapolation of weak-coupling calculations to strong coupling is not *a priori* justified. Nevertheless, it turns out that for small departures from extremality, weak-coupling results still reproduce semiclassical answers for entropy, temperature, and luminosity.

### Near-Extremal Entropy

Nonextremal properties are best understood for the D1–D5–$N$ system on $T^4 \times S^1$. In the orbifold limit, the conformal field theory which describes the low-energy dynamics is equivalent to a gas of strings which are wound around the $S^1$ and which can oscillate along the $T^4$. The total winding number is $k = Q_1Q_5$ and may be achieved by sets of strings which are multiply wound in various ways. As argued below, entropically the most favored configuration is a single long string wound around $Q_1Q_5$ times. Thus, the thermodynamics may be analyzed exactly along the lines of the fundamental string in the previous section. The thermodynamic relations are given by [22] with $f = 4$ and $L = 2\pi RQ_1Q_5$. The extremal state consists entirely of right movers and $E = E_R = N/R$. Substituting these values in [22] yields the correct formula for the microscopic entropy

$$S^{\text{3-charge}}_{\text{micro}} = 2\pi\sqrt{Q_1Q_5N} \qquad [27]$$

The same expression follows if $f = 4Q_1Q_5$ and $L = 2\pi R$ corresponding to $Q_1Q_5$ singly wound strings. However, for statistical methods to hold, the entropy must be much larger than the number of flavors. The ratio of the entropy to the number of flavors is $S/f \sim \sqrt{N/Q_1Q_5}$ for multiple singly wound strings and is not guaranteed to be large when all of $Q_1$, $Q_5$, $N$ are large. On the other hand, this ratio is $S/f \sim \sqrt{Q_1Q_5N}$ for the long string. This shows that the long string is always entropically favored.

A departure from the extremal state is achieved by adding a left-moving momentum $2\pi n/L$ as well as a right-moving momentum $2\pi n/L$ to the extremal state, thus adding energy to the system but maintaining the total momentum. For the long string, this yields

$$S_R = 2\pi\sqrt{Q_1Q_5N + n}, \qquad S_L = 2\pi\sqrt{n} \qquad [28]$$

For small departures from extremality, $n \ll N$, the expressions for the total entropy and temperature as a function of the excess energy $\Delta E = 2n/Q_1Q_5$ agree exactly with the near-extremal Bekenstein–Hawking entropy and the Hawking temperature of the classical solution, as shown by Callan and Maldacena (1996) and by Horowitz and Strominger.

The necessity of the long string appears in another important physical consideration. For statistical mechanics to be valid, the specific heat of the system has to be larger than unity. This implies that for the case considered here the energy gap $\Delta E$ must be larger than $1/RQ_1Q_5$, which is precisely what the long string yields.

### Hawking Radiation

A nonextremal state described above is unstable, since a left mover can annihilate a right mover into a closed-string mode which may leave the brane system and propagate to the asymptotic region. The resulting closed-string state will be in a thermal state whose temperature is the physical temperature of the initial state. This process is the microscopic

description of Hawking radiation. The decay rate is related to the absorption cross section of the corresponding mode by the principle of detailed balance, encoded in eqn [4].

From the point of view of the classical solution, the absorption cross section can be calculated by solving the linearized wave equation in the background geometry and calculating the ratio of the incident and reflected waves. It follows from these calculations that at low energies, absorption (and hence emission) are dominated by massless minimally coupled scalars. In fact, for any spherically symmetric black hole in any number of dimensions, there is a general theorem which ensures that the low-energy limit of this absorption cross section is exactly equal to the horizon area.

In the microscopic model for the three-charge black hole, this absorption cross section may be calculated by the usual rules of quantum mechanics. In the long-string limit and in the approximation that the modes on the long string form a dilute gas, the result has been derived by Das and Mathur (1996):

$$\sigma(\omega) = \frac{2\pi G_{10} Q_1 Q_5}{V} \omega \frac{e^{\omega/T} - 1}{(e^{\omega/2T_R} - 1)(e^{\omega/2T_L} - 1)} \quad [29]$$

where $V$ is the volume of the $T^4$ and $T$ is the physical temperature given by [20]. For a near-extremal hole $T_R \gg T_L$, so that $T \sim 2T_L$. Then in the extreme low-energy limit $\omega \ll T_R$, so that the corresponding Bose factor may be approximated as $1/(e^{\omega/2T_R} - 1) \sim 2T_R/\omega$. The cross section [29] becomes

$$\sigma = \frac{4\pi Q_1 Q_5 G_{10} T_R}{V} = \frac{4 G_{10} S_R}{(2\pi R) V} = 4 G_5 S_{\text{extremal}} = A_H \quad [30]$$

where $G_5$ is the five-dimensional Newton's gravitational constant. We have used the relation [22] with $L = 2\pi R Q_1 Q_5$ and $f = 4$. The fact that in the near-extremal limit $S_R$ is simply the extremal entropy and the fact that the extremal entropy reproduces the Bekenstein–Hawking formula has been used as well. Thus, the microscopic cross section exactly reproduces the semiclassical result at low energies. Even more remarkably, the full cross section [29] agrees with the semiclassical answer for the gray-body factor for parameters which correspond to the dilute-gas regime, as shown by Maldacena and Strominger.

It is rather surprising that the results for microscopic absorption cross section calculated at weak coupling agree with the semiclassical answers, since the relevant process involves states which are not supersymmetric and therefore a naive extrapolation to strong coupling is not *a priori* justified. There are strong indications, however, that low-energy nonrenormalization theorems are at work. This agreement has been established not only for black holes with finite-horizon areas, but also for other systems with no horizons – most significantly, a set of parallel 3-branes – and forms the basis for Maldacena's conjecture about AdS/CFT Correspondence (see AdS/CFT Correspondence).

## Effects of Higher-Derivative Terms

The classical low-energy limit of string theory is supergravity. The effects of the massive modes of the string as well as effect of string loops is to add terms to the supergravity action which involve higher number of spacetime derivatives, for example, terms containing higher powers of the curvature. In the presence of such terms, the Bekenstein–Hawking formula for black hole entropy [2] receives corrections which can be calculated in a systematic fashion. It turns out that for a class of extremal black holes, this corrected entropy as computed in the modified supergravity is also in exact agreement with a microscopic calculation.

One example of this agreement is provided by four-dimensional extremal black holes in type IIA string theory compactified on a Calabi–Yau manifold. These are obtained by wrapping D4 branes on three different 4-cycles on the Calabi–Yau and having in addition a number of D0 branes. Let $p^A$, $A = 1, \ldots, 3$ denote the three D4 charges and $q_0$ denote the D0 charge. The microscopic entropy of the BPS state can be computed by embedding this in M-theory:

$$S_{\text{micro}}^{\text{CY-Black hole}} = 2\pi \sqrt{\frac{1}{6} |q_0| (C_{ABC} p^A p^B p^C + c_{2A} p^A)} \quad [31]$$

where $C_{ABC}$ is the intersection number of the 4-cycles and $c_2$ denotes the second Chern class of the Calabi–Yau space. When all the charges $p^A$ are large, the term involving $c_2$ is subdominant. In this case, the result agrees with the Bekenstein–Hawking entropy of the corresponding classical solution. When the charges are not all large (so that the second term is appreciable), the curvatures of the supergravity solution become large at the horizon and higher-derivative corrections to the action cannot be ignored. In this particular case, it turns out that these higher-derivative corrections are string-loop corrections and can be computed using general properties of $N = 2$ supersymmetry, so that one can compute corrections to near-horizon geometry. Furthermore, one has to now modify the

expression for macroscopic entropy using the formalism of Wald. Putting these together, it is found that the macroscopic entropy following from the modified supergravity is in exact agreement with [31]. This subject is reviewed in Mohaupt (2000).

These methods have also been applied to the problem of two-charge black holes in heterotic string theory on $T6$ or, equivalently, type IIA on $K3 \times T^2$ (Dabholkar 2004). Recall that in this case the horizon of the usual supergravity solution is singular. It has been found that leading-order higher-derivative corrections smoothen out the horizon into a $AdS_2 \times S^2$ spacetime and the modified expression for the macroscopic entropy is again in exact agreement with the microscopic answer [23].

## Geometry of Microstates

A satisfactory solution of the information-loss paradox requires a much more detailed understanding of black holes in string theory. The discussion above shows that black holes have microstates which may be described well in the weak-coupling regime. It is interesting to ask whether there is a description of these microstates in the strong-coupling regime in terms of the effective geometry perceived by suitable probes. This question has been answered for the two-charge system in great detail (see Mathur (2004)). It turns out that the D1–D5 microstates can be described by perfectly smooth metrics with no horizons, and they asymptote to the standard two-charge metric discussed above. The location of the erstwhile stretched horizon marks the point where the different microstates start differing from each other significantly. Since each such geometry does not have a horizon, neither does it have any entropy – this is consistent with their identification with nondegenerate microstates. Indeed, the number of such microstates correctly accounts for the microscopic entropy. Whether a similar picture holds for the three-charge system remains to be seen in detail, although there are some indications that this may be true. In this approach, it is not yet fully understood how a horizon emerges and why the entropy scales as the horizon area.

## Outlook

One key feature of the understanding of black hole statistical mechanics from the dynamics of branes is the fact that a problem in gravity is mapped to a problem in a theory without gravity, for example, open-string field theory. In fact, the closed strings in the bulk are already contained in the spectrum of the open strings. This is a consequence of the basic duality between open strings and closed strings. Furthermore, the open-string theory lives in a lower-dimensional spacetime. This is a manifestation of the holographic principle. As argued by Maldacena, the presence of a horizon implies that the low-energy limit retains all the modes of the closed strings near the horizon, while it truncates the open-string theory to a gauge theory. Open–closed duality then reduces to gauge–string duality. This provides a strong evidence that black holes obey the normal laws of quantum mechanics and hence their time evolution is unitary.

One of the most outstanding problems in the subject is a proper understanding of neutral black holes. Most of the quantitative results described above depend on supersymmetry, which allows extrapolation of weak-coupling answers to the strong-coupling domain. Some of these results can be extended to situations which have small departures from supersymmetry, for example, near-extremal black holes. States corrresponding to neutral black holes are, however, far from supersymmetry and known calculational techniques fail. There are good reasons to expect, however, that the general philosophy – in particular the holographic principle – is still valid. Finally, so far string theory has been able to attack problems of eternal black holes. A satisfactory understanding of the information-loss problem requires an understanding of the dynamics of black hole formation and subsequent evaporation. Unfortunately, very little is known about this at the moment.

*See also:* AdS/CFT Correspondence; Black Hole Mechanics; Supergravity; Superstring Theories.

## Glossary

**ADM** (Arnowitt–Deser–Misner) **mass** – Mass of a gravitational background which is asymptotically flat.

**AdS$_n$** (anti-de Sitter space) – A space (or spacetime) with constant negative curvature in $n$ dimensions.

**BPS state** (Bogomolny–Prasad–Sommerfeld state) – In a theory of extended supersymmetry, a state that is invariant under a nontrivial subalgebra of the full supersymmetry algebra. These states always carry conserved charges, and supersymmetry determines the mass exactly in terms of the charges.

**Calabi–Yau space** – Complex Kahler manifold with vanishing first Chern class.

**Compactify** (n. compactification) – To consider a field or string theory in a spacetime some of whose spatial dimensions are compact.

**Dirichlet boundary condition** – The boundary condition which fixes the value of a field on the boundary.

**Duality** Equivalence of systems which appear to be distinct. For string theories, such equivalences relate string theories on different spacetimes as well as theories with different coupling constants.

**Einstein–Hilbert action** – The standard action for gravity which leads to Einstein's equation, $S = (1/16\pi G)\int d^d x \sqrt{g} R$, where $R$ is the Ricci scalar, $g$ denotes the determinant of the metric, and $G$ is Newton's gravitational constant.

**Instanton** – A classical solution of Euclidean field theory with finite action.

**Kaluza–Klein gauge field** – In a compactified theory, the gauge field which arises from the metric of the higher-dimensional theory.

**K3** – The unique Calabi–Yau manifold in four dimensions having an SU(2) holonomy.

**Loop levels** – In a Feynman diagram expansion of a field theory, terms which contribute in higher orders of the Planck constant $\hbar$.

**Macroscopic entropy** – Entropy associated with gravitational backgrounds via the Bekenstein–Hawking formula or its generalization.

**Microscopic entropy** – Entropy which follows from the degeneracy of states of a system via Boltzmann's relation.

**Minimally coupled scalar** – A scalar field whose equation of motion is the standard Klein–Gordon equation where the derivatives are covariant derivatives.

**Neveu–Schwarz/Neveu–Schwarz states** – In type I and II string theories, bosonic closed-string states whose left- and right-moving parts are bosonic.

**No-hair theorem** – A theorem in general relativity which states that black holes with nonsingular horizons are uniquely characterized by their mass, angular momenta, and charges which can couple to long-range gauge fields.

**Orbifold** – A coset space $M/G$ where $G$ is a group of discrete symmetries of a manifold $M$. If $G$ has a fixed point, the space is singular.

**p-Form** – A fully antisymmetric $p$-index tensor.

**Ramond–Ramond states** – In type I and II string theories, bosonic closed-string states whose left- and right-moving parts are fermionic.

**Reissner–Nordstrom black hole** – Black hole solution of general relativity with electric Maxwell charge.

**$S^n$** – $n$-Dimensional sphere.

**Supergravity** – Supersymmetric extension of general relativity.

**Supersymmetry** – A symmetry between bosons and fermions.

**Threshold bound state** – A bound state which is marginally bound, that is, the binding energy is zero.

**Tree level** – In a Feynman diagram expansion of a field theory, terms which contribute to lowest order of the Planck constant $\hbar$.

**U($N$)** – The group of $N \times N$ unitary matrices. If the determinant is unity, the subgroup is called SU($N$).

## Further Reading


Callan CG and Maldacena M (1996) D-brane approach to black hole quantum mechanics. *Nuclear Physics B* 472: 591 (arXiv:hep-th/9602043).

Dabholkar A (2004) Exact counting of black hole microstates, arXiv:hep-th/0409148.

Das SR and Mathur SD (1996) Comparing decay rates for black holes and D-branes. *Nuclear Physics B* 478: 561 (arXiv:hep-th/9606185).

Das SR and Mathur SD (2001) The quantum physics of black holes: results from string theory. *Annual Review of Nuclear and Particle Science* 50: 153 (arXiv:gr-qc/0105063).

David JR, Mandal G, and Wadia SR (2002) Microscopic formulation of black holes in string theory. *Physics Reports* 369: 549 (arXiv:hep-th/0203048).

Dijkgraaf R, Moore GW, and Verlinde E (1996) Elliptic genera of symmetric products and second quantized strings. *Communications in Mathematical Physics* 185: 197 (arXiv:hep-th/9608096).

't Hooft G (1990) The black hole interpretation of string theory. *Nuclear Physics B* 335: 138.

Johnson C (2003) *D-Branes*. Cambridge: Cambridge University Press.

Maldacena JM (1996) Black holes in string theory, arXiv:hep-th/9607235.

Maldacena J, Strominger A, and Witten E (1997) Black hole entropy in M-theory. *Journal of High Energy Physics* 9712: 002 (arXiv:hep-th/9711053).

Mathur SD (2004) Where are the states of a black hole?, arXiv:hep-th/0401115.

Mohaupt T (2000) Black hole entropy, special geometry and strings. *Fortschritte der Physik* 49: 3 (arXiv:hep-th/0007195).

Sen A (1995) Extremal black holes and elementary string states. *Modern Physics Letters A* 10: 2081.

Strominger A and Vafa C (1996) Microscopic origin of the Bekenstein–Hawking entropy. *Physics Letters B* 379: 99 (arXiv:hep-th/9601029).

Susskind L (1993) Some speculations about black hole entropy in string theory, arXiv:hep-th/9309145.

Wald R (1994) *Quantum Field Theory In Curved Space-Time and Black Hole Thermodynamics*. Chicago, IL: University of Chicago Press.

# Breaking Water Waves

**A Constantin**, Trinity College, Dublin,
Republic of Ireland



## Introduction

Watching the sea or a lake it is often possible to trace a wave as it propagates on the water's surface. One can roughly distinguish two types of breaking waves. All waves break while reaching the shore but certain waves break far from the shore. In the first case, the change in water depth or the presence of an obstacle (e.g., a rock) seems to cause wave breaking, while for certain waves within the second category, these factors appear not to be essential. It is a matter of observation that for many waves that break in the open water a drastic increase in their slope near breaking is noticeable. This leads us to the following mathematical definition: the wave profile gradually steepens as it propagates until it develops a point where the slope is vertical and the wave is said to have broken (Whitham 1980). Throughout this article, we are concerned with wave breaking that is not caused by a drastic change of the topography of the bottom; for a discussion of wave breaking at the beach we refer to Johnson (1997). The governing equations for water waves (see the next section) are too difficult to be dealt with in their full generality. Therefore, to gain some insight, one has to find simpler models that are more tractable mathematically. Investigating the properties of the model, certain predictions can be made. The conclusions reached will reflect reality only to some limited extent. The value of a model depends on the number and the degree of accuracy of physically useful deductions that can be made from it – the "truth" of the model is meaningless as all experiments contain inaccuracies and effects other than those accounted for (while deriving the model) cannot be totally excluded. We intend to discuss the way in which a recent model due to Camassa and Holm (1993) can lead to a better understanding of breaking water waves. Firstly we survey a few classical nonlinear partial differential equations that model the propagation of water waves over a flat bed (within the confines of the linear theory one cannot cope with the wave breaking phenomenon) and discuss their relevance to the study of breaking waves. We then analyze the breaking of waves within the context of the Camassa–Holm equation: existence of breaking waves, criteria that guarantee that a certain initial shape develops into a breaking wave, specific features of wave breaking (blow-up rate and blow-up set for certain types of breaking waves). We conclude the presentation with a discussion of the way in which solutions to the Camassa–Holm equation can be continued after wave breaking.

## The Governing Equations

The water waves that one typically sees propagating on the surface of the sea or on a lake are, as a matter of common experience, approximately two dimensional. That is, the motion is identical in any direction parallel to the crest line. To describe these waves, it suffices to consider a cross section of the flow that is perpendicular to the crest line. Choose Cartesian coordinates $(x, y)$ with the $y$-axis pointing vertically upwards and the $x$-axis being the direction of wave propagation, while the origin lies at the mean water level. Let $(u(t,x,y), v(t,x,y))$ be the velocity field of the flow, let $y = -d$ be the flat bed (for some fixed $d > 0$), and let $y = \eta(t,x)$ be the water's free surface. Homogeneity (constant density) is a physically reasonable assumption for gravity waves (Johnson 1997), and it implies the equation of mass conservation

$$u_x + v_y = 0 \qquad [1]$$

The inviscid setting is realistic since experimental evidence confirms that the length scales associated with an adjustment of the velocity distribution due to laminar viscosity or turbulent mixing are long compared to typical wavelengths. Under the assumption of inviscid flow the equation of motion is Euler's equation

$$\begin{aligned} u_t + uu_x + vu_y &= -P_x \\ v_t + uv_x + vv_y &= -P_y - g \end{aligned} \qquad [2]$$

where $P(t,x,y)$ denotes the pressure and $g$ is the gravitational constant of acceleration. The free surface decouples the motion of the water from that of the air so that (Johnson 1997) the dynamic boundary condition

$$P = P_0 \quad \text{on } y = \eta(t,x) \qquad [3]$$

must hold if we neglect surface tension, where $P_0$ is the (constant) atmospheric pressure. Moreover, since the same particles always form the free surface, we have the kinematic boundary condition

$$v = \eta_t + u\eta_x \quad \text{on } y = \eta(t,x) \qquad [4]$$

On the flat bed we have the kinematic boundary condition

$$v = 0 \quad \text{on } y = -d \qquad [5]$$

expressing the fact that the flow is tangent to the horizontal bed (or, equivalently, that water cannot penetrate the rigid bed). The governing equations for water waves are [1]–[5]. Other than the fact that they are highly nonlinear, a main difficulty in analyzing the governing equations lies in the fact that we deal with a free boundary problem: the free surface $y=\eta(t,x)$ is not specified *a priori*. In our discussion, we suppose that initially (at time $t=0$), a disturbance of the flat surface of still water was created and we analyze the subsequent motion of the water. The balance between the restoring gravity force and the inertia of the system governs the evolution of the mass of water and our primary objective is the behavior of the free surface.

An important category of flows are those of zero vorticity, characterized by the additional assumption

$$u_y = v_x \qquad [6]$$

The vorticity of a flow, $\omega = u_y - v_x$, measures the local spin or rotation of a fluid element. In flows for which [6] holds the local whirl is completely absent and for this reason such flows are called irrotational. Relation [6] ensures the existence of a velocity potential, namely a function $\phi(t,x,y)$ defined up to a constant via

$$\phi_x = u, \qquad \phi_y = v$$

Notice that [1] ensures that $\phi$ is a harmonic function, that is, $(\partial_x^2 + \partial_y^2)\phi = 0$. In this way, the powerful methods of complex analysis become available for the study of irrotational flows. Thus, while most water flows are with vorticity, the study of irrotational flows can be defended mathematically on grounds of beauty. Concerning the physical relevance of irrotational water flows, experimental evidence indicates that for waves entering a region of still water the assumption of irrotational flow is realistic (Johnson 1997). Moreover, as a consequence of Kelvin's circulation theorem (Acheson 1990), a water flow that is irrotational initially has to be irrotational at all later times. It is thus reasonable to consider that water motions starting from rest will remain irrotational at later times.

## Nonlinear Model Equations

Starting from the governing equations [1]–[6] one can derive a variety of model equations using the nondimensionalization and scaling approach: a suitable set of nondimensional variables is introduced, which, after scaling, leads to the appearance of parameters. The sizes and relative sizes of these parameters then govern the type of phenomenon that is of interest. An asymptotic expansion in one or several parameters yields an equation that is usually of significance in some region of space/time. The aim of this process is to obtain a simpler model that can be used to gain some understanding and to make some predictions for specific physical processes. This scaling method yields the Korteweg–de Vries (KdV) equation

$$\eta_t + \eta\eta_x + \eta_{xxx} = 0, \quad t > 0,\, x \in \mathbb{R} \qquad [7]$$

as a model for the unidirectional propagation of shallow water waves over a flat bed (Johnson 1997). In [7] the function $\eta(t,x)$ represents the height of the water's free surface above the flat bed. We would like to emphasize that the "shallow water" regime does not refer to water of insignificant depth – it indicates that the typical wavelength is much larger than the typical depth (e.g., tidal waves are considered to be shallow water waves although they affect the motion of the deep sea). The KdV model admits the solitary wave solutions

$$\eta_c(t,x) = 3c\,\mathrm{sech}^2\left(\frac{\sqrt{c}}{2}(x-ct)\right), \quad c \in \mathbb{R} \qquad [8]$$

For any fixed $c > 0$, the profile $\eta_c$ propagates without change of form at constant speed $c$ on the surface on the water, that is, it represents a traveling wave. Since the profiles [8] of the traveling waves drop rapidly to the undisturbed water level $\eta = 0$ ahead and behind the crest of the wave, $\eta_c$ are called solitary waves. Notice that [8] shows that taller solitary waves travel faster. They have other special properties: an initial profile consisting of two solitary waves, with the taller preceding the smaller one, evolves in such a way that the taller wave catches up the other, there is a period of complicated nonlinear interaction but eventually both solitary waves emerge completely unscathed! This special type of nonlinear interaction (the superposition principle is not valid since KdV is a nonlinear equation) in which solitary waves regain their form upon collision occurs only for special equations, in which case the solitary waves are called solitons. A further interesting property of the KdV model, relevant for the understanding of the interaction of solitons, is the fact that it is completely integrable (McKean 1998): there is a transformation which converts the equation into an infinite sequence of linear ordinary differential equations which can be trivially integrated. Moreover, the KdV-solitons $\eta_c$ are stable: an initial profile that is close to the form of a soliton will evolve into a wave that at any later times has a form close to that of a soliton (Benjamin 1972). Despite all these intriguing features of the KdV-model, for all initial profiles $x \mapsto \eta(0,x)$ within the Sobolev space $H^1(\mathbb{R})$ of square-integrable functions with a square-integrable distributional derivative, eqn [7] has a unique solution

defined for all times $t \geq 0$ (cf. Kenig *et al.* (1996)) so that the KdV model cannot be used to shed light on the wave breaking phenomenon.

Whitham (1980) suggested the equation

$$\eta_t + \eta\eta_x + \int_{\mathbb{R}} k(x-y)\eta_y(t,y)\mathrm{d}y = 0 \qquad [9]$$

for the free surface profile $x \mapsto \eta(t,x)$, with the singular kernel

$$k(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \left(\frac{\tanh(\xi)}{\xi}\right)^{1/2} \mathrm{e}^{\mathrm{i}\xi x}\mathrm{d}\xi$$

to model wave breaking. It can be shown (see Constantin and Escher (1998) and references therein) that [9] describes wave breaking: there are smooth initial profiles $x \mapsto \eta(0,x)$ such that the resulting unique solution of [9] exists on a maximal time interval $[0,T)$ with

$$\sup_{(t,x)\in[0,T)\times\mathbb{R}} \{\eta(t,x)\} < \infty$$

$$\inf_{x\in\mathbb{R}} \{\eta_x(t,x)\} \to -\infty \quad \text{as} \quad t \uparrow T$$

(the solution remains bounded but its slope becomes infinite in finite time). However, in contrast to the KdV model, eqn [9] is not integrable and does not possess soliton solutions. As emphasized by Whitham (1980), it is intriguing to find models for water waves which exhibit both soliton interaction and wave breaking.

The Camassa–Holm equation

$$\eta_t - \eta_{txx} + 3\eta\eta_x = 2\eta_x\eta_{xx} + \eta\eta_{xxx} \qquad [10]$$

was first obtained by Fokas and Fuchssteiner (1981/82) as a nonlinear partial differential equation with infinitely many conservation laws. Camassa and Holm (1993) derived [10] as a model for shallow water waves, established that the equation possesses soliton solutions and found that it is formally integrable (for a discussion of the integrability issues we refer to Constantin (2001), and Lenells (2002)). Moreover, the solitons of [10] are stable (Constantin and Strauss 2003). An astonishing plentitude of structures is tied into the Camassa–Holm equation: [10] is a re-expression of geodesic flow on the diffeomorphism group (Constantin 2000, Kouranbaeva 1999), a property that can be used to show that the least action principle holds in the sense that there is a unique flow transforming a wave profile into a nearby profile within the class of flows that minimize the kinetic energy (see the discussion in Constantin (2000) and Constantin and Kolev (2003)). Interestingly, the Camassa–Holm equation also models wave breaking. More precisely (see the discussion in Constantin (2000)), for any initial data $x \mapsto \eta_0(x) = \eta(0,x)$ in $H^3(\mathbb{R})$ there is a unique solution of [10] defined on some maximal time interval $[0,T)$ and the solution stays uniformly bounded on $[0,T)$ with

$$\lim_{t\uparrow T}\left(\inf_{x\in\mathbb{R}}\{\eta_x(t,x)\}(T-t)\right) = -2 \text{ if } T < \infty$$

In addition to this, for a large class of initial data, there is precisely one point where the slope of the wave becomes infinite at breaking time (Constantin 2000): if $\eta_0 \not\equiv 0$ is odd and such that $\eta_0(x) - \eta_0''(x) \geq 0$ for all $x \leq 0$, then the corresponding wave $t \mapsto [x \mapsto \eta(t,x)]$ will break in finite time $T < \infty$ and

$$\lim_{t\uparrow T} \eta_x(t,0) = -\infty$$

whereas

$$|\eta_x(t,x)| \leq K + K\frac{\cosh(x)}{|\sinh(x)|}$$

$$t \in [0,T), \quad x \neq 0$$

for some constant $K > 0$. Thus, the Camassa–Holm model is an integrable infinite-dimensional Hamiltonian system with stable solitons and eqn [10] admits also breaking waves as local solutions (see Constantin and Escher (1998) and McKean (1998) and references therein for further results on wave breaking for the Camassa–Holm equation).

We conclude our discussion by pointing out that it is possible to continue solutions of the Camassa–Holm equation past the breaking time. For this purpose it is convenient to rewrite [10] as the nonlinear nonlocal conservation law

$$\eta_t + \eta\eta_x + \frac{1}{2}\partial_x \int_{\mathbb{R}} \mathrm{e}^{-|x-y|}\left(\eta^2 + \frac{\eta_x^2}{2}\right)\mathrm{d}y = 0 \qquad [11]$$

reminiscent to some extent to the form of [7] and [9] and obtained by formally applying the operator $(1-\partial_x^2)^{-1}$ to [10] in view of the fact that

$$(1-\partial_x^2)^{-1}f = P * f \quad \text{for} \quad f \in L^2(\mathbb{R})$$

the kernel of the convolution being

$$P(x) = \tfrac{1}{2}\mathrm{e}^{-|x|}, \quad x \in \mathbb{R}$$

By introducing a new set of independent and dependent variables it is possible to resolve all singularities due to wave breaking in the sense that [11] is transformed into a semilinear system, the unique solution of which can be obtained as a fixed point of a contractive operator (Bressan and Constantin 2005). In terms of [11], a semigroup of global conservative solutions (in the sense that the total energy

$$\frac{1}{2}\int_{\mathbb{R}} (\eta^2 + \eta_x^2)\mathrm{d}x$$

equals a constant, for almost every time), depending continuously on the initial data $\eta(0,\cdot) \in H^1(\mathbb{R})$, is thus constructed.

*See also:* Compressible Flows: Mathematical Theory; Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Integrable Systems: Overview; Interfaces and Multicomponent Fluids.

## Further Reading

Acheson DJ (1990) *Elementary Fluid Dynamics*. New York: Oxford University Press.
Benjamin TB (1992) The stability of solitary waves. *Proceedings of the Royal Society of London Series A* 328: 153–183.
Bressan A and Constantin A (2005) Global conservative solutions of the Camassa–Holm equation, *Preprints on Conservation Laws* 2005-016 (www.math.ntnu.no/conservation/2005/016) .
Camassa R and Holm DD (1993) A new integrable shallow water equation with peaked solitons. *Physical Review Letters* 71: 1661–1664.
Constantin A (2000) Existence of permanent and breaking waves for a shallow water equation: a geometric approach. *Annales de l'Institut Fourier (Grenoble)* 50: 321–362.
Constantin A (2001) On the scattering problem for the Camassa–Holm equation. *Proceedings of the Royal Society of London Series A* 457: 953–970.
Constantin A and Escher J (1998) Wave breaking for nonlinear nonlocal shallow water equations. *Acta Mathematica* 181: 229–243.
Constantin A and Kolev B (2003) Geodesic flow on the diffeomorphism group of the circle. *Commentarii Mathematici Helvetica* 78: 787–804.
Constantin A and Strauss WA (2000) Stability of peakons. *Communications on Pure and Applied Mathematics* 53: 603–610.
Fokas AS and Fuchssteiner B (1981/82) Symplectic structures, their Bäcklund transformations and hereditary symmetries. *Physica D* 4: 47–66.
Gesztesy F and Holden H (2003) *Soliton Equations and their Algebro-Geometric Solutions*. Cambridge: Cambridge University Press.
Johnson RS (1997) *A Modern Introduction to the Mathematical Theory of Water Waves*. Cambridge: Cambridge University Press.
Johnson RS (2002) Camassa–Holm, Korteweg–de Vries and related models for water waves. *Journal of Fluid Mechanics* 455(2002): 63–82.
Kenig CE, Ponce G, and Vega LA (1996) A bilinear estimate with applications to the KdV equation. *Journal of the American Mathematical Society* 9: 573–603.
Kouranbaeva S (1999) The Camassa–Holm equation as a geodesic flow on the diffeomorphism group. *Journal of Mathematical Physics* 40: 857–868.
Lenells J (2002) The scattering approach for the Camassa–Holm equation. *Journal of Nonlinear Mathematical Physics* 9: 389–393.
McKean HP (1979) Integrable systems and algebraic curves. In: *Global Analysis*, Lecture Notes in Mathematics, vol. 755, pp. 83–200. Berlin: Springer.
McKean HP (1998) Breakdown of a shallow water equation. *Asian Journal of Mathematics* 2: 867–874.
Whitham GB (1980) *Linear and Nonlinear Waves*. New York: Wiley.

# BRST Quantization

**M Henneaux**, Université Libre de Bruxelles, Bruxelles, Belgium



## Introduction

The BRST symmetry was originally introduced in the seminal papers by Becchi *et al*. (1976) and Tyutin (1975) for Yang–Mills gauge theories as a tool for controlling the renormalization of the models in a consistent (gauge-independent) way. This symmetry was discovered as a residual symmetry of the gauge-fixed action. It was realized later that, in fact, the BRST construction is quite general, in the sense that it covers arbitrary gauge theories and not just Yang–Mills gauge models. Furthermore, it is intrinsic, in that no gauge choice is actually necessary to define it.

The purpose of this review is to explain the general, intrinsic features of the BRST formalism applicable to "any" gauge theory. The proper setting for discussing these issues is that of homological algebra (Stasheff (1998), and references therein). This article first explains the necessary algebraic material underlying the construction and then illustrates it in the cases of the Hamiltonian BRST formalism and the Lagrangian BRST formalism.

## A Result from Homological Algebra

The main result of homological algebra needed in the BRST construction deals with a differential complex $\mathcal{C}$ with two gradings. The first grading is an $\mathbb{N}$-degree and is called the "resolution degree," or "r-degree." The second grading is a $\mathbb{Z}$-degree and is called the total ghost number. It is denoted by gh. We assume that there are two odd derivations $\delta$ and $s_0$ that have the following properties:

$$r(\delta) = -1, \qquad \mathrm{gh}(\delta) = 1$$
$$r(s_0) = 0, \qquad \mathrm{gh}(s_0) = 1 \qquad [1]$$

and

$$\delta^2 = 0, \quad s_0\delta + \delta s_0 = 0, \quad s_0^2 = -[\delta, s_1] \qquad [2]$$

for some derivation $s_1$ of r-degree 1 and ghost number 1. The bracket $[\cdot,\cdot]$ is the graded commutator – in this specific case, the anticommutator. We also assume that the homology of $\delta$ vanishes at nonzero value of the r-degree, both in the original complex $\mathcal{C}$,

$$H_k(\delta, \mathcal{C}) = 0, \quad k > 0 \qquad [3]$$

(which is equivalent to $\delta a = 0$, $r(a) > 0 \Rightarrow a = \delta b$) and in the space of derivations,

$$[\alpha, \delta] = 0, \quad r(\alpha) \neq 0 \Rightarrow \alpha = [\beta, \delta] \qquad [4]$$

where $\alpha$ and $\beta$ are both derivations in $\mathcal{C}$. The r-degree of a homogeneous linear operator $\alpha$ is defined through $r(\alpha(x)) = r(\alpha) + r(x)$ for any element $x \in \mathcal{C}$ and is negative when $\alpha$ decreases the r-degree.

In $H_0(\delta, \mathcal{C})$, the (odd) derivation $s_0$ defines a differential. The cohomology of $s_0$ modulo $\delta$, denoted $H^k(s_0, H_0(\delta, \mathcal{C}))$, is the cohomology of $s_0$ in $H_0(\delta, \mathcal{C})$. It is explicitly defined through the cocycle condition

$$s_0 a = \delta m \qquad [5]$$

with coboundaries of the form

$$s_0 b + \delta n \qquad [6]$$

The central result underlying the BRST construction is:

**Theorem 1** *Given the above setting, there exists an odd derivation s in $\mathcal{C}$ with the following properties*:

$$s = \delta + s_0 + s_1 + \cdots \qquad [7]$$

$$r(s_k) = k, \qquad \mathrm{gh}(s_k) = 1 \qquad [8]$$

$$s^2 = 0 \qquad [9]$$

*Furthermore, one has*

$$H^k(s, \mathcal{C}) = H^k(s_0, H_0(\delta, \mathcal{C})) \qquad [10]$$

The proof is straightforward (see, e.g., Henneaux and Teitelboim (1992)). In particular, the proof of [10] is a standard spectral sequence argument with a sequence that collapses after the second step. It is interesting to note that, contrary to $s_0$, which is only a differential modulo $\delta$,$s$ is a true differential. The construction of $s$ provides a model for $H^k(s_0, H_0(\delta, \mathcal{C}))$. The differential $s$ is not unique, but this does not affect the subsequent discussion.

In physical applications, the total ghost number is a derived quantity. The primary gradings are the resolution degree and the "filtration degree" called the pure ghost number and denoted pgh. It is an $\mathbb{N}$-degree and one has

$$\mathrm{gh} = \mathrm{pgh} - r \qquad [11]$$

The r-degree is known as the antighost or antifield number, depending on the context (see below). When $r(x) = 0$, one has $\mathrm{gh}(x) = \mathrm{pgh}(x)$. Since the pure ghost number is non-negative, this implies that

$$H^k(s, \mathcal{C}) = 0, \quad k < 0 \qquad [12]$$

## A Geometric Application

### Geometric Setting

Theorem 1 is relevant to the following situation. Consider a surface $\Sigma$ in a manifold $M$, defined by equations

$$f_a = 0 \qquad [13]$$

which may or may not be independent. (We assume for definiteness that the variables in $M$ are bosonic, that is, that $M$ is an ordinary manifold – as opposed to a supermanifold. The graded case can be covered without difficulty by including appropriate sign factors at the relevant places.) Assume that $\Sigma$ is partitioned by orbits generated by vector fields $\boldsymbol{X}_\alpha$ defined everywhere in $M$, tangent to $\Sigma$ and closing on $\Sigma$ in the Lie bracket,

$$[\boldsymbol{X}_\alpha, \boldsymbol{X}_\beta] = C^\gamma{}_{\alpha\beta}\boldsymbol{X}_\gamma + \text{“more”} \qquad [14]$$

where "more" denotes terms that vanish on $\Sigma$. We assume, for simplicity, that the vector fields $\boldsymbol{X}_\alpha$ are linearly independent of $\Sigma$, although this is not necessary. The formalism can be developed in the nonindependent case, but it then requires more variables. We are interested in the quotient space $\Sigma/\mathcal{O}$ of the surface $\Sigma$ by the orbits. To guide the geometrical intuition, we shall assume that this quotient space is a smooth manifold (the fiber of the orbits, etc.), and we shall suggestively adopt notations adapted to this best possible case. The approach, being purely algebraic, is in fact more general. (Accordingly, the notations should be understood with a liberal mind.)

The aim here is to describe the algebra of "observables," that is, the algebra $C^\infty(\Sigma/\mathcal{O})$ of functions on the quotient space $\Sigma/\mathcal{O}$. The terminology "observables" anticipates the physical situation discussed below, where the orbits are the "gauge orbits." In order to describe algebraically the algebra of observables, one observes that this algebra is obtained

through a two-step procedure. First, one restricts the functions from $M$ to $\Sigma$. Second, one imposes the invariance condition along the orbits. To each of these steps corresponds a separate differential.

### Longitudinal Complex

The longitudinal complex is associated with the second step. One can consider on $\Sigma$ an "exterior derivative operator $D$ along the gauge orbits." This operator is defined on functions on $\Sigma$ as

$$Df = \mathbf{X}_\alpha(f)C^\alpha \qquad [15]$$

where the 1-forms $C^\alpha$ dual to the $\mathbf{X}$'s are called ghosts. In the physical context, the form-degree is the pgh described earlier, and so $\mathrm{pgh}(C^\alpha) = 1$. The action of $D$ on the ghosts is given by

$$DC^\alpha = -\tfrac{1}{2}C^\gamma{}_{\alpha\beta}C^\alpha C^\beta \qquad [16]$$

The longitudinal complex $\mathcal{L}_\Sigma$ is the complex of exterior forms along the gauge orbits. In our representation used here, it is given by the space of polynomials in the ghosts $C^\alpha$ with coefficients that are functions on $\Sigma$. The exterior derivative $D$ is defined on this space by extending the formulas [15] and [16] so that it is an odd derivation. One clearly has (on $\Sigma$)

$$D^2 = 0 \qquad [17]$$

The functions on the quotient space $\Sigma/\mathcal{O}$ are just the elements of the zeroth cohomological group $H^0(D, \mathcal{L}_\Sigma)$,

$$H^0(D, \mathcal{L}_\Sigma) = C^\infty(\Sigma/\mathcal{O}) \qquad [18]$$

In general, $H^k(D, \mathcal{L}_\Sigma) \neq 0$.

### Koszul–Tate Differential $\delta$

The Koszul–Tate differential $\delta$ implements the first step in the reduction procedure. More precisely, it provides an algebraic resolution of the algebra $C^\infty(\Sigma)$ of the smooth functions on the surface $\Sigma$.

That algebra can be identified with the quotient algebra

$$C^\infty(\Sigma) = C^\infty(M)/\mathcal{N} \qquad [19]$$

where $\mathcal{N}$ is the ideal of functions that vanish on $\Sigma$. The Koszul–Tate complex $\mathcal{K}$ is defined by adding one new generator for each equation $f_a = 0$ defining $\Sigma$, denoted $t_a^*$ and assigned r-degree 1. In the algebra $C^\infty(M) \otimes \wedge(t_a^*)$ (where $\wedge(t_a^*)$ is the exterior algebra on $t^*$), one defines $\delta$ through

$$\delta f = 0 \quad \forall f \in C^\infty(M), \qquad \delta t_a^* = f_a \qquad [20]$$

and extends it as an odd derivation. It is clear that $r(\delta) = -1$ and that $\delta^2 = 0$. Because the functions on $M$ are annihilated by $\delta$, they are clearly cycles at r-degree zero. Because the left-hand side $f_a$ of the equations $f_a = 0$ are exact (equal to $\delta t_a^*$), the ideal $\mathcal{N}$ coincides with the set of boundaries in degree zero.
Thus,

$$H_0(\delta, \mathcal{K}) = C^\infty(\Sigma) \qquad [21]$$

We see accordingly that $\delta$ successfully enforces the restriction to the surface $\Sigma$ through its homology in degree zero.

However, if the equations $f_a = 0$ are not independent, this is not the end of the story. Indeed, any identity $Z_A^a f_a = 0$ on the functions $f_a$ leads to a nontrivial cycle $Z_A^a t_a^*$ in r-degree 1, $\delta(Z_A^a t_a^*) = 0$. This is undesirable. To cure this drawback, one introduces further generators $t_A^*$ in r-degree 2, one for each identity $Z_A^a f_a = 0$, and defines

$$\delta t_A^* = Z_A^a t_a^*, \quad r(t_A^*) = 2 \qquad [22]$$

in order to "kill" the unwanted cycles $Z_A^a t_a^*$. The Koszul complex $\mathcal{K}$ is thus enlarged to contain these new (even) variables and redefined as

$$\mathcal{K} = C^\infty(M) \otimes \wedge(t_a^*) \otimes S(t_A^*) \qquad [23]$$

where $S(t_A^*)$ is the symmetric algebra in $t_A^*$. The operator $\delta$ is extended to $\mathcal{K}$ as an odd derivation. One has $\delta^2 = 0$ and the property [21] is unaffected by the inclusion of the new generators. Furthermore, by construction,

$$H_1(\delta, \mathcal{K}) = 0 \qquad [24]$$

If there is no "identity on the identities," we shall assume that the process stops. Otherwise, one needs to introduce further generators in r-degree 3 and possibly higher. When all the appropriate variables are included, there is no homology at higher r-degree. Thus,

$$H_k(\delta, \mathcal{K}) = 0, \quad k > 0 \qquad [25]$$

### Combining $\delta$ with $D$

We now turn to the problem of combining the Koszul–Tate complex with the longitudinal complex, so as to implement the full reduction. To that end, we define $\mathcal{C}$ by adding the ghosts to $\mathcal{K}$,

$$\mathcal{C} = \mathcal{K} \otimes \wedge(C_\alpha) = 0 \qquad [26]$$

We then extend the action of the Koszul–Tate differential in the simplest way which preserves all gradings, namely

$$\delta C_\alpha = 0 \qquad [27]$$

It is clear that the homology of $\delta$ in $\mathcal{C}$ is given by

$$H_0(\delta, \mathcal{C}) = \mathcal{L}_\Sigma, \qquad H_k(\delta, \mathcal{C}) = 0 \quad (k > 0) \qquad [28]$$

One can also extend the longitudinal derivative $D$ to the whole complex $\mathcal{C}$ because the vector fields $\boldsymbol{X}_\alpha$ are defined throughout $M$ and so, the definitions [15] and [16] make sense in $\mathcal{C}$. One defines the action of $D$ on the generators $t^*$ by requiring that

$$D\delta + \delta D = 0 \qquad [29]$$

This is easily verified to be possible. However, the (odd) derivation so obtained fails to be a differential in $\mathcal{C}$ when the vector fields $X_\alpha$ do not close off the surface $\Sigma$. In that case, the gauge transformations are not integrable off $\Sigma$; one says that they form an "open algebra." One has then $D^2 = 0$ only on $\Sigma$, or, more precisely,

$$D^2 = -\delta s_1 - s_1 \delta \qquad [30]$$

for some (odd) derivation $s_1$ (that vanishes in the "closed algebra" case). But this situation is precisely the one discussed earlier, with the Koszul–Tate differential being indeed $\delta$, as anticipated by the notation, and the longitudinal differential $D$ playing the role of $s_0$ (the degrees also match). Applying the theorem discussed there, we can conclude:

**Theorem 2** *There exists a differential $s$ in $\mathcal{C}$,*

$$s = \delta + D + s_1 + \cdots, \qquad s^2 = 0 \qquad [31]$$

*such that*

$$H^0(s, \mathcal{C}) = C^\infty(\Sigma/\mathcal{O}) \qquad [32]$$

This is an immediate consequence of Theorem 1 and eqns [18] and [28]. The differential $s$ is known in the physical applications described below as the BRST differential.

## Hamiltonian BRST Construction

As a first application of the above setting, we consider the Hamiltonian description of gauge systems. As already known, gauge systems are characterized in the Hamiltonian description by constraints and, for this reason, are called "constrained Hamiltonian systems." Furthermore, the gauge transformations generate gauge orbits on the constraint surface and the physical observables are the functions on the quotient space of the constraint surface by the gauge orbits.

A further important feature arises in the Hamiltonian formalism: the gauge transformations are canonical transformations that are generated by the first-class constraints. Assuming that all the second-class constraints have been eliminated and that the bracket being used is the Dirac bracket, one sees that there is a vector field $\boldsymbol{X}_\alpha$ for each constraint function $f_a, \alpha \equiv a$. (The functions $f_a$ are thus assumed to be independent since the vector fields $\boldsymbol{X}_\alpha$ are assumed to be so. If not, further variables are needed, but the analysis proceeds along the same ideas.)

This implies, in turn, that there is a pairing between the ghosts $C^a$ associated with the longitudinal exterior derivative and the generators $t_a^*$ of the Koszul–Tate complex. This pairing enables one to extend the bracket structure defined on the phase space to the pairs $(C^a, t_a^*)$ by declaring that these are canonically conjugate. The variables $t_a^*$ are the momenta conjugate to the ghosts, $[t_a^*, C^b] = \delta_a^b$. Accordingly, the complex $\mathcal{C}$ relevant to the Hamiltonian situation,

$$\mathcal{C} = C^\infty(P) \otimes \wedge(C^a) \wedge (t_a^*) \qquad [33]$$

has a phase-space structure (here, $P \equiv M$ is the manifold obtained after eliminating the second-class constraints, equipped with the Dirac bracket). The space $\mathcal{C}$ is known as the "extended phase space." The r-degree is called "antighost number" in the Hamiltonian context.

By the general theorem described in the previous section, one knows that the cohomology at gh = 0 of the BRST differential is isomorphic to the algebra of the observables. Thus, there are two alternative ways to describe this physical algebra, either through reduction, by eliminating the redundant (gauge) variables, or cohomologically in an extended space containing additional variables, the ghosts, and their momenta.

There is an additional interesting feature of the BRST construction in the Hamiltonian case: the BRST transformation is a canonical transformation in the extended phase space, in the sense that

$$sF = [\Omega, F] \qquad [34]$$

for some "BRST generator" $\Omega$ of ghost number 1 $(F, \Omega \in \mathcal{C})$. The nilpotency $s^2$ of the BRST differential is equivalent to

$$[\Omega, \Omega] = 0 \qquad [35]$$

That $s$ is canonically generated implies that the cohomological BRST groups come with a natural bracket structure: the Poisson bracket of the extended phase space passes on to the BRST cohomological groups. In particular, $H^0(s, \mathcal{C})$, equipped with this bracket structure, is isomorphic (as Poisson algebra) to the algebra of physical observables.

## Lagrangian BRST Construction

The analysis of the Lagrangian BRST construction, due to Batalin and Vilkovisky (1981) ("antifield formalism"), proceeds in the same way because the covariant description of the space of observables involves also the same geometric ingredients. The surface $\Sigma$ is now the "stationary surface," that is, the space of solutions to the equations of motion. The space $M$ in which it is embedded is the space of all field histories. The gauge symmetry acts on this space. Furthermore, the gauge vector fields are tangent to $\Sigma$ since a solution is mapped on a solution by a gauge transformation. The integral submanifolds are the gauge orbits. The observables are the functions on the quotient space.

Since the equations of motion follow from an action principle, there are as many equations as there are fields $\varphi^i$. The corresponding generators $t_a^*$ in the Koszul–Tate complex (at degree 1) are called "antifields conjugate to the fields" and are denoted $\varphi_i^*$. The r-degree is known as "antifield" (or also "antighost") number. The gauge symmetry of the action implies Noether identities on the equations of motion. These are, therefore, not independent. According to the above general discussion, there are further generators in the Koszul–Tate complex, at degree 2. More precisely, there are as many new generators in degree 2 as there are Noether identities or independent gauge symmetries. These are called antifields conjugate to the ghosts and denoted $C_\alpha^*$.

In the longitudinal complex, one has the ghosts $C^\alpha$, with as many ghosts as there are gauge symmetries. Thus, the BRST complex is the space

$$\mathcal{C} = C^\infty(M) \otimes \wedge(C^\alpha) \otimes \wedge(\varphi_i^*) \otimes S(C_\alpha^*) \qquad [36]$$

where $M$ is the space of all field histories. There is now a natural pairing between the original field variables $\varphi^i$ and the antifields $\varphi_i^*$, as well as between the ghosts $C^\alpha$ and the antifields $C_\alpha^*$. One thus defines a bracket in which the fields $\varphi^i$ and the ghosts $C^\alpha$ on the one hand, and the antifields $\varphi_i^*$ and $C_\alpha^*$ on the other, are declared to be conjugate. This bracket is denoted by parentheses,

$$(\varphi^i, \varphi_j^*) = \delta_j^i, \qquad (C^\alpha, C_\beta^*) = \delta_\beta^\alpha \qquad [37]$$

However, since the bracket pairs variables with degrees that add up to $-1$, it is in fact an "odd bracket," called the "antibracket."

The BRST differential is again canonically generated, but this time in the antibracket,

$$sF = (S, F), \quad F \in \mathcal{C} \qquad [38]$$

where the generator $S$ is an even function of the fields, the ghosts and the antifields, with $\mathrm{gh} = 0$ (the ghost number is carried by the odd antibracket). The nilpotency $s^2 = 0$ of the BRST differential is equivalent to the crucial "master equation,"

$$(S, S) = 0 \qquad [39]$$

Because the BRST differential is canonically generated, there is a natural bracket in cohomology. This bracket is not the Poisson bracket of observables (at $\mathrm{gh} = 0$) because it changes the ghost number by one unit. One can, however, relate it to the Poisson bracket of observables (Barnich and Henneaux 1996); furthermore, it plays an important role in the study of the consistent deformations of the action.

## Spacetime Locality

In the context of local field theory, one is often interested in a particular class of functions of the field histories, namely the so-called space of local functionals. A local functional is, by definition, the integral of a local $n$-form (where $n$ is the spacetime dimension). A local $n$-form reads, in local coordinates,

$$\omega = f(x)\ \mathrm{d}^n x \qquad [40]$$

where $f(x)$ depends on the fields at $x$ as well as on a finite number of their derivatives. When the ghosts and the antifields are included, the local functions depend on them in the same way.

The previous general cohomological result was derived in the space of all function(al)s, without locality restriction. When changing the space of cochains, one may change the cohomology. For instance, a local functional which is BRST-trivial in the space of all functionals may become nontrivial in the space of local functionals. This indeed happens here because the homology of the Koszul–Tate differentials usually no longer vanishes at strictly positive r-degree in the space of local functionals, where it is related to local conservation laws. As a result, the analysis of the BRST cohomology in the space of local functionals is an interesting and nontrivial problem. In particular, the cohomological groups $H^k(s)$ in the space of local functionals may not vanish at negative ghost numbers.

## BRST Quantization

The quantization of a dynamical system can proceed along different lines. For gauge models, the path-integral approach is most efficiently pursued in the context of the antifield formalism. We shall briefly outline here the general principles underlying the

operator approach, which is based on the Hamiltonian formalism.

In the operator approach, all the variables, including the ghosts and the conjugate momenta, are realized as operators in a space endowed with a nonpositive-definite inner product (because of the ghosts and the gauge modes). Real dynamical variables become formally Hermitian operators. Ignoring anomalies, the BRST generator $\Omega$ becomes an operator that fulfills the conditions

$$\Omega^* = \Omega, \qquad \Omega^2 = 0 \qquad [41]$$

(which allows for nontrivial solutions $\Omega \neq 0$ because the inner product is not positive definite). The second relation is a consequence of the classical Poisson bracket relation $[\Omega,\Omega] = 0$ and the fact that the graded Poisson bracket of two odd objects becomes the anticommutator.

To remove the ghost and gauge redundancy, which has no physical content, one must impose a condition that selects physical states. The appropriate condition is motivated by the general cohomological result connecting the BRST cohomology with the algebra of physical observables. One imposes the condition

$$\Omega|\psi\rangle = 0 \qquad [42]$$

Because of [41], states of the form $\Omega|\chi\rangle$ are solutions of [42], but they have a vanishing inner product with any other physical states, including themselves. They are called null states. The physical states are given by the BRST state cohomology. The physical operators are given by the BRST operator cohomology at $gh = 0$ and induce a well-defined action in the state cohomology. In particular, the Hamiltonian, being gauge invariant in the original theory, is represented by a BRST cohomological class, so that the time evolution maps physical states on physical states.

The whole scheme is (formally) consistent because exact BRST operators have vanishing matrix elements between states annihilated by the BRST operator $\Omega$, while null states $|\phi\rangle$ are such that $\langle\psi|A|\phi\rangle = 0$ whenever $A$ is a BRST-closed operator, $[A, \Omega] = 0$, and $|\psi\rangle$ a physical state. Problems may arise, however, if the classical relations $[\Omega, \Omega] = 0$ and $[H,\Omega] = 0$ are not satisfied in presence of extra terms of order $\hbar$,that is,

$$\Omega^2 \neq 0 \quad \text{or} \quad H\Omega + \Omega H \neq 0 \qquad [43]$$

In such cases, one says that they are anomalies. These are usually fatal to the consistency of the theory.

## Some Applications

The number of applications of the BRST formalism is so large that it would be out of place to try being exhaustive here. Some of its main successes are outlined here, with suggestions for "Further reading."

### Renormalization of Gauge Theories

First, there is the original context of perturbative renormalization and anomalies for gauge theories of the Yang–Mills type. The relevant cohomology here is the BRST cohomology in the space of local functionals involving the fields, the ghosts, and the antifields. The antifields are also known in this context as Zinn-Justin sources for the BRST variations of the fields and ghosts, since Zinn-Justin was the first to introduce them (with that meaning). Many authors have contributed to the full computation of the local BRST cohomology. A review is given in Barnich *et al.* (2000), where extensions to other theories are also indicated.

### String Theory

Modern string theory would be inconceivable without the BRST formalism. This started with the pioneering paper by Kato and Ogawa (1983), where the critical dimension of the bosonic string was derived from the condition that $\Omega^2$ should vanish (quantum mechanically), and where it was shown that the string physical states could be identified with the state BRST cohomology. The reader is referred to excellent monographs on modern string theory (see "Further reading").

### Deformations of Gauge Models

The study of consistent deformations of a given gauge theory (i.e., the problem of introducing consistent couplings) is also efficiently dealt with in the BRST context. References to applications may be found in Henneaux (1998).

*See also:* Anomalies; Batalin–Vilkovisky Quantization; BF Theories; Constrained Systems; Functional Integration in Quantum Physics; Graded Poisson Algebras; Indefinite Metric; Perturbative Renormalization Theory and BRST; Quantum Chromodynamics; Quantum Field Theory: A Brief Introduction; Renormalization: General Theory; String Field Theory; Supermanifolds; Topological Sigma Models.

## Further Reading

Barnich G, Brandt F, and Henneaux M (2000) Local BRST cohomology in gauge theories. *Physics Reports* 338: 439.

Barnich G and Henneaux M (1996) Isomorphisms between the Batalin–Vilkovisky antibracket and the Poisson bracket. *Journal of Mathematical Physics* 37: 5273.

Batalin IA and Vilkovisky GA (1977) Relativistic *S*-matrix of dynamical systems with boson and fermion constraints. *Physics Letters* B69: 309.

Batalin IA and Vilkovisky GA (1981) Gauge algebra and quantization. *Physics Letters* B102: 27.

Becchi C, Rouet A, and Stora R (1976) Renormalization of gauge theories. *Annals of Physics, NY* 98: 287.

Fradkin ES and Vilkovisky GA (1975) Quantization of relativistic systems with constraints. *Physics Letters* B55: 224.

Green MB, Schwarz JH, and Witten E (1987) *Superstring Theory*, vols. 1 and 2. Cambridge: Cambridge University Press.

Henneaux M (1998) Consistent interactions between gauge fields: the cohomological approach. *Contemporary Mathematics* 219: 93.

Henneaux M and Teitelboim C (1992) *Quantization of Gauge Systems*. Princeton: Princeton University Press.

Kato M and Ogawa K (1983) Covariant quantization of strings based on BRS invariance. *Nuclear Physics* B212: 443.

Kugo T and Ojima I (1979) Local covariant operator formalism of nonabelian gauge theories and quark confinement problem. *Progress of Theoretical Physics (Suppl.)* 66: 1.

Polchinski J (1998) *String Theory*,vols. 1 and 2. Cambridge: Cambridge University Press.

Stasheff JD (1998) The (secret?) homological algebra of the Batalin–Vilkovisky approach. *Contemporary Mathematics* 219: 195.

Tyutin IV (1975) Gauge invariance in field theory and statistical physics in the operator formalism. Preprint Lebedev-75-39.

# C

## $C^*$-Algebras and their Classification

**G A Elliott**, University of Toronto, Toronto, Canada



The study of algebras of Hilbert space operators, closed under the adjoint operation and in the weak operator topology, was begun by John von Neumann shortly after the discovery of quantum mechanics, and partly with the aim of understanding the monolithic ideas proposed by Heisenberg and Schrödinger.

Seventy-five years later, the theory of these algebras has become a monolith in its own right (*see* von Neumann Algebras: Introduction, Modular Theory and Classification Theory; von Neumann Algebras: Subfactor Theory), with more internal structure and with more external reference to physics and, as it turns out, to other areas of mathematics than could possibly have been imagined at the outset. (The most striking example of an application to mathematics is perhaps the discovery of the Jones knot polynomial (*see* The Jones Polynomial); note that this has also had repercussions for physics.)

Twenty-five years after the beginning of the theory of von Neumann algebras, as these algebras are now called, Gelfand and Naimark noticed that a second class of algebras of operators on a Hilbert space, closed under the adjoint operation, was worthy of study, namely those closed in the norm topology. Gelfand and Naimark made two important discoveries concerning this class of operator algebras, now called $C^*$-algebras.

First, Gelfand and Naimark showed that, in the commutative case, at least when the $C^*$-algebra is considered only up to isomorphism – with its identity as a concrete algebra of operators suppressed – the information contained in a $C^*$-algebra is purely topological. More precisely, Gelfand and Naimark showed that the category of unital commutative $C^*$-algebras, with unit-preserving algebra homomorphisms (these necessarily preserve the adjoint operation), is equivalent in a contravariant way (i.e., with reversal of arrows) to the category of compact Hausdorff spaces, with continuous maps. The compact space associated with a unital commutative $C^*$-algebra under the Gelfand–Naimark correspondence may be viewed as the space of maximal proper ideals, with a natural topology (the hull-kernel, or Jacobson, topology), and is called the spectrum. This space may also be viewed as the set of (unital, linear, multiplicative) maps from the algebra into the complex numbers, in which case the topology is that of pointwise convergence.

Second, using this result, Gelfand and Naimark proved that arbitrary $C^*$-algebras could be axiomatized in a simple way abstractly, as $^*$-algebras – that is, as algebras over the complex numbers with a conjugate linear anti-automorphism of order 2 – with certain special properties. It is now known that the only property that needs to be assumed is the existence of a (necessarily unique) Banach space norm related to the $^*$-algebra structure by means of the so-called $C^*$-algebra identity:

$$\|x^*x\| = \|x^*\|\,\|x\| \qquad [1]$$

This is clearly related to – and in fact implies – the normed algebra inequality

$$\|xy\| \leq \|x\|\,\|y\| \qquad [2]$$

One reason that the Gelfand–Naimark axiomatization of $C^*$-algebras is important is that it underlines how natural it is to consider a $C^*$-algebra abstractly, i.e., independently of any particular representation. Indeed, while one of the fundamental phenomena of von Neumann algebra theory (discovered by Murray and von Neumann) is that, essentially – in rather a strong sense – there is only one way to represent a given von Neumann algebra on a Hilbert space (and there is even a canonical way, called the standard representation!), it is an equally fundamental phenomenon of $C^*$-algebra theory that, except in extremely special cases, this is no longer true.

For instance, although the $C^*$-algebra of compact operators on a given Hilbert space has, up to unitary equivalence, only a single irreducible representation – this is what underlies the fact, proved by von Neumann, referred to as the uniqueness of the

Heisenberg commutation relations for a quantum-mechanical system with finitely many degrees of freedom – as soon as one considers a physical system with infinitely many degrees of freedom, one finds that the naturally associated $C^*$-algebra has infinitely many – indeed, uncountably many – unitary equivalence classes of irreducible representations, and it is impossible to parametrize these in any reasonable way.

This striking dichotomy presents itself also in other contexts, more elementary perhaps than the physics of infinitely many degrees of freedom. Consider the dynamical system consisting of a circle and a fixed rotation acting on it. If the rotation is of finite order – i.e., if the angle is a rational multiple of $2\pi$ – then the naturally associated $C^*$-algebra is relatively easy to study. In the case of angle zero, it is the unital commutative $C^*$-algebra with Gelfand–Naimark spectrum the torus. In the general case of a rational angle, the space of unitary equivalence classes of irreducible representations is still naturally parametrized by the torus. (And this is the same as the space of primitive ideals – the kernels of the irreducible representations – with the Jacobson topology.)

In the irrational case – the case of a rotation by an irrational multiple of $2\pi$ (still elementary from a geometrical point of view; note that the calendar is based on such a system!) – the irreducible representations are no longer parametrized up to unitary equivalence by the torus – and the space of primitive ideals consists of a single point – the $C^*$-algebra is simple. (But it is decidedly not simple to study!)

This fundamental dichotomy in the classification of $C^*$-algebras – conjectured by Gaarding and Wightman in the quantum-mechanical setting and by Mackey in the geometrical one – was established by Glimm. Glimm proved (in the setting of separability; most of his results were generalized later to the nonseparable case) that a large number of *a priori* different ways that a $C^*$-algebra could behave well were in fact one and the same behavior: either all present for a given $C^*$-algebra, or all catastrophically absent!

Some of the properties considered by Glimm, and shown to be equivalent (for a separable $C^*$-algebra) were as follows. First of all, every representation of the $C^*$-algebra on a Hilbert space should be of type I, i.e., should generate a von Neumann algebra of type I. (A von Neumann algebra was said by Murray and von Neumann to be of type I if it contained a minimal projection of central support one, i.e., a projection not contained in a proper direct summand and minimal with this property.) Second, in every irreducible representation (not necessarily injective) on a Hilbert space, the image of the $C^*$-algebra should contain the compact operators. Third, any two irreducible representations with the same kernel should be unitarily equivalent. Fourth, it should be possible to parametrize the unitary equivalence classes of irreducible representations by a real number in a natural way (respecting the natural Borel structure introduced by Mackey).

The first of the equivalent properties listed above, that all representations of a $C^*$-algebra should be of type I, suggested a name for the property – that the $C^*$-algebra itself should be of type I. This property of a $C^*$-algebra, identified by Glimm – or, rather, its opposite, which as mentioned above is much more common (just as irrational numbers are more common than rationals, or systems with infinitely many degrees of freedom are, at least in theory, much more common than those with finitely many degrees of freedom) – is a fundamental unifying principle of nature.

Besides commutative $C^*$-algebras – as mentioned above, just another way of looking at topological spaces (compact Hausdorff spaces, that is) – and besides the $C^*$-algebra associated to a rotation or to a physical system with infinitely many degrees of freedom, what are some of the naturally occurring examples of $C^*$-algebras – of type I or not!

First, let us take a closer look at what arises from a system with infinitely many degrees of freedom – in the fermion case. As shown by Jordan and Wigner, one obtains what, as a $C^*$-algebra, is very easy to describe, namely, just the infinite tensor product in the category of unital $C^*$-algebras of copies of the algebra of $2 \times 2$ matrices over the complex numbers. As it happens, in work earlier than that referred to above, Glimm had considered such infinite tensor product $C^*$-algebras, also allowing the components to be matrix algebras of order different from two. This raised a problem of classification – for those $C^*$-algebras, all of which were simple and not of type I. (The only simple unital $C^*$-algebra of type I is a single matrix algebra, or a finite tensor product of matrix algebras!)

In a pioneering classification paper (the first paper on the classification of $C^*$-algebras being perhaps that of Gelfand and Naimark, in which the commutative case was described), Glimm obtained the classification of infinite tensor products of matrix algebras, showing that it was a direct extension of the classification of finite tensor products, i.e., just of the matrix algebras themselves. As described later by Dixmier, Glimm's classification was as follows. Given a sequence $n_1, n_2, \ldots$ of natural numbers (equal to one or more), form the infinite product in a natural way – just by keeping track of the total number of times each prime number appears in the

finite products $n_1 \ldots n_k$ (a multiplicity which may be either finite or infinite). Call such a formal infinite product a generalized integer – or, perhaps, a supernatural number! Two (countably) infinite tensor products of matrix algebras are isomorphic (just as in the finite tensor product case) if and only if the corresponding supernatural numbers are equal.

In formulating Glimm's classification of infinite tensor products of matrix algebras in this way, Dixmier pointed out that each supernatural number determines a subgroup of the rational numbers (those with denominator dividing the supernatural number) and that every subgroup of the rational numbers containing the integers arises in this way. He then gave an alternative derivation of Glimm's theorem by recovering this subgroup of the rational numbers as a natural invariant of the algebra, namely, as the subgroup generated by the values on projections of the unique normalized trace. (By a trace is meant here a unitarily invariant positive linear functional.) This could even be interpreted as an alternative statement of Glimm's theorem.

Soon afterwards, Bratteli considered an extension of Glimm's class of $C^*$-algebras, namely, the inductive limits of arbitrary sequences of finite-dimensional $C^*$-algebras, and gave a classification of these algebras in terms of the embedding multiplicity data in the sequences. This was exactly analogous to the original classification of Glimm, but now vastly more complex, with the multiplicity data of the sequence encoded in what is now called a Bratteli diagram. (Note that a finite-dimensional $C^*$-algebra is just a direct sum of matrix algebras over the complex numbers.) Bratteli diagrams have proved to be very important, and in particular have been shown by Putnam and others to be useful for the study of minimal homeomorphisms of the Cantor set.

Bratteli's extension of Glimm's tensor product classification was followed by a corresponding extension by the present author of Dixmier's approach to Glimm's result. It was no longer possible to express the appropriate data in terms of traces (even in the case of a unique normalized trace). Instead, the present author recalled the concept of equivalence of projections introduced by Murray and von Neumann forty years earlier, together with the fact, proved by Murray and von Neumann, that equivalence is compatible with addition of orthogonal projections. (Two projections in a $^*$-algebra are equivalent if they are equal to $x^*x$ and $xx^*$ for some element $x$.) The resulting elementary invariant – the set of equivalence classes of projections with the operation of addition whenever defined (whenever the equivalence classes to be added have orthogonal representatives) – one might refer to this as a local abelian semigroup – which was used by Murray and von Neumann to divide von Neumann algebras into what they called types I, II, and III – was shown by the author to determine Bratteli's algebras up to isomorphism.

Bratteli called his algebras approximately finite-dimensional $C^*$-algebras, or AF algebras. The author referred to his invariant simply as the range of the (abstract) dimension, and pointed out that this structure determined an enveloping ordered abelian group, which he called the dimension group. It was soon noticed that the dimension group was related to the $K$-group introduced by Grothendieck in algebraic geometry (*see* $K$-Theory), and by Atiyah and Hirzebruch (*see* $K$-Theory) in topology.

Grothendieck's $K$-group was defined for an arbitrary ring with unit, and Atiyah and Hirzebruch in effect considered the special case of the ring of continuous functions on a compact Hausdorff space – in other words, a commutative $C^*$-algebra – in the process showing that the deep phenomenon of Bott periodicity could be expressed in terms of this invariant. The invariant itself (see below) is essentially the same as that of Murray and von Neumann. In the special case that the ring is an AF algebra, the $K$-group coincides with the dimension group. (The $K$-group has a natural ordered, or pre-ordered, structure, although this was often suppressed.)

Let us consider the definition of the $K$-group of a not necessarily unital $C^*$-algebra; it is in this setting that the statement of Bott periodicity attains its simplest form.

First, in the unital case, one constructs the abelian local semigroup (addition just partially defined) of Murray–von Neumann equivalence classes of projections, as described above in the case of an AF algebra. Let us call this the dimension range. As stated above, for AF algebras this is all that needs to be done – the enveloping group of the dimension range is already the $K$-group. In the general case, one must repeat the construction for the algebra of $2 \times 2$ matrices over the given algebra, with the given algebra considered as embedded as the upper left-hand corner of the matrix algebra. The dimension range of the given algebra then maps naturally into (but not necessarily onto) the dimension range of the matrix algebra. One should then repeat this construction, doubling the order of the matrix algebra at every stage (or, alternatively, increasing it just by one). The enveloping group of the (algebraic) inductive limit of this sequence of local semigroups is then the $K$-group of the given algebra. (Alternatively, one may just consider immediately the $^*$-algebra of all infinite matrices over the given

$C^*$-algebra with only finitely many nonzero entries, and form the dimension range of this $^*$-algebra – and the enveloping group of this abelian local semigroup, now in fact a semigroup.)

In the case of a nonunital $C^*$-algebra, one adjoins a unit (as may be done, for instance, by representing the $C^*$-algebra faithfully on a Hilbert space, and showing that the $C^*$-algebra obtained by adjoining the identity operator is independent of the representation – actually, one need only check that the $^*$-algebra structure is unique, as the $C^*$-algebra norm on a $C^*$-algebra is always determined by the $^*$-algebra structure). The $K$-group of the resulting unital $C^*$-algebra then maps naturally into the $K$-group of the natural one-dimensional quotient, and the kernel of this map is, for reasons that will become clearer later, defined to be the $K$-group of the nonunital algebra.

Atiyah and Hirzebruch in fact referred to the $K$-group of the $C^*$-algebra as $K_0$ – the reason being that there is another very natural group to consider, namely, the $K$-group of the suspension of the $C^*$-algebra. (The suspension, $SA$, of a $C^*$-algebra $A$ is defined as the $C^*$-algebra of all continuous functions from the real line $\mathbb{R}$ into $A$ which converge to zero at $\pm\infty$, with the pointwise $^*$-algebra operations and the supremum norm. It may also be defined as the (unique) $C^*$-algebra tensor product $A \otimes C_0(\mathbb{R})$, where $C_0(\mathbb{R})$ denotes the suspension of the $C^*$-algebra $\mathbb{C}$ of complex numbers.) Denoting the $K_0$-group of the suspension of a given $C^*$-algebra by $K_1$, one might expect this process to continue, but in fact it is periodic $(K_0, K_1, K_0, K_1, \ldots)$. Bott periodicity states that there is a natural isomorphism of $K_2$ with $K_0$. ($C^*$-algebras can also be defined with the field of real numbers as scalars, and in this case the period of Bott periodicity is eight.)

Another way of stating Bott periodicity, or, more precisely, of embedding it into the $K$-theory of $C^*$-algebras, is as follows. Given a short exact sequence of $C^*$-algebras,

$$0 \to J \to A \to A/J \to 0 \qquad [3]$$

i.e., given a $C^*$-algebra $A$ and a closed two-sided ideal $J$ (the quotient $^*$-algebra is then a $C^*$-algebra with the quotient norm) – $A$ is sometimes referred to as an extension of $J$ by $A/J$ – consider the natural short (not necessarily exact) sequences

$$K_0(J) \to K_0(A) \to K_0(A/J) \qquad [4]$$

and

$$K_1(J) \to K_1(A) \to K_1(A/J) \qquad [5]$$

($K_0$ and $K_1$ are functors!). There exist natural connecting maps $K_1(A/J) \to K_0(J)$ and $K_0(A/J) \to K_1(J)$ – the first referred to as the index map, and the second (sometimes referred to as the odd-order index map) obtained from this immediately from Bott periodicity (as stated above) – such that the periodic six-term sequence

$$\begin{array}{ccccc} K_0(J) & \to & K_0(A) & \to & K_0(A/J) \\ \uparrow & & & & \downarrow \\ K_1(A/J) & \leftarrow & K_1(A) & \leftarrow & K_1(J) \end{array}$$

is exact. (The periodicity stated above can also be recovered from this.)

Given that the functor $K_0$ classifies AF algebras, one might expect the functor $K_1$ to be useful for classification purposes also. In fact, this is the case. (Indeed, as shown by Brown, the $K_1$-functor is already important for the theory of AF algebras – in spite of, or even because of (!), the fact that the $K_1$-group of an AF algebra is zero.) Using the six-term exact sequence of Bott periodicity described above, corresponding to an extension of $C^*$-algebras, together with results of the present author, Brown showed that any extension of one AF algebra by another is again an AF algebra.

A rather large class of simple unital $C^*$-algebras has by now been classified by means of the invariants $K_0$ and $K_1$ – together with the class of the unit in $K_0$, and the order (or pre-order) structure on $K_0$ – and also taking into account the compact convex set of tracial states on the $C^*$-algebra (a positive linear functional on a $C^*$-algebra is called a trace if it has the same value on $x^*x$ and $xx^*$ for every element $x$, and a tracial state if it is a state, that is, has norm 1, or has value 1 on the unit in the case the algebra has a unit). In addition to the set of tracial states, together with its natural topology and convex structure, one should also keep track of the natural pairing between traces and $K_0$ (any trace on a unital $C^*$-algebra has the same value on two equivalent projections – equal to $x^*x$ and $xx^*$ for some element $x$ – and hence gives rise to an additive real-valued functional on $K_0$).

In terms of these invariants (which might, broadly speaking, be called $K$-theoretical), it has been possible to classify the simple unital $C^*$-algebras (not of type I) arising as inductive limits (i.e., as the completions of increasing unions) of sequences of finite direct sums of matrix algebras over separable commutative $C^*$-algebras, these assumed to have spectra of dimension at most three, on the one hand (work of the present author together with Guihua Gong and Liangqing Li, a culmination of earlier work of these authors together with a number of others), and, on the other hand, it has been possible (work of Kirchberg and Phillips, also based on earlier work by a number of authors) to classify the

$C^*$-algebra tensor products (in a natural sense) of these $C^*$-algebras with what is called the Cuntz $C^*$-algebra $O_\infty$ (see below). In the first of these two cases, the compact convex set of tracial states – always a Choquet simplex – is an arbitrary (metrizable) such space.

In the second case, this space is empty (as it is for $O_\infty$ in particular). In both cases, $K_0$ and $K_1$ are arbitrary countable abelian groups, with the proviso that $K_0$ is not the sum of a torsion group and a cyclic group. In the first case, the order structure on $K_0$, the class of the unit element, and the pairing of $K_0$ with the space of traces have certain special properties; as it turns out, these can be expressed in a simple way. (The class of the unit need only be positive and nonzero.) In the second case, the order structure on $K_0$ is degenerate – every element is positive – and the class of the unit can be arbitrary (including zero!).

Let us just note that the Cuntz $C^*$-algebra $O_\infty$ is the unital $C^*$-algebra generated by an infinite sequence $s_1, s_2, \ldots$ of isometries with orthogonal ranges (in other words, elements $s_i$ such that $s_i^* s_i$ is the unit and $s_j^* s_i = 0$ if $j \neq i$). One need not require the $C^*$-algebra to have the universal property with respect to these generators and relations as it is in fact unique (up to an isomorphism preserving these generators). In particular, this $C^*$-algebra is simple. (If one considers a finite sequence of isometries with orthogonal ranges, and assumes in addition that the sum of these is the unit, one also obtains a simple $C^*$-algebra, the Cuntz $C^*$-algebra $O_n$, $n = 2, 3, \ldots$). The $K_0$-group and $K_1$-group of $O_\infty$ are, respectively, $\mathbb{Z}$ and 0. (The $K_0$-group and $K_1$-groups of $O_n$ for $n = 2, 3, \ldots$ are, respectively, $\mathbb{Z}/(n-1)\mathbb{Z}$ and 0.)

Both classes of $C^*$-algebras considered in the classification result stated above, although described in rather a concrete way (in terms of inductive limits and tensor products), can also be characterized axiomatically, in a way that makes it clear that they are, in fact, much more general than they seem. (These axiomatizations are due to Lin and to Kirchberg and Phillips. Typically, the abstract axioms are easier to establish in a given case than the inductive limit form described above.)

In view of this, and the fact that one of the axioms is a notion of amenability (the analogous property for $C^*$-algebras of a notion that has also been considered for von Neumann algebras) and since amenable von Neumann algebras (on a separable Hilbert space) have been classified completely (in remarkable work of Connes, together with many others, starting with Murray and von Neumann – and, one must also mention, ending with Haagerup, who settled a particularly stubborn case), it is natural to ask whether the $K$-theoretical invariants described above might be sufficient to classify all amenable separable $C^*$-algebras, say, those which are simple and unital.

The work of Villadsen has shown that additional invariants must in fact be considered, if one is to deal with arbitrary amenable simple $C^*$-algebras, and this has been confirmed in subsequent work of Rørdam and of Toms. (Villadsen's examples were obtained by removing the condition of low dimension on the spectra of the commutative $C^*$-algebras appearing in the inductive limit decomposition considered above.) The very nature of these authors' work, however, has been to introduce additional invariants, all of which it seems natural to consider as, broadly speaking, $K$-theoretical. (And all of which, as it happens, are already familiar.)

The question of the classifiability, in terms of simple invariants ($K$-theoretical in nature, at least in the broad sense, and including the spectrum which is indispensable in the nonsimple case), of all (separable) amenable $C^*$-algebras would therefore still appear to be on the agenda.

Already, in any case, just like the analogous question for von Neumann algebras (now settled), this question would appear to have had a noticeable influence on the development of the subject – not least in underlining the importance of $K$-theoretical methods, which have proved to be pertinent both in connection with the index theory of differential operators on geometrical structures – from foliations to fractals – and in connection with questions in physics, related to quantum statistical mechanics (*see* e.g., Quantum Hall Effect), to quantum field theory (e.g., the standard model), and even to string theory and $M$-theory.

*See also:* Axiomatic Quantum Field Theory; Bosons and Fermions in External Fields; The Jones Polynomial; $K$-Theory; Positive Maps on $C^*$-Algebras; Quantum Hall Effect; von Neumann Algebras: Introduction, Modular Theory, and Classification Theory; von Neumann Algebras: Subfactor Theory.

## Further Reading

Davidson KR (1996) *$C^*$-Algebras by Example*. Fields Institute Monographs, 6. Providence, RI: American Mathematical Society.

Dixmier J (1969) *Les $C^*$-Algèbres et leurs Répresentations*, 2nd edn. Paris: Gauthier–Villars.

Elliott GA (1995) The classification problem for amenable $C^*$-algebras. In: Chatterji SD (ed.) *Proceedings of the International Congress of Mathematicians*, vols. 1, 2, pp. 922–932. (Zürich, 1994). Basel: Birkhäuser.

Evans DE and Kawahigashi Y (1998) *Quantum Symmetries on Operator Algebras*. Oxford: Oxford University Press.
Fillmore PA (1996) *A User's Guide to Operator Algebras*. New York: Wiley.
Kadison RV and Ringrose J (1983–92) *Fundamentals of the Theory of Operator Algebras* (4 volumes). New York: Academic Press.
Lin H (2001) *An Introduction to the Classification of Amenable C*-Algebras*. Singapore: World Scientific.
Pedersen GK (1979) *C*-Algebras and their Automorphism Groups*, London Math. Soc. Monographs. London: Academic Press.
Rørdam M (2002) *Classification of Nuclear, Simple C*-Algebras*, Encyclopaedia of Mathematical Sciences, vol. 126, pp. 1–145. Berlin: Springer.
Sakai S (1971) *C*-Algebras and W*-Algebras*. Berlin: Springer.

# Calibrated Geometry and Special Lagrangian Submanifolds

**D D Joyce**, University of Oxford, Oxford, UK



## Calibrated Geometry

"Calibrated geometry," introduced by Harvey and Lawson (1982), is the study of special classes of "minimal submanifolds" $N$ of a Riemannian manifold $(M,g)$, defined using a closed form $\varphi$ on $M$ called a calibration. For example, if $(M,J,g)$ is a Kähler manifold with Kähler form $\omega$, then complex $k$-submanifolds of $M$ are calibrated with respect to $\varphi=\omega^k/k!$. Another important class of calibrated submanifolds are special Lagrangian submanifolds in Calabi–Yau manifolds, which is the focus of the section "Special Lagrangian geometry."

### Calibrations and Calibrated Submanifolds

We begin by defining "calibrations" and "calibrated submanifolds."

**Definition 1** Let $(M,g)$ be a Riemannian manifold. An "oriented tangent $k$-plane" $V$ on $M$ is a vector subspace $V$ of some tangent space $T_xM$ to $M$ with $\dim V=k$, equipped with an orientation. If $V$ is an oriented tangent $k$-plane on $M$ then $g|_V$ is a Euclidean metric on $V$; so, combining $g|_V$ with the orientation on $V$ gives a natural volume form $\mathrm{vol}_V$ on $V$, which is a $k$-form on $V$.

Now let $\varphi$ be a closed $k$-form on $M$. $\varphi$ is said to be a calibration on $M$, if for every oriented $k$-plane $V$ on $M$, $\varphi|_V \le \mathrm{vol}_V$. Here, $\varphi|_V=\alpha\cdot\mathrm{vol}_V$ for some $\alpha\in\mathbb{R}$, and $\varphi|_V\le\mathrm{vol}_V$ if $\alpha\le 1$. Let $N$ be an oriented submanifold of $M$ with dimension $k$. Then each tangent space $T_xN$ for $x\in N$ is an oriented tangent $k$-plane. We say that $N$ is a calibrated submanifold if $\varphi|_{T_xN}=\mathrm{vol}_{T_xN}$ for all $x\in N$.

It is easy to show that calibrated submanifolds are automatically "minimal submanifolds." We prove this in the compact case, but noncompact calibrated submanifolds are locally volume-minimizing as well.

**Proposition 2** *Let $(M,g)$ be a Riemannian manifold, $\varphi$ a calibration on $M$, and $N$ a compact $\varphi$-submanifold in $M$. Then $N$ is volume-minimizing in its homology class.*

**Proof** Let $\dim N=k$, and let $[N]\in H_k(M,\mathbb{R})$ and $[\varphi]\in H^k(M,\mathbb{R})$ be the homology and cohomology classes of $N$ and $\varphi$. Then

$$[\varphi]\cdot[N]=\int_{x\in N}\varphi|_{T_xN}=\int_{x\in N}\mathrm{vol}_{T_xN}=\mathrm{Vol}(N)$$

since $\varphi|_{T_xN}=\mathrm{vol}_{T_xN}$ for each $x\in N$, as $N$ is a calibrated submanifold. If $N'$ is any other compact $k$-submanifold of $M$ with $[N']=[N]$ in $H_k(M,\mathbb{R})$, then

$$[\varphi]\cdot[N]=[\varphi]\cdot[N']=\int_{x\in N'}\varphi|_{T_xN'}\le\int_{x\in N'}\mathrm{vol}_{T_xN'}=\mathrm{Vol}(N')$$

since $\varphi|_{T_xN'}\le\mathrm{vol}_{T_xN'}$ because $\varphi$ is a calibration. The last two equations give $\mathrm{Vol}(N)\le\mathrm{Vol}(N')$. Thus, $N$ is volume-minimizing in its homology class. □

Now let $(M,g)$ be a Riemannian manifold with a calibration $\varphi$, and let $\iota:N\to M$ be an immersed submanifold. Whether $N$ is a $\varphi$-submanifold depends upon the tangent spaces of $N$. That is, it depends on $\iota$ and its first derivative. So, for $N$ to be calibrated with respect to $\varphi$ is a first-order partial differential equation on $\iota$. But if $N$ is calibrated then $N$ is minimal, and for $N$ to be minimal is a second-order partial differential equation on $\iota$.

One moral is that the calibrated equations, being first order, are often easier to solve than the minimal submanifold equations, which are second order. So calibrated geometry is a fertile source of examples of minimal submanifolds.

### Calibrated Submanifolds and Special Holonomy

A calibration $\varphi$ on $(M,g)$ is only interesting if there exist plenty of $\varphi$-submanifolds $N$ in $M$, locally or globally. Since $\varphi|_{T_xN}=\mathrm{vol}_{T_xN}$ for each $x\in N$, $\varphi$-submanifolds will be abundant only if the family $\mathcal{F}_\varphi$ of calibrated tangent $k$-planes $V$ with $\varphi|_V=\mathrm{vol}_V$

is "reasonably large" – say, if $\mathcal{F}_\varphi$ has small codimension in the family of all tangent $k$-planes $V$ on $M$. A maximally boring example is the $k$-form $\varphi = 0$, which is a calibration but has no calibrated tangent $k$-planes, so no $\varphi$-submanifolds.

Thus, most calibrations $\varphi$ will have few or no $\varphi$-submanifolds, and only special calibrations $\varphi$ with $\mathcal{F}_\varphi$ large will have interesting calibrated geometries. Now the field of Riemannian holonomy groups is a natural companion for calibrated geometry, because it gives a simple way to generate interesting calibrations $\varphi$ which automatically have $\mathcal{F}_\varphi$ large.

Let $G \subset \mathrm{O}(n)$ be a possible holonomy group of a Riemannian metric. In particular, we can take $G$ to be one of the holonomy groups $\mathrm{U}(m)$, $\mathrm{SU}(m)$, $\mathrm{Sp}(m)$, $G_2$, or Spin(7) from Berger's classification. Then $G$ acts on the $k$-forms $\Lambda^k(\mathbb{R}^n)^*$ on $\mathbb{R}^n$, so we can look for $G$-invariant $k$-forms on $\mathbb{R}^n$. Suppose $\varphi_0$ is a nonzero, $G$-invariant $k$-form on $\mathbb{R}^n$.

By rescaling $\varphi_0$ we can be arrange that for each oriented $k$-plane $U \subset \mathbb{R}^n$, we have $\varphi_0|_U \le \mathrm{vol}_U$, and that $\varphi_0|_U = \mathrm{vol}_U$ for at least one such $U$. Let $H$ be the stabilizer subgroup of this $U$ in $G$. Then $\varphi_0|_{\gamma\cdot U} = \mathrm{vol}_{\gamma\cdot U}$ by $G$-invariance, so $\gamma \cdot U$ is a calibrated $k$-plane for all $\gamma \in G$. Thus, the family $\mathcal{F}_0$ of $\varphi_0$-calibrated $k$-planes in $\mathbb{R}^n$ contains $G/H$, so it is "reasonably large," and it is likely that the calibrated submanifolds will have an interesting geometry.

Now let $M$ be a manifold of dimension $n$, and $g$ a metric on $M$ with Levi-Civita connection $\nabla$ and holonomy group $G$. Then there is a $k$-form $\varphi$ on $M$ with $\nabla\varphi = 0$, corresponding to $\varphi_0$. Hence $\mathrm{d}\varphi = 0$, and $\varphi$ is closed. Also, the condition $\varphi_0|_U \le \mathrm{vol}_U$ for all oriented $k$-planes $U$ in $\mathbb{R}^n$ implies that $\varphi|_V \le \mathrm{vol}_V$ for all oriented tangent $k$-planes $V$ in $M$. Thus, $\varphi$ is a calibration on $M$. The family $\mathcal{F}_\varphi$ of calibrated tangent $k$-planes on $M$ fibers over $M$ with fiber $\mathcal{F}_0$; so, it is "reasonably large."

This gives a general method for finding interesting calibrations on manifolds with reduced holonomy. Here are the most significant examples.

- Let $G = \mathrm{U}(m) \subset \mathrm{O}(2m)$. Then $G$ preserves a 2-form $\omega_0$ on $\mathbb{R}^{2m}$. If $g$ is a metric on $M$ with holonomy $\mathrm{U}(m)$, then $g$ is Kähler with complex structure $J$, and the 2-form $\omega$ on $M$ associated to $\omega_0$ is the Kähler form of $g$.

  One can show that $\omega$ is a calibration on $(M,g)$, and the calibrated submanifolds are exactly the "holomorphic curves" in $(M,J)$. More generally, $\omega^k/k!$ is a calibration on $M$ for $1 \le k \le m$, and the corresponding calibrated submanifolds are the complex $k$-dimensional submanifolds of $(M,J)$.
- Let $G = \mathrm{SU}(m) \subset \mathrm{O}(2m)$. Then $G$ preserves a complex volume form $\Omega_0 = \mathrm{d}z_1 \wedge \cdots \wedge \mathrm{d}z_m$ on $\mathbb{C}^m$. Thus, a Calabi–Yau $m$-fold $(M,g)$ with $\mathrm{Hol}(g) = \mathrm{SU}(m)$ has a holomorphic volume form $\Omega$. The real part $\mathrm{Re}\,\Omega$ is a calibration on $M$, and the corresponding calibrated submanifolds are called special Lagrangian submanifolds.
- The group $G_2 \subset \mathrm{O}(7)$ preserves a 3-form $\varphi_0$ and a 4-form $*\varphi_0$ on $\mathbb{R}^7$. Thus, a Riemannian 7-manifold $(M,g)$ with holonomy $G_2$ comes with a 3-form $\varphi$ and 4-form $*\varphi$, which are both calibrations. The corresponding calibrated submanifolds are called associative 3-folds and coassociative 4-folds.
- The group $\mathrm{Spin}(7) \subset \mathrm{O}(8)$ preserves a 4-form $\Omega_0$ on $\mathbb{R}^8$. Thus a Riemannian 8-manifold $(M,g)$ with holonomy Spin(7) has a 4-form $\Omega$, which is a calibration. The $\Omega$-submanifolds are called Cayley 4-folds.

It is an important general principle that to each calibration $\varphi$ on an $n$-manifold $(M,g)$ with special holonomy constructed in this way, there corresponds a constant calibration $\varphi_0$ on $\mathbb{R}^n$. Locally, $\varphi$-submanifolds in $M$ resemble the $\varphi_0$-submanifolds in $\mathbb{R}^n$, and have many of the same properties. Thus, to understand the calibrated submanifolds in a manifold with special holonomy, it is often a good idea to start by studying the corresponding calibrated submanifolds of $\mathbb{R}^n$.

In particular, singularities of $\varphi$-submanifolds in $M$ will be locally modeled on singularities of $\varphi_0$-submanifolds in $\mathbb{R}^n$. (In the sense of geometric measure theory, the tangent cone at a singular point of a $\varphi$-submanifold in $M$ is a conical $\varphi_0$-submanifold in $\mathbb{R}^n$.) So by studying singular $\varphi_0$-submanifolds in $\mathbb{R}^n$, we may understand the singular behavior of $\varphi$-submanifolds in $M$.

## Special Lagrangian Geometry

We now focus on one class of calibrated submanifolds, special Lagrangian submanifolds in Calabi–Yau manifolds. Calabi–Yau 3-folds are used to make the spacetime vacuum in string theory, and special Lagrangian 3-folds are the classical versions of A-branes, or supersymmetric 3-cycles, in Calabi–Yau 3-folds. Special Lagrangian geometry aroused great interest amongst string theorists because of its rôle in the SYZ conjecture, providing a geometric basis for "mirror symmetry" of Calabi–Yau 3-folds.

### Calabi–Yau Manifolds

Here is our definition of Calabi–Yau manifold. Readers are warned that there are several different definitions of Calabi–Yau manifolds in use in the literature. Ours is unusual in regarding $\Omega$ as part of the given structure.

**Definition 3** Let $m \geq 2$. A Calabi–Yau $m$-fold is a quadruple $(M, J, g, \Omega)$ such that $(M, J)$ is a compact $m$-dimensional complex manifold, $g$ a Kähler metric on $(M, J)$ with Kähler form $\omega$, and $\Omega$ a holomorphic $(m, 0)$-form on $M$ called the holomorphic volume form, which satisfies

$$\omega^m/m! = (-1)^{m(m-1)/2}(\mathrm{i}/2)^m \Omega \wedge \bar{\Omega} \qquad [1]$$

The constant factor in [1] is chosen to make $\mathrm{Re}\,\Omega$ a calibration. It follows from [1] that $g$ is Ricci-flat, $\Omega$ is constant under the Levi-Civita connection, and the holonomy group of $g$ has $\mathrm{Hol}(g) \subseteq \mathrm{SU}(m)$.

Let $(M, J)$ be a compact, complex manifold, and $g$ a Kähler metric on $M$, with Ricci curvature $R_{ab}$. Define the Ricci form $\rho$ of $g$ by $\rho_{ac} = J_a^b R_{bc}$. Then $\rho$ is a closed real $(1, 1)$-form on $M$, with de Rham cohomology class $[\rho] = 2\pi c_1(M) \in H^2(M, \mathbb{R})$, where $c_1(M)$ is the first Chern class of $M$ in $H^2(M, \mathbb{Z})$. The Calabi conjecture specifies which closed $(1, 1)$-forms can be the Ricci forms of a Kähler metric on $M$.

**The Calabi conjecture** *Let $(M, J)$ be a compact, complex manifold, and $g'$ a Kähler metric on $M$, with Kähler form $\omega'$. Suppose that $\rho$ is a real, closed $(1, 1)$-form on $M$ with $[\rho] = 2\pi c_1(M)$. Then there exists a unique Kähler metric $g$ on $M$ with Kähler form $\omega$, such that $[\omega] = [\omega'] \in H^2(M, \mathbb{R})$, and the Ricci form of $g$ is $\rho$.*

Note that $[\omega] = [\omega']$ says that $g$ and $g'$ are in the same Kähler class. The conjecture was posed by Calabi in 1954, and was eventually proved by Yau in 1976. Its importance to us is that when the canonical bundle $K_M$ is trivial, so that $c_1(M) = 0$, we can take $\rho \equiv 0$, and then $g$ is Ricci-flat. Since $K_M$ is trivial, it has a nonzero holomorphic section, a holomorphic $(m, 0)$-form $\Omega$. As $g$ is Ricci-flat, it follows that $\nabla\Omega = 0$, where $\nabla$ is the Levi-Civita connection of $g$. Rescaling $\Omega$ by a complex constant makes [1] hold, and then $(M, J, g, \Omega)$ is a Calabi–Yau $m$-fold. This proves:

**Theorem 4** *Let $(M, J)$ be a compact complex $m$-manifold with $K_M$ trivial. Then every Kähler class on $M$ contains a unique Ricci-flat Kähler metric $g$. There exists a holomorphic $(m, 0)$-form $\Omega$, unique up to change of phase $\Omega \mapsto \mathrm{e}^{\mathrm{i}\theta}\Omega$, such that $(M, J, g, \Omega)$ is a Calabi–Yau $m$-fold.*

Using algebraic geometry, one can produce many examples of complex $m$-folds $(M, J)$ satisfying these conditions, such as the Fermat $(m + 2)$-tic

$$\{[z_0, \ldots, z_{m+1}] \in \mathbb{CP}^{m+1} : z_0^{m+2} + \cdots + z_{m+1}^{m+2} = 0\} \qquad [2]$$

Therefore, Calabi–Yau $m$-folds are very abundant.

### Special Lagrangian Submanifolds

**Definition 5** Let $(M, J, g, \Omega)$ be a Calabi–Yau $m$-fold. Then $\mathrm{Re}\,\Omega$ is a calibration on the Riemannian manifold $(M, g)$. An oriented real $m$-dimensional submanifold $N$ in $M$ is called a special Lagrangian submanifold (SL $m$-fold) if it is calibrated with respect to $\mathrm{Re}\,\Omega$.

Here is an alternative definition of SL $m$-folds. It is often more useful than Definition 5.

**Proposition 6** *Let $(M, J, g, \Omega)$ be a Calabi–Yau $m$-fold, with Kähler form $\omega$, and $N$ a real $m$-dimensional submanifold in $M$. Then $N$ admits an orientation making it into an SL $m$-fold in $M$ if and only if $\omega|_N \equiv 0$ and $\mathrm{Im}\,\Omega|_N \equiv 0$.*

Regard $N$ as an immersed submanifold, with immersion $\iota : N \to M$. Then $[\omega|_N]$ and $[\mathrm{Im}\,\Omega|_N]$ are unchanged under continuous variations of the immersion $\iota$. Thus, $[\omega|_N] = [\mathrm{Im}\,\Omega|_N] = 0$ is a necessary condition not just for $N$ to be special Lagrangian, but also for any isotopic submanifold $N'$ in $M$ to be special Lagrangian. This proves:

**Corollary 7** *Let $(M, J, g, \Omega)$ be a Calabi–Yau $m$-fold, and $N$ a compact real $m$-submanifold in $M$. Then a necessary condition for $N$ to be isotopic to a special Lagrangian submanifold $N'$ in $M$ is that $[\omega|_N] = 0$ in $H^2(N, \mathbb{R})$ and $[\mathrm{Im}\,\Omega|_N] = 0$ in $H^m(N, \mathbb{R})$.*

### Deformations of Compact SL *m*-Folds

The deformation theory of compact special Lagrangian manifolds was studied by McLean (1998), who proved the following result:

**Theorem 8** *Let $(M, J, g, \Omega)$ be a Calabi–Yau $m$-fold, and $N$ a compact special Lagrangian $m$-fold in $M$. Then the moduli space $\mathcal{M}_N$ of special Lagrangian deformations of $N$ is a smooth manifold of dimension $b^1(N)$, the first Betti number of $N$.*

*Sketch proof*. Suppose for simplicity that $N$ is an embedded submanifold. There is a natural orthogonal decomposition $TM|_N = TN \oplus \nu$, where $\nu \to N$ is the normal bundle of $N$ in $M$. As $N$ is Lagrangian, the complex structure $J : TM \to TM$ gives an isomorphism $J : \nu \to TN$. But the metric $g$ gives an isomorphism $TN \cong T^*N$. Composing these two gives an isomorphism $\nu \cong T^*N$.

Let $T$ be a small tubular neighborhood of $N$ in $M$. Then we can identify $T$ with a neighborhood of the zero section in $\nu$. Using the isomorphism $\nu \cong T^*N$, we have an identification between $T$ and a neighborhood of the zero section in $T^*N$. This can be chosen to identify the Kähler form $\omega$ on $T$ with the natural symplectic

structure on $T^*N$. Let $\pi: T \to N$ be the obvious projection.

Under this identification, submanifolds $N'$ in $T \subset M$ which are $C^1$ close to $N$ are identified with the graphs of small smooth sections $\alpha$ of $T^*N$. That is, submanifolds $N'$ of $M$ close to $N$ are identified with 1-forms $\alpha$ on $N$. We need to know: which 1-forms $\alpha$ are identified with SL $m$-folds $N'$?

Now, $N'$ is special Lagrangian if $\omega|_{N'} \equiv \operatorname{Im}\Omega|_{N'} \equiv 0$. But $\pi|_{N'}: N' \to N$ is a diffeomorphism, so we can push $\omega|_{N'}$ and $\operatorname{Im}\Omega|_{N'}$ down to $N$, and regard them as functions of $\alpha$. Calculation shows that

$$\pi_*(\omega|_{N'}) = \mathrm{d}\alpha \quad \text{and} \quad \pi_*(\operatorname{Im}\Omega|_{N'}) = F(\alpha, \nabla\alpha)$$

where $F$ is a nonlinear function of its arguments. Thus, the moduli space $\mathcal{M}_N$ is locally isomorphic to the set of small 1-forms $\alpha$ on $N$ such that $\mathrm{d}\alpha \equiv 0$ and $F(\alpha, \nabla\alpha) \equiv 0$.

Now it turns out that $F$ satisfies $F(\alpha, \nabla\alpha) \approx \mathrm{d}(*\alpha)$ when $\alpha$ is small. Therefore, $\mathcal{M}_N$ is locally approximately isomorphic to the vector space of 1-forms $\alpha$ with $\mathrm{d}\alpha = \mathrm{d}(*\alpha) = 0$. But by Hodge theory, this is isomorphic to the de Rham cohomology group $H^1(N, \mathbb{R})$, and is a manifold with dimension $b^1(N)$.

To carry out this last step rigorously requires some technical machinery: one must work with certain Banach spaces of sections of $T^*N$, $\Lambda^2 T^*N$ and $\Lambda^m T^*N$, use elliptic regularity results to prove that the map $\alpha \mapsto (\mathrm{d}\alpha,\ F(\alpha, \nabla\alpha))$ has closed image in these Banach spaces, and then use the implicit function theorem for Banach spaces to show that the kernel of the map is what is expected.

### Obstructions to Existence of Compact SL *m*-Folds

Let $\{(M, J_t, g_t, \Omega_t) : t \in (-\epsilon, \epsilon)\}$ be a smooth one-parameter family of Calabi–Yau $m$-folds. Suppose $N_0$ is an SL $m$-fold in $(M, J_0, g_0, \Omega_0)$. When can we extend $N_0$ to a smooth family of SL $m$-folds $N_t$ in $(M, J_t, g_t, \Omega_t)$ for $t \in (-\epsilon, \epsilon)$?

By Corollary 7, a necessary condition is that $[\omega_t|_{N_0}] = [\operatorname{Im}\Omega_t|_{N_0}] = 0$ for all $t$. Our next result shows that locally, this is also a sufficient condition.

**Theorem 9** *Let $\{(M, J_t, g_t, \Omega_t) : t \in (-\epsilon, \epsilon)\}$ be a smooth one-parameter family of Calabi–Yau m-folds, with Kähler forms $\omega_t$. Let $N_0$ be a compact* SL *m-fold in $(M, J_0, g_0, \Omega_0)$, and suppose that $[\omega_t|_{N_0}] = 0$ in $H^2(N_0, \mathbb{R})$ and $[\operatorname{Im}\Omega_t|_{N_0}] = 0$ in $H^m(N_0, \mathbb{R})$ for all $t \in (-\epsilon, \epsilon)$. Then $N_0$ extends to a smooth one-parameter family $\{N_t : \mathrm{t} \in (-\delta, \delta)\}$, where $0 < \delta \le \epsilon$ and $N_t$ is a compact* SL *m-fold in $(M, J_t, g_t, \Omega_t)$.*

This can be proved using similar techniques to Theorem 8. Note that the condition $[\operatorname{Im}\Omega_t|_{N_0}] = 0$ for all $t$ can be satisfied by choosing the phases of the $\Omega_t$ appropriately, and if the image of $H_2(N, \mathbb{Z})$ in $H_2(M, \mathbb{R})$ is zero, then the condition $[\omega|_N] = 0$ holds automatically.

Thus, the obstructions $[\omega_t|_{N_0}] = [\operatorname{Im}\Omega_t|_{N_0}] = 0$ in Theorem 9 are actually fairly mild restrictions, and SL $m$-folds should be considered as pretty stable under small deformations of the Calabi–Yau structure.

**Remark** The deformation and obstruction theory of compact SL $m$-folds are extremely well behaved compared to many other moduli space problems in differential geometry. In other geometric problems (such as the deformations of complex structures on a complex manifold, or pseudoholomorphic curves in an almost-complex manifold, or instantons on a Riemannian 4-manifold), the deformation theory often has the following general structure.

There are vector bundles $E$, $F$ over a compact manifold $M$, and an elliptic operator $P: C^\infty(E) \to C^\infty(F)$, usually first order. The kernel $\operatorname{Ker} P$ is the set of infinitesimal deformations, and the cokernel Coker $P$ the set of obstructions. The actual moduli space $\mathcal{M}$ is locally the zeros of a nonlinear map $\Psi: \operatorname{Ker} P \to \operatorname{Coker} P$.

In a generic case, $\operatorname{Coker} P = 0$, and then the moduli space $\mathcal{M}$ is locally isomorphic to $\operatorname{Ker} P$, and so is locally a manifold with dimension $\operatorname{ind}(P)$. However, in nongeneric situations $\operatorname{Coker} P$ may be nonzero, and then the moduli space $\mathcal{M}$ may be nonsingular, or have an unexpected dimension.

However, SL $m$-folds do not follow this pattern. Instead, the obstructions are topologically determined, and the moduli space is always smooth, with dimension given by a topological formula. This should be regarded as a minor mathematical miracle.

### Mirror Symmetry and the SYZ Conjecture

Mirror symmetry is a mysterious relationship between pairs of Calabi–Yau 3-folds $M, \hat{M}$, arising from a branch of physics known as string theory, and leading to some very strange and exciting conjectures about Calabi–Yau 3-folds, many of which have been proved in special cases.

In the beginning (the 1980s), mirror symmetry seemed mathematically completely mysterious. But there are now two complementary conjectural theories, due to Kontsevich and Strominger–Yau–Zaslow, which explain mirror symmetry in a fairly mathematical way. Probably both are true, at some level. The second proposal, due to Strominger, Yau, and Zaslow (1996), is known as the SYZ conjecture. Here is an attempt to state it.

**The SYZ conjecture** *Suppose $M$ and $\hat{M}$ are mirror Calabi–Yau 3-folds. Then (under some additional conditions), there should exist a compact topological 3-manifold $B$ and surjective, continuous maps $f : M \to B$ and $\hat{f} : \hat{M} \to B$, such that*

(i) *There exists a dense open set $B_0 \subset B$, such that for each $b \in B_0$, the fibers $f^{-1}(b)$ and $\hat{f}^{-1}(b)$ are nonsingular special Lagrangian 3-tori $T^3$ in $M$ and $\hat{M}$. Furthermore, $f^{-1}(b)$ and $\hat{f}^{-1}(b)$ are in some sense dual to one another.*

(ii) *For each $b \in \Delta = B \backslash B_0$, the fibers $f^{-1}(b)$ and $\hat{f}^{-1}(b)$ are expected to be singular special Lagrangian 3-folds in $M$ and $\hat{M}$.*

The fibrations $f$ and $\hat{f}$ are called special Lagrangian fibrations, and the set of singular fibers $\Delta$ is called the discriminant. In part (i), the nonsingular fibers of $f$ and $\hat{f}$ are supposed to be dual tori. What does this mean?

On the topological level, we can define duality between two tori $T, \hat{T}$ to be a choice of isomorphism $H^1(T, \mathbb{Z}) \cong H_1(\hat{T}, \mathbb{Z})$. We can also define duality between tori equipped with flat Riemannian metrics. Write $T = V/\Lambda$, where $V$ is a Euclidean vector space and $\Lambda$ a lattice in $V$. Then the dual torus $\hat{T}$ is defined to be $V^*/\Lambda^*$, where $V^*$ is the dual vector space and $\Lambda^*$ the dual lattice. However, there is no notion of duality between nonflat metrics on dual tori.

Strominger, Yau, and Zaslow argue only that their conjecture holds when $M, \hat{M}$ are close to the "large complex structure limit." In this case, the diameters of the fibers $f^{-1}(b), \hat{f}^{-1}(b)$ are expected to be small compared to the diameter of the base space $B$, and away from singularities of $f, \hat{f}$, the metrics on the nonsingular fibers are expected to be approximately flat. So, part (i) of the SYZ conjecture says that for $b \in B \backslash B_0$, $f^{-1}(b)$ is approximately a flat Riemannian 3-torus, and $\hat{f}^{-1}(b)$ is approximately the dual flat Riemannian torus.

Mathematical research on the SYZ conjecture has followed two broad approaches. The first could be described as symplectic topological. For this, we treat $M, \hat{M}$ just as symplectic manifolds and $f$, $\hat{f}$ just as Lagrangian fibrations. We also suppose $B$ is a smooth 3-manifold and $f, \hat{f}$ are smooth maps. Under these simplifying assumptions, Mark Gross, Wei-Dong Ruan, and others have built up a beautiful, detailed picture of how dual SYZ fibrations work at the global topological level.

The second approach could be described as local geometric. Here, we try to take the special Lagrangian condition seriously from the outset, and focus on the local behavior of special Lagrangian submanifolds, and especially their singularities, rather than on global topological questions. In addition, we are intrested in what fibrations of generic Calabi–Yau 3-folds might look like.

There is now a well-developed theory of SL $m$-folds with isolated singularities modeled on cones (Joyce 2003a). This is applied to SL fibrations and the SYZ conjecture in Joyce (2003a, b), leading to the tentative conclusions that for generic Calabi–Yau 3-folds $M$, special Lagrangian fibrations $f : M \to B$ will be only piecewise smooth, and have discriminants $\Delta$ of real codimension 1 in $B$, in contrast to smooth fibrations which have $\Delta$ of codimension 2. We also argue that for generic mirrors $M, \hat{M}$ and $f, \hat{f}$, the discriminants $\Delta, \hat{\Delta}$ cannot be homeomorphic and so do not coincide. This contradicts part (ii) above.

A better way to formulate the SYZ conjecture may be in terms of families of mirror Calabi–Yau 3-folds $M_t, \hat{M}_t$ and fibrations $f_t : M_t \to B$, $\hat{f}_t : \hat{M}_t \to B$ for $t \in (0, \epsilon)$ which approach the "large complex structure limit" as $t \to 0$. Then we could require the discriminants $\Delta_t$, $\hat{\Delta}_t$ of $f_t, \hat{f}_t$ to converge to some common, codimension 2 limit $\Delta_0$ as $t \to 0$.

It is an important, and difficult, open problem to construct examples of special Lagrangian fibrations of compact, holonomy SU(3) Calabi–Yau 3-folds. None are currently known.

*See also:* Minimal submanifolds; Mirror Symmetry: A Geometric Survey; Moduli Spaces: An Introduction; Riemannian Holonomy Groups and Exceptional Holonomy.

## Further Reading


Gross M, Huybrechts D, and Joyce D (2003) *Calabi–Yau Manifolds and Related Geometries*, Universitext Series, Berlin: Springer.

Harvey R and Lawson HB (1982) Calibrated geometries. *Acta Mathematica* 148: 47–157.

Joyce DD (2000) *Compact Manifolds with Special Holonomy*. Oxford: Oxford University Press.

Joyce DD (2003a) Special Lagrangian submanifolds with isolated conical singularities. V. Survey and applications. *Journal of Differential Geometry* 63: 279–347, math.DG/0303272.

Joyce DD (2003b) Singularities of special Lagrangian fibrations and the SYZ conjecture. *Communications in Analysis and Geometry* 11: 859–907, math.DG/0011179.

Joyce DD (2003c) U(1)-invariant special Lagrangian 3-folds in $\mathbb{C}^3$ and special Lagrangian fibrations. *Turkish Mathematical Journal* 27: 99–114, math.DG/0206016.

McLean RC (1998) Deformations of calibrated submanifolds. *Communications in Analysis and Geometry* 6: 705–747.

Strominger A, Yau S-T, and Zaslow E (1996) Mirror symmetry is T-duality. *Nuclear Physics B* 479: 243–259, hep-th/9606040.

# Calogero–Moser–Sutherland Systems of Nonrelativistic and Relativistic Type

**S N M Ruijsenaars**, Centre for Mathematics and Computer Science, Amsterdam, The Netherlands



## Introduction

Systems of Calogero–Moser–Sutherland (CMS) type form a class of finite-dimensional dynamical systems that are integrable both at the classical and at the quantum level. The CMS systems describe $N$ point particles moving on a line or on a ring, interacting via pair potentials that are specific functions of four types, namely rational (I), hyperbolic (II), trigonometric (III), and elliptic (IV). They occur not only in a nonrelativistic (Galilei-invariant), but also in a relativistic (Poincaré-invariant) setting. Thus, one can distinguish a hierarchy of 16 physically distinct versions (classical/quantum, nonrelativistic/relativistic, type I–IV), the most general one being the quantum relativistic type IV system.

The nonrelativistic systems date back to pioneering work by Calogero, Sutherland, and Moser in the early 1970s. The pair potential structure of the interaction can be encoded in the root system $A_{N-1}$, and there also exist integrable versions for all of the remaining root systems. The classical systems are given by $N$ Poisson commuting Hamiltonians with a polynomial dependence on the particle momenta $p_1, \ldots, p_N$. Accordingly, the quantum versions are described by $N$ commuting Hamiltonians that are partial differential operators.

The relativistic systems were introduced in the mid-1980s, at the classical level by Ruijsenaars and Schneider, and at the quantum level by Ruijsenaars. They converge to the nonrelativistic systems in the limit $c \to \infty$, where $c$ is the speed of light. Again, the systems can be related to the root system $A_{N-1}$, and they admit integrable versions for other root systems. All of the commuting classical Hamiltonians depend exponentially on generalized momenta $p_1, \ldots, p_N$. Hence, the associated commuting quantum Hamiltonians are analytic difference operators.

The above integrable systems can be further generalized by allowing supersymmetry or internal degrees of freedom ("spins"), coupled in quite special ways to retain integrability. In this article, however, the focus is on the 16 versions of the $A_{N-1}$-symmetric CMS systems without internal degrees of freedom. The primary aim is to acquaint the reader with their definition and integrability, and with their most prominent features and interrelationships. Second, we intend to give a rough sketch of the state of the art concerning explicit solutions for the various versions. This involves a concretization of the action-angle maps and eigenfunction transforms that simultaneously diagonalize the commuting dynamics, paying special attention to their remarkable duality properties.

It is beyond the scope of this article to review the hundreds of papers specifically dealing with CMS type systems, let alone the much larger literature where they play some role. Indeed, the systems have been encountered in a great many different contexts and they are related to a host of other integrable systems in various ways. Accordingly, they can be studied from the perspective of various subfields of mathematics and theoretical physics. First some of these perspectives and relations to seemingly quite different topics will be mentioned before embarking on the far more focused survey.

Staying first within the confines of the CMS type systems, some nonobvious limits yielding other familiar finite-dimensional integrable systems will be mentioned. To begin with, all of the $A_{N-1}$ type systems give rise to systems with a Toda type (exponential "nearest neighbor") interaction via a suitable limiting transition (basically a strong-coupling limit). This leads to integrable $N$-particle systems with a classical/quantum, nonrelativistic/relativistic, nonperiodic/periodic version; starting from the quantum relativistic periodic Toda system, the remaining seven versions can be obtained by suitable limits.

Next, we recall that the quantum system of $N$ nonrelativistic bosons on the line or ring interacting via a pair potential of $\delta$-function type is soluble via a Bethe ansatz, with the "line version" exhibiting quantum soliton behavior (factorized scattering). It has been shown that there exist scaling limits of eigenfunctions for suitable CMS systems that give rise to the latter Bethe type eigenfunctions for $N = 2$, while convergence for $N > 2$ is plausible, but has not been demonstrated thus far.

Via suitable analytic continuations preserving reality/formal self-adjointness, one can arrive at CMS systems with more than one species of particle (particles and "antiparticles"). Likewise, analytic continuations and appropriate limits of CMS systems associated with root sytems other than $A_{N-1}$ lead to a further proliferation of $N$-dimensional integrable systems. Typically, such limits refer either

to the commuting Hamiltonians (the Toda limit being a case in point) or to the joint eigenfunctions (as exemplified by the $\delta$-function system limit); it seems difficult to control both sets of quantities at once.

Starting from the spin type CMS systems, another kind of limit can be taken. Specifically, by "freezing" the particles at equilibrium positions, it is possible to arrive at integrable spin chains of Haldane–Shastry and Inozemtsev type.

At this point, it is expedient to insert a brief remark on finite-dimensional integrable systems. As the term suggests, one may expect that, with due effort, such systems can be "integrated," or, equivalently, "solved." But it should be noted that the latter terms (let alone the qualifier "due effort") have no unambiguous mathematical meaning. Certainly, "solving" involves obtaining explicit information on the action-angle map and joint eigenfunction transform at the classical and quantum level, resp., but *a priori* it is not at all clear how far one can proceed.

Focusing again on the CMS systems and their relatives, it should be stressed that, in many cases, one is still far removed from a complete solution, especially for the elliptic CMS systems. In this regard the previous remark serves not only as a caveat, but also to make clear why the various vantage points provided by different subfields in mathematics and physics are crucial: typically, they yield complementary insights and distinct representations for solutions, serving different purposes.

To be sure, in first approximation the mathematics involved at the classical and quantum level is symplectic geometry and Hilbert space theory, resp. In point of fact, however, far more ingredients have turned out to be quite natural and useful. On the classical level, these include the theory of groups, Lie algebras and symmetric spaces, linear algebra and spectral theory, Riemann surface theory, and more generally algebraic geometry.

On the quantum level, the viewpoint of harmonic analysis on symmetric spaces is particularly natural and fruitful for the nonrelativistic CMS systems and their arbitrary root-system versions, whereas quantum groups/algebras/symmetric spaces can be tied in with the relativistic systems and their versions for other root systems. (The $c \to \infty$ limit amounts to the $q \to 1$ limit in the quantum group picture.) As a matter of fact, the whole area of special functions and their $q$-analogs is intimately related to the quantum CMS type systems (cf. also the last section of this article). Finally, the occurrence of commuting analytic difference operators in the relativistic ($q \neq 1$) systems leads to largely uncharted territory in the intersection of the theory of Hilbert space eigenfunction expansions and the theory of linear analytic difference equations.

The study of the thermodynamics ($N \to \infty$ limit with temperature $\geq 0$ and density $\geq 0$ fixed) associated with the trigonometric and elliptic CMS systems and their spin cousins yields its own circle of problems. It was initiated by Sutherland three decades ago, and even though a host of results on partition functions, correlation functions, fractional statistics, strong–weak coupling duality, relations to Yangians, etc., have meanwhile been obtained, many questions are still open. This area also has links with random-matrix theory, but the input from this field is thus far limited to certain discrete couplings.

The above $N$-dimensional integrable systems are related to a great many infinite-dimensional integrable systems, both at the classical and at the quantum level. On the one hand, there are structural analogs that have been used to advantage in the study of CMS systems, including Lax pair and $R$-matrix formulations, zero-curvature representations, bi-Hamiltonian formalism, Bäcklund transformations, time discretizations, and tools such as Baker–Akhiezer functions, Bethe ansatz, separation of variables, and Baxter-type $Q$-operators.

On the other hand, there are striking physical similarities between various soliton field theories (a prominent one being the sine-Gordon field theory) and infinite soliton lattices (in particular several Toda type lattices), and the CMS systems for special parameter values. Particularly conspicuous are the ties between the classical CMS systems and the KP and two-dimensional Toda hierarchies. The latter relations actually extend beyond the solitons, including rational and theta function solutions.

CMS systems are relevant in various other contexts not yet mentioned. A prominent one among these is a class of supersymmetric gauge field theories. In this quantum context, the classical CMS systems have surfaced in the description of moduli spaces encoding the vacuum structure (Seiberg–Witten theory). Equally surprising, certain classical CMS systems (with internal degrees of freedom) have found a second application in a quantum context, namely in the description of quantum chaos (level repulsion).

We conclude this introduction by listing additional disparate subjects where connections with CMS type systems have been found. These include the theory of Sklyanin, affine Hecke, Kac–Moody, Virasoro and $W$-algebras, equations of Knizhnik–Zamolodchikov, Yang–Baxter, Witten–Dijkgraaf–Verlinde–Verlinde, and Painlevé type, Gaudin,

Hitchin, Wess–Zumino, matrix and quasi-exactly solvable models, Dunkl–Cherednik and Polychronakos operators, the quantum Hall effect and quantum transport, two-dimensional Yang–Mills theory, functional equations, integrable mappings, Huygens' principle, and the bispectral problem.

## Classical Nonrelativistic CMS Systems

A system of $N$ nonrelativistic equal-mass $m$ particles on the line interacting via pair potentials can be described by a Hamiltonian

$$H = \frac{1}{2m}\sum_{j=1}^{N} p_j^2 + \sum_{1\le j<k\le N} V(x_j - x_k), \quad m > 0 \quad [1]$$

The CMS systems are defined by four distinct choices of pair potential. The simplest choice reads

$$V(x) = g^2/mx^2, \quad g > 0 \quad (\mathrm{I}) \quad [2]$$

Hence, the coupling constant $g$ has dimension [action] (the product of [position] and [momentum]). This potential is clearly repulsive. Thus, each initial state in the phase space

$$\Omega = \{(x,p) \in \mathbb{R}^{2N} \,|\, x \in G\} \quad [3]$$

where $G$ is the configuration space

$$G = \{x \in \mathbb{R}^N \,|\, x_N < \cdots < x_1\} \quad [4]$$

is a scattering state.

The next level is given by the hyperbolic choice

$$V(x) = g^2\nu^2/m\sinh^2(\nu x), \quad \nu > 0 \quad (\mathrm{II}) \quad [5]$$

Hence, $\nu$ has dimension $[\text{position}]^{-1}$, and the previous system arises by taking $\nu$ to 0. It is clear that [5] yields again a repulsive particle system, so that each state in $\Omega$ given by [3] is a scattering state.

The highest level in the hierarchy is the elliptic level, where

$$V(x) = g^2\wp(x;\omega,\omega')/m, \quad \omega, -\mathrm{i}\omega' > 0 \quad (\mathrm{IV}) \quad [6]$$

and $\wp(x;\omega,\omega')$ denotes the Weierstrass $\wp$-function with periods $2\omega$ and $2\omega'$. It is beyond the scope of this article to elaborate on the elliptic regime, even though it is of considerable interest. It reappears in later sections as the most general regime in which integrability holds true. Indeed, a prominent feature of the elliptic case [6] is that it can be specialized both to the hyperbolic case [5] and to the trigonometric case, given by

$$V(x) = g^2\nu^2/m\sin^2(\nu x) \quad (\mathrm{III}) \quad [7]$$

To obtain the hyperbolic specialization, one should take $\omega' = \mathrm{i}\pi/2\nu$ and send $\omega$ to $\infty$; then [6] reduces to [5] (up to an additive constant). Likewise, [7] results from [6] by choosing $\omega = \pi/2\nu$ and taking $-\mathrm{i}\omega'$ to $\infty$.

The physical picture associated with the trigonometric and elliptic systems is quite different from that of the rational and hyperbolic ones. Of course, the potentials [7] and [6] are again repulsive, but now the internal motion is confined and oscillatory. More specifically, due to energy conservation the phase spaces

$$\Omega_{\mathrm{III}} = G_{\mathrm{III}} \times \mathbb{R}^N, \quad G_{\mathrm{III}} = \{x_N < \cdots < x_1, x_1 - x_N < \pi/\nu\} \quad [8]$$

$$\Omega_{\mathrm{IV}} = G_{\mathrm{IV}} \times \mathbb{R}^N, \quad G_{\mathrm{IV}} = \{x_N < \cdots < x_1, x_1 - x_N < 2\omega\} \quad [9]$$

are left invariant by the flow generated by the trigonometric and elliptic $N$-particle Hamiltonian, resp.

Alternatively, one may interpret the trigonometric Hamiltonian as describing particles constrained to move on a circle and interacting via the inverse square potential [2]. In this picture, the quantities $2\nu x_1, \ldots, 2\nu x_N$ are viewed as angular positions on the circle, and one needs a suitable quotient of the phase space [8] by a discrete group action to describe a state of the system.

Turning to integrability aspects, we begin by noting that the total momentum Hamiltonian

$$P = \sum_{j=1}^{N} p_j \quad [10]$$

obviously Poisson commutes with the above defining Hamiltonians of the systems. For $N = 2$, therefore, integrability is plain. It is possible to write down explicitly the higher commuting Hamiltonians for $N > 2$ as well but, in the nonrelativistic setting, it is more illuminating to characterize them as the power traces or (equivalently) the symmetric functions of a so-called Lax matrix.

The Lax matrix is an $N \times N$ matrix-valued function on the phase space of the system. It plays a pivotal role not only for understanding integrability, but also for setting up an action-angle transformation. The latter issue is discussed again later. Here the more conspicuous features of the Lax matrix will be explained, focusing on the type II system for expository ease. Then one can choose

$$L_{jj} = p_j, \; L_{jk} = \mathrm{i}g\nu/\sinh\nu(x_j - x_k), \quad j,k = 1,\ldots,N, \; j \neq k \quad [11]$$

Thus, $L$ is Hermitean and we have

$$\mathrm{tr}\, L = P, \qquad \mathrm{tr}\, L^2 = 2mH \quad [12]$$

(The rational Lax matrix results from [11] by taking $\nu \to 0$, and the trigonometric one by taking $\nu \to i\nu$. The elliptic Lax matrix has a similar structure, but it involves an extra "spectral" parameter.)

Although not obvious, it is true that all of the power traces

$$H_k = \frac{1}{k}\operatorname{tr} L^k, \quad k = 1, \ldots, N \qquad [13]$$

are in involution (i.e., Poisson commute). One way to understand this involves the so-called Lax pair equation associated with the Hamiltonian flow generated by $H = H_2/m$. This involves a second $N \times N$ matrix function given by

$$\begin{aligned} M_{jj} &= \sum_{l \neq j} \frac{-ig\nu^2}{m \sinh^2 \nu(x_j - x_l)} \\ M_{jk} &= \frac{ig\nu^2 \cosh \nu(x_j - x_k)}{m \sinh^2 \nu(x_j - x_k)} \\ & j \neq k \end{aligned} \qquad [14]$$

When the positions and momenta in $L$ and $M$ evolve according to the $H$-flow, one has

$$\dot{L}_t = [M_t, L_t] \qquad [15]$$

where $[\cdot, \cdot]$ is the matrix commutator. (Indeed, [15] amounts to the Hamilton equations, as is readily checked.) Since $M$ is anti-Hermitean, it is not difficult to derive from this Lax pair equation that the flow is isospectral: $L_t$ is related to $L_0$ by a unitary transformation $L_t = U_t L_0 U_t^*$ obtained from $M_t$, so that the spectrum of $L_t$ is time independent.

This argument already shows the existence of $N$ conserved quantities under the $H$-flow, namely the $N$ eigenvalues of $L$. It is, however, simpler to work with either the power traces $H_k$ given by [13] or with the symmetric functions $S_k$ of $L$, given by

$$\det(\mathbf{1}_N + \lambda L) = \sum_{k=0}^{N} \lambda^k S_k \qquad [16]$$

These Hamiltonians depend only on the eigenvalues of $L$, so they are also conserved under the flow. Note that

$$S_1 = P, \qquad S_2 = P^2 - mH \qquad [17]$$

To see why these Hamiltonians are in involution, one can invoke the long-time asymptotics of the $H$-flow. It reads

$$\begin{aligned} & p(t) \sim \hat{p}, \qquad \hat{p}_N < \cdots < \hat{p}_1, \\ & x_j(t) \sim x_j^+ + t\hat{p}_j/m, \\ & j = 1, \ldots, N, \ t \to \infty \end{aligned} \qquad [18]$$

Accordingly, one gets

$$L_t \sim \operatorname{diag}(\hat{p}_1, \ldots, \hat{p}_N) = L_\infty, \quad t \to \infty \qquad [19]$$

Since the time evolution is a canonical transformation and the Poisson brackets $\{H_k, H_l\}$ are time independent (by the Jacobi identity), it now readily follows from [19] that they vanish. (Indeed, $H_k$ and $H_l$ reduce to power traces of $L_\infty$, and the asymptotic momenta $\hat{p}_1, \ldots, \hat{p}_N$ Poisson commute.)

## Quantum Nonrelativistic CMS Systems

The canonical quantization prescription

$$p_j \to -i\hbar\partial/\partial x_j, \quad j = 1, \ldots, N \qquad [20]$$

($\hbar$ being the Planck constant) gives rise to an unambiguous quantum Hamiltonian

$$H = -\frac{\hbar^2}{2m}\sum_{j=1}^{N} \partial_j^2 + \sum_{1 \le j < k \le N} V(x_j - x_k) \qquad [21]$$

for any classical Hamiltonian [1]. Thus, the defining Hamiltonians of the above systems give rise to well-defined partial differential operators (PDOs), which act on suitable dense subspaces of the Hilbert space $L^2(G_\kappa, dx), \kappa = \mathrm{I}, \ldots, \mathrm{IV}$, with $G_{\mathrm{I}}$ and $G_{\mathrm{II}}$ given by $G$ in [4], and $G_{\mathrm{III}}, G_{\mathrm{IV}}$ by [8] and [9], respectively.

We recall that there is no general result ensuring that a classically integrable system admits an integrable quantum version. More precisely, when one substitutes [20] in $N$ Poisson commuting Hamiltonians, it need not be true that they commute as quantum operators, even when no ordering ambiguities are present. For the power trace Hamiltonians such ambiguities do occur. (For example, [11] gives rise to a term in $H_3$ proportional to $p_1/\sinh^2 \nu(x_1 - x_2)$.) On the other hand, no noncommuting factors occur in the quantization of $S_1, \ldots, S_N$. To verify this, one need only note that $S_k$ equals the sum of all $k \times k$ principal minors of $L$, cf. [16]; choosing a diagonal element $p_j$ in a summand, one therefore has no dependence on $x_j$ in the remaining factors, hence no ordering ambiguity.

As a result, the prescription [20] yields $N$ unambiguous operators $S_k(x, -i\hbar\nabla)$, which are moreover formally self-adjoint on $L^2(G_\kappa, \mathrm{d}x)$ for each of the four cases $\kappa = \mathrm{I}, \ldots, \mathrm{IV}$. Although by no means obvious, it is true that these operators do commute. Thus, integrability is preserved under quantization of the above systems. Now the power traces of a matrix can be expressed as polynomials in the symmetric functions (via the Newton

identities), so this yields an ordering ensuring that the quantized power traces commute as well.

Just as the action-angle transformation for a classically integrable system "diagonalizes" all of the Poisson commuting Hamiltonians at once (in the sense that the transformed Hamiltonians depend only on the action variables), one expects that there exists a unitary operator that transforms all of the commuting Hamiltonians to diagonal form. In the classical setting, the existence of this diagonalizing map follows (under suitable technical restrictions) from the Liouville–Arnold theorem, whereas in the quantum context the existence of such a joint eigenfunction transformation is a far more delicate issue. This problem is briefly discussed later again, noting here that the solutions obtained to date vary considerably in completeness and "explicitness" for the four regimes.

## Classical Relativistic CMS Systems

The nonrelativistic spacetime symmetry group is the Galilei group. Its Lie algebra is represented by the time translation generator $H$ given by [1], space translation generator $P$ given by [10], and the Galilei boost generator

$$B = -m\sum_{j=1}^{N} x_j \qquad [22]$$

More precisely, the Poisson brackets are given by

$$\{H, P\} = 0, \quad \{H, B\} = P, \quad \{P, B\} = Nm \qquad [23]$$

so that the last bracket does not vanish (as is the case for the Galilei Lie algebra). This deviation is inconsequential, however, since the constant $Nm$ (central extension) yields trivial Hamilton equations.

The relativistic spacetime symmetry group (Poincaré group) yields a Lie algebra that differs from [23] only in $Nm$ being replaced by $H/c^2$, where $c$ is the speed of light. Clearly, the functions

$$H = mc^2\sum_{j=1}^{N}\cosh\left(\frac{p_j}{mc}\right)$$
$$P = mc\sum_{j=1}^{N}\sinh\left(\frac{p_j}{mc}\right) \qquad [24]$$

together with $B$ given by [22] give rise to these altered Poisson brackets. Physically, these three generators describe a system of $N$ relativistic free mass-$m$ particles in terms of their rapidities $p_j/mc$.

A natural ansatz to take interaction into account now reads

$$H = mc^2\sum_{j=1}^{N}\cosh\left(\frac{p_j}{mc}\right)V_j(x)$$
$$P = mc\sum_{j=1}^{N}\sinh\left(\frac{p_j}{mc}\right)V_j(x) \qquad [25]$$
$$V_j(x) = \prod_{k\neq j} f(x_j - x_k)$$

Indeed, it is plain that this still entails

$$\{H, B\} = P, \qquad \{P, B\} = H/c^2 \qquad [26]$$

But to obtain a relativistic particle system, the time and space translations must also commute. The corresponding requirement $\{H, P\} = 0$ yields a severe constraint on the "pair potential" function $f(x)$ in [25] whenever $N > 2$. (For $N = 2$, one gets $\{H, P\} = 0$ irrespective of the choice of $f$.)

As it turns out, the vanishing requirement is satisfied when

$$f^2(x) = a + b\wp(x) \qquad [27]$$

where $a$, $b$ are constants and $\wp(x)$ is the Weierstrass function already encountered. Taking, for example, $a, b > 0$, one can take the positive square root of the right-hand side of [27]. This choice of $f(x)$ yields the defining Hamiltonian of the relativistic elliptic system (type IV). In the three degenerate cases, it is convenient to choose

$$f(x) = \begin{cases} (1 + g^2/m^2c^2x^2)^{1/2} & \text{(I)} \\ (1 + \sin^2(\nu g/mc)/\sinh^2(\nu x))^{1/2} & \text{(II)} \\ (1 + \sinh^2(\nu g/mc)/\sin^2(\nu x))^{1/2} & \text{(III)} \end{cases} \qquad [28]$$

It is an elementary exercise to check that this implies

$$\lim_{c\to\infty}(H - Nmc^2) = H_{\text{nr}}, \quad \lim_{c\to\infty} P = P_{\text{nr}} \qquad [29]$$

where $H_{\text{nr}}$ and $P_{\text{nr}}$ are the above nonrelativistic time and space translation generators. Hence, the defining Hamiltonians of the relativistic systems reduce to their nonrelativistic counterparts in the limit $c \to \infty$.

The special character of the function [27] makes itself felt not only in ensuring Poincaré invariance, but also in entailing integrability. To begin with, note that the functions

$$S_{\pm N} = \exp\left(\pm\beta\sum_{j=1}^{N} p_j\right), \quad \beta = 1/mc \qquad [30]$$

commute with $H$ and $P$, so that integrability for $N=3$ is plain. More generally, the Hamiltonians

$$S_{\pm l} = \sum_{\substack{I\subset\{1,\ldots,N\}\\ |I|=l}} \exp\left(\pm\beta\sum_{j\in I} p_j\right)\prod_{\substack{j\in I\\ k\notin I}} f(x_j - x_k), \quad l=1,\ldots,N \qquad [31]$$

can be shown to mutually commute. Clearly, one has

$$S_{-l} = S_{-N}S_{N-l}, \quad l=1,\ldots,N-1 \qquad [32]$$

and

$$H = (S_1 + S_{-1})/2m\beta^2, \quad P = (S_1 - S_{-1})/2\beta \qquad [33]$$

As anticipated by the notation, the functions $S_1,\ldots,S_N$ may be viewed as the symmetric functions of a Lax matrix. More precisely, in the elliptic case this is true up to multiplicative constants that depend on a spectral parameter occurring in the Lax matrix. As before, only the Lax matrix for the type II system is specified here. In this case, one can dispense with the spectral parameter and choose

$$L_{jk} = e_j C_{jk} e_k, \quad j,k=1,\ldots,N \qquad [34]$$

where

$$e_j = \exp(\nu x_j + \beta p_j/2)\prod_{l\neq j} f(x_j - x_l)^{1/2} \qquad [35]$$

$$C_{jk} = \exp(-\nu(x_j + x_k))\frac{\sinh(\mathrm{i}\beta\nu g)}{\sinh \nu(x_j - x_k + \mathrm{i}\beta g)} \qquad [36]$$

In [35], $f(x)$ is the type II function given by [28]. The matrix $C$ arises from Cauchy's matrix $1/(w_j - z_k)$ via a suitable substitution, and Cauchy's identity

$$\det\left(\frac{1}{w_j - z_k}\right)_{j,k=1}^{N} = \prod_{j=1}^{N}\frac{1}{w_j - z_j}\prod_{1\le j<k\le N}\frac{(w_j - w_k)(z_j - z_k)}{(w_j - z_k)(z_j - w_k)} \qquad [37]$$

ensures that [34] yields the Hamiltonians $S_l$ of [31].

To conclude this section, we point out that the relation

$$L = \mathbf{1}_N + \beta L_{\mathrm{nr}} + O(\beta^2), \quad \beta\to 0 \qquad [38]$$

where $L_{\mathrm{nr}}$ denotes the nonrelativistic Lax matrix [11], can be used to deduce the involutivity of the nonrelativistic Hamiltonians from that of their relativistic counterparts.

## Quantum Relativistic CMS Systems

When the canonical quantization prescription [20] is applied to the classical Hamiltonians [31] with $f(x)=1$, one obtains commuting quantum operators whose action is exemplified by

$$\exp\left(-\frac{\hbar}{mc}\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}x}\right)F(x) = F\left(x - \mathrm{i}\frac{\hbar}{mc}\right) \qquad [39]$$

That is, the operators act on functions that have an analytic continuation in $x_1,\ldots,x_N$ from the real line $\mathbb{R}$ to a strip around $\mathbb{R}$ in the complex plane $\mathbb{C}$, whose width is at least $2\hbar/mc$.

Operators of this type are called analytic difference operators (henceforth AΔOs). The choice $f(x)=1$ amounts to the free case $g=0$ in [28]. For $g\neq 0$, however, the canonical quantization exemplified by [39] yields noncommuting AΔOs. Thus, the factor ordering following from [31] would entail that integrability breaks down at the quantum level.

As mentioned before, there is no general result guaranteeing that a different ordering that preserves integrability exists. Even so, this is true in the present case. Specifically, the function $f(x)$ can be factorized as $f_+(x)f_-(x)$, and then the AΔOs

$$S_{\pm l} = \sum_{\substack{I\subset\{1,\ldots,N\}\\ |I|=l}} \prod_{\substack{j\in I\\ k\notin I}} f_{\mp}(x_j - x_k) \times \exp\left(\mp \mathrm{i}\hbar\beta\sum_{j\in I}\partial_j\right)\prod_{\substack{j\in I\\ k\notin I}} f_{\pm}(x_j - x_k) \qquad [40]$$

do commute. In the elliptic case [27], this factorization involves the Weierstrass $\sigma$-function, and commutativity can be encoded in a sequence of functional equations satisfied by the $\sigma$-function. For the type I–III systems the pertinent factorization of [28] is given by

$$f_{\pm}(x) = \begin{cases} (1 \pm \mathrm{i}\beta g/x)^{1/2} & \text{(I)} \\ (\sinh\nu(x\pm \mathrm{i}\beta g)/\sinh\nu x)^{1/2} & \text{(II)} \\ (\sin\nu(x\pm \mathrm{i}\beta g)/\sin\nu x)^{1/2} & \text{(III)} \end{cases} \qquad [41]$$

(Here one has $g > 0$, and the choice of square root is such that $f_{\pm}(x)\to 1$ for $g\downarrow 0$.)

The nonrelativistic limit $c\to\infty$ of the quantum Hamiltonians [33] can be determined by expanding $S_1$ and $S_{-1}$ in a power series in $\beta = 1/mc$. In this way, one obtains once more [29], except for a small, but crucial change in $H_{\mathrm{nr}}$: instead of the coupling constant dependence $g^2$ in the potential energy, one gets $g(g-\hbar)$. The extra term arises from the action of the term linear in $\beta$ in the expansion of the exponential on the term linear in $\beta$ in the expansion of the functions $f_{\pm}(x)$.

From the perspective of the nonrelativistic quantum CMS systems, the change $g^2\to g(g-\hbar)$ appears ad hoc. As it transpires, however, the different

dependence on $g$ ensures that the eigenfunctions of $H_{\mathrm{nr}}$ depend on $g$ in a far simpler way. This will become clear shortly.

## Action-Angle Transforms and Duality

Under certain technical assumptions, any integrable system given by $N$ independent Poisson commuting Hamiltonians $S_1(x,p),\ldots,S_N(x,p)$ on a $2N$-dimensional phase space admits local canonical transformations to action-angle variables. Like the spectral theorem on the quantum level, this structural result is of limited practical value. Indeed, just as the spectral theorem yields no concrete information concerning eigenfunctions, bound-state energies, scattering, etc., associated with a given self-adjoint Hamiltonian, the Liouville–Arnold theorem only yields general insight in the type of motion that can occur and the geometric character of the local maps (in terms of invariant tori).

To fully comprehend ("solve") a given integrable system, one should render the associated action-angle map as concrete as possible. For the CMS type systems, a complete solution to this problem has only been achieved for the systems of type I–III. The motion in the trigonometric systems is oscillatory, so that a closeup via the action-angle transform involves extensive geometric constructions. By contrast, the type I and II systems are scattering systems, and here the action-angle map can be tied in with the classical wave maps (Møller transformations).

We now sketch some salient features of the action-angle maps for systems of type I and II. In all cases the map (denoted $\Phi$) is a canonical transformation from the phase space $\Omega$ (eqn [3]) with 2-form $\mathrm{d}x \wedge \mathrm{d}p$ to the phase space

$$\hat{\Omega} = \{(\hat{x},\hat{p}) \in \mathbb{R}^{2N} \mid \hat{p} \in G\} \qquad [42]$$

with 2-form $\mathrm{d}\hat{x} \wedge \mathrm{d}\hat{p}$. Thus, the actions $\hat{p}_1,\ldots,\hat{p}_N$ vary over $G$ given by [4] and the "angles" $\hat{x}_1,\ldots,\hat{x}_N$ over $\mathbb{R}$. Consequently, $\hat{\Omega}$ amounts to $\Omega$ with $x$ and $p$ interchanged.

As should be the case, the transformed commuting Hamiltonians

$$\hat{S}_k = S_k \cdot \Phi^{-1}, \quad k = 1,\ldots,N \qquad [43]$$

depend only on the action vector $\hat{p}$. To be specific, they arise from $S_k(x,p)$ by taking $g=0$ (no interaction, hence no $x$ dependence) and substituting $p \to \hat{p}$. Indeed, the actions $\hat{p}_k$ are the $t \to \infty$ limits of the momenta $p_k(t)$, where the $t$ dependence refers to the defining Hamiltonian of the system.

As it happens, the Lax matrix $L$ is of decisive importance to concretize the action-angle map $\Phi$, and in particular to reveal its hidden duality properties. The starting point is a commutation relation of $L(x,p)$ with a diagonal matrix $A(x)$ given by

$$A(x) = \mathrm{diag}(\mathrm{d}(x_1),\ldots,\mathrm{d}(x_N))$$
$$\mathrm{d}(y) = \begin{cases} y & \text{(I)} \\ \exp(2\nu y) & \text{(II)} \end{cases} \qquad [44]$$

Obviously, the symmetric functions $\check{D}_k(x)$ of $A(x)$ yield an integrable system on $\Omega$, so the Hamiltonians

$$D_k(\hat{x},\hat{p}) = (\check{D}_k \circ \Phi^{-1})(\hat{x},\hat{p}), \quad k = 1,\ldots,N \qquad [45]$$

yield an integrable system on the action-angle phase space $\hat{\Omega}$. The crux of the matter is now that these systems are familiar: they are also systems of type I and II!

To be specific, let us denote the dual systems just described by a caret, and the nonrelativistic/relativistic systems by a suffix nr/rel, resp. Then the duality properties alluded to are given by

$$\hat{\mathrm{I}}_{\mathrm{nr}} \simeq \mathrm{I}_{\mathrm{nr}}, \qquad \hat{\mathrm{I}}_{\mathrm{rel}} \simeq \mathrm{II}_{\mathrm{nr}}$$
$$\hat{\mathrm{II}}_{\mathrm{nr}} \simeq \mathrm{I}_{\mathrm{rel}}, \qquad \hat{\mathrm{II}}_{\mathrm{rel}} \simeq \mathrm{II}_{\mathrm{rel}} \qquad [46]$$

and $\Phi^{-1}$ serves as the action-angle map for the dual systems.

In order to sketch why this state of affairs holds true for the $\mathrm{II}_{\mathrm{rel}}$ system, recall that its Lax matrix is given by [34]. From this, one readily checks the commutation relation

$$\coth(\mathrm{i}\beta\nu g)[A,L] = 2e \otimes e - (AL + LA) \qquad [47]$$

Since $L$ is Hermitean, there exists a unitary $U$ diagonalizing $L$. It can now be shown that the spectrum of $L$ is positive and nondegenerate, and that $U^*e$ has nonzero components. The gauge ambiguity in $U$ (given by a permutation matrix and diagonal phase matrix) can, therefore, be fixed by requiring

$$U^*LU = \mathrm{diag}(\exp(\beta\hat{p}_1),\ldots,\exp(\beta\hat{p}_N)),$$
$$\hat{p}_N < \cdots < \hat{p}_1 \qquad [48]$$

$$(U^*e)_j > 0, \quad j = 1,\ldots,N \qquad [49]$$

A suitable reparametrization of $U^*e$ then yields the "angle" vector $\hat{x}$.

As a consequence, $U^*AU$ becomes a function of $\hat{x}$ and $\hat{p}$. In detail, one finds

$$(U^*AU)(\hat{x},\hat{p}) = L(\beta/2, 2\nu; \hat{p},\hat{x})^{\mathrm{T}} \qquad [50]$$

where $L(\nu,\beta;x,p)$ is given by [34] and $T$ denotes the transpose. Therefore, the "dual Lax matrix" $\hat{A} = U^*AU$ is essentially equal to $L$, explaining the self-duality $\hat{\mathrm{II}}_{\mathrm{rel}} \simeq \mathrm{II}_{\mathrm{rel}}$ announced above.

With the action-angle transform under explicit control, much more can be said about the solutions to Hamilton's equations for each of the commuting Hamiltonians, both as regards finite times and as regards long-time asymptotics (scattering). It is beyond the scope of this article to enlarge on this, but it is worth mentioning that the scattering reveals the solitonic character of the particles. Indeed, the set of asymptotic momenta $\hat{p}_1, \ldots, \hat{p}_N$ is conserved under the scattering and the asymptotic position shifts are factorized in terms of pair shifts. A quite remarkable feature of the type I systems is that the shifts actually vanish ("billiard ball" scattering).

## Eigenfunction Transforms and Duality

Both at the relativistic and at the nonrelativistic level the commuting quantum Hamiltonians $S_1, \ldots, S_N$ are formally self-adjoint on the Hilbert space $L^2(G_\kappa, \mathrm{d}x), \kappa = \mathrm{I}, \ldots, \mathrm{IV}$. Thus, it may be expected that it is possible to construct a unitary eigenfunction transform

$$\Phi_\kappa : L^2(G_\kappa, \mathrm{d}x) \to L^2(\hat{G}_\kappa, \mathrm{d}\mu_\kappa(p)), \quad \kappa = \mathrm{I}, \ldots, \mathrm{IV} \qquad [51]$$

diagonalizing $S_k$ as multiplication by a real-valued function $M_k(p)$. Here $\hat{G}_\kappa$ encodes the joint spectrum and $\mathrm{d}\mu_\kappa(p)$ is a suitable measure on $\hat{G}_\kappa$.

Obviously, this expectation is borne out in the free case $g = 0$. Then, $\Phi_\kappa$ is basically Fourier transformation, its kernel consisting of a sum of joint eigenfunctions

$$\exp(-\mathrm{i}x \cdot \sigma(p)/\hbar), \quad \sigma \in S_N \qquad [52]$$

with $\sigma$ ranging over the permutation group $S_N$. For $\kappa = \mathrm{I}, \mathrm{II}$, one can take $G_\kappa = \hat{G}_\kappa = G$ (eqn [4]) and $\mathrm{d}\mu_\kappa(p) = \mathrm{d}p$. Here one gets

$$M_k(p) = \sum_{1 \le i_1 < \cdots < i_k \le N} \begin{cases} p_{i_1} \cdots p_{i_k} \\ \exp(\beta p_{i_1}) \cdots \exp(\beta p_{i_k}) \end{cases} \qquad [53]$$

in the nonrelativistic and relativistic case, resp. For $\kappa = \mathrm{III}, \mathrm{IV}$, one needs to take into account periodic boundary conditions on the walls of $G_\kappa$, yielding a discrete joint spectrum after the center-of-mass motion is omitted. (With the above choices of $G_{\mathrm{III}}$ and $G_{\mathrm{IV}}$, cf. [8] and [9], the center-of-mass motion is a free motion along the line, so the total momentum still varies continuously.) Of course, the diagonalized $S_k$ are once more given by [53], since the kernel of $\Phi_\kappa$ consists of free boson states.

Taking next $g > 0$, the above expectation has not been confirmed for all of the eight regimes involved. This is not only because in some cases not even the existence of joint eigenfunctions has been shown, but also because in the relativistic case the unitarity of $\Phi_{\mathrm{II}}$ and $\Phi_{\mathrm{IV}}$ already breaks down for $N = 2$ when $g$ increases beyond a critical value, cf. [57] below. It is quite likely that this happens for $N > 2$ as well, but this is not readily apparent from the current fragmentary knowledge on joint eigenfunctions for $N > 2$.

The only two cases where the $g > 0$ joint eigenfunction transform is of an elementary nature are the $\mathrm{III}_{\mathrm{nr}}$ and $\mathrm{III}_{\mathrm{rel}}$ cases. Indeed, the joint eigenfunctions describing the internal motion are of the form

$$\psi_n(x) = W(x)^{1/2} P_n(x), \quad n \in \mathbb{N}^{N-1} \qquad [54]$$

Here,

$$W(x) = \prod_{1 \le j < k \le N} w(x_j - x_k) \qquad [55]$$

is a positive weight function on $G_{\mathrm{III}}$ and the $P_n(x)$ are multivariable orthogonal polynomials. Thus, $P_n(x)$ is a finite linear combination of the above free boson states, with $p$ in [52] a linear function of $n$. For the $\mathrm{III}_{\mathrm{nr}}$ case, these eigenfunctions were already found by Sutherland. (Here, the functions $P_n(x)$ amount to polynomials, often called the Jack polynomials, which arose in a statistics context.) The $\mathrm{III}_{\mathrm{rel}}$ polynomials may be viewed as the special $A_{N-1}$ case of Macdonald's orthogonal $q$-polynomials for arbitrary root systems, with

$$q = \exp(-2\hbar\beta\nu) \qquad [56]$$

(Note that $q$ converges to 1 both in the nonrelativistic limit $c \to \infty$ and in the classical limit $\hbar \to 0$.)

For the $\mathrm{II}_{\mathrm{nr}}$ case, the joint eigenfunctions were found and studied a couple of decades ago by Heckman and Opdam, yielding a multivariable hypergeometric transform. Indeed, for $N = 2$, the eigenfunctions can be expressed in terms of the hypergeometric function ${}_2F_1$, as has been known since the early days of quantum mechanics. Likewise, the arbitrary-$N$ $\mathrm{I}_{\mathrm{nr}}$ joint eigenfunction transform (studied in detail by de Jeu) can be viewed as a multivariable Hankel transform, the $N = 2$ kernel being essentially a Hankel function.

Much less is known concerning $\mathrm{IV}_{\mathrm{nr}}$ eigenfunctions, and *a fortiori* for the associated transform $\Phi_{\mathrm{IV}}$. For $N = 2$ the time-independent Schrödinger equation amounts to the Lamé equation. Hence, solutions are Lamé functions that can be studied in particular via Fuchs theory (regular singularities). A far more explicit form of the eigenfunctions dates back to work by Hermite in the nineteenth century. More precisely, provided the $g$ dependence of the

defining Hamiltonian is changed from $g^2$ to $g(g-\hbar)$ (a change already encountered above), Hermite's results apply to couplings $g=l\hbar$, $l=2,3,4,\ldots$ His eigenfunctions have a structure that is nowadays referred to as the Bethe ansatz. For the same $g$ values and arbitrary $N$, $H_{\mathrm{nr}}$ eigenfunctions of Bethe ansatz type were found and studied by Felder and Varchenko, but even for these $g$ values much remains to be done to achieve a complete understanding of the $\Phi_{\mathrm{IV}}$ transform.

A quite different approach, due to Komori and Takemura, does yield rather detailed information on $\Phi_{\mathrm{IV}}$ for arbitrary $g>0$. The key feature of their strategy is to view the $\mathrm{IV}_{\mathrm{nr}}$ case as a perturbation of the $\mathrm{III}_{\mathrm{nr}}$ case. This entails, however, that the validity of their results is restricted to large imaginary period of the $\wp$-function.

For the $\mathrm{IV}_{\mathrm{rel}}$ system, there are only rather complete results on $\Phi_{\mathrm{IV}}$ for $N=2$. More specifically, the eigenfunction transform is known to be unitary for

$$g \in [0, \hbar + \pi/\beta\nu] \qquad [57]$$

and a dense set in a corresponding parameter space. (For $g$ outside this interval, unitarity is violated.) The kernel of $\Phi_{\mathrm{IV}}$ involves eigenfunctions of Bethe ansatz structure. For $g=l\hbar$, $l=2,3,\ldots$ and arbitrary $N$, Bethe ansatz type $H_{\mathrm{rel}}$ eigenfunctions were found by Billey, generalizing the Felder–Varchenko results mentioned above.

It remains to discuss the $\mathrm{I}_{\mathrm{rel}}$ and $\mathrm{II}_{\mathrm{rel}}$ systems. To this end, we first recall the classical dualities [46]. It is natural to expect that these dualities are still present at the quantum level. For the $\mathrm{I}_{\mathrm{nr}}$ case, this is readily confirmed: the transform is indeed invariant under interchange of $x$ and $p$. In fact, the $N=2$ center-of-mass Hankel transform even depends only on $(x_1-x_2)(p_1-p_2)$, so that self-duality is manifest in this case.

More generally, for $N=2$ the expected dualities [46] are indeed present. The $\mathrm{II}_{\mathrm{nr}}$ ${}_2F_1$ transform satisfies the $\mathrm{I}_{\mathrm{rel}}$ analytic difference equation in $p_1-p_2$ due to the contiguous relations obeyed by ${}_2F_1$. The $\mathrm{II}_{\mathrm{rel}}$ transform is only unitary when $g$ is restricted by [57], and it is indeed self-dual in the same sense as the action-angle map (Ruijsenaars).

Turning finally to the case $N>2$, the multi-variable hypergeometric transform $\Phi_{\mathrm{II}}$ does have the expected duality property. More specifically, its inverse diagonalizes the commuting $\mathrm{I}_{\mathrm{rel}}$ A$\Delta$Os (Chalykh). For $\mathrm{II}_{\mathrm{rel}}$ with $N>2$ and $g=l\hbar$, $l=2,3,\ldots$, Chalykh also finds elementary joint eigenfunctions with the expected self-duality. To date, no Hilbert space results for the $N>2$ $\mathrm{II}_{\mathrm{rel}}$ case have been obtained.

To conclude, we mention that the soliton scattering behavior at the classical level is preserved under quantization in all cases where this can be checked. That is, no new momenta are created in the scattering process and the $S$-matrix is factorized as a product of pair $S$-matrices. Moreover, for the type I cases, the $S$-matrix is a momentum-independent (but $g$-dependent) phase, as a quantum analog of the classical billiard ball scattering.

*See also:* Bethe Ansatz; Classical *r*-Matrices, Lie Bialgebras, and Poisson Lie Groups; Functional Equations and Integrable Systems; Integrable Discrete Systems; Integrable Systems and Algebraic Geometry; Integrable Systems in Random Matrix Theory; Integrable Systems: Overview; Isochronous Systems; Ordinary Special Functions; *q*-Special Functions; Quantum Calogero–Moser Systems; Seiberg–Witten Theory; Separation of Variables for Differential Equations; Sine-Gordon Equation; Toda Lattices.

## Further Reading

Babelon O, Bernard D, and Talon M (2003) *Introduction to Classical Integrable Systems*. Cambridge: Cambridge University Press.

Calogero F (1971) Solution of the one-dimensional $N$-body problem with quadratic and/or inversely quadratic pair potentials. *Journal of Mathematical Physics* 12: 419–436.

Calogero F (2001) *Classical Many-Body Problems Amenable to Exact Treatments*. Berlin: Springer.

van Diejen JF and Vinet L (eds.) (2000) *Calogero–Moser–Sutherland Models*. Berlin: Springer.

Fock V, Gorsky A, Nekrasov N, and Rubtsov V (2000) Duality in integrable systems and gauge theories. *Journal of High Energy Physics* 7(28): 1–39.

Marshakov A (1999) *Seiberg–Witten Theory and Integrable Systems*. Singapore: World Scientific.

Moser J (1975) Three integrable Hamiltonian systems connected with isospectral deformations. *Advances in Mathematics* 16: 197–220.

Olshanetsky MA and Perelomov AM (1981) Classical integrable finite-dimensional systems related to Lie algebras. *Physics Reports* 71: 313–400.

Olshanetsky MA and Perelomov AM (1983) Quantum integrable systems related to Lie algebras. *Physics Reports* 94: 313–404.

Ruijsenaars SNM (1987) Complete integrability of relativistic Calogero–Moser systems and elliptic function identities. *Communications in Mathematical Physics* 110: 191–213.

Ruijsenaars SNM (1999) Systems of Calogero–Moser type. In: Semenoff G and Vinet L (eds.) *Proceedings of the 1994 Banff Summer School Particles and Fields*, pp. 251–352. Berlin: Springer.

Ruijsenaars SNM and Schneider H (1986) A new class of integrable systems and its relation to solitons. *Annals of Physics (NY)* 170: 370–405.

Sutherland B (1972) Exact results for a quantum many-body problem in one dimension II. *Physical Review A* 5: 1372–1376.

# Canonical General Relativity

**C Rovelli**, Université de la Méditerranée et Centre de Physique Théorique, Marseilles, France



## Introduction

Lagrangian formulations of general relativity (GR) were found by Hilbert and by Einstein himself, almost immediately after the discovery of the theory. The construction of Hamiltonian formulations of GR, on the other hand, has taken much longer, and has required decades of theoretical research.

The first such formulations were developed by Dirac and by Bergmann and his collaborators, in the 1950s. Their cumbersome formalism was simplified by the introduction of new variables: first by Arnowit, Deser, and Misner in the 1960s and then by Ashtekar in the 1980s. A large number of variants and improvements of these formalisms have been developed by many other authors. Most likely the process is not over, and there is still much to learn about the canonical formulation of GR.

A number of reasons motivate the study of canonical GR. In general, the canonical formalism can be an important step towards quantum theory; it allows the identification of the physical degrees of freedom, and the gauge-invariant states and observables of theory; and it is an important tool for analyzing formal aspects of the theory such as its Cauchy problem. All these issues are highly nontrivial, and present open problems, in GR.

In turn, the structural peculiarity and the conceptual novelty of GR have motivated re-analyses and extensions of the canonical formalism itself.

The following sections discuss the source of the peculiar difficulty of canonical GR, and summarize the formulations of the theory that are most commonly used.

## The Origin of the Difficulties

The reason for the complexity of the Hamiltonian formulation of GR is not so much in the intricacy of its nonlinear field equations; rather, it must be found in the conceptual novelty introduced by GR at the very foundation of the structure of mechanics.

The dynamical systems considered before GR can be formulated in terms of states evolving in time. One assumes that a time variable $t$ can be measured by a physical clock, and that certain observable quantities $A$ of the system can be measured at every instant of time. If we know the state $s$ of the system at some initial time, the theory predicts the value $A(t)$ of these quantities for any given later instant of time $t$. The space of the possible initial states $s$ is the phase space $\Gamma_0$. Observables are real functions on $\Gamma_0$. Infinitesimal time evolution can be represented as a vector field in $\Gamma_0$. This vector field is determined by the Hamiltonian, which is also a function on $\Gamma_0$. The integral lines $s(t)$ of this vector field determine the time evolution $A(t) = A(s(t))$ of the observables.

This conceptual structure is very general. It can be easily adapted to special-relativistic systems. However, it is not general enough for general-relativistic systems. GR is *not* formulated as the evolution of states and observables in a preferred time variable which can be measured by a physical clock. Rather, it is formulated as the *relative* (common) evolution of many observable quantities. Accordingly, in GR there is no quantity playing the same role as the conventional Hamiltonian. In fact, the canonical Hamiltonian density that one obtains from a Legendre transformation from a Lagrangian vanishes identically in GR.

The origin of this peculiar behavior of the theory is the following. The field equations are written as evolution equations in a time coordinate $t$. However, they are invariant under arbitrary changes of $t$. That is, if we replace $t$ with an arbitrary function $t' = t'(t)$ in a solution of the field equations, we obtain another solution. This underdetermination does not lead to a lack of predictivity in GR, because we do not interpret the variable $t$ as the measurable reading of a physical clock, as we do in non-general-relativistic theories. Rather, we interpret $t$ as a nonobservable mathematical parameter, void of physical significance. Accordingly, the notions of "state at a given time" and "value of an observable at a given time" are very unnatural in GR.

A Hamiltonian formulation of GR requires a version of the canonical formalism sufficiently general to deal with this broader notion of evolution. Generalizations of the Hamiltonian formalism have been developed by many authors, such as Dirac (see below), Souriau, Arnold, Witten, and many others. The first step in this direction was taken by Lagrange himself: Lagrange gave a time-independent interpretation of the phase space as the space $\Gamma$ of the solutions of the equations of motion (modulo gauges). As we shall see, however, consensus is still lacking on a fully satisfactory formalism.

## Dirac Theory of Constrained Systems

Dirac has developed a Hamiltonian theory for mechanical systems with constraints, precisely in

view of its application to GR. Dirac's theory is beautiful, finds vast applications, and it is still commonly taken as the basis to discuss Hamiltonian GR, although GR does not fit very naturally into Dirac's scheme. In the following, only the part of Dirac's theory relevant for GR is summarized.

Consider a Lagrangian system with Lagrangian variables $q^i$, with $i = 1, \ldots, n$. Call $v^i$ the corresponding velocities. Let the system be defined by the Lagrangian $L(q^i, v^i)$. The momenta are defined as functions of $q^i$ and $v^i$ by $p_i(q^i, v^i) = \partial L(q^i, v^i)/\partial v^i$. The canonical Hamiltonian $H(q^i, p_i) = v^i(q^i, p_i)p_i - L(q^i, v^i(q^i, p_i))$ (summation over repeated indices is understood) is obtained by inverting the function $p_i(q^i, v^i)$ and expressing the velocities as functions of the momenta $v^i(q^i, p_i)$. The phase space $\Gamma_0$ is the space of the variables $(q^i, p_i)$. Infinitesimal time evolution is given by the vector field $V = v^i(q^i, p_i)\partial/\partial q^i + f_i(q^i, p_i)\partial/\partial p_i$, where velocities and forces are given by the Hamilton equations $v^i = \partial H/\partial p_i$ and $f_i = -\partial H/\partial q^i$.

More formally, the 2-form $\omega = \mathrm{d}p_i \wedge \mathrm{d}q^i$ endows $\Gamma_0$ with a symplectic structure. In the presence of such a structure, every function $A$ determines a vector field $V_A$, defined by $i_{V_A}\omega = -\mathrm{d}A$. By integrating this field, we have a flow in $\Gamma_0$, called the flow generated by $A$. Time evolution is the flow generated by the Hamiltonian. Given two functions $A$ and $B$, their Poisson brackets are defined by the function $\{A, B\} = -V_A(B) = V_B(A)$. Therefore, the time evolution of an observable $A$ satisfies $\mathrm{d}A/\mathrm{d}t = \{A, H\}$. A dynamical system is completely characterized by the set $(\Gamma_0, \omega, \mathcal{A}, H)$, where $\mathcal{A} = (A_1, \ldots, A_N)$ is the ensemble of the observables.

A constrained system, in the sense of Dirac, is a system for which the image of the function $v^i \rightarrow p_i(q^i, v^i)$ is smaller than $R^n$. We can characterize the image $\mathcal{I}$ of the map $(q^i, v^i) \rightarrow (q^i, p_i)$ with a set of equations on $\Gamma_0$

$$C_\alpha(q^i, p_i) = 0 \qquad [1]$$

where $\alpha = 1, \ldots, m'$. These are called the primary constraints.

The "constraint surface" $C$ is the largest subspace of $\mathcal{I}$ which is preserved by time evolution. It can be characterized by adding additional constraints, still of the form (1), with $\alpha = m' + 1, \ldots, m$. These additional constraints, called secondary constraints, can be computed as the Poisson brackets of the primary constraints with the Hamiltonian (plus the Poisson brackets of these secondary constraints with the Hamiltonian, and so on, until the Poisson brackets of all the constraints with the Hamiltonian vanish on in $C$). We say that an equation holds weakly if it holds on $C$.

A constrained system is "first class" if the Poisson brackets of the constraints among themselves vanishes weakly. Maxwell theory and GR are first-class constrained systems. In a first-class constrained system, the constraints generate flows that preserve $C$ and foliate it into "orbits." The space of these orbits is called the physical phase space (see Figure 1).

This flow is interpreted as a "gauge" transformation, namely as a change of mathematical description of the same physical state. As first observed by Dirac, such interpretation is necessary if we demand a deterministic physical evolution, for the following reason. A first-class constrained system is a system in which the time evolution $q^i(t)$ of the Lagrangian variables is not completely determined by the equations of motion. (The relation between constraints and underdetermination of the evolution is simple to understand. In a Lagrangian system, the number of equations of motion is equal to the number of Lagrangian variables. If one of these equations is a constraint (between the initial velocities and initial coordinates), then one evolution equation is missing.) To recover a deterministic physical evolution, we must interpret two "mathematical" states that can evolve from the same initial data, as describing the same "physical" state. As shown by Dirac, the transformations generated by the constraints are precisely the ones that implement such an identification.

It follows that the physical states must be identified with the equivalence classes of the points of $C$ under the gauge transformations generated by the constraints, namely with the orbits of their flow. It is easy to show that (locally) there is a unique symplectic 2-form $\omega_{\mathrm{ph}}$ on $\Gamma_{\mathrm{ph}}$ such that its pullback to $C$ is equal to the pullback of $\omega$ to $C$ ($i_*\omega = \pi_*\omega_{\mathrm{ph}}$, see Figure 1). Physical observables $A_{\mathrm{ph}}$ are functions on $C$ that are gauge invariant, namely constant on

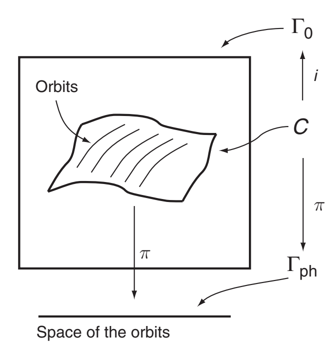



**Figure 1** The structure of a first-class constrained system. $\Gamma_0$: phase space, $C$: constraint surface, $\Gamma_{\mathrm{ph}}$: physical phase space; $i$: imbedding of $C$ in $\Gamma$; $\pi$ projection to orbit space (sending each point into its orbit).

the orbits. That is, they are functions on $\Gamma_{\mathrm{ph}}$. The Hamiltonian is a physical observable. The dynamical system $(\Gamma_{\mathrm{ph}}, \omega_{\mathrm{ph}}, \mathcal{A}_{\mathrm{ph}}, H)$, where $\mathcal{A}_{\mathrm{ph}}$ is the ensemble of the physical observables, is a complete description of the physical system, called the gauge-invariant formulation, with no more constraints or gauges.

For instance, the phase space of Maxwell theory is coordinatized by the Maxwell potential $A_\mu(\boldsymbol{x}), \mu = 0, 1, 2, 3$, and its conjugate momentum $E^\mu(\boldsymbol{x})$. Since the time derivative of $A_0$ does not appear in the Maxwell action, the primary constraint is

$$E^0(\boldsymbol{x}) = 0 \qquad [2]$$

The secondary constraint turns out to be the Gauss law,

$$\partial_a E^a(\boldsymbol{x}) = 0 \qquad [3]$$

where $a = 1, 2, 3$. The first generates arbitrary transformations of $A_0$, while the second generates the time-independent gauge transformations $\delta A_a(\boldsymbol{x}) = \partial_a \lambda(\boldsymbol{x})$. The pair $(A_0, \pi^0)$ can be dropped altogether, since it is formed by a pure gauge variable and a variable constrained to vanish. The (gauge-invariant) Hamiltonian is $H = 1/8\pi \int \mathrm{d}^3\boldsymbol{x}\,(E^a E_a + B^a B_a)$, where $B^a = \epsilon^{abc}\partial_b A_c$ is the magnetic field and $E^a$ is easily recognized as the electric field. $E^a$ and $B_a$ are the physical observables.

## General Structure of GR Constraints

GR fits into Dirac theory with a certain difficulty. Since the constraints are the generators of the gauge invariances, it is easy to determine their structure in GR. The gauge invariances of GR are given by the coordinate transformations $x^\mu \to x'^\mu = f^\mu(x)$, where $x = (\boldsymbol{x}, t)$. Accordingly, we have four primary constraints $\pi^\mu = 0$, analogous to [2], and four secondary constraints $C_\mu(\boldsymbol{x}) = 0$, analogous to [3]. These are usually separated into the three "momentum" constraints

$$C_a(\boldsymbol{x}) = 0 \qquad [4]$$

which generate fixed-time spatial coordinate transformations and the "Hamiltonian" constraint

$$C(\boldsymbol{x}) = 0 \qquad [5]$$

which generates changes in the $t$ coordinate.

The metric $g_{\mu\nu}(x)$ that represents the gravitational field in Einstein's original formulation has ten independent components per point. Each first-class constraint indicates that one Lagrangian variable is a gauge degree of freedom. The physical degrees of freedom of GR are therefore $(10 - 4 - 4) = 2$ per point. In the linearized theory, these are the two degrees of freedom that describe the two polarizations of a gravitational wave of given momentum. Formulations of GR in which there are additional gauge invariances (such as Cartan's tetrad formulation, see below) have, accordingly, more constraints.

Since the Hamiltonian generates evolution in the Lagrangian evolution parameter $t$, and since such evolution can be obtained as a gauge transformation, it follows that the Hamiltonian is a constraint in GR. The vanishing of the Hamiltonian is a characteristic feature of general-relativistic systems. The Hamiltonian structure of GR is therefore determined by its phase space and its constraints. The gauge-invariant formulation of the theory is given just by the set $(\Gamma_{\mathrm{ph}}, \omega_{\mathrm{ph}}, \mathcal{A}_{\mathrm{ph}})$ and no Hamiltonian. The physical interpretation of this structure is discussed in the last section.

## ADM Formalism

In Einstein's formulation, the Lagrangian variable of GR is the metric field $g_{\mu\nu}(\boldsymbol{x}, t)$ (here we use the signature $[-,+,+,+]$). Arnowit, Deser, and Misner have introduced the following change of variables:

$$q_{ab} = g_{ab}, \quad N = 1/\sqrt{-g^{00}}, \quad N^a = q^{ab} g_{a0} \qquad [6]$$

where $q^{ab}$ is the inverse of the three-dimensional metric $q_{ab}$, used henceforth to raise and lower space indices $a, b = 1, 2, 3$. This is equivalent to writing the invariant interval in the form

$$\mathrm{d}s^2 = -N^2\,\mathrm{d}t^2 + q_{ab}(\mathrm{d}x^a + N^a\,\mathrm{d}t)(\mathrm{d}x^b + N^b\,\mathrm{d}t)$$

These variables have an interesting geometric interpretation. Consider a family of spacelike ("ADM") surfaces $\Sigma_t$ defined by $t = \text{constant}$. $q_{ab}$ is the 3-metric induced on the surface. $N$ is called the "lapse" function and $N^a$ is called the "shift" function. Their geometrical interpretation is illustrated in Figure 2.

When written in terms of these variables, the action of GR takes the form

$$S[q_{ab}, N, N^a] = \int \mathrm{d}^4x \sqrt{q} N[R + k_{ab}k^{ab} - k^2]$$

where $q = \det q_{ab}$ and $R$ are the determinant and the Ricci scalar of the metric $q_{ab}$;

$$k_{ab} = \frac{1}{2N}(\partial_t q_{ab} - D_a N_b - D_b N_a)$$

is the extrinsic curvature of the constant time surface; and $D_a$ is the covariant derivative of $q_{ab}$. This action is independent of the time derivatives of

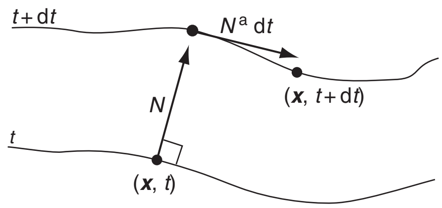



**Figure 2** The geometrical interpretation of the lapse $N(\boldsymbol{x}, t)$ and shift $N^a(\boldsymbol{x}, t)$ fields. Two ADM surfaces, defined by the values $t$ and $t + dt$, are displayed. $N(\boldsymbol{x}, t)dt$ is the proper length of the vector joining the two surfaces, normal to the first surface at $(\boldsymbol{x}, t)$. This is the proper time lapsed between the two surfaces for an observer at rest on the first surface at $(\boldsymbol{x}, t)$. The quantity $dx^a = N^a(\boldsymbol{x}, t)dt$ is the shift (the displacement) between the endpoint of this vector and the point $(\boldsymbol{x}, t + dt)$ having the same spacial coordinates as $(\boldsymbol{x}, t)$.

$N$ and $N^a$. The conjugate momenta $\pi$ and $\pi_a$ of these quantities are therefore the primary constraints and the pairs $(\pi, N)$ and $(\pi_a, N^a)$ can be taken out of the phase space as for the pair $(E^0, A_0)$ in the Maxwell example. We can therefore take the 3-metric $q_{ab}(\boldsymbol{x})$ and its conjugate momentum $p^{ab}(\boldsymbol{x})$ as the canonical variables of GR. The momentum is related to the "velocity" $\partial_t q_{ab}$, by

$$p^{ab} = \sqrt{q}(k^{ab} - kq^{ab})$$

where $k = k_{ab}q^{ab}$.

The secondary constraints [4] and [5] turn out to be

$$C_a = \sqrt{q}D_b\left(\frac{1}{\sqrt{q}}p^b{}_a\right) = 0 \qquad [7]$$

and

$$C = \frac{1}{\sqrt{q}}\left(p^{ab}p_{ab} - \frac{1}{2}p^2\right) - \sqrt{q}R = 0 \qquad [8]$$

where $p = p^{ab}q_{ab}$

If the two fields $q_{ab}(\boldsymbol{x}, t)$ and $p^{ab}(\boldsymbol{x}, t)$ satisfy the Hamilton equations

$$\frac{\partial q_{ab}(\boldsymbol{x}, t)}{\partial t} = \{q_{ab}(\boldsymbol{x}, t), H(t)\} \qquad [9]$$

$$\frac{\partial p^{ab}(\boldsymbol{x}, t)}{\partial t} = \{p^{ab}(\boldsymbol{x}, t), H(t)\} \qquad [10]$$

where

$$H(t) = \int d^3x\, N(\boldsymbol{x}, t)C[q_{ab}(\boldsymbol{x}, t),\, p^{ab}(\boldsymbol{x}, t)] + N^a(\boldsymbol{x}, t)C_a[q_{ab}(\boldsymbol{x}, t),\, p^{ab}(\boldsymbol{x}, t)]$$

with arbitrary functions $N(\boldsymbol{x}, t), N^a(\boldsymbol{x}, t)$, then the metric $g_{\mu\nu}(\boldsymbol{x}, t)$, defined from $q_{ab}, N, N^a$ by eqn [6], is the general solution of the vacuum Einstein equation $\text{Ricci}[g] = 0$. Therefore, these equations provide a Hamiltonian form of the Einstein field equation.

## Tetrad Formalism

The tetrad formalism, developed by Cartan, Weyl, and Schwinger, has definite advantages with respect to the metric formalism. It allows the coupling of fermion fields to GR and is, therefore, needed to couple the standard model to GR. In the tetrad formalism, the gravitational field is represented by four covariant fields $e^I_\mu(x)$, where $I, J, \ldots = 0, 1, 2, 3$ are flat Lorentz indices raised and lowered with the Minkowski metric $\eta_{IJ} = \text{diag}[-1, +1, +1, +1]$. The relation with the metric formalism is given by

$$g_{\mu\nu} = \eta_{IJ}e^I_\mu e^J_\nu$$

In this formulation, GR has an additional local SO(3,1) gauge invariance, given by local Lorentz transformations on the $I$ indices. The corresponding canonical formalism is usually defined in a gauge in which $e^i_0 = 0$, where $i, j, \ldots = 1, 2, 3$ are flat three-dimensional indices raised and lowered with the $\delta_{ij} = \text{diag}[+1, +1, +1]$. In this gauge, the Lorentz group is reduced to the local SO(3) group of spatial transformations, and the ADM variable are defined by

$$e^I_\mu = \begin{pmatrix} N & N^i \\ 0 & e^i_a \end{pmatrix} \qquad [11]$$

where $N^i = e^i_a N^a$. This is equivalent to writing the invariant interval in the form

$$ds^2 = -N^2\, dt^2 + (e_{ai}\, dx^a + N_i\, dt)\left(e^i_b\, dx^b + N^i\, dt\right)$$

The reduced canonical variables can be taken to be the field $e^i_a(\boldsymbol{x})$ that represents the "triad" of the ADM surface, and its conjugate momentum $p^a_i(\boldsymbol{x})$. Their relation with the three-dimensional metric variables is given by transforming internal indices into tangent indices with the triad field $e^i_a$ and its inverse $e^a_i$. In particular,

$$q_{ab} = \delta_{ij}e^i_a e^j_b \qquad [12]$$

$$p^{ab} = e^{bi}p^a_i \qquad [13]$$

Also, for later reference,

$$k^i_a \equiv e^{ib}k_{ab} = \frac{2}{\det e}\left(p^i_a - \tfrac{1}{2}e^i_a p\right) \qquad [14]$$

where $p = e^i_a p^a_i$.

The momentum and Hamiltonian constraints are the same as in the ADM formulation, with $q_{ab}$ and $p^{ab}$ expressed in terms of the triad variables. The additional constraint that generates the internal rotations is

$$G_i = \epsilon_{ijk}e^j_a\, p^{ak} = 0 \qquad [15]$$

## Ashtekar Formalism

The Ashtekar formalism simplifies the form of the constraints and casts GR in a form having the same kinematics as Yang–Mills theory. With its variants, it is widely used in nonperturbative quantum gravity, in particular in the loop formulation (*see* Loop Quantum Gravity). It can be obtained from the tetrad canonical formalism by the canonical transformation

$$A^i_a = \tfrac{1}{2}\epsilon^i_{jk}\omega^{jk}_a + \mathrm{i}k^i_a \qquad [16]$$

$$E^a_i = \det e\, e^a_i \qquad [17]$$

where $\omega^{ij} = \omega^{ij}_a\, \mathrm{d}x^a$ is the (torsion-free) spin connection of the triad 1-form field $e^i = e^i_a\, \mathrm{d}x^a$, determined by the Cartan equation

$$\mathrm{d}e^i + \omega^i_k \wedge e^k = 0$$

The "electric" field $E$ is real, while the Sen–Ashtekar connection $A^i = A^i_a\, \mathrm{d}x^a$ is complex and satisfies the reality condition

$$A^i + \overline{A^i} = 2\Gamma^i[e] \qquad [18]$$

The connection $A^i$ has a simple geometrical interpretation. It is the pullback $A_{ai} = \omega^{(+)}_{a0i}$ on the $t = 0$ ADM surface of the self-dual part

$$\omega^{(+)}_{IJ} = \frac{1}{2}\left(\omega_{IJ} - \frac{\mathrm{i}}{2}\,\epsilon_{IJ}{}^{KL}\omega_{KL}\right)$$

of the four-dimensional torsion free spin connection $\omega^{IJ}_\mu$ determined by the tetrad field $e^I_\mu$.

In terms of these fields, the constraint equations can be written in the form

$$G_i = D_a E^a_i = 0 \qquad [19]$$

$$C_a = F^i_{ab} E^a_i = 0 \qquad [20]$$

$$C = \epsilon_{ijk} F^i_{ab} E^{ja} E^{kb} = 0 \qquad [21]$$

where $D_a$ is the covariant derivative and $F_{ab}$ is the curvature defined by the connection $A$. The first of these constraints is the nonabelian version of the Gauss law [3]: it is the gauge constraint of Yang–Mills theory. The constraints are polynomial in the canonical variables.

These equations are often written using a basis $\tau_i$ in the su(2) Lie algebra, and defining the su(2) connection $A = A^i\tau_i$ and the su(2)-valued vector field $E^a = E^{ai}\tau_i$. In terms of these fields the constraints can be written in the form

$$G = D_a E^a = 0$$

$$C_a = \mathrm{tr}[F_{ab}E^a] = 0$$

$$C = \mathrm{tr}[F_{ab}E^a E^b] = 0$$

where the trace is on su(2).

A variant of this formalism commonly used in quantum gravity is obtained by replacing [16] with the Barbero connection

$$A^i_a = \tfrac{1}{2}\,\epsilon^i_{jk}\,\omega^{jk}_a + \gamma\, k^i_a \qquad [22]$$

where $\gamma$ is an arbitrary complex number, called the Immirzi parameter. In terms of this connection, [21] is replaced by

$$C = \epsilon_{ijk} F^i_{ab} E^{ja} E^{kb} + \frac{1+\gamma^2}{4}\det e(k_{ab}k^{ab} - k^2) = 0$$

where $e^i_a$ and $k_{ab}$ are given as function of $E$ and $A$ by [22] and [17]. The choice $\gamma = 1$, with the constraint [19]–[21], gives the canonical formulation of Euclidean GR.

All the formulations described extend readily to matter couplings. The structure of the constraints remains the same – with additional constraints corresponding to matter gauge invariances, if any. The GR constraints are modified by the addition of matter terms. In particular, the Hamiltonian constraint $C$ and the momentum constraint $C_a$ are modified by the addition of terms determined by the energy density and the momentum density of the matter, respectively. In the Ashtekar formulation, a fermion field modifies the Gauss law constraint by the addition of a torsion term.

## Evolution

In the gauge-invariant canonical structure of GR, there is no explicit time flow generated by a Hamiltonian. If the formalism is utilized just in order to express the Einstein equation in first-order canonical form, this is not a difficulty, because evolution in the coordinate time is generated by the constraints. On the other hand, if we are interested in understanding the structure of states, observables, and evolution of GR, the situation appears to be puzzling. An additional complication arises from the fact that virtually no gauge-invariant observable $A_{\mathrm{ph}}$ is known explicitly as a function on the phase space. These issues become especially relevant when the canonical formalism is taken as a starting point for quantization. How is physical evolution represented in canonical GR?

The first relevant observation is that the gauge-invariant phase space $\Gamma_{\mathrm{ph}}$ is better understood as a phase space in the sense of Lagrange: namely as the space $\Gamma$ of the solutions of the equations of motion modulo gauges, rather than a space of instantaneous states. Recall that in GR the notion of "instantaneous state" is rather unnatural.

In the ADM formulation, for instance, an orbit on the constraint surface of GR can be understood as the ensemble of all possible values that the variables

$(q_{ab}(\boldsymbol{x}), p^{ab}(\boldsymbol{x}))$ can take on arbitrary spacelike ADM surfaces embedded in a given solution of the Einstein equation. Motion along the orbit (which has dimension $4 \times \infty^3$) corresponds to arbitrary deformations of the surface.

Physical applications of classical GR deal with relations between "partial observables." A partial observable is any variable physical quantity that can be measured, even if its value cannot be determined from the knowledge of the physical state. An example of partial observable in nonrelativistic mechanics is given precisely by the nonrelativistic time $t$. Partial observables are represented in GR as functions on $\Gamma_0$. A physical state in $\Gamma_{\text{ph}}$ determines an orbit in $C$, and therefore a set of relations between partial observables (see Figure 1). That is, it determines the possible values that the partial observables can take "when" and "where" other partial observables have given values. All physical predictions of classical GR can be expressed in this form.

One of the partial observables can be selected to play the role of a physical clock time, and evolution can be expressed in terms of such clock time. In general, it is difficult – if not impossible – to find a clock time observable in terms of which evolution is a proper conventional Hamiltonian evolution. Matter couplings partially simplify the task. For instance, if the motion of planet Earth is coupled to GR, then proper time along this motion from a significative event on Earth, which is a partial observable, can be a convenient clock time. In pure gravity, the "York time" defined as the trace of the extrinsic curvature $T_{\text{Y}} = k$, on ADM surfaces where $k$ is spatially constant, has been extensively and effectively used as a clock time in formal analysis of the theory. A Hamiltonian that generates evolution in a given clock time $T$ can be formally obtained by solving the Hamiltonian constraint with respect to a momentum $P_T$ conjugate to $T$. Such "reparametrizations" of the relative evolution of the partial observables can be useful to analyze equations and to help intuition, but they are by no means necessary to have a well-defined interpretation of the theory.

Another possibility to introduce a preferred time flow is to consider asymptotically flat solutions of the field equations. In this case, one can define a nonvanishing Hamiltonian, given by a boundary integral at spacial infinity. This Hamiltonian generates evolution in an asymptotic Minkowski time. This choice is convenient for describing observations performed from a large distance on isolated gravitational systems. Many general-relativistic physical observations do not belong to this category.

Various other techniques to define a fully generally covariant canonical formalism have been explored. Among these: definitions of the physical symplectic structure directly on the space of the solutions of the field equations; generalization of the initial and final surfaces to boundaries of compact spacetime regions; construction of "evolving constants of motion," namely families of gauge-invariant observables depending on a clock time parameter; multisymplectic formalisms that treats space and time derivatives on a more equal footing; and others. Many of these techniques are attempts to overcome the unequal way in which time and space dependence are treated in the conventional Hamiltonian formalism.

GR has deeply modified our understanding of space and time. An extension of the canonical formalism of mechanics, compatible with such a modification, is needed, but consensus on the way (or even the possibility) of formulating a fully satisfactory general-relativistic extension of Hamiltonian mechanics is still lacking.

*See also:* Asymptotic Structure and Conformal Infinity; Constrained Systems; General Relativity: Overview; Loop Quantum Gravity; Quantum Cosmology; Quantum Geometry and its Applications; Spin Foams; Wheeler–De Witt Theory.

## Further Reading


Arnowitt R, Deser S, and Misner CW (1962) The dynamics of general relativity. In: Witten L (ed.) *Gravitation: An Introduction to Current Research*, p. 227. New York: Wiley.

Ashtekar A (1991) *Non-Perturbative Canonical Gravity*. Singapore: World Scientific.

Bergmann P (1989) The canonical formulation of general relativistic theories: the early years, 1930–1959. In: Howard D and Stachel J (eds.) *Einstein and the History of General Relativity*. Boston: Birkhäuser.

Dirac PAM (1950) Generalized Hamiltonian dynamics. *Canadian Journal of Mathematical Physics* 2: 129–148.

Dirac PAM (1958) The theory of gravitation in Hamiltonian form. *Proceedings of the Royal Society of London, Series A* 246: 333.

Dirac PAM (1964) *Lectures on Quantum Mechanics*. New York: Belfer Graduate School of Science, Yeshiva University.

Gotay MJ, Isenberg J, Marsden JE, and Montgomery R (1998) Momentum maps and classical relativistic fields. Part 1: Covariant field theory. Archives: physics/9801019.

Hanson A, Regge T, and Teitelboim C (1976) *Constrained Hamiltonian Systems*. Rome: Academia Nazionale dei Lincei.

Henneaux M and Teitelboim C (1972) *Quantization of Gauge Systems*. Princeton: Princeton University Press.

Isham CJ (1993) Canonical quantum gravity and the problem of time. In: Ibort LA and Rodriguez MA (eds.) *Recent Problems in Mathematical Physics, Salamanca*, Dordrecht: Kluwer Academic.

Lagrange JL (1808) *Mémories de la première classe des sciences mathematiques et physiques*. Paris: Institute de France.

Rovelli C (2004) *Quantum Gravity*. Cambridge: Cambridge University Press.

Souriau JM (1969) *Structure des Systemes Dynamics*. Paris: Dunod.

# Capacities Enhanced by Entanglement

**P Hayden**, McGill University, Montreal, QC, Canada



## Introduction

Shared entanglement between a sender and receiver can significantly improve the usefulness of a quantum channel for the communication of either classical or quantum data. Superdense coding and teleportation provide the most well-known examples of this improvement; free entanglement doubles the classical capacity of a noiseless quantum channel and makes it possible for a noiseless classical channel to send quantum data. In fact, the entanglement-assisted classical and quantum capacities of a quantum channel are in many senses simpler and better behaved than their unassisted counterparts (Holevo 1998, Schumacher and Westmoreland 1997, Devetak 2005). Most importantly, these capacities can be calculated using simple formulas and finite optimization procedures (Bennett *et al.* 1999, 2002). No such finite procedure is known for either of the unassisted capacities. Moreover, the entanglement-assisted classical and quantum capacities are related by a simple factor of 2. The unassisted capacities, in contrast, have completely different formulas. In fact, the simple factor of 2 generalizes to a statement known as the quantum reverse Shannon theorem, which governs the rate at which one quantum channel can simulate another (Bennett *et al.* 2005). The answer is given by the ratio of the entanglement-assisted capacities.

### Notation

Quantum systems will be denoted by $A$, $B$, and so on as well as their variants such as $A'$ and $\hat{A}$. The choice of letter will generally indicate which party holds a given system, with $A$ reserved for the sender, Alice, and $B$ for the receiver, Bob. Given a quantum system $C$, $C^{\otimes n}$ will often be written as $C^n$. These symbols will be used to denote both the Hilbert space of the quantum system and the set of density operators on that system. Thus, a quantum channel $\mathcal{N}: A' \to B$ refers to a trace-preserving, completely positive (TPCP) map from the operators on the Hilbert space of $A'$ to those of $B$. $\mathrm{id}^C$ refers to the identity channel on $C$. The map $\mathcal{N} \otimes \mathrm{id}^C$ will frequently be abbreviated to $\mathcal{N}$ in order to simplify long expressions. Likewise, the density operator $|\varphi\rangle\langle\varphi|$ of a pure quantum state $|\varphi\rangle$ will be abbreviated to $\varphi$. $\pi^C$ will refer to the maximally mixed state on $C$ and $\pi_d$ to the maximally mixed state on a specified $d$-dimensional quantum system.

For a given quantum state $\varphi^{AB}$ on the composite system $AB$, $\varphi^A = \mathrm{tr}_B\, \varphi^{AB}$ and

$$H(A)_\varphi = H(\varphi^A) = -\mathrm{tr}(\varphi^A \log_2 \varphi^A) \qquad [1]$$

is the von Neumann entropy of $\varphi^A$, while

$$H(A|B)_\varphi = -I_c(A\rangle B) = H(AB)_\varphi - H(B)_\varphi \qquad [2]$$

is its conditional entropy and

$$I(A;B)_\varphi = H(A)_\varphi + H(B)_\varphi - H(AB)_\varphi \qquad [3]$$

its mutual information.

## Entanglement-Assisted Classical and Quantum Capacities

The entanglement-assisted classical capacity of a quantum channel $\mathcal{N}: A' \to B$ is the optimal rate at which classical information can be communicated through the channel while in addition making use of an unlimited number of maximally entangled states.

The formal definition proceeds as follows. Alice and Bob are assumed to share $nS$ ebits in the form of a maximally entangled state $|\Phi\rangle^{\tilde{A}\tilde{B}}$ of Schmidt rank $2^{nS}$. Conditioned on her message $m \in \{1, 2, \ldots, 2^{nR}\}$, Alice will apply an encoding operation $\mathcal{E}_m : \tilde{A} \to A'^n$. Bob's decoding is given by a POVM $\{\Lambda_m\}_{m=1}^{2^{nR}}$ on the composite system $\tilde{B}B^n$. The procedure is said to have maximum probability of error $\epsilon$ if

$$\max_m \mathrm{tr}\left[\Lambda_m(\mathcal{N}^{\otimes n} \circ \mathcal{E}_m)(\Phi)\right] \geq 1 - \epsilon \qquad [4]$$

These elements, illustrated in Figure 1, consisting of the shared entanglement, as well as the encoding and decoding operations meeting the criterion of eqn [4], are called a $(2^{nR}, 2^{nS}, n, \epsilon)$ entanglement-assisted classical code for the channel $\mathcal{N}$. A rate $R$ is said to be achievable if there exists a choice of $S \geq 0$ and a sequence of entanglement-assisted classical codes $(2^{nR}, 2^{nS}, n, \epsilon_n)$ with $\epsilon_n \to 0$. The entanglement-assisted

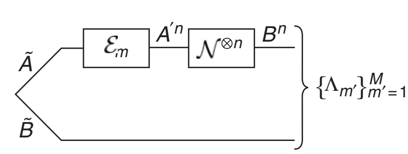



**Figure 1** Circuit representation of the elements of an entanglement-assisted classical code for the channel $\mathcal{N}$. Alice encodes message $m$ by applying the operation $\mathcal{E}_m$ to her half of the shared entanglement. Bob decodes by applying the POVM $\{\Lambda_{m'}\}$ on the output of the channel and his half of the shared entanglement.

classical capacity $C_E(\mathcal{N})$ of $\mathcal{N}$ is defined to be the supremum over all achievable rates.

**Theorem 1** (Bennett *et al.* 1999, 2002). *The entanglement-assisted classical capacity $C_E$ of a quantum channel $\mathcal{N}: A' \to B$ is given by*

$$C_E(\mathcal{N}) = \max_{\sigma} I(A;B)_\sigma \qquad [5]$$

*where the maximization is over states $\sigma^{AB} = \mathcal{N}(\varphi^{AA'})$ arising from the channel by acting on the $A'$ half of any pure state $|\varphi\rangle^{AA'}$.*

The theorem bears a strong formal resemblance to Shannon's noisy coding theorem for the classical capacity of a classical noisy channel. There the capacity formula is also given by an optimization of the mutual information, but over joint distributions between the input and output alphabets arising from the action of the channel. Such a joint distribution cannot exist in general for a quantum channel because the no-cloning theorem excludes the possibility of the input and output existing simultaneously. Equation [5] instead refers to a natural substitute for the joint input–output distribution: a quantum state arising from the quantum channel acting on half of an entangled pure state.

Another point worth stressing is that, unlike the known formulas for the unassisted classical and quantum capacities of a quantum channel, eqn [5] refers to only a single use of $\mathcal{N}$ instead of the limit of many uses, $\mathcal{N}^{\otimes n}$. The formula can therefore readily be used to evaluate $C_E$ for any channel of interest.

Consider, for example, the $d$-dimensional depolarizing channel

$$\mathcal{D}_p(\rho) = (1-p)\rho + p\pi_d \qquad [6]$$

that with probability $p$ completely randomizes the input but otherwise leaves the input invariant. For such channels, the maximum is achieved by choosing a maximally entangled state for $|\varphi\rangle^{AA'}$, yielding

$$C_E(\mathcal{D}_p) = 2\log_2 d - h_{d^2}\left(1 - p\frac{d^2-1}{d^2}\right) \qquad [7]$$

where for any $0 \leq q \leq 1$ and integer $r \geq 1$,

$$h_r(q) = -q\log_2 q - (1-q) \times \log_2\left(\frac{1-q}{r-1}\right) \qquad [8]$$

is the Shannon entropy of the distribution $(q, (1-q)/(r-1), \ldots, (1-q)/(r-1))$.

Entanglement assistance also simplifies the relationship between the classical and quantum capacities of a channel. Proceeding as before to formally define the quantum capacity, Alice and Bob are again assumed to share a maximally entangled state $|\Phi\rangle^{\tilde{A}\tilde{B}}$ of Schmidt rank $2^{nS}$. Alice's encoding operation will be a TPCP map $\mathcal{E}: \hat{A}\tilde{A} \to A'^n$ acting on an input system $\hat{A}$ and her half of the shared entanglement, $\tilde{A}$. Bob's decoding will likewise be a TPCP map $\mathcal{D}: \tilde{B}B^n \to \hat{B}$ acting on the output of the channel, $B^n$, and his half of the shared entanglement, $\tilde{B}$. $\hat{A}$ and $\hat{B}$ are assumed to be isomorphic quantum systems of some fixed dimension $2^{nQ}$. The procedure is said to have subspace fidelity $1-\epsilon$ if

$$^{\hat{B}}\langle\varphi|\left(\mathcal{D}\circ\mathcal{N}^{\otimes n}\circ\mathcal{E}\right)\left(\Phi^{\tilde{A}\tilde{B}}\otimes\varphi^{\hat{A}}\right)|\varphi\rangle^{\hat{B}} \geq 1-\epsilon \qquad [9]$$

for all $|\varphi\rangle^{\hat{A}} \in \hat{A}$. These elements, illustrated in Figure 2, are together called a $(2^{nQ}, 2^{nS}, n, \epsilon)$ entanglement-assisted quantum code for the channel $\mathcal{N}$. A rate $Q$ is said to be achievable if there exists a choice of $S \geq 0$ and a sequence of entanglement-assisted quantum codes $(2^{nR}, 2^{nS}, n, \epsilon_n)$ with $\epsilon_n \to 0$. The entanglement-assisted quantum capacity $Q_E(\mathcal{N})$ of $\mathcal{N}$ is defined to be the supremum over all achievable rates.

There is considerable freedom in the definition of the entanglement-assisted quantum capacity. It could, for example, be defined as the largest amount of maximal entanglement that can be generated using the channel, minus the entanglement consumed during the protocol itself. Alternatively, the fidelity criterion eqn [9] could be strengthened to require that $\mathcal{D}\circ\mathcal{N}^{\otimes n}\circ\mathcal{E}$ preserve not only pure states on $\hat{A}$ but any entanglement between $\hat{A}$ and a reference system. All of these variants yield the same capacity formula:

$$Q_E(\mathcal{N}) = \tfrac{1}{2}C_E(\mathcal{N}) \qquad [10]$$

This equivalence is a direct consequence of the existence of the teleportation and superdense coding protocols. When maximal entanglement is available, teleportation converts the ability to send classical data into the ability to send quantum data at half the classical rate. Conversely, by consuming

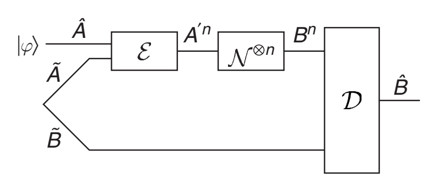



**Figure 2** Circuit representation of the elements of an entanglement-assisted quantum code for the channel $\mathcal{N}$. $\mathcal{E}$ is Alice's encoding operation, which acts on both her input state and her half of the shared entanglement. Bob decodes using a quantum operation $\mathcal{D}$ acting on the output of the channel and his half of the shared entanglement.

maximal entanglement, superdense coding converts the ability to send quantum data into the ability to send classical data at double the quantum rate.

## Sketch of Proof

The proof of a capacity theorem can usually be broken into two parts, achievability and optimality. The achievability part demonstrates the existence of a sequence of codes reaching the prescribed rate while the optimality part shows that it is impossible to do better.

The main idea in the achievability proof can be understood by studying the special case where $\varphi^{A'} = \pi^{A'}$. Let $d_{A'} = \dim A'$ and $\{U_j\}_{j=1}^{d_{A'}^{2n}}$ be a set of Weyl operators for $A'^n$. The relevant property of these operators is that averaging over them implements the constant map: for all density operators $\rho$,

$$\frac{1}{d_{A'}^{2n}} \sum_{j=1}^{d_{A'}^{2n}} U_j \rho U_j^\dagger = \pi^{A'^n} \qquad [11]$$

Consider the state $\sigma_j$ that arises if Alice acts with $U_j$ on the $A'^n$ half of a rank-$d_{A'}^n$ maximally entangled state $|\varphi\rangle^{AA'^n}$ and then sends the $A'^n$ half of the resulting state through $\mathcal{N}$. (Note that here $A'^n$ also plays the role of $\tilde{A}$.) The entropy of the resulting state is

$$H(\sigma_j) = H\Big(\mathcal{N}((U_j \otimes I_{\tilde{B}})\varphi(U_j^\dagger \otimes I_{\tilde{B}}))\Big) \qquad [12]$$

$$= H(\mathcal{N}(\varphi)) \qquad [13]$$

since $U_j$ does not change the local density operator on $A'^n$.

On the other hand, if Alice selects a value of $j$ from the uniform distribution, then the resulting average input state to the channel will be

$$\pi^{A'^n} \otimes \pi^A = \varphi^{A'^n} \otimes \varphi^A \qquad [14]$$

and the corresponding average output state will be $\mathcal{N}(\varphi^{A'^n}) \otimes \varphi^A$, which has entropy

$$H(\mathcal{N}(\varphi^{A'^n})) + H(\varphi^A) \qquad [15]$$

Therefore, the Holevo quantity of the ensemble of output states, defined as the entropy of the average state minus the average of the entropies of the individual output states, will be equal to

$$H(\varphi^A) + H\Big(\mathcal{N}(\varphi^{A'^n})\Big) - H\Big(\mathcal{N}(\varphi^{AA'^n})\Big) \qquad [16]$$

This is precisely the quantity $I(A;B)_\sigma$ for the state $\mathcal{N}(\varphi^{AA'^n})$ since the channel $\mathcal{N}$ transforms the $A'^n$ system into $B$. Moreover, if Bob is given the $A$ part of the maximally entangled state, then this is the Holevo quantity of an ensemble of states that can be produced by Alice acting on half of a shared entangled state and then sending her half through the channel. Invoking the Holevo–Schumacher–Westmoreland (HSW) theorem for the classical capacity (Holevo 1998, Schumacher and Westmoreland 1997) therefore completes the proof; using coding, the Holevo quantity is an achievable communication rate.

The proof that eqn [5] is optimal involves a series of entropy manipulations similar to the optimality proofs for the unassisted classical and quantum capacities. From the point of view of quantum information, the truly unusual part of the proof is the demonstration that it is unnecessary to consider multiple copies of $\mathcal{N}$ (Cerf and Adami 1997). Specifically, let

$$f(\mathcal{N}) = \max_\sigma I(A;B)_\sigma \qquad [17]$$

where the maximization is defined as in Theorem 1. Techniques analogous to those used for the unassisted capacities yield the upper bound

$$C_E(\mathcal{N}) \le \lim_{n\to\infty} \frac{1}{n} f(\mathcal{N}^{\otimes n}) \qquad [18]$$

Unlike the unassisted case, however, a relatively easy argument shows that

$$f(\mathcal{N}_1 \otimes \mathcal{N}_2) = f(\mathcal{N}_1) + f(\mathcal{N}_2) \qquad [19]$$

(The analogous statement is an important conjecture for the classical capacity and is known to be false for the quantum capacity (DiVincenzo *et al.* 1998).) As a result, $C_E(\mathcal{N}) \le f(\mathcal{N})$, which is the optimality part of Theorem 1.

To see the origin of eqn [19], it will be helpful to invoke Stinespring's theorem to write $\mathcal{N}_j = \mathrm{tr}_{E_j} \mathcal{U}_j^{B_j E_j}$, where $\mathcal{U}_j : A'_j \to B_j E_j$ is an isometry. Fix a state $|\varphi\rangle^{AA'_1A'_2}$ and let $\sigma = (\mathcal{U}_1 \otimes \mathcal{U}_2)(\varphi)$. Equation [19] follows from the fact that

$$\begin{aligned} I(A;B_1B_2)_\sigma &\le I(AB_2E_2;B_1)_\sigma \\ &\quad + I(AB_1E_1;B_2)_\sigma \end{aligned} \qquad [20]$$

Simply redefining $A$ to be $AB_2E_2$ shows that the first term of the right-hand side is upper bounded by $f(\mathcal{N}_1)$. The second term, likewise, is upper bounded by $f(\mathcal{N}_2)$. Equation [20] is itself equivalent to the inequality

$$\begin{aligned} H(B_1B_2|E_1E_2)_\sigma &+ H(B_1B_2)_\sigma \\ &\le H(B_1|E_1)_\sigma + H(B_2|E_2)_\sigma \\ &\quad + H(B_1)_\sigma + H(B_2)_\sigma \end{aligned} \qquad [21]$$

The inequality $H(B_1B_2)_\sigma \le H(B_1)_\sigma + H(B_2)_\sigma$ holds by the subadditivity of the von Neumann entropy.

Repeated applications of the strong subadditivity inequality, moreover, lead to the inequality

$$H(B_1B_2|E_1E_2)_\sigma \leq H(B_1|E_1)_\sigma + H(B_2|E_2)_\sigma \qquad [22]$$

Together, they prove eqn [20] and, thence, eqn [19]. The intuitive meaning of this "single-letterization" is unclear, but regardless, it is interesting to note that the proof involved invoking a pair of purifying environment systems, $E_1$ and $E_2$, and studying the entropy relationships between the true outputs of the channel and the environment's share.

## The Quantum Reverse Shannon Theorem

A strong argument can be made that the entanglement-assisted capacity of a quantum channel is the most important capacity of that channel and that all the other capacities are, in some sense, of less significance. The fact that it is unnecessary to distinguish between the classical and quantum entanglement-assisted capacities because they are related by a factor of 2 is a hint in that direction, as is the simple, single-letter formula for $C_E(\mathcal{N})$.

A more general argument can be made by considering the problem of having one channel simulate another. Indeed, the quantum capacity of a quantum channel is simply the optimal rate at which that channel can simulate the noiseless channel $\mathrm{id}_2$ on a single qubit. Likewise, the classical capacity of a quantum channel is its optimal rate for simulation of a qubit dephasing channel

$$\rho \mapsto |0\rangle\langle 0|\rho|0\rangle\langle 0| + |1\rangle\langle 1|\rho|1\rangle\langle 1| \qquad [23]$$

In this spirit, the fact that $C_E(\mathcal{N}) = 2Q_E(\mathcal{N})$ can be re-expressed in the form

$$Q_E(\mathcal{N}) = \frac{C_E(\mathcal{N})}{C_E(\mathrm{id}_2)} \qquad [24]$$

Equivalently, when entanglement is free, the optimal rate at which $\mathcal{N}$ can simulate a noiseless qubit channel is given by the ratio between the entanglement-assisted classical capacities of $\mathcal{N}$ and $\mathrm{id}_2$. The quantum reverse Shannon theorem generalizes this statement to the simulation of arbitrary channels in the presence of free entanglement.

Suppose that Alice and Bob would like to use $\mathcal{N}_1 : A' \to B$ to simulate another channel $\mathcal{N}_2 : A' \to B$. Fix an input state $\varphi^{A'}$ and let $|\varphi\rangle^{AA'^n}$ be a purification of $(\varphi^{A'})^{\otimes n}$. As always, assume that Alice and Bob share a maximally entangled state $|\Phi\rangle^{\tilde{A}\tilde{B}}$ of Schmidt rank $2^{nS}$. Alice's encoding operation will be a TPCP map $\mathcal{E} : \tilde{A}A'^n \to A'^m$ acting on $n$ copies of the input system $A'$ and her half of the shared entanglement, $\tilde{A}$. Bob's decoding will likewise be a TPCP map $\mathcal{D} : B^m\tilde{B} \to B^n$ acting on $m$ copies of the output of the channel, and his half of the shared entanglement, $\tilde{B}$. This procedure is said to $\epsilon$-simulate $\mathcal{N}_2^{\otimes n}$ on $(\varphi^{A'})^{\otimes n}$ if

$$F\Big(\mathcal{N}_2^{\otimes n}(\varphi^{AA'^n}), (\mathcal{D} \circ \mathcal{N}_1^{\otimes m} \circ \mathcal{E})(\Phi^{\tilde{A}\tilde{B}} \otimes \varphi^{AA'^n})\Big) \geq 1 - \epsilon \qquad [25]$$

where $F$ is the mixed state fidelity $F(\rho, \sigma) = (\mathrm{tr}\sqrt{\rho^{1/2}\sigma\rho^{1/2}})^2$. The entire procedure, illustrated in Figure 3, is said to be a $(2^{nS}, m, n, \epsilon)$ entanglement-assisted simulation of $\mathcal{N}_2$ by $\mathcal{N}_1$. A rate $R$, measured in copies of $\mathcal{N}_2$ per copy of $\mathcal{N}_1$, is said to be achievable for $\varphi^{A'}$ if there exists a choice of $S \geq 0$ and a sequence of $(2^{nS}, m_n, n, \epsilon_n)$ entanglement-assisted simulations with $n/m_n \to R$ while $\epsilon_n \to 0$.

The quantum reverse Shannon theorem states that the entanglement-assisted capacity completely governs the achievable simulation rates.

**Theorem 2** (Winter 2004, Bennett *et al.*). *Given two channels $\mathcal{N}_1 : A' \to B$ and $\mathcal{N}_2 : A' \to B$, $R$ is an achievable simulation rate for $\mathcal{N}_2$ by $\mathcal{N}_1$ and all input states $\varphi^{A'}$ if and only if*

$$R \leq \frac{C_E(\mathcal{N}_1)}{C_E(\mathcal{N}_2)} \qquad [26]$$

Note that the form of eqn [26] ensures that the simulation is asymptotically reversible: if a channel $\mathcal{N}_1$ is used to simulate $\mathcal{N}_2$ and the simulation is then used to simulate $\mathcal{N}_1$ again, then the overall rate becomes

$$\frac{C_E(\mathcal{N}_1)}{C_E(\mathcal{N}_2)}\frac{C_E(\mathcal{N}_2)}{C_E(\mathcal{N}_1)} = 1 \qquad [27]$$

Thus, in the presence of free entanglement and for a known input density operator of the form $(\varphi^{A'})^{\otimes n}$, a single parameter, the entanglement-assisted classical capacity, suffices to completely characterize the asymptotic properties of a quantum channel.

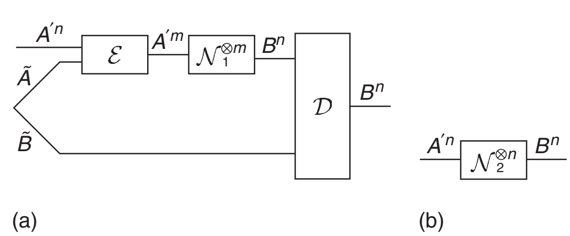



**Figure 3** Circuit representation of an entanglement-assisted simulation of $\mathcal{N}_2$ by $\mathcal{N}_1$. (a) The simulation circuit, with Alice's encoding operation $\mathcal{E}$ acting on $n$ copies of $A'$ and Bob's decoding operation producing $n$ copies of $B$. (b) The circuit that the protocol is intended to simulate. As stated, the quantum reverse Shannon theorem allows the simulation circuit to depend on the density operator of the input state restricted to $A'^n$.

Moreover, since two channels that are asymptotically equivalent without free entanglement will surely remain equivalent if free entanglement is permitted, eqn [26] gives essentially the only possible nontrivial, single-parameter asymptotic characterization of quantum channels. This is the sense in which the entanglement-assisted capacity should be regarded as the most important capacity of a quantum channel.

The proof of the quantum reverse Shannon theorem is quite involved, but some of its features can be understood without much work. First, note that by the optimality statement of the entanglement-assisted classical capacity, the desired simulation can exist only if eqn [26] holds. Otherwise, composing the simulation of $\mathcal{N}_2$ by $\mathcal{N}_1$ with a sequence of codes achieving $C_E(\mathcal{N}_2)$ would result in a sequence of codes beating the capacity formula for $\mathcal{N}_1$.

Similarly, note that one method to simulate a channel $\mathcal{N}_1$ using $\mathcal{N}_2$ is to first use $\mathcal{N}_2$ to simulate the noiseless channel and then use the simulated noiseless channel to simulate $\mathcal{N}_1$. Since the achievable rates for the first step are characterized by the entanglement-assisted capacity theorem, proving the achievability part of Theorem 2 reduces to finding protocols for simulating a general noisy quantum channel $\mathcal{N}_2$ by a noiseless one. That perhaps sounds like a strange goal, but nonetheless is the difficult part of the quantum reverse Shannon theorem.

It is likely that the quantum reverse Shannon theorem can be extended to cover other types of inputs than the known tensor power states $(\varphi^{A'})^{\otimes n}$. The most desirable form of the theorem would be one valid for all possible input density operators on $A'^{\otimes n}$, providing a single simulation procedure dependent only on the channels and not the input state. It is known that without modifying the form of the free entanglement, this most ambitious form of the theorem fails, but it is conjectured that the full-strength theorem does hold provided very large amounts of entanglement are supplied in the form of the so-called embezzling states (van Dam and Hayden 2003).

## Relationships between Protocols

There is another sense in which the entanglement-assisted capacity can be viewed as the fundamental capacity of a quantum channel: an efficient protocol for achieving the entanglement-assisted capacity can be converted into protocols achieving the unassisted quantum and classical capacities, or at least very close variants thereof.

An efficient protocol in this case refers to one that does not waste entanglement. Suppose that $\mathcal{N}: A' \to B$ can be written $\mathrm{tr}_E \mathcal{U}^{BE}$ for some isometry $\mathcal{U}^{BE}$. Let $|\varphi\rangle^{AA'}$ be a pure state and $|\sigma\rangle^{ABE} = \mathcal{U}^{BE}|\varphi\rangle^{AA'}$ the corresponding purified channel output state. Careful analysis of the entanglement-assisted classical communication protocol achieving the rate $I(A;B)_\sigma$ leads to an entanglement-assisted quantum communication protocol consuming entanglement at the rate $(1/2)I(A;E)_\sigma$ ebits per use of $\mathcal{N}$ and yielding communication at the rate of $(1/2)I(A;B)_\sigma$ qubits per use $\mathcal{N}$. The protocol achieving this goal is known as the "father" (Devetak *et al.* 2004).

If the entanglement consumed in the father were actually supplied by quantum communication from Alice to Bob, then the net rate of quantum communication produced by the resulting protocol would be $(1/2)I(A;B)_\sigma - (1/2)I(A;E)_\sigma$ qubits from Alice to Bob, that is, the total produced minus the total consumed.

This quantity, how much more information $B$ has about $A$ than $E$ does, can be simplified using an interesting identity. Since $|\sigma\rangle^{ABE}$ is pure,

$$I(A;E)_\sigma = H(A)_\sigma + H(E)_\sigma - H(AE)_\sigma \qquad [28]$$

$$= H(A)_\sigma + H(AB)_\sigma - H(B)_\sigma \qquad [29]$$

Expanding $I(A;B)_\sigma$ and canceling terms then reveals that

$$\begin{aligned}\tfrac{1}{2}I(A;B) - \tfrac{1}{2}I(A;E) &= -H(A|B)_\sigma \\ &= I_c(A\rangle B)_\sigma \qquad [30]\end{aligned}$$

where the function $I_c$ is known as the coherent information. After optimizing over input states and multiple channel uses, this is precisely the formula for the unassisted quantum capacity of a quantum channel (Devetak 2005). Thus, the net rate of qubit communication for the protocol derived from the father exactly matches the rates necessary to achieve the unassisted quantum capacity. The only caveat is that the protocol derived from the father uses quantum communication catalytically, meaning that some communication needs to be invested in order to get a gain of $I_c(A\rangle B)$. For the unassisted quantum capacity, no investment is necessary. Nonetheless, detailed analysis of the situation reveals that the amount of catalytic communication required can be reduced to an amount sublinear in the number of channel uses, meaning the rate of required investment can be made arbitrarily small. In this sense, the father protocol essentially generates the optimal protocols for the unassisted quantum capacity.

Protocols achieving the unassisted classical capacity can be constructed in a similar way. In this case, one starts from an ensemble $\mathcal{E} = \{p_j, \mathcal{N}(\psi_j^{A'})\}$ of states generated by the channel. Achievability of

the unassisted classical capacity formula follows from achievability of rates of the form

$$\chi(\mathcal{E}) = H\Big(\sum_j p_j \mathcal{N}(\psi_j^{A'})\Big) - \sum_j p_j H\Big(\mathcal{N}(\psi_j^{A'})\Big) \qquad [31]$$

for arbitrary ensembles of output states. Consider the channel

$$\widetilde{\mathcal{N}}(\rho) = \sum_j \langle j|\rho|j\rangle \cdot \mathcal{N}(\psi_j) \qquad [32]$$

and input state $|\varphi\rangle^{AA'} = \sum_j \sqrt{p_j}|j\rangle^A|j\rangle^{A'}$. If $\sigma = \widetilde{\mathcal{N}}(\varphi)$, then $I(A;B)_\sigma$ is equal to $\chi(\mathcal{E})$. Thus, there are protocols consuming entanglement that achieve the classical communications rate $\chi(\mathcal{E})$ for the modified channel $\widetilde{\mathcal{N}}$. Because the channel $\widetilde{\mathcal{N}}$ includes an orthonormal measurement which destroys all entanglement between $A$ and $B$, however, it can be argued that any entanglement used in such a protocol could be replaced by shared randomness, which could then in turn be eliminated by a standard derandomization argument. The net result is a procedure for choosing rate $\chi(\mathcal{E})$ codes for the channel $\mathcal{N}$ consisting of states of the form $\psi_{j_1} \otimes \cdots \otimes \psi_{j_n}$, which is the essence of the achievability proof for the unassisted classical capacity.

This may seem like an unnecessarily cumbersome and even circular approach to the unassisted classical capacity given that the proof sketched above for the entanglement-assisted classical capacity itself invokes the unassisted result in the form of the HSW theorem. The approach becomes more satisfying when one learns that simple and direct proofs of the father protocol exist that completely bypass the HSW theorem (Abeyesinghe *et al.* 2005).

Thus, the entanglement-assisted communication protocols can be easily transformed into their unassisted analogs, confirming the central place of entanglement-assisted communication in quantum information theory.

## Acknowledgmnts

The author is grateful to the inventors of the quantum reverse Shannon theorem for letting him discuss their results prior to their publication and to Jon Yard for a careful reading of the manuscript. This work has been supported by the Canadian Institute for Advanced Research, the Canada Research Chairs program, and Canada's NSERC.

*See also:* Capacity for Quantum Information; Channels in Quantum Information Theory; Entanglement; Finite Weyl Systems; Quantum Channels: Classical Capacity; Quantum Entropy.

## Further Reading

Abeyesinghe A, Devetak I, Hayden P, and Winter A (2005) Fully quantum Slepian–Wolf (in preparation).

Bennett CH, Devetak I, Harrow AW, Shor PW, and Winter A (2005) The quantum Reverse Shannon Theorem (in preparation).

Bennett CH, Shor PW, Smolin JA, and Thapliyal AV (1999) Entanglement-assisted classical capacity of noisy quantum channels. *Physical Review Letters* 83: 3081 (arXiv.org:quant-ph/9904023).

Bennett CH, Shor PW, Smolin JA, and Thapliyal AV (2002) Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. *IEEE Transactions on Information Theory* 48(10): 2637 (arXiv.org:quant-ph/0106052).

Cerf N and Adami C (1997) Von Neumann capacity of noisy quantum channels. *Physical Review A* 56: 3470 (arXiv.org:quant-ph/9609024).

Devetak I (2005) The private classical capacity and quantum capacity of a quantum channel. *IEEE Transactions on Information Theory* 51(1): 44 (arXiv.org/0304127).

Devetak I, Harrow AW, and Winter A (2004) A family of quantum protocols. *Physical Review Letters* 93: 230504 (arXiv.org:quant-ph/0308044).

DiVincenzo DP, Smolin JA, and Shor PW (1998) Quantum channel capacity of very noisy channels. *Physical Review A* 57: 830 (arXiv.org:quantph/9706061).

Holevo AS (1998) The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory* 44: 269–273.

Schumacher B and Westmoreland MD (1997) Sending classical information via noisy quantum channels. *Physical Review A* 56: 131–138.

van Dam W and Hayden P (2003) Universal entanglement transformation without communication. *Physical Review A* 67: 060302 (arXiv.org:quant-ph/0201041).

Winter A (2004) Extrinsic and instrinsic data in quantum measurements: asymptotic convex decomposition of positive operator valued measures. *Communications in Mathematical Physics* 244(1): 157 (arXiv.org:quantph/0109050).

# Capacity for Quantum Information

**D Kretschmann**, Technische Universität Braunschweig, Braunschweig, Germany



## Introduction

Any processing of quantum information, be it storage or transfer, can be represented as a quantum channel: a completely positive and trace-preserving map that transforms states (density matrices) on the sender's end of the channel into states on the receiver's end. Very often, the channel $S$ that sender and receiver (conventionally called Alice and Bob, respectively) would like to implement is not readily available, typically due to detrimental noise effects, limited technology, or insufficient funding. They may then try to simulate $S$ with some other channel $T$, which they happen to have at their disposal. The quantum channel capacity $Q(T,S)$ of $T$ with respect to $S$ quantifies how well this simulation can be performed, in the limit of long input strings, so that Alice and Bob can take advantage of collective pre- and post-processing (cf. Figure 1). Higher capacities may result if Alice and Bob are allowed to use additional resources in the process, such as classical side channels or a bunch of maximally entangled pairs shared between them.

Quantum capacity thus gives the ultimate benchmarks for the simulation of one quantum channel by another and for the optimal use of auxiliary resources. Together with the compression rate of a quantum source (*see* Source Coding in Quantum Information Theory), it lies at the heart of quantum information theory.

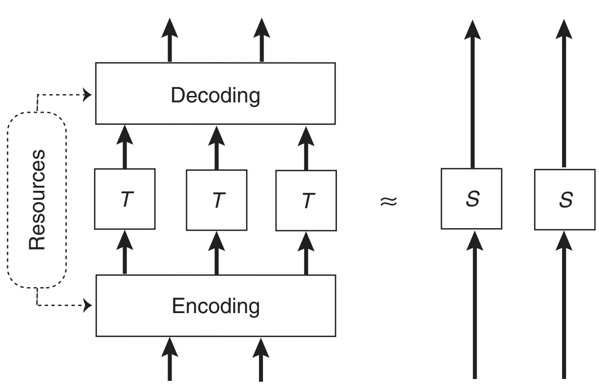



**Figure 1** Equipped with collective encoding and decoding operations (and perhaps some auxiliary resources), $n=3$ instances of the channel $T$ simulate $m=2$ instances of the channel $S$. The transmission rate of the above scheme is 2/3. Capacity is the largest such rate, in the limit of long messages and optimal encoding and decoding.

In a very typical scenario, Alice and Bob would like to implement the ideal (noiseless) quantum channel $S = \mathrm{id}$: they are interested in sending quantum states undistorted over some distance, or want to store them safely for some period of time, so that all the precious quantum correlations are preserved. The capacity $Q(T) \equiv Q(T, \mathrm{id})$ is then the maximal number of qubit transmissions per use of the channel, taken in the limit of long messages and using collective encoding and decoding schemes asymptotically eliminating all transmission errors. This is what is generally called the *quantum capacity* of the channel $T$, and it is our main focus in this article. Little is known so far about the quantum capacity for the simulation of other (nonideal) channels (cf. the section "Related capacities").

In remarkable contrast to the classical setting, quantum channel capacities are very much affected by additional resources. This leads to unexpected and fascinating applications such as teleportation and dense coding. But it also results in a bewildering variety of inequivalent channel capacities, which still hold many challenges for future research.

### Notation

A quantum channel which transforms input systems on a Hilbert space $\mathcal{H}_\mathcal{A}$ into output systems on a (possibly different) Hilbert space $\mathcal{H}_\mathcal{B}$ is represented (in Schrödinger picture) by a completely positive and trace-preserving linear map $T : \mathcal{B}_*(\mathcal{H}_\mathcal{A}) \to \mathcal{B}_*(\mathcal{H}_\mathcal{B})$, where $\mathcal{B}_*(\mathcal{H})$ denotes the space of trace class operators on the Hilbert space $\mathcal{H}$ (*see* Channels in Quantum Information Theory). We write $\mathcal{A}$ instead of $\mathcal{B}_*(\mathcal{H}_\mathcal{A})$ to streamline the presentation, and $\mathcal{A}^n$ for the $n$-fold tensor product $\mathcal{B}_*(\mathcal{H}_\mathcal{A})^{\otimes n}$.

It is evident that the definition of channel capacity requires the comparison of different quantum channels. A suitable distance measure is the *norm of complete boundedness* (or cb-norm, for short), denoted by $\| \cdot \|_{\mathrm{cb}}$. For two channels $T$ and $S$, the distance $(1/2)\|T - S\|_{\mathrm{cb}}$ can be defined as the largest difference between the overall probabilities in two statistical quantum experiments differing only by exchanging one use of $S$ by one use of $T$. These experiments may involve entangling the systems on which the channels act with arbitrary further systems; hence the cb-norm remains a valid distance-measure if the given channel is only part of a larger system. Equivalently, we may set $\|T\|_{\mathrm{cb}} := \sup_n \|T \otimes \mathrm{id}_n\|$, where $\|R\| := \sup_{\|\varrho\|_1 \le 1} \|R(\varrho)\|_1$

denotes the norm of linear operators, and $\|\varrho\|_1 := \mathrm{tr}\sqrt{\varrho^*\varrho}$ is the trace norm on the space of trace-class operators $\mathcal{B}_*(\mathcal{H})$.

We use base two logarithms throughout, and we write $\mathrm{ld}\, x := \log_2 x$ and $\exp_2 x := 2^x$.

## Quantum Channel Capacity

The intuitive concept underlying quantum channel capacity is made rigorous in the following definition:

**Definition 1** A positive number $R$ is called *achievable rate* for the quantum channel $T: \mathcal{A} \to \mathcal{B}$ with respect to the quantum channel $S: \mathcal{A}' \to \mathcal{B}'$ iff for any pair of integer sequences $(n_\nu)_{\nu\in\mathbb{N}}$ and $(m_\nu)_{\nu\in\mathbb{N}}$ with $\lim_{\nu\to\infty} n_\nu = \infty$ and $\overline{\lim}_{\nu\to\infty} \frac{m_\nu}{n_\nu} \leq R$ we have

$$\lim_{\nu\to\infty} \inf_{D,E} \|DT^{\otimes n_\nu}E - S^{\otimes m_\nu}\|_{\mathrm{cb}} = 0 \qquad [1]$$

the infimum taken over all encoding channels $E$ and decoding channels $D$ with suitable domain and range. The *channel capacity* $Q(T,S)$ of $T$ with respect to $S$ is defined to be the supremum of all achievable rates. The *quantum capacity* is the special case $Q(T) := Q(T, \mathrm{id}_2)$, with $\mathrm{id}_2$ being the ideal qubit channel.

In this article, we mainly concentrate on channels between finite-dimensional systems. This is enough to bring out the basic ideas. Many of the concepts and results discussed here can be generalized to *Gaussian channels*, which play a central role as building blocks for quantum optical communication lines (Holevo and Werner 2001, Eisert and Wolf).

There is considerable freedom in the definition of quantum channel capacity, at least for ideal reference channels (Kretschmann and Werner 2004). In particular, the encoding channels $E$ in eqn [1] may always be restricted to isometric embeddings.

In addition, it is not necessary to check an infinite number of pairs of sequences $(n_\nu)_{\nu\in\mathbb{N}}$ and $(m_\nu)_{\nu\in\mathbb{N}}$ when testing a given rate $R$, as Definition 1 would suggest. Instead, it is enough to find one such pair which achieves the rate $R$ infinitely often, $\overline{\lim}_{\nu\to\infty} m_\nu/n_\nu = R$.

Without affecting the capacity, the cb-norm $\|T\|_{\mathrm{cb}}$ may be replaced by the unstabilized operator norm $\|T\|$ or by fidelity measures, which are in general much easier to compute. In particular, one might choose the *minimum fidelity*,

$$F(T) := \min_{\|\psi\|=1} \langle\psi|T(|\psi\rangle\langle\psi|)|\psi\rangle \qquad [2]$$

or even the *average fidelity*,

$$\bar{F}(T) := \int \langle\psi|T(|\psi\rangle\langle\psi|)|\psi\rangle \, \mathrm{d}\psi \qquad [3]$$

Unfortunately, this equivalence is restricted to capacities with noiseless reference channel $S = \mathrm{id}$. In the vicinity of other (nonideal) channels, equivalence of the stabilized and unstabilized error criteria may be lost. Of course, the comparison of channels is ultimately based on the comparison of a state to its image, and here the pure states are the worst case. Hence, the remarkable insensitivity of the quantum capacity to the choice of the error criterion stems from the observation that the comparison between an arbitrary state and a pure state is rather insensitive to the criterion used.

Instead of requiring the error quantity in eqn [1] to approach zero in the large block limit $\nu \to \infty$, one might feel tempted to impose that the errors vanish completely for some sufficiently large block length, since this is the standard setup in the theory of quantum error correction (*see* Quantum Error Correction and Fault Tolerance). While it is true that errors can always be assumed to vanish exponentially in eqn [1], requiring perfect correction may completely change the picture: if a channel has some small positive probability for depolarization, the same also holds for its tensor powers, and no such channel allows the perfect transmission of even one qubit. Hence, the capacity for perfect correction will vanish for such channels, while the standard capacity (in accordance with Definition 1) will be close to maximal, $Q(T) \approx 1$. The existence of perfect error-correcting codes thus gives lower bounds on the channel capacity, but is not required for a positive transfer rate.

In the other extreme, one might sometimes feel inclined to tolerate (small) finite errors in the transmission. For some $\varepsilon > 0$, we define $Q_\varepsilon(T)$ exactly like the quantum capacity in Definition 1, but require only that the error quantity in eqn [1] falls below $\varepsilon$ for some sufficiently large $\nu$. Obviously, $Q_\varepsilon(T) \geq Q(T)$ for any quantum channel $T$. We also have $\lim_{\varepsilon\to 0} Q_\varepsilon(T) = Q(T)$ (Kretschmann and Werner 2004). In the classical setting, even a *strong converse* is known: if $\varepsilon > 0$ is small enough, one cannot achieve bigger rates by allowing small errors, that is, $C_\varepsilon(T) = C(T)$. It is still undecided whether an analogous property holds for the quantum capacity $Q(T)$.

## Related Capacities

This article is chiefly concerned with the quantum capacity of a quantum channel. A variety of other

capacities have been derived from Definition 1 by either amending the channel $S$ to be simulated, or allowing Alice and Bob to make use of additional resources. Their interrelations are reviewed in Bennett *et al.* (2004)

Much interest has been devoted to the hybrid problem of transmitting classical information undistorted over noisy quantum channels. The classical capacity $C(T)$ of a quantum channel $T$ is discussed in the article Quantum Channels: Classical Capacity of this Encyclopedia. It is obtained by choosing the ideal one-bit channel rather than the one-qubit channel as the standard of reference in Definition 1. Encoding channels $E$ and decoding channels $D$ are then restricted to preparations and measurements, respectively. Since a quantum channel can also be employed to send classical information, we have $C(T) \geq Q(T)$. There are, obviously, examples in which this inequality is strict: the *entanglement-breaking* channel $T(\varrho) = \sum_j \langle j|\varrho|j\rangle\, |j\rangle\langle j|$ is composed of a measurement in the orthonormal basis $\{|j\rangle\}_j$, followed by a preparation of the corresponding basis states. It destroys all the entanglement between the sender and a reference system, implying $Q(T) = 0$. Yet all the basis states $|j\rangle$ are transmitted undistorted, which is enough to guarantee that $C(T) = 1$.

Definition 1 also applies to purely classical channels, and thus to the setting of Shannon's information theory. A classical channel $T$ between two $d$-level systems is completely specified by the $d \times d$ matrix $(T_{xy})_{x,y=1}^{d}$ of transition probabilities. For these channels the cb-norm difference is just (twice) the maximal error probability:

$$\|\mathrm{id} - T\|_{\mathrm{cb}} = 2\sup_x\{1 - T_{xx}\}$$

which is the standard error criterium for classical information transfer.

Dense coding and teleportation suggest that entanglement is a powerful resource for information transfer. It doubles the classical channel capacity of a noiseless channel, and it allows to send quantum information over purely classical channels. Surprisingly, the *entanglement-assisted capacities* are often simpler and better behaved than their unassisted counterparts. Unlike the classical and quantum capacities proper, they are relatively easy to calculate using finite optimization procedures, and there has recently been significant progress in understanding the simulation rates for nonideal channels in this scenario (*see* Capacities Enhanced by Entanglement).

The quantum channel capacity is unaffected by entanglement-breaking side channels. In particular, classical forward communication alone cannot enhance it. However, unlike in the purely classical case, both the quantum and classical channel capacity (but not the entanglement-assisted capacity) may increase under classical feedback.

## Elementary Properties

The capacity of a composite channel $T_1 \circ T_2$ cannot be bigger than the capacity of the channel with the smallest bandwidth. This in turn suggests that simulating a concatenated channel is in general easier than simulating any of the individual channels. These relations are known as *bottleneck inequalities*:

$$Q(T_1 \circ T_2, S) \leq \min\{Q(T_1, S), Q(T_2, S)\} \qquad [4]$$

$$Q(T, S_1 \circ S_2) \geq \max\{Q(T, S_1), Q(T, S_2)\} \qquad [5]$$

Instead of running $T_1$ and $T_2$ in succession, we may also run them in parallel. In this case, the capacity can be shown to be *superadditive*,

$$Q(T_1 \otimes T_2, S) \geq Q(T_1, S) + Q(T_2, S) \qquad [6]$$

For the standard ideal channels, we even have additivity. The same holds true if both $S$ and one of the channels $T_1, T_2$ are noiseless, the third channel being arbitrary. However, results on the activation of bound-entangled states seem to suggest that the inequality in eqn [6] may be strict for some channels (*see* Entanglement).

Finally, the *two-step coding inequality* tells us that by using an intermediate channel in the coding process we cannot increase the transmission rate:

$$Q(T_1, T_2) \geq Q(T_1, T_3)\, Q(T_3, T_2) \qquad [7]$$

Applying eqn [7] twice with $T_2 = \mathrm{id}$ and $T_3 = \mathrm{id}$ immediately yields upper and lower bounds on the channel capacity with nonideal reference channel,

$$\frac{Q(T_1)}{Q(T_2)} \geq Q(T_1, T_2) \geq Q(T_1)\, Q(\mathrm{id}, T_2) \qquad [8]$$

The evaluation of the lower bound in eqn [8] then requires efficient protocols for simulating a noisy channel $T_2$ with a noiseless resource.

There are special cases in which the quantum channel capacity can be evaluated relatively easily, the most relevant one being the noiseless channel $\mathrm{id}_n$, where by the subscript $n$ we denote the dimension of the underlying Hilbert space. In this case, we have

$$Q(\mathrm{id}_n, \mathrm{id}_m) = \frac{\mathrm{ld}\, n}{\mathrm{ld}\, m} \qquad [9]$$

The lower bound $Q(\mathrm{id}_n, \mathrm{id}_m) \geq \mathrm{ld}n/\mathrm{ld}m$ is immediate from counting dimensions. To establish the upper bound, we use the fact that a noiseless quantum channel cannot simulate itself with a rate

exceeding unity: $Q(\mathrm{id}_m, \mathrm{id}_m) \leq 1$. This is just the upper bound we want to prove for the special case $n = m$, and it can be extended to the general case with the help of the two-step coding inequality [7]: $Q(\mathrm{id}_m, \mathrm{id}_n)\, Q(\mathrm{id}_n, \mathrm{id}_m) \leq Q(\mathrm{id}_m, \mathrm{id}_m) \leq 1$, implying $Q(\mathrm{id}_n, \mathrm{id}_m) \leq 1/Q(\mathrm{id}_m, \mathrm{id}_n) \leq \mathrm{ld}\, n/\mathrm{ld}\, m$, where in the last step we have applied the lower bound with the roles of $n$ and $m$ interchanged.

Combining eqn [9] with the two-step coding inequality [7], we see that for any channel $T$

$$Q(T, \mathrm{id}_n) = \frac{\mathrm{ld}\, m}{\mathrm{ld}\, n} Q(T, \mathrm{id}_m) \qquad [10]$$

which shows that quantum channel capacities relative to noiseless channels of different dimensionality only differ by a constant factor. Fixing the dimensionality of the reference channel then only corresponds to a choice of units. Conventionally, the ideal qubit channel $\mathrm{id}_2$ is chosen as a standard of reference, as in Definition 1 above, thereby fixing the unit "bit."

The upper bound on the capacity of ideal channels can also be obtained from a general upper bound on quantum capacities (Holevo and Werner 2001), which has the virtue of being easily calculated in many situations. It involves the *transposition map*, which we denote by $\Theta$, defined as matrix transposition with respect to some fixed orthonormal basis. The transposition is positive but not completely positive, and thus does not describe a physical channel (*see* Channels in Quantum Information Theory). We have $\|\Theta\|_{\mathrm{cb}} = d$ for a $d$-level system. For any channel $T$ and small $\varepsilon > 0$,

$$Q(T) \leq Q_\varepsilon(T) \leq \mathrm{ld}\ \|T\Theta\|_{\mathrm{cb}} =: Q_\Theta(T) \qquad [11]$$

where $Q_\varepsilon$ is the finite error capacity introduced in the section "Quantum channel capacity."

The upper bound $Q_\Theta(T)$ has some remarkable properties, which make it a capacity-like quantity in its own right. For example, it is exactly additive,

$$Q_\Theta(S \otimes T) = Q_\Theta(S) + Q_\Theta(T) \qquad [12]$$

for any pair $S, T$ of channels, and it satisfies the bottleneck inequality:

$$Q_\Theta(ST) \leq \min\{Q_\Theta(S),\ Q_\Theta(T)\}$$

Moreover, it coincides with the quantum capacity on ideal channels, $Q_\Theta(\mathrm{id}_n) = Q(\mathrm{id}_n) = \mathrm{ld}\, n$, and it vanishes whenever $T\Theta$ is completely positive. In particular, if $\mathrm{id} \otimes T$ maps any entangled state to a state with positive partial transpose, we have $Q_\Theta(T) = 0$.

## State–Channel Duality

Quantum capacity is closely related to the *distillable entanglement*, which is the optimal rate $m/n$ at which $n$ copies of a given bipartite quantum state $\varrho$ shared between Alice and Bob can be asymptotically converted into $m$ maximally entangled qubit pairs (*see* Entanglement). Similar to the quantum capacity, the definition involves the large block limit $n, m \to \infty$ and an optimization over all conceivable distillation protocols. These may consist of several rounds of local quantum operations and (forward or two-way) classical communication. The one-way and two-way distillable entanglement of $\varrho$ will be denoted by $D_1(\varrho)$ and $D_2(\varrho)$, respectively.

Suppose that Alice and Bob are connected by a quantum channel $T$ and run such a one-way distillation protocol on (many copies of) the state $\varrho_T := (T \otimes \mathrm{id})|\Omega\rangle\langle\Omega|$, where $|\Omega\rangle := (1/\sqrt{d_{\mathcal{A}}}) \sum_i |i, i\rangle$ is maximally entangled on $\mathcal{H}_{\mathcal{A}} \otimes \mathcal{H}_{\mathcal{A}'}$. If the distillation yields maximally entangled qubits at positive rate $R$, Alice may apply the standard teleportation scheme to send arbitrary quantum states to Bob undistorted at that same rate $R$. Like the distillation protocol itself, teleportation requires classical forward communication, which however does not affect the channel capacity (cf. the section "Related capacities"). Thus, $Q(T) \geq D_1(\varrho_T)$. If two-way distillation is allowed, we have $Q_2(T) \geq D_2(\varrho_T)$ for the capacity $Q_2(T)$ assisted by two-way classical side communication.

Conversely, if Alice and Bob use a bipartite quantum state $\varrho$ shared between them as a substitute for the maximally entangled state $|\Omega\rangle$ in the standard teleportation protocol, they will implement some noisy quantum channel $T_\varrho$. If this channel allows to transfer quantum information at nonvanishing rate $R$, Alice may share maximally entangled states with Bob at that same rate $R$. Consequently, $D_1(\varrho) \geq Q(T_\varrho)$ and $D_2(\varrho) \geq Q_2(T_\varrho)$.

These relations (Bennett *et al.* 1996) allow to bound channel capacities in terms of distillable entanglement and vice versa. If the two maps $T \mapsto \varrho_T$ and $\varrho \mapsto T_\varrho$ are mutually inverse, we even have $D_1(\varrho) = Q(T_\varrho)$ and $D_2(\varrho) = Q_2(T_\varrho)$. In this case, the duality $\varrho \rightleftharpoons T_\varrho$ is the physical implementation of Jamiolkowski's isomorphism between bipartite states and channels (*see* Channels in Quantum Information Theory). This has been shown (Horodecki *et al.* 1999) to hold for *isotropic states*, which are invariant under the group of all $\overline{U} \otimes U$ transformations, where $\overline{U}$ is the complex conjugate of the unitary $U$. The corresponding channels are partly depolarizing.

In general, $T_{\varrho_T} \neq T$. However, the so-called *conclusive teleportation* allows us to implement $T$ at least probabilistically, resulting in the relation

$$\frac{1}{d_{\mathcal{A}}^2} Q(T) \leq D_1(\varrho_T) \leq Q(T) \qquad [13]$$

The duality [13] can be applied to show that both the unassisted and the two-way quantum capacities are continuous in any open set of channels having nonvanishing capacities (Horodecki and Nowakowski 2005).

## Coding Theorems

Computing channel capacities straight from Definition 1 is a tricky business. It involves optimization in systems of asymptotically many tensor factors, and can only be performed in special cases, like the noiseless channels in the section "Elementary properties." *Coding theorems* aspire to reduce this problem to an optimization over a low-dimensional space. They usually come in two parts: the *converse* provides an upper bound on the channel capacity (typically in terms of some entropic expression), while the *direct* part consists of a coding scheme that attains this bound. By Shannon's celebrated coding theorem, the classical capacity of a classical noisy channel can be obtained from a maximization of the *mutual information* over all joint input–output distributions.

For the quantum channel capacity, the relevant entropic quantity is the *coherent information*,

$$I_c(T, \varrho) := H(T(\varrho)) - H\big(T \otimes \mathrm{id}(|\psi_\varrho\rangle\langle\psi_\varrho|)\big) \qquad [14]$$

where $H$ denotes the von Neumann entropy: $H(\varrho) = -\mathrm{tr}\,\varrho\,\mathrm{ld}\varrho$, and $\psi_\varrho \in \mathcal{H}_A \otimes \mathcal{H}_{A'}$ is a purification of the density operator $\varrho \in \mathcal{A}$. The coherent information does not increase under quantum operations, $I_c(S \circ T, \varrho) \le I_c(T, \varrho)$ for any quantum channel $S$ and state $\varrho \in \mathcal{A}$. This is the *data processing inequality* (Barnum *et al.* 1998), which shows that the regularized coherent information provides an upper bound on the quantum channel capacity: if Alice and Bob have a coding scheme for the channel $T$ with capacity $Q(T)$, $n$ channel uses allow them to share a maximally entangled state of size $\sim \exp_2 n\,Q(T)$. The coherent information of this state equals $\sim n\,Q(T)$, and was no larger prior to Bob's decoding.

Recently, Devetak (2005) developed a coding scheme to show that this bound is in fact attainable. Different proofs were outlined by Lloyd and Shor.

**Theorem 1** *For every quantum channel $T$,*

$$Q(T) = \lim_{n\to\infty} \frac{1}{n} \max_{\varrho} I_c(T^{\otimes n}, \varrho) \qquad [15]$$

Unlike the classical or quantum mutual information, coherent information is strictly superadditive for some channels (DiVincenzo *et al.* 1998). Hence, taking the limit $n \to \infty$ in eqn [15] is indeed required, and in general the evaluation of the capacity formula [15] still demands the solution of asymptotically large variational problems. This should be contrasted with the entanglement-assisted capacities $C_E(T) = 2Q_E(T)$ (where a simple nonregularized coding theorem is known to hold, *see* Capacities Enhanced by Entanglement) and the capacity for classical information $C(T)$ (where additivity is conjectured but not proved, *see* Quantum Channels: Classical Capacity). Even a maximization of the single-shot coherent information $I_c(T, \varrho)$ appears to be a difficult optimization problem, since this quantity is neither convex nor concave and may have multiple local maxima (Shor 2003). Thus, even for simple-looking systems like the qubit depolarizing channel, so far we only have upper and lower bounds on the quantum channel capacity, but do not yet know how to compute its exact value.

We now sketch Devetak's proof of Theorem 1, assuming only some familiarity with Holevo–Schumacher–Westmoreland (HSW) random codes for the classical channel capacity (*see* Quantum Channels: Classical Capacity). It is easily seen from Stinespring's dilation theorem (*see* Channels in Quantum Information Theory) that a noiseless quantum channel provides perfect security against eavesdropping. This is one of the characteristic traits of quantum mechanics and lies at the heart of quantum cryptography. In his proof, Devetak showed a way to turn this around and upgrade coding schemes for private classical information to quantum channel codes.

The relation between quantum information transfer over a channel $T : \mathcal{A} \to \mathcal{B}$ and privacy against eavesdropping is best understood in terms of the *companion channel* $T_\mathcal{E} : \mathcal{A} \to \mathcal{E}$. $T_\mathcal{E}$ arises from a given Stinespring isometry $V : \mathcal{H}_\mathcal{A} \to \mathcal{H}_\mathcal{B} \otimes \mathcal{H}_\mathcal{E}$ of $T \equiv T_\mathcal{B}$ by interchanging the roles of the output system $\mathcal{B}$ and the environment $\mathcal{E}$:

$$T_\mathcal{B}(\varrho) = \mathrm{tr}_\mathcal{E} V \varrho V^* \rightleftharpoons T_\mathcal{E}(\varrho) = \mathrm{tr}_\mathcal{B} V \varrho V^* \qquad [16]$$

The channel $T_\mathcal{E}$ describes the information flow into the environment $\mathcal{E}$, a system we assume to be under complete control of a potential eavesdropper, Eve say. The setup for private classical information transfer (including the definition of rates and capacity) is then exactly the same as for the classical channel capacity (*see* Quantum Channels: Classical Capacity), but the protocols now have to satisfy the additional requirement that $T_\mathcal{E}$ releases (almost) no information to the environment. This can be achieved by randomizing over $\nu_\mathcal{E} \sim \exp_2 n\,\chi(T_\mathcal{E}, \{p_i, \varrho_i\})$ code words of a standard HSW code of total size $\sim \exp_2 n\,\chi(T_\mathcal{B}, \{p_i, \varrho_i\})$, where $\{p_i, \varrho_i\}$ is the quantum ensemble from which a set of random code words

$\{\sigma_{k,l}\}_{k=1,l=1}^{\nu_\mathcal{B},\nu_\mathcal{E}}$ is generated. The appearance of the *Holevo bound*

$$\chi(T,\{p_i,\varrho_i\}) := H\left(\sum_i p_i T(\varrho_i)\right) - \sum_i p_i H(T(\varrho_i)) \quad [17]$$

in the dimension of both these code spaces can be understood from the size of the relevant typical subspaces (Devetak and Winter 2004).

The randomization guarantees that the remaining $\nu_\mathcal{B} \sim \exp_2 n(\chi(T_\mathcal{B}) - \chi(T_\mathcal{E}))$ code words are almost indistinguishable to Eve:

$$\left\| \frac{1}{\nu_\mathcal{E}} \sum_{l=1}^{\nu_\mathcal{E}} T_\mathcal{E}^{\otimes n}(\sigma_{kl} - \sigma_{jl}) \right\|_1 \leq \varepsilon, \quad \forall j,k = 1,\ldots,\nu_\mathcal{B} \quad [18]$$

The net transfer rate for private classical information is then $R \sim \chi(T_\mathcal{B}) - \chi(T_\mathcal{E})$, which is just the total transfer rate for the channel Alice $\to$ Bob reduced by the transfer rate Alice $\to$ Eve.

Remarkably, if $\varrho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ is a decomposition of $\varrho \in \mathcal{A}$ into pure states, the private transfer rate exactly equals the coherent information,

$$\begin{aligned} I_c(T_\mathcal{B},\varrho) &= H(T_\mathcal{B}(\varrho)) - H(T_\mathcal{E}(\varrho)) \\ &= \chi(T_\mathcal{B}) - \chi(T_\mathcal{E}) \end{aligned} \quad [19]$$

The so-called *entropy exchange*

$$H(T_\mathcal{E}(\varrho)) = H\big(T_\mathcal{B} \otimes \mathrm{id}(|\psi_\varrho\rangle\langle\psi_\varrho|)\big)$$

quantifies the extent to which a formerly pure ancilla state becomes mixed via interaction with the signal states. Equation[19] then nicely reflects the intuition that for high-rate quantum information transfer the signal states should not entangle too much with the environment. In fact, for an almost noiseless channel the entropy exchange nearly vanishes, and the optimized coherent information almost attains the maximal value 1, while for nearly depolarizing channels we have $I_c(T_\mathcal{B},\varrho) \approx -H(\varrho) \leq 0$.

So far, we have sketched a protocol for private classical information transfer. Devetak's *coherentification* allows to pass from the transmission of classical messages to the transmission of coherent superpositions. This technique has also been applied to obtain entanglement distillation protocols from secret key distillation, and offers a unified view on the secret classical resources and their quantum counterparts (Devetak and Winter 2004, Devetak *et al.* 2004).

In order to transfer quantum information, Alice will only need to send one half of a maximally entangled state of dimensionality $\sim \exp_2 n I_c(T_\mathcal{B},\varrho)$. As described in the previous section, teleportation then allows her to transfer arbitrary quantum states from a subspace of that size.

Given a set of pure state code words $\{|\varphi_{kl}\rangle\}_{k=1,l=1}^{\nu_\mathcal{B},\nu_\mathcal{E}}$ of a private classical information protocol, for entanglement transfer Alice prepares the input state

$$|\Phi\rangle_{\mathcal{A}'\mathcal{A}} = \frac{1}{\sqrt{\nu_\mathcal{B}}} \sum_{k=1}^{\nu_\mathcal{B}} |k\rangle_{\mathcal{A}'} \otimes \frac{1}{\sqrt{\nu_\mathcal{E}}} \sum_{l=1}^{\nu_\mathcal{E}} |\varphi_{kl}\rangle_\mathcal{A} \quad [20]$$

where $\mathcal{A}'$ denotes a reference system that Alice keeps in her lab. On his share of the resulting output state $|\Phi'\rangle_{\mathcal{A}'\mathcal{B}\mathcal{E}}$ Bob will then employ the corresponding measurement operators $\{M_{kl}\}_{k,l=1}^{\nu_\mathcal{B},\nu_\mathcal{E}}$ to implement the coherent measurement

$$V_M\,|\varphi\rangle_\mathcal{B} := \sum\nolimits_{kl} \sqrt{M_{kl}}|\varphi\rangle_\mathcal{B} \otimes |kl\rangle_{\mathcal{B}_1\mathcal{B}_2}$$

which places the measurement outcomes into some reference system $\mathcal{B}_1 \otimes \mathcal{B}_2$. Any measurement which identifies the output with high probability only slightly disturbs the output state, and thus Bob's coherent measurement leaves the total system in an approximation of the state

$$|\Phi''\rangle = \frac{1}{\sqrt{\nu_\mathcal{B}\nu_\mathcal{E}}} \sum_{k=1,l=1}^{\nu_\mathcal{B},\nu_\mathcal{E}} |k\rangle_{\mathcal{A}'}|k\rangle_{\mathcal{B}_1}|l\rangle_{\mathcal{B}_2}|\varphi'_{kl}\rangle_{\mathcal{B}\mathcal{E}} \quad [21]$$

in which Eve and Bob are still entangled. A completely depolarizing channel $T_\mathcal{E}$ would directly yield a factorized output state $\mathcal{B} \otimes \mathcal{E}$ here. Although the randomization in eqn [18] does not necessarily result in complete depolarization, there is a controlled unitary operation which Bob may apply to effectively decouple Eve's system, resulting in the output state $\sim (1/\sqrt{\nu_\mathcal{B}}) \sum_k |kk\rangle_{\mathcal{A}'\mathcal{B}_1} \otimes \mathcal{E}$, which is the maximally entangled state of size $\nu_\mathcal{B} \sim \exp_2 n I_c(T_\mathcal{B},\varrho)$ required for teleportation. The direct part of the capacity theorem then follows by applying the above coding scheme to large blocks and maximizing over (pure) input ensembles, concluding the proof.

Devetak's proof of the coding theorem seems to indicate that the private classical capacity $C_p(T)$ equals the quantum capacity $Q(T)$ for every quantum channel $T$. However, for the coherentification protocol, we have restricted the private coding schemes to pure state input ensembles, and thus we can only conclude that $Q(T) \leq C_p(T)$. The existence of bound-entangled states with positive one-way distillable secret key rate (Horodecki *et al.* 2005) implies that this inequality can be strict. A general procedure does exist to retrieve (almost) all the information from the output of a noisy quantum channel that releases (almost) no information to the environment. But this requires a stronger form of privacy than eqn [18].

## Quantum Channels with Memory

This article has so far been restricted to *memoryless* quantum channels, in which successive channel inputs are acted on independently. Messages of $n$ symbols are then processed by the tensor product channel $T^{\otimes n}$, as in Definition 1 and illustrated in **Figure 1**. In many real-world applications, the assumption of having uncorrelated noise cannot be justified, and memory effects need to be taken into account. For a quantum channel $T$ with register input $\mathcal{A}$ and register output $\mathcal{B}$, such effects are conveniently modeled (Bowen and Mancini 2004) by introducing an additional memory system $\mathcal{M}$, so that now $T:\mathcal{M}\otimes\mathcal{A}\rightarrow\mathcal{B}\otimes\mathcal{M}$ is a completely positive and trace-preserving map with two input systems and two output systems. Long messages with $n$ signal states will then be processed by the concatenated channel $T_n:\mathcal{M}\otimes\mathcal{A}^n\rightarrow\mathcal{B}^n\otimes\mathcal{M}$. In such a concatenation, the memory system is passed on from one channel application to the next, and thus introduces (classical or quantum) correlations between consecutive register inputs.

Remarkably, this relatively simple model can be shown (Kretschmann and Werner 2005) to encompass every reasonable physical process: every stationary channel $S:\mathcal{A}^\infty\rightarrow\mathcal{B}^\infty$ which turns an infinite string of input states (on the quasilocal algebra $\mathcal{A}^\infty$) into an infinite string of output states on $\mathcal{B}^\infty$ and satisfies the *causality* constraint is in fact a concatenated memory channel. Causality here means that the outputs of the stationary channel $S$ at given time $t_0$ do not depend on inputs at times $t > t_0$. **Figure 2** illustrates the *structure theorem* for causal stationary quantum channels. In general, it produces not only the memory channel $T$ with memory algebra $\mathcal{M}$, but also a map $R$ describing the influence of input states in the remote past. Intuitively, such a map is often not needed, because memory effects decrease in time: the memory channel $T$ is called *forgetful* if outputs at a large time $t$ depend only weakly on the memory initialization at time zero. In fact, memory effects can be shown to die out even exponentially. The set of these channels is open and dense in the set of quantum memory channels. Hence, generic memory channels are forgetful.

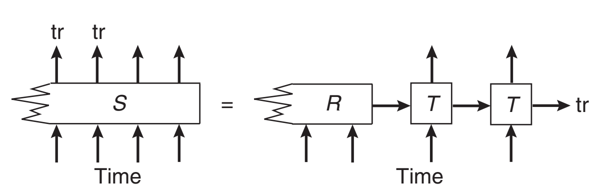



**Figure 2** By the structure theorem, a causal automaton $S$ can be decomposed into a chain of concatenated memory channels $T$ plus some input initializer $R$. Evaluation with the partial trace tr means that the corresponding output is ignored.

The capacity of memory channels is defined in complete analogy to the memoryless case, replacing the $n$-fold tensor product $T^{\otimes n}$ in Definition 1 by the $n$-fold concatenation $T_n$. The coding theorems for (private) classical and quantum information can then be extended from the memoryless case to the very important class of forgetful channels (Kretschmann and Werner 2005).

Nonforgetful channels call for universal coding schemes, which apply irrespective of the initialization of the input memory. Such schemes are presently known only for very special cases.

## Acknowledgmnts

The author thanks the members of the quantum information group at TU Braunschweig for their careful reading of the manuscript and many helpful suggestions. He also gratefully acknowledges the funding from Deutsche Forschungsgemeinschaft (DFG).

*See also:* Capacities Enhanced by Entanglement; Channels in Quantum Information Theory; Entanglement; Positive Maps on $C^*$-Algebras; Quantum Channels: Classical Capacity; Quantum Error Correction and Fault Tolerance; Source Coding in Quantum Information Theory.

## Further Reading

Barnum H, Nielsen MA, and Schumacher B (1998) Information transmission through a noisy quantum channel. *Physical Review A* 57: 4153 (quant-ph/9702049).

Bennett CH, Devetak I, Shor PW, and Smolin JA (2004) Inequalities and separations among assisted capacities of quantum channels, quant-ph/0406086.

Bennett CH, DiVincenzo DP, Smolin JA, and Wootters WK (1996) Mixed-state entanglement and quantum error correction. *Physical Review A* 54: 3824 (quant-ph/9604024).

Bowen G and Mancini S (2004) Quantum channels with a finite memory. *Physical Review A* 69: 012306 (quant-ph/0305010).

Devetak I (2005) The private classical information capacity and quantum information capacity of a quantum channel. *IEEE Transactions on Information Theory* 51: 44 (quant-ph/0304127).

Devetak I, Harrow AW, and Winter A (2004) A family of quantum protocols. *Physical Review Letters* 93: 230504 (quant-ph/0308044).

Devetak I and Winter A (2004) Relating quantum privacy and quantum coherence: an operational approach. *Physical Review Letters* 93: 080501 (quant-ph/0307053).

DiVincenzo DP, Shor PW, and Smolin JA (1998) Quantum channel capacities of very noisy channels. *Physical Review A* 57: 830 (quant-ph/9706061).

Eisert J and Wolf MM Gaussian quantum channels. In Cerf N, Leuchs G, and Polzik E (eds.) *Quantum Information with*

*Continuous Variables of Atoms and Light*. London: Imperial College Press (in preparation)(quant-ph/0505151).
Holevo AS and Werner RF (2001) Evaluating capacities of bosonic Gaussian channels. *Physical Review A* 63: 032312 (quant-ph/9912067).
Horodecki M, Horodecki P, and Horodecki R (1999) General teleportation channel, singlet fraction, and quasidistillation. *Physical Review A* 60: 1888 (quant-ph/9807091).
Horodecki P and Nowakowski ML (2005) Simple test for quantum channel capacity, quant-ph/0503070.
Horodecki K, Pankowski L, Horodecki M, and Horodecki P (2005) Low dimensional bound entanglement with one-way distillable cryptographic key, quant-ph/0506203.
Kretschmann D and Werner RF (2004) Tema con variazioni: quantum channel capacity. *New Journal of Physics* 6: 26 (quant-ph/0311037).
Kretschmann D and Werner RF (2005) Quantum channels with memory. *Physical Review A* 72: 062323 (quant-ph/0502106).
Shor PW (2003) Capacities of quantum channels and how to find them. *Mathematical Programming* 97: 311 (quant-ph/0304102).

# Capillary Surfaces

**R Finn**, Stanford University, Stanford, CA, USA



## Historical and Conceptual Background

A capillary surface is the interface separating two fluids that lie adjacent to each other and do not mix. Examples of such surfaces are the upper surface of liquid partially filling a vertical cylinder (capillary tube), the surface of a liquid drop resting in equilibrium on a tabletop (sessile drop) and the surface of a liquid drop hanging from a ceiling (pendent drop); further instances are the surface of a falling raindrop, the bounding surface of the liquid in the fuel tank of a spaceship, and the interface formed by a fluid mass rotating within another fluid. This last example extends to the problem of rotating stars.

Interfaces separating fluids and solids share some of the physical attributes of capillary surfaces, and the study of wetted portions of rigid "support surfaces" becomes essential for describing global behavior of capillary configurations. However, some significant distinctions appear that change the formal structure of the problems, and must be accounted for in the theory.

Phenomena governed by capillarity pervade all of daily life, and most are so familiar as to escape special notice. By contrast, throughout the eighteenth century and presumably earlier, great attention centered on the rise of liquid in a narrow glass circular-cylindrical tube dipped vertically into a liquid reservoir (Figure 1); this striking event had a dramatic impact that confounded intuition. Clarification of the behavior became one of the major problems challenging the scientific world of the time, and was not achieved during that period. The term "capillary," adapted from the Latin "capillus" for hair, was applied to the phenomenon since it was observed only for tubes with very fine openings; the

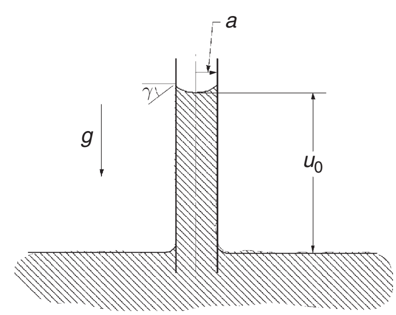



**Figure 1** Capillary tube in infinite reservoir, in downward gravity field.

more general usage adopted in the definition above derives from the recognition of a class of phenomena with a common physical basis.

The first recorded observations concerning capillarity seem due to Aristoteles *c.* 350 BC. He wrote that "a broad flat body, even of heavy material, will float on water, however a narrow thin one such as a needle will always sink." Any reader with access to a needle and a glass of water will have little difficulty refuting the assertion. Remarkably, the error in reasoning seems not to have been pointed out for almost 2000 years, when Galileo addressed the problem in his *Discorsi*, about 1600. The only substantive studies till that time are apparently those of Leonardo da Vinci a hundred years earlier. Leonardo introduced reasoning close in spirit to that of current literature; however, the Calculus was not available to him, and he was not in a position to develop his ideas in quantitative ways.

### Young's Contribution

The later discovery of the Calculus provided a driving impetus guiding many new studies during the eighteenth century. But despite the enormity of that weapon, it did not on its own suffice, and initial quantitative success had to await two initiatives

taken by Thomas Young in 1805. Young based his studies on the concept of surface tension that had been introduced by von Segner half a century earlier. Segner hypothesized that every curve on a fluid/fluid interface $S$ experiences on both its sides an orthogonal force $\sigma$ per unit length, which (for given temperature) depends only on the materials and is directed into the tangent planes on the respective sides. The presence of such forces can be indicated by simple experiments. They become clearly evident in the case of thin (soap) films spanning a frame, in which case there is an easily observed orthogonal pull on the frame, see the section "Dual interpretation of $\sigma$: distinction between fluids and solids." Young made two basic conceptual contributions (Y1, Y2):

Y1. *Relation of pressure jump across a free interface to mean curvature and surface tension.*

Consider a piece of surface $S$ in the shape of a spherical bowl of radius $R$, separating two immiscible fluid media, as in Figure 2. In equilibrium, any pressure difference $\delta p$ across $S$ must be balanced by a tension $\sigma$ on its rim $\Gamma$. If $S$ projects to a disk of (small) radius $r$ on the plane tangent to $S$ at the symmetry point, we are led to

$$\pi r^2 \delta p \simeq 2\pi r \sigma \sin\vartheta \qquad [1]$$

where $\vartheta$ is inclination of $S$ at the rim, relative to the plane. We thus find at the base point

$$\delta p = 2\sigma \frac{d \sin\vartheta}{dr} = 2\sigma \frac{1}{R} \qquad [2]$$

Young then went on to consider a general $S$, without symmetry hypothesis. Letting $1/R_1, 1/R_2$ denote the planar curvatures at a point in $S$ of two normal sections in orthogonal directions, he asserted that

$$\delta p = 2\sigma \frac{1}{2}\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \equiv 2\sigma H \qquad [3]$$

where $H$ is the mean curvature of $S$ at the point. Although Young provided no formal justification for this step, we can establish it with the aid of a general formula from differential geometry that was not known in his lifetime:

$$\int_S 2H\mathbf{N}\, dS = \oint_\Gamma \mathbf{n}\, ds \qquad [4]$$

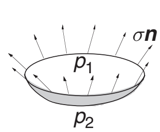



**Figure 2** Pressure change across fluid element, balanced by surface tension.

where $\mathbf{N}$ is a unit normal on $S$, and $\mathbf{n}$ is unit conormal (as indicated in Figure 2) on $\Gamma$. Multiplying both sides of [4] by $\sigma$, the right-hand side becomes the net surface tension force on $S$. Since that must equal the net balancing pressure force, we obtain

$$\int_S (\delta p - 2\sigma H)\mathbf{N}\, dS = 0 \qquad [5]$$

Letting the diameter of $S$ tend to zero, the assertion follows.

We emphasize here the implicit assumption above, that $\sigma$ is a constant depending only on the particular materials, and not on the shape of $S$. This author knows of no source in which that is clearly established, although experiments and experience provide some *a posteriori* justification. See the further comments under Y2, and later in sections "Gauss' contribution: the energy method" and "Dual interpretation of $\sigma$: distinction between fluids and solids."

Y2. *The capillary contact angle.*

Young asserted that there are surface tensions for solid/fluid interfaces analogous to those just introduced, and again depending only on the materials. This assertion is erroneous, as was suggested in writings of Bikerman and of others, and more recently established in a definitive example by Finn. Using his premise, Young attempted to characterize the *contact angle* $\gamma$ made by the fluid surface with a rigid boundary, by requiring that the net tangential component of the three surface tension vectors vanish at the triple interface; this leads to the often employed but incorrect "Young diagram," see Figure 3, and the relation

$$\cos\gamma = \frac{\sigma_1 - \sigma_2}{\sigma_0} \qquad [6]$$

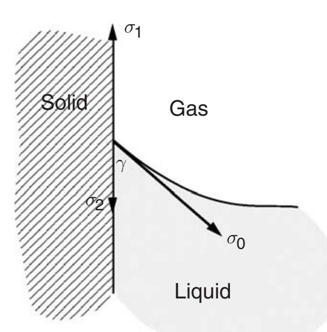



**Figure 3** Young diagram; balance of tangential forces. Residual normal force remains.

for $\cos\gamma$ in terms of the magnitudes of the three "surface tensions.'' Young concluded that the contact angle depends only on the materials, and in no other way on the conditions of the problem. This basic assertion is by a fortuitous accident correct, as follows from the contribution by Gauss described below; it underlies all modern theory.

Using Y1 and Y2, Young produced the first verifiable prediction for the rise height $u_0$ in the circular capillary tube of Figure 1. He assumed the interface to be spherical, so that $H$ is constant and $a = \cos\gamma/H$. He assumed vanishing outside pressure. According to classic laws of hydrostatics, $\delta p = \rho g u_0 = 2\sigma H$ by Y1, where $\rho$ is fluid density; there follows the celebrated relation, presented entirely in words in his 1805 article:

$$u_0 \approx \frac{2\cos\gamma}{\kappa a}, \qquad \kappa = \frac{\rho g}{\sigma} \qquad [7]$$

Young scorned the mathematical method, and made a point of deriving and publishing his results on capillarity without use of any mathematical symbols. This personal idiosyncrasy causes his publications to be something of a challenge to read.

### The Laplace Contribution

In 1806, Laplace published the first analytical expression for the mean curvature of a surface $u(x, y)$, and showed that the expression can be written as a divergence. He obtained the equation

$$\operatorname{div} Tu \equiv 2H, \quad Tu \equiv \frac{\nabla u}{\sqrt{1 + |\nabla u|^2}} \qquad [8]$$

Thus, if $H$ is known from geometrical or physical considerations, as it is for the capillary tube in the example just considered, one finds a second-order (nonlinear) equation for the surface height of any solution as a graph. The equation is elliptic for any function $u(x, y)$ inserted into the coefficients, however not uniformly so; the particular nonuniformity leads to some striking and unusual behavior of its solutions, as we shall see. With the aid of [8], Laplace improved the Young estimate [7] to

$$u_0 \approx \frac{2\cos\gamma}{\kappa a} - \left[\frac{1}{\cos\gamma} - \frac{2}{3}\left(\frac{1 - \sin^3\gamma}{\cos^3\gamma}\right)\right] a \qquad [9]$$

Both Young and Laplace proposed their formulas for "narrow tubes", but neither gave any quantitative indication of what "narrow" should signify. Note that whenever $0 \le \gamma < \pi/2$, [9] becomes negative when the nondimensional *Bond Number* $B = \kappa a^2$ exceeds 8; since $u$ is known to be positive in the indicated range for $\gamma$, [9] provides no information in that case, whereas [7] is still of some value. Nevertheless, [9] is asymptotically exact and consists of the first two terms of the formal expansion in powers of $a$; that was first proved by D Siegel in 1980, almost 200 years following the discovery of the formulas. In 1968, P Concus extended the formal expansion for the height to the entire traverse $0 < r < a$. F Brulois (1981) and independently E Miersemann (1994) proved the expansion to be asymptotic to every order. Explicit bounds for the rise height above and below, making quantitative the notion of "narrow," were obtained by Finn.

Laplace supplied the first detailed mathematical investigations into the behavior of capillary surfaces, applying his ideas to many specific examples. His underlying motivation apparently derived at least partly from astronomical problems, and he published his contributions in two "Suppléments" to the tenth volume of his *Mécanique Céleste*.

### Gauss' Contribution: The Energy Method

Young and Laplace both based their reasonings on force-balance arguments, which at best were unclear and at worst conceptually wrong. In 1830, Gauss took up the problem anew from a variational point of view, using the Johann Bernoulli principle of virtual work. To do so, he attempted to characterize both surface energies and bulk fluid energies in terms of postulated particle attractions and repulsions. In an astonishing 30 pages, he essentially introduced foundations of modern potential theory, of measure theory, and of thermodynamics. He ended up with elaborate expressions that could not readily be applied, and which at least to some extent he did not use. He asserted, for example, that the bulk internal energy would be proportional to volume, which for an incompressible fluid is constant under admissible deformations, and on that basis he ignored the bulk energy term completely. His procedures then led him, in an independent and more convincing way, to the identical equation and boundary condition that had been produced by his predecessors. It must, of course, be remarked that his justification for ignoring the bulk energy term would not be correct for a compressible liquid (see the section "Compressibility"), and it is open to some

question for the central motivating problem of a capillary tube dipped into an infinite liquid bath, in which event there is no volume constraint.

The material that follows is guided by the ideas of Gauss; however, I have found it advantageous to replace his elaborate hypotheses on particle attractions and repulsions by a simpler phenomenological reasoning as to the nature of the energy terms to be expected.

To fix ideas, we consider a semi-infinite cylinder of general section $\Omega$ and of homogeneous material, closed at the bottom, situated vertically in a downward gravity field $g$ per unit mass, and partly filled with an incompressible liquid of density $\rho$ covering the bottom (a more exact discussion, taking account of compressibility, is indicated below in the section "Compressibility"). We assume an equilibrium fluid configuration with the liquid bounded above by an ideally thin interface $S: u(x, y)$ (see Figure 4). We distinguish the energy terms that occur:

1. *Surface energy*. This is the energy required to create the surface interface $S$. We can characterize it by noting that fluid particles within or exterior to the liquid are attracted equally to neighboring particles in all directions; however, at the surface $S$ there is a differential attraction, to particles of the exterior medium (such as air) above, or to the liquid below (see Figure 5). Thus, particles in the interface are pulled orthogonally to $S$. In general, for a liquid–gas interface, significant work will be done only on the liquid and those particles will be pulled toward the liquid; otherwise, the liquid would evaporate across the interface and disappear. The work done in that (infinitesimal) motion is proportional to the area of $S$, so that for the surface energy $E_S$ we obtain

$$E_S = \sigma \int_\Omega \sqrt{1 + |\nabla u|^2} \mathrm{d}x \qquad [10]$$

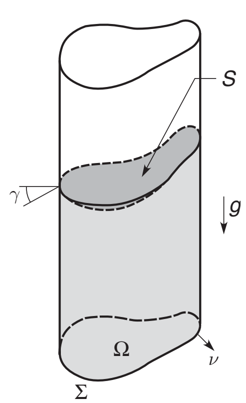



**Figure 4** Liquid in cylindrical capillary tube, of general section $\Omega$. Reproduced with permission from the American Institute of Aeronautics and Astronautics.

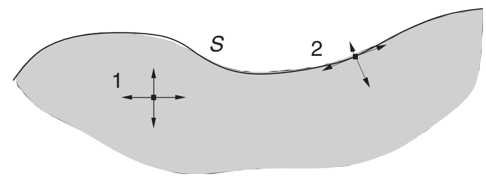



**Figure 5** Attractions on a fluid element: (1) interior to the fluid; (2) on the surface interface.

The constant $\sigma$ has the dimensions of force per unit length, and turns out to be the surface tension of the interface. We note from [10] its dual interpretation as areal energy density on $S$, arising from formation of that surface. This alternative interpretation lends conceptual support to the supposition that $\sigma$ is constant on $S$. See the section "Dual interpretation of $\sigma$: distinction between fluids and solids."

Implicit in the above discussion are deep premises about the nature of the forces acting within the fluid. Essentially these forces must be perceptible only at infinitesimal distances, and grow rapidly with decreasing distance. Forces both of attraction and of repulsion must be present. The recognition of the need for such forces can be traced back to Newton. Quantitative postulates as to their precise nature were introduced by van der Waals in the late nineteenth century, and the topic remains still in active study. Since these forces appear at molecular distance levels, their introduction leads inevitably to questions of statistical mechanics. Additionally, our discussion of work done in forming the surface implicitly assumes a compressible transition layer there, in conflict with our treatment of $S$ as an ideally thin interface bounding an incompressible fluid. In these senses, it is striking that [10] – which is in accord with classical constructions – could be obtained via global qualitative postulates concerning a continuum in static equilibrium, in which the specific nature of the forces is not introduced.

Rayleigh measured the thickness of the surface interface between water and air to be of molecular size, thus providing experimental justification for the procedure adopted.

2. *Wetting energy*. A similar discussion applies at the interface separating the liquid and solid at the cylinder walls; however, this time the net attraction can be in either direction, as particles from neither medium can migrate significantly into the other. For the wetting energy $E_W$, we write, with $\Sigma$ the boundary of $\Omega$,

$$E_W = -\beta\sigma \oint_\Sigma u \, \mathrm{d}s \qquad [11]$$

We designate $\beta$ as the *relative adhesion coefficient* of the liquid–gas–solid configuration. We assume that the cylinder walls are of homogeneous material, so that $\beta$ will be constant. In general, $\beta$ is a difference of factors that apply on the walls at the two interfaces, with the liquid and with the external medium.

3. *Gravitational energy*. The work done in lifting an amount of liquid $\rho\delta h\delta\Omega$ against the gravity field from the base level to a height $h$ in a vertical tube of small section $\delta\Omega$ is $\rho gh\delta h\delta\Omega$. Thus, the work done in filling that tube up to the surface height $u$ is $(\rho gu^2/2)\delta\Omega$, and the total gravitational energy is

$$E_G = \frac{\rho g}{2}\int_\Omega u^2\,\mathrm{d}x \qquad [12]$$

4. *Volume constraint*. In the configuration considered the volume is to be unvaried during admissible deformations; we take account of the constraint by introducing a Lagrange parameter $\lambda$, and an additional "energy" term

$$E_V = \lambda\sigma\int_\Omega u\,\mathrm{d}x \qquad [13]$$

According to the principle of virtual work, the sum $E$ of the above energies must remain unvaried in any deformation that respects all mechanical constraints other than the volume constraint. We choose a deformation $u \to u + \varepsilon\eta$, with $\eta$ smooth in the closure of $\Omega$, which determines a functional $E(\varepsilon)$. From $E'(0) = 0$ follows

$$\int_\Omega\left\{\nabla\eta\cdot\frac{\nabla u}{\sqrt{1+|\nabla u|^2}} + \eta(\kappa u+\lambda)\right\}\mathrm{d}x - \beta\oint_\Sigma \eta\,\mathrm{d}s = 0 \qquad [14]$$

from which

$$\int_\Omega \eta\{-\operatorname{div}Tu + (\kappa u+\lambda)\}\mathrm{d}x + \oint_\Sigma \eta(\nu\cdot Tu - \beta)\mathrm{d}s = 0 \qquad [15]$$

with $Tu \equiv \nabla u/\sqrt{1+|\nabla u|^2}$, and with $\nu$ the unit exterior normal on $\Sigma$. Choosing first $\eta$ to have compact support in $\Omega$, the boundary term vanishes, and the "fundamental lemma" of the calculus of variations yields

$$\operatorname{div}Tu = \kappa u + \lambda, \qquad \kappa = \rho g/\sigma \qquad [16]$$

throughout $\Omega$. Thus, the area integral in [15] vanishes for any $\eta$. We are therefore free to choose $\eta$ as we wish on the boundary, and the fundamental lemma now yields $\nu\cdot Tu = \beta$ on $\Sigma$. We now note that for any liquid surface $u(x,y)$ there holds

$$\nu\cdot Tu = \cos\gamma \qquad [17]$$

on $\Sigma$, where $\gamma$ is the angle between the cylinder wall and the surface $S$, measured within the liquid. Since $\beta$ is assumed to be constant, that is so also for $\gamma$. It is a physical constant: the contact angle, that must be measured in an independent experiment, and cannot be prescribed in advance or calculated within the scope of the theory.

The constant $\beta$, originally introduced as a general proportionality constant, is now characterized as $\beta = \cos\gamma$. We thus see that a physical surface of the form envisaged is possible only if $-1 \le \beta \le 1$. Physically, one expects that if $\beta < -1$ the liquid will separate from the walls, while, if $\beta > 1$, the liquid will spread over the walls as a thin film.

Equation [16] and boundary condition [17] provide a nonlinear second-order equation that is elliptic for any function $u(x,y)$, and also a nonlinear transversality condition on the boundary, for determining the surface interface $S$. The expression $\operatorname{div}Tu$ is exactly twice the mean curvature of the surface $S$. If $\kappa \neq 0$ then $\lambda$ can be eliminated by addition of a constant to $u$. The problem [16]–[17] for the fluid in a vertical cylindrical capillary tube of general section becomes thus a geometrical one: to find a surface whose mean curvature is a prescribed function of position in space, and which meets the cylindrical boundary walls in a prescribed angle $\gamma$.

In the absence of gravity, [16] takes the form

$$\operatorname{div}Tu = 2H \qquad [18]$$

for a surface of constant mean curvature $H$. The constant $H$ is determined by integrating [18] over $\Omega$, and using [17]:

$$2H = \frac{|\Sigma|\cos\gamma}{|\Omega|} \qquad [19]$$

where $|\Sigma|$ and $|\Omega|$ denote the respective perimeter and area, and thus $H$ is independent of volume. From the known uniqueness up to an additive constant of the solutions of [18], [17] it follows that the shape of the solution surface is independent of volume. That result holds also for [16], [17] in view of the possibility to eliminate $\lambda$ from the equation by addition of a constant, and the uniqueness of the solutions of the resulting equation.

Equations [16]–[17] or [18]–[17] are appropriate for determining capillary surfaces that are graphs

$u(x, y)$ over a base domain $\Omega$. More generally, any surface $S$ in 3-space satisfies the equation

$$\Delta x = 2HN \quad [20]$$

where $H$ is its scalar mean curvature and $N$ is a unit normal vector on $S$. Here $\Delta$ is the "intrinsic Laplacian" in the metric of $S$. This is the appropriate relation to be applied in situations for which the physical surface folds over itself and cannot be expressed globally as a graph. The formal simplicity of [20] is deceptive; the challenges arising from the nonlinearity in the equation can be formidable, and very little general theory is as yet available.

### Dual Interpretation of $\sigma$: Distinction between Fluids and Solids

We have already remarked the duality in connection with eqn [10] above. It can be made explicit with a simple experiment proposed by Dupré. One makes a rigid frame with a sliding bar of length $l$, as in Figure 6, and dips the frame into soap solution. On lifting the frame from the solution the opening will be filled with a soap film, and one finds a force $F = 2\sigma l$ on the bar, directed orthogonal to the bar (the factor 2 appears since the film has two sides). The work done in sliding the bar a distance $\delta x$ is $\delta F = 2\sigma l \delta x$, which can also be written $\delta F = 2\sigma \delta A$ with $\delta A$ an element of area. In this sense, the two interpretations of $\sigma$ are formally equivalent, for fluid/fluid interfaces.

The equivalence cannot be extended to solid/fluid interfaces. Consider a rigid spherical ball of generic material and radius $R$, freely floating in an infinite liquid bath in a gravity-free environment, see Figure 7a. It can be shown that the unique symmetric solution to the problem is a horizontal surface, as in the figure. A variational procedure as above shows that if $e_0, e_1, e_2$ are the interfacial energy densities associated with the three interfaces, then

$$\cos \gamma = \frac{e_1 - e_2}{e_0} \quad [21]$$

in formal analogy with the Young relation [6]. But $e_1, e_2$ cannot be interpreted as interfacial forces whose net tangential component cancels that of $e_0$.

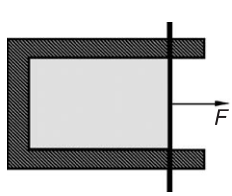
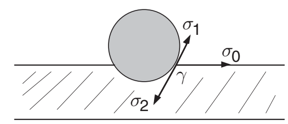



**Figure 6** Dupré apparatus for exhibiting surface tension.



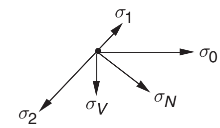




**Figure 7** (a) Floating spherical ball; presumed "Young" forces. (b) Normal and vertical components of Young forces; contradiction to presumed equilibrium.

To do so would lead to a net downward force $\sigma_v$ on the ball (see Figure 7b), contradicting the supposed equilibrium state.

## Mathematical and Physical Predictions: Experiments

In the following sections, we study the kinds of behavior imposed on a surface $S$ by the requirement that it appear as solution of one of the indicated equations and boundary conditions. Some of these properties are quite surprising in the context of classically expected behavior of solutions of equations of mathematical physics. The mathematical predictions were, however, corroborated in certain cases experimentally, as we discuss below.

### Uniqueness and Nonuniqueness

We begin by considering uniqueness questions. We start with a semi-infinite capillary tube, closed at the bottom, to be partially filled with a prescribed volume of (incompressible) liquid making contact angle $\gamma$ on the container walls (Figure 8a). If $\kappa \geq 0$, any solution is uniquely determined. That is a quite general theorem, valid for a wide class of domains $\Omega$ including all piecewise smooth domains (at the corners of which data of the form [17] cannot be prescribed); formally, data can be omitted on any boundary set of linear Hausdorff measure zero. In this result, no growth conditions need be imposed near the boundary (note that such a statement would be false for solutions of the Laplace equation under Dirichlet boundary conditions).

Next we consider a sessile liquid drop on a horizontal plate (Figure 8b). Again the solution is uniquely determined by the volume and by $\gamma$, although the known proof differs greatly from that of the other case.

We now consider a smooth deformation of the base plane, depending on a parameter $t$, which carries it into the cylinder; that can be done in such a way that the supporting surface is at all times "bowl-shaped," as in Figure 8c. Since the bowl formation tends to restrict the possible deformations

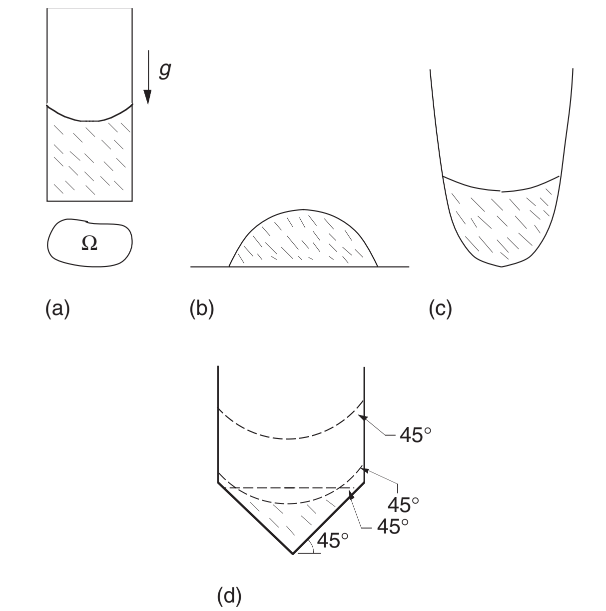



**Figure 8** Support configurations: (a) capillary tube, general section; (b) horizontal plate; (c) convex surface appearing during deformation of horizontal plate to capillary tube; and (d) Nonuniqueness of configuration appearing during convex deformation. Reproduced from *Mathematics Intelligencer* 24(3) 2002 21–33 with permission from Springer-Verlag Heidelberg.

of the fluid consistent with smooth contact with the supporting rigid surface, one might expect that the corresponding capillary surface $S(t)$, arising from the identical fluid mass, will for each $t$ be uniquely determined.

That is however not true, even for symmetric configurations. We can see that from the configuration of Figure 8d, consisting of a vertical circular cylinder whose base is a 45° cone. We assume a contact angle $\gamma = 45°$ and adjust the radius so that a horizontal surface lying just below the cylinder/cone juncture provides the prescribed volume. This is a formal solution surface. Now fill the configuration with a larger volume, so that the contact line will lie above the juncture. The upper surface will no longer be flat, in view of the 45° contact angle, and takes an appearance as indicated in the figure. Finally, we decrease the fluid volume, keeping all other parameters unchanged. As noted above, the upper surface moves rigidly downward, and it is clear that if the original surface is close enough to the juncture line, then the prescribed volume will be attained before the contact line reaches the juncture. Thus, uniqueness fails.

In this construction as just described, the bounding surface is not smooth; however, one sees easily that the procedure continues to work if the edge and vertex are smoothed locally. In fact, one can carry the procedure to a striking conclusion; by appropriate smoothing, one can construct a bounding surface admitting an entire continuum of distinct solution interfaces, all with the same contact angle and enclosing the same fluid volume (Gulliver and Hildebrandt; Finn). This can be done for any gravity field. Figure 9 illustrates seven members of the family of interfaces, in the particular case $\kappa = 0$.

The question immediately arises as to which if any of the continuum of surfaces will be seen in an experiment. In fact, it can be proved that none of the indicated surfaces is mechanically stable (Finn, Concus and Finn, Wente). Since the indicated family includes all symmetric surfaces that are stationary for the energy functional, we find that any stable stationary configuration must be asymmetric. Thus, we have obtained an example of symmetry breaking, in which all conditions of the problem are symmetric, but for which all physically acceptable solutions are asymmetric.

These results were subjected to computational test by M Callahan using the Surface Evolver software, to experimental test by M Weislogel in a drop tower, and to experimental test by S Lucid in the Mir Space Station. The results of the latter experiment are compared in Figure 10 with the computer calculations. In both cases, both a local minimizer (potato chip) and a presumed global minimizer (spoon) were observed.

The seven surface interfaces indicated in Figure 9 all provide the same sum of surface and wetting energy, and bound the same volume of fluid. They all satisfy an eqn [18] with constant $H$, in accordance with hypotheses of incompressibility and vanishing gravity. Thus, formally, all configurations have identical mechanical energy. The surfaces

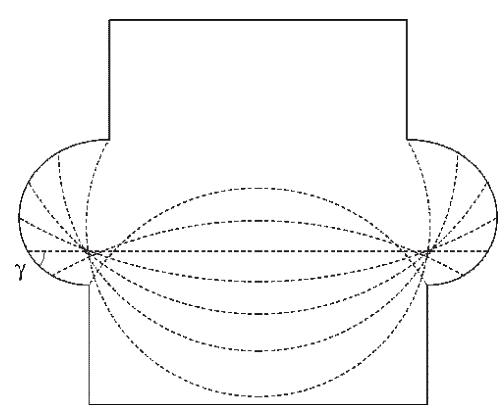



**Figure 9** Seven spherical capillary interfaces in an "exotic" container of homogeneous material in zero gravity. All interfaces bound the same volume and have the same sum of free surface and wetting energies. If all pressures above the interfaces are the same, then the pressures below them successively increase as the curvature vectors of the vertical sections change from upwardly to downwardly directed. Reproduced from Mathematics Intelligences 24(3) 2002 21–33 with permission from Springer-Verlag Heidelberg.

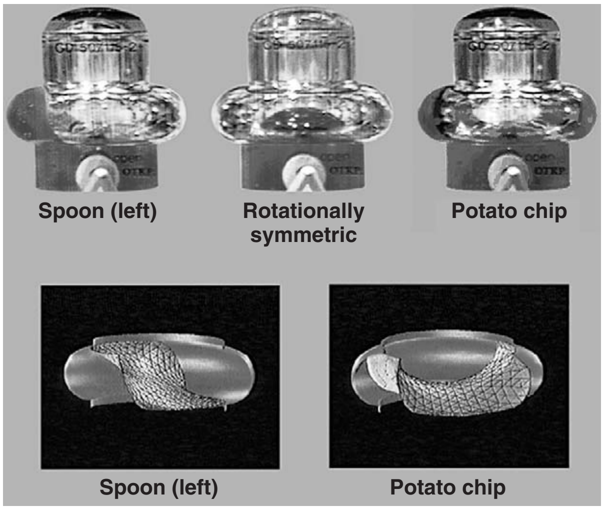



**Figure 10** Symmetry breaking in exotic container, $g = 0$. Below: calculated presumed global minimizer (spoon) and local minimizer (potato chip). Above: experiment on Mir: symmetric insertion of fluid (center); spoon (left); potato chip (right). This is a grayscale version of a color figure reproduced from *Journal of Fluid Mechanics*, 224: 383–94, (1991) with permission of Cambridge University Press.

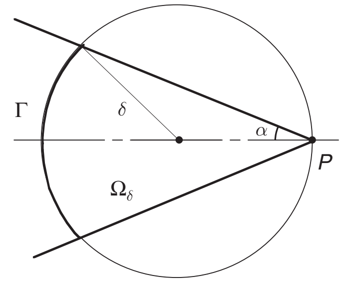



**Figure 11** Wedge domain. Reproduced from Finn R "Capillary Surface Interfaces" in *Notices of AMS* 46 No.7 (1999) with permission of the American Mathematical Society.

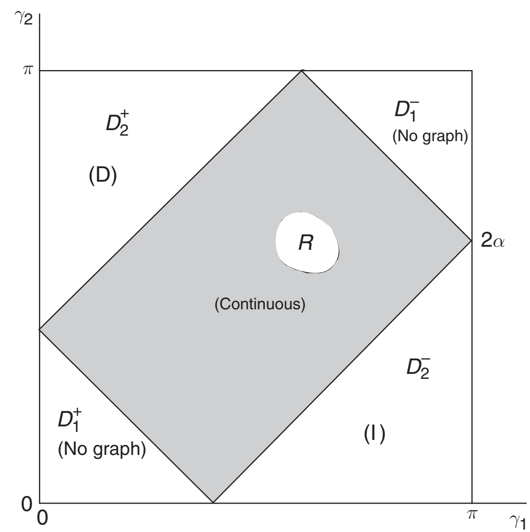



**Figure 12** Domain $R$ of data yielding continuous normal to capillary surface in wedge of opening $2\alpha < \pi$. The symbols D and I are clarified in the section "Behavior at a corner point." Reproduced from "Capillary Wedges Revisited" in *SIAM J. Math. Anal.* 27 No.1 (1996) 56–69 with permission from SIAM.

are all spherical caps; however, the radii $R$ of the caps vary considerably. According to Y1 above, the pressure change across each interface is $\Delta p = 2\sigma/R$. Since one may assume the outer region to be a vacuum with zero pressure for all caps, we find that the pressures within the fluids vary greatly among the configurations. One would thus expect that work is done within the fluid in passing from one configuration to another, a circumstance we have excluded by hypothesis when determining the family. From this point of view, the (customary) hypothesis of incompressibility that was used in determining the family is put into significant question; we examine this point in some detail in the section "Compressibility."

#### Discontinuous Dependence I

Capillary surfaces can exhibit striking discontinuous dependence on the defining data. As initial example, we consider the behavior of a solution of [18]–[17] at a protruding corner point $P$ of the domain $\Omega$ of definition. For simplicity, we assume the corner bounded locally by straight segments, meeting in an opening angle $2\alpha < \pi$, thus forming locally a wedge domain. In anticipation of material to follow, we assume contact angles $\gamma_1$ and $\gamma_2$ on the respective sides, $0 \leq \gamma_1, \gamma_2 \leq \pi$. One can show that a necessary condition for a solution surface over a domain $\Omega_\delta$ as in Figure 11 to have a continuous normal vector up to $P$ is that the data point $(\gamma_1, \gamma_2)$ lie in the closure of the rectangle $R$ of Figure 12. (This figure includes also additional material anticipating the section "Drops in wedges").

For data points interior to $R$, this criterion also suffices for the existence of at least one such solution surface, for any prescribed $H$; such surfaces can in fact be produced explicitly as spherical caps (planes if $H = 0$). It remains to discuss what can occur with data arising from the remaining four subregions of the square.

If $(\gamma_1, \gamma_2) \in D_1^{\pm}$, then there is no solution to [18]–[17] in any neighborhood of the corner point $P$. On the other hand, an explicit solution for any $H > 0$ can be found as a lower spherical cap on the segment $\gamma_1 + \gamma_2 = \pi - 2\alpha$ that separates $D_1^+$ from $R$ (see Figure 13, which indicates the equatorial circle). Correspondingly, if $H < 0$ then an explicit solution can be found on the separation line between $D_1^-$ and $R$. Thus, there is a

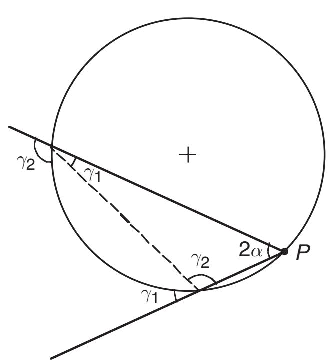


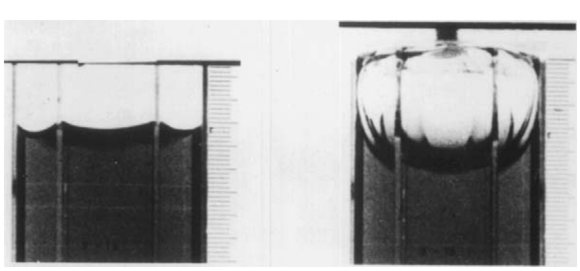

**Figure 13** Construction of solution as lower hemisphere; $\gamma_1 + \gamma_2 = \pi - 2\alpha, H > 0$. Reproduced from "Capillary Wedges Revisited" in *SIAM J. Math. Anal.* 27 No.1 (1996) 56–69 with permission from SIAM.

discontinuous change in behavior in crossing from $R$ to either of the $D_1$ regions.

This behavior was put to experimental test by W Masica, who considered the case $0 < \gamma_1 = \gamma_2 = \gamma < \pi/2$ near the crossing point $\gamma = \gamma_{cr}$ with $D_1^+$, for which $\alpha + \gamma_{cr} = \pi/2$. He partially filled a regular hexagonal cylinder of acrylic plastic, successively with two different liquids, making respective contact angles greater or less than $\gamma_{cr}$ with the plastic. For each liquid, Masica then allowed the cylinder to fall in a 132 m drop tower. Figure 14 compares the two configurations after about 5 s of free fall. In the case $\gamma > \gamma_{cr}$ he obtained the spherical-cap solution, which in this case covers the entire base domain $\Omega$ and appears as an explicit solution of [18]–[17]. When $\gamma < \gamma_{cr}$, the liquid rose to the top of the cylinder near the edges, filling out the edges over the corner points. The surface interface $S$ does not cover $\Omega$, but instead folds back over itself, doubly covering a portion of $\Omega$. Thus, a physical surface appears as it must, but it is not a solution of [18] over $\Omega$.

### Discontinuous Dependence II

About 1970, M Miranda raised informally the question, whether a capillary tube $Z_0$, whose section



(a) (b)

**Figure 14** Liquid in hexagonal cylinder, during free fall in drop tower: (a) $\alpha + \gamma > \pi/2$; (b) $\alpha + \gamma < \pi/2$.

$\Omega_0$ lies strictly interior to a section $\Omega_1$ of a tube $Z_1$, will raise liquid from an infinite reservoir in a downward directed gravity field to a higher level over $\Omega_0$ than will $Z_1$ over that subdomain of its section. That is true if both cylinders are circular, and in the intervening years its correctness was established in a number of other cases of particular interest.

Finn and Kosmodem'yanskii, Jr. showed, however, by example that the assertion fails in a large range of cases, and in fact can fail with arbitrarily large height differences, uniformly over $\Omega_0$. Beyond that, the construction exhibits a strikingly discontinuous change of behavior, under perturbations of a disk as inner domain. Perhaps more remarkably, the assertion can hold with the inner domain a disk, but with discontinuous reversal of behavior as the disk is perturbed to neighboring disks. That was shown in a form of the example given later by Finn, and illustrated in Figure 15. Here the outer domain $\Omega_1$ is polygonal, with sides that extend to be tangent to a unit disk $\Omega_0$, as indicated. The angle $\gamma$ is to be chosen so that $0 \le \pi/2 - \gamma \le \alpha_{\min}$, where $\alpha_{\min}$ is the smallest of the interior vertex half-angles of $\Omega_1$. In view of the assumed infinite fluid reservoir, there is no volume constraint, and the governing equation [16] takes the form

$$\operatorname{div} Tu = \kappa u, \qquad \kappa = \rho g/\sigma > 0 \qquad [22]$$

Taking at first the inner domain to be $\Omega_0$, it can be shown that for the corresponding solutions $u^0$ and $u^1$ of [22], there holds $u^0 > u^1$ over $\Omega_0$ for

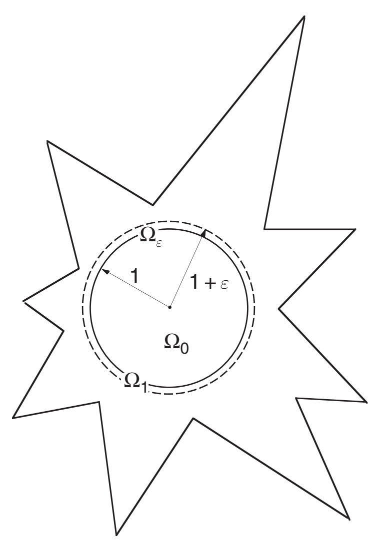



**Figure 15** Discontinuous reversal of limiting height behavior. All sides of the polygonal domain $\Omega_1$ are tangent to the unit disk $\Omega_0$. For the corresponding solution heights $u^0$ in $\Omega_0$, $u^\varepsilon$ in the disk $\Omega_\varepsilon$ of radius $1+\varepsilon$, and $u^1$ in $\Omega_1$, there holds $u^1 - u^0 < 0$, for any downward gravity. But $\lim_{\kappa\to 0}(u^1 - u^\varepsilon) = +\infty$, for any $\varepsilon > 0$.

any $\kappa > 0$, and thus the Miranda question has a positive answer for that configuration. But if we replace $\Omega_0$ by a concentric disk $\Omega_\varepsilon \subset \Omega_1$ of radius $1+\varepsilon$, we find

$$\left| \left\{ \inf_{\Omega_\varepsilon} u^1(x;\kappa) - \sup_{\Omega_\varepsilon} u^\varepsilon(x;\kappa) \right\} - \frac{2\varepsilon}{1+\varepsilon} \frac{\cos\gamma}{\kappa} \right| < \frac{1-\sin\omega}{\cos\gamma} + (1+\varepsilon)\frac{1-\sin\gamma}{\cos\gamma} \qquad [23]$$

where $\omega = \arccos(\cos\gamma/\sin\alpha)$, and $u^\varepsilon$ is the solution of [22], [17] in $\Omega_\varepsilon$. Since $\kappa$ does not appear on the right side of [23], there follows in particular that for any $\varepsilon > 0$, there holds

$$\lim_{\kappa\to 0} \left\{ \inf_{\Omega_\varepsilon} u^1(x;\kappa) - \sup_{\Omega_\varepsilon} u^\varepsilon(x;\kappa) \right\} = \infty \qquad [24]$$

In particular, a negative answer to Miranda's question appears for all gravity sufficiently small. But as observed above, a positive answer occurs in $\Omega_0$, for any positive gravity. Thus, the limiting behavior as $\kappa \to 0$ changes discontinuously, as $\varepsilon \to 0$. We find that the two limiting procedures cannot be interchanged: for any $x \in \Omega_0$, we obtain

$$\begin{aligned} &\lim_{\varepsilon\to 0}\lim_{\kappa\to 0}\left\{u^1(x;\kappa) - u^\varepsilon(x;\kappa)\right\} = +\infty. \\ &\lim_{\kappa\to 0}\lim_{\varepsilon\to 0}\left\{u^1(x;\kappa) - u^\varepsilon(x;\kappa)\right\} \equiv \text{const.} < 0 \end{aligned} \qquad [25]$$

## Existence Questions I

For the general equation [20] there is an established literature on existence of surfaces containing a prescribed space curve. There is very little literature relating to the capillarity boundary condition that the solution surface $S$ meet a prescribed "support" surface $W$ in a prescribed angle $\gamma$. The existence of at least one such surface interior to a prescribed sufficiently smooth closed space domain was proved by Almgren, and then Taylor proved smoothness at the contact curve. These are abstract theorems that are basic for the theory but in general do not provide specific information in particular cases of interest.

Special interest attaches to the nonparametric cases [16] or [18] with boundary condition [17], especially in view of the discontinuous behavior properties described above. These cases were studied in depth by a number of authors, with results that put the above examples into some perspective.

M Emmer proved the existence of a unique solution of [16]–[17] for any compact $\Omega$ having Lipschitz boundary with Lipschitz constant $L$ such that $\sqrt{1+L^2}\cos\gamma < 1-\varepsilon$ for some $\varepsilon > 0$. Finn and Gerhardt (F and G) extended this condition, and showed in particular that solutions exist in general in piecewise smooth $\Omega$. This result contrasts with the zero-gravity case [18] discussed in the section "Existence questions II," for which solutions fail to exist when $\sqrt{1+L^2}\cos\gamma > 1$ at a protruding corner (see the section "Discontinuous dependence I"). However, in the cases $\sqrt{1+L^2}\cos\gamma > 1$ studied by F and G the solution $u(x)$ is necessarily unbounded in the corner. This condition is equivalent to $\alpha < |\gamma - \pi/2|$ at the corner. Concus and Finn showed that if $\alpha \geq |\gamma - \pi/2|$ in a neighborhood $\Omega_\delta$ of a corner with rectilinear sides, as indicated in Figure 11, then the solution $u(x)$ satisfies

$$|u(x;\kappa)| < \frac{2}{\kappa\delta} + \delta \qquad [26]$$

independent of $\alpha, \gamma$ in the range considered. Here it is assumed that [16] is normalized so that $\lambda = 0$; when $\kappa \neq 0$ this can always be achieved by adding a constant to $u$. On the other hand, if $\alpha < |\gamma - \pi/2|$, then

$$u(x;\kappa) \approx \frac{\cos\vartheta - \sqrt{k^2 - \sin^2\vartheta}}{k\kappa r} \qquad [27]$$

where $k = \sin\alpha/\cos\gamma$ and $\vartheta$ is polar angle relative to a bisector at the vertex; hence $u$ becomes unbounded as $O(1/r)$. Thus, the behavior changes discontinuously as the configuration for which $\alpha = |\gamma - \pi/2|$ is crossed.

This prediction was corroborated by T Coburn in a "kitchen sink" experiment in the Medical School at Stanford University. Coburn formed a wedge using two sheets of acrylic plastic, resting on a glass plate, and inserted a drop of distilled water at the base of the wedge. Initially, the wedge was opened sufficiently that $\alpha + \gamma \geq \pi/2$, and he obtained the configuration of **Figure 16a**, with the maximum height slightly lower than that indicated by [26]. By closing down the angle slightly, the liquid rose to over ten times that height, as shown in **Figure 16b**. This experiment was later repeated by Weislogel under laboratory conditions; it incidentally establishes the contact angle of water and acrylic plastic in the Earth's atmosphere as $80° \pm 2°$.

The indicated procedure provides in general a very accurate way to measure contact angles, when the angle is not far from $\pi/2$. For $\gamma$ near zero or $\pi$ in the Earth's gravity field, the discontinuity is confined to a microscopic neighborhood of the vertex, and can be difficult to observe. This technical difficulty was addressed by Fischer and Finn, who introduced "canonical proboscis" domains, the theory of which was further developed by Finn and

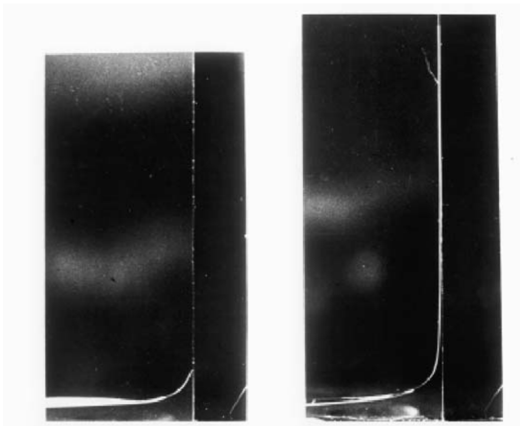



**Figure 16** Distilled water in wedges formed by acrylic plastic plates; $g > 0$. (a) $\alpha + \gamma > \pi/2$; (b) $\alpha + \gamma < \pi/2$. Reproduced from P Concus and R Finn, "On Capillary Free Surfaces in a Gravitational Field" in *Acta Math* 132 (1974) 207–223 with permission of Institut Mittag-Loeffler.

Leise and by Finn and Marek. For such domains the change in behavior is not strictly discontinuous, but it is nearly so, and it extends over large portions of the cylinder section, so that it is easily observable. Concus, Finn, and Weislogel conducted space experiments, demonstrating the feasibility of the method as a means for measuring contact angles in general ranges.

In [26]–[27] no growth conditions at the corner are imposed; the estimates hold for every solution defined in $\Omega_\delta$ and assuming the prescribed data on the side walls, with no data prescribed at the vertex. The formula [27] is the initial term of a formal asymptotic expansion of the solution, in powers of $r$. Miersemann obtained the complete expansion, asymptotic to every order, when $\alpha < |\gamma - \pi/2|$. He obtained somewhat less complete information in the bounded case [26].

Chen, Finn, and Miersemann provided a form of [27] that is applicable for any data $(\gamma_1, \gamma_2)$ on the respective sides of the wedge, that arise from the $D_1^{\pm}$ regions of **Figure 12**. Lancaster and Siegel and independently Chen, Finn, and Miersemann showed that if $-2\alpha \leq \gamma_1 + \gamma_2 - \pi \leq 2\alpha$, then every solution is bounded at the vertex. This result holds also for the zero gravity eqn [18].

In the case of [18], Concus and Finn showed that in the $D_1^{\pm}$ regions no solution exists, regardless of $H$. Again, this result holds without growth conditions.

From these considerations and from remarks in the section "Discontinuous dependence I" follows that for data in $D_2^{\pm}$, all solutions either of [18] or of [16] are bounded but have discontinuous derivatives at the vertex $P$. Extrapolating from the behavior of particular computed solutions, Concus and Finn conjectured that all solutions of [18] or of [16] that arise from data in $D_2^{\pm}$ are discontinuous at $P$. A number of attempts to prove or to disprove this conjecture have till now been unsuccessful.

An existence theorem for [16]–[17] alternative to that of Emmer was obtained independently by Ural'tseva, using a very different approach. This procedure yielded smoothness estimates up to the boundary, but required a hypothesis of boundary smoothness, so that the result does not mesh with the discontinuous dependence behavior as does that of Emmer. Later versions of the existence result, again under boundary smoothness requirements, were given by Gerhardt, Spruck, and Simon and Spruck. In the procedure introduced by Emmer, the boundary trace is shown to exist only in a very weak sense (which, however, suffices for a uniqueness proof). The later work can be adapted to show that the Emmer solutions are smooth on the smooth parts of $\partial\Omega$.

None of the above procedures provides existence for the zero gravity case [18]. As we shall see in the following section, that is not an accident of the methods, but reflects subtle properties of the equations.

## Existence Questions II

We consider here the zero-gravity case [18], over a domain $\Omega$ bounded by a piecewise smooth curve $\Sigma$, under the boundary condition [17]. Integrating [18] over $\Omega$ and using [17], we find $2H|\Omega| = |\Sigma| \cos\gamma$. Let $\Omega^* \subset \Omega, \Sigma^* = \Sigma \cap \partial\Omega^*, \Gamma = \Omega \cap \partial\Omega^*$. The same procedure over $\Omega^*$, using that $|Tu| < 1$ for any $u(x, y)$, leads to the bound

$$\Phi[\Gamma; \gamma] > 0 \qquad [28]$$

where $\Phi$ is defined by

$$\Phi[\Gamma; \gamma] \equiv |\Gamma| - |\Sigma^*| \cos\gamma + 2H|\Omega^*| \qquad [29]$$

The inequality [28] must hold for any choice of $\Omega^* \subset \Omega$. This provides a necessary condition for existence of a solution to [18]–[17] in $\Omega$. E Giusti showed that when $\Omega^*$ is interpreted in a generalized sense as a Caccioppoli set, the condition [28] becomes also sufficient for existence.

It is easy to give specific examples of convex analytic domains $\Omega$, in which subdomains $\Omega^*$ can be found such that [28] fails. Thus, the general existence results for [16] do not carry over to [18], regardless of local domain smoothness. Nevertheless, in many cases of interest (e.g., a circular disk or an ellipse that is not too eccentric), solutions of [18]–[17] do exist for any $\gamma$ and are well behaved. Finn investigated the condition [28] in general by showing the existence of a system of arcs $\{\Gamma\} \subset \Omega$

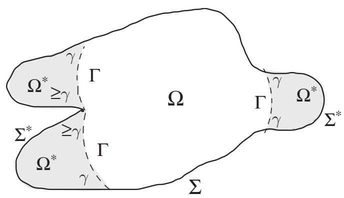



**Figure 17** Extremal configuration for the functional $\Phi$.

that minimize $\Phi$. All such arcs are circular of radius $1/2H$, and meet $\Sigma$ either at smooth points in an angle $\gamma$, or else at a reentrant corner point in an angle $\gamma^* \geq \gamma$, measured on the side of $\Gamma$ opposite to that into which the curvature vector points (Figure 17). All minimizing configurations are bounded by arcs of that form, although not all such configurations minimize. In a typical situation one will encounter only a finite number of such arcs, in which case only a finite number of cases need be examined. If $\Phi > 0$ in each such case, then a solution of [18]–[17] exists for the given $\Omega$ and $\gamma$. It may occur that no such arcs exist; we then observe that since $\Phi[\emptyset;\gamma] = \Phi[\Sigma;\gamma] = 0$, $\Phi$ cannot become nonpositive for any $\Omega^* \subset \Omega$ unless a minimizing $\Gamma$ can be found in $\Omega$, contradicting the assumed nonexistence of minimizers. Thus, the criterion is then vacuously satisfied, and we conclude that a solution of [18]–[17] exists.

One has, of course, to ask what happens physically in cases for which $\Phi[\Gamma;\gamma] \leq 0$ for some $\Gamma$ as above. The possible modes of behavior were studied in particular cases by Tam and later by deLazzer, Langbein, Dreyer, and Rath; Finn and Neel characterized the general case. Formally, the fluid rises to infinity throughout domains $\Omega^*$ of the form indicated, but with $H$ replaced by a value $H^- < H$; on the opposite side of the circular arcs $\Gamma$, the fluid is asymptotic to the vertical cylinders over $\Gamma$. In a physical situation, the fluid will rise to the top of the container in a nearly cylindrical region adjacent to a portion of the container walls, approximating the indicated behavior and partially wetting the top of the container. One sees that behavior in Figure 14b, in which the fluid fills out regions adjacent to the corners. An analogous configuration would still be observed if the corners were smoothed locally. If insufficient fluid is available, a portion of the base $\Omega$ could become unwetted.

### Behavior at a Corner Point

Lancaster and Siegel (L and S) studied the behavior of the limits (which they designate by $Ru$) of bounded solutions of [16] or of [18] along radial segments tending to a corner point $P$ of a domain $\Omega$. These limits can exhibit remarkable idiosyncratic behavior. For simplicity of exposition, we restrict ourselves here to rectilinear boundary segments at $P$, and assume constant boundary angles $\gamma_1, \gamma_2 \neq 0, \pi$ on the two sides. L and S prove first that the limits $Ru$ exist and vary continuously with direction of approach; then they show the existence of "fan" regions of directions adjacent to those of the sides, in which the limits are constant independent of direction, see Figure 18. They obtain that if the opening angle $2\alpha$ at $P$ satisfies $2\alpha < \pi$, then for data in the rectangle $R$ of Figure 12 the fans overlap (see Figure 18a), so that the solution is necessarily continuous at $P$. For data in $D_2^+$, the solution decreases from the $\gamma_1$ side $\Sigma_1$ to the $\gamma_2$ side $\Sigma_2$ ("D" behavior), subject to the Concus–Finn conjecture (see the section "Existence questions I"), with the reverse behavior ("I") in $D_2^-$. Concus and Finn showed that if $2\alpha < \pi$ then in $D_1^\pm$ there is no bounded solution of [16]–[17] or [18]–[17] as a graph. For [16]–[17], unbounded solutions do however exist for such data (see the section "Existence questions I").

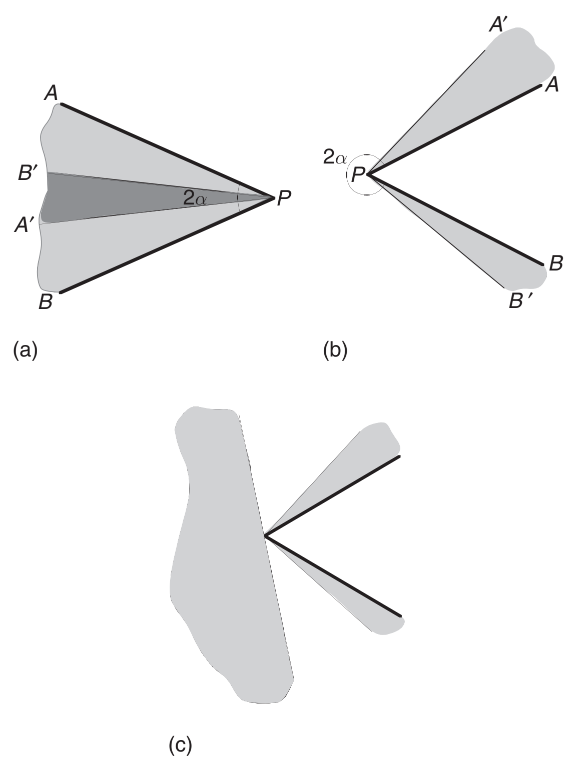



**Figure 18** (a) Fan domains $APA'$ and $BPB'$ of constant limiting values; $2\alpha < \pi$ so that the fans overlap when data are in $R$. (b) $2\alpha > \pi$; case 1. Fans $APA'$ and $BPB'$ of constant radial limits appear. Limiting value changes strictly monotonically as approach direction changes from $A'P$ to $B'P$. (c) $2\alpha > \pi$; case 2. In addition to the two fans adjacent to the sides of the wedge, a half plane of constant radial limits appears.

If $2\alpha > \pi$, then the fans do not overlap, and in fact continuity at $P$ cannot in general be expected. Outside the indicated fan regions adjacent to the wedge sides, the limit values either change strictly monotonically with angle of approach, as in Figure 18b, or else they do so except for approaches within a third, central fan, which covers a full half-space, and interior to which the limiting values again remain constant, see Figure 18c. L and S give an example under which that behavior actually occurs. Remarkably, in the example the prescribed data are the same on both boundary segments. The solution is nevertheless discontinuous at $P$, with an interval in which the radial limit increases, another interval in which it decreases, two fans of constant limit adjacent to the sides, and a fan of breadth $\pi$ in-between.

General conditions for continuity at a reentrant corner ($2\alpha > \pi$) have not yet been established. L and S give a sufficient condition, depending on a hypothesis of symmetry. Since no such hypothesis is needed when $2\alpha < \pi$, one might at first expect it to be superfluous. However, Shi and Finn showed that by introducing an asymmetric domain perturbation that in an asymptotic sense can be arbitrarily small, the solution can be made discontinuous at $P$. That can be done without affecting any other hypotheses of the L and S theorem.

In as yet unpublished work, D Shi characterized all possible behaviors at a reentrant corner, subject to the validity of the Concus–Finn conjecture at a protruding corner. If $\kappa \geq 0$ then all solutions of [16] or of [18] in a neighborhood of $P$ in $\Omega$ are bounded at $P$. The further behavior depends on the particular data, and is indicated in Figure 19. Note the analogy with Figure 12, although the interpretations in the figures differ in detail. Here the symbol I denotes strictly increasing from the side $\Sigma_1$ to $\Sigma_2$, except on the fan regions of constant limits; ID denotes constancy on a fan adjacent to $\Sigma_1$, then strictly increasing, then constancy on a fan of opening $\pi$, then strictly decreasing, then constancy on a fan adjacent to $\Sigma_2$. D and DI are defined analogously. All cases can be realized in particular configurations.

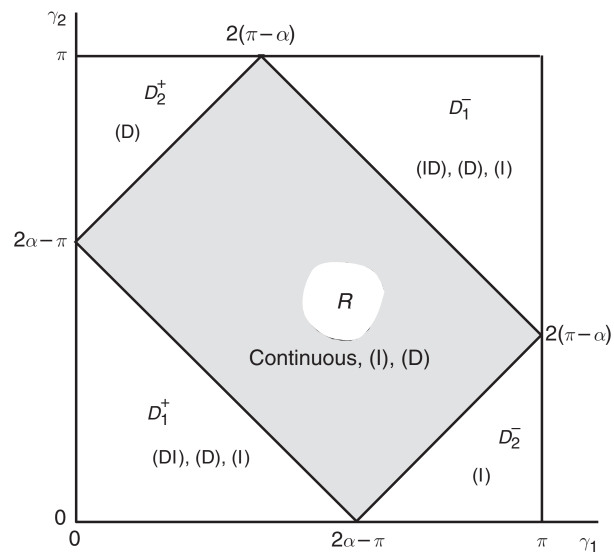



**Figure 19** $\pi < 2\alpha < 2\pi$. Possible modes of behavior. Reproduced with permission from the *Pacific Journal of Mathematics.*

#### Drops in Wedges

Closely related to the material just discussed is the question of the possible configurations of a connected drop of liquid placed into a wedge formed by intersecting plates of possibly differing materials, in the absence of gravity. Thus, one has distinct contact angles $\gamma_1, \gamma_2$ on the two plates. Finn and McCuan showed that if $(\gamma_1, \gamma_2) \in R$ then the only possibility is that the drop surface $S$ is part of a sphere. For data in $D_1^{\pm}$, no such drop can exist, barring exotically singular behavior at the vertex points where the edge of the wedge meets $S$.

For data in $D_2^{\pm}$ the situation is less clear. Concus, Finn, and McCuan (CFM) showed that local behavior exhibiting such data is indeed possible; however, they conjectured that such behavior cannot occur for simple drops. In conjunction with the above results, they were led to the conjecture that the free surface $S$ of any liquid drop in a planar wedge, that meets the wedge in exactly two vertices and the wedge faces in constant contact angles $\gamma_1, \gamma_2$, is necessarily spherical. Here it is supposed only that $0 \leq \gamma_1, \gamma_2 \leq \pi$.

The behavior of a drop of prescribed volume, as the data move from the midpoint of $R$ to the $D$ regions along parallels to the sides of $R$, is displayed in Figure 20. As one moves into the $D_2^{\pm}$ regions, the drop detaches from one side of the wedge and becomes a spherical cap resting on a single planar surface, in accord with the above conjecture. As $D_1^-$ is approached, the liquid becomes a drop of very large radius that fills out a long thin region in the wedge, and disappears to infinity as the boundary of $R$ is crossed. However, as $D_1^+$ is entered, the configuration transforms smoothly into a spherical liquid bridge, connecting the two faces of the wedge without contacting the wedge line.

#### Stability Questions

A number of authors, for example, Langbein, Vogel, Finn and Vogel, Steen, and Zhou, have studied the

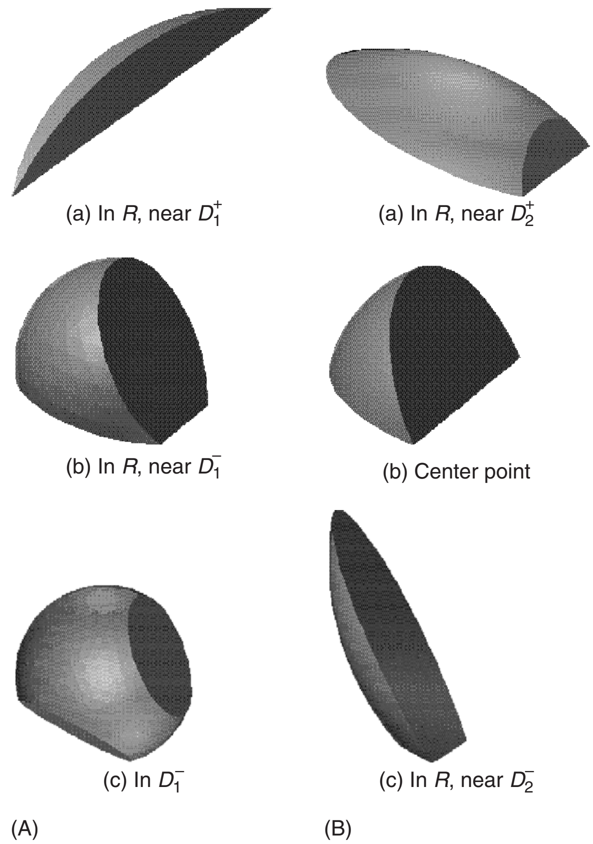



**Figure 20** (A) Drop configurations in wedge with opening angle $2\alpha = 50°$, for three data positions on the line $\gamma_1 = \gamma_2 = \gamma$ (a) $\gamma = 70°$ (in $R$, near $D_1^+$); (b) $\gamma = 90°$ (in $R$, near $D_1^-$); (c) $\gamma = 110°$ (in $D_1^-$). The first two cases yield edge blobs, the third a spherical tube that does not contact the edge line. (B) Drop configurations in a wedge of opening angle $2\alpha = 50°$, for three data choices in $R$, on the line $\gamma_1 = \pi - \gamma_2 = \gamma$; (a) $\gamma = 70°$ (near $D_2^+$); (b) $\gamma = 90°$ (center of $R$); (c) $\gamma = 35°$ (near $D_2^-$). As $D_2^{\pm}$ is entered, original boundary conditions can no longer be satisfied by spherical drop, but configuration changes smoothly into drop on single plane, with prescribed data for that plane. Reproduced with permission from Concus P, Finn R and McCuan J (2001) Liquid bridges, edge blobs, and Scherk-type capillary surfaces. *Indiana University Mathematics Journal* 50: 411–441.

stability of liquid drops trapped between parallel plates, forming an annular liquid bridge joining the plates under the capillarity boundary condition of prescribed contact angles $\gamma_1, \gamma_2$ on the respective plates. These studies consider the effects of disturbances within the fluid, assuming the plates are rigid and perfectly parallel. CFM show that from the point of view of physical prediction, the results of these studies may be open to some question. Specifically, they show that unless the drop is initially of spherical form, then infinitesimal tilting of one of the plates always results in a discontinuous transition of the drop form. Depending on the particular data, the transition can be to a spherical drop; however, it can also occur that the tilting causes the entire fluid to disappear to infinity in the wedge.

CFM proved that if a connected liquid mass with spherical outer surface $S$ cuts off areas $|W_1|, |W_2|$ from plates $\Pi_1, \Pi_2$ which it meets in angles $\gamma_1, \gamma_2$, as in Figure 20, then

$$-\sum_1^2 |W_j| \cos \gamma_j + |S| = \frac{3|V|}{R} \qquad [30]$$

where $|S|$ denotes area of the spherical free surface interface, $|V|$ the enclosed volume, and $R$ the radius. An immediate consequence is that the mechanical energy $E$ of the configuration is

$$E = \frac{3\sigma|V|}{R} \qquad [31]$$

where $\sigma$ is surface tension. Using this result, they show that if a spherical liquid mass meets two wedge faces in angles $\gamma_1, \gamma_2$ in the absence of gravity, then the configuration has smaller mechanical energy than does any connected liquid mass of the same volume that meets only one of the faces in the contact angle for that face. In turn, the drop on a single face has smaller energy than does a spherical ball of the same volume that meets no face. Note that in all zero-gravity cases for which stability relative to plate tilting can be expected, the liquid mass must be spherical.

## Compressibility

Until very recently, all literature on capillarity was based on a hypothesis that the body of the fluid is incompressible. Indeed, from the point of view of macroscopic mechanical measurements, most liquids are nearly incompressible. But all liquids are also to some extent compressible, and this property was even conceptually essential in our characterization in the section "Gauss' contribution: the energy method" of the surface energy, even for the nominally incompressible case. It is as yet unclear to what extent the compressibility properties of the bulk liquid will influence the physical predictions of the theory. In this connection, see the remarks at the end of the section "Uniqueness and nonuniqueness."

### The Equations I

Finn derived two possible equations extending [16] and [17], arising from different modelings. Both characterize equilibrium points as stationary points for the mechanical energy, and both are based on a hypothesized pressure–density relation $\rho = \rho_0 + \chi(p - p_0)$. The first equation takes account of the change in density with height, arising from

the gravity field. For a container consisting of a semi-infinite vertical cylinder, closed at the bottom, one obtains

$$\operatorname{div} Tu = \frac{\rho_0 g}{\sigma} u + \chi g(1 - \cos\omega) + \lambda \qquad [32]$$

where $\omega$ is the angle between the upward directed surface normal and the vertical axis, and $\lambda$ is to be determined by a volume constraint. Athanassenas and Finn proved that for a general smooth domain $\Omega$, prescribed $\gamma$, and prescribed fluid mass $M$ subject to the restriction

$$M < \rho_0 |\Omega| / \chi g \qquad [33]$$

there exists exactly one solution of [32] achieving the boundary data $\gamma$.

The condition [33] is necessary for existence with the prescribed mass.

The methods used for this theorem do not permit regularity conditions to be relaxed to allow domains with corner points. An approximation procedure yields an existence theorem for such cases, however the uniqueness proof then fails; it can be replaced by a weaker result, estimating the difference between two eventual solutions: Let $u$, $v$, be solutions of [32] in a piecewise smooth domain $\Omega$, and suppose $\nu \cdot Tu \le \nu \cdot Tv$ on $\Sigma = \partial\Omega$ except at the corner points, where no data are prescribed. Then

$$u \le v + \chi\sigma/\rho_0 \qquad [34]$$

throughout $\Omega$.

Note that in this result, no growth condition is imposed at the corner points. It can happen that both $u$ and $v$ are unbounded at a corner point; nevertheless, [34] holds uniformly over $\Omega$.

The solutions of [32] emulate many of the characteristics of solutions of [16]. Notably, there is again a dichotomy of behavior, depending on opening angle $2\alpha$ at a corner point, with all solutions either bounded, or unbounded with growth like $1/r$.

### The Equations II

If in addition to taking account of the change of density with height, one accounts for the energy change due to expansion or contraction of volume elements with changing density, one is led to the equation

$$\operatorname{div} Tu = \frac{\rho_0 - \chi p_0}{\sigma\chi}(e^{\chi g u} - 1) + \chi g(1 - \cos\omega) + \Lambda \qquad [35]$$

Here the changes from the incompressible case are much more significant than for [32]. In order to ensure stable behavior of solutions, it seems appropriate to impose the condition $\rho_0 > \chi p_0$. The general existence theorem above can no longer be expected; it is possible to give explicit examples of analytic domains, and constant data $\gamma$, for which no solution of the problem exists. Thus, even in a large downward gravity field, the solutions can emulate the behavior of solutions of [18]. That can happen, however, only for data $\gamma$ exceeding $\pi/2$. The condition [33] is again necessary for existence.

For eqn [34], $\Lambda$ cannot be eliminated by addition of a constant to the solution, and its determination creates a new level of difficulty toward solution of the physical existence question. Athanassenas and Finn proved unique existence of solutions of [35], [17] for a capillary tube of general smooth section $\Omega$ dipped into an infinite liquid bath (which corresponds to $\Lambda = 0$), when $0 \le \gamma \le \pi/2$. If $\gamma > \pi/2$ then solutions do not always exist; it can happen that the surface moves down to the bottom of the tube, regardless of the depth of immersion. Under a hypothesis of radial symmetry, Finn and Luli were able to prove the existence of solutions with prescribed mass in a semi-infinite cylinder closed at the bottom, in the range $0 \le \gamma < \pi$, and uniqueness if $0 \le \gamma \le \pi/2$. Note that in this case, values $\gamma > \pi/2$ are not excluded. For large enough mass, the surface will always cover the base of the tube.

## Closing Remarks

This brief survey is intended only as a general indication of the current state of the theory; much material of interest could not be included. Nor have we addressed hysteresis effects on contact angle. Detailed references to the material discussed and also to further information can be found in the articles listed below. More recent publications can be located by following links in MathSciNet or Zentralblatt.

## Acknowledgmnt

I owe a special debt of thanks to my colleague Paul Concus, who read the material in detail and provided many effectual suggestions, leading to a much-improved exposition.

*See also:* Compressible Flows: Mathematical Theory; Interfaces and Multicomponent Fluids; Newtonian Fluids and Thermohydraulics.

## Further Reading

References for text material and for further reading are cited in the expository articles:

Finn R (2002a) *Milan Journal of Mathematics* 70: 1–23.
Finn R (2002b) *Mathematical Intelligencer* 24: 21–33.

**Cartan Model** *see* Equivariant Cohomology and the Cartan Model

# Cauchy Problem for Burgers-Type Equations

**G M Henkin**, Université P.-M. Curie, Paris VI, Paris, France



## Burgers Type Equations

We consider here two types of equations: the scalar partial differential equations (PDEs) of the form

$$\frac{\partial f}{\partial t} + \varphi(f)\frac{\partial f}{\partial x} = \varepsilon\frac{\partial^2 f}{\partial x^2}, \quad \varepsilon > 0 \qquad [1]$$

$f = f(x, t)$, $x \in \mathbb{R}$, $t \in \mathbb{R}_+$, and the scalar difference–differential equations of the form

$$\frac{\partial F}{\partial t} + \varphi(F)\frac{F(x,t) - F(x-\varepsilon,t)}{\varepsilon} = 0, \quad \varepsilon > 0 \qquad [2]$$

$F = F(x, t)$, $x \in \mathbb{R}$, $t \in \mathbb{R}_+$.

Equation [1] for the case of linear $f \mapsto \varphi(f)$ was called as Burgers equation by Hopf (1950), who justified this by the remark: "equation was first

$$\frac{\partial f}{\partial t} + f\frac{\partial f}{\partial x} = \varepsilon\frac{\partial^2 f}{\partial x^2}$$

introduced by J. M. Burgers (1940) as a simplest model to the differential equations of fluid flow". In fact, eqn [1] for linear $\varphi(f)$ was introduced earlier in 1915 by Bateman. Equation [1] for general $\varphi(f)$ appeared later in very different models, for example, in the model for displacement of oil by water, in a model of road traffic, etc.

For $\varphi(f) = a + b \cdot f$, Hopf and Cole have studied [1] basing on the substitution

$$f = -\frac{1}{b}\left(a + \varepsilon\frac{\partial g}{\partial x}\Big/ g\right)$$

reducing [1] to the heat equation

$$\frac{\partial y}{\partial t} = \varepsilon\frac{\partial^2 g}{\partial x^2}$$

This transformation (often called as the Hopf–Cole transform) appeared for the first time in 1906 in the book of Forsyth "Theory of differential equations."

Equation [2] first appeared for $\varphi(F) = a + b \cdot F$, $\varepsilon = 1$, $x = n \in \mathbb{Z}$, in Levi, Ragnisco, Bruchi (1983) as a semidiscrete equation reducible to the linear equation

$$\frac{\mathrm{d}G_n(t)}{\mathrm{d}t} = a(G_{n-1}(t) - G_n(t))$$

by the substitution

$$F(n,t) = -\frac{a}{b}\left(\frac{G_n(t) - G_{n-1}(t)}{G_n(t)}\right)$$

Equation [2] for general $\varphi(F)$ was introduced by Henkin, Polterovich (1991) for the description of a Schumpeterian evolution of industry. For any $\varepsilon > 0$, one can consider [2] as the family of difference–differential equations, depending on the parameter $\theta = \{x/\varepsilon\} \in [0, 1)$, where $\{x/\varepsilon\}$ denotes the fractional part of $x/\varepsilon$. For physical applications of [1] (see Gelfand (1959), Landan and Lifschitz (1968), Lax (1973)), the inviscid case ($\varepsilon = +0$) is the most interesting. But, for some special physical models and for some social and biological applications (see Henkin, Polterovich (1991), Serre (1999)), the interesting case concerns eqn [2] with $\varepsilon = 1$ and $x \in \mathbb{Z}$.

The results considered in this article concern mainly the Cauchy problem for eqns [1] and [2] with initial data $f(x,0)$, $F(x,0)$ satisfying the conditions

$$\begin{gathered} f(x,0) \to \alpha^{\pm}, \quad x \to \pm\infty \\ \int_{-\infty}^{0} |f(x,0) - \alpha^-|\mathrm{d}x \\ + \int_0^{\infty} |\alpha^+ - f(x,0)|\mathrm{d}x < \infty \end{gathered} \qquad [3]$$

and correspondingly

$$\begin{gathered} F(k\varepsilon + \theta\varepsilon) \to \alpha^{\pm}, \quad k \to \pm\infty \\ \sum_{k=-\infty}^{0} |F(k\varepsilon + \theta\varepsilon, 0) - \alpha^-| \\ + \sum_{k=0}^{\infty} |\alpha^+ - F(k\varepsilon + \theta\varepsilon, 0)| < \infty \end{gathered} \qquad [4]$$

where $\alpha^- \le \alpha^+$, $\theta \in [0, 1)$ and the mapping $\theta \mapsto \{F(k\varepsilon + \theta\varepsilon, 0) - \alpha^{\operatorname{sgn} k}, k \in \mathbb{Z}\} \in l^1$ is smooth.

The standard classical questions concerning Cauchy problems [1], [3] and [2], [4], namely those relating to existence, unicity, regularity, and conservation laws are well established (see Oleinik (1959), and Serre (1999)). This section formulates only those which are essential for the study of asymptotic behavior of solutions $f(x,t)$ and $F(x,t)$, when $t \to \infty$ or $\varepsilon \to 0$, and of the relation between vanishing viscosity and difference scheme approximations for inviscid Burgers type equations.

One can see that asymptotic behavior of solutions of [2], [4] when $\varepsilon \to +0$ is not the same as the asymptotic behavior of [1], [3] when $\varepsilon \to +0$, in spite of fact that in the limiting case $\varepsilon = +0$ both [1] and [2] look identical. It can be explained by the fact that eqn [2] can be interpreted as a semidiscrete approximation of the nonconservative (nonphysical) equation

$$\frac{\partial F}{\partial t} + \varphi(F)\frac{\partial F}{\partial x} = \frac{\varepsilon}{2}\varphi(F)\frac{\partial^2 F}{\partial x^2}$$

However, the problem [2], [4] can be naturally transformed into conservative (physical) initial problem. Indeed, the substitution

$$f = \int_0^F \frac{\mathrm{d}y}{\varphi(y)}$$

(under condition of integrability of $1/\varphi(y)$) transforms [2] into the equation

$$\frac{\partial f(x,t)}{\partial t} + \frac{\psi(f(x,t)) - \psi(f(x-\varepsilon,t))}{\varepsilon} = 0 \qquad [5]$$

where $\psi'(f) = \varphi(F)$. Equation [5] is the so-called monotone one-sided semidiscrete approximation of conservative viscous equation,

$$\frac{\partial f}{\partial t} + \varphi(F)\frac{\partial f}{\partial x} = \frac{\varepsilon}{2}\frac{\partial}{\partial x}\left(\varphi(F)\frac{\partial f}{\partial x}\right) \qquad [6]$$

where

$$f(x,0) \to \int_0^{\alpha^{\pm}} \frac{\mathrm{d}y}{\varphi(y)}, \quad x \to \pm\infty$$

The results of finite-difference approximations for nonlinear conservation laws (see A. Harten, J. Hyman, P. Lax (1976)) explain both the similarity of behavior of [6] and [5] as well as some difference in the behavior of [1] and [2].

For further exposition the following assumption is useful:

**Assumption 1** Let $\varphi$ in [1], [2] be a positive and continuously differentiable function on the interval $[\alpha^-, \alpha^+]$. Let $\varphi'$ have only isolated zeros.

From references one can deduce the following general properties of Cauchy problems [1], [3] and [2], [4].

**Theorem 0** *Under Assumption 1, we have:*

(i) *There exists a unique (weak) solution $f(x,t)$, $x \in \mathbb{R}$, $t \in \mathbb{R}_+$ of the problem [1], [3]; this solution is necessarily smooth for $t > 0$; besides, it satisfies the following conservation laws for $t > 0$:*

$$f(x,t) \to \alpha^-, \quad x \to -\infty$$
$$f(x,t) \to \alpha^+, \quad x \to +\infty$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\int_0^{\infty}(\alpha^+ - f(x,t))\right]\mathrm{d}x - \int_{-\infty}^0 (f(x,t) - \alpha^-)\mathrm{d}x = \int_{\alpha^-}^{\alpha^+} \varphi(y)\mathrm{d}y$$

*Moreover, if the initial value $f(x,0)$ is nondecreasing as a function of $x$, then solution $f(x,t)$ is nondecreasing as a function of $x$ for all $t \geq 0$.*

(ii) *There exists a unique solution $F(x,t)$ $x \in \mathbb{R}$, $t \in \mathbb{R}_+$ of the problem [2], [4]; this solution is smooth for $t > 0$; besides, it satisfies the following conservation laws for $t > 0$ and $\theta \in [0, 1)$:*

$$F(k\varepsilon + \theta\varepsilon, t) \to \alpha^-, \quad k \to -\infty$$
$$F(k\varepsilon + \theta\varepsilon, t) \to \alpha^+, \quad k \to +\infty$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\sum_{k=1}^{\infty}\int_{F(k\varepsilon+\theta\varepsilon,t)}^{\alpha^+}\frac{\mathrm{d}y}{\varphi(y)} - \sum_{k=-\infty}^{0}\int_{\alpha^-}^{F(k\varepsilon+\theta\varepsilon,t)}\frac{\mathrm{d}y}{\varphi(y)}\right] = \alpha^+ - \alpha^-$$

*Moreover, if for some $\theta \in [0, 1)$ the $F(k\varepsilon + \theta\varepsilon, 0)$ is nondecreasing as a function of $k \in \mathbb{Z}$ then solution $F(k\varepsilon + \theta\varepsilon, t)$ is also nondecreasing as a function of $k \in \mathbb{Z}$ for all $t \geq 0$ and the same $\theta$.*

## Gelfand's Problem and Iljin–Oleinik Theorem

The main results considered in this article are related to the following problem, formulated explicitly by Gelfand (1959): to find the asymptotic ($t \to \infty$) of the solution $f(x, t)$ of the eqn [1] with the initial condition

$$f(x,0) = \begin{cases} \alpha^{\pm}, & \text{if } \pm x > \pm x^{\pm} \\ f^0(x), & \text{if } x \in [x^-, x^+] \end{cases} \qquad [7]$$

where $\alpha^- \leq \alpha^+$.

Gelfand found a solution to this problem for the inviscid case $\varepsilon = +0$ with initial conditions $f(x, 0) = \alpha^-$ if $x < 0$, and $f(x, 0) = \alpha^+$ if $x \geq 0$ (see below), and remarked that it would be interesting to prove that the main term of the asymptotic ($t \to \infty$) of $f(x, t)$ satisfying [1], [7] coincides with the solution of [1], [7] for $\varepsilon = +0$.

Gelfand's problem admits natural extension for eqn [2] with the initial conditions

$$F(x,0)=\alpha^{\pm}, \qquad \text{if } \pm x > \pm x^{\pm}$$
$$F(x,0)=F^{0}(x), \quad \text{if } x\in[x^{-},x^{+}] \qquad [8]$$

Let us introduce, for $u \in [\alpha^-, \alpha^+]$, the function $\psi(u) = -\int_{\alpha^-}^{u} \varphi(y)\mathrm{d}y$. Let the function $\hat{\psi}(u)$, $u \in [\alpha^-, \alpha^+]$, be upper bound of the convex set

$$\{(u,v)\colon v\le \psi(u),\ u\in[\alpha^{-},\alpha^{+}]\}$$

By Assumption 1, the set $s=\{u\in[\alpha^-, \alpha^+]\colon \psi(u) < \hat{\psi}(u)\}$ is the finite union of intervals, $s=(\alpha^-,\beta_0)\cup(\alpha_1,\beta_1)\cup\cdots(\alpha_L,\alpha^+)$, where $\alpha^-=\alpha_0\le\beta_0\le\alpha_1<\beta_1\cdots\le\alpha_L\le\beta_L=\alpha^+$.

Let us define the function $\hat{f}(x, t)$ by

$$\hat{f}(x,t)=\begin{cases}\alpha^{-}, & \text{if } x<\varphi(\alpha^{-})\cdot t\\ (\hat{\psi}')^{(-1)}(x/t), & \text{if } \varphi(\alpha^{-})\cdot t\le x\le\varphi(\alpha^{+})\cdot t\\ \alpha^{+}, & \text{if } x>\varphi(\alpha^{+})\cdot t\end{cases}$$

where in the case $\hat{\psi}'(u)\equiv\xi_l$, $u\in(\alpha_l,\beta_l)$, $l=0, 1,\ldots,L$; also, by definition, $(\hat{\psi}')^{(-1)}(\xi_l)=[\alpha_l,\beta_l]$.

**Theorem 1** (Gelfand) *The solution* $f(x, t)$ *of the problem* [1], [7] *for the case* $\varepsilon=+0$ *and initial conditions* $f(x,0)=\alpha^{\pm}$, *if* $\pm x>0$, *has the explicit form*: $f(x, t)=\hat{f}(x, t)$.

The analogous statement is valid also for the problem [2], [8] if, in the construction above, one takes

$$\Psi(u)=\int_{0}^{u}\frac{\mathrm{d}y}{\varphi(y)}$$

instead of $\psi(u)$, $u\in[\alpha^-, \alpha^+]$.

The Gelfand problem for [1], [3] and [1], [7] with monotonic $\varphi(f)$ was solved by Iljin and Oleinik (1960). In the case $\alpha^-=\alpha^+$, the solution of this problem follows from an earlier work of Lax (1957). For the case of linear $\varphi(f)$, the solution of this problem follows from an earlier work of Hopf (1950).

For semidiscrete initial problems [2], [4] and [2], [8], the analog of the asymptotic results of Hopf and Iljin–Oleinik have been obtained and applied by Henkin and Polterovich (1991).

The case of increasing $\varphi(f)$ has been studied in detail. In this case, for both initial problems [1], [3] and [2], [4], there is uniform convergence of solutions $f(x, t)$ and $F(x, t)$ to the so-called rarefaction profile

$$g(x/t)=\begin{cases}\alpha^{\pm}, & \pm x>\varphi(\alpha^{\pm})t\\ \varphi^{(-1)}(x/t), & x\in[\varphi(\alpha^{-})\cdot t,\varphi(\alpha^{+})\cdot t]\end{cases}$$

$t\to\infty$ (see Iljin and Oleinik (1960) and Henkin and Polterovich (1991)). More precise result in this case about convergence to the so-called N-wave has been obtained by Dafermos (1977) and Liu (1978).

For the case of a general $\varphi(f)$, in particular, for the case of nonincreasing $\varphi(f)$, we need the notion of shock profile. Following Serre (1999), three definitions can be introduced.

**Definition** The initial problem [1], [3] (correspondingly, [2], [4]) admits $(\alpha^-, \alpha^+)$-shock profile $(\alpha^-<\alpha^+)$ if there exists a traveling-wave solution of this equation, that is, of the form $f=\tilde{f}(x-ct)$ (correspondingly, $F=\tilde{F}(x-Ct)$), such that $\tilde{f}(x)\to\alpha^{\pm}$ when $x\to\pm\infty$ (correspondingly, $\tilde{F}(x)\to\alpha^{\pm}$ when $x\to\pm\infty$).

From the results of Gelfand (1959) and Oleinik (1959), it follows that initial problem [1], [3] admits $(\alpha^-, \alpha^+)$-shock profile iff

$$c=\frac{1}{\alpha^{+}-\alpha^{-}}\int_{\alpha^{-}}^{\alpha^{+}}\varphi(y)\mathrm{d}y < \frac{1}{u-\alpha^{-}}\int_{\alpha^{-}}^{u}\varphi(y)\mathrm{d}y, \quad \forall u\in(\alpha^{-},\alpha^{+}) \qquad [9]$$

From the results of Henkin and Polterovich (1991) and Belenky (1990), it follows that initial problem [2], [4] admits $(\alpha^-, \alpha^+)$-shock profile iff

$$\frac{1}{C}=\frac{1}{\alpha^{+}-\alpha^{-}}\int_{\alpha^{-}}^{\alpha^{+}}\frac{\mathrm{d}y}{\varphi(y)} > \frac{1}{u-u^{-}}\int_{\alpha^{-}}^{u}\frac{\mathrm{d}y}{\varphi(y)}, \quad \forall u\in(\alpha^{-},\alpha^{+}) \qquad [10]$$

In the case $\varepsilon=+0$, the equality in [9] and [10] is called the Rankine–Hugoniot condition, the inequality in [9] and [10] is called the entropy condition (or the Gelfand–Oleinik condition).

**Definition** For initial problem [1], [3] (correspondingly, [2], [4]) admitting $(\alpha^-, \alpha^+)$-shock profile and for $\varepsilon=+0$, we will call by shock waves the weak solutions of [1], [3] (correspondingly, [2], [5], [4]) of the form

$$\tilde{f}^{\alpha^{\pm}}(x-ct)=\alpha^{\pm}, \quad \text{if } \pm x\ge\pm ct$$
$$\tilde{F}^{\alpha^{\pm}}(x-Ct)=\alpha^{\pm}, \quad \text{if } \pm x\ge\pm Ct$$

where $c$, $C$ satisfy Rankine–Hugoniot and entropy conditions [9], [10].

**Definition** The $(\alpha^-, \alpha^+)$-shock profile for [1] (correspondingly, for [2]) is called strict if in addition to [9], [10] we have the Lax (1954) condition:

$$\varphi(\alpha^{+})<c<\varphi(\alpha^{-}) \qquad [11]$$

and correspondingly

$$\varphi(\alpha^{+})<C<\varphi(\alpha^{-}) \qquad [12]$$

The $(\alpha^-, \alpha^+)$-shock profile for [1] or [2] is called semicharacteristic if one of the inequalities in [11] or [12] is strict and the other is an equality. This profile is called characteristic if both inequalities in [11] or [12] are equalities.

One can check (Iljin and Oleinik 1960, Henkin and Polterovich 1991) that if in addition to Assumption 1 the function $\varphi$ on $[\alpha^-, \alpha^+]$ is nonconstant and nonincreasing then eqn [1] (correspondingly, [2]) admits a strict $(\alpha^-, \alpha^+)$-shock profile.

The main result of Iljin–Oleinik (1960) for eqn [1] and analogous statement of Henkin and Polterovich (1991) for eqn [2] can be presented as follows.

**Theorem 2**

(i) *Let the initial problem* [1], [3] *admit a strict* $(\alpha^-, \alpha^+)$*-shock profile* $\tilde{f}$. *Let* $f(x,t)$, $x \in \mathbb{R}$, $t \in \mathbb{R}_+$, *be a solution of* [1], [3]. *Then there exists* $d_0 \in \mathbb{R}$

$$\sup_{x\in\mathbb{R}} |f(x,t) - \tilde{f}(x - ct - d_0)| \to 0, \quad t \to \infty \quad [13]$$

*The value of* $d_0$ *is determined uniquely by relation*

$$\int_{-\infty}^{\infty} \{f(x,0) - \tilde{f}(x - d_0)\}\, dx = 0$$

(ii) *Let the initial problem* [2], [4] *admit a strict* $(\alpha^-, \alpha^+)$*-shock profile* $\tilde{F}$. *Let* $F(x, t)$, $x \in \mathbb{R}$, $t \in \mathbb{R}_+$ *be a solution of* [2], [4]. *Then there exists continuous function* $D_0(\theta)$, $\theta \in [0, 1)$, *such that*

$$\sup_{x\in\mathbb{R}} |F(x,t) - \tilde{F}(x - Ct - D_0(\{x/\varepsilon\})| \to 0, \quad t \to \infty \quad [14]$$

*The function* $D_0(\theta)$, $\theta \in [0, 1]$, *is determined uniquely from relation*

$$\sum_{k=-\infty}^{\infty} \{\Phi(F(n,0)) - \Phi(\tilde{F}(n - D_0))\} = 0$$

*where*

$$\Phi(F) = \int_F^A \frac{dy}{\varphi(y)}, \quad F < A, \ \tilde{F} < A$$

(iii) *If in conditions* (i) *and* (ii), *we take* $\varepsilon = +0$ *then there exist* $d_0$, $D_0$ *such that* $\forall \delta > 0$, *we have*

$$\begin{aligned} &\sup_{x \ge ct+d_0+\delta} |\alpha^+ - f(x,t)| + \sup_{x \le ct+d_0-\delta} |\alpha^- - f(x,t)| \to 0, \quad t \to \infty \\ &\sup_{x \ge Ct+D_0+\delta} |\alpha^+ - F(x,t)| + \sup_{x \le Ct+D_0-\delta} |\alpha^- - F(x,t)| \to 0, \quad t \to \infty \end{aligned} \quad [15]$$

*The values of* $d_0$ *and* $D_0$ *are determined by*

$$\int_{-\infty}^{d_0} (f(x,0) - \alpha^-)\, dx + \int_{d_0}^{\infty} (f(x,0) - \alpha^+)\, dx = 0$$

$$\int_{-\infty}^{D_0} (F(x,0) - \alpha^-)\, dx + \int_{D_0}^{\infty} (F(x,0) - \alpha^+)\, dx = 0$$

**Remarks**

(i) The statements of Theorem 2 give a positive answer to Gelfand's question for the case of initial problem [1], [3] and [2], [4], admitting strict shock profiles.

(ii) For linear $\varphi(f) = a + bf$, $a > 0$, $a + b\alpha^+ > 0$, $b < 0$, the traveling waves $\tilde{f}$, $\tilde{F}$ for [1], [3] and [2], [4] can be found explicitly:

$$\tilde{f} = \alpha^- + \frac{\alpha^+ - \alpha^-}{1 + \exp\{-p(x - ct)\}}$$

$$c = a + \frac{b}{2}(\alpha^+ + \alpha^-), \quad p = \frac{(\alpha^- - \alpha^+)b}{2\varepsilon}$$

$$\tilde{F} = \alpha^- + \frac{\alpha^+ - \alpha^-}{1 + \exp\{-P(x - Ct)\}}$$

$$C = b \Big/ \ln\frac{a + b\alpha^+}{a + b\alpha^-}, \quad P = \frac{\gamma}{\varepsilon}\ln\frac{a + b\alpha^-}{a + b\alpha^+}$$

where

$$b\gamma = (\alpha + b\alpha^-)\left(1 - \left(\frac{a + b\alpha^+}{a + b\alpha^-}\right)^{\gamma}\right)$$

(iii) For initial problems [1], [7] and [2], [8], $\alpha^+ > \alpha^-$, the asymptotic convergence statements [13]–[15] admit the precise asymptotic estimates (see Iljin and Oleinik (1960) for [1], [7]:

$$\sup_{x\in\mathbb{R}} |f(x,t) - \tilde{f}(x - ct - d_0)| = O(e^{-\gamma t}) \quad \gamma > 0, \varepsilon > 0 \quad [16]$$

$$\sup_{x\in\mathbb{R}} |F(x,t) - \tilde{F}(x - Ct - D_0(\{x/\varepsilon\}))| = O(e^{-\gamma t}) \quad \gamma > 0, \varepsilon > 0 \quad [17]$$

$$\begin{aligned} &f(x,t) = \alpha^\pm \quad \text{for} \quad \pm x > \pm(ct + d_0) \\ &t \ge t_0, \ \varepsilon = +0 \\ &F(x,t) = \alpha^\pm \quad \text{for} \quad \pm x > \pm(Ct + D_0) \\ &t \ge t_0, \ \varepsilon = +0 \end{aligned} \quad [18]$$

Theorem 2(i) is proved basing on the following idea. Let $f$ satisfy the initial problem [1], [3] and let

$\tilde{f}(x - ct + d_0)$ be $(\alpha^-, \alpha^+)$-shock profile for [1], satisfying condition [13]. Put

$$\delta(x,t) = \int_{-\infty}^{x} \{f(y,t) - \tilde{f}(y - ct - d_0)\}\mathrm{d}y$$

The function $\delta(x, t)$ satisfies the nonlinear parabolic equation

$$\frac{\partial \delta}{\partial t} + \varphi(\kappa\tilde{f} + (1-\kappa)f)\frac{\partial \delta}{\partial x} = \varepsilon\frac{\partial^2 \delta}{\partial x^2}$$

where $\kappa(x, t)$ is some smooth function of $(x, t)$ with values in $[0, 1]$.

Besides, by conservation law of Theorem 0(i), we have $\delta(x, t) \to 0$, $x \to \pm\infty$, $\forall t \geq 0$.

Estimates basing on maximum principle and appropriate comparison statements give that $\delta(x, t) \rightrightarrows 0$, $x \in \mathbb{R}$, $t \to \infty$. It implies that

$$f(x,t) - \tilde{f}(x - ct - d_0) \rightrightarrows 0, \quad x \in \mathbb{R},\ t \to \infty$$

Theorem 2(ii) is proved in a similar way. Let $F(n, t)$ satisfy the initial problem [2], [4] with $x = n \in \mathbb{Z}$, $\varepsilon = 1$, $\theta = \{x\} = 0$, and let $\tilde{F}(n - Ct - D_0)$ be $(\alpha^-, \alpha^+)$-shock profile for [2], satisfying condition [14]. Put

$$\Delta(n,t) = \sum_{-\infty}^{n}\{\Phi(F(n,t)) - \Phi(\tilde{F}(n - Ct - D_0))\}$$

Then function $\Delta(n, t)$ satisfies the semidiscrete parabolic equation

$$\frac{\mathrm{d}\Delta(n,t)}{\mathrm{d}t} = \varphi(\Phi^{(-1)}(\kappa\Phi(F) + (1-\kappa)\Phi(\tilde{F})))(\Delta(n-1,t) - \Delta(n,t))$$

where $\kappa(n, t)$ is some function with values in $[0, 1]$.

Besides, by conservation law of Theorem 0(ii), we have

$$\Delta(n,t) \to 0, \quad n \to \pm\infty,\ \forall\, t \geq 0$$

Estimates, basing on generalized maximum principle and comparison statements, give that $\Delta(n, t) \rightrightarrows 0$, $n \in \mathbb{Z}$, $t \to \infty$. It implies that

$$F(n,t) - \tilde{F}(n - Ct + D_0) \rightrightarrows 0, \quad n \in \mathbb{Z},\ t \to \infty$$

**Remark** For the cases of nonstrict shock profiles (characteristic or semicharacteristic) the statements of Theorem 2 are not valid. The reason is that, under initial conditions [3], [4] for any $d_0$ and $D_0$, we have

$$\int_{-\infty}^{\infty}\{f(x,o) - \tilde{f}(x - d_0)\}\mathrm{d}x = \infty$$

and, correspondingly,

$$\sum_{-\infty}^{\infty}\{\Phi(\tilde{F}(k\varepsilon + \theta\varepsilon - D_0) - \Phi(F(k\varepsilon + \theta\varepsilon, 0))\} = \infty$$

So, the crucial argument, related to conservation law, does not hold.

One can extend the important Theorems 2(i), 2(ii) for the case of nonstrict shock profiles in two different ways: by changing conditions of these theorems or by changing conclusions of these theorems.

The first method (started by Mei, Matsumura, and Nishihara in 1994) was completed by the following $L^1$-asymptotic stability result (Serre 2004).

**Theorem 3** (Freistühler–Serre). *Let eqns [1], [2] admit $(\alpha^-, \alpha^+)$-shock profiles and $\tilde{f}$, $\tilde{F}$ – the corresponding train-wave solutions of [1], [2]. Let $f(x, t)$, $F(n, t)$, $x \in \mathbb{R}$, $n \in \mathbb{Z}$, $t \in \mathbb{R}_+$ be solutions of eqns [1], [2] with such initial conditions that*

$$\int_{-\infty}^{\infty}|f(x,0) - \tilde{f}(x)|\mathrm{d}x < \infty$$

$$\sum_{-\infty}^{\infty}|F(n,0) - \tilde{F}(n)| < \infty$$

*Then*

$$\int_{-\infty}^{\infty}|f(x,t) - \tilde{f}(x - ct - d_0)|\mathrm{d}x \to 0$$

*and, correspondingly,*

$$\sum_{-\infty}^{\infty}|F(n,t) - \tilde{F}(n - Ct - D_0)| \to 0, \quad t \to \infty$$

*where constants $d_0$ and $D_0$ are calculated from the same relations as in Theorem 2.*

**Remark** For the inviscid case $\varepsilon = +0$, the statement of Theorem 3 is still valid for equations admitting strict shock profiles, but generally is not valid for equations admitting only nonstrict shock profiles (see Serre (2004)).

The second method permits, keeping initial conditions [3], [4], to localize the positions of viscous shock waves for generalized Burgers equations (see the next section).

## Asymptotic Behavior of Solutions of Generalized Burgers Equations

The main current interest and the main difficulty in the study of Gelfand's problem for generalized Burgers equations consist in the following question formulated explicitly for initial problem [1], [3] by Liu *et al.* (1998): “In the Cauchy problem there is

the question of determining the location of viscous shock waves". A similar question and related conjecture were formulated by Henkin and Potterovich (1999) for the initial problem [2], [4].

For solving this problem, it is important to solve it first for the Burgers type equations admitting nonstrict shock profiles.

**Theorem 4** (Henkin–Shananin–Tumanov).

(i) *Let the initial problem* [1], [3] *admit the nonstrict* $(\alpha^-, \alpha^+)$*-shock profile* [9] *and* $\tilde{f}(x - ct)$ *be a corresponding traveling-wave solution. Let*

$$\varphi'(\alpha^-) \neq 0, \quad \text{if} \quad \varphi(\alpha^-) = c$$
$$\varphi'(\alpha^+) \neq 0, \quad \text{if} \quad \varphi(\alpha^+) = c$$

*Let* $f(x, t)$ *be a solution of* [1], [3]. *Then there exist constants* $\gamma_0$ *and* $d_0$ *such that*

$$\sup_{x\in\mathbb{R}} |f(x,t) - \tilde{f}(x - ct - \epsilon\gamma_0 \ln t - d_0)| \to 0, \quad t \to \infty$$

*where*

$$(\alpha^+ - \alpha^-)\cdot\gamma_0 = \begin{cases} -1/\varphi'(\alpha^+), & \text{if } \varphi(\alpha^-) > c = \varphi(\alpha^+) \\ 1/\varphi'(\alpha^-), & \text{if } \varphi(\alpha^-) = c > \varphi(\alpha^+) \\ 1/\varphi'(\alpha^-) - 1/\varphi'(\alpha^+), & \text{if } \varphi(\alpha^-) = c = \varphi(\alpha^+) \end{cases}$$

(ii) *Let the initial problem* [2], [4] *with* $\epsilon = 1$ *admit the nonstrict* $(\alpha^-, \alpha^+)$*-shock profile* [10] *and* $\tilde{F}(n - Ct)$ *be a corresponding traveling-wave solution. Let*

$$\varphi'(\alpha^-) \neq 0, \quad \text{if } \varphi(\alpha^-) = C$$
$$\varphi'(\alpha^+) \neq 0, \quad \text{if } \varphi(\alpha^+) = C$$

*Let* $F(n, t)$ *be a solution of* [2], [4]. *Let*

$$\Delta F(n,0) \overset{\text{def}}{=} F(n,0) - F(n-1,0) \geq 0$$

*Then there exist constants* $\Gamma_0$ *and* $D_0$ *such that*

$$\sup_{n\in\mathbb{Z}} |F(n,t) - \tilde{F}(n - Ct - \Gamma_0 \ln t - D_0)| \to 0, \quad t \to \infty$$

*where*

$$(\alpha^+ - \alpha^-)\cdot\Gamma_0 = \begin{cases} -C/(2\varphi'(\alpha^+)), & \text{if } \varphi(\alpha^-) > C = \varphi(\alpha^+) \\ C/(2\varphi'(\alpha^-)), & \text{if } \varphi(\alpha^-) = C > \varphi(\alpha^+) \\ (C/2)[-1/\varphi'(\alpha^+) + 1/\varphi'(\alpha^-)], & \text{if } \varphi(\alpha^-) = C = \varphi(\alpha^+) \end{cases}$$

**Remarks**

(i) One could think that nonstrict shock profiles as in Theorem 4 can appear only in exceptional cases. But Proposition 2 and Theorem 5 below show, on the contrary, that characteristic shock profiles and, as a consequence, the behavior of initial problems [1], [3] and [2], [4] as in Theorem 4 are rather a rule than an exception.

(ii) The statement of Theorem 4(i) (and also of Theorem 5(i)) below) disprove the Gelfand hope that the main term of asymptotic ($t \to \infty$) of $f(x, t)$, satisfying [1], [7], coincides with the solution of [1], [7] for $\epsilon = +0$ with the same initial condition. Indeed, in conditions of Theorem 4, we have $\varphi(\alpha^-) = c$ or $\varphi(\alpha^+) = c$, but $\varphi'(\alpha^-) \neq \varphi'(\alpha^+)$; then for any $\epsilon > 0$ the traveling wave $\tilde{f}(x - ct - \epsilon\gamma_0 \ln t - d_0)$ for [1], [3], concentrated near the point $x_\epsilon(t) = ct + \epsilon\gamma_0 \ln t + d_0$, moves away ($t \to \infty$) from the shockwave for [1], [7] for $\epsilon = +0$, concentrated near the point $x_0(t) = ct + o(\ln t)$, where $o(\ln t)/\ln t \to 0, \ t \to \infty$.

(iii) Theorem 4 (and also Theorem 5 below) also illustrate another interesting phenomenon: for the case $\varphi'(\alpha^-) \neq \varphi'(\alpha^+)$, one has asymptotic convergence of the solution of [1], [3] (correspondingly of [2], [4]) to the traveling wave $\tilde{f}(x - ct - \epsilon\gamma_0 \ln t - d_0)$ (correspondingly $\tilde{F}(x - Ct - \epsilon\Gamma_0 \ \ln t - D_0)$), which does not satisfy eqn [1] or correspondingly eqn [2]. Such a phenomenon was first discovered by Liu and Yu (1997) in the special boundary-value problem for the classical Burgers equations, if $u(x, t)$ satisfies the following conditions:

$$\text{if } u_t + u\cdot u_x = u_{xx},\ u(0,t) = 1,\ u(\infty,t) = -1,$$
$$u(x,0) = -th\frac{x}{2}, \text{ then}$$
$$|u(x,t) + th\frac{1}{2}(x - \ln(1+t))| \to 0,\ t \to \infty,\ x \geq 0$$

Theorem 4 is proved in basing on the following idea. Let $f(x, t)$ satisfy [1], [3] and $F(n, t)$ satisfy [2], [4]. Let $\tilde{f}(x - ct)$ be the traveling wave for [1], [3] and $\tilde{F}(n - Ct)$ be the traveling wave for [2], [4]. Suppose that $\varphi(\alpha^-) > c = C = \varphi(\alpha^+)$. Let $d_A(t)$ and $D_A(t)$, $A > 0$ be functions such that

$$\int_{ct-A\sqrt{t}}^{ct+A\sqrt{t}} \{f(x,t) - \tilde{f}(x - ct - d_A(t))\}\mathrm{d}x = 0 \qquad [19]$$

and, correspondingly,

$$\sum_{k=[Ct-A\sqrt{t}]}^{[Ct+A\sqrt{t}]} \{\Phi(F(k,t)) - \Phi(\tilde{F}(k - Ct - D_A(t)))\} + (Ct + A\sqrt{t} - [Ct + A\sqrt{t}])(\Phi(F(Ct + A\sqrt{t}] + 1, t)) - \Phi(\tilde{F}([Ct + A\sqrt{t}] + 1 - Ct + D_A(t)) = 0 \qquad [20]$$

The relations [9], [20] can be called "localized conservation law." The proof contains two difficult parts.

The first part consists in proving that for $A > 2\sqrt{c}$ (correspondingly, $A > 2\sqrt{C}$) the following asymptotics are valid:

$$d_A(t) = \frac{\epsilon \cdot \ln t}{(\alpha^- - \alpha^+)\varphi'(\alpha^+)} + d^0 + o(1), \quad t \to \infty$$
$$D_A(t) = \frac{C \ln t}{2(\alpha^- - \alpha^+)\varphi'(\alpha^+)} + D^0 + o(1), \quad t \to \infty \qquad [21]$$

where $d^0$, $D^0$ are independent of $A$.

The second part gives the following convergence statements:

$$\sup_{x \in [ct - A\sqrt{t}, ct + A\sqrt{t}]} \Big| \int_{ct - A\sqrt{t}}^{x} \{f(y,t) - \tilde{f}(y - ct - d_A(t))\} \mathrm{d}y \Big| \to 0, \quad t \to \infty$$
$$\sup_{x \in [Ct - A\sqrt{t}, Ct + A\sqrt{t}]} \Big| \sum_{k=[Ct - A\sqrt{t}]}^{n} \{\Phi(F(k,t)) - \Phi(\tilde{F}(k - Ct - D_A(t)))\} \Big| \to 0, \quad t \to \infty$$

The precise *a priori* estimates of local solutions of [1], [2] play an important role in the proof. An example of such an estimate, also useful for further results, is given below.

**Proposition 1** *Let, in eqn [2],* $C = \varphi(0) > 0$, $\epsilon = 1$, $0 \le \varphi'(0) < \gamma_0$, $\bar{x} \stackrel{\text{def}}{=} (x - Ct)/\sqrt{Ct}$. *Let the function* $F(x, t)$, *defined in the domain* $\Omega_0 = \{(x, t)\colon a_1 < \bar{x} < a_2\}$, $a_2 > 0$, *satisfy eqn [2],*

$$\Delta F(x,t) \stackrel{\text{def}}{=} F(x,t) - F(x-1,t) \ge 0$$
$$|F(x,t)| \le \frac{\Gamma}{\sqrt{Ct}}, \quad (x,t) \in \Omega_0, \; t \ge t_0$$

*Then*

$$\Delta F(x,t) \le \frac{B \cdot \Gamma}{Ct}, \qquad (x,t) \in \Omega_0, \; t \ge t_0$$

*where*

$$B = B_0\left[a_2 + \left(\frac{1}{d} + \frac{\gamma_0 \Gamma}{C}\right)(1 + \ln(1 + a_2))\right]$$
$$d = \min(\bar{x} - a_1, a_2 - \bar{x})$$

*and* $B_0$ *is an absolute constant.*

It is interesting to compare *a priori* estimate of Proposition 1 with some similar (but less precise) estimates in the theory of classical quasilinear parabolic equations (Ladyzhenskaya *et al.* 1968).

We will formulate now the general conjecture concerning asymptotic behavior of solutions of initial problems [1], [3] and [2], [4] and some partial results which confirm this conjecture. To simplify formulation we admit the following.

**Assumption 2** Let $\hat{\psi}(u)$ and $\hat{\Psi}(u)$ be upper bounds of the convex hulls for the graphs of

$$\psi(u) = -\int_{\alpha^-}^{u} \varphi(y)\mathrm{d}y$$

and

$$\Psi(u) = \int_{\alpha^-}^{u} \frac{\mathrm{d}y}{\varphi(y)}$$

respectively, with $u \in [\alpha^-, \alpha^+]$. We suppose that

$$s = \{u \in [\alpha^-, \alpha^+] : \psi(u) < \hat{\psi}(u)\}$$
$$= (\alpha^-, \beta_0) \cup (\alpha_1, \beta_1) \cup \cdots (\alpha_L, \alpha^+)$$

where

$$\alpha^- = \alpha_0 < \beta_0 < \alpha_1 < \beta_1 < \cdots < \alpha_L < \beta_L = \alpha^+$$

or, correspondingly,

$$S = \{u \in [\alpha^-, \alpha^+] : \Psi(u) < \hat{\Psi}(u)\}$$
$$= (\alpha^-, b_0) \cup (a_1, b_1) \cup \cdots (a_M, \alpha^+)$$

where

$$\alpha^- = a_0 < b_0 < a_1 < b_1 < \cdots < a_M < b_M = \alpha^+$$

In addition, we suppose that $\varphi'(\alpha_l) \ne 0, \varphi'(\beta_l) \ne 0, l = 0, 1, \ldots, L$ or, correspondingly, $\varphi'(a_m) \ne 0, \varphi'(b_m) \ne 0$, $m = 0, 1, \ldots, M$.

**Proposition 2** (Weinberger 1990, Henkin and Polterovich 1999). *Under Assumptions 1, 2, one has:*

(i) *If* $u \in [\alpha^-, \alpha^+] \setminus s$ *and, correspondingly,* $u \in [\alpha^-, \alpha^+] \setminus S$, *then following functions are well defined:*

$$g_l\left(\frac{x}{t}\right) = \begin{cases} \beta_l, & \text{if } x < \varphi(\beta_l) \cdot t \\ \varphi^{(-1)}(x/t), & \text{if } \varphi(\beta_l) \cdot t \le x \le \varphi(\alpha_{l+1}) \cdot t \\ \alpha_{l+1}, & \text{if } x > \varphi(\alpha_{l+1}) \cdot t, \end{cases} \quad l = 0, 1, \ldots, L$$

*and, correspondingly,*

$$G_l\left(\frac{x}{t}\right) = \begin{cases} b_m, & \text{if } x < \varphi(b_m) \cdot t \\ \varphi^{(-1)}(x/t), & \text{if } \varphi(b_m) \cdot t \le x \le \varphi(a_{m+1}) \cdot t \\ a_m, & \text{if } x > \varphi(a_{m+1}) \cdot t, \end{cases} \quad m = 0, 1, \ldots, M$$

(ii) *For any interval* $(\alpha_l, \beta_l) \subset s$ *and, correspondingly,* $(a_m, b_m) \subset S$ *there exist traveling waves* $\tilde{f}_l(x - c_l t)$ *for [1] with overfall* $(\alpha_l, \beta_l)$ *and,*

*correspondingly*, $\tilde{F}_m(x - C_m t)$ for [2] *with overfall* $(a_m, b_m)$, *where*

$$c_l = \frac{1}{\beta_l - \alpha_l}\int_{\alpha_l}^{\beta_l} \varphi(y)\mathrm{d}y$$
$$c_l = \varphi(\beta_l), \quad l = 0, \ldots, L-1$$
$$c_l = \varphi(\alpha_l), \quad l = 1, \ldots, L$$

*and*, *correspondingly*,

$$C_m^{-1} = \frac{1}{b_m - a_m}\int_{a_m}^{b_m} \frac{\mathrm{d}y}{\varphi(y)}$$
$$C_m = \varphi(b_m), \quad m = 0, \ldots, M-1$$
$$C_m = \varphi(a_m), \quad m = 1, \ldots, M$$

**Conjecture** (Henkin and Polterovich 1994, 1999, Henkin and Shananin 2004). Let

$$\begin{aligned}\tilde{f}&(x, t, \gamma_0, \ldots, \gamma_L)\\ &= \sum_{l=0}^{L} \tilde{f}_l(x - c_l t - \varepsilon\gamma_l(t)) + \sum_{l=0}^{L-1} g_l\left(\frac{x}{t}\right) - \sum_{l=0}^{L-1}\beta_l\\ &\quad - \sum_{l=1}^{L}\alpha_l, \quad L \geq 1\end{aligned}$$

$$\begin{aligned}\tilde{F}&(n\varepsilon, t, \Gamma_0, \ldots, \Gamma_M)\\ &= \sum_{m=0}^{M} \tilde{F}_m(n\varepsilon - C_m t - \varepsilon\Gamma_m(t)) + \sum_{m=0}^{M-1} G_m\left(\frac{n\varepsilon}{t}\right)\\ &\quad - \sum_{m=0}^{M-1} b_m - \sum_{m=1}^{M} a_m, \quad M \geq 1\end{aligned}$$

Then under Assumptions 1, 2, the following statements are valid:

(i) For any solution $f(x, t)$, $x \in \mathbb{R}$, $t \in \mathbb{R}_+$, of initial problem [1], [3], there exist shift-functions $\gamma_l(t)$:

$$\gamma_l^- \ln t + O(1) \leq \gamma_l(t) \leq \gamma_l^+ \ln t + O(1)$$
$$0 \leq \gamma_l^- \leq \gamma_l^+ < \infty, \quad l = 0, 1, \ldots, L$$

such that

$$\sup_{x\in\mathbb{R}} |f(x, t) - \tilde{f}(x, t, \gamma_0, \gamma_1, \ldots, \gamma_L)| \to 0, \quad t \to \infty$$

(ii) Moreover, in (i) one can take

$$\gamma_l^- = \gamma_l^+ = \frac{\varepsilon}{(\beta_l - \alpha_l)} \times \begin{cases} -\dfrac{1}{\varphi'(\beta_l)}, & \text{if } l = 0 < L, \varphi(\alpha_0) \neq \varphi(\beta_0)\\ \dfrac{1}{\varphi'(\alpha_l)} - \dfrac{1}{\varphi'(\beta_l)}, & \text{if } 0 < l < L\\ \dfrac{1}{\varphi'(\alpha_l)}, & \text{if } l = L > 0, \varphi(\alpha_L) \neq \varphi(\beta_L)\end{cases}$$

(iii) For any solution $F(n\varepsilon, t)$, $n \in \mathbb{Z}$, $t \in \mathbb{R}_+$, of initial problem [2], [4], there exist shift-functions $\Gamma_m(t)$:

$$\Gamma_m^- \ln t + O(1) \leq \Gamma_m(t) \leq \Gamma_m^+ \ln t + O(1)$$
$$0 \leq \Gamma_m^- \leq \Gamma_m^+ < \infty, \quad l = 0, 1, \ldots, L$$

such that

$$\sup_{n\in\mathbb{Z}} |F(n\varepsilon, t) - \tilde{F}(n\varepsilon, t, \Gamma_0, \Gamma_1, \ldots, \Gamma_M)| \to 0, \quad t \to \infty$$

(iv) Moreover, in (iii) one can take

$$\Gamma_m^- = \Gamma_m^+ = \frac{C_m}{(b_m - a_m)} \times \begin{cases} -\dfrac{1}{\varphi'(b_m)}, & \text{if } m = 0 < M, \varphi(a_0) \neq \varphi(b_0)\\ \dfrac{1}{\varphi'(a_m)} - \dfrac{1}{\varphi'(b_m)}, & \text{if } 0 < m < M\\ \dfrac{1}{\varphi'(a_m)}, & \text{if } m = M > 0, \varphi(a_M) \neq \varphi(b_M)\end{cases}$$

The main result confirming formulated conjectures is the following.

**Theorem 5** (Henkin and Shananin). *Conjecture* (i) *for* $L = 1$ *and corresponding conjecture* (iii) *for* $M = 1$ *are true, that is,for solution of initial problem* [1], [3] *there exist shift functions* $\gamma_l(t) = O(\ln t)$ *such that for* $t \to \infty$ *we have*

$$f(x, t) \mapsto \begin{cases} \tilde{f}_0(x - c_0 t - \varepsilon\gamma_0(t)), & \text{if } x \leq c_0 t\\ \varphi^{(-1)}(x/t), & \text{if } c_0 t \leq x \leq c_1 t\\ \tilde{f}_1(x - c_1 t - \varepsilon\gamma_1(t)), & \text{if } x \geq c_1 t\end{cases}$$

*and for solution of initial problem* [2], [4] *there exist shift functions* $\Gamma_m(t) = O(\ln t)$ *such that for* $t \to \infty$ *we have*

$$F(n\varepsilon, t) \mapsto \begin{cases} \tilde{F}_0(n\varepsilon - C_0 t - \varepsilon\Gamma_0(t)), & \text{if } n\varepsilon \leq C_0 t\\ \varphi^{(-1)}(n\varepsilon/t), & \text{if } C_0 t \leq n\varepsilon \leq C_1 t\\ \tilde{F}_1(n\varepsilon - C_1 t - \varepsilon\Gamma_1(t)), & \text{if } n\varepsilon \geq C_1 t\end{cases}$$

The proof of Theorem 5 is of the same nature as the proof of Theorem 4.

**Remarks**

(i) The proof of stronger Conjectures (ii) and (iv) for $L = 1$ or $M = 1$ are in preparation.
(ii) The numerical results, Rykova and Spivak (preprint, 2004), confirm conjecture (iii) for $M = 2$.
(iii) The results of Weinberger (1990) and Henkin and Polterovich (1999) confirm convergence statements of Conjectures (i), (iii) for all $L$ and $M$, but only on the intervals of rarefaction

profiles: $x \in [\varphi(\beta_l)t, \varphi(\alpha_{l+1})t]$ or, correspondingly, $x \in [\varphi(b_m)t, \varphi(a_{m+1})t]$, $t > 0$.

The problem of finding asymptotics ($t \to \infty$) of solutions of (viscous) conservation laws has been posed originally not only for generalized Burgers equations but also for systems of conservation laws in one spatial variable (see Gelfand (1959)). In this direction many important results on existence and asymptotic stability of viscous shock profiles (continuous and discrete) have been obtained and applied (see Benzoni-Gavage (2004), Lax (1973), Serre (1999), Zumbrun and Howard (1998) and references therein). The results of type of Theorems 4,5 have not yet been obtained for systems of conservation laws.

It is also very interesting to study asymptotic behavior of scalar (viscous) conservation laws in several spatial variables (continuous or discrete), basing on the asymptotic properties of Burgers type equations. In this direction there have been several important results and problems (see Bauman and Phillips (1986), Henkin and Polterovich (1991), Hoff and Zumbrun (2000), Serre (1999), Weinberger (1990), and references therein).

## Further Reading


Bauman P and Phillips D (1986b) Large-time behavior of solutions to a scalar conservation law in several space dimensions. *Transactions of the American Mathematical Society* 298: 401–419.

Belenky V (1990) Diagram of growth of a monotonic function and a problem of their reconstruction by the diagram. Preprint, CEMI Academy of Science, Moscow, 1–44 (in Russian).

Benzoni-Gavage S (2002a) Stability of semi-discrete shock profiles by means of an Evans function in infinite dimension. *J.Dyn. Diff. Equations* 14: 613–674.

Burgers JM (1940) Application of a model system to illustrate some points of the statistical theory of free turbulence. *Proc. Acad. Sci. Amsterdam* 43: 2–12.

Dafermos CM (1977) Characteristics in hyperbolic conservation laws. A study of structure and the asymptotic behavior of solutions. In: Knops RJ (ed.) *Nonlinear Analysis and Mechanics: Heriot–Watt Symposium*, vol. 17, pp. 1–58. Research Notes in Mathematics, London: Pitman.

Gelfand IM (1959) Some problems in the theory of quasilinear equations. *Usp. Mat. Nauk* 14: 87–158 (in Russian). ((1963) *American Mathematical Society Translations 33*).

Harten A, Hyman JM, and Lax PD (1976) On finite-difference approximations and entropy conditions for shocks. *Communications in Pure and Applied Mathematics* 29: 297–322.

Henkin GM and Polterovich VM (1991) Schumpeterian dynamics as a nonlinear wave theory. *Journal of Mathematical Economics* 20: 551–590.

Henkin GM and Polterovich VM (1999) A difference-differential analogue of the Burgers equation and some models of economic development. *Discrete and Continuous Dynamical Systems* 5: 697–728.

Henkin GM and Shananin AA (2004) Asymptotic behavior of solutions of the Cauchy problem for Burgers type equations. *Journal Mathématiques Pure et Appliquée* 83: 1457–1500.

Henkin GM, Shananin AA, and Tumanov AE (2005) Estimates for solutions of Burgers type equations and some applications. *Journal Mathématiques Pure et Appliquée* 84: 717–752.

Hoff D and Zumbrun K (2000) Asymptotic behavior of multi-dimensional viscous shock fronts. *Indiana University Mathematical Journal* 49: 427–474.

Hopf E (1950) The partial differential equation $u_t + uu_x = \mu u_{xx}$. *Communications in Pure and Applied Mathematics* 3: 201–230.

Iljin AM and Oleinik OA (1960) Asymptotic behavior of the solutions of the Cauchy problem for some quasilinear equations for large values of time. *Mat. Sbornik* 51: 191–216 (in Russian).

Ladyzhenskaya OA, Solonnikov VA, and Ural'ceva NN (1968) *Linear and Quasilinear Equations of Parabolic Type*. Amer. Math.Soc.Transl. Monogr. vol. 23. Providence, RI.

Landau LD and Lifschitz EM (1968) *Fluid Mechanics*. Elmsford, NY: Pergamon.

Lax PD (1954) Weak solutions of nonlinear hyperbolic equation and their numerical computation. *Communications in Pure and Applied Mathematics* 7: 159–193.

Lax PD (1957) Hyperbolic systems of conservation laws, II. *Communications in Pure and Applied Mathematics* 10: 537–566.

Lax PD , (1973) Hyperbolic systems of conservation laws and the mathematical theory of shock waves. Conference Board of the Mathematical Science, Monograph 11. SIAM.

Levi D, Ragnisco O, and Brushi M (1983) Continuous and discrete matrix Burgers Hierarchies. *Nuovo Cimento* 74: 33–51.

Liu T-P (1978) Invariants and asymptotic behavior of solutions of a conservation law. *Proceedings of American Mathematical Society* 71: 227–231.

Liu T-P, Matsumura A, and Nishihara K (1998) Behaviors of solutions for the Burgers equation with boundary corresponding to rarefaction waves. *SIAM Journal of Mathematical Analysis* 29: 293–308.

Liu T-P and Yu S-H (1997) Propagation of stationary viscous Burgers shock under the effect of boundary. *Archieves for Rational and Mechanical Analysis* 139: 57–92.

Oleinik OA (1959) Uniqueness and stability of the generalized solution of the Cauchy problem for a quasi-linear equation. *Usp.Mat.Nauk* 14: 165–170. ((1963) American Mathematical Society Translations 33).

Serre D (1999) *Systems of Conservation Laws, I*. Cambridge: Cambridge University Press.

Serre D (2004) $L^1$-stability of nonlinear waves in scalar conservation laws. In: Dafermos C and Feireisl E (eds.) *Handbook of Differential Equations*, pp. 473–553. Elsevier.

Weinberger HF (1990) Long-time behavior for a regularized scalar conservation law in the absence of genuine nonlinearity. *Annales de L'institut Henri Poincare (C) Analyse Nonlineaire*.

Zumbrun K and Howard D (1998) Poinwise semigroup methods and stability of viscous shock waves. *Indiana University Mathematical Journal* 47: 63–185.

# Cellular Automata

**M Bruschi**, Università di Roma "La Sapienza", Rome, Italy
**F Musso**, Università "Roma Tre", Rome, Italy



## What is a Cellular Automaton?

Cellular automata (CAs) were first introduced by J von Neumann in his investigation of "complexity," following an inspired suggestion by S Ulam. But in the last 50 years they have been investigated and used in a number of fields; widely different terminologies have been used by researchers that now it is difficult even to give a precise general definition of a CA. Thus, some definitions and approximations are in order.

First a broad definition:

1. have a number of cells (boxes);
2. at any (discrete) time step, any cell can present itself in a certain "state" among a finite number of different states;
3. the state of any cell can change (evolve) from a time step to the subsequent time step; and
4. there is a rule (evolution law, EL) which determines this transition.

Note that the number of cells can be finite or infinite; the cells can be arranged on a line, on a surface, in the ordinary three-dimensional (3D) space, or possibly in a hyperspace (in any case, the cells can be numbered); the different states of a cell can be denoted by integer numbers but, in different contexts of application of CAs, different imaginative pictures have also been used (e.g., different colors, dead and living cells, number of balls in a box, etc.); the evolution of a CA proceeds in finite time steps (time is also discrete); the EL, provided that it is effective on any possible configuration of a given CA (computability), is otherwise completely arbitrary (indeed, there are not only deterministic and probabilistic ELs, but also those that "evolve" in time – following a meta-EL, which in turn can be deterministic or probabilistic).

Consider some examples of CAs.

**Example 1** (CA1) Consider a linear array of seven boxes (cells; one can number them $c(i)$, $i = 1, 2, \ldots, 7$). Each box can be empty or it can contain a ball (so there are just two states for each cell). Given a configuration of this CA at time $t$, what happens at time $t + 1$ (EL)?

(i) the state of the first box c(1) never changes;
(ii) for each other box $c(i)$, $i = 2, 3, \ldots, 7$;
(iia) if the box is empty and the box on its left is empty then put a ball in the box;
(iib) if there is a ball in the box and also there is a ball in the box on its left then empty the box.

An example of the evolution of such a rather trivial CA is given in Figure 1.

A more precise notation can now be established.

First, let us denote the state of a cell at time $t$ by a "state function," say $S$. According to the point (iib) above, the number of possible states is arbitrary but finite: denote this number by the positive integer $M$ ($M > 1$). Then $S$ takes values on a finite field, say $\mathbb{Z}_M = \mathbb{Z}/M\mathbb{Z} = \{0, 1, 2, \ldots, M - 1\}$ (in plain words, we have denoted the $M$ states for the CA by the first $M$ non-negative integers). Different cells can be labeled with a progressive number: $c(n), n = n_1, n_1 + 1, \ldots, n_2 - 1, n_2$; possibly, in case of an infinite number of cells, one has $n_1 \longrightarrow -\infty$ and/or $n_2 \longrightarrow +\infty$. In the case of $n_1 = -\infty, n_2 = \infty$, one speaks of a unidimensional CA. Of course, the field $S$ depends on $n$ as well as on time (remember that, for a CA, "time" is a discrete variable: $t = 0, 1, 2, \ldots$). The field $S(n, t)$ describes completely the CA. If the EL is deterministic, then one can determine (compute) $S(n, t)$ step by step for $t > 0$ from the initial configuration $S(n, 0)$ (initial datum, ID). Consider only static ELs, namely those that do not change in time. A further distinction can be made: there are ELs such that the future state of the generic cell, $S(n, t + 1)$, depends on the whole current configuration of the CA (these are called nonlocal ELs) and there are ELs for which $S(n, t + 1)$ depends only on

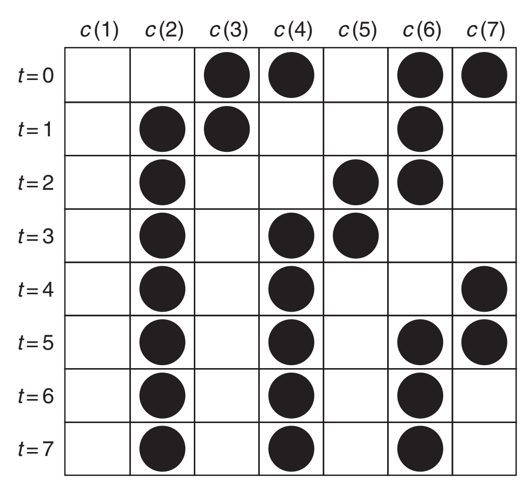



**Figure 1** A seven time-step evolution of CA1 starting from a given ID ($t = 0$). Note that a stable configuration has been reached at $t = 6$.

the current state of a finite number, say $N$, of cells (local ELs):

$$\{S(n+k_i,t)\},\ i=1,2,\ldots,N,k_i\in\mathbb{Z} \Longrightarrow S(n,t+1) \quad [1]$$

Note that, in principle, the set of cells that determine, according to the EL, the future state of the generic cell $n$, could depend on $n$, namely one can have $N=N(n)$, as well $k_i=k_i(n), i=1,2,\ldots,N(n)$ (see CA2 below). In any case, such a set of cells is called the interaction set (IS). Moreover, the distance from the cell $n$ of the farthest cell in the IS is called the range $R$ (of the interaction): $R=\max(|k_i|)$. If $\text{IS}\equiv\{c(n-R),c(n-R+1),\ldots c(n),\ldots c(n+R-1), c(n+R)\}$, then this IS is called a neighborhood of range $R$. It is, moreover, clear that, for unidimensional CA, there exists at least one infinite subset of cells that have the same state. If there is only one such subset, then it is called the vacuum set and the state of its cells is called vacuum state: let $V$ denote the value of this state ($0\leq V<M,\quad S(n,t)\underset{n\to\pm\infty}{\longrightarrow}V$).

**Example 2** (CA2) An example of CA with *n-dependent* IS ($M=2, R=3, V=0$). This is the EL: the cell $c(n)$ changes its state ($0\to1, 1\to0$) iff

(i) $n$ is even and at least one of the two cells on its left is not in the vacuum state;
(ii) $n$ is odd and one or three of the three cells on its right are not in the vacuum state.

An example of the evolution of such a CA is given in Figure 2.

Usually, only a subclass of ELs is considered for which the phenomenon of vacuum excitation cannot occur. Namely, during the evolution of the CA, an infinite subset of the vacuum set cannot change its state in just one time step. In other words: if the set of cells starting from the first cell and ending with the last one for which $S(n,t)\neq V$ be called population set (PS), then PS is a finite set at each time.

Of course, one can easily devise an EL for which this is not true; nevertheless, the EL itself is still valid (computable), for instance,

**Example 3** (CA3) This is an unidimensional CA, namely there are infinite cells on a line ($n\in\mathbb{Z}$). The cells have $M$ states and $V=0$; the EL reads:

the state of each cell cycles in the set of available states ($0\to1, 1\to2,\ldots,M-2\to M-1, M-1\to0$)

Note that the range $R$ is zero, there is a vacuum excitation; nevertheless, the EL is effective.

Deterministic, static, and local ELs that do not give rise to vacuum excitation are called normal ELs (NELs).

Since $M$, $N$ are finite for an NEL, one can give the NEL itself as a table, considering every possible configuration of the IS and specifying the outcome for each configuration (note that there are $M^N$ possible configurations).

**Example 4** (CA4) $n\in\mathbb{Z}, M=2, V=0, \text{IS}\equiv\{c(n), c(n-1), c(n+2)\}, N=3, R=2$. The EL is:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $S(n,t)$ | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| $S(n-1,t)$ | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| $S(n+2,t)$ | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| $S(n,t+1)$ | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 |

[2]

An example of the evolution of such a CA is given in Figure 3.

However, these NELs can also be given in an alternative representation (more useful in view of the extensions of the concept of CA itself, see below). Namely, an NEL can be given as a discrete-time EL for the state function $S(n,t)$ in the finite field $\mathbb{Z}_M=\{0,1,2,\ldots,M-1\}$.

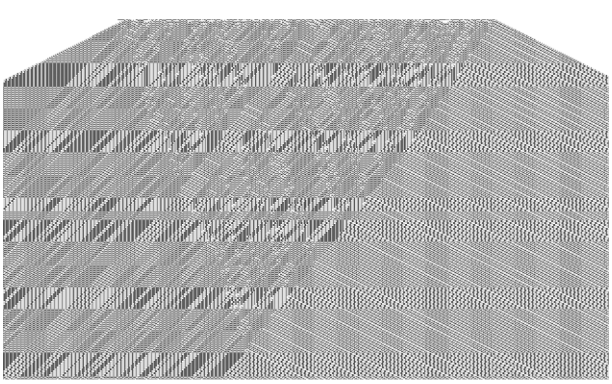


**Figure 2** Three hundred and eighty time steps of CA2, starting from a random chosen initial configuration. Note the left–right asymmetry due to the asymmetry of its IS and EL.

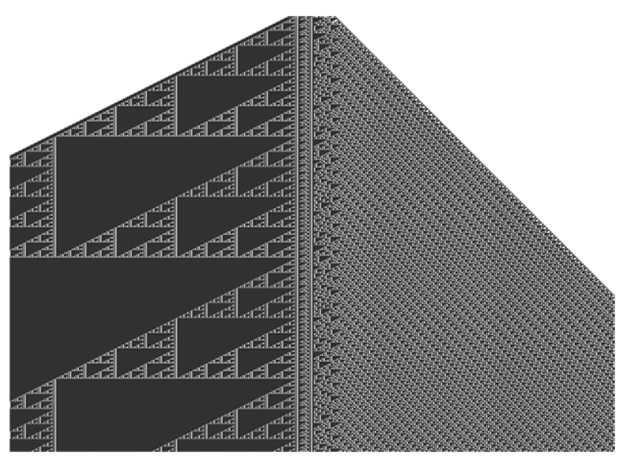


**Figure 3** Four hundred and sixty-one time steps of CA4, starting from a random chosen PS of 50 cells.

For example, the NEL above for CA4 can be expressed as follows:

$$S(n,t+1) \overset{2}{=} S(n-1,t) + S(n,t) + S(n+2,t) + S(n,t)S(n+2,t) + S(n-1,t)S(n,t)S(n+2,t) \quad [3]$$

Here and in the following, the symbol $\overset{M}{=}$ denotes a congruence mod $M$.

Another example is the following.

**Example 5** (CA5) $n \in \mathbb{Z}, M=3, N=3, V=0, R=1$, IS $\equiv \{c(n-1), c(n), c(n+1)\}$. The NEL is:

$$S(n,t+1) \overset{3}{=} S(n-1,t) + S(n,t) + S(n+1,t) + 2S(n-1,t)S(n+1,t) \quad [4]$$

An example of the evolution of such a CA is given in Figure 4.

### Classification of ELs

Considering a CA with given $M > 1, N \geq 1$, the number $L$ of possible deterministic, static ELs is

$$L(M,N) = M^{(M^N)} \quad [5]$$

Of course, this number can be very large for relatively small values of $M$ and $N$ also. Nevertheless, it is a finite positive integer, so that, for given $M$, $N$, one could denote every EL by an integer number and investigate the typical behavior of each EL. A considerable reduction of this number is obtained if one limits attention to totalistic ELs, namely to those whose outcome depends only on the global configuration of the IS, often just on

$$\sigma(n,t) = \sum_{i=1}^{N} S(n+k_i); \quad i=1,2,\ldots,N, \, k_i \in \mathbb{Z} \quad [6]$$

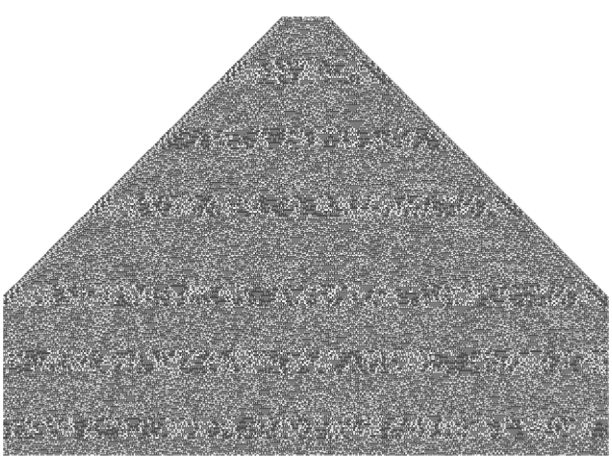


**Figure 4** Four hundred and sixty-one time steps of CA5, starting from a random chosen PS of 50 cells.

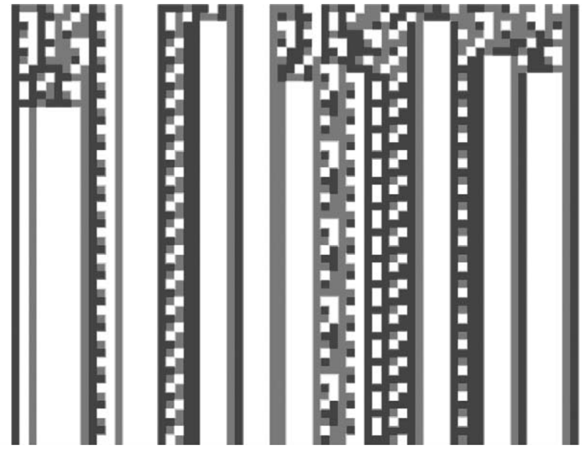


**Figure 5** A class-1 CA: every ID rapidly evolves to periodic structures; $M=3, V=0, R=2$, EL: $S(n,t+1) \overset{3}{=} S(n,t) + S(n-1,t)S(n+2,t)$.

Deep and extensive computer investigations have been exploited for unidimensional CAs with small values of $M$, $N$. Surprisingly enough, it seems that the typical behavior of all these CAs can be (roughly and heuristically) classified in just four classes (Wolfram 2002):

- Class 1 (simple): possibly after a complicated transient, simple patterns emerge.
- Class 2 (fractalic): possibly after a transient, overall regular nested structures are obtained.
- Class 3 (chaotic): complicated but seemingly random behavior.
- Class 4 (complex): possibly after a transient, localized structures emerge that interact in complex ways.

Due to the looseness of the above definitions, perhaps a better way to distinguish between classes is to train one's eye. Consider some examples of CAs for each class: the typical behavior of class-1 CA is shown in Figures 5 and 6, of class-2 CA in Figures 7 and 8, of class-3 CA in Figures 4 and 9, of class-4 CA in Figures 10 and 11. Note, however, that often one has "mixed type" CA: for example, CA4 is of class 1 on the right and of class 2 on the left (see Figure 3); Figure 12 exhibits a CA where the typical behaviors of classes 2 and 3 are superimposed.

## Extensions

The concept of a CA is so simple that many extensions of the above-sketched definition of a CA can be easily devised. A (nonexhaustive) survey of such extensions follows.

**Figure 6** A class-1 CA, a random ID vanishes after 337 time steps, $M=5, V=0, R=2$, EL: $S(n,t+1) \overset{5}{=} S(n-1,t)S(n-2,t)+S(n+1,t)S(n+2,t)+S(n-1,t)S(n+1,t)+S(n-2,t)S(n+2,t)$.

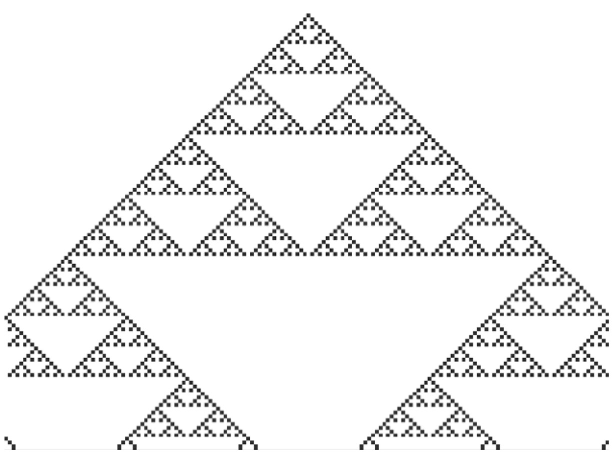


**Figure 7** A class-2 CA: Sierpinsky triangles appear; $M=2$, $V=0, R=1$, EL: $S(n,t+1) \overset{2}{=} S(n-1,t)+S(n+1,t)$.

### Vector CA

In this extension, the state function $S(n,t)$ is considered as a "vector," namely $S(n,t) \equiv (S_1(n,t), S_2(n,t), \dots S_L(n,t)), L$ being a positive integer. Each component $S_l(n,t) (l=1,2,\dots,L)$ takes values in a finite field, say $\mathbb{Z}_{M_l} = \{0, 1, 2, \dots, M_l - 1\}$, and evolves, according to some EL, interacting with the other components. Of course, one can give separately the time evolution for each component; however, it is also possible to give a global representation of a vector CA, introducing a global function $\tilde{S}(n,t)$ that takes values in the finite field $\mathbb{Z}_M, M = M_1 M_2 \dots M_L$; for example,

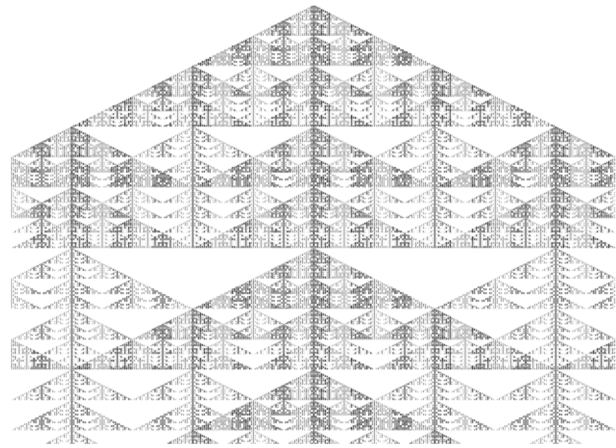


**Figure 8** A class-2 CA: a double fractal structure appears; $M=4$, $V=0, R=2$, EL: $S(n,t+1) \overset{4}{=} S(n-2,t)+S(n,t)+S(n+2,t)$.

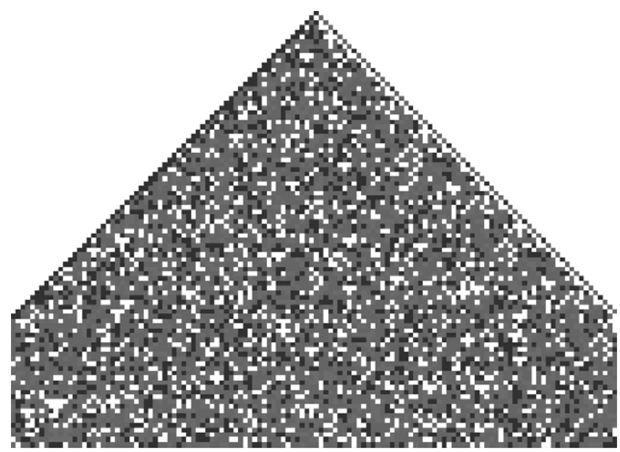


**Figure 9** A class-3 CA: $M=5, V=0, R=2$, EL: $S(n,t+1) \overset{5}{=} 2S(n-1,t)+S(n+1,t)+S(n,t)(S(n+1,t)+S(n+2,t))+S(n-1,t)S(n+1,t)$.

$$\tilde{S}(n,t) = S_L(n,t) + \sum_{l=1}^{L-1} \left( S_l(n,t) \prod_{k>l}^{L} M_k \right) \qquad [7]$$

Thus, in a sense, vector CAs are still usual CAs with a complicated EL.

**Example 6** (CA6) A two-component vector CA:

$$S_1(n,t+1) \overset{M_1}{=} S_1(n,t)S_1(n+1,t) + (M_1-1)S_2(n-1,t)S_2(n,t) + c_1 \qquad [8]$$

$$S_2(n,t+1) \overset{M_2}{=} S_1(n-1,t)S_2(n,t) + S_1(n,t)S_2(n+1,t) + c_2 \qquad [9]$$

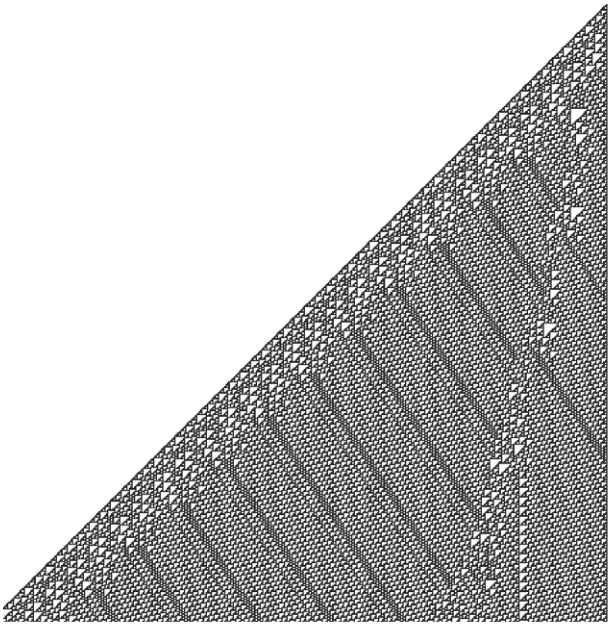


**Figure 10** A class-4 CA (Wolfram CA 110): $M=2$, $V=0$, $R=1$, EL: $S(n,t+1)\overset{2}{=}S(n,t)+S(n+1,t)+S(n,t)S(n+1,t)+S(n-1,t)$ $S(n,t)S(n+1,t)$.

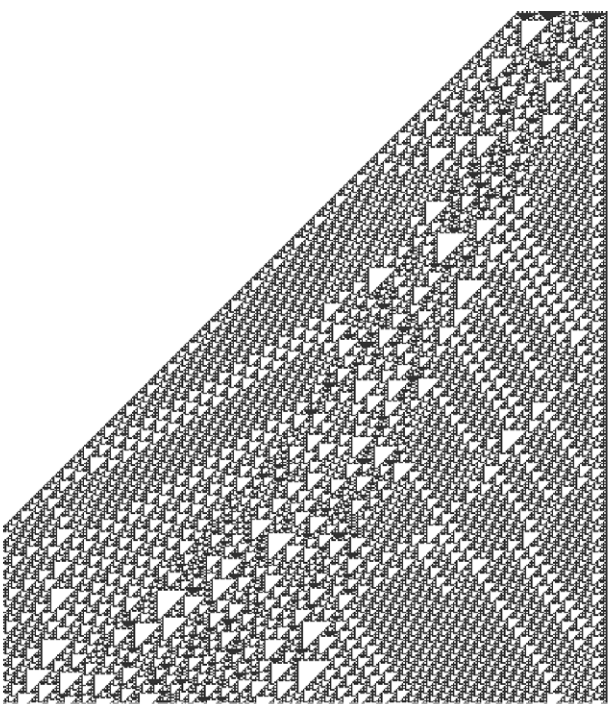


**Figure 11** A class-4 CA. Note the interacting moving structures on the left and on the right; note also the apparently chaotic behavior in the center; $M=2, V=0, R=2$, EL: $S(n,t+1)\overset{2}{=}$ $S(n,t)+S(n+1,t)+S(n-1,t)S(n+2,t)$.

The global behavior of this CA can be expressed, for example, through the global state function

$$\tilde{S}(n,t) = M_2 S_1(n,t) + S_2(n,t) \qquad [10]$$

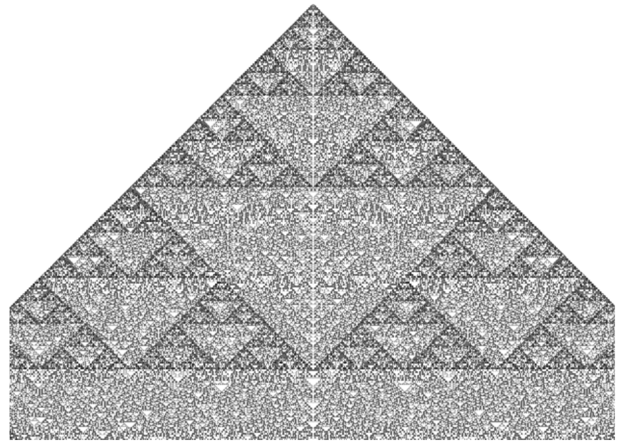


**Figure 12** A mixed-class CA: a fractalic structure is superimposed on a chaotic one; $M=4$, $V=0$, $R=2$, EL: $S(n,t+1)\overset{2}{=}$ $S(n,t)(S(n-2,t)+S(n+2,t))+S(n-1,t)S(n+1,t)$.

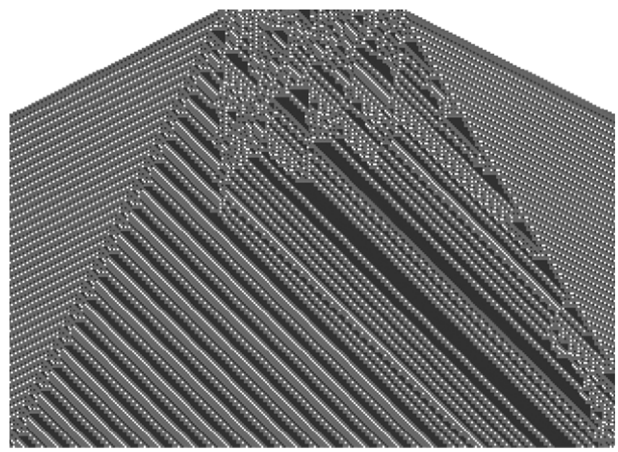


**Figure 13** Global behavior of the vector CA6.

Obviously, $\tilde{S}\in\mathbb{Z}_M$ with $M=M_1M_2$. Figure 13 represents the global behavior of this CA with $M_1=2, M_2=3, c_1=1, c_2=1, V=0$.

Note that this CA can be considered as an extension of the celebrated quadratic map.

### Multidimensional CA

Up to now we have considered CAs with finite number of cells (finite CAs) or with an infinite number of cells arranged on a line (unidimensional CAs). Now we consider CAs with cells arranged on a surface, usually a plane (bidimensional CAs), or on 3-space (tridimensional CAs), or even on a hyperspace (multidimensional CAs). In any case, if the number of cells is finite, the evolution of such CAs, according to an NEL, must end up to a final cycle: this is due to the finiteness of the "phase space" (thus, these CAs should be classified as class 1; however, note that, if the "phase space" is large enough, the dynamics of

such CAs can still be very rich). Usually, one considers an infinite number of cells tessellating the whole $s$-space, $s = 2, 3, \ldots$ (e.g., squares or hexagons on the plane, cubic cells in 3-space). The changes in the previous notation and definitions are plain: for example, for a bidimensional CA, the state function depends now on two discrete "space" variables $(S(n_1, n_2, t), n_1 \in \mathbb{Z}, n_2 \in \mathbb{Z})$; furthermore, there is a greater freedom in choosing a neighborhood of range $R$. Two most-used neighborhoods of range 1 are shown below:

$$
\begin{array}{c}
\text{Neumann neighborhood} \\
\begin{array}{ccccc}
\bigcirc & \bigcirc & \bigcirc & \bigcirc & \bigcirc \\
\bigcirc & \bigcirc & \otimes & \bigcirc & \bigcirc \\
\bigcirc & \otimes & \odot & \otimes & \bigcirc \\
\bigcirc & \bigcirc & \otimes & \bigcirc & \bigcirc \\
\bigcirc & \bigcirc & \bigcirc & \bigcirc & \bigcirc
\end{array} \\
\text{Moore–Conway neighborhood} \\
\begin{array}{ccccc}
\bigcirc & \bigcirc & \bigcirc & \bigcirc & \bigcirc \\
\bigcirc & \otimes & \otimes & \otimes & \bigcirc \\
\bigcirc & \otimes & \odot & \otimes & \bigcirc \\
\bigcirc & \otimes & \otimes & \otimes & \bigcirc \\
\bigcirc & \bigcirc & \bigcirc & \bigcirc & \bigcirc
\end{array}
\end{array}
\qquad [11]
$$

The most famous (and interesting) bidimensional CA is "Life", introduced by J H Conway, which is discussed next.

**Example 7** (CA "Life"; Moore–Conway neighborhood, $V = 0, M = 2$). A cell in the vacuum state 0 is called "dead"; a cell in the state 1 is called "alive." The EL is as follows:

(i) If a cell is dead at time $t$, it comes alive at time $t + 1$ if and only if exactly three of its eight neighbors are alive at time $t$ (reproduction).
(ii) If a cell is alive at time $t$, it dies at time $t + 1$ if and only if fewer than two (loneliness) or more than three (overcrowding) neighbors are alive at time $t$.

Clearly, this is a totalistic NEL. Now considering the explicit form of $\sigma$ (see [6]):

$$
\sigma(n_1, n_2, t) = -S(n_1, n_2, t) + \sum_{k_1=-1}^{1} \sum_{k_2=-1}^{1} S(n_1 + k_1, n_2 + k_2, t) \qquad [12]
$$

the above EL can be simply expressed as:

$$
S(n_1, n_2, t + 1) = \delta_{3,\sigma} + \delta_{2,\sigma} S(n_1, n_2, t) \qquad [13]
$$

where $\delta_{3,\sigma}$ is the Kroenecker symbol.

Life is a class-4 CA; it exhibits a rich variety of interesting structures: stable structures, oscillators (periodic structures), gliders and ships (moving structures), emitters and absorbers (namely, structures that, after a time period, reconstitute themselves, but meanwhile they have emitted or adsorbed moving structures). These structures are essential to prove that Life can be used to construct a universal Turing machine (see below). One can get a rough idea of such "richness" from Figure 14.

As in the previous case of vector CA, one could object that also multidimensional CAs are not true extensions of the unidimensional CAs. Indeed, since the whole set of cells is still a countable set, one could number the cells with just a discrete "space" variable (say $n \in \mathbb{Z}$ ). For example, in the case of a square tessellation of the plane, we could enumerate the cells in the plane starting from the origin as follows:

$$
\begin{array}{ccccccc}
 & & & 22 & \longrightarrow & & \\
 & & & & & & \\
 & & & 21 & 20 & 19 & 18 \\
-13 & -12 & -11 & 4 & 5 & 6 & 17 \\
-14 & -9 & -10 & 3 & 2 & 7 & 16 \\
-15 & -8 & -1 & 0 & 1 & 8 & 15 \\
-16 & -7 & -2 & -3 & 10 & 9 & 14 \\
-17 & -6 & -5 & -4 & 11 & 12 & 13 \\
-18 & -19 & -20 & -21 & & & \\
 & & & & & & \\
 & & \longleftarrow & -22 & & &
\end{array}
\qquad [14]
$$

Thus, any multidimensional CA could in principle be viewed as a unidimensional one. Of course, one has to pay a price for this: ISs and ELs that are simple for a multidimensional CA become cumbersome for its unidimensional version and vice versa.

### Higher Time Derivatives

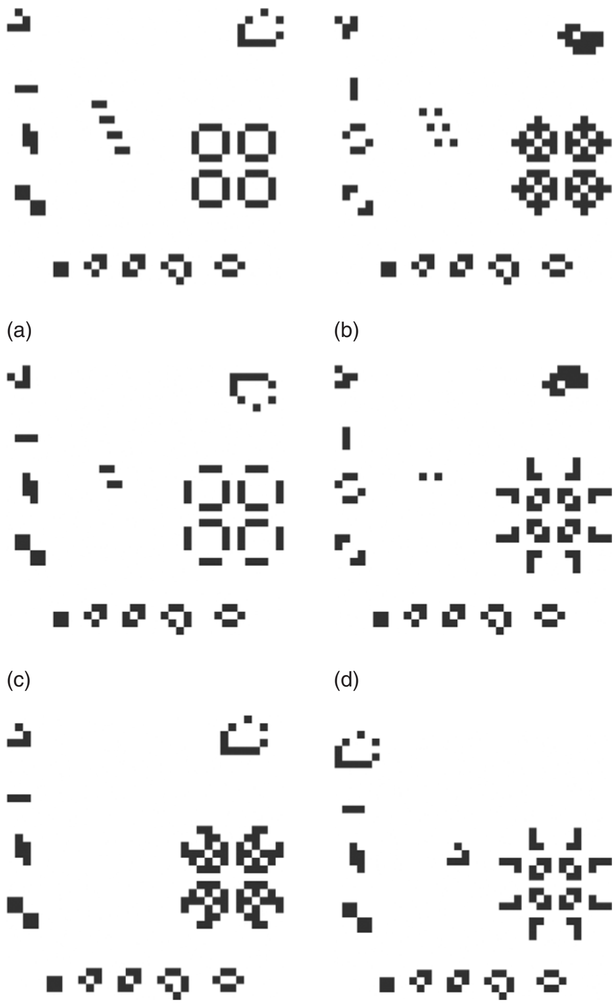
Up to now, we have considered CAs whose evolved state $S(t + 1)$ depends only on the state $S(t)$, namely the state of the CA itself at the previous time step. In other words the EL involves just the first (discrete) time derivative (1_CA). One can easily extend all the previous definitions to consider higher-order discrete time derivatives ($K$_CA). Of course, the ID and the IS for such a CA involve the state of the CA at $K$ subsequent time steps.

An example of a unidimensional 2_CA is given below.

**Example 8** (CA7) $M = 3, V = 0, R = 1$. The EL is:

$$
S(n, t + 1) \overset{3}{=} S(n - 1, t) + S(n, t - 1) + S(n + 1, t) \qquad [15]
$$

An example of the evolution of such a CA is given in Figure 15.



(a) (b) (c) (d) (e) (f)

**Figure 14** CA "Life": (a) Time 0. Near the lower border, five stable structures (from the left to the right: a "block", a "boat", a "ship", a "loaf", a "beehive"); near the left border three "blinkers" (period-2 oscillators); near the right corner, a symmetric structure that, in one time step, evolves into a "pulsar" (a period-3 oscillator), on the left-up corner a "glider" (a moving structure); on the right-up corner a "medium weight spaceship" (another moving structure); in the center, a configuration that vanishes in a few time steps. (b) Time 1. The structures on the lower border are unchanged, the blinkers, the glider, and the space ship are in an intermediate state, on the right border, the pulsar starts to pulse. (c) Time 2. The three blinkers on the left border are again in their original configurations (periodic structure with period 2), the pulsar, the glider and the spaceship are in another intermediate state. (d) Time 3. The pulsar is in its second state, the glider and the spaceship in their third, the structure in the center is going to vanish. (e) Time 4. The pulsar has completed its pulsation (period-3 oscillator, see **Figure 14b**); the structure in the center has vanished, the glider and the spaceship have recovered their original configurations (see **Figure 14a**) but meanwhile they have moved of a cell in four time steps (1\4 of the highest velocity attainable by a moving structure in a CA of range 1). The glider is moving downward and to the right, the space ship in horizontal to the left. (f) Time 60. The space ship has almost completed its crossing, the glider has reached the center and it is in a collision route with the pulsar.

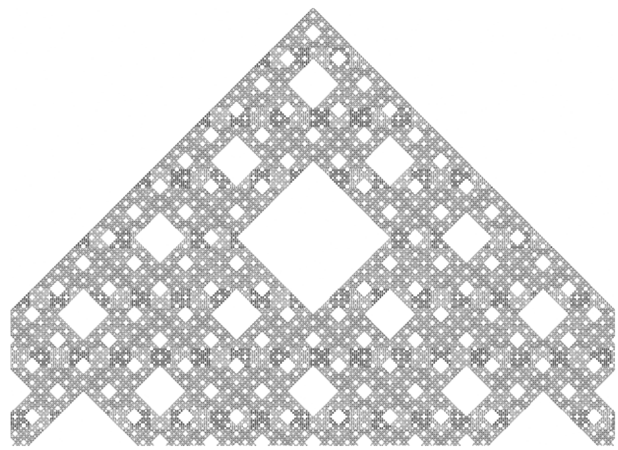


**Figure 15** CA7, clearly a class-2 CA.

It is plain that taking a suitable continuum limit of a $K$_CA one gets a partial differential equation of order $K$ for the evolution. However, there are also special and interesting CAs, called "filter" CAs, that in a suitable continuum limit end up in integral evolution equations. For a filter unidimensional CA, the evolved state at the cell $n$, $S(n,t+1)$, depends also on the (already) evolved states of the cells on its left (or right): for example, an NEL of the type

$$\begin{aligned} &S(n,t+1) \stackrel{M}{=} F(S(n+k_i,t), S(n-\tilde{k}_j,t+1)) \\ &\quad i=1,2,\ldots,N; \quad k_i \in \mathbb{Z} \\ &\quad j=1,2,\ldots,\tilde{N}; \quad \tilde{k}_j \in \mathbb{N} \qquad [16] \end{aligned}$$

is still valid (computable). Extensions to $K$_CAs or vector CAs or multidimensional CA are plain. Very often filter CAs exhibit a class-4 behavior with particle-like structures moving and interacting in a complex way; see the following example and examples in the next section.

**Example 9** (CA8) $M=2, V=0, R=2$. The EL is:

$$\begin{aligned} S(n,t+1) \stackrel{2}{=} &S(n-1,t-1)S(n-2,t) \\ &+ S(n,t) + S(n+1,t)S(n+2,t) \qquad [17] \end{aligned}$$

An example of the evolution of such a CA is given in Figure 16.

### Invertible CA

For most of the ELs there is a loss of information in the course of the evolution (see, e.g., Figures 5 and 6). Indeed, different definitions of "CA entropy" have been introduced to measure the "randomness" in the behavior of a given CA. However, since CAs are important in physical

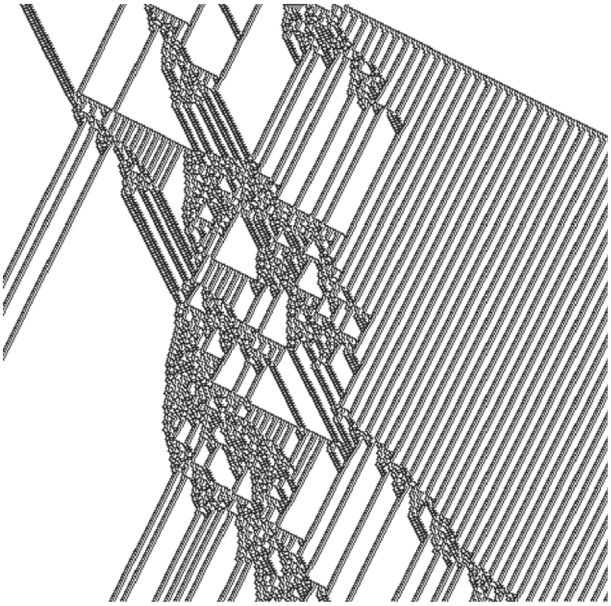


**Figure 16** CA8, a "filter" CA. Note the emerging of particle-like structures moving to the left and to the right and interacting in complex ways.

modeling as well as in cryptography and data compression, there is great interest in a special subclass of CAs which are "invertible" (time reversible). Namely, for an "invertible" CA following a given EL and starting from an arbitrary ID, there exists an "inverse" EL such that one can recover the ID from the evolved states. Invertible CAs can be easily devised in the case of $K_CA\,(K > 1)$. For example, if $K = 2, 3 \ldots$, one can consider ELs of the form

$$S(n,t+1) \stackrel{M}{=} S(n,t-K+1) + F\Big(S(n+k_i^j, t-j)\Big) \quad [18a]$$

where

$$\begin{aligned} &i = 1, 2, \ldots, N^j; \quad k_i^j \in \mathbb{Z} \\ &j = 0, 1, 2, \ldots, K-2 \end{aligned} \quad [18b]$$

and $F$ is an arbitrary polynomial function.

It is then clear that the inverse EL reads

$$\begin{aligned} \tilde{S}(n,\tilde{t}+1) \stackrel{M}{=} &\tilde{S}(n,\tilde{t}-K+1) \\ &+ (M-1)F\Big(\tilde{S}(n+k_i^j, \tilde{t}+j-K+2\Big) \end{aligned} \quad [19]$$

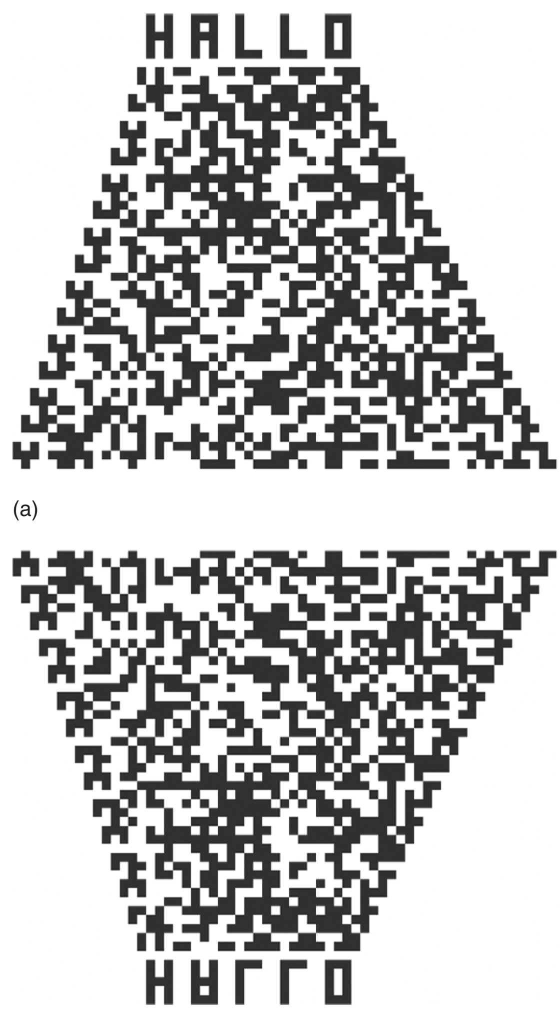
Indeed, if an arbitrary ID evolves according to the EL [18], then applying the inverse EL [19] to $K$ subsequent evolved states (taken in reversed order), eventually the original ID is recovered (in reversed order) (see the following example).

**Example 10** (CA9) A 6_CA: $M = 2, V = 0, R = 1$. The EL is:

$$\begin{aligned} S(n,t+1) \stackrel{2}{=} &S(n,t-5) + S(n,t-3) + S(n+1,t-2) \\ &+ S(n-1,t-1) \\ &+ S(n,t-2)S(n+1,t-2) \\ &+ S(n,t)S(n-1,t) \end{aligned} \quad [20]$$

The inverse EL, according to [19], reads (Figure 17)

$$\begin{aligned} \tilde{S}(n,\tilde{t}+1) \stackrel{2}{=} &\tilde{S}(n,\tilde{t}-5) + \tilde{S}(n,\tilde{t}-1) + \tilde{S}(n+1,\tilde{t}-2) \\ &+ \tilde{S}(n-1,\tilde{t}-3) \\ &+ \tilde{S}(n,\tilde{t}-2)\tilde{S}(n+1,\tilde{t}-2) \\ &+ \tilde{S}(n,\tilde{t}-4)\tilde{S}(n-1,\tilde{t}-4) \end{aligned} \quad [21]$$




(a)

(b)

**Figure 17** CA9, a 6_CA: (a) a 50 time-step evolution from a peculiar ID; (b) a 50 time-step evolution of the inverse EL, starting from the last six configurations of Figure 17a (taken in inverse order); the ID of Figure 17a is recovered (in inverse order).

Of course, more complicated invertible ELs can be devised. Invertible ELs can be also easily devised for "filter" CA, for example, if an NEL for a "filter" CA reads

$$S(n,t+1) \overset{M}{=} S(n,t) + F(S(n+k_i,t), S(n-\tilde{k}_j,t+1)) \quad [22]$$

where $k_i$ and $\tilde{k}_j$ are positive integers $(i=1,2,\ldots,N; j=1,2,\ldots,\tilde{N})$ and $F$ is an arbitrary (polynomial) function, then it is invertible and the inverse NEL reads

$$\tilde{S}(n,\tilde{t}+1) \overset{M}{=} \tilde{S}(n,\tilde{t}) + (M-1) \times F(\tilde{S}(n+k_i,\tilde{t}+1), \tilde{S}(n-\tilde{k}_j,\tilde{t})) \quad [23]$$

Note that [22] is computable starting from $n=-\infty$, whereas [23] is computable starting from $n=+\infty$.

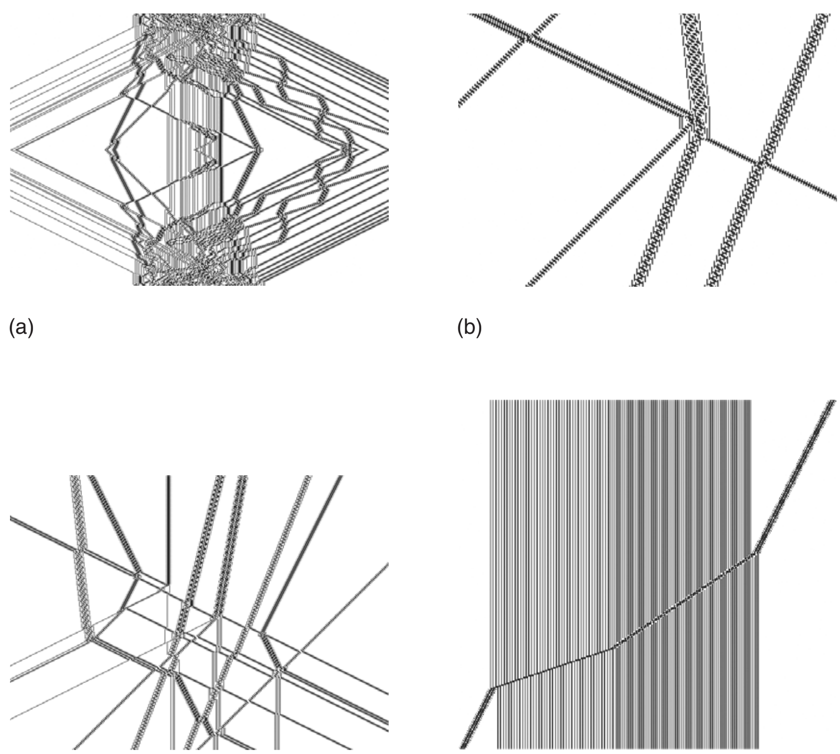
**Example 11** (CA10) A 1.5_CA, $M=2$, $V=0$, $R=3$. The EL is:

$$\begin{aligned} S(n,t+1) \overset{2}{=} & S(n,t) + S(n-3,t+1)S(n-2,t+1) \\ & + S(n+2,t)S(n+3,t) \\ & + S(n-2,t+1)S(n-1,t+1) \\ & + S(n+1,t)S(n+2,t) \end{aligned} \quad [24]$$

Note that this EL is of the form [22]; therefore, it is invertible (see **Figure 18a**). According to [23], the inverse EL reads:

$$\begin{aligned} \tilde{S}(n,\tilde{t}+1) \overset{2}{=} & S(n,\tilde{t}) + \tilde{S}(n+3,\tilde{t}+1)\tilde{S}(n+2,\tilde{t}+1) \\ & + \tilde{S}(n-2,t)\tilde{S}(n-3,t) \\ & + \tilde{S}(n+2,\tilde{t}+1)\tilde{S}(n+1,\tilde{t}+1) \\ & + \tilde{S}(n-1,\tilde{t})\tilde{S}(n-2,\tilde{t}) \end{aligned} \quad [25]$$

This CA exhibits a very rich dynamics: any complex ID rapidly decays in a great variety of coherent particle-like structures, steady or moving to the right or



(a) (b) (c) (d)

**Figure 18** CA10: (a) 230 time-step evolution, then the inverse EL is applied for 230 further time step in order to recover the initial configuration. (b) Collisions between different kinds of particle-like coherent moving structures. The last collision (on the right) is a solitonic one: the interaction produces just a phase shift, preserving number, shape, and velocities of the involved "particles." (c) "Particles" moving with different velocities and interacting in complex ways (solitonic collisions, particle creations and annihilations). (d) A particle goes through a nonhomogeneous medium and undergoes refraction by the medium itself.

to the left with different velocities. The interactions between different particles may be solitonic (the particles emerge unchanged but shifted) or annihilation–creation phenomena can occur (see Figures 18a–d).

## Applications of CAs

### CAs as Universal Constructors and Turing Machines

In the 1950s, von Neumann, who contributed to the development of the first computer (ENIAC), decided to work out a mathematical theory of automata. Such a theory was finalized to give an answer to the following question: is it possible to build an automaton such that it allows universal computation (i.e., it embodies a universal Turing machine) and, moreover, it is able to build (in order of decreasing generality)

1. an arbitrary automata (universal constructor);
2. a copy of itself (self-reproducing); and
3. an automaton that is itself a universal Turing machine (constructor)?

The last question von Neumann had intention to address was if in the process of automata self-reproduction (if possible) a process of evolution could take place, that is, if a simpler automaton could generate a more complex one.

In the beginning, the idea of von Neumann was to describe, using mathematical axioms, an automaton moving inside a warehouse and selecting various elementary spare parts (e.g., “muscles,” switches, rigid girders) and then assembling them into a new automaton. While this original idea was very realistic, it was also very difficult to pursue, so that von Neumann, following a suggestion by Ulam, decided to consider his questions in the more abstract framework of CAs.

The particular CA he considered is an infinite square CA with 29 possible states. The transition rule is dependent upon the cell to update and its north, east, south, and west neighbor cell (the von Neumann neighborhood). Among the 29 possible states there is one state that is “quiescent” (the vacuum state).

von Neumann proved the existence of a configuration of $\sim$50 000 cells immersed in a sea of quiescent states that embodies a universal Turing machine and that is a universal constructor. An infinite one-dimensional “tape” is used to store a description of the automaton to build. The universal constructor reads the description on the tape, develops a “constructing arm” that builds the configuration described on the tape in an unoccupied part of the cellular space, makes a copy of the tape and finally attaches it to the newly built automaton and retracts the constructing arm. When on the tape, it stores a description of the universal constructor itself, then it self-reproduces. The total size of the self-reproducing automaton amounts to $\sim$200 000 cells. (Some computer simulations of von Neumann self-reproducing automaton are available on the web.)

Since von Neumann’s CA is a very complex one, it led researchers to think that a CA able to simulate a universal Turing machine should also be quite complex. The perspective changed completely after the introduction of CA Life. Conway was looking for a simple CA with a possible rich dynamics; however, it was subsequently realized that Life was much more complicated that anyone could have thought. Finally, thanks to the development of faster computers that allowed visualization of the evolution of quite large populations and through the contribution of a large number of researchers, it was proved that a universal Turing machine could be embedded in Life.

The discovery that even a simple CA such as Life could incorporate a universal Turing machine led to the question whether it could be possible to build a universal Turing machine inside a simple one-dimensional CA. This is indeed the case: up to now, the simplest CA capable of universal computation is the W110 CA (see Figure 10), as proved recently by Cook after a conjecture formulated by Wolfram in 1985.

### CAs for Computer Simulations

One of the major applications of CAs is the computer simulation of various dynamical processes. Even if CAs were not invented for this purpose, they possess peculiarities that make them particularly suitable for this task. The main advantage of using a CA for a dynamical simulation is due to their completely discrete nature that allows exact simulations on a computer. Thus, any spurious effect due to rounding errors is ruled out. Another advantage is that the EL of a CA can be seen as a function between finite sets. For this reason, one can specify the EL through a “lookup table” (see [2]): then when running the simulations, the computer has only to access the table instead of computing the function every time, shortening considerably the computation time. Another great advantage of CAs in computer simulations is that, for their very nature (at least for local EL), they can be implemented on parallel machines. These two concepts are at the basis of dedicated computers for CAs simulations developed by Toffoli, Margolus, and co-workers (CAM series). The possibility to use efficiently parallel computers for CA simulation could prove

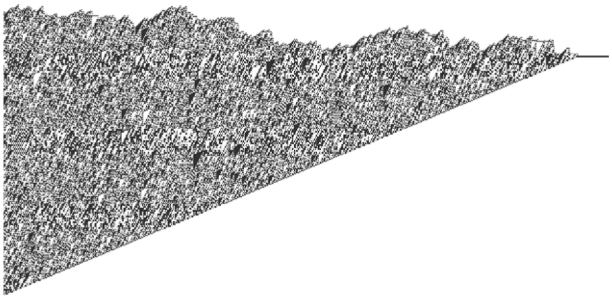


**Figure 19** A CA that "computes" the $3n+1$ Collatz–Ulam map. The ID for the CA is the initial number for the iterated map (binary notation, order $2^{300}$, randomly chosen, displayed on the left vertical axis). The CA, according to the Collatz conjecture, ends up to the final stable configuration (horizontal line on the right for the CA, $1 \rightarrow 4 \rightarrow 2 \rightarrow 1$ for the map).

to be fundamental when computer speeds approach saturation. Moreover, CAs themselves can mimic parallel computations, see, for example, Figure 19, where a nonlocal CA "computes" very efficiently the celebrated Collatz–Ulam $3n+1$ map.

### CAs in Physics

Since Newton, physics has been described through differential equations and continuous functions. However, such a mathematical description is not fit for simulation on a computer, and some discretizations must be considered. First, one has to discretize space and time passing from differential equations to (finite systems of) finite difference equations; second, one has to round off the values of the functions to store them in the memory of the computer. The main drawback of this procedure is that in chaotic systems such approximations can rapidly lead to great differences between the real and the simulated behavior. As already noticed, this problem does not appear in CA. Thus, one would like to use this good characteristic of CAs in physical modeling taking due account of the continuous nature of the physics involved. This requires attention and ingenuity in constructing reliable CA models for physical processes. For example, this goal has been achieved in the so-called lattice gas automata (LGAs).

LGAs are CA models for the microscopic dynamics of fluids and gases. The thermodynamic limit of these CAs yields the correct continuous functions for the macroscopic quantities (density, pressure, viscosity, etc.).

The first step toward LGAs was the discovery that the HPP model developed in the 1970s by Hardy, Pomeau and De Pazzis was in fact a CA. The HPP model describes the behavior of a fluid (or a gas) in a plane. The configuration space is given by a bidimensional square lattice and the particles are described by arrows lying on the edges of the lattices and pointing to some vertex (see Figure 20a).

The particles are assumed to be all identical and with the same velocity, and particles on the same edge with the same direction are not allowed (exclusion principle). The EL prescribes that particles move with unitary velocity along the edges in the direction pointed by the arrow (free flight) unless there are exactly two particles on the edges connected to a given vertex and they point in opposite directions (collision); in this case they are replaced by two arrows pointing outward on the previously empty edges (see Figure 20b). Clearly, the EL conserves the number and the momentum of the particles.

The HPP model can be described algebraically. The admissible particle velocities are just

$$c_1 = +\hat{x}, \quad c_2 = +\hat{y}, \quad c_3 = -\hat{x}, \quad c_4 = -\hat{y} \qquad [26]$$

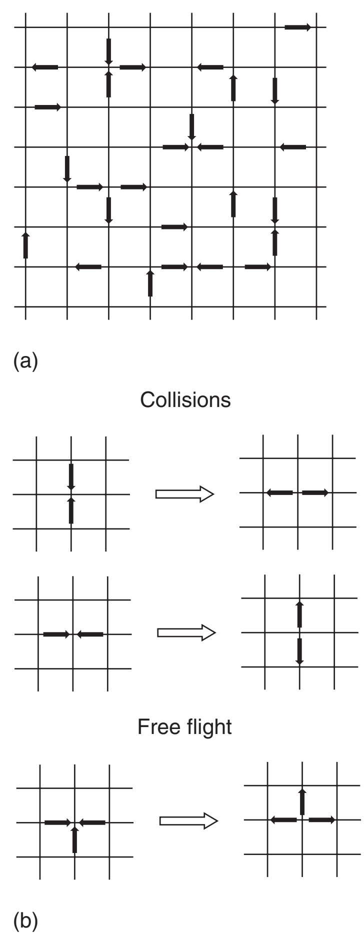



**Figure 20** (a) An example of configuration for the HPP model. (b) Head on collisions and three particle collisions in the HPP model.

Accordingly, only four bits $n_j(\boldsymbol{x},t)$, $j=1,2,3,4$, are required to denote the presence (1) or the absence (0) of a particle with velocity $c_j$ pointing vertex $\boldsymbol{x}$ at time $t$. The dynamical rule for HPP can be written in the form

$$n_j(\boldsymbol{x}+c_j,t+1)=n_j(\boldsymbol{x},t)+\omega_j(\boldsymbol{x},t) \qquad [27]$$

where term $n_j(\boldsymbol{x},t)$ on the right-hand side accounts for the free flight of particles, while $\omega_j(\boldsymbol{x},t)$ modifies the trajectories in the case of collisions. The $\omega_j$ are determined by the state of the system according to the following rules:

$$\omega_1=-n_1(1-n_2)n_3(1-n_4) + (1-n_1)n_2(1-n_3)n_4 \qquad [28a]$$

$$\omega_2=-n_2(1-n_3)n_4(1-n_1) + (1-n_2)n_3(1-n_4)n_1 \qquad [28b]$$

$$\omega_3=-n_3(1-n_4)n_1(1-n_2) + (1-n_3)n_4(1-n_1)n_2 \qquad [28c]$$

$$\omega_4=-n_4(1-n_1)n_2(1-n_3) + (1-n_4)n_1(1-n_2)n_3 \qquad [28d]$$

It is plain that eqns [27] and [28] can be interpreted as the EL for a CA.

In the thermodynamic limit, the equations governing the dynamics of the macroscopic quantities of the fluid are given by the continuity equation and by anisotropic Navier–Stokes equations. The anisotropy in the Navier–Stokes equations is due to the fact that the invariance group of the square lattice is too small. This problem was solved by Frisch, Hasslacher, and Pomeau in 1986, with the introduction of the FPP model. It turns out that a hexagonal lattice has enough symmetries to recover the isotropic Navier–Stokes equations in the thermodynamic limit. So, the FPP model is an example of a model where even if the microscopic dynamics is almost a caricature of the real dynamics, the thermodynamic limit gives rise to the correct physical equations.

CAs have been used to simulate many other physical processes (unfortunately, there is no space here for a sufficiently elaborate description). The principal fields of application are: percolation theory, magnetism, diffusion phenomena, sandpiles, models of earthquakes, crystal growth, etc.

The more intriguing aspect of some even simple CAs (e.g., CA9, CA10: see Figures 16 and 18) is their very rich particle-like dynamics. For instance, the existence of solitonic collisions suggested that the techniques recently developed to find and treat "integrable" nonlinear dynamical systems (nonlinear continuous and discrete evolution equations, many-body problems) could profitably be extended to find "integrable" CAs. Indeed, many such CAs have been found that exhibit "solitons" and are endowed with nontrivial conservation laws (of course, this is very important in physical modeling). Moreover, the above-cited similarity between certain CA behaviors and elementary particle physics phenomena suggests that the fundamental structure of reality (at the Planck level) could indeed be that of a CA (cells of Plank length, discrete time flow): attempts to construct this underlying CA physics have been pursued.

### Other Applications

CAs exhibit a great plasticity, which makes them well suited to model systems in a wide range of fields. This is mainly due to the fact that CAs with very simple rules can also simulate universal Turing machines, so that they can exhibit a very rich and complicated overall dynamics (in principle, one could simulate any dynamical system using a simple CA). There is another reason for the wide applicability of CA modeling even outside of physics: namely, it is well known that algorithms, not differential equations, are better instruments to schematize dynamical processes for complex and organized systems. Since simple algorithms can be naturally implemented on CAs, the latter are very useful for realizing simple models and simulations in many fields: biology, economics, ecology, neural networks, traffic models, etc.

Moreover, applications of CAs in informatics and specifically in cryptography and data compression have been investigated.

*See also:* Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Generic Properties of Dynamical Systems; Integrable Systems: Overview.

## Further Reading


Berlekamp ER, Conway JH, and Guy R (1982) *Winning Ways for Your Mathematical Plays*. London: Academic Press.

Boghosian BM (1999) Lattice gases and cellular automata. *Future Generation Computer Systems* 16: 171–185.

Boon JP, Dab D, Kapral R, and Lawniczak A (1996) Lattice gas automata for reactive systems. *Physics Reports* 273: 55–147.

Burks AW (ed.) (1970) *Essay on Cellular Automata*. Urbana: University of Illinois Press.

Chopard B and Droz M (1998) *Cellular Automata Modeling of Physical Systems*. Cambridge: Cambridge University Press.

Doolen G (ed.) (1990) *Lattice Gas Methods for Partial Differential Equations*. New York: Addison-Wesley.

Gardner M (1983) *Wheels, Life, and Other Mathematical Amusements*. New York: W H Freeman.

Jackson EA (1990) *Perspectives of Nonlinear Dynamics*. Cambridge: Cambridge University Press.
Toffoli T and Margolus N (1987) *Cellular Automata Machines – A New Enviroment for Modeling*. Cambridge: The MIT Press.
von Neumann J (1966) In: Burks AW (ed.) *Theory of Self-Reproducing Automata*. Urbana: University of Illinois Press.
Wolfram S (2002) *A New Kind of Science*. Champaign: Wolfram Media.

# Central Manifolds, Normal Forms

**P Bonckaert**, Universiteit Hasselt, Diepenbeek, Belgium



## Introduction

We consider differentiable dynamical systems generated by a diffeomorphism or a vector field on a manifold. We restrict to the finite-dimensional case, although some of the ideas can also be developed in the general case (Vanderbauwhede and Iooss 1992). We also restrict to the behavior near a stationary point or a periodic orbit of a flow.

Let the origin 0 of $\boldsymbol{R}^n$ be a stationary point of a $C^1$ vector field $X$, that is, $X(0)=0$. We consider the linear approximation $A=\mathrm{d}X(0)$ of $X$ at 0 and its spectrum $\sigma(A)$, which we decompose as $\sigma(A)=\sigma_s\cup\sigma_c\cup\sigma_u$, where $\sigma_s$ resp. $\sigma_c$ resp. $\sigma_u$ consists of those eigenvalues with real part $<0$ resp. $=0$ resp. $>0$. If $\sigma_c=\emptyset$ then there is no central manifold, and the stationary point 0 is called hyperbolic. Let $E_s, E_c$, and $E_u$ be the linear $A$-invariant subspaces corresponding to $\sigma_s$ resp. $\sigma_c$ resp. $\sigma_u$. Then $\boldsymbol{R}^n=E_s\oplus E_c\oplus E_u$. We look for corresponding $X$-invariant manifolds in the neighborhood of 0, in the form of graphs of maps. More precisely:

**Theorem 1** *Let the vector field $X$ above be of class $C^r$ $(1\leq r<\infty)$. There exist map germs $\phi_{ss}:(E_s,0)\to E_c\oplus E_u$, $\phi_{sc}:(E_s\oplus E_c,0)\to E_u$, $\phi_{uu}:(E_u,0)\to E_s\oplus E_c$, $\phi_{cu}:(E_c\oplus E_u,0)\to E_s$, and $\phi_c:(E_c,0)\to E_s\oplus E_u$ of class $C^r$ such that the graphs of these maps are invariant for the flow of $X$. Moreover, these maps are of class $C^r$, and their linear approximation at 0 is zero, that is, their graphs are tangent to, respectively, $E_s, E_s\oplus E_c, E_u, E_c\oplus E_u$, and $E_c$. If $X$ is of class $C^\infty$ then $\phi_{ss}$ and $\phi_{uu}$ are also of class $C^\infty$. If $X$ is analytic then $\phi_{ss}$ and $\phi_{uu}$ are also analytic.*

The graph of $\phi_c$ is called the (local) central (or, center) manifold of $X$ at 0 and it is often denoted by $W^c$. Thus, it is an invariant manifold of $X$ tangent at the generalized eigenspace of $\mathrm{d}X(0)$ corresponding to the eigenvalues having zero real part.

## (Non) uniqueness, Smoothness

Most proofs in the literature (Vanderbauwhede 1989) use a cutoff in order to construct globally defined objects, and then obtain the invariant graph as the solution of some fixed-point problem of a contraction in an appropriate function space. Although this solution is unique for the globalized problem, this is not the case at the germ level: another cutoff may produce a different germ of a central manifold. In other words, locally a central manifold might not be unique, as is easily seen on the planar example $x^2\partial/\partial x-y\partial/\partial y$. On the other hand, the $\infty$-jet of the map $\phi_c$, in case of a $C^\infty$ vector field, is unique, so if there would exist an analytic central manifold then this last one is unique; in the foregoing example, it is the $x$-axis. But for the (polynomial) example $(x-y^2)\partial/\partial x+y^2\partial/\partial y$ one can calculate that the $\infty$-jet of $x=\phi_c(y)$ is given by $j_\infty\phi_c(y)=\sum_{n\geq 1} n!y^{n+1}$, which has a vanishing radius of convergence, so there is no analytic central manifold. On the other hand, by the Borel theorem we can choose a $C^\infty$-representative for $\phi_c$. This can be generalized in the planar case:

**Proposition 1** *If $n=2$ and if $X$ is $C^\infty$ and if the $\infty$-jet of $X$ in the direction of the central manifold is nonzero, then this central manifold is $C^\infty$. In particular, if $X$ is analytic then the central manifold is either an analytic curve of stationary points or is a $C^\infty$ curve along which $X$ has a nonzero jet.*

For proofs and additional reading, the reader is referred to Aulbach (1992). In general, a central manifold is not necessarily $C^\infty$ (van Strien 1979, Arrowsmith and Place 1990): for the system in $\boldsymbol{R}^3$ given by

$$(x^2-z^2)\frac{\partial}{\partial x}+(y+x^2-z^2)\frac{\partial}{\partial y}+0\cdot\frac{\partial}{\partial z}$$

one can find a $C^k$ central manifold for every $k$ but there is no $C^\infty$ central manifold. Indeed, in this case the domain of definition of $\phi_c$ shrinks to zero when $k$ tends to infinity.

## Central Manifold Reduction

The importance of a central manifold lies in the principle of central manifold reduction, which roughly says that for local bifurcation phenomena it is enough to study the behavior on the central manifold, that is, if two vector fields, restricted to their central manifolds, have homeomorphic integral curve portraits, and if the dimensions of $E_s$ and $E_u$ are equal, then the two vector fields have homeomorphic integral curve portraits in $\boldsymbol{R}^n$, at least locally near 0. Let us be more precise:

**Theorem 2** *Let $m$ be the dimension of $E_c$. There exists $p$, $0 \leq p \leq n-m$, such that $X$ is locally $C^0$-conjugate to*

$$X' = \sum_{i=1}^{m} \tilde{X}_i(z_1, \ldots, z_m)\frac{\partial}{\partial z_i} + \sum_{i=m+1}^{m+p} z_i \frac{\partial}{\partial z_i} - \sum_{i=m+p+1}^{n} z_i \frac{\partial}{\partial z_i}$$

*where $(z_1, \ldots, z_m)$ is a coordinate system on a central manifold, $(z_1, \ldots, z_n)$ is a coordinate system on $\boldsymbol{R}^n$ extending $(z_1, \ldots, z_m)$ and $\sum_{i=1}^{m} \tilde{X}_i\, \partial/\partial z_i$ is the restriction of $X$ to a central manifold. Moreover, if*

$$Y = \sum_{i=1}^{m} \tilde{Y}_i(z_1, \ldots, z_m)\frac{\partial}{\partial z_i} + \sum_{i=m+1}^{m+p} z_i \frac{\partial}{\partial z_i} - \sum_{i=m+p+1}^{n} z_i \frac{\partial}{\partial z_i}$$

*and if $\sum_{i=1}^{m} \tilde{Y}_i\, \partial/\partial z_i$ is $C^0$-equivalent (resp. $C^0$-conjugate) to $\sum_{i=1}^{m} \tilde{X}_i\, \partial/\partial z_i$ then $X$ is $C^0$-equivalent (resp. -conjugate) to $Y$.*

For a proof and further reading (a generalization) see Palis and Takens (1977).

In case that more smoothness than just $C^0$ is needed, we have the principle of normal linearization along the central manifold. More concretely, let $x$ denote a coordinate in the central manifold and let $y$ be a complementary variable, that is, let $X = X_c \partial/\partial x + X_h \partial/\partial y$. We define the normally linear part along the central manifold by

$$NX := X_c(x, 0)\frac{\partial}{\partial x} + \frac{\partial X_h}{\partial y}(x, 0) \cdot y \frac{\partial}{\partial y}$$

Under certain nonresonance conditions (Takens 1971, Bonckaert 1997) on the real parts of the eigenvalues of $\mathrm{d}X(0)$, there exists a $C^r$ local conjugacy between $X$ and $NX$ for each $r \in \boldsymbol{N}$ (assuming $X$ to be of class $C^\infty$). If there are resonances, then one can conjugate with the so-called seminormal or renormal form containing higher-order terms (see Bonckaert (1997, 2000) and references therein; here one can also find results for cases where extra constraints should be respected, like symmetry, reversibility, or invariance of some given foliation etc.).

### Parameters

Having an eigenvalue with zero real part is ungeneric, so in bifurcation problems we consider $p$-parameter families $X_\lambda$ near, say, $\lambda = 0$. With respect to the results above, we remark that such a family can be considered as a vector field near $(0,0) \in \boldsymbol{R}^n \times \boldsymbol{R}^p$ tangent to the leaves $\boldsymbol{R}^n \times \{\lambda\}$. In fact, the parameter direction $\boldsymbol{R}^p$ is contained in $E_c$. In all the results mentioned, this structure "of being a family" is respected. For example, in Theorem 2 we replace $\tilde{X}_i(z_1, \ldots, z_m)$ by $\tilde{X}_i(z_1, \ldots, z_m, \lambda)$. Hence, if $\tilde{X}_\lambda$ is a versal unfolding of $\tilde{X}_0$ then $X_\lambda$ is a versal unfolding of $X_0$. By this, the search for versal unfoldings is reduced to the unfolding of singularities whose linear approximation at 0 has a purely imaginary spectrum.

### Diffeomorphisms, Periodic Orbits

A completely analogous theory can be developed for fixed points of diffeomorphisms $f : (\boldsymbol{R}^n, 0) \to \boldsymbol{R}^n$. Here we split up the spectrum of the linear part $L = \mathrm{d}f(0)$ at 0 as $\sigma(L) = \sigma_s \cup \sigma_c \cup \sigma_u$, where $\sigma_s$ resp. $\sigma_c$ resp. $\sigma_u$ consists of those eigenvalues with modulus $<1$ resp. $=1$ resp. $>1$. This theory can be applied to the time-$t$ map of a vector field (and will give the same invariant manifolds) and to the Poincaré map of a transversal section of a periodic orbit of a vector field (Chow *et al.* 1994).

## Normal Forms

The general idea of a normal form is to put a (complicated) system into a form "as simple as possible" by means of a change of coordinates. This idea was already developed to a great extent by H Poincaré. Simple examples are: (1) putting a square matrix into Jordan form, (2) the flow box theorem (Arrowsmith and Place 1990) near a nonsingular point. Depending on the context and on the purpose of the simplification, this concept may vary greatly. It depends on the kind of changes of coordinates that are tolerated (linear, polynomial, formal series, smooth, analytic) and on the possible structures that must be preserved (e.g., symplectic, volume-preserving, symmetric, reversible etc.). Let us restrict to local normal forms, that is, in the vicinity of a stationary point of a vector field or a diffeomorphism (the latter can be

applied to the Poincaré map of a periodic orbit). We concentrate on the simplification of the Taylor series. The general idea is to apply consecutive polynomial changes of variables; at each step we simplify terms of a degree higher than in the step before. The ideal simplification would be to put all higher-order terms to zero, which would (at least at the level of formal series) linearize the system. But as soon as there are resonances (see below), this is impossible: the planar system $2x\partial/\partial x + (y + x^2)\partial/\partial y$ cannot be formally linearized.

### Setting

Let $X$ be a $C^{r+1}$ vector field defined on a neighborhood of $0 \in \boldsymbol{R}^n$, and denote $A = \mathrm{d}X(0)$ (its linear approximation at 0). The Taylor expansion of $X$ at 0 takes the form

$$X(x) = A \cdot x + \sum_{k=2}^{r} X_k(x) + O(|x|^{r+1})$$

where $X_k \in H^k$, the space of vector fields whose components are homogeneous polynomials of degree $k$. The classical formal normal-form theorem is as follows. We define the operator $L_A$ on $H^k$ by putting $L_A h(x) = \mathrm{d}h(x) \cdot A \cdot x - A \cdot h(x)$; one calls $L_A$ the homological operator. One checks that $L_A(H^k) \subset H^k$. One also denotes this by ad $A(h)(x)$: see further in the Lie algebra setting. Let $R^k$ be the range of $L_A$, that is, $R^k = L_A(H^k)$. Let $G^k$ denote any complementary subspace to $R^k$ in $H^k$. The formal normal-form theorem states, under the above settings:

**Theorem 3** (Chow *et al.* 1994, Dumortier 1991) *There exists a composition of near identity changes of variables of the form*

$$x = y + \xi^k(y) \qquad [1]$$

*where the components of $\xi^k$ are homogeneous polynomials of degree $k$, such that the vector field $X$ is transformed into*

$$Y(y) = A \cdot y + \sum_{k=2}^{r} g_k(y) + O(|y|^{r+1})$$

*where $g_k \in G^k, k = 2, \ldots, r$.*

Sometimes this theorem is applied to the restriction of a vector field to its central manifold, for reasons explained in the last section. This is the reason why we did not assume $X$ to be $C^\infty$; in the latter case one can let $r \to \infty$ and obtain a normal form on the level of formal Taylor series (also called $\infty$-jets). Using a theorem of Borel, we infer the existence of a $C^\infty$ change of variables $\phi$ such that the Taylor series of $\phi_*(X)$ is $A \cdot y + \sum_{k=2}^{\infty} g_k(y)$. For practical computations, it is often appropriate to first simplify the linear part $A$ and to diagonalize it whenever possible. Hence, it is convenient to use a complexified setting and to use complex polynomials or power series. One can show that all involved changes of variables preserve the property of "being a complex system coming from a real system," that is, at the final stage we can return to a real system (see, e.g., Arrowsmith and Place (1990) for a more precise mathematical description).

Hence, we can assume that $A$ is an upper triangular matrix. Let the eigenvalues be $\lambda_1, \ldots, \lambda_n$. It can be calculated that the eigenvalues of $L_A$, as an operator $H^k \to H^k$, are then the numbers $\langle \lambda, \alpha \rangle - \lambda_j$ where $\alpha \in \boldsymbol{N}^n$, $\sum_{j=1}^{n} \alpha_j = k$ and $1 \le j \le n$. Hence, if these would all be nonzero then $B^k = H^k$, and then we have an ideal simplification, that is, all $g_k$ equal to zero. However, if such a number is zero, that is,

$$\langle \lambda, \alpha \rangle - \lambda_j = 0 \qquad [2]$$

it is called a resonance between the eigenvalues. In such a case, we have to choose a complementary space $G^k$. From linear algebra it follows that one can always choose

$$G^k = \ker(L_{A^*}) \qquad [3]$$

where $A^*$ is the adjoint operator. But this choice [3] is not unique and is, from the computational point of view, not always optimal, especially if there are nilpotent blocks. This fact has been exploited by many authors. A typical example is the case where $A = y\partial/\partial x$. On the other hand, if $A$ is semisimple we can choose the complementary space to be $\ker(L_A)$, so $L_A g_k = 0$; we can assume it to be the (complex) diagonal$[\lambda_1, \ldots, \lambda_n]$. In that case we can be more explicit as follows. Let $e_j = \partial/\partial x_j$ denote the standard basis on $\boldsymbol{C}^n$. For a monomial one can calculate that

$$L_A(x^\alpha e_j) = (\langle \lambda, \alpha \rangle - \lambda_j) x^\alpha e_j \qquad [4]$$

If the latter is zero, then the monomial is called resonant. This implies that the normal form can be chosen so that it only contains resonant monomials.

Putting a system into normal form not only simplifies the original system, it also gives more geometric insight on the Taylor series. To be more precise, suppose (for simplicity, this can be generalized (Dumortier 1997)) that $A$ is semisimple. One can calculate that the condition $L_A g_k = 0$ implies: $\exp(-At) g_k(\exp(At)x) = g_k(x)$ for all $t \in \boldsymbol{R}$. This means that $g_k$ is invariant for the one-parameter group $\exp(At)$. A typical example in the plane is: $A$ has eigenvalues $\mathrm{i}\lambda, -\mathrm{i}\lambda$. Note that the (only) resonances are $\langle(\mathrm{i}\lambda, -\mathrm{i}\lambda), (p+1, p)\rangle - \mathrm{i}\lambda = 0$ and

$\langle(\mathrm{i}\lambda, -\mathrm{i}\lambda),(p,p+1)\rangle + \mathrm{i}\lambda = 0$ for all $p \in \boldsymbol{N}$. We suppose that the original system was real, that is, on $\boldsymbol{R}^2$; we can choose linear coordinates such that for $z = x + \mathrm{i}y$, $\bar{z} = x - \mathrm{i}y$ the linear part is $A = \text{diagonal}[\mathrm{i}\lambda, -\mathrm{i}\lambda]$. Applying the remarks above, we conclude that the normal form only contains the monomials $(z\bar{z})^p z\partial/\partial z$ and $(z\bar{z})^p \bar{z}\partial/\partial\bar{z}$. The geometric interpretation here is that these monomials are invariant for rotations around $(0,0)$. This can also be seen on the real variant of this: the Taylor series of the (real) normalized system has the form $(\lambda + f(x^2+y^2))(x\partial/\partial y - y\partial/\partial x) + g(x^2+y^2)(x\partial/\partial x + y\partial/\partial y)$ and is invariant for rotations. Warning: the dynamic behavior of a formal normal form in the central manifold can be very different from that of the original vector field, since we are only looking at the formal level. A trivial example is (take $f = g = 0$ in the foregoing example) $X(x,y) = \lambda(x\partial y - y\partial x) - \exp(-1/(x^2))\partial/\partial x$, where orbits near $(0,0)$ spiral to $(0,0)$, whereas the normal form is just a linear rotation. This difference is due to the so-called flat terms, that is, the difference between the transformed vector field and a $C^\infty$-realization of its normalized Taylor series (or polynomial). In case of analyticity of $X$, one can ask for analyticity of the normalizing transformation $\phi$. Generically, this is not the case in many situations. The precise meaning of this "genericity condition" is too elaborate to explain in this brief review article. We provide some suggestions for further reading in the next section. One could roughly say that, in the central manifold, the normal form has too much symmetry and is too poor to model more complicated dynamics of the system, which can be "hidden in the flat terms." To quote Il'yashenko (1981): "In the theory of normal forms of analytic differential equations, divergence is the rule and convergence the exception ...."

In many applications, we want to preserve some extra structure, such as a symplectic structure, a volume form, some symmetry, reversibility, some projection etc.; the case of a projection is important since it includes vector fields depending on a parameter. Sometimes a superposition of these structures appears (e.g., a family of volume-preserving systems). We would like that the normal-form procedure respects this structure at each step. One can often formulate this in terms of vector fields belonging to some Lie subalgebra $\mathcal{L}_0$. The idea is then to use changes of variables like [1], where $\xi_k$ is then generated by a vector field in $\mathcal{L}_0$. This will guarantee that all changes of variables are "compatible" with the extra structure. Unlike the general case where we could work with monomials as in [4], we will have to consider vector fields $h_k$ in $\mathcal{L}_0$ whose components are homogeneous polynomials of degree $k$. If this can be done, one says that $\mathcal{L}_0$ respects the grading by the homogeneous polynomials. In order to fix ideas, suppose that $\mathcal{L}_0$ are the divergence-free planar vector fields. Note that a monomial $x^i y^j \partial/\partial x$ is not divergence free. We can instead use time mappings of homogeneous vector fields of the form $a(q+1)x^{p+1}y^q\partial/\partial x - a(p+1)x^p\ y^{q+1}\partial/\partial y$. Up to terms of higher order we can use the time-one map of $h_k$ instead of $x + h_k(x)$. In case that one asks for a $C^\infty$-realization of the normalizing transformation, we need an extra assumption on the extra structure, that is, on $\mathcal{L}_0$, called the Borel property: denote by $J_{\infty,0}$ the set of formal series such that each truncation is the Taylor polynomial of an element of $\mathcal{L}_0$. The extra assumption is: each element of $J_{\infty,0}$ must be the Taylor series of a $C^\infty$ vector field in $\mathcal{L}_0$. It can be proved (Broer 1981) that the following structures respect the grading and satisfy the Borel property: being an $r$-parameter family, respecting a volume form on $\boldsymbol{R}^n$, being a Hamiltonian vector field ($n$ even), and being reversible for a linear involution.

One could consider other types of grading of the Lie-algebras involved.

This method, using the framework of the so-called filtered Lie algebras, is explained and developed systematically in a more general and abstract context in Broer (1981).

In nonlocal bifurcations, such as near a homoclinic loop, for example, it is not enough to perform central manifold reduction near the singularity: a simplified smooth model in a full neighborhood of the singularity is often needed, for example, in order to compute Poincaré maps.

Let us start with the "purely" hyperbolic case (i.e., $\dim E_c = 0$). First we compute the formal normal form such as the above. If there are no resonances [2] then we can formally linearize the vector field $X$. If $X$ is $C^\infty$ then a classical theorem of Sternberg (1958) states that this linearization can be realized by a $C^\infty$ change of variables (i.e., no more flat terms remaining). In case there are resonances, we must allow nonlinear terms: the resonant monomials. In this case we can also reduce $C^\infty$ to this normal form. Using the same methods, it is also possible to reduce to a polynomial normal form, but this time using $C^k (k < \infty)$ changes of variables. More precisely, if $k$ is a given number and if we write the vector field as $X = X_N + R_N$, where $X_N$ is the Taylor polynomial up to order $N$ (which can be assumed to be in normal form) and where $R_N(x) = O(|x|^{N+1})$, then for $N$ sufficiently large there is a $C^k$ change of variables conjugating $X$ to $X_N$ near 0. The number $N$ depends on the spectrum of $A = \mathrm{d}X(0)$. An elegant proof of these facts can be found in Il'yashenko and Yakovenko (1991). For the case when extra structure must be

preserved, see Bonckaert (1997), which also deals with the partially hyperbolic case ($\dim E_c \geq 1$). As already remarked above, the case of a parameter-dependent family can be regarded as a partially hyperbolic stationary point preserving this extra structure.

The question of an analytic normal form, also in the hyperbolic case, leads to convergence questions and calls upon the so-called small-divisor problems. The classical results are due to Poincaré and Siegel. Let us summarize them; they are formulated in the complex analytic setting:

**Theorem 4**

(i) *If the convex hull of the spectrum of A does not contain $0 \in \mathbf{C}$ then X can locally be put into normal form by an analytic change of variables. Moreover, this normal form is polynomial.*

(ii) *If the spectrum $\{\lambda_1, \ldots, \lambda_n\}$ of A satisfies the condition that there exists $C > 0$ and $\mu > 0$ such that for any $m \in \mathbf{N}^n$ with $\sum_j m_j \geq 2$:*

$$|\langle(\lambda_1, \ldots, \lambda_n), m\rangle - \lambda_j| \geq \frac{C}{|m|^\mu} \qquad [5]$$

*for $1 \leq j \leq n$ then X can be locally linearized by an analytic change of variables.*

Note that case (i) contains the case where 0 is a hyperbolic source or sink. This case (i) in Theorem 4 can be extended if there are parameters: if $X$ depends analytically on a parameter $\varepsilon \in \mathbf{C}^p$ near $\varepsilon = 0$ then the change of variables is also analytic in $\varepsilon$; moreover, the normal form is then a polynomial in the space variables whose coefficients are analytically dependent on the parameter $\varepsilon$.

For case (ii) this is surely not the case, since the condition [5] is fragile: a small distortion of the parameter generically causes resonances, be it of a high order. To fix ideas, consider $n = 2$ and suppose $\lambda_1 < 0 < \lambda_2$. By a generic but arbitrary small perturbation, we can have that the ratio of these eigenvalues becomes a negative rational number $-p/q$, which gives a resonance of the form [2] with $j = 1$ and $\alpha = (q + 1, p)$, so [5] is violated.

So analytic linearization, or even a polynomial analytic normal form, is ungeneric for families of such hyperbolic stationary points. The search for analytic normal forms, that is, simplified models, for families is still under investigation. A first simplification is obtained via the stable and unstable manifold from Theorem 1, that is, the graphs of $\phi_{ss}$ and $\phi_{uu}$. When $X$ is analytic near 0 then these manifolds are also analytic. So, up to an analytic change of variables, we can assume that $E_s$ and $E_u$ are invariant, which gives a simplification of the expression of $X$. Moreover, there is analytic dependence on parameters.

For local diffeomorphisms there are completely similar theorems pertaining to all the cases considered above.

## Concluding Remarks

The concept of central manifold can be extended to more general invariant sets (see Chow *et al.* (2000) and references therein). It can also be extended to the infinite-dimensional case and can be applied to partial differential equations (Vanderbauwhede and Iooss 1992).

Concerning the generic divergence of normalizing transformations, the reader is referred to Broer and Takens (1989), Bruno (1989), Il'yashenko (1981), and Il'yashenko and Pyartli (1991). Although the power series giving the normalizing transformation generally diverges, the study of the dynamics is often performed by truncating the normal form at a certain order. Recently, Iooss and Lombardi (2005) considered the question as to what an optimal truncation is. It is shown, in case $dX(0)$ is semisimple, that the order of the normal form can be optimized so that the remainder satisfies some estimate shrinking exponentially fast to zero as a function of the radius of the domain.

Concerning normal forms preserving the Hamiltonian structure, see Birkhoff (1966) and Siegel and Moser (1995) for a starting point; this is an extended subject on its own, sometimes called Birkhoff normal form, and it would require another review article.

Further simplifications of the normal form can sometimes be obtained by taking into account nonlinear terms (instead of just $A$) in order to obtain reductions of higher-order terms (see Gaeta (2002) and especially the references therein).

Applications of normal forms and central manifolds to bifurcation theory have been explained in Dumortier (1991).

*See also:* Averaging Methods; Bifurcation Theory; Dynamical Systems and Thermodynamics; Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Finite Group Symmetry Breaking; Korteweg–de Vries Equation and Other Modulation Equations; Multiscale Approaches; Normal Forms and Semiclassical Approximation; Symmetry and Symmetry Breaking in Dynamical Systems.

## Further Reading

Arrowsmith D and Place C (1990) *Dynamical Systems*. Cambridge: Cambridge University Press.

Aulbach B (1992) One-dimensional center manifolds are $C^\infty$. *Results in Mathematics* 21: 3–11.

Birkhoff GD (1966) *Dynamical Systems*. With an addendum by Jurgen Moser. American Mathematical Society Colloquium Publications, vol. IX. Providence, RI: American Mathematical Society.
Bonckaert P (1997) Conjugacy of vector fields respecting additional properties. *Journal of Dynamical and Control Systems* 3: 419–432.
Bonckaert P (2000) Symmetric and reversible families of vector fields near a partially hyperbolic singularity. *Ergodic Theory and Dynamical Systems* 20: 1627–1638.
Broer H (1981) Formal normal forms for vector fields and some consequences for bifurcations in the volume preserving case. In: *Dynamical Systems and Turbulence, Warwick 1980*, vol. 898, Lecture Notes in Mathematics. New York: Springer.
Broer H and Takens F (1989) Formally symmetric normal forms and genericity. *Dynamics Reported. A Series in Dynamical Systems and their Applications* 2: 11–18.
Bruno AD (1989) *Local Methods in Nonlinear Differential Equations*. New York: Springer.
Chow S-N, Li C, and Wang D (1994) *Normal Forms and Bifurcations of Planar Vector Fields*. Cambridge: Cambridge University Press.
Chow S-N, Liu W, and Yi Y (2000) Center manifolds for invariant sets. *Journal of Differential Equations* 168: 355–385.
Dumortier F (1991) Local study of planar vector fields: singularities and their unfoldings. In: Van Groesen E and De Jager EM (eds.) *Structures in Dynamics, Studies in Mathematical Physics*, vol. 2, pp. 161–241. Amsterdam: Elsevier.
Gaeta G (2002) Poincaré normal and renormalized forms. *Acta Applicandae Mathematicae* 70(1–3): 113–131 (symmetry and perturbation theory).
Il'yashenko YS (1981) In the theory of normal forms of analytic differential equations violating the conditions of Bryuno divergence is the rule and convergence the exception. *Moscow University Mathematical Bulletin* 36(2): 11–18.
Il'yashenko YS and Pyartli AS (1986) Materialization of resonances and divergence of normalizing series for polynomial differential equations. *Journal of Mathematical Sciences* 32(3): 300–313.
Il'yashenko YS and Yakovenko SY (1991) Finitely smooth normal forms of local families of diffeomorphisms and vector fields. *Russian Mathematical Surveys* 46: 1–43.
Iooss G and Lombardi E (2005) Polynomial normal forms with exponentially small remainder for analytic vector fields. *Journal of Differential Equations* 212: 1–61.
Palis J and Takens F (1977) Topological equivalence of normally hyperbolic dynamical systems. *Topology* 16(4): 335–345.
Siegel CL and Moser JK (1971) *Lectures on Celestial Mechanics*, (reprint 1995). Berlin: Springer.
Sternberg S (1958) On the structure of local homeomorphisms of Euclidean *n*-space. II. *American Journal of Mathematics* 80: 623–631.
Takens F (1971) Partially hyperbolic fixed points. *Topology* 10: 133–147.
Vanderbauwhede A (1989) Center manifolds, normal forms and elementary bifurcations. In: Kirchgraber U and Walther O (eds.) *Dynamics Reported*, vol. 2, pp. 89–169. New York: Wiley.
Vanderbauwhede A and Iooss G (1992) Center manifold theory in infinite dimensions. In: Jones CKRT *et al.* (eds.) *Dynamics Reported*, vol. 1, New Series, pp. 125–163. Berlin: Springer.

# Channels in Quantum Information Theory

**M Keyl**, Università di Pavia, Pavia, Italy



## Introduction

Consider a typical quantum system such as a string of ions in a trap. To "process" the quantum information the ions carry, we have to perform in general many steps of a quite different nature. Typical examples are: free time evolution (including unwanted but unavoidable interactions with the environment), controlled time evolution (e.g., the application of a "quantum gate" in a quantum computer), preparations and measurements. Each processing step can be described by a channel which transforms input systems into output system of a possibly different type (e.g., a measurement transforms quantum systems into classical information).

### Systems, States, and Algebras

To get a unified mathematical description of systems of different physical nature, it is useful to consider $C^*$-algebras (which are, in our case, always finite dimensional): quantum systems can be represented in terms of the algebra $\mathcal{B}(\mathcal{H})$ of (bounded) operators on the Hilbert space $\mathcal{H} = \mathbb{C}^d$; for classical information we have to choose the set $\mathcal{C}(X)$ of (continuous), complex-valued functions on the finite alphabet $X$; and the tensor product of both $\mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X)$ describes hybrid systems which are half-classical and half-quantum. Assume now that $\mathcal{A}$ is one of these algebras. Effects (i.e., yes/no measurements on the system in question) are then described by $A \in \mathcal{A}$ satisfying $0 \leq A \leq \mathbb{1}$, states are positive, normalized linear functionals $\omega : \mathcal{A} \to \mathbb{C}$, and the probability to get the result "yes" during an $A$ measurement on a system in the state $\omega$ is given by $\omega(A)$. Since $\mathcal{A}$ is assumed to be finite dimensional, each state $\omega$ on $\mathcal{B}(\mathcal{H})$ is represented by a density operator $\rho$, that is, $\omega(A) = \mathrm{tr}(\rho A)$. Likewise, a state $\omega$ on $\mathcal{C}(X)$ has the form $\omega(A) = \sum_x A(x) p_x$, where $(p_x)_{x \in X}$ denotes a probability distribution on $X$, and a state $\omega$ on $\mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X)$ is described by a sequence $(\rho_x)_{x \in X}$ of positive (trace-class) operators on $\mathcal{B}(\mathcal{H})$ with $\sum_x \mathrm{tr}(\rho_x) = 1$ such that $\omega(A) = \sum_x \mathrm{tr}(\rho_x A_x)$. Here

we have used the fact that an element $A \in \mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X)$ can be represented in a canonical way by a sequence $(A_x)_{x\in X}$ of operators on $\mathcal{H}$. The set of states will be denoted in the following by $\mathcal{S}(\mathcal{A})$ and the set of effects by $\mathcal{E}(\mathcal{A})$.

### Completely Positive Maps

Our aim is now to get a mathematical object which can be used to describe a channel. To this end, consider two $C^*$-algebras, $\mathcal{A}, \mathcal{B}$, describing the input and output system, respectively, and an effect $A \in \mathcal{B}$ of the output system. If we invoke first a channel which transforms $\mathcal{A}$ systems into $\mathcal{B}$ systems, and measure $A$ afterwards on the output systems, we end up with a measurement of an effect $T(A)$ on the input systems. Hence, we get a map $T : \mathcal{E}(\mathcal{B}) \to \mathcal{E}(\mathcal{A})$ which completely describes the channel (note that the direction of the mapping arrow is reversed compared to the natural ordering of processing). Alternatively, we can look at the states and interpret a channel as a map $T^* : \mathcal{S}(\mathcal{A}) \to \mathcal{S}(\mathcal{B})$ which transforms $\mathcal{A}$ systems in the state $\rho \in \mathcal{S}(\mathcal{A})$ into $\mathcal{B}$ systems in the state $T^*(\rho)$. To distinguish between both maps, we can say that $T$ describes the channel in the Heisenberg picture and $T^*$ in the Schrödinger picture. On the level of the statistical interpretation, both points of view should coincide of course, that is, the probabilities $(T^*\rho)(A)$ and $\rho(TA)$ to get the result "yes" during an $A$ measurement on $\mathcal{B}$ systems in the state $T^*\rho$, respectively a $TA$ measurement on $\mathcal{A}$ systems in the state $\rho$, should be the same. Since $(T^*\rho)(A)$ is linear in $A$, we see immediately that $T$ must be an affine map, that is, $T(\lambda_1 A_1 + \lambda_2 A_2) = \lambda_1 T(A_1) + \lambda_2 T(A_2)$ for each convex linear combination $\lambda_1 A_1 + \lambda_2 A_2$ of effects in $\mathcal{B}$, and this in turn implies that $T$ can be extended naturally to a linear map, which we will identify in the following with the channel itself, that is, we say that $T$ *is* the channel.

Let us now change slightly our point of view and start with a linear operator $T : \mathcal{A} \to \mathcal{B}$. To be a channel, $T$ must map effects to effects, that is, $T$ has to be positive: $T(A) \geq 0 \,\forall A \geq 0$ and bounded from above by $\mathbb{1}$, that is, $T(\mathbb{1}) \leq \mathbb{1}$. In addition, it is natural to require that two channels in parallel are again a channel. More precisely, if two channels $T : \mathcal{A}_1 \to \mathcal{B}_1$ and $S : \mathcal{A}_2 \to \mathcal{B}_2$ are given, we can consider the map $T \otimes S$ which associates to each $A \otimes B \in \mathcal{A}_1 \otimes \mathcal{A}_2$ the tensor product $T(A) \otimes S(B) \in \mathcal{B}_1 \otimes \mathcal{B}_2$. It is natural to assume that $T \otimes S$ is a channel which converts composite systems of type $\mathcal{A}_1 \otimes \mathcal{A}_2$ into $\mathcal{B}_1 \otimes \mathcal{B}_2$ systems. Hence, $S \otimes T$ should be positive as well.

**Definition 1** Consider two observable algebras $\mathcal{A}, \mathcal{B}$ and a linear map $T : \mathcal{A} \to \mathcal{B} \subset \mathcal{B}(\mathcal{H})$.

(i) $T$ is called positive if $T(A) \geq 0$ holds for all positive $A \in \mathcal{A}$.
(ii) $T$ is called completely positive (CP) if $T \otimes \mathrm{Id} : \mathcal{A} \otimes \mathcal{B}(\mathbb{C}^n) \to \mathcal{B}(\mathcal{H}) \otimes \mathcal{B}(\mathbb{C}^n)$ is positive for all $n \in \mathbb{N}$. Here Id denotes the identity map on $\mathcal{B}(\mathbb{C}^n)$.
(iii) $T$ is called unital if $T(\mathbb{1}) = \mathbb{1}$ holds.

Consider now the map $T^* : \mathcal{B}^* \to \mathcal{A}^*$ which is dual to $T$, that is, $T^*\rho(A) = \rho(TA)$ for all $\rho \in \mathcal{B}^*$ and $A \in \mathcal{A}$. It is called the Schrödinger-picture representation of the channel $T$, since it maps states to states provided $T$ is unital. (Complete) positivity can be defined in the Schrödinger picture as in the Heisenberg picture, and we immediately see that $T$ is (completely) positive iff $T^*$ is.

It is natural to ask whether the distinction between positivity and complete positivity is really necessary, that is, whether there are positive maps which are not CP. If at least one of the algebras $\mathcal{A}$ or $\mathcal{B}$ is classical, the answer is no: each positive map is CP in this case. If both algebras are quantum however, complete positivity is not implied by positivity alone. The most prominent example for this fact is the transposition map.

If item (ii) holds only for a fixed $n \in \mathbb{N}$, the map $T$ is called $n$-positive. This is obviously a weaker condition than complete positivity. However, $n$-positivity implies $m$-positivity for all $m \leq n$, and for $\mathcal{A} = \mathcal{B}(\mathbb{C}^d)$ complete positivity is implied by $n$-positivity, provided $n \geq d$ holds.

Let us consider now the question whether a channel should be unital or not. We have already mentioned that $T(\mathbb{1}) \leq \mathbb{1}$ must hold since effects should be mapped to effects. If $T(\mathbb{1})$ is not equal to $\mathbb{1}$, we get $\rho(T\mathbb{1}) = T^*\rho(\mathbb{1}) < \mathbb{1}$ for the probability to measure the effect $\mathbb{1}$ on systems in the state $T^*\rho$, but this is impossible for channels which produce an output with certainty, because $\mathbb{1}$ is the effect which is always true. In other words, if a CP map is not unital, it describes a channel which sometimes produces no output at all and $T(\mathbb{1})$ is the effect which measures whether we have got an output. We will assume henceforth that channels are unital if nothing else is explicitly stated.

## Quantum Channels

In this section we will discuss some basic properties of CP maps which transform quantum systems into quantum systems, in particular the Stinespring theorem, which constitutes the most important structural result. For a more detailed presentation, including generalizations to more general input/

output algebras the reader should consult the textbook by Paulsen (2002).

### The Stinespring Theorem

Hence consider channels between quantum systems, i.e., $\mathcal{A}=\mathcal{B}(\mathcal{H}_1)$ and $\mathcal{B}=\mathcal{B}(\mathcal{H}_2)$. A fairly simple example (not necessarily unital) is given in terms of an operator $V:\mathcal{H}_1 \to \mathcal{H}_2$ by $\mathcal{B}(\mathcal{H}_1) \ni A \mapsto VAV^* \in \mathcal{B}(\mathcal{H}_2)$. A second example is the restriction to a subsystem, which is given in the Heisenberg picture by $\mathcal{B}(\mathcal{H}) \ni A \mapsto A \otimes \mathbb{1}_{\mathcal{K}} \in \mathcal{B}(\mathcal{H} \otimes \mathcal{K})$. Finally the composition $S \circ T = ST$ of two channels is again a channel. The following theorem says that each channel can be represented as a composition of these two examples [7].

**Theorem 2** (Stinespring dilation theorem). *Every completely positive map $T:\mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ has the form*

$$T(A) = V^*(A \otimes \mathbb{1}_{\mathcal{K}})V \qquad [1]$$

*with an additional Hilbert space $\mathcal{K}$ and an operator $V:\mathcal{H}_2 \to \mathcal{H}_1 \otimes \mathcal{K}$. Both (i.e., $\mathcal{K}$ and $V$) can be chosen such that the span of all $(A \otimes \mathbb{1})V\phi$ with $A \in \mathcal{B}(\mathcal{H}_1)$ and $\phi \in \mathcal{H}_2$ is dense in $\mathcal{H}_1 \otimes \mathcal{K}$. This particular decomposition is unique (up to unitary equivalence) and is called the* minimal *decomposition.*

By introducing a family $|\chi_j\rangle\langle\chi_j|$ of one-dimensional projectors with $\sum_j |\chi_j\rangle\langle\chi_j| = \mathbb{1}$, we can define the "Kraus operators" $\langle\psi, V_j\phi\rangle = \langle\psi \otimes \chi_j, V\phi\rangle$. In terms of these, we can rewrite eqn [1] in the following form (Kraus 1983):

**Corollary 3** (Kraus form). *Every CP map $T:\mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ can be written in the form*

$$T(A) = \sum_{j=1}^{N} V_j^* A V_j \qquad [2]$$

*with operators $V_j:\mathcal{H}_2 \to \mathcal{H}_1$.*

To get a third representation of channels, consider the Stinespring form [1] of $T$ and a vector $\psi \in \mathcal{K}$ such that $U(\phi \otimes \psi) = V(\phi)$ can be extended to a unitary map $U:\mathcal{H} \otimes \mathcal{K} \to \mathcal{H} \otimes \mathcal{K}$. It is then easy to see that the dual $T^*$ of $T$ can be written as:

**Corollary 4** (Ancilla form). *Assume that $T:\mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ is a channel. Then there is a Hilbert space $\mathcal{K}$, a pure state $\rho_0$, and a unitary map $U:\mathcal{H} \otimes \mathcal{K} \to \mathcal{H} \otimes \mathcal{K}$ such that*

$$T^*(\rho) = \mathrm{tr}_{\mathcal{K}}(U(\rho \otimes \rho_0)U^*) \qquad [3]$$

*holds.*

This representation of a channel has a (seemingly) very nice physical interpretation, because we can look at eqn [3] as the unitary interaction of the system with an unobservable environment, which is initially in the state $\rho_0$. The problem, however, is that there is a great arbitrariness in the choice of $U$ and $\rho_0$. This is the weakness of the ancilla form compared to the Stinespring representation.

Finally, let us state a related result. It characterizes all decompositions of a given completely positive map into completely positive summands. By analogy with results for states on abelian algebras (i.e., probability measures), we will call it a Radon–Nikodym theorem (see Arveson (1969) for a proof).

**Theorem 5** (Radon–Nikodym theorem). *Let $T_x:\mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2), x \in X$ be a family of CP maps and let $V:\mathcal{H}_2 \to \mathcal{H}_1 \otimes \mathcal{K}$ be the Stinespring operator of $\bar{T} = \sum_x T_x$; then there are uniquely determined positive operators $F_x$ in $\mathcal{B}(\mathcal{K})$ with $\sum_x F_x = \mathbb{1}$ and*

$$T_x(A) = V^*(A \otimes F_x)V \qquad [4]$$

### The Jamiołkowski Isomorphism

The subject of this section is a relation between CP maps and states of bipartite systems, first discovered by Jamiołkowski (1972), and which is very useful in translating properties of bipartite systems into properties of positive maps and vice versa.

The idea is based on the following setup. Alice and Bob share a bipartite system in a maximally entangled state

$$\chi = \frac{1}{\sqrt{d}} \sum_{\alpha=1}^{d} e_\alpha \otimes e_\alpha \in \mathcal{H} \otimes \mathcal{H} \qquad [5]$$

(where $e_1, \ldots, e_d$ denote an orthonormal basis of $\mathcal{H}$). Alice applies to her subsystem a channel $T:\mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$ while Bob does nothing. At the end of the processing, the overall system ends up in a state

$$R_T = (T \otimes \mathrm{Id})|\chi\rangle\langle\chi| \qquad [6]$$

Mathematically, eqn [6] makes sense if $T$ is only linear but not necessarily positive or CP (but then $R_T$ is not positive either). If we denote the space of all linear maps from $\mathcal{B}(\mathcal{H})$ into $\mathcal{B}(\mathcal{H}')$ by $\mathcal{L}$, we get a map

$$\mathcal{L} \ni T \mapsto R_T \in \mathcal{B}(\mathcal{K} \otimes \mathcal{H}) \qquad [7]$$

which is easily shown to be linear (i.e., $R_{\mu T + \lambda S} = \mu R_T + \lambda R_S$ for all $\lambda, \mu \in \mathbb{C}$ and all $T, S \in \mathcal{L}$). Furthermore, this map is bijective, hence a linear isomorphism.

**Theorem 6** *The map defined in eqns [7] and [6] is a linear isomorphism. The inverse map is given by*

$$\mathcal{B}(\mathcal{H}\otimes\mathcal{H}') \ni \rho \mapsto T_\rho \in \mathcal{L} \qquad [8]$$

*with*

$$\langle e'_\mu, T_\rho(\sigma)e'_\nu\rangle = d\,\mathrm{tr}\Big(\rho(|e'_\nu\rangle\langle e'_\mu| \otimes \sigma^{\mathrm{T}})\Big) \qquad [9]$$

*where $e'_1,\ldots,e'_{d'} \in \mathcal{H}'$ denote an (arbitrary) orthonormal basis of $\mathcal{H}'$ and the transposition of $\sigma$ is defined with respect to the basis $e_\alpha, \alpha = 1,\ldots,d$ used to define $\chi$ in [5].*

From the definition of $R_T$ in eqn [6], it is obvious that $R_T$ is positive, if $T$ is CP. To see that the converse is also true is not as trivial (because a transposition is involved), but it requires only a short calculation, which is omitted here. Hence, we get:

**Corollary 7** *The operator $R_T$ is positive, iff the map $T$ is CP.*

## Examples

Let us return now to the general case (i.e., arbitrary input and output algebras) and discuss several examples.

### Channels Under Symmetry

It is often useful to consider channels with special symmetry properties. To be more precise, consider a group $G$ and two unitary representations $\pi_1, \pi_2$ on the Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. A channel $T:\mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ is called covariant (with respect to $\pi_1$ and $\pi_2$) if

$$T[\pi_1(U)A\pi_1(U)^*] = \pi_2(U)T[A]\pi_2(U)^* \quad \forall A \in \mathcal{B}(\mathcal{H}_1)\,\forall U \in G \qquad [10]$$

holds. The general structure of covariant channels is governed by a fairly powerful variant of Stinesprings theorem (Keyl and Werner 1999).

**Theorem 8** *Let $G$ be a group with finite-dimensional unitary representations $\pi_j : G \to \mathrm{U}(\mathcal{H}_j)$ and $T:\mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ a $\pi_1, \pi_2$-covariant channel.*

(i) *Then there is a finite-dimensional unitary representation $\tilde{\pi} : G \to \mathrm{U}(\mathcal{K})$ and an operator $V:\mathcal{H}_2 \to \mathcal{H}_1 \otimes \mathcal{K}$ with $V\pi_2(U) = \pi_1(U) \otimes \tilde{\pi}(U)V$ and $T(A) = V^*A \otimes \mathbb{1}V$.*
(ii) *If $T = \sum_\alpha T^\alpha$ is a decomposition of $T$ in CP and covariant summands, there is a decomposition $\mathbb{1} = \sum_\alpha F^\alpha$ of the identity operator on $\mathcal{K}$ into positive operators $F^\alpha \in \mathcal{B}(\mathcal{K})$ with $[F^\alpha, \tilde{\pi}(g)] = 0$ such that $T^\alpha(X) = V^*(X \otimes F^\alpha)V$.*

The most prominent examples of covariant channels arise with $\mathcal{H}_1 = \mathcal{H}_2 = \mathbb{C}^d, G = \mathrm{U}(d)$ and $\pi_1(U) = \pi_2(U) = U$. All channels of this type are of the form

$$T(A) = (1-\vartheta)A + \vartheta d^{-1}\mathrm{tr}(A)\mathbb{1} \quad \text{with } \vartheta \in [0, d^2/(d^2-1)] \qquad [11]$$

and are known as "depolarizing channels." They often serve as a standard model for noise. Two particular cases are the ideal channel arising with $\vartheta = 0$, and the completely depolarizing channel ($\vartheta = 1$) which erases all information. If we choose $\pi_2(U) = \bar{U}$ (where the bar denotes complex conjugate) instead of $\pi_2(U) = U$, we get

$$T(A) = \frac{\vartheta}{d+1}\left[\mathrm{tr}(A)\mathbb{1} + A^{\mathrm{T}}\right] + \frac{1-\vartheta}{d-1}\left[\mathrm{tr}(A)\mathbb{1} - A^{\mathrm{T}}\right], \quad \vartheta \in [0,1] \qquad [12]$$

If we map these channels to states of bipartite systems (using the Jamiołkowski isomorphism from the last section), we get "Isotropic states" from eqn [11] and "Werner states" from [12].

### Classical Channels

The classical analog to a quantum operation is a channel $T:\mathcal{C}(X) \to \mathcal{C}(Y)$ which describes the transmission or manipulation of classical information. As already mentioned in the subsection "Completely positive maps," positivity and complete positivity are equivalent in this case. Hence, we have to assume only that $T$ is positive and unital. Obviously, $T$ is characterized by its matrix elements $T_{xy} = \delta_y(Te_x)$, where $\delta_y \in \mathcal{C}^*(X)$ denotes the Dirac measure at $y \in Y$ and $e_x \in \mathcal{C}(X)$ is the canonical basis in $\mathcal{C}(X)$. More precisely, $\delta_y$ and $e_x$ denote, respectively, the probability distribution and the function on $X$, given by

$$\delta_y = (\delta_{xy})_{x\in X} \quad \text{and} \quad e_x(y) = \delta_{xy} \qquad [13]$$

We will keep this notation up to the end of this article. Positivity and normalization of $T$ imply that $0 \le T_{xy} \le 1$ and

$$1 = \delta_y(\mathbb{1}) = \delta_y(T(\mathbb{1})) = \delta_y\left[T\left(\sum_x e_x\right)\right] = \sum_x T_{xy} \qquad [14]$$

holds. Hence the family $(T_{xy})_{x\in X}$ is a probability distribution on $X$ and $T_{xy}$ is, therefore, the transition probability to get the information $x \in X$ at the output side of the channel if $y \in Y$ was sent.

### Observables

Let us consider now a channel which transforms quantum information $\mathcal{B}(\mathcal{H})$ into classical information $\mathcal{C}(X)$. Since positivity and complete positivity are again equivalent, we just have to look at a positive and unital map $E:\mathcal{C}(X) \to \mathcal{B}(\mathcal{H})$. With the canonical basis $e_x, x \in X$, of $\mathcal{C}(X)$, we get a family $E_x = E(e_x)$, $x \in X$, of positive operators $E_x \in \mathcal{B}(\mathcal{H})$ with $\sum_{x\in X} E_x = \mathbb{1}$. Hence, the $E_x$ form a positive operator valued (POV) measure, i.e., an observable. If, on the other hand, a POV measure $E_x \in \mathcal{B}(\mathcal{H})$, $x \in X$, is given, we can define a quantum-to-classical channel $E:\mathcal{C}(X) \to \mathcal{B}(\mathcal{H})$ by

$$E(f) = \sum_{x\in X} f(x)E_x \qquad [15]$$

This shows that the observable $E_x, x \in X$, and the channel $E$ can be identified.

### Preparations

Let us now exchange the role of $\mathcal{C}(X)$ and $\mathcal{B}(\mathcal{H})$; in other words, let us consider a channel $R:\mathcal{B}(\mathcal{H}) \to \mathcal{C}(X)$ with a classical input and a quantum output algebra. In the Schrödinger picture, we get a family of density matrices $\rho_x := R^*(\delta_x) \in \mathcal{B}^*(\mathcal{H}), x \in X$, where $\delta_x \in \mathcal{C}^*(X)$ denotes again the Dirac measure on $X$. Hence, we get a parameter-dependent preparation that can be used to encode the classical information $x \in X$ into the quantum information $\rho_x \in \mathcal{B}^*(\mathcal{H})$.

### Instruments

An observable describes only the statistics of measuring results, but does not contain information about the state of the system after the measurement. To get a description which fills this gap, we have to consider channels which operate on quantum systems and produce hybrid systems as output, that is, $T:\mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X) \to \mathcal{B}(\mathcal{K})$. Following Davies (1976), we will call such an object an instrument. From $T$ we can derive the subchannel

$$\mathcal{C}(X) \ni f \mapsto T(\mathbb{1} \otimes f) \in \mathcal{B}(\mathcal{K}) \qquad [16]$$

which is the observable measured by $T$, that is, $\mathrm{tr}(T(\mathbb{1} \otimes e_x)\rho)$ is the probability to measure $x \in X$ on systems in the state $\rho$. On the other hand, we get for each $x \in X$ a quantum channel (which is not unital)

$$\mathcal{B}(\mathcal{H}) \ni A \mapsto T_x(A) = T(A \otimes e_x) \in \mathcal{B}(\mathcal{K}) \qquad [17]$$

It describes the operation performed by the instrument $T$ if $x \in X$ was measured. More precisely, if a measurement on systems in the state $\rho$ gives the result $x \in X$, we get (up to normalization) the state $T_x^*(\rho)$ after the measurement, while

$$\mathrm{tr}\big(T_x^*(\rho)\big) = \mathrm{tr}\big(T_x^*(\rho)\mathbb{1}\big) = \mathrm{tr}(\rho T(\mathbb{1} \otimes e_x)) \qquad [18]$$

is (again) the probability to measure $x \in X$ on $\rho$. The instrument $T$ can be expressed in terms of the operations $T_x$ by

$$T(A \otimes f) = \sum_x f(x)T_x(A) \qquad [19]$$

Hence, we can identify $T$ with the family $T_x, x \in X$. Finally, we can consider the second marginal of $T$

$$\mathcal{B}(\mathcal{H}) \ni A \mapsto T(A \otimes \mathbb{1}) = \sum_{x\in X} T_x(A) \in \mathcal{B}(\mathcal{K}) \qquad [20]$$

It describes the operation we get if the outcome of the measurement is ignored.

The best-known example of an instrument is a von Neumann–Lüders measurement associated with a PV measure given by family of projections $E_x, x = 1, \ldots, d$; for example, the eigenprojections of a self-adjoint operator $A \in \mathcal{B}(\mathcal{H})$. It is defined as the channel

$$T : \mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X) \to \mathcal{B}(\mathcal{H}) \quad \text{with } X = \{1, \ldots, d\} \text{ and } T_x(A) = E_x A E_x \qquad [21]$$

Hence, we get the final state $\mathrm{tr}(E_x\rho)^{-1}E_x\rho E_x$ if we measure the value $x \in X$ on systems initially in the state $\rho$ – this is well known from quantum mechanics.

### Parameter-Dependent Operations

Let us change now the role of $\mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X)$ and $\mathcal{B}(\mathcal{K})$; in other words, consider a channel $T:\mathcal{B}(\mathcal{K}) \to \mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X)$ with hybrid input and quantum output. It describes a device which changes the state of a system depending on the additional classical information. As for an instrument, $T$ decomposes into a family of (unital!) channels $T_x:\mathcal{B}(\mathcal{K}) \to \mathcal{B}(\mathcal{H})$ such that we get $T^*(\rho \otimes p) = \sum_x p_x T_x^*(\rho)$ in the Schrödinger picture. Physically, $T$ describes a parameter-dependent operation: depending on the classical information $x \in X$, the quantum information $\rho \in \mathcal{B}(\mathcal{K})$ is transformed by the operation $T_x$.

Finally, we can consider a channel $T:\mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X) \to \mathcal{B}(\mathcal{K}) \otimes \mathcal{C}(Y)$ with hybrid input and output to get a parameter-dependent instrument: similarly to the above discussion, we can define a family of instruments $T_y:\mathcal{B}(\mathcal{H}) \otimes \mathcal{C}(X) \to \mathcal{B}(\mathcal{K}), y \in Y$, by the equation $T^*(\rho \otimes p) = \sum_y p_y T_y^*(\rho)$. Physically, $T$ describes the following device: it receives the classical information $y \in Y$ and a quantum system in the state $\rho \in \mathcal{B}^*(\mathcal{K})$ as input. Depending on $y$, a measurement with the instrument $T_y$ is performed, which in turn produces the measuring value $x \in X$ and leaves the quantum system in the state (up to normalization) $T_{y,x}^*(\rho)$; with $T_{y,x}$ given as in eqn [17] by $T_{y,x}(A) = T_y(A \otimes e_x)$.

*See also:* Capacities Enhanced by Entanglement; Capacity for Quantum Information; Entanglement; Optimal Cloning of Quantum States; Positive Maps on $C^*$-Algebras; Quantum Channels: Classical Capacity; Quantum Dynamical Semigroups; Quantum Entropy; Quantum Spin Systems; Source Coding in Quantum Information Theory.

## Further Reading


Arveson W (1969) Subalgebras of $C^*$-algebras. *Acta Mathematica* 123: 141–224.

Davies EB (1976) *Quantum Theory of Open Systems*. London: Academic Press.

Jamiołkowski A (1972) Linear transformations which preserve trace and positive semidefiniteness of operators. *Reports on Mathematical Physics* 3: 275–278.

Keyl M and Werner RF (1999) Optimal cloning of pure states, testing single clones. *Journal of Mathematical Physics* 40: 3283–3299.

Kraus K (1983) *States Effects and Operations*. Berlin: Springer.

Paulsen VI (2002) *Completely Bounded Maps and Dilations*. Cambridge: Cambridge University Press.

Stinespring WF (1955) Positive functions on $C^*$-algebras. *Proceedings of the American Mathematical Society* 6: 211–216.


# Chaos and Attractors

**R Gilmore**, Drexel University, Philadelphia, PA, USA



## Introduction

Chaos is a type of behavior that can be exhibited by a large class of physical systems and their mathematical models. These systems are deterministic. They are modeled by sets of coupled nonlinear ordinary differential equations (ODEs):

$$\dot{x}_i = \frac{\mathrm{d}x_i}{\mathrm{d}t} = f_i(x; c) \qquad [1]$$

called dynamical systems. The coordinates $x$ designate points in a state space or phase space. Typically, $x \in R^n$ or some $n$-dimensional manifold for some $n \geq 3$, and $c \in R^k$ are called control parameters. They describe parameters that can be controlled in physical systems, such as pumping rates in lasers or flow rates in chemical mixing reactions. The most important mathematical property of dynamical systems is the uniqueness theorem, which states that there is a unique trajectory through every point at which $f(x; c)$ is continuous and Lipschitz and $f(x; c) \neq 0$. In particular, two distinct periodic orbits cannot have any points in common.

The properties of dynamical systems are governed, in lowest order, by the number, stability, and distribution of their fixed points, defined by $\dot{x}_i = f_i(x; c) = 0$. It can happen that a dynamical system has no stable fixed points and no stable limit cycles ($x(t) = x(t+T)$, some $T > 0$, all $t$). In such cases, if the solution is bounded and recurrent but not periodic, it represents an unfamiliar type of attractor. If the system exhibits "sensitivity to initial conditions" ($|x(t) - y(t)| \sim e^{\lambda t}|x(0) - y(0)|$ for $|x(0) - y(0)| = \epsilon$ and $\lambda > 0$ for most $x(0)$), the solution set is called a "chaotic attractor." If the attractor has fractal structure, it is called a "strange attractor."

Tools to study strange attractors have been developed that depend on three types of mathematics: geometry, dynamics, and topology.

Geometric tools attempt to study the metric relations among points in a strange attractor. These include a spectrum of fractal dimensions. These real numbers are difficult to compute, require very long, very clean data sets, provide a number without error estimates for which there is no underlying statistical theory, and provide very little information about the attractor.

Dynamical tools include estimation of Lyapunov exponents and a Lyapunov dimension. They include globally averaged exponents and local Lyapunov exponents. These are eigenvalues related to the different stretching ($\lambda > 0$) and squeezing ($\lambda < 0$) eigendirections in the phase space. To each globally averaged Lyapunov exponent $\lambda_i$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, there corresponds a "partial dimension" $\epsilon_i$, $0 \leq \epsilon_i \leq 1$, with $\epsilon_i = 1$ if $\lambda_i \geq 0$. The Lyapunov dimension is the sum of the partial dimensions $d_{\mathrm{L}} = \sum_{i=1}^{n} \epsilon_i$. That the partial dimension $\epsilon_i = 1$ for $\lambda_i \geq 0$ indicates that the flow is smooth in the stretching ($\lambda_i > 0$) and flow directions and fractal in the squeezing ($\lambda_i < 0$) directions with $\epsilon_i < 1$. Dynamical indices provide some useful information about a strange attractor. In particular, they can be used to estimate some fractal properties of a strange attractor, but not vice versa.

Topological tools are very powerful for a restricted class of dynamical systems. These are dynamical systems in three dimensions ($n = 3$). For such systems there are three Lyapunov exponents $\lambda_1 > \lambda_2 > \lambda_3$, with $\lambda_1 > 0$ describing the stretching direction and responsible for "sensitivity to initial conditions," $\lambda_2 = 0$ describing the direction of the flow, and $\lambda_3 < 0$ describing the squeezing direction

and responsible for "recurrence." Strange attractors are generated by dissipative dynamical systems, which satisfy the additional condition $\lambda_1 + \lambda_2 + \lambda_3 < 0$. For such attractors, $\epsilon_1 = \epsilon_2 = 1$ and $\epsilon_3 = \lambda_1/|\lambda_3|$ by the Kaplan–Yorke conjecture, so that $d_L = 2 + \epsilon_3 = 2 + \lambda_1/|\lambda_3|$.

A number of tools from classical topology have been exploited to probe the structure of strange attractors in three dimensions. These include the Gauss linking number, the Euler characteristic, the Poincaré–Hopf index theorem, and braid theory. More recent topological contributions include several definitions for entropy, the development of a theory for knot holders or braid holders (also called branched manifolds), the Birman–Williams theorem for these objects, and relative rotation rates, a topological index for individual periodic orbits and orbit pairs.

Three-dimensional strange attractors are remarkably well understood; those in higher dimensions are not. As a result, the description that follows is largely restricted to strange attractors with $d_L < 3$ that exist in $R^3$ or other three-dimensional manifolds (e.g., $R^2 \times S^1$). The obstacle to progress in higher dimensions is the lack of a higher-dimensional analog of the Gauss linking number for orbit pairs in $R^3$.

## Overview

The program described below has two objectives:

1. classify the global topological structure of strange attractors in $R^3$; and
2. determine the "perestroikas" (changes) that such attractors can undergo as experimental conditions or control parameters change.

Four levels of structure are required to complete this program. Each is topological and discretely quantifiable. This provides a beautiful interaction between a rigidity of structure, demanded by topological constraints, and freedom within this rigidity. These four levels of structure are:

1. basis sets of orbits,
2. branched manifolds or knot holders,
3. bounding tori, and
4. embeddings of bounding tori.

## Branched Manifolds: Stretching and Squeezing

A strange attractor is generated by the repetition of two mechanisms: stretching and squeezing. Stretching occurs in the directions identified by the positive Lyapunov exponents and squeezing occurs in the directions identified by the negative Lyapunov exponents. In $R^3$ there is one stretching direction and one squeezing direction.

A simple stretch-and-squeeze mechanism that nature appears to be very fond of is illustrated in Figure 1. In this illustration, a cube of initial conditions at (a) is advected by the flow in a short time to (b). During this process, the cube is deformed by being stretched ($\lambda_1 > 0$). It also shrinks in a transverse direction ($\lambda_3 < 0$). During the initial phase of this deformation, two nearby points typically separate exponentially in time. If they were to continue to separate exponentially for all times, the invariant set would not be bounded. Therefore, this separation cannot continue indefinitely, and in fact it must somehow reverse itself after some time because the motion is recurrent. The mechanism shown in Figure 1 involves folding, which begins between (b) and (c) and continues through to (d). Squeezing occurs where points from distant parts of the attractor approach each other exponentially, as at (d). Finally, the cube, shown deformed at (d), returns to the neighborhood of initial conditions (a). This process repeats itself and builds up the strange attractor. As can be inferred from this figure, the strange attractor constructed by the repetitive process is smooth in the expanding ($\lambda_1$) and flow ($\lambda_2 = 0$) directions but fractal in the squeezing ($\lambda_3$) direction. The attractor's fractal dimension is $\epsilon_1 + \epsilon_2 + \epsilon_3 = 2 + \epsilon_3 = 2 + \lambda_1/|\lambda_3|$.

Figure 1 summarizes the boundedness and recurrence conditions that were introduced to define strange attractors, and illustrates one stretching and squeezing mechanism that occurs repetitively to build up the fractal structure of the strange attractor

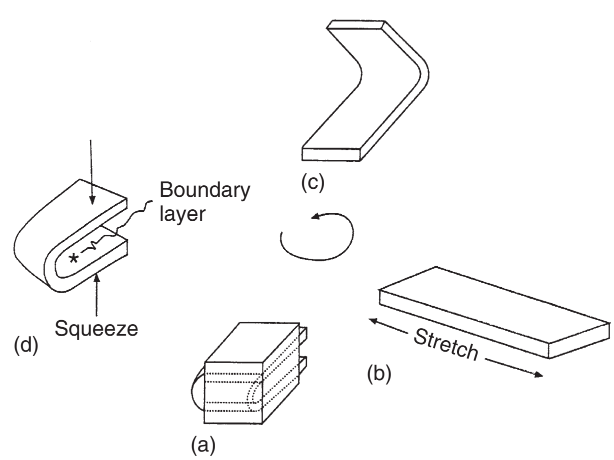



**Figure 1** A common stretch-and-fold mechanism generates many experimentally observed strange attractors. *The Topology of Chaos*; R Gilmore and M Lefranc; Copyright © 2002, Wiley. This material is used by permission of John Wiley & Sons, Inc.

and to organize all the (unstable) periodic orbits in it in a unique way. The particular mechanism shown in Figure 1 is called a stretch-and-fold mechanism. Other mechanisms involve stretch and roll, and tear and squeeze.

The stretch-and-squeeze mechanisms are well summarized by the cartoons shown in Figure 2. On the left, a cube of initial conditions (top) is deformed under the flow. The flow is downward. Stretching occurs in one direction (horizontal) and shrinking occurs in a transverse direction (perpendicular to the page). In the limit of extreme shrinking ($\lambda_3 \rightarrow -$"$\infty$"), the dynamics of the stretching part of the flow is represented by the two-dimensional surface shown on the bottom left. This surface fails to be a manifold because of the singularity, called a splitting point. This singularity represents an initial condition that flows to an unstable fixed point with at least one stable direction. On the right (squeezing), two distant cubes of initial conditions (top) in the flow are deformed and brought to each other's proximity under the flow (middle). In the limit of extreme dissipation, two two-dimensional surfaces representing inflows are joined at a branch line to a single surface representing an outflow. This surface fails to be a manifold because of the branch line, which is a singularity of a different kind. Points below the branch line in this representation of the flow (on the

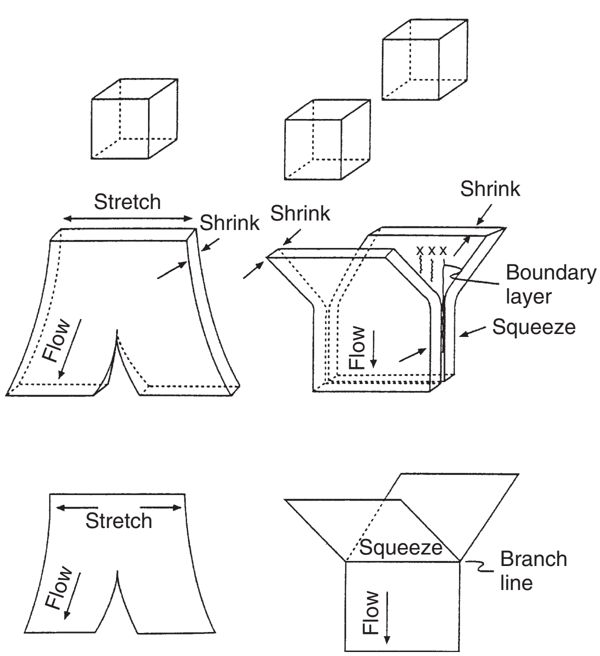



**Figure 2** Left: The stretch mechanism is modeled by a two-dimensional surface with a splitting point singularity. Right: The squeeze mechanism is modeled by a two-dimensional surface with a branch line singularity. *The Topology of Chaos*; R Gilmore and M Lefranc; Copyright © 2002, Wiley. This material is used by permission of John Wiley & Sons, Inc.

outflow side of the branch line) have two preimages above the branch line, one in each inflow sheet. This structure generates positive entropy.

A beautiful theorem of Birman and Williams justifies the use of the two cartoons shown at the bottom of Figure 2 to characterize strange attractors in $R^3$. As preparation for the theorem, Birman and Williams introduced an important identification for the nongeneric or atypical points that "are not sensitive to initial conditions"

$$x \sim y \quad \text{if} \quad |x(t) - y(t)| \stackrel{t\rightarrow\infty}{\longrightarrow} 0 \qquad [2]$$

That is, two points in a strange attractor are identified if they have asymptotically the same future. In practice, this amounts to projecting the flow down along the stable ($\lambda_3 < 0$) direction onto a two-dimensional surface described by the stretching ($\lambda_1 > 0$) and the flow ($\lambda_2 = 0$) directions. This surface is not a manifold because of lower-dimensional singularities: splitting points and branch lines. The two-dimensional surface has many names, for example, knot holder (because it holds the periodic orbits that exist in abundance in strange attractors), braid holders, templates, branched manifolds. The flow, restricted to this surface, is called a semiflow. Under the semiflow, points in the branched manifold have a unique future but do not have a unique past. The degree of nonuniqueness is measured by the topological entropy of the dynamical system. The Birman–Williams theorem is:

**Theorem** *Assume that a flow* $\Phi_t$

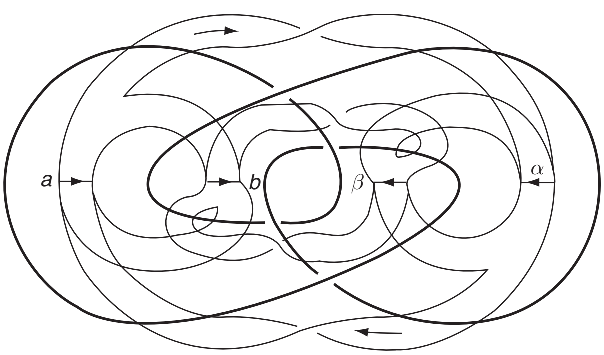
(i) *on* $R^3$ *is dissipative* ($\lambda_1 > 0$, $\lambda_2 = 0$, $\lambda_3 < 0$ *and* $\lambda_1 + \lambda_2 + \lambda_3 < 0$);
(ii) *generates a hyperbolic strange attractor* (*the eigenvectors of the local Lyapunov exponents* $\lambda_1$, $\lambda_2$, $\lambda_3$ *span everywhere on the attractor*).

*Then the projection* [2] *maps the strange attractor* $\mathcal{SA}$ *to a branched manifold* $\mathcal{BM}$ *and the flow* $\Phi_t$ *on* $\mathcal{SA}$ *to a semiflow* $\hat{\Phi}_t$ *on* $\mathcal{BM}$ *in* $R^3$. *The periodic orbits in* $\mathcal{SA}$ *under* $\Phi_t$ *correspond* 1:1 *with the periodic orbits in* $\mathcal{BM}$ *under* $\hat{\Phi}_t$ *with perhaps one or two specified exceptions. On any finite subset of orbits the correspondence can be taken via isotopy.*

The beauty of this theorem is that it guarantees that a flow $\Phi_t$ that generates a (fractal) strange attractor $\mathcal{SA}$ can be continuously deformed to a new flow $\hat{\Phi}_t$ on a simple two-dimensional structure $\mathcal{BM}$. During this deformation, periodic orbits are neither created nor destroyed. The uniqueness theorem for ODEs is satisfied during the deformation, so orbit segments do not pass through each other. As a result, the topological organization of all the

unstable periodic orbits in the strange attractor is the same as the topological organization of all the unstable periodic orbits in the branched manifold. In fact, the branched manifold (knot holder) defines the topological organization of all the unstable periodic orbits that it supports. Topological organization is defined by the Gauss linking number and the relative rotation rates, another braid index.

The significance of this theorem is that strange attractors can be characterized – in fact classified – by their branched manifolds. Figure 3 shows a branched manifold "for a figure-8 knot" as well as the figure-8 knot itself (dark curve). If a constant current is sent through a conducting wire tied into the shape of a figure-8 knot, a discrete countable set of magnetic field lines will be closed. These closed field lines can be deformed onto the two-dimensional surface shown in Figure 3. Each of the eight branches of this branched manifold can be named. One way to do this specifies the two branch lines that are joined by the branch in the sense of the flow (e.g., $(a\alpha)$ and $(\beta\alpha)$ (but not $(a\beta)$). Every closed field line can be labeled by a symbol sequence that is unique up to cyclic permutation. This symbol sequence provides a symbolic name for the orbit. For example, $(a\alpha)(\alpha\beta)(\beta b)(ba)$ is a period-4 orbit. The structure of a branched manifold is determined in part by a transition matrix $T$. The matrix element $T_{ij}$ is 1 if the transition from branch $i$ to branch $j$ is allowed, 0 otherwise. The transition matrix for the figure-8 branched manifold is shown in Figure 3.




| | ab | aα | αβ | αa | ba | bβ | βα | βb |
|---|---|---|---|---|---|---|---|---|
| ab | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| aα | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| αβ | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| αa | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| ba | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| bβ | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| βα | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| βb | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |

**Figure 3** Figure-8 knot (dark curve) and the figure-8 branched manifold. Transition matrix for the eight branches of the figure-8 branched manifold is also shown. Flow direction is shown by arrows. *The Topology of Chaos*; R Gilmore and M Lefranc; Copyright © 2002, Wiley. This material is used by permission of John Wiley & Sons, Inc.

The Birman–Williams theorem is stronger than its statement suggests. More systems satisfy the statement of the theorem than do the assumptions of the theorem. The figure-8 knot, and its attendant magnetic field, is not dissipative – in fact, it is not even a dynamical system, yet the closed loops can be isotoped to the figure-8 knot holder. There are other ways in which the Birman–Williams theorem is stronger than its statement suggests.

It is apparent from Figure 3 that the figure-8 branched manifold can be built up Lego© fashion from the two basic building blocks shown in Figure 2. This is more generally true. Every branched manifold can be built up, Lego© fashion, from the stretch (with a splitting point singularity) and the squeeze (with a branch line singularity) building blocks, subject to the following two conditions:

1. outputs flow to inputs and
2. there are no free ends.

The figure-8 branched manifold is built up from four stretch and four squeeze building blocks. As a result, there are eight branches and four branch lines.

Two often-studied strange attractors are shown in Figures 4 and 5. Figure 4 shows the details of the Rössler dynamical system. A similar spectrum of features is shown in Figure 5 for the Lorenz equations. The knot holder in Figure 5e is obtained from the caricature in Figure 5d by twisting the right-hand lobe by $\pi$ radians.

Branched manifolds can be used to characterize all three-dimensional strange attractors. Branched manifolds that classify the strange attractors generated by four familiar sets of equations (for some control parameter values) are shown in Figure 6. The sets of equations, and one set of parameter values that generate strange attractors, are presented in Table 1.

The beauty of this topological classification of strange attractors is that it is apparent, just by inspection, that there is no smooth change of variables that will map any of these systems to any of the others for the parameter values shown.

Branched manifolds can be described algebraically. In Figure 7 we provide the algebraic

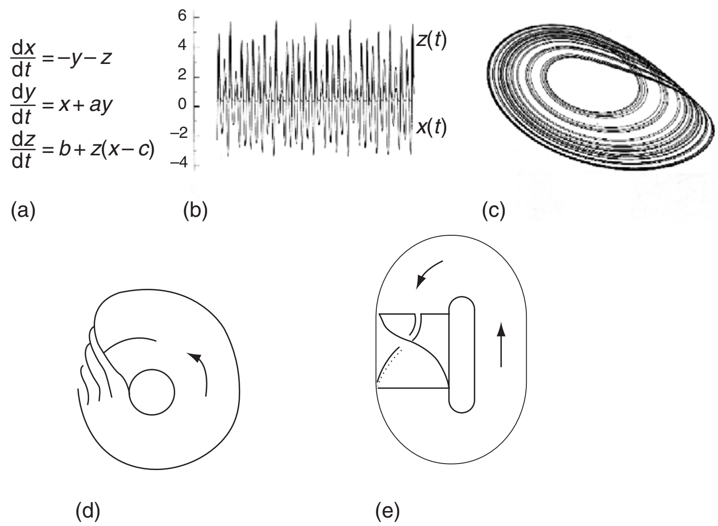



**Figure 4** The Rössler dynamical system. (a) Rössler equations. (b) Time series $z(t)$ and $x(t)$ generated by these equations, and (c) projection of the strange attractor onto the $x$–$y$ plane. (d) Caricature of the flow and (e) knot holder derived directly from the caricature. Control parameter values $(a, b, c) = (2.0, 4.0, 0.398)$. *The Topology of Chaos*; R Gilmore and M Lefranc; Copyright © 2002, Wiley. This material is used by permission of John Wiley & Sons, Inc.

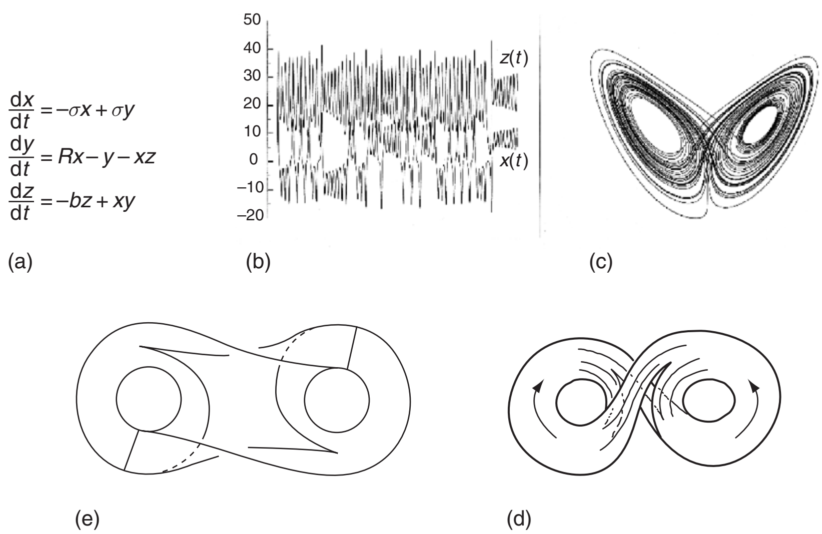



**Figure 5** (a) Lorenz equations. (b) Time series $x(t)$ and $z(t)$ generated by these equations, and (c) projection of the strange attractor onto the $x$–$y$ plane. (d) Caricature of the flow and (e) knot holder derived directly from the caricature by rotating the right-hand lobe by $\pi$ radians. Control parameter values $(R, \sigma, b) = (26.0, 10.0, 8/3)$. *The Topology of Chaos*; R Gilmore and M Lefranc; Copyright © 2002, Wiley. This material is used by permission of John Wiley & Sons, Inc.

description of two branched manifolds. Figure 7a shows the branched manifold that describes experimental data generated by many physical systems. The mechanism is a simple stretch-and-fold deformation with zero global torsion that generates a typical Smale horseshoe. There are two branches. The diagonal elements of the matrix identify the local torsion of the flow through the corresponding branch, measured in units of $\pi$. Branch 0 has no local torsion, and branch 1 shows a half-twist and has local torsion +1. The off-diagonal matrix elements are twice the linking number of the period-1 orbits in the corresponding pair of branches. Since the period-1 orbits in these two branches do not link, the off-diagonal matrix elements are 0. The period-1 orbits in the branches labeled 1 and 2 in Figure 7b have linking number +1, so the off-diagonal matrix elements are $T(1, 2) = T(2, 1) = 2 \times +1$. The array identifies the order (above, below) that the two branches are joined at the branch line, the smaller the value, the closer to the viewer. These two pieces of information, four integers in Figure 7a and eight in

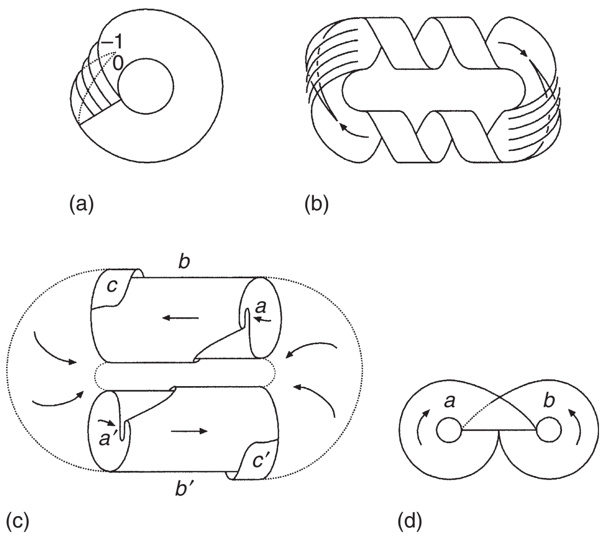



**Figure 6** Branched manifolds for four standard sets of equations: (a) Rössler equations, (b) periodically driven Duffing equations, (c) periodically driven van der Pol equations, and (d) Lorenz equations. *The Topology of Chaos*; R Gilmore and M Lefranc; Copyright © 2002, Wiley. This material is used by permission of John Wiley & Sons, Inc.

**Table 1** Four sets of equations that generate strange attractors

| *Dynamical system* | *ODEs* | *Parameter values* |
|---|---|---|
| Rössler | $\dot{x} = -y - z$<br>$\dot{y} = x + ay$<br>$\dot{z} = b + z(x - c)$ | $(a, b, c) = (2.0, 4.0, 0.398)$ |
| Duffing | $\dot{x} = y$<br>$\dot{y} = -\delta y - x^3 + x + A\sin(\omega t)$ | $(\delta, A, \omega) = (0.4, 0.4, 1.0)$ |
| van der Pol | $\dot{x} = by + (c - dy^2)x$<br>$\dot{y} = -x + A\sin(\omega t)$ | $(b, c, d, A, \omega) = (0.7, 1.0, 10.0, 0.25, \pi/2)$ |
| Lorenz | $\dot{x} = -\sigma x + \sigma y$<br>$\dot{y} = Rx - y - xz$<br>$\dot{z} = -bz + xy$ | $(R, \sigma, b) = (26.0, 10.0, 8/3)$ |

Figure 7b, serve to determine the topological organization of all the unstable periodic orbits in any strange attractor with either branched manifold.

The periodic orbits are identified by a repeating symbol sequence of least period $p$, which is unique up to cyclic permutation. The symbol sequence consists of a string of integers, sequentially identifying the branches through which the orbit passes. For a branched manifold with two branches, there are two symbols. The number of orbits of period $p, N(p)$, obeys the recursion relation

$$pN(p) = 2^p - \sum_{1=k|p}^{k \leq p/2} kN(k) \qquad [3]$$

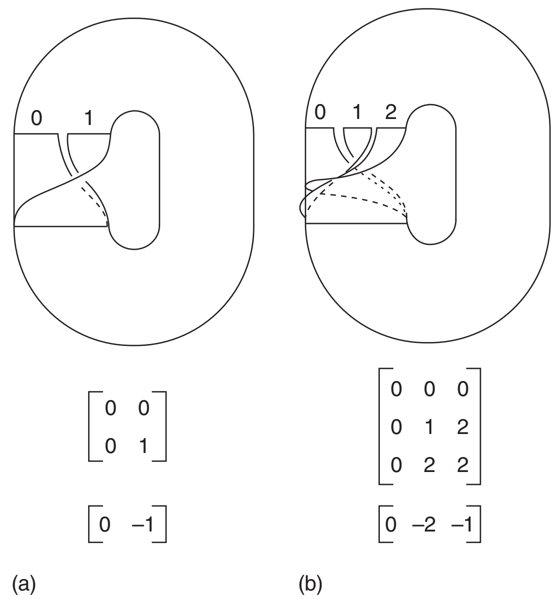



**Figure 7** Branched manifolds are described algebraically. The diagonal matrix elements describe the twist of each branch. The off-diagonal matrix elements are twice the linking number of the period-1 orbits in each of the two branches. The array describes the order in which the branches are connected at the branch line. (a) Smale horseshoe branched manifold. (b) Beginning of a "gateau roulé" (jelly roll) branched manifold.

Table 2 shows the number of orbits of period $p \leq 20$ for the branched manifolds with two and three branches shown in Figure 7. The number of orbits of period $p$ grows exponentially with $p$, and the limit $h_T = \lim_{p \to \infty} \log(N(p))/p$ defines the topological entropy $h_T$ for the branched manifold. The limits are $\ln 2$ and $\ln 3$ for the branched manifolds with two and three branches, respectively. The linking numbers of orbits up to period 5 in the Smale horseshoe branched manifold are shown in Table 3, which identifies each of the orbits by its symbol sequence (e.g., 00111).

**Table 2** Number of orbits of period $p$ on the branched manifolds with two and three branches, shown in **Figure 7**. The integers $N_3(p)$ are constructed by replacing $2^p$ by $3^p$ in eqn [3]

| *Period* $p$ | *Two branches* $N_2(p)$ | *Three branches* $N_3(p)$ | *Period* $p$ | *Two branches* $N_2(p)$ | *Three branches* $N_3(p)$ |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 11 | 186 | 16 104 |
| 2 | 1 | 3 | 12 | 335 | 44 220 |
| 3 | 2 | 8 | 13 | 630 | 122 640 |
| 4 | 3 | 18 | 14 | 1 161 | 341 484 |
| 5 | 6 | 48 | 15 | 2 182 | 956 576 |
| 6 | 9 | 116 | 16 | 4 080 | 2 690 010 |
| 7 | 18 | 312 | 17 | 7 710 | 7 596 480 |
| 8 | 30 | 810 | 18 | 14 532 | 21 522 228 |
| 9 | 56 | 2184 | 19 | 27 954 | 61 171 656 |
| 10 | 99 | 5880 | 20 | 52 377 | 174 336 264 |

**Table 3** Linking numbers of orbits to period 5 in the Smale horseshoe branched manifold with zero global torsion

| | | 0 | 1 | $2_1$ | $3_1$ | $3_1$ | $4_1$ | $4_2$ | $4_2$ | $5_1$ | $5_1$ | $5_2$ | $5_2$ | $5_3$ | $5_3$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 1 | 0 | 0 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 |
| $2_1$ | 01 | 0 | 1 | 1 | 2 | 2 | 3 | 2 | 2 | 4 | 4 | 3 | 3 | 2 | 2 |
| $3_1$ | 011 | 0 | 1 | 2 | 2 | 3 | 4 | 3 | 3 | 5 | 5 | 5 | 5 | 3 | 3 |
| $3_1$ | 001 | 0 | 1 | 2 | 3 | 2 | 4 | 3 | 3 | 5 | 5 | 4 | 4 | 3 | 3 |
| $4_1$ | 0111 | 0 | 2 | 3 | 4 | 4 | 5 | 4 | 4 | 8 | 8 | 7 | 7 | 4 | 4 |
| $4_2$ | 0011 | 0 | 1 | 2 | 3 | 3 | 4 | 3 | 4 | 5 | 5 | 5 | 5 | 4 | 4 |
| $4_2$ | 0001 | 0 | 1 | 2 | 3 | 3 | 4 | 4 | 3 | 5 | 5 | 5 | 5 | 4 | 4 |
| $5_1$ | 01111 | 0 | 2 | 4 | 5 | 5 | 8 | 5 | 5 | 8 | 10 | 9 | 9 | 5 | 5 |
| $5_1$ | 01101 | 0 | 2 | 4 | 5 | 5 | 8 | 5 | 5 | 10 | 8 | 8 | 8 | 5 | 5 |
| $5_2$ | 00111 | 0 | 2 | 3 | 5 | 4 | 7 | 5 | 5 | 9 | 8 | 6 | 7 | 5 | 5 |
| $5_2$ | 00101 | 0 | 2 | 3 | 5 | 4 | 7 | 5 | 5 | 9 | 8 | 7 | 6 | 5 | 5 |
| $5_3$ | 00011 | 0 | 1 | 2 | 3 | 3 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 4 | 5 |
| $5_3$ | 00001 | 0 | 1 | 2 | 3 | 3 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 5 | 4 |

Tables of linking numbers have been used successfully to identify mechanisms that nature uses to generate chaotic data. This analysis procedure is called topological analysis. Segments of data are identified that closely approximate unstable periodic orbits existing in the strange attractor. These data segments are then embedded in $R^3$. Each orbit is given a trial identification (symbol sequence). Their pairwise linking numbers are computed either by counting signed crossings or using the time-parametrized data segments and estimating the integers numerically using the Gauss linking integral

$$\mathrm{Link}(A, B) = \frac{1}{4\pi} \oint \oint \frac{\boldsymbol{r}_A(t_1) - \boldsymbol{r}_B(t_2)}{|\boldsymbol{r}_A(t_1) - \boldsymbol{r}_B(t_2)|^3}\, \mathrm{d}\boldsymbol{r}_A(t_1) \times \mathrm{d}\boldsymbol{r}_B(t_2)$$

This table of experimental integers is compared with the table of linking numbers for orbits with the same symbolic name on a trial branched manifold. This procedure serves to identify the branched manifold and refine the symbolic identifications of the experimental orbits, if necessary. The procedure is vastly overdetermined. For example, the linking numbers of only three low-period orbits serve to identify the four pieces of information required to specify a branched manifold with two branches. Since six or more surrogate periodic orbits can typically be extracted from experimental data, providing $\binom{6}{2} = 15$ or more linking numbers, this topological analysis procedure has built-in self-consistency checks, unlike analysis procedures based on geometric and dynamical tools.

## Basis Sets of Orbits

A branched manifold determines the topological organization of all the periodic orbits that it supports. Whenever a low-dimensional strange attractor is subjected to topological analysis, it is always the case that fewer periodic orbits are present and identified than are allowed by the branched manifold that classifies it. This is the case for strange attractors generated by experimental data as well as strange attractors generated by ODEs. The full spectrum occurs only in the hyperbolic limit, which has never been seen.

The orbits that are present are organized exactly as in the hyperbolic limit – that is, as determined by the underlying branched manifold. As control parameters change, the strange attractor undergoes perestroikas. New orbits are created and/or old orbits are annihilated in direct or inverse period-doubling and saddle–node bifurcations. The orbits that are present are always organized as determined by the branched manifold. Orbits are not created or annihilated independently of each other. Rather, there is a partial order ("forcing order") involved in orbit creation and annihilation. This partial order is poorly understood for general branched manifolds. It is much better understood for the two-branch Smale horseshoe branched manifold.

The forcing diagram for this branched manifold is shown in Figure 8 for orbits up to period 8. It is typically the case that the existence of one orbit in a strange attractor forces the presence of a spectrum of additional orbits. Forcing is transitive, so if orbit $A$ forces orbit $B (A \Rightarrow B)$ and $B$ forces $C$, then $A$ forces $C$: if $A \Rightarrow B$ and $B \Rightarrow C$ then $A \Rightarrow C$. For this reason, it is sufficient to show only the first-order forcing in this figure. The orbits shown are labeled by their period and the order in which they are created in a particular highly dissipative limit of the dynamics: the logistic map (U-sequence order in Figure 8). For example, $5_2$ describes the second (pair) of period-5 orbits created in the

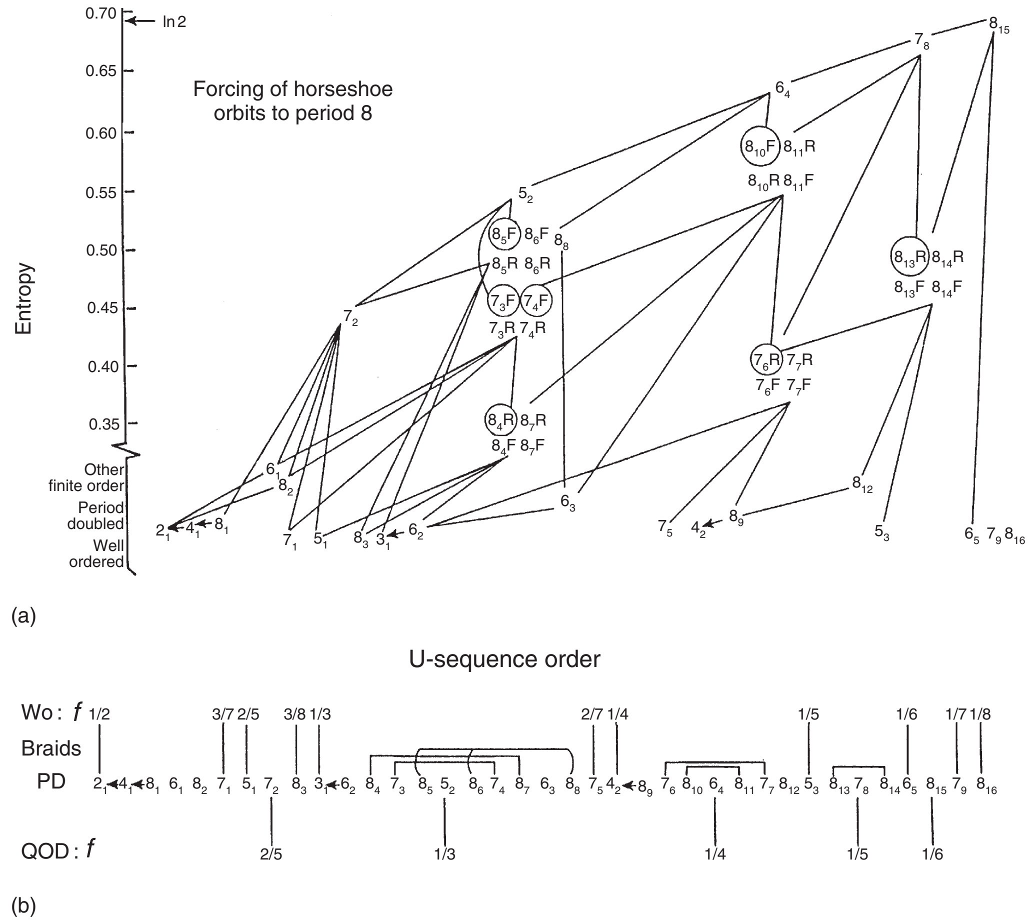



**Figure 8** (a) Forcing diagram for orbits up to period 8 in the Smale horseshoe branched manifold. (b) The sequence ("universal order") in which orbits are created in the highly dissipative limit, which is the logistic map. *The Topology of Chaos*; R Gilmore and M Lefranc; Copyright © 2002, Wiley. This material is used by permission of John Wiley & Sons, Inc.

logistic map in the transition from simple, nonchaotic behavior to fully chaotic (hyperbolic) behavior.

The orbits in the forcing diagram are organized according to their one-dimensional entropy (horizontal axis, U-sequence order) and their two-dimensional entropy (vertical axis). Nonchaotic ("laminar") behavior occurs at the lower left of this figure, where both entropies are zero. Fully chaotic behavior occurs at the upper right, where both entropies are ln 2. As control parameters change, a dynamical system that can exhibit chaos generated by a stretch-and-fold mechanism follows a path in the forcing diagram from the lower left to the upper right. Each such path is a "route to chaos." The Smale horseshoe mechanism exhibits many different routes to chaos: each follows a different path in the forcing diagram.

The state of a strange attractor at any stage in its route to chaos can be specified by a "basis set of orbits." This is a set of orbits whose presence forces the existence of all other orbits that can concurrently be found in the attractor, up to any finite period. The basis set of orbits can be constructed algorithmically. The algorithm is as follows:

1. Write down all the orbits that are present in order of increasing two-dimensional entropy from left to right.
2. For orbits with the same two-dimensional entropy, order by increasing one-dimensional entropy.
3. Remove the "highest" (rightmost) orbit from this list, together with all the orbits that it forces. This is the first basis orbit.
4. Of the orbits remaining, again remove the rightmost and all the orbits that it forces. This is the second basis orbit.
5. Continue until all orbits have been removed.

For any finite period, the above algorithm terminates because there is only a finite number of orbits. For example, if the orbit $5_2$ is present as well as all orbits with lower one-dimensional entropy, the basis set is $8_7R, 7_6, 7_4F, 8_6F, 8_8, 5_2$. As control parameters change, a strange attractor undergoes perestroikas that are quantitatively determined by changes in the basis sets of orbits.

## Bounding Tori

As experimental conditions or control parameters change, strange attractors can undergo "grosser" perestroikas than those that can be described by a change in the basis set of orbits. This occurs when new orbits are created that cannot be contained on the initial branched manifold – for example, when orbits are created that must be described by a new symbol. This is seen experimentally in the transition from horseshoe type dynamics to gateau roulé type dynamics. This involves the addition of a third branch to the branched manifold with two branches, as shown in Figures 7a and 7b. Strange attractors can undergo perestroikas described by the addition of new branches to, or deletion of old branches from, a branched manifold. These perestroikas are in a very real sense "grosser" than the perestroikas that can be described by changes in the basis sets of orbits on a fixed branched manifold.

There is a structure that provides constraints on the allowed bifurcations of branched manifolds (creation/annihilation of branches), which is analogous to the constraints that a branched manifold provides on the bifurcations and topological organization of the periodic orbits that can exist on it. This structure is called a bounding torus.

Bounding tori are constructed as follows. The semiflow on a branched manifold is "inflated" or "blown up" to a flow on a thin open set in $R^3$ containing this branched manifold. The boundary of this open set is a two-dimensional surface. Such surfaces have been classified. They are uniquely tori of genus $g$; $g=0$ (sphere), $g=1$ (tire tube), $g=2,3,\ldots$. The torus of genus $g$ has Euler characteristic $\chi=2-2g$. The flow is into this surface. The flow, restricted to the surface, exhibits a singularity wherever it is normal to the surface. At such singularities the stability is determined by the local Lyapunov exponents: $\lambda_1>0$ and $\lambda_3<0$, since the flow direction ($\lambda_2=0$) is normal to the surface. As a result, all singularities are saddles; so, by the Poincaré–Hopf theorem, the number of singularities is strongly related to the genus. The number is $2(g-1)$.

The flow, restricted to the genus-$g$ surface, can be put into canonical form and these canonical forms can be classified. The classification involves projection of the genus-$g$ torus onto a two-dimensional surface. The planar projection consists of a disk with outer boundary and $g$ interior holes. All singularities can be placed on the interior holes. The flow on the interior holes without singularities is in the same direction as the flow on the exterior boundary. Interior holes with singularities have an even number, $4,6,\ldots$. Some canonical forms are shown in Figure 9.

Poincaré sections have been used to simplify the study of flows in low-dimensional spaces by effectively reducing the dimension of the dynamics. In three dimensions, a Poincaré surface of section for a strange attractor is a minimal two-dimensional surface with the property that all points in the attractor intersect this surface transversally an infinite number of times under the flow. The Poincaré surface need not be connected and in fact is often not connected.

The Poincaré section for the flow in a genus-$g$ torus consists of the union of $g-1$ disjoint disks ($g\geq 3$) or is a single disk ($g=1$). The locations of the disks are determined algorithmically, as shown in Figure 9. The interior circles without singularities are labeled by capital letters $A,B,C,\ldots$ and those with singularities are labeled with lowercase letters $a,b,c,\ldots$ The components of the global Poincaré surface of section are numbered sequentially $1,2,\ldots,g-1$, in the order they are encountered when traversing the outer boundary in the direction of the flow, starting from any point on that boundary. Each component of the global Poincaré surface of section connects (in the projection) an interior circle without singularities to the exterior boundary. There is one component between each successive encounter of the flow with

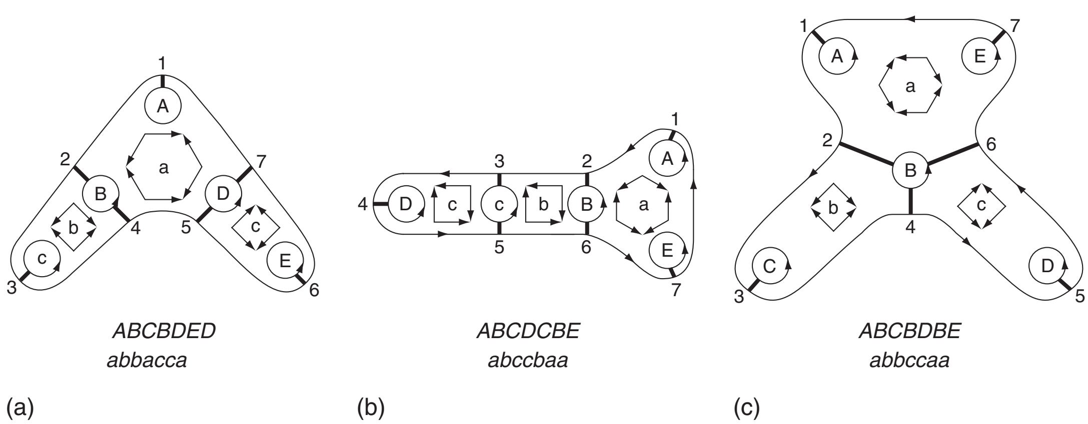



**Figure 9** Three inequivalent canonical forms of genus 8 are shown. Each is identified by a "period-7 orbit" and its dual. Reprinted figure with permission from *Physical Review E*, 69, 056206, 2004. Copyright (2004) by the American Physical Society.

holes that have singularities. Heavy lines are used to show the location of the seven components of the global Poincaré surface of section for each of the three inequivalent genus-8 canonical forms shown in Figure 9. The structure of the flow is summarized by a transition matrix. For the canonical form shown in Figure 9c the transition matrix is

$$T = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

where $T_{i,j} = 1$ if the flow can proceed directly from component $i$ to component $j$, 0 otherwise.

Bounding tori, dressed with flows, can be labeled. In fact, two dual labeling schemes are possible. Following the outer boundary in the direction of the flow, one encounters the $g - 1$ components of the global Poincaré surface of section sequentially, the interior holes without singularities at least once each, and the interior holes with singularites at least twice each. The canonical form (genus-$g$ torus dressed with a flow) on the genus-8 bounding torus shown in Figure 9a can be labeled by the sequence in which the holes without singularities are encountered (*ABCBDED*) or the order in which the holes with singularities are encountered (*abbacca*). Both sequences contain $g - 1$ symbols. These labels are unique up to cyclic permutation.

Symbol sequences for canonical forms for bounding tori act in many ways like symbol sequences for periodic orbits on branched manifolds. Although there is a 1:1 correspondence between bounded closed two-dimensional surfaces in $R^3$ and genus $g$, the number of canonical forms grows rapidly with $g$, as shown in Table 4. In fact, the number, $N(g)$, grows exponentially and can even be assigned an entropy:

**Table 4** Number of canonical bounding tori as a function of genus $g$

| $g$ | $N(g)$ | $g$ | $N(g)$ | $g$ | $N(g)$ |
|---|---|---|---|---|---|
| 3 | 1 | 9 | 15 | 15 | 2 211 |
| 4 | 1 | 10 | 28 | 16 | 5 549 |
| 5 | 2 | 11 | 67 | 17 | 14 290 |
| 6 | 2 | 12 | 145 | 18 | 36 824 |
| 7 | 5 | 13 | 368 | 19 | 96 347 |
| 8 | 6 | 14 | 870 | 20 | 252 927 |

$$\lim_{g \to \infty} \frac{\ln(N(g))}{g - 1} = \ln 3 \qquad [5]$$

In some sense, canonical forms that constrain branched manifolds within them behave like branched manifolds that constrain periodic orbits on them.

Every strange attractor that has been studied in $R^3$ has been described by a canonical bounding torus that contains it. This classification is shown in Table 5.

Branched manifold perestroikas are constrained by bounding tori as follows. Each branch line of any branched manifold can be moved into one of the $g - 1$ components of the global Poincaré surface of section. Any branched manifold contained in a genus-$g$ bounding torus ($g \geq 3$) must have at least one branch between each pair of components of the global Poincaré surface of section between which the flow is allowed, as summarized by the canonical form's transition matrix. New branches can only be added in a way that is consistent with the canonical form's transition matrix, continuity requirements, and the no intersection condition.

**Table 5** All known strange attractors of dimension $d_L < 3$ are bounded by one of the standard dressed tori. Dual labels for the bounding tori depend on $g - 1$ symbols describing holes with or without singularities

| *Strange attractor* | *Holes w/o singularites* | *Holes with singularities* | *Genus* |
|---|---|---|---|
| Rossler, Duffing, Burke, and Shaw | $A$ | | 1 |
| Various lasers, gateau roulé | $A$ | | 1 |
| Neuron with subthreshold oscillations | $A$ | | 1 |
| Shaw–van der Pol | $A$ | | 1 |
| Lorenz, Shimizu–Morioka, Rikitake | $AB$ | $aa$ | 3 |
| $\mathcal{C}_2$ covers of Rossler | $AB$ | $a^2$ | 3 |
| $\mathcal{C}_2$ cover of Lorenz[a] | $ABCD$ | $a^4$ | 5 |
| $\mathcal{C}_2$ cover of Lorenz[b] | $ABCB$ | $abba$ | 5 |
| 2 → 1 Image of figure-8 branched manifold | $ABCB$ | $ab(ab)^{-1}$ | 5 |
| Figure-8 branched manifold | $AEBECEDE$ | $a^2b^2c^2d^2$ | 9 |
| $\mathcal{C}_n$ covers of Rossler | $AB \cdots N$ | $a^n$ | $n+1$ |
| $\mathcal{C}_n$ cover of Lorenz[a] | $AB \cdots (2N)$ | $a^{2n}$ | $2n+1$ |
| $\mathcal{C}_n$ cover of Lorenz[b] | $(AZ)(BZ) \cdots (NZ)$ | $a^2b^2 \cdots n^2$ | $2n+1$ |
| Multispiral attractors | $A(B \cdots M)N(B \cdots M)^{-1}$ | $(ab \cdots m)(ab \cdots m)^{-1}$ | $2m+1$ |

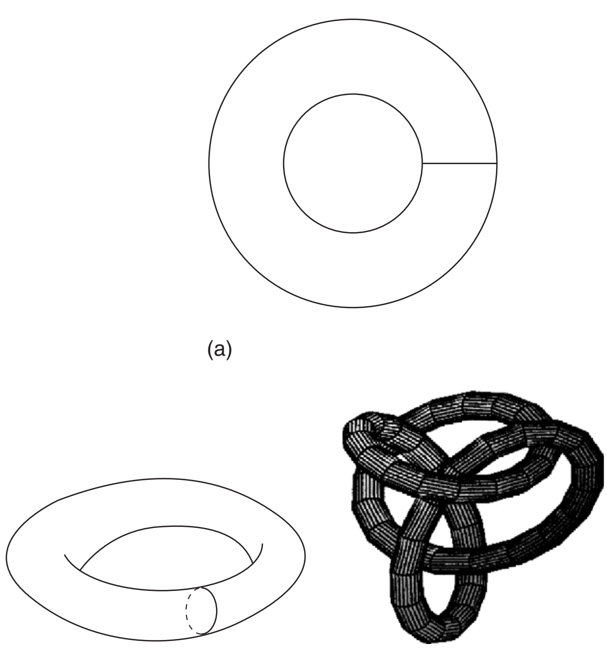
[a]Rotation axis through origin.
[b]Rotation axis through one focus.

In the simplest case, $g = 1$, a third branch can be added to a branched manifold with two branches only if its local torsion differs by $\pm 1$ from the adjacent branch. In addition, the ordering of the new branch must be consistent with the continuity and no intersection (ODE uniqueness theorem) requirements.

## Embeddings of Bounding Tori

The last level of topological structure needed for the classification of strange attractors in $R^3$ describes their embeddings in $R^3$. The classification using genus-$g$ bounding tori is intrinsic – that is, the canonical form shows how the flow looks from inside the torus. Strange attractors, and the tori that bound them, are actually embedded in $R^3$. For a complete classification, we must specify not only the canonical form but also how this form sits in $R^3$.

This program has not yet been completed, but we illustrate it with the genus-1 bounding torus in Figure 10. Figure 10a shows the canonical form, and two different embeddings of it in $R^3$. The embedding on the left is unknotted. The embedding on the right is knotted like a figure-8 knot. Extrinsic embeddings of genus-1 tori are described by tame knots in $R^3$, and tame knots can be used as "centerlines" for extrinsically embedded genus-1 tori. Higher-genus ($g \geq 3$) canonical forms – intrinsic genus-$g$ tori dressed with a canonical flow – have a larger (but discrete) variety of extrinsic embeddings in $R^3$.



(a) (b) (c)

**Figure 10** (a) Canonical form for genus-1 bounding torus. Extrinsic embeddings of the torus into $R^3$ that are (b) unknotted and (c) knotted like the figure-8 knot.

## The Embedding Question

The mechanism that nature uses to generate chaotic behavior in physical systems is not directly observable, and must be deduced by examining the data that are generated. Typically, the data consist of a single scalar time series that is discretely recorded: $x_i$, $i = 1, 2, \ldots$. In order to exhibit a strange attractor, a mapping of the data into $R^N$ must also be constructed. If the attractor is low dimensional ($d_L < 3$), one can hope that a mapping into $R^3$ can be constructed that exhibits no self-intersections or other degeneracies. Such a map is called an embedding. Once an embedding in $R^3$ is available, a topological analysis can be carried out. The analysis reveals the mechanism that underlies the creation of the embedded strange attractor.

But how do you know that the mechanism that generates the observed, embedded strange attractor has anything to do with the mechanism nature used to generate the experimental data?

If the embedding is contained in a genus-1 bounding torus, then the topological mechanism that generates the data, as defined by some unknown branched manifold $\mathcal{BM}_{\mathrm{EXP}}$, and the topological mechanism that is identified from the embedded strange attractor $\mathcal{BM}_{\mathrm{EMB}}$, are identical up to three degrees of freedom: parity, global torsion, and the knot type. As a result, in this case (genus-1) a topological analysis of embedded data does reveal nature's hidden secrets.

*See also:* Ergodic theory; Fractal dimensions in dynamics; Generic Properties of Dynamical Systems; Gravitational *N*-body Problem (Classical); Homeomorphisms and Diffeomorphisms of the Circle; Homoclinic phenomena; Inviscid Flows; Lyapunov Exponents and Strange Attractors; Nonequilibrium Statistical Mechanics (Stationary): Overview; Random Algebraic Geometry, Attractors and Flux Vacua; Random Matrix Theory in Physics; Regularization for Dynamical Zeta Functions; Singularity and Bifurcation Theory; Symmetry and Symmetry Breaking in Dynamical Systems; Synchronization of Chaos.

## Further Reading


Abraham R and Shaw CD (1992) *Dynamics: The Geometry of Behavior*, Studies in Nonlinearity, 2nd edn. Reading, MA: Addison-Wesley.

Eckmann J-P and Ruelle D (1985) Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics* 57(3): 617–656.

Gilmore R (1998) Topological analysis of chaotic dynamical systems. *Reviews of Modern Physics* 70(4): 1455–1529.

Gilmore R and Lefranc M (2002) *The Topology of Chaos, Alice in Stretch and Squeezeland*. New York: Wiley.

Gilmore R and Letellier C (2006) *The Symmetry of Chaos Alice in the Land of Mirrors*. Oxford: Oxford University Press.

Gilmore R and Pei X (2001) The topology and organization of unstable periodic orbits in Hodgkin–Huxley models of receptors with subthreshold oscillations. In: Moss F and Gielen S (eds.) *Handbook of Biological Physics, Neuro-informatics, Neural Modeling*, vol. 4, pp. 155–203. Amsterdam: North-Holland.

Ott E (1993) *Chaos in Dynamical Systems*. Cambridge: Cambridge University Press.

Solari HG, Natiello MA, and Mindlin GB (1996) *Nonlinear Physics and Its Mathematical Tools*. Bristol: IoP Publishing.

Tufillaro NB, Abbott T, and Reilly J (1992) *An Experimental Approach to Nonlinear Dynamics and Chaos*. Reading, MA: Addison-Wesley.

# Characteristic Classes

**P B Gilkey**, University of Oregon, Eugene, OR, USA
**R Ivanova**, University of Hawaii Hilo, Hilo, HI, USA
**S Nikčević**, SANU, Belgrade, Serbia and Montenegro



## Vector Bundles

Let $\mathrm{Vect}_k(M, \mathbb{F})$ be the set of isomorphism classes of real ($\mathbb{F} = \mathbb{R}$) or complex ($\mathbb{F} = \mathbb{C}$) vector bundles of rank $k$ over a smooth connected $m$-dimensional manifold $M$. Let

$$\mathrm{Vect}(M, \mathbb{F}) = \bigcup_k \mathrm{Vect}_k(M, \mathbb{F})$$

### Principal Bundles – Examples

Let $H$ be a Lie group. A fiber bundle

$$\rho : \mathcal{P} \to M$$

with fiber $H$ is said to be a principal bundle if there is a right action of $H$ on $\mathcal{P}$ which acts transitively on the fibers, that is, if $\mathcal{P}/H = M$. If $H$ is a closed subgroup of a Lie group $G$, then the natural projection $G \to G/H$ is a principal $H$ bundle over the homogeneous space $G/H$. Let $\mathrm{O}(k)$ and $\mathrm{U}(k)$ denote the orthogonal and unitary groups, respectively. Let $S^k$ denote the unit sphere in $\mathbb{R}^{k+1}$. Then we have natural principal bundles:

$$\mathrm{O}(k) \subset \mathrm{O}(k+1) \to S^k$$
$$\mathrm{U}(k) \subset \mathrm{U}(k+1) \to S^{2k+1}$$

Let $\mathbb{RP}^k$ and $\mathbb{CP}^k$ denote the real and complex projective spaces of lines through the origin in $\mathbb{R}^{k+1}$ and $\mathbb{C}^{k+1}$, respectively. Let

$$\mathbb{Z}_2 = \{\pm \mathrm{Id}\} \subset \mathrm{O}(k)$$
$$S^1 = \{\lambda \cdot \mathrm{Id} : |\lambda| = 1\} \subset \mathrm{U}(k)$$

One has $\mathbb{Z}_2$ and $S^1$ principal bundles:

$$\mathbb{Z}_2 \to S^{k-1} \to \mathbb{RP}^{k-1}$$
$$S^1 \to S^{2k-1} \to \mathbb{CP}^{k-1}$$

### Frames

A frame $\mathbf{s} := (s_1, \ldots, s_k)$ for $V \in \mathrm{Vect}_k(M, \mathbb{F})$ over an open set $\mathcal{O} \subset M$ is a collection of $k$ smooth sections to $V|_{\mathcal{O}}$ so that $\{s_1(P), \ldots, s_k(P)\}$ is a basis for the fiber $V_P$ of $V$ over any point $P \in \mathcal{O}$. Given such a frame $\mathbf{s}$, we can construct a local trivialization which identifies $\mathcal{O} \times \mathbb{F}^k$ with $V|_{\mathcal{O}}$ by the mapping

$$(P; \lambda_1, \ldots, \lambda_k) \to \lambda_1 s_1(P) + \cdots + \lambda_k s_k(P)$$

Conversely, given a local trivialization of $V$, we can take the coordinate frame

$$s_i(P) = P \times (0, \ldots, 0, 1, 0, \ldots, 0)$$

Thus, frames and local trivializations of $V$ are equivalent notions.

### Simple Covers

An open cover $\{\mathcal{O}_\alpha\}$ of $M$, where $\alpha$ ranges over some indexing set $A$, is said to be a simple cover if any finite intersection $\mathcal{O}_{\alpha_1} \cap \cdots \cap \mathcal{O}_{\alpha_k}$ is either empty or contractible.

Simple covers always exist. Put a Riemannian metric on $M$. If $M$ is compact, then there exists a uniform $\delta > 0$ so that any geodesic ball of radius $\delta$ is geodesically convex. The intersection of geodesically convex sets is either geodesically convex (and hence contractible) or empty. Thus, covering $M$ by a finite number of balls of radius $\delta$ yields a simple cover. The argument is similar even if $M$ is not compact where an infinite number of geodesic balls is used and the radii are allowed to shrink near $\infty$.

### Transition Cocycles

Let $\mathrm{Hom}(\mathbb{F}, k)$ be the set of linear transformations of $\mathbb{F}^k$ and let $\mathrm{GL}(\mathbb{F}, k) \subset \mathrm{Hom}(\mathbb{F}, k)$ be the group of all invertible linear transformations.

Let $\{\mathbf{s}_\alpha\}$ be frames for a vector bundle $V$ over some open cover $\{\mathcal{O}_\alpha\}$ of $M$. On the intersection $\mathcal{O}_\alpha \cap \mathcal{O}_\beta$, one may express $\mathbf{s}_\alpha = \psi_{\alpha\beta} \mathbf{s}_\beta$, that is

$$s_{\alpha,i}(P) = \sum_{1 \le j \le k} {\psi_{\alpha\beta,i}}^j(P) s_{\beta,j}(P)$$

Gilmore R and Letellier C (2006) *The Symmetry of Chaos Alice in the Land of Mirrors*. Oxford: Oxford University Press.
Gilmore R and Pei X (2001) The topology and organization of unstable periodic orbits in Hodgkin–Huxley models of receptors with subthreshold oscillations. In: Moss F and Gielen S (eds.) *Handbook of Biological Physics, Neuro-informatics, Neural Modeling*, vol. 4, pp. 155–203. Amsterdam: North-Holland.
Ott E (1993) *Chaos in Dynamical Systems*. Cambridge: Cambridge University Press.
Solari HG, Natiello MA, and Mindlin GB (1996) *Nonlinear Physics and Its Mathematical Tools*. Bristol: IoP Publishing.
Tufillaro NB, Abbott T, and Reilly J (1992) *An Experimental Approach to Nonlinear Dynamics and Chaos*. Reading, MA: Addison-Wesley.

# Characteristic Classes

**P B Gilkey**, University of Oregon, Eugene, OR, USA
**R Ivanova**, University of Hawaii Hilo, Hilo, HI, USA
**S Nikčević**, SANU, Belgrade, Serbia and Montenegro



## Vector Bundles

Let $\text{Vect}_k(M, \mathbb{F})$ be the set of isomorphism classes of real ($\mathbb{F} = \mathbb{R}$) or complex ($\mathbb{F} = \mathbb{C}$) vector bundles of rank $k$ over a smooth connected $m$-dimensional manifold $M$. Let

$$\text{Vect}(M, \mathbb{F}) = \bigcup_k \text{Vect}_k(M, \mathbb{F})$$

### Principal Bundles – Examples

Let $H$ be a Lie group. A fiber bundle

$$\rho : \mathcal{P} \to M$$

with fiber $H$ is said to be a principal bundle if there is a right action of $H$ on $\mathcal{P}$ which acts transitively on the fibers, that is, if $\mathcal{P}/H = M$. If $H$ is a closed subgroup of a Lie group $G$, then the natural projection $G \to G/H$ is a principal $H$ bundle over the homogeneous space $G/H$. Let $\text{O}(k)$ and $\text{U}(k)$ denote the orthogonal and unitary groups, respectively. Let $S^k$ denote the unit sphere in $\mathbb{R}^{k+1}$. Then we have natural principal bundles:

$$\text{O}(k) \subset \text{O}(k+1) \to S^k$$
$$\text{U}(k) \subset \text{U}(k+1) \to S^{2k+1}$$

Let $\mathbb{RP}^k$ and $\mathbb{CP}^k$ denote the real and complex projective spaces of lines through the origin in $\mathbb{R}^{k+1}$ and $\mathbb{C}^{k+1}$, respectively. Let

$$\mathbb{Z}_2 = \{\pm\text{Id}\} \subset \text{O}(k)$$
$$S^1 = \{\lambda \cdot \text{Id} : |\lambda| = 1\} \subset \text{U}(k)$$

One has $\mathbb{Z}_2$ and $S^1$ principal bundles:

$$\mathbb{Z}_2 \to S^{k-1} \to \mathbb{RP}^{k-1}$$
$$S^1 \to S^{2k-1} \to \mathbb{CP}^{k-1}$$

### Frames

A frame $s := (s_1, \ldots, s_k)$ for $V \in \text{Vect}_k(M, \mathbb{F})$ over an open set $\mathcal{O} \subset M$ is a collection of $k$ smooth sections to $V|_{\mathcal{O}}$ so that $\{s_1(P), \ldots, s_k(P)\}$ is a basis for the fiber $V_P$ of $V$ over any point $P \in \mathcal{O}$. Given such a frame $s$, we can construct a local trivialization which identifies $\mathcal{O} \times \mathbb{F}^k$ with $V|_{\mathcal{O}}$ by the mapping

$$(P; \lambda_1, \ldots, \lambda_k) \to \lambda_1 s_1(P) + \cdots + \lambda_k s_k(P)$$

Conversely, given a local trivialization of $V$, we can take the coordinate frame

$$s_i(P) = P \times (0, \ldots, 0, 1, 0, \ldots, 0)$$

Thus, frames and local trivializations of $V$ are equivalent notions.

### Simple Covers

An open cover $\{\mathcal{O}_\alpha\}$ of $M$, where $\alpha$ ranges over some indexing set $A$, is said to be a simple cover if any finite intersection $\mathcal{O}_{\alpha_1} \cap \cdots \cap \mathcal{O}_{\alpha_k}$ is either empty or contractible.

Simple covers always exist. Put a Riemannian metric on $M$. If $M$ is compact, then there exists a uniform $\delta > 0$ so that any geodesic ball of radius $\delta$ is geodesically convex. The intersection of geodesically convex sets is either geodesically convex (and hence contractible) or empty. Thus, covering $M$ by a finite number of balls of radius $\delta$ yields a simple cover. The argument is similar even if $M$ is not compact where an infinite number of geodesic balls is used and the radii are allowed to shrink near $\infty$.

### Transition Cocycles

Let $\text{Hom}(\mathbb{F}, k)$ be the set of linear transformations of $\mathbb{F}^k$ and let $\text{GL}(\mathbb{F}, k) \subset \text{Hom}(\mathbb{F}, k)$ be the group of all invertible linear transformations.

Let $\{s_\alpha\}$ be frames for a vector bundle $V$ over some open cover $\{\mathcal{O}_\alpha\}$ of $M$. On the intersection $\mathcal{O}_\alpha \cap \mathcal{O}_\beta$, one may express $s_\alpha = \psi_{\alpha\beta} s_\beta$, that is

$$s_{\alpha,i}(P) = \sum_{1 \le j \le k} \psi_{\alpha\beta,i}{}^j(P) s_{\beta,j}(P)$$

The maps $\psi_{\alpha\beta}: \mathcal{O}_\alpha \cap \mathcal{O}_\beta \to \mathrm{GL}(\mathbb{F}, k)$ satisfy

$$\begin{aligned} &\psi_{\alpha\alpha} = \mathrm{Id} \quad \text{on } \mathcal{O}_\alpha \\ &\psi_{\alpha\beta} = \psi_{\alpha\gamma}\psi_{\gamma\beta} \quad \text{on } \mathcal{O}_\alpha \cap \mathcal{O}_\beta \cap \mathcal{O}_\gamma \end{aligned} \qquad [1]$$

Let $G$ be a Lie group. Maps belonging to a collection $\{\psi_{\alpha\beta}\}$ of smooth maps from $\mathcal{O}_\alpha \cap \mathcal{O}_\beta$ to $G$ which satisfy eqn [1] are said to be transition cocycles with values in $G$; if $G \subset \mathrm{GL}(\mathbb{F}, k)$, they can be used to define a vector bundle by making appropriate identifications.

### Reducing the Structure Group

If $G$ is a subgroup of $\mathrm{GL}(\mathbb{F}, k)$, then $V$ is said to have a $G$-structure if we can choose frames so the transition cocycles belong to $G$; that is, we can reduce the structure group to $G$.

Denote the subgroup of orientation-preserving linear maps by

$$\mathrm{GL}^+(\mathbb{R}, k) := \{\psi \in \mathrm{GL}(\mathbb{R}, k): \det(\psi) > 0\}$$

If $V \in \mathrm{Vect}_k(M, \mathbb{R})$, then $V$ is said to be orientable if we can choose the frames so that

$$\psi_{\alpha\beta} \in \mathrm{GL}^+(\mathbb{R}, k)$$

Not every real vector bundle is orientable; the first Stiefel–Whitney class $sw_1(V) \in H^1(M; \mathbb{Z}_2)$, which is defined later, vanishes if and only if $V$ is orientable. In particular, the Möbius line bundle over the circle is not orientable.

Similarly, a real (resp. complex) bundle $V$ is said to be Riemannian (resp. Hermitian) if we can reduce the structure group to the orthogonal group $\mathrm{O}(k) \subset \mathrm{GL}(\mathbb{R}, k)$ (resp. to the unitary group $\mathrm{U}(k) \subset \mathrm{GL}(\mathbb{C}, k)$).

We can use a partition of unity to put a positive-definite symmetric (resp. Hermitian symmetric) fiber metric on $V$. Applying the Gram–Schmidt process then constructs orthonormal frames and shows that the structure group can always be reduced to $\mathrm{O}(k)$ (resp. to $\mathrm{U}(k)$); if $V$ is a real vector bundle, then the structure group can be reduced to the special orthogonal group $\mathrm{SO}(k)$ if and only if $V$ is orientable.

### Lifting the Structure Group

Let $\tau$ be a representation of a Lie group $H$ to $\mathrm{GL}(\mathbb{F}, k)$. One says that the structure group of $V$ can be lifted to $H$ if there exist frames $\{s_\alpha\}$ for $V$ and smooth maps $\phi_{\alpha\beta}: \mathcal{O}_\alpha \cap \mathcal{O}_\beta \to H$, so $\tau\phi_{\alpha\beta} = \psi_{\alpha\beta}$ where eqn [1] holds for $\phi$.

### Spin Structures

For $k \geq 3$, the fundamental group of $\mathrm{SO}(k)$ is $\mathbb{Z}_2$. Let $\mathrm{Spin}(k)$ be the universal cover of $\mathrm{SO}(k)$ and let

$$\tau : \mathrm{Spin}(k) \to \mathrm{SO}(k)$$

be the associated double cover; set $\mathrm{Spin}(2) = S^1$ and let $\tau(\lambda) = \lambda^2$. An oriented bundle $V$ is said to be spin if the transition functions can be lifted from $\mathrm{SO}(k)$ to $\mathrm{Spin}(k)$; this is possible if and only if the second Stiefel–Whitney class of $V$, which is defined later, vanishes. There can be inequivalent spin structures, which are parametrized by the cohomology group $H^1(M; \mathbb{Z}_2)$.

### The Tangent Bundle of Projective Space

The tangent bundle $T\mathbb{RP}^m$ of real projective space is orientable if and only if $m$ is odd; $T\mathbb{RP}^m$ is spin if and only if $m \equiv 3 \bmod 4$. If $m \equiv 3 \bmod 4$, there are two inequivalent spin structures on this bundle as $H^1(\mathbb{RP}^m; \mathbb{Z}_2) = \mathbb{Z}_2$.

The tangent bundle $T\mathbb{CP}^m$ of complex projective space is always orientable; $T\mathbb{CP}^m$ is spin if and only if $m$ is odd.

### Principal and Associated Bundles

Let $H$ be a Lie group and let

$$\phi_{\alpha\beta} : \mathcal{O}_\alpha \cap \mathcal{O}_\beta \to H$$

be a collection of smooth functions satisfying the compatibility conditions given in eqn [1]. We define a principal bundle $\mathcal{P}$ by gluing $\mathcal{O}_\alpha \times H$ to $\mathcal{O}_\beta \times H$ using $\phi$:

$$(P, h)_\alpha \sim (P, \phi_{\alpha\beta}(P)h)_\beta \quad \text{for } P \in \mathcal{O}_\alpha \cap \mathcal{O}_\beta$$

Because right multiplication and left multiplication commute, right multiplication gives a natural action of $H$ on $\mathcal{P}$:

$$(P, h)_\alpha \cdot \tilde{h} := (P, h \cdot \tilde{h})_\alpha$$

The natural projection $\mathcal{P} \to \mathcal{P}/H = M$ is an $H$ fiber bundle.

Let $\tau$ be a representation of $H$ to $\mathrm{GL}(\mathbb{F}, k)$. For $\xi \in \mathcal{P}, \lambda \in \mathbb{F}^k$, and $h \in H$, define a gluing

$$(\xi, \lambda) \sim (\xi \cdot h^{-1}, \tau(h)\lambda)$$

The associated vector bundle is then given by

$$\mathcal{P} \times_\tau \mathbb{F}^k := \mathcal{P} \times \mathbb{F}^k / \sim$$

Clearly, $\{\tau\phi_{\alpha\beta}\}$ are the transition cocycles of the vector bundle $\mathcal{P} \times_\tau \mathbb{F}^k$.

### Frame Bundles

If $V$ is a vector bundle, the associated principal $\mathrm{GL}(\mathbb{F}, k)$ bundle is the bundle of all frames; if $V$ is given an inner product on each fiber, then the associated principal $\mathrm{O}(k)$ or $\mathrm{U}(k)$ bundle is the bundle of orthonormal frames. If $V$ is an oriented Riemannian vector bundle, the associated principal $\mathrm{SO}(k)$ bundle is the bundle of oriented orthonormal frames.

### Direct Sum and Tensor Product

Fiber-wise direct sum (resp. tensor product) defines the direct sum (resp. tensor product) of vector bundles:

$$\begin{aligned} \oplus : &\mathrm{Vect}_k(M, \mathbb{F}) \times \mathrm{Vect}_n(M, \mathbb{F}) \\ &\to \mathrm{Vect}_{k+n}(M, \mathbb{F}) \\ \otimes : &\mathrm{Vect}_k(M, \mathbb{F}) \times \mathrm{Vect}_n(M, \mathbb{F}) \\ &\to \mathrm{Vect}_{kn}(M, \mathbb{F}) \end{aligned}$$

The transition cocycles of the direct sum (resp. tensor product) of two vector bundles are the direct sum (resp. tensor product) of the transition cocycles of the respective bundles.

The set of line bundles $\mathrm{Vect}_1(M, \mathbb{F})$ is a group under $\otimes$. The unit in the group is the trivial line bundle $\mathbb{1} := M \times \mathbb{F}$; the inverse of a line bundle $L$ is the dual line bundle $L^* := \mathrm{Hom}(L, \mathbb{F})$ since

$$L \otimes L^* = \mathbb{1}$$

### Pullback Bundle

Let $\rho : V \to M$ be the projection associated with $V \in \mathrm{Vect}_k(M, \mathbb{F})$. If $f$ is a smooth map from $N$ to $M$, then the pullback bundle $f^*V$ is the vector bundle over $N$ which is defined by setting

$$f^*V := \{(P, v) \in N \times V : f(P) = \rho(v)\}$$

The fiber of $f^*V$ over $P$ is the fiber of $V$ over $f(P)$.

Let $\{s_\alpha\}$ be local frames for $V$ over an open cover $\{\mathcal{O}_\alpha\}$ of $M$. For $P \in f^{-1}(\mathcal{O}_\alpha)$, define

$$\{f^* s_\alpha\}(P) := (P, s_\alpha(f(P)))$$

This gives a collection of frames for $f^*V$ over the open cover $\{f^{-1}(\mathcal{O}_\alpha)\}$ of $N$. Let

$$f^*\psi_{\alpha\beta} := \psi_{\alpha\beta} \circ f$$

be the pullback of the transition functions. Then

$$\begin{aligned} \{f^* s_\alpha\}(P) &= (P, \psi_{\alpha\beta}(f(P)) s_\beta(f(P))) \\ &= \{(f^*\psi_{\alpha\beta})(f^* s_\beta)\}(P) \end{aligned}$$

This shows that the pullback of the transition functions for $V$ are the transition functions of the pullback $f^*(V)$.

### Homotopy

Two smooth maps $f_0$ and $f_1$ from $N$ to $M$ are said to be homotopic if there exists a smooth map $F : N \times I \to M$ so that $f_0(P) = F(P, 0)$ and so that $f_1(P) = F(P, 1)$. If $f_0$ and $f_1$ are homotopic maps from $N$ to $M$, then $f_1^*V$ is isomorphic to $f_2^*V$.

Let $[N, M]$ be the set of all homotopy classes of smooth maps from $N$ to $M$. The association $V \to f^*V$ induces a natural map

$$[N, M] \times \mathrm{Vect}_k(M, \mathbb{F}) \to \mathrm{Vect}_k(N, \mathbb{F})$$

If $M$ is contractible, then the identity map is homotopic to the constant map $c$. Consequently, $V = \mathrm{Id}^*V$ is isomorphic to $c^*V = M \times \mathbb{F}^k$. Thus, any vector bundle over a contractible manifold is trivial. In particular, if $\{\mathcal{O}_\alpha\}$ is a simple cover of $M$ and if $V \in \mathrm{Vect}(M, \mathbb{F})$, then $V|_{\mathcal{O}_\alpha}$ is trivial for each $\alpha$. This shows that a simple cover is a trivializing cover for every $V \in \mathrm{Vect}(M, \mathbb{F})$.

### Stabilization

Let $\mathbb{1} \in \mathrm{Vect}_1(M, \mathbb{F})$ denote the isomorphism class of the trivial line bundle $M \times \mathbb{F}$ over an $m$-dimensional manifold $M$. The map $V \to V \oplus \mathbb{1}$ induces a stabilization map

$$s : \mathrm{Vect}_k(M, \mathbb{F}) \to \mathrm{Vect}_{k+1}(M, \mathbb{F})$$

which induces an isomorphism

$$\begin{aligned} &\mathrm{Vect}_k(M, \mathbb{R}) = \mathrm{Vect}_{k+1}(M, \mathbb{R}) \quad \text{for } k > m \\ &\mathrm{Vect}_k(M, \mathbb{C}) = \mathrm{Vect}_{k+1}(M, \mathbb{C}) \quad \text{for } 2k > m \end{aligned} \qquad [2]$$

These values of $k$ comprise the stable range.

### The *K*-Theory

The direct sum $\oplus$ and tensor product $\otimes$ make $\mathrm{Vect}(M, \mathbb{F})$ into a semiring; we denote the associated ring defined by the Grothendieck construction by $\mathrm{K}\mathbb{F}(M)$. If $V \in \mathrm{Vect}(M, \mathbb{F})$, let $[V] \in \mathrm{K}\mathbb{F}(M)$ be the corresponding element of $K$-theory; $\mathrm{K}\mathbb{F}(M)$ is generated by formal differences $[V_1] - [V_2]$; such formal differences are called virtual bundles.

The Grothendieck construction (*see* $K$-theory) introduces nontrivial relations. Let $S^m$ denote the standard sphere in $\mathbb{R}^{m+1}$. Since

$$T(S^m) \oplus \mathbb{1} = (m+1)\mathbb{1}$$

we can easily see that $[TS^m] = m[\mathbb{1}]$ in $\mathrm{K}\mathbb{R}(S^m)$, despite the fact that $T(S^m)$ is not isomorphic to $m\mathbb{1}$ for $m \neq 1, 3, 7$.

Let $L$ denote the nontrivial real line bundle over $\mathbb{RP}^k$. Then $T\mathbb{RP}^k \oplus \mathbb{1} = (k+1)L$, so

$$[T\mathbb{RP}^k] = (k+1)[L] - [\mathbb{1}]$$

The map $V \to \text{Rank}(V)$ extends to a surjective map from $\text{K}\mathbb{F}(M)$ to $\mathbb{Z}$. We denote the associated ideal of virtual bundles of virtual rank 0 by

$$\widetilde{\text{K}\mathbb{F}}(M) := \ker(\text{Rank})$$

In the stable range, $V \to [V] - k[\mathbb{1}]$ identifies

$$\begin{aligned} \text{Vect}_k(M,\mathbb{R}) &= \widetilde{\text{K}\mathbb{R}}(M) \quad \text{if } k > m \\ \text{Vect}_k(M,\mathbb{C}) &= \widetilde{\text{K}\mathbb{C}}(M) \quad \text{if } 2k > m \end{aligned} \qquad [3]$$

These groups contain nontrivial torsion. Let $L$ be the nontrivial real line bundle over $\mathbb{RP}^k$. Then

$$\widetilde{\text{K}\mathbb{R}}(\mathbb{RP}^k) = \mathbb{Z}\cdot\{[L]-[\mathbb{1}]\}/2^{\nu(k)}\mathbb{Z}\{[L]-[\mathbb{1}]\}$$

where $\nu(k)$ is the Adams number.

### Classifying Spaces

Let $\text{Gr}_k(\mathbb{F},n)$ be the Grassmannian of $k$-dimensional subspaces of $\mathbb{F}^n$. By mapping a $k$-plane $\pi$ in $\mathbb{F}^n$ to the corresponding orthogonal projection on $\pi$, we can identify $\text{Gr}_k(\mathbb{F},n)$ with the set of orthogonal projections of rank $k$:

$$\{\xi \in \text{Hom}(\mathbb{F}^n)\colon \xi^2 = \xi,\ \xi^* = \xi,\ \text{tr}(\xi) = k\}$$

There is a natural associated tautological $k$-plane bundle

$$V_k(\mathbb{F},n) \in \text{Vect}_k(\text{Gr}_k(\mathbb{F},n),\mathbb{F})$$

whose fiber over a $k$-plane $\pi$ is the $k$-plane itself:

$$V_k(\mathbb{F},n) := \{(\xi,x) \in \text{Hom}(\mathbb{F}^n)\times\mathbb{F}^n : \xi x = x\}$$

Let $[M,\text{Gr}_k(\mathbb{F},n)]$ denote the set of homotopy equivalence classes of smooth maps $f$ from $M$ to $\text{Gr}_k(\mathbb{F},n)$. Since $[f_1]=[f_2]$ implies that $f_1^*V$ is isomorphic to $f_2^*V$, the association

$$f \to f^*V_k(\mathbb{F},n) \in \text{Vect}_k(M,\mathbb{F})$$

induces a map

$$[M,\text{Gr}_k(\mathbb{F},n)] \to \text{Vect}_k(M,\mathbb{F})$$

This map defines a natural equivalence of functors in the stable range:

$$\begin{aligned} [M,\text{Gr}_k(\mathbb{R},\nu+k)] &= \text{Vect}_k(M,\mathbb{R}) \quad \text{for } \nu > m \\ [M,\text{Gr}_k(\mathbb{C},\nu+k)] &= \text{Vect}_k(M,\mathbb{C}) \quad \text{for } 2\nu > m \end{aligned} \qquad [4]$$

The natural inclusion of $\mathbb{F}^n$ in $\mathbb{F}^{n+1}$ induces natural inclusions

$$\begin{aligned} \text{Gr}_k(\mathbb{F},n) &\subset \text{Gr}_k(\mathbb{F},n+1) \\ V_k(\mathbb{F},n) &\subset V_k(\mathbb{F},n+1) \end{aligned} \qquad [5]$$

Let $\text{Gr}_k(\mathbb{F},\infty)$ and $V_k(\mathbb{F},\infty)$ be the direct limit spaces under these inclusions; these are the infinite-dimensional Grassmannians and classifying bundles, respectively. The topology on these spaces is the weak or inductive topology. The Grassmannians are called classifying spaces. The isomorphisms of eqn [4] are compatible with the inclusions of eqn [5] and we have

$$[M,\text{Gr}_k(\mathbb{F},\infty)] = \text{Vect}_k(M,\mathbb{F}) \qquad [6]$$

### Spaces with Finite Covering Dimension

A metric space $X$ is said to have a covering dimension at most m if, given any open cover $\{\mathcal{U}_\alpha\}$ of $X$, there exists a refinement $\{\mathcal{O}_\beta\}$ of the cover so that any intersection of more than $m+1$ of the $\{\mathcal{O}_\beta\}$ is empty. For example, any manifold of dimension $m$ has covering dimension at most $m$. More generally, any $m$-dimensional cell complex has covering dimension at most $m$.

The isomorphisms of [2]–[4], and [6] continue to hold under the weaker assumption that $M$ is a metric space with covering dimension at most $m$.

## Characteristic Classes of Vector Bundles

### The Cohomology of $\text{Gr}_k(\mathbb{F},\infty)$

The cohomology algebras of the Grassmannians are polynomial algebras on suitably chosen generators:

$$\begin{aligned} H^*(\text{Gr}_k(\mathbb{R},\infty);\mathbb{Z}_2) &= \mathbb{Z}_2[\text{sw}_1,\ldots,\text{sw}_k] \\ H^*(\text{Gr}_k(\mathbb{C},\infty);\mathbb{Z}) &= \mathbb{Z}[c_1,\ldots,c_k] \end{aligned} \qquad [7]$$

### The Stiefel–Whitney Classes

Let $V \in \text{Vect}_k(M,\mathbb{R})$. We use eqn [6] to find $\Psi : M \to \text{Gr}_k(\mathbb{R},\infty)$ which classifies $V$; the map $\Psi$ is uniquely determined up to homotopy and, using eqn [7], one sets

$$\text{sw}_i(V) := \Psi^*\text{sw}_i \in H^i(M;\mathbb{Z}_2)$$

The total Stiefel–Whitney class is then defined by

$$\text{sw}(V) = 1 + \text{sw}_1(V) + \cdots + \text{sw}_k(V)$$

The Stiefel–Whitney class has the properties:

1. If $f : X_1 \to X_2$, then $f^*(\text{sw}(V)) = \text{sw}(f^*V)$.
2. $\text{sw}(V \oplus W) = \text{sw}(V)\text{sw}(W)$.
3. If $L$ is the Möbius bundle over $S^1$, then $sw_1(L)$ generates $H^1(S^1;\mathbb{Z}_2) = \mathbb{Z}_2$.

The cohomology algebra of real projective space is a truncated polynomial algebra:

$$H^*(\mathbb{RP}^k;\mathbb{Z}_2) = \mathbb{Z}_2[x]/x^{k+1} = 0$$

Since $T\mathbb{RP}^k \oplus 1 = (k+1)L$, one has

$$\begin{aligned}\mathrm{sw}(T\mathbb{RP}^k) &= (1+x)^{k+1} \\ &= 1 + kx + \frac{(k+1)k}{2}x^2 + \cdots \end{aligned} \qquad [8]$$

### Orientability and Spin Structures

The Stiefel–Whitney classes have real geometric meaning. For example, $\mathrm{sw}_1(V) = 0$ if and only if $V$ is orientable; if $\mathrm{sw}_1(V) = 0$, then $\mathrm{sw}_2(V) = 0$ if and only if $V$ admits a spin structure. With reference to the discussion on the tangent bundle or projective space, eqn [8] yields

$$\mathrm{sw}_1(T\mathbb{RP}^k) = \begin{cases} 0 & \text{if } k \equiv 0 \bmod 2 \\ x & \text{if } k \equiv 1 \bmod 2 \end{cases}$$

Thus, $\mathbb{RP}^k$ is orientable if and only if $k$ is odd. Furthermore,

$$\mathrm{sw}_2(T\mathbb{RP}^k) = \begin{cases} 0 & \text{if } k \equiv 3 \bmod 4 \\ x & \text{if } k \equiv 1 \bmod 4 \end{cases}$$

Thus, $T\mathbb{RP}^k$ is spin if and only if $k \equiv 3 \bmod 4$.

### Chern Classes

Let $V \in \mathrm{Vect}_k(M, \mathbb{C})$. We use eqn [6] to find $\Psi : M \to \mathrm{Gr}_k(\mathbb{C}, \infty)$ which classifies $V$; the map $\Psi$ is uniquely determined up to homotopy and, using eqn [7], one sets

$$c_i(V) := \Psi^* c_i \in H^{2i}(M; \mathbb{Z})$$

The total Chern class is then defined by

$$c(V) := 1 + c_1(V) + \cdots + c_k(V)$$

The Chern class has the properties:

1. If $f : X_1 \to X_2$, then $f^*(c(V)) = c(f^*V)$.
2. $c(V \oplus W) = c(V)c(W)$.
3. Let $L$ be the classifying line bundle over $S^2 = \mathbb{CP}^1$. Then $\int_{S^2} c_1(L) = -1$.

The cohomology algebra of complex projective space also is a truncated polynomial algebra

$$H^*(\mathbb{CP}^k; \mathbb{Z}) = \mathbb{Z}[x]/x^{k+1}$$

where $x = c_1(L)$ and $L$ is the complex classifying line bundle over $\mathbb{CP}^k = \mathrm{Gr}_1(\mathbb{C}, k+1)$. If $T_c\mathbb{CP}^k$ is the complex tangent bundle, then

$$c(T_c\mathbb{CP}^k) = (1+x)^{k+1}$$

### The Pontrjagin Classes

Let $V$ be a real vector bundle over a topological space $X$ of rank $r = 2k$ or $r = 2k+1$. The Pontrjagin classes $p_i(V) \in H^{4i}(X; \mathbb{Z})$ are characterized by the properties:

1. $p(V) = 1 + p_1(V) + \cdots + p_k(V)$.
2. If $f : X_1 \to X_2$, then $f^*(p(V)) = p(f^*V)$.
3. $p(V \oplus W) = p(V)p(W)$ mod elements of order 2.
4. $\int_{\mathbb{CP}^2} p_1(T\mathbb{CP}^2) = 3$.

We can complexify a real vector bundle $V$ to construct an associated complex vector bundle $V_{\mathbb{C}}$. We have

$$p_i(V) := (-1)^i c_{2i}(V_{\mathbb{C}})$$

Conversely, if $V$ is a complex vector bundle, we can construct an underlying real vector bundle $V_{\mathbb{R}}$ by forgetting the underlying complex structure. Modulo elements of order 2, we have

$$p(V_{\mathbb{R}}) = c(V)c(V^*)$$

Let $T\mathbb{CP}^k$ be the real tangent bundle of complex projective space. Then

$$p(T\mathbb{CP}^k) = (1 - x^2)^{k+1}$$

### Line Bundles

Tensor product makes $\mathrm{Vect}_1(M, \mathbb{F})$ into an abelian group. One has natural equivalences of functors which are group homomorphisms:

$$\begin{aligned} \mathrm{sw}_1 &: \mathrm{Vect}_1(M, \mathbb{R}) \to H^1(M; \mathbb{Z}_2) \\ c_1 &: \mathrm{Vect}_1(M, \mathbb{C}) \to H^2(M; \mathbb{Z}) \end{aligned}$$

A real line bundle $L$ is trivial if and only if it is orientable or, equivalently, if $\mathrm{sw}_1(L)$ vanishes. A complex line bundle $L$ is trivial if and only if $c_1(L) = 0$. There are nontrivial vector bundles with vanishing Stiefel–Whitney classes of rank $k > 1$. For example, $\mathrm{sw}_i(TS^k) = 0$ for $i > 0$ despite the fact that $TS^k$ is trivial if and only if $k = 1, 3, 7$.

## Curvature and Characteristic Classes

### de Rham Cohomology

We can replace the coefficient group $\mathbb{Z}$ by $\mathbb{C}$ at the cost of losing information concerning torsion. Thus, we may regard $p_i(V) \in H^{4i}(M; \mathbb{C})$ if $V$ is real or $c_i(V) \in H^{2i}(M; \mathbb{C})$ if $V$ is complex. Let $M$ be a smooth manifold. Let $C^\infty \Lambda^p M$ be the space of smooth $p$-forms and let

$$d : C^\infty \Lambda^p M \to C^\infty \Lambda^{p+1} M$$

be the exterior derivative. The de Rham cohomology groups are then defined by

$$H^p_{\mathrm{deR}}(M) := \frac{\ker(d : C^\infty\Lambda^p M \to C^\infty\Lambda^{p+1}M)}{\mathrm{im}(d : C^\infty\Lambda^{p-1}M \to C^\infty\Lambda^p M)}$$

The de Rham theorem identifies the topological cohomology groups $H^p(M;\mathbb{C})$ with the de Rham cohomology groups $H^p_{\mathrm{deR}}(M)$ which are given differential geometrically.

Given a connection on $V$, the Chern–Weyl theory enables us to compute Pontrjagin and Chern classes in de Rham cohomology in terms of curvature.

### Connections

Let $V$ be a vector bundle over $M$. A connection

$$\nabla : C^\infty(V) \to C^\infty(T^*M \otimes V)$$

on $V$ is a first-order partial differential operator which satisfies the Leibnitz rule, that is, if $s$ is a smooth section to $V$ and if $f$ is a smooth function on $M$,

$$\nabla(fs) = \mathrm{d}f \otimes s + f\nabla s$$

If $X$ is a tangent vector field, we define

$$\nabla_X s = \langle X, \nabla s\rangle$$

where $\langle\cdot,\cdot\rangle$ denotes the natural pairing between the tangent and cotangent spaces. This generalizes to the bundle setting the notion of a directional derivative and has the properties:

1. $\nabla_{fX}s = f\nabla_X s$.
2. $\nabla_X(fs) = X(f)s + f\nabla_X s$.
3. $\nabla_{X_1+X_2}s = \nabla_{X_1}s + \nabla_{X_2}s$.
4. $\nabla_X(s_1 + s_2) = \nabla_X s_1 + \nabla_X s_2$.

### The Curvature 2-Form

Let $\omega_p$ be a smooth $p$-form. Then

$$\nabla : C^\infty(\Lambda^p M \otimes V) \to C^\infty(\Lambda^{p+1}M \otimes V)$$

can be extended by defining

$$\nabla(\omega_p \otimes s) = \mathrm{d}\omega_p \otimes s + (-1)^p\omega_p \wedge \nabla s$$

In contrast to ordinary exterior differentiation, $\nabla^2$ need not vanish. We set

$$\Omega(s) := \nabla^2 s$$

This is not a second-order partial differential operator; it is a zeroth-order operator, that is,

$$\begin{aligned}\Omega(fs) &= \mathrm{dd}f \otimes s - \mathrm{d}f \wedge \nabla s + \mathrm{d}f \wedge \nabla s + f\nabla^2 s\\ &= f\Omega(s)\end{aligned}$$

The curvature operator $\Omega$ can also be computed locally. Let $(s_i)$ be a local frame. Expand

$$\nabla s_i = \sum_j \omega_i^j \otimes s_j$$

to define the connection 1-form $\omega$. One then has

$$\nabla^2 s_i = \left(\mathrm{d}\omega_i^j - \omega_i^k \wedge \omega_k^j\right) \otimes s_k$$

and so

$$\Omega_i^j = \mathrm{d}\omega_i^j - \omega_i^k \wedge \omega_k^j$$

If $\tilde{s} = \psi_i^{js} j$ is another local frame, we compute

$$\tilde{\omega} = \mathrm{d}gg^{-1} + g\omega g^{-1} \quad \text{and} \quad \tilde{\Omega} = g\Omega g^{-1}$$

Although the connection 1-form $\omega$ is not tensorial, the curvature is an invariantly defined 2-form-valued endomorphism of $V$.

### Unitary Connections

Let $(\cdot\,,\cdot)$ be a nondegenerate Hermitian inner product on $V$. We say that $\nabla$ is a unitary connection if

$$(\nabla s_1, s_2) + (s_1, \nabla s_2) = d(s_1, s_2)$$

Such connections always exist and, relative to a local orthonormal frame, the curvature is skew-symmetric, that is,

$$\Omega + \Omega^* = 0$$

Thus, $\Omega$ can be regarded as a 2-form-valued element of the Lie algebra of the structure group, $\mathrm{O}(V)$ in the real setting or $\mathrm{U}(V)$ in the complex setting.

### Projections

We can always embed $V$ in a trivial bundle $1^\nu$ of dimension $\nu$; let $\pi_V$ be the orthogonal projection on $V$. We project the flat connection to $V$ to define a natural connection on $V$. For example, if $M$ is embedded isometrically in the Euclidean space $\mathbb{R}^\nu$, this construction gives the Levi-Civita connection on the tangent bundle $TM$. The curvature of this connection is then given by

$$\Omega = \pi_V\ \mathrm{d}\pi_V\ \mathrm{d}\pi_V$$

Let $V_P$ be the fiber of $V$ over a point $P \in M$. The inclusion $i : V \subset \mathbb{R}^n$ defines the classifying map $f : P \to Gr_k(\mathbb{R}, n)$ where we set

$$f(P) = i(V_P)$$

### Chern–Weyl Theory

Let $\nabla$ be a Riemannian connection on a real vector bundle $V$ of rank $k$. We set

$$p(\Omega) := \det\left(I + \frac{1}{2\pi}\Omega\right)$$

Let $\Omega^{\mathrm{T}}$ denote the transpose matrix of differential form. Since $\Omega + \Omega^{\mathrm{T}} = 0$, the polynomials of odd degree in $\Omega$ vanish and we may expand

$$p(\Omega) = 1 + p_1(\Omega) + \cdots + p_r(\Omega)$$

where $k = 2r$ or $k = 2r + 1$ and the differential forms $p_i(\Omega) \in C^\infty\Lambda^{4i}(M)$ are forms of degree $4i$.

Changing the gauge (i.e., the local frame) replaces $\Omega$ by $g\Omega g^{-1}$ and hence $p(\Omega)$ is independent of the local frame chosen. One can show that $\mathrm{d}p_i(\Omega) = 0$; let $[p_i(\Omega)]$ denote the corresponding element of de Rham cohomology. This is independent of the particular connection chosen and $[p_i(\Omega)]$ represents $p_i(V)$ in $H^{4i}(M;\mathbb{C})$.

Similarly, let $V$ be a complex vector bundle of rank $k$ with a Hermitian connection $\nabla$. Set

$$c(\Omega) := \det\left(I + \frac{\sqrt{-1}}{2\pi}\Omega\right) = 1 + c_1(\Omega) + \cdots + c_k(\Omega)$$

Again $c_i(\Omega)$ is independent of the local gauge and $\mathrm{d}c_i(\Omega) = 0$. The de Rham cohomology class $[c_i(\Omega)]$ represents $c_i(V)$ in $H^{2i}(M;\mathbb{C})$.

### The Chern Character

The total Chern character is defined by the formal sum

$$\begin{aligned}\mathrm{ch}(\Omega) &:= \mathrm{tr}(\mathrm{e}^{\sqrt{-1}\Omega/2\pi}) \\ &= \sum_\nu \frac{(\sqrt{-1})^\nu}{(2\pi)^\nu \nu!}\mathrm{tr}(\Omega^\nu) \\ &= \mathrm{ch}_0(\Omega) + \mathrm{ch}_1(\Omega) + \cdots\end{aligned}$$

Let $\mathrm{ch}(V) = [\mathrm{ch}(\Omega)]$ denote the associated de Rham cohomology class; it is independent of the particular connection chosen. We then have the relations

$$\mathrm{ch}(V \oplus W) = \mathrm{ch}(V) + \mathrm{ch}(W)$$
$$\mathrm{ch}(V \otimes W) = \mathrm{ch}(V)\mathrm{ch}(W)$$

The Chern character extends to a ring isomorphism from $KU(M) \otimes \mathbb{Q}$ to $H^e(M;\mathbb{Q})$, which is a natural equivalence of functors; modulo torsion, $K$ theory and cohomology are the same functors.

### Other Characteristic Classes

The Chern character is defined by the exponential function. There are other characteristic classes which appear in the index theorem that are defined using other generating functions that appear in index theory. Let $\boldsymbol{x} := (x_1, \ldots)$ be a collection of indeterminates. Let $s_\nu(\boldsymbol{x})$ be the $\nu$th elementary symmetric function;

$$\prod_\nu (1 + x_\nu) = 1 + s_1(\boldsymbol{x}) + s_2(\boldsymbol{x}) + \cdots$$

For a diagonal matrix $A := \mathrm{diag}(\lambda_1, \ldots)$, denote the normalized eigenvalues by $x_j := \sqrt{-1}\lambda_j/2\pi$. Then

$$c(A) = \det\left(1 + \frac{\sqrt{-1}}{2\pi}A\right) = 1 + s_1(\boldsymbol{x}) + \cdots$$

Thus, the Chern class corresponds in a certain sense to the elementary symmetric functions.

Let $f(\boldsymbol{x})$ be a symmetric polynomial or more generally a formal power series which is symmetric. We can express $f(\boldsymbol{x}) = F(s_1(\boldsymbol{x}), \ldots)$ in terms of the elementary symmetric functions and define $f(\Omega) = F(c_1(\Omega), \ldots)$ by substitution. For example, the Chern character is defined by the generating function

$$f(\boldsymbol{x}) := \sum_{\nu=1}^{k} \mathrm{e}^{x_\nu}$$

The Todd class is defined using a different generating function:

$$\begin{aligned}\mathrm{td}(\boldsymbol{x}) &:= \prod_\nu x_\nu(1 - \mathrm{e}^{-x_\nu})^{-1} \\ &= 1 + \mathrm{td}_1(\boldsymbol{x}) + \cdots\end{aligned}$$

If $V$ is a real vector bundle, we can define some additional characteristic classes similarly. Let $\{\pm\sqrt{-1}\lambda_1, \ldots\}$ be the nonzero eigenvalues of a skew-symmetric matrix $A$. We set $x_j = -\lambda_j/2\pi$ and define the Hirzebruch polynomial $L$ and the $\hat{A}$ genus by

$$\begin{aligned}L(\boldsymbol{x}) &:= \prod_\nu \frac{x_\nu}{\tanh(x_\nu)} \\ &= 1 + L_1(\boldsymbol{x}) + L_2(\boldsymbol{x}) + \cdots \\ \hat{A}(\boldsymbol{x}) &:= \prod_\nu \frac{x_\nu}{2\sinh((1/2)x_\nu)} \\ &= 1 + \hat{A}_1(\boldsymbol{x}) + \hat{A}_2(\boldsymbol{x}) + \cdots\end{aligned}$$

The generating functions

$$\frac{x}{\tanh(x)} \quad \text{and} \quad \frac{x}{2\sinh((1/2)x)}$$

are even functions of $x$, so the ambiguity in the choice of sign in the eigenvalues plays no role. This defines characteristic classes

$$L_i(V) \in H^{4i}(M;\mathbb{C}) \quad \text{and} \quad \hat{A}_i(V) \in H^{4i}(M;\mathbb{C})$$

### Summary of Formulas

We summarize below some of the formulas in terms of characteristic classes:

1. $c_1(\Omega) = \dfrac{\sqrt{-1}\mathrm{tr}(\Omega)}{2\pi}$,
2. $c_2(\Omega) = \dfrac{1}{8\pi^2}\{\mathrm{tr}(\Omega^2) - \mathrm{tr}(\Omega)^2\}$,
3. $p_1(\Omega) = -\dfrac{1}{8\pi^2}\mathrm{tr}(\Omega^2)$,
4. $\mathrm{ch}(V) = k + \left\{c_1 + \dfrac{c_1^2 - 2c_2}{2} + \cdots\right\}(V)$,
5. $\mathrm{td}(V) = \left\{1 + \dfrac{c_1}{2} + \dfrac{(c_1^2 + c_2)}{12} + \dfrac{c_1 c_2}{24} + \cdots\right\}(V)$,
6. $\hat{A}(V) = \left\{1 - \dfrac{p_1}{24} + \dfrac{7p_1^2 - 4p_2}{5760} + \cdots\right\}(V)$,
7. $L(V) = \left\{1 + \dfrac{p_1}{3} + \dfrac{7p_2 - p_1^2}{45} + \cdots\right\}(V)$,
8. $\mathrm{td}(V \oplus W) = \mathrm{td}(V)\mathrm{td}(W)$,
9. $\hat{A}(V \oplus W) = \hat{A}(V)\hat{A}(W)$,
10. $L(V \oplus W) = L(V)L(W)$.

### The Euler Form

So far, this article has dealt with the structure groups $\mathrm{O}(k)$ in the real setting and $\mathrm{U}(k)$ in the complex setting. There is one final characteristic class which arises from the structure group $\mathrm{SO}(k)$. Suppose $k = 2n$ is even. While a real antisymmetric matrix $A$ of shape $2n \times 2n$ cannot be diagonalized, it can be put in block off 2-diagonal form with blocks,

$$\begin{pmatrix} 0 & \lambda_\nu \\ -\lambda_\nu & 0 \end{pmatrix}$$

The top Pontrjagin class $p_n(A) = x_1^2 \cdots x_n^2$ is a perfect square. The Euler class

$$e_{2n}(A) := x_1 \cdots x_n$$

is the square root of $p_n$. If $V$ is an oriented vector bundle of dimension $2n$, then

$$e_{2n}(V) \in H^{2n}(M;\mathbb{C})$$

is a well-defined characteristic class satisfying $e_{2n}(V)^2 = p_n(V)$.

If $V$ is the underlying real oriented vector bundle of a complex vector bundle $W$,

$$e_{2n}(V) = c_n(W)$$

If $M$ is an even-dimensional manifold, let $e_m(M) := e_m(TM)$. If we reverse the local orientation of $M$, then $e_m(M)$ changes sign. Consequently, $e_m(M)$ is a measure rather than an $m$-form; we can use the Riemannian measure on $M$ to regard $e_m(M)$ as a scalar. Let $R_{ijkl}$ be the components of the curvature of the Levi-Civita connection with respect to some local orthonormal frame field; we adopt the convention that $R_{1221} = 1$ on the standard sphere $S^2$ in $\mathbb{R}^3$. If $\varepsilon^{I,J} := (e^I, e^J)$ is the totally antisymmetric tensor, then

$$e_{2n} := \sum_{I,J} \frac{\varepsilon^{I,J} R_{i_1 i_2 j_2 j_1} \cdots R_{i_{m-1} i_m j_m j_{m-1}}}{(8\pi)^n n!}$$

Let $\mathcal{R} := R_{ijji}$ and $\rho_{ij} := R_{ikkj}$ be the scalar curvature and the Ricci tensor, respectively. Then

$$e_2 = \frac{1}{4\pi}\mathcal{R}$$
$$e_4 = \frac{1}{32\pi^2}(\mathcal{R}^2 - 4|\rho|^2 + |R|^2)$$

## Characteristic Classes of Principal Bundles

Let $\mathfrak{g}$ be the Lie algebra of a compact Lie group $G$. Let $\pi : \mathcal{P} \to M$ be a principal $G$ bundle over $M$. For $\xi \in \mathcal{P}$, let

$$\mathcal{V}_\xi := \ker \pi_* : T_\xi\mathcal{P} \to T_{\pi\xi}M \quad \text{and} \quad \mathcal{H}_\xi := \mathcal{V}_\xi^\perp$$

be the vertical and horizontal distributions of the projection $\pi$, respectively. We assume that the metric on $\mathcal{P}$ is chosen to be $G$-invariant and such that $\pi_* : \mathcal{H}_\xi \to T_{\pi\xi}M$ is an isometry; thus, $\pi$ is a Riemannian submersion. If $F$ is a tangent vector field on $M$, let $\mathcal{H}F$ be the corresponding vertical lift. Let $\rho_\mathcal{V}$ be orthogonal projection on the distribution $\mathcal{V}$. The curvature is defined by

$$\Omega(F_1, F_2) = \rho_\mathcal{V}[\mathcal{H}(F_1), \mathcal{H}(F_2)]$$

the horizontal distribution $\mathcal{H}$ is integrable if and only if the curvature vanishes. Since the metric is $G$-invariant, $\Omega(F_1, F_2)$ is invariant under the group action. We may use a local section $s$ to $P$ over a contractible coordinate chart $\mathcal{O}$ to split $\pi^{-1}\mathcal{O} = \mathcal{O} \times G$. This permits us to identify $\mathcal{V}$ with $TG$ and to regard $\Omega$ as a $\mathfrak{g}$-valued 2-form. If we replace the section $s$ by a section $\tilde{s}$, then $\tilde{\Omega} = g\Omega g^{-1}$ changes by the adjoint action of $G$ on $\mathfrak{g}$.

If $V$ is a real or complex vector bundle over $M$, we can put a fiber metric on $V$ to reduce the structure group to the orthogonal group $\mathrm{O}(r)$ in the real setting or the unitary group $\mathrm{U}(r)$ in the complex setting. Let $\mathcal{P}_V$ be the associated frame bundle. A Riemannian connection $\nabla$ on $V$ induces an invariant splitting of $T\mathcal{P}_V = \mathcal{V} \oplus \mathcal{H}$ and defines a natural

metric on $\mathcal{P}_V$; the curvature $\Omega$ of the connection $\nabla$ defined here agrees with the definition previously.

Let $\mathcal{Q}(G)$ be the algebra of all polynomials on $\mathfrak{g}$ which are invariant under the adjoint action. If $Q \in \mathcal{Q}(G)$, then $Q(\Omega)$ is well defined. One has $dQ(\Omega) = 0$. Furthermore, the de Rham cohomology class $Q(P) := [Q(\Omega)]$ is independent of the particular connection chosen. We have

$$\begin{aligned}\mathcal{Q}(\mathrm{U}(k)) &= \mathbb{C}[c_1, \ldots, c_k]\\ \mathcal{Q}(\mathrm{SU}(k)) &= \mathbb{C}[c_2, \ldots, c_k]\\ \mathcal{Q}(\mathrm{O}(2k)) &= \mathbb{C}[p_1, \ldots, p_k]\\ \mathcal{Q}(\mathrm{O}(2k+1)) &= \mathbb{C}[p_1, \ldots, p_k]\\ \mathcal{Q}(\mathrm{SO}(2k)) &= \mathbb{C}[p_1, \ldots, p_k, e_k]/e_k^2 = p_k\\ \mathcal{Q}(\mathrm{SO}(2k+1)) &= \mathbb{C}[p_1, \ldots, p_k]\end{aligned}$$

Thus, for this category of groups, no new characteristic classes ensue. Since the invariants are Lie-algebra theoretic in nature,

$$\mathcal{Q}(\mathrm{Spin}(k)) = \mathcal{Q}(\mathrm{SO}(k))$$

Other groups, of course, give rise to different characteristic rings of invariants.

## Acknowledgmnts

Research of P Gilkey was partially supported by the MPI (Leipzig, Germany), that of R Ivanova by the UHH Seed Money Grant, and of S Nikčević by MM 1646 (Serbia), DAAD (Germany), and Dierks von Zweck Stiftung (Esen, Germany).

*See also:* Cohomology Theories; Gerbes in Quantum Field theory; Instantons: Topological Aspects; *K*-Theory; Mathai-Quillen Formalism; Riemann Surfaces.

## Further Reading


Besse AL (1987) Einstein manifolds. *Ergebnisse der Mathematik und ihrer Grenzgebiete (3)* [*Results in Mathematics and Related Areas (3)*], p. 10. Berlin: Springer-Verlag.

Bott R and Tu LW (1982) Differential forms in algebraic topology. *Graduate Texts in Mathematics*, p. 82. New York–Berlin: Springer-Verlag.

Chern S (1944) A simple intrinsic proof of the Gauss–Bonnet formula for closed Riemannian manifolds. *Annals of Mathematics* 45: 747–752.

Chern S (1945) On the curvatura integra in a Riemannian manifold. *Annals of Mathematics* 46: 674–684.

Conner PE and Floyd EE (1964) Differentiable periodic maps. *Ergebnisse der Mathematik und ihrer Grenzgebiete, N.F.*, Band 33. New York: Academic Press; Berlin–Göttingen–Heidelberg: Springer-Verlag.

de Rham G (1950) Complexes à automorphismes et homéomorphie différentiable (French). *Ann. Inst. Fourier Grenoble* 2: 51–67.

Eguchi T, Gilkey PB, and Hanson AJ (1980) Gravitation, gauge theories and differential geometry. *Physics Reports* 66: 213–393.

Eilenberg S and Steenrod N (1952) *Foundations of Algebraic Topology*. Princeton, NJ: Princeton University Press.

Greub W, Halperin S, and Vanstone R (1972) *Connections, Curvature, and Cohomology. Vol. I: De Rham Cohomology of Manifolds and Vector Bundles*. Pure and Applied Mathematics, vol. 47. New York–London: Academic Press.

Hirzebruch F (1956) Neue topologische Methoden in der algebraischen Geometrie (German). *Ergebnisse der Mathematik und ihrer Grenzgebiete (N.F.)*, Heft 9. Berlin–Göttingen–Heidelberg: Springer-Verlag.

Husemoller D (1966) *Fibre Bundles*. New York–London–Sydney: McGraw-Hill.

Karoubi M (1978) *K-theory. An introduction*. Grundlehren der Mathematischen Wissenschaften, Band 226. Berlin–New York: Springer-Verlag.

Kobayashi S (1987) *Differential Geometry of Complex Vector Bundles*. Publications of the Mathematical Society of Japan, 15. Kanô Memorial Lectures, 5. Princeton, NJ: Princeton University Press; Tokyo: Iwanami Shoten.

Milnor JW and Stasheff JD (1974) *Characteristic Classes*. Annals of Mathematics Studies, No. 76. Princeton, NJ: Princeton University Press; Tokyo: University of Tokyo Press.

Steenrod NE (1962) *Cohomology Operations*. Lectures by NE Steenrod written and revised by DBA Epstein. Annals of Mathematics Studies, No. 50. Princeton, NJ: Princeton University Press.

Steenrod NE (1951) *The Topology of Fibre Bundles*. Princeton Mathematical Series, vol. 14. Princeton, NJ: Princeton University Press.

Stong RE (1968) *Notes on Cobordism Theory*. Mathematical Notes. Princeton, NJ: Princeton University Press; Tokyo: University of Tokyo Press.

Weyl H (1939) *The Classical Groups. Their Invariants and Representations*. Princeton, NJ: Princeton University Press.


# Chern–Simons Models: Rigorous Results

**A N Sengupta**, Louisiana State University, Baton Rouge, LA, USA



## Introduction

The relationship between topological invariants and functional integrals from quantum Chern–Simons theory discovered by Witten (1989) raised several challenges for mathematicians. Most of the tremendous amount of mathematical activity generated by Witten's discovery has been concerned primarily with issues that arise after one has accepted the functional integral as a formal object. This has left, as an important challenge, the task of giving rigorous meaning to the functional integrals themselves and to rigorously derive their relation to topological invariants. The present article will discuss efforts to put the functional integral itself on a rigorous basis.

## Chern–Simons Functional Integrals

We shall describe here the typical Chern–Simons functional integral. For the purposes of this article, we will confine ourselves to a simpler setting rather than the most general possible one. In fact, we shall work with fields over three-dimensional Euclidean space $\mathbb{R}^3$ (instead of a general 3-manifold).

The typical Chern–Simons functional integral is of the form

$$\int_{\mathcal{A}} e^{i(k/4\pi)S_{CS}(A)} W_{C_1,R_1}(A) \ldots W_{C_n,R_n}(A) DA \qquad [1]$$

Our objective in this section will be to specify what the terms in this formal integral mean. Very briefly, the integration is with respect to a formal "Lebesgue measure" on $\mathcal{A}$, an infinite-dimensional space of geometric objects $A$ called connections over $\mathbb{R}^3$ with values in the Lie algebra $LG$ of a group $G$. In the first term in the integrand, in the exponent, $k$ is a real number, and $S_{CS}(A)$ is the Chern–Simons action for the connection $A$. Each term $W_{C_i,R_i}(A)$ is a Wilson loop observable, the trace in some representation $R_i$ of the holonomy of the connection $A$ around the loop $C_i$. The entire integral, formal though it may be, provides an invariant associated with the system of loops $C_1, \ldots, C_n$.

Let $G$ be a compact Lie group; for ease of exposition, let us take $G$ to be a closed, connected subgroup of $U(n)$. Thus, each element of $G$ is an $n \times n$ complex matrix $g$ with $g^*g = I$, the identity. The Lie algebra $LG$ consists of all $n \times n$ matrices $A$ which are skew-Hermitian, that is, satisfy $A^* = -A$, and for which $e^{tA} \in G$ for all real numbers $t$. On $LG$ there is a convenient inner product given by

$$\langle A, B \rangle = \mathrm{tr}(AB^*)$$

This inner product is invariant under the conjugation action of the group $G$ on its Lie algebra $LG$.

By a connection over $\mathbb{R}^3$ we shall mean a $C^\infty$ 1-form with values in $LG$. The set of all connections is an affine (in our case, actually a linear) space $\mathcal{A}$. If $A \in \mathcal{A}$, then define

$$S_{CS}(A) = \int_{\mathbb{R}^3} \mathrm{tr}(A \wedge dA + \tfrac{2}{3} A \wedge A \wedge A) \qquad [2]$$

This is, up to constant multiple, the Chern–Simons action functional.

Let $A$ be a connection and consider a piecewise smooth path

$$C : [0, 1] \to \mathbb{R}^3$$

With this one can associate a $G$-valued path $[0,1] \to G : t \mapsto g(t) \in G$ satisfying the differential equation

$$g'(t)g(t)^{-1} = -A(C'(t))$$

subject to the initial condition $g(0) = I$, the identity. The path $t \mapsto g(t)$ describes parallel transport along $C$ by the connection $A$. If $C$ is a loop then the final value $g(1)$ is the holonomy of $A$ around $C$. If $R$ is a representation of $G$ on some finite-dimensional vector space then the trace of $R(g(1))$ is the Wilson loop observable:

$$W_{C,R}(A) = \mathrm{tr}(R(g(1))) \qquad [3]$$

Thus, we have specified the meaning of the terms appearing in the formal integral [1], where $C_1, \ldots, C_n$ of eqn [1] form a link (a family of nonintersecting, imbedded loops) in $\mathbb{R}^3$ and $R_1, \ldots, R_n$ are finite-dimensional representations of $G$. Witten showed that, at least for suitable values of $k$, integrals of this form ought to produce topological invariants, which he identified, for the link.

The integral [1] is problematic for several reasons. First, there is no reasonable and useful analog of Lebesgue measure on an infinite-dimensional space. Even if one were to regularize this measure in some simple way, one would run into the problem that the measure would not live on the space of smooth connections, and so the integrand would become meaningless.

There are several different approaches to a mathematical interpretation of [1]. The approach that is often taken in practice is to simply ignore the analytical problem and define the value of the integral [1] to be what Witten's calculations have given. One approach, used, for instance, by Bar-Natan (1995) is to expand the integrand in a series and relate each individual integral in this expansion separately to topological invariants. Discrete approximation procedures to the continuum integral have also been explored. In the abelian case, infinite-dimensional oscillatory integral techniques have been used to understand the functional integral. Fröhlich and King (1999) showed the possibility of interpreting parallel transport using ideas from stochastic differential equations. Such an approach has been used successfully in the case of two-dimensional Yang–Mills theory, where the functional integral actually corresponds to integration with respect to a measure. In this article, we focus on a method of understanding the normalized Chern–Simons functional integral in terms of infinite-dimensional distribution theory and examining some ideas for understanding Wilson loop expectation values in this setting.

## Infinite Dimensional Distributions

Let $(x^0, x^1, x^2)$ denote the usual coordinates on $\mathbb{R}^3$. Gauge symmetry, an issue which will not be examined here, may be used to simplify the problem of the Chern–Simons integral. In particular, one

need only focus on connections which vanish in the $x^2$-direction, that is, connections of the form $A = A_0 dx^0 + A_1 dx^1$. For such $A$, the triple wedge-product term in the Chern–Simons action disappears, and we are left with the quadratic expression:

$$S_{CS}(A) = \int_{\mathbb{R}^3} \mathrm{tr}(A \wedge dA) \qquad [4]$$

This is good, since the functional integral now involves a quadratic exponent and so stands a good chance of rigorous realization, just as Gaussian measure can be given rigorous meaning in infinite dimensions. However, in the Chern–Simons situation, there is no hope of actually getting a measure, not even a complex measure.

The next best thing to a measure is a distribution or "generalized function." A distribution over a space $Y$ is a continuous linear functional on a topological vector space of functions on $Y$. Thus, the objective is to realize the Chern–Simons functional integral as a continuous linear functional on some space of test functions over $\mathcal{A}$ (more precisely, on an extension of $\mathcal{A}$). Before turning to the specific case of the Chern–Simons integral, let us examine some elements of the theory of infinite-dimensional distributions, in as much as they are relevant to our needs.

Let us consider a Hilbert space $\mathcal{E}_0$, and a positive Hilbert–Schmidt operator $T$ on $\mathcal{E}_0$. For each integer $p \geq 0$, let $\mathcal{E}_p = T^p(\mathcal{E}_0)$, which is a Hilbert space with the inner product $\langle x, y\rangle_p = \langle T^{-p}x, T^{-p}y\rangle$. Then we have the chain of inclusions

$$\mathcal{E} = \bigcap_{p\geq 1} \mathcal{E}_p \subset \cdots \mathcal{E}_2 \subset \mathcal{E}_1 \subset \mathcal{E}_0 \qquad [5]$$

with each inclusion $\mathcal{E}_{p+1} \to \mathcal{E}_p$ being Hilbert–Schmidt. Let $\mathcal{E}_{-p} = \mathcal{E}'_p$ be the topological dual of $\mathcal{E}_p$, the space of continuous linear functionals on $\mathcal{E}_p$, and let $\mathcal{E}'$ be the topological dual of $\mathcal{E}$, where the latter is given the topology generated by all the norms $\|\cdot\|_p$. Then we have the inclusions

$$\mathcal{E}_0 \simeq \mathcal{E}'_0 \subset \mathcal{E}_{-1} \subset \mathcal{E}_{-2} \subset \cdots \subset \mathcal{E}' = \bigcup_{p\geq 0} \mathcal{E}_{-p} \qquad [6]$$

For each $x \in \mathcal{E}$ there is the evaluation map $\hat{x} : \mathcal{E}' \to \mathbb{R} : \phi \mapsto \phi(x)$. A very special case of a general theorem of Minlos guarantees that on the dual $\mathcal{E}'$ there is a measure $\mu$ on the sigma algebra generated by all the functions $\hat{x}$ such that each $\hat{x}$ is a Gaussian random variable of mean zero and variance $|x|_0^2$, that is,

$$\int_{\mathcal{E}'} e^{it\hat{x}}\, d\mu = e^{-t^2|x|_0^2/2}$$

for all $x \in \mathcal{E}$ and $t \in \mathbb{R}$. This measure $\mu$ is the standard Gaussian measure on $\mathcal{E}'$ for the infinite-dimensional nuclear space $\mathcal{E}$.

The inner products $\langle\cdot,\cdot\rangle_p$ give rise to a nuclear space structure on function spaces over $\mathcal{E}$. Let $\mathcal{U}$ be the algebra of functions on $\mathcal{E}'$ generated by the exponentials $e^{\lambda\hat{x}}$, with $x$ running over $\mathcal{E}$ and $\lambda$ over $\mathbb{C}$. For each $p \geq 0$, there is an inner product $\langle\langle\cdot,\cdot\rangle\rangle_p$ on $\mathcal{U}$ such that

$$\left\langle\left\langle e^{\lambda\hat{x}-\lambda^2|x|_p^2/2}, e^{\mu\hat{y}-\mu^2|y|_p^2/2}\right\rangle\right\rangle_p = e^{\lambda\bar{\mu}\langle x,y\rangle_p} \qquad [7]$$

For $p = 0$ the left-hand side coincides with the $L^2(\mu)$ inner product. Let $[\mathcal{E}]_p$ be the Hilbert space completion of $\mathcal{U}$ in the $\langle\langle\cdot,\cdot\rangle\rangle_p$ inner product. Then

$$\cdots [\mathcal{E}]_3 \subset [\mathcal{E}]_2 \subset [\mathcal{E}]_1 \subset [\mathcal{E}]_0 = L^2(\mathcal{E}', \mu) \qquad [8]$$

Let $[\mathcal{E}] = \cap_{p\geq 0} [\mathcal{E}]_p$, equipped with topology from all the norms $\|\cdot\|_p$, and $[\mathcal{E}]'$ its topological dual. Elements of $[\mathcal{E}]'$, being continuous linear functionals on the "test function space" $[\mathcal{E}]$, are called distributions over $\mathcal{E}$, in the language of white-noise analysis.

A fundamental tool in the study of infinite-dimensional distributions is the $S$-transform. This generalizes the traditional Segal–Bargmann transform from the $L^2$-setting to the context of distributions. Let $\mathcal{E}_c$ be the complexification of $\mathcal{E}$. The inner product $\langle\cdot,\cdot\rangle_0$ on $\mathcal{E}$ extends to a complex-bilinear pairing $\mathcal{E}_c \times \mathcal{E}_c \to \mathbb{C} : (z, w) \mapsto z \cdot w$. The evaluation pairing $\mathcal{E}' \times \mathcal{E} \to \mathbb{R}$ also extends naturally to the complexifications. For $\Phi$ a distribution belonging to $[\mathcal{E}]'$, define a function $S\Phi$ on $\mathcal{E}$ by

$$S\Phi(z) = \Phi(c_z)$$

for all $z \in \mathcal{E}_c$. Here $c_z$ is the coherent state function on $\mathcal{E}'$ given by $c_z(\phi) = e^{\phi(z)-(1/2)z\cdot z}$. A fundamental and useful result in white-noise analysis, due originally to Potthoff and Streit, specifies the range of the transform $S$ and allows reconstruction of a distribution $\Phi$ from the function $S\Phi$. Briefly, the range of $S$ consists of functions which are holomorphic, in an appropriate sense, and have at most quadratic exponential growth. In particular, this theorem implies that a function of the form $z \mapsto e^{az\cdot z}$, for any constant $a$, is in the range of $\Phi$.

## Rigorous Realization of Chern–Simons Integrals

We return to the Chern–Simons context. As mentioned earlier, gauge symmetry may be invoked to reduce the space of connections to the smaller space:

$$\mathcal{E} = X \oplus X \qquad [9]$$

where $X = \mathcal{S}(\mathbb{R}^3) \otimes LG$ is the space of rapidly decreasing functions with values in the Lie algebra $LG$. Let

$$T_1 = \left(-\frac{d^2}{dx^2} + \frac{x^2}{4}\right)^{-1}$$

as a linear operator on $L^2(\mathbb{R}^3), T_2 = T_1^{\otimes 3} \otimes I$ the induced operator on $L^2(\mathbb{R}^3) \otimes LG$, and $T = T_2 \oplus T_2$. Then, as described in the preceding section, we have the space $\mathcal{E}$ and its dual $\mathcal{E}'$. There is then the standard Gaussian measure $\mu$ on $\mathcal{E}'$, and the space $[\mathcal{E}]'$ of distributions over $\mathcal{E}'$.

The normalized Chern–Simons integral may be viewed as a linear functional

$$\Phi_{\text{CS}} : F \mapsto \frac{1}{N} \int_{\mathcal{E}} e^{i(k/4\pi)S_{\text{CS}}(A)} F(A) D\mathcal{A} \qquad [10]$$

where $N$ is a "normalizing" factor. Rigorous meaning can be given to this by first formally working out what the $S$-transform of $\Phi_{\text{CS}}$ ought to be. Calculation shows that $S\Phi$ is indeed a holomorphic function on $\mathcal{E}_c$ of quadratic growth. The Potthoff–Streit theorem then implies that $\Phi_{\text{CS}}$ does exist as a distribution in the space $[\mathcal{E}]'$. Let us examine this in some more detail.

As before, we take $A$ to be of the form $A = A_0 dx^0 + A_1 dx^1$, with the component $A_2$ equal to 0. Integration by parts shows that

$$\frac{k}{4\pi} S_{\text{CS}}(A) = -\frac{k}{2\pi} \int_{\mathbb{R}^3} \text{tr}(A_0 \partial_2 A_1)\, \text{dvol} \qquad [11]$$

A formal computation reveals that $S(\Phi_{\text{CS}})(j)$ should be given by

$$\exp\left(\frac{2\pi i}{k} \text{tr}\left(j_0 \partial_2^{-1} j_1\right)\right) \qquad [12]$$

where $j = (j_0, j_1)$, and

$$\partial_2^{-1} f(x) = \frac{1}{2} \int ds [1_{(-\infty, x_2]}(s) - 1_{[x_2, \infty)}(s)]\, f(x^0, x^1, s)$$

The Potthoff–Streit criterion implies the existence of a distribution $\Phi_{\text{CS}}$, whose $S$-transform is given by the above expression.

The distribution $\Phi_{\text{CS}}$ is, however, not a sufficiently powerful object to allow determination of the Wilson loop expectations that one would really like to have. For instance, $\Phi_{\text{CS}}$ does not live on the space of smooth connections and so the meaning of parallel transport needs to be defined. The state of knowledge, at the rigorous level, at this point is still evolving, with progress reported by A. Hahn. We describe some ideas for the Wilson loop expectations in the following.

The strategy for defining parallel transport along a path is to smear out the path by means of bump functions and essentially replace the path by a path of test functions in $\mathcal{E}$. The description given here is mainly for the case of abelian $G$. Choose first a $C^\infty$ non-negative bump function $\psi$ on $\mathbb{R}^3$, vanishing outside the unit ball and having $L^1$ norm equal to 1. For $\epsilon > 0$, let $\psi^\epsilon$ be the scaled bump function given by $\psi^\epsilon(x) = \epsilon^{-3}\psi(x/\epsilon)$. Next, for a smooth loop $[0,1] \to l(t) = (l_0(t), l_1(t), l_2(t))$, let $l^\epsilon(t) = \psi^\epsilon(\cdot - l(t))$, the scaled bump function centered now at the path point $l(t)$. Now consider a generalized connection $A = (A_0, A_1) \in \mathcal{E}'$. Set

$$B_A^{l^\epsilon}(t) = A_0(l^\epsilon(t)) l'(t)_0 + A_1(l^\epsilon(t)) l'(t)_1 \qquad [13]$$

The equation of parallel transport can be reformulated as a differential equation for a matrix-valued path $t \mapsto P_A^{l^\epsilon}(t)$ satisfying

$$\frac{d}{dt} P_A^{l^\epsilon}(t) + B_A^{l^\epsilon}(t) P_A^{l^\epsilon}(t) = 0 \qquad [14]$$

and the initial condition $P_A^{l^\epsilon}(t) = I$. With this smearing, one can consider functions of the form

$$W_\epsilon(L; A) = \prod_{i=1}^{n} \text{tr}(P^{l_i^\epsilon}(A)) \qquad [15]$$

for a link $L$ consisting of loops $l_1, \ldots, l_n$, instead of the classical Wilson loop variable.

At this stage, it would be natural to consider taking $\epsilon \downarrow 0$ in $\Phi(W_\epsilon(L))$. However, this is still problematic. A further regularization is needed, roughly corresponding to the geometric notion of framing. In the definition of $\Phi_{\text{CS}}$, alteration is made to the quadratic form $Q(j, j)$ in the exponent which appears in the expression for $S(\Phi_{\text{CS}})$, replacing it with $Q(j, \phi_s^* j)$, where $\{\phi_s\}_{s>0}$ is a family of suitable diffeomorphisms of $\mathbb{R}^3$, with $\phi_0$ being the identity. In a sense, this splits a single loop $l$ into $l$ and a neighboring loop $\phi_s \circ l$. At the end, one has to take $s \downarrow 0$. The resulting limiting value is the expected link-invariant. We shall not go into the case of nonabelian $G$, which is more complex, for which work continues to be in progress.

Infinite-dimensional distributions can be used to formulate a rigorous theory for normalized Chern–Simons functional integrals. The more specific questions raised by the Wilson-loop integrals in this setting opens up new problems for further developments in the distribution theory, connecting geometry, topology, and infinite-dimensional analysis.

## Acknowledgments

This research is supported by US NSF grant DMS-0201683.

*See also:* BF Theories; Feynman Path Integrals; Fractional Quantum Hall Effect; Knot Theory and Physics; Large-*N* and Topological Strings; Large-*N* Dualities; Quantum 3-Manifold Invariants; Quantum Hall Effect; Spin Foams; String Field Theory; Topological Quantum Field Theory: Overview; Twistor Theory: Some Applications.

## Further Reading


Albeverio S, Hahn A, and Sengupta AN (2003) Chern–Simons theory, Hida distributions, and state models. *Infinite Dimensional Analysis Quantum Probability and Related Topics* 6: 65–81.

Albeverio S and Schäfer J (1994) Abelian Chern–Simons theory and linking numbers via oscillatory integrals. *Journal of Mathematical Physics (N.Y.)* 36 (suppl. 5): 2135–2169.

Albeverio S and Sengupta A (1997) A mathematical construction of the non-Abelian Chern–Simons functional integral. *Communications in Mathematical Physics* 186: 563–579.

Altschuler D and Freidel L (1997) Vassiliev Knot invariants and Chern–Simons perturbation theory to all orders. *Communications in Mathematical Physics* 187: 261–287.

Atiyah M (1990) *The Geometry and Physics of Knot Polynomials*. Cambridge: Cambridge University Press.

Bar-Natan D (1995) Perturbative Chern–Simons theory. *Journal of Knot Theory and its Ramifications* 4: 503.

Chern S-S and Simons J (1974) Characteristic forms and geometric invariants. *Annals of Mathematics* 99: 48–69.

Fröhlich J and King C (1989) The Chern–Simons theory and Knot polynomials. *Communications in Mathematical Physics* 126: 167–199.

Kondratiev Yu, Leukert P, Potthoff J, Streit L, and Westerkamp W (1996) Generalized functionals in Gaussian spaces – the characterization theorem revisited. *Journal of Functional Analysis* 141 (suppl. 2): 301–318.

Kuo H-H (1996) *White Noise Distribution Theory*. Boca Raton, FL: CRC Press.

Landsman NP, Pflaum M, and Schlichenmaier M (2001) *Quantization of Singular Symplectic Quotients*. Basel–Boston–Berlin: Birkhäuser.

Leukert P and Schäfer J (1996) A rigorous construction of Abelian Chern–Simons path integrals using White Noise analysis. *Rev. Math. Phys.* 8 (suppl. 3): 445–456.

Sen Samik, Sen Siddhartha, Sexton JC, and Adams DH (2000) Geometric discretization scheme applied to the Abelian Chern–Simons theory. *Physical Review E* 61: 3174–5185.

Simon B (1971) Distributions and their Hermite expansions. *Journal of Mathematical Physics (N.Y.)* 12: 140–148.

Witten E (1989) Quantum field theory and the Jones polynomial. *Communications in Mathematical Physics* 121: 351–399.


# Classical Groups and Homogeneous Spaces

**S Gindikin**, Rutgers University, Piscataway, NJ, USA



Classical groups are Lie groups corresponding to three classical geometries – linear, metric, and symplectic. Let us start with the complex field $\mathbb{C}$. We consider the linear space $\mathbb{C}^n$ and the group $GL(n;\mathbb{C})$ of its automorphisms – nondegenerate (invertible) linear transformations. The complex linear metric space is the space $\mathbb{C}^n$ endowed by a nondegenerate symmetric bilinear form; the orthogonal group $O(n;\mathbb{C})$ is the subgroup in $GL(n;\mathbb{C})$ of automorphisms of this structure. If, for $n = 2l$, we replace the symmetric form by a nondegenerate skew-symmetric form, we obtain the linear symplectic space and the group $Sp(l;\mathbb{C})$ of its automorphisms – the symplectic group.

A fundamental observation of nineteenth century geometry was that the transfer from the complex field to the real one, gives not only three corresponding groups for $\mathbb{R}$ but a much reacher collection of real forms of complex classical groups: unitary, pseudounitary, pseudoorthogonal, etc. (see below). Classical geometries correspond to homogeneous manifolds with classical groups of transformations. Geometers understood that this produces a very reach world of non-Euclidean geometries, including the first example of non-Euclidean geometry – hyperbolic geometry. Some classical algebraic theories through such an approach obtain a geometrical interpretation (see below the consideration of the cone of symmetric positive forms). Between classical manifolds there are Minkowski space, Grassmannians, and multidimensional analogs of the disk and the half-plane. A substantial part of this theory is a matrix geometry, which serves as a background for matrix analysis. A rich geometry on classical manifolds with many symmetries is a background for a rich multidimensional analysis with many explicit formulas. Classical geometries, starting with Minkowski geometry, have appeared in some problems of mathematical physics.

A crucial technical fact is the embedding of the classical groups in the class of semisimple Lie groups; it gives a very strong unified method to work with semisimple groups and corresponding geometries – the method of roots. Nevertheless, some special realizations and constructions for classical groups can also be very useful. A very impressive example is the twistors of Penrose, where an initial construction is the realization of points of four-dimensional Minkowski space as lines in three-dimensional complex projective space. We mention below some general facts about semisimple groups and homogeneous manifolds, but the focus will be on special possibilities for the classical groups. The class of simple Lie groups contains, besides the classical groups, only a finite number of exceptional groups which are also very interesting and are connected, in particular, with noncommutative and nonassociative geometries; they have applications to mathematical physics.

## Complex Groups and Homogeneous Manifolds

### Complex Classical Groups

The complete linear group $\mathrm{GL}(n;\mathbb{C})$ is the group of nongenerate matrices $g$ of order $n$ ($\det g \neq 0$) and the special linear group $\mathrm{SL}(n;\mathbb{C})$ is its subgroup of matrices with the determinant equal 1 (unimodular condition). The unimodular condition kills the one-dimensional center, perhaps, leaving only a finite center. We realize the direct products of several copies of complete linear groups with different dimensions, for example, $\mathrm{GL}(k;\mathbb{C}) \times \mathrm{GL}(l;\mathbb{C})$, as the groups of the blockdiagonal nondegenerate matrices. The letter S always means that we take matrices with determinant 1. So the notation $\mathrm{S}(\mathrm{L}(k;\mathbb{C}) \times \mathrm{L}(l;\mathbb{C}))$ means that we take blockdiagonal matrices with blocks of sizes $k$, $l$ and with the determinant 1.

Let $I$ be a nondegenerate symmetric matrix of order $n$; then the orthogonal group $\mathrm{O}(n;\mathbb{C})$ is the subgroup in $\mathrm{GL}(n;\mathbb{C})$ of matrices preserving the corresponding symmetric form so that

$$g^{\top} I g = I$$

These matrices can have the determinant $\pm 1$. The special orthogonal group $\mathrm{SO}(n;\mathbb{C})$ is the subgroup of orthogonal matrices with determinant 1. Different $I$'s give isomorphic orthogonal groups since they are all linearly equivalent. If we take as $I$ the unit matrix $E = E_n$, then we receive the group of orthogonal matrices in the classical sense: $g^{\top} g = E$.

If $n = 2l$ and we replace in this definition the symmetric matrix $I$ by a nondegenerate skew-symmetric matrix $J$, we obtain the symplectic group $\mathrm{Sp}(l;\mathbb{C})$. Again, different $J$'s give isomorphic groups. The typical example of $J$ is

$$J = \begin{pmatrix} 0 & E_l \\ -E_l & 0 \end{pmatrix}$$

It is convenient then to represent matrices $g$ as

$$g = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

where the blocks $A$, $B$, $C$, $D$ are matrices of order $l$. Then the symplectic condition is that $A^{\top} D - C^{\top} A = E$ and matrices $A^{\top} C$ and $D^{\top} B$ are symmetric. If $C = 0$ then $D = (A^{\top})^{-1}$ and $A^{-1} B$ is a symmetric matrix. In this way, we have in $\mathrm{Sp}(l;\mathbb{C})$ a subgroup $P$ of blocktriangular matrices of a very simple structure; it is an example of subgroups which are called parabolic.

There are two principal classes of homogeneous spaces with complex semisimple Lie groups: flag manifolds and Stein manifolds.

#### Flag Manifolds

These homogeneous spaces $F = G/P$ with semisimple (in our case with classical) groups $G$ have parabolic subgroups $P$ as the isotropy subgroups. The group $G = \mathrm{GL}(n;\mathbb{C})$ transitively acts on the flag manifolds $F(n_1, \ldots, n_r), 0 < n_1 < \cdots < n_r < n$, whose elements are $(n_1, \ldots, n_r)$-flags – sequences of embedded subspaces in $\mathbb{C}^n$ of the dimensions $(n_1, \ldots, n_r)$. The isotropy subgroup $P = P(n_1, \ldots, n_r)$ is the subgroup of blocktriangle matrices with the diagonal blocks of sizes $k_1, \ldots, k_{r+1}, k_j = (n_j - n_{j-1}), n_0 = 0, n_{r+1} = n$. The flag manifolds are compact complex manifolds. The matrices proportional to the unit matrix $E_n$ act trivially and we can consider instead of the action of $G = \mathrm{GL}(n;\,\mathbb{C})$ the transitive action of $G = \mathrm{SL}(n;\mathbb{C})$.

Let us pay particular attention to two extremal cases. The first one is the case of the maximal flag manifold when we have the sequence of all integers $(1, 2, 3, \ldots, n-1)$ – complete flags; the subgroup $P$ in this case is called Borelian. Another case is minimal flag manifolds with $r = 1$ (for them the unipotent radical of the parabolic subgroups is commutative). Then in the case of $\mathrm{SL}(n;\mathbb{C})$ the sequence has only one element $n_1 = k < n$ and we have Grassmannian manifolds $\mathrm{Gr}_{\mathbb{C}}(k;n) = F(k)$ of $k$-dimensional subspaces in $\mathbb{C}^n$. If $k = 1$ or $k = n-1$, we obtain the dual realizations of the complex projective space $\mathbb{C}\mathrm{P}^{n-1}$. We can interpret points of $\mathrm{Gr}_{\mathbb{C}}(k;n)$ also as $(k-1)$-dimensional planes in $\mathbb{C}P^{n-1}$.

We can define points of the projective space $\mathbb{C}\mathrm{P}^{n-1}$ by homogeneous coordinates – as the equivalency classes $(z \sim cz, z \in \mathbb{C}^n \setminus \{0\}, c \in \mathbb{C} \setminus 0)$. For the Grassmannians we can similarly use matrix homogeneous coordinates (Stiefel's coordinates): classes of $(k \times n)$-matrices $Z \in \mathrm{Mat}(k, n)$ of the maximal rank $k$ relative to the equivalency

$$Z \sim uZ, \quad u \in \mathrm{GL}(k;\mathbb{C})$$

The rows of a matrix $Z$ correspond to a base in subspace with the homogeneous coordinate $Z$; the left multiplication on a matrix $u$ replaces this base, but does not change the subspace. The group $\mathrm{GL}(n;\mathbb{C})$ acts by right multiplications:

$$Z \mapsto Zg$$

and this action preserves the equivalency classes. Suppose $k \leq n-k$ and the left $k$-minor of $Z$ is not zero. Such matrices give the dense coordinate chart $\mathbb{C}^{k(n-k)}$: we can pick in the equivalency classes the representatives $(E_k, z), z \in \mathrm{Mat}(k, n-k)$, and consider the matrices $z$ as (inhomogeneous) local coordinates. In the inhomogeneous coordinates the

action of the group has a matrix fractional linear form: let

$$g = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$
$$A \in \mathrm{Mat}(k), \qquad D \in \mathrm{Mat}(n-k),$$
$$B \in \mathrm{Mat}(k, n-k), \qquad C \in \mathrm{Mat}(n-k, k)$$

Then we have the transformation in inhomogeneous coordinates:

$$z \mapsto (A + zC)^{-1}(B + zD)$$

The condition $C = 0$ defines the parabolic subgroup which has affine action in inhomogeneous coordinates which is transitive in the coordinate chart. In such a way the Grassmannian is a compactification of $\mathbb{C}^{k(n-k)}$ (realized as a space of $k \times (n-k)$ matrices). If $n = 2k$, we can consider it as the compactification of the space of square matrices $z$ of order $k$ with the flat generalized conformal structure defined by translations of the isotropy cone $\{\det z = 0\}$.

There are similar constructions of flag manifolds for other classical groups. We will consider only the minimal flag manifolds. For $O(2k; \mathbb{C})$ we consider the isotropic Grassmannian $\mathrm{Gr}^I_{\mathbb{C}}(2k; \mathbb{C})$ of isotropic $k$-subspaces relative to the symmetric form $I$. We take the matrix realization of $\mathrm{Gr}_{\mathbb{C}}(k; 2k)$, using Stiefel's homogeneous coordinates, and add the matrix equation

$$ZIZ^{\top} = 0$$

which is well defined in the homogeneous coordinates (compatible with the equivalency classes) and defines isotropic subspaces relative to $I$. This matrix cone is preserved by the subgroup $O(2k; \mathbb{C}) \subset GL(2k; \mathbb{C})$ corresponding to the matrix $I$. If we take the symmetric matrix

$$I = \begin{pmatrix} 0 & E_k \\ E_k & 0 \end{pmatrix}$$

then in inhomogeneous coordinates ($z$ is a square $k$-matrix) this equation is transformed into the condition that the matrix $z$ is skew-symmetric. So, in a natural sense, the isotropic Grassmannian is the compactification of the linear space of skew-symmetric matrices $\mathrm{Alt}(k) = \mathbb{C}^N, N = k(k-1)/2$.

A similar construction makes sense for the symplectic group: if we replace the symmetric form $I$ with the skew-symmetric form $J$, we obtain the equation of the matrix cone representing the Lagrangian Grassmannian $\mathrm{Gr}^L_{\mathbb{C}}(k; 2k)$ of Lagrangian subspaces in $2k$-dimensional linear symplectic space. If we were to choose $J$ as above, then in the (inhomogeneous) coordinate chart we obtain the condition that the matrix $z$ is symmetric. Thus, we have the (dense) coordinate chart on the Lagrangian Grassmannian $\mathbb{C}^N = \mathrm{Sym}(k), N = k(k+1)/2$ – the linear space of symmetric matrices.

There is one more type of minimal flag manifolds for the orthogonal group $SO(n; \mathbb{C})$ – the quadric $Q$ in the projective space:

$$I(z) = zIz^{\top} = 0$$

where rows $z \in \mathbb{C}^n \setminus \{0\}$ represent, in homogeneous coordinates, points in $\mathbb{C}P^{n-1}$. If $I = E_n$ we have the equation $(z_1)^2 + \cdots + (z_n)^2 = 0$. This quadric is the complex compact conformal flat manifold $\mathbb{C}C^N, N = n-2$; it is the compactification of $\mathbb{C}^N$ endowed with the flat conformal structure corresponding to the quadratic isotropic cone. The parabolic group is generated by linear conformal transformations and translations. On the quadric $Q$ the conformal structure is defined by intersections of tangent spaces with $Q$. Apparently, this structure is invariant relative to the natural action of $SO(n; \mathbb{C})$.

### Classical Stein Manifolds

Such homogeneous complex manifolds $X = G/H$ have complex reductive isotropy subgroups $H$. Contrary to the flag manifolds which are compact, these manifolds are Stein ones and there are many holomorphic functions on them. The typical examples for $G = GL(n; \mathbb{C})$ are homogeneous spaces $S(k_1, \ldots, k_{r+1}), n = k_1 + \cdots + k_{r+1}$, for which the isotropy subgroups are blockdiagonal matrices with the blocks of sizes $k_1, \ldots, k_{r+1}$. Then points of the manifold can be realized as generic sets of subspaces $L_j \subset \mathbb{C}^n$, $\dim L_j = k_j, 1 \le j \le r+1$ or, what is equivalent, generic sets of $(k_j - 1)$-dimensional planes in $\mathbb{C}P^{n-1}$. Since the isotropy subgroup of such a homogeneous space is a subgroup of the parabolic subgroup $P(n_1, \ldots, n_r)$, $k_j = n_j - n_{j-1}$, we have the natural fibering $S(k_1, \ldots, k_{r+1}) \to F(n_1, \ldots, n_r)$ (it is simple to see this geometrically: the $i$th subspace of a flag in the base is the direct sum of first $i$ subspaces representing a point in the fiber). This is a convenient tool to apply complex analysis on $S$ to the compact manifold $F$ where there are no nontrivial holomorphic functions. Let us emphasize that such a connection exists only for special classes of classical Stein manifolds.

Let us pay special attention to the subclass of symmetric Stein manifolds. For such manifolds $X$, the isotropy subgroup $H$ is fixed relative to a holomorphic involutive automorphism of $G$. Complex semisimple Lie groups $G$ (including classical ones) are symmetric Stein manifolds relative to the action of their square $G \times G$ by left and right multiplications.

Classical Stein manifolds for $\mathrm{SL}(n;\mathbb{C})$ considered above are symmetric if $r=1$ and we have the manifold of pairs of subspaces of complimentary dimensions intersecting only on $\{0\}$. The simplest example is the manifold of pairs of different points of the projective line $\mathbb{C}\mathrm{P}^1$. Let us point out again that the transition to the generic pairs of points transforms the compact complex manifold without nonconstant holomorphic functions into a Stein manifold with a large collection of holomorphic functions.

Some other examples of symmetric Stein manifolds are connected with classical geometry and linear algebra. The affine hyperboloid in $\mathbb{C}^n$,

$$Q(z)=1$$

is a symmetric space for $G=\mathrm{O}(n;\mathbb{C}), H=\mathrm{O}(n-1;\mathbb{C})$. We can compare it with the projective quadric $Q(z)=0$ which is a minimal flag manifold. Let us remark that there is a duality here: it is possible to interpret points of the hyperboloid of dimension $n$ as generic hyperplane sections of the projective quadric of dimension $n-1$.

The space $X$ of complex symmetric matrices of order $n$ with determinant 1 is symmetric for the group $\mathrm{SL}(n;\mathbb{C})$ which acts by the changes of variables in the corresponding quadratic forms:

$$z\mapsto g^{\top}zg,\ g\in \mathrm{SL}(n;\mathbb{C})$$

The transitive action reflects the possibility of transforming such a form into a sum of squares. The isotropy subgroup is $\mathrm{SO}(n;\mathbb{C})$.

The Stein symmetric manifold $X=\mathrm{SO}(n;\mathbb{C})/\mathrm{S}(\mathrm{O}(k;\mathbb{C})\times \mathrm{O}(n-k;\mathbb{C}))$ is realized as the manifold of $k$-dimensional subspaces in $\mathbb{C}^n$ on which the restriction of the principal symmetric form $I$ is nondegenerate.

#### Isomorphisms in Small Dimensions

Isomorphisms of classical groups in small dimensions produce isomorphisms of some classical homogeneous manifolds. Such isomorphisms were very important in the history of geometry; below are a few examples. We will consider local isomorphisms (up to a finite center). We have $\mathrm{SL}(2;\mathbb{C})\cong \mathrm{SO}(3;\mathbb{C})$. Let us realize $\mathbb{C}^3$ as the space of symmetric matrices $z$ of order 2. Then, as we remarked above, the two-dimensional submanifold $X$ of matrices with determinant 1 is the symmetric Stein manifold for the group $\mathrm{SL}(2;\mathbb{C})$. On the other hand, we can take $\det z$ as the quadratic symmetric form $I$ in $\mathbb{C}^3$; then $X$ is the hyperboloid for this form and the action of $\mathrm{SL}(2;\mathbb{C})$ on symmetric matrices gives the orthogonal transformations relative to this form $I$.

Similarly, we can interpret the local isomorphism $\mathrm{SO}(4;\mathbb{C})\cong \mathrm{SL}(2;\mathbb{C})\times \mathrm{SL}(2;\mathbb{C})$. We realize $\mathbb{C}^4$ as the space of square matrices $z$ of order 2 with the symmetric quadratic form $I(z,z)=\det(z)$. Then left and right multiplications of $z$ on unimodular matrices ($z\mapsto uzv, u,v\in \mathrm{SL}(2;\mathbb{C})$) induce orthogonal transforms for the form $I$ and any orthogonal transform can be represented in such a form (one can see it by the calculation of dimensions).

The local isomorphism $\mathrm{SL}(4;\mathbb{C})\cong \mathrm{SO}(6;\mathbb{C})$ has a slightly more complicated nature. Let us consider the Grassmannian $\mathrm{Gr}_{\mathbb{C}}(2;4)$ of lines in the projective space $\mathbb{C}P^3$ with $2\times 4$ matrices $Z$ as matrix homogeneous coordinates. Let $p_{ij}, i<j$, be the minors of $Z$ with $i$th and $j$th columns. They are called Plücker coordinates on $\mathrm{Gr}_{\mathbb{C}}(2;4)$: the equivalency class of $Z$ is defined by the sequence of six numbers $p=(p_{ij}, 1\le i<j\le j)\ne(0,\ldots,0)$ up to a constant factor. Thus, we have an imbedding of $\mathrm{Gr}_{\mathbb{C}}(2;4)$ in the projective space $\mathbb{C}P^5$. The image will be the quadric

$$p_{12}p_{34}-p_{13}p_{24}+p_{14}p_{24}=0$$

Thus, we have the isomorphism of two flag manifolds and the action of $\mathrm{SL}(4;\mathbb{C})$ on the Grassmannian transforms in orthogonal transformations of four-dimensional quadric in $\mathbb{C}P^5$. The Plücker coordinates can be defined for any Grassmannian, but they do not produce in other cases some isomorphisms with other flag manifolds; nevertheless, they realize them as intersections of quadrics in projective spaces.

## Compact Classical Homogeneous Manifolds

Compact classical groups $\mathrm{U}(n), \mathrm{SU}(n), \mathrm{O}(n), \mathrm{SO}(n), \mathrm{Sp}(l)$ are maximal compact subgroups in the corresponding classical complex groups $\mathrm{GL}(n;\mathbb{C}), \mathrm{SL}(n;\mathbb{C}), \mathrm{O}(n;\mathbb{C}), \mathrm{SO}(n;\mathbb{C}), \mathrm{Sp}(l;\mathbb{C})$. This condition defines them up to an isomorphism. They are fixed subgroups of some antiholomorphic involutive automorphisms. The unitary groups $\mathrm{U}(n)$ and $\mathrm{SU}(n)$ are the groups of unitary matrices ($g^*g=E$,) correspondingly, of unitary matrices with determinant 1. As the compact orthogonal group we can take the intersection $\mathrm{U}(n)\cap \mathrm{O}(n;\mathbb{C})$. For the standard form $I$, it will be the group of real orthogonal matrices: $g^{\top}g=E$ (so the involution in $\mathrm{O}(n;\mathbb{C})$ is the conjugation $g\mapsto \bar{g}$). Similarly, we can take $\mathrm{Sp}(l)=\mathrm{SU}(2l)\cap \mathrm{Sp}(l;\mathbb{C})$ (then the involution is $g\mapsto -J\bar{g}J$).

Compact classical groups act on compact homogeneous Riemann manifolds. There are two mechanisms connecting compact and complex homogeneous manifolds. We observe the first possibility in the case of flag manifolds which are

compact. We considered them so far relative to the action of complex (noncompact) groups. It turns out that on the flag manifold $F = G/P$ the maximal compact subgroup $U \subset G$ continues to be transitive: so we can consider flag manifolds also as being homogeneous with compact groups. Then $F = U/C$, where $C$ is the centralizer of a torus in $U$. There is a Kähler metric on $F$, invariant relative to $U$. Thus, $G$ is the group of all automorphisms of $F$ as the complex manifold, but $U$ is the group of its automorphisms as the Kähler manifold. It defines two sides of geometry of flag manifolds: complex and Kähler. Flag manifolds are the only compact homogeneous Kähler manifolds with semisimple Lie groups (the class of all compact Kähler manifolds also contains locally flat compact manifolds – toruses). In the example considered above we have $F(n_1, \ldots, n_r) = \mathrm{SU}(n)/\mathrm{S}(\mathrm{U}(k_0) \times \cdots \mathrm{U}(k_r))$. In the language of Stiefel (homogeneous) coordinates, we fix a positive Hermitian form in $\mathbb{C}^n$ and characterize subspaces by orthonormal bases. For $r = 1$ we have Grassmannians $\mathrm{Gr}_{\mathbb{C}}(k; n)$, in particular the projective space $\mathbb{C}\mathrm{P}^{n-1}$ which we consider relative to the action of the unitary groups. Relative to this action they are Hermitian symmetric spaces. In the case of minimal flag manifolds for other groups the action of maximal compact subgroups also defines on them the structure of compact Hermitian symmetric spaces. Let us emphasize that relative to noncompact groups of biholomorphic automorphisms $G$, the minimal flag manifolds (including the Grassmannians) are not symmetric.

In the case of homogeneous Stein manifolds $X = G/H$, the picture is different: the maximal compact subgroups have no open orbits. There are totally real orbits which are the compact forms of $X$: $X_{\mathbb{R}} = G_{\mathbb{R}}/H_{\mathbb{R}}$, where $G_{\mathbb{R}}$ and $H_{\mathbb{R}}$ are compact forms of $G$ and $H$, respectively. It is the canonical embedding of compact homogeneous manifolds in their complexifications. The important special case is the embedding of compact symmetric manifolds in the Stein symmetric manifolds – their complexifications.

For compact symmetric manifolds $X = U/K$ the groups $U, K$ are compact Lie groups and elements of $K$ are fixed for an involutive automorphism $\sigma$ such that $K$ contains the connected component of the subgroup of all fixed elements of $\sigma$. This possibility to connect several symmetric manifolds with one involution is illustrated by the next example. The sphere $\mathrm{S}^{n-1} \subset \mathbb{R}^n$ is the symmetric space $\mathrm{SO}(n)/\mathrm{SO}(n-1)$; the real projective space $\mathbb{R}\mathrm{P}^{n-1}$ is $\mathrm{SO}(n)/\mathrm{O}(n-1)$. Here $\mathrm{SO}(n-1)$ is the connected component of $O(n-1)$ and $S^{n-1}$ is a double covering of $\mathbb{R}\mathrm{P}^{n-1}$. A few more examples, the real Grassmannian $\mathrm{Gr}_{\mathbb{R}}(k; n)$ of $k$-subspaces in $\mathbb{R}^n$ can be defined as $\mathrm{SO}(n)/\mathrm{S}(\mathrm{O}(k) \times \mathrm{O}(n-k))$. This representation corresponds to the characterization of subspaces by orthonormal bases. The consideration of arbitrary bases defines the action of the larger group $GL(n; \mathbb{R})$ on $\mathrm{Gr}_{\mathbb{R}}(k; n)$. Relative to this action, the real Grassmannian is not symmetric since the isotropy subgroup is parabolic and is not involutive. Such a possibility to extend the group is typical for a class of compact symmetric manifolds called symmetric $R$-spaces. They are real forms of Hermitian compact symmetric manifolds (minimal flag manifolds). Let us also mention compact symmetric spaces $\mathrm{SU}(n)/\mathrm{SO}(n)$, which is the compact form of the space of unimodular symmetric matrices and can be presented by the submanifold of unitary matrices in it. Also, all compact Lie groups $G$ are symmetric spaces relative to the action of $G \times G$.

## Noncompact Riemannian Symmetric Manifolds

This class of symmetric manifolds has the strongest connections with classical mathematics. Let us consider noncompact real semisimple Lie groups – real forms of complex semisimple Lie groups. They correspond to antiholomorphic involutions in complex groups.

Between real forms of $\mathrm{SL}(\mathbb{C}, n)$ there are real and quaternionic unimodular groups $\mathrm{SL}(\mathbb{R}, n), \mathrm{SL}(\mathbb{H}, n)$ and pseudounitary groups $\mathrm{SU}(p, q)$ of complex matrices preserving a Hermitian form $H$ of the signature $(p, q)$. The complex orthogonal group has as real forms, in particular, pseudoorthogonal groups $\mathrm{SO}(p, q)$ of real matrices preserving a quadratic form of the signature $(p, q)$.

Let $G$ be a real simple Lie group and $K$ be its maximal compact subgroup. Then $X = G/K$ is a Riemann symmetric manifold of noncompact type; $K$ is defined by an involutive automorphism of $G$. Therefore, in irreducible situation there is a correspondence between noncompact Riemann symmetric manifolds and real simple noncompact Lie groups. $K$-orbits on $X$ are parametrized by points of the orbit on $X$ of a maximal abelian subgroup $A$ – the Cartan subgroup of the symmetric space $X$. Its dimension $l$ is the important invariant of $X$ – its rank. The algebraic base for geometry of $X$ is the Iwasawa decomposition

$$G = KAN$$

where $N$ is a maximal unipotent subgroup (in a natural sense compatible with $A$). Then the parabolic subgroup $P = AN$ is transitive on $X$.

### Symmetric Cones

Let us start with $X = \mathrm{GL}(n, \mathbb{R})/\mathrm{O}(n)$. This manifold corresponds to the classical theory of quadratic forms: $X$ can be realized as the manifold $\mathrm{Sym}_+(n)$ of symmetric positive matrices $x \gg 0$ of order $n$ (corresponding to positive quadratic forms). Then the transitivity of $\mathrm{GL}(n; \mathbb{R})$ on $X$ corresponds to the possibility to transform positive forms to a sum of squares. The sufficiency of triangle matrices for such transformations corresponds to the transitivity on $X = \mathrm{Sym}_+(n)$ of the parabolic subgroup $P$ of (upper) triangle matrices with positive diagonal elements. So $A$ is the group of diagonal matrices with positive elements and the submanifold of diagonal matrices in $X$ parametrizes $K$-orbits. The general fact about $A$-parametrization in this example is the classical fact about the reduction of quadratic forms to diagonal form by orthogonal transformations.

There are complex and quaternionic versions of this picture. The symmetric manifold $X = \mathrm{GL}(n; \mathbb{C})/\mathrm{U}(n)$ is realized as the manifold $\mathrm{Herm}_+(n)$ of positive complex Hermitian matrices (forms) and similarly classical facts of linear algebra on Hermitian quadratic forms are transformed into geometrical statements on symmetric spaces. Let us emphasize that we consider here the group $\mathrm{GL}(n; \mathbb{C})$ as the real group. The same situation exists with the manifold $\mathrm{Herm}_+(\mathbb{H}, n)$ of positive quaternionic Hermitian matrices, which is the symmetric manifold for the real group $\mathrm{GL}(n; \mathbb{H})$.

These three manifolds can be included in an impressive geometrical structure. They all are convex homogeneous cones $V$ in linear spaces $\mathbb{R}^N$ which are self-dual ($V = V^*$) relative to a bilinear form $\langle \cdot, \cdot \rangle$. Let us recall that

$$V^* = \{x; \langle x, y \rangle > 0, y \in \overline{V} \setminus 0\}$$

Here $\overline{V}$ is the closure of $V$. So these three symmetric manifolds are linear homogeneous self-dual cones.

There is only one more type of classical homogeneous self-dual cones – quadratic (Lorentzian) cones

$$L_n = \{x \in \mathbb{R}^{n+1};\ x_1^2 - x_2^2 - \cdots - x_{n+1}^2 > 0, x_1 > 0\}$$

which is also called the future light cone (the condition $x_1 < 0$ defines the past light cone). The group of linear automorphisms of this cone is $\mathrm{SO}(1, n) \times \mathbb{R}^+$; the first factor is the Lorentz group.

There is also one exceptional 27-dimensional cone; it is possible to interpret this cone as the cone of positive Hermitian matrices of third order over Cayley numbers. There is a very nice structural theory of homogeneous self-dual cones; it is convenient to develop this theory in the language of Jordan algebras (Faraut and Koranyi 1994). Such cones participate as elements of explicit constructions of other classes of symmetric spaces (see below).

Following Siegel, it is possible to connect with homogeneous self-dual cones multidimensional versions of Euler integrals ($\Gamma$- and $B$-functions) (Faraut and Koranyi 1994). They have many applications, including those to integral formulas for complex symmetric domains.

### Riemann Symmetric Manifolds of Rank 1

The first example of non-Euclidean geometry is connected with the Riemann symmetric manifolds of rank 1 – hyperbolic spaces; $X = \mathrm{SO}(1, n)/\mathrm{O}(n)$ is the hyperbolic space of dimension $n$. It can be realized as the upper sheet of the two-sheeted hyperboloid:

$$x_0^2 - x_1^2 - \cdots - x_n^2 = 1, x_0 > 0$$

Pseudoorthogonal linear transformations from $\mathrm{SO}(1, n)$ preserve this surface; they play the role of hyperbolic motions. The equivalent realization is in the real ball $x_1^2 - \cdots - x_n^2 < 1$ relative to the projective transformations preserving this ball.

Another example of a Riemann symmetric manifold of rank 1 is the complex hyperbolic symmetric space $X = \mathrm{SU}(1; n)/U(n)$. It can similarly be realized either as the hyperboloid

$$|z_0|^2 - |z_1|^2 - \cdots - |z_n|^2 = 1$$

in $\mathbb{C}^{n+1}$ relative to pseudounitary linear transformations or as the complex ball $|z_1|^2 + \cdots + |z_n|^2 < 1$ relative to complex projective transformations preserving it. There are also quaternionic hyperbolic spaces which are realized as the quaternionic balls in the quaternionic projective spaces. These three series exhaust all classical Riemann symmetric manifolds of rank 1 of noncompact type. There is only one exceptional symmetric manifold of rank 1: it has the dimension 16 and can be interpreted as a two-dimensional ball for Cayley numbers.

### Classical Symmetric Domains in $\mathbb{C}^n$ (Cartan Domains)

Riemann symmetric manifolds of noncompact type which admit an invariant complex structure also have an invariant Hermitian form corresponding to the Riemann metrics. For this reason, we will call them noncompact Hermitian symmetric manifolds (we considered above the compact Hermitian symmetric manifolds). They are Stein manifolds, but as opposed to symmetric Stein manifolds, which we considered above, they are homogeneous relative to real groups. The condition for a Riemann symmetric

manifold $X = G/K$ to be Hermitian is that $K$ has an one-dimensional center. All Hermitian symmetric manifolds of noncompact type can be realized as bounded domains in $\mathbb{C}^n$ (but, of course, not all their holomorphic automorphisms extend in $\mathbb{C}^n$). In the case of classical manifolds, these domains are called Cartan's domains: Cartan gave their explicit matrix realizations.

The nature of groups of holomorphic automorphisms of symmetric domains $X = G/K \subset \mathbb{C}^N$ is explained by Cartan's duality. Each such domain (Hermitian symmetric manifold of noncompact type) admits an embedding in a Hermitian symmetric manifold of compact type $X_{\mathbb{C}}$ such that the complexification $G_{\mathbb{C}}$ of $G$ is the group of holomorphic automorphisms of $X_{\mathbb{C}}$ (correspondingly, $D$ is an open $G$-orbit on $X_{\mathbb{C}}$). Moreover, $X$ lies inside a (Zariski open) coordinate chart $\mathbb{C}^N$, which is an orbit of a parabolic subgroup.

The simplest example is the complex ball $\mathbb{C}B^n$ (complex hyperbolic space) imbedded in the complex projective space $\mathbb{C}P^n$. The affine chart $\mathbb{C}^n$ is the orbit of the parabolic subgroup of affine transformations. Let us consider more complicated examples.

Let $X_{\mathbb{C}}$ be the Grassmannian $\mathrm{Gr}_{\mathbb{C}}(k;n), q = n - k \geq p$; we will use matrix homogeneous coordinates $Z$ – $k \times n$ matrices – for the description of the symmetric domain. Then $G_{\mathbb{C}} = \mathrm{SL}(n;\mathbb{C})$. Let us take its real form $G = \mathrm{SU}(k;q), k + q = n$. We fix a Hermitian form $H$ of the signature $(k, q)$ and realize $G$ as the group of matrices preserving $H$:

$$gHg^* = H$$

Then $X = X_{k,q} = \mathrm{SU}(k,q)/\mathrm{S}(\mathrm{U}(k) \times \mathrm{U}(q))$ can be realized as the domain in the Grassmannian

$$ZHZ^* \gg 0$$

so that this Hermitian matrix of order $k$ must be positive. It is essential that this condition is invariant relative to multiplications of $Z$ on nondegenerate matrices $u$ on the left and, therefore, it is a well-defined condition in homogeneous coordinates.

Let us specify the choice of $H$:

$$H_1 = \begin{pmatrix} E_k & 0 \\ 0 & -E_q \end{pmatrix}$$

Then the corresponding domain $X_1$ is defined in inhomogeneous coordinates $Z = (E_k, z), z \in \mathrm{Mat}(k, q)$, by the condition

$$E_k - zz^* \gg 0$$

This matrix ball lies completely in the coordinate chart $\mathbb{C}^{kq}$. Its rank is equal to $\min(k, q)$. Thus, we have the realization of this Hermitian symmetric space as a bounded domain in $\mathbb{C}^N, N = kq$. In the case $k = 1$, we have the usual (scalar) complex ball. Let us remark that the edge of the boundary (Shilov's boundary) is the compact symmetric space

$$zz^* = E_k$$

with the group of automorphisms $\mathrm{S}(\mathrm{U}(k) \times \mathrm{U}(q))$ (the isotropy subgroup of $X$). For $k = q$ the edge coincides with the set of unitary matrix $\mathrm{U}(k)$.

Different forms $H$ of the signature $(k, q)$ are linearly equivalent and they correspond to different (biholomorphically equivalent) realizations of this Hermitian symmetric spaces. Let us, in the beginning, set $k = q$; the inhomogeneous matrix coordinates are square matrices of order $k$. Let us take the form

$$H_2 = \begin{pmatrix} 0 & \mathrm{i}E_k \\ -\mathrm{i}E_k & 0 \end{pmatrix}$$

Then, in inhomogeneous matrix coordinates, we have the domain $X_2$:

$$\frac{1}{\mathrm{i}}(z - z^*) \gg 0$$

(complex matrices with positive skew-Hermitian parts). This domain (but not its boundary) lies in the chart. It has the structure of the tube domain $T = \mathbb{R}^n + \mathrm{i}V, n = k^2$, corresponding to the symmetric cone of positive Hermitian matrices (we take the space of such matrices as a real form of $\mathbb{C}^n$). The group of affine transformations of the tube domain:

$$z \mapsto uzu^* + a, \quad u \in \mathrm{GL}(k;\mathbb{C}), a \in \mathrm{Herm}(k)$$

is transitive on $X_2$; it is the parabolic subgroup in $\mathrm{SU}(k, q)$.

The biholomorphic equivalency of the realizations of $X$ corresponding to different $H$ is induced by the equivalency of these forms. We have

$$H_2 = \lambda H_1 \lambda^*, \quad \lambda = \frac{\sqrt{2}}{2}\begin{pmatrix} E_k & -\mathrm{i}E_k \\ -\mathrm{i}E_k & E_k \end{pmatrix}$$

Then the transform $Z \mapsto Z\lambda$ transforms $X_2$ in $X_1$. In inhomogeneous coordinates it is the fractional linear matrix transform

$$z \mapsto \mathrm{i}(z + \mathrm{i}E_k)^{-1}(z - \mathrm{i}E_k)$$

It is the matrix version of the classical Cayley transform. Similarly, we can write down the inverse transform.

If $q \neq k$, then there is also an analog of the tube realization. Let $r = q - k > 0$ and

$$H_2 = \begin{pmatrix} 0 & \mathrm{i}E_k & 0 \\ -\mathrm{i}E_k & 0 & 0 \\ 0 & 0 & -E_r \end{pmatrix}$$

Let us represent the inhomogeneous coordinates as $z = (E_k, w, u), w \in \mathrm{Mat}(k), u \in \mathrm{Mat}(k, r)$. Then the domain $X_2$ is defined by the condition

$$\frac{1}{\mathrm{i}}(w - w^*) - uu^* \gg 0$$

This is an example of Siegel domains of the second kind (Pyatetskii-Shapiro 1969). This domain has a transitive group of affine transformations:

$$\begin{gathered}(w, u) \mapsto (w + a + 2ub^* + bb^*, u + b)\\ a \in \mathrm{Herm}(k), \quad b \in \mathrm{Mat}(k, r)\\ (w, u) \mapsto (cwc^*, cu) \quad c \in \mathrm{GL}\,(k; \mathbb{C})\end{gathered}$$

This class of symmetric domains in Grassmannians is called Cartan's domains of the first class. There are similar constructions for minimal flag domains (compact Hermitian symmetric spaces) with other groups. Let us consider the Lagrangian Grassmannian $\mathrm{Gr}^L_{\mathbb{C}}(k; 2k)$ corresponding to the form $J$ above. Here $G_{\mathbb{C}} = \mathrm{Sp}(k, \mathbb{C})$. Its real form $G = \mathrm{Sp}(k; \mathbb{R})$ can be realized as the subgroup of complex symplectic matrices preserving a Hermitian form $H$ of the signature $(k, k)$. In other words, we intersect the domains from the last example with the Lagrangian Grassmannians. We consider the coordinate chart with inhomogeneous coordinates – symmetric matrices $z \in \mathrm{Sym}(k)$. For $H_1$ we have the domain of symmetric matrices $z$ with the condition

$$E_k - z\bar{z} \gg 0, z = z^\top$$

This bounded realization is called Siegel's disk. For $H_2$ the real form is the group of real symplectic matrices and $X_2$ is the domain

$$\Im z = \frac{1}{2\mathrm{i}}(z - \bar{z}) \gg 0, \quad z = z^\top$$

of complex symmetric matrices with positive imaginary parts; it is called Siegel's half-plane. This is the third class of Cartan's domains. There are Siegel domains of second kind connecting with the cones of positive symmetric matrices; some of them are homogeneous, but they are never symmetric.

There are two more series of classical minimal flag manifolds: the isotropic Grassmannians and quadrics. They both contain the dual bounded symmetric domains (Cartan's domains of second and fourth classes correspondingly). Some of these domains in the isotropic Grassmannians admit the realizations as tubes with the cone of positive Hermitian quaternionic matrices and others as Siegel domains of the second kind corresponding to the same cones.

Symmetric domains in quadrics can be realized as tube domains with the Lorentzian (light) cones. The corresponding tubes are called the future (past) tube, depending on which light cone was taken. Let us consider this construction. The group of holomorphic automorphisms of these domains is $G = \mathrm{SO}(2; n)$ – the conformal extension of the Lorentz group. To realize this group, let us fix a real symmetric matrix $Q$ of signature $(2, n)$ and the group is the group of linear transformations preserving simultaneously the quadratic symmetric and Hermitian forms with this matrix $Q$:

$$g^\top Qg = Q, \qquad g^* Qg = Q$$

The standard realization corresponds to the diagonal matrix $Q$ with the diagonal $(1, 1, -1, \ldots, -1)$. Cartan's domains of the fourth class are connected components of the manifold

$$ZQZ^\top = 0, \qquad ZQZ^* > 0$$

where rows $Z$ are homogeneous coordinates in the projective space $\mathbb{C}\mathrm{P}^{n+1}$. In other words, we consider a domain on the quadric in the projective space (which is the complex flat conformal space $\mathbb{C}\mathrm{C}^n$). For the standard $Q$ the domain will lie in the coordinate chart; thus it is the bounded realization. For the tube realization, we take

$$Q = \begin{pmatrix} 0 & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & E_n \end{pmatrix}$$

Let $Z = (z_0, z_1, w_1, \ldots, w_n)$, $w = u + \mathrm{i}v$, $q(s, t) = s_1t_1 - s_2t_2 - \cdots - \cdots - s_nt_n$ and we consider the affine chart $\mathbb{C}^{n+1} = \{z_0 = 1\}$. We have

$$\begin{gathered}ZQZ^\top = 2z_1 + q(w, w) = 0\\ ZQZ^* = 2\Re z_1 + q(w, \bar{w}) > 0\end{gathered}$$

The first condition gives $2\Re z_1 = q(v, v) - q(u, u)$ and then the second condition gives the final description of the considered set in $\mathbb{C}^n_w$:

$$q(v, v) = v_1^2 - v_2^2 - \cdots - v_n^2 > 0, \quad w = u + \mathrm{i}v$$

as the union of the future and the past tubes ($T_\pm = \{v_1 \gtrless 0\}$). The edge $\mathbb{R}^n$ of these tubes ($v = 0$) has the structure of the Minkowski space corresponding to the form $q$. The parabolic subgroup is the affine conformal group of the Minkowski space. It includes the Poincaré group and is transitive on tubes. The complete group of holomorphic automorphisms of tubes $G = \mathrm{SO}(2, n)$ is the group of all (not only affine) conformal transformations of the Minkowski space. The complete edge of these symmetric domains in the quadric $\mathbb{C}\mathrm{C}^n$ is the conformal compactification of the Minkowski space (a compact symmetric $R$-space with the compact group $\mathrm{S}(\mathrm{O}(2) \times \mathrm{O}(n))$ on which the noncompact group $\mathrm{SO}(2, n)$ also acts).

In addition to four Cartan's classes of classical domains there are two exceptional symmetric domains in the dimensions 27 and 16 (dual to two exceptional compact Hermitian symmetric spaces of these dimensions). The first of them can be realized as the tube domain corresponding to the exceptional cone of positive Hermitian matrices with Cayley numbers of order 3 (the dimension 27) and another can be realized as a Siegel domain of the second kind associated with the eight-dimensional future tube. It is possible, using Γ-function of self-dual homogeneous cones, to write explicit Bergman and Cauchy–Szego integral formulas.

## Noncompact Symmetric *R*-Spaces

There are several other interesting noncompact symmetric manifolds. Let us mention the noncompact symmetric $R$-spaces which are real forms of complex symmetric domains. The typical example is the domain of real square matrices $x \in \mathrm{Mat}(k)$:

$$E_k - xx^\top \gg 0$$

The condition is that this symmetric matrix is positive. It is the Riemann symmetric space with the group $G = \mathrm{SO}(k,k)$. It can be embedded in the real Grassmannian $\mathrm{Gr}_{\mathbb{R}}(k;2k)$ with the matrix homogeneous coordinates $X \in \mathrm{Mat}_{\mathbb{R}}(k,2k)$ and the group $\mathrm{SL}(2k;\mathbb{R})$ acting of $X$ by right multiplications. Let

$$I_1 = \begin{pmatrix} E_k & 0 \\ 0 & -E_k \end{pmatrix}$$

and $\mathrm{SO}(k,k)$ be the subgroup of matrices preserving the quadratic form $I_1$: $gI_1g^\top = I_1$. This group will preserve the domain $XI_1X^\top \gg 0$ and, in the inhomogeneous coordinates $X = (E_k, x), x \in \mathrm{Mat}_{\mathbb{R}}(k)$, it will be exactly the same as the domain above. The group $\mathrm{SO}(k,k)$ acts by matrix fractional linear transformations. This domain is the real form on Siegel's ball. If we replace the form on

$$I_2 = \begin{pmatrix} 0 & E_k \\ E_k & 0 \end{pmatrix}$$

then we realize our symmetric manifold as the domain

$$x + x^\top \gg 0$$

So, the symmetric part of the matrix $x$ must be positive. This realization is homogeneous relative to the linear automorphisms: $x \mapsto axa^\top + b$, $a \in \mathrm{GL}(k;\mathbb{R})$, $b = -b^\top$. A similar construction exists for rectangular matrices.

## Geometry of Isomorphisms in Small Dimensions

We connected above several local isomorphisms of complex classical groups with some geometrical facts. Let us mention now several similar examples for real groups. We start from isomorphisms of symmetric cones. The cone $\mathrm{Sym}_+(2)$ of symmetric positive matrices of second order is (linearly) isomorphic to the future light cone $L(2)$. The comparison of the groups of automorphisms gives the local isomorphism

$$\mathrm{SL}(2;\mathbb{R}) \cong SO(1;2)$$

This isomorphism corresponds also to the isomorphism of two classical realizations of hyperbolic plane – of Poincaré and Klein. Let us also mention that the isomorphism $\mathrm{SL}(2,\mathbb{R}) \cong \mathrm{SU}(1,1)$ corresponds to the holomorphic equivalency of the disk and the upper half-plane. The isomorphism $\mathrm{Herm}_+(2) = L(3)$ corresponds to the presentation of any Hermitian matrix of the order 2 in Pauli's coordinates,

$$z = \begin{pmatrix} t - x_1 & x_2 + ix_3 \\ x_2 - ix_3 & t + x_1 \end{pmatrix}$$

Then, $\det z = t^2 - x_1^2 - x_2^2 - x_3^2$. To compare the groups of automorphisms, we receive

$$\mathrm{SL}(2,\mathbb{C}) \cong SO(1,3)$$

Similarly, in the quaternionic case, the isomorphism of the cones $\mathrm{Herm}_+(2,\mathbb{H})$ gives the isomorphism

$$\mathrm{SL}(2,\mathbb{H}) \cong \mathrm{SO}(1,5)$$

The linear isomorphism of cones produces the holomorphic isomorphism of corresponding tubes and their groups of holomorphic automorphisms. So each of these three isomorphisms gives automatically one more isomorphism. Let us give it for the first two cones:

$$\mathrm{Sp}(2,\mathbb{R}) \cong \mathrm{SO}(2,2), \qquad \mathrm{SU}(2,2) \cong \mathrm{SO}(2,3)$$

We just compared the descriptions of automorphisms of classical tubes from above.

Considering $\det(x)$ as the quadratic form of signature $(2,2)$ on $\mathrm{Mat}(2) \simeq \mathbb{R}^4$, we obtain

$$\mathrm{SO}(2,2) \cong \mathrm{SL}(2,\mathbb{R}) \times \mathrm{SL}(2,\mathbb{R})$$

Each of local isomorphisms in the complex case has different real forms which admit some geometrical interpretations. We mentioned above two real forms of the isomorphism $\mathrm{SL}(4;\mathbb{C}) \cong \mathrm{SO}(6;\mathbb{C})$. The isomorphism for $\mathrm{SO}(2,2)$ admits another interpretation in the language of Plücker's coordinates: points of the quadric in $\mathbb{R}P^5$ of the signature $(2,3)$ can be interpreted as (complex) lines in $\mathbb{C}P^3$ which lie on a Hermitian quadric of the signature $(2,2)$ (Gindikin

1983). The isomorphism above for the group $SL(2, \mathbb{H})$ also corresponds to Hopf's fibering of $\mathbb{C}P^3$ on complex lines over the sphere $S^4$ or the isomorphism $S^4$ and the quaternionic projective line $\mathbb{H}P^1$. In all these cases, isomorphisms of homogeneous manifolds intertwine the actions of locally isomorphic groups.

## Pseudo-Riemann Symmetric Manifolds

We obtain the next broad class of homogeneous manifolds if we preserve conditions that the group $G$ is a real semisimple one, the isotropy subgroup $H$ is involutive, but we remove the restriction that $H$ must be (maximal) compact. Such symmetric manifolds are often called semisimple pseudo-Riemann symmetric manifolds (since there are also pseudo-Riemann symmetric manifolds whose groups are not semisimple). This class of spaces contains symmetric Stein manifolds $X_{\mathbb{C}} = G_{\mathbb{C}}/H_{\mathbb{C}}$. Each semisimple symmetric manifold $X = G/H$ admits complexification as a symmetric Stein manifold. Each real semisimple Lie group $G$ is symmetric relative to the group $G \times G$.

The simplest family of semisimple symmetric manifolds is the family of all hyperboloids of all signatures

$$H_{p,q} = \{x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_n^2 = 1\}$$

with the groups $SO(p, q)$. Their complexifications are complex hyperboloids. There are two types of Riemann manifolds in these families: compact ones – spheres and noncompact ones – two-sheeted hyperboloids; all others are pseudo-Riemann.

The Cartan duality holds for pseudo-Hermitian symmetric manifolds: they are domains in compact Hermitian symmetric manifolds (minimal flag manifolds) $Z = G_{\mathbb{C}}/P_{\mathbb{C}}$. They are open orbits of real forms $G$ of the groups of holomorphic automorphisms $G_{\mathbb{G}}$. We construct examples of such manifolds if we consider one of the above-described realizations of noncompact Hermitian symmetric manifolds (through matrix homogeneous coordinates) and replace the condition of positivity with the condition that the symmetric (Hermitian) matrix in the definition has a fixed nondegenerate signature $(\mathrm{i}, k - \mathrm{i})$. We can call such pseudo-Hermitian symmetric manifolds satellites of Hermitian ones. Correspondingly, we can consider nonconvex tubes, for example, the set $T$ of such symmetric matrices whose imaginary parts have the signature $(\mathrm{i}, n - \mathrm{i})$. This domain is linear homogeneous, but it is not symmetric; to receive the symmetric manifold we need to extend the nonconvex tube by a manifold of smaller dimension (which plays a role of infinity).

There are pseudo-Hermitian symmetric manifolds which are not satellites of Hermitian ones. Let us give an interesting example. The group $SL(2p, \mathbb{R})$ has two open orbits on the Grassmannian $\mathrm{Gr}_{\mathbb{C}}(p; 2p)$ which are both pseudo-Hermitian symmetric spaces. Let us consider as above the Stiefel coordinates $Z \in \mathrm{Mat}_{\mathbb{C}}(p, 2p)$ and let $Z = X + \mathrm{i}Y$. Then the orbits are defined by the conditions

$$\det \begin{pmatrix} X \\ Y \end{pmatrix} \gtrless 0$$

In the intersection with the coordinate chart $Z = (E, z), z \in \mathrm{Mat}_{\mathbb{C}}(p), z = x + \mathrm{i}y$, we have the conditions

$$\det y \gtrless 0$$

Therefore, we obtain (nonconvex) tube domains in $\mathbb{C}^N = \mathrm{Mat}_{\mathbb{C}}(p), N = p^2$, corresponding to nonconvex homogeneous cones $V_{\pm}$ of real matrices with positive (negative) determinants. These tubes do not coincide with the symmetric manifolds which include also some sets of small dimensions outside of the coordinate chart (on "infinity"). There are other homogeneous nonconvex cones such that corresponding tube domains are Zariski open parts of pseudo-Hermitian symmetric spaces (D'Atri and Gindikin 1993). Between these cones are cones of nondegenerate skew-symmetric matrices, of skew-Hermitian quaternionic matrices. We again observe strong connections with classical mathematics. Not all pseudo-Hermitian symmetric manifolds admit such tube realizations of dense parts. Analysis in pseudo-Hermitian symmetric manifolds is very interesting: we consider there instead of holomorphic functions $\bar{\partial}$-cohomology of some degree.

Geometric relations between different symmetric manifolds are usually important for analytic applications since they can produce some nontrivial integral transformations. In a broad sense, such transforms are considered in integral geometry (Gelfand *et al.* 2003). An important example is duality between some compact Hermitian symmetric manifolds (when points in one of them are interpreted as submanifolds in another one). The simplest example is the projective duality between dual copies of projective spaces or, more generally, the realization of points of Grassmannians as projective planes. Such a duality can induce a duality between orbits of real forms of groups. In a special case, it can be a duality between Hermitian and pseudo-Hermitian symmetric manifolds.

Here is one important example. Let us consider in the projective space $\mathbb{C}\mathrm{P}^{2k-1}$ the domain $D$ which in

homogeneous coordinates – rows $z = (z_0, z_1, \ldots, z_n)$ are defined by the equation $zHz^* > 0$, where $H$ is a Hermitian form of the signature $(k, k)$, for example,

$$|z_0|^2 + \cdots + |z_k|^2 - |z_{k+1}|^2 - \cdots - |z_n|^2 > 0$$

This domain is $(k-1)$-pseudoconcave and it contains $(k-1)$-dimensional complex compact cycles, namely $(k-1)$-dimensional planes. The manifold of these planes is exactly the domain $X$ in the Grassmannian $\mathrm{Gr}_{\mathbb{C}}(k; 2k)$ (of projective $(k-1)$-planes) which is the noncompact Hermitian symmetric space – the orbit of the group $\mathrm{SU}(k, k)$ (see above). This picture is the geometrical basis for a deep analytic construction. In the domain $D$ the spaces of $(k-1)$-dimensional $\bar{\partial}$-cohomology are infinite dimensional for some coefficients. Their integration on $(k-1)$-planes (the Penrose transform) gives sections of corresponding vector bundles on $X$. The images are described by differential equations – generalized massless equations. The basic twistor theory corresponds to $k = 2$ when $X$ is isomorphic to four-dimensional future tube (see above).

Similar dual realizations of Hermitian symmetric manifolds exist only in special cases. The twistor realization of four-dimensional future tube was possible since the Grassmannian $\mathrm{Gr}_{\mathbb{C}}(2; 4)$ is isomorphic to the quadric in $\mathbb{C}P^5$. This does not work for the future tubes of bigger dimensions but there is another possibility (Gindikin 1998). Let us have the quadric $Q_{n-1} \subset \mathbb{C}P^n$ be defined in the homogeneous coordinates by the equation

$$\square(z) = (z_0)^2 - (z_1)^2 - \cdots - (z_n)^2 = 0$$

and $z \cdot \zeta$ is the bilinear form. As already mentioned, the set of (nondegenerate) hyperplane sections

$$\zeta \cdot z = 0, \quad \zeta \in \mathbb{C}^{n+1}, \ \square(\zeta) = 1$$

of $Q_{n-1}$ is the corresponding hyperboloid $H_n$. Thus, we have the duality between a flag manifold (the quadric $Q_{n-1}$) and a symmetric Stein manifold (the hyperboloid $H_n$) with the same group $\mathrm{SO}(n+1, \mathbb{C})$; they have different dimensions.

The group $\mathrm{SO}(1, n)$ has two orbits on $Q_{n-1}$: the real quadric $Q_{\mathbb{R}} = \{z \in Q_{n-1}; \Im(z) = 0\}$ and its complement $X = Q_{n-1} \backslash Q_{\mathbb{R}}$. Hyperplane sections which do not intersect $Q_{\mathbb{R}}$ (lie at $X$) correspond such $\zeta \in H_n$ that

$$\square(\Re(z)) > 0$$

This set has two connected components $D_\pm$ which are biholomorphically equivalent to the future and past tubes $T_\pm$ of the dimension $n$. Let us emphasize that their group of automorphisms is $\mathrm{SO}(2, n)$ in spite of the fact that this group acts neither on $X$ nor on $H_n$. Such an extension of the symmetry group is a very interesting phenomenon. It happens for several other symmetric manifolds, but is not a general fact. This geometrical construction gives a possibility to construct a multidimensional version of the Penrose transform from $(n-2)$-dimensional $\bar{\partial}$-cohomology with different coefficients into solutions of massless equations on the future (past) tubes.

The last duality is connected with some general geometrical construction. We mentioned that each of the Riemann symmetric manifolds $X = G/K$ admits a canonical embedding in the symmetric Stein manifold $X_{\mathbb{C}} = G_{\mathbb{C}}/K_{\mathbb{C}}$. It turns out that $X$ has in $X_{\mathbb{C}}$ a canonical Stein neighborhood – the complex crown $\Omega(X)$ such that many analytic objects on $X$ can be holomorphically extended on the crown (Gindikin 2002). For example, all solutions of all invariant differential equations on $X$ (which are elliptic) admit such holomorphic extension. In the last example, $D_+$ is the crown of the Riemann symmetric space which is defined, in $H_n$, by the condition $\Im(\zeta) = 0, \Re(\zeta_0) > 0$.

Symmetric manifolds are distinguished from most other homogeneous manifolds by a very rich geometry which is a background for deep analytic considerations. There are several important nonsymmetric homogeneous manifolds. We already mentioned flag manifolds and Stein homogeneous manifolds with complex semisimple Lie groups which can be nonsymmetric. Pseudo-Riemann symmetric manifolds are open orbits of real groups on compact Hermitian symmetric spaces. It turns out that open orbits on other flag manifolds also produce interesting homogeneous manifolds. Let $F = G_{\mathbb{C}}/P_{\mathbb{C}}$ be a flag manifold. Flag domains are open orbits of a real form $G$ on $F$. Of course, pseudo-Hermitian symmetric manifolds are a special case of this construction. Let us consider a simple example with $G_{\mathbb{C}} = \mathrm{SL}(3; \mathbb{C})$ and $P$ – the triangle group. Then points of $F$ are pairs {a point $z$ and a line $l$ passing through it}. Let $G = \mathrm{SU}(2; 1)$; it has two open orbits on $\mathbb{C}\mathrm{P}^2$: the complex ball $D$ and its complementary $D^{\mathrm{C}}$. On $F$, the group $G$ has three open orbits (flag domains): in the first $z \in D$, $l$ is arbitrary; in the second $l \subset D^{\mathrm{C}}$; in the third $z \in D^{\mathrm{C}}, l$ intersects $D$. They are all 1-pseudoconcave. In one-dimensional $\bar{\partial}$-cohomology of these flag domains with coefficients in line bundles, are realized all three discrete series of unitary representations of $\mathrm{SU}(2, 1)$. For arbitrary semisimple Lie groups, all discrete series of representations can also be realized in $\bar{\partial}$-cohomology of flag domains. Crowns of Riemann symmetric spaces which we just mentioned parametrize cycles (complex compact submanifolds)

in flag domains. Some general version of the Penrose transform connects through the integration along cycles cohomology in flag domains with holomorphic solutions of some differential equations in crowns (generalized massless equations).

*See also:* Combinatorics: Overview; Compact Groups and their Representations; Lie Groups: General Theory; Pseudo-Riemannian Nilpotent Lie Groups; Several Complex Variables: Compact Manifolds; Stability of Minkowski Space; Symmetry Classes in Random Matrix Theory; Twistor Theory: Some Applications; Twistors.

## Further Reading


Akhiezer D (1990) Homogeneous complex manifolds. In: Gindikin S and Henkin G (eds.) *Several Complex Variables IV*, vol. 10, Encyclopaedia of Mathematical Science. New York: Springer.

D'Atri J and Gindikin S (1993) Siegel domain realization of pseudo-Hermitian symmetric manifolds. *Geometriae Dedicata* 46: 91–126.

Faraut J and Koranyi A (1994) *Analysis on Symmetric Cones*. Oxford: Oxford University Press.

Gelfand I, Gindikin S, and Graev M (2003) *Selected Topics in Integral Geometry*. Providence, RI: American Mathematical Society.

Gindikin S (1983) The complex universe of Roger Penrose. *Mathematical Intellingencer* 5(1): 27–35.

Gindikin S (1998) SO(1; *n*)-twistors. *Journal of Geometry and Physics* 26: 26–36.

Gindikin S (2002) Some remarks on complex crowns of real symmetric spaces. *Acta Mathematica Applicata* 73(1–2): 95–101.

Helgasson S (1978) *Differential Geometry, Lie Groups and Symmetric Spaces*. New York: Academic Press.

Helgasson S (1994) *Geometric Analysis on Symmetric Spaces*. Providence, RI: American Mathematical Society.

Onishchik A and Vinberg E (1993) Lie Groups and Lie Algebras I Foundations of Lie Theory. In: Onishchik A (ed.) *Lie Groups and Lie Algebras*. Encyclopaedia of Mathematical Sciences, vol. 20. New York: Springer.

Pyatetskii-Shapiro I (1969) *Automorphic Functions and Geometry of Classical Domains*. Amsterdam: Gordon and Breach.


# Classical *r*-Matrices, Lie Bialgebras, and Poisson Lie Groups

**M A Semenov-Tian-Shansky**, Steklov Institute of Mathematics, St. Petersburg, Russia, and Université de Bourgogne, Dijon, France



## Introduction

The notion of "classical *r*-matrices" has emerged as a by-product of the quantum inverse scattering method (which was developed mainly by L D Faddeev and his team in their work at the Steklov Mathematical Institute in Leningrad); it has given a new insight into the study of Hamiltonian structures associated with classical integrable systems solvable by the classical inverse scattering method and its generalizations. Important classification results for classical *r*-matrices are due to Belavin and Drinfeld. Based on the initial results of Sklyanin, Drinfeld introduced the important concepts of "Poisson Lie groups" and "Lie bialgebras" which arise as a semiclassical approximation in the study of quantum groups.

A Poisson group is a Lie group $G$ equipped with a Poisson bracket such that the multiplication $m: G \times G \to G$ is a Poisson mapping. A Poisson bracket on $G$ with this property is called multiplicative. More explicitly, let $\lambda_x, \rho_x$ be the left and right translation operators in $C^\infty(G)$ by an element $x \in G, \lambda_x\varphi(y) = \varphi(xy), \rho_x\varphi(y) = \varphi(yx)$. Multiplication in $G$ is a Poisson mapping, if for any $\varphi, \psi \in C^\infty(G)$, we have

$$\{\varphi, \psi\}(xy) = \{\lambda_x\varphi, \lambda_x\psi\}(y) + \{\rho_y\varphi, \rho_y\psi\}(x) \qquad [1]$$

Note that in general, multiplicative brackets are neither left nor right invariant; in other words, for fixed $x$ translation operators $\lambda_x, \rho_x$ do not preserve Poisson brackets.

Multiplicative Poisson brackets naturally arise in the study of integrable systems which admit the so-called "zero-curvature representation." The study of zero-curvature equations, and in particular, of the Poisson properties of the associated monodromy map, was the main source of nontrivial examples (associated with classical *r*-matrices, classical Yang–Baxter equations, and factorizable Lie bialgebras). The special class of multiplicative Poisson brackets encountered in this context is closely related to factorization problems in Lie groups (in particular, the matrix Riemann problem); these problems represent the key tools in constructing solutions of zero-curvature equations.

The equivalent definition of Poisson Lie groups uses the dual language of "Hopf algebras." Let $A = F(G)$ be the commutative algebra of (smooth) functions on a Lie group $G$ equipped with the standard coproduct $\Delta: A \to A \otimes A$

$$\Delta\varphi(x, y) = \varphi(xy), \qquad \varphi \in F(G), \quad x, y \in G$$

as usual, we identify the (topological) tensor product $F(G) \otimes F(G)$ with $F(G \times G)$. The multiplicative

Poisson bracket on $G$ equips $F(G)$ with the structure of a Poisson–Hopf algebra, that is

$$\Delta\{\varphi, \psi\} = \{\Delta\varphi, \Delta\psi\} \qquad [2]$$

Equation [2] is the starting point for the study of relations between Poisson groups and quantum groups. Following the general philosophy of deformation quantization, we can look for a deformation $A_h$ of the commutative Hopf algebra $A$ with the deformation germ determined by the Poisson structure on $A$ satisfying eqn [2]. The fundamental theorem (conjectured by Drinfeld and proved by Etingof and Kazhdan) asserts that any Poisson algebra associated with a Poisson group admits a formal quantization (in the category of Hopf algebras).

## Poisson Groups and Lie Bialgebras

Let $G$ be a Lie group with Lie algebra $\mathfrak{g}$ equipped with a multiplicative Poisson bracket. Any Poisson bracket is bilinear in differentials of functions; it is convenient to express it by means of right- or left-invariant differentials. For $\varphi \in F(G)$ set

$$\langle \nabla\varphi(x), X\rangle = (\mathrm{d}/\mathrm{d}t)_{t=0}\varphi(\mathrm{e}^{tX}x),$$
$$\langle \nabla'\varphi(x), X\rangle = (\mathrm{d}/\mathrm{d}t)_{t=0}\varphi(x\mathrm{e}^{tX}),$$
$$X \in \mathfrak{g}, \nabla\varphi(x), \nabla'\varphi(x) \in \mathfrak{g}^*$$

Let us define the Poisson operator $\eta : G \to \mathrm{Hom}(\mathfrak{g}^*, \mathfrak{g})$ by setting

$$\{\varphi, \psi\}(x) = \langle \eta(x)\nabla\varphi(x), \nabla\psi\rangle \qquad [3]$$

For a finite-dimensional Lie algebra, we can identify $\mathrm{Hom}(\mathfrak{g}^*, \mathfrak{g})$ with $\mathfrak{g}\otimes\mathfrak{g}$; the skew symmetry of Poisson bracket implies that $\eta \in \mathfrak{g}\wedge\mathfrak{g}$. By an abuse of language, the same identification is traditionally used for infinite-dimensional algebras (e.g., for loop algebras) as well. Of course, in the latter case, the corresponding Poisson tensors are represented by singular kernels which do not lie in the algebraic tensor product and should be regarded as distributions.

Multiplicativity of Poisson bracket on $G$ implies a functional equation for $\eta$

$$\eta(xy) = (\mathrm{Ad}\, x \otimes \mathrm{Ad}\, x)\cdot\eta(y) + \eta(x) \qquad [4]$$

which means that $\eta$ is a 1-cocycle on $G$ (with values in $\mathfrak{g}\wedge\mathfrak{g}$). By setting

$$\delta(X) = \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)_{t=0}\eta(\mathrm{e}^{tX}), \quad X \in \mathfrak{g}$$

we conclude from eqn [4] that $\delta : \mathfrak{g} \to \mathfrak{g}\wedge\mathfrak{g}$ is a 1-cocycle on $\mathfrak{g}$, that is,

$$\begin{aligned}\delta([X, Y]) = {}& [X\otimes I + I\otimes X, \delta(Y)] \\ & - [Y\otimes I + I\otimes Y, \delta(X)]\end{aligned}$$

Equation [4] implies that $\eta(e) = 0$, that is, a multiplicative Poisson structure is identically zero at the unit element. Its linearization at this point induces the structure of a Lie algebra on the cotangent space $T_e^*G \simeq \mathfrak{g}^*$; namely, for any $\xi, \xi' \in \mathfrak{g}^*$, choose $\varphi, \varphi' \in F(G)$ in such a way that $\nabla_e\varphi = \xi, \nabla_e\varphi' = \xi'$, and set

$$[\xi, \xi']_* = \nabla_e\{\varphi, \varphi'\} \qquad [5]$$

It is easy to see that $\langle [\xi, \xi']_*, X\rangle = \langle \xi\wedge\xi', \delta(X)\rangle$, which proves that the bracket is well defined, while eqn [5] implies the Jacobi identity.

**Definition 1** Let $\mathfrak{g}, \mathfrak{g}^*$ be a pair of linear spaces set in duality; $(\mathfrak{g}, \mathfrak{g}^*)$ is called a Lie bialgebra if both $\mathfrak{g}$ and $\mathfrak{g}^*$ are Lie algebras and the mapping $\delta : \mathfrak{g} \to \mathfrak{g}\otimes\mathfrak{g}$ which is dual to the commutator map $[\,,]_* : \mathfrak{g}^*\otimes\mathfrak{g}^* \to \mathfrak{g}^*$ is a 1-cocycle on $\mathfrak{g}$.

Thus if $G$ is a Poisson–Lie group, the pair $(\mathfrak{g}, \mathfrak{g}^*)$ is a Lie bialgebra (called the "tangent Lie bialgebra" of $G$). Poisson–Lie groups form a category in which the morphisms are Lie group homomorphisms, which are also Poisson mappings. A morphism $(\mathfrak{g}, \mathfrak{g}^*) \rightsquigarrow (\mathfrak{h}, \mathfrak{h}^*)$ in the category of Lie bialgebras is a Lie algebra homomorphism $\mathfrak{g} \to \mathfrak{h}$ such that the dual map $\mathfrak{h}^* \to \mathfrak{g}^*$ is again a Lie algebra homomorphism. It is easy to see that morphisms of Poisson groups induce morphisms of their tangent bialgebras. The converse is also true.

**Theorem 1**

(*i*) *Let* $(\mathfrak{g}, \mathfrak{g}^*)$ *be a Lie bialgebra,* $G$ *a connected, simply connected Lie group with Lie algebra* $\mathfrak{g}$. *There is a unique multiplicative Poisson bracket on G such that* $(\mathfrak{g}, \mathfrak{g}^*)$ *is its tangent Lie bialgebra.*
(*ii*) *Morphisms of Lie bialgebras induce Poisson mappings of the corresponding Poisson groups.*

Basically, the theorem asserts that a Poisson tensor is uniquely restored from the infinitesimal cocycle on the corresponding Lie algebra; moreover, the obstruction for the Jacobi identity vanishes globally if this is true for its infinitesimal part at the unit element of the group.

It is important to observe that Lie bialgebras possess a remarkable symmetry: if $(\mathfrak{g}, \mathfrak{g}^*)$ is a Lie bialgebra, the same is true for $(\mathfrak{g}^*, \mathfrak{g})$. Hence, the dual group $G^*$ (which corresponds to $\mathfrak{g}^*$) also carries a multiplicative Poisson bracket. The duality theory for Lie bialgebras, based on the key notion of the Drinfeld double, is discussed in the next section.

## Classical *r*-Matrices and Special Classes of Lie Bialgebras

The general classification problem for Lie bialgebras is unfeasible (e.g., classification of abelian Lie bialgebras includes classification of all Lie algebras). In applications, one mainly deals with important special classes of Lie bialgebras, of which factorizable Lie bialgebras are probably the most important. In a sense, this class may be regarded as exhaustive, since (as explained below) any Lie bialgebra is canonically embedded into a factorizable one. Various other special classes discussed in literature are "coboundary bialgebras," "triangular bialgebras," and "quasitriangular bialgebras."

The Lie bialgebra $(\mathfrak{g}, \mathfrak{g}^*, \delta)$ is called a coboundary bialgebra if the cobracket $\delta$ is a trivial 1-cocycle on $\mathfrak{g}$, that is,

$$\delta(X) = [X \otimes I + I \otimes X, r] \quad \text{for all } X \in \mathfrak{g} \qquad [6]$$

the constant element $r \in \mathfrak{g} \wedge \mathfrak{g}$ is called the "classical $r$-matrix." If $\mathfrak{g}$ is semisimple, $H^1(\mathfrak{g}, V) = 0$ for any $\mathfrak{g}$-module $V$ by the classical Whitehead theorem, and hence all Lie bialgebra structures on $\mathfrak{g}$ are of coboundary type. The associated Lie bracket on $\mathfrak{g}^*$ is given by the formula

$$[\xi, \xi']_* = \mathrm{ad}^*_{\mathfrak{g}}\, r\xi \cdot \xi' - \mathrm{ad}^*_{\mathfrak{g}}\, r\xi' \cdot \xi \qquad [7]$$

where we identified $r \in \mathfrak{g} \wedge \mathfrak{g}$ with a skew-symmetric linear operator $r : \mathfrak{g}^* \to \mathfrak{g}$. The restrictions imposed on $r$ by the Jacobi identity are formulated in terms of the so-called "Yang–Baxter tensor" $[[r, r]] \in \mathfrak{g} \wedge \mathfrak{g} \wedge \mathfrak{g}$, which is a quadratic expression in $r$. To define it, let us mark different factors in tensor products, for example, $\mathfrak{g} \otimes \mathfrak{g} \otimes \mathfrak{g}$, by fixed numbers 1, 2, 3, ... which indicate their place; for simplicity, we assume that $\mathfrak{g}$ is embedded in an associative algebra $\mathcal{A}$ with a unit. The embeddings are defined as

$$i_{12}, i_{23}, i_{13} : \mathfrak{g} \otimes \mathfrak{g} \longrightarrow \mathcal{A} \otimes \mathcal{A} \otimes \mathcal{A}$$

by setting $i_{12}(X \otimes Y) = X \otimes Y \otimes I$, and similarly in other cases. For $a \in \mathfrak{g} \otimes \mathfrak{g}$, we put $i_{12}(a) = a_{12}$, etc. Set

$$[[r, r]] = [r_{12}, r_{13}] + [r_{12}, r_{23}] + [r_{13}, r_{23}] \qquad [8]$$

The commutators in the RHS are computed in the associative algebra $\mathcal{A} \otimes \mathcal{A} \otimes \mathcal{A}$; it is easy to check that the result does not depend on the choice of the embedding $\mathfrak{g} \hookrightarrow \mathcal{A}$.

**Proposition 1** *The Jacobi identity for* $[\,,\,]_*$ *is valid if and only if* $[[r, r]]$ *is* $\mathrm{ad}\,\mathfrak{g}$*-invariant, that is, if*

$$[X \otimes I \otimes I + I \otimes X \otimes I + I \otimes I \otimes X, [[r]]] = 0 \quad \text{for all } X \in \mathfrak{g}$$

A coboundary Lie bialgebra with $[[r, r]] \in (\overset{3}{\wedge}\mathfrak{g})^{\mathfrak{g}}$ is called "quasitriangular"; it is called "triangular" if $r$ satisfies the classical Yang–Baxter equation $[[r, r]] = 0$. (Both terms come from another name of the classical Yang–Baxter equation, the "classical triangle equation.")

When a Lie algebra $\mathfrak{g}$ admits a nondegenerate invariant inner product, the class of quasitriangular Lie bialgebra structures on $\mathfrak{g}$ admits an important specialization. Let $\mathfrak{g} \otimes \mathfrak{g}^* \simeq \mathfrak{g} \otimes \mathfrak{g}$ be the natural isomorphism induced by the inner product. Let $I \in \mathfrak{g} \otimes \mathfrak{g}^*$ be the canonical element; its image $t \in \mathfrak{g} \otimes \mathfrak{g}$ under this isomorphism is called the "tensor Casimir element." Clearly, $t \in (S^2\mathfrak{g})^{\mathfrak{g}}$ and, moreover, $[t_{12}, t_{23}] \in (\overset{3}{\wedge}\mathfrak{g})^{\mathfrak{g}}$. When $\mathfrak{g}$ is semisimple, the mapping $(S^2\mathfrak{g})^{\mathfrak{g}} \to (\overset{3}{\wedge}\mathfrak{g})^{\mathfrak{g}} : s \mapsto [s_{12}, s_{23}]$ is an isomorphism; in particular, if $\mathfrak{g}$ is simple, both spaces are one dimensional and generated by a tensor Casimir (which is unique up to a scalar multiple). A Lie bialgebra $(\mathfrak{g}, r)$ is called factorizable if $r \in \mathfrak{g} \wedge \mathfrak{g}$ satisfies the modified classical Yang–Baxter equation

$$[r, r] = c[t_{12}, t_{23}], \quad c = \text{const} \neq 0 \qquad [9]$$

The convenient normalization is $c = -1/4$ (it can be achieved by an appropriate normalization of $r$). Instead of dealing with the modified Yang–Baxter equation, we may relax the antisymmetry condition imposed on $r$. Set $r_\pm = r \pm (1/2)t \in \mathfrak{g} \otimes \mathfrak{g}$. Since $t$ is $\mathrm{ad}\,\mathfrak{g}$-invariant, the symmetric part of $r_\pm$ drops out from the cobracket; on the other hand, one has $[[r_\pm, r_\pm]] = 0$. Regarding $r_\pm$ as a linear operator, $r_\pm \in \mathrm{Hom}(\mathfrak{g}^*, \mathfrak{g})$, we get the following important result:

**Proposition 2** *Let* $(\mathfrak{g}, \mathfrak{g}^*)$ *be a factorizable Lie bialgebra.*

(*i*) *The mappings* $r_\pm \in \mathrm{Hom}(\mathfrak{g}^*, \mathfrak{g})$ *are Lie algebra homomorphisms*; *moreover*, $r_+^* = -r_-$.
(*ii*) *The combined mapping*

$$i_r : \mathfrak{g}^* \to \mathfrak{g} \oplus \mathfrak{g} : X \mapsto (r_+X, r_-X)$$

*is a Lie algebra embedding.*
(*iii*) *Any* $X \in \mathfrak{g}$ *admits a unique decomposition* $X = X_+ - X_-$ *with* $(X_+, X_-) \in \mathrm{Im}\, i_r$.

The additive decomposition in a factorizable Lie bialgebra gives rise to a multiplicative factorization problem in the associated Lie group. Namely, $i_r$ may be extended to a Lie group embedding $i_r : G^* \to G \times G$ and any $x \in G$, which is sufficiently close to the unit element, admits a decomposition $x = x_+ x_-^{-1}$ with $(x_+, x_-) \in \mathrm{Im}\; i_r$.

Any Lie bialgebra $(\mathfrak{g}, \mathfrak{g}^*)$ admits a canonical embedding into a larger Lie bialgebra (called its

"double") which is already factorizable. Namely, set $\mathfrak{d} = \mathfrak{g} \oplus \mathfrak{g}^*$ as a linear space and equip it with the natural inner product,

$$\langle\langle (X, F), (X', F') \rangle\rangle = \langle F, X' \rangle + \langle F', X \rangle \qquad [10]$$

**Theorem 2**

(*i*) *There exists a unique structure of the Lie algebra on* $\mathfrak{d}$ *such that*: (*a*) $\mathfrak{g}, \mathfrak{g}^* \subset \mathfrak{d}$ *are Lie subalgebras.* (*b*) *The inner product* [10] *is invariant.*

(*ii*) *Let* $P_{\mathfrak{g}}, P_{\mathfrak{g}^*}$ *be the projection operators onto* $\mathfrak{g}, \mathfrak{g}^* \subset \mathfrak{d}$ *parallel to the complementary subalgebra. Set* $r_+^{\mathfrak{d}} = P_{\mathfrak{g}}, r_-^{\mathfrak{d}} = -P_{\mathfrak{g}^*}$; *then* $(\mathfrak{d}, r_\pm^{\mathfrak{d}})$ *is a factorizable Lie bialgebra.*

(*iii*) *The inclusion map* $(\mathfrak{g}, \mathfrak{g}^*) \rightsquigarrow (\mathfrak{d}, \mathfrak{d}^*)$ *is a homomorphism of Lie bialgebras and the dual inclusion map* $(\mathfrak{g}^*, \mathfrak{g}) \rightsquigarrow (\mathfrak{d}, \mathfrak{d}^*)$ *is an antihomomorphism.*

Conversely, let $\mathfrak{a}$ be a Lie algebra equipped with a nondegenerate invariant inner product, $\mathfrak{a}_\pm \subset \mathfrak{a}$ its Lie subalgebras such that (*i*) $\mathfrak{a}_\pm$ are isotropic with respect to inner product, (*ii*) $\mathfrak{a} = \mathfrak{a}_+ \dot{+} \mathfrak{a}_-$ as a linear space. The triple $(\mathfrak{a}, \mathfrak{a}_+, \mathfrak{a}_-)$ is called a "Manin triple." Let $P_\pm$ be the projection operators onto $\mathfrak{a}_\pm$ in this decomposition. Set $r_\pm = \pm P_\pm$. Then $(\mathfrak{a}, r_\pm)$ is a factorizable Lie bialgebra; moreover, $\mathfrak{a}_+$ and $\mathfrak{a}_-$ are set into duality by the inner product in $\mathfrak{a}$ and inherit the structure of a Lie bialgebra, and $\mathfrak{a}$ is their double.

If $(\mathfrak{g}, \mathfrak{g}^*)$ is itself a factorizable Lie bialgebra, its double admits a simple explicit description. Set $\mathfrak{d} = \mathfrak{g} \oplus \mathfrak{g}$ (direct sum of Lie algebras); let us equip $\mathfrak{d}$ with the inner product

$$\langle\langle (X, X'), (Y, Y') \rangle\rangle = \langle X, Y \rangle - \langle Y, Y' \rangle$$

Let $\mathfrak{g}^\delta \subset \mathfrak{d}$ be the diagonal subalgebra; we identify $\mathfrak{g}^*$ with the embedded subalgebra $i_r(\mathfrak{g}^*) \subset \mathfrak{d}$.

**Proposition 3**

(*i*) $(\mathfrak{d}, \mathfrak{g}^\delta, i_r(\mathfrak{g}^*))$ *is a Manin triple.*

(*ii*) *As a Lie algebra,* $\mathfrak{d} = \mathfrak{g} \oplus \mathfrak{g}$ *is isomorphic to the double of* $\mathfrak{g}$.

Key examples of factorizable Lie bialgebras are associated with semisimple Lie algebras and their loop algebras.

1. Let $\mathfrak{k}$ be a compact semisimple Lie algebra: $\mathfrak{g} = \mathfrak{k}_{\mathbb{C}}$ its complexification regarded as a real Lie algebra, $\sigma \in \operatorname{Aut} \mathfrak{g}$ the Cartan involution which fixes $\mathfrak{k}$, and $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ the associated Cartan decomposition. Fix a real split Cartan subalgebra $\mathfrak{a} \subset \mathfrak{p}$ and the associated Iwasawa decomposition $\mathfrak{g} = \mathfrak{k} \dot{+} \mathfrak{a} \dot{+} \mathfrak{n}$; put $\mathfrak{s} = \mathfrak{a} \dot{+} \mathfrak{n}$. Let $B$ be the complex Killing form on $\mathfrak{g}$; let us equip $\mathfrak{g}$ with the real inner product $(X, Y) = \operatorname{Im} B(X, Y)$, then $(\mathfrak{g}, \mathfrak{k}, \mathfrak{s})$ is a Manin triple. Hence, any compact semisimple Lie group $K$ carries a natural Poisson structure; its double $G = D(K)$ is the complex group $G = K_{\mathbb{C}}$ (regarded as a real Lie group). The associated factorization problem in $G$ is the Iwasawa decomposition $G = KAN$, which exists globally.
2. Let $\mathfrak{g}$ be a real split semisimple Lie algebra, $\mathfrak{h}$ its Cartan subalgebra, and $\Delta_+$ a system of positive roots. Fix an invariant inner product on $\mathfrak{g}$ which is positive on $\mathfrak{h}$, and let $\{e_\alpha; \alpha \in \pm\Delta_+\}$ be the root vectors normalized in such a way that $(e_\alpha, e_{-\alpha}) = 1$. Let
$$\mathfrak{n}_\pm = \bigoplus_{\alpha \in \Delta_+} \mathbb{R} \cdot e_{\pm\alpha}$$
Fix an orthonormal basis $\{H_i\}$ in $\mathfrak{h}$; let $P_\pm, P_0$ be the projection operators onto $\mathfrak{n}_\pm, \mathfrak{h}$ in the Bruhat decomposition $\mathfrak{g} = \mathfrak{n}_- \dot{+} \mathfrak{h} \dot{+} \mathfrak{n}_+$. The standard Lie bialgebra structure on $\mathfrak{g}$ is given by the *r*-matrices $r_\pm = \pm P_\pm \pm \frac{1}{2} P_0$. In tensor notation,
$$r_\pm = \pm \sum_{\alpha \in \Delta_+} e_\alpha \wedge e_{-\alpha} \pm \frac{1}{2} \sum_i H_i \otimes H_i \qquad [11]$$
Let $\mathfrak{b}_\pm = \mathfrak{h} \dot{+} \mathfrak{n}_\pm$ be the opposite Borel subalgebras; the inner product in $\mathfrak{g}$ sets them into duality, and $(\mathfrak{b}_+, \mathfrak{b}_-)$ is a Lie sub-bialgebra in $(\mathfrak{g}, \mathfrak{g}^*)$. Let $G$ be the connected, simply connected Lie group associated with $\mathfrak{g}, B_\pm = HN_\pm$ its opposite Borel subgroups which correspond to $\mathfrak{b}_\pm$. Let $p: B_\pm \to B_\pm / N_\pm \simeq H$ be the canonical projection. The associated factorization problem in $G$, $g = b_+ b_-^{-1}$, $(b_+, b_-) \in B_+ \times B_-, p(b_+) = p(b_-)^{-1}$, is closely related to the Bruhat decomposition; it is solvable for all $g$ in the open Bruhat cell $B_+ N_- \subset G$.
3. Let $L\mathfrak{g} = \mathfrak{g} \otimes \mathbb{C}((z))$ be the loop algebra of a finite dimensional semisimple Lie algebra $\mathfrak{g}$, as usual we denote the ring of formal Laurent series by $\mathbb{C}((z))$. Put $L\mathfrak{g}_+ = \mathfrak{g} \otimes \mathbb{C}[[z]]$, $L\mathfrak{g}_- = \mathfrak{g} \otimes z^{-1}\mathbb{C}[z^{-1}]$. Fix an invariant inner product on $\mathfrak{g}$ and equip $L\mathfrak{g}$ with the inner product
$$\langle\langle X, Y \rangle\rangle = \operatorname{Res}_{z=0} \langle X(z), Y(z) \rangle \, dz$$
Then $(L\mathfrak{g}, L\mathfrak{g}_+, L\mathfrak{g}_-)$ is a Manin triple. The associated classical *r*-matrix is called "rational *r*-matrix"; in tensor notation, it is represented by a singular kernel
$$r(z, z') = \frac{t}{z - z'}$$
where $t \in \mathfrak{g} \otimes \mathfrak{g}$ is the tensor Casimir, which is essentially the Cauchy kernel.
4. Let us assume that $\mathfrak{g} = \mathfrak{sl}(n)$; in this case, the loop algebra $L\mathfrak{g}$ admits a nontrivial decomposition

associated with the so-called "elliptic *r*-matrix." Set

$$I_1 = \mathrm{diag}(1, \varepsilon, \ldots, \varepsilon^{n-1}),$$

$$I_2 = \begin{pmatrix} 0 & 1 & & \ldots & 0 \\ & 0 & 1 & & \\ \vdots & & & \ddots & \vdots \\ & & & \ddots & 1 \\ 1 & 0 & \ldots & & 0 \end{pmatrix}, \quad \varepsilon = e^{2\pi i/n} \qquad [12]$$

Put $\mathbb{Z}_n^2 = \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$; for $a = (a_1, a_2) \in \mathbb{Z}_n^2$, set $I_a = I_1^{a_1} I_2^{a_2}$; matrices $I_a$ define an irreducible projective representation of $\mathbb{Z}_n^2$ (they form the so-called "finite Heisenberg group)". Let us denote the elliptic curve of modulus $\tau$ by $\mathcal{E} = \mathbb{C}/\mathbb{Z} + \tau\mathbb{Z}$ and let $P \to \mathcal{E}$ be the *n*-dimensional holomorphic vector bundle with flat connection and with monodromies given by

$$z \mapsto z + 1 : h_1 = \mathrm{Ad}\, I_1, \qquad z \mapsto z + \tau : h_2 = \mathrm{Ad}\, I_2$$

Let $\mathcal{G}_{\mathcal{E}} \subset L\mathfrak{g}$ be the subspace of Laurent expansions at zero of the global meromorphic sections of $P$ with a unique pole at $0 \in \mathcal{E}$. Then $(L\mathfrak{g}, L\mathfrak{g}_+, \mathcal{G}_{\mathcal{E}})$ is again a Manin triple. The associated classical *r*-matrix is the kernel of a singular integral operator which associates a meromorphic section of $P$ to its principal part at 0. Explicitly, it is given by

$$r(z - z') = \frac{1}{n} \sum_{a,b=0}^{n-1} \zeta\left(\frac{z - z'}{n} - a - b\tau\right) \times (\mathrm{Ad}\, I_{a,b} \otimes I) \cdot t \qquad [13]$$

where $\zeta$ is the Weierstrass zeta function.

5. Let $\mathfrak{g}$ be an arbitrary semisimple Lie algebra again. Let us equip the loop algebra $L\mathfrak{g}$ with the inner product

$$\langle\langle X, Y \rangle\rangle_0 = \mathrm{Res}_{z=0} \langle X(z), Y(z) \rangle z^{-1}\, dz$$

Set $\mathcal{N}_+ = \mathfrak{n}_+ \dot{+} \mathfrak{g} \otimes z\mathbb{C}[[z]]$, $\mathcal{N}_- = \mathfrak{n}_- \dot{+} \mathfrak{g} \otimes z^{-1}\mathbb{C}[z^{-1}]$. We have $L\mathfrak{g} = \mathcal{N}_+ \dot{+} \mathfrak{h} \dot{+} \mathcal{N}_-$, where we identify $\mathfrak{h}, \mathfrak{n}_\pm \subset \mathfrak{g}$ with the corresponding subalgebras of constant loops in $L\mathfrak{g}$. Let $P_\pm$, $P_0$ be the projection operators onto $\mathcal{N}_\pm, \mathfrak{h}$ in this decomposition and $r_\pm = \pm P_\pm \pm (1/2)P_0$. The classical *r*-matrices $r_\pm$ define on $L\mathfrak{g}$ the structure of a factorizable Lie bialgebra. The associated tensor kernels are called the trigonometric classical *r*-matrices.

Classical *r*-matrices described above are associated with factorization problems in the infinite-dimensional loop groups: matrix Riemann problems or matrix Cousin problems (in the elliptic case). Belavin and Drinfeld have given a complete classification of factorizable Lie bialgebra structures for semisimple Lie algebras; in the loop algebra case, the problem they solved consists of classification of all meromorphic solutions of the classical Yang–Baxter equation. In other words, we assume that the distribution kernel associated with the classical *r*-matrix is represented by a meromorphic function (of two complex variables). Up to an equivalence, any such solution depends only on one variable and belongs to the rational, trigonometric, or elliptic type (in the latter case, the underlying Lie algebra is necessarily $\mathfrak{sl}(n)$). Classification of solutions in the elliptic case is completely rigid; in the trigonometric case, the moduli space is finite dimensional and admits an explicit description. In the rational case, the classification is somewhat less explicit (it has been completed by A Stolin under some nondegeneracy condition). Contrary to to the popular belief, there are many other structures of a factorizable Lie bialgebra on loop algebras, for which the associated *r*-matrices are given by more singular distribution kernels.

## Poisson Lie Groups

If the tangent Lie bialgebra of a Poisson Lie group is of coboundary type, the cocycle $\eta$ is also trivial, $\eta(\mathfrak{g}) = r - \mathrm{Ad}\,\mathfrak{g} \otimes \mathrm{Ad}\,\mathfrak{g} \cdot r$. Hence, the Poisson bracket on $G$ is given by

$$\{\varphi, \psi\} = \langle r, \nabla'\varphi \wedge \nabla'\psi \rangle - \langle r, \nabla\varphi \wedge \nabla\psi \rangle, \quad r \in \mathfrak{g} \wedge \mathfrak{g}$$

where $\nabla\varphi, \nabla'\varphi \in \mathfrak{g}^*$ are left and right differentials of $\varphi \in C^\infty(G)$. This is the so-called "Sklyanin bracket". Let us assume that $G$ is a matrix group; its affine ring generated by evaluation functions $\phi_{ij}$ which assign to $L \in G$ its matrix coefficients, $\phi_{ij}(L) = L_{ij}$. The Poisson bracket on $G$ is completely determined by its values on $\phi_{ij}$. Explicitly, we get

$$\{\phi_{ij}, \phi_{km}\}(L) = [r, L \otimes L]_{ikjm} \qquad [14]$$

the commutator in the RHS is in $\mathrm{Mat}(n^2)$. By a variation of language, evaluating functions and their values on a generic element $L \in G$ are denoted by the same letter; using tensor notation to suppress matrix indices, we get

$$\{L_1, L_2\} = [r, L_1 L_2], \quad L_1 = L \otimes I,\ L_2 = I \otimes L \quad [15]$$

In the case of loop algebras, these Poisson bracket relations take the form

$$\{L_1(\lambda), L_2(\mu)\} = [r(\lambda, \mu), L_1(\lambda) L_2(\mu)]$$

Let us assume that $G$ is factorizable and the associated factorization problem is globally solvable. The Poisson bracket on the dual group $G^* \simeq$

$i_r(G^*) \subset G \times G$ may be characterized in terms of the matrix coefficients of $(h_+, h_-) = i_r(h)$, or of their quotient $h = h_+ h_-^{-1}$. Explicitly, we get

$$\{h_\pm^1, h_\pm^2\}_* = [r, h_\pm^1 h_\pm^2], \quad \{h_+^1, h_-^2\}_* = [r_+, h_+^1 h_-^2] \quad [16]$$

$$\{h_1, h_2\}_* = r h_1 h_2 + h_1 h_2 r - h_2 r_+ h_1 - h_1 r_- h_2, \quad r = \tfrac{1}{2}(r_+ + r_-) \quad [17]$$

The key question in the geometry of Poisson groups consists in description of symplectic leaves in $G, G^*$. This question is already nontrivial when $G^*$ is abelian (and hence may be identified with the dual of the Lie algebra $\mathfrak{g} = \mathrm{Lie}(G)$). The Poisson bracket on $\mathfrak{g}^*$ is linear; this is the well-known Lie–Poisson (alias, Beresin–Kirillov–Kostant) bracket. Its symplectic leaves coincide with the orbits of the coadjoint representation of $G$ in $\mathfrak{g}^*$. The natural way to prove this fundamental result (which goes back to Lie) is to consider first the natural action of $G$ on the cotangent bundle $T^*G \simeq G \times \mathfrak{g}^*$; this action is Hamiltonian, and the coadjoint orbits arise as a result of Hamiltonian reduction associated with this action. The generalization of the theory of coadjoint orbits to the case of arbitrary Poisson groups starts with the notion of symplectic double, which is the nonlinear analog of the cotangent bundle.

Let $D$ be the double of $(G, G^*)$; assume for simplicity that $D = G \cdot G^*$ globally and hence the associated factorization problem is always solvable. Let $r_\mathfrak{d} = (1/2)(P_\mathfrak{g} - P_{\mathfrak{g}^*})$. Set

$$\{\varphi, \psi\}_\pm = \langle r_\mathfrak{d} \nabla\varphi, \nabla\psi \rangle \pm \langle r_\mathfrak{d} \nabla'\varphi, \nabla'\psi \rangle \quad [18]$$

The bracket $\{\,,\}_-$ is the usual Sklyanin bracket which defines the structure of a Poisson group on $D$, while $\{\,,\}_+$ is nondegenerate and defines a symplectic structure on $D$. Let us denote the copies of $D$ equipped with the bracket $\{\,,\}_\pm$ by $D_\pm$. The bracket on $D_+$ is not multiplicative, but it is covariant with respect to the action of $D_-$ by left and right translations; in other words, the natural mappings $D_- \times D_+ \to D_+$ and $D_+ \times D_- \to D_+$, associated with multiplication in $D$, preserve Poisson brackets. Since $G, G^* \subset D_-$ are Poisson subgroups, natural actions $G \times D_+ \to D_+$ and $G^* \times D_+ \to D_+$ by left and right translations are Poisson mappings. Consider the natural projections

$$\begin{array}{ccc} & D_+ & \\ {}^{\pi}\swarrow & & \searrow^{\pi'} \\ G^* \simeq D/G & & G\backslash D \simeq G^* \end{array} \qquad \begin{array}{ccc} & D_+ & \\ {}^{p}\swarrow & & \searrow^{p'} \\ G \simeq D/G^* & & G^*\backslash D \simeq G \end{array}$$

onto the space of left and right coset classes. It is easy to see that functions on $D_+$ which are constant on each projection fiber are closed with respect to the Poisson bracket. This means that the quotient spaces inherit the Poisson structure. Moreover, the maps $\pi, \pi'$ and $p, p'$ form the so-called "dual pairs", that is, the algebras of functions which are constant on the fibers of $\pi$ and $\pi'$ (or of $p$ and $p'$) are mutual centralizers of one another in the big Poisson algebra $F(D_+)$. Since $D = G \cdot G^* = G^* \cdot G$, we have $G^*/D \simeq G$, $G/D \simeq G^*$; it is easy to check that the quotient Poisson structure induced on $G, G^*$ coincides with the original one. Applying the fundamental theorem on dual pairs of Poisson mappings (going back to S. Lie), we conclude that symplectic leaves in $G$ and $G^*$, respectively, coincide with the orbits of $G^*$ (respectively, $G$) in these quotient spaces. The actions $G \times G^* \to G^*$, $G^* \times G \to G$ are called "dressing transformations". Unit elements in $G$ and $G^*$ are fixed points of dressing transformations; their linearizations at the tangent spaces $T_eG^* \simeq \mathfrak{g}^*$, $T_eG \simeq \mathfrak{g}$ coincide with the coadjoint actions of $G$ and $G^*$, respectively.

When $D \neq G \cdot G^*$ (i.e., the factorization problem in $D$ is not always solvable), dressing actions are still well defined as global transformations of the quotient spaces; in this case $G$, $G^*$ may be identified with open cells in $D/G^*$, $D/G$, respectively, which means that dressing action on $G$, $G^*$ is, in general, incomplete.

If the group $G$ is factorizable, symplectic leaves in the dual group $G^*$ admit a nice uniform description: since in this case $D = G \times G$ and $G \subset D$ is the diagonal subgroup, the quotient $D/G$ may be modeled on $G$ itself. The quotient Poisson bracket in this realization coincides with [17], while the dressing action coincides with conjugation in $G$ (and is independent of $r$). Hence, symplectic leaves in $D/G$ coincide with conjugacy classes in $G$; the equivalence of this model with $G^*$ (equipped with the bracket [16]) is provided by the factorization map. The description of symplectic leaves in $G$ is more subtle (and already crucially depends on the choice of $r$!); for semisimple Lie groups with the standard Poisson structure, it is related to the geometry of double Bruhat cells.

For loop groups with rational, trigonometric, or elliptic $r$-matrices, dressing action is associated with auxiliary factorization problems in the loop group. Roughly speaking, symplectic leaves correspond to rational loops with prescribed singularities. Many important examples have been described in connection with integrable lattice systems, although a complete classification theorem is still not available. For $\mathfrak{g} = \mathfrak{sl}(2)$, the elliptic Manin triple described earlier leads to the Poisson structure on the group of "elliptic loops" with values in SL(2); its simplest symplectic leaves (corresponding to loops with simple poles) are associated with a remarkable Poisson algebra, the Sklyanin algebra (with four generators and two Casimir functions), which admits an interesting explicit quantization.

Dressing action is a nontrivial example of a Poisson group action. In general, such actions are not Hamiltonian in the usual sense; the appropriate generalization is provided by the notion of the nonabelian moment map. Let $G \times \mathcal{M} \to \mathcal{M}$ be an action of a Poisson group $G$ on a Poisson manifold $\mathcal{M}, \mathfrak{g} \to \mathrm{Vect}\,\mathcal{M}$, the associated homomorphism of Lie algebras. A mapping $\mu : \mathcal{M} \to G^*$ is called the nonabelian moment map associated with this action, if for any $X \in \mathfrak{g}$ and $\varphi \in F(\mathcal{M})$, we have

$$X \cdot \varphi = \langle \mu^{-1} \{\mu, \varphi\}_{\mathcal{M}}, X \rangle$$

In this case, $G \times \mathcal{M} \to \mathcal{M}$ is *a fortiori* a Poisson map. Both dressing actions $G^* \times G \to G$ and $G \times G^* \to G^*$ admit nonabelian moment maps, which are just the identity maps $\mu = \mathrm{id}_G$ and $\mu^* = \mathrm{id}_{G^*}$. For compact Poisson groups, the nonabelian moment map has good convexity properties, which generalize the convexity properties of the ordinary moment map for Hamiltonian group actions.

The general theory of homogeneous Poisson spaces has some peculiarities. Typically, the $G$-covariant Poisson structure on a given homogeneous space is not unique (when it exists); this is true already for principal homogeneous spaces (a simple example is provided by the symplectic double $D_+$). Let $G$ be a Poisson Lie group, $(\mathfrak{g}, \mathfrak{g}^*)$ its tangent Lie bialgebra, $\mathfrak{d}$ its double, $U$ its Lie subgroup, $\mathfrak{u} = \mathrm{Lie}\, U$. A subalgebra $\mathfrak{l} \subset \mathfrak{d}$ is called Lagrangian if it is isotropic with respect to the canonical inner product in $\mathfrak{d}$. The general classification result, according to Drinfeld, asserts that there is a bijection between $G$-covariant Poisson structures on $G/U$ and the set of all Lagrangian subalgebras $\mathfrak{l} \subset \mathfrak{d}$ such that $\mathfrak{l} \cap \mathfrak{g} = \mathfrak{u}$. Various nontrivial examples arise, notably in the study of integrable systems. For instance, the geometric proof of the factorization theorem for lattice zero-curvature equation, which is stated in the following section, uses a different Poisson structure on the double (the so-called "twisted symplectic double).”

## Applications to Integrable Systems

The definition of Poisson–Lie groups was motivated by key examples which arise in the theory of integrable systems. In applications, one often deals with nonlinear differential equations which may be written in the form of the so-called "lattice zero curvature equations"

$$\frac{\mathrm{d}L_m}{\mathrm{d}t} = L_m M_m - M_{m+1} L_m, \quad m \in \mathbb{Z} \qquad [19]$$

where $L_m, M_m$ are matrices, possibly depending on an additional parameter (or, more generally, abstract linear operators). Equations [19] give the compatibility conditions for the auxiliary linear system

$$\psi_{m+1} = L_m \psi_m, \quad \frac{\mathrm{d}\psi_m}{\mathrm{d}t} = -M_m \psi_m, \qquad m \in \mathbb{Z} \quad [20]$$

The use of finite-difference operators associated with a one-dimensional lattice, as in [20], is particularly well suited for the study of "multiparticle" lattice models. Let we assume that the "potential" $L_m$ in [20] is periodic, $L_{m+N} = L_m$; the period $N$ may be interpreted as the number of copies of an "elementary" system. It is natural to presume that "Lax matrices" $L_m$ in [19] are elements of a matrix Lie group $G$ (or of a loop group, if they depend on an extra parameter). The auxiliary linear problem [20] leads to a family of dynamical systems on $G^N$ which remain integrable for any $N$. Let $T : G^N \to G$ be the "monodromy map" which assigns to the set $L_1, \ldots, L_N$ of local Lax matrices their ordered product $T_L = L_N L_{N-1} \cdots L_1$. Let us assume that $G$ is equipped with the Sklyanin bracket associated with a factorizable $r$-matrix $r$. Then $T$ is a Poisson map. Let $I(G)$ be the algebra of central functions on $G$; for $\varphi \in I(G)$, set $H_\varphi = \varphi \circ T$. All functions $H_\varphi, \varphi \in I(G)$ are in involution with respect to the product Poisson bracket on $G^N$ and give rise to lattice zero-curvature equations of the same form as [19]; for a given $\varphi$, we may choose the $M$-matrix in either of the two forms:

$$M_m^{\pm} = r_{\pm}\left(\psi_m \nabla\varphi(T_L)\psi_m^{-1}\right), \quad \psi_m = \prod_{1 \le k \le m} L_k$$

Let $L_m(t), m = 1, \ldots, N$, be the integral curve of this equation which starts at $L_m^0$. The construction of this curve reduces to the factorization problem associated with the chosen $r$-matrix. Explicitly, we get

$$L_m(t) = g_{m+1}(t)_+^{-1} L_m^0 g_m(t)_+ = g_{m+1}(t)_-^{-1} L_m^0 g_m(t)_-$$

where $(g_m(t)_+, g_m(t)_-)$ is the curve in $G^*$ which solves the factorization problem

$$g_m(t)_+ g_m(t)_-^{-1} = {}^0\psi_m \exp(t\nabla\varphi(T(L^0)))\, {}^0\psi_m^{-1},$$
$${}^0\psi_m = \psi_m(L^0)$$

This result exhibits the double role of the $r$-matrix. On the one hand, it serves to define the Poisson structure on $G^N$ which is adapted to the study of lattice zero-curvature equations; in particular, the dynamical flow associated with these equations is automatically confined to symplectic leaves in $G^N$. (In applications, $G$ is usually a loop group equipped with a factorizable $r$-matrix; despite the fact that $\dim G = \infty$, it admits plenty finite-dimensional symplectic leaves.) In its second incarnation, the $r$-matrix serves to define the factorization problem which solves these zero-curvature equations. In the loop

group case, this is a matrix Riemann problem; its explicit solution is based on the study of the spectral curve associated with the "monodromy matrix" $T_L$ and uses the technique of algebraic geometry.

The monodromy map $T : G^N \to G$ may be regarded as a nonabelian moment map associated with an action of the dual Lie algebra $\mathfrak{g}^*$ on the phase space. This action actually extends to an action of the (local) Lie group $G^*$ which transforms solutions into solutions again. This is the prototype "dressing" action (originally defined by Zakharov and Shabat in their study of zero-curvature equations related to Riemann–Hilbert problems). Dressing provides an effective tool to produce new solutions of zero-curvature equations from the "trivial" ones; it was also the first nontrivial example of a Poisson group action.

*See also:* Affine Quantum Groups; Bicrossproduct Hopf Algebras and Noncommutative Spacetime; Bi-Hamiltonian Methods in Soliton Theory; Deformations of the Poisson Bracket on a Symplectic Manifold; Functional Equations and Integrable Systems; Hamiltonian Fluid Dynamics; Hopf Algebras and $q$-Deformation Quantum Groups; Integrable Systems and Recursion Operators on Symplectic and Jacobi Manifolds; Integrable Systems: Overview; Lie, Symplectic and Poisson Groupoids, and their Lie Algebroids; Multi-Hamiltonian Systems; Poisson Reduction; Recursion Operators in Classical Mechanics; Toda Lattices; Yang–Baxter Equations.

## Further Reading


Babelon O, Bernard D, and Talon M (2003) *Introduction to Classical Integrable Systems*. Cambridge: Cambridge University Press.

Belavin AA and Drinfel'd VG (1984) Triangle equations and simple Lie algebras. In: *Mathematical physics reviews*, vol. 4, *Soviet Scientific Reviews Section C Mathematical Physics Reviews*, pp. 93–165. Chur: Harwood Academic Publishers, Reprinted in 1998, Classic Reviews in Mathematics and Mathematical Physics, vol. 1. Amsterdam: Harwood Academic Publishers.

Chari V and Pressley A (1995) *A Guide to Quantum Groups*. Cambridge: Cambridge University Press.

Drinfeld VG (1987) Quantum groups. In: *Proceedings of the International Congress of Mathematicians, (Berkeley, Calif., 1986)* vol. 1, pp. 798–820. Providence, RI: American Mathematical Society.

Etingof P and Schiffman O (1998) *Lectures on Quantum Groups*. Boston: International Press.

Frenkel E, Reshetikhin N, and Semenov-Tian-Shansky MA (1998) Drinfeld–Sokolov reduction for difference operators and deformations of W-algebras. I. The case of Virasoro algebra. *Communications in Mathematical Physics* 192(3): 605–629.

Lu J-H (1991) Momentum mappings and reduction of Poisson actions. *Symplectic Geometry, Groupoids, and Integrable Systems (Berkeley, CA, 1989), Mathematical de Sciences Research Institute Publications* vol. 20: 209–226. New York: Springer.

Lu J-H and Weinstein A (1990) Poisson–Lie groups, dressing transformations, and Bruhat decompositions. *Journal of Differential Geometry* 31(2): 501–526.

Reshetikhin N (2000) Characteristic systems on Poisson–Lie groups and their quantization. In: *Integrable Systems: From Classical to Quantum (Montréal, QC, 1999)*, CRM Proceedings Lecture Notes, vol. 26, pp. 165–188. Providence, RI: American Mathematical Society.

Reshetikhin NY and Semenov-Tian-Shansky MA (1990) Central extensions of quantum current groups. *Letters in Mathematical Physics* 19(2): 133–142.

Reyman AG and Semenov-Tian-Shansky MA (1994) Group-theoretical methods in the theory of finite-dimensional integrable systems. In: *Encyclopaedia of Mathematical Sciences, Dynamical Systems VII*, ch. 2, vol. 16, pp. 116–225. Berlin: Springer.

Semenov-Tian-Shansky MA (1994) Lectures on $R$-matrices, Poisson–Lie groups and integrable systems. In: Babelon O, Cartier P, and Kosmann-Schwarzbach Y (eds.) *Lectures on Integrable Systems (Sophia-Antipolis, 1991)*, pp. 269–317. River Edge: World Scientific.

Terng C-L and Uhlenbeck K (1998) Poisson actions and scattering theory for integrable systems. In: *Surveys in Differential Geometry: Integrable Systems*, pp. 315–402. Lectures on geometry and topology, sponsored by Lehigh University's *Journal of Differential Geometry*. A supplement to the *Journal of Differential Geometry*. Edited by Chuu Lian Terng and Karen Uhlenbeck. Surveys in Differential Geometry IV, Boston: International Press.


# Clifford Algebras and Their Representations

**A Trautman**, Warsaw University, Warsaw, Poland



## Introduction

### Introductory and Historical Remarks

Clifford (1878) introduced his "geometric algebras" as a generalization of Grassmann algebras, complex numbers, and quaternions. Lipschitz (1886) was the first to define groups constructed from "Clifford numbers" and use them to represent rotations in a Euclidean space. Cartan discovered representations of the Lie algebras $\mathrm{so}_n(\mathbb{C})$ and $\mathrm{so}_n(\mathbb{R}), n > 2$, that do not lift to representations of the orthogonal groups. In physics, Clifford algebras and spinors appear for the first time in Pauli's nonrelativistic theory of the "magnetic electron." Dirac (1928), in his work on the relativistic wave equation of the electron, introduced matrices that provide a representation of the Clifford algebra of Minkowski space. Brauer and Weyl (1935) connected the Clifford and Dirac ideas with Cartan's spinorial representations of Lie algebras; they found, in any number of dimensions, the spinorial, projective representations of the orthogonal groups.

Clifford algebras and spinors are implicit in Euclid's solution of the Pythagorean equation $x^2 - y^2 + z^2 = 0$, which is equivalent to

$$\begin{pmatrix} y-x & z \\ z & y+x \end{pmatrix} = 2\begin{pmatrix} p \\ q \end{pmatrix}(p \quad q) \qquad [1]$$

and gives $x = q^2 - p^2$, $y = p^2 + q^2$, $z = 2pq$. If the numbers appearing in [1] are real, then this equation can be interpreted as providing a representation of a vector $(x, y, z) \in \mathbb{R}^3$, null with respect to a quadratic form of signature $(1, 2)$, as the "square" of a spinor $(p, q) \in \mathbb{R}^2$. The pure spinors of Cartan (1938) provide a generalization of this observation to higher dimensions.

Multiplying the square matrix in [1] on the left by a real, $2 \times 2$ unimodular matrix, on the right by its transpose, and taking the determinant, one arrives at the exact sequence of group homomorphisms:

$$1 \to \mathbb{Z}_2 \to \mathrm{SL}_2(\mathbb{R}) = \mathrm{Spin}^0_{1,2} \to \mathrm{SO}^0_{1,2} \to 1$$

Multiplying the same matrix by

$$\varepsilon = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \qquad [2]$$

on the left and computing the square of the product, one obtains

$$\begin{pmatrix} z & x+y \\ x-y & -z \end{pmatrix}^2 = (x^2 - y^2 + z^2)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

This equation is an illustration of the idea of representing a quadratic form as the square of a linear form in a Clifford algebra. Replacing $y$ by $\mathrm{i}y$, one arrives at complex spinors, the Pauli matrices,

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \mathrm{i}\varepsilon, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$\mathrm{Spin}_3 = \mathrm{SU}_2$, etc.

This article reviews Clifford algebras, the associated groups, and their representations, for quadratic spaces over complex or real numbers. These notions have been generalized by Chevalley (1954) to quadratic spaces over arbitrary number fields.

## Notation

If $S$ is a vector space over $K = \mathbb{R}$ or $\mathbb{C}$, then $S^*$ denotes its dual, that is, the vector space over $K$ of all $K$-linear maps from $S$ to $K$. The value of $\omega \in S^*$ on $s \in S$ is sometimes written as $\langle s, \omega \rangle$. The transpose of a linear map $f: S_1 \to S_2$ is the map $f^*: S_2^* \to S_1^*$ defined by $\langle s, f^*(\omega)\rangle = \langle f(s), \omega\rangle$ for every $s \in S_1$ and $\omega \in S_2^*$. If $S_1$ and $S_2$ are complex vector spaces, then a map $f: S_1 \to S_2$ is said to be semilinear if it is $\mathbb{R}$-linear and $f(\mathrm{i}s) = -\mathrm{i}f(s)$. The complex conjugate of a finite-dimensional complex vector space $S$ is the complex vector space $\bar{S}$ of all semilinear maps from $S^*$ to $\mathbb{C}$. There is a natural semilinear isomorphism (complex conjugation) $S \to \bar{S}$, $s \mapsto \bar{s}$ such that $\langle \omega, \bar{s}\rangle = \overline{\langle s, \omega\rangle}$ for every $\omega \in S^*$. The space $\bar{\bar{S}}$ can be identified with $S$ and then $\bar{\bar{s}} = s$. The spaces $(\bar{S})^*$ and $\overline{S^*}$ are identified. If $f: S_1 \to S_2$ is a complex-linear map, then there is the complex-conjugate map $\bar{f}: \bar{S}_1 \to \bar{S}_2$ given by $\bar{f}(\bar{s}) = \overline{f(s)}$ and the Hermitian conjugate map $f^\dagger \overset{\mathrm{def}}{=} \bar{f}^*: S_1 \to \bar{S_2}^*$. A linear map $A: S \to \bar{S}^*$ such that $A^\dagger = A$ is said to be Hermitian. $K(N)$ denotes, for $K = \mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$, the set of all $N$ by $N$ matrices with elements in $K$.

## Real, Complex, and Quaternionic Structures

A real structure on a complex vector space $S$ is a complex-linear map $C: S \to \bar{S}$ such that $\bar{C}C = \mathrm{id}_S$. A vector $s \in S$ is said to be real if $\bar{s} = C(s)$. The set of all real vectors is a real vector space; its real dimension is the same as the complex dimension of $S$.

A complex-linear map $C: S \to \bar{S}$ such that $\bar{C}C = -\mathrm{id}_S$ defines on $S$ a quaternionic structure; a necessary condition for such a structure to exist is that the complex dimension $m$ of $S$ be even, $m = 2n$, $n \in \mathbb{N}$. The space $S$ with a quaternionic structure can be made into a right vector space over the field $\mathbb{H}$ of quaternions. In the context of quaternions, it is convenient to represent the imaginary unit of $\mathbb{C}$ as $\sqrt{-1}$. Multiplication on the right by the quaternion unit i is realized as the multiplication (on the left) by $\sqrt{-1}$. If j and $\mathrm{k} = \mathrm{ij}$ are the other two quaternion units and $s \in S$, then one puts $s\mathrm{j} = \bar{C}(\bar{s})$ and $s\mathrm{k} = s\mathrm{ij}$.

A real vector space $S$ can be complexified by forming the tensor product $\mathbb{C} \otimes_{\mathbb{R}} S = S \oplus \mathrm{i}S$.

The realification of a complex vector space $S$ is the real vector space having $S$ as its set of vectors so that $\dim_{\mathbb{R}} S = 2\dim_{\mathbb{C}} S$. The complexification of a realification of $S$ is the "double" $S \oplus S$ of the original space.

## Inner-Product Spaces and Their Groups

**Definitions: quadratic and symplectic spaces** A bilinear map $B: S \times S \to K$ on a vector space $S$ over $K$ is said to make $S$ into an inner-product space. To save on notation, one also writes $B: S \to S^*$ so that $\langle s, B(t)\rangle = B(s, t)$ for all $s, t \in S$. The group of automorphisms of an inner-product space,

$$\mathrm{Aut}(S, B) = \{R \in \mathrm{GL}(S) | R^* \circ B \circ R = B\}$$

is a Lie subgroup of the general linear group $\mathrm{GL}(S)$. An inner-product space $(S, B)$ is said here to be

quadratic (resp., symplectic) if $B$ is symmetric (resp., antisymmetric and nonsingular). A quadratic space is characterized by its quadratic form $s \mapsto B(s,s)$. For $K=\mathbb{C}$, a Hermitian map $A: S \to \bar{S}^*$ defines a Hermitian scalar product $A(s,t) = \langle \bar{s}, A(t) \rangle$.

An orthogonal space is defined here as a quadratic space $(S,B)$ such that $B: S \to S^*$ is an isomorphism. The group of automorphisms of an orthogonal space is the orthogonal group $\mathrm{O(S,B)}$. The group of automorphisms of a symplectic space is the symplectic group $\mathrm{Sp}(S,B)$. The dimension of a symplectic space is even. If $S = K^{2n}$ is a symplectic space over $K=\mathbb{R}$ or $\mathbb{C}$, then its symplectic group is denoted by $\mathrm{Sp}_{2n}(K)$. Two quaternionic symplectic groups appear in the list of spin groups of low-dimensional spaces:

$$\mathrm{Sp}_2(\mathbb{H}) = \{a \in \mathbb{H}(2) \mid a^\dagger a = I\}$$

and

$$\mathrm{Sp}_{1,1}(\mathbb{H}) = \{a \in \mathbb{H}(2) \mid a^\dagger \sigma_z a = \sigma_z\}$$

Here $a^\dagger$ denotes the matrix obtained from $a$ by transposition and quaternionic conjugation.

**Contractions, frames, and orthogonality** From now on, unless otherwise specified, $(V,g)$ is a quadratic space of dimension $m$. Let $\wedge V = \oplus_{p=0}^{m} \wedge^p V$ be its exterior (Grassmann) algebra. For every $v \in V$ and $w \in \wedge V$ there is the contraction $g(v) \rfloor w$ characterized as follows. The map $V \times \wedge V \to \wedge V$, $(v,w) \mapsto g(v) \rfloor w$, is bilinear; if $x \in \wedge^p V$, then $g(v) \rfloor (x \wedge w) = (g(v) \rfloor x) \wedge w + (-1)^p x \wedge (g(v) \rfloor w)$ and $g(v) \rfloor v = g(v,v)$.

A frame $(e_\mu)$ in a quadratic space $(V,g)$ is said to be a quadratic frame if $\mu \neq \nu$ implies $g(e_\mu, e_\nu) = 0$.

For every subset $W$ of $V$ there is the orthogonal subspace $W^\perp$ containing all vectors that are orthogonal to every element of $W$.

If $(V,g)$ is a real orthogonal space, then there is an orthonormal frame $(e_\mu)$, $\mu = 1, \ldots, m$, in $V$ such that $k$ frame vectors have squares equal to $-1$, $l$ frame vectors have squares equal to 1 and $k + l = m$. The pair $(k,l)$ is the signature of $g$. The quadratic form $g$ is said to be neutral if the orthogonal space $(V,g)$ admits two maximal totally null subspaces $W$ and $W'$ such that $V = W \oplus W'$. Such a space $V$ is $2n$-dimensional, either complex or real with $g$ of signature $(n,n)$. A Lorentzian space has maximal totally null subspaces of dimension 1 and a Euclidean space, characterized by a definite quadratic form, has no null subspaces. The Minkowski space is a Lorentzian space of dimension 4.

If $(V,g)$ is a complex orthogonal space, then an orthonormal frame $(e_\mu)$, $\mu = 1, \ldots, m$, can be chosen in $V$ so that, defining $g_{\mu\nu} = g(e_\mu, e_\nu)$, one has $g_{\mu\mu} = (-1)^{\mu+1}$ and, if $\mu \neq \nu$, then $g_{\mu\nu} = 0$.

If $A: S \to \bar{S}^*$ is a Hermitian isomorphism, then there is a (pseudo)unitary frame $(e_\alpha)$ in $S$ such that the matrix $A_{\bar{\alpha}\beta} = A(e_\alpha, e_\beta)$ is diagonal, has $p$ 1's and $q$ $-1$'s on the diagonal, $p + q = \dim S$. If $p = q$, then $A$ is said to be neutral. $A$ is definite if either $p$ or $q = 0$.

## Algebras

**Definitions** An algebra over $K$ is a vector space $\mathcal{A}$ over $K$ with a bilinear map $\mathcal{A} \times \mathcal{A} \to \mathcal{A}$, $(a,b) \mapsto ab$, which is distributive with respect to addition. The algebra is associative if $(ab)c = a(bc)$ holds for all $a,b,c \in \mathcal{A}$. It is commutative if $ab = ba$ for all $a,b \in \mathcal{A}$. An element $1_\mathcal{A}$ is the unit of $\mathcal{A}$ if $1_\mathcal{A} a = a 1_\mathcal{A} = a$ holds for every $a \in \mathcal{A}$.

From now on, unless otherwise specified, the bare word algebra denotes a finite-dimensional, associative algebra over $K = \mathbb{R}$ or $\mathbb{C}$, with a unit element. If $S$ is an $N$-dimensional vector space over $K$, then the set $\mathrm{End}\, S$ of all endomorphisms of $S$ is an $N^2$-dimensional algebra over $K$, the product being defined by composition; if $f, g \in \mathrm{End}\, S$, then one writes $fg$ instead of $f \circ g$; the unit of $\mathrm{End}\, S$ is the identity map $I$. By definition, homomorphisms of algebras map units into units. The map $K \to \mathcal{A}$, $a \mapsto a 1_\mathcal{A}$ is injective and one identifies $K$ with its image in $\mathcal{A}$ by this map so that the unit can be represented by $1 \in K \subset \mathcal{A}$. A set $\mathcal{B} \subset \mathcal{A}$ is said to generate $\mathcal{A}$ if every element of $\mathcal{A}$ can be represented as a linear combination of products of elements of $\mathcal{B}$. For example, if $V$ is a vector space over $K$, then its tensor algebra

$$\mathcal{T}(V) = \oplus_{p=0}^{\infty} \otimes^p V$$

is an (infinite-dimensional) algebra over $K$ generated by $K \oplus V$. The algebra of all $N \times N$ matrices with entries in an algebra $\mathcal{A}$ is denoted by $\mathcal{A}(N)$. Its unit element is the unit matrix $I$. In particular, $\mathbb{R}(N)$, $\mathbb{C}(N)$, and $\mathbb{H}(N)$ are algebras over $\mathbb{R}$. The algebra $\mathbb{R}(2)$ is generated by the set $\{\sigma_x, \sigma_z\}$. As a vector space, the algebra $\mathbb{R}(2)$ is spanned by the set $\{I, \sigma_x, \varepsilon, \sigma_z\}$.

The direct sum $\mathcal{A} \oplus \mathcal{B}$ of the algebras $\mathcal{A}$ and $\mathcal{B}$ over $K$ is an algebra over $K$ such that its underlying vector space is $\mathcal{A} \times \mathcal{B}$ and the product is defined by $(a,b) \cdot (a',b') = (aa', bb')$ for every $a, a' \in \mathcal{A}$ and $b, b' \in \mathcal{B}$. Similarly, the product in the tensor product algebra $\mathcal{A} \otimes_K \mathcal{B}$ is defined by

$$(a \otimes b) \cdot (a' \otimes b') = aa' \otimes bb' \qquad [3]$$

For example, if $\mathcal{A}$ is an algebra over $\mathbb{R}$, then the tensor product algebra $\mathbb{R}(N) \otimes_{\mathbb{R}} \mathcal{A}$ is isomorphic to $\mathcal{A}(N)$ and

$$K(N) \otimes_K K(N') = K(NN') \quad [4]$$

for $K = \mathbb{R}$ or $\mathbb{C}$ and $N, N' \in \mathbb{N}$. There are isomorphisms of algebras over $\mathbb{R}$:

$$\begin{aligned} \mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} &= \mathbb{C} \oplus \mathbb{C} \\ \mathbb{C} \otimes_{\mathbb{R}} \mathbb{H} &= \mathbb{C}(2) \\ \mathbb{H} \otimes_{\mathbb{R}} \mathbb{H} &= \mathbb{R}(4) \end{aligned} \quad [5]$$

An algebra over $\mathbb{R}$ can be complexified by complexifying its underlying vector space; it follows from [5] that $\mathbb{C}(2)$ is the complex algebra obtained by complexification of the real algebra $\mathbb{H}$.

The center of an algebra $\mathcal{A}$ is the set

$$\mathcal{Z}(\mathcal{A}) = \{a \in \mathcal{A} \mid ab = ba \ \forall \ b \in \mathcal{A}\}$$

The center is a commutative subalgebra containing $K$. An algebra over $K$ is said to be central if its center coincides with $K$. The algebras $\mathbb{R}(N)$ and $\mathbb{H}(N)$ are central over $\mathbb{R}$. The algebra $\mathbb{C}(N)$ is central over $\mathbb{C}$, but not over $\mathbb{R}$.

**Simplicity and representations** Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be subsets of the algebra $\mathcal{A}$. Define $\mathcal{B}_1\mathcal{B}_2 = \{b_1 b_2 \mid b_1 \in \mathcal{B}_1, b_2 \in \mathcal{B}_2\}$. A vector subspace $\mathcal{B}$ of $\mathcal{A}$ is said to be a left (resp., right) ideal of $\mathcal{A}$ if $\mathcal{A}\mathcal{B} \subset \mathcal{B}$ (resp., $\mathcal{B}\mathcal{A} \subset \mathcal{B}$). A two-sided ideal – or simply an ideal – is a left and right ideal. An algebra $\mathcal{A} \neq \{0\}$ is said to be simple if its only two-sided ideals are $\{0\}$ and $\mathcal{A}$.

For example, the algebras $\mathbb{R}(N)$ and $\mathbb{H}(N)$ are simple over $\mathbb{R}$; the algebra $\mathbb{C}(N)$ is simple when considered as an algebra over both $\mathbb{R}$ and $\mathbb{C}$; every associative, finite-dimensional simple algebra over $\mathbb{R}$ or $\mathbb{C}$ is isomorphic to one of them.

A representation of an algebra $\mathcal{A}$ over $K$ in a vector space $S$ over $K$ is a homomorphism of algebras $\rho : \mathcal{A} \to \operatorname{End} S$. If $\rho$ is injective, then the representation is said to be faithful. For example, the regular representation $\rho$: $\mathcal{A} \to \operatorname{End} A$ of an algebra $\mathcal{A}$, defined by $\rho(a)b = ab$ for all $a, b \in \mathcal{A}$, is faithful. A vector subspace $T$ of the vector space $S$ carrying a representation $\rho$ of $\mathcal{A}$ is said to be invariant for $\rho$ if $\rho(a)T \subset T$ for every $a \in \mathcal{A}$; it is proper if distinct from both $\{0\}$ and $S$. For example, a left ideal of $\mathcal{A}$ is invariant for the regular representation. Given an invariant subspace $T$ of $\rho$ one can reduce $\rho$ to $T$ by forming the representation $\rho_T : \mathcal{A} \to \operatorname{End} T$, where $\rho_T(a)s = \rho(a)s$ for every $a \in \mathcal{A}$ and $s \in T$. A representation is irreducible if it has no proper invariant subspaces.

A linear map $F : S_1 \to S_2$ is said to intertwine the representations $\rho_1 : \mathcal{A} \to \operatorname{End} S_1$ and $\rho_2 : \mathcal{A} \to \operatorname{End} S_2$ if $F\rho_1(a) = \rho_2(a)F$ holds for every $a \in \mathcal{A}$. If $F$ is an isomorphism, then the representations $\rho_1$ and $\rho_2$ are said to be equivalent, $\rho_1 \sim \rho_2$. The following two propositions are classical:

**Proposition (A)**

(i) *An algebra over $K$ is simple if and only if it admits a faithful irreducible representation in a vector space over $K$. Such a representation is unique, up to equivalence.*

(ii) *The complexification of a central simple algebra over $\mathbb{R}$ is a central simple algebra over $\mathbb{C}$.*

For real algebras, one often considers complex representations, that is, representations in complex vector spaces. Two such representations $\rho_1 : \mathcal{A} \to \operatorname{End} S_1$ and $\rho_2 : \mathcal{A} \to \operatorname{End} S_2$ are said to be complex equivalent if there is a complex isomorphism $F : S_1 \to S_2$ intertwining the representations; they are real equivalent if there is an isomorphism among the realifications of $S_1$ and $S_2$, intertwining the representations. For example, $\mathbb{C}$, considered as an algebra over $\mathbb{R}$, has two complex-inequivalent representations in $\mathbb{C}$: the identity representation and its complex conjugate. The realifications of these representations, given by $\mathrm{i} \mapsto \varepsilon$ and $\mathrm{i} \mapsto -\varepsilon$, respectively, are real equivalent: they are intertwined by $\sigma_z$. The real algebra $\mathbb{H}$, being central simple, has only one, up to complex equivalence, representation in $\mathbb{C}^2$: every such representation is equivalent to the one given by

$$\mathrm{i} \mapsto \sigma_x/\sqrt{-1}, \qquad \mathrm{j} \mapsto \sigma_y/\sqrt{-1}, \qquad \mathrm{k} \mapsto \sigma_z/\sqrt{-1}$$

This representation extends to an injective homomorphism of algebras $i : \mathbb{H}(N) \to \mathbb{C}(2N)$ which is used to define the quaternionic determinant of a matrix $a \in \mathbb{H}(N)$ as $\det_{\mathbb{H}}(a) = \det i(a)$, so that $\det_{\mathbb{H}}(a) \geqslant 0$ and $\det_{\mathbb{H}}(ab) = \det_{\mathbb{H}}(a)\det_{\mathbb{H}}(b)$ for every $a, b \in \mathbb{H}(N)$. In particular, if $q \in \mathbb{H}$ and $\lambda, \mu \in \mathbb{R}$, then $\det_{\mathbb{H}}(q) = \bar{q}q$ and

$$\det_{\mathbb{H}} \begin{pmatrix} \lambda & q \\ -\bar{q} & \mu \end{pmatrix} = (\lambda\mu + \bar{q}q)^2 \quad [6]$$

There are quaternionic unimodular groups $\mathrm{SL}_N(\mathbb{H}) = \{a \in \mathbb{H}(N) \mid \det_{\mathbb{H}}(a) = 1\}$. For example, the group $\mathrm{SL}_1(\mathbb{H})$ is isomorphic to $\mathrm{SU}_2$ and $\mathrm{SL}_2(\mathbb{H})$ is a noncompact, 15-dimensional Lie group, one of the spin groups in six dimensions.

**Antiautomorphisms and inner products** An automorphism of an algebra $\mathcal{A}$ is a linear isomorphism $\alpha$: $\mathcal{A} \to \mathcal{A}$ such that $\alpha(ab) = \alpha(a)\alpha(b)$. An invertible element $c \in \mathcal{A}$ defines an inner automorphism $\mathrm{Ad}(c) \in \mathrm{GL}(\mathcal{A})$, $\mathrm{Ad}(c)a = cac^{-1}$. Complex conjugation in $\mathbb{C}$, considered as an algebra over $\mathbb{R}$, is an automorphism that is not inner. An antiautomorphism of an

algebra $\mathcal{A}$ is a linear isomorphism $\beta:\mathcal{A}\to\mathcal{A}$ such that $\beta(ab)=\beta(b)\beta(a)$ for all $a,b\in\mathcal{A}$. An (anti)automorphism $\beta$ is involutive if $\beta^2=\mathrm{id}$. For example, conjugation of quaternions defines an involutive antiautomorphism of $\mathbb{H}$.

Let $\rho:\mathcal{A}\to\mathrm{End}\,S$ be a representation of an algebra with an involutive antiautomorphism $\beta$. There is then the contragredient representation $\check{\rho}:\mathcal{A}\to\mathrm{End}\,S^*$ given by $\check{\rho}(a)=(\rho(\beta(a)))^*$. If, moreover, $\mathcal{A}$ is central simple and $\rho$ is faithful irreducible, then there is an isomorphism $B:S\to S^*$ intertwining $\rho$ and $\check{\rho}$ which is either symmetric, $B^*=B$, or antisymmetric, $B^*=-B$. It defines on $S$ the structure of an inner-product space. This structure extends to $\mathrm{End}\,S$: there is a symmetric isomorphism $B\otimes B^{-1}:\mathrm{End}\,S\to(\mathrm{End}\,S)^*=\mathrm{End}\,S^*$ given, for every $f\in\mathrm{End}\,S$, by $(B\otimes B^{-1})(f)=BfB^{-1}$.

Let $K^\times=K\backslash\{0\}$ be the multiplicative group of the field $K$. Given a simple algebra $\mathcal{A}$ with an involutive antiautomorphism $\beta$, one defines $N(a)=\beta(a)a$ and the group

$$\mathcal{G}(\beta)=\{a\in\mathcal{A}\,|\,N(a)\in K^\times\}$$

Let $\rho:\mathcal{A}\to\mathrm{End}\,S$ be the faithful irreducible representation as above, then, for $a\in\mathcal{A}$ and $s,t\in S$, one has

$$B(\rho(a)s,\rho(a)t)=N(a)B(s,t)$$

If $a\in\mathcal{G}(\beta)$ and $\lambda\in K^\times$, then $\lambda a\in\mathcal{G}(\beta)$ and the norm $N$ satisfies $N(\lambda a)=\lambda^2N(a)$. The inner product $B$ is invariant with respect to the action of the group

$$\mathcal{G}_1(\beta)=\{a\in\mathcal{G}(\beta)\,|\,N(a)=1\}$$

**Proposition (B)** *Let $\mathcal{A}$ be a central simple algebra over $K$ with an involutive antiautomorphism $\beta$ and a faithful irreducible representation $\rho$ so that*

$$\check{\rho}(a)=B\rho(a)B^{-1}$$

*The map $h:\mathcal{A}\times\mathcal{A}\to K$ defined by*

$$h(a,b)=\mathrm{tr}\,\rho(\beta(a)b)$$

is bilinear, symmetric, and nondegenerate. The map *$\rho$ is an isometry of the quadratic space $(\mathcal{A},h)$ on its image in the quadratic space* $(\mathrm{End}\,S, B\otimes B^{-1})$.

## Graded Algebras

**Definitions** An algebra $\mathcal{A}$ is said to be $\mathbb{Z}$-graded (resp., $\mathbb{Z}_2$-graded) if there is a decomposition of the underlying vector space $\mathcal{A}=\oplus_{p\in\mathbb{Z}}\mathcal{A}^p$ (resp., $\mathcal{A}=\mathcal{A}^0\oplus\mathcal{A}^1$) such that $\mathcal{A}^p\mathcal{A}^q\subset\mathcal{A}^{p+q}$. In a $\mathbb{Z}_2$-graded algebra, it is understood that $p+q$ is reduced mod 2. If $a\in\mathcal{A}^p$, then $a$ is said to be homogeneous of degree $p$. The exterior algebra $\wedge V$ of a vector space $V$ is $\mathbb{Z}$-graded. Every $\mathbb{Z}$-graded algebra becomes $\mathbb{Z}_2$-graded when one reduces the degree of every element mod 2. A graded isomorphism of graded algebras is an isomorphism that preserves the grading.

A $\mathbb{Z}_2$-grading of $\mathcal{A}$ is characterized by the involutive automorphism $\alpha$ such that, if $a\in\mathcal{A}^p$, then $\alpha(a)=(-1)^pa$. From now on, grading means $\mathbb{Z}_2$-grading unless otherwise specified. The elements of $\mathcal{A}^0$ (resp., $\mathcal{A}^1$) are said to be even (resp., odd). It is often convenient to denote the graded algebra as

$$\mathcal{A}^0\to\mathcal{A} \qquad [7]$$

Given such an algebra over $K$ and $N\in\mathbb{N}$, one constructs the graded algebra $\mathcal{A}^0(N)\to\mathcal{A}(N)$. Two graded algebras over $K$, $\mathcal{A}^0\to\mathcal{A}$ and $\mathcal{A}'^0\to\mathcal{A}'$ are said to be of the same type if there are integers $N$ and $N'$ such that the algebras $\mathcal{A}^0(N)\to\mathcal{A}(N)$ and $\mathcal{A}'^0(N')\to\mathcal{A}'(N')$ are graded isomorphic. The property of being of the same type is an equivalence relation in the set of all graded algebras over $K$.

Given an algebra $\mathcal{A}$, one constructs two "canonical" graded algebras as follows:

1. the double algebra

$$\mathcal{A}\to\mathcal{A}\oplus\mathcal{A}$$

graded by the "swap" automorphism, $\alpha(a_1,a_2)=(a_2,a_1)$ for $a_1,a_2\in\mathcal{A}$;

2. the algebra

$$\mathcal{A}\oplus\mathcal{A}\to\mathcal{A}(2)$$

is defined by declaring the diagonal (resp., antidiagonal) elements of $\mathcal{A}(2)$ to be even (resp., odd).

The real algebra $\mathbb{R}(2)$ has also another grading, given by the involutive automorphism $\alpha$ such that $\alpha(a)=\varepsilon a\varepsilon^{-1}$, where $a\in\mathbb{R}(2)$ and $\varepsilon$ is as in [2]. In this case, [7] reads

$$\mathbb{C}\to\mathbb{R}(2)$$

There are also graded algebras over $\mathbb{R}$:

$$\mathbb{R}\to\mathbb{C},\quad \mathbb{C}\to\mathbb{H},\quad \text{and}\quad \mathbb{H}\to\mathbb{C}(2)$$

The grading of the last algebra can be defined by declaring the Pauli matrices and i$I$ to be odd.

**Super Lie algebras** A super Lie algebra is a graded algebra $\mathcal{A}$ such that the product $(a,b)\mapsto[a,b]$ is super anticommutative, $[a,b]=-(-1)^{pq}[b,a]$, and satisfies the super Jacobi identity,

$$[a,[b,c]]=[[a,b],c]+(-1)^{pq}[b,[a,c]]$$

for every $a\in\mathcal{A}^p$, $b\in\mathcal{A}^q$ and $c\in\mathcal{A}$. To every graded associative algebra $\mathcal{A}$ there corresponds a super Lie algebra $\mathcal{GLA}$: its underlying vector space and grading are as in $\mathcal{A}$ and the product, for $a\in\mathcal{A}^p$

and $b \in \mathcal{A}^q$, is given as the supercommutator $[a, b] = ab - (-1)^{pq} ba$.

**Supercentrality and graded simplicity** A graded algebra $\mathcal{A}$ over $K$ is supercentral if $\mathcal{Z}(\mathcal{A}) \cap \mathcal{A}^0 = K$. The algebra $\mathbb{R} \to \mathbb{C}$ is supercentral, but the real ungraded algebra $\mathbb{C}$ is not central.

A subalgebra $\mathcal{B}$ of a graded algebra $\mathcal{A}$ is said to be a graded subalgebra if $\mathcal{B} = \mathcal{B} \cap \mathcal{A}^0 \oplus \mathcal{B} \cap \mathcal{A}^1$. A graded ideal of $\mathcal{A}$ is an ideal that is a graded subalgebra. A graded algebra $\mathcal{A} \neq \{0\}$ is said to be graded simple if it has no graded ideals other than $\{0\}$ and $\mathcal{A}$. The double algebra of a simple algebra is graded simple, but not simple.

**The graded tensor product** Let $\mathcal{A}$ and $\mathcal{B}$ be graded algebras; the tensor product of their underlying vector spaces admits a natural grading, $(\mathcal{A} \otimes \mathcal{B})^p = \oplus_q \mathcal{A}^q \otimes \mathcal{B}^{p-q}$. The product defined in [3] makes $\mathcal{A} \otimes \mathcal{B}$ into a graded algebra. There is another "super" product in the same graded vector space given by

$$(a \otimes b) \cdot (a' \otimes b') = (-1)^{pq} aa' \otimes bb'$$

for $a' \in \mathcal{A}^p$ and $b \in \mathcal{B}^q$. The resulting graded algebra is referred to as the graded tensor product and denoted by $\mathcal{A} \hat{\otimes} \mathcal{B}$. For example, if $V$ and $W$ are vector spaces, then the Grassmann algebra $\wedge(V \oplus W)$ is isomorphic to $\wedge V \hat{\otimes} \wedge W$.

## Clifford Algebras

### Definitions: The Universal Property and Grading

The Clifford algebra associated with a quadratic space $(V, g)$ is the quotient algebra

$$\mathcal{C}\ell(V, g) = \mathcal{T}(V)/\mathcal{J}(V, g) \qquad [8]$$

where $\mathcal{J}(V, g)$ is the ideal in the tensor algebra $\mathcal{T}(V)$ generated by all elements of the form $v \otimes v - g(v, v)1_{\mathcal{T}(V)}$, $v \in V$.

The Clifford algebra is associative with a unit element denoted by 1. One denotes by $\kappa$ the canonical map of $\mathcal{T}(V)$ onto $\mathcal{C}\ell(V, g)$ and by $ab$ the product of two elements $a, b \in \mathcal{C}\ell(V, g)$ so that $\kappa(P \otimes Q) = \kappa(P)\kappa(Q)$ for $P, Q \in \mathcal{T}(V)$. The map $\kappa$ is injective on $K \oplus V$, and one identifies this subspace of $\mathcal{T}(V)$ with its image under $\kappa$. With this identification, for all $u, v \in V$, one has

$$uv + vu = 2g(u, v)$$

Clifford algebras are characterized by their universal property described in the following proposition.

**Proposition (C)** *Let $\mathcal{A}$ be an algebra with a unit $1_{\mathcal{A}}$ and let $f : V \to \mathcal{A}$ be a* Clifford map, *that is, a linear map such that $f(v)^2 = g(v, v)1_{\mathcal{A}}$ for every $v \in V$. There then exists a homomorphism $\hat{f}: \mathcal{C}\ell(V, g) \to \mathcal{A}$ of algebras with units, an extension of $f$, so that $f(v) = \hat{f}(v)$ for every $v \in V$.*

As a corollary, one obtains

**Proposition (D)** *If $f$ is an isometry of $(V, g)$ into $(W, h)$, then there is a homomorphism of algebras $\mathcal{C}\ell(f) : \mathcal{C}\ell(V, g) \to \mathcal{C}\ell(W, h)$ extending $f$ so that there is the commutative diagram*

$$\begin{array}{ccc} \mathcal{C}\ell(V, g) & \xrightarrow{\mathcal{C}\ell(f)} & \mathcal{C}\ell(W, h) \\ \uparrow & & \uparrow \\ V & \underset{f}{\longrightarrow} & W \end{array}$$

For example, the isometry $v \mapsto -v$ extends to the involutive main automorphism $\alpha$ of $\mathcal{C}\ell(V, g)$, defining its $\mathbb{Z}_2$-grading:

$$\mathcal{C}\ell(V, g) = \mathcal{C}\ell^0(V, g) \oplus \mathcal{C}\ell^1(V, g)$$

The algebra $\mathcal{C}\ell(V, g)$ admits also an involutive canonical antiautomorphism $\beta$ characterized by $\beta(1) = 1$ and $\beta(v) = v$ for every $v \in V$.

### The Vector Space Structure of Clifford Algebras

Referring to proposition (D), let $\mathcal{A} = \mathrm{End}(\wedge V)$ and, for every $v \in V$ and $w \in \wedge V$, put $f(v)w = v \wedge w + g(v) \rfloor w$, then $f : V \to \mathrm{End}(\wedge V)$ is a Clifford map and the map

$$i : \mathcal{C}\ell(V, g) \to \wedge V \qquad [9]$$

given by $i(a) = \hat{f}(a)1_{\wedge V}$ is an isomorphism of vector spaces. This proves

**Proposition (E)** *As a* vector space, *the algebra $\mathcal{C}\ell(V, g)$ is isomorphic to the exterior algebra $\wedge V$.*

If $V$ is $m$-dimensional, then $\mathcal{C}\ell(V, g)$ is $2^m$-dimensional. The linear isomorphism [9] defines a $\mathbb{Z}$-grading of the vector space underlying the Clifford algebra: if $i(a_k) \in \wedge^k V$, then $a_k$ is said to be of Grassmann degree $k$. Every element $a \in \mathcal{C}\ell(V, g)$ decomposes into its Grassmann components, $a = \sum_{k \in \mathbb{Z}} a_k$. The Clifford product of two elements of Grassmann degrees $k$ and $l$ decomposes as follows: $a_k b_l = \sum_{p \in \mathbb{Z}} (a_k b_l)_p$, and $(a_k b_l)_p = 0$ if $p < |k - l|$ or $p \equiv k - l + 1 \bmod 2$ or $p > m - |m - k - l|$.

One often uses [9] to identify the vector spaces $\wedge V$ and $\mathcal{C}\ell(V, g)$; this having been done, one can write, for every $v \in V$ and $a \in \mathcal{C}\ell(V, g)$,

$$va = v \wedge a + g(v) \rfloor a \qquad [10]$$

so that $[v, a] = 2g(v) \rfloor a$, where $[\,,\,]$ is the supercommutator. It defines a super Lie algebra structure in the vector space $K \oplus V$. The quadratic form defined by $g$ need not be nondegenerate; for example, if it is the

0-form, then [10] shows that the Clifford and exterior multiplications coincide and $\mathcal{C}\ell(V,0)$ is isomorphic, as an algebra, to the Grassmann algebra.

### Complexification of Real Clifford Algebras

**Proposition (F)** *If $(V,g)$ is a real quadratic space, then the algebras $\mathbb{C}\otimes\mathcal{C}\ell(V,g)$ and $\mathcal{C}\ell(\mathbb{C}\otimes V,\mathbb{C}\otimes g)$ are isomorphic, as graded algebras over $\mathbb{C}$.*

From now on, through the end of the article, one assumes that $(V,g)$ is an orthogonal space over $K=\mathbb{R}$ or $\mathbb{C}$.

The Clifford algebra associated with the orthogonal space $\mathbb{C}^m$ is denoted by $\mathcal{C}\ell_m$. The Clifford algebra associated with the orthogonal space $(\mathbb{R}^{k+l},g)$, where $g$ is of signature $(k,l)$, is denoted by $\mathcal{C}\ell_{k,l}$, so that $\mathbb{C}\otimes\mathcal{C}\ell_{k,l}=\mathcal{C}\ell_{k+l}$.

### Relations between Clifford Algebras in Spaces of Adjacent Dimensions

Consider an orthogonal space $(V,g)$ over $K$ and the one-dimensional orthogonal space $(K,h_1)$, having a unit vector $w\in K$, $h_1(w,w)=\varepsilon$, where $\varepsilon=1$ or $-1$. The map $V\ni v\mapsto vw\in\mathcal{C}\ell^0(V\oplus K,g\oplus h_1)$ satisfies $(vw)^2=-\varepsilon g(v,v)$ and extends to the isomorphism of algebras $\mathcal{C}\ell(V,-\varepsilon g)\to\mathcal{C}\ell^0(V\oplus K,g\oplus h_1)$. This proves

**Proposition (G)** *There are isomorphisms of algebras: $\mathcal{C}\ell_m\to\mathcal{C}\ell^0_{m+1}$ and $\mathcal{C}\ell_{k,l}\to\mathcal{C}\ell^0_{k+1,l}$.*

Consider the orthogonal space $(K^2,h)$ with a neutral $h$ such that, for $\lambda,\mu\in K$, one has $\langle(\lambda,\mu),h(\lambda,\mu)\rangle=\lambda\mu$. The map

$$K^2\to K(2),\qquad(\lambda,\mu)\mapsto\begin{pmatrix}0&\lambda\\ \mu&0\end{pmatrix}$$

has the Clifford property and establishes the isomorphisms represented by the horizontal arrows in the diagram

$$\begin{array}{ccc}\mathcal{C}\ell(K^2,h)&\to&K(2)\\ \uparrow&&\uparrow\\ \mathcal{C}\ell^0(K^2,h)&\to&K\oplus K\end{array}\qquad[11]$$

**Proposition (H)** *If $(K^2,h)$ is neutral and $(V,g)$ is over $K$, then the algebra $\mathcal{C}\ell(V\oplus K^2,g\oplus h)$ is isomorphic to the algebra $\mathcal{C}\ell(V,g)\otimes K(2)$. Specifically, there are isomorphisms*

$$\begin{aligned}\mathcal{C}\ell_{k+1,l+1}&=\mathcal{C}\ell_{k,l}\otimes\mathbb{R}(2)\\ \mathcal{C}\ell_{m+2}&=\mathcal{C}\ell_m\otimes\mathbb{C}(2)\end{aligned}\qquad[12]$$

### The Chevalley Theorem and the Brauer–Wall Group

If $(V,g)$ and $(W,h)$ are quadratic spaces over $K$, then their sum is the quadratic space $(V\oplus W,g\oplus h)$ characterized by $g\oplus h:V\oplus W\to V^*\oplus W^*$ so that $(g\oplus h)(v,w)=(g(v),h(w))$. By noting that the map $V\oplus W\ni(v,w)\mapsto v\otimes1+1\otimes w\in\mathcal{C}\ell(V,g)\hat{\otimes}\mathcal{C}\ell(W,h)$ has the Clifford property, Chevalley proved

**Proposition (I)** *The algebra $\mathcal{C}\ell(V\oplus W,g\oplus h)$ is isomorphic to the algebra $\mathcal{C}\ell(V,g)\hat{\otimes}\mathcal{C}\ell(W,h)$.*

The type of the (graded) algebra $\mathcal{C}\ell(V\oplus W,g\oplus h)$ depends only on the types of $\mathcal{C}\ell(V,g)$ and $\mathcal{C}\ell(W,h)$. The Chevalley theorem (I) shows that the set of types of Clifford algebras over $K$ forms an abelian group for a multiplication induced by the graded tensor product. The unit of this Brauer–Wall group of $K$ is the type of the algebra $\mathcal{C}\ell(K^2,h)$ described in [11]; for a full account with proofs, see Wall (1963).

### The Volume Element and the Centers

Let $e=(e_\mu)$ be an orthonormal frame in $(V,g)$. The volume element associated with $e$ is

$$\eta=e_1e_2\cdots e_m$$

If $\eta'$ is the volume element associated with another orthonormal frame $e'$ in the same orthogonal space, then either $\eta'=\eta$ ($e$ and $e'$ are of the same orientation) or $\eta'=-\eta$ ($e$ and $e'$ are of opposite orientation). For $K=\mathbb{C}$, one has $\eta^2=1$; for $K=\mathbb{R}$ and $g$ of signature $(k,l)$ one has

$$\eta^2=(-1)^{(1/2)(k-l)(k-l+1)}\qquad[13]$$

It is convenient to define $\iota\in\{1,i\}$ so that $\eta^2=\iota^2$. For every $v\in V$ one has $v\eta=(-1)^{m+1}\eta v$. The structure of the centers of Clifford algebras is as follows:

**Proposition (J)** *If $m$ is even, then $\mathcal{Z}(\mathcal{C}\ell(V,g))=K$ and $\mathcal{Z}(\mathcal{C}\ell^0\ (V,g))=K\oplus K\eta$. If $m$ is odd, then $\mathcal{Z}(\mathcal{C}\ell(V,g))=K\oplus K\eta$ and $\mathcal{Z}(\mathcal{C}\ell^0(V,g))=K$.*

*The graded algebra $\mathcal{C}\ell(V,g)$ is supercentral for every $m$.*

### The Structure of Clifford Algebras

**The complex case** Using [4] one obtains from [11] and [12] the isomorphisms of algebras

$$\mathcal{C}\ell^0_{2n+1}=\mathcal{C}\ell_{2n}=\mathbb{C}(2^n)\qquad[14]$$

$$\mathcal{C}\ell_{2n+1}=\mathcal{C}\ell^0_{2n+2}=\mathbb{C}(2^n)\oplus\mathbb{C}(2^n)\qquad[15]$$

for $n=0,1,2,\ldots$. Therefore, there are only two types of complex Clifford algebras, represented by $\mathbb{C}\to\mathbb{C}\oplus\mathbb{C}$ and $\mathbb{C}\oplus\mathbb{C}\to\mathbb{C}(2)$: the Brauer–Wall group of $\mathbb{C}$ is $\mathbb{Z}_2$.

**The real case** In view of proposition (I) and $\mathcal{C}\ell_{1,1} = \mathbb{R}(2)$, the algebra $\mathcal{C}\ell_{k,l}$ is of the same type as $\mathcal{C}\ell_{k-l,0}$ if $k > l$ and of the same type as $\mathcal{C}\ell_{0,l-k}$ if $k < l$. Since $\mathcal{C}\ell_{k,l} \hat{\otimes} \mathcal{C}\ell_{l,k} = \mathcal{C}\ell_{k+l,k+l}$, the type of $\mathcal{C}\ell_{l,k}$ is the inverse of the type of $\mathcal{C}\ell_{k,l}$. The algebra $\mathcal{C}\ell^0_{4,0} \to \mathcal{C}\ell_{4,0}$ is isomorphic to $\mathbb{H} \oplus \mathbb{H} \to \mathbb{H}(2)$: if $x = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \subset \mathcal{C}\ell_{4,0}$, and $q = \mathrm{i}x_1 + \mathrm{j}x_2 + \mathrm{k}x_3 + x_4 \in \mathbb{H}$, then an isomorphism is obtained from the Clifford map $f$,

$$f(x) = \begin{pmatrix} 0 & q \\ -\bar{q} & 0 \end{pmatrix} \qquad [16]$$

In view of [13], the volume element $\eta$ satisfies $\eta^2 = 1$. By replacing $-\bar{q}$ with $\bar{q}$ in [16], one shows that $\mathcal{C}\ell_{0,4}$ is also isomorphic to $\mathbb{H}(2)$. The map $\mathbb{R}^4 \times \mathbb{R}^{k+l} \to \mathbb{H}(2) \otimes \mathcal{C}\ell_{k,l}$ given by $(x, y) \mapsto f(x) \otimes 1 + \eta \otimes y$ has the Clifford property and establishes the isomorphism of algebras $\mathcal{C}\ell_{k+4,l} = \mathbb{H} \otimes \mathcal{C}\ell_{k,l}$. Since, similarly, $\mathcal{C}\ell_{k,l+4} = \mathbb{H} \otimes \mathcal{C}\ell_{k,l}$, one obtains the isomorphism

$$\mathcal{C}\ell_{k+4,l} = \mathcal{C}\ell_{k,l+4}$$

Therefore,

$$\mathcal{C}\ell_{k+8,l} = \mathcal{C}\ell_{k+4,l+4} = \mathcal{C}\ell_{k,l+8} = \mathcal{C}\ell_{k,l} \otimes \mathbb{R}(16)$$

and the algebras $\mathcal{C}\ell_{k,l}$, $\mathcal{C}\ell_{k+8,l}$, and $\mathcal{C}\ell_{k,l+8}$ are all of the same type. This double periodicity of period 8 is subsumed by saying that real Clifford algebras can be arranged on a "spinorial chessboard." The type of $\mathcal{C}\ell^0_{k,l} \to \mathcal{C}\ell_{k,l}$ depends only on $k - l \bmod 8$; the eight types have the following low-dimensional algebras as representatives: $\mathcal{C}\ell_{1,0}$, $\mathcal{C}\ell_{2,0}$, $\mathcal{C}\ell_{3,0}$, $\mathcal{C}\ell_{4,0} = \mathcal{C}\ell_{0,4}$, $\mathcal{C}\ell_{0,3}$, $\mathcal{C}\ell_{0,2}$, and $\mathcal{C}\ell_{0,1}$. The Brauer–Wall group of $\mathbb{R}$ is $\mathbb{Z}_8$, generated by the type of $\mathcal{C}\ell^0_{1,0} \to \mathcal{C}\ell_{1,0}$, that is, by $\mathbb{R} \to \mathbb{C}$. Bearing in mind the isomorphism $\mathcal{C}\ell_{k,l} = \mathcal{C}\ell^0_{k+1,l}$ and abbreviating $\mathbb{C} \to \mathbb{R}(2)$ to $\mathbb{C} \to \mathbb{R}$, etc., one can arrange the types of real Clifford algebras in the form of a "spinorial clock":

$$\begin{array}{ccccc} \mathbb{R} & \overset{7}{\rightarrow} & \mathbb{R} \oplus \mathbb{R} & \overset{0}{\rightarrow} & \mathbb{R} \\ 6\uparrow & & & & \downarrow 1 \\ \mathbb{C} & & & & \mathbb{C} \\ 5\uparrow & & & & \downarrow 2 \\ \mathbb{H} & \underset{4}{\leftarrow} & \mathbb{H} \oplus \mathbb{H} & \underset{3}{\leftarrow} & \mathbb{H} \end{array} \qquad [17]$$

**Proposition (K)** *Recipe for determining* $\mathcal{C}\ell^0_{k,l} \to \mathcal{C}\ell_{k,l}$:

(i) *find the integers* $\mu$ *and* $\nu$ *such that* $k - l = 8\mu + \nu$ *and* $0\ \nu\ 7$;
(ii) *from the spinorial clock, read off* $\mathcal{A}^0_\nu \to \nu\mathcal{A}_\nu$ *and compute the real dimensions*, $\dim \mathcal{A}^0_\nu = 2^{\tau^0}$ *and* $\dim \mathcal{A}_\nu = 2^\tau$; *and*
(iii) *form* $\mathcal{C}\ell^0_{k,l} = \mathcal{A}^0_\nu(2^{(1/2)(k+l-1-\tau^0)})$ *and* $\mathcal{C}\ell_{k,l} = \mathcal{A}_\nu(2^{(1/2)(k+l-\tau)})$.

The spinorial clock is symmetric with respect to the reflection in the vertical line through its center; this is a consequence of the isomorphism of algebras $\mathcal{C}\ell_{k,l+2} = \mathcal{C}\ell_{l,k} \otimes \mathbb{R}(2)$.

Note that the "abstract" algebra $\mathcal{C}\ell_{k,l}$ carries, in general, less information than the Clifford algebra defined in [8], which contains $V$ as a distinguished vector subspace with the quadratic form $v \mapsto v^2 = g(v, v)$. For example, the algebras $\mathcal{C}\ell_{8,0}$, $\mathcal{C}\ell_{4,4}$, and $\mathcal{C}\ell_{0,8}$ are all graded isomorphic.

### Theorem on Simplicity

From general theory (Chevalley 1954) or by inspection of [14], [15], and [17], one has

**Proposition (L)** *Let* $m$ *be the dimension of the orthogonal space* $(V, g)$ *over* $K$.

(i) *If* $m$ *is even (resp., odd), then the algebra* $\mathcal{C}\ell(V, g)$ *(resp.,* $\mathcal{C}\ell^0(V, g)$*) over* $K$ *is central simple.*
(ii) *If* $K = \mathbb{C}$ *and* $m$ *is odd (resp., even), then the algebra* $\mathcal{C}\ell(V, g)$ *(resp.,* $\mathcal{C}\ell^0(V, g)$*) is the direct sum of two isomorphic complex central simple algebras.*
(iii) *If* $K = \mathbb{R}$ *and* $m$ *is odd (resp., even), then the algebra* $\mathcal{C}\ell(V, g)$ *(resp.,* $\mathcal{C}\ell^0(V, g)$*) when* $\eta^2 = 1$ *is the direct sum of two isomorphic central simple algebras and when* $\eta^2 = -1$ *is simple with a center isomorphic to* $\mathbb{C}$.

## Representations

### The Pauli, Cartan, Dirac, and Weyl Representations

**Odd dimensions** Let $(V, g)$ be of dimension $m = 2n + 1$ over $K$. From propositions (A) and (L) it follows that the central simple algebra $\mathcal{C}\ell^0(V, g)$ has a unique, up to equivalence, faithful, and irreducible representation in the complex $2^n$-dimensional vector space $S$ of Pauli spinors. By putting $\sigma(\eta) = \iota I$ it is extended to a Pauli representation $\sigma : \mathcal{C}\ell(V, g) \to \operatorname{End} S$. Given an orthonormal frame $(e_\mu)$ in $V$, Pauli endomorphisms (matrices if $S$ is identified with $\mathbb{C}^{2^n}$) are defined as $\sigma_\mu = \sigma(e_\mu) \in \operatorname{End} S$. The representations $\sigma$ and $\sigma \circ \alpha$ are complex inequivalent. For $K = \mathbb{C}$ none of them is faithful; their direct sum is the faithful Cartan representation of $\mathcal{C}\ell(V, g)$ in $S \oplus S$. For $K = \mathbb{R}$ and $(1/2)(k - l - 1)$ even, the representations $\sigma$ and $\sigma \circ \alpha$ are real equivalent and faithful. On computing $\beta(\eta)$ one finds that the contragredient representation $\check{\sigma}$ is equivalent to $\sigma$ for $n$ even and to $\sigma \circ \alpha$ for $n$ odd.

**Even dimensions** Similarly, for $(V, g)$ of dimension $m = 2n$ over $K$, the central simple algebra $\mathcal{C}\ell(V, g)$ has a unique, up to equivalence, faithful, and

irreducible representation $\gamma : \mathcal{C}\ell(V,g) \to \operatorname{End} S$ in the $2^n$-dimensional complex vector space $S$ of Dirac spinors. The Dirac endomorphisms (matrices) are $\gamma_\mu = \gamma(e_\mu)$. Put $\Gamma = \iota\gamma(\eta)$ so that $\Gamma^2 = I$: the matrix $\Gamma$ generalizes the familiar $\gamma_5$. The Dirac representation $\gamma$ restricted to $\mathcal{C}\ell^0(V,g)$ decomposes into the sum $\gamma_+ \oplus \gamma_-$ of two irreducible representations in the vector spaces

$$S_\pm = \{s \in S \,|\, \Gamma s = \pm s\}$$

of Weyl (chiral) spinors. The elements of $S_+$ are said to be of opposite chirality with respect to those of $S_-$. The transpose $\Gamma^*$ defines a similar split of $S^*$. The representations $\gamma_+$ and $\gamma_-$ are never complex-equivalent, but they are real equivalent and faithful for $K = \mathbb{R}$ and $(1/2)(k-l)$ odd.

The representations $\gamma \circ \alpha$ and $\check{\gamma}$ are both equivalent to $\gamma$. It is convenient to describe simultaneously the properties of the transpositions of the Pauli and Dirac matrices; let $\rho_\mu$ be either the Pauli matrices for $V$ of dimension $2n+1$ or the Dirac matrices for $V$ of dimension $2n$. There is a complex isomorphism $B : S \to S^*$ such that

$$\rho^*_\mu = (-1)^n B \rho_\mu B^{-1} \qquad [18]$$

In the case of the Dirac matrices, the factor $(-1)^n$ in [18] implies that this equation also holds for $\Gamma$ in place of $\rho_\mu$. The isomorphism $B$ preserves (resp., changes) the chirality of Weyl spinors for $n$ even (resp., odd). Every matrix of the form $B\gamma_{\mu_1} \ldots \gamma_{\mu_p}$, where

$$1 \leqslant \mu_1 < \cdots < \mu_p 2n \qquad [19]$$

is either symmetric or antisymmetric, depending on $p$ and the symmetry of $B$. A simple argument, based on counting the number of such products of one symmetry, leads to the equation

$$B^* = (-1)^{(1/2)n(n+1)} B$$

valid in dimensions $2n$ and $2n+1$.

**Inner products on spinor spaces** Let $S$ be the complex vector space of Dirac or Pauli spinors associated with $(V,g)$ over $K$. The isomorphism $B$: $S \to S$ defines on $S$ an inner product $B(s,t) = \langle s, B(t)\rangle$, $s,t \in S$, which is orthogonal for $m \equiv 0,1,6,$ or $7 \bmod 8$ and symplectic for $m \equiv 2,3,4,$ or $5 \bmod 8$. For $m \equiv 0 \bmod 4$, this product restricts to an inner product on the space of Weyl spinors that is orthogonal for $m \equiv 0 \bmod 8$ and symplectic for $m \equiv 4 \bmod 8$. For $m \equiv 2 \bmod 4$, the map $B$ defines the isomorphisms $B_\pm : S_\pm \to S^*_\mp$.

**Example** One of the most used representations $\gamma$: $\mathcal{C}\ell_{3,1} \to \mathbb{C}(4)$ is given by the Dirac matrices

$$\gamma_1 = \begin{pmatrix} 0 & \sigma_x \\ -\sigma_x & 0 \end{pmatrix}, \qquad \gamma_2 = \begin{pmatrix} 0 & \sigma_y \\ -\sigma_y & 0 \end{pmatrix}$$

$$\gamma_3 = \begin{pmatrix} 0 & \sigma_z \\ -\sigma_z & 0 \end{pmatrix}, \qquad \gamma_4 = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix} \qquad [20]$$

### Change Conjugation and Majorana Spinors

Throughout this section and next, one assumes $K = \mathbb{R}$ so that, given a representation $\rho : \mathcal{C}\ell(V,g) \to \operatorname{End} S$,one can form the complex- ("charge") conjugate representation $\bar{\rho} : \mathcal{C}\ell(V,g) \to \operatorname{End} \bar{S}$ defined by $\bar{\rho}(a) = \overline{\rho(a)}$ and the Hermitian conjugate representation $\rho^\dagger : \mathcal{C}\ell(V,g) \to \operatorname{End} \bar{S}^*$, where $\rho^\dagger(a) = \overline{\check{\rho}(a)}$.

**Even dimensions** The representations $\bar{\gamma}$ and $\gamma$ are equivalent: there is an isomorphism $C : S \to \bar{S}$ such that

$$\overline{\gamma}_\mu = C\gamma_\mu C^{-1} \qquad [21]$$

The automorphism $\bar{C}C$ is in the commutant of $\gamma$; it is, therefore, proportional to $I$ and, by a change of scale, one can achieve $\bar{C}C = I$ for $k - l \equiv 0$ or $6 \bmod 8$ and $\bar{C}C = -I$ for $k - l \equiv 2$ or $4 \bmod 8$.

The spinor $s_c = C^{-1}\bar{s} \in S$ is the charge conjugate of $s \in S$. If $\psi : V \to S$ is a solution of the Dirac equation

$$(\gamma^\mu(\partial_\mu - \mathrm{i}qA_\mu) - \kappa)\psi = 0$$

for a particle of electric charge $q$, then $\psi_c$ is a solution of the same equation with the opposite charge. Since

$$\overline{\Gamma} = \iota^2 C\Gamma C^{-1}$$

charge conjugation preserves (resp., changes) the chirality of Weyl spinors for $(1/2)(k-l)$ even (resp., odd).

If $\bar{C}C = I$, then

$$\operatorname{Re} S = \{s \in S \,|\, s_c = s\}$$

is a real vector space of dimension $2^n$, the space of Dirac–Majorana spinors. The representation $\gamma$ is real: restricted to $\operatorname{Re} S$ and expressed with respect to a frame in this space, it is given by real $2^n \times 2^n$ matrices. For $k - l \equiv 0 \bmod 8$ the representations $\gamma_+$ and $\gamma_-$ are both real: in this case there are Weyl–Majorana spinors.

**Odd dimensions** On computing $\overline{\sigma(\eta)}$ one finds that the conjugate representation $\bar{\sigma}$ is equivalent to $\sigma$

(resp., $\sigma \circ \alpha$) if $\eta^2 = 1$ (resp., $\eta^2 = -1$). There is an isomorphism $C : S \to \bar{S}$ such that

$$\overline{\sigma}_\mu = (-1)^{(1/2)(k-l+1)} C \sigma_\mu C^{-1} \qquad [22]$$

and $\bar{C}C = I$ (resp., $\bar{C}C = -I$) for $k - l \equiv 1$ or $7 \bmod 8$ (resp., $k - l \equiv 3$ or $5 \bmod 8$). For $k - l \equiv 1 \bmod 8$, the restriction of the Pauli representation to $\mathcal{C}\ell^0_{k,l}$ is real and the Pauli matrices are pure imaginary; for $k - l \equiv 7 \bmod 8$, the Pauli representations of $\mathcal{C}\ell_{k,l}$ are both real and so are the Pauli matrices. In both these cases there are Pauli–Majorana spinors.

### Hermitian Scalar Products and Multivectors

For $m = k + l$ odd and $C$ as in [22], the map $A = \bar{B}C : S \to \bar{S}^*$ intertwines the representations $\sigma^\dagger$ and $\sigma$ (resp., $\sigma \circ \alpha$) for $k$ even (resp., odd),

$$\sigma^\dagger_\mu = (-1)^k A \sigma_\mu A^{-1}$$

By rescaling of $B$, the map $A$ can be made Hermitian. The corresponding Hermitian form $s \mapsto A(s, s)$ is definite if and only if $k$ or $l = 0$; otherwise, it is neutral.

For $m = k + l$ even, the representations $\gamma^\dagger$ and $\gamma$ are equivalent and one can define a Hermitian isomorphism $A : S \to \bar{S}^*$ so that

$$\gamma^\dagger_\mu = A \gamma_\mu A^{-1} \qquad [23]$$

The isomorphism $A' = A\Gamma$ intertwines the representations $\gamma^\dagger$ and $\gamma \circ \alpha$; it can also be made Hermitian by rescaling. The Hermitian form $A(s, s)$ is definite for $k = 0$ and $A'(s, s)$ is definite for $l = 0$; otherwise, these forms are neutral. For example, in the familiar representation [20], one has $A = \gamma_4$, a neutral form.

For $p = 0, 1, \ldots, m = 2n$, two spinors $s$ and $t \in S$ define the $p$-vector with components

$$A_{\mu_1 \ldots \mu_p}(s, t) = \langle \bar{s}, A \gamma_{\mu_1} \ldots \gamma_{\mu_p} t \rangle \qquad [24]$$

where the indices are as in [19]. The Hermiticity of $A$ and [23] imply

$$\overline{A_{\mu_1 \ldots \mu_p}(s, t)} = (-1)^{(1/2)p(p-1)} A_{\mu_1 \ldots \mu_p}(t, s)$$

In view of $\Gamma^\dagger = (-1)^k A \Gamma A^{-1}$, the map $A$ defines, for $k$ even, a nondegenerate Hermitian scalar product on the spaces $S_\pm$ whereas $A(s, t) = 0$ if $s$ and $t$ are Weyl spinors of opposite chiralities. For $k$ odd, $A$ changes the chirality.

## The Radon–Hurwitz Numbers

**Proposition** (M) *For every integer $m > 0$, the algebra $\mathcal{C}\ell_{m,0}$ has an irreducible real representation $\rho$ of dimension $2^{\chi(m)}$, where $\chi(m)$ is the mth Radon–Hurwitz number given by*

| $m =$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $\chi(m) =$ | 1 | 2 | 2 | 3 | 3 | 3 | 3 | 4 |

*and $\chi(m + 8) = \chi(m) + 4$. The matrices $\rho_\mu \in \mathbb{R}(2^{\chi(m)})$, $\mu = 1, \ldots, m$, defining these representations satisfy*

$$\rho_\mu \rho_\nu + \rho_\nu \rho_\mu = -2\delta_{\mu\nu} I$$

*and can be chosen so as to be antisymmetric. In all dimensions other than $m \equiv 3 \bmod 4$ the representations are faithful.*

*For $m \equiv 2$ and $4 \bmod 8$ (resp., $m \equiv 1, 3,$ and $5 \bmod 8$) the representations $\rho$ are the realifications of the corresponding Dirac (resp., Pauli) representations. In dimensions $m \equiv 0$ and $6 \bmod 8$ (resp., $m \equiv 7 \bmod 8$) the Dirac (resp., Pauli) representations themselves are real.*

## Inductive Construction of Representations

An inductive construction of the Pauli representations

$$\sigma : \mathcal{C}\ell_{n-1,n} \to \mathbb{R}(2^{n-1}), \quad n = 1, 2, \ldots$$

and of the Dirac representations

$$\gamma : \mathcal{C}\ell_{n,n} \to \mathbb{R}(2^n), \quad n = 1, 2, \ldots$$

is as follows.

1. In dimension 1, put $\sigma_1 = 1$.
2. Given $\sigma_\mu \in \mathbb{R}(2^{n-1})$, $\mu = 1, \ldots, 2n - 1$, define

$$\gamma_\mu = \begin{pmatrix} 0 & \sigma_\mu \\ \sigma_\mu & 0 \end{pmatrix} \quad \text{for } \mu = 1, \ldots, 2n - 1$$

and

$$\gamma_{2n} = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$$

3. Given $\gamma_\mu \in \mathbb{R}(2^n)$, $\mu = 1, \ldots, 2n$, define $\sigma_\mu = \gamma_\mu$ for $\mu = 1, \ldots, 2n$, and $\sigma_{2n+1} = \gamma_1 \cdots \gamma_{2n}$.

All entries of these matrices are either 0, 1, or $-1$; therefore, they can be used to construct representations of Clifford algebras of orthogonal spaces over any commutative field of characteristic $\neq 2$.

By induction, one has $\sigma^*_\mu = (-1)^{\mu+1} \sigma_\mu$. Therefore, the isomorphisms appearing in [18] are $B = \gamma_2 \gamma_4 \cdots \gamma_{2n}$ for both $m = 2n$ and $2n + 1$.

By multiplying some of the matrices $\sigma_\mu$ or $\gamma_\mu$ by the imaginary unit, one obtains complex representations of the Clifford algebras associated with the quadratic

forms of other signatures. For example, in dimension 3, $(\sigma_1, i\sigma_2, \sigma_3)$ are the Pauli matrices. In dimension 4, multiplying $\gamma_2$ by i one obtains the Dirac matrices for $g$ of signature $(1,3)$, in the "chiral representation":

$$\gamma_1 = \begin{pmatrix} 0 & \sigma_x \\ \sigma_x & 0 \end{pmatrix}, \qquad \gamma_2 = \begin{pmatrix} 0 & \sigma_y \\ \sigma_y & 0 \end{pmatrix}$$
$$\gamma_3 = \begin{pmatrix} 0 & \sigma_z \\ \sigma_z & 0 \end{pmatrix}, \qquad \gamma_4 = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} \qquad [25]$$

To obtain the real Majorana representation one uses the following fact:

**Proposition (N)** *If the matrix* $C \in \mathbb{R}(2^n)$ *is such that* $C^2 = I$ *and* [21] *holds, then the matrices* $(I + iC)\gamma_\mu(I + iC)^{-1}$, $\mu = 1, \ldots, 2n$, {\it are real}.

For the matrices [25], one can take $C = \gamma_1\gamma_3\gamma_4$ to obtain

$$\gamma'_1 = \begin{pmatrix} 0 & \sigma_x \\ \sigma_x & 0 \end{pmatrix}, \qquad \gamma'_2 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}$$
$$\gamma'_3 = \begin{pmatrix} 0 & \sigma_z \\ \sigma_z & 0 \end{pmatrix}, \qquad \gamma'_4 = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$$

The real representations described in proposition (M) can be obtained by the following direct inductive construction. Consider the following seven real antisymmetric and anticommuting $8 \times 8$ matrices:

$$\begin{aligned} &\rho_1 = \sigma_z \otimes I \otimes \varepsilon, && \rho_2 = \sigma_z \otimes \varepsilon \otimes \sigma_x \\ &\rho_3 = \sigma_z \otimes \varepsilon \otimes \sigma_z, && \rho_4 = \sigma_x \otimes \varepsilon \otimes I \\ &\rho_5 = \sigma_x \otimes \sigma_x \otimes \varepsilon, && \rho_6 = \sigma_x \otimes \sigma_z \otimes \varepsilon \\ &\rho_7 = \varepsilon \otimes I \otimes I \end{aligned} \qquad [26]$$

For $m = 4, 5, 6,$ and $7$ the matrices $\rho_1, \ldots, \rho_m$ generate the representations of $C\ell_{m,0}$ in $\mathbb{R}^8$. The eight matrices $\theta_\mu = \sigma_x \otimes \rho_\mu, \mu = 1, \ldots, 7,$ and $\theta_8 = \varepsilon \otimes I \otimes I \otimes I$ give the required representation of $C\ell_{8,0}$ in $\mathbb{R}^{16}$. By dropping the first factor in $\rho_1, \rho_2, \rho_3$, one obtains the matrices generating a representation of $C\ell_{3,0}$ in $\mathbb{R}^4$, etc. The symmetric matrix $\Theta = \theta_1 \cdots \theta_8 = \sigma_z \otimes I \otimes I \otimes I$ anticommutes with all the $\theta$s and $\Theta^2 = I$. If the matrices $\rho_\mu \in \mathbb{R}(2^{\chi(m)})$ correspond to a representation of $C\ell_{m,0}$, then the $m + 8$ matrices $\Theta \otimes \rho_1, \ldots, \Theta \otimes \rho_m, \theta_1 \otimes I, \ldots, \theta_8 \otimes I$ generate the required representation of $C\ell_{m+8,0}$.

## Vector Fields on Spheres and Division Algebras

It is known that even-dimensional spheres have no nowhere-vanishing tangent vector fields. All such fields on odd-dimensional spheres can be constructed with the help of the representation $\rho$ described in proposition (M). Given a positive even integer $N$, let $m$ be the largest integer such that $N = 2^{\chi(m)}p$, where $p$ is an odd integer. Consider the unit sphere $\mathbb{S}_{N-1} = \{x \in \mathbb{R}^N \mid ||x|| = 1\}$ of dimension $N - 1$. For $v \in \mathbb{R}^m$, put $\rho'(v) = \rho(v) \otimes I$, where $I \in \mathbb{R}(p)$ is the unit matrix. Since $\rho(v)$ is antisymmetric, so is the matrix $\rho'(v) \in \mathbb{R}(N)$. Therefore, for every $x \in \mathbb{S}_{N-1}$, the vector $\rho'(v)x$ is orthogonal to $x$. The map $x \mapsto \rho'(v)x$ defines a vector field on $\mathbb{S}_{N-1}$ that vanishes nowhere unless $v = 0$: the $(N-1)$-sphere admits a set of $m$ tangent vector fields which are linearly independent at every point. Using methods of algebraic topology, it has been shown that this method gives the maximum number of linearly independent tangent vector fields on spheres.

If $m = 1, 3,$ or $7$, then $m + 1 = 2^{\chi(m)}$ and, for these values of $m$, the sphere $\mathbb{S}_m$ is parallelizable. Moreover, one can then introduce in $\mathbb{R}^{m+1}$ the structure of an algebra $A_m$ as follows. Put $\rho_0 = I$. If $e_0 \in \mathbb{R}^{m+1}$ is a unit vector and $e_\mu = \rho_\mu(e_0)$, then $(e_0, e_1, \ldots, e_m)$ is an orthonormal frame in $\mathbb{R}^{m+1}$. The product of $x = \sum_{\mu=0}^m x_\mu e_\mu$ and $y = \sum_{\mu=0}^m y_\mu e_\mu$ is defined to be

$$x \cdot y = \sum_{\mu,v=0}^{m} x_\mu y_v \rho_\mu(e_v)$$

so that $e_0$ is the unit element for this product. Defining $\mathrm{Re}\,x = x_0 e_0$, $\mathrm{Im}\,x = x - \mathrm{Re}\,x$, $\bar{x} = \mathrm{Re}\,x - \mathrm{Im}\,x$, one has $\bar{x} \cdot x = e_0||x||^2$ and $\bar{x} \cdot (x \cdot y) = (\bar{x} \cdot x) \cdot y$, so that $x \cdot y = 0$ implies $x = 0$ or $y = 0$: $A_m$ is a normed algebra without zero divisors. The algebras $A_1$ and $A_3$ are isomorphic to $\mathbb{C}$ and $\mathbb{H}$, respectively, and $A_7$ is, by definition, the algebra $\mathbb{O}$ of octonions discovered by Graves and Cayley. The algebra $\mathbb{O}$ is nonassociative; its multiplication table is obtained with the help of [26].

## Spinor Groups

Let $(V, g)$ be a quadratic space over $K$. If $u \in V$ is not null, then it is invertible as an element of $C\ell(V, g)$ and the map $v \mapsto -uvu^{-1}$ is a reflection in the hyperplane orthogonal to $u$. The orthogonal group $\mathrm{O}(V, g) = \mathrm{O}(V, -g) = \{R \in \mathrm{GL}(V) \mid R^* \circ g \circ R = g\}$ is generated by the set of all such reflections. A spinor group $G$ is a subset of $C\ell(V, g)$ that is a group with respect to multiplication induced by the product in the algebra, with a homomorphism $\rho: G \to \mathrm{GL}(V)$ whose image contains the connected component $\mathrm{SO}^0(V, g)$ of the group of rotations of $(V, g)$. In the case of real quadratic spaces, one considers also spinor groups that are subsets of $\mathbb{C} \otimes C\ell(V, g)$ with similar properties. By restriction, every

representation of $C\ell(V,g)$ or $\mathbb{C}\otimes C\ell(V,g)$ gives spinor representations of the spinor groups it contains.

### Pin Groups

It is convenient to define a unit vector $v\in V\subset C\ell(V,g)$ to be such that $v^2=1$ for $V$ complex and $v^2=1$ or $-1$ for $V$ real. The group $\mathrm{Pin}(V,g)$ is defined as the subgroup of $\mathrm{Cpin}(V,g)$ consisting of products of all finite sequences of unit vectors. Defining now the twisted adjoint representation $\widetilde{\mathrm{Ad}}$ by $\widetilde{\mathrm{Ad}}(a)v=\alpha(a)va^{-1}$, one ontains the exact sequence

$$1\to\mathbb{Z}_2\to\mathrm{Pin}(V,g)\overset{\widetilde{\mathrm{Ad}}}{\to}\mathrm{O}(V,g)\to 1 \qquad [27]$$

If $\dim V$ is even, then the adjoint representation $\mathrm{Ad}(a)v=ava^{-1}$ also yields an exact sequence like [27]; if it is odd, then the image of Ad is $\mathrm{SO}(V,g)$ and the kernel is the four-element group $\{1,\,-1,\eta,\,-\eta\}$.

Given an orthonormal frame $(e_\mu)$ in $(V,g)$ and $a\in\mathrm{Pin}(V,g)$, one defines the orthogonal matrix $R(a)=(R_\mu^\nu(a))$ by

$$\widetilde{\mathrm{Ad}}(a)e_\mu=e_\nu R_\mu^\nu(a) \qquad [28]$$

If $(V,g)$ is complex, then the algebras $C\ell(V,g)$ and $C\ell(V,\,-g)$ are isomorphic; this induces an isomorphism of the groups $\mathrm{Pin}(V,g)$ and $\mathrm{Pin}(V,\,-g)$. If $V=\mathbb{C}^m$, then this group is denoted by $\mathrm{Pin}_m(\mathbb{C})$. If $V=\mathbb{R}^{k+l}$ and $g$ of signature $(k,l)$, then one writes $\mathrm{Pin}(V,g)=\mathrm{Pin}_{k,l}$. A similar notation is used for the groups spin, see below.

### Spin Groups

The spin group $\mathrm{Spin}(V,g)=\mathrm{Pin}(V,g)\cap C\ell^0(V,g)$ is generated by products of all sequences of an even number of unit vectors. Since the algebras $C\ell^0(V,g)$ and $C\ell^0(V,\,-g)$ are isomorphic, so are the groups $\mathrm{Spin}(V,g)$ and $\mathrm{Spin}(V,\,-g)$. Since $\alpha(a)=a$ for $a\in\mathrm{Spin}(V,g)$, the twisted adjoint representation reduces to the adjoint representation and yields the exact sequence

$$1\to\mathbb{Z}_2\to\mathrm{Spin}(V,g)\xrightarrow{\mathrm{Ad}}\mathrm{SO(V,g)}\to 1 \qquad [29]$$

For $V=\mathbb{C}^m$, the spin group is denoted by $\mathrm{Spin}_m(\mathbb{C})$. Since $\mathrm{Spin}_m(\mathbb{C})\subset G_1(\beta)$, the bilinear form $B$ is invariant with respect to the action of this group.

### Spin$^0$ Groups

The connected component $\mathrm{Spin}^0(V,g)$ of the group $\mathrm{Spin}(V,g)$ coincides with $\mathrm{Spin}(V,g)$ if either the quadratic space $(V,g)$ is complex or real and $kl=0$. In signature $(k,l)$, the connect group $\mathrm{Spin}^0_{k,l}$ is generated in $C\ell^0_{k,l}$ by all products of the form $u_1\ldots u_{2_p}v_1\ldots v_{2_q}$ such that $u_i^2=-1$ and $v_j^2=1$. The connected groups $\mathrm{Spin}_{m,0}$ and $\mathrm{Spin}_{0,m}$ are isomorphic and denoted by $\mathrm{Spin}_m$. Since $\mathrm{Spin}^0_{k,l}\subset G_1(\beta)$, the Hermitian form $A$ and the bilinear form $B$ are invariant with respect to the action of this group. Moreover, for $k+l$ even, from [24] and [28] there follows the transformation law of multivectors formed from pairs of spinors,

$$A_{\mu_1\cdots\mu_p}(\gamma(a)s,\gamma(a)t) = A_{\nu_1\ldots\nu_p}(s,t)R^{\nu_1}_{\mu_1}(a^{-1})\ldots R^{\nu_p}_{\mu_p}(a^{-1})$$

Consider $\mathrm{Spin}^0(V,g)$ and assume that either $V$ is complex of dimension $\geqslant 2$ or real with $k$ or $l\geqslant 2$. Then there are two unit orthogonal vectors $e_1,e_2\in V$ such that $(e_1,e_2)^2=-1$. The vector $u(t)=e_1\cos t+\,e_2\sin t$ is obtained from $e_1$ by rotation in the plane span $\{e_1,e_2\}$ by the angle $t\in\mathbb{R}$. The curve $t\mapsto e_1u(t)$, $0\le t\le\pi$, connects the elements $1$ and $-1$ of $\mathrm{Spin}^0(V,g)$. Its image in $\mathrm{SO}^0(V,g)$, that is, the curve $t\mapsto\mathrm{Ad}(e_1u(t))$, $0\le t\le\pi$, is closed: $\mathrm{Ad}(1)=\mathrm{Ad}(-1)$. This fact is often expressed by saying that "a spinor undergoing a rotation by $2\pi$ changes sign." There is no homomorphism – not even a continuous map – $f:\mathrm{SO}^0(V,g)\to\mathrm{Spin}^0(V,g)$ such that $\mathrm{Ad}\circ f=\mathrm{id}$.

### Spin$^c$ Groups

For the purposes of physics, to describe charged fermions, and in the theory of the Seiberg–Witten invariants, one needs the Spin$^c$ groups that are spinorial extensions of the real orthogonal groups by the group $\mathrm{U}_1$ of "phase factors." Assume $V$ to be real and $g$ of signature $(k,l)$ so that the sequence [29] can be written as

$$1\to\mathbb{Z}_2\to\mathrm{Spin}_{k,l}\to\mathrm{SO}_{k,l}\to 1$$

Define the action of $\mathbb{Z}_2=\{1,\,-1\}$ in $\mathrm{Spin}_{k,l}\times\mathrm{U}_1$ so that $(-1)(a,z)=(-a,-z)$. The quotient $(\mathrm{Spin}_{k,l}\times\mathrm{U}_1)/\mathbb{Z}_2=\mathrm{Spin}^c_{k,l}$ yields the extensions

$$1\to\mathrm{U}_1\to\mathrm{Spin}^c_{k,l}\to\mathrm{SO}_{k,l}\to 1$$

and

$$1\to\mathrm{Spin}_{k,l}\to\mathrm{Spin}^c_{k,l}\to\mathrm{U}_1\to 1$$

For example, $\mathrm{Spin}_3=\mathrm{SU}_2$ and $\mathrm{Spin}^c_3=\mathrm{U}_2$.

### Spin Groups in Dimensions $\leqslant$6

The connected components of spin groups associated with orthogonal spaces of dimension $\leqslant 6$ are isomorphic to classical groups. They can be explicitly described starting from the following observations.

Consider the four-dimensional vector space (of twistors) $T$ over $K$, with a volume element $\mathrm{vol} \in \wedge^4 T$. The six-dimensional vector space $V = \wedge^2 T$ has a scalar product $g$ defined by $g(u,v)\mathrm{vol} = 2u \wedge v$ for $u, v \in V$. The quadratic form $g(u,u)$ is the Pfaffian, $\mathrm{Pf}(u)$. If $u \in V$ is represented by the corresponding isomorphism $T^* \to T$ and $a \in \mathrm{End}\, T$, then $\mathrm{Pf}(aua^*) = \det a \mathrm{Pf}(u)$. The last formula shows $\mathrm{Spin}^0(V,g) = \mathrm{SL}(T)$, so that $\mathrm{Spin}_6(\mathbb{C}) = \mathrm{SL}_4(\mathbb{C})$. For $K = \mathbb{R}$, the Pfaffian is of signature $(3,3)$, so that $\mathrm{Spin}^0_{3,3} = \mathrm{SL}_4(\mathbb{R})$. A non-null vector $v \in V$ defines a symplectic form on $T^*$. The five-dimensional vector space $v^\perp \subset V$ is invariant with respect to the symplectic group $\mathrm{Sp}(T^*, u) = \mathrm{Spin}^0(v^\perp, \mathrm{Pf}|v^\perp)$. This shows that $\mathrm{Spin}_5(\mathbb{C}) = \mathrm{Sp}_4(\mathbb{C})$ and $\mathrm{Spin}^0_{2,3} = \mathrm{Sp}_4(\mathbb{R})$. Spin groups for other signatures in real dimensions 6 and 5 are obtained by considering appropriate real subspaces of $\mathbb{C}^6$ and $\mathbb{C}^5$, respectively. For example, [6] is used to show that $\mathrm{Spin}^0_{1,5} = \mathrm{SL}_2(\mathbb{H})$.

Spin groups in dimensions 4 and lower are similarly obtained from the observation that det is a quadratic form on the four-dimensional space $K(2)$ and $C\ell^0(K(2), \det) = K(2) \oplus K(2)$.

Several spin groups are listed below.

*The complex spin groups*

$$\mathrm{Spin}_2(\mathbb{C}) = \mathbb{C}^\times, \qquad \mathrm{Spin}_3(\mathbb{C}) = \mathrm{SL}_2(\mathbb{C})$$
$$\mathrm{Spin}_4(\mathbb{C}) = \mathrm{SL}_2(\mathbb{C}) \times \mathrm{SL}_2(\mathbb{C})$$
$$\mathrm{Spin}_5(\mathbb{C}) = \mathrm{Sp}_4(\mathbb{C})$$
$$\mathrm{Spin}_6(\mathbb{C}) = \mathrm{SL}_4(\mathbb{C})$$

*The real, compact spin groups*

$$\mathrm{Spin}_2 = \mathrm{U}_1, \qquad \mathrm{Spin}_3 = \mathrm{SU}_2$$
$$\mathrm{Spin}_4 = \mathrm{SU}_2 \times \mathrm{SU}_2, \qquad \mathrm{Spin}_5 = \mathrm{Sp}_2(\mathbb{H})$$
$$\mathrm{Spin}_6 = \mathrm{SU}_4$$

*The groups* $\mathrm{Spin}^0_{k,l}$ *for* $1 \leqslant k \leqslant l$ *and* $k + l \leq 6$

$$\mathrm{Spin}^0_{1,1} = \mathbb{R}^\times, \qquad \mathrm{Spin}^0_{1,2} = \mathrm{SL}_2(\mathbb{R})$$
$$\mathrm{Spin}^0_{1,3} = \mathrm{SL}_2(\mathbb{C})$$
$$\mathrm{Spin}^0_{2,2} = \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$$
$$\mathrm{Spin}^0_{1,4} = \mathrm{Sp}_{1,1}(\mathbb{H})$$
$$\mathrm{Spin}^0_{2,3} = \mathrm{Sp}_4(\mathbb{R}), \qquad \mathrm{Spin}^0_{1,5} = \mathrm{SL}_2(\mathbb{H})$$
$$\mathrm{Spin}^0_{2,4} = \mathrm{SU}_{2,2}$$
$$\mathrm{Spin}^0_{3,3} = \mathrm{SL}_4(\mathbb{R})$$

*See also:* Dirac Operator and Dirac Field; Index Theorems; Relativistic Wave Equations Including Higher Spin Fields; Spinors and Spin Coefficients; Twistors.

## Further Reading


Adams JF (1981) Spin (8), triality, $F_4$ and all that. In: Hawking SW and Roček M (eds.) *Superspace and Supergravity*. Cambridge: Cambridge University Press.

Atiyah MF, Bott R, and Shapiro A (1964) Clifford modules. *Topology* 3(suppl. 1): 3–38.

Baez JC (2002) The octonions. *Bulletin of the American Mathematical Society* 39: 145–205.

Brauer R and Weyl H (1935) Spinors in *n* dimensions. *American Journal of Mathematics* 57: 425–449.

Budinich P and Trautman A (1988) *The Spinorial Chessboard*,-Trieste Notes in Physics. Berlin: Springer.

Cartan É (1938) *Théorie des spineurs*. Actualités Scientifiques et Industrielles, No. 643 et 701. Paris: Hermann (English transl.:*The Theory of Spinors*. Paris: Hermann, 1966).

Chevalley C (1954) *The Algebraic Theory of Spinors*. New York: Columbia University Press.

Clifford WK (1878) Applications of Grassmann's extensive algebra. *American Journal of Mathematics* 1: 350–358.

Clifford WK (1882) On the classification of geometric algebras. In: Tucker R (ed.) *Mathematical Papers by William Kingdon Clifford*, pp. 397–401. London: Macmillan.

Dirac PAM (1928) The quantum theory of the electron. *Proceedings of the Royal Society of London A* 117: 610–624.

Eckmann B (1942) Gruppentheoretische Beweis des Satzes von Hurwitz–Radon über die Komposition quadratischer Formen. *Commentarii Mathematici Helvetici* 15: 358–366.

Karoubi M (1968) Algèbres de Clifford et *K*-théorie. *Annales Scientifiques de l'École Normale Superieure 4ème sér* 1: 161–270.

Lipschitz RO (1886) *Untersuchungen über die Summen von Quadraten*. Berlin: Max Cohen und Sohn.

Lounesto P (2001) *Clifford Algebras and Spinors*, 2nd edn. London Math. Soc. Lecture Note Series, vol. 286. Cambridge: Cambridge University Press.

Pauli W (1927) Zur Quantenmechanik des magnetischen Elektrons. *Z. Physik* 43: 601–623.

Penrose R and MacCallum MAH (1973) Twistor theory: an approach to the quantisation of fields and space-time. *Physics Report* 6C(4): 241–316.

Porteous IR (1995) *Clifford Algebras and the Classical Groups*, Cambridge Studies in Advanced Mathematics, vol. 50. Cambridge: Cambridge University Press.

Postnikov MM (1986) *Lie groups and Lie algebras*. Mir: Moscow.

Sudbery A (1987) Division algebras (pseudo)orthogonal groups and spinors. *Journal of Physics* A17: 939–955.

Trautman A (1997) Clifford and the "square root" ideas. *Contemporary Mathematics* 203: 3–24.

Trautman A and Trautman K (1994) Generalized pure spinors. *Journal of Geometry and Physics* 15: 1–22.

Wall CTC (1963) Graded Brauer groups. *Journal für die Reine und Angewandte Mathematik* 213: 187–199.

# Cluster Expansion

**R Kotecký**, Charles University, Prague,
Czech Republic, and the University of Warwick, UK



## Introduction

The method of cluster expansions in statistical physics provides a systematic way of computing power series for thermodynamic potentials (logarithms of partition funtions) as well as correlations. It originated from the works of Mayer and others devoted to expansions for dilute gas.

## Mayer Expansion

Consider a system of interacting particles with Hamiltonian

$$H_N(\boldsymbol{p}_1,\ldots,\boldsymbol{p}_N,\boldsymbol{r}_1,\ldots,\boldsymbol{r}_N) = \sum_{i=1}^{N}\frac{\boldsymbol{p}_i^2}{2m} + \sum_{i,j=1}^{N}\Phi(\boldsymbol{r}_i-\boldsymbol{r}_j) \qquad [1]$$

where $\Phi$ is a *stable* and *regular* pair potential. Namely, we assume that there exists $B \geq 0$ such that

$$\sum_{i,j=1}^{N}\Phi(\boldsymbol{r}_i-\boldsymbol{r}_j) \geq -BN \qquad [2]$$

for all $N = 2, 3, \ldots$ and all $(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_N) \in \mathbb{R}^{3N}$, and that

$$C(\beta) = \int \left|e^{-\beta\Phi(\boldsymbol{r})} - 1\right| d^3\boldsymbol{r} < \infty \qquad [3]$$

for some $\beta > 0$ (and hence all $\beta > 0$). Basic thermodynamic quantities are given in terms of the *grand-canonical partition function*

$$Z(\beta,\lambda,V) = \sum_{N=0}^{\infty}\frac{z^N}{N!}\int_{\mathbb{R}^{3N}\times V^N} e^{-\beta H_N}\frac{\prod d^3\boldsymbol{p}_i\prod d^3\boldsymbol{r}_i}{h^{3N}} = \sum_{N=0}^{\infty}\frac{\lambda^N}{N!}\int_{V^N} e^{-\beta\sum_{i,j}\Phi(\boldsymbol{r}_i-\boldsymbol{r}_j)}\prod d^3\boldsymbol{r}_i \qquad [4]$$

In the second expression we absorbed the factor resulting from the integration over impulses into (configurational) activity $\lambda = (2\pi m/\beta h^2)^{3/2} z$. In particular, the *pressure* $p$ and the *density* $\rho$ are defined by the thermodynamic limits (with $V \to \infty$ in the sense of Van Hove)

$$p(\beta,\lambda) = \frac{1}{\beta}\lim_{V\to\infty}\frac{1}{|V|}\log Z(\beta,\lambda,V) \qquad [5]$$

and

$$\rho(\beta,\lambda) = \lim_{V\to\infty}\frac{1}{|V|}\lambda\frac{\partial}{\partial\lambda}\log Z(\beta,\lambda,V) \qquad [6]$$

*Mayer series* are the expansions of $p$ and $\rho$ in powers of $\lambda$:

$$\beta p(\beta,\lambda) = \sum_{n=1}^{\infty} b_n\lambda^n \qquad [7]$$

and

$$\rho(\beta,\lambda) = \sum_{n=1}^{\infty} n b_n\lambda^n \qquad [8]$$

Mayer's idea for a systematic computation of coefficients $b_n$ was based on a reformulation of partition function $Z(\beta,\lambda,V)$ in terms of *cluster integrals*. Introducing the function

$$f(\boldsymbol{r}) = e^{-\beta\Phi(\boldsymbol{r})} - 1 \qquad [9]$$

and using $\mathcal{G}[N]$ to denote the set of all graphs on $N$ vertices $\{1,\ldots,N\}$, we get

$$Z(\beta,\lambda,V) = \sum_{N=0}^{\infty}\frac{\lambda^N}{N!}\int_{V^N}\prod_{i,j=1}^{N}\left(1+f(\boldsymbol{r}_i-\boldsymbol{r}_j)\right)\prod d^3\boldsymbol{r}_i = \sum_{N=0}^{\infty}\frac{\lambda^N}{N}\sum_{g\in\mathcal{G}[N]} w(g) \qquad [10]$$

where

$$w(g) = \int_{V^N}\prod_{\{i,j\}\in g} f(\boldsymbol{r}_i-\boldsymbol{r}_j)\prod d^3\boldsymbol{r}_i \qquad [11]$$

Observing that the weight $w$ is multiplicative in connected components (clusters) $g_1,\ldots,g_k$ of the graph $g$,

$$w(g) = \prod_{\ell=1}^{k} w(g_\ell) \qquad [12]$$

we can rewrite

$$Z(\beta,\lambda,V) = \sum_{N=0}^{\infty}\frac{\lambda^N}{N!}\sum_{\{g_l\}}\prod_{g\in\mathcal{G}} w(g) \qquad [13]$$

with the sum running over all disjoint collections $\{g_l\}$ of connected graphs with vertices in $\{1,\ldots,N\}$. A straightforward exponential expansion can be used to show that, at least in the sense of formal power series,

$$\log Z(\beta,\lambda,V) = \sum_{n=1}^{\infty}\frac{\lambda^n}{n!}\sum_{g\in\mathcal{C}[n]} w(g) \qquad [14]$$

where $\mathcal{C}[n]$ is the set of all connected graphs on $n$ vertices. Using $b_n^{(V)}$ to denote the coefficients

$$b_n^{(V)} = \frac{1}{|V|}\frac{1}{n!}\sum_{g\in\mathcal{C}[n]} w(g) \qquad [15]$$

and observing that the limits $\lim_{V\to\infty}(1/|V|)w(g)$ of cluster integrals exist, we get $b_n = \lim_{V\to\infty} b_n^{(V)}$. The convergence of Mayer series can be controlled directly by combinatorial estimates on the coefficients $b_n^{(V)}$. As a result, the diameter of convergence of the series [7] and [8] can be proved to be at least $(C(\beta)e^{2\beta B+1})^{-1}$. A less direct proof is based on an employment of linear integral Kirkwood–Salsburg equations in a suitable Banach space of correlation functions.

Similar combinatorial methods are available also for evaluation of coefficients of the *virial expansion* of pressure in powers of gas density,

$$\beta p(\beta,\rho) = \sum_{n=1}^{\infty}\beta_n\rho^n \qquad [16]$$

obtained by inverting [8] (notice that $b_1 = 1$) and inserting it into [7]. One is getting $\beta_n = \lim_{V\to\infty}\beta_n^{(V)}$ with

$$\beta_n^{(V)} = \frac{1}{|V|}\frac{1}{n!}\sum_{g\in\mathcal{B}[n]} w(g) \qquad [17]$$

where $\mathcal{B}[n] \subset \mathcal{C}[n]$ is the set of all 2-connected graphs on $\{1,\ldots,n\}$; namely, those graphs that cannot be split into disjoint subgraphs by erasing one vertex (and all adjacent edges). The diameter of convergence of the virial expansion turns out to be no less than $(C(\beta)e(e^{2\beta B}+1))^{-1}$.

## Abstract Polymer Models

An application of the ideas of Mayer expansions to lattice models is based on a reformulation of the partition function in terms of a *polymer model*, a formulation akin to [13] above. Namely, the partition function is rewritten as a sum over collections of pairwise compatible geometric objects – polymers. Most often, the compatibility means simply their disjointness.

While the reformulation of "physical partition function" in terms of a polymer model (including the definition of compatibility) depends on particularities of a given lattice model and on the considered region of parameters – high-temperature, low-temperature, large external fields, etc. – the essence and results of cluster expansion may be conveniently formulated in terms of an *abstract polymer model*.

Let $G = (V, E)$ be any (possibly infinite) countable graph and suppose that a map $w: V \to \mathbb{C}$ is given. Vertices $v \in V$ are called *abstract polymers*, with two abstract polymers connected by an edge in the graph $G$ called *incompatible*. We shall refer to $w(v)$ as to the *weight* of the abstract polymer $v$. For any finite $W \subset V$, we consider the induced subgraph $G[W]$ of $G$ spanned by $W$ and define

$$Z_W(w) = \sum_{I\subset W}\prod_{v\in I} w(v) \qquad [18]$$

Here the sum runs over all collections $I$ of compatible abstract polymers – or, in other words, the sum is over all *independent sets* $I$ of vertices in $W$ (no two vertices in $I$ are connected by an edge).

The partition function $Z_W(w)$ is an entire function in $w = \{w(v)\}_{v\in W} \in \mathbb{C}^{|W|}$ and $Z_W(0) = 1$. Hence, it is nonvanishing in some neighborhood of the origin $w = 0$ and its logarithm is, on this neighbourhood, an analytic function yielding a convergent Taylor series

$$\log Z_W(w) = \sum_{X\in\mathcal{X}(W)} a_W(X)w^X \qquad [19]$$

Here, $\mathcal{X}(W)$ is the set of all multi-indices $X: W \to \{0\,1,\ldots\}$ and $w^X = \prod_v w(v)^{X(v)}$. Inspecting the formula for $a_W(X)$ in terms of corresponding derivatives of $\log Z_W(w)$, it is easy to show that the Taylor coefficients $a_W(X)$ actually do not depend on $W: a_W(X) = a_{\text{supp}}\, X(X)$, where $\text{supp}\, X = \{v \in V: X\,(v) \neq 0\}$. As a result, one is getting the existence of coefficients $a(X)$ such that

$$\log Z_W(w) = \sum_{X\in\mathcal{X}(W)} a(X)w^X \qquad [20]$$

for every finite $W \subset V$.

The coefficients $a(X)$ can be obtained explicitly. One can pass from [18] to [20] in a similar way as passing from [10] to [13]. The starting point is to replace the restriction to compatible collections of abstract polymers in the sum [18] by the factor $\prod_{v,v'\in W}(1+F(v,v'))$ with

$$F(v,v') = \begin{cases} 0 \text{ if } v \text{ and } v' \text{ are compatible} \\ -1 \text{ otherwise } (v \text{ and } v' \\ \quad \text{connected by an edge from } G) \end{cases} \qquad [21]$$

and to expand the product afterwards. The resulting formula is

$$a(X) = (X!)^{-1}\sum_{H\subset G(X)} (-1)^{|E(H)|} \qquad [22]$$

Here, $G(X)$ is the graph with $|X| = \sum|X(v)|$ vertices induced from $G[\text{supp}\, X]$ by replacing each of its vertices $v$ by the complete graph on $|X(v)|$ vertices and $X!$ is the multifactorial $X! = \prod_{v\in\text{supp}\, X} X(v)!$. The sum is over all connected subgraphs $H \subset G(X)$ spanned by the set of vertices of $G(X)$ and $|E(H)|$ is the number of edges of the graph $H$.

A useful property of the coefficients $a(X)$ is their *alternating sign*,

$$(-1)^{|X|+1}a(X) \geq 0 \quad [23]$$

More important than an explicit form of the coefficients $a(X)$ are the convergence criteria for the series [20]. One way to proceed is to find direct combinatorial bounds on the coefficients as expressed by [22]. While doing so, one has to take into account the cancelations arising in view of the presence of terms of opposite signs in [22]. Indeed, disregarding them would lead to a failure since, as it is easy to verify, the number of connected graphs on $|X|$ vertices is bounded from below by $2^{(|X|-1)(|X|-2)/2}$. An alternative approach is to prove the convergence of [20] on polydisks $\mathcal{D}_{W,\boldsymbol{R}} = \{w : |w(v)| \leq R(v) \text{ for } v \in W\}$ by induction in $|W|$, once a proper condition on the set of radii $\boldsymbol{R} = \{R(v); v \in V\}$ is formulated. The most natural for the inductive proof (leading in the same time to the strongest claim) turns out to be the Dobrushin condition:

*There exists a function* $r : V \to [0, 1)$ *such that*, *for each* $v \in V$

$$R(v) \leq r(v) \prod_{v' \in \mathcal{N}(v)} (1 - r(v')) \quad [24]$$

Here $\mathcal{N}(v)$ is the set of vertices $v' \in V$ adjacent in graph $G$ to the vertex $v$.

Using $\mathcal{X}$ to denote the set of all multi-indices $X : V \to \{0, 1, \ldots\}$ with finite $|X| = \sum |X(v)|$ and saying that $X \in \mathcal{X}$ is a cluster if the graph $G$(supp $X$) is connected, we can summarize the cluster expansion claim for an abstract polymer model in the following way:

**Theorem** (Cluster expansion). *There exists a function* $a : \mathcal{X} \to \mathbb{R}$ *that is nonvanishing only on* clusters, *so that for any sequence of diameters* $\boldsymbol{R}$ *satisfying the condition* [24] *with a sequence* $\{r(v)\}$, *the following holds true*:

(i) *For every finite* $W \subset V$, *and any contour weight* $w \in \mathcal{D}_{W,\boldsymbol{R}}$, *one has* $Z_W(w) \neq 0$ *and*

$$\log Z_W(w) = \sum_{X \in \mathcal{X}(W)} a(X) w^X$$

(ii) $\sum_{X \in \mathcal{X} : \operatorname{supp} X \ni v} |a(X)||w|^X \leq -\log(1 - r(v))$.

Notice that, we have got not only an absolute convergence of the Taylor series of $\log Z_W$ in the closed polydisk $\mathcal{D}_{W,\boldsymbol{R}}$, but also the bound (ii) (uniform in $W$) on the sum over all terms containing a fixed vertex $v$. Such a bound turns out to be very useful in applications of cluster expansions. It yields, eventually, bounds on various error terms, avoiding a need of an explicit evaluation of the number of clusters of "given size."

The restriction to compatible collections of polymers can be actually relaxed. Namely, replacing [25] by

$$Z_W(w) = \sum_{W' \subset W} \prod_{v \in W'} w(v) \prod_{v, v' \in W'} U(v, v') \quad [25]$$

with $U(v, v') \in [0, 1]$ (soft repulsive interaction), and the condition [24] by

$$R(v) \leq r(v) \prod_{v' \neq v} \frac{1 - r(v')}{1 - U(v, v') r(v')} \quad [26]$$

one can prove that the partition function $Z_W(w)$ does not vanish on the polydisk $\mathcal{D}_{W,\boldsymbol{R}}$ implying thus that the power series of $\log Z_W(w)$ converges absolutely on $\mathcal{D}_{W,\boldsymbol{R}}$.

Polymers that arise in typical applications are geometric objects endowed with a "support" in the considered lattice, say $\mathbb{Z}^d, d \geq 1$, and their weights satisfy the condition of translation invariance. Cluster expansions then yield an explicit power series for the *pressure* (resp. free energy) in the thermodynamic limit as well as its finite-volume approximation.

To formulate it for an abstract polymer model, we assume that for each $x \in \mathbb{Z}^d$, an isomorphism $\tau_x : G \to G$ is given and that with each abstract polymer $v \in V$ a finite set $\Lambda(v) \subset \mathbb{Z}^d$ is associated so that $\Lambda(\tau_x(v)) = \Lambda(v) + x$ for every $v \in V$ and every $x \in \mathbb{Z}^d$. For any finite $W \subset V$ and any multi-index $X$, let $\Lambda(W) = \cup_{v \in W} \Lambda(v)$ and $\Lambda(X) = \Lambda(\operatorname{supp}(X))$. On the other hand, for any finite $\Lambda \subset \mathbb{Z}^d$, let $W(\Lambda) = \{v \in V : \Lambda(v) \subset \Lambda\}$. Assuming also that the weight $w : V \to \mathbb{C}$ is translation invariant – that is, $w(v) = w(\tau_x(v))$ for every $v \in V$ and every $x \in \mathbb{Z}^d$ – we get an explicit expression for the "pressure" of abstract polymer model in the thermodynamic limit

$$p = \lim_{\Lambda \to \infty} \frac{1}{|\Lambda|} \log Z_{W(\Lambda)}(w) = \sum_{X : \Lambda(X) \ni 0} \frac{a(X) w^X}{|\Lambda(X)|} \quad [27]$$

In addition, the finite-volume approximation can be explicitly evaluated, yielding

$$\log Z_{W(\Lambda)}(w) = p|\Lambda| + \sum_{X : \Lambda(X) \cap \Lambda^c \neq \emptyset} a(X) w^X \frac{|\Lambda(X) \cap \Lambda|}{|\Lambda(X)|} \quad [28]$$

Using the claim (ii), the second term can be bounded by const. $|\partial \Lambda|$.

## Cluster Expansions for Lattice Models

There is a variety of applications of cluster expansions to lattice models. As noticed above, the first step is always to rewrite the model in terms of a polymer representation.

### High-Temperature Expansions

Let us illustrate this point in the simplest case of the Ising model. Its partition function in volume $\Lambda \subset \mathbb{Z}^d$, with free boundary conditions and vanishing external field, is

$$Z_\Lambda(\beta) = \sum_{\sigma_\Lambda} \exp\left\{ \sum_{\substack{x,y\in\Lambda \\ |x-y|=1}} \sigma_x\sigma_y \right\} \qquad [29]$$

Using the identity

$$e^{\beta\sigma_x\sigma_y} = \cosh\,\beta + \sigma_x\sigma_y \sinh\,\beta \qquad [30]$$

it can be rewritten in the form

$$Z_\Lambda(\beta) = 2^{|\Lambda|}(\cosh\,\beta)^{|B(\Lambda)|} \sum_B (\tanh\,\beta)^{|B|} \qquad [31]$$

Here, the sum runs over all subsets $B$ of the set $B(\Lambda)$ of all bonds in $\Lambda$ (pairs of nearest-neighbor sites from $\Lambda$) such that each site is contained in an even number of bonds from $B$. Using $\Lambda(B)$ to denote the set of sites contained in bonds from $B$, we say that $B_1, B_2 \subset B(\Lambda)$ are disjoint if $\Lambda(B_1) \cap \Lambda(B_2) = \emptyset$. Splitting now $B$ into a collection $\mathcal{B} = \{B_1, \ldots, B_k\}$ of its connected components called (high-temperature) *polymers* and using $\mathcal{B}(\Lambda)$ to denote the set of all polymers in $\Lambda$, we are getting

$$Z_\Lambda(\beta) = 2^{|\Lambda|}(\cosh\,\beta)^{|B(\Lambda)|} \sum_{\mathcal{B}\subset\mathcal{B}(\Lambda)} \prod_{B\in\mathcal{B}} (\tanh\,\beta)^{|B|} \qquad [32]$$

with the sum running over all collections $\mathcal{B}$ of mutually disjoint polymers. This expression is exactly of the form [18], once we define compatibility of polymers by their disjointness. Introducing the weights

$$w(B) = (\tanh\,\beta)^{|B|} \qquad [33]$$

and taking the set $\mathcal{B}(\Lambda)$ of all polymers in $\Lambda$ for $W$, we get the *polymer representation* $Z_\Lambda(\beta) = 2^{|\Lambda|}(\cosh\,\beta)^{|B(\Lambda)|} Z_{\mathcal{B}(\Lambda)}(w)$.

To apply the cluster expansion theorem, we have to find a function $r$ such that the right-hand side of [24] is positive and yields thus the radius of a polydisk of convergence. Taking $r(B) = \epsilon^{|B|}$ with a suitable $\epsilon$, we get

$$\prod_{B'\in N(B)} (1 - r(B')) \geq e^{-2|B|} \qquad [34]$$

allowing to choose $R(B) = r(B)e^{-2|B|} = (\epsilon e^{-2})^{|B|}$. Indeed, to verify [34] we just notice that the number of polymers of size $n$ containing a fixed site is bounded by $\kappa^n$ with a suitable constant $\kappa$. Thus,

$$\sum_{B':\,\Lambda(B')\ni x} \epsilon^{|B'|} \leq \sum_{n=1}^{\infty} \kappa^n \epsilon^n \leq 1 \qquad [35]$$

once $\epsilon$ is sufficiently small, and thus

$$\sum_{B'\in N(B)} \epsilon^{|B|} \leq |\Lambda(B)| \leq |B| \qquad [36]$$

yielding [34] ($1 - t > e^{-2t}$ for $t < 1/2$). To have $w \in \mathcal{D}_{W,\boldsymbol{R}}$ (for any $W$) is, for $R(B) = (\epsilon e^{-2})^{|B|}$, sufficient to take $\beta \leq \beta_0$ with $\tanh\beta_0 = \epsilon e^{-2}$.

As a consequence, for $\beta \leq \beta_0$ we can use the cluster expansion theorem to obtain a convergent power series in powers of $\tanh\,\beta$. In particular, using $\Lambda(X) = \cup_{B\in\text{supp}X}\,\Lambda(B)$, we get the pressure by the explicit formula

$$\beta p(\beta) = \log\,2 + d\,\log(\cosh\,\beta) + \sum_{X:\Lambda(X)\ni x} \frac{a(X)}{|\Lambda(X)|} w^X \qquad [37]$$

for any fixed $x \in \mathbb{Z}^d$ (by translation invariance of the contributing terms, the choice of $x$ is irrelevant). The function $\beta p(\beta)$ is analytic on the region $\beta \leq \beta_0$ since it is obtained as a uniformly absolutely convergent series of analytic terms $(\tanh\,\beta)^{|X|}$.

This type of *high-temperature cluster expansion* can be extended to a large class of models with Boltzmann factor in the form $\exp\{-\beta\sum_A U_A(\phi)\}$, where $\phi = (\phi_x; x \in \mathbb{Z}^d)$ is the configuration with *a priori* on-site probability distribution $\nu(\mathrm{d}\phi_x)$ and $U_A$, for any finite $A \subset \mathbb{Z}^d$, are the multi-site interactions (depending only on $(\phi_x; x \in A)$). Using the Mayer trick we can rewrite

$$\exp\left\{-\beta \sum_{A\subset\Lambda} U_A(\phi)\right\} = \prod_A (1 + f_A(\phi)) \qquad [38]$$

with $f_A(\phi) = \exp\{-\beta U_A(\phi)\} - 1$. Expanding the product we will get a polymer representation with polymers $\mathcal{A}$ consisting of connected collections $\mathcal{A} = (A_1, \ldots, A_k)$ with weights

$$w(\mathcal{A}) = \int \prod_{A\in\mathcal{A}} f_A(\phi) \prod_{x\in\cup_{A\in\mathcal{A}}A} \nu(\mathrm{d}\phi_x) \qquad [39]$$

under appropriate bounds on the interactions $U_A$ and for $\beta$ small enough, using $\Lambda(\mathcal{A})$ to denote the set $\cup_{A\in\mathcal{A}}A$, we get,

$$\sum_{\mathcal{A}:\Lambda(\mathcal{A})\ni x} |w(\mathcal{A})| \leq 1 \qquad [40]$$

This assumption allows, as before in the case of the high-temperature Ising model, to apply the cluster expansion theorem yielding an explicit series expansion for the pressure.

### Correlations

Cluster expansions can be applied for evaluation of decay of correlations. Let us consider, for the class of models discussed above, the expectation

$$\langle\Psi\rangle_\Lambda = \frac{1}{Z_\Lambda} \int \Psi(\phi)\, e^{-\beta H_\Lambda(\phi)} \prod_{x\in\Lambda} \nu(\mathrm{d}\phi_x) \qquad [41]$$

with $H_\Lambda(\phi) = \sum_{A\subset\Lambda} U_A(\phi)$ and a function $\Psi$ depending only on variables $\phi_x$ on sites $x$ from a finite set $S \subset \Lambda \subset \mathbb{Z}^d$.

A convenient way of evaluating the expectation starts with introduction of the modified partition function

$$Z_{\Lambda,\Psi}(\alpha) = Z_\Lambda + \alpha Z_{\Lambda,\Psi} = Z_\Lambda(1 + \alpha\langle\Psi\rangle_\Lambda) \quad [42]$$

Clearly,

$$\langle\Psi\rangle_\Lambda = \left.\frac{\mathrm{d}\log Z_{\Lambda,\Psi}(\alpha)}{\mathrm{d}\alpha}\right|_{\alpha=0} \quad [43]$$

Thus, one may get an expression for the expectation $\langle\Psi\rangle_\Lambda$, by forming a polymer representation of $Z_{\Lambda,\Psi}(\alpha)$ and isolating terms linear in $\alpha$ in the corresponding cluster expansion. For the first step, in the just cited high-temperature case with general multi-site interactions, we first enlarge the original set $\mathcal{A}(\Lambda)$ of all polymers in $\Lambda$ (consisting of connected collections $\mathcal{A} = (A_1, \ldots, A_k)$) to $\mathcal{W}_S(\Lambda) = \mathcal{A}(\Lambda) \cup \mathcal{A}_S(\Lambda)$, where $\mathcal{A}_S(\Lambda)$ is the set of all collections $(\mathcal{A}_1, \ldots, \mathcal{A}_k)$ of polymers such that each of them intersects the set $S$ (polymers $(\mathcal{A}_1, \ldots, \mathcal{A}_k)$ are "glued" by $S$ into a single entity). Compatibility is defined as before by disjointness; in addition, any two collections from $\mathcal{A}_S(\Lambda)$ are declared to be incompatible as well as any polymer $\mathcal{A}$ from $\mathcal{A}(\Lambda)$ intersecting $S$ is considered to be incompatible with any collection from $\mathcal{A}_S(\Lambda)$. Defining now $w_\alpha(\mathcal{A}) = w(\mathcal{A})$ for $\mathcal{A} \in \mathcal{A}(\Lambda)$ and

$$w_\alpha(\mathcal{A}) = \alpha \int \Psi(\phi)\mathrm{e}^{-\beta H_\Lambda(\phi)} \prod_{x\in\cup_{A\in\mathcal{A}_1\cup\cdots\cup\mathcal{A}_k} A\cup S} \nu(\mathrm{d}\phi_x) \quad [44]$$

for $\mathcal{A} = (\mathcal{A}_1, \ldots, \mathcal{A}_k) \in \mathcal{A}_S(\Lambda)$, we get $Z_{\Lambda,\Psi}(\alpha)$ exactly in the form [18],

$$Z_{\Lambda,\Psi}(\alpha) = \sum_{\mathcal{I}\subset\mathcal{W}_S(\Lambda)} \prod_{\mathcal{A}\in\mathcal{I}} w_\alpha(\mathcal{A}) \quad [45]$$

As a result, we have

$$\log \mathcal{Z}_{\Lambda,\Psi}(\alpha) = \sum_{X\in\mathcal{X}(\mathcal{W}_S(\Lambda))} a(X) w_\alpha^X \quad [46]$$

allowing easily to isolate terms linear in $\alpha$: namely, the terms with multi-indices $X$ with $\operatorname{supp} X \cap \mathcal{A}_S(\Lambda)$ consisting of a single collection, say $\mathcal{A}_0$, that occurs with multiplicity one, $X(\mathcal{A}_0) = 1$. Explicitly, using

$$\mathcal{X}_{S,\mathcal{A}_0}(\Lambda) = \{X \in \mathcal{X}(\mathcal{W}_S(\Lambda)) : \operatorname{supp} X \cap \mathcal{A}_S(\Lambda) = \{\mathcal{A}_0\}, X(\mathcal{A}_0) = 1\} \quad [47]$$

we get

$$\langle\Psi\rangle_\Lambda = \sum_{\mathcal{A}_0\in\mathcal{A}_S(\Lambda)} \sum_{X\in\mathcal{X}_{S,\mathcal{A}_0}(\Lambda)} a(X) w^X \quad [48]$$

It is easy to show that, for sufficiently small $\beta$, the series on the right-hand side is absolutely convergent even if we extend $\mathcal{A}_S(\Lambda)$ to $\mathcal{A}_S = \cup_\Lambda \mathcal{A}_S(\Lambda)$ and $\mathcal{X}_{S,\mathcal{A}_0}(\Lambda)$ to $\mathcal{X}_{S,\mathcal{A}_0} = \cup_\Lambda \mathcal{X}_{S,\mathcal{A}_0}(\Lambda)$. As a result, we have an explicit expression for the limiting expectation $\langle\Psi\rangle$ in terms of an absolutely convergent power series. This can be immediately applied to show that $|\langle\Psi\rangle - \langle\Psi\rangle_\Lambda|$ decay exponentially in distance between $S$ and the complement of $\Lambda$. Indeed, it suffices to find a suitable bound on $\sum_X |\mathrm{a}(X)||w|^X$ with the sum running over all clusters $X$ reaching from the set $S$ to $\Lambda^c$. To this end one does not need to evaluate explicitly the number of clusters of given "diameter" $\operatorname{diam}(X) = \sum_{\mathcal{A}} X(\mathcal{A}) \operatorname{diam}(\Lambda(\mathcal{A})) = m$ with $m \geq \operatorname{dist}(S, \Lambda^c)$. The needed estimate is actually already contained in the condition (ii) from the cluster expansion theorem. It just suffices to choose a suitable $k$ and assume that $\beta$ is small enough to assure validity of (40) in a stronger form, $\sum_{\mathcal{A}:\Lambda(\mathcal{A})\ni x} |w(\mathcal{A})|K^{|\Lambda(\mathcal{A})|} \leq 1$, yielding eventually

$$\sum_{X:\operatorname{diam}(X)\geq\operatorname{dist}(S,\Lambda^c)} |a(X)||w|^X \leq K^{-\operatorname{dist}(S,\Lambda^c)} |S| \sum_{X:\cup_{\mathcal{A}\in\operatorname{supp}X}\Lambda(\mathcal{A})\ni x} |a(X)||w|^X K^{\sum X(\mathcal{A})|\Lambda(\mathcal{A})|} \leq |S| K^{-\operatorname{dist}(S,\Lambda^c)} \quad [49]$$

Exponential decay of correlations $\langle\Psi_1;\Psi_2\rangle_\Lambda = \langle\Psi_1\Psi_2\rangle_\Lambda - \langle\Psi_1\rangle_\Lambda\langle\Psi_2\rangle_\Lambda$ (and the limiting $\langle\Psi_1;\Psi_2\rangle$) in distance between supports of $\Psi_1$ and $\Psi_2$ can be established in a similar way by isolating terms proportional to $\alpha_1\alpha_2$ in the cluster expansion of $\log Z_{\Lambda,\Psi_1,\Psi_2}(\alpha_1, \alpha_2)$ with

$$Z_{\Lambda,\Psi_1,\Psi_2}(\alpha_1,\alpha_2) = Z_\Lambda(1 + \alpha_1\langle\Psi_1\rangle_\Lambda + \alpha_2\langle\Psi_2\rangle_\Lambda + \alpha_1\alpha_2\langle\Psi_1\Psi_2\rangle_\Lambda) \quad [50]$$

The resulting claim can be readily generalized to one about the decay of the correlation $\langle\Psi_1;\ldots;\Psi_k\rangle$ in terms of the shortest tree connecting supports $S_1, \ldots, S_k$ of the functions $\Psi_1, \ldots, \Psi_k$.

### Low-Temperature Expansions

Finally, in some models with symmetries, we can apply cluster expansion also at low temperatures. Let us illustrate it again in the case of Ising model. This time, we take the partition function $Z_\Lambda^+(\beta)$ with plus boundary conditions. First, let us define for each nearest-neighbor bond $\langle x, y\rangle$ its dual as the $(d-1)$-dimensional closed unit hypercube orthogonal to the segment from $x$ to $y$ and bisecting it at its center. For a given configuration $\sigma_\Lambda$, we consider the boundary of the regions of constant spins consisting of the union $\partial(\sigma_\Lambda)$ of all hypercubes that are dual to nearest-neighbor bonds $\langle x, y\rangle$ for which $\sigma_x \neq \sigma_y$. The *contours corresponding to* $\sigma_\Lambda$ are now defined as the connected components of $\partial(\sigma_\Lambda)$. Notice that, under the fixed boundary condition, there is a one-to-one correspondence between configurations $\sigma_\Lambda$ and sets $\Gamma$ of mutually compatible (disconnected) contours in $\Lambda$.

Observing that the number of faces in $\partial(\sigma_\Lambda)$ is just the sum of the areas $|\gamma|$ of the contours $\gamma \in \Gamma$, we get the polymer representation

$$Z_\Lambda^+(\beta) = e^{\beta|E(\Lambda)|} \sum_\Gamma \exp\left(-\beta \sum_{\gamma\in\Gamma} |\gamma|\right) \qquad [51]$$

where the sum is over all collections of disjoint contours in $\Lambda$. Here $E(\Lambda)$ is the set of all bonds $\langle x, y\rangle$ with at least one endpoint $x, y$ in $\Lambda$.

The condition [24] with $r(\gamma) = \epsilon^\gamma$ yields a similar bound on the weights $w(\gamma) = e^{-\beta|\gamma|}$ as in the high-temperature expansion. To verify it, for $\beta$ sufficiently large, boils down to the evaluation of number of contours of size $n$ that contain a fixed site.

As a result, we can employ the cluster expansion theorem to get

$$\log Z_\Lambda^+(\beta) = \beta|E(\Lambda)| + \sum_{X:X\in\mathcal{X}(\mathcal{C}(\Lambda))} a(X)w^X \qquad [52]$$

with an explicit formula for the limit

$$\beta p(\beta) = \beta d + \sum_{X:A(X)\ni 0} \frac{a(X)}{|A(X)|} w^X \qquad [53]$$

Here, $A(X)$ is the set of sites attached to contours from supp $X$,

$$A(X) = \cup_{\gamma\in \text{supp}\, X} A(\gamma) \qquad [54]$$

with

$$A(\gamma) = \{x \in \mathbb{Z}^d \mid \text{such that } \text{dist}(x,\gamma) \le 1/2\} \qquad [55]$$

As a consequence of the fact that [53] is, for large $\beta$, an absolutely convergent sum of analytic terms $a(X)w^X = a(X)e^{-\beta\sum_\gamma X(\gamma)|\gamma|}$ (considered as functions of $\beta$), the function $\beta p(\beta)$ is, for large $\beta$, analytic in $\beta$.

The fact that one can explicitly express the difference $\log Z_\Lambda^+(\beta) - |\Lambda|\beta p(\beta)$ (cf. [28]) found numerous applications in situations where one needs an accurate evaluation of the influence of the boundary of the region $\Lambda$ on the partition function. One such example is a study of microscopic behavior of interfaces. The main idea is to use the explicit expression in the form

$$\begin{aligned} Z_\Lambda^+(\beta) &= \exp\{\beta p(\beta)|\Lambda|\} \exp\left\{\sum_{X:A(X)\cap\Lambda^c\neq\emptyset} a(X)w^X \frac{|A(X)\cap\Lambda|}{|A(X)|}\right\} \\ &= \exp\{\beta p(\beta)|\Lambda|\} \prod_{X:A(X)\cap\Lambda^c\neq\emptyset} (1+f_X) \end{aligned} \qquad [56]$$

Noticing that

$$f_X = \exp\left\{a(X)w^X \frac{|A(X)\cap\Lambda|}{|A(X)|}\right\} - 1$$

does not vanish only if $A(X)\cap\Lambda \neq \emptyset$, we can expand the product to obtain "decorations" of the boundary $\partial\Lambda$ by clusters $f_X$. In the case of interface these clusters can be incorporated into the weight of interface, while on a fixed boundary they yield a "wall free energy."

The possibility of the (low-temperature) polymer representation of the partition function in terms of contours is based on the $+ \leftrightarrow -$ symmetry of the Ising model. In absence of such a symmetry, cluster expansions can still be used, but in the framework of Pirogov–Sinai theory (*see* Pirogov–Sinai Theory).

## Bibliographical Notes

Cluster expansions originated from the works of Ursell, Yvon, Mayer, and others and were first studied in terms of formal power series. The combinatorial and enumeration problems considered in this framework were summarized in Uhlenbeck and Ford (1962). For related topics in modern language, see Bergeron *et al.* (1998). The convergence results for Mayer and virial expansions for dilute gas were first proved in the works of Penrose, Lebowitz, Groenveld, and Ruelle (see Ruelle (1969) for a detailed survey). General polymer models on lattice were discussed by Gruber and Kunz (1971) (see also Simon (1993) for discussion of high-temperature and low-temperature cluster expansions of lattice models). Abstract polymer models were introduced in Kotecký and Preiss (1986). An elegant proof of a general claim presented by Dobrushin (1996) was further extended and summarized by Scott and Sokal (2005). We follow their reformulation of the Dobrushin condition. Cluster expansions with a view on applications in quantum field theory are reviewed in Brydges (1986).

*See also:* Phase Transitions in Continuous Systems; Pirogov–Sinai Theory; Wulff Droplets.

## Further Reading

Bergeron F, Labelle G, and Leroux P (1998) *Combinatorial Species and Tree-Like Structures*, Coll. Encyclopaedia of Mathematics and Its Applications, vol. 67. Cambridge, MA: Cambridge University Press.

Brydges DC (1986) A short course on cluster expansions. In: Osterwalder K and Stora R (eds.) *Critical Phenomena, Random Systems, Gauge Theories*, pp. 129–183. Les Houches, Session XLIII, 1984. Amsterdam/New York: Elsevier.

Dobrushin RL (1996) Estimates of semi-invariants for the Ising model at low temperatures. In: Dobrushin RL, Minlos RA, Shukin MA, and Vershik AM (eds.) *Topics in Statistical and Theoretical Physics*, pp. 59–81. Providence, RI: American Mathematical Society.

Gruber C and Kunz H (1971) General properties of polymer systems. *Communications Mathematical Physics* 22: 133–161.

Kotecký R and Preiss D (1986) Cluster expansion for abstract polymer models. *Communications in Mathematical Physics* 103: 491–498.

Ruelle D (1969) *Statistical Mechanics: Rigorous Results*, The Mathematical Physics Monograph Series. Reading, MA: Benjamin.

Scott AD and Sokal AD (2005) The repulsive lattice gas, the independent-set polynomial, and the Lovász local lemma. *Journal of Statistical Physics* 118: 1151–1261.

Simon B (1993) *The Statistical Mechanics of Lattice Gases*, Princeton Series in Physics, vol. 1. Princeton: Princeton University Press.

Uhlenbeck GE and Ford GW (1962) The theory of linear graphs with applications to the theory of the virial development of the properties of gases. In: de Boer J and Uhlenbeck GE (eds.) *Studies in Statistical Mechanics*, vol. I, Amsterdam: North-Holland.

# Coherent States

**S T Ali**, Concordia University, Montreal, QC, Canada



## Introduction

Very generally, a family of coherent states is a set of continuously labeled quantum states, with specific mathematical and physical properties, in terms of which arbitrary quantum states can be expressed as linear superpositions. Since coherent states are continuously labeled, they form overcomplete sets of vectors in the Hilbert space of states. Originally these states were introduced into physics by Schrödinger (1926), as a family of quantum states in terms of which the transition from quantum to classical mechanics could be conveniently studied. These states have the minimal uncertainty property, in the sense that they saturate the Heisenberg uncertainty relations. The name coherent state was applied when these states were rediscovered in the context of quantum optical radiation by Glauber, Klauder, and Sudarshan. It was demonstrated that in these states the correlation functions of the quantum optical field factorize as they do in classical optics, so that the optical field has a near-classical behavior, with the optical beam being coherent. In this article, we shall refer to these originally studied coherent states as canonical coherent states (CCS).

The canonical coherent states, apart from their use in quantum optics, have also been found to be extremely useful in computations in atomic and molecular physics, in quantum statistical mechanics, and in certain areas of mathematics and mathematical physics, including harmonic analysis, symplectic geometry, and quantization theory. Their wide applicability has prompted the search for other families of states sharing similar mathematical and physical properties. These other families of states are usually called generalized coherent states, even when there is no link to optical coherence in such studies.

## Some Properties of CCS

In addition to the minimal uncertainty property, the canonical coherent states have a number of analytical and group-theoretical properties which are taken as starting points in looking for generalizations. We now define the canonical coherent states mathematically and enumerate a few of these properties.

Suppose that the vectors $|0\rangle, |1\rangle, \ldots, |n\rangle, \ldots,$ correspond to quantum states of $0, 1, \ldots, n, \ldots,$ excitons, respectively. The Hilbert space of these states, in which they form an orthonormal basis, is often known as Fock space. The canonical coherent states are then defined in terms of this basis, for each complex number $z$, by the analytic expansion:

$$|z\rangle = \mathrm{e}^{-|z|^2/2} \sum_{n=0}^{\infty} \frac{z^n}{\sqrt{n!}} |n\rangle \qquad [1]$$

The states $|z\rangle$ are normalized to unity: $\langle z|z\rangle = 1$. They satisfy the formal eigenvalue equation

$$a|z\rangle = z|z\rangle \qquad [2]$$

where $a$ is the annihilation operator for excitons, which acts on the basis vectors (Fock states) $|n\rangle$ as follows:

$$a|n\rangle = \sqrt{n}|n-1\rangle \qquad [3]$$

Its adjoint $a^\dagger$ has the action

$$a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle \qquad [4]$$

and

$$[a, a^\dagger] = aa^\dagger - a^\dagger a = I \qquad [5]$$

$I$ being the identity operator on Fock space. Introducing the self-adjoint operators $Q$ and $P$, of position and momentum, respectively,

$$Q = \frac{a + a^\dagger}{\sqrt{2}}, \qquad P = \frac{a - a^\dagger}{\mathrm{i}\sqrt{2}} \qquad [6]$$

it is possible to demonstrate the minimal uncertainty property referred to above (we take $\hbar = 1$):

$$\langle \Delta Q\rangle\langle \Delta P\rangle = \tfrac{1}{2} \qquad [7]$$

where for any observable $A$,

$$\langle \Delta A\rangle = \left[\langle z|A^2|z\rangle - \langle z|A|z\rangle^2\right]^{1/2}$$

is its dispersion in the state $|z\rangle$.

One can also prove the resolution of the identity,

$$\int_{\mathbb{C}} |z\rangle\langle z| \frac{\mathrm{d}q\ \mathrm{d}p}{2\pi} = I \qquad [8]$$

where $z = (1/\sqrt{2})(q - \mathrm{i}p)$ has been written in terms of its real and imaginary parts $(1/\sqrt{2})q$ and $(1/\sqrt{2})p$, respectively. The above operator integral is to be understood in the weak sense, as will be explained later. Equation [8] incorporates the mathematical fact that the set of vectors $|z\rangle$ is overcomplete in the Hilbert space. Indeed, using [8] any vector $|\phi\rangle$ in the Hilbert space can be written as a linear (integral) superposition of these states:

$$|\phi\rangle = \int_{\mathbb{C}} \overline{\Psi(z)}|z\rangle \frac{\mathrm{d}q\ \mathrm{d}p}{2\pi}$$

where $\Psi$ is the component function, $\Psi(z) = \langle\phi|z\rangle$. Thus, the coherent states $|z\rangle$ form a continuously labeled total set of vectors in the Hilbert space and since this space is separable, they are an overcomplete set.

Analytic properties of the vectors $|z\rangle$ emerge when the scalar product $\langle\phi|z\rangle$ is taken with respect to an arbitrary vector $|\phi\rangle$ in Fock space. From [1] it is clear that

$$F(z) = \langle\phi|z\rangle = \mathrm{e}^{-|z|^2/2} f(z)$$

where $f$ is an entire analytic function in the complex variable $z$. Moreover, the mapping $\phi \mapsto f$ is an isometric embedding of the Fock space onto the Hilbert space of analytic functions, with respect to the norm

$$\|f\| = \left[\int_{\mathbb{C}} |f(z)|^2 \mathrm{d}\mu(z, \bar{z})\right]^{1/2} \qquad [9]$$

defined by the measure $\mathrm{d}\mu(z,\bar{z}) = (1/2\pi)\mathrm{e}^{-|z|^2}\,\mathrm{d}q\,\mathrm{d}p$.

Group-theoretical properties of the CCS can be demonstrated by noting that

$$|n\rangle = \frac{(a^\dagger)^n}{\sqrt{n!}}|0\rangle \text{ and } a|0\rangle = 0$$

using which [1] can be recast into the form

$$|z\rangle = \mathrm{e}^{-|z|^2/2}\mathrm{e}^{za^\dagger}|0\rangle = U(z)|0\rangle$$
$$U(z) = \mathrm{e}^{za^\dagger} - \bar{z}a \qquad [10]$$

The vectors $|z\rangle$ and the unitary operator $U(z)$ can be reexpressed in terms of the real variables $q$, $p$ and the operators Q, $P$ as

$$|z\rangle = |q,p\rangle = U(q,p)|0\rangle$$
$$U(q,p) = \mathrm{e}^{\mathrm{i}(pQ - qP)} \qquad [11]$$

The operators $U(q,p)$ realize a (projective) unitary, irreducible representation of the Weyl–Heisenberg group, which is the group whose Lie algebra has the generators $Q$, $P$, and $I$, obeying the commutation relations $[Q,P] = \mathrm{i}I$. The existence of the resolution of the identity [8] is the statement of the fact that this representation is square integrable (a notion which will be elaborated upon in the section "Some examples") which gives us the next paradigm for building coherent states, namely by the action, on a fixed vector, of the unitary operators of a square-integrable representation of a locally compact group.

The above range of properties, which are enjoyed by the CCS, cannot all be expected to hold when looking for generalizations. It then becomes necessary to adopt one or other of these properties as the starting point and to proceed from there. In so doing, it is best first to set down a general definition of coherent states, involving a minimal mathematical structure. Motivated more by possible applications to physics, we do this in the following section.

## General Definition

Let $\mathfrak{H}$ be an abstract, separable Hilbert space over the complexes, $X$ a locally compact space and $\mathrm{d}\nu$ a measure on $X$. Let $|x,i\rangle$ be a family of vectors in $\mathfrak{H}$, defined for each $x$ in $X$ and $i = 1, 2, 3, \ldots, N$, where $N$ is usually a finite integer, although it could also be infinite. We assume that this set of vectors possesses the following properties:

1. For each $i$, the mapping $x \mapsto |x,i\rangle$ is weakly continuous, that is, for each vector $|\phi\rangle$ in $\mathfrak{H}$, the function $\Psi_i(x) = \langle x,i|\phi\rangle$ is continuous (in the topology of $X$).
2. For each $x$ in $X$, the vectors $|x,i\rangle, i = 1, 2, \ldots, N$, are linearly independent.
3. The resolution of the identity

$$\sum_{i=1}^{N} \int_X |x,i\rangle\langle x,i| \mathrm{d}\nu(x) = I_{\mathfrak{H}} \qquad [12]$$

holds in the weak sense on the Hilbert space $\mathfrak{H}$, that is, for any two vectors $|\phi\rangle, |\psi\rangle$ in $\mathfrak{H}$, the following equality holds:

$$\sum_{i=1}^{N} \int_X \langle\phi|x,i\rangle\langle x,i|\psi\rangle \mathrm{d}\nu(x) = \langle\phi|\psi\rangle$$

A set of vectors $|x,i\rangle$ satisfying the above three properties is called a family of generalized vector coherent states. In case $N = 1$, the set is called a family of generalized coherent states. Sometimes the resolution of the identity condition is replaced by a weaker

condition, with the vectors $|x, i\rangle$ simply forming a total set in $\mathfrak{H}$ and the functions $F_i(x) = \langle x, i|\phi\rangle$, as $|\phi\rangle$ runs through $\mathfrak{H}$, forming a reproducing kernel Hilbert space. Alternatively, the identity on the right-hand side of [12] could also be replaced by a bounded, positive operator $T$ with bounded inverse. In this case, the term frame is also used for the family of generalized coherent states. For physical applications, however, the resolution of the identity condition is always assumed to hold, although the measure $d\nu$ could be of a very general nature (possibly also singular). The objective in all these cases is to ensure that an arbitrary vector $|\phi\rangle$ be expressible as a linear (integral) combination of these vectors. Indeed, [12] is immediately seen to imply that

$$|\phi\rangle = \sum_{i=1}^{N} \int_X \Psi_i(x)|x, i\rangle \mathrm{d}\nu(x) \qquad [13]$$

where $\Psi_i(x) = \langle x, i|\phi\rangle$.

Associated to a family of generalized coherent states on a Hilbert space $\mathfrak{H}$, there is an intrinsic isomorphism between this space and a Hilbert space of (in general, vector valued) continuous functions over $X$. Using this isomorphism, it is always possible to look upon coherent states as a family of continuous functions which are square integrable with respect to the measure $d\nu$. To demonstrate this, we note that, in view of [12], for each vector $|\phi\rangle$ in $\mathfrak{H}$, the vector-valued function $\boldsymbol{\Psi}(x)$ on $x$, with components $\Psi_i(x) = \langle x, i|\phi\rangle$, $i = 1, 2, \ldots, N$, satisfies the norm condition

$$\sum_{i=1}^{N} \int_X |\Psi_i(x)|^2 \mathrm{d}\nu(x) = \|\phi\|^2_{\mathfrak{H}}$$

This means that the set of vectors $\boldsymbol{\Psi}$, as $|\phi\rangle$ runs through $\mathfrak{H}$, is a closed subspace of the Hilbert space $L^2_{\mathbb{C}^N}(X, d\nu)$ of $N$-vector-valued functions on $x$. Let us denote this subspace by $\mathfrak{H}_K$ and note that this space is a reproducing kernel Hilbert space with a matrix-valued kernel $\boldsymbol{K}(x, y)$ having matrix elements

$$K(x, y)_{ij} = \langle x, i|y, j\rangle, \quad i, j = 1, 2, \ldots, N \qquad [14]$$

and enjoying the properties

$$K(x, y)_{ij} = \overline{K(y, x)_{ji}}, \quad K(x, x)_{ii} > 0 \qquad [15]$$

and

$$\sum_{\ell=1}^{N} \int_X K(x, z)_{i\ell} K(z, y)_{\ell j} \mathrm{d}\nu(z) = K(x, y)_{ij} \qquad [16]$$

If $\boldsymbol{e}^i$, $i = 1, 2, \ldots, N$, are the vectors constituting the canonical basis of $\mathbb{C}^N$, then for each $x$ in $X$ and $i = 1, 2, \ldots, N$, the vector-valued function $\boldsymbol{\xi}^i_x$ on $X$, defined by $\boldsymbol{\xi}^i_x(y) = \boldsymbol{K}(y, x)\boldsymbol{e}^i$, is the image in $\mathfrak{H}_K$ of the generalized vector coherent state $|x, i\rangle$, under the above-mentioned isometry. The vectors $\boldsymbol{\xi}^i_x$ span the space $\mathfrak{H}_K$ and for an arbitrary element $\boldsymbol{\Psi}$ of this Hilbert space, the reproducing property [16] of the kernel implies the relation

$$\int_X \boldsymbol{K}(x, y)\boldsymbol{\Psi}(y)\mathrm{d}\nu(y) = \boldsymbol{\Psi}(x) \qquad [17]$$

Conversely, given any reproducing kernel Hilbert space, with a kernel satisfying the relations [15] and [16], generalized coherent states can be constructed as above in terms of this kernel. Mathematically, therefore, generalized coherent states are just the set of vectors naturally defined by the kernel in a reproducing kernel Hilbert space.

## Some Examples

We present in this section some of the more commonly used types of coherent states, as illustrations of the general structure given above.

A large class of generalizations of the canonical coherent states [1] is obtained by a simple modification of their analytic structure. Let $x_1 \leq x_2 \leq \cdots \leq x_n \leq \cdots$ be an infinite sequence of positive numbers ($x_1 \neq 0$). Define $x_n! = x_1 x_2 \cdots x_n$ and by convention set $x_0! = 1$. In the same Fock space in which the CCS were described, we now define the related deformed or nonlinear coherent states via the analytic expansion

$$|z\rangle = \mathcal{N}(|z|^2)^{-1/2} \sum_{n=0}^{\infty} \frac{z^n}{\sqrt{x_n!}} |n\rangle \qquad [18]$$

The normalization factor $\mathcal{N}(|z|^2)$ is chosen so that $\langle z\,|\,z\rangle = 1$. These generalized coherent states are overcomplete in the Fock space and satisfy a resolution of the identity of the type

$$\int_{\mathcal{D}} |z\rangle\langle z|\mathcal{N}(|z|^2)\mathrm{d}\nu(z, \overline{z}) = I \qquad [19]$$

$\mathcal{D}$ being an open disk in the complex plane of radius $L$, the radius of convergence of the series $\sum_{n=0}^{\infty}(z^n/\sqrt{x_n!})$. (In the case of the CCS, $L = \infty$.) The measure $d\nu$ is generically of the form $\mathrm{d}\theta\, \mathrm{d}\lambda(r)$ (for $z = r e^{i\theta}$), where $d\lambda$ is related to the $x_n!$ through the moment condition

$$\frac{x_n!}{2\pi} = \int_0^L r^{2n}\, \mathrm{d}\lambda(r), \quad n = 0, 1, 2, \ldots \qquad [20]$$

This means that once the quantities $x_n!$ are specified, the measure $d\lambda$ is to be determined by solving the

moment problem [20], which of course may not always have a solution. This puts a constraint on the type of sequences $\{x_n\}$ which may be used in the construction.

Once again, we see that for an arbitrary vector $|\phi\rangle$ in the Fock space, the function $F(z) = \langle\phi\,|\,z\rangle$, of the complex variable $z$, is of the form $F(z) = \mathcal{N}(|z|^2)^{-1/2} f(z)$, where $f$ is an analytic function on the domain $\mathcal{D}$. The reproducing kernel associated to these coherent states is

$$\begin{aligned} K(\bar{z}, z') &= \langle z|z'\rangle \\ &= \left[\mathcal{N}(|z|^2)\mathcal{N}(|z'|^2)\right]^{-1/2} \sum_{n=0}^{\infty} \frac{(\bar{z}z')^n}{x_n!} \end{aligned} \qquad [21]$$

By analogy with [2], one can define a generalized annihilation operator $A$ by its action on the vectors $|z\rangle$,

$$A|z\rangle = z|z\rangle \qquad [22]$$

and its adjoint operator $A^\dagger$. These act on the Fock states $|n\rangle$ as follows:

$$\begin{aligned} A|n\rangle &= \sqrt{x_n}|n-1\rangle \\ A^\dagger|n\rangle &= \sqrt{x_{n+1}}|n+1\rangle \end{aligned} \qquad [23]$$

Depending on the exact values of the quantities $x_n$, these two operators, together with the identity $I$ and all their commutators, could generate a wide range of algebras including various deformed quantum algebras. The term nonlinear, as often applied to these generalized coherent states, comes again from quantum optics, where many such families of states are used in studying the interaction between the radiation field and atoms, and the strength of the interaction itself depends on the frequency of radiation. Of course, these coherent states will not in general have either the group-theoretical or the minimal uncertainty properties of the CCS.

The following is an example of generalized coherent states of the above type, built over the unit disk, $\mathcal{D} = \{z \in \mathbb{C} \,|\, |z| < 1\}$: on the Fock space, we define the states

$$|z\rangle = (1-r^2)^\kappa \sum_{n=0}^{\infty} \left[\frac{(2\kappa)_n}{n!}\right]^{1/2} z^n|n\rangle \quad r = |z| \qquad [24]$$

where $\kappa = 1,\ 3/2,\ 2,\ 5/2, \ldots$, and

$$\begin{aligned} (a)_m &= \frac{\Gamma(a+m)}{\Gamma(a)} \\ &= a(a+1)(a+2)\cdots(a+m-1) \end{aligned}$$

Comparing [24] with [18] we see that $x_n = n/(2\kappa + n - 1)$ so that $\lim_{n\to\infty} x_n = 1$. Thus, the infinite sum is convergent for any $z$ lying in the unit disk. These generalized coherent states arise from representations of the group SU(1, 1) belonging to the discrete series, each irreducible representation being labeled by a specific value of the index $\kappa$. The associated Hilbert space of functions, analytic on the unit disk, is a subspace of $L^2(\mathcal{D}, \mathrm{d}\mu_\kappa)$, with

$$\begin{aligned} \mathrm{d}\mu_\kappa(z, \bar{z}) &= (2\kappa - 1)\frac{(1-r^2)^{2\kappa-2}}{\pi} r\, \mathrm{d}r\, \mathrm{d}\theta \\ z &= r e^{\mathrm{i}\theta} \end{aligned}$$

which can be obtained by solving the moment problem [20]. The resolution of the identity satisfied by these states is

$$\frac{2\kappa - 1}{\pi} \int_{\mathcal{D}} |z\rangle\langle z| \frac{r\,\mathrm{d}r\,\mathrm{d}\theta}{(1-r^2)^2} = I \qquad [25]$$

The associated generalized creation and annihilation operators are

$$\begin{aligned} A|n\rangle &= \sqrt{\frac{n}{2\kappa + n - 1}}|n-1\rangle \\ A^\dagger|n\rangle &= \sqrt{\frac{n+1}{2\kappa + n}}|n+1\rangle \end{aligned} \qquad [26]$$

so that, clearly, $[A, A^\dagger] \neq I$.

Operators $A$ and $A^\dagger$ of the general type defined in [23] are also known as ladder operators. When such operators appear as generators of representations of Lie algebras, their eigenvectors (see [22]) are usually called Barut–Girardello coherent states. As an example, the representation of the Lie algebra of SU(1,1) on the Fock space is generated by the three operators $K_+, K_-$, and $K_3$, which satisfy the commutation relations

$$[K_3, K_\pm] = \pm K_\pm, \quad [K_-, K_+] = 2K_3 \qquad [27]$$

They act on the vectors $|n\rangle$ as follows:

$$\begin{aligned} K_-|n\rangle &= \sqrt{n(2\kappa + n - 1)}|n-1\rangle \\ K_+ &= K_-^\dagger \\ K_3|n\rangle &= (\kappa + n)|n\rangle \end{aligned} \qquad [28]$$

Thus, $K_-|0\rangle = 0$ and

$$|n\rangle = \frac{1}{\sqrt{n!(2\kappa)_n}} K_+^n|0\rangle$$

The Barut–Girardello coherent states $|z\rangle$ are now defined as the formal eigenvectors of the ladder operator $K_-$:

$$K_-|z\rangle = z|z\rangle, \quad z \in \mathbb{C} \qquad [29]$$

They have the analytic form

$$|z\rangle = \frac{|z|^{2\kappa-1}}{\sqrt{I_{2\kappa-1}(2|z|)}} \sum_{n=0}^{\infty} \frac{z^n}{\sqrt{n!(2\kappa + n - 1)!}}|n\rangle \qquad [30]$$

where $I_\nu(x)$ is the order-$\nu$ modified Bessel function of the first kind. These coherent states satisfy the resolution of the identity,

$$\frac{2}{\pi}\int_{\mathbb{C}} |z\rangle\langle z| K_{2\kappa-1}(2r) I_{2\kappa-1}(2r) r\, \mathrm{d}r\, \mathrm{d}\theta = I \qquad z = r\mathrm{e}^{\mathrm{i}\theta} \qquad [31]$$

where again, $K_\nu(x)$ is the order-$\nu$ modified Bessel function of the second kind.

A nonanalytic extension of the expression [18] is often used to define generalized coherent states associated to physical Hamiltonians having pure point spectra. These coherent states, known as Gazeau–Klauder coherent states, are labeled by action–angle variables. Suppose that we are given the physical Hamiltonian $H = \sum_{n=0}^{\infty} E_n |n\rangle\langle n|$, with $E_0 = 0$, that is, it has the energy eigenvalues $E_n$ and eigenvectors $|n\rangle$, which we assume to form an orthonormal basis for the Hilbert space of states $\mathfrak{H}$. Let us write the eigenvalues as $E_n = \omega\epsilon_n$ by introducing a sequence of dimensionless quantities $\{\epsilon_n\}$ ordered as: $0 = \epsilon_0 < \epsilon_1 < \epsilon_2 < \cdots$. Then, for all $J \geq 0$ and $\gamma \in \mathbb{R}$, the Gazeau–Klauder coherent states are defined as

$$|J,\gamma\rangle = \mathcal{N}(J)^{-1/2} \sum_{k=0}^{\infty} \frac{J^{n/2}\mathrm{e}^{-\mathrm{i}\epsilon_n\gamma}}{\sqrt{\epsilon_n!}} |n\rangle \qquad [32]$$

where again $\mathcal{N}$ is a normalization factor, which turns out to be dependent on $J$ only. These coherent states satisfy the temporal stability condition

$$\mathrm{e}^{-\mathrm{i}Ht}|J,\gamma\rangle = |J,\gamma+\omega t\rangle \qquad [33]$$

and the action identity

$$\langle J,\gamma|H|J,\gamma\rangle_{\mathfrak{H}} = \omega J \qquad [34]$$

While these generalized coherent states do form an overcomplete set in $\mathfrak{H}$, the resolution of the identity is generally not given by an integral relation of the type [12].

For the second set of examples of generalized coherent states, we take the group-theoretical structure of the CCS as the point of departure. Let $G$ be a locally compact group and suppose that it has a continuous, irreducible representation on a Hilbert space $\mathfrak{H}$ by unitary operators $U(g)$, $g \in G$. This representation is called square integrable if there exists a nonzero vector $|\psi\rangle$ in $\mathfrak{H}$ for which the integral

$$c(\psi) = \int_G |\langle\psi|U(g)\psi\rangle|^2 \,\mathrm{d}\mu(g) \qquad [35]$$

converges. Here $\mathrm{d}\mu$ is a Haar measure of $G$, which for definiteness, we take to be the left-invariant measure. (The value of the above integral is independent of whether the left- or the right-invariant measure is used, so we could just as well have used the right-invariant measure.) A vector $|\psi\rangle$, satisfying [35], is said to be admissible, and it can be shown that the existence of one such vector guarantees the existence of an entire dense set of such vectors in $\mathfrak{H}$. Moreover, if the group $G$ is unimodular, that is, if the left- and the right-invariant measures coincide, then the existence of one admissible vector implies that every vector in $\mathfrak{H}$ is admissible. Given a square-integrable representation and an admissible vector $|\psi\rangle$, let us define the vectors

$$|g\rangle = \frac{1}{\sqrt{c(\psi)}} U(g)|\psi\rangle \qquad [36]$$

for all $g$ in the group $G$. These vectors are to be seen as the analogs of the canonical coherent states [11], written there in terms of the representation of the Weyl–Heisenberg group. Next, it can be shown that the resolution of the identity

$$\int_G |g\rangle\langle g| \mathrm{d}\mu(g) = I_{\mathfrak{H}} \qquad [37]$$

holds on $\mathfrak{H}$. Thus, the vectors $|g\rangle$ constitute a family of generalized coherent states. The functions $F(g) = \langle g|\phi\rangle$ for all vectors $|\phi\rangle$ in $\mathfrak{H}$ are square integrable with respect to the measure $\mathrm{d}\mu$ and the set of such functions, which in fact are continuous in the topology of $G$, forms a closed subspace of $L^2(G, \mathrm{d}\mu)$. Furthermore, the mapping $\phi \mapsto F$ is a linear isometry between $\mathfrak{H}$ and $L^2(G, \mathrm{d}\mu)$ and under this isometry the representation U gets mapped to a subrepresentation of the left regular representation of $G$ on $L^2(G, \mathrm{d}\mu)$.

A typical example of the above construction is provided by the affine group, $G_{\mathrm{Aff}}$. This is the group of all $2 \times 2$ matrices of the type

$$g = \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \qquad [38]$$

$a$ and $b$ being real numbers with $a \neq 0$. We shall also write $g = (b, a)$. This group is nonunimodular, with the left-invariant measure being given by $\mathrm{d}\mu(b,a) = (1/a^2)\, \mathrm{d}b\, \mathrm{d}a$. (The right-invariant measure is $(1/a)\, \mathrm{d}b\, \mathrm{d}a$.) The affine group has a unitary irreducible representation on the Hilbert space $L^2(\mathbb{R}, dx)$. Vectors in $L^2(\mathbb{R}, \mathrm{d}x)$ are measurable functions $\phi(x)$ of the real variable $x$ and the (unitary) operators $U(b,a)$ of this representation act on them in the manner

$$(U(b,a)\phi)(x) = \frac{1}{\sqrt{|a|}}\phi\left(\frac{x-b}{a}\right) \qquad [39]$$

If $\psi$ is a function in $L^2(\mathbb{R}, dx)$ such that its Fourier transform $\widehat{\psi}$ satisfies the condition

$$\int_{\mathbb{R}} \frac{|\widehat{\psi}(k)|^2}{|k|}\, \mathrm{d}k < \infty \qquad [40]$$

then it can be shown to be an admissible vector, that is,

$$c(\psi) = \int_{G_{\mathrm{Aff}}} |\langle \psi | U(b,a)\psi\rangle|^2 \frac{\mathrm{d}b\ \mathrm{d}a}{a^2} < \infty$$

Thus, following the general construction outlined above, the vectors

$$|b,a\rangle = \frac{1}{\sqrt{c(\psi)}} U(b,a)\psi, \quad (b,a) \in G_{\mathrm{Aff}} \qquad [41]$$

define a family of generalized coherent states and one has the resolution of the identity

$$\int_{G_{\mathrm{Aff}}} |b,a\rangle\langle b,a| \frac{\mathrm{d}b\,\mathrm{d}a}{a^2} = I \qquad [42]$$

on $L^2(\mathbb{R}, \mathrm{d}x)$.

In the signal-analysis literature a vector satisfying the admissibility condition [40] is called a mother wavelet and the generalized coherent states [41] are called wavelets. Signals are then identified with vectors $|\phi\rangle$ in $L^2(\mathbb{R}, \mathrm{d}x)$ and the function

$$F(b,a) = \langle b,a|\phi\rangle \qquad [43]$$

is called the continuous wavelet transform of the signal $\phi$.

There exist alternative ways of constructing generalized coherent states using group representations. For example, the Perelomov method is based on the observation that the vector $|0\rangle$, appearing in the construction of the canonical coherent states in [10] and [11] using the representation of the Weyl–Heisenberg group, is invariant up to a phase, under the action of its center. Consequently, the coherent states $|z\rangle$, as written in [10], are labeled, not by elements of the group itself, but only by the points in the quotient space of the group by its (central) phase subgroup. Generally, let $G$ be a locally compact group and $U$ a unitary irreducible representation of it on the Hilbert space $\mathfrak{H}$. We do not assume $U$ to be square integrable. We fix a vector $|\psi\rangle$ in $\mathfrak{H}$, of unit norm and denote by $H$ the subgroup of $G$ consisting of all elements $h$ for which

$$U(h)|\psi\rangle = \mathrm{e}^{\mathrm{i}\omega(h)}|\psi\rangle \qquad [44]$$

where $\omega$ is a real-valued function of $h$. Let $X = G/H$ be the left-coset space and $x$ an arbitrary element in $X$. Choosing a coset representative $g(x) \in G$, for each coset $x$, we define the vectors

$$|x\rangle = U(g(x))|\psi\rangle \qquad [45]$$

in $\mathfrak{H}$. The dependence of these vectors on the specific choice of the coset representative $g(x)$, is only through a phase. Thus, if instead of $g(x)$ we took a different representative $g(x)' \in G$ for the same coset $x$, then since $g(x)' = g(x)h$ for some $h \in H$, in view of [44] we would have $U(g(x)')|\psi\rangle = \mathrm{e}^{\mathrm{i}\omega(h)}|x\rangle$. Hence, quantum mechanically, both $|x\rangle$ and $U(g(x)')|\psi\rangle$ represent the same physical state and in particular, the projection operator $|x\rangle\langle x|$ depends only on the coset. Vectors $|x\rangle$, defined in this manner, are called Gilmore–Perelomov coherent states. Since $U$ is assumed to be irreducible, the set of all these vectors as $x$ runs through $G/H$ is dense in $\mathfrak{H}$. In this definition of generalized coherent states, no resolution of the identity is postulated. However, if $X$ carries an invariant measure, under the natural action of $G$, and if the formal operator $B$ defined as

$$B = \int_X |x\rangle\langle x|\, \mathrm{d}\mu(x)$$

is bounded, then it is necessarily a multiple of the identity and a resolution of the identity is again retrieved.

The Perelomov construction can be used to define coherent states for any locally compact group. On the other hand, there exist other constructions of generalized coherent states, using group representations, which generalize the notion of square integrability to homogeneous spaces of the group. Briefly, in this approach one starts with a unitary irreducible representation $U$ and attempts to find a vector $|\psi\rangle$, a subgroup $H$ and a section $\sigma : G/H \longrightarrow G$ such that

$$\int_{G/H} |x\rangle\langle x|\, \mathrm{d}\mu(x) = T \qquad [46]$$

where $|x\rangle = U(\sigma(x))|\psi\rangle$, $T$ is a bounded, positive operator with bounded inverse and $\mathrm{d}\mu$ is a quasi-invariant measure on $X = G/H$. It is not assumed that $|\psi\rangle$ be invariant up to a phase under the action of $H$ and clearly, the best situation is when $T$ is a multiple of the identity. Although somewhat technical, this general construction is of enormous versatility for semidirect product groups of the type $\mathbb{R}^n \rtimes K$, where $K$ is a closed subgroup of $\mathrm{GL}(n, \mathbb{R})$. Thus, it is useful for many physically important groups, such as the Poincaré or the Euclidean group, which do not have square-integrable representations in the sense of the earlier definition (see eqn [35]). The integral condition [46] ensures that any vector $|\phi\rangle$ in $\mathfrak{H}$ can be written in terms of the $|x\rangle$. Indeed, it

is easy to see that one has the integral representation of a vector,

$$|\phi\rangle = \int_X \Psi(x)|x\rangle \, \mathrm{d}\mu(x)$$
$$\Psi(x) = \langle x|T^{-1}\phi\rangle$$

in terms of the generalized coherent states.

The canonical coherent states satisfy the minimal uncertainty relation [7]. It is possible to build families of coherent states by generalizing from this condition. To do this, one typically starts with two self-adjoint generators in the Lie algebra of a particular group representation and then looks for appropriate eigenvectors of a complex combination of these two generators. For two self-adjoint operators $B$ and $C$ on a Hilbert space $\mathfrak{H}$, satisfying the commutation relation $[B, C] = \mathrm{i}D$ and any normalized vector $\phi$ in $\mathfrak{H}$, one can prove the Heisenberg uncertainty relation

$$(\Delta B)^2(\Delta C)^2 \geq \frac{\langle D\rangle^2}{4} \qquad [47]$$

where $\langle X\rangle = \langle\phi|X\phi\rangle$ and $(\Delta X)^2 = \langle X^2\rangle - \langle X\rangle^2$, for any operator $X$ on $\mathfrak{H}$. More generally, one can prove the Schrödinger–Robertson uncertainty relation

$$(\Delta B)^2(\Delta C)^2 \geq \frac{1}{4}\left[\langle D\rangle^2 + \langle F\rangle^2\right] \qquad [48]$$

where $\langle F\rangle = \langle BC + CB\rangle - 2\langle B\rangle\langle C\rangle$ measures the correlation between $B$ and $C$ in the state $\phi$. If $\langle F\rangle = 0$, the above relation reduces to the Heisenberg uncertainty relation. On the other hand, if $\langle D\rangle = 0$, the Heisenberg uncertainty relations become redundant. Suppose now that $B$ and $C$ are two self-adjoint elements of the Lie algebra in the unitary irreducible representation of a Lie group and we look for states $|\phi\rangle$ which minimize the uncertainty relation [48], that is, for which the equality holds. It turns out that such states can be found by considering the linear combination $B + \mathrm{i}\lambda C$, for a fixed complex number $\lambda$, and solving the formal eigenvalue equation

$$[B + \mathrm{i}\lambda C]|z, \lambda\rangle = z|z, \lambda\rangle$$
$$\text{with} \quad z = \langle B\rangle + \mathrm{i}\lambda\langle C\rangle \qquad [49]$$

Solutions to this equation for which $|\lambda| = 1$ are called squeezed states, since in this case $\Delta B \neq \Delta C$. Generally, the states $|z, \lambda\rangle$ are known as intelligent states. As an example, for the operators $Q$ and $P$ in [6], for which one has

$$(\Delta Q)^2(\Delta P)^2 \geq \frac{1}{4}\left[1 + \langle F\rangle^2\right]$$

taking the combination $Q + \mathrm{i}\lambda P$, one obtains the minimal uncertainty states,

$$|z, \lambda\rangle = \mathcal{N}(z, \lambda)^{-1/2} \, \mathrm{e}^{-w(a^\dagger)^2/2} \, \mathrm{e}^{(z/\sqrt{2})(1+w)a^\dagger}|0\rangle \qquad [50]$$

$\mathcal{N}(z, \lambda)$ being a normalization constant and $w = (1-\lambda)/(1+\lambda)$. The case $\lambda = -1$ does not lead to any solutions, while $\lambda = 1$ gives the canonical coherent states [10]. For real $\lambda \neq 1$ the above states are the well-known squeezed states of quantum optics.

Our final example is that of a family of vector coherent states, which will be obtained essentially by replacing the complex variable $z$ in [18] by a matrix variable. We choose the domain $\Omega = \mathbb{C}^{2\times 2}$ (all $2\times 2$ complex matrices), equipped with the measure

$$\mathrm{d}\nu(\mathfrak{z}, \mathfrak{z}^\dagger) = \frac{\mathrm{e}^{-\mathrm{tr}[\mathfrak{z}\mathfrak{z}^\dagger]}}{\pi^4} \prod_{k,j=1}^{2} \mathrm{d}x_{kj} \wedge \mathrm{d}y_{kj}$$

where $\mathfrak{z}$ is an element of $\Omega$ and $z_{kj} = x_{kj} + \mathrm{i}y_{kj}$ are its entries. One can then prove the matrix orthogonality relation

$$\begin{aligned}\int_\Omega & \mathfrak{z}^k \mathfrak{z}^{\dagger\ell} \, \mathrm{d}\nu(\mathfrak{z}, \mathfrak{z}^\dagger) \\ &= \frac{1}{2}\int_\Omega \mathrm{tr}[\mathfrak{z}^k \mathfrak{z}^{\dagger\ell}] \, \mathrm{d}\nu(\mathfrak{z}, \mathfrak{z}^\dagger)\mathbb{I}_2 \\ &= b(k)\mathbb{I}_2, \quad k, \ell = 0, 1, 2, \ldots, \infty\end{aligned} \qquad [51]$$

$\mathbb{I}_2$ being the $2\times 2$ identity matrix and

$$b(k) = \frac{(k+3)!}{2(k+1)(k+2)}$$
$$k = 1, 2, 3, \ldots, \quad b(0) = 1 \qquad [52]$$

Consider the Hilbert space $\tilde{\mathfrak{H}} = L^2_{\mathbb{C}^2}(\Omega, \mathrm{d}\nu)$ of square integrable, two-component vector-valued functions on $\Omega$ and in it consider the vectors $|\boldsymbol{\Psi}^i_k\rangle$, $i = 1, 2$, $k = 0, 1, 2, \ldots, \infty$, defined by the $\mathbb{C}^2$-valued functions,

$$\boldsymbol{\Psi}^i_k(\mathfrak{z}^\dagger) = \frac{1}{\sqrt{b(k)}}\mathfrak{z}^{\dagger k}\chi^i \qquad [53]$$

where the vectors $\chi^i, i = 1, 2$, form an orthonormal basis of $\mathbb{C}^2$. By virtue of [51], the vectors $|\boldsymbol{\Psi}^i_k\rangle$ constitute an orthonormal set in $\tilde{\mathfrak{H}}$, that is,

$$\langle\boldsymbol{\Psi}^i_k|\boldsymbol{\Psi}^j_\ell\rangle_{\tilde{\mathfrak{H}}} = \delta_{k\ell}\,\delta_{ij}$$

Denote by $\mathfrak{H}_K$ the Hilbert subspace of $\tilde{\mathfrak{H}}$ generated by this set of vectors. This can be shown to be a reproducing kernel Hilbert space of analytic

functions in the variable $\mathfrak{z}^\dagger$, with the matrix valued kernel $K : \Omega \times \Omega \longmapsto C^{2\times 2}$:

$$K(\mathfrak{z}'^\dagger, \mathfrak{z}) = \sum_{i=1}^{2}\sum_{k=0}^{\infty} \mathbf{\Psi}^i_k(\mathfrak{z}'^\dagger)\mathbf{\Psi}^i_k(\mathfrak{z}^\dagger)^\dagger = \sum_{i=1}^{2}\sum_{k=0}^{\infty} \frac{\mathfrak{z}'^{\dagger k}\mathfrak{z}^k}{b(k)} \quad [54]$$

Vector coherent states in $\mathfrak{H}_K$ are then naturally associated to this kernel and are given by

$$|\mathfrak{z}, i\rangle = \sum_{j=1}^{2}\sum_{k=0}^{\infty} \frac{\chi^{j\dagger}\mathfrak{z}^k\chi^i}{\sqrt{b(k)}}|\mathbf{\Psi}^j_k\rangle$$
$$\text{that is,}\ \ |\mathfrak{z}, i\rangle(\mathfrak{z}'^\dagger) = K(\mathfrak{z}'^\dagger, \mathfrak{z})\chi^i \quad [55]$$

for $i = 1, 2$ and all $\mathfrak{z}$ in $\Omega$. They satisfy the resolution of the identity

$$\sum_{i=1}^{2}\int_\Omega |\mathfrak{z}, i\rangle\langle\mathfrak{z}, i| \mathrm{d}\nu(\mathfrak{z}, \mathfrak{z}^\dagger) = I_{\mathfrak{H}_K} \quad [56]$$

The expression for the $|\mathfrak{z}, i\rangle$ in [55], involving the sum, should be compared to [18], of which it is a direct analog.

## Some Applications of Coherent States

Generalized coherent states have many applications in physics, signal analysis, and mathematics, of which we mention a few here. As an example of an application of deformed coherent states, we take

$$x_n = \left[\frac{q^n - q^{-n}}{q - q^{-1}}\right]^{1/2}, \quad q > 0 \quad [57]$$

in the definition of these states in [18]. It is then easy to see that the operators $A$ and $A^\dagger$, defined in [23], satisfy the $q$-deformed commutation relation

$$AA^\dagger - qA^\dagger A = q^{-N} \quad [58]$$

where $N$ is the usual number operator, which acts on the Fock states as $N|n\rangle = n|n\rangle$. Clearly, in the limit as $q \to 1$, these $q$-deformed coherent states go over to the canonical coherent states, with the operators $A$ and $A^\dagger$ becoming the usual creation and annihilation operators $a$ and $a^\dagger$, respectively. The operators $A$ and $A^\dagger$ and the commutation relation [58] describe a system of $q$-deformed oscillators, which have been used to describe, for example, the vibrations of polyatomic molecules. The potential energy between the atoms of such a molecule has anharmonic terms, leading to a deformation of the usual oscillator algebra, generated by the operators $a$ and $a^\dagger$.

As already mentioned, generalized coherent states are widely used in signal analysis. The wavelet transform $F(b, a) = \langle b, a|\phi\rangle$, introduced in [43], is a time–frequency transform, in which the parameter $b$ is identified with time and $1/a$ with frequency. Wavelet transforms are used extensively to analyze, encode, and reconstruct signals arising in many different branches of physics, engineering, seismography, electronic data processing, etc. Similarly, the canonical coherent states, as written in [11], give rise to the transform $F(q, p) = \langle q, p \mid \phi\rangle$. Again, if $q$ is interpreted as time and $p$ as frequency, then this is just the windowed Fourier transform, also used extensively in signal processing. More general wavelets, from higher-dimensional affine groups, are used to analyze higher-dimensional signals, while wavelet like transforms from other groups have been used to study signals exhibiting different geometries. In particular, wavelet transforms from spherical geometries have been applied to the study of brain signals and to astrophysical data.

Our final example is taken from quantization theory. A quantization technique is a method for performing the transition from a given classical mechanical system to its quantum counterpart. Many methods have been developed to accomplish this and the use of coherent states is one of them. Suppose that we are given a family of coherent states $|x\rangle$ in a Hilbert space $\mathfrak{H}$, where the set $X$ from which $x$ is taken is a classical phase space. This means that $X$ is a symplectic manifold with an associated 2-form $\omega$, which defines a Poisson bracket on the set of observables of the classical system, which are real-valued functions on $X$. There is a natural measure $\mathrm{d}\omega$, defined on $X$ by the 2-form $\omega$. Let us assume that the coherent states $|x\rangle$ satisfy a resolution on the identity with respect to this measure:

$$\int_X |x\rangle\langle x| \mathrm{d}\omega(x) = I_{\mathfrak{H}}$$

In this case, the coherent states may be used to quantize the observables of the classical system in the following way: let $f$ be a real-valued function on $X$, representing a classical observable and suppose that the formal operator

$$\widehat{f} = \int_X f(x)|x\rangle\langle x| \mathrm{d}\omega(x) \quad [59]$$

is well defined as a self-adjoint operator on $\mathfrak{H}$. Then we may take the operator $\widehat{f}$ to be the quantized observable corresponding to the classical observable $f$. Suppose that we have two such operators, $\widehat{f}$ and $\widehat{g}$,

corresponding to the two classical observables $f$ and $g$, which have the Poisson bracket $\{f, g\}$, defined via the 2-form $\omega$. We then check if the quantization condition

$$\widehat{\{f, g\}} = \frac{2\pi}{\mathrm{i}h}[\widehat{f}, \widehat{g}] \qquad [60]$$

where $h$ is Planck's constant, is satisfied. Generally this will be the case for a certain number of classical observables. This method of quantization has been most successfully used for manifolds $X$ which have a (complex) Kähler structure. Over such a manifold, one can define a Hilbert space of analytic functions, which has a reproducing kernel and hence a naturally associated set of coherent states. As a specific example, we take the case of canonical coherent states [11]. We can identify the complex plane $\mathbb{C}$ with the phase space $\mathbb{R}^2$ of a free classical particle having a single degree of freedom. The measure $\mathrm{d}\omega$ in this case is just $(1/2\pi)\mathrm{d}q\,\mathrm{d}p$. If we now quantize the classical observables $f(q,p) = q$ and $f(q,p) = p$, of position and momentum, respectively, using the canonical coherent states, we obtain the two operators

$$\begin{aligned} Q &= \int_{\mathbb{R}^2} q|q,p\rangle\langle q,p| \frac{\mathrm{d}q\,\mathrm{d}p}{2\pi} \\ P &= \int_{\mathbb{R}^2} p|q,p\rangle\langle q,p| \frac{\mathrm{d}q\,\mathrm{d}p}{2\pi} \end{aligned} \qquad [61]$$

It can be verified that these two operators satisfy the canonical commutation relations $[Q, P] = \mathrm{i}I_{\mathfrak{H}}$, as required.

*See also:* Solitons and Kac–Moody Lie Algebras; Wavelets: Mathematical Theory.

## Further Reading


Ali ST, Antoine J-P, and Gazeau J-P (2000) *Coherent States, Wavelets and Their Generalizations*. New York: Springer.

Ali ST and Engliš M (2005) Quantization methods – a guide for physicists and analysts. *Reviews in Mathematical Physics* 17: 391–490.

Brif C (1997) SU(2) and SU(1,1) algebra eigenstates: a unified analytic approach to coherent and intelligent states. *International Journal of Theoretical Physics* 36: 1651–1682.

Klauder JR and Sudarshan ECG (1968) *Fundamentals of Quantum Optics*. New York: Benjamin.

Klauder JR and Skagerstam BS (1985) *Coherent States – Applications in Physics and Mathematical Physics*. Singapore: World Scientific.

Perelomov AM (1986) *Generalized Coherent States and their Applications*. Berlin: Springer.

Schrödinger E (1926) Der stetige Übergang von der Mikro- zur Makromechanik. *Naturwissenschaften* 14: 664–666.

Sivakumar S (2000) Studies on nonlinear coherent states. *Journal of Optics B: Quantum Semiclass. Opt.* 2: R61–R75.

Zhang W-M, Feng DH, and Gilmore RG (1990) Coherent states: theory and some applications. *Reviews of Modern Physics* 62: 867–927.


# Cohomology Theories

**U Tillmann**, University of Oxford, Oxford, UK



## Introduction

The origins of cohomology theory are found in topology and algebra at the beginning of the last century but since then it has become a tool of nearly every branch of mathematics. It's a way of life! Naturally, this article can only give a glimpse at the rich subject. We take here the point of view of algebraic topology and discuss only the cohomology of spaces.

Cohomology reflects the global properties of a manifold, or more generally of a topological space. It has two crucial properties: it only depends on the homotopy type of the space and is determined by local data. The latter property makes it in general computable.

To illustrate the interplay between the local and global structure, consider the Euler characteristic of a compact manifold; as will be explained below, cohomology is a refinement of the Euler characteristic. For simplicity, assume that the manifold $M$ is a surface and that we have chosen a way of dividing the surface into triangles. The Euler characteristic is then defined to be

$$\chi(M) = F - E + V$$

where $F$ denotes the number of faces, $E$ the number of edges, and $V$ the number of vertices in the triangulation. Remarkably, this number does not depend on the triangulation. Yet, this simple, easy to compute number can already distinguish the different types of closed, oriented surfaces: for the sphere we have $\chi = 2$, the torus $\chi = 0$, and in general for any surface $M_g$ of genus $g$

$$\chi(M_g) = 2 - 2g$$

The Euler characteristic also tells us something about the geometry and analysis of the manifold. For example, the total curvature of a surface is equal to its Euler characteristic. This is the Gauss–Bonnet theorem and an analogous result holds in higher dimensions. Another striking result is the Poincaré–Hopf theorem which equates the Euler characteristic with the total index of a vector field and thus gives strong restrictions on what kind of vector fields can exist on a manifold. This interplay between global analysis and topology has been one of the most exciting and fruitful research areas and is most powerfully expressed in the celebrated Atiyah–Singer index theorem, which determines the analytic index of an elliptic operator, such as the Dirac operator on a spin manifold, in terms of cohomology classes.

## Chain Complexes and Homology

There are several different geometric definitions of the cohomology of a topological space. All share some basic algebraic structure which we will explain first.

A "chain complex" $(C_*, \partial_*)$

$$\cdots C_{i+1} \xrightarrow{\partial_{i+1}} C_i \xrightarrow{\partial_i} C_{i-1} \cdots \xrightarrow{\partial_1} C_0 \qquad [1]$$

is a collection of vector spaces (or $R$-modules more generally) $C_i, i \geq 0$, and linear maps ($R$-module maps) $\partial_i : C_i \to C_{i-1}$ with the property that for all $i$

$$\partial_i \circ \partial_{i+1} = 0 \qquad [2]$$

The scalar fields one tends to consider are the rationals $\mathbb{Q}$, reals $\mathbb{R}$, complex numbers $\mathbb{C}$, or a primary field $\mathbb{Z}_{\mathrm{p}}$, while the most important ring $R$ is the ring of integers $\mathbb{Z}$ though we will also consider localizations such as $\mathbb{Z}[1/p]$, which has the effect of suppressing any $p$-primary torsion information.

Of particular interest are the elements in $C_i$ that are mapped to zero by $\partial_i$, the $i$-dimensional "cycles," and those that are in the image of $\partial_{i+1}$, the $i$-dimensional "boundaries." Because of [2], every boundary is a cycle, and we may define the quotient vector space ($R$-module), the $i$th-dimensional homology,

$$H_i(C_*; \partial_*) := \frac{\ker \partial_i}{\mathrm{im} \partial_{i+1}} \qquad [3]$$

$(C_*, \partial_*)$ is "exact" if all its cycles are boundaries. Homology thus measures to what extent the sequence [1] fails to be exact.

### Simplicial Homology

A triangulation of a surface gives rise to its "simplicial" chain complex: Taking coefficients in $\mathbb{Z}, \mathbb{C}_2, \mathbb{C}_1, \mathbb{C}_0$ are the free abelian groups generated by the set of faces, edges, and vertices, respectively; $C_i = \{0\}$ for $i \geq 3$. The map $\partial_2$ assigns to a triangle the sum of its edges; $\partial_1$ maps an edge to the sum of its endpoints. If we are working with $\mathbb{Z}_2$ coefficients, this defines for us a chain complex as [2] is clearly satisfied; in general, one needs to keep track of the orientations of the triangles and edges and take sums with appropriate signs (cf. [6] below). An easy calculation shows that for an oriented, closed surface $M_g$ of genus $g$, we have

$$\begin{aligned} H_0(M_g, \mathbb{Z}) &= \mathbb{Z} \\ H_1(M_g; \mathbb{Z}) &= \mathbb{Z}^{2g} \\ H_2(M_g; \mathbb{Z}) &= \mathbb{Z} \\ H_i(M_g; \mathbb{Z}) &= 0 \quad \text{for } i \geq 3 \end{aligned} \qquad [4]$$

Note that the Euler characteristic can be recovered as the alternating sum of the rank of the homology groups:

$$\chi(M) = \sum_{i=0}^{\dim M} (-1)^i \operatorname{rk} H_i(M; \mathbb{Z}) \qquad [5]$$

Every smooth manifold $M$ has a triangulation, so that its simplicial homology can be defined just as above. More generally, simplicial homology can be defined for any simplicial space, that is, a space that is built up out of points, edges, triangles, tetrahedra, etc. Formula [5] remains valid for any compact manifold or simplicial space.

### Singular Homology

Let $X$ be any topological space, and let $\triangle^n$ be the oriented $n$-simplex $[v_0, \ldots, v_n]$ spanned by the standard basis vectors $v_i$ in $\mathbb{R}^{n+1}$. The set of singular $n$-chains $S_n(X)$ is the free abelian group on the set of continuous maps $\sigma : \triangle^n \longrightarrow X$. The boundary of $\sigma$ is defined by the alternating sum of the restriction of $\sigma$ to the faces of $\triangle^n$:

$$\partial_n(\sigma) := \sum_{i=0}^{n} (-1)^{-i} \sigma|_{[v_0, \ldots, \hat{v}_i, \ldots, v_n]} \qquad [6]$$

One easily checks that the boundary of a boundary is zero, and hence $(S_*(X), \partial_*)$ defines a chain complex. Its homology is by definition the singular homology $H_*(X; \mathbb{Z})$ of $X$. For any simplicial space, the inclusion of the simplicial chains into the singular chains induces an isomorphism of homology groups. In particular, this implies that the simplicial homology of a manifold, and hence its Euler characteristic do not depend on its triangulation.

If in the definition of simplicial and singular homology we take free $R$-modules (where $R$ may

also be a field) instead of free abelian groups, we get the homology $H_*(X;R)$ of $X$ with coefficients in $R$. The "universal-coefficient theorem" describes the homology with arbitrary coefficients in terms of the homology with integer coefficients. In particular, if $R$ is a field of characteristic zero,

$$\dim H_n(X;R) = \operatorname{rk} H_n(X;\mathbb{Z})$$

### Basic Properties of Singular Homology

While simplicial homology (and the more efficient cellular homology which we will not discuss) is easier to compute and easier to understand geometrically, singular homology lends itself more easily to theoretical treatment.

1. *Homotopy invariance*. Any continuous map $f:X\to Y$ induces a map on homology $f_*:H_*(X;R)\to H_*(Y;R)$ which only depends on the homotopy class of $f$.

In particular, a homotopy equivalence $f:X\to Y$ induces an isomorphism in homology. So, for example, the inclusion of the circle $S^1$ into the punctured plane $\mathbb{C}\backslash\{0\}$ is a homotopy equivalence, and thus

$$\begin{aligned} H_i(\mathbb{C}\backslash\{0\};R) &\simeq H_i(S^1;R) \\ &= \begin{cases} \mathbb{Z} & \text{for } i=0,1 \\ 0 & \text{for } i\geq 2 \end{cases} \end{aligned}$$

For the one point space we have $H_0(\text{pt};R)=R$. Define reduced homology by $\tilde{H}_*(X;R):=\ker(H_*(X;R)\to H_*(\text{pt};R))$.

2. *Dimension axiom*. $\tilde{H}_i(\text{pt};R)=0$ for all $i$.

More generally, it follows immediately from the definition of simplicial homology that the homology of any $n$-dimensional manifold is zero in dimensions larger than $n$.

We mentioned in the introduction that homology depends only on local data. This is made precise by the

3. *Mayer–Vietoris theorem. Let $X=A\cup B$ be the union of two open subspaces. Then the following sequence is exact*:

$$\begin{aligned} \cdots \longrightarrow H_n(A\cap B;R) &\longrightarrow H_n(A;R)\oplus H_n(B;R) \\ &\longrightarrow H_n(X;R) \xrightarrow{\partial} H_{n-1}(A\cap B;R) \\ &\longrightarrow \cdots \longrightarrow H_0(X;R)\to 0 \end{aligned}$$

On the level of chains, the first map is induced by the diagonal inclusion, while the second map takes the difference between the first and second summands. Finally, $\partial$ takes a cycle $c=a+b$ in the chains of $X$ that can be expressed as the sum of a chain $a$ in $A$ and $b$ in $B$ to $\partial c:=\partial_n a=-\partial_n b$. For example, consider two cones, $A$ and $B$, on a space $X$ and identify them at the base $X$ to define the suspension $\Sigma X$ of $X$. Then $\Sigma X=A\cup B$ with $A,B\simeq \text{pt}$ and $A\cap B\simeq X$. The boundary map $\partial$ is then an isomorphism:

$$\tilde{H}_n(X;R)\simeq H_{n+1}(\Sigma X;R) \quad \text{for all } n\geq 0 \qquad [7]$$

From this one can easily compute the homology of a sphere. First note that

$$\tilde{H}_0(X;\mathbb{Z})=\mathbb{Z}^{k-1}$$

where $k$ is the number of connected components in $X$. Also, $S^n\simeq \Sigma S^{n-1}\simeq\cdots\simeq\Sigma^n S^0$. Thus, by [7],

$$H_n(S^n;\mathbb{Z})\simeq\mathbb{Z} \quad \text{and} \quad \tilde{H}_*(S^n;\mathbb{Z})=0 \quad \text{for } *\neq n \qquad [8]$$

If $Y$ is a subspace of $X$, relative homology groups $H_*(X,Y;R)$ can be defined as the homology of the quotient complex $S_*(X)/S_*(Y)$. When $Y$ has a good neighborhood in $X$ (i.e., it is a neighborhood deformation retract in $X$), then, by the "excision theorem,"

$$H_*(X,Y;R)\simeq\tilde{H}_*(X/Y;R)$$

where $X/Y$ denotes the quotient space of $X$ with $Y$ identified to a point. There is a long exact sequence

$$\begin{aligned} \cdots\longrightarrow H_n(Y;R)\longrightarrow H_n(X;R)\longrightarrow H_n(X,Y;R) \\ \xrightarrow{\partial} H_{n-1}(Y;R)\longrightarrow\cdots\longrightarrow H_0(X,Y;R)\longrightarrow 0 \end{aligned}$$

This and the Mayer–Vietoris sequence give two ways of breaking up the problem of computing the homology of a space into computing the homology of related spaces. An iteration of this process leads to the powerful tool of spectral sequences (*see* Spectral Sequences).

### Relation to Homotopy Groups

Let $\pi_1(X,x_0)$ denote the fundamental group of $X$ relative to the base point $x_0$. These are the based homotopy classes of based maps from a circle to $X$.

$$\text{If } X \text{ is connected, then } H_1(X;\mathbb{Z}) \text{ is the abelianization of } \pi_1(X,x_0) \qquad [9]$$

Indeed, every map from a (triangulated) sphere to $X$ defines a cycle and hence gives rise to a homology class. This defines the Hurewicz map $h:\pi_*(X;x_0)\to H_*(X;\mathbb{Z})$. In general there is no good description of its image. However, if $X$ is $k$-connected with $k\geq 1$, then $h$ induces an isomorphism in dimension $k+1$ and an epimorphism in dimension $k+2$.

Though [9] indicates that homology cannot distinguish between all homotopy types, the fundamental group is in a sense the only obstruction to this. A simple form of the "Whitehead theorem" states:

**Theorem** *If a map $f: X \to Y$ between two simplicial complexes with trivial fundamental groups induces an isomorphism on all homology groups, then it is a homotopy equivalence.*

*Warning*: This does not imply that two simply connected spaces with isomorphic homology groups are homotopic! The existence of the map $f$ inducing this isomorphism is crucial and counterexamples can easily be constructed.

## Dual Chain Complexes and Cohomology

The process of dualizing itself cannot be expected to yield any new information. Nevertheless, the cohomology of a space, which is obtained by dualizing its simplicial chain complex, carries important additional structure: it possesses a product, and moreover, when the coefficients are a primary field, it is an algebra over the rich Steenrod algebra. As with homology we start with the algebraic setup.

Every chain complex $(C_*, \partial_*)$ gives rise to a dual chain complex $(C^*, \partial^*)$ where $C^i = \hom_R(C_i, R)$ is the dual $R$-module of $C_i$; because of [2], the composition of two dual boundary morphisms $\partial^{i+1}: C^i \to C^{i+1}$ is trivial. Hence we may define the $i$th dimensional cohomology group as

$$H^i(C^*, \partial^*) := \frac{\ker \partial^{i+1}}{\operatorname{im} \partial^i} \qquad [10]$$

Evaluation $(\sigma, \phi) \mapsto \phi(\sigma)$ descends to a dual pairing

$$H_n(C_*, \partial_*) \otimes_R H^n(C^*, \partial^*) \longrightarrow R$$

and when $R$ is a field, this identifies the cohomology groups as the duals of the homology groups. More generally, the universal-coefficient theorem relates the two. A simple version states: let $(C_*, \partial_*)$ be a chain complex of free abelian groups (such as the simplicial or singular chain complexes) with finitely generated homology groups. Then,

$$H^i(C^*, \partial^*) \simeq H_i^{\text{free}}(C_*, \partial_*) \oplus H_{i-1}^{\text{tor}}(C_*, \partial) \qquad [11]$$

where $H_*^{\text{tor}}$ denotes the torsion subgroup of $H_*$ and $H_*^{\text{free}}$ denotes the quotient group $H_*/H_*^{\text{tor}}$.

### Singular Cohomology

The dual $S^*(X)$ of the singular chain complex of a space $X$ carries a natural pairing, the cup product, $\cup: S^p(X) \otimes S^q(X) \to S^{p+q}(X)$ defined by

$$(\phi_1 \cup \phi_2)(\sigma) := \phi_1(\sigma|_{[v_0, \ldots, v_p]})\phi_2(\sigma|_{[v_p, \ldots, v_{p+q}]})$$

This descends to a multiplication on cohomology groups and makes $H^*(X; R) := \bigoplus_{n \geq 0} H^n(X; R)$ into an associative, graded commutative ring: $u \cup v = (-1)^{\deg u \deg v} v \cup u$.

The "Künneth theorem" gives some geometric intuition for the cup product. A simple version states: for spaces $X$ and $Y$ with $H^*(Y; R)$ a finitely generated free $R$-module, the cup product defines an isomorphism of graded rings

$$H^*(X; R) \otimes_R H^*(Y; R) \longrightarrow H^*(X \times Y; R)$$

For example, for a sphere, all products are trivial for dimension reasons. Hence,

$$H^*(S^n; \mathbb{Z}) = \bigwedge\nolimits^*(x) \qquad [12]$$

is an exterior algebra on one generator $x$ of degree $n$. On the other hand, the cohomology of the $n$-dimensional torus $T^n$ is an exterior algebra on $n$ degree-1 generators,

$$H^*(T^n; \mathbb{Z}) = \bigwedge\nolimits^*(x_1, \ldots, x_n) \qquad [13]$$

The dual pairing can be generalized to the slant or cap product

$$\cap : H_n(X; R) \otimes_R H^i(X; R) \longrightarrow H_{n-i}(X; R)$$

defined on the chain level by the formula $(\sigma, \phi) \mapsto \phi(\sigma|_{[v_0, \ldots, v_i]})\sigma|_{[v_i, \ldots, v_n]}$.

### Steenrod Algebra

The cup product on the chain level is homotopy commutative, but not commutative. Steenrod used this defect to define operations

$$Sq^i : H^n(X; \mathbb{Z}_2) \longrightarrow H^{n+i}(X; \mathbb{Z}_2)$$

for all $i \geq 0$ which refine the cup-squaring operation: when $n = i$, then $Sq^n(x) = x \cup x$. These are natural group homomorphisms which commute with suspension. Furthermore, they satisfy the Cartan and Adem Relations

$$Sq^n(x \cup y) = \sum_{i+j=n} Sq^i(x) \cup Sq^j(y)$$

$$Sq^i Sq^j = \sum_{k=0}^{[i/2]} \binom{j-k-1}{i-2k} Sq^{i+j-k} Sq^k$$

for $i \leq 2j$

The mod-2 Steenrod algebra $\mathcal{A}$ is then the free $\mathbb{Z}_2$-algebra generated by the Steenrod squares $Sq^i, i \geq 0$, subject only to the Adem relations. With the help of Adem's relations, Serre and Cartan found a $\mathbb{Z}_2$-basis for $\mathcal{A}$:

$$\{Sq^I := Sq^{i_1} \cdots Sq^{i_n} | i_j \geq 2i_{j+1} \quad \text{for all } j\}$$

The Steenrod algebra is also a Hopf algebra with a commutative comultiplication $\Delta : \mathcal{A} \to \mathcal{A} \otimes \mathcal{A}$ induced by

$$\Delta(Sq^n) := \sum_{i+j=n} Sq^i \otimes Sq^j$$

The Cartan relation implies that the mod-2 cohomology of a space is compatible with the comultiplication, that is, $H^*(X; \mathbb{Z}_2)$ is an algebra over the Hopf algebra $\mathcal{A}$. There are odd primary analogs of the Steenrod algebra based on the reduced $p$th power operations

$$P^i : H^n(X; \mathbb{Z}_p) \to H^{n+2i(p-1)}(X; \mathbb{Z}_p)$$

with similar properties to $\mathcal{A}$.

One of the most striking applications of the Steenrod algebra can be found in the work of Adams on the "vector fields on spheres problem": for each $n$, find the greatest number $k$, denoted $K(n)$, such that there is a $k$-field on the $(n-1)$-sphere $S^{n-1}$. Recall that a $k$-field is an ordered set of $k$ pointwise linear independent tangent vector fields. If we write $n$ in the form $n = 2^{4a+b}(2s+1)$ with $0 \le b < 4$, Adams proved that $K(n) = 2^b + 8a - 1$. In particular, when $n$ is odd, $K(n) = 0$. We give an outline of the proof for this special case in the next section.

- The failure of associativity of the cup product at the chain level gives rise to secondary operations, the so-called "Massey products."

## Cohomology of Smooth Manifolds

A smooth manifold $M$ of dimension $n$ can be triangulated by smooth simplices $\sigma : \Delta^n \to M$. If $M$ is compact, oriented, without boundary, the sum of these simplices define a homology cycle $[M]$, the fundamental class of $M$. The most remarkable property of the cohomology of manifolds is that they satisfy "Poincaré duality": taking cap product with $[M]$ defines an isomorphism:

$$D := [M] \cap : H^k(M; \mathbb{Z}) \xrightarrow{\simeq} H_{n-k}(M; \mathbb{Z}) \quad \text{for all } k \qquad [14]$$

In particular, for connected manifolds, $H^n(M; \mathbb{Z}) \simeq \mathbb{Z}$; and every map $f : M' \to M$ between oriented, compact closed manifolds of the same dimension has a degree: $f^* : H^*(M; \mathbb{Z}) \to H^*(M'; \mathbb{Z})$ is multiplication by an integer $\deg(f)$, the degree of $f$. For smooth maps, the degree is the number of points in the inverse image of a generic point $p \in M$ counted with signs:

$$\deg(f) = \sum_{p' \in f^{-1}(p)} \text{sign}(p')$$

where $\text{sign}(p')$ is +1 or –1 depending on whether $f$ is orientation preserving or reversing in a neighborhood of $p'$. For example, a complex polynomial of degree $d$ defines a map of the two-dimensional sphere to itself of degree $d$: a generic point has $n$ points in its inverse image and the map is locally orientation preserving. On the other hand, a map of $S^{n-1}$ induced by a reflection of $\mathbb{R}^n$ reverses orientation and has degree –1. Thus, as degrees multiply on composing maps, the antipodal map $x \mapsto -x$ has degree $(-1)^n$. As an application we prove:

*Every tangent vector field on an even-dimensional sphere $S^{n-1}$ has a zero.*

**Proof** Assume $v(x)$ is a vector field which is nonzero for all $x \in S^{n-1}$. Then $x$ is perpendicular to $v(x)$, and after rescaling, we may assume that $v(x)$ has length 1. The function $F(x,t) = \cos(t)x + \sin(t)\, v(x)$ is a well-defined homotopy from the identity map $(t=0)$ to the antipodal map $(t=\pi)$. But this is impossible as homotopic maps induce the same map in (co)homology and we have already seen that the degree of the identity map is 1 while the degree of the antipodal map is $(-1)^n = -1$ when $n$ is odd.

- It is well known that two self-maps of a sphere of any dimension are homotopic if and only if they have the same degree, that is, $\pi_n(S^n) \simeq \mathbb{Z}$ for $n \ge 1$.
- When $M$ is not orientable, $[M]$ still defines a cycle in homology with $\mathbb{Z}_2$-coefficients, and $[M]\cap$ defines an isomorphism between the cohomology and homology with $\mathbb{Z}_2$ coefficients.
- As $[M]$ represents a homology class, so does every other closed (orientable) submanifold of $M$. It is however not the case that every homology class can be represented by a submanifold or linear combinations of such.

Cohomology is a contravariant functor. Poincaré duality however allows us to define, for any $f : M' \to M$ between oriented, compact, closed manifolds of arbitrary dimensions, a "transfer" or "Umkehr map,"

$$f^! := D^{-1} f_* D' : H^*(M'; \mathbb{Z}) \to H^{*-c}(M; \mathbb{Z})$$

which lowers the degree by $c = \dim M' - \dim M$. It satisfies the formula

$$f^!(f^*(x) \cup y) = x \cup f^!(y)$$

for all $x \in H^*(M; \mathbb{Z})$ and $y \in H^*(M'; \mathbb{Z})$. When $f$ is a covering map then $f^!$ can be defined on the chain level by

$$f^!(x)(\sigma) := x\Big(\sum_{f(\tilde{\sigma})=\sigma} \tilde{\sigma}\Big)$$

where $x \in C^*(M')$ and $\sigma \in C_*(M)$.

### de Rham Cohomology

If $x_1, \dots, x_n$ are the local coordinates of $\mathbb{R}^n$, define an algebra $\Omega^*$ to be the algebra generated by symbols $dx_1, \dots, dx_n$ subject to the relations $dx_i dx_j = -dx_j dx_i$ for all $i, j$. We say $dx_{i_1} \cdots dx_{i_q}$ has degree $q$. The differential forms on $\mathbb{R}^n$ are the algebra

$$\Omega^*(\mathbb{R}^n) := \{C^\infty \text{functions on } \mathbb{R}^n\} \otimes_{\mathbb{R}} \Omega^*$$

The algebra $\Omega^*(\mathbb{R}^n) = \bigoplus_{q=0}^n \Omega^q(\mathbb{R}^n)$ is naturally graded by degree. There is a differential operator $d : \Omega^q(\mathbb{R}^n) \to \Omega^{q+1}(\mathbb{R}^n)$ defined by

1. if $f \in \Omega^0(\mathbb{R}^n)$, then $df = \sum (\partial f / \partial x_i) dx_i$
2. if $\omega = \sum f_I dx_I$, then $d\omega = \sum df_I dx_I$

$I$ stands here for a multi-index. For example, in $\mathbb{R}^3$ the differential assigns to 0-forms ( = functions) the gradient, to 1-forms the curl, and to 2-forms the divergence. An easy exercise shows that $d^2 = 0$ and the $q$th de Rham cohomology of $\mathbb{R}^n$ is the vector space

$$H^q_{\text{de R}}(\mathbb{R}^n) = \frac{\ker d : \Omega^q(\mathbb{R}^n) \to \Omega^{q+1}(\mathbb{R}^n)}{\text{im } d : \Omega^{q-1}(\mathbb{R}^n) \to \Omega^q(\mathbb{R}^n)}$$

More generally, the de Rham complex $\Omega^*(M)$ and its cohomology $H^*_{\text{de R}}(M)$ can be defined for any smooth manifold $M$.

Let $\sigma$ be a smooth, singular, real $(q+1)$-chain on $M$, and let $\omega \in \Omega^q(M)$. Stokes theorem then says

$$\int_{\partial\sigma} \omega = \int_\sigma d\omega$$

and therefore integration defines a pairing between the $q$th singular homology and the $q$th de Rahm cohomology of $M$. This pairing is exact and thus de Rahm cohomology is isomorphic to singular cohomology with real coefficients:

$$H^*_{\text{de R}}(M) \simeq (H_*(M; \mathbb{R}))^* \simeq H^*(M; \mathbb{R})$$

Let $\Omega^*_c(M)$ denote the subcomplex of compactly supported forms and $H^*_c(M)$ its cohomology. Integration with respect to the first $i$ coordinates defines a map

$$\Omega^*_c(\mathbb{R}^n) \to \Omega^{*-i}_c(\mathbb{R}^{n-i})$$

which induces an isomorphism in cohomology; note in particular $H^n_c(\mathbb{R}^n) = \mathbb{R}$. More generally, when $E \to M$ is an $i$-dimensional orientable, real vector bundle over a compact, orientable manifold $M$, integration over the fiber gives the "Thom isomorphism":

$$H^*_c(E) \simeq H^{*-i}_c(M) \simeq H^{*-i}_{\text{de R}}(M)$$

For orientable fiber bundles $F \to M' \xrightarrow{f} M$ with compact, orientable fiber $F$, integration over the fiber provides another definition of the transfer map

$$f^! : H^*_{\text{de R}}(M') \to H^{*-i}_{\text{de R}}(M)$$

### Hodge Decomposition

Let $M$ be a compact oriented Riemannian manifold of dimension $n$. The Hodge star operator, $*$, associates to every $q$-form an $(n-q)$-form. For $\mathbb{R}^n$ and any orthonormal basis $\{e_1, \dots, e_n\}$, it is defined by setting

$$*(e_1 \wedge \cdots \wedge e_q) := \pm e_{p+1} \wedge \cdots \wedge e_n$$

where one takes $+$ if the orientation defined by $\{e_1, \dots, e_n\}$ is the same as the given one, and $-$ otherwise. Using local coordinate charts this definition can be extended to $M$. Clearly, $*$ depends on the chosen metric and orientation of $M$. If $M$ is compact, we may define an inner product on the $q$-forms by

$$(\omega, \omega') := \int_M \omega \wedge *\omega'$$

With respect to this inner product $*$ is an isometry. Define the codifferential via

$$\delta := (-1)^{np+n+1} * d* : \Omega^q(M) \to \Omega^{q-1}(M)$$

and the Laplace–Beltrami operator via

$$\Delta := \delta d + d\delta$$

The codifferential satisfies $\delta^2 = 0$ and is the adjoint of the differential. Indeed, for $q$-forms $\omega$ and $(q+1)$-forms $\omega'$:

$$(d\omega, \omega') = (\omega, \delta\omega') \qquad [15]$$

It follows easily that $\Delta$ is self-adjoint, and furthermore,

$$\Delta\omega = 0 \quad \text{if and only if} \quad d\omega = 0 \text{ and } \delta\omega = 0 \qquad [16]$$

A form $\omega$ satisfying $\Delta\omega = 0$ is called "harmonic." Let $\mathcal{H}^q$ denote the subspace of all harmonic $q$-forms. It is not hard to prove the "Hodge decomposition theorem":

$$\Omega^q = \mathcal{H}^q \oplus \text{im}\, d \oplus \text{im}\, \delta$$

Furthermore, by adjointness [15], a form $\omega$ is closed only if it is orthogonal to $\text{im}\,\delta$. On calculating the de Rham cohomology we can also ignore the summand $\text{im}\, d$ and find that:

*Each de Rham cohomology class on a compact oriented Riemannian manifold M contains a unique harmonic representative, that is,* $H^q_{\text{de R}}(M) \simeq \mathcal{H}^q$.

*Warning*: This is an isomorphism of vector spaces and in general does not extend to an isomorphism of algebras.

## Examples

We list the cohomology of some important examples.

#### Projective Spaces

Let $\mathbb{R}P^n$ be real projective space of dimension $n$. Then,

$$H^*(\mathbb{R}P^n;\mathbb{Z}_2)=\mathbb{Z}_2[x]/(x^{n+1})$$

is a stunted polynomial ring on a generator $x$ of degree 1.

Similarly, let $\mathbb{C}P^n$ and $\mathbb{H}P^n$ denote complex and quaternionic projective space of real dimensions $2n$ and $4n$, respectively. Then,

$$H^*(\mathbb{C}P^n;\mathbb{Z})=\mathbb{Z}[y]/(y^{n+1})$$
$$H^*(\mathbb{H}P^n;\mathbb{Z})=\mathbb{Z}[z]/(z^{n+1})$$

are stunted polynomial rings with $\deg(y)=2$ and $\deg(z)=4$.

#### Lie Groups

Let $G$ be a compact, connected Lie group of rank $l$, that is, the dimension of the maximal torus of $G$ is $l$. Then,

$$H^*(G,\mathbb{Q})\simeq \bigwedge_{\mathbb{Q}}^{*}[a_{2d_1-1},a_{2d_2-1},\ldots,a_{2d_l-1}]$$

where $|a_i|=i$ and $d_1,\ldots,d_l$ are the fundamental degrees of $G$ which are known for all $G$. Often this structure lifts to the integral cohomology. In particular we have:

$$H^*_{\text{free}}(SO(2k+1);\mathbb{Z}))\simeq \bigwedge_{\mathbb{Z}}^{*}[a_3,a_7,\ldots,a_{4k-1}]$$
$$H^*_{\text{free}}(SO(2k);\mathbb{Z}))\simeq \bigwedge_{\mathbb{Z}}^{*}[a_1,a_7,\ldots,a_{4k-5},a_{2k-1}]$$
$$H^*(U(k);\mathbb{Z})\simeq \bigwedge_{\mathbb{Z}}^{*}[a_1,a_3,\ldots,a_{2k-1}]$$

#### Classifying Spaces

For any group $G$ there exists a classifying space $BG$, well defined up to homotopy. Classifying spaces are of central interest to geometers and topologists for the set of isomorphism classes of principal $G$-bundles over a space $X$ is in one-to-one correspondence with the set of homotopy classes of maps from $X$ to $BG$. In particular, every cohomology class $c\in H^*(BG;R)$ defines a characteristic class of principle $G$-bundles $E$ over $X$: if $E$ corresponds to the map $f_E:X\to BG$, then $c(E):=f_E^*(c)$.

$BG$ can be constructed as the space of $G$-orbits of a contractible space $EG$ on which $G$ acts freely. Thus, for example,

$$B\mathbb{Z}=\mathbb{R}/\mathbb{Z}\simeq S^1$$
$$B\mathbb{Z}_2=(\lim_{n\to\infty}S^n)/\mathbb{Z}_2\simeq\mathbb{R}P^\infty$$
$$BS^1=(\lim_{n\to\infty}S^{2n+1})/S^1\simeq\mathbb{C}P^\infty$$

and more generally, infinite Grassmannian manifolds are classifying spaces for linear groups. When $G$ is a compact connected Lie group,

$$H^*(BG;\mathbb{Q})\simeq\mathbb{Q}[x_{2d_1},\ldots,x_{2d_l}]$$

with $d_i$ as above and $|x_i|=i$. In particular,

$$H^*(BSO(2k+1);\mathbb{Z}[1/2])\simeq\mathbb{Z}[1/2][p_1,p_2,\ldots,p_k]$$
$$H^*(BSO(2k);\mathbb{Z}[1/2])\simeq\mathbb{Z}[1/2][p_1,p_2,\ldots,p_{k-1},e_k]$$
$$H^*(BU(k);\mathbb{Z})\simeq\mathbb{Z}[c_1,c_2,\ldots,c_k]$$

where the Pontryagin, Euler, and Chern classes have degree $|p_i|=4i$, $|e_k|=2k$, and $|c_i|=2i$, respectively.

#### Moduli Spaces

Let $\mathcal{M}_g^n$ be the space of Riemann surfaces of genus $g$ with $n$ ordered, marked points. There are naturally defined classes $\kappa_i$ and $e_1,\ldots,e_n$ of degree $2i$ and 2, respectively. By Harer–Ivanov stability and the recent proof of the Mumford conjecture (Madsen–Weiss, preprint 2004), there is an isomorphism up to degree $*<3g/2$ of the rational cohomology of $\mathcal{M}_g^n$ with

$$\mathbb{Q}[\kappa_1,\kappa_2,\ldots]\otimes\mathbb{Q}[e_1,\ldots,e_n]$$

The rational cohomology vanishes in degrees $*>4g-5$ if $n=0$, and $*>4g-4+n$ if $n>0$. Though the stable part of the cohomology is now well understood, the structure of the unstable part, as proposed by Faber (Viehweg 1999), remains conjectural.

## Generalized Cohomology Theories

The three basic properties of singular homology appropriately dualized, hold of course also for cohomology. Furthermore, they (essentially) determine (co)homology uniquely as a functor from the category of simplicial spaces and continuous functions to the category of abelian groups. If we drop the dimension axiom (2), we are left with homotopy invariance (1), and the Mayer–Vietoris sequence (3). Abelian group valued functors satisfying (1) and (3) are so called “generalized (co)homology theories.”

$K$-theory and cobordism theory are two well-known examples but there are many more.

### *K*-Theory

The geometric objects representing elements in complex $K$-theory $K^0(X)$ are isomorphism classes of finite dimensional complex vector bundles $E$ over $X$. Vector bundles $E, E'$ can be added to form a new bundle $E \oplus E'$ over $X$, and $K^0(X)$ is just the group completion of the arising monoid. Thus, for example, for the point space we have $K^0(\mathrm{pt}) = \mathbb{Z}$. Tensor product of vector bundles $E \otimes E'$ induces a multiplication on $K$-theory making $K^*(X)$ into a graded commutative ring.

In many ways $K$-theory is easier than cohomology. In particular, the groups are 2-periodic: all even degree groups are isomorphic to the reduced $K$-theory group $\tilde{K}^0(X) := \mathrm{coker}(K^0(\mathrm{pt}) = \mathbb{Z} \to K^0(X))$, and all odd degree groups are isomorphic to $K^{-1}(X) := \tilde{K}^0(\Sigma X)$.

The theory of characteristic classes gives a close relation between the two cohomology theories. The Chern character map, a rational polynomial in the Chern classes, defines

$$\begin{aligned} \mathrm{ch} : K^0(X) \otimes_{\mathbb{Z}} \mathbb{Q} &\to H^{\mathrm{even}}(X; \mathbb{Q}) \\ &:= \bigoplus_{k \geq 0} H^{2k}(X; \mathbb{Q}) \end{aligned}$$

an isomorphism of rings. Thus, the $K$-theory and cohomology of a space carry the same rational information. But they may have different torsion parts. This became an issue in string theory when D-brane charges which had formerly been thought of as differential forms (and hence cohomology classes) were later reinterpreted more naturally as $K$-theory classes by Witten 1998)

- There are real and quaternionic $K$-theory groups which are 8-periodic.

### Cobordism Theory

The geometric objects representing an element in the oriented cobordism group $\Omega^n_{SO}(X)$ are pairs $(M, f)$ where $M$ is a smooth, orientable $n$-dimensional manifold and $f: M \to X$ is a continuous map. Two pairs $(M, f)$ and $(M', f')$ represent the same cobordism class if there exists a pair $(W, F)$ where $W$ is an $(n+1)$-dimensional, smooth, oriented manifold with boundary $\partial W = M \cup -M'$ such that $F: W \to X$ restricts to $f$ and $f'$ on the boundary $\partial W$. Disjoint union and Cartesian product of manifolds define an addition and multiplication so that $\Omega^*_{SO}(X)$ is a graded, commutative ring.

- Similarly, unoriented, complex, or spin cobordism groups can be defined.

### Elliptic Cohomology

Quillen proved that complex cobordism theory is universal for all complex oriented cohomology theories, that is, those cohomology theories that allow a theory of Chern classes. In a complex oriented theory, the first Chern class of the tensor product of two line bundles can be expressed in terms of the first Chern class of each of them via a two-variable power series: $c_1(E \otimes E') = F(c_1(E), c_1(E'))$. $F$ defines a formal group law and Quillen's theorem asserts that the one arising from complex cobordism theory is the universal one.

Vice versa, given a formal group law, one may try to construct a complex oriented cohomology theory from it. In particular, an elliptic curve gives rise to a formal group law and an elliptic cohomology theory. Hopkins *et al.* have described and studied an inverse limit of these elliptic theories, which they call the theory of topological modular forms, tmf, as the theory is closely related to modular forms. In particular, there is a natural map from the groups $\mathrm{tmf}_{2n}(\mathrm{pt})$ to the group of modular forms of weight $n$ over $\mathbb{Z}$. After inverting a certain element (related to the discriminant), the theory becomes periodic with period $24^2 = 576$.

Witten (1998) showed that the purely theoretically constructed elliptic cohomology theories should play an important role in string theory: the index of the Dirac operator on the free loop space of certain manifolds should be interpreted as an element of it. But unlike for ordinary cohomology, $K$-theory, and cobordism theory we do not (yet) know a good geometric object representing elements in this theory without which its use for geometry and analysis remains limited. Segal speculated some 20 years ago that conformal field theories should define such geometric objects. Though progress has been made, the search for a good geometric interpretation of elliptic cohomology (and tmf) remains an active and important research area.

### Infinite Loop Spaces

Brown's representability theorem implies that for each (reduced) generalized cohomology theory $h^*$ we can find a sequence of spaces $E_*$ such that $h^n(X)$ is the set of homotopy classes $[X, E_n]$ from the space $X$ to $E_n$ for all $n$. Recall that the Mayer–Vietoris sequence implies that $h^n(X) \simeq h^{n+1}(\Sigma X)$. The suspension functor $\Sigma$ is adjoint to the based loop space functor $\Omega$ which takes a space $X$ to the space of based maps from the circle to $X$. Hence,

$$\begin{aligned} h^n(X) &= [X, E_n] = [\Sigma X, E_{n+1}] \\ &= [X, \Omega E_{n+1}] \end{aligned}$$

and it follows that every generalized cohomology theory is represented by an infinite loop space

$$E_0 \simeq \Omega E_1 \simeq \cdots \simeq \Omega^n E_n \simeq \cdots$$

Vice versa, any such infinite loop space gives rise to a generalized cohomology theory.

One may think of infinite loop spaces as the abelian groups up to homotopy in the strongest sense. Indeed, ordinary cohomology with integer coefficients is represented by

$$\mathbb{Z} \simeq \Omega S^1 \simeq \Omega^2 \mathbb{C}P^\infty \simeq \cdots \simeq \Omega^n K(n, \mathbb{Z}) \simeq \cdots$$

where by definition the Eilenberg–MacLane space $K(n, \mathbb{Z})$ has trivial homotopy groups for all dimensions not equal to $n$ and $\pi_n K(n, \mathbb{Z}) = \mathbb{Z}$. Complex $K$-theory is represented by

$$\mathbb{Z} \times BU \simeq \Omega(U) \simeq \Omega^2(BU) \simeq \Omega^3(U) \simeq \cdots$$

This is Bott's celebrated "periodicity theorem." Finally, oriented cobordism theory is represented by

$$\Omega^\infty MSO := \lim_{n\to\infty} \Omega^n \mathrm{Th}(\gamma_n)$$

where $\gamma_n \to BSO_n$ is the universal $n$-dimensional vector space over the Grassmannian manifold of oriented $n$-planes in $\mathbb{R}^\infty$, and $\mathrm{Th}(\gamma_n)$ denotes its Thom space.

A good source of infinite loop spaces are symmetric monoidal categories. Indeed every infinite loop space can be constructed from such a category: the symmetric monoidal structure gives the corresponding homotopy abelian group structure. For example, the category of finite-dimensional, complex vector spaces and their isomorphisms gives rise to $\mathbb{Z} \times BU$. To give another example, in quantum field theory, one considers the $(d+1)$-dimensional cobordism category with objects the compact, oriented $d$-dimensional manifolds, and their $(d+1)$-dimensional cobordisms as morphisms. Disjoint union of manifolds makes this category into a symmetric monoidal category. The associated infinite loop space and hence generalized cohomology theory has recently been identified as a $(d+1)$-dimensional slice of oriented cobordism theory (Galatius *et al.* preprint 2005).

*See also:* Characteristic Classes; Equivariant Cohomology and the Cartan Model; Functional Equations and Integrable Systems; Index Theorems; Intersection Theory; *K*-Theory; Moduli Spaces: An Introduction; Riemann Surfaces; Spectral Sequences.

## Further Reading

Adams F (1978) Infinite Loop spaces. *Annals of Mathematical Studies* 90: PUP.
Bott R and Tu L (1982) *Differential Forms in Algebraic Topology*. Springer.
Galatius, Madsen, Tillmann, Weiss (2005).
Hatcher A (2002) *Algebraic Topology*, (http://www.math.cornell.edu). Cambridge: Cambridge University Press.
Madsen, Weiss (2004).
Mosher R and Tangora M (1968) *Cohomology Operations and Applications in Homotopy Theory*. Harper and Row.
Viehweg (1999) Aspects of Mathematics E33.
Witten (1998) Journal of Higher Energy Physics 12.
Witten (1998) Springer Lecture Notes in Mathematics, vol. 1326.

# Combinatorics: Overview

**C Krattenthaler**, Universität Wien, Vienna, Austria



## Introduction

Combinatorics is a vast field which enters particularly in a crucial way in statistical physics. There, it is particularly the enumerative problems that are of importance. Therefore, in this article, we shall mainly concentrate on the enumerative aspects of combinatorics. We first recall the basic terminology, in particular the basic combinatorial objects and numbers, together with the simplest facts about them. We then provide introductions into the most important techniques of enumeration: the generating function technique, Redfield–Pólya theory, methods of solving functional equations of combinatorial origin, methods of asymptotic enumeration, the theory of heaps, and the transfer matrix method. The subsequent sections then discuss specific problem circles with relation to statistical physics more closely. We discuss lattice path problems, explain Kasteleyn's method of enumerating perfect matchings and tilings, present the fundamental theorems on nonintersecting paths, and provide an introduction into the research field involving vicious walkers, plane partitions, rhombus tilings, alternating sign matrices, six-vertex configurations, and fully packed loop configurations. Finally, we explain how one should treat binomial and hypergeometric series, which frequently arise in enumeration problems.

## Basic Combinatorial Terminology

In this section we review basic combinatorial notions and facts. The reader can find a more detailed treatment and further results, for example, in chapter 1 of Stanley (1986).

The basic combinatorial choice problems and their solutions are: there are $2^n$ subsets of an $n$-element set. There are $\binom{n}{k}$ $k$-element subsets of an $n$-element set. Given an alphabet $\mathcal{A} = \{a_1, a_2, \ldots\}$, a word is a (finite or infinite) sequence of elements of $\mathcal{A}$. Usually, a finite word is written in the form $w_1 w_2 \ldots w_n$ (with $w_i \in \mathcal{A}$). Out of the letters $\{1, 2, \ldots, k\}$, one can build $k^n$ words of length $n$. Out of the letters $\{1, 2, \ldots, k\}$, one can build $\binom{n+k-1}{n}$ increasing sequences of length $n$. The number of permutations of an $n$-element set is $n!$. The set of permutations of $\{1, 2, \ldots, n\}$ is denoted by $\mathfrak{S}_n$. The number of permutations of an $n$-element set with exactly $k$ cycles is the Stirling number of the first kind, $s(n, k)$. These numbers are given as the expansion coefficients of falling factorials,

$$x(x-1)\cdots(x-n+1) = \sum_{k=0}^{n} (-1)^{n-k} s(n,k) x^k$$

or in form of the double (formal) power series

$$\sum_{n,k \geq 0} s(n,k) x^k \frac{y^n}{n!} = (1+y)^x$$

A partition of a set is a collection of pairwise disjoint subsets the union of which is the complete set. The subsets in the collection are called the blocks of the partition. The total number of partitions of an $n$-element set is the Bell number $B_n$. These numbers are given by

$$\sum_{n \geq 0} B_n \frac{x^n}{n!} = \mathrm{e}^{\mathrm{e}^x - 1}$$

The number of partitions of an $n$-element set into exactly $k$ blocks is the Stirling number of the second kind, $S(n, k)$. These numbers are given by

$$\sum_{n,k \geq 0} S(n,k) x^k \frac{y^n}{n!} = \mathrm{e}^{x(\mathrm{e}^y - 1)}$$

or, explicitly, by

$$S(n,k) = \frac{1}{k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} j^n$$

A composition of a positive integer $n$ is a representation of $n$ as a sum $n = s_1 + s_2 + \cdots + s_k$ of other positive integers $s_i$, where the order of the summands matters. The total number of compositions of $n$ is $2^{n-1}$. The number of compositions of $n$ with exactly $k$ summands is $\binom{n-1}{k-1.}$ A partition of a positive integer $n$ is a representation of $n$ as a sum $n = \lambda_1 + \lambda_2 + \cdots + \lambda_k$ of other positive integers $\lambda_i$, where the order of the summands does matter. Thus, we may assume that the summands are ordered, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$. This is the motivation to write partitions most often in the form of tuples $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ the entries of which are weakly decreasing. The summands of a partition are called the parts of the partition. Let $p(n)$ denote the number of partitions of $n$. These numbers are given by

$$\sum_{n=0}^{\infty} p(n) x^n = \frac{1}{\prod_{i=1}^{\infty} (1 - x^i)}$$

If $p(n, k)$ denotes the number of partitions of $n$ into at most $k$ parts, then we have

$$\sum_{n=0}^{\infty} p(n,k) x^n = \frac{1}{\prod_{i=1}^{k} (1 - x^i)}$$

Finally, if $p(n, k, m)$ denotes the number of partitions of $n$ into at most $k$ parts, all of which are at most $m$, then

$$\sum_{n \geq 0} p(n,k,m) x^n = \frac{(1 - x^{k+m})(1 - x^{k+m-1}) \cdots (1 - x^{m+1})}{(1 - x^k)(1 - x^{k-1}) \cdots (1 - x)}$$

The expression on the right-hand side is called q-binomial coefficient, and is denoted by $\left[ \begin{matrix} k+m \\ k \end{matrix} \right]_x$.

Partitions are frequently encoded in terms of their Ferrers diagrams. The Ferrers diagram of a partition $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_\ell)$ is an array of cells with $\ell$ left-justified rows and $\lambda_i$ cells in row $i$. For example, the diagram in **Figure 1** is the Ferrers diagram of the partition $(3, 3, 2)$.

A lattice path P in $\mathbb{Z}^d$ (where $\mathbb{Z}$ denotes the set of integers) is a path in the $d$-dimensional integer lattice $\mathbb{Z}^d$ which uses only points of the lattice, that is, it is a sequence $(P_0, P_1, \ldots, P_l)$, where $P_i \in \mathbb{Z}^d$ for all $i$. The vectors $\overrightarrow{P_0 P_1}, \overrightarrow{P_1 P_2}, \ldots, \overrightarrow{P_{l-1} P_l}$ are called the steps of $P$. The number of steps, $l$, is called the length of $P$. **Figure 2** shows a lattice path in $\mathbb{Z}^2$ of length 11.

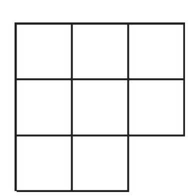


**Figure 1** A Ferrers diagram.

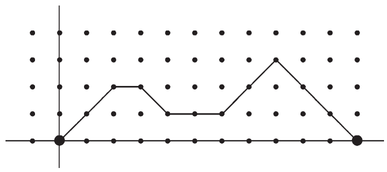


**Figure 2** A Motzkin path.

A Dyck path is a lattice path in the integer plane $\mathbb{Z}^2$ consisting of up-steps $(1,1)$ and down-steps $(1,-1)$, which starts at the origin, never passes below the $x$-axis, and ends on the $x$-axis. See Figure 3 for an example.

The number of Dyck paths of length $2n$ is the Catalan number

$$C_n = \frac{1}{n+1}\binom{2n}{n}$$

The generating function (see the next section for an introduction to the theory of generating functions) for these numbers is

$$\sum_{n=0}^{\infty} C_n x^n = \frac{1-\sqrt{1-4x}}{2x} \qquad [1]$$

The reader is referred to exercise 6.19 in Stanley (1999) for countless occurrences of the Catalan numbers.

A Motzkin path is a lattice path in the integer plane $\mathbb{Z}^2$ consisting of up-steps $(1,1)$, level steps $(1,0)$, and down-steps $(1,-1)$, which starts at the origin, never passes below the $x$-axis, and ends on the $x$-axis. The path in Figure 2 is in fact a Motzkin path. The number of Motzkin paths of length $n$ is the Motzkin number

$$M_n = \sum_{k\geq 0} \frac{1}{k+1}\binom{2k}{k}\binom{n}{2k}$$

The generating function for these numbers is

$$\sum_{n=0}^{\infty} M_n x^n = \frac{1-x-\sqrt{1-2x-3x^2}}{2x^2} \qquad [2]$$

The reader is referred to exercise 6.38 in Stanley (1999) for numerous occurrences of the Motzkin numbers.

A Schröder path is a lattice path in the integer plane $\mathbb{Z}^2$ consisting of horizontal steps $(1,0)$ and

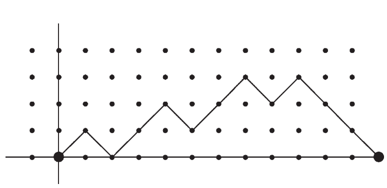


**Figure 3** A Dyck path.

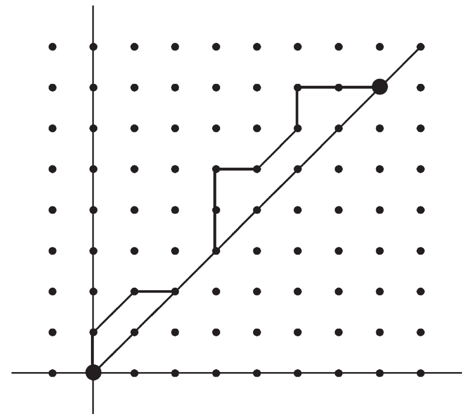


**Figure 4** A Schröder path.

vertical steps $(0,1)$, which starts at the origin, never passes below the diagonal $x=y$, and ends on the diagonal $x=y$. See Figure 4 for an example.

The number of Schröder paths of length $n$ is the (large) Schröder number

$$S_n = \sum_{k\geq 0} \frac{1}{k+1}\binom{2k}{k}\binom{n+k}{2k}$$

The generating function for these numbers is

$$\sum_{n=0}^{\infty} S_n x^n = \frac{1-x-\sqrt{1-6x+x^2}}{2x} \qquad [3]$$

The reader is referred to exercise 6.39 in Stanley (1999) for numerous occurrences of the Schröder numbers.

There is another famous sequence of numbers which we did not touch yet, the Fibonacci numbers $F_n$. They are given by

$$F_n = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+1}$$

with generating function

$$\sum_{n=0}^{\infty} F_n x^n = \frac{1}{1-x-x^2} \qquad [4]$$

They also occur in numerous places. For example, the number $F_n$ counts all paths on the integers $\mathbb{Z}$ from 0 to $n$ with steps $(1,0)$ and $(2,0)$.

An undirected graph G consists of vertices and edges. An edge is a two-element subset of the vertices, which, however, is thought of as a line or curve connecting the two vertices. See Figure 5a for an example. The usual notation for a graph $G$ is $G=(V,E)$, where $V$ is the set of vertices and $E$ is the set of edges of $G$. A graph is planar if it is

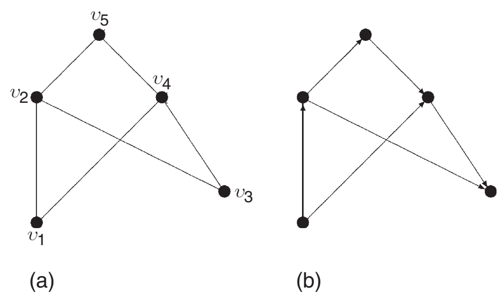



**Figure 5** (a) An undirected graph. (b) A directed graph.

embedded in the plane (sphere) in such a way that the curves which mark the edges do not intersect in their interiors. There can be several different ways to embed the same graph in the plane (or in another surface). When we speak of a planar graph then we assume the graph already to be embedded in a given way. For example, the graph in Figure 5 is not a planar graph, by its drawing. However, there is a different embedding which is planar (namely, all embeddings which put the vertex $v_3$ above the vertex $v_5$ and leave the other vertices as they are). A tree is a graph without any cycles.

A directed graph (or digraph) $G$ consists of vertices and arcs (which are sometimes also called directed edges). An arc is a pair of vertices, which, however, is thought of an arrow pointing from the first vertex of the pair to the second. See Figure 5b for an example. The usual notation for a directed graph $G$ is again $G=(V,E)$, where $V$ is the set of vertices and $E$ is the set of arcs of $G$. All other notions explained for undirected graphs have analogous meanings for directed graphs.

Graphs can be labeled, in which case each vertex is assigned a label, or unlabeled. The (undirected) graph in Figure 5a is labeled, whereas the (directed) graph in Figure 5b is unlabeled.

## Generating Functions

Generating functions are the very basic tools of enumeration. For introductions to this technique, from different points of view, the reader is referred to Bergeron *et al.* (1998), Flajolet and Sedgewick (chapter 1 in the reference listed in "Further reading" section), and Stanley (1998, chapter 1; 1999, chapter 4).

Let $\mathcal{A}$ be a set of (unlabeled) objects. Each object $a$ in $\mathcal{A}$ has a certain size, $|a|$, which is a non-negative integer. Let us also assume that there is only a finite number of objects from $\mathcal{A}$ of a given size. Let $a_n$ be the number of objects from $\mathcal{A}$ of size $n$. The (ordinary) generating function for $\mathcal{A}$ is the formal power series

$$F_{\mathcal{A}}(x) = \sum_{a\in\mathcal{A}} x^{|a|} = \sum_{n=0}^{\infty} a_n x^n$$

("formal" means that $x$ is just an indeterminate, not a real or complex number. One can compute with formal power series in the same way as with analytic series, only that convergence issues do not arise, respectively that "convergence" has a different meaning; cf. Stanley (1998, section 1.1)) Typical examples are **`Sets`** (the collection containing all "unlabeled sets," that is, all objects of the form $\{\bullet,\bullet,\ldots,\bullet\}$, including the empty set, where the size of $\{\bullet,\bullet,\ldots,\bullet\}$ is the number of $\bullet$'s), **`Sequences`** (the collection containing all "unlabeled sequences," that is, all objects of the form $(\bullet,\bullet,\ldots,\bullet)$, including the empty sequence), **`Cycles`** (unlabeled cycles), with respective generating function

$$F_{\texttt{Sets}}(x) = F_{\texttt{Sequences}}(x) = \frac{1}{1-x} \qquad F_{\texttt{Cycles}}(x) = \frac{x}{1-x} \tag{5}$$

or **`Trees`** (unlabeled trees).

If $\mathcal{A}$ and $\mathcal{B}$ are two sets of objects, one can define several other sets of objects using them. The union of $\mathcal{A}$ and $\mathcal{B}$, written $\mathcal{A}\cup\mathcal{B}$, has as a groundset the disjoint union of $\mathcal{A}$ and $\mathcal{B}$, and the size of an element from $\mathcal{A}$ is its size in $\mathcal{A}$, while the size of an element from $\mathcal{B}$ is its size in $\mathcal{B}$. We have

$$F_{\mathcal{A}\cup\mathcal{B}}(x) = F_{\mathcal{A}}(x) + F_{\mathcal{B}}(x) \tag{6}$$

The product of $\mathcal{A}$ and $\mathcal{B}$, written $\mathcal{A}\times\mathcal{B}$, has as a groundset the set of pairs $\mathcal{A}\times\mathcal{B}$, and the size of an element $(a,b)$ from $\mathcal{A}\times\mathcal{B}$ is the sum of the sizes of $a$ (in $\mathcal{A}$) and of $b$ (in $\mathcal{B}$). We have

$$F_{\mathcal{A}\times\mathcal{B}}(x) = F_{\mathcal{A}}(x)\cdot F_{\mathcal{B}}(x) \tag{7}$$

The substitution of two sets $\mathcal{A}$ and $\mathcal{B}$ of objects can only be defined in certain circumstances, and only in certain more restrictive circumstances the generating function for the substitution can be computed by substituting the generating functions for $\mathcal{A}$ and $\mathcal{B}$. Let us assume that any object $a$ from $\mathcal{A}$ of size $n$, by its structure, has $n$ atoms (nodes). For example, if $\mathcal{A}$ is a certain set of trees, where the size of a tree is the number of leaves in the tree, then we may take, as the atoms, the leaves of the tree. In this situation, the substitution of $\mathcal{B}$ *in* $\mathcal{A}$, denoted by $\mathcal{A}(\mathcal{B})$, is the set of objects which arises by replacing the atoms of objects from $\mathcal{A}$ by objects from $\mathcal{B}$ in all possible ways. The size of an object from $\mathcal{A}(\mathcal{B})$ is the sum of the sizes of the objects from $\mathcal{B}$ that it

contains. In order that $\mathcal{A}(\mathcal{B})$ contains only a finite number of objects of a given size, we must assume that $\mathcal{B}$ contains no elements of size 0. If, in addition, the atoms of any element $a$ from $\mathcal{A}$ inherit an order (e.g., if $\mathcal{A}$ is a set of binary trees, then the leaves of a binary tree are ordered in a natural way from "left" to "right"), then we have

$$F_{\mathcal{A}(\mathcal{B})}(x) = F_{\mathcal{A}}(F_{\mathcal{B}}(x)) \qquad [8]$$

However, this equation is not true in general. The general formula comes out of Redfield–Pólya theory (see [21] and [24]) and requires the notion of cycle index series. For example, if $\mathcal{B}$ is the set of connected (unlabeled) graphs, $\mathcal{A}$ is **Sets**, so that $\mathcal{A}(\mathcal{B})$ is the set of all (connected and disconnected) graphs, then [8] is not true, but what is true is

$$F_{\texttt{Sets}(\mathcal{B})} = \exp\left(F_{\mathcal{B}}(x) + \tfrac{1}{2}F_{\mathcal{B}}(x^2) + \tfrac{1}{3}F_{\mathcal{B}}(x^3) + \cdots\right) \qquad [9]$$

This holds, in fact, for any set $\mathcal{B}$ of unlabeled objects. (This is seen by combining [24], [17], and [21].)

Next we deal with the enumeration of labeled objects. Let $\mathcal{A}$ be a set of labeled objects, again, each object $a$ with a certain size $|a|$ which is a nonnegative integer. "Labeled" means that each object of size $n$, by its structure, comes with $n$ atoms (nodes) which are labeled $1, 2, \ldots, n$. For example, $\mathcal{A}$ may be the set of all labeled graphs, where the size of a graph is the number of its vertices, and where the vertices are labeled $1, 2, \ldots, n$. Again, we assume that there is only a finite number of objects from $\mathcal{A}$ of a given size. Let $a_n$ be the number of objects from $\mathcal{A}$ of size $n$. The exponential generating function for $\mathcal{A}$ is the formal power series

$$E_{\mathcal{A}}(x) = \sum_{a \in \mathcal{A}} \frac{x^{|a|}}{|a|!} = \sum_{n=0}^{\infty} a_n \frac{x^n}{n!}$$

Typical examples are **Sets** (the collection containing all "labeled sets," that is all objects of the form $\{1, 2, \ldots, n\}$, including the empty set), **Permutations**, **Cycles** (labeled cycles), with respective generating functions

$$E_{\texttt{Sets}}(x) = \exp(x) \qquad [10]$$

$$E_{\texttt{Permutations}}(x) = \frac{1}{1-x} \qquad [11]$$

$$E_{\texttt{Cycles}}(x) = \log \frac{1}{1-x} \qquad [12]$$

or **Trees** (labeled trees). The explicit form of the generating function for **Trees** is discussed in the section "Solving equations for generating functions: the Lagrange inversion formula and the kernel method."

If $\mathcal{A}$ and $\mathcal{B}$ are two sets of objects, one defines again several other sets of objects using them. The union of $\mathcal{A}$ and $\mathcal{B}$, written $\mathcal{A} \cup \mathcal{B}$, has as a groundset the disjoint union of $\mathcal{A}$ and $\mathcal{B}$, and the size of an element from $\mathcal{A}$ is its size in $\mathcal{A}$, while the size of an element from $\mathcal{B}$ is its size in $\mathcal{B}$. We have

$$E_{\mathcal{A} \cup \mathcal{B}}(x) = E_{\mathcal{A}}(x) + E_{\mathcal{B}}(x) \qquad [13]$$

To define the product of $\mathcal{A}$ and $\mathcal{B}$, written $\mathcal{A} \times \mathcal{B}$, we cannot simply take $\mathcal{A} \times \mathcal{B}$ as a groundset, we must also say something about the labeling of the objects. So, as a groundset we take all pairs $(a, b)$ with $a \in \mathcal{A}$ and $b \in \mathcal{B}$, but labeled in all possible ways by $1, 2, \ldots, |a| + |b|$ such that the order of labels assigned to $a$ respects the original order of labels of $a$, and the same for $b$. The size of such an element $(a,b)$ is again the sum of the sizes of $a$ (in $\mathcal{A}$) and of $b$ (in $\mathcal{B}$). We have

$$E_{\mathcal{A} \times \mathcal{B}}(x) = E_{\mathcal{A}}(x) \cdot E_{\mathcal{B}}(x) \qquad [14]$$

Since, in the labeled world, objects come automatically with atoms, the substitution of two sets $\mathcal{A}$ and $\mathcal{B}$ of objects can now always be defined. The substitution of $\mathcal{B}$ in $\mathcal{A}$, denoted by $\mathcal{A}(\mathcal{B})$, is the set of objects which arises by replacing the atoms of objects from $\mathcal{A}$ by objects from $\mathcal{B}$ in all possible ways, and labeling the substituted objects in all possible ways by $1, 2, \ldots, \sum_b |b|$ (the sum being over the objects from $\mathcal{B}$ which were put in the places of the atoms) that are consistent with the original labelings of the objects from $\mathcal{B}$. The size of an object from $\mathcal{A}(\mathcal{B})$ is the sum of the sizes of the objects from $\mathcal{B}$ that it contains. In order that $\mathcal{A}(\mathcal{B})$ contains only a finite number of objects of a given size, we must assume that $\mathcal{B}$ contains no elements of size 0. Then we have

$$E_{\mathcal{A}(\mathcal{B})}(x) = E_{\mathcal{A}}(E_{\mathcal{B}}(x)) \qquad [15]$$

An example of a composition is

$$\texttt{Permutations} = \texttt{Sets}(\texttt{Cycles})$$

Thus, from [15] we have

$$E_{\texttt{Permutations}}(x) = E_{\texttt{Sets}}(E_{\texttt{Cycles}}(x))$$

corresponding to the identity

$$\frac{1}{1-x} = \exp(\log 1/(1-x))$$

Another manifestation of the composition rule is, for example, the fact (which is sometimes called the "exponential principle") that, if one takes the log of the partition function for some maps, the result is the partition function for the connected maps among them.

All of the above can be generalized to a weighted setting. Namely, if $\mathcal{A}$ is a set of objects (labeled or unlabeled), and if $w:\mathcal{A}\to R$ is a weight function from $\mathcal{A}$ into some ring $R$, then all of the above remains true, if we replace the definitions of $F_{\mathcal{A}}(x)$ and $E_{\mathcal{A}}(x)$ above by the weighted sums

$$F_{\mathcal{A}}(x) = \sum_{a\in\mathcal{A}} w(a)x^{|a|}$$

and

$$E_{\mathcal{A}}(x) = \sum_{a\in\mathcal{A}} w(a)\frac{x^{|a|}}{|a|!}$$

respectively, if in the definition of the union of $\mathcal{A}$ and $\mathcal{B}$ we define the weight of an object to be its weight in $\mathcal{A}$, respectively $\mathcal{B}$, if in the definition of the product of $\mathcal{A}$ and $\mathcal{B}$ we define the weight of an object $(a, b)$ to be the product of the weights of $a$ and $b$, and if in the definition of the substitution we define the weight of an object in $\mathcal{A}(\mathcal{B})$ as the product of the weights of the objects from $\mathcal{B}$ that were put in place of the atoms.

## Redfield–Pólya Theory of Colored Enumeration

The natural and uniform environment for the separate treatment of generating functions for unlabeled and labeled objects in the last section is the theory for counting colored objects founded by Redfield and Pólya, in the modern treatment through cycle index series due to Joyal. We refer the reader to Bergeron *et al.* (1998, appendix 1), de Bruijn (1981), and Stanley (1999, chapter 7) for further reading.

Let $\mathcal{A}$ be a set of labeled objects with the constraint that there is only a finite number of objects of a given size. The cycle index series for $\mathcal{A}$ is the formal multivariable series

$$Z_{\mathcal{A}}(x_1,x_2,\ldots) = \sum_{n=0}^{\infty}\frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n}\mathrm{fix}_\sigma(\mathcal{A})x_1^{c_1(\sigma)}x_2^{c_2(\sigma)}x_3^{c_3(\sigma)}\cdots \qquad [16]$$

where $\mathrm{fix}_\sigma(\mathcal{A})$ is the number of objects $a$ from $\mathcal{A}$ that remain invariant when the labels are permuted according to the permutation $\sigma$ (in particular, if $\sigma\in\mathfrak{S}_n$, the size of $a$ must be $n$ in order that $\sigma$ can be applied to the labels), and where $c_i(\sigma)$ denotes the number of cycles of length $i$ of $\sigma$.

In most cases, it is difficult to obtain compact expressions for the cycle index series. However, for our familiar families of objects, compact expressions are available:

$$Z_{\mathtt{Sets}}(x_1,x_2,\ldots) = \exp\left(x_1+\frac{x_2}{2}+\frac{x_3}{3}+\cdots\right) \qquad [17]$$

$$Z_{\mathtt{Permutations}}(x_1,x_2,\ldots) = \prod_{i=1}^{\infty}\frac{1}{1-x_i} \qquad [18]$$

$$Z_{\mathtt{Cycles}}(x_1,x_2,\ldots) = \sum_{i=1}^{\infty}\frac{\Phi(i)}{i}\log\frac{1}{1-x_i} \qquad [19]$$

where $\Phi(i)$ is the Euler totient function (the number of positive integers $j\le i$ relatively prime to $i$).

What makes the cycle index series so fundamental is the fact that the generating functions from the last section are specializations of it. Namely, the exponential generating function for $\mathcal{A}$ is equal to

$$E_{\mathcal{A}}(x) = Z_{\mathcal{A}}(x,0,0,\ldots) \qquad [20]$$

If, given the set of labeled objects $\mathcal{A}$, we produce a set of unlabeled objects $\tilde{\mathcal{A}}$ by taking all the objects from $\mathcal{A}$ but forgetting the labels, then the ordinary generating function for $\tilde{\mathcal{A}}$ is another specialization of the cycle index series,

$$F_{\tilde{\mathcal{A}}}(x) = Z_{\mathcal{A}}(x,x^2,x^3,\ldots) \qquad [21]$$

The cycle index series satisfies the following properties with respect to union, product and composition of sets of objects:

$$Z_{\mathcal{A}\cup\mathcal{B}}(x_1,x_2,\ldots) = Z_{\mathcal{A}}(x_1,x_2,\ldots) + Z_{\mathcal{B}}(x_1,x_2,\ldots) \qquad [22]$$

$$Z_{\mathcal{A}\times\mathcal{B}}(x_1,x_2,\ldots) = Z_{\mathcal{A}}(x_1,x_2,\ldots) \times Z_{\mathcal{B}}(x_1,x_2,\ldots) \qquad [23]$$

$$Z_{\mathcal{A}(\mathcal{B})}(x_1,x_2,\ldots) = Z_{\mathcal{A}}(Z_{\mathcal{B}}(x_1,x_2,x_3,\ldots), Z_{\mathcal{B}}(x_2,x_4,x_6\ldots), Z_{\mathcal{B}}(x_3,x_6,x_9,\ldots),\ldots) \qquad [24]$$

Similar to the theory of generating functions surveyed in the last section, one can also develop a weighted version of the cycle index series. Given a set of labeled objects $\mathcal{A}$, where each object $a$ is assigned a weight $w(a)$, one changes the definition [16] insofar as $\mathrm{fix}_\sigma(\mathcal{A})$ gets replaced by the weighted sum $\sum_{\sigma(a)=a} w(a)$, where $\sigma(a)$ means the object arising from $a$ by permuting the labels according to $\sigma$. Then all the above formulas remain true in this weighted setting.

Cycle index series are instrumental in the enumeration of colored objects. The basic situation is that we have given a set $\tilde{\mathcal{A}}$ of unlabeled objects so that every object of size $n$ comes with $n$ atoms (nodes). For example, we may think of $\tilde{\mathcal{A}}$ as the set of cycles. We are now going to color each atom by a

color from the set of colors $C$. The question that we pose is: *how many different colored objects of a given size are there?* In our example, if $C$ consists of the two colors "black" and "white," then we are asking the question of how many necklaces one can make out of $n$ pearls that can be black or white. In terms of generating functions, we want to compute

$$\Gamma_{\tilde{\mathcal{A}}}(x) = \sum_{c} x^{|c|}$$

where the sum is over all colored objects $c$ that one can obtain by coloring the objects from $\tilde{\mathcal{A}}$.

The central result of Redfield–Pólya theory is that, if $\mathcal{A}$ is the set of labeled objects that one obtains from $\tilde{\mathcal{A}}$ by labeling the objects of $\tilde{\mathcal{A}}$ in all possible ways, then

$$\Gamma_{\tilde{\mathcal{A}}}(x) = Z_{\mathcal{A}}(|C|x, |C|x^2, |C|x^3, \ldots)$$

There is again a weighted version. One allows the objects $a$ from $\tilde{\mathcal{A}}$ to have weight $w(a) \in R$. Moreover, one assumes a weight function $f: C \to R$ on the colors with values in the ring $R$. One defines the weight of a colored object obtained by coloring the atoms of $a$ to be $w(a)$ multiplied by the product of all $f(\gamma)$, where $\gamma$ ranges over all the colors of the atoms (including repetitions of colors). Let $\Gamma_{\tilde{\mathcal{A}}}(w, f)$ denote the sum of all the weights of all colored objects obtained from $\tilde{\mathcal{A}}$. Then

$$\Gamma_{\tilde{\mathcal{A}}}(w,f) = Z_{\mathcal{A}}\left(\sum_{c\in C} f(c), \sum_{c\in C} f(c)^2, \sum_{c\in C} f(c)^3, \ldots\right)$$

We remark that these results cover also the case of enumeration of objects under a group action. This includes the enumeration of objects on which we impose certain symmetries. See Bergeron *et al.* (1998, appendix 1), de Bruijn (1981), and Stanley (1999, chapter 7) for more details. The enumeration of asymmetric objects is the subject of an ongoing research program (cf. Labelle and Lamathe (2004)).

## Solving Equations for Generating Functions: The Lagrange Inversion Formula and the Kernel Method

In this section, we describe two methods to solve functional equations for generating functions. The Lagrange inversion makes it possible (in some situations) to find explicit expressions for the coefficients of an implicitly given series. The kernel method (and its extensions), on the other hand, is a powerful method to obtain an explicit expression for an implicitly given function. We refer the reader to Flajolet and Sedgewick, (section VII.5 of the reference in "Further reading" section) for further reading.

In many situations it will happen that, when we apply the methods from the last section, we end up with a functional equation for the generating function $f(x) = \sum_{n=0}^{\infty} f_n x^n$ that we wanted to compute. For example, if $t_n$ denotes the number of labeled rooted trees with $n$ nodes, and if we write $T(x) = \sum_{n=1}^{\infty} t_n x^n/n!$, then, by applying a straightforward decomposition of a tree into its root and its set of subtrees attached to the root, we obtain the equation

$$T(z) = z\exp(T(z)) \qquad [25]$$

How does one solve such an equation? As a matter of fact, for $T(z)$, there is no expression in terms of known functions. However, the Lagrange inversion formula enables one to find the coefficients $t_n/n!$ of $T(z)$ explicitly. The theorem reads as follows.

**Theorem** *Let $g(x)$ be a formal Laurent series containing only a finite number of negative powers of $x$, and let $f(x)$ be a formal power series without constant term. If we expand $g(x)$ in powers of $f(x)$,*

$$g(x) = \sum_{k} c_k f^k(x) \qquad [26]$$

*then the coefficients $c_n$ are given by*

$$c_n = \frac{1}{n}[x^{-1}]g'(x)f^{-n}(x) \quad \text{for } n \neq 0 \qquad [27]$$

*or, alternatively, by*

$$c_n = [x^{-1}]g(x)f'(x)f^{-n-1}(x) \qquad [28]$$

*Here, $[x^n]h(x)$ denotes the coefficient of $x^n$ in the power series $h(x)$.*

With this theorem in hand, eqn [25] is easy to solve. We write it in the form

$$T(x)\exp(-T(x)) = x \qquad [29]$$

We want to know the coefficients in the expansion $T(x) = \sum_{n=0}^{\infty} t_n x^n/n!$. Since, by [29], $T(x)$ is the compositional inverse of $x\exp(-x)$, substitution of $x\exp(-x)$ instead of $x$ gives

$$x = \sum_{n=0}^{\infty} \frac{t_n}{n!}(x\exp(-x))^n$$

This equation is in the form [26] with $f(x) = x\exp(-x)$ and $g(x) = x$. Hence, by [27], we obtain

$$\frac{t_n}{n!} = \frac{1}{n}[x^{-1}](x\exp(-x))^{-n} = \frac{1}{n}[x^{n-1}]\exp(nx) = \frac{n^{n-1}}{n!}$$

and, thus, $t_n = n^{n-1}$.

The second method to solve functional equations which we explain in this section is the kernel method. We illustrate the method by an example. Let us consider the problem of counting Dyck paths of length $2n$ (see the section "Basic combinatorial terminology"). Rather than attempting to arrive at a solution of the problem directly, we consider the more general problem of counting the number $a_{n,k}$ of paths consisting of steps (1, 1) and (1, −1), which start at the origin, never drop below $y=0$, have length $n$, and end at height $k$. We then form the bivariate generating function $F(u,x)=\sum_{n,k\geq 0} a_{n,k}x^n u^k$. We then have the functional equation

$$F(u,x)=1+xuF(u,x)+\frac{x}{u}(F(u,x)-F(0,x)) \quad [30]$$

since a path can be empty (this explains the term 1), it can end by a step (1,1) (this explains the term $xuF(u)$), or it can end by a step (1,−1). The latter can only happen if the path before that last step did not end at height 0. The generating function for these paths is $F(u,x)-F(0,x)$, and this explains the third term in the eqn [30]. In fact, we may replace [30] by

$$F(u,x)=1+xuF(u,x)+\frac{x}{u}(F(u,x)-F_1(x)) \quad [31]$$

because [31] implies that $F_1(x)=F(0,x)$.

The idea of the kernel method is to get rid of the unknown series $F(u,x)$. This is possible because $F(u,x)$ occurs linearly in [31], which can be rewritten as

$$F(u,x)\left(1-xu-\frac{x}{u}\right)=1-\frac{x}{u}F_1(x) \quad [32]$$

We simply equate the coefficient of $F(u,x)$ in this equation to zero,

$$1-xu-\frac{x}{u}=0$$

solve this for $u$,

$$u=\frac{1-\sqrt{1-4x^2}}{2x}$$

(the other solution for $u$ makes no sense in [31]), and substitute this back in [32], to obtain

$$F_1(x)=\frac{1-\sqrt{1-4x^2}}{2x^2}$$

the familiar generating function [2] for the Catalan numbers. Now, by substituting this result in [31], we can even compute the full series $F(u,x)$.

While this was certainly a complicated, and unusual, way to compute the Catalan numbers, this approach generalizes when one considers paths with different step sets (see section VII.5 of the Flajolet and Sedgewick reference in "Further reading" section). In a more general situation, one has a functional equation

$$P(F(u,x),F_1(x),\ldots,F_k(x),x,u)=0 \quad [33]$$

where $F(u,x)$ appears linearly, as well as the unknown series $F_1(x),\ldots,F_k(x)$, whereas $x$ and $u$ appear rationally. It is clear that one can apply the same technique, namely collecting all the terms involving $F(u,x)$, equating the coefficient of $F(u,x)$ to zero, solving for $u$ and substituting back in [33]. If there is more than one function $F_i(x)$, then this will only give one equation for $F_i(x)$. However, when equating the coefficient of $F(u,x)$, which was a polynomial equation, there can be more solutions. (That was actually also the case in our example, although only one solution could be used.) All these solutions can be substituted in [33] to give many more equations for $F_i(x)$. The kernel method will work if we have enough equations to determine the unknown functions $F_i(x)$ (see the Flajolet and Sedgewick reference, section VII.5 for further details). In the variant of the "obstinate kernel method," more equations are produced in more sophisticated ways. The method has been largely extended by Bousquet–Mélou and co-workers to cover equations of the form [33], where $P$ is a polynomial such that eqn [33] determines all involved series uniquely. This extension covers in particular the so-called quadratic method due to Brown, which is of great significance in the work of Tutte on the enumeration of maps. We refer the reader to Bousquet–Mélou and Jehanne (2005) and the references given there for these extensions.

## Extracting Asymptotic Information from Generating Functions

There is powerful machinery available to extract the asymptotic behavior of the coefficients of a power series out of analytic properties of the power series. We describe the corresponding methods, singularity analysis and the saddle point method in this section. The survey by Odlyzko (1995) and the Flajolet and Sedgewick reference in "Further reading" are excellent sources for further reading, which, in particular, contain several other methods which we cannot cover here for reasons of limited space.

Let us suppose that we are interested in the asymptotic behavior of the sequence $(f_n)_{n\geq 0}$ of real (or complex) numbers as $n$ tends to infinity. Let us suppose that the power series $f(z)=\sum_{n=0}^{\infty} f_n z^n$ converges in some neighborhood of the origin. (If this series converges only at $z=0$, then either one has to try to scale, that is, for example, look at the

power series $f(z) = \sum_{n=0}^{\infty} f_n z^n/n!$ instead, or one must apply methods other than singularity analysis or the saddle point method. In the latter case, depending on the nature of the coefficients $f_n$, this may be the Euler–Maclaurin or the Poisson summation formulas, the Mellin transform technique, or other direct methods. The reader is referred to Odlyzko (1995) and the Flajolet and Sedgewick reference.) The idea is then to consider $f(z)$ as a complex function in $z$ (and extend the range of $f$ beyond the disk of convergence about the origin), and to study the singularities of $f(z)$. (The point at infinity can also be a singularity.) The upshot is that the singularities of $f(z)$ with smallest modulus dictate the asymptotic behavior of the coefficients $f_n$. These singularities of smallest modulus are called the dominating singularities.

If there is an infinite number of dominant singularities, then one has to try the circle method. We refer the reader to Andrews (1976) and Ayoub (1963) for details of this method.

If there is a finite number of dominant singularities, then there can be again two different situations, depending on whether these are "small" or "large" singularities. Roughly speaking, a singularity is small if the function $f(z)$ grows at most polynomially when $z$ approaches the singularity, otherwise it is "large." A typical example of a small singularity is $z = 1/4$ in $(1-4z)^{-1/2}$, whereas a typical example of a large singularity is $z = \infty$ in $\exp(x)$ or $z = 1$ in $\exp(1/(1-z))$.

The method to apply for small singularities is the method of singularity analysis as developed by Flajolet and Odlyzko. (Singularity analysis implies Darboux's method, which occurs frequently in the literature, and, thus, supersedes it.) For the sake of simplicity, we consider first only the case of a unique dominant singularity. We shall address the issue of several dominant singularities shortly. Furthermore, we assume the singularity to be $z = 1$, again for the sake of simplicity of presentation. The general result can then be obtained by rescaling $z$.

The basic idea is the transfer principle:

$$\text{If } f(z) \underset{z\to 1}{=} \sigma(z) + O(\tau(z)) \quad \text{then}$$
$$f_n \underset{n\to\infty}{=} \sigma_n + O(\tau_n) \qquad [34]$$

where $\sigma(z) = \sum_{n=0}^{\infty} \sigma_n z^n$ is a linear combination of standard functions of the form $(1-z)^{-\alpha}$, or logarithmic variants, and $\tau(z) = \sum_{n=0}^{\infty} \tau_n z^n$ also lies in the scale (see sections VI.3,4 of the Flajolet and Sedgewick reference for the exact statement). The expansion for $f(z)$ in [34] is called the singular expansion of $f(z)$. For the above-mentioned standard functions, we have

$$[z^n](1-z)^{-\alpha}\left(\frac{1}{z}\log\frac{1}{1-z}\right)^{\beta} \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}(\log n)^{\beta}\left(1 + \frac{C_1}{1!}\frac{\beta}{\log n} + \frac{C_2}{2!}\frac{\beta(\beta-1)}{(\log n)^2} + \cdots\right) \qquad [35]$$

where $[z^n]g(z)$ denotes the coefficient of $z^n$ in $g(z)$, and where

$$C_k = \Gamma(\alpha)\frac{d^k}{ds^k}\frac{1}{\Gamma(s)}\bigg|_{s=\alpha}$$

If $\alpha$ is a nonpositive integer, then this expansion has to be taken with care (cf. section VI.2 of the Flajolet and Sedgewick reference).

To see how this works, consider the example $f_n = \sum_{k=0}^{n}\binom{2k}{k}$. We have

$$\sum_{n=0}^{\infty} f_n z^n = \frac{1}{(1-z)\sqrt{1-4z}}$$

The function on the right-hand side is meromorphic in all of $\mathbb{C}$ (where $\mathbb{C}$ denotes the complex numbers), with singularities at $z = 1$ and $z = 1/4$. The dominant singularity is $z = 1/4$. We determine the singular expansion of $f(z)$ about $z = 1/4$,

$$f(z) = \frac{4}{3}(1-4z)^{-1/2} - \frac{4}{9}(1-4z)^{1/2} + \frac{4}{27}(1-4z)^{3/2} + O\left((1-4z)^{5/2}\right)$$

(We stopped the expansion after three terms. The farther we go, the more terms can we compute of the asymptotic expansion for $f_n$.) Hence, we obtain

$$\begin{aligned} f_n &= 4^n\left(\frac{4}{3}\frac{n^{-1/2}}{\Gamma(1/2)}\left(1 - \frac{1}{8n} + \frac{1}{128n^2}\right) - \frac{4}{9}\frac{n^{-3/2}}{\Gamma(-1/2)}\left(1 + \frac{3}{8n}\right) + \frac{4}{27}\frac{n^{-5/2}}{\Gamma(-3/2)} + O\left(n^{-7/2}\right)\right) \\ &= \frac{4^n}{\sqrt{\pi n}}\left(\frac{4}{3} + \frac{1}{18n} + \frac{11}{288n^2} + O\left(\frac{1}{n^3}\right)\right) \end{aligned}$$

If there are several small dominant singularities (but only a finite number of them), then one simply applies the above procedure for all of them and, to obtain the desired asymptotic expansion, one adds up the corresponding contributions.

The method to apply for large singularities is the saddle point method. For the following considerations, we assume that $f(z)$ is analytic in $|z| < R \le \infty$. At the heart of the saddle point method lies Cauchy's formula

$$f_n = [z^n]f(z) = \frac{1}{2\pi\,\mathrm{i}}\int_{\mathcal{C}} \frac{f(z)}{z^{n+1}}\,\mathrm{d}z \qquad [36]$$

for writing the $n$th coefficient in the power series expansion of $f(z)$. Here, $\mathcal{C}$ is some simple closed contour around the origin that stays in the range $|z| < R$. The idea is to exploit the fact that we are free to deform the contour. The aim is to choose a contour such that the main contribution to the integral in [36] comes from a very tiny part of the contour, whereas the contribution of the rest is negligible. This will be possible if we put the contour through a saddle point of the integrand $f(z)/z^{n+1}$. Under suitable conditions, the main contribution will then come from the small passage of the path through the saddle point, and the contribution of the rest will be negligible.

In practice, the saddle point method is not always straightforward to apply, but has to be adapted to the specific properties of the function $f(z)$ that we are encountering. We refer the reader to the corresponding chapters in the Flajolet and Sedgewick reference and Odlyzko (1995) for more details. There is one important exception though, namely the Hayman admissible functions. We will not reproduce the definition of Hayman admissibility because it is cumbersome (cf. section VIII.5 in the Flajolet and Sedgewick reference and definition 12.4 of Odlyzko (1995)). However, in many applications, it is not even necessary to go back to it because of the closure properties of Hayman admissible functions. Namely, it is known (cf. Odlyzko (1995), theorem 12.8) that $\exp(p(z))$ is Hayman admissible in $|z| < \infty$ for any polynomial $p(z)$ with real coefficients as long as the coefficients $a_n$ of the Taylor series of $\exp(p(z))$ are positive for all sufficiently large $n$ (thus, e.g., $\exp(z)$ is Hayman admissible), and it is known that, if $f(z)$ and $g(z)$ are Hayman admissible in $|z| < R \le \infty$, then $\exp(f(z))$ and $f(z)g(z)$ are also (thus, e.g., $\exp(\exp(z) - 1)$ is Hayman admissible).

The central result of Hayman's theory is the following: if $f(z) = \sum_{n\ge 0} f_n z^n$ is Hayman admissible in $|z| < R$, then

$$f_n \sim \frac{f(r_n)}{r_n^n\sqrt{2\pi b(r_n)}} \quad \text{as} \quad n \to \infty \qquad [37]$$

where $r_n$ is the unique solution for large $n$ of the equation $a(r) = n$ in $(R_0, R)$, with $a(r) = rf'(r)/f(r)$, $b(r) = ra'(r)$, and a suitably chosen constant $R_0 > 0$.

This result covers only the first term in the asymptotic expansion. There is an even more sophisticated theory due to Harris and Schoenfeld, which allows one to also find a complete asymptotic expansion. We refer the reader to section VIII.5 of the Flajolet and Sedgewick reference and Odlyzko (1995) for more details.

Methods for the asymptotic analysis of multivariable generating functions are also available (see the corresponding chapters in Flajolet and Sedgewick, Odlyzko (1995) and the recent important development surveyed in the Pemantle and Wilson reference listed in "Further reading"). We add that both the method of singularity analysis and Hayman's theory of admissible functions have been made largely automatic, and that this has been implemented in the Maple program `gdev` (see "Further reading").

## The Theory of Heaps

The theory of heaps, developed by Viennot, is a geometric rendering of the theory of the partial commutation monoid of Cartier and Foata, which is now most often called the Cartier–Foata monoid. Its importance stems from the fact that several objects which appear in statistical physics, such as Motzkin paths, animals, respectively polyominoes, or Lorentzian triangulations (see the Viennot and James reference in "Further reading" and the references therein), are in bijection with heaps.

Informally, a heap is what we would imagine. We take a collection of "pieces," say $B_1, B_2, \ldots$, and put them one upon the other, sometimes also sideways, to form a "heap," see Figure 6.

There, we imagine that pieces can only move vertically, so that the heap in Figure 6 would indeed form a stable arrangement. Note that we allow several copies of a piece to appear in a heap. (This means that they differ only by a vertical translation.) For example, in Figure 6 there appear two copies of $B_2$. Under these assumptions, there are pieces which can move past each other, and others which cannot. For example, in Figure 6, we can move the piece $B_6$ higher up, thus moving it higher than $B_1$ if we wish. However, we cannot move $B_7$ higher than $B_6$,

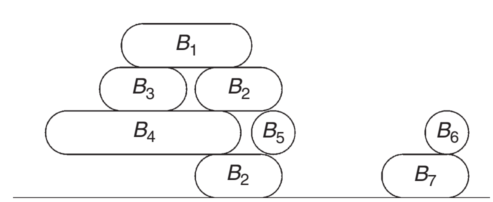



**Figure 6** A heap of pieces.

because $B_6$ blocks the way. On the other hand, we can move $B_7$ past $B_1$ (thus taking $B_6$ with us). Thus, a rigorous way to introduce heaps is by beginning with a set $\mathcal{B}$ of pieces (in our example, $\mathcal{B} = \{B_1, B_2, \ldots, B_7\}$), and we declare which pieces can be moved past another and which cannot. We indicate this by a symmetric relation $\mathcal{R}$: we write $a\mathcal{R}b$ to indicate that a cannot move past $b$ (and vice versa). When we consider a word $a_1a_2\ldots a_n$ of pieces, $a_i \in \mathcal{B}$, we think of it as putting first $a_1$, then putting $a_2$ on top of it (and, possibly, moving it past $a_1$), then putting $a_3$ on top of what we already have, etc. We declare two words to be equivalent if one arises from the other by commuting adjacent letters which are not in relation. A heap is then an equivalence class of words under this equivalence relation. What we have described just now is indeed the original definition of Cartier and Foata.

The class of heaps which occurs most frequently in applications is the class of heaps of monomers and dimers, which we now introduce. Let $\mathcal{B} = M \cup D$, where $\mathcal{M} = \{m_0, m_1, \ldots\}$ is the set of monomers and $\mathcal{D} = \{d_1, d_2, \ldots\}$ is the set of dimers. We think of a monomer $m_i$ as a point, symbolized by a circle, with $x$-coordinate $i$, see Figure 7. We think of a dimer $d_i$ as two points, symbolized by circles, with $x$-coordinates $i-1$ and $i$ which are connected by an edge, see Figure 7. We impose the relations $m_i\mathcal{R}m_i$, $m_i\mathcal{R}d_i$, $m_i\mathcal{R}d_{i+1}$, $i = 0, 1, \ldots, d_i\,\mathcal{R}d_j$, $i-1 \le j \le i$, and extend $\mathcal{R}$ to a symmetric relation. Figure 8 shows two heaps of momomers and dimers.

For example, Motzkin paths are in bijection with heaps of monomers and dimers. To see this, given a Motzkin path, we read the steps of the path from

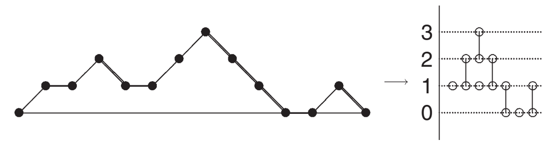



**Figure 9** Bijection between Motzkin paths and heaps of monomers and dimers.

the beginning to the end. Whenever we read a level-step at height $h$, we make it into a monomer with $x$-coordinate $h$, whenever we read a down-step from height $h$ to height $h-1$, we make it into a dimer whose endpoints have $x$-coordinates $h-1$ and $h$. Up-steps are ignored. Figure 9 shows an example. In the figure, the heap is not in "standard" fashion, in the sense that the $x$-axis is not shown as a horizontal line but as a vertical line (cf. Figure 7). But it could be easily transformed into "standard" fashion by a simple reflection with respect to a line of slope 1.

Lattice animals on the triangular lattice and on the quadratic lattice are also in bijection with heaps, this time with heaps consisting entirely out of dimers. Given an animal, one simply replaces each vertex of the animal by a dimer, see Figures 10 and 11. While in the case of animals on the triangular lattice this gives a constraintless bijection (see Figure 10), in the case of the quadratic lattice this sets up a bijection with heaps of dimers in which two dimers of the same type can never be placed directly one over the other (see Figure 11). For example, two dimers $d_5$, one placed directly over the other (as they occur in Figure 10), are forbidden under this rule.

Next we make heaps into a monoid by introducing a composition of heaps. (A monoid is a set with a binary operation which is associative.) Intuitively, given two heaps $H_1$ and $H_2$, the composition of $H_1$ and $H_2$, the heap $H_1 \circ H_2$, is the heap which results

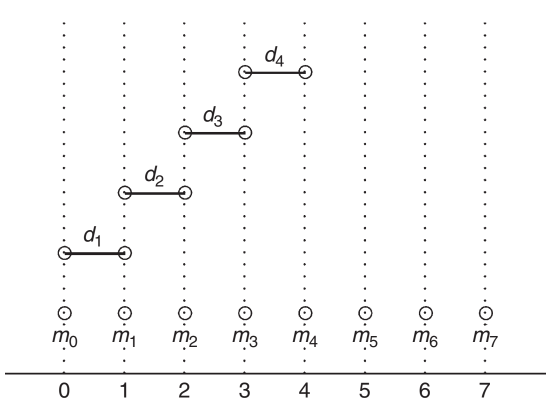



**Figure 7** Monomers and dimers.

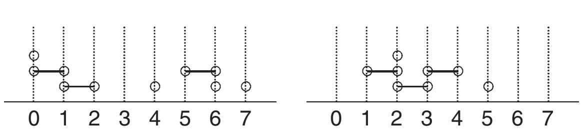



**Figure 8** Two heaps of monomers and dimers.

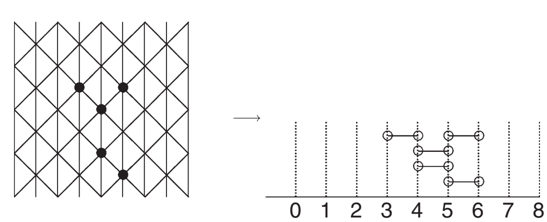



**Figure 10** Bijection between animals and heaps of dimers.

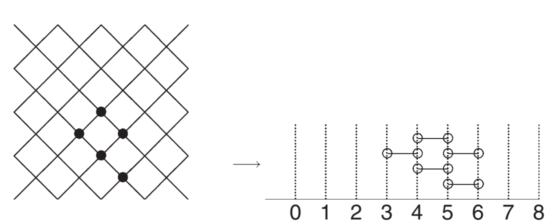



**Figure 11** Bijection between animals and heaps of dimers.

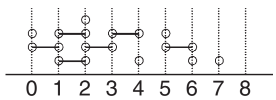


**Figure 12** The composition of the heaps in Figure 8.

by putting $H_2$ on top of $H_1$. In terms of words, the composition of two heaps is the equivalence class of the concatenation $uw$, where $u$ is a word from the equivalence class of $H_1$, and $w$ is a word from the equivalence class of $H_2$.

The composition of the two heaps in Figure 8 is shown in Figure 12.

Given pieces $\mathcal{B}$ with relation $\mathcal{R}$, let $\mathcal{H}(\mathcal{B},\mathcal{R})$ be the set of all heaps consisting of pieces from $\mathcal{B}$, including the empty heap, the latter denoted by $\emptyset$. It is easy to see that the composition makes $(\mathcal{H}(\mathcal{B},\mathcal{R}),\circ)$ into a monoid with unit $\emptyset$.

For the statement of the main theorem in the theory of heaps, we need two more terms. A trivial heap is a heap consisting of pieces all of which are pairwise unrelated. Figure 13a shows a trivial heap consisting of monomers and dimers. A pyramid is a heap with exactly one maximal (= topmost) element. Figure 13a shows a pyramid consisting of monomers and dimers. Finally, if $H$ is a heap, then we write $|H|$ for the number of pieces in $H$.

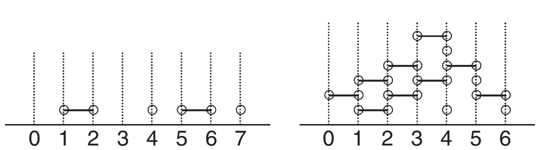
In applications, heaps will have weights, which are defined by introducing a weight $w(B)$ for each piece $B$ in $\mathcal{B}$, and by extending the weight $w$ to all heaps $H$ by letting $w(H)$ denote the product of all weights of the pieces in $H$ (multiplicities of pieces included).

Let $\mathcal{M}$ be a subset of the pieces $\mathcal{B}$. Then, the generating function for all heaps with maximal pieces contained in $\mathcal{M}$ is given by

$$\sum_{\substack{H\in\mathcal{H}(\mathcal{B},\mathcal{R})\\ \text{maximal pieces}\subseteq\mathcal{M}}} w(H)=\frac{\sum_{T\in\mathcal{T}(\mathcal{B}\backslash\mathcal{M},\mathcal{R})}(-1)^{|T|}w(T)}{\sum_{T\in\mathcal{T}(\mathcal{B},\mathcal{R})}(-1)^{|T|}w(T)} \qquad [38]$$

where $\mathcal{T}(\mathcal{B},\mathcal{R})$ denotes the set of all trivial heaps with pieces from $\mathcal{B}$. In particular, the generating function for all heaps is given by

$$\sum_{H\in\mathcal{H}(\mathcal{B},\mathcal{R})} w(H)=\frac{1}{\sum_{T\in\mathcal{T}(\mathcal{B},\mathcal{R})}(-1)^{|T|}w(T)} \qquad [39]$$



(a) (b)

**Figure 13** (a) A trivial heap. (b) A pyramid.

Furthermore, if $\mathcal{P}(\mathcal{B},\mathcal{R})$ denotes the set of all pyramids with pieces from $\mathcal{B}$, then

$$\sum_{P\in\mathcal{P}(\mathcal{B},\mathcal{R})}\frac{w(P)}{|P|}=\log\left(\sum_{H\in\mathcal{H}(\mathcal{B},\mathcal{R})} w(H)\right) \qquad [40]$$

where $|P|$ is the number of pieces of $P$. (As the reader will have guessed, this is a consequence of the "exponential principle" mentioned in the section "generating functions.")

## The Transfer Matrix Method

The transfer matrix method (cf. Stanley (1986), chapter 4 for further reading) applies whenever we are able to build the combinatorial objects that we are interested in by moving on a finite number of states in a step-by-step fashion, where the current step does not depend on the previous ones. (In statistical language, we are considering a finite-state Markov chain.) For example, Motzkin paths which are constrained to stay between two parallel lines, say between $y=0$ and $y=K$, can be described in such a way: the states are the heights $0,1,\ldots,K$, and, if we are in state $h$, then in the next step we are allowed to move to states $h+1, h$, or $h-1$, except that from state 0 we cannot move to $-1$ (there is no state $-1$), and when we are in state $K$ we cannot move to $K+1$ (there is no state $K+1$).

For describing the general situation, let $G=(V,E)$ be a directed graph with vertex set $V$ and edge set $E$. Let $w_n(u,v)$ denote the number of walks from vertex $u$ to vertex $v$ along edges of $G$. To compute these numbers, we consider the adjacency matrix of $G, A(G)$. By definition, using our notation, $A(G)=(w_1(u,v))_{u,v\in V}$. Obviously, $(w_n(u,v))_{u,v\in V}=(A(G))^n$. Thus,

$$\left(\sum_{n=0}^{\infty} w_n(u,v)x^n\right)_{u,v\in V}=\sum_{n=0}^{\infty}(A(G))^n x^n = (I_n-A(G)x)^{-1}$$

where $I_n$ is the $n\times n$ identity matrix. In other words, the generating functions $\sum_{n=0}^{\infty} w_n(u,v)x^n$ for the walk numbers between $u$ and $v$ form the entries of a matrix which is the inverse matrix of $I_n-A(G)x$. By elementary linear algebra,

$$\sum_{n=0}^{\infty} w_n(u,v)\,x^n = \frac{(-1)^{\#u+\#v}\det(I_n-A(G)x)_{v,u}}{\det(I_n-A(G)x)} \qquad [41]$$

where $\det(I_n-A(G)x)_{v,u}$ is the minor of $I_n-A(G)x$ with the row indexed by $v$ and the column indexed

by $u$ omitted, and where $\#u$ denotes the row number of $u$ and similarly for $\#v$. A weighted version could also be developed in the same way, where we put a weight $w(e)$ on each edge, and the weight of a walk is the product of the weights of all its edges.

In particular, the expression [41] is a rational function in $x$. Then, by the basic theorem on rational generating functions (cf. Stanley (1986), section 4.1), the number $w_n(u, v)$ can be expressed as a sum $\sum_{i=1}^{d} P_i(n)\gamma_i^n$, where the $\gamma_i$'s are the different roots of the polynomial $\det(xI_n - A(G))$, and $P_i(n)$ is a polynomial of degree less than the multiplicity of the root $\gamma_i$. (The $P_i(n)$'s depend on $u$ and $v$, whereas the $\gamma_i$'s do not.) If there exists a unique root $\gamma_j$ with maximal modulus, then this implies that, asymptotically as $n \to \infty, w_n(u, v) \sim P_j(n)\gamma_j^n$.

## Lattice Paths

Recall from the section on basic combinatorial terminology that a lattice path P in $\mathbb{Z}^d$ is a path in the $d$-dimensional integer lattice $\mathbb{Z}^d$ which uses only points of the lattice, that is, it is a sequence $(P_0, P_1, \ldots, P_l)$, where $P_i \in \mathbb{Z}^d$ for all $i$. The vectors $\overrightarrow{P_0P_1}, \overrightarrow{P_1P_2}, \ldots, \overrightarrow{P_{l-1}P_l}$ are called the steps of $P$. The number of steps, $l$, is called the length of $P$.

The enumeration of lattice paths has always been an intensively studied topic in statistics, because of their importance in the study of random walks, of rank order statistics for nonparametric testing, and of queueing processes. The reader is referred to Feller (1957) and particularly Mohanty's (1979) book, which is a rich source for enumerative results on lattice paths, albeit in a statistical language. We review the most important results in this section. Most of these concern two-dimensional lattice paths, that is, the case $d = 2$.

To begin with, we consider paths in the integer plane $\mathbb{Z}^2$ consisting of horizontal and vertical unit steps in the positive direction. Clearly, the number of all (unrestricted) paths from the origin to $(n, m)$ is the binomial coefficient $\binom{n+m}{n}$. By the reflection principle, which is commonly attributed to D André (see, e.g., Comtet (1974) p. 22), it follows that the number of paths from the origin to $(n, m)$ which do not pass above the line $y = x + t$, where $m \leq n + t$, is given by

$$\binom{n+m}{n} - \binom{n+m}{n+t+1} \qquad [42]$$

Roughly, the reflection principle sets up a bijection between the paths from the origin to $(n, m)$ which do pass above the line $y = x + t$ and *all* paths from $(-t-1, t+1)$ to $(n, m)$, by reflecting the path portion between the origin and the last touching point on $y = x + t + 1$ in this latter line. Thus, the result of the enumeration problem is the number of all paths from (0, 0) to $(n, m)$, which is given by the binomial coefficient $\binom{n+m}{n}$, minus the number of all paths from $(-t-1, t+1)$ to $(n, m)$, which is given by the binomial coefficient $\binom{n+m}{n+t+1}$, whence the formula [42].

If one considers more generally paths bounded by the line $my = nx + t$, no compact formula is known. It seems that the most conceptual way to approach this problem is through the so-called kernel method (see the section on solving equations for generating functions), which, in combination with the saddle point method, allows one also to obtain strong asymptotic results. There is one special instance, however, which has a "nice" formula. The number of all lattice paths from the origin to $(n, m)$ which never pass above $x = \mu y$, where $\mu$ is a positive integer, is given by

$$\frac{n - \mu m + 1}{n + m + 1}\binom{n+m+1}{m} \qquad [43]$$

The most elegant way to prove this formula is by means of the cycle lemma of Dvoretzky and Motzkin (see Mohanty (1979), p. 9 where the cycle lemma occurs under the name of "penetrating analysis").

Iteration of the reflection principle shows that the number of paths from the origin to $(n, m)$ which stay between the lines $y = x + t$ and $y = x + s$ (being allowed to touch them), where $t \geq 0 \geq s$ and $n + t \geq m \geq n + s$, is given by the finite (!) sum (see, e.g., Mohanty (1979), p. 6)

$$\sum_{k \in \mathbb{Z}} \left( \binom{n+m}{n - k(t-s+2)} - \binom{n+m}{n - k(t-s+2) + t + 1} \right) \qquad [44]$$

The enumeration of lattice paths restricted to regions bounded by hyperplanes has also been considered for other regions, such as quadrants, octants, and rectangles, as well as in higher dimensions. A general result due to Gessel and Zeilberger, and Biane, independently, on the number of lattice paths in a chamber (alcove) of an (affine) reflection group (see Krattenthaler (2003) for the corresponding references and pointers to further results) shows how far one can go when one uses the reflection principle. In particular, this result covers [42] and [44], the enumeration of lattice paths in quadrants, octants, rectangles, and many other results that have

appeared (before and after) in the literature. We present a particularly elegant (and frequently occurring) special case. (In reflection group language, it corresponds to the reflection group of "type $A_{n-1}$." See Humphreys (1990) for terminology and information on reflection groups.)

Let $A=(a_1,a_2,\ldots,a_d)$ and $E=(e_1,e_2,\ldots,e_d)$ be points in $\mathbb{Z}^d$ with $a_1 \geq a_2 \geq \cdots \geq a_d$ and $e_1 \geq e_2 \geq \cdots \geq e_d$. The number of all paths from A to E in the integer lattice $\mathbb{Z}^d$, which consist of positive unit steps and which stay in the region $x_1 \geq x_2 \geq \cdots \geq x_d$, equals

$$\left(\sum_{i=1}^{d}(e_i - a_i)\right)! \det_{1\leq i,j\leq d}\left(\frac{1}{(e_i - a_j - i + j)!}\right) \qquad [45]$$

The counting problem of the theorem is equivalent to numerous other counting problems. It has been originally formulated as an $n$-candidate ballot problem, but it is as well equivalent to counting the number of standard Young tableaux of a given shape. In the case that all $a_j$'s are equal, the determinant does in fact evaluate into a closed-form product. In Young tableaux theory, a particular way to write the result is known as the hook-length formula (see, e.g., Stanley (1999), corollary 7.21.6).

We return to lattice paths in the plane, mentioning some more closely related results. The first is a result of Mohanty (1979, section 4.2), which expresses the number of all lattice paths from the origin to $(n,m)$ which touch the line $y=x+t$ exactly r times, never crossing it, as the difference

$$\binom{n+m-r}{n+t-1} - \binom{n+m-r}{n+t}, \quad r \geq 1 \qquad [46]$$

Not forbidding that the paths cross the bounding line, we arrive at the problem of counting the lattice paths from the origin to $(n,m)$, which cross the main diagonal $y=x$ exactly $r$ times, the answer being

$$\begin{cases} \dfrac{m-n+2r+1}{m+n+1}\dbinom{m+n+1}{n-r} & \text{if } m>n \\[2ex] \dfrac{2r+2}{n}\dbinom{2n}{n-r-1} & \text{if } m=n \end{cases} \qquad [47]$$

Next, we give the number of lattice paths from the origin to $(n,\ n)$ which have $2r$ steps on one side of the line $y=x$, as

$$\binom{2r}{r}\binom{2n-2r}{n-r} \qquad [48]$$

a result due to Sparre Andersen. We refer the reader to Mohanty (1979, chapter 3) for further results in this direction.

Enumerating lattice paths with a fixed number of maximal straight pieces (which correspond to runs), is intimately connected to another basic enumeration problem concerning lattice paths: the enumeration of lattice paths having a fixed number of turns. An effective way to attack the latter problem is by means of two-rowed arrays (see the survey article by Krattenthaler (1997), where in particular analogs of the reflection principle for two-rowed arrays are developed. These imply formulas for the number of lattice paths with fixed starting points and endpoints and a fixed number of north-east (respectively east–north) turns, for unrestricted paths, as well as for paths bounded by lines. (A north–east turn in a lattice path is a point where the direction changes from "north" to "east." An east–north turn is defined analogously.) In particular, analogs of [42]–[44] are known when the number of north–east (respectively east–north) turns is fixed.

These formulas imply for example (see again Krattenthaler (1997, section 3.5)) that the number of lattice paths from the origin to $(n,n)$ which never pass above the line $y=x+t$ and have exactly $2r$ maximal straight pieces is given by

$$2\binom{n-1}{r-1}^2 - \binom{n+t-1}{r-2}\binom{n-t-1}{r} - \binom{n+t-1}{r-1}\binom{n-t-1}{r-1} \qquad [49]$$

with a similar result for the case of $2r+1$ maximal straight pieces. (If $t=0$, the numbers in [49] become

$$\frac{1}{n}\binom{n}{r}\binom{n}{r-1}$$

and they are known as the Narayana numbers.) Furthermore, they imply that the number of lattice paths from the origin to $(n,n)$ which never pass above the line $y=x+t$ and never below the line $y=x-t$ and have exactly $2r$ maximal straight pieces is given by

$$\sum_{k=-\infty}^{\infty}\left\{2\binom{n-2kt-1}{r+k-1}\binom{n+2kt-1}{r-k-1} - \binom{n-2kt+t-1}{r+k-2}\binom{n+2kt-t-1}{r-k} - \binom{n-2kt+t-1}{r+k-1}\binom{n+2kt-t-1}{r-k-1}\right\} \qquad [50]$$

with a similar result for the case of $2r+1$ maximal straight pieces.

The most general boundary for lattice paths that one can imagine is the restriction that it stays

between two given (fixed) paths. Let us assume that the horizontal steps of the upper (fixed) path are at heights $a_1 \leq a_2 \leq \cdots \leq a_n$, whereas the horizontal steps of the lower (fixed) path are at heights $b_1 \leq b_2 \leq \cdots \leq b_n$, $a_i \geq b_i$, $i = 1, 2, \ldots, n$. Then the number of all paths from $(0, b_1)$ to $(n, a_n)$ satisfying the property that for all $i = 1, 2, \ldots, n$ the height of the $i$th horizontal step is between $b_i$ and $a_i$ is given by the determinant

$$\det_{1 \leq i,j \leq n}\left(\binom{a_i - b_j + 1}{j - i + 1}\right) \qquad [51]$$

In the statistical literature, this formula is often known as "Steck's formula," but it is actually a special case of a much more general theorem due to Kreweras. A generalization of [51] to higher-dimensional paths was given by Handa and Mohanty (see Mohanty (1979, section 2.4)).

Next, we consider three-step lattice paths in the integer plane $\mathbf{Z}^2$, that is, paths consisting of up-steps $(1, 1)$, level steps $(1, 0)$, and down-steps $(1, -1)$. The particular problem that we are interested in is to count such three-step paths starting at $(0, r)$ and ending at $(\ell, s)$, which do not pass below the $x$-axis and do not pass above the horizontal line $y = K$. Furthermore, we assign the weight 1 to an up-step, the weight $b_h$ to a level-step at height $h$, and the weight $\lambda_h$ to a down-step from height $h$ to $h - 1$. The weight $w(P)$ of a path $P$ is defined as the product of the weights of all its steps. Then we have the following result, which can be obtained by the transfer matrix method described in the last section.

Define the sequence $(p_n(x))_{n \geq 0}$ of polynomials by

$$xp_n(x) = p_{n+1}(x) + b_n p_n(x) + \lambda_n p_{n-1}(x) \quad \textit{for } n \geq 1 \qquad [52]$$

with initial conditions $p_0(x) = 1$ and $p_1(x) = x - b_0$. Furthermore, define $(\mathrm{S}p_n(x))_{n \geq 0}$ to be the sequence of polynomials which arises from the sequence $(p_n(x))$ by replacing $\lambda_i$ by $\lambda_{i+1}$ and $b_i$ by $b_{i+1}, i = 0, 1, 2, \ldots,$ everywhere in the three-term recurrence [52] and in the initial conditions. Finally, given a polynomial $p(x)$ of degree $n$, we denote the corresponding reciprocal polynomial $x^n p(1/x)$ by $p^*(x)$.

With the weight $w$ defined as before, the generating function $\sum_P w(P)x^{\ell(P)}$, where the sum is over all three-step paths which start at $(0, r)$, terminate at height $s$, do not pass below the $x$-axis, and do not pass above the line $y = K$, is given by

$$\begin{cases} \dfrac{x^{s-r} p_r^*(x) S^{s+1} p_{K-s}^*(x)}{p_{K+1}^*(x)}, & r \leq s \\ \lambda_r \cdots \lambda_{s+1} \dfrac{x^{r-s} p_s^*(x) S^{r+1} p_{K-r}^*(x)}{p_{K+1}^*(x)}, & r \geq s \end{cases} \qquad [53]$$

The sequence of polynomials $(p_n(x))_{n \geq 0}$ is in fact a sequence of orthogonal polynomials (cf. Koekoek and Swarttouw (1998) and Szegő (1959)).

We remark that in the case that $r = s = 0$ there is also an elegant expression for the generating function due to Flajolet (see section V.2 of the Flajolet and Sedgewick reference in "Further reading") in terms of a continued fraction.

In order to solve our problem, we just have to extract the coefficient of $x^\ell$ in [53]. By a partial fraction expansion, a formula of the type

$$\sum_m c_m \xi_m^\ell \qquad [54]$$

results, where the $\xi_m$'s are the zeroes of $p_{K+1}(x)$, and the $c_m$'s are some coefficients, only a finite number of them being nonzero.

It should be noted that, because of the many available parameters (the $b_n$'s and $\lambda_n$'s), by appropriate specializations one can also obtain numerous results about enumerating three-step paths according to various statistics, such as the number of touchings on the bounding lines, etc.

There are two important special cases in which a completely explicit solution in terms of elementary functions can be given.

The first case occurs for $b_i = 0$ and $\lambda_i = 1$ for all $i$. In this case, the polynomials $p_n(x)$ defined by the three-term recurrence [52] are Chebyshev polynomials of the second kind, $p_n(x) = U_n(x/2)$. (The Chebyshev polynomial of the second kind $U_n(x)$ is defined by $U_n(\cos t) = \sin((n+1)t)/\sin t$ (see Koekoek and Swarttouw (1998) for almost exhaustive information on these polynomials and, more generally, on hypergeometric orthogonal polynomials)). The result which is then obtained from the general theorem (clearly, the zeros of $U_n(x)$ are $x = \cos(2k\pi/(n+1)), k = 1, 2, \ldots, n,$ and therefore the partial fraction expansion of [53] is easily determined) is that the number of lattice paths from $(0, r)$ to $(\ell, s)$ with only up- and down-steps, which always stay between the $x$-axis and the line $y = K$, is given by (see also Feller (1957, chapter XIV, eqn [5.7])

$$\frac{2}{K+2}\sum_{k=1}^{K+1}\left(2\cos\frac{\pi k}{K+2}\right)^\ell \times \sin\frac{\pi k(r+1)}{K+2}\sin\frac{\pi k(s+1)}{K+2} \qquad [55]$$

a formula which goes back to Lagrange.

The second case occurs for $b_i = 1$ and $\lambda_i = 1$ for all $i$. In this case, the polynomials $p_n(x)$ defined by the three-term recurrence [52] are again Chebyshev polynomials of the second kind, $p_n(x) = U_n((x-1)/2)$. The result which is then

obtained from the general theorem is that the number of three-step lattice paths from $(0, r)$ to $(\ell, s)$, which always stay between the $x$-axis and the line $y = K$, is given by

$$\frac{2}{K+2}\sum_{k=1}^{K+1}\left(2\cos\frac{\pi k}{K+2}+1\right)^{\ell} \times \sin\frac{\pi k(r+1)}{K+2}\sin\frac{\pi k(s+1)}{K+2} \qquad [56]$$

## Perfect Matchings and Tilings

In this section we consider the problem of counting the perfect matchings of a graph. For an introduction into the problem, and into methods to solve it, as well as for a report on recent developments, we refer the reader to Propp (1999).

Let $G = (V, E)$ be a finite loopless graph with vertex set $V$ and edge set $E$. A matching (also called 1-factor in graph theory) is a subset of the edges with the property that no two edges share a vertex. A matching is perfect if it covers all the edges. Let $M(G)$ denote the number of perfect matchings of the graph $G$. More generally, we could assign a weight $w(e)$ to each edge $e$ of the graph and define the weight of a matching to be the product of the weights of all its edges. Let $M_w(G)$ denote the sum of all weights of all matchings of the graph $G$.

Kasteleyn's method for determining $M(G)$, respectively $M_w(G)$, makes use of determinants and Pfaffians. Recall that the Pfaffian $\mathrm{Pf}(A)$ of a triangular array $A = (a_{i,j})_{1\le i<j\le 2n}$ is defined by

$$\mathrm{Pf}(A) = \sum_m (\mathrm{sgn}\, m) \prod_{\{i,j\}\in m} \alpha_{i,j} \qquad [57]$$

where the sum is over all perfect matchings of the complete graph on vertices $\{1, 2, \ldots, 2n\}$, and where the product is over all edges $\{i, j\}, i < j$, of $m$. The sign $\mathrm{sgn}\, m$ of $m$ is $(-1)^{\#\text{crossings of } m}$, where a crossing is a pair $(\{i, j\}, \{k, l\})$ of edges such that $i < k < j < l$. Usually, one extends the triangular array $A$ to a matrix by setting $a_{j,i} = -a_{i,j}, i < j$, and $a_{i,i} = 0$ for all $i$. Then, abusing notation, we identify the triangular array with the skew-symmetric matrix $A = (a_{i,j})_{1\le i,j\le 2n}$. The Pfaffian satisfies the following useful properties:

$$\mathrm{Pf}(B^t AB) = \det(B)\,\mathrm{Pf}(A)$$

and

$$\mathrm{Pf}(A)^2 = \det(A) \qquad [58]$$

The latter equality shows in particular that Pfaffians are very close to determinants. They do, in fact, generalize determinants since

$$\mathrm{Pf}\begin{pmatrix} 0 & B \\ -B & 0 \end{pmatrix} = \det B \qquad [59]$$

for any square matrix $B$.

Thus, given a graph with vertices $v_1, v_2, \ldots, v_{2n}$, specializing $a_{i,j}$ to the weight of the edge between $v_i$ and $v_j$, if it exists, and setting $a_{i,j} = 0$ otherwise in the definition of the Pfaffian, we obtain almost $M_w(G)$, the only difference is that there could be signs in front of the individual terms of the sum, whereas in $M_w(G)$ the sign in front of each term must be $+$. (The object obtained by omitting the sign in [57] is called Hafnian. Unfortunately, in contrast to the Pfaffian, it does not have any nice properties and it is therefore extremely difficult to compute.) Kasteleyn's idea is to circumvent this problem by orienting the edges of the graph, defining signed weights of the edges, in such a way that the Pfaffian of the array with signed weights produces exactly $M_w(G)$.

More precisely, given a (weighted) graph $G$ with vertices $v_1, v_2, \ldots, v_{2n}$, we make it into an oriented (weighted) graph $\overrightarrow{G}$. That is, if there is an edge between $v_i$ and $v_j$, $e_{i,j}$ say, we orient it either from $v_i$ to $v_j$ or the other way. Now we define the signed adjacency matrix $A(\overrightarrow{G})$ of $\overrightarrow{G}$ by letting its $(i, j)$-entry to be $+w(e_{i,j})$ if there is an edge from $v_i$ to $v_j$ oriented that way, $-w(e_{i,j})$ if there is an edge from $v_j$ to $v_i$ oriented that way, and 0 if there is no edge between $v_i$ and $v_j$. Such an orientation is called Pfaffian if

$$\mathrm{Pf}(A(\overrightarrow{G})) = \pm M_w(G)$$

Clearly, the question remains whether a Pfaffian orientation can be found for a given graph. In general, this is an open question. However, Kasteleyn shows that for planar graphs such a Pfaffian orientation can always be found. Moreover, he shows that any orientation of a planar graph which has the property that around any face bounded by $4k$ edges an odd number of edges is oriented in either direction and that around any face bounded by $4k + 2$ edges an even number of edges is oriented in either direction is Pfaffian.

For bipartite graphs (i.e., for graphs in which the set of vertices can be split into two disjoint sets such that all the edges connect the vertex of one of these sets to a vertex of the other), the situation is even nicer. This is because for a bipartite graph $G$ in which both parts of the bipartition of the vertices are of the same size (otherwise, there is no perfect matching), any signed

adjacency matrix $A(\overrightarrow{G})$ has the block form of the matrix on the left-hand side of [59] and, hence, the Pfaffian reduces to a determinant. More precisely, let $G$ be a bipartite graph with vertex set $V = U \cup W$, $U = \{u_1, u_2, \ldots, u_n\}$ and $W = \{w_1, w_2, \ldots, w_n\}$, with edges connecting some $u_i$ to some $w_j$. Given a Pfaffian orientation $\overrightarrow{G}$, we build the signed bipartite adjacency matrix $B(\overrightarrow{G}) = (b_{i,j})_{1 \le i,j \le n}$ of $\overrightarrow{G}$ by setting $b_{i,j} = +w(e_{i,j})$ if there is an edge from $u_i$ to $w_j$ oriented that way, $-w(e_{i,j})$ if there is an edge from $u_j$ to $w_i$ oriented that way, and 0 if there is no edge between $u_i$ and $w_j$. Then we have

$$\det(B(\overrightarrow{G})) = \pm M_w(G)$$

In particular, this holds for any bipartite planar graph. See Robertson *et al.* (1999) for a structural description about which (not necessarily planar) bipartite graphs admit a Pfaffian orientation.

Kasteleyn's construction in the planar case has been generalized to graphs on surfaces of any genus $g$ in Dolbilin *et al.* (1996), Galluccio and Loebl (1999), and Tesler (2000), independently. As predicted by Kasteleyn, the solution is in terms of a linear combination of $4^g$ Pfaffians.

With the help of his method, Kasteleyn computed the number of dimer coverings of an $m \times n$ rectangle. (A dimer is a $2 \times 1$ rectangle. Thus, this is equivalent to counting the number of perfect matchings on the $m \times n$ grid graph. The formula was independently found by Temperley and Fisher.) The result is

$$\prod_{i=1}^{m} \prod_{j=1}^{n} \left( 2\cos\frac{\pi i}{m+1} + 2\sqrt{-1}\cos\frac{\pi j}{n+1} \right)$$

For even $m$ and $n$, the formula can be rewritten as

$$\prod_{i=1}^{m/2} \prod_{j=1}^{n/2} \left( 4\cos^2\frac{\pi i}{m+1} + 4\cos^2\frac{\pi j}{n+1} \right)$$

There is a similar rewriting if one of $m$ or $n$ is odd. (If both $m$ and $n$ are odd, there is no dimer covering.)

For further reading and references *see* Dimer Problems and Kuperberg (1998).

## Nonintersecting Paths

Let $G = (V, E)$ be a directed acyclic graph with vertices $V$ and directed edges $E$. Furthermore, we are given a function $w$ which assigns a weight $w(x)$ to every vertex or edge $x$. Let us define the weight $w(P)$ of a walk $P$ in the graph by $\prod_e w(e) \prod_v w(v)$, where the first product is over all edges $e$ of the walk $P$ and the second product is over all vertices $v$ of $P$. We denote the set of all walks in $G$ from $u$ to $v$ by $\mathcal{P}(u \to v)$, and the set of all families $(P_1, P_2, \ldots, P_n)$ of walks, where $P_i$ runs from $u_i$ to $v_i, i = 1, 2, \ldots, n$, by $\mathcal{P}(\boldsymbol{u} \to \boldsymbol{v})$, with $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$ and $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$. The symbol $\mathcal{P}^+(\boldsymbol{u} \to \boldsymbol{v})$ stands for the set of all families $(P_1, P_2, \ldots, P_n)$ in $\mathcal{P}(\boldsymbol{u} \to \boldsymbol{v})$ with the additional property that no two walks share a vertex. We call such families of walk(er)s "vicious walkers" or, alternatively, "nonintersecting paths." The weight $w(\boldsymbol{P})$ of a family $\boldsymbol{P} = (P_1, P_2, \ldots, P_n)$ of walks is defined as the product $\prod_{i=1}^{n} w(P_i)$ of all the weights of the walks in the family. Finally, given a set $M$ with weight function $w$, we write $\mathrm{GF}(\mathcal{M}; w)$ for the generating function $\sum_{x \in \mathcal{M}} w(x)$.

We need two further notations before we are able to state the Lindström–Gessel–Viennot theorem. (For references and historical remarks, we refer the reader to footnote 5 in Krattenthaler (2005a).) As earlier, the symbol $\mathfrak{S}_n$ denotes the symmetric group of order $n$. Given a permutation $\sigma \in \mathfrak{S}_n$, we write $\boldsymbol{u}_\sigma$ for $(u_{\sigma(1)}, u_{\sigma(2)}, \ldots, u_{\sigma(n)})$. Then

$$\sum_{\sigma \in \mathfrak{S}_n} (\mathrm{sgn}\ \sigma) \cdot \mathrm{GF}(\mathcal{P}^+(\boldsymbol{u}_\sigma \to \boldsymbol{v}); w) = \det_{1 \le i,j \le n} \left( \mathrm{GF}(\mathcal{P}(u_j \to v_i); w) \right) \qquad [60]$$

Most often, this theorem is applied in the case where the only permutation $\sigma$ for which vicious walks exist is the identity permutation, so that the sum on the left-hand side reduces to a single term that counts all families $(P_1, P_2, \ldots, P_n)$ of vicious walks, the $i$th walk $P_i$ running from $A_i$ to $E_i, i = 1, 2, \ldots, n$. This case occurs, for example, if for any pair of walks $(P, Q)$ with $P$ running from $u_a$ to $v_d$ and $Q$ running from $u_b$ to $v_c$, $a < b$ and $c < d$, it is true that $P$ and $Q$ must have a common vertex. Explicitly, in that case we have

$$\mathrm{GF}(\mathcal{P}^+(\boldsymbol{u} \to \boldsymbol{v}); w) = \det_{1 \le i,j \le n} \left( \mathrm{GF}(\mathcal{P}(u_j \to v_i); w) \right) \quad [61]$$

If the starting points or/and the endpoints are not fixed, then the corresponding number is given by a Pfaffian, a result obtained by Okada and Stembridge (see Bressoud (1999) for references). For a set $\mathcal{A}$ of starting points, let $\mathcal{P}^+(\mathcal{A} \to \boldsymbol{v})$ denote the set of all families $(P_1, P_2, \ldots, P_{2n})$ of nonintersecting lattice paths, where $P_i$ runs from some point of $\mathcal{A}$ to $v_i, i = 1, 2, \ldots, 2n$. Furthermore, let us suppose that the elements of $\mathcal{A} = \{u_1, u_2, \ldots\}$ are ordered in such a way that for any pair of walks $(P, Q)$ with $P$ running from $u_a$ to $v_d$ and $Q$ running from $u_b$ to $v_c$, $a < b$ and $c < d$, it is true that $P$ and $Q$ must have a common vertex. (This is the same condition as the one which makes [61] valid, with the only difference that, here,

the number of $u_i$'s could be larger than the number of $v_i$'s.) Then,

$$\mathrm{GF}(\mathcal{P}^+(\mathcal{A}\to \boldsymbol{v});w) = \underset{1\le i,j\le 2n}{\mathrm{Pf}}\left(\sum_{a<b}\left(\mathrm{GF}(\mathcal{P}(u_a\to v_i);w)\mathrm{GF}(\mathcal{P}(u_b\to v_j);w) - \mathrm{GF}(\mathcal{P}(u_b\to v_i);w)\mathrm{GF}(\mathcal{P}(u_a\to v_j);w)\right)\right) \quad [62]$$

If the number of paths is odd, then one can use the same formula by adding an artificial point to the endpoints and to the set of starting points $\mathcal{A}$. There is also a theorem by Okada and Stembridge which covers the case that starting points and endpoints vary. Refinements when the number of turns is fixed can be found in Krattenthaler (1997).

## Vicious Walkers, Plane Partitions, Rhombus Tilings, and Fully Packed Loop Configurations

In this section we describe the interrelations between four frequently appearing objects in statistical mechanics and combinatorics: vicious walkers, plane partitions, rhombus tilings, and fully packed loop configurations.

Given a lattice, vicious walkers, as introduced by Fisher (1984), are particles which move on lattice sites in such a way that two particles never occupy the same lattice site. Models of vicious walkers have been the object of numerous studies from various points of view. Rather than accomplishing the impossible task of providing a complete overview of references, the reader is referred to the basic reference Fisher (1984) and to Krattenthaler (2005a) for further pointers to the literature.

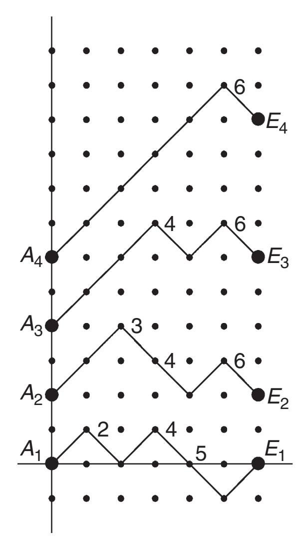
Most of the known results apply for vicious walkers on the line. There are in fact two different models: in the random turns vicious walker model, $n$ walkers move on the integral points of the real line in such a way that at each tick of the clock exactly one walker moves to the right or to the left, whereas in the lock step vicious walker model $n$ walkers move on the integral points of the real line in such a way that at each tick of the clock each walker moves to the right or to the left.

The first model is equivalent to a model of one walker in $\mathbb{Z}^n$ ($\mathbb{Z}$ denoting the set of integers) which at each tick of the clock moves a positive or negative unit step in the direction of one of the coordinate axes, always staying in the wedge $x_1 > x_2 > \cdots > x_n$. This point of view was already put forward by Fisher (1984). However, this problem belongs to the problem of counting paths in chambers of reflection groups discussed in the section "Lattice paths."

The second model could also be realized as a single walker model (cf. Krattenthaler (2003)). However, most often it is realized as a model of $n$ paths in the plane consisting of steps $(1,1)$ and $(1,-1)$ with the property that no two paths have a point in common. In this picture, the $x$-axis becomes the time line, the $k$th path doing an up-step $(1,1)$ from $(t-1,y)$ to $(t,y+1)$ meaning that the $k$th particle moves to the left at time $t$, whereas the $k$th path doing a down-step $(1,-1)$ from $(t-1,y)$ to $(t,y-1)$ meaning that the $k$th particle moves to the right at time $t$.

The reader should consult Figure 14a for an example. (The labelings should be ignored at this point.) Clearly, what we encounter here is a particular instance of the nonintersecting paths of the last section. Therefore, for fixed starting points and endpoints, formula [61] applies, whereas if the starting points vary and the endpoints are fixed, it is formula [62] that applies.

At this point, the links to the other objects, semistandard tableaux and plane partitions (cf. Bressoud (1999)), emerge. A filling of the cells of the Ferrers diagram of $\lambda$ with elements of the set $\{1,2,\ldots\}$, which is weakly increasing along rows and strictly increasing along columns is called a (semistandard) tableau of shape $\lambda$. Figure 14b shows such a semistandard tableau of shape $(4,3,2)$. In fact, vicious walkers and semistandard tableaux are equivalent objects. To see this, first label down-steps by the $x$-coordinate of their endpoint, so that a step from $(a-1,b)$ to $(a,b-1)$ is labeled by $a$, see Figure 14a. Then, out of the labels of the $j$th path, form the $j$th column of the corresponding tableau,




(a)

| 2 | 3 | 4 | 6 |
|---|---|---|---|
| 4 | 4 | 6 | |
| 5 | 6 | | |

(b)

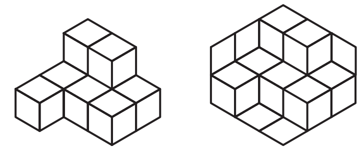
**Figure 14** (a) Vicious walkers. (b) A tableau.

see Figure 14b. The resulting array of numbers is indeed a semistandard tableau. This can be readily seen, since the entries are trivially strictly increasing along columns, and they are weakly increasing along rows because the paths do not touch each other. Thus, problems of enumerating vicious walkers can be translated into tableau enumeration problems, and vice versa.

The significance of semistandard tableaux lies particularly in the representation theory for classical groups, *see* Classical Groups and Homogenous Spaces and Compact Groups and Their Representations. Namely, the irreducible characters for $GL(n,\mathbb{C})$ and $SL(n,\mathbb{C})$, the Schur functions, are generating functions for semistandard tableaux of a given shape. If the entries of the $i$th row of a semistandard tableau are required to be at least $2i-1$, then one speaks of symplectic tableaux, and the irreducible characters for $Sp(2n,\mathbb{C})$ are generating functions for symplectic tableaux of a given shape. We refer the reader to Krattenthaler *et al.* (2000) for more information on these topics.

Objects which are very close to semistandard tableaux are plane partitions. According to MacMahon, a plane partition of shape $\lambda$ is a filling of the Ferrers diagram of $\lambda$ with non-negative integers which is weakly decreasing along rows and columns. See Figure 15b for an example of a plane partition of shape $(3,3,3)$. In particular, semistandard tableaux and plane partitions of rectangular shape are actually equivalent. For, let $T$ be a semistandard tableau of rectangular shape. Then, from each element of the $i$th row we subtract $i$. Finally, the obtained array is rotated by 180°. As a result, we obtain a plane partition. See Figure 15 for a semistandard tableau and a plane partition which correspond to each other under these transformations.

On the other hand, plane partitions can also be realized as three-dimensional objects, by interpreting each entry in the array as a pile of unit cubes of the size of the entry. For example, the plane partition in Figure 15 corresponds to the pile of cubes in Figure 16a. But then, forgetting the three-dimensional view, by embedding the picture in a minimally bounding hexagon, and by filling the emerging empty regions by rhombi of unit length in the unique way this is possible, we obtain a rhombus tiling of a hexagon in which opposite sides have the same length, see Figure 16b.



(a) (b)

**Figure 16** (a) A plane partition; three-dimensional view. (b) A rhombus tiling.

From the rhombus tiling, there is then again an elegant way to go to nonintersecting paths: we mark the mid-points of the edges along two opposite sides, see Figure 17a. Now we draw lattice paths which connect points on different sides, by "following" along the other lozenges, as indicated in Figure 17a by the dashed lines. Clearly, the resulting paths are nonintersecting, that is, no two paths have a common vertex. If we slightly distort the underlying lattice, we get orthogonal paths with horizontal and vertical steps in the positive direction, see Figure 17b.

Rhombus tilings, on their part, are equivalent to perfect matchings of hexagonal graphs. To see this, one places the tiling on the underlying triangular grid, see Figure 18a. Then one places a bond into each rhombus, so that it connects the mid-points of the two triangles out of which the rhombus is composed, see Figure 18b. Finally, one forgets the contour of the tiling, but instead one introduces all the other edges which connect mid-points of adjacent triangles of the triangular grid, see Figure 18c. Thus, one arrives at a perfect matching of the hexagonal graph consisting of the edges connecting mid-points of triangles.

Because of these various connections, enumeration problems for vicious walkers, plane partitions, tableaux, rhombus tilings can be approached by the different methods which are available for the various objects: the determinant theorem from the section "Nonintersecting paths," together with determinant evaluation techniques (cf. the survey Krattenthaler (2005b)), apply, as well as the "Kasteleyn method" from the section "Perfect

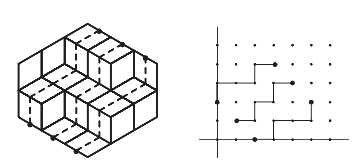
| 1 | 1 | 2 |
|---|---|---|
| 3 | 3 | 3 |
| 4 | 5 | 5 |

(a)

| 2 | 2 | 1 |
|---|---|---|
| 1 | 1 | 1 |
| 1 | 0 | 0 |

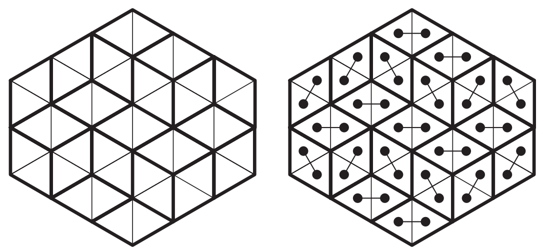
(b)

**Figure 15** (a) A semistandard tableau. (b) A plane partition.

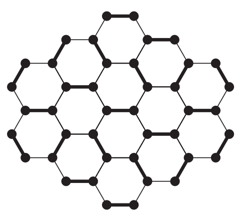


(a) (b)

**Figure 17** (a) A rhombus tiling. (b) A family of nonintersecting paths.



(a) (b)



(c)

**Figure 18** (a) A rhombus tiling. (b) Bonds in rhombi. (c) A perfect matching of a hexagonal graph.

matchings and tilings," and also methods from character theory for the classical groups. All of these methods have been applied extensively (see the surveys by Kenyon (2003), Propp (1999), and Krattenthaler *et al.* (2000)), the first and third more frequently for exact enumeration, while the second particularly for asymptotic studies. It should be noted that methods from random matrix theory also apply in certain situations, see Johansson (2002). *See* Growth Processes in Random Matrix Theory and Random Matrix Theory in Physics.

In fact, we missed mentioning a further object, from statistical physics, which in some cases is equivalent to vicious walkers, etc.: fully packed loop configurations. (Fully packed loop configurations are in bijection with six-vertex configurations, see the next section.) If one imposes certain "connectivity constraints" on fully packed loop configurations, then one can construct bijections with rhombus tilings and, hence, with nonintersecting paths and with the other objects discussed in this section. The reader is referred to Di Francesco *et al.* (2004) and references therein.

Having explained the various connections, we cite some fundamental results in the area. (We refer the reader to Bressoud (1999) and Stanley (1999, chapter 7).) MacMahon proved that the number of all plane partitions contained in an $a \times b \times c$ box (when viewed in three dimensions) is equal to

$$\prod_{i=1}^{a}\prod_{j=1}^{b}\prod_{k=1}^{c}\frac{i+j+k-1}{i+j+k-2} \qquad [63]$$

Thus, the number of rhombus tilings of a hexagon with side lengths $a,b,c,a,b,c$ is given by the same number, as well as the number of all vicious walkers $(P_1, P_2, \ldots, P_a)$, where $P_i$ runs from $(0, 2i)$ to $(b+c, b-c+2i), i=1,2,\ldots,a$. More generally, the number of semistandard tableaux of shape $\lambda$ with entries at most $m$ is given by the hook-content formula

$$\prod_{u\in\lambda}\frac{c(u)+m}{h(u)} \qquad [64]$$

where $u$ ranges over all the cells of the Ferrers diagram of $\lambda$, with $c(u)$ being the content of $u$, defined as the difference of the column number and the row number of $u$, and with $h(u)$ being the hook length of $u$, defined as the number of cells in the hook of $u$, the latter consisting of the cells to the right of $u$ in the same row and below $u$ in the same column, including $u$. Thus, this also gives a formula for the number of all vicious walkers $(P_1, P_2, \ldots, P_a)$, where $P_i$ runs from $(0, 2i)$ to $(N, h_i)$. See Krattenthaler *et al.* (2000, section 2) for details. There it is also explained that a Schur function summation formula, together with an analog of the hook-content formula for special orthogonal characters, proves that the number of all vicious walkers $(P_1, P_2, \ldots, P_a)$, where $P_i$ runs from $(0, 2i)$ for $N$ steps is given by

$$\prod_{1\le i\le j\le N}\frac{a+i+j-1}{i+j-1} \qquad [65]$$

The reader is referred to the references given in this section for many more results, in particular, on the enumeration of plane partitions with symmetry, the enumeration of rhombus tilings of regions other than hexagons, and the enumeration of vicious walkers with various starting points and endpoints, under various constraints.

## Six-Vertex Model and Alternating-Sign Matrices

An alternating-sign matrix is a square matrix of 0's, 1's and –1's for which the sum of entries in each row and in each column is 1 and the nonzero entries of each row and of each column alternate in sign. For instance,

$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

is a $5 \times 5$ alternating-sign matrix. Zeilberger proved that the number of $n \times n$ alternating-sign matrices is given by

$$\prod_{i=0}^{n-1} \frac{(3i+1)!}{(n+i)!} \qquad [66]$$

and he went on to prove the finer version that the number of $n \times n$ alternating-sign matrices with the (unique) 1 in the first row in position $j$ is given by

$$\frac{\binom{n+j-2}{n-1}\binom{2n-j-1}{n-1}}{\binom{3n-2}{n-1}} \prod_{i=0}^{n-1} \frac{(3i+1)!}{(n+i)!} \qquad [67]$$

The first number is also equal to the number of totally symmetric self-complementary plane partitions contained in the $(2n) \times (2n) \times (2n)$ box, but there is no intrinsic explanation why this is so. We refer the reader to Bressoud (1999) for an exposition of these results, and for pointers to the literature containing further unexplained connections between alternating-sign matrices and plane partitions.

While the first result was achieved by a brute-force constant-term approach, the second result is based on the observation that alternating-sign matrices are in bijection with configurations in the six-vertex model on the square grid under domain-wall boundary conditions. This then allowed one to use a formula due to Izergin for the partition function for these six-vertex configurations. Similar formulas for variations of the model have been found by Kuperberg, and by Razumov and Stroganov (see Razumov and Stroganov (2005) and references therein).

A configuration in the six-vertex model is an orientation of edges of a 4-regular graph (i.e., at each vertex there meet exactly four edges) such that at each vertex two edges are oriented towards the vertex and two are oriented away from the vertex. Thus, there are six possible vertex configurations, giving the name of the model, see **Figure 19**. To go from one object to the other, one uses the translation between local configurations at a vertex and entries in alternating-sign matrices indicated in the figure. An example of the correspondence can be found in **Figure 20**.

Another manifestation of alternating-sign matrices and six-vertex configurations are fully packed loop configurations. A fully packed loop configuration on a graph is a collection of edges such that each vertex is

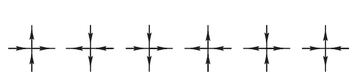
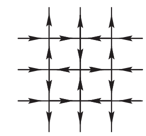



**Figure 19** The six vertex configurations.

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & -1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$



(a) (b)

**Figure 20** (a) An alternating-sign matrix. (b) A six-vertex configuration.

incident to exactly two edges. One obtains a fully packed loop configuration out of a six-vertex configuration by dividing the square lattice into its even and odd sublattice denoted by $A$ and $B$, respectively. Instead of arrows, only those edges are drawn that, on sublattice $A$, point inward and, on sublattice $B$, point outward. The reader is referred to de Gier (2005) and Di Francesco *et al.* (2004) for further reading.

The story of alternating-sign matrices and their connection to the six-vertex model is given a vivid account in Bressoud (1999), with further important results by Kuperberg, Okada, Razumov and Stroganov, referenced in Razumov and Stroganov (2005).

Fully packed loop configurations seem to play an important role in the explicit form of the ground-state vectors of certain Hamiltonians in the dense $O(1)$ loop model. The corresponding conjectures are surveyed in de Gier (2005). There is important progress on these conjectures by Di Francesco and Zinn–Justin (2005, and references therein).

## Binomial Sums and Hypergeometric Series

When dealing with enumerative problems, it is inevitable to deal with binomial sums, that is, sums in which the summands are products/quotients of binomial coefficients and factorials, such as, for example,

$$\sum_{k=0}^{n} \binom{2k}{k}\binom{2n-2k}{n-k}$$

In most cases, the right environment in which one should work is the theory of (generalized) hypergeometric series. These are defined as follows:

$${}_rF_s\left[\begin{matrix} a_1,\ldots,a_r \\ b_1,\ldots,b_s \end{matrix}; z\right] = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_r)_k}{(b_1)_k \cdots (b_s)_k} \frac{z^k}{k!}$$

where $(\alpha)_k = \alpha(\alpha+1)(\alpha+2)\cdots(\alpha+k-1)$ for $k > 0$, and $(\alpha)_0 = 1$. The symbol $(\alpha)_k$ is called the Pochhammer symbol or shifted factorial. For in-depth treatments of the subject, we refer the reader

to Andrews *et al.* (1999), Gasper and Rahman (2004), and Slater (1966).

Hypergeometric series can be characterized as series in which the quotient of the $(k+1)$st by the $k$th summand is a rational function in $k$. This is also the way to convert binomial sums into their hypergeometric form (respectively to see if this is possible; in most cases it is): form the quotient of the $(k+1)$st by the $k$th summand and read off the parameters $a_1, \ldots, a_r, b_1, \ldots, b_s$, and the argument $z$ from the factorization of the numerator and the denominator polynomials of the rational function, out of these form the corresponding hypergeometric series, and multiply the series by the summand for $k=0$. This is, in fact, a completely routine task, and, indeed, computer algebra programs such as Maple and Mathematica do this automatically.

The reason why hypergeometric series are much more fundamental than the binomial sums themselves is that there are hundreds of ways to write the same sum using binomial coefficients and factorials, whereas there is just one hypergeometric form, that is, hypergeometric series are a kind of normal form for binomial sums. In particular, given a specific binomial sum, it is a hopeless enterprise to scan through all the identities available in the literature for this sum. There may be an identity for it, but perhaps written differently. On the contrary, given a specific hypergeometric series, the list of available identities which apply to this series is usually not large, and tables of such identities can be set up in a systematic way. This has been done (cf. Slater (1966); the most comprehensive table available to this date is contained in the manual of the Mathematica package HYP – see "Further reading"), and scanning through these tables is largely facilitated by the use of the Mathematica package HYP.

We give here some of the most important identities for hypergeometric series. Aside from the binomial theorem, the most important summation formulas are: the Gauß ${}_2F_1$-summation formula

$$ {}_2F_1\left[\begin{matrix} a, b \\ c \end{matrix}; 1\right] = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} $$

provided $\Re(c-a-b) > 0$,
the Pfaff–Saalschütz summation formula

$$ {}_3F_2\left[\begin{matrix} a, b, -n \\ c, 1+a+b-c-n \end{matrix}; 1\right] = \frac{(c-a)_n (c-b)_n}{(c)_n (c-a-b)_n} $$

provided $n$ is a non-negative integer, and
the Dougall summation formula

$$ {}_7F_6\left[\begin{matrix} a, a/2+1, b, c, d, 1+2a-b-c-d+n, -n \\ a/2, 1+a-b, 1+a-c, 1+a-d, \\ -a+b+c+d-n, a+1+n \end{matrix}; 1\right] $$
$$ = \frac{(1+a)_n (1+a-b-c)_n (1+a-b-d)_n (1+a-c-d)_n}{(1+a-b)_n (1+a-c)_n (1+a-d)_n (1+a-b-c-d)_n} $$

provided $n$ is a non-negative integer.

Some of the most important transformation formulas are
the Euler transformation formula

$$ {}_2F_1\left[\begin{matrix} a, b \\ c \end{matrix}; z\right] = (1-z)^{c-a-b} {}_2F_1\left[\begin{matrix} c-a, c-b \\ c \end{matrix}; z\right] $$

provided $|z| < 1$,
the Kummer transformation formula

$$ {}_3F_2\left[\begin{matrix} a, b, c \\ d, e \end{matrix}; 1\right] = \frac{\Gamma(e)\Gamma(d+e-a-b-c)}{\Gamma(e-a)\Gamma(d+e-b-c)} \times {}_3F_2\left[\begin{matrix} a, d-b, d-c \\ d, d+e-b-c \end{matrix}; 1\right] $$

provided both series converge,
and the Whipple transformation formulas

$$ {}_4F_3\left[\begin{matrix} a, b, c, -n \\ e, f, 1+a+b+c-e-f-n \end{matrix}; 1\right] = \frac{(e-a)_n (f-a)_n}{(e)_n (f)_n} \times {}_4F_3\left[\begin{matrix} -n, a, 1+a+c-e-f-n, 1+a+b-e-f-n \\ 1+a+b+c-e-f-n, 1+a-e-n, 1+a-f-n \end{matrix}; 1\right] \quad [68] $$

where $n$ is a non-negative integer, and

$$ {}_7F_6\left[\begin{matrix} a, 1+\frac{a}{2}, b, c, d, e, -n \\ \frac{a}{2}, 1+a-b, 1+a-c, 1+a-d, 1+a-e, 1+a+n \end{matrix}; 1\right] = \frac{(1+a)_n (1+a-d-e)_n}{(1+a-d)_n (1+a-e)_n} \times {}_4F_3\left[\begin{matrix} 1+a-b-c, d, e, -n \\ 1+a-b, 1+a-c, -a+d+e-n \end{matrix}; 1\right] \quad [69] $$

provided $n$ is a non-negative integer.

Since about 1990, for the verification of binomial and hypergeometric series, there are automatic tools available. The book by Petkovšek *et al.* (1996) is an excellent introduction into these aspects. The philosophy is as follows. Suppose we are given a binomial or hypergeometric series $S(n) = \sum_k F(n,k)$. The Gosper–Zeilberger algorithm (see "Further reading") (cf. Petkovšek *et al.* (1996); a simplified version was presented in the reference Zeilberger in "Further reading") will find a linear recurrence

$$A_0(n)S(n) + A_1(n)S(n+1) + \cdots + A_d(n)S(n+d) = C(n) \qquad [70]$$

for some $d$, where the coefficients $A_i(n)$ are polynomials in $n$, and where $C(n)$ is a certain function in $n$, with proof!

If, for example, we suspected that $S(n) = \text{RHS}(n)$, where $\text{RHS}(n)$ is some closed-form expression, then we just have to verify that $\text{RHS}(n)$ satisfies the recurrence [70] and check $S(n) = \text{RHS}(n)$ for sufficiently many initial values of $n$ to have a proof for the identity $S(n) = \text{RHS}(n)$ for all n. On the other hand, if $\text{RHS}(n)$ was a different sum, then we would apply the algorithm to find a recurrence for $\text{RHS}(n)$. If it turns out to be the same recurrence then, again, a check of $S(n) = \text{RHS}(n)$ for a few initial values will provide a full proof of $S(n) = \text{RHS}(n)$ *for all* $n$.

Even in the case that we do not have a conjectured expression $\text{RHS}(n)$, this is not the end of the story. Given a recurrence of the type [70], the Petkovšek algorithm (see "Further reading") (cf. Petkovšek *et al.* (1996)) is able to find a closed-form solution (where "closed form" has a precise meaning), respectively tell that there is no closed-form solution.

The fascinating point about both algorithms is that neither do we have to know what the algorithm does internally nor do we have to check that. For the Petkovšek algorithm, this is obvious anyway because, once the computer says that a certain expression is a solution of [70], it is a routine matter to check that. This is less obvious for the Gosper–Zeilberger algorithm. However, what the Gosper–Zeilberger algorithm does is, for a given sum $S(n) = \sum_k F(n,k)$, it finds polynomials $A_0(n)$, $A_1(n), \ldots, A_d(n)$ and an expression $G(n,k)$ (which is, in fact, a rational multiple of $F(n,k)$), such that

$$A_0(n)F(n,k) + A_1(n)F(n+1,k) + \cdots + A_d(n)F(n+d,k) = G(n,k+1) - G(n,k) \qquad [71]$$

for some $d$. Because of the properties of $F(n,k)$ and $G(n,k)$, which are part of the theory, this is an identity which can be directly verified by clearing all common factors and checking the remaining identity between rational functions in $n$ and $k$. However, we may now sum both sides of [71] over $k$ to obtain a recurrence of the form [70].

Algorithms for multiple sums are also available (see "Further reading"). They follow ideas by Wilf and Zeilberger (1992) (of which a simplified version is presented in a Mohammed and Zeilberger preprint (see "Further reading")); however, they run more quickly in capacity problems. Schneider (2005) is currently developing a very promising new algorithmic approach to the automatic treatment of multisums. *See* q-Special Functions and Statistical Mechanics and Combinatorial Problems.

*See also:* Classical Groups and Homogeneous Spaces; Compact Groups and Their Representations; Dimer Problems; Growth Processes in Random Matrix Theory; Ordinary Special Functions; *q*-Special Functions; Saddle Point Problems; Statistical Mechanics and Combinatorial Problems.

## Further Reading

http://algo.inria.fr This site includes, among its libraries, the Maple program **gdev**.

Andrews GE (1976) The Theory of Partitions, *Encyclopedia of Mathematics and Its Applications*, vol. 2. (reprinted by Cambridge University Press, Cambridge, 1998). Reading: Addison–Wesley.

Andrews GE, Askey RA, and Roy R (1999) In: Rota GC (ed.) *Special Functions,* Encyclopedia of Mathematics and Its Applications, vol. 71. Cambridge: Cambridge University Press.

Ayoub R (1963) An Introduction to the Analytic Theory of Numbers. *Mathematical Surveys*, vol. 10, Providence, RI: American Mathematical Society.

Bergeron F, Labelle G, and Leroux P (1998) *Combinatorial Species and Tree-Like Structures*. Cambridge: Cambridge University Press.

Bousquet-Mélou M and Jehanne A (2005), Polynomial equations with one catalytic variable, algebraic series, and map enumeration. *Preprint*, ar$\chi$iv:math.CO/0504018.

Bressoud DM (1999) *Proofs and Confirmations – The Story of the Alternating Sign Matrix Conjecture*. Cambridge: Cambridge University Press.

de Bruijn NG (1964) Pólya's theory of counting. In: Beckenbach EF (ed.) *Applied Combinatorial Mathematics*, New York: Wiley, (reprinted by Krieger, Malabar, Florida, 1981).

Comtet L (1974) *Advanced Combinatorics*. Dordrecht: Reidel.

Dolbilin NP, Mishchenko AS, Shtan'ko MA, Shtogrin MI, and Zinoviev YuM (1996) Homological properties of dimer configurations for lattices on surfaces. *Functional Analysis and its Application* 30: 163–173.

Feller W (1957) *An Introduction to Probability Theory and Its Applications*, vol. 1, 2nd edn. New York: Wiley.

Fisher ME (1984) Walks, walls, wetting and melting. *Journal of Statistical Physics* 34: 667–729.

Flajolet P and Sedgewick R, *Analytic Combinatorics*, book project, available at http://algo.inria.fr.

Di Francesco P, Zinn-Justin P and Zuber J.-B. (2004), Determinant formulae for some tiling problems and application to fully packed loops, *Preprint*, ar$\chi$iv:math-ph/0410002.

Di Francesco P and Zinn-Justin P (2005), Quantum Knizhnik–Zamolodchikov equation, generalized Razumov–Stroganov sum rules and extended Joseph polynomials. *Preprint*, ar$\chi$iv:math-ph/0508059.

Galluccio A and Loebl M (1999) On the theory of Pfaffian orientations I. Perfect matchings and permanents. *Electronic Journal of Combinatorics* 6: Article #R6, 18 pp.

http://www.fmf.uni-lj.si – website of Faculty of Mathematics of University of Ljubljana. A Mathematica implementation by Marko Petkovšek is available here.

Gasper G and Rahman M (2004) *Basic Hypergeometric Series*, 2nd edn. Encyclopedia of Mathematics and Its Applications, vol. 96. Cambridge: Cambridge University Press.

de Gier J (2005) Loops matchings and alternating-sign matrices. *Discrete Mathematics* 365–388.

Humphreys JE (1990) *Reflection Groups and Coxeter Groups*. Cambridge: Cambridge University Press.

Johansson K (2002) Non-intersecting paths, random tilings and random matrices. *Probability Theory and Related Fields* 123: 225–280.

Kenyon R (2003) *An Introduction to the Dimer Model*, Lecture Notes for a Short Course at the ICTP, 2002; ar$\chi$iv:math.CO/0310326.

Koekoek R and Swarttouw RF, *The Askey–scheme of hypergeometric orthogonal polynomials and its q-analogue*, TU Delft, The Netherlands, 1998; on the www: http://aw.twi.tudelft.nl.

Krattenthaler C (1997) The enumeration of lattice paths with respect to their number of turns. In: Balakrishnan N (ed.) *Advances in Combinatorial Methods and Applications to Probability and Statistics*, pp. 29–58. Boston: Birkhäuser.

Krattenthaler C (2003), Asymptotics for random walks in alcoves of affine Weyl groups. *Preprint*, ar$\chi$iv:math.CO/0301203.

Krattenthaler C (2005a), Watermelon configurations with wall interaction: exact and asymptotic results. *Preprint*, ar$\chi$iv:math.CO/0506323.

Krattenthaler C (2005b) Advanced determinant calculus: a complement. *Linear Algebra Applications* 411: 68–166.

Krattenthaler C, Guttmann AJ, and Viennot XG (2000) Vicious walkers, friendly walkers and Young tableaux II: with a wall. *Journal of Physics A: Mathematical and General* 33: 8835–8866.

Kuperberg G (1998) An exploration of the permanent-determinant method. *Electronic Journal of Combinatorics* 5: Article #R46, 34 pp.

Labelle G and Lamathe C (2004) A shifted asymmetry index series. *Advances in Applied Mathematics* 32: 576–608.

Mohammed M and Zeilberger D (2005) Multi-variable Zeilberger and Almkvist–Zeilberger algorithms and the sharpening of Wilf–Zeilberger theory. *Advanced Applications in Mathematics* (to appear).

Mohanty SG (1979) *Lattice Path Counting and Applications*. New York: Academic Press.

Odlyzko AM (1995) Asymptotic enumeration methods. In: Graham RL, Grötschel M, and Lovász L (eds.) *Handbook of Combinatorics*, pp. 1063–1229. Amsterdam: Elsevier.

Pemantle R and Wilson MC, Twenty combinatorial examples of asymptotics derived from multivariate generating functions. *Preprint*, available at http://www.cs.auckland.ac.nz.

Petkovšek M, Wilf H, and Zeilberger D (1996) $A = B$ Wellesley: Peters AK.

http://www.mat.univie.ac.at – Website of Faculty of Mathematics, University of Vienna. It provides the manual of the Mathematica package HYP.

Propp J (1999) Enumeration of matchings: problems and progress. In: Billera L, Björner A, Greene C, Simion R, and Stanley RP (eds.) *New Perspectives in Algebraic Combinatorics*, Mathematical Sciences Research Institute Publications, vol. 38, pp. 255–291. Cambridge: Cambridge University Press.

Razumov AV and Stroganov YG (2005) Enumeration of quarter-turn symmetric alternating-sign matrices of odd order. Preprint, ar$\chi$iv:math-ph/0507003.

Robertson N, Seymour PD, and Thomas R (1999) Permanents, Pfaffian orientations, and even directed circuits. *Annals of Mathematics* 150(2): 929–975.

Schneider C (2005) A new Sigma approach to multi-summation. *Advances in Applied Mathematics* 34(4): 740–767.

Slater LJ (1966) *Generalized Hypergeometric Functions*. Cambridge: Cambridge University Press.

Stanley RP (1986) *Enumerative Combinatorics*, Pacific Grove, CA: Wadsworth & Brooks/Cole, (reprinted by Cambridge University Press, Cambridge, 1998).

Stanley RP (1999) *Enumerative Combinatorics*, vol. 2. Cambridge: Cambridge University Press.

Szegö G (1959) *Orthogonal Polynomials*, American Mathematical Society Colloquium Publications, vol. 23. New York. Providence RI: American Mathematical Society.

Tesler G (2000) Matchings in graphs on non-oriented surfaces. *Journal of Combinatorial Theory Series B* 78: 198–231.

http://www.risc.uni.linz.ac.at – website of RISC (Research Institute for Symbolic Computation). Mathematica implementations written by Peter Paule and Markus Schorn, and Axel Riese and Kurt Wegschaider are available here.

http://www.math.rutgers.edu – website of Department of Mathematics, Rutgers University. Computer implementations written by D Zeilberger are available here.

Viennot X and James W Heaps of segments, q-Bessel functions in square lattice enumeration and applications in quantum gravity. *Preprint*.

Wilf HS and Zeilberger D (1992) An algorithmic proof theory for hypergeometric (ordinary and "$q$") multisum/integral identities. *Inventiones Mathematicae* 108: 575–633.

Zeilberger D (2005) Deconstructing the Zeilberger algorithm. *Journal of Difference Equations and Applications* 11: 851–856.

# Compact Groups and Their Representations

**A Kirillov**, University of Pennsylvania, Philadelphia, PA, USA
**A Kirillov, Jr.**, Stony Brook University, Stony Brook, NY, USA



In this article, we describe the structure and representation theory of compact Lie groups. Throughout the article, $G$ is a compact real Lie group with Lie algebra $\mathfrak{g}$. Unless otherwise stated, $G$ is assumed to be connected. The word "group" will always mean a "Lie group" and the word "subgroup" will mean a closed Lie subgroup. The notation Lie($H$) stands for the Lie algebra of a Lie group $H$. We assume that the reader is familiar with the basic facts of the theory of Lie groups and Lie algebras, which can be found in Lie Groups: General Theory, or in the books listed in the bibliography.

## Examples of Compact Lie Groups

Examples of compact groups include

- finite groups,
- quotient groups $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$, or more generally, $V/L$, where $V$ is a finite-dimensional real vector space and $L$ is a lattice in $V$, that is, a discrete subgroup generated by some basis in $V$ – groups of this type are called “tori”; it is known that every commutative connected compact group is a torus;
- unitary groups $\mathrm{U}(n)$ and special unitary groups $\mathrm{SU}(n), n \geq 2$;
- orthogonal groups $\mathrm{O}(n)$ and $\mathrm{SO}(n), n \geq 3$; and
- the groups $U(n, \mathbb{H}), n \geq 1$, of unitary quaternionic transformations, which are isomorphic to $\mathrm{Sp}(n) := \mathrm{Sp}(n, \mathbb{C}) \cap \mathrm{SU}(2n)$.

The groups $\mathrm{O}(n)$ have two connected components, one of which is $\mathrm{SO}(n)$. The groups $\mathrm{SU}(n)$ and $\mathrm{Sp}(n)$ are connected and simply connected.

The groups $\mathrm{SO}(n)$ are connected but not simply connected: for $n \geq 3$, the fundamental group of $\mathrm{SO}(n)$ is $\mathbb{Z}_2$. The universal cover of $\mathrm{SO}(n)$ is a simply connected compact Lie group denoted by $\mathrm{Spin}(n)$. For small $n$, we have isomorphisms: $\mathrm{Spin}(3) \simeq \mathrm{SU}(2), \mathrm{Spin}(4) \simeq \mathrm{SU}(2) \times \mathrm{SU}(2), \mathrm{Spin}(5) \simeq \mathrm{Sp}(4)$, and $\mathrm{Spin}(6) \simeq \mathrm{SU}(4)$.

## Relation to Semisimple Lie Algebras and Lie Groups

### Reductive Groups

A Lie algebra $\mathfrak{g}$ is called

- “simple” if it is nonabelian and has no ideals different from $\{0\}$ and $\mathfrak{g}$ itself;
- “semisimple” if it is a direct sum of simple ideals; and
- “reductive” if it is a direct sum of semisimple and commutative ideals.

We call a connected Lie group $G$ “simple” or “semisimple” if $\mathrm{Lie}(G)$ has this property.

**Theorem 1** *Let $G$ be a connected compact Lie group and $\mathfrak{g} = \mathrm{Lie}(G)$. Then*

(i) *The Lie algebra $\mathfrak{g} = \mathrm{Lie}(G)$ is reductive: $\mathfrak{g} = \mathfrak{a} \oplus \mathfrak{g}'$, where $\mathfrak{a}$ is abelian and $\mathfrak{g}' = [\mathfrak{g}, \mathfrak{g}]$ is semisimple.*
(ii) *The group $G$ can be written in the form $G = (A \times K)/Z$, where $A$ is a torus, $K$ is a connected, simply connected compact semisimple Lie group, and $Z$ is a finite central subgroup in $A \times K$.*
(iii) *If $G$ is simply connected, it is a product of simple compact Lie groups.*

The proof of these results is based on the fact that the Killing form of $\mathfrak{g}$ is negative semidefinite.

**Example 1** The group $\mathrm{U}(n)$ contains as the center the subgroup $C$ of scalar matrices. The quotient group $\mathrm{U}(n)/C$ is simple and isomorphic to $\mathrm{SU}(n)/\mathbb{Z}_n$. The presentation of Theorem 1 in this case is

$$\begin{aligned} \mathrm{U}(n) &= \left(\mathbb{T}^1 \times \mathrm{SU}(n)\right)/\mathbb{Z}_n \\ &= (C \times \mathrm{SU}(n))/(C \cap \mathrm{SU}(n)) \end{aligned}$$

For the group $\mathrm{SO}(4)$ the presentation is $(\mathrm{SU}(2) \times \mathrm{SU}(2))/\{\pm(1 \times 1)\}$.

This theorem effectively reduces the study of the structure of connected compact groups to the study of simply connected compact simple Lie groups.

### Complexification of a Compact Lie Group

Recall that for a real Lie algebra $\mathfrak{g}$, its complexification is $\mathfrak{g}_{\mathbb{C}} = \mathfrak{g} \otimes \mathbb{C}$ with obvious commutator. It is also well known that $\mathfrak{g}_{\mathbb{C}}$ is semisimple or reductive iff $\mathfrak{g}$ is semisimple or reductive, respectively. There is a subtlety in the case of simple algebras: it is possible that a real Lie algebra is simple, but its complexification $\mathfrak{g}_{\mathbb{C}}$ is only semisimple. However, this problem never arises for Lie algebras of compact groups: if $\mathfrak{g}$ is a Lie algebra of a real compact Lie group, then $\mathfrak{g}$ is simple if and only if $\mathfrak{g}_{\mathbb{C}}$ is simple.

The notion of complexification for Lie groups is more delicate.

**Definition 1** Let $G$ be a connected real Lie group with Lie algebra $\mathfrak{g}$. A complexification of $G$ is a connected complex Lie group $G_{\mathbb{C}}$ (i.e., a complex manifold with a structure of a Lie group such that group multiplication is given by a complex analytic map $G_{\mathbb{C}} \times G_{\mathbb{C}} \to G_{\mathbb{C}}$), which contains $G$ as a closed subgroup, and such that $\mathrm{Lie}(G_{\mathbb{C}}) = \mathfrak{g}_{\mathbb{C}}$. In this case, we will also say that $G$ is a real form of $G_{\mathbb{C}}$.

It is not obvious why such a complexification exists at all; in fact, for arbitrary real group it may not exist. However, for compact groups we do have the following theorem.

**Theorem 2** *Let $G$ be a connected compact Lie group. Then it has a unique complexification $G_{\mathbb{C}} \supset G$. Moreover, the following properties hold:*

(i) *The inclusion $G \subset G_{\mathbb{C}}$ is a homotopy equivalence. In particular, $\pi_1(G) = \pi_1(G_{\mathbb{C}})$ and the quotient space $G_{\mathbb{C}}/G$ is contractible.*
(ii) *Every complex finite-dimensional representation of $G$ can be uniquely extended to a complex analytic representation of $G_{\mathbb{C}}$.*

Since the Lie algebra of a compact Lie group $G$ is reductive, we see that $G_{\mathbb{C}}$ must be reductive; if $G$ is semisimple or simple, then so is $G_{\mathbb{C}}$. The natural question is whether every complex reductive group can be obtained in this way. The following theorem gives a partial answer.

**Theorem 3** *Every connected complex semisimple Lie group $H$ has a compact real form: there is a compact real subgroup $K \subset H$ such that $H = K_{\mathbb{C}}$. Moreover, such a compact real form is unique up to conjugation.*

**Example 2**

(i) The unitary group $\mathrm{U}(n)$ is a compact real form of the group $\mathrm{GL}(n, \mathbb{C})$.
(ii) The orthogonal group $\mathrm{SO}(n)$ is a compact real form of the group $\mathrm{SO}(n, \mathbb{C})$.
(iii) The group $\mathrm{Sp}(n)$ is a compact real form of the group $\mathrm{Sp}(n, \mathbb{C})$.
(iv) The universal cover of $\mathrm{GL}(n, \mathbb{C})$ has no compact real form.

These results have a number of important applications. For example, they show that study of representations of a semisimple complex group $H$ can be replaced by the study of representations of its compact form; in particular, every representation is completely reducible (this argument is known as Weyl's unitary trick).

### Classification of Simple Compact Lie Groups

Theorem 1 essentially reduces such classification to classification of simply connected simple compact groups, and Theorems 2 and 3 reduce it to the classification of simple complex Lie algebras. Since the latter is well known, we get the following result.

**Theorem 4** *Let $G$ be a connected, simply connected simple compact Lie group. Then $\mathfrak{g}_{\mathbb{C}}$ must be a simple complex Lie algebra and thus can be described by a Dynkin diagram of one the following types: $A_n, B_n, C_n, D_n, E_6, E_7, E_8, F_4, G_2$.*

*Conversely, for each Dynkin diagram in the above list, there exists a unique, up to isomorphism, simply connected simple compact Lie group whose Lie algebra is described by this Dynkin diagram.*

For types $A_n, \ldots, D_n$, the corresponding compact Lie groups are well-known classical groups shown in the table below:

| $A_n, n \geq 1$ | $B_n, n \geq 2$ | $C_n, n \geq 3$ | $D_n, n \geq 4$ |
|---|---|---|---|
| $\mathrm{SU}(n+1)$ | $\mathrm{Spin}(2n+1)$ | $\mathrm{Sp}(n)$ | $\mathrm{Spin}(2n)$ |

The restrictions on $n$ in this table are made to avoid repetitions which appear for small values of $n$. Namely, $A_1 = B_1 = C_1$, which gives $\mathrm{SU}(2) = \mathrm{Spin}(3) = \mathrm{Sp}(1)$; $D_2 = A_1 \cup A_1$, which gives $\mathrm{Spin}(4) = \mathrm{SU}(2) \times \mathrm{SU}(2)$; $B_2 = C_2$, which gives $\mathrm{SO}(5) = \mathrm{Sp}(4)$; and $A_3 = D_3$, which gives $\mathrm{SU}(4) = \mathrm{Spin}(6)$. Other than that, all entries are distinct.

Exceptional groups $E_6, \ldots, G_2$ also admit explicit geometric and algebraic descriptions which are related to the exceptional nonassociative algebra $\mathbb{O}$ of the so-called octonions (or Cayley numbers). For example, the compact group of type $G_2$ can be defined as a subgroup of $\mathrm{SO}(7)$ which preserves an almost-complex structure on $S^6$. It can also be described as the subgroup of $\mathrm{GL}(7, \mathbb{R})$ which preserves one quadratic and one cubic form, or, finally, as a group of all automorphisms of $\mathbb{O}$.

## Maximal Tori

### Main Properties

In this section, $G$ is a compact connected Lie group.

**Definition 2** A "maximal torus" in $G$ is a maximal connected commutative subgroup $T \subset G$.

The following theorem lists the main properties of maximal tori.

**Theorem 5**

(i) *For every element $g \in G$, there exists a maximal torus $T \ni g$.*
(ii) *Any two maximal tori in $G$ are conjugate.*
(iii) *If $g \in G$ commutes with all elements of a maximal torus $T$, then $g \in T$.*
(iv) *A connected subgroup $H \subset G$ is a maximal torus iff the Lie algebra* $\mathrm{Lie}(H)$ *is a maximal abelian subalgebra in* $\mathrm{Lie}(G)$.

**Example 3** Let $G = \mathrm{U}(n)$. Then the set $T$ of diagonal unitary matrices is a maximal torus in $G$; moreover, every maximal torus is of this form after a suitable unitary change of basis. In particular, this implies that every element in $G$ is conjugate to a diagonal matrix.

**Example 4** Let $G = \mathrm{SO}(3)$. Then the set $D$ of diagonal matrices is a maximal commutative subgroup in $G$, but not a torus. Here $D$ consists of four elements and is not connected.

### Maximal Tori and Cartan Subalgebras

The study of maximal tori in compact Lie groups is closely related to the study of Cartan subalgebras in reductive complex Lie algebras. For convenience of readers, we briefly recall the appropriate definitions

here; details can be found in Serre (2001) or in Lie Groups: General Theory.

**Definition 3** Let $\mathfrak{a}$ be a complex reductive Lie algebra. A Cartan subalgebra $\mathfrak{h} \subset \mathfrak{a}$ is a maximal commutative subalgebra consisting of semisimple elements.

Note that for general Lie algebras Cartan subalgebra is defined in a different way; however, for reductive algebras the definition given above is equivalent to the standard one.

A choice of a Cartan subalgebra gives rise to the so-called root decomposition: if $\mathfrak{h} \subset \mathfrak{a}$ is a Cartan subalgebra in a complex reductive Lie algebra, then we can write

$$\mathfrak{a} = \mathfrak{h} \oplus \left( \bigoplus_{\alpha \in R} \mathfrak{a}_\alpha \right) \qquad [1]$$

where

$$\mathfrak{a}_\alpha = \{x \in \mathfrak{a} |\ \mathrm{ad}\, h.x = \langle \alpha, h \rangle x \, \forall h \in \mathfrak{h}\}$$
$$R = \{\alpha \in \mathfrak{h}^* - \{0\} | \mathfrak{a}_\alpha \neq 0\} \subset \mathfrak{h}^*$$

The set $R$ is called the "root system" of $\mathfrak{a}$ with respect to Cartan subalgebra $\mathfrak{h}$; elements $\alpha \in R$ are called "roots." We will also frequently use elements $\alpha^\vee \in \mathfrak{h}$ defined by $\langle \alpha^\vee, \beta \rangle = 2(\alpha, \beta)/(\alpha, \alpha)$ where $( \, , \, )$ is a nondegenerate invariant bilinear form on $\mathfrak{a}^*$ and $\langle \, , \, \rangle$ is the pairing between $\mathfrak{a}$ and $\mathfrak{a}^*$. It can be shown that so defined $\alpha^\vee$ does not depend on the choice of the form $( \, , \, )$.

**Theorem 6** *Let $G$ be a connected compact Lie group with Lie algebra $\mathfrak{g}$, and let $T \subset G$ be a maximal torus in $G$, $\mathfrak{t} = \mathrm{Lie}(T) \subset \mathfrak{g}$. Let $\mathfrak{g}_\mathbb{C}, G_\mathbb{C}$ be the complexification of $\mathfrak{g}$, $G$ as in Theorem 2.*

*Let $\mathfrak{h} = \mathfrak{t}_\mathbb{C} \subset \mathfrak{g}_\mathbb{C}$. Then $\mathfrak{h}$ is a Cartan subalgebra in $\mathfrak{g}_\mathbb{C}$, and the corresponding root system $R \subset \mathrm{i}\mathfrak{t}^*$. Conversely, every Cartan subalgebra in $\mathfrak{g}_\mathbb{C}$ can be obtained as $\mathfrak{t}_\mathbb{C}$ for some maximal torus $T \subset G$.*

### Weights and Roots

Let $G$ be semisimple. Recall that the root lattice $Q \subset \mathrm{i}\mathfrak{t}^*$ is the abelian group generated by roots $\alpha \in R$, and let the coroot lattice $Q^\vee \subset \mathrm{i}\mathfrak{t}$ be the abelian group generated by coroots $\alpha^\vee, \alpha \in R$. Define also the weight and coweight lattices by

$$P = \{\lambda | \langle \alpha^\vee, \lambda \rangle \in \mathbb{Z} \quad \forall \alpha \in R\} \subset \mathrm{i}\mathfrak{t}^*$$
$$P^\vee = \{t | \langle t, \alpha \rangle \in \mathbb{Z} \quad \forall \alpha \in R\} \subset \mathrm{i}\mathfrak{t},$$

where $\langle \cdot, \cdot \rangle$ is the pairing between $\mathfrak{t}$ and the dual vector space $\mathfrak{t}^*$.

It follows from the definition of root system that we have inclusions

$$Q \subset P \subset \mathrm{i}\mathfrak{t}^*$$
$$Q^\vee \subset P^\vee \subset \mathrm{i}\mathfrak{t} \qquad [2]$$

Both $P, Q$ are lattices in $\mathrm{i}\mathfrak{t}^*$; thus, the index $(P:Q)$ is finite. It can be computed explicitly: if $\alpha_i$ is a basis of the root system, then the fundamental weights $\omega_i$ defined by

$$\langle \alpha_i^\vee, \omega_j \rangle = \delta_{ij}$$

form a basis of $P$. The simple roots $\alpha_i$ are related to fundamental weights $\omega_j$ by the Cartan matrix $A$: $\alpha_i = \sum A_{ij}\omega_j$. Therefore, $(P:Q) = (P^\vee : Q^\vee) = |\det A|$.

Definitions of $P, Q, P^\vee, Q^\vee$ also make sense when $\mathfrak{g}$ is reductive but not semisimple. However, in this case they are no longer lattices: $\mathrm{rk} Q < \dim \mathfrak{t}^*$, and $P$ is not discrete.

We can now give more precise information about the structure of the maximal torus.

**Lemma 1** *Let $T$ be a compact connected commutative Lie group, and $\mathfrak{t} = \mathrm{Lie}(T)$ its Lie algebra. Then the exponential map is surjective and preimage of unit is a lattice $L \subset \mathfrak{t}$. There is an isomorphism of Lie groups*

$$\exp : \mathfrak{t}/L \to T$$

*In particular, $T \simeq \mathbb{R}^r/\mathbb{Z}^r = \mathbb{T}^r, r = \dim T$.*

Let $X(T) \subset \mathrm{i}\mathfrak{t}^*$ be the lattice dual to $(2\pi \mathrm{i})^{-1}L$:

$$X(T) = \{\lambda \in \mathrm{i}\mathfrak{t}^* | \langle \lambda, l \rangle \in 2\pi \mathrm{i}\mathbb{Z} \ \forall l \in L\} \qquad [3]$$

It is called the "character lattice" for $T$ (see the subsection "Examples of representations").

**Theorem 7** *Let $G$ be a compact connected Lie group, and let $T \subset G$ be a maximal torus in $G$.*

*Then $Q \subset X(T) \subset P$. Moreover, the group $G$ is uniquely determined by the Lie algebra $\mathfrak{g}$ and the lattice $X(T) \in \mathrm{i}\mathfrak{t}^*$ which can be any lattice between $Q$ and $P$.*

**Corollary** *For a given complex semisimple Lie algebra $\mathfrak{a}$, there are only finitely many (up to isomorphism) compact connected Lie groups $G$ with $\mathfrak{g}_\mathbb{C} = \mathfrak{a}$.*

*The largest of them is the simply connected group, for which $T = \mathfrak{t}/2\pi \mathrm{i} Q^\vee, X(T) = P$; the smallest is the so-called "adjoint group," for which $T = \mathfrak{t}/2\pi \mathrm{i} P^\vee$, $X(T) = Q$.*

**Example 5** Let $G = \mathrm{U}(n)$. Then $\mathrm{i}\mathfrak{t} = \{$real diagonal matrices$\}$. Choosing the standard basis of matrix

units in it, we identify it $\simeq \mathbb{R}^n$, which also allows us to identify $\mathrm{it}^* \simeq \mathbb{R}^n$. Under this identification,

$$Q = \left\{(\lambda_1, \ldots, \lambda_n) | \lambda_i \in \mathbb{Z}, \sum \lambda_i = 0\right\}$$
$$P = \left\{(\lambda_1, \ldots, \lambda_n) | \lambda_i \in \mathbb{R}, \lambda_i - \lambda_j \in \mathbb{Z}\right\}$$
$$X(T) = \mathbb{Z}^n$$

Note that $Q, P$ are not lattices: $Q \simeq \mathbb{Z}^{n-1}$, $P \simeq \mathbb{R} \times \mathbb{Z}^{n-1}$.

Now let $G = \mathrm{SU}(n)$. Then $\mathrm{it}^* = \mathbb{R}^n / \mathbb{R} \cdot (1, \ldots, 1)$, and $Q, P$ are the images of $Q, P$ for $G = \mathrm{U}(n)$ in this quotient. In this quotient they are lattices, and $(P : Q) = n$. The character lattice in this case is $X(T) = P$, since $\mathrm{SU}(n)$ is simply connected. The adjoint group is $\mathrm{PSU}(n) = \mathrm{SU}(n)/C$, where $C = \{\lambda \cdot \mathrm{id} | \lambda^n = 1\}$ is the center of $\mathrm{SU}(n)$.

### Weyl Group

Let us fix a maximal torus $T \subset G$. Let $N(T) \subset G$ be the normalizer of $T$ in $G$: $N(T) = \{g \in G \,|\, gTg^{-1} = T\}$. For any $g \in N(T)$ the transformation $A(g) : t \mapsto gtg^{-1}$ is an automorphism of $T$. According to Theorem 5, this automorphism is trivial iff $g \in T$. So in fact, it is the quotient group $N(T)/T$ which acts on $T$.

**Definition 4** The group $W = N(T)/T$ is called the "Weyl group" of $G$.

Since the Weyl group acts faithfully on $\mathfrak{t}$ and $\mathfrak{t}^*$, it is common to consider $W$ as a subgroup in $\mathrm{GL}(\mathfrak{t}^*)$. It is known that $W$ is finite.

The Weyl group can also be defined in terms of Lie algebra $\mathfrak{g}$ and its complexification $\mathfrak{g}_{\mathbb{C}}$.

**Theorem 8** *The Weyl group coincides with the subgroup in* $\mathrm{GL}(\mathrm{it}^*)$ *generated by reflections* $s_\alpha : x \mapsto x - (2(\alpha, x))/(\alpha, \alpha), \alpha \in R$, *where, as before*, $(\,,\,)$ *is a nondegenerate invariant bilinear form on* $\mathfrak{g}^*$.

**Theorem 9**

(i) *Two elements* $t_1, t_2 \in T$ *are conjugate in* $G$ *iff* $t_2 = w(t_1)$ *for some* $w \in W$.
(ii) *There exists a natural homeomorphism of quotient spaces* $G/\mathrm{Ad}G \simeq T/W$, *where* $\mathrm{Ad}G$ *stands for action of* $G$ *on itself by conjugation. (Note, however, that these quotient spaces are not manifolds: they have singularities.)*
(iii) *Let us call a function* $f$ *on* $G$ *central if* $f(hgh^{-1}) = f(g)$ *for any* $g, h \in G$. *Then the restriction map gives an isomorphism*

$$\{\textit{continuous central functions on } G\} \simeq \{W - \textit{invariant continuous functions on } T\}$$

**Example 6** Let $G = \mathrm{U}(n)$. The set of diagonal unitary matrices is a maximal torus, and the Weyl group is the symmetric group $S_n$ acting on diagonal matrices by permutations of entries. In this case, Theorem 9 shows that if $f(U)$ is a central function of a unitary matrix, then $f(U) = \tilde{f}(\lambda_1, \ldots, \lambda_n)$, where $\lambda_i$ are eigenvalues of $U$ and $\tilde{f}$ is a symmetric function in $n$ variables.

## Representations of Compact Groups

### Basic Notions

By a representation of $G$ we understand a pair $(\pi, V)$, where $V$ is a complex vector space and $\pi$ is a continuous homomorphism $G \to \mathrm{Aut}(V)$. This notation is often shortened to $\pi$ or $V$. In this article, we only consider finite-dimensional (f.d.) representations; in this case, the homomorphism $\pi$ is automatically smooth and even real-analytic.

We associate to any f.d. representation $(\pi, V)$ of $G$ the representation $(\pi_*, V)$ of the Lie algebra $\mathfrak{g} = \mathrm{Lie}(G)$ which is just the derivative of the map $\pi : G \to \mathrm{Aut}V$ at the unit point $e \in G$. In terms of the exponential map, we have the following commutative diagram:

$$\begin{array}{ccc} G & \xrightarrow{\pi} & \mathrm{Aut}V \\ {\scriptstyle\exp}\uparrow & & \uparrow{\scriptstyle\exp} \\ \mathfrak{g} & \xrightarrow{\pi_*} & \mathrm{End}V \end{array}$$

Choosing a basis in $V$, we can write the operators $\pi(g)$ and $\pi_*(X)$ in matrix form and consider $\pi$ and $\pi_*$ as matrix-valued functions on $G$ and $\mathfrak{g}$. The diagram above means that

$$\pi(\exp X) = e^{\pi_*(X)} \qquad [4]$$

Recall that if $G$ is connected, simply connected, then every representation of $\mathfrak{g}$ can be uniquely lifted to a representation of $G$. Thus, classification of representations of connected simply connected Lie groups is equivalent to the classification of representations of Lie algebras.

Let $(\pi_1, V_1)$ and $(\pi_2, V_2)$ be two representations of the same group $G$. An operator $A \in \mathrm{Hom}(V_1, V_2)$ is called an "intertwining operator," or simply an "intertwiner," if $A \circ \pi_1(g) = \pi_2(g) \circ A$ for all $g \in G$. Two representations are called "equivalent" if they admit an invertible intertwiner. In this case, using an appropriate choice of bases, we can write $\pi_1$ and $\pi_2$ by the same matrix-valued function.

Let $(\pi, V)$ be a representation of $G$. If all operators $\pi(g), g \in G$, preserve a subspace $V_1 \subset V$, then the restrictions $\pi_1(g) = \pi(g)|_{V_1}$ define a "subrepresentation" $(\pi_1, V_1)$ of $(\pi, V)$. In this case, the quotient space $V_2 = V/V_1$ also has a canonical structure of a representation, called the "quotient representation."

A representation $(\pi, V)$ is called "reducible" if it has a nontrivial (different from $V$ and $\{0\}$) subrepresentation. Otherwise it is called "irreducible."

We call representation $(\pi, V)$ "unitary" if $V$ is a Hilbert space and all operators $\pi(g), g \in G$, are unitary, that is, given by unitary matrices in any orthonormal basis. We use a short term "unirrep" for a "unitary irreducible representation."

### Main Theorems

The following simple but important result was one of the first discoveries in representation theory. It holds for representations of any group, not necessarily compact.

**Theorem 10** (Schur lemma). *Let $(\pi_i, V_i), i = 1, 2$, be any two irreducible finite-dimensional representations of the same group $G$. Then any intertwiner $A : V_1 \to V_2$ is either invertible or zero.*

**Corollary 1** *If $V$ is an irreducible f.d. representation, then any intertwiner $A : V \to V$ is scalar: $A = c \cdot \mathrm{id}, c \in \mathbb{C}$.*

**Corollary 2** *Every irreducible representation of a commutative group is one dimensional.*

The following theorem is one of the fundamental results of the representation theory of compact groups. Its proof is based on the technique of invariant integrals on a compact group, which will be discussed in the next section.

**Theorem 11**

(i) *Any f.d. representation of a compact group is equivalent to a unitary representation.*

(ii) *Any f.d. representation is completely reducible: it can be decomposed into direct sum*

$$V = \bigoplus n_i V_i$$

*where $V_i$ are pairwise nonequivalent unirreps. Numbers $n_i \in \mathbb{Z}_+$ are called "multiplicities."*

### Examples of Representations

The representation theory looks rather different for abelian (i.e., commutative) and nonabelian groups. Here we consider two simplest examples of both kinds.

Our first example is a one-dimensional compact connected Lie group. Topologically, it is a circle which we realize as a set $\mathbb{T} \simeq \mathrm{U}(1)$ of all complex numbers $t$ with absolute value 1.

Every unirrep of $\mathbb{T}$ is one dimensional; thus, it is just a continuous multiplicative map $\pi$ of $\mathbb{T}$ to itself. It is well known that every such map has the form

$$\pi_k(t) = t^k \quad \text{for some } k \in \mathbb{Z}$$

The collection of all unirreps of $\mathbb{T}$ is itself a group, called "Pontrjagin dual" of $\mathbb{T}$ and denoted by $\widehat{\mathbb{T}}$. This group is isomorphic to $\mathbb{Z}$.

By Theorem 11, any f.d. representation $\pi$ of $\mathbb{T}$ is equivalent to a direct sum of one-dimensional unirreps. So, an equivalence class of $\pi$ is defined by the multiplicity function $\mu$ on $\widehat{\mathbb{T}} = \mathbb{Z}$ taking nonnegative values:

$$\pi \simeq \sum_{k \in \mathbb{Z}} \mu(k) \cdot \pi_k$$

The many-dimensional case of compact connected abelian Lie group can be treated in a similar way. Let $T$ be a torus, that is, an abelian compact group, $\mathfrak{t} = \mathrm{Lie}(T)$. Then every irreducible representation of $T$ is one dimensional and thus is defined by a group homomorphism $\chi : T \to \mathbb{T}^1 = \mathrm{U}(1)$. Such homomorphisms are called "characters" of $T$. One easily sees that such characters themselves form a group (Pontrjagin dual of $T$). If we denote by $L$ the kernel of the exponential map $\mathfrak{t} \to T$ (see Lemma 1), one easily sees that every character has a form

$$\chi(\exp(t)) = \mathrm{e}^{\langle t, \lambda \rangle}, \quad t \in \mathfrak{t}, \; \lambda \in X(T)$$

where $X(T) \subset \mathrm{i}\mathfrak{t}^*$ is the lattice defined by [3]. Thus, we can identify the group of characters $\widehat{T}$ with $X(T)$. In particular, this shows that $\widehat{T} \simeq \mathbb{Z}^{\dim T}$.

The second example is the group $G = \mathrm{SU}(2)$, the simplest connected, simply connected nonabelian compact Lie group. Topologically, $G$ is a three-dimensional sphere since the general element of $G$ is a matrix of the form

$$g = \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}, \quad a, b \in \mathbb{C}, \; |a|^2 + |b|^2 = 1$$

Let $V$ be two-dimensional complex vector space, realized by column vectors $\binom{u}{v}$. The group $G$ acts naturally on $V$. This action induces the representation $\Pi$ of $G$ in the space $S(V)$ of all polynomials in $u$, $v$. It is infinite dimensional, but has many f.d. subrepresentations. In particular, let $S^k(V)$, or simply $S^k$, be the space of all homogeneous polynomials of degree $k$. Clearly, $\dim S^k = k + 1$.

It turns out that the corresponding f.d. representations $(\Pi_k, S^k), k \geq 0$, are irreducible, pairwise nonequivalent, and exhaust the set $\widehat{G}$ of all unirreps.

Some particular cases are of special interest:

1. $k = 0$. The space $V_0$ consists of constant functions and $\Pi_0$ is the trivial one-dimensional representation: $\Pi_0(g) \equiv 1$.
2. $k = 1$. The space $V_1$ is identical to $V$ and $\Pi_1$ is just the tautological representation $\pi(g) \equiv g$.
3. $k = 2$. The space $V_2$ is spanned by monomials $u^2, uv, v^2$. The remarkable fact is that this

representation is equivalent to a real one. Namely, in the new basis

$$x = \frac{u^2+v^2}{2}, \qquad y = \frac{u^2-v^2}{2\mathrm{i}}, \qquad z = \mathrm{i}uv$$

we have

$$\Pi_2\begin{pmatrix} a & b \\ -\overline{b} & \overline{a}\end{pmatrix} = \begin{pmatrix} \mathrm{Re}(a^2+b^2) & 2\mathrm{Im}(ab) & \mathrm{Im}(b^2-a^2) \\ 2\mathrm{Im}(a\overline{b}) & |a|^2-|b|^2 & 2\mathrm{Re}(a\overline{b}) \\ \mathrm{Im}(a^2+b^2) & 2\mathrm{Re}(ab) & \mathrm{Re}(a^2-b^2)\end{pmatrix}$$

This formula defines a homomorphism $\Pi_2 : \mathrm{SU}(2) \to \mathrm{SO}(3)$. It can be shown that this homomorphism is surjective, and its kernel is the subgroup $\{\pm 1\} \subset \mathrm{SU}(2)$:

$$1 \to \{\pm 1\} \hookrightarrow \mathrm{SU}(2) \xrightarrow{\Pi_2} \mathrm{SO}(3) \to 1$$

The simplest way to see it is to establish the equivalence of $\Pi_2$ with the adjoint representation of $G$ in $\mathfrak{g}$. The corresponding intertwiner is

$$S^2 \ni (\alpha+\mathrm{i}\gamma)u^2 + 2\mathrm{i}\beta uv + (\alpha-\mathrm{i}\gamma)v^2 \longleftrightarrow \begin{pmatrix} \mathrm{i}\beta & \alpha+\mathrm{i}\gamma \\ -\alpha+\mathrm{i}\gamma & -\mathrm{i}\beta\end{pmatrix} \in \mathfrak{g}$$

Note that SU(2) and SO(3) are the only compact groups associated with the Lie algebra $\mathfrak{sl}(2,\mathbb{C})$.

The group $G$ contains the subgroup $H$ of diagonal matrices, isomorphic to $\mathbb{T}^1$. Consider the restriction of $\Pi_n$ to $\mathbb{T}^1$. It splits into the sum of unirreps $\pi_k$ as follows:

$$\mathrm{Res}^G_{T^1}\Pi_n = \sum_{s=0}^{s=[n/2]} \pi_{n-2s}$$

The characters $\pi_k$ which enter this decomposition are called the weights of $\Pi_n$. The collection of all weights (together with multiplicities) forms a multiset in $\widehat{\mathbb{T}}$ denoted by $P(\Pi_n)$ or $P(S^n)$.

Note the following features of this multiset:

1. $P(\Pi_n)$ is invariant under reflection $k \mapsto -k$.
2. All weights of $\Pi_n$ are congruent modulo 2.
3. The nonequivalent unirreps have different multisets of weights.

Below we show how these features are generalized to all compact connected Lie groups.

## Fourier Transform

### Haar Measure and Invariant Integral

The important feature of compact groups is the existence of the so-called "invariant integral," or "average."

**Theorem 12** *For every compact Lie group $G$, there exists a unique measure $\mathrm{d}g$ on $G$, called "Haar measure," which is invariant under left shifts $L_g : h \mapsto gh$ and satisfies $\int_G \mathrm{d}g = 1$.*

*In addition, this measure is also invariant under right shifts $h \mapsto hg$ and under involution $h \mapsto h^{-1}$.*

Invariance of the Haar measure implies that for every integrable function $f(g)$, we have

$$\int_G f(g)\,\mathrm{d}g = \int_G f(hg)\,\mathrm{d}g = \int_G f(gh)\,\mathrm{d}g = \int_G f(g^{-1})\,\mathrm{d}g$$

For a finite group $G$, the integral with respect to the Haar measure is just averaging over the group:

$$\int_G f(g)\,\mathrm{d}g = \frac{1}{|G|}\sum_{g\in G} f(g)$$

For compact connected Lie groups, the Haar measure is given by a differential form of top degree which is invariant under right and left translations.

For a torus $T^n = \mathbb{R}^n/\mathbb{Z}^n$ with real coordinates $\theta_k \in \mathbb{R}/\mathbb{Z}$ or complex coordinates $t_k = \mathrm{e}^{2\pi\mathrm{i}\theta_k}$, the Haar measure is $\mathrm{d}^n\theta := \mathrm{d}\theta_1\mathrm{d}\theta_2\cdots\mathrm{d}\theta_n$ or

$$\mathrm{d}^n t := \prod_{k=1}^{n} \frac{\mathrm{d}t_k}{2\pi\mathrm{i}t_k}$$

In particular, consider a central function $f$ (see Theorem 9). Since every conjugacy class contains elements of the maximal torus $T$ (see Theorem 5), such a function is determined by its values on $T$, and the integral of a central function can be reduced to integration over $T$. The resulting formula is called "Weyl integration formula." For $G = \mathrm{U}(n)$ it looks as follows:

$$\int_{U(n)} f(g)\mathrm{d}g = \frac{1}{n!}\int_T f(t)\prod_{i<j}|t_i - t_j|^2\mathrm{d}^n t$$

where $T$ is the maximal torus consisting of diagonal matrices

$$t = \mathrm{diag}(t_1,\ldots,t_n), \qquad t_k = \mathrm{e}^{2\pi\mathrm{i}\theta_k}$$

and $\mathrm{d}^n t$ is defined above.

Weyl integration formula for arbitrary compact group $G$ can be found in Simon (1996) or Bump (2004, section 18).

The main applications of the Haar measure are the proof of complete reducibility theorem (Theorem 11) and orthogonality relations (see below).

### Orthogonality Relations and Peter–Weyl Theorem

Let $V_1, V_2$ be unirreps of a compact group $G$. Taking any linear operator $A : V_1 \to V_2$ and averaging the expression $A(g) := \pi_2(g^{-1}) \circ A \circ \pi_1(g)$ over

$G$, we get an intertwining operator $\langle A \rangle = \int_G A(g)\mathrm{d}g$. Comparing this fact with the Schur lemma, one obtains the following fundamental results.

Let $(\pi, V)$ be any unirrep of a compact group $G$. Choose any orthonormal basis $\{v_k, 1 \le k \le \dim V\}$ in $V$ and denote by $t^V_{kl}$, or $t^\pi_{kl}$, the function on $G$ defined by

$$t^V_{kl}(g) = (\pi(g)v_l, v_k)$$

The functions $t^V_{kl}$ are called "matrix elements" of the unirrep $(\pi, V)$.

**Theorem 13** (Orthogonality relations)

- (i) *The matrix elements $t^V_{kl}$ are pairwise orthogonal and have norm* $(\dim V)^{-1/2}$ *in* $L^2(G, \mathrm{d}g)$.
- (ii) *The matrix elements corresponding to equivalent unirreps span the same subspace in* $L^2(G, \mathrm{d}g)$.
- (iii) *The matrix elements of two nonequivalent unirreps are orthogonal.*
- (iv) *The linear span of all matrix elements of all unirreps is dense in* $\mathrm{C}(G), C^\infty(G)$, *and in* $L^2(G, \mathrm{d}g)$ (*generalized Peter–Weyl theorem*).

In particular, this theorem implies that the set $\widehat{G}$ of equivalence classes of unirreps is countable. For an f.d. representation $(\pi, V)$ we introduce the character of $\pi$ as a function

$$\chi_\pi(g) = \mathrm{tr}\pi(g) = \sum_{k=1}^{\dim V} t^\pi_{kk}(g) \qquad [5]$$

It is obviously a central function on $G$.

**Remark** Traditionally, in representation theory the word "character" has two different meanings: (1) a multiplicative map from a group to U(1), and (2) the trace of a representation operator $\pi(g)$. For one-dimensional representations both notions coincide.

From the orthogonality relations we get the following result.

**Corollary** *The characters of unirreps of $G$ form an orthonormal basis in the subspace of central functions in* $L^2(G, \mathrm{d}g)$.

### Noncommutative Fourier Transform

The noncommutative Fourier transform on a compact group $G$ is defined as follows. Let $\widehat{G}$ denote the set of equivalence classes of unirreps of $G$. Choose for any $\lambda \in \widehat{G}$ a representation $(\pi_\lambda, V_\lambda)$ of class $\lambda$ and an orthonormal basis in $V_\lambda$. Denote by $d(\lambda)$ the dimension of $V_\lambda$.

We introduce the Hilbert space $L^2(\widehat{G})$ as the space of matrix-valued functions on $\widehat{G}$ whose value at a point $\lambda \in \widehat{G}$ belongs to $\mathrm{Mat}_{d(\lambda)}(\mathbb{C})$. The norm is defined as

$$\|F\|^2_{L^2(\widehat{G})} = \sum_{\lambda \in \widehat{G}} d(\lambda) \cdot \mathrm{tr}(F(\lambda)F(\lambda)^*)$$

For a function $f$ on $G$ define its Fourier transform $\widetilde{f}$ as a matrix-valued function on $\widehat{G}$:

$$\widetilde{f}(\lambda) = \int_G f(g^{-1})\pi_\lambda(g)\mathrm{d}g$$

Note that in the case $G = \mathbb{T}^1$ this transform associates to a function $f$ the set of its Fourier coefficients. In general this transform keeps some important features of Fourier coefficients.

**Theorem 14**

- (i) *For a function $f \in L^1(G, \mathrm{d}g)$ the Fourier transform $\widetilde{f}$ is well defined and bounded* (*by matrix norm*) *function on* $\widehat{G}$.
- (ii) *For a function* $f \in L^1(G, \mathrm{d}g) \cap L^2(G, \mathrm{d}g)$ *the following analog of the Plancherel formula holds*:

$$\|f\|^2_{L^2(G,\mathrm{d}g)} := \int_G |f(g)|^2 \mathrm{d}g = \sum_{\lambda \in \widehat{G}} d(\lambda) \cdot \mathrm{tr}(\widetilde{f}(\lambda)\widetilde{f}(\lambda)^*) =: \|\widetilde{f}\|^2_{L^2(\widehat{G})}$$

- (iii) *The following inversion formula expresses $f$ in terms of $\widetilde{f}$*:

$$f(g) = \sum_{\lambda \in \widehat{G}} d(\lambda) \cdot \mathrm{tr}(\widetilde{f}(\lambda)\pi_\lambda(g))$$

- (iv) *The Fourier transform sends the convolution to the matrix multiplication*:

$$\widetilde{f_1 * f_2} = \widetilde{f}_1 \cdot \widetilde{f}_2$$

*where the convolution product $*$ is defined by*

$$(f_1 * f_2)(h) = \int_G f_1(hg) f_2(g^{-1})\,\mathrm{d}g$$

Note the special case of the inversion formula for $g = e$:

$$f(e) = \sum_{\lambda \in \widehat{G}} d(\lambda) \cdot \mathrm{tr}(\widetilde{f}(\lambda)),$$

or

$$\delta(g) = \sum_{\lambda \in \widehat{G}} d(\lambda) \cdot \chi_\lambda(g)$$

where $\delta(g)$ is Dirac's delta-function: $\int_G f(g)\delta(g)\,\mathrm{d}g = f(e)$. Thus, we get a presentation of Dirac's delta-function as a linear combination of characters.

## Classification of Finite-Dimensional Representations

In this section, we give a classification of unirreps of a connected compact Lie group $G$.

### Weight Decomposition

Let $G$ be a connected compact group with maximal torus $T$, and let $(\pi, V)$ be a f.d. representation of $G$. Restricting it to $T$ and using complete reducibility, we get the following result.

**Theorem 15** *The vector space $V$ can be written in the form*

$$V = \bigoplus_{\lambda \in X(T)} V_\lambda, \\ V_\lambda = \{v \in V | \pi_*(t)v = \langle \lambda, t \rangle v \ \forall t \in \mathfrak{t}\} \qquad [6]$$

*where $X(T)$ is the character group of $T$ defined by* [3].
*The spaces $V_\lambda$ are called* "weight subspaces," *vectors $v \in V_\lambda$* – "weight vectors" *of weight $\lambda$. The set*

$$P(V) = \{\lambda \in X(T) | V_\lambda \neq \{0\}\} \qquad [7]$$

*is called the* "set of weights" *of $\pi$, or the* "spectrum" *of $\mathrm{Res}_T^G \pi$, and*

$$\mathrm{mult}_{(\pi,V)}(\lambda) := \dim V_\lambda$$

*is called the* "multiplicity" *of $\lambda$ in $V$.*

The next theorem easily follows from the definition of the Weyl group.

**Theorem 16** *For any f.d. representation $V$ of $G$, the set of weights with multiplicities is invariant under the action of the Weyl group*:

$$w(P(V)) = P(V), \qquad \mathrm{mult}_{(\pi,V)}(\lambda) = \mathrm{mult}_{(\pi,V)}(w(\lambda))$$

*for any $w \in W$.*

### Classification of Unirreps

Recall that $R$ is the root system of $\mathfrak{g}_\mathbb{C}$. Assume that we have chosen a basis of simple roots $\alpha_1, \ldots, \alpha_r \subset R$. Then $R = R_+ \cup R_-$; roots $\alpha \in R_+$ can be written as a linear combination of simple roots with positive coefficients, and $R_- = -R_+$.

A (not necessarily f.d.) representation of $\mathfrak{g}_\mathbb{C}$ is called a "highest-weight representation" if it is generated by a single vector $v \in V_\lambda$ (the highest-weight vector) such that $\mathfrak{g}_\alpha v = 0$ for all positive roots $\alpha \in R_+$.

It can be shown that for every $\lambda \in X(T)$, there is a unique irreducible highest-weight representation of $\mathfrak{g}_\mathbb{C}$ with highest weight $\lambda$, which is denoted $L(\lambda)$. However, this representation can be infinite dimensional; moreover, it may not be possible to lift it to a representation of $G$.

**Definition 5** A weight $\lambda \in X(T)$ is called "dominant" if $\langle \lambda, \alpha_i^\vee \rangle \in \mathbb{Z}_+$ for any simple root $\alpha_i$. The set of all dominant weights is denoted by $X_+(T)$.

**Theorem 17**

(i) *All weights of $L(\lambda)$ are of the form $\mu = \lambda - \Sigma n_i \alpha_i$, $n_i \in \mathbb{Z}_+$.*
(ii) *Let $\lambda \in X_+$. Then the irreducible highest-weight representation $L(\lambda)$ is f.d. and lifts to a representation of $G$.*
(iii) *Every irreducible f.d. representation of $G$ is of the form $L(\lambda)$ for some $\lambda \in X_+$.*

Thus, we have a bijection {unirreps of $G$} $\leftrightarrow X_+$.

**Example 7** Let $G = \mathrm{SU}(2)$. There is a unique simple root $\alpha$ and the unique fundamental weight $\omega$, related by $\alpha = 2\omega$. Therefore, $X_+ = \mathbb{Z}_+ \cdot \omega$ and unirreps are indexed by non-negative integers. The representation with highest weight $k \cdot \omega$ is precisely the representation $\Pi_k$ constructed in the subsection "Examples of representations."

**Example 8** Let $G = \mathrm{U}(n)$. Then $X = \mathbb{Z}^n$, and $X_+ = \{(\lambda_1, \ldots, \lambda_n) \in \mathbb{Z}^n \,|\, \lambda_1 \geq \cdots \geq \lambda_n\}$. Such objects are well known in combinatorics: if we additionally assume that $\lambda_n \geq 0$, then such dominant weights are in bijection with partitions with $n$ parts. They can also be described by "Young diagrams" with $n$ rows (see Fulton and Harris (1991)).

### Explicit Construction of Representations

In addition to description of unirreps as highest-weight representations, they can also be constructed in other ways. In particular, they can be defined analytically as follows. Let $B = HN_+$ be the Borel subgroup in $G_\mathbb{C}$; here $H = \exp \mathfrak{h}$, $N_+ = \exp \sum_{\alpha \in R_+} (\mathfrak{g}_\mathbb{C})_\alpha$. For $\lambda \in \mathfrak{h}^*$, let $\chi_\lambda : B \to \mathbb{C}^\times$ be a multiplicative map defined by

$$\chi_\lambda(hn) = \mathrm{e}^{\langle h, \lambda \rangle} \qquad [8]$$

**Theorem 18** (Cartan–Borel–Weil). *Let $\lambda \in X(T)$. Denote by $V(\lambda)$ the space of complex-analytic functions on $G_\mathbb{C}$ which satisfy the following transformation property*:

$$f(gb) = \chi_\lambda^{-1}(b) f(g), \quad g \in G_\mathbb{C}, \ b \in B$$

*The group $G_\mathbb{C}$ acts on $V(\lambda)$ by left shifts*:

$$(\pi(g)f)(h) = f(g^{-1}h) \qquad [9]$$

*Then*

(i) $V(\lambda) \neq \{0\}$ *iff* $-\lambda \in X_+$.

(ii) *If* $-\lambda \in X_+$, *the representation of* $G$ *in* $V(\lambda)$ *is equivalent to* $L(w_0(\lambda))$, *where* $w_0 \in W$ *is the unique element of the Weyl group which sends* $R_+$ *to* $R_-$.

This theorem can also be reformulated in more geometric terms: the spaces $V(\lambda)$ are naturally interpreted as spaces of global sections of appropriate line bundles on the "flag variety" $\mathcal{B} = G_{\mathbb{C}}/B = G/T$.

For classical groups, irreducible representations can also be constructed explicitly as the subspaces in tensor powers $(\mathbb{C}^n)^{\otimes k}$, transforming in a certain way under the action of the symmetric group $S_k$.

## Characters and Multiplicities

### Characters

Let $(\pi, V)$ be a f.d. representation of $G$ and let $\chi_\pi$ be its character as defined by [5]. Since $\chi_\pi$ is central, and every element in $G$ is conjugate to an element of $T$, $\chi_\pi$ is completely determined by its restriction to $T$, which can be computed from the weight decomposition [6]:

$$\chi_\pi|_T = \sum_{\lambda \in X(T)} \dim V_\lambda \cdot e_\lambda = \sum_{\lambda \in X(T)} \mathrm{mult}_\pi \lambda \cdot e_\lambda \qquad [10]$$

where $e_\lambda$ is the function on $T$ defined by $e_\lambda(\exp(t)) = \mathrm{e}^{\langle t, \lambda \rangle}$, $t \in \mathfrak{t}$. Note that $e_{\lambda+\mu} = e_\lambda e_\mu$ and that $e_0 = 1$.

### Weyl Character Formula

**Theorem 19** (Weyl character formula). *Let* $\lambda \in X_+$. *Then*

$$\chi_{L(\lambda)} = \frac{A_{\lambda+\rho}}{A_\rho}, \qquad A_\mu = \sum_{w \in W} \varepsilon(w) e_{w(\mu)}$$

*where, for* $w \in W$, *we denote* $\varepsilon(w) = \det w$ *considered as a linear map* $\mathfrak{t}^* \to \mathfrak{t}^*$, *and* $\rho = (1/2) \sum_{R_+} \alpha$.

In particular, computing the value of the character at point $t = 0$ by L'Hopital's rule, it is possible to deduce the following formula for the dimension of irreducible representations:

$$\dim L(\lambda) = \prod_{\alpha \in R_+} \frac{\langle \alpha^\vee, \lambda + \rho \rangle}{\langle \alpha^\vee, \rho \rangle} \qquad [11]$$

**Example 9** Let $G = \mathrm{SU}(2)$. Then Weyl character formula gives, for irreducible representation $\Pi_k$ with highest weight $k \cdot \omega$,

$$\chi_{\Pi_k} = \frac{x^{k+1} - x^{-(k+1)}}{x - x^{-1}} = x^k + x^{k-2} + \cdots + x^{-k}, \quad x = e_\omega$$

which implies $\dim \Pi_k = k + 1$.

Weyl character formula is equivalent to the following formula for weight multiplicities, due to Kostant:

$$\mathrm{mult}_{L(\lambda)} \mu = \sum_{w \in W} \varepsilon(w) K(w(\lambda + \rho) - \rho - \mu)$$

where $K$ is Kostant's partition function: $K(\tau)$ is the number of ways of writing $\tau$ as a sum of positive roots (with repetitions).

For classical Lie groups such as $G = \mathrm{U}(n)$, there are more explicit combinatorial formulas for weight multiplicities; for $\mathrm{U}(n)$, the answer can be written in terms of the number of "Young tableaux" of a given shape. Details can be found in Fulton and Harris (1991).

### Tensor Product Multiplicities

Let $(\pi, V)$ be a f.d. representation of $G$. By complete reducibility, one can write $V = \Sigma n_\lambda L(\lambda)$. The coefficients $n_\lambda$ are called multiplicities; finding them is an important problem in many applications. In particular, a special case of this is finding the multiplicities in tensor product of two unirreps:

$$L(\lambda) \otimes L(\mu) = \sum N^\nu_{\lambda\mu} L(\nu)$$

Characters provide a practical tool for computing multiplicities: since characters of unirreps are linearly independent, multiplicities can be found from the condition that $\chi_V = \Sigma n_\lambda \chi_{L(\lambda)}$. In particular,

$$\chi_{L(\lambda)} \chi_{L(\mu)} = \sum N^\nu_{\lambda\mu} \chi_{L(\nu)}$$

**Example 10** For $G = \mathrm{SU}(2)$, tensor product multiplicities are given by

$$\Pi_n \otimes \Pi_m = \oplus \Pi_l$$

where the sum is taken over all $l$ such that $|m - n| \leq l \leq m + n$, $m + n + l$ is even.

For $G = \mathrm{U}(n)$, there is an algorithm for finding the tensor product multiplicities, formulated in the language of Young tableaux (Littlewood–Richardson rule). There are also tables and computer programs for computing these multiplicities; some of them are listed in the bibliography.

*See also:* Classical Groups and Homogeneous Spaces; Combinatorics: Overview; Equivariant Cohomology and

the Cartan Model; Finite Group Symmetry Breaking; Lie Groups: General Theory; Ljusternik–Schnirelman Theory; Noncommutative Geometry and the Standard Model; Optimal Cloning of Quantum States; Ordinary Special Functions; Quasiperiodic Systems; Symmetry Classes in Random Matrix Theory.

## Further Reading

Bump D (2004) *Lie Groups*. New York: Springer.

Bröcker T and tom Dieck T (1995) *Representations of Compact Lie Groups*, Graduate Texts in Mathematics, vol. 98. New York: Springer.

Fulton W and Harris J (1991) *Representation Theory*. New York: Springer.

Knapp A (2002) *Lie Groups beyond an Introduction*, 2nd edn. Boston: Birkhaüser.

LiE: A Computer algebra package for Lie group computations, available from http://young.sp2mi.univ-poitiers.fr

McKay WG, Patera J, and Rand DW (1990) *Tables of Representations of Simple Lie Algebras, vol. I. Exceptional Simple Lie Algebras.* Montreal: CRM.

Serre J-P (2001) *Complex Semisimple Lie Algebras*. Berlin: Springer.

Simon B (1996) *Representations of Finite and Compact Groups*. Providence, RI: American Mathematical Society.

Zelobenko DP (1973) *Compact Lie Groups and Their Representations*. Providence, RI: American Mathematical Society.

# Compactification of Superstring Theory

**M R Douglas**, Rutgers, The State University of New Jersey, Piscataway, NJ, USA



## Introduction

Superstring theories and M-theory, at present the best candidate quantum theories which unify gravity, Yang–Mills fields, and matter, are directly formulated in ten and eleven spacetime dimensions. To obtain a candidate theory of our four-dimensional universe, one must find a solution of one of these theories whose low-energy physics is well described by a four-dimensional effective field theory (EFT), containing the well-established standard model (SM) of particle physics coupled to Einstein's general relativity (GR). The standard paradigm for finding such solutions is compactification, along the lines originally proposed by Kaluza and Klein in the context of higher-dimensional general relativity. One postulates that the underlying $D$-dimensional spacetime is a product of four-dimensional Minkowski spacetime, with a $(D-4)$-dimensional compact and small Riemannian manifold $K$. One then finds that low-energy physics effectively averages over $K$, leading to a four-dimensional EFT whose field content and Lagrangian are determined in terms of the topology and geometry of $K$.

Of the huge body of prior work on this subject, the part most relevant for string/M-theory is supergravity compactification, as in the limit of low energies, small curvatures and weak coupling, the various string theories and M-theory reduce to ten- and eleven-dimensional supergravity theories. Many of the qualitative features of string/M-theory compactification, and a good deal of what is known quantitatively, can be understood simply in terms of compactification of these field theories, with the addition of a few crucial ingredients from string/M-theory. Thus, most of this article will restrict attention to this case, leaving many "stringy" topics to the articles on conformal field theory, topological string theory, and so on. We also largely restrict attention to compactifications based on Ricci-flat compact spaces. There is an equally important class in which $K$ has positive curvature; these lead to anti-de Sitter (AdS) spacetimes and are discussed in the article on AdS/CFT (*see* AdS/CFT Correspondence).

After a general review, we begin with compactification of the heterotic string on a three complex dimensional Calabi–Yau manifold. This was the first construction which led convincingly to the SM, and remains one of the most important examples. We then survey the various families of compactifications to higher dimensions, with an eye on the relations between these compactifications which follow from superstring duality. We then discuss some of the phenomena which arise in the regimes of large curvature and strong coupling. In the final section, we bring these ideas together in a survey of the various known four-dimensional constructions.

## General Framework

Let us assume we are given a $D$- $(=d+k)$ dimensional field theory $\mathcal{T}$. A compactification is then a $D$-dimensional spacetime which is topologically the product of a $d$-dimensional spacetime with an $k$-dimensional manifold $K$, the compactification or "internal" manifold, carrying a Riemannian metric and with definite expectation values for all other fields in $\mathcal{T}$. These must solve the equations of motion, and preserve $d$-dimensional Poincaré invariance (or, perhaps another $d$-dimensional symmetry group).

The most general metric ansatz for a Poincaré invariant compactification is

$$G_{IJ} = \begin{pmatrix} f & \eta_{\mu\nu} & 0 \\ & 0 & G_{ij} \end{pmatrix}$$

where the tangent space indices are $0 \le I < d + k = D, 0 \le \mu < d$, and $1 \le i \le k$. Here $\eta_{\mu\nu}$ is the Minkowski metric, $G_{ij}$ is a metric on $K$, and $f$ is a real-valued function on $K$ called the "warp factor."

As the simplest example, consider pure $D$-dimensional GR. in this case, Einstein's equations reduce to Ricci flatness of $G_{IJ}$. Given our metric ansatz, this requires $f$ to be constant, and the metric $G_{ij}$ on $K$ to be Ricci flat. Thus, any $K$ which admits such a metric, for example, the $k$-dimensional torus, will lead to a compactification.

Typically, if a manifold admits a Ricci-flat metric, it will not be unique; rather there will be a moduli space of such metrics. Physically, one then expects to find solutions in which the choice of Ricci-flat metric is slowly varying in $d$-dimensional spacetime. General arguments imply that such variations must be described by variations of $d$-dimensional fields, governed by an EFT. Given an explicit parametrization of the family of metrics, say $G_{ij}(\phi^\alpha)$ for some parameters $\phi^\alpha$, in principle the EFT could be computed explicitly by promoting the parameters to $d$-dimensional fields, substituting this parametrization into the $D$-dimensional action, and expanding in powers of the $d$-dimensional derivatives. In pure GR, we would find the four-dimensional effective Lagrangian

$$\mathcal{L}_{\mathrm{EFT}} = \int \mathrm{d}^k y \sqrt{\det G(\phi)} R^{(4)} + \sqrt{\det G(\phi)} G^{ik}(\phi) G^{jl}(\phi) \frac{\partial G_{ij}}{\partial \phi^\alpha} \frac{\partial G_{kl}}{\partial \phi^\beta} \partial_\mu \phi^\alpha \partial_\mu \phi^\beta + \cdots \qquad [1]$$

While this is easily evaluated for $K$ a symmetric space or torus, in general a direct computation of $\mathcal{L}_{\mathrm{EFT}}$ is impossible. This becomes especially clear when one learns that the Ricci-flat metrics $G_{ij}$ are not explicitly known for the examples of interest. Nevertheless, clever indirect methods have been found that give a great deal of information about $\mathcal{L}_{\mathrm{EFT}}$; this is much of the art of superstring compactification. However, in this section, let us ignore this point and continue as if we could do such computations explicitly.

Given a solution, one proceeds to consider its small perturbations, which satisfy the linearized equations of motion. If these include exponentially growing modes (often called "tachyons"), the solution is unstable. (Note that this criterion is modified for AdS compactifications). The remaining perturbations can be divided into massless fields, corresponding to zero modes of the linearized equations of motion on $K$, and massive fields, the others. General results on eigenvalues of Laplacians imply that the masses of massive fields depend on the diameter of $K$ as $m \sim 1/\mathrm{diam}(K)$, so at energies far smaller than $m$, they cannot be excited (this is not universal; given strong negative curvature on $K$, or a rapidly varying warp factor, one can have perturbations of small nonzero mass). Thus, the massive fields can be "integrated out," to leave an EFT with a finite number of fields. In the classical approximation, this simply means solving their equations of motion in terms of the massless fields, and using these solutions to eliminate them from the action. At leading order in an expansion around a solution, these fields are zero and this step is trivial; nevertheless, it is useful in making a systematic definition of the interaction terms in the EFT.

As we saw in pure GR, the configuration space parametrized by the massless fields in the EFT, is the moduli space of compactifications obtained by deforming the original solution. Thus, from a mathematical point of view, low-energy EFT can be thought of as a sort of enhancement of the concept of moduli space, and a dictionary set up between mathematical and physical languages. To give its next entry, there is a natural physical metric on moduli space, defined by restriction from the metric on the configuration space of the theory $\mathcal{T}$; this becomes the sigma-model metric for the scalars in the EFT. Because the theories $\mathcal{T}$ arising from string theory are geometrically natural, this metric is also natural from a mathematical point of view, and one often finds that much is already known about it. For example, the somewhat fearsome two derivative terms in eqn [1], are (perhaps) less so when one realizes that this is an explicit expression for the Weil–Petersson metric on the moduli space of Ricci-flat metrics. In any case, knowing this dictionary is essential for taking advantage of the literature.

Another important entry in this dictionary is that the automorphism group of a solution translates into the gauge group in the EFT. This can be either continuous, leading to the gauge symmetry of Maxwell and Yang–Mills theories, or discrete, leading to discrete gauge symmetry. For example, if the metric on $K$ has continuous isometry group $G$, the resulting EFT will have gauge symmetry $G$, as in the original example of Kaluza and Klein with $K \cong S^1$ and $G \cong \mathrm{U}(1)$. Mathematically, these phenomena of "enhanced symmetry" are often treated using the languages of equivariant theories (cohomology, K-theory, etc.), stacks, and so on.

To give another example, obstructed deformations (solutions of the linearized equations which do not correspond to elements of the tangent space of the true moduli space) correspond to scalar fields which, while massless, appear in the effective potential in a way which prevents giving them expectation values. Since the quadratic terms $V''$ are masses, this dependence must be at cubic or higher order.

While the preceding concepts are general and apply to compactification of all local field theories, string and M-theory add some particular ingredients to this general recipe. In the limits of small curvatures and weak coupling, string and M-theory are well described by the ten- and 11-dimensional supergravity theories, and thus the string/M-theory discussion usually starts with Kaluza–Klein compactification of these theories, which we denote I, IIa, IIb, HE, HO and M. Let us now discuss a particular example.

## Calabi–Yau Compactification of the Heterotic String

Contact with the SM requires finding compactifications to $d=4$ either without supersymmetry, or with at most $N=1$ supersymmetry, because the SM includes chiral fermions, which are incompatible with $N>1$. Let us start with the $E_8 \times E_8$ heterotic string or "HE" theory. This choice is made rather than HO because only in this case can we find the SM fermion representations as subrepresentations of the adjoint of the gauge group.

Besides the metric, the other bosonic fields of the HE supergravity theory are a scalar $\Phi$ called the dilaton, Yang–Mills gauge potentials for the group $G \equiv E_8 \times E_8$, and a 2-form gauge potential $B$ (often called the "Neveu–Schwarz" or "NS" 2-form) whose defining characteristic is that it minimally couples to the heterotic string world-sheet. We will need their gauge field strengths below: for Yang–Mills, this is a 2-form $F^a_{IJ}$ with $a$ indexing the adjoint of Lie $G$, and for the NS 2-form this is a 3-form $H_{IJK}$. Denoting the two Majorana–Weyl spinor representations of $SO(1,9)$ as $S$ and $C$, then the fermions are the gravitino $\psi_I \in S \otimes V$, a spin 1/2 "dilatino" $\lambda \in C$, and the adjoint gauginos $\chi^a \in S$. We use $\Gamma_I$ to denote Dirac matrices contracted with a "zehnbein," satisfying $\{\Gamma_I, \Gamma_J\} = 2G_{IJ}$, and $\Gamma_{IJ} = (1/2)[\Gamma_I, \Gamma_J]$, etc.

A local supersymmetry transformation with parameter $\epsilon$ is then

$$\delta\psi_I = D_I\epsilon + \tfrac{1}{8}H_{IJK}\Gamma^{JK}\epsilon \qquad [2]$$

$$\delta\lambda = \partial_I\Phi\Gamma^I\epsilon - \tfrac{1}{12}H_{IJK}\Gamma^{IJK}\epsilon \qquad [3]$$

$$\delta\chi^a = F^a_{IJ}\Gamma^{IJ}\epsilon \qquad [4]$$

We now assume $N=1$ supersymmetry. An unbroken supersymmetry is a spinor $\epsilon$ for which the left-hand side is zero, so we seek compactifications with a unique solution of these equations.

We first discuss the case $H=0$. Setting $\delta\psi_\mu$ in eqn [2] to zero, we find that the warp factor $f$ must be constant. The vanishing of $\delta\psi_i$ requires $\epsilon$ to be a covariantly constant spinor. For a six-dimensional $M$ to have a unique such spinor, it must have SU(3) holonomy; in other words, $M$ must be a Calabi–Yau manifold. In the following, we use basic facts about their geometry.

The vanishing of $\delta\lambda$ then requires constant dilaton $\Phi$, while the vanishing of $\delta\chi^a$ requires the gauge field strength $F$ to solve the hermitian Yang–Mills equations,

$$F^{2,0} = F^{0,2} = F^{1,1} = 0$$

By the theorem of Donaldson and Uhlenbeck–Yau, such solutions are in one-to-one correspondence with $\mu$-stable holomorphic vector bundles with structure group $H$ contained in the complexification of $G$. Choose such a bundle $E$; by the general discussion above, the commutant of $H$ in $G$ will be the automorphism group of the connection on $E$ and thus the low-energy gauge group of the resulting EFT. For example, since $E_8$ has a maximal $E_6 \times SU(3)$ subgroup, if $E$ has structure group $H = SL(3)$, there is an embedding such that the unbroken gauge symmetry is $E_6 \times E_8$, realizing one of the standard grand unified groups $E_6$ as a factor.

The choice of $E$ is constrained by anomaly cancellation. This discussion (Green *et al.* 1987) modifies the Bianchi identity for $H$ to

$$\mathrm{d}H = \mathrm{tr}\ R \wedge R - \frac{1}{30}\sum_a F^a \wedge F^a \qquad [5]$$

where $R$ is the matrix of curvature 2-forms. The normalization of the $F \wedge F$ term is such that if we take $E \cong TK$ the holomorphic tangent bundle of $K$, with isomorphic connection, then using the embedding we just discussed, we obtain a solution of eqn [5] with $H=0$.

Thus, we have a complete solution of the equations of motion. General arguments imply that supersymmetric Minkowski solutions are stable, so the small fluctuations consist of massless and massive fields. Let us now discuss a few of the massless fields. Since the EFT has $N=1$ supersymmetry, the massless scalars live in chiral multiplets, which are local coordinates on a complex Kähler manifold.

First, the moduli of Ricci-flat metrics on $K$ will lead to massless scalar fields: the complex structure

moduli, which are naturally complex, and Kähler moduli, which are not. However, in string compactification the latter are complexified to the periods of the 2-form $B + iJ$ integrated over a basis of $H_2(K, \mathbb{Z})$, where $J$ is the Kähler form and $B$ is the NS 2-form. In addition, there is a complex field pairing the dilaton (actually, $\exp(-\Phi)$) and the "model-independent axion," the scalar dual in $d=4$ to the 2-form $B_{\mu\nu}$. Finally, each complex modulus of the holomorphic bundle $E$ will lead to a chiral multiplet. Thus, the total number of massless uncharged chiral multiplets is $1 + h^{1,1}(K) + h^{2,1}(K) + \dim H^1(K, \text{End }(E))$.

Massless charged matter will arise from zero modes of the gauge field and its supersymmetric partner spinor $\chi^a$. It is slightly easier to discuss the spinor, and then appeal to supersymmetry to get the bosons. Decomposing the spinors of SO(6) under SU(3), one obtains $(0, p)$ forms, and the Dirac equation becomes the condition that these forms are harmonic. By the Hodge theorem, these are in one-to-one correspondence with classes in Dolbeault cohomology $H^{0,p}(K, V)$, for some bundle $V$. The bundle $V$ is obtained by decomposing the spinor into representations of the holonomy group of $E$. For $H = \text{SU}(3)$, the decomposition of the adjoint under the embedding of $\text{SU}(3) \times E_6$ in $E_8$,

$$248 = (8,1) + (1,78) + (3,27) + (\bar{3},\bar{27}) \qquad [6]$$

implies that charged matter will form "generations" in the 27, of number $\dim H^{0,1}(K, E)$, and "antigenerations" in the $\bar{27}$, of number $\dim H^{0,1}(K, \bar{E}) = \dim H^{0,2}(K, E)$. The difference in these numbers is determined by the Atiyah–Singer index theorem to be

$$N_{\text{gen}} \equiv N_{27} - N_{\bar{27}} = \tfrac{1}{2} c_3(E)$$

In the special case of $E \cong TK$, these numbers are separately determined to be $N_{27} = h^{1,1}$ and $N_{\bar{27}} = h^{2,1}$, so their difference is $\chi(K)/2$, half the Euler number of $K$. In the real world, this number is $N_{\text{gen}} = 3$, and matching this under our assumptions so far is very constraining.

Substituting these zero modes into the ten-dimensional Yang–Mills action and integrating, one can derive the $d=4$ EFT. For example, the cubic terms in the superpotential, usually called Yukawa couplings after the corresponding fermion–boson interactions in the component Lagrangian, are obtained from the cubic product of zero modes

$$\int_K \Omega \wedge \text{tr}(\phi_1 \wedge \phi_2 \wedge \phi_3)$$

where $\Omega$ is the holomorphic $\phi_i \in H^{0,1}(K, \text{Rep } E)$ are the zero modes, and tr arises from decomposing the $E_8$ cubic group invariant.

Note the very important fact that this expression only depends on the cohomology classes of the $\phi_i$ (and $\Omega$). This means the Yukawa couplings can be computed without finding the explicit harmonic representatives, which is not possible (we do not even know the explicit metric). More generally, one expects to be able to explicitly compute the superpotential and all other holomorphic quantities in the effective Lagrangian solely from "topological" information (the Dolbeault cohomology ring, and its generalizations within topological string theory). On the other hand, computing the Kähler metric in an $N=1$ EFT is usually out of reach as it would require having explicit normalized zero modes. Most results for this metric come from considering closely related compactifications with extended supersymmetry, and arguing that the breaking to $N=1$ supersymmetry makes small corrections to this.

There are several generalizations of this construction. First, the necessary condition to solve eqn [5] is that the left-hand side be exact, which requires

$$c_2(E) = c_2(TK) \qquad [7]$$

This allows for a wide variety of $E$'s to be used, so that $N_{\text{gen}} = 3$ can be attained with many more $K$'s. This class of models is often called "(0, 2) compactification" to denote the world-sheet supersymmetry of the heterotic string in these backgrounds. One can also use bundles with larger structure group; for example, $H = SL(4)$ leads to unbroken $\text{SO}(10) \times E_8$, and $H = \text{SL}(5)$ leads to unbroken $\text{SU}(5) \times E_8$.

The subsequent breaking of the grand unified group to the SM gauge group is typically done by choosing $K$ with nontrivial $\pi_1$, so that it admits a flat line bundle $W$ with nontrivial holonomy (usually called a "Wilson line"). One then uses the bundle $E \otimes W$ in the above discussion, to obtain the commutant of $H \otimes W$ as gauge group. For example, if $\pi_1(K) \cong \mathbb{Z}_5$, one can use $W$ whose holonomy is an element of order 5 in SU(5), to obtain as commutant the SM gauge group $\text{SU}(3) \times \text{SU}(2) \times \text{U}(1)$.

Another generalization is to take the 3-form $H \neq 0$. This discussion begins by noting that, for supersymmetry, we still require the existence of a unique spinor $\epsilon$; however, it will no longer be covariantly constant in the Levi-Civita connection. One way to structure the problem is to note that the right-hand side of eqn [2] takes the form of a connection with torsion; the resulting equations have been discussed mathematically in (Li and Yau 2004).

Another recent approach to these compactifications (Gauntlett 2004) starts out by arguing that $\epsilon$ cannot vanish on $K$, so it defines a weak SU(3) structure, a local reduction of the structure group of

$T\,K$ to SU(3) which need not be integrable. This structure must be present in all $N=1, d=4$ supersymmetric compactifications and there are hopes that it will lead to a useful classification of the possible local structures and corresponding partial differential equations (PDEs) on $K$.

## Higher-Dimensional and Extended Supersymmetric Compactifications

While there are similar quasirealistic constructions which start from the other string theories and M-theory, before we discuss these, let us give an overview of compactifications with $N \geq 2$ supersymmetry in four dimensions, and in higher dimensions. These are simpler analog models which can be understood in more depth; their study led to one of the most important discoveries in string/M-theory, the theory of superstring duality.

As before, we require a covariantly constant spinor. For Ricci-flat $K$ with other background fields zero, this requires the holonomy of $K$ to be one of trivial, SU($n$), Sp($n$), or the exceptional holonomies $G_2$ or Spin(7). In Table 1 we tabulate the possibilities with spacetime dimension $d$ greater or equal to 3, listing the supergravity theory, the holonomy type of $K$, and the type of the resulting EFT: dimension $d$, total number of real supersymmetry parameters $Ns$, and the number of spinor supercharges $N$ (in $d=6$, since left- and right-chirality Majorana spinors are inequivalent, there are two numbers).

The structure of the resulting supergravity EFTs is heavily constrained by $Ns$. We now discuss the various possibilities.

**Table 1** String/M-theories, holonomy groups and the resulting supersymmetry

| *Theory* | *Holonomy* | *d* | *Ns* | *N* |
|---|---|---|---|---|
| M, II | *Torus* | Any | 32 | Max |
| M | SU(2) | 7 | 16 | 1 |
| | SU(3) | 5 | 8 | 1 |
| | $G_2$ | 4 | 4 | 1 |
| | Sp(4) | 3 | 6 | 3 |
| | SU(4) | 3 | 4 | 2 |
| | Spin(7) | 3 | 2 | 1 |
| IIa | SU(2) | 6 | 16 | (1, 1) |
| | SU(3) | 4 | 8 | 2 |
| | $G_2$ | 3 | 4 | 2 |
| IIb | SU(2) | 6 | 16 | (0, 2) |
| | SU(3) | 4 | 8 | 2 |
| | $G_2$ | 3 | 4 | 2 |
| HE, HO, I | *Torus* | Any | 16 | Max/2 |
| | SU(2) | 6 | 8 | 1 |
| | SU(3) | 4 | 4 | 1 |
| | $G_2$ | 3 | 2 | 1 |

### $Ns = 32$

Given the supersymmetry algebra, if such a supergravity exists, it is unique. Thus, toroidal compactifications of $d=11$ supergravity, IIa and IIb supergravity lead to the same series of maximally supersymmetric theories. Their structure is governed by the exceptional Lie algebra $E_{11-d}$; the gauge charges transform in a fundamental representation of this algebra, while the scalar fields parametrize a coset space $G/H$, where $G$ is the maximally split real form of the Lie group $E_{11-d}$, and $H$ is a maximal compact subgroup of $G$. Nonperturbative duality symmetries lead to a further identification by a maximal discrete subgroup of $G$.

### $Ns = 16$

This supergravity can be coupled to maximally supersymmetric super Yang–Mills theory, which given a choice of gauge group $G$ is unique. Thus, these theories (with zero cosmological constant and thus allowing super-Poincaré symmetry) are uniquely determined by the choice of $G$.

In $d=10$, the choices $E_8 \times E_8$ and $\mathrm{Spin}(32)/\mathbb{Z}_2$ which arise in string theory, are almost uniquely determined by the Green–Schwarz anomaly cancellation analysis. Compactification of these HE, HO and type I theories on $T^n$ produces a unique theory with moduli space

$$\mathbb{R}^+ \times \mathrm{SO}(n, n+16; \mathbb{Z})\backslash \mathrm{SO}(n, n+16; \mathbb{R})/\mathrm{SO}(n, \mathbb{R}) \times \mathrm{SO}(n+16, \mathbb{R}) \qquad [8]$$

In Kaluza–Klein (KK) reduction, this arises from the choice of metric $g_{ij}$, the antisymmetric tensor $B_{ij}$ and the choice of a flat $E_8 \times E_8$ or $\mathrm{Spin}(32)/\mathbb{Z}_2$ connection on $T^n$, while a more unified description follows from the heterotic string world-sheet analysis. Here the group $\mathrm{SO}(n, n+16)$ is defined to preserve an even self-dual quadratic form $\eta$ of signature $(n, n+16)$; for example, $\eta = (-E_8) \oplus (-E_8) \oplus I \oplus I \oplus I$, where $I$ is the form of signature (1,1) and $E_8$ is the Cartan matrix. In fact, all such forms are equivalent under orthogonal integer similarity transformation; so, the resulting EFT is unique. It has a rank $16+2n$ gauge group, which at generic points in moduli space is $\mathrm{U}(1)^{16+2n}$, but is enhanced to a nonabelian group $G$ at special points. To describe $G$, we first note that a point $p$ in moduli space determines an $n$-dimensional subspace $V_p$ of $\mathbb{R}^{16+2n}$, and an orthogonal subspace $V_p^\perp$ (of varying dimension). Lattice points of length squared –2 contained in $V_p^\perp$ then correspond to roots of the Lie algebra of $G_p$.

The other compactifications with $Ns=16$ is M-theory on K3 and its further toroidal reductions, and IIb on K3. M-theory compactification to $d=7$ is dual to heterotic on $T^3$, with the same moduli space and enhanced gauge symmetry. As we discuss at the end of the section "Stringy and quantum corrections," the extra massless gauge bosons of enhanced gauge symmetry are M2 branes wrapped on 2-cycles with topology $S^2$. For such a cycle to have zero volume, the integral of the Kähler form and holomorphic 2-form over the cycle must vanish; expressing this in a basis for $H^2(K3, \mathbb{R})$ leads to exactly the same condition we discussed for enhanced gauge symmetry above. The final result is that all such K3 degenerations lead to one- of the two-dimensional canonical singularities, of types A, D or E, and the corresponding EFT phenomenon is the enhanced gauge symmetry of corresponding Dynkin type A, D, or E.

IIb on K3 is similar, but reducing the self-dual Ramond–Ramond (RR) 4-form potential on the 2-cycles leads to self-dual tensor multiplets instead of Maxwell theory. The moduli space is eqn [8] but with $n=5$, not $n=4$, incorporating periods of RR potentials and the $SL(2, \mathbb{Z})$ duality symmetry of IIb theory.

One may ask if the $Ns=16$ I/HE/HO theories in $d=8$ and $d=9$ have similar duals. For $d=8$, these are obtained by a pretty construction known as "F-theory." Geometrically, the simplest definition of F-theory is to consider the special case of M-theory on an elliptically fibered Calabi–Yau, in the limit that the Kähler modulus of the fiber becomes small. One check of this claim for $d=8$ is that the moduli space of elliptically fibered K3s agrees with eqn [8] with $n=2$.

Another definition of F-theory is the particular case of IIb compactification using Dirichlet 7-branes, and orientifold 7-planes. This construction is $T$-dual to the type I theory on $T^2$, which provides its simplest string theory definition. As discussed in Polchinski (1999), one can think of the open strings giving rise to type I gauge symmetry as living on 32 Dirichlet 9-branes (or D9-branes) and an orientifold nineplane. $T$-duality converts Dirichlet and orientifold $p$-branes to $(p-1)$-branes; thus this relation follows by applying two $T$-dualities.

These compactifications can also be parametrized by elliptically fibered Calabi–Yaus, where $K$ is the base, and the branes correspond to singularities of the fibration. The relation between these two definitions follows fairly simply from the duality between M-theory on $T^2$, and IIb string on $S^1$. There is a partially understood generalization of this to $d=9$.

Finally, these constructions admit further discrete choices, which break some of the gauge symmetry. The simplest to explain is in the toroidal compactification of I/HE/HO. The moduli space of theories we discussed uses flat connections on the torus which are continuously connected to the trivial connection, but in general the moduli space of flat connections has other components. The simplest example is the moduli space of flat $E_8 \times E_8$ connections on $S^1$, which has a second component in which the holonomy exchanges the two $E_8$'s. On $T^3$, there are connections for which the holonomies cannot be simultaneously diagonalized. This structure and the M-theory dual of these choices is discussed in (de Boer *et al*. 2001).

### $Ns=8$, $d<6$

Again, the gravity multiplet is uniquely determined, so the most basic classification is by the gauge group $G$. The full low-energy EFT is determined by the matter content and action, and there are two types of matter multiplets. First, vector multiplets contain the Yang–Mills fields, fermions and $6-d$ scalars; their action is determined by a prepotential which is a $G$-invariant function of the fields. Since the vector multiplets contain massless adjoint scalars, a generic vacuum in which these take nonzero distinct vacuum expectation values (VEVs) will have $U(1)^r$ gauge symmetry, the commutant of $G$ with a generic matrix (for $d<5$, while there are several real scalars, the potential forces these to commute in a supersymmetric vacuum). Vacua with this type of gauge symmetry breaking, which does not reduce the rank of the gauge group, are usually referred to as on a "Coulomb branch" of the moduli space. To summarize, this sector can be specified by $n_V$, the number of vector multiplets, and the prepotential $\mathcal{F}$, a function of the $n_V$ VEVs which is cubic in $d=5$, and holomorphic in $d=4$.

Hypermultiplets contain scalars which parametrize a quaternionic Kähler manifold, and partner fermions. Thus, this sector is specified by a $4n_H$ real dimensional quaternionic Kähler manifold. The $G$ action comes with triholomorphic moment maps; if nontrivial, VEVs in this sector can break gauge symmetry and reduce it in rank. Such vacua are usually referred to as on a "Higgs branch."

The basic example of these compactifications is M-theory on a Calabi–Yau 3-fold ($CY_3$). Reduction of the 3-form leads to $h^{1,1}(K)$ vector multiplets, whose scalar components are the CY Kähler moduli. The CY complex structure moduli pair with periods of the 3-form to produce $h^{2,1}(K)$ hypermultiplets. Enhanced gauge symmetry then appears when the

$CY_3$ contains ADE singularities fibered over a curve, from the same mechanism involving wrapped M2 branes we discussed under $Ns = 16$. If degenerating curves lead to other singularities (e.g., the ODP or "conifold"), it is possible to obtain extremal transitions which translate physically into Coulomb–Higgs transitions. Finally, singularities in which surfaces degenerate lead to nontrivial fixed-point theories.

Reduction on $S^1$ leads to IIa on $CY_3$, with the spectrum above plus a "universal hypermultiplet" which includes the dilaton. Perhaps the most interesting new feature is the presence of world-sheet instantons, which correct the metric on vector multiplet moduli space. This metric satisfies the restrictions of special geometry and thus can be derived from a prepotential.

The same theory can be obtained by compactification of IIb theory on the mirror $CY_3$. Now vector multiplets are related to the complex structure moduli space, while hypermultiplets are related to Kähler moduli space. In this case, the prepotential derived from variation of complex structure receives no instanton corrections, as we discuss in the next section.

Finally, one can compactify the heterotic string on $K3 \times T^{6-d}$, but this theory follows from toroidal reduction of the $d = 6$ case we discuss next.

#### *Ns* = 8, *d* = 6

These supergravities are similar to $d < 6$, but there is a new type of matter multiplet, the self-dual tensor (in $d < 6$ this is dual to a vector multiplet). Since fermions in $d = 6$ are chiral, there is an anomaly cancellation condition relating the numbers of the three types of multiplets (Aspinwall 1996, section 6.6),

$$n_{\mathrm{H}} - n_{\mathrm{V}} + 29 n_{\mathrm{T}} = 273 \qquad [9]$$

One class of examples is the heterotic string compactified on K3. In the original perturbative constructions, to satisfy eqn [7], we need to choose a vector bundle with $c_2(V) = \chi(K3) = 24$. The resulting degrees of freedom are a single self-dual tensor multiplet and a rank-16 gauge group. More generally, one can introduce $N_{5B}$ heterotic 5-branes, which generalize eqn [7] to $c_2(E) + N_{5B} = c_2(TK)$. Since this brane carries a self-dual tensor multiplet, this series of models is parametrized by $n_{\mathrm{T}}$. They are connected by transitions in which an $E_8$ instanton shrinks to zero size and becomes a 5-brane; the resulting decrease in the dimension of the moduli space of $E_8$ bundles on K3 agrees with eqn [9].

Another class of examples is F-theory on an elliptically fibered $CY_3$. These are related to M-theory on an elliptically fibered $CY_3$ in the same general way we discussed under $Ns = 16$. The relation between F-theory and the heterotic string on K3 can be seen by lifting M-theory-heterotic duality; this suggests that the two constructions are dual only if the $CY_3$ is a K3 fibration as well. Since not all elliptically fibered $CY_3$s are K3 fibered, the F-theory construction is more general.

We return to $d = 4$ and $Ns = 4$ in the final section. The cases of $Ns < 4$ which exist in $d \leq 3$ are far less studied.

## Stringy and Quantum Corrections

The $D$-dimensional low-energy effective supergravity actions on which we based our discussion so far are only approximations to the general story of string/M-theory compactification. However, if Planck's constant is small, $K$ is sufficiently large, and its curvature is small, then they are controlled approximations.

In M-theory, as in any theory of quantum gravity, corrections are controlled by the Planck scale parameter $M_{\mathrm{P}}^{D-2}$, which sits in front of the Einstein term of the $D$-dimensional effective Lagrangian, and plays the role of $\hbar$. In general, this is different from the four-dimensional Planck scale, which satisfies $M_{\mathrm{P}4}^2 = \mathrm{Vol}(K) M_{\mathrm{P}}^{D-2}$. After taking the low-energy limit $E \ll M_{\mathrm{P}}$, the remaining corrections are controlled by the dimensionless parameters $l_{\mathrm{P}}/R$, where $R$ can any characteristic length scale of the solution: a curvature radius, the length of a nontrivial cycle, and so on.

In string theory, one usually thinks of the corrections as a double series expansion in $g_{\mathrm{s}}$, the dimensionless (closed) string coupling constant, and $\alpha'$, the inverse string tension parameter, of dimensions $(\text{length})^2$. The ten-dimensional Planck scale is related to these parameters as $M_{\mathrm{P}}^8 = 1/g_{\mathrm{s}}^2 (\alpha')^4$, up to a constant factor that depends on conventions.

Besides perturbative corrections, which have power-like dependence on these parameters, there can be world sheet and "brane" instanton corrections. For example, a string world sheet can wrap around a topologically nontrivial spacelike 2-cycle $\Sigma$ in $K$, leading to an instanton correction to the effective action which is suppressed as $\exp(-\mathrm{Vol}(\Sigma)/2\pi\alpha')$. More generally, any $p$-brane wrapping a $p$-cycle can produce a similar effect. As for which terms in the effective Lagrangian receive corrections, this depends largely on the number and symmetries of the fermion zero modes on the instanton world volumes.

Let us start by discussing some cases in which one can argue that these corrections are not present.

First, extended supersymmetry can serve to eliminate many corrections. This is analogous to the familiar result that the superpotential in $d = 4, N = 1$ supersymmetric field theory does not receive (or "is protected from") perturbative corrections, and in many cases follows from similar formal arguments. In particular, supersymmetry forbids corrections to the potential and two derivative terms in the $Ns = 32$ and $Ns = 16$ Lagrangians.

In $Ns = 8$, the superpotential is protected, but the two derivative terms can receive corrections. However, there is a simple argument which precludes many such corrections – since vector multiplet and hypermultiplet moduli spaces are decoupled, a correction whose control parameter sits in (say) a vector multiplet, cannot affect hypermultiplet moduli space. This fact allows for many exact computations in these theories.

As an example, in IIb on $CY_3$, the metric on vector multiplet moduli space is precisely eqn [1] as obtained from supergravity (in other words, the Weil–Petersson metric on complex structure moduli space). First, while in principle it could have been corrected by world-sheet instantons, since these depend on Kähler moduli which sit in hypermultiplets, it is not. The only other instantons with the requisite zero modes to modify this metric are wrapped Dirichlet branes. Since in IIb theory these wrap even-dimensional cycles, they also depend on Kähler moduli and thus leave vector moduli space unaffected.

As previously discussed, for K3-fibered $CY_3$, this theory is dual to the heterotic string on $K3 \times T^2$. There, the vector multiplets arise from Wilson lines on $T^2$, and reduce to an adjoint multiplet of $N = 2$ supersymmetric Yang–Mills theory. Of course, in the quantum theory, the metric on this moduli space receives instanton corrections. Thus, the duality allows deriving the exact moduli space metric, and many other results of the Seiberg–Witten theory of $N = 2$ gauge theory, as aspects of the geometry of Calabi–Yau moduli space.

In $Ns = 4$, only the superpotential is protected, and that only in perturbation theory; it can receive nonperturbative corrections. Indeed, it appears that this is fairly generic, suggesting that the effective potentials in these theories are often sufficiently complicated to exhibit the structure required for supersymmetry breaking and the other symmetry breakings of the SM. Understanding this is an active subject of research.

We now turn from corrections to novel physical phenomena which arise in these regimes. While this is too large a subject to survey here, one of the basic principles which governs this subject is the idea that string/M-theory compactification on a singular manifold $K$ is typically consistent, but has new light degrees of freedom in the EFT, not predicted by KK arguments. We implicitly touched on one example of this in the discussion of M-theory compactification on K3 above, as the space of Ricci-flat K3 metrics has degeneration limits in which curvatures grow without bound, while the volumes of 2-cycles vanish. On the other hand, the structure of $Ns = 16$ supersymmetry essentially forces the $d = 7$ EFT in these limits to be nonsingular. Its only noteworthy feature is that a nonabelian gauge symmetry is restored, and thus certain charged vector bosons and their superpartners become massless.

To see what is happening microscopically, we must consider an M-theory membrane (or 2-brane), wrapped on a degenerating 2-cycle. This appears as a particle in $d = 7$, charged under the vector potential obtained by reduction of the $D = 11$ 3-form potential. The mass of this particle is the volume of the 2-cycle multiplied by the membrane tension, so as this volume shrinks to zero, the particle becomes massless. Thus, the physics is also well defined in 11 dimensions, though not literally described by 11-dimensional supergravity.

This phenomenon has numerous generalizations. Their common point is that, since the essential physics involves new light degrees of freedom, they can be understood in terms of a lower-dimensional quantum theory associated with the region around the singularity. Depending on the geometry of the singularity, this is sometimes a weakly coupled field theory, and sometimes a nontrivial conformal field theory. Occasionally, as in IIb on K3, the lightest wrapped brane is a string, leading to a "little string theory" (Aharony 2000).

## $N = 1$ Supersymmetry in Four Dimensions

Having described the general framework, we conclude by discussing the various constructions which lead to $N = 1$ supersymmetry. Besides the heterotic string on a $CY_3$, these compactifications include type IIa and IIb on orientifolds of $CY_3$, the related F-theory on elliptically fibered Calabi–Yau 4-folds ($CY_4$), and M-theory on $G_2$ manifolds. Let us briefly spell out their ingredients, the known nonperturbative corrections to the superpotential, and the duality relations between these constructions.

To start, we recap the heterotic string construction. We must specify a $CY_3 K$, and a bundle $E$ over $K$ which admits a Hermitian Yang–Mills connection. The gauge group $G$ is the commutant of the structure group of $E$ in $E_8 \times E_8$ or $\mathrm{Spin}(32)/\mathbb{Z}_2$,

while the chiral matter consists of metric moduli of $K$, and fields corresponding to a basis for the Dolbeault cohomology group $H^{0,1}(K, \mathrm{Rep}\, E)$ where Rep $E$ is the bundle $E$ embedded into an $E_8$ bundle and decomposed into $G$-reps.

There is a general (though somewhat formal) expression for the superpotential,

$$W = \int \Omega \wedge + \mathrm{tr}\left(\bar{A}\bar{\partial}\bar{A} + \tfrac{2}{3}\bar{A}^3\right) + \int \Omega \wedge H^{(3)} + W_{\mathrm{NP}} \qquad [10]$$

The first term is the holomorphic Chern–Simons action, whose variation enforces the $F^{0,2} = 0$ condition. The second is the "flux superpotential," while the third term is the nonperturbative corrections. The best understood of these arise from supersymmetric gauge theory sectors. In some, but not all, cases, these can be understood as arising from gauge theoretic instantons, which can be shown to be dual to heterotic 5-branes wrapped on $K$. Heterotic world-sheet instantons can also contribute.

The HO theory is $S$-dual to the type I string, with the same gauge group, realized by open strings on Dirichlet 9-branes. This construction involves essentially the same data. The two classes of heterotic instantons are dual to D1- and D5-brane instantons, whose world-sheet theories are somewhat simpler.

If the $\mathrm{CY}_3\, K$ has a fibration by tori, by applying $T$-duality to the fibers along the lines discussed for tori under $Ns = 16$ above, one obtains various type II orientifold compactifications. On an elliptic fibration, double $T$-duality produces a IIb compactification with D7s and O7s. Using the relation between IIb theory on $T^2$ and F-theory on K3 fiberwise, one can also think of this as an F-theory compactification on a K3-fibered $\mathrm{CY}_4$. More generally, one can compactify F theory on any elliptically fibered 4-fold to obtain $N = 1$. These theories have D3-instantons, the $T$-duals of both the type I D1- and D5-brane instantons.

The theory of mirror symmetry predicts that all $\mathrm{CY}_3$s have $T^3$ fibration structures. Applying the corresponding triple $T$-duality, one obtains a IIa compactification on the mirror $\mathrm{CY}_3\, \tilde{K}$, with D6-branes and O6-planes. Supersymmetry requires these to wrap special Lagrangian cycles in $\tilde{K}$. As in all Dirichlet brane constructions, enhanced gauge symmetry arises from coincident branes wrapping the same cycle, and only the classical groups are visible in perturbation theory. Exceptional gauge symmetry arises as a strong coupling phenomenon of the sort described in the previous section. The superpotential can also be thought of as mirror to eqn [10], but now the first term is the sum of a real Chern–Simons action on the special Lagrangian cycles, with disk world-sheet instanton corrections, as studied in open string mirror symmetry. The gauge theory instantons are now D2-branes.

Using the duality relation between the IIa string and 11-dimensional M-theory, this construction can be lifted to a compactification of M-theory on a seven-dimensional manifold $L$, which is an $S^1$ fibration over $K$. The D6 and O6 planes arise from singularities in the $S^1$ fibration. Generically, $L$ can be smooth, and the only candidate in Table 1 for such an $N = 1$ compactification is a manifold with $G_2$ holonomy; therefore, $L$ must have such holonomy. Finally, both the IIa world-sheet instantons and the D2-brane instantons lift to membrane instantons in M-theory.

This construction implicitly demonstrates the existence of a large number of $G_2$ holonomy manifolds. Another way to arrive at these is to go back to the heterotic string on $K$, and apply the duality (again under $Ns = 16$) between heterotic on $T^3$ and M-theory on K3 to the $T^3$ fibration structure on $K$, to arrive at M-theory on a K3-fibered manifold of $G_2$ holonomy. Wrapping membranes on 2-cycles in these fibers, we can see enhanced gauge symmetry in this picture fairly directly. It is an illuminating exercise to work through its dual realizations in all of these constructions.

Our final construction uses the interpretation of the strong coupling limit of the HE theory as M-theory on a one-dimensional interval $I$, in which the two $E_8$ factors live on the two boundaries. Thus, our original starting point can also be interpreted as the heterotic string on $K \times I$. This construction is believed to be important physically as it allows generalizing a heterotic string tree-level relation between the gauge and gravitational couplings which is phenomenologically disfavored. One can relate it to a IIa orientifold as well, now with D8- and O8-branes.

These multiple relations are often referred to as the "web" of dualities. They lead to numerous relations between compactification manifolds, moduli spaces, superpotentials, and other properties of the EFTs, whose full power has only begun to be appreciated.

## Suggestions for further reading

Original references for all but the most recent of these topics can be found in the following textbooks and proceedings. We have also referenced a few research articles which are good starting points for the more recent literature. There are far more reviews than we could reference here, and a partial listing of these appears at http://www.slac.stanford.edu/spires/reviews/

*See also:* Brane Construction of Gauge Theories; Random Algebraic Geometry, Attractors and Flux Vacua;

String Theory: Phenomenology; Superstring Theories; Two-Dimensional Conformal Field Theory and Vertex Operator Algebras; Viscous Incompressible Fluids: Mathematical Theory.

## Further Reading


Aharony O (2000) A brief review of "little string theories." *Classical and Quantum Gravity* 17: 929–938.

Aspinwall PS (1996) K3 surfaces and string duality, 1996 preprint, arXiv:hep-th/9611137.

Bachas C *et al.* (eds.) (2002) *Les Houches 2001: Unity from Duality: Gravity, Gauge Theory and Strings*. Berlin: Springer.

de Boer J *et al.* (2002) Triples, fluxes, and strings. *Advances in Theoretical and Mathematical Physics* 4: 995.

Connes A and Gawędzki K (eds.) (1998) *Les Houches 1995: Quantum Symmetries*. Amsterdam: North-Holland.

Deligne P *et al.* (eds.) (1999) *Quantum Fields and Strings: A Course for Mathematicians*. Providence, RI: American Mathematical Society.

Douglas M *et al.* (eds.) (2004) *Strings and Geometry: Proceedings of the 2002 Clay School*. Providence, RI: American Mathematical Society.

Gauntlett J (2004) Branes, calibrations and supergravity. In: Douglas M *et al.* (eds.) *Strings and Geometry*, pp. 79–126. Providence, RI: American Mathematical Society.

Green MB, Schwarz JH, and Witten E (1987) *Superstring Theory*, 2 vols. Cambridge: Cambridge University Press.

Li J and Yau S-T (2004) The existence of supersymmetric string theory with torsion, 2004 preprint, arXiv:hep-th/0411136.

Polchinski J (1998) *String Theory*, 2 vols. Cambridge: Cambridge University Press.


# Compressible Flows: Mathematical Theory

**G-Q Chen**, Northwestern University, Evanston, IL, USA



## Introduction

The Euler equations for compressible fluids consist of the conservation laws of mass, momentum, and energy:

$$\partial_t \rho + \nabla_{\boldsymbol{x}} \cdot \boldsymbol{m} = 0, \quad \boldsymbol{x} \in \boldsymbol{R}^d \qquad [1]$$

$$\partial_t \boldsymbol{m} + \nabla_{\boldsymbol{x}} \cdot \left( \frac{\boldsymbol{m} \otimes \boldsymbol{m}}{\rho} \right) + \nabla_{\boldsymbol{x}} p = 0 \qquad [2]$$

$$\partial_t E + \nabla_{\boldsymbol{x}} \cdot \left( \frac{\boldsymbol{m}}{\rho}(E + p) \right) = 0 \qquad [3]$$

Equivalently, these correspond to the general form of nonlinear hyperbolic systems of conservation laws:

$$\partial_t \boldsymbol{u} + \nabla_{\boldsymbol{x}} \cdot \boldsymbol{f}(\boldsymbol{u}) = 0, \quad \boldsymbol{x} \in \boldsymbol{R}^d, \ \boldsymbol{u} \in \boldsymbol{R}^n \qquad [4]$$

System [1]–[3] is closed by the following constitutive relations:

$$p = p(\rho, e), \qquad E = \frac{1}{2}\frac{|\boldsymbol{m}|^2}{\rho} + \rho e \qquad [5]$$

In [1]–[3] and [5], $\tau = 1/\rho$ is the deformation gradient (specific volume for fluids, strain for solids), $\boldsymbol{v} = (v_1, \ldots, v_d)^\top$ is the fluid velocity with $\rho \boldsymbol{v} = \boldsymbol{m}$ the momentum vector, $p$ is the scalar pressure, and $E$ is the total energy with $e$ the internal energy which is a given function of $(\tau, p)$ or $(\rho, p)$ defined through thermodynamical relations. The other two thermodynamic variables are temperature $\theta$ and entropy $S$. If $(\rho, S)$ are chosen as independent variables, then the constitutive relations can be written as

$$(e, p, \theta) = (e(\rho, S), p(\rho, S), \theta(\rho, S)) \qquad [6]$$

governed by $\theta \, \mathrm{d}S = \mathrm{d}e + p\mathrm{d}\tau = \mathrm{d}e - p\mathrm{d}\rho/\rho^2$. For polytropic gases,

$$\begin{aligned} p &= p(\rho, S) = \kappa \rho^\gamma \mathrm{e}^{S/c_v} \\ e &= \frac{p}{(\gamma - 1)\rho} \\ \theta &= \frac{p}{R\rho} \end{aligned} \qquad [7]$$

where $R > 0$ may be taken to be the universal gas constant divided by the effective molecular weight of the particular gas, $c_v > 0$ is the specific heat at constant volume, $\gamma = 1 + R/c_v > 1$ is the adiabatic exponent, and $\kappa$ can be any positive constant under scaling.

The most important criterion of applicability of any mathematical model is its well-posedness: existence, uniqueness, and stability. The well-posedness theory for compressible fluid flows is far from being complete, and many further issues are still unexplored. In particular, the global existence and uniqueness of solutions in $\boldsymbol{R}^d$, $d \geq 2$, is still a major open problem, and only partial results shed some lights on the amazing complexity of the problem. Below, we will mainly focus on the well-posedness issues with emphasis on the Cauchy problem, the initial value problem:

$$\boldsymbol{u}|_{t=0} = \boldsymbol{u}_0 \qquad [8]$$

first for inviscid compressible fluid flows and then for viscous compressible fluid flows.

Throughout this article, where a cited reference is not shown in the "Further reading" section, it may usually be found by consulting Bressan (2000),

Chen (2005), Dafermos (2005), Feireisl (2004), Lions (1986, 1988) or Liv (2000).

## Inviscid Compressible Fluid Flows: Euler Equations

Solutions to the Euler equations [1]–[3] are generically discontinuous functions obeying the Clausius–Duhem inequality, the second law of thermodynamics:

$$\partial_t(\rho S) + \nabla_{\boldsymbol{x}} \cdot (\boldsymbol{m}S) \geq 0 \qquad [9]$$

in the sense of distributions. Such discontinuous solutions are called entropy solutions.

When a flow is isentropic, that is, entropy $S$ is a uniform constant $S_0$ in the flow, then the Euler equations for the flow take the simpler form:

$$\begin{aligned} \partial_t\rho + \nabla_{\boldsymbol{x}} \cdot \boldsymbol{m} = 0 \\ \partial_t\boldsymbol{m} + \nabla_{\boldsymbol{x}} \cdot (\boldsymbol{m} \otimes \boldsymbol{m}/\rho) + \nabla_{\boldsymbol{x}}p = 0 \end{aligned} \qquad [10]$$

where the pressure is a function of the density, $p = p(\rho, S_0)$, with constant $S_0$. For a polytropic gas,

$$p(\rho) = \kappa\rho^\gamma, \quad \gamma > 1 \qquad [11]$$

where $\kappa$ can be any positive constant by scaling. This system can be derived from [1] to [3] as follows: for smooth solutions of [1]–[3], entropy $S(\rho, \boldsymbol{m}, E)$ is conserved along fluid particle trajectories, that is,

$$\partial_t(\rho S) + \nabla_{\boldsymbol{x}} \cdot (\boldsymbol{m}S) = 0$$

If the entropy is initially a uniform constant and the solution remains smooth, then the energy equation can be eliminated and entropy $S$ keeps the same constant in later time. Thus, under constant initial entropy, a smooth solution of [1]–[3] satisfies the equations in [10]. Furthermore, solutions of system [10] are also a good approximation to solutions of system [1]–[3] even after shocks form, since the entropy increases across a shock to the third order in wave strength for solutions of [1]–[3], while in [10] the entropy is constant. Moreover, system [10] is an excellent model for the isothermal fluid flow with $\gamma = 1$ and for the shallow-water flow with $\gamma = 2$. For such barotropic flows (i.e., $p = p(\rho)$), the energy equation [3] serves as an entropy inequality (see Lax (1973)):

$$\partial_t E + \nabla_{\boldsymbol{x}} \cdot (\boldsymbol{m}(E + p(\rho))/\rho) \leq 0$$
in the sense of distributions

In the one-dimensional case, system [1]–[3] in Eulerian coordinates is

$$\begin{aligned} &\partial_t\rho + \partial_x m = 0, \qquad \partial_t m + \partial_x\left(m^2/\rho + p\right) = 0 \\ &\partial_t E + \partial_x(m(E + p)/\rho) = 0 \end{aligned} \qquad [12]$$

The system above can be rewritten in Lagrangian coordinates:

$$\begin{aligned} &\partial_t\tau - \partial_x v = 0, \qquad \partial_t v + \partial_x p = 0 \\ &\partial_t(e + v^2/2) + \partial_x(pv) = 0 \end{aligned} \qquad [13]$$

with $v = m/\rho$, where the coordinates $(t, x)$ are the Lagrangian coordinates, which are different from the Eulerian coordinates for [12]; for simplicity of notations, we do not distinguish them. For the barotropic case, systems [12] and [13] reduce to

$$\partial_t\rho + \partial_x m = 0, \qquad \partial_t m + \partial_x\left(m^2/\rho + p\right) = 0 \qquad [14]$$

and

$$\partial_t\tau - \partial_x v = 0, \qquad \partial_t v + \partial_x p = 0 \qquad [15]$$

respectively, where pressure $p = p(\rho) = \tilde{p}(\tau)$, $\tau = 1/\rho$. The solutions of [12] and [13], as well as [14] and [15], are equivalent even for entropy solutions with vacuum where $\rho = 0$.

The potential flow is well known in transonic aerodynamics, beyond the isentropic approximation [10] from [1] to [3]. Denote $D_t = \partial_t + \sum_{k=1}^d v_k\partial_{x_k}$ the convective derivative along fluid particle trajectories. From [1] to [3], we have

$$D_t S = 0 \qquad [16]$$

and, by taking the curl of the momentum equations,

$$D_t\left(\frac{\omega}{\rho}\right) = \frac{\omega}{\rho} \cdot \nabla_{\boldsymbol{x}}\boldsymbol{v} + \frac{p_S(\rho, S)}{\rho^3}\nabla_{\boldsymbol{x}}\rho \times \nabla_{\boldsymbol{x}}S \qquad [17]$$

The identities [16] and [17] imply that a smooth solution of [1]–[3] which is both isentropic and irrotational at time $t = 0$ remains isentropic and irrotational for all later times, as long as this solution stays smooth. Then, the conditions $S = S_0 = \text{const.}$ and $\omega = \text{curl}_{\boldsymbol{x}}\boldsymbol{v} = 0$ are reasonable for smooth solutions. For a smooth irrotational solution, we integrate the $d$-momentum equations in [10] through Bernoulli's law:

$$\partial_t\boldsymbol{v} + \nabla_{\boldsymbol{x}}(|\boldsymbol{v}|^2/2) + \nabla_{\boldsymbol{x}}h(\rho) = 0$$

where $h'(\rho) = p_\rho(\rho, S_0)/\rho$. On a simply connected space region, the condition $\text{curl}_{\boldsymbol{x}}\,\boldsymbol{v} = 0$ implies that there exists $\Phi$ such that $\boldsymbol{v} = \nabla_{\boldsymbol{x}}\Phi$. Then,

$$\begin{aligned} \partial_t\rho + \nabla_{\boldsymbol{x}} \cdot (\rho\nabla_{\boldsymbol{x}}\Phi) = 0 \\ \partial_t\Phi + \tfrac{1}{2}|\nabla_{\boldsymbol{x}}\Phi|^2 + h(\rho) = K \end{aligned} \qquad [18]$$

for some constant $K$. From the second equation in [18], we have

$$\rho(D\Phi) = h^{-1}(K - (\partial_t\Phi + \tfrac{1}{2}|\nabla_{\boldsymbol{x}}\Phi|^2))$$

Then, system [18] can be rewritten as the following time-dependent potential flow equation of second order:

$$\partial_t \rho(D\Phi) + \nabla_x \cdot (\rho(D\Phi)\nabla_x \Phi) = 0 \qquad [19]$$

For a steady solution $\Phi = \Phi(x)$, that is, $\partial_t \Phi = 0$, we obtain the celebrated steady potential flow equation of aerodynamics:

$$\nabla_x \cdot (\rho(\nabla_x \Phi)\nabla_x \Phi) = 0 \qquad [20]$$

In applications in aerodynamics, [18] or [19] is used for discontinuous solutions, and the empirical evidence is that entropy solutions of [18] or [19] are fairly good approximations to entropy solutions for [1]–[3] provided that (1) the shock strengths are small, (2) the curvature of shock fronts is not too large, and (3) there is a small amount of vorticity in the region of interest. Model [19] or [18] is an excellent model to capture multidimensional shock waves by ignoring vorticity waves, while the incompressible Euler equations are an excellent model to capture multidimensional vorticity waves by ignoring shock waves.

### Local Well-Posedness for Classical Solutions

Consider the Cauchy problem for the Euler equations [1]–[3] with Cauchy data [8]:

> Assume that $u_0 : \mathbf{R}^d \to \mathcal{D}$ is in $H^s \cap L^\infty$ with $s > d/2 + 1$. Then, for the Cauchy problem [1]–[3] and [8], there exists a finite time $T = T(\|u_0\|_s, \|u_0\|_{L^\infty}) \in (0, \infty)$ such that there is a unique, stable bounded classical solution $u \in C^1([0, T] \times \mathbf{R}^d)$ with $u(t, x) \in \mathcal{D}$ for $(t, x) \in [0, T] \times \mathbf{R}^d$ and $u \in C([0, T]; H^s) \cap C^1([0, T]; H^{s-1})$. Moreover, the interval $[0, T)$ with $T < \infty$ is the maximal interval of the classical $H^s$ existence for [1]–[3] if and only if either $\|(u_t, \nabla_x u)\|_{L^\infty} \to \infty$ or $u(t, x)$ escapes every compact subset $K \Subset \mathcal{D}$ as $t \to T$.

This local existence can be established by relying solely on the elementary linear existence theory for symmetric hyperbolic systems with smooth coefficients (cf. Majda (1984)), or by the abstract semigroup theory (Kato 1975).

### Formation of Singularities

For the one-dimensional case, singularities include the development of shock waves and formation of vacuum states. For the multidimensional case, the situation is much more complicated: besides shock waves and vacuum states, singularities can also be generated from vortex sheets, focusing and breaking of waves, among others.

Consider the Cauchy problem of the Euler equations [1]–[3] in $\mathbf{R}^3$ for polytropic gases with smooth initial data:

$$(\rho, v, S)|_{t=0} = (\rho_0, v_0, S_0)(x)$$
$$\rho_0(x) > 0, \ x \in \mathbf{R}^3 \qquad [21]$$

satisfying $(\rho_0, v_0, S_0)(x) = (\bar{\rho}, 0, \bar{S})$ for $|x| \ge L$, where $\bar{\rho} > 0$, $\bar{S}$, and $L$ are given constants. The equations possess a unique local $C^1$ solution $(\rho, v, S)(t, x)$ with $\rho(t, x) > 0$ provided that the initial data [21] is sufficiently regular. The support of the smooth disturbance $(\rho_0(x) - \bar{\rho}, v_0(x), S_0(x) - \bar{S})$ propagates with speed at most $\sigma = \sqrt{p_\rho(\bar{\rho}, \bar{S})}$ (the sound speed), that is,

$$(\rho, v, S)(t, x) = (\bar{\rho}, 0, \bar{S}) \quad \text{if } |x| \ge L + \sigma t \qquad [22]$$

Define

$$P(t) = \int_{\mathbf{R}^3} \left( p(\rho(t, x), S(t, x))^{1/\gamma} - p(\bar{\rho}, \bar{S})^{1/\gamma} \right) \mathrm{d}x$$
$$F(t) = \int_{\mathbf{R}^3} (\rho v)(t, x) \cdot x \, \mathrm{d}x$$

which, roughly speaking, measure the entropy and the radial component of momentum. Then, if $(\rho, v, S)(t, x)$ is a $C^1$ solution of [1]–[3] and [21] for $0 < t < T$, and

$$P(0) \ge 0, \qquad F(0) > \alpha\sigma R^4 \max_x \rho_0(x)$$
$$\text{with} \quad \alpha = 16\pi/3 \qquad [23]$$

then the lifespan $T$ of the $C^1$ solution is finite (Sideris 1985).

To illustrate a way in which the conditions in [23] may be satisfied, consider the initial data: $\rho_0 = \bar{\rho}$, $S_0 = \bar{S}$. Then $P(0) = 0$, and [23] holds if

$$\int_{|x|<R} v_0(x) \cdot x \, \mathrm{d}x > \alpha\sigma R^4$$

Comparing both sides, one finds that the initial velocity must be supersonic in some region relative to the sound speed at infinity. The formation of a singularity (presumably a shock wave) is detected as the disturbance overtakes the wave front forcing the front to propagate with supersonic speed.

Singularities are formed even without the condition of largeness, such as [23], being satisfied. For example, if $S_0(x) \ge \bar{S}$ and, for some $0 < R_0 < R$,

$$\int_{|x|>r} |x|^{-1}(|x| - r)^2(\rho_0(x) - \bar{\rho}) \, \mathrm{d}x > 0$$
$$\int_{|x|>r} |x|^{-3}(|x|^2 - r^2)\rho_0(x) v_0(x) \cdot x \, \mathrm{d}x \ge 0 \qquad [24]$$

for $R_0 < r < R$, then the lifespan $T$ of the $C^1$ solution of [1]–[3] and [21] is finite. The

assumptions in [24] mean that, in an average sense, the gas must be slightly compressed and outgoing directly behind the wave front.

### Local Well-Posedness for Shock-Front Solutions

For a general hyperbolic system of conservation laws [4], shock-front solutions are discontinuous, piecewise smooth entropy solutions with the following structure:

1. There exists a $C^2$ spacetime hypersurface $\mathcal{S}(t)$ defined in $(t, \boldsymbol{x})$ for $0 \le t \le T$ with spacetime normal $(\nu_t, \nu_{\boldsymbol{x}}) = (\nu_t, \nu_1, \ldots, \nu_d)$ as well as two $C^1$ vector-valued functions: $\boldsymbol{u}^+(t, \boldsymbol{x})$ and $\boldsymbol{u}^-(t, \boldsymbol{x})$, defined on respective domains $\mathcal{D}^+$ and $\mathcal{D}^-$ on either side of the hypersurface $\mathcal{S}(t)$ and satisfying $\partial_t \boldsymbol{u}^\pm + \nabla_{\boldsymbol{x}} \cdot \boldsymbol{f}(\boldsymbol{u}^\pm) = 0$ in $\mathcal{D}^\pm$;
2. The jump across the hypersurface $\mathcal{S}(t)$ satisfies the Rankine–Hugoniot condition:

$$\{\nu_t(\boldsymbol{u}^+ - \boldsymbol{u}^-) + \nu_{\boldsymbol{x}} \cdot (\boldsymbol{f}(\boldsymbol{u}^+) - \boldsymbol{f}(\boldsymbol{u}^-))\}|_{\mathcal{S}} = 0$$

For [4], the surface $\mathcal{S}$ is not known in advance and must be determined as part of the solution of the problem; thus, the two equations in (1)–(2) describe a multidimensional, highly nonlinear, free-boundary-value problem. The initial data yielding shock-front solutions is defined as follows. Let $\mathcal{S}_0$ be a smooth hypersurface parametrized by $\alpha$, and let $\nu(\alpha) = (\nu_1, \ldots, \nu_d)(\alpha)$ be a unit normal to $\mathcal{S}_0$. Define the piecewise smooth initial values for respective domains $\mathcal{D}_0^+$ and $\mathcal{D}_0^-$ on either side of the hypersurface $\mathcal{S}_0$ as

$$\boldsymbol{u}_0(\boldsymbol{x}) = \begin{cases} \boldsymbol{u}_0^+(\boldsymbol{x}), & \boldsymbol{x} \in \mathcal{D}_0^+ \\ \boldsymbol{u}_0^-(\boldsymbol{x}), & \boldsymbol{x} \in \mathcal{D}_0^- \end{cases} \qquad [25]$$

It is assumed that the initial jump in [25] satisfies the Rankine–Hugoniot condition, that is, there is a smooth scalar function $\sigma(\alpha)$ so that

$$\begin{aligned} &-\sigma(\alpha)\big(\boldsymbol{u}_0^+(\alpha) - \boldsymbol{u}_0^-(\alpha)\big) \\ &+ \nu(\alpha) \cdot \big(\boldsymbol{f}\big(\boldsymbol{u}_0^+(\alpha)\big) - \boldsymbol{f}\big(\boldsymbol{u}_0^-(\alpha)\big)\big) = 0 \end{aligned} \qquad [26]$$

and that $\sigma(\alpha)$ does not define a characteristic direction, that is,

$$\sigma(\alpha) \ne \lambda_i\big(\boldsymbol{u}_0^\pm\big), \quad \alpha \in \bar{\mathcal{S}}_0,\ 1 \le i \le n \qquad [27]$$

where $\lambda_i$, $i = 1, \ldots, n$, are the eigenvalues of [4]. It is natural to require that $\mathcal{S}(0) = \mathcal{S}_0$.

Consider the Euler equations [1]–[3] in $\boldsymbol{R}^3$ for polytropic gases with piecewise smooth initial data:

$$(\rho, \boldsymbol{v}, E)|_{t=0} = \begin{cases} \big(\rho_0^+, \boldsymbol{v}_0^+, E^+\big)(\boldsymbol{x}), & \boldsymbol{x} \in \mathcal{D}_0^+ \\ \big(\rho_0^-, \boldsymbol{v}_0^-, E^-\big)(\boldsymbol{x}), & \boldsymbol{x} \in \mathcal{D}_0^- \end{cases} \qquad [28]$$

Assume that $\mathcal{S}_0$ is a smooth compact surface in $\boldsymbol{R}^3$ and that $(\rho_0^+, \boldsymbol{v}_0^+, E_0^+)(\boldsymbol{x})$ belongs to the uniform local Sobolev space $H^s_{\mathrm{ul}}(\mathcal{D}_0^+)$, while $(\rho_0^-, \boldsymbol{v}_0^-, E_0^-)(\boldsymbol{x})$ belongs to the Sobolev space $H^s(\mathcal{D}_0^-)$, for some fixed $s \ge 10$. Assume also that there is a function $\sigma(\alpha) \in H^s(\mathcal{S}_0)$ so that [26] and [27] hold, and the compatibility conditions up to order $s - 1$ are satisfied on $\mathcal{S}_0$ by the initial data, together with the entropy condition:

$$\begin{aligned} &\boldsymbol{v}_0^+ \cdot \nu(\alpha) + \sqrt{p_\rho(\rho_0^+, S_0^+)} < \sigma(\alpha) \\ &< \boldsymbol{v}_0^- \cdot \nu(\alpha) + \sqrt{p_\rho(\rho_0^-, S_0^-)} \end{aligned} \qquad [29]$$

Then, there are a $C^2$ hypersurface $\mathcal{S}(t)$ and $C^1$ functions $(\rho^\pm, \boldsymbol{v}^\pm, E^\pm)(t, \boldsymbol{x})$ defined for $t \in [0, T]$, with $T$ sufficiently small, so that

$$(\rho, \boldsymbol{v}, E)(t, \boldsymbol{x}) = \begin{cases} (\rho^+, \boldsymbol{v}^+, E^+)(t, \boldsymbol{x}), & (t, \boldsymbol{x}) \in \mathcal{D}^+ \\ (\rho^-, \boldsymbol{v}^-, E^-)(t, \boldsymbol{x}), & (t, \boldsymbol{x}) \in \mathcal{D}^- \end{cases} \qquad [30]$$

is the discontinuous shock-front solution of the Cauchy problem [1]–[3] and [28]. Here a vector function $\boldsymbol{u}$ is in $H^s_{\mathrm{ul}}$, provided that there exists some $r > 0$ so that $\max_{\boldsymbol{y} \in \boldsymbol{R}^d} \|w_{r,\boldsymbol{y}} \boldsymbol{u}\|_{H^s} < \infty$ with $w_{r,\boldsymbol{y}}(\boldsymbol{x}) = w((\boldsymbol{x} - \boldsymbol{y})/r)$, where $w \in C_0^\infty(\boldsymbol{R}^d)$ is a function so that $w(\boldsymbol{x}) \ge 0$, $w(\boldsymbol{x}) = 1$ when $|\boldsymbol{x}| \le 1/2$, and $w(\boldsymbol{x}) = 0$ when $|\boldsymbol{x}| > 1$.

The compatibility conditions are needed in order to avoid the formation of discontinuities in higher derivatives along other characteristic surfaces emanating from $\mathcal{S}_0$: Once the main condition [26] is satisfied, the compatibility conditions are automatically guaranteed for a wide class of initial data. The idea of the proof is to use the existence of a strictly convex entropy and the symmetrization of [4]; the shock-front solutions are defined as the limit of a convergent classical iteration scheme based on a linearization by using the theory of linearized stability for shock fronts (Majda 1984). The uniform existence time of shock-front solutions in shock strength can be achieved (Métivier 1990).

### Global Theory in $L^\infty$ for the Isentropic Euler Equations for $x \in \boldsymbol{R}$

Consider the Cauchy problem for [14] with initial data:

$$(\rho, m)|_{t=0} = (\rho_0, m_0)(x) \qquad [31]$$

where $\rho_0$ and $m_0$ are in the physical region $\{(\rho, m) : \rho \ge 0, |m| \le C_0 \rho\}$ for some $C_0 > 0$. System [14] is strictly hyperbolic at the states with $\rho > 0$, and strict hyperbolicity fails at the vacuum states $V := \{(\rho, m/\rho) : \rho = 0, |m/\rho| < \infty\}$. Then, we have:

1. There exists a global solution $(\rho, m)(t, x)$ of the Cauchy problem [14] and [31] satisfying

$$0 \le \rho(t, x) \le C, \qquad |m(t, x)| \le C\rho(t, x) \qquad [32]$$

for some $C > 0$ depending only on $C_0$ and $\gamma$, and the entropy inequality

$$\partial_t \eta(\rho, m) + \partial_x q(\rho, m) \leq 0 \qquad [33]$$

in the sense of distributions for any convex weak entropy–entropy flux pair $(\eta, q)$, that is,

$$\nabla q(\rho, m) = \nabla \eta(\rho, m) \nabla f(\rho, m)$$

with

$$\nabla^2 \eta(\rho, m) \geq 0 \quad \text{and} \quad \eta|_V = 0$$

2. The solution operator $(\rho, m)(t, \cdot) = S_t(\rho_0, m_0)(\cdot)$, determined by (1), is compact in $L^1_{\text{loc}}(\boldsymbol{R})$ for $t > 0$;
3. Furthermore, if $(\rho_0, m_0)(x)$ is periodic with period $P$, then there exists a global periodic solution $(\rho, m)(t, x)$ with [32] such that $(\rho, m)(t, x)$ asymptotically decays to

$$\frac{1}{|P|} \int_P (\rho_0, m_0)(x) \mathrm{d}x$$

in $L^1$.

The convergence of the Lax–Friedrichs scheme, the Godunov scheme, and the vanishing viscosity method for system [14] have also been established.

The results are based on a compensated compactness framework to replace the BV compactness framework. For a gas obeying the $\gamma$-law, the case $\gamma = (N+2)/N$, $N \geq 5$ odd, was first studied by DiPerna (1983), and the case $1 < \gamma \leq 5/3$ for usual gases was first solved by Chen (1986) and Ding-Chen-Luo (1985). The cases $\gamma \geq 3$ and $5/3 < \gamma < 3$ were treated by Lions–Perthame–Tadmor (1994) and Lions–Perthame–Souganidis (1996), respectively. The case of general pressure laws was solved by Chen–LeFloch (2000, 2003). All the results for entropy solutions to [14] in Eulerian coordinates can equivalently be presented as the corresponding results for entropy solutions to [15] in Lagrangian coordinates. The isothermal case $\gamma = 1$ was treated by Huang–Wang (2002).

### Global Theory in BV for the Adiabatic Euler Equations for $x \in \boldsymbol{R}$

Consider the Euler equations [13] for polytropic gases with the Cauchy data:

$$(\tau, v, S)|_{t=0} = (\tau_0, v_0, S_0)(x) \qquad [34]$$

Then we have (Liu 1977, Temple 1981, Chen and Wagner 2003):

> Let $K \subset \{(\tau, v, S) : \tau > 0\}$ be a compact set in $\boldsymbol{R}_+ \times \boldsymbol{R}^2$, and let $N \geq 1$ be any constant. Then there exists a constant $C_0 = C_0(K, N)$, independent of $\gamma \in (1, 5/3]$, such that, for every initial data $(\tau_0, v_0, S_0) \in K$ with $TV_{\boldsymbol{R}}(\tau_0, v_0, S_0) \leq N$, when
>
> $$(\gamma - 1) TV_{\boldsymbol{R}}(\tau_0, v_0, S_0) \leq C_0 \quad \text{for any } \gamma \in (1, 5/3]$$
>
> the Cauchy problem [13] and [34] has a global entropy solution $(\tau, v, S)(t, x)$ which is bounded and satisfies
>
> $$TV_{\boldsymbol{R}}(\tau, v, S)(t, \cdot) \leq C\, TV_{\boldsymbol{R}}(\tau_0, v_0, S_0)$$
>
> for some constant $C > 0$ independent of $\gamma$.

This result specially includes that for the barotropic case (Nishida 1968, Nishida–Smoller 1973, DiPerna 1973). Some efforts in the direction of relaxing the requirement of small total variation have been made. Some extensions to the initial-boundary value problems have also been made. In addition, an entropy solution in BV with periodic data or compact support decays when $t \to 0$. Furthermore, even for a general hyperbolic system [4] for $x \in R$, we have:

> If the initial data functions $\boldsymbol{u}_0(x)$ and $\boldsymbol{v}_0(x)$ have sufficiently small total variation and $\boldsymbol{u}_0 - \boldsymbol{v}_0 \in L^1(\boldsymbol{R})$, then, for the corresponding exact Glimm, or wave-front tracking, or vanishing viscosity solutions $\boldsymbol{u}(t, x)$ and $\boldsymbol{v}(t, x)$ of the Cauchy problem [4] and [8], there exists a constant $C > 0$ such that
>
> $$\|\boldsymbol{u}(t, \cdot) - \boldsymbol{v}(t, \cdot)\|_{L^1(\boldsymbol{R})} \leq C \|\boldsymbol{u}_0 - \boldsymbol{v}_0\|_{L^1(\boldsymbol{R})} \quad \text{for all } t > 0 \qquad [35]$$

An immediate consequence is that the whole sequence of the approximate solutions constructed by the Glimm (1965) scheme, as well as the wave-front tracking method and the vanishing viscosity method, converges to a unique entropy solution of [4] and [8] when the mesh size or the viscosity coefficient tends to zero. More detailed discussions and extensive references about the $L^1$-stability of BV entropy solutions and related topics can be found in Bressan (2000) and Dafermos (2000); also see Chen and Wang (2002). Furthermore, the Riemann solution is unique and asymptotically stable in the class of entropy solutions to [13] with large variation satisfying only one physical entropy inequality (Chen-Frid-Li 2002).

### Multidimensional Steady Theory

The mathematical study of two-dimensional steady supersonic flows past wedges, whose vertex angles are less than the critical angle, can date back to the 1940s, since the stability of such flows is fundamental in applications (cf. Courant–Friedrichs (1948)). Local solutions around the wedge vertex were first constructed (Gu 1962, Schaeffer 1976, Li 1980).

Such global potential solutions were constructed when the wedge has some convexity, or is a small perturbation of the straight wedge with fast decay in the flow direction (Chen 2001, Chen-Xin-Yin 2002), or is piecewise smooth which is a small perturbation of straight wedge (Zhang 2003). For the two-dimensional steady supersonic flows governed by the full Euler equations past Lipschitz wedges, it indicates (Chen-Zhang-Zhu 2005a) that, when the wedge vertex angle is less than the critical angle, the strong shock front emanating from the wedge vertex is nonlinearly stable in structure globally, although there may be many weak shocks and vortex sheets between the wedge boundary and the strong shock front, under the BV perturbation of the wedge so that the total variation of the tangent function along the wedge boundary is suitably small. This asserts that any supersonic shock for the wedge problem is nonlinearly stable.

A self-similar gas flow past an infinite cone in $\boldsymbol{R}^3$ with small vertex angle is also nonlinearly stable upon the BV perturbation of the obstacle (Lien-Liu 1999). It is still open for the nonlinear stability when the infinite cone in $\boldsymbol{R}^3$ has arbitrary vertex angle. The stability issues of supersonic vertex sheets have been studied by classical linearized stability analysis, large-scale numerical simulations, and asymptotic analysis. In particular, the nonlinear development of instabilities of supersonic vortex sheets at high Mach number was predicted as time evolves (Woodward 1985, Artola-Majda 1989). In contrast with the prediction of evolution instability, steady supersonic vortex sheets, as time-asymptotics, are stable globally in structure, even under the BV perturbation of the Lipschitz walls, although there may be many weak shocks and supersonic vortex sheets away from the strong vortex sheet (Chen-Zhang-Zhu 2005b).

Transonic shock problems for steady fluid flows are important in applications (cf. Courant and Friedrichs (1948)). A program on the existence and stability of multidimensional transonic shocks has been initiated and three new analytical approaches have been developed (Chen-Feldman 2003, 2004). The transonic problems include the existence and stability of transonic shocks in the whole $\boldsymbol{R}^d$, the existence and stability of transonic flows past finite or infinite nozzles, the stability of transonic flows past infinite nonsmooth wedges, and the existence of regular shock reflection solutions. The first approach is an iteration scheme based on the nondegeneracy of the free boundary condition: the jump of the normal derivative of a solution across the free boundary has a strictly positive lower bound (Chen-Feldman 2003, 2004), which works for the nonlinear equations whose coefficients may depend on not only the solution itself but also the gradients of the solution. The second approach is a partial hodograph procedure, with which the existence and stability of multidimensional transonic shocks that are not nearly orthogonal to the flow direction can be handled (Chen-Feldman 2004): one of the main ingredients in this approach is to employ a partial hodograph transform to reduce the free boundary problem into a conormal boundary value problem for the corresponding nonlinear equations of divergence form and then develop techniques to solve the conormal boundary value problem. When the regularity of the steady perturbation is $C^{3,\alpha}$ or higher, the third approach is to employ the implicit function theorem to deal with the existence and stability problem. Another iteration approach, which works well for the two-dimensional equations whose coefficients depend only on the solution itself, has also been developed (Canic-Keyfitz-Lieberman 2000).

Further longstanding open problems include the existence of global transonic flows past an airfoil or a smooth obstacle (Morawetz 1956–58, 1985).

#### Multidimensional Unsteady Problems

Now we present some multidimensional time-dependent problems with a simplifying feature that the data (domain and/or the initial data) coupled with the structure of the underlying equations obey certain geometric structure so that the multidimensional problems can be reduced to lower-dimensional problems with more complicated couplings. Different types of geometric structure call for different techniques.

The Euler equations for compressible fluids with geometric structure describe many important fluid flows, including spherically symmetric flows and self-similar flows. Such geometric flows are motivated by many physical problems such as shock diffractions, supernovas formation in stellar dynamics, inertial confinement fusion, and underwater explosions. For the initial data with large amplitude having geometric structure, the required physical insight is: (1) whether the solution has the same geometric structure globally and (2) whether the solution blows up to infinity in a finite time. These questions are not easily understood in physical experiments and numerical simulations, especially for the blow-up, because of the limited capacity of available instruments and computers.

The first type of geometric structure is spherical symmetry. A criterion for $L^\infty$ Cauchy data functions of arbitrarily large amplitude was observed to guarantee the existence of spherically symmetric solutions in $L^\infty$ in the large for the isentropic flows, which model outgoing blast waves and large-time asymptotic solutions (Chen 1997). On the other hand, it is evident that the density blows up as $|\boldsymbol{x}| \to 0$ in general, especially for the focusing case; the singularity at the origin makes the problem truly multidimensional due to the reflection of waves from infinity and their strengthening as they move radially inwards. One of the important open questions is to understand the order of singularity, $\rho(t, |\boldsymbol{x}|) \sim |\boldsymbol{x}|^{-\alpha}$, at the origin for bounded Cauchy data.

The second type of geometric structure is self-similarity, that is, the solutions with initial data functions that give rise to self-similar solutions, especially including Riemann solutions. Compressible flow equations in $\boldsymbol{R}^d$, $d \geq 2$, with one or more linearly degenerate modes of wave propagation have additional difficulties. In that case, the global flow is governed by a reduced (self-similar) system which is of composite (hyperbolic–elliptic) type in the subsonic region. The linearly degenerate waves give rise to one or more families of degenerate characteristics which remain real in the subsonic region. In some cases, the reduced equations couple an elliptic (degenerate elliptic) problem for the density with a hyperbolic (transport) equation for the vorticity.

An important prototype for both practical applications and the theory of multidimensional complex wave patterns is the problem of diffraction of a shock wave which is incident along an inclined ramp (see Glimm and Majda (1991)). When a plane shock hits a wedge head-on, a self-similar reflected shock moves outward as the original shock moves forward. The computational and asymptotic analysis shows that various patterns of reflected shocks may occur, including regular reflection and (simple, double, and complex) Mach reflections. The main part or whole reflected shock is a transonic shock in the self-similar coordinates, for which the corresponding equation changes the type from hyperbolic to elliptic across the shock. There are few rigorous mathematical results on the global existence and stability of shock reflection solutions and the transition among regular, simple Mach, double Mach, and complex Mach reflections for the potential flow equation [19] and the full Euler equations [1]–[3]. Some results were recently obtained for simplified models including the transonic small-disturbance equation near the reflection point and the pressure gradient equation when the wedge is close to a flat wall.

For the potential flow equation [19], a self-similar solution is a solution of the form: $\Psi = t\phi(\boldsymbol{y})$, $\boldsymbol{y} = \boldsymbol{x}/t$. Letting $\varphi(\boldsymbol{y}) = -\boldsymbol{y}^2/2 + \phi(\boldsymbol{y})$, then the system can be rewritten in the form of a second-order equation of mixed hyperbolic–elliptic type in $\boldsymbol{y} \in \boldsymbol{R}^d$ by scaling:

$$\nabla_{\boldsymbol{y}} \cdot (\rho(|\nabla_{\boldsymbol{y}}\varphi|^2, \varphi)\nabla_{\boldsymbol{y}}\varphi) + d\rho(|\nabla_{\boldsymbol{y}}\varphi|^2, \varphi) = 0 \qquad [36]$$

with $\rho(q^2, z) = (1 - (q^2 + 2z)/2)^{1/(\gamma-1)}$. Equation [36] at $|\nabla_{\boldsymbol{y}}\varphi| = q$ is hyperbolic (pseudosupersonic) if $\rho(q^2, z) + q\rho_q(q^2, z) < 0$ and elliptic (pseudosubsonic) if $\rho(q^2, z) + q\rho_q(q^2, z) > 0$. Under this framework, the nature of the shock reflection pattern has been explored for weak incident shocks (strength $b$) and small wedge angles $2\theta_{\mathrm{w}}$ by a number of different scalings, a study of mixed equations, and matching asymptotics for the different scalings, where the parameter $\beta = c_1\theta_{\mathrm{w}}^2/b(\gamma + 1)$ ranges from 0 to $\infty$ and $c_1$ is the speed of sound behind the incident shock (Morawetz 1994). For $\beta > 2$, a regular reflection of both strong and weak kinds is possible as well as a Mach reflection; for $\beta < 1/2$, a Mach reflection occurs and the flow behind the reflection is subsonic and can be constructed in principle (with an elliptic problem) and matched; and for $1/2 < \beta < 2$, the flow behind a Mach reflection may be transonic which is a solution of a nonlinear boundary-value problem of mixed type. The basic pattern of reflection has been shown to be an almost semicircular shock issuing, for a regular reflection, from the reflection point on the wedge and, for a Mach reflection, matched with a local interaction flow. Some related observations were also made (Keller-Blank 1951, Hunter-Keller 1984, Hunter 1988). It is important to establish rigorous proofs. Recently, a rigorous existence proof was established for global solutions to shock reflection by large-angle wedges in Chen and Feldman (2005).

### Analytical Frameworks for Entropy Solutions

The recent great progress for entropy solutions for one-dimensional time-dependent Euler equations and two-dimensional steady Euler equations, based on BV, $L^1$, or even $L^\infty$ estimates, naturally arises the expectation that a similar approach may also be effective for the multidimensional Euler equations, or more generally, hyperbolic systems of conservation laws, especially,

$$\|\boldsymbol{u}(t, \cdot)\|_{\mathrm{BV}} \leq \mathrm{C}\|\boldsymbol{u}_0\|_{\mathrm{BV}} \qquad [37]$$

Unfortunately, this is not the case. The necessary condition for [37] to be held for $p \neq 2$ (Rauch 1986) is

$$\nabla f_k(\boldsymbol{u})\nabla f_l(\boldsymbol{u}) = \nabla f_l(\boldsymbol{u})\nabla f_k(\boldsymbol{u}) \quad \text{for all } k, l = 1, 2, \ldots, d \qquad [38]$$

The analysis suggests that only systems in which the commutativity relation [38] holds offer any hope for treatment in the framework of BV. This special case includes the scalar case $n = 1$ and the case of one space dimension $d = 1$. Beyond that, it contains very few systems of physical interest.

In this regard, it is important to identify effective analytical frameworks for studying entropy solutions of the multidimensional Euler equations [1]–[3], which are not in BV. Naturally, we want to approach the questions of existence, stability, uniqueness, and long-time behavior of entropy solutions with as much generality as possible. For this purpose, a theory of divergence-measure fields to construct such a global framework has been developed for studying entropy solutions (Chen-Frid 1999, 2000, Chen-Torres 2005, Chen-Torres-Ziemer 2005). For more details, see Chen (2005).

## Viscous Compressible Fluid Flows: Navier–Stokes Equations

Compressible fluid flows that are viscous and conduct heat are governed by the following Navier–Stokes equations:

$$\partial_t \rho + \nabla_{\boldsymbol{x}} \cdot \boldsymbol{m} = 0, \quad \boldsymbol{x} \in \boldsymbol{R}^d \qquad [39]$$

$$\partial_t \boldsymbol{m} + \nabla_{\boldsymbol{x}} \cdot \left(\frac{\boldsymbol{m} \otimes \boldsymbol{m}}{\rho}\right) + \nabla_{\boldsymbol{x}} p = \nabla_{\boldsymbol{x}} \cdot \boldsymbol{\Sigma} \qquad [40]$$

$$\partial_t E + \nabla_{\boldsymbol{x}} \cdot \left(\frac{\boldsymbol{m}}{\rho}(E + p)\right) = \nabla_{\boldsymbol{x}} \cdot \left(\frac{\boldsymbol{m}}{\rho} \cdot \boldsymbol{\Sigma}\right) - \nabla_{\boldsymbol{x}} \cdot \boldsymbol{q} \qquad [41]$$

Here, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\nabla_{\boldsymbol{x}}\boldsymbol{v}, \rho, \theta)$ is the viscous stress tensor which is symmetric from the conservation of angular momentum and $\boldsymbol{q}$ is the heat flux. If the fluid is isotropic and the viscous tensor $\boldsymbol{\Sigma}$ is a linear function of $\nabla_{\boldsymbol{x}}\boldsymbol{v}$ and invariant under a change of reference frame (translation and rotation), then we deduce from elementary algebraic manipulations that necessarily

$$\boldsymbol{\Sigma} = \lambda(\rho, \theta)\nabla_{\boldsymbol{x}} \cdot \boldsymbol{v} + 2\mu(\rho, \theta)D \qquad [42]$$

which corresponds to the Newtonian fluids, where $D = (\nabla_{\boldsymbol{x}}\boldsymbol{v} + (\nabla_{\boldsymbol{x}}\boldsymbol{v})^{\top})/2$ is the deformation tensor and $\lambda$ and $\mu$ are the Lamé viscosity coefficients. Furthermore, since the fluid is isotropic, we are led to the Fourier law:

$$\boldsymbol{q} = -k(\rho, \theta, |\nabla_{\boldsymbol{x}}\theta|)\nabla_{\boldsymbol{x}}\theta$$

for scalar function $k$ which, in most cases, is taken to be simply a function of $\rho$ and $\theta$, or even a constant called the thermal conduction coefficient. Again, system [39]–[41] is closed by the constitutive relations in [5]. The equation for entropy $S$ is

$$\partial_t(\rho S) + \nabla_{\boldsymbol{x}} \cdot \left(\boldsymbol{m}S + \frac{\boldsymbol{q}}{\theta}\right) = \frac{\boldsymbol{\Sigma}(\nabla_{\boldsymbol{x}}\boldsymbol{v}) : \nabla_{\boldsymbol{x}}\boldsymbol{v}}{\theta} - \frac{\boldsymbol{q} \cdot \nabla_{\boldsymbol{x}}\theta}{\theta^2} \qquad [43]$$

The second law of thermodynamics indicates that the right-hand side of [43] should be non-negative which yields the restriction:

$$k(\rho, \theta, |\nabla_{\boldsymbol{x}}\theta|) \geq 0, \quad \mu \geq 0, \quad \lambda + 2\mu/d \geq 0$$

The case $\mu > 0$ and $\lambda + \mu > 0$ is the viscous case with heat conductivity $k > 0$. In particular, the kinetic theory indicates that the Stokes relationship should hold, namely $\lambda = -2\mu/d$ and the adiabatic component $\gamma = 5/3$ for monatomic gases.

In mathematical viscous fluid dynamics, an important model is the barotropic model for viscous fluids, that is, $p = p(\rho)$. Then, the specific energy $E$ can be taken in the form of $E = (1/2)\rho|\boldsymbol{v}|^2 + \rho e(\rho)$ with $e'(\rho) = p(\rho)/\rho^2$. For classical solutions, the energy of a barotropic flow satisfies the equality:

$$\partial_t E + \nabla_{\boldsymbol{x}} \cdot ((E + p)\boldsymbol{v}) = \nabla_{\boldsymbol{x}} \cdot (\boldsymbol{\Sigma}\boldsymbol{v}) - \boldsymbol{\Sigma} : \nabla_{\boldsymbol{x}}\boldsymbol{v}$$

which is now a direct consequence of [39] and [40].

The question of local existence of classical solutions to [39]–[41] for regular initial data was addressed by Nash (1962), where there is no indication whether or not these solutions exist for all times.

In the case of one space dimension, the well-posedness is largely settled. The basic result for the existence of classical solutions is that of Kazhikhov (1976); see Lions (1998) and Feireisl (2004) for extensive references. The discontinuous solutions have been constructed (Shelukhin 1979, Serre 1986, Hoff 1987, Chen-Hoff-Trivisa 2000).

For the Navier–Stokes equations in $\boldsymbol{R}^3$ with general equation of state, the global classical solutions for the Cauchy problem and various initial-boundary value problems whose initial data is small around a constant state have been

constructed (Matsumura-Nishida 1980, 1983). The approach is to obtain *a priori* estimates via energy methods for extending the local solution or for a difference method globally. These results have been extended to the Cauchy problem or the initial-boundary value problems with small discontinuous initial data (Hoff 1997).

For the Navier–Stokes equations in $\boldsymbol{R}^d$ for barotropic flows with [11] and large initial data, the global existence of solutions containing vacuum for the Cauchy problem or various initial-boundary value problems was first established by Lions (1998) for $\gamma \geq 3/2$ if $d=2$, $\gamma \geq 9/5$ if $d=3$, and $\gamma > d/2$ if $d \geq 4$. The gap was closed by Feireisl–Novotný–Petzeltová (2001) for the full range $\gamma > d/2$. These results have been extended to the full Navier–Stokes equations describing the motion of a general compressible, viscous, and heat conducting fluid (see Feireisl (2004)). The physically relevant isothermal case, $\gamma = 1$, is completely open even if $d=2$. The only large data existence result is that for radially symmetric data (Hoff 1992). The general case $\gamma \geq 1$ and $d=3$ for radially symmetric data was solved only recently (Jiang-Zhang 2001).

The lower-bound estimate on the density is a delicate issue. Weak solutions containing vacuum for the isentropic viscous flows with constant viscosity are unstable in general (Hoff-Serre 1991). Hence, it is important to see whether vacuum will never develop if the initial data is away from vacuum; this has been shown for the one-dimensional case for large initial data and for the multidimensional case with small data. On the other hand, from the kinetic theory, if solutions contain vacuum, then the viscosity coefficients in the Navier–Stokes equations should depend on the density near vacuum; this indeed stabilizes the solutions for the one-dimensional case.

The stability of viscous shock waves has been studied for the one-dimensional case (see Liu (2000) and the references therein). The compressible–incompressible limits from the isentropic compressible to incompressible Navier–Stokes equations when the Mach number tends to zero have been established for arbitrarily weak solutions (Lions-Masmoudi 1998) and for smooth solutions and a class of initial data functions (Hoff 1998).

The inviscid limits from the Navier–Stokes equations to the Euler equations have been established as long as the solutions of the Euler equations are smooth, when the viscosity and heat conductivity coefficients tend to zero (Klainerman-Majda 1982). It is completely open for general entropy solutions, even in the one-dimensional case.

*See also:* Breaking Water Waves; Capillary Surfaces; Fluid Mechanics: Numerical Methods; Geophysical Dynamics; Incompressible Euler Equations: Mathematical Theory; Inviscid Flows; Magnetohydrodynamics; Newtonian Fluids and Thermohydraulics; Non-Newtonian Fluids; Partial Differential Equations: Some Examples; Stability of Flows; Viscous Incompressible Fluids: Mathematical Theory.

## Further Reading

Bressan A (2000) *Hyperbolic Systems of Conservation Laws: The One-Dimensional Cauchy Problem*. Oxford: Oxford University Press.

Chen G-Q (2005) Euler equations and related hyperbolic conservation laws. In: Dafermos CM and Feireisl E (eds.) *Handbook of Differential Equations II: Evolutionary Differential Equations*, Chapter 1, pp. 1–104. Amsterdam: Elsevier.

Chen G-Q and Wang D (2002) The Cauchy problem for the Euler equations for compressible fluids. In: Friedlander S and Serre D (eds.) *Handbook of Mathematical Fluid Dynamics*, vol. 1, ch. 5, pp. 421–543. Amsterdam: Elsevier Science B.V.

Courant R and Friedrichs KO (1948) *Supersonic Flow and Shock Waves*. New York: Springer.

Dafermos CM (2005) *Hyperbolic Conservation Laws in Continuum Physics* (2nd edn). Berlin: Springer.

Feireisl E (2004) *Dynamics of Viscous Compressible Fluids*. Oxford: Oxford University Press.

Glimm J (1965) Solutions in the large for nonlinear hyperbolic system of equations. *Communications on Pure and Applied Mathematics* 18: 95–105.

Glimm J and Majda A (1991) *Multidimensional Hyperbolic Problems and Computations*. New York: Springer.

Lax PD (1973) *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*. Philadelphia: SIAM.

Lions PL (1996, 1998) *Mathematical Topics in Fluid Mechanics*, vols. 1–2. New York: Oxford University Press.

Liu T-P (2000) *Hyperbolic and Viscous Conservation Laws*, CBMS-NSF RCSAM, vol. 72. Philadelphia: SIAM.

Majda A (1984) *Compressible Fluid Flow and Systems of Conservation Laws in Several Space Variables*. New York: Springer.

# Computational Methods in General Relativity: The Theory

**M W Choptuik**, University of British Columbia, Vancouver, Canada



## Conventions and Units

This article adopts many of the conventions and notations of Misner, Thorne, and Wheeler (1973) – hereafter denoted MTW – including metric signature $(-+++)$; definitions of Christoffel symbols and curvature tensors (up to index permutations permitted by standard symmetries of the tensors in a coordinate basis); the use of Greek indices $\alpha, \beta, \gamma, \ldots$, ranging over the spacetime coordinate values $(0,1,2,3) \rightarrow (t, x^1, x^2, x^3)$, to denote the components of spacetime tensors such as $g_{\mu\nu}$; the similar use of Latin indices $i, j, k, \ldots$, ranging over the spatial coordinate values $(1,2,3) \rightarrow (x^1, x^2, x^3)$, for spatial tensors such as $\gamma_{ij}$; the use of the Einstein summation convention for both types of indices; the use of standard Kronecker delta symbols (tensors), $\delta^{\mu}{}_{\nu}$ and $\delta^{i}{}_{j}$; the choice of geometric units, $G = c = 1$; and, finally, the normalization of the matter fields implicit in the choice of the constant $8\pi$ in [1].

The majority of the equations that appear in this article are tensor equations, or specific components of tensor equations, written in traditional index (not abstract index) form. Thus, these equations are generally valid in any coordinate system, $(t, x^i)$, but, of course do require the introduction of a coordinate basis and its dual. This approach is also largely a matter of convention, since all of what follows can be derived in a variety of fashions, some of them purely geometrical, and there are also approaches to numerical relativity based, for example, on frames rather than coordinate bases.

This article departs from MTW in its use of $\alpha, \beta^i$, and $\gamma_{ij}$ to denote the lapse, shift, and spatial metric, respectively, rather than MTW's $N$, $N^i$, and ${}^{(3)}g_{ij}$.

Finally, the operations of partial differentiation with respect to coordinates $x^{\mu}$, $t$, and $x^i$ are denoted $\partial_{\mu}$, $\partial_t$, and $\partial_i$, respectively.

## Introduction

The numerical analysis of general relativity, or numerical relativity, is concerned with the use of computational methods to derive approximate solutions to the Einstein field equations

$$G_{\mu\nu} = 8\pi T_{\mu\nu} \qquad [1]$$

Here, $G_{\mu\nu}$ is the Einstein tensor – that contracted piece of the Riemann curvature tensor that has vanishing divergence – and $T_{\mu\nu}$ is the stress tensor of the matter content of the spacetime. $T_{\mu\nu}$ likewise has vanishing divergence, an expression of the principle of local conservation of stress–energy that general relativity embodies.

The elegant tensor formulation [1] belies the fact that, ultimately, the field equations are generically a complicated and nonlinear set of partial differential equations (PDEs) for the components of the spacetime metric tensor, $g_{\mu\nu}(x^{\alpha})$, in some coordinate system $x^{\alpha}$. Moreover, implicit in a numerical solution of [1] is the numerical solution of the equations of motion for any matter fields that couple to the gravitational field – that is, that contribute to $T_{\mu\nu}$. The reader is reminded that it is a hallmark of general relativity that, in principle, all matter fields – including massless ones such as the electromagnetic field – contribute to $T_{\mu\nu}$.

Now, in the $3+1$ approach to general relativity that is described below, the task of solving the field equations [1] is formulated as an initial-value or Cauchy problem. Specifically, the spacetime metric, $g_{\mu\nu}(x^{\alpha}) = g_{\mu\nu}(t, x^k)$, which encodes all geometric information concerning the spacetime, $\mathcal{M}$, is viewed as the time history, or dynamical evolution, of the spatial metric, $\gamma_{ij}(0, x^k)$, of an initial spacelike hypersurface, $\Sigma(0)$. In any practical calculation, the degree to which the matter fields "back-react" on the gravitational field, that is, contribute to $T_{\mu\nu}$ substantially enough to cause perturbations in $g_{\mu\nu}$ at or above the desired accuracy threshold, will thus depend on the specifics of the initial configuration.

In astrophysics, there are relatively few well-identified environments in which it is generally thought to be crucial to the faithful emulation of the physics that the matter fields be fully coupled to the gravitational field. However, both observationally and theoretically, the existence of gravitationally compact objects is quite clear. Gravitationally compact means that a star with mass, $M$, has a radius, $R$, comparable to its Schwarzschild radius, $R_M$, which is defined by

$$R_M = \frac{2G}{c^2} M \approx 10^{-27}\,\mathrm{kg\,m^{-1}} \qquad [2]$$

Here, and only here, $G$ and $c$ – Newton's gravitational constant and the speed of light, respectively – have been explicitly reintroduced. The fact that $R_M/R$ is about $10^{-6}$ and $10^{-9}$ at the surfaces of the sun and earth, respectively, is a reminder of just how

weak gravity is in the locality of Earth. However, as befits anything of Einsteinian nature, the weakness of gravity is relative, so that at the surface of a neutron star, one would find

$$\frac{R_M}{R} \sim 0.4 \qquad [3]$$

while for black holes, one has

$$\frac{R_M}{R} = 1 \qquad [4]$$

In such circumstances, gravity is anything but weak! Furthermore, in situations where the matter–energy distribution has a highly time-dependent quadrupole moment – such as occurs naturally with a compact-binary system (i.e., a gravitationally bound two-body system, in which each of the bodies is either a black hole or a neutron star) – the dynamics of the gravitational field, including, crucially, the dynamics of the radiative components of the gravitational field, can be expected to dominate the dynamics of the overall system, matter included. For scenarios such as these, it should come as no surprise that the solution of the combined gravitohydrodynamical system begs for numerical analysis.

In addition, both from the physical and mathematical perspectives, it is also natural to study the strong, field dynamic regimes ($R \to R_M$ and/or $v \to c$, where $v$ is the typical speed characterizing internal bulk motion of the matter) of general relativity within the context of a variety of matter models. Typical processes addressed by these theoretical studies include the process of black hole formation, end-of-life events for various types of model stars, and, again, the interaction, including collisions, of gravitationally compact objects. Note that it is another hallmark of general relativity that highly dynamical spacetimes need not contain any matter; indeed, the interaction of two black holes – the natural analog of the Kepler problem in relativity – is a vacuum problem; that is, it is described by a solution of [1] with $T_{\mu\nu} = 0$.

Motivated in significant part by the large-scale efforts currently underway to directly detect gravitational radiation (gravitational waves), much of the contemporary work in numerical relativity is focused on precisely the problem of the late phases of compact-binary inspiral and merger. Such binaries are expected to be the most likely candidates for early detection by existing instruments such as TAMA, GEO, VIRGO, LIGO, and, more likely, by planned detectors including LIGO II and LISA (see, e.g., Hough and Rowan (2000)). Detailed and accurate predictions of expected waveforms from these events – using the techniques of numerical relativity – have the potential to substantially hasten the discovery process, on the basis of the general principle that if one knows what signal to look for, it is much easier to extract that signal from the experimental noise.

The computational task facing numerical relativists who study problems such as binary inspiral is formidable. In particular, such problems are intrinsically "3D," to use the CFD (computational fluid dynamics) nomenclature in which time dependence is always assumed. That is, the PDEs that must be solved govern functions, $F(t, x^k)$, that depend on all three spatial coordinates, $x^k$, as well as on time, $t$. Unfortunately, even a cursory description of 3D work in numerical relativity as it stands at this time is far beyond the scope of this article.

What follows, then, is an outline of a traditional approach to numerical relativity that underpins many of the calculations from the early years of the field (1970s and 1980s), most of which were carried out with simplifying restrictions to either spherical symmetry or axisymmetry. The mathematical development, which will hereafter be called the $3+1$ approach to general relativity, has the advantage of using tensors and an associated tensor calculus that are reasonably intuitive for the physicist. This "standard" $3+1$ approach is also sufficient in many instances (particularly those with symmetry) in the sense that it leads to well-posed sets of PDEs that can be discretized and then solved computationally in a convergent (stable) fashion. In addition, a thorough understanding of the $3+1$ approach will be of significant help to the reader wishing to study any of the current literature in numerical relativity, including the 3D work.

However, the reader is strongly cautioned that the blind application of any of the equations that follow, especially in a 3D context, may well lead to "ill-posed systems," numerical analysis of which is useless. Anyone specifically interested in using the methods of numerical relativity to generate discrete, approximate solutions to [1], particularly in the generic 3D case, is thus urged to first consult one of the comprehensive reviews of numerical relativity that continue to appear at fairly regular intervals (see, e.g., Lehner (2001), or Baumgarte and Shapiro (2003)). Most such references will also provide a useful overview of many of the most popular numerical techniques that are currently being used to discretize (convert to algebraic form) the Einstein equations, as well as the main algorithms that are used to solve the resulting discrete equations. These subjects are not

described below, not least since discussion of the available discretization techniques only makes sense in the context of PDEs of specific systems with specific boundary conditions, while there is only space here to describe the general mathematical setting for $3+1$ numerical relativity.

## The $3+1$ Spacetime Split

At least at the current time, computations in numerical relativity are restricted to the case of globally hyperbolic spacetimes. A spacetime (four-dimensional pseudo-Riemannian manifold), $\mathcal{M}_\Sigma$, endowed with a metric, $g_{\mu\nu}$, is globally hyperbolic if there is at least one edgeless, spacelike hypersurface, $\Sigma(0)$, that serves as a Cauchy surface. That is, provided that the initial data for the gravitational field are set consistently on $\Sigma(0)$ – so that the four constraint equations are satisfied (see below) – the entire metric $g_{\mu\nu}(t, x^i)$ can be determined from the field equations [1] (with appropriate boundary conditions), and thus, so can the complete geometric structure of the spacetime manifold.

To be sure, global hyperbolicity is restrictive. It excludes, for example, the highly interesting Gödel universe. However, particularly from the point of view of studying asymptotically flat solutions (or solutions asymptotic to any of the currently popular cosmologies), as is usually the case in astrophysics, the requirement of global hyperbolicity is natural.

The $3+1$ split is based on the complete foliation of $\mathcal{M}_\Sigma$ based on level surfaces of a scalar function, $t$ – the time function. That is, the $t=\text{const.}$ slices, are three-dimensional spacelike (Riemannian) hypersurfaces, and, as $t$ ranges from $-\infty$ to $+\infty$, completely fill the spacetime manifold, $\mathcal{M}_\Sigma$. In order for the $\Sigma(t)$ to be everywhere spacelike, $t$ must be everywhere timelike:

$$g_{\mu\nu}\nabla^\mu t \nabla^\nu t < 0 \qquad [5]$$

Here $\nabla_\mu$ is the spacetime covariant derivative operator compatible with the four metric, $g_{\mu\nu}$, thus satisfying $\nabla_\alpha g_{\mu\nu}=0$, and $g^{\mu\nu}$ is the inverse metric tensor, which satisfies $g^{\mu\alpha}g_{\alpha\nu}=\delta^\mu{}_\nu$. The reader is reminded that $\delta^\mu{}_\nu$ is a Kronecker delta symbol; that is, $\delta^\mu{}_\nu$ has the value 1 if $\mu=\nu$, and the value 0 otherwise.

Furthermore, the scalar function $t$ is now adopted as the temporal coordinate, so that $x^\mu=(t, x^i)$, where the $x^i$ are the three spatial coordinates. As noted implicitly above, since the problem under consideration is a pure Cauchy evolution, the range of $t$ should nominally be infinite, both to the future as well as to the past; that is, the solution domain is

$$-\infty < t < \infty \qquad [6]$$

$$|X| \equiv \left(\gamma_{ij}x^i x^j\right)^{1/2} < \infty \qquad [7]$$

However, this assumes that one has global existence for arbitrarily strong initial data, which is decidedly not always the case in general relativity. Indeed, "continued" or "catastrophic" gravitational collapse – that is, the process of black hole formation – signaled, in modern language, by the appearance of a trapped surface, inexorably leads to a physical singularity, which – the somewhat vague nature of the singularity theorems of Penrose, Hawking, and others notwithstanding – in actual numerical computations invariably turns out to be "catastrophic" in terms of Cauchy evolution.

Such behavior in time-dependent nonlinear PDEs is quite familiar in the mathematical community at large, where it is frequently known as finite-time blow-up (or finite-time singularity). However, despite the fact that such behavior is one of the most fascinating aspects of solutions of the Einstein equations, the following discussion will be, implicitly at least, restricted to the case of weak initial data, that is, to initial data for which there is global existence.

With the manifold $\mathcal{M}_\Sigma$ sliced into an infinite stack of spacelike hypersurfaces, $\Sigma(t)$, attention shifts to any single surface, as well as to the manner in which such a generic surface is embedded in the spacetime.

First, each spacelike hypersurface, $\Sigma(t)$, is itself a three-dimensional Riemannian differential manifold with a metric $\gamma_{ij}(t, x^k)$. (Note that in this discussion, the symbol $t$ is to be understood to represent any specific value of coordinate time.) From this metric, one can construct an inverse metric, $\gamma^{ij}(t, x^k)$, defined, as usual, so that

$$\gamma^{ik}\gamma_{kj} = \delta^i{}_j \qquad [8]$$

Associated with the spatial metric, $\gamma_{ij}$, is a natural spatial covariant derivative operator, $D_i$, that is compatible with $\gamma_{ij}$:

$$D_k\gamma_{ij} = 0 \qquad [9]$$

With the spatial metric, $\gamma_{ij}$, and its inverse, $\gamma^{ij}$, in hand, the standard formulas of tensor analysis can be applied to compute the usual suite of geometrical tensors. All tensors thus computed, and indeed, all tensors defined intrinsically to the

hypersurfaces $\Sigma(t)$ are called "spatial" tensors, and have their indices (if any) raised and lowered with $\gamma^{ij}$ and $\gamma_{ij}$, respectively.

Thus, the Christoffel symbols of the second kind, $\Gamma^i{}_{jk}$, are given by

$$\Gamma^i{}_{jk} = \frac{1}{2}\gamma^{il}\left(\partial_k\gamma_{lj} + \partial_j\gamma_{lk} - \partial_l\gamma_{jk}\right) \qquad [10]$$

Note that these quantities are symmetric in their last two indices

$$\Gamma^i{}_{jk} = \Gamma^i{}_{kj} \qquad [11]$$

and that they can be used, as usual, in explicit calculation of the action of the spatial covariant derivative operator on an arbitrary tensor. In particular, for the special cases of a spatial vector, $V^i$, and a covector (1-form), $W_i$, one has

$$D_i V^j = \partial_i V^j + \Gamma^j{}_{ik} V^k \qquad [12]$$

and

$$D_i W_j = \partial_i W_j - \Gamma^k{}_{ij} W_k \qquad [13]$$

respectively.

Given the Christoffel symbols, the components of the spatial Riemann tensor, denoted here $\mathcal{R}_{ijk}{}^l$, are computed using

$$\mathcal{R}_{ijk}{}^l = \partial_j\Gamma^l{}_{ik} - \partial_i\Gamma^l{}_{jk} + \Gamma^m{}_{ik}\Gamma^l{}_{mj} - \Gamma^m{}_{jk}\Gamma^l{}_{mi} \qquad [14]$$

Finally, the Ricci tensor, $\mathcal{R}^i{}_j$, and Ricci scalar, $\mathcal{R}$, are defined in the usual fashion

$$\mathcal{R}^i{}_j = \gamma^{ik}\mathcal{R}_{kj} = \gamma^{ik}\mathcal{R}_{klj}{}^l \qquad [15]$$

$$\mathcal{R} = \gamma^{ij}\mathcal{R}_{ij} \qquad [16]$$

The reader should again note that all of the tensors just defined "live" on each and every single spacelike hypersurface, $\Sigma(t)$, and are thus known as hypersurface-intrinsic quantities. In particular, the spatial Riemann tensor, $\mathcal{R}_{ijk}{}^l$, which encodes all intrinsic geometric information about $\Sigma(t)$, in no way depends on how the slice is embedded in the spacetime $\mathcal{M}_\Sigma$.

The next step in the 3 + 1 approach involves rewriting the fundamental spacetime line element for the squared proper distance, $ds^2$, between two spacetime events, $\mathcal{P}$ and $\mathcal{Q}$, having coordinates $x^\mu$ and $x^\mu + dx^\mu$, respectively,

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu \qquad [17]$$

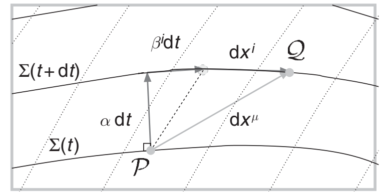



**Figure 1** Spacetime displacement in the $3+1$ approach, following Misner, Thorne, and Wheeler (1973). Solid lines represent surfaces of constant time, $t$; that is, each solid line represents a single spacelike hypersurface, $\Sigma(t)$. Dotted lines denote trajectories of constant spatial coordinate, that is, trajectories with $x^k = \text{const}$. The lapse function, $\alpha(t, x^k)$, encodes the (local) ratio between elapsed coordinate time, $dt$, and elapsed proper time, $d\tau = \alpha\, dt$, for an observer moving normal to the slices (i.e., for an observer with a 4-velocity, $u^\mu$, identical to the hypersurface normal, $n^\mu$). Similarly, the shift vector, $\beta^i(t, x^k)$, describes the shift, $\beta^i(t, x^i)\, dt$, in trajectories of constant spatial coordinate – the dotted lines in the figure – relative to motion perpendicular to the slices. The $3+1$ form of the line element [18] then follows immediately from an application of the spacetime version of the Pythagorean theorem.

As Figure 1 illustrates, a quick route to the 3 + 1 decomposition of the above expression, and thus of the tensor $g_{\mu\nu}$ itself, is based on an application of the "four-dimensional Pythagorean theorem." In setting up the calculation, one naturally identifies four functions, the scalar lapse, $\alpha(t, x^k)$, and the vector shift, $\beta^i(t, x^k)$, that encode the full coordinate (gauge) freedom of the theory. That is, complete specification of the lapse and shift is equivalent to completely fixing the spacetime coordinate system.

In light of the above discussion, and again referring to Figure 1, one readily deduces the 3 + 1 decomposition of the spacetime line element:

$$ds^2 = -\alpha^2 dt^2 + \gamma_{ij}\left(dx^i + \beta^i dt\right)\left(dx^j + \beta^j dt\right) \qquad [18]$$

A rearranged form of this last expression is also often seen in the literature:

$$ds^2 = \left(-\alpha^2 + \beta_k\beta^k\right)dt^2 + 2\beta_k dx^k dt + \gamma_{ij}dx^i dx^j \qquad [19]$$

The following useful identifications of the "time–time," "time–space," and "space–space" pieces of the spacetime metric, $g_{\mu\nu}$, follow immediately from [19]:

$$g_{00} = -\alpha^2 + \beta^i\beta_i \qquad [20]$$

$$g_{0i} = g_{i0} = \beta_i = \gamma_{ik}\beta^k \qquad [21]$$

$$g_{ij} = \gamma_{ij} \qquad [22]$$

This last relation is an example of a useful general result; the purely spatial components, $Q_{ijk\cdots}$, of a

completely covariant, but otherwise arbitrary, spacetime tensor, $Q_{\alpha\beta\gamma\cdots}$, constitute the components of a completely covariant spatial tensor.

A straightforward calculation, which provides a good exercise in the use of the $3+1$ calculus, yields the following equally useful identifications for various pieces of the inverse spacetime metric: $g^{\alpha\beta}$

$$g^{00} = -\alpha^{-2} \quad [23]$$

$$g^{0i} = g^{i0} = \alpha^{-2}\beta^i \quad [24]$$

$$g^{ij} = \gamma^{ij} - \alpha^{-2}\beta^i\beta^j \quad [25]$$

Since the Einstein field equations are equations with, loosely speaking, geometry on one side and matter on the other, tensors built from matter fields must also be decomposed. In particular, it is conventional to define tensors, $\rho$, $j_i$, and $S_{ij}$ that result from various projections of the spacetime stress energy tensor, $T_{\mu\nu}$, onto the hypersurface:

$$\rho \equiv n_\mu n_\nu T^{\mu\nu} \quad [26]$$

$$j_i \equiv -n_\mu T^\mu{}_i \quad [27]$$

$$S_{ij} \equiv T_{ij} \quad [28]$$

For observers with 4-velocities $u^\mu$ equal to $n^\mu$, and only for those observers with $u^\mu = n^\mu$, the above quantities have the interpretation of the locally and instantaneously measured energy density, momentum density, and spatial stresses, respectively. As with the geometric quantities, all of the matter variables, $\rho$, $j_i$, and $S_{ij}$ defined in [26]–[28] are spatial tensors and thus have their indices (if any) raised and lowered with the 3-metric. Note that the identification $S_{ij} = T_{ij}$ is another illustration of the general result mentioned in the context of the previous identification of $\gamma_{ij}$ and $g_{ij}$.

Finally, observing that time parameters are naturally defined in terms of level surfaces (equipotential surfaces), it should be no surprise that the covariant components, $n_\mu$, of the hypersurface normal field,

$$n_\mu = (-\alpha, 0, 0, 0) \quad [29]$$

are simpler than the components, $n^\mu$, of the normal itself,

$$n^\mu = \left(\alpha^{-1}, \alpha^{-1}\beta^i\right) \quad [30]$$

and, in fact, eqn [29] can also be deduced from a quick study of **Figure 1**.

In the $3+1$ approach, in addition to the 3-metric, $\gamma_{ij}(t, x^k)$, and coordinate functions, $\alpha(t, x^i)$ and $\beta(t, x^i)$, it is convenient to introduce an additional rank-2 symmetric spatial tensor, $K_{ij}(t, x^k)$, known as the extrinsic curvature (or second fundamental form). This additional tensor is analogous to a time derivative of $\gamma_{ij}(t, x^k)$, or, from a Hamiltonian perspective, to a variable that is dynamically conjugate to $\gamma_{ij}(t, x^k)$.

As the name suggests, the extrinsic curvature describes the manner in which the slice $\Sigma(t)$ is embedded in the manifold (to be contrasted with $\mathcal{R}_{ijk}{}^l$ defined by [14] which is, as mentioned previously, completely insensitive to the manner in which the hypersurface is embedded in $\mathcal{M}_\Sigma$).

Geometrically, $K_{ij}$ is computed by calculating the spacetime gradient of the normal covector field, $n_\mu$, and projecting the result on to the hypersurface,

$$K_{ij} = -\tfrac{1}{2}\nabla_i n_j \quad [31]$$

where it must be stressed that $\nabla_\mu$ is the spacetime covariant derivative operator compatible with the 4-metric, $g_{\alpha\beta}$; that is, $\nabla_\mu g_{\alpha\beta} = 0$. A straightforward tensor calculus calculation then yields the following, which can be viewed as a definition of the $K_{ij}$:

$$K_{ij} = \frac{1}{2\alpha}\left(\partial_t \gamma_{ij} + D_i\beta_j + D_j\beta_i\right) \quad [32]$$

Here, $D_i$ is the spatial covariant metric, compatible with $\gamma_{ij}(D_k\gamma_{ij} = 0)$, that was defined previously. Observe that this equation can be easily solved for $\partial_t\gamma_{ij}$ (this will be done below), and thus, in the $3+1$ approach it is [32] that is the origin of the evolution equations for the 3-metric components, $\gamma_{ij}$.

## Einstein's Equations in $3+1$ Form

### The Constraint Equations

As is well known, as a result of the coordinate (gauge) invariance of the theory, general relativity is overdetermined in a sense completely analogous to the situation in electrodynamics with the Maxwell equations. One of the ways that this situation is manifested is via the existence of the constraint equations of general relativity. Briefly, starting from the naive view that the ten metric functions, $g_{\mu\nu}(t, x^i)$, that completely determine the spacetime geometry are all dynamical – that is, that they satisfy second-order-in-time equations of motion – one finds that the Einstein equations do not provide dynamical equations of motion for the lapse, $\alpha$, or the shift, $\beta^i$. Rather, four of the field equations [1] are equations of constraint for the "true" dynamical variables of the theory, $\{\gamma_{ij}, \partial_t\gamma_{ij}\}$, or, equivalently, $\{\gamma_{ij}, K^i{}_j\}$. Note that in the following, the mixed form, $K^i{}_j$, is at times used – again by convention – as the principal representation of the extrinsic curvature tensor (instead of $K_{ij}$ as previously, or $K^{ij}$).

Thus, four of the components of [1] can be written in the form

$$C^{\mu}\left(\gamma_{ij}, K^{i}{}_{j}, \partial_k \gamma_{ij}, \partial_l \partial_k \gamma_{ij}, \partial_k K^{i}{}_{j}\right) = T^{\mu} \qquad [33]$$

where $T^{\mu}$ depends only on the matter content in the spacetime. Note that in addition to having no dependence on $\partial_t^2 \gamma_{ij}$, the constraints are also independent of $\alpha$ and $\beta^i$.

If the Einstein equations [1] are to hold throughout the spacetime, then the constraints [33] must hold on each and every spacelike hypersurface, $\Sigma(t)$, including, crucially, the initial hypersurface, $\Sigma(0)$. From the point of view of Cauchy evolution, this means that the 12 functions, $\{\gamma_{ij}(0, x^k), K^i{}_j(0, x^k)\}$, constituting the gravitational part of the initial data, are not completely freely specifiable, but must satisfy the four constraints

$$C^{\mu}\left(\gamma_{ij}(0, x^k), K^i{}_j(0, x^k), \ldots\right) = T^{\mu}(0, x^k) \qquad [34]$$

However, provided initial data that do satisfy the equations is chosen, then – as consistency of the theory demands – the dynamical equations of motion for the $\{\gamma_{ij}, K^i{}_j\}$ (eqns [37] and [38] below) guarantee that the constraints will be satisfied on all future (or past) hypersurfaces, $\Sigma(t)$. In this internal self-consistency, the geometrical Bianchi identities, $\nabla_{\mu} G^{\mu\nu} = 0$, and the local conservation of stress energy, $\nabla_{\mu} T^{\mu\nu} = 0$, play crucial roles.

In the $3+1$ approach, as one would expect, the constraint equations further naturally subdivide into a scalar equation

$$\mathcal{R} - K_{ij}K^{ij} + K^2 = 16\pi\rho \qquad [35]$$

and a (spatial) vector equation

$$D_j K^{ij} - D^i K = 8\pi j^i \qquad [36]$$

where the energy and momentum densities, $\rho$ and $j^i = \gamma^{ik} j_k$, are given by [26]–[28]. Equations [35] and [36] are often known as the Hamiltonian and momentum constraint, respectively, not least since the behavior of their solutions as $X \equiv \sqrt{\gamma_{ij} x^i x^j} \to \infty$ encodes the conserved mass and linear momentum (four numbers) that can be defined in asymptotically flat spacetimes.

In a general $3+1$ coordinate system, and with an appropriate choice of variables, the constraints can be written as a set of quasilinear elliptic equations for four of the $\{\gamma_{ij}, K^i{}_j\}$ (or, more properly, for certain algebraic combinations of the $\{\gamma_{ij}, K^i{}_j\}$). Thus, especially for 2D and 3D calculations, the setting of initial data for the Cauchy problem in general relativity is itself a highly nontrivial mathematical and computational exercise. Readers wishing more details on this subject are directed to the comprehensive review by Cook (2000).

### The Evolution Equations

As discussed above, in the $3+1$ form of the Einstein equations [1], the spatial metric, $\gamma_{ij}$, and the extrinsic curvature, $K^i{}_j$, are viewed as the dynamical variables for the gravitational field. The remainder of the $3+1$ equations are thus two sets of six first-order-in-time evolution equations; one set for $\gamma_{ij}$,

$$\begin{aligned}\partial_t \gamma_{ij} = &- 2\alpha\gamma_{ik}K^k{}_j + \beta^k \partial_k \gamma_{ij} \\ &+ \gamma_{ik}\partial_j \beta^k + \gamma_{kj}\partial_i \beta^k\end{aligned} \qquad [37]$$

and the other set for $K^i{}_j$,

$$\begin{aligned}\partial_t K^i{}_j = &\beta^k \partial_k K^i{}_j - \partial_k \beta^i K^k{}_j + \partial_j \beta^k K^i{}_k - D^i D_j \alpha \\ &+ \alpha\left(\mathcal{R}^i{}_j + K K^i{}_j + 8\pi\left(\tfrac{1}{2}\delta^i{}_j(S-\rho) - S^i{}_j\right)\right)\end{aligned} \qquad [38]$$

As also noted previously, the evolution equations [37] for the spatial metric components, $\gamma_{ij}$, follow from the definition of the extrinsic curvature [31]. The derivation of the equations for the extrinsic curvature, on the other hand, require lengthy, but well-documented, manipulations of the spatial components of the field equations [1].

### The (Naive) Cauchy Problem

A naive statement of the Cauchy problem for $3+1$ numerical relativity is thus as follows: fix a specified number, $N$, of matter fields $\xi^A(t, x^k)$, $A = 1, 2, \ldots, N$, all minimally coupled to the gravitational field, with a total stress tensor, $T_{\mu\nu}$, given by

$$T_{\mu\nu} = \sum_{A=1}^{N} T^A_{\mu\nu} \qquad [39]$$

where $T^A_{\mu\nu}$ is the stress tensor corresponding to the matter field $\xi^A$. Choose a topology for $\Sigma(0)$ (e.g., $\mathcal{R}^3$ with asymptotically flat boundary conditions; $T^3$, with no boundaries, etc.) This also fixes the topology of $\mathcal{M}_\Sigma$ to be $\mathbb{R}\times$the topology of $\Sigma(0)$.

Next, freely specify eight of the 12 $\{\gamma_{ij}(0, x^k), K^i{}_j(0, x^k)\}$, as well as initial values, $\xi^A(0, x^k)$, for the matter fields. Then determine the remaining four dynamical gravitational fields from the constraints [35] and [36]. This completes the initial data specification.

One must now choose a prescription for the kinematical (coordinate) functions, $\alpha$ and $\beta^i$, so that either explicitly or implicitly, they are completely fixed; for the case of implicit specification, this may well mean that the coordinate functions themselves will satisfy PDEs, which, furthermore, can be of essentially any type in practice (i.e., elliptic, hyperbolic, parabolic, ...). Finally, with consistent initial data, $\{\gamma_{ij}(0, x^k), K^i{}_j(0, x^k); \xi_A(0, x^k)\}$, in hand, and with a prescription for the coordinate functions, the evolution

equations [37] and [38] can be used to advance the dynamical variables forward or backward in time.

The above description is naive since, apart from a consistent mathematical specification, the most crucial issue in the solution of a time-dependent PDE as a Cauchy problem is that the problem be "well posed." Roughly speaking, this means that solutions do not grow without bound ("blow-up") without physical cause, and that small, smooth changes to initial data yield correspondingly small, smooth changes to the evolved data. In short, the Cauchy problem must be stable, and whether or not a particular subset of the equations displayed in this section yields a well-posed problem is a complicated and delicate issue, especially in the generic 3D case. The reader is thus again cautioned against blind application of any of the equations displayed in this article.

### Boundary Conditions

In principle, because all spacelike hypersurfaces, $\Sigma(t)$, in a pure Cauchy evolution are edgeless – and provided that the initial data $\{\gamma_{ij}(0,x^k), K^i{}_j(0,x^k);\ \xi_A(0,x^k)\}$ is consistent with asymptotic flatness, or whatever other condition is appropriate given the topology of the $\Sigma(t)$ – there are essentially no boundary conditions to be imposed on the dynamical variables, $\{\gamma_{ij}(t,x^k), K^i{}_j(t,x^k)\}$, during Cauchy evolution. Note that asymptotic flatness generally requires that

$$\lim_{X\to\infty} \gamma_{ij} = f_{ij} + O\left(\frac{1}{X}\right) \qquad [40]$$

and

$$\lim_{X\to\infty} K^i{}_j = O\left(\frac{1}{X^2}\right) \qquad [41]$$

where $X$ is defined by

$$X \equiv \sqrt{\gamma_{ij}x^i x^j} \qquad [42]$$

as previously, and $f_{ij}$ is the flat 3-metric. Similarly, should the lapse, $\alpha$, and shift, $\beta$, be constrained by elliptic PDEs – as is frequently the case in practice – then the only natural place to set boundary conditions is at spatial infinity, and then, provided that the frame at spatial infinity is inertial, with coordinate time $t$ measuring proper time, one should have

$$\lim_{X\to\infty} \alpha = 1 + O\left(\frac{1}{X}\right) \qquad [43]$$

and

$$\lim_{X\to\infty} \beta^i = O\left(\frac{1}{X}\right) \qquad [44]$$

It is critical to note at this point, however, that in the vast bulk of past and current work in numerical relativity, including most of the ongoing work in 3D, the Einstein equations [1] have been solved, not as a pure Cauchy problem, but as a mixed initial-value/boundary-value (IBVP) problem. That is, in the discretization process in which the continuum equations [1] are replaced with algebraic equations, the continuum domain [6]–[7] is typically replaced with a truncated spatial domain

$$|x^i| \le X^i_{\max} \qquad [45]$$

where the $X^i_{\max}$ are *a priori* specified constants (parameters of the computational solution) that define the extremities of the "computational box." As one might expect, the theory underlying stability and well-posedness of IBVP problems – especially for differential systems as complicated as [1] – is even more involved than for the pure initial-value case, and is another very active area of research in both mathematical and numerical relativity (see, e.g., Friedrich and Nagy (1999)).

*See also:* Critical Phenomena in Gravitational Collapse; Einstein Equations: Initial Value Formulation; Fluid Mechanics: Numerical Methods; General Relativity: Overview; Geometric Analysis and General Relativity; Gravitational Waves; Hamiltonian Reduction of Einstein's Equations; Magnetohydrodynamics; Spacetime Topology, Causal Structure and Singularities; Symmetric Hyperbolic Systems and Shock Waves.

## Further Reading

Baumgarte T and Shapiro SL (2001) Numerical relativity and compact binaries. *Physics Reports* 376: 41–131.

Cook G (2000) Initial data for numerical relativity. *Living Reviews of Relativity* 3: 5 (irr-2000-5).

Font JA (2003) Numerical hydrodynamics in general relativity. *Living Reviews of Relativity* 6: 4 (irr-2003-4).

Frauendiener J (2004) Conformal infinity. *Living Reviews of Relativity* 7: 1 (irr-2004-1).

Friedrich H and Nagy G (1999) The initial boundary value problem for Einstein's vacuum field equation. *Communications in Mathematical Physics* 201: 619–655.

Hough J and Rowan S (2000) Gravitational wave detection by interferometry (ground and space). *Living Reviews of Relativity* 3: 3 (irr-2000-3).

Lehner L (2001) Numerical relativity: a review. *Classical and Quantum Gravity* 18: R25–R86.

Misner CW, Thorne KS, and Wheeler JA (1973) *Gravitation*. San Francisco: W.H. Freeman.

Reula OA (1998) Hyperbolic methods for Einstein's equations. *Living Reviews of Relativity* 1: 3 (irr-1998-3).

Winicour J (2001) Characteristic evolution and matching. *Living Reviews of Relativity* 4: 3 (irr-2001-3).

**Confinement** *see* Quantum Chromodynamics

**Conformal Geometry** *see* Two-dimensional Conformal Field Theory and Vertex Operator Algebras

**Conservation Laws** *see* Symmetries and Conservation Laws

# Constrained Systems

**M Henneaux**, Université Libre de Bruxelles, Brussels, Belgium



## Introduction

Consider a dynamical system with coordinates $q^i$ $(i = 1, \ldots, n)$ and Lagrangian $L(q^i, \dot{q}^i)$ (field theory is formally covered by regarding the spatial coordinates as a continuous index). When going to the Hamiltonian formulation, it is usually assumed that the Legendre transformation between the velocities $\dot{q}^i$ and the momenta

$$p_i = \frac{\partial L}{\partial \dot{q}^i} \qquad [1]$$

can be inverted to yield the velocities as functions of the $q$'s and the $p$'s. This "regular" situation occurs for most systems appearing in standard classical mechanics and enables one to proceed to the Hamiltonian formulation of the theory without difficulty.

In field theory, however, the regular case is the exception rather than the rule. This is due to gauge invariance and first-order Lagrangians.

- *Gauge invariance* A system possesses gauge symmetries if it is invariant under transformations that involve arbitrary functions of time (gauge transformations). In that case, the solution of the equations of motion with given initial data is not unique, since it is always possible to perform a gauge transformation in the course of the evolution without changing the initial data. It is then clear that the Legendre transformation cannot be invertible, for if it were, one could rewrite the equations of motion in the standard canonical form $\dot{q}^i = \partial H/\partial p_i, \dot{p}_i = -\partial H/\partial q^i$. These canonical equations are in normal form and have a unique solution for given initial data, which would contradict the presence of a gauge symmetry.

  A simple example that illustrates this phenomenon is given by the following model for three variables $q^1$, $q^2$, and $\lambda$, the Lagrangian of which reads

  $$L = \tfrac{1}{2}\left((\dot{q}^1 - \lambda)^2 + (\dot{q}^2 - \lambda)^2\right) \qquad [2]$$

  This model is inspired by electromagnetism: the variables $q^1$ and $q^2$ play a role somewhat similar to that of the spatial components of the vector potential, while $\lambda$ corresponds to the temporal component. The Lagrangian is invariant under the gauge transformations

  $$q^1 \to q^1 + \varepsilon, \qquad q^2 \to q^2 + \varepsilon, \qquad \lambda \to \lambda + \dot{\varepsilon} \qquad [3]$$

  where $\varepsilon$ is an arbitrary function of time. The conjugate momenta are

  $$p_1 = \dot{q}^1 - \lambda, \qquad p_2 = \dot{q}^2 - \lambda, \qquad \pi_\lambda = 0$$

  One cannot invert the Legendre transformation since one cannot express the velocity $\dot{\lambda}$ in terms of the momenta.
- *First-order Lagrangians* Fermionic fields obey first-order equations. Their Lagrangian is linear in the derivatives, so that the conjugate momenta $p_i$ depend on the coordinates $q^i$ only. It is then clearly impossible to express the velocities in terms of the momenta through the Legendre transformation. More generally, any first-order Lagrangian with or without gauge symmetry leads to a noninvertible Legendre transformation.

A simple system that exhibits this feature is described by the Lagrangian

$$L = z^2\dot{z}^1 - \frac{1}{2}(z^2)^2 \qquad [4]$$

for two bosonic degrees of freedom ($z^1$, $z^2$). This is in fact the canonical form of the Lagrangian for a free particle in one dimension ($z^2$ is the momentum conjugate to the position $z^1$): the system is already in Hamiltonian form. There is no gauge invariance, but because the Lagrangian is first order, the Legendre transformation with [4] as starting point,

$$p_1 = z^2, \qquad p_2 = 0 \qquad [5]$$

is non invertible for the velocities (which do not even appear in the formulas for the momenta).

Dirac showed how to develop the Hamiltonian formalism in the case when the Legendre transformation is not invertible. One can still reformulate the equations in phase space and write them in terms of brackets with the Hamiltonian, but a new major feature emerges, namely the canonical variables are no longer free. Rather, the permissible phase-space points are constrained to be on the so-called "constrained surface." For this reason, systems for which the Legendre transformation is not invertible are also called "constrained Hamiltonian systems." We shall adopt this terminology here.

The purpose of this article is to explain the main ideas underlying the Dirac method. To simplify the discussions and to focus on the features peculiar to the Dirac construction, we shall assume as a rule that all necessary smoothness conditions are fulfilled by the functions, surfaces, etc., appearing in the formalism. How to develop the analysis when some of the smoothness conditions are not fulfilled is of definite interest but goes beyond the scope of this review. We shall also assume, for definiteness, that all the variables are bosonic in order to avoid straightforward but somewhat cumbersome sign factors in the formulas.

## General Theory

### Dirac Algorithm

**Primary constraints** When the Legendre transformation [1] cannot be inverted, the momenta $p_i$'s do not span an $n$-dimensional space but are constrained by relations

$$\phi_m(q,p) = 0, \quad m = 1, \ldots, M \qquad [6]$$

which follow from their definition. These equations reduce to identities when the momenta are replaced by their expression [1] in terms of the coordinates and the velocities. They are called primary constraints. We shall assume that the matrix

$$\frac{\partial(\phi_m)}{\partial(p_i, q^i)}$$

is everywhere of constant (maximum) rank $M$ on the phase-space surface defined by eqns [6] which is assumed to be smooth. This surface is of dimension $2n - M$.

**Canonical Hamiltonian** The next step in the Dirac procedure is to define the canonical Hamiltonian $H$ through

$$H = \dot{q}^i p_i - L \qquad [7]$$

As shown by Dirac, $H$ can be re-expressed as a function $H(q, p)$ of the momenta and the coordinates, even when the Legendre transformation is not invertible: the canonical Hamiltonian $H$ depends on the velocities only through the $p_i$'s. Furthermore, the original equations of motion in Lagrangian form are equivalent to the Hamiltonian equations

$$\dot{q}^i = \frac{\partial H}{\partial p_i} + u^m \frac{\partial \phi_m}{\partial p_i} \qquad [8]$$

$$\dot{p}_i = -\frac{\partial H}{\partial q^i} - u^m \frac{\partial \phi_m}{\partial q^i} \qquad [9]$$

$$\phi_m(q, p) = 0 \qquad [10]$$

where the $u^m$'s are parameters, some of which will be determined through the consistency algorithm to be discussed shortly. (In [7]–[9] and everywhere below, there is a summation over the repeated indices.)

**Secondary constraints** The equations of motion [8] and [9] can be rewritten as

$$\dot{F} = [F, H] + u^m[F, \phi_m] \qquad [11]$$

where $F = F(q,p)$ is any function of the canonical variables. Here, the Poisson bracket is defined as usual by

$$[G, F] = \frac{\partial G}{\partial q^i}\frac{\partial F}{\partial p_i} - \frac{\partial G}{\partial p_i}\frac{\partial F}{\partial q^i} \qquad [12]$$

If one takes for $F$ one of the primary constraints $\phi_m$, one should get zero, $\dot{\phi}_m = 0$. This yields the consistency conditions

$$[\phi_m, H] + u^{m'}[\phi_m, \phi_{m'}] = 0 \qquad [13]$$

These conditions can imply further restrictions on the canonical variables and/or impose conditions on the

variables $u^m$. Any new relation $X(q,p)=0$ on the canonical variables leads, in turn, to a further consistency condition $\dot{X}=[X,H]+u^{m'}[X,\phi_{m'}]=0$, which can bring in either further restriction on the constraint surface or fix more variables $u^m$. Constraints that follow from the consistency algorithm are called "secondary constraints." Finally, one is left with a certain number of secondary constraints, which are denoted by $\phi_k=0$, $k=M+1,\ldots,M+K$. We assume again that all the constraints (primary and secondary) define a smooth surface, called the "constraint surface," and fulfill the condition that $\partial(\phi_k)/\partial(q^i,p_i)$ is of maximum rank $J\equiv M+K$ on the constraint surface. (We also assume for simplicity that there is no branching in the consistency algorithm.)

**Restrictions on the $u$'s** Having a complete set of constraints

$$\phi_j=0, \quad j=1,\ldots,M+K\equiv J \qquad [14]$$

we can now investigate more precisely the restrictions on the variables $u^m$. These read

$$[\phi_j,H]+u^m[\phi_j,\phi_m]\approx 0, \quad j=1,\ldots,J \qquad [15]$$

where the notation $\approx$ means "equal modulo the constraints." In [15], $m$ is summed from 1 to $M$. Equations [15] are a set of $J$ linear, inhomogeneous equations for the $u$'s, with coefficients that are functions of the canonical variables $q^i$, $p_i$. The general solution of this system is of the form

$$u^m=U^m+u^aV_a^m \qquad [16]$$

where $U^m$ is a particular solution and where the $V_a^m$ $(a=1,\ldots,A)$ provide a complete set of independent solutions of the homogeneous system

$$V_a^m[\phi_j,\phi_m]\approx 0 \qquad [17]$$

The coefficients $u^a (a=1,\ldots,A)$ are completely arbitrary.

We thus see the emergence of another new feature in the theory, in addition to the appearance of constraints. It is that the general solution of the equations of motion may contain arbitrary functions of time (when $A\neq 0$), in agreement with the possible presence of a gauge symmetry.

## First- and Second-Class Constraints

**First- and second-class functions** A function $F(q,p)$ is called a first-class function if it generates a canonical transformation that maps the constraint surface on itself. Thus, $F(q,p)$ is first class if its Poisson brackets with all the constraints vanish weakly (i.e., are zero on the constraint surface),

$$[F,\phi_j]\approx 0, \quad j=1,\ldots,J \qquad [18]$$

A function is second class otherwise, that is, if there is at least one constraint $\phi_j$ such that $[F,\phi_j]\neq 0$ (not even weakly). Second-class functions generate canonical transformations that do not leave the constraint surface invariant. Since canonical transformations that map the constraint surface on itself form a group, the Poisson bracket of two first-class functions is itself a first-class function.

Because the system is constrained to lie on the constraint surface, the only allowed canonical transformations are those that are generated by first-class functions. The importance of the distinction between first-class and second-class functions stems from this elementary fact. Note, in particular, that the time evolution is generated – as it should – by a first-class generator since the equations of motion [11] can be rewritten as

$$\dot{F}\approx[F,H']+u^a[F,V_a^m\phi_m] \qquad [19]$$

with

$$H'=H+U^m\phi_m \qquad [20]$$

One has both $[H',\phi_m]\approx 0$ and $[V_a^m\phi_m,\ \phi_j]\approx 0$.

**Splitting of the constraints** One can separate the constraints between first-class and second-class constraints. This can be achieved by considering the matrix $C_{jj'}$ of the Poisson bracket of the constraints,

$$C_{jj'}=[\phi_j,\phi_{j'}], \quad j,j'=1,\ldots,J \qquad [21]$$

One has the following theorem due to Dirac.

**Theorem 1** *If* $\det C_{jj'}\approx 0$, *there exists at least one first-class constraint among the* $\phi_j$*'s.*

**Proof** Straightforward: if $\det C_{jj'}\approx 0$, one can find a nontrivial solution $\lambda^j$ of $\lambda^jC_{jj'}\approx 0$. The corresponding constraint $\lambda^j\phi_j$ is easily verified to be first class.

By redefining the constraints as $\phi_j\rightarrow\bar{\phi}_j=a_j^{j'}\phi_{j'}$ with $a_j^{j'}(q,p)$ invertible, one can bring the Poisson brackets of the constraints to the form

$$[\gamma_a,\gamma_b]=0, \qquad [\gamma_a,\chi_\alpha]=0, \qquad [\chi_\alpha,\chi_\beta]=C_{\alpha\beta} \qquad [22]$$

with $(\bar{\phi}_j)\equiv(\gamma_a,\chi_\alpha)$ and where the matrix $C_{\alpha\beta}$ is invertible. (We assume, for simplicity, throughout that the rank of the matrix $C_{jj'}$ is constant on the constraint surface ("regular case").) In this representation, the constraints are completely split into first-class constraints $(\gamma_a)$ and second-class

constraints ($\chi_\alpha$): there is no first-class constraint left among the $\chi_\alpha$'s, and the set $\{\gamma_a\}$ exhausts all the first-class constraints. Note that now the index $a = 1, \ldots, A, A+1, \ldots, \bar{A}$ runs over all (primary and secondary) first-class constraints.

This separation of the constraints into first-class and second-class constraints is quite important because, as already seen above, the first-class constraints generate admissible canonical transformations, while the second-class constraints do not.

For a bosonic system, the matrix $C_{\alpha\beta}$ is antisymmetric. As $C_{\alpha\beta}$ is invertible, this implies that the number of second-class constraints is even. In the fermionic case, $C_{\alpha\beta}$ is symmetric (in the fermionic sector) and, therefore, the number of second-class constraints can be even or odd.

**First-class constraints and gauge symmetries** The first-class constraints not only map the constraint surface on itself, but generate, in fact, transformations that do not change the physical state of the system, that is, gauge transformations. Indeed, the presence of arbitrary functions in the solutions of the equations of motion indicates that the $q$'s and the $p$'s involve some redundancy and are not all physically distinct. Only those phase-space functions whose time evolution does not depend on the arbitrary functions $u^a$ are observables.

That the first-class constraints generate gauge transformations is rather clear in the case of the first-class primary constraints, since these appear explicitly in the generator of the time evolution multiplied by arbitrary functions. That it also holds for the first-class secondary constraints is known as the "Dirac conjecture." This conjecture can be proved under reasonable assumptions (see, e.g., Henneaux *et al.* 1990). The reason that the secondary first-class constraints also correspond to gauge transformations is that they appear in the brackets of the Hamiltonian with the primary first-class constraints. Thus, different choices of arbitrary functions $u^a$ in the dynamical equations of motion will lead to phase-space points that differ by a canonical transformation whose generator involves the secondary first-class constraints as well.

In any case, as noted below, one must identify the phase-space points in the same orbit generated by all the first-class constraints (primary and secondary) in order to get a reduced space with a symplectic structure ("reduced phase space"). For this reason, one postulates that the first-class constraints always generate gauge transformations, even for systems which are counterexamples to the Dirac conjecture (i.e., in that case, one defines the gauge transformations as being the transformations generated by the first-class constraints).

The extended Hamiltonian $H_E$ is defined to be the sum of the first-class Hamiltonian [20] and of all the first-class constraints $\gamma_a$ multiplied by an arbitrary Lagrange multiplier,

$$H_E = H' + v^a \gamma_a \qquad [23]$$

(with $a$ summed from 1 to $\bar{A}$). It is the generator of the time evolution in which the complete gauge symmetry is fully displayed.

**Elimination of second-class constraints – Dirac brackets** Second-class constraints do not generate permissible canonical transformations, since they do not map the constraint surface on itself. For this reason, it is convenient to eliminate them. This can consistently be done by using the Dirac brackets instead of the Poisson brackets. By definition, the Dirac bracket $[F, G]_D$ of two phase-space functions $F$ and $G$ is given by

$$[F, D]_D = [F, G] - [F, \chi_\alpha] C^{\alpha\beta} [\chi_\beta, G] \qquad [24]$$

where $C^{\alpha\beta}$ is the inverse to $C_{\alpha\beta}$,

$$C^{\alpha\beta} C_{\beta\gamma} = \delta^\alpha_\gamma$$

(which exists since the $\chi_\alpha$'s are second class). As shown by Dirac, the bracket [24] is indeed a bracket (antisymmetry, derivation property, and Jacobi identity). Furthermore, it fulfills the crucial property that the Dirac bracket of anything with any second-class constraint is zero,

$$[F, \chi_\alpha]_D = 0 \quad (F \text{ arbitrary}) \qquad [25]$$

Thus, one can consistently eliminate the second-class constraints and replace the Poisson bracket by the Dirac bracket. Once this is done, one has fewer canonical variables and only first-class constraints remain (if any). It also follows from the definition that the Dirac bracket of two first-class functions is equal to their Poisson bracket.

**Gauge conditions** One can push the reduction procedure further and eliminate the first-class constraints by means of gauge conditions. Gauge conditions $C_a = 0$ are conditions on the phase-space variables which do not follow from the Lagrangian and which have the property that they cut each gauge orbit once and only once. Since the gauge transformations are generated by the first-class constraints, this requirement is (locally) equivalent to

$$[C_a, \gamma_b] \varepsilon^b \approx 0 \Rightarrow \varepsilon^b \approx 0 \qquad [26]$$

That is, the constraints $(\gamma_a, C_b)$ form together a second-class system: there is no first-class constraint left once the conditions $C_a = 0$ are included. One can then eliminate all the constraints and gauge conditions and introduce the corresponding Dirac bracket. For gauge-invariant functions, this Dirac bracket coincides with the original Poisson bracket.

The reduced phase space is the unconstrained space obtained after this reduction, equipped with the Dirac bracket. It has dimension $2n - s - 2\bar{A}$, where $2n$ is the dimension of the original phase space, $s$ is the number of second-class constraints, and $\bar{A}$ is the number of first-class constraints. In the bosonic case, this number is even (as it should) because $s$ is even. One sees that "first-class constraints strike twice" since they need gauge conditions.

The observables of the theory are the reduced phase-space functions. They form a Poisson algebra, the relevant reduced phase-space bracket being the Dirac bracket associated with all the constraints and gauge conditions. The symplectic structure defined in the reduced phase space is nondegenerate because one has removed all the first-class constraints.

The definition of reduced phase space given above is useful in practice but has the conceptual drawback of relying on gauge conditions. This approach does not display clearly its intrinsic significance and, furthermore, in the case of the so-called Gribov problems (global obstructions to cutting each gauge orbit once and only once), may yield the incorrect expectation that the reduced phase space does not exist. We shall provide a more intrinsic definition below, which does not involve gauge conditions.

**Examples**

*First example* (see eqn [2]). There is here one primary constraint, namely $\pi_\lambda = 0$. The canonical Hamiltonian is $(1/2)((p_1)^2 + (p_2)^2) + \lambda(p_1 + p_2)$. The consistency algorithm yields the secondary constraint $p_1 + p_2 = 0$ and no condition on the $u$'s. The constraints are first class. They generate the gauge transformations $q^1 \to q^1 + \varepsilon$, $q^2 \to q^2 + \varepsilon$, and $\lambda \to \lambda + \eta$, which coincide with the Lagrangian gauge transformations if one identifies $\eta$ with $\dot{\varepsilon}$ ($\varepsilon$ and $\dot{\varepsilon}$ are, of course, independent at any given time). One can fix the gauge by means of the gauge conditions $\lambda = 0$, $q^1 + q^2 = 0$. The reduced phase space is two-dimensional and the observables can be identified with the functions of the gauge-invariant variables $(1/2)(q^1 - q^2)$ and $p_1 - p_2$, which are conjugate. Any other gauge condition leads to the same reduced phase space.

*Second example* (see eqn [4]). The primary constraints are $p_1 - z^2 = 0$ and $p_2 = 0$ and define a two-dimensional plane in the four-dimensional phase space $(z^1, z^2, p_1, p_2)$. The consistency algorithm forces $u^1 = z^2$ and $u^2 = 0$ and does not bring any further constraint. The constraints are second class since $[p_2, p_1 - z^2] = 1$. One can eliminate $p_1$ and $p_2$ through the constraints. The Dirac brackets of the remaining variables vanish, except $[z^1, z^2] = 1$. The reduced phase is the space of the $z$'s, with $z^2$ conjugate to $z^1$. The Hamiltonian is the free-particle Hamiltonian , $H = (1/2)(z^2)^2$. Thus, one recovers the original description which was already in Hamiltonian form. (The recognition that a system is already in first-order form often enables one to shortcut some aspects of the Dirac procedure by not introducing the unnecessary momenta which would in any case be eliminated in the end.)

## Quantization

The phase space of physical interest is the reduced phase space and the physical algebra is the algebra of the observables. The quantization of the theory then amounts to quantizing the algebra of the observables. This can be achieved along two different lines:

1. *Reduce then quantize*: In this direct approach, one represents as quantum operators only the reduced phase-space functions. There is no operator associated with non-gauge-invariant functions.
2. *Quantize then reduce*: In this approach, one represents as quantum operators the bigger algebra of functions of all the phase-space variables. One must then take into account the constraints. The second-class constraints are enforced as operator equations, which is consistent with the correspondence rule that the commutator in the quantum theory is $i\hbar$ times the Dirac bracket,

$$AB - BA = i\hbar [A, B]_{\mathrm{D}} \qquad [27]$$

(plus higher-order terms in $\hbar$). The first-class constraints are implemented in a more subtle way. It would be inconsistent to impose them as operator equations since in general $[\gamma_a, F]_{\mathrm{D}} \neq 0$ (even in the Dirac bracket). What one does is to impose them as conditions on the physical states: these are defined as the states annihilated by the first-class constraints,

$$\gamma_a |\psi\rangle = 0 \qquad [28]$$

For simple systems, it is easy to verify that the two procedures are equivalent. There is yet another

approach, in which one extends the system rather than reduce it. This is the Becchi–Rouet–Stora–Tyutin (BRST) approach, in which the new variables are called ghosts.

## Geometric Description

We defined above first-class and second-class constraints through algebraic means. It turns out that these definitions also have a geometrical interpretation, which sheds considerable insight into their nature.

The phase-space symplectic 2-form $\sigma$ induces, by pullback, a 2-form $\sigma_\Sigma$ on the constraint surface $\Sigma$. While $\sigma$ is of maximal rank, this may not be the case for the induced $\sigma_\Sigma$, which may be degenerate. In fact, the rank of $\sigma_\Sigma$ fails to be equal to the maximum rank $2n - J$ (where $J$ is the total number of constraints) by precisely the number $\bar{A}$ of first-class constraints.

Indeed, the Hamiltonian vector fields $X_{\gamma a}$ associated with the first-class constraints are tangent to the constraint surface $\Sigma$ and are null eigenvectors of $\sigma_\Sigma$,

$$\sigma_\Sigma(X_{\gamma a}, Y) = 0 \quad \forall Y \text{ tangent to } \Sigma \qquad [29]$$

as an immediate consequence of the first-class property. Here, all first-class constraints (primary and secondary) yield a null eigenvector. The integral surfaces of the vector fields $X_{\gamma a}$ are the gauge orbits. The reduced phase space is nothing else but the quotient space of the constraint surface by the gauge orbits. The 2-form induced in the quotient space is invertible because one has removed all degeneracy directions (including the ones associated with secondary first-class constraints). Reaching the reduced phase space falls under the scope of Hamiltonian reduction. The observables are the functions on the reduced phase space.

Thus, the reduced phase space is obtained through a two-step procedure. First, one restricts the functions to functions on the constraint surface $\Sigma$. One may view the algebra $C^\infty(\Sigma)$ of smooth functions on $\Sigma$ as the quotient algebra $C^\infty(P)/\mathcal{N}$ of the algebra $C^\infty(P)$ of smooth phase-space functions by the ideal $\mathcal{N}$ of phase-space functions that vanish on the constraint surface $\sigma$. The second step in the reduction procedure is to impose the gauge-invariant condition on the functions in $C^\infty(\Sigma)$, that is, to impose that they are constant along the gauge orbits $\mathcal{O}$. Assuming all necessary smoothness and regularity conditions to be fulfilled (i.e., that the orbits fiber which is, for instance, the case if the gauge orbits are the orbits of a free and proper group action), one may denote the algebra of observables as $C^\infty(\Sigma/\mathcal{O})$. This algebra is a Poisson algebra because the induced 2-form on the quotient space $\Sigma/\mathcal{O}$ is nondegenerate. The algebraic description of the observables underlies the BRST construction.

It is interesting to note that in the covariant approach to phase space, a similar two-step reduction procedure occurs. What plays the role of the constraint surface is the stationary surface in the space of all histories $q^i(t)$ of the dynamical variables. The gauge symmetry acts on this space and the reduced phase space is just the quotient space. One can establish the equivalence of the two descriptions (Barnich *et al.* 1991).

*See also:* Batalin–Vilkovisky Quantization; BRST Quantization; Canonical General Relativity; Operads; Perturbative Renormalization Theory and BRST; Quantum Dynamics in Loop Quantum Gravity; Quantum Field Theory: A Brief Introduction.

## Further Reading

Anderson JL and Bergmann PG (1951) Constraints in covariant field theories. *Physical Review* 83: 1018.

Barnich G, Henneaux M, and Schomblond C (1991) On the covariant description of the canonical formalism. *Physical Review D* 44: 939.

Dirac PAM (1950) Generalized Hamiltonian dynamics. *Canadian Journal of Mathematics* 2: 129.

Dirac PAM (1967) *Lectures on Quantum Mechanics*. New York: Academic Press.

Flato M, Lichnerowicz A, and Sternheimer D (1976) Deformations of Poisson brackets, Dirac brackets and applications. *Journal of Mathematical Physics* 17: 1754.

Hanson A, Regge T, and Teitelboim C (1976) *Constrained Hamiltonian Systems*. Rome: Accad. Naz. dei Lincei.

Henneaux M and Teitelboim C (1992) *Quantization of Gauge Systems*. Princeton: Princeton University Press.

Henneaux M, Teitelboim C, and Zanelli J (1990) Gauge invariance and degree of freedom count. *Nuclear Physics B* 332: 169.

Marsden JE and Weinstein A (1974) Reduction of symplectic manifolds with symmetry. *Reports on Mathematical Physics* 5: 121.

# Constructive Quantum Field Theory

**G Gallavotti**, Università di Roma "La Sapienza," Rome, Italy



## Euclidean Quantum Fields

The construction of a relativistic quantum field is still an open problem for fields in spacetime dimension $d \geq 4$. The conceptual difficulty that sometimes led to fear an incompatibility between nontrivial quantum systems and special relativity has however been solved in the case of dimension $d = 2, 3$ although, so far, has not influenced the corresponding debate on the foundations of quantum mechanics, still much alive.

It began in the early 1960s with Wightman's work on the axioms and the attempts at understanding the mathematical aspects of renormalization theory and with Hepps' renormalization theory for scalar fields. The breakthrough idea was, perhaps, Nelson's realization that the problem could really be studied in Euclidean form. A solution in dimensions $d = 2, 3$ has been obtained in the 1960s and 1970s through a remarkable series of papers by Nelson, Glimm, Jaffe, and Guerra. While the works of Nelson and Guerra relied on the "Euclidean approach" (see below) and on $d = 2$, the early works of Glimm and Jaffe dealt with $d = 3$ making use of the "Minkowskian approach" (based on second quantization) but making already use of a "multiscale analysis" technique. The latter received great impulsion and systematization by the adoption of Wilson's views and methods on renormalization: in physics terminology, renormalization group methods; a point of view taken here following the Euclidean approach. The solution dealt initially with scalar fields but it has been subsequently considerably extended.

The Euclidean approach studies quantum fields through the following problems:

1. existence of the functional integrals defining the generating functions (see below) of the probability distribution of the interacting fields in finite volume: the "ultraviolet stability problem,"
2. existence of the infinite-volume limit of the generating functions: the "infrared problem," and
3. check that the infinite volume generating functions satisfy the axioms needed to pass from the Euclidean, probabilistic, formulation to a Minkowskian formulation guaranteeing the existence of the Hamiltonian operator, relativistic covariance, Ruelle–Haag scattering theory: the "reconstruction problem."

The characteristic problem for the construction of quantum fields is (1) and here attention will be confined to it with the further restriction to the paradigmatic massive scalar fields cases. The dimension $d$ of the spacetime will be $d = 2, 3$ unless specified otherwise.

Given a cube $\Lambda$ of side $L$, $\Lambda \subset \mathbb{R}^d$, consider the following functional integral on the space of the fields on $\Lambda$, that is, on functions $\varphi_{\boldsymbol{\xi}}^{(\leq N)}$ defined for $\boldsymbol{\xi} \in \Lambda$,

$$Z_N(\Lambda, f) = \int \exp\left(-\int_\Lambda \left(\lambda_N \varphi_{\boldsymbol{\xi}}^{(\leq N)^4} + \mu_N \varphi_{\boldsymbol{\xi}}^{(\leq N)^2} + \nu_N + f_{\boldsymbol{\xi}} \varphi_{\boldsymbol{\xi}}^{(\leq N)}\right) \mathrm{d}\boldsymbol{\xi}\right) P_N(\mathrm{d}\varphi^{(\leq N)}) \qquad [1]$$

The fields $\varphi_{\boldsymbol{\xi}}^{(\leq N)}$ are called "Euclidean" fields with ultraviolet cutoff $N > 0$, $f_{\boldsymbol{\xi}}$ is a smooth function with compact support bounded by $|f_{\boldsymbol{\xi}}| \leq 1$ (for definiteness), the constants $\lambda_N > 0, \mu_N, \nu_N$ are called "bare couplings," and $P_N$ is a Gaussian probability distribution defining the free-field distribution with mass $m$ and ultraviolet cutoff $N$; the probability distribution $P_N$ is determined by its "covariance" $C_{\boldsymbol{\xi},\boldsymbol{\eta}}^{(\leq N)} \stackrel{\text{def}}{=} \int \varphi_{\boldsymbol{\xi}}^{(\leq N)} \varphi_{\boldsymbol{\eta}}^{(\leq N)} \, \mathrm{d}P_N$, which in the physics literature is called a "propagator," given by

$$C_{\boldsymbol{\xi},\boldsymbol{\eta}}^{(\leq N)} = \frac{1}{(2\pi)^d} \sum_{\boldsymbol{n} \in \mathbb{Z}^d} \int \frac{e^{i\boldsymbol{p}\cdot(\boldsymbol{\xi}-\boldsymbol{\eta}+\boldsymbol{n}L)}}{\boldsymbol{p}^2 + m^2} \chi_N(|\boldsymbol{p}|) \mathrm{d}^d \boldsymbol{p} \qquad [2]$$

The sum over the integers $\boldsymbol{n} \in \mathbb{Z}^d$ is introduced so that the field $\varphi_{\boldsymbol{\xi}}^{(\leq N)}$ is periodic over the box $\Lambda$: this is not really necessary as in the limit $L \to \infty$ either translation invariance would be recovered or lack of it properly understood, but it makes the problem more symmetric and generates a few technical simplifications; here $\chi_N(z)$ is a "regularizer" and a standard choice is

$$\chi_N(|\boldsymbol{p}|) = \frac{m^2(\gamma^{2N} - 1)}{\boldsymbol{p}^2 + \gamma^{2N} m^2}$$

with $\gamma > 1$, which is such that

$$\frac{\chi_N(|\boldsymbol{p}|)}{\boldsymbol{p}^2 + m^2} \equiv \frac{1}{\boldsymbol{p}^2 + m^2} - \frac{1}{\boldsymbol{p}^2 + \gamma^{2N} m^2} \equiv \sum_{h=1}^{N} \left(\frac{1}{\boldsymbol{p}^2 + \gamma^{2(h-1)} m^2} - \frac{1}{\boldsymbol{p}^2 + \gamma^{2h} m^2}\right) \qquad [3]$$

here $\gamma > 1$ can be chosen arbitrarily: so $\gamma = 2$. If $d > 3$, the above regularization will not be sufficient and a $\chi_N$ decaying faster than $\boldsymbol{p}^{-2}$ would be needed.

A simple estimate yields, if $\varepsilon \in (0,1)$ is fixed and $c$ is suitably chosen,

$$\left|C^{(\leq N)}_{\xi,\eta}\right| \leq c\gamma^{(d-2)N} e^{-m|\xi-\eta|}$$
$$\left|C^{(\leq N)}_{\xi,\eta} - C^{(\leq N)}_{\xi,\eta'}\right| \leq c\gamma^{(d-2)N}(\gamma^N m|\eta-\eta'|)^{\varepsilon} \quad [4]$$

with $\gamma^{(d-2)N}$ interpreted as $N$ if $d=2$.

The

$$\zeta(f) = \log\frac{Z_N(\Lambda,f)}{Z_N(\Lambda,0)}$$

defines a "generating function" of a probability distribution $P_{\text{int}}$ over the fields on $\Lambda$ which will be called the "distribution with $\varphi^4$-interaction" regularized on $\Lambda$ and at length scale $m^{-1}\gamma^{-N}$: the integral, in [1],

$$V_N\left(\varphi^{(\leq N)}\right) \overset{\text{def}}{=} -\int_\Lambda \left(\lambda_N \varphi_\xi^{(\leq N)^4} + \mu_N \varphi_\xi^{(\leq N)^2} + \nu_N + f_\xi \varphi_\xi^{(\leq N)}\right) d^d\xi \quad [5]$$

will be called the "interaction potential" with external field $f$. The regularization is introduced to guarantee that the integral [1], $\int e^{V_N}\, dP_N$, is well defined if $\lambda_N > 0$. The momenta of $P_{\text{int}}$ are the functional derivatives of $\zeta(f)$: they are called "Schwinger functions."

The problem (1) can now be made precise: it is to show the existence of $\lambda_N, \mu_N, \nu_N$ so that the limit

$$\lim_{N\to\infty} \frac{Z_N(\Lambda,f)}{Z_N(\Lambda,0)}$$

exists for all $f$ and is not Gaussian, that is, it is not the exponential of a quadratic form in $f$: which would be the case if $\lambda_N, \mu_N \to 0$ fast enough: the last requirement is of course essential because the Gaussian case describes, in the physical interpretation, free fields and noninteracting particles, that is, it is trivial. Note that $\nu_N$ does not play a role: its introduction is useful to be able to study separately the numerator and the denominator of the fraction

$$\frac{Z_N(\Lambda,f)}{Z_N(\Lambda,0)}$$

For more details, the reader is referred to Wightman and Gärding (1965), Streater and Wightman (1964), Nelson (1966), Guerra (1972), Osterwalder and Schrader (1973), and Simon (1974).

## The Regularized Free Field

Since the propagator, see [4], decays exponentially over a scale $m^{-1}$ and is smooth over a scale $m^{-1}\gamma^{-N}$, the fields $\varphi_\xi^{(\leq N)}$ sampled with distribution $P_N$ are rather singular objects. Their properties cannot be described by a single length scale: they are extremely large for large $N$, take independent values only beyond distances of order $m^{-1}$ but, at the same time, they look smooth only on the much smaller scale $m^{-1}\gamma^{-N}$. Their essential feature is that fixed $\varepsilon < 1$, for example, $\varepsilon = 1/2$, with $P_N$-probability 1 there is $B > 0$ such that (interpreting $\gamma^{(d-2)/2N}$ as $N$ if $d=2$)

$$\left|\varphi_\xi^{(\leq N)}\right| \leq B\gamma^{N(d-2)/2}$$
$$\left|\varphi_\xi^{(\leq N)} - \varphi_\eta^{(\leq N)}\right| < B\gamma^{N(d-2)/2}\left(\gamma^N m|\xi-\eta|\right)^{\varepsilon/2} \quad [6]$$

and furthermore the probability of the relations in [6] will be $N$-independent, that is, $\varphi_\xi^{(\leq N)}$ are bounded and roughly of size $\gamma^{N(d-2)/2}$ as $N \to \infty$ and, on a very small length scale $m^{-1}\gamma^{-N}$, almost constant.

Substantial control on the field $\varphi_\xi^{(\leq N)}$ statistically sampled with distribution $P_N$ can be obtained by decomposing it, through [3], into "components of various scales": that is, as a sum of statistically mutually independent fields whose properties are entirely characterized by a single scale of length. This means that they have size of order 1 and are independent and smooth on the same length scale.

Assuming the side of $\Lambda$ to be an integer multiple of $m^{-1}$, let $\mathcal{Q}_h$ be a pavement of $\Lambda$ into boxes of side $m^{-1}\gamma^{-h}$, imagined hierarchically arranged so that the boxes of $\mathcal{Q}_h$ are exactly paved by those of $\mathcal{Q}_{h+1}$.

Define $z_\xi^{(h)}$ to be the random field with propagator $C_{\xi,\eta}^{(h)}$ with Fourier transform

$$\sum_{n\in\mathbb{Z}^d}\left(\frac{1}{p^2+\gamma^{-2}m^2} - \frac{1}{p^2+m^2}\right) e^{in\cdot p L\gamma^h}$$

so that $\varphi_\xi^{(\leq N)}$ and its propagator $C_{\xi,\eta}^{(\leq N)}$ can be represented, see [2], [3], as

$$\varphi_\xi^{(\leq N)} \equiv \sum_{h=1}^N \gamma^{h(d-2)/2} z^{(h)}_{\gamma^h\xi}$$
$$C_{\xi,\eta}^{(\leq N)} = \sum_{h=1}^N \gamma^{h(d-2)} C^{(h)}_{\gamma^h\xi,\gamma^h\eta} \quad [7]$$

where the fields $z^{(h)}$ are independently distributed Gaussian fields. Note that the fields $z^{(h)}$ are also almost identically distributed because their propagator is obtained by periodizing over the period $\gamma^h L$ the same function

$$\overline{C}^{(0)}_{\xi,\eta} \overset{\text{def}}{=} \int \frac{e^{ip\cdot(\xi-\eta)}dp}{(2\pi)^d}\left(\frac{1}{p^2+\gamma^{-2}m^2} - \frac{1}{p^2+m^2}\right)$$

that is, their propagator is

$$C^{(h)}_{\xi,\eta} = \sum_{n\in\mathbb{Z}^d} \overline{C}^{(0)}_{\xi,\eta+\gamma^h nL}$$

The reason why they are not exactly equally distributed is that the field $z^{(h)}_{\xi}$ is periodic with period $\gamma^h L$ rather than $L$. But proceeding with care the sum over $n$ in the above expressions can be essentially ignored: this is a little price to pay if one wants translation invariance built in the analysis since the beginning.

The representation [7] defines a "multiscale representation" of the field $\varphi^{(\leq N)}_{\xi}$. Smoothness properties for the field $\varphi^{(\leq N)}_{\xi}$ can be read from those of its "components" $z^{(h)}$. Define, for $\Delta\in\mathcal{Q}_0$,

$$\left\|z^{(h)}\right\|_{\Delta} = \max_{\substack{\xi\in\Delta,\eta\in\Lambda\\ |\xi-\eta|\leq m^{-1}}} \left( \left|z^{(h)}_{\xi}\right| + \tau \frac{\left|z^{(h)}_{\xi} - z^{(h)}_{\eta}\right|}{|\xi-\eta|^{1/4}} \right) \qquad [8]$$

and $\tau$ will be chosen $\tau=0$ or $\tau=1$ as needed (in practice $\tau=0$ if $d=2$ and $\tau=1$ if $d=3$): $\tau=1$ will allow us to discuss some smoothness properties of the fields which will be necessary (e.g., if $d=3$). Then the size $\|z\|_{\Delta}$ of any field $z^{(h)}$, for all $h\geq 1$, is estimated by

$$P\left(\max_{\Delta\subset\mathcal{Q}_0} \|z\|_{\Delta} \leq B\right) \geq \mathrm{e}^{-c\mathrm{e}^{-c'B^2}|\Lambda|}$$
$$P(\|z\|_{\Delta} \geq B_{\Delta}, \forall\Delta\in\mathcal{D}) \leq \prod_{\Delta\in\mathcal{D}} c\mathrm{e}^{-c'B^2_{\Delta}} \qquad [9]$$

where $P$ is the Gaussian probability distribution of $z$, $\mathcal{D}$ is any collection of boxes $\Delta\in\mathcal{Q}_0$ and $c, c'>0$ are suitable constants. The [9] imply in particular [6]. The estimates [9] follow from the Markovian nature of the Gaussian field $z^{(h)}$, that is, from the fact that the propagator is the Green's function of an elliptic operator (of fourth order, see the first of [3]), with constant coefficients which implies also the inequalities (fixing $\varepsilon\in(0,1)$)

$$\left|C^{(h)}_{\xi,\eta}\right| \equiv \left|\int z_{\xi} z_{\eta} P(\mathrm{d}z)\right| \leq c\mathrm{e}^{-mc'|\xi-\eta|}$$
$$\left|C^{(h)}_{\xi,\eta} - C^{(h)}_{\xi,\eta'}\right| \leq c(m|\eta-\eta'|)^{\varepsilon} \qquad [10]$$

where $|\xi-\eta|$ is reinterpreted as the distance between $\xi,\eta$ measured over the periodic box $\gamma^h\Lambda$ (hence $|\xi-\eta|$ differs from the ordinary distance only if the latter is of the order of $\gamma^h L$). The interpretation of [10] is that $z^{(h)}_{\xi}$ are essentially bounded variables which, on scale $\sim m^{-1}$, are essentially constant and furthermore beyond length $\sim m^{-1}$ are essentially independently distributed.

For more details, the reader is referred to Wilson (1970, 1972) and Gallavotti (1981, 1985).

## Perturbation Theory

The naive approach to the problem is to fix $\lambda_N\equiv\lambda>0$ and to develop $Z_N(\Lambda,f)$ or, more conveniently and equivalently, $(1/|\Lambda|)\log Z_N(\Lambda,f)$ in powers of $\lambda$. If one fixes *a priori* $\mu_N,\nu_N$ independent of $N$, however, even a formal power series is not possible: this is trivially due to the divergence of the coefficients of the power series, already to second order, for generic $f$ in the limit $N\to\infty$. Nevertheless it is possible to determine $\mu_N(\lambda),\nu_N(\lambda)$ as functions of $N$ and $\lambda$ so that a formal power series exists (to all orders in $\lambda$): this is the key result of renormalization theory.

To find the perturbative expansion, the simplest is to use a graphical representation of the coefficients of the power expansion in $\lambda,\mu_N,\nu_N,f$ and the Gaussian integration rules which yield (after a classical computation) that the coefficient of $\lambda^n\mu^p_N f_{\xi_1}\ldots f_{\xi_r}$ is obtained by considering the graph elements shown in **Figure 1**, where the segments will be called half-lines and the graph elements will be called, respectively, "coupling" or "$\varphi^4$-vertex," "mass vertex," "vacuum vertex," and "external vertex."

The half-lines of the graph elements are considered distinct (i.e., imagine a label attached to distinguish them). Then consider all possible *connected* graphs $G$ obtained by first drawing, respectively, $n,p,r$ graph elements in **Figure 1**, which are not vacuum vertices, with their nodes marked by points in $\Lambda$ named $\xi_1,\ldots,\xi_n$, $\xi_{n+1},\ldots,\xi_{n+p+r}$; and form all possible graphs obtained by attaching pairs of half-lines emerging from the vertices of the graph elements. These are the "nontrivial graphs." Furthermore, consider also the single "trivial" graph formed just by the third graph element and consisting of a single point. All graphs obtained in this way are particular Feynman graphs.

Given a nontrivial graph $G$ (there are many of them) we define its value to be the product

$$W_G(\xi_1,\ldots,\xi_n,\xi_{n+1},\ldots,\xi_{n+p+r})$$
$$= (-1)^{n+p+r}\frac{\lambda^n\mu^p_N\prod f_{\xi_{n+p+j}}}{n!p!r!}\prod_{\ell} C^{(\leq N)}_{\xi_\ell,\eta_\ell} \qquad [11]$$

where the last product runs over all pairs $\ell=(\xi_\ell,\eta_\ell)$ of half-lines of $G$ that are joined and connect two vertices labeled by points $\xi_\ell,\eta_\ell$: "call line of $G$" any such pair. If the graph consists of the single vacuum

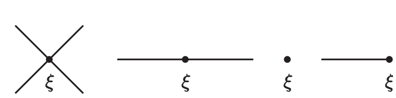



**Figure 1** The graph elements to representing $\varphi^{(\leq N)4}_{\xi}$, $\varphi^{(\leq N)2}$, a constant $\varphi^{(\leq N)}$.

vertex its value will be $\nu_N$. The series for $(1/|\Lambda|)\log Z_N(\Lambda, f)$ is then

$$-\nu_N + \frac{1}{|\Lambda|}\sum_G \int W_G(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n+p+r}) \prod_{j=1}^{n+p+r} \mathrm{d}\boldsymbol{\xi}_j \qquad [12]$$

and the integral will be called the integrated graph value.

Suppose first that $\mu_N = \nu_N = 0$. Then if a graph $G$ contains subgraphs like in Figure 2, the corresponding respective contribution to the integral in [12] (considering only the integrals over $\boldsymbol{\eta}$ and suitably taking care of the combinatorial factors) is a factor obtained by integrating over $\boldsymbol{\xi}$ the quantities

$$-6\lambda C^{(\leq N)}_{\alpha\xi} C^{(\leq N)}_{\xi\xi} C^{(\leq N)}_{\xi\beta}$$

$$\text{or} \quad \frac{4^2 \cdot 3!}{2!}\lambda^2 C^{(\leq N)}_{\alpha\xi} \int C^{(\leq N)3}_{\xi\eta} C^{(\leq N)}_{\eta\beta} \mathrm{d}\boldsymbol{\eta} \qquad [13]$$

which if $d = 3$ diverge as $N \to \infty$ as $\gamma^N$ or, respectively, as $N$; the second factor does not diverge in dimension $d = 2$ while the first still diverges as $N$. The divergences arise from the fact that as $\boldsymbol{\xi} - \boldsymbol{\eta} \to 0$ the propagator behaves as $|\boldsymbol{\xi} - \boldsymbol{\eta}|^{-N}$ if $d = 3$ or as $-\log|\boldsymbol{\xi} - \boldsymbol{\eta}|$ if $d = 2$, all the way until saturation occurs at distance $|\boldsymbol{\xi} - \boldsymbol{\eta}| \simeq m^{-1}\gamma^{-N}$: for this reason the latter divergences are called "ultraviolet divergences."

However, if we set $\mu_N \neq 0$, then for every graph containing a subgraph like those in Figure 2 there is another one identical except that the points $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are connected via a mass vertex, see Figure 1, with the vertex in $\boldsymbol{\xi}$, by a line $\boldsymbol{\alpha\xi}$ and a line $\boldsymbol{\xi\beta}$; the new graph value receives a contribution from the mass vertex inserted in $\boldsymbol{\xi}$ between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ simply given by a factor $-\mu_N$. Therefore if we fix, for $d = 3$,

$$\mu_N = -6\lambda C^{(\leq N)}_{\xi\xi} + \frac{4^2 \cdot 3!}{2}\lambda^2 \times \int_\Lambda C^{(\leq N)3}_{\xi\eta} \mathrm{d}\boldsymbol{\eta} \stackrel{\text{def}}{=} -6\lambda C^{(\leq N)}_{\xi\xi} + \delta\mu_N \qquad [14]$$

we can simply consider graphs which do not contain any mass graph element and in which there are no subgraphs like the first in Figure 2 while the subgraphs like the second in Figure 2 do not contribute a factor $\int C^{(\leq N)}_{\alpha\xi} C^{(\leq N)3}_{\xi\eta} C^{(\leq N)}_{\eta\beta} \mathrm{d}\boldsymbol{\eta}$ but a renormalized factor $\int C^{(\leq N)}_{\alpha\xi} C^{(\leq N)3}_{\xi\eta} (C^{(\leq N)}_{\eta\beta} - C^{(\leq N)}_{\xi\beta}) \mathrm{d}\boldsymbol{\eta}$. If $d = 2$, we only need to define $\mu_N$ as the first term on the right-hand side (RHS) of [14] and we can leave the subgraphs like the second in Figure 2 as they are (without any renormalization).

Graphs without external lines are called vacuum graphs and there are a few such graphs which are divergent. Namely, if $d = 3$, they are the first three drawn in Figure 3; furthermore, if $\mu_N$ is set to the above nonzero value a new vacuum graph, the fourth in Figure 3, can be formed. Such graphs contribute to the graph value, respectively, the terms in the sum

$$-3\lambda C^{(\leq N)2}_{\xi_1,\xi_1} + \frac{4!}{2}\lambda^2 \int C^{(\leq N)4}_{\xi_1\xi_2} \mathrm{d}\boldsymbol{\xi}_2 - \frac{2^3 \cdot 3!^3}{3!}\lambda^3 \times \int C^{(\leq N)2}_{\xi_1\xi_2} C^{(\leq N)2}_{\xi_2\xi_3} C^{(\leq N)2}_{\xi_3\xi_1} \mathrm{d}\boldsymbol{\xi}_2 \mathrm{d}\boldsymbol{\xi}_3 - \mu_N C^{(\leq N)}_{\xi_1\xi_1} \qquad [15]$$

and diverge, respectively, as $\gamma^{2N}, \gamma^N, N, \gamma^{2N}$ if $d = 3$ while, if $d = 2$, only the first and the last (see [14]) diverge, like $N^2$.

Therefore, if we fix $\nu_N$ as minus the quantity in [15] we can disregard graphs like those in Figure 3; if $d = 2$ $\nu_N$ can be defined to be the sum of the first and last terms in [15].

The formal series in $\lambda$ and $f$ thus obtained is called the "renormalized series" for the field $\varphi^4$ in dimension $d = 2$ or, respectively, $d = 3$. Note that with the given definitions and choices of $\mu_N, \nu_N$ the only graphs $G$ that need to be considered to construct the expansion in $\lambda$ and $f$ are formed by the first and last graph elements in Figure 1, paying attention that the graphs in Figure 3 do not contribute and, if $d = 3$, the graphs with subgraphs like the second in Figure 2 have to be computed with the modification described.

In the next section, it will be shown that the above are the only sources of divergences as $N \to \infty$ and therefore the problem of studying [1] is solved at the level of formal power series by the subtraction in [14]. This also shows that giving a meaning to the series thus obtained is likely to be much easier if $d = 2$ than if $d = 3$.

The coefficients of order $k$ of the expansion in $\lambda$ of $(1/|\Lambda|)\log Z_N(\Lambda, f)$ can be ordered by the number $2n$ of vertices representing external fields: and have the form $\int S^{(k)}_{2n}(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{2n}) \prod_{i=1}^{2n} (f_{\xi_i} \mathrm{d}\boldsymbol{\xi}_i)$: the kernels $S^{(k)}_{2n}$ are the Schwinger functions of order $2n$, see the section "Euclidean quantum fields."

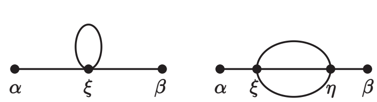



**Figure 2** Divergent subgraphs, if $d = 3$. If $d = 2$ only the first diverges.

**Figure 3** Divergent vacuum graphs.

**Remark** If $d=4$, the regularization at cutoff $N$ in [2] is not sufficient as in the subtraction procedure smoothness of the first derivatives of the field $\varphi^{(\leq N)}$ is necessary, while the regularization [2] does not even imply [6], that is, not even Hölder continuity. A higher regularization (i.e., using a $\chi_N$ like the square of the $\chi_N$ in [3]). Furthermore, the subtractions discussed in the case $d=3$ are not sufficient to generate a formal power series and many more subtractions are needed: for instance, graphs with a subgraph like the one in Figure 4 would give a contribution to the graph value which is a factor

$$\lambda^2 \ell_N \stackrel{\text{def}}{=} \frac{2\cdot 6^2}{2!}\lambda^2 \int_\Lambda C_{\xi\eta}^{(\leq N)2} \mathrm{d}\eta$$

also divergent as $N\to\infty$ proportionally to $N$. Although this divergence could be canceled by changing $\lambda$ into $\lambda_N = \lambda + \lambda^2 \ell_N$ the previously discussed cancelations would be affected and a change in the value of $\mu_N$ would become necessary; furthermore, the subtraction in [14] will not be sufficient to make finite the graphs, not even to second order in $\lambda$, unless a new term $-\alpha_N \int (\partial_\xi \varphi_\xi^{(\leq N)})^2 \,\mathrm{d}\boldsymbol{\xi}$ with $\alpha_N = (1/2)\lambda^2 \int \partial_\eta C_{\xi\eta}^{(\leq N)3} (\boldsymbol{\xi}-\boldsymbol{\eta})^2$ is added in the exponential in [1].

But all this will not be enough and still new divergences, proportional to $\lambda^3$, will appear.

And so on indefinitely, the consequence being that it will be necessary to define $\lambda_N, \mu_N, \alpha_N, \nu_N$ as formal power series in $\lambda$ (with coefficients diverging as $N\to\infty$) in order to obtain a formal power series in $\lambda$ for [1] in which all coefficients have a finite limit as $N\to\infty$. Thus, the interpretation of the formal renormalized series in the case $d=4$ is substantially different and naturally harder than the cases $d=2,3$. Beyond formal perturbation expansions, the case $d=4$ is still an open problem: the most widespread conjecture is that the series cannot be given a meaning other than setting to 0 all coefficients of $\lambda^j, j>0$. In other words, the conjecture claims, there should be no nontrivial solution to the ultraviolet problem for scalar $\varphi^4$ fields in $d=4$. But this is far from being proved, even at a heuristic level. The situation is simpler if $d\geq 5$: in such cases, it is impossible to find formal power series in $\lambda$ for $(1/|\Lambda|)\log Z_N(\Lambda, f)$, even allowing $\lambda_N, \mu_N, \alpha_N, \nu_N$ to be formal power series in $\lambda$ with divergent coefficients.

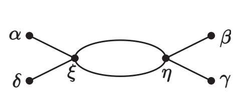



**Figure 4** The simplest new divergent subgraph on $d=4$.

The distinctions between the cases $d=2,3,4,>4$ explain the terminology given to the $\varphi^4$-scalar field theories calling them super-renormalizable if $d=2,3$, renormalizable if $d=4$ and nonrenormalizable if $d>4$. Since the (divergent) coefficients in the formal power series defining $\lambda_N, \mu_N, \alpha_N, \nu_N$ are called counter-terms, the $\varphi^4$-scalar fields require finitely many counter-terms (see [14]) in the super-renormalizable cases and infinitely many in the renormalizable case. The nonrenormalizable cases ($d>4$) cannot be treated in a way analogous to the renormalizable ones.

For more details, the reader is referred to Gallavotti (1985), Aizenman (1982), and Fröhlich (1982).

## Finiteness of the Renormalized Series, $d=2,3$: "Power Counting"

Checking that the renormalized series is well defined to all orders is a simple dimensional estimate characteristic of many multiscale arguments that in physics have become familiar with the name of "renormalization group arguments."

Consider a graph $G$ with $n+r$ vertices built over $n$ graph elements with vertices $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n$ each with four half-lines and $r$ graph elements with vertices $\boldsymbol{\xi}_{n+1}, \ldots, \boldsymbol{\xi}_{n+r}$ representing the external fields: as remarked in the previous section, these are the only graphs to be considered to form the renormalized series.

Develop each propagator into a sum of propagators as in [7]. The graph $G$ value will, as a consequence, be represented as a sum of values of new graphs obtained from $G$ by adding scale labels on its lines and the value of the graph will be computed as a product of factors in which a line joining $\boldsymbol{\xi\eta}$ and bearing a scale label $h$ will contribute with $C_{\xi\eta}^{(h)}$ replacing $C_{\xi\eta}^{(\leq N)}$. To avoid proliferation of symbols, we shall call the graphs obtained in this way, i.e., with the scale labels attached to each line, still $G$: no confusion should arise as we shall, henceforth, only consider graphs $G$ with each line carrying also a scale label.

The scale labels added on the lines of the graph $G$ allow us to organize the vertices of $G$ into "clusters": a cluster of scale $h$ consists in a maximal set of vertices (of the graph elements in the graph) connected by lines of scale $h'\geq h$ among which one at least has scale $h$.

It is convenient to consider the vertices of the graph elements as "trivial" clusters of highest scale: conventionally call them clusters of scale $N+1$.

The clusters can be of "first generation" if they contain only trivial clusters, of "second generation"

if they contain only clusters which are trivial or of the first generation, and so on.

Imagine to enclose in a box the vertices of graph elements inside a cluster of the first generation and then into a larger box the vertices of the clusters of the second generation and so on: the set of boxes ordered by inclusion can then be represented by a rooted tree graph whose nodes correspond to the clusters and whose "top points" are nodes representing the trivial clusters (i.e., the vertices of the graph).

If the maximum number of nodes that have to be crossed to reach a top point of the tree starting from a node $v$ is $n_v$ ($v$ included and the top nodes included), then the node $v$ represents a cluster of the $n_v$th generation. The first node before the root is a cluster containing all vertices of $G$ and the root of the tree will not be considered a node and it can conventionally bear the scale label 0: it represents symbolically the value of the graph.

For instance, in Figure 5 a tree $\theta$ is drawn: its nodes correspond to clusters whose scale is indicated next to them; in the second part of the drawing, the trivial clusters as well as the clusters of the first generation are enclosed into boxes.

Then consider the next generation clusters, that is, the clusters which only contain clusters of the first generation or trivial ones, and draw boxes enclosing all the graph vertices that can be reached from each of them by descending the tree, etc. Figure 6 represents all boxes (of any generation) corresponding to the nodes of the tree in Figure 5. The representations of the clusters of a graph $G$ by a tree or by hierarchically ordered boxes (see Figures 5 and 6) are completely equivalent provided inside each box not representing a top point of the tree the scale $h_v$ of the corresponding cluster $v$ is marked. For

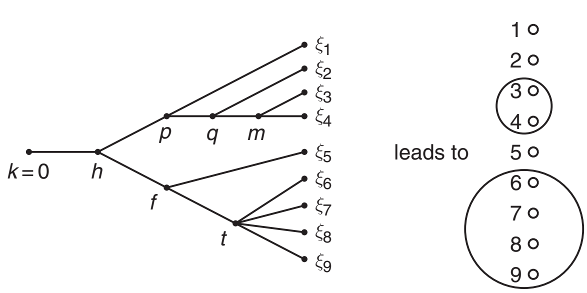



**Figure 5** A tree and its clusters of generation 1 and 2.

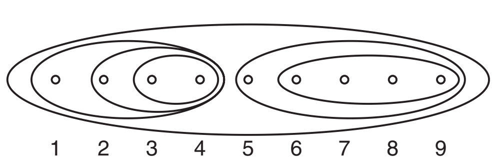



**Figure 6** All clusters of any generation for the tree in Figure 5.

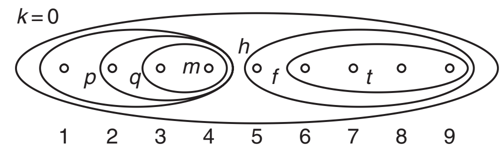



**Figure 7** The clusters in Figure 6 after affixing the scale labels.

instance, in the case of Figure 6 one gets Figure 7. By construction, if two top points $\xi$ and $\eta$ are inside the same box $b_v$ of scale $h_v$ but not in inner boxes, then there is a path of graph lines joining $\xi$ and $\eta$ all of which have scales $\geq h_v$ and one at least has scale $h_v$.

Given a graph $G$, fix one of its points $\xi_1$ (say) and integrate the absolute value of the graph over the positions of the remaining points. The exponential decay of the propagators implies that if a point $\eta$ is linked to a point $\eta'$ by a line of scale $h$ the integration over the position of $\eta'$ is essentially constrained to extend only over a distance $\gamma^{-h}m^{-1}$. Furthermore, the maximum size of the propagator associated with a line of scale $h$ is bounded proportionally to $\gamma^{(d-2)h}$. Therefore, recalling that $|f_\xi|$ is suppose bounded by 1, the mentioned integral can be immediately bounded by

$$\frac{\lambda^n}{n!r!}C^{n+r}I \stackrel{\text{def}}{=} \frac{\lambda^n C^{n+r}}{n!r!}\prod_{\ell}\gamma^{(d-2)/2h_\ell}\prod_{v}\gamma^{-\mathrm{d}h_v(s_v-1)} \qquad [16]$$

where, $C$ being a suitable constant, the first product is over the half-lines $\ell$ composing the graph lines and the second is over the tree nodes (i.e., over the clusters of the graph $G$), $s_v$ is the number of subclusters contained in the cluster $v$ but not in inner clusters; and in [16] the scale of a half-line $\ell$ is $h_\ell$ if $\ell$ is paired with another half-line to form a line $\ell$ (in the graph $G$) of scale label $h_\ell$.

Denoting by $v'$ the cluster immediately containing $v$ in $G$, by $n_v^{\text{inner}}$ the number of half-lines in the cluster $v$, by $n_v, r_v$ the numbers of graph elements of the first type or of the fourth type in Figure 1 with vertices in the cluster $v$, and denoting by $n_v^e$ the number of lines which are not in the cluster $v$ but have one extreme on a vertex in $v$ ("lines external to $v$"), the identities ($k=0$)

$$\begin{aligned}&\sum_{v>\text{root}}(h_v-k)(s_v-1)\\&\quad\equiv\sum_{v>\text{root}}(h_v-h_{v'})(n_v+r_v-1)\\&\sum_{v>\text{root}}(h_v-k)n_v^{\text{inner}}\equiv\sum_{v>\text{root}}(h_v-h_{v'})\widetilde{n}_v^{\text{inner}}\end{aligned} \qquad [17]$$

with

$$\widetilde{n}_v^{\text{inner}} \stackrel{\text{def}}{=} 4n_v + r_v - n_v^e$$

hold, so that the estimate [16] can be elaborated into

$$I \le \prod_{v>r} \gamma^{-\rho_v(h_v - h_{v'})}$$
$$\rho_v \overset{\text{def}}{=} -d + (4-d)n_v + r_v\frac{d+2}{2} + \frac{d-2}{2}n_v^e \qquad [18]$$

where $h_{v'} = k = 0$ if $v$ is the first nontrivial node (i.e., $v' = \text{root}$), and an estimate of the integral of the absolute value of the graphs $G$ with given tree structure but different scale labels is proportional to $\Sigma_{\{h_v\}} I < \infty$ if (and only if) $\rho_v > 0, \forall v$.

But there may be clusters $v$ with only two external lines $n_v^e = 2$ and two graph vertices inside: for which $\rho_v = 0$. However, this can happen only if $d = 3$ and in only one case: namely if the graph $G$ contains a subgraph of the second type in Figure 2 and the three intermediate lines form a cluster $v$ of scale $h_v$ while the other two lines are external to it: hence on scale $h' > h_v$. In this case, one has to remember that the subtraction in the previous section has led to a modification of the contribution of such a subgraph to the value of the graph (integrated over the position labels of the vertices). As discussed in the previous section, the change amounts to replacing the propagator $C^{(h')}_{\eta,\beta}$ by $C^{(h')}_{\eta,\beta} - C^{(h')}_{\xi,\beta}$.

This improves, in [18], the estimate of the contribution of the line joining $\boldsymbol{\eta}$ to $\boldsymbol{\beta}$ from being proportional to $\int C^{(\le h_v)3}_{\xi\eta} C^{(\le h')}_{\eta\beta}\, d\boldsymbol{\eta}$ to being proportional to $\int C^{(\le h_v)3}_{\xi\eta}(C^{(\le h')}_{\eta\beta} - C^{(\le h')}_{\xi\beta})\, d\boldsymbol{\eta}$; and this changes the contribution of the line $\boldsymbol{\eta\beta}$ from $\gamma^{(d-2)h'}$ to $\int e^{-m\gamma^{h_v}|\boldsymbol{\xi}-\boldsymbol{\eta}|} (\gamma^{h'}|\boldsymbol{\xi}-\boldsymbol{\eta}|)^{1/2}\, d\boldsymbol{\eta}$ because $C^{(h')}$ is regular on scale $\gamma^{-h'}m^{-1}$, see [10] with $\varepsilon = 1/2$.

Since $\boldsymbol{\xi}, \boldsymbol{\eta}$ are in a cluster of higher scale $h_v$ this means that the estimate is improved by $\gamma^{-(1/2)(h_v - h')}$. In terms of the final estimate, this means that $\rho_v$ in [18] can be improved to $\overline{\rho}_v = \rho_v + 1/2$ for the clusters for which $\rho_v = 0$. Hence, the integrated value of the graph $G$ (after taking also into account the integration over the initially selected vertex $\boldsymbol{\xi}_1$, trivially giving a further factor $|\Lambda|$ by translation invariance), and summed over the possible scale labels is bounded proportionally to $|\Lambda|\Sigma_{\{h_v\}} I < \infty$ once the estimate of $I$ is improved as described.

Note that the graphs contributing to the perturbation series for $(1/|\Lambda|)\log Z_N(\Lambda, f)$ to order $\lambda^n$ are finitely many because the number $r$ of external vertices is $r \le 2n + 2$ (since graphs must be connected). Hence, the perturbation series is finite to all orders in $\lambda$.

The above is the renormalizability proof of the scalar $\varphi^4$-fields in dimension $d = 2, 3$. The theory is renormalizable even if $d = 4$ as mentioned in the remark at the end of the previous section. The analysis would be very similar to the above: it is just a little more involved power-counting argument.

For more details, the reader is referred to Hepp (1966), Gallavotti (1985), sections 8 and 16.

## Asymptotic Freedom ($d = 2, 3$). Heuristic Analysis

Finiteness to all orders of the perturbation expansions is by no means sufficient to prove the existence of the ultraviolet limit for $Z_N(\Lambda, f)$ or for $(1/|\Lambda|) \log Z_N(\Lambda, f)$: and *a priori* it might not even be necessary. For this purpose, the first step is to check uniform (upper and lower) boundedness of $Z_N(\Lambda, f)$ as $N \to \infty$.

The reason behind the validity of a bound $e^{|\Lambda| E_-(\lambda,f)} \le Z_N(\Lambda, f) \le e^{|\Lambda| E_+(\lambda,f)}$ with $E_\pm(\lambda, f)$ cutoff independent has been made very clear after the introduction of the renormalization group methods in field theory. The approach studies the integral $Z_N(\Lambda, f)$, recursively, decomposing the field $\varphi^{(\le N)}_{\xi}$ into its regular components $z^{(h)}_{\xi}$, see [7], and integrating first over $z^{(N)}$, then over $z^{(N-1)}$ and so on.

The idea emerges naturally if the potential $V_N$ in [1] and [4] is written in terms of the "normalized" variables $X^{(N)}_{\xi} \overset{\text{def}}{=} \gamma^{-N(d-2)/2}\varphi^{(\le N)}_{\xi}$, see [6]; here if $d = 2$ the factor $\gamma^{(d-2)/2N}$ is interpreted as $N^{1/2}$.

The key remark is that as far as the integration over the small-scale component $z^{(N)}$ is concerned the field $X^{(N)}_{\xi}$ is a sum of two fields of size of order 1 (statistically),

$$X^{(N)}_{\xi} \equiv z^{(N)}_{\gamma^N \xi} + \gamma^{-(d-2)/2} X^{(N-1)}_{\xi}$$

if $d = 2$ this becomes

$$X^{(N)}_{\xi} \equiv \frac{1}{N^{1/2}} z^{(N)}_{\gamma^N \xi} + \frac{(N-1)^{1/2}}{N^{1/2}} X^{(N-1)}_{\xi}$$

and it can be considered to be smooth on scale $m^{-1}\gamma^{-N}$ (also statistically). Hence, approximately constant and of size of order O(1) on the small cubes $\Delta$ of volume $\gamma^{-dN} m^{-d}$ of the pavement $\mathcal{Q}_N$ introduced before [7]; at the same time it can be considered to take (statistically) independent values on different cubes of $\mathcal{Q}_N$. This is suggested by the inequalities [8]–[10].

Therefore, it is natural to decompose the potential $V_N$, see [5], as a sum over the small cubes $\Delta$ of volume $\gamma^{-dN} m^{-d}$ of the pavement $\mathcal{Q}_N$ as (see [14] for the definition of $\mu_N, \nu_N$), taking henceforth $m = 1$,

$$V_N(z^{(N)}) \overset{\text{def}}{=} -\sum_{\Delta \in \mathcal{Q}_N} \gamma^{-Nd} \int_\Delta \Big( \lambda \gamma^{2(d-2)N} X^{(N)\,4}_{\xi} + \mu_N \gamma^{(d-2)N} X^{(N)\,2}_{\xi} + \nu_N + f_\xi \gamma^{(d-2)/2N} X^{(N)}_{\xi} \Big) \frac{d\xi}{|\Delta|} \qquad [19]$$

where $\gamma^{(d-2)N}$ is interpreted as $N$ if $d=2$. Hence, if $d=3$ it is

$$V_N(z^{(N)}) \stackrel{\text{def}}{=} -\sum_{\Delta\in\mathcal{Q}_N} \gamma^{-N} \int_\Delta \Big(\lambda X_\xi^{(N)\,4} + \overline{\mu}_N X_\xi^{(N)\,2} + \overline{\nu}_N + f_\xi \gamma^{-\frac{3}{2}N} X_\xi^{(N)}\Big) \frac{\mathrm{d}\xi}{|\Delta|} \qquad [20]$$

where

$$\overline{\mu}_N \stackrel{\text{def}}{=} (-6\lambda c_N + \lambda^2 N \gamma^{-N} c'_N),$$
$$\overline{\nu}_N \stackrel{\text{def}}{=} 3\lambda c_N^2 + \lambda^2 \gamma^{-N} b_N + \lambda^3 N \gamma^{-2N} b'_N$$

and $c_N, c'_N, b_N, b'_N$, computable from [15] and [14], admit a limit as $N\to\infty$. While if $d=2$ it is

$$V_N(z^{(N)}) \stackrel{\text{def}}{=} -\sum_{\Delta\in\mathcal{Q}_N} N^2\gamma^{-2N} \int_\Delta \Big(\lambda X_\xi^{(N)\,4} + \overline{\mu}_N X_\xi^{(N)\,2} + \overline{\nu}_N + f_\xi N^{-\frac{3}{2}} X_\xi^{(N)}\Big) \frac{\mathrm{d}\xi}{|\Delta|} \qquad [21]$$

where $\overline{\mu}_N \stackrel{\text{def}}{=} -6\lambda c_N$ and $\overline{\nu}_N = 3\lambda c_N^2$ and $c_N$, computable from [13], admits a limit as $N\to\infty$.

The fields $z^{(N)}$ and $X^{(N-1)}$ can be considered constant over boxes $\Delta\in\mathcal{Q}_N$: $z_\xi^{(N)} = s_\Delta, X_\xi^{(N-1)} = x_\Delta$ for $\xi\in\Delta$ and the $s_\Delta$ can be considered statistically independent on the scale of the lattice $\mathcal{Q}_N$.

Therefore, [20] and [21] show that integration over $z^{(N)}$ in the integral defining $Z_N(\Lambda, f)$ is not too different from the computation of a partition function of a lattice continuous spin model in which the "spins" are $s_\Delta$ and, most important, interact extremely weakly if $N$ is large. In fact, the coupling constants are of order of a power of $|X^{(N-1)}|$ times $O(\gamma^{-N})$ if $d=3$ ($O(N^2\gamma^{-2N})$ if $d=2$), or of order $O(\gamma^{-N(d+2)/2}\max|f_\xi|)$, no matter how large $\lambda$ and $f$.

This says that the smallest scale fields are extremely weakly coupled. The fields $X^{(N-1)}$ can be regarded as external fields of size that will be called $B_{N-1}$, of order 1 or even allowed to grow with a power of $N$, see [6]. Their presence in $V_N$ does not affect the size of the couplings, as far as the analysis of the integral over $z^{(N)}$ is concerned, because the couplings remain exponentially small in $N$, see [20] and [21], being at worst multiplied by a power of $B_{N-1}$, i.e., changed by a factor which is a power of $N$.

The smallness of the coupling at small scale is a property called "asymptotic freedom." Once fields and coordinates are "correctly scaled," the real size of the coupling becomes manifest, that is, it is extremely small and the addends in $V_N$ proportional to the "counter-terms" $\mu_N, \nu_N$, which looked divergent when the fields were not properly scaled, are in fact of the same order or much smaller than the main $\varphi^4$-term.

Therefore, the integration over $z^{(N)}$ can be, heuristically, performed by techniques well established in statistical mechanics (i.e., by straightforward perturbation expansions): at least if the field $X_\xi^{(\leq N-1)}$ is smooth and bounded, as prescribed by [6], with $B=B_{N-1}$ growing as a power of $N$. In this case, denoting symbolically the integration over $z^{(N)}$ by $P$ or by $\langle\ldots\rangle$, it can be expected that it should give

$$\int \mathrm{e}^{V_N}\,\mathrm{d}P\Big(z^{(N)}\Big) \equiv \mathrm{e}^{V_{j;N-1}+\overline{\mathcal{R}}(j,N)|\Lambda|} \qquad [22]$$

where $V_{j;N-1}$ is the Taylor expansion of $\log\int \mathrm{e}^{V_N}\,\mathrm{d}P(z^{(N)})$ in powers of $\lambda$ (hence essentially in the very small parameter $\lambda\gamma^{-(4-d)N}$) truncated at order $j$, that is,

$$V_{1;N-1} = [\langle V_N\rangle]^{\leq 1}$$
$$V_{2;N-1} = \left[\langle V_N\rangle + \frac{(\langle V_N^2\rangle - \langle V_N\rangle^2)}{2!}\right]^{\leq 2}$$
$$V_{3;N-1} = \left[\langle V_N\rangle + \frac{(\langle V_N^2\rangle - \langle V_N\rangle^2)}{2!} + \frac{\Big(\langle V_N(\langle V_N^2\rangle - \langle V_N\rangle^2)\rangle - \langle V_N(\langle V_N^2\rangle - \langle V_N\rangle^2)\rangle\Big)}{3!}\right]^{\leq 3}, \ldots \qquad [23]$$

where $[\cdot]^{\leq j}$ denotes truncation to order $j$ in $\lambda$, and $\overline{\mathcal{R}}(j,N)$ is a remainder (depending on $\varphi_\xi^{(\leq N-1)}$) which can be expected to be estimated, for $d=2, 3$, by

$$|\overline{\mathcal{R}}(j,N)| \leq \mathcal{R}(j,N) \stackrel{\text{def}}{=} C_j B_N^{4j} (\lambda N^2 \gamma^{-(4-d)N})^{j+1} \gamma^{dN} \qquad [24]$$

for suitable constants $C_j$, that is, a remainder estimated by the $(j+1)$th power of the coupling times the number of boxes of scale $N$ in $\Lambda$. The relations [22]–[24] result from a naive Taylor expansion (in $\lambda$ of the $\log\int \mathrm{e}^{V_N}\,\mathrm{d}P(z^{(N)})$, taking into account that, in $V_N$ as a function of $z^{(N)}$, the $z^{(N)}$'s appear multiplied by quantities at most of size $\leq \lambda\gamma^{4-d}N^2B_N^3$, by [20] and [21] if $|X^{(N-1)}|\leq B_{N-1}$). In a statistical mechanics model for a lattice spin system, such a calculation of $Z_N$ would lead to a mean-field equation of state once the remainder was neglected.

The peculiarity of field theory is that a relation like [22] and [24] has to be applied again to $V_{j;N-1}$ to perform the integration over $z^{(N-1)}$ and define $V_{j;N-2}$ and, then, again to $V_{j;N-2}\ldots$. Therefore, it will be essential to perform the integral in [22] to an order (in $\lambda$) high enough so that the bound $\mathcal{R}(j,N)$ can be

summed over $N$: this requires (see [24]) an explicit calculation of [23] pushed at least to order $j=1$ if $d=2$ or to order $j=3$ if $d=3$; furthermore it is also necessary to check that the resulting $V_{j;N-1}$ can still be interpreted as low-coupling spin model so that [22] can be iterated with $N-1$ replacing $N$ and then with $N-2$ replacing $N-1, \ldots$.

The first necessary check towards a proof of the discussed heuristic "expectations" is that, defining recursively $V_{j;h}$ from $V_{j,h+1}$ for $h=N-1,\ldots,1,0$ by [23] with $V_N$ replaced by $V_{j;h+1}$ and $V_{j;N-1}$ replaced by $V_{j;h}$, the couplings between the variables $z^{(h)}$ do not become "worse" than those discussed in the case $h=N$. Furthermore, the field $\varphi_\xi^{(\leq N-1)}$ has a high probability of satisfying [6], but fluctuations are possible: hence the $\mathcal{R}$-estimate has to be combined with another one dealing with the large fluctuations of $X_\xi^{(N-1)}$ which has to be shown to be "not worse."

For more details, the reader is referred to Gallavotti (1978, 1985) and Benfatto and Gallavotti (1995).

## Effective Potentials and Their Scale (In)Dependence

To analyze the first problem mentioned at the end of the previous section, define $V_{j;h}$ by [23] with $V_N$ replaced by $V_{j;h+1}$ for $h=N-1,N-2,\ldots,0$. The quantities $V_{j;h}$, which are called "effective potentials" on scale $h$ (and order $j$), turn out to be in a natural sense scale independent: this is a consequence of renormalizability, realized by Wilson as a much more general property which can be checked, in the very special cases considered here with $d=2,3$, at fixed $j$ by induction, and in the super-renormalizable models considered here it requires only an elementary computation of a few Gaussian integrals as the case $j=3$ (or even $j=1$ if $d=2$) is already sufficient for our purposes.

It can also be (more easily) proved for general $j$ by a dimensional argument parallel to the one presented earlier to check finiteness of the renormalized series. The derivation is elementary but it should be stressed that, again, it is possible only because of the special choice of the counter-terms $\mu_N,\nu_N$. If $d=3$, the boundedness and smoothness of the fields $\varphi^{(\leq h)}$ and $z^{(h)}$ expressed by the second of [6] and of [10] is essential; while if $d=2$ the smoothness is not necessary.

The structure of $V_{j;h}$ is conveniently expressed in terms of the fields $X_\xi^{(h)}$, as a sum of three terms $V_h^{(\mathrm{rel})}$ (standing for "relevant" part), $V_h^{(\mathrm{irr})}$ (standing for "irrelevant" part), and a "field independent" part $E(j,h)|\Lambda|$.

The relevant part in $d=2$ is simply of the form [21] with $h$ replacing $N$: call it $V_h^{(\mathrm{rel},1)}$. If $d=3$, it is given by [20] with $h$ replacing $N$ plus, for $h<N$, a second "nonlocal" term

$$V_h^{(\mathrm{rel},2)} \stackrel{\mathrm{def}}{=} \frac{4^2\,3!}{2!\,2!}\lambda^2 \int \left(C_{\eta\eta'}^{(\leq h)\,3} - C_{\eta\eta'}^{(\leq N)\,3}\right) \times \left(\varphi_{\eta}^{(\leq h)} - \varphi_{\eta'}^{(\leq h)}\right)^2 \mathrm{d}\eta\mathrm{d}\eta'$$

which is conveniently expressed in terms of a "nonlocal" field

$$Y_{\eta\eta'}^{(h)} \stackrel{\mathrm{def}}{=} \frac{\varphi_{\eta}^{(\leq h)} - \varphi_{\eta'}^{(\leq h)}}{(\gamma^h|\eta-\eta'|)^{\frac{1}{4}}}$$

as $V_h^{(\mathrm{rel})} = V_h^{(\mathrm{rel},1)} + V_h^{(\mathrm{rel},2)}$ with

$$V_h^{(\mathrm{rel},2)} \stackrel{\mathrm{def}}{=} -\lambda^2\gamma^{-2h} \sum_{\Delta,\Delta'\in\mathcal{Q}_h} \int_{\Delta\times\Delta'} Y_{\eta\eta'}^{(h)2} A_{\eta\eta'}^{(h)} \times \mathrm{e}^{-c'\gamma^h|\eta-\eta'|} \frac{\mathrm{d}\eta\mathrm{d}\eta'}{|\Delta|\,|\Delta'|} \qquad [25]$$

where

$$0 < a \leq \left(\frac{A_{\eta\eta'}^{(h)}}{(\gamma^h|\eta-\eta'|)^{3-(1/2)}}\right)_N < a'$$

with $a,a',c'>0$ and the subscript $N$ means that the expression in parenthesis "saturates at scale $N$", i.e., its denoninator becomes $\gamma^{(3-(1/2))(h-N)}$ as $|\eta-\eta'|\to 0$.

The expression [25] is not the full part of the potential $V_{j;h}$ which is of second order in the fields: there are several other contributions which are collected below as "irrelevant."

It should be stressed that "irrelevant" is a traditional technical term: by no means it should suggest "negligibility." On the contrary, it could be maintained that the whole purpose of the theory is to study the irrelevant terms. The irrelevant part of the potential can be better designated as the "driven part," as its behavior is "controlled" by the relevant part: although initially $V_{j;h}$, $h=N$, contains no irrelevant terms, it eventually contains them for $h<N$ and they keep getting generated as $h$ diminishes. Furthermore, the part of the irrelevant terms generated at scale $h_0 \leq N$ becomes very small at scales $h \ll h_0$ so that the irrelevant part of $V_{j;h}$ at small $h$ (e.g., at $h=0$, i.e., on the "physical scale" of the observer) only depend on the relevant terms in a few scales near $h$.

It also turns out that the Schwinger functions are simply related to the irrelevant terms.

The irrelevant part of the effective potential can be expressed as a finite sum of integrals of

monomials in the fields $X^{(h)}_\xi$ if $d=2$, or in the fields $X^{(h)}_\xi$ and $Y^{(h)}_{\eta\eta'}$ if $d=3$, which can be written as $V^{(\mathrm{irr})}_{j;h}$ given by

$$\int \left( \prod_{k=1}^{p} X^{(h)\, n_k}_{\xi_k} \prod_{k'=1}^{q} Y^{(h)\, n'_{k'}}_{\eta_{k'}\eta'_{k'}} \right) \mathrm{e}^{-\gamma^h c' d(\xi_1,\ldots,\eta'_q)} \lambda^n \gamma^{-ht} \times W(\xi_1,\ldots,\eta'_q) \prod_{k=1}^{p} \frac{\mathrm{d}\xi_k}{|\Delta_k|} \prod_{k'=1}^{q} \frac{\mathrm{d}\eta_{k'}\mathrm{d}\eta'_{k'}}{|\Delta^1_{k'}|\,|\Delta^2_{k'}|} \qquad [26]$$

with the integral extended to products $\Delta_1 \times \cdots \times \Delta_p \times \cdots \times (\Delta^1_q \times \Delta^2_q)$ of boxes $\Delta \in \mathcal{Q}_h$, and $d(\xi_1,\ldots,\eta'_q)$ is the length of the shortest tree graph that connects all the $p+2q>0$ points, the exponents $n,t$ are $\geq 2$, and $t$ is $\geq 3$ if $q>0$; the kernel $W$ depends on all coordinates $\xi_1,\ldots,\eta'_q$ and it is bounded above by $C_j \prod_{k'=1}^{q} A_{\eta_{k'}\eta'_{k'}}$ for some $C_j$; the sums $\sum n_k + \sum n'_{k'}$ cannot exceed $4j$. The test functions $f$ do not appear in [26] because by assumption they are bounded by 1: but $W$ depends on the $f$'s as well.

The field-independent part is simply the value of $\log Z_N(\Lambda,f)$ computed by the perturbation analysis in the section "Perturbation theory" up to order $j$ in $\lambda$ but using as propagator $(C^{(\leq N)} - C^{(\leq h)})$: thus, $E(j,h)$ is a constant depending on $N$ but uniformly bounded as $N\to\infty$ (because of the renormalizability proved in the section "Perturbation theory").

If $d=2$, there is no need to introduce the nonlocal fields $Y^{(h)}$ and in [26] one can simply take $q=0$, and the relevant part also can be expressed by omitting the term $V^{(\mathrm{rel},2)}_h$ in [25]: unlike the $d=3$ case, the estimate on the kernels $W$ by an $N$-independent $C_j$ holds uniformly in $h$ without having to introduce $Y$. For $d=2$, it will therefore be supposed that $V^{(\mathrm{rel},2)}_h \equiv 0$ in [25] and $q=0$ in [26].

It is not necessary to have more information on the structure of $V_{j;h}$ even though one can find simple graphical rules, closely related to the ones in the section "Perturbation theory," to construct the coefficients $W$ in full detail. The $W$ depend, of course, on $h$ but the uniformity of the bound on $W$ is the only relevant property and in this sense the effective potentials are said to be (almost) "scale independent."

The above bounds on the irrelevant part can be checked by an elementary direct computation if $j\leq 3$: in spite of its "elementary character," the uniformity in $h\leq N$ is a result ultimately playing an essential role in the theory together with the dominance of the relevant part over the irrelevant one which, once the fields are properly scaled, is "much smaller" (by a factor of order $\gamma^{-h}$, see [26]), at least if $h$ is large.

**Remarks**

(i) Checking scale independence for $j=1$ is just checking that $\int P(\mathrm{d}z^{(h)}) V_{1;h} = V_{1;h-1}$. Note that

$$V_{1;h} \stackrel{\mathrm{def}}{=} \int_\Lambda \lambda\left(\varphi_\xi^{(\leq h)4} - 6C_{00}^{(\leq h)}\varphi_\xi^{(\leq h)2} + 3C_{00}^{(\leq h)2}\right)\mathrm{d}\xi$$

hence, calling $:\varphi_\xi^{(\leq h)4}:$ the polynomial in the integral (Wick's monomial of order 4), we have here an elementary Gaussian integral ("martingale property of Wick monomials"). Note the essential role of the counter-terms. For $j>1$, the computation is similar but it involves higher-order polynomials (up to $4j$) and the distinction between $d=2$ and $d=3$ becomes important.

(ii) $V_{j;0}$ contains only the field-independent part $E(j,0)|\Lambda|$ (see above) which is just a number (as there are no fields of scale 0): by the above definitions, it is identical to the perturbative expansion truncated to $j$th order in $\lambda$ of $\log Z_N(\Lambda,f)$, well defined as discussed earlier.

## Nonperturbative Renormalization: Small Fields

Having introduced the notion of effective potential $V_{j;h}$, of order $j$ and scale $h$, satisfying the bounds (described after [26]) on the kernels $W$ representing it, the problem is to estimate the remainder in [22] and find its relation with the value [24] given by the heuristic Taylor expansion. Assume $\lambda<1$ to avoid distinguishing this case from that with $\lambda\geq 1$ which would lead to very similar estimates but to different $\lambda$-dependence on some constants.

Define $\chi_B(z^{(h)})=1$ if $\|z^{(h)}\|_\Delta \leq Bh^2$ for all $\Delta\in\mathcal{Q}_h$, see [8], and 0 otherwise; then the following lemma holds:

**Lemma 1** *Let $\|X^{(h)}\|_\Delta$ be defined as [8] with $z$ replaced by $X$ and suppose $\|X^{(h)}\|_\Delta \leq Bh^4$ for all $\Delta$ then, for all $j\geq 1$, it is*

$$\int \mathrm{e}^{V_{j;h+1}}\chi_B(z^{(h+1)})\mathrm{d}P(z^{(h+1)}) = \mathrm{e}^{V_{j;h}+\mathcal{R}'(j,h+1)|\Lambda|} \qquad [27]$$

*with, for suitable constants $c_-, c'_-$,*

$$|\mathcal{R}'_-(j;h+1)| \leq \mathcal{R}_-(j;h+1) \stackrel{\mathrm{def}}{=} \mathcal{R}(j;h+1) + c_-\mathrm{e}^{-c'_-B^2(h+1)^2}$$

*and $\mathcal{R}(j;h+1)$ given by [24] with $h+1$ in place of $N$.*

Since $Z_N(\Lambda,f) \geq \int \mathrm{e}^{V_N}\prod_{h=1}^N \chi_B(z^{(h)})P(\mathrm{d}z^{(h)})$ this immediately gives a lower bound on $E=(1/|\Lambda|)\log Z_N(\Lambda,f)$: in fact if $\chi_B(\|z^{(h')}\|)=1$ for

$h'=1,\ldots,h$, then $\|X^{(h)}\|_\Delta \le c\,Bh'^4$ for some $c$ so that, by recursive application of Lemma 1, $Z_N(\Lambda,f)\ge e^{V_{j,0}-\sum_{h=1}^{N}\mathcal{R}_-(j,h)|\Lambda|}$. By the remark at the end of the previous section, given $j$ the lower bound on $E$ just described agrees with the perturbation expansion of $E=(1/|\Lambda|)\log Z_N(\Lambda,f)$ truncated to order $j$ (in $\lambda$) up to an error bounded by $\sum_{h=1}^{\infty}\mathcal{R}_-(j,h)$.

**Remark** The problem solved by Lemma 1 is usually referred to as the small-field problem, to contrast it with the large-field problem discussed later. The proof of the lemma is a simple Taylor expansion in $\lambda\gamma^{-h}$ if $d=3$ or in $\lambda h^2\gamma^{-2h}$ if $d=2$ to order $j$ (in $\lambda$). The constraint on $z^{(h+1)}$ makes the integrations over $z^{(h+1)}$, necessary to compute $V_{j;h}$ from $V_{j;h+1}$, not Gaussian. But the tail estimates [9], together with the Markov property of the distribution of $z^{(h)}$ can be used to estimate the difference with respect to the Gaussian unconstrained integrations of $z^{(h+1)}$: and the result is the addition of the small "tail error" changing $\mathcal{R}$ into $\mathcal{R}_-$ in [27]. The estimate of the main part of the remainder $\mathcal{R}$ would be obvious if the fields $z^{(h)}$ were independent on boxes of scale $\gamma^{-h}$: they are not independent but they are Markovian and the estimate can be done by taking into account the Markov property.

For more details, the reader is referred to Wilson (1970, 1972), Gallavotti (1978, 1981, 1985), and Benfatto *et al.* (1978).

## Nonperturbative Renormalization: Large Fields, Ultraviolet Stability

The small-field estimates are not sufficient to obtain ultraviolet stability: to control the cases in which $|X^{(h)}_{\boldsymbol{\xi}}|>Bh^4$ for some $\boldsymbol{\xi}$ or some $h$, or $|Y^{(h)}_{\boldsymbol{\xi}\boldsymbol{\eta}}|>Bh^4$ for some $|\boldsymbol{\xi}-\boldsymbol{\eta}|<\gamma^{-h}$, a further idea is necessary and it rests on making use of the assumption that $\lambda>0$ which, in a sense to be determined, should suppress the contribution to the integral defining $Z_N(\Lambda,f)$ coming from very large values of the field. Assume also $\lambda<1$ for the same reasons advanced in the section "Effective potentials and their scale (in)dependence."

Consider first $d=2$. Let $\mathcal{D}_N$ be the "large-field region" where $|X^{(N)}_{\boldsymbol{\xi}}|>BN^4$ and let $V_N(\Lambda/\mathcal{D}_N)$ be the integral defining the potential in [21] extended to the region $\Lambda/\mathcal{D}_N$, complement of $\mathcal{D}_N$. This region is typically very irregular (and random as $X$ itself is random with distribution $P_N$).

An upper bound on the integral defining $Z_N(\Lambda,f)$ is obtained by simply replacing $e^{V_N}$ by $e^{V_N(\Lambda/\mathcal{D}_N)}$ because in $\mathcal{D}_N$ the first term in the integrand in [21] is $\le -\lambda N^2\gamma^{2N}(BN^4)<0$ and it *overwhelmingly* dominates on the remaining terms whose value is bounded by a similar expression with a smaller power of $N$. Then if $\mathcal{E}^c\stackrel{\text{def}}{=}\Lambda/\mathcal{E}$ denotes the complement in $\Lambda$ of a set $\mathcal{E}\subset\Lambda$:

**Lemma 2** *Let $d=2$. Define $V_h(\mathcal{D}^c_h)$ to be given by the expression [22] with the integrals extending over $\Delta_j/\mathcal{D}_h$ and define $\mathcal{R}(j,h+1)$ by [24]. Then*

$$\int e^{V_{h+1}(\mathcal{D}^c_{h+1})}\,dP\left(z^{(h+1)}\right)=e^{V_h(\mathcal{D}^c_h)+\overline{\mathcal{R}}_+(j,h+1)|\Lambda|}\qquad[28]$$

*where* $|\overline{\mathcal{R}}_+(j,h+1|\le\mathcal{R}_+(j,h+1\stackrel{\text{def}}{=}\mathcal{R}(j;h+1)+c_+e^{-c'_+B^2(h+1)^2}$ *with suitable* $c_+,c'_+$.

**Remark** Lemma 2 is genuinely not perturbative and making essential use of the positivity of $\lambda$. Below the analysis of the proof of the lemma, which consists essentially in its reduction to Lemma 1, is described in detail. It is perhaps the most interesting part and the core of the theory of the proof that truncating the expansion in $\lambda$ of $(1/|\Lambda|)\log Z_N(\Lambda,f)$ to order $j$ gives as a result an estimate exact to order $\lambda^{j+1}$ of $(1/|\Lambda|)\log Z_N(\Lambda,f)$.

Let $R_N$ be the cubes $\Delta\in\mathcal{Q}_N$ in which there is at least one point $\boldsymbol{\xi}$ where $|z^{(N)}_{\boldsymbol{\xi}}|\ge BN^2$. By definition, the region $\mathcal{D}_N/\mathcal{D}_{N-1}$ is covered by $R_N$.

Remark that in the region $\mathcal{D}_{N-1}/R_N$ the field $X^{(N-1)}$ is large but $z_N$ is not large so that $X^{(N)}$ is still very large: this is so because the bounds set to define the regions $\mathcal{D}$ and $R$ are quite different being $BN^4$ and $BN^2$, respectively. Hence, if a point is in $\mathcal{D}_{N-1}$ and not in $R_N$, then the field $X^{(N)}$ must be of the order $\gg BN^3$. Therefore, by positivity of the $\lambda\varphi^{(\le N)4}_{\boldsymbol{\xi}}$ term (which dominates all other terms so that $V^{(N)}(\varphi^{(\le N)}_{\boldsymbol{\xi}})<0$ for $\boldsymbol{\xi}\in\mathcal{D}_N\cup(\mathcal{D}_{N-1}/R_N))$ we can replace $V_N(\mathcal{D}^c_N)$ by $V((\mathcal{D}_N\cup(\mathcal{D}_{N-1}/R_N))^c)$, for the purpose of obtaining an upper bound.

Furthermore, modulo a suitable correction, it is possible to replace $V((\mathcal{D}_N\cup(\mathcal{D}_{N-1}/R_N))^c)$ by $V((\mathcal{D}_{N-1}\cup R_N)^c)$: because the integrand in $V_N$ is bounded below by

$$-b\lambda\gamma^{-2N}N^2$$

if $d=2$ (by $-b\lambda\gamma^{-N}$ if $d=3$), for some $b$, so that the points in $R_N$ can at most lower $V((\mathcal{D}_N\cup(\mathcal{D}_{N-1}/R_N))^c)$ by $-b\lambda N^2\ \gamma^{-(4-d)N}\#(R_N)$ if $\#R_N$ is the number of boxes of $\mathcal{Q}_N$ in $R_N$ and $V(\varphi_{\boldsymbol{\xi}})$ is bounded below by its minimum: thus,

$$V((\mathcal{D}_{N-1}\cup R_N)^c)+b\lambda N^2\gamma^{(4-d)N}\#(R_N)$$

is an upper bound to $V((\mathcal{D}_N\cup(\mathcal{D}_{N-1}/R_N))^c)$.

In the complement of $\mathcal{D}_{N-1}\cup R_N$, all fields are "small"; if $X^{(N-1)}$ and $R_N$ are fixed this region is not random (as a function of $z^{(N)}$) any more. Therefore,

if $X^{(N-1)}, R_N$ are fixed the integration over $z^{(N)}$, conditioned to having $z^{(N)}$ fixed (and large) in the region $R_N$, is performed by means of the same argument necessary to prove Lemma 1 (essentially a Taylor expansion in $\lambda\gamma^{-(4-d)N}$). The large size of $z^{(N)}$ in $R_N$ does not affect too much the result because on the boundary of $R_N$ the field $z^{(N)}$ is $\leq BN^2$ (recalling that $z^{(N)}$ is continuous) and since the variable $z^{(N)}$ is Markovian, the boundary effect decays exponentially from the boundary $\partial R_N$: it adds a quantity that can be shown to be bounded by the number of boxes in $R_N$ on the boundary of $R_N$, hence by $\#R_N$, times $b'(N-1)^2\gamma^{-(4-d)}(B(N-1)^4)^4$ for some $b'$.

The result of the integration over $z^{(N)}$ of $e^{V_N((\mathcal{D}_N \cup (\mathcal{D}_{N-1}/R_N))^c)}$ conditioned to the large-field values of $z^{(N)}$ in $R_N$ leads to an upper bound on $\int e^{V_N} P(dz^{(N)})$ as

$$\sum_{R_N} e^{V_{j;N-1}(\mathcal{D}^c_{N-1})+\mathcal{R}(j,N)|\Lambda|} \times \prod_{\Delta\in R_N} \left(c\, e^{-c'(BN^2)^2} e^{+c''\lambda\gamma^{-(4-d)N}N^2(BN^4)^4}\right)^{\#R_N} \qquad [29]$$

where $c, c', c''$ are suitable constants: this is explained as follows.

1. Taylor expansion (in $\lambda$) of the integral $e^{V_N((\mathcal{D}_{N-1}\cup R_N)^c)+b\lambda N^2\gamma^{-(4-d)N}\#(R_N)}$ (which, by construction, is an upper bound on $e^{V_N(\mathcal{D}^c_N)}$) with respect to the field $z^{(N)}$, conditioned to be fixed and large in $R_N$, would lead to an upper bound as

$$e^{V_{j;N-1}((\mathcal{D}_{N-1}\cup R_N)^c)+\mathcal{R}'(j,N)|\Lambda|+b''\lambda(BN^4)^4\gamma^{(4-d)N}\#(R_N)}$$

   with $\mathcal{R}'$ equal to [24] possibly with some $C'_j$ replacing $C_j$. The second exponential on the RHS of [29] arises partly from the above correction $b''\lambda(BN^4)^4\gamma^{-(4-d)N}\#(R_N)$ and partly from a contribution of similar form explained in (3) below.
2. Integration over the large conditioning fields fixed in $R_N$ is controlled by the second estimate in [9] (the tail estimate): the first factors in parentheses in [29] is the tail estimate just mentioned, i.e., the probability that $z^{(N)}$ is large in the region $R_N$. The second factor is only partly explained in (1) above.
3. Without further estimates, the bound [29] would contain $V_{j;N-1}((\mathcal{D}_{N-1}\cup R_N)^c)$ rather than $V_{j;N-1}(\mathcal{D}^c_{N-1})$. Hence, there is the need to change the potential $V_{j;N-1}((\mathcal{D}_{N-1}\cup R_N)^c)$ by "reintroducing" the contribution due to the fields in $R_N/\mathcal{D}_{N-1}$ in order to reconstruct $V_{j;N-1}(\mathcal{D}^c_{N-1})$. Reintroducing this part of the potential costs a quantity like $b'\lambda N^2\gamma^{(4-d)N}(BN^4)^4\#(R_N)$ (because the reintroduction occurs in the region $R_N/\mathcal{D}_{N-1}$ which is covered by $R_N$ and in such points the field $X^{(N-1)}_\xi$ is *not large*, being bounded by $B(N-1)^4$); so that their contribution to the effective potential is still dominated by the $\varphi^4$-term and therefore by $\gamma^{-(4-d)N}$ times a power of $BN^4$ times the volume of $R_N$ (in units $\gamma^{-N}$, i.e., $\#R_N$). All this is taken care of by suitably fixing $c''$.

Note that the sum over $R_N$ of [29] is

$$\left(1 + c\, e^{-c'B^2N^4} e^{+c''\lambda\gamma^{-(4-d)N}N^2(BN^4)^4}\right)^{\gamma^{dN}|\Lambda|}$$

(because $\Lambda$ contains $|\Lambda|\gamma^{dN}$ cubes of $\mathcal{Q}_N$); hence, it is bounded above by $e^{c_+ e^{-c'_+B^2N^2}}$ for suitably defined $c_+, c'_+$.

The same argument can be repeated for $V_{j;h}(\mathcal{D}^c_h)$ with any $h$ if $V_{j;h}(\mathcal{D}^c_h)$ is defined by the sum over $\Delta$'s in $\mathcal{Q}_h$ of the same integrals as those in [25] and [26] with $\Delta_j/\mathcal{D}_h$ replacing $\Delta_j$ in the integration domains.

Applying Lemma 1 recursively with $j \geq 1$ (if $d=3$ it would become necessary to take $j\geq 3$), it follows that there exist $N$-independent upper and lower bounds $E_\pm|\Lambda|$ on $\log Z(\Lambda, f)$ of the form $V_{j;0} \pm \sum_{h=1}^{\infty}(\mathcal{R}(j,h) + c_\pm e^{-c'_\pm B^2h^2})|\Lambda|$ for $c_\pm, c'_\pm > 0$ suitably chosen and $\lambda$-independent for $\lambda < 1$. By the remark at the end of Sec.6, given $j$, the bounds just described agree with the perturbation expansion $E(j,0)|\Lambda| \equiv V_{j;0}$ of $\log Z(\Lambda, f)$ truncated to order $j$ (in $\lambda$) up to the remainders $\pm\sum_{h=1}^{\infty}\mathcal{R}_\pm(j,h)$. Hence, if $B$ is chosen proportional to $\log_+\lambda^{-1} \stackrel{\text{def}}{=} \log(e+\lambda^{-1})$, the upper and lower bounds coincide to order $j$ in $\lambda$ with the value obtained by truncating to order $j$ the perturbative series.

The latter remark is important as it implies not only that the bounds are finite (by the section "Perturbation theory") but also that the function $(1/|\Lambda|)\log Z(\Lambda, f)$ is not quadratic in $f$: already to order 1 in $\lambda$ it is quartic in $f$ (containing a term equal to $-\lambda(\int C_{\xi,0}f_\xi d\xi)^4$).

The latter property is important as it excludes that the result is a "Gaussian" generating function. Thus, the outline of the proof of Lemma 2, which together with Lemma 1 forms the core of the analysis of the ultraviolet stability for $d=2$, is completed.

If $d=3$, more care is needed because (very mild) smoothness, like the considered Hölder continuity with exponent 1/4, of $z$, $X$ is necessary to obtain the key scale independence property discussed in earlier: therefore, the natural measure of the size of $z^{(h)}$ and $X^{(h)}$ in a box $\Delta \in \mathcal{Q}_h$ is no longer the maximum of $|z^{(h)}_\xi|$ or of $|X^{(h)}_\xi|$. The region $\mathcal{D}_h$ becomes more

involved as it has to consist of the points $\boldsymbol{\xi}$ where $|X^{(h)}_{\boldsymbol{\xi}}| > Bh^4$ *and* of the pairs $\boldsymbol{\eta}, \boldsymbol{\eta}'$ where

$$|Y_{\boldsymbol{\eta},\boldsymbol{\eta}'}| \equiv \frac{\left|X^{(h)}_{\boldsymbol{\eta}} - X^{(h)}_{\boldsymbol{\eta}'}\right|}{(\gamma^h|\boldsymbol{\eta} - \boldsymbol{\eta}'|)^{\frac{1}{4}}} > Bh^4$$

i.e., it is not just a subset of $\Lambda$.

However, if $d = 3$, the relevant part also contains the negative term $V^{(\mathrm{rel},2)}$, see [25]: and since it dominates over all other terms which contain a $Y$-field (because their couplings [25] are smaller by about $\gamma^{-h}$), the argument given for $d = 2$ can be adapted to the new situation. Two regions $\mathcal{D}^1_h, \mathcal{D}^2_h$ will be defined: the first consists of all the points $\boldsymbol{\xi}$ where $|X^{(h)}_{\boldsymbol{\xi}}| > Bh^4$ and the second of all the pairs $\boldsymbol{\eta}, \boldsymbol{\eta}'$ where $|Y^{(h)}_{\boldsymbol{\eta},\boldsymbol{\eta}'}| > Bh^4$. The region $R_h$ will be the collection of all $\Delta \in \mathcal{Q}_h$, where $\|z^{(h)}\|_\Delta > Bh^2$, see [8] with $\tau = 0$. Then $V(\mathcal{D}^c_h)$ will be defined as the sum of the integrals in [25] and [26] with the integrals over $\boldsymbol{\xi}_i$ further restricted to $\boldsymbol{\xi}_i \notin \mathcal{D}^1_h$ and those over the pairs $\boldsymbol{\eta}_i, \boldsymbol{\eta}'_i$ are further restricted to $(\boldsymbol{\eta}_i, \boldsymbol{\eta}'_i) \notin \mathcal{D}^2_h$. With the new settings, Lemma 2 can be proved also for $d = 3$ along the same lines as in the $d = 2$ case.

For more details, the reader is referred to Wilson (1970, 1972), Benfatto *et al.* (1978), and Gallavotti (1981).

## Ultraviolet Limit, Infrared Behavior, and Other Applications

The results on the ultraviolet stability are nonperturbative, as no assumption is made on the size of $\lambda$ (the assumption $\lambda < 1$ has been imposed in the last two sections only to obtain simpler expressions for the $\lambda$-dependence of various constants): nevertheless the multiscale analysis has allowed us to use perturbative techniques (i.e., the Taylor expansion in Lemmata 1, 2) to find the solution. The latter procedure is the essence of the renormalization group methods: they aim at reducing a difficult multiscale problem to a sequence of simple single-scale problems. Of course, in most cases, it is difficult to implement the approach and the scalar quantum fields in dimensions 2, 3 are among the simplest examples. The analysis of the beta function and of the running couplings, which appear in essentially all renormalization group applications, does not play a role here (or, better, their role is so inessential that it has even been possible to avoid mentioning them). This makes the models somewhat special from the renormalization group viewpoint: the running couplings at length scale $h$, if introduced, would tend exponentially to 0 as $h \to \infty$; unlike what happens in the most interesting renormalization group applications in which they either tend to zero only as powers of $h$ or do not tend to zero at all.

The multiscale analysis method, i.e., the renormalization group method, in a form close to the one discussed here has been applied very often since its introduction in physics and it has led to the solution of several important problems. The following is not an exhaustive list and includes a few open questions.

1. The arguments just discussed imply, with minor extra work that $Z_N(\Lambda, f)$ as $N \to \infty$ not only admit uniform upper and lower bounds but also that the limit as $N \to \infty$ actually exists and it is a $C^\infty$ function of $\lambda$, $f$. Its $\lambda$ and $f$-derivatives at $\lambda = 0$ and $f = 0$ are given by the formal perturbation calculation. In some cases, it is even possible to show that the formal series for $Z_N(\Lambda, f)$ in powers of $\lambda$ is Borel summable.
2. The problem of removing the infrared cutoff (i.e., $\Lambda \to \infty$) is in a sense more a problem of statistical mechanics. In fact, it can be solved for $d = 2, 3$ by a typical technique used in statistical mechanics, the "cluster expansion." This is not intended to mean that it is technically an easy task: understanding its connection with the low-density expansions and the possibility of using such techniques has been a major achievement that is not discussed here.
3. The third problem mentioned in the introduction, that is, checking the axioms so that the theory could be interpreted as a quantum field theory is a difficult problem which required important efforts to control and which is not analyzed here. An introduction to it can be its analysis in the $d = 2$ case.
4. Also the problem of keeping the ultraviolet cutoff and removing the infrared cutoff while the parameter $m^2$ in the propagator approaches 0 is a very interesting problem related to many questions in statistical mechanics at the critical point.
5. Field theory methods can be applied to various statistical mechanics problems away from criticality: particularly interesting is the theory of the neutral Coulomb gas and of the dipole gas in two dimensions.
6. The methods can be applied to Fermi systems in field theory as well as in equilibrium statistical mechanics. The understanding of the ground state in not exactly soluble models of spinless fermions in one dimension at small coupling is one of the results. And via the transfer matrix theory it has led to the understanding of nontrivial critical behavior in two-dimensional models that are not exactly soluble (like Ising next-nearest-neighbor or Ashkin–Teller model). Fermi systems are of particular interest also because in their analysis the large-fields problem is absent, but this great

technical advantage is somewhat offset by the anticommutation properties of the fermionic fields, which do not allow us to employ probabilistic techniques in the estimates.

7. An outstanding open problem is whether the scalar $\varphi^4$-theory is possible and nontrivial in dimension $d=4$: this is a case of a renormalizable not asymptotically free theory. The conjecture that many support is that the theory is necessarily trivial (i.e., the function $Z_N(\Lambda, f)$ becomes necessarily a Gaussian in the limit $N \to \infty$). One of the main problems is the choice of the ultraviolet cut-off; unlike the $d=2,3$ cases in which the choice is a matter of convenience it does not seem that the issue of triviality can be settled without a careful analysis of the choice and of the role of the ultraviolet cut-off.
8. Very interesting problems can be found in the study of highly symmetric quantum fields: gauge invariance presents serious difficulties to be studied (rigorously or even heuristically) because in its naive forms it is incompatible with regularizations. Rigorous treatments have been in some cases possible and in few cases it has been shown that the naive treatment is not only not rigorous but it leads to incorrect results.
9. In connection with item (8) an outstanding problem is to understand relativistic pure gauge Higgs fields in dimension $d=4$: the latter have been shown to be ultraviolet stable but the result has not been followed by the study of the infrared limit.
10. The classical gauge theory problem is quantum electrodynamics, QED, in dimension 4: it is a renormalizable theory (taking into account gauge invariance) and its perturbative series truncated after the first few orders give results that can be directly confronted with experience, giving very accurate predictions. Nevertheless, the model is widely believed to be incomplete: in the sense that, if treated rigorously, the result would be a field describing free noninteracting assemblies of photons and electrons. It is believed that QED can make sense only if embedded in a model with more fields, representing other particles (e.g., the standard model), which would influence the behavior of the electromagnetic field by providing an effective ultraviolet cutoff high enough for not altering the predictions on the observations on the time and energy scales on which present (and, possibly, future over a long time span) experiments are performed. In dimension $d=3$, QED is super-renormalizable, once the gauge symmetry is properly taken into account, and it can be studied with the techniques described above for the scalar fields in the corresponding dimension.

In general, constructive quantum field theory seems to be in a deep crisis: the few solutions that have been found concern very special problems and are very demanding technically; the results obtained have often not been considered to contribute appreciably to any "progress." And many consider that the work dedicated to the subject is not worth the results that one can even hope to obtain. Therefore, in recent years, attempts have been made to follow other paths: an attitude that in the past usually did not lead, in general to great achievements but that is always tempting and worth pursuing because the rare major progresses made in physics resulted precisely by such changes of attitude, leaving aside developments requiring work which was too technical and possibly hopeless: just to mention an important case, one can recall quantum mechanics which disposed of all attempts at understanding the observed atomic levels quantization on the basis of refined developments of classical electromagnetism.

For more details, the reader is referred to Nelson (1966), Guerra (1972), Glimm *et al.* (1973), Glimm and Jaffe (1981), Simon (1974), Benfatto *et al.* (1978, 2003), Aizenman (1982), Gawedzky and Kupiainen (1983, 1985a, b), Balaban (1983), and Giuliani and Mastropietro (2005).

*See also:* Algebraic Approach to Quantum Field Theory; Axiomatic Quantum Field Theory; Euclidean Field Theory; Integrability and Quantum Field Theory; Perturbation Theory and its Techniques; Quantum Field Theory: A Brief Introduction; Scattering, Asymptotic Completeness and Bound States.

## Further Reading

Aizenman M (1982) Geometric analysis of $\varphi^4$-fields and Ising models. *Communications in Mathematical Physics* 86: 1–48.

Balaban T (1983) $(\text{Higgs})_{3,2}$ quantum fields in a finite volume. III. Renormalization. *Communications in Mathematical Physics* 88: 411–445.

Benfatto G, Cassandro M, Gallavotti G *et al.* (1978) Some probabilistic techniques in field theory. *Communications in Mathematical Physics* 59: 143–166.

Benfatto G, Cassandro M, Gallavotti G *et al.* (1980) Ultraviolet stability in Euclidean scalar field theories. *Communications in Mathematical Physics* 71: 95–130.

Benfatto G and Gallavotti G (1995) *Renormalization Group*, pp. 1–143. Princeton: Princeton University Press.

Benfatto G, Giuliani A, and Mastropietro V (2003) Low temperature analysis of two dimensional Fermi systems with symmetric Fermi surface. *Annales Henry Poincaré* 4: 137–193.

De Calan C and Rivasseau V (1981) Local existence of the Borel transform in euclidean $\phi_4^4$. *Communications in Mathematical Physics* 82: 69–100.

Fröhlich J (1982) On the triviality of $\lambda\phi_d^4$ theories and the approach to the critical point in $d \geq 4$ dimensions. *Nuclear Physics B* 200: 281–296.

Gallavotti G (1978) Some aspects of renormalization problems in statistical mechanics. *Memorie dell' Accademia dei Lincei* 15: 23–59.
Gallavotti G (1981) Elliptic operators and Gaussian processes. In: *Aspects Statistiques et Aspects Physiques des Processus Gaussiens*, pp. 349–360. Colloques Internat. C.N.R.S, St. Flour. Publications du CNRS, Paris.
Gallavotti G (1985) Renormalization theory and ultraviolet stability via renormalization group methods. *Reviews of Modern Physics* 57: 471–569.
Gawedzky K and Kupiainen A (1983) Block spin renormalization group for dipole gas and $(\partial\phi)^4$. *Annals of Physics* 147: 198–243.
Gawedzky K and Kupiainen A (1985a) Gross–Neveu model through convergent perturbation expansion. *Communications in Mathematical Physics* 102: 1–30.
Gawedzky K and Kupiainen A (1985b) Massless lattice $\phi_4^4$ theory: rigorous control of a renormalizable asymptotically free model. *Communications in Mathematical Physics* 99: 197–252.
Giuliani A and Mastropietro V (2005) Anomalous universality in the anisotropic Ashkin–Teller model. *Communications in Mathematical Physics* 256: 681–735.
Glimm J, Jaffe A, and Spencer T (1973) Velo G and Wightman A (eds.) *Constructive Field theory*, Lecture Notes in Physics, vol. 25, pp. 132–242. New York: Springer.
Glimm J and Jaffe A (1981) *Quantum Physics*. Springer.
Guerra F (1972) Uniqueness of the vacuum energy density and Van Hove phenomena in the infinite volume limit for two-dimensional self-coupled Bose fields. *Physical Review Letters* 28: 1213–1215.
Hepp K (1966) *Théorie de la rénormalization*. Lecture Notes in Physics, vol. 2. Heidelberg: Springer.
Nelson E (1966) A quartic interaction in two dimensions. In: Goodman R and Segal I (eds.) *Mathematical Theory of Elementary Particles*, pp. 69–73. Cambridge: M.I.T.
Osterwalder K and Schrader R (1973) Axioms for Euclidean Green's functions. *Communications in Mathematical Physics* 31: 83–112.
Simon B (1974) *The* $P(\varphi)_2$ *Euclidean (Quantum) Field Theory*. Princeton: Princeton University Press.
Streater RF and Wightman AS (1964) *PCT, Spin, Statistics and All That*. Benjamin-Cummings (reprinted Princeton University Press, 2000).
Wightman AS and Gärding L (1965) Fields as operator-valued distributions in relativistic quantum theory. *Arkiv för Fysik* 28: 129–189.
Wilson KG (1970) Model of coupling constant renormalization. *Physical Review D* 2: 1438–1472.
Wilson KG (1972) Renormalization of a scalar field in strong coupling. *Physical Review D* 6: 419–426.

# Contact Manifolds

**J B Etnyre**, University of Pennsylvania, Philadelphia, PA, USA



## Introduction

Contact geometry has been seen to underly many physical phenomena and is related to many other mathematical structures. Contact structures first appeared in the work of Sophus Lie on partial differential equations. They reappeared in Gibbs' work on thermodynamics, Huygens' work on geometric optics, and in Hamiltonian dynamics. More recently, contact structures have been seen to have relations with fluid mechanics, Riemannian geometry, and low-dimensional topology, and these structures provide an interesting class of subelliptic operators.

After summarizing the basic definitions, examples, and facts concerning contact geometry, this article discusses the connections between contact geometry and symplectic geometry, Riemannian geometry, complex geometry, analysis, and dynamics. The article ends by discussing two of the most-studied connections with physics: Hamiltonian dynamics and geometric optics. References for other important topics in contact geometry (e.g., thermodynamics, fluid dynamics, holomorphic curves, and open book decompositions) are provided in the "Further reading" section.

## Basic Definitions and Examples

A hyperplane field $\xi$ on a manifold $M$ is a codimension-1 sub-bundle of the tangent bundle $TM$. Locally, a hyperplane field can always be described as the kernel of a 1-form. In other words, for every point in $M$ there is a neighborhood $U$ and a 1-form $\alpha$ defined on $U$ such that the kernel of the linear map $\alpha_x : T_xM \to \mathbb{R}$ is $\xi_x$ for all $x$ in $U$. The form $\alpha$ is called a local defining form for $\xi$. A contact structure on a $(2n+1)$-dimensional manifold $M$ is a "maximally nonintegrable hyperplane field" $\xi$. The hyperplane field $\xi$ is maximally nonintegrable if for any (and hence every) locally defining 1-form $\alpha$ for $\xi$ the following equation holds:

$$\alpha \wedge (\mathrm{d}\alpha)^n \neq 0 \qquad [1]$$

(this means that the form is, pointwise, never equal to 0). Geometrically, the nonintegrability of $\xi$ means that no hypersurface in $M$ can be tangent to $\xi$ along an open subset of the hypersurface. Intuitively, this means that the hyperplanes "twist too much" to be tangent to hypersurfaces (Figure 1). The pair $(M, \xi)$

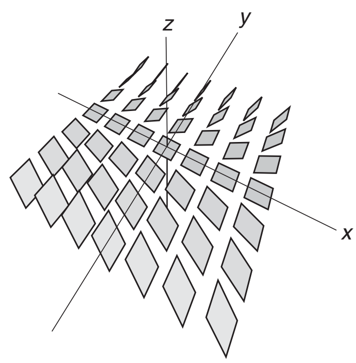



**Figure 1** The standard contact structure on $\mathbb{R}^3$ given as the kernel of $dz - y\,dx$. Courtesy of Stephan Schönenberger.

is called a contact manifold and any locally defining form $\alpha$ for $\xi$ is called a contact form for $\xi$.

**Example 1** The most basic example of a contact structure can be seen on $\mathbb{R}^{2n+1}$ as the kernel of the 1-form $\alpha = \mathrm{d}z - \sum_{i=1}^{n} y_i\, \mathrm{d}x_i$, where the coordinates on $\mathbb{R}^{2n+1}$ are $(x_1, y_1, \ldots, x_n, y_n, z)$. This example is shown in Figure 1 when $n = 1$.

**Example 2** Recall that on the cotangent space of any $n$-manifold $M$, there is a canonical 1-form $\lambda$, called the Liouville form. If $(q_1, \ldots, q_n)$ are local coordinates on $M$, then any 1-form can be expressed as $\sum_{i=1}^{n} p_i\, \mathrm{d}q_i$, so $(q_1, p_1, \ldots, q_n, p_n)$ are local coordinates on $T^*M$. In these coordinates,

$$\lambda = \sum_{i=1}^{n} p_i \pi^* \mathrm{d}q_i \qquad [2]$$

where $\pi : T^*M \to M$ is the natural projection map. The 1-jet space of $M$ is the manifold $J^1(M) = T^*M \times \mathbb{R}$ and can be considered as a bundle over $M$. The 1-jet space has a natural contact structure given as the kernel of $\alpha = \mathrm{d}z - \lambda$, where $z$ is the coordinate on $\mathbb{R}$. Note that if $M = \mathbb{R}^n$ then we recover the previous example.

**Example 3** The (oriented) projectivized cotangent space of a manifold $M$ is the set $P^*M$ of nonzero covectors in $T^*M$ where two covectors are identified if they differ by a positive real number, that is,

$$P^*M = (T^*M \setminus \{0\})/\mathbb{R}_+ \qquad [3]$$

where $\{0\}$ is the zero section of $T^*M$ and $\mathbb{R}_+$ denotes the positive real numbers. If $M$ has a metric then $P^*M$ can be easily identified with the space of unit covectors. Considering $P^*M$ as unit covectors, we can restrict the canonical 1-form $\lambda$ to $P^*M$ to get a 1-form $\alpha$ whose kernel defines a contact structure $\xi$ on $P^*M$. (Although there is no canonical contact form on $P^*M$, the contact structure $\xi$ is still well defined.) Note that if $M$ is compact then so is $P^*M$; so this gives examples of contact structures on compact manifolds.

If $\alpha$ and $\alpha'$ are two locally defining 1-forms for $\xi$, then there is a nonzero function $f$ such that $\alpha' = f\alpha$. Thus, $\alpha' \wedge (\mathrm{d}\alpha')^n = f^{n+1}\alpha \wedge (\mathrm{d}\alpha)^n$ is a nonzero top dimensional form on $M$ and if $n$ is odd then the orientation defined by the local defining form is independent of the actual form. Hence, when $n$ is odd, a contact structure defines an orientation on $M$ (this is independent of whether or not $\xi$ is orientable!). If $M$ had a preassigned orientation (and $n$ is odd), then the contact structure is called "positive" if it induces the given orientation and "negative" otherwise. One should be careful when reading the literature, as some authors build positive into their definition of contact structure, especially when $n = 1$. If there is a globally defined 1-form $\alpha$ whose kernel defines $\xi$, then $\xi$ is called transversally orientable or co-orientable. This is equivalent to the bundle $\xi$ being orientable when $n$ is odd or when $n$ is even and $M$ is orientable. In this article the discussion is restricted to transversely orientable contact structures.

Suppose that $\alpha$ is a contact form for $\xi$, then eqn [1] implies that $\mathrm{d}\alpha|_\xi$ is a symplectic form on $\xi$. This is one sense in which a contact structure is like an odd-dimensional analog of a symplectic structure.

A submanifold $L$ of a contact manifold $(M, \xi)$ is called Legendrian if $\dim M = 2\dim L + 1$ and $T_pL \subset \xi_p$.

**Example 4** A fiber in the unit cotangent bundle with the contact structure from Example 3 is a Legendrian sphere.

**Example 5** Let $f : M \to \mathbb{R}$ be a function. Then $j_1(f)(q) = (q, \mathrm{d}f_q, f(q))$ is a section of the 1-jet space $J^1(M)$ of $M$; it is called the 1-jet of $f$. If $s$ is any section of the 1-jet space, then it is Legendrian if and only if it is the 1-jet of a function.

This observation is the basis for Lie's study of partial differential equations. More specifically, a first-order partial differential equation on $M$ can be considered as giving an algebraic equation on $J^1(M)$. Then, a section of $J^1(M)$ satisfying this algebraic equation corresponds to the 1-jet of a solution to the original partial differential equation if and only if it is Legendrian.

Recently, Legendrian submanifolds have been much studied. There are various classification results in three dimensions and several striking existence results in higher dimensions.

## Local Theory

The natural equivalence between contact structures is contactomorphism. Two contact structures $\xi_0$ and

$\xi_1$ on manifolds $M_0$ and $M_1$, respectively, are contactomorphic if there is a diffeomorphism $f: M_0 \to M_1$ such that $f_*(\xi_0) = \xi_1$. All contact structures are locally contactomorphic. In particular, we have the following theorem.

**Theorem 1** (Darboux's Theorem). *Suppose $\xi_i$ is a contact structure on the manifold $M_i, i = 0, 1$, and $M_0$ and $M_1$ have the same dimension. Given any points $p_0$ and $p_1$ in $M_0$ and $M_1$, respectively, there are neighborhoods $N_i$ of $p_i$ in $M_i$ and a contactomorphism from $(N_0, \xi_0|_{N_0})$ to $(N_1, \xi_1|_{N_1})$. Moreover, if $\alpha_i$ is a contact form for $\xi_i$ near $p_i$, then the contactomorphism can be chosen to pull $\alpha_1$ back to $\alpha_0$.*

Thus, locally all contact structures (and contact forms!) look like the one given in Example 1 above.

Furthermore, contact structures are "local in time." That is, compact deformations of contact structures do not produce new contact structures.

**Theorem 2** (Gray's theorem). *Let $M$ be an oriented $(2n+1)$-dimensional manifold and $\xi_t, t \in (0, 1)$, a family of contact structures on $M$ that agree off of some compact subset of $M$. Then there is a family of diffeomorphisms $\phi_t : M \to M$ such that $(\phi_t)_* \xi_t = \xi_0$.*

In particular, on a compact manifold, all deformations of contact structures come from diffeomorphisms of the underlying manifold. The theorem is not true if the contact structures do not agree off of a compact set. For example, there is a one-parameter family of noncontactomorphic contact structures on $S^1 \times R^2$.

## Existence and Classification

The existence of contact structures on closed odd-dimensional manifolds is quite difficult. However, Gromov has shown that contact structures on open manifolds obey an h-principle. To explain this, we note that if $(M^{2n+1}, \xi)$ is a co-oriented contact manifold then the tangent bundle of $M$ can be written as $\xi \oplus \mathbb{R}$ and thus the structure group of $TM$ can be reduced to $U(n)$ (since $\xi$ has a conformal symplectic structure on it). Such a reduction of the structure group is called an almost contact structure on $M$. Clearly, a contact structure on $M$ induces an almost contact structure. If $M$ is an open manifold, Gromov proved that the inclusion of the space of co-oriented contact structures on $M$ into the space of almost contact structures on $M$ is a weak homotopy equivalence. In particular, if an open manifold meets the necessary algebraic condition for the existence of an almost contact structure, then the manifold has a co-oriented contact structure.

Lutz and Martinet proved a similar, but weaker, result for oriented closed 3-manifolds. More specifically, every closed oriented 3-manifold admits a co-oriented contact structure and in fact has at least one for every homotopy class of plane field. There has been much progress on classifying contact structures on 3-manifolds and here an interesting dichotomy has appeared. Contact structures break into one of two types: tight or overtwisted. Overtwisted contact structures obey an h-principle and are in general easy to understand. Tight contact structures have a more subtle, geometric nature. In higher dimensions there is much less known about the existence (or classification) of contact structures.

## Relations with Symplectic Geometry

Let $(X, \omega)$ be a symplectic manifold. A vector field $v$ satisfying

$$L_v \omega = \omega \qquad [4]$$

(where $L_v \omega$ is the Lie derivative of $\omega$ in the direction of $v$) is called a symplectic dilation. A compact hypersurface $M$ in $(X, \omega)$ is said to have "contact type" if there exists a symplectic dilation $v$ in a neighborhood of $M$ that is transverse to $M$. Given a hypersurface $M$ in $(X, \omega)$, the characteristic line field $LM$ in the tangent bundle of $M$ is the symplectic complement of $TM$ in $TX$. (Since $M$ is codimension 1, it is coisotropic; thus, the symplectic complement lies in $TM$ and is one dimensional.)

**Theorem 3** *Let $M$ be a compact hypersurface in a symplectic manifold $(X, \omega)$ and denote the inclusion map $i: M \to X$. Then $M$ has contact type if and only if there exists a 1-form $\alpha$ on $M$ such that $d\alpha = i^* \omega$ and the form $\alpha$ is never zero on the characteristic line field.*

If $M$ is a hypersurface of contact type, then the 1-form $\alpha$ is obtained by contracting the symplectic dilation $v$ into the symplectic form: $\alpha = \iota_v \omega$. It is easy to verify that the 1-form $\alpha$ is a contact form on $M$. Thus, a hypersurface of contact type in a symplectic manifold inherits a co-oriented contact structure.

Given a co-orientable contact manifold $(M, \xi)$, its symplectization $\mathrm{Symp}(M, \xi) = (X, \omega)$ is constructed as follows. The manifold $X = M \times (0, \infty)$, and given a global contact form $\alpha$ for $\xi$ the symplectic form is $\omega = d(t\alpha)$, where $t$ is the coordinate on $\mathbb{R}$. (The symplectization is also equivalently defined as $(M \times \mathbb{R}, d(e^t \alpha))$.)

**Example 6** The symplectization of the standard contact structure on the unit cotangent bundle

(see Example 3) is the standard symplectic structure on the complement of the zero section in the cotangent bundle.

The symplectization is independent of the choice of contact from $\alpha$. To see this, fix a co-orientation for $\xi$ and note the manifold $X$ which can be identified (in many ways) with the sub-bundle of $T^*M$ whose fiber over $x \in M$ is

$$\{\beta \in T_x^*M : \beta(\xi_x) = 0 \text{ and } \beta > 0 \text{ on vectors positively transverse to } \xi_x\} \quad [5]$$

and restricting $d\lambda$ to this subspace yields a symplectic form $\omega$, where $\lambda$ is the Liouville form on $T^*M$ defined in Example 2. A choice of contact form $\alpha$ fixes an identification of $X$ with the sub-bundle of $T^*M$ under which $d(t\alpha)$ is taken to $d\lambda$.

The vector field $v = \partial/\partial t$ on $(X, \omega)$ is a symplectic dilation that is transverse to $M \times \{1\} \subset X$. Clearly, $\iota_v \omega|_{M\times\{1\}} = \alpha$. Thus, we see that any co-orientable contact manifold can be realized as a hypersurface of contact type in a symplectic manifold. In summary, we have the following theorem.

**Theorem 4** *If $(M, \xi)$ is a co-oriented contact manifold, then there is a symplectic manifold* $\mathrm{Symp}(M, \xi)$ *in which $M$ sits as a hypersurface of contact type. Moreover, any contact form $\alpha$ for $\xi$ gives an embedding of $M$ into* $\mathrm{Symp}(M, \xi)$ *that realizes $M$ as a hypersurface of contact type.*

We also note that all the hypersurfaces of contact type in $(X, \omega)$ look locally, in $X$, like a contact manifold sitting inside its symplectification.

**Theorem 5** *Given a compact hypersurface $M$ of contact type in a symplectic manifold $(X, \omega)$ with the symplectic dilation given by $v$, there is a neighborhood of $M$ in $X$ symplectomorphic to a neighborhood of $M \times \{1\}$ in* $\mathrm{Symp}(M, \xi)$ *where the symplectization is identified with $M \times (0, \infty)$ using the contact form $\alpha = \iota_v \omega|_M$ and $\xi = \ker \alpha$.*

## The Reeb Vector Field and Riemannian Geometry

Let $(M, \xi)$ be a contact manifold. Associated to a contact form $\alpha$ for $\xi$ is the Reeb vector field $v_\alpha$. This is the unique vector field satisfying

$$\iota_{v_\alpha} \alpha = 1 \quad \text{and} \quad \iota_{v_\alpha} d\alpha = 0 \quad [6]$$

One may readily check that $v_\alpha$ is transverse to the contact hyperplanes and the flow of $v_\alpha$ preserves $\xi$ (in fact, it preserves $\alpha$). These two conditions characterize Reeb vector fields; that is, a vector field $v$ is the Reeb vector field for some contact form for $\xi$ if and only if it is transverse to $\xi$ and its flow preserves $\xi$.

The fundamental question concerning Reeb vector fields asks if its flow has a (contractible) periodic orbit. A paraphrazing of the Weinstein conjecture asserts a positive answer to this question. Most progress on this conjecture has been made in dimension 3 where H Hofer has proved the existence of periodic orbits for all Reeb fields on $S^3$ and on 3-manifolds with essential spheres (i.e., embedded $S^2$'s that do not bound a 3-ball in the manifold). Relations with Hamiltonian dynamics are discussed below.

Recall, from Example 3, that a Riemannian metric $g$ on a manifold $M$ provides an identification of the (oriented) projectivized cotangent bundle $P^*M$ with the unit cotangent bundle. Considered as a subset of $T^*M$, $P^*M$ inherits not only a contact structure but also a contact form $\alpha$ (by restricting the Liouville form). Let $v_\alpha$ be the associated Reeb vector field. The metric $g$ also provides an identification of the tangent and cotangent bundles of $M$. Thus, $P^*M$ may be considered as the unit tangent bundle of $M$. Let $w_g$ be the vector field on the unit tangent bundle generating the geodesic flow on $M$.

**Theorem 6** *The Reeb vector field $v_\alpha$ is identified with geodesic flow field $w_g$ when $P^*M$ is identified with the unit tangent space using the metric $g$.*

## Relations with Complex Geometry and Analysis

Let $X$ be a complex manifold with boundary and denote the induced complex structure on $TX$ by $J$. The complex tangencies $\xi$ to $M = \partial X$ are described by the equation $d\phi \circ J = 0$, where $\phi$ is a function defined in a neighborhood of the boundary such that 0 is a regular value and $\phi^{-1}(0) = M$. The form $L(v, w) = -d(d\phi \circ J)(v, Jw)$, for $v, w \in \xi$, is called the Levi form, and when $L(v, w)$ is positive (negative) definite, then $X$ is said to have strictly pseudoconvex (pseudoconcave) boundary. The hyperplane field $\xi$ will be a contact structure if and only if $d(d\phi \circ J)$ is a nondegenerate 2-form on $\xi$ (if and only if $L(v, w)$ is definite). A well-studied source of examples comes from Stein manifolds.

**Example 7** Let $X$ be a complex manifold and again let $J$ denote the induced complex structure on $TX$. From a function $\phi : X \to \mathbb{R}$, we can define a 2-form $\omega = -d(d\phi \circ J)$ and a symmetric form $g(v, w) = \omega(v, Jw)$. If this symmetric form is positive definite, the function $\phi$ is called "strictly plurisubharmonic." The manifold $X$ is a Stein manifold if $X$

admits a proper strictly plurisubharmonic function $\phi : X \to \mathbb{R}$. An important result says that $X$ is Stein if and only if it can be realized as a closed complex submanifold of $C^n$. Clearly any noncritical level set of $\phi$ gives a contact manifold.

Contact manifolds also give rise to an interesting class of differential operators. Specifically, a contact structure $\xi$ on $M$ defines a symbol-filtered algebra of pseudodifferential operators $\Psi_\xi^*(M)$, called the "Heisenberg calculus." Operators in this algebra are modeled on smooth families of convolution operators on the Heisenberg group. An important class of operators of this type are the "sum-of-squares" operators. Locally, the highest-order part of such an operator takes the form

$$L = \sum_{j=1}^{2n} v_j^2 + \mathrm{i} a v_\alpha \qquad [7]$$

where $\{v_1, \ldots, v_{2n}\}$ is a local framing for the contact field and $v_\alpha$ is a Reeb vector field. This operator belongs to $\Psi_\xi^2(M)$ and is subelliptic for $a$ outside a discrete set.

## Hamiltonian Dynamics

Given a symplectic manifold $(X, \omega)$, a function $H : X \to \mathbb{R}$ will be called a Hamiltonian. (Only autonomous Hamiltonians are discussed here.) The unique vector field satisfying

$$\iota_{v_H} \omega = -\mathrm{d}H$$

is called the Hamiltonian vector field associated to $H$. Many problems in classical mechanics can be formulated in terms of studying the flow of $v_H$ for various $H$.

**Example 8** If $(X, \omega) = (\mathbb{R}^{2n}, \mathrm{d}\lambda)$, where $\lambda$ is from Example 2, then the flow of the Hamiltonian vector field is given by

$$\dot{q} = \frac{\partial H}{\partial p}, \qquad \dot{p} = -\frac{\partial H}{\partial q}$$

A standard fact says that the flow of $v_H$ preserves the level sets of $H$.

**Theorem 7** *If $M$ is a level set of $H$ corresponding to a regular value and $M$ is a hypersurface of contact type, then the trajectories of $v_H$ and of the Reeb vector field (associated to $M$ in Theorem 3) agree.*

Thus under suitable hypothesis, Hamiltonian dynamics is a reparametrization of Reeb dynamics. In particular, searching for periodic orbits in such a Hamiltonian system is equivalent to searching for periodic orbits in a Reeb flow. Thus in this context, Weinstein's conjecture asserts a positive answer to the questions: Does the Hamiltonian flow along a regular level set of contact type have a periodic orbit? Viterbo proved that the answer was yes if the hypersurface is compact and in $(\mathbb{R}^{2n}, \omega = \mathrm{d}\alpha)$. Other progress has been made by studying Reeb dynamics.

## Geometric Optics

In this section, we study the propagation of light (or various other disturbances) in a medium (for the moment, we do not specify the properties of this medium). The medium will be given by a three-dimensional manifold $M$. Given a point $p$ in $M$ and $t > 0$, let $I_p(t)$ be the set of all points to which light can travel in time $\leq t$. The wave front of $p$ at time $t$ is the boundary of this set and is denoted as $\Phi_p(t) = \partial I_p(t)$.

**Theorem 8** (Huygens' principle). *$\Phi_p(t + t')$ is the envelope of the wave fronts $\Phi_q(t')$ for all $q \in \Phi_p(t)$.*

This is best understood in terms of contact geometry. Let $\pi : (T^*M \backslash \{0\}) \to P^*M$ be the natural projection (see Example 3) and let $S$ be any smooth sub-bundle of $T^*M \backslash \{0\}$ that is transverse to the radial vector field in each fiber and for which $\pi|_S : S \to P^*M$ is a diffeomorphism. The restriction of the Liouville form to $S$ gives a contact form $\alpha$ and a corresponding Reeb vector field $v$. Given a subset $F$ of $M$ with a well-defined tangent space at every point set

$$L_F = \{p \in S : \pi(p) \in F \ \text{and} \ p(w) = 0 \ \text{for all} \ w \in T_{\pi(p)}F\} \qquad [8]$$

The set $L_F$ is a Legendrian submanifold of $S$ and is called the "Legendrian lift" of $F$. If $L$ is a generic Legendrian submanifold in $S$, then $\pi(L)$ is called the front projection of $L$ and $L_{\pi(L)} = L$. Given a Legendrian submanifold $L$, let $\Psi_t(L)$ be the Legendrian submanifold obtained from $L$ by flowing along $v$ for time $t$.

**Example 9** Given a metric $g$ on $M$, Fermat's principle says that light travels along geodesics. Thus, if $S$ is the unit cotangent bundle, then using $g$ to identify the geodesic flow with the Reeb flow one sees that light will travel along trajectories of the Reeb vector field. Given a point $p$ in $M$, the Legendrian submanifold $L_p$ is a sphere sitting in $T_p^*M$. The Huygens principle follows from the observation that $\Phi_p(t) = \pi(\Psi_t(L_p))$.

Using the more general $S$ discussed above, one can generalize this example to light traveling in a medium that is nonhomogeneous (i.e., the speed differs from point to point in $M$) and anisotropic (i.e., the speed differs depending on the direction of travel).

*See also:* Hamiltonian Fluid Dynamics; Integrable Systems and Recursion Operators on Symplectic and Jacobi Manifolds; Minimax Principle in the Calculus of Variations.

## Further Reading


Aebisher B, Borer M, Kälin M, Leuenberger Ch, and Reimann HM (1994) *Symplectic Geometry*, Progress in Mathematics, vol. 124. Basel: Birkhäuser.

Arnol'd VI (1989) *Mathematical Methods of Classical Mechanics*, Graduate Texts in Mathematics, vol. 60, xvi+516, pp. 163–179. New York: Springer.

Arnol'd VI (1990) *Contact Geometry: The Geometrical Method of Gibbs's Thermodynamics*, Proceedings of the Gibbs Symposium. (New Haven, CT, 1989), pp. 163–179. Providence, RI: American Mathematical Society.

Beals R and Greiner P (1988) Calculus on Heisenberg manifolds. *Annals of Mathematics Studies* 119.

Eliashberg Y, Givental A, and Hofer H (2000) *Introduction to Symplectic Field Theory*, GAFA 2000 (Tel Aviv, 1999), Geom. Funct. Anal. 2000, Special Volume, Part II, pp. 560–673.

Etnyre J. Legendrian and transversal knots. *Handbook of Knot Theory* (in press).

Etnyre J (1998) *Symplectic Convexity in Low-Dimensional Topology*, Symplectic, Contact and Low-Dimensional Topology (Athens, GA, 1996), Topology Appl., vol. 88, No. 1–2, pp. 3–25.

Etnyre J and Ng L (2003) *Problems in Low Dimensional Contact Topology*, Topology and Geometry of Manifolds (Athens, GA, 2001), pp. 337–357, Proc. Sympos. Pure Math., vol. 71. Providence, RI: American Mathematical Society.

Geiges H Contact geometry. *Handbook of Differential Geometry*, vol. 2 (in press).

Geiges H (2001a) *Contact Topology in Dimension Greater than Three*, European Congress of Mathematics, vol. II (Barcelona, 2000), Progress in Mathematics, vol. 202, pp. 535–545. Basel: Birkhäuser.

Geiges H (2001b) A brief history of contact geometry and topology. *Expositiones Mathematicae* 19(1): 25–53.

Ghrist R and Komendarczyk R (2001) Topological features of inviscid flows. *An Introduction to the Geometry and Topology of Fluid Flows* (Cambridge, 2000), 183–201, NATO Sci. Ser. II Math. Phys. Chem., vol. 47. Dordrecht: Kluwer Academic.

Giroux E (2002) *Géométrie de contact: de la dimension trois vers les dimensions sup*érieures, Proceedings of the International Congress of Mathematicians, vol. II (Beijing, 2002), pp. 405–414. Beijing: Higher Ed. Press.

Hofer H and Zehnder E (1994) *Symplectic Invariants and Hamiltonian Dynamics*, Birkhäuser Advanced Texts: Basler Lehrbücher, pp. xiv+341. Basel: Birkhäuser.

Taylor ME (1984) *Noncommutative Microlocal Analysis, Part I*, Mem Amer. Math. Soc., 52, no. 313. American Mathematical Society.


# Control Problems in Mathematical Physics

**B Piccoli**, Istituto per le Applicazioni del Calcolo, Rome, Italy



## Introduction

Control Theory is an interdisciplinary research area, bridging mathematics and engineering, dealing with physical systems which can be "controlled," that is, whose evolution can be influenced by some external agent. A general model can be written as

$$y(t) = A(t, y(0), u(\cdot)) \qquad [1]$$

where $y$ describes the state variables, $y(0)$ the initial condition, and $u(\cdot)$ the control function. Thus, eqn [1] means that the state at time $t$ depends on the initial condition but also on some parameters $u$ which can be chosen as function of time. To be precise, there are some control problems which are not of evolutionary type; however, in this presentation we restrict ourselves to this case.

One has to distinguish among the control set $U$ where the control function can take values: $u(t) \in U$, and the space of control functions, $\mathcal{U}$, to which each control function should belong: $u(\cdot) \in \mathcal{U}$. Thus, for example, we may have $U = R^m$ and $\mathcal{U} = L^\infty([0, T], R^m)$.

There are various problems one can formulate regarding systems of type [1], among which:

*Controllability* Given any two states $y_0$ and $y_1$ determine a control function $u(\cdot)$ such that for some time $t > 0$ we have $y_1 = A(t, y_0, u(\cdot))$.

*Optimal control* Consider a cost function $J(y(\cdot), u(\cdot))$ depending both on the evolutions of $y$ and $u$ and determine a control function $\tilde{u}(\cdot)$ and a trajectory $\tilde{y}(t) = A(t, y_0, \tilde{u}(\cdot))$ such that $\tilde{y}(\cdot)$ steers the system from $y_0$ to $y_1$, as before, and the cost $J$ is minimized (or maximized).

*Stabilization* We say that $\bar{y}$ is an equilibrium if there exists $\bar{u} \in U$ such that $A(t, \bar{y}, \bar{u}) = \bar{y}$ for every $t > 0$ (here $\bar{u}$ indicates also the constant in time control function). Determine the control $u$ as function of the state $y$ so that $\bar{y}$ is a (Lyapunov) stable equilibrium for the uncontrolled dynamical system $y(t) = A(t, y(0), u(y(\cdot)))$.

*Observability* Assume that we can observe not the state $y$, but a function $\phi(y)$ of the state. Determine conditions on $\phi$ so that the state $y$ can be reconstructed from the evolution of $\phi(y)$ choosing $u(\cdot)$ suitably.

For the sake of simplicity, we restrict ourselves mainly to the first two problems and just mention

some facts about the others. Also, we focus on two cases:

*Control of ordinary differential equations* (*ODEs*) In this case $t \in R, y \in R^n$, $U$ is a set, typically $U \subset R^m$, and $A$ is determined by a controlled ODE

$$\dot{y} = f(t, y, u) \qquad [2]$$

A typical example in mathematical physics is the control of mechanical systems (Bloch 2003, Bullo and Lewis 2005).

*Control of partial differential equations* (*PDEs*) In this case $t \in R, x \in R^n, y(x)$ belongs to a Banach functional space, for example, $H^s(R^{n+1}, R), U$ is a functional space, and $A$ is determined by a controlled PDE,

$$F(t, x, y, y_t, y_{x_1}, \ldots, y_{x_n}, y_t, \ldots, u) = 0 \qquad [3]$$

A typical example in mathematical physics is the control of wave equation using boundary conditions, see below.

There are various other possible situations we do not treat here: "stochastic control," when $y$ is a random variable and $A$ defined by a (controlled) stochastic differential equation; "discrete time control," where $t \in N$; "hybrid control," where $t$ and $y$ may have both discrete and continuous components, and so on.

As shown above, the control law can be assigned in (at least) two basically different ways. In open-loop form, as a function of time: $t \to u(t)$, and in closed-loop form or feedback, as a function of the state: $y \to u(y)$. For example, in optimal control we look for a control $\tilde{u}(t)$ in open-loop form, while in stabilization we search for a feedback control $u(y)$. The open-loop control depends on $y(0)$, while a feedback control can stabilize regardless of the initial condition.

**Example 1** A point with unit mass moves along a straight line; if a controller is able to apply an external force $u$, then, calling $y_1(t), y_2(t)$, respectively, the position and the velocity of the point at time $t$, the motion is described by the control system

$$(\dot{y}_1, \dot{y}_2) = (y_2, u) \qquad [4]$$

It is easy to check that the feedback control $u(y_1, y_2) = -y_1 - y_2$ stabilizes the system asymptotically to the origin, that is, for every initial data $(\bar{y}_1, \bar{y}_2)$, the solution of the corresponding Cauchy problem satisfies $\lim_{t\to\infty}(y_1, y_2)(t) = (0, 0)$.

Another simple problem consists in driving the point to the origin with zero velocity in minimum time from given initial data. It is quite easy to see that the optimal strategy is to accelerate towards the

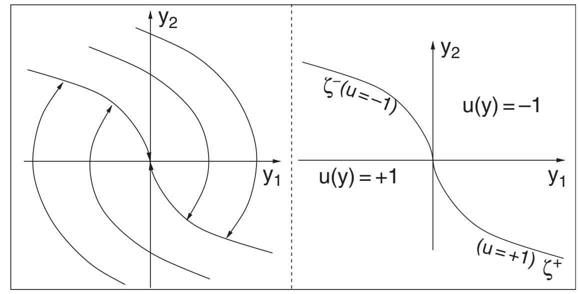



**Figure 1** Example 1. The simplest example of (a) optimal synthesis and (b) corresponding feedback.

origin with maximum force on some interval $[0, t]$ and then to decelerate with maximum force to reach the origin at velocity zero. The set of optimal trajectories is depicted in Figure 1a: they can be obtained using the following discontinuous feedback, see Figure 1b. Define the curves $\zeta^{\pm} = \{(y_1, y_2) : \mp y_2 > 0, y_1 = \pm y_2^2\}$ and let $\zeta$ be defined as the union $\zeta^{\pm} \cup \{0\}$. We define $A^+$ to be the region below $\zeta$ and $A^-$ the one above. Then the feedback is given by

$$u(x) = \begin{cases} +1 & \text{if } (y_1, y_2) \in A^+ \cup \zeta^+ \\ -1 & \text{if } (y_1, y_2) \in A^- \cup \zeta^- \\ 0 & \text{if } (y_1, y_2) = (0, 0) \end{cases}$$

**Example 2** Consider a (one-dimensional) vibrating string of unitary length with a fixed endpoint. The model for the motion of the displacement of the string with respect to the rest position is given by

$$y_{tt} + \Delta y = 0, \qquad y(t, 0) = 0 \qquad [5]$$

with initial data

$$y(0, \cdot) = y_0, \qquad y_t(0, \cdot) = y_1 \qquad [6]$$

Assume that we can control the position of the second endpoint; then,

$$y(t, 1) = u(t) \qquad [7]$$

for some control function $u(\cdot) \in R$.

Let us introduce another key concept: the reachable set at time $t$ from $\bar{y}$ is the set

$$R(t; \bar{y}) = \{A(t, \bar{y}, u(\cdot)) : u(\cdot) \in \mathcal{U}\}$$

Various problems can be formulated in terms of reachable sets, for example, controllability requires that for every $\bar{y}$ the union of all $R(t; \bar{y})$ as $t \to \infty$ includes the entire space. The dependence of $R(t; \bar{y})$ on time $t$ and on the set of controls $\mathcal{U}$ is also a subject of investigation: one may ask whether the same points in $R(t; \bar{y})$ can be reached by using controls which are piecewise constant, or take values within some subsets of $U$.

## Control of ODEs

For most proofs we refer to Agrachev and Sachkov (2004) and Sontag (1998).

### Controllability

Consider first the case of a linear system:

$$\dot{y} = Ay + Bu, \quad u \in U, \ \ y(0) = y_0 \qquad [8]$$

where $y, y_0 \in R^n$, $U \subset R^m$, $A$ is an $n \times n$ matrix and $B$ an $n \times m$ matrix. We have the following property of reachable sets:

**Theorem 1** *If $U$ is compact convex then the reachable set $R(t)$ for [8] is compact and convex.*

A control system [8] is controllable if taking $U = R^m$ we have $R(t) = R^n$ for every $t > 0$. By linearity, this is equivalent to requiring the reachable set to be a neighborhood of the origin in case of bounded controls. Define the controllability matrix to be the $n \times nm$ matrix

$$C(A, B) = (B, AB, \ldots, A^{n-1}B)$$

Controllability is characterized by the following:

**Theorem 2** (Kalman controllability theorem). *The linear system [8] is controllable if and only if* $\mathrm{rank}(C(A, B)) = n$.

For linear systems, there exists a duality between controllability and observability in the sense of the following theorem:

**Theorem 3** *Consider the linear control system [8] and assume to observe the variable $z(y) = Cy$ for some $p \times n$ matrix $C$. Then, observability holds if and only if the linear system $\dot{y} = A^t y + C^t v$ is controllable.*

There exists no characterization of controllability for nonlinear systems as for linear ones, but we have the linearization result:

**Theorem 4** *A nonlinear system is locally controllable if its linearization is. The converse is false.*

There are many results for the important class of control–affine systems

$$\dot{y} = f_0(y) + \sum_{i=1}^{m} f_i(y) u_i \qquad [9]$$

where $f_0, \ldots, f_m$ are smooth vector fields on $R^n$ and $U = R^m$. In general, there exists no explicit representation for the trajectories of [9], in terms of integrals of the control as it happens for linear systems. Still, a rich mathematical theory has been developed applying techniques and ideas from differential geometry: the so-called geometric control theory. The main idea is that controllability (and properties of optimal trajectories) is determined by the Lie algebra generated by vector fields $f_i$. For example:

**Theorem 5** (Lie-algebraic rank condition). *Let $\mathcal{L}$ be the Lie algebra generated by the vector fields $f_i, i = 1, \ldots, m$, and assume $f_0 = 0$. If $\mathcal{L}(y)$ is of dimension $n$ at every point $y$ then the system is controllable.*

We refer to Agrachev and Sachkov (2004) and Jurdjevic (1997) for general presentation of geometric control theory and give a simple example to show how Lie brackets characterize reachable directions.

**Example 3** Consider the Brockett integrator

$$\dot{y}_1 = u_1, \qquad \dot{y}_2 = u_2, \qquad \dot{y}_3 = u_1 y_2 - u_2 y_1$$

Starting from the origin, using constant controls, we can move along curves tangent to the $y_1 y_2$ plane. However, let $f_1 = (1, 0, y_2)$ and $f_2 = (0, 1, -y_1)$ (fields corresponding to constant controls); then their Lie bracket is given by

$$[f_1, f_2](0) = (Df_2 \cdot f_1 - Df_2 \cdot f_2)(0) = (0, 0, -2)$$

Moving for time $t$ first along the integral curve of $f_1$, then of $f_2$, then of $-f_1$, and finally of $-f_2$, we reach a point $t^2[f_1, f_2](0) + o(t^2)$ along the vertical direction $y_3$. This corresponds to say that the system satisfies LARC.

### Optimal Control

The theory of optimal control has developed in three main directions:

*Existence* of optimal controls, under various assumptions on $L, f, U$. When the sets $F(t, y)$ are convex, optimal solutions can be constructed following the direct method of Tonelli for the calculus of variations, that is, as limits of minimizing sequences: the two main ingredients are compactness and lower-semicontinuity. If convexity does not hold, existence is not granted in general but for special cases.

*Necessary conditions* for the optimality of a control $u(\cdot)$. The major result in this direction is the celebrated "Pontryagin maximum principle" (PMP) which extends the Euler–Lagrange equation to control systems, and the Weierstrass necessary conditions for a strong local minimum in the calculus of variations. Various extensions and other necessary conditions are now available (Agrachev and Sachkov 2004).

*Sufficient conditions* for optimality. The standard procedure resorts to embedding the optimal control problem in a family of problems, obtained by

varying the initial conditions. One defines the value function $V$ by

$$V(t, \bar{y}) = \inf J(y(\cdot), u(\cdot))$$

where the inf is taken over the set of trajectories and controls satisfying $y(t) = \bar{y}$. Under suitable assumptions, $V$ is the solution to a first-order Hamilton–Jacobian PDE. The lack of regularity of the value function $V$ has long provided a major obstacle to a rigorous mathematical analysis, solved by the theory of viscosity solutions (Bardi and Capuzzo Dolcetta 1997). Another method consists in building an optimal synthesis, that is, a collection of trajectory–control pairs.

**Pontryagin maximum principle** Consider a general autonomous control system:

$$\dot{y} = f(y, u) \qquad [10]$$

where $y \in R^n$ and $u \in U$ compact subset of $R^m$. We assume to have regularity of $f$ guaranteeing existence and uniqueness of trajectories for every $u(\cdot) \in \mathcal{U}$. For a fixed $T > 0$, an optimal control problem in Mayer form is given by

$$\min_{u(\cdot)\in\mathcal{U}} \psi(y(T, u)), \quad y(0) = \bar{y} \qquad [11]$$

where $\psi$ is the final cost and $\bar{y}$ the initial condition. More generally, one can consider also the Lagrangian cost $\int L(y, u)\mathrm{d}t$ and reduce to this case by adding a variable $y_0(0) = 0$ and $\dot{y}_0 = L$.

The well-known PMP provides, under suitable assumptions, a necessary condition for optimality in terms of a lift of the candidate optimal trajectory to the cotangent bundle. For problems as [11], PMP can be stated as follows:

**Theorem 6** *Let $u^*(\cdot)$ be a (bounded) admissible control whose corresponding trajectory $y^*(\cdot) = y(\cdot, u^*)$ is optimal. Call $p : [0, T] \mapsto R^n$ the solution of the adjoint linear equation*

$$\begin{aligned} \dot{p}(t) &= -p(t) \cdot D_y f(y^*(t), u^*(t)) \\ p(T) &= \nabla\psi(y^*(T)) \end{aligned} \qquad [12]$$

*Then the maximality condition*

$$p(t) \cdot f(y^*(t), u^*(t)) = \max_{\omega\in U} p(t) \cdot f(y^*(t), \omega) \qquad [13]$$

*holds for almost every time $t \in [0, T]$.*

Notice that the conclusion of the theorem can be interpreted by saying that the pair $(y, p)$ satisfies the system:

$$\dot{y} = \frac{\partial H(y^*, p, u^*)}{\partial p}, \qquad \dot{p} = -\frac{\partial H(y^*, p, u^*)}{\partial y}$$

where $H(y, p, u) = \langle p, f(y, u)\rangle$. This is a pseudo–Hamiltonian system, since $H$ also depends on $u^*$. Alternatively, one can define the maximized Hamiltonian

$$\mathcal{H}(y, p) = \max_u \langle p, f(y, u)\rangle$$

but $\mathcal{H}$ may fail to be smooth. Another difficulty lies in the fact that an initial condition is given for $y$ and a final condition is given for $\lambda$.

The proof of PMP relies on a special type of variations, called needle variations, of a reference trajectory. Given a candidate optimal control $u^*$ and corresponding trajectory $y^*$, a time $\tau$ of approximate continuity for $f(y^*(\cdot), u^*(\cdot))$ and $\omega \in U$, a needle variation is a family of controls $u_\varepsilon$ obtained by replacing $u^*$ with $\omega$ on the interval $[\tau - \varepsilon, \tau]$. A needle variation gives rise to a variation $v$ of the trajectory satisfying the variational equation

$$\dot{v}(t) = D_y f(y^*(t), u^*(t)) \cdot v(t) \qquad [14]$$

in classical sense only after time $\tau$. Recently Piccoli and Sussmann (2000) introduced a setting in which needle and other variations happen to be differentiable.

One may also consider some final (or initial) constraint:

$$(T, y(T)) \in S \qquad [15]$$

where $S \subset R \times R^n$ (and $T$ not fixed). In this case, the final condition for $p$ is more complicated as well as the proof of PMP. It is interesting to note the many connections between PMP and classical mechanics framework well illustrated by Bloch (2003) and Jurdjevic (1997).

**Value function and HJB equation** In this section we consider the minimization problem

$$\inf_{u\in\mathcal{U}} \psi(T, y(T, u)) \qquad [16]$$

for the control system

$$\dot{y} = f(t, y, u), \quad u(t) \in U \quad \text{a.e.} \qquad [17]$$

subject to the terminal constraints [15], where $S \subset R^{n+1}$ is a closed target set.

**Theorem 7** (PDE of dynamic programming). *Assume that the value function $V$, for [15]–[17], is $C^1$ on some open set $\Omega \subseteq R \times R^n$, not intersecting the target set $S$. Then $V$ satisfies the Hamilton–Jacobi equation*

$$\begin{aligned} &V_s(s, y) + \min_{\omega\in U}\{V_y(s, y) \cdot f(s, y, \omega)\} = 0 \\ &\forall (s, y) \in \Omega \end{aligned} \qquad [18]$$

Equation [18] is called the Hamilton–Jacobi–Bellman (HJB) equation, after Richard Bellman. In general,

however, $V$ fails to be differentiable: this is the case for Example 1 along the lines $\zeta^{\pm}$. To isolate $V$ as the unique solution of the HJB equation, one has to resort to the concept of viscosity solution. The dynamic programming and HJB equation apparatus applies also to stochastic problems for which the equation happens to be parabolic, because of the Ito formula.

**Optimal syntheses** Roughly speaking, an optimal synthesis is a collection of optimal trajectories, one for each initial condition $\bar{y}$. Geometric techniques provide a systematic method to construct syntheses:

**Step 1** Study the properties of optimal trajectories via PMP and other necessary conditions.
**Step 2** Determine a (finite-dimensional) sufficient family for optimality, that is, a class of trajectories (satisfying PMP) containing all possible optimal ones.
**Step 3** Construct a synthesis selecting one trajectory for every initial condition in such a way as to cover the state space in a regular fashion.
**Step 4** Prove that the synthesis of Step 3 is indeed optimal.

One of the main problems in step 2 is the possible presence of optimal controls with an infinite number of discontinuities, known as Fuller phenomenon. The key concept of regular synthesis, of step 3, was introduced by Boltianskii and recently refined by Piccoli and Sussmann (2000) to include Fuller phenomena. The above strategy works only in some special cases, for example for two-dimensional minimum-time problems (Boscain and Piccoli 2004): we report below an example.

**Example 4** Consider the problem of orienting in minimum time a satellite with two orthogonal rotors: the speed of one rotor is controlled, while the second rotor has constant speed. This problem is modelled by a left-invariant control system on SO(3):

$$\dot{y} = y(F + uG), \quad y \in \mathrm{SO}(3), \quad |u| \leq 1$$

where $F$ and $G$ are two matrices of so(3), the Lie algebra of SO(3). Using the isomorphism of Lie algebras $(\mathrm{SO}(3), [.,.]) \sim (R^3, \times)$, the condition that the rotors are orthogonal reads: $\mathrm{trace}(F \cdot G) = 0$. If we are interested to orient only a fixed semi-axis then we project the system on the sphere $S^2$:

$$\dot{y} = y(F + uG), \quad y \in S^2, \quad |u| \leq 1$$

In this case, $F + G$ and $F - G$ are rotations around two fixed axes and, if the angle between these two axes is less than $\pi/2$, every optimal trajectory is a finite concatenation of arcs corresponding to constant control +1 or –1. The "optimal synthesis" can be obtained by the feedback shown in **Figure 2**.

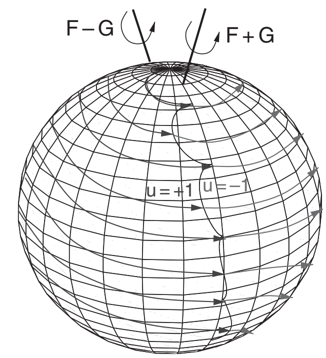



**Figure 2** Optimal feedback for Example 4.

## Control of PDEs

The theory for control of models governed by PDEs is, as expected, much more ramified and much less complete. An exhaustive resume of the available results is not possible in short space, thus we focus on Example 2 and few others to illustrate some techniques to treat control problems and give various references (see also Fursikov and Imanuvilov (1996), Komornik (1994), and Lasiecka and Triggiani (2000), and references therein).

Besides the variety of control problems illustrated in the Introduction, for PDE models one can consider different ways of applying the control, for example:

*Boundary control* One consider the system [3] (with $F$ independent of $u$) and impose the condition $y(t,x) = u(t,x)$ to hold for every time $t$ and every $x$ in some region. Usually, we assume $y(t)$ to be defined bounded region $\Omega$ and the control acts on some set $\Gamma \subset \partial\Omega$. Obviously, also Neumann conditions are natural as $\partial_\nu y = u$ where $\nu$ is the exterior normal to $\Omega$.

*Internal control* One consider the system [3] with $F$ depending on $u$. Thus, the control acts on the equation directly.

*Other controls* There are various other control problems one may consider as Galerkin-type approximation and control of some finite family of modes. An interesting example is given by Coron (2002), where the position of a tank is controlled to regulate the water level inside.

### Control of a Vibrating String

We consider Example 2, but various results hold for hyperbolic linear systems in general. First consider the uncontrolled system

$$z_{tt} = \Delta z, \quad z(0,t) = z(1,t) = 0 \qquad [19]$$

A first integral is the energy given by

$$E(t) = \frac{1}{2} \int \left[ |z_x|^2 + |z_t|^2 \right] \mathrm{d}x$$

Then we say that the system [19] is observable at time $T$ if there exists $C(T)$ such that

$$E(0) \leq C(T) \int_0^T |z_x(1,t)|^2 \, dt$$

which means that if we observe zero displacement on the right end for time $T$ then the solution has zero energy and hence vanishes. In this case, the system is observable for every time $T \geq 2$: this is precisely the time taken by a wave to travel from the right end point to the left one and backward.

Thanks to a duality as for the finite-dimensional case, observability of [19] is equivalent to null controllability for [5]–[7], that is, to the property that for every initial conditions $y_0, y_1$ there exists a control $u(\cdot)$ such that the corresponding solution verifies $y(x,T) = y_t(x,T) = 0$. More precisely, the desired control is given by $u(t) = \tilde{z}_x(1,t)$, where $\tilde{z}$ is the solution of [19] minimizing the functional (over $L^2 \times H^{-1}$)

$$J(z(\cdot,0), z_t(\cdot,0)) = \frac{1}{2}\int_0^T |z_x(1,t)|^2 \, dt + \int y_0 z_t(\cdot,0) dx - \int y_1 z(\cdot,0) \, dx$$

One can check that this functional is continuous and convex, and the coercivity is granted by the observability of [19]; thus, a minimum exists by the direct method of Tonelli. This is an example of the method known as Hilbert's uniqueness method introduced by Lions (1988).

In the multidimensional case, controllability can be characterized by imposing a condition on the region $\Gamma \subset \partial\Omega$ on which the control acts. More precisely, rays of geometric optics in $\Omega$ should intersect $\Gamma$ (Zuazua 2005).

If we consider infinite-time horizon $T = +\infty$ and introduce the functional

$$J = \int_0^{+\infty} \|y\|^2 \, dt + N \int u^2 \, dt \, dx$$

then the optimal control is determined as follows. If $(y, p)$ is a solution of the optimality system: [5]–[6] with $y = 0$ outside $\Gamma$ and

$$p_{tt} - \Delta p + y = 0, \ \partial_\nu p + Ny = 0 \quad \text{on } \Gamma$$
$$p = 0 \quad \text{on } \partial\Omega$$

then $u = y$ on $\Gamma$ (Lions 1988, Zuazua 2005).

### Controllability via Return Method of Coron

As we saw in Theorem 4, a nonlinear system may be controllable even if its linearization is not. In this case, controllability can be proved by the return method of Coron, which consists in finding a trajectory $y$ such that the following hold:

1. $y(0) = y(T) = 0$;
2. the linearized system around $y$ is controllable.

Then by implicit-function theorem, local controllability is granted, that is, there exits $\varepsilon > 0$ such that for every data $y_0, y_1$ of norm less than $\varepsilon$, there exists a control steering the system from $y_0$ to $y_1$ in time $T$.

This method does not give many advantages in the finite-dimensional case, but permits to obtain excellent results for PDE systems such as Euler, Navier–Stokes, Saint–Venant, and others (Coron 2002).

### Control of Schrödinger Equation

Consider the issue of designing an efficient transfer of population between different atomic or molecular levels using laser pulses. The mathematical description consists in controlling the Schrödinger equation. Many results are available in the finite-dimensional case. Finite-dimensional closed quantum systems are in fact left-invariant control systems on SU($n$), or on the corresponding Hilbert sphere $S^{2n-1} \subset \mathbf{C}^n$, where $n$ is the number of atomic or molecular levels, and powerful techniques of geometric control are available both for what concerns controllability and optimal control (Agrachev and Sachkov 2004, Boscain and Piccoli 2004, Jurdjevic 1997).

Recent papers consider the minimum-time problem with unbounded controls as well as minimization of the energy of transition. Boscain *et al.* (2002) have applied the techniques of sub-Riemannian geometry on Lie groups and of optimal synthesis on two-dimensional manifolds to the population transfer problem in a three-level quantum system driven by two external fields of arbitrary shape and frequency.

Although many results are available for finite-dimensional systems, only few controllability properties have been proved for the Schrödinger equation as a PDE, and in particular no satisfactory global controllability results are available at the moment.

## Further Reading


Agrachev A and Sachkov Y (2004) *Control from a Geometric Perspective*. Springer.

Bardi M and Capuzzo Dolcetta I (1997) *Optimal Control and Viscosity Solutions of Hamilton–Jacobi–Bellman Equations*. Boston: Birkhauser.

Bloch AM (2003) *Nonholonomic Mechanics and Control*, with the collaboration of J. Baillieul, P. Crouch and J. Marsden, with scientific input from P. S. Krishnaprasad, R. M. Murray and D. Zenkov. New York: Springer.

Boscain U and Piccoli B (2004) Optimal Synthesis for Control Systems on 2-D Manifolds. *Springer SMAI*, vol. 43. Heidelberg: Springer.

Boscain U, Chambrion T, and Gauthier J-P (2002) On the $K+P$ problem for a three-level quantum system: optimality implies resonance. *Journal of Dynamical and Control Systems* 8: 547–572.
Bullo F and Lewis AD (2005) *Geometric Control of Mechanical Systems*. New York: Springer.
Coron JM (2002) Return method: some application to flow control. *Mathematical Control Theory, Part 1, 2 (Trieste, 2001)*. In: Agrachev A (ed.) *ICTP Lecture Notes*, vol. VIII. Trieste: Abdus Salam Int. Cent. Theoret. Phys.
Fursikov AV and Imanuvilov O Yu (1996) *Controllability of Evolution Equations*. Lecture Notes Series, vol. 34. Seoul: Seoul National University.
Jurdjevic V (1997) *Geometric Control Theory*. Cambridge: Cambridge University Press.
Komornik V (1994) *Exact Controllability and Stabilization. The Multiplier Method*. Chichester: Wiley.
Lasiecka I and Triggiani R (2000) *Control theory for Partial Differential Equations: Continuous and Approximation Theories*. Cambridge: Cambridge University Press.
Lions JL (1988) Exact controllability, stabilization and perturbations for distributed systems. *SIAM Review* 30: 1–68.
Piccoli B and Sussmann HJ (2000) Regular synthesis and sufficiency conditions for optimality. *SIAM Journal of Control Optimization* 39: 359–410.
Sontag ED (1998) *Mathematical Control Theory*. New York: Springer.
Zuazua E (2005) Propagation, observation and conrol of wave approximatex by finite difference methods. *SIAM Review* 47: 197–243.

# Convex Analysis and Duality Methods

**G Bouchitté**, Université de Toulon et du Var, La Garde, France



## Introduction

Convexity is an important notion in nonlinear optimization theory as well as in infinite-dimensional functional analysis. As will be seen below, very simple and powerful tools will be derived from elementary duality arguments (which are by-products of the Moreau–Fenchel transform and Hahn–Banach theorem). We will emphasize on applications to a large range of variational problems. Some arguments of measure theory will be skipped.

## Basic Convex Analysis

In the following, we denote by $X$ a normed vector space, and by $X^*$ the topological dual of $X$. If a topology different from the normed topology is used on $X$, we will denote it by $\tau$. For every $x \in X$ and $A \subset X$, $\mathcal{V}_x$ denotes the open neighborhoods of $x$ and $\operatorname{int} A$, $\operatorname{cl} A$, respectively, the interior and the closure of $A$. We deal with extended real-valued functions $f : X \to \mathbb{R} \cup \{+\infty\}$. We denote by $\operatorname{dom} f = f^{-1}(\mathbb{R})$ and by $\operatorname{epi} f = \{(x,\alpha) \in X \times \mathbb{R} : f(x) \le \alpha\}$ the domain and the epigraph of $f$, respectively. We say that $f$ is proper if $\operatorname{dom} f \neq \emptyset$. Recall that $f$ is convex if for every $(x,y) \in X^2$ and $t \in [0,1]$, there holds

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y)$$
$$\text{(by convention } \infty + a = +\infty)$$

The notion of convexity for a subset $A \subset X$ is recovered by saying that $\chi_A$ is convex, where its indicator function $\chi_A$ is defined by setting

$$\chi_A(x) = \begin{cases} 0 & \text{if } x \in A \\ +\infty & \text{otherwise} \end{cases}$$

### Continuity and Lower-Semicontinuity

A first consequence of the convexity is the continuity on the topological interior of the domain. We refer for instance to Borwein and Lewis (2000) for a proof of

**Theorem 1** *Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be convex and proper. Assume that $\sup_U f < +\infty$, where $U$ is a suitable open subset of $X$. Then $f$ is continuous and locally Lipschitzian on all* $\operatorname{int}(\operatorname{dom} f)$.

As an immediate corollary, a convex function on a normed space is continuous provided it is majorized by a locally bounded function. In the finite-dimensional case, it is easily deduced that a finite-valued convex function $f : \mathbb{R}^d \to \mathbb{R}$ is locally Lipschitz. Furthermore, by Aleksandrov's theorem, $f$ is almost everywhere twice differentiable and the non-negative Hessian matrix $\nabla^2 f$ coincides with the absolutely continuous part of the distributional Hessian matrix $D^2 f$ (it is a Radon measure taking values in the non-negative symmetric matrices).

However, in infinite-dimensional spaces, for ensuring compactness properties (as, e.g., in condition (ii) of Theorem 4 below), we need to use weak topologies and the situation is not so simple. A major idea consists in substituting the continuity property with lower-semicontinuity.

**Definition 2** A function $f : X \to \mathbb{R} \cup \{+\infty\}$ is $\tau$-l.s.c. at $x_0 \in X$ if for all $\alpha \in \mathbb{R}$, there exists $U \in \mathcal{V}_{x_0}$ such that $f > \alpha$ on $U$. In particular, $f$ will be l.s.c. on all $X$ provided $f^{-1}((r, +\infty))$ is open for every $r \in \mathbb{R}$.

**Remark 3**

(i) The following sequential notion can be also used: $f$ is $\tau$-sequentially l.s.c. at $x_0$ if

$$\forall (x_n) \subset X \; x_n \xrightarrow{\tau} x_0 \implies \liminf_{n \to +\infty} f(x_n) \geq f(x_0)$$

It turns out that this notion (weaker in general) is equivalent to the previous one provided $x_0$ admits a countable basis of neighborhoods.

(ii) A well-known consequence of Hahn–Banach theorem is that, for convex functions, the lower-semicontinuity property with respect to the normed topology of $X$ is equivalent to the weak (or weak sequential) lower-semicontinuity.

**Theorem 4** (Existence). *Let* $f : X \to \mathbb{R} \cup \{+\infty\}$ *be proper, such that*

(i) $f$ *is* $\tau$*-l.s.c.,*
(ii) $\forall r \in \mathbb{R}, f^{-1}((-\infty, r])$ *is* $\tau$*-relatively compact.*

*Then there is* $\bar{x} \in X$ *such that* $f(\bar{x}) = \inf f$ *and* $\operatorname{argmin} f := \{x \in X | f(x) = \inf f\}$ *is* $\tau$*-compact.*

In practice, the choice of the topology $\tau$ is ruled by the condition (ii) above. For example, if $X$ is a reflexive infinite-dimensional Banach space and if $f$ is coercive (i.e., $\lim_{\|x\| \to \infty} f(x) = +\infty$), we may take for $\tau$ the weak topology (but never the normed topology). This restriction implies in practice that the first condition in Theorem 4 may fail. In this case, it is often useful to substitute $f$ with its lower-semicontinuous (l.s.c.) envelope.

**Definition 5** Given a topology $\tau$, the relaxed function $\bar{f} (= \bar{f}^\tau)$ is defined as

$$\bar{f}(x) = \sup\{g(x) | g : X \to \mathbb{R} \cup \{+\infty\},\; g \text{ is } \tau\text{-}l.s.c., g \leq f\}$$

It is easy to check that $f$ is $\tau$-l.s.c. at $x_0$ if and only if $\bar{f}(x_0) = f(x_0)$. Futhermore,

$$\bar{f}(x) = \sup_{U \in \mathcal{V}_u} \inf_U f, \quad \operatorname{epi} \bar{f} = \operatorname{cl}_{(X \times \mathbb{R})} (\operatorname{epi} f)$$

We can now state the relaxed version of Theorem 1.4.

**Theorem 6** (Relaxation). *Let* $f : X \to \mathbb{R} \cup \{+\infty\}$, *then*: $\inf f = \inf \bar{f}$. *Assume further that, for all real* $r$, $f^{-1}((-\infty, r])$ *is* $\mathcal{T}$*-relatively compact; then* $f$ *attains its minimum and* $\operatorname{argmin} f = \operatorname{argmin} \bar{f} \cap \{x \in X | f(x) = \bar{f}(x)\}$.

### Moreau–Fenchel Conjugate

The duality between $X$ and $X^*$ will be denoted by the symbol $\langle \cdot | \cdot \rangle$. If $X$ is a Euclidian space, we identify $X^*$ with $X$ via the scalar product denoted $(\cdot | \cdot)$.

**Definition 7** Let $f : X \to \mathbb{R} \cup \{+\infty\}$. The Moreau–Fenchel conjugate $f^* : X^* \to \mathbb{R} \cup \{+\infty\}$ of $f$ is defined by setting, for every $x^* \in X^*$:

$$f^*(x^*) = \sup\{\langle x | x^* \rangle - f(x) | x \in X\}$$

In a symmetric way, if $f^*$ is proper on $X^*$, we define the biconjugate $f^{**} : X \to \mathbb{R} \cup \{+\infty\}$ by setting

$$f^{**}(x) = \sup\{\langle x | x^* \rangle - f^*(x^*) | x^* \in X^*\}$$

As a consequence, the so-called Fenchel inequality holds:

$$\langle x | x^* \rangle \leq f(x) + f^*(x^*), \quad (x, x^*) \in X \times X^*$$

Notice that $f$ does not need to be convex. However, if $f$ is convex, then $f^*$ agrees with the Legendre–Fenchel transform.

**Definition 8** Let $f : X \to \mathbb{R} \cup \{+\infty\}$. The subdifferential of $f$ at $x$ is the possibly void subset of $\partial f(x) \subset X^*$ defined by

$$\partial f(x) := \{x^* \in X^* \colon f(x) + f^*(x^*) = \langle x, x^* \rangle\}$$

It is easy to check that $\partial f(x)$ is convex and weak-star closed. Moreover, if $f$ is convex and has a differential (or Gateaux derivative) $f'(x)$ at $x$, then $\partial f(x) = \{f'(x)\}$. After summarizing some elementary properties of the Fenchel transform, we give examples in $\mathbb{R}^d$ or in infinite-dimensional spaces.

**Lemma 9**

(i) $f^*$ *is convex, l.s.c. with respect to the weak star topology of* $X^*$.
(ii) $f^*(0) = -\inf f$ *and* $f \geq g \Rightarrow f^* \leq g^*$.
(iii) $(\inf_i f_i)^* = \sup_i f_i^*$, *for every family* $\{f_i\}$.
(iv) $f^{**}(x) = \sup\{g(x)$: $g$ affine continuous on $X$ and $g \leq f\}$ (*by convention, the supremum is identically* $-\infty$ *if no such* $g$ *exists*).

**Proof** (i) This assertion is a direct consequence of the fact that $f^*$ can be written as the supremum of functions $g_x$, where $g_x := \langle x | \cdot \rangle - f(x)$. Clearly, these functions are affine and weakly star-continuous on $X^*$. The assertions (ii), (iii) are trivial. To obtain (iv), it is enough to observe that an affine function $g$ of the form $g(x) = \langle x, x^* \rangle - \beta$ satisfies $g \leq f$ iff $f^*(x^*) \leq \beta$. □

**Example 1** Let $f : X \to \mathbb{R}$, be defined by

$$f(x) = \frac{1}{p} \|x\|_X^p, \quad 1 < p < +\infty$$

then,

$$f^*(x^*) = \frac{1}{p'} \|x^*\|_{X^*}^{p'}, \quad \text{with } \frac{1}{p} + \frac{1}{p'} = 1$$

whereas, for $p=1$, we find $f^* = \chi_{B^*}$, where $B^* = \{\|x^*\| \le 1\}$.

**Example 2** Let $A \in \mathbb{R}^{d^2}_{\mathrm{sym}}$ be a symmetric positive-definite matrix and let $f(x) := (1/2)(Ax\,|\,x)(x \in \mathbb{R}^d)$. Then, for all $y \in \mathbb{R}^d$, we have $f^*(y) = (1/2)(A^{-1}y\,|\,y)$. Notice that if $A$ has a negative eigenvalue, then $f^* \equiv +\infty$.

Particular examples on $R^d$ are also very popular. For instance:

**Minimal surfaces**

$$f(x) = \sqrt{1+|x|^2}$$

$$f^*(y) = \begin{cases} -\sqrt{1-|y|^2} & \text{if } |y| \le 1 \\ +\infty & \text{otherwise} \end{cases}$$

**Entropy**

$$f(x) = \begin{cases} x \log x & \text{if } x \in \mathbb{R}_+ \\ +\infty & \text{otherwise} \end{cases}, \quad f^*(y) = \exp(y-1)$$

**Example 3** Let $C \subset X$ be convex, and let $f = \chi_C$. Then,

$$f^*(x^*) = \sigma_C(c^*) = \sup_{x\in C} \langle x | x^* \rangle$$
$$\text{(support function of } C)$$

Notice that if $M$ is a subspace of $X$, then $(\chi_M)^* = \chi_{M^\perp}$. We specify now a particular case of interest.

Let $\Omega$ be a bounded open subset of $\mathbb{R}^n$. Take $X = C_0(\bar{\Omega}; \mathbb{R}^d)$ to be the Banach space of continuous functions on the compact $\bar{\Omega}$) with values in $\mathbb{R}^d$. As usual, we identify the dual $X^*$ with the space $\mathcal{M}_b(\bar{\Omega}; \mathbb{R}^d)$ of $\mathbb{R}^d$-valued Borel measures on $\bar{\Omega}$ with finite total variation. Let $K$ be a closed convex of $R^d$ such that $0 \in K$. Then $\rho^0_K(\xi) := \sup\{(\xi\,|\,z) \colon z \in K\}$ is a non-negative convex l.s.c. and positively 1-homogeneous function on $\mathbb{R}^d$ (e.g., $\rho_K$ is the Euclidean norm if $K$ is the unit ball of $\mathbb{R}^d$). Let us define $C := \{\varphi \in X \colon \varphi(x) \in K, \forall x \in \Omega\}$. Then, we have

$$(\chi_C)^*(\lambda) = \int_\Omega \rho^0_K(\lambda) := \int_\Omega \rho^0_K\left(\frac{d\lambda}{d\theta}\right)\theta(dx) \qquad [1]$$

where $\theta$ is any non-negative Radon measure such that $\lambda \ll \theta$ (the choice of $\theta$ is indifferent). In the case where $K$ is the unit ball, we recover the total variation of $\lambda$.

**Example 4** (Integral functionals). Given $1 \le p < +\infty$, $(\Omega, \mu, \mathcal{T})$ a measured space and $\varphi : \Omega \times \mathbb{R}^d \to [0, +\infty]$ a $\mathcal{T} \otimes B_{\mathbb{R}^d}$-measurable integrand. Then the partial conjugate $\varphi^*(x, z^*) := \sup\{\langle z\,|\,z^*\rangle - \varphi(x,z) \colon z \in \mathbb{R}^d\}$ is a convex measurable integrand. Let us define

$$I_\varphi : u \in (L^p_\mu)^d \to \int_\Omega \varphi(x, u(x))d\mu \in \mathbb{R} \cup \{+\infty\}$$

and assume that $I_\varphi$ is proper. Then there holds $(I_\varphi)^* = I_{\varphi^*}$, where

$$(I_\varphi)^* : v \in (L^{p'}_\mu)^d \to \int_\Omega \varphi^*(x, v(x))d\mu$$

## Duality Arguments

### Two Key Results

The first result related to the biconjugate $f^{**}$ is a consequence of the Hahn–Banach theorem. Recalling the assertion (v) of Lemma 9, we notice that the existence of an affine minorant for $f$ is equivalent to the properness of $f^*$ (i.e., $\exists x^*_0 \in X^* \colon f^*(x^*_0) < +\infty$).

**Theorem 10** *Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be convex and proper. Then*

(i) *$f$ is l.s.c. at $x_0$ if and only if $f^*$ is proper and $f^{**}(x_0) = f(x_0)$. In particular, the lower-semicontinuity of $f$ on all $X$ is equivalent to the identity $f \equiv f^{**}$.*
(ii) *If $f^*$ is proper, then $f^{**} = \bar{f}$.*

**Proof** We notice that by Lemma 9, $f^{**} \le f$ and $f^{**}$ is l.s.c (even for the weak topology). Therefore, $f^{**} \le \bar{f}$ and, moreover, $f$ is l.s.c. at $x_0$ if $f^{**}(x_0) \ge f(x_0)$. Conversely, if $f$ is l.s.c. at $x_0$, for every $\alpha_0 < f(x_0)$, there exists a neighborhood $V$ of $x_0$ such that $V \times (-\infty, \alpha_0) \cap \operatorname{epi} f = \emptyset$. It follows that $\overline{\operatorname{epi} f}$ is a proper closed convex subset of $X \times \mathbb{R}$ which does not intersect the compact singleton $\{(x_0, \alpha_0)\}$. By applying the Hahn–Banach strict separation theorem, there exists $(x^*_0, \beta_0) \in X^* \times \mathbb{R}$ such that

$$\langle x_0, x^*_0\rangle + \alpha_0\beta_0 < \langle x, x^*_0\rangle + \alpha\beta_0$$
$$\text{for all } (x, \alpha) \in \operatorname{epi} f$$

Taking $\alpha \to \infty$ and $x \in \operatorname{dom} f$, we find $\beta_0 \ge 0$. In fact, $\beta_0 > 0$ as the strict inequality above would be violated for $x = x_0$. Eventually, we obtain that $f$ is minorized by the affine continuous function $g(x) = -\langle x - x_0, x^*_0/\beta\rangle + \alpha_0$. Thus, we conclude that $f^*$ is proper and that $f^{**}(x_0) \ge \alpha_0$.

The assertion (ii) is a direct consequence of the equivalence in (i). □

**Theorem 11** *Let X be a normed space and let $f: X \to [0, +\infty]$ be a convex and proper function; assume that f is continuous at 0, then*

(i) *$f^*$ achieves its minimum on $X^*$*
(ii) *$f(0) = f^{**}(0) = -\inf f^*$*

**Proof**

(i) Let $M$ be an upper bound of $f$ on the ball $\{\|x\| \leq R\}$. Then

$$f^*(x^*) \geq \sup\{\langle x, x^*\rangle - f(x) \colon \|x\| \leq R\} \\ \geq R\|x^*\|_{X^*} - M$$

Hence, for every $r$, the set $\{x^* \in X^* \colon f^*(x^*) \leq r\}$ is bounded, thus $\tau$-relatively compact, where $\tau$ is the weak-star topology on $X^*$. By assertion (i) of Lemma 9, $f^*$ is $\tau$-l.s.c. and Theorem 4 applies.

(ii) By Theorem 10, since $f$ is convex proper and l.s.c. at $x_0 = 0$, we have $f(0) = f^{**}(0) = -\inf f^*$.

□

### Some Useful Consequences

**Proposition 12** (Conjugate of a sum). *Let $f, g: X \to \mathbb{R} \cup \{+\infty\}$ be convex such that*

$$\exists x_0 \in X : f \text{ is continuous at } x_0 \text{ and } g(x_0) < +\infty \qquad [2]$$

*Then*

(i) $$(f+g)^*(x^*) = \inf_{x_1^* + x_2^* = x^*} \{f^*(x_1^*) + g^*(x_2^*)\}$$

*(the equality holds in $\bar{\mathbb{R}}$).*

(ii) *If both sides of the equality in* (i) *are finite, then the infimum in the right-hand side is achieved.*

**Proof** Without any loss of generality, we may assume that $x^* = 0$ (we reduce to this case by substituting $g$ with $g - \langle \cdot, x^*\rangle$). We let

$$h(p) = \inf\{f(x+p) + g(x) | x \in X\}$$

Noticing that $(p, x) \mapsto f(x+p) + g(x)$ is convex, we infer that $h(p)$ is convex as well. As $h$ is majorized by the function $p \mapsto f(x_0 + p) + g(x_0)$, which by [2] continuous at 0, we deduce from Theorems 1 and 11 that $h(0) = h^{**}(0)$ and that $h^*$ achieves its infimum. Now $h(0) = \inf(f+g) = -(f+g)^*(0)$ and

$$h^*(p^*) = \sup\{\langle p, p^*\rangle - h(p) \colon p \in X\} \\ = \sup\{\langle p, p^*\rangle - f(x+p) - g(x) \colon x \in X, p \in X\} \\ = g^*(-p^*) + f^*(p^*)$$

The assertions (i), (ii) follow since $-h^{**}(0) = \min h^* = \min\{g^*(-p^*) + f^*(p^*)\}$.

□

**Proposition 13** (Composition). *Let $X, Y$ be two Banach spaces and $A: X \mapsto Y$ a linear operator with dense domain $D(A)$. Let $\Psi: Y \to \mathbb{R} \cup \{+\infty\}$ be a convex l.s.c. function and let $F \mapsto X$ be the convex functional defined by*

$$F(u) = \begin{cases} \Psi(Au) & \text{if } u \in D(A) \\ +\infty & \text{otherwise} \end{cases}$$

*Assume that there exists $u_0 \in D(A)$ such that $\Psi$ is continuous at $Au_0$. Then*

(i) *The Fenchel conjugate of F is given by*

$$\forall f \in X^*, \quad F^*(f) = \inf\{\Psi^*(\sigma) \colon \sigma \in Y^*, A^*\sigma = f\}$$

*where, if both sides of the equality are finite, the infimum on the right-hand side is achieved.*

(ii) *If, in addition, Y is reflexive and $\Psi$ is l.s.c. coercive, we have*

$$\bar{F}(u) = F^{**}(u) = \inf\{\Psi(p) | (u, p) \in \overline{G(A)}\} \qquad [3]$$

*where G(A) denotes the graph of A.*

**Proof**

(i) Define $H, K: X \times Y \to \mathbb{R} \cup \{+\infty\}$ by

$$H(u, p) = \chi_{G(A)}(u, p), \qquad K(u, p) = \Psi(p)$$

Then we have the identity $F^*(f) = (H+K)^*(f, 0)$, where the conjugate of $H + K$ is taken with respect to the duality $(X \times Y, X^* \times Y^*)$. From the assumption, $K$ is continuous at $(u_0, Au_0) \in \operatorname{dom} H$. By Proposition 12, we obtain

$$(H+K)^*(f, 0) \\ = \inf_{(g,\sigma) \in X^* \times Y^*} \{K^*(f - g, \sigma) + H^*(g, -\sigma)\}$$

After a simple computation, it is easy to check that

$$H^*(g, -\sigma) = \begin{cases} 0 & \text{if } A^*\sigma = f \\ +\infty & \text{otherwise} \end{cases}$$
$$K^*(f - g, \sigma) = \begin{cases} \Psi^*(\sigma) & \text{if } g = f \\ +\infty & \text{otherwise} \end{cases}$$

(ii) Let $J(u) := \inf\{\Psi(p) \colon (u, p) \in \overline{G(A)}\}$. As observed for $F^*$ in the proof of (i), we have the identity $J^*(f) = (H+K)^*(f, 0)$. Therefore, in view of Theorem 10, $\bar{F} = F^{**} = J^{**}$ and it is enough to prove that $J$ is convex l.s.c. proper. Let us consider a sequence $(u_n)$ in $X$ converging to some $u \in X$. Without any loss of generality, we may assume that $\liminf J(u_n) = \lim J(u_n) < +\infty$. Then there is a sequence $(p_n)$ such that, for every $n$, $(u_n, p_n) \in \bar{G}(A)$ and $J(u_n) \geq \psi(u_n) - 1/n$. As $\psi$ is coercive, $\{p_n\}$ is bounded in the reflexive space $Y$ and possibly passing to a subsequence,

we may assume that $p_n$ converges weakly to some $p$. Since $\overline{G(A)}$ is a (weakly) closed subspace of $X \times Y$, we infer that $(u, p)$ as the limit of $(u_n, p_n)$ still belongs to $\overline{G(A)}$. Thus, we conclude, thanks to the (weak) lower-semicontinuity of $\Psi$

$$\liminf_n J(u_n) = \lim_n \Psi(p_n) \geq \Psi(p) \geq J(u)$$

□

An immediate consequence of Propositions 12 and 13 is the following variant:

**Proposition 14** *Under the same notation as in Proposition 13, let $\Phi : X \to \mathbb{R} \cup \{+\infty\}$ be a convex function and assume that there exists $u_0 \in D(A)$ such that $F(u_0) < +\infty$ and $\Psi$ is continuous at $Au_0$. Then we have*

$$\inf_{u \in X}\{\phi(u) + \Psi(Au)\} = \sup_{\sigma \in Y^*}\{-\phi^*(-A^*\sigma) - \Psi^*(\sigma)\}$$

*where the supremum on the right-hand side is achieved. Furthermore, a pair $(\bar{u}, \bar{\sigma})$ is optimal if and only if it satisfies the relations: $\bar{\sigma} \in \partial\Psi(A\bar{u})$ and $-A^*\bar{\sigma} \in \partial\phi(\bar{u})$.*

**Remark 15** From the assertion (ii) of Proposition 13, we may conclude that $F$ is l.s.c. whenever the operator $A$ is closed. If now $A$ is merely closable (with closure denoted by $\bar{A}$), we obtain

$$\bar{F}(u) = \begin{cases} G(\bar{A}u) & \text{if } u \in \text{ dom } \bar{A} \\ +\infty & \text{otherwise} \end{cases}$$

This is the typical situation when $F$ is an integral functional defined on smooth functions of the kind

$$F(u) = \int_\Omega f(x, \nabla u)\, dx$$

where $\Omega$ is an bounded open subset of $\mathbb{R}^n, f : \Omega \times \mathbb{R}^n \to \mathbb{R}$ is a convex integrand with quadratic growth (i.e., $c|z|^2 \leq f(x,z) \leq C(1 + |z|^2$ for suitables $C \geq c > 0$). Then $X = L^2(\Omega), Y = L^2(\Omega; \mathbb{R}^n)$,

$$G(v) = \int_\Omega f(x, v(x))\, dx$$

and $A : u \in C^1(\Omega) \mapsto \nabla u \in L^2(\Omega; \mathbb{R}^n)$. It turns out that $A$ is closable and that the domain of $\bar{A}$ characterizes the Sobolev space $W^{1,2}(\Omega)$ on which $\bar{A}$ coincides with the distributional gradient operator.

The situation is more involved if we consider

$$F(u) = \int_\Omega f(x, \nabla u)\, d\mu$$

$\mu$ is a possibly concentrated Radon measure supported on $\Omega$. In general, the operator $A : u \in C^1(\Omega) \subset L^2_\mu(\Omega) \mapsto \nabla u \in L^2_\mu(\Omega; \mathbb{R}^n)$ is not closable and we need to come back to the general formula [3]. The general structure of $\overline{G(A)}$ has been given in Bouchitté *et al.* (1997) and Bouchitté and Fragalà (2002, 2003), namely

$$(u, \xi) \in \overline{G(A)} \Longleftrightarrow u \in W^{1,2}_\mu, \exists \eta \in L^2_\mu(\Omega; \mathbb{R}^n):$$
$$\xi = \nabla_\mu u + \eta, \ \eta(x) \in T_\mu(x)^\perp$$

where $T_\mu(x), \nabla_\mu(x)$ are suitable notions of tangent space and tangential gradient with respect to $\mu$, and $W^{1,2}_\mu$ denotes the domain of the extended tangential gradient operator.

**Remark 16** The assertion (ii) of Proposition 13 is not valid in the nonreflexive case. In particular, for

$$F(u) = \int_\Omega f(x, \nabla u) dx$$

where $f(x, \cdot)$ has a linear growth at infinity, we need to take $Y$ as the space of $\mathbb{R}^n$-values vector measures on $\Omega$ and the relaxed functional $F^{**}$ needs to be indentified on the space $BV(\Omega)$ of integrable functions with bounded variations. The computation of $F^{**}$ is a delicate problem for which we refer to Bouchitté and Dal Maso (1993) and Bouchitté and Valadier (1998).

**Remark 17** By duality techniques, it is possible also to handle variational integrals of the kind

$$F(u) = \int_\Omega f(x, u(x), \nabla u(x)) dx$$

even if the dependence of $f(x, u, z)$ with respect to $u$ is nonconvex. The idea consists in embedding the space $BV(\Omega)$ in the larger space $BV(\Omega \times \mathbb{R})$ through the map $u \mapsto 1_u$, where $1_u$ is the characteristic function defined on $\Omega \times R$ by setting

$$1_u(x, t) := \begin{cases} 1 & \text{if } u(x) > t \\ 0 & \text{otherwise} \end{cases}$$

Then it is possible to show, under suitable conditions on the integrand $f$, that there exists a convex l.s.c., 1-homogeneous functional $G : BV(\Omega \times \mathbb{R}) \to \mathbb{R} \cup \{+\infty\}$ such that $\bar{F}(u) = G(1_u)$. This functional $G$ is constructed as in the Example 3 taking $C$ to be a suitable convex subset of $C^0(\Omega \times \mathbb{R})$. This nice new idea has been the key tool of the calibration method developed recently (Alberti *et al.* 2003).

## Convex Variational Problems in Duality

### Finite-Dimensional Case

We sketch the duality scheme in two cases.

**Linear programming** Let $c \in \mathbb{R}^n, b \in \mathbb{R}^m$ and $A$ an $m \times n$ matrix. We denote by $A^{\mathrm{T}}$ the transpose matrix. We consider the linear program

$$(\mathcal{P}) \qquad \inf\{(c|x)\colon x \geq 0,\ Ax \leq b\}$$

and its perturbed version ($p \in \mathbb{R}^m$)

$$h(p) := \inf\{(c|x)\colon x \geq 0,\ Ax + p \leq b\}$$

An easy computation gives

$$\forall y \in \mathbb{R}^m, \quad h^*(y) = \begin{cases} -(b|y) & \text{if } A^{\mathrm{T}}y + c \leq 0, y \geq 0 \\ +\infty & \text{otherwise} \end{cases} \qquad [4]$$

**Lemma 18** *Assume that* $\inf(\mathcal{P})$ *is finite. Then*:

(i) $h$ *is convex proper and l.s.c. at* 0.
(ii) $(\mathcal{P})$ *has at least one solution.*

**Proof** We introduce the $(n+m) \times (m+1)$ matrix $B$ defined by

$$B := \begin{pmatrix} c^{\mathrm{T}} & 0 \\ A & I_m \end{pmatrix}$$

($I_m$ is the $m$-dimensional identity matrix). Denote $\{b_1, b_2, \ldots, b_{n+m}\} \subset \mathbb{R}^{m+1}$ the columns of $B$ and $K$ the convex cone $K := \{\sum_{j=1}^{j=n+m} \lambda_j b_j\colon \lambda_j \geq 0\}$. By Farkas lemma, this cone $K$ is closed.

(i) Let $\alpha := \liminf\{h(p)\colon p \to 0\}$. We have to prove that $\alpha \geq h(0) = \inf \mathcal{P}$. Let $\{p_\varepsilon\}$ be a sequence in $\mathbb{R}^m$ such that $p_\varepsilon \to 0$ and $h(p_\varepsilon) \to \alpha$. By the definition of $h$, we may choose $x_\varepsilon \geq 0$ such that $Ax_\varepsilon \leq b$ and $(c\,|\,x_\varepsilon) \to \alpha$. Then we see that the column vector $\tilde{x_\varepsilon}$ associated with $(x_\varepsilon, b - Ax_\varepsilon) \in \mathbb{R}^{n+m}$ satisfies: $B\tilde{x_\varepsilon} \in K$ and

$$B\tilde{x_\varepsilon} \to \begin{pmatrix} \alpha \\ b \end{pmatrix}$$

Therefore,

$$\begin{pmatrix} \alpha \\ b \end{pmatrix} \in K$$

and there exists $\tilde{x} = (x, x')$ such that $x \geq 0, x' \geq 0$, $(c\,|\,x) = \alpha$ and $Ax + x' = b$. It follows that $x$ is admissible for $(\mathcal{P})$ and then $(c\,|\,x) = \alpha \geq h(0)$.

(ii) We repeat the proof of (i) choosing $p_\varepsilon = 0$ so that $\alpha = \inf(\mathcal{P})$. □

Thanks to the assertion (i) in Lemma 18, we deduce from Theorem 10 that $\inf(\mathcal{P}) = h(0) = h^{**}(0) = \sup -h^*$. Recalling [4], we therefore consider the dual problem:

$$(\mathcal{P}^*) \qquad \sup\{-b \cdot y\colon\ y \geq 0,\ A^{\mathrm{T}} + c \geq 0\}$$

**Theorem 19** *The following assertions are equivalent*:

(i) $(\mathcal{P})$ *has a solution.*
(ii) $(\mathcal{P}^*)$ *has a solution.*
(iii) *There exists* $(x_0, y_0) \in \mathbb{R}^n_+ \times \mathbb{R}^m_+$ *such that* $Ax_0 \leq b, A^{\mathrm{T}}y_0 + c \geq 0$.

*In this case, we have* $\min(\mathcal{P}) = \max(\mathcal{P}^*)$ *and an admissible pair* $(\bar{x}, \bar{y})$ *is optimal if and only if* $c \cdot \bar{x} = -b \cdot \bar{y}$ *or, equivalently, satisfies the complementarity relations*: $(A\bar{x} - b) \cdot \bar{y} = (A^{\mathrm{T}}\bar{y} + c) \cdot \bar{x} = 0$.

**Convex programming** Let $f, g_1, \ldots, g_m : X \to \mathbb{R}$ be convex l.s.c. functions and the optimization problem

$$(\mathcal{P}) \qquad \inf\{f(x)\colon\ g_j(x) \leq 0,\ j = 1, 2 \ldots, m\}$$

Here $X = \mathbb{R}^n$ or any Banach space. As before, we introduce the value function

$$p \in \mathbb{R}^m, \quad h(p) := \inf\{f(x)\colon g_j(x) + p_j \leq 0\ j \in 1, 2, \ldots, m\}$$

and compute its Fenchel conjugate:

$$\lambda \in \mathbb{R}^m, \quad h^*(\lambda) = \begin{cases} -\inf\limits_{x \in X}\{L(x, \lambda)\} & \text{if } \lambda \geq 0 \\ +\infty & \text{otherwise} \end{cases}$$

where $L(x, \lambda) := f(x) + \sum \lambda_i g_i(x)$ is the so-called Lagrangian. We notice that $h$ is convex and that the equality $h(0) = h^{**}(0)$ is equivalent to the zero-duality gap relation

$$\inf_x \sup_\lambda L(x, \lambda) = \sup_\lambda \inf_x L(x, \lambda)$$

This condition is fulfilled, in particular, if we make the following qualification assumption (ensuring that $h$ is continuous at 0 and Theorem 11 applies):

$$\exists x_0 \in X : f \text{ continuous at } x_0, g_j(x_0) < 0, \forall j \qquad [5]$$

**Theorem 20** *Assume that* [5] *holds. Then* $\bar{x}$ *is optimal for* $(\mathcal{P})$ *if and only if there exist Lagrangian multipliers* $\bar{\lambda}_1, \bar{\lambda}_2, \ldots \bar{\lambda}_m$ *in* $\mathbb{R}_+$ *such that*

$$\bar{x} \in \operatorname*{argmin}_X \left( f + \sum_j \bar{\lambda}_j g_j \right), \quad \bar{\lambda}_j g_j(\bar{x}) = 0, \ \forall j$$

Notice that the existence of such a solution $\bar{x}$ is ensured if, for example, $X = \mathbb{R}^n$ and if, for some $k > 0$, the function $f + k\sum_j g_j$ is coercive.

#### Primal–Dual Formulations in Mechanics

We present here the example of elasticity which motivated the pioneering work by J J Moreau on convex duality techniques. Further examples can be found in Ekeland and Temam (1976). An elastic body is placed in a bounded domain $\Omega \subset \mathbb{R}^n$ whose boundary $\Gamma$ consists of two disjoint parts $\Gamma = \Gamma_0 \cup \Gamma_1$. The unknown $u : \Omega \to \mathbb{R}^n$ (deformation) satisfies a Dirichlet condition $u = 0$ on $\Gamma_0$, where the body is clamped. The system is subjected to a surface load $g \in L^2(\Gamma_1; \mathbb{R}^n)$ and to a volumic load $f \in L^2(\Omega; \mathbb{R}^n)$. The static equilibrium problem has the following variational formulation:

$$(\mathcal{P}) \qquad \inf_{u=0 \text{ on } \Gamma_0} \left\{ \int_\Omega j(x, e(u))\, dx - \int_\Omega f \cdot u\, dx - \int_{\Gamma_1} g \cdot u\, d\mathcal{H}^{n-1} \right\}$$

where $e(u) := (1/2)(u_{i,j} + u_{j,i})$ denotes the symmetric strain tensor and $j : (x, z) \in \Omega \times \mathbb{R}^{n^2}_{\text{sym}} \to \mathbb{R}_+$ is a convex integrand representing the local elastic behavior of the material. We assume a quadratic growth as in Remark 15 (in the case of linear elasticity, an isotropic homogeneous material is characterized by the quadratic form

$$j(x, z) = \frac{\lambda}{2} |\text{tr}(z)|^2 + \mu |z|^2$$

$\lambda, \mu$ being the Lamé constants).

We apply Proposition 14 with $X = W^{1,2}(\Omega; \mathbb{R}^n)$, $Y = L^2(\Omega; \mathbb{R}^{n^2}_{\text{sym}})$, $Au = e(u)$ and where we set

$$\Phi(u) = \begin{cases} -\int_\Omega f \cdot u\, dx \\ \quad -\int_{\Gamma_1} g \cdot u\, d\mathcal{H}^{n-1} & \text{if } u = 0 \text{ on } \Gamma_0 \\ +\infty & \text{otherwise} \end{cases}$$

$$\Psi(v) = \int_\Omega j(x, v)\, dx$$

After some computations, we may write the supremum appearing in Proposition 14 as our dual problem

$$(\mathcal{P}^*) \qquad \sup \left\{ -\int_\Omega j^*(x, \sigma)\, dx : \ \sigma \in L^2(\Omega; \mathbb{R}^{n^2}_{\text{sym}}), -\text{div}\, \sigma = f \text{ on } \Omega, \sigma \cdot n = g \text{ on } \Gamma_1 \right\}$$

where $j^*$ is the Moreau–Fenchel conjugate with respect to the second argument and $n(x)$ denotes the exterior unit normal on $\Gamma$. The matrix-valued map $\sigma$ is called the stress tensor and $j^*$ the stress potential. Note that the boundary conditions for $\sigma n$ have to be understood in the sense of traces.

**Theorem 21** *The problems* $(\mathcal{P})$ *and* $(\mathcal{P}^*)$ *have solutions and we have the equality*: $\inf(\mathcal{P}) = \sup(\mathcal{P}^*)$. *Futhermore, a pair* $(\bar{u}, \bar{\sigma})$ *is optimal if and only if it satisfies the following system*:

$$\begin{array}{lll} -\text{div}\, \bar{\sigma} = f & \text{on } \Omega & (\textit{equlibrium}) \\ \bar{\sigma}(x) \in \partial j(x, e(\bar{u})) & \text{a.e. on } \Omega & (\textit{constitutive law}) \\ \bar{u} = 0 & \text{a.e. on } \Gamma_0 & \\ \bar{\sigma} n = g & \text{on } \Gamma_1 & \end{array}$$

## Duality in Mass Transport Problems

#### General Cost Functions

Let $X, Y$ be a compact metric space and $c : X \times Y \to [0, +\infty)$ a continuous cost function. We denote by $\mathcal{P}(X), \mathcal{P}(X \times Y)$ the sets of probability measures on $X$ and $X \times Y$, respectively. Given two elements $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$, we denote by $\Gamma(\mu, \nu)$ the subset of probability measures in $\mathcal{P}(X \times Y)$ whose marginals are, respectively, $\mu$ and $\nu$. Identified as a subset of $(C^0(X \times Y))^*$ (the space of signed Radon measures on $X \times Y$), it is convex and weakly-star compact. The Monge–Kantorovich formulation of the mass transport problem reads as follows:

$$T_c(\mu, \nu) := \inf \left\{ \int_{X \times Y} c(x, y) \gamma(dx dy) : \gamma \in \Gamma(\mu, \nu) \right\} \qquad [6]$$

This formulation, where the infimum is achieved (as we minimize an l.s.c. functional on a compact set for the weak star topology), is already a relaxation of the initial Monge mass transport problem,

$$\inf_T \left\{ \int_X c(x, Tx) \mu(dx) : T^\#(\mu) = \nu \right\}$$

where the infimum is searched among all transports maps $T : X \mapsto Y$ pushing forward $\mu$ on $\nu$ (i.e., such that $\mu(T^{-1}(B) = \nu(B)$ for all Borel subset $B \subset Y$). This is equivalent to restricting the infimum in [6] to the subclass $\{\gamma_T\} \subset \Gamma(\mu, \nu)$, where

$$\langle \gamma_T, \phi(x, y) \rangle := \int_X \phi(x, Tx) \mu(dx)$$

In order to find a dual problem for [6], we fix $\nu \in \mathcal{P}(Y)$ and consider the functional $F : \mathcal{M}_b(X) \to [0, +\infty)$ defined by

$$F(\mu) = \begin{cases} T_c(\mu, \nu) & \text{if } \mu \geq 0, \mu(X) = 1 \\ +\infty & \text{otherwise} \end{cases}$$

($\mathcal{M}_b(X)$ denote the Banach space of (bounded) signed Radon measures on X).

**Lemma 22** *F is convex, weakly-star l.s.c. and proper. Its Moreau–Fenchel conjugate is given by*

$$\forall \varphi \in C^0(X), \quad F^*(\varphi) = -\int_Y \varphi^c(y) \nu(dy)$$

*where*

$$\varphi^c(y) := \inf\{c(x,y) - \varphi(x) \colon x \in X\}$$

**Proof** The convexity property is obvious and the properness follows from the fact that

$$F(\mu) \le \int_{X\times Y} c(x,y)\ \mu\otimes\nu(\mathrm{d}x\mathrm{d}y)$$

Let $\mu_n$ be such that $\mu_n \rightharpoonup \mu$ (weakly star). We may assume that $\liminf_n F(\mu_n) = \lim_n F(\mu_n) := \alpha$ is finite. Then $\mu_n$ and the associated optimal $\gamma_n$ are probability measures on $X$ and on $X\times Y$, respectively. As $X$ and $Y$ are compact, possibly passing to a subsequence, we may assume that $\gamma_n \rightharpoonup \gamma$, and clearly we have $\gamma \in \Gamma(\mu,\nu)$. Since $c(x,y)$ is l.s.c. non-negative, we conclude that

$$\begin{aligned}\liminf_n F(\mu_n) &= \liminf_n \int_{X\times Y} c(x,y)\gamma_n(\mathrm{d}x\mathrm{d}y)\\ &\ge \int_{X\times Y} c(x,y)\,\gamma(\mathrm{d}x\mathrm{d}y)\\ &= F(\mu)\end{aligned}$$

Let us compute now $F^*(\varphi)$. We have

$$\begin{aligned}-F^*(\varphi) &= \inf\left\{\int_{X\times Y} c(x,y)\gamma(\mathrm{d}x\mathrm{d}y) - \int_X \varphi\,\mathrm{d}\mu \colon \mu\in\mathcal{P}(X), \gamma\in\Gamma(\mu,\nu)\right\}\\ &= \inf\left\{\int_{X\times Y}(c(x,y)-\varphi(x))\gamma(\mathrm{d}x\mathrm{d}y)\colon \gamma\in\Gamma(\mu,\nu)\right\}\\ &\ge \int_Y \varphi^c(y)\,\nu(\mathrm{d}y)\end{aligned}$$

To prove that the last inequality is actually an equality, we observe that, for every $y\in Y$ and $\varphi\in C^0(X)$, the minimum of the l.s.c. function $c(\,\cdot\,,y)-\varphi$ is attained on the compact set $X$ and there exists a Borel selection map $S(y)$ such that $\varphi^c(y) = c(S(y),y) - \varphi(S(y)$ for all $y\in Y$. We obtain the desired equality by choosing $\gamma$ defined, for every test $\psi$, by

$$\int_{X\times Y}\psi(x,y)\gamma(\mathrm{d}x\mathrm{d}y) := \int_Y \psi(S(y),y)\nu(\mathrm{d}y)$$

□

We observe that, for every $\varphi\in C^0(X)$, the function $\varphi^c$ introduced in Lemma 22 is continuous (use the uniform continuity of $c$) and therefore the pair $(\varphi,\varphi^c)$ belong to the class

$$\mathcal{F}_c := \{(\varphi,\Psi)\in C^0(X)\times C^0(Y)\colon \varphi(x)+\psi(y)\le c(x,y)\}$$

Let us introduce the dual problem of [6]:

$$\sup\left\{\int_X \varphi\,\mathrm{d}\mu + \int_Y \psi\,\mathrm{d}\nu \colon (\varphi,\psi)\in\mathcal{F}_c\right\} \qquad [7]$$

We will say that $(\varphi,\psi)\in\mathcal{F}_c$ is a pair of c-concave conjugate functions if $\psi=\varphi^c$ and $\psi^c=\varphi$ (where symmetrically $\psi^c(x) := \inf\{c(x,y)-\psi(x)\colon y\in Y\}$). Checking the latter condition amounts to verifying that $\varphi$ enjoys the so-called c-concavity property $\varphi^{cc}=\varphi$ (in general, we have only $\varphi^{cc}\ge\varphi$, whereas $\varphi^{ccc}=\varphi^c$). We refer for instance to Villani (2003) for further details about this $c$-duality.

Now, by exploiting Theorem 10 and Lemma 22, we obtain a very simple proof of Kantorovich duality theorem:

**Theorem 23** *The following duality formula holds*:

$$T_c(\mu,\nu) = \sup\left\{\int_X \varphi\,\mathrm{d}\mu + \int_Y \psi\,\mathrm{d}\nu \colon (\varphi,\psi)\in\mathcal{F}_c\right\}$$

*Moreover, the supremum in the right-hand side member is achieved by a pair* $(\bar\varphi,\bar\psi)$ *of conjugate c-concave functions such that, for any optimal* $\bar\gamma$ *in* [6], *there holds* $\bar\varphi(x)+\bar\psi(y) = c(x,y)$, $\bar\gamma$*-a.e.*

**Proof** By Theorem 10 and Lemma 22, we have

$$\begin{aligned}T_c(\mu,\nu) &= F^{**}(\mu)\\ &= \sup\left\{\int_X \varphi\,\mathrm{d}\mu + \int_Y \varphi^c\,\mathrm{d}\nu \colon \varphi\in C^0(X)\right\}\\ &\le \sup\left\{\int_X \varphi\,\mathrm{d}\mu + \int_Y \psi\,\mathrm{d}\nu\colon\ (\varphi,\psi)\in\mathcal{F}_c\right\}\\ &\le T_c(\mu,\nu)\end{aligned}$$

where the last inequality follows from the definition of $\mathcal{F}_c$. Therefore, $\inf[6] = \sup[7]$. Furthermore, on the right-hand side of first equality, we increase the supremum by substituting $\varphi$ with $\varphi^{cc}$ (recall that $\varphi^{ccc}=\varphi^c$). Thus,

$$\sup[7] = \sup\left\{\int_X \varphi\,\mathrm{d}\mu + \int_Y \varphi^c\,\mathrm{d}\nu \colon \varphi\in C^0(X),\ \varphi\ c\text{-concave}\right\}$$

Take a maximizing sequence $(\varphi_n,\varphi_n^c)$ of $c$-concave conjugate functions. It is easy to check that $\{f_n\}$ is equicontinuous on $X$: this follows from the $c$-concavity property and from the uniform continuity of $c$ (observe that $\varphi_n(x_1)-\varphi_n(x_2) = \varphi_n^{cc}(x_1)-\varphi_n^{cc}(x_2) \le \sup_Y\{c(x_1,\,\cdot\,)-c(x_2,\,\cdot\,)\}$). Then, by Ascoli's theorem, possibly passing to subsequences, we may assume that: $\varphi_n - c_n$ converges uniformly to some continuous function $\bar\varphi$ where $\{c_n\}$ is a suitable sequence of reals. Then, one checks that $\bar\varphi$ is still $c$-concave and that $(\varphi_n-c_n)^c = \varphi_n^c + c_n$ converges uniformly to

$\bar{\varphi}^c$. Thus, recalling that $\mu(X) = \nu(Y) = 1$, we deduce that

$$\begin{aligned}\sup[7] &= \lim_n \left( \int_X \varphi_n \,\mathrm{d}\mu + \int_Y \varphi_n^c \,\mathrm{d}\nu \right) \\ &= \lim_n \left[ \int_X (\varphi_n - c_n) \,\mathrm{d}\mu + \int_Y (\varphi_n^c + c_n) \,\mathrm{d}\nu \right] \\ &= \int_X \bar{\varphi} \,\mathrm{d}\mu + \int_Y \bar{\varphi}^c \,\mathrm{d}\nu\end{aligned}$$

The last assertion is a consequence of the extremality relation:

$$\begin{aligned}0 &= \inf[6] - \sup[7] \\ &= \int_{X\times Y} \big(c(x,y) - \bar{\varphi}(x) - \bar{\psi}(y)\big)\bar{\gamma}(\mathrm{d}x\mathrm{d}y)\end{aligned}$$

□

**Remark 24**

(i) In their discrete version (i.e., $\mu, \nu$ are atomic measures), problems [6] and [7] can be seen as particular linear programming problems (see the section "Finite-dimensional case").

(ii) The case $X = Y \subset \mathbb{R}^n$ and $c(x,y) = (1/2)|x-y|^2$ is important. In this case, the notion of $c$-concavity is linked to convexity and the Fenchel transform since, for every $\varphi \in C^0(X)$, one has

$$\frac{|\cdot|^2}{2} - \varphi^c = \left( \frac{|\cdot|^2}{2} - \varphi \right)^*$$

Then if $(\bar{\varphi}, \bar{\varphi}^c)$ is a solution of [7], we find that

$$\varphi_0(x) := \frac{|x|^2}{2} - \bar{\varphi}(x)$$

is convex continuous and that the extremality condition: $\bar{\varphi}(x) + \bar{\varphi}^c(y) = c(x,y)$ is equivalent to Fenchel equality $\varphi_0(x) + \varphi_0^*(y) = (x|y)$. Therefore, any optimal $\bar{\gamma}$ is supported in the graph of the subdifferential map $\partial\varphi_0$. In the case where $\mu$ is absolutely continuous with respect to the Lebesgue measure, it is then easy to deduce that the optimal $\bar{\gamma}$ is unique and that $\bar{\gamma} = \gamma_{T_0}$, where $T_0 = \nabla\varphi_0$ is the unique gradient (a.e. defined) of a convex function such that $\nabla\varphi_0^\sharp(\mu) = \nu$. This is a celebrated result by Y Brenier (see, e.g., the monographs by Evans (1997) and Villani (2003)).

### The Distance Case

In the following, we assume that $X = Y$ and that $c(x,y)$ is a semidistance. As an immediate consequence of the triangular inequality, we have the following equivalence:

$$\begin{aligned}\varphi \quad c\text{-concave} &\Leftrightarrow \varphi(x) - \varphi(y) \le c(x,y), \ \forall (x,y) \\ &\Leftrightarrow \ \varphi^c = -\varphi\end{aligned}$$

Let us denote $\mathrm{Lip}_1(X) := \{u \in C^0(X) \colon u(x) - u(y) \le c(x,y)\}$. The first assertion of Theorem 23 becomes the Kantorovich–Rubintein duality formula:

$$T_c(\mu,\nu) = \max\left\{ \int_X u \,\mathrm{d}(\mu - \nu) \colon u \in \mathrm{Lip}_1(X) \right\} \qquad [8]$$

As it appears, $T_c(\mu,\nu)$ depends only on the difference $f = \mu - \nu$, which belongs to the space $\mathcal{M}_0(X)$ of signed measure on $X$ with zero average. Defining $N(f) := T_c(f^+, f^-)$ provides a seminorm (Kantorovich norm) on $\mathcal{M}_0(X)$ (it turns out that $\mathcal{M}_0(X)$ is not complete and that in general its completion is a strict subspace of the dual of $\mathrm{Lip}(X)$).

We will now specialize to the case where $X$ is a compact manifold equipped with a geodesic distance. This will allow us to link the original problem to another primal–dual formulation closer to that considered in the section "Primal–dual formulation in mechanics" and yielding to a connection with partial differential equations. As a model example, let us assume that $K = \bar{\Omega}$, where $\Omega$ is a bounded connected open subset of $\mathbb{R}^n$ with a Lipschitz boundary. Let $\Sigma \subset \bar{\Omega}$ be a compact subset (on which the transport will have zero cost) and define

$$\begin{aligned}c(x,y) := \inf\{ &\mathcal{H}^1(S \setminus \Sigma) \colon \\ &S \text{ Lipschitz curve joining } x \text{ to } y, S \subset \bar{\Omega}\}\end{aligned} \qquad [9]$$

where $\mathcal{H}^1$ denotes the one-dimensional Hausdorff measure (length). It is easy to check that

$$c(x,y) = \min\{\delta_\Omega(x,y), \delta_\Omega(x,\Sigma) + \delta_\Omega(y,\Sigma)\}$$

where $\delta_\Omega(x,y)$ is the geodesic distance on $\Omega$ (induced by the Euclidean norm). Furthermore, the following characterization holds:

$$\begin{aligned}&u \in \mathrm{Lip}_1(X) \Longleftrightarrow u \in W^{1,\infty}(\Omega), \\ &\quad |\nabla u| \le 1 \text{ a.e. in } \Omega, \ u = cte \text{ on } \Sigma\end{aligned} \qquad [10]$$

Since $f := \mu - \nu$ is balanced, the value of the constant on $\Sigma$ in [10] is irrelevant and can be set to 0. Thus we may rewrite the right hand side member of [8] in a equivalent way as

$$\max\left\{ \int_{\bar{\Omega}} u \,\mathrm{d}f \colon u \in W^{1,\infty}(\Omega), \right. \left. |\nabla u| \le 1 \text{ a.e. on } \Omega, \ u = 0 \text{ on } \Sigma \right\} \qquad [11]$$

We will now derive a new dual problem for [11] by using Proposition 14. To this aim, we consider

$X = C^1(\bar{\Omega})$ (as a closed subspace of $W^{1,\infty}(\Omega)$), $Y = C^0(\bar{\Omega};\mathbb{R}^n)$, $Y^* = \mathcal{M}_b(\bar{\Omega};\mathbb{R}^n)$ and the operator $A: u \in X \mapsto \nabla u \in Y$.

**Theorem 25** *Let* $\mu, \nu \in \mathcal{P}(\bar{\Omega})$, $f = \mu - \nu$ *and* $c$ *defined by* [9]. *Then,*

$$T_c(\mu,\nu) = \min\left\{ \int_{\bar{\Omega}} |\lambda| : \lambda \in \mathcal{M}_b(\bar{\Omega};\mathbb{R}^n), -\operatorname{div}\lambda = f \text{ on } \bar{\Omega}\setminus\Sigma \right\} \quad [12]$$

*where the divergence condition is intended in the sense that*

$$\int_{\bar{\Omega}} \lambda \cdot \nabla\varphi = \int_{\bar{\Omega}} \varphi\, df$$

*for all* $\varphi \in C^\infty$ *compactly supported in* $\mathbb{R}^n \setminus \Sigma$.

**Proof** (*sketch*) We apply Proposition 14 with $\phi(u) = -\int_{\bar{\Omega}} u\, df$ if $u = 0$ on $\Sigma$ ($+\infty$ otherwise), $A = \nabla$, and $\psi(v) = 0$ if $|v| \le 1$ on $\bar{\Omega}$ ($+\infty$ otherwise). We obtain that the minimum $\alpha$ in [12] is reached and that $\alpha = \beta$, where

$$-\beta := \inf\left\{ -\int_{\bar{\Omega}} u\, df : u \in C^1(\bar{\Omega}), |\nabla u| \le 1 \text{ on } \Omega\ u = 0 \text{ on } \Sigma \right\}$$

To prove that $\beta = T_c(\mu,\nu) = \sup$ (11), we consider a maximizer $\bar{u}$ in [11] and prove that it can be approximated uniformly by a sequence $\{u_n\}$ of functions in $C^1(\bar{\Omega})$ which satisfy the same constraints. This technical part is done by truncation and convolution arguments (we refer to Bouchitté *et al.* (2003) for details). □

**Remark 26** By localizing the integral identity associated with [12], it is possible to deduce the optimality conditions which characterize optimal pairs $(\bar{u}, \bar{\lambda})$ for [11], [12] (without requiring any regularity). This is done by using a weak notion of tangential gradient with respect to a measure (see Bouchitté *et al.* (1997) and Bouchitté and Fragalà (2002)). If $\bar{\lambda} = \bar{\sigma}\, dx$ where $\sigma \in L^1(\Omega;\mathbb{R}^n)$ and if $\Sigma \subset \partial\Omega$, then we find that $\bar{\sigma} = a\nabla\bar{u}$, where the pair $(\bar{u}, a)$ solves the following system:

$$\begin{aligned} -\operatorname{div}(a\nabla\bar{u}) &= f \text{ on } \Omega && (\textit{diffusion equation}) \\ |\nabla\bar{u}| &= 1 \text{ a.e. on } \{a > 0\} && (\textit{eikonal equation}) \\ u &= 0 \text{ a.e. on } \Sigma && \\ \frac{\partial u}{\partial n} &= 0 \text{ on } \Sigma && \end{aligned}$$

**Remark 27** Given a solution $\bar{\gamma}$ for [6], we can construct a solution $\bar{\lambda}$ for [12] by selecting for every $(x,y) \in \operatorname{spt}(\bar{\gamma})$ a geodesic curve $S_{xy}$ joining $x$ and $y$ (possibly passing through the free-cost zone $\Sigma$) and by setting, for every test $\phi$:

$$\langle \bar{\lambda}, \phi \rangle := \int_{\bar{\Omega}\times\bar{\Omega}} \left( \int_{S_{xy}} \phi \cdot \tau_{S_{xy}}\, d\mathcal{H}^1 \right) \bar{\lambda}(dx dy)$$

where $\tau_{S_{xy}}$ denote the unit oriented tangent vector (see Bouchitté and Buttazzo (2001)). It is also possible to show (see Ambrosio (2003)) that any solution $\bar{\lambda}$ can be represented as before through a particular solution $\bar{\gamma}$. As a consequence, the support of any solution $\bar{\gamma}$ of [12] is supported in the geodesic envelope of the set $\operatorname{spt}(\mu) \cup \operatorname{spt}(\nu) \cup \Sigma$. However, we stress the fact that, in general, there is no uniqueness at all of the optimal triple $(\bar{\gamma}, \bar{u}, \bar{\lambda})$ for [6], [11] and [12].

**Remark 28** An approximation procedure for particular solutions of problems [11], [12] can be obtained by solving a $p$-Laplace equation and then by sending $p$ to infinity. Precisely, consider the solution $u_p \in W^{1,p}(\Omega)$ of

$$\begin{aligned} -\operatorname{div}(|\nabla u|^{p-2}\nabla u) &= f && \text{on } \bar{\Omega}\setminus\Sigma \\ u &= 0 && \text{on } \Sigma \end{aligned}$$

which, for $p > n$, exists (due to the compact embedding $W^{1,p}(\Omega) \subset C^0(\bar{\Omega})$) and is unique. In Bouchitté *et al.* (2003) it is proved that the sequence $\{(u_p, \sigma_p)\}$, where $\sigma_p = |\nabla u_p|^{p-2}\nabla u_p$, is relatively compact in $\mathcal{M}_b(\bar{\Omega};\mathbb{R}^n) \times C^0(\bar{\Omega}$ (weakly star with respect to the first component) and that every cluster point $(\bar{u}, \bar{\lambda})$ solves [11], [12]. It is an open problem to know whether or not such a cluster point is unique. If the answer is "yes," the process described above would select one optimal pair among all possible solutions. As far as problem [11] is concerned, this problem is connected with the theory of viscosity solutions for the infinite Laplacian (see Evans (1997)) although this theory does not provide an answer as it erases the role of the source term $f$. On the other hand, a new entropy selection principle should be found for the solutions of dual problem [12]. In fact, the following partial result holds: let $E: \mathcal{M}_b(\bar{\Omega};\mathbb{R}^n) \to \mathbb{R} \cup \{+\infty\}$ be the functional defined by

$$E(\lambda) := \begin{cases} \int_\Omega |\sigma| \log(|\sigma|)\, dx & \text{if } \lambda \ll dx \text{ and } \sigma = \dfrac{d\lambda}{d|\lambda|} \\ +\infty & \text{otherwise} \end{cases}$$

Assume that [12] admits at least one solution $\lambda_0$ such that $E(\lambda_0) < +\infty$. Then it can be shown that

the sequence $\{\sigma_p\}$ does converge weakly-star to $\bar{\lambda}$, the unique minimizer of the problem

$$\inf\{E(\lambda): \lambda \text{ solution of } [12]\}$$

The general case, in particular when all optimal measures are singular, is open.

**Remark 29** Variational problems [11], [12] have important counterparts in the theory of elasticity and in optimal design problems (see Bouchitté and Buttazo (2001)). They read, respectively, as

$$\max\left\{\int_{\bar{\Omega}} u \cdot \mathrm{d}f: \ u \in \cap_{p>1} W^{1,p}(\Omega; \mathbb{R}^n), \right.$$
$$\left. \nabla u(x) \in K \text{ a.e. on } \Omega, \ u = 0 \text{ on } \Sigma\right\}$$
$$\min\left\{\int_{\bar{\Omega}} \rho_K^0(\lambda): \lambda \in \mathcal{M}_b(\bar{\Omega}; \mathbb{R}^{n^2}_{\mathrm{sym}}), \right.$$
$$\left. -\operatorname{div}\lambda = f \text{ on } \bar{\Omega}\backslash\Sigma\right\}$$

where $K \subset \mathbb{R}^{n^2}_{\mathrm{sym}})$ is a convex compact subset of symmetric second-order tensors associated with the elastic material, $\rho_K^0(\xi) = \sup\{\xi \cdot z: z \in K\}$ is convex positively 1-homogeneous and the functional on measures $\int_{\bar{\Omega}} \rho_K^0(\lambda)$ is intended in the sense given in [1]. A celebrated example is given by Michell's problem (Michell 1904) where $n = 2$ and $K := \{z \in \mathbb{R}^{n^2}_{\mathrm{sym}}, |\rho(z)| \le 1\}$, $\rho(z)$ being the largest singular value of $z$. The potential $\rho_K^0$ is given by the nondifferentiable convex function $\rho_K^0(\xi) = \tau_1(\xi) + \tau_2(\xi)$, where the $\tau_i(\xi)$'s are the singular values of $\xi$.

Unfortunately, it is not known if the vector variational problem above can be linked to an optimal transportation problem of the type [6], even if the analogous of equivalence [10] does exist in the Michell's case, namely (for $\Omega$ convex):

$$\rho(e(u)) \le 1 \quad \text{on } \Omega$$
$$\iff |(u(x) - u(y)|x - y)| \le |x - y|^2, \ \forall (x, y)$$

## Further Reading


Alberti G, Bouchitté G, and Dal Maso G (2003) The calibration method for the Mumford–Shah functional and free-discontinuity problems. *Calculus of Variations and Partial Differential Equations* 16(3): 299–333.

Ambrosio L (2003) Lecture notes on optimal transport problems. In: *Mathematical Aspects of Evolving Interfaces (Funchal 2000)*, Lecture Notes in Mathematics, vol. 1812, pp. 1–52. Berlin: Springer.

Borwein M and Lewis SA (2000) *Convex Analysis and Nonlinear Optimization. Theory and Examples*, CMS Series. Berlin: Springer.

Bouchitté G and Buttazzo G (2001) Characterization of optimal shapes and masses through Monge–Kantorovich equations. *Journal of the European Mathematical Society* 3: 139–168.

Bouchitté G, Buttazzo G, and De Pascale L (2003) A p-Laplacian approximation for some mass optimization problems. *Journal of Optimization Theory and Applications* 118: 1–25.

Bouchitté G, Buttazzo G, and Seppecher P (1997) Energies with respect to a measure and applications to low dimensional structures. *Calculus of Variations and Partial Differential Equations* 5: 37–54.

Bouchitté G and Dal Maso G (1993) Integral representation and relaxation of convex local functionals on $BV(\Omega)$. *Annali della Scuola Superiore di Pisa* 20(4): 483–533.

Bouchitté G and Fragalà I (2002) Variational theory of weak geometric structures: the measure method and its applications. *Variational Methods for Discontinuous Structures, Ser. PNLDE*, vol. 51, pp. 19–40. Basel: Birkhäuser.

Bouchitté G and Fragalà I (2003) Second order energies on thin structures: variational theory and non-local effects. *Journal of Functional Analysis* 204(1): 228–267.

Bouchitté G and Valadier M (1988) Integral representation of convex functionals on a space of measures. *Journal of Functional Analysis* 80: 398–420.

Ekeland I and Temam R (1976) *Analyse convexe et problèmes variationnels*. Paris: Dunod-Gauthier Villars.

Evans LC (1997) Partial differential equations and Monge–Kantorovich mass transfer. In: Bott R, Jaffe A, Jerison D, Lutsztig G, Singer I, and Yau JT (eds.) *Current Developments in Mathematics*, pp. 65–126. Cambridge.

Michell AGM (1904) The limits of economy of material in frame structures. *Philosophical Magazine and Journal of Science* 6: 589–597.

Rockafellar RT (1970) *Convex Analysis*. Princeton: Princeton University Press.

Villani C (2003) *Topics in Optimal transportation*, Graduate studies in Mathematics, vol. 58. Providence, RI: AMS.


**Cosmic Censorship** *see* Spacetime Topology, Causal Structure and Singularities

# Cosmology: Mathematical Aspects

**G F R Ellis**, University of Cape Town,
Cape Town, South Africa



## Introduction

Mathematical cosmology focuses on the geometrical and mathematical aspects of the study of the universe as a whole. Because the structure of spacetime (with metric tensor $g_{ab}(x^j)$) is governed by gravity, with matter and energy causing spacetime curvature according to the nonlinear gravitational field equations of the theory of general relativity, it has its roots in differential geometry. It is to be distinguished from the three other major aspects of modern cosmology, namely astrophysical cosmology, high-energy physics cosmology, and observational cosmology; see Peacock (1999) for these aspects.

The Einstein field equations (EFEs) are

$$R_{ab} - \frac{1}{2}Rg_{ab} + \Lambda g_{ab} = \kappa T_{ab} \qquad [1]$$

where $R_{ab}$ is the Ricci tensor, $R$ the Ricci scalar, $T_{ab}$ the matter tensor, $\Lambda$ the cosmological constant, and $\kappa$ the gravitational constant. Cosmological models differ from generic solutions of these equations in that they have preferred world lines in spacetime associated with the motion of matter and distribution of radiation (Ellis 1971). This is a classic case of a broken symmetry: the underlying equations [1] are locally Lorentz invariant but their solutions are not. These preferred world lines, characterized by a unit 4-velocity vector $u^a$, are associated at late times with "fundamental observers," and a key aspect of cosmological modeling is determining the observational relations such observers would determine through astronomical observations.

The dynamics of cosmological models is determined by their matter content. This is usually represented in simplified form, often using the "perfect-fluid" approximation to represent the effect of matter or radiation; that is,

$$T_{ab} = (\rho + p)u_a u_b + p g_{ab} \qquad [2]$$

where $\rho$ is the energy density and $p$ the pressure, and the matter 4-velocity $u_b$ is the preferred cosmological 4-velocity. This description can include a scalar field $\phi$ with dynamics governed by the Klein–Gordon equation, provided $u_a$ is normal to spacelike surfaces $\{\phi = \text{const}\}$. Suitable equations of state describe the nature of the matter envisaged (e.g., $p = 0$ for baryons, whereas $p = \rho/3$ for radiation); in the case of a scalar field with potential $V(\phi)$ and spacelike surfaces $\{\phi = \text{const.}\}$, on choosing $u^a$ orthogonal to these surfaces, the stress tensor has a perfect-fluid form with $\rho = (1/2)\dot{\phi}^2 + V(\phi)$, $p = (1/2)\dot{\phi}^2 - V(\phi)$. A cosmological constant $\Lambda$ can be represented as a perfect fluid with $\rho + p = 0$, $\Lambda = p$. More general matter may involve a momentum flux density $q_a$ and anisotropic pressures $\pi_{ab}$ (Ehlers 1961). Whatever the nature of the matter, it will usually be required to satisfy energy conditions (Hawking and Ellis 1973). All realistic matter has a positive inertial mass density:

$$\rho + p > 0 \qquad [3]$$

(note that realistic cosmological models are nonempty), whereas all ordinary matter has a positive gravitational mass density:

$$\rho + 3p > 0 \qquad [4]$$

but this is not necessarily true for a scalar field or effective cosmological constant.

Mathematical cosmology (Ellis and van Elst 1999) studies (1) generic properties of solutions with a preferred 4-velocity field and matter content as indicated above, (2) the standard FLRW models, (3) approximate FLRW solutions, and (4) other exact and approximate cosmological solutions. The ultimate underlying issue is (5) the origin of the universe. We look at these in turn. We aim to use covariant methods as far as possible, to avoid being misled by coordinate effects, and to obtain exact solutions and exact results as far as possible, because approximate methods can be misleading in the case of these nonlinear field equations.

## Exact Properties

We can split the equations into spacelike and timelike parts relative to the 4-velocity $u^a$, obtaining the $(1+3)$ covariant dynamical equations and identities in terms of the fluid shear $\sigma_{ab}$, vorticity $\omega_{ab}$, expansion $\Theta = u^a{}_{;a}$, and acceleration $a^b = u_{a;b}u^b$ (Ehlers 1961, Ellis 1971, Ellis and van Elst 1999). The energy density of a perfect fluid obeys the conservation equation

$$\dot{\rho} = -3(\rho + p)\frac{\dot{S}}{S} \qquad [5]$$

with extra terms occurring in the case of more complex matter. From the momentum equations, pressure-free solutions are geodesic ($a^b = 0$). The crucial Raychaudhuri–Ehlers equation for the

time derivative of the expansion (Ehlers 1961) can be written as

$$3\frac{\ddot{S}}{S} = 2(\omega^2 - \sigma^2) + a^b_{;b} - \frac{\kappa}{2}(\rho + 3p) + \Lambda \qquad [6]$$

where the representative length scale $S$ is defined by $\Theta = 3\dot{S}/S$. This is the basis of the "fundamental singularity theorem": if in an expanding universe $\omega = 0 = a^b$ and the combined matter present satisfies [4], with $\Lambda \leq 0$, then there was a singularity where $S \to 0$ a finite time $t_0 < 1/H_0$ ago, $H_0 = (\dot{S}/S)_0$ being the present value of the Hubble constant. The energy density will diverge there, so this is a spacetime singularity: an origin of physics, matter, and spacetime itself. However, the deduction does not follow if there is rotation or acceleration, which could conceivably avoid the singularity, so this result is by itself inconclusive for realistic cosmologies.

The vorticity obeys conservation laws analogous to those in Newtonian theory (Ehlers 1961). Vorticity-free solutions ($\omega = 0$) occur whenever the fluid flow lines are hypersurface-orthogonal in spacetime, that is, there exists a cosmic time function for the comoving observers, which will measure proper time along the flow lines if additionally the fluid flow is geodesic. The rate of change of shear is related to the conformal curvature (Weyl) tensor, which represents the free gravitational field, and which splits into an electric part $E_{ab}$ and a magnetic part $H_{ab}$ in close analogy with electromagnetic theory. Shear-free solutions ($\sigma = 0$) are very special because they strongly constrain the Weyl tensor; indeed if the flow is shear free and geodesic, then it either does not expand ($\Theta = 0$), or does not rotate ($\omega = 0$) (Ellis 1967). The set of cosmological observations associated with generic cosmological models has been characterized in power series form by Kristian and Sachs (1966), and that result has been extended to general models by Ellis *et al.* (1985).

The local regularity of the theory is expressed in existence and uniqueness theorems for the EFEs, provided the matter behavior is well defined through prescription of suitable equations of state (Hawking and Ellis 1973). However, in general the theory breaks down in the large, and this feature is specified by the Hawking–Penrose singularity theorems, predicting the existence of a geodesic incompleteness of spacetime under conditions applicable to realistic cosmological models satisfying the energy conditions given by eqns [3] and [4] (Hawking and Ellis 1973, Tipler *et al.* 1980). However, the conclusion does not follow if the energy conditions are not satisfied. Furthermore, the deduction follows only if the gravitational field equations remain valid to arbitrarily early times; but we would in fact expect that, at high enough energy densities, quantum gravity would take over from classical gravity, so whether or not there was indeed a singularity would depend on the nature of the as yet unknown theory of quantum gravity. The cash value of the singularity theorems then is the implication that, when the energy conditions are satisfied, one would indeed be involved in such a quantum gravity realm in the very early universe.

## The Standard Friedmann–Lemaître Models

The standard models of cosmology are the Friedmann–Lemaître (FL) models with Robertson–Walker (RW) geometry: that is, they are exactly spatially homogeneous and locally isotropic, invariant under a $G_6$ of isometries (Robertson 1933, Ehlers 1961). They have a unique cosmic time function $t$, with space sections $\{t = \text{const.}\}$ of constant spatial curvature orthogonal to the uniquely preferred 4-velocity $u^a$. The fluid acceleration, vorticity, and shear all vanish, and all physical quantities depend only on the time coordinate $t$. They can be represented by a metric with scale factor $S(t)$:

$$\begin{aligned} ds^2 &\equiv g_{ab}dx^a dx^b \\ &= -dt^2 + S^2(t)\{dr^2 + f^2(r)(d\theta^2 + \sin^2\theta\, d\phi^2)\} \end{aligned} \qquad [7]$$

in comoving coordinates $(x^a) = (t, r, \theta, \phi)$, where $f(r) = \{\sin r, r, \sinh r\}$ if $\{k = +1, 0, -1\}$, and the matter is a perfect fluid with 4-velocity vector $u^a = dx^a/ds = \delta^a_0$. The curvature of the space sections $\{t = \text{const.}\}$ is $K = k/S^2$; these 3-spaces are necessarily closed (compact) if they are positively curved ($k = +1$), but may be open or closed in the flat ($k = 0$) and negatively curved ($k = -1$) cases, depending on their topology (Lachieze-Rey and Luminet 1995).

Matter obeys the conservation equation [5], whose outcome depends on the equation of state; for baryons $\rho = M/S^3$, whereas for radiation $\rho = M/S^4$, where $M$ is a constant. The dynamics of the models is governed by the Raychaudhuri equation

$$3\frac{\ddot{S}}{S} = -\frac{\kappa}{2}(\rho + 3p) + \Lambda \qquad [8]$$

which has the Friedmann equation

$$\frac{3\dot{S}^2}{S^2} = \kappa\rho + \Lambda - \frac{3k}{S^2} \qquad [9]$$

as a first integral whenever $\dot{S} \neq 0$. Depending on the matter components present, one can qualitatively

characterize the dynamical behavior of these models (Robertson 1933) and find exact and approximate solutions to these equations as well as phase planes representing the relation of the different models to each other; for example, Ehlers and Rindler (1989) give the phase planes for models with noninteracting matter and radiation and an arbitrary cosmological constant. Universes with maxima or minima in $S(t)$ can only occur if $k=+1$; when $\Lambda=0$, the universe recollapses in the future iff $k=+1$. Static solutions are possible only if $k=+1$ and (assuming [4]) $\Lambda>0$. The simplest expanding solutions are the Einstein–de Sitter universes with $k=0=\Lambda$.

Equation [8] is a special case of [6], with corresponding implications: if the combined matter present satisfies [4], with $\Lambda\leq 0$, then there must have been an initial singularity, or at least the universe must have emerged from a quantum gravity domain. The temperature would have been arbitrarily high in the past, so there was a hot big bang era in the early universe where matter and radiation were in equilibrium with each other at very high temperatures that rapidly fell as the universe expanded. Many physical processes took place then, in particular nucleosynthesis of light elements took place at $\sim 10^9$ K. Decoupling of matter and radiation took place at a temperature of $\sim 4000$ K, followed by formation of stars and galaxies (see Peacock (1999) for a discussion of these physical processes). The black-body radiation emitted by the surface of last scattering at 4000 K is observed by us today as cosmic black-body radiation (CBR) at a temperature of 2.75 K.

One can determine observational relations for these models such as the magnitude–redshift relation for "standard candles" at recent times from the EFEs (Sandage 1961). The aim of observations is to determine the Hubble constant $H_0$, dimensionless deceleration parameter $q_0=-(3/H_0^2)(\ddot{S}/S)_0$, and normalized density parameters $\Omega_{0i}=\kappa\rho_{0i}/3H_0^2$ for each component of matter present. The spatial curvature and the cosmological constant then follow from [6] and [9]; also the present scale factor $S_0$ is determined if $k\neq 0$. The universe is of positive spatial curvature $(k=+1)$ iff $\Omega_0\equiv\Omega_{\rm m}+\Omega_\Lambda>1$, where $\Omega_{\rm m}\equiv\sum_i\Omega_{0i}$, $\Omega_\Lambda=\Lambda/3H_0^2$. Current observations indicate $\Omega_{\rm m}\simeq 0.3, \Omega_\Lambda\simeq 0.7, \Omega_0\simeq 1.02\pm 0.02$. Because the nucleosynthesis results limit the baryon density to a very low value $(\Omega_{0\rm b}\simeq 0.02)$, which is about the same as the density of luminous matter, this indicates the dominant presence of both nonbaryonic dark matter and a repulsive force corresponding to either a cosmological constant or varying scalar field (dark energy).

Crucial causal limitations occur because of the existence of particle horizons (Rindler 1956), the nature of which is most clear when represented in conformal diagrams (Hawking and Ellis 1973, Tipler *et al.* 1980). These result from the fact that light can only proceed a finite distance in the finite time since the origin of the universe, and imply that for a standard radiation-dominated hot-big-bang early universe, regions of larger than $\sim 1^\circ$ angular size on the surface of last scattering, which emits the CBR, are causally disconnected: hence, no causal process since the start of the universe can account for the extreme isotropy of the CBR ($\Delta T/T\simeq 10^{-5}$ over the whole sky, once a dipole anisotropy $\Delta T/T\simeq 10^{-3}$ due to our local velocity relative to the cosmological rest frame is allowed for). This is the "horizon problem," one of the driving forces behind the theory of "inflation" (Guth 1981): the idea that, in the very early universe, a slow-rolling scalar field led to a brief exponential expansion through at least 50 e-folds (during which time the spacetime was approximately de Sitter), thus smoothing the universe and solving the horizon problem (Guth 1981, Peacock 1999). This is possible because a scalar field can violate the energy condition [3] and so allows acceleration: $\ddot{S}>0$. Consequently, there are now many studies of the dynamics of FLRW solutions driven by scalar fields and the subsequent decay of these scalar fields into radiation. One interesting point is that one can obtain exact solutions of this kind for arbitrarily chosen evolutions $S(t)$, provided they satisfy a restriction on the magnitude of $\dot{S}^2$, by running the field equations backwards to determine the needed potential $V(\phi)$ (Ellis and Madsen 1991). The inflationary paradigm is dominant in present-day theoretical cosmology, but suffers from the problem that it is not in fact a well-defined theory, for there is no single accepted proposal for the physical nature of the effective scalar field underlying the supposed exponential expansion; rather there are numerous competing proposals. As the inflaton has not yet been identified, this theory is not yet soundly linked to well-established physics.

## Approximate FL Solutions

The real universe is, of course, not exactly FL, and studies of structure formation depend on studies of solutions that are approximately FL models – they are realistic ("lumpy") universe models. These enable detailed studies of observable properties such as CBR anisotropies and gravitational lensing induced by matter inhomogeneities, and of the development of those inhomogeneities from quantum fluctuations in the very early universe that then get expanded to very large scales by inflation.

The key problem here is that apart from the standard coordinate freedom allowed in general relativity, there is a serious gauge issue: the background FL model is not uniquely determined by the realistic universe model; however, the magnitudes of many perturbed quantities depend on how it is fitted into the lumpy model. For example, the density perturbation $\delta\rho$ is determined pointwise by the equation

$$\delta\rho(x^i) \equiv \rho(x^i) - \overline{\rho}(x^i)$$

where $\overline{\rho}(x^i)$ is the background density. But by altering the correspondence between the background and realistic models (specifically, by the choice of surfaces $\overline{\rho}(x^i) = \text{const.}$ in the realistic model) one can assign that quantity any value, including zero (if one chooses $\overline{\rho}(x^i) = \rho(x^i)$). This is the "gauge problem."

One can handle it by using standard variables and keeping close track of the gauge freedom at all times. However, one then ends up with higher-order equations than necessary because some of the perturbation modes present are pure gauge modes with no physical significance. Alternatively, one can fix the gauge by some unique specification of how the background model is fitted into the realistic model, but there is no agreement on a unique way to do this, and different choices give different answers. The preferable resolution is to use gauge-invariant variables, either coordinate based (Bardeen 1980) or covariant, based on the $(1+3)$ covariant decomposition of spacetime quantities mentioned above (Ellis and Bruni 1989), in either case resulting in perturbation equations without gauge freedom and of order corresponding to the physical degrees of freedom. The key point in the latter approach is to choose covariant variables that vanish in the background spacetime; they are then automatically gauge invariant. Realistic structure formation studies carry out this process for a mixture of matter components with different average velocities, and extend to a kinetic theory description of the background radiation (see Ellis and van Elst (1999) and references therein). The outcome is a prediction of the CBR anisotropy power spectrum, determined by the inhomogeneities in the gravitational field and the motions of the matter components at decoupling (Sachs and Wolfe 1967). This spectrum can then be compared with observations and used in determining the values of the cosmological parameters mentioned above (see Peacock 1999).

One crucial issue is why it is reasonable to use a perturbed FL model for the observable region of the universe. The key argument is that this is plausible because of the high isotropy of all observations around us when averaged on a sufficiently large spatial scale, and particularly the very low anisotropy of the CBR. The Ehlers–Geren–Sachs (EGS) theorem (Ehlers *et al.* 1968) provides a sound basis for this argument: it shows that if freely propagating CBR (obeying the Liouville equation) is exactly isotropic in an expanding universe domain $\mathcal{U}$,then the universe is exactly FL in that domain (i.e., it has exactly the RW spatially homogenous and isotropic geometry there), the point being that any inhomogeneities in the matter distribution between us and the surface of last scattering will produce anisotropies in the CBR temperature we measure. But that result does not apply to the real universe, because the CBR is not exactly isotropic. The "almost EGS" theorem (Stoeger *et al.* 1995) shows that this result is stable: almost isotropic CBR in the domain $\mathcal{U}$ implies that the universe is almost-FL in that domain. The application to the real universe comes by making a weak Copernican assumption: "we assume we are not special, so all observers in $\mathcal{U}$ (taken to be the visible part of the universe) will also see almost isotropic CBR, just as we do." The result then follows. A further argument for homogeneity of the universe comes from postulating "uniform thermal histories" (Bonnor and Ellis 1986), but that argument is yet to be completed and applied in a practical way.

## Anisotropic and Inhomogeneous Models

The FL universes are geometrically extremely special. We wish further to understand the full range of possible universe models, their dynamical behaviors, and which of them might, at some epoch, realistically represent the real universe. This enables us to see how the approximate FL models fit into this wider set of possibilities, and under what circumstances they are attractors in this set of cosmologies.

Exact solutions are characterized by their spacetime symmetries. Symmetries are characterized by the dimension $s$ of the surfaces of homogeneity and the dimension $q$ of the isotropy group at a general point, together giving the dimension $r = s + t$ of the group of isometries $G_r$ (at special points, such as a center of symmetry, $s$ can decrease and $q$ increase but always so that $r$ stays unchanged). In the case of a cosmological model, because the 4-velocity $u^a$ is invariant under isotropies, the only possible dimensions for the isotropy group are $q = 3, 1, 0$; whereas the dimension $t$ of the surfaces of homogeneity can take any value from 4 to 0. This gives the basis for a classification of cosmological spacetimes (Ellis 1967, Ellis and van Elst 1999).

When $q = 3$, we have isotropic solutions – there are no preferred spatial directions – and it is then a theorem that they must be spatially homogeneous FL universes (Ehlers 1961). When $q = 1$, we

have locally rotationally symmetric (LRS) solutions, with precisely one preferred spacelike direction at a generic point (Ellis 1967). When $q = 0$, the solutions are anisotropic in that there can be no continuous group of rotations leaving the solution invariant; however, there can be discrete isotropies in some special cases.

When $t = 4$,we have spacetime homogeneous solutions, with all physical quantities constant; they cannot expand (by [5] and [3]). Nevertheless, two cases are of interest. For $q = 1$ ($r = 5$) we find the Gödel universe, rotating everywhere with constant vorticity, which illustrates important causal anomalies (Gödel 1949, Hawking and Ellis 1973). For $q = 3$ ($r = 6$), we find the Einstein "static universe" (Einstein 1917), the unique nonexpanding FL model with $k = 1$ and $\Lambda > 0$. It is of interest because it could possibly represent the asymptotic initial state of nonsingular inflationary universe models (Ellis *et al.* 2003). The higher-symmetry models (de Sitter and anti-de Sitter universes with higher-dimensional isotropy groups) are not included here because they do not obey the energy condition [3] – they are empty universes, which can be interesting asymptotic states but are not by themselves good cosmological models.

When $t = 3$, we have spatially homogeneous evolving universe models. For $q = 0$ ($r = 3$), there are a large family of Bianchi universes, spatially homogeneous but anisotropic, characterized into nine types according to the structure constants of the Lie algebra of the three-dimensional symmetry group $G_3$. These can be "orthogonal": the fluid flow is orthogonal to the surfaces of homogeneity, or "tilted"; the latter case can have fluid rotation or acceleration, but the former cannot. They exhibit a large variety of behaviors, including power-law, oscillatory, and nonscalar singularities (Tipler *et al.* 1980). A vexed question is whether truly chaotic behavior occurs in Bianchi IX models. The behavior of large families of these models has been characterized in dynamical systems terms (Wainwright and Ellis 1996), showing the intriguing way that higher-symmetry solutions provide a "skeleton" that guides the behavior of lower-symmetry solutions in the space of spacetimes. Many Bianchi models can be shown to isotropize at late times, particularly if viscosity is present; thus, they are asymptotic to the FL universes in the far future. In some cases, Bianchi models exhibit intermediate isotropization: they are much like FL models for a large part of their life, but are very different from it both at very early and very late stages of their evolution. These could be good models of the real universe. An important theorem by Wald (1983) shows that a cosmological constant will tend to isotropize Bianchi solutions at late times. This is an indication that inflation can succeed in making anisotropic early states resemble FL models at later times. Observational properties like element abundances and CBR anisotropy patterns can be worked out in these models (some of them develop a characteristic isolated "hot spot" in the CBR sky). For $q = 1$ ($r = 4$), we have spatially homogeneous LRS models, either Kantowski Sachs or Bianchi universes, and again observations can be worked out in detail and phase planes developed showing their dynamical behavior, often isotropizing at late times. There are orthogonal and tilted cases, the latter possibly involving nonscalar singularities. For $q = 3$ ($r = 6$), we have the isotropic FL models, discussed above. Both the LRS and isotropic cases could be good models of the real universe.

When $t = 2$, we have inhomogeneous evolving models. This is a very large family, but the LRS ($q = 1, r = 2$) cases have been examined in detail; in the case of pressure-free matter, these are the Tolman–Bondi inhomogeneous models (Bondi 1947) that can be integrated exactly, and have been used for many interesting astrophysical and cosmological studies. Krasiński (1997) gives a very complete catalog of these and lower-symmetry inhomogeneous models and their uses in cosmology. A considerable challenge is the dynamical systems analysis for generic inhomogeneous models, needed to properly understand the early evolution of generic universe models (Uggla *et al.* 2003), and hence to determine what is generic behavior.

## The Origin of the Universe

The issue underlying all this is what led to the initial conditions for the universe, for example, providing the starting conditions for inflation. There are many approaches to studying the quantum gravity phase of cosmology, including the Wheeler–de Witt equation, the path-integral approach, string cosmology, pre-big bang theory, brane cosmology, the ekpyrotic universe, the cyclic universe, and loop quantum gravity approaches. These lie beyond the purview of the present article, except to say that they are all based on unproven extrapolations of known physics. The physically possible paths will become clearer as the nature of quantum gravity is elucidated.

It is pertinent to note that there exist nonsingular realistic cosmological solutions, possible in the light of the violations of the energy condition enabled by the supposed scalar fields that underlie inflationary universe theory. These nonsingular solutions can even avoid the quantum gravity era (Ellis *et al.* 2003). However, they have very fine-tuned initial conditions, which is nowadays considered as a disadvantage; but

there is no proof that whatever processes led to the existence of the universe preferred generic rather than fine-tuned conditions; this is a philosophical rather than physical assumption. It may well be that, as regards the start of the universe, the options are that either an initial singularity occurred, or the initial conditions were very finely tuned and allowed an infinitely existing universe. Investigation of whether this conjecture is in fact valid, and if so which is the best option, are intriguing open topics.

*See also:* Einstein Equations: Exact Solutions; Einstein–Cartan Theory; General Relativity: Experimental Tests; General Relativity: Overview; Gravitational Lensing; Lie Groups: General Theory; Newtonian Limit of General Relativity; Quantum Cosmology; Shock Wave Refinement of the Friedman–Robertson–Walker Metric; Spacetime Topology, Causal Structure and Singularities; String Theory: Phenomenology.

## Further Reading

Bardeen JM (1980) *Physical Review D* 22: 1882.
Bondi H (1947) *Monthly Notices of the Royal Astronomical Society* 107: 410.
Bonnor WB and Ellis GFR (1986) *Monthly Notices of the Royal Astronomical Society* 218: 605.
Ehlers J (1961) *Abh Mainz Akad Wiss u Lit* (translated in Gen Rel Grav 25: 1225, 1993).
Ehlers J, Geren P, and Sachs RK (1968) *Journal of Mathematical Physics* 9: 1344.
Ehlers J and Rindler W (1989) *Monthly Notices of the Royal Astronomical Society* 238: 503.
Einstein A (1917) *Sitz Ber Preuss Akad Wiss* (translated in *The Principle of Relativity*, 1993). Dover.
Ellis GFR (1967) *Journal of Mathematical Physics* 8: 1171.
Ellis GFR (1971) In: Sachs RK (ed.) *General Relativity and Cosmology*, Proc. Int. School of Physics "Enrico Fermi," Course XLVII, p. 104. Academic Press.
Ellis GFR and Bruni M (1989) *Physical Review D* 40: 1804.
Ellis GFR and van Elst H (1999) In: Lachieze-Ray M (ed.) *Theoretical and Observational Cosmology*, vol. 541 [gr-qc/9812046], Nato Series C: Mathematical and Physical Sciences: Kluwer.
Ellis GFR and Madsen M (1991) *Classical and Quantum Gravity* 8: 667.
Ellis GFR, Murugan J, and Tsagas CG (2003) gr-qc/0307112.
Ellis GFR, Nel SD, Stoeger W, Maartens R, and Whitman AP (1985) *Physics Reports* 124(5 and 6): 315.
Gödel K (1949) *Reviews of Modern Physics* 21: 447.
Guth A (1981) *Physical Review D* 23: 347.
Hawking SW and Ellis GFR (1973) *The Large Scale Structure of Space Time*. Cambridge: Cambridge University Press.
Krasiński A (1997) *Inhomogeneous Cosmological Models*. Cambridge: Cambridge University Press.
Kristian J and Sachs RK (1966) *The Astrophysical Journal* 143: 379.
Lachieze-Rey M and Luminet JP (1995) *Physics Reports* 254: 135–214.
Robertson HP (1933) *Reviews of Modern Physics* 5: 62.
Peacock JA (1999) *Cosmological Physics*. Cambridge: Cambridge University Press.
Rindler W (1956) *Monthly Notices of the Royal Astronomical Society* 116: 662.
Sachs RK and Wolfe A (1967) *Astrophysical Journal* 147: 73.
Sandage A (1961) *Astrophysical Journal* 133: 355.
Stoeger W, Maartens R, and Ellis GFR (1995) *Astrophysical Journal* 443: 1.
Tipler FJ, Clarke CJS, and Ellis GFR (1980) In: Held A (ed.) *General Relativity and Gravitation: One Hundred Years after the Birth of Albert Einstein,* vol. 2, p. 97. Plenum.
Uggla C, van Elst H, Wainwright J, and Ellis GFR (2003) *Physical Review D* gr-qc/0304002 (to appear).
Wainwright J and Ellis GFR (eds.) (1996) *The Dynamical Systems Approach to Cosmology*. Cambridge: Cambridge University Press.
Wald RM (1983) *Physical Review D* 28: 2118.

# Cotangent Bundle Reduction

**J-P Ortega**, Université de Franche-Comté, Besançon, France
**T S Ratiu**, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland



## Introduction

The general symplectic reduction theory (*see* Symmetry and Symplectic Reduction) becomes much richer and has many applications if the symplectic manifold is the cotangent bundle $(T^*Q, \Omega_Q = -\boldsymbol{d}\Theta_Q)$ of a manifold $Q$. The canonical 1-form $\Theta_Q$ on $T^*Q$ is given by $\Theta_Q(\alpha_q)(V_{\alpha_q}) = \alpha_q(T_{\alpha_q}\pi_Q(V_{\alpha_q}))$, for any $q \in Q$, $\alpha_q \in T_q^*Q$, and tangent vector $V_{\alpha_q} \in T_{\alpha_q}(T^*Q)$, where $\pi_Q : T^*Q \to Q$ is the cotangent bundle projection and $T_{\alpha_q}\pi_Q : T_{\alpha_q}(T^*Q) \to T_qQ$ is its tangent map (or derivative) at $q$. In natural cotangent bundle coordinates $(q^i, p_i)$, we have $\Theta_Q = p_i\boldsymbol{d}q^i$ and $\Omega_Q = \boldsymbol{d}q^i \wedge \boldsymbol{d}p_i$.

Let $\Phi : G \times Q \to Q$ be a left smooth action of the Lie group $G$ on the manifold and $Q$. Denote by $g \cdot q = \Phi(g, q)$ the action of $g \in G$ on the point $q \in Q$ and by $\Phi_g : Q \to Q$ the diffeomorphism of $Q$ induced by $g$. The lifted left action $G \times T^*Q \to T^*Q$, given by $g \cdot \alpha_q = T^*_{g\cdot q}\Phi_{g^{-1}}(\alpha_q)$ for $g \in G$ and $\alpha_q \in T_q^*Q$, preserves $\Theta_Q$, and admits the equivariant momentum map $\boldsymbol{J} : T^*Q \to \mathfrak{g}^*$ whose expression is $\langle \boldsymbol{J}(\alpha_q), \xi \rangle = \alpha_q((\xi_Q(q))$, where $\xi \in \mathfrak{g}$, the Lie algebra of $G$, $\langle\, , \rangle : \mathfrak{g}^* \times \mathfrak{g} \to \mathbb{R}$ is the duality pairing between the dual $\mathfrak{g}^*$ and $\mathfrak{g}$, and $\xi_Q(q) = \mathrm{d}\Phi(\exp t\xi, q)/\mathrm{d}t|_{t=0}$ is the value of the

infinitesimal generator vector field $\xi_Q$ of the $G$-action at $q \in Q$ (*see* Hamiltonian Group Actions and Symmetries and Conservation Laws). Throughout this article, it is assumed that the $G$-action on $Q$, and hence on $T^*Q$, is free and proper. Recall also that $((T^*Q)_\mu, (\Omega_Q)_\mu)$ denotes the reduced manifold at $\mu \in \mathfrak{g}^*$ (*see* Symmetry and Symplectic Reduction), where $(T^*Q)_\mu := \mathbf{J}^{-1}(\mu)/G_\mu$ is the orbit space of the $G_\mu$-action on the momentum level manifold $\mathbf{J}^{-1}(\mu)$ and $G_\mu := \{g \in G \mid \mathrm{Ad}_g^*\mu = \mu\}$ is the isotropy subgroup of the coadjoint representation of $G$ on $\mathfrak{g}^*$. The left-coadjoint representation of $g \in G$ on $\mu \in \mathfrak{g}^*$ is denoted by $\mathrm{Ad}_{g^{-1}}^*\mu$.

Cotangent bundle reduction at zero is already quite interesting and has many applications. Let $\rho: Q \to Q/G$ be the $G$-principal bundle projection defined by the proper free action of $G$ on $Q$, usually referred to as the shape space bundle. Zero is a regular value of $\mathbf{J}$ and the map $\varphi_0 : ((T^*Q)_0, (\Omega_Q)_0) \to (T^*(Q/G), \Omega_{Q/G})$ given by $\varphi_0([\alpha_q])(T_q\rho(v_q)) := \alpha_q(v_q)$, where $\alpha_q \in \mathbf{J}^{-1}(0)$, $[\alpha_q] \in (T^*Q)_0$, and $v_q \in T_qQ$, is a well-defined symplectic diffeomorphism.

This theorem generalizes in two nontrivial ways when one reduces at a nonzero value of $\mathbf{J}$: an embedding and a fibration theorem.

## Embedding Version of Cotangent Bundle Reduction

Let $\mu \in \mathfrak{g}^*, Q_\mu := Q/G_\mu, \rho_\mu : Q \to Q_\mu$ the projection onto the $G_\mu$-orbit space, $\mathfrak{g}_\mu := \{\xi \in \mathfrak{g} \mid \mathrm{ad}_\xi^*\mu = 0\}$ the Lie algebra of the coadjoint isotropy subgroup $G_\mu$, where $\mathrm{ad}_\xi\eta := [\xi, \eta]$ for any $\xi, \eta \in \mathfrak{g}$, $\mathrm{ad}_\xi^* : \mathfrak{g}^* \to \mathfrak{g}^*$ the dual map, $\mu' := \mu|_{\mathfrak{g}_\mu} \in \mathfrak{g}_\mu^*$ the restriction of $\mu$ to $\mathfrak{g}_\mu$, and $((T^*Q)_\mu, (\Omega_Q)_\mu)$ the reduced space at $\mu$. The induced $G_\mu$-action on $T^*Q$ admits the equivariant momentum map $\mathbf{J}^\mu : T^*Q \to \mathfrak{g}_\mu^*$ given by $\mathbf{J}^\mu(\alpha_q) = \mathbf{J}(\alpha_q)|_{\mathfrak{g}_\mu}$. Assume there is a $G_\mu$-invariant 1-form $\alpha_\mu$ on $Q$ with values in $(\mathbf{J}^\mu)^{-1}(\mu')$. Then there is a unique closed 2-form $\beta_\mu$ on $Q_\mu$ such that $\rho_\mu^*\beta_\mu = \mathbf{d}\alpha_\mu$. Define the magnetic term $B_\mu := \pi_{Q_\mu}^*\beta_\mu$, where $\pi_{Q_\mu}$ : $T^*Q_\mu \to Q_\mu$ is the cotangent bundle projection, which is a closed 2-form on $T^*Q_\mu$. Then the map $\varphi_\mu : ((T^*Q)_\mu, (\Omega_Q)_\mu) \to (T^*Q_\mu, \Omega_{Q_\mu} - B_\mu)$ given by $\varphi_\mu([\alpha_q])(T_q\rho_\mu(v_q)) := (\alpha_q - \alpha_\mu(q))(v_q)$, for $\alpha_q \in \mathbf{J}^{-1}(\mu)$, $[\alpha_q] \in (T^*Q)_\mu$, and $v_q \in T_qQ$, is a symplectic embedding onto a submanifold of $T^*Q_\mu$ covering the base $Q_\mu$. The embedding $\varphi_\mu$ is a diffeomorphism onto $T^*Q_\mu$ if and only if $\mathfrak{g} = \mathfrak{g}_\mu$. If the 1-form $\alpha_\mu$ takes values in the smaller set $\mathbf{J}^{-1}(\mu)$ then the image of $\varphi_\mu$ is the the vector sub-bundle $[T\rho_\mu(VQ)]^\circ$ of $T^*Q_\mu$, where $VQ \subset TQ$ is the vertical vector sub-bundle consisting of vectors tangent to the $G$-orbits, that is, its fiber at $q \in Q$ equals $V_qQ = \{\xi_Q(q) \mid \xi \in \mathfrak{g}\}$, and $^\circ$ denotes the annihilator relative to the natural duality pairing between $TQ_\mu$ and $T^*Q_\mu$. Note that if $\mathfrak{g}$ is abelian or $\mu = 0$, the embedding $\varphi_\mu$ is always onto and thus the reduced space is again, topologically, a cotangent bundle.

It should be noted that there is a choice in this theorem, namely the 1-form $\alpha_\mu$. Whereas the reduced symplectic space $((T^*Q)_\mu, (\Omega_Q)_\mu)$ is intrinsic, the symplectic structure on the space $T^*Q_\mu$ depends on $\alpha_\mu$. The theorem above states that no matter how $\alpha_\mu$ is chosen, there is a symplectic diffeomorphism, which also depends on $\alpha_\mu$, of the reduced space onto a submanifold of $T^*Q_\mu$.

## Connections

The 1-form $\alpha_\mu$ is usually obtained from a left connection on the principal bundle $\rho_\mu : Q \to Q/G_\mu$ or $\rho : Q \to Q/G$. A left connection 1-form $A \in \Omega^1(Q; \mathfrak{g})$ on the left principal $G$-bundle $\rho : Q \to Q/G$ is a Lie algebra-valued 1-form $A : TQ \to \mathfrak{g}$, where $\mathfrak{g}$ denotes the Lie algebra of $G$, satisfying the conditions $A(\xi_Q) = \xi$ for all $\xi \in \mathfrak{g}$ and $\mathcal{A}(T_q\Phi_g(v)) = \mathrm{Ad}_g(\mathcal{A}(v))$ for all $g \in G$ and $v \in T_qQ$, where $\mathrm{Ad}_g$ denotes the adjoint action of $G$ on $\mathfrak{g}$. The horizontal vector sub-bundle $HQ$ of the connection $A$ is defined as the kernel of $A$, that is, its fiber at $q \in Q$ is the subspace $H_q := \ker A(q)$. The map $v_q \mapsto \mathrm{ver}_q\ (v_q) := [A(q)(v_q)]_Q(q)$ is called the vertical projection, while the map $v_q \mapsto \mathrm{hor}_q(v_q) := v_q - \mathrm{ver}_q(v_q)$ is called the horizontal projection. Since for any vector $v_q \in T_qQ$ we have $v_q = \mathrm{ver}_q(v_q) + \mathrm{hor}_q(v_q)$, it follows that $TQ = HQ \oplus VQ$ and the maps $\mathrm{hor}_q : T_qQ \to H_qQ$ and $\mathrm{ver}_q : T_qQ \to V_qQ$ are projections onto the horizontal and vertical subspaces at every $q \in Q$.

Connections can be equivalently defined by the choice of a sub-bundle $HQ \subset TQ$ complementary to the vertical sub-bundle $VQ$ satisfying the following $G$-invariance property: $H_{g \cdot q}Q = T_q\Phi_g(H_qQ)$ for every $g \in G$ and $q \in Q$. The sub-bundle $HQ$ is called, as before, the horizontal sub-bundle and a connection 1-form $A$ is defined by setting $A(q)(\xi_Q(q) + u_q) = \xi$, for any $\xi \in \mathfrak{g}$ and $u_q \in H_qQ$.

The curvature of the connection $A$ is the Lie algebra-valued 2-form on $Q$ defined by $B(u_q, v_q) = \mathbf{d}A(\mathrm{hor}_q(u_q), \mathrm{hor}_q(v_q))$. When one replaces vectors in the exterior derivative with their horizontal projections, then the result is called the exterior covariant derivative and the preceding formula for $B$ is often written as $B = DA$. Curvature measures the lack of integrability of the horizontal distribution, namely $B(u, v) = -A([\mathrm{hor}(u), \mathrm{hor}(v)])$ for any two vector fields $u$ and $v$ on $Q$. The Cartan structure equations state that $B(u, v) = \mathbf{d}\mathcal{A}(u, v) - [\mathcal{A}(u), \mathcal{A}(v)]$, where the bracket on the right hand side is the Lie bracket in $\mathfrak{g}$.

Since the connection $A$ is a Lie algebra-valued 1-form, for each $\mu \in \mathfrak{g}^*$ the formula $\alpha_\mu(q) := A(q)^*(\mu)$, where $A(q)^* : \mathfrak{g}^* \to T_q^*Q$ is the dual of the linear map $A(q) : T_qQ \to \mathfrak{g}$, defines a usual 1-form on $Q$. This 1-form $\alpha_\mu$ takes values in $\boldsymbol{J}^{-1}(\mu)$ and is equivariant in the following sense: $\Phi_g^* \alpha_\mu = \alpha_{\mathrm{Ad}_g^* \mu}$ for any $g \in G$.

### Magnetic Terms and Curvature

There are two methods to construct the 1-form $\alpha_\mu$ from a connection. The first is to start with a connection 1-form $A^\mu \in \Omega^1(Q; \mathfrak{g}_\mu)$ on the principal $G_\mu$-bundle $\rho_\mu : Q \to Q/G_\mu$. Then the 1-form $\alpha_\mu := \langle \mu|_{\mathfrak{g}_\mu}, A^\mu \rangle \in \Omega^1(Q)$ is $G_\mu$-invariant and has values in $(\boldsymbol{J}^\mu)^{-1}(\mu|_{\mathfrak{g}_\mu})$. The magnetic term $B_\mu$ is the pullback to $T^*(Q/G_\mu)$ of the $\mu|_{\mathfrak{g}_\mu}$-component $\boldsymbol{d}\alpha_\mu$ of the curvature of $A^\mu$ thought of as a 2-form on the base $Q/G_\mu$.

The second method is to start with a connection $A \in \Omega^1(Q, \mathfrak{g})$ on the principal bundle $\rho : Q \to Q/G$, to define $\alpha_\mu := \langle \mu, A \rangle \in \Omega^1(Q)$, and to observe that this 1-form is $G_\mu$-invariant and has values in $\boldsymbol{J}^{-1}(\mu)$. The magnetic term $B_\mu$ is in this case the pullback to $T^*(Q/G_\mu)$ of the $\mu$-component $\boldsymbol{d}\alpha_\mu$ of the curvature of $A$ thought of as a 2-form on the base $Q/G_\mu$.

### The Mechanical Connection

If $(Q, \langle\langle \, , \rangle\rangle)$ is a Riemannian manifold and $G$ acts by isometries, there is a natural connection on the bundle $\rho : Q \to Q/G$, namely, define the horizontal space at a point to be the metric orthogonal to the vertical space. This connection is called the mechanical connection and its horizontal bundle consists of all vectors $v_q \in TQ$ such that $\boldsymbol{J}(\langle\langle v_q, \cdot \rangle\rangle) = 0$.

To determine the Lie algebra-valued 1-form $A$ of this connection, the notion of locked inertia tensor needs to be introduced. This is the linear map $\mathbb{I}(q) : \mathfrak{g} \to \mathfrak{g}^*$ depending smoothly on $q \in Q$ defined by the identity $\langle \mathbb{I}(q)\xi, \eta \rangle = \langle\langle \xi_Q(q), \eta_Q(q) \rangle\rangle$ for any $\xi, \eta \in \mathfrak{g}$. Since the $G$-action is free, each $\mathbb{I}(q)$ is invertible. The connection 1-form whose horizontal space was defined above is given by $A(q)(v_q) = \mathbb{I}(q)^{-1}(\boldsymbol{J}(\langle\langle v_q, \cdot \rangle\rangle))$.

Denote by $K : T^*Q \to \mathbb{R}$ the kinetic energy of the metric $\langle\langle , \rangle\rangle$ on the cotangent bundle, that is, $K(\langle\langle v_q, \cdot \rangle\rangle) := (1/2)\|v_q\|^2$. The 1-form $\alpha_\mu = A(\,\cdot\,)^*\mu$ is characterized for the mechanical connection $A$ by the condition $K(\alpha_\mu(q)) = \inf\{K(\beta_q) \mid \beta_q \in \boldsymbol{J}^{-1}(\mu) \cap T_q^*Q\}$.

### The Amended Potential

A simple mechanical system is a Hamiltonian system on a cotangent bundle $T^*Q$ whose Hamiltonian function is the the sum of the kinetic energy of a Riemannian metric on $Q$ and a potential function $V : Q \to \mathbb{R}$. If there is a Lie group $G$ acting on $Q$ by isometries and leaving the potential invariant, then we have a simple mechanical system with symmetry. The amended or effective potential $V_\mu : Q \to \mathbb{R}$ at $\mu \in \mathfrak{g}^*$ is defined by $V_\mu := H \circ \alpha_\mu$, where $\alpha_\mu$ is the 1-form associated to the mechanical connection. Its expression in terms of the locked moment of inertia tensor is given by $V_\mu(q) := V(q) + (1/2)\langle \mu, \mathbb{I}(q)^{-1}\mu \rangle$. The amended potential naturally induces a smooth function $\widehat{V}_\mu \in C^\infty(Q/G_\mu)$.

The fundamental result about simple mechanical systems with symmetry is the following. The push-forward by the embedding $\varphi_\mu : ((T^*Q)_\mu, (\Omega_Q)_\mu) \to (T^*Q_\mu, \Omega_{Q_\mu} - B_\mu)$ of the reduced Hamiltonian $H_\mu \in C^\infty((T^*Q)_\mu)$ of a simple mechanical system $H = K + V \circ \pi_Q \in C^\infty(T^*Q)$ is the restriction to the vector sub-bundle $\varphi_\mu((T^*Q)_\mu) \subset T^*(Q/G_\mu)$, which is also a symplectic submanifold of $(T^*(Q/G_\mu), \Omega_{Q/G_\mu} - B_\mu)$, of the simple mechanical system on $T^*(Q/G_\mu)$ whose kinetic energy is given by the quotient Riemannian metric on $Q/G_\mu$ and whose potential is $\widehat{V}_\mu$. However, Hamilton's equations on $T^*(Q/G_\mu)$ for this simple mechanical system are computed relative to the magnetic symplectic form $\Omega_{Q/G_\mu} - B_\mu$.

There is a wealth of applications starting from this classical theorem to mechanical systems, spanning such diverse areas as topological characterization of the level sets of the energy–momentum map to methods of proving nonlinear stability of relative equilibria (block-diagonalization of the stability form in the application of the energy–momentum method).

### Fibration Version of Cotangent Bundle Reduction

There is a second theorem that realizes the reduced space of a cotangent bundle as a locally trivial bundle over shape space $Q/G$. This version is particularly well suited in the study of quantization problems and in control theory. The result is the following. Assume that $G$ acts freely and properly on $Q$. Then the reduced symplectic manifold $(T^*Q)_\mu$ is a fiber bundle over $T^*(Q/G)$ with fiber the coadjoint orbit $\mathcal{O}_\mu$. How this is related to the Poisson structure of the quotient $(T^*Q)/G$ will be discussed later.

### The Kaluza–Klein Construction

The extra term in the symplectic form of the reduced space is called a magnetic term because it has this interpretation in electromagnetism. To understand why $B_\mu$ is called a magnetic term, consider the problem of a particle of mass $m$ and charge $e$ moving in $\mathbb{R}^3$ under the influence of a given

magnetic field $\boldsymbol{B} = B_x \boldsymbol{i} + B_y \boldsymbol{j} + B_z \boldsymbol{k}, \operatorname{div} \boldsymbol{B} = 0$. The Lorentz force law (written in the International System) gives the equations of motion

$$m \frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} = e\boldsymbol{v} \times \boldsymbol{B} \qquad [1]$$

where $e$ is the charge and $\boldsymbol{v} = (\dot{x}, \dot{y}, \dot{z}) = \dot{\boldsymbol{q}}$ is the velocity of the particle. What is the Hamiltonian description of these equations?

There are two possible answers to this question. To formulate them, associate to the divergence free vector field $\boldsymbol{B}$ the closed 2-form $B = B_x \boldsymbol{d}y \wedge \boldsymbol{d}z - B_y \boldsymbol{d}x \wedge \boldsymbol{d}z + B_z \boldsymbol{d}x \wedge \boldsymbol{d}y$. Also, write $\boldsymbol{B} = \operatorname{curl} \boldsymbol{A}$ for some other vector field $\boldsymbol{A} = (A_x, A_y, A_z)$ on $\mathbb{R}^3$, called the magnetic potential.

*Answer 1* Take on $T^*\mathbb{R}^3$ the symplectic form $\Omega_B = \boldsymbol{d}x \wedge \boldsymbol{d}p_x + \boldsymbol{d}y \wedge \boldsymbol{d}p_y + \boldsymbol{d}z \wedge \boldsymbol{d}p_z - eB$, where $(p_x, p_y, p_z) = \boldsymbol{p} := m\boldsymbol{v}$ is the momentum of the particle, and $h = m\|\boldsymbol{v}\|^2/2 = m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)/2$ is the Hamiltonian, the kinetic energy of the particle. A direct verification shows that $\boldsymbol{d}h = \Omega_B(X_h, \cdot\,)$, where

$$X_h = \dot{x}\frac{\partial}{\partial x} + \dot{y}\frac{\partial}{\partial y} + \dot{z}\frac{\partial}{\partial z} + e(B_z\dot{y} - B_y\dot{z})\frac{\partial}{\partial p_x} + e(B_x\dot{z} - B_z\dot{x})\frac{\partial}{\partial p_y} + e(B_y\dot{x} - B_z\dot{x})\frac{\partial}{\partial p_z} \qquad [2]$$

which gives the equations of motion [1].

*Answer 2* Take on $T^*\mathbb{R}^3$ the canonical symplectic form $\Omega = \boldsymbol{d}x \wedge \boldsymbol{d}p_x + \boldsymbol{d}y \wedge \boldsymbol{d}p_y + \boldsymbol{d}z \wedge \boldsymbol{d}p_z$ and the Hamiltonian $h_A = \|\boldsymbol{p} - e\boldsymbol{A}\|^2/2m$. A direct verification shows that $\boldsymbol{d}h_A = \Omega(X_{h_A}, \cdot\,)$, where $X_{h_A}$ has the same expression [2].

Next we show how the magnetic term in the symplectic form $\Omega_B$ is obtained by reduction from the Kaluza–Klein system. Let $Q = \mathbb{R}^3 \times S^1$ with the circle $G = S^1$ acting on $Q$, only on the second factor. Identify the Lie algebra $\mathfrak{g}$ of $S^1$ with $\mathbb{R}$. Since the infinitesimal generator of this action defined by $\xi \in \mathfrak{g} = \mathbb{R}$ has the expression $\xi_Q(\boldsymbol{q}, \theta) = (\boldsymbol{q}, \theta; \boldsymbol{0}, \xi)$, if $TS^1$ is trivialized as $S^1 \times \mathbb{R}$, a momentum map $\boldsymbol{J} : T^*Q = \mathbb{R}^3 \times S^1 \times \mathbb{R}^3 \times \mathbb{R} \to \mathfrak{g}^* = \mathbb{R}$ is given by $\boldsymbol{J}(\boldsymbol{q}, \theta; \boldsymbol{p}, p)\xi = (\boldsymbol{p}, p) \cdot (\boldsymbol{0}, \xi) = p\xi$, that is, $\boldsymbol{J}(\boldsymbol{q}, \theta; \boldsymbol{p}, p) = p$. In this case, the coadjoint action is trivial, so for any $\mu \in \mathfrak{g}^* = \mathbb{R}$, we have $G_\mu = S^1, \mathfrak{g}_\mu = \mathbb{R}$, and $\mu' = \mu$. The 1-form $\alpha_\mu = \mu(A_x\boldsymbol{d}x + A_y\boldsymbol{d}y + A_z\boldsymbol{d}z + \boldsymbol{d}\theta) \in \Omega^1(Q)$, where $\boldsymbol{d}\theta$ denotes the length 1-form on $S^1$, is clearly $G_\mu = S^1$-invariant, has values in $\boldsymbol{J}^{-1}(\mu) = \{(\boldsymbol{q}, \theta; \boldsymbol{p}, \mu) \mid \boldsymbol{q}, \boldsymbol{p} \in \mathbb{R}^3, \theta \in S^1\}$, and its exterior differential equals $\boldsymbol{d}\alpha_\mu = \mu B$. Thus, the closed 2-form $\beta_\mu$ on the base $Q_\mu = Q/G_\mu = Q/S^1 = \mathbb{R}^3$ equals $\mu B$ and hence the magnetic term, that is, the closed 2-form $B_\mu = \pi^*_{Q_\mu}\beta_\mu$ on $T^*Q_\mu = T^*\mathbb{R}^3$, is also $\mu B$ since $\pi_{Q_\mu} : Q = \mathbb{R}^3 \times S^1 \to Q/G_\mu = \mathbb{R}^3$ is the projection. Therefore, the reduced space $(T^*Q)_\mu$ is symplectically diffeomorphic to $(T^*\mathbb{R}^3, \boldsymbol{d}x \wedge \boldsymbol{d}p_x + \boldsymbol{d}y \wedge \boldsymbol{d}p_y + \boldsymbol{d}z \wedge \boldsymbol{d}p_z - \mu B)$, which coincides with the phase space in Answer 1 if we put $\mu = e$. This also gives the physical interpretation of the momentum map $\boldsymbol{J} : T^*Q = \mathbb{R}^3 \times S^1 \times \mathbb{R}^3 \times \mathbb{R} \to \mathfrak{g}^* = \mathbb{R}$, $\boldsymbol{J}(\boldsymbol{q}, \theta; \boldsymbol{p}, p) = p$ and hence of the variable conjugate to the circle variable $\theta$: $p$ represents the charge. Moreover, the magnetic term in the symplectic form is, up to a charge factor, the magnetic field.

The kinetic energy Hamiltonian

$$h(\boldsymbol{q}, \theta; \boldsymbol{p}, p) := \frac{1}{2m}\|\boldsymbol{p}\|^2 + \frac{1}{2}p^2$$

of the Kaluza–Klein metric, that is, the Riemannian metric obtained by keeping the standard metrics on each factor and declaring $\mathbb{R}^3$ and $S^1$ orthogonal, induces the reduced Hamiltonian

$$h_\mu(\boldsymbol{q}) = \frac{1}{2m}\|\boldsymbol{p}\|^2 + \frac{1}{2}\mu^2$$

which, up to the constant $\mu^2/2$, equals the kinetic energy Hamiltonian in Answer 1. Note that this reduced system is not the geodesic flow of the Euclidean metric because of the presence of the magnetic term in the symplectic form. However, the equations of motion of a charged particle in a magnetic field are obtained by reducing the geodesic flow of the Kaluza–Klein metric.

A similar construction is carried out in Yang–Mills theory where $A$ is a connection on a principal bundle and $B$ is its curvature. Magnetic terms also appear in classical mechanics. For example, in rotating systems the Coriolis force (up to a dimensional factor) plays the role of the magnetic term.

## Reconstruction of Dynamics for Cotangent Bundles

A general reconstruction method of the dynamics from the reduced dynamics was given in (*see* Symmetry and Symplectic Reduction). For cotangent bundles, using the mechanical connection, this method simplifies considerably.

Start with the following general situation. Let $G$ act freely on the configuration manifold $Q$; let $h : T^*Q \to \mathbb{R}$ be a $G$-invariant Hamiltonian, $\mu \in \mathfrak{g}^*$, $\alpha_q \in \boldsymbol{J}^{-1}(\mu)$, and $c_\mu(t)$ the integral curve of the reduced system with initial condition $[\alpha_q] \in (T^*Q)_\mu$ given by the reduced Hamiltonian function $h_\mu : (T^*Q)_\mu \to \mathbb{R}$. In terms of a connection $A \in \Omega^1(\boldsymbol{J}^{-1}(\mu); \mathfrak{g}_\mu)$ on the left $G_\mu$-principal bundle $\boldsymbol{J}^{-1}(\mu) \to (T^*Q)_\mu$ the reconstruction procedure proceeds in four steps:

- *Step 1*: Horizontally lift the curve $c_\mu(t) \in (T^*Q)_\mu$ to a curve $d(t) \in \boldsymbol{J}^{-1}(\mu)$ with $d(0) = \alpha_q$.
- *Step 2*: Set $\xi(t) = A(d(t))(X_h(d(t))) \in \mathfrak{g}_\mu$.

- *Step 3*: With $\xi(t) \in \mathfrak{g}_\mu$ determined in step 2, solve the nonautonomous differential equation $\dot{g}(t) = T_e L_{g(t)} \xi(t)$ with initial condition $g(0) = e$, where $L_g$ denotes left translation on $G$; this is the step that involves "quadratures" and is the main obstacle to finding explicit formulas.
- *Step 4*: The curve $c(t) = g(t) \cdot d(t)$, with $d(t)$ found in step 1 and $g(t)$ found in step 3 is the integral curve of $X_h$ with initial condition $c(0) = \alpha_q$.

This method depends on the choice of the connection $A \in \Omega^1(\mathbf{J}^{-1}(\mu); \mathfrak{g}_\mu)$. Here are several particular cases when this procedure simplifies.

(a) *One-dimensional coadjoint isotropy group*. If $G_\mu = S^1$ or $G_\mu = \mathbb{R}$, identify $\mathfrak{g}_\mu$ with $\mathbb{R}$ via the map $a \in \mathbb{R} \leftrightarrow a\zeta \in \mathfrak{g}_\mu$, where $\zeta \in \mathfrak{g}_\mu, \zeta \neq 0$, is a generator of $\mathfrak{g}_\mu$. Then a connection 1-form on the $S^1$ (or $\mathbb{R}$) principal bundle $\mathbf{J}^{-1}(\mu) \to (T^*Q)_\mu$ is the 1-form $A$ on $\mathbf{J}^{-1}(\mu)$ given by $A = (1/\langle \mu, \zeta \rangle)\theta_\mu$, where $\theta_\mu$ is the pullback of the canonical 1-form $\theta \in \Omega^1(T^*Q)$ to the submanifold $\mathbf{J}^{-1}(\mu)$. The curvature of this connection is the 2-form on $(T^*Q)_\mu$ given by $\operatorname{curv}(A) = -(1/\langle \mu, \zeta \rangle)\omega_\mu$, where $\omega_\mu$ is the reduced symplectic form on $(T^*Q)_\mu$. In this case, the curve $\xi(t) \in \mathfrak{g}_\mu$ in step 2 is given by $\xi(t) = \Lambda[h](d(t))$, where $\Lambda \in \mathfrak{X}(T^*Q)$ is the Liouville vector field characterized by the property of being the unique vector field on $T^*Q$ that satisfies the relation $\mathbf{d}\theta(\Lambda, \cdot) = \theta$. In canonical coordinates $(q^i, p_i)$ on $T^*Q, \Lambda = p_i \frac{\partial}{\partial p_i}$.

(b) *Induced connection*. Any connection $\mathcal{A} \in \Omega^1(Q; \mathfrak{g}_\mu)$ on the left principal bundle $Q \to Q/G_\mu$ induces a connection $A \in \Omega^1(\mathbf{J}^{-1}(\mu); \mathfrak{g}_\mu)$ by $A(\alpha_q) \times (V_{\alpha_q}) := \mathcal{A}(q)(T_{\alpha_q}\pi_Q(V_{\alpha_q}))$, where $q \in Q, \alpha_q \in T_q^*Q$, $V_{\alpha_q} \in T_{\alpha_q}(T^*Q)$, and $\pi_Q : T^*Q \to Q$ is the cotangent bundle projection. In this case, the curve $\xi(t) \in \mathfrak{g}_\mu$ in step 2 is given by $\xi(t) = \mathcal{A}(q(t))(\mathbb{F}h(d(t))$, where $q(t) := \pi_Q(d(t))$ is the base integral curve and the vector bundle morphism $\mathbb{F}h : T^*Q \to TQ$ is the fiber derivative of $h$ given by

$$\mathbb{F}h(\alpha_q)(\beta_q) := \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} h(\alpha_q + t\beta_q)$$

for any $\alpha_q, \beta_q \in T_a^*Q$. Two particular instances of this situation are noteworthy.

(b1) Assume that the Hamiltonian $h$ is that of a simple mechanical system with symmetry. Choosing $A$ to be the mechanical connection $A_{\text{mech}}$, the curve $\xi(t) \in \mathfrak{g}_\mu$ in step 2 is given by $\xi(t) = A_{\text{mech}}(q(t))\,(\langle\langle d(t), \cdot \rangle\rangle)$.

(b2) If $Q = G$ is a Lie group, $\dim G_\mu = 1$, and $\zeta$ is a generator of $\mathfrak{g}_\mu$, then the connection $A \in \Omega^1(G)$ can be chosen to equal $\mathcal{A}(g) := (1/\langle \mu, \zeta \rangle)$ $T_g^* R_{g^{-1}}(\mu)$, where $\zeta$ is a generator of $\mathfrak{g}_\mu$ and $R_g$ is right translation on $G$.

(c) *Reconstruction of dynamics for simple mechanical systems with symmetry*. The case of simple mechanical systems with symmetry deserves special attention since several steps in the reconstruction method can be simplified. For simple mechanical systems, the knowledge of the base integral curve $q(t)$ suffices to determine the entire integral curve on $T^*Q$. Indeed, if $h = K + V \circ \pi_q$ is the Hamiltonian, the Legendre transformation $\mathbb{F}h : T^*Q \to TQ$ determines the Lagrangian system on $TQ$ given by $\ell(u_q) = (1/2)\|u_q\|^2 - V(u_q)$, for $u_q \in T_qQ$. Lagrange's equations are second-order and thus the evolution of the velocities is given by the time derivative $\dot{q}(t)$ of the base integral curve. Since $\mathbb{F}h = (\mathbb{F}\ell)^{-1}$, the solution of the Hamiltonian system is given by $\mathbb{F}\ell(\dot{q}(t))$. Using the explicit expression of the mechanical connection and the notation given in the general procedure, the method of reconstruction simplifies to the following steps. To find the integral curve $c(t)$ of the simple mechanical system with $G$-symmetry $h = K + V \circ \pi_Q$ on $T^*Q$ with initial condition $c(0) = \alpha_q \in T_q^*Q$, knowing the integral curve $c_\mu(t)$ of the reduced Hamiltonian system on $(T^*Q)_\mu$ given by the reduced Hamiltonian function $h_\mu : (T^*Q)_\mu \to \mathbb{R}$ with initial condition $c_\mu(0) = [\alpha_q]$ one proceeds in the following manner. Recall the symplectic embedding $\varphi_\mu : ((T^*Q)_\mu, (\Omega_Q)_\mu) \to (T^*(Q/G_\mu), \Omega_{Q/G_\mu} - B_\mu)$. The curve $\varphi_\mu(c_\mu(t)) \in T^*(Q/G_\mu)$ is an integral curve of the Hamiltonian system on $(T^*(Q/G_\mu), \Omega_{Q/G_\mu} - B_\mu)$ given by the function that is the sum of the kinetic energy of the quotient Riemannian metric and the quotient amended potential $\widehat{V}_\mu$. Let $q_\mu(t) := \pi_{Q/G_\mu}(c_\mu(t))$ be the base integral curve of this system, where $\pi_{Q/G_\mu} : T^*(Q/G_\mu) \to Q/G_\mu$ is the cotangent bundle projection.

- *Step 1*: Relative to the mechanical connection $A_{\text{mech}} \in \Omega^1(Q; \mathfrak{g}_\mu)$, horizontally lift $q_\mu(t) \in Q/G_\mu$ to a curve $q_h(t) \in Q$ passing through $q_h(0) = q$.
- *Step 2*: Determine $\xi(t) \in \mathfrak{g}_\mu$ from the algebraic system $\langle\langle \xi(t)_Q(q_h(t)), \eta_Q(q_h(t)) \rangle\rangle = \langle \mu, \eta \rangle$ for all $\eta \in \mathfrak{g}_\mu$, where $\langle\langle \cdot, \cdot \rangle\rangle$ is the $G$-invariant kinetic energy Riemannian metric on $Q$. This implies that $\dot{q}_h(0)$ and $\xi(0)_Q(q)$ are the horizontal and vertical components of the vector $\alpha_q^\sharp \in T_qQ$ which is associated by the metric $\langle\langle \cdot, \cdot \rangle\rangle$ to the initial condition $\alpha_q$.
- *Step 3*: Solve $\dot{g}(t) = T_e L_{g(t)} \xi(t)$ in $G_\mu$ with initial condition $g(0) = e$.
- *Step 4*: The curve $q(t) := g(t) \cdot q_h(t)$, with $q_h(t)$ and $g(t)$ determined in steps 2 and 4, respectively, is the base integral curve of the simple mechanical system with symmetry defined by the function $h$ satisfying $q(0) = 0$. The curve $(\mathbb{F}h)^{-1}(\dot{q}(t)) \in T^*Q$ is the integral curve of this system with initial

condition $c(0)=\alpha_q$. In addition, $q'(t)=g(t)\cdot(\dot{q}_h(t)+\xi(t)_Q(q_h(t)))$ is the horizontal plus vertical decomposition relative to the connection induced on $\boldsymbol{J}^{-1}(\mu)\to(T^*Q)_\mu$ by the mechanical connection $A_{\mathrm{mech}}\in\Omega^1(Q;\mathfrak{g}_\mu)$.

There are several important situations when step 3, the main obstruction to an explicit solution of the reconstruction problem, can be carried out. We shall review some of them below.

(c1) *The case* $G_\mu=S^1$. If $G_\mu$ is abelian, the equation in step 3 has the solution $g(t)=\exp\int_0^t\xi(s)\mathrm{d}s$. If, in addition, $G_\mu=S^1$, then $\xi(s)$ can be explicitly determined by step 2. Indeed, if $\zeta\in\mathfrak{g}_\mu$ is a generator of $\mathfrak{g}_\mu$, writing $\xi(s)=a(s)\zeta$ for some smooth real-valued function $a$ defined on some open interval around the origin, the algebraic equation in step 2 implies that $\langle\langle a(s)\xi(t)_Q(q_h(t)),\zeta_Q(q_h(t))\rangle\rangle=\langle\mu,\zeta\rangle$, which gives $a(s)=\langle\mu,\zeta\rangle/\|\zeta_Q(q_h(s))\|^2$. Therefore, the base integral curve of the solution of the simple mechanical system with symmetry on $T^*Q$ passing through $q$ is

$$q(t)=\exp\left(\langle\mu,\zeta\rangle\int_0^t\frac{\mathrm{d}s}{\|\zeta_Q(q_h(s))\|^2}\zeta\right)\cdot q_h(t)$$

and

$$\dot{q}(t)=\exp\left(\langle\mu,\zeta\rangle\int_0^t\frac{\mathrm{d}s}{\|\zeta_Q(q_h(s))\|^2}\zeta\right)\times\left(\dot{q}_h(t)+\frac{\langle\mu,\zeta\rangle}{\|\zeta_Q(q_h(s))\|^2}\zeta_Q(q_h(t))\right)$$

(c2) *The case of compact Lie groups*. An obvious situation when the differential equation in step 3 can be solved is if $\xi(t)=\xi$ for all $t$, where $\xi$ is a given element of $\mathfrak{g}_\mu$. Then the solution is $g(t)=\exp(t\xi)$. However, step 2 puts certain restrictions under this hypothesis, because it requires that $\langle\langle\xi(t)_Q(q_h(t)),\eta_Q(q_h(t))\rangle\rangle=\langle\mu,\eta\rangle$ for any $\eta\in\mathfrak{g}_\mu$. This is satisfied if there is a bilinear nondegenerate form $(\cdot,\cdot)$ on $\mathfrak{g}$ satisfying $(\zeta,\eta)=\langle\langle\zeta_Q(q),\eta_Q(q)\rangle\rangle$ for all $q\in Q$ and $\zeta,\eta\in\mathfrak{g}$. This implies that $(\cdot,\cdot)$ is positive definite and invariant under the adjoint action of $G$ on $\mathfrak{g}$, so semisimple Lie algebras of noncompact type are excluded. If $G$ is compact, which ensures the existence of a positive adjoint invariant inner product on $\mathfrak{g}$, and $Q=G$, this condition implies that the kinetic energy metric is invariant under the adjoint action. There are examples in which such conditions are natural, such as in Kaluza–Klein theories. Thus, if $G$ is a compact Lie group and $(\cdot,\cdot)$ is a positive-definite metric invariant under the adjoint action of $G$ on $\mathfrak{g}$ satisfying $(\zeta,\eta)=\langle\langle\zeta_Q(q),\eta_Q(q)\rangle\rangle$ for all $q\in Q$ and $\zeta,\eta\in\mathfrak{g}$, then the element $\xi(t)$ in step 2 can be chosen to be constant and is determined by the identity $(\xi,\cdot)=\mu|_{\mathfrak{g}_\mu}$ on $\mathfrak{g}_\mu$. The solution of the equation on step 3 is then $g(t)=\exp(t\xi)$.

(c3) *The case when* $\dot{\xi}(t)$ *is proportional to* $\xi(t)$. Try to find a real-valued function $f(t)$ such that $g(t)=\exp(f(t)\xi(t))$ is a solution of the equation $\dot{g}(t)=T_eL_{g(t)}\xi(t)$ with $f(0)=0$. This gives, for small $t$, the equation $\dot{f}(t)\xi(t)+f(t)\dot{\xi}(t)=\xi(t)$, that is, it is necessary that $\xi(t)$ and $\dot{\xi}(t)$ be proportional. So, if $\dot{\xi}(t)=\alpha(t)\xi(t)$ for some known smooth function $\alpha(t)$, then this gives $f(t)=\int_0^t\exp(\int_t^s\alpha(r)\mathrm{d}r)\,\mathrm{d}s$.

(c4) *The case of* $G_\mu$ *solvable*. Write $g(t)=\exp(f_1(t)\xi_1)\exp(f_2(t)\xi_2)\cdots\exp(f_n(t)\xi_n)$, for some basis $\{\xi_1,\xi_2,\ldots,\xi_n\}$ of $\mathfrak{g}_\mu$ and some smooth real-valued functions $f_i$, $i=1,2,\ldots,n$, defined around zero. It is known that if $G_\mu$ is solvable, the equation in step 3 can be solved by quadratures for the $f_i$.

### Reconstruction Phases for Simple Mechanical Systems with $S^1$ Symmetry

Consider a simple mechanical system with symmetry $G$ on the Riemannian manifold $(Q,\langle\langle\cdot,\cdot\rangle\rangle)$ with $G$-invariant potential $V\in C^\infty(Q)$. If $\mu\in\mathfrak{g}^*$, let $V_\mu$ be the amended potential and $\widehat{V}_\mu\in C^\infty(Q/G_\mu)$ the induced function on the base. Let $c:[0,T]\to T^*Q$ be an integral curve of the system with Hamiltonian $h=K+V\circ\pi_Q$ and suppose that its projection $c_\mu:[0,T]\to(T^*Q)_\mu$ to the reduced space is a closed integral curve of the reduced system with Hamiltonian $h_\mu$. The reconstruction phase associated to the loop $c_\mu(t)$ is the group element $g\in G_\mu$, satisfying the identity $c(T)=g\cdot c(0)$. We shall present two explicit formulas of the reconstruction phase for the case when $G_\mu=S^1$. Let $\zeta\in\mathfrak{g}_\mu=\mathbb{R}$ be a generator of the coadjoint isotropy algebra and write $c(T)=\exp(\varphi\zeta)\cdot c(0)$; in this case, $\varphi$ is identified with the reconstruction phase and, as we shall see in concrete mechanical examples, it truly represents an angle.

If $G_\mu=S^1$, the $G_\mu$-principal bundle $\pi_\mu:\boldsymbol{J}^{-1}(\mu)\to(T^*Q)_\mu:=\boldsymbol{J}^{-1}(\mu)/G_\mu$ admits two natural connections: $A=(1/\mu\zeta)\theta_\mu\in\Omega^1(\boldsymbol{J}^{-1}(\mu))$, where $\theta_\mu$ is the pullback of the canonical 1-form on the cotangent bundle to the momentum level submanifold $\boldsymbol{J}^{-1}(\mu)$, and $\pi_Q^*A_{\mathrm{mech}}\in\Omega^1(\boldsymbol{J}^{-1}(\mu))$. There is no reason to choose one connection over the other and thus there are two natural formulas for the reconstruction phase in this case. Let $c_\mu(t)$ be a periodic orbit of period $T$ of the reduced system and denote also by $h_\mu$ the value of the Hamiltonian function on it.

Assume that $D$ is a two-dimensional surface in $(T^*Q)_\mu$ whose boundary is the loop $c_\mu(t)$. Since the manifolds $(T^*Q)_\mu$ and $T^*(Q/S^1)$ are diffeomorphic (but not symplectomorphic), it makes sense to consider the base integral curve $q_\mu(t)$ obtained by projecting $c_\mu(t)$ to the base $Q/S^1$, which is a closed curve of period $T$. Denote by

$$\langle \widehat{V}_\mu \rangle := \frac{1}{T}\int_0^T \widehat{V}_\mu(q_\mu(t))\,\mathrm{d}t$$

the average of $\widehat{V}_\mu$ over the loop $q_\mu(t)$. Let $q_h(t)\in Q$ be the $A_{\mathrm{mech}}$-horizontal lift of $q_\mu(t)$ to $Q$ and let $\chi$ be the $A_{\mathrm{mech}}$-holonomy of the loop $q_\mu(t)$ measured from $q(0)$, the base point of $c(0)$; its expression is given by $\exp\chi = \exp(-\iint_D B)$, where $B$ is the curvature of the mechanical connection. Denote by $\omega_\mu$ the reduced symplectic form on $(T^*Q)_\mu$. With these notations the phase $\varphi$ is given by

$$\begin{aligned}\varphi &= \frac{1}{\mu\zeta}\iint_D \omega_\mu + \frac{2(h_\mu - \langle \widehat{V}_\mu\rangle)T}{\mu\zeta}\\ &= \chi + \mu\zeta\int_0^T \frac{\mathrm{d}s}{\|\zeta_Q(q_h(s))\|^2}\end{aligned} \qquad [3]$$

The first terms in both formulas are the so-called geometric phases because they carry only geometric information given by the connection, whereas the second terms are called the dynamic phases since they encapsulate information directly linked to the Hamiltonian. The expression of the total phase as a sum of a geometric and a dynamic phase is not intrinsic and is connection dependent. It can even happen that one of these summands vanishes. We shall consider now two concrete examples: the free rigid body and the heavy top.

#### Reconstruction Phases for the Free Rigid Body

The motion of the free rigid body is a geodesic with respect to a left-invariant Riemannian metric on SO(3) given by the moment of inertia of the body. The phase space of the free rigid body motion is $T^*\mathrm{SO}(3)$ and a momentum map $\boldsymbol{J}: T^*\mathrm{SO}(3)\to\mathbb{R}^3$ of the lift of left translation to the cotangent bundle is given by right translation to the identity element. We have identified here $\mathfrak{so}(3)$ with $\mathbb{R}^3$ by the Lie algebra isomorphism $\boldsymbol{x}\in(\mathbb{R}^3,\times)\mapsto\hat{\boldsymbol{x}}\in(\mathfrak{so}(3),[\cdot,\cdot])$, where $\hat{\boldsymbol{x}}(\boldsymbol{y}) = \boldsymbol{x}\times\boldsymbol{y}$, and $\mathfrak{so}(3)^*$ with $\mathbb{R}^3$ by the inner product on $\mathbb{R}^3$. The reduced manifold $\boldsymbol{J}^{-1}(\mu)/G_\mu$ is identified with the sphere $S^2_{\|\mu\|}$ in $\mathbb{R}^3$ of radius $\|\mu\|$ with the symplectic form $\omega_\mu = -\mathrm{d}S/\|\mu\|$, where $\mathrm{d}S$ is the standard area form on $S^2_{\|\mu\|}$ and $G_\mu \cong S^1$ is the group of rotations around the axis $\mu$. These concentric spheres are the coadjoint orbits of the Lie–Poisson space $\mathfrak{so}(3)^*$ and represent the level sets of the Casimir functions that are all smooth functions of $\|\Pi\|^2$, where $\Pi\in\mathbb{R}^3$ denotes the body angular momentum.

The Hamiltonian of the rigid body on the Lie–Poisson space $T^*\mathrm{SO}(3)/\mathrm{SO}(3)\cong\mathbb{R}^3$ is given by

$$h(\Pi) := \frac{1}{2}\left(\frac{\Pi_1^2}{I_1}+\frac{\Pi_2^2}{I_2}+\frac{\Pi_3^2}{I_3}\right)$$

where $I_1, I_2, I_3 > 0$ are the principal moments of inertia of the body. Let $\mathbb{I} := \mathrm{diag}(I_1, I_2, I_3)$ denote the moment of inertia tensor diagonalized in a principal-axis body frame. The Lie–Poisson bracket on $\mathbb{R}^3$ is given by $\{f,g\}(\Pi) = -\Pi\cdot(\nabla f(\Pi)\times\nabla g(\Pi))$ and the equation of motions are $\dot{\Pi} = \Pi\times\Omega$, where $\Omega\in\mathbb{R}^3$ is the body angular velocity given in terms of $\Pi$ by $\Omega_i := \Pi/I_i$, for $i = 1,2,3$, that is, $\Omega = \mathbb{I}^{-1}\Pi$. The trajectories of the these equations are found by intersecting a family of homothetic energy ellipsoids with the angular momentum concentric spheres. If $I_1 > I_2 > I_3$, one immediately sees that all orbits are periodic with the exception of four centers (the two possible rotations about the long and the short moment of inertia axis of the body), two saddles (the two rotations about the middle moment of inertia axis of the body), and four heteroclinic orbits connecting the two saddles.

Suppose that $\Pi(t)$ is a periodic orbit on the sphere $S^2_{\|\mu\|}$ with period $T$. After time $T$, by how much has the rigid body rotated in space? The answer to this question follows directly from [3]. Taking $\zeta = \mu/\|\mu\|$ and the potential $v\equiv 0$ we get

$$\begin{aligned}\varphi &= -\Lambda + \frac{2h_\mu T}{\|\mu\|}\\ &= \iint_D \frac{2\|\mathbb{I}\Pi(s)\|^2 - (\Pi(s)\cdot\mathbb{I}\Pi(s))(\mathrm{tr}\,\mathbb{I})}{(\Pi(s)\cdot\mathbb{I}\Pi(s))^2}\,\mathrm{d}s\\ &\quad + \|\mu\|^3\int_0^T \frac{\mathrm{d}s}{(\Pi(s)\cdot\mathbb{I}\Pi(s))}\end{aligned}$$

where $D$ is one of the two spherical caps on $S^2_{\|\mu\|}$ whose boundary is the periodic orbit $\Pi(t)$, $h_\mu$ is the value of the total energy on the solution $\Pi(t)$, and $\Lambda$ is the oriented solid angle, that is,

$$\Lambda := -\frac{1}{\|\mu\|}\iint_D \omega_\mu, \qquad |\Lambda| = \frac{\mathrm{area}D}{\|\mu\|^2}$$

#### Reconstruction Phases for the Heavy Top

The heavy top is a simple mechanical systems with symmetry $S^1$ on $T^*\mathrm{SO}(3)$ whose Hamiltonian function is given by $h(\alpha_h) := (1/2)\|\alpha_h^\sharp\|^2 + Mg\ell\boldsymbol{k}\cdot h\chi$, where $h\in\mathrm{SO}(3), \alpha_h\in T_h^*\mathrm{SO}(3), \boldsymbol{k}$ is the unit vector of the spatial $Oz$ axis (pointing in the direction opposite to

that of the gravity force), $M \in \mathbb{R}$ is the total mass of the body, $g \in \mathbb{R}$ is the value of the gravitational acceleration, the fixed point about which the body moves is the origin, and $\chi$ is the unit vector of the straight line segment of length $\ell$ connecting the origin to the center of mass of the body. This Hamiltonian is left invariant under rotations about the spatial $Oz$ axis. A momentum map induced by this $S^1$-action is given by $\boldsymbol{J} : T^*\mathrm{SO}(3) \to \mathbb{R}, \boldsymbol{J}(\alpha_h) = T_e^* L_h(\alpha_h) \cdot \boldsymbol{k}$; recall that $T_e^* L_h(\alpha_h) =: \Pi \in \mathbb{R}^3$ is the body angular momentum. The reduced space $\boldsymbol{J}^{-1}(\mu)/S^1$ is generically the cotangent bundle of the unit sphere endowed with the symplectic structure given by the sum of the canonical form plus a magnetic term; equivalently, this is the coadjoint orbit in the dual of the Euclidean Lie algebra $\mathfrak{se}(3)^* = \mathbb{R}^3 \times \mathbb{R}^3$ given by $\mathcal{O}_\mu = \{(\Pi, \Gamma) \mid \Pi \cdot \Gamma = \mu, \|\Gamma\|^2 = 1\}$. The projection map $\boldsymbol{J}^{-1}(\mu) \to \mathcal{O}_\mu$ implementing the symplectic diffeomorphism between the reduced space and the coadjoint orbit in $\mathfrak{se}(3)^*$ is given by $\alpha_h \mapsto (\Pi, \Gamma) := (T_e^* L_h(\alpha_h), h^{-1}\boldsymbol{k})$. The orbit symplectic form $\omega_\mu$ on $\mathcal{O}_\mu$ has the expression $\omega_\mu(\Pi, \Gamma)((\Pi \times \boldsymbol{x} + \Gamma \times \boldsymbol{y}, \Gamma \times \boldsymbol{x}), (\Pi \times \boldsymbol{x}' + \Gamma \times \boldsymbol{y}', \Gamma \times \boldsymbol{x}')) = -\Pi \cdot (\boldsymbol{x} \times \boldsymbol{x}') - \Gamma \cdot (\boldsymbol{x} \times \boldsymbol{y}' - \boldsymbol{x}' \times \boldsymbol{y})$ for any $\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{y}, \boldsymbol{y}' \in \mathbb{R}^3$. The heavy-top equations $\dot{\Pi} = \Pi \times \Omega + Mg\ell\Gamma \times \chi, \dot{\Gamma} = \Gamma \times \Omega$ are Lie–Poisson equations on $\mathfrak{se}(3)^*$ for the Hamiltonian $h(\Pi, \Gamma) = (1/2)\Pi \cdot \Omega + Mg\ell\Gamma \cdot \chi$ and the Lie–Poisson bracket $\{f, g\}(\Pi, \Gamma) = -\Pi \cdot (\nabla_\Pi f \times \nabla_\Pi g) - \Gamma \cdot (\nabla_\Pi f \times \nabla_\Gamma g - \nabla_\Pi g \times \nabla_\Gamma f)$, where $\nabla_\Pi$ and $\nabla_\Gamma$ denote the partial gradients.

Let $(\Pi(t), \Gamma(t))$ be a periodic orbit of period $T$ of the heavy-top equations. After time $T$, by how much has the heavy top rotated in space? The answer is provided by [3]:

$$\begin{aligned}\varphi &= \frac{1}{\mu}\iint_{\mathcal{D}} \omega_\mu + \frac{1}{\mu}\left(2h_\mu T - 2Mg\ell \int_0^T \Gamma(s) \cdot \chi \mathrm{d}s\right) \\ &= \iint_D \frac{2\|\mathbb{I}\Gamma(s)\|^2 - (\Gamma(s) \cdot \mathbb{I}\Gamma(s))(\operatorname{tr} \mathbb{I})}{(\Gamma(s) \cdot \mathbb{I}\Gamma(s))^2} \mathrm{d}s \\ &\quad + \int_0^T \frac{\mathrm{d}s}{\Gamma(s) \cdot \mathbb{I}\Gamma(s)}\end{aligned}$$

where $D$ is the spherical cap on the unit sphere whose boundary is the closed curve $\Gamma(t)$ and $\mathcal{D}$ is a two-dimensional submanifold of the orbit $\mathcal{O}_\mu$ bounded by the closed integral curve $(\Pi(t), \Gamma(t))$. The first terms in each summand represent the geometric phase and the second terms the dynamic phase.

### Gauged Poisson Structures

If the Lie group $G$ acts freely and properly on a smooth manifold $Q$, then $(T^*Q)/G$ is a quotient Poisson manifold (*see* Poisson Reduction), where the quotient is taken relative to the (left) lifted cotangent action. The leaves of this Poisson manifold are the orbit reduced spaces $\boldsymbol{J}^{-1}(\mathcal{O}_\mu)/G$, where $\mathcal{O}_\mu \subset \mathfrak{g}^*$ is the coadjoint $G$-orbit through $\mu \in \mathfrak{g}^*$ (*see* Symmetry and Symplectic Reduction). Is there an explicit formula for this reduced Poisson bracket on a manifold diffeomorphic to $(T^*Q)/G$? It turns out that this question has two possible answers, once a connection on the principal bundle $\pi : Q \to Q/G$ is introduced. The discussion below will also link to the fibration version of cotangent bundle reduction.

In order to present these answers, we review two bundle constructions. Let $G$ act freely and properly on the manifold $P$ and consider the a (left) principal $G$-bundle $\rho : P \to P/G := M$. Let $\tau : N \to M$ be a surjective submersion. Then the pullback bundle $\tilde{\rho} : (n, p) \in \tilde{P} := \{(n, p) \in N \times P \mid \rho(p) = \tau(n)\} \mapsto n \in N$ over $N$ is also a principal (left) $G$-bundle relative to the action $g \cdot (n, p) := (n, g \cdot p)$.

If there is a (left) $G$-action a manifold $V$, then the diagonal $G$-action $g \cdot (p, v) = (g \cdot p, g \cdot v)$ on $P \times V$ is also free and proper and one can form the associated bundle $P \times_G V := (P \times V)/G$ which is a locally trivial fiber bundle $\rho_E : [p, v] \in E := P \times_G V \mapsto \rho(p) \in M$ over $M$ with fibers diffeomorphic to $V$. Analogously, one can form the associated fiber bundle $\rho_{\tilde{E}} : \tilde{E} := \tilde{P} \times_G V \to N$. Summarizing, the associated bundle $\tilde{E} = \tilde{P} \times_G V \to N$ is obtained from the principal bundle $\rho : P \to M$, the surjective submersion $\tau : N \to M$, and the $G$-manifold $V$ by pullback and association, in this order.

These operations can be reversed. First, form the associated bundle $\rho_E : E = P \times_G V \to M$ and then pull it back by the surjective submersion $\tau : N \to M$ to $N$ to get the pullback bundle $\tilde{\rho}_E : \tilde{E} \to N$. The map $\Phi : \tilde{P} \times_G V \to \tilde{E}$ defined by $\Phi([(n, p), v]) := (n, [p, v])$ is an isomorphism of locally trivial fiber bundles.

These general considerations will be used now to realize the quotient Poisson manifold $(T^*Q)/G$ in two different ways. Let $Q$ be a manifold and $G$ a Lie group (with Lie algebra $\mathfrak{g}$) acting freely and properly on it. Let $A \in \Omega^1(Q; \mathfrak{g})$ be a connection 1-form on the left $G$-principal bundle $\pi : Q \to Q/G$. Pull back the $G$-bundle $\pi : Q \to Q/G$ by the cotangent bundle projection $\pi_{Q/G} : T^*(Q/G) \to Q/G$ to $T^*(Q/G)$ to obtain the $G$-principal bundle $\tilde{\pi}_{Q/G} : (\alpha_{[q]}, q) \in \tilde{Q} := \{(\alpha_{[q]}, q) \mid [q] = \pi(q), q \in Q\} \mapsto \alpha_{[q]} \in T^*(Q/G)$. This bundle is isomorphic to the annihilator $(VQ)^\circ \subset T^*Q$ of the vertical bundle $VQ := \ker T\pi \subset TQ$. Next, form the coadjoint bundle $\rho_S : S := \tilde{Q} \times_G \mathfrak{g}^* \to T^*(Q/G)$ of $\tilde{Q}, \rho_S((\alpha_{[q]}, q), \mu) = \alpha_{[q]}$, that is, the associated vector bundle to the $G$-principal bundle $\tilde{Q} \to T^*(Q/G)$ given by the coadjoint representation of $G$ on $\mathfrak{g}^*$. The connection-dependent map $\Phi_A : S \to (T^*Q)/G$ defined by $\Phi_A([(\alpha_{[q]}, q), \mu]) := [T_q^*\pi(\alpha_{[q]}) + A(q)^*\mu]$, where $q \in Q, \alpha_q \in T_q^*Q$, and

$\mu \in \mathfrak{g}^*$, is a vector bundle isomorphism over $Q/G$. The Sternberg space is the Poisson manifold $(S, \{\cdot, \cdot\}_S)$, where $\{\cdot, \cdot\}_S$ is the pullback to $S$ by $\Phi_A$ of the quotient Poisson bracket on $(T^*Q)/G$.

Next, we proceed in the opposite order. Construct first the coadjoint bundle $\rho_{\tilde{\mathfrak{g}}^*} : [q, \mu] \in \tilde{\mathfrak{g}}^* := Q \times_G \mathfrak{g}^* \mapsto [q] \in Q/G$ associated to the principal bundle $\pi : Q \to Q/G$ and then pull it back by the cotangent bundle projection $\pi_{Q/G} : T^*(Q/G) \to Q/G$ to $T^*(Q/G)$ to obtain the vector bundle $\rho_W : W := \{(\alpha_{[q]}, [q, \mu]) \mid \pi_{Q/G}(\alpha_{[q]}) = \rho_{\tilde{\mathfrak{g}}^*}([q, \mu]) = [q]\}$, $\rho_W(\alpha_{[q]}, [q, \mu]) = \alpha_{[q]}$ over $T^*(Q/G)$. Note that $W = T^*(Q/G) \oplus \tilde{\mathfrak{g}}^*$ and hence $W$ is also a vector bundle over $Q/G$. Let $HQ$ be the horizontal sub-bundle defined by the connection $A$; thus, $TQ = HQ \oplus VQ$, where $H_qQ := \ker A(q)$. For each $q \in Q$, the linear map $T_q\pi|_{H_qQ} : H_qQ \to T_{[q]}(Q/G)$ is an isomorphism. Let $\mathrm{hor}_q := (T_q\pi|_{H_qQ})^{-1} : T_{[q]}(Q/G) \to H_qQ \subset T_qQ$ be the horizontal lift operator induced by the connection $A$. Thus, $\mathrm{hor}_q^* : T_q^*Q \to T_{[q]}^*(Q/G)$ is a linear surjective map whose kernel is the annihilator $(H_qQ)^\circ$ of the horizontal space. The connection-dependent map $\Psi_A : (T^*Q)/G \to W$ defined by $\Psi_A([\alpha_q]) := (\mathrm{hor}_q^*(\alpha_q), [q, \boldsymbol{J}(\alpha_q)])$, where $q \in Q$, $\alpha_q \in T_q^*Q$, and $\boldsymbol{J} : T^*Q \to \mathfrak{g}^*$ is the momentum map of the lifted action, $\langle \boldsymbol{J}(\alpha_q), \xi \rangle = \alpha_q((\xi_Q(q))$ for $\xi \in \mathfrak{g}$, is a vector bundle isomorphism over $Q/G$ and $\Psi_A \circ \Phi_A = \Phi$. The Weinstein space is the Poisson manifold $(W, \{\cdot, \cdot\}_W)$, where $\{\cdot, \cdot\}_W$ is the push-forward by $\Psi_A$ of the Poisson bracket of $(T^*Q)/G$. In particular, $\Phi : S \to W$ is a connection independent Poisson diffeomorphism. The Poisson brackets on $S$ and on $W$ are called gauged Poisson brackets. They are expressed explicitly in terms of various covariant derivatives induced on $S$ and on $W$ by the connection $A \in \Omega^1(Q; \mathfrak{g})$.

Recall that the connection $A$ on the principal bundle $\pi : Q \to Q/G$ naturally induces connections on pullback bundles and affine connections on associated vector bundles. Thus, both $S$ and $W$ carry covariant derivatives induced by $A$. They are given, according to general definitions, in the cases under consideration, by:

- If $f \in C^\infty(S)$, $s = [(\alpha_{[q]}, q), \mu] \in S$, and $v_{\alpha_{[q]}} \in T_{\alpha_{[q]}}T^*(Q/G)$, then $\boldsymbol{d}_{\tilde{A}}^S f(s) \in T_{\alpha_{[q]}}^* T^*(Q/G)$ is defined by $\boldsymbol{d}_{\tilde{A}}^S f(s)(v_{\alpha_{[q]}}) := \boldsymbol{d}f(s)(T_{((\alpha_{[q]}, q), \mu)}\pi_{\tilde{Q}\times\mathfrak{g}^*}((v_{\alpha_{[q]}}, \mathrm{hor}_q(T_{\alpha_{[q]}}\tau(v_{\alpha_{[q]}}))), 0))$ where $\pi_{\tilde{Q}\times\mathfrak{g}^*} : \tilde{Q} \times \mathfrak{g}^* \to \tilde{Q} \times_G \mathfrak{g}^* = S$ is the orbit map. The symbol $\boldsymbol{d}_{\tilde{A}}^S$ signifies that this is a covariant derivative on the associated bundle $S$ induced by the connection $\tilde{A}$ on the principal $G$-pullback bundle $\tilde{Q} \to T^*(Q/G)$. This connection $\tilde{A}$ is the pullback connection defined by $A$.
- If $f \in C^\infty(W)$, $w = (\alpha_{[q]}, [q, \mu]) \in W$, and $v_{\alpha_{[q]}} \in T_{\alpha_{[q]}}T^*(Q/G)$, then $\tilde{\nabla}_A^W f(w) \in T_{\alpha_{[q]}}^* T^*(Q/G)$ is defined by $\tilde{\nabla}_A^W f(w)(v_{\alpha_{[q]}}) = \boldsymbol{d}f(w)(v_{\alpha_{[q]}}, T_{(q,\mu)}\pi_{Q\times\mathfrak{g}^*}(\mathrm{hor}_q(T_{\alpha_{[q]}}\tau_{Q/G}(v_{\alpha_{[q]}})), 0))$ where $\pi_{Q\times\mathfrak{g}^*} : Q \times \mathfrak{g}^* \to Q \times_G \mathfrak{g}^* = \tilde{\mathfrak{g}}^*$ is the orbit map. The symbol $\tilde{\nabla}_A^W$ signifies that this is a covariant derivative on the pullback bundle $W$ induced by the covariant derivative $\nabla_A$ on the coadjoint bundle $\tilde{\mathfrak{g}}^*$. This covariant derivative $\nabla_A$ is induced on $\tilde{\mathfrak{g}}^*$ by the connection $A$.
- For $f \in C^\infty(W)$, we have $\boldsymbol{d}_{\tilde{A}}^S(f \circ \Phi) = (\tilde{\nabla}_A^W f) \circ \Phi$.

To write the two gauged Poisson brackets on $S$ and on $W$ explicitly, we denote by $\tilde{\mathfrak{g}} = Q \times_G \mathfrak{g}$ the adjoint bundle of $\pi : Q \to Q/G$, by $\Omega_{Q/G}$ the canonical symplectic structure on $T^*(Q/G)$, by $B \in \Omega^2(Q; \mathfrak{g})$ the curvature of $A$, and by $\mathcal{B}$ the $\tilde{\mathfrak{g}}$-valued 2-form $\mathcal{B} \in \Omega^2(Q/G; \tilde{\mathfrak{g}})$ on the base $Q/G$ defined by $\mathcal{B}([q])(u_{[q]}, v_{[q]}) = [q, B(q)(u_q, v_q)]$, for any $u_q, v_q \in T_qQ$ that satisfy $T_q\pi(u_q) = u_{[q]}$ and $T_q\pi(v_q) = v_{[q]}$. Note that both $S^*$ and $W^*$ are Lie algebra bundles, that is, their fibers are Lie algebras and the fiberwise Lie bracket operation depends smoothly on the base point. If $f \in C^\infty(S)$, denote by $\delta f/\delta s \in S^* = \tilde{Q} \times_G \mathfrak{g}$ the usual fiber derivative of $f$. Similarly, if $f \in C^\infty(W)$ denote by $\delta f/\delta w \in W^*$ the usual fiber derivative of $f$. Finally, $\sharp : T^*(T^*(Q/G)) \to T(T^*(Q/G))$ is the vector bundle isomorphism induced by $\Omega_{Q/G}$. The Poisson bracket of $f, g \in C^\infty(S)$ is given by

$$\begin{aligned}\{f, g\}_S(s) &= \Omega_{Q/G}(\alpha_{[q]})\left(\boldsymbol{d}_{\tilde{A}}^S f(s)^\sharp, \boldsymbol{d}_{\tilde{A}}^S g(s)^\sharp\right) \\ &\quad - \left\langle s, \left[\frac{\delta f}{\delta s}, \frac{\delta g}{\delta s}\right]\right\rangle \\ &\quad + \left\langle v, (\pi_{Q/G}^*\mathcal{B})(\alpha_{[q]})\left(\boldsymbol{d}_{\tilde{A}}^S f(s)^\sharp, \boldsymbol{d}_{\tilde{A}}^S g(s)^\sharp\right)\right\rangle\end{aligned}$$

where $v = [q, \mu] \in \tilde{\mathfrak{g}}^*$. The Poisson bracket $f, g \in C^\infty(W)$ is given by

$$\begin{aligned}\{f, g\}_W(w) &= \Omega_{Q/G}(\alpha_{[q]})\left(\tilde{\nabla}_A^W f(w)^\sharp, \tilde{\nabla}_A^W g(w)^\sharp\right) \\ &\quad - \left\langle w, \left[\frac{\delta f}{\delta w}, \frac{\delta g}{\delta w}\right]\right\rangle \\ &\quad + \left\langle v, (\pi_{Q/G}^*\mathcal{B})(\alpha_{[q]})\left(\tilde{\nabla}_A^W f(w)^\sharp, \tilde{\nabla}_A^W g(w)^\sharp\right)\right\rangle\end{aligned}$$

Note that their structure is of the form: "canonical" bracket plus a (left) "Lie–Poisson" bracket plus a curvature coupling term.

### The Symplectic Leaves of the Sternberg and Weinstein Spaces

The map $\varphi_A : \tilde{Q} \times \mathfrak{g}^* \to T^*Q$ given by $\varphi_A((\alpha_{[q]}, q), \mu) := T_q^*\pi(\alpha_{[q]}) + A(q)^*\mu$, where $((\alpha_{[q]}, q), \mu) \in \tilde{Q} \times \mathfrak{g}^*$, is a $G$-equivariant diffeomorphism; the $G$-action on $T^*Q$ is by cotangent lift and on $\tilde{Q} \times \mathfrak{g}^*$ is $g \cdot ((\alpha_{[q]}, q), \mu) = ((\alpha_{[q]}, g \cdot q), \mathrm{Ad}_{g^{-1}}^*\mu)$. The pullback $\boldsymbol{J}_A$

of the momentum map to $\tilde{Q}\times\mathfrak{g}^*$ has the expression $J_A((\alpha_{[q]},q),\mu)=\mu$, so if $\mathcal{O}\subset\mathfrak{g}^*$ is a coadjoint orbit we have $J_A^{-1}(\mathcal{O})=\tilde{Q}\times\mathcal{O}$, and hence the orbit reduced manifold $J_A^{-1}(\mathcal{O})/G$, whose connected components are the symplectic leaves of $S$, equals $\tilde{Q}\times_G\mathcal{O}$. Its symplectic form is the Sternberg minimal coupling form $\tilde{\omega}_{\mathcal{O}}^{-}+\rho_S^*\Omega_{Q/G}$.

In this formula, the 2-form $\tilde{\omega}_{\mathcal{O}}^{-}$ has not been defined yet. It is uniquely defined by the identity $\pi_{\tilde{Q}\times\mathfrak{g}^*}\tilde{\omega}_{\mathcal{O}}^{-}=\boldsymbol{d}\widehat{A}+\Pi_{\mathcal{O}}\omega_{\mathcal{O}}^{-}$, where $\omega_{\mathcal{O}}^{-}$ is the minus orbit symplectic form on $\mathcal{O}$ (*see* Symmetry and Symplectic Reduction), $\Pi_{\mathcal{O}}:\tilde{Q}\times\mathcal{O}\to\mathcal{O}$ is the projection on the second factor, and $\widehat{A}\in\Omega^2(\tilde{Q}\times\mathcal{O})$ is the 2-form given by $\widehat{A}((\alpha_{[q]},q),\mu)\quad((u_{\alpha_{[q]}},v_q),\nu)=-\langle\mu,A(q)(v_q)\rangle$ for $((\alpha_{[q]},q),\mu)\in\tilde{Q}\times\mathcal{O},(u_{\alpha_{[q]}},v_q)\in T_{(\alpha_{[q]},q)}\tilde{Q}$, and $\nu\in\mathfrak{g}^*$.

The symplectic leaves of the Weinstein space $W$ are obtained by pushing forward by $\Phi$ the symplectic leaves of the Sternberg space. They are the connected components of the symplectic manifolds $(T^*(Q/G)\oplus(Q\times_G\mathcal{O}),\Pi^*_{T^*(Q/G)}\Omega_{Q/G}+\Pi^*_{Q\times_G\mathcal{O}}\omega^{-}_{Q\times_G\mathcal{O}})$, where $\mathcal{O}$ is a coadjoint orbit in $\mathfrak{g}^*$, $\Omega_{Q/G}$ is the canonical symplectic form on $T^*(Q/G)$, $\omega^{-}_{Q\times_G\mathcal{O}}$ is a closed 2-form on $Q\times_G\mathcal{O}$ to be defined below, and $\Pi_{T^*(Q/G)}:T^*(Q/G)\oplus(Q\times_G\mathcal{O})\to T^*(Q/G),\Pi_{Q\times_G\mathcal{O}}:T^*(Q/G)\oplus(Q\times_G\mathcal{O})\to Q\times_G\mathcal{O}$ are the projections. The closed 2-form $\omega^{-}_{Q\times_G\mathcal{O}}\in\Omega^2(Q\times_G\mathcal{O})$ is uniquely determined by the identity $\pi^*_{Q\times\mathcal{O}}\omega^{-}_{Q\times_G\mathcal{O}}=\omega^{-}_{Q\times\mathcal{O}}$, where $\pi_{Q\times\mathcal{O}}:Q\times\mathcal{O}\to Q\times_G\mathcal{O}$ is the orbit space projection, $\omega^{-}_{Q\times\mathcal{O}}\in\Omega^2(Q\times\mathcal{O})$ is closed and given by $\omega^{-}_{Q\times\mathcal{O}}(q,\mu)((u_q,-\mathrm{ad}^*_\xi\mu),(v_q,-\mathrm{ad}^*_\eta\mu)):=-\boldsymbol{d}(\mathcal{A}\times\mathrm{id}_{\mathcal{O}})(q,\mu)((u_q,-\mathrm{ad}^*_\xi\mu),(v_q,-\mathrm{ad}^*_\eta\mu))+\omega^{-}_{\mathcal{O}}(\mu)(\mathrm{ad}^*_\xi\mu,\mathrm{ad}^*_\eta\mu)$, and $\mathcal{A}\times\mathrm{id}_{\mathcal{O}}\in\Omega^1(Q\times\mathfrak{g}^*)$ is given by $(\mathcal{A}\times\mathrm{id}_{\mathcal{O}})(q,\mu)(u_q,-\mathrm{ad}^*_\xi\mu)=\langle\mu,\mathcal{A}(q)(u_q)\rangle$, for $q\in Q,\mu\in\mathfrak{g}^*,u_q,v_q\in T_qQ,\xi,\eta\in\mathfrak{g}$.

Thus, on the Sternberg and Weinstein spaces, both the Poisson bracket as well as the symplectic form on the leaves have explicit connection dependent formulas (*see* Gauge Theory: Mathematical Applications for a general treatment of gauge theories).

*See also:* Gauge Theory: Mathematical Applications; Hamiltonian Group Actions; Poisson Reduction; Symmetries and Conservation Laws; Symmetry and Symplectic Reduction.

## Further Reading


Abraham R and Marsden JE (1978) *Foundations of Mechanics*, 2nd edn. Reading, MA: Addison-Wesley.

Guichardet A (1984) On rotation and vibration motions of molecules. *Annales de l'Institut Henri Poincaré. Physique Théorique* 40: 329–342.

Iwai T (1987) A geometric setting for classical molecular dynamics. *Annales de l'Institut Henri Poincaré. Physique Théorique*. 47: 199–219.

Kummer M (1981) On the construction of the reduced phase space of a Hamiltonian system with symmetry. *Indiana University Mathematics Journal* 30: 281–291.

Lewis D, Marsden JE, Montgomery R, and Ratiu TS (1986) The Hamiltonian structure for dynamic free boundary problems. *Physica D* 18: 391–404.

Marsden JE, Montgomery R, and Ratiu TS (1990) Reduction, symmetry, and phases in mechanics. *Memoirs of the American Mathematical Society* 88(436).

Marsden JE, Misiołek G, Ortega J-P, Perlmutter M, and Ratiu TS (2005) *Hamiltonian Reduction by Stages*, Lecture Notes in Mathematics. Springer.

Marsden JE and Perlmutter M (2000) The orbit bundle picture of cotangent bundle reduction. *Comptes Rendus Mathématiques de l'Académie des Sciences. La Société Royale du Canada* 22: 33–54.

Marsden JE and Ratiu TS (2003) *Introduction to Mechanics and Symmetry,* 2nd edn. second printing; 1st edn. (1994), Texts in Applied Mathematics, vol. 17. New York: Springer.

Montgomery R (1984) Canonical formulations of a particle in a Yang–Mills field. *Letters in Mathematical Physics* 8: 59–67.

Montgomery R (1991) How much does a rigid body rotate? A Berry's phase from the eighteenth century. *American Journal of Physics* 59: 394–398.

Montgomery R, Marsden JE, and Ratiu TS (1984) Gauged Lie Poisson structures. In: Marsden J (ed.) *Fluids and Plasmas: Geometry and Dynamics*, Contemporary Mathematiques, vol. 28, pp. 101–114. Providence, RI: American Mathematical Society.

Satzer WJ (1977) Canonical reduction of mechanical systems invariant under abelian group actions with an application to celestial mechanics. *Indiana, University Mathematics Journal* 26: 951–976.

Simo JC, Lewis D, and Marsden JE (1991) The stability of relative equilibria. Part I: The reduced energy–momentum method. *Archive for Rational Mechanics and Analysis* 115: 15–59.

Smale S (1970) Topology and mechanics. *Inventiones Mathematicae* 10: 305–331, 11: 45–64.

Sternberg S (1977) Minimal coupling and the symplectic mechanics of a classical particle in the presence of a Yang–Mills field. *Proceedings of the National Academy of Sciences* 74: 5253–5254.

Weinstein A (1978) A universal phase space for particles in Yang–Mills fields. *Letters in Mathematical Physics* 2: 417–420.

Zaalani N (1999) Phase space reduction and Poisson structure. *Journal of Mathematical Physics* 40(7): 3431–3438.

# Critical Phenomena in Gravitational Collapse

**C Gundlach**, University of Southampton, Southampton, UK



## Introduction

Sufficiently dense concentrations of mass–energy in general relativity collapse irreversibly and form black holes. More precisely, the singularity theorems state that once a closed trapped surface has developed, some world lines will only extend to a finite length in the future – they end in a spacetime singularity. Furthermore, the cosmic censorship hypothesis states that this singularity is hidden away inside a black hole. One can, therefore, classify initial data in general relativity which describe an isolated system with no black hole present into those which remain regular, and those which form a black hole during their evolution.

Theorems on the stability of Minkowski spacetime, and similar results for some types of matter coupled to gravity, imply that sufficiently weak (in some technical sense) initial data will remain regular. On the other hand, no necessary or sufficient criterion for black hole formation is known. For very strong data the existence of a closed trapped surface implies black hole formation, but although the data themselves may be regular, the trapped surface must already be inside the black hole. Between the very weak and very strong regime, there is a middle regime of initial data for which one cannot decide if they will or will not form a black hole, other than evolving them in time.

The threshold between collapse and dispersion was first explored systematically by Choptuik (1992). He concentrated on the simple model of a spherically symmetric massless scalar (matter) field $\phi(r,t)$. In this model, the scalar-field matter must either form a black hole, or disperse to infinity – it cannot form stable stars. Choptuik explored the space of initial data by means of one-parameter families of initial data which interpolate between strong data (say with large parameter $p$) that form a black hole and weak data (with small $p$) that disperse. The critical value $p_*$ of the parameter $p$ can be found for each family by evolving many data sets from that family. Near the black hole threshold, Choptuik found the following phenomena:

1. *Mass scaling*. By fine-tuning the initial data to the threshold along any one-parameter family, one can make arbitrarily small black holes. Near the threshold, the black hole mass scales as
$$M \simeq C(p - p_*)^\gamma \quad \text{for } p \geq p_* \qquad [1]$$
for the black hole mass $M$ in the limit $p \to p_*$ from above.
2. *Universality*. While $p_*$ and $C$ depend on the particular one-parameter family of data, the critical exponent $\gamma$ has a universal value, $\gamma \simeq 0.374$, for all one-parameter families of scalar-field data. Furthermore, for a finite time in a finite region of space, the solutions generated by all near-critical data approach one and the same solution $\phi_*$, called the critical solution:
$$\phi(r, t) \simeq \phi_*\left(\frac{r}{L}, \frac{t - t_*}{L}\right) \qquad [2]$$
The constants $t_*$ and $L$ depend again on the family of initial data, but $\phi_*(r,t)$ is universal. This universal phase ends when the evolution decides between black hole formation and dispersion. The universal critical solution is approached by any initial data that are sufficiently close to the black hole threshold, on either side, and from any one-parameter family.
3. *Scale-echoing*. The critical solution $\phi_*(r,t)$ is unchanged when one rescales space and time by a factor $e^\Delta$:
$$\phi_*(r, t) = \phi_*\left(e^\Delta r, e^\Delta t\right) \qquad [3]$$
where $\Delta \simeq 3.44$ for the scalar field.

The same phenomena were quickly discovered in many other types of matter coupled to gravity, and even in vacuum gravity (where gravitational waves can form black holes). The echoing period $\Delta$ and critical exponent $\gamma$ depend on the type of matter, but the existence of the phenomena appears to be generic. For some types of matter (e.g., perfect fluid matter), the critical solution is continuously scale invariant (or continuously self-similar, CSS) in the sense that
$$\phi_*(r, t) = \phi_*(r/t) \qquad [4]$$
rather than scale-periodic (or discretely self-similar, DSS) as in [3]. (We use the notation $\phi_*(x)$ for the function of one variable $r/t$.) We have described scale invariance and scale-echoing here in terms of coordinates, but these do admit geometric, coordinate-invariant definitions, which are not restricted to spherical symmetry.

There is also another kind of critical behavior at the black hole threshold. Here, too, the evolution goes through a universal critical solution, but it is static, rather than scale invariant. As a consequence, the mass of black holes near the threshold takes a universal finite value (some fixed fraction of the mass of the critical solution), instead of showing power-law

scaling. In an analogy with first- and second-order phase transitions in statistical mechanics, the critical phenomena with a finite mass at the black hole threshold are called type I, and the critical phenomena with power-law scaling of the mass are called type II.

At this point, we characterize the degree of rigor of the various parts of the theory that is summarized in this article. Critical phenomena were discovered in the numerical time evolution of generic asymptotically flat initial data. Numerical evolution of many elements of a specific one-parameter family, and fine-tuning to the black hole threshold along that family showed self-similarity and mass scaling near the threshold. Doing this for a number of randomly chosen one-parameter families suggests that these phenomena, and in particular the echoing scale $\Delta$ and mass-scaling exponent $\gamma$, are universal between initial data within one model (e.g., the spherical scalar field). Numerical experiments, however, can only explore a finite-dimensional subspace of the infinite-dimensional space of initial data (phase space) of the field theory, and so cannot prove universality.

We go further by applying the theory of dynamical systems to general relativity. The arguments summarized in the next section would be difficult to make rigorous, as the dynamical system under consideration is infinite dimensional, but they suggest a focus on fixed points of the dynamical system and their linear perturbations. Even though the dynamical systems motivation is not mathematically rigorous, the linearized analysis itself is a well-defined problem that can be solved numerically to essentially arbitrary precision. This proves universality on a perturbative level, and provides numerical values of $\Delta$ and $\gamma$. A combination of the global dynamical systems analysis and perturbative analysis even predicts further critical exponents for black hole charge and angular momentum. Finally, critical phenomena have been discovered in a number of systems (different types of matter and symmetry restrictions), and this suggests that they may be generic for some large class of field theories (although details such as the numerical values of $\gamma$ and $\Delta$ do depend on the system), but there is no conclusive evidence for this at present.

## The Dynamical Systems Picture

When we consider general relativity as an infinite-dimensional dynamical system, a solution curve is a spacetime. Points along the curve are Cauchy surfaces in the spacetime, which can be thought of as moments of time. An important difference between general relativity and other field theories is that the same spacetime can be sliced in many different ways, none of which is preferred. Therefore, to turn general relativity into a dynamical system, one has to fix a slicing (and in practice also coordinates on each slice). In the example of the spherically symmetric massless scalar field, using polar slicing and an area radial coordinate $r$, a point in phase space can be characterized by the two functions

$$Z = \left\{ \phi(r),\ r\frac{\partial \phi}{\partial t}(r) \right\} \qquad [5]$$

In spherical symmetry, there are no degrees of freedom in the scalar field, and Cauchy data for the metric can be reconstructed from $Z$ using the Einstein constraints.

The phase space consists of two halves: initial data whose time evolution always remains regular, and data which contain a black hole or form one during time evolution. The numerical evidence collected from individual one-parameter families of data suggests that the black hole threshold that separates the two is a smooth hypersurface. The mass-scaling law [1] can, therefore, be restated without explicit reference to one-parameter families. Let $P$ be any function on phase space such that data sets with $P > 0$ form black holes, and data with $P < 0$ do not, and which is analytic in a neighborhood of the black hole threshold $P = 0$. The black hole mass as a function on phase space is then given by

$$M \simeq F(P)\ P^{\gamma} \qquad [6]$$

for $P > 0$, where $F(P) > 0$ is an analytic function.

Consider now the time evolution in this dynamical system, near the threshold ("critical surface") between black hole formation and dispersion. A phase-space trajectory that starts out in a critical surface by definition never leaves it. A critical surface is, therefore, a dynamical system in its own right, with one dimension fewer. If it has an attracting fixed point, such a point is called a critical point. It is an attractor of codimension 1, and the critical surface is its basin of attraction. The fact that the critical solution is an attractor of codimension 1 is visible in its linear perturbations: it has an infinite number of decaying perturbation modes tangential to (and spanning) the critical surface, and a single growing mode not tangential to the critical surface.

Any trajectory beginning near the critical surface, but not necessarily near the critical point, moves almost parallel to the critical surface toward the critical point. As the phase point approaches the critical point, its movement parallel to the surface

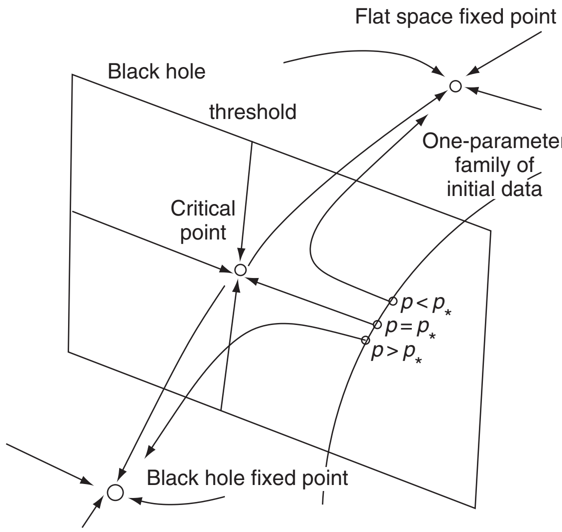



**Figure 1** The phase-space picture for the black hole threshold in the presence of a critical point. The arrow lines are time evolutions, corresponding to spacetimes. The line without an arrow is not a time evolution, but a one-parameter family of initial data that crosses the black hole threshold at $p = p_*$. (Reproduced with permission from Gundlach C (2003) Critical phenomena in gravitational collapse. *Physics Reports* 376: 339–405.)

slows down, while its distance and velocity out of the critical surface are still small. The phase point spends sometime moving slowly near the critical point. Eventually, it moves away from the critical point in the direction of the growing mode, and ends up on an attracting fixed point.

This is the origin of universality: any initial data set that is close to the black hole threshold (on either side) evolves to a spacetime that approximates the critical spacetime for sometime. When it finally approaches either the dispersion fixed point or the black hole fixed point, it does so on a trajectory that appears to be coming from the critical point itself. All near-critical solutions are passing through one of these two funnels. All details of the initial data have been forgotten, except for the distance from the black hole threshold: the closer the initial phase point is to the critical surface, the more the solution curve approaches the critical point, and the longer it will remain close to it.

In all systems that have been examined, the black hole threshold contains at least one critical point. A fixed point of the dynamical system represents a spacetime with an additional continuous symmetry that generic solutions do not have. If the critical spacetime is time independent in the usual sense, we have type I critical phenomena; if the symmetry is scale invariance, we have type II critical phenomena. The attractor within the critical surface may also be a limit cycle, rather than a fixed point. In spacetime terms this corresponds to a discrete symmetry (DSS rather than CSS in type II, or a pulsating critical solution, rather than a stationary one, in type I).

## Self-Similarity and Mass Scaling

Type II critical phenomena occur where the critical solution is scale invariant (self-similar, CSS or DSS). Using suitable spacetime coordinates, a CSS solution can be characterized as independent of a time coordinate $\tau$ which is also a logarithmic scale. Similarly, a DSS solution can be characterized as periodic in $\tau$. For example, starting from the scale periodicity [3] in polar-radial coordinates, we replace $r$ and $t$ by new coordinates

$$x \equiv -\frac{r}{t - t_*}, \qquad \tau \equiv -\ln\left(-\frac{t - t_*}{L}\right) \qquad [7]$$

where the accumulation time $t_*$ and scale $L$ must be matched to the one-parameter family under consideration. $\tau$ has been defined so that it increases as $t$ increases and approaches $t_*$ from below. It is useful to think of $r$, $t$, and $L$ as having dimension length in units $c = G = 1$, and of $x$ and $\tau$ as dimensionless. Choptuik's observation, expressed in these coordinates, is that in any near-critical solution there is a spacetime region where the fields $Z$ are well approximated by the critical solution, or

$$Z(x, \tau) \simeq Z_*(x, \tau) \qquad [8]$$

with

$$Z_*(x, \tau + \Delta) = Z_*(x, \tau) \qquad [9]$$

Note that the time parameter of the dynamical system must be chosen as $\tau$ if a CSS solution is to be a fixed point, or a DSS solution a cycle. More generally (going beyond spherical symmetry), on any self-similar spacetime one can introduce coordinates $x^\mu = (\tau, x^1, x^2, x^3)$ in which the metric is of the form

$$g_{\mu\nu} = \mathrm{e}^{-2\tau}\bar{g}_{\mu\nu} \qquad [10]$$

and where $\bar{g}_{\mu\nu}$ is independent of $\tau$ for a CSS spacetime, and periodic in $\tau$ for a DSS spacetime. These coordinates are not unique.

The critical exponent $\gamma$ can be calculated from the linear perturbations of the critical solution. In order to keep the notation simple, the discussion will be restricted to a critical solution that is spherically symmetric and CSS, which is correct, for example, for perfect-fluid matter.

Let us assume that we have fine-tuned initial data close to the black hole threshold so that in a region the resulting spacetime is well approximated by the CSS critical solution. This part of the spacetime

corresponds to the section of the phase-space trajectory that lingers near the critical point. In this region, we can linearize around $Z_*$. As $Z_*$ does not depend on $\tau$, its linear perturbations can depend on $\tau$ only exponentially. Labeling the perturbation modes by $i$, a single mode perturbation is of the form

$$\delta Z = C_i e^{\lambda_i \tau} Z_i(x) \qquad [11]$$

In the near-critical regime, we can therefore approximate the solution as

$$Z(x,\tau) \simeq Z_*(x) + \sum_{i=0}^{\infty} C_i(p)\, e^{\lambda_i \tau} Z_i(x) \qquad [12]$$

The notation $C_i(p)$ is used because the perturbation amplitudes $C_i$ depend on the initial data, and hence on the parameter $p$ that controls the initial data.

If $Z_*$ is a critical solution, by definition there is exactly one $\lambda_i$ with positive real part (in fact, it is purely real), say $\lambda_0$. As $t \to t_*$ from below, which corresponds to $\tau \to \infty$, all other perturbations decay and can be neglected. By definition, the critical solution corresponds to $p = p_*$, and so we must have $C_0(p_*) = 0$. Linearizing around $p_*$, we obtain

$$Z(x,\tau) \simeq Z_*(x) + \left.\frac{dC_0}{dp}\right|_{p_*} (p - p_*)\, e^{\lambda_0 \tau}\, Z_0(x) \qquad [13]$$

in a region of the spacetime.

Now we extract Cauchy data at one particular value of $\tau$ within that region, namely at $\tau_p$ defined by

$$\left.\frac{dC_0}{dp}\right|_{p_*} |p - p_*| e^{-\lambda_0 \tau_p} \equiv \epsilon \qquad [14]$$

where $\epsilon$ is an arbitrary small constant, so that

$$Z(x,\tau_p) \simeq Z_*(x) \pm \epsilon\, Z_0(x) \qquad [15]$$

where $\pm$ is the sign of $p - p_*$, left behind because by definition $\epsilon$ is positive. As $\tau$ increases from $\tau_p$, the growing perturbation becomes nonlinear and the approximation [13] breaks down. Then either a black hole forms (say for the positive sign), or the solution disperses (for the negative sign). We need not follow this nonlinear evolution in detail to find the black hole mass scaling in the former case: dimensional analysis is sufficient. Going back to coordinates $t$ and $r$, we have

$$Z(r, t_p) \simeq Z_*\left(\frac{r}{L_p}\right) \pm \epsilon\, Z_0\left(\frac{r}{L_p}\right) \qquad [16]$$

where

$$L_p \equiv L e^{-\tau_p} \qquad [17]$$

These Cauchy data at $t = t_p$ depend on the initial data at $t = 0$ only through the overall scale $L_p$, and through the sign in front of $\epsilon$. If the field equations themselves are scale invariant, or asymptotically scale invariant at scales $L_p$ and smaller, the black hole mass, which has dimensions of length in gravitational units, must be proportional to the initial data scale $L_p$, the only length scale that is present. Therefore,

$$M \propto L_p \propto (p - p_*)^{1/\lambda_0} \qquad [18]$$

and we have found the critical exponent to be $\gamma = 1/\lambda_0$.

## The Analogy with Statistical Mechanics

The existence of a threshold where a qualitative change takes place, universality, scale invariance, and critical exponents suggest that there is a mathematical analogy between type II critical phenomena and critical phase transitions in statistical mechanics.

In equilibrium statistical mechanics, observable macroscopic quantities, such as the magnetization of a ferromagnetic material, are derived as statistical averages over microstates of the system. The expected value of an observable is

$$\langle A \rangle = \sum_{\text{microstates}} A(\text{microstate})\, e^{-H(\text{microstate},\mu)} \qquad [19]$$

The Hamiltonian $H$ depends on the parameters $\mu$, which comprise the temperature, parameters characterizing the system such as interaction energies of the constituent molecules, and macroscopic forces such as the external magnetic field. The objective of statistical mechanics is to derive relations between the macroscopic quantities $A$ and parameters $\mu$.

Phase transitions in thermodynamics are thresholds in the space of external forces $\mu$ at which the macroscopic observables $A$, or one of their derivatives, change discontinuously. In a ferromagnetic material at high temperatures, the magnetization $\boldsymbol{m}$ of the material (alignment of atomic spins) is determined by the external magnetic field $\boldsymbol{B}$. At low temperatures, the material shows a spontaneous magnetization even at zero external field, which breaks rotational symmetry. With increasing temperature, the spontaneous magnetization $\boldsymbol{m}$ decreases and vanishes at the Curie temperature $T_*$ as

$$|\boldsymbol{m}| \sim (T_* - T)^{\gamma} \qquad [20]$$

In the presence of a very weak external field, the spontaneous magnetization aligns itself with the external field $\boldsymbol{B}$, while its strength is, to leading order, independent of $\boldsymbol{B}$. The function $\boldsymbol{m}(\boldsymbol{B}, T)$,

therefore, changes discontinuously at $\boldsymbol{B}=0$. The line $\boldsymbol{B}=0$ for $T < T_*$ is, therefore, a line of first-order phase transitions between the possible directions of the spontaneous magnetization (in a one-dimensional system, between $\boldsymbol{m}$ up and $\boldsymbol{m}$ down). This line ends at the critical point $(\boldsymbol{B}=0, T=T_*)$ where the order parameter $|\boldsymbol{m}|$ vanishes. The role of $\boldsymbol{B}=0$ as the critical value of $\boldsymbol{B}$ is obscured by the fact that $\boldsymbol{B}=0$ is singled out by symmetry.

A critical phase transition involves scale-invariant physics. One sign of this is that fluctuations appear on a large range of length scales between the underlying atomic scale and the scale of the sample. In particular, the atomic scale, and any dimensionful parameters associated with that scale, must become irrelevant at the critical point. This can be taken as the starting point for obtaining properties of the system at the critical point.

One first defines a semigroup acting on microstates: the renormalization group. Its action is to group together a small number of particles as a single particle of a fictitious new system, using some averaging procedure. Alternatively, this can also be done in Fourier space. One then defines a dual action of the renormalization group on the space of Hamiltonians by demanding that the partition function is invariant under the renormalization group action:

$$\sum_{\text{microstates}} \mathrm{e}^{-H} = \sum_{\text{microstates}'} \mathrm{e}^{-H'} \qquad [21]$$

The renormalized Hamiltonian $H'$ is in general more complicated than the original one, but it can be approximated by a fixed expression where only a finite number of parameters $\mu$ are adjusted. Fixed points of the renormalization group correspond to Hamiltonians with the parameters $\mu$ at their critical values. The critical value of any dimensional parameter $\mu$ must be zero (or infinity). Only dimensionless combinations can have nontrivial critical values.

The behavior of thermodynamical quantities at the critical point is in general not trivial to calculate. But the action of the renormalization group on length scales is given by its definition. The blowup of the correlation length $\xi$ at the critical point is, therefore, the easiest critical exponent to calculate. We make contact with critical phenomena in gravitational collapse by considering the time evolution in coordinates $(\tau, x)$ as a renormalization group action. The calculation of the critical exponent for the black hole mass $M$ is the precise analog of the calculation of the critical exponent for the correlation length $\xi$, substituting $T_* - T$ for $p - p_*$, and taking into account that the $\tau$-evolution in critical collapse is toward smaller scales, while the renormalization group flow goes toward larger scales: therefore, $\xi$ diverges at the critical point, while $M$ vanishes.

We have shown above that the black hole mass is controlled by one global function $P$ on phase space. Clearly, $P$ is the gravity equivalent of $T - T_*$ in the ferromagnet. But it is tempting to speculate (Gundlach 2002)that there is also a gravity equivalent of the external magnetic field $\boldsymbol{B}$, which gives rise to a second independent critical exponent. At least in some situations, the angular momentum of the initial data can play this role. Note that, like $\boldsymbol{B}$, angular momentum is a vector, with a critical value that is zero because all other values break rotational symmetry. Furthermore, the final black hole can have nonvanishing angular momentum, which must depend on the angular momentum of the initial data. The former is analogous to the magnetization $\boldsymbol{m}$, the latter to the external field $\boldsymbol{B}$. It can be shown that this analogy holds perturbatively for small angular momentum. Future numerical simulations will show if it goes further.

## Universality and Cosmic Censorship

Critical phenomena in gravitational collapse first generated interest because a complicated self-similar structure and dimensionless numbers $\gamma$ and $\Delta$ arise from generic initial data evolved by quite simple field equations. Another point of interest is the rather detailed analogy of phenomena in a deterministic field theory with critical phase transitions in statistical mechanics. But critical phenomena are important for general relativity mostly for a different reason.

Black holes are among the most important solutions of general relativity because of their universality: the black hole uniqueness theorems state that stable black holes are completely determined by their mass, angular momentum, and electric charge – the Kerr–Newman family of black holes. Perturbation theory shows that any perturbations of black holes from the Kerr–Newman solutions must be radiated away.

Critical solutions have a similar importance because they are generic intermediate states of the evolution that are also independent of the initial data. An important distinction is that critical solutions depend on the matter model, and are therefore less universal than black holes. However, critical phenomena in gravitational collapse seem to arise in axisymmetric vacuum spacetimes, and so are apparently not linked to the

presence of matter. Furthermore, they also arise in perfect-fluid matter with the equation of state $p = \rho/3$, which is that of an ultrarelativistic gas. This is a good approximation for matter at very high density, such as in the big bang. This is important because critical phenomena probe arbitrarily large matter densities or spacetime curvatures as the initial data are fine-tuned to the black hole threshold. At even higher densities, presumably on the Planck scale, scale invariance is again broken by quantum-gravity effects, and so critical phenomena will end there.

The cosmic censorship conjecture states that naked singularities do not arise from suitably generic initial data for suitably well-behaved matter. Critical phenomena in gravitational collapse have forced a tightening of this conjecture. Type II (self-similar) critical solutions contain a naked singularity, that is, a point of infinite spacetime curvature from which information can reach a distant observer. (By contrast, the singularity inside a black hole is hidden from distant observers.) On a kinematical level, this could be seen already from the form [10] of the metric. Because the critical solution is the end state for all initial data that are exactly on the black hole threshold, all initial data on the black hole threshold form a naked singularity. As type II critical phenomena appear to be generic at least in spherical symmetry, this means that in generic self-gravitating systems, the space of regular initial data that form naked singularities is larger than expected, namely of codimension 1. Excluding naked singularities from generic initial data may be the sharpest version of cosmic censorship one can now hope to prove.

Another point of interest in critical collapse is that it allows one to make a small region of arbitrarily high curvature from finite-curvature initial data. This may be a route for probing quantum-gravity effects. Similarly, one can make black holes that are much smaller than any length scale present in the initial data or the matter equation of state. An application has been suggested for this in cosmology, where primordial black holes could have masses much smaller than the Hubble scale at which they are created, rather than of the order of this scale.

## Outlook

Critical phenomena in gravitational collapse are now well understood in spherical symmetry, both theoretically and in numerical simulations. In some matter models, the phenomenology is quite complicated, but it still fits into the basic picture outlined here.

The crucial question as to what happens beyond spherical symmetry remains largely unanswered at the time of writing. Perturbation theory around spherical symmetry suggests that critical phenomena are not restricted to exactly spherical situations. This is also supported by simulations in axisymmetric (highly nonspherical) vacuum gravity. Other simulations of nonspherical gravitational collapse which cover the necessary range of spacetime scales required to see critical phenomena are only just becoming available, and the results are not yet clear-cut. For collapse with angular momentum, no high-resolution calculations have yet been carried out. As the necessary techniques become available, one should be prepared for numerical simulations to make dramatic extensions or corrections to the picture of critical collapse drawn up here.

*See also:* Computational Methods in General Relativity: The Theory; Spacetime Topology, Causal Structure and Singularities; Stability of Minkowski Space; Stationary Black Holes.

## Further Reading

Abrahams AM and Evans CR (1993) Critical behavior and scaling in vacuum axisymmetric gravitational collapse. *Physical Review Letters* 70: 2980–2983.

Choptuik MW (1993) Universality and scaling in gravitational collapse of a massless scalar field. *Physical Review Letters* 70: 9–12.

Choptuik MW (1999) Critical behavior in gravitational collapse. *Progress of Theoretical Physics* 136 (suppl.): 353–365.

Evans CR and Coleman JS (1994) Critical phenomena and self-similarity in the gravitational collapse of radiation fluid. *Physical Review Letters* 72: 1782–1785.

Gundlach C (1999) *Living Reviews in Relativity* 2: 4 (published electronically at http://www.livingreviews.org).

Gundlach C (2002) Critical gravitational collapse with angular momentum: from critical exponents to universal scaling functions. *Physical Review D* 65: 064019.

# Current Algebra

**G A Goldin**, Rutgers University, Piscataway, NJ, USA



## Introduction

Certain commutation relations among the current density operators in quantum field theories define an infinite-dimensional Lie algebra. The original current algebra of Gell-Mann described weak and electromagnetic currents of the strongly interacting particles (hadrons), leading to the Adler–Weisberger formula and other important physical results. This helped inspire mathematical and quantum-theoretic developments such as the Sugawara model, light cone currents, Virasoro algebra, the mathematical theory of affine Kac–Moody algebras, and nonrelativistic current algebra in quantum and statistical physics. Lie algebras of local currents may be the infinitesimal representations of loop groups, local current groups or gauge groups, diffeomorphism groups, and their semidirect products or other extensions. Broadly construed, current algebra thus leads directly into the representation theory of infinite-dimensional groups and algebras. Applications have ranged across conformally invariant field theory, vertex operator algebras, exactly solvable lattice and continuum models in statistical physics, exotic particle statistics and $q$-commutation relations, hydrodynamics and quantized vortex motion. This brief survey describes but a few highlights.

## Relativistic Local Current Algebra for Hadrons

To model superfluidity, Landau had proposed in 1941 a quantum hydrodynamics fundamentally based on local fluid densities and currents as (operator) dynamical variables. However, current algebra came into its own in theoretical physics with the ideas of Gell-Mann in the early 1960s. The basic concept, in the era just preceding quantum chromodynamics (QCD), was that even without knowing the Lagrangian governing hadron dynamics in detail, exact kinematical information – the local symmetry – could still be encoded in an algebra of currents. The local (vector and axial vector) current density operators, expressed where possible in terms of underlying quantized field operators in Hilbert space, were to form two octets of Lorentz 4-vectors, with each octet corresponding to the eight generators of the compact Lie group SU(3).

More specifically (Adler and Dashen 1968), let $\mathcal{F}_a^{\mu}(x), a = 1, 2, \ldots, 8, \mu = 0, 1, 2, 3,$ be an octet of hadronic vector currents, where as usual $x = (x^{\nu}) = (x^0, \boldsymbol{x})$ denotes a point in four-dimensional spacetime. Likewise, introduce an axial vector octet $\mathcal{F}_a^{5\mu}(x)$. Unless otherwise specified, we use natural units, where $\hbar = 1$ and $c = 1$. Define the corresponding charges $F_a$ and $F_a^5$ to be the space integrals of the time components of these currents, that is,

$$F_a(x^0) = \int \mathrm{d}^3x \mathcal{F}_a^0(x^0, \boldsymbol{x})$$
$$F_a^5(x^0) = \int \mathrm{d}^3x \mathcal{F}_a^{50}(x^0, \boldsymbol{x}) \qquad [1]$$

where $\mathrm{d}^3x = \mathrm{d}x^1\,\mathrm{d}x^2\,\mathrm{d}x^3$. Then $F_1, F_2, F_3$ are the three components $I_1, I_2, I_3$ of the isotopic spin, and $Y = (2\sqrt{3}/3)F_8$ is the hypercharge. The usual electromagnetic current $J_{\mathrm{em}}^{\mu}(x^0, \boldsymbol{x})$ is given by

$$J_{\mathrm{em}}^{\mu} = q\left(\mathcal{F}_3^{\mu} + \frac{\sqrt{3}}{3}\mathcal{F}_8^{\mu}\right) \qquad [2]$$

where $q$ is the unit elementary charge, and the total charge is given by $Q = \int \mathrm{d}^3x\, J_{\mathrm{em}}^0(x^0, \boldsymbol{x}) = q(I_3 + Y/2)$. The hadronic part of the weak current entering an effective Lagrangian can be written as

$$J_{\mathrm{w}}^{\mu} = \left[\left(\mathcal{F}_1^{\mu} - \mathcal{F}_1^{5\mu}\right) + \mathrm{i}\left(\mathcal{F}_2^{\mu} - \mathcal{F}_2^{5\mu}\right)\right]\cos\theta_{\mathrm{C}}$$
$$+ \left[\left(\mathcal{F}_4^{\mu} - \mathcal{F}_4^{5\mu}\right) + \mathrm{i}\left(\mathcal{F}_5^{\mu} - \mathcal{F}_5^{5\mu}\right)\right]\sin\theta_{\mathrm{C}} \qquad [3]$$

where $\theta_{\mathrm{C}}$ is the Cabibbo angle (determined experimentally to be $\sim 0.27$ rad). The terms with $\mathcal{F}_1 - \mathcal{F}_1^5$ and $\mathcal{F}_2 - \mathcal{F}_2^5$ are strangeness conserving, those with $\mathcal{F}_4 - \mathcal{F}_4^5$ and $\mathcal{F}_5 - \mathcal{F}_5^5$ are not.

The main current algebra hypothesis is that the time components $\mathcal{F}^0$ and $\mathcal{F}^{50}$ of these octets satisfy the equal-time commutation relations:

$$[\mathcal{F}_a^0(x^0, \boldsymbol{x}), \mathcal{F}_b^0(y^0, \boldsymbol{y})]_{x^0=y^0} = \mathrm{i}\delta^{(3)}(\boldsymbol{x} - \boldsymbol{y})\sum_d c_{abd}\mathcal{F}_d^0(x^0, \boldsymbol{x})$$
$$[\mathcal{F}_a^0(x^0, \boldsymbol{x}), \mathcal{F}_b^{50}(y^0, \boldsymbol{y})]_{x^0=y^0} = \mathrm{i}\delta^{(3)}(\boldsymbol{x} - \boldsymbol{y})\sum_d c_{abd}\mathcal{F}_d^{50}(x^0, \boldsymbol{x}) \qquad [4]$$
$$[\mathcal{F}_a^{50}(x^0, \boldsymbol{x}), \mathcal{F}_b^{50}(y^0, \boldsymbol{y})]_{x^0=y^0} = \mathrm{i}\delta^{(3)}(\boldsymbol{x} - \boldsymbol{y})\sum_d c_{abd}\mathcal{F}_d^0(x^0, \boldsymbol{x})$$

where the $c_{abd}$ are structure constants of the Lie algebra of SU(3), antisymmetric in the indices. Since current commutators relate bilinear expressions to

linear ones, they fix the normalizations of the currents. The chiral currents $\mathcal{F}_a^{L\mu} = (1/2)(\mathcal{F}_a^\mu - \mathcal{F}_a^{5\mu})$ and $\mathcal{F}_a^{R\mu} = (1/2)(\mathcal{F}_a^\mu + \mathcal{F}_a^{5\mu})$ commute with each other, so that the local current algebra decomposes into two independent pieces.

The Dirac $\delta$-functions in eqns [4] require that $\mathcal{F}_a^0$ and $\mathcal{F}_a^{50}$ be interpreted as (unbounded) operator-valued distributions; while the fixed-time condition suggests these should make mathematical sense as three-dimensional distributions, with $x^0$ held constant. Such distributions may be modeled on the test-function space $\mathcal{D}$ of real-valued, compactly supported, $C^\infty$ functions on the spacelike hyperplane $\boldsymbol{R}^3$. For functions $f_a, f_a^5 \in \mathcal{D}$, one has formally the "smeared currents" that are expected to be bona fide (unbounded) operators in Hilbert space; suppressing $x^0$,

$$\begin{aligned} \mathcal{F}_a^0(f_a) &= \int_{\boldsymbol{R}^3} \mathrm{d}^3x f_a(\boldsymbol{x}) \mathcal{F}_a^0(x^0, \boldsymbol{x}) \\ \mathcal{F}_a^{50}(f_a^5) &= \int_{\boldsymbol{R}^3} \mathrm{d}^3x f_a^5(\boldsymbol{x}) \mathcal{F}_a^{50}(x^0, \boldsymbol{x}) \end{aligned} \qquad [5]$$

Equations [4] then become

$$\begin{aligned} [\mathcal{F}_a^0(f_a), \mathcal{F}_b^0(f_b)] &= [\mathcal{F}_a^{50}(f_a), \mathcal{F}_b^{50}(f_b)] \\ &= \mathrm{i} \sum_d \mathcal{F}_d^0(c_{abd} f_a f_b) \\ [\mathcal{F}_a^0(f_a), \mathcal{F}_b^{50}(f_b)] &= \mathrm{i} \sum_d \mathcal{F}_d^{50}(c_{abd} f_a f_b) \end{aligned} \qquad [6]$$

Let $g(\boldsymbol{x})$ be a $C^\infty$ map from $\boldsymbol{R}^3$ to the Lie algebra $\mathcal{G}$ of chiral $SU(3) \times SU(3)$, equal to zero outside a compact set. The set of all such $\mathcal{G}$-valued functions forms an infinite-dimensional Lie algebra under the pointwise bracket, $[g, g'](\boldsymbol{x}) = [g(\boldsymbol{x}), g'(\boldsymbol{x})]$. Let us call this Lie algebra $\mathrm{map}_0(\boldsymbol{R}^3, \mathcal{G})$, where the subscript 0 indicates the condition of compact support when that is applicable (on compact manifolds, we omit the subscript). Expanding $g(\boldsymbol{x})$ with respect to a fixed basis of $\mathcal{G}$, we straightforwardly identify the map $g$ with the two octets of test functions $f_a$ and $f_a^5$. Then, defining $\mathcal{F}(g) = \sum_a \mathcal{F}_a^0(f_a) + \sum_a \mathcal{F}_a^{50}(f_a^5)$, eqns [6] are interpreted (for fixed $x^0$) as a representation $\mathcal{F}$ of $\mathrm{map}_0(\boldsymbol{R}^3, \mathcal{G})$.

Integrating out the spatial variables entirely using eqns [1] leads to a representation at $x^0$ of $\mathcal{G}$ by the charges $F_a$ and $F_a^5$. The Adler–Weisberger sum rule was first derived (in 1965) from the commutation relations of these charges, together with the assumption of a partially conserved axial-vector current (PCAC). It connected nucleon $\beta$-decay coupling with pion–nucleon scattering cross sections, agreeing well with experiment. Various low-energy theorems followed, also in accord with experiment. Shortly thereafter, Adler was able to eliminate the PCAC assumption, and derived a further sum rule going beyond an experimental test of the algebra of charges to test the actual local current algebra. Here, the prediction pertained to structure functions in the deep inelastic scattering of neutrinos. This was elaborated by Bjorken to inelastic electron scattering. On the theoretical side, the study of the chiral current in perturbation theory led into the theory of anomalies. All these ideas were highly influential in subsequent theoretical work (Treiman *et al.* 1985, Mickelsson 1989).

It is a natural idea to try to extend eqns [4] or [6], which elegantly express the combined ideas of locality and symmetry, to an equal-time commutator algebra that would also include the space components of the local currents $\mathcal{F}_a^k, k = 1, 2, 3$. One may write without difficulty the commutators of the charges in [1] with these space components:

$$\begin{aligned} [F_a(x^0), \mathcal{F}_b^k(x^0, \boldsymbol{x})] &= [F_a^5(x^0), \mathcal{F}_b^{5k}(x^0, \boldsymbol{x})] \\ &= \mathrm{i} \sum_d c_{abd} \mathcal{F}_d^k(x^0, \boldsymbol{x}) \\ [F_a(x^0), \mathcal{F}_b^{5k}(x^0, \boldsymbol{x})] &= [F_a^5(x^0), \mathcal{F}_b^k(x^0, \boldsymbol{x})] \\ &= \mathrm{i} \sum_d c_{abd} \mathcal{F}_d^{5k}(x^0, \boldsymbol{x}) \end{aligned} \qquad [7]$$

But the commutator of the local time component with the local space component of the current cannot be merely the obvious extrapolation from eqns [4] and [7], that is, it cannot be

$$\begin{aligned} &[\mathcal{F}_a(x^0, \boldsymbol{x}), \mathcal{F}_b^k(y^0, \boldsymbol{y})]_{x^0 = y^0} \\ &\quad = \mathrm{i}\delta^{(3)}(\boldsymbol{x} - \boldsymbol{y}) \sum_d c_{abd} \mathcal{F}_d^k(x^0, \boldsymbol{x}) \end{aligned}$$

and so forth. Under very general conditions, for a relativistic theory based on local quantum fields or local observables, additional "Schwinger terms" are required on the right-hand sides of such commutators (Renner 1968).

Well-known difficulties in specifying the Schwinger terms are associated with the fact that operator-valued distributions are singular when regarded as if they were functions of spacetime points. Thus, the product of two distributions at a point is often singular or undefined. When the currents forming a local current algebra are written as normal-ordered products of field operator distributions and their derivatives, the Schwinger terms in their commutation relations may be calculated, for example, by "splitting points" in the arguments of the underlying fields, and subsequently letting the separation tend toward zero. The general form of a Schwinger term typically involves the derivative of a $\delta$-function times an operator. This may be a multiple of the identity (i.e., a $c$-number) or not, depending on the underlying

field-theoretic model. Furthermore, when the number of spacetime dimensions is greater than $1+1$, the $c$-number Schwinger terms turn out to be infinite. Hence, we do not obtain this way a bona- fide infinite-dimensional, equal-time commutator algebra comprising all the components of the local currents.

## Sugawara, Kac–Moody, and Virasoro Algebras

Since equations such as [4] and [6] are not explicitly dependent on how the currents are constructed from underlying canonical fields, one has the possibility of writing a theory entirely in terms of self-adjoint currents as the dynamical variables, bypassing the field operators entirely, and expressing a Hamiltonian operator directly in terms of such local currents. This is in the spirit of approaches to quantum field theory based on local algebras of observables. It suggests consideration of relativistic current algebras with finite c-number or operator Schwinger terms in $s+1$ dimensions, $s \geq 1$.

The Sugawara model, which is of this type, turned out to be one of the most influential of those proposed in the late 1960s and early 1970s. Henceforth, let $G$ be a compact Lie group, and $\mathcal{G}$ its Lie algebra; let $F_a, a=1,\ldots,\dim\mathcal{G}$, be a basis for $\mathcal{G}$, with $[F_a, F_b] = \mathrm{i}\Sigma_d c_{abd} F_d$. The Sugawara current algebra, at the fixed time $x^0 = y^0$ (which, from here on, we suppress in the notation), is given by

$$
\begin{aligned}
[J_a^0(\boldsymbol{x}), J_b^0(\boldsymbol{y})] &= \mathrm{i}\delta^{(3)}(\boldsymbol{x}-\boldsymbol{y})\sum_d c_{abd} J_d^0(\boldsymbol{x}) \\
[J_a^0(\boldsymbol{x}), J_b^k(\boldsymbol{y})] &= \mathrm{i}\delta^{(3)}(\boldsymbol{x}-\boldsymbol{y})\sum_d c_{abd} J_d^k(\boldsymbol{x}) \\
&\quad + \mathrm{i}c\delta_{ab}\frac{\partial}{\partial x^k}\delta^{(3)}(\boldsymbol{x}-\boldsymbol{y})I \\
[J_a^k(\boldsymbol{x}), J_b^\ell(\boldsymbol{y})] &= 0 \quad (k,\ell = 1,2,3)
\end{aligned}
\qquad [8]
$$

where $J_a^\mu = (J_a^0, J_a^k), k = 1,2,3$, is again a 4-vector, $c$ is a finite constant, and $I$ is the identity operator. The time components in eqns [8] behave like the local currents in eqns [4]. The Schwinger term is a $c$-number, while setting the commutators of the space components to zero is the simplest choice consistent with the Jacobi identity. The Sugawara Hamiltonian is given in terms of the local currents by the formal expression:

$$
H = \frac{1}{2c}\sum_a \int_{R^3} \mathrm{d}^3x \left[ J_a^0(\boldsymbol{x})^2 + \sum_{k=1}^3 J_a^k(\boldsymbol{x})^2 \right] \qquad [9]
$$

where the pointwise products of the currents require interpretation in the particular representation. This Hamiltonian leads to current conservation equations for the $J_a^\mu$.

Related to the Sugawara current algebra, with $s=1$ and the spatial dimension compactified, are affine Kac–Moody and Virasoro algebras (Goddard and Olive 1986, Kac 1990). Consider the infinite-dimensional Lie algebra map$(S^1, \mathcal{G})$ of smooth functions from the circle to $\mathcal{G}$ under the pointwise bracket. This is also called a loop algebra. Referring to the basis $F_a$, define $T_a^{(m)}$ for integer $m$ to be the Fourier function $\theta \to F_a \exp[-\mathrm{i}m\theta]$. The pointwise bracket in map$(S^1, \mathcal{G})$ gives $[T_a^{(m)}, T_b^{(n)}] = \mathrm{i}\Sigma_d c_{abd} T_d^{(m+n)}$ for these generators. The corresponding (untwisted) affine Kac–Moody algebra is a (uniquely defined, nontrivial) one-dimensional central extension of this loop algebra – that is, the new generator commutes with all elements of the Lie algebra and, in an irreducible representation, must be a multiple of the identity. In such a representation, the new bracket can be written as

$$
[T_a^{(m)}, T_b^{(n)}] = \mathrm{i}\sum_d c_{abd} T_d^{(m+n)} + km\delta_{ab}\delta_{m,-n}I \qquad [10]
$$

where $k$ is a constant. Here, $T_a^{(m=0)}$ is again a representation of $\mathcal{G}$. Self-adjointness of the local currents in the representation imposes the condition $T_a^{(m)^*} = T_a^{(-m)}$.

Now the compactly supported $C^\infty$ (tangent) vector fields on a $C^\infty$ manifold $M$ form a natural Lie algebra under the Lie bracket, denoted by vect$_0(M)$. In local Euclidean coordinates, for $\boldsymbol{g}_1, \boldsymbol{g}_2 \in$ vect$_0(M)$, one can write this bracket as

$$
[\boldsymbol{g}_1, \boldsymbol{g}_2] = \boldsymbol{g}_1 \cdot \nabla \boldsymbol{g}_2 - \boldsymbol{g}_2 \cdot \nabla \boldsymbol{g}_1 \qquad [11]
$$

As the affine Kac–Moody algebras are central extensions of the algebra of $\mathcal{G}$-valued functions on $S^1$, so are Virasoro algebras central extensions of the algebra of vector fields on $S^1$. Let $L^{(m)}$ denote the (complexified) vector field described by $\exp[-\mathrm{i}m\theta](1/\mathrm{i})\partial/\partial\theta$, for integer $m$. These generators then satisfy $[L^{(m)}, L^{(n)}] = (m-n)L^{(m+n)}$. Adjoining to the Lie algebra of vector fields a new central element (commuting with all the $L^{(m)}$), the Virasoro bracket in an irreducible representation is given by the formula

$$
\begin{aligned}
[L^{(m)}, L^{(n)}] &= (m-n)L^{(m+n)} \\
&\quad + c\frac{(m+1)m(m-1)}{12}\delta_{m,-n}I
\end{aligned}
\qquad [12]
$$

where the numerical coefficient $c$ is called the Virasoro central charge; self-adjointness of the currents imposes $L^{(m)^*} = L^{(-m)}$. It is straightforward to verify that eqn [12] satisfies the Jacobi identity. The special form of the central term in the Virasoro current algebra results from the Gelfand–Fuks cohomology on the algebra of vector fields.

The Kac–Moody and Virasoro algebras, both modeled on $S^1$, may be combined to form a natural semidirect sum of Lie algebras, with the additional bracket

$$[T_a^{(m)}, L^{(n)}] = mT_a^{(m+n)} \qquad [13]$$

Roughly speaking, the Kac–Moody generators correspond to Fourier transforms of charge densities on $S^1$, whereas the Virasoro generators correspond to Fourier transforms of infinitesimal motions in $S^1$. The central extensions provide the finite, $c$-number Schwinger terms. These structures have important application to light cone current algebra, conformally invariant quantum field theories in $(1+1)$-dimensional spacetime, the quantum theory of strings, exactly solvable models in statistical mechanics, and many other domains.

Of greatest physical importance, both in quantum field theory and statistical mechanics, are those irreducible, self-adjoint representations of the Virasoro algebra known as highest weight representations, where the spectrum of the operator $L^{(m=0)}$ is bounded below. In these applications, one represents a pair of Virasoro algebras by mutually commuting sets of operators $L^{(m)}$ and $\bar{L}^{(m)}$. In the quantum theory, for example, one takes the total energy $H \propto \bar{L}^{(0)} + L^{(0)}$, and the total momentum $P \propto \bar{L}^{(0)} - L^{(0)}$. In a highest weight representation, there is a unique eigenstate of $L^{(0)}$ having the lowest eigenvalue $h$; for this "vacuum" $|h\rangle, L^{(m)}|h\rangle = 0,\ m > 0$.

Friedan, Qiu, and Shenker showed in 1984 that highest weight representations are characterized by a class of specific, non-negative values of the central charge $c$ and, correspondingly, of $h$: either $c \geq 1$ (and $h \geq 0$) or $c = 1 - 6(\ell+2)^{-1}(\ell+3)^{-1}$, $\ell = 1, 2, 3, \ldots$ (and $h$ assumes a corresponding, specified set of values for each value of $\ell$). In a beautiful application to the study of the critical behavior of well-known statistical systems, in which the generator of dilations is proportional to $\bar{L}^{(0)} + L^{(0)}$, they discovered a direct correspondence with permitted values of the central charge; thus, $c = 1/2$ for the Ising model, $c = 7/10$ for the tricritical Ising model, $c = 4/5$ for the three-state Potts model, and $c = 6/7$ for the tricritical three-state Potts model.

## Current Algebras and Groups

Local current algebras may be exponentiated to obtain corresponding infinite-dimensional topological groups (Pressley and Segal 1986, Mickelsson 1989, Kac 1990). Let $G$ be a Lie group whose Lie algebra is $\mathcal{G}$. The algebra $\mathrm{map}_0(M, \mathcal{G})$, consisting of smooth, compactly supported $\mathcal{G}$-valued functions on $M$ under the pointwise bracket, exponentiates to the local current group $\mathrm{Map}_0(M, G)$, consisting of smooth maps from $M$ to $G$ that are the identity outside a compact set in $M$, under the pointwise group operation. When $M$ is taken to be the four-dimensional spacetime manifold (rather than a spacelike hyperplane), the local current group modeled on $M$ is mathematically a gauge group for nonabelian gauge field theory.

Likewise, the algebra $\mathrm{vect}_0(M)$ exponentiates to the group $\mathrm{Diff}_0(M)$ of compactly supported $C^\infty$ diffeomorphisms of $M$ (under composition). The Kac–Moody and Virasoro algebras exponentiate to central extensions of the loop group $\mathrm{Map}(S^1, G)$ and the diffeomorphism group $\mathrm{Diff}(S^1)$, respectively. The semidirect sums of the Lie algebras are the infinitesimal generators of semidirect products of the groups.

Under appropriate technical conditions, self-adjoint representations of current algebras generate (and may be obtained from) continuous unitary representations of the corresponding groups. The needed technical conditions have to do with the existence of a dense set of analytic vectors belonging to a common, dense invariant domain of essential self-adjointness for the currents.

## Nonrelativistic Current Algebra

In nonrelativistic local current algebra, Schwinger terms do not appear. In 1968, Dashen and Sharp defined (at fixed time $t$, suppressed in the present notation) a mass density $\rho(\boldsymbol{x}) = m\psi^*(\boldsymbol{x})\psi(\boldsymbol{x})$ and a momentum density $\boldsymbol{J}(\boldsymbol{x}) = (\hbar/2\mathrm{i})\{\psi^*(\boldsymbol{x})\nabla\psi(\boldsymbol{x}) - [\nabla\psi^*(\boldsymbol{x})]\psi(\boldsymbol{x})\}$, where $\psi$ is a second-quantized canonical field; here we keep $\hbar$ in the notation. The resulting equal-time algebra is the semidirect sum:

$$\begin{aligned}
&[\rho(\boldsymbol{x}), \rho(\boldsymbol{y})] = 0 \\
&[\rho(\boldsymbol{x}), J^k(\boldsymbol{y})] = -\mathrm{i}\hbar \frac{\partial}{\partial x^k}[\delta^{(3)}(\boldsymbol{x}-\boldsymbol{y})\rho(\boldsymbol{x})] \\
&[J^k(\boldsymbol{x}), J^\ell(\boldsymbol{y})] = \mathrm{i}\hbar\left\{\frac{\partial}{\partial y^k}[\delta^{(3)}(\boldsymbol{x}-\boldsymbol{y})J^\ell(\boldsymbol{y})] \right. \\
&\qquad\qquad \left. - \frac{\partial}{\partial x^\ell}[\delta^{(3)}(\boldsymbol{x}-\boldsymbol{y})J^k(\boldsymbol{x})]\right\}
\end{aligned} \qquad [14]$$

Since this current algebra is independent of whether $\psi$ obeys commutation or anticommutation relations, the information as to particle statistics (Bose or Fermi) is not encoded in the Lie algebra itself but in the choice of its representation (up to unitary equivalence). Again interpreting $\rho$ and $J^k$ as operator-valued distributions, define $\rho(f) = \int_{\boldsymbol{R}^3} \mathrm{d}^3x\, f(\boldsymbol{x})\rho(\boldsymbol{x})$ and $J(\boldsymbol{g}) = \int_{\boldsymbol{R}^3} \mathrm{d}^3x\, \Sigma_{k=1}^3 g^k(\boldsymbol{x})J^k(\boldsymbol{x})$, where $f$ and the

components $g^k$ of the vector field $\boldsymbol{g}$ belong to the function-space $\mathcal{D}$. Then the Lie algebra becomes

$$
\begin{aligned}
&[\rho(f_1), \rho(f_2)] = 0 \\
&[\rho(f), J(\boldsymbol{g})] = \mathrm{i}\hbar\rho(\boldsymbol{g}\cdot\nabla f) \\
&[J(\boldsymbol{g}_1), J(\boldsymbol{g}_2)] = -\mathrm{i}\hbar J([\boldsymbol{g}_1, \boldsymbol{g}_2])
\end{aligned} \qquad [15]
$$

Equations [15] are a representation by self-adjoint operators of the semidirect sum of the abelian Lie algebra $\mathcal{D}$ with $\mathrm{vect}_0(\boldsymbol{R}^3)$. The corresponding group is the natural semidirect product of the space $\mathcal{D}$ (regarded as an abelian topological group under addition) with $\mathrm{Diff}_0(\boldsymbol{R}^3)$.

The construction generalizes to a general manifold $M$ or manifold with boundary (in place of $\boldsymbol{R}^3$), and to a general set of charge densities that generate the local Lie algebra $\mathrm{map}_0(M, \mathcal{G})$. When $M = S^1$, we have the Kac–Moody and Virasoro algebras with central charge zero. However, $L^{(0)}$ in the nonrelativistic $(1+1)$-dimensional quantum theories is proportional to the total momentum $P$, and thus is unbounded above and below.

The continuous unitary representations of $\mathrm{Diff}_0(M)$, or its semidirect product with a local current group at fixed time, thus describe nonrelativistic quantum systems (Albeverio *et al.* 1999, Goldin 2004). The unitary representation $V(\phi), \phi \in \mathrm{Diff}_0(M)$, satisfies $V(\phi_r^{\boldsymbol{g}}) = \exp[\mathrm{i}(r/\hbar)J(\boldsymbol{g})]$, where $r \in \boldsymbol{R}$ and $\phi_r^{\boldsymbol{g}}$ is the one-parameter flow in $\mathrm{Diff}_0(M)$ generated by the vector field **g**. Such a representation may be described very generally by means of a measure $\mu$ on a configuration space $\Delta$, quasi-invariant under a group action of $\mathrm{Diff}_0(M)$ on $\Delta$, together with a unitary 1-cocycle $\chi$ on $\mathrm{Diff}_0(M) \times \Delta$. The Hilbert space for the representation is $\mathcal{H} = L^2_{\mathrm{d}\mu}(\Delta, \mathcal{W})$, which is the space of measurable functions $\Psi(\gamma), \gamma \in \Delta$, taking values in an inner product space $\mathcal{W}$, and square integrable with respect to $\mu$. The unitary representation $V$ is given by

$$
[V(\phi)\Psi](\gamma) = \chi_\phi(\gamma)\Psi(\phi\gamma)\sqrt{\frac{\mathrm{d}\mu_\phi}{\mathrm{d}\mu}(\gamma)} \qquad [16]
$$

where $\phi\gamma$ denotes the group action $\mathrm{Diff}_0(M) \times \Delta \to \Delta$; $\mu_\phi$ is the measure on $\Delta$ transformed by $\phi$ (which, by the quasi-invariance of $\mu$, is absolutely continuous with respect to $\mu$); $\mathrm{d}\mu_\phi/\mathrm{d}\mu$ is the Radon–Nikodym derivative; and $\chi_\phi(\gamma) : \mathcal{W} \to \mathcal{W}$ is a system of unitary operators in $\mathcal{W}$ obeying the cocycle equation

$$
\chi_{\phi_1\phi_2}(\gamma) = \chi_{\phi_1}(\gamma)\chi_{\phi_2}(\phi_1\gamma) \qquad [17]
$$

Equations [16] and [17] hold outside sets of $\mu$-measure zero in $\Delta$. Given the quasi-invariant measure $\mu$ on $\Delta$, one may always choose $\mathcal{W} = \mathbf{C}$ and $\chi_\phi(\gamma) \equiv 1$ to obtain a unitary group representation on complex-valued wave functions; but inequivalent cocycles describe unitarily inequivalent representations.

The configuration space $\Delta^{(N)}, N = 1, 2, 3, \ldots,$ consists of $N$-point subsets of $\boldsymbol{R}^s$, and $\mu^{(N)}$ is the (local) Lebesgue measure on $\Delta^{(N)}$. The corresponding diffeomorphism group and local current algebra representations describe $N$ identical quantum particles in $s$-dimensional space. When $\chi \equiv 1$, we have bosonic exchange symmetry. Inequivalent cocycles on $\Delta^{(N)}$ are obtained (for $s \geq 2$) by inducing (generalizing Mackey's method) from inequivalent unitary representations of the fundamental group $\pi_1[\Delta^{(N)}]$. For $s \geq 3$, this fundamental group is the symmetric group $S_N$ of particle permutations; the odd representation of $S_N, N \geq 2$, gives fermionic exchange symmetry, while the higher-dimensional representations are associated with particles satisfying the parastatistics of Greenberg and Messiah.

When $s = 2$, however, $\pi_1[\Delta^{(N)}]$ is the braid group $B_N$. Goldin, Menikoff, and Sharp obtained induced representations of the current algebra describing the intermediate statistics proposed by Leinaas and Myrheim for identical particles in 2-space. Such excitations, subsequently termed "anyons" by Wilczek and characterized as charge-flux tube composites, are important constructs in the theory of surface phenomena such as the quantum Hall effect, and anyonic statistics has also been applied to the study of high-$T_{\mathrm{c}}$ superconductivity. Current algebra representations induced by higher-dimensional representations of $B_N$ describe the statistics of "plektons." Similarly, current algebra in nonsimply connected space describes the Aharonov–Bohm effect.

Let $\psi^*(h) = \int_{\boldsymbol{R}^s} \mathrm{d}^s x\, h(\boldsymbol{x})\psi^*(\boldsymbol{x})$ denote the smeared creation field. Let the indexed set of representations $\rho_N, J_N, N = 0, 1, 2, \ldots,$ satisfying the current algebra [15], act in Hilbert spaces $\mathcal{H}_N$, where $\psi^*(h) : \mathcal{H}_N \to \mathcal{H}_{N+1}, \psi(h) : \mathcal{H}_{N+1} \to \mathcal{H}_N, \psi(h)|\mathcal{H}_0 = 0$, so that $\psi^*$ and $\psi$ intertwine the $N$-particle diffeomorphism group representations. Let $\rho(f)$ and $J(\boldsymbol{g})$ act on $\bigoplus_{N=0}^{\infty} \mathcal{H}_N$, so that $\rho(f)\Psi_N = \rho_N(f)\Psi_N, J(\boldsymbol{g})\Psi_N = J_N(\boldsymbol{g})\Psi_N$. Then conditions for a Fock space hierarchy are specified by commutator brackets of the fields with the currents:

$$
\begin{aligned}
&[\rho(f), \psi^*(h)] = \psi^*(\rho_{N=1}(f)h) \\
&[J(\boldsymbol{g}), \psi^*(h)] = \psi^*(J_{N=1}(\boldsymbol{g})h)
\end{aligned} \qquad [18]
$$

The local creation and annihilation fields for anyons in $\boldsymbol{R}^2$, obeying [18], satisfy $q$-commutation relations, where $q$ is the relative phase change associated with a single counterclockwise exchange of two anyons,

and the $q$-commutator $[A, B]_q = AB - qBA$. These relations generalize the canonical commutation ($q = 1$) and anticommutation ($q = -1$) relations of quantum field theory.

When $\Delta$ is the configuration space of infinite but locally finite subsets of $\boldsymbol{R}^s$, nonrelativistic current algebra describes the physics of infinite gases in continuum classical or quantum statistical mechanics. Here, the most important kinds of measures $\mu$ are Poisson measures (associated with gases of noninteracting particles at fixed average density) or Gibbsian measures (associated with translation-invariant two-body interactions). These measures describe equilibrium states and correlation functions in the classical case, and specify the current algebra representations in the quantum theory.

The group of volume-preserving diffeomorphisms was taken by Arnold as the symmetry group of an ideal, classical, incompressible fluid, and Marsden and Weinstein described the hydrodynamics of such a fluid using the Lie–Poisson bracket associated with the nonrelativistic current algebra of divergenceless vector fields. The idea of using this algebra to study quantized fluid motion, included in the program proposed by Rasetti and Regge, formed the basis of the subsequent study of quantized vortex structures in superfluids from the point of view of geometric quantization on coadjoint orbits of the diffeomorphism group. This leads to quantum configuration spaces whose elements are no longer sets of points – for example, spaces of vortex filaments in $\boldsymbol{R}^2$, or ribbons and tubes in $\boldsymbol{R}^3$.

*See also:* Algebraic Approach to Quantum Field Theory; Electroweak Theory; Quantum Chromodynamics; Solitons and Kac–Moody Lie Algebras; Symmetries in Quantum Field Theory: Algebraic Aspects; Toda Lattices; Two-Dimensional Conformal Field Theory and Vertex Operator Algebras.

## Further Reading

Adler SL and Dashen RF (1968) *Current Algebras and Applications to Particle Physics*. New York: Benjamin.

Albeverio S, Kondratiev YuG, and Röckner M (1999) Diffeomorphism groups and current algebras: configuration space analysis in quantum theory. *Reviews in Mathematical Physics* 11: 1–23.

Arnol'd VI and Khesin BA (1998) *Topological Methods in Hydrodynamics*. Applied Mathematical Sciences, vol. 125. Berlin: Springer.

Gell-Mann M and Ne'eman Y (2000) *The Eightfold Way (1964) (reissued)*. Cambridge, MA: Perseus Publishing.

Goddard P and Olive D (1986) Kac–Moody and Virasoro algebras in relation to quantum physics. *International Journal of Modern Physics A* 1: 303–414.

Goldin GA (1996) Quantum vortex configurations. *Acta Physica Polonica B* 27: 2341–2355.

Goldin GA (2004) Lectures on diffeomorphism groups in quantum physics. In: Govaerts J, Hounkonnou N, and Msezane AZ (eds.) *Contemporary Problems in Mathematical Physics: Proceedings of the Third International Workshop*, pp. 3–93. Singapore: World Scientific.

Goldin GA and Sharp DH (1991) The diffeomorphism group approach to anyons. *International Journal of Modern Physics B* 5: 2625–2640.

Ismagilov RS (1996) *Representations of Infinite-Dimensional Groups*. Translations of Mathematical Monographs, vol. 152. Providence, RI: American Mathematical Society.

Kac V (1990) *Infinite Dimensional Lie Algebras*. Cambridge: Cambridge University Press.

Marsden J and Weinstein A (1983) Coadjoint orbits, vortices, and Clebsch variables for incompressible fluids. *Physica D* 7: 305–323.

Mickelsson J (1989) *Current Algebras and Groups*. New York: Plenum.

Ottesen JT (1995) *Infinite Dimensional Groups and Algebras in Quantum Physics*. Berlin: Springer.

Pressley A and Segal G (1986) *Loop Groups*. Oxford: Oxford University Press.

Renner B (1968) *Current Algebras and Their Applications*. Oxford: Pergamon.

Sharp DH and Wightman AS (eds.) (1974) *Local Currents and Their Applications*. New York: Elsevier.

Treiman SB, Jackiw R, Zumino B, and Witten E (1985) *Current Algebra and Anomalies*. Singapore: World Scientific.

# D

## Deformation Quantization

**A C Hirshfeld**, Universität Dortmund, Dortmund, Germany



### Introduction

Deformation quantization is an alternative way of looking at quantum mechanics. Some of its techniques were introduced by the pioneers of quantum mechanics, but it was first proposed as an autonomous theory in a paper in *Annals of Physics* (Bayen *et al.* 1978). More recent reviews treat modern developments (HH I 2001, Dito and Sternheimer 2002, Zachos 2002).

Deformation quantization concentrates on the central physical concepts of quantum theory: the algebra of observables and their dynamical evolution. Because it deals exclusively with functions of phase-space variables, its conceptual break with classical mechanics is less severe than in other approaches. It formulates the correspondence principle very precisely which played such an important role in the historical development.

Although this article deals mainly with nonrelativistic bosonic systems, deformation quantization is much more general. For inclusion of fermions and the Dirac equation see (Hirshfeld *et al.* 2002b). The fermionic degrees of freedom may, in special cases, be obtained from the bosonic ones by supersymmetric extension (Hirshfeld *et al.* 2004). For applications to field theory, see Hirshfeld *et al.* (2002). For the relation to Hopf algebras see Hirshfeld *et al.* (2003), and to geometric algebra, see Hirshfeld *et al.* (2005).

The observables of a physical system, such as the Hamilton function, are smooth real-valued functions on phase space. Physical quantities of the system at some time, such as the energy, are calculated by evaluating the Hamilton function at the point $x_0 = (q_0, p_0)$ in phase space that characterizes the state of the system at this time (we assume for the moment, a one-particle system). The mathematical expression for this operation is

$$E = \int H(q,p)\delta^{(2)}(q - q_0, p - p_0)\, dq\, dp \quad [1]$$

where $\delta^{(2)}$ is the two-dimensional Dirac delta function. The observables of the dynamical system are functions on the phase space, the states of the system are positive functionals on the observables (here the Dirac delta functions), and we obtain the value of the observable in a definite state by the operation shown in eqn [1].

In general, functions on a manifold are multiplied by each other in a pointwise manner, that is, given two functions $f$ and $g$, their product $fg$ is the function

$$(fg)(x) = f(x)g(x) \quad [2]$$

In the context of classical mechanics, the observables build a commutative algebra, called the commutative "classical algebra of observables."

In Hamiltonian mechanics there is another way to combine two functions on phase space in such a way that the result is again a function on the phase space, namely by using the Poisson bracket

$$\begin{aligned} \{f,g\}(q,p) &= \sum_{i=1}^{n} \left( \frac{\partial f}{\partial q_i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q_i} \right)\bigg|_{q,p} \\ &= f\left( \overleftarrow{\partial}_q \overleftarrow{\partial}_p - \overleftarrow{\partial}_p \overrightarrow{\partial}_p \right) g \end{aligned} \quad [3]$$

in an abbreviated notation.

The notation can be further abbreviated by using $x$ to represent points of the phase-space manifold, $x = (x_1, \ldots, x_{2n})$, and introducing the Poisson tensor $\alpha^{ij}$, where the indices $i, j$ run from 1 to $2n$. In canonical coordinates $\alpha^{ij}$ is represented by the matrix

$$\alpha = \begin{pmatrix} 0 & -I_n \\ I_n & 0 \end{pmatrix} \quad [4]$$

where $I_n$ is the $n \times n$ identity matrix. Then eqn [3] becomes

$$\{f, g\}(x) = \alpha^{ij}\partial_i f(x)\, \partial_j g(x) \quad [5]$$

where $\partial_i = \partial/\partial x_i$.

For a general observable,

$$\dot{f} = \{f, H\} \quad [6]$$

Because $\alpha$ transforms like a tensor with respect to coordinate transformations, eqn [5] may also be written in noncanonical coordinates. In this case the components of $\alpha$ need not be constants, and may depend on the point of the manifold at which they are evaluated. But in Hamiltonian mechanics, $\alpha$ is still required to be invertible. A manifold equipped with a Poisson tensor of this kind is called a symplectic manifold. In general, the tensor $\alpha$ is no longer required to be invertible, but it nevertheless suffices to define Poisson brackets via eqn [5], and these brackets are required to have the properties

1. $\{f, g\} = -\{g, f\}$,
2. $\{f, gh\} = \{f, g\}h + g\{f, h\}$, and
3. $\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0$.

Property (1) implies that the Poisson bracket is antisymmetric, property (2) is referred to as the Leibnitz rule, and property (3) is called the Jacobi identity. The Poisson bracket used in Hamiltonian mechanics satisfies all these properties, but we now abstract these properties from the concrete prescription of eqn [3], and a Poisson manifold $(M, \alpha)$ is defined as a smooth manifold $M$ equipped with a Poisson tensor $\alpha$, whose components are no longer necessarily constant, such that the bracket defined by eqn [5] has the above properties. It turns out that such manifolds provide a better context for treating dynamical systems with symmetries. In fact, they are essential for treating gauge-field theories, which govern the fundamental interactions of elementary particles.

## Quantum Mechanics and Star Products

The essential difference between classical and quantum mechanics is Heisenberg's uncertainty relation, which implies that in the latter, states can no longer be represented as points in phase space. The uncertainty is a consequence of the noncommutativity of the quantum mechanical observables. That is, the commutative classical algebra of observables must be replaced by a noncommutative quantum algebra of observables.

In the conventional approach to quantum mechanics, this noncommutativity is implemented by representing the quantum mechanical observables by linear operators in Hilbert space. Physical quantities are then represented by eigenvalues of these operators, and physical states are related to the operator eigenfunctions. Although these entities are somehow related to their classical counterparts, to which they are supposed to reduce in an appropriate limit, the precise relationship has remained obscure, one hundred years after the beginnings of quantum mechanics. Textbooks refer to the correspondence principle, which guided the pioneers of the subject. Attempts to give this idea a precise formulation by postulating a specific relation between the classical Poisson brackets of observables and the commutators of the corresponding quantum mechanical operators, as undertaken, for example, by Dirac and von Neumann, encountered insurmountable difficulties, as pointed out by Groenewold in 1946 in an unjustly neglected paper (Groenewold 1948). In the same paper Groenewold also wrote down the first explicit representation of a "star product" (see eqn [11]), without however realizing the potential of this concept for overcoming the difficulties that he wanted to resolve.

In the deformation quantization approach, there is no such break when going from the classical system to the corresponding quantum system; we describe the quantum system by using the same entities that are used to describe the classical system. The observables of the system are described by the same functions on phase space as their classical counterparts. Uncertainty is realized by describing physical states as distributions on phase space that are not sharply localized, in contrast to the Dirac delta functions which occur in the classical case. When we evaluate an observable in some definite state according to the quantum analog of eqn [1] (see eqn [24]), values of the observable in a whole region contribute to the number that is obtained, which is thus an average value of the observable in the given state. Noncommutativity is incorporated by introducing a noncommutative product for functions on phase space, so that we get a new noncommutative quantum algebra of observables. The systematic work on deformation quantization stems from Gerstenhaber's seminal paper, where he introduced the concept of a star product of smooth functions on a manifold (Gerstenhaber 1964).

For applications to quantum mechanics, we consider smooth complex-valued functions on a Poisson manifold. A star product $f * g$ of two such functions is a new smooth function, which, in general, is described by an infinite power series:

$$f * g = fg + (\mathrm{i}\hbar)C_1(f, g) + O(\hbar^2) = \sum_{n=0}^{\infty} (\mathrm{i}\hbar)^n C_n(f, g) \qquad [7]$$

The first term in the series is the pointwise product given in eqn [2], and $(\mathrm{i}\hbar)$ is the deformation parameter, which is assumed to be varying continuously. If $\hbar$ is identified with Planck's constant, then what varies is really the magnitude of the

action of the dynamical system considered in units of $\hbar$: the classical limit holds for systems with large action. In this limit, which we express here as $\hbar \to 0$, the star product reduces to the usual product. In general, the coefficients $C_n$ will be such that the new product is noncommutative, and we consider the noncommutative algebra formed from the functions with this new multiplication law as a deformation of the original commutative algebra, which uses pointwise multiplication of the functions.

The expressions $C_n(f, g)$ denote functions made up of the derivatives of the functions $f$ and $g$. It is obvious that without further restrictions of these coefficients, the star product is too arbitrary to be of any use. Gerstenhaber's discovery was that the simple requirement that the new product be associative imposes such strong requirements on the coefficients $C_n$ that they are essentially unique in the most important cases (up to an equivalence relation, as discussed below). Formally, Gerstenhaber required that the coefficients satisfy the following properties:

1. $\sum_{j+k=n} C_j(C_k(f,g),h) = \sum_{j+k=n} C_j(f, C_k(g,h))$,
2. $C_0(f,g) = fg$, and
3. $C_1(f,g) - C_1(g,f) = \{f,g\}$.

Property (1) guarantees that the star product is associative: $(f*g)*h = f*(g*h)$. Property (2) means that in the limit $\hbar \to 0$, the star product $f*g$ agrees with the pointwise product $fg$. Property (3) has at least two aspects: (i) mathematically, it anchors the new product to the given structure of the Poisson manifold and (ii) physically, it provides the connection between the classical and quantum behavior of the dynamical system. Define a commutator by using the new product:

$$[f,g]_* = f*g - g*f \tag{8}$$

Property (3) may then be written as

$$\lim_{\hbar\to 0} \frac{1}{i\hbar}[f,g]_* = \{f,g\} \tag{9}$$

Equation [9] is the correct form of the correspondence principle. In general, the quantity on the left-hand side of eqn [9] reduces to the Poisson bracket only in the classical limit. The source of the mathematical difficulties that previous attempts to formulate the correspondence principle encountered was related to trying to enforce equality between the Poisson bracket and the corresponding expression involving the quantum mechanical commutator. Equation [9] shows that such a relation in general only holds up to corrections of higher order in $\hbar$.

For physical applications we usually require the star product to be Hermitean: $\overline{f*g} = \bar{g}*\bar{f}$, where $\bar{f}$ denotes the complex conjugate of $f$. The star products considered in this article have this property.

For a given Poisson manifold, it is not clear *a priori* if a star product for the smooth functions on the manifold actually exists, that is, whether it is at all possible to find coefficients $C_n$ that satisfy the above list of properties. Even if we find such coefficients, it it still not clear that the series they define through eqn [7] yields a smooth function. Mathematicians have worked hard to answer these questions in the general case. For flat Euclidian spaces, $M = \mathbb{R}^{2n}$, a specific star product has long been known. In this case, the components of the Poisson tensor $\alpha^{ij}$ can be taken to be constants. The coefficient $C_1$ can then be chosen antisymmetric, so that

$$C_1(f,g) = \tfrac{1}{2}\alpha^{ij}(\partial_i f)(\partial_j g) = \tfrac{1}{2}\{f,g\} \tag{10}$$

by property (3) above. The higher-order coefficients may be obtained by exponentiation of $C_1$. This procedure yields the Moyal star product (Moyal 1949):

$$f *_M g = f \exp\left(\left(\frac{i\hbar}{2}\right)\alpha^{ij}\overleftarrow{\partial}_i\overrightarrow{\partial}_j\right) g \tag{11}$$

In canonical coordinates, eqn [11] becomes

$$(f *_M g)(q,p) = f(q,p)\exp\left(\frac{i\hbar}{2}(\overleftarrow{\partial}_q\overrightarrow{\partial}_p - \overleftarrow{\partial}_p\overrightarrow{\partial}_q)\right) g(q,p) \tag{12}$$

$$= \sum_{m,n=0}^{\infty}\left(\frac{i\hbar}{2}\right)^{m+n}\frac{(-1)^m}{m!n!}(\partial_p^m\partial_q^n f)(\partial_p^n\partial_q^m g) \tag{13}$$

We now come to the question of uniqueness of the star product on a given Poisson manifold. Two star products $*$ and $*'$ are said to be "$c$-equivalent" if there exists an invertible transition operator

$$T = 1 + \hbar T_1 + \cdots = \sum_{n=0}^{\infty}\hbar^n T_n \tag{14}$$

where the $T_n$ are differential operators that satisfy

$$f *' g = T^{-1}((Tf)*(Tg)) \tag{15}$$

It is known that for $M = \mathbb{R}^{2n}$ all admissible star products are $c$-equivalent to the Moyal product. The concept of $c$-equivalence is a mathematical one ($c$ stands for cohomology (Gerstenhaber 1964)); it does not by itself imply any kind of physical equivalence, as shown below.

Another expression for the Moyal product is a kind of Fourier representation:

$$(f *_M g)(q,p) = \frac{1}{\hbar^2\pi^2}\int \mathrm{d}q_1\,\mathrm{d}q_2\,\mathrm{d}p_1\,\mathrm{d}p_2 f(q_1,p_1)g(q_2,p_2) \times \exp\left[\frac{2}{\mathrm{i}\hbar}(p(q_1-q_2)+q(p_2-p_1) + (q_2p_1-q_1p_2)\right] \qquad [16]$$

Equation [16] has an interesting geometrical interpretation. Denote points in phase space by vectors, for example, in two dimensions:

$$\mathbf{r} = \begin{pmatrix} q \\ p \end{pmatrix}, \qquad \mathbf{r}_1 = \begin{pmatrix} q_1 \\ p_1 \end{pmatrix}, \qquad \mathbf{r}_2 = \begin{pmatrix} q_2 \\ p_2 \end{pmatrix} \qquad [17]$$

Now, consider the triangle in phase space spanned by the vectors $\mathbf{r}-\mathbf{r}_1$ and $\mathbf{r}-\mathbf{r}_2$. Its area (symplectic volume) is

$$\begin{aligned} A(\mathbf{r},\mathbf{r}_1,\mathbf{r}_2) &= \tfrac{1}{2}(\mathbf{r}-\mathbf{r}_1)\wedge(\mathbf{r}-\mathbf{r}_2) \\ &= \tfrac{1}{2}[p(q_2-q_1)+q(p_1-p_2)+(q_1p_2-q_2p_1)] \end{aligned} \qquad [18]$$

which is proportional to the exponent in eqn [16]. Hence, we may rewrite eqn [16] as

$$(f*g)(\mathbf{r}) = \int \mathrm{d}\mathbf{r}_1\,\mathrm{d}\mathbf{r}_2 f(\mathbf{r}_1)g(\mathbf{r}_2)\exp\left[\frac{4\mathrm{i}}{\hbar}A(\mathbf{r},\mathbf{r}_1,\mathbf{r}_2)\right] \qquad [19]$$

## Deformation Quantization

The properties of the star product are well adapted for describing the noncommutative quantum algebra of observables. We have already discussed the associativity and the incorporation of the classical and semiclassical limits. Note that the characteristic nonlocality feature of quantum mechanics is also explicit. In the expression for the Moyal product given in eqn [13], the star product of the functions $f$ and $g$ at the point $x=(q,p)$ involves not only the values of the functions $f$ and $g$ at this point, but also all higher derivatives of these functions at $x$. But for a smooth function, knowledge of all the derivatives at a given point is equivalent to the knowledge of the function on the entire space. In the integral expression of eqn [16], we also see that knowledge of the functions $f$ and $g$ on the whole phase space is necessary to determine the value of the star product at the point $x$.

The $c$-equivalent star products correspond to different quantization schemes. Having chosen a quantization scheme, the quantities of interest for the quantum system may be calculated. It turns out that different quantization schemes lead to different spectra for the observables. The choice of a specific quantization scheme can only be motivated by further physical requirements. In the simple example we discuss below, the classical system is completely specified by its Hamilton function. In more general cases, one may have to decide what constitutes a sufficiently large set of good observables for a complete specification of the system (Bayen *et al.* 1978).

A state is characterized by its energy $E$; the set of all possible values for the energy is called the spectrum of the system. The states are described by distributions on phase space called projectors. The state corresponding to the energy $E$ is denoted by $\pi_E(q,p)$. These distributions are normalized:

$$\frac{1}{2\pi\hbar}\int \pi_E(q,\,p)\mathrm{d}q\,\mathrm{d}p = 1 \qquad [20]$$

and idempotent:

$$(\pi_E * \pi_{E'})(q,\,p) = \delta_{E,E'}\pi_E(q,p) \qquad [21]$$

The fact that the Hamilton function takes the value $E$ when the system is in the state corresponding to this energy is expressed by the equation

$$(H*\pi_E)(q,\,p) = E\pi_E(q,\,p) \qquad [22]$$

Equation [22] corresponds to the time-independent Schrödinger equation, and is sometimes called the "$*$-genvalue equation." The spectral decomposition of the Hamilton function is given by

$$H(q,\,p) = \sum_E E\pi_E(q,\,p) \qquad [23]$$

where the summation sign may indicate an integration if the spectrum is continuous. The quantum mechanical version of eqn [1] is

$$\begin{aligned} E &= \frac{1}{2\pi\hbar}\int (H*\pi_E)(q,\,p)\mathrm{d}q\,\mathrm{d}p \\ &= \frac{1}{2\pi\hbar}\int H(q,p)\,\pi_E(q,\,p)\mathrm{d}q\,\mathrm{d}p \end{aligned} \qquad [24]$$

where the last expression may be obtained by using eqn [16] for the star product.

The time-evolution function for a time-independent Hamilton function is denoted by $\mathrm{Exp}(Ht)$, and the fact that the Hamilton function is the generator of the time evolution of the system is expressed by

$$\mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Exp}(Ht) = H*\mathrm{Exp}(Ht) \qquad [25]$$

This equation corresponds to the time-dependent Schrödinger equation. It is solved by the star exponential:

$$\mathrm{Exp}(Ht) = \sum_{n=0}^{\infty} \frac{1}{n!} \left( \frac{-\mathrm{i}t}{\hbar} \right)^n (H*)^n \qquad [26]$$

where $(H*)^n = \underbrace{H*H*\cdots*H}_{n\,\mathrm{times}}$. Because each state of definite energy $E$ has a time evolution $\exp(\mathrm{i}Et/\hbar)$, the complete time-evolution function may be written in the form

$$\mathrm{Exp}(Ht) = \sum_E \pi_E \mathrm{e}^{-\mathrm{i}Et/\hbar} \qquad [27]$$

This expression is called the "Fourier–Dirichlet expansion" for the time-evolution function.

Questions concerning the existence and uniqueness of the star exponential as a $C^\infty$ function and the nature of the spectrum and the projectors again require careful mathematical analysis. The problem of finding general conditions on the Hamilton function $H$ which ensure a reasonable physical spectrum is analogous to the problem of showing, in the conventional approach, that the symmetric operator $\hat{H}$ is self-adjoint and finding its spectral projections.

## The Simple Harmonic Oscillator

As an example of the above procedure, we treat the simple one-dimensional harmonic oscillator characterized by the classical Hamilton function

$$H(q,p) = \frac{p^2}{2m} + \frac{m\omega^2}{2} q^2 \qquad [28]$$

In terms of the holomorphic variables

$$a = \sqrt{\frac{m\omega}{2}} \left( q + \mathrm{i}\frac{p}{m\omega} \right), \quad \bar{a} = \sqrt{\frac{m\omega}{2}} \left( q - \mathrm{i}\frac{p}{m\omega} \right) \qquad [29]$$

the Hamilton function becomes

$$H = \omega a \bar{a} \qquad [30]$$

Our aim is to calculate the time-evolution function. We first choose a quantization scheme characterized by the normal star product

$$f *_N g = f \mathrm{e}^{\hbar \overleftarrow{\partial}_a \overrightarrow{\partial}_{\bar{a}}} g \qquad [31]$$

we then have

$$\bar{a} *_N a = a\bar{a}, \qquad a *_N \bar{a} = a\bar{a} + \hbar \qquad [32]$$

so that

$$[a, \bar{a}]_{*_N} = \hbar \qquad [33]$$

Equation [25] for this case is

$$\mathrm{i}\hbar \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{Exp}_N(Ht) = (H + \hbar\omega\bar{a}\partial_{\bar{a}}) \mathrm{Exp}_N(Ht) \qquad [34]$$

with the solution

$$\mathrm{Exp}_N(Ht) = \mathrm{e}^{-a\bar{a}/\hbar} \exp\left( \mathrm{e}^{-\mathrm{i}\omega t} a\bar{a}/\hbar \right) \qquad [35]$$

By expanding the last exponential in eqn [35], we obtain the Fourier–Dirichlet expansion

$$\mathrm{Exp}_N(Ht) = \mathrm{e}^{-a\bar{a}/\hbar} \sum_{n=0}^{\infty} \frac{1}{\hbar^n n!} \bar{a}^n a^n \, \mathrm{e}^{-\mathrm{i}n\omega t} \qquad [36]$$

From here, we can read off the energy eigenvalues and the projectors describing the states by comparing coefficients in eqns [27] and [36]:

$$\pi_0^{(N)} = \mathrm{e}^{-a\bar{a}/\hbar} \qquad [37]$$

$$\pi_n^{(N)} = \frac{1}{\hbar^n n!} \pi_0 \bar{a}^n a^n = \frac{1}{\hbar^n n} \bar{a}^n *_N \pi_0^{(N)} *_N a^n \qquad [38]$$

$$E_n = n\hbar\omega \qquad [39]$$

Note that the spectrum obtained in eqn [39] does not include the zero-point energy. The projector onto the ground state $\pi_0^{(N)}$ satisfies

$$a *_N \pi_0^{(N)} = 0 \qquad [40]$$

The spectral decomposition of the Hamilton function (eqn [23]) is in this case

$$H = \sum_{n=0}^{\infty} n\hbar\omega \left( \frac{1}{\hbar^n n!} \mathrm{e}^{-a\bar{a}/\hbar} \bar{a}^n a^n \right) = \omega a \bar{a} \qquad [41]$$

We now consider the Moyal quantization scheme. If we write eqn [12] in terms of holomorphic coordinates, we obtain

$$f *_M g = f \exp\left( \frac{\hbar}{2} (\overleftarrow{\partial}_a \overrightarrow{\partial}_{\bar{a}} - \overleftarrow{\partial}_{\bar{a}} \overrightarrow{\partial}_a) \right) g \qquad [42]$$

Here, we have

$$a *_M \bar{a} = a\bar{a} + \frac{\hbar}{2}, \qquad \bar{a} *_M a = a\bar{a} - \frac{\hbar}{2} \qquad [43]$$

and again

$$[a, \bar{a}]_{*_M} = \hbar \qquad [44]$$

The value of the commutator of two phase-space variables is fixed by property (3) of the star product,

and cannot change when one goes to a $c$-equivalent star product. The Moyal star product is $c$-equivalent to the normal star product with the transition operator

$$T = \mathrm{e}^{-(\hbar/2)\vec{\partial}_a \vec{\partial}_{\bar{a}}} \quad [45]$$

We can use this operator to transform the normal product version of the $*$-genvalue equation, eqn [22], into the corresponding Moyal product version according to eqn [15]. The result is

$$\begin{aligned} H *_{M} \pi_n^{(M)} &= \omega\left(\bar{a} *_{M} a + \frac{\hbar}{2}\right) *_{M} \pi_n^{(M)} \\ &= \hbar\omega\left(n + \tfrac{1}{2}\right)\pi_n^{(M)} \end{aligned} \quad [46]$$

with

$$\pi_0^{(M)} = T\pi_0^{(N)} = 2\mathrm{e}^{-2a\bar{a}/\hbar} \quad [47]$$

$$\pi_n^{(M)} = T\pi_n^{(N)} = \frac{1}{\hbar^n n}\bar{a}^n *_{M} \pi_0^{(M)} *_{M} a^n \quad [48]$$

The projector onto the ground state $\pi_0^{(M)}$ satisfies

$$a *_{M} \pi_0^{(M)} = 0 \quad [49]$$

We now have, for the spectrum,

$$E_n = \left(n + \tfrac{1}{2}\right)\hbar\omega \quad [50]$$

which is the textbook result. We conclude that for this problem, the Moyal quantization scheme is the correct one.

The use of the Moyal product in eqn [25] for the star exponential of the harmonic oscillator leads to the following differential equation for the time evolution function:

$$\begin{aligned} &\mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Exp}_M(Ht) \\ &= \left(H - \frac{(\hbar\omega)^2}{4}\partial_H - \frac{(\hbar\omega)^2}{4}H\partial_H^2\right)\mathrm{Exp}_M(Ht) \end{aligned} \quad [51]$$

The solution is

$$\mathrm{Exp}_M(Ht) = \frac{1}{\cos(\omega t/2)}\exp\left[\left(\frac{2H}{\mathrm{i}\hbar\omega}\right)\tan\left(\frac{\omega t}{2}\right)\right] \quad [52]$$

This expression can be brought into the form of the Fourier–Dirichlet expansion of eqn [27] by using the generating function for the Laguerre polynomials:

$$\frac{1}{1+s}\exp\left[\frac{zs}{1+s}\right] = \sum_{n=0}^{\infty} s^n(-1)^n L_n(z) \quad [53]$$

with $s = \mathrm{e}^{-\mathrm{i}\omega t}$. The projectors then become

$$\pi_n^{(M)} = 2(-1)^n \mathrm{e}^{-2H/\hbar\omega} L_n\left(\frac{4H}{\hbar\omega}\right) \quad [54]$$

which is equivalent to the expression already found in eqn [48].

## Conventional Quantization

One usually finds the observables characterizing some quantum mechanical system by starting from the corresponding classical system, and then, either by guessing or by using some more or less systematic method, and finding the corresponding representations of the classical quantities in the quantum system. The guiding principle is the correspondence principle: the quantum mechanical relations are supposed to reduce somehow to the classical relations in an appropriate limit. Early attempts to systematize this procedure involved finding an assignment rule $\Theta$ that associates to each phase-space function $f$ a linear operator in Hilbert space $\hat{f} = \Theta(f)$ in such a way that in the limit $\hbar \to 0$, the quantum mechanical equations of motion go over to the classical equations. Such an assignment cannot be unique, because even though an operator that is a function of the basic operators $\hat{Q}$ and $\hat{P}$ reduces to a unique phase-space function in the limit $\hbar \to 0$, there are many ways to assign an operator to a given phase-space function, due to the different orderings of the operators $\hat{Q}$ and $\hat{P}$ that all reduce to the original phase-space function. Different ordering procedures correspond to different quantization schemes. It turns out that there is no quantization scheme for systems with observables that depend on the coordinates or the momenta to a higher power than quadratic which leads to a correspondence between the quantum mechanical and the classical equations of motion, and which simultaneously strictly maintains the Dirac–von Neumann requirement that $(1/\mathrm{i}\hbar)[\hat{f}, \hat{g}] \leftrightarrow \{f, g\}$. Only within the framework of deformation quantization does the correspondence principle acquire a precise meaning.

A general scheme for associating phase-space functions and Hilbert space operators, which includes all of the usual orderings, is given as follows: the operator $\Theta_\lambda(f)$ corresponding to a given phase-space function $f$ is

$$\Theta_\lambda(f) = \int \tilde{f}(\xi, \eta)\mathrm{e}^{-\mathrm{i}(\xi\hat{Q} + \eta\hat{P})}\mathrm{e}^{\lambda(\xi,\eta)}\mathrm{d}\xi\,\mathrm{d}\eta \quad [55]$$

where $\tilde{f}$ is the Fourier transform of $f$, and $(\hat{Q}, \hat{P})$ are the Schrödinger operators that correspond to the phase-space variables $(q, p)$; $\lambda(\xi, \eta)$ is a quadratic form:

$$\lambda(\xi, \eta) = \frac{\hbar}{4}(\alpha\eta^2 + \beta\xi^2 + 2\mathrm{i}\gamma\xi\eta) \quad [56]$$

Different choices for the constants $(\alpha, \beta, \gamma)$ yield different operator ordering schemes.

The relation between operator algebras and star products is given by

$$\Theta(f)\Theta(g) = \Theta(f * g) \qquad [57]$$

where $\Theta$ is a linear assignment of the kind discussed above. Different assignments, which correspond to different operator orderings, correspond to $c$-equivalent star products. It demonstrates that the quantum mechanical algebra of observables is a representation of the star product algebra. Because in the algebraic approach to quantum theory all the information concerning the quantum system may be extracted from the algebra of observables, specifying the star product completely determines the quantum system.

The inverse procedure of finding the phase-space function that corresponds to a given operator $\hat{f}$ is, for the special case of Weyl ordering, given by

$$f(q,p) = \int \langle q + \tfrac{1}{2}\xi|\hat{f}|q - \tfrac{1}{2}\xi\rangle \mathrm{e}^{-\mathrm{i}\xi p/\hbar}\,\mathrm{d}\xi \qquad [58]$$

When using holonomic coordinates, it is convenient to work with the coherent states

$$\hat{a}|a\rangle = a|a\rangle, \quad \langle\bar{a}|\hat{a}^\dagger = \langle\bar{a}|\bar{a} \qquad [59]$$

These states are related to the energy eigenstates of the harmonic oscillator

$$|n\rangle = \frac{1}{\sqrt{n!}}\hat{a}^{\dagger^n}|0\rangle \qquad [60]$$

by

$$|a\rangle = e^{-\frac{1}{2}a\bar{a}/\hbar}\sum_{n=0}^{\infty}\frac{a^n}{\sqrt{n!}}|n\rangle,$$
$$\langle\bar{a}| = e^{-\frac{1}{2}a\bar{a}/\hbar}\sum_{n=0}^{\infty}\frac{\bar{a}^n}{\sqrt{n!}}\langle n| \qquad [61]$$

In normal ordering, we obtain the phase space function $f(a,\bar{a})$ corresponding to the operator $\hat{f}$ by just taking the matrix element between coherent states:

$$f(a,\bar{a}) = \langle\bar{a}|f(\hat{a},\hat{a}^\dagger)|a\rangle \qquad [62]$$

For holomorphic coordinates, it is easy to show

$$\pi_n^{(N)}(a,\bar{a}) = \frac{1}{\hbar^n}\langle\bar{a}|n\rangle\langle n|a\rangle = \frac{1}{\hbar^n n!}(\bar{a}a)^n \mathrm{e}^{-\bar{a}a/\hbar} \qquad [63]$$

in agreement with eqn [38] for the normal star product projectors.

The star exponential $\mathrm{Exp}(Ht)$ and the projectors $\pi_n$ are the phase-space representations of the time-evolution operator $\exp(-\mathrm{i}\hat{H}t/\hbar)$ and the projection operators $\hat{\rho}_n = |n\rangle\langle n|$, respectively. Weyl ordering corresponds to the use of the Moyal star product for quantization and normal ordering to the use of the normal star product. In the density matrix formalism, we say that the projection operator is that of a pure state, which is characterized by the property of being idempotent: $\hat{\rho}_n^2 = \hat{\rho}_n$ (compare eqn [21]). The integral of the projector over the momentum gives the probability distribution in position space:

$$\begin{aligned}\frac{1}{2\pi\hbar}&\int \pi_n^{(M)}(q,p)\mathrm{d}p \\ &= \frac{1}{2\pi\hbar}\int \langle q+\xi/2|n\rangle\langle n|q-\xi/2\rangle \mathrm{e}^{-\mathrm{i}\xi p/\hbar}\mathrm{d}\xi\,\mathrm{d}p \\ &= \langle q|n\rangle\langle n|q\rangle = |\psi_n(q)|^2\end{aligned} \qquad [64]$$

and the integral over the position gives the probability distribution in momentum space:

$$\frac{1}{2\pi\hbar}\int \pi_n^{(M)}(q,p)\mathrm{d}q = \langle p|n\rangle\langle n|p\rangle = |\tilde{\psi}_n(p)|^2 \qquad [65]$$

The normalization is

$$\frac{1}{2\pi\hbar}\int \pi_n^{(M)}(q,p)\mathrm{d}q\,\mathrm{d}p = 1 \qquad [66]$$

which is the same as eqn [20]. Applying these relations to the ground-state projector of the harmonic oscillator, eqn [47] shows that this is a minimum-uncertainty state. In the classical limit $\hbar \to 0$, it goes to a Dirac $\delta$-function. The expectation value of the Hamiltonian operator is

$$\begin{aligned}\frac{1}{2\pi\hbar}\int (H *_M \pi_n^{(M)})(q,p)\mathrm{d}q\,\mathrm{d}p &= \int \langle q|\hat{H}\hat{\rho}_n|q\rangle \mathrm{d}q \\ &= \mathrm{tr}(\hat{H}\hat{\rho}_n)\end{aligned} \qquad [67]$$

which should be compared to eqn [24].

## Quantum Field Theory

A real scalar field is given in terms of the coefficients $a(\mathbf{k}), \bar{a}(\mathbf{k})$ by

$$\phi(x) = \int \frac{\mathrm{d}^3 k}{(2\pi)^{3/2}\sqrt{2\omega_\mathbf{k}}}\left[a(\mathbf{k})\mathrm{e}^{-ikx} + \bar{a}(\mathbf{k})\mathrm{e}^{ikx}\right] \qquad [68]$$

where $\hbar\omega_\mathbf{k} = \sqrt{\hbar^2 k^2 + m^2}$ is the energy of a single-quantum of the field. The corresponding quantum field operator is

$$\Phi(x) = \int \frac{\mathrm{d}^3 k}{(2\pi)^{3/2}\sqrt{2\omega_\mathbf{k}}}\left[\hat{a}(\mathbf{k})\mathrm{e}^{-\mathrm{i}kx} + \hat{a}^\dagger(\mathbf{k})\mathrm{e}^{ikx}\right] \qquad [69]$$

where $\hat{a}(\mathbf{k}), \hat{a}^\dagger(\mathbf{k})$ are the annihilation and creation operators for a quantum of the field with momentum $\hbar\mathbf{k}$. The Hamiltonian is

$$H = \int \mathrm{d}^3 k \hbar\omega_\mathbf{k}\hat{a}^\dagger(\mathbf{k})\hat{a}(\mathbf{k}) \qquad [70]$$

$N(\mathbf{k}) = \hat{a}^\dagger(\mathbf{k})\hat{a}(\mathbf{k})$ is interpreted as the number operator, and eqn [70] is then just the generalization of eqn [39], the expression for the energy of the harmonic oscillator in the normal ordering scheme, for an infinite number of degrees of freedom. Had we chosen the Weyl-ordering scheme, it would have resulted in (by the generalization of eqn [50]) an infinite vacuum energy. Hence, requiring the vacuum energy to vanish implies the choice of the normal ordering scheme in free field theory. In the framework of deformation quantization, this requirement leads to the choice of the normal star product for treating free scalar fields: only for this choice is the star product well defined.

Currently, in realistic physical field theories involving interacting relativistic fields we are limited to perturbative calculations. The objects of interest are products of the fields. The analog of the Moyal product of eqn [11] for systems with an infinite number of degrees of freedom is

$$\phi(x_1) * \phi(x_2) * \cdots * \phi(x_n) = \exp\left[\frac{1}{2}\sum_{i<j}\int \mathrm{d}^4x\,\mathrm{d}^4y \frac{\delta}{\delta\phi_i(x)}\Delta(x-y)\frac{\delta}{\delta\phi_j(y)}\right] \times \phi_1(x_1), \ldots, \phi_n(x_n)|_{\phi_i=\phi} \qquad [71]$$

where the expressions $\delta/\delta\phi(x)$ indicate functional derivatives. Here, we have used the antisymmetric Schwinger function:

$$\Delta(x-y) = [\Phi(x), \Phi(y)] \qquad [72]$$

The Schwinger function is uniquely determined by relativistic invariance and causality from the equal-time commutator

$$\left[\Phi(x), \dot{\Phi}(y)\right]\big|_{x^0=y^0} = \mathrm{i}\hbar\delta^{(3)}(\mathbf{x}-\mathbf{y}) \qquad [73]$$

which is the characterization of the canonical structure in the field theoretic framework.

The Moyal product is, however, not the suitable star product to use in this context. In relativistic quantum field theory, it is necessary to incorporate causality in the form advocated by Feynman: positive frequencies propagate forward in time, whereas negative frequencies propagate backwards in time. This property is achieved by using the Feynman propagator:

$$\Delta_F(x) = \begin{cases} \Delta^+(x) & \text{for } x^0 > 0 \\ -\Delta^-(x) & \text{for } x^0 < 0 \end{cases} \qquad [74]$$

where $\Delta^+(x), \Delta^-(x)$ are the propagators for the positive and negative frequency components of the field, respectively. In operator language

$$\Delta_F(x-y) = \mathcal{T}(\Phi(x)\Phi(y)) - \mathcal{N}(\Phi(x)\Phi(y)) \qquad [75]$$

where $\mathcal{T}$ indicates the time-ordered product of the fields and $\mathcal{N}$ the normal-ordered product. Because the second term in eqn [75] is a normal-ordered product with vanishing vacuum expectation value, the Feynman propagator may be simply characterized as the vacuum expectation value of the time-ordered product of the fields. The antisymmetric part of the positive frequency propagator is the Schwinger function:

$$\Delta^+(x) - \Delta^+(-x) = \Delta^+(x) + \Delta^-(x) = \Delta(x) \qquad [76]$$

The fact that going over to a $c$-equivalent product leaves the antisymmetric part of the differential operator in the exponent of eqn [71] invariant suggests that the use of the positive frequency propagator instead of the Schwinger function merely involves the passage to a $c$-equivalent star product. This is indeed easy to verify. The time-ordered product of the operators is obtained by replacing the Schwinger function $\Delta(x-y)$ in eqn [72] by the $c$-equivalent positive frequency propagator $\Delta^+(x-y)$, restricting the time integration to $x^0 > y^0$, as in eqn [74], and symmetrizing the integral in the variables $x$ and $y$, which brings in the negative frequency propagator $\Delta^-(x-y)$ for times $x^0 < y^0$. Then eqn [71] becomes Wick's theorem, which is the basic tool of relativistic perturbation theory. In operator language

$$\mathcal{T}(\Phi(x_1), \ldots, \Phi(x_n)) = \exp\left[\frac{1}{2}\int \mathrm{d}^4x\,\mathrm{d}^4y \frac{\delta}{\delta\Phi(x)}\Delta_F(x-y)\frac{\delta}{\delta\Phi(y)}\right] \times \mathcal{N}(\Phi(x_1), \ldots, \Phi(x_n)) \qquad [77]$$

Another interesting relation between deformation quantization and quantum field theory has been uncovered by studies of the Poisson–Sigma model. This model involves a set of scalar fields $X^i$ which map a two-dimensional manifold $\Sigma_2$ onto a Poisson space $M$, as well as generalized gauge fields $A_i$, which are 1-forms on $\Sigma_2$ mapping to 1-forms on $M$. The action is given by

$$S_{PS} = \int_{\Sigma_2} (A_i dX^i + \alpha^{ij} A_i A_j) \qquad [78]$$

where $\alpha^{ij}$ is the Poisson structure of $M$. A remarkable formula was found (Cattaneo and Felder 2000):

$$(f * g)(x) = \int DXDA f(X(1))g(X(2))\mathrm{e}^{\mathrm{i}S_{PS}/\hbar} \qquad [79]$$

where $f$, $g$ are functions on $M$, $*$ is Kontsevich's star product (Kontsevich 1997), and the functional integration is over all fields $X$ that satisfy the boundary condition $X(\infty) = x$. Here $\Sigma_2$ is taken to be a disk in $\mathbb{R}^2$; 1, 2, and $\infty$ are three points on its circumference. By expanding the functional integral in eqn [79] according to the usual rules of perturbation theory, one finds that the coefficients of the powers of $\hbar$ reproduce the graphs

and weights that characterize Kontsevich's star product. For the case in which the Poisson tensor is invertible, we can perform the Gaussian integration in eqn [79] involving the fields $A_i$. The result is

$$(f * g)(x) = \int DXf(X(1))g(X(2)) \exp\left[\frac{i}{\hbar}\int \Omega_{ij} dX^i dX^j\right] \quad [80]$$

Equation [80] is formally similar to eqn [16] for the Moyal product, to which the Kontsevich product reduces in the symplectic case. Here $\Omega_{ij} = (\alpha^{ij})^{-1}$ is the symplectic 2-form, and $\int \Omega_{ij} dX^i\, dX^j$ is the symplectic volume of the manifold $M$. To make this relationship exact, one must integrate out the gauge degrees of freedom in the functional integral in eqn [79]. Since the Poisson-sigma model represents a topological field theory there remains only a finite-dimensional integral, which coincides with the integral in eqn [80].

*See also:* Deformations of the Poisson Bracket on a Symplectic Manifold; Deformation Quantization and Representation Theory; Deformation Theory; Fedosov Quantization; Noncommutative Geometry from Strings; Operads; Quantum Field Theory: A Brief Introduction; Schrödinger Operators.

## Further Reading

Bayen F, Flato M, Fronsdal C, Lichnerowicz A, and Sternheimer D (1978) Deformation theory and quantization I, II. *Annals of Physics (NY)* 111: 61–110, 111–151.

Cattaneo AS and Felder G (2000) A path integral approach to the Kontsevich quantization formula. *Communications in Mathematical Physics* 212: 591–611.

Dito G and Sternheimer D (2002) Deformation Quantization: genesis, developments, and metamorphoses. In: Halbout G (ed.) *Deformation Quantization*, IRMA Lectures in Mathematical Physics, vol. I, pp. 9–54, (Walter de Gruyter, Berlin, 2002), math.QA/02201168.

Gerstenhaber M (1964) On the deformation of rings and algebras. *Annals of Mathematics* 79: 59–103.

Groenewold HJ (1946) On the principles of elementary quantum mechanics. *Physica* 12: 405–460.

Hirshfeld A and Henselder P (2002a) Deformation quantization in the teaching of quantum mechanics. *American Journal of Physics* 70(5): 537–547.

Hirshfeld A and Henselder P (2002b) Deformation quantization for systems with fermions. *Annals of Physics* 302: 59–77.

Hirshfeld A and Henselder P (2002c) Star products and perturbative field theory. *Annals of Physics (NY)* 298: 382–393.

Hirshfeld A and Henselder P (2003) Star products and quantum groups in quantum mechanics and field theory. *Annals of Physics* 308: 311–328.

Hirshfeld A, Henselder P, and Spernat T (2004) Cliffordization, spin, and fermionic star products. *Annals of Physics (NY)* 314: 75–98.

Hirshfeld A, Henselder P, and Spernat T (2005) Star products and geometric algebra. *Annals of Physics (NY)* 317: 107–129.

Kontsevich M (1997) Deformation quantization of Poisson manifolds, q-alg/9709040.

Moyal JE (1949) Quantum mechanics as a statistical theory. *Proceedings of Cambridge Philosophical Society* 45: 99–124.

Zachos C (2001) Deformation quantization: quantum mechanics lives and works in phase space, hep-th/0110114.

# Deformation Quantization and Representation Theory

**S Waldmann**, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany



## The Quantization Problem

Though quantum theory for the classical phase space $\mathbb{R}^{2n}$ is well established by means of what usually is called canonical quantization, physics demands to go beyond $\mathbb{R}^{2n}$: On the one hand, systems with constraints lead by phase-space reduction to classical phase spaces different from $\mathbb{R}^{2n}$; in general one ends up with a symplectic or even Poisson manifold. Thus, one needs to quantize geometrically nontrivial phase spaces. On the other hand, field theories and thermodynamical systems require to pass from $\mathbb{R}^{2n}$ to infinitely many degrees of freedom, where one faces additional analytical difficulties. Both types of difficulties combine for gauge field theories and gravity, whence it is clear that quantization is still one of the most important issues in mathematical physics.

One possibility (among many others) is to use the structural similarity between the classical and quantum observable algebras. In both cases the observables constitute a complex $*$-algebra: in the classical case it is commutative with the additional structure of a Poisson bracket, whereas in the quantum case the algebra is noncommutative. In deformation quantization, one tries to pass from the classical observables to the quantum observables by a deformation of the algebraic structures.

## From Canonical Quantization to Star Products

Let us briefly recall canonical quantization and the ordering problem. In order to "quantize" classical

observables like the polynomials on $\mathbb{R}^{2n}$ to $q^k, p_l$, one assigns the operators

$$q^k \mapsto \varrho(q^k) = Q^k = (q \mapsto q^k \psi(q)) \qquad [1]$$

$$p_l \mapsto \varrho(p_l) = P_l = \left(q \mapsto \frac{\hbar}{\mathrm{i}} \frac{\partial \psi}{\partial q^l}(q)\right) \qquad [2]$$

for $k, l = 1, \ldots, n$, defined on a suitable domain in $L^2(\mathbb{R}^n, \mathrm{d}^n q)$. For simplicity, we choose $C_0^\infty(\mathbb{R}^n)$ as domain. The well-known ordering problem is encountered if one wants to also quantize higher polynomials. One convenient (although not the only) possibility is Weyl's total symmetrization rule, that is, for a monomial like $q^2 p$ we take the quantization

$$\begin{aligned} \varrho_{\mathrm{Weyl}}(q^2 p) &= \tfrac{1}{3}(Q^2 P + QPQ + PQ^2) \\ &= -\mathrm{i}\hbar q^2 \frac{\partial}{\partial q} - \mathrm{i}\hbar q \end{aligned} \qquad [3]$$

This can be written in the more explicit form:

$$\varrho_{\mathrm{Weyl}}(f) = \sum_{r=0}^{\infty} \frac{1}{r!} \left(\frac{\hbar}{\mathrm{i}}\right)^r \frac{\partial^r (Nf)}{\partial p_{i_1} \cdots \partial p_{i_r}}\bigg|_{p=0} \frac{\partial^r}{\partial q^{i_1} \cdots \partial q^{i_r}} \qquad [4]$$

with

$$N = \exp\left(\frac{\hbar}{2\mathrm{i}} \Delta\right) \quad \text{and} \quad \Delta = \frac{\partial^2}{\partial q^i \partial p_i}$$

Using [4] one can easily extend $\varrho_{\mathrm{Weyl}}$ to all functions $f \in C^\infty(\mathbb{R}^{2n})$ which are polynomial in the momentum variables only and have an arbitrary smooth dependence on the position variables. This Poisson subalgebra of $C^\infty(\mathbb{R}^{2n})$ certainly covers all classical observables of physical interest. Denoting these observables by $\mathrm{Pol}(T^*\mathbb{R}^n)$, one obtains a linear isomorphism

$$\varrho_{\mathrm{Weyl}} : \mathrm{Pol}(T^*\mathbb{R}^n) \xrightarrow{\cong} \mathrm{Diffop}(\mathbb{R}^n) \qquad [5]$$

into the differential operators with smooth coefficients, called Weyl symbol calculus. Other orderings would result in a different linear isomorphism like [5], for example, the standard ordering is obtained by simply omitting the operator $N$ in [4].

Using [5], one can pull back the operator product of $\mathrm{Diffop}(\mathbb{R}^n)$ to obtain a new product $\star_{\mathrm{Weyl}}$ for $\mathrm{Pol}(T^*\mathbb{R}^n)$, that is

$$f \star_{\mathrm{Weyl}} g = \varrho_{\mathrm{Weyl}}{}^{-1}(\varrho_{\mathrm{Weyl}}(f)\varrho_{\mathrm{Weyl}}(g)) \qquad [6]$$

which is called the Weyl–Moyal star product. Explicitly, one has

$$f \star_{\mathrm{Weyl}} g = \mu \circ \exp\left(\frac{\mathrm{i}\hbar}{2}\left(\frac{\partial}{\partial q^k} \otimes \frac{\partial}{\partial p_k} - \frac{\partial}{\partial p_k} \otimes \frac{\partial}{\partial q^k}\right)\right) f \otimes g \qquad [7]$$

where $\mu(f \otimes g) = fg$ is the commutative product. Clearly, for $f, g \in \mathrm{Pol}(T^*\mathbb{R}^n)$ the exponential series terminates after finitely many terms. If one now wants to extend further to all smooth functions, then [7] is only a formal power series in $\hbar$. Since on a manifold one does not have *a priori* a nice distinguished class of functions like $\mathrm{Pol}(T^*\mathbb{R}^n)$, one indeed has to generalize in this direction if a geometric framework is desired. This observation and the simple fact, that $\star_{\mathrm{Weyl}}$ satisfies all the following properties, lead to the definition of a formal star product by Bayen *et al.* (1978):

**Definition 1** A formal star product on a Poisson manifold $(M, \pi)$ is an associative $\mathbb{C}[[\lambda]]$-bilinear product

$$f \star g = \sum_{r=0}^{\infty} \lambda^r C_r(f, g) \qquad [8]$$

for $f, g \in C^\infty(M)[[\lambda]]$ such that

1. $C_0(f, g) = fg$ and $C_1(f, g) - C_1(g, f) = \mathrm{i}\{f, g\}$,
2. $1 \star f = f = f \star 1$, and
3. $C_r$ is a bidifferential operator.

If in addition $\overline{f \star g} = \bar{g} \star \bar{f}$, then $\star$ is called Hermitian.

Clearly, $\star_{\mathrm{Weyl}}$ defines a Hermitian star product for $\mathbb{R}^{2n}$. The first condition is called the correspondence principle in deformation quantization and the formal parameter $\lambda = \bar{\lambda}$ corresponds to Planck's constant $\hbar$ once a convergence scheme is established.

If $S = \mathsf{id} + \sum_{r=1}^{\infty} \lambda^r S_r$ is a formal series of differential operators with $S_r 1 = 0$ for $r \geq 1$, then it is easy to see that

$$f \star' g = S^{-1}(Sf \star Sg) \qquad [9]$$

defines again a star product which is Hermitian if $\star$ is Hermitian and if in addition $\overline{Sf} = S\bar{f}$. In particular, the operator $N$, as before, serves for the transition from $\star_{\mathrm{Weyl}}$ to the standard-ordered star product $\star_{\mathrm{Std}}$ obtained the same way from the standard-ordered quantization. Thus, [9] can be seen as the abstract notion of changing the ordering prescription, even if no operator representation has been specified. Two star products related by such an equivalence transformation are called equivalent and $*$-equivalent in the Hermitian case.

One main advantage of formal deformation quantization is that one has very strong existence and classification results:

**Theorem 2** *On every Poisson manifold there exists a star product.*

The above theorem was first shown by deWilde and Lecomte (1983) for the symplectic case and

independently by Fedosov (1985) and Omori, Maeda, and Yoshioka (1991). In 1997, Kontsevich was able to prove the general Poisson case by showing his profound formality theorem. The full classification of star products up to equivalence was first obtained for the symplectic case by Nest and Tsygan (1995) and independently by Deligne (1995), Bertelson, Cahen, and Gutt (1997), and Weinstein and Xu (1997). The general Poisson case again follows from Kontsevich's formality. In particular, in the symplectic case, star products are classified by their characteristic class

$$c : \star \mapsto c(\star) \in \frac{[\omega]}{\mathrm{i}\lambda} + \mathrm{H}^2_{\mathrm{deRham}}(M, \mathbb{C})[[\lambda]] \qquad [10]$$

As conclusion one can state that for the price of formal power series in $\hbar$ one obtains in formal deformation quantization a very general and well-understood picture of the observable algebra for the quantum version of any classical system described kinematically by a Poisson manifold. It turns out that already in this framework one can discuss dynamics as well by use of a Heisenberg equation formulated with $\star$. Moreover, the quantization of symmetries described by Hamiltonian Lie group or Lie algebra actions has been extensively studied.

For a physical theory of quantization, however, there are still at least two ingredients missing. On the one hand, one has to overcome the formal power series expansion in $\hbar$. This problem is, in principle, on the same footing as any perturbative approach to quantum theory and thus no easy answer can be expected to hold in general. In particular examples, however, such as the Weyl–Moyal star product, it can easily be solved. These issues together with the corresponding questions about a spectral calculus are best studied in the framework of Rieffel's strict deformation quantization based on a more $C^*$-algebraic formulation of the deformation problem. On the other hand, the observable algebra is not enough to describe a quantum system: one also needs to have a notion for the states. It turns out that already in the formal framework one has a physically reasonable notion of states as discussed by Bordemann and Waldmann (1998).

## States and Representations

The notion of states in deformation quantization is adapted from the $C^*$-algebraic world and based on the notion of positive functionals. Recall that for a $*$-algebra $\mathcal{A}$ over $\mathbb{C}$ a linear functional $\omega : \mathcal{A} \to \mathbb{C}$ is called positive if $\omega(a^*a) \geq 0$. For formal deformation quantization, things are slightly more subtle as now one has to consider $\mathbb{C}[[\lambda]]$-linear functionals

$$\omega : (C^\infty(M)[[\lambda]], \star) \longrightarrow \mathbb{C}[[\lambda]] \qquad [11]$$

where $\star$ is assumed to be a Hermitian star product in the following. Then the positivity is understood in the sense of formal power series where $a \in \mathbb{R}[[\lambda]]$ is called positive if $a = \sum_{r=r_0}^{\infty} \lambda^r a_r$ with $a_{r_0} > 0$. Thus, we can make sense out of the following requirement:

**Definition 3** Let $\star$ be a Hermitian star product on $M$. A $\mathbb{C}[[\lambda]]$-linear functional $\omega : C^\infty(M)[[\lambda]] \to \mathbb{C}[[\lambda]]$ is called positive with respect to $\star$ if

$$\omega(\overline{f} \star f) \geq 0 \qquad [12]$$

and it is called a state if, in addition, $\omega(1) = 1$.

In fact, $\omega(f)$ is interpreted as the expectation value of the observable $f$ in the state $\omega$. The positivity [12] ensures that the usual uncertainty relations between expectation values hold.

Sometimes it is convenient to consider positive functionals only defined on a (proper) $*$-ideal in $C^\infty(M)[[\lambda]]$, for instance, $C_0^\infty(M)[[\lambda]]$.

Since in some situations one wants more general formal series than just power series, it is convenient to embed the above definition of states into a larger and more algebraic context: consider an ordered ring $\mathsf{R}$, that is, a commutative, associative, unital ring $\mathsf{R}$ together with a distinguished subset $\mathsf{P} \subset \mathsf{R}$ (the positive elements) such that $\mathsf{R}$ is the disjoint union $-\mathsf{P} \dot{\cup} \{0\} \dot{\cup} \mathsf{P}$, and we have $\mathsf{P} \cdot \mathsf{P} \subseteq \mathsf{P}$ and $\mathsf{P} + \mathsf{P} \subseteq \mathsf{P}$. Then $\mathsf{C} = \mathsf{R}(\mathrm{i})$ denotes the ring extension by a square root i of –1 and consider $*$-algebras $\mathcal{A}$ over $\mathsf{C}$. Clearly, this generalizes the cases $\mathsf{R} = \mathbb{R}$, where $\mathsf{C} = \mathbb{C}$, as well as $\mathsf{R} = \mathbb{R}[[\lambda]]$, where $\mathsf{C} = \mathbb{C}[[\lambda]]$. In this way, one provides a framework where $C^*$-algebras, $*$-algebras over $\mathbb{C}$, and formal Hermitian star products can be treated on the same footing. It is clear that the definition of a positive functional immediately extends to $\omega : \mathcal{A} \to \mathsf{C}$ for such a ring $\mathsf{C}$.

**Example 4**

(i) For the Wick star product on $\mathbb{R}^{2n} \cong \mathbb{C}^n$, defined by

$$f \star_{\mathrm{Wick}} g = \sum_{r=0}^{\infty} \frac{(2\lambda)^r}{r!} \frac{\partial^r f}{\partial z^{i_1} \cdots \partial z^{i_r}} \frac{\partial^r g}{\partial \overline{z}^{i_1} \cdots \partial \overline{z}^{i_r}} \qquad [13]$$

the $\delta$-functional $\delta : f \mapsto f(0)$ is positive. Note, however, that $\delta$ is not positive for $\star_{\mathrm{Weyl}}$.

(ii) For the Weyl–Moyal star product $\star_{\mathrm{Weyl}}$ the Schrödinger functional

$$\omega(f) = \int_{\mathbb{R}^n} f(q, p = 0)\mathrm{d}^n q \qquad [14]$$

defined on the $*$-ideal $C_0^\infty(\mathbb{R}^{2n})[[\lambda]]$, is positive.

(iii) For any connected symplectic manifold $(M, \omega)$ and any Hermitian star product $\star$, there exists a unique normalized trace functional

$$\begin{aligned} \mathsf{tr} : C_0^\infty(M)[[\lambda]] &\longrightarrow \mathbb{C}[[\lambda]] \\ \mathsf{tr}(f \star g) &= \mathsf{tr}(g \star f) \end{aligned} \qquad [15]$$

with zeroth order equal to the integration over $M$ with respect to the Liouville measure $\Omega = \omega^n$. Then this trace is positive as well, $\mathsf{tr}(\bar{f} \star f) \geq 0$.

Having a notion for states as expectation-value functionals is still not enough to formulate quantum theory. One main feature of quantum states, the superposition principle, is not yet implemented. In particular, forming convex combinations like $\omega = c_1\omega_1 + c_2\omega_2$, with $c_1, c_2 \geq 0$ and $c_1 + c_2 = 1$, does not give a superposition of $\omega_1$ and $\omega_2$ but a mixed stated. Hence, one needs an additional linear structure on the states whence we look for a $*$-representation $\pi$ of the observable algebra $\mathcal{A}$ on a pre-Hilbert space $\mathcal{H}$ over $\mathsf{C}$ such that the states $\omega_1, \omega_2$ can be written as vector states $\omega_i(a) = \langle \phi_i, \pi(a)\phi_i \rangle$ for some unit vectors $\phi_1, \phi_2 \in \mathcal{H}$. Then one can build superpositions of the vectors $\phi_1, \phi_2$ in the usual way. While this is the well-known argument in any quantum theory based on the observable algebras, for deformation quantization one first has to make sense out of the above notions, since now $\mathsf{R} = \mathbb{R}[[\lambda]]$ is only an ordered ring. This can actually be done in a consistent way as demonstrated and exemplified by Bordemann, Bursztyn, Waldmann, and others.

We recall the basic results: A pre-Hilbert space $\mathcal{H}$ over $\mathsf{C}$ is a $\mathsf{C}$-module with a $\mathsf{C}$-sesquilinear inner product $\langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathsf{C}$ such that $\langle \phi, \psi \rangle = \overline{\langle \psi, \phi \rangle}$ and $\langle \phi, \phi \rangle > 0$ for $\phi \neq 0$. This makes sense since $\mathsf{R}$ is ordered. An operator $A : \mathcal{H}_1 \to \mathcal{H}_2$ is called adjointable if there exists an operator $A^* : \mathcal{H}_2 \to \mathcal{H}_1$ such that $\langle A\phi, \psi \rangle_2 = \langle \phi, A^*\psi \rangle_1$ for all $\phi \in \mathcal{H}_1$, $\psi \in \mathcal{H}_2$. The set of adjointable operators is denoted by $\mathfrak{B}(\mathcal{H}_1, \mathcal{H}_2)$, and $\mathfrak{B}(\mathcal{H}) = \mathfrak{B}(\mathcal{H}, \mathcal{H})$ turns out to be a $*$-algebra over $\mathsf{C}$. This allows one to define a $*$-representation $\pi$ of $\mathcal{A}$ on $\mathcal{H}$ to be a $*$-homomorphism $\pi : \mathcal{A} \to \mathfrak{B}(\mathcal{H})$. An intertwiner $T$ between two $*$-representations $(\mathcal{H}_1, \pi_1)$ and $(\mathcal{H}_2, \pi_2)$ is an operator $T \in \mathfrak{B}(\mathcal{H}_1, \mathcal{H}_2)$ with $T\pi_1(a) = \pi_2(a)T$ for all $a \in \mathcal{A}$. This defines the category $*$-$\mathsf{Rep}(\mathcal{A})$ of $*$-representations of $\mathcal{A}$.

Let us now recall that a positive linear functional $\omega$ can be written as an expectation value for a vector state in some representation. This is the well-known Gelfand–Naimark–Segal (GNS) construction from operator algebra theory which can be transferred to this purely algebraic context (Bordemann and Waldmann 1998). First recall that any positive linear functional $\omega : \mathcal{A} \to \mathsf{C}$ satisfies the Cauchy–Schwarz inequality

$$\omega(a^*b)\,\overline{\omega(a^*b)} \leq \omega(a^*a)\omega(b^*b) \qquad [16]$$

and $\omega(a^*b) = \overline{\omega(b^*a)}$. If $\mathcal{A}$ is unital, which will always be assumed for simplicity, then $\omega(a^*) = \overline{\omega(a)}$ follows. Then

$$\mathcal{J}_\omega = \{a \in \mathcal{A} \mid \omega(a^*a) = 0\} \qquad [17]$$

is a left ideal in $\mathcal{A}$, the so-called Gel'fand ideal, and hence $\mathcal{H}_\omega = \mathcal{A}/\mathcal{J}_\omega$ is a left $\mathcal{A}$-module with module structure denoted by $\pi_\omega(a)\psi_b = \psi_{ab}$, where $\psi_b \in \mathcal{H}_\omega$ denotes the equivalence class of $b \in \mathcal{A}$. Finally, $\langle \psi_b, \psi_c \rangle = \omega(b^*c)$ turns $\mathcal{H}_\omega$ into a pre-Hilbert space and $\pi_\omega$ becomes a $*$-representation, the GNS representation with respect to $\omega$. Moreover, $\psi_1 \in \mathcal{H}_\omega$ is a cyclic vector, $\psi_b = \pi_\omega(b)\psi_1$, with the property

$$\omega(a) = \langle \psi_1, \pi_\omega(a)\psi_1 \rangle \qquad [18]$$

These properties characterize the GNS representation $(\mathcal{H}_\omega, \pi_\omega, \psi_1)$ up to unitary equivalence.

**Example 5** We can now apply this construction to the three basic examples and obtain the following well-known representations as GNS representations:

(i) The GNS representation corresponding to the $\delta$-functional and the Wick star product is (unitarily equivalent to) the formal Bargmann–Fock representation. Here $\mathcal{H}_\delta = \mathbb{C}[[\bar{y}^1, \ldots, \bar{y}^n]][[\lambda]]$ with inner product

$$\begin{aligned} \langle \phi, \psi \rangle = \sum_{r=0}^{\infty} \frac{(2\lambda)^r}{r!} &\overline{\frac{\partial^r \phi}{\partial \bar{y}^{i_1} \cdots \partial \bar{y}^{i_r}}(0)} \\ &\times \frac{\partial^r \psi}{\partial \bar{y}^{i_1} \cdots \partial \bar{y}^{i_r}}(0) \end{aligned} \qquad [19]$$

and $\pi_\delta$ is explicitly given by

$$\begin{aligned} \pi_\delta(f) = \sum_{r,s=0}^{\infty} \frac{(2\lambda)^r}{r!s!} &\frac{\partial^{r+s} f}{\partial z^{i_1} \cdots \partial z^{i_r} \partial \bar{z}^{j_1} \cdots \partial \bar{z}^{j_s}}(0) \\ &\times \bar{y}^{j_1} \cdots \bar{y}^{j_s} \frac{\partial^r}{\partial \bar{y}^{i_1} \cdots \partial \bar{y}^{i_r}} \end{aligned} \qquad [20]$$

In particular, $\pi_\delta(z^i) = 2\lambda\partial/\partial\bar{y}^i$ and $\pi_\delta(\bar{z}^i) = \bar{y}^i$ are the annihilation and creation operators and [20] gives the Wick (or normal) ordering. This basic example has been extended to arbitrary Kähler manifolds by Bordemann and Waldmann (1998).

(ii) The Weyl–Moyal star product $\star_{\text{Weyl}}$ and the Schrödinger functional $\omega$ as in [14] give the usual Schrödinger representation as GNS representation. We obtain $\mathcal{H}_\omega = C_0^\infty(\mathbb{R}^n)[[\lambda]]$ with inner product

$$\langle \phi, \psi \rangle = \int_{\mathbb{R}^n} \overline{\phi(q)} \psi(q) \, \mathrm{d}^n q \qquad [21]$$

and $\pi_\omega(f) = \varrho_{\text{Weyl}}(f)$ as in [4] with $\hbar$ replaced by $\lambda$. The Schrödinger representation as a particular case of a GNS representation has been generalized to arbitrary cotangent bundles including representations on sections of line bundles over the configuration space (Dirac's representation for magnetic monopoles) by Bordemann, Neumaier, Pflaum, and Waldmann (1999, 2003). In this context, the WKB expansion can also be formulated.

(iii) For the positive trace $\mathsf{tr}$, the GNS pre-Hilbert is simply the space $\mathcal{H}_{\mathsf{tr}} = C_0^\infty(M)[[\lambda]]$ with inner product $\langle f, g \rangle = \mathsf{tr}(\overline{f} \star g)$. The corresponding GNS representation is the left regular representation $\pi_{\mathsf{tr}}(f)g = f \star g$. Note that in this case the commutant of the representation is (anti-)isomorphic to the observable algebra and given by all the right multiplications. Thus, $\pi_{\mathsf{tr}}$ is highly reducible and the size of the commutant indicates a "thermodynamical" interpretation of this representation. Indeed, one can take this GNS representation, and more general for arbitrary KMS functionals, as a starting point of a preliminary version of a Tomita–Takesaki theory for deformation quantization as shown by Waldmann (1999).

After these fundamental examples, we now reconsider the question of superpositions: in general, two (pure) states $\omega_1, \omega_2$ cannot be realized as vector states inside a single irreducible representation. One encounters superselection rules. Usually, for instance, in algebraic quantum field theory, the existence of superselection rules indicates the presence of charges. In particular, it is not sufficient to consider one single representation of the observable algebra $\mathcal{A}$. Instead, one has to investigate (as good as possible) all superselection sectors of the representation theory $*\text{-}\mathsf{Rep}(\mathcal{A})$ of $\mathcal{A}$ and find physically motivated criteria to select distinguished representations. In usual quantum mechanics on $\mathbb{R}^{2n}$, this turns out to be rather simple, thanks to the (nontrivial) uniqueness theorem of von Neumann: one has a unique irreducible representation of the Weyl algebra up to unitary equivalence. In infinite dimensions or in topologically nontrivial situations, however, von Neumann's theorem does not apply and one indeed has superselection rules.

In deformation quantization, some parts of these superselection rules have been understood well: again, for cotangent bundles $T^*Q$, one can classify the unitary equivalence classes of Schrödinger-like representations on $C_0^\infty(Q)[[\lambda]]$ by topological classes of nontrivial vector potentials. Thus, one arrives at the interpretation of the Aharonov–Bohm effect as superselection rule where the classification is essentially given by $\mathrm{H}^1_{\text{deRham}}(Q, \mathbb{C})/2\pi\mathrm{i}\ \mathrm{H}^1_{\text{deRham}}(Q, \mathbb{Z})$.

## General Representation Theory

Although it is very much desirable to determine the structure and the superselection sectors in $*\text{-}\mathsf{Rep}(\mathcal{A})$ completely, this is only achievable in the very simplest examples. Moreover, for formal star products, many artifacts due to the purely algebraic nature have to be expected: the Bargmann–Fock and Schrödinger representation in Example 5 are unitarily inequivalent and thus define a superselection rule, even the pre-Hilbert spaces are nonisomorphic. However, these artifacts vanish immediately when one imposes the suitable convergence conditions together with appropriate topological completions (von Neumanns's theorem). Given such problems, it is very difficult to find "hard" superselection rules which indeed have physical significance already at the formal level. Nevertheless, the example of the Aharonov–Bohm effect shows that this is possible. In any case, new techniques for investigating $*\text{-}\mathsf{Rep}(\mathcal{A})$ have to be developed. It turns out that comparing $*\text{-}\mathsf{Rep}(\mathcal{A})$ with some other $*\text{-}\mathsf{Rep}(\mathcal{B})$ is much simpler but still gives some nontrivial insight in the structure of the representation theory. Here the Morita theory provides a highly sophisticated tool.

The classical notion of Morita equivalence as well as Rieffel's more specialized strong Morita equivalence for $C^*$-algebras have been transferred to deformation quantization and, more generally, to $*$-algebras $\mathcal{A}$ over $\mathsf{C} = \mathsf{R}(\mathrm{i})$ by Bursztyn and Waldmann (2001). The aim is to construct functors

$$\mathsf{F} : *\text{-}\mathsf{Rep}(\mathcal{A}) \longrightarrow *\text{-}\mathsf{Rep}(\mathcal{B}) \qquad [22]$$

which allow us to compare these categories and determine whether they are equivalent. But even if they are not equivalent, functors such as [22] are interesting. As example, one considers the situation of classical phase space reduction $M \rightsquigarrow M_{\text{red}}$ as it is present in every constraint system or gauge theory. Suppose one succeeded with the (highly nontrivial) problem of quantizing both classical phase spaces in a reasonable way whence one has quantum observable algebras $\mathcal{A}$ and $\mathcal{A}_{\text{red}}$. Then, of course, a relation between $*\text{-}\mathsf{Rep}(\mathcal{A})$ and $*\text{-}\mathsf{Rep}(\mathcal{A}_{\text{red}})$ is of

particular physical interest although one cannot expect both representation theories to be equivalent: $\mathcal{A}$ contains additional but physically irrelevant structure leading to possibly "more" representations.

To get a clear picture of the Morita theory, one has to extend the notion of $*$-representations to the following framework: for an auxiliary $*$-algebra $\mathcal{D}$ over $\mathsf{C}$, one defines a pre-Hilbert right $\mathcal{D}$-module to be a right $\mathcal{D}$-module $\mathcal{H}$ together with a $\mathsf{C}$-sesquilinear $\mathcal{D}$-valued inner product $\langle \cdot , \cdot \rangle : \mathcal{H} \times \mathcal{H} \to \mathcal{D}$ such that $\langle \phi, \psi \rangle^*$ and $\langle \phi, \psi \cdot d \rangle = \langle \phi, \psi \rangle d$ for $d \in \mathcal{D}$ and such that $\langle \cdot , \cdot \rangle$ is completely positive. This means $(\langle \phi_i, \phi_j \rangle) \in M_n(\mathcal{D})^+$ for all $\phi_1, \ldots, \phi_n$, where, in general, an algebra element $a \in \mathcal{A}$ is called positive, $a \in \mathcal{A}^+$, if $\omega(a) \geq 0$ for all positive linear functionals $\omega : \mathcal{A} \to \mathsf{C}$.

Then one defines $\mathfrak{B}(\mathcal{H})$ analogously as for pre-Hilbert spaces leading to a definition of a $*$-representation $\pi$ of $\mathcal{A}$ on a pre-Hilbert right $\mathcal{D}$-module $\mathcal{H}$. The corresponding category of $*$-representations is denoted by $*\text{-}\mathsf{Rep}_{\mathcal{D}}(\mathcal{A})$. Clearly, elements in $*\text{-}\mathsf{Rep}_{\mathcal{D}}(\mathcal{A})$ are in particular $(\mathcal{A}, \mathcal{D})$-bimodules.

The advantage is that now one has a tensor product $\hat{\otimes}$ taking care of the inner products as well. For $*$-algebras $\mathcal{A}, \mathcal{B}, \mathcal{C}$, one has a functor

$$\hat{\otimes} : *\text{-}\mathsf{Rep}_{\mathcal{B}}(\mathcal{C}) \times *\text{-}\mathsf{Rep}_{\mathcal{A}}(\mathcal{B}) \longrightarrow *\text{-}\mathsf{Rep}_{\mathcal{A}}(\mathcal{C}) \quad [23]$$

which, on objects, is essentially given by $\otimes_{\mathcal{B}}$. In fact, for $\mathcal{F} \in *\text{-}\mathsf{Rep}_{\mathcal{B}}(\mathcal{C})$ and $\mathcal{E} \in *\text{-}\mathsf{Rep}_{\mathcal{A}}(\mathcal{B})$, one defines on the $(\mathcal{C}, \mathcal{A})$-bimodule $\mathcal{F} \otimes_{\mathcal{B}} \mathcal{E}$ an $\mathcal{A}$-valued inner product by $\langle x \otimes \phi, y \otimes \psi \rangle = \langle \phi, \langle x, y \rangle \cdot \psi \rangle$, which turns out to be well defined and completely positive again. Then $\mathcal{F} \hat{\otimes} \mathcal{E}$ is $\mathcal{F} \otimes_{\mathcal{B}} \mathcal{E}$ equipped with this inner product modulo its possibly nonempty degeneracy space.

By fixing one of the arguments of $\hat{\otimes}$, one obtains the functor of Rieffel induction of $*$-representations

$$\mathsf{R}_{\mathcal{E}} : *\text{-}\mathsf{Rep}_{\mathcal{D}}(\mathcal{A}) \longrightarrow *\text{-}\mathsf{Rep}_{\mathcal{D}}(\mathcal{B}) \quad [24]$$

where $\mathcal{E} \in *\text{-}\mathsf{Rep}_{\mathcal{A}}(\mathcal{B})$ is fixed and $\mathsf{R}_{\mathcal{E}}(\mathcal{H}) = \mathcal{E} \hat{\otimes} \mathcal{H}$ for $\mathcal{H} \in *\text{-}\mathsf{Rep}_{\mathcal{D}}(\mathcal{A})$.

The idea of strong Morita equivalence is then to search for such bimodules $\mathcal{E}$ where $\mathsf{R}_{\mathcal{E}}$ gives an equivalence of categories. In detail, this is accomplished by the following definition, where, for simplicity, only unital $*$-algebras are considered.

**Definition 6** A $(\mathcal{B}, \mathcal{A})$-bimodule $\mathcal{E}$ is called a strong Morita equivalence bimodule if it is equipped with completely positive inner products $\langle \cdot , \cdot \rangle_{\mathcal{A}}$ and $\langle \cdot , \cdot \rangle_{\mathcal{B}}$ such that both inner products are full, in the sense that

$$\mathsf{C}\text{-span}\{\langle x, y \rangle_{\mathcal{A}} | x, y \in \mathcal{E}\} = \mathcal{A} \quad [25]$$

and analogously for $\langle \cdot , \cdot \rangle_{\mathcal{B}}$, and compatible, in the sense that

$$\langle b \cdot x, y \rangle_{\mathcal{A}} = \langle x, b^* \cdot y \rangle_{\mathcal{A}}, \quad \langle x \cdot a, y \rangle_{\mathcal{B}} = \langle x, y \cdot a^* \rangle_{\mathcal{B}} \quad [26]$$

$$\langle x, y \rangle_{\mathcal{B}} \cdot z = x \cdot \langle y, z \rangle_{\mathcal{A}} \quad [27]$$

In this case, $\mathcal{A}$ and $\mathcal{B}$ are called strongly Morita equivalent.

It turns out that this is indeed an equivalence relation and that strong Morita equivalence implies the equivalence of the representation theories:

**Theorem 7** *For unital $*$-algebras over $\mathsf{C}$, strong Morita equivalence is an equivalence relation.*

**Theorem 8** *If $\mathcal{E}$ is a strong Morita equivalence bimodule, then $\mathsf{R}_{\mathcal{E}}$ as in [24] is an equivalence of categories.*

**Example 9** The fundamental example in Morita theory is that a unital $*$-algebra $\mathcal{A}$ is strongly Morita equivalent to the matrices $M_n(\mathcal{A})$ via the $(M_n(\mathcal{A}), \mathcal{A})$-bimodule $\mathcal{A}^n$ where the inner product is $\langle x, y \rangle_{\mathcal{A}} = \sum_{i=1}^n x_i^* y_i$ and $\langle \cdot , \cdot \rangle_{M_n(\mathcal{A})}$ is uniquely determined by the compatibility condition [27].

An efficient way to encode the whole Morita theory of unital $*$-algebras over $\mathsf{C}$ is to collect all strong Morita equivalence bimodules modulo isometric isomorphisms of bimodules. Then the tensor product $\hat{\otimes}$ makes this into a "large" groupoid whose units are the $*$-algebras themselves. This so-called Picard groupoid $\mathsf{Pic}$ then encodes everything one can say about strong Morita equivalence. In particular, the orbits of this groupoid are precisely the strong Morita equivalence classes of $*$-algebras. The isotropy groups are the Picard groups $\mathsf{Pic}(\mathcal{A})$ which generalize the (outer) automorphism groups.

## Strong Morita Equivalence of Star Products

This section considers star products from the viewpoint of the Morita equivalence. Here one can show that for $\mathcal{A} = (C^\infty(M)[[\lambda]], \star)$, the possible candidates of equivalence bimodules are formal power series of sections $\Gamma^\infty(E)[[\lambda]]$ of vector bundles $E \to M$. This follows as, on the one hand, strong Morita equivalence is compatible with the classical limit $\lambda = 0$ in the sense that it implies strong Morita equivalence of the classical limits. On the other hand, any (classical or quantum) equivalence bimodule is finitely generated and projective as right $\mathcal{A}$-module. Thus, by the Serre–Swan theorem one obtains the sections of a vector bundle in the

classical limit. Now one can show that every vector bundle can uniquely (up to equivalence) be deformed such that $\Gamma^\infty(E)[[\lambda]]$ becomes a right $\mathcal{A}$-module. Thus, the only thing to be computed is which deformation $\star'$ is induced by this deformation of $E$ for the endomorphisms $\Gamma^\infty(\mathsf{End}(E))[[\lambda]]$, since one can show that then the result will always be a strong Morita equivalence bimodule. The inner products come from deformations of a Hermitian fiber metric on $E$.

Since every vector bundle $E \to M$ can be deformed in this manner in an essentially unique way, we arrive at a general global construction of a noncommutative field theory where the fields are sections of $E$ endowed with a deformed bimodule structure. In the case where $M$ is even a symplectic manifold, a simple extension of Fedosov's construction of a star product $\star$ gives a rather explicit formula for the deformed bimodule structure of $\Gamma^\infty(E)[[\lambda]]$ including a construction of the deformation $(\Gamma^\infty(\mathsf{End}(E))[[\lambda]], \star')$ which acts from the left. As usual in Fedosov's approach, the construction depends functorially on the choice of a connection $\nabla^E$ for $E$.

Returning to the question of strong Morita equivalence of star products, we see that the vector bundle $E$ has to be a line bundle $L$ since only in this case we have $\Gamma^\infty(\mathsf{End}(E)) \cong C^\infty(M)$. Since the deformation of the Hermitian fiber metric is always possible and since two equivalent Hermitian star products are always $*$-equivalent, one can show that strong Morita equivalence is already implied by ring-theoretic Morita equivalence (the converse is true in general).

**Theorem 10** *Star products are strongly Morita equivalent if and only if they are Morita equivalent.*

An analogous statement holds for $C^*$-algebras, known as Beer's theorem (1982).

In the symplectic case, the characteristic class $c(\star')$ of the induced star product $\star'$ can be computed explicitly leading to the following classification by Bursztyn and Waldmann (2002):

**Theorem 11** *Let $\star, \star'$ be star products on a symplectic manifold $M$. Then $\star'$ is (strongly) Morita equivalent to $\star$ if and only if there exists a symplectomorphism $\psi$ such that*

$$\psi^* c(\star') - c(\star) \in 2\pi \mathrm{i} \mathrm{H}^2_{\mathrm{deRham}}(M, \mathbb{Z}) \qquad [28]$$

A similar result in the general Poisson case was given by Jurčo, Schupp, and Wess (2002) based on Kontsevich's formality theorem. This approach is motivated by a careful investigation of noncommutative (scalar) field theories.

Finally, it is worth mentioning that [28] has a very simple physical interpretation. Consider again a cotangent bundle $T^*Q$ with a topologically nontrivial configuration space $Q$, for example, $\mathbb{R}^3 \setminus \{0\}$. Then there is a canonical Weyl-type star product $\star_{\mathrm{Weyl}}$ depending on the choice of a connection $\nabla$ and an integration density $\mu > 0$, generalizing [7] to a curved situation. Now let $B$ be a magnetic field, modeled as a closed 2-form on $Q$. Minimal coupling leads to a new star product $\star^B_{\mathrm{Weyl}}$ describing an electrically charged particle moving in $Q$ in the external field $B$. Then the two star products $\star_{\mathrm{Weyl}}$ and $\star^B_{\mathrm{Weyl}}$ are (strongly) Morita equivalent if and only if the magnetic field satisfies Dirac's integrality condition for the (possibly nontrivial) magnetic charges described by $B$. Thus, Dirac's condition is responsible for the very strong statement that the quantizations with and without magnetic field are Morita equivalent. In particular, the $*$-representation theories of $\star_{\mathrm{Weyl}}$ and $\star^B_{\mathrm{Weyl}}$ are equivalent. Even more specifically, using $B$ to construct a line bundle $L \to Q$ one obtains the result that Dirac's $*$-representation of $\star^B_{\mathrm{Weyl}}$ on $\Gamma^\infty_0(L)[[\lambda]]$ is precisely the Rieffel induction of the Schrödinger representation of $\star_{\mathrm{Weyl}}$ on $C^\infty_0(Q)[[\lambda]]$.

*See also:* Aharonov–Bohm Effect; Algebraic Approach to Quantum Field Theory; Deformation Quantization; Deformation Theory; Deformations of the Poisson Bracket on a Symplectic Manifold; Fedosov Quantization.

## Further Reading


Bayen F, Flato M, Frønsdal C, Lichnerowicz A, and Sternheimer D (1978) Deformation theory and quantization. *Annals of Physics* 111: 61–151.

Bertelson M, Cahen M, and Gutt S (1997) Equivalence of star products. *Class. Quant. Grav.* 14: A93–A107.

Bieliavsky P, Dito G, Maeda Y, and Waldmann S (2002) The deformation quantization homepage. http://idefix.physik.uni-freiburg.de/ ~star/en/index.html (regularly updated online bibliography and short introductory articles).

Bordemann M and Waldmann S (1998) Formal GNS construction and states in deformation quantization. *Communications in Mathematical Physics* 195: 549–583.

Bordemann M, Neumaier N, Pflaum MJ, and Waldmann S (2003) On representations of star product algebras over cotangent spaces on Hermitian line bundles. *Journal of Functional Analysis* 199: 1–47.

Bursztyn H and Waldmann S (2001) Algebraic Rieffel induction, formal Morita equivalence and applications to deformation quantization. *Journal of Geometrical Physics* 37: 307–364.

Bursztyn H and Waldmann S (2002) The characteristic classes of Morita equivalent star products on symplectic manifolds. *Communications in Mathematical Physics* 228: 103–121.

Deligne P (1995) Déformations de l'algèbre des fonctions d'une variété symplectique: comparaison entre Fedosov et DeWilde, Lecomte. *Sel. Math. New Series* 1(4): 667–697.

DeWilde M and Lecomte PBA (1983) Existence of star-products and of formal deformations of the Poisson Lie algebra of arbitrary symplectic manifolds. *Letters in Mathematical Physics* 7: 487–496.

Dito G and Sternheimer D (2002) Deformation quantization: genesis, developments and metamorphoses. In: Halbout G (ed.) *Deformation Quantization*, IRMA Lectures in Mathematics and Theoretical Physics, vol. 1, pp. 9–54. Berlin: Walter de Gruyter.

Fedosov BV (1986) Quantization and the index. *Soviet Physics Doklady* 31(11): 877–878.

Gutt S (2000) Variations on deformation quantization. In: Dito G and Sternheimer D (eds.) *Conférence Moshé Flato 1999. Quantization, Deformations, and Symmetries*, Mathematical Physics Studies, vol. 21, pp. 217–254. Dordrecht: Kluwer Academic.

Jurco B, Schupp P, and Wess J (2002) Noncommutative line bundles and Morita equivalence. *Letters of Mathematical Physics* 61: 171–186.

Kontsevich M (2003) Deformation quantization of Poisson manifolds. *Letters in Mathematical Physics* 66: 157–216.

Nest R and Tsygan B (1995) Algebraic index theorem. *Communications in Mathematical Physics* 172: 223–262.

Omori H, Maeda Y, and Yoshioka A (1991) Weyl manifolds and deformation quantization. *Advanced Mathematics* 85: 224–255.

Waldmann S (2002) On the representation theory of deformation quantization. In: Halbout G (ed.) *Deformation Quantization*, IRMA Lectures in Mathematics and Theoretical Physics, vol. 1, pp. 107–133. Berlin: Walter de Gruyter.

Waldmann S (2005) States and representation theory in deformation quantization. *Reviews of Mathematical Physics* 17: 15–75.

Weinstein A and Xu P (1998) Hochschild cohomology and characteristic classes for star-products. In: Khovanskij A, Varchenko A, and Vassiliev V (eds.) *Geometry of Differential Equations*, pp. 177–194. Dedicated to VI Arnol'd on the occasion of his 60th birthday. Providence, RI: American Mathematical Society.


# Deformation Theory

**M J Pflaum**, Johann Wolfgang Goethe-Universität, Frankfurt, Germany



## Introduction and Historical Remarks

In mathematical deformation theory one studies how an object in a certain category of spaces can be varied as a function of the points of a parameter space. In other words, deformation theory thus deals with the structure of families of objects like varieties, singularities, vector bundles, coherent sheaves, algebras, or differentiable maps. Deformation problems appear in various areas of mathematics, in particular in algebra, algebraic and analytic geometry, and mathematical physics. According to Deligne, there is a common philosophy behind all deformation problems in characteristic zero. It is the goal of this survey to explain this point of view. Moreover, we will provide several examples with relevance for mathematical physics.

Historically, modern deformation theory has its roots in the work of Grothendieck, Artin, Quillen, Schlessinger, Kodaira–Spencer, Kuranishi, Deligne, Grauert, Gerstenhaber, and Arnol'd. The application of deformation methods to quantization theory goes back to Bayen–Flato–Fronsdal–Lichnerowicz–Sternheimer, and has led to the concept of a star product on symplectic and Poisson manifolds. The existence of such star products has been proved by de Wilde–Lecomte and Fedosov for symplectic and by Kontsevich for Poisson manifolds.

Recently, Fukaya and Kontsevich have found a far-reaching connection between general deformation theory, the theory of moduli, and mirror symmetry. Thus, deformation theory comes back to its origins, which lie in the desire to construct moduli spaces. Briefly, a moduli problem can be described as the attempt to collect all isomorphism classes of spaces of a certain type into one single object, the moduli space, and then to study its geometric and analytic properties. The observations by Fukaya and Kontsevich have led to new insight into the algebraic geometry of mirror varieties and their application to string theory.

## Basic Definitions and Examples

Deformation theory is based on the notion of a ringed space, so we briefly recall its definition.

**Definition 1** Let $\Bbbk$ be a field. By a $\Bbbk$-ringed space one understands a topological space $X$ together with a sheaf $\mathcal{A}$ of unital $\Bbbk$-algebras on $X$. The sheaf $\mathcal{A}$ will be called the structure sheaf of the ringed space. In case each of the stalks $\mathcal{A}_x, x \in X$, is a local algebra, that is, has a unique maximal ideal $\mathfrak{m}_x$, one calls $(X, \mathcal{A})$ a locally $\Bbbk$-ringed space. Likewise, one defines a commutative $\Bbbk$-ringed space as a ringed space such that the stalks of the structure sheaf are all commutative.

Given two $\Bbbk$-ringed spaces $(X, \mathcal{A})$ and $(Y, \mathcal{B})$, a morphism from $(X, \mathcal{A})$ to $(Y, \mathcal{B})$ is a pair $(f, \varphi)$, where

$f: X \to Y$ is a continuous mapping and $\varphi: f^{-1}\mathcal{B} \to \mathcal{A}$ a morphism of sheaves of algebras. This means in particular that for every point $x \in X$ there is a homomorphism of algebras $\varphi_x: \mathcal{B}_{f(x)} \to \mathcal{A}_x$ induced by $\varphi$. Under the assumption that both ringed spaces are local, $(f, \varphi)$ is called a morphism of locally ringed spaces, if each $\varphi_x$ is a homomorphism of local $\Bbbk$-algebras, that is, maps the maximal ideal of $\mathcal{B}_{f(x)}$ to the one of $\mathcal{A}_x$.

Clearly, $\Bbbk$-ringed spaces (resp. locally or commutative $\Bbbk$-ringed spaces) together with their morphisms form a category. The following is a list of examples of ringed spaces, in particular of those which will be needed later.

**Example 2**

(i) Denote by $\mathcal{C}^\infty$ the sheaf of smooth functions on $\mathbb{R}^n$, by $\mathcal{C}^\omega$ the sheaf of real analytic functions, and let $\mathcal{O}$ be the sheaf of holomorphic functions on $\mathbb{C}^n$. Then $(\mathbb{R}^n, \mathcal{C}^\infty)$, $(\mathbb{R}^n, \mathcal{C}^\omega)$, and $(\mathbb{C}^n, \mathcal{O})$ are ringed spaces over $\mathbb{R}$ resp. $\mathbb{C}$.

(ii) A differentiable manifold of dimension $n$ can be understood as a locally $\mathbb{R}$-ringed space $(M, \mathcal{C}^\infty_M)$ which locally is isomorphic to $(\mathbb{R}^n, \mathcal{C}^\infty)$. Likewise, a real analytic manifold is a ringed space $(M, \mathcal{C}^\omega_M)$ which locally can be modeled by $(\mathbb{R}^n, \mathcal{C}^\omega)$, and a complex manifold is an $(M, \mathcal{O}_M)$ which locally looks like $(\mathbb{C}^n, \mathcal{O})$.

(iii) Let $D$ be a domain in $\mathbb{C}^n$, and $\mathcal{J}$ an ideal sheaf in $\mathcal{O}_D$ of finite type, which means that $\mathcal{J}$ is locally finitely generated over $\mathcal{O}_D$. Let $Y$ be the support of the quotient sheaf $\mathcal{O}_D/\mathcal{J}$. The pair $(Y, \mathcal{O}_Y)$, where $\mathcal{O}_Y$ denotes the restriction of $\mathcal{O}_D/\mathcal{J}$ to $Y$, then is a ringed space, called a complex model space. A complex space now is a ringed space $(X, \mathcal{O}_X)$ which locally looks like a complex model space (cf. Grauert and Remmert 1984).

(iv) Let $\Bbbk$ be an algebraically closed field, and $\mathbb{A}^n$ the affine space over $\Bbbk$ of dimension $n$. Then $\mathbb{A}^n$, together with the sheaf of regular functions, is a ringed space.

(v) Given a ring $A$, its spectrum Spec $A$ together with the sheaf of regular functions $\mathcal{O}_A$ forms a ringed space (cf. (Hartshorne (1997), section II.2)). One calls (Spec $A, \mathcal{O}_A$) an affine scheme. More generally, a scheme is a ringed space $(X, \mathcal{O}_X)$ which locally can be modeled by affine schemes.

(vi) Finally, if $A$ is a local $\Bbbk$-algebra, the pair $(*, A)$ can be understood as a locally ringed space. With $A$ the algebra of formal power series $\Bbbk[[t]]$ over one variable $t$, this example plays an important role in the theory of formal deformations of algebras.

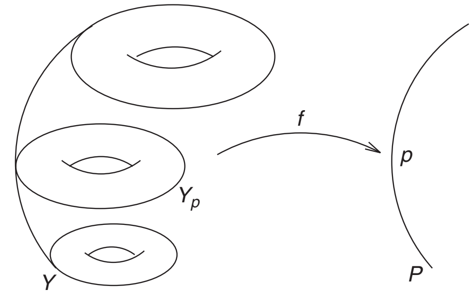



**Figure 1** A fibered space.

**Definition 3** A morphism $(f, \varphi): (Y, \mathcal{B}) \to (P, \mathcal{S})$ of ringed spaces is called fibered, if the following conditions are fulfilled:

(i) $(P, \mathcal{S})$ is a commutative locally ringed space;
(ii) $f: Y \to P$ is surjective; and
(iii) $\varphi_y: \mathcal{S}_{f(y)} \to \mathcal{B}_y$ maps $\mathcal{S}_{f(y)}$ into the center of $\mathcal{B}_y$ for each $y \in Y$.

The *fiber* of $(f, \varphi)$ over a point $p \in P$ then is the ringed space $(Y_p, \mathcal{B}_p)$ defined by

$$Y_p = f^{-1}(p), \qquad \mathcal{B}_p = \mathcal{B}_{|f^{-1}(p)}/\mathfrak{m}_p\mathcal{B}_{|f^{-1}(p)}$$

where $\mathfrak{m}_p$ is the maximal ideal of $\mathcal{S}_p$ which acts on $\mathcal{B}_{|f^{-1}(p)}$ via $\varphi$.

A fibered morphism of ringed spaces can be pictured in Figure 1.

Additionally to this intuitive picture, conditions (i)–(iii) imply that the stalks $\mathcal{B}_y$ are central extensions of $\mathcal{B}_y/\mathfrak{m}_{f(y)}\mathcal{B}_y$ by $\mathcal{S}_{f(y)}$.

**Definition 4** Let $(P, \mathcal{S})$ be a commutative locally ringed space over a field $\Bbbk$ with $P$ connected, let $*$ be a fixed point in $P$, and $(X, \mathcal{A})$ a $\Bbbk$-ringed space. A deformation of $(X, \mathcal{A})$ over the parameter space $(P, \mathcal{S})$ with distinguished point $*$ then is a fibered morphism $(f, \varphi): (Y, \mathcal{B}) \to (P, \mathcal{S})$ over $\Bbbk$ together with an isomorphism $(i, \iota): (X, \mathcal{A}) \to (Y_*, \mathcal{B}_*)$ such that for all $p \in P$ and $y \in f^{-1}(p)$ the homomorphism $\varphi_y: \mathcal{S}_p \to \mathcal{B}_y$ is flat.

The condition of flatness in the definition of a deformation serves as a substitute for "local triviality" and works also in the presence of singularities. (see Palamodov (1990), section 3) for a discussion of this point.

In the remainder of this section, we provide a list of some of the most important deformation problems in mathematics, and show how these can be formulated within the above language.

### Products of $\Bbbk$-Ringed Spaces

Let $(X, \mathcal{A})$ be any $\Bbbk$-ringed space and $(P, \mathcal{S})$ a $\Bbbk$-scheme. For any closed point $* \in P$, the product

$(X \times P, \mathcal{B}) = (X, \mathcal{A}) \times_{\Bbbk} (P, \mathcal{S})$ then is a flat deformation of $(X, \mathcal{A})$ with distinguished point $*$. This can be seen easily from the fact that $\mathcal{B}_{(x,p)} = \mathcal{A}_x \otimes_{\Bbbk} \mathcal{S}_p$ for every $x \in X$ and $p \in P$.

### Families of Matrices as Deformations

Let $(P, \mathcal{O}_P)$ be a complex space with distinguished point $*$ and $A_P : P \to \mathrm{Mat}(n \times n, \mathbb{C})$ a holomorphic family of complex $n \times n$ matrices over $P$. By the following construction, $A_P$ can be understood as a deformation, more precisely as a deformation of the matrix $A := A_P(*)$. Let $Y$ be the graph of $A_P$ in the product space $P \times \mathrm{Mat}(n \times n, \mathbb{C})$ and $f : Y \to P$ be the restriction of the projection onto the first coordinate. Define the sheaf $\mathcal{B}$ as the inverse image sheaf $f^{-1}\mathcal{S}$, and let $\varphi$ be the sheaf morphism which for every $y \in Y$ is induced by the identity map $\varphi_y : \mathcal{S}_{f(y)} \to \mathcal{B}_y := \mathcal{S}_{f(y)}$. It is then immediately clear that $(f, \varphi)$ is a deformation of the fiber $f^{-1}(*)$ and that this fiber coincides with the matrix $A$.

Now let $A$ be an arbitrary complex $n \times n$-matrix, and choose a $\mathrm{GL}(n, \mathbb{C})$-slice through $A$, that is, a submanifold $P$ containing $A$ which is transversal to the $\mathrm{GL}(n, \mathbb{C})$-orbit through $A$. Hereby, it is assumed that $\mathrm{GL}(n, \mathbb{C})$ acts by the adjoint action on $\mathrm{Mat}(n \times n, \mathbb{C})$. The family $A_P$ given by the canonical embedding $P \hookrightarrow \mathrm{Mat}(n \times n, \mathbb{C})$ now is a deformation of $A$. The germ of this deformation at $*$ is versal in the sense defined in the next section.

### Deformation of a Scheme à la Grothendieck

Assume that $(P, \mathcal{S})$ is a connected scheme over $\Bbbk$. A deformation of a scheme $(X, \mathcal{A})$ then is a deformation $(f, \varphi) : (Y, \mathcal{B}) \to (P, \mathcal{S})$ in the sense defined above, together with the requirement that $f : Y \to P$ is a proper map, that is, $f^{-1}(K)$ is compact for every compact $K \subset P$. As a particular example, consider the $\Bbbk$-scheme $Y = \mathrm{Spec}\, \Bbbk[x, y, t]/(xy - t)$. It gives rise to a fibration $Y \to \mathrm{Spec}\, \Bbbk[t]$, whose fibers $Y_a$ with $a \in \Bbbk$ are hyperbolas $xy = a$, when $a \neq 0$, and consist of the two axes $x = 0$ and $y = 0$, when $a = 0$. For $\Bbbk = \mathbb{R}$, this deformation can be illustrated as in Figure 2.

For further information on this and similar examples, see Hartshorne (1977), in particular example 3.3.2.

### Deformation of a Complex Space

According to Grothendieck, one understands by a deformation of a complex space $(X, \mathcal{A})$ a morphism of complex spaces $(f, \varphi) : (Y, \mathcal{B}) \to (P, \mathcal{S})$ which is both a proper flat morphism of complex spaces and a deformation of $(X, \mathcal{A})$ as a ringed space. In case $(X, \mathcal{A})$ and $(P, \mathcal{S})$ are complex manifolds and if $P$ is

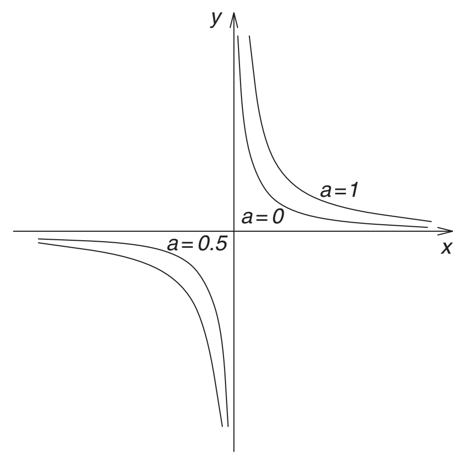



**Figure 2** Deformation of the coordinate axes.

connected, each of the fibers $Y_p$ is a compact complex manifold. Moreover, the family $(Y_p)_{p \in P}$ then is a family of compact complex manifolds in the sense of Kodaira–Spencer (cf. Palamodov (1990)).

### Deformation of Singularities

Let $p$ be a point of some $\mathbb{C}^n$. Two complex spaces $(X, \mathcal{O}_X) \subset (\mathbb{C}^n, \mathcal{O})$ and $(X', \mathcal{O}_{X'}) \subset (\mathbb{C}^n, \mathcal{O})$ with $x \in X \cap X'$ are then called germ equivalent at $x$ if there exists an open neighborhood $U \in \mathbb{C}^n$ of $x$ such that $X \cap U = X' \cap U$. Obviously, germ equivalence at $x$ is an equivalence relation indeed. We denote the equivalence class of $X$ by $[X]_x$. Clearly, if $[X]_x = [X']_x$, then one has $\mathcal{O}_{X,x} = \mathcal{O}_{X',x}$ for the stalks at $x$. By a singularity one understands a pair $([X]_x, \mathcal{O}_{X,x})$. In the literature, such a singularity is often denoted by $(X, x)$. The singularity $(X, x)$ is called nonsingular or regular if $\mathcal{O}_{X,x}$ is isomorphic to an algebra of convergent power series $\mathbb{C}\{z_1, \ldots, z_d\}$. A deformation of a complex singularity $(X, x)$ over a complex germ $(P, *)$ is a morphism of ringed spaces $([Y]_x, \mathcal{O}_{Y,x}) \to ([P]_*, \mathcal{O}_{P,*})$ which is induced by a holomorphic map and which is a deformation of $([X]_x, \mathcal{O}_{X,x})$ as a ringed space. See Artin (1976) and the overview article by Greuel (1992) for further details and a variety of examples.

### First-Order Deformation of Algebras

Consider a $\Bbbk$-algebra $A$ and the truncated polynomial algebra $S = \Bbbk[\varepsilon]/\varepsilon^2\Bbbk[\varepsilon]$. Furthermore, let $\alpha : A \times A \to A$ be a Hochschild 2-cocycle of $A$; in other words, assume that the relation

$$a_1\alpha(a_2, a_3) - \alpha(a_1a_2, a_3) + \alpha(a_1, a_2a_3) - \alpha(a_1, a_2)a_3 = 0 \qquad [1]$$

holds for all $a_1, a_2, a_3 \in A$. Then one can define a new $k$-algebra $B$, whose underlying linear structure

is isomorphic to $A \otimes_{\mathbb{k}} S$ and whose product is given by the following construction: any element $b \in B$ can be written uniquely in the form $b = a_0 + a_1\varepsilon$, with $a_0, a_1 \in A$. Then the product of $b = a_0 + a_1\varepsilon \in B$ and $b' = a_0' + a_1'\varepsilon \in B$ is given by

$$b \cdot b' = a_0 a_0' + \left[\alpha(a_0, a_0') + a_0 a_1' + a_1 a_0'\right]\varepsilon \qquad [2]$$

By condition [1], this product is associative. One thus obtains a flat deformation $\Delta: S \to B$ of the algebra $A$ and calls it the first-order or infinitesimal deformation of $A$ along the Hochschild cocycle $\alpha$. For further information on this and the connection between deformation theory and Hochschild cohomology, see the overview article by Gerstenhaber and Schack (1986).

### Formal Deformation of an Algebra

Let us generalize the preceding example and explain the concept of a formal deformation of an algebra by Gerstenhaber. Assume again $A$ to be an arbitrary $\mathbb{k}$-algebra and choose bilinear maps $\alpha_n : A \times A \to A$ for $n \in \mathbb{N}$ such that $\alpha_0$ is the product on $A$ and $\alpha_1$ is a Hochschild cocycle. Furthermore, let $S$ be the algebra $\mathbb{k}[[t]]$ of formal power series in one variable over $\mathbb{k}$. Then define on the linear space $B = A[[t]]$ of formal power series in one variable with coefficients in $A$ the following bilinear map:

$$\star : B \times B \to B$$
$$\left(\sum_{n\in\mathbb{N}} a_n t^n, \sum_{n\in\mathbb{N}} b_n t^n\right) \mapsto \sum_{n\in\mathbb{N}} \sum_{\substack{k,l,m\in\mathbb{N}\\ k+l+m=n}} \alpha_m(a_k, b_l) t^n \qquad [3]$$

If $B$ together with $\star$ becomes a $\mathbb{k}$-algebra or, in other words, if $\star$ is associative, one can easily see that it gives a flat deformation of $A$ over $S = \mathbb{k}[[t]]$. In that case, one says that $B$ is a formal deformation of $A$ by the family $(\alpha_n)_{n\in\mathbb{N}}$. Contrarily to the preceding example, there might not exist for every Hochschild cocycle $\alpha$ on $A$ a formal deformation $B$ of $A$ defined by a family $(\alpha_n)_{n\in\mathbb{N}}$ such that $\alpha_1 = \alpha$. In case it exists, we will say that the deformation $B$ of $A$ is in the direction of $\alpha$. If the third Hochschild cohomology group $\mathrm{H}^3(A, A)$ vanishes, there exists for every Hochschild cocycle $\alpha$ on $A$ a deformation $B$ of $A$ in the direction of $\alpha$ (see again Gerstenhaber and Schack (1986) for further details).

### Formal Deformation Quantization of Symplectic and Poisson Manifolds

Let us consider the last two examples for the case where $A$ is the algebra $\mathcal{C}^\infty(M)$ of smooth functions on a symplectic or Poisson manifold $M$. Then the Poisson bracket $\{,\}$ gives a Hochschild cocycle on $\mathcal{C}^\infty(M)$. There exists a first-order deformation of $\mathcal{C}^\infty(M)$ along $(1/2\mathrm{i})\{,\}$ and, even though $\mathrm{HH}^3(A, A)$ might not always vanish, a deformation quantization of $M$, that means a formal deformation of $\mathcal{C}^\infty(M)$ in the direction of the Poisson bracket $(1/2\mathrm{i})\{,\}$. For the symplectic case, this fact has been proved first by deWilde–Lecomte using methods from Hochschild cohomology theory. A more geometric and intuitive proof has been given by Fedosov (1996). The Poisson case has been settled in the work of Kontsevich (2003) (see also the section "Deformation quantization of Poisson manifolds").

### Quantized Universal Enveloping Algebras According to Drinfeld

A quantized universal enveloping algebra for a complex Lie algebra $\mathfrak{g}$ is a Hopf algebra $A$ over $\mathbb{C}[[t]]$ such that $A$ is a topologically free $\mathbb{C}[[t]]$-module (i.e., $A = (A/tA)[[t]]$ as left $\mathbb{C}[[t]]$-module) and $A/tA$ is the universal enveloping algebra $U\mathfrak{g}$ of $\mathfrak{g}$. Because $A$ is a topologically free $\mathbb{C}[[t]]$-module, $A$ is a flat $\mathbb{C}[[t]]$-module and thus a deformation of $U\mathfrak{g}$ over $\mathbb{C}[[t]]$. See Drinfel'd (1986) and the monograph by Kassel (1995) for further details and examples of quantized universal enveloping algebras.

### Quantum Plane

Consider the tensor algebra $T = \bigoplus_{n\in\mathbb{N}} (\mathbb{R}^2)^{\otimes n}$ of the two-dimensional real vector space $\mathbb{R}^2$, and let $(x, y)$ be the canonical basis of $\mathbb{R}^2$. Then form the tensor product sheaf $\mathcal{T}_{\mathbb{C}^*} = T \otimes_{\mathbb{R}} \mathcal{O}_{\mathbb{C}^*}$ and let $\mathcal{I}_{\mathbb{C}^*}$ be the ideal sheaf in $\mathcal{T}_{\mathbb{C}^*}$ generated by the relation

$$x \otimes y - zy \otimes x = 0 \qquad [4]$$

where $z : \mathbb{C}^* \to \mathbb{C}$ is the identity function. The quotient sheaf $\mathcal{B} = \mathcal{B}_{\mathbb{C}^*} = \mathcal{T}_{\mathbb{C}^*}/\mathcal{I}_{\mathbb{C}^*}$ then is a sheaf of $\mathbb{C}$-algebras and an $\mathcal{O}_{\mathbb{C}^*}$-module. Using eqn [4] now move all occurrences of $x$ in an element of $\mathcal{B}_{\mathbb{C}^*}$ to the right of all $y$'s. Since $1/z$ is an element of $\mathcal{O}(\mathbb{C}^*)$, one can thus show that $\mathcal{B}_{\mathbb{C}^*}$ is a free $\mathcal{O}_{\mathbb{C}^*}$-module. Hence, $\mathcal{B}_{\mathbb{C}^*}$ is flat over $\mathcal{O}_{\mathbb{C}^*}$. Further, it is easy to see that for every $q \in \mathbb{C}^*$ the $\mathbb{C}$-algebra $A_q = \mathcal{B}_q/\mathfrak{m}_q\mathcal{B}_q$ is freely generated by elements $x$, $y$ with relations

$$x \otimes y - qy \otimes x = 0 \qquad [5]$$

We call $A_q$ the q-deformed quantum plane and $B = \mathcal{B}(\mathbb{C}^*)$ the over $\mathbb{C}^*$ universally deformed quantum plane. Altogether, one can interpret $B$ as a deformation of $A_q$ over $\mathbb{C}^*$, in particular as a deformation of $A_1 = T \otimes_{\mathbb{R}} \mathbb{C} = \mathbb{C}[x, y]$, the algebra of complex polynomials in two generators.

In the same way, one can deform function algebras on higher-dimensional vector spaces as well as function algebras on certain Lie groups. In this manner, one obtains the quantum group

$SU_q(2)$ as a deformation of a Hopf algebra of functions on SU(2). See, for example, the work of Faddeev–Reshetikhin–Takhtajan (1990), Manin (1988) and Wess–Zumino (1990) for more information on $q$-deformations of vector spaces, Lie groups, differential calculi, etc.

## Versal Deformations

In this section, and the ones that follow, we consider only germs of deformations, that is, deformations over parameter spaces of the form $(*, S)$. This means in particular that the structure sheaf only consists of its stalk $S$ at $*$, a commutative local $\Bbbk$-algebra. Let us now suppose that the sheaf morphism $\varphi : (Y, \mathcal{B}) \to (*, S)$ (over the canonical map $Y \to *$) is a deformation of the ringed space $(X, \mathcal{A})$ and that $\tau : T \to S$ is a homomorphism of commutative local $\Bbbk$-algebras. Then the sheaf morphism $\tau^*\varphi : \mathcal{B} \otimes_S T \to T$ with $(\tau^*\varphi)_y(t) = 1 \otimes t$ for $y \in Y$ and $t \in T$ is a deformation of $(X, \mathcal{A})$ over the parameter space $(*, T)$. One says that the deformation $\tau^*\varphi$ is induced by the homomorphism $\tau$.

**Definition 5** A deformation $\varphi : (Y, \mathcal{B}) \to S$ of $(X, \mathcal{A})$ is called versal if every (germ of a) deformation of $(X, \mathcal{A})$ is isomorphic to a deformation germ induced by a homomorphism of $\Bbbk$-algebras $\tau : T \to S$. A versal deformation is called universal, if the inducing homomorphism $\tau : T \to S$ is unique, and *miniversal* if $S$ is of minimal dimension.

**Example 6**

(i) In the section "Families of matrices as deformations," the construction of a versal deformation of a complex matrix $A$ has been sketched.
(ii) According to Kuranishi, every compact complex manifold has a versal deformation by an analytic germ. See Kuranishi (1971) for a detailed exposition and the section "The Kodaira–Spencer algebra controlling deformations of compact complex manifolds" for a description of the principal ideas.
(iii) Grauert has shown that for isolated singularities there exists a versal analytic deformation.
(iv) By the work of Douady–Verdier, Grauert, and Palamodov one knows that for every compact complex space there exists a miniversal analytic deformation. One of the essential methods in the existence proof hereby is Palamodov's construction of the cotangent complex (see Palamodov (1990).
(v) Bingener (1987) has further established Palomodov's approach and thus has provided a unified and quite general method for constructing versal deformations in analytic geometry.
(vi) Fialowski–Fuchs have constructed miniversal deformations of Lie algebras.

## Schlessinger's Theorem

According to Grothendieck, spaces in algebraic geometry are represented by functors from a category of commutative rings to the category of sets. In this picture, an affine algebraic variety $X$ over the base field $\Bbbk$ and with coordinate ring $A$ is equivalently described by the functor $\mathrm{Hom}_{\mathrm{alg}}(A, -)$ defined on the category of commutative $\Bbbk$-algebras. As will be shown by examples in the next section, versal deformations are often encoded by functors representing spaces. More precisely, a deformation problem leads to a so-called functor of Artin rings, which means a covariant functor $F$ from the category of (local) Artinian $\Bbbk$-algebras to the category of sets such that the set $F(\Bbbk)$ has exactly one element. The question now arises as to under which conditions the functor $F$ is representable, that is, there exists a commutative $\Bbbk$-algebra $A$ such that $F \cong \mathrm{Hom}_{\mathrm{alg}}(A, -)$. In the work of Schlessinger (1968), the structure of functors of Artin rings has been studied in detail. Moreover, criteria have been established, when such a functor is pro-presentable, which means that it can be represented by a complete local algebra $\hat{A}$, where "completeness" is understood with respect to the $\mathfrak{m}$-adic topology. Because of its importance for deformation theory, we will state Schlessinger's theorem in this section. Before we come to its details, let us recall some notation.

**Definition 7** By an Artinian $\Bbbk$-algebra over a field $\Bbbk$ one understands a commutative $\Bbbk$-algebra $R$ which satisfies the following descending chain condition:

(Dec) Every descending chain $I_1 \supset \cdots \supset I_k \supset I_{k+1} \supset \cdots$ of ideals in $R$ becomes stationary.

Among others, an Artinian algebra $R$ has the following properties:

1. $R$ is Noetherian, that is, it satisfies the ascending chain condition.
2. Every prime ideal in $R$ is maximal.
3. (Chinese remainder theorem) $R$ is isomorphic to a finite product $\Pi_{i=1}^{n} R_i$, where each $R_i$ is a local Artinian algebra.
4. Every maximal ideal $\mathfrak{m}$ of $R$ is nilpotent, that is, $\mathfrak{m}^k = 0$ for some $k \in \mathbb{N}$.
5. Every quotient $R/\mathfrak{m}^k$ with $\mathfrak{m}$ maximal is finite dimensional.

**Definition 8** Assume that $f : B \to A$ is a surjective homomorphism in the category $\Bbbk$-$\mathsf{Alg}_{\mathsf{l,Art}}$ of local Artinian $\Bbbk$-algebras. Then $f$ is called a small extension if ker $f$ is a nonzero principal ideal $(b)$ in $B$ such that $\mathfrak{m}b = (0)$, where $\mathfrak{m}$ is the maximal ideal of $B$.

**Theorem 9** (Schlessinger (1968, theorem 2.11)). *Let F be a functor of Artin rings* (*over the base field* $\Bbbk$). *Assume that* $A' \to A$ *and* $A'' \to A$ *are morphisms in* $\Bbbk$-$\mathsf{Alg}_{\mathsf{l,Art}}$, *and consider the map*

$$F(A' \times_A A'') \to F(A') \times_{F(A)} F(A'') \qquad [6]$$

*Then F is pro-representable if and only if F has the following properties*:

(H1) *The map* [6] *is a surjection whenever* $A'' \to A$ *is a small extension.*
(H2) *The map* [6] *is a bijection*, *when* $A = \Bbbk$ and $A'' = \Bbbk[\varepsilon]$.
(H3) *One has* $\dim_\Bbbk(t_F) < \infty$ *for the tangent space* $t_F := F(\Bbbk[\varepsilon])$.
(H4) *For every small extension* $A' \to A$, *the map*

$$F(A' \times_A A') \to F(A') \times_{F(A)} F(A')$$

*is an isomorphism.*

Suppose that the functor $F$ satisfies conditions (H1)–(H4), and let $\hat{A}$ be an arbitrary complete local $\Bbbk$-algebra. By Yoneda's lemma, every element

$$\xi = \operatorname*{proj\,lim}_{n \in \mathbb{N}} \xi_n \in \hat{A} = \operatorname*{proj\,lim}_{n \in \mathbb{N}} \hat{A}/\mathfrak{m}^n \hat{A}$$

induces a natural transformation

$$\mathrm{Hom}_{\mathrm{alg}}(\hat{A}, -) \to F, \quad \left(u : \hat{A} \to R\right) \mapsto F(u_n)(\xi_n) \quad [7]$$

where $n \in \mathbb{N}$ is chosen large enough such that the homomorphism $u : \hat{A} \to R$ factors through some $u_n : \hat{A}/\mathfrak{m}^n \to R$. This is possible indeed, since $R$ is Artinian. In the course of the proof of Schlessinger's theorem, $\hat{A}$ and the element $\xi \in \hat{A}$ are now constructed in such a way that [7] is an isomorphism.

## Differential Graded Lie Algebras and Deformation Problems

According to a philosophy going back to Deligne "every deformation problem in characteristic zero is controlled by a differential graded Lie algebra, with quasi-isomorphic differential graded Lie algebras giving the same deformation theory" (cf. Goldman and Millson (1988), p. 48). In the following, we will explain the main idea of this concept and apply it to two particular examples.

### Differential Graded Lie Algebras

**Definition 10** By a graded algebra over a field $\Bbbk$ one understands a graded $\Bbbk$-vector space $A^\bullet = \bigoplus_{k \in \mathbb{Z}} A^k$ together with a bilinear map

$$\mu : A^\bullet \times A^\bullet \to A^\bullet, \quad (a, b) \mapsto a \cdot b = \mu(a, b)$$

such that $A^k \cdot A^l \subset A^{k+l}$ for all $k, l \in \mathbb{Z}$. The graded algebra $A^\bullet$ is called associative if $(ab)c = a(bc)$ for all $a, b, c \in A^\bullet$.

A graded subalgebra of $A^\bullet$ is a graded subspace $B^\bullet = \bigoplus_{k \in \mathbb{Z}} B^k \subset A^\bullet$ which is closed under $\mu$, a graded ideal is a graded subalgebra $I^\bullet \subset A^\bullet$ such that $I^\bullet \cdot A^\bullet \subset I^\bullet$ and $A^\bullet \cdot I^\bullet \subset I^\bullet$.

A homomorphism between graded algebras $A^\bullet$ and $B^\bullet$ is a homogeneous map $f : A^\bullet \to B^\bullet$ of degree 0 such that $f(a \cdot b) = f(a) \cdot f(b)$ for all $a, b \in A^\bullet$.

From now on, assume that $\Bbbk$ has characteristic $\neq 2, 3$. A graded Lie algebra then is a graded $\Bbbk$-vector space $\mathfrak{g}^\bullet = \bigoplus_{k \in \mathbb{Z}} \mathfrak{g}^k$ together with a bilinear map

$$[\cdot, \cdot] : \mathfrak{g}^\bullet \times \mathfrak{g}^\bullet \to \mathfrak{g}^\bullet, \quad (a, b) \mapsto [a, b]$$

such that the following axioms hold true:

1. $[\mathfrak{g}^k, \mathfrak{g}^l] \subset \mathfrak{g}^{k+l}$ for all $k, l \in \mathbb{Z}$.
2. $[\xi, \zeta] = -(-1)^{kl}[\zeta, \xi]$ for all $\xi \in \mathfrak{g}^k, \zeta \in \mathfrak{g}^l$.
3. $(-1)^{k_1 k_3}[[\xi_1, \xi_2], \xi_3] + (-1)^{k_2 k_1}[[\xi_2, \xi_3], \xi_1] + (-1)^{k_3 k_2}[[\xi_3, \xi_1], \xi_2] = 0$ for all $\xi_i \in \mathfrak{g}^{k_i}$ with $i = 1, 2, 3$.

By axiom (1), it is clear that a graded Lie algebra is in particular a graded algebra. So the above-defined notions of a graded ideal, homomorphism, etc., apply as well to graded Lie algebras.

**Example 11** Let $A^\bullet = \bigoplus_{k \in \mathbb{Z}} A^k$ be a graded associative algebra. Then $A^\bullet$ becomes a graded Lie algebra with the bracket

$$[a, b] = ab - (-1)^{kl} ba \quad \text{for } a \in A^k \text{ and } b \in A^l$$

The space $A^\bullet$ regarded as a graded Lie algebra is often denoted by $\mathfrak{lie}^\bullet(A^\bullet)$.

**Definition 12** A linear map $D : A^\bullet \to A^\bullet$ defined on a graded algebra $A^\bullet$ is called a *derivation* of *degree l* if

$$D(ab) = (Da)b + (-1)^{kl} a(Db) \quad \text{for all } a \in A^k \text{ and } b \in A^\bullet$$

A graded (Lie) algebra $A^\bullet$ together with a derivation $d$ of degree 1 is called a differential graded (Lie) algebra if $d \circ d = 0$. Then $(A^\bullet, d)$ becomes a cochain complex. Since ker $d$ is a graded

subalgebra of $A^\bullet$ and im $d$ a graded ideal in ker $d$, the cohomology space

$$H^\bullet(A^\bullet, d) = \ker d / \mathrm{im}\, d$$

inherits the structure of a graded (Lie) algebra from $A^\bullet$.

Let $f : A^\bullet \to B^\bullet$ be a homomorphism of differential graded (Lie) algebras $(A^\bullet, d)$ and $(B^\bullet, \partial)$. Assume further that $f$ is a cochain map, that is, that $f \circ d = \partial \circ f$. Then one says that $f$ is quasi-isomorphism or that the differential graded (Lie) algebras $A^\bullet$ and $B^\bullet$ are quasi-isomorphic if the induced homomorphism on the cohomology level $\overline{f} : H^\bullet(A^\bullet, d) \to H^\bullet(B^\bullet, \partial)$ is an isomorphism. Finally, a differential graded (Lie) algebra $(A^\bullet, d)$ is called formal if it is quasi-isomorphic to its cohomology $(H^\bullet(A^\bullet, d), 0)$.

### Maurer–Cartan Equation

Assume that $(\mathfrak{g}^\bullet, [\cdot,\cdot], d)$ is a differential graded Lie algebra over $\mathbb{C}$. Define the space $\mathcal{MC}(\mathfrak{g}^\bullet)$ of solutions of the Maurer–Cartan equation by

$$\mathcal{MC}(\mathfrak{g}^\bullet) := \{\omega \in \mathfrak{g}^1 \mid d\omega - \tfrac{1}{2}[\omega, \omega] = 0\} \qquad [8]$$

In case the differential graded Lie algebra $\mathfrak{g}^\bullet$ is nilpotent, this space naturally possesses a groupoid structure or, in other words, a set of arrows which are all invertible. The reason for this is that, under the assumption of nilpotency, the space $\mathfrak{g}^0$ is equipped with the Campbell–Hausdorff multiplication

$$\mathfrak{g}^0 \times \mathfrak{g}^0 \to \mathfrak{g}^0, \quad (X, Y) \mapsto \log(\exp X, \exp Y)$$

and the group $\mathfrak{g}^0$ acts on $\mathfrak{g}^1$ by the exponential function. More precisely, in this situation one can define for two objects $\alpha, \beta \in \mathcal{MC}(\mathfrak{g}^\bullet)$ the space of arrows $\alpha \to \beta$ as the set of all $\lambda \in \mathfrak{g}^0$ such that $\exp \lambda \cdot \alpha = \beta$.

We have now the means to define for every complex differential graded Lie algebra $\mathfrak{g}^\bullet$ its deformation functor $\mathrm{Def}_{\mathfrak{g}^\bullet}$. This functor maps the category of local Artinian $\mathbb{C}$-algebras to the category of groupoids and is defined on objects as follows:

$$\mathrm{Def}_{\mathfrak{g}^\bullet}(R) := \mathcal{MC}(\mathfrak{g}^\bullet \otimes \mathfrak{m}) \qquad [9]$$

Hereby, $R$ is a complex local Artinian algebra, and $\mathfrak{m}$ its maximal ideal. Note that since $R$ is Artinian, $\mathfrak{g}^\bullet \otimes \mathfrak{m}$ is a nilpotent differential graded Lie algebra, hence $\mathrm{Def}_{\mathfrak{g}^\bullet}(R)$ carries a groupoid structure as constructed above. Clearly, $\mathrm{Def}_{\mathfrak{g}^\bullet}$ is also a functor of Artin rings as defined in the previous section.

With appropriate choices of the differential graded Lie algebra $\mathfrak{g}^\bullet$, essentially all deformation problems from the section "Basic definitions and examples" can be recovered via a functor of the form $\mathrm{Def}_{\mathfrak{g}^\bullet}$. Below, we will show in some detail how this works for two examples, namely the deformation theory of complex manifolds and the deformation quantization of Poisson manifolds. But before we come to this, let us state a result which shows how the deformation functor behaves under quasi-isomorphisms of the underlying differential graded Lie algebra. This result is crucial in a sense that it allows to equivalently describe a deformation problem with controlling $\mathfrak{g}^\bullet$ by any other differential graded Lie algebra within the quasi-isomorphism class of $\mathfrak{g}^\bullet$. So, in particular in the case where the differential graded Lie algebra is formal, one often obtains a direct solution of the deformation problem.

**Theorem 13** (Deligne, Goldman–Millson). *Assume that $f : \mathfrak{g}^\bullet \to \mathfrak{h}^\bullet$ is a quasi-isomorphism of differential graded Lie algebras. For every local Artinian $\mathbb{C}$-algebra $R$ the induced functor $f_* : \mathrm{Def}_{\mathfrak{g}^\bullet}(R) \to \mathrm{Def}_{\mathfrak{h}^\bullet}(R)$ then is an equivalence of groupoids.*

### The Kodaira–Spencer Algebra Controlling Deformations of Compact Complex Manifolds

Let $M$ be a compact complex $n$-dimensional manifold. Recall that then the complexified tangent bundle $T_{\mathbb{C}}M$ has a decomposition into a holomorphic tangent bundle $T^{1,0}M$ and an antiholomorphic tangent bundle $T^{0,1}M$. This leads to a decomposition of the space of complex $n$-forms into the spaces $\Omega^{p,q}M$ of forms on $M$ of type $(p, q)$. More generally, a smooth subbundle $J^{0,1} \subset T_{\mathbb{C}}M$ which induces a decomposition of the form $T_{\mathbb{C}}M = J^{1,0} \oplus J^{0,1}$, where $J^{1,0} := \overline{J^{0,1}}$, is called an almost complex structure on $M$. Clearly, the decomposition of $T_{\mathbb{C}}M$ into the holomorphic and antiholomorphic part is an almost complex structure, and an almost complex structure which is induced by a complex structure is called integrable. Assume that an almost complex structure $J^{0,1}$ is given on $M$ and that it has finite distance to the complex structure on $M$. The latter means that the restriction $\varrho_J^{0,1}$ of the projection $\varrho : T_{\mathbb{C}}M \to T^{0,1}M$ along $T^{1,0}M$ to the subbundle $J^{0,1}$ is an isomorphism. Denote by $\beta$ the inverse of $\varrho_J^{0,1}$, and let $\omega \in \Omega^{0,1}(M, T^{1,0}M)$ be the composition $-\varrho \circ \beta$. One checks immediately that every almost complex structure with finite distance to the complex structure on $M$ is uniquely characterized by a section $\omega \in \Omega^{0,1}(M, T^{1,0}M)$ and that every element of $\Omega^{0,1}(M, T^{1,0}M)$ comes from an almost complex structure on $M$.

As a consequence of the Newlander–Nirenberg theorem, one can now show that the almost

complex structure $J^{0,1}$ resp. $\omega$ is integrable if and only if the equation

$$\overline{\partial}\omega - \tfrac{1}{2}[\omega,\omega] = 0 \qquad [10]$$

is fulfilled. But this is nothing else than the Maurer–Cartan equation in the Kodaira–Spencer differential graded Lie algebra

$$(\mathfrak{L}^\bullet, \overline{\partial}, [\cdot,\cdot]) = \left( \bigoplus_{p\in\mathbb{N}} \Omega^{0,p}(M, T^{1,0}M), \overline{\partial}, [\cdot\,,\cdot] \right)$$

Hereby, $\Omega^{0,p}(M, T^{1,0}M)$ denotes the $T^{1,0}M$-valued differential forms on $M$ of type $(0,p), \overline{\partial}$: $\Omega^{0,p}(M,T^{1,0}M) \to \Omega^{0,p+1}(M,T^{1,0}M)$ the Dolbeault operator, and $[\,\cdot\,,\cdot\,]$ is induced by the Lie bracket of holomorphic vector fields. As a consequence of these considerations, deformations of the complex manifold $M$ can equivalently be described by families $(\omega_p)_{p\in P} \subset \mathfrak{L}^1$ which satisfy eqn [10] and $\omega_* = 0$. Thus, it remains to determine the associated deformation functor $\mathrm{Def}_{\mathfrak{L}^\bullet}$.

According to Schlessinger's theorem, the functor $\mathrm{Def}_{\mathfrak{L}^\bullet}$ is pro-representable. Hence, there exists a local $\mathbb{C}$-algebra $R_{\mathfrak{L}^\bullet}$ complete with respect to the $\mathfrak{m}$-adic topology such that

$$\mathrm{Def}_{\mathfrak{L}^\bullet}(R) = \mathrm{Hom}_{\mathrm{alg}}(R_{\mathfrak{L}^\bullet}, R) \qquad [11]$$

for every local Artinian $\mathbb{C}$-algebra $R$. Moreover, by Artin's theorem, there exists a "convergent" solution of the Maurer–Cartan equation, that is, $R_{\mathfrak{L}^\bullet}$ can be replaced in eqn. [11] by a ring $\overline{R}_{\mathfrak{L}^\bullet}$ representing an analytic germ.

**Theorem 14** (Kodaira–Spencer, Kuranishi). *The ringed space* $(\overline{R}_{\mathfrak{L}^\bullet}, (0))$ *is a miniversal deformation of the complex structure on* $M$.

### Deformation Quantization of Poisson Manifolds

Let $A$ be an associative $\mathbb{k}$-algebra with $\mathrm{char}\,\mathbb{k} = 0$. Put for every integer $k \geq -1$

$$\mathfrak{g}^k := \mathrm{Hom}_{\mathbb{k}}(A^{\otimes(k+1)}, A)$$

Then $\mathfrak{g}^\bullet$ becomes a graded vector space. Let us impose a differential and a bracket on $\mathfrak{g}^\bullet$. The differential is the usual *Hochschild coboundary* $b : \mathfrak{g}^k \to \mathfrak{g}^{k+1}$,

$$\begin{aligned} bf(a_0 \otimes \cdots \otimes a_{k+1}) &:= a_0 f(a_1 \otimes \cdots \otimes a_{k+1}) \\ &\quad + \sum_{i=0}^{k} (-1)^{i+1} f(a_0 \otimes \cdots \otimes a_i a_{i+1} \otimes \cdots \otimes a_{k+1}) \\ &\quad + (-1)^k f(a_0 \otimes \cdots \otimes a_k) a_{k+1} \end{aligned}$$

The bracket is the Gerstenhaber bracket

$$[\cdot,\cdot] : \mathfrak{g}^{k_1} \times \mathfrak{g}^{k_2} \to \mathfrak{g}^{k_1+k_2}$$
$$[f_1, f_2] := f_1 \circ f_2 - (-1)^{k_1 k_2} f_2 \circ f_1$$

where

$$\begin{aligned} &f_1 \circ f_2(a_0 \otimes \cdots \otimes a_{k_1+k_2}) \\ &:= \sum_{i=0}^{k_1} (-1)^{ik_2} f_1\big(a_0 \otimes \cdots \otimes a_{i-1} \otimes f_2(a_i \otimes \cdots \otimes a_{i+k_2}) \\ &\qquad \otimes a_{i+k_2+1} \otimes \cdots \otimes a_{k_1+k_2}\big) \end{aligned}$$

The triple $(\mathfrak{g}^\bullet, b, [\,\cdot\,,\cdot\,])$ then is a differential graded Lie algebra.

Consider the Maurer–Cartan equation $b\gamma - (1/2)[\gamma,\gamma] = 0$ in $\mathfrak{g}^1$. Obviously, it is equivalent to the equality

$$\begin{aligned} &a_0\gamma(a_1,a_2) - \gamma(a_0 a_1, a_2) + \gamma(a_0, a_1 a_2) - \gamma(a_0,a_1)a_2 \\ &\quad = \gamma(\gamma(a_0,a_1),a_2) - \gamma(a_0, \gamma(a_1,a_2)) \\ &\qquad \text{for } a_0, a_1, a_2 \in A \end{aligned} \qquad [12]$$

If one defines now for some $\gamma \in \mathfrak{g}^1$ the bilinear map $m : A \times A \to A$ by $m(a,b) = ab + \gamma(a,b)$, then [12] implies that $m$ is associative if and only if $\gamma$ satisfies the Maurer–Cartan equation.

Let us apply these observations to the case where $A$ is the algebra $\mathcal{C}^\infty(M)[[t]]$ of formal power series in one variable with coefficients in the space of smooth functions on a Poisson manifold $M$. By (a variant of) the theorem of Hochschild–Kostant–Rosenberg and Connes, one knows that in this case the cohomology of $(\mathfrak{g}^\bullet, b)$ is given by formal power series with coefficients in the space $\Gamma^\infty(\Lambda^\bullet TM)$ of antisymmetric vector fields. Now, $\Gamma^\infty(\Lambda^\bullet TM)$ carries a natural Lie algebra bracket as well, namely the Schouten bracket. Thus, one obtains a second differential graded Lie algebra $(\Gamma^\infty(\Lambda^\bullet TM)[[t]], 0, [\,\cdot\,,\cdot\,])$. Unfortunately, the projection onto cohomology $(\mathfrak{g}^\bullet, b) \to \Gamma^\infty(\Lambda^\bullet TM)[[t]]$ does not preserve the natural brackets, hence is not a quasi-isomorphism in the category of differential graded Lie algebras. It has been the fundamental observation by Kontsevich that this defect can be cured as follows.

**Theorem 15** (Kontsevich 2003). *For every Poisson manifold M the differential graded Lie algebra* $(\mathfrak{g}^\bullet, b, [\,\cdot\,,\cdot])$ *is formal in the sense that there exists a quasi-isomorphism* $(\mathfrak{g}^\bullet, b, [\,\cdot\,,\cdot]) \to (\Gamma^\infty(\Lambda^\bullet TM)[[t]], 0, [\,\cdot\,,\cdot])$ *in the category of* $L^\infty$*-algebras.*

Note that the theorem only claims the existence of a quasi-isomorphism in the category of $L^\infty$-algebras or, in other words, in the category of homotopy Lie algebras. This is a notion somewhat weaker than a differential graded Lie algebra, but Theorem 13 also holds in the context of $L^\infty$-algebras.

Since the solutions of the Maurer–Cartan equation in $(\Gamma^\infty(\Lambda^\bullet TM)[[t]], 0, [\cdot\,,\cdot])$ are exactly the formal paths of Poisson bivector fields on $M$, Kontsevich's formality theorem entails:

**Corollary 16** *Every Poisson manifold has a formal deformation quantization.*

*See also:* Deformation Quantization; Deformation Quantization and Representation Theory; Deformations of the Poisson Bracket on a Symplectic Manifold; Fedosov Quantization; Holonomic Quantum Fields; Operads.

## Further Reading


Artin M (1976) *Lectures on Deformations of Singularities*. Lectures on Mathematics and Physics, vol. 54. Bombay: Tata Institute of Fundamental Research. II.

Bayen F, Flato M, Fronsdal C, Lichnerowicz A, and Sternheimer D (1978) Deformation theory and quantization, I and II. *Annals of Physics* 111: 61–151.

Bingener J (1987) *Lokale Modulräume in der analytischen Geometrie*, Band 1 und 2, Aspekte der Mathematik Braunschweig: Vieweg.

Drinfel'd VG (1986) *Quantum Groups*, Proc. of the ICM, 1986, pp. 798–820.

Faddeev–Reshetikhin–Takhtjan (1990) Quantization of Lie groups and Lie algebras. *Leningrad Mathematical Journal* 1(1): 193–225.

Fedosov B (1996) *Deformation Quantization and Index Theory*. Berlin: Akademie-Verlag.

Grauert H and Remmert R (1984) *Coherent Analytic Sheaves*. Grundlehren der Mathematischen Wissenschaften, vol. 265. Berlin: Springer.

Gerstenhaber M and Schack S (1986) Algebraic cohomology and deformation theory. Deformation theory of algebras and structures and applications, Nato Advanced Study Institute, Castelvecchio-Pascoli/Italy. NATO ASI series, C247: 11–264.

Goldman WM and Millson JJ (1988) *The Deformation Theory of Representations of Fundamental Groups of Compact Kähler Manifolds*, Publ. Math. Inst. Hautes Étud. Sci, vol. 67, pp. 43–96.

Greuel G-M (1992) *Deformation und Klassifikation von Singularitäten und Moduln*, Jahresbericht der DMV, Jubiläumstag., 100 Jahre DMV, Bremen/Dtschl. 1990, 177–238.

Hartshorne R (1977) *Algebraic Geometry*. Graduate Texts in Mathematics, vol. 52. Berlin: Springer.

Kassel Ch (1995) *Quantum Groups*, Graduate Texts in Mathematics, vol. 155. New York: Springer.

Kontsevich M (2003) Deformation quantization of Poisson manifolds. *Letters in Mathematical Physics* 66(3): 157–216.

Kuranishi M (1971) *Deformations of Compact Complex Manifolds*, Séminaire de Mathématiques Supérieures. Été 1969 Montreal: Les Presses de l'Université de Montreal.

Manin (1988) *Quantum Groups and Non-Commutative Geometry*. Montreal: Les Publications CRM.

Palamodov VP (1990) Deformation of complex spaces. In: Gindikin SG and Khenkin GM (eds.) *Several Complex Variables IV, Encyclopaedia of Mathematical Sciences*, vol. 10. Berlin: Springer.

Schlessinger M (1968) Functors of Artin rings. *Transactions of the American Mathematical Society* 130: 208–222.

Wess–Zumino (1990) Covariant differential calculus on the quantum hyperplane. *Nuclear Physics B, Proceedings Supplement* 18B: 302–312.


# Deformations of the Poisson Bracket on a Symplectic Manifold


**S Gutt**, Université Libre de Bruxelles, Brussels, Belgium

**S Waldmann**, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany




## Introduction to Deformation Quantization

The framework of classical mechanics, in its Hamiltonian formulation on the motion space, employs a symplectic manifold (or more generally a Poisson manifold). Observables are families of smooth functions on that manifold $M$. The dynamics is defined in terms of a Hamiltonian $H \in C^\infty(M)$ and the time evolution of an observable $f_t \in C^\infty(M \times \mathbb{R})$ is governed by the equation: $(\mathrm{d}/\mathrm{d}t)f_t = -\{H, f_t\}$.

The quantum-mechanical framework, in its usual Heisenberg's formulation, employs a Hilbert space (states are rays in that space). Observables are families of self-adjoint operators on the Hilbert space. The dynamics is defined in terms of a Hamiltonian $H$, which is a self-adjoint operator, and the time evolution of an observable $A_t$ is governed by the equation $\mathrm{d}A_t/\mathrm{d}t = (\mathrm{i}/\hbar)[H, A_t]$.

Quantization of a classical system is a way to pass from classical to quantum results. A first idea for quantization is to define a correspondence $Q : f \mapsto Q(f)$ mapping a function $f$ to a self-adjoint operator $Q(f)$ on a Hilbert space $\mathscr{H}$ in such a way that $Q(1) = \mathrm{Id}$ and $[Q(f), Q(g)] = \mathrm{i}\hbar Q(\{f, g\})$. Unfortunately, there is no such correspondence defined on all smooth functions on $M$ when one puts an irreducibility requirement (which is necessary not to violate Heisenberg's principle).

Different mathematical treatments of quantization have appeared:

- Geometric quantization of Kostant and Souriau: first, prequantization of a symplectic manifold $(M, \omega)$ where one builds a Hilbert space and a correspondence $Q$ defined on all smooth functions on $M$ but with no irreducibility; second, polarization to "cut down the number of variables."
- Berezin's quantization where one builds on a particular class of symplectic manifolds (some

Kähler manifolds) a family of associative algebras using a symbolic calculus, that is, a dequantization procedure.

- Deformation quantization introduced by Flato, Lichnerowicz, and Sternheimer in 1976 where they "suggest that quantization be understood as a deformation of the structure of the algebra of classical observables rather than a radical change in the nature of the observables."

This deformation approach to quantization is part of a general deformation approach to physics (a seminal idea stressed by Flato): one looks at some level of a theory in physics as a deformation of another level.

Deformation quantization is defined in terms of a star product which is a formal deformation of the algebraic structure of the space of smooth functions on a Poisson manifold. The associative structure given by the usual product of functions and the Lie structure given by the Poisson bracket are simultaneously deformed.

In this article we concentrate on some mathematical results concerning deformations of the Poisson bracket on a symplectic manifold, classification of star products on symplectic manifolds, group actions on star products, convergence properties of some star products, and star products on cotangent bundles.

## Deformations of the Poisson Bracket on a Symplectic Manifold

**Definition 1** A Poisson bracket defined on the space of smooth functions on a manifold $M$ is an $\mathbb{R}$-bilinear map on $C^\infty(M), (u,v) \mapsto \{u,v\}$ such that for any $u,v,w \in C^\infty(M)$:

(i) $\{u,v\} = -\{v,u\}$;
(ii) $\{\{u,v\},w\} + \{\{v,w\},u\} + \{\{w,u\},v\} = 0$;
(iii) $\{u,vw\} = \{u,v\}w + \{u,w\}v$.

A Poisson bracket is given in terms of a contravariant skew-symmetric 2-tensor $P$ on $M$ (called the Poisson tensor) by $\{u,v\} = P(\mathrm{d}u \wedge \mathrm{d}v)$. The Jacobi identity for the Poisson bracket is equivalent to the vanishing of the Schouten bracket $[P,P] = 0$. (The Schouten bracket is the extension – as a graded derivation for the exterior product – of the bracket of vector fields to skew-symmetric contravariant tensor fields.) A Poisson manifold $(M,P)$ is a manifold $M$ with a Poisson bracket defined by $P$.

A particular class of Poisson manifolds, essential in classical mechanics, is the class of "symplectic manifolds." If $(M,\omega)$ is a symplectic manifold (i.e., $\omega$ is a closed nondegenerate 2-form on $M$) and if $u,v \in C^\infty(M)$, the Poisson bracket of $u$ and $v$ is

$$\{u,v\} := X_u(v) = \omega(X_v, X_u)$$

where $X_u$ denotes the Hamiltonian vector field corresponding to the function $u$, that is, such that $\mathrm{i}(X_u)\omega = \mathrm{d}u$. In coordinates the components of the Poisson tensor $P^{ij}$ form the inverse matrix of the components $\omega_{ij}$ of $\omega$.

Duals of Lie algebras form the class of linear Poisson manifolds. If $\mathfrak{g}$ is a Lie algebra, then its dual $\mathfrak{g}^*$ is endowed with the Poisson tensor $P$ defined by $P_\xi(X,Y) := \xi([X,Y])$ for $X, Y \in \mathfrak{g} = (\mathfrak{g}^*)^* \sim (T_\xi \mathfrak{g}^*)^*$.

**Definition 2** A Poisson deformation of the Poisson bracket on a Poisson manifold $(M,P)$ is a Lie algebra deformation of $(C^\infty(M), \{\,,\})$ which is a derivation in each argument, that is, of the form $\{u,v\}_\nu = P_\nu(\mathrm{d}u, \mathrm{d}v)$, where $P_\nu = P + \sum \nu^k P_k$ is a series of skew-symmetric contravariant 2-tensors on $M$ (such that $[P_\nu, P_\nu] = 0$).

Two Poisson deformations $P_\nu$ and $P'_\nu$ of the Poisson bracket $P$ on a Poisson manifold $(M,P)$ are equivalent if there exists a formal path in the diffeomorphism group of $M$, starting at the identity, that is, a series $T = \exp D = \mathrm{Id} + \sum_j (1/j!) D^j$ for $D = \sum_{r\geq 1} \nu^r D_r$ where the $D_r$ are vector fields on $M$, such that

$$T\{u,v\}_\nu = \{Tu, Tv\}'_\nu$$

where $\{u,v\}_\nu = P_\nu(\mathrm{d}u,\mathrm{d}v)$ and $\{u,v\}'_\nu = P'_\nu(\mathrm{d}u,\mathrm{d}v)$.

**Proposition 3** (Flato *et al.* 1975, Lecomte 1987). *On a symplectic manifold* $(M,\omega)$, *any Poisson deformation of the Poisson bracket corresponds to a series of closed 2-forms on M*, $\Omega_\nu = \omega + \sum_{r>0} \nu^r \omega_r$ *and is given by*

$$\{u,v\}_\nu = P_\nu(\mathrm{d}u,\mathrm{d}v) = \Omega_\nu\left(X^\nu_u, X^\nu_v\right)$$

*with* $\mathrm{i}(X^\nu_u)\Omega_\nu = \mathrm{d}u$. *The equivalence classes of Poisson deformations of the Poisson bracket P are parametrized by* $H^2(M;\mathbb{R})[[\nu]]$.

Poisson deformations are used in classical mechanics to express some constraints on the system. To deal with quantum mechanics, Flato *et al.* (1976) introduced star products. These give, by skew-symmetrization, Lie deformations of the Poisson bracket.

**Definition 4** A "star product" on $(M,P)$ is an $\mathbb{R}[\![\nu]\!]$-bilinear associative product $*$ on $C^\infty(M)[\![\nu]\!]$ given by

$$u * v = u *_\nu v := \sum_{r\geq 0} \nu^r C_r(u,v)$$

for $u, v \in C^\infty(M)$ (we consider here real-valued functions; the results for complex-valued functions are similar), such that $C_0(u,v) = uv$, $C_1(u,v) - C_1(v,u) = \{u,v\}$, $1 * u = u * 1 = u$.

When the $C_r$'s are bidifferential operators on $M$, one speaks of a differential star product. When each $C_r$ is a bidifferential operator of order at most $r$ in each argument, one speaks of a natural star product.

One finds in the literature other normalizations for the skew-symmetric part of $C_1$ such as $(i/2)\{\,,\}$; these amount to a rescaling of the parameter $\nu$. For physical applications, in the above convention for the formal parameter, $\nu$ corresponds to $\mathrm{i}\hbar$, where $\hbar$ is Planck's constant.

In the case of complex-valued functions, one can add the further requirement that the complex conjugation is a $^*$-involution for $*$, that is, $\overline{f * g} = \bar{g} * \bar{f}$. According to the interpretation of $\nu$ as being $\mathrm{i}\hbar$, we have to require $\bar{\nu} = -\nu$. Star products satisfying this additional property are called symmetric or Hermitian.

A star product can also be defined not on the whole of $C^\infty(M)$ but on a subspace $N$ which is stable under pointwise multiplication and Poisson bracket.

The simplest example of a deformation quantization is the Moyal product for the Poisson structure $P$ on a vector space $V = \mathbb{R}^m$ with constant coefficients:

$$P = \sum_{i,j} P^{ij} \partial_i \wedge \partial_j, \quad P^{ij} = -P^{ji} \in \mathbb{R}$$

where $\partial_i = \partial/\partial y^i$ is the partial derivative in the direction of the coordinate $y^i$, $i = 1, \ldots, n$. The formula for the Moyal product is

$$(u *_M v)(z) = \exp\left(\frac{\nu}{2} P^{rs} \partial_{y^r} \partial_{y'^s}\right)(u(y)v(y'))\big|_{y=y'=z} \qquad [1]$$

When $P$ is nondegenerate (so $V = \mathbb{R}^{2n}$), the space of formal power series of polynomials on $V$ with Moyal product is called the formal Weyl algebra $W = (S(V)[[\nu]], *_M)$.

Let $\mathfrak{g}^*$ be the dual of a Lie algebra $\mathfrak{g}$. The algebra of polynomials on $\mathfrak{g}^*$ is identified with the symmetric algebra $S(\mathfrak{g})$. One defines a new associative law on this algebra by a transfer of the product $\circ$ in the universal enveloping algebra $U(\mathfrak{g})$, via the bijection between $S(\mathfrak{g})$ and $U(\mathfrak{g})$ given by the total symmetrization $\sigma$:

$$\sigma : S(\mathfrak{g}) \to U(\mathfrak{g}) : X_1 \ldots X_k \mapsto \frac{1}{k!} \sum_{\rho \in S_k} X_{\rho(1)} \circ \cdots \circ X_{\rho(k)}$$

Then $U(\mathfrak{g}) = \oplus_{n \geq 0} U_n$, where $U_n := \sigma(S^n(\mathfrak{g}))$ and we decompose an element $u \in U(\mathfrak{g})$; accordingly $u = \sum u_n$. We define, for $P \in S^p(\mathfrak{g})$ and $Q \in S^q(\mathfrak{g})$,

$$P * Q = \sum_{n \geq 0} (\nu)^n \sigma^{-1}((\sigma(P) \circ \sigma(Q))_{p+q-n}) \qquad [2]$$

This yields a differential star product on $\mathfrak{g}^*$ (Gutt 1983). This star product can be written with an integral formula (for $\nu = 2\pi\mathrm{i}$)(Drinfeld 1987):

$$u * v(\xi) = \int_{\mathfrak{g} \times \mathfrak{g}} \hat{u}(X)\hat{v}(Y) \mathrm{e}^{2\mathrm{i}\pi\langle \xi, \mathrm{CBH}(X,Y)\rangle} \, \mathrm{d}X \, \mathrm{d}Y$$

where $\hat{u}(X) = \int_{\mathfrak{g}^*} u(\eta) \mathrm{e}^{-2\mathrm{i}\pi\langle \eta, X\rangle}$ and CBH denotes Campbell–Baker–Hausdorff formula for the product of elements in the group in a logarithmic chart ($\exp X \exp Y = \exp \mathrm{CBH}(X, Y)\ \forall\ X, Y \in \mathfrak{g}$). We call this the standard (or CBH) star product on the dual of a Lie algebra.

De Wilde and Lecomte (1983) proved that on any symplectic manifold there exists a differential star product. Fedosov (1994) gave a recursive construction of a star product on a symplectic manifold $(M, \omega)$ constructing flat connections on the Weyl bundle. Omori *et al.* (1991) gave an alternative proof of existence of a differential star product on a symplectic manifold, gluing local Moyal star products. In 1997, Kontsevich gave a proof of the existence of a star product on any Poisson manifold and gave an explicit formula for a star product for any Poisson structure on $V = \mathbb{R}^m$. This appeared as a consequence of the proof of his formality theorem.

### Fedosov's Construction of Star Products

Fedosov's construction gives a star product on a symplectic manifold $(M, \omega)$, when one has chosen a symplectic connection and a sequence of closed 2-forms on $M$. The star product is obtained by identifying the space $C^\infty(M)[[\nu]]$ with an algebra of flat sections of the so-called Weyl bundle endowed with a flat connection whose construction is related to the choice of the sequence of closed 2-forms on $M$.

**Definition 5** The symplectic group $\mathrm{Sp}(n, \mathbb{R})$ acts by automorphisms on the formal Weyl algebra $W$. If $(M, \omega)$ is a symplectic manifold, we can form its bundle $F(M)$ of symplectic frames which is a principal $\mathrm{Sp}(n, \mathbb{R})$-bundle over $M$. The associated bundle $\mathscr{W} = F(M) \times_{\mathrm{Sp}(n,\mathbb{R})} W$ is a bundle of associative algebras on $M$ called the Weyl bundle. Sections of the Weyl bundle have the form of formal series

$$a(x, y, \nu) = \sum_{2k+l \geq 0} \nu^k a_{(k)i_1 \ldots i_l}(x) y^{i_1} \cdots y^{i_l}$$

where the coefficients $a_{(k)}$ are symmetric covariant $l$-tensor fields on $M$. The product of two sections taken pointwise makes the space of sections into an algebra, and in terms of the above representation of sections the multiplication has the form

$$(a \circ b)(x, y, \nu) = \left( \exp\left( \frac{\nu}{2} P^{ij} \frac{\partial}{\partial y^i} \frac{\partial}{\partial z^j} \right) a(x, y, \nu) b(x, z, \nu) \right) \Big|_{y=z}$$

Note that the center of this algebra coincides with $C^\infty(M)[[\nu]]$.

A symplectic connection on $M$ is a linear torsion-free connection $\nabla$ such that $\nabla\omega = 0$.

**Remark 6** It is well known that such connections always exist but, unlike the Riemannian case, are not unique. To see the existence, take any torsion-free connection $\nabla'$ and set $T(X, Y, Z) = (\nabla'_X \omega)(Y, Z)$. Define $S$ by $\omega(S(X, Y), Z) = (1/3)(T(X, Y, Z) + T(Y, X, Z))$, then $\nabla_X Y = \nabla'_X Y + S(X, Y)$ defines a symplectic connection.

The connection $\nabla$ induces a covariant derivative on sections of the Weyl bundle, denoted $\partial$. The idea is to try to modify it to have zero curvature. Consider $Da = \partial a - \delta(a) - (1/\nu)[r, a]$, where $r$ is a 1-form with values in $\mathscr{W}$, with $[a, a'dx] = (a \circ a' - a' \circ a)dx$ and $\delta(a) = (1/\nu)[\sum_{ij} \omega_{ij} y^i dx^j, a]$.

**Theorem 7** (Fedosov 1994). *For a given series $\Omega = \sum_{i \geq 1} \nu^i \omega_i$ of closed 2-forms on M, there is a unique $r \in \Gamma(\mathscr{W} \otimes \Lambda^1)$ satisfying some normalization condition, so that $Da = \partial a - \delta(a) - (1/\nu)[r, a]$ is flat. For any $a_\circ \in C^\infty(M)[[\nu]]$, there is a unique a in the subspace $\mathscr{W}_D$ of flat sections of $\mathscr{W}$, such that $a(x, 0, \nu) = a_0(x, \nu)$. The use of this linear isomorphism to transport the algebra structure of $\mathscr{W}_D$ to $C^\infty(M)[[\nu]]$ defines the star product of Fedosov $*_{\nabla,\Omega}$.*

Writing $*_{\nabla,\Omega} = \sum_{i \geq 0} \nu^r C_r^\Omega$, $C_r^\Omega$ only depends on $\omega_i$ for $i < r$ and $C_{r+1}^\Omega(u, v) = c\omega_r(X_u, X_v) + \tilde{C}_{r+1}(u, v)$, where $c \in \mathbb{R}$ and the last term does not depend on $\omega_r$.

## Classification of Star Products on a Symplectic Manifold

Star products on a manifold $M$ are examples of deformations of associative algebras (in the sense of Gerstenhaber). Their study uses the Hochschild cohomology of the algebra (here $C^\infty(M)$ with values in $C^\infty(M)$) where $p$-cochains are $p$-linear maps from $(C^\infty(M))^p$ to $C^\infty(M)$ and where the Hochschild coboundary operator maps the $p$-cochain $C$ to the $(p+1)$-cochain

$$\begin{aligned}(\partial C)(u_0, \ldots, u_p) = {} & u_0 C(u_1, \ldots, u_p) \\ & + \sum_{r=1}^{p} (-1)^r C(u_0, \ldots, u_{r-1}u_r, \ldots, u_p) \\ & + (-1)^{p+1} C(u_0, \ldots, u_{p-1})u_p\end{aligned}$$

For differential star products, we consider differential cochains given by differential operators on each argument. The associativity condition for a star product at order $k$ in the parameter $\nu$ reads

$$\begin{aligned}(\partial C_k)(u, v, w) = \sum_{r+s=k, r,s>0} & (C_r(C_s(u, v), w) \\ & - C_r(u, C_s(v, w)))\end{aligned}$$

If one has cochains $C_j$, $j < k$ such that the star product they define is associative to order $k - 1$, then the right-hand side above is a cocycle ($\partial(\text{RHS}) = 0$) and one can extend the star product to order $k$ if it is a coboundary ($\text{RHS} = \partial(C_k)$).

Denoting by $m$ the usual multiplication of functions, and writing $* = m + C$, where $C$ is a formal series of multidifferential operators, the associativity also reads $\partial C = [C, C]$ where the bracket on the right-hand side is the graded Lie algebra bracket on $D_{\text{poly}}(M)[[\nu]] = \{\text{multidifferential operators}\}$.

**Theorem 8** (Vey 1975). *Every differential p-cocycle C on a manifold M is the sum of the coboundary of a differential $(p - 1)$-cochain and a 1-differential skew-symmetric p-cocycle A: $C = \partial B + A$. In particular, a cocycle is a coboundary if and only if its total skew-symmetrization, which is automatically 1-differential in each argument, vanishes. Given a connection $\nabla$ on M, B can be defined from C by universal formulas* (Cahen and Gutt 1982). *Also*

$$H^p_{\text{diff}}(C^\infty(M), C^\infty(M)) = \Gamma(\Lambda^p TM)$$

The similar result about continuous cochains is due to Connes (1985). In the somewhat pathological case of completely general cochains, the full cohomology is not known.

**Definition 9** Two star products $*$ and $*'$ on $(M, P)$ are said to be equivalent if there is a series of linear operators on $C^\infty(M)$, $T = \text{Id} + \sum_{r=1}^\infty \nu^r T_r$ such that

$$T(f * g) = Tf *' Tg \qquad [3]$$

Remark that the $T_r$ automatically vanish on constants since 1 is a unit for $*$ and for $*'$.

If $*$ and $*'$ are equivalent differential star products, then the equivalence is given by differential operators $T_r$; if they are natural, the equivalence is given by $T = \text{Exp}\, E$ with $E = \sum_{r=1}^\infty \nu^r E_r$, where the $E_r$ are differential operators of order at most $r + 1$.

Nest and Tsygan (1995), then Deligne (1995) and Bertelson *et al.* (1995, 1997) proved that any differential star product on a symplectic manifold $(M, \omega)$ is equivalent to a Fedosov star product and that its equivalence class is parametrized by the corresponding element in $H^2(M; \mathbb{R})[[\nu]]$.

Kontsevich (IHES preprint 97) proved that the coincidence of the set of equivalence classes of star and Poisson deformations is true for general Poisson manifolds:

**Theorem 10** (Kontsevich). *The set of equivalence classes of differential star products on a Poisson manifold* $(M, P)$ *can be naturally identified with the set of equivalence classes of Poisson deformations of* $P$: $P_\nu = P\nu + P_2\nu^2 + \cdots \in \Gamma(X, \wedge^2 T_X)[[\nu]]$, $[P_\nu, P_\nu] = 0$.

Deligne (1995) defines cohomological classes associated to differential star products on a symplectic manifold; this leads to an intrinsic way to parametrize the equivalence class of such a differential star product. The characteristic class $c(*)$ is given in terms of the skew-symmetric part of the term of order 2 in $\nu$ in the star product and in terms of local ("$\nu$-Euler") derivations of the form $D = \nu(\partial/\partial\nu) + X + \sum_{r\geq 1}\nu^r D'_r$. This characteristic class has the following properties:

- The map $C$ from equivalence classes of star products on $(M, \omega)$ to the affine space $-[\omega]/\nu + H^2(M; \mathbb{R})[\![\nu]\!]$ mapping $[*]$ to $c(*)$ is a bijection.
- The characteristic class is natural relative to diffeomorphisms and is equivariant under a change of parameter (Gutt and Rawnsley 1995).
- The characteristic class $c(*)$ coincides (cf. Deligne (1995) and Neumaier (1999)) for Fedosov-type star products with their characteristic class introduced by Fedosov as the de Rham class of the curvature of the generalized connection used to build them (up to a sign and factors of 2).

Index theory has been introduced in the framework of deformation quantization by Fedosov (1996) and by Nest and Tsygan (1995, 1996). We refer to the papers of Bressler, Nest, and Tsygan for further developments in that subject. A first tool in that theory is the existence of a trace for the deformed algebra; this trace is essentially unique in the framework of symplectic manifolds (an elementary proof is given in Karabegov (1998) and Gutt and Rawnsley (2003)); the trace is not unique for more general Poisson manifolds.

**Definition 11** A homomorphism from a differential star product $*$ on $(M, P)$ to a differential star product $*'$ on $(M', P')$ is an $\mathbb{R}$-linear map $A : C^\infty(M)[\![\nu]\!] \to C^\infty(M')[\![\nu]\!]$, continuous in the $\nu$-adic topology, such that

$$A(u * v) = Au *' Av$$

It is an isomorphism if the map is bijective.

Any isomorphism between two differential star products on symplectic manifolds is the combination of a change of parameter and a $\nu$-linear isomorphism. Any $\nu$-linear isomorphism between two star products $*$ on $(M, \omega)$ and $*'$ on $(M', \omega')$ is the combination of the action on functions of a symplectomorphism $\psi : M' \to M$ and an equivalence between $*$ and the pullback via $\psi$ of $*'$. It exists if and only if those two star products are equivalent, that is, if and only if $(\psi^{-1})^* c(*') = c(*)$, where $(\psi^{-1})^*$ denotes the action of $\psi^{-1}$ on the second de Rham cohomology space. In particular, a symplectomorphism $\psi$ of a symplectic manifold can be extended to a $\nu$-linear automorphism of a given differential star product on $(M, \omega)$ if and only if $(\psi)^* c(*) = c(*)$ (Gutt and Rawnsley 1999).

The notion of homomorphism and its relation to modules has been studied by Bordemann (2004).

The link between the notion of star product on a symplectic manifold and symplectic connections already appears in the seminal paper of Bayen *et al.* (1978), and was further developed by Lichnerowicz (1982), who showed that any Vey star product (i.e., a star product defined by bidifferential operators whose principal symbols at each order coincide with those of the Moyal star product) determines a unique symplectic connection. Fedosov's construction yields a Vey star product on any symplectic manifold starting from a symplectic connection and a formal series of closed 2-forms on the manifold. Furthermore, any star product is equivalent to a Fedosov star product and the de Rham class of the formal 2-form determines the equivalence class of the star product. On the other hand, many star products which appear in natural contexts (e.g., cotangent bundles or Kähler manifolds) are not Vey star products but are natural star products.

**Theorem 12** (Gutt and Rawnsley 2004). *Any natural star product on a symplectic manifold* $(M, \omega)$ *determines uniquely*

(i) *A symplectic connection* $\nabla = \nabla(*)$.
(ii) *A formal series of closed 2-forms* $\Omega = \Omega(*) \in \nu\Lambda^2(M)[\![\nu]\!]$.
(iii) *A formal series* $E = \sum_{r\geq 1}\nu^r E_r$ *of differential operators of order* $\leq r+1$ ($E_2$ *of order* $\leq 2$), *with* $E_r u = \sum_{k=2}^{r+1}(E_r^{(k)})^{i_1\ldots i_k}\nabla^k_{i_1\ldots i_k} u$, *where the* $E_r^{(k)}$ *are symmetric contravariant* $k$*-tensor fields*

*such that*

$$u * v = \exp -E\big((\exp Eu) *_{\nabla,\Omega} (\exp Ev)\big) \qquad [4]$$

*We denote* $* = *_{\nabla,\Omega,E}$. *If* $\tau$ *is a diffeomorphism of* $M$ *then the data for* $\tau \cdot *$ *is* $\tau\cdot\nabla$, $\tau\cdot\Omega$, *and* $\tau \cdot E$. *In*

*particular, a vector field X is a derivation of a natural star product* $*$, *if and only if* $\mathscr{L}_X\omega=0$, $\mathscr{L}_X\Omega=0$, $\mathscr{L}_X\nabla=0$, *and* $\mathscr{L}_X E=0$.

## Group Actions on Star Products

Symmetries in quantum theories are automorphisms of an algebra of observables. In the framework where quantization is defined in terms of a star product, a symmetry $\sigma$ of a star product $*$ is an automorphism of the $\mathbb{R}[\![\nu]\!]$-algebra $C^\infty(M)[\![\nu]\!]$ with multiplication given by $*$:

$$\sigma(u * v) = \sigma(u) * \sigma(v), \qquad \sigma(1) = 1$$

where $\sigma$, being determined by what it does on $C^\infty(M)$, will be a formal series $\sigma(u)=\sum_{r\geq 0}\nu^r\sigma_r(u)$ of linear maps $\sigma_r\cdot C^\infty(M)\to C^\infty(M)$. We denote by $\mathrm{Aut}_{\mathbb{R}[\![\nu]\!]}(M,*)$ the set of those symmetries.

Any such automorphism $\sigma$ of $*$ then can be written as $\sigma(u)=T(u\circ\tau^{-1})$, where $\tau$ is a Poisson diffeomorphism of $(M,P)$ and $T=\mathrm{Id}+\sum_{r\geq 1}\nu^r T_r$ is a formal series of linear maps. If $*$ is differential, then the $T_r$ are differential operators; if $*$ is natural, then $T=\mathrm{Exp}\,E$ with $E=\sum_{r\geq 1}\nu^r E_r$ and $E_r$ is a differential operator of order at most $r+1$.

If $\sigma_t$ is a one-parameter group of symmetries of the star product $*$, then its generator $D$ will be a derivation of $*$. Denote the Lie algebra of $\nu$-linear derivations of $*$ by $\mathrm{Der}_{\mathbb{R}[\![\nu]\!]}(M,*)$.

An action of a Lie group $G$ on a star product $*$ on a Poisson manifold $(M,P)$ is a homomorphism $\sigma: G\to \mathrm{Aut}_{\mathbb{R}[\![\nu]\!]}(M,*)$; then $\sigma_g=(\tau_g)^{-1*}+\mathrm{O}(\nu)$ and there is an induced Poisson action $\tau$ of $G$ on $(M,P)$.

Given a Poisson action $\tau$ of $G$ on $(M,P)$, a star product is said to be "invariant" under $G$ if all the $(\tau_g)^{-1*}$ are automorphisms of $*$.

An action of a Lie group $G$ on $*$ induces a homomorphism of Lie algebras $D:\mathfrak{g}\to\mathrm{Der}_{\mathbb{R}[\![\nu]\!]}(M,*)$. For each $\xi\in\mathfrak{g}, D_\xi=\xi^*+\sum_{r\geq 1}\nu^r D^r_\xi$, where $\xi^*$ is the fundamental vector field on $M$ defined by $\tau$; hence,

$$\xi^*(x) = \frac{\mathrm{d}}{\mathrm{d}t}|_0\tau(\exp - t\xi)x)$$

Such a homomorphism $D:\mathfrak{g}\to\mathrm{Der}_{\mathbb{R}[\![\nu]\!]}(M,*)$ is called an action of the Lie algebra $\mathfrak{g}$ on $*$.

**Proposition 13** (Arnal *et al.* 1983). *Given* $D:\mathfrak{g}\to\mathrm{Der}_{\mathbb{R}[\![\nu]\!]}(M,*)$ *a homomorphism so that for each* $\xi\in\mathfrak{g}$, $D_\xi=\xi^*+\sum_{r\geq 1}\nu^r D^r_\xi$, *where* $\xi^*$ *are the fundamental vector fields on M defined by an action* $\tau$ *of G on M and the* $D^r_\xi$ *are differential operators, then there exists a local homomorphism* $\sigma: U\subset G\to\mathrm{Aut}_{\mathbb{R}[\![\nu]\!]}(M,*)$ *so that* $\sigma_*=D$.

If we want the analog in our framework to the requirement that operators should correspond to the infinitesimal actions of a Lie algebra, we should ask the derivations to be inner so that functions are associated to the elements of the Lie algebra.

A derivation $D\in\mathrm{Der}_{\mathbb{R}[\![\nu]\!]}(M,*)$ is said to be essentially inner or Hamiltonian if $D=(1/\nu)\mathrm{ad}_* u$ for some $u\in C^\infty(M)[\![\nu]\!]$. We call an action of a Lie group almost $*$-Hamiltonian if each $D_\xi$ is essentially inner; this is equivalent to the knowledge of a linear map $\lambda:\mathfrak{g}\to C^\infty(M)[\![\nu]\!]\ \xi\mapsto\lambda_\xi$ so that $\mathrm{ad}_*(1/\nu)[\lambda_\xi,\lambda_\eta]_*=\mathrm{ad}_*\,\lambda_{[\xi,\eta]}$.

We say the action is $*$-Hamiltonian if $\lambda_\xi$ can be chosen to make

$$\mathfrak{g}\to C^\infty(M)[\![\nu]\!], \quad \xi\mapsto\lambda_\xi$$

a homomorphism of Lie algebras, where $C^\infty(M)[\![\nu]\!]$ is endowed with the bracket $(1/\nu)[\,,]_*$. Such a homomorphism is called a quantization in Arnal *et al.* (1983) and is called a generalized moment map in Bordemann *et al.* (1998).

When a map $\mu^0:\mathfrak{g}\to C^\infty(M)$ is a generalized moment map, that is,

$$\frac{1}{\nu}\left(\mu^0_\xi*\mu^0_\eta-\mu^0_\eta*\mu^0_\xi\right)=\mu^0_{[\xi,\eta]}$$

the star product is said to be covariant under $\mathfrak{g}$.

When a map $\mu:\mathfrak{g}\to C^\infty(M)[\![\nu]\!]$ is a generalized moment map, so that $D_\xi$ has no terms in $\nu$ of degree $>0$, thus $D_\xi=\xi^*$, this map is called a quantum moment map (Xu 1998). Clearly in that situation, the star product is invariant under the action of $\mathfrak{g}$ on $M$.

Covariant star products have been considered to study representations theory of some classes of Lie groups in terms of star products. In particular, an autonomous star formulation of the theory of representations of nilpotent Lie groups has been given by Arnal and Cortet (1984, 1985).

Consider a differential star product $*$ on a symplectic manifold, admitting an algebra $\mathfrak{g}$ of vector fields on $M$ consisting of derivations of $*$, and assume there is a symplectic connection $\nabla$ which is invariant under $\mathfrak{g}$; then $*$ is equivalent, through an equivariant equivalence ($T$ with $\mathscr{L}_X T=0$), to a Fedosov star product $*_{\nabla,\Omega}$; this yields to a classification of such invariant star products (Bertelson *et al.* 1998).

**Proposition 14** (Kravchenko, Gutt and Rawnsley, Müller-Bahns, Neumaier, and Hamachi). *Consider a Fedosov star product* $*_{\nabla,\Omega}$ *on a symplectic manifold. A vector field X is a derivation of* $*_{\nabla,\Omega}$ *if and only if* $\mathscr{L}_X\omega=0, \mathscr{L}_X\Omega=0$, *and* $\mathscr{L}_X\nabla=0$. *A vector field X is an inner derivation of* $*=*_{\nabla,\Omega}$ *if*

*and only if* $\mathscr{L}_X \nabla = 0$ *and there exists a series of functions* $\lambda_X$ *such that*

$$\mathrm{i}(X)\omega - \mathrm{i}(X)\Omega = \mathrm{d}\lambda_X$$

*In this case* $X(u) = (1/\nu)(\mathrm{ad}_* \lambda_X)(u)$.

On a symplectic manifold $(M, \omega)$, a vector field $X$ is an inner derivation of the natural star product $* = *_{\nabla,\Omega,E}$ if and only if $\mathscr{L}_X \nabla = 0, \mathscr{L}_X E = 0$, and there exists a series of functions $\lambda_X$ such that

$$\mathrm{i}(X)\omega - \mathrm{i}(X)\Omega = \mathrm{d}\lambda_X$$

Then $X = (1/\nu)\mathrm{ad}_* \mu_X$ with $\mu_X = \mathrm{Exp}(E^{-1})\lambda_X$.

Let $G$ be a compact Lie group of symplectomorphisms of $(M, \omega)$ and $\mathfrak{g}$ the corresponding Lie algebra of symplectic vector fields on $M$. Consider a star product $*$ on $M$ which is invariant under $G$. The Lie algebra $\mathfrak{g}$ consists of inner derivations for $*$ if and only if there exists a series of functions $\lambda_X$ and a representative $(1/\nu)(\omega - \Omega)$ of the characteristic class of $*$ such that $\mathrm{i}(X)\omega - \mathrm{i}(X)\Omega = \mathrm{d}\lambda_X$.

Star products which are invariant and covariant are used in the problem of reduction: this is a device in symplectic geometry which allows one to reduce the number of variables. An important issue in quantization is to know if and how quantization commutes with reduction. This problem has been studied by Fedosov for the action of a compact group on the particular star products constructed by him with trivial characteristic class ( $*_{\nabla,0}$ ). Here, one indeed obtains some "quantization commutes with reduction" statements. More generally, Bordemann, Herbig, and Waldmann considered covariant star products. In this case, one can construct a classical and quantum BRST complex whose cohomology describes the algebra of observables for the reduced system. While this is well known classically – at least under some regularity assumptions on the group action – for the quantized situation, the nontrivial question is whether the quantum BRST cohomology is "as large as" the classical one. Clearly, from the physical point of view, this is crucial. It turns out that whereas for strongly invariant star products one indeed obtains a quantization of the reduced phase space, in general the quantum BRST cohomology might be too small. More general situations of reduction have also been discussed by, for example, Bordemann as well as Cattaneo and Felder, when a coisotropic (i.e., first class) constraint manifold is given.

## Convergence of Some Star Products on a Subclass of Functions

Let $(M, P)$ be a Poisson manifold and let $*$ be a differential star product on it with 1 acting as the identity. Observe that if there exists a value $k$ of $\nu$ such that

$$u * v = \sum_{r=0}^{\infty} \nu^r C_r(u, v)$$

converges (for the pointwise convergence of functions), for all $u, v \in C^\infty(M)$, to $F_k(u, v)$ in such a way that $F_k$ is associative, then $F_k(u, v) = uv$. This is easy to see as the order of differentiation in the $C_r$ necessarily is at least $r$ in each argument and thus the Borel lemma immediately gives the result. So assuming "too much" convergence kills all deformations. On the other hand, in any physical situation, one needs some convergence properties to be able to compute the spectrum of quantum observables in terms of a star product (as in Bayen *et al.* 1978).

In the example of Moyal star product on the symplectic vector space $(\mathbb{R}^{2n}, \omega)$, the formal formula

$$(u *_M v)(z) = \exp\left(\frac{\nu}{2} P^{rs} \partial_{x^r} \partial_{y^s}\right)(u(x)v(y))\Big|_{x=y=z}$$

obviously converges when $u$ and $v$ are polynomials. On the other hand, there is an integral formula for Moyal star product given by

$$(u * v)(\xi) = (\pi\hbar)^{-2n} \int u(\xi')v(\xi'') \times \exp\left(\frac{2\mathrm{i}}{\hbar}\Big(\omega(\xi, \xi'') + \omega(\xi'' + \xi') + \omega(\xi', \xi)\Big)\right) \mathrm{d}\xi' \, \mathrm{d}\xi''$$

and this product $*$ gives a structure of associative algebra on the space of rapidly decreasing functions $\mathscr{S}(\mathbb{R}^{2n})$. The formal formula converges (for $\nu = \mathrm{i}\hbar$) in the topology of $\mathscr{S}'$ for $u$ and $v$ with compactly supported Fourier transform.

Some works have been done about convergence of star products.

- The method of quantization of Kähler manifolds due to Berezin as the inverse of taking symbols of operators, to construct on Hermitian symmetric spaces star products which are convergent on a large class of functions on the manifold (Moreno, Cahen Gutt, and Rawnsley, Karabegov, Schlichenmaier).

- The constructions of operator representations of star products (Fedosov, Bordemann, Neumaier, and Waldmann).
- The work of Rieffel and the notion of strict deformation quantization. Examples of strict (Fréchet) quantization have been given by Omori, Maeda, Niyazaki, and Yoshioka, and by Bieliavsky.

### Convergence of Berezin-Type Star Products on Hermitian Symmetric Spaces

The method to construct a star product involves making a correspondence between operators and functions using coherent states, transferring the operator composition to the symbols, introducing a suitable parameter into this Berezin composition of symbols, taking the asymptotic expansion in this parameter on a large algebra of functions, and then showing that the coefficients of this expansion satisfy the cocycle conditions to define a star product on the smooth functions (Cahen *et al.* 1995). The idea of an asymptotic expansion appears in Berezin (1975) and in Moreno and Ortega-Navarro (1983, 1986).

This asymptotic expansion exists for compact $M$, and defines an associative multiplication on formal power series in $k^{-1}$ with coefficients in $C^\infty(M)$ for compact coadjoint orbits. For $M$ a Hermitian symmetric space of compact type and more generally for compact coadjoint orbits (i.e., flag manifolds), this formal power series converges on the space of symbols (Karabegov 1998).

For general Hermitian symmetric spaces of noncompact type, using their realization as bounded domains, one defines an analogous algebra of symbols of polynomial differential operators.

Reshetikhin and Takhtajan have constructed an associative formal star product given by an asymptotic expansion on any Kähler manifold. This they do in two steps, first building an associative product for which 1 is not a unit element, then passing to a star product.

We denote by $(L,\nabla,h)$ a quantization bundle for the Kähler manifold $(M,\omega,J)$ (i.e., a holomorphic line bundle $L$ with connection $\nabla$ admitting an invariant Hermitian structure $h$, such that the curvature is $\operatorname{curv}(\nabla)=-2i\pi\omega$). We denote by $\mathscr{H}$ the Hilbert space of square-integrable holomorphic sections of $L$ which we assume to be nontrivial. The coherent states are vectors $e_q\in\mathscr{H}$ such that

$$s(x)=\langle s,e_q\rangle q,\quad \forall q\in\mathscr{L}_x,\ x\in M,\ s\in\mathscr{H}$$

($\mathscr{L}$ is the complement of the zero section in $L$). The function $\epsilon(x)=|q|^2\|e_q\|^2, q\in\mathscr{L}_x$, is well defined and real analytic.

Let $A:\mathscr{H}\to\mathscr{H}$ be a bounded linear operator and let

$$\widehat{A}(x)=\frac{\langle Ae_q,e_q\rangle}{\langle e_q,e_q\rangle},\quad q\in\mathscr{L}_x,\ x\in M$$

be its symbol. The function $\widehat{A}$ has an analytic continuation to an open neighborhood of the diagonal in $M\times\bar{M}$ given by

$$\widehat{A}(x,y)=\frac{\langle Ae_{q'},e_q\rangle}{\langle e_{q'},e_q\rangle},\quad q\in\mathscr{L}_x,\ q'\in\mathscr{L}_y$$

which is holomorphic in $x$ and antiholomorphic in $y$. We denote by $\hat{E}(L)$ the space of symbols of bounded operators on $\mathscr{H}$. We can extend this definition of symbols to some unbounded operators provided everything is well defined.

The composition of operators on $\mathscr{H}$ gives rise to an associative product $*$ for the corresponding symbols:

$$(\widehat{A}*\widehat{B})(x)=\int_M \widehat{A}(x,y)\widehat{B}(y,x)\psi(x,y)\epsilon(y)\frac{\omega^n(y)}{n!}$$

where

$$\psi(x,y)=\frac{|\langle e_{q'},e_q\rangle|^2}{\|e_{q'}\|^2\|e_q\|^2},\quad q\in\mathscr{L}_x,\ q'\in\mathscr{L}_y$$

is a globally defined real analytic function on $M\times M$ provided $\epsilon$ has no zeros ($\psi(x,y)\le 1$ everywhere, with equality where the lines spanned by $e_q$ and $e_{q'}$ coincide).

Let $k$ be a positive integer. The bundle $(L^k=\otimes^k L,\nabla^k,h^k)$ is a quantization bundle for $(M,k\omega,J)$ and we denote by $\mathscr{H}^k$ the corresponding space of holomorphic sections and by $\hat{E}(L^k)$ the space of symbols of linear operators on $\mathscr{H}^k$. We let $\epsilon^{(k)}$ be the corresponding function. We say that the quantization is regular if $\epsilon^{(k)}$ is a nonzero constant for all non-negative $k$ and if $\psi(x,y)=1$ implies $x=y$. (Remark that if the quantization is homogeneous, all $\epsilon^{(k)}$ are constants.)

**Theorem 15** (Cahen *et al.*). *Let $(M,\omega,J)$ be a Kähler manifold and $(L,\nabla,h)$ be a regular quantization bundle over M. Let $\widehat{A},\widehat{B}$ be in $\mathscr{B}$, where $\mathscr{B}\subset C^\infty(M)$ consists of functions f which have an analytic continuation in $M\times\bar{M}$ so that $f(x,y)\psi(x,y)^l$ is globally defined, smooth and bounded on $K\times M$ and $M\times K$ for each compact subset K of M for some positive power l. Then*

$$(\widehat{A}*_k\widehat{B})(x)=\int_M \widehat{A}(x,y)\widehat{B}(y,x)\psi^k(x,y)\epsilon^{(k)}k^n\frac{\omega^n}{n!}(y)$$

*defined for k sufficiently large, admits an asymptotic expansion in $k^{-1}$ as $k \to \infty$*

$$(\widehat{A} *_k \widehat{B})(x) \sim \sum_{r \geq 0} k^{-r} C_r(\widehat{A}, \widehat{B})(x)$$

*and the cochains $C_r$ are smooth bidifferential operators, invariant under the automorphisms of the quantization and determined by the geometry alone. Furthermore, $C_0(\widehat{A}, \widehat{B}) = \widehat{A}\widehat{B}$ and $C_1(\widehat{A}, \widehat{B}) - C_1(\widehat{B}, \widehat{A}) = (\mathrm{i}/\pi)\{\widehat{A}, \widehat{B}\}$.*

*If M is a flag manifold, this defines a star product on $C^\infty(M)$ and the $*_k$ product of two symbols is convergent (it is a rational function of k without pole at infinity) (cf. Karabegov in that generality).*

*If $\mathscr{D}$ be a bounded symmetric domain and $\mathscr{E}$ the algebra of symbols of polynomial differential operators on a homogeneous holomorphic line bundle L over $\mathscr{D}$ which gives a realization of a holomorphic discrete series representation of $G_0$, then for f and g in $\mathscr{E}$ the Berezin product $f *_k g$ has an asymptotic expansion in powers of $k^{-1}$ which converges to a rational function of k. The coefficients of the asymptotic expansion are bidifferential operators which define an invariant and covariant star product on $C^\infty(\mathscr{D})$.*

## Star Products on Cotangent Bundles

Since from the physical point of view cotangent bundles $\pi : T^*Q \to Q$ over some configuration space $Q$, endowed with their canonical symplectic structure $\omega_0$, are one of the most important phase spaces, any quantization scheme should be tested and exemplified for this class of classical mechanical systems.

We first recall that on $T^*Q$ there is a canonical vector field $\xi$, the Euler or Liouville vector field which is locally given by $\xi = p_k(\partial/\partial p_k)$. Here and in the following, we use local bundle coordinates $(q^k, p_k)$ induced by local coordinates $x^k$ on $Q$. Using $\xi$ we can characterize those functions $f \in C^\infty(T^*Q)$ which are polynomial in the fibers of degree $k$ by $\xi f = kf$. They are denoted by $\mathrm{Pol}^k(T^*Q)$, whereas $\mathrm{Pol}^\bullet(T^*Q)$ denotes the subalgebra of all functions which are polynomial in the fibers. Clearly, most of the physically relevant observables such as the kinetic energy, potentials, and generators of point transformations are in $\mathrm{Pol}^\bullet(T^*Q)$. Moreover, $\mathrm{Pol}^\bullet(T^*Q)$ is a Poisson subalgebra with

$$\left\{\mathrm{Pol}^k(T^*Q), \mathrm{Pol}^\ell(T^*Q)\right\} \subseteq \mathrm{Pol}^{k+\ell-1}(T^*Q) \qquad [5]$$

since $\mathscr{L}_\xi \omega_0 = \omega_0$ is conformally symplectic.

All this suggests that for a quantization of $T^*Q$, the polynomials $\mathrm{Pol}^\bullet(T^*Q)$ should play a crucial role. In deformation quantization this is accomplished by the notion of a homogeneous star product (De Wilde and Lecomte 1983). If the operator

$$\mathsf{H} = \nu \frac{\partial}{\partial \nu} + \mathcal{L}_\xi \qquad [6]$$

is a derivation of a formal star product $\star$, then $\star$ is called homogeneous. It immediately follows that $\mathrm{Pol}(T^*Q)[\nu] \subseteq C^\infty(T^*Q)[[\nu]]$ is a subalgebra over the ring $\mathbb{C}[\nu]$ of polynomials in $\nu$. Hence for homogeneous star products, the question of convergence (in general quite delicate) has a simple answer.

Let us now describe a simple construction of a homogeneous star product (following Bordemann *et al.* (1998)). We choose a torsion-free connection $\nabla$ on $Q$ and consider the operator of the symmetrized covariant derivative, locally given by

$$\mathsf{D} = \mathrm{d}x^k \vee \nabla_{\partial/\partial x^k} : \Gamma^\infty(\mathrm{S}^k T^*Q) \to \Gamma^\infty(\mathrm{S}^{k+1} T^*Q) \qquad [7]$$

Clearly, $\mathsf{D}$ is a global object and a derivation of the symmetric algebra $\bigoplus_{k=0}^\infty \Gamma^\infty(\mathrm{S}^k T^*Q)$. Let now $f \in \mathrm{Pol}^\bullet(T^*Q)$ and $\psi \in C^\infty(Q)$ be given. Then one defines the standard-ordered quantization $\varrho_{\mathrm{Std}}(f)$ of $f$ with respect to $\nabla$ to be the differential operator $\varrho_{\mathrm{Std}}(f) : C^\infty(Q) \to C^\infty(Q)$ locally given by

$$\varrho_{\mathrm{Std}}(f)\psi = \sum_{r=0}^\infty \frac{(-\nu)^r 1}{r!} \left. \frac{\partial^r f}{\partial p_{k_1} \cdots \partial p_{k_r}} \right|_{p=0} \times \mathrm{i_s}\left(\frac{\partial}{\partial x^{k_1}}\right) \cdots \mathrm{i_s}\left(\frac{\partial}{\partial x^{k_r}}\right) \frac{1}{r!} \mathsf{D}^r \psi \qquad [8]$$

where $\mathrm{i_s}$ denotes the symmetric insertation of vector fields in symmetric forms. Again, this is independent of the coordinate system $x^k$. The infinite sum is actually finite as long as $f \in \mathrm{Pol}^\bullet(T^*Q)$ whence we can safely set $\nu = \mathrm{i}\hbar$ in this case. Indeed, [8] is the well-known symbol calculus for differential operators and it establishes a linear bijection

$$\varrho_{\mathrm{Std}} : \mathrm{Pol}^\bullet(T^*Q) \to \mathrm{DiffOp}(C^\infty(Q)) \qquad [9]$$

which generalizes the usual canonical quantization in the flat case of $T^*Q = T^*\mathbb{R}^n = \mathbb{R}^{2n}$. Using this linear bijection, we can define a new product $\star_{\mathrm{Std}}$ for $\mathrm{Pol}^\bullet(T^*Q)$ by

$$f \star_{\mathrm{Std}} g = \varrho_{\mathrm{Std}}^{-1}(\varrho_{\mathrm{Std}}(f)\varrho_{\mathrm{Std}}(g)) = \sum_{r=0}^\infty \nu^r C_r(f, g) \qquad [10]$$

It is now easy to see that $\star_{\mathrm{Std}}$ fulfills all requirements of a homogeneous star product except for the fact that the $C_r(\,\cdot\,,\cdot\,)$ are bidifferential. In this approach

this is far from being obvious as we only worked with functions polynomial in the fibers so far. Nevertheless, it is true whence $\star_{\mathrm{Std}}$ indeed defines a star product for $C^\infty(T^*Q)[[\nu]]$.

In fact, there is a different characterization of $\star_{\mathrm{Std}}$ using a slightly modified Fedosov construction: first one uses $\nabla$ to define a torsion-free symplectic connection on $T^*Q$ by a fairly standard lifting. Moreover, using $\nabla$ one can define a standard-ordered fiberwise product $\circ_{\mathrm{Std}}$ for the formal Weyl algebra bundle over $T^*Q$, being the starting point of the Fedosov construction of star products. With these two ingredients one finally obtains $\star_{\mathrm{Std}}$ from the Fedosov construction with the big advantage that now the order of differentiation in the $C_r$ can easily be determined to be $r$ in each argument, whence $\star_{\mathrm{Std}}$ is even a natural star product. Moreover, $C_r$ differentiates the first argument only in momentum directions which reflects the standard ordering.

Already in the flat situation the standard ordering is not an appropriate quantization scheme from the physical point of view as it maps real-valued functions to differential operators which are not symmetric in general. To pose this question in a geometric framework, we have to specify a positive density $\mu \in \Gamma^\infty(|\Lambda^n|T^*Q)$ on the configuration space $Q$ first, as for functions there is no invariant meaning of integration. Specifying $\mu$ we can consider the pre-Hilbert space $C_0^\infty(Q)$ with inner product

$$\langle \phi, \psi \rangle = \int_Q \bar{\phi}\psi\mu \qquad [11]$$

Now the adjoint with respect to [11] of $\varrho_{\mathrm{Std}}(f)$ can be computed explicitly. We first consider the second-order differential operator

$$\Delta = \frac{\partial^2}{\partial q^k \partial p_k} + p_k \Gamma^k_{\ell m} \frac{\partial^2}{\partial p_\ell \partial p_m} + \Gamma^k_{k\ell} \frac{\partial}{\partial p_\ell} \qquad [12]$$

where $\Gamma^k_{\ell m}$ are the Christoffel symbols of $\nabla$. In fact, $\Delta$ is defined independently of the coordinates and coincides with the Laplacian of the pseudo-Riemannian metric on $T^*Q$ which is obtained from the natural pairing of vertical and horizontal spaces defined by using $\nabla$. Moreover, we need the 1-form $\alpha$ defined by $\nabla_X \mu = \alpha(X)\mu$ and the corresponding vertical vector field $\alpha^{\mathrm{v}} \in \Gamma^\infty(T(T^*Q))$ locally given by $\alpha^{\mathrm{v}} = \alpha_k(\partial/\partial p_k)$. Then

$$\varrho_{\mathrm{Std}}(f)^\dagger = \varrho_{\mathrm{Std}}(N^2\bar{f}), \quad N = \mathrm{e}^{(\nu/2)(\Delta + \alpha^{\mathrm{v}})} \qquad [13]$$

Note that due to the curvature contributions, this statement is a highly nontrivial partial integration compared to the flat case. Note also that for $f \in \mathrm{Pol}(T^*Q)[\nu]$, we have $Nf \in \mathrm{Pol}(T^*Q)[\nu]$ as well, and $N$ commutes with H. As in the flat case this allows one to define a Weyl-ordered quantization by

$$\varrho_{\mathrm{Weyl}}(f) = \varrho_{\mathrm{Std}}(Nf) \qquad [14]$$

together with a so-called Weyl-ordered star product

$$f \star_{\mathrm{Weyl}} g = N^{-1}(Nf \star_{\mathrm{Std}} Ng) \qquad [15]$$

which is now a Hermitian and homogeneous star product such that $\varrho_{\mathrm{Weyl}}$ becomes a $^*$-representation of $\star_{\mathrm{Weyl}}$, that is, we have $\varrho_{\mathrm{Weyl}}(f \star_{\mathrm{Weyl}} g) = \varrho_{\mathrm{Weyl}}(f)\,\varrho_{\mathrm{Weyl}}(g)$ and $\varrho_{\mathrm{Weyl}}(f)^\dagger = \varrho_{\mathrm{Weyl}}(\bar{f})$. Note that in the flat case this is precisely the Moyal star product $*_M$ from [1].

The star products $\star_{\mathrm{Std}}$ and $\star_{\mathrm{Weyl}}$ have been extensively studied by Bordemann, Neumaier, Pflaum, and Waldmann and provide now a well-understood quantization on cotangent bundles. We summarize a few highlights of this theory:

1. In the particular case of a Levi-Civita connection $\nabla$ for some Riemannian metric $g$ and the corresponding volume density $\mu_g$, the 1-form $\alpha$ vanishes. This simplifies the operator $N$ and describes the physically most interesting situation.
2. If the configuration space is a Lie group $G$, then its cotangent bundle $T^*G \cong G \times \mathfrak{g}^*$ is trivial by using, for example, left-invariant 1-forms. In this case the star products $\star_{\mathrm{Weyl}}$ and $\star_{\mathrm{Std}}$ restrict to the CBH star product on $\mathfrak{g}^*$. Moreover, $\star_{\mathrm{Weyl}}$ coincides with the star product found by Gutt (1983) on $T^*G$.
3. Using the operator $N$ one can interpolate between the two different ordering descriptions $\varrho_{\mathrm{Std}}$ and $\varrho_{\mathrm{Weyl}}$ by inserting an additional ordering parameter $\kappa$ in the exponent, that is, $N_\kappa = \exp(\nu\kappa(\Delta + \alpha^{\mathrm{v}}))$. Thus, one obtains $\kappa$-ordered representations $\varrho_\kappa$ together with corresponding $\kappa$-ordered star products $\star_\kappa$, where $\kappa = 0$ corresponds to standard ordering and $\kappa = 1/2$ corresponds to Weyl ordering. For $\kappa = 1$, one obtains antistandard ordering and in general one has the relation $\overline{f \star_\kappa g} = \bar{g} \star_{1-\kappa} \bar{f}$ as well as $\varrho_\kappa(f)^\dagger = \varrho_{1-\kappa}(\bar{f})$.
4. One can describe also the quantization of an electrically charged particle moving in a magnetic background field $B$. This is modeled by a closed 2-form $B \in \Gamma^\infty(\Lambda^2 T^*Q)$ on $Q$. Using local vector potentials $A \in \Gamma^\infty(T^*Q)$ with $B = \mathrm{d}A$ locally, and by minimal coupling, one obtains a star product $\star_B$ which depends only on $B$ and not on the local potentials $A$. It will be equivalent to $\star_{\mathrm{Weyl}}$ if and only if $B$ is exact. In general, its characteristic class is, up to a factor, given by the class $[B]$ of the magnetic field $B$. While the observable

algebra always exists, a Schrödinger-like representation of $\star_B$ only exists if $B$ satisfies the usual integrality condition. In this case, there exists a representation on sections of a line bundle whose first Chern class is given by $[B]$. This manifests Dirac's quantization condition for magnetic charges in deformation quantization. Another equivalent interpretation of this result is obtained by Morita theory: the star products $\star_{\text{Weyl}}$ and $\star_B$ are Morita equivalent if and only if $B$ satisfies Dirac's integrality condition.

5. Analogously, one can determine the unitary equivalence classes of representations for a fixed, exact magnetic field $B$. It turns out that the representations depend on the choice of the global vector potential $A$ and are unitarily equivalent if the difference between the two vector potentials satisfies an integrality condition known from the Aharonov–Bohm effect. This way, the Aharonov–Bohm effect can be formulated within the representation theory of deformation quantization.
6. There are several variations of the representations $\varrho_{\text{Std}}$ and $\varrho_{\text{Weyl}}$. In particular, one can construct a representation on half-forms instead of functions, thereby avoiding the choice of the integration density $\mu$. Moreover, all the Weyl-ordered representations can be understood as GNS representations coming from a particular positive functional, the Schrödinger functional. For $\varrho_{\text{Weyl}}$ this functional is just the integration over the configuration space $Q$.
7. All the (formal) star products and their representations can be understood as coming from formal asymptotic expansions of integral formulas. From this point of view, the formal representations and star products are a particular kind of global symbol calculus.
8. At least for a projectible Lagrangian submanifold $L$ of $T^*Q$, one finds representations of the star product algebras on the functions on $L$. This leads to explicit formulas for the WKB expansion corresponding to this Lagrangian submanifold.
9. The relation between configuration space symmetries, the corresponding phase-space reduction, and the reduced star products has been analyzed extensively by Kowalzig, Neumaier, and Pflaum.

*See also:* Classical *r*-Matrices, Lie Bialgebras, and Poisson Lie Groups; Deformation Quantization; Deformation Quantization and Representation Theory; Deformation Theory; Fedosov Quantization; Operads.

## Further Reading

Bayen F, Flato M, Frønsdal C, Lichnerowicz A, and Sternheimer D (1978) Deformation theory and quantization. *Annals of Physics* 111: 61–151.

Cattaneo A (Notes By Indelicato D) Formality and star products. In: Gutt S, Rawnsley J, and Sternheimer D (eds.) *Poisson Geometry, Deformation Quantisation and Group Representations*. LMS Lecture Note Series 323, pp. 79–144. Cambridge: Cambridge University Press.

Dito G and Sternheimer D (2002) Deformation quantization: genesis, developments and metamorphoses. In: Halbout G (ed.) *Deformation Quantization*, IRMA Lectures in Mathematics and Theoretical Physics, vol. 1, pp. 9–54. Berlin: Walter de Gruyter.

Gutt S (2000) Variations on deformation quantization. In: Dito G and Sternheimer D (eds.) *Conférence Moshé Flato 1999. Quantization, Deformations, and Symmetries*, Mathematical Physics Studies, vol. 21, pp. 217–254. Dordrecht: Kluwer Academic.

Waldmann S (2005) States and representations in deformation quantization. *Reviews in Mathematical Physics* 17: 15–75.

# $\bar{\partial}$-Approach to Integrable Systems

**P G Grinevich**, L D Landau Institute for Theoretical Physics, Moscow, Russia



## Introduction

The $\bar{\partial}$-approach is one of the most generic methods for constructing solutions of completely integrable systems. Taking into account that most soliton systems are represented as compatibility condition for a set of linear differential operators (Lax pairs, zero-curvature representations, L–A–B Manakov triples), it is sufficient to construct these operators. Such compatible families can be defined by presenting their common eigenfunctions. If it is possible to show that some analytic constraints imply that a function is a common eigenfunction of a family of operators, solutions of original nonlinear system are also generated.

The main idea of the $\bar{\partial}$ method is to impose the following analytic constraints: if $\lambda$ denotes the spectral parameter and $\boldsymbol{x}$ the physical variables, then, for arbitrary fixed values $\boldsymbol{x}$, the $\partial_{\bar{\lambda}}$ derivative of the wave function is expressed as a linear combination of the wave functions at other values of $\lambda$ with $\boldsymbol{x}$-independent coefficients. In specific examples, this property is either derived from the

direct spectral transform or imposed *a priori*. Of course, the specific realization of this scheme depends critically on the nonlinear system.

The origin of the $\bar{\partial}$-method came from the following observation. A solution of the one-dimensional inverse-scattering problem (the problem of reconstructing the potential by discrete spectrum and scattering amplitude at positive energies) for the one-dimensional time-independent Schrödinger operator

$$L = -\partial_x^2 + u(x) \qquad [1]$$

was obtained by Gelfand, Levitan, and Marchenko in the 1950s. It essentially used analytic continuation of the wave function from the real momenta to the complex ones. If the potential $u(x)$ decays sufficiently fast as $|x| \to \infty$, then the eigenfunction equation

$$L\psi(k,x) = k^2\psi(k,x) \qquad [2]$$

has two solutions $\psi_+(k,x)$ and $\psi_-(k,x)$ such that

1. $\psi_\pm(k,x) = (1 + o(1))e^{ikx}$ as $x \to \pm\infty$.
2. The functions $\psi_+(k,x), \psi_-(k,x)$ are holomorphic in $k$ in the upper half-plane and the lower half-plane, respectively.

Existence of analytic continuation to complex momenta is typical for one-dimensional systems. But in the multidimensional case the situation is different. For example, wave functions for the mutlidimensional Schrödinger operator constructed by Faddeev are well defined for all complex momenta $\boldsymbol{k}$, but they are nonholomorphic in $\boldsymbol{k}$, and they become holomorphic only after restriction to some special one-dimensional subspaces. The last property was one of the key points in the Faddeev approach.

Beals and Coifman (1981–82) and Ablowitz *et al.* (1983) discovered that departure from holomorphicity for multidimensional wave functions can be interpreted as spectral data. Such spectral transforms proved to be very natural and suit perfectly the purposes of the soliton theory. Some other famous methods, including the Riemann–Hilbert problem, can be interpreted as special reductions of the $\bar{\partial}$ method.

## Nonlocal $\bar{\partial}$-Problem and Local $\bar{\partial}$-Problem

The most generic formulation of the $\bar{\partial}$-method is the nonlocal $\bar{\partial}$-problem. Assume that the following data is given:

1. A rational $n \times n$ matrix-valued normalization function $\eta(\lambda)$.
2. An $n \times n$ matrix-valued function $g(\lambda, \boldsymbol{x})$ (it describes the dynamics) such that
   - $g(\lambda, \boldsymbol{x})$ depends on the spectral parameter $\lambda \in \mathbb{C}$ and "physical" variables $\boldsymbol{x} = (x_1, \ldots, x_N)$; the physical variables $x_k$ are either continuous ($x_k$ belongs to a domain in $\mathbb{R}$ or in $\mathbb{C}$) or discrete ($x_k$ takes integer values);
   - $g(\lambda, \boldsymbol{x})$ is analytic in $\lambda$, defined for all $\lambda \in \mathbb{C}$, except for a finite number of singular points, and is single valued; and
   - $\det g(\lambda, \boldsymbol{x})$ has only finite number of zeros.

   For problems with continuous physical variables the typical form of $g(\lambda, \boldsymbol{x})$ is $g(\lambda, \boldsymbol{x}) = \exp(\sum_i x_j K_j(\lambda))$, where $K_j(\lambda)$ are meromorphic matrices, mutually commuting for all $\lambda$. The discrete variables are usually encoded in orders of poles and zeros.
3. An $n \times n$ matrix-valued function $R(\lambda, \mu)$ – the "generalized spectral data." Usually, it is a regular function of four real variables $\Re\lambda, \Im\lambda, \Re\mu, \Im\mu$. (We write this as a function of two complex variables, but we do not assume it to be holomorphic. It would be more precise to write it as $R(\lambda, \bar{\lambda}, \mu, \bar{\mu})$, but to avoid long notations we omit the $\bar{\lambda}$, $\bar{\mu}$ dependence.) To avoid analytical complications, the function $R(\lambda, \mu)$ is usually assumed to vanish as $\lambda$ or $\mu$ tend to singular points of $\eta(\lambda)$, $g(\lambda, \boldsymbol{x})$.

Then the wave function $\Psi$ is defined by the data using the following properties:

1. $\Psi = \Psi(\lambda, \boldsymbol{x})$ takes values in complex $n \times n$ matrices:

$$\Psi(\lambda, \boldsymbol{x}) = \begin{bmatrix} \psi_{11}(\lambda, \boldsymbol{x}) & \ldots & \psi_{1n}(\lambda, \boldsymbol{x}) \\ \vdots & \cdots & \vdots \\ \psi_{n1}(\lambda, \boldsymbol{x}) & \ldots & \psi_{nn}(\lambda, \boldsymbol{x}) \end{bmatrix} \qquad [3]$$

2. For all $\lambda \in \mathbb{C}$ outside the singular points, the $\eta(\lambda)$, $g(\lambda, \boldsymbol{x})$ wave function $\Psi$ satisfies the $\bar{\partial}$-equation of inverse-spectral problem,

$$\frac{\partial\Psi(\lambda, \boldsymbol{x})}{\partial\bar{\lambda}} = \iint_{\mu\in\mathbb{C}} d\mu \wedge d\bar{\mu}\, \Psi(\mu, \boldsymbol{x}) R(\lambda, \mu) \qquad [4]$$

   It is important that condition [4] is $\boldsymbol{x}$-independent.
3. The function $\chi(\lambda, \boldsymbol{x}) - \eta(\lambda)$, where $\chi(\lambda, \boldsymbol{x}) = \Psi(\lambda, \boldsymbol{x}) g^{-1}(\lambda, \boldsymbol{x})$, is regular for all $\lambda \in \mathbb{C}$ and

$$\chi(\lambda, \boldsymbol{x}) - \eta(\lambda) \to 0 \quad \text{as } |\lambda| \to \infty \qquad [5]$$

The wave function $\Psi(\lambda, \boldsymbol{x})$ is calculated by employing the data $\eta(\lambda), g(\lambda, \boldsymbol{x}), R(\lambda, \mu)$ using the following procedure. Taking into account that the

functions $\eta(\lambda)$, $g(\lambda, \boldsymbol{x})$ are holomorphic in $\lambda$, eqn [4] can be rewritten in terms of $\chi(\lambda, \boldsymbol{x})$:

$$\frac{\partial[\chi(\lambda, \boldsymbol{x}) - \eta(\lambda)]}{\partial \bar{\lambda}} = \iint_{\mu \in \mathbb{C}} \mathrm{d}\mu \wedge \mathrm{d}\bar{\mu}\, \chi(\mu, \boldsymbol{x}) g(\lambda, \boldsymbol{x}) \times R(\lambda, \mu) g^{-1}(\mu, \boldsymbol{x}) \qquad [6]$$

The right-hand side of [6] is regular; therefore, this relation is valid for all complex $\lambda$ values.

Equation [6] with the boundary condition [5] is equivalent to the following integral equation:

$$\chi(\lambda, \boldsymbol{x}) = \eta(\lambda) + \frac{1}{2\pi \mathrm{i}} \iint_{\nu \in \mathbb{C}} \frac{\mathrm{d}\nu \wedge \mathrm{d}\bar{\nu}}{\nu - \lambda} \times \iint_{\mu \in \mathbb{C}} \mathrm{d}\mu \wedge \mathrm{d}\bar{\mu}\, \chi(\mu, \boldsymbol{x}) g(\nu, \boldsymbol{x}) \times R(\nu, \mu) g^{-1}(\mu, \boldsymbol{x}) \qquad [7]$$

This equation can be derived using the generalized Cauchy formula. Let $f(z)$ be a smooth (not necessarily holomorphic) function in a bounded domain $D$ in the complex plane. Then

$$f(z) = \frac{1}{2\pi \mathrm{i}} \oint_{\partial D} \frac{\mathrm{d}\chi}{\chi - z} f(\chi) + \frac{1}{2\pi \mathrm{i}} \iint_{D} \frac{\mathrm{d}\chi \wedge \mathrm{d}\bar{\chi}}{\chi - z} \frac{\partial f(\chi)}{\partial \bar{\chi}} \qquad [8]$$

If the kernel $g(\lambda, \boldsymbol{x}) R(\lambda, \mu) g^{-1}(\mu, \boldsymbol{x})$ is sufficiently good (e.g., it is sufficient to assume, that $(1 + |\lambda|)^{1+\epsilon} g(\lambda, \boldsymbol{x}) R(\lambda, \mu) g^{-1}(\mu, \boldsymbol{x})(1 + \mu\bar{\mu})^2$, $\epsilon > 0$, is a continuous function at both finite and infinite points), then we have a Fredholm equation (the operator on the right-hand side of [7] is compact). If it has no unit eigenvalues, eqn [7] is uniquely solvable. But, for some values of $\boldsymbol{x}$, one of the eigenvalues may become equal to 1, and it results in singularities of solutions.

If the norm of the integral operator is smaller than 1, eqn [7] is uniquely solvable. To generate solutions that are regular for all values of physical variables, it is natural to restrict the class of admissible spectral data by assuming the kernel $g(\lambda, \boldsymbol{x}) R(\lambda, \mu) g^{-1}(\mu, \boldsymbol{x})$ to be bounded in $\boldsymbol{x}$ for all $\lambda, \mu$. In the scalar case $n = 1$, this restriction implies:

$$R(\lambda, \mu) = 0 \quad \text{for all } \lambda, \mu \text{ such that } g(\lambda, \boldsymbol{x}) g^{-1}(\mu, \boldsymbol{x}) \text{ is unbounded in } \boldsymbol{x} \qquad [9]$$

For specific examples like the Kadomtsev–Petviashvili-II (KP-II), direct scattering transform automatically generates spectral data satisfying [9]. In KP-II, [9] implies

$$R(\lambda, \mu) = A(\lambda)\delta(\lambda - \mu) + T(\lambda)\delta(\lambda - \bar{\mu}) \qquad [10]$$

The coefficient $A(\lambda)$ can be eliminated by multiplying the wave function to an appropriate function of $\lambda$; therefore, in standard texts, $A(\lambda) \equiv 0$.

If for every $\lambda$ the kernel $R(\lambda, \mu)$ is equal to 0 everywhere except at finite number of points $\mu_1(\lambda), \ldots, \mu_k(\lambda)$, one has the so-called local $\bar{\partial}$-problem. Such kernels are rather typical.

## Examples of Soliton Systems Integrable by the $\bar{\partial}$-Problem Method

Let us discuss some important examples.

### The KP-II Hierarchy

The first nontrivial equation from the KP hierarchy has the following form:

$$(u_t + 6uu_x - u_{xxx})_x = 3\alpha^2 u_{yy} \qquad [11]$$

From a physical point of view, the case of real $\alpha^2$ is the most interesting one. Equation [11] is called KP-I if $\alpha^2 = -1$ and KP-II if $\alpha^2 = +1$. The Lax pair for KP-II reads:

$$[L, A] = 0$$

where

$$\begin{aligned} L &= \partial_y - \partial_x^2 + u(x, y, t) \\ A &= \partial_t - 4\partial_x^3 + 6u(x, y, t)\partial_x + 3u_x(x, y, t) + 3w(x, y, t) \end{aligned} \qquad [12]$$

The Cauchy problem for initial data $u(x, y, 0)$ decaying at infinity is solved by using the nonlocal Riemann problem for KP-I and local $\bar{\partial}$-problem for KP-II. The wave function is assumed to be scalar valued ($n = 1$), and $\bar{\partial}$-equation [4] takes the following form:

$$\frac{\partial \Psi(\lambda, x, y, t)}{\partial \bar{\lambda}} = T(\lambda) \Psi(\bar{\lambda}, x, y, t) \qquad [13]$$

The wave function $\Psi(\lambda, x, y, t)$ is assumed to be regular for finite $\lambda$'s and to have the following essential singularity as $\lambda = \infty$:

$$\Psi(\lambda, x, y, t) = \exp(\lambda x + \lambda^2 y + 4\lambda^3 t)(1 + o(1)) \qquad [14]$$

Equivalently, $\eta(\lambda) \equiv 1$ and the function $g(\lambda, x, t)$ has one essential singularity at $\lambda = \infty$,

$$g(\lambda, x, t) = \exp(\lambda x + \lambda^2 y + 4\lambda^3 t) \qquad [15]$$

Higher times $t_k$ from the KP hierarchy are incorporated into this scheme by assuming that

$$g(\lambda, \boldsymbol{t}) = \exp\left(\sum_{k=1}^{\infty} \lambda^k t_k\right) \qquad [16]$$

Here $x = t_1$, $y = t_2$, $t = 4t_3$.

Equation [13] was originally derived (Ablowitz *et al.* 1983) from the direct spectral transform. If the potential $u(x, y)$ is sufficiently small and $u(x, y) = O(1/(x^2 + y^2)^{1+\epsilon})$ for $x^2 + y^2 \to \infty$, then the wave function $\Psi(\lambda, x, y)$ for the $L$-operator [12]

$$\begin{aligned} &L\Psi(\lambda, x, y) = 0 \\ &\quad \Psi(\lambda, x, y) = \exp(\lambda x + \lambda^2 y)[1 + o(1)] \\ &\quad \text{for } x^2 + y^2 \to \infty \end{aligned} \qquad [17]$$

can be constructed by solving the following integral equation for the pre-exponent $\chi(\lambda, x, y) = \Psi(\lambda, x, y) \exp(-\lambda x - \lambda^2 y)$:

$$\begin{aligned} \chi(\lambda, x, y) = 1 - \iint G(\lambda, x - x', y - y') u(x, y) \\ \times \chi(\lambda, x, y)\, \mathrm{d}x'\, \mathrm{d}y' \end{aligned} \qquad [18]$$

where the Green function $G(\lambda, x, y)$ is given by

$$G(\lambda, x, y) = \frac{1}{4\pi^2} \iint \frac{\mathrm{e}^{\mathrm{i}(p_x x + p_y y)}}{p_x^2 + \mathrm{i}p_y - 2\mathrm{i}\lambda p_x} \mathrm{d}p_x\, \mathrm{d}p_y \qquad [19]$$

It is not holomorphic in $\lambda$, but

$$\frac{\partial G(\lambda, x, y)}{\partial \bar{\lambda}} = -\frac{\mathrm{i}}{2\pi} \mathrm{sgn}(\Re\lambda)\, \mathrm{e}^{-2\mathrm{i}\Re\lambda x - 4\mathrm{i}\Re\lambda\Im\lambda y} \qquad [20]$$

The nonholomorphicity of $G(\lambda, x, y)$ results in the special nonholomorphicity of $\Psi(\lambda, x, y)$ of the form [13].

**Remark** We see that one function of two real variables $T(\lambda)$ is sufficient to solve the Cauchy problem in the plane. But it is also possible to construct solutions of KP-II starting from generic nonlocal kernels $R(\lambda, \mu)$ (to guarantee at least local existence of solutions, it is enough to assume that $R(\lambda, \mu)$ is small and has finite support). It looks like a paradox, but the situation is exactly the same in the linear case. In the standard Fourier method, only exponents with real momenta are used, but local solutions can be constructed as combinations of exponents with arbitrary complex momenta.

#### Novikov–Veselov Hierarchy and Two-Dimensional Schrödinger Operator

Equations from this hierarchy admit representation in terms of Manakov L–A–B triples,

$$\frac{\partial L}{\partial t_n} = [A_n, L] + B_n L \qquad [21]$$

where

$$\begin{aligned} L &= -4\partial_z \partial_{\bar{z}} + u(z, t) \\ A_n &= 2^{2n+1}\left(\partial_z^{2n+1} + \partial_{\bar{z}}^{2n+1}\right) + \cdots \end{aligned} \qquad [22]$$

The order of $B_n$ is smaller than $2n + 1$. In particular, for $n = 1$,

$$\begin{aligned} A_1 &= 8\left(\partial_z^3 + \partial_{\bar{z}}^3\right) + 2(w\partial_z + \bar{w}\partial_{\bar{z}}) \\ B_1 &= w_z + \bar{w}_{\bar{z}} \end{aligned} \qquad [23]$$

$$u_t = 8\partial_z^3 u + 8\partial_{\bar{z}}^3 u + 2\partial_{\bar{z}}(uw) + 2\partial_z(u\bar{w}) \qquad [24]$$

where

$$u(z, t) = \bar{u}(z, t), \qquad \partial_{\bar{z}} w(z, t) = -3\partial_{\bar{z}} u(z, t) \qquad [25]$$

This hierarchy is integrated using the scattering transform at zero energy for the two-dimensional Schrödinger operator $L$. If Cauchy data with asymptotic

$$u(z) \to -E_0, \quad w(z) \to 0, \qquad \text{for } |z| \to \infty \qquad [26]$$

is considered, the scattering transform for the operator $\tilde{L} = L + E_0$ with the potential $\tilde{u}(z) = u(z) + E_0$ at fixed energy $E_0$ and decaying at infinity is used. In fact, the fixed-energy scattering problem is one of the basic problems of mathematical physics, and the Novikov–Veselov hierarchy can be treated as an infinite-dimensional Abel symmetry algebra for this problem. The scattering transform essentially depends on the sign of $E_0$. The case $E_0 = 0$, studied by Boiti, Leon, Manna, and Pempinelli is the most complicated from the analytic point of view, and we do not discuss it now.

If $E_0 < 0$, the wave function satisfies a pure local $\bar{\partial}$-relation:

$$\frac{\partial \Psi(\lambda, z)}{\partial \bar{\lambda}} = T(\lambda)\Psi\left(\frac{1}{\bar{\lambda}}, z\right) \qquad [27]$$

with $\eta(\lambda) \equiv 1$, and

$$g(\lambda, z) = \mathrm{e}^{(\kappa/2)(\lambda\bar{z} + z/\lambda)}, \quad \kappa^2 = -E_0 \qquad [28]$$

Starting from generic spectral data $T(\lambda)$, one obtains a fixed-energy eigenfunction for a second-order operator,

$$\begin{aligned} \tilde{L}\Psi(\lambda, z) &= E_0 \Psi(\lambda, z) \\ L &= -4\partial_z \partial_{\bar{z}} + V(z)\partial_z + \tilde{u}(z) \end{aligned} \qquad [29]$$

To generate pure potential operators ($V(z) \equiv 0$), it is necessary to impose additional symmetry constraints of the spectral data (see the section "Reductions on the $\bar{\partial}$ data").

If $E_0 > 0$, there are two types of generalized spectral data – $\bar{\partial}$-data and nonlocal Riemann problem data. The wave function satisfies the $\bar{\partial}$-relation:

$$\frac{\partial \Psi(\lambda, z)}{\partial \bar{\lambda}} = T(\lambda)\Psi\left(-\frac{1}{\bar{\lambda}}, z\right), \quad |\lambda| \neq 1 \qquad [30]$$

and has a jump at the unit circle $|\lambda| = 1$. The boundary values $\Psi_{\pm}(\lambda, z) = \Psi(\lambda(1 \pm 0), z)$ are connected by the following relation:

$$\Psi_{+}(\lambda, z) = \Psi_{-}(\lambda, z) + \oint_{|\mu|=1} R(\lambda, \mu)\Psi_{-}(\mu, z)|d\mu| \qquad [31]$$

$$g(\lambda, z) = e^{i(\kappa/2)(\lambda\bar{z} + z/\lambda)}, \quad \kappa^2 = E_0 \qquad [32]$$

Constraints on the spectral data associated with pure potential operators were found by Novikov and Grinevich for $R(\lambda, \mu)$, and by Manakov and Grinevich for $T(\lambda)$. Existence of two different types of generalized scattering data has a very transparent physical meaning: there is a one-to-one correspondence between the classical scattering amplitude at energy $E_0$ and the nonlocal Riemann problem data $R(\lambda, \mu)$. The $\bar{\partial}$-data $T(\lambda)$ can be treated as a complete set of additional parameters enumerating all potentials with a given scattering amplitude at one energy.

### Davey–Stewartson-II and Ishimori-I Equations

The Davey–Stewartson-II (DS-II) equation

$$i\partial_t q + 2\left(\partial_z^2 + \partial_{\bar{z}}^2\right)q + (g + \bar{g})q = 0 \qquad [33]$$

$$\partial_{\bar{z}} g = -\kappa \partial_z |q|^2 \qquad [34]$$

$$q = q(z, t), \qquad g = g(z, t), \qquad z = x + iy \qquad [35]$$

can be treated as an integrable $(2 + 1)$-dimensional extension of nonlinear Schrödinger equation. The Ishimori-I equation

$$\partial_t S + S \times \left(\partial_x^2 S + \partial_y^2 S\right) + \partial_x w \partial_y S + \partial_y w \partial_x S = 0 \qquad [36]$$

$$\partial_x^2 w - \partial_y^2 w + 2S(\partial_x S \times \partial_y S) = 0 \qquad [37]$$

$$S = S(x, y, t), \quad S = (S_1, S_2, S_3), \quad S^2 = 1 \qquad [38]$$

is an integrable $(2 + 1)$-dimensional extension of the Heisenberg magnetic equation. Both systems are integrated by using the following zero-curvature representation:

$$\begin{pmatrix} \partial_{\bar{z}} & 0 \\ 0 & \partial_z \end{pmatrix}\Psi = \frac{1}{2}\begin{pmatrix} 0 & q(z,t) \\ \kappa\overline{q(z,t)} & 0 \end{pmatrix}\Psi \qquad [39]$$

$$\partial_t \Psi = \begin{pmatrix} 2i\partial_z^2 + ig & iq_{\bar{z}} - iq\partial_{\bar{z}} \\ -i\kappa\bar{q}_z + i\kappa q\partial_z & -2i\partial_{\bar{z}}^2 - ig \end{pmatrix}\Psi \qquad [40]$$

The wave function satisfies the following "scattering" equation:

$$\begin{pmatrix} \partial_{\bar{k}} & 0 \\ 0 & \partial_k \end{pmatrix}\Psi^{T} = \begin{pmatrix} 0 & \kappa\bar{b}(k) \\ b(k) & 0 \end{pmatrix}\Psi^{T} \qquad [41]$$

Here $\Psi^{T}$ denotes the transposed matrix. Let us point out the amazing symmetry between the direct and inverse transforms.

### Discrete Systems

In the examples discussed above, continuous variables are "encoded" in essential singularities of $g(\lambda, \boldsymbol{x})$. Discrete variables correspond to orders of zeros and poles. For example, assuming that the function $g(\lambda, \boldsymbol{t})$ in the KP integration scheme depends on extra continuous variables $t_{-1}, t_{-2}, \ldots, t_{-n}, \ldots$ and discrete variable $t_0 = n$,

$$g(\lambda, \boldsymbol{t}) = \lambda^{t_0} \exp\left(\sum_{k \neq 0}^{\infty} \lambda^k t_k\right) \qquad [42]$$

one obtains solutions of the so-called two-dimensional Toda–KP hierarchy.

Assume that we have a nonlocal $\bar{\partial}$-problem for a scalar function with $\eta \equiv 1$ and

$$g(\lambda, n_1, \ldots, n_k) = \prod_{j=1}^{k}\left(\frac{\lambda - P_k}{\lambda - Q_k}\right)^{n_k} \qquad [43]$$

The wave function defines a map $\mathbb{Z}^k \to \mathbb{C}^N$,

$$(n_1, \ldots, n_k) \to (\Psi(\lambda_1, n_1, \ldots, n_k), \ldots, \Psi(\lambda_N, n_1, \ldots, n_k)) \qquad [44]$$

where $\lambda_1, \ldots, \lambda_N$ are some points in $\mathbb{C}$. This construction generates the so-called quadrilateral lattices (each two-dimensional face is planar).

### Multidimensional Problems

The $\bar{\partial}$-approach can also be applied to multidimensional inverse-scattering problems, but typically the scattering data are overdetermined and satisfy additional nonlinear compatibility conditions. For

example, the Faddeev wave functions for the $n$-dimensional stationary Schr̈odinger operator

$$\left(-\partial_{x_1}^2 - \cdots - \partial_{x_n}^2 + u(\boldsymbol{x})\right)\Psi(\boldsymbol{k},\boldsymbol{x}) = (\boldsymbol{k}\cdot\boldsymbol{k})\Psi(\boldsymbol{k},\boldsymbol{x}) \quad [45]$$

$$\Psi(\boldsymbol{k},\boldsymbol{x}) = e^{i\boldsymbol{k}\cdot\boldsymbol{x}}(1+o(1)) \quad [46]$$

in the nonphysical domain $\boldsymbol{k}_I \neq 0$ ($\boldsymbol{k}_R$ and $\boldsymbol{k}_I$ denote the real and imaginary parts of $\boldsymbol{k}$, respectively) satisfy the following $\bar{\partial}$-equation:

$$\frac{\partial\Psi(\boldsymbol{k},\boldsymbol{x})}{\partial\bar{k}_j} = -2\pi\int_{\boldsymbol{\xi}\in\mathbb{R}^n}\xi_j h(\boldsymbol{k},\boldsymbol{k}_R+\boldsymbol{\xi})\Psi(\boldsymbol{k}+\boldsymbol{\xi},\boldsymbol{x}) \times\delta(\boldsymbol{\xi}\cdot\boldsymbol{\xi}+2\boldsymbol{k}\cdot\boldsymbol{\xi})\mathrm{d}\xi_1\cdots\mathrm{d}\xi_n \quad [47]$$

The characterization of admissible spectral data $h(\boldsymbol{k},\boldsymbol{l}), \boldsymbol{k}\in\mathbb{C}^n, \boldsymbol{l}\in\mathbb{R}^n$ is based on the following compatibility equation:

$$\frac{\partial h(\boldsymbol{k},\boldsymbol{l})}{\partial\bar{k}_j} + \frac{1}{2}\frac{\partial h(\boldsymbol{k},\boldsymbol{l})}{\partial l_j} = -2\pi\int_{\boldsymbol{\xi}\in\mathbb{R}^n}\xi_j h(\boldsymbol{k},\boldsymbol{k}_R+\boldsymbol{\xi})h(\boldsymbol{k}+\boldsymbol{\xi},\boldsymbol{l}) \times\delta(\boldsymbol{\xi}\cdot\boldsymbol{\xi}+2\boldsymbol{k}\cdot\boldsymbol{\xi})\mathrm{d}\xi_1\cdots\mathrm{d}\xi_n \quad [48]$$

More details can be found in Novikov and Henkin (1987).

## Reductions on the $\bar{\partial}$-Data

The most generic $\bar{\partial}$-data usually result in solutions from wrong functional class (they may, e.g., be complex or singular), or extra constraints on the auxiliary linear operators are necessary to obtain solutions of the zero-curvature representation. For example, to obtain real KP-II solutions using the local $\bar{\partial}$-problem [13], the following reduction on the $\bar{\partial}$-data should be implied:

$$T(\bar{\lambda}) = -\bar{T}(\lambda) \quad [49]$$

It can be easily derived from the direct transform. But it is not always the case. For example, the selection of pure potential two-dimensional Schrödinger operators originally was not so evident. To formulate the answer, it is convenient to introduce a new function $b(\lambda), T(\lambda) = b(\lambda)\pi\,\mathrm{sgn}(\lambda\bar{\lambda}-1)/\bar{\lambda}$.

For $E_0 < 0$, the following constraints select real potential operators:

$$b\left(-\frac{1}{\bar{\lambda}}\right) = b(\lambda), \qquad b\left(\frac{1}{\bar{\lambda}}\right) = \bar{b}(\lambda) \quad [50]$$

In some situations, the problem of finding appropriate reductions is the most difficult part of the integration procedure. It is true not only for the $\bar{\partial}$-approach, but also for other techniques including the finite-gap method. For example, the inverse-spectral transform for the two-dimensional Schrödinger operator was first developed for finite-gap (quasiperiodic) potentials and only later for the decaying ones. For operators with finite gap at one energy the first-order terms were constructed by Dubrovin, Krichever, and Novikov in 1976, but only in 1984 the potentiality reduction was found by Novikov and Veselov.

## Nonsingular Solutions

As mentioned above, one can construct regular solutions by choosing sufficiently small (in an appropriate norm) scattering data. But for some special systems the regularity follows automatically from reality reductions. For example, for arbitrary large $\bar{\partial}$-data, real KP-II solutions constructed by the local $\bar{\partial}$-problem [13] with reduction [49] are regular. The proof is based on the theory of generalized analytic functions (in the Vekua sense). Another example is the two-dimensional Schr̈odinger operator at a fixed negative energy $E_0 < 0$. The potentiality and reality constraints imply that the potential is nonsingular for arbitrary large $T(\lambda)$. But, unfortunately, the $\bar{\partial}$-problem with regular data covers only a part of the space of potentials. In fact, each such operator possesses a strictly positive real eigenfunction at the level $E_0$, exponentially growing in all directions (it also follows from the generalized analytic functions theory). Existence of such function implies that the whole discrete spectrum is located above the energy $E_0$, and it gives a restriction on the potential. (For more details, see the review by Grinevich (2000).)

## Some Explicit Solutions

The generic $\bar{\partial}$-problem results in potentials that could not be expressed in terms of elementary or standard special functions. But for degenerate kernels, a solution of the inverse-spectral problem can be written explicitly. For example, if

$$R(\lambda,\mu) = \sum_{k=1}^{n} r_k(\lambda)s_k(\mu) \quad [51]$$

the wave function and solutions can be expressed in quadratures.

In particular, if all $r_k(\lambda)$ and $s_k(\mu)$ are $\delta$-functions, $r_k(\lambda) = R_k\delta(\lambda-\lambda_k)$ and $s_k(\mu) = S_k\delta(\mu-\mu_k)$, the wave function is rational in $\lambda$ and can be expressed as a rational combination of exponents of $x_k$. If for some $k$ and $l$, $\lambda_k = \mu_l$, this procedure needs some

regularization. For example, it is possible to assume, that $\delta(\lambda-\lambda_0)/(\lambda-\lambda_0)=0$.

If for all $k$, $\lambda_k=\mu_k$, the $\bar{\partial}$-problem generates rational in $\boldsymbol{x}$ solutions (lumps). It is possible to show that, the Novikov–Veselov real rational solitons for $E_0>0$ are always nonsingular, decay at $\infty$ as $1/(x^2+y^2)$, and the potential $u(z)$ has zero scattering in all directions for the waves with energy $E_0$.

## The $\bar{\partial}$-Problem on Riemann Surfaces

In all examples discussed above, the spectral variable is defined in a Riemann sphere. It is natural to generalize it by considering wave functions depending on a spectral parameter defined on a Riemann surface of higher genus. Spectral transforms of such type arise in the theory of localized perturbation of periodic solutions. Assume that the KP-II potential $u(x,y)$ has the form

$$u(x,y)=u_0(x,y)+u_1(x,y) \qquad [52]$$

where $u_0(x,y)$ is a real nonsingular finite-gap potential and $u_1(x,y)$ decays sufficiently fast at infinity. Denote by $\Psi_0(\gamma,x,y)$ the wave function of the operator $L_0=\partial_y-\partial_x^2+u_0(x,y)$, where $\gamma\in\Gamma$, the spectral curve $\Gamma$ is a compact Riemann surface of genus $g$ with a distinguished point $\infty$. In addition to essential singularity at the point $\infty$, the wave function $\Psi_0(\gamma,x,y)$ has $g$ simple poles at points $\gamma_1,\ldots,\gamma_g$ and is holomorphic in $\gamma$ outside these singular points. For a real nonsingular potential, $\Gamma$ is an $M$-curve, that is, there exists an antiholomorphic involution $\tau:\Gamma\to\Gamma, \tau\infty=\infty$, the set of fixed points form $g+1$ ovals $a_0,\ldots,a_g,\infty\in a_0,\gamma_k\in a_k$. The wave function $\Psi(\gamma,x,y)$ of the perturbed operator $L=\partial_y-\partial_x^2+u(x,y)$ is defined at the same spectral curve $\Gamma$, but it is not holomorphic any more. It has the following properties:

1. At the point $\infty$, the wave function $\Psi(\gamma,x,y)$ has an essential singularity: $\Psi(\gamma,x,y)=\Psi_0(\gamma,x,y)(1+o(1))$.
2. In the neighborhoods of the points $\gamma_k,\Psi(\gamma,x,y)$ can be written as a product of a continuous function by a simple pole at $\gamma_k$.
3. The wave function $\Psi(\gamma,x,y)$ satisfies the $\bar{\partial}$ equation

$$\frac{\partial\Psi(\gamma,x,y,t)}{\partial\bar{\gamma}}\mathrm{d}\bar{\gamma}=T(\gamma)\Psi(\tau\gamma,x,y,t) \qquad [53]$$

where the (0, 1)-form $T(\gamma)=t(\gamma)\mathrm{d}\bar{\gamma}$ is regular outside the divisor points $\gamma_k$ and in the neighborhood of $\gamma_k$ it possible to define local coordinate such that $t(\gamma)=\mathrm{sgn}(\Im\gamma)t_1(\gamma)\overline{(\gamma-\gamma_k)}/(\gamma-\gamma_k), t_1(\gamma)$ is regular.

A solution of the inverse problem can be obtained by using appropriate analogs of Cauchy kernels on Riemann surfaces.

## Quasiclassical Limit

The systems integrable by the $\bar{\partial}$-method usually describe integrable systems with high-order derivatives. It is well known that by applying some limiting procedures to integrable systems one can construct new completely integrable equations, but integration methods for these equations are based on completely different analytic tools. One of most important examples is the theory of dispersionless hierarchies. The limiting procedure for the $\bar{\partial}$-problem (quasiclassical $\partial$-problem) was developed by Konopelchenko and collaborators. In the KP case, the quasiclassical limit of the wave function $\Psi(\lambda,\boldsymbol{t})$ is assumed to have the following form:

$$\Psi\left(\lambda,\frac{\boldsymbol{t}}{\epsilon}\right)=\hat{\chi}(\lambda,\boldsymbol{t},\epsilon)\exp\left(\frac{S(\lambda,\boldsymbol{t})}{\epsilon}\right) \qquad [54]$$

It is possible to show that the function $S(\lambda,\boldsymbol{t})$ satisfies a Beltrami-type equation:

$$\frac{S(\lambda,\boldsymbol{t})}{\partial\bar{\lambda}}=W\left(\lambda,\frac{S(\lambda,\boldsymbol{t})}{\partial\lambda}\right) \qquad [55]$$

which is treated as a dispersionless limit of [13]. Higher-order corrections were also discussed in the literature (see Konopelchenko and Moro (2003)).

*See also:* Boundary-Value Problems for Integrable Equations; Integrable Systems and Algebraic Geometry; Integrable Systems and the Inverse Scattering Method; Integrable Systems: Overview; Integrable Systems and Discrete Geometry; Riemann–Hilbert Methods in Integrable Systems.

## Further Reading


Ablowitz MJ, Bar Jaakov D, and Fokas AS (1983) On the inverse scattering of the time-dependent Schrödinger equation and the associated Kadomtsev–Petviashvili equation. *Studies in Applied Mathematics* 69(2): 135–143.

Beals R and Coifman RR (1981–82) Scattering, transformations spectrales et equations d'evolution nonlineare. I, II: Seminaire Goulaouic – Meyer–Schwartz – Exp. 22; Exp. 21. Palaiseau: Ecole Polytechnique.

Beals R and Coifman RR (1985) Multidimensional inverse scattering and nonlinear partial differential equations. *Proceedings of Symposia in Pure Mathematics* 43: 45–70.

Bogdanov LV and Konopelchenko BG (1995) Lattice and $q$-difference Darboux–Zakharov–Manakov systems via $\bar{\partial}$-dressing method. *Journal of Physics A* 28(5): L173–L178.

Bogdanov LV and Manakov SV (1988) The nonlocal $\bar{\partial}$ problem and $(2+1)$-dimensional soliton equations. *Journal of Physics A* 21(10): L537–L544.

Grinevich PG (2000) The scattering transform for the two-dimensional Schrodinger operator with a potential that decreases at infinity at fixed nonzero energy. *Russian Mathematical Surveys* 55(6): 1015–1083.

Konopelchenko B and Moro A (2003) Quasi-classical $\bar{\partial}$-dressing approach to the weakly dispersive KP hierarchy. *Journal of Physics A* 36(47): 11837–11851.

Nachman AJ and Ablowitz MJ (1984) A multidimensional inverse-scattering method. *Studies in Applied Mathematics* 71(3): 243–250.

Novikov RG and Henkin GM (1987) The $\partial$-equation in the multidimensional inverse scattering problem. *Russian Mathematical Surveys* 42(4): 109–180.

# Derived Categories

**E R Sharpe**, University of Utah, Salt Lake City, UT, USA



## Introduction

In this article we shall briefly outline derived categories and their relevance for physics. Derived categories (and their enhancements) classify off-shell states in a two-dimensional topological field theory on Riemann surfaces with boundary known as the open-string B model. We briefly review pertinent aspects of that topological field theory and its relation to derived categories, the Bondal–Kapranov enhancement and its relation to the open-string B model, as well as B model twists of two-dimensional theories known as Landau–Ginzburg models, and how information concerning stability of D-branes is encoded in this language. We concentrate on more physical aspects of derived categories; for a very readable short review concentrating on the mathematics, see, for example, Thomas (2000).

## Sheaves and Derived Categories in the Open-String B Model

Derived categories are mathematical constructions which are believed to be related to D-branes in the open-string B model. We shall begin by briefly reviewing the B model, as well as D-branes.

The A and B models are two-dimensional topological field theories, closely related to nonlinear sigma models, which are supersymmetrizations of theories summing over maps from a Riemann surface (the world sheet of the string) into some "target space" $X$. In both the A and B models, one considers only certain special correlation functions, involving correlators closed under the action of a nilpotent scalar operator known as the "BRST operator," $Q$, which is part of the original supersymmetry transformations. In considering the pertinent correlation functions, only certain types of maps contribute. The A model has the properties of being invariant under complex structure deformations of the target space $X$, and its pertinent correlation functions are computed by summing over holomorphic maps into the target $X$. The A model will not be relevant for us here. The B model has the properties of being invariant under Kähler moduli of the target $X$, and its pertinent correlation functions are computed by summing over constant maps into the target $X$. In the closed-string B model, the states of the theory are counted by the cohomology groups $H^*(X, \Lambda^* TX)$, where $X$ is constrained to be Calabi–Yau. The BRST operator in the B model $Q$ can be identified with $\overline{\partial}$ for many purposes. The open-string B model is the same topological field theory, but now defined on a Riemann surface with boundaries. As with all open-string theories, we specify boundary conditions on the fields, which force the ends of the string to live on some submanifold of the target, and we associate to the boundaries degrees of freedom (known as the Chan–Paton factors) which describe a (possibly twisted) vector bundle over the submanifold. In the case of the B model, the submanifold is a complex submanifold, and the vector bundle is forced to be a holomorphic vector bundle over that submanifold.

To lowest order, that combination of a submanifold $S$ of $X$ together with a (possibly twisted) holomorphic vector submanifold, is a "D-brane" in the open-string B model. We shall denote the twisted bundle by $\mathcal{E} \otimes \sqrt{K_S}$, where $K_S$ is the canonical bundle of $S$, and the $\sqrt{K_S}$ factor is an explicit incorporation of something known as the Freed–Witten anomaly. Now, if $i: S \hookrightarrow X$ is the inclusion map, then to this D-brane we can associate a sheaf $i_* \mathcal{E}$.

Technically, a sheaf is defined by associating sets, or modules, or rings, to each open set on the underlying space, together with restriction maps saying how data associated to larger open sets restricts to smaller open sets, obeying the obvious consistency conditions, together with some gluing conditions that say how local sections can be patched back together. A vector bundle defines a

sheaf by associating to any open set sections of the bundle over that open set. Sheaves of the form "$i_*\mathcal{E}$" look like, intuitively, vector bundles over submanifolds, with vanishing fibers off the submanifold. A more detailed discussion of sheaves is beyond the scope of this article; see instead, for example, Sharpe (2003).

To "associate a sheaf" means finding a sheaf such that physical properties of the D-brane system are well modeled by mathematics of the sheaf. (In particular, the physical definition of D-brane has, on the face of it, nothing at all to do with the mathematical definition of a sheaf, so one cannot directly argue that they are the same, but one can still use one to give a mathematical model of the other.) For example, the spectrum of open-string states in the B model stretched between two D-branes, associated to sheaves $i_*\mathcal{E}$ and $j_*\mathcal{F}$, turn out to be calculated by a cohomology group known as $\mathrm{Ext}^*_X(i_*\mathcal{E}, j_*\mathcal{F})$.

There are many more sheaves not of the form $i_*\mathcal{E}$, that is, that do not look like vector bundles over submanifolds. It is not known in general whether they also correspond to (on-shell) D-branes, but in some special cases the answer has been worked out. For example, structure sheaves of nonreduced schemes turn out to correspond to D-branes with nonzero nilpotent Higgs vevs.

For a set of ordinary D-branes, the description above suffices. However, more generally one would like to describe collections of D-branes and anti-D-branes, and tachyons. An anti-D-brane has all the same physical properties as an ordinary D-brane, modulo the fact that they try to annihilate each other. The open-string spectrum between coincident D-branes and anti-D-branes contains tachyons. One can give an (off-shell) vacuum expectation value to such tachyons, and then the unstable brane–antibrane–tachyon system will evolve to some other, usually simpler, configuration. For example, given a single D-brane wrapped on a curve, with trivial line bundle, and an anti-D-brane wrapped on the same curve, with line bundle $\mathcal{O}(-1)$, and a nonzero tachyon $\mathcal{O}(-1) \to \mathcal{O}$, then one expects that the system will dynamically evolve to a smaller D-brane sitting at a point on the curve.

Now, one would like to find some mathematics that describes such systems, and gives information about the endpoints of their evolution. Technically, one would like to classify universality classes of world-sheet boundary renormalization group flow.

It has been conjectured that derived categories of sheaves provide such a classification. To properly explain derived categories is well beyond the scope of this article (see instead the "Further reading" section at the end), but we shall give a short outline below.

Mathematically, derived categories of sheaves concern complexes of sheaves, that is, sets of sheaves $\mathcal{E}_i$ together with maps $d_i : \mathcal{E}_i \to \mathcal{E}_{i+1}$

$$\cdots \longrightarrow \mathcal{E}_i \xrightarrow{d_i} \mathcal{E}_{i+1} \xrightarrow{d_{i+1}} \mathcal{E}_{i+2} \xrightarrow{d_{i+2}} \cdots$$

such that $d_{i+1} \circ d_i = 0$. A category is defined by a collection of "objects" together with maps between the objects, known as morphisms. In a derived category of coherent sheaves, the objects are complexes of sheaves, and the maps are equivalence classes of maps between complexes.

Physically, if the complex consists of locally free sheaves (equivalently, vector bundles), then we can associate a brane/antibrane/tachyon system, by identifying the $\mathcal{E}_i$ for $i$ even, say, with D-branes, and the $\mathcal{E}_i$ for $i$ odd with anti-D-branes. If the $\mathcal{E}_i$ are all locally free sheaves, then there are tachyons between the branes and antibranes, and we can identify the $d_i$'s with those tachyons. In the open-string world-sheet theory, giving a tachyon a vacuum expectation value modifies the BRST operator $Q$, and a necessary condition for the new theory to still be a topological field theory is that $Q^2 = 0$, a condition which turns out to imply that $d_{i+1} \circ d_i = 0$.

To re-create the structure of a derived category, we need to impose some equivalence relations. To see what sorts of equivalence relations one would like to impose, note the following. Physically, we would like to identify, for example, a configuration consisting of a brane, an antibrane, and a tachyon, which we can describe as a complex

$$\mathcal{O}(-D) \longrightarrow \mathcal{O}$$

with a one-element complex

$$\mathcal{O}_{\mathrm{D}}$$

corresponding to the D-brane which we believe is the endpoint of the evolution of the brane/antibrane configuration.

One natural mathematical way to create identifications of this form is to identify complexes that differ by "quasi-isomorphisms," meaning, a set of maps $(f^n : C^n \to D^n)$ compatible with $d$'s, and inducing an isomorphism $\tilde{f}^n : H^n(C) \cong H^n(D)$ on the cohomologies of the complexes. In particular, in the example above, there is a natural set of maps

$$\begin{array}{ccc} \mathcal{O}(-D) & \longrightarrow & \mathcal{O} \\ \downarrow & & \downarrow \\ 0 & \longrightarrow & \mathcal{O}_{\mathrm{D}} \end{array}$$

that define a quasi-isomorphism. More generally, in homological algebra, one typically does computations by replacing ordinary objects with projective or injective resolutions, that is, complexes with special properties, in which the desired computation becomes trivial, and defining the result for the original object to be the same as the result for the resolution. To formalize this procedure, one would like a mathematical setup in which objects and their projective and injective resolutions are isomorphic.

However, to define an equivalence relation, one usually needs an isomorphism, and the quasi-isomorphisms above are not, in general, isomorphisms. Creating an equivalence from nonisomorphisms, to resolve this problem, can be done through a process known as "localization" (generalizing the notion of localization in commutative algebra). The resulting equivalence relations on maps between complexes define the derived category.

The derived category is a category whose objects are complexes, and whose morphisms $C^{\cdot} \to D^{\cdot}$ are equivalence classes of pairs $(s, t)$ where $s : G^{\cdot} \to C^{\cdot}$ is a quasi-isomorphism between $C^{\cdot}$ and another complex $G^{\cdot}$, and $t : G^{\cdot} \to D^{\cdot}$ is a map of complexes. We take two such pairs $(s, t), (s', t')$ to be equivalent if there exists another pair $(r, h)$ between the auxiliary complexes $G^{\cdot}, G^{\cdot\prime}$, making the obvious diagram commute. This is, in a nutshell, what is meant by localization, and by working with such equivalence classes, this allows us to formally invert maps that are otherwise noninvertible. (We encourage the reader to consult the "Further reading" section for more details.)

Mathematically, this technology gives a very elegant way to rethink, for example, homological algebra. There is a notion of a derived functor, a special kind of functor between derived categories, and notions from homological algebra such as Ext and Tor can be re-expressed as cohomologies of the image complexes under the action of a derived functor, thus replacing cohomologies with complexes.

Physically, looking back at the physical realization of complexes, we see a basic problem: different representatives of (isomorphic) objects in the derived category are described by very different physical theories. For example, the sheaf $\mathcal{O}_D$ corresponds to a single D-brane, defined by a two-dimensional boundary conformal field theory (CFT), whereas the brane/antibrane/tachyon collection $\mathcal{O}(-D) \to \mathcal{O}$ is defined by a massive nonconformal two-dimensional theory. These are very different physical theories. If we want "localization on quasi-isomorphisms" to happen in physics, we have to explain which properties of the physical theories we are interested in, because clearly the entire physical theories are not and cannot be isomorphic.

Although the entire physical theories are not isomorphic, we can hope that under renormalization group flow, the theories will become isomorphic. That is certainly the physical content of the statement that the brane/antibrane system $\mathcal{O}(-D) \to \mathcal{O}$ should describe the D-brane corresponding to $\mathcal{O}_D$ – after world-sheet boundary renormalization group flow, the nonconformal two-dimensional theory describing the brane/antibrane system becomes a CFT describing a single D-brane.

More globally, this is the general prescription for finding physical meanings of many categories: we can associate physical theories to particular types of representatives of isomorphism classes of objects, and then although distinct representatives of the same object may give rise to very different physical theories, those physical theories at least lie in the same universality class of world-sheet renormalization group flow. In other words, (equivalence classes of) objects are in one-to-one correspondence with universality classes of physical theories.

Showing such a statement directly is usually not possible – it is usually technically impractical to follow renormalization group flow explicitly. There is no symmetry reason or other basic physics reason why renormalization group flow must respect quasi-isomorphism. The strongest constraint that is clearly applied by physics is that renormalization group flow must preserve D-brane charges (Chern characters, or more properly, $K$-theory), but objects in a derived category contain much more information than that.

However, indirect tests can be performed, and because many indirect tests are satisfied, the result is generally believed.

The reader might ask why it is not more efficient to just work with the cohomology complexes $H^{\cdot}(C)$ themselves, rather than the original complexes. One reason is that the original complexes contain more information than the cohomology – passing to cohomology loses information. For example, there exist examples of complexes that have the same cohomology, yet are not quasi-isomorphic, and so are not identified within the derived category, and so physically are believed to lie in different universality classes of boundary renormalization group flow.

Another motivation for relating physics to derived categories is Kontsevich's approach to mirror symmetry. Mirror symmetry relates pairs of Calabi–Yau manifolds, of the same dimension, in a fashion such that easy classical computations in one Calabi–Yau

are mapped to difficult "quantum" computations involving sums over holomorphic curves in the other Calabi–Yau. Because of this property, mirror symmetry has proved a fertile ground for algebraic geometers to study. Kontsevich proposed that mirror symmetry should be understood as a relation between derived categories of coherent sheaves on one Calabi–Yau and derived Fukaya categories on the other Calabi–Yau. At the time he made this proposal, no one had any idea how either could be realized in physics, but since that time, physicists have come to believe that Kontsevich was secretly talking about D-branes in the A and B models.

## Bondal–Kapranov Enhancements

Mathematically derived categories are not quite as ideal as one would like. For example, the cone construction used in triangulated categories does not behave functorially – the cone depends upon the representative of the equivalence class defining an object in a derived category, and not just the object itself.

Physically, our discussion of brane/antibrane systems was not the most general possible. One can give vacuum expectation values to more general vertex operators, not just the tachyons.

Curiously, these two issues solve each other. By incorporating a more general class of boundary vertex operators, one realizes a more general mathematical structure, due to Bondal and Kapranov, which repairs many of the technical deficiencies of ordinary derived categories. Ordinary complexes are replaced by generalized complexes in which arrows can map between non-neighboring elements of the complex. Schematically, the BRST operator is deformed by boundary vacuum expectation values to the form

$$Q = \overline{\partial} + \sum_a \phi_a$$

and demanding that the BRST operator square to zero implies that

$$\sum_a \overline{\partial}\phi_a + \sum_{a,b} \phi_b \circ \phi_a = 0$$

which is the same as the condition for a generalized complex. Note that for ordinary complexes, the condition above factors into

$$\overline{\partial}\phi_n = 0$$
$$\phi_{n+1} \circ \phi_n = 0$$

which yields an ordinary complex of sheaves (Figure 1).

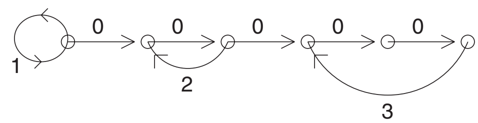



**Figure 1** 1. Example of generalized complex. Each arrow is labeled by the degree of the corresponding vertex operator.

## Landau–Ginzburg Models

So far we have described how derived categories are relevant to geometric compactifications, that is, sigma models on Calabi–Yau manifolds. However, there are also "nongeometric" theories – CFTs that do not come from sigma models on manifolds, of which Landau–Ginzburg models and their orbifolds are prominent examples. Landau–Ginzburg models can also be twisted into topological field theories, and the B-type topological twist of (an orbifold of) a Landau–Ginzburg model is believed to be isomorphic, as a topological field theory, to the B model obtained from a nonlinear sigma model, of the form we outlined earlier. Landau–Ginzburg models have a very different form than nonlinear sigma models, and so sometimes there can be practical computational advantages to working with one rather than the other.

A Landau–Ginzburg model is an ungauged sigma model with a nonzero superpotential (a holomorphic function over the target space that defines a bosonic potential and Yukawa couplings). (In "typical" cases, the target space is a vector space.) Because of the superpotential, a Landau–Ginzburg model is a massive theory – not itself a CFT, but many Landau–Ginzburg models are believed to flow to CFTs under the renormalization group.

In formulating open strings based on Landau–Ginzburg models, naive attempts fail because of something known as the Warner problem: if the superpotential is nonzero, then the obvious ways to try to define the theory on a Riemann surface with boundary have the undesirable property that the supersymmetry transformations only close up to a nonzero boundary term, proportional to derivatives of the superpotential. In order to find a description of open strings in which the supersymmetry transformations close, one must take a very nonobvious formulation of the boundary data. Specifically, to solve the Warner problem, one is led to work with pairs of matrices whose product is proportional to the superpotential.

This method of solving the Warner problem is known as matrix factorization, and D-branes in this theory are defined by the factorization chosen, that is, the choice of pairs of matrices. In simple cases, we can be more explicit as follows. Choose a set of

polynomials $F_\alpha, G_\alpha$ such that the Landau–Ginzburg superpotential $W$ is given by

$$W = \sum_\alpha F_\alpha G_\alpha + \text{constant.}$$

The $F_\alpha$ and $G_\alpha$ are used to define the boundary action – the $F$'s appear as part of the boundary superpotential and the $G$'s appear as part of the supersymmetry transformations of boundary fermi multiplets. The $F_\alpha$ and $G_\alpha$, that is, the factorization of $W$, determine the D-brane in the Landau–Ginzburg theory. We can also think of having a pair of holomorphic vector bundles $\mathcal{E}_1, \mathcal{E}_2$ of the same rank, and interpret $F$ and $G$ as holomorphic sections of $\mathcal{E}_1^\vee \otimes \mathcal{E}_2$ and $\mathcal{E}_2^\vee \otimes \mathcal{E}_1$, respectively, obeying $FG \propto W \cdot \text{Id}$ and $GF \propto W \cdot \text{Id}$, up to additive constants.

Although a Landau–Ginzburg model is not the same thing as a sigma model on a Calabi–Yau, orbifolds of Landau–Ginzburg models are often on the same Kähler moduli space. Perhaps, the most famous example of this relates sigma models on quintic hypersurfaces in $\boldsymbol{P}^4$ to a $\boldsymbol{Z}_5$ orbifold of a Landau–Ginzburg model over $\boldsymbol{C}^5$ with five chiral superfields $x_1, x_2, x_3, x_4, x_5$, and a superpotential of the form

$$W = x_1^5 + x_2^5 + x_3^5 + x_4^5 + x_5^5 + \psi x_1 x_2 x_3 x_4 x_5$$

for $\psi$ a complex number, corresponding to the equation of the degree-5 hypersurface in $\boldsymbol{P}^4$. The (complexified) Kähler moduli space in this example is a $\boldsymbol{P}^1$, with the sigma model on the quintic at one pole, the zero-volume limit of the sigma model along the equator, and the Landau–Ginzburg orbifold at the opposite pole.

Since the closed-string topological B model is independent of Kähler moduli, and the sigma model on the quintic and the Landau–Ginzburg orbifold above lie on the same Kähler moduli space, one would expect them both to have the same spectrum of D-branes, and indeed this is believed to be true.

## Pi-Stability

So far we have discussed D-branes in the topological B model, a topological twist of a physical sigma model. If we untwist back to a physical sigma model, then the stability of those D-branes becomes an issue.

To begin to understand what we mean by stability in this context, consider a set of $N$ D-branes wrapped on, say, a K3 surface, at large radius (so that world-sheet instanton corrections are small). On the world volume of the D-branes, we have a rank-$N$ vector bundle, and in the physical theory on that world volume we have a consistency condition for supersymmetric vacua, that the vector bundle be "Mumford–Takemoto stable." To understand what is meant by this condition on a Kähler manifold, let $\omega$ denote the Kähler form, and define the "slope" $\mu$ of a vector bundle $\mathcal{E}$ on a manifold $X$ of complex dimension $n$ to be given by

$$\mu(\mathcal{E}) = \frac{\int_X \omega^{n-1} \wedge c_1(\mathcal{E})}{\text{rank } \mathcal{E}}$$

where $\omega$ is the Kähler form. Then, we say that $\mathcal{E}$ is (semi-)stable if for all subsheaves $\mathcal{F}$ satisfying certain consistency conditions, $\mu(\mathcal{F})(\leq) < \mu(\mathcal{E})$.

Since the slope of a bundle depends upon the Kähler form, whether a given bundle is Mumford–Takemoto stable depends upon the metric. In general, on a Kähler manifold, the Kähler cone breaks up into subcones, with a different moduli space of (stable) holomorphic vector bundles in each subcone.

This is a mathematical notion of stability, but it also corresponds to physical stability, at least in a regime in which quantum corrections are small. If a given bundle is only stable in a proper subset of the Kähler cone, then when it reaches the boundary of the subcone in which it is stable, the gauge field configuration that satisfies the Donaldson–Uhlenbeck–Yau partial differential equation splits into a sum of two separate bundles. In a heterotic string compactification, this leads to a low-energy enhanced U(1) gauge symmetry and D-terms which realize the change in moduli space. In D-branes, this means the formerly bound state of D-branes (described by an irreducible holomorphic vector bundle) becomes only marginally bound; a decay becomes possible.

Pi-stability is a proposal for generalizing the considerations above to D-branes no longer wrapping the entire Calabi–Yau, and including quantum corrections.

In order to define pi-stability, we must first introduce a notion of grading $\varphi$ of a D-brane. Specifically, for a D-brane wrapped on the entire Calabi–Yau $X$ with holomorphic vector bundle $\mathcal{E}$, the grading is defined as the mirror to the expression $\int_X \text{ch}(\mathcal{E}) \wedge \Pi$, where $\Pi$ encodes the periods. Close to the large-radius limit, this has the form:

$$\varphi(\mathcal{E}) = \frac{1}{\pi} \text{Im} \log \int_X \exp(B + \mathrm{i}\omega) \wedge \text{ch}(\mathcal{E}) \times \wedge \sqrt{\text{td}(TX)} + \cdots$$

where $B$ is a 2-form, the "B field." As defined $\varphi$ is clearly $S^1$-valued; however, we must choose a

particular sheet of the log Riemann surface, to obtain an $\boldsymbol{R}$-valued function.

This notion of grading of D-branes is an ansatz, introduced as part of the definition of pi-stability. Physically, it is believed that the difference in grading between two D-branes corresponds to the fractional charge of the boundary-condition-changing vacuum between the two D-branes, though we know of no convincing first-principles derivation of that statement. In particular, unlike closed-string computations, the degree of the Ext group element corresponding to a particular boundary R-sector state is not always the same as the $U(1)_R$ charge – for example, it is often determined by the $U(1)_R$ charge minus the charge of the vacuum. The grading gives us the mathematical significance of that vacuum charge. This mismatch between Ext degrees and $U(1)_R$ charges is necessary for the grading to make sense: Ext group degrees are integral, after all, yet we want the grading to be able to vary continuously, so the grading had better not be the same as an Ext group degree.

Given an $\boldsymbol{R}$-valued function from a particular definition of log in the definition of $\varphi$ above, the statement of pi-stability is then that for all subsheaves $\mathcal{F}$, as in the statement of Mumford–Takemoto stability,

$$\varphi(\mathcal{F}) \leq \varphi(\mathcal{E})$$

Before trying to understand the physical meaning of $\varphi$, or the extension of these ideas to derived categories, let us try to confirm that Mumford–Takemoto stability emerges as a limit of pi-stability.

For simplicity, suppose that $X$ is a Calabi–Yau 3-fold. Then, for large Kähler form $\omega$, we can expand $\varphi(\mathcal{E})$ as,

$$\varphi(\mathcal{E}) \approx \frac{1}{\pi} \operatorname{Im} \log\left[-\frac{\mathrm{i}}{3!}\int_X \omega^3 (\operatorname{rk}\, \mathcal{E})\right] + \frac{3}{\pi}\frac{\int_X \omega^2 \wedge c_1(\mathcal{E})}{\int_X \omega^3(\operatorname{rk}\, \mathcal{E})} + \cdots$$

Thus, we see that to leading order in the Kähler form $\omega$, $\varphi(\mathcal{F}) \leq \varphi(\mathcal{E})$ if and only if

$$\frac{\int_X \omega^2 \wedge c_1(\mathcal{F})}{\operatorname{rk}\, \mathcal{F}} \leq \frac{\int_X \omega^2 \wedge c_1(\mathcal{E})}{\operatorname{rk}\, \mathcal{E}}$$

which is precisely the statement of Mumford–Takemoto stability on a 3-fold $X$.

One can define a notion of (classical) stability for more general sheaves, but what one wants is to apply pi-stability to derived categories, not just sheaves.

However, there is a technical problem that limits such an extension. Specifically, in a derived category, there is no meaningful notion of "subobject." Thus, a notion of stability formulated in terms of subobjects cannot be immediately applied to derived categories. There are two (equivalent) workarounds to this issue that have been discussed in the math and physics literatures, which can be briefly summarized as follows:

1. One workaround involves picking a subcategory of the derived category that does allow you to make sense of subobjects. Such a structure is known, loosely, as a "T-structure," and so one can imagine formulating stability by first picking a T-structure, then specifying a slope function on the elements of the subcategory picked out by the subcategory.
2. Another (equivalent) workaround is to work with a notion of "relative stability." Instead of speaking about whether a D-brane is stable against decay into any other object, one only speaks about whether it is stable against decay into pairs of specified objects.

In this fashion, one can make sense of pi-stability for derived categories.

*See also:* Fourier–Mukai Transform in String Theory; Mirror Symmetry: A Geometric Survey; Spectral Sequences; Superstring Theories; Topological Quantum Field Theory: Overview.

## Further Reading

Bondal A and Kapranov M (1991) Enhanced triangulated categories. *Mathematics of the USSR Sbornik* 70: 92–107.

Bridgeland T (2002) Stability conditions on triangulated categories, math.AG/0212237.

Caldararu A (2005) Derived categories of sheaves: a skimming, math.AG/0501094.

Hartshorne R (1966) *Residues and Duality*. Lecture Notes in Mathematics, vol. 20. Berlin: Springer.

Kapustin A and Li Y (2003) D-branes in Landau–Ginzburg models and algebraic geometry. *Journal of High Energy Physics*, 0312: 005 (hep-th/0210296).

Kontsevich M (1995) Homological algebra of mirror symmetry. In: Chatterji SD (ed.) *Proceedings of International Congress of Mathematicians* (Zurich, 1994), pp. 120–139, (alg-geom/9411018). Boston: Birkhäuser.

Sharpe E (2003) Lectures on D-branes and sheaves, hep-th/0307245.

Thomas R (2000) Derived categories for the working mathematician, math.AG/0001045.

Weibel C (1994) *An Introduction to Homological Algebra*. Cambridge Studies in Advanced Mathematics, vol. 38. Cambridge: Cambridge University Press.

# Determinantal Random Fields

**A Soshnikov**, University of California at Davis, Davis, CA, USA



## Introduction

The theory of random point fields has its origins in such diverse areas of science as life tables, particle physics, population processes, and communication engineering. A standard reference to the subject is the monograph by Daley and Vere-Jones (1988).

This article is concerned with a special class of random point fields, introduced by Macchi in the mid-1970s. The model that Macchi considered describes the statistical distribution of a fermion system in thermal equilibrium. Macchi proposed to call the new class of random point processes the fermion random point processes. The characteristic property of this family of random point processes is the condition that $k$-point correlation functions have the form of determinants built from a correlation kernel. This implies that the particles obey the Pauli exclusion principle. Until the mid-1990s, fermion random point processes attracted only a limited interest in mathematics and physics communities, with the exception of two important works by Spohn (1987) and Costin–Lebowitz (1995). This situation changed dramatically at the end of the last century, as the subject greatly benefited from the newly discovered connections to random matrix theory, representation theory, random growth models, combinatorics, and number theory. Things are rapidly developing at the moment. Even the terminology has not yet set in stone. Many experts currently use the term "determinantal random point fields" instead of "fermion random point fields." We follow this trend in our article.

This article is intended as a short introduction to the subject. The next section builds a mathematical framework and gives a formal mathematical definition of the determinantal random point fields. Then we discuss examples of determinantal random point fields from quantum mechanics, random matrix theory, random growth models, combinatorics, and representation theory. This is followed by a discussion of the ergodic properties of translation-invariant determinantal random point fields. We discuss the Gibbsian property of determinantal random point fields. Central-limit theorem type results for the counting functions and similar linear statistics are also discussed. The final section is devoted to some generalizations of determinantal point fields, namely immanantal and Pfaffian random point fields.

## Mathematical Framework

We start by building a standard mathematical framework for the theory of random point processes. Let $E$ be a one-particle space and $X$ a space of finite or countable configurations of particles in $E$. In general, $E$ can be a separable Hausdorff space. However, for our purposes it suffices to consider $E=\mathbb{R}^d$ or $E=\mathbb{Z}^d$. We usually assume in this section that $E=\mathbb{R}^d$, with the understanding that all constructions can be easily extended to the discrete case. We assume that each configuration $\xi=(x_i), x_i\in E$, $i\in\mathbb{Z}^1$ (or $i\in\mathbb{Z}^1_+$ for $d>1$), is locally finite. In other words, for every compact $K\subset E$, the number of particles in $K$, $\#_K(\xi)=\#(x_i\in K)$ is finite.

In order to introduce a $\sigma$-algebra of measurable subsets of $X$, we first define the cylinder sets. Let $B\subset E$ be a bounded Borel set and let $n\geq 0$. We call $C_n^B=\{\xi\in X:\#_B(\xi)=n\}$ a cylinder set. We define $\mathcal{B}$ as a $\sigma$-algebra generated by all cylinder sets (i.e., $\mathcal{B}$ is the minimal $\sigma$-algebra that contains all $C_n^B$).

**Definition 1** A random point field is a triplet $(X,\mathcal{B},\Pr)$, where Pr is a probability measure on $(X,\mathcal{B})$.

It was observed in the 1960–1970s (see, e.g., Lenard (1973, 1975)), that in many cases the most convenient way to define a probability measure on $(X,\mathcal{B})$ is via the point correlation functions. Let $E=\mathbb{R}^d$, equipped with the underlying Lebesgue measure.

**Definition 2** Locally integrable function $\rho_k:E^k\to\mathbb{R}^1_+$ is called a $k$-point correlation function of the random point field $(X,\mathcal{B},\Pr)$ if, for any disjoint bounded Borel subsets $A_1,\ldots,A_m$ of $E$ and for any $k_i\in\mathbb{Z}^1_+, i=1,\ldots,m$, $\sum_{i=1}^m k_i=k$, the following formula holds:

$$\mathbb{E}\prod_{i=1}^m\frac{(\#_{A_i})!}{(\#_{A_i}-k_i)!}=\int_{A_1^{k_1}\times\cdots\times A_m^{k_m}}\rho_k(x_1,\ldots,x_k)\mathrm{d}x_1\cdots\mathrm{d}x_k \qquad [1]$$

where by $\mathbb{E}$ we denote the mathematical expectation with respect to Pr. In particular, $\rho_1(x)$ is the particle density, since

$$\mathbb{E}\#_A=\int_A\rho_1(x)\mathrm{d}x$$

for any bounded Borel $A\subset E$. In general, $\rho_k(x_1,\ldots,x_k)$ has the following probabilistic interpretation. Let $[x_1,x_i+\mathrm{d}x_i], i=1,\ldots,k$, be infinitesimally small boxes around $x_i$, then $\rho_k(x_1,x_2,\ldots,x_k)$ $\mathrm{d}x_1\cdots\mathrm{d}x_k$ is the probability to find a particle in each of these boxes.

In the discrete case $E = \mathbb{Z}^d$, the construction of a random point field is very similar. The probability space $X$ and the $\sigma$-algebra $\mathcal{B}$ are constructed essentially in the same way as before. Moreover, in the discrete case, the set of the countable configurations of particles can be identified with the set of all subsets of $E$. Therefore, $X = \{0, 1\}^E$, and $\mathcal{B}$ is generated by the events $\{C_x, x \in E\}$, where $C_x = \{x \in \xi\}$. The $k$-point correlation function $\rho(x_1, \dots, x_k)$ is then just a probability that a configuration $\xi$ contains the sites $x_1, \dots, x_k$. In other words, $\rho_k(x_1, \dots, x_k) = \Pr(\bigcap_{i=1}^k C_{x_i})$. In particular, the one-point correlation function $\rho_1(x), x \in \mathbb{Z}^d$, is the probability that a configuration contains the site $x$, that is, $\rho_1(x) = \Pr(C_x)$.

The problem of the existence and the uniqueness of a random point field defined by its correlation functions was studied by Lenard (1973–1975). It is not surprising that Lenard's papers revealed many parallels to the classical moment problem. In particular, the random point field is uniquely defined by its correlation functions if the distribution of random variables $\{\#_A\}$ for bounded Borel sets $A$ is uniquely determined by its moments.

In this article we study a special class of random point fields introduced by Macchi (1975). To shorten the exposition, we give the definitions only in the continuous case $E = \mathbb{R}^d$. In the discrete case, the definitions are essentially the same.

Let $K : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ be an integral locally trace-class operator. The last condition means that for any compact $B \subset \mathbb{R}^d$ the operator $K\chi_B$ is trace class, where $\chi_B(x)$ is an indicator of $B$. The kernel of $K$ is defined up to a set of measure zero in $\mathbb{R}^d \times \mathbb{R}^d$. For our purposes, it is convenient to choose it in such a way that for any bounded measurable $B$ and any positive integer $n$

$$\mathrm{tr}((\chi_B K \chi_B)) = \int_B K(x, x)\mathrm{d}x \qquad [2]$$

We refer the reader to Soshnikov (2000, p. 927) for the discussion. We are now ready to define a determinantal (fermion) random point field on $\mathbb{R}^d$.

**Definition 3** A random point field on $E$ is said to be determinantal (or fermion) if its $n$-point correlation functions are of the form

$$\rho_n(x_1, \dots, x_n) = \det\big(K(x_i, x_j)\big)_{1 \le i \le n} \qquad [3]$$

**Remark 1** If the kernel is Hermitian-symmetric, then the non-negativity of $n$-point correlation functions implies that the kernel $K(x, y)$ is non-negative definite and, therefore $K$ must be a non-negative operator. It should be noted, however, that there exist determinantal random point fields corresponding to non-Hermitian kernels (see, e.g., [18] later). The kernel $K(x, y)$ is usually called a correlation kernel of the determinantal random point process.

In the Hermitian case, the necessary and sufficient conditions on the operator $K$ to define a determinantal random point filed were established by Soshnikov (2000); see also Macchi (1975).

**Theorem 1** *Hermitian locally trace class operator $K$ on $L^2(E)$ determines a determinantal random point field if and only if $0 \le K \le 1$ (in other words, both $K$ and $1 - K$ are non-negative operators). If the corresponding random point field exists, it is unique.*

The main technical part of the proof is the following proposition.

**Proposition 1** *Let $(X, B, P)$ be a determinantal random point field with the Hermitian-symmetric correlation kernel $K$. Let $f$ be a non-negative continuous function with compact support. Then*

$$\mathbb{E}e^{\langle \xi, f\rangle} = \det\Big(\mathrm{Id} - (1 - \mathrm{e}^{-f})^{1/2} K (1 - \mathrm{e}^{-f})^{1/2}\Big) \qquad [4]$$

*where $\langle \xi, f\rangle$ is the value of the linear statistics defined by the test function $f$ on the configuration $\xi = (x_i)$; in other words, $\langle \xi, f\rangle = \Sigma_i f(x_i)$.*

**Remark 2** The right-hand side (RHS) of [4] is well defined as the Fredholm determinant of a trace-class operator. Letting $f = \sum_{i=1}^k s_i \chi I_i$, one obtains $\mathbb{E}e^{\langle \xi, f\rangle} = \mathbb{E}\prod_{i=1}^k z_i^{\#_{I_i}}$, with $z_i = \mathrm{e}^{s_i}$. In this case, the left-hand side (LHS) of [4] becomes the generating function of the joint distribution of the counting random variables $\#_{I_i}, i = 1, \dots, k$.

Unfortunately, there are very few known results in the non-Hermitian case. In particular, the necessary and sufficient condition on $K$ for the existence of the determinantal random point field with the non-Hermitian correlation kernel is not known.

We end this section with the introduction of the Janossy densities (a.k.a. density distributions, exclusion probability densities, etc.) of a random point field.

The term Janossy densities in the theory of random point processes was introduced by Srinivasan in 1969, who referred to the 1950 paper by Janossy on particle showers. Let us assume that all point correlation functions exist and are locally integrable, and let $I$ be a bounded Borel subset of

$\mathbb{R}^d$. Intuitively, one can think of the Janossy density $\mathcal{J}_{k,I}(x_1,\ldots,x_k)$, $x_1,\ldots,x_k \in I$, as

$$\mathcal{J}_{k,I}(x_1,\ldots,x_k)\prod_{i=1}^{k}\mathrm{d}x_i = \Pr\{\text{there are exactly } k \text{ particles in } I \text{ and there is a particle in each of the } k \text{ infinitesimal boxes } (x_i, x_i+\mathrm{d}x_i),\ i=1,\ldots,k\} \quad [5]$$

To give a formal definition, we express point correlation functions in terms of Janossy densities and vice versa:

$$\rho_k(x_1,\ldots,x_k) = \sum_{j=1}^{\infty}\frac{1}{j!}\int_{I^j}\mathcal{J}_{k+j,I}(x_1,\ldots,x_k,x_{k+1},\ldots,x_{k+j}) \times \mathrm{d}x_{k+1}\ldots\mathrm{d}x_{k+j} \quad [6]$$

$$\mathcal{J}_{k,I}(x_1,\ldots,x_k) = \sum_{j=0}^{\infty}\frac{(-1)^j}{j!}\int_{I^j}\rho_{k+j}(x_1,\ldots,x_k,x_{k+1},\ldots,x_{k+j}) \times \mathrm{d}x_{k+1}\cdots\mathrm{d}x_{k+j} \quad [7]$$

A very useful property of the Janossy densities is that

$$\Pr\{\text{there are exactly } k \text{ particles in } I\} = \frac{1}{k!}\int_{I^k}\mathcal{J}_{k,I}(x_1,\ldots,x_k)\mathrm{d}x_1\cdots\mathrm{d}x_k \quad [8]$$

In the case of determinantal random point fields, Janossy densities also have a determinantal form, namely

$$\mathcal{J}_{k,I}(x_1,\ldots,x_k) = \det(\mathrm{Id}-K_I)\cdot\det\big(L_I(x_i,x_j)\big)_{1\le i,j\le k} \quad [9]$$

In the last equation, $K_I$ is the restriction of the operator $K$ to the $L^2(I)$. In other words, $K_I(x,y) = \chi_I(x)K(x,y)\chi_I(y)$, where $\chi_I$ is the indicator of $I$. The operator $L_I$ is expressed in terms of $K_I$ as $L_I = (\mathrm{Id}-K_I)^{-1}K_I$. For further results on the Janossy densities of determinantal random point processes we refer the reader to Soshnikov (2004) and references therein.

## Examples of Determinantal Random Point Fields

### Fermion Gas

Let $H = -\mathrm{d}^2/\mathrm{d}x^2 + V(x)$ be a Schrödinger operator with discrete spectrum on $L^2(E)$. We denote by $\{\varphi_\ell\}_{\ell=0}^{\infty}$ an orthonormal basis of the eigenfunctions, $H\varphi_\ell = \lambda_\ell\cdot\varphi_\ell, \lambda_0 < \lambda_1 \le \lambda_2 \le \cdots$. To define a Fermi gas, we consider the $n$th exterior power of $H$, $\wedge^n(H)\colon \wedge^n(L^2(E)) \to \wedge^n(L^2(E))$, where $\wedge^n(L^2(E))$ is the space of square-integrable antisymmetric functions of $n$ variables and $\wedge^n(H) = \sum_{i=1}^{n}(-\mathrm{d}^2/\mathrm{d}x_i^2 + V(x_i))$. The eigenstates of the Fermi gas are given by the normalized Slater determinants

$$\psi_{k_1,\ldots,k_n}(x_1,\ldots,x_n) = \frac{1}{\sqrt{n!}}\sum_{\sigma\in S_n}(-1)^\sigma\prod_{i=1}^{n}\varphi_{k_i}(x_{\sigma(i)}) = \frac{1}{\sqrt{n!}}\det(\varphi_{k_i}(x_j))_{1\le i,j\le n} \quad [10]$$

where $0 \le k_1 < k_2 < \cdots < k_n$. A probability distribution of $n$ particles in the Fermi gas is given by the squared absolute value of the eigenstate:

$$p(x_1,\ldots,x_n) = |\psi(x_1,\ldots,x_n)|^2 = \frac{1}{n!}\det\big(\varphi_{k_i}(x_j)\big)_{1\le i,j\le n} \times \det\Big(\overline{\varphi_{k_j}(x_i)}\Big)_{1\le i,j\le n} = \frac{1}{n!}\det\big(K_n(x_i,x_j)\big)_{1\le i,j\le n} \quad [11]$$

where $K_n(x,y) = \sum_{i=1}^{n}\varphi_{k_i}(x)\overline{\varphi_{k_i}(y)}$ is the kernel of the orthogonal projector onto the subspace spanned by the $n$ eigenfunctions $\{\varphi_{k_i}\}$ of $H$. The $n$-dimensional probability distribution [11] defines a determinantal random point field with $n$ particles. The $k$-point correlation functions are given by

$$\rho_k^{(n)}(x_1,\ldots,x_n) = \frac{n!}{(n-k)!}\int p_n(x_1,\ldots,x_n) \times \mathrm{d}x_{k+1}\cdots\mathrm{d}x_n = \det\big(K_n(x_1,x_j)\big)_{1\le i,j\le k} \quad [12]$$

### Random Matrix Models

Some of the most important ensembles of random matrices fall into the class of determinantal random point processes.

The archetypal ensemble of Hermitian random matrices is a so-called Gaussian unitary ensemble (GUE). Let us consider the space of $n\times n$ Hermitian matrices $\{A = (A_{ij})_{1\le i,j\le n}, \mathrm{Re}(A_{ij}) = \mathrm{Re}(A_{ji}), \mathrm{Im}(A_{ij}) = -\mathrm{Im}(A_{ji})\}$. A GUE random matrix is defined by its probability distribution

$$P(\mathrm{d}A) = \mathrm{const}_n\cdot\exp(-\mathrm{tr}A^2)\mathrm{d}A \quad [13]$$

where $\mathrm{d}A$ is a Lebesgue measure, that is, $\mathrm{d}A = \prod_{i<j}\mathrm{dRe}(A_{ij})\,\mathrm{dIm}(A_{ij})\prod_{k=1}^{n}\mathrm{d}A_{kk}$. The eigenvalues of a random Hermitian matrix are real random

variables, whose joint probability distribution is a determinantal random point process of $n$ particles on the real line. The correlation kernel has the Christoffel–Darboux form built from the Hermite polynomials.

The GUE ensemble of random matrices is invariant under the unitary transformation $A \to UAU^{-1}$, $U \in U(n)$. An important generalization of [13] that preserves the unitary invariance is

$$P(\mathrm{d}A) = \mathrm{const}_n \exp(-\mathrm{tr}\,V(A))\mathrm{d}A \qquad [14]$$

where, for example, $V(x)$ is a polynomial of even degree with positive leading coefficients. The correlation functions of the eigenvalues in [14] are again determinantal, and the Hermite polynomials in the correlation kernel have to be replaced by the orthonormal polynomials with respect to the weight $\exp(-V(x))$. For details, we refer the reader to the monographs by Mehta (2004) and Deift (2000).

There are many other ensembles of random matrices for which the joint distribution of the eigenvalues has determinantal point correlation functions: classical compact groups with respect to the Haar measure, complex non-Hermitian Gaussian random matrices, positive Hermitian random matrices of the Wishart type, and chains of correlated Hermitian matrices. We refer the reader to Soshnikov (2000) for more information.

## Discrete Translation-Invariant Determinantal Random Point Fields

Let $g : \mathbb{T}^d \to [0,1]$ be a Lebesgue-measurable function on the $d$-dimensional torus $\mathbb{T}^d$. Assume that $0 \le g \le 1$. A configuration $\xi$ in $\mathbb{Z}^d$ can be thought of as a 0–1 function on $\mathbb{Z}^d$, that is, $\xi(x) = 1$ if $x \in \xi$ and $\xi(x) = 0$ otherwise. We define a $\mathbb{Z}^d$-invariant probability measure Pr on the Borel sets of $X = \{0,1\}^{\mathbb{Z}^d}$ in such a way that

$$\rho_k(x_1, \ldots, x_k) = \Pr(\xi(x_1) = 1, \ldots, \xi(x_k) = 1) := \det\big(\hat{g}(x_i - x_j)\big)_{1 \le i,j \le k} \qquad [15]$$

for $x_1, \ldots, x_k \in \mathbb{Z}^d$. In the above formula, $\{g(n)\}$ are the Fourier coefficients of $g$, that is, $g(x) = \sum_n \hat{g}(n) \mathrm{e}^{in\cdot x}$. It is clear from Definition 3 that [15] defines a determinantal random point field on $\mathbb{Z}^d$ with the translation-invariant kernel $K(x,y) = \hat{g}(x - y)$. Below we discuss several examples that fall into this category. For further discussion we refer the reader to Lyons (2003) and Soshnikov (2000).

1. In the trivial case when $g$ is identically a constant $p \in [0,1]$, we obtain the i.i.d. Bernoulli probability measure.
2. The edges of the uniform spanning tree in $\mathbb{Z}^2$ parallel to the horizontal axis can be viewed as the determinantal random point field in $Z^2$ with

$$g(x,y) = \frac{\sin^2 \pi x}{\sin^2 \pi x + \sin^2 \pi y}$$

   Similarly, the edges of the uniform spanning forest in $Z^d$ parallel to the $x_1$-axis correspond to the function

$$g(x_1, \ldots, x_d) = \frac{\sin^2 \pi x_1}{\sum_{i=1}^d \sin^2 \pi x_i}$$

   (the uniform spanning forest on $\mathbb{Z}^d$ is a tree only for $d \le 4$). The result is due to Burton and Pemantle (1993).
3. Let $d = 1$ and $\gamma$ be a parameter between 0 and 1. Consider

$$g(x) = \frac{(1-\gamma)^2}{|\mathrm{e}^{2\pi i x} - \gamma|^2}$$

   The corresponding probability measure is a renewal process and

$$K(n) = \hat{g}(n) = \frac{1-\gamma}{1+\gamma}\gamma^{|n|}$$

   (see Soshnikov (2000)).
4. The process with $g(x) = \chi_I(x)$, where $I$ is an arbitrary arc of a unit circle, appeared in the work of Borodin and Olshanski (2000). The corresponding correlation kernel is known as the discrete sine kernel. The determinantal random point process on $\mathbb{Z}^1$ with the discrete sine kernel describes the typical form of large Young diagrams "in the bulk" (see the next subsection).
5. The discrete sine correlation kernel with $g = \chi_{[0,1/2]}$ appeared in the zig-zag process (Johansson 2002) derived from the uniform domino tilings in the plane. It corresponds to $g = \chi_{[0,1/2]}$.

## Determinantal Measures on Partitions

By a partition of $n = 1, 2, \ldots$ we understand a collection of non-negative integers $\lambda = (\lambda_1, \ldots, \lambda_m)$ such that $\lambda_1 + \cdots + \lambda_m = n$ and $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_m$. We shall use a notation Par$(n)$ for the set of all partitions of $n$.

The Plancherel measure $M_n$ on the set Par$(n)$ is defined as

$$M_n(\lambda) = \frac{(\dim \lambda)^2}{n!} \qquad [16]$$

where $\dim \lambda$ is the dimension of the corresponding irreducible representation of the symmetric group

$S_n$. Let $\text{Par} = \bigsqcup_{n=0}^{\infty} \text{Par}(n)$. Consider a probability measure $M^\theta$ on Par

$$M^\theta(\lambda) = e^{-\theta}\frac{\theta^n}{n!}M_n(\lambda) \quad \text{where}$$
$$\lambda \in \text{Par}(n),\ n = 0,1,2,\ldots,\quad 0 \le \theta < \infty \qquad [17]$$

$M^\theta$ is called the Poissonization of the measures $M_n$. The analysis of the asymptotic properties of $M_n$ and $M^\theta$ has been important in connection to the famous Ulam problem and related questions in representation theory.

It was shown by Borodin and Okounkov (2000), and, independently, Johansson (2001) that $M^\theta$ is a determinantal random point field. The corresponding correlation kernel $K$ (in the so-called modified Frobenius coordinates) is a so-called discrete Bessel kernel on $\mathbb{Z}^1$,

$$K(x,y) = \begin{cases} \sqrt{\theta}\,\dfrac{J_{|x|-1/2}(2\sqrt{\theta})J_{|y|+1/2}(2\sqrt{\theta}) - J_{|x|+1/2}(2\sqrt{\theta})J_{|y|-1/2}(2\sqrt{\theta})}{|x|-|y|} \\ \text{if } xy > 0 \\ \sqrt{\theta}\,\dfrac{J_{|x|-1/2}(2\sqrt{\theta})J_{|y|-1/2}(2\sqrt{\theta}) - J_{|x|+1/2}(2\sqrt{\theta})J_{|y|+1/2}(2\sqrt{\theta})}{x-y} \\ \text{if } xy < 0 \end{cases} \qquad [18]$$

where $J_x(\cdot)$ is the Bessel function of order $x$. One can observe that the kernel $K(x,y)$ is not Hermitian, but the restriction of this kernel to the positive and negative semiaxis is Hermitian.

$M^\theta$ is a special case of an infinite parameter family of probability measures on Par, called the Schur measures, and defined as

$$M(\lambda) = \frac{1}{Z}s_\lambda(x)s_\lambda(y) \qquad [19]$$

where $s_\lambda$ are the Schur functions, $x = (x_1, x_2, \ldots)$ and $y = (y_1, y_2, \ldots)$ are parameters such that

$$Z = \sum_{\lambda \in \text{Par}} s_\lambda(x)s_\lambda(y) = \prod_{i,j}(1 - x_i y_j)^{-1} \qquad [20]$$

is finite and $\{x_i\}_{i=1}^\infty = \overline{\{y_i\}_{i=1}^\infty}$. It was shown by Okounkov (2001), that the Schur measures belong to the class of the determinantal random point fields.

### NonIntersecting Paths of a Markov Process

Let $p_{t,s}(x,y)$ be the transition probability of a Markov process $\xi(t)$ on $\mathbb{R}$ with continuous trajectories and let $(\xi_1(t), \xi_2(t), \ldots, \xi_n(t))$ be $n$ independent copies of the process. A classical result of Karlin and McGregor (1959) states that if $n$ particles start at the positions $x_1^{(0)} < x_2^{(0)} < \cdots < x_n^{(0)}$, then the probability density of their joint distribution at time $t_1 > 0$, given that their paths have not intersected for all $0 \le t \le t_1$, is equal to

$$\pi_{t_1}(x_1^{(1)}, \ldots, x_n^{(1)}) = \frac{1}{Z}\det(p_{0,t_1}(x_i^{(0)}, x_j^{(1)}))_{i,j=1}^n$$

provided the process $(\xi_1(t), \xi_2(t), \ldots, \xi_n(t))$ in $\mathbb{R}^n$ has a strong Markovian property.

Let $0 < t_1 < t_2 < \cdots < t_{M+1}$. The conditional probability density that the particles are in the positions $x_1^{(1)} < x_2^{(1)} < \cdots < x_n^{(1)}$ at time $t_1$, at the positions $x_1^{(2)} < x_2^{(2)} < \cdots < x_n^{(2)}$ at time $t_2, \ldots,$ at the positions $x_1^{(M)} < x_2^{(M)} < \cdots < x_n^{(M)}$ at time $t_M$, given that at time $t_{M+1}$ they are at the positions $x_1^{(M+1)} < x_2^{(M+1)} < \cdots < x_n^{(M+1)}$ and their paths have not intersected, is then equal to

$$\pi_{t_1,t_2,\ldots,t_M}(x_1^{(1)}, \ldots, x_n^{(M)}) = \frac{1}{Z_n, M}\prod_{l=0}^{M}\det(p_{t_l,t_{l+1}}(x_i^{(l)}, x_j^{(l+1)}))_{i,j=1}^n \qquad [21]$$

where $t_0 = 0$.

It is not difficult to show that [21] can be viewed as a determinantal random point process (see, e.g., Johansson (2003).

The formulas of a similar type also appeared in the papers by Johansson, Prähofer, Spohn, Ferrari, Forrester, Nagao, Katori, and Tanemura in the analysis of polynuclear growth models, random walks on a discrete circle, and related problems.

## Ergodic Properties

As before, let $(X, \mathcal{B}, \Pr)$ be a random point field with a one-particle space $E$. Hence, $X$ is a space of the locally finite configurations of particles in $E$, $\mathcal{B}$ a Borel $\sigma$-algebra of measurable subsets of $X$, and Pr a probability measure on $(X, B)$. Throughout this section, we assume $E = \mathbb{R}^d$ or $\mathbb{Z}^d$. We define an action $\{T^t\}_{t \in E}$ of the additive group $E$ on $X$ in the following natural way:

$$T^t : X \to X,\ (T^t\xi)_i = (\xi)_i + t \qquad [22]$$

We recall that a random point field $(X, \mathcal{B}, P)$ is called translation invariant if, for any $A \in \mathcal{B}$, any $t \in E$, $\Pr(T^{-t}A) = \Pr(A)$. The translation invariance of the correlation kernel $K(x,y) = K(x-y, 0) =: K(x-y)$ implies the translation invariance of $k$-point correlation functions

$$\rho_k(x_1 + t, \ldots, x_k + t) = \rho_k(x_1, \ldots, x_k)$$
$$\text{a.e. } k = 1, 2, \ldots,\quad t \in E \qquad [23]$$

which, in turn, implies the translation invariance of the random point field. The ergodic properties

of such point fields were studied by several mathematicians (Soshnikov 2000, Shirai and Takahashi, 2003, Lyons and Steif 2003). The first general result in this direction was obtained by Soshnikov (2000).

**Theorem 2** *Let* $(X, B, P)$ *be a determinantal random point field with a translation-invariant correlation kernel. Then the dynamical system* $(X, B, P, \{T^t\})$ *is ergodic, has the mixing property of any multiplicity and its spectra is absolutely continuous.*

We refer the reader to the article on ergodic theory for the definitions of ergodicity, mixing property, absolute continuous spectrum of the dynamical system, etc.

In the discrete case [15], $E = \mathbb{Z}^d$, more is known. Lyons and Steif (2003) proved that the shift dynamical system is Bernoulli, that is, it is isomorphic (in the ergodic theory sense) to an i.i.d. process. Under the additional conditions $\mathrm{Spec}(K) \subset (0,1)$ and $\sum_n |n||K(n)|^2 < \infty$, Shirai and Takahashi (2003a) proved the uniform mixing property.

## Gibbsian Properties

Costin and Lebowitz (1995) were the first to question the Gibbsian nature of the determinantal random point fields; they studied the continuous determinantal random point process on $\mathbb{R}^1$ with a so-called sine correlation kernel

$$K(x, y) = \frac{\sin(\pi(x-y))}{\pi(x-y)}$$

The first rigorous result (in the discrete case) was established by Shirai and Takahashi (2003b).

**Theorem 3** *Let E be a countable discrete space and K a symmetric bounded operator on* $l^2(E)$. *Assume that* $\mathrm{Spec}(K) \subset (0,1)$. *Then* $(X, B, P)$ *is a Gibbs measure with the potential U given by* $U(x|\xi) = -\log(J(x,x) - \langle J_\xi^{-1} j_\xi^x, j_\xi^x \rangle)$, *where* $x \in E, \xi \in X$, $\{x\} \cap \xi = \emptyset$. *Here* $J(x,y)$ *stands for the kernel of the operator* $J = (\mathrm{Id} - K)^{-1} K$, *and we set* $J_\xi = (J(y,z))_{y,z \in \xi}$ *and* $j_\xi^x = (J(x,y))_{y \in \xi}$.

We recall that the Gibbsian property of the probability measure $P$ on $(X, \mathcal{B})$ means that

$$E[F|\mathcal{B}_{\Lambda^c}](\xi) = \frac{1}{Z_{\Lambda,\xi}} \sum_{\eta \subset \Lambda} \mathrm{e}^{-U(\eta|\xi_{\Lambda^c})} F(\eta \cup \xi_{\Lambda^c})$$

where $\Lambda$ is a finite subset of $E$, $\mathcal{B}_{\Lambda^c}$ is the $\sigma$-algebra generated by the $\mathcal{B}$-measurable functions with the support outside of $\Lambda$, $E[F|\mathcal{B}_{\Lambda^c}]$ is the conditional mathematical expectation of the integrable function $F$ on $(X, \mathcal{B}, P)$ with respect to the $\sigma$-algebra $\mathcal{B}_{\Lambda^c}$. The potential $U$ is uniquely defined by the values of $U(x, \xi)$, as follows from the following recursive relation:

$$\begin{aligned} U(\{x_1, \ldots, x_n\}|\xi) &= U(x_n|\{x_1, \ldots, x_{n-1}\} \cup \xi) \\ &\quad + U(x_{n-1}|\{x_1, \ldots, x_{n-2}\} \cup \xi) \\ &\quad + \cdots + U(x_1|\xi) \end{aligned}$$

For additional information about the Gibbsian property, *see* Introductory Articles: Equilibrium Statistical Mechanics. Much less is known in the continuous case. Some generalized form of Gibbsianness, under quite restrictive conditions, was recently established by Georgii and Yoo (2004).

## Central Limit Theorem for Counting Function

In this section, we discuss the central-limit theorem type results for the linear statistics. The first important result in this direction was established by Costin and Lebowitz in 1995, who proved the central-limit theorem for the number of particles in the growing box, $\#_{[-L,L]}, L \to \infty$, in the case of the determinantal random point process on $\mathbb{R}^1$ with the sine correlation kernel

$$K(x, y) = \frac{\sin(\pi(x-y))}{\pi(x-y)}$$

Below we formulate the Costin–Lebowitz theorem in its general form due to Soshnikov (1999, 2000).

**Theorem 4** *Let E be a one-particle space,* $\{0 \le K_t \le 1\}$ *a family of locally trace-class operators in* $L^2(E), \{(X, \mathcal{B}, P_t)\}$ *a family of the corresponding determinantal random point fields in E, and* $\{I_t\}$ *a family of measurable subsets in E such that*

$$\begin{aligned} &\mathrm{Var}\#_{I_t} \\ &\quad = \mathrm{tr}(K_t \cdot \chi_{I_t} - (K_t \cdot \chi_{I_t})^2) \to \infty \quad \text{as } t \to \infty \end{aligned} \qquad [24]$$

*Then the distribution of the normalized number of particles in* $I_t$ *(with respect to* $P_t$*) converges to the normal law, that is,*

$$\frac{\#_{I_t} - \mathbb{E}\#_{I_t}}{\sqrt{\mathrm{Var}\#_{I_t}}} \xrightarrow{w} N(0,1)$$

An analogous result holds for the joint distribution of the counting functions $\{\#_{I_{t1}}, \ldots, \#_{I_{tk}}\}$, where $I_t^1, \ldots, I_t^k$ are disjoint measurable subsets in $E$.

The proof of the Costin–Lebowitz theorem uses the $k$-point cluster functions. In the determinantal case, the cluster functions have a simple form

$$r_k(x_1, \ldots, x_k) = (-1)^l \frac{1}{l} \sum_{\sigma \in S_k} K(x_{\sigma(1)}, x_{\sigma(2)}) K(x_{\sigma(2)}, x_{\sigma(3)}) \cdots \times K(x_{\sigma(k)}, x_{\sigma(1)}) \qquad [25]$$

The importance of the cluster function stems from the fact that the integrals of the $k$-point cluster function over the $k$-cube with a side $I$ can be expressed as a linear combination of the first $k$ cumulants of the counting random variable $\#_I$. In other words,

$$\int_{I \times \ldots \times I} r_k(x_1, \ldots, x_k)\, dx_1 \cdots dx_k = \sum_{l=1}^{k} \beta_{kl} C_l(\#_I) \qquad [26]$$

It follows from [25] that the integral at the LHS of [26] equals, up to a factor $(-1)^l (l-1)!$, to the trace of the $k$th power of the restriction of $K$ to $I$. This allows one to estimate the cumulants of the counting random variable $\#_I$. For details, we refer the reader to Soshnikov (2000). The central-limit theorem for a general class of linear statistics, under some technical assumptions on the correlation kernel was proved in Soshnikov (2002). Finally, we refer the reader to Soshnikov (2000) for the functional central-limit theorem for the empirical distribution function of the nearest spacings.

## Generalizations: Immanantal and Pfaffian Point Processes

In this section, we discuss two important generalizations of the determinantal point processes.

### Immanantal Processes

Immanantal random point processes were introduced by P Diaconis and S N Evans in 2000. Let $\lambda$ be a partition of $n$. Denote by $\chi^\lambda$ the character of the corresponding irreducible representation of the symmetric group $S_n$. Let $K(x, y)$, be a non-negative-definite, Hermitian kernel. An immanantal random point process is defined through the correlation functions

$$\rho_k(x_1, \ldots, x_k) = \sum_{\sigma \in S_n} \chi^\lambda(\sigma) \prod_{i=1}^{n} K(x_i, x_{\sigma(i)}) \qquad [27]$$

In other words, the correlation functions are given by the immanants of the matrix with the entries $K(x_i, x_j)$. We will denote the RHS of [27] by $K^\lambda[x_1, \ldots, x_n]$.

In the special case $\lambda = (1^n)$ (i.e., $\lambda$ consists of $n$ parts, all of which equal to 1), one obtains that $\chi^\lambda(\sigma) = (-1)^\sigma$, and $K^\lambda[x_1, \ldots, x_n] = \det(K(x_i, x_j))$. Therefore, in the case $\lambda = (1^n)$ the random point process with the correlation functions [27] is a determinantal random point process. When $\lambda = (n)$ (i.e., the permutation has only one part, namely $n$) we have $\chi^\lambda = 1$ identically, and $K^\lambda[x_1, \ldots, x_n] = \mathrm{per}(K(x_i, x_j))$, the permanent of the matrix $K(x_i, x_j)$. The corresponding random point process is known as the boson random point process.

### Pfaffian Processes

Let

$$K(x, y) = \begin{pmatrix} K_{11}(x, y) & K_{12}(x, y) \\ K_{21}(x, y) & K_{22}(x, y) \end{pmatrix}$$

be an antisymmetric $2 \times 2$ matrix-valued kernel, that is, $K_{ij}(x, y) = -K_{ji}(y, x)$, $i, j = 1, 2$. The kernel defines an integral operator acting on $L^2(E) \bigoplus L^2(E)$, which we assume to be locally trace class. A random point process on $E$ is called Pfaffian if its point correlation functions have a Pfaffian form

$$\rho_k(x_1, \ldots, x_k) = \mathrm{pf}(K(x_i, x_j))_{i,j=1,\ldots,k}, \quad k \geq 1 \qquad [28]$$

The RHS of [28] is the Pfaffian of the $2k \times 2k$ antisymmetric matrix (since each entry $K(x_i, x_j)$ is a $2 \times 2$ block). Determinantal random point processes is a special case of the Pfaffian processes, corresponding to the matrix kernel of the form

$$K(x, y) = \begin{pmatrix} 0 & \tilde{K}(x, y) \\ -\tilde{K}(y, x) & 0 \end{pmatrix}$$

where $\tilde{K}$ is a scalar kernel. The most well known examples of the Pfaffian random point processes, that cannot be reduced to determinantal form are $\beta = 1$ and $\beta = 4$ polynomial ensembles of random matrices and their limits (in the bulk and at the edge of the spectrum), as the size of a matrix goes to infinity.

## Acknowledgment

The research of A Soshnikov was supported in part by the NSF grant DMS-0405864.

*See also:* Dimer Problems; Ergodic Theory; Growth Processes in Random Matrix Theory; Integrable Systems in Random Matrix Theory; Percolation Theory; Quantum Ergodicity and Mixing of Eigenfunctions; Random Matrix Theory in Physics; Random Partitions; Statistical Mechanics and Combinatorial Problems; Symmetry Classes in Random Matrix Theory.

## Further Reading

Borodin A and Olshanski G (2000) Distribution on partitions, point processes, and the hypergeometric kernel. *Communications in Mathematical Physics* 211: 335–358.

Daley DJ and Vere-Jones D (1988) *An Introduction to the Theory of Point Processes*. New York: Springer.

Diaconis P and Evans SN (2000) Immanants and finite point processes. *Journal of Combinatorial Theory Series A* 91(1–2): 305–321.

Georgii H-O and Yoo HJ (2005) Conditional intensity and Gibbsianness of determinantal point processes. *Journal of Statistical Physics* 118(91/92): 55–84.

Johansson K (2003) Discrete polynuclear growth and determinantal processes. *Communications in Mathematical Physics* 242(1–2): 277–329.

Lyons R (2003) Determinantal probability measures. *Publications Mathématiques Institut de Hautes Études Scientifiques* 98: 167–212.

Lyons R and Steif J (2003) Stationary determinantal processes: phase multiplicity, Bernoillicity, entropy, and domination. *Duke Mathematical Journal* 120(3): 515–575.

Macchi O (1975) The coincidence approach to stochastic point processes. *Advances in Applied Probability* 7: 83–122.

Okounkov A (2001) Infinite wedge and random partitions. *Selecta Mathematica. New Series* 7(1): 57–81.

Shirai T and Takahashi Y (2003a) Random point fields associated with certain Fredholm determinants, I. Fermion, Poisson and boson point processes. *Journal of Functional Analysis* 205(2): 414–463.

Shirai T and Takahashi Y (2003b) Random point fields associated with certain Fredholm determinants, II. Fermion shifts and their ergodic and Gibbs properties. *The Annals of Probability* 31(3): 1533–1564.

Soshnikov A (2000) Determinantal random point fields. *Russian Mathematical Surveys* 55: 923–975.

Soshnikov A (2002) Gaussian limit for determinantal random point fields. *The Annals of Probability* 30(1): 171–187.

Soshnikov A (2004) Janossy densities of coupled random matrices. *Communications in Mathematical Physics* 251: 447–471.

Spohn H (1987) Interacting Brownian particles: a study of Dyson's model. In: Papanicolau G (ed.) *Hydrodynamic Behavior and Interacting Particle Systems*, IMA Vol. Math. Appl. vol. 9, pp. 157–179. New York: Springer.

Tracy CA and Widom H (1998) Correlation functions, cluster functions, and spacing distributions for random matrices. *Journal of Statistical Physics* 92(5/6): 809–835.

# Diagrammatic Techniques in Perturbation Theory

**G Gentile**, Università degli Studi "Roma Tre", Rome, Italy



## Introduction

Consider the dynamical system on $\mathbb{R}^d$ described by the equation

$$\dot{u} = \frac{\mathrm{d}u}{\mathrm{d}t} = G(u) + \varepsilon F(u) \qquad [1]$$

where $F, G : \mathcal{S} \subset \mathbb{R}^d \to \mathbb{R}^d$ are analytic functions and $\varepsilon$ a real (small) parameter. Suppose also that for $\varepsilon = 0$ a solution $u_0 : \mathbb{R} \to \mathcal{S}$ (for some initial condition $u_0(0) = \bar{u}$) is known.

We look for a solution of [1] which is a perturbation of $u_0$, that is, for a solution $u$ which can be written in the form $u = u_0 + U$, with $U = O(\varepsilon)$ and $U(0) = \bar{U} \equiv u(0) - \bar{u}$. Then we consider the variational equation

$$\dot{U} = M(t)U + \Phi(t), \quad M_{ij}(t) = \partial_{u_i} G_j(u_0(t)) \qquad [2]$$

where $\Phi(t) = \tilde{\Phi}(u_0(t), U)$, with $\tilde{\Phi}(u_0, U) = G(u_0 + U) - G(u_0) - \partial_u G(u_0)U + \varepsilon F(u_0 + U)$. By defining the Wronskian matrix $W$ as the solution of the matrix equation $\dot{W} = M(t)W$ such that $W(0) = \mathbb{1}$ (the columns of $W$ are given by $d$ independent solutions of the linear equation $\dot{u} = M(t)u$), we can write

$$U(t) = W(t)\bar{U} + W(t)\int_0^t \mathrm{d}\tau W^{-1}(\tau)\Phi(\tau) \qquad [3]$$

If we expect the solution $U$ to be of order $\varepsilon$, we can try to write it as a Taylor series in $\varepsilon$, that is,

$$U(t) = \sum_{k=1}^{\infty} \varepsilon^k U^{(k)} \qquad [4]$$

and, by inserting [4] into [3] and equating the coefficients with the same Taylor order, we obtain

$$U^{(k)}(t) = W(t)\bar{U}^{(k)} + W(t)\int_0^t \mathrm{d}\tau\, W^{-1}(\tau)\Phi^{(k)}(\tau) \qquad [5]$$

where $\Phi^{(k)}(t)$ is defined as

$$\Phi^{(1)}(t) = F(u_0(t))$$

$$\Phi^{(k)}(t) = \sum_{p=2}^{\infty} \frac{1}{p!}\frac{\partial^p G}{\partial u^p}(u_0(t)) \sum_{k_1+\cdots+k_p=k} U^{(k_1)}\cdots U^{(k_p)} + \sum_{p=1}^{\infty}\frac{1}{p!}\frac{\partial^p F}{\partial u^p}(u_0(t)) \times \sum_{k_1+\cdots+k_p=k-1} U^{(k_1)}\cdots U^{(k_p)} \quad k \geq 2 \qquad [6]$$

Hence $\Phi^{(k)}(t)$ depends only on coefficients of orders strictly less than $k$. In this way, we obtain an algorithm useful for constructing the solution recursively, so that the problem is solved, up to (substantial) convergence problems.

## Historical Excursus

The study of a system like [1] by following the strategy outlined above can be hopeless if we do not make some further assumptions on the types of motions we are looking for.

We shall see later, in a concrete example, that the coefficients $U^{(k)}(t)$ can increase in time, in a $k$-dependent way, thus preventing the convergence of the series for large $t$. This is a general feature of this class of problems: if no care is taken in the choice of the initial datum, the algorithm can provide a reliable description of the dynamics only for a very short time.

However, if one looks for solutions having a special dependence on time, things can work better. This happens, for instance, if one looks for quasiperiodic solutions, that is, functions which depend on time through the variable $\psi = \omega t$, with $\omega \in \mathbb{R}^N$ a vector with rationally independent components, that is such that $\omega \cdot \nu \neq 0$ for all $\nu \in \mathbb{Z}^N \setminus \{0\}$ (the dot denotes the standard inner product, $\omega \cdot \nu = \omega_1\nu_1 + \cdots + \omega_N\nu_N$). A typical problem of interest is: what happens to a quasiperiodic solution $u_0(t)$ when a perturbation $\varepsilon F$ is added to the unperturbed vector field $G$, as in [1]? Situations of this type arise when considering perturbations of integrable systems: a classical example is provided by planetary motion in celestial mechanics.

Perturbation series such as [4] have been extensively studied by astronomers in order to obtain a more accurate description of the celestial motions compared to that following from Kepler's theory (in which all interactions between planets are neglected and the planets themselves are considered as points). In particular, we recall the works of Newcomb and Lindstedt (series such as [4] are now known as Lindstedt series). At the end of the nineteenth century, Poincaré showed that the series describing quasiperiodic motions are well defined up to any perturbation order $k$ (at least if the perturbation is a trigonometric polynomial), provided that the components of $\omega$ are assumed to be rationally independent: this means that, under this condition, the coefficients $U^{(k)}(t)$ are defined for all $k \in \mathbb{N}$. However, Poincaré also showed that, in general, the series are divergent; this is due to the fact that, as seen later, in the perturbation series small divisors $\omega \cdot \nu$ appear, which, even if they do not vanish, can be arbitrarily close to zero.

The convergence of the series can be proved indeed (more generally for analytic perturbations, or even those that are differentiably smooth enough) by assuming on $\omega$ a stronger nonresonance condition, such as the Diophantine condition

$$|\omega \cdot \nu| > \frac{C_0}{|\nu|^\tau} \quad \forall \nu \in \mathbb{Z}^N \setminus \{0\} \qquad [7]$$

where $|\nu| = |\nu_1| + \cdots + |\nu_N|$, and $C_0$ and $\tau$ are positive constants. We note that the set of vectors satisfying [7] for some positive constant $C_0$ have full measure in $\mathbb{R}^N$ provided one takes $\tau > N - 1$.

Such a result is part of the Kolmogorov–Arnold–Moser (KAM) theorem, and it was first proved by Kolmogorov in 1954, following an approach quite different fom the one described here. New proofs were given in 1962 by Arnol'd and by Moser, but only very recently, in 1988, Eliasson gave a proof in which a bound $C^k$ is explicitly derived for the coefficients $U^{(k)}(t)$, again implying convergence for $\varepsilon$ to be small enough.

Eliasson's work was not immediately known widely, and only after publication of papers by Gallavotti and by Chierchia and Falcolini, in which Eliasson's ideas were revisited, did his work become fully appreciated. The study of perturbation series [4] employs techniques very similar to those typical of a very different field of mathematical physics, the quantum field theory, even if such an analogy was stressed and used to full extent only in subsequent papers.

The techniques have so far been applied to a wide class of problems of dynamical systems: a list of original results is given at the end.

## A Paradigmatic Example

Consider the case $\mathcal{S} = \mathcal{A} \times \mathbb{T}^N$, with $\mathcal{A}$ an open subset of $\mathbb{R}^N$, and let $\mathcal{H}_0 : \mathcal{A} \to \mathbb{R}$ and $f : \mathcal{A} \times \mathbb{T}^N \to \mathbb{R}$ be two analytic functions. Then consider the Hamiltonian system with Hamiltonian $\mathcal{H}(A,\alpha) = \mathcal{H}_0(A) + \varepsilon f(A,\alpha)$. The corresponding equations describe a dynamical system of the form [1], with $u = (A,\alpha)$, which can be written explicitly:

$$\begin{cases} \dot{A} = -\varepsilon \partial_\alpha f(A,\alpha) \\ \dot{\alpha} = \partial_A \mathcal{H}_0(A) + \varepsilon \partial_A f(A,\alpha) \end{cases} \qquad [8]$$

Suppose, for simplicity, $\mathcal{H}_0(A) = A^2/2$ and $f(A,\alpha) = f(\alpha)$, where $A^2 = A \cdot A$. Then, we obtain for $\alpha$ the following closed equation:

$$\ddot{\alpha} = -\varepsilon \partial_\alpha f(\alpha) \qquad [9]$$

while $A$ can be obtained by direct integration once [9] has been solved. For $\varepsilon = 0$, [9] gives trivially $\alpha = \alpha_0(t) \equiv \alpha_0 + \omega t$, where $\omega = \partial_A \mathcal{H}_0(A_0) = A_0$ is

called the rotation (or frequency) vector. Hence, for $\varepsilon = 0$ all solutions are quasiperiodic. We are interested in the preservation of quasiperiodic solutions when $\varepsilon \neq 0$.

For $\varepsilon \neq 0$, we can write, as in [3],

$$\alpha = \alpha_0(t) + a(t), \quad a(t) = \sum_{k=1}^{\infty} \varepsilon^k a^{(k)}(t) \qquad [10]$$

where $a^{(k)}$ is determined as the solution of the equation

$$a^{(k)} = t\bar{A}^{(k)} + \bar{a}^{(k)}(t) - \int_0^t \mathrm{d}\tau \int_0^\tau \mathrm{d}\tau' [\partial_\alpha f(\alpha(\tau'))]^{(k-1)} \qquad [11]$$

with $[\partial_\alpha f(\alpha(\tau'))]^{(k-1)}$ expressed as in [6].

The quasiperiodic solutions with rotation vector $\omega$ could be written as a Fourier series, by expanding

$$a^{(k)}(t) = \sum_{\nu \in \mathbb{Z}^N} \mathrm{e}^{\mathrm{i}\nu\cdot\omega t} a_\nu^{(k)} \qquad [12]$$

with $\omega$ as before. If the series [10], with the Taylor coefficients as in [12], exists, it will describe a quasiperiodic solution analytic in $\varepsilon$, and in such a case we say that it is obtained by continuation of the unperturbed one with rotation vector $\omega$, that is $\alpha_0(t)$.

Suppose that the integrand $[\partial_\alpha f(\alpha(\tau'))]^{(k-1)}$ in [11] has vanishing average. Then the integral over $\tau'$ in [11] produces a quasiperiodic function, which in general has a nonvanishing average, so that the integral over $\tau$ produces a quasiperiodic function plus a term linear in $t$. If we choose $\bar{A}^{(k)}$ in [11] so as to cancel out exactly the term linear in time, we end up with a quasiperiodic function. In Fourier space, an explicit calculation gives, for all $\nu \neq 0$,

$$a_\nu^{(1)} = \frac{1}{(\omega\cdot\nu)^2}\mathrm{i}\nu f_\nu$$
$$a_\nu^{(k)} = \frac{1}{(\omega\cdot\nu)^2}\sum_{p=1}^{\infty}\sum_{\substack{k_1+\cdots+k_p=k-1\\ \nu_0+\nu_1+\cdots+\nu_p=\nu}} \frac{(\mathrm{i}\nu_0)^{p+1}}{p!} a_{\nu_1}^{(k_1)}\ldots a_{\nu_p}^{(k_p)}$$
$$k \geq 2 \qquad [13]$$

which again is suitable for an iterative construction of the solution. The coefficients $a_0^{(k)}$ are left undetermined, and we can fix them (arbitrarily) as identically vanishing.

Of course, the property that the integrand in [11] has zero average is fundamental; otherwise, terms increasing as powers of $t$ would appear (the so-called secular terms). Indeed, it is easy to realize that, if this happened, to order $k$ terms proportional to $t^{2k}$ could be present, thus requiring, at best, $|\varepsilon| < |t|^{-2}$ for convergence up to time $t$. This would exclude *a fortiori* the possibility of quasiperiodic solutions.

The aforementioned property of zero average can be verified only if the rotation vector is nonresonant, that is, if its components are rationally independent or, more particularly, if the Diophantine condition [7] is satisfied. Such a result was first proved by Poincaré, and it holds irrespective of how the parameters $\bar{a}^{(k)}$ appearing in [11] are fixed. This reflects the fact that quasiperiodic motions take place on invariant surfaces (KAM tori), which can be parameterized in terms of the angle variables $\alpha(t)$, so that the values $\bar{a}^{(k)}$ contribute to the initial phases, and the latter can be arbitrarily fixed.

The recursive equations [13] can be suitably studied by introducing a diagrammatic representation, as explained below.

## Graphs and Trees

A (connected) graph $\mathbb{G}$ is a collection of points, called vertices, and lines connecting all of them. We denote with $V(\mathbb{G})$ and $L(\mathbb{G})$ the set of vertices and the set of lines, respectively. A path between two vertices is a minimal subset of $L(\mathbb{G})$ connecting the two vertices. A graph is planar if it can be drawn in a plane without graph lines crossing.

A tree is a planar graph $\mathbb{G}$ containing no closed loops (cycles); in other words, it is a connected acyclic graph. One can consider a tree $\mathbb{G}$ with a single special vertex $\mathrm{v}_0$: this introduces a natural partial ordering on the set of lines and vertices, and one can imagine that each line carries an arrow pointing toward the vertex $\mathrm{v}_0$. We can add an extra oriented line $\ell_0$ connecting the special vertex $\mathrm{v}_0$ to another point which will be called the root of the tree; the added line will be called the root line. In this way, we obtain a rooted tree $\theta$ defined by $V(\theta) = V(\mathbb{G})$ and $L(\theta) = L(\mathbb{G}) \cup \ell_0$. A labeled tree is a rooted tree $\theta$ together with a label function defined on the sets $V(\theta)$ and $L(\theta)$.

Two rooted trees which can be transformed into each other by continuously deforming the lines in the plane in such a way that the latter do not cross each other (i.e., without destroying the graph structure) will be said to be equivalent. This notion of equivalence can also be extended to labeled trees, simply by considering equivalent two labeled trees if they can be transformed into each other in such a way that the labels also match.

Given two vertices $\mathrm{v}, \mathrm{w} \in V(\theta)$, we say that $\mathrm{w} \prec \mathrm{v}$ if $\mathrm{v}$ is on the path connecting $\mathrm{w}$ to the root line. One

can identify a line with the vertices it connects; given a line $\ell = (\mathrm{v}, \mathrm{w})$, one says that $\ell$ enters v and exits w. For each vertex v, we define the branching number as the number $p_\mathrm{v}$ of lines entering v.

The number of unlabeled trees with $k$ vertices can be bounded by the number of random walks with $2k$ steps, that is, by $4^k$.

The labels are as follows: with each vertex v we associate a mode label $\nu_\mathrm{v} \in \mathbb{Z}^N$, and with each line we associate a momentum $\nu_\ell \in \mathbb{Z}^N$, such that the momentum of the line leaving the vertex v is given by the sum of the mode labels of all vertices preceding v (with v being included): if $\ell = (\mathrm{v}', \mathrm{v})$ then $\nu_\ell = \sum_{\mathrm{w} \preceq \mathrm{v}} \nu_\mathrm{w}$. Note that for a fixed unlabeled tree the branching labels are uniquely determined, and, for a given assignment of the mode labels, the momenta of the lines are also uniquely determined.

Define

$$V_\mathrm{v} = \frac{(\mathrm{i}\nu_\mathrm{v})^{p_\mathrm{v}+1}}{p_\mathrm{v}!} f_{\nu_\mathrm{v}}, \qquad g_\ell = \frac{1}{(\omega \cdot \nu_\ell)^2} \qquad [14]$$

where the tensor $V_\mathrm{v}$ is referred to as the node factor of v and the scalar $g_\ell$ as the propagator of the line $\ell$. One has $|f_\nu| \le F\,\mathrm{e}^{-\kappa|\nu|}$, for suitable positive constants $F$ and $\kappa$, by the analyticity assumption. Then one can check that the coefficients $a_\nu^{(k)}$, defined in [12], for $\nu \neq 0$, can be expressed in terms of trees as

$$a_\nu^{(k)} = \sum_{\theta \in \Theta_\nu^{(k)}} \mathrm{Val}(\theta)$$
$$\mathrm{Val}(\theta) = \left( \prod_{\mathrm{v} \in V(\theta)} V_\mathrm{v} \right) \left( \prod_{\ell \in L(\theta)} g_\ell \right) \qquad [15]$$

where $\Theta_\nu^{(k)}$ denotes the set of all inequivalent trees with $k$ vertices and with momentum $\nu$ associated with the root line, while the coefficients $a_0^{(k)}$ can be fixed $a_0^{(k)} = 0$ for all $k \ge 1$, by the arbitrariness of the initial phases previously remarked. The property that $[\partial_\alpha f(\alpha(\tau'))]^{(k-1)}$ in [11] has zero average for all $k \ge 1$ implies that for all lines $\ell \in L(\theta)$ one has $g_\ell = (\omega \cdot \nu_\ell)^{-2}$ only for $\nu_\ell \neq 0$, whereas $g_\ell = 1$ for $\nu_\ell = 0$, so that the numerical values $\mathrm{Val}(\theta)$ are well defined for all trees $\theta$. If $a_0^{(k)} = 0$ for all $k \ge 1$, then $\nu_\ell \neq 0$ for all $\ell \in L(\theta)$.

The proof of [15] can be performed by induction on $k$. Alternatively, we can start from the recursive definition [13], whereby the trees naturally arise in the following way.

Represent graphically the coefficient $a_\nu^{(k)}$ as in Figure 1; to keep track of the labels $k$ and $\nu$, we assign $k$ to the black bullet and $\nu$ to the line. For $k = 1$, the black bullet is meant as a grey vertex (like the ones appearing in Figure 3).

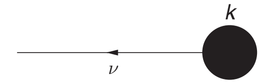



**Figure 1** Graphical representation of $a_\nu^{(k)}$.

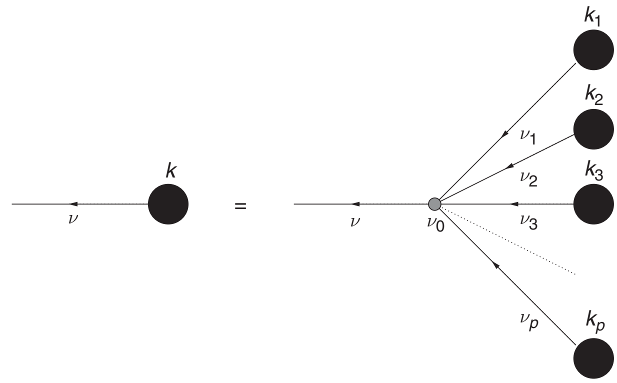



**Figure 2** Graphical representation of the recursive equation [13].

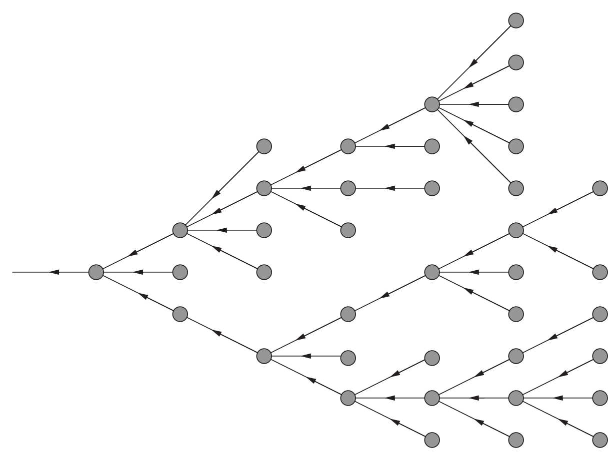


**Figure 3** An example of tree to be summed over in [15] for $k = 39$. The labels are not explicitly shown. The momentum of the root line is $\nu$, so that the mode labels satisfy the constraint $\sum_{v \in V(\theta)} \nu_v = \nu$.

Then recursive equation [13] can be graphically represented as the diagram in Figure 2, provided that we associate with the (grey) vertex $\mathrm{v}_0$ the node factor $V_{\mathrm{v}_0}$, with $\nu_{\mathrm{v}_0} = \nu_0$ and $p_{\mathrm{v}_0} = p$ denoting the number of lines entering $\mathrm{v}_0$, and with the lines $\ell_i$, $i = 1, \ldots, p$, entering $\mathrm{v}_0$ the momenta $\nu_{\ell_i}$, respectively. Of course, the sums over $p$ and over the possible assignments of the labels $\{k_i\}_{i=1}^p$ and $\{\nu_i\}_{i=0}^p$ are understood. Each black bullet on the right-hand side of Figure 2, together with its exiting line looks like the diagram on the left-hand side, so that it represents $a_{\nu_i}^{(k_i)}$, $i = 1, \ldots, p$. Note that Figure 2 has to be interpreted in the following way: if one associates with the diagram as drawn in the right-hand side a numerical value (as

described above) and one sums all the values over the assignments of the labels, then the resulting quantity is precisely $a_\nu^{(k)}$.

The (fundamental) difference between the black bullets on the right- and left-hand sides is that the labels $k_i$ of the latter are strictly less than $k$, hence we can iterate the diagrammatic decomposition simply by expressing again each $a_{\nu_i}^{(k_i)}$ as $a_\nu^{(k)}$ in [13], and so on, until one obtains a tree with $k$ grey vertices and no black bullets; see Figure 3, where the labels are not explicitly written. This corresponds to the tree expansion [15].

Any tree appearing in [15] is an example of what physicists call a Feynman graph, while the diagrammatic rules one has to follow in order to associate to the tree $\theta$ its right numerical value $\mathrm{Val}(\theta)$ are usually called the Feynman rules for the model under consideration. Such a terminology is borrowed from quantum field theory.

## Multiscale Analysis and Clusters

Suppose we replace [9] with $\alpha = \varepsilon\partial_\alpha f(\alpha)$, so that no small divisors appear (that is, $g_\ell = 1$ in [14]). Then convergence is easily proved for $\varepsilon$ small enough, since (by using the identity $\sum_{\mathrm{v}\in V(\theta)} p_\mathrm{v} = k-1$ and the inequality $\mathrm{e}^{-x}x^k/k! \le 1$ for all $x \in \mathbb{R}_+$ and all $k \in \mathbb{N}$), one finds

$$\prod_{\mathrm{v}\in v(\theta)} |V_\mathrm{v}| \le \left(\frac{4^2 F}{\kappa^2}\right)^k \mathrm{e}^{-\kappa|\nu|/4} \left(\prod_{\mathrm{v}\in v(\theta)} \mathrm{e}^{-\kappa|\nu_\mathrm{v}|/4}\right) \qquad [16]$$

and the sum over the mode labels can be performed by using the exponential decay factors $\mathrm{e}^{-\kappa|\nu_\mathrm{v}|/4}$, while the sum over all possible unlabeled trees gives $4^k$. In particular, analyticity in $t$ follows.

Of course, the interesting case is when the propagators are present. In such a case, even if no division by zero occurs, as $\omega\cdot\nu_\ell \neq 0$ (by the assumed Diophantine condition [13] and the absence of secular terms discussed previously), the quantities $\omega\cdot\nu_\ell$ in [14] can be very small.

Then we can introduce a scale $h$ characterizing the size of each propagator: we say that a line $\ell$ has scale $h_\ell = h \ge 0$ if $\omega\cdot\nu_\ell$ is of order $2^{-h}C_0$ and scale $h_\ell = -1$ if $\omega\cdot\nu_\ell$ is greater than $C_0$ (of course, a more formal definition can be easily envisaged, for which the reader is referred to the original papers). Then, we can bound $|\omega\cdot\nu_\ell| \ge 2^{-h}C_0$ for any $\ell \in L(\theta)$, and write

$$\prod_{\ell\in L(\theta)} |g_\ell| \le C_0^{-2k} \prod_{h=0}^{\infty} 2^{2hN_h(\theta)}$$
$$\le C_0^{-2k} 2^{2h_0 k} \exp\left(\sum_{h=h_0}^{\infty} 2\log 2\, hN_h(\theta)\right) \qquad [17]$$

where $N_h(\theta)$ is the number of lines in $L(\theta)$ with scale $h$ and $h_0$ is a (so far arbitrary) positive integer. The problem is then reduced to that of finding an estimate for $N_h(\theta)$.

To identify which kinds of tree are the source of problems, we introduce the notion of a cluster and a self-energy graph. A cluster $T$ with scale $h_T$ is a connected set of nodes linked by a continuous path of lines with the same scale label $h_T$ or a lower one and which is maximal, namely all the lines not belonging to $T$ but connected to it have scales higher than $h_T$ and at least one line in $T$ has scale $h_T$. An inclusion relation is established between clusters, in such a way that the innermost clusters are the clusters with lowest scale, and so on. Each cluster $T$ can have an arbitrary number of lines coming into it (entering lines), but only one or zero lines coming out from it (exiting line): lines of $T$ which either enter or exit $T$ are called external lines. A cluster $T$ with only one entering line $\ell_T^2$ and with one exiting line $\ell_T^1$ such that one has $\nu_{\ell_T^1} = \nu_{\ell_T^2}$ will be called a self-energy graph (SEG) or resonance. In such a case, the line $\ell_T^1$ is called a resonant line. Examples of clusters and SEGs are suggested by the bubbles in Figure 4; the mode labels are not represented, whereas the scales of the lines are explicitly written.

If $S_h(\theta)$ is the number of SEGs whose resonant lines have scales $h$, then $N_h^*(\theta) = N_h(\theta) - S_h(\theta)$ will denote the number of nonresonant lines with scale $h$.

A fundamental result, known as Siegel–Bryuno lemma, shows that, for some positive constant $c$, one has

$$N_h^*(\theta) \le 2^{h/\tau} c \sum_{\mathrm{v}\in V(\theta)} |\nu_\mathrm{v}| \qquad [18]$$

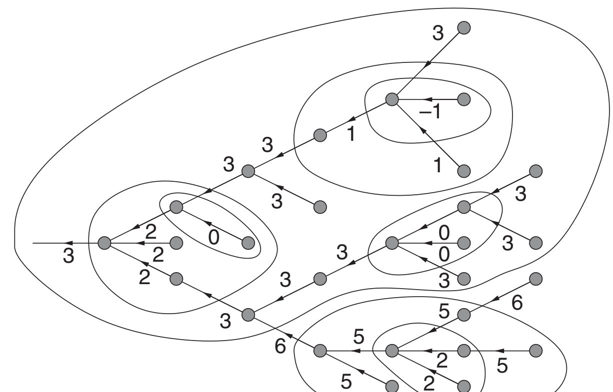



**Figure 4** Examples of clusters and SEGs. Note that the tree itself is a cluster (with scale 6), and each of the two clusters with one entering and one exiting lines is a SEG only if the momenta of its external lines are equal to each other.

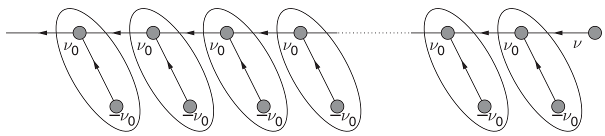



**Figure 5** Example of tree whose value grows like a factorial.

which, if inserted into [17] instead of $N_h(\theta)$, would give a convergent series; then $h_0$ should be chosen in such a way that the sum of the series in [17] is less than, say, $\kappa\sum_{v\in V(\theta)}|\nu_v|/8$.

The bound [18] is a very deep one, and was originally proved by Siegel for a related problem (Siegel's problem), in which, in the formalism followed here, SEGs do not occur; such a bound essentially shows that accumulation of small divisors is possible only in the presence of SEGs. A possible tree with $k$ vertices whose value can be proportional to some power of $k!$ is represented in Figure 5, where a chain of $(k-1)/2$ SEGs, $k$ odd, is drawn with external lines carrying a momentum $\nu$ such that $\omega\cdot\nu \approx C_0|\nu|^{-\tau}$.

In order to take into account the resonant lines, we have to add a factor $(\omega\cdot\nu_\ell)^{-2}$ for each resonant line $\ell$. It is a remarkable fact that, even if there are trees whose value cannot be bounded as a constant to the power $k$, there are compensations (that is, partial cancellations) between the values of all trees with the same number of vertices, such that the sum of all such trees admits a bound of this kind.

The cancellations can be described graphically as follows. Consider a tree $\theta$ with a SEG $T$. Then take all trees which can be obtained by shifting the external lines of $T$, that is, by attaching such lines to all possible vertices internal to $T$, and sum together the values of all such trees. An example is given in Figure 6. The corresponding sum turns out to be proportional to $(\omega\cdot\nu)^2$, if $\nu$ is the momentum of the resonant line of $T$, and such a factor compensates exactly the propagator of this line. The argument above can be repeated for all SEGs: this requires a little care because there are SEGs which are inside some other SEGs. Again, for details and a more formal discussion, the reader is referred to original papers.

The conclusion is that we can take into account the resonant lines: this simply adds an extra constant raised to the power $k$, so that an overall estimate $C^k$, for some $C>0$, holds for $U^{(k)}(t)$, and the convergence of the series follows.

## Other Examples and Applications

The discussion carried out so far proves a version of the KAM theorem, for the system described by [9], and it is inspired by the original papers by Eliasson (1996) and, mostly, by Gallavotti (1994).

Here we list some problems in which original results have been proved by means of the diagrammatic techniques described above, or by some variants of them. These are discussed in the following.

The first generalization one can think of is the problem of conservation in quasi-integrable systems of resonant tori (that is, invariant tori whose frequency vectors have rationally dependent components). Even if most of such tori disappear as an effect of the perturbation, some of them are conserved as lower-dimensional tori, which, generically, become of either elliptic or hyperbolic or mixed type according to the sign of $\varepsilon$ and the perturbation. With techniques extending those described here (introducing also, in particular, a suitable resummation procedure for divergent series), this has been done by Gallavotti and Gentile; see Gallavotti *et al.* (2004) and Gallavotti and Gentile (2005) for an account.

An expansion like the one considered so far can be envisaged also for the motions occurring on the stable and unstable manifolds of hyperbolic lower-dimensional tori for perturbations of Hamiltonians describing a system of rotators (as in the previous case) plus $n$ pendulum-like systems. In such a case, the function $G(u)$ has a less simple form. For $n=1$, one can look for solutions which depend on time through two variables, $\psi=\omega t$ and $x=e^{-gt}$, with $(\omega,g)\in\mathbb{R}^{N+1}$, and $\omega$ Diophantine as before and $g$ related to the timescale of the pendulum. This has been worked out by Gallavotti (1994), and then used by Gallavotti *et al.* (1999) to study a class of three-timescale systems, in order to obtain a lower

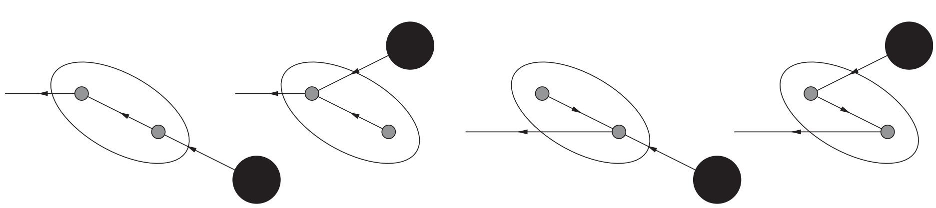


**Figure 6** Example of SEGs whose values have to be summed together in order to produce the cancellation discussed in the text. The mode labels are all fixed.

bound on the homoclinic angles (i.e., the angles between the stable and unstable manifolds of hyperbolic tori which are preserved by the perturbation). The formalism becomes a little more involved, essentially because of the entries of the Wronskian matrix appearing in [5]. In such a case, the unperturbed solution $u_0(t)$ corresponds to the rotators moving linearly with rotation vector $\omega$ and the pendulum moving along its separatrix; a nontrivial fact is that if $g_0$ denotes the Lyapunov exponent of the pendulum in the absence of the perturbation, then one has to look for an expansion in $x = e^{-gt}$ with $g = g_0 + O(\varepsilon)$, because the perturbation changes the value of such an exponent.

The same techniques have also been applied to study the relation of the radius of convergence of the standard map, an area-preserving diffeomorphism from the cylinder to itself, which has been widely studied in the literature since the original papers by Greene and by Chirikov, both appeared in 1979, with the arithmetical properties of the rotation vector (which is, in this case, just a number). In particular, it has been proved that the radius of convergence is naturally interpolated through a function of the rotation number known as Bryuno function (which has been introduced by Yoccoz as the solution of a suitable functional equation completely independent of the dynamics); see Berretti and Gentile (2001) for a review of results of this and related problems.

Also the generalized Riccati equation $\dot{u} - iu^2 - 2if(\omega t) + i\varepsilon^2 = 0$, where $\omega \in \mathbb{T}^d$ is Diophantine and $f$ is an analytic periodic function of $\psi = \omega t$, has been studied with the diagrammatic technique by Gentile (2003). Such an equation is related to two-level quantum systems (as first used by Barata), and existence of quasiperiodic solutions of the generalized Riccati equation for a large measure set $\mathcal{E}$ of values of $\varepsilon$ can be exploited to prove that the spectrum of the corresponding two-level system is pure point for those values of $\varepsilon$; analogously, one can prove that, for fixed $\varepsilon$, one can impose some further nonresonance conditons on $\omega$, still leaving a full measure set, in such a way that the spectrum is pure point. (We note, in addition, that, technically, such a problem is very similar to that of studying conservation of elliptic lower-dimensional tori with one normal frequency.)

Finally we mention a problem of partial differential equations, where, of course, the scheme described above has to be suitably adapted: this is the study of periodic solutions for the nonlinear wave equation $u_{tt} - u_{xx} + mu = \varphi(u)$, with Dirichlet boundary conditions, where $m$ is a real parameter (mass) and $\varphi(u)$ is a strictly nonlinear analytic odd function. Gentile and Mastropietro (2004) reproduced the result of Craig and Wayne for the existence of periodic solutions for a large measure set of periods, and, in a subsequent paper by the same authors with Procesi (2005), an analogous result was proved in the case $m = 0$, which had previously remained an open problem in literature.

*See also:* Averaging Methods; Integrable Systems and Discrete Geometry; KAM Theory and Celestial Mechanics; Stability Theory and KAM.

## Further Reading

Berretti A and Gentile G (2001) Renormalization group and field theoretic techniques for the analysis of the Lindstedt series. *Regular and Chaotic Dynamics* 6: 389–420.

Chierchia L and Falcolini C (1994) A direct proof of a theorem by Kolmogorov in Hamiltonian systems. *Annali della Scuola Normale Superiore di Pisa* 21: 541–593.

Eliasson LH (1996) Absolutely convergent series expansions for quasi periodic motions. *Mathematical Physics Electronic Journal* 2, paper 4 (electronic), Preprint 1988.

Gallavotti G (1994) Twistless KAM tori, quasi flat homoclinic intersections, and other cancellations in the perturbation series of certain completely integrable Hamiltonian systems. A review. *Reviews in Mathematical Physics* 6: 343–411.

Gallavotti G, Bonetti F, and Gentile G (2004) *Aspects of the Ergodic, Qualitative and Statistical Theory of Motion*. Berlin: Springer.

Gallavotti G and Gentile G (2005) Degenerate elliptic tori. *Communications in Mathematical Physics* 257: 319–362.

Gallavotti G, Gentile G, and Mastropietro V (1999) Separatrix splitting for systems with three time scales. *Communications in Mathematical Physics* 202: 197–236.

Gentile G (2003) Quasi-periodic solutions for two-level systems. *Communications in Mathematical Physics* 242: 221–250.

Gentile G and Mastropietro V (2004) Construction of periodic solutions of the nonlinear wave equation with Dirichlet boundary conditions by the Lindstedt series method. *Journal de Mathématiques Pures et Appliquées* 83: 1019–1065.

Gentile G, Mastropietro V, and Procesi M (2005) Periodic solutions for completely resonant nonlinear wave equations with Dirichlet boundary conditions. *Communications in Mathematical Physics* 256: 437–490.

Harary F and Palmer EM (1973) *Graphical Enumeration*. New York: Academic Press.

Poincaré H (1892–99) *Les méthodes nouvelles de la mécanique céleste*, vol. I–III. Paris: Gauthier-Villars.

# Dimer Problems

**R Kenyon**, University of British Columbia, Vancouver, BC, Canada



## Definitions

The dimer model arose in the mid-twentieth century as an example of an exactly solvable statistical mechanical model in two dimensions with a phase transition. It is used to model a number of physical processes: free fermions in 1 dimension, the two-dimensional Ising model, and various other two-dimensional statistical-mechanical models at restricted parameter values, such as the 6- and 8-vertex models and $O(n)$ models. A number of observable quantities such as the "height function" and densities of motifs have been shown to have conformal invariance properties in the scaling limit (when the lattice spacing tends to zero).

Recently, the model is also used as an elementary model of crystalline surfaces in $\mathbb{R}^3$.

A dimer covering, or perfect matching, of a graph is a set of edges ("dimers") which covers every vertex exactly once. In other words, it is a pairing of adjacent vertices (see Figure 1a which is a dimer covering of an $8 \times 8$ grid). Dimer coverings of a grid are sometimes represented as domino tilings, that is, tilings with $2 \times 1$ rectangles (Figure 1b). The dimer model is the study of the set of dimer coverings of a graph. Typically, the underlying graph is taken to be a regular lattice in two dimensions, for example, the square grid or the honeycomb lattice, or a finite part of such a lattice.

Dimer coverings of the honeycomb graph are in bijection with tilings of plane regions with 60° rhombi, also known as lozenges (see Figure 2). These tilings in turn are projections of piecewise-linear surfaces in $\mathbb{R}^3$ composed of unit squares in the 2-skeleton of $\mathbb{Z}^3$. So one can think of honeycomb dimer coverings as modeling discrete surfaces in $\mathbb{R}^3$. These surfaces are monotone in the sense that the orthogonal projection to the plane $P_{111} = \{(x, y, z) | x + y + z = 0\}$ is injective.

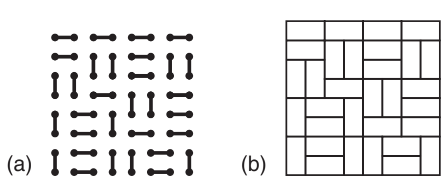



**Figure 1** A dimer covering of a grid and the corresponding domino tiling.

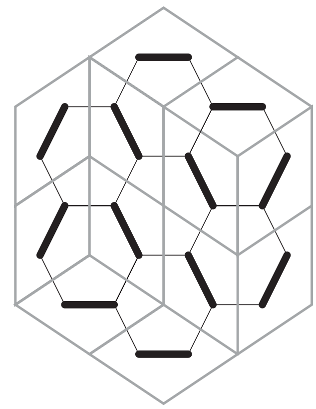


**Figure 2** Honeycomb dimers (solid) and the corresponding "lozenge" tilings (gray).

Other models related to the dimer model are:

- The spanning tree model on planar graphs. The set of spanning trees on a planar graph is in bijection with the set of dimer coverings on an associated bipartite planar graph. Conversely, dimer coverings of a bipartite planar graph are in bijection with directed spanning trees on an associated graph.
- The Ising model on a planar graph with zero external field can be modeled with dimers on an associated planar graph.
- Plane partitions (three-dimensional versions of integer partitions). Viewing a plane partition along the (1, 1, 1)-direction, one sees a lozenge tiling of the plane.
- Annihilating random walks in one dimension can be modeled with dimers on an associated planar graph.
- The monomer-dimer model, where one allows a certain density of holes (monomers) in a dimer covering. This model is unsolved at present, although some partial results have been obtained.

### Gibbs Measures

The most general setting in which the dimer model can be solved is that of an arbitrary planar graph with energies on the edges. We define here the corresponding measure.

Let $G = (V, E)$ be a graph and $\mathcal{M}(G)$ the set of dimer coverings of $G$. Let $\mathcal{E}$ be a real-valued function on the edges of $G$, with $\mathcal{E}(e)$ representing the energy associated to a dimer on the bond $e$. One defines the energy of a dimer covering as the sum of the energies of those bonds covered with dimers.

The partition function of the model on $(G, \mathcal{E})$ is then the sum

$$Z = \sum_{C \in \mathcal{M}(G)} e^{-\mathcal{E}(C)/kT}$$

where the sum is over dimer coverings. In what follows we will take $kT = 1$ for simplicity. Note that $Z$ depends on both $G$ and $\mathcal{E}$.

The partition function is well defined for a finite graph and defines the Gibbs measure, which is by definition the probability measure $\mu = \mu_{\mathcal{E}}$ on the set $\mathcal{M}(G)$ of dimer coverings satisfying $\mu(C) = (1/Z)e^{-\mathcal{E}(C)}$ for a covering $C$.

For an infinite graph $G$ with fixed energy function $\mathcal{E}$, a Gibbs measure on $\mathcal{M}(G)$ is by definition any measure which is a limit of the Gibbs measures on a sequence of finite subgraphs which fill out $G$. There may be many Gibbs measures on an infinite graph, since this limit typically depends on the sequence of finite graphs. When $G$ is an infinite periodic graph (and $\mathcal{E}$ is periodic as well), it is natural to consider translation-invariant Gibbs measures; one can show that in the case of a bipartite, periodic planar graph the translation-invariant and ergodic Gibbs measures form a two-parameter family – see Theorem 3 below.

For a translation-invariant Gibbs measure $\nu$ which is a limit of Gibbs measures on an increasing sequence of finite graphs $G_n$, one can define the partition function per vertex of $\nu$ to be the limit

$$Z = \lim_{n \to \infty} Z(G_n)^{1/|G_n|}$$

where $|G_n|$ is the number of vertices of $G_n$. The free energy, or surface tension, of $\nu$ is $-\log Z$.

## Combinatorics

### Partition Function

One can compute the partition function for dimer coverings on a finite planar graph $G$ as the Pfaffian (square root of the determinant) of a certain antisymmetric matrix, the Kasteleyn matrix. The Kasteleyn matrix is an oriented adjacency matrix of $G$, indexed by the vertices $V$: orient the edges of a graph embedded in the plane so that each face has an odd number of clockwise oriented edges. Then define $K = (K_{vv'})$ with

$$K_{vv'} = \pm e^{-\mathcal{E}(vv')}$$

if $G$ has an edge $vv'$, with a sign according to the orientation of that edge, and $K_{vv'} = 0$ if $v, v'$ are not adjacent. We then have the following result of Kasteleyn:

**Theorem 1** $Z = |\mathrm{Pf}(K)| = \sqrt{|\det K|}$.

Here $\mathrm{Pf}(K)$ denotes the Pfaffian of $K$.

Such an orientation of edges (which always exists for planar graphs) is called a Kasteleyn orientation; any two such orientations can be obtained from one another by a sequence of operations consisting of reversing the orientations of all edges at a vertex.

If $G$ is a bipartite graph, that is, the vertices can be colored black and white with no neighbors having the same color, then the Pfaffian of $K$ is the determinant of the submatrix whose rows index the white vertices and columns index the black vertices. For bipartite graphs, instead of orienting the edges one can alternatively multiply the edge weights by a complex number of modulus 1, with the condition that the alternating product around each face (the first, divided by the second, times the third, as so on) is real and negative.

For nonplanar graphs, one can compute the partition function as a sum of Pfaffians; for a graph embedded on a surface of Euler characteristic $\chi$, this requires in general $2^{2-\chi}$ Pfaffians.

### Local Statistics

The inverse of the Kasteleyn matrix can be used to compute the local statistics, that is, the probability that a given set of edges occurs in a random dimer covering (random with respect to the Gibbs measure $\mu$).

**Theorem 2** *Let $S = \{(v_1, v_2), \ldots, (v_{2k-1}, v_{2k})\}$ be a set of edges of $G$. The probability that all these edges occur in a $\mu$-random covering is*

$$\Pr(S) = \left(\prod_{i=1}^{k} K_{v_{2i-1}, v_{2i}}\right) \mathrm{Pf}_{2k \times 2k}((K^{-1})_{v_i, v_j})$$

Again, for bipartite graphs the Pfaffian can be made into a determinant.

### Heights

**Bipartite graphs** Suppose $G$ is a bipartite planar graph. A 1-form on $G$ is simply a function on the set of oriented edges which is antisymmetric with respect to reversing the edge orientation: $f(-e) = -f(e)$ for an edge $e$. A 1-form can be identified with a flow: just flow by $f(e)$ along oriented edge $e$. The divergence of the flow $f$ is then $d^* f$. Let $\Omega$ be the space of flows on edges of $G$, with divergence 1 at each white vertex and divergence –1 at each black vertex, and such that the flow along each edge from white to black is in [0, 1]. From a dimer covering $M$ one can construct such a flow $\omega(M) \in \Omega$: just flow

one unit along each dimer, and zero on the remaining edges. The set $\Omega$ is a convex polyhedron in $\mathbb{R}^E$ and its vertices can be seen to be exactly the dimer coverings.

Given any two flows $\omega_1, \omega_2 \in \Omega$, their difference is a divergence-free flow. Its dual $(\omega_1 - \omega_2)^*$ (or conjugate flow) defined on the planar dual of $G$ is therefore the gradient of a function $h$ on the faces of $G$, that is, $(\omega_1 - \omega_2)^* = \mathrm{d}h$, where $h$ is well defined up to an additive constant.

When $\omega_1$ and $\omega_2$ come from dimer coverings, $h$ is integer valued, and is called the height difference of the coverings. The level sets of the function $h$ are just the cycles formed by the union of the two matchings. If we fix a "base point" covering $\omega_0$ and a face $f_0$ of $G$, we can then define the height function of any dimer covering (with flow $\omega$) to be the function $h$ with value zero at $f_0$ and which satisfies $\mathrm{d}h = (\omega - \omega_0)^*$.

**Nonbipartite graphs** On a nonbipartite planar graph the height function can be similarly defined modulo 2. Fix a base covering $\omega_0$; for any other covering $\omega$, the superposition of $\omega_0$ and $\omega$ is a set of cycles and doubled edges of $G$; the function $h$ is constant on the complementary components of these cycles and changes by 1 mod 2 across each cycle. We can think of the height modulo 2 as taking two values, or spins, on the faces of $G$, and the dimer chains are the spin-domain boundaries. In particular, dimers on a nonbipartite graph model can in this way model the Ising model on an associated dual planar graph.

## Thermodynamic Limit

By periodic planar graph we mean a graph $G$, with energy function on edges, for which translations by elements of $\mathbb{Z}^2$ or some other rank-2 lattice $\Gamma \subset \mathbb{R}^2$ are isomorphisms of $G$ preserving the edge energies, and such that the quotient $G/\mathbb{Z}^2$ is a finite graph. Without loss of generality we can take $\Gamma = \mathbb{Z}^2$. The standard example is $G = \mathbb{Z}^2$ with $\mathcal{E} \equiv 0$, which we refer to as "dimers on the grid." However, other examples display different global behaviors and so it is worthwhile to remain in this generality.

For a periodic planar graph $G$, an ergodic probability measure on $\mathcal{M}(G)$ is one which is translation invariant (the measure of a set is the same as any $\mathbb{Z}^2$-translate of that set) and whose invariant subsets have measure 0 or 1.

We will be interested in probability measures which are both ergodic and Gibbs (we refer to them as ergodic Gibbs measures, dropping the term "probability"). When $G$ is bipartite, there are multiple ergodic Gibbs measures (see Theorem 3 below). When $G$ is nonbipartite, it is conjectured that there is a single ergodic Gibbs measure.

In the remainder of this section we assume that $G$ is bipartite, and assume also that the $\mathbb{Z}^2$-action preserves the coloring of the edges as black and white (simply pass to an index-2 sublattice if not).

For integer $n > 0$ let $G_n = G/n\mathbb{Z}^2$, a finite graph on a torus (in other words, with periodic boundary conditions). For a dimer covering $M$ of $G_n$, we define $(h_x, h_y) \in \mathbb{Z}^2$ to be the horizontal and vertical height change of $M$ around the torus, that is, the net flux of $\omega(M) - \omega_0$ across a horizontal, respectively vertical, cut around the torus (in other words, $h_x, h_y$ are the horizontal and vertical periods around the torus of the 1-form $\omega(M) - \omega_0$). The characteristic polynomial $P(z, w)$ of $G$ is by definition

$$P(z, w) = \sum_{M \in \mathcal{M}(G_1)} \mathrm{e}^{-\mathcal{E}(M)} z^{h_x} w^{h_y} (-1)^{h_x h_y}$$

here the sum is over dimer coverings $M$ of $G_1 = G/\mathbb{Z}^2$, and $h_x, h_y$ depend on $M$. The polynomial $P$ depends on the base point $\omega_0$ only by a multiplicative factor involving a power of $z$ and $w$. From this polynomial most of the large-scale behavior of the ergodic Gibbs measures can be extracted.

The Gibbs measure on $G_n$ converges as $n \to \infty$ to the (unique) ergodic Gibbs measure $\mu$ with smallest free energy $F = -\log Z$. The unicity of this measure follows from the strict concavity of the free energy of ergodic Gibbs measures as a function of the slope, see below. The free energy $F$ of the minimal free energy measure is

$$F = -\frac{1}{(2\pi \mathrm{i})^2} \int_{S^1 \times S^1} \log P(z, w) \frac{\mathrm{d}z}{z} \frac{\mathrm{d}w}{w}$$

that is, minus the Mahler measure of $P$.

For any translation-invariant measure $\nu$ on $\mathcal{M}(G)$, the average slope $(s, t)$ of the height function for $\nu$-almost every tiling is by definition the expected horizontal and vertical height change over one fundamental domain, that is, $s = \mathbb{E}[h(f + (1, 0)) - h(f)]$ and $t = \mathbb{E}[h(f + (0, 1)) - h(f)]$ where $f$ is any face. This quantity $(s, t)$ lies in the Newton polygon of $P(z, w)$ (the convex hull in $\mathbb{R}^2$ of the set of exponents of monomials of $P$). In fact, the points in the Newton polygon are in bijection with the ergodic Gibbs measures on $\mathcal{M}(G)$:

**Theorem 3** *When $G$ is a periodic bipartite planar graph, any ergodic Gibbs measure has average slope $(s, t)$ lying in $N(P)$. Moreover, for every point $(s, t) \in N(P)$ there is a unique ergodic Gibbs measure $\mu(s, t)$ with that average slope.*

In particular, this gives a complete description of the set of all ergodic Gibbs measures. The ergodic Gibbs measure $\mu(s,t)$ of slope $(s,t)$ can be obtained as the limit of the Gibbs measures on $G_n$, when one conditions the configurations to have a particular slope approximating $(s,t)$.

### Ronkin Function and Surface Tension

The Ronkin function of $P$ is a map $R:\mathbb{R}^2 \to \mathbb{R}$ defined for $(B_x, B_y) \in \mathbb{R}^2$ by

$$R(B_x, B_y) = \frac{1}{(2\pi i)^2} \int_{S^1 \times S^1} \log P(z e^{B_x}, w e^{B_y}) \frac{dz}{z} \frac{dw}{w}$$

The Ronkin function is convex and its graph is piecewise linear on the complement of the amoeba $\mathbb{A}(P)$ of $P$, which is the image of the zero set $\{(z,w) \in \mathbb{C}^2 \mid P(z,w)=0\}$ under the map $(z,w) \mapsto (\log|z|, \log|w|)$ (see Figures 3 and 4 for an example).

The free energy $F(\mu(s,t))$ of $\mu(s,t)$, as a function of $(s,t) \in N(P)$, is the Legendre dual of the Ronkin function of $P(z,w)$: we have

$$F(\mu(s,t)) = R(B_x, B_y) - sB_x - tB_y$$

where

$$s = \frac{\partial R(B_x, B_y)}{\partial B_x}, \qquad t = \frac{\partial R(B_x, B_y)}{\partial B_y}$$

The continuous map $\nabla R:\mathbb{R}^2 \to N(P)$ which takes $(B_x, B_y)$ to $(s,t)$ is injective on the interior of $\mathbb{A}(P)$, collapses each bounded complementary component of $\mathbb{A}(P)$ to an integer point in the interior of $N(P)$, and collapses each unbounded complementary component of $\mathbb{A}(P)$ to an integer point on the boundary of $N(P)$.

Under the Legendre duality, the facets in the graph of the Ronkin function (i.e., maximal regions on which $R$ is linear) give points of nondifferentiability of the free energy $F$, as defined on $N(P)$. We refer to these points of nondifferentiability as "cusps." Cusps occur only at integer slopes $(s,t)$ (see Figure 5 for the free energy associated to the Ronkin function in Figure 4).

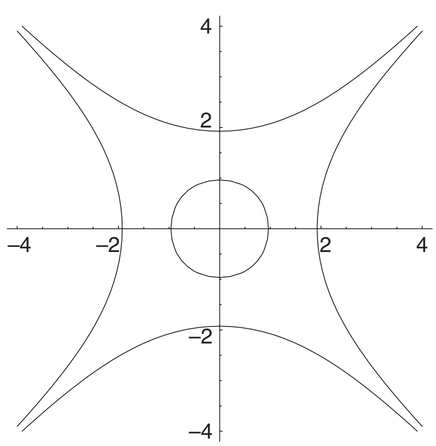



**Figure 3** The amoeba of $P(z,w)=5+z+1/z+w+1/w$, which is the characteristic polynomial for dimers on the periodic "square-octagon" lattice.

**Figure 4** Minus the Ronkin function of $P(z,w)=5+z+1/z+w+1/w$.

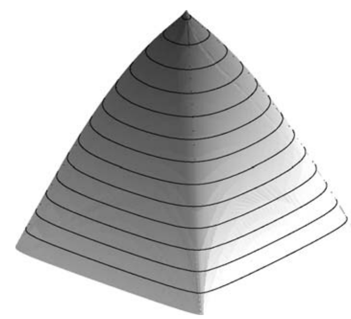


**Figure 5** (Negative of) the free energy for dimers on the square-octagon lattice.

By Theorem 3, the coordinates $(B_x, B_y)$ can also be used to parametrize the set of Gibbs measures $\mu(s,t)$ (but only those with slope $(s,t)$ in the interior of $N(P)$ or on the corners of $N(P)$ and boundary integer points). This parametrization is not one-to-one since when $(B_x, B_y)$ varies in a complementary component of the amoeba, the measure $\mu(s,t)$ does not change. On the interior of the amoeba the parametrization is one-to-one.

The remaining Gibbs measures, whose slopes are on the boundary of $N(P)$, can be obtained by taking limits of $(B_x, B_y)$ along the "tentacles" of the amoeba.

### Phases

The Gibbs measures $\mu(s,t)$ can be partitioned into three classes, or phases, according to the behavior of the fluctuations of the height function. If we measure the height at two distant points $x_1$ and $x_2$ in $G$, the average height difference, $\mathbb{E}[h(x_1)-h(x_2)]$, is a linear function of $x_1 - x_2$ determined by the average slope of the measure. The height fluctuation is defined to be the random variable $h(x_1)-h(x_2)-\mathbb{E}[h(x_1)-h(x_2)]$. This random variable depends on

the two points and we are interested in its behavior when $x_1$ and $x_2$ are far apart.

We say $\mu(s,t)$ is

1. "Frozen" if the height fluctuations are bounded almost surely.
2. "Rough" (or "liquid") if the covariance in the height function $\mathbb{E}[h(x_1)h(x_2)] - \mathbb{E}[h(x_1)]\mathbb{E}[h(x_2)]$ is unbounded as $|x_1 - x_2| \to \infty$.
3. "Smooth" (or "gaseous") if the covariance of the height function is bounded but the height fluctuations are unbounded.

The height fluctuations can be related to the decay of the entries of $K^{-1}$, which are in turn related to the decay of the Fourier coefficients of $1/P$. In particular, we have

**Theorem 4** *The measure $\mu(s,t)$ is respectively frozen, rough, or smooth according to whether $(B_x, B_y) = (\nabla R)^{-1}(s,t)$ is in the closure of an unbounded complementary component of $\mathbb{A}(p)$, in the interior of $\mathbb{A}(P)$, or in the closure of a bounded component of $\mathbb{A}(P)$.*

The characteristic polynomials $P$ which occur in the dimer model are not arbitrary: their algebraic curves $\{P = 0\}$ are all of a special type known as Harnack curves, which are characterized by the fact that the map from the zero-set of $P$ in $\mathbb{C}^2$ to its amoeba in $\mathbb{R}^2$ is at most two-to-one. In fact:

**Theorem 5** *By varying the edge energies all Harnack curves can be obtained as the characteristic polynomial of a planar dimer model.*

### Local Statistics

In the thermodynamic limit (on a periodic planar graph), local statistics of dimer coverings for the Gibbs measure of minimal free energy can be obtained from the limit of the inverse of the Kasteleyn matrix on the finite toroidal graphs $G_n$. This in turn can be computed from the Fourier coefficients of $1/P$.

As an example, let $G$ be the square grid $\mathbb{Z}^2$ and take $\mathcal{E} = 0$ (which corresponds to the uniform measure on configurations for finite graphs). An appropriate choice of signs for the Kasteleyn matrix is to put weights $1, -1$ on alternate horizontal edges and $i, -i$ on alternate vertical edges in such a way that around each white vertex the weights are cyclically $1, i, -1, -i$. For this choice of signs we have

$$K^{-1}_{(0,0),(x,y)} = \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi}\frac{\mathrm{e}^{-\mathrm{i}(\theta x+\phi y)}\mathrm{d}\theta\,\mathrm{d}\phi}{2\sin\theta + 2i\sin\phi}$$

This integral can be evaluated explicitly (see Figure 6 for values of $K^{-1}_{(0,0),(x,y)}$ near the origin; by

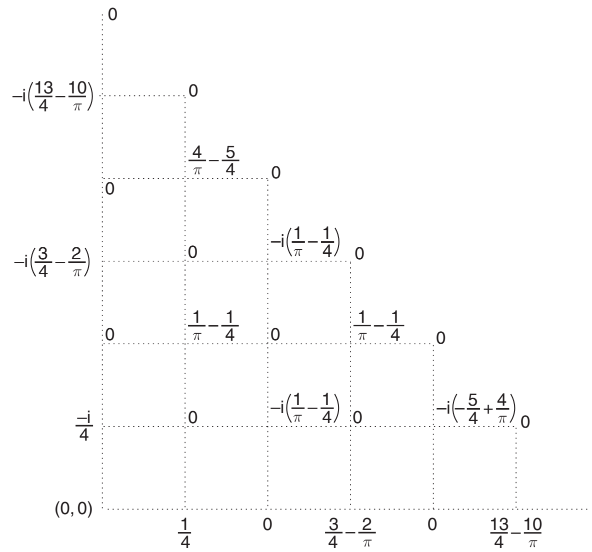



**Figure 6** Values of $K^{-1}$ on $Z^2$ with zero energies.

translation invariance $K^{-1}_{(x',y'),(x,y)} = K^{-1}_{(0,0),(x-x',y-y')}$) and values in other quadrants can be obtained by $K^{-1}_{(0,0),(x,y)} = -\mathrm{i}K^{-1}_{(0,0),(-y,x)}$).

As a sample computation, using Theorem 2, the probability that the dimer covering the origin points to the right and, simultaneously, the one covering $(0,1)$ points upwards is

$$K_{(0,0),(1,0)}K_{(0,2),(0,1)}\det\begin{pmatrix} K^{-1}_{(0,0),(1,0)} & K^{-1}_{(0,0),(0,1)} \\ K^{-1}_{(0,2),(1,0)} & K^{-1}_{(0,2),(0,1)}\end{pmatrix}$$

$$= 1\cdot(-\mathrm{i})\cdot\det\begin{pmatrix}\frac{1}{4} & \frac{-\mathrm{i}}{4} \\ -\frac{1}{4}+\frac{1}{\pi} & \frac{\mathrm{i}}{4}\end{pmatrix} = \frac{1}{4\pi}$$

Another computation which follows is the decay of the edge covariances. If $e_1, e_2$ are two edges at distance $d$, then $\Pr(e_1 \& e_2) - \Pr(e_1)\Pr(e_2)$ decays quadratically in $1/d$, since $K^{-1}((0,0),(x,y))$ decays like $1/(|x| + |y|)$.

## Scaling Limits

The scaling limit of the dimer model is the limit when the lattice spacing tends to zero.

Let us define the scaling limit in the following way. Let $\epsilon\mathbb{Z}^2$ be the square grid scaled by $\epsilon$, so the lattice mesh size is $\epsilon$. Fix a Jordan domain $U \subset \mathbb{R}^2$ and consider for each $\epsilon$ a subgraph $U_\epsilon$ of $\epsilon\mathbb{Z}^2$, bounded by a simple polygon, which tends to $U$ as $\epsilon \to 0$. We are interested in limiting properties of random dimer coverings of $U_\epsilon$, in the limit as $\epsilon \to 0$, for example, the fluctuations of the height function and edge densities.

The limit depends on the (sequence of) boundary conditions, that is, on the exact choice of approximating regions $U_\epsilon$. By changing $U_\epsilon$ one can change the limiting rescaled height function along the boundary. It is conjectured that the limit of the height function along the boundary of $U_\epsilon$ (scaled by $\epsilon \ldots$ and assuming this limit exists) determines essentially all of the limiting behavior in the interior, in particular the limiting local statistics.

Therefore, let $u$ be a real-valued continuous function on the boundary of $U$. Consider a sequence of subgraphs $U_\epsilon$ of $\epsilon\mathbb{Z}^2$, as $\epsilon \to 0$ as above, and whose height function along the boundary, when scaled by $\epsilon$, is approximating $u$. We discuss the limit of the model in this setting.

### Crystalline Surfaces

The height function allows us to view dimer coverings as random surfaces in $\mathbb{R}^3$: to a dimer covering of $G$, one associates the graph of its height function, extended in a piecewise linear fashion over the edges and faces of the dual $G^*$. These surfaces are then piecewise linear random surfaces, which resemble crystal surfaces in the sense that microscopically (on the scale of the lattice) they are rough, whereas their long-range behavior is smooth and facetted, as we now describe.

In the scaling limit, boundary conditions as described in the last paragraph of the previous section are referred to as "wire-frame" boundary conditions, since the graph of the height function can be thought of as a (random) surface spanning the wire frame defined by its boundary values.

In the scaling limit, there is a law of large numbers which says that the Gibbs measure on random surfaces (which is unique since we are dealing with a finite graph) concentrates, for fixed wire-frame boundary conditions, on a single surface $S_0$. That is, as the lattice spacing $\epsilon$ tends to zero, with probability tending to 1 the random surface lies close to a limiting surface $S_0$. The surface $S_0$ is the unique surface which minimizes the total surface tension, or free energy, for its fixed boundary values, that is, minimizes the integral over the surface of the $F(\mu(s,t))$, where $(s,t)$ is the slope of the surface at the point being integrated over. Existence and unicity of the minimizer follow from the strict convexity of the free energy/surface tension as a function of the slope.

At a point where the free energy has a cusp, the crystal surface $S_0$ will in general have a facet, that is, a region on which it is linear. Outside of the facets, one expects that $S_0$ is analytic, since the free energy is analytic outside the cusps.

### Fluctuations

While the scaled height function $\epsilon h$ in the scaling limit converges to its mean value $h_0$ (whose graph is the surface $S_0$), the fluctuations of the unrescaled height function $h - (1/\epsilon)h_0$ will converge in law to a random process on $U$.

In the simplest setting, that of honeycomb dimers with $\mathcal{E} \equiv 0$, and in the absence of facets, the height fluctuations converge to a continuous Gaussian process, the image of the Gaussian free field on the unit disk $\mathbb{D}$ under a certain diffeomorphism $\Phi$ (depending on $h_0$) of $\mathbb{D}$ to $U$.

In the particular case $h_0 = 0$, $\Phi$ is the Riemann map from $\mathbb{D}$ to $U$ and the law of the height fluctuations is just the Gaussian free field on $U$ (defined to be the Gaussian process whose covariance kernel is the Dirichlet Green's function). The conformal invariance of the Gaussian free field is the basis for a number of conformal invariance properties of the honeycomb dimer model.

### Densities of Motifs

Another observable of interest is the density field of a motif. A motif is a finite collection of edges, taken up to translation. For example, consider, for the square grid, the "L" motif consisting of a horizontal domino and a vertical domino aligned to form an "L," which we showed above to have a density $1/4\pi$ in the thermodynamic limit. The probability of seeing this motif at any given place is $1/4\pi$. However, in the scaling limit one can ask about the fluctuations of the occurrences of this motif: in a large ball around a point $x$, what is the distribution of $N_L - A/4\pi$, where $N_L$ is the number of occurrences of the motif, and $A$ is the area of the ball? These fluctuations form a random field, since there is a long-range correlation between occurrences of the motif.

It is known that on $\mathbb{Z}^2$, for the minimal free energy ergodic Gibbs measure, the rescaled density field

$$\frac{1}{\sqrt{A}}\left(N_L - \frac{A}{4\pi}\right)$$

converges as $\epsilon \to 0$ weakly to a Gaussian random field which is a linear combination of a directional derivative of the Gaussian free field and an independent white noise. A similar result holds for other motifs.

The joint distribution of densities of several motifs can also be shown to be Gaussian.

*See also:* Combinatorics: Overview; Determinantal Random Fields; Growth Processes in Random Matrix Theory; Statistical Mechanics and Combinatorial Problems; Statistical Mechanics of Interfaces.

## Further Reading

Cohn H, Kenyon R, and Propp J (2001) A variational principle for domino tilings. *Journal of American Mathematical Society* 14(2): 297–346.

Kasteleyn P (1967) Graph theory and crystal physics. In: *Graph Theory and Theoretical Physics*, pp. 43–110. London: Academic Press.

Kenyon R, Okounkov A, and Sheffield S (2005) Dimers and Amoebae. *Annals of Mathematics*, math-ph/0311005.

# Dirac Fields in Gravitation and Nonabelian Gauge Theory

**J A Smoller**, University of Michigan, Ann Arbor, MI, USA



## Introduction

In this article we describe some recent results (Finster *et al.* 1999a,b, 2000 a–c, 2002a) concerning the existence of both particle-like, and black hole solutions of the coupled Einstein–Dirac–Yang–Mills (EDYM) equations. We show that there are stable globally defined static, spherically symmetric solutions. We also show that for static black hole solutions, the Dirac wave function must vanish identically outside the event horizon. The latter result indicates that the Dirac particle (fermion) must either enter the black hole or tend to infinity.

The plan of the article is as follows. The next section describes the background material. It is followed by a discussion of the coupled EDYM equations for static, spherically symmetric particle-like and black hole solutions. The final section of the article is devoted to a discussion of these results.

## Background Material

### Einstein's Equations

We begin by describing the Einstein equation for the gravitational field (for more details, see, e.g., Adler *et al.* (1975)). We first note Einstein's hypotheses of general relativity (GR):

(E1) The gravitational field is the metric $g_{ij}$ in $3+1$ spacetime dimensions. The metric is assumed to be symmetric.
(E2) At each point in spacetime, the metric can be diagonalized as diag(–1,1,1,1).
(E3) The equations which describe the gravitational field should be covariant; that is, independent of the choice of coordinate system.

The hypothesis (E1) is Einstein's brilliant insight, whereby he "geometrizes" the gravitational field. (E2) means that there are inertial frames at each point (but not globally), and guarantees that special relativity (SR) is included in GR, while (E3) implies that the gravitational field equations must be tensor equations; that is, coordinates are an artifact, and physics should not depend on the choice of coordinates.

### Einstein's Equations of GR

The metric $g_{ij}=g_{ij}(x), i,j=0,1,2,3, x=(x^0,x^1,x^2,x^3)$, $x^0=ct$ ($c$=speed of light, $t$=time), is the metric tensor defined on four-dimensional spacetime. Einstein's equations are ten (tensor) equations for the unknown metric $g_{ij}$ (gravitational field), and take the form

$$R_{ij}-\tfrac{1}{2}Rg_{ij}=\sigma T_{ij} \qquad [1]$$

where the left-hand side $G_{ij}=R_{ij}-\frac{1}{2}Rg_{ij}$ is the Einstein tensor and depends only on the geometry, $\sigma=8\pi G/c^4$, where $G$ is Newton's gravitational constant, while $T_{ij}$, the energy–momentum tensor, represents the source of the gravitational field, and encodes the distribution of matter. (The word "matter" in GR refers to everything which can produce a gravitational field, including elementary particles, electromagnetic or Yang–Mills (YM) fields. From the Bianchi identities in geometry (cf. Adler *et al.* (1975)), the (covariant) divergence of the Einstein tensor, $G_{ij}$, vanishes identically, namely

$$G^j_{i,j}=0$$

so, on solutions of Einstein's equations,

$$T^j_{i,j}=0$$

and this in turn expresses the conservation of energy and momentum. The quantities which comprise the Einstein tensor are given as follows: first, from the metric tensor $g_{ij}$, we form the Levi-Civita connection $\Gamma^k_{ij}$ defined by:

$$\Gamma^k_{ij}=\frac{1}{2}g^{k\ell}\left(\frac{\partial g_{\ell j}}{\partial x^i}-\frac{\partial g_{i\ell}}{\partial x^j}-\frac{\partial g_{ij}}{\partial x^\ell}\right)$$

where ($4\times 4$ matrix) $[g^{k\ell}]=[g_{k\ell}]^{-1}$, and summation convention is employed; namely, an index which appears as both a subscript and a superscript is to be summed from 0 to 3. With the aid of $\Gamma^k_{ij}$, we can

construct the celebrated Riemann curvature tensor $R^i_{qk\ell}$:

$$R^i_{qk\ell} = \frac{\partial \Gamma^i_{q\ell}}{\partial x^k} - \frac{\partial \Gamma^i_{qk}}{\partial x^\ell} + \Gamma^i_{pk}\Gamma^p_{q\ell} - \Gamma^i_{p\ell}\Gamma^p_{qk}$$

Finally, the terms $R_{ij}$ and $R$ which appear in the Einstein tensor $G_{ij}$ are given by

$$R_{ij} = R^s_{isj}$$

(the Ricci tensor), and

$$R = g^{ij}R_{ij}$$

is the scalar curvature.

From the above definitions, one sees at once the enormous complexity of the Einstein equations. For this reason, one usually seeks solutions which have a high degree of symmetry, and in what follows, in this section, we shall only consider static, spherically symmetric solutions; that is, solutions which depend only on $r = |x| = \sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2}$. In this case, the metric $g_{ij}$ takes the form

$$\mathrm{d}s^2 = -T(r)^2\mathrm{d}t^2 + A(r)^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2 \qquad [2]$$

where $\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2$ is the standard metric on the unit 2-sphere, $r,\theta,\varphi$ are the usual spherical coordinates, and $t$ denotes time.

### Black Hole Solutions

Consider the problem of finding the gravitational field outside a ball of mass $M$ in $\mathbb{R}^3$; that is, there is no matter exterior to the ball. Solving Einstein's equations $G_{ij}{}' = 0$ gives the famous Schwarzschild solution (1916):

$$\mathrm{d}s^2 = -\left(1 - \frac{2m}{r}\right)c^2\mathrm{d}t^2 + \left(1 - \frac{2m}{r}\right)^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2 \qquad [3]$$

where $m = GM/c^2$. Since $2m$ has the dimensions of length, it is called the Schwarzschild radius. Observe that when $r = 2m$, the metric is singular; namely, $g_{tt} = 0$ and $g_{rr} = \infty$. By transforming the metric [2] to the so-called Kruskal coordinates (cf. Adler *et al.* (1975)), one observes that the Schwarzschild sphere $r = 2m$ has the physical characteristics of a black hole: light and nearby particles can enter the region $r < 2m$, nothing can exit this region, and there is an intrinsic (nonremovable) singularity at the center $r = 0$.

For the general metric [2], we define a black hole solution of Einstein's equations to be a solution which satisfies, for some $\rho > 0$,

$$A(\rho) = 0, \quad A(r) > 0 \text{ if } r > \rho$$

$\rho$ is called the radius of the black hole, or the event horizon.

### Yang–Mills Equations

The YM equations generalize Maxwell's equations. To see how this comes about, we first write Maxwell's equations in an invariant way. Thus, let $A$ denote a scalar-valued 1-form:

$$A = A_i\mathrm{d}x^i, \quad A_i \in \mathbb{R}$$

which is called the electromagnetic potential (by physicists), or a connection (by geometers). The electromagnetic field (curvature) is the 2-form

$$F = \mathrm{d}A$$

In local coordinates,

$$F = F_{ij}\mathrm{d}x^i \wedge \mathrm{d}x^j, \quad F_{ij} = \frac{\partial A_j}{\partial x^i} - \frac{\partial A_i}{\partial x^j}$$

In this framework, Maxwell's equations are given by

$$\mathrm{d}^\star F = 0, \qquad \mathrm{d}F = 0 \qquad [4]$$

where $\star$ is the Hodge star operator, mapping 2-forms to 2-forms (in $\mathbb{R}^4$), and is defined by

$$({}^\star F)_{k\ell} = \tfrac{1}{2}\sqrt{|g|}\varepsilon_{ijk\ell}F^{ij}$$

where $g = \det(g_{ij})$ and $\varepsilon_{ijk\ell}$ is the completely antisymmetric symbol defined by $\varepsilon_{ijk\ell} = \mathrm{sgn}(ijk\ell)$. As usual, indices are raised (or lowered) via the metric, so that, for example,

$$F^{ij} = g^{\ell i}g^{mj}F_{jm}$$

It is important to notice that ${}^\star F$ depends on the metric. Note also that Maxwell's equations are linear equations for the $A_i$'s.

The YM equations generalize Maxwell's equations and can be described as follows. With each YM field (described below) is associated a compact Lie group $G$ called the gauge group. For such $G$, we denote its Lie algebra by $\mathfrak{g}$, defined to be the tangent space at the identity of $G$. Now let $A$ be a $\mathfrak{g}$-valued 1-form

$$A = A_i\mathrm{d}x^i$$

where each $A_i$ is in $\mathfrak{g}$. In this case, the curvature 2-form is defined by

$$F = \mathrm{d}A + A \wedge A$$

or, in local coordinates,

$$F_{ij} = \frac{\partial A_j}{\partial x^i} - \frac{\partial A_i}{\partial x^j} + [A_i, A_j]$$

The commutator $[A_i, A_j] = 0$ if $G$ is an abelian group, but is generally nonzero if $G$ is a matrix group. In this framework, the YM equations can be written in the form $d^\star F = 0$, where now d is an appropriately defined covariant exterior derivative. For Maxwell's equations, the gauge group $G = U(1)$ (the circle group $\{e^{i\theta}: \theta \in \mathbb{R}\}$) so $\mathfrak{g}$ is abelian and we recover Maxwell's equations from the YM equations. Observe that if $G$ is nonabelian, then the YM equations $d^\star F = 0$ are nonlinear equations for the connection coefficients $A_i$.

### The Dirac Equation in Curved Spacetime

The Dirac equation is a generalization of Schrödinger's equation, in a relativistic setting (Bjorken and Drell 1964). It thus combines quantum mechanics with the theory of relativity. In addition, the Dirac equation also describes the intrinsic "spin" of fermions and, for this reason, solutions of the Dirac equation are often called spinors.

The Dirac equation can be written as

$$(G - m)\Psi = 0 \qquad [5]$$

where $G$ is the Dirac operator, $m$ is the mass of the Dirac particle (fermion), and $\Psi$ is a complex-valued 4-vector called the wave function, or spinor. The Dirac operator $G$ is of the form

$$G = iG^j(x)\frac{\partial}{\partial x^j} + B(x) \qquad [6]$$

where $G^j$ as well as $B$ are $4 \times 4$ matrices, $m$ is the (rest) mass of the fermion, and $i = \sqrt{-1}$. The Dirac equation is thus a linear equation for the spinors. The $G^j$ (called Dirac matrices) and the Lorentzian metric $g_{ij}$ are related by

$$g^{jk} I = \tfrac{1}{2}\{G^j, G^k\} \qquad [7]$$

where $\{G^j, G^k\}$ is the anticommutator

$$\{G^j, G^k\} = G^j G^k + G^k G^j$$

Thus, the Dirac matrices depend on the underlying metric in four-dimensional spacetime.

Suppose that $H$ is a spacelike hypersurface in $\mathbb{R}^4$, with future-directed normal vector $\nu = \nu(x)$, and let $d\mu$ be the invariant measure on $H$ induced by the metric $g_{ij}$. We define a scalar product on solutions $\Psi, \Phi$ of the Dirac equation by

$$\langle \Psi | \Phi \rangle = \int_H \bar{\Psi} G^j \Phi \nu_j \, d\mu \qquad [8]$$

This scalar product is positive definite, and because of current conservation (cf. Finster (1988))

$$\nabla_j \bar{\Psi} G^j \Phi = 0$$

it is also independent of $H$. By generalizing the expression (due to Dirac), $\bar{\Psi}\gamma^0\Psi = |\Psi|^2$, in Minkowski space, where $\gamma^0$ and $\bar{\Psi}$, the adjoint spinors, are defined by

$$\gamma^0 = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & -\mathbf{1} \end{pmatrix}, \qquad \bar{\Psi} = \Psi^* \gamma^0$$

where $*$ denotes complex conjugation, and $\mathbf{1}$ is the $2 \times 2$ identity matrix, the quantity $\bar{\Psi} G^j \Psi \nu_j$ is interpreted as the probability density of the Dirac particle. We normalize solutions of the Dirac equation by requiring

$$\langle \Psi | \Psi \rangle = 1 \qquad [9]$$

## Spherically Symmetric EDYM Equations

In the remainder of this article we assume that all fields are spherically symmetric, so they depend only on the variable $r = |x|$. In this case, the Lorentzian metric in polar coordinates $(t, r, \theta, \varphi)$ takes the form [2]. The Dirac wave function can be (Finster *et al.* 2000b) described by two real functions, $(\alpha(r), \beta(r))$, and the potential $W(r)$ corresponds to the magnetic component of an SU(2) YM field. As shown in Finster *et al.* (2000b), the EDYM equations are

$$\sqrt{A}\alpha' = \frac{w}{r}\alpha - (m + \omega T)\beta \qquad [10]$$

$$\sqrt{A}\beta' = (-m + \omega T)\alpha - \frac{w}{r}\beta \qquad [11]$$

$$rA' = 1 - A - \frac{1}{e^2}\frac{(1 - w^2)^2}{r^2} - 2\omega T^2(\alpha^2 + \beta^2) - \frac{2}{e^2}Aw'^2 \qquad [12]$$

$$2rA'\frac{T'}{T} = -1 + A + \frac{1}{e^2}\frac{(1 - w^2)^2}{r^2} + 2mT(\alpha^2 - \beta^2) - 2\omega T^2(\alpha^2 + \beta^2) + 4\frac{T}{r}w\alpha\beta - \frac{2}{e^2}Aw'^2 \qquad [13]$$

$$rAw'' = -(1 - w^2)w + e^2 rT\alpha\beta - r^2\frac{A'T - 2AT'}{2T}w' \qquad [14]$$

Equations [10] and [11] are the Dirac equations, [12] and [13] are the Einstein equations, and [14] is

the YM equation. The constants $m, \omega$, and $e$ denote, respectively, the rest mass of the Dirac particle, its energy, and the YM coupling constant.

### Nonexistence of Black Hole Solutions

Let the surface $r = \rho > 0$ represent a black hole event horizon:

$$A(\rho) = 0, \quad A(r) > 0 \text{ if } r > \rho \qquad [15]$$

In this case, the normalization condition [9] is replaced by

$$\int_{r_0}^{\infty} (\alpha^2 + \beta^2) \frac{\sqrt{T}}{A}\, dr < \infty, \quad \text{for every } r_0 > \rho \qquad [16]$$

In addition, we assume that the following global conditions hold:

$$\lim_{r\to\infty} r\big(1 - A(r)\big) = M < \infty \qquad [17]$$

(finite mass),

$$\lim_{r\to\infty} T(r) = 1 \qquad [18]$$

(gravitational field is asymptotically flat Minkowskian), and

$$\lim_{r\to\infty} \big(w(r)^2, w'(r)\big) = (1, 0) \qquad [19]$$

(the YM field is well behaved).

Concerning the event horizon $r = \rho$, we make the following regularity assumptions:

1. The volume element $\sqrt{|\det g_{ij}|} = |\sin\theta| r^2 A^{-1} T^{-2}$ is smooth and nonzero on the horizon; that is,

$$T^{-2}A^{-1},\ T^2 A \in C^1([\rho, \infty))$$

2. The strength of the YM field $F_{ij}$ is given by

$$\operatorname{tr}(F_{ij}F^{ij}) = \frac{2Aw'^2}{r^4} + \frac{(1 - w^2)^2}{r^4}$$

(cf. Bartnik and McKinnon 1988). We assume that this scalar is bounded near the horizon; that is, outside the event horizon and near $r = \rho$, assume that

$$w \text{ and } Aw'^2 \text{ are bounded} \qquad [20]$$

3. The function $A(r)$ is monotone increasing outside of and near the event horizon.

As discussed in Finster *et al.* (1999a), if assumption 1 or 2 were violated, then an observer freely falling into the black hole would feel strong forces when crossing the horizon. Assumption 3 is considerably weaker than the corresponding assumption in Finster *et al.* (1999b), where, indeed, it was assumed that the function $A(r)$ obeyed a power law $A(r) = c(r-\rho)^s + \mathcal{O}((r-\rho)^{s+1})$, with positive constants $c$ and $s$, for $r > \rho$.

The main result in this subsection is the following theorem:

**Theorem 1** *Every black hole solution of the EDYM equations [10]–[14] satisfying the regularity conditions 1–3 cannot be normalized and coincides with a Bartnik–McKinnon (BM) black hole of the corresponding Einstein–Yang–Mills (EYM) equations; that is, the spinors $\alpha$ and $\beta$ must vanish identically outside the event horizon.*

**Remark** Smoller and Wasserman (1998) proved that any black hole solution of the EYM equations that has finite mass (i.e., that satisfies [17]) must be one of the BM black hole solutions (Bartnik and McKinnon 1988) whose existence was first demonstrated in Smoller *et al.* (1993). Thus, amending the EYM equations by taking quantum-mechanical effects into account – in the sense that both the gravitational and YM fields can interact with Dirac particles – does not yield any new types of black hole solutions.

The present strategy in proving this theorem is to assume that we have a black hole solution of the EDYM equations [10]–[18] satisfying assumptions 1–3, where the spinors do not vanish identically outside of the black hole. We shall show that this leads to a contradiction. The proof is broken up into two cases: either $A^{-1/2}$ is integrable or nonintegrable near the event horizon. We shall only discuss the proof for the case when $A^{-1/2}$ is integrable near the event horizon, leaving the alternate case for the reader to view in Finster *et al.* (2000a).

If $A^{-1/2}$ is integrable, then one shows that there are positive constants $c, \varepsilon$ such that

$$c \le \alpha^2(r) + \beta^2(r) \le \frac{1}{c}, \quad \text{if} \quad \rho < r < \rho + \varepsilon \qquad [21]$$

Indeed, multiplying [10] by $\alpha$, and [11] by $\beta$ and adding gives an estimate of the form

$$\sqrt{A}(\alpha^2 + \beta^2)' \le \kappa(\alpha^2 + \beta^2)$$

Upon dividing by $\sqrt{A}(\alpha^2 + \beta^2)$ and integrating from $r > \rho$ to $\rho + \varepsilon$ gives

$$|\log(\alpha^2 + \beta^2)(\rho + \varepsilon) - \log(\alpha^2 + \beta^2)(r)| \le \text{const.}$$

from which the desired result follows. Next, from [12] and [13],

$$r(AT^2)' = 4 - \omega T^4(\alpha^2 + \beta^2) + T^3\left[2m(\alpha^2 - \beta^2) + \frac{4w}{r}\alpha\beta\right] - \frac{4}{e^2}(Aw')^2T^2 \qquad [22]$$

Using assumption 2 together with the last theorem, we see that the coefficients of $T^4, T^3$, and $T^2$ on the right-hand side of [21] are bounded near $\rho$, and from assumption 1 the left-hand side of [21] is bounded near $\rho$. Since assumption 1 implies $T(r) \to \infty$ as $r \searrow \rho$, we see that $\omega = 0$. Since $\omega = 0$, the Dirac equations simplify and we can show that $\alpha\beta$ is a positive decreasing function which tends to 0 as $r \to \infty$. Then the YM equation can be written in the form

$$r^2(Aw')' = -w(1-w)^2 + e^2\frac{r(T\sqrt{A})\alpha\beta}{\sqrt{A}} + r^2\frac{(AT^2)'}{2AT^2}(Aw') \qquad [23]$$

From assumption 2, $Aw'^2$ is bounded so $A^2w'^2 \to 0$ as $r \searrow \rho$. Thus, from [22] we can write, for $r$ near $\rho$,

$$(Aw')'(r) \geq c_1 + \frac{c_2}{\sqrt{A(r)}}$$

where $c_1$ and $c_2$ are positive constants. Using this inequality, we can show that for $r$ near $\rho$,

$$A(r) = (r - \rho)B(r)$$

where $0 < \lim_{r \searrow \rho} B(r) < \infty$. It follows that $A(\rho) = 0$ and $A'(\rho) > 0$. Thus, the Einstein metric has the same qualitative features as the Schwarzschild metric near the event horizon. Hence, the metric singularity can be removed via a Kruskal transformation (Adler *et al.* 1975). In these Kruskal coordinates, the YM potential is continuous and bounded (as is easily verified). As a consequence, the arguments in Finster *et al.* (2000c) go through and show that the spinors must vanish identically outside the horizon. For this, one must note that continuous zero-order terms in the Dirac operator are irrelevant for the derivation of the matching conditions in Finster *et al.* 2000c, section 2.4). Thus, the matching conditions (equations (2.31), (2.34) of Finster *et al.* (2000c)) are valid without changes in the presence of our YM field. Using conservation of the (electromagnetic) Dirac current and its positivity in timelike directions, the arguments in Finster *et al.* (2000c, section 4) all carry over. This completes the proof.

We have thus proved that the only black hole solutions of our EDYM equations are the BM black holes; that is, the spinors must vanish identically. In other words, the EDYM equations do not admit normalizable black hole solutions. Thus, in the presence of quantum-mechanical Dirac particles, static and spherically symmetric black hole solutions do not exist. Another interpretation of these our result is that Dirac particles can only either disappear into the black hole or escape to infinity. These results were proved under very weak regularity assumptions on the form of the event horizon (see assumptions 1–3).

## Particle-Like Solutions

By a particle-like (bound state) solution of the (SU(2)) EDYM equations, we mean a smooth solution of eqns [10]–[14], which is defined for all $r \geq 0$, and satisfies condition [9], which explicitly becomes

$$\int_0^r (\alpha^2 + \beta^2)\frac{\sqrt{T}}{A}\,dr = 1 \qquad [24]$$

In addition, we demand that [17]–[19] also hold. It is easily shown that, near $r = 0$, we must have

$$w(r) = 1 - \frac{\lambda}{2}r^2 + \mathcal{O}(r^2) \qquad [25]$$

where $\lambda$ is a real parameter. From this, via a Taylor expansion, one finds that

$$\begin{aligned} \alpha(r) &= \alpha_1 r + \mathcal{O}(r^3) \\ \beta(r) &= \tfrac{1}{2}(\omega T_0 - m)\alpha_1 r^2 + \mathcal{O}(r^3) \end{aligned} \qquad [26]$$

$$A(r) = 1 + \mathcal{O}(r^2), \quad T(r) = T_0 + \mathcal{O}(r^2) \qquad [27]$$

with two parameters $\alpha_1$ and $T_0 > 0$. Using linearity of the Dirac equation, we can always assume that $\alpha_1 > 0$.

Under all realistic conditions, the coupling of Dirac particles to the YM field (describing the weak or strong interactions) is much stronger than the coupling to the gravitational field. Thus, we are particularly intrested in the case of weak gravitational coupling. As shown in Finster *et al.* (2000b), the gravitational field is essential for the formation of bound states. However, for arbitrarily weak gravitational coupling, we can hope to find bound states. It is even conceivable that these bound-state solutions might have a well-defined limit when the gravitational coupling tends to zero, if we let the YM coupling go to infinity at the same time. Our idea is that this limiting case might yield a system of equations which is simpler than the full EDYM system, and can thus serve as a physically interesting starting point for the analysis of the coupled

interactions described by the EDYM equations. Expressed in dimensionless quantities, we shall thus consider the limits

$$m^2\kappa \to 0 \quad \text{and} \quad e^2 \to \infty \qquad [28]$$

That is, we ask whether weak gravitational coupling can give rise to bound states. Using numerical methods, we find particle-like solutions which are stable, even for arbitrarily weak gravitational coupling.

Now assuming that [27] holds (weak gravitational coupling), so that $(A, T) \approx (1, 1)$, then we find that the Dirac equations have a meaningful limit only under the assumptions that $\alpha$ converges and that

$$m\beta(r) \to \hat{\beta}(r), \quad m^2(T(r) - 1) \to \varphi \qquad [29]$$
$$m(\omega - m) \to E$$

with two real functions $\hat{\beta}, \varphi$ and a real parameter $E$. Multiplying [29] with $m$ and taking the limits [28] as well as $A, T \to 1$, the Dirac equations become

$$\alpha' = \frac{w}{r}\alpha - 2\hat{\beta} \qquad [30]$$

$$\hat{\beta}' = (E + \varphi)\alpha - \frac{w}{r}\hat{\beta} \qquad [31]$$

We next consider the YM equation [14]. The last term in [14] drops out in the limit of weak gravitational coupling [27]. The second summand converges only under the assumption that

$$\frac{e^2}{m} \to q \qquad [32]$$

with $q$ a real parameter, playing the role of an "effective" coupling constant. Together with [27], this implies that $m \to \infty$. The YM equations thus have the limit

$$r^2 w'' = -(1 - w)^2 w + qr\alpha\hat{\beta} \qquad [33]$$

In order to get a well-defined and nontrivial limit of the Einstein equations [13] and [14], we need to assume that the parameter $m^3\kappa$ has a finite, nonzero limit. Since this parameter has the dimension of inverse length, we can arrange by a scaling of our coordinates that

$$m^3\kappa \to 1 \qquad [34]$$

We differentiate the $T$-equation [13] with respect to $r$ and substitute [12]. Taking the limits [28] and [33], a straightforward calculation yields the equation

$$r^2\Delta\varphi = -\alpha^2 \qquad [35]$$

where $\Delta = r^{-2}\partial_r(r^2\partial_r)$ is the radial Laplacian in Euclidean $\mathbb{R}^3$. Indeed, this equation can be regarded as Newton's equation with the Newtonian potential $\varphi$. Thus, the limiting case [34] for the gravitational field corresponds to taking the Newtonian limit. Finally, the normalization condition [16] reduces to

$$\int_0^\infty \alpha(r)^2 \, dr = 1 \qquad [36]$$

The boundary conditions [17]–[19], [24]–[26] are transformed into

$$w(r) = 1 - \frac{\lambda}{2}r^2 + \mathcal{O}(r^3), \quad \lim_{r\to\infty} w(r) = \pm 1 \qquad [37]$$

$$\alpha(r) = \alpha_1 r + \mathcal{O}(r^3), \quad \hat{\beta}(r) = \mathcal{O}(r^3) \qquad [38]$$

$$\varphi(r) = \varphi_0 + \mathcal{O}(r^3), \quad \lim_{r\to\infty} \varphi(r) < \infty \qquad [39]$$

with the three parameters $\lambda, \alpha_1$, and $\varphi_0$. We point out that the limiting system contains only one coupling constant $q$. According to [31] and [33], $q$ is in dimensionless form given by

$$e^2 m^2 \kappa \to q \qquad [40]$$

Hence, in dimensionless quantities, the limit [17] describes the situation where the gravitational coupling goes to zero, while the YM coupling constant goes to infinity like $e^2 \sim (m^2\kappa)^{-1}$. Therefore, this limiting case is called the reciprocal coupling limit (RCL). The reciprocal coupling system is given by eqns [29], [30], [32], and [34] together with the normalization conditions [35] and the boundary conditions [36]–[38]. According to [28], the parameter $E$ coincides up to a scaling factor with $\omega - m$, and thus has the interpretation as the (properly scaled) energy of the Dirac particle. As in Newtonian mechanics, the potential $\varphi$ is determined only up to a constant $\mu \in \mathbb{R}$; namely, the reciprocal limit equations are invariant under the transformation

$$\varphi \to \varphi + \mu, \qquad E \to E - \mu \qquad [41]$$

To simplify the connection between the EDYM equations, and the RCL equations, we introduce a parameter $\varepsilon$ in such a way that as $\varepsilon \to 0$, EDYM $\to$ RCL; namely,

$$\varepsilon = \frac{m^2\kappa}{e^2}$$

Notice that $\varepsilon$ describes the relative strength of gravity versus the YM interaction. For realistic physical situations, the gravitational coupling is weak; namely, $m^2\kappa \ll 1$, but the YM coupling constant is of order $1 : e^2 \sim 1$. So we investigate the parameter range $\varepsilon \ll 1, q \sim 0$. These form the starting points for the numeric below.

We seek stable bound states for weak gravitational coupling. For this purpose, we consider the total binding energy

$$B = M - m \qquad [42]$$

where $M$ is the ADM mass defined by [17] and $m$ is the rest mass of the Dirac particle. $B$ is thus the amount of energy set free when the binding is broken. If $B < 0$, then energy is needed to break up the binding. According to Lee (1987), a solution is stable if $B < 0$. In order to find solutions of the RCL equations with $B < 0$, Lee's treatment and a new two-parameter shooting method (Finster *et al.* 2000b) can be used. Stable solutions of these RCL equations then follow (see Finster *et al.* (2000b) for details).

We now turn to the full EDYM equations. Here are the key steps of our method:

1. Find solutions which are small perturbations of the limiting (RCL) solutions.
2. Trace these solutions by gradually changing the coupling constants.
3. This should yield a one-parameter family of solutions which are "far" from the known limiting solutions.

The point is that we use the RCL solutions as a starting point for numerics, and we "continue" these solutions to solutions of the full EDYM equations.

To be somewhat more specific, we see that if we fix $\varepsilon$ and $q$, we have two parameters:

$$\alpha_1 = \alpha'(0) \quad \text{and} \quad E = \omega - m$$

and two conditions at $\infty$:

$$\alpha^2 + \beta^2 \to 0, \qquad w^2 \to 1$$

We consider the EDYM equations with weaker side conditions

$$0 < \lambda^2 \equiv \int_0^\infty (\alpha^2 + \beta^2)\frac{\sqrt{T}}{A}\,\mathrm{d}r < \infty$$
$$0 < \tau = \lim_{r\to\infty} T(r) < \infty$$
$$\lim_{r\to\infty} w^2(r) = 1$$
$$\rho = \lim_{r\to\infty} r(1 - A(r)) < \infty$$

Then we rescale these solutions to obtain the true side conditions via the transformations

$$\tilde{\alpha}(r) = \sqrt{\tau}\lambda^{-2}\alpha(\lambda^{-2}r)$$
$$\tilde{\beta}(r) = \sqrt{\tau}\lambda^{-2}\beta(\lambda^{-2}r)$$
$$\tilde{A}(r) = A(\lambda^{-2}r), \quad \tilde{T}(r) = \tau^{-1}T(\lambda^{-2}r)$$
$$\tilde{m} = \lambda^{-2}m, \quad \tilde{\omega} = \tau\lambda^{-2}\omega$$
$$\tilde{\kappa} = \lambda^6\kappa, \quad \tilde{e}^2 = \lambda^2 e^2$$

## Discussion

In this article we have considered the SU(2) EDYM equations. Our first result shows that the only black hole solutions of these equations are the BM black holes; that is, the spinors must vanish identically outside of the black hole. In other words, the EDYM equations do not admit normalizable black hole solutions. Thus, as mentioned earlier, this result indicates that the Dirac particle either enters the black hole or escapes to infinity. Two recent publications (Finster *et al.* 2002a,b) we consider the Cauchy problem for a massive Dirac equation in a charged, rotating-black-hole geometry (the non-extreme Kerr–Newman black hole), with compactly supported initial data outside the black hole. We prove that, in this case, the probability that the Dirac particle lies in any compact set tends to zero as $t \to \infty$. This means that the Dirac particle indeed either enters the black hole or tends to infinity. We also show that the wave function decays at a rate $t^{-5/6}$ on any compact set outside of the event horizon.

For particle-like solutions of the SU(2) EDYM equations, we find stable bound states for arbitrarily weak gravitational coupling. This shows that as weak as the gravitational interaction is, it has a regularizing effect on the equations. The stability of particle-like solutions of the EDYM equations is in sharp contrast to the EYM equations, where the particle-like solutions are all unstable (Straumann and Zhou 1990).

## Acknowledgment

This work was partially supported by the NSF, Contract Number DMS-010-3998.

*See also:* Abelian and Nonabelian Gauge Theories using Differential Forms; Black Hole Mechanics; Bosons and Fermions in External Fields; Dirac Operator and Dirac Field; Einstein Equations: Exact Solutions; Einstein Equations: Initial Value Formulation; Noncommutative Geometry and the Standard Model; Relativistic Wave Equations Including Higher Spin Fields; Symmetry Classes in Random Matrix Theory.

## Further Reading

Adler R, Bazin M, and Schiffer M (1975) *Introduction to General Relativity*, 2nd edn. New York: McGraw-Hill.

Bartnik R and McKinnon J (1988) Particle-like solutions of the Einstein–Yang–Mills equations. *Physical Review Letters* 61: 141–144.

Bjorken J and Drell S (1964) *Relativistic Quantum Mechanics*. New York: McGraw-Hill.

Finster F (1988) Local U(2, 2) symmetry in relativistic quantum mechanics. *Journal of Mathematical Physics* 39: 6276–6290.

Finster F, Smoller J, and Yau S-T (1999a) Particle-like solutions of the Einstein–Dirac equations. *Physical Review D* 59: 104020/1–104020/19.

Finster F, Smoller J, and Yau S-T (2000a) The interaction of Dirac particles with non-abelian gauge fields and gravitation-black holes. *Michigan Mathematical Journal* 47: 199–208.
Finster F, Smoller J, and Yau S-T (2000b) The interaction of Dirac particles with non-abelian gauge fields and gravitation-bound states. *Nuclear Physics B* 584: 387–414.
Finster F, Smoller J, and Yau S-T (1999b) Nonexistence of black hole solutions for a spherically symmetric, static Einstein–Dirac–Maxwell system. *Communications in Mathematical Physics* 205: 249–262.
Finster F, Smoller J, and Yau S-T (2000c) Nonexistence of time-periodic solutions of the Dirac equation in a Reissner–Nordström black hole background. *Journal of Mathematical Physics* 41: 2173–2194.
Finster F, Kamran N, Smoller J, and Yau S-T (2002a) The long time dynamics of Dirac particles in the Kerr–Newman black hole geometry. *Advances in Theoretical and Mathematical Physics* 7: 25–52.
Finster F, Kamran N, Smoller J, and Yau S-T (2002b) Decay rates and probability estimates Dirac particles in the Kerr–Newman black hole geometry. *Communications in Mathematical Physics* 230: 201–244.
Lee TD (1987) Mini-soliton stars. *Physical Review D* 25: 3640–3657.
Smoller J, Wasserman A, and Yau S-T (1993) Existence of infinitely-many smooth global solutions of the Einstein–Yang/Mills equations. *Communications in Mathematical Physics* 151: 303–325.
Smoller J and Wasserman A (1998) Extendability of solutions of the Einstein–Yang/Mills equations. *Communications in Mathematical Physics* 194: 707–733.
Straumann N and Zhou Z (1990) Instability of the Bartnik–McKinnon solution of the Einstein–Yang–Mills equations. *Physics Letters B* 237: 353–356.

# Dirac Operator and Dirac Field

**S N M Ruijsenaars**, Centre for Mathematics and Computer Science, Amsterdam, The Netherlands



## Introduction

The Dirac equation arose in the early days of quantum mechanics, inspired by the problem of taking special relativity into account in the quantum mechanical description of a freely moving electron. From the outset, however, Dirac looked for an equation that also accomodated the electron spin and that could be modified to include interaction with an external electromagnetic field. The equation he discovered satisfies all of these requirements. On the other hand, when it is rewritten in Hamiltonian form, the spectrum of the resulting Dirac operator includes not only the desired interval $[mc^2, \infty)$ (where $m$ is the electron mass and $c$ the speed of light), but also an interval $(-\infty, -mc^2]$.

Dirac himself already considered this negative part of the spectrum as unphysical, since no such negative energies had been observed and their presence would entail instability of the electron. This physical flaw of the "first-quantized" description of a relativistic electron led to the introduction of "second quantization," as encoded in quantum field theory. In the field-theoretic version of the Dirac theory, the unphysical negative energies are obviated by a prescription that originated in Dirac's hole theory.

Specifically, Dirac postulated that the negative-energy states of his equation were occupied by a sea of unobservable particles, the Pauli principle forbidding an occupancy greater than one. In this heuristic picture, the annihilation of a negative-energy electron yields a hole in the sea, observable as a new type of positive-energy particle with the same mass, but opposite charge. This led Dirac to predict that the electron should have an oppositely charged partner.

His prediction was soon confirmed experimentally, the partner of the negatively charged electron showing up as the positively charged positron. More generally, all electrically charged particles (not only spin-1/2 particles described by the Dirac equation) have turned out to have oppositely charged antiparticles. Furthermore, some electrically neutral particles also have distinct antiparticles.

Returning to the second-quantized Dirac theory, this involves a Dirac quantum field in which the creation/annihilation operators of negative-energy states are replaced by annihilation/creation operators of positive-energy holes, resp. The hole theory substitution therefore leads to a Hilbert space (called Fock space) that accomodates an arbitrary number of particles and antiparticles with the same mass and opposite charge.

Soon after the introduction of the Dirac equation (which dates from 1928), it turned out that the number of particles and antiparticles is not conserved in a high-energy collision. Such creation and annihilation processes admit a natural description in the Fock spaces associated with relativistic quantum field theories. The very comprehensive mathematical description of real-world elementary particle phenomena that is now called the standard model arose some 30 years ago, and has been abundantly confirmed by experiment ever since. It involves

various relativistic quantum fields with nonlinear interactions. The Dirac quantum field is an essential ingredient, inasmuch as it is used to describe all spin-1/2 particles and antiparticles in the model (including quarks, electrons, neutrinos etc.).

After this survey (which is not only very brief, but also biased toward the physical concepts at issue), the contents of this article will be sketched. The free Dirac equation associated with the physical Minkowski spacetime $\mathbb{R}^4$ is first detailed. The exposition and notation are slightly unconventional in some respects. This is because we are partly preparing the ground for a mathematically precise account of the second-quantized version of the free Dirac theory. For example, momentum space (as opposed to position space) is emphasized, since the variable $\boldsymbol{x}$ in the Dirac equation does not have a clear physical significance and should be discarded in the Hilbert space formulation of the second-quantized Dirac field. The latter acts on a Fock space of multi-particle and -antiparticle wave functions depending on momentum and spin variables, and the spacetime dependence of the Dirac field is solely a consequence of relativistic covariance. (In particular, the variable $\boldsymbol{x}$ in the Dirac field $\Psi(t, \boldsymbol{x})$ should not be viewed as the position of particles and antiparticles created and annihilated by the field.)

To be sure, there is much more to the Dirac theory than its free first- and second-quantized versions for Minkowski spacetime $\mathbb{R}^4$. The primary purpose here is, however, to present these foundational versions in some detail. A much more sketchy account of further developments can be found in subsequent sections. First, the one-particle theory is reconsidered. Generalizations of the free theory to arbitrary dimensions and Euclidean settings are sketched and interactions with external fields are described, touching on various aspects and applications.

The next focus is on relations with index theory that arise when the massless Euclidean Dirac operator is generalized to geometric settings, namely $l$-dimensional Riemannian manifolds allowing a spin structure. We illustrate the general Atiyah–Singer index theory for the Dirac framework with some simple examples for $l=1$ (Toeplitz operators) and $l=2$ (the manifold $S^1 \times S^1$).

More information on the many-particle Dirac theory appears in the final section. Brief remarks on the Dirac field in interaction with other quantized fields are followed by an elaboration of the far simpler situation of the Dirac field interacting with external fields. Among the $S$-operators corresponding to such fields there is a special class of unitary matrix multipliers; the external field then vanishes for $t < 0$ and equals the pure gauge field corresponding to the unitary matrix for $t \geq 0$. Specializing to an even spacetime dimension and choosing special "kink" type unitaries, the associated Fock-space quadratic forms can be made to converge to the free Dirac field.

As mentioned already, Dirac's second quantization procedure was invented to get rid of the unphysical negative energies of the first-quantized (one-particle) theory. It is an amazing fact that the resulting formalism for the simplest case (namely the massless Dirac operator in a two-dimensional spacetime) can be exploited for quite different purposes. In particular, this setting can be tied in with various soliton equations and the representation theory of certain infinite-dimensional groups and Lie algebras. In conclusion, some of these applications are briefly sketched, namely the construction of special solutions to the Kadomtsev–Petviashvili (KP) equation (including the KP solitons and finite-gap solutions) and special representations of Kac–Moody and Virasoro algebras.

## The Free One-Particle Dirac Equation in $\mathbb{R}^4$

The free time-dependent Dirac equation is a linear hyperbolic evolution equation for a function $\Psi(t, \boldsymbol{x})$ on spacetime $\mathbb{R}^4$ with values in $\mathbb{C}^4$. It involves four $4 \times 4$ matrices $\gamma^\mu$, $\mu = 0, 1, 2, 3$, satisfying the $\gamma$-algebra

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}\mathbf{1}_4, \quad g = \mathrm{diag}(1, -1, -1, -1) \quad [1]$$

Using the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}$$
$$\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad [2]$$

one can choose for example

$$\gamma^0 = \begin{pmatrix} 0 & \mathbf{1}_2 \\ \mathbf{1}_2 & 0 \end{pmatrix}, \quad \gamma^k = \begin{pmatrix} 0 & -\sigma_k \\ \sigma_k & 0 \end{pmatrix}$$
$$k = 1, 2, 3 \quad [3]$$

Now the free Dirac equation reads

$$(\mathrm{i}\hbar\gamma^0\partial_t + \mathrm{i}\hbar c\gamma \cdot \nabla - mc^2\mathbf{1}_4)\Psi(t, \boldsymbol{x}) = 0 \quad [4]$$

where $\hbar$ is Planck's constant, $c$ the speed of light, and $m$ the particle mass. Using from now on units so that $\hbar = c = 1$, this can be abbreviated as

$$\left(\mathrm{i}\sum_{\mu=0}^{3}\gamma^{\mu}\partial_{\mu} - m\right)\Psi(x) = 0, \quad x = (x^0, x^1, x^2, x^3)$$
$$\partial_\mu = \partial/\partial x^\mu, \quad \mu = 0, 1, 2, 3 \qquad [5]$$

The relativistic invariance of this equation can be understood as follows. First, since the equation does not explicitly involve the spacetime coordinates, it is invariant under spacetime translations. (If $\Psi(t, \boldsymbol{x})$ solves [5], then also $\Psi(t - a_0, \boldsymbol{x} - \boldsymbol{a})$ is a solution for all $(a_0, \boldsymbol{a}) \in \mathbb{R}^4$.) Second, it is invariant under Lorentz transformations (rotations and boosts). Indeed, if $\Psi(x)$ is a solution and $L \in \mathrm{SO}(1,3)$, then $S(L)\Psi(L^{-1}x)$ solves [5] too, where $S(L)$ denotes a (suitably normalized) matrix satisfying

$$S(L)^{-1}\gamma^{\mu}S(L) = \sum_{\nu=0}^{3} L^{\mu}{}_{\nu}\gamma^{\nu} \qquad [6]$$

(The matrices $\gamma^{\mu\prime}$ on the right-hand side of [6] satisfy the $\gamma$-algebra [1]. From this, the existence of a representation $S(L)$ of SO(1, 3) satisfying [6] is readily deduced.)

As a consequence, the Poincaré group (inhomogeneous Lorentz group) acts in a natural way on the space of solutions to the time-dependent Dirac equation, expressing its independence of the choice of inertial frame. For quantum mechanical purposes, however, one needs to choose a frame and use the associated time variable to rewrite the equation as a Hilbert space evolution equation.

The relevant Hilbert space $\check{\mathcal{H}}$ is the space of four-component functions that are square integrable over space,

$$\check{\mathcal{H}} = L^2(\mathbb{R}^3, \mathrm{d}\boldsymbol{x}) \otimes \mathbb{C}^4 \qquad [7]$$

To obtain a self-adjoint Hamiltonian on $\check{\mathcal{H}}$, one multiplies [5] by $\gamma^0$ and introduces the Hermitian matrices

$$\beta = \gamma^0, \qquad \alpha^k = \gamma^0\gamma^k, \quad k = 1, 2, 3 \qquad [8]$$

Then, one obtains the Schrödinger type equation

$$\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}\psi = \check{H}\psi \qquad [9]$$

where $\check{H}$ is the Dirac operator,

$$\check{H} = -\mathrm{i}\alpha\cdot\nabla + \beta m \qquad [10]$$

Under Fourier transformation,

$$\mathcal{F} : \check{\mathcal{H}} \to L^2(\mathbb{R}^3, \mathrm{d}\boldsymbol{p}) \otimes \mathbb{C}^4$$
$$\psi(\boldsymbol{x}) \mapsto \phi(\boldsymbol{p}) = (2\pi)^{-3/2}\int_{\mathbb{R}^3} \mathrm{d}\boldsymbol{x}\ \exp(-\mathrm{i}\boldsymbol{x}\cdot\boldsymbol{p})\psi(\boldsymbol{x}) \qquad [11]$$

eqn [9] turns into

$$\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}\phi = D(\boldsymbol{p})\phi, \quad D(\boldsymbol{p}) = \boldsymbol{\alpha}\cdot\boldsymbol{p} + \beta m \qquad [12]$$

The matrix $D(\boldsymbol{p})$ is Hermitian and has square $E_p^2\mathbf{1}_4$, where $E_p$ is the relativistic energy,

$$E_p = (\boldsymbol{p}\cdot\boldsymbol{p} + m^2)^{1/2} \qquad [13]$$

corresponding to a momentum $\boldsymbol{p}$. Now, we have

$$U_{\mathrm{C}}\overline{D(-\boldsymbol{p})} = -D(\boldsymbol{p})U_{\mathrm{C}} \qquad [14]$$

where $U_{\mathrm{C}}$ is the charge conjugation matrix,

$$U_{\mathrm{C}} = \mathrm{i}\gamma^2 \qquad [15]$$

Hence, the four eigenvalues of $D(\boldsymbol{p})$ are given by $E_p$, $E_p$, $-E_p$, and $-E_p$. Therefore, the matrices

$$P_{\pm}(\boldsymbol{p}) = \frac{1}{2}\left(\mathbf{1}_4 \pm \frac{D(\boldsymbol{p})}{E_p}\right) \qquad [16]$$

are projections on the positive and negative spectral subspaces of $D(\boldsymbol{p})$.

As orthonormal base for the positive-energy subspace, we can now choose

$$w_{+,j}(\boldsymbol{p}) = \left(\frac{2E_p}{E_p + m}\right)^{1/2} P_+(\boldsymbol{p})b_j, \quad j = 1, 2 \qquad [17]$$

where

$$b_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\1\\0\end{pmatrix}, \quad b_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\1\\0\\1\end{pmatrix} \qquad [18]$$

Next, setting

$$w_{-,j}(\boldsymbol{p}) = U_{\mathrm{C}}\overline{w_{+,j}}(\boldsymbol{p}), \quad j = 1, 2 \qquad [19]$$

an orthonormal base $w_{-,1}(-\boldsymbol{p})$, $w_{-,2}(-\boldsymbol{p})$ for the negative-energy subspace of $D(\boldsymbol{p})$ is obtained; cf. [14].

The upshot is that the time-independent Dirac equation

$$\check{H}\psi = E\psi \qquad [20]$$

gives rise to bounded eigenfunctions

$$\begin{aligned} e_{+,j}(\boldsymbol{x},\boldsymbol{p}) &= (2\pi)^{-3/2}\exp(\mathrm{i}\boldsymbol{x}\cdot\boldsymbol{p})w_{+,j}(\boldsymbol{p}), \quad j=1,2 \\ e_{-,j}(\boldsymbol{x},\boldsymbol{p}) &= (2\pi)^{-3/2}\exp(-\mathrm{i}\boldsymbol{x}\cdot\boldsymbol{p})w_{-,j}(\boldsymbol{p}), \quad j=1,2 \end{aligned} \qquad [21]$$

with eigenvalues $E=E_p$ and $E=-E_p$, resp. Clearly, they are not square-integrable, but they can be used as the kernel of a unitary transformation between $\check{\mathcal{H}}$ (7) and the Hilbert space

$$\begin{aligned} \mathcal{H} &= \mathcal{H}_+ \oplus \mathcal{H}_- = P_+\mathcal{H} \oplus P_-\mathcal{H} \\ \mathcal{H}_+, \mathcal{H}_- &= L^2(\mathbb{R}^3, \mathrm{d}\boldsymbol{p}) \otimes \mathbb{C}^2 \end{aligned} \qquad [22]$$

Specifically, we have

$$\begin{aligned} &W : \mathcal{H} \to \check{\mathcal{H}} \\ &f(\boldsymbol{p}) = (f_+(\boldsymbol{p}), f_-(\boldsymbol{p})) \\ &\mapsto \psi(\boldsymbol{x}) = \sum_{\delta=+,-}\sum_{j=1,2}\int_{\mathbb{R}^3}\mathrm{d}\boldsymbol{p}\, e_{\delta,j}(\boldsymbol{x},\boldsymbol{p})f_{\delta,j}(\boldsymbol{p}) \end{aligned} \qquad [23]$$

which entails

$$(W^{-1}\psi)_{\delta,j}(\boldsymbol{p}) = \int_{\mathbb{R}^3}\mathrm{d}\boldsymbol{x}\,\overline{e_{\delta,j}}(\boldsymbol{x},\boldsymbol{p})\cdot\psi(\boldsymbol{x}) \qquad [24]$$

(Here and throughout this article, a bar denotes complex conjugation.)

From the above, it is clear that the Dirac Hamiltonian $\check{H}$ acting on the Hilbert space $\check{\mathcal{H}}$ is unitarily equivalent to the multiplication operator on $\mathcal{H}$ [22] given by

$$(Hf)_\delta(\boldsymbol{p}) = \delta E_p f_\delta(\boldsymbol{p}), \quad \delta=+,- \qquad [25]$$

Indeed, $W$ is a diagonalizing transformation for $\check{H}$, the relation

$$H = W^{-1}\check{H}W \qquad [26]$$

yielding an explicit realization of the spectral theorem.

Using the same notational convention, the momentum, charge conjugation, parity, and time-reversal operators on $\check{\mathcal{H}}$, given by

$$(\check{P}_k\psi)(\boldsymbol{x}) = -\mathrm{i}\partial_k\psi(\boldsymbol{x}) \qquad [27]$$

$$(\check{C}\psi)(\boldsymbol{x}) = U_C\bar{\psi}(\boldsymbol{x}) \qquad [28]$$

$$(\check{P}\psi)(\boldsymbol{x}) = U_P\psi(-\boldsymbol{x}), \quad U_P = \gamma^0 \qquad [29]$$

$$(\check{T}\psi)(\boldsymbol{x}) = U_T\bar{\psi}(\boldsymbol{x}), \quad U_T = \gamma^1\gamma^3 \qquad [30]$$

transform into the operators

$$(P_kf)_\delta(\boldsymbol{p}) = \delta p_k f_\delta(\boldsymbol{p}), \quad \delta=+,- \qquad [31]$$

$$(Cf)_\delta(\boldsymbol{p}) = \bar{f}_{-\delta}(\boldsymbol{p}), \quad \delta=+,- \qquad [32]$$

$$(Pf)_\delta(\boldsymbol{p}) = \delta f_\delta(-\boldsymbol{p}), \quad \delta=+,- \qquad [33]$$

$$(Tf)_\delta(\boldsymbol{p}) = \mathrm{i}\sigma_2\bar{f}_\delta(-\boldsymbol{p}), \quad \delta=+,- \qquad [34]$$

Note that $P_k$, $P$, and $T$ leave the positive- and negative-energy subspaces $\mathcal{H}_+$ and $\mathcal{H}_-$ invariant, whereas $C$ interchanges them.

To conclude this section, we describe some salient features of the unitary representation of the (identity component of the) Poincaré group on $\mathcal{H}$, which follows from the representation on solutions to [5] already sketched. The spacetime translations over $a \in \mathbb{R}^4$ are represented by the unitary operator $\exp(-\mathrm{i}a_0H + \mathrm{i}\boldsymbol{a}\cdot\boldsymbol{P})$; explicitly,

$$\begin{aligned} &(\exp(-\mathrm{i}a_0H + \mathrm{i}\boldsymbol{a}\cdot\boldsymbol{P})f)_\delta(\boldsymbol{p}) \\ &\quad = \exp(-\mathrm{i}\delta(a_0E_p - \boldsymbol{a}\cdot\boldsymbol{p}))f_\delta(\boldsymbol{p}), \quad \delta=+,- \end{aligned} \qquad [35]$$

The representation of the Lorentz group involves unitary $2\times2$ matrices $U(k, A)$, where $k$ is an arbitrary 4-vector satisfying $k^\mu k_\mu = 1$ and $A$ the matrix in SL(2, $\mathbb{C}$) representing $L \in$ SO(1, 3). (Recall that SL(2, $\mathbb{C}$) can be viewed as a 2-fold cover of SO(1, 3).) In particular, $U(k, A)$ does not depend on $k$ for rotations,

$$U(k, A) = A^*, \quad \forall A \in \mathrm{SU}(2) \qquad [36]$$

(Here and henceforth, we use $*$ to denote the Hermitian adjoint of matrices and operators.) For boosts, however, there is dependence on the vector $k$, which is the image of the vector (1, **0**) under the boost. We refrain from a more detailed description of $U(k, A)$, as this would carry us too far afield.

The unitary SO(1, 3) representation leaves the decomposition $\mathcal{H} = P_+\mathcal{H} \oplus P_-\mathcal{H}$ invariant. On the positive-energy subspace $\mathcal{H}_+$, it is given by

$$(U(L)f)_+(\boldsymbol{p}) = \left(\frac{p_0^L}{p_0}\right)^{1/2} U\left(\frac{p}{m}, A\right)^* f_+(\boldsymbol{p}^L) \qquad [37]$$

where

$$p = (E_p, \boldsymbol{p}), \quad p^L = L^{-1}p \qquad [38]$$

On $\mathcal{H}_-$, it is given by the complex-conjugate representation,

$$(U(L)f)_-(\boldsymbol{p}) = \left(\frac{p_0^L}{p_0}\right)^{1/2} U\left(\frac{p}{m}, A\right)^{\mathrm{t}} f_-(\boldsymbol{p}^L) \qquad [39]$$

just as for the spacetime translations, cf. [35]. (The superscript t is used to denote the transpose matrix.) This feature is crucial for the second-quantized Dirac theory, which is discussed next.

## The Free Dirac Field in $\mathbb{R}^4$

The free Dirac field is an operator-valued distribution on a Fock space that describes an arbitrary number of spin-1/2 particles and antiparticles in terms of momentum space wave functions. Since spin-1/2 particles are fermions (which encodes the Pauli exclusion principle), an $M$-particle wave function $F^{+,\dots,+}_{j_1,\dots,j_M}(\boldsymbol{p}_1,\dots,\boldsymbol{p}_M)$ (where $j_l \in \{1, 2\}$ is the spin index) is antisymmetric under any interchange of a pair $(j_i, \boldsymbol{p}_i)$ and $(j_k, \boldsymbol{p}_k)$. Likewise, $N$-antiparticle wave functions $F^{-,\dots,-}_{k_1,\dots,k_N}(\boldsymbol{q}_1,\dots,\boldsymbol{q}_N)$ are antisymmetric. But a wave function $F^{+-}_{j,k}(\boldsymbol{p},\boldsymbol{q})$ describing a particle–antiparticle pair need not have any symmetry property, since a particle and an antiparticle can be distinguished by their charge.

The relevant Fock space is therefore the tensor product of two antisymmetric Fock spaces built over the one-particle and one-antiparticle spaces $L^2(\mathbb{R}^3, \mathrm{d}\boldsymbol{p}) \otimes \mathbb{C}^2$. For later purposes, it is important to view these spaces as the summands $\mathcal{H}_+$ and $\mathcal{H}_-$ of the space $\mathcal{H}$ from the previous section. Thus, the arena for the free Dirac field is the Hilbert space

$$\mathcal{F}_a(\mathcal{H}) \simeq \mathcal{F}_a(\mathcal{H}_+) \otimes \mathcal{F}_a(\mathcal{H}_-) \qquad [40]$$

where, for example,

$$\mathcal{F}_a(\mathcal{H}) = (\mathbb{C} \oplus \mathcal{H} \oplus (\mathcal{H} \otimes \mathcal{H})_a \oplus \cdots)^- \qquad [41]$$

where the bar denotes the completion of the infinite direct sum in the obvious inner product. The tensor $(1, 0, 0, \dots)$ is viewed as the vacuum (the "filled Dirac sea") and denoted by $\Omega$.

To get around in Fock space, one employs the creation and annihilation operators $c^{(*)}(f), f \in \mathcal{H}$. The creation operator $c^*(f), f \in \mathcal{H}$, is defined by linear and continuous extension of its action on the vacuum $\Omega$ and on elementary antisymmetric tensors, recursively given by

$$\begin{aligned} &c^*(f)\Omega = f, \quad c^*(f)f_1 = f \wedge f_1, \dots \\ &\quad c^*(f)f_1 \wedge \cdots \wedge f_N = f \wedge f_1 \wedge \cdots \wedge f_N, \dots \end{aligned} \qquad [42]$$

Its adjoint, the annihilation operator $c(f)$, satisfies

$$\begin{aligned} &c(f)\Omega = 0, \quad c(f)f_1 = (f, f_1)\Omega, \dots \\ &c(f)f_1 \wedge \cdots \wedge f_N = \sum_{j=1}^{N} (-)^{j-1}(f, f_j) \\ &\qquad \times f_1 \wedge \cdots \wedge \widehat{f_j} \wedge \cdots \wedge f_N, \dots \end{aligned} \qquad [43]$$

Accordingly, the operators $c^{(*)}(f)$ satisfy the canonical anticommutation relations (CARs) over $\mathcal{H}$,

$$\begin{aligned} &\{c(f), c(g)\} = 0, \\ &\quad \{c(f), c^*(g)\} = (f, g), \quad \forall f, g \in \mathcal{H} \end{aligned} \qquad [44]$$

where $\{A, B\}$ denotes the anticommutator $AB + BA$. (From this, one readily deduces that $c^{(*)}(f)$ is bounded with norm $\|f\|$.)

Next, recalling the direct sum decomposition [22], a notation change

$$c^{(*)}(P_+f) \to a^{(*)}(P_+f), \quad c^{(*)}(P_-f) \to b^{(*)}(P_-f) \qquad [45]$$

is made, thus indicating that $a^{(*)}$ and $b^{(*)}$ should be viewed as the creation/annihilation operators of particles and antiparticles, resp. Since $\mathcal{H}_+$ and $\mathcal{H}_-$ are copies of $L^2(\mathbb{R}^3, \mathrm{d}\boldsymbol{p}) \otimes \mathbb{C}^2$, a given function $(f_1(\boldsymbol{p}), f_2(\boldsymbol{p}))$ in the latter space can occur both as an argument of $a^{(*)}(\cdot)$ and of $b^{(*)}(\cdot)$; it can also be viewed as a smearing function for unsmeared quantities $a_j^{(*)}(\boldsymbol{p})$ and $b_j^{(*)}(\boldsymbol{p})$, $j = 1, 2$, that are often referred to as operators as well (even though they are only quadratic forms). Thus, one has, for example,

$$\begin{aligned} b^*(f) &= \sum_{j=1}^{2} \int_{\mathbb{R}^3} \mathrm{d}\boldsymbol{p}\, b_j^*(\boldsymbol{p}) f_j(\boldsymbol{p}) \\ b(f) &= \sum_{j=1}^{2} \int_{\mathbb{R}^3} \mathrm{d}\boldsymbol{p}\, b_j(\boldsymbol{p}) \bar{f}_j(\boldsymbol{p}) \end{aligned} \qquad [46]$$

As explained shortly, the smeared time-zero Dirac field takes the form

$$\Phi(f) = a(P_+f) + b^*(KP_-f), \quad f \in \mathcal{H} \qquad [47]$$

Here and below, $K$ denotes complex conjugation on $\mathcal{H}$, $\mathcal{H}_+$, and $\mathcal{H}_-$. Just as the operators $c^{(*)}(f)$, the operators $\Phi^{(*)}(f)$ satisfy the CARs over $\mathcal{H}$,

$$\begin{aligned} &\{\Phi(f), \Phi(g)\} = 0 \\ &\{\Phi(f), \Phi^*(g)\} = (f, g), \quad \forall f, g \in \mathcal{H} \end{aligned} \qquad [48]$$

as is readily verified using [44]–[45]. But this $\Phi$-representation is not unitarily equivalent to the $c$-representation [44]. This becomes clear in particular from the consideration of a crucial type of CAR automorphism that is considered next.

To this end, we fix a unitary operator $U$ on $\mathcal{H}$. Then it is plain that the operators

$$\tilde{c}^{(*)}(f) = c^{(*)}(Uf) \qquad [49]$$

$$\tilde{\Phi}^{(*)}(f) = \Phi^{(*)}(Uf) \qquad [50]$$

also satisfy the CARs. The CAR-algebra automorphism $c^{(*)}(f) \mapsto \tilde{c}^{(*)}(f)$ can be unitarily implemented in $\mathcal{F}_a(\mathcal{H})$, since one has

$$\tilde{c}^{(*)}(f) = \Gamma(U)c^{(*)}(f)\Gamma(U^*) \qquad [51]$$

where $\Gamma(U)$ denotes the Fock-space product operator corresponding to $U$. Thus, for example,

$$\begin{aligned} &\Gamma(U)\Omega = \Omega, \quad \Gamma(U)f = Uf, \ldots \\ &\quad \Gamma(U)f_1 \wedge \cdots \wedge f_N = Uf_1 \wedge \cdots \wedge Uf_N, \ldots \end{aligned} \qquad [52]$$

For the CAR automorphism $\Phi^{(*)}(f) \mapsto \tilde{\Phi}^{(*)}(f)$ this is not true, however. Rewriting it in terms of the annihilation and creation operators $a^{(*)}$ and $b^{(*)}$ via [47], it amounts to a linear transformation (Bogoliubov transformation), whose unitary implementability has been clarified several decades ago. To be specific, the necessary and sufficient condition for unitary implementability is that the off-diagonal parts

$$U_{+-} = P_+ U P_-, \qquad U_{-+} = P_- U P_+ \qquad [53]$$

in the $2 \times 2$ matrix decomposition of operators on $\mathcal{H}$ be Hilbert–Schmidt operators. Therefore, no problem arises when $U$ is diagonal with respect to this decomposition. Indeed, in that case one can choose as unitary implementer the product operator

$$\tilde{\Gamma}(U) = \Gamma(U_{++}) \otimes \Gamma(KU_{--}K) \qquad [54]$$

(cf. the tensor product structure [40] of $\mathcal{F}_a(\mathcal{H})$).

In particular, the automorphism

$$\Phi(f) \mapsto \Phi(\mathrm{e}^{\mathrm{i}tH}f) \qquad [55]$$

where $H$ is the free diagonalized Dirac Hamiltonian [25], is implemented by the operator

$$\tilde{\Gamma}(\mathrm{e}^{\mathrm{i}tH}) = \Gamma(\mathrm{e}^{\mathrm{i}tE}) \otimes \Gamma(\mathrm{e}^{\mathrm{i}tE}) \qquad [56]$$

where $E$ denotes multiplication by $E_p$ on $\mathcal{H}_+$ and $\mathcal{H}_-$. The change of CAR representation, therefore, entails that the unphysical negative energies of the one-particle theory are replaced by positive energies of antiparticles. Hence, we obtain a mathematically precise version of Dirac's hole theory substitution $b_j(\boldsymbol{p}) \to b_j^*(\boldsymbol{p})$, $b_j^*(\boldsymbol{p}) \to b_j(\boldsymbol{p})$.

More generally, if one chooses for $U$ the Poincaré group representation (given by [35] and [37]–[39]), then the Fock-space implementer [54] is the tensor product of two product operators with the same action on $\mathcal{F}_a(L^2(\mathbb{R}^3, \mathrm{d}\boldsymbol{p}) \otimes \mathbb{C}^2)$. Observe that this is also true for the Fock-space version $\tilde{\Gamma}(T) = \Gamma(T)$ of the time-reversal operator [34]. By contrast, the Fock-space parity operator $\tilde{\Gamma}(P) = \Gamma(P)$ gives rise to two product operators with slightly different actions, cf. [33]. Accordingly, particles and antiparticles have opposite parity.

The map

$$\Phi(f) \mapsto \Phi(Cf)^* \qquad [57]$$

also yields a CAR automorphism. It is unitarily implemented by the Fock-space charge-conjugation operator

$$\mathcal{C} = \Gamma\left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right) \qquad [58]$$

which interchanges particles and antiparticles. Notice that $\mathcal{C}$ is unitary, whereas $C$ is antiunitary.

It remains to establish the precise relation of the above to the customary free Dirac field $\Psi(t, \boldsymbol{x})$. This is a quadratic form on $\mathcal{F}_a(\mathcal{H})$ given by

$$\begin{aligned} \Psi(t, \boldsymbol{x}) = (2\pi)^{-3/2} \int_{\mathbb{R}^3} \mathrm{d}\boldsymbol{p} \sum_{j=1,2} \Big( a_j(\boldsymbol{p}) w_{+,j}(\boldsymbol{p}) \mathrm{e}^{-\mathrm{i}E_p t + \mathrm{i}\boldsymbol{p}\cdot\boldsymbol{x}} \\ + b_j^*(\boldsymbol{p}) w_{-,j}(\boldsymbol{p}) \mathrm{e}^{\mathrm{i}E_p t - \mathrm{i}\boldsymbol{p}\cdot\boldsymbol{x}} \Big) \end{aligned} \qquad [59]$$

(Its expectation $\langle F_1, \Psi(t, \boldsymbol{x}) F_2 \rangle$ is, for example, well defined for $F_1$, $F_2$ in the dense subspace of $\mathcal{F}_a(\mathcal{H})$ that consists of vectors with finitely many particles and antiparticles and wave functions in Schwartz space.) It satisfies the time-dependent Dirac equation

$$\mathrm{i}\partial_t \Psi = (-\mathrm{i}\alpha \cdot \nabla + \beta m)\Psi \qquad [60]$$

in the sense of quadratic forms. Furthermore, smearing it with a function $\overline{\psi}(\boldsymbol{x})$ in the Hilbert space $\check{\mathcal{H}}$ (7), we obtain

$$\begin{aligned} &\int_{\mathbb{R}^3} \mathrm{d}\boldsymbol{x} \overline{\psi}(\boldsymbol{x}) \cdot \Psi(t, \boldsymbol{x}) \\ &\quad = \Phi(\mathrm{e}^{\mathrm{i}tH} W^{-1} \psi) \\ &\quad = \tilde{\Gamma}(\mathrm{e}^{\mathrm{i}tH}) \Phi(W^{-1}\psi) \tilde{\Gamma}(\mathrm{e}^{-\mathrm{i}tH}) \quad \psi \in \check{\mathcal{H}} \end{aligned} \qquad [61]$$

As announced, the time evolution of the free Dirac field is, therefore, given by the unitary one-parameter group [56], whose generator (the second-quantized Dirac Hamiltonian) has spectrum $\{0\} \cup [m, \infty)$.

The Dirac field $\Psi(t, \boldsymbol{x})$ can also be smeared with a test function $F(t, \boldsymbol{x})$ in the Schwartz space $S(\mathbb{R}^4)^4$, yielding a bounded operator

$$\Psi(F) = \int_{\mathbb{R}^4} \mathrm{d}x F(x) \cdot \Psi(x) \qquad [62]$$

Then one obtains the relativistic covariance relation

$$\tilde{\Gamma}(U(a, L)) \Psi(F) \tilde{\Gamma}(U(a, L))^* = \Psi(F^{a,L}) \qquad [63]$$

where

$$F^{a,L}(x) = S(L^{-1})^t F(L^{-1}(x - a)) \qquad [64]$$

and $U(a, L)$ denotes the Poincaré group representation on $\mathcal{H}$, cf. [35] and [37]–[39]. Likewise, one gets the inversion formulas

$$\tilde{\Gamma}(I)\Psi(F)\tilde{\Gamma}(I)^* = \Psi(F_I), \quad I = P, T \qquad [65]$$

with

$$F_P(t, \boldsymbol{x}) = U_P^t F(t, -\boldsymbol{x}), \quad F_T(t, \boldsymbol{x}) = U_T^t \bar{F}(-t, \boldsymbol{x}) \qquad [66]$$

while the Fock-space charge-conjugation operator [58] transforms the Dirac field as

$$\mathcal{C}\Psi(F)\mathcal{C} = \Psi(F_C)^* \qquad [67]$$

with

$$F_C(x) = U_C^t \bar{F}(x) \qquad [68]$$

Finally, let us consider the global U(1) gauge transformations $f \mapsto e^{i\phi} f$, where $\phi \in \mathbb{R}$ and $f \in \mathcal{H}$. They can be implemented by

$$\tilde{\Gamma}(e^{i\phi}) = \Gamma(e^{i\phi}) \otimes \Gamma(e^{-i\phi}) \qquad [69]$$

and one has

$$\tilde{\Gamma}(e^{i\phi})\Psi(F)\tilde{\Gamma}(e^{i\phi})^* = \Psi(F_\phi) \qquad [70]$$

with

$$F_\phi(x) = e^{-i\phi} F(x) \qquad [71]$$

The generator $Q$ of the one-parameter group $\phi \mapsto \tilde{\Gamma}(e^{i\phi})$ is the charge operator: on wave functions describing $N_+$ particles and $N_-$ antiparticles, it has eigenvalue $N_+ - N_-$.

## More on the One-Particle Dirac Theory

Even for the free one-particle setting, the account given earlier is far from complete. To begin with, the free Dirac equation admits a specialization to massless particles. In the Weyl representation of the $\gamma$-algebra adopted above, the choice $m = 0$ entails that the $\boldsymbol{p}$-space equation [12] decouples into two $2 \times 2$ equations for spinors that can be labeled by their chirality ("handedness"). This refers to their eigenvalue with respect to the chirality matrix

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} \mathbf{1}_2 & 0 \\ 0 & -\mathbf{1}_2 \end{pmatrix} \qquad [72]$$

and this notion derives from the noninvariance of the separate $2 \times 2$ equations under parity. (A positive-chirality spinor is mapped to a negative-chirality spinor under the parity operator $\check{P}$ (33) and vice versa.) Since the weak interaction breaks parity symmetry, the two $2 \times 2$ equations (often called Weyl equations) do have physical relevance. Indeed, the associated quantum fields are a crucial ingredient of the standard model.

Next, we point out that it is possible to switch to a representation in which the gamma matrices are real. This so-called Majorana representation is convenient (but not indispensable) in the description of neutral spin-1/2 particles. By definition, such particles are equal to their antiparticles, so that the second-quantized formalism of the previous section must be adapted: one needs the neutral CAR algebra over $\mathcal{H}$ (also known as self-dual CAR).

For various purposes, it is important to formulate the free Dirac equation for a spacetime whose spatial dimension is arbitrary. Then one needs, first of all, gamma matrices satisfying the (Minkowski) Clifford algebra relations

$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}\mathbf{1}_\Lambda, \quad g = \mathrm{diag}(1, -\mathbf{1}_n) \qquad [73]$$

where $n$ is the space dimension and the minimal size $\Lambda \times \Lambda$ of the gamma matrices is to be determined.

Clearly, for $n = 1$ and $n = 2$, one can take $\Lambda = 2$, choosing, for example,

$$\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \gamma^1 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$
$$\gamma^2 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \qquad [74]$$

to fulfill [73]. For $n = 4$, one can take $\Lambda = 4$, just as for $n = 3$, supplementing [1] with the matrix $i\gamma^5$, cf. [72].

More generally, for $n = 2N - 1$ and $n = 2N$, one can take $\Lambda = 2^N$ in [73]. Indeed, a representation on the $2^N$-dimensional fermion Fock space $\mathcal{F}_a(\mathbb{C}^N)$ (cf. [41]) is readily constructed using the creation and annihilation operators described in the previous section. Once this has been taken care of, most of the discussion on the free one-particle Dirac equation in $\mathbb{R}^4$ can be easily generalized. Of special importance in this regard is the straightforward adaptation of the formulas [7]–[26], which form the foundation for the second-quantized version. Indeed, the discussion of the last section applies nearly verbatim for arbitrary spacetime dimension.

In several applications, the so-called Euclidean version of the free Dirac theory in spacetime dimension $n + 1$ is important. Basically, this version is obtained upon replacing $i\partial_0$ by $\partial_{n+1}$ in the Dirac equation, a substitution that changes the character of the equation from hyperbolic to elliptic. Provided that the mass vanishes, the Euclidean Dirac equation admits a reinterpretation as a time-independent zero-eigenvalue Weyl equation in a Minkowski spacetime of dimension $n + 2$. (This equation is often called the zero-mode equation.)

Let us now turn to the description of the interaction with an external electromagnetic potential $A_\mu(t, \boldsymbol{x})$. This can be taken into account via the minimal substitution,

$$\partial_\mu \to \partial_\mu + \mathrm{i}eA_\mu \qquad [75]$$

also known as the covariant derivative, in the time-dependent Dirac equation [5].

For the electron in the Coulomb field of a nucleus of charge $Ze$, one has

$$A_k = 0,\ k = 1,2,3, \quad A_0 = -\frac{Ze}{4\pi|\boldsymbol{x}|} \qquad [76]$$

and the time-independent equation

$$\left(-\mathrm{i}\boldsymbol{\alpha}\cdot\nabla + \beta m - \frac{Ze^2}{4\pi|\boldsymbol{x}|}\right)\psi = E\psi \qquad [77]$$

can be solved explicitly. This leads to a bound-state spectrum that is more accurate than its nonrelativistic counterpart. In particular, one finds that energy levels that are degenerate in the nonrelativistic theory split up into slightly different levels. The resulting fine structure of the Dirac levels can be understood as a consequence of the coupling between the spin of the electron and its orbital motion.

In spite of this better agreement with the experimental levels, the physical interpretation of the Dirac electron in a Coulomb field is enigmatic. This is not only because of the persistence of the negative-energy states of the free theory (which turn into scattering states), but also because of unphysical properties of the position operator. More general time-independent external fields (such as step potentials $A_0(\boldsymbol{x})$ with a step height larger than $2m$) can cause transitions between positive- and negative-energy states (Klein paradox). This phenomenon is enhanced when time dependence is allowed. In particular, any external field that is given by functions in $C_0^\infty(\mathbb{R}^4)$ leads to a scattering operator $S$ on the one-particle space $\mathcal{H}$ [22] that has nonzero off-diagonal parts $S_{\pm\mp}$. Hence, a positive-energy wave packet scattering at such a time- and space-localized field has a nonzero probability to show up as a negative-energy wave packet.

When one tensors the one-particle space $\check{\mathcal{H}}$ with an internal symmetry space $\mathbb{C}^k$, one can also couple external Yang–Mills fields $A_\mu$ taking values in the $k \times k$ matrices via the substitution [75]. (From a geometric viewpoint, this can be rephrased as tensoring the spinor bundle with a vector bundle equipped with a connection $A$.) The generalization of this external gauge field coupling to a Minkowski spacetime or Euclidean space of arbitrary dimension is straightforward. An adaptation of the resulting interacting one-particle Dirac theory in arbitrary dimension to quite general geometric settings also yields a crucial starting point for index theory.

Before turning to the latter area, we conclude this section with another striking application of the one-particle framework, namely the massless Dirac equation in two spacetime dimensions with special external fields. Specifically, the relevant Dirac operator is of the form

$$\begin{pmatrix} \mathrm{i}\frac{\mathrm{d}}{\mathrm{d}x} & -\mathrm{i}q(x) \\ \mathrm{i}r(x) & -\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}x} \end{pmatrix} \qquad [78]$$

where $r(x)$ and $q(x)$ are not necessarily real valued. (Note that this operator is in general not self-adjoint.) With suitable restrictions on $r$ and $q$, the direct and inverse scattering theory associated with the Dirac operator [78] can be applied to various nonlinear PDEs in two spacetime dimensions to solve their Cauchy problems in considerable detail. As a crucial special case, initial conditions yielding vanishing reflection give rise to soliton solutions for the pertinent equation.

The first example in this framework was found by Zakharov and Shabat (the nonlinear Schrödinger equation); with other choices of $r$ and $q$ several other soliton PDEs (including the sine-Gordon and modified Korteweg–de Vries equations) were handled by Ablowitz, Kaup, Newell, and Segur, who studied a quite general class of external fields $r$ and $q$.

## The Dirac Operator and Index Theory

Thus far, we have considered various versions of the Dirac operator associated with the spaces $\mathbb{R}^l$ for some $l \geq 1$. For applications in the area of index theory, however, one needs to generalize this base manifold. Indeed, one can define a Dirac operator for any $l$-dimensional oriented Riemannian manifold $\mathcal{M}$ that admits a spin structure. This is a lifting of the transition functions of the tangent bundle $T\mathcal{M}$ (which may be assumed to take values in $\mathrm{SO}(l)$) to the simply connected twofold cover $\mathrm{Spin}(l)$ (taking $l \geq 3$).

Choosing first $l = 2N + 1$, the spin group has a faithful irreducible representation on $\mathbb{C}^{2^N}$. Hence, one obtains a $\mathbb{C}^{2^N}$-bundle over $\mathcal{M}$, the spinor bundle. The Levi-Civita connection on $\mathcal{M}$ derived from the metric can now be lifted to a connection on the spinor bundle. From the covariant derivative corresponding to the spin connection and the

Clifford algebra generators $\gamma^1, \ldots, \gamma^l$, one can then construct a first-order elliptic differential operator that acts on sections of the spinor bundle. (For the case $\mathcal{M} = \mathbb{R}^{2N+1}$ with its Euclidean metric, this construction yields the massless positive-chirality Dirac operator acting on wave functions with $2^N$ components, as considered above.)

The massless Dirac operator thus obtained is self-adjoint as an operator on the $L^2$-space $\mathcal{H}$ associated with the spinor bundle, and it has infinite-dimensional positive and negative spectral subspaces $\mathcal{H}_+$ and $\mathcal{H}_-$. (In this section the check accent on position-space quantities is omitted.) Specializing to the case of compact $\mathcal{M}$, a continuous map from $\mathcal{M}$ to $\mathbb{C}^*$ gives rise to a Fredholm operator on $\mathcal{H}_+$, and more generally a continuous map from $\mathcal{M}$ to $\mathrm{GL}(k, \mathbb{C})$ yields a Fredholm operator on $\mathcal{H}_+ \otimes \mathbb{C}^k$.

For a smooth map, the Fredholm index of this operator can be written in terms of an integral over $\mathcal{M}$ involving certain closed differential forms. The value of this integral does not change when exact forms are added, since $\mathcal{M}$ has no boundary. Hence, one is dealing with de Rham cohomology classes. In this context, the class involved ("characteristic class") is determined by the Riemann curvature tensor of $\mathcal{M}$ and the topological ("winding") characteristics of the map.

The simplest example of this state of affairs arises for $l = 1$ and $\mathcal{M} = S^1$ with its obvious spin structure (periodic boundary conditions). Writing $\psi \in \mathcal{H} = L^2(S^1)$ as

$$\psi(z) = \sum_{n \in \mathbb{Z}} a_n z^n, \quad z \in S^1 \qquad [79]$$

the Dirac operator $H$ on $\mathcal{H}$ reads

$$H = z \frac{\mathrm{d}}{\mathrm{d}z} \qquad [80]$$

It has eigenfunctions $z^n, n \in \mathbb{Z}$. Thus, we may choose

$$(P_+\psi)(z) = \sum_{n \geq 0} a_n z^n, \quad (P_-\psi)(z) = \sum_{n < 0} a_n z^n \qquad [81]$$

As a consequence, the functions in $\mathcal{H}_+ (\mathcal{H}_-)$ are $L^2$-boundary values of holomorphic functions in $|z| < 1$ ($|z| > 1$). Operators of the form

$$T_\psi = P_+ M_\psi P_+ \qquad [82]$$

where $\psi$ is a continuous function on $S^1$ and $M_\psi$ denotes multiplication by $\psi$, are called Toeplitz operators. It is not hard to see that they are Fredholm (viewed as operators on $\mathcal{H}_+$), provided that $\psi$ does not vanish on $S^1$. (Recall a bounded operator $B$ is Fredholm if it has finite-dimensional kernel $K$ and cokernel $C$. Its Fredholm index is given by

$$\mathrm{index}(B) = \dim K - \dim C \qquad [83]$$

and is norm continuous and invariant under addition of a compact operator.) Assuming $\psi(S^1) \subset \mathbb{C}^*$ from now on, the curve $\psi(S^1)$ has a well-defined winding number $w(\psi)$ with respect to the origin. The equality

$$\mathrm{index}(T_\psi) = -w(\psi) \qquad [84]$$

between objects from the area of analysis on the left-hand side and from the areas of topology and geometry on the right-hand side is the simplest example of an Atiyah–Singer type index formula. When $\psi$ is not only continuous but also smooth, the index formula can be rewritten as

$$\mathrm{index}(T_\psi) = -\frac{1}{2\pi \mathrm{i}} \int_{S^1} \frac{\mathrm{d}\psi}{\psi} \qquad [85]$$

yielding a characteristic class version.

It should be noted that the operator $M_\psi$ on $\mathcal{H}$ has a bounded inverse $M_{1/\psi}$ when $0 \notin \psi(S^1)$, hence a trivially vanishing index. Therefore, the compression [82] involving the spectral projection of the Dirac operator is needed to get a nonzero index. Observe also that the equality [84] is quite easily verified for the case $\psi(z) = z^n$, since $T_\psi$ yields a power of the right ($n > 0$) or left ($n < 0$) shift on $P_+\mathcal{H} \simeq l^2(\mathbb{N})$.

We proceed to the case of even-dimensional manifolds, $l = 2N$. Then the fiber $\mathbb{C}^{2^N}$ of the spinor bundle splits into a direct sum of even and odd spinors, corresponding to two distinct representations of $\mathrm{Spin}(2N)$ on $\mathbb{C}^{2^{N-1}}$. (Here it is assumed that $N > 3$; recall the Lie algebra isomorphisms $\mathrm{so}(4) \simeq \mathrm{so}(3) \oplus \mathrm{so}(3)$ and $\mathrm{so}(6) \simeq \mathrm{su}(4)$.) With respect to this decomposition, the Dirac operator can be written as

$$H = \begin{pmatrix} 0 & \mathcal{D}^* \\ \mathcal{D} & 0 \end{pmatrix} \qquad [86]$$

where $\mathcal{D}$ and $\mathcal{D}^*$ are again first-order elliptic differential operators expressed in terms of Clifford algebra generators and the spin connection. Tensoring the spinor bundle with a vector bundle equipped with a connection $A$, one can define a Dirac operator on the tensor product which involves $A$ and takes the form

$$H_A = \begin{pmatrix} 0 & \mathcal{D}_A^* \\ \mathcal{D}_A & 0 \end{pmatrix} \qquad [87]$$

with respect to the even/odd spinor decomposition. Once more, the index of $\mathcal{D}_A$ (viewed as a Fredholm operator between two different Hilbert spaces) can be expressed as an integral over $\mathcal{M}$ involving characteristic classes that depend on the curvatures of the two connections.

Probably the simplest example of the constructions just sketched is given by the torus $\mathcal{M} = S^1 \times S^1$ with its flat metric. Employing the above coordinate and spin structure on $S^1$, one can take

$$\mathcal{H} = L^2(S^1 \times S^1) \otimes \mathbb{C}^2, \quad \mathcal{D} = z_1 \frac{\partial}{\partial z_1} + \mathrm{i} z_2 \frac{\partial}{\partial z_2} \qquad [88]$$

Since the curvature vanishes, the index theorem for this situation implies $\mathrm{index}(\mathcal{D}) = 0$. (Note that this is also plain from [88]: both kernel and cokernel of $\mathcal{D}$ are spanned by the constant sections.) On the other hand, when one tensors the spinor bundle with a line bundle with connection $A$, the index formula reads

$$\mathrm{index}(\mathcal{D}_A) = -\frac{1}{2\pi} \int_{S^1 \times S^1} F \qquad [89]$$

where $F$ is the curvature 2-form corresponding to $A$.

The Atiyah–Singer index theorem for Dirac operators has far-reaching applications. It can be used to derive other results in this area, such as the Gauss–Bonnet–Chern theorem, the Hirzebruch signature theorem, and (when $\mathcal{M}$ is a Kähler manifold) Riemann–Roch type theorems. From this, one can obtain information on various questions, such as the existence of positive scalar curvature metrics or zeros of vector fields on $\mathcal{M}$. Other applications include insights on topological invariants of manifolds obtained from "simple" manifolds (such as spheres and tori) by glueing or covering operations. This hinges on the additive properties of the index that are clear from its being given by an integral over the manifold. Conversely, the integrality of Fredholm indices can be used to deduce that certain rational cohomology classes are actually integral on manifolds that admit the structure that is required for the pertinent index theorem to apply, that certain manifolds do not admit such structures, since one knows that the relevant class is not integral, etc.

## More on the Dirac Field

As mentioned earlier, the free-field formalism can be easily generalized to an arbitrary spacetime dimension $d$. For $d > 4$, however, no renormalizable interacting quantum field models involving the Dirac field are known. For the physical case $d = 4$ the standard model involves various Dirac fields interacting with quantized gauge fields and Klein–Gordon fields. Although its perturbation theory is renormalizable, its mathematical existence is to date wide open.

It is far beyond the scope of this article to elaborate on the analytical difficulties of relativistic quantum field theories, let alone those associated with the standard model. Even for $d = 2$ and 3, a nonperturbative construction of interacting quantum field models involving the Dirac field is an extremely difficult enterprise. Apart from some rigorous results on certain self-interacting Dirac field models, the only interacting model that is reasonably well understood from the constructive field theory viewpoint is the Yukawa model for $d = 2$ and 3. This describes the interaction between the Dirac field $\Psi$ and a Klein–Gordon field $\phi$, the interaction term being formally given by $g(\Psi^* \gamma^0 \Psi)\phi$.

On the other hand, the interaction of the quantized Dirac field with external classical fields is much more easily understood and analytically controlled. As a bonus, within this context, one can make contact with various issues of physical and mathematical relevance. We now proceed to sketch the external-field framework and some of its applications.

Let us first consider the addition of an external field term $gV(t, \boldsymbol{x})$ to the free Dirac operator $\check{H}$ on

$$\check{\mathcal{H}} = L^2(\mathbb{R}^n, \mathrm{d}\boldsymbol{x}) \otimes \mathbb{C}^\Lambda \otimes \mathbb{C}^k \qquad [90]$$

We assume from now on that the coupling $g$ is real and that $V$ is a self-adjoint $k\Lambda \times k\Lambda$ matrix-valued function on spacetime $\mathbb{R}^{n+1}$ with matrix elements that are in $C_0^\infty(\mathbb{R}^{n+1})$. Then the (interaction picture) scattering operator $S$ exists. It is unitary and has off-diagonal Hilbert–Schmidt parts $S_{\pm\mp}$, so that a unitary Fock-space $S$-operator $\tilde{\Gamma}(S)$ implementing the Bogoliubov transformation generated by $S$ exists:

$$\tilde{\Gamma}(S)\Phi(f)\tilde{\Gamma}(S)^* = \Phi(Sf), \quad \forall f \in \mathcal{H} \qquad [91]$$

The arbitrary phase in $\tilde{\Gamma}(S)$ can be fixed by requiring that the vacuum expectation value of $\tilde{\Gamma}(S)$ be positive. More precisely, this number is generically nonzero and satisfies

$$|(\Omega, \tilde{\Gamma}(S)\Omega)| = \det(1 + T_S)^{-1/2} \qquad [92]$$

where $T_S$ is a positive trace class operator determined by $S$.

The vector $\tilde{\Gamma}(S)\Omega$ is a superposition of wave functions with an equal and arbitrary number of particles and antiparticles. More generally, the Fock-space $S$-operator $\tilde{\Gamma}(S)$ leaves the subspaces of $\mathcal{F}_a(\mathcal{H})$ with a fixed eigenvalue $q \in \mathbb{Z}$ of the charge operator $Q$ invariant, and can create and

annihilate an arbitrary number of particle–antiparticle pairs.

The unitary propagator $U(T_1, T_2)$ corresponding to $V(t, \boldsymbol{x})$ does not have Hilbert–Schmidt off-diagonal parts (unless the spacetime dimension is sufficiently small and special external fields are chosen). Even so, the diagonal parts are Fredholm with vanishing index, and the off-diagonal parts are compact. Omitting the ill-defined determinantal factor, these properties imply that one obtains a renormalized quadratic form $\tilde{\Gamma}_{\text{rcn}}(U(T_1, T_2))$ satisfying the implementing relation

$$\tilde{\Gamma}_{\text{rcn}}(U(T_1, T_2))\Phi(f) = \Phi(U(T_1, T_2)f)\tilde{\Gamma}_{\text{rcn}}(U(T_1, T_2)), \quad \forall f \in \mathcal{H} \qquad [93]$$

in the quadratic form sense.

The above unitary operators on $\mathcal{H}$ yield Fredholm diagonal parts whose indices vanish. (They are norm continuous in $g$ and reduce to the identity for $g = 0$.) This is why their Fock-space implementers leave the charge sectors invariant. Indeed, for a unitary operator $U$ on $\mathcal{H}$ with compact off-diagonal parts the implementer maps the charge-$q$ sector to the charge-$(q + q(U))$ sector, where

$$q(U) = \text{index}(U_{--}) \qquad [94]$$

Specializing to the case

$$n = 2N - 1, \quad \Lambda = 2\lambda, \quad \lambda = 2^{N-1} \qquad [95]$$

a unitary $(k\Lambda \times k\Lambda)$-matrix multiplier $\check{U}$ on $\check{\mathcal{H}}$ does not have compact off-diagonal parts in general. But when it is of the form

$$\check{U} = \begin{pmatrix} \mathbf{1}_\lambda \otimes u_+(\boldsymbol{x}) & 0 \\ 0 & \mathbf{1}_\lambda \otimes u_-(\boldsymbol{x}) \end{pmatrix} \qquad [96]$$

with respect to the chiral decomposition (the generalization of the $\gamma^5$-decomposition [72] to even spacetime dimension), then it suffices for compactness of the off-diagonal parts that the matrices $u_\pm(\boldsymbol{x}) \in \text{U}(k)$ are continuous and converge to $\mathbf{1}_k$ for $|\boldsymbol{x}| \to \infty$.

Viewing $\mathbb{R}^{2N-1}$ as arising from $S^{2N-1}$ via stereographic projection, the latter unitaries can be viewed as continuous maps from $S^{2N-1}$ to $\text{U}(k)$, reducing to $\mathbf{1}_k$ at the north pole. As such, they yield elements of the homotopy group $\pi_{2N-1}(\text{U}(k))$. By virtue of Bott's periodicity theorem, the latter group equals $\mathbb{Z}$ for $k \geq N$. Thus, the maps $u_\pm$ have a well-defined "winding number" $w(u_\pm) \in \mathbb{Z}$ for $k \geq N$. From the index formula

$$\text{index}(U_{--}) = w(u_+) - w(u_-) \qquad [97]$$

and [94] one now deduces that one can obtain implementers $\tilde{\Gamma}_{\text{rcn}}(U)$ effecting a nonzero charge change from unitary maps with nonzero winding number.

In particular, choosing $k = \lambda = 2^{N-1} \geq N$, there exist quite special "kink maps"

$$u_{\epsilon,\boldsymbol{a}}(\boldsymbol{x}) \in \text{U}(\lambda), \quad \epsilon > 0, \ \boldsymbol{a} \in \mathbb{R}^{2N-1} \qquad [98]$$

with winding number 1 and such that the quadratic form implementers of the unitary multiplication operators

$$\check{U}_{+,\epsilon,\boldsymbol{a}} = -\begin{pmatrix} \mathbf{1}_\lambda \otimes u_{\epsilon,\boldsymbol{a}}(\boldsymbol{x}) & 0 \\ 0 & \mathbf{1}_\lambda \otimes \mathbf{1}_\lambda \end{pmatrix}$$
$$\check{U}_{-,\epsilon,\boldsymbol{a}} = -\begin{pmatrix} \mathbf{1}_\lambda \otimes \mathbf{1}_\lambda & 0 \\ 0 & \mathbf{1}_\lambda \otimes u_{\epsilon,\mathbf{a}}(-\boldsymbol{x}) \end{pmatrix} \qquad [99]$$

converge to (a linear combination of the chiral components of) the free Dirac field $\Psi(0, \boldsymbol{a})$ as the kink size parameter $\epsilon$ goes to 0.

For the special case $N = 1$, one can take

$$u_{\epsilon,a}(x) = \frac{x - a - \text{i}\epsilon}{x - a + \text{i}\epsilon} \qquad [100]$$

and the off-diagonal parts of $U_{\pm,\epsilon,a}$ are actually Hilbert–Schmidt. Thus, the implementers can be chosen to be unitary operators. But to get convergence to the Dirac field components $\Psi(0, a)_\pm$ as $\epsilon \to 0$, the unitary implementers $\tilde{\Gamma}(U_{\pm,\epsilon,a})$ should be renormalized by a multiplicative factor.

For the $N = 1$ case, the unitary multipliers [96] give rise to loop groups. Indeed, requiring

$$\lim_{x \to \pm\infty} u_\delta(x) = \mathbf{1}_k, \quad \delta = +, - \qquad [101]$$

we are dealing with continuous maps $S^1 \to \text{U}(k)$. From the viewpoint of the Dirac theory, these groups are local gauge groups. The convergence to the Dirac field just sketched can be used to great advantage to clarify the structure of the corresponding Fock-space gauge groups. Their Lie algebras yield representations of Kac–Moody algebras, a topic which is considered shortly.

Before doing so, it should be pointed out that under some mild smoothness assumptions all of the above unitary matrix multipliers can also be viewed as $S$-operators associated with very special external fields. Indeed, the gauge-transformed Dirac operator

$$\check{H}_U = \check{U}^* \check{H} \check{U} \qquad [102]$$

is of the form

$$\check{H}_U = \check{H} + V(\boldsymbol{x}) \qquad [103]$$

where $V(\boldsymbol{x})$ is a self-adjoint $k\Lambda \times k\Lambda$ matrix on $\mathbb{R}^{2N-1}$ (a "pure gauge" field). If one now defines a time-dependent external field by

$$V(t, \boldsymbol{x}) = \begin{cases} V(\boldsymbol{x}), & t \geq 0 \\ 0, & t < 0 \end{cases} \qquad [104]$$

then $\check{U}$ equals the $S$-operator for $V(t, \boldsymbol{x})$. (Equivalently, $\check{U}$ is the $t \to \infty$ wave operator for the time-independent external field $V(\boldsymbol{x})$.)

To conclude this section we sketch some applications of the second-quantized Dirac formalism for the special case $N = 1$, $m = 0$, and positive chirality. Even though we could stick to the massless positive-chirality Dirac operator $-\mathrm{i}\,\mathrm{d}/\mathrm{d}x$ on the line, it is simpler and more natural to start from its counterpart on the circle already considered in the last section, cf. [80]. (Under the Cayley transform, the positive- and negative-energy subspaces of $-\mathrm{i}\,\mathrm{d}/\mathrm{d}x$ on $L^2(\mathbb{R})$ correspond to those of $z\mathrm{d}/\mathrm{d}z$ on $L^2(S^1)$, given by [81].) Letting $z = \mathrm{e}^{\mathrm{i}\theta}$, we then obtain

$$\begin{aligned} &\check{H} = -\mathrm{i}\,\mathrm{d}/\mathrm{d}\theta, \quad \check{\mathcal{H}} = L^2([0, 2\pi], \mathrm{d}\theta) \\ &\mathcal{H} = l^2(\mathbb{Z}), \quad \mathcal{H}_+ = l^2(\mathbb{N}), \quad \mathcal{H}_- = l^2(\mathbb{Z}_-) \end{aligned} \qquad [105]$$

and a corresponding Dirac field

$$\Psi(t, \theta) = (2\pi)^{-1/2} \left( \sum_{n=0}^{\infty} a_n \mathrm{e}^{-\mathrm{i}nt + \mathrm{i}n\theta} + \sum_{n=1}^{\infty} b_{-n}^* \mathrm{e}^{\mathrm{i}nt - \mathrm{i}n\theta} \right)$$
$$(t, \theta) \in \mathbb{R} \times [0, 2\pi] \qquad [106]$$

where

$$a_l = c(e_l),\ l \geq 0, \quad b_l = c(e_l),\ l < 0 \qquad [107]$$

and $\{e_l\}_{l \in \mathbb{Z}}$ is the canonical basis of $l^2(\mathbb{Z})$.

Consider now the group $\mathrm{GL}(\mathcal{H})$ of bounded operators on $\mathcal{H}$ with bounded inverses. The transformation

$$\begin{aligned} &\Phi^*(f) \mapsto \Phi^*(Gf), \quad \Phi(f) \mapsto \Phi(G^{-1*}f) \\ &f \in \mathcal{H}, \quad G \in GL(\mathcal{H}) \end{aligned} \qquad [108]$$

leaves the CAR [48] invariant. Provided that $G$ belongs to the subgroup

$$\begin{aligned} &G_2(\mathcal{H}) \\ &= \{G \in \mathrm{GL}(\mathcal{H}) \,|\, G_{\pm\mp} \text{ Hilbert–Schmidt}\} \end{aligned} \qquad [109]$$

there exists an implementer $\tilde{\Gamma}(G)$ on $\mathcal{F}_a(\mathcal{H})$:

$$\begin{aligned} &\tilde{\Gamma}(G)\Phi^*(f) = \Phi^*(Gf)\tilde{\Gamma}(G), \\ &\tilde{\Gamma}(G)\Phi(f) = \Phi(G^{-1*}f)\tilde{\Gamma}(G), \quad \forall f \in \mathcal{H} \end{aligned} \qquad [110]$$

In particular, the multiplication operator

$$\exp(h(x)), \quad h(x) = \sum_{k=1}^{\infty} x_k z^{-k}, \quad z = \mathrm{e}^{\mathrm{i}\theta} \qquad [111]$$

belongs to $G_2(\mathcal{H})$ provided the sequence $x_k$ vanishes sufficiently fast as $k \to \infty$. Thus, one obtains an implementer $\tilde{\Gamma}(\mathrm{e}^{h(x)})$, the so-called KP evolution operator. This designation is justified by the vacuum expectation value

$$\tau(x) = (\Omega, \tilde{\Gamma}(\mathrm{e}^{h(x)})\tilde{\Gamma}(G)\Omega), \quad G \in G_2(\mathcal{H}) \qquad [112]$$

being a tau-function solving the hierarchy of KP evolution equations in Hirota bilinear form, as first shown by Sato and his Kyoto school. For example, the KP equation itself,

$$u_{yy} = \partial_x\left(\tfrac{4}{3}u_t - 2uu_x - \tfrac{1}{3}u_{xxx}\right) \qquad [113]$$

has the bilinear form

$$\left(\frac{\partial^4}{\partial y_1^4} - 4\frac{\partial}{\partial y_1}\frac{\partial}{\partial y_3} + 3\frac{\partial^2}{\partial y_2^2}\right)\tau(x+y)\tau(x-y)\Big|_{y=0} = 0 \qquad [114]$$

the relation being given by

$$x_1 = x, \quad x_2 = y, \quad x_3 = t, \quad u = 2\partial_1^2 \ln \tau \qquad [115]$$

The class of solutions to [113] thus obtained includes not only the rational and soliton solutions (which correspond to choosing $\check{G}$ as multiplication by a rational function of $z = \mathrm{e}^{\mathrm{i}\theta}$ that does not vanish on $S^1$), but also the finite-gap solutions associated with compact Riemann surfaces. Moreover, for suitable subgroups of $G_2(\mathcal{H})$, one obtains tau-functions for related soliton hierarchies, including the Korteweg–de Vries, Boussinesq and Hirota–Satsuma hierarchies. Even though the class of solutions associated with $G_2(\mathcal{H})$ via the Dirac formalism is large, it should be noted that from the perspective of the Cauchy problem for the pertinent evolution equations the solutions are nongeneric, inasmuch as the initial data are real-analytic functions.

Finally, we consider Lie algebra representations related to the above special starting point [105] for the second-quantized Dirac framework. Assume that $\exp(tA)$ is a one-parameter group of bounded operators on $\mathcal{H}$ with generator $A$ in the Lie algebra of $G_2(\mathcal{H})$,

$$\begin{aligned} &g_2(\mathcal{H}) \\ &= \{A \text{ bounded} \,|\, A_{\pm\mp} \text{ Hilbert–Schmidt}\} \end{aligned} \qquad [116]$$

Then one can take

$$\tilde{\Gamma}(\exp(tA)) = \exp(t\mathrm{d}\tilde{\Gamma}(A)) \qquad [117]$$

where $\mathrm{d}\tilde{\Gamma}(A)$ is the Fock-space operator uniquely determined up to an additive constant by its commutation relation

$$[\mathrm{d}\tilde{\Gamma}(A), \Phi^*(f)] = \Phi^*(Af), \quad \forall f \in \mathcal{H} \qquad [118]$$

with the smeared Dirac field $\Phi^*(f)$. Fixing the constant by requiring

$$(\Omega, \mathrm{d}\tilde{\Gamma}(A)\Omega) = 0 \qquad [119]$$

the map $A \mapsto \mathrm{d}\tilde{\Gamma}(A)$ satisfies the Lie algebra relations

$$[\mathrm{d}\tilde{\Gamma}(A), \mathrm{d}\tilde{\Gamma}(B)] = \mathrm{d}\tilde{\Gamma}([A,B]) + C(A,B)\mathbf{1} \qquad [120]$$

so that the term

$$C(A,B) = \mathrm{tr}(A_{-+}B_{+-} - B_{-+}A_{+-}) \qquad [121]$$

encodes a central extension of the Lie algebra $g_2(\mathcal{H})$ [116].

The developments sketched in the previous paragraph are in fact independent of the specific form of the Hilbert space $\mathcal{H}$ and its $\mathcal{H}_+/\mathcal{H}_-$ decomposition. But the special feature of the choice [105] and its $S^1 \to \mathbb{R}$ analog is that the smeared Dirac current

$$\int_0^{2\pi} \mathrm{d}\theta\, \psi(\theta) : \Psi^*(0,\theta)\Psi(0,\theta) :, \quad \psi \in C^\infty(S^1) \qquad [122]$$

(where the double dots denote normal ordering – the replacement of terms involving $b_k b_l^*$ by $-b_l^* b_k$) is of the form $\mathrm{d}\tilde{\Gamma}(A_\psi)$ with $A_\psi \in g_2(\mathcal{H})$ determined by $\psi$. (For spacetime dimension $d > 2$, this is no longer true, as the Hilbert–Schmidt condition is violated.) Moreover, [120] reduces to

$$[\mathrm{d}\tilde{\Gamma}(A_\psi), \mathrm{d}\tilde{\Gamma}(A_\phi)] = C(A_\psi, A_\phi)\mathbf{1} \qquad [123]$$

with the central extension explicitly given by

$$C(A_\psi, A_\phi) = \frac{\mathrm{i}}{2\pi}\int_0^{2\pi} \mathrm{d}\theta\, \psi'(\theta)\phi(\theta) \qquad [124]$$

We have just sketched the details of the (simplest version of the) Dirac current algebra: the term [124] is commonly known as the Schwinger term, so that the central extension featuring in [120]–[121] may be viewed as a generalization. The above setup can also be slightly generalized so as to obtain representations of the Virasoro algebra, which is a central extension of the Lie algebra of polynomial vector fields on $S^1$. The general framework has a quite similar version for the neutral Dirac field (Majorana field), described in terms of the self-dual CAR algebra. In the neutral setting, one can construct the Neveu–Schwarz and Ramond representations of the Virasoro algebra, which are crucial in string theory.

Tensoring $\check{\mathcal{H}}$ with an internal symmetry space $\mathbb{C}^k$ and starting from the Lie algebra of rational maps $S^1 \to \mathrm{sl}(k, \mathbb{C})$, $z \mapsto M(z)$, with poles occurring solely at $z = 0$ and $z = \infty$ (regarded as multiplication operators on $L^2(S^1)^k$), the Fock-space counterparts obtained via the $\mathrm{d}\tilde{\Gamma}$-operation yield representations of the Kac–Moody Lie algebra $A^{(1)}_{k-1}$. Specifically, on the charge-0 sector of $\mathcal{F}_a(\mathcal{H})$, one obtains the so-called basic representation, whereas the charge-$q$ sectors with $q = 1, \ldots, k-1$, yield the fundamental representations. Using the neutral version of Dirac's second quantization, one can also obtain the basic and a fundamental representation of the Kac–Moody algebras $B^{(1)}_l$ (for $k = 2l+1$) and $D^{(1)}_l$ (for $k = 2l$).

*See also:* Bosons and Fermions in External Fields; Clifford Algebras and Their Representations; Current Algebra; Dirac Fields in Gravitation and Nonabelian Gauge Theory; Gerbes in Quantum Field Theory; Holonomic Quantum Fields; Index Theorems; Quantum Field Theory in Curved Spacetime; Quantum Chromodynamics; Random Walks in Random Environments; Relativistic Wave Equations Including Higher Spin Fields; Solitons and Kac–Moody Lie Algebras; Spinors and Spin Coefficients; Symmetry Classes in Random Matrix Theory.

## Further Reading


Bjorken JD and Drell SD (1964) *Relativistic Quantum Mechanics.* New York: McGraw-Hill.

Carey AL and Ruijsenaars SNM (1987) On fermion gauge groups, current algebras and Kac–Moody algebras. *Acta Applicandae Mathematicae* 10: 1–86.

Date E, Jimbo M, Kashiwara M, and Miwa T (1983) Transformation groups for soliton equations. In: Jimbo M and Miwa T (eds.) *Proceedings of RIMS Symposium, Nonlinear Integrable Systems – Classical Theory and Quantum Theory*, pp. 39–119. Singapore: World Scientific.

Dirac PAM (1928) The quantum theory of the electron. *Proceedings of the Royal Society of London. Series A* 117: 610–624.

Dirac PAM (1928) The quantum theory of the electron, II. *Proceedings of the Royal Society of London. Series A* 118: 351–361.

Glimm J and Jaffe A (1981) *Quantum Physics*. New York: Springer.

Itzykson C and Zuber JB (1980) *Quantum Field Theory*. New York: McGraw-Hill.

Pressley A and Segal G (1986) *Loop Groups*. Oxford: Clarendon.

Rose ME (1961) *Relativistic Electron Theory*. New York: Wiley.

Ruijsenaars SNM (1989) Index formulas for generalized Wiener–Hopf operators and boson–fermion correspondence in 2*N* dimensions. *Communications in Mathematical Physics* 124: 553–593.

Schweber SS (1961) *An Introduction to Relativistic Quantum Field Theory*. Evanston, IL: Row-Peterson.

Streater RF and Wightman AS (1964) *PCT, Spin and Statistics, and All That*. New York: Benjamin.

Thaller B (1992) *The Dirac Equation*. New York: Springer.

# Dispersion Relations

**J Bros**, CEA/DSM/SPhT, CEA/Saclay, Gif-sur-Yvette, France



## Introduction

Dispersion relations constitute a basic chapter of mathematical physics which covers various types of classical and quantum scattering phenomena and illustrates in a typical way the importance of general principles in theoretical physics, among which causality plays a major role. Each such phenomenon is described in terms of a scattering amplitude $F(\omega)$, which is a complex-valued function of a frequency variable $\omega$; in quantum physics, this variable becomes an energy variable called $E$ (or $s$ in particle physics), as it follows from the fundamental de Broglie relation $E = \hbar\omega$. The real and imaginary parts of $F(\omega)$, which are called respectively the dispersive part $D(\omega)$ and the absorptive part $A(\omega)$ of $F$, have well-defined physical interpretations for all these phenomena; they represent quantities which are essentially accessible to measurements. The term dispersion relations refers to linear integral equations which relate the functions $D(\omega)$ and $A(\omega)$; such integral equations are always closely related to the Cauchy integral representation of a subjacent holomorphic function $\hat{F}(\omega^{(c)})$ of the complexified frequency (or energy) variable $\omega^{(c)}$. $\hat{F}(\omega^{(c)})$ is called the holomorphic scattering function or in short the scattering function, and the scattering amplitude appears as the boundary value of the latter, taken at positive real values of $\omega$ from the upper half-plane of $\omega^{(c)}$, namely

$$F(\omega) = \lim_{\varepsilon \to 0} \hat{F}(\omega + i\varepsilon), \quad \varepsilon > 0$$

Historically, the first relations of that type to be obtained were the Kramers–Krönig relations (1926), which concern the propagation of light in a dielectric medium. In this basic example, $F(\omega)$ represents the complex refractive index of the medium $n'(\omega) = n(\omega) + i\kappa(\omega)$ for a monochromatic wave with frequency $\omega$. The dispersive part $D(\omega)$ is the real refractive index $n(\omega)$, which is the inverse ratio of the phase velocity of the wave in the medium to its velocity $c$ in the vacuum: the fact that it depends on the frequency $\omega$ corresponds precisely to the phenomenon of dispersion of light in a dielectric medium. A slab of the latter thus appears as a prototype of a macroscopic scatterer. The absorptive part $A(\omega)$ is the rate of exponential damping $\kappa(\omega)$ of the wave, caused by the absorption of energy in the medium.

It has appeared much later that for many scattering phenomena, dispersion relations can be derived from an appropriate set of general physical principles. This means that inside a certain axiomatic framework these relations are model independent with respect to the detailed structure of the scatterer or to the detailed type of particle interaction in the quantum case.

In a very short and oversimplifying way, the following logical scheme holds. At first, one can say that any mathematical formulation of a physical principle of causality results in support-type properties with respect to a time variable $t$ of an appropriate "causal structural function" $R(t)$ of the physical system considered: typically, such a causal function should vanish for negative values of $t$. It follows that its Fourier transform $\tilde{R}$ admits an analytic continuation $\tilde{R}^{(c)}$ in the upper half-plane of the corresponding conjugate variable, interpreted as a frequency (or an energy in the quantum case): here is the general reason for the occurrence of complex frequencies and of holomorphic functions of such variables. In fact, the relevant holomorphic scattering function $\hat{F}(\omega^{(c)})$ always appears as generated by $\tilde{R}^{(c)}$ via some (more or less sophisticated) procedure: in the simplest case, $\hat{F}$ coincides with $\tilde{R}^{(c)}$ itself, but this is not so in general. Finally, the derivation of suitable analyticity and boundedness properties of $\hat{F}(\omega^{(c)})$ in a domain whose typical form is the upper half-plane, allows one to apply a Cauchy-type integral representation to this function; the dispersion relations directly follow from the latter.

The first part of this article aims to describe the most typical dispersion relations and their link with the Cauchy integral. It then presents two basic illustrations of these relations, which are: (1) in classical physics, the Kramers–Krönig relations mentioned above, and (2) in quantum physics, the dispersion relations for the forward scattering of equal-mass particles. The aim of the subsequent parts is to give as complete as possible accounts of the derivation of the relevant analyticity domains inside appropriate axiomatic frameworks which, respectively, contain the previous two examples. The simplest axiomatic framework is the one which governs all the phenomena of linear response: in the latter, the proof of analyticity and dispersion relations most easily follows the logical line sketched above. It will be presented together with its application to the derivation of

the Kramers–Krönig relations. The rest of the article is devoted to the derivation of the so-called crossing analyticity domains which are the relevant background of dispersion relations for the two-particle scattering (or collision) amplitudes in particle physics. This derivation relies on the general axiomatic framework of relativistic quantum field theory (QFT) (*see* Axiomatic Quantum Field Theory) and more specifically on the "analytic program in complex momentum space" of the latter. This framework, whose rigorous mathematical form has been settled around 1960, represents the safest conceptual approach for describing the particle collision processes in a range of energies which covers by far all those that can be produced and will be produced in the accelerators for several decades. A simple account of the field-theoretical axiomatic framework and of the logical line of the derivation of dispersion relations will be presented here for the simplest kinematical situations. A broader presentation of the analytic program including an extended class of analyticity properties for the general structure functions and (two-particle and multiparticle) collision amplitudes in QFT can also be found in this encyclopedia (*see* Scattering in Relativistic Quantum Field Theory: The Analytic Program). For brevity, we shall not treat here the derivation of dispersion relations in the framework of nonrelativistic potential theory. Concerning the latter, the interested reader can refer to the book by Nussenzweig (1972). A collection of old basic papers on field-theoretical dispersion relations can be found in the review book edited by Klein (1961). For a recent and well-documented review of the multiplicity of versions and applications of dispersion relations and their experimental checking, the reader can consult the article by Vernov (1996).

## Typical Dispersion Relations

The possibility of defining the scattering function $\hat{F}(\omega^{(c)})$ in the full upper half-plane and of exploiting the corresponding boundary value $F$ of $\hat{F}$ on the negative part as well as on the positive part of the real axis will depend on the framework of considered phenomena. For the moment, we do not consider the more general situations which also occur in particle physics and will be described later ("crossing domains" and "quasi-dispersion-relations").

In the simplest cases, the real and imaginary parts $D$ and $A$ of $F$ are extended to negative values of the variable $\omega$ via additional symmetry relations resulting from appropriate "reality conditions." As a typical and basic example, there occurs the symmetry relation $\overline{\hat{F}(\omega^{(c)})} = \hat{F}(-\overline{\omega^{(c)}})$, (with $\omega^{(c)}$ and $-\overline{\omega^{(c)}}$ in the upper half-plane) and correspondingly $D(\omega) = D(-\omega)$, $A(\omega) = -A(-\omega)$ on the reals; we shall call ($S$) this symmetry relation.

The simplest case of dispersion relations is then obtained when $D$ and $A$ are linked by the reciprocal Hilbert transformations:

$$D(\omega) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} A(\omega') \frac{1}{\omega' - \omega}\, d\omega' \qquad [1a]$$

$$A(\omega) = -\frac{1}{\pi} P \int_{-\infty}^{+\infty} D(\omega') \frac{1}{\omega' - \omega}\, d\omega' \qquad [1b]$$

where $P$ denotes Cauchy's principal value, defined for any differentiable function $\varphi(x)$ (sufficiently regular at infinity) by

$$P \int_{-\infty}^{+\infty} \frac{\varphi(x)}{x}\, dx = \lim_{\varepsilon \to 0} \left[ \int_{-\infty}^{-\varepsilon} \varphi(x) \frac{dx}{x} + \int_{\varepsilon}^{+\infty} \varphi(x) \frac{dx}{x} \right] \qquad [2]$$

As a matter of fact, the pair of equations [1a], [1b] is equivalent to the following relation for $F \doteq D + iA$:

$$F(\omega) = \frac{1}{2i\pi} \int_{-\infty}^{+\infty} F(\omega') \lim_{\varepsilon \to 0} \frac{1}{\omega' - \omega - i\varepsilon}\, d\omega' \qquad [3]$$

The latter is obtained as a limiting case of the Cauchy formula

$$\hat{F}(\omega^{(c)}) = \frac{1}{2i\pi} \int_{-\infty}^{+\infty} \frac{F(\omega')}{\omega' - \omega^{(c)}}\, d\omega' \qquad [4]$$

expressing the fact that $\hat{F}$ is holomorphic and sufficiently decreasing at infinity in the upper half-plane $\mathcal{I}_+$ of the complex variable $\omega^{(c)}$ and that $F(\omega)$ is the boundary value of $\hat{F}(\omega^{(c)})$ on all the reals.

Finally, one checks that in view of the symmetry relation ($S$), the Hilbert integral relations between $D$ and $A$ given above reduce to the following dispersion relations:

$$D(\omega) = \frac{2}{\pi} P \int_{0}^{+\infty} A(\omega') \frac{\omega'}{\omega'^2 - \omega^2}\, d\omega' \qquad [5a]$$

$$A(\omega) = -\frac{2\omega}{\pi} P \int_{0}^{+\infty} D(\omega') \frac{1}{\omega'^2 - \omega^2}\, d\omega' \qquad [5b]$$

### Two Basic Examples

1. *The Kramers–Krönig relation in classical optics* It will be shown in the next part that the complex refractive index $n'(\omega) = n(\omega) + i\kappa(\omega)$ of a dielectric medium is the boundary value of a holomorphic

function $\hat{n}'(\omega^{(c)})$ in $\mathcal{I}_+$ satisfying the symmetry relation $(S)$, and such that the integral $\int_{-\infty}^{\infty} |\hat{n}'(\omega + i\eta) - 1|^2 \, d\omega$ is uniformly bounded for all $\eta > 0$.

It follows that all the previous relations are satisfied by the function $\hat{F}(\omega^{(c)}) = \hat{n}'(\omega^{(c)}) - 1$. In particular, the real refractive index $n(\omega)$ and the "extinction coefficient" $\beta(\omega) \doteq 2\omega\kappa(\omega)/c$ ($c$ being the velocity of light in the vacuum) are linked by the following Kramers–Krönig dispersion relation (corresponding to eqn [5a]):

$$n(\omega) - 1 = \frac{c}{2\pi} P \int \frac{\beta(\omega')}{\omega'^2 - \omega^2} \, d\omega' \qquad [6]$$

2. *Dispersion relation for the forward two-particle scattering amplitude in relativistic quantum physics* One considers the following collision phenomenon in particle physics. A particle $\Pi_2$ with mass $m$, called the target and sitting at rest in the laboratory, is collided by an identical particle $\Pi_1$ with relativistic energy $\omega$ larger than $m$ ($= mc^2$; in high-energy physics, one usually chooses units such that $c = 1$). After the collision, the particle $\Pi_1$ is scattered in all possible directions, $\theta$, of space, according to a certain quantum scattering amplitude $T_\theta(\omega)$, whose modulus is essentially the rate of probability for detecting $\Pi_1$ in the direction $\theta$. The forward scattering amplitude $T_0(\omega)$ corresponds to the detection of $\Pi_2$ in the forward longitudinal direction with respect to its incidence direction towards the target. Let us also assume that the particles carry no charge of any kind, so that each particle coincides with its "antiparticle." In that case, $T_0(\omega)$ is shown to be the boundary value of a scattering function $\hat{T}_0(\omega^{(c)})$ enjoying the following properties:

1. it is a holomorphic function in $\mathcal{I}_+$ satisfying the symmetry relation $(S)$;
2. its behavior at infinity in $\mathcal{I}_+$ is such that the integral

$$\int_{-\infty}^{\infty} \left| \frac{\hat{T}_0(\omega + i\eta)}{(\omega + i\eta)^2} \right|^2 d\omega$$

is uniformly bounded for all $\eta > 0$; and
3. under more specific assumptions on the mass spectrum of the subjacent theory, the "absorptive part" $A(\omega) \doteq \operatorname{Im} T_0(\omega)$ vanishes for $|\omega| < m$.

Then by applying eqn [5a] to the function $D(\omega) \doteq \operatorname{Re}[(T_0(\omega) - T_0(0))/\omega^2]$ (regular at $\omega = 0$), one obtains the following dispersion relation:

$$\operatorname{Re} T_0(\omega) = T_0(0) + \frac{2\omega^2}{\pi} P \int_m^{+\infty} A(\omega') \frac{1}{\omega'(\omega'^2 - \omega^2)} \, d\omega' \qquad [7]$$

**Remark** In view of (3), the scattering function $\hat{T}_0(\omega^{(c)})$ admits an analytic continuation as an even function of $\omega^{(c)}$ (still called $\hat{T}_0$) in the cut-plane $\boldsymbol{C}_m^{(\text{cut})} \doteq \boldsymbol{C} \backslash \{\omega \in \boldsymbol{R}; |\omega| \geq m\}$. In fact, in view of $(S)$ and (3), the boundary value $T_0$ of $\hat{T}_0$ satisfies the relation $T_0(\omega) = T_0(-\omega)$ in the real interval $\delta_m \doteq \{\omega \in \boldsymbol{R}; -m < \omega < m\}$. Let us then introduce the function $\hat{T}_0^-(\omega^{(c)}) \doteq \hat{T}_0(-\omega^{(c)})$ as a holomorphic function of $\omega^{(c)}$ in $\mathcal{I}_-$: one sees that the boundary values of $\hat{T}_0$ and $\hat{T}_0^-$ from the respective domains $\mathcal{I}_+$ and $\mathcal{I}_-$ coincide on $\delta_m$ and therefore admit a common analytic continuation throughout this real interval (in view of "Painlevé's lemma" or "one-dimensional edge-of-the-wedge theorem"). One also notes that in view of $(S)$ the extended function $\hat{T}_0$ satisfies the "reality condition" $\overline{\hat{T}_0(\omega^{(c)})} = \hat{T}_0(\overline{\omega^{(c)}})$ in $\boldsymbol{C}_m^{(\text{cut})}$. The fact that $\hat{T}_0$ is well defined as an even holomorphic function in the cut-plane $\boldsymbol{C}_m^{(\text{cut})}$ has been established in the general framework of QFT, as explained in the last part of this article.

## Phenomena of Linear Response: Causality and Dispersion Relations in the Classical Domain

The subsequent axiomatic framework and results (due to J S Toll (1952, 1956)) concern any physical system which exhibits the following type of phenomena: whenever it receives some excitation signal, called the input and represented by a real-valued function of time $f_{\text{in}}(t)$ with compact support, the system emits a response signal, called the output and represented by a corresponding real-valued function $f_{\text{out}}(t)$, in such a way that the following postulates are satisfied:

(P1) *Linearity*. To every linear combination of inputs $a_1 f_{\text{in},1} + a_2 f_{\text{in},2}$, there corresponds the output $a_1 f_{\text{out},1} + a_2 f_{\text{out},2}$.

(P2) *Reproductibility or time-translation invariance*. Let $\tau$ be a time-translation parameter taking arbitrary real values; to every "time-translated input" $f_{\text{in}}^{(\tau)}(t) \doteq f_{\text{in}}(t - \tau)$, there corresponds the output $f_{\text{out}}^{(\tau)}(t) \doteq f_{\text{out}}(t - \tau)$.

(P3) *Causality*. The effect cannot precede the cause, namely if $t_{\text{in}}$ and $t_{\text{out}}$ denote respectively the lower bounds of the supports of $f_{\text{in}}(t)$ and $f_{\text{out}}(t)$, then there always holds the inequality $t_{\text{in}} \leq t_{\text{out}}$.

(P4) *Continuity of the response*. There exists some continuity inequality which expresses the fact that a certain norm of the output is majorized by a corresponding norm of the input. The case of an $L_2$-norm inequality of the

form $|f_{\text{out}}| \leq |f_{\text{in}}|$ is particularly significant: when the norm $|f| \doteq [\int |f(t)|^2 \, \mathrm{d}t]^{1/2}$ is interpretable as an energy (for the output as well as for the input), it acquires the meaning of a "dissipation" property of the system.

The postulate of linear dependence (P1) of $f_{\text{out}}$ with respect to $f_{\text{in}}$ is obviously satisfied if the response is described by any general kernel $K(t, t')$ such that the following formula makes sense:

$$f_{\text{out}}(t) = \int_{-\infty}^{+\infty} K(t, t') f_{\text{in}}(t') \mathrm{d}t' \qquad [8]$$

Conversely, the existence of a distribution kernel $K$ can be established rigorously under the continuity assumption postulated in (P4) by using the Schwartz nuclear theorem. In full generality (see our comment in the next paragraph), the kernel $K(t, t')$ appears to be a tempered distribution in the pair of variables $(t, t')$ and the previous integral formula holds in the sense of distributions, which means that both sides of eqn [8] must be considered as tempered distributions (in $t$) acting on any smooth test-function $g(t)$ in the Schwartz space $\mathcal{S}$. (Note, for instance, that the trivial linear application $f_{\text{out}} = f_{\text{in}}$ is represented by the kernel $K(t, t') = \delta(t - t')$).

From the reproductibility postulate (P2), it follows that the distribution $K$ can be identified with a distribution of the single variable $\tau = t - t'$, namely $K(t, t') \doteq R(t - t')$. Moreover, the real-valuedness condition imposed to the pairs $(f_{\text{in}}, f_{\text{out}})$ entails that $R$ is real. Finally, the causality postulate (P3) implies that the support of the distribution $R$ is contained in the positive real axis, so that one can write, in the sense of distributions,

$$f_{\text{out}}(t) = \int_{-\infty}^{t} R(t - t') f_{\text{in}}(t') \mathrm{d}t' \qquad [9]$$

The convolution kernel $R(t - t')$ is typically what one calls in physics a "retarded kernel."

If we now introduce the frequency variable $\omega$, which is the conjugate of the time variable $t$, by the Fourier transformation

$$\tilde{f}(\omega) = \int_{-\infty}^{+\infty} f(t) \, \mathrm{e}^{\mathrm{i}\omega t} \, \mathrm{d}t$$

we see that the convolution equation [9] is equivalent to the following one:

$$\tilde{f}_{\text{out}}(\omega) = \tilde{R}(\omega) \tilde{f}_{\text{in}}(\omega) \qquad [10]$$

In the latter, the Fourier transform $\tilde{R}(\omega)$ of $R$ is a tempered distribution, which is the boundary value from the upper half-plane $\mathcal{I}_+$ of a holomorphic function $\tilde{R}^{(\text{c})}(\omega^{(\text{c})})$, called the Fourier–Laplace transform of $R$. $\tilde{R}^{(\text{c})}$ is defined for all $\omega^{(\text{c})} = \omega + \mathrm{i}\eta$, with $\eta > 0$, by the following formula in which the exponential is a good test-function for the distribution $R$ (since exponentially decreasing for $t \to +\infty$):

$$\tilde{R}^{(\text{c})}(\omega^{(\text{c})}) = \int_{0}^{+\infty} R(t) \mathrm{e}^{\mathrm{i}\omega^{(\text{c})} t} \, \mathrm{d}t \qquad [11]$$

More precisely, the tempered-distribution character of $R$ is strictly equivalent to the fact that $\tilde{R}^{(\text{c})}$ is of moderate growth both at infinity and near the reals in $\mathcal{I}_+$, namely that it satisfies a majorization of the following form for some real positive numbers $p$ and $q$:

$$|\tilde{R}^{(\text{c})}(\omega + \mathrm{i}\eta)| \leq C \frac{(1 + |\omega|^2 + \eta^2)^q}{\eta^p} \qquad [12]$$

We thus conclude from eqn [10] that each phenomenon of linear response is represented very simply in the frequency variable by the multiplicative operator $\tilde{R}(\omega)$, whose analytic continuation $\tilde{R}^{(\text{c})}(\omega^{(\text{c})})$ is called the (causal) response function.

### A Typical Illustration: The Damped Harmonic Oscillator

We consider the motion $x = x(t)$ of a damped harmonic oscillator of mass $m$ submitted to an external force $F(t)$. The force is the input ($f_{\text{in}} = F$) and the resulting motion is the output, namely $f_{\text{out}}(t) = x(t)$. All the previous general postulates (P1)–(P4) are then satisfied, but this particular model is, of course, governed by its dynamical equation

$$x''(t) + 2\gamma x'(t) + \omega_0^2 x(t) = \frac{F(t)}{m} \qquad [13]$$

where $\omega_0$ is the eigenfrequency of the oscillator and $\gamma$ is the damping constant ($\gamma > 0$). The relevant solution of this second-order differential equation with constant coefficients is readily obtained in terms of the Fourier transforms $\tilde{x}(\omega)$ of $x(t)$ and $\tilde{F}(\omega)$ of $F(t)$. One can in fact replace eqn [13] by the equivalent equation

$$(-\omega^2 - 2\mathrm{i}\gamma\omega + \omega_0^2)\tilde{x}(\omega) = \frac{\tilde{F}(\omega)}{m} \qquad [14]$$

whose solution is of the form [10], namely $\tilde{x}(\omega) = \tilde{R}(\omega)\tilde{F}(\omega)$, with

$$\begin{aligned} \tilde{R}(\omega) &= -\frac{\tilde{F}(\omega)}{m(\omega^2 + 2\mathrm{i}\gamma\omega - \omega_0^2)} \\ &= -\frac{\tilde{F}(\omega)}{m(\omega - \omega_1)(\omega - \omega_2)} \end{aligned} \qquad [15a]$$

$$\omega_{1,2} = \pm(\omega_0^2 - \gamma^2)^{1/2} - \mathrm{i}\gamma \qquad [15b]$$

It is clear that the rational function defined by eqns [15] admits an analytic continuation in the full complex plane of $\omega^{(c)}$ minus the pair of simple poles $(\omega_1, \omega_2)$ which lie in the lower half-plane. In particular, it is holomorphic (and decreasing at infinity) in $\mathcal{I}_+$, as expected from the previous general result. Moreover, this example suggests that for any particular phenomenon of linear response, the details of the dynamics are encoded in the singularities of the holomorphic scattering function $\tilde{R}^{(c)}(\omega^{(c)})$, which all lie in the lower half-plane. The validity of a dispersion relation only expresses the analyticity (and decrease at infinity) of that function in the upper half-plane, which is model independent.

**Remark** The same mathematical analysis applies to any electric oscillatory circuit, in which the capacitance, inductance, and resistance are involved in place of the parameters $m$, $\omega_0$ and $\gamma$: $f_{\text{in}}$ and $f_{\text{out}}$ correspond respectively to an external electric potential and to the current induced in the circuit; the response function is the admittance of the circuit.

### Application to the Kramers–Krönig Relation

The background of the Kramers–Krönig relation [6], namely the analyticity and boundedness properties of the complex refractive index $\hat{n}'(\omega^{(c)})$ in $\mathcal{I}_+$, is provided by the previous axiomatic framework. However, it is not the quantity $\hat{n}'(\omega^{(c)})$ itself but appropriate functions of the latter which play the role of causal response functions; two phenomena can in fact be exhibited, which both contribute to proving the relevant properties of $\hat{n}'(\omega^{(c)})$.

1. *Propagation of light in a dielectric slab with thickness* $\delta$. One considers the wave front $f_{\text{in}}(t)$ of an incoming wave normally incident upon the slab, with Fourier decomposition

$$f_{\text{in}}(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \tilde{f}(\omega)\, \mathrm{e}^{-\mathrm{i}\omega t}\, \mathrm{d}\omega \qquad [16]$$

After having traveled through the medium, it gives rise to an outgoing wave $f_{\text{out}}(t)$ on the exit face of the slab, whose Fourier decomposition can be written as follows (provided the thickness $\delta$ of the slab is very small):

$$f_{\text{out}}(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \tilde{f}(\omega)\, \mathrm{e}^{-\mathrm{i}\omega(t - n'(\omega)/c\delta)}\, \mathrm{d}\omega \qquad [17]$$

In the latter, the real part of $n'(\omega)/c$ is the inverse of the light velocity in the medium, while its imaginary part takes into account the exponential damping of the wave. The output $f_{\text{out}}$ thus appears as a causal linear response with respect to $f_{\text{in}}$ (since $f_{\text{out}}$ "starts after" $f_{\text{in}}$). According to the general formula [10], the corresponding response function $\tilde{R}_\delta^{(c)}$ can be directly computed from eqns [16] and [17], which yields:

$$\tilde{R}_\delta^{(c)}(\omega^{(c)}) = \mathrm{e}^{\mathrm{i}\omega^{(c)}\hat{n}'(\omega^{(c)})\delta/c} \qquad [18]$$

In view of the previous axiomatic analysis, $\tilde{R}_\delta^{(c)}$ has to be holomorphic and of moderate growth in $\mathcal{I}_+$, and since this holds for all $\delta$'s sufficiently small, it can be shown that the function $\hat{n}'(\omega^{(c)})$ itself is holomorphic and of moderate growth in $\mathcal{I}_+$ (no logarithmic singularity can be produced).

2. *Polarization of the medium produced by an electric field*. The dielectric polarization signal $P(t)$ produced at a point of a medium by an external electric field $E(t)$ is also a phenomenon of linear response which obeys the postulates (P1)–(P4); the corresponding formula [10] reads

$$\tilde{P}(\omega) = \chi'(\omega)\tilde{E}(\omega) \qquad [19a]$$

where $\chi'$ is the complex dielectric susceptibility of the medium, which is related to $n'$ by Maxwell's relation

$$\tilde{\chi}'(\omega) = \frac{[n'^2(\omega) - 1]}{4\pi} \qquad [19b]$$

One thus recovers the fact that $\chi'$ admits an analytic continuation in $\mathcal{I}_+$; one can also show by a physical argument that $\tilde{\chi}'(\omega)$, and thereby $n'(\omega) - 1$, tends to zero as a constant divided by $\omega^2$ when $\omega$ tends to infinity. This behavior at infinity extends to $\hat{n}'(\omega^{(c)}) - 1$ in $\mathcal{I}_+$ in view of the Phragmen–Lindelöf theorem, since $\hat{n}'$ is known (from (1)) to be of moderate growth. This justifies the analytic background of Kramers–Krönig's relation.

## From Relativistic QFT to the Dispersion Relations of Particle Physics: Historical Considerations and General Survey

In the quantum domain, the derivation of dispersion relations for the two-particle scattering (or collision) amplitudes of particle physics has represented, since 1956 and throughout the 1960s, an important conceptual progress for the theoretical treatment of that branch of physics. These phenomena are described in a quantum-theoretical framework in which the basic kinematical variables are the energies and momenta of the particles involved. These variables play the role of the frequency of light in the optical scattering phenomena. Moreover, since large energies and momenta are involved, which allow the occurrence of particle creation

according to the conservation laws of special relativity, it is necessary to use a relativistic quantum-mechanical framework. Around 1950, the success of the quantum electrodynamics formalism for computing the electron–photon, electron–electron, and electron–positron scattering amplitudes revealed the importance of the concept of relativistic quantum field for the understanding of particle physics. However, the methods of perturbation theory, which had ensured the success of quantum electrodynamics in view of the small value of the coupling parameter of that theory (namely the electric charge of the electron), were at that time inapplicable to the strong nuclear interaction phenomena of high-energy physics. This failure motivated an important school of mathematical physicists for working out a model-independent axiomatic approach of relativistic QFT (e.g., Lehmann, Symanzik, Zimmermann (1954), Wightman (1956), and Bogoliubov (1960); *see* Axiomatic Quantum Field Theory). Their main purpose was to provide a conceptually satisfactory treatment of relativistic quantum collisions, at least for the case of massive particles. Among various postulates expressing the invariance of the theory under the Poincaré group in an appropriate quantum-mechanical Hilbert-space framework, the approach basically includes a certain formulation of the principle of causality, called microcausality or local commutativity. This axiomatic approach of QFT was followed by a conceptually important variant, namely the algebraic approach to QFT (Haag, Kastler, Araki 1960), whose most important developments are presented in the book by Haag (1992) (*see* Algebraic Approach to Quantum Field Theory). From the historical viewpoint, and in view of the analyticity properties that they also generate, one can say that all these (closely related) approaches parallel the axiomatic approach of linear response phenomena with, of course, a much higher degree of complexity. In particular, the characterization of scattering (or collision) amplitudes in terms of appropriate structure functions of the basic quantum fields of the theory is a nontrivial preliminary step which was taken at an early stage of the theory under the name of "asymptotic theory and reduction formulae" (Lehmann, Symanzik, Zimmermann 1954–57, Haag–Ruelle 1962, Hepp 1965). There again, in the field-theoretical axiomatic framework, causality generates analyticity through Fourier–Laplace transformation, but several complex variables now play the role which was played by the complex frequency in the axiomatics of linear response phenomena: they are obtained by complexifying the relativistic energy–momentum variables of the (Fourier transforms of the) quantum fields involved in the high-energy collision processes. In fact, the holomorphic functions which play the role of the causal response function $\tilde{R}(\omega)$ are the QFT structure functions or "Green functions in energy–momentum space." The study of all possible analyticity properties of these functions resulting from the QFT axiomatic framework is called the analytic program (*see* Scattering in Relativistic Quantum Field Theory: The Analytic Program). The primary basic scope of the latter concerns the derivation of analyticity properties for the scattering functions of two-particle collision processes, which appears to be a genuine challenge for the following reason. The basic Einstein relation $E = mc^2$, which applies to all the incoming and outgoing particles of the collisions, operates as a geometrical constraint on the corresponding physical energy–momentum vectors: according to the Minkowskian geometry, the latter have to belong to mass hyperboloids, which define the so-called "mass shell" of the collision considered. It is on the corresponding complexified mass-shell manifold that the scattering functions are required to be defined as holomorphic functions. In the analytic program of QFT, the derivation of such analyticity domains and of corresponding dispersion relations in the complex plane of the squared total energy variable, $s$, of each given collision process then relies on techniques of complex geometry in several variables. As a matter of fact, the scattering amplitude is a function (or distribution) of two variables $F(s,t)$, where $t$ is a second important variable, called the squared momentum transfer, which plays the role of a fixed parameter for the derivation of dispersion relations in the variable $s$. The value $t = 0$ corresponds to the special kinematical situation which has been described above (for the case of equal-mass particles $\Pi_1$ and $\Pi_2$) under the name of forward scattering and the variable $s$ is a simple affine function of the energy $\omega$ of the colliding particle $\Pi_1$ in the laboratory Lorentz frame, (namely $s = 2m^2 + 2m\omega$ in the equal-mass case). It is for the corresponding scattering amplitude $T_0(\omega) \doteq F_0(s) \doteq F(s,t)_{|t=0}$ that a dispersion relation such as eqn [7] can be derived, although this derivation is far from being as simple as for the phenomena of linear response in classical physics: even in that simplest case, it already necessitates the use of analytic completion techniques in several complex variables. The first proof of this dispersion relation was performed by K Symanzik in 1956. In the case of general kinematical situations of measurements, the direction of observation of the scattered particle includes a nonzero angle with the incidence direction, which always corresponds to a negative value of $t$. The derivation of dispersion relations at fixed $t = t_0 < 0$, namely for the scattering amplitude

$F_{t_0}(s) \doteq F(s,t)_{|t=t_0}$ requires further arguments of complex geometry, and it is submitted to subtle limitations of the form $t_1 < t_0 \le 0$, where $t_1$ depends on the mass spectrum of the particles involved in the theory. The first rigorous proof of dispersion relations at $t < 0$ was performed by N N Bogoliubov in 1960.

Three conceptually important features of the dispersion relations in particle physics deserve to be pointed out.

1. In comparison with the dispersion relations of classical optics, a feature which appears to be new is the so-called "crossing property," which is characteristic of high-energy physics since it relies basically on the relativistic kinematics. According to that property, the boundary values of the analytic scattering function $\hat{F}_t(s)$ at positive and negative values of $s$ from the respective half-planes $\mathrm{Im}\, s > 0$ and $\mathrm{Im}\, s < 0$ are interpreted, respectively, as the scattering amplitudes of two physically different collision processes, which are deduced from each other by replacing the incident particle by the corresponding antiparticle; one also says that "these two collision processes are related by crossing." A typical example is provided by the proton–proton and proton–antiproton collisions, whose scattering amplitudes are therefore mutually related by the property of analytic continuation. This type of relationship between the values of the scattering function at positive and negative values of $s$ generalizes in a nontrivial way the symmetry relation ($S$) satisfied by the forward scattering function $\hat{T}_0(\omega^{(c)})$ when each particle coincides with its antiparticle (see the second basic example above). No nontrivial crossing property holds in that special case and the fact that $\hat{T}_0$ is an even function of $\omega^{(c)}$ precisely expresses the identity of the two-collision processes related by crossing. In the general case, for $t = 0$ as well as for $t = t_0 < 0$ for any value of $t_0$, the analyticity domain that one obtains for the scattering function is not the full cut-plane of $s$: in its general form, a "crossing domain" may exclude some bounded region $B_{t_0}$ from the cut-plane, but it always contains an infinite region which is the exterior of a circle minus cuts along the two infinite parts of the real $s$-axis (Bros, Epstein, Glaser 1965): these cuts are along the physical regions of the two collision processes related by crossing. In that general case, the scattering function $\hat{F}_t(s)$ still satisfies what can be called a quasi-dispersion-relation, in which the right-hand side contains an additional Cauchy integral, taken along the boundary of $B_t$.

2. A second important feature concerns the behavior at large values of $s$ of the scattering functions $\hat{F}_t(s)$ in their analyticity domain. As indicated in the presentation of the second basic example, a "precise-increase" property was expected to be satisfied by the forward scattering amplitude $T_0(\omega)$ for $\omega$ (or $s$) tending to infinity. This "precise-increase" property implied the necessity of writing the corresponding dispersion relation [7] for the function $(T_0(\omega) - T_0(0))/\omega^2$: this is what one calls a "dispersion relation with a subtraction." As a matter of fact, the existence of such restrictive bounds on the total cross sections at high energies had been discovered in 1961 by M Froissart: his derivation relied basically on the use of the unitarity of the scattering operator (expressing the quantum principle of conservation of probabilities), but also on a strong analyticity postulate for the scattering function not implied by the general field-theoretical approach (namely the Mandelstam domain of "double dispersion relations"). In the general framework of QFT, Froissart-type bounds appeared to be closely linked to a further nontrivial extension of the range of "admissible" values of $t$ for which $\hat{F}_t(s)$ can be analytically continued in a cut-plane or crossing domain. In fact, the extension of this range to positive (i.e., "unphysical") and even complex values of $t$, and as a second step the proof of Froissart-type bounds in $s(\log s)^2$ for $F_t(s)$ at all these admissible values of $t$, were performed in 1966 by A Martin. They rely on a subtle conspiracy of the analyticity properties deduced from the QFT axiomatic framework and of positivity and unitarity properties expressing the basic Hilbertian structure of the quantum collision theory. The consequence of these bounds on the exact form of the dispersion relations is that, as in formula [7] of the case $t = 0$, it is justified to write a (the so-called "subtracted") dispersion relation for $(F_t(s) - F_t(0) - sF_t'(0))/s^2$: for the general case when the crossing property replaces the symmetry ($S$), such a dispersion relation involves two subtractions (since $F_t'(0) \neq 0$). Detailed information concerning the interplay of analyticity and unitarity on the mass shell and the derivation of refined forms of dispersion relations and various boundedness properties for the scattering functions are given in the book by Martin (1969).

3. Constraints imposed by dispersion relations and experimental checks. The conceptual importance of dispersion relations incorporating the above features (1) and (2) is displayed by such spectacular application as the relationship between the high-energy behaviors of proton–proton and proton–antiproton cross sections. Even though the closest forms of relationship between these cross sections (e.g., the existence of equal high-energy limits) necessitate for their proof some extra assumption concerning, for instance, the behavior

of the ratio between the dispersive and absorptive parts of the forward scattering amplitude, one can speak of an actual model-independent implication of general QFT that imposes nontrivial constraints on phenomena. Otherwise stated, checking experimentally the previous type of relationship up to the limits of high energies imposed by the present technology of accelerators constitutes an indirect, but important test of the validity of the general principles of QFT.

As a matter of fact, it has also appeared frequently in the literature of high-energy physics during the last 40 years that the Froissart bound by itself was considered as a key criterion to be satisfied by any sensible phenomenological model in particle physics. As already stated above, the Froissart bound is one of the deepest consequences of the analytic program of general QFT, since its derivation also incorporates in the most subtle way the quantum principle of probability conservation. Would it be only for the previous basic results, the derivation of dispersion relations (and, more generally, the results of the analytic program) in QFT appear as an important conceptual bridge between a fundamental theoretical framework of relativistic quantum physics and the phenomenology of high-energy particle physics.

## Basic Concepts and Main Steps in the QFT Derivation of Dispersion Relations

The rest of this article outlines the derivation of the analytic background of dispersion relations for the forward scattering amplitudes in the framework of axiomatic QFT. After a brief introduction on relativistic scattering processes and the problematics of causality in particle physics, it gives an account of the Wightman axioms and the simplest reduction formula which relates the forward scattering amplitude to a retarded product of the field operators. Then it describes how the latter can be used for justifying a certain type of analyticity domain for the forward scattering functions, namely a crossing domain or in the best cases a cut-plane in the squared energy variable $s$. This is the basic result that allows one to write dispersion relations (or quasi-dispersion-relations) at $t=0$; the exact form of the latter, including at most two subtractions, relies on the use of Hilbertian positivity and of the unitarity of the scattering operator.

### Relativistic Quantum Scattering as a Phenomenon of Linear Response

Collisions of quantum particles may be seen as phenomena of linear response, but in a way which differs greatly from what has been previously described.

**Particles in Minkowskian geometry** Each state of a relativistic classical particle with mass $m$ is characterized by its energy–momentum vector or 4-momentum $p=(p_0,\boldsymbol{p})$ satisfying the mass-shell condition $p^2 \doteq p_0^2-\boldsymbol{p}^2=m^2$ (in units such that $c=1$). In view of the condition of positivity of the energy $p_0>0$ the "physical mass shell" thus coincides with the positive sheet $H_m^+$ of the mass hyperboloid $H_m$ with equation $p^2=m^2$.

The set of all energy–momentum configurations characterizing the collisions of two relativistic classical particles with initial (resp. final) 4-momenta $p_1,p_2$ (resp. $p_1',p_2'$) is the mass-shell manifold $\mathcal{M}$ defined by the conditions

$$p_i^2=m^2,\quad p_i'^2=m^2,\quad p_{i,0}>0,\quad p'_{i,0}>0,\quad i=1,2$$
$$p_1+p_2=p_1'+p_2'$$

where the latter equation expresses the relativistic law of total energy–momentum conservation. $\mathcal{M}$ is an eight-dimensional manifold, invariant under the (six-dimensional) Lorentz group: the orbits of this group that constitute a foliation of $\mathcal{M}$ are parametrized by two variables, namely the squared total energy $s=(p_1+p_2)^2=(p_1'+p_2')^2$ and the squared momentum transfer $t=(p_1-p_1')^2=(p_2-p_2')^2$ (or $u=(p_1-p_2')^2=4m^2-s-t$). In these variables, called the Mandelstam variables, the "physical region" $\Phi$ of the collision is represented by the set of pairs $(s,t)$ (or triplets $(s,t,u)$ with $s+t+u=4m^2$) such that $t\leq 0, u\leq 0$, and therefore $s\geq 4m^2$.

Correspondingly, each state of a relativistic quantum particle with mass $m$ is characterized by a wave packet $\hat{f}(p)$ on $H_m^+$, which is an element of unit norm of $L_2(H_m^+;\mu_m(p))$, with $\mu_m(p)=d\boldsymbol{p}/(\boldsymbol{p}^2+m^2)^{1/2}$. In Minkowskian spacetime with coordinates $x=(x_0,\boldsymbol{x})$, any such state is represented by a wave function $f(x)$ whose Fourier transform is the tempered distribution (with support in $H_m^+$) $\hat{f}(p)\times\delta(p^2-m^2)$: $f(x)$ is a positive-energy solution of the Klein–Gordon equation $(\partial^2/\partial x_0^2-\Delta_{\boldsymbol{x}}+m^2)f(x)=0$. A free two-particle state is a symmetric wave packet $\hat{f}(p_1,p_2)$ on $H_m^+\times H_m^+$ in the Hilbert space $L_2(H_m^+\times H_m^+;\mu_m\otimes\mu_m)$.

**Scattering kernels as response kernels: distribution character** While the input to be considered is a free wave packet $\hat{f}_{\text{in}}(p_1,p_2)$ on $H_m^+\times H_m^+$, representing the preparation of an initial two-particle state, the output corresponds to the detection of a final two-particle state also characterized by a wave packet $\hat{g}_{\text{out}}(p_1',p_2')$ on $H_m^+\times H_m^+$. In quantum mechanics, linearity is linked to the "superposition principle" of states,

which allows one to state that collisions are described by a certain bilinear form $(\hat{f}_{\text{in}}, \hat{g}_{\text{out}}) \to S(\hat{f}_{\text{in}}, \hat{g}_{\text{out}})$, called the "scattering matrix." This bilinear form is bicontinuous with respect to the Hilbertian norms of the wave packets, and it then results from the Schwartz nuclear theorem that it is represented by a distribution kernel $S(p_1, p_2; p'_1, p'_2)$, namely a tempered distribution with support contained in $\mathcal{M}$, in such a way that (formally)

$$S(\hat{f}_{\text{in}}, \hat{g}_{\text{out}}) = \int \hat{f}_{\text{in}}(p_1, p_2)\overline{\hat{g}_{\text{out}}}(p'_1, p'_2) S(p_1, p_2; p'_1, p'_2) \times \mu_m(p_1)\mu_m(p_2)\mu_m(p'_1)\mu_m(p'_2) \quad [20]$$

If there were no interaction, $S(\hat{f}_{\text{in}}, \hat{g}_{\text{out}})$ would reduce to the Hilbertian scalar product $<\hat{g}_{\text{out}}, \hat{f}_{\text{in}}>$ in $L_2(H_m^+ \times H_m^+; \mu_m \otimes \mu_m)$ and the corresponding kernel $S$ would be the identity kernel

$$I(p_1, p_2; p'_1, p'_2) = \frac{1}{2}\left[\delta(p_1 - p'_1)\delta(p_2 - p'_2) + \delta(p_1 - p'_2)\delta(p_2 - p'_1)\right]$$

In the general case, the interaction is therefore described by the scattering kernel $T(p_1, p_2; p'_1, p'_2) \doteq S(p_1, p_2; p'_1, p'_2) - I(p_1, p_2; p'_1, p'_2)$. The action of $T$ as a bilinear form (defined in the same way as the action of $S$ in eqn [20]) may be seen as the quantum analog of the classical response formula [10]. Note, however, the difference in the mathematical treatment of the output: instead of being considered as the direct response $(\tilde{f}_{\text{out}})$ to the input, it is now explored by Hilbertian duality in terms of detection wave packets $\hat{g}_{\text{out}}$, in conformity with the principles of quantum theory. Finally, in view of the invariance of the collision process under the Lorentz group, the scattering kernel $T$ is constant along the orbits of this group in $\mathcal{M}$ and it then defines a distribution $F(s, t) \doteq T(p_1, p_2; p'_1, p'_2)$ with support in the physical region $\Phi$: this is what is called the scattering amplitude.

**What becomes of causality?** One can show that the positive-energy solutions of the Klein–Gordon equation cannot vanish in any open set of Minkowski spacetime; they necessarily spread out in the whole spacetime. This makes it impossible to formulate a causality condition comparable to eqn [9] in terms of the spacetime wave functions $f_{\text{in}}$ and $g_{\text{out}}$ corresponding to the input and output wave packets $\hat{f}_{\text{in}}, \hat{g}_{\text{out}}$. In this connection, it is, however, appropriate to note that (after various attempts of "weak causality conditions") a certain condition called "macrocausality" (Iagolnitzer and Stapp 1969; see the book by Iagolnitzer (1992)) has been shown to be equivalent to some local properties of analyticity of the scattering kernel $T$; but it is not our purpose to develop that point here for two reasons: (1) the interpretation of that condition is rather involved, because it integrates a very weak form of causality together with the spatial short-range character of the strong nuclear interactions between the elementary particles; (2) the domains of analyticity obtained are by far too small with respect to those necessary for writing dispersion relations. The reason for this failure is that the scattering kernel only represents an asymptotic quantum observable, in the sense that it is intended to describe observations far apart from the extremely small spacetime region where the particles strongly interact, namely in regions where this interaction is asymptotically small. Although well adapted to what is actually observed in the detection experiments, the concept of scattering kernel is not sufficient for describing the fundamental interactions of physics: it must be enriched by other theoretical concepts which might explicitly take into account the microscopic interactions in spacetime. This motivates the introduction of quantum fields as basic quantities in particle physics.

### Relativistic Quantum Fields: Microcausality and the Retarded and Advanced Kernels; Analyticity in Complex Energy–Momentum Space

By an idealization of the concept of quantum electromagnetic field and a generalization to all types of microscopic interactions of matter, one considers that all the phenomena involving such interactions can be described by fields $\Phi_i(x)$, whose amplitude can, in principle, be measured in arbitrarily small regions of Minkowski spacetime. In the quantum framework, one is thus led to the notion of local observable $\mathcal{O}$ (emphasized as a basic concept in the axiomatic approach of Araki, Haag, and Kastler). In the Wightman field-theoretical framework, a local observable corresponds to the measuring process of a ponderated average of a field $\Phi_i(x)$ of the form $\mathcal{O} \doteq \Phi_i[f] = \int \Phi_i(x) f(x)\, dx$. In the latter, $f(x)$ denotes a smooth real-valued test-function with (arbitrary) compact support $K$ in spacetime; the observable $\mathcal{O}$ is then said to be localized in $K$. Each observable $\mathcal{O} = \Phi_i(f)$ has to be a self-adjoint (unbounded) operator acting in (a dense domain of) the Hilbert space $\mathcal{H}$ generated by all the states of the system of fundamental fields $\{\Phi_i\}$; therefore, the correct mathematical concept of relativistic quantum field $\Phi(x)$ is an "operator-valued tempered distribution on Minkowski spacetime." Here the additional "temperateness assumption" is a convenient technical assumption which in particular allows the passage to the energy–momentum space by making use of the Fourier transformation.

In this QFT framework, it is natural to express a certain form of causality by assuming that two observables $\Phi(f)$ and $\Phi(f')$ commute if the supports of $f$ and $f'$ are spacelike-separated regions in spacetime, which means that no signal with velocity smaller or equal to the velocity of light can propagate from either one of these regions to the other. This expresses the idea that these two observables should be independent, that is, "compatible as quantum observables." This postulate is equivalent to the following condition, called microcausality or local commutativity, and understood in the sense of operator-valued tempered distributions:

$$[\Phi(x_1), \Phi(x_1')] = 0, \quad \text{for } (x_1 - x_1')^2 < 0 \qquad [21]$$

where $(x_1 - x_1')^2$ is the squared Minkowskian pseudonorm of $x = x_1 - x_1' = (x_0, \boldsymbol{x})$, namely $x^2 = x_0^2 - \boldsymbol{x}^2$. It follows that for every admissible pair of states $\Psi, \Psi'$ in $\mathcal{H}$, the tempered distribution

$$C_{\Psi,\Psi'}(x_1, x_1') \doteq \langle \Psi, [\Phi(x_1), \Phi(x_1')]\Psi' \rangle \qquad [22]$$

has its support contained in the union of the sets $\mathcal{V}^+ : x_1 - x_1' \in \overline{V^+}$ and $\mathcal{V}^- : x_1 - x_1' \in \overline{V^-}$, where $\overline{V^+}$ and $\overline{V^-}$ are, respectively, the closures of the forward and backward cones $V^+ \doteq \{x = (x_0, \boldsymbol{x}); x_0 > |\boldsymbol{x}|\}$, $V^- \doteq -V^+$ in Minkowski spacetime. It is always possible to decompose the previous distribution as

$$C_{\Psi,\Psi'}(x_1, x_1') = R_{\Psi,\Psi'}(x_1, x_1') - A_{\Psi,\Psi'}(x_1, x_1') \qquad [23]$$

in such a way that the supports of the distributions $R_{\Psi,\Psi'}(x_1, x_1')$ and $A_{\Psi,\Psi'}(x_1, x_1')$ belong, respectively, to $\mathcal{V}^+$ and $\mathcal{V}^-$. $R_{\Psi,\Psi'}$ and $A_{\Psi,\Psi'}$ are called, respectively, retarded and advanced kernels and they are often formally expressed (for convenience) as follows:

$$R_{\Psi,\Psi'}(x_1, x_1') = \theta(x_{1,0} - x_{1,0}')C_{\Psi,\Psi'}(x_1, x_1')$$
$$A_{\Psi,\Psi'}(x_1, x_1') = -\theta(x_{1,0} - x_{1,0}')C_{\Psi,\Psi'}(x_1, x_1')$$

in terms of the Heaviside step function $\theta(t)$ of the time-coordinate difference $t = x_{1,0} - x_{1,0}'$. For every pair $(\Psi, \Psi'), R_{\Psi,\Psi'}(x_1, x_1')$ appears as a relativistic generalization of the retarded kernel $R(t - t')$ of eqn [10]: its support property in spacetime, similar to the support property of $R$ in time, expresses a relativistic form of causality, or "Einstein causality."

There exists a several-variable extension of the theory of Fourier–Laplace transforms of tempered distributions which is based on a formula similar to eqn [11]. We introduce the vector variables $X = (x_1 + x_1')/2, x = x_1 - x_1'$ and a complex 4-momentum $k = p + iq = (k_0, \boldsymbol{k})$ as the conjugate vector variable of $x$ with respect to the Minkowskian scalar product $k \doteq k_0 x_0 - \boldsymbol{k} \cdot \boldsymbol{x}$, and we define

$$\tilde{R}^{(c)}_{\Psi,\Psi'}(k, X) = \int_{\overline{V}^+} R_{\psi,\Psi'}\left(X + \frac{x}{2}, X - \frac{x}{2}\right) e^{ik\cdot x}\, dx \qquad [24]$$

Since $q \cdot x > 0$ for all pairs $(q, x)$ such that $q \in V^+$, $x \in \overline{V}^+$, it follows that $\tilde{R}^{(c)}_{\Psi,\Psi'}(k, X)$ is holomorphic with respect to $k$ in the domain $\mathcal{T}^+$ containing all $k = p + iq$ such that $q$ belongs to $V^+$. Moreover, in the limit $q \to 0$ this holomorphic function tends (in the sense of distributions) to the Fourier transform $\tilde{R}_{\Psi,\Psi'}(p, X)$ of $R_{\Psi,\Psi'}(X + x/2, X - x/2)$ with respect to $x$. The domain $\mathcal{T}^+$, which is called the "forward tube," is the analog of the domain $\mathcal{I}_+$ of the $\omega$-plane; bounds of moderate type comparable to those of [12] apply to the holomorphic function $\tilde{R}^{(c)}_{\Psi,\Psi'}$ in $\mathcal{T}^+$. Similarly, the advanced kernel $A_{\Psi,\Psi'}(X + x/2, X - x/2)$ admits a Fourier–Laplace transform $\tilde{A}^{(c)}_{\Psi,\Psi'}(k, X)$, which is holomorphic and of moderate growth in the "backward tube" $\mathcal{T}^-$ containing all $k = p + iq$ such that $q$ belongs to $V^-$. In view of [23], the Fourier transform $\tilde{C}_{\Psi,\Psi'}(p, X)$ of $C_{\Psi,\Psi'}(X + x/2, X - x/2)$ then appears as the difference between the boundary values of $\tilde{R}^{(c)}_{\Psi_1,\Psi_2}$ and $\tilde{A}^{(c)}_{\Psi_1,\Psi_2}$ on the reals (from the respective domains $\mathcal{T}^+$ and $\mathcal{T}^-$).

### The Field-Theoretical Axiomatic Framework and the Passage from the Structure Functions of QFT to the Scattering Kernels (Case of Forward Scattering)

**The postulates (Wightman axioms)** Apart from the causality postulate, which we have already presented above in view of its distinguished role for generating analyticity properties in complex energy–momentum space, the field-theoretical axiomatic approach to collision theory is based on the following postulates (for all the fundamental developments of axiomatic field theory, the interested reader may consult the books by Streater and Wightman (1980) and by Jost (1965); *see* Axiomatic Quantum Field Theory).

1. There exists a unitary representation $g \to \mathcal{U}(g)$ of the Poincaré group $G$ in the Hilbert space of states $\mathcal{H}$; in this representation, the abelian subgroup of translations of space and time has a Lie algebra whose generators are interpreted as the four self-adjoint (commuting) operators $P_\mu$ of total energy–momentum of the system.
2. The quantum field operators $\Phi(x)$ transform covariantly under that representation; in the simplest case of scalar fields (considered here), $\Phi(gx) = U(g)\Phi(x)U(g^{-1})$.
3. There exists a unique state $\Omega$, called the vacuum, such that the action of all polynomials of field operators on $\Omega$ generates a dense subset of $\mathcal{H}$;

moreover, $\Omega$ is assumed to be invariant under the representation $\mathcal{U}$ of $G$, and thereby such that $P_\mu \Omega = 0$.

4. Spectral condition or positivity of energy in all physical states. The joint spectrum $\Sigma$ of the operators $P_\mu$ is contained in the closed forward cone $\overline{V^+}$ of energy–momentum space. In order to perform the collision theory of massive particles, one needs a more detailed "mass-gap assumption": $\Sigma$ is the union of the origin $O$, of one or several positive sheets of hyperboloid $H^+_{m_i}$ and of a region $V^+_M$ defined by the conditions $p^2 \geq M^2, p_0 > 0$, with $M$ larger than all the $m_i$.

The Hilbert space $\mathcal{H}$ is correspondingly decomposed as the direct sum of the vacuum subspace (or zero-particle subspace) generated by $\Omega$, of subspaces of stable one-particle states with masses $m_i$ isomorphic to $L_2(H^+_{m_i}, \mu_{m_i})$, and of a remaining subspace $\mathcal{H}'$. As a result of the construction of "asymptotic states," $\mathcal{H}'$ can be shown to contain two subspaces $\mathcal{H}'_{\text{in}}$ and $\mathcal{H}'_{\text{out}}$, generated, respectively, by $N$-particle incoming states (with $N$ arbitrary and $\geq 2$) and by $N$-particle outgoing states. The collision operator $S$ is then defined as the partially isometric operator from $\mathcal{H}'_{\text{out}}$ onto $\mathcal{H}'_{\text{in}}$, which maps a reference basis of outgoing states onto the corresponding basis of incoming states.

**An independent postulate: asymptotic completeness** (*see* **Scattering, Asymptotic Completeness and Bound States and Scattering in Relativistic Quantum Field Theory: Fundamental Concepts and Tools**) The theory is said to satisfy the property of asymptotic completeness if all the states of $\mathcal{H}$ can be interpreted as superpositions of various $N$-particle states (either in the incoming or in the outgoing state basis), namely if one has $\mathcal{H}' = \mathcal{H}'_{\text{in}} = \mathcal{H}'_{\text{out}}$. This property is not implied by the previous postulates on quantum fields, but its physical interpretation and its role in the analytic program are of primary importance (see Scattering in Relativistic Quantum Field Theory: The Analytic Program). Let us simply note here that asymptotic completeness implies as a by-product the unitarity property of the collision operator $S$ on the full Hilbert space $\mathcal{H}'$ (i.e., $SS^* = S^*S = I$).

**Connection between retarded kernel and scattering kernel for the forward scattering case; a simple "reduction formula"** We consider the scattering of a particle $\Pi_1$ with mass $m_1$ on a target consisting of a particle $\Pi_2$ with mass $m_2$ and denote by $T(p_1, p_2; p'_1, p'_2)$ the corresponding scattering kernel (defined similarly as for the case of equal-mass particles considered earlier). Equations [22]–[24] are then applied to the case when $\Psi$ and $\Psi'$ coincide with a one-particle state of $\Pi_2$ at rest, namely with 4-momentum $p_2 = p'_2$ along the time axis: $p_2 = ((p_2)_0, 0), (p_2)_0 = m_2$. This describes in a simple way the case of forward scattering, since in view of the energy–momentum conservation law $p_1 + p_2 = p'_1 + p'_2$, the choice $p_2 = p'_2$ also implies that $p_1 = p'_1$. (The possibility of restricting the distribution $T(p_1, p_2; p'_1, p'_2)$ to such fixed values of the energy–momenta is shown to be mathematically well justified). The advantage of this simple case is that the corresponding kernels [22], [23] of $(x_1, x'_1)$ are invariant under spacetime translations and therefore depend only on $x$ (and not on $X$). We can thus rewrite eqns [22], [23] with simplified notations as follows:

$$\begin{aligned} C_{p_2}(x) &\doteq <p_2, \left[\Phi\left(\frac{x}{2}\right), \Phi\left(-\frac{x}{2}\right)\right] p_2> \\ &= R_{p_2}(x) - A_{p_2}(x) \end{aligned} \qquad [25]$$

which can be shown to give correspondingly by Fourier transformation

$$\begin{aligned} \tilde{C}_{p_2}(p) &= <p_2, \tilde{\Phi}(p)\tilde{\Phi}(-p)p_2> - <p_2, \tilde{\Phi}(-p)\tilde{\Phi}(p)p_2> \\ &= \tilde{R}_{p_2}(p) - \tilde{A}_{p_2}(p) \end{aligned} \qquad [26]$$

If the particle $\Pi_1$ appears in the asymptotic states of the field $\Phi$, the scattering kernel $T(p_1, p_2; p'_1, p'_2)$ is then given in the forward configurations $p_1 = p'_1 \in H^+_{m_1}, p_2 = p'_2 \in H^+_{m_2}$, by the following reduction formula in which $s = (p_1 + p_2)^2$:

$$\begin{aligned} F_0(s) &= T(p_1, p_2; p_1, p_2) \\ &= \left[(p_1^2 - m_1^2)\tilde{R}_{p_2}(p_1)\right]_{|H^+_{m_1}} \end{aligned} \qquad [27]$$

## Analyticity Domains in Energy–Momentum Space: From the "Primitive Off-Shell Domains" of QFT to the Crossing Manifolds on the Mass Shell

For simplicity, we shall restrict ourselves to the consideration of forward scattering amplitudes, namely to the derivation of crossing analyticity domains and (quasi-)dispersion relations at $t = 0$ for two-particle collision processes of the form $\Pi_1 + \Pi_2 \to \Pi_1 + \Pi_2$, $\Pi_1$ and $\Pi_2$ being given massive particles with arbitrary spins and charges.

**The holomorphic function $H_{p_2}(k)$ and its primitive domain $D$. Nontriviality of dispersion relations for the scattering amplitudes** As suggested by eqn [24], we can exploit the analyticity properties of the Fourier–Laplace transforms of the retarded and advanced kernels $R_{p_2}$ and $A_{p_2}$: in fact, $\tilde{R}_{p_2}(p)$ and

$\tilde{A}_{p_2}(p)$ are, respectively, the boundary values of the holomorphic functions

$$\begin{aligned}\tilde{R}^{(c)}_{p_2}(k) &= \int_{\overline{V}^+} R_{p_2}(x)\,\mathrm{e}^{\mathrm{i}k\cdot x}\,\mathrm{d}x \\ \tilde{A}^{(c)}_{p_2}(k) &= \int_{\overline{V}^-} A_{p_2}(x)\,\mathrm{e}^{\mathrm{i}k\cdot x}\,\mathrm{d}x\end{aligned} \qquad [28]$$

from the corresponding domains $\mathcal{T}^+$ and $\mathcal{T}^-$. According to the reduction formula [27], it is appropriate to consider correspondingly the functions $H^+_{p_2}(k) \doteq (k^2 - m_1^2)\tilde{R}^{(c)}_{p_2}(k)$ and $H^-_{p_2}(k) \doteq (k^2 - m_1^2)\tilde{A}^{(c)}_{p_2}(k)$, which are also, respectively, holomorphic in $\mathcal{T}^+$ and $\mathcal{T}^-$. Then the forward scattering amplitude $F_0(s) = F(s, 0) = T(p_1, p_2; p_1, p_2)$ appears as the restriction to the hyperboloid sheet $p \in H^+_{m_1}$ of the boundary value $H^+_{p_2}(p)$ of $H^+_{p_2}(k)$ on the reals.

Moreover, it can be seen that the two boundary values $H^+_{p_2}(p) = (p^2 - m_1^2)\tilde{R}_{p_2}(p)$ and $H^-_{p_2}(p) = (p^2 - m_1^2)\tilde{A}_{p_2}(p)$ coincide as distributions in the region

$$\begin{aligned}\mathcal{R} = \{p \in \mathbf{R}^4; (p + p_2)^2 < (m_1 + m_2)^2;\\ (p - p_2)^2 < (m_1 + m_2)^2\}\end{aligned} \qquad [29]$$

This follows from the intermediate expression in eqn [26] and from the fact that a state of the form $(p^2 - m_1^2)\Phi(\pm p)p_2 >$ is a state of energy–momentum $\pm p + p_2$ and therefore vanishes (in view of the spectral condition) if $(\pm p + p_2)^2 < (m_1 + m_2)^2$ (here we also use a simplifying assumption according to which no one-particle bound state is present in this channel).

The situation obtained concerning the holomorphic functions $H^+_{p_2}(k)$ and $H^-_{p_2}(k)$ parallels (in complex dimension four) the case of a pair of holomorphic functions in the upper and lower half-planes whose boundary values on the reals coincide on a certain interval playing the role of $\mathcal{R}$. As in this one-dimensional case there is a theorem, called the "edge-of-the-wedge theorem" (see below), which implies that $H^+_{p_2}(k)$ and $H^-_{p_2}(k)$ have a common analytic continuation $H_{p_2}(k)$: this function is holomorphic in a domain $D$ which is the union of $\mathcal{T}^+, \mathcal{T}^-$ and of a complex neighborhood of $\mathcal{R}$; $D$ is called the primitive domain of $H_{p_2}(k)$.

Moreover, it follows from the postulate of invariance of the field $\Phi(x)$ under the action of the Poincaré group (see postulate (2)) that the holomorphic function $H_{p_2}(k)$ only depends of the two complex variables $\zeta = k^2 (= k_0^2 - \boldsymbol{k}^2)$ and $k \cdot p_2$ or equivalently $s = (k + p_2)^2 = \zeta + m_2^2 + 2k \cdot p_2$; it thus defines a corresponding holomorphic function $\hat{H}_{p_2}(\zeta, s) \doteq H_{p_2}(k)$ in the image of $D$ in these variables.

In view of the reduction formula [27], the scattering function $\hat{F}_0(s)$ should appear as the restriction of the holomorphic function $\hat{H}_{p_2}(\zeta, s)$ to the physical mass-shell value $\zeta = m_1^2$. However, it turns out that the section of $D$ by the complex mass-shell manifold $\mathcal{M}^{(c)}$ with equation $k^2 = m_1^2$ is empty: this geometrical fact is responsible for the nontriviality of the proof of dispersion relations for the physical quantity $\hat{F}_0(s)$ on the mass shell. In fact, the tube $T^+ \cup T^-$ which constitutes the basic part of the domain $D$ and is given by the field-theoretical microcausality postulate, is a "purely off-shell" complex domain, as it can be easily checked: if a complex point $k = p + \mathrm{i}q$ is such that $q^2 > 0$, the corresponding squared mass $\zeta = k^2 = p^2 - q^2 + 2\mathrm{i}p \cdot q$ is real if and only if $p \cdot q = 0$, which implies $p^2 < 0$ (i.e., $p$ spacelike) and therefore $\zeta = p^2 - q^2 < 0$.

**"Off-shell dispersion relations" as a first step** The starting point, which is easy to obtain from the domain $D$, is the analyticity of the holomorphic function $\hat{H}_{p_2}(\zeta, s)$ in a cut-plane of the variable $s$ for all negative values of the squared mass variable $\zeta$. This cut-plane $\Delta_\zeta$ is always the complement in $\mathbf{C}$ (i.e., the complex $s$-plane) of the union of the $s$-cut ($s$ real $\geq (m_1 + m_2)^2$) and of the $u$-cut ($u = 2\zeta + 2m_2^2 - s$ real $\geq (m_1 + m_2)^2$). This analyticity property thus justifies "off-shell dispersion relations" at fixed negative values of $\zeta$ for the field-theoretical structure function $\hat{H}_{p_2}(\zeta, s)$.

The latter property and the subsequent analysis concerning the process of analytic continuation of $\hat{H}_{p_2}$ to positive values of $\zeta$ will be more easily understood geometrically if one reduces the complex space of $k$ to a two-dimensional complex space, which is legitimate in view of the equality $H_{p_2}(k) = \hat{H}_{p_2}(\zeta, s)$.

Having chosen the $k_0$-axis along $p_2$, we reduce the orthogonal space coordinates $\boldsymbol{k}$ of $k$ to the radial variable $k_r$. One thus gets the following expressions of the variables $\zeta$ and $s$ (resp. $u$):

$$\zeta = k_0^2 - k_r^2, \quad s = \zeta + m_2^2 + 2m_2 k_0$$
$$(\text{resp. } u = \zeta + m_2^2 - 2m_2 k_0)$$

Then we can write $\hat{H}_{p_2}(\zeta, s) \doteq \mathrm{H}_{p_2}(k_0, k_r) = \mathrm{H}_{p_2}(k_0, -k_r)$, and describe the image $D_r$ of the domain $D$ in the variables $k = (k_0, k_r) = p + \mathrm{i}q$ as $T_r^+ \cup T_r^- \cup \mathcal{N}(\mathcal{R}_r)$, where:

1. $T_r^\pm$ is defined by the condition $q^2 \doteq q_0^2 - q_r^2 > 0$, $q_0 > 0$ or $q_0 < 0$,
2. $\mathcal{N}$ is a complex neighborhood of the real region $\mathcal{R}_r$ defined as follows. Let $h_s^+, h_u^-$ be the two branches of hyperbolae with respective equations:

$$h_s^+ : \ (p_0 + m_2)^2 - p_r^2 = (m_1 + m_2)^2, \ p_0 + m_2 > 0$$
$$h_u^- : \ (p_0 - m_2)^2 - p_r^2 = (m_1 + m_2)^2, \ p_0 - m_2 < 0$$

Then $\mathcal{R}_r$ is the intersection of the region situated below $h_s^+$ and of the region situated above $h_u^-$.

Let us now consider any complex hyperbola $h^{(c)}[\zeta]$ with equation $k^2 \doteq k_0^2 - k_r^2 = \zeta$. On such a complex curve either one of the variables $k_0$ or $s$ or $u$ is a good parameter for holomorphic functions which are even in $k_r$, like $\mathrm{H}_{p_2}(k_0, k_r)$. If $\zeta$ is real, any complex point $k = p + iq$ of $h^{(c)}[\zeta]$ is such that $p^2$ and $q^2$ have opposite signs (since $p \cdot q = 0$). Therefore, the sign of $q^2$ is always opposite to the sign of $\zeta(= p^2 - q^2)$: if $\zeta$ is negative, all the complex points of $h^{(c)}[\zeta]$ thus belong to $T_r^+ \cup T_r^-$; the union of all these points with the real points of $h^{(c)}[\zeta]$ in $\mathcal{R}_r$ is therefore a subset of $D_r$, which is represented in the complex plane of $s$ by the cut-plane $\Delta_\zeta$. The function $\hat{H}_{p_2}(\zeta, s)$ is therefore analytic (and univalent) in $\Delta_\zeta$ for each $\zeta < 0$. Moreover, the existence of moderate bounds of type [12] on $H_{p_2}$ in $D$ (resulting from the temperateness assumption) then implies the validity of dispersion relations (with subtractions) for $\hat{H}_{p_2}(\zeta, s)$ in $\Delta_\zeta$.

**The problem of analytic completion to the complex mass-shell hyperbola $h^{(c)}[m_1^2]$: what is provided by the Jost–Lehmann–Dyson domain** A basic fact in complex geometry in $n$ variables, with $n \geq 2$, is the existence of a distinguished class of domains, called holomorphy domains: for each domain $\mathcal{U}$ in this class, there exists at least one function which is holomorphic in $\mathcal{U}$ and cannot be analytically continued at any point of the boundary of $\mathcal{U}$. In one dimension, every domain is a holomorphy domain. In dimension larger than one, a general domain $\mathcal{U}$ is not a holomorphy domain, but it admits a holomorphy envelope $\hat{\mathcal{U}}$, which is a holomorphy domain containing $\mathcal{U}$, such that every function holomorphic in $\mathcal{U}$ admits an analytic continuation in $\hat{\mathcal{U}}$.

It turns out that the domain $D_r$ considered above in the last subsection) is not a holomorphy domain; its holomorphy envelope $\hat{D}_r$ (obtained geometrically by Bros, Messiah, and Stora in 1961) coincides with a domain introduced by Jost–Lehmann (1957) and Dyson (1958) by methods of wave equations. This domain can be characterized as the union of $D_r$ with all the complex points of all the hyperbolae with equations $(k_0 - a)^2 - (k_r - b)^2 = c^2$ (for all $a, b, c$ real, including the complex straight lines for which $c = 0$) whose both branches have a nonempty intersection with the real region $\mathcal{R}_r$.

In particular, one easily sees that all the hyperbolae $h^{(c)}[\zeta]$ with $0 \leq \zeta < m_1^2$ belong to the previous class. It follows that for any $\zeta$ in this positive interval, the function $\hat{H}_{p_2}(\zeta, s)$ can still be analytically continued as a holomorphic function of $s$ in the cut-plane $\Delta_\zeta$ and thereby satisfies the corresponding dispersion relations.

The physical mass shell hyperbola $h^{(c)}[m_1^2]$ thus appears as a limiting case of the previous family (for $\zeta$ tending to $m_1^2$ from below). The analyticity of $\hat{H}_{p_2}(m_1^2, s)$ in $\Delta_{m_1^2}$ can then be justified provided one knows that this function is analytic at at least one point of $\Delta_{m_1^2}$: but this additional information results from a more thorough exploitation of the analyticity properties resulting from the QFT postulates. This will be now briefly outlined below.

**Further information coming from the four-point function in complex momentum space** It is possible to obtain further analyticity properties of $\hat{H}_{p_2}(\zeta, s) \doteq H_{p_2}(k)$ by considering the latter as the restriction to the submanifold $k_1 = -k_3 = k$; $k_2 = -k_4 = p_2$ of a master analytic function $H_4(k_1, k_2, k_3, k_4)$, called the four-point function of the field $\Phi$ in complex energy–momentum space (*see* Scattering in Relativistic Quantum Field Theory: The Analytic Program). This function is holomorphic in a well-defined primitive domain $D_4$ of the linear submanifold $k_1 + k_2 + k_3 + k_4 = 0$. It is then possible to compute some local parts situated near the reals of the holomorphy envelope of $D_4$, which implies, as a by-product, that the function $\hat{H}_{p_2}(\zeta, s)$ can be analytically continued in a set $\Sigma$ of the form

$$\Sigma = \{(\zeta, s);\ \zeta \in \delta,\ s \in \mathcal{V}_{s_1}(\zeta)\} \cup \{(\zeta, s);\ \zeta \in \delta,\ u = 2\zeta + 2m_2^2 - s \in \mathcal{V}_{u_1}(\zeta)\} \quad [30]$$

with the following specifications:

1. $\delta$ is a domain in the $\zeta$-plane, which is a complex neighborhood of a real interval of the form $-a < \zeta < M_1^2$; here $M_1$ denotes a spectral mass threshold in the theory such that $M_1 > m_1$;
2. for each $\zeta, \mathcal{V}_{s_1}(\zeta)$ (resp. $\mathcal{V}_{u_1}(\zeta)$) is a cut-neighborhood in the $s$-plane of the real half-line $s > s_1$ (resp. of the half-line $u = 2\zeta + 2m_2^2 - s$ real $> u_1$); $s_1$ and $u_1$ denote appropriate real numbers independent of $\zeta$.

**The final analytic completion: crossing domains on $h^{(c)}[m_1^2]$. Dispersion relations for $\pi_0$–$\pi_0$ meson scattering and "quasi-dispersion-relations" for proton–proton scattering** We now wish to describe briefly the final step of analytic completion, which displays the existence of a "quasi-cut-plane domain" in $s$ for the function $\hat{H}_{p_2}(m_1^2, s)$, even in the more general case when the $s$-cut and $u$-cut are associated with different scattering channels, whose respective mass thresholds $s = M_{12}^2$ and $u = M_{12}'^2$ are unequal.

This general situation may occur as soon as one charged particle $\Pi_1$ of the $s$-channel is replaced by the corresponding antiparticle $\overline{\Pi}_1$ in the $u$-channel, in contrast with the case of neutral particles (like the $\pi_0$ meson) which coincide with their own antiparticles. Here it is important to note that the two real branches $h^+[m_1^2]$ and $h^-[m_1^2]$ of the mass shell hyperbola $h^{(c)}[m_1^2]$ correspond, respectively, to the physical region of the "direct scattering channel" of the reaction $\Pi_1 + \Pi_2 \to \Pi_1 + \Pi_2$ with squared total energy $s$, and to the physical region of the "crossed scattering channel" of the reaction $\overline{\Pi}_1 + \Pi_2 \to \overline{\Pi}_1 + \Pi_2$ with squared total energy $u$. A typical and important example is the case of proton–proton scattering in the $s$-channel, where $M_{12}$ equals twice the mass $m(=m_1=m_2)$ of the proton, while the corresponding $u$-channel refers to the proton–antiproton scattering, whose threshold $M'_{12}$ equals twice the mass $\mu$ of the $\pi$ meson.

In that general case, the analysis of the subsection "'Off-shell dispersion relations' as a first step" still applies, so that the function $\hat{H}_{p_2}(\zeta, s)$ is always analytic in a set of the form

$$S_0 = \{(\zeta, s);\ -a < \zeta < 0,\ s \in \Delta_\zeta\} \qquad [31]$$

Then, the additional information described above in the last subsection allows one to use the following crucial property of analytic completion, which we call

**Crossing lemma** *If a function $G(\zeta, s)$ is holomorphic in a domain which contains the union of the sets $\Sigma$ and $S_0$ (see eqns [30] and [31]), then it admits an analytic continuation in a set of the following form*:

$$\{(\zeta, s);\ \zeta \in \delta,\ s \in \Delta_\zeta;\ |s - \zeta - m_2^2| = |u - \zeta - m_2^2| > R(\zeta)\}$$

By applying this property to the function $\hat{H}_{p_2}(\zeta, s)$ and restricting $\zeta$ to the mass-shell value $m_1^2$ which belongs to $\delta$, one obtains the analyticity of the scattering function $\hat{F}_0(s) \doteq \hat{H}_{p_2}(m_1^2, s)$ in a crossing domain of the complex mass shell hyperbola $h^{(c)}[m_1^2]$: the crossing between the two physical regions $h^+[m_1^2]$ $(s \geq M_{12}^2)$ and $h^-[m_1^2]$ $(u \geq M'^2_{12})$ is ensured by a complex domain of $h^{(c)}[m_1^2]$ whose image in the $s$-plane is the "cut-neighborhood of infinity" $\{s; s \in \Delta_{m_1^2}, |s - m_1^2 - m_2^2| = |u - m_1^2 - m_2^2| > R(m_1^2)\}$. Note that the relevant boundary values of $\hat{F}_0$ for obtaining the scattering amplitudes of the two collision processes with respective physical regions $h^+[m_1^2]$ and $h^-[m_1^2]$ have to be taken from the respective sides $\operatorname{Im} s > 0$ and $\operatorname{Im} u = -\operatorname{Im} s > 0$ of the corresponding $s$- and $u$-cuts.

It is only for the neutral case, where $M_{12} = M'_{12} = m_1 + m_2$, that a more favorable scenario occurs, as explained earlier: in this case, the interval $\{\zeta \in ]-a, 0[\}$ of the set $S_0$ is replaced by $\{\zeta \in ]-a, m_1^2[\}$, so that the whole cut-plane domain $\Delta_{m_1^2}$ is obtained in the result of the previous crossing lemma. The scattering amplitudes of $\pi_0$–$\pi_0$ meson scattering and of $\pi$ meson–proton scattering enjoy this property and, therefore, satisfy genuine dispersion relations in which the scattering function is even (see the second basic example described at the beginning of this article). In the general case of crossing domains obtained above, corresponding Cauchy integral relations have been written and used under the name of "quasi-dispersion-relations."

**Complementary results** Some comments can now be added concerning the passage from the purely geometrical results (i.e., analyticity domains) described above to the writing of precise (quasi-) dispersion relations with two subtractions:

*Polynomial bounds and dispersion relations with N subtractions* The previous methods of analytic completion also allow one to control the bounds at infinity in the relevant complex domains. As it has been noticed after eqn [24], the Fourier–Laplace transforms of the retarded and advanced kernels, and thereby the holomorphic functions $H^{\pm}_{p_2}(k)$ discussed at the start of this section are bounded at most by a power of a suitable norm of $k$ in their respective tubes $\mathcal{T}^{\pm}$. Correspondingly, the holomorphic function $H_{p_2}(k)$ (resp. $\mathrm{H}_{p_2}(k_0, k_r)$) admits the same type of bound in its primitive analyticity domain $D$ (resp. $D_r$). These bounds are a consequence of the tempered distribution character of the structure functions of the fields which is built-in in the Wightman field-theoretical framework. Then it can be checked that in the holomorphy envelope $\hat{D}_r$ of $D_r$, and thereby in the cut-plane (or crossing) domains obtained in the intersection of $\hat{D}_r$ and of the complex mass shell $h^{(c)}[m_1^2]$, the same type of power bound is still valid: $\hat{F}_0(s)$ is therefore bounded by some power $|s|^{N-1}$ of $|s|$ and thus satisfies a (quasi-)dispersion relation with $N$ subtractions. The same type of argument holds for all the similar cut-domains (or crossing domains) in $s$ obtained for $\hat{F}_t(s)$ for all negative value of $t$.

It is also worthwhile to mention that a similar remarkable (since not at all predictable) result was also obtained in the Haag, Kastler, and Araki framework of algebraic QFT (Epstein, Glaser, Martin, 1969; *see* Scattering in Relativistic Quantum Field Theory: The Analytic Program for further comments).

In this connection, one can also mention a more recent result. In the Buchholz–Fredenhagen axiomatic approach of charged fields (1982), in which

locality is replaced by the more general notion of "stringlike locality" (*see* Algebraic Approach to Quantum Field Theory, Axiomatic Quantum Field Theory, and Scattering in Relativistic Quantum Field Theory: Fundamental Concepts and Tools), a proof of forward dispersion relations has again been obtained (Bros, Epstein, 1994).

*The extension of the analyticity domains by positivity and the derivation of bounds by unitarity* (Martin 1966; see the book by Martin (1969)). The following ingredients have been used:

1. Positivity conditions on the absorptive part of $F(s,t)$, which are expressed by the infinite set of inequalities $(\mathrm{d}/\mathrm{d}t)^n \operatorname{Im} F(s,t)_{|t=0} \geq 0$ (for all integers $n$),
2. The existence of a two-dimensional complex neighborhood of some point $(s=s_0, t=0)$ in the analyticity domain resulting from QFT.

The following results have then been obtained:

(a) It is justified to differentiate the forward (subtracted) dispersion relations with respect to $t$ at any order.
(b) $\hat{F}(s,t)$ can be analytically continued in a fixed circle $|t| < t_{\max}$ for all values of $s$. The latter implies the extension of dispersion relations in $s$ to positive (and complex) values of $t$.
(c) In a last step, the use of unitarity conditions for the "partial waves" $f_\ell(s)$ of $F(s,t)$ (*see* Scattering in Relativistic Quantum Field Theory: The Analytic Program) allows one to obtain Froissart-type bounds on the scattering amplitudes and thereby to justify the writing of (quasi-)dispersion relations with at most two subtractions for all the admissible values of $t$.

*See also:* Algebraic Approach to Quantum Field Theory; Axiomatic Quantum Field Theory; Perturbation Theory and its Techniques; Scattering in Relativistic Quantum Field Theory: The Analytic Program; Scattering, Asymptotic Completeness and Bound States; Scattering in Relativistic Quantum Field Theory: Fundamental Concepts and Tools.

## Further Reading


Haag R (1992) *Local Quantum Physics*.Berlin: Springer.

Iagolnitzer D (1992) *Scattering in Quantum Field Theories: The Axiomatic and Constructive Approaches*. Princeton Series in Physics. Princeton: Princeton University Press.

Jost R (1965) *The General Theory of Quantized Fields*. AMS, Providence, RI: American Mathematical Society.

Klein L (ed.) (1961) *Dispersion Relations and the Abstract Approach to Field Theory*. New York: Gordon and Breach.

Martin A (1969) *Scattering Theory: Unitarity, Analyticity and Crossing*, Lecture Notes in Physics. Berlin: Springer.

Nussenzweig HM (1972) *Causality and Dispersion Relations*. New York: Academic Press.

Streater RF and Wightman AS (1964, 1980) *PCT, Spin and Statistics, and all that*. Princeton: Princeton University Press.

Vernov YuS (1996) *Dispersion Relations in the Historical Aspect*, IHEP Publications, Protvino Conf.: Fundamental Problems of High Energy Physics and Field Theory.


# Dissipative Dynamical Systems of Infinite Dimension


**M Efendiev and S Zelik**, Universität Stuttgart, Stuttgart, Germany

**A Miranville**, Université de Poitiers, Chasseneuil, France




## Introduction

A dynamical system (DS) is a system which evolves with respect to the time. To be more precise, a DS $(S(t), \Phi)$ is determined by a phase space $\Phi$ which consists of all possible values of the parameters describing the state of the system and an evolution map $S(t): \Phi \to \Phi$ that allows one to find the state of the system at time $t > 0$ if the initial state at $t=0$ is known. Very often, in mechanics and physics, the evolution of the system is governed by systems of differential equations. If the system is described by ordinary differential equations (ODEs),

$$\frac{\mathrm{d}}{\mathrm{d}t} y(t) = F(t, y(t)), \quad y(0) = y_0, \\ y(t) := (y_1(t), \ldots, y_N(t)) \qquad [1]$$

for some nonlinear function $F: \mathbb{R}_+ \times \mathbb{R}^N \to \mathbb{R}^N$, we have a so-called finite-dimensional DS. In that case, the phase space $\Phi$ is some (invariant) subset of $\mathbb{R}^N$ and the evolution operator $S(t)$ is defined by

$$S(t)y_0 := y(t), \quad y(t) \text{ solves } [1] \qquad [2]$$

We also recall that, in the case where eqn [1] is autonomous (i.e., does not depend explicitly on the time), the evolution operators $S(t)$ generate a semigroup on the phase space $\Phi$, that is,

$$S(t_1 + t_2) = S(t_1) \circ S(t_2), \quad t_1, t_2 \in \mathbb{R}_+ \qquad [3]$$

Now, in the case of a distributed system whose initial state is described by functions $u_0 = u_0(x)$ depending on the spatial variable $x$, the evolution is usually governed by partial differential equations (PDEs) and the corresponding phase space $\Phi$ is some infinite-dimensional function space (e.g., $\Phi := L^2(\Omega)$ or $\Phi := L^\infty(\Omega)$ for some domain $\Omega \subset \mathbb{R}^N$.) Such DSs are usually called infinite dimensional.

The qualitative study of DSs of finite dimensions goes back to the beginning of the twentieth century, with the pioneering works of Poincaré on the $N$-body problem (one should also acknowledge the contributions of Lyapunov on the stability and of Birkhoff on the minimal sets and the ergodic theorem). One of the most surprising and significant facts discovered at the very beginning of the theory is that even relatively simple equations can generate very complicated chaotic behaviors. Moreover, these types of systems are extremely sensitive to initial conditions (the trajectories with close but different initial data diverge exponentially). Thus, in spite of the deterministic nature of the system (we recall that it is generated by a system of ODEs, for which we usually have the unique solvability theorem), its temporal evolution is unpredictable on timescales larger than some critical time $T_0$ (which depends obviously on the error of approximation and on the rate of divergence of close trajectories) and can show typical stochastic behaviors. To the best of our knowledge, one of the first ODEs for which such types of behaviors were established is the physical pendulum parametrically perturbed by time-periodic external forces,

$$y''(t) + \sin(y(t))(1 + \varepsilon \sin(\omega t)) = 0 \qquad [4]$$

where $\omega$ and $\varepsilon > 0$ are physical parameters. We also mention the more recent (and more relevant for our topic) famous example of the Lorenz system which is defined by the following system of ODEs in $\mathbb{R}^3$:

$$\begin{cases} x' = \sigma(y - x) \\ y' = -xy + rx - y \\ z' = xy - bz \end{cases} \qquad [5]$$

where $\sigma, r,$ and $b$ are some parameters. These equations are obtained by truncation of the Navier–Stokes equations and give an approximate description of a horizontal fluid layer heated from below. The warmer fluid formed at the bottom tends to rise, creating convection currents. This is similar to what happens in the Earth's atmosphere. For a sufficiently intense heating, the time evolution has a sensitive dependence on the initial conditions, thus representing a very irregular and chaotic convection. This fact was used by Lorenz to justify the so-called "butterfly effect," a metaphor for the imprecision of weather forecast.

The theory of DSs in finite dimensions had been extensively developed during the twentieth century, due to the efforts of many famous mathematicians (such as Anosov, Arnold, LaSalle, Sinai, Smale, etc.) and, nowadays, much is known on the chaotic behaviors in such systems, at least in low dimensions. In particular, it is known that, very often, the trajectories of a chaotic system are localized, up to a transient process, in some subset of the phase space having a very complicated fractal geometric structure (e.g., locally homeomorphic to the Cartesian product of $R^m$ and some Cantor set) which, thus, accumulates the nontrivial dynamics of the system (the so-called strange attractor). The chaotic dynamics on such sets are usually described by symbolic dynamics generated by Bernoulli shifts on the space of sequences. We also note that, nowadays, a mathematician has a large amount of different concepts and methods for the extensive study of concrete chaotic DSs in finite dimensions. In particular, we mention here different types of bifurcation theories (including the KAM theory and the homoclinic bifurcation theory with related Shilnikov chaos), the theory of hyperbolic sets, stochastic description of deterministic processes, Lyapunov exponents and entropy theory, dynamical analysis of time series, etc.

We now turn to infinite-dimensional DSs generated by PDEs. A first important difficulty which arises here is related to the fact that the analytic structure of a PDE is essentially more complicated than that of an ODE and, in particular, we do not have in general the unique solvability theorem as for ODEs, so that even finding the proper phase space and the rigorous construction of the associated DS can be a highly nontrivial problem. In order to indicate the level of difficulties arising here, it suffices to recall that, for the three-dimensional Navier–Stokes system (which is one of the most important equations of mathematical physics), the required associated DS has not been constructed yet. Nevertheless, there exists a large number of equations for which the problem of the global existence and uniqueness of a solution has been solved. Thus, the question of extending the highly developed finite-dimensional DS theory to infinite dimensions arises naturally.

One of the first and most significant results in that direction was the development of the theory of integrable Hamiltonian systems in infinite dimensions and the explicit resolution (by inverse-scattering methods) of several important conservative equations

of mathematical physics (such as the Korteweg–de Vries (and the generalized Kadomtsev–Petiashvilli hierarchy), the sine-Gordon, and the nonlinear Schrödinger equations). Nevertheless, it is worth noting that integrability is a very rare phenomenon, even among ODEs, and this theory is clearly insufficient to understand the dynamics arising in PDEs. In particular, there exist many important equations which are essentially out of reach of this theory.

One of the most important classes of such equations consists of the so-called dissipative PDEs which are the main subject of our study. As hinted by this denomination, these systems exhibit some energy dissipation process (in contrast to conservative systems for which the energy is preserved) and, of course, in order to have nontrivial dynamics, these models should also account for the energy income. Roughly speaking, the complicated chaotic behaviors in such systems usually arise from the interaction of the following mechanisms:

1. energy dissipation in the higher part of the Fourier spectrum;
2. external energy income in its lower part;
3. energy flux from lower to higher Fourier modes provided by the nonlinear terms of the equation.

We chose not to give a rigorous definition of a dissipative system here (although the concepts of energy dissipation and related dissipative systems are more or less obvious from the physical point of view, they seem too general to have an adequate mathematical definition). Instead, we only indicate several basic classes of equations of mathematical physics which usually exhibit the above behaviors.

The first example is, of course, the Navier–Stokes system, which describes the motion of a viscous incompressible fluid in a bounded domain $\Omega$ (we will only consider here the two-dimensional case $\Omega \subset \mathbb{R}^2$, since the adequate formulation in three dimensions is still an open problem):

$$\begin{cases} \partial_t u - (u, \nabla_x)u = \nu\Delta_x u + \nabla_x p + g(x) \\ \operatorname{div} u = 0, u|_{t=0} = u_0, u|_{\partial\Omega} = 0 \end{cases} \qquad [6]$$

Here, $u(t,x) = (u_1(t,x), u_2(t,x))$ is the unknown velocity vector, $p = p(t,x)$ is the unknown pressure, $\Delta_x$ is the Laplacian with respect to $x$, $\nu > 0$ and $g$ are given kinematic viscosity and external forces, respectively, and $(u, \nabla_x)u$ is the inertial term ($[(u,\nabla_x)u]_i = \sum_{j=1}^{2} u_j \partial_{x_j} u_i$, $i = 1, 2$). The unique global solvability of [6] has been proved by Ladyzhenskaya. Thus, this equation generates an infinite-dimensional DS in the phase space $\Phi$ of divergence-free square-integrable vector fields.

The second example is the damped nonlinear wave equation in $\Omega \subset \mathbb{R}^n$:

$$\begin{gathered} \partial_t^2 u + \gamma\partial_t u - \Delta_x u + f(u) = 0 \\ u|_{\partial\Omega} = 0, \quad u|_{t=0} = u_0, \quad \partial_t u|_{t=0} = u_0' \end{gathered} \qquad [7]$$

which models, for example, the dynamics of a Josephson junction driven by a current source (sine-Gordon equation). It is known that, under natural sign and growth assumptions on the nonlinear interaction function $f$, this equation generates a DS in the energy phase space $E$ of pairs of functions $(u, \partial_t u)$ such that $\partial_t u$ and $\nabla_x u$ are square integrable.

The last class of equations that we will consider here consists of reaction–diffusion systems in a domain $\Omega \subset \mathbb{R}^n$:

$$\partial_t u = a\Delta_x u - f(u), \quad u|_{t=0} = u_0 \qquad [8]$$

(endowed with Dirichlet ($u|_{\partial\Omega} = 0$) or Neumann ($\partial_n u|_{\partial\Omega} = 0$) boundary conditions), which describes some chemical reaction in $\Omega$. Here, $u = (u^1, \ldots, u^N)$ is an unknown vector-valued function which describes the concentrations of the reactants, $f(u)$ is a given interaction function, and $a$ is a diffusion matrix. It is known that, under natural assumptions on $f$ and $a$, these equations also generate an infinite-dimensional DS, for example, in the phase space $\Phi := [L^\infty(\Omega)]^n$.

We emphasize once more that the phase spaces $\Phi$ in all these examples are appropriate infinite-dimensional function spaces. Nevertheless, it was observed in experiments that, up to a transient process, the trajectories of the DS considered are localized inside a "very thin" invariant subset of the phase space having a complicated geometric structure which, thus, accumulates all the nontrivial dynamics of the system. It was conjectured a little later that these invariant sets are, in some proper sense, finite dimensional and that the dynamics restricted to these sets can be effectively described by a finite number of parameters. Thus (when this conjecture is true), in spite of the infinite-dimensional initial phase space, the effective dynamics (reduced to this invariant set) is finite dimensional and can be studied by using the algorithms and concepts of the classical finite-dimensional DS theory. In particular, this means that the infinite dimensionality plays here only the role of (possibly essential) technical difficulties, which cannot, however, produce any new dynamical phenomena which are not observed in the finite-dimensional theory.

The above finite-dimensional reduction principle of dissipative PDEs in bounded domains has been given solid mathematical grounds (based on the concept of the so-called global attractor) over the

last three decades, starting from the pioneering papers of Ladyzhenskaya. This theory is considered in more detail here.

The finite-dimensional reduction theory has some limitations. Of course, the first and most obvious restriction of this principle is the effective dimension of the reduced finite-dimensional DS. Indeed, it is known that, typically, this dimension grows at least linearly with respect to the volume $\mathrm{vol}(\Omega)$ of the spatial domain $\Omega$ of the DS considered (and the growth of the size of $\Omega$ is the same (up to a rescaling) as the decay of the viscosity coefficient $\nu$ or the diffusion matrix $a$, see eqns [6]–[8]). So, for sufficiently large domains $\Omega$, the reduced DS can be too large for reasonable investigations.

The next, less obvious, but much more essential, restriction is the growing spatial complexity of the DS. Indeed, as shown by Babin–Buinimovich, the spatial complexity of the system (e.g., the number of topologically different equilibria) grows exponentially with respect to $\mathrm{vol}(\Omega)$. Thus, even in the case of relatively small dimensions, the reduced system can be out of reasonable investigations, due to its extremely complicated structure.

Therefore, the approach based on the finite-dimensional reduction does not look so attractive for large domains. It seems, instead, more natural, at least from the physical point of view, to replace large bounded domains by their limit unbounded ones (e.g., $\Omega = \mathbb{R}^n$ or cylindrical domains). Of course, this approach requires a systematic study of dissipative DSs associated with PDEs in unbounded domains.

The dynamical study of PDEs in unbounded domains started from the pioneering paper of Kolmogorov–Petrovskij–Piskunov, in which the traveling wave solutions of reaction–diffusion equations in a strip were constructed and the convergence of the trajectories (for specific initial data) to this traveling wave solutions were established. Starting from this, many results on the dynamics of PDE in unbounded domains have been obtained. However, for a long period, the general features of such dynamics remained completely unclear. The main problems arising here are:

1. the essential infinite dimensionality of the DS considered (absence of any finite-dimensional reduction), which leads to essentially new dynamical effects that are not observed in finite-dimensional theories;
2. the additional spatial "unbounded" directions lead to the so-called spatial chaos and the interaction between spatial and temporal chaotic modes generates the spatio-temporal chaos, which also has no analog in finite dimensions.

Nevertheless, several ideas are mentioned in the following which (from authors' point of view) were the most important for the development of these topics. The first one is the pioneering paper of Kirchgässner, in which dynamical methods were applied to the study of the spatial structure of solutions of elliptic equations in cylinders (which can be considered as equilibria equations for evolution PDEs in unbounded cylindrical domains). The second is the Sinai–Buinimovich model of spacetime chaos in discrete lattice DSs. Finally, the third is the adaptation of the concept of a global attractor to unbounded domains by Abergel and Babin–Vishik.

We note that the situation on the understanding of the general features of the dynamics in unbounded domains, however, seems to have changed in the last several years, due to the works of Collet–Eckmann and Zelik. This is the reason why a section of this review is devoted to a more detailed discussion on this topic.

Other important questions are the object of current studies and we only briefly mention some of them. We mention for instance, the study of attractors for nonautonomous systems (i.e., systems in which the time appears explicitly). This situation is much more delicate and is not completely understood; notions of attractors for such systems have been proposed by Chepyzhov–Vishik, Haraux and Kloeden–Schmalfuss. We also mention that theories of (global) attractors for non-well-posed problems have been proposed by Babin–Vishik, Ball, Chepyzhov–Vishik, Melnik–Valero, and Sell.

## Global Attractors and Finite-Dimensional Reduction

### Global Attractors: The Abstract Setting

As already mentioned, one of the main concepts of the modern theory of DSs in infinite dimensions is that of the global attractor. We give below its definition for an abstract semigroup $S(t)$ acting on a metric space $\Phi$, although, without loss of generality, the reader may think that $(S(t), \Phi)$ is just a DS associated with one of the PDEs ([6]–[8]) described in the introduction.

To this end, we first recall that a subset $K$ of the phase space $\Phi$ is an attracting set of the semigroup $S(t)$ if it attracts the images of all the bounded subsets of $\Phi$, that is, for every bounded set $B$ and every $\varepsilon > 0$, there exists a time $T$ (depending in general on $B$ and $\varepsilon$) such that the image $S(t)B$ belongs to the

$\varepsilon$-neighborhood of $K$ if $t \geq T$. This property can be rewritten in the equivalent form

$$\lim_{t\to\infty} \mathrm{dist}_H(S(t)B, K) = 0 \qquad [9]$$

where $\mathrm{dist}_H(X, Y) := \sup_{x\in\mathbb{X}} \inf_{y\in\mathbb{Y}} d(x, y)$ is the nonsymmetric Hausdorff distance between subsets of $\Phi$.

The following definition of a global attractor is due to Babin–Vishik.

**Definition 1** A set $\mathcal{A} \subset \Phi$ is a global attractor for the semigroup $S(t)$ if

(i) $\mathcal{A}$ is compact in $\Phi$;
(ii) $\mathcal{A}$ is strictly invariant: $S(t)\mathcal{A} = \mathcal{A}$, for all $t \geq 0$;
(iii) $\mathcal{A}$ is an attracting set for the semigroup $S(t)$.

Thus, the second and third properties guarantee that a global attractor, if it exists, is unique and that the DS reduced to the attractor contains all the nontrivial dynamics of the initial system. Furthermore, the first property indicates that the reduced phase space $\mathcal{A}$ is indeed "thinner" than the initial phase space $\Phi$ (we recall that, in infinite dimensions, a compact set cannot contain, e.g., balls and should thus be nowhere dense).

In most applications, one can use the following attractor's existence theorem.

**Theorem 1** *Let a DS $(S(t), \Phi)$ possess a compact attracting set and the operators $S(t): \Phi \to \Phi$ be continuous for every fixed t. Then, this system possesses the global attractor $\mathcal{A}$ which is generated by all the trajectories of $S(t)$ which are defined for all $t \in \mathbb{R}$ and are globally bounded.*

The strategy for applying this theorem to concrete equations of mathematical physics is the following. In a first step, one verifies a so-called dissipative estimate which has usually the form

$$\|S(t)u_0\|_\Phi \leq Q(\|u_0\|_\Phi)\,\mathrm{e}^{-\alpha t} + C_*, \quad u_0 \in \Phi \qquad [10]$$

where $\|\cdot\|_\Phi$ is a norm in the function space $\Phi$ and the positive constants $\alpha$ and $C_*$ and the monotonic function $Q$ are independent of $t$ and $u_0 \in \Phi$ (usually, this estimate follows from energy estimates and is sometimes even used in order to "define" a dissipative system). This estimate obviously gives the existence of an attracting set for $S(t)$ (e.g., the ball of radius $2C_*$ in $\Phi$), which is, however, noncompact in $\Phi$. In order to overcome this problem, one usually derives, in a second step, a smoothing property for the solutions, which can be formulated as follows:

$$\|S(1)u_0\|_{\Phi_1} \leq Q_1(\|u_0\|_\Phi), \quad u_0 \in \Phi \qquad [11]$$

where $\Phi_1$ is another function space which is compactly embedded into $\Phi$. In applications, $\Phi$ is usually the space $L^2(\Omega)$ of square integrable functions, $\Phi_1$ is the Sobolev space $H^1(\Omega)$ of the functions $u$ such that $u$ and $\nabla_x u$ belong to $L^2(\Omega)$ and estimate [11] is a classical smoothing property for solutions of parabolic equations (for hyperbolic equations, a slightly more complicated asymptotic smoothing property should be used instead of [11]).

Since the continuity of the operators $S(t)$ usually arises no difficulty (if the uniqueness is proven), then the above scheme gives indeed the existence of the global attractor for most of the PDEs of mathematical physics in bounded domains.

### Dimension of the Global Attractor

In this subsection, we start by discussing one of the basic questions of the theory: in which sense is the dynamics on the global attractor finite dimensional? As already mentioned, the global attractor is usually not a manifold, but has a rather complicated geometric structure. So, it is natural to use the definitions of dimensions adopted for the study of fractal sets here. We restrict ourselves to the so-called fractal (or box-counting, entropy) dimension, although other dimensions (e.g., Hausdorff, Lyapunov, etc.) are also used in the theory of attractors.

In order to define the fractal dimension, we first recall the concept of Kolmogorov's $\varepsilon$-entropy, which comes from the information theory and plays a fundamental role in the theory of DSs in unbounded domains considered in the next section.

**Definition 2** Let $\mathcal{A}$ be a compact subset of a metric space $\Phi$. For every $\varepsilon > 0$, we define $N_\varepsilon(K)$ as the minimal number of $\varepsilon$-balls which are necessary to cover $\mathcal{A}$. Then, Kolmogorov's $\varepsilon$-entropy $\mathcal{H}_\varepsilon(\mathcal{A}) = \mathcal{H}_\varepsilon(\mathcal{A}, \Phi)$ of $\mathcal{A}$ is the digital logarithm of this number, $\mathcal{H}_\varepsilon(\mathcal{A}) := \log_2 N_\varepsilon(\mathcal{A})$. We recall that $\mathcal{H}_\varepsilon(\mathcal{A})$ is finite for every $\varepsilon > 0$, due to the Hausdorff criterium. The fractal dimension $d_\mathrm{f}(\mathcal{A}) \in [0, \infty]$ of $\mathcal{A}$ is then defined by

$$d_\mathrm{f}(\mathcal{A}) := \limsup_{\varepsilon\to 0} \mathcal{H}_\varepsilon(\mathcal{A}) / \log_2 1/\varepsilon \qquad [12]$$

We also recall that, although this dimension coincides with the usual dimension of the manifold for Lipschitz manifolds, it can be noninteger for more complicated sets. For instance, the fractal dimension of the standard ternary Cantor set in [0, 1] is $\ln 2 / \ln 3$.

The so-called Mané theorem (which can be considered as a generalization of the classical Yitni embedding theorem for fractal sets) plays an important role in the finite-dimensional reduction theory.

**Theorem 2** *Let $\Phi$ be a Banach space and $\mathcal{A}$ be a compact set such that $d_f(\mathcal{A}) < N$ for some $N \in \mathbb{N}$. Then, for "almost all" $(2N+1)$-dimensional planes $L$ in $\Phi$, the corresponding projector $\Pi_L : \Phi \to L$ restricted to the set $\mathcal{A}$ is a Hölder continuous homeomorphism.*

Thus, if the finite fractal dimensionality of the attractor is established, then, fixing a hyperplane $L$ satisfying the assumptions of the Mané theorem and projecting the attractor $\mathcal{A}$ and the DS $S(t)$ restricted to $\mathcal{A}$ onto this hyperplane ($\bar{\mathcal{A}} := \Pi_L \mathcal{A}$ and $\bar{S}(t) := \Pi_L \circ S(t) \circ \Pi_L^{-1}$), we obtain, indeed, a reduced DS $(\bar{S}(t), \bar{\mathcal{A}})$ which is defined on a finite-dimensional set $\bar{\mathcal{A}} \subset L \sim \mathbb{R}^{2N+1}$. Moreover, this DS will be Hölder continuous with respect to the initial data.

### Estimates on the Fractal Dimension

Obviously, good estimates on the dimension of the attractors in terms of the physical parameters are crucial for the finite-dimensional reduction described above, and (consequently) there exists a highly developed machinery for obtaining such estimates. The best-known upper estimates are usually obtained by the so-called volume contraction method, which is based on the study of the evolution of infinitesimal $k$-dimensional volumes in the neighborhood of the attractor (and, if the DS considered contracts the $k$-dimensional volumes, then the fractal dimension of the attractor is less than $k$). Lower bounds on the dimension are usually based on the observation that the global attractor always contains the unstable manifolds of the (hyperbolic) equilibria. Thus, the instability index of a properly constructed equilibrium gives a lower bound on the dimension of the attractor.

In the following, several estimates for the classes of equations given in the introduction are formulated, beginning with the most-studied case of the reaction–diffusion system [8]. For this system, sharp upper and lower bounds are known, namely

$$C_1 \mathrm{vol}(\Omega) \le d_f(\mathcal{A}) \le C_2 \mathrm{vol}(\Omega) \qquad [13]$$

where the constants $C_1$ and $C_2$ depend on $a$ and $f$ (and, possibly, on the shape of $\Omega$), but are independent of its size. The same types of estimates also hold for the hyperbolic equation [7]. Concerning the Navier–Stokes system [6] in general two-dimensional domains $\Omega$, the asymptotics of the fractal dimension as $\nu \to 0$ is not known. The best-known upper bound has the form $d_f(\mathcal{A}) \le C\nu^{-2}$ and was obtained by Foias–Temam by using the so-called Lieb–Thirring inequalities. Nevertheless, for periodic boundary conditions, Constantin–Foias–Temam and Liu obtained upper and lower bounds of the same order (up to a logarithmic correction):

$$C_1 \nu^{-4/3} \le d_f(\mathcal{A}) \le C_2 \nu^{-4/3}(1 + \ln(\nu^{-1}))^{1/3} \qquad [14]$$

### Global Lyapunov Functions and the Structure of Global Attractors

Although the global attractor has usually a very complicated geometric structure, there exists one exceptional class of DS for which the global attractor has a relatively simple structure which is completely understood, namely the DS having a global Lyapunov function. We recall that a continuous function $\mathcal{L} : \Phi \to \mathbb{R}$ is a global Lypanov function if

1. $\mathcal{L}$ is nonincreasing along the trajectories, that is, $\mathcal{L}(S(t)u_0) \le \mathcal{L}(u_0)$, for all $t \ge 0$;
2. $\mathcal{L}$ is strictly decreasing along all nonequilibrium solutions, that is, $\mathcal{L}(S(t)u_0) = \mathcal{L}(u_0)$ for some $t > 0$ and $u_0$ implies that $u_0$ is an equilibrium of $S(t)$.

For instance, in the scalar case $N = 1$, the reaction–diffusion equations [8] possess the global Lyapunov function $\mathcal{L}(u_0) := \int_\Omega [a|\nabla_x u_0(x)|^2 + F(u_0(x))]\mathrm{d}x$, where $F(v) := \int_0^v f(u)\,\mathrm{d}u$. Indeed, multiplying eqn [8] by $\partial_t u$ and integrating over $\Omega$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{L}(u(t)) = -2\|\partial_t u(t)\|^2_{L^2(\Omega)} \le 0 \qquad [15]$$

Analogously, in the scalar case $N = 1$, multiplying the hyperbolic equation [7] by $\partial_t u(t)$ and integrating over $\Omega$, we obtain the standard global Lyapunov function for this equation.

It is well known that, if a DS posseses a global Lyapunov function, then, at least under the generic assumption that the set $\mathcal{R}$ of equilibria is finite, every trajectory $u(t)$ stabilizes to one of these equilibria as $t \to +\infty$. Moreover, every complete bounded trajectory $u(t)$, $t \in \mathbb{R}$, belonging to the attractor is a heteroclinic orbit joining two equilibria. Thus, the global attractor $\mathcal{A}$ can be described as follows:

$$\mathcal{A} = \bigcup_{u_0 \in \mathcal{R}} \mathcal{M}^+(u_0) \qquad [16]$$

where $\mathcal{M}^+(u_0)$ is the so-called unstable set of the equilibrium $u_0$ (which is generated by all heteroclinic orbits of the DS which start from the given equilibrium $u_0 \in \mathcal{A}$). It is also known that, if the equilibrium $u_0$ is hyperbolic (generic assumption), then the set $\mathcal{M}^+(u_0)$ is a $\kappa$-dimensional submanifold of $\Phi$, where $\kappa$ is the instability index of $u_0$. Thus, under the generic hyperbolicity assumption on the equilibria, the

attractor $\mathcal{A}$ of a DS having a global Lyapunov function is a finite union of smooth finite-dimensional submanifolds of the phase space $\Phi$. These attractors are called regular (following Babin–Vishik).

It is also worth emphasizing that, in contrast to general global attractors, regular attractors are robust under perturbations. Moreover, in some cases, it is also possible to verify the so-called transversality conditions (for the intersection of stable and unstable manifolds of the equilibria) and, thus, verify that the DS considered is a Morse–Smale system. In particular, this means that the dynamics restricted to the regular attractor $\mathcal{A}$ is also preserved (up to homeomorphisms) under perturbations.

A disadvantage of the approach of using a regular attractor is the fact that, except for scalar parabolic equations in one space dimension, it is usually extremely difficult to verify the "generic" hyperbolicity and transversality assumptions for concrete values of the physical parameters and the associated hyperbolicity constants, as a rule, cannot be expressed in terms of these parameters.

### Inertial Manifolds

It should be noted that the scheme for the finite-dimensional reduction described above has essential drawbacks. Indeed, the reduced system $(\bar{S}(t), \bar{A})$ is only Hölder continuous and, consequently, cannot be realized as a DS generated by a system of ODEs (and reasonable conditions on the attractor $\mathcal{A}$ which guarantee the Lipschitz continuity of the Mané projections are not known). On the other hand, the complicated geometric structure of the attractor $\mathcal{A}$ (or $\bar{A}$) makes the use of this finite-dimensional reduction in computations hazardous (in fact, only the heuristic information on the number of unknowns which are necessary to capture all the dynamical effects in approximations can be extracted).

In order to overcome these problems, the concept of an inertial manifold (which allows one to embed the global attractor into a smooth manifold) has been suggested by Foias–Sell–Temam. To be more precise, a Lipschitz finite-dimensional manifold $\mathbb{M} \subset \Phi$ is an inertial manifold for the DS $(S(t), \Phi)$ if

1. $\mathbb{M}$ is semiinvariant, that is, $S(t)\mathbb{M} \subset \mathbb{M}$, for all $t \geq 0$;
2. $\mathbb{M}$ satisfies the following asymptotic completeness property: for every $u_0 \in \Phi$, there exists $v_0 \in \mathbb{M}$ such that

$$\|S(t)u_0 - S(t)v_0\|_\Phi \leq Q(\|u_0\|_\Phi)e^{-\alpha t} \qquad [17]$$

where the positive constant $\alpha$ and the monotonic function $Q$ are independent of $u_0$.

We can see that an inertial manifold, if it exists, confirms in a perfect way the heuristic conjecture on the finite dimensionality formulated in the introduction. Indeed, the dynamics of $S(t)$ restricted to an inertial manifold can be, obviously, described by a system of ODEs (which is called the inertial form of the initial PDE). On the other hand, the asymptotic completeness gives (in a very strong form) the equivalence of the initial DS $(S(t), \Phi)$ with its inertial form $(S(t), \mathbb{M})$. Moreover, in turbulence, the existence of an inertial manifold would yield an exact interaction law between the small and large structures of the flow.

Unfortunately, all the known constructions of inertial manifolds are based on a very restrictive condition, the so-called spectral gap condition, which requires arbitrarily large gaps in the spectrum of the linearization of the initial PDE and which can usually be verified only in one space dimension. So, the existence of an inertial manifold is still an open problem for many important equations of mathematical physics (including in particular the two-dimensional Navier–Stokes equations; some nonexistence results have also been proven by Mallet–Paret).

### Exponential Attractors

We first recall that Definition 1 of a global attractor only guarantees that the images $S(t)B$ of all the bounded subsets converge to the attractor, without saying anything on the rate of convergence (in contrast to inertial manifolds, for which this rate of convergence can be controlled). Furthermore, as elementary examples show, this convergence can be arbitrarily slow, so that, until now, we have no effective way for estimating this rate of convergence in terms of the physical parameters of the system (an exception is given by the regular attractors described earlier for which the rate of convergence can be estimated in terms of the hyperbolicity constants of the equilibria. However, even in this situation, it is usually very difficult to estimate these constants for concrete equations). Furthermore, there exist many physically relevant systems (e.g., the so-called slightly dissipative gradient systems) which have trivial global attractors, but very rich and physically relevant transient dynamics which are automatically forgotten under the global-attractor approach. Another important problem is the robustness of the global attractor under perturbations. In fact, global attractors are usually only upper semicontinuous under

perturbations (which means that they cannot explode) and the lower semicontinuity (which means that they cannot also implode) is much more delicate to prove and requires some hyperbolicity assumptions (which are usually impossible to verify for concrete equations).

In order to overcome these difficulties, Eden–Foias–Nicolaenko–Temam have introduced an intermediate object (between inertial manifolds and global attractors), namely an exponential attractor (also called an inertial set).

**Definition 3** A compact set $\mathcal{M} \subset \Phi$ is an exponential attractor for the DS $(S(t), \Phi)$ if

(i) $\mathcal{M}$ has finite fractal dimension: $d_f(\mathcal{M}) < \infty$;
(ii) $\mathcal{M}$ is semi-invariant: $S(t)\mathcal{M} \subset \mathcal{M}$, for all $t \geq 0$;
(iii) $\mathcal{M}$ attracts exponentially the images of all the bounded sets $B \subset \phi$:

$$\mathrm{dist}_H(S(t)B, \mathcal{M}) \leq Q(\|B\|_\Phi)e^{-\alpha t} \qquad [18]$$

where the positive constant $\alpha$ and the monotonic function $Q$ are independent of $B$.

Thus, on the one hand, an exponential attractor remains finite dimensional (like the global attractor) and, on the other hand, estimate [18] allows one to control the rate of attraction (like an inertial manifold). We note, however, that the relaxation of strict invariance to semi-invariance allows this object to be nonunique. So, we have here the problem of the "best choice" of the exponential attractor. We also mention that an exponential attractor, if it exists, always contains the global attractor.

Although the initial construction of exponential attractors is based on the so-called squeezing property (and requires Zorn's lemma), we formulate below a simpler construction, due to Efendiev–Miranville–Zelik, which is similar to the method proposed by Ladyzhenskaya to verify the finite dimensionality of global attractors. This is done for discrete times and for a DS generated by iterations of some map $S:\Phi \to \Phi$, since the passage from discrete to continuous times usually arises no difficulty (without loss of generality, the reader may think that $S = S(1)$ and $(S(t), \Phi)$ is one of the DS mentioned in the introduction).

**Theorem 3** *Let the phase space $\Phi_0$ be a closed bounded subset of some Banach space $H$ and let $H_1$ be another Banach space compactly embedded into $H$. Assume also that the map $S:\Phi_0 \to \Phi_0$ satisfies the following "smoothing" property*:

$$\|Su_1 - Su_2\|_{H_1} \leq K\|u_1 - u_2\|_H, \quad u_1, u_2 \in \Phi_0 \qquad [19]$$

*for some constant $K$ independent of $u_i$. Then, the DS $(S, \Phi_0)$ possesses an exponential attractor.*

In applications, $\Phi_0$ is usually a bounded absorbing/attracting set whose existence is guaranteed by the dissipative estimate [10], $H := L^2(\Omega)$ and $H_1 := H^1(\Omega)$. Furthermore, estimate [19] simply follows from the classical parabolic smoothing property, but now applied to the equation of variations (as in [11], hyperbolic equations require a slightly more complicated analogue of [19]). These simple arguments show that exponential attractors are as general as global attractors and, to the best of our knowledge, exponential attractors exist indeed for all the equations of mathematical physics for which the finite dimensionality of the global attractor can be established. Moreover, since $\mathcal{A} \subset \mathcal{M}$, this scheme can also be used to prove the finite dimensionality of global attractors.

It is finally worth emphasizing that the control on the rate of convergence provided by [18] makes exponential attractors much more robust than global attractors. In particular, they are upper and lower semicontinuous under perturbations (of course, up to the "best choice," since they are not unique), as shown by Efendiev–Miranville–Zelik.

## Essentially Infinite-Dimensional Dynamical Systems – The Case of Unbounded Domains

As already mentioned in the introduction, the theory of dissipative DS in unbounded domains is developing only now and the results given here are not as complete as for bounded domains. Nevertheless, we indicate below several of the most interesting (from our point of view) results concerning the general description of the dynamics generated by such problems by considering a system of reaction–diffusion equations [8] in $\mathbb{R}^n$ with phase space $\Phi = L^\infty(\mathbb{R}^n)$ as a model example (although all the results formulated below are general and depend weakly on the choice of equation).

### Generalization of the Global Attractor and Kolmogorov's $\varepsilon$-Entropy

We first note that Definition 1 of a global attractor is too strong for equations in unbounded domains. Indeed, as seen earlier, the compactness of the attractor is usually based on the compactness of the embedding $H^1(\Omega) \subset L^2(\Omega)$, which does not hold in unbounded domains. Furthermore, an attractor, in the sense of Definition 1, does not exist for most of the interesting examples of eqns [8] in $\mathbb{R}^n$.

It is natural to use instead the concept of the so-called locally compact global attractor which is well adapted to unbounded domains. This attractor $\mathcal{A}$ is only bounded in the phase space $\Phi = L^\infty(\mathbb{R}^n)$, but its restrictions $\mathcal{A}|_\Omega$ to all bounded domains $\Omega$ are compact in $L^\infty(\Omega)$. Moreover, the attraction property should also be understood in the sense of a local topology in $L^\infty(\mathbb{R}^n)$. It is known that this generalized global attractor $\mathcal{A}$ exists indeed for problem [8] in $\mathbb{R}^n$ (of course, under some "natural" assumptions on the nonlinearity $f$ and the diffusion matrix $a$). As for bounded domains, its existence is based on the dissipative estimate [10], the smoothing property [11], and the compactness of the embedding $H^1_{\mathrm{loc}}(\mathbb{R}^n) \subset L^2_{\mathrm{loc}}(\mathbb{R}^n)$ (we need to use the local topology only to have this compactness).

The next natural question that arises here is how to control the "size" of the attractor $\mathcal{A}$ if its fractal dimension is infinite (which is usually the case in unbounded domains). One of the most natural ways to handle this problem (which was first suggested by Chepyzhov–Vishik in the different context of uniform attractors associated with nonautonomous equations in bounded domains and appears as extremely fruitful for the theory of dissipative PDE in unbounded domains) is to study the asymptotics of Kolmogorov's $\varepsilon$-entropy of the attractor. Actually, since the attractor $\mathcal{A}$ is compact only in a local topology, it is natural to study the entropy of its restrictions, say, to balls $B^R_{x_0}$ of $\mathbb{R}^n$ of radius $R$ centered at $x_0$ with respect to the three parameters $R, x_0$, and $\varepsilon$. A more or less complete answer to this question is given by the following estimate:

$$\mathcal{H}_\varepsilon(\mathcal{A}|_{B^R_{x_0}}) \le C(R + \log_2 1/\varepsilon)^n \log_2 1/\varepsilon \qquad [20]$$

where the constant $C$ is independent of $\varepsilon \le 1$, $R$, and $x_0$. Moreover, it can be shown that this estimate is sharp for all $R$ and $\varepsilon$ under the very weak additional assumption that eqn [8] possesses at least one exponentially unstable spatially homogeneous equilibrium.

Thus, formula [20] (whose proof is also based on a smoothing property for the equation of variations) can be interpreted as a natural generalization of the heuristic principle of finite dimensionality of global attractors to unbounded domains. It is also worth recalling that the entropy of the embedding of a ball $\mathcal{B}_k$ of the space $C^k(B^R_{x_0})$ into $C(B^R_{x_0})$ has the asymptotic $\mathcal{H}_\varepsilon(\mathcal{B}) \sim C_R(1/\varepsilon)^{n/k}$, which is essentially worse than [20]. So, [20] is not based on the smoothness of the attractor $\mathcal{A}$ and, therefore, reflects deeper properties of the equation.

## Spatial Dynamics and Spatial Chaos

The next main difference with bounded domains is the existence of unbounded spatial directions which can generate the so-called spatial chaos (in addition to the "usual" temporal chaos arising under the evolution). In order to describe this phenomenon, it is natural to consider the group $\{T_h, h \in \mathbb{R}^n\}$ of spatial translations acting on the attractor $\mathcal{A}$:

$$(T_h u_0)(x) := u_0(x+h), \quad T_h : \mathcal{A} \to \mathcal{A} \qquad [21]$$

as a DS (with multidimensional "times" if $n > 1$) acting on the phase space $\mathcal{A}$ and to study its dynamical properties.

In particular, it is worth noting that the lower bounds on the $\varepsilon$-entropy that one can derive imply that the topological entropy of this spatial DS is infinite and, consequently, the classical symbolic dynamics with a finite number of symbols is not adequate to clarify the nature of chaos in [21]. In order to overcome this difficulty, it was suggested by Zelik to use Bernoulli shifts with an infinite number of symbols, belonging to the whole interval $\omega \in [0, 1]$. To be more precise, let us consider the Cartesian product $\mathbb{M}_n := [0,1]^{\mathbb{Z}^n}$ endowed with the Tikhonov topology. Then, this set can be interpreted as the space of all the functions $v : \mathbb{Z}^n \to [0,1]$, endowed with the standard local topology. We define a DS $\{\mathcal{T}_l, l \in \mathbb{Z}^n\}$ on $\mathbb{M}_n$ by

$$(\mathcal{T}_l v)(m) := v(m+l), \quad v \in \mathbb{M}_n,\ l, m \in \mathbb{Z}^n \qquad [22]$$

Based on this model, the following description of spatial chaos was obtained.

**Theorem 4** *Let eqn [8] in $\Omega = \mathbb{R}^n$ possess at least one exponentially unstable spatially homogeneous equilibrium. Then, there exist $\alpha > 0$ and a homeomorphic embedding $\tau : \mathbb{M}_n \to \mathcal{A}$ such that*

$$T_{\alpha l} \circ \tau(v) = \tau \circ \mathcal{T}_l(v), \quad \forall l \in \mathbb{Z}^n,\ v \in \mathbb{M}_n \qquad [23]$$

Thus, the spatial dynamics, restricted to the set $\tau(\mathbb{M}_n)$, is conjugated to the symbolic dynamics on $\mathbb{M}_n$. Moreover, there exists a dynamical invariant (the so-called mean toplogical dimension) which is always finite for the spatial DS [22] and strictly positive for the Bernoulli scheme $\mathbb{M}_n$. So, the embedding [23] clarifies, indeed, the nature of chaos arising in the spatial DS [21].

## Spatio-Temporal Chaos

To conclude, we briefly discuss an extension of Theorem 4, which takes into account the temporal modes and, thus, gives a description of the spatio-temporal chaos. In order to do so, we first note

that the spatial DS [21] commutes obviously with the temporal evolution operators $S(t)$ and, consequently, an extended $(n+1)$-parametric semigroup $\{\mathbb{S}(t,h), (t,h) \in \mathbb{R}_+ \times \mathbb{R}^n\}$ acts on the attractor:

$$\mathbb{S}(t,h) := S(t) \circ T_h, \qquad \mathbb{S}(t,h) : \mathcal{A} \to \mathcal{A} \\ t \in \mathbb{R}_+, \quad h \in \mathbb{R}^n \qquad [24]$$

Then, this semigroup (interpreted as a DS with multidimensional times) is responsible for all the spatio-temporal dynamical phenomena in the initial PDE [8] and, consequently, the question of finding adequate dynamical characteristics is of a great interest. Moreover, it is also natural to consider the subsemigroups $\mathbb{S}_{V_k}(t,h)$ associated with the $k$-dimensional planes $V_k$ of the spacetime $\mathbb{R}_+ \times \mathbb{R}^n$, $k < n+1$.

Although finding an adequate description of the dynamics of [24] seems to be an extremely difficult task, some particular results in this direction have already been obtained. Thus, it has been proved by Zelik that the semigroup [24] has finite topological entropy and the entropy of its subsemigroups $\mathbb{S}_{V_k}(t,h)$ is usually infinite if $k < n+1$. Moreover (adding a natural transport term of the form $(L, \nabla_x)u$ to eqn [8]), it was proved that the analog of Theorem 4 holds for the subsemigroups $\mathbb{S}_{V_n}(t,h)$ associated with the $n$-dimensional hyperplanes $V_n$ of the spacetime. Thus, the infinite-dimensional Bernoulli shifts introduced in the previous subsection can be used to describe the temporal evolution in unbounded domains as well.

In particular, as a consequence of this embedding, the topological entropy of the initial purely temporal evolution semigroup $S(t)$ is also infinite, which indicates that (even without considering the spatial directions) we have indeed here essential new levels of dynamical complexity which are not observed in the classical DS theory of ODEs.

*See also:* Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Ergodic Theory; Evolution Equations: Linear and Nonlinear; Fractal Dimensions in Dynamics; Inviscid flows; Lyapunov Exponents and Strange Attractors.

## Further Reading

Babin AV and Vishik MI (1992) *Attractors of Evolution Equations*. Amsterdam: North-Holland.

Chepyzhov VV and Vishik MI (2002) Attractors for Equations of Mathematical Physics. *American Mathematical Society Colloquium Publications*, vol. 49. Providence, RI: American Mathematical Society.

Faddeev LD and Takhtajan LA (1987) Hamiltonian Methods in the Theory of Solitons. *Springer Series in Soviet Mathematics*. Berlin: Springer.

Hale JK (1988) Asymptotic Behavior of Dissipative Systems. *Mathematical Surveys and Monographs*, vol. 25. Providence, RI: American Mathematical Society.

Katok A and Hasselblatt B (1995) Introduction to the Modern Theory of Dynamical Systems. *Encyclopedia of Mathematics and its Applications*, vol. 54. Cambridge: Cambridge University Press.

Ladyzhenskaya OA (1991) *Attractors for Semigroups and Evolution Equations*. Cambridge: Cambridge University Press.

Temam R (1997) Infinite-Dimensional Dynamical Systems in Mechanics and Physics. *Applied Mathematical Sciences*, 2nd edn., vol. 68. pp. New York: Springer.

Zelik S (2004) Multiparametrical semigroups and attractors of reaction–diffusion equations in $\mathbb{R}^n$. *Proceedings of the Moscow Mathematical Society* 65: 69–130.

**Donaldson Invariants** *see* Gauge Theoretic Invariants of 4-Manifolds

# Donaldson–Witten Theory

**M Mariño**, CERN, Geneva, Switzerland



## Introduction

Since they were introduced by Witten in 1988, topological quantum field theories (TQFTs) have had a tremendous impact in mathematical physics (see Birmingham *et al.* (1991) and Cordes *et al.* for a review). These quantum field theories are constructed in such a way that the correlation functions of certain operators provide topological invariants of the spacetime manifold where the theory is defined. This means that one can use the methods and insights of quantum field theory in order to obtain information about topological invariants of low-dimensional manifolds.

Historically, the first TQFT was Donaldson–Witten theory, also called topological Yang–Mills theory. This theory was constructed by Witten (1998) starting from $\mathcal{N}=2$ super Yang–Mills by a procedure called

"topological twisting." The resulting model is topological and the famous Donaldson invariants of 4-manifolds are then recovered as certain correlation functions in the topological theory. The analysis of Witten (1998) did not indicate any new method to compute the invariants, but in 1994 the progress in understanding the nonperturbative dynamics of $\mathcal{N}=2$ theories (Seiberg and Witten 1994 a, b) led to an alternative way of computing correlation functions in Donaldson–Witten theory. As Witten (1994) showed, Donaldson–Witten theory can be reduced to another, simpler topological theory consisting of a twisted abelian gauge theory coupled to spinor fields. This theory leads to a different set of 4-manifold invariants, the so-called "Seiberg–Witten invariants," and Donaldson invariants can be expressed in terms of these invariants through Witten's "magic formula." The connection between Seiberg–Witten and Donaldson invariants was streamlined and extended by Moore and Witten by using the so-called $u$-plane integral (Moore and Witten 1998). This has led to a rather complete understanding of Donaldson–Witten theory from a physical point of view.

In this article we provide a brief review of Donaldson–Witten theory. First, we describe the construction of the model, from both a mathematical and a physical point of view, and state the main results for the Donaldson–Witten generating functional. In the next section, we present the basic results of the $u$-plane integral of Moore and Witten and sketch how it can be used to solve Donaldson–Witten theory. In the final section, we mention some generalizations of the basic framework. For a complete exposition of Donaldson–Witten theory, the reader is referred to the book by Labastida and Mariño (2005). A short review of the $u$-plane integral can be found in Mariño and Moore (1998a).

## Donaldson–Witten Theory: Basic Construction and Results

### Donaldson–Witten Theory According to Donaldson

Donaldson theory as formulated in Donaldson (1990), Donaldson and Kronheimer (1990), and Friedman and Morgan (1991) starts with a principal $G=\mathrm{SO}(3)$ bundle $V\to X$ over a compact, oriented, Riemannian 4-manifold $X$, with fixed instanton number $k$ and Stiefel–Whitney class $w_2(V)$ (SO(3) bundles on a 4-manifold are classified up to isomorphism by these topological data). The moduli space of anti-self-dual (ASD) connections is then defined as

$$\mathcal{M}_{\mathrm{ASD}}=\{A:F^+(A)=0\}/\mathcal{G} \qquad [1]$$

where $F^+(A)$ is the self-dual part of the curvature, and $\mathcal{G}$ is the group of gauge transformations. To construct the Donaldson polynomials, one considers the universal bundle

$$\boldsymbol{P}=(V\times\mathcal{A}^*)/(\mathcal{G}\times G) \qquad [2]$$

where $\mathcal{A}^*$ is the space of irreducible $G$-connections on $V$. This is a $G$-bundle over $\mathcal{B}^*\times X$, where $\mathcal{B}^*=\mathcal{A}^*/\mathcal{G}$ is the space of irreducible connections modulo gauge transformations, and as such has a Pontrjagin class

$$p_1(\boldsymbol{P})\in H^*(\mathcal{B}^*)\otimes H^*(X) \qquad [3]$$

One can then obtain differential forms on $\mathcal{B}^*$ by taking the slant product of $p_1(\boldsymbol{P})$ with homology classes in $X$. In this way we obtain the Donaldson map:

$$\mu: H_i(X)\longrightarrow H^{4-i}(\mathcal{B}^*) \qquad [4]$$

After restriction to $\mathcal{M}_{\mathrm{ASD}}$, we obtain the following differential forms on the moduli space of ASD connections:

$$\begin{aligned} x\in H_0(X)&\to\mathcal{O}(x)\in H^4(\mathcal{M}_{\mathrm{ASD}})\\ S\in H_2(X)&\to I_2(S)\in H^2(\mathcal{M}_{\mathrm{ASD}})\end{aligned} \qquad [5]$$

If the manifold $X$ has $b_1(X)\neq 0$, there are also cohomology classes associated to 1-cycles and 3-cycles, but we will not consider them here.

We can now formally define the Donaldson invariants as follows. Consider the space

$$\boldsymbol{A}(X)=\mathrm{Sym}(H_0(X)\oplus H_2(X)) \qquad [6]$$

with a typical element written as $x^\ell S_{i_1}\cdots S_{i_p}$. The Donaldson invariant corresponding to this element of $\boldsymbol{A}(X)$ is the following intersection number:

$$\begin{aligned}&\mathcal{D}_X^{w_2(V),k}(x^\ell S_{i_1}\cdots S_{i_p})\\ &\quad=\int_{\mathcal{M}_{\mathrm{ASD}}}\mathcal{O}^\ell\wedge I_2(S_{i_1})\wedge\cdots\wedge I_2(S_{i_p})\end{aligned} \qquad [7]$$

where $\mathcal{M}_{\mathrm{ASD}}$ is the moduli space of ASD connections with second Stiefel–Whitney class $w_2(V)$ and instanton number $k$. The integral in [7] will be different from zero only if the degrees of the forms add up to $\dim(\mathcal{M}_{\mathrm{ASD}})$.

It is very convenient to pack all Donaldson invariants in a generating functional. Let $\{S_i\}_{i=1,\ldots,b_2}$ be a basis of 2-cycles. We introduce the formal sum

$$S=\sum_{i=1}^{b_2}v_iS_i \qquad [8]$$

where $v_i$ are complex numbers. We then define the Donaldson–Witten generating functional as

$$Z_{\mathrm{DW}}^{w_2(V)}(p, v_i) = \sum_{k=0}^{\infty} \mathcal{D}_X^{w_2(V),k}(\mathrm{e}^{px+S}) \qquad [9]$$

where on the right-hand side we are summing over all instanton numbers, that is, we are summing over all topological configurations of the SO(3) gauge field with a fixed $w_2(V)$. This gives a formal power series in $p$ and $v_i$.

For $b_2^+(X) > 1$, the generating functional [9] is a diffeomorphism invariant of $X$; therefore, it is potentially a powerful tool in four-dimensional topology. When $b_2^+(X) = 1$, Donaldson invariants are metric dependent. The metric dependence can be described in more detail as follows. Define the period point as the harmonic 2-form satisfying

$$*\omega = \omega, \qquad \omega^2 = 1 \qquad [10]$$

which depends on the conformal class of the metric. As the conformal class of the metric varies, $\omega$ describes a curve in the cone

$$V_+ = \{\omega \in H^2(X, \boldsymbol{R}) : \omega^2 > 0\} \qquad [11]$$

Let $\zeta \in H^2(X)$ satisfy

$$\zeta \equiv w_2(V) \bmod 2, \quad \zeta^2 < 0, \ (\zeta, \omega) = 0 \qquad [12]$$

Such an element $\zeta$ defines a "wall" in $V_+$:

$$W_\zeta = \{\omega : (\zeta, \omega) = 0\} \qquad [13]$$

The complements of these walls are called "chambers," and the cone $V_+$ is then divided in chambers separated by walls. A class $\zeta$ satisfying [12] is the first Chern class associated to a reducible solution of the ASD equations, and it causes a singularity in moduli space: the Donaldson invariants jump when we pass through such a wall. Therefore, when $b_2^+(X) = 1$, Donaldson invariants are metric independent in each chamber. A basic problem in Donaldson–Witten theory is to determine the jump in the generating function as we cross a wall,

$$Z_+^\zeta(p, S) - Z_-^\zeta(p, S) = WC_\zeta(p, S) \qquad [14]$$

The jump term $WC_\zeta(p, S, \delta)$ is usually called the "wall-crossing" term.

The basic goal of Donaldson theory is to study the properties of the generating functional [9] and to compute it for different 4-manifolds $X$. On the mathematical side, many results have been obtained on $Z_{\mathrm{DW}}$, and some of them can be found in Donaldson and Kronheimer (1990), Friedman and Morgan (1991), Stern (1998), and Göttsche (1996). On the other hand, Donaldson theory can be formulated as a topological field theory, and many of these results can be obtained by using quantum field theory techniques. This will be our main focus for the rest of the article.

### Donaldson–Witten Theory According to Witten

Witten (1988) constructed a twisted version of $\mathcal{N} = 2$ super Yang–Mills theory which has a nilpotent Becchi–Rouet–Stora–Tyutin (BRST) charge (modulo gauge transformations)

$$\overline{\mathcal{Q}} = \epsilon^{\dot{\alpha}\dot{A}} Q_{\dot{\alpha}\dot{A}} \qquad [15]$$

where $Q_{\dot{\alpha}\dot{A}}$ are the supersymmetric (SUSY) charges. Here $\dot{\alpha}$ is a chiral spinor index and $A$ has its origin in the SU(2) $\mathcal{R}$-symmetry. The field content of the theory is the standard twisted $\mathcal{N} = 2$ vector multiplet:

$$A, \psi_\mu = \psi_{\alpha\dot{\alpha}}, \phi, \quad D_{\mu\nu}^+, \chi_{\mu\nu}^+ = \overline{\psi}_{\dot{\alpha}\dot{\beta}}, \quad \overline{\phi}, \eta = \overline{\psi}^{\dot{\alpha}}{}_{\dot{\alpha}} \quad [16]$$

where $(1/2) D_{\mu\nu}^+ \, \mathrm{d}x^\mu \, \mathrm{d}x^\nu$ is a self-dual 2-form derived from the auxiliary fields, etc. All fields are valued in the adjoint representation of the gauge group. After twisting, the theory is well defined on any Riemannian 4-manifold, since the fields are naturally interpreted as differential forms and the $\overline{\mathcal{Q}}$ charge is a scalar (Witten 1988).

The observables of the theory are $\overline{\mathcal{Q}}$ cohomology classes of operators, and they can be constructed from 0-form observables $\mathcal{O}^{(0)}$ using the descent procedure. This amounts to solving the equations

$$d\mathcal{O}^{(i)} = \{\overline{\mathcal{Q}}, \mathcal{O}^{(i+1)}\}, \quad i = 0, \ldots, 3 \qquad [17]$$

The integration over $i$-cycles $\gamma^{(i)}$ in $X$ of the operators $\mathcal{O}^{(i)}$ is then an observable. These descent equations have a canonical solution: the 1-form-valued operator $K_{\alpha\dot{\alpha}} = -\mathrm{i}\delta_\alpha^A Q_{\dot{\alpha}\dot{A}}/4$ verifies

$$d = \{\overline{\mathcal{Q}}, K\} \qquad [18]$$

as a consequence of the supersymmetry algebra. The operators $\mathcal{O}^{(i)} = K^i \mathcal{O}^{(0)}$ solve the descent equations [17] and are canonical representatives. When the gauge group is SU(2), the observables are obtained by the descent procedure from the operator

$$\mathcal{O} = \mathrm{tr}(\phi^2) \qquad [19]$$

The topological descendant $\mathcal{O}^{(2)}$ is given by

$$\mathcal{O}^{(2)} = -\frac{1}{2}\mathrm{tr}\left(\frac{1}{\sqrt{2}}\phi(F_{\mu\nu}^- + D_{\mu\nu}^+) - \frac{1}{4}\psi_\mu\psi_\nu\right)\mathrm{d}x^\mu \wedge \mathrm{d}x^\nu \quad [20]$$

and the resulting observable is

$$I_2(S) = \int_S \mathcal{O}^{(2)} \qquad [21]$$

$\mathcal{O}$ and $I_2(S)$ correspond to the cohomology classes in [5]. One of the main results of Witten (1988) is that the semiclassical approximation in the twisted $\mathcal{N}=2$ Yang–Mills theory is exact. The semiclassical evaluation of correlation functions of the observables above leads directly to the definition of Donaldson invariants, and the generating functional [9] can be written as a correlation function of the twisted theory. One then has

$$Z_{\mathrm{DW}}^{w_2(V)}(p,S) = \left\langle \exp\left(p\mathcal{O} + I_2(S)\right)\right\rangle \qquad [22]$$

### Results for the Donaldson–Witten Generating Function

The basic results that have emerged from the physical approach to Donaldson–Witten theory are the following.

1. The Donaldson–Witten generating functional is in general the sum of the two terms,

$$Z_{\mathrm{DW}} = Z_u + Z_{\mathrm{SW}} \qquad [23]$$

(We have omitted the Stiefel–Whitney class for convenience.) The first term, $Z_u$, is called the "$u$-plane integral." It is given by a complicated integral over $\mathbf{C}$ which can be written, in turn, as an integral over a fundamental domain of the congruence subgroup $\Gamma^0(4)$ of $\mathrm{SL}(2,\mathbf{Z})$. $Z_u$ depends only on the cohomology ring of $X$, and therefore does not contain any information beyond the one provided by classical topology. Finally, $Z_u$ vanishes if $b_2^+(X) > 1$, and it is responsible for the wall-crossing behavior of $Z_{\mathrm{DW}}$ when $b_2^+(X)=1$.

2. The second term of [23], $Z_{\mathrm{SW}}$, is called the Seiberg–Witten contribution. This contribution involves the Seiberg–Witten invariants of $X$, which are obtained by considering the moduli problem defined by the Seiberg–Witten monopole equations (Witten 1994b):

$$\begin{aligned} F^+_{\dot\alpha\dot\beta} + 4\mathrm{i}\bar{M}_{(\dot\alpha}M_{\dot\beta)} &= 0 \\ D_L^{\alpha\dot\alpha}M_{\dot\alpha} &= 0 \end{aligned} \qquad [24]$$

In these equations, $M_{\dot\alpha}$ is a section of the spinor bundle $S^+\otimes L^{1/2}$, $L$ is the determinant line bundle of a $\mathrm{Spin}_c$ structure on $X$, $F^+_{\dot\alpha\dot\beta} = \bar\sigma^{\mu\nu}_{\dot\alpha\dot\beta}F^+_{\mu\nu}$ is the self-dual part of the curvature of a U(1) connection on $L$, and $D_L$ is the Dirac operator for the bundle $S^+\otimes L^{1/2}$. The solutions of these equations modulo gauge equivalence form the moduli space $\mathcal{M}_S W$, and the Seiberg–Witten invariants are defined by integrating suitable differential forms on this moduli space. We will label $\mathrm{Spin}_c$ structures by the class $\lambda = c_1(L^{1/2}) \in H^2(X,\mathbf{Z}) + w_2(X)/2$. We say that $\lambda$ is a Seiberg–Witten basic class if the corresponding Seiberg–Witten invariants are not all zero. If $\mathcal{M}_S W$ is zero dimensional, the Seiberg–Witten invariant depends only on the $\mathrm{Spin}_c$ structure associated to $\lambda = c_1(L^{1/2})$, and is denoted by $\mathrm{SW}(\lambda)$.

3. A manifold $X$ is said to be of Seiberg–Witten simple type if all the Seiberg–Witten basic classes have a zero-dimensional moduli space. For simply connected 4-manifolds of Seiberg–Witten simple type and with $b_2^+(X) > 1$, Witten determined the Seiberg–Witten contribution and proposed the following "magic formula" for $Z_{\mathrm{DW}}$ (Witten 1994b):

$$Z_{\mathrm{DW}} = 2^{1+7\chi/4+11\sigma/4}\sum_\lambda \mathrm{e}^{2\mathrm{i}\pi(\lambda_0\cdot\lambda+\lambda_0^2)}\Big[\mathrm{e}^{2p+S^2/2}\mathrm{e}^{2(S,\lambda)} + \mathrm{i}^{\chi_h - w_2(V)^2}\mathrm{e}^{-2p-S^2/2}\mathrm{e}^{-2\mathrm{i}(S,\lambda)}\Big]\mathrm{SW}(\lambda) \qquad [25]$$

In this equation, $\chi,\sigma$ are the Euler characteristic and signature of $X$, respectively, $\chi_{\mathrm{h}} = (\chi+\sigma)/4$ is the holomorphic Euler characteristic of $X$, and $\lambda_0$ is an integer lifting of $w_2(V)$. This formula generalizes previous results by Witten (1994a) for Kähler manifolds. It also follows from this formula that the Donaldson–Witten generating function of simply connected 4-manifolds of Seiberg–Witten simple type and with $b_2^+(X) > 1$ satisfies

$$\left(\frac{\partial^2}{\partial p^2} - 4\right)Z_{\mathrm{DW}} = 0$$

which is the Donaldson simple type condition introduced by Kronheimer and Mrowka (1994).

4. Using the $u$-plane integral, one can find explicit expressions for $Z_{\mathrm{DW}}$ in more general situations (like non-simply-connected manifolds or manifolds which are not of Seiberg–Witten simple type).

In the next section we explain the formalism of the $u$-plane integral introduced by Moore and Witten (1998), which makes possible a detailed derivation of the above results.

## The $u$-Plane Integral

### Definition of the $u$-Plane Integral

The evaluation of the Donaldson–Witten generating function can be made by using the results of Seiberg and Witten (1994 a, b) on the low-energy dynamics of SU(2), $\mathcal{N}=2$ Yang–Mills theory. In their work, Seiberg and Witten determined the exact low-energy effective action of the model up to two derivatives.

From a physical point of view, there are certainly corrections to this effective action which are difficult to evaluate. Fortunately, the computation in the twisted version of the theory can be done by just considering the Seiberg–Witten effective action. This is because the correlation functions in the twisted theory are invariant under rescalings of the metric, so we can evaluate them in the limit of large distances or equivalently of very low energies. The effective action up to two derivatives is sufficient for that purpose.

One way of describing the main result of the work of Seiberg and Witten is that the moduli space of $\bar{\mathcal{Q}}$-fixed points of the twisted SO(3) $\mathcal{N}=2$ theory on a compact 4-manifold has two branches, which we refer to as the Coulomb and Seiberg–Witten branches. On the Coulomb branch the expectation value

$$u = \frac{\langle \mathrm{tr}\,\phi^2 \rangle}{16\pi^2}$$

breaks $\mathrm{SO(3)} \to \mathrm{U(1)}$ via the standard Higgs mechanism. The Coulomb branch is simply a copy of the complex $u$-plane. The low-energy effective theory on this branch is simply the abelian $\mathcal{N}=2$ gauge theory. However, at two points, $u=\pm 1$, there is a singularity where the moduli space meets a second branch, the Seiberg–Witten branch. At these points, the effective action is given by the magnetic dual of the U(1), $\mathcal{N}=2$ gauge theory coupled to a monopole matter hypermultiplet. Therefore, this branch consists of solutions to the Seiberg–Witten equations [24].

Since the manifold $X$ is compact, the partition function of the twisted theory is a sum over "all" vacuum states. Equation [23] then follows. In this equation, $Z_u$ comes from "integrating over the $u$-plane," while $Z_{\mathrm{SW}}$ corresponds to the points $u=\pm 1$. As we stated before, $Z_u$ vanishes for manifolds of $b_2^+(X) > 1$, but once this piece has been determined an argument originally presented at Moore and Witten (1998) allows one to derive the form of $Z_{\mathrm{SW}}$ as well for arbitrary $b_2^+(X) \geq 1$.

The computation of $Z_u$ is presented in detail in Moore and Witten (1998). The starting point of the computation is the untwisted low-energy theory, which has been described in detail in Seiberg and Witten (1994 a, b) and Witten (1995). It is an $\mathcal{N}=2$ theory characterized by a prepotential $\mathcal{F}$ which depends on an $\mathcal{N}=2$ vector multiplet. The effective gauge coupling is given by $\tau(a)=\mathcal{F}''(a)$, where $a$ is the scalar component of the vector multiplet. The Euclidean Lagrange density for the $u$-plane theory can be obtained simply by twisting the physical theory. It can be written as

$$\begin{aligned} &\frac{\mathrm{i}}{6\pi} K^4 \mathcal{F}(a) + \frac{1}{16\pi}\left\{\overline{\mathcal{Q}}, \overline{\mathcal{F}}''\chi(D+F_+)\right\} \\ &\quad - \frac{\mathrm{i}\sqrt{2}}{32\pi}\left\{\overline{\mathcal{Q}}, \overline{\mathcal{F}}' d * \psi\right\} - \frac{\sqrt{2}\mathrm{i}}{3\times 2^5\pi} \\ &\quad \times \left\{\mathcal{Q}, \overline{\mathcal{F}}'''\chi_{\mu\nu}\chi^{\nu\lambda}\chi_\lambda{}^{\mu}\right\}\sqrt{g}d^4x \\ &\quad + \mathcal{A}(u)\mathrm{tr}R\wedge R + \mathcal{B}(u)\mathrm{tr}R\wedge\tilde{R} \end{aligned} \qquad [26]$$

where $\mathcal{A}(u), \mathcal{B}(u)$ describe the coupling to gravity, and after integration of the corresponding differential forms we obtain terms proportional to the signature $\sigma$ and Euler characteristic $\chi$ of $X$. The data of the low-energy effective action can be encoded in an elliptic curve of the form

$$y^2 = x^3 - ux^2 + \tfrac{1}{4}x \qquad [27]$$

and $\tau$ is the modulus of the curve. The monodromy group of this curve is $\Gamma^0(4)$. All the quantities involved in the action can be obtained by integrating a certain meromorphic differential on the curve, and they can be expressed in terms of modular forms.

As for the operators, we have $u=\mathcal{O}(P)$ by definition. We may then obtain the 2-observables from the descent procedure. The result is that $I(S)\to \tilde{I}(S)=\int_S K^2 u = \int_S (\mathrm{d}u/\mathrm{d}a)(D_+ + F_-)+\cdots$. Here $D_+$ is the auxiliary field. Although one has $I(S)\to\tilde{I}(S)$ in going from the microscopic theory to the effective theory, it does not necessarily follow that $I(S_1)I(S_2)\to\tilde{I}(S_1)\tilde{I}(S_2)$ because there can be contact terms. If $S_1$ and $S_2$ intersect, then in passing to the low-energy theory we integrate out massive modes. This can induce delta function corrections to the operator product expansion modifying the mapping to the low-energy theory as follows:

$$\exp(I(S)) \to \exp\left(\tilde{I}(S) + S^2 T(u)\right) \qquad [28]$$

where $T(u)$ is the contact term. Such contact terms were observed in Witten (1994a) and studied in detail in Losev *et al.* (1998). It can be shown that

$$T(u) = -\frac{1}{24}E_2(\tau)\left(\frac{\mathrm{d}u}{\mathrm{d}a}\right)^2 + \frac{1}{3}u \qquad [29]$$

where $E_2(\tau)$ is Eisenstein's series and $\mathrm{d}a/\mathrm{d}u$ is one of the periods of the elliptic curve [27].

The final result of Moore and Witten is the following expression:

$$Z_u(p,S) = \int_C \frac{\mathrm{d}u\,\mathrm{d}\bar{u}}{y^{1/2}}\mu(\tau)\mathrm{e}^{2pu+S^2\hat{T}(u)}\Psi \qquad [30]$$

Here,

$$\mu(\tau) = \frac{\mathrm{d}\bar{\tau}}{\mathrm{d}\bar{u}}\left(\frac{\mathrm{d}a}{\mathrm{d}u}\right)^{1-(1/2)\chi}\Delta^{\sigma/8}$$
$$\hat{T}(u) = T(u) + \frac{(\mathrm{d}u/\mathrm{d}a)^2}{8\pi y} \qquad [31]$$

where $y = \mathrm{Im}\,\tau$ and $\Delta$ is the discriminant of the curve [27]. The quantity $\Psi$ is essentially a Narain–Siegel theta function associated to the lattice $H^2(X, \boldsymbol{Z})$. Notice that this lattice is Lorentzian and has signature $(1, (-1)^{b_2^-(X)})$ (since $b_2^+(X) = 1$). The self-dual projection of a 2-form $\lambda$ can be done with the period point $\omega$ as $\lambda_+ = (\lambda, \omega)\omega$. The lattice is shifted by half the second Stiefel–Whitney class of the bundle, $w_2(V)$, that is,

$$\Gamma = H^2(X, \boldsymbol{Z}) + \tfrac{1}{2} w_2(V)$$

and

$$\Psi = \exp\left[-\frac{1}{8\pi y}\left(\frac{\mathrm{d}u}{\mathrm{d}a}\right)^2 S_-^2\right] \mathrm{e}^{2\pi \mathrm{i}\lambda_0^2} \sum_{\lambda\in\Gamma} (-1)^{(\lambda-\lambda_0)\cdot w_2(X)}$$
$$\times \left[(\lambda, \omega) + \frac{\mathrm{i}}{4\pi y}\frac{\mathrm{d}u}{\mathrm{d}a}(S, \omega)\right]$$
$$\times \exp\left[-\mathrm{i}\pi\bar{\tau}(\lambda_+)^2 - \mathrm{i}\pi\tau(\lambda_-)^2 - \mathrm{i}\frac{\mathrm{d}u}{\mathrm{d}a}(S, \lambda_-)\right] \qquad [32]$$

Here, $w_2(X)$ is the second Stiefel–Whitney class of $X$, and $\lambda_0$ is a choice of lifting of $w_2(V)$ to $H^2(X, \boldsymbol{Z})$. This expression can be extended to the non-simply-connected case (see Mariño and Moore (1999) and Moore and Witten (1998)). The study of the $u$-plane integral leads to a systematic derivation of many important results in Donaldson–Witten theory. We will discuss in detail two such applications, Göttsche's wall-crossing formula and Witten's "magic formula."

### Wall-Crossing Formula

As shown by Moore and Witten, the $u$-plane integral is well defined and does not depend on the period point (hence on the metric on $X$) except for discontinuous behavior at walls. There are two kinds of walls, associated, respectively, to the singularities at $u = \infty$ (the semiclassical region of the underlying Yang–Mills theory) and at $u = \pm 1$, given by

$$u = \infty\colon\ \lambda_+ = 0, \lambda \in H^2(X, \boldsymbol{Z}) + \tfrac{1}{2} w_2(V)$$
$$u = \pm 1\colon\ \lambda_+ = 0, \lambda \in H^2(X, \boldsymbol{Z}) + \tfrac{1}{2} w_2(X) \qquad [33]$$

The first type of walls is precisely the one that appears in Donaldson theory on manifolds of $b_2^+(X) = 1$. The discontinuity of the $u$-plane integral at these walls can be easily computed from eqn [33]:

$$\mathrm{WC}_{\zeta=2\lambda}(p, S)$$
$$= -\frac{\mathrm{i}}{2}(-1)^{(\lambda-\lambda_0, w_2(X))}\mathrm{e}^{2\pi\mathrm{i}\lambda_0^2}\Big[q^{-\lambda^2/2} h_\infty(\tau)^{-2}\vartheta_4^\sigma f_\infty^{-1}$$
$$\times \exp\{2pu_\infty + S^2 T_\infty - \mathrm{i}(\lambda, S)/h_\infty\}\Big]_{q^0} \qquad [34]$$

This expression involves the modular forms $h_\infty, f_{2\infty}, u_\infty$, and $T_\infty$ (the subscript $\infty$ refers to the fact that they are computed at the "electric" frame which is appropriate for the Seiberg–Witten curve at $u \to \infty$). They can be written in terms of Jacobi theta functions $\vartheta_i(q)$, with $q = \mathrm{e}^{2\pi\mathrm{i}\tau}$, and their explicit expression is

$$h_\infty(q) = \tfrac{1}{2}\vartheta_2(q)\vartheta_3(q)$$
$$f_\infty(q) = \frac{\vartheta_2(q)\vartheta_3(q)}{2\vartheta_4^8(q)}$$
$$u_\infty(q) = \frac{\vartheta_2^4(q) + \vartheta_3^4(q)}{2(\vartheta_2(q)\vartheta_3(q))^2} \qquad [35]$$
$$T_\infty(q) = -\frac{1}{24}\frac{E_2(q)}{h_\infty^2(q)} + \frac{1}{3}u_\infty(q)$$

The subindex $q^0$ means that in the expansion in $q$ of the modular forms, we pick the constant term. The formula [34] agrees with the formula of Göttsche (1996) for the wall crossing of the Donaldson–Witten generating functional.

### The Seiberg–Witten Contribution and Witten's Magic Formula

At $u = \pm 1, Z_u$ jumps at the second type of walls [33], which are called Seiberg–Witten (SW) walls. In fact, these walls are labeled by classes $\lambda \in H^2(X; \boldsymbol{Z}) + (1/2)w_2(X)$, which correspond to Spin$_c$ structures on $X$. At these walls, the Seiberg–Witten invariants have wall-crossing behavior. Since the Donaldson polynomials do not jump at SW walls, it must happen that the change of $Z_u$ at $u = \pm 1$ is canceled by the change of $Z_{\mathrm{SW}}$. As shown by Moore and Witten, this actually allows one to obtain a precise expression for $Z_{\mathrm{SW}}$ for general 4-manifolds of $b_2^+(X) \geq 1$.

On general grounds, $Z_{\mathrm{SW}}$ is given by the sum of the generating functionals at $u = \pm 1$. These involve a magnetic $\mathrm{U}(1), \mathcal{N} = 2$ vector multiplet coupled to a hypermultiplet (the monopole field). The twisted

Lagrangian for such a system involves the magnetic prepotential $\widetilde{\mathcal{F}}_D(a_D)$, and it can be written as

$$\{\bar{\mathcal{Q}}, W\} + \frac{\mathrm{i}}{16\pi}\tilde{\tau}_D F\wedge F + p(u)\mathrm{tr}R\wedge *R + \ell(u)\mathrm{tr}R\wedge R - \frac{\mathrm{i}\sqrt{2}}{32\pi}\frac{\mathrm{d}\tilde{\tau}_D}{\mathrm{d}a_D}(\psi\wedge\psi)\wedge F + \frac{\mathrm{i}}{3\times 2^7\pi}\frac{\mathrm{d}^2\tilde{\tau}_D}{\mathrm{d}a_D^2}\psi\wedge\psi\wedge\psi\wedge\psi \quad [36]$$

where $\tilde{\tau}_D = \widetilde{\mathcal{F}}_D''(a_D)$. Using the cancellation of wall crossings, one can actually compute the functions $\widetilde{\mathcal{F}}_D(a_D), p(u), \ell(u)$ and determine the precise form of the Seiberg–Witten contributions. One finds that a $\mathrm{Spin}_c$ structure $\lambda$ at $u=1$ gives the following contribution to the Donaldson–Witten generating functional:

$$Z_{\mathrm{SW}}^{u=1,\lambda} = \frac{\mathrm{SW}(\lambda)}{16}\mathrm{e}^{2\mathrm{i}\pi(\lambda_0^2-\lambda_0\cdot\lambda)} \times\left[q_D^{-\lambda^2/2}\frac{\vartheta_2^{8+\sigma}}{a_D h_M}\left(-2\mathrm{i}\frac{a_D}{h_M^2}\right)^{\chi_{\mathrm{h}}} \times\exp\left(2pu_M + \mathrm{i}(\lambda,S)/h_M + S^2 T_M\right)\right]_{q_D^0} \quad [37]$$

Here, $a_D, h_M, u_M$, and $T_M$ are modular forms that can be expressed as well in terms of Jacobi theta functions $\vartheta_i(q_D)$, where $q_D = \exp(2\pi\mathrm{i}\tau_D)$. The subscript $M$ refers to the monopole point, and they are related by an $S$-transformation to the quantities obtained in the "electric" frame at $u\to\infty$. Their explicit expression is

$$\begin{aligned} a_D(q_D) &= -\frac{\mathrm{i}}{6}\frac{2E_2(q_D)-\vartheta_3^4(q_D)-\vartheta_4^4(q_D)}{\vartheta_3(q_D)\vartheta_4(q_D)} \\ h_M(q_D) &= \frac{1}{2\mathrm{i}}\vartheta_3(q_D)\vartheta_4(q_D) \\ u_M(q_D) &= \frac{1}{2}\frac{\vartheta_3^4(q_D)+\vartheta_4^4(q_D)}{(\vartheta_3(q_D)\vartheta_4(q_D))^2} \\ T_M(q) &= -\frac{1}{24}\frac{E_2(q_D)}{h_M^2(q_D)} + \frac{1}{3}u_M(q_D) \end{aligned} \quad [38]$$

The contribution at $u=-1$ is related to the contribution at $u=1$ by a $u\to -u$ symmetry:

$$Z_{u=-1}(p,S) = \mathrm{e}^{-2\pi\mathrm{i}\lambda_0^2}\mathrm{i}^{(\chi+\sigma)/4} Z_{u=1}(-p,-\mathrm{i}S) \quad [39]$$

If the manifold has $b_2^+(X) > 1$ and is of Seiberg–Witten simple type, [37] reduces to

$$(-1)^{\chi_{\mathrm{h}}} 2^{1+7\chi/4+11\sigma/4}\mathrm{e}^{2p+S^2/2}\mathrm{e}^{-2(S,\lambda)} \times \mathrm{e}^{2\mathrm{i}\pi(\lambda_0^2-\lambda_0\cdot\lambda)}\mathrm{SW}(\lambda) \quad [40]$$

This leads to Witten's "magic formula" [25] which expresses the Donaldson invariants in terms of Seiberg–Witten invariants.

### Other Applications of the $u$-Plane Integral

The $u$-plane integral makes possible to derive other results on the Donaldson–Witten generating functional.

*The blow-up formula*. This relates the function $Z_{\mathrm{DW}}$ on $X$ to $Z_{\mathrm{DW}}$ on the blown-up manifold $\widehat{X}$. The $u$-plane integral leads directly to the general blow-up formula of Fintushel and Stern (1996).

*Direct evaluations*. The $u$-plane integral can be evaluated directly in many cases, and this leads to explicit formulas for the Donaldson–Witten generating functional of certain 4-manifolds with $b_2^+(X)=1$, on certain chambers, and in terms of modular forms. For example, there are explicit formulas for the Donaldson–Witten generating functional of product ruled surfaces of the form $S^2\times\Sigma_g$ in the limiting chambers in which $S^2$ or $\Sigma_g$ are very small (Moore and Witten 1998, Mariño and Moore 1999). Moore and Witten (1998) have also derived an explicit formula for the Donaldson invariants of $\boldsymbol{CP}^2$ in terms of Hurwitz class numbers.

## Extensions of Donaldson–Witten Theory

Donaldson–Witten theory is a twisted version of SU(2), $\mathcal{N}=2$ Yang–Mills theory. The twisting of more general $\mathcal{N}=2$ gauge theories, involving other gauge groups and/or matter content, leads to other topological field theories that give interesting generalizations of Donaldson–Witten theory. We now briefly list some of these extensions and their most important properties.

*Higher-rank theories*. The extension of Donaldson–Witten to other gauge groups has been studied in detail in Mariño and Moore (1998b) and Losev *et al.* (1998). One can study the higher-rank generalization of the $u$-plane integral, and as shown in Mariño and Morre (1998b), this leads to a fairly explicit formula for the Donaldson–Witten generating function in the SU($N$) case, for manifolds with $b_2^+ > 1$ and of Seiberg–Witten simple type. Mathematically, higher-rank generalizations of Donaldson theory turn out to be much more complicated, but they can be studied. In particular, higher-rank generalizations of the Donaldson invariants can be defined and computed (Kronheimer 2004), and the results so far agree with the predictions of Mariño and Moore (1998b). Unfortunately it seems that these higher-rank generalizations do not contain new topological information, besides the one encoded in the Seiberg–Witten invariants.

*Theories with matter*. Twisted SU(2), $\mathcal{N}=2$ theories with hypermultiplets lead to generalizations of Donaldson–Witten theory involving nonabelian

monopole equations (see Mariño (1997) and Labastida and Mariño (2005) for a review of these models and some of their properties). The $u$-plane integral leads to explicit formulas for the generating functionals of these theories, which for manifolds of $b_2^+ > 1$ can be written in terms of Seiberg–Witten invariants. Again, no new topological information seems to be encoded in these theories. One can however exploit new physical phenomena arising in the theories with hypermultiplets (in particular, the presence of superconformal points) to obtain new information about the Seiberg–Witten invariants (see Mariño *et al.* (1999) for these developments).

*Vafa–Witten theory*. The so-called Vafa–Witten theory is a close cousin of Donaldson–Witten theory, and was introduced by Vafa and Witten (1994) as a topological twist of $\mathcal{N}=4$ Yang–Mills theory. In some cases, the partition function of this theory counts the Euler characteristic of the moduli space of instantons on the 4-manifold $X$. For a review of some properties of this theory, see Lozano (1999).

*See also:* Duality in Topological Quantum Field Theory; Mathai–Quillen Formalism; Seiberg–Witten Theory; Topological Quantum Field Theory: Overview.

## Further Reading


Birmingham D, Blau M, Rakowski M, and Thompson G (1991) Topological field theories. *Physics Reports* 209: 129.

Cordes S, Moore G, and Rangoolam S (1994) Lectures on 2D Yang–Mills theory, equivariant cohomology and topological field theory (arXiv:hep-th/9411210).

Donaldson SK (1990) Polynomial invariants for smooth four-manifolds. *Topology* 29: 257.

Donaldson SK and Kronheimer PB (1990) *The Geometry of Four-Manifolds*. Oxford: Clarendon.

Fintushel R and Stern RJ (1996) The blowup formula for Donaldson invariants. *Annals of Mathematics* 143: 529 (arXiv:alg-geom/9405002).

Friedman R and Morgan JW (1991) *Smooth Four-Manifolds and Complex Surfaces*. New York: Springer.

Göttsche L (1996) Modular forms and Donaldson invariants for four-manifolds with $b_+ = 1$. *Journal of American Mathematical Society* 9: 827 (arXiv:alg-geom/9506018).

Kronheimer P (2004) Four-manifold invariants from higher-rank bundles (arXiv:math.GT/0407518).

Kronheimer P and Mrowka T (1994) Recurrence relations and asymptotics for four-manifolds invariants. *Bulletin of the American Mathematical Society* 30: 215.

Kronheimer P and Mrowka T (1995) Embedded surfaces and the structure of Donaldson's polynomials invariants. *Journal of Differential Geometry* 33: 573.

Labastida J and Mariño M (2005) *Topological Quantum Field Theory and Four-Manifolds*. New York: Springer.

Losev A, Nekrasov N, and Shatashvili S (1998) Issues in topological gauge theory. *Nuclear Physics B* 534: 549 (arXiv:hep-th/9711108).

Lozano C (1999) Duality in topological quantum field theories (arXiv:hep-th/9907123).

Mariño M (1997) The geometry of supersymmetric gauge theories in four dimensions (arXiv:hep-th/9701128).

Mariño M and Moore G (1998a) Integrating over the Coulomb branch in $\mathcal{N}=2$ gauge theory. *Nuclear Physics Proceedings Supplement* 68: 336 (arXiv:hep-th/9712062).

Mariño M and Moore GW (1998b) The Donaldson–Witten function for gauge groups of rank larger than one. *Communications in Mathematical Physics* 199: 25 (arXiv:hep-th/9802185).

Mariño M and Moore GW (1999) Donaldson invariants for non-simply connected manifolds. *Communications in Mathematical Physics* 203: 249 (arXiv:hep-th/9804104).

Mariño M, Moore GW, and Peradze G (1999) Superconformal invariance and the geography of four-manifolds. *Communications in Mathematical Physics* 205: 691 (arXiv:hep-th/9812055).

Moore G and Witten E (1998) Integration over the $u$-plane in Donaldson theory. *Advances in Theoretical and Mathematical Society* 1: 298 (arXiv:hep-th/9709193).

Seiberg N and Witten E (1994a) Electric – magnetic duality, monopole condensation, and confinement in $\mathcal{N}=2$ supersymmetric Yang–Mills theory. *Nuclear Physics B* 426: 19.

Seiberg N and Witten E (1994b) Electric – magnetic duality, monopole condensation, and confirement in $\mathcal{N}=2$ supersymmetric Yang–Mills theory – erratum. *Nuclear Physics B* 430: 485 (arXiv:hep-th/9407087).

Stern R (1998) Computing Donaldson invariants. In: *Gauge Theory and the Topology of Four-Manifolds*, IAS/Park City Mathematical Series. Providence, RI: American Mathematical Society.

Vafa C and Witten E (1994) A strong coupling test of S duality. *Nuclear Physics B* 431: 3 (arXiv:hep-th/9408074).

Witten E (1988) Topological quantum field theory. *Communications in Mathematical Physics* 117: 353.

Witten E (1994a) Supersymmetric Yang–Mills theory on a four manifold. *Journal of Mathematical Physics* 35: 5101 (arXiv:hep-th/9403195).

Witten E (1994b) Monopoles and four manifolds. *Mathematical Research Letters* 1: 769 (arXiv:hep-th/9411102).

Witten E (1995) On S duality in Abelian gauge theory. *Selecta Mathematica* 1: 383 (arXiv:hep-th/9505186).

# Duality in Topological Quantum Field Theory

**C Lozano**, INTA, Madrid, Spain
**J M F Labastida**, CSIC, Madrid, Spain



## Introduction

There have been many exciting interactions between physics and mathematics in the past few decades. A prominent role in these interactions has been played by certain field theories, known as topological quantum field theories (TQFTs). These are quantum field theories whose correlation functions are metric independent and, in fact, compute certain mathematical invariants (Birmingham *et al.* 1991, Cordes *et al.* 1996, Labastida and Lozano 1998).

Well-known examples of TQFTs are, in two dimensions, the topological sigma models (Witten 1988a), which are related to Gromov–Witten invariants and enumerative geometry; in three dimensions, Chern–Simons theory (Witten 1989), which is related to knot and link invariants; and in four dimensions, topological Yang–Mills theory (or Donaldson–Witten theory) (Witten 1988b), which is related to the Donaldson invariants. The two- and four-dimensional theories above are examples of cohomological (also Witten-type) TQFTs. As such, they are related to an underlying supersymmetric quantum field theory (the $\mathcal{N}=2$ nonlinear sigma model, and the $\mathcal{N}=2$ supersymmetric Yang–Mills theory, respectively) and there is no difference between the topological and the standard version on flat space. However, when one considers curved spaces, the topological version differs from the supersymmetric theory on flat space in that some of the fields have modified Lorentz transformation properties (spins). This unconventional spin assignment is also known as twisting, and it comes about basically to preserve supersymmetry on curved space. In fact, the twisting gives rise to at least one nilpotent scalar supercharge $\mathcal{Q}$, which is a certain linear combination of the original (spinor) supersymmetry generators.

In these theories the energy momentum tensor is $\mathcal{Q}$-exact, that is,

$$T_{\mu\nu} = \{\mathcal{Q}, \Lambda_{\mu\nu}\}$$

for some $\Lambda_{\mu\nu}$, which (barring potential anomalies) leads to the statement that the correlation functions of operators in the cohomology of $\mathcal{Q}$ are all metric independent. Furthermore, the corresponding path integrals are localized to field configurations that are annihilated by $\mathcal{Q}$, and this typically leads to some moduli problem related to the computation of certain mathematical invariants.

On the other hand, in Chern–Simons theory, as a representative of the so-called Schwarz-type topological theories, the topological character is manifest: one starts with an action which is explicitly independent of the metric on the 3-manifold, and thus correlation functions of metric-independent operators are topological invariants as long as quantization does not introduce any undesired metric dependence.

Even though the primary motivation for introducing TQFTs may be to shed light onto awkward mathematical problems, they have proved to be a valuable tool to gain insight into many questions of interest in physics as well. One such question where TQFTs can (and in fact do) play a role is duality. In what follows, an overview of the manifestations of duality is provided in the context of TQFTs.

## Duality

The notion of duality is at the heart of some of the most striking recent breakthroughs in physics and mathematics. In broad terms, a duality (in physics) is an equivalence between different (and often complementary) descriptions of the same physical system. The prototypical example is electric–magnetic (abelian) duality. Other, more sophisticated, examples are the various string-theory dualities, such as $T$-duality (and its more specialized realization, mirror symmetry) and strong/weak coupling $S$-duality, as well as field theory dualities such as Montonen–Olive duality and Seiberg–Witten effective duality.

Also, the original 't Hooft conjecture, stating that SU($N$) gauge theories are equivalent (or dual), at large $N$, to string theories, has recently been revived by Maldacena (1998) by explicitly identifying the string-theory duals of certain (supersymmetric) gauge theories.

One could wonder whether similar duality symmetries work for TQFTs as well. As noted in the following, this is indeed the case.

In two dimensions, topological sigma models come under two different versions, known as types A and B, respectively, which correspond to the two different ways in which $\mathcal{N}=2$ supersymmetry can be twisted in two dimensions. Computations in each model localize on different moduli spaces and, for a given target manifold, give different results, but it turns out that if one considers mirror pairs of Calabi–Yau manifolds,

computations in one manifold with the A-model are equivalent to computations in the mirror manifold with the B-model.

Also, in three dimensions, a program has been initiated to explore the duality between large $N$ Chern–Simons gauge theory and topological strings, thereby establishing a link between enumerative geometry and knot and link invariants (Gopakumar and Vafa 1998).

Perhaps the most impressive consequences of the interplay between duality and TQFTs have come out in four dimensions, on which we will focus in what follows.

## Duality in Twisted $\mathcal{N}=2$ Theories

As mentioned above, topological Yang–Mills theory (or Donaldson–Witten theory) can be constructed by twisting the pure $\mathcal{N}=2$ supersymmetric Yang–Mills theory with gauge group SU(2). This theory contains a gauge field $A$, a pair of chiral spinors $\lambda_1$, $\lambda_2$, and a complex scalar field $B$. The twisted theory contains a gauge field $A$, bosonic scalars $\lambda$, $\phi$, a Grassman-odd scalar $\eta$, a Grassman-odd vector $\psi$, and a Grassman-odd self-dual 2-form $\chi$.

On a 4-manifold $X$, and for gauge group $G$, the twisted action has the form

$$\begin{aligned}\mathcal{S} = \int_X \mathrm{d}^4x\sqrt{g}\,\mathrm{tr}\Big(& F^{+^2} - \mathrm{i}\chi^{\mu\nu}D_\mu\psi_\nu + \mathrm{i}\eta D_\mu\psi^\mu \\ &+\frac{1}{4}\phi\{\chi_{\mu\nu},\chi^{\mu\nu}\} + \frac{\mathrm{i}}{4}\lambda\{\psi_\mu,\psi^\mu\} - \lambda D_\mu D^\mu\phi \\ &+\frac{\mathrm{i}}{2}\phi\{\eta,\eta\} + \frac{1}{8}[\lambda,\phi]^2\Big)\end{aligned} \qquad [1]$$

where $F^+$ is the self-dual part of the Yang–Mills field strength $F$. The action [1] is invariant under the transformations generated by the scalar supercharge $\mathcal{Q}$:

$$\begin{aligned}\{\mathcal{Q},A_\mu\} &= \psi_\mu, & \{\mathcal{Q},\chi_{\mu\nu}\} &= F^+_{\mu\nu} \\ \{\mathcal{Q},\psi\} &= \mathrm{d}_A\phi, & \{\mathcal{Q},\eta\} &= \mathrm{i}[\lambda,\phi] \\ \{\mathcal{Q},\phi\} &= 0, & \{\mathcal{Q},\lambda\} &= \eta\end{aligned} \qquad [2]$$

In these transformations, $\mathcal{Q}^2$ is a gauge transformation with gauge parameter $\phi$, modulo field equations. Observables are, therefore, related to the $G$-equivariant cohomology of $\mathcal{Q}$ (i.e., the cohomology of $\mathcal{Q}$ restricted to gauge invariant operators). Auxiliary fields can be introduced so that the action [1] is $\mathcal{Q}$-exact, that is,

$$\mathcal{S} = \{\mathcal{Q},\Lambda\} \qquad [3]$$

for $\Lambda$ a certain functional of the fields of the theory which comes under the name of gauge fermion, a BRST-inspired terminology which reflects the formal resemblance of topological cohomological field theories with some aspects of the BRST approach to the quantization of gauge theories. Before constructing the topological observables of the theory, we begin by pointing out that for each independent Casimir of the gauge group $G$ it is possible to construct an operator $W_0$, from which operators $W_i$ can be defined recursively through the descent equations $\{\mathcal{Q},W_i\}=\mathrm{d}W_{i-1}$. For example, for the quadratic Casimir,

$$W_0 = \frac{1}{8\pi^2}\mathrm{tr}(\phi^2) \qquad [4]$$

which generates the following family of operators:

$$\begin{aligned}W_1 &= \frac{1}{4\pi^2}\mathrm{tr}(\phi\psi) \\ W_2 &= \frac{1}{4\pi^2}\mathrm{tr}\left(\frac{1}{2}\psi\wedge\psi + \phi\wedge F\right) \\ W_3 &= \frac{1}{4\pi^2}\mathrm{tr}(\psi\wedge F)\end{aligned} \qquad [5]$$

Using these one defines the following observables:

$$\mathcal{O}^{(k)} = \int_{\gamma_k} W_k \qquad [6]$$

where $\gamma_k \in H_k(X)$ is a $k$-cycle on the 4-manifold $X$. The descent equations imply that they are $\mathcal{Q}$-closed and depend only on the homology class of $\gamma_k$.

Topological invariants are constructed by taking vacuum expectation values of products of the operators $\mathcal{O}^{(k)}$:

$$\begin{aligned}&\left\langle \mathcal{O}^{(k_1)}\mathcal{O}^{(k_2)}\cdots\mathcal{O}^{(k_p)}\right\rangle \\ &\quad = \int \mathcal{O}^{(k_1)}\mathcal{O}^{(k_2)}\cdots\mathcal{O}^{(k_p)}\,\mathrm{e}^{-\mathcal{S}/e^2}\end{aligned} \qquad [7]$$

where the integration has to be understood on the space of field configurations modulo gauge transformations, and $e$ is a coupling constant. Standard arguments show that due to the $\mathcal{Q}$-exactness of the action $\mathcal{S}$, the quantities obtained in [7] are independent of $e$. This implies that the observables of the theory can be obtained either in the weak-coupling limit $e\to 0$ (also short-distance or ultraviolet regime, since the $\mathcal{N}=2$ theory is asymptotically free), where perturbative methods apply, or in the strong-coupling (also long-distance or infrared) limit $e\to\infty$, where one is forced to consider a nonperturbative approach.

In the weak-coupling limit one proves that the correlation functions [7] descend to polynomials in the product cohomology of the moduli space of anti-self-dual (ASD) instantons $H_{k_1}(\mathcal{M}_{\mathrm{ASD}})\times H_{k_2}(\mathcal{M}_{\mathrm{ASD}})\times\cdots\times H_{k_p}(\mathcal{M}_{\mathrm{ASD}})$, which are precisely

the Donaldson polynomial invariants of $X$. However, the weak- coupling analysis does not add any new ingredient to the problem of the actual computation of the invariants. The difficulties that one has to face in the field theory representation are similar to those in ordinary Donaldson theory.

Nevertheless, the field theory connection is very important since in this theory the strong- and weak-coupling limits are exact, and therefore the door is open to find a strong-coupling description which could lead to a new, simpler representation for the Donaldson invariants.

This alternative strategy was pursued by Witten (1994a), who found the strong-coupling realization of the Donaldson–Witten theory after using the results on the strong-coupling behavior of $\mathcal{N}=2$ supersymmetric gauge theories which he and Seiberg (Seiberg and Witten 1994a–c) had discovered. The key ingredient in Witten's derivation was to assume that the strong-coupling limit of Donaldson–Witten theory is equivalent to the "sum" over the twisted effective low-energy descriptions of the corresponding $\mathcal{N}=2$ physical theory. This "sum" is not entirely a sum, as in general it has a part which contains a continuous integral. The "sum" is now known as integration over the $u$-plane after the work of Moore and Witten (1998). Witten's (1994a) assumption can be simply stated as saying that the weak-/strong-coupling limit and the twist commute. In other words, to study the strong-coupling limit of the topological theory, one first untwists, then works out the strong-coupling limit of the physical theory and, finally, one twists back. From such a viewpoint, the twisted effective (strong-coupling) theory can be regarded as a TQFT dual to the original one. In addition, one could ask for the dual moduli problem associated to this dual TQFT. It turns out that in many interesting situations ($b_2^+(X)>1$) the dual moduli space is an abelian system corresponding to the Seiberg–Witten or monopole equations (Witten 1994a). The topological invariants associated with this new moduli space are the celebrated Seiberg–Witten invariants.

Generalizations of Donaldson–Witten theory, with either different gauge groups and/or additional matter content (such as, e.g., twisted $\mathcal{N}=2$ Yang–Mills multiplets coupled to twisted $\mathcal{N}=2$ matter multiplets) are possible, and some of the possibilities have in fact been explored (see Moore and Witten (1998) and references therein). The main conclusion that emerges from these analyses is that, in all known cases, the relevant topological information is captured by the Seiberg–Witten invariants, irrespectively of the gauge group and matter content of the theory under consideration. These cases are not reviewed here, but rather the attention is turned to the twisted theories which emerge from $\mathcal{N}=4$ supersymmetric gauge theories.

## Duality in Twisted $\mathcal{N}=4$ Theories

Unlike the $\mathcal{N}=2$ supersymmetric case, the $\mathcal{N}=4$ supersymmetric Yang–Mills theory in four dimensions is unique once the gauge group $G$ is fixed. The microscopic theory contains a gauge or gluon field, four chiral spinors (the gluinos) and six real scalars. All these fields are massless and take values in the adjoint representation of the gauge group. The theory is finite and conformally invariant, and is conjectured to have a duality symmetry exchanging strong and weak coupling and exchanging electric and magnetic fields, which extends to a full SL(2, $\mathbb{Z}$) symmetry acting on the microscopic complexified coupling (Montonen and Olive 1977)

$$\tau=\frac{\theta}{2\pi}+\frac{4\pi \mathrm{i}}{e^2} \tag{8}$$

As in the $\mathcal{N}=2$ case, the $\mathcal{N}=4$ theory can be twisted to obtain a topological model, only that, in this case, the topological twist can be performed in three inequivalent ways, giving rise to three different TQFTs (Vafa and Witten 1994). A natural question to answer is whether the duality properties of the $\mathcal{N}=4$ theory are shared by its twisted counterparts and, if so, whether one can take advantage of the calculability of topological theories to shed some light on the behavior and properties of duality.

The answer is affirmative, but it is instructive to clarify a few points. First, as mentioned above, the topological observables in twisted $\mathcal{N}=2$ theories are independent of the coupling constant $e$, so the question arises as to how the twisted $\mathcal{N}=4$ theories come to depend on the coupling constant. As it turns out, twisted $\mathcal{N}=2$ supersymmetric gauge theories have an off-shell formulation such that the TQFT action can be expressed as a $\mathcal{Q}$-exact expression, where $\mathcal{Q}$ is the generator of the topological symmetry. Actually, this is true only up to a topological $\theta$-term $\int_X \mathrm{tr}(F\wedge F)$,

$$\begin{aligned}\mathcal{S}=&\frac{1}{2e^2}\int_X \sqrt{g}\,\mathrm{d}^4x\{\mathcal{Q},\Lambda\}\\&-2\pi \mathrm{i}\tau\frac{1}{16\pi^2}\int_X \mathrm{tr}(F\wedge F)\end{aligned} \tag{9}$$

for some $\Lambda$. However, the $\mathcal{N}=2$ supersymmetric gauge theories possess a global U(1) chiral symmetry which is generically anomalous, so one can actually

get rid of the $\theta$-term with a chiral rotation. As a result of this, the observables in the topological theory are insensitive to $\theta$-terms (and hence to $\tau$ and $e$) up to a rescaling.

On the other hand, in $\mathcal{N}=4$ supersymmetric gauge theories $\theta$-terms are observable. There is no chiral anomaly and these terms cannot be shifted away as in the $\mathcal{N}=2$ case. This means that in the twisted theories one might have a dependence on the coupling constant $\tau$, and that – up to anomalies – this dependence should be holomorphic (resp. antiholomorphic if one reverses the orientation of the 4-manifold). In fact, on general grounds, one would expect for the partition functions of the twisted theories on a 4-manifold $X$ and for gauge group $G$ to take the generic form

$$Z_X(G) = q^{-c(X,G)} \sum_k q^k \chi(\mathcal{M}_k) \qquad [10]$$

where $q = e^{2\pi i \tau}$, $c$ is a universal constant (depending on $X$ and $G$), $k = (1/16\pi^2)\int_X \mathrm{tr}(F \wedge F)$ is the instanton number, and $\chi(\mathcal{M}_k)$ encodes the topological information corresponding to a sector of the moduli space of the theory with instanton number $k$.

Now we can be more precise as to how we expect to see the Montonen–Olive duality in the twisted $\mathcal{N}=4$ theories. First, under $\tau \to -1/\tau$ the gauge group $G$ gets exchanged with its dual group $\hat{G}$. Correspondingly, the partition functions should behave as modular forms

$$Z_G(-1/\tau) = \pm\kappa(X, G)\tau^w Z_{\hat{G}}(\tau) \qquad [11]$$

where $\kappa$ is a constant (depending on $X$ and $G$), and the modular weight $w$ should depend on $X$ in such a way that it vanishes on flat space.

In addition to this, in the $\mathcal{N}=4$ theory all the fields take values in the adjoint representation of $G$. Hence, if $H^2(X, \pi_1(G)) \neq 0$, it is possible to consider nontrivial $G$/Center($G$) gauge configurations with discrete magnetic 't Hooft flux through the 2-cycles of $X$. In fact, $G$/Center($G$) bundles on $X$ are classified by the instanton number and a characteristic class $v \in H^2(X, \pi_1(G))$. For example, if $G = \mathrm{SU}(2)$, we have $\hat{G} = \mathrm{SU}(2)/\mathbb{Z}_2 = \mathrm{SO}(3)$ and $v$ is the second Stiefel–Whitney class $w_2(E)$ of the gauge bundle $E$. This Stiefel–Whitney class can be represented in de Rham cohomology by a class in $H^2(X, \mathbb{Z})$ defined modulo 2, that is, $w_2(E)$ and $w_2(E) + 2\omega$, with $\omega \in H^2(X, \mathbb{Z})$, represent the same 't Hooft flux, so if $w_2(E) = 2\lambda$, for some $\lambda \in H^2(X, \mathbb{Z})$, then the gauge configuration is trivial in SO(3) (it has no 't Hooft flux).

Similarly, for $G = \mathrm{SU}(N)$ (for which $\hat{G} = \mathrm{SU}(N)/\mathbb{Z}_N$), one can fix fluxes in $H^2(X, \mathbb{Z}_N)$ (the corresponding Stiefel–Whitney class is defined modulo $N$). One has, therefore, a family of partition functions $Z_v(\tau)$, one for each magnetic flux $v$. The SU($N$) partition function is obtained by considering the zero flux partition function (up to a constant factor), while the (dual) $\mathrm{SU}(N)/\mathbb{Z}_N$ partition function is obtained by summing over all $v$, and both are to be exchanged under $\tau \to -1/\tau$. The action of SL(2, $\mathbb{Z}$) on the $Z_v$ should be compatible with this exchange, and thus the $\tau \to -1/\tau$ operation mixes the $Z_v$ by a discrete Fourier transform which, for $G = \mathrm{SU}(N)$ reads

$$Z_v(-1/\tau) = \pm\kappa(X, G)\tau^w \sum_u e^{2\pi i u \cdot v/N} Z_u(\tau) \qquad [12]$$

We are now in a position to examine the (three) twisted theories in some detail. For further details and references, the reader is referred to Lozano (1999).

The first twisted theory considered here possesses only one scalar supercharge (and hence comes under the name of "half-twisted theory"). It is a nonabelian generalization of the Seiberg–Witten abelian monopole theory, but with the monopole multiplets taking values in the adjoint representation of the gauge group. The theory can be perturbed by giving masses to the monopole multiplets while still retaining its topological character. The resulting theory is the twisted version of the mass-deformed $\mathcal{N}=4$ theory, which preserves $\mathcal{N}=2$ supersymmetry and whose low-energy effective description is known. This connection with $\mathcal{N}=2$ theories, and its topological character, makes it possible to go to the long-distance limit and compute in terms of the twisted version of the low-energy effective description of the supersymmetric theory. Below, we review how the $u$-plane approach works for gauge group SU(2).

The twisted theory for gauge group SU(2) has a U(1) global symmetry (the ghost number) which has an anomaly $-3(2\chi + 3\sigma)/4$ on gravitational backgrounds (i.e., on curved manifolds). Nontrivial topological invariants are thus obtained by considering the vacuum expectation value of products of observables with ghost numbers adding up to $-3(2\chi + 3\sigma)/4$. The relevant observables for this theory and gauge group SU(2) or SO(3) are precisely the same as in the Donaldson–Witten theory (eqns [4] and [5]). In addition to this, it is possible to enrich the theory by including sectors with nontrivial nonabelian electric and magnetic 't Hooft fluxes which, as pointed out above, should behave under SL(2, $\mathbb{Z}$) duality in a well-defined fashion.

The generating function for these correlation functions is given as an integration over the moduli space of vacua of the physical theory (the $u$-plane), which, for generic values of the mass parameter, forms a one-dimensional complex compact manifold (described by a complex variable customarily denoted by $u$, hence the name), which parametrizes a family of elliptic curves that encodes all the relevant information about the low-energy effective description of the theory. At a generic point in the moduli space of vacua, the only contribution to the topological correlation functions comes from a twisted $\mathcal{N}=2$ abelian vector multiplet. Additional contributions come from points in the moduli space where the low-energy effective description is singular (i.e., where the associated elliptic curve degenerates).

Therefore, the total contribution to the generating function thus consists of an integration over the moduli space with the singularities removed – which is nonvanishing for $b_2^+(X)=1$ (Moore and Witten 1998) only – plus a discrete sum over the contributions of the twisted effective theories at each of the three singularities of the low-energy effective description (Seiberg and Witten 1994a, b, c). The effective theory at a given singularity contains, together with the appropriate dual photon multiplet, one charged hypermultiplet, which corresponds to the state becoming massless at the singularity. The complete effective action for these massless states also contains certain measure factors and contact terms among the observables, which reproduce the effect of the massive states that have been integrated out as well as incorporate the coupling to gravity (i.e., explicit nonminimal couplings to the metric of the 4-manifold). How to determine these *a priori* unknown functions was explained in Moore and Witten (1998). The idea is as follows. At points on the $u$-plane where the (imaginary part of the) effective coupling diverges, the integral is discontinuous at anti-self-dual abelian gauge configurations. This is commonly referred to as "wall crossing." Wall crossing can take place at the singularities of the moduli space – the appropriate local effective coupling $\tau_{\text{eff}}$ diverges there – and, in the case of the asymptotically free theories, at the point at infinity – the effective electric coupling diverges owing to asymptotic freedom.

On the other hand, the final expression for the invariants can exhibit a wall-crossing behavior at most at $u\to\infty$, so the contribution to wall crossing from the integral at the singularities at finite values of $u$ must cancel against the contributions coming from the effective theories there, which also display wall-crossing discontinuities. Imposing this cancelation fixes almost completely the unknown functions in the contributions to the topological correlation functions from the singularities. The final result for the contributions from the singularities (which give the complete answer for the correlation functions when $b_2^+(X)>1$) is written explicitly and completely in terms of the fundamental periods $\mathrm{d}a/\mathrm{d}u$ (written in the appropriate local variables) and the discriminant of the elliptic curve comprising the Seiberg–Witten solution for the physical theory. For simply connected spin 4-manifolds of simple type the generating function is given by

$$
\begin{aligned}
\langle \mathrm{e}^{p\mathcal{O}+I(S)}\rangle_v &= 2^{(\nu/2+(2\chi+3\sigma)/8)} m^{-(3\nu+7\sigma)/8}(\eta(\tau))^{-12\nu} \\
&\times\Bigg\{(\kappa_1)^{\nu}\left(\frac{\mathrm{d}a}{\mathrm{d}u}\right)_1^{-(\nu+\sigma/4)} \mathrm{e}^{2pu_1+S^2T_1} \\
&\times\sum_x \delta_{[x/2],v}\, n_x\, \mathrm{e}^{(\mathrm{i}/2)(\mathrm{d}u/\mathrm{d}a)_1 x\cdot S} \\
&+2^{-b_2/2}(-1)^{\sigma/8}(\kappa_2)^{\nu}\left(\frac{\mathrm{d}a}{\mathrm{d}u}\right)_2^{-(\nu+\sigma/4)} \\
&\times \mathrm{e}^{2pu_2+S^2T_2}\sum_x(-1)^{v\cdot x/2} n_x\, \mathrm{e}^{(\mathrm{i}/2)(\mathrm{d}u/\mathrm{d}a)_2 x\cdot S} \\
&+2^{-b_2/2}\mathrm{i}^{-v^2}(\kappa_3)^{\nu}\left(\frac{\mathrm{d}a}{\mathrm{d}u}\right)_3^{-(\nu+\sigma/4)} \mathrm{e}^{2pu_3+S^2T_3} \\
&\times\sum_x(-1)^{v\cdot x/2} n_x \mathrm{e}^{(\mathrm{i}/2)(\mathrm{d}u/\mathrm{d}a)_3 x\cdot S}\Bigg\}
\end{aligned}
\qquad [13]
$$

where $x$ is a Seiberg–Witten basic class (and $n_x$ is the corresponding Seiberg–Witten invariant), $m$ is the mass parameter of the theory, $\nu=(\chi+\sigma)/4$, $v\in H^2(X,\mathbb{Z}_2)$ is a 't Hooft flux, $S$ is the formal sum $S=\Sigma_a\alpha_a\Sigma_a$ (and, correspondingly, $I(S)=\sum_a \alpha_a I(\Sigma_a)$, with $I(\Sigma_a)=\int_{\Sigma_a} W_2$), where $\{\Sigma_a\}_{a=1,\ldots,b_2(X)}$ form a basis of $H_2(X)$ and $\alpha_a$ are constant parameters, while $\eta(\tau)$ is the Dedekind function, $\kappa_i=(\mathrm{d}u/\mathrm{d}q_{\text{eff}})_{u=u_i}$ (with $q_{\text{eff}}=\exp(2\pi\mathrm{i}\tau_{\text{eff}})$, and $\tau_{\text{eff}}$ is the ratio of the fundamental periods of the elliptic curve), and the contact terms $T_i$ have the form

$$
T_i=-\frac{1}{12}\left(\frac{\mathrm{d}u}{\mathrm{d}a}\right)_i^2+E_2(\tau)\frac{u_i}{6}+\frac{m^2}{72}E_4(\tau) \qquad [14]
$$

with $E_2$ and $E_4$ the Einstein series of weights 2 and 4, respectively. Evaluating the quantities in [13] gives the final result as a function of the physical parameters $\tau$ and $m$, and of topological data of $X$ as the Euler characteristic $\chi$, the signature $\sigma$ and the basic classes $x$. The expression [13] has to be understood as a formal power series in $p$ and $\alpha_a$, whose coefficients give the vacuum expectation values of products of $\mathcal{O}=W_0$ and $I(\Sigma_a)$.

The generating function [13] has nice properties under the modular group. For the partition function $Z_v$,

$$\begin{aligned} Z_v(\tau+1) &= (-1)^{\sigma/8} \mathrm{i}^{-v^2} Z_v(\tau) \\ Z_v(-1/\tau) &= 2^{-b_2/2}(-1)^{\sigma/8}\left(\frac{\tau}{\mathrm{i}}\right)^{-\chi/2} \\ &\quad \times \sum_w (-1)^{w\cdot v} Z_w(\tau) \end{aligned} \qquad [15]$$

Also, with $Z_{\mathrm{SU}(2)} = 2^{-1} Z_{v=0}$ and $Z_{\mathrm{SO}(3)} = \sum_v Z_v$,

$$\begin{aligned} Z_{\mathrm{SU}(2)}(\tau+1) &= (-1)^{\sigma/8} Z_{\mathrm{SU}(2)}(\tau) \\ Z_{\mathrm{SO}(3)}(\tau+2) &= Z_{\mathrm{SO}(3)}(\tau) \\ Z_{\mathrm{SU}(2)}(-1/\tau) &= (-1)^{\sigma/8} 2^{-\chi/2} \tau^{-\chi/2} Z_{\mathrm{SO}(3)}(\tau) \end{aligned} \qquad [16]$$

Notice that the last of these three equations corresponds precisely to the strong–weak coupling duality transformation conjectured by Montonen and Olive (1977).

As for the correlation functions, one finds the following behavior under the inversion of the coupling:

$$\begin{aligned} &\left\langle \frac{1}{8\pi^2} \mathrm{tr}\phi^2 \right\rangle_\tau^{\mathrm{SU}(2)} = \langle \mathcal{O} \rangle_\tau^{\mathrm{SU}(2)} = \frac{1}{\tau^2} \langle \mathcal{O} \rangle_{-1/\tau}^{\mathrm{SO}(3)} \\ &\left\langle \frac{1}{8\pi^2} \int_S \mathrm{tr}(2\phi F + \psi \wedge \psi) \right\rangle_\tau^{\mathrm{SU}(2)} = \langle I(S) \rangle_\tau^{\mathrm{SU}(2)} \\ &\quad = \frac{1}{\tau^2} \langle I(S) \rangle_{-1/\tau}^{\mathrm{SO}(3)} \\ &\langle I(S) I(S) \rangle_\tau^{\mathrm{SU}(2)} = \left(\frac{\tau}{\mathrm{i}}\right)^{-4} \langle I(S) I(S) \rangle_{-1/\tau}^{\mathrm{SO}(3)} \\ &\qquad + \frac{\mathrm{i}}{2\pi} \frac{1}{\tau^3} \langle \mathcal{O} \rangle_{-1/\tau}^{\mathrm{SO}(3)} \sharp(S \cap S) \end{aligned} \qquad [17]$$

Therefore, as expected, the partition function of the twisted theory transforms as a modular form, while the topological correlation functions turn out to transform covariantly under SL(2, $\mathbb{Z}$), following a pattern which can be reproduced with a far more simple topological abelian model.

The second example considered next is the Vafa–Witten (1994) theory. This theory possesses two scalar supercharges, and has the unusual feature that the virtual dimension of its moduli space is exactly zero (it is an example of balanced TQFT), and therefore the only nontrivial topological observable is the partition function itself. Furthermore, the twisted theory does not contain spinors, so it is well defined on any compact, oriented 4-manifold.

Now this theory computes, with the subtleties explained in Vafa and Witten (1994), the Euler characteristic of instanton moduli spaces. In fact, in this case in the generic partition function [10],

$$Z_X(G) = q^{-c(X,G)} \sum_k q^k \chi(\mathcal{M}_k) \qquad [18]$$

$\chi(\mathcal{M}_k)$ is the Euler characteristic of a suitable compactification of the $k$th instanton moduli space $\mathcal{M}_k$ of gauge group $G$ in $X$.

As in the previous example, it is possible to consider nontrivial gauge configurations in $G$/Center ($G$) and compute the partition function for a fixed value of the 't Hooft flux $v \in H^2(X, \pi_1(G))$. In this case, however, the Seiberg–Witten approach is not available, but, as conjectured by Vafa and Witten, one can nevertheless carry out computations in terms of the vacuum degrees of freedom of the $\mathcal{N}=1$ theory which results from giving bare masses to all the three chiral multiplets of the $\mathcal{N}=4$ theory. It should be noted that a similar approach was introduced by Witten (1994b) to obtain the first explicit results for the Donaldson–Witten theory just before the far more powerful Seiberg–Witten approach was available.

As explained in detail by Vafa and Witten (1994), the twisted massive theory is topological on Kähler 4-manifolds with $h^{2,0} \neq 0$, and the partition function is actually invariant under the perturbation. In the long-distance limit, the partition function is given as a finite sum over the contributions of the discrete massive vacua of the resulting $\mathcal{N}=1$ theory. In the case at hand, it turns that, for $G = \mathrm{SU}(N)$, the number of such vacua is given by the sum of the positive divisors of $N$. The contribution of each vacuum is universal (because of the mass gap), and can be fixed by comparing with known mathematical results (Vafa and Witten 1994). However, this is not the end of the story. In the twisted theory, the chiral superfields of the $\mathcal{N}=4$ theory are no longer scalars, so the mass terms cannot be invariant under the holonomy group of the manifold unless one of the mass parameters be a holomorphic 2-form $\omega$. (Incidentally, this is the origin of the constraint $h^{(2,0)} \neq 0$ mentioned above.) This spatially dependent mass term vanishes where $\omega$ does, and we will assume as in Vafa and Witten (1994) and Witten (1994b) that $\omega$ vanishes with multiplicity 1 on a union of disjoint, smooth complex curves $C_i, i = 1, \ldots, n$ of genus $g_i$ which represent the canonical divisor $K$ of $X$. The vanishing of $\omega$ introduces corrections involving $K$ whose precise form is not known *a priori*. In the $G = \mathrm{SU}(2)$ case, each of the $\mathcal{N}=1$ vacua bifurcates along each of the components $C_i$ of the canonical divisor into two strongly coupled massive vacua. This vacuum degeneracy is believed to stem from the spontaneous breaking of a $\mathbb{Z}_2$ chiral symmetry which is unbroken in bulk (see, e.g., Vafa and Witten (1994) and Witten (1994b)).

The structure of the corrections for $G = \mathrm{SU}(N)$ (see [19] below) suggests that the mechanism at work in this case is not chiral symmetry breaking.

Indeed, near any of the $C_i$'s, there is an $N$-fold bifurcation of the vacuum. A plausible explanation for this degeneracy could be found in the spontaneous breaking of the center of the gauge group (which for $G=\mathrm{SU}(N)$ is precisely $\mathbb{Z}_N$). In any case, the formula for SU($N$) can be computed (at least when $N$ is prime) along the lines explained by Vafa and Witten (1994) and assuming that the resulting partition function satisfies a set of nontrivial constraints which are described below.

Then, for a given 't Hooft flux $v \in H^2(X, \mathbb{Z}_N)$, the partition function for gauge group SU($N$) (with prime $N$) is given by

$$Z_v = \left(\sum_{\boldsymbol{\varepsilon}} \delta_{v, w_N(\boldsymbol{\varepsilon})} \prod_{i=1}^{n} \prod_{\lambda=0}^{N-1} \left(\frac{\chi_\lambda}{\eta}\right)^{(1-g_i)\delta_{\varepsilon_i,\lambda}}\right) \times \left(\frac{1}{N^2} G(q^N)\right)^{\nu/2} + N^{1-b_1} \times \sum_{m=0}^{N-1}\left[\prod_{i=1}^{n}\left(\sum_{\lambda=0}^{N-1}\left(\frac{\chi_{m,\lambda}}{\eta}\right)^{1-g_i} \mathrm{e}^{(2\mathrm{i}\pi/N)\lambda v\cdot[C_i]_N}\right)\right] \times \mathrm{e}^{\mathrm{i}\pi((N-1)/N)mv^2}\left(\frac{G(\alpha^m q^{1/N})}{N^2}\right)^{\nu/2} \quad [19]$$

where $\alpha = \exp(2\pi \mathrm{i}/N), G(q) = \eta(q)^{24}$ (with $\eta(q)$ the Dedekind function), $\chi_\lambda$ are the SU($N$) characters at level 1 and $\chi_{m,\lambda}$ are certain linear combinations thereof. $[C_i]_N$ is the reduction modulo $N$ of the Poincaré dual of $C_i$, and

$$w_N(\boldsymbol{\varepsilon}) = \sum_{i=1}^{n} \varepsilon_i [C_i]_N \quad [20]$$

where $\varepsilon_i = 0, 1, \ldots, N-1$ are chosen independently.

Equation [19] has the expected properties under the modular group:

$$Z_v(\tau+1) = \mathrm{e}^{(\mathrm{i}\pi/12)N(2\chi+3\sigma)}\mathrm{e}^{-\mathrm{i}\pi((N-1)/N)v^2} Z_v(\tau)$$
$$Z_v(-1/\tau) = N^{-b_2/2}\left(\frac{\tau}{\mathrm{i}}\right)^{-\chi/2} \times \sum_u \mathrm{e}^{(2\mathrm{i}\pi u\cdot v/N)} Z_u(\tau) \quad [21]$$

and also, with $Z_{\mathrm{SU}(N)} = N^{b_1-1} Z_0$ and $Z_{\mathrm{SU}(N)/\mathbb{Z}_N} = \sum_v Z_v$,

$$Z_{\mathrm{SU(N)}}(-1/\tau) = N^{-\chi/2}\left(\frac{\tau}{\mathrm{i}}\right)^{-\chi/2} Z_{\mathrm{SU}(N)/\mathbb{Z}_N}(\tau) \quad [22]$$

which is, up to some correction factors that vanish in flat space, the original Montonen–Olive conjecture!

There is a further property to be checked which concerns the behavior of [19] under blow-ups. This property was heavily used by Vafa and Witten (1994) and demanding it in the present case was essential in deriving the above formula. Blowing up a point on a Kähler manifold $X$ replaces it with a new Kähler manifold $\widehat{X}$ whose second cohomology lattice is $H^2(\widehat{X}, \mathbb{Z}) = H^2(X, \mathbb{Z}) \oplus I^-$, where $I^-$ is the one-dimensional lattice spanned by the Poincaré dual of the exceptional divisor $B$ created by the blow-up. Any allowed $\mathbb{Z}_N$ flux $\widehat{v}$ on $\widehat{X}$ is of the form $\widehat{v} = v \oplus r$, where $v$ is a flux in $X$ and $r = \lambda B$, $\lambda = 0, 1, \ldots, N-1$. The main result concerning [19] is that under blowing up a point on a Kähler 4-manifold with canonical divisor as above, the partition functions for fixed 't Hooft fluxes have a factorization as

$$Z_{\widehat{X},\widehat{v}}(\tau_0) = Z_{X,v}(\tau_0)\frac{\chi_\lambda(\tau_0)}{\eta(\tau_0)} \quad [23]$$

Precisely the same behavior under blow-ups of the partition function [19] has been proved for the generating function of Euler characteristics of instanton moduli spaces on Kähler manifolds. This should not come as a surprise since, as mentioned above, on certain 4-manifolds, the partition function of Vafa–Witten theory computes the Euler characteristics of instanton moduli spaces. Therefore, [19] can be seen as a prediction for the Euler numbers of instanton moduli spaces on those 4-manifolds.

Finally, the third twisted $\mathcal{N}=4$ theory also possesses two scalar supercharges, and is believed to be a certain deformation of the four-dimensional BF theory, and as such it describes essentially intersection theory on the moduli space of complexified gauge connections. In addition to this, the theory is "amphicheiral," which means that it is invariant to a reversal of the orientation of the spacetime manifold. The terminology is borrowed from knot theory, where an oriented knot is said to be amphicheiral if, crudely speaking, it is equivalent to its mirror image. From this property, it follows that the topological invariants of the theory are completely independent of the complexified coupling constant $\tau$.

*See also:* Donaldson–Witten Theory; Electric–Magnetic Duality; Hopf Algebras and $q$-Deformation Quantum Groups; Large-$N$ and Topological Strings; Seiberg–Witten Theory; Topological Quantum Field Theory: Overview.

## Further Reading

Birmingham D, Blau M, Rakowski M, and Thompson G (1991) Topological field theories. *Physics Reports* 209: 129–340.

Cordes S, Moore G, and Rangoolam S (1996) Lectures on 2D Yang–Mills theory, equivariant cohomology and topological field theory. In: David F, Ginsparg P, and Zinn-Justin J (eds.) *Fluctuating Geometries in Statistical Mechanics and Field Theory*, Les Houches Session LXII, p. 505, hep-th/9411210. Elsevier.

Gopakumar R and Vafa C (1998) Topological gravity as large $N$ topological gauge theory. *Advances in Theoretical and Mathematical Physics* 2: 413–442, hep-th/9802016; On the gauge theory/geometry correspondence. *Advances in Theoretical and Mathematical Physics* 3: 1415–1443, hep-th/9811131.
Labastida JMF and Lozano C (1998) Lectures on topological quantum field theory. In: Falomir H, Gamboa R, and Schaposnik F (eds.) *Proceedings of the CERN-Santiago de Compostela-La Plata Meeting*, Trends in Theoretical Physics, pp. 54–93, hep-th/9709192. New York: American Institute of Physics.
Lozano C (1999) *Duality in Topological Quantum Field Theories*. Ph.D. thesis, U. Santiago de Compostela, arXiv: hep-th/ 9907123.
Maldacena J (1998) The large $N$ limit of superconformal field theories and supergravity. *Advances in Theoretical and Mathematical Physics* 2: 231–252, hep-th/9711200.
Montonen C and Olive D (1977) Magnetic monopoles as gauge particles? *Physics Letters B* 72: 117–120.
Moore G and Witten E (1998) Integration over the $u$-plane in Donaldson theory. *Advances in Theoretical and Mathematical Physics* 1: 298–387, hep-th/9709193.
Seiberg N and Witten E (1994a) Electric–magnetic duality, monopole condensation, and confinement in $\mathcal{N}=2$ supersymmetric Yang–Mills theory. *Nuclear Physics B* 426: 19–52, hep-th/9407087.
Seiberg N and Witten E (1994b) Electric–magnetic duality, monopole condensation, and confinement in $\mathcal{N}=2$ supersymmetric Yang–Mills theory. – erratum. *Nuclear Physics B* 430: 485–486.
Seiberg N and Witten E (1994c) Monopoles, duality and chiral symmetry breaking in $\mathcal{N}=2$ supersymmetric QCD. *Nuclear Physics B* 431: 484–550, hep-th/9408099.
Vafa C and Witten E (1994) A strong coupling test of $S$-duality. *Nuclear Physics B* 431: 3–77, hep-th/9408074.
Witten E (1988b) Topological quantum field theory. *Communications in Mathematical Physics* 117: 353–386.
Witten E (1988a) Topological sigma models. *Communications in Mathematical Physics* 118: 411–449.
Witten E (1989) Quantum field theory and the Jones polynomial. *Communications in Mathematical Physics* 121: 351–399.
Witten E (1994a) Monopoles and four-manifolds. *Mathematical Research Letters* 1: 769–796, hep-th/9411102.
Witten E (1994b) Supersymmetric Yang–Mills theory on a four-manifold. *Journal of Mathematical Physics* 35: 5101–5135, hep-th/9403195.

# Dynamical Systems and Thermodynamics

**A Carati, L Galgani and A Giorgilli**, Università di Milano, Milan, Italy



## Introduction

The relations between thermodynamics and dynamics are dealt with by statistical mechanics. For a given dynamical system of Hamiltonian type in a classical framework, it is usually assumed that a dynamical foundation for equilibrium statistical mechanics, namely for the use of the familiar Gibbs ensembles, is guaranteed if one can prove that the system is ergodic, that is, has no integrals of motion apart from the Hamiltonian itself. One of the main consequences is then that classical mechanics fails in explaining thermodynamics at low temperatures (e.g., the specific heats of crystals or of polyatomic molecules at low temperatures, or the black body problem), because the classical equilibrium ensembles lead to equipartition of energy for a system of weakly coupled oscillators, against Nernst's third principle. This is actually the problem that historically led to the birth of quantum mechanics, equipartition being replaced by Planck's law. At a given temperature $T$, the mean energy of an oscillator of angular frequency $\omega$ is not $k_B T$ ($k_B$ being the Boltzmann constant), and thus is not independent of frequency (equipartition), but decreases to zero exponentially fast as frequency increases.

Thus, the problem of a dynamical foundation for classical statistical mechanics would be reduced to ascertaining whether the Hamiltonian systems of physical interest are ergodic or not. It is just in this spirit that many mathematical works were recently addressed at proving ergodicity for systems of hard spheres, or more generally for systems which are expected to be not only ergodic but even hyperbolic. However, a new perspective was opened in the year 1955, with the celebrated paper of Fermi, Pasta, and Ulam (FPU), which constituted the last scientific work of Fermi.

The FPU paper was concerned with numerical computations on a system of $N$ (actually, 32 or 64) equal particles on a line, each interacting with the two adjacent ones through nonlinear springs, certain boundary conditions having been assigned (fixed ends). The model mimics a one-dimensional crystal (or also a string), and can be described in the familiar way as a perturbation of a system of $N$ normal modes, which diagonalize the corresponding linearized system. The initial conditions corresponded to the excitation of only a few low-frequency modes, and it was expected that energy would rather quickly flow to the high-frequency modes, thus establishing equipartition of energy, in agreement with the predictions of classical equilibrium statistical mechanics. But this did not occur within the available computation times, and the

energy rather appeared to remain confined within a packet of low-frequency modes having a certain width, as if being in a state of apparent equilibrium of a nonstandard type. This fact can be called "the FPU paradox." In the words of Ulam, written as a comment in *Fermi's Collected Papers*, this is described as follows: "The results of the computations were interesting and quite surprising to Fermi. He expressed the opinion that they really constituted a little discovery in providing intimations that the prevalent beliefs in the universality of mixing and thermalization in nonlinear systems may not be always justified."

The FPU paper immediately had a very strong impact on the theory of dynamical systems, because it motivated all the modern theory of infinite-dimensional integrable systems and solitons (KdV equation), starting from the works of Zabusky and Kruskal (1965). But in this way the FPU paradox was somehow enhanced, because the FPU system turned out to be associated to the class of integrable systems, namely the systems having a number of integrals of motion equal to the number of degrees of freedom, which are in a sense the most antithermodynamic systems. The merit of establishing a bridge towards ergodicity goes to Izrailev and Chirikov (1966). Making reference to the most advanced results then available in the perturbation theory for nearly integrable systems (KAM theory), these authors pointed out that ergodicity, and thus equipartition, would be recovered if one took initial data with a sufficiently large energy. And this was actually found to be the case. Moreover, it turned out that their work, and its subsequent completion by Shepelyanski, was often interpreted as supporting the conjecture that the FPU paradox would disappear in the thermodynamic limit (infinitely many particles, with finite density and energy density). The opposite conjecture was advanced in the year 1970 by Bocchieri, Scotti, Bearzi, and Loinger, and its relevance for the relations between classical and quantum mechanics was immediately pointed out by Cercignani, Galgani, and Scotti. A long debate then followed. Possibly, some misunderstandings occurred, because in the discussions concerning the dynamical aspects of the problem reference was generally made to notions involving infinite times. In fact, it had not yet been conceived that the FPU equilibrium might actually be an apparent one, corresponding to some type of intermediate metaequilibrium state. This was for the first time suggested by researchers in Parisi's group in the year 1982. The analogy of such a situation with that occurring in glasses was pointed out more recently.

In the present article, the state of the art of the FPU problem is discussed. The thesis of the present authors is that the FPU phenomenon survives in the thermodynamic limit, in the last mentioned sense, namely that at sufficiently low temperatures there exists a kind of metaequilibrium state surviving for extremely long times. The corresponding thermodynamics turns out to be different from the standard one predicted by the equilibrium ensembles, inasmuch as it presents qualitatively some quantum-like features (typically, specific heats in agreement with Nernst's third principle). The key point, with respect to equilibrium statistical mechanics, is that the internal thermodynamic energy should be identified not with the whole mechanical energy, but only with a suitable fraction of it, to be identified through its dynamical properties, as was suggested more than a century ago by Boltzmann himself, and later by Nernst.

Here, it is first discussed why nearly integrable systems can be expected to present the FPU phenomenon. Then the latter is illustrated. Finally, some hints are given for the corresponding thermodynamics.

## Nearly Integrable versus Hyperbolic Systems, and the Question of the Rates of Thermalization

As mentioned above, it is usually assumed that the problem of providing a dynamical foundation to classical statistical mechanics is reduced to the mathematical problem of ascertaining whether the Hamiltonian systems of physical interest are ergodic or not. However, there remains open a subtler problem. Indeed, the notion of ergodicity involves the limit of an infinite time (time averages should converge to ensemble averages as $t \to \infty$), while intermediate times might be relevant. In this connection it is convenient to distinguish between two classes of dynamical systems, namely the hyperbolic and the nearly integrable ones.

The first class, in a sense the prototype of chaotic systems, should include the systems of hard spheres (extensively studied after the classical works of Sinai), or more generally the systems of mass points with mutual repulsive interactions. For such systems it can be expected that the time averages of the relevant dynamical quantities in an extremely short time converge to the corresponding ensemble averages, so that the classical equilibrium ensembles could be safely used.

A completely different situation occurs for the dynamical systems such as the FPU systems, which are nearly integrable, that is, are perturbations of

systems having a number of integrals of motion equal to the number of degrees of freedom. Indeed, in such a case ergodicity means that the addition of an interaction, no matter how small, makes an integrable system lose all of its integrals of motion, apart from the Hamiltonian itself. And, in fact, this quite remarkable property was already proved to be generic by Poincaré, through a set of considerations which had a fundamental impact on the theory of dynamical systems itself. In view of its importance for the foundations of statistical mechanics, the proof given by Poincaré was reconsidered by Fermi, who added a subtle contribution concerning the role of single invariant surfaces. It is just to such a paper that Ulam makes reference in his comment to the FPU work mentioned above, when he says: "Fermi's earlier interest in the ergodic theory is one motive" for the FPU work.

The point is that the picture which looks at the ergodicity induced on an integrable system by the addition of a perturbation, no matter how small, somehow lacks continuity. One might expect that, in situations in which the nonlinear interaction which destroys the integrals of motion is very small (i.e., at low temperatures), the underlying integrable structure should somehow be still appreciable, in some continuous way. In fact, continuity should be recovered by making a question of times, namely by considering the rates of thermalization (to use the very FPU phrase), or equivalently the relaxation times, namely the times needed for the time averages of the relevant dynamical quantities to converge to the corresponding ensemble averages. By continuity, one clearly expects that the relaxation times diverge as the perturbation tends to zero. But more complicated situations might occur, as, for example, the existence of two (or more) relevant timescales. The point of view that timescales of different orders of magnitude might occur in dynamical systems (with the exhibition of an interesting example) and that this might be relevant for statistical mechanics, was discussed by Poincaré himself in the year 1906. Indeed, he denotes as "first-order very large time" a time which is sufficient for a system to reach a "provisional equilibrium," whereas he denotes as "second-order very large time" a time which is necessary for the system to reach its "definitive equilibrium."

## The FPU Phenomenon: Historical and Conceptual Developments

We now illustrate the FPU phenomenon, following essentially its historical development. We will make reference to Figures 1–8, which are the results of numerical integrations of the FPU dynamical system. If $x_1, \ldots, x_N$ denote the positions of the particles (of unitary mass), or more precisely the displacements from their equilibrium positions, and $p_i$ the corresponding momenta, the Hamiltonian is

$$H = \sum_{i=1}^{N} \frac{p_i^2}{2} + \sum_{i=1}^{N+1} V(r_i)$$

where $r_i = x_i - x_{i-1}$ and one has taken a potential $V(r) = r^2/2 + \alpha r^3/3 + \beta r^4/4$ depending on two positive parameters $\alpha$ and $\beta$. Boundary conditions with fixed ends, namely $x_0 = x_{N+1} = 0$, are considered. We recall that the angular frequencies

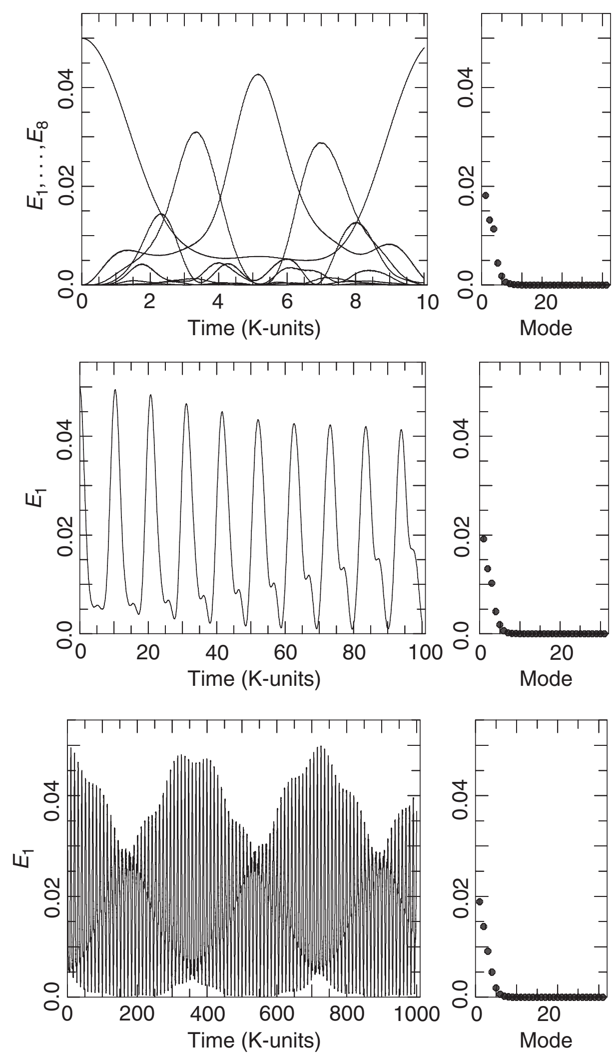



**Figure 1** The FPU paradox: normal-mode energies $E_j$ versus time (left) and energy spectrum, namely time average of $E_j$ versus $j$ (right) for three different timescales. The energy, initially given to the lowest-frequency mode, does not flow to the high-frequency modes within the accessible observation time. Here, $N = 32$ and $E = 0.05$.

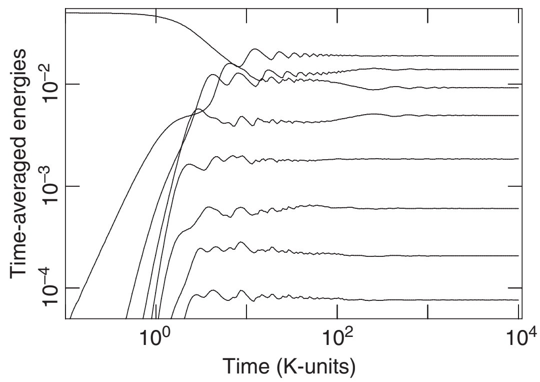



**Figure 2** The FPU paradox: time averages of the energies of the modes 1, 2, …, 8 (from top to bottom) versus time for the same run as **Figure 1**. The spectrum has reached an apparent equilibrium, different from that of equipartition predicted by classical equilibrium statistical mechanics. An exponential decay of the tail is clearly exhibited.

of the corresponding normal modes are $\omega_j = 2\sin[j\pi/2(N+1)]$, with $j=1,\ldots,N$; it is thus convenient to take as time unit the value $\pi$, which is essentially, for any $N$, the period of the fastest normal mode.

The original FPU result is illustrated in Figures 1 and 2. Here $N=32, \alpha=\beta=1/4$, and the total energy is $E=0.05$; the energy was given initially to the first normal mode (with vanishing potential energy). Three timescales (increasing from top to bottom) are considered, the top one corresponding to the timescale of the original FPU paper. In the boxes on the left the energies $E_j(t)$ of modes $j$ are reported versus time ($j=1,\ldots,8$ at top, $j=1$ at center and bottom). In the boxes on the right we report the corresponding spectra, namely the time average (up to the respective final times) of the energy of mode $j$ versus $j$, for $1 \le j \le N$. In Figure 2 we report, for the same run of Figure 1, the time averages of the energies of the various modes versus time; this figure corresponds to the last one of the original FPU work. The facts to be noticed in connection with these two figures are the following: (1) the spectrum (namely the distribution of energy among the modes, in time average) appears to have relaxed very quickly to some form, which remains essentially unchanged up to the maximum observed time; (2) there is no global equipartition, but only a partial one, because the energy remains confined within a group of low-frequency modes, which form a small packet of a certain definite width; and (3) the time evolutions of the mode energies appear to be of quasiperiodic type, since longer and longer quasiperiods can be observed as the total time increases.

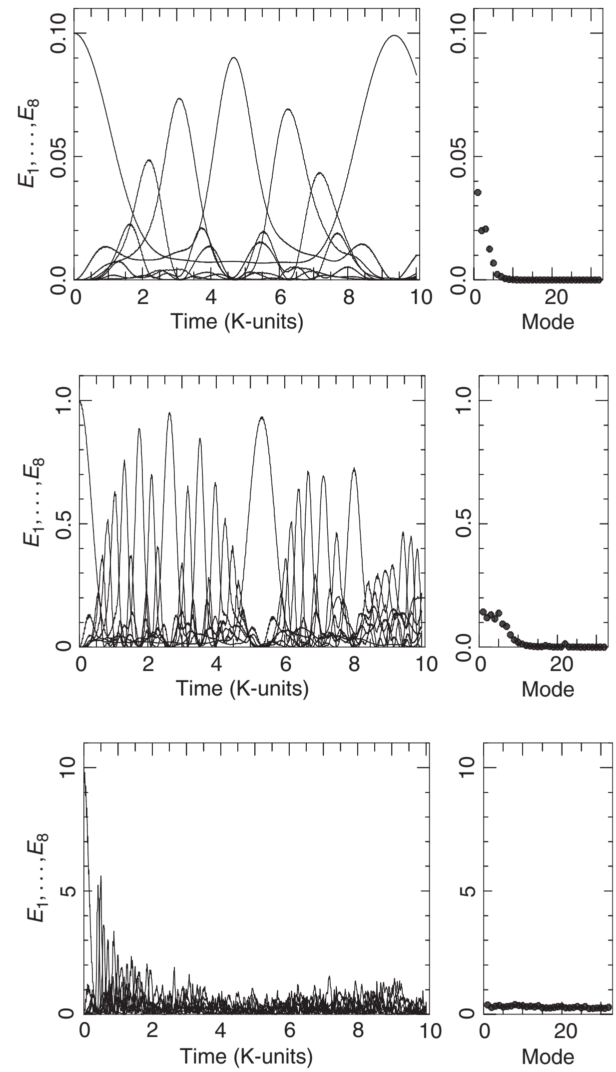



**Figure 3** The Izrailev–Chirikov contribution: for a fixed observation time, equipartition is attained if the initial energy $E$ is high enough. Here, from top to bottom, $E=0.1, 1, 10$.

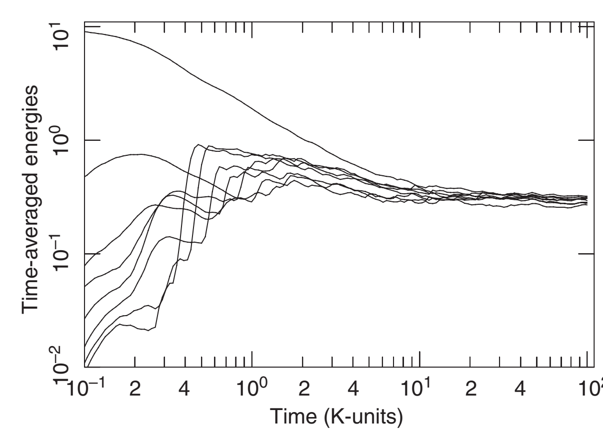



**Figure 4** The Izrailev–Chirikov contribution: time averages of the mode energies versus time for the same run as at bottom of **Figure 3**.

After the works of Zabusky and Kruskal, by which the FPU system was somehow assimilated to an integrable system, the bridge toward ergodicity was made by Izrailev and Chirikov (1966), through the idea that there should exist a stochasticity threshold. Making reference to KAM theory, which had just been formulated in the framework of perturbation theory for nearly integrable systems, their main remark was as follows. It is known that KAM theory, which essentially guarantees a behavior similar to that of an integrable system, applies only if the perturbation is smaller than a certain threshold; on the other hand, in the FPU model the natural perturbation parameter is the energy $E$ of the system. Thus, the FPU phenomenon can be expected to disappear above a certain threshold energy $E_c$. This is indeed the case, as illustrated in **Figures 3 and 4**. The parameters $\alpha, \beta$ and the class of initial data are as in **Figure 1**. In **Figure 3** the total time is kept fixed (at 10 000 units), whereas the energy $E$ is increased in passing from top to bottom, actually from $E = 0.1$ to $E = 1$ and $E = 10$. One sees that at $E = 10$ equipartition is attained within the given observation time; correspondingly, the motion of the modes visually appears to be nonregular. The approach to equipartition at $E = 10$ is clearly exhibited in **Figure 4**, where the time averages of the energies are reported versus time.

There naturally arose the problem of the dependence of the threshold $E_c$ on the number $N$ of degrees of freedom (and also on the class of initial data). Certain semianalytical considerations of Izrailev and Chirikov were generally interpreted as suggesting that the threshold should vanish in the thermodynamic limit for initial excitations of high-frequency modes. Recently, Shepelyanski completed the analysis by showing that the threshold should vanish also for initial excitations of the low-frequency modes, as in the original FPU work (see, however, the subsequent paper by Ponno mentioned below). If this were true, the FPU phenomenon would disappear in the thermodynamic limit. In particular, the equipartition principle would be dynamically justified at all temperatures.

The opposite conjecture was advanced by Bocchieri *et al.* (1970). This was based on numerical calculations, which indicated that the energy threshold should be proportional to $N$, namely that the FPU phenomenon persists in the thermodynamic limit provided the specific energy $\epsilon = E/N$ is below a critical value $\epsilon_c$, which should be definitely nonvanishing. Actually, the computations were performed on a slightly different model, in which nearby particles were interacting through a more physical Lennard-Jones potential. By taking concrete values having a physical significance, namely the values commonly assumed for argon, for the threshold of the specific energy they found the value $\epsilon_c \simeq 0.04 V_0$, where $V_0$ is the depth of the Lennard-Jones potential well. This corresponds to a critical temperature of the order of a few kelvin. The relevance of such a conjecture (persistence of the FPU phenomenon in the thermodynamic limit) was soon strongly emphasized by Cercignani, Galgani, and Scotti, who also tried to establish a connection between the FPU spectrum and Planck's distribution.

Up to this point, the discussion was concerned with the alternative whether the FPU system is ergodic or not, and thus reference was made to properties holding in the limit $t \to \infty$. Correspondingly, one was making reference to KAM theory, namely to the possible existence of surfaces ($N$-dimensional tori) which should be dynamically invariant (for all times). The first paper in which attention was drawn to the problem of estimating the relaxation times to equilibrium was by Fucito *et al.* (1982). The model considered was actually a different one (the so-called $\phi^4$ model), but the results can also be extended to the FPU model. Analytical and numerical indications were given for the existence of two timescales. In a short time the system was found to relax to a state characterized by an FPU-like spectrum, with a plateau at the low frequencies, followed by an exponential tail. This, however, appeared as being a sort of metastable state. In their words: "The nonequilibrium spectrum may persist for extremely long times, and may be mistaken for a stationary state if the observation time is not sufficiently long." Indeed, on a second much larger timescale the slope of the exponential tail was found to increase logarithmically with time, with a rate which decreases to zero with the energy. This is an indication that the time for equipartition should increase as an exponential with the inverse of the energy.

This is indeed the picture that the present authors consider to be essentially correct, being supported by very recent numerical computations, and by analytical considerations. Curiously enough, however, such a picture was not fully appreciated until quite recently. Possibly, the reason is that the scientific community had to wait until becoming acquainted with two relevant aspects of the theory of dynamical systems, namely Nekhoroshev theory and the relations between KdV equation and resonant normal-form theory.

The first step was the passage from KAM theory to Nekhoroshev theory. Let us recall that, whereas in KAM theory one looks for surfaces which are invariant (for all times), in Nekhoroshev theory one

looks instead for a kind of weak stability involving finite times, albeit "extremely long" ones, as they are found to increase as stretched exponentials with the inverse of the perturbative parameter. Thus, one meets with situations in which one can have instability over infinite times, while having a kind of practical stability up to exponentially long times. Notice that Nekhoroshev's theory was formulated only in the year 1974, and that it started to be known in the West only in the early 1980s, just because of its interest for the FPU problem. Another interesting point is that just in those years one started to become acquainted with a related historical fact. Indeed, the idea that equipartition might require extremely long times, so that one would be confronted with situations of a practical lack of equipartition, has in fact a long tradition in statistical mechanics, going back to Boltzmann and Jeans, and later (in connection with sound dispersion in gases of polyatomic molecules) to Landau and Teller.

In this way the idea of the existence of extremely long relaxation times to equipartition came to be accepted. The ingredient that was still lacking is the idea of a quick relaxation to a metastable state. The importance of this should not be overlooked. Indeed, without it one cannot at all have a thermodynamics different from the standard equilibrium one corresponding to equipartition. This was repeatedly emphasized, against Jeans, by Poincaré on general grounds and by Nernst on empirical grounds. The full appreciation of this latter ingredient was obtained quite recently (although it had been clearly stated by Fucito *et al.* (1982)). A first hint in this direction came from the realization (see **Figure 5**) of a deep analogy between the FPU phenomenon and the phenomenology of glasses. Then there came a strong numerical indication by Berchialla, Galgani, and Giorgilli. Finally, from the analytical point of view, there was a suitable revisitation (by Ponno) of the traditional connection between the FPU system and the KdV equation with its solitons. The relevant points are the following: (1) the KdV equation describes well the solutions of the FPU problem (for initial data of FPU type) only on a "short" timescale, which increases as a power of $1/\epsilon$, and so describes only a first process of quick relaxation; (2) the corresponding spectrum has a very definite analytical form, the energy being spread up to a maximal frequency $\bar{\omega}(\epsilon) \simeq \epsilon^{1/4}$ and then decaying exponentially; and (3) the relevant formulas contain the energy only through the specific energy $\epsilon$, and thus can be expected to hold also in the thermodynamic limit. It should be mentioned, however, that all the results of an

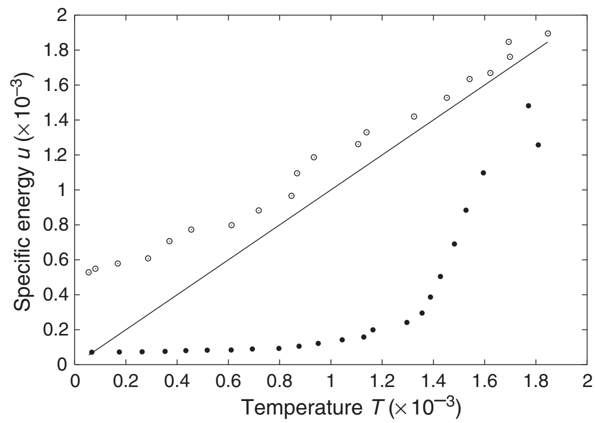



**Figure 5** Analogy with glasses: the specific energy $u$ of an FPU system is plotted versus temperature $T$ for a cooling process (upper curve) and a heating process (lower curve). The FPU system is kept in contact with a heat reservoir, whose temperature is changed at a given rate. At low temperatures the system does not have time to reach the equilibrium curve $u = T$ (with $k_B = 1$).

analytic type mentioned above have a purely formal character, because up to now none of them was proved, in the thermodynamic limit, in the sense of rigorous perturbation theory. This requires a suitable readaptation of the known techniques, which is currently being obtained both in connection with Nekhoroshev's theorem (in order to explain the extreme slowness of a possible final approach to equilibrium) and in connection with the normal-form theory for partial differential equations (in order to explain the fast relaxation to the metaequilibrium state).

In conclusion, for the case of initial conditions of the FPU type (excitation of a few low-frequency modes) the situation seems to be as follows. The first phenomenon that occurs in a "short" time (of the order of $(1/\epsilon)^{3/4}$ is a quick relaxation to the formation of what can be called a "natural packet" of low-frequency modes extending up to a certain maximal frequency $\bar{\omega} \simeq \epsilon^{1/4}$. This is a phenomenon which has nothing to do with any diffusion in phase space. In fact, it shows up also for an integrable system such as a Toda lattice (as will be illustrated below), and should be described by a suitable resonant normal form related to the KdV equation. One has then to take into account the fact that the domain of the frequencies in the FPU model is bounded ($\omega < 2$ in the chosen units). Now, as the function $\bar{\omega}(\epsilon)$ is monotonic, this fact leads to the existence of a critical value $\epsilon_c$ of the specific energy $\epsilon$, defined by $\bar{\omega}(\epsilon_c) = 2$. Indeed, for $\epsilon > \epsilon_c$ the quick relaxation process leads altogether to equipartition. Below the threshold, instead, the same quick process leads to the formation of an FPU-like spectrum,

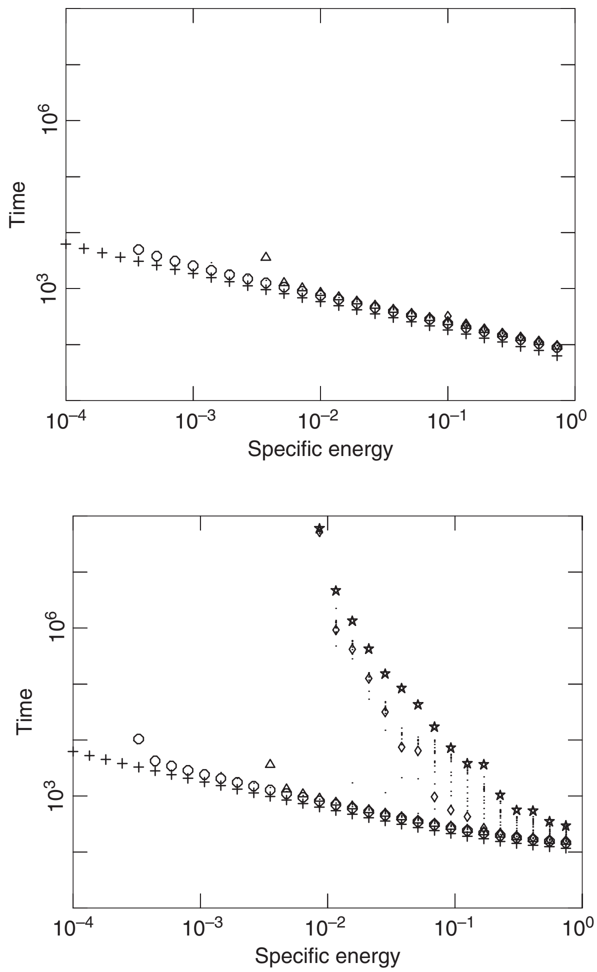



**Figure 6** Time needed to form a packet versus specific energy for the FPU model (bottom) and the corresponding Toda model (top). Different symbols refer to packets of different width. The existence of two timescales below a critical specific energy in the FPU model is exhibited.

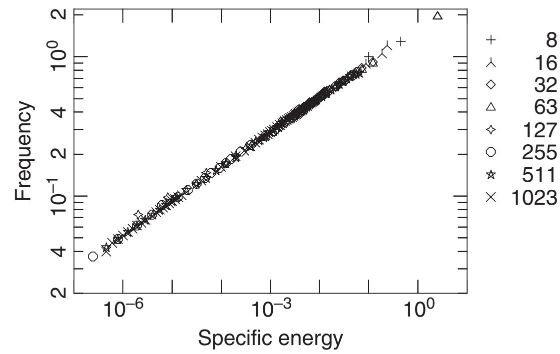



**Figure 7** Width of the natural packet versus specific energy, for *N* ranging from 8 to 1023. Reproduced from Berchialla L, Galgani L, and Giorgilli A (2004) Localization of energy in FPU chains, *Discr. Cont. Dyn. Systems B* 11: 855–866, with permission from American Institute of Mathematical Sciences.

involving only modes of sufficiently low frequency. This should, however, be a metastable state (which might be mistaken for a stationary one), which should be followed, on a second timescale, by a relaxation to the final equilibrium, through a sort of Arnol'd diffusion requiring extremely long Nekhoroshev-like times. This is actually the way in which the old idea of a threshold, originally conceived in terms of KAM tori, is now recovered even for ergodic systems, in terms of timescales.

The existence of a process of quick relaxation, and of a threshold in the above-mentioned sense, is illustrated in Figures 6 and 7. In Figure 6 the lower part refers to the FPU model, while the upper one refers to a corresponding Toda model. The latter is in a sense the prototype of an integrable nonlinear system; with respect to the FPU case, the difference is that the potential $V(r)$ is now exponential. The parameters of the exponential were chosen so that the two models coincide up to cubic terms in the potential. With the energy given to the lowest-frequency mode, the figure shows the time needed in order that energy spreads up to a mode $\bar{k}$, for several values of $\bar{k}$, as a function of $\epsilon$. It is seen that in the Toda model (top) there is formed a packet extending up to rather well-defined width, and that this occurs within a relaxation time increasing as a power of $1/\epsilon$. An analogous phenomenon occurs for the FPU model (bottom). The only difference is that, below a critical specific energy $\epsilon_c \simeq 0.1$, there exists a subsequent relaxation time to equipartition, which involves a time growing faster than any inverse power of $\epsilon$. Such a second phenomenon is due to the nonintegrable character of the FPU model. In Figure 7 the width of the natural packet for the FPU model is exhibited, by reporting the frequency $\bar{\omega}$ of its highest mode as a function of $\epsilon$. As one sees, the numerical results clearly indicate the existence of a relation $\bar{\omega} \simeq \epsilon^{1/4}$, which holds for a number of degrees of freedom $N$ ranging from 8 to 1023. This is actually the law which is predicted by resonant normal-form theory.

## Boltzmann and Nernst Revisited

All the results illustrated above refer to initial data of FPU type, namely with an excitation of a few low-frequency modes. However, from the point of view of statistical mechanics, such initial data are exceptional, and one should rather consider initial data extracted from the Gibbs distribution at a certain temperature. One can then couple the FPU system to a heat bath at a slightly different temperature, and look at the spectrum of the FPU system after a certain time. The result, for the case of a heat bath at a higher temperature, is shown in

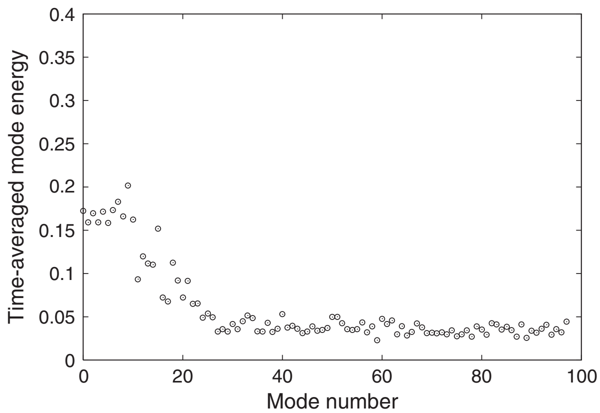



**Figure 8** A case of an FPU system initially at equilibrium and thus in equipartition. Spectrum of the FPU system after it was kept in contact with a heat reservoir at a higher temperature.

Figure 8. Clearly, here one has a situation similar to that occurring for initial data of FPU type, because only a packet of low-frequency modes exhibits a reaction, each of its modes actually adapting itself to the temperature of the bath, whereas the high-frequency modes do not react at all, that is, remain essentially frozen.

This capability of reacting to external disturbances (which seems to pertain only to a fraction of the mechanical energy initially inserted into the system) can be characterized in a quantitative way through an estimate of the fluctuations of the total energy of the FPU system. This is indeed the sense of the fluctuation–dissipation theorem, the precursor of which is perhaps the contribution of Einstein to the first Solvay conference (1911). Through such a method, the specific heat of the FPU system is estimated (apart from a numerical factor) by the time average of $[E(t) - E(0)]^2$, where $E(t)$ is the energy, at time $t$, of the FPU system in dynamical contact with a heat bath (at the same temperature from which the initial data are extracted). Usually, in the spirit of ergodic theory, one looks at the infinite-time limit of such a quantity. But in the spirit of the metastable picture described above, one can check whether the time average presents a previous stabilization to some value smaller than the one predicted at equilibrium. Such a result, which is in qualitative agreement with the third principle, has indeed been obtained (by Carati and Galgani) recently.

In conclusion, in situations of metaequilibrium such as those existing in the FPU model at low temperatures, a thermodynamics can still be formulated. Indeed, by virtue of the quick relaxation process described above, the time averages of the relevant quantities are found to stabilize in rather short times. In this way, one overcomes the critique of Poincaré to Jeans, namely that one cannot have a thermodynamics at all if reference is made only to the existence of extremely long relaxation times to the final equilibrium. A relaxation to a “provisional equilibrium” within a “first-order very large time” (to quote Poincaré) is required . The difference with respect to the standard equilibrium thermodynamics relies now in the mechanical interpretation of the first principle. Indeed, the internal thermodynamic energy is identified not with the whole mechanical energy, but just with that fraction of it which is capable of reacting in short times to the external perturbations.

This is the way in which the old idea of Boltzmann (and Jeans) might perhaps be presently implemented. For what concerns the fraction of the mechanical energy which is not included in the thermodynamic internal energy, as not being able to react in relatively short times, this should somehow play the role of a zero-point energy. This was suggested in the year 1971 by C Cercignani. But in fact, such a concept was put forward by Nernst himself in an extremely speculative work in 1916, where he also advanced the concept that, for a system of oscillators of a given frequency, there should exist both dynamically ordered (*geordnete*) and dynamically chaotic (*ungeordnete*) motions, the latter being prevalent above a certain energy threshold. According to him, this fact should be relevant for a dynamical understanding of the third principle and of Planck’s law.

It is well known that the modern theory of dynamical systems has led to familiarity with the (sometimes abused) notions of order and chaos and of a transition between them. One might say that the FPU work just forced the scientific community to take into account such notions in connection with the principle of equipartition of energy. It is really fascinating to see that the same notions, with the same terminology, had already been introduced much earlier on purely thermodynamic grounds, in connection with the relations between classical and quantum mechanics.

*See also:* Boundary Control Method and Inverse Problems of Wave Propagation; Central Manifolds, Normal Forms; Ergodic Theory; Fourier Law; Gravitational *N*-body Problem (Classical); Newtonian Fluids and Thermohydraulics; Nonequilibrium Statistical Mechanics: Interaction between Theory and Numerical Simulations; Quantum Statistical Mechanics: Overview; Regularization for Dynamical Zeta Functions; Stability Theory and KAM; Toda Lattices; Weakly Coupled Oscillators.

## Further Reading

Benettin G, Galgani L, and Giorgilli A (1984) Boltzmann’s ultraviolet cutoff and Nekhoroshev’s theorem on Arnol’d diffusion. *Nature* 311: 444–445.

Bocchieri P, Scotti A, Bearzi B, and Loinger A (1970) Anharmonic chain with Lennard-Jones interaction. *Physical Review A* 2: 2013–2019.

Boltzmann L (1966) *Lectures on Gas Theory* (translated by Brush SG, Vol. II, Sects. 43–45), Berkeley: The University of California Press.

Fermi E (1923) Beweiss das ein mechanisches Normalsystem im allgemeinen quasi-ergodisch ist. In: Fermi E (ed.) (1965) *Note e Memorie (Collected Papers)*, no. 11, vol. I, pp. 79–87. Roma: Accademia Nazionale dei Lincei.

Fermi E, Pasta J, and Ulam S (1955) Studies of nonlinear problems. In: Fermi E (ed.) (1965) *Note e Memorie (Collected Papers)*, no. 266, vol. II, pp. 977–988. Roma: Accademia Nazionale dei Lincei.

Fucito E, Marchesoni F, Marinari E, Parisi G, Peliti L *et al.* (1982) Approach to equilibrium in a chain of nonlinear oscillators. *Journal de Physique* 43: 707–713.

Galgani L and Scotti A (1972) Recent progress in classical nonlinear dynamics. *Rivista del Nuovo Cimento* 2: 189–209.

Izrailev FM and Chirikov BV (1966) Statistical properties of a nonlinear string. *Soviet Physics Doklady X* 11: 30–32.

Nernst W (1916) Über einen Versuch, von quantentheoretischen Betrachtungen zur Annahme stetiger Energieänderungen zurückzukehren. *Verhandlungen der Deutschen Physikalischen Gesellschaft* 18: 83.

Poincaré H (1996) Reflexions sur la théorie cinétique des gas. *Journal de Physique Theoretical Application*. 5: 369–403 (in (1954) Oeuvres de Henry Poincaré, Tome IX: 597–619. Paris: Gauthier-Villars).

Ponno A (2003) Soliton theory and the Fermi–Pasta–Ulam problem in the thermodynamic limit. *Europhysics Letters* 64: 606–612.

Shepelyansky DI (1997) Low-energy chaos in the Fermi–Pasta–Ulam problem. *Nonlinearity* 10: 1331–1338.

Zabusky NJ and Kruskal MD (1965) Interaction of “solitons” in a collisionless plasma and the recurrence of initial states. *Physical Review Letters* 15: 240–243.

# Dynamical Systems in Mathematical Physics: An Illustration from Water Waves

**O Goubet**, Université de Picardie Jules Verne, Amiens, France



## Introduction

The purpose of this article is to describe some basic problems related to the interplay between dynamical systems and mathematical physics. Since it is impossible to be exhaustive in these topics, the focus here is on water-wave models. These mathematical models are described by partial differential equations that can be understood as dynamical systems in a suitable infinite-dimensional phase space.

We will not address the original equations for two-dimensional (2D) surface water waves, even if we know that dynamical-system methods can help to exhibit some solitary waves for the equations. The reader is referred to relevant articles in this encyclopedia for details. Another approach is to seek these 2D surface water waves as saddle points for some Hamiltonians, which too is discussed elsewhere in this work.

This article presents these arguments on some asymptotical models for the propagation of surface water waves.

## Asymptotical Models in Hydrodynamics

To begin with, consider an irrotational fluid in a canal that is governed by the Euler equations and that is subject to gravitational forces. For a canal of finite depth, Boussinesq (1877) and Korteweg–de Vries (KdV) (1890) obtained the following model for unidirectional long waves:

$$u_t + u_x + u_{xxx} + uu_x = 0 \qquad [1]$$

Sometimes we drop the $u_x$ term on the left-hand side of [1], thanks to a suitable change of coordinates. Alternatively, we can also deal with the so-called generalized KdV equation, which reads

$$u_t + u_{xxx} + u^k u_x = 0 \qquad [2]$$

where $k$ is a positive integer. There are also other models designed to represent long waves in shallow water. Let us introduce the regularized long-wave equation (also referred to as the Benjamin–Bona–Mahony equation) that reads

$$u_t - u_{txx} + u_x + uu_x = 0 \qquad [3]$$

or the Camassa–Holm equation

$$u_t - u_{txx} + 3uu_x = 2u_x u_{xx} + uu_{xxx} \qquad [4]$$

For deep water, a well-known model was introduced by Zakharov (1968)

$$\mathrm{i}u_t + u_{xx} + \varepsilon|u|^2 u = 0 \qquad [5]$$

which describes the slow modulations of wave packets. Here the unknown $u(x, t)$ takes values in $\mathbb{C}$, and this nonlinear Schrödinger equation is in fact a system. In these equations, $\varepsilon$ is either 1 or –1; throughout this article, we shall refer to the former case as the focusing case and to the latter

as the defocusing case. We may also substitute $|u|^{2p}u$ in the nonlinear term in [5] to obtain alternate models.

The variable $t$ represents the time and the space variable $x$ belongs either to $\mathbb{R}$ or to a finite interval when we are dealing with periodic flows.

The above models are intended to describe the propagation of unidirectional waves. For two-way waves, see Bona *et al.* (2002).

Actually, these equations feature particular solutions, the so-called traveling waves. Let us recall, for instance, that for generalized KdV equation [2] these solutions are

$$u(t,x) = Q_c(x - ct) \qquad [6]$$

$$Q_c(x) = c^{1/p}Q(\sqrt{c}x) \qquad [7]$$

$$Q(x) = (3\,\mathrm{ch}^{-2}(px))^{1/p} \qquad [8]$$

These so-called solitons (**Figure 1**) move to the right without changing their shape; $c$ is the speed of propagation. In real life, this phenomenon was observed by Russel (1834). Riding his horse, he was able to follow for miles the propagation of such a wave on the canal from Edinburgh to Glasgow. On the other hand, Camassa–Holm equations are designed to describe the propagation of peaked solitons as shown in **Figure 2**.

Focusing nonlinear Schrödinger equations also feature solitary waves that read $u_\omega(t,x) = \exp(\mathrm{i}\omega t)Q(x)$, where $Q$ is solution to

$$Q_{xx} - \omega Q + Q^{2p+1} = 0 \qquad [9]$$

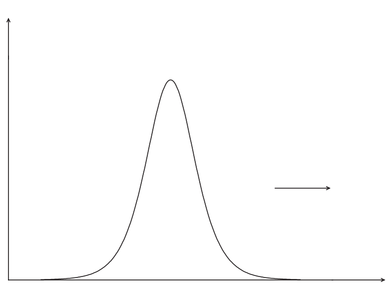


**Figure 1** A soliton.

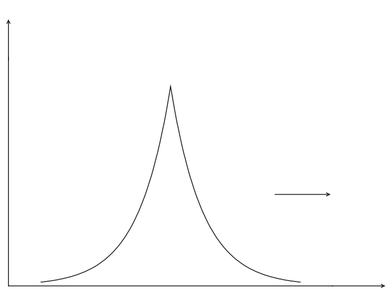


**Figure 2** Peaked soliton.

There are numerous examples of equations or systems of equations that model 2D surface water waves. Among all these models, a first issue is to identify the relevant models insofar as the dynamical properties are concerned. Indeed, we address here the question of stability of solitary waves (up to the symmetries of the equation). For instance, the orbital stability for cubic Schrödinger reads: for any $\varepsilon > 0$, there exists a neighborhood $\Omega$ of $u_\omega(x,0)$ such that any trajectory starting from $\Omega$ satisfies

$$\sup_t \inf_\theta \inf_y ||u(t) - \exp(\mathrm{i}\theta)u_\omega(t, . - y)||_{H^1} \le \varepsilon \qquad [10]$$

Another issue consists in the interaction of $N$ solitons. Schneider and Wayne (2000) have addressed the issue of the validity of water-wave models when this interaction is concerned.

Assume now that the validity of these models is granted. To consider [1] or [5] as a dynamical system, the next issue is then to consider the initial-value problem.

## The Initial-Value Problem

Let us supplement these equations with initial data $u_0$ in some Sobolev space. We shall consider either

$$H^s(\mathbb{R}) = \left\{ u; \int_{\mathbb{R}} (1 + |\xi|^2)^s |\hat{u}(\xi)|^2 \,\mathrm{d}\xi < +\infty \right\} \qquad [11]$$

in the case where $x$ belongs to the whole line, or the corresponding Sobolev space with periodic boundary conditions. It should be examined whether these equations provide a continuous flow $S(t): u_0 \to u(t)$ in these functional spaces (at least locally in time). We would like to point out that for each Sobolev space under consideration, we may have a different flow. This fact is at the heart of infinite-dimensional dynamical systems.

The initial-value problem was a challenge for decades for low norms, that is for small $s$. The last breakthrough was performed by Bourgain (1993). Let us present the method for KdV equation. Consider $U(t)u_0$ the solution of the Airy equation

$$u_t + u_{xxx} = 0, \quad u(0) = u_0 \qquad [12]$$

Without going into further details, the idea is to perform a fixed-point argument to the Duhamel's form of the equation,

$$u(t) = U(t)u_0 - \frac{1}{2}\int_0^t U(t-s)\partial_x(u^2(s))\mathrm{d}s \qquad [13]$$

in a suitable mixed-spacetime Banach space whose norm reads $\|U(-t)u(t,x)\|_{H_t^b H_x^\rho}$. This relies on fine properties in harmonic analysis. Thanks to this method, we know that the Schrödinger equation [5] and the KdV equation [1] are well-posed in, respectively, $H^s(\mathbb{R}), s \geq 0$ and $H^s(\mathbb{R}), s > -3/4$, locally in time. For the periodic case, the results are slightly different. We would like to point out that both KdV and nonlinear Schrödinger equations provide semigroups $S(t)$ that do not feature smoothing effect. A trajectory that starts from $H^s$ remains in $H^s$; indeed, we can also solve these partial differential equations backward in time.

The next issue is to determine if these flows are defined for all times. Loosely speaking, the following alternative holds true: either the local flow in $H^s$ extends to a global one, or some blow-up phenomenon occurs, that is, $||S(t)u_0||_{H^s}$ collapses in finite time.

To this end, let us observe that, for instance, the mass $\int_{\mathbb{R}} |u(x)|^2\, dx$ is conserved for both KdV and nonlinear Schrödinger flows. Therefore, one can prove that the solutions in $L^2$ are global in time. It is worthwhile to observe that the Bourgain method also provides some global existence results below the energy norm.

Consider now the flow of the solutions in $H^1$. The second invariant for nonlinear Schrödinger equations reads

$$\int_{\mathbb{R}} |u_x(x)|^2 \mathrm{d}x - \frac{\varepsilon}{p+1} |u(x)|^{2p+2} \mathrm{d}x \qquad [14]$$

Therefore, the local solutions in $H^1$ extend to global ones in the defocusing case ($\varepsilon = -1$). In the focusing case, the situation is more contrasted. The solution is global if the nonlinearity is less than an $H^1$-critical value ($p = 2$ for Schrödinger, and $k = 4$ for generalized KdV equation). This critical value depends on some Sobolev embeddings as

$$\int_{\mathbb{R}} |u(x)|^{2p+2} \mathrm{d}x \leq C_p \|u\|_{L^2}^{p+1} \|u_x\|_{L^2}^{p} \qquad [15]$$

Therefore, since the mass is constant, the second invariant controls the $H^1$ norm of the solution if $p < 2$. Note that the critical power of the nonlinearity depends also on the dimension of the space; it is the cubic Schrödinger that is critical in $H^1(\mathbb{R}^2)$. It is well known that, for some initial data, blow-up phenomena can occur for 2D cubic Schrödinger equations. Moreover, the behavior of blow-up solutions is more or less understood. This analysis was performed using the conformal invariance of the equation. For quintic Schrödinger equation, which is critical in 1D, this conformal invariance states that if $u(t,x)$ is solution, then

$$v(t,x) = |t|^{-1/2} \exp\left(\frac{\mathrm{i}x^2}{4t}\right) \bar{u}\left(\frac{1}{t}, \frac{x}{t}\right) \qquad [16]$$

is also solution.

On the other hand, for the generalized KdV equation, there is no conformal invariance and the blow-up issue had been open for years. There was some numerical evidence that blow-up can occur for $k = 4$. Recently, Martel and Merle (2002) have given a complete description of the blow-up profile for this equation. Their methods are quite complex and rely on an ejection of mass at infinity in a suitable coordinate system.

In the discussion so far we have presented some quantities that are invariant by the flow of the solutions. This is related to the Hamiltonian structure of the dynamical systems under consideration.

## Hamiltonian Systems in Hydrodynamics

The study of Hamiltonian systems has developed beyond celestial mechanics (the famous $n$-body problems) to other fields in mathematical physics. We focus here on dynamical systems that read

$$u_t = J \frac{\partial}{\partial u} H(u) \qquad [17]$$

where $H$ is the Hamiltonian and $J$ some skew-symmetric operator. For instance, [1] is a Hamiltonian system with $J = \partial_x$ (i.e., an unbounded skew-symmetric operator) and

$$H(u) = \frac{1}{2} \int (u^2 - u_x^2) \mathrm{d}x + \frac{1}{6} \int u^3 \mathrm{d}x \qquad [18]$$

There is a subclass of Hamiltonian systems that are integrable by inverse-scattering methods. For instance, [1] belongs to this class. Indeed, these methods give a complete description of the asymptotics when $t \to \pm\infty$. It is well known (Deift and Zhou 1993) that, asymptotically, any solution to KdV equation consists of a wave train moving to the right in the physical space up to a dispersive part moving to the left.

On the other hand, a generic Hamiltonian system is not integrable. The study of the asymptotics and of the dynamical properties of such a system deserves another analysis. We say that a system features asymptotic completeness if there exist $u_+$

and $u_-$ such that the solution $u(t)$ of [17] supplemented with initial data $u_0$ satisfies

$$\|u(t) - U(t)u_+\| \to 0 \quad [19]$$

$$\|u(t) - U(t)u_-\| \to 0 \quad [20]$$

when, respectively, $t \to +\infty$ or $t \to -\infty$. Here $U(t)u_0$ is the solution of the free equation, that is, the associated linear equation, supplemented with initial data $u_0$; for instance, the Airy equation is the free equation related to the KdV equation. The operators $u_- \to u_0 \to u_+$ are called wave operators. This is related to the Bohr's transition in quantum mechanics. Loosely speaking, we are able to prove these scattering properties for high powers in the nonlinearity for subcritical defocusing Schrödinger equations.

The asymptotics of trajectories can be more complicated. Let us recall that the stability of traveling waves is also an important issue in understanding the dynamical properties of these models. For instance, let us point out that Martel and Merle proved the asymptotic stability of the sum of $N$ solitons for KdV in the subcritical case.

Beyond these asymptotics we are interested in the case where the permanent regime is chaotic (or turbulent). A scenario is that there exist quasiperiodic solutions of arbitrarily order $N$ for the system under consideration. The next challenge about these Hamiltonian systems is to apply the Kolmogorov–Arnol'd–Moser theory to exhibit this type of solutions to systems like [17]. Here we restrict our discussion to the case of bounded domains, with either periodic or homogeneous Dirichlet conditions. Then, let us introduce the following definition: a solution is quasiperiodic if there exist a finite number $N$ of frequencies $\omega_k$ such that

$$u(t, x) = \sum_{l=1}^{N} u_l(x) \exp(\mathrm{i}\omega_k t) \quad [21]$$

This extends the case of periodic solutions ($N = 1$), which are isomorphic to the torus. To prove the existence of such structures, one idea is then to imbed $N$-dimensional invariant tori into the phase space of solutions. One may approximate the infinite-dimensional Hamiltonian by a sequence of finite ones and consider the convergence of iterated symplectic transformations, or one solves directly some nonlinear functional equation. Actually, the difficulty is that resonances can occur. Resonances occur when there are some linear combinations of the frequencies that vanish (or that are arbitrarily close to 0). This introduces a small divisor problem in a phase space that has infinite dimension. To overcome these difficulties, a Nash–Moser scheme can be implemented (Craig 1996). There are numerous such open problems. For instance, let us observe that known results are essentially only for the case where the dimension of the ambient space is 1. On the other hand, quasiperiodic solutions correspond to $N$-dimensional invariant tori for the flow of solutions; one may seek for Lagrangian invariant tori that correspond to the case where $N = +\infty$. Current research is directed towards extending this analysis.

Another issue is to seek invariant measures for these Hamiltonian dynamical systems, as in statistical mechanics. Bourgain was successful in performing this analysis for some nonlinear Schrödinger equations either in the case of periodic boundary conditions or in the whole space. This result is an important step in the ergodic analysis of our Hamiltonian dynamical systems. This could explain the Poincaré recurrence phenomena observed numerically for these types of equations: some particular solutions seem to come back to their initial state after a transient time. This point will not be developed here.

All these results are properties of conservative dynamical systems. We now address the case when some dissipation takes place.

## Dissipative Water-Wave Models

To model the effect of viscosity on 2D surface water waves, we go back to a flow governed by the Navier–Stokes equations and we proceed to obtain damped equations (Ott and Sudan 1970, Kakutani and Matsuuchi 1975). In fact, the damping in KdV equations can be either a diffusion term that leads to study the equation

$$u_t + u_{xxx} + uu_x = \nu u_{xx} \quad [22]$$

where $\nu$ is a positive number analogous to the viscosity, or a zero-order term $-\nu u$ on the right-hand side of [22]. In the first case, we obtain a KdV–Burgers equation that has some smoothing effect in time. In the second case, we have a zero-order dissipation term. A nonlocal term would be $\nu\mathcal{F}^{-1}(|\xi|^{2\beta}\hat{u}(\xi))$ for $\beta \in [0, 1]$, where $\mathcal{F}(u) = \hat{u}$ denotes the Fourier transform of $u$.

A first issue concerning damped water-wave equations is to estimate the decay rate of the solutions towards the equilibrium (no decay) when $t \to +\infty$. For [22] the ultimate result is that, for initial data $u_0 \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, the $L^2$ norm of the solution decays like $t^{-1/4}$ (Amick *et al.* 1989). Energy methods have been developed to handle these problems, as the Shonbeck's splitting method.

The center manifold theory is another approach that is employed in dynamical systems. The aim is to prove the existence of a finite-dimensional manifold that is invariant (in a neighborhood of the origin) by the flow of the solutions and that attracts the other trajectories with high speed. Therefore, this manifold, and the trajectories therein, monitor the decay rate of the solutions towards the origin. The construction of such a manifold relies on splitting properties of the spectrum of the associated linearized operator (Gallay and Wayne 2002). Using a suitable change of variables (that moves the continuous spectrum away from the origin), Gallay and Wayne were able to construct such a manifold in an infinite-dimensional phase space.

Another issue is the understanding of the dynamics for damped–forced water-wave equations as

$$u_t + uu_x + u_{xxx} + \nu u = f(x) \qquad [23]$$

The dynamical system approach is the attractor theory (Temam 1997). Equations such as [23] provide dissipative semigroups $S(t)$ in some energy spaces. The theory has developed for years and we know that these dynamical systems feature global attractors. A global attractor is a compact subset in the energy space under consideration which is invariant by the flow of the solutions and that attracts all the trajectories when $t \to +\infty$. Moreover, if we deal with periodic boundary conditions, this global attractor has finite fractal (or Hausdorff) dimension. This dimension depends on the data concerning $\nu$ and $f$.

Actually, eqn [23] provides semigroups either in $L^2(\mathbb{R}), H^1(\mathbb{R})$, or in $H^2(\mathbb{R})$. These three dynamical systems feature global attractors $\mathcal{A}_0, \mathcal{A}_1, \mathcal{A}_2$. From the viewpoint of physics, the attractors describe the permanent regime of the flow. One may wonder if this permanent regime depends on the space chosen for the mathematical study. Eventually, the last result for this issue establish that $\mathcal{A}_0 = \mathcal{A}_1 = \mathcal{A}_2$. This property is equivalent to prove the asymptotical smoothing effect for the associated semigroup: even if $S(t)$ is not a smoothing operator for finite $t$, then all solutions converge to a smooth set when $t$ goes to the infinity.

All these results are for subcritical nonlinearities. As already noted, dissipation provides smoothing at infinity. Nevertheless, damping does not prevent blow-up. Let us illuminate this by the following result due to Tsutsumi (1984). The damped Schrödinger equation

$$\mathrm{i}u_t + \mathrm{i}\nu u + u_{xx} + |u|^{2p}u = 0 \qquad [24]$$

features blow-up solutions in $H^1(\mathbb{R})$ for $p > 2$, even if all solutions are damped in $L^2(\mathbb{R})$ with exponential speed.

This completes the discussion of damped–forced water-wave equations. We now consider equations that are forced with a random forcing term.

## Stochastic Water-Wave Models

During the modeling process that led to KdV or Schrödinger equations from Euler equation, we have neglected some low-order terms. We now model these terms by a noise and we are led to a new randomly forced dynamical system that reads

$$u_t + u_x + u_{xxx} + uu_x = \gamma\dot{\xi} \qquad [25]$$

Here one may assume that $\xi(x, t)$ is a Gaussian process with correlations

$$\mathbb{E}(\dot{\xi}(x,t)\dot{\xi}(y,s)) = \delta_{x-y}\delta_{t-s} \qquad [26]$$

that is, a spacetime white noise. The parameter $\gamma$ is the amplitude of the process. Unfortunately, due to the lack of smoothing effect of KdV or Schrödinger equations, it is more convenient to work with a noise that is correlated in space, satisfying

$$\mathbb{E}(\dot{\xi}(x,t)\dot{\xi}(y,s)) = c(x-y)\delta_{t-s} \qquad [27]$$

here $c(x - y)$ is some smooth ansatz for $\delta_{x-y}$, defined from some Hilbert–Schmidt kernel $K$ as

$$c(x-y) = \int_{\mathbb{R}} K(x,z)K(y,z)\,\mathrm{d}z$$

We also consider random perturbation of focusing Schrödinger equation, which reads either

$$u_t + \mathrm{i}u_{xx} + \mathrm{i}|u|^{2p}u = u\dot{\xi} \qquad [28]$$

(which represents a multiplicative noise) or

$$u_t + \mathrm{i}u_{xx} + \mathrm{i}|u|^{2p}u = \mathrm{i}\gamma\dot{\xi} \qquad [29]$$

(which is an additive noise). In the former case, the noise acts as a potential, while in the latter case it represents a forcing term. These equations also model the propagation of waves in an inhomogeneous medium.

Research is in progress to study these stochastic dynamical systems. To begin with, the theory of the

initial-value problem has to be established in this new context (see, e.g., de Bouard and Debussche (2003)).

One challenge is to understand the effect of noise on dynamical properties of the particular solutions described above, for instance, the solitary waves for Schrödinger equation, either in the subcritical case $p < 2$ or in the critical case $p = 2$ and beyond.

Results obtained both theoretically and numerically on the influence of the noise on blow-up phenomena (random process) for generalized Schrödinger equations are likely almost-sure results.

On the one hand, if the noise is additive and the power supercritical, $p > 1$, there is some numerical evidence that a spacetime white noise can delay or even prevent the blow-up. However, if the noise is not so irregular (as for the correlated in space noise described above) it seems that any solution blows up in finite time.

de Bouard and Debussche have proved that for either an additive or a multiplicative noise, any smooth and localized (in space) initial data give rise to a trajectory that collapses in arbitrarily small time with a positive probability. This contrasts with the deterministic case, where only particular initial data could lead to blow-up trajectories. Actually, the noise enforces that any trajectory must pass through this blow-up region, with a positive probability.

## Acknowledgment

The author acknowledges Anne de Bouard and Vicentiu Radulescu for helpful comments and remarks.

*See also:* Bifurcations in Fluid Dynamics; Bifurcation Theory; Breaking Water Waves; Cellular Automata; Central Manifolds, Normal Forms; Dissipative Dynamical Systems of Infinite Dimension; Fractal Dimensions in Dynamics; Hamilton–Jacobi Equations and Dynamical Systems: Variational Aspects; Metastable States; Newtonian Fluids and Thermohydraulics; Nonlinear Schrödinger Equations; Quantum Calogero–Moser Systems; Random Dynamical Systems; Scattering, Asymptotic Completeness and Bound States; Stability Problems in Celestial Mechanics; Stochastic Resonance; Symmetry and Symplectic Reduction.

## Further Reading

Bona JL, Chen M, and Saut J-C (2002) Boussinesq equations and other systems for small-amplitude long waves in nonlinear dispersive media. I. Derivation and linear theory. *Journal of Nonlinear Sciences* 12(4): 283–318.

de Bouard A and Debussche A (2003) The stochastic nonlinear Schrödinger equation in $H^1$. *Stochastic Analysis Applications* 21(1): 97–126.

Bourgain J (1999) Global solutions of nonlinear Schrödinger equations. *American Mathematical Society Colloquium Publications, 46*. Providence, RI: American Mathematical Society.

Buffoni B, Séré É, and Toland JF (2003) Surface water waves as saddle points of the energy. *Calcul of Variations in Partial Differential Equations* 17(2): 199–220.

Craig W (1996) KAM theory in infinite dimensions. Dynamical systems and probabilistic methods in partial differential equations (Berkeley, CA, 1994), 31–46. *Lectures in Applied Mathematics,* 31. Providence, RI: American Mathematical Society.

Dangelmayr G, Fiedler B, Kirchgässner K, and Mielke A (1996) Dynamics of nonlinear waves in dissipative systems: reduction, bifurcation and stability. With a contribution by G. Raugel. *Pitman Research Notes in Mathematics Series, 352*. Harlow: Longman.

Dias F and Iooss G (2003) Water-waves as a spatial dynamical system. In: Friedlander S and Serre D (eds.) *Handbook of Mathematical Fluid Dynamics*, chap 10, pp. 443–499. Elsevier.

Gallay T and Wayne CE (2002) Invariant manifolds and the long-time asymptotics of the Navier–Stokes and vorticity equations on $\mathbb{R}^2$. *Archives for Rational and Mechanics Analysis* 163(3): 209–258.

Martel Y and Merle F (2002) Stability of blow-up profile and lower bounds for blow-up rate for the critical generalized KdV equation. *Annals of Mathematics (2)* 155(1): 235–280.

Schneider G and Wayne CE (2000) The long-wave limit for the water wave problem. I. The case of zero surface tension. *Communication in Pure and Applied Mathematics* 53(12): 1475–1535.

Sulem C and Sulem P-L (1999) The nonlinear Schrödinger equation. Self-focusing and wave collapse. *Applied Mathematical Sciences, 139*. New York: Springer-Verlag.

Temam R (1997) Infinite-dimensional dynamical systems in mechanics and physics. Second edition. *Applied Mathematical Sciences, 68*. New York: Springer-Verlag.

Tsutsumi M (1984) Nonexistence of global solutions to the Cauchy problem for the damped nonlinear Schrödinger equations. *SIAM Journal of Mathematics Analysis* 15(2): 357–366.

Whitam GB (1999) Linear and nonlinear waves. *Pure and Applied Mathematics*. New York: Wiley.

# E

## Effective Field Theories

**G Ecker**, Universität Wien, Vienna, Austria



### Introduction

Effective field theories (EFTs) are the counterpart of the "theory of everything." They are the field theoretical implementation of the quantum ladder: heavy degrees of freedom need not be included among the quantum fields of an EFT for a description of low-energy phenomena. For example, we do not need quantum gravity to understand the hydrogen atom nor does chemistry depend upon the structure of the electromagnetic interaction of quarks.

EFTs are approximations by their very nature. Once the relevant degrees of freedom for the problem at hand have been established, the corresponding EFT is usually treated perturbatively. It does not make much sense to search for an exact solution of the Fermi theory of weak interactions. In the same spirit, convergence of the perturbative expansion in the mathematical sense is not an issue. The asymptotic nature of the expansion becomes apparent once the accuracy is reached where effects of the underlying "fundamental" theory cannot be neglected any longer. The range of applicability of the perturbative expansion depends on the separation of energy scales that define the EFT.

EFTs pervade much of modern physics. The effective nature of the description is evident in atomic and condensed matter physics. The present article will be restricted to particle physics, where EFTs have become important tools during the last 25 years.

### Classification of EFTs

A first classification of EFTs is based on the structure of the transition from the "fundamental" (energies $> \Lambda$) to the "effective" level (energies $< \Lambda$).

1. *Complete decoupling* The fundamental theory contains heavy and light degrees of freedom. Under very general conditions (decoupling theorem, Appelquist and Carazzone 1975) the effective Lagrangian for energies $\ll \Lambda$, depending only on light fields, takes the form

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{d\leq 4} + \sum_{d>4} \frac{1}{\Lambda^{d-4}} \sum_{i_d} g_{i_d} O_{i_d} \qquad [1]$$

The heavy fields with masses $> \Lambda$ have been "integrated out" completely. $\mathcal{L}_{d\leq 4}$ contains the potentially renormalizable terms with operator dimension $d \leq 4$ (in natural mass units where Bose and Fermi fields have $d = 1$ and 3/2, respectively), the $g_{i_d}$ are coupling constants and the $O_{i_d}$ are monomials in the light fields with operator dimension $d$. In a slightly misleading notation, $\mathcal{L}_{d\leq 4}$ consists of relevant and marginal operators, whereas the $O_{i_d}$ $(d > 4)$ are denoted irrelevant operators. The scale $\Lambda$ can be the mass of a heavy field (e.g., $M_W$ in the Fermi theory of weak interactions) or it reflects the short-distance structure in a more indirect way.

2. *Partial decoupling* In contrast to the previous case, the heavy fields do not disappear completely from the EFT but only their high-momentum modes are integrated out. The main area of application is the physics of heavy quarks. The procedure involves one or several field redefinitions introducing a frame dependence. Lorentz invariance is not manifest but implies relations between coupling constants of the EFT (reparametrization invariance).

3. *Spontaneous symmetry breaking* The transition from the fundamental to the effective level occurs via a phase transition due to spontaneous symmetry breaking generating (pseudo-)Goldstone bosons. A spontaneously broken symmetry relates processes with different numbers of Goldstone bosons. Therefore, the distinction between renormalizable ($d \leq 4$) and nonrenormalizable ($d > 4$) parts in the effective Lagrangian [1] becomes meaningless. The effective Lagrangian of type 3 is generically nonrenormalizable. Nevertheless, such Lagrangians define perfectly consistent quantum field theories at sufficiently low energies. Instead of the operator dimension as in [1], the number of derivatives of the fields and the number of symmetry-breaking

insertions distinguish successive terms in the Lagrangian. The general structure of effective Lagrangians with spontaneously broken symmetries is largely independent of the specific physical realization (universality). There are many examples in condensed matter physics, but the two main applications in particle physics are electroweak symmetry breaking and chiral perturbation theory (both discussed later) with the spontaneously broken global chiral symmetry of quantum chromodynamics QCD.

Another classification of EFTs is related to the status of their coupling constants.

A. *Coupling constants can be determined by matching the EFT with the underlying theory at short distances*. The underlying theory is known and Green functions can be calculated perturbatively at energies $\sim\Lambda$ both in the fundamental and in the effective theory. Identifying a minimal set of Green functions fixes the coupling constants $g_{i_d}$ in eqn [1] at the scale $\Lambda$. Renormalization group equations can then be used to run the couplings down to lower scales. The nonrenormalizable terms in the Lagrangian [1] can be fully included in the perturbative analysis.
B. *Coupling constants are constrained by symmetries only*.
   - The underlying theory and therefore also the EFT coupling constants are unknown. This is the case of the standard model (SM) (see the next section). A perturbative analysis beyond leading order only makes sense for the known renormalizable part $\mathcal{L}_{d\leq 4}$. The nonrenormalizable terms suppressed by powers of $\Lambda$ are considered at tree level only. The associated coupling constants $g_{i_d}$ serve as bookmarks for new physics. Usually, but not always (cf., e.g., the subsection "Noncommutative spacetime"), the symmetries of $\mathcal{L}_{d\leq 4}$ are assumed to constrain the couplings.
   - The matching cannot be performed in perturbation theory even though the underlying theory is known. This is the generic situation for EFTs of type 3 involving spontaneous symmetry breaking. The prime example is chiral perturbation theory as the EFT of QCD at low energies.

## The SM as an EFT

With the possible exception of the scalar sector to be discussed in the subsection "Electroweak symmetry breaking" the SM is very likely the renormalizable part of an EFT of type 1B. Except for nonzero neutrino masses, the SM Lagrangian $\mathcal{L}_{d\leq 4}$ in [1] accounts for physics up to energies of roughly the Fermi scale $G_F^{-1/2} \simeq 300\,\text{GeV}$.

Since the SM works exceedingly well up to the Fermi scale where the electroweak gauge symmetry is spontaneously broken it is natural to assume that the operators $O_{i_d}$ with $d > 4$, made up from fields representing the known degrees of freedom and including a single Higgs doublet in the SM proper, should be gauge invariant with respect to the full SM gauge group $SU(3)_c \times SU(2)_L \times U(1)_Y$. An almost obvious constraint is Lorentz invariance that will be lifted in the next subsection, however.

These requirements limit the Lagrangian with operator dimension $d = 5$ to a single term (except for generation multiplicity), consisting only of a left-handed lepton doublet $L_L$ and the Higgs doublet $\Phi$:

$$O_{d=5} = \epsilon_{ij}\epsilon_{kl} L_{iL}^{\top} C^{-1} L_{kL} \Phi_j \Phi_l + \text{h.c.} \qquad [2]$$

This term violates lepton number and generates nonzero Majorana neutrino masses. For a neutrino mass of 1 eV, the scale $\Lambda$ would have to be of the order of $10^{13}$ GeV if the associated coupling constant in the EFT Lagrangian [1] is of order 1.

In contrast to the simplicity for $d = 5$, the list of gauge-invariant operators with $d = 6$ is enormous. Among them are operators violating baryon or lepton number that must be associated with a scale much larger than 1 TeV. To explore the territory close to present energies, it therefore makes sense to impose baryon and lepton number conservation on the operators with $d = 6$. Those operators have all been classified (Buchmüller and Wyler 1986) and the number of independent terms is of the order of 80. They can be grouped in three classes.

The first class consists of gauge and Higgs fields only. The corresponding EFT Lagrangian has been used to parametrize new physics in the gauge sector constrained by precision data from LEP. The second class consists of operators bilinear in fermion fields, with additional gauge and Higgs fields to generate $d = 6$. Finally, there are four-fermion operators without other fields or derivatives. Some of the operators in the last two groups are also constrained by precision experiments, with a certain hierarchy of limits. For lepton and/or quark flavor conserving terms, the best limits on $\Lambda$ are in the few TeV range, whereas the absence of neutral flavor changing processes yields lower bounds on $\Lambda$ that are several orders of magnitude larger. If there is new physics in the TeV range flavor changing neutral transitions must be strongly suppressed, a powerful constraint on model building.

It is amazing that the most general renormalizable Lagrangian with the given particle content accounts

for almost all experimental results in such an impressive manner. Finally, we recall that many of the operators of dimension 6 are also generated in the SM via radiative corrections. A necessary condition for detecting evidence for new physics is therefore that the theoretical accuracy of radiative corrections matches or surpasses the experimental precision.

### Noncommutative Spacetime

Noncommutative geometry arises in some string theories and may be expected on general grounds when incorporating gravity into a quantum field theory framework. The natural scale of noncommutative geometry would be the Planck scale in this case without observable consequences at presently accessible energies. However, as in theories with large extra dimensions the characteristic scale $\Lambda_{NC}$ could be significantly smaller. In parallel to theoretical developments to define consistent noncommutative quantum field theories (short for quantum field theories on noncommutative spacetime), a number of phenomenological investigations have been performed to put lower bounds on $\Lambda_{NC}$.

Noncommutative geometry is a deformation of ordinary spacetime where the coordinates, represented by Hermitian operators $\hat{x}_\mu$, do not commute:

$$\left[\hat{x}_\mu, \hat{x}_\nu\right] = \mathrm{i}\theta_{\mu\nu} \qquad [3]$$

The antisymmetric real tensor $\theta_{\mu\nu}$ has dimensions length$^2$ and it can be interpreted as parametrizing the resolution with which spacetime can be probed. In practically all applications, $\theta_{\mu\nu}$ has been assumed to be a constant tensor and we may associate an energy scale $\Lambda_{NC}$ with its nonzero entries:

$$\Lambda_{NC}^{-2} \sim \theta_{\mu\nu} \qquad [4]$$

There is to date no unique form for the noncommutative extension of the SM. Nevertheless, possible observable effects of noncommutative geometry have been investigated. Not unexpected from an EFT point of view, for energies $\ll \Lambda_{NC}$, noncommutative field theories are equivalent to ordinary quantum field theories in the presence of nonstandard terms containing $\theta_{\mu\nu}$ (Seiberg–Witten map). Practically all applications have concentrated on effects linear in $\theta_{\mu\nu}$.

Kinetic terms in the Lagrangian are in general unaffected by the noncommutative structure. New effects arise therefore mainly from renormalizable $d = 4$ interactions terms. For example, the Yukawa coupling $g_Y \bar{\psi}\psi\phi$ generates the following interaction linear in $\theta_{\mu\nu}$:

$$\mathcal{L}_Y^{NC} = g_Y \theta_{\mu\nu}\left(\partial^\mu\bar{\psi}\partial^\nu\psi\phi + \partial^\mu\bar{\psi}\psi\partial^\nu\phi + \bar{\psi}\partial^\mu\psi\partial^\nu\phi\right) \qquad [5]$$

These interaction terms have operator dimension 6 and they are suppressed by $\theta_{\mu\nu} \sim \Lambda_{NC}^{-2}$. The major difference to the previous discussion on physics beyond the SM is that there is an intrinsic violation of Lorentz invariance due to the constant tensor $\theta_{\mu\nu}$. In contrast to the previous analysis, the terms with dimension $d > 4$ do not respect the symmetries of the SM.

If $\theta_{\mu\nu}$ is indeed constant over macroscopic distances, many tests of Lorentz invariance can be used to put lower bounds on $\Lambda_{NC}$. Among the exotic effects investigated are modified dispersion relations for particles, decay of high-energy photons, charged particles producing Cerenkov radiation in vacuum, birefringence of radiation, a variable speed of light, etc. A generic signal of noncommutativity is the violation of angular momentum conservation that can be searched for at the Large Hadron Collider (LHC) and at the next linear collider.

Lacking a unique noncommutative extension of the SM, unambiguous lower bounds on $\Lambda_{NC}$ are difficult to establish. However, the range $\Lambda_{NC} \lesssim 10\,\mathrm{TeV}$ is almost certainly excluded. An estimate of the induced electric dipole moment of the electron (noncommutative field theories violate CP in general to first order in $\theta_{\mu\nu}$) yields $\Lambda_{NC} \gtrsim 100\,\mathrm{TeV}$. On the other hand, if the SM were CP invariant, noncommutative geometry would be able to account for the observed CP violation in $K^0 - \overline{K^0}$ mixing for $\Lambda_{NC} \sim 2\,\mathrm{TeV}$.

### Electroweak Symmetry Breaking

In the SM, electroweak symmetry breaking is realized in the simplest possible way through renormalizable interactions of a scalar Higgs doublet with gauge bosons and fermions, a gauged version of the linear $\sigma$ model.

The EFT version of electroweak symmetry breaking (EWEFT) uses only the experimentally established degrees of freedom in the SM (fermions and gauge bosons). Spontaneous gauge symmetry breaking is realized nonlinearly, without introducing additional scalar degrees of freedom. It is a low-energy expansion where energies and masses are assumed to be small compared to the symmetry-breaking scale. From both perturbative and nonperturbative arguments we know that this scale cannot be much bigger than 1 TeV. The Higgs model can be viewed as a specific example of an EWEFT as long as the Higgs boson is not too light (heavy-Higgs scenario).

The lowest-order effective Lagrangian takes the following form:

$$\mathcal{L}_{EWSB}^{(2)} = \mathcal{L}_B + \mathcal{L}_F \qquad [6]$$

where $\mathcal{L}_F$ contains the gauge-invariant kinetic terms for quarks and leptons including mass terms. In addition to the kinetic terms for the gauge bosons $\boldsymbol{W}_\mu, B_\mu$, the bosonic Lagrangian $\mathcal{L}_B$ contains the characteristic lowest-order term for the would-be-Goldstone bosons:

$$\mathcal{L}_B = \mathcal{L}_{\text{gauge}}^{\text{kin}} + \frac{v^2}{4}\langle D_\mu U^\dagger D^\mu U\rangle \qquad [7]$$

with the gauge-covariant derivative

$$\begin{aligned} D_\mu U &= \partial_\mu U - \mathrm{i}gW_\mu U + \mathrm{i}g' U\hat{B}_\mu \\ W_\mu &= \frac{\vec{\tau}}{2}\boldsymbol{W}_\mu, \qquad \hat{B}_\mu = \frac{\tau_3}{2}B_\mu \end{aligned} \qquad [8]$$

where $\langle\ldots\rangle$ denotes a (two-dimensional) trace. The matrix field $U(\phi)$ carries the nonlinear representation of the spontaneously broken gauge group and takes the value $U = \mathbb{1}$ in the unitary gauge. The Lagrangian [6] is invariant under local $SU(2)_L \times U(1)_Y$ transformations:

$$\begin{aligned} W_\mu &\to g_L W_\mu g_L^\dagger + \frac{\mathrm{i}}{g} g_L \partial_\mu g_L^\dagger \\ \hat{B}_\mu &\to \hat{B}_\mu + \frac{\mathrm{i}}{g'} g_R \partial_\mu g_R^\dagger \\ f_L &\to g_L f_L, \quad f_R \to g_R f_R, \quad U \to g_L U g_R^\dagger \end{aligned} \qquad [9]$$

with

$$\begin{aligned} g_L(x) &= \exp(\mathrm{i}\vec{\alpha}_L(x)\vec{\tau}/2) \\ g_R(x) &= \exp(\mathrm{i}\alpha_Y(x)\tau_3/2) \end{aligned}$$

and $f_{L(R)}$ are quark and lepton fields grouped in doublets.

As is manifest in the unitary gauge $U = \mathbb{1}$, the lowest-order Lagrangian of the EWEFT just implements the tree-level masses of gauge bosons ($M_W = M_Z \cos\theta_W = vg/2$, $\tan\theta_W = g'/g$) and fermions but does not carry any further information about the underlying mechanism of spontaneous gauge symmetry breaking. This information is first encoded in the couplings $a_i$ of the next-to-leading-order Lagrangian

$$\mathcal{L}_{\text{EWSB}}^{(4)} = \sum_{i=0}^{14} a_i O_i \qquad [10]$$

with monomials $O_i$ of $O(p^4)$ in the low-energy expansion. The Lagrangian [10] is the most general CP and $SU(2)_L \times U(1)_Y$ invariant Lagrangian of $O(p^4)$.

Instead of listing the full Lagrangian, we display three typical examples:

$$\begin{aligned} O_0 &= \frac{v^2}{4}\langle TV_\mu\rangle^2 \\ O_3 &= -g\langle W_{\mu\nu}[V^\mu, V^\nu]\rangle \\ O_5 &= \langle V_\mu V^\mu\rangle^2 \end{aligned} \qquad [11]$$

where

$$\begin{aligned} T &= U\tau_3 U^\dagger, \qquad V_\mu = D_\mu U U^\dagger \\ W_{\mu\nu} &= \frac{\mathrm{i}}{g}[\partial_\mu - \mathrm{i}gW_\mu, \partial_\nu - \mathrm{i}gW_\nu] \end{aligned} \qquad [12]$$

In the unitary gauge, the monomials $O_i$ reduce to polynomials in the gauge fields. The three examples in eqn [11] start with quadratic, cubic, and quartic terms in the gauge fields, respectively. The strongest constraints exist for the coefficients of quadratic contributions from the Large Electron–Positron collider LEP1, less restrictive ones for the cubic self-couplings from LEP2, and none so far for the quartic ones.

## Heavy-Quark Physics

EFTs in this section are derived from the SM and they are of type 2A in the classification introduced previously. In a first step, one integrates out $W$, $Z$, and top quark. Evolving down from $M_W$ to $m_b$, large logarithms $\alpha_s(m_b)\ln(M_W^2/m_b^2)$ are resummed into the Wilson coefficients. At the scale of the $b$-quark, QCD is still perturbative, so that at least a part of the amplitudes is calculable in perturbation theory. To separate the calculable part from the rest, the EFTs below perform an expansion in $1/m_Q$, where $m_Q$ is the mass of the heavy quark.

Heavy-quark EFTs offer several important advantages.

1. Approximate symmetries that are hidden in full QCD appear in the expansion in $1/m_Q$.
2. Explicit calculations simplify in general, for example, the summing of large logarithms via renormalization group equations.
3. The systematic separation of hard and soft effects for certain matrix elements (factorization) can be achieved much more easily.

### Heavy-Quark Effective Theory

Heavy-quark effective theory (HQET) is reminiscent of the Foldy–Wouthuysen transformation (nonrelativistic expansion of the Dirac equation). It is a systematic expansion in $1/m_Q$, when $m_Q \gg \Lambda_{QCD}$, the scale parameter of QCD. It can be applied to processes where the heavy quark remains essentially on shell: its velocity $v$ changes only by small amounts $\sim\Lambda_{QCD}/m_Q$. In the hadron rest frame, the heavy quark is almost at rest and acts as a quasistatic source of gluons.

More quantitatively, one writes the heavy-quark momentum as $p^\mu = m_Q v^\mu + k^\mu$, where $v$ is the hadron 4-velocity ($v^2 = 1$) and $k$ is a residual

momentum of $O(\Lambda_{QCD})$. The heavy quark field $Q(x)$ is then decomposed with the help of energy projectors $P_v^{\pm} = (1 \pm \not{v})/2$ and employing a field redefinition:

$$\begin{aligned} Q(x) &= e^{-im_Q v \cdot x}(h_v(x) + H_v(x)) \\ h_v(x) &= e^{im_Q v \cdot x} P_v^+ Q(x) \\ H_v(x) &= e^{im_Q v \cdot x} P_v^- Q(x) \end{aligned} \qquad [13]$$

In the hadron rest frame, $h_v(x)$ and $H_v(x)$ correspond to the upper and lower components of $Q(x)$, respectively. With this redefinition, the heavy-quark Lagrangian is expressed in terms of a massless field $h_v$ and a "heavy" field $H_v$:

$$\begin{aligned} \mathcal{L}_Q &= \bar{Q}(i\not{D} - m_Q)Q \\ &= \overline{h_v}\, iv \cdot D\, h_v - \overline{H_v}(iv \cdot D + 2m_Q)H_v \\ &\quad + \text{mixed terms} \end{aligned} \qquad [14]$$

At the semiclassical level, the field $H_v$ can be eliminated by using the QCD field equation $(i\not{D} - m_Q)Q = 0$ yielding the nonlocal expression

$$\mathcal{L}_Q = \overline{h_v} iv \cdot D h_v + \overline{h_v}\, i\not{D}_\perp \frac{1}{iv \cdot D + 2m_Q - i\epsilon} i\not{D}_\perp h_v \qquad [15]$$

with $D_\perp^\mu = (g^{\mu\nu} - v^\mu v^\nu)D_\nu$. The field redefinition in [13] ensures that, in the heavy-hadron rest frame, derivatives of $h_v$ give rise to small momenta of $O(\Lambda_{QCD})$ only. The Lagrangian [15] is the starting point for a systematic expansion in $m_Q$.

To leading order in $1/m_Q (Q = b, c)$, the Lagrangian

$$\mathcal{L}_{b,c} = \overline{b_v}\, iv \cdot D\, b_v + \overline{c_v} iv \cdot D c_v \qquad [16]$$

exhibits two important approximate symmetries of HQET: the flavor symmetry $SU(2)_F$ relating heavy quarks moving with the same velocity and the heavy-quark spin symmetry generating an overall SU(4) spin-flavor symmetry. The flavor symmetry is obvious and the spin symmetry is due to the absence of Dirac matrices in [16]: both spin degrees of freedom couple to gluons in the same way. The simplest spin-symmetry doublet consists of a pseudoscalar meson $H$ and the associated vector meson $H^*$. Denoting the doublet by $\mathcal{H}$, the matrix elements of the heavy-to-heavy transition current are determined to leading order in $1/m_Q$ by a single form factor, up to Clebsch–Gordan coefficients:

$$\langle \mathcal{H}(v') | \overline{h_{v'}} \Gamma h_v | \mathcal{H}(v) \rangle \sim \xi(v \cdot v') \qquad [17]$$

$\Gamma$ is an arbitrary combination of Dirac matrices and the form factor $\xi$ is the so-called Isgur–Wise function. Moreover, since $\overline{h_v}\gamma^\mu h_v$ is the Noether current of heavy-flavor symmetry, the Isgur–Wise function is fixed in the no-recoil limit $v' = v$ to be $\xi(v \cdot v' = 1) = 1$. The semileptonic decays $B \to Dl\nu_l$ and $B \to D^* l\nu_l$ are therefore governed by a single normalized form factor to leading order in $1/m_Q$, with important consequences for the determination of the Cabibbo–Kobayashi–Maskawa (CKM) matrix element $V_{cb}$.

The HQET Lagrangian is superficially frame dependent. Since the SM is Lorentz invariant, the HQET Lagrangian must be independent of the choice of the frame vector $v$. Therefore, a shift in $v$ accompanied by corresponding shifts of the fields $h_v$ and of the covariant derivatives must leave the Lagrangian invariant. This reparametrization invariance is unaffected by renormalization and it relates coefficients with different powers in $1/m_Q$.

### Soft-Collinear Effective Theory

HQET is not applicable in heavy-quark decays where some of the light particles in the final state have momenta of $O(m_Q)$, for example, for inclusive decays like $B \to X_s\gamma$ or exclusive ones like $B \to \pi\pi$. In recent years, a systematic heavy-quark expansion for heavy-to-light decays has been set up in the form of soft-collinear effective theory (SCET).

SCET is more complicated than HQET because now the low-energy theory involves more than one scale. In the SCET Lagrangian, a light quark or gluon field is represented by several effective fields. In addition to the soft fields $h_v$ in [15], the so-called collinear fields enter that have large energy and carry large momentum in the direction of the light hadrons in the final state.

In addition to the frame vector $v$ of HQET ($v = (1, 0, 0, 0)$ in the heavy-hadron rest frame), SCET introduces a lightlike reference vector $n$ in the direction of the jet of energetic light particles (for inclusive decays), for example, $n = (1, 0, 0, 1)$. All momenta $p$ are decomposed in terms of light-cone coordinates $(p_+, p_-, p_\perp)$ with

$$p^\mu = \frac{n \cdot p}{2}\bar{n}^\mu + \frac{\bar{n} \cdot p}{2} n^\mu + p_\perp^\mu = p_+^\mu + p_-^\mu + p_\perp^\mu \qquad [18]$$

where $\bar{n} = 2v - n = (1, 0, 0, -1)$. For large energies, the three light-cone components are widely separated, with $p_- = O(m_Q)$ being large while $p_\perp$ and $p_+$ are small. Introducing a small parameter $\lambda \sim p_\perp/p_-$, the light-cone components of (hard-)collinear particles scale like $(p_+, p_-, p_\perp) = m_Q(\lambda^2, 1, \lambda)$. Thus, there are three different scales in the problem compared to only two in HQET. For exclusive decays, the situation is even more involved.

The SCET Lagrangian is obtained from the full theory by an expansion in powers of $\lambda$. In addition to the heavy quark field $h_v$, one introduces soft as well as collinear quark and gluon fields by field

redefinitions so that the various fields have momentum components that scale appropriately with $\lambda$.

Similar to HQET, the leading-order Lagrangian of SCET exhibits again approximate symmetries that can lead to a reduction of form factors describing heavy-to-light decays. As in HQET, reparametrization invariance implements Lorentz invariance and results in stringent constraints on subleading corrections in SCET.

An important result of SCET is the proof of factorization theorems to all orders in $\alpha_s$. For inclusive decays, the differential rate is of the form

$$\mathrm{d}\Gamma \sim HJ \times S \qquad [19]$$

where $H$ contains the hard corrections. The so-called jet function $J$ sensitive to the collinear region is convoluted with the shape function $S$ representing the soft contributions. At leading order, the shape function drops out in the ratio of weighted decay spectra for $B \to X_u l \nu_l$ and $B \to X_s \gamma$ allowing for a determination of the CKM matrix element $V_{ub}$. Factorization theorems have become available for an increasing number of processes, most recently also for exclusive decays of $B$ into two light mesons.

#### Nonrelativistic QCD

In HQET the kinetic energy of the heavy quark appears as a small correction of $O(\Lambda^2_{\mathrm{QCD}}/m_Q)$. For systems with more than one heavy quark, the kinetic energy cannot be treated as a perturbation in general. For instance, the virial theorem implies that the kinetic energy in quarkonia $\bar{Q}Q$ is of the same order as the binding energy of the bound state.

NRQCD, the EFT for heavy quarkonia, is an extension of HQET. The Lagrangian for NRQCD coincides with HQET in the bilinear sector of the heavy-quark fields but it also includes quartic interactions between quarks and antiquarks. The relevant expansion parameter in this case is the relative velocity between $Q$ and $\bar{Q}$. In contrast to HQET, there are at least three widely separate scales in heavy quarkonia: in addition to $m_Q$, the relative momentum of the bound quarks $p \sim m_Q v$ with $v \ll 1$ and the typical kinetic energy $E \sim m_Q v^2$. The main challenges are to derive the quark–antiquark potential directly from QCD and to describe quarkonium production and decay at collider experiments. In the abelian case, the corresponding EFT for quantum electrodynamics (QED) is called NRQED that has been used to study electromagnetically bound systems like the hydrogen atom, positronium, muonium, etc.

In NRQCD only the hard degrees of freedom with momenta $\sim m_Q$ are integrated out. Therefore, NRQCD is not enough for a systematic computation of heavy-quarkonium properties. Because the nonrelativistic fluctuations of order $m_Q v$ and $m_Q v^2$ have not been separated, the power counting in NRQCD is ambiguous in higher orders.

To overcome those deficiencies, two approaches have been put forward: potential NRQCD (pNRQCD) and velocity NRQCD (vNRQCD). In pNRQCD, a two-step procedure is employed for integrating out quark and gluon degrees of freedom:

$$\begin{array}{cc} \mathrm{QCD} & \Lambda > m_Q \\ \Downarrow & \\ \mathrm{NRQCD} & m_Q > \Lambda > m_Q v \\ \Downarrow & \\ \mathrm{pNRQCD} & m_Q v > \Lambda > m_Q v^2 \end{array}$$

The resulting EFT derives its name from the fact that the four-quark interactions generated in the matching procedure are the potentials that can be used in Schrödinger perturbation theory. It is claimed that pNRQCD can also be used in the nonperturbative domain where $\alpha_s(m_Q v^2)$ is of order 1 or larger. The advantage would be that also charmonium becomes accessible to a systematic EFT analysis.

The alternative approach of vNRQCD is only applicable in the fully perturbative regime when $m_Q \gg m_Q v \gg m_Q v^2 \gg \Lambda_{\mathrm{QCD}}$ is valid. It separates the different degrees of freedom in a single step leaving only ultrasoft energies and momenta of $O(m_Q v^2)$ as continuous variables. The separation of larger scales proceeds in a similar fashion as in HQET via field redefinitions. A systematic nonrelativistic power counting in the velocity $v$ is implemented.

### The Standard Model at Low Energies

At energies below 1 GeV, hadrons – rather than quarks and gluons – are the relevant degrees of freedom. Although the strong interactions are highly nonperturbative in the confinement region, Green functions and amplitudes are amenable to a systematic low-energy expansion. The key observation is that the QCD Lagrangian with $N_f = 2$ or 3 light quarks,

$$\begin{aligned} \mathcal{L}_{\mathrm{QCD}} &= \bar{q}\big(\mathrm{i}\not{D} - \mathcal{M}_q\big)q - \tfrac{1}{4}G^{\alpha}_{\mu\nu}G^{\alpha\mu\nu} + \mathcal{L}_{\text{heavy quarks}} \\ &= \overline{q_{\mathrm{L}}}\mathrm{i}\not{D}q_{\mathrm{L}} + \overline{q_{\mathrm{R}}}\mathrm{i}\not{D}q_{\mathrm{R}} - \overline{q_{\mathrm{L}}}\mathcal{M}_q q_{\mathrm{R}} \\ &\quad - \overline{q_{\mathrm{R}}}\mathcal{M}_q q_{\mathrm{L}} + \cdots \\ q_{\mathrm{R,L}} &= \tfrac{1}{2}(1 \pm \gamma_5)q, \quad q^{\top} = (ud[s]) \end{aligned} \qquad [20]$$

exhibits a global symmetry

$$\underbrace{\mathrm{SU}(N_f)_{\mathrm{L}} \times \mathrm{SU}(N_f)_{\mathrm{R}}}_{\text{chiral group } G} \times \mathrm{U}(1)_{\mathrm{V}} \times \mathrm{U}(1)_{\mathrm{A}} \qquad [21]$$

in the limit of $N_f$ massless quarks ($\mathcal{M}_q = 0$). At the hadronic level, the quark number symmetry $U(1)_V$ is realized as baryon number. The axial $U(1)_A$ is not a symmetry at the quantum level due to the abelian anomaly.

Although not yet derived from first principles, there are compelling theoretical and phenomenological arguments that the ground state of QCD is not even approximately chirally symmetric. All evidence, such as the existence of relatively light pseudoscalar mesons, points to spontaneous chiral symmetry breaking $G \to SU(N_f)_V$, where $SU(N_f)_V$ is the diagonal subgroup of $G$. The resulting $N_f^2 - 1$ (pseudo-)Goldstone bosons interact weakly at low energies. In fact, Goldstone's theorem ensures that purely mesonic or single-baryon amplitudes vanish in the chiral limit ($\mathcal{M}_q = 0$) when the momenta of all pseudoscalar mesons tend to zero. This is the basis for a systematic low-energy expansion of Green functions and amplitudes. The corresponding EFT (type 3B in our classification) is called chiral perturbation theory (CHPT) (Weinberg 1979, Gasser and Leutwyler 1984, 1985).

Although the construction of effective Lagrangians with nonlinearly realized chiral symmetry is well understood, there are some subtleties involved. First of all, there may be terms in a chiral-invariant action that cannot be written as the four-dimensional integral of an invariant Lagrangian. The chiral anomaly for $SU(3) \times SU(3)$ bears witness of this fact and gives rise to the Wess–Zumino–Witten action. A general theorem to account for such exceptional cases is due to D'Hoker and Weinberg (1994). Consider the most general action for Goldstone fields with symmetry group $G$, spontaneously broken to a subgroup $H$. The only possible non-$G$-invariant terms in the Lagrangian that give rise to a $G$-invariant action are in one-to-one correspondence with the generators of the fifth cohomology group $\mathcal{H}^5(G/H; \mathbb{R})$ of the coset manifold $G/H$. For the relevant case of chiral $SU(N)$, the coset space $SU(N)_L \times SU(N)_R / SU(N)_V$ is itself an $SU(N)$ manifold. For $N \geq 3, \mathcal{H}^5(SU(N); \mathbb{R})$ has a single generator that corresponds precisely to the Wess–Zumino–Witten term.

At a still deeper level, one may ask whether chiral-invariant Lagrangians are sufficient (except for the anomaly) to describe the low-energy structure of Green functions as dictated by the chiral Ward identities of QCD. To be able to calculate such Green functions in general, the global chiral symmetry of QCD is extended to a local symmetry by the introduction of external gauge fields. The following invariance theorem (Leutwyler 1994) provides an answer to the above question. Except for the anomaly, the most general solution of the Ward identities for a spontaneously broken symmetry in Lorentz-invariant theories can be obtained from gauge-invariant Lagrangians to all orders in the low-energy expansion. The restriction to Lorentz invariance is crucial: the theorem does not hold in general in nonrelativistic effective theories.

### Chiral Perturbation Theory

The effective chiral Lagrangian of the SM in the meson sector is displayed in Table 1. The lowest-order Lagrangian for the purely strong interactions is given by

$$\mathcal{L}_{p^2} = \frac{F^2}{4}\langle D_\mu U D^\mu U^\dagger \rangle + \frac{F^2 B}{2}\langle (s + ip)U^\dagger + (s - ip)U \rangle \qquad [22]$$

with a covariant derivative $D = \partial_\mu U - i(v_\mu + a_\mu)U + iU(v_\mu - a_\mu)$. The first term has the familiar form [7] of the gauged nonlinear $\sigma$ model, with the matrix field $U(\phi)$ transforming as $U \to g_R U g_L^\dagger$ under chiral rotations. External fields $v_\mu, a_\mu, s, p$ are introduced for constructing the generating functional of Green functions of quark currents. To implement explicit chiral symmetry breaking, the scalar field $s$ is set equal to the quark mass matrix $\mathcal{M}_q$ at the end of the calculation.

The leading-order Lagrangian has two free parameters $F$, $B$ related to the pion decay constant and to the quark condensate, respectively:

$$F_\pi = F\left[1 + O(m_q)\right]$$
$$\langle 0|\bar{u}u|0\rangle = -F^2 B\left[1 + O(m_q)\right] \qquad [23]$$

The Lagrangian [22] gives rise to $M_\pi^2 = B(m_u + m_d)$ at lowest order. From detailed studies of pion–pion scattering (Colangelo *et al.* 2001), we know that the leading term accounts for at least 94% of the pion mass. This supports the standard counting of CHPT,

**Table 1** The effective chiral Lagrangian of the SM in the meson sector

| $\mathcal{L}_{\text{chiral order}}$ (# *of LECs*) | *Loop order* |
|---|---|
| $\mathcal{L}_{p^2}(2) + \mathcal{L}^{\Delta S=1}_{G_F p^2}(2) + \mathcal{L}^{em}_{e^2p^0}(1) + \mathcal{L}^{emweak}_{G_8 e^2 p^0}(1)$ | $L = 0$ |
| $+\mathcal{L}_{p^4}(10) + \mathcal{L}^{odd}_{p^6}(32) + \mathcal{L}^{\Delta S=1}_{G_8 p^4}(22) + \mathcal{L}^{\Delta S=1}_{G_{27} p^4}(28)$ | $L = 1$ |
| $+\mathcal{L}^{em}_{e^2p^2}(14) + \mathcal{L}^{emweak}_{G_8 e^2 p^2}(14) + \mathcal{L}^{leptons}_{e^2 p}(5)$ | |
| $+\mathcal{L}_{p^6}(90)$ | $L = 2$ |

The numbers in brackets refer to the number of independent couplings for $N_f = 3$.. The parameter-free Wess–Zumino–Witten action $S_{WZW}$ that cannot be written as the four-dimensional integral of an invariant Lagrangian must be added.

with quark masses booked as $O(p^2)$ like the two-derivative term in [22].

The effective chiral Lagrangian in Table 1 contains the following parts:

1. strong interactions: $\mathcal{L}_{p^2}, \mathcal{L}_{p^4}, \mathcal{L}_{p^6}^{\text{odd}}, \mathcal{L}_{p^6} + S_{\text{WZW}}$;
2. nonleptonic weak interactions to first order in the Fermi coupling constant $G_F$: $\mathcal{L}_{G_F p^2}^{\Delta S=1}$, $\mathcal{L}_{G_8 p^4}^{\Delta S=1}, \mathcal{L}_{G_{27} p^4}^{\Delta S=1}$;
3. radiative corrections for strong processes: $\mathcal{L}_{e^2 p^0}^{\text{em}}, \mathcal{L}_{e^2 p^2}^{\text{em}}$;
4. radiative corrections for nonleptonic weak decays: $\mathcal{L}_{G_8 e^2 p^0}^{\text{emweak}}, \mathcal{L}_{G_8 e^2 p^2}^{\text{emweak}}$; and
5. radiative corrections for semileptonic weak decays: $\mathcal{L}_{e^2 p}^{\text{leptons}}$.

Beyond the leading order, unitarity and analyticity require the inclusion of loop contributions. In the purely strong sector, calculations have been performed up to next-to-next-to-leading order. Figure 1 shows the corresponding skeleton diagrams of $O(p^6)$, with full lowest-order tree structures to be attached to propagators and vertices. The coupling constants of the various Lagrangians in Table 1 absorb the divergences from loop diagrams leading to finite renormalized Green functions with scale-dependent couplings, the so-called low-energy constants (LECs). As in all EFTs, the LECs parametrize the effect of "heavy" degrees of freedom that are not represented explicitly in the EFT Lagrangian. Determination of those LECs is a major task for CHPT. In addition to phenomenological information, further theoretical input is needed. Lattice gauge theory has already furnished values for some LECs. To bridge the gap between the low-energy domain of CHPT and the perturbative domain of QCD, large-$N_c$ motivated interpolations with meson resonance exchange have been used successfully to pin down some of the LECs.

Especially in cases where the knowledge of LECs is limited, renormalization group methods provide valuable information. As in renormalizable quantum field theories, the leading chiral logs $(\ln M^2/\mu^2)^L$ with a typical meson mass $M$, renormalization scale $\mu$ and loop order $L$ can in principle be determined from one-loop diagrams only. In contrast to the renormalizable situation, new derivative structures (and quark mass insertions) occur at each loop order preventing a straightforward resummation of chiral logs.

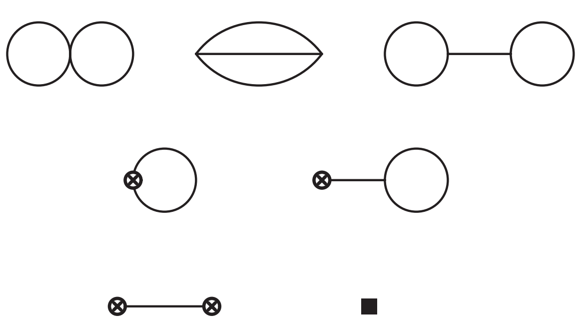


**Figure 1** Skeleton diagrams of $O(p^6)$.. Normal vertices are from $\mathcal{L}_{p^2}$, crossed circles and the full square denote vertices from $\mathcal{L}_{p^4}$ and $\mathcal{L}_{p^6}$, respectively.

Among the many applications of CHPT in the meson sector are the determination of quark mass ratios and the analysis of pion–pion scattering where the chiral amplitude of next-to-next-to-leading order has been combined with dispersion theory (Roy equations). Of increasing importance for precision physics (CKM matrix elements, $(g-2)_\mu, \ldots$) are isospin-violating corrections including radiative corrections, where CHPT provides the only reliable approach in the low-energy region. Such corrections are also essential for the analysis of hadronic atoms like pionium, a $\pi^+\pi^-$ bound state.

CHPT has also been applied extensively in the single-baryon sector. There are several differences to the purely mesonic case. For instance, the chiral expansion proceeds more slowly and the nucleon mass $m_N$ provides a new scale that does not vanish in the chiral limit. The formulation of heavy-baryon CHPT was modeled after HQET integrating out the nucleon modes of $O(m_N)$. To improve the convergence of the chiral expansion in some regions of phase space, a manifestly Lorentz-invariant formulation has been set up more recently (relativistic baryon CHPT). Many single-baryon processes have been calculated to fourth order in both approaches, for example, pion–nucleon scattering. With similar methods as in the mesonic sector, hadronic atoms like pionic or kaonic hydrogen have been investigated.

### Nuclear Physics

In contrast to the meson and single-baryon sectors, amplitudes with two or more nucleons do not vanish in the chiral limit when the momenta of Goldstone mesons tend to zero. Consequently, the power counting is different in the many-nucleon sector. Multinucleon processes are treated with different EFTs depending on whether all momenta are smaller or larger than the pion mass.

In the very low energy regime $|\vec{p}| \ll M_\pi$, pions or other mesons do not appear as dynamical degrees of freedom. The resulting EFT is called "pionless EFT" and it describes systems like the deuteron, where the typical nucleon momenta are $\sim\sqrt{m_N B_d} \simeq 45\,\text{MeV}$ ($B_d$ is the binding energy of the deuteron). The Lagrangian for the strong interactions between two nucleons has the form

$$\mathcal{L}_{NN} = C_0 \left(N^\top P_i N\right)^\dagger N^\top P_i N + \cdots \qquad [24]$$

where $P_i$ are spin–isospin projectors and higher-order terms contain derivatives of the nucleon fields. The existence of bound states implies that at least part of the EFT Lagrangian must be treated nonperturbatively. Pionless EFT is an extension of effective-range theory that has long been used in nuclear physics. It has been applied successfully especially to the deuteron but also to more complicated few-nucleon systems like the Nd and n$\alpha$ systems. For instance, precise results for Nd scattering have been obtained with parameters fully determined from NN scattering. Pionless EFT has also been applied to the so-called halo nuclei, where a tight cluster of nucleons (like $^4$He) is surrounded by one or more "halo" nucleons.

In the regime $|\vec{p}| > M_\pi$, the pion must be included as a dynamical degree of freedom. With some modifications in the power counting, the corresponding EFT is based on the approach of Weinberg (1990, 1991), who applied the usual rules of the meson and single-nucleon sectors to the nucleon–nucleon potential (instead of the scattering amplitude). The potential is then to be inserted into a Schrödinger equation to calculate physical observables. The systematic power counting leads to a natural hierarchy of nuclear forces, with only two-nucleon forces appearing up to next-to-leading order. Three- and four-nucleon forces arise at third and fourth order, respectively.

Significant progress has been achieved in the phenomenology of few-nucleon systems. The two- and $n$-nucleon ($3 \leq n \leq 6$) sectors have been pushed to fourth and third order, respectively, with encouraging signs of "convergence." Compton scattering off the deuteron, $\pi$d scattering, nuclear parity violation, solar fusion, and other processes have been investigated in the EFT approach. The quark mass dependence of the nucleon–nucleon interaction has also been studied.

*See also:* Anomalies; Electroweak Theory; High $T_c$ Superconductor Theory; Noncommutative Geometry and the Standard Model; Operator Product Expansion in Quantum Field Theory; Perturbation Theory and its Techniques; Quantum Chromodynamics; Quantum Electrodynamics and its Precision Tests; Renormalization: General Theory; Seiberg–Witten Theory; Standard Model of Particle Physics; Symmetries and Conservation Laws; Symmetry Breaking in Field Theory.

## Further Reading

Appelquist T and Carazzone J (1975) Infrared singularities and massive fields. *Physical Review D* 11: 2856.

Bedaque PF and van Kolck U (2002) Effective field theory for few nucleon systems. *Annual Review of Nuclear and Particle Science* 52: 339 (nucl-th/0203055).

Brambilla N, Pineda A, Soto J, and Vairo A (2004) Effective field theories for quarkonium. *Reviews of Modern Physics* (in press) (hep-ph/0410047).

Buchmüller W and Wyler D (1986) Effective Lagrangian analysis of new interactions and flavour conservation. *Nuclear Physics B* 268: 621.

Colangelo G, Gasser J, and Leutwyler H (2001) $\pi\pi$ scattering. *Nuclear Physics B* 603: 125 (hep-ph/0103088).

D'Hoker E and Weinberg S (1994) General effective actions. *Physical Review D* 50: 6050 (hep-ph/9409402).

Ecker G (1995) Chiral perturbation theory. *Progress in Particle and Nuclear Physics* 35: 1 (hep-ph/9501357).

Feruglio F (1993) The chiral approach to the electroweak interactions. *International Journal of Modern Physics A* 8: 4937 (hep-ph/9301281).

Gasser J and Leutwyler H (1984) Chiral perturbation theory to one loop. *Annals of Physics* 158: 142.

Gasser J and Leutwyler H (1985) Chiral perturbation theory: expansions in the mass of the strange quark. *Nuclear Physics B* 250: 465.

Hinchliffe I, Kersting N, and Ma YL (2004) Review of the phenomenology of noncommutative geometry. *International Journal of Modern Physics A* 19: 179 (hep/0205040).

Hoang AH (2002) Heavy quarkonium dynamics. In: Shifman M (ed.) *At the Frontier of Particle Physics/Handbook of QCD*, vol. 4. Singapore: World Scientific. (hep-ph/0204299).

Leutwyler H (1994) On the foundations of chiral perturbation theory. *Annals of Physics* 235: 165 (hep-ph/9311274).

Mannel T (2004) *Effective Field Theories in Flavour Physics*, Springer Tracts in Modern Physics, vol. 203. Berlin: Springer.

Manohar AV and Wise MB (2000) Heavy quark physics. *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.* 10: 1.

Meißner U (2005) Modern theory of nuclear forces. *Nuclear Physics A* 751: 149 (nucl-th/0409028).

Pich A (1999) Effective field theory. In: Gupta R *et al.* (eds.) *Proc. of Les Houches Summer School 1997: Probing the Standard Model of Particle Interactions*. Amsterdam: Elsevier. (hep-ph/9806303).

Scherer S (2002) Introduction to chiral perturbation theory. *Advances in Nuclear Physics* 27: 277 (hep-ph/0210398).

Weinberg S (1979) Phenomenological Lagrangians. *Physica A* 96: 327.

Weinberg S (1990) Nuclear forces from chiral Lagrangians. *Physics Letters B* 251: 288.

Weinberg S (1991) Effective chiral Lagrangians for nucleon–pion interactions and nuclear forces. *Nuclear Physics B* 363: 3.

# Eigenfunctions of Quantum Completely Integrable Systems

**J A Toth**, McGill University, Montreal, QC, Canada



## Introduction

This article is an introduction to eigenfunctions of quantum completely integrable (QCI) systems. For these systems, one can understand asymptotics of eigenfunctions better than for other systems, so it is natural to study them. It is useful to begin the discussion with the most important geometric example given by the quantum Hamiltonian, $P_1 = -\sqrt{\Delta}$. We fix a basis of eigenfunctions, $\varphi_j, j = 1, 2, \ldots$, with

$$-\sqrt{\Delta}\varphi_j = \lambda_j\varphi_j, \quad \langle\varphi_i, \varphi_j\rangle = \delta_{ij}$$

and assume that there exist functionally independent (pseudo)differential operators $P_2, \ldots, P_n$ with the property that

$$[P_i, P_j] = 0, \quad i, j = 1, \ldots, n$$

In this case, $P_1$ is said to be QCI and the operators, $P_k, k = 1, \ldots, n$, can be simultaneously diagonalized. It is therefore natural to study the special basis of Laplace eigenfunctions which are joint eigenvectors of the $P_k's$. From now on, the $\phi_j$'s are always assumed to be joint eigenfunctions of the commuting operators, $P_k, k = 1, \ldots, n$. The classical observables corresponding to the operators $P_k, k = 1, \ldots, n$, are the respective principal symbols, $p_k \in C^\infty(T^*M), j = 1, \ldots, n$. In particular, the bicharacteristic flow of $p_1(x, \xi) = |\xi|_g$ is the classical "geodesic flow"

$$G_t : T^*M \longrightarrow T^*M$$

Examples of manifolds with QCI Laplacians include tori and spheres of revolution, Liouville metrics on tori and spheres, large families of metrics on homogeneous spaces, as well as hyperellipsoids with distinct axes in arbitrary dimension. There are also many inhomogeneous QCI examples (see the next section). It is of interest to understand the asymptotics of both eigenvalues and eigenfunctions. There is a large literature devoted to eigenvalue asymptotics, including trace formulas and Bohr–Sommerfeld rules (see Colin de Verdiere (1994a, b), Helffer and Sjoestrand (1990), and Colin de Verdiere and Vu Ngoc (2003)). We will concentrate here on the corresponding problem of determining eigenfunction asymptotics. The key property of eigenfunctions in the QCI case is localization in phase space, $T^*M$. This allows one to study more effectively the concentration and blow-up properties than in any other setting. It is important to contrast this with, for example, the situation in the ergodic case. Moreover, in the QCI case, there is a particularly strong connection between dynamics of the geodesic flow, $G_t : T^*M \to T^*M$, and the asymptotics of individual eigenfunctions. In the general case, one can usually only relate the dynamics to spectral averages, such as in the trace formula (Duistermaat and Guillemin 1975).

For the most part, the literature on eigenfunction asymptotics addresses the following basic problems:

1. determining sharp upper and lower bounds for $\varphi_j$ as $\lambda_j \to \infty$ and
2. describing the link between the blow-up properties of $\varphi_j$ as $\lambda_j \to \infty$ and the dynamics of the geodesic flow, $G_t$.

The starting point in the study of eigenfunction asymptotics in the QCI case is the fact that the joint eigenfunctions, $\varphi_j$, have masses that localize on the level sets, $\mathcal{P}^{-1}(b) := \{(x, \xi) \in T^*M; p_j(x, \xi) = b_j, j = 1, \ldots, n\}$. Moreover, by the Liouville–Arnol'd theorem, for generic levels (indexed by $b \in \mathbb{R}$),

$$\mathcal{P}^{-1}(b) = \sum_{k=1}^{m} \Lambda_k(b) \qquad [1]$$

where the $\Lambda_k(b) \subset T^*M$ are Lagrangian tori. The affine symplectic coordinates in a neighborhood of $\Lambda_k(b)$ are called "action-angle variables" $(\theta_1^{(k)}, \ldots, \theta_n^{(k)}; I_1^{(k)}, \ldots, I_n^{(k)}) \in \mathbb{T}^n \times \mathbb{R}^n$. Written in terms of these coordinates, the classical Hamilton equations defining the geodesic flow assume the form

$$\frac{d\theta}{dt} = F(I), \qquad \frac{dI}{dt} = 0$$

and this system of ordinary differential equations (ODEs) is solved by quadrature. This explains why one refers to such systems as completely integrable. At the quantum level, one can construct semiclassical Lagrangian distributions,

$$\Phi_\lambda(x) := \int_{\mathbb{R}^n} e^{i\lambda\varphi^{(k)}(x,\eta)} a(x, \eta; \lambda)\, d\eta \qquad [2]$$

which microlocally concentrate on $\Lambda^{(k)}(b)$ as $\lambda \to \infty$ and satisfy $P_j\Phi_\lambda = b_j\lambda\Phi_\lambda + \mathcal{O}(\lambda^{-\infty})$ in $L^2(M)$. An important fact is that the actual joint eigenfunctions, $\varphi_j$, are approximated to $\mathcal{O}(\lambda^{-\infty})$-accuracy in $L^2(M)$ by suitable linear combinations of the quasimodes, $\Phi_\lambda$. However, there are subtleties underlying this correspondence which are often neglected in the physics literature:

3. The actual joint eigenfunctions $\phi_j$ localize on the level sets $\mathcal{P}^{-1}(b)$ which usually consist of many

connected components. Consequently, the eigenfunctions are approximated by (sometimes large) linear combinations of Lagrangian quasimodes attached to the different component tori. The precise splitting of mass amongst these different components is a difficult and, in general, unsolved problem in microlocal tunneling.

4. The local torus foliation given by action-angle variables tends to degenerate and Lagrangian quasimodes are no longer approximate solutions to the (joint) eigenvalue equations near the singularities of the foliation. The singularities and their relative configurations can be complicated (Colin de Verdiere and Vu Ngoc 2003) and most of the interesting asymptotic blow-up properties of eigenfunctions tend to be associated with these degeneracies. The main tool for studying joint eigenfunctions near degeneracies is the quantum analog of the Eliasson normal form (Eliasson 1984, Vu Ngoc 2000). We will refer to this as the "quantum Birkhoff normal form" (QBNF).

## Background on QCI Systems

Let $(M^n, g)$ be a compact, closed Riemannian manifold and $P_1 := Op_\hbar(p_1)$ be a formally self-adjoint, elliptic (in the classical sense) $\hbar$-pseudodifferential operator. In local coordinates, the Schwarz kernel of $P_1$ is of the form,

$$P_1(x, y; \hbar) = (2\pi\hbar)^{-n} \int_{\mathbb{R}^n} e^{i\langle x-y,\xi\rangle/\hbar} p_1(x, \xi; \hbar)\, d\xi$$

where $p_1(x,\xi;\hbar) \in S^{0,m}_{cl}(T^*M)$; that is, $p_1(x,\xi;\hbar) \sim \sum_{j=0}^{\infty} p_{1,j}(x,\xi)\hbar^j$ with $\partial_x^\alpha \partial_\xi^\beta p_{1,j}(x,\xi) = \mathcal{O}_{\alpha,\beta}\langle\xi\rangle^{m-j-|\beta|}$ (Dimassi and Sjoestrand 1999). It is often convenient to work with $\hbar$-pseudodifferential operators rather than their classical counterparts. In the homogeneous case, one chooses $\hbar^{-1} \in \mathrm{Spec}\sqrt{\Delta}$.

$P_1 \in Op_\hbar(S^{0,m}_{cl})$ is said to be QCI if there exist self-adjoint $P_j = Op_\hbar(p_j) \in Op_\hbar(S^{0,m'}_{cl}), j = 2, \ldots, n$, for some $m'$ with $[P_i, P_j] = 0, i, j = 1, \ldots, n$, such that $dp_1 \wedge \cdots \wedge dp_n \neq 0$ on a dense open subset, $\Omega_{reg} \subset T^*M$, and $P_1^2 + \cdots + P_n^2$ is elliptic in the classical sense. There are many inhomogeneous QCI examples including quantum Euler, Lagrange, and Kowalevsky tops together with quantum Neumann and Rosochatius oscillators in arbitrary dimension.

Since $\{p_i, p_j\} = 0$, the joint Hamilton flow of the $p_i$'s induces a symplectic $\mathbb{R}^n$-action on $T^*M$:

$$\Phi_t : T^*M \longrightarrow T^*M$$
$$\Phi_t(x,\xi) = \exp t_1 H_{p_1} \circ \cdots \circ \exp t_1 H_{p_n}(x,\xi)$$
$$t = (t_1, \ldots, t_n) \in R^n$$

The associated moment map is just

$$\mathcal{P} : T^*M - 0 \to R^n, \quad \mathcal{P} = (p_1, \ldots, p_n)$$

We denote the image $\mathcal{P}(T^*M - 0)$ by $\mathcal{B}$, the regular values (resp. singular values) by $\mathcal{B}_{reg}$ (resp. $\mathcal{B}_{sing}$) of the moment map.

To establish bounds for the joint eigenfunctions of $P_1, \ldots, P_n$, one imposes a "finite-complexity" assumption (Toth and Zelditch 2002) on the classical integrable system. This condition holds for all systems of interest in physics. To describe it, for each $b = (b^{(1)}, \ldots, b^{(n)}) \in \mathcal{B}$, let $m_{cl}(b)$ denote the number of $R^n$-orbits of the joint flow $\Phi_t$ on the level set $\mathcal{P}^{-1}(b)$. Then, the finite-complexity condition says that for some $M_0 > 0$,

$$m_{cl}(b) < M_0 (\forall b \in \mathcal{B})$$

In addition, when $\mathcal{P}$ is proper,

$$\mathcal{P}^{-1}(b) = \sum_{k=1}^{m_{cl}(b)} \Lambda_k(b) \qquad [3]$$

for any $b \in \mathcal{B}_{reg}$, where the $\Lambda_k(b)$ are Lagrangian tori. The starting point for analyzing joint eigenfunctions is the following correspondence principle (Zelditch 1990) which makes the eigenfunction localization alluded to in the introduction more precise:

**Theorem 1** *Let $Op_\hbar(a) \in Op_\hbar(S^0_{cl})(T^*M)$ and $P_j, j = 1, \ldots, n$, be a QCI system of commuting operators. Then, for every $b \in \mathcal{B}_{reg}$, there exists a subsequence of joint eigenfunctions $\varphi_\mu(x) := \varphi(x;\mu(\hbar))$ with $\hbar \in (0, \hbar_0]$ and joint eigenvalues $\mu(\hbar) = (\mu_1(\hbar), \ldots, \mu_n(\hbar)) \in \mathrm{Spec}(P_1, \ldots, P_n)$ with $|\mu(\hbar) - b| = \mathcal{O}(\hbar)$ such that*

$$\langle Op_\hbar(a)\varphi_\mu, \varphi_\mu\rangle = |c(\hbar)|^2 \int_{\Lambda(b)} a(x,\xi)\, d\mu_b + \mathcal{O}(\hbar)$$

*Here, $d\mu_b$ denotes Lebesgue measure on the torus, $\Lambda(b)$.*

The proof of Theorem 1 follows from the $\hbar$-microlocal, regular quantum normal construction near $\Lambda(b)$ (see the section "Birkhoff normal forms").

## Blow-Up of Eigenfunctions: Qualitative Results

Before discussing quantitative bounds for joint eigenfunctions, it is useful to prove qualitative results. Here, we review only the homogeneous case where $P_1 = \hbar\sqrt{\Delta}$, although the general case can be dealt with similarly (Toth and Zelditch 2002). Two well-known QCI examples which exhibit extremes in eigenfunction concentration are the round sphere and the flat torus. In the case of the sphere, the zonal harmonics blow-up like $\lambda^{1/2}$ at the poles, whereas, in the case of the flat torus, all the joint eigenfunctions are uniformly bounded. The rest of the article will be

essentially devoted to understanding these extreme blow-up properties (and intermediate ones) more systematically. When discussing blow-up of eigenfunctions, it is natural to start with the following:

**Question** Do there exist QCI manifolds (other than the flat torus) for which all eigenfunctions are uniformly bounded in $L^\infty$?

Toth and Zelditch (2002) have proved that, up to coverings, the flat torus is the only example with uniformly bounded eigenfunctions. Their argument used the correspondence principle in Theorem 1 combined with some deep results from symplectic geometry. To deal with the issue of multiplicities, it is convenient to define

$$L^\infty(\lambda; g) = \sup_{\varphi \in V_\lambda} \|\varphi\|_{L^\infty}$$

where $V_\lambda = \{\varphi; P_1\varphi_\lambda = \lambda\varphi_\lambda\}$ and it is assumed that $\|\varphi\|_{L^2} = 1$.

**Theorem 2** (Toth and Zelditch 2002). *Suppose that $P_1 = \sqrt{\Delta}$ is QCI on a compact, Riemannian manifold $(M, g)$ and suppose that the corresponding moment map satisfies the finite-complexity condition. Then, if $L^\infty(\lambda, g) = \mathcal{O}(1), (M, g)$ is flat.*

The proof of Theorem 2 follows by contradiction: that is, one assumes that all eigenfunctions are uniformly bounded. There are two main steps in the proof of Theorem 2: the first is entirely analytic and uses the correspondence principle in Theorem 1 and uniform boundedness to determine the topology of $M$. The second step uses two deep results from symplectic topology/geometry to determine the metric, $g$, up to coverings.

Using a local Weyl law argument and the finite-multiplicity assumption, it can be shown that for each $b \in \mathcal{B}_{reg}$, there exists a subsequence, $\varphi_\mu$, of joint eigenfunctions such that Proposition 1 holds with

$$|c(\hbar; b)|^2 \geq \frac{1}{C}$$

where $C > 0$ is a uniform constant not depending on $b \in \mathcal{B}_{reg}$. With this subsequence, one applies Theorem 1 with $a(x, \xi) = V(x) \in C^\infty(M)$. It then easily follows by the boundedness assumption that for $\hbar$ sufficiently small and appropriate constants $C_0, C_1 > 0$,

$$\left| \int_{\Lambda(b)} \left(\pi^*_{\Lambda(b)} V\right) \mathrm{d}\mu_b \right| \leq \frac{1}{C_0} \int_M |V(x)|\, |\varphi_\mu(x)|^2 \,\mathrm{dVol}(x) \leq \frac{1}{C_1} \int_M |V(x)| \,\mathrm{dVol}(x) \qquad [4]$$

where $\pi_{\Lambda(b)}$ denotes the restriction of the canonical projection $\pi : T^*M \to M$ to the Lagrangian $\Lambda(b)$. The estimate in [4] is equivalent to the statement,

$$(\pi_{\Lambda(b)})_*(\mathrm{d}\mu_b) \ll \mathrm{dVol}(x)$$

where given two Borel measures $\mathrm{d}\mu$ and $\mathrm{d}\nu$, one writes $\mathrm{d}\mu \ll \mathrm{d}\nu$ if $\mathrm{d}\mu$ is absolutely continuous with respect to $\mathrm{d}\nu$. Consequently, $\pi_{\Lambda(b)} : \Lambda(b) \to M$ has no singularities and thus, up to coverings, $M$ is topologically a torus (since $\Lambda(b)$ is).

Since there are many QCI systems on $n$-tori, it still remains to determine how the uniform-boundedness condition constrains the metric geometry of $(M, g)$. First, by a classical result of Mane, if $T^*M$ possesses a $C^1$-foliation by Lagrangians, $(M, g)$ cannot have conjugate points. By the first step in the proof, it follows that under the uniform-boundedness assumption, $M$ is a topological torus and $T^*M$ possesses a smooth foliation by Lagrangian tori. Consequently, $(M, g)$ has no conjugate points. Finally, the Burago–Ivanov proof of the Hopf conjecture says that metric tori without conjugate points are flat. Therefore, $(M, g)$ is flat.

Consistent with Theorem 2, one can show (Toth and Zelditch 2003, Lerman and Shirokova 2002) that if $(M, g)$ is integrable and not a flat torus, then there must exist a compact $\Phi_t$-orbit (i.e., an orbit of the joint flow of $X_{p_j}, j = 1, \ldots, n$) with dim $= k < n$. In the QCI case, these "singular" orbits trap eigenfunction mass for appropriate subsequences. To understand this statement in detail, it is necessary to review QBNF constructions in the context of QCI systems.

## Birkhoff Normal Forms

There are several excellent expositions on the topic of Birkhoff normal forms in the literature (see, e.g., Guillemin (1996), Iatchenko *et al.* (2002), and Zelditch (1998)), which discuss both the classical and quantum constructions. Here, we discuss the aspects which are most relevant for QCI systems.

Consider the Schrödinger operator, $P(x; \hbar D_x) = -\hbar^2(\mathrm{d}^2/\mathrm{d}x^2) + V(x)$ with $V(x + 2\pi) = V(x)$ acting on $C^\infty(\mathbb{R}/2\pi\mathbb{Z})$. Assume that the potential, $V(x)$, is Morse and that $x = 0$ is a potential minimum with $V(0) = V'(0) = 0$ and $\Omega \subset T^*(\mathbb{S}^1)$ an open neighborhood containing $(0, 0)$. In its simplest incarnation, the classical Birkhoff normal-form theorem says that for small enough $\Omega$, there exists a symplectic diffeomorphism, $\kappa^{-1} : (\Omega; (0, 0)) \longrightarrow (\Omega; (0, 0)); \kappa^{-1} : (x, \xi) \mapsto (y, \eta)$, and a (locally defined) function $F_0 \in C^\infty(\mathbb{R})$ such that

$$(p \circ \kappa)(y, \eta) = F_0(\eta^2 + y^2) \qquad [5]$$

provided $(y,\eta)\in\Omega$. At the quantum level, the analogous QBNF expansion says that there exist microlocally unitary $\hbar$-Fourier integral operators, $U(\hbar): C^\infty(\Omega)\to C^\infty(\Omega)$ and a classical symbol $F(x,\hbar)\sim\sum_{j=0}^{\infty}F_j(x)\hbar^j$, such that

$$U(\hbar)^*\circ P(\hbar)\circ U(\hbar)={}_\Omega F(\hat{I}_e;\hbar) \qquad [6]$$

with $\hat{I}_e=\hbar^2D_y^2+y^2$. Given two $\hbar$-pseudos $P$ and $Q$, the notation $P={}_\Omega Q$ means that $\|\chi(P-Q)\|_{L^2}\|_0=\mathcal{O}(\hbar^\infty)$ and $\|(P-Q)\chi\|_0=\mathcal{O}(\hbar^\infty)$, for any $\chi\in C_0^\infty(\Omega)$. Since it can be easily shown that eigenfunctions $\varphi_\mu$, with $\mu(\hbar)=\mathcal{O}(\hbar^\beta)$, $0<\beta\le 1$, localize very sharply near $x=0$, from the $\hbar$-microlocal unitary equivalence in [6], the eigenfunction and eigenvalue asymptotics (including trace formulas) can all be determined by working with the model operator on the right-hand side (RHS) of [6]. Moreover, on the model side, the eigenfunctions and eigenvalues are explicitly known.

At a potential maximum, there exist classical and quantum normal forms analogous to [5] and [6] (see Helffer and Sjoestrand (1990) and Colin de Verdiere and Parisse (1994a)) except that the harmonic oscillator action operator, $\hat{I}_e$, is replaced by the hyperbolic action operator,

$$\hat{I}_h=\hbar\left(yD_y+\tfrac{1}{2}\right) \qquad [7]$$

The 1D Schrödinger operator is the simplest example of a QCI system where $(0,0)\in T^*\mathbb{S}^1$ is a nondegenerate critical point of the classical Hamiltonian, $H(x,\xi)=\xi^2+V(x)$. Under a mild nondegeneracy hypothesis (Vu Ngoc 2000), there is an analogous normal form for arbitrary QCI systems which is valid near nondegenerate rank $k<n$ orbits of the joint flow, $\Phi_t$. At the classical level, this result is due to Eliasson (1984) and the quantum analog is due to Vu-Ngoc (2000). To state the result is general, one has to define the appropriate model operators: these are $\hat{I}_e$ and $\hat{I}_h$ together with the loxodromic model operators $\Re\hat{I}_{ch}=\hbar D_\theta, \Im\hat{I}_{ch}=\hbar\rho D_\rho+\hbar/i$, where $(\rho,\theta)$ denote polar coordinates in $\mathbb{R}^2$. The local model phase space for a rank $k<n$ orbit, $\mathcal{O}_k$, is just $T^*(\mathbb{T}^k)\times T^*(\mathbb{R}^{n-k})$. In this case, the QBNF says that, for a sufficiently small neighborhood, $\mathcal{G}_k$ of $\mathcal{O}_k$, there exists a family of $\hbar$-Fourier integral operators, $U_\kappa: C^\infty(\mathcal{G}_k)\to C^\infty(T^*(\mathbb{T}^k)\times T^*(\mathbb{R}^{n-k}))$ and symbols $f_j(\hbar)\sim\sum_{j=0}^{\infty}f_j\hbar^j$, such that

$$\begin{aligned} U_\kappa^*P_jU_\kappa &={}_{\mathcal{G}_k}\mathcal{M}_\hbar\cdot(Q_1-f_1(\hbar),\ldots,Q_n-f_n(\hbar)) \\ U_\kappa^*\circ U_\kappa &={}_{\mathcal{G}_k}\mathrm{Id} \end{aligned} \qquad [8]$$

Here, $\mathcal{M}_\hbar$ is a microlocally invertible matrix of $\hbar$-pseudodifferential operators commuting with the $Q_j$'s, and the $Q_j$'s are to be chosen from the list of model operators $\{\hat{I}_{ch},\hat{I}_h,\hat{I}_e,\hat{I}_{reg}\}$, where $\hat{I}_{reg}=(\hbar D_{\theta_1},\ldots,\hbar D_{\theta_k})$ denotes the regular model operator acting along the $k$-dimensional orbit, $\mathcal{O}_k$. Moreover, if $(y_1,\ldots,y_{n-k},\eta_1,\ldots,\eta_{n-k})\in T^*(\mathbb{R}^{n-k})$ denote the symplectic model coordinates, then the $Q_j$'s act in separate, complementary $(y_1,\ldots,y_{n-k})$-variables. The main point here is that [8] is actually a convergent normal norm in $\hbar$ in the sense that error terms in [8] are $\mathcal{O}(\hbar^\infty)$. In contrast (Guillemin 1996, Iatchenko *et al.* 2002, Zelditch 1998), the general Birkhoff normal form is only formal in the sense that error terms vanish to successively higher orders along the orbit, $\mathcal{O}_k$, but are not necessarily small in terms of the spectral parameter, $\hbar$.

Using [8], it can be shown that the joint eigenfunctions, $\varphi_\mu$, are microlocally determined in terms of the $\hbar$-Fourier integral operators, $U_\kappa$, and certain model eigenfunctions. More precisely,

$$U_\kappa^*\varphi_\mu(\theta,y;\hbar)={}_{\mathcal{G}_k}c(\hbar)\,\mathrm{e}^{\mathrm{i}m\theta}\cdot[u_h\cdot u_{ch}\cdot u_e](y;\hbar) \qquad [9]$$

where $m\in\mathbb{Z}+1/4, c(\hbar)\in\mathbb{C}(\hbar)$. The generalized eigenfunctions of the model operators, $\hat{I}_h,\hat{I}_{ch},\hat{I}_e$, acting transversely to the orbit $\mathcal{O}_k$ are $u_h(y;\mu,\hbar)=c_+(\hbar)|y|_+^{-1/2+\mathrm{i}\mu/\hbar}+c_-(\hbar)|y|_-^{-1/2+\mathrm{i}\mu/\hbar}$, $u_{ch}(\rho,\theta;t,k,\hbar)=\rho^{\mathrm{i}t/\hbar-1}\mathrm{e}^{\mathrm{i}k\theta}$ and $u_e(y;n,\hbar)=H_n(\hbar^{-1/2}y)$, where $H_n(y)$ is the $n$th Hermite function.

## Eigenfunction Lower Bounds: Quantitative Results

Let $\mathcal{O}_k$ be a singular rank $k<n$ orbit as in the previous section. From the qualitative results of the first section, it follows that there must exist joint eigenfunctions, $\varphi_\mu$, of the commuting operators, $P_j, j=1,\ldots,n$, which blow up along the orbit, $\mathcal{O}_k$. To obtain quantitative results, one could try to determine the $L^p\to L^q$ mapping properties of the $\hbar$-Fourier integral operator, $U_\kappa$. However, since the canonical transformation $\kappa$ to normal form can be complicated, this method is quite cumbersome. To avoid this complication (Toth and Zelditch 2003), it suffices to compute $L^2$-masses only, but on scales of order $\hbar^\delta$ where $0\le\delta<1/2$. Let $\pi(\mathcal{G}_k(\hbar^\delta))$ be the configuration space projection of the $\hbar^\delta$-radius tube $\mathcal{G}_k(\hbar^\delta)\supset\mathcal{O}_k$. Since

$$\|\varphi_\mu\|_{L^\infty}^2\cdot\mathrm{Vol}(\pi(\mathcal{G}_k(\hbar^\delta)))\ge\int_{\mathcal{G}_k(\hbar^\delta)}|\varphi_\mu|^2\,\mathrm{dVol} \qquad [10]$$

one is reduced to estimating $\int_{\mathcal{G}_k(\hbar^\delta)}|\phi_\mu|^2\,\mathrm{dVol}$ from below. To bound this integral from below, it suffices to

1. reduce the estimate to one involving only the model eigenfunctions in the Birkhoff normal form and
2. estimate the normalizing $\hbar$-dependent constant $c(\hbar)$ in [9].

To prove (1) one introduces a cutoff function $\chi(x,\xi;\hbar^{\delta}) \in C_0^{\infty}(\mathcal{G}_k(\hbar^{\delta}))$ and is identically equal to one near $\mathcal{O}_k$. Then, since $\pi^{-1}(\pi(\mathcal{G}_k(\hbar^{\delta}))) \supset \mathcal{G}_k(\hbar^{\delta})$, from the Garding inequality, it follows that

$$\int_{\mathcal{G}_k(\hbar^{\delta})} |\varphi_\mu|^2 \, \mathrm{dVol} \gg \langle Op_\hbar(\chi(x,\xi;\hbar^{\delta}))\varphi_\mu, \varphi_\mu\rangle \quad [11]$$

In light of the QBNF result in [8], the computation of the matrix element on the RHS of [11] is reduced to a corresponding computation for the $L^2$-normalized model eigenfunctions. Since the $U_\kappa$'s are microlocally unitary, it follows that

$$\langle Op_\hbar(\chi(x,\xi;\hbar^{\delta}))\varphi_\mu, \varphi_\mu\rangle \sim_{\hbar\to 0^+} C(\delta)\cdot |c(\hbar)|^2 \quad [12]$$

Here, the constant $C(\delta)>0$ depends only on the scale of the cutoff function. It finally remains to deal with (2). Bounding the size of $|c(\hbar)|$ from below amounts to estimating the $L^2$-mass of the joint eigenfunction $\varphi_\mu$ which must be trapped near the orbit, $\mathcal{O}_k$. Using a local (singular) Weyl law argument, it is shown in Toth and Zelditch (2003) that

$$|c(\hbar)|^2 \gg |\log \hbar|^{-\beta} \quad [13]$$

where $\beta>0$ indexes the number of hyperbolic and loxodromic model operators. The final result quantifies blow-up along a compact orbit:

**Theorem 3** (Toth and Zelditch 2003). *Let $\mathcal{O}_k$ be a rank $k<n$ orbit of the joint flow $\Phi_t$. If this orbit is compact and nondegenerate, then there exists a subsequence of $L^2$-normalized joint eigenfunctions $\varphi_{\lambda_{j_k}}, k=1,2,\ldots,$ of the QCI system $P_j, j=1,\ldots,n$, such that for any $\epsilon>0$,*

$$\|\varphi_{\lambda_{j_k}}\|_{L^\infty} \gg_\epsilon \lambda_{j_k}^{(n-k/4)-\epsilon}$$

By using the semiclassical scale $\hbar^{1/2}|\log\hbar|^{1/2}$, one can (slightly) improve the lower bound in Theorem 3 to $\|\varphi_{\lambda_{j_k}}\|_{L^\infty} \gg_\epsilon \lambda_{j_k}^{n-k/4}|\log\lambda_{j_k}|^{-\alpha}$ for some $\alpha\geq 0$ (see Sogge *et al.* (2005)).

When $(M,g)$ is not flat, there must exist a singular, compact orbit of dimension $k$ with $1\leq k\leq n-1$ and so, as an immediate corollary of Theorem 3, it follows that for some $\alpha\geq 0$,

$$L^\infty(\lambda;g) \gg \lambda^{1/4}|\log\lambda|^{-\alpha} \quad [14]$$

Since the bound in [14] is highly dependent on dimension, establishing the existence of high-codimension singular orbits would strengthen the estimate substantially. However, this appears to be a difficult and open problem.

## Maximal Blow-Up of Modes and Quasimodes

We review here a number of converses to a recent result of Sogge and Zelditch (2002) on Riemannian manifolds $(M,g)$ with maximal eigenfunction growth. These authors proved that if there exists a sequence of $L^2$-normalized eigenfunctions of the Laplacian $\Delta$ of $(M,g)$ whose $L^\infty$-norms are comparable to zonal spherical harmonics on $S^n$, then there must exist a point comparable to the north pole of $S^n$, that is, a recurrent point $z$ such that a positive measure of geodesics emanating from $z$ return to it at a fixed time $T$. The most extreme kind of recurrent point is a "blow-down point" of period $T$, where by definition all geodesics leaving $z$ return to $z$ at time $T$, that is, form geodesic loops. Poles of surfaces of revolution are blow-down points where all geodesic loops at $z$ are smoothly closed, while umbilic points of triaxial ellipsoids are examples of blow-down points where all but two geodesic loops are not smoothly closed. On real-analytic manifolds, all recurrent points are blow-down points. The converse question is the following: what kind of mode (eigenfunction) or quasimode growth must occur when a blow-down point exists?

Sogge *et al.* (2005) proved that maximal quasimode growth (Colin de Verdiere 1977) implies the existence of a blow-down point. This generalizes the main result of Sogge and Zelditch (2002) from modes (which one rarely understands) to quasimodes (which one often understands better). Conversely, existence of a blow-down point insures near-maximal quasimode growth, that is, here, maximal up to logarithmic factors. If one assumes that the geodesic flow $G^t: T^*M \to T^*M$ of $(M,g)$ is completely integrable and that $\dim M = 2$, then the results of Sogge *et al.* (2005) show that actual eigenfunctions have near maximal blow-up. Examples show that, in general, blow-up points do not necessarily cause modes to have near-maximal blow-up.

An important geometric invariant of a blow-down point is the first-return map to the cotangent fiber over the blow-down point:

$$G_z^T : S_z^*M \to S_z^*M \quad [15]$$

$G_z^T$ is also an important analytic invariant: the blow-up rate of modes or quasimodes, specifically the occurrence of the logarithmic factors, depends on the fixed-point structure of this map. When all geodesic loops at $z$ are smoothly closed, that is, when the first-return map is the identity, then there exist quasimodes of maximal growth. When the first-return map has fixed points, the maximal growth is modified by logarithmic factors.

To put these results in context, we first recall the local Weyl law of Avakumovich–Levitan (Duistermaat and Guillemin 1975), which states that

$$\sum_{\lambda_\nu \le \lambda} |\varphi_\nu(x)|^2 = (2\pi)^{-n} \int_{p(x,\xi)\le\lambda} \mathrm{d}\xi + R(\lambda, x) \qquad [16]$$

with uniform remainder bounds

$$|R(\lambda, x)| \le C\lambda^{n-1} \quad x \in M$$

It follows that

$$L^\infty(\lambda, g) = 0(\lambda^{(n-1)/2}) \qquad [17]$$

on any compact Riemannian manifold. Riemannian manifolds for which the equality

$$L^\infty(\lambda, g) = \Omega(\lambda^{(n-1)/2}) \qquad [18]$$

is achieved for some subsequence of eigenfunctions are said to be of maximal eigenfunction growth. In addition to modes, and almost inseparable from them, are the quasimodes of the Laplacian (Colin de Verdiere 1977). As the name suggests, quasimodes are approximate eigenfunctions. The crudest type of quasimode is quasimode $\{\psi_k\}$ of order 0, namely a sequence of $L^2$-normalized functions which solve

$$\|(\Delta - \mu_k)\psi_k\|_{L^2} = O(1)$$

for a sequence of quasieigenvalues $\mu_k$. By the spectral theorem, it follows that there must exist true eigenvalues in the interval $[\mu_k - \delta, \mu_k + \delta]$ for some $\delta > 0$. $(M, g)$ is said to have maximal 0-order quasimode growth if there exists a sequence of quasimodes of order 0 for which $\|\psi_k\|_{L^\infty} = \Omega(\lambda^{(n-1)/2})$. There are analogous definitions for more refined quasimodes, for example, quasimodes of higher order or (most refined) quasimodes defined by oscillatory integrals. It is natural to include quasimodes in this study because they often reflect the geometry and dynamics of the geodesic flow more strongly than actual modes. For quasimodes, there is the following result:

**Theorem 4** (Sogge *et al.* 2005). Let $(M^n, g)$ *be a compact Riemannian manifold with Laplacian* $\Delta$. *Then*:

(i) *If there exists a quasimode sequence* $\{(\psi_k, \mu_k)\}$ *of order 0 with* $\|\psi_k\|_{L^\infty} = \Omega(\mu_k^{(n-1)/2})$, *then there exists a recurrent point* $z \in M$ *for the geodesic flow. If* $(M, g)$ *is real analytic, then there exists a blow-down point.*

(ii) *Conversely, if there exists a blow-down point and if the map* $G_z^T = \mathrm{id}$, *then there exists a quasimode sequence* $\{(\psi_k, \mu_k)\}$ *of order 0 with* $\|\psi_k\|_{L^\infty} = \Omega(\mu_k^{n-1})$.

(iii) *Let* $n = 2$ *and* $(M^n, g)$ *be real analytic. Then, if* $G_z^T$ *has a finite number of nondegenerate fixed points, there exists a quasimode sequence* $\{(\psi_k, \mu_k)\}$ *of order 0 with* $\|\psi_k\|_{L^\infty} = \Omega(\mu_k^{1/2} \times |\log \mu_k|^{-1/2})$.

The assumption that $G_z^T = \mathrm{id}$ is the same as saying that all geodesics leaving $z$ smooth close up at $z$ again. As mentioned above, poles of surfaces of revolution have this property. On the contrary, the umbilic points of triaxial ellipsoids in $\mathbb{R}^3$ are blow-down points for which $G_z^T \ne \mathrm{id}$. That is, every geodesic leaving an umbilic point returns at the same time, but only two closed geodesics in this family are closed, and they give rise to fixed points of $G_z^t$. One can show (see Toth 1996) that there exists a sequence of eigenfunctions in this case for which $L^\infty(g, \lambda) \sim \lambda^{1/2} |\log \lambda|^{-1/2}$. Hence, the above result is sharp. Moreover, it is clear from the proof that the fixed points are responsible for the logarithmic correction to maximal eigenfunction growth: they cause a change in the normal form of the Laplacian near the blow-down point.

Theorem 4 illustrates the intimate connection between maximal blow-up of quasimodes and existence of blow-down points. It is natural to ask, however, when blow-down points cause blow-up in modes, that is, actual eigenfunctions. As mentioned above, this is not generally the case and some further mechanism is needed to ensure it. In the case of QCI surfaces, one can prove:

**Theorem 5** (Sogge *et al.* 2005). *Let* $(M, g)$ *be a smooth, compact surface,* $P_1 = \sqrt{\Delta}$, $P_2$ *be an Eliasson nondegenerate QCI system on* $M$ *and* $\varphi_k$ *be an* $L^2$*-normalized joint eigenfunction of* $P_1, P_2$ *with* $\sqrt{\Delta}\varphi_k = \lambda_k \varphi_k$. *Suppose that there exists a blow-down point* $z \in M$ *for the geodesic flow* $G_t := \exp tX_{p_1}$. *Then, there exists a subsequence of (joint) Laplace eigenfunctions,* $\varphi_{j_k}, k = 1, 2, \ldots$, *such that for any* $\epsilon > 0$,

$$\|\varphi_{j_k}\|_{L^\infty} \gg_\epsilon \lambda_{j_k}^{(1/2)-\epsilon}$$

The role of complete integrability is to force joint eigenfunctions to localize on level sets of the moment map and thus to blow up at blow-down points. The proofs of Theorems 4 and 5 are similar. To prove the latter, by the same reasoning as in the orbit case (Theorem 3), one needs to bound from below the integral

$$\int_{B(z;\hbar^\delta)} |\varphi_\mu|^2 \mathrm{dVol} \qquad [19]$$

for an appropriate subsequence of $\varphi_\mu$s, where $B(z;\hbar^\delta)$ denotes a ball of radius $\hbar^\delta$ centered at the blow-down point, $z \in M$. The blow-down condition implies that $S^*_z M \subset \mathcal{P}^{-1}(b)$ for some $b \in \mathcal{B}$. The relevant subsequence of eigenfunctions, $\varphi_\mu$, are the ones with joint eigenvalues satisfying $|\mu(\hbar) - b| = \mathcal{O}(\hbar)$. Since the eigenfunctions $\varphi_\mu$ are microlocally concentrated on the set $\mathcal{P}^{-1}(b)$, by Gärding,

$$\int_{B(z;\hbar^\delta)} |\varphi_\mu|^2 \mathrm{dVol} \gg \langle Op_\hbar(\chi(x,\xi;\hbar^\delta))\varphi_\mu, \varphi_\mu\rangle \quad [20]$$

where $\chi(x,\xi,\hbar^\delta)$ is a cutoff localized on an $\hbar^\delta$-neighborhood of $\Omega = \pi^{-1}(z) \cap \mathcal{P}^{-1}(b)$. The matrix elements on the RHS of [20] are estimated by passing to QBNF. The subtlety here lies in the choice of scale, $\delta$. For $0 < \delta < 1/2$, the $\hbar$-pseudodifferential operators $Op_\hbar(\chi(x,\xi;\hbar^\delta))$ are contained in a standard calculus (Dimassi and Sjoestrand 1999) and so they automatically satisfy the $\hbar$-Egorov theorem. In particular, the passage to normal form by conjugating with the $U_\kappa$'s is automatic. The crucial point here is that to obtain the (near)-maximal blow-up near a blow-down point $z \in M$, one needs to able to choose $0 \le \delta < 1$. Using second-microlocal methods similar to the ones in Sjoestrand and Zworski (1999), it is shown in Sogge *et al.* (2005) that the blow-down geometry implies that the microlocal cutoffs are contained in an $\hbar$-pseudodifferential operator calculus and, in particular, the relevant $\hbar$-Egorov theorem needed to pass to QBNF is satisfied for any $0 \le \delta < 1$. Then, by explicit computation for the model eigenfunctions, one can show that

$$Op_\hbar(\chi(x,\xi;\hbar^\delta))\varphi_\mu, \varphi_\mu\rangle \gg_\delta \hbar^\delta \quad [21]$$

for any $\delta$ with $0<\delta<1$. The result in Theorem 5 then follows from the bound

$$\|\varphi_\mu\|^2_{L^\infty} \cdot \mathrm{Vol}(B(z;\hbar^\infty)) \gg_\delta \hbar^\delta \quad [22]$$

where one takes $\delta$ arbitrarily close to 1. By analyzing the $U_\kappa s$ carefully (Sogge *et al.* 2005), the lower bound in Theorem 5 can be improved slightly by replacing the $\lambda^{-\epsilon}$ by $|\log\lambda|^{-\alpha}$ for some $\alpha > 0$, although the sharp constant, $\alpha > 0$, appears to be difficult to determine in general. In cases where the geometry of the first-return map, $G^T_z$, is particularly simple, one can sometimes get sharp $|\log\lambda|$-power improvements in Theorem 5 (see Theorem 4 (iii)).

## Eigenfunction Upper Bounds: Quantitative Results

In light of the $\Omega$-bounds in Theorem 5, it is natural to ask whether there are analogous upper bounds for $L^\infty(\lambda;g)$ in the QCI case. The following result holds in the case of real-analytic surfaces:

**Theorem 6** (Sogge *et al.* 2005). *Let $(M,g)$ be a real-analytic Riemannian 2-manifold and $P_1 = \sqrt{\Delta}$ and $P_2$ be a QCI system on $(M,g)$ where, the principal symbol, $p_2$, of $P_2$ is a metric form on $T^*M$.*

(i) *If $M \cong T^2$,*

$$L^\infty(\lambda;g) = \mathcal{O}(\lambda^{1/4})$$

(ii) *If $M \cong S^2$, let $M_{\mathrm{rec}}$ be the set of completely recurrent points for the geodesic flow, $G_t : T^*M \to T^*M$ and let $\Omega_{\mathrm{rec}} \subset M$ be an open neighborhood of $M_{\mathrm{rec}}$. Then,*

$$L^\infty(\lambda;g)|_{M-\Omega_{\mathrm{rec}}} = \mathcal{O}(\lambda^{1/4})$$

An old result of Kozlov says that if the surface $(M,g)$ is analytic, then topologically either $M \cong S^2$ or $M \cong T^2$, so that the estimates in Theorem 6 cover all possible cases in two dimensions. The assumptions in Theorem 6 are satisfied in many examples including surfaces of revolution, Liouville surfaces, and ellipsoids with distinct axes in $\mathbb{R}^3$.

The proof of Theorem 6 follows from a pointwise (joint) trace formula argument (Duistermaat and Guillemin 1975). Namely, in Sogge *et al.* (2005), it is shown that if there are no blow-down points for $G_t$, then for appropriate $\rho \in \mathcal{S}(R)$ with $\rho \ge 0$ and $\hat\rho \in C_0^\infty(R)$,

$$\sum_{j=1}^{\infty} \rho\Big(\hbar^{-1}\Big[\mu_j^{(1)}(\hbar) - b_1\Big]\Big) \cdot \rho\Big(\hbar^{-1}\Big[\mu_j^{(2)}(\hbar) - b_2\Big]\Big) \times |\varphi_\mu(x;\hbar)|^2 = \mathcal{O}(\hbar^{-1/2}) \quad [23]$$

where the estimate in [23] is uniform in $x \in M$ and locally uniform in $b = (b_1, b_2) \in \mathcal{B}$. Part (ii) follows from this. To prove part (i), one applies a simple homological argument to show that if $M \cong T^2$, there cannot exist blow-down points for the geodesic flow (see also Sogge and Zelditch (2002)).

## Open Problems

Most questions related to eigenfunction blow-up are completely open and general results are rare (Sogge and Zelditch 2002). Specific results/conjectures in the ergodic case can be found in Quantum Ergodicity and Mixing of Eigenfunctions. We would like to point out here some specific questions related to the above results in the QCI case:

1. All the known examples with blow-down points turn out to be integrable. Is this necessarily always the case?
2. Does the maximal bound $L^\infty(\lambda;g) \sim \lambda^{(n-1)/2}$ necessarily imply that $(M,g)$ is QCI?

3. At the other extreme, does the minimal bound $L^\infty(\lambda; g) \sim 1$ necessarily imply that $(M, g)$ is flat, or do there exist nonflat manifolds (which are necessarily not QCI) satisfying $L^\infty(\lambda; g) \sim 1$?

## Acknowledgmnt

The research of J A Toth was partially supported by NSERC grant OGP0170280 and a William Dawson Fellowship.

*See also:* Functional Equations and Integrable Systems; Quantum Ergodicity and Mixing of Eigenfunctions.

## Further Reading


Colin de Verdiere Y and Parisse B (1994a) Equilibre instable en régime semi-classique I: concentration microlocale. *Communications in Partial Differential Equations* 19: 1535–1563.

Colin de Verdiere Y and Parisse B (1994b) Equilibre instable en régime semi-classique II: conditions de Bohr–Sommerfeld. *Ann. Inst. Henri Poincaré. Phys. Theor.* 61(3): 347–367.

Colin de Verdiere Y and Vu Ngoc S (2003) Singular Bohr–Sommerfeld rules for 2D integrable systems. *Annales Scientifiques de L'école Normale Supérieure* 4(36): 1–55.

Colin de Verdiere Y (1977) Quasi-modes sur les variétés Riemanniennes compactes. *Inventiones Mathematicae* 43: 15–52.

Dimassi M and Sjoestrand J (1999) Spectral Asymptotics in the Semi-Classical Limit. *London Math. Soc. Lecture Notes*, vol. 1. 268. Cambridge: Cambridge University Press.

Duistermaat J and Guillemin V (1975) The spectrum of positive elliptic operators and periodic bicharacteristics. *Inventiones Mathematicae* 29: 39–79.

Eliasson LH (1984) *Hamiltonian Systems with Poisson Commuting Integrals*. Ph.D. thesis, University of Stockholm.

Guillemin V (1996) Wave-trace invariants. *Duke Mathematical Journal* 83(2): 287–352.

Helffer B and Sjoestrand J (1990) Semiclassical analysis of Harper's equation III. *Bull. Soc. Math. France*, Mémoire No. 39 (1990).

Iatchenko A, Sjoestrand J, and Zworski M (2002) Birkhoff normal forms in semi-classical inverse problems. *Mathematical Research Letters* 9(2–3): 337–362.

Lerman E and Shirokova N (2002) Completely integrable torus actions on symplectic cones. *Mathematical Research Letters* 9(1): 105–115.

Sjoestrand J and Zworski M (1999) Asymptotic distribution of resonances for convex obstacles. *Acta Mathematica* 183: 191–253.

Sogge CD, Toth JA, and Zelditch S (2005) Blow-up of modes and quasimodes at blow-down points. Preprint.

Sogge CD and Zelditch S (2002) Riemannian manifolds with maximal eigenfunction growth. *Duke Mathematical Journal* 114(3): 387–437.

Toth JA (1996) Eigenfunction localization in the quantized rigid body. *Journal of Differential Geometry* 43(4): 844–858.

Toth JA and Zelditch S (2002) Riemannian manifolds with uniformly bounded eigenfunctions. *Duke Mathematical Journal* 111(1): 97–132.

Toth JA and Zelditch S (2003) $L^p$-norms of eigenfunctions in the completely integrable case. *Annales Henri Poincaré* 4: 343–368.

Vu-Ngoc S (2000) Formes normales semi-classiques des systèmes complètement integrables au voisinage d'un point critique de l'application moment. *Asymptotic Analysis* 24(3–4): 319–342.

Zelditch S (1990) Quantum transition amplitudes for ergodic and completely integrable systems. *Journal of Functional Analysis* 94: 415–436.

Zelditch S (1998) Wave invariants for non-degenerate closed geodesics. *Geometric and Functional Analysis* 8: 179–217.


# Eight Vertex and Hard Hexagon Models

**P A Pearce**, University of Melbourne, Parkville, VIC, Australia



## Introduction

The goal of statistical mechanics is to calculate the macroscopic properties of matter from a knowledge of the fundamental interactions between the constituent microscopic components. For simplicity, let us assume discrete states. The mathematical problem, as formulated by Gibbs, is then to calculate the partition function

$$Z_N = \sum_{\text{states } \sigma} e^{-\beta H(\sigma)} \qquad [1]$$

where $\beta = 1/k_B T$ is the inverse temperature, $k_B$ is the Boltzmann constant, and the Hamiltonian $H$ describes the interaction energy of the state $\sigma$ of the $N$ constituent degrees of freedom. The formidable nature of the problem ensues from the fact that $Z_N$ is needed in the limit of an arbitrarily large system to obtain the bulk free energy $\psi(T)$ or partition function per site $\kappa$ in the thermodynamic limit

$$-\beta\psi(T) = \lim_{N\to\infty} \frac{1}{N} \log Z_N = \log \kappa \qquad [2]$$

This limit generally exists because the free energy of a finite system is extensive, that is, it grows proportionally with the system size. Once the bulk free energy is known, the other thermodynamic potentials are obtained, in principle, by taking derivatives with respect to the temperature $T$ and other thermodynamic fields such as the volume $V$ or the external magnetic field $h$. Phase transitions and the accompanying critical phenomena are associated with singularities of the bulk free energy as a function of the thermodynamic fields. Up until the beginning of the 1970s, there were

only a handful of two-dimensional lattice models that had yielded exact solution, most notably, the Ising model (free-fermion or dimer model), the spherical model, the square ice, and six-vertex models. This situation changed dramatically with Baxter's solution of the eight-vertex and hard-hexagon models. The methods developed by Baxter make it possible to solve an infinite plethora of two-dimensional lattice models. In this article, we compare and contrast the remarkable properties of these two prototypical models that played such a pivotal role in the emergence of the modern theory of Yang–Baxter integrability.

## Definition of the Models

### Eight-Vertex Model

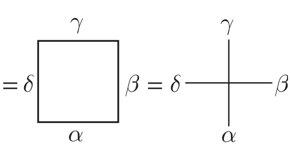
The eight-vertex model emerged from the study of two-dimensional ferroelectrics. The local degrees of freedom are arrow states $\alpha, \beta, \gamma, \delta = \pm 1$ which live on the edges of the elementary faces of the square lattice and describe the local polarization within the ferroelectric material. Of the 16 possible configurations around a face, the local configurations of an elementary square face are restricted to the eight configurations shown in Figure 1.

The partition function is

$$Z_N = \sum_{\text{arrow states}} \prod_{\text{faces}} W\begin{pmatrix} & \gamma & \\ \delta & & \beta \\ & \alpha & \end{pmatrix} \qquad [3]$$

where the Boltzmann face weights are given alternative graphical representations as a face or vertex

$$W\begin{pmatrix} & \gamma & \\ \delta & & \beta \\ & \alpha & \end{pmatrix} = \qquad [4]$$



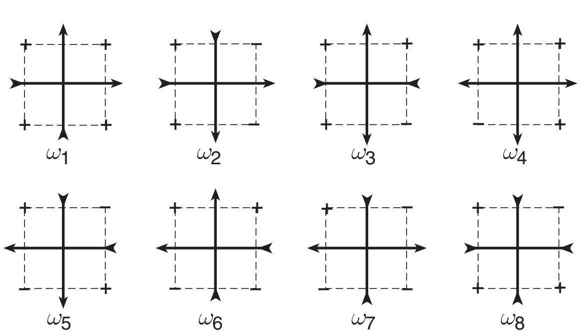
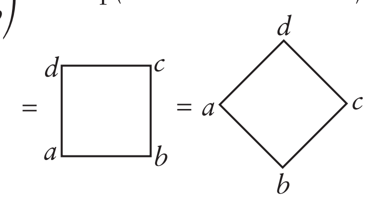




**Figure 1** The eight vertex configurations of the eight-vertex model showing one of the two corresponding configurations of the related Ising model. The model is solvable in the symmetric case, $\omega_1 = \omega_5, \omega_2 = \omega_6, \omega_3 = \omega_7, \omega_4 = \omega_8$, when the Boltzmann weights are equal in pairs under arrow reversal.

In the face representation, the arrow states are often called bond variables. Formally, the Hamiltonian is a sum over local energies $H = \sum_{\text{faces}} E(\alpha, \beta, \gamma, \delta)$, where $W(\alpha, \beta, \gamma, \delta) = \exp(-\beta E(\alpha, \beta, \gamma, \delta))$ but we use face weights since $E$ is infinite for excluded configurations. The general eight-vertex model includes many other ferroelectric models including the rectangular Ising model, Slater's model of potassium dihydrogen phosphate (KDP), the Rys F model of an antiferroelectric, the square ice model and the six-vertex model solved by Lieb. In the case of the six-vertex model, $\omega_4 = \omega_8 = 0$, so arrows are conserved with "two in" and "two out" at each vertex.

The eight-vertex model can be formulated as an Ising model with spins $a, b, c, d = \pm 1$ at the corners of the elementary faces and Boltzmann face weights

$$W\begin{pmatrix} d & c \\ a & b \end{pmatrix} = R \exp(Kac + Lbd + Mabcd) \qquad [5]$$




The four independent vertex weights are related to $R, K, L, M$ by

$$\begin{aligned} \omega_1 &= \omega_5 = Re^{K+L+M} \\ \omega_2 &= \omega_6 = Re^{-K-L+M} \\ \omega_3 &= \omega_7 = Re^{K-L-M} \\ \omega_4 &= \omega_8 = Re^{-K+L-M} \end{aligned} \qquad [6]$$

This is not the usual rectangular Ising model since it involves four-spin interactions in addition to two-spin interactions. The spins and arrows are related by

$$\alpha = ab, \quad \beta = bc, \quad \gamma = cd, \quad \delta = da \qquad [7]$$

This mapping is one-to-two, since we can arbitrarily fix one spin somewhere on the lattice. It follows that $Z_{\text{Ising}} = 2Z_{\text{vertex}}$. The eight-vertex model obviously includes the six-vertex ($\omega_4 = \omega_8$) and the rectangular Ising models ($M = 0$). Although it is not at all obvious, the three-spin Ising model is also included as a special case ($K = M, L = 0$).

Notice that the eight-vertex face weights are invariant under spin reversal of the spins on either diagonal. This $\mathbb{Z}_2 \times \mathbb{Z}_2$ symmetry, which the eight-vertex model shares with the Ashkin–Teller model, is peculiar because it allows the model to exhibit continuously varying critical exponents. Because of symmetries and duality, it is sufficient to consider the regime $\omega_1 > \omega_2 + \omega_3 + \omega_4$ with $\omega_2, \omega_3, \omega_4 > 0$. In terms of spins, this corresponds to the

ferromagnetically ordered phase; in terms of vertices or arrows, this corresponds to the ferroelectric phase. The eight-vertex model is critical on the four surfaces

$$\begin{aligned} \omega_1 &= \omega_2 + \omega_3 + \omega_4, \qquad & \omega_2 &= \omega_1 + \omega_3 + \omega_4 \\ \omega_3 &= \omega_1 + \omega_2 + \omega_4, \qquad & \omega_4 &= \omega_1 + \omega_2 + \omega_3 \end{aligned} \quad [8]$$

A convenient parameter to measure the departure from criticality $t = (T - T_c)/T_c$ is

$$\begin{aligned} t = -\frac{1}{16\omega_1\omega_2\omega_3\omega_4}[(\omega_1 - \omega_2 - \omega_3 - \omega_4) \\ \times (\omega_1 - \omega_2 + \omega_3 + \omega_4) \\ \times (\omega_1 + \omega_2 - \omega_3 + \omega_4) \\ \times (\omega_1 + \omega_2 + \omega_3 - \omega_4)] \end{aligned} \quad [9]$$

Because of the unusual four-spin interaction, it is difficult to realize the eight-vertex model experimentally in the laboratory.

### Hard-Hexagon Model

The hard-hexagon model is a two-dimensional lattice model of a gas of hard nonoverlapping particles. The particles are placed on the sites of a triangular lattice with nearest-neighbor exclusion so that no two particles are together or adjacent. Effectively, the triangular lattice is partially covered with nonoverlapping hard tiles of hexagonal shape. Let us draw the triangular lattice as a square lattice with one set of diagonals as in Figure 2. The partition function for the hard-hexagon model is

$$Z_N = \sum_{n=0}^{N} z^n g(n, N) \quad [10]$$

where $z > 0$ is the activity and $g(n, N)$ is the number of ways of placing $n$ particles on the $N$ sites such that no two particles are together or adjacent. To each lattice site $j$, assign a spin or occupation

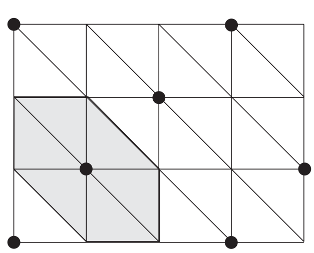


**Figure 2** The triangular lattice drawn as a square lattice with one set of diagonals. The close-packed arrangement of particles (solid circles) fills one of the three independent sublattices. One of the nonoverlapping hard hexagons is shown shaded. At low activities, the hard hexagons are sparsely scattered on the lattice with no preferential occupation of a particular sublattice.

number $\sigma_j$; if the site is empty, $\sigma_j = 0$; if the site is full, $\sigma_j = 1$. The partition function can then be written in terms of spins as

$$Z_N = \sum_{\text{spins } \sigma} \prod_{\langle ij \rangle} z^{(\sigma_i + \sigma_j)/6}(1 - \sigma_i\sigma_j) \quad [11]$$

where the product is over all bonds $\langle ij \rangle$ of the triangular lattice and the sum is over all configurations of the $N$ spins or occupation numbers $\sigma_j = 0, 1$. The exponent of $z$ arises because the activity is shared out between the six bonds incident at each site. The remaining term, $(1 - \sigma_i\sigma_j) = 0, 1$, ensures that neighboring sites are not occupied simultaneously by excluding such terms from the sum.

The activity $z$ gives the *a priori* probability of finding a particle at a given site and can be written as $z = e^{-\beta\mu}$, where $\mu$ is the chemical potential. The density of particles increases monotonically as the activity increases but only a third of the total lattice sites can be occupied. At low activities, there are only a few particles scattered randomly so the $S_3$ sublattice symmetry of the triangular lattice is preserved. However, at higher activities approaching the close-packing limit, there is a sudden change and one of the three sublattices is preferentially occupied so the $S_3$ sublattice symmetry is spontaneously broken. This dramatic change signals an order–disorder phase transition at some critical value $z_c$ of the activity. The system is disordered below the critical activity but is ordered above it. The fundamental problem is to obtain the statistical properties of this model such as the bulk free energy and the sublattice densities

$$\begin{aligned} \rho_k &= \langle \sigma_k \rangle \\ &= \{\text{fraction of spins sitting on} \\ &\qquad \text{sublattice } k = 1, 2, 3\} \end{aligned} \quad [12]$$

in the thermodynamic limit $N \to \infty$. The mean density is

$$\rho = (\rho_1 + \rho_2 + \rho_3)/3 \leq 1/3 \quad [13]$$

Assuming that sublattice $k = 1$ is preferentially occupied, an order parameter is defined by

$$R = \rho_1 - \rho_2 \quad [14]$$

The order parameter vanishes in the disordered regime but is nonzero in an ordered regime. Notice that the symmetry between sublattices $k = 2$ and 3 is not broken.

Unlike the eight-vertex model, the hard-hexagon model can be realized by a physical system in the laboratory, namely helium adsorbed on a graphite surface. The graphite substrate is composed of hexagonal cells formed by six carbon atoms with

an interatom distance of 2.46 Å. Energetically, the adsorbed helium atoms prefer to sit in the potential well at the center of the hexagonal cells. The diameter of the helium atom, however, is 2.56 Å, which precludes the simultaneous occupation of neighboring cells by excluded volume effects. Some beautiful experiments carried out by Bretz indicate that this system undergoes a phase transition. Indeed, Bretz took precise measurements of the specific heat as the temperature or, equivalently, the activity $z$, is varied, and obtained a symmetric power-law divergence at the critical point

$$C \sim |z - z_c|^{\alpha}, \quad \alpha \approx 0.36 \qquad [15]$$

with critical exponent $\alpha$ close to 1/3. Of course, one does not actually see divergences experimentally. Rather, it is the presence of dramatic peaks in the specific heat that are the hallmarks of a second-order transition.

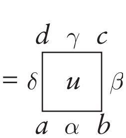
## Yang–Baxter Equations and Commuting Transfer Matrices

### Yang–Baxter Equations

The eight-vertex and hard-hexagon models were solved by Rodney Baxter at the beginning of the 1970s and 1980s, respectively. Although the two models are quite different in nature, they are quintessential of exactly solvable lattice models. The seminal work of Baxter gives a precise criterion to decide if a two-dimensional lattice model is exactly solvable: it is exactly solvable if its local face weights satisfy the celebrated Yang–Baxter equation. We present a general formulation of the Yang–Baxter equations and commuting transfer matrices and then show how Baxter implemented these for the eight-vertex and hard-hexagon models.

The first important step in the exact solution of a two-dimensional lattice model is the parametrization of the Boltzmann weights in terms of a distinguished variable $u$ called the spectral parameter. Typically, critical models involve trigonometric or hyperbolic functions and off-critical models involve elliptic functions of the spectral parameter. In terms of $u$, the local Boltzmann weights of a general two-dimensional lattice model take the form

$$W\left(\begin{matrix} d & \gamma & c \\ \delta & & \beta \\ a & \alpha & b \end{matrix}\middle|\, u\right) = \qquad [16]$$

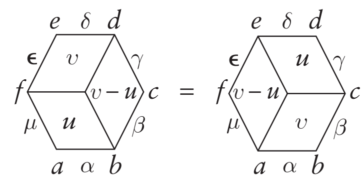



where the allowed values of the spins $a, b, c, \ldots$ and arrows (or bond variables) $\alpha, \beta, \gamma, \ldots$ may be restricted by certain constraints. The spins $a, b, c, d$ are absent for the eight-vertex model and the arrows $\alpha, \beta, \gamma, \delta$ are absent for the hard-hexagon model.

The general Yang–Baxter equations take the following algebraic and graphical forms:

$$\sum_{g,\eta,\xi,\zeta} W\left(\begin{matrix} f & \zeta & g \\ \mu & & \eta \\ a & \alpha & b \end{matrix}\middle|\, u\right) W\left(\begin{matrix} e & \delta & d \\ \epsilon & & \xi \\ f & \zeta & g \end{matrix}\middle|\, v\right) W\left(\begin{matrix} d & \gamma & c \\ \xi & & \beta \\ g & \eta & b \end{matrix}\middle|\, v-u\right)$$

$$= \sum_{g,\eta,\xi,\zeta} W\left(\begin{matrix} e & \eta & g \\ \epsilon & & \xi \\ f & \mu & a \end{matrix}\middle|\, v-u\right) W\left(\begin{matrix} g & \zeta & c \\ \xi & & \beta \\ a & \alpha & b \end{matrix}\middle|\, v\right) W\left(\begin{matrix} e & \delta & d \\ \eta & & \gamma \\ g & \zeta & c \end{matrix}\middle|\, u\right)$$



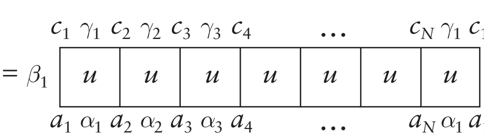

[17]

Graphically, this equation can be interpreted as saying that the diamond-shaped face with spectral parameter $v - u$ can be pushed through from the right to the left with the effect of interchanging the spectral parameters $u$ and $v$ in the remaining two faces.

### Commuting Transfer Matrices

A square lattice is built up row-by-row using the row transfer matrix $T(u)$ with matrix elements

$$\langle a, \alpha | \, T(u) \, | c, \gamma \rangle = \sum_{\beta_1, \beta_2, \ldots, \beta_N = \pm 1} \prod_{j=1}^{N} W\left(\begin{matrix} c_j & \gamma_j & c_{j+1} \\ \beta_j & & \beta_{j+1} \\ a_j & \alpha_j & a_{j+1} \end{matrix}\middle|\, u\right) \qquad [18]$$

= 

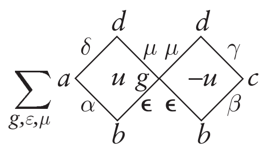

[19]

Here there are $N$ columns, and periodic boundary conditions are applied so that $a_{N+1} = a_1, \beta_{N+1} = \beta_1$, and so on. The significance of the Yang–Baxter equations is that they imply a one-parameter family of commuting transfer matrices

$$\boldsymbol{T}(u)\boldsymbol{T}(v) = \boldsymbol{T}(v)\boldsymbol{T}(u) \qquad [20]$$

Pictorially, the product on the left is represented by two rows, one above the other, the lower row with spectral parameter $u$ and the upper row with spectral parameter $v$. The matrix product implies

that the spins and arrows on the intervening row are summed out. Inserting a diamond-shaped face with spectral parameter $v-u$ and then using the local Yang–Baxter equation to progressively push it from right to left around the period interchanges all of the spectral parameters $u$ with the spectral parameter $v$. At the end, the diamond-shaped face is removed again. This heuristic argument was made rigorous by Baxter, who showed quite generally, and for the eight-vertex and hard-hexagon models in particular, that the diamond faces are in fact invertible:




$$= \rho(u)\ \delta(a, c)\ \delta(\alpha, \beta)\ \delta(\gamma, \delta) \qquad [21]$$

independent of $b, d$ where the scalar function $\rho(u)$ is model dependent. This equation is called the inversion relation.

Invariably, the existence of commuting transfer matrices leads to functional equations satisfied by the transfer matrices. Typically, the transfer matrices can be simultaneously diagonalized and so the functional equations can be solved for the eigenvalues of the transfer matrices. Mathematically, this is where Yang–Baxter techniques derive their power. For example, building up the lattice row-by-row, we see that the partition function of an $M \times N$ lattice is

$$Z_{MN} = \operatorname{tr} \boldsymbol{T}(u)^M = \sum_n T_n(u)^M \qquad [22]$$

where $T_n(u)$ are the eigenvalues of $\boldsymbol{T}(u)$. Typically, by the Perron–Frobenius theorem, the largest eigenvalue $T_0(u)$ is real, positive, and nondegenerate:

$$T_0(u) > |T_1(u)| \geq |T_2(u)| \geq \cdots \qquad [23]$$

Consequently,

$$\begin{aligned} -\beta\psi &= \lim_{N\to\infty} \lim_{M\to\infty} \frac{1}{MN} \log \sum_n T_n(u)^M \\ &= \lim_{N\to\infty} \frac{1}{N} \log T_0(u) \end{aligned} \qquad [24]$$

Thus the calculation of the bulk free energy is reduced to the problem of finding the largest eigenvalue of the transfer matrix.

### Parametrization of the Eight-Vertex Model

Using the spin formulation of the eight-vertex model, Baxter showed that two transfer matrices $\boldsymbol{T}(K, L, M), \boldsymbol{T}(K', L', M')$ commute whenever

$$\Delta(K, L, M) = \Delta(K', L', M') \qquad [25]$$

where

$$\begin{aligned} \Delta(K, L, M) = \sinh 2K \sinh 2L \\ + \tanh 2M \cosh 2K \cosh 2L \end{aligned} \qquad [26]$$

If $M$ and $\Delta$ are regarded as fixed, this is seen to be a symmetric biquadratic relation between $e^{2K}$ and $e^{2L}$ and is naturally parametrized in terms of elliptic functions. Unfortunately, many different notations and conventions for these elliptic functions appear in the literature which can be confusing to the uninitiated. Let

$$s = \frac{\vartheta_1(u)}{\vartheta_1(\lambda)}, \quad s_\pm = \frac{\vartheta_1(\lambda \pm u)}{\vartheta_1(\lambda)}, \quad \eta = \frac{\vartheta_1^2(\lambda)}{\vartheta_4^2(\lambda)} \qquad [27]$$

$$c = \frac{\vartheta_4(u)}{\vartheta_4(\lambda)}, \quad c_\pm = \frac{\vartheta_4(\lambda \pm u)}{\vartheta_4(\lambda)}, \quad \mu = \frac{\vartheta_4(0)}{\vartheta_4(\lambda)} \qquad [28]$$

where $\vartheta_1(u) = \vartheta_1(u, q)$ and $\vartheta_4(u) = \vartheta_4(u, q)$ are standard elliptic theta functions of nome $q$. Then the vertex weights can be parametrized as

$$\begin{aligned} \omega_1 &= R\mu^{-1} c c_-, & \omega_2 &= R\eta\mu^{-1} s s_- \\ \omega_3 &= R\mu^{-1} c s_-, & \omega_4 &= R\mu^{-1} c_- s \end{aligned} \qquad [29]$$

In the ferromagnetic regime $u$, $\lambda$, and $\tau$ are all pure imaginary with $0 < q < 1$ and $0 < \operatorname{Im} u < \operatorname{Im} \lambda < (\pi/2) \operatorname{Im} \tau$. The critical line occurs in the limit $q \to 1$. In this sense, we are using a low-temperature elliptic parametrization. Another elliptic parametrization, which is useful to study the critical limit, is obtained by transforming to the conjugate nome $q'$. If $q = e^{-\pi\epsilon}$ then the conjugate nome is defined by $q' = e^{-\pi/\epsilon}$ so that $q' \to 0$ as $q \to 1$.

We regard the crossing parameter $\lambda$ as constant, $u$ as a variable, and write the transfer matrix as $\boldsymbol{T}(u)$. It follows from this parametrization that $M$ and $\Delta$ are constants, independent of $u$. Furthermore, any two transfer matrices $\boldsymbol{T}(u)$ and $\boldsymbol{T}(v)$ commute and hence $\boldsymbol{T}(u)$ is a one-parameter family of commuting transfer matrices. For interest, we point out that the integrable $XYZ$ quantum spin chain belongs to this family. Specifically, the logarithmic derivative of the eight-vertex transfer matrix yields

$$\frac{d}{du} [\log \boldsymbol{T}(u)]_{u=0} = H_{XYZ} \qquad [30]$$

where

$$H_{XYZ} = -\frac{1}{2} \sum_{j=1}^{N} \left( J_x \sigma_j^x \sigma_{j+1}^x + J_y \sigma_j^y \sigma_{j+1}^y + J_z \sigma_j^z \sigma_{j+1}^z \right) \qquad [31]$$

and $\sigma_j^x, \sigma_j^y, \sigma_j^z$ are the usual Pauli spin matrices.

### Parametrization of the Hard-Hexagon Model

Actually, Baxter did not solve the hard-hexagon model directly. Instead, he solved a generalized hard-hexagon model, which is a model of hard squares with interactions along the diagonals of the elementary squares as shown in Figure 3. This in turn corresponds to the $A_4$ case of the more general solvable $A_L$ restricted solid-on-solid (RSOS) models of Andrews, Baxter, and Forrester.

The face weights of the generalized hard-hexagon model are

$$W\begin{pmatrix} d & c \\ a & b \end{pmatrix} = mz^{(a+b+c+d)/4}t^{-a+b-c+d}(1-ab) \times (1-bc)(1-cd)(1-da) \times \exp(Lac+Mbd) \qquad [32]$$

Here the activity $z$ has been shared out between the four faces adjacent to a site, $m$ is a trivial normalization constant, and $t$ is a gauge parameter that cancels out of the partition function and transfer matrix. The anisotropy between $L$ and $M$ introduces an additional parameter which will play the role of the spectral parameter $u$. In fact, using the Yang–Baxter equation, Baxter showed that this model is exactly solvable on the manifold

$$z = (1-e^{-L})(1-e^{-M})/(e^{L+M}-e^{L}-e^{M}) \qquad [33]$$

Specifically, two transfer matrices $\boldsymbol{T}(z,L,M)$ and $\boldsymbol{T}(z',L',M')$ commute whenever

$$\Delta(z,L,M) = \Delta(z',L',M') \qquad \Delta(z,L,M) = z^{-1/2}(1-ze^{L+M}) \qquad [34]$$

The hard-hexagon model is recovered in the limit $L=0$, $M=-\infty$ which forbids simultaneous occupation of sites joined by one set of diagonals. In this special limit, the activity $z$ is unconstrained. It is curious to note that the pure hard-square model with $L=M=0$ is not solvable.

Eliminating $z$ between the above relations gives a symmetric biquadratic relation between $e^L$ and $e^M$, which is naturally parametrized in terms of elliptic functions. Choosing $m$ and $t$ appropriately, the Boltzmann weights are

$$W\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \frac{\theta(2\lambda+u)}{\theta(2\lambda)}$$
$$W\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = W\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \frac{\theta(u)}{[\theta(\lambda)\theta(2\lambda)]^{1/2}}$$
$$W\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = W\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \frac{\theta(\lambda-u)}{\theta(\lambda)} \qquad [35]$$
$$W\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{\theta(2\lambda-u)}{\theta(2\lambda)}$$
$$W\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \frac{\theta(\lambda+u)}{\theta(\lambda)}$$

Here the crossing parameter is $\lambda=\pi/5$, $-\lambda<u<2\lambda$, and

$$\theta(u) = \theta(u,q^2) = \sin u\prod_{n=1}^{\infty}(1-q^{2n}e^{2iu}) \times (1-q^{2n}e^{-2iu})(1-q^{2n}) \qquad [36]$$

is a nonstandard elliptic theta function of nome $q^2$. Despite the deceiving notation, the nome $q^2$ lies in the range $-1<q^2<1$ and is determined by the relation

$$\Delta^2 = \left[\frac{\theta(\lambda)}{\theta(2\lambda)}\right]^5 = z(1-ze^{L+M})^2 \qquad [37]$$

Regarding $q^2$ as fixed and $u$ as a variable, it follows that $\boldsymbol{T}(u)$ is a one-parameter family of commuting transfer matrices.

The regimes relevant to the hard-hexagon model are:

$$\begin{aligned} &\text{Regime I (disordered)}: && -1<q^2<0,\ -\lambda<u<0 \\ &\text{Regime II (triangular ordered)}: && 0<q^2<1,\ -\lambda<u<0 \end{aligned} \qquad [38]$$

The borderline case $q^2=0$ corresponds to a line of critical points. The original hard-hexagon model is obtained in the limit $u\to-\lambda=-\pi/5$, so it follows that the critical point occurs at

$$z_c = \left(\frac{1+\sqrt{5}}{2}\right)^5 = \frac{1}{2}(11+5\sqrt{5}) \qquad [39]$$

Away from criticality the activity is related to the nome $q^2$ by

$$z = z_c\prod_{n=1}^{\infty}\left[\frac{1-2q^{2n}\cos(4\pi/5)+q^{4n}}{1-2q^{2n}\cos(2\pi/5)+q^{4n}}\right]^5 \qquad [40]$$

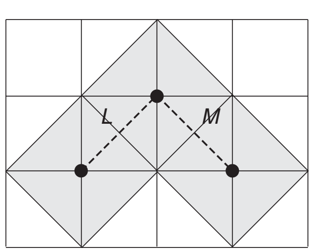



**Figure 3** Interacting hard squares showing the diagonal interactions $L$ and $M$. The hard-hexagon model corresponds to the limit $L=0, M=-\infty$..

## Functional Equations

### Baxter's $T-Q$ Relation

In a *tour de force* Baxter showed that the transfer matrix of the eight-vertex model satisfies the functional equation

$$T(u)Q(u) = \phi(u)Q(u-\lambda) + \phi(u-\lambda)Q(u+\lambda) \quad [41]$$

where $\phi(u) = (cs)^N = [\vartheta_1(u)\vartheta_4(u)/\vartheta_1(\lambda)\vartheta_4(\lambda)]^N$ and $Q(u)$ is an auxiliary family of mutually commuting transfer matrices satisfying $[Q(u), Q(v)] = [Q(u), T(v)] = 0$. In principle, these equations, which are intimately related to the Bethe ansatz, can be solved to obtain all the eigenvalues of the transfer matrix. Without entering into the intricacies of solving these equations, we summarize the results for the partition function per site $\kappa$, correlation length $\xi$, and interfacial tension $\sigma$. As we have seen, the largest eigenvalue of the transfer matrix yields $\kappa$. The interfacial tension $\sigma$ and correlation length $\xi$ were obtained, respectively, by Baxter and by Johnson, Krinsky, and McCoy by integrating over (continuous) bands of eigenvalues. In the ferromagnetic regime, their results are

$$\log(\kappa/\omega_1) = \sum_{n=1}^{\infty} \frac{x^{-n}(x^{2n}-q^n)^2(x^n+x^{-n}-z^n-z^{-n})}{n(1-q^{2n})(1+x^{2n})} \quad [42]$$

$$\xi^{-1} = -\tfrac{1}{2}\log k(x^2), \qquad \sigma = k_B T/\xi \quad [43]$$

where $x = e^{\pi i \lambda/2}$, $z = x^{-1}e^{\pi i u}$, and $k$ is the elliptic modulus of nome $x^2$:

$$k(x^2) = 4x\prod_{n=1}^{\infty}\left(\frac{1+x^{4n}}{1+x^{4n-2}}\right)^4 \quad [44]$$

Detailed analysis shows that near $T_c$ the free energy $\psi$ in general behaves as

$$\psi \sim \cot(\pi^2/2\bar{\mu})t^{\pi/\bar{\mu}} \sim t^{2-\alpha}, \quad t \to 0 \quad [45]$$

where $t = (T - T_c)/T_c$,

$$\tan(\bar{\mu}/2) = (\omega_3\omega_4/\omega_1\omega_2)^{1/2} = e^{-2M} \quad [46]$$

and $\alpha = 2 - \pi/\bar{\mu}$ with $0 < \bar{\mu} < \pi$. Exceptional cases occur, however, if $\pi/\bar{\mu}$ is an integer. This occurs, for example, in the case of the rectangular Ising model ($M = 0, \bar{\mu} = \pi/2$), which exhibits a logarithmic singularity in the specific heat ($\alpha = 0_{\log}$). Similarly, using $\log k(x^2) \sim (-t)^{\pi/2\bar{\mu}}$, the other associated critical exponents are

$$\xi^{-1} \sim (-t)^{\nu}, \quad \sigma \sim (-t)^{\mu}, \quad \nu = \mu = \pi/2\bar{\mu} \quad [47]$$

Notice that, due to the special symmetries of the eight-vertex model, these critical exponents vary continuously as the four-spin interaction is varied. This violates the universality hypothesis, which asserts that the exponents should only depend on the dimensionality and symmetries and not on the details of the interactions. Suzuki has suggested that it is more natural to use the inverse correlation length $\xi^{-1}$, rather than the temperature difference $T - T_c$, to measure the departure from criticality with the effect that it is the renormalized critical exponents

$$\hat{\alpha} = (2-\alpha)/\nu, \quad \hat{\beta} = \beta/\nu, \quad \hat{\mu} = \mu/\nu \quad [48]$$

that are independent of the details of the interactions.

### Hard-Hexagon Functional Equation

Baxter and Pearce showed that the normalized row transfer matrix of the generalized hard-hexagon model,

$$t(u) = \left[\frac{\theta(u+2\lambda)\theta(\lambda)}{\theta(u+\lambda)\theta(u-2\lambda)}\right]^N T(u) \quad [49]$$

satisfies the simple functional equation

$$t(u)t(u+\lambda) = I + t(u-2\lambda) \quad [50]$$

where $\lambda = \pi/5$. Since $T(u)$ is a commuting family of matrices, this equation can be solved for the eigenvalues $T(u)$ to obtain the partition function per site $\kappa$, correlation length $\xi$, and interfacial tension $\sigma$. Let $p = |q^2|$, $s = |q^2|^{5/6}$, then the results are summarized as

$$\kappa = \begin{cases} \kappa_c \displaystyle\prod_{n=1}^{\infty} \frac{[1-2q^{2n}\cos(4\pi/5)+q^{4n}]^2(1-q^{2n})^2(1-p^{5n})(1-p^{10(2n-1)/3})^3}{[1-2q^{2n}\cos(2\pi/5)+q^{4n}]^3(1-p^{5n/3})^3(1-p^{10(2n-1)})}, & z \le z_c \\ \kappa_c \displaystyle\prod_{n=1}^{\infty} \frac{[1-2q^{2n}\cos(4\pi/5)+q^{4n}]^2(1-q^{2n})^2(1-p^{5n})}{[1-2q^{2n}\cos(2\pi/5)+q^{4n}]^3(1-p^{5n/3})^3}, & z > z_c \end{cases} \quad [51]$$

$$\kappa_c = \left[\frac{27(25+11\sqrt{5})}{250}\right]^{1/2} \quad [52]$$

$$e^{-\xi^{-1}} = \begin{cases} \prod_{n=1}^{\infty} \frac{1-\sqrt{3}s^{2n-1}+s^{4n-2}}{1+\sqrt{3}s^{2n-1}+s^{4n-2}}, & z \le z_c \\ \min_{0<\alpha<\pi} \left[\prod_{n=1}^{\infty} \frac{1-2s^{2n-1}\cos\left(\frac{2\pi}{3}-\alpha\right)+s^{4n-2}}{1-2s^{2n-1}\cos\left(\frac{2\pi}{3}+\alpha\right)+s^{4n-2}}\right]^2, & z > z_c \end{cases} \quad [53]$$

$$\sigma = \begin{cases} 0, & z \le z_c \\ \frac{1}{2}k_B T/\xi, & z > z_c \end{cases} \quad [54]$$

It follows that $\kappa(z)$, $\xi(z)$, and $\sigma(z)$ are analytic functions of $z$, except at the critical point $z = z_c$.

The associated critical exponents

$$\psi \sim (z-z_c)^{2-\alpha}, \qquad \xi \sim (z-z_c)^{-\nu}$$
$$\sigma \sim (z-z_c)^{-\mu}, \quad \alpha = 1/3, \quad \nu = \mu = 5/6 \quad [55]$$

agree with experiments on helium adsorbed on graphite.

## Corner Transfer Matrices

The one-point functions and order parameters of the eight-vertex and hard-hexagon models were obtained by Baxter by using corner transfer matrices (CTMs). The idea is to build up the square lattice quadrant-by-quadrant as shown in Figure 4. The partition function and one-point function are then

$$Z = \operatorname{tr} \boldsymbol{ABCD}, \qquad \langle \sigma_1 \rangle = \frac{\operatorname{tr} \boldsymbol{SABCD}}{\operatorname{tr} \boldsymbol{ABCD}} \quad [56]$$

where $\boldsymbol{S}$ is the diagonal matrix with entries $S_{\sigma,\sigma} = \sigma_0$ and the entries $A_{\sigma,\sigma'}$ are labeled by half-rows of spins $\sigma = (\sigma_0, \sigma_1, \sigma_2, \dots)$ and $\sigma' = (\sigma_0, \sigma'_1, \sigma'_2, \dots)$.

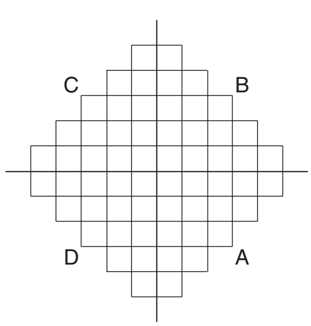



**Figure 4** The square lattice divided into four quadrants corresponding to the CTMs ***A***, ***B***, ***C***, ***D***. The spin at the center is $\sigma_0$.. The spins on the boundaries are fixed by the boundary conditions.

The CTMs have some remarkable properties. If the Boltzmann weights are invariant under reflections about the diagonals, as is the case for the eight-vertex model, Baxter argued that, in the limit of a large lattice,

$$\boldsymbol{A}(u) = \boldsymbol{C}(u) = \boldsymbol{B}(\lambda - u) = \boldsymbol{D}(\lambda - u) \quad [57]$$

where $\boldsymbol{A}(u)$ is a commuting family of matrices. Since these are block matrices in the center spin $\sigma_0$, they also commute with $\boldsymbol{S}$. Moreover, Baxter showed that the eigenvalues of $\boldsymbol{A}(u)$ are exponentials of the form

$$A(u)_\sigma = m_\sigma \exp(uE_\sigma) \quad [58]$$

where the constants $m_\sigma$ and $E_\sigma$ can be evaluated in the low-temperature limit. It follows that

$$\langle \sigma_0 \rangle = \frac{\sum_\sigma \sigma_0 m_\sigma^4 e^{2\lambda E_\sigma}}{\sum_\sigma m_\sigma^4 e^{2\lambda E_\sigma}} \quad [59]$$

When the Boltzmann weights do not exhibit symmetry about the diagonals, which is the case for hard hexagons, the above arguments need to be modified.

### One-Point Functions of the Eight-Vertex Model

For the eight-vertex model, Baxter showed that

$$m_\sigma = 1, \quad E_\sigma = \frac{1}{2}\pi i \sum_{j=1}^{N} jH(\sigma_{j-1}, \sigma_j, \sigma_{j+1})$$
$$H(\sigma_{j-1}, \sigma_j, \sigma_{j+1}) = 1 - \sigma_{j-1}\sigma_{j+1} \quad [60]$$

subject to the boundary condition $\sigma_N = \sigma_{N+1} = +1$. Introducing a new set of spins

$$\mu_j = \sigma_{j-1}\sigma_{j+1}, \quad j = 1, 2, \dots, N \quad [61]$$

we have $\sigma_0 = \mu_1\mu_3\mu_5 \dots$. Setting $s = (xz)^{1/2} = e^{\pi i u/2}$, $t = (x/z)^{1/2} = e^{\pi i(\lambda - u)/2}$ and taking the limit of large $N$, the diagonalized matrices are direct products of $2 \times 2$ matrices:

$$\boldsymbol{S} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \otimes \cdots \quad [62]$$

$$\boldsymbol{A}(u) = \boldsymbol{C}(u) = \begin{pmatrix} 1 & 0 \\ 0 & s \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & s^2 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & s^3 \end{pmatrix} \otimes \cdots \quad [63]$$

$$\boldsymbol{B}(u) = \boldsymbol{D}(u) = \begin{pmatrix} 1 & 0 \\ 0 & t \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & t^2 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & t^3 \end{pmatrix} \otimes \cdots \quad [64]$$

It follows that the magnetization is

$$\langle \sigma_0 \rangle = \prod_{n=1}^{\infty} \frac{1 - x^{4n-2}}{1 + x^{4n+2}} = (k')^{1/4} = (1 - k^2)^{1/8} \quad [65]$$

where $k' = k'(x^2)$ is the conjugate elliptic modulus of nome $x^2$ and the associated critical exponent is

$$\langle \sigma_0 \rangle \sim (-t)^\beta, \quad \beta = \pi/16\bar{\mu} \qquad [66]$$

The polarization of the eight-vertex model is

$$\langle \alpha \rangle = \langle \sigma_0 \sigma_1 \rangle = \prod_{n=1}^{\infty} \left( \frac{1 - x^{2n}}{1 + x^{2n}} \frac{1 + q^n}{1 - q^n} \right)^2 \qquad [67]$$

This cannot be obtained by a direct application of CTMs but was conjectured by Baxter and Kelland and subsequently derived by Jimbo, Miwa, and Nakayashiki using difference equations.

### One-Point Functions of the Hard-Hexagon Model

For hard hexagons, the working is more complicated because one must keep track of the sublattice of the central spin $\sigma_0$, but fascinating connections emerge with the Rogers–Ramanujan functions:

$$\begin{aligned} G(x) &= \prod_{n=1}^{\infty} \frac{1}{(1 - x^{5n-4})(1 - x^{5n-1})} \\ H(x) &= \prod_{n=1}^{\infty} \frac{1}{(1 - x^{5n-3})(1 - x^{5n-2})} \end{aligned} \qquad [68]$$

For hard hexagons, Baxter showed that

$$\rho_k = \frac{\operatorname{tr} \boldsymbol{S}(\boldsymbol{A}_k \boldsymbol{B}_k)^2}{\operatorname{tr} (\boldsymbol{A}_k \boldsymbol{B}_k)^2} = \frac{\sum_\sigma \sigma_0 r_0^{2\sigma_0} w_0^{2E_\sigma}}{\sum_\sigma r_0^{2\sigma_0} w_0^{2E_\sigma}} \qquad [69]$$

where $k = 1, 2, 3$ labels the sublattice of the triangular lattice. Here the spin configurations $\sigma = (\sigma_0, \sigma_1, \sigma_3, \ldots)$ with $\sigma_j = 0, 1$ are subject to the constraint $\sigma_j \sigma_{j+1} = 0$ for all $j$. If $|q^2| = \mathrm{e}^{-\epsilon}$ and $g(x) = H(x)/G(x)$ then

$$\begin{aligned} &x = -\mathrm{e}^{-\pi^2/5\epsilon}, \; r_0^2 = -x/g(x), \; w_0 = -x^3 &&\text{for } z \le z_c \\ &x = \mathrm{e}^{-4\pi^2/5\epsilon}, \; r_0^2 = x^{-1} g(x), \; w_0 = x^{-3/2} &&\text{for } z > z_c \end{aligned} \qquad [70]$$

and

$$E_\sigma = \begin{cases} \sum_{j=1}^{\infty} j(\sigma_j - s_j), & z \le z_c \\ \sum_{j=1}^{\infty} j(\sigma_j - \sigma_{j-1}\sigma_{j+1} \\ \qquad - s_j + s_{j-1}s_{j+1}), & z > z_c \end{cases} \qquad [71]$$

For large $N$, $\sigma_j \to s_j$, where the ground-state values $s_j$ determined by the boundary conditions are

$$z \le z_c : \quad s_j = 0 \qquad [72]$$

$$z > z_c : \quad s_{3j+k} = 1, \; s_{3j+k\pm 1} = 0, \; k = 1, 2, 3 \qquad [73]$$

After applying some Rogers–Ramanujan identities and introducing the elliptic functions

$$\begin{aligned} Q(x) &= \prod_{n=1}^{\infty} (1 - x^n) \\ P(x) &= \frac{Q(x)}{Q(x^2)} = \prod_{n=1}^{\infty} (1 - x^{2n-1}) \end{aligned} \qquad [74]$$

the expressions for the sublattice densities simplify in the limit of large $N$ giving

$$\rho = \rho_1 = \rho_2 = \rho_3 = -\frac{x G(x) H(x^6) P(x^3)}{P(x^2)}, \quad z \le z_c \qquad [75]$$

in the disordered fluid phase and

$$\begin{aligned} \rho_1 &= \frac{H(x) Q(x) [G(x) Q(x) + x^2 H(x^9) Q(x^9)]}{Q(x^3)^2} \\ \rho_2 = \rho_3 &= \frac{x^2 H(x) H(x^9) Q(x) Q(x^9)}{Q(x^3)^2} \end{aligned} \qquad [76]$$

$$\begin{aligned} R = \rho_1 - \rho_2 &= \frac{Q(x) Q(x^5)}{Q(x^3)^2} \\ &= \prod_{n=1}^{\infty} \frac{(1 - x^n)(1 - x^{5n})}{(1 - x^{3n})^2}, \qquad z > z_c \end{aligned} \qquad [77]$$

in the triangular ordered phase. In principle, the dependence on $x$ can be eliminated by observing that

$$z = \begin{cases} -x[H(x)/G(x)]^5, & z \le z_c \\ x^{-1}[G(x)/H(x)]^5, & z > z_c \end{cases} \qquad [78]$$

In practice, this is quite nontrivial. Although it is far from obvious, because $x \to 1$ is a subtle limit, the critical exponent associated with the order parameter $R$ is

$$R \sim (z - z_c)^\beta \sim (q^2)^\beta, \quad \beta = 1/9 \qquad [79]$$

## Summary

Baxter's exact solutions of the eight-vertex and hard-hexagon models have been reviewed. These prototypical examples clearly illustrate the mathematical power and elegance of commuting transfer matrices and Yang–Baxter techniques. The results for the principal thermodynamic quantities, including free energies, correlation lengths, interfacial tensions, and one-point functions, have been summarized. For convenience in comparison, the associated critical exponents are collected in Table 1. All these exponents confirm the hyperscaling relation $2 - \alpha = d\nu$ for lattice dimensionality $d = 2$.

More recently, Yang–Baxter techniques have been applied to solve an infinite variety of lattice models in two dimensions. Commuting transfer methods have

**Table 1** Comparison of the exactly calculated critical exponents of the rectangular Ising, eight-vertex and hard-hexagon models. The rectangular Ising model corresponds to the special case $\bar{\mu}=\pi/2$ of the eight-vertex model. The eight-vertex exponents vary continuously with $0<\bar{\mu}<\pi$.. The critical exponents of the hard-hexagon model, with its $S_3$ symmetry, lie in the universality class of the three-state Potts model.

| *Model* | $\alpha$ | $\beta$ | $\nu$ | $\mu$ |
|---|---|---|---|---|
| Rectangular Ising | $0_{\log}$ | 1/8 | 1 | 1 |
| Eight vertex | $2-\pi/\bar{\mu}$ | $\pi/16\bar{\mu}$ | $\pi/2\bar{\mu}$ | $\pi/2\bar{\mu}$ |
| Hard hexagons | 1/3 | 1/9 | 5/6 | 5/6 |

also been adapted to study integrable boundaries and associated boundary critical behavior. Lastly, it should be mentioned that, in the continuum scaling limit, there are deep connections with conformal field theory and integrable quantum field theory. On the one hand, the lattice can often provide a convenient way to regularize the infinities that occur in these continuous field theories. On the other hand, the field theories can predict and explain the universal properties of lattice models such as critical exponents.

*See also:* Bethe Ansatz; Boundary Conformal Field Theory; Hopf Algebras and *q*-Deformation Quantum Groups; Integrability and Quantum Field Theory; *q*-Special Functions; Quantum Spin Systems; Two-Dimensional Ising Model; Yang–Baxter Equations.

## Further Reading

Andrews GE, Baxter RJ and Forrester PJ (1984) Eight-Vertex SOS model and generalized Rogers–Ramanujan-type identities. *Journal of Statistical Physics* 35: 193.

Baxter RJ (1971) 8 Vertex model in lattice statistics. *Physics Review Letters* 26: 832–833.

Baxter RJ (1972) Partition function of 8 vertex lattice model. *Annals of Physics* 70: 193–228.

Baxter RJ (1972) One-dimensional anisotropic Heisenberg chain. *Annals of Physics* 70: 323–337.

Baxter RJ (1973) 8-vertex model in lattice statistics and one-dimensional anisotropic Heisenberg chain: 1. Some fundamental eigenvectors; 2. Equivalence to a generalized ice-type lattice model; 3. Eigenvectors of transfer matrix and Hamiltonian. *Annals of Physics* 76: 1–24, 25–47, 48–71.

Baxter RJ (1980) Hard hexagons – exact solution. *Journal of Physics* A13: L61.

Baxter RJ (1981) Rogers–Ramanujan identities in the hard hexagon model. *Journal of Statistical Physics* 26: 427–452.

Baxter RJ (1982) *Exactly Solved Models in Statistical Mechanics*. London: Academic Press.

Baxter RJ (1982) The inversion relation method for some two-dimensional exactly solved models in lattice statistics. *Journal of Statistical Physics* 28: 1.

Baxter RJ (2001) Completeness of the Bethe ansatz for the six and eight-vertex models. *Journal of Statistical Physics* 108: 1–48.

Baxter RJ (2004) The six and eight-vertex models revisited. *Journal of Statistical Physics* 116: 43–66.

Baxter RJ and Kelland SB (1974) Spontaneous polarization of 8-vertex model. *Journal of Physics* C7: L403–L406.

Baxter RJ and Pearce PA (1982) Hard hexagons – interfacial tension and correlation length. *Journal of Physics* A15: 897–910.

Baxter RJ and Pearce PA (1983) Hard squares with diagonal attractions. *Journal of Physics* A16: 2239.

Bazhanov VV and Reshetikhin NY (1989) Critical RSOS models and conformal field theory. *International Journal of Modern Physics* A4: 115.

Bretz M (1977) Ordered helium films on highly uniform graphite – finite-size effects, critical parameters, and 3-state Potts model. *Physical Review Letters* 38: 501–505.

Fabricius K and McCoy BM (2003) New developments in the eight vertex model. *Journal of Statistical Physics* 111: 323–337.

Fabricius K and McCoy BM (2004) Functional equations and fusion matrices for the eight-vertex model. *Publications of the Research Institute for Mathematical Sciences* 40: 905–932.

Forrester PJ and Baxter RJ (1985) Further exact solutions of the eight-vertex SOS model and generalizations of the Rogers–Ramanujan identities. *Journal of Statistical Physics* 38: 435–472.

Jimbo M (1990) *Yang–Baxter Equation in Integrable Systems*. Advanced Series in Mathematical Physics, vol. 10. Singapore: World Scientific.

Jimbo M and Miwa T (1995) *Algebraic Analysis of Solvable Lattice Models*. Regional Conference Series in Mathematics, No. 85. Providence: American Mathematical Society.

Johnson JD, Krinsky S and McCoy BM (1973) Vertical arrow correlation length in 8-vertex model and low-lying excitations of XYZ Hamiltonian. *Physical Review A* 8: 2526–2547.

Klümper A and Pearce PA (1991) Analytic calculation of scaling dimensions: tricritical hard squares and critical hard hexagons. *Journal of Statistical Physics* 64: 13.

Klümper A and Pearce PA (1992) Conformal weights of RSOS lattice models and their fusion hierarchies. *Physica A* 183: 304.

Lieb EH (1967) Residual entropy of square ice. *Physical Review* 162: 162.

Lieb EH (1967) Exact solution of F model of an antiferroelectric. *Physical Review Letters* 18: 1046.

Lieb EH (1967) Exact solution of 2-dimensional Slater KDP model of a ferroelectric. *Physics Review Letters* 19: 108.

Lieb EH and Wu FY (1972) Two-dimensional ferroelectric models. In: Domb C and Green MS (eds.) *Phase Transitions and Critical Phenomena*, vol. 1, pp. 321–490. London: Academic Press.

McCoy BM (1999) The Baxter revolution. *The Physicist* 36(6): 210–214.

Pearce PA and Baxter RJ (1984) Deviations from critical density in the generalized hard hexagon model. *Journal of Physics* A17: 2095–2108.

# Einstein Equations: Exact Solutions

**Jiří Bičák**, Charles University, Prague, Czech Republic and Albert Einstein Institute, Potsdam, Germany



## Introduction

Even in a linear theory like Maxwell's electrodynamics, in which sufficiently general solutions of the field equations can be obtained, one needs a good sample, a useful kit, of explicit exact fields like the homogeneous field, the Coulomb monopole field, the dipole, and other simple solutions, in order to gain a physical intuition and understanding of the theory. In Einstein's general relativity, with its nonlinear field equations, the discoveries and analyses of various specific explicit solutions revealed most of the unforeseen features of the theory. Studies of special solutions stimulated questions relevant to more general situations, and even after the formulation of a conjecture about a general situation, newly discovered solutions can play a significant role in verifying or modifying the conjecture. The cosmic censorship conjecture assuming that "singularities forming in a realistic gravitational collapse are hidden inside horizons" is a good illustration.

Albert Einstein presented the final version of his gravitational field equations (or the Einstein's equations, EEs) to the Prussian Academy in Berlin on 18 November 1915:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu} \qquad [1]$$

Here, the spacetime metric tensor $g_{\mu\nu}(x^\rho), \mu, \nu, \rho, \ldots = 0, 1, 2, 3$, determines the invariant line element $g = g_{\mu\nu}\, dx^\mu\, dx^\nu$, and acts also as a dynamical variable describing the gravitational field; the Ricci tensor $R_{\mu\nu} = g^{\rho\sigma} R_{\rho\mu\sigma\nu}$, where $g^{\mu\rho} g_{\rho\nu} = \delta^\mu_\nu$, is formed from the Riemann curvature tensor $R_{\rho\mu\sigma\nu}$; both depend nonlinearly on $g_{\alpha\beta}$ and $\partial_\mu g_{\alpha\beta}$, and linearly on $\partial_\mu \partial_\nu g_{\alpha\beta}$; the scalar curvature $R = g^{\mu\nu} R_{\mu\nu}$. $T_{\mu\nu}(x^\rho)$ is the energy–momentum tensor of matter ("sources"); and Newton's gravitational constant $G$ and the velocity of light $c$ are fundamental constants. If not stated otherwise, we use the geometrized units in which $G = c = 1$, and the same conventions as in Misner *et al.* (1973) and Wald (1984). For example, in the case of perfect fluid with density $\rho$, pressure $p$, and 4-velocity $U^\mu$, the energy–momentum tensor reads $T_{\mu\nu} = (\rho + p)\; U_\mu U_\nu + p g_{\mu\nu}$. To obtain a (local) solution of [1] in coordinate patch $\{x^\rho\}$ means to find "physically plausible" (i.e., complying with one of the positive-energy conditions) functions $\rho(x^\rho), p(x^\rho), U_\mu(x^\rho)$, and metric $g_{\mu\nu}(x^\rho)$ satisfying [1]. In vacuum $T_{\mu\nu} = 0$ and [1] implies $R_{\mu\nu} = 0$.

In 1917, Einstein generalized [1] by adding a cosmological term $\Lambda g_{\mu\nu}$ ($\Lambda = \text{const.}$):

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu} \qquad [2]$$

A homogeneous and isotropic static solution of [2] (with metric [8], $k = +1, a = \text{const.}$), in which the "repulsive effect" of $\Lambda > 0$ compensates the gravitational attraction of incoherent dust ("uniformly distributed galaxies") – the Einstein static universe – marked the birth of modern cosmology. Although it is unstable and lost its observational relevance after the discovery of the expansion of the universe in the late 1920s, in 2004 a "fine-tuned" cosmological scenario was suggested according to which our universe starts asymptotically from an initial Einstein static state and later enters an inflationary era, followed by a standard expansion epoch (*see* Cosmology: Mathematical Aspects). There are many other examples of "old" solutions which turned out to act as asymptotic states of more general classes of models.

## Invariant Characterization and Classification of the Solutions

### Algebraic Classification

The Riemann tensor can be decomposed as

$$R_{\alpha\beta\gamma\delta} = C_{\alpha\beta\gamma\delta} + E_{\alpha\beta\gamma\delta} + G_{\alpha\beta\gamma\delta} \qquad [3]$$

where $E$ and $G$ are constructed from $R_{\alpha\beta}$, $R$, and $g_{\alpha\beta}$ (see, e.g., Stephani *et al.* (2003)); the Weyl conformal tensor $C_{\alpha\beta\gamma\delta}$ can be considered as the "characteristic of the pure gravitational field" since, at a given point, it cannot be determined in terms of the matter energy–momentum tensor $T_{\alpha\beta}$ (as $E$ and $G$ can using EEs). Algebraic classification is based on a classification of the Weyl tensor. This is best formulated using two-component spinors $\alpha_A (A = 1, 2)$, in terms of which any Weyl spinor $\Psi_{ABCD}$ determining $C_{\alpha\beta\gamma\delta}$ can be factorized:

$$\Psi_{ABCD} = \alpha_{(A} \beta_B \gamma_C \delta_{D)} \qquad [4]$$

brackets denote symmetrization; each of the spinors determines a principal null direction, say, $k^\alpha = \alpha^A \bar{\alpha}^{A'}$ (*see* Spinors and Spin Coefficients). The Petrov–Penrose classification is based on coincidences among these directions. A solution is of type *I* (general case), *II*, *III*, and *N* ("null") if all null directions are different, or two, three, and all four coincide, respectively. It is of type *D* ("degenerate")

if there are two double null directions. The equivalent tensor equations are simplest for type $N$:

$$C_{\alpha\beta\gamma\delta}k^{\delta} = 0, \qquad C_{\alpha\beta\gamma\delta}C^{\alpha\beta\gamma\delta} = 0, \qquad C_{\alpha\beta\gamma\delta}C^{*\alpha\beta\gamma\delta} = 0 \qquad [5]$$

where $C^{*}_{\alpha\beta\gamma\delta} = (1/2)\epsilon_{\alpha\beta\rho\sigma}C^{\rho\sigma}{}_{\gamma\delta}$, $\epsilon$ is the Levi-Civita pseudotensor.

### Classification According to Symmetries

Most of the available solutions have some exact continuous symmetries which preserve the metric. The corresponding group of motions is characterized by the number and properties of its Killing vectors $\xi^{\alpha}$ satisfying the Killing equation $(£_{\xi}g)_{\alpha\beta} = \xi_{\alpha;\beta} + \xi_{\beta;\alpha} = 0$ ($£$ is the Lie derivative) and by the nature (spacelike, timelike, or null) of the group orbits. For example, axisymmetric, stationary fields possess two commuting Killing vectors, of which one is timelike. Orbits of the axial Killing vector are closed spacelike curves of finite length, which vanishes at the axis of symmetry. In cylindrical symmetry, there exist two spacelike commuting Killing vectors. In both cases, the vectors generate a two-dimensional abelian group. The two-dimensional group orbits are timelike in the stationary case and spacelike in the cylindrical symmetry.

If a timelike $\xi^{\alpha}$ is hypersurface orthogonal, $\xi_{\alpha} = \lambda\Phi_{,\alpha}$ for some scalar functions $\lambda, \Phi$, the spacetime is "static." In coordinates with $\xi = \partial_t$, the metric is

$$g = -e^{2U}dt^2 + e^{-2U}\gamma_{ik}dx^i dx^k \qquad [6]$$

where $U, \gamma_{ik}$ do not depend on $t$. In vacuum, $U$ satisfies the potential equation $U^{:a}_{:a} = 0$, the covariant derivatives (denoted by :) are with respect to the three-dimensional metric $\gamma_{ik}$. A classical result of Lichnerowicz states that if the vacuum metric is smooth everywhere and $U \to 0$ at infinity, the spacetime is flat (for refinements, see Anderson (2000)).

In cosmology, we are interested in groups whose regions of transitivity (points can be carried into one another by symmetry operations) are three-dimensional spacelike hypersurfaces (homogeneous but anisotropic models of the universe). The three-dimensional simply transitive groups $G_3$ were classified by Bianchi in 1897 according to the possible distinct sets of structure constants but their importance in cosmology was discovered only in the 1950s. There are nine types: Bianchi I to Bianchi IX models. The line element of the Bianchi universes can be expressed in the form

$$g = -dt^2 + g_{ab}(t)\omega^a\omega^b \qquad [7]$$

where the time-independent 1-forms $\omega^a = E^a_{\alpha}dx^{\alpha}$ satisfy the relations $d\omega^a = -(1/2)C^a_{bc}\omega^b \wedge \omega^c$, $d$ is the exterior derivative and $C^a_{bc}$ are the structure constants (*see* Cosmology: Mathematical Aspects for more details).

The standard Friedmann–Lemaître–Robertson–Walker (FLRW) models admit in addition an isotropy group SO(3) at each point. They can be represented by the metric

$$g = -dt^2 + [a(t)]^2\left(\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\varphi^2)\right) \qquad [8]$$

in which $a(t)$, the "expansion factor," is determined by matter via EEs, the curvature index $k = -1, 0, +1$, the three-dimensional spaces $t$=const. have a constant curvature $K = k/a^2; r \in [0,1]$ for closed ($k = +1$) universe, $r \in [0, \infty)$ in open ($k = 0, -1$) universes (for another description (*see* Cosmology: Mathematical Aspects).

There are four-dimensional spacetimes of constant curvature solving EEs [2] with $T_{\mu\nu} = 0$: the Minkowski, de Sitter, and anti-de Sitter spacetimes. They admit the same number [10] of independent Killing vectors, but interpretations of the corresponding symmetries differ for each spacetime.

If $\xi^{\alpha}$ satisfies $£_{\xi}g_{\alpha\beta} = 2\Phi g_{\alpha\beta}$, $\Phi$ = const., it is called a homothetic (Killing) vector. Solutions with proper homothetic motions, $\Phi \neq 0$, are "self-similar." They cannot in general be asymptotically flat or spatially compact but can represent asymptotic states of more general solutions. In Stephani *et al.* (2003), a summary of solutions with proper homotheties is given; their role in cosmology is analyzed by Wainwright and Ellis (eds.) (1997); for mathematical aspects of symmetries in general relativity, see Hall (2004).

There are other schemes for invariant classification of exact solutions (reviewed in Stephani *et al.* (2003)): the algebraic classification of the Ricci tensor and energy–momentum tensor of matter; the existence and properties of preferred vector fields and corresponding congruences; local isometric embeddings into flat pseudo-Euclidean spaces, etc.

## Minkowski (M), de Sitter (dS), and Anti-de Sitter (AdS) Spacetimes

These metrics of constant (zero, positive, negative) curvature are the simplest solutions of [2] with $T_{\mu\nu} = 0$ and $\Lambda = 0, \Lambda > 0, \Lambda < 0$, respectively. The standard topology of M is $R^4$. The dS has the topology $R^1 \times S^3$ and is best represented as a four-dimensional hyperboloid $-v^2 + w^2 + x^2 + y^2 + z^2 = (3/\Lambda)$ in a five-dimensional flat space with metric $g = -dv^2 + dw^2 + dx^2 + dy^2 + dz^2$. The AdS has the topology $S^1 \times R^3$; it is a four-dimensional hyperboloid $-v^2 - w^2 + x^2 + y^2 + z^2 = -(3/\Lambda)$, $\Lambda < 0$, in flat five-dimensional space

with signature (− , − , + , +, +). By unwrapping the circle $S^1$ and considering the universal covering space, one gets rid of closed timelike lines.

These spacetimes are all conformally flat and can be conformally mapped into portions of the Einstein universe (*see* Asymptotic Structure and Conformal Infinity). However, their conformal structure is globally different. In M, one can go to infinity along timelike/null/spacelike geodesics and reach five qualitatively different sets of points: future/past timelike infinity $i^{\pm}$, future/past null infinity $\mathcal{I}^{\pm}$, and spacelike infinity $i^0$. In dS, there are only past and future conformal infinities $\mathcal{I}^-, \mathcal{I}^+$, both being spacelike (on the Einstein cylinder, the dS spacetime is a "horizontal strip" with $\mathcal{I}^+/\mathcal{I}^-$ as the "upper/lower circle"). The conformal infinity in AdS is timelike.

As a consequence of spacelike $\mathcal{I}^{\pm}$ in dS, there exist both particle (cosmological) and event horizons for geodesic observers (Hawking and Ellis 1973). dS plays a (doubly) fundamental role in the present-day cosmology: it is an approximate model for inflationary paradigm near the big bang and it is also the asymptotic state (at $t \to \infty$) of cosmological models with a positive cosmological constant. Since recent observations indicate that $\Lambda > 0$, it appears to describe the future state of our universe. AdS has come recently to the fore due to the "holographic" conjecture (*see* AdS/CFT Correspondence).

Christodoulou and Klainermann, and Friedrich proved that M, dS, and AdS are stable with respect to general, nonlinear (though "weak") vacuum perturbations – result not known for any other solution of EEs (*see* Stability of Minkowski Space).

## Schwarzschild and Reissner–Nordström Metrics

These are spherically symmetric spacetimes – the $SO_3$ rotation group acts on them as an isometry group with spacelike, two-dimensional orbits. The metric can be brought into the form

$$g = -e^{2\nu} dt^2 + e^{2\lambda} dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2) \qquad [9]$$

$\nu(t, r), \lambda(t, r)$ must be determined from EEs. In vacuum, we are led uniquely to the Schwarzschild metric

$$g = -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2) \qquad [10]$$

where $M = \text{const.}$ has to be interpreted as mass, as test particle orbits show. The spacetime is static at $r > 2M$, that is, outside the Schwarzschild radius at $r = 2M$, and asymptotically $(r \to \infty)$ flat.

Metric [10] describes the exterior gravitational field of an arbitrary (static, oscillating, collapsing, or expanding) spherically symmetric body (spherically symmetric gravitational waves do not exist). It is the most influential solution of EEs. The essential tests of general relativity – perihelion advance of Mercury, deflection of both optical and radio waves by the Sun, and signal retardation – are based on [10] or rather on its expansion in $M/r$. Space missions have been proposed that could lead to measurements of "post-post-Newtonian" effects (*see* General Relativity: Experimental Tests, and Misner *et al.* (1973)). The full Schwarzschild metric is of importance in astrophysical processes involving compact stars and black holes.

Metric [10] describes the spacetime outside a spherical body collapsing through $r = 2M$ into a spherical black hole. In **Figure 1**, the formation of an event horizon and trapped surfaces is indicated in ingoing Eddington–Finkelstein coordinates $(v, r, \theta, \varphi)$ where $v = t + r + 2M \log(r/2M - 1)$ so that $(v, \theta, \varphi) = \text{const.}$ are ingoing radial null geodesics. The interior of the star is described by another metric (e.g., the Oppenheimer–Snyder collapsing dust solution – see below). The Kruskal extension of the Schwarzschild solution, its compactification, the concept of the bifurcate Killing horizon, etc., are analyzed

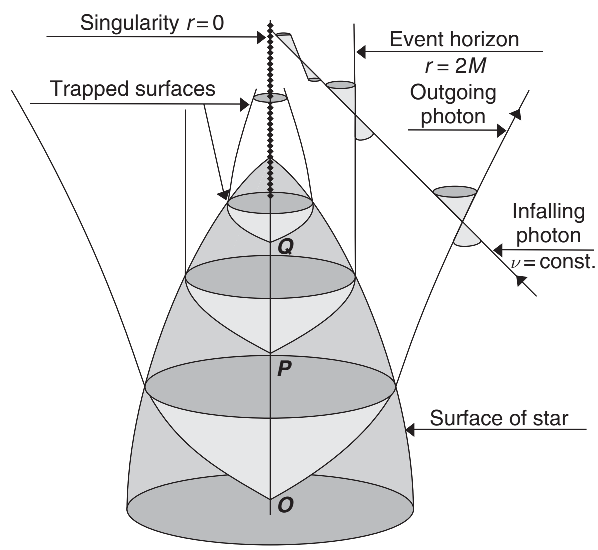



**Figure 1** Gravitational collapse of a spherical star (the interior of the star is shaded). The light cones of three events, *O*, *P*, *Q*, at the center of the star, and of three events outside the star are illustrated. The event horizon, the trapped surfaces, and the singularity formed during the collapse are also shown. Although the singularity appears to lie along the direction of time, from the character of the light cone outside the star but inside the event horizon we can see that it has a spacelike character. Reproduced from Bičák J (2000) Selected solutions of Einstein's field equations: their role in general relativity and astrophysics. In: Schmidt BG (ed.) *Einstein's Field Equations and their physical Implications*, Lecture Notes in Physics, vol. 540, pp. 1–126. Heildelberg: Springer, with permission from Springer-Verlag.

in Stationary Black Holes and in Misner *et al.* (1973), Hawking and Ellis (1973), and Bičák (2000).

The Reissner–Nordström solution describes the exterior gravitational and electromagnetic fields of a spherical body with mass $M$ and charge $Q$. The energy-momentum tensor on the right-hand side of EE [2] is that of the electromagnetic field produced by the charge; the field satisfies the curved-space Maxwell equations. The metric reads

$$g = -\left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right)dt^2 + \left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right)^{-1} dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2) \qquad [11]$$

The analytic extension of the electrovacuum metric [11] is qualitatively different from the Kruskal extension of the Schwarzschild metric. In the case $Q^2 > M^2$ there is a "naked singularity" (visible from $r \to \infty$) at $r = 0$ where curvature invariants diverge. If $Q^2 < M^2$, the metric describes a (generic) static charged black hole with two event horizons at $r = r_\pm = M \pm (M^2 - Q^2)^{1/2}$. The Killing vector $\partial/\partial t$ is null at the horizons, timelike at $r > r_+$ and $r < r_-$, but spacelike between the horizons. The character of the extended spacetime is best seen in the compactified form, **Figure 2**, in which world-lines of radial light rays are 45° lines. Again, two infinities (right and left, in regions I and III) arise (as in the Kruskal–Schwarzschild diagram, *see* Stationary Black Holes), however, the maximally extended geometry consists of an infinite chain of asymptotically flat regions connected by "wormholes" between the singularities at $r = 0$. In contrast to the Schwarzschild singularity, the singularities are timelike – they do not block the way to the future. The inner horizon $r = r_-$ represents a Cauchy horizon for a typical initial hypersurface like $\Sigma$ (**Figure 2**): what is happening in regions $V$ is in general influenced not only by data on $\Sigma$ but also at the singularities. The Cauchy horizon is unstable (for references, see Bičák (2000) and recent work by Dafermos (2005)).

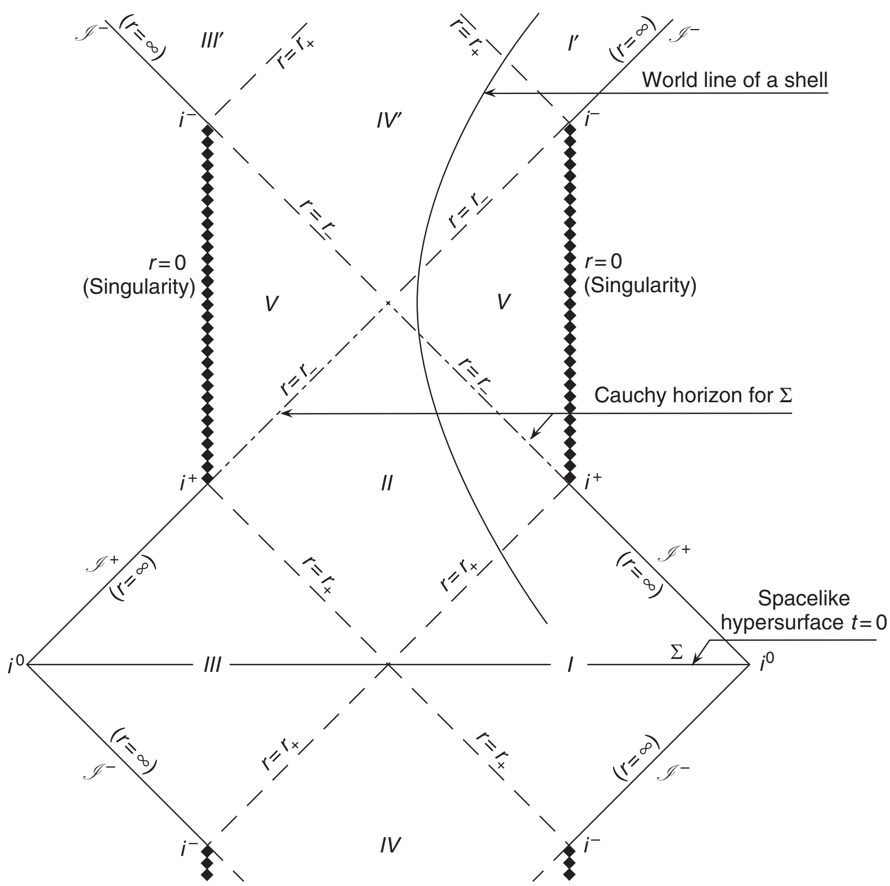



**Figure 2** The compactified Reissner–Nordström spacetime representing a non-extreme black hole consists of an infinite chain of asymptotic regions ("universes") connected by "wormholes" between timelike singularities. The world-line of a shell collapsing from "universe" *I* and re-emerging in "universe" *I′* is indicated. The inner horizon at $r = r_-$ is the Cauchy horizon for a spacelike hypersurface $\Sigma$. It is unstable and thus it will very likely prevent such a process. Reproduced from Bičák J (2000) Selected solutions of Einstein's field equations: their role in general relativity and astrophysics. In: Schmidt BG (ed.) *Einstein's Field Equations and their Physical Implications*, Lecture Notes in Physics, vol. 540, pp. 1–126. Heildelberg: Springer, with permission from Springer-Verlag.

For $M^2 = Q^2$ the two horizons coincide at $r_+ = r_- = M$. Metric [11] describes extreme Reissner–Nordström black holes. The horizon becomes degenerate and its surface gravity vanishes (*see* Stationary Black Holes). Extreme black holes play a significant role in string theory (Ortín 2004).

## Stationary Axisymmetric Solutions

Assume the existence of two commuting Killing vectors – timelike $\xi^\alpha$ and axial $\eta^\alpha$ ($\xi^\alpha \xi_\alpha < 0, \eta^\alpha \eta_\alpha > 0$), $\xi^\alpha$ normalized at (asymptotically flat) infinity, $\eta^\alpha$ at the rotation axis. They generate two-dimensional orbits of the group $G_2$. Assume there exist 2-spaces orthogonal to these orbits. This is true in vacuum and also in case of electromagnetic fields or perfect fluids whose 4-current or 4-velocity lies in the surfaces of transitivity of $G_2$ (e.g., toroidal magnetic fields are excluded). The metric can then be written in Weyl's coordinates $(t, \rho, \varphi, z)$

$$g = -e^{2U}(dt + A d\varphi)^2 + e^{-2U}[e^{2k}(d\rho^2 + dz^2) + \rho^2 d\varphi^2] \qquad [12]$$

$U$, $k$, and $A$ are functions of $\rho, z$.

The most celebrated vacuum solution of the form [12] is the Kerr metric for which $U$, $k$, $A$ are ratios of simple polynomials in spheroidal coordinates (simply related to $(\rho, z)$). The Kerr solution is characterized by mass $M$ and specific angular momentum $a$. For $a^2 > M^2$, it describes an asymptotically flat spacetime with a naked singularity. For $a^2 \leq M^2$, it represents a rotating black hole that has two horizons which coalesce into a degenerate horizon for $a^2 = M^2$ – an extreme Kerr black hole. The two horizons are located at $r_\pm = M \pm (M^2 - a^2)^{1/2}$ ($r$ being the Boyer–Lindquist coordinate (*see* Stationary Black Holes)). As with the Reissner–Nordström black hole, the singularity inside is timelike and the inner horizon is an (unstable) Cauchy horizon. The analytic extension of the Kerr metric resembles **Figure 2** (see Frolov and Novikov (1998), Hawking and Ellis (1973), Misner *et al.* (1973), Ortín (2004), Semerák *et al.* (2002), Stephani *et al.* (2003), and Wald (1984) for details).

Thanks to the black hole uniqueness theorems (*see* Stationary Black Holes), the Kerr metric is the unique solution describing all rotating black holes in vacuum. If the cosmic censorship conjecture holds, Kerr black holes represent the end states of gravitational collapse of astronomical objects with supercritical masses. According to prevalent views, they reside in the nuclei of most galaxies. Unlike with a spherical collapse, there are no exact solutions available which would represent the formation of a Kerr black hole. However, starting from metric [12] and identifying, for example, $z = b =$ const. and $z = -b$ (with the region $-b < z < b$ being cut off), one can construct thin material disks which are physically plausible and can be the sources of the Kerr metric even for $a^2 > M^2$ (see Bičák (2000) for details).

In a general case of metric [12], EEs in vacuum imply the "Ernst equation" for a complex function $f$ of $\rho$ and $z$:

$$(\Re f)\left[f_{,\rho\rho} + f_{,zz} + \frac{1}{\rho} f_{,\rho}\right] = f_{,\rho}^2 + f_{,z}^2 \qquad [13]$$

or, equivalently, $(\Re f)\triangle f = (\nabla f)^2$, where $f = e^{2U} + ib$, $U$ enters [12], and $b(\rho, z)$ is a "potential" for $A(\rho, z)$: $A_{,\rho} = \rho e^{-4U} b_{,z}$, $A_{,z} = -\rho e^{-4U} b_{,\rho}$; $k(\rho, z)$ in [12] can be determined from $U$ and $b$ by quadratures. Tomimatsu and Sato (TS) exploited symmetries of [13] to construct metrics generalizing the Kerr metric. Replacing $f$ by $\xi = (1 - f)/(1 + f)$, one finds that in case of the Kerr metric $\xi^{-1}$ is a linear function in the prolate spheroidal coordinates, whereas for TS solutions $\xi$ is a quotient of higher-order polynomials. A number of other solutions of eqn [13] were found but they are of lower significance than the Kerr solution (cf. Stephani *et al.* (2003), Chapter 20).

These solutions inspired "solution-generating methods" in general relativity. The Ernst equation can be regarded as the integrability condition of a system of linear differential equations. The problem of solving such a system can be reformulated as the Riemann–Hilbert problem in complex function theory (*see* Riemann–Hilbert Problem and Integrable Systems: Overview). We refer to Stephani *et al.* (2003) and Belinski and Verdaguer (2001) where these techniques using Bäcklund transformations, inverse-scattering method, etc., are also applied in the nonstationary context of two spacelike Killing vectors (waves, cosmology). In the stationary case, all asymptotically flat, stationary, axisymmetric vacuum solutions can, in principle, be generated. It is known how to generate fields with given values of multipole moments, though the required calculations are staggering. By solving the Riemann–Hilbert problem with appropriate boundary data, Neugebauer and Meinel constructed the exact solution representing a rigidly rotating thin disk of dust (cf. Stephani *et al.* (2003) and Bičák (2000)).

A subclass of metrics [12] is formed by static Weyl solutions with $A = b = 0$. Equation [13] then becomes the Laplace equation $\Delta U = 0$. The nonlinearity of EEs enters only the equations for $k$: $k_{,\rho} = \rho(U_{,\rho}^2 - U_{,z}^2)$, $k_{,z} = 2\rho U_{,\rho} U_{,z}$. The class contains some explicit solutions of interest: the "linear superposition" of collinear particles with string-like singularities between them which keep the system in static equilibrium; solutions representing external

fields of counter-rotating disks, for example, those which are "inspired" by galactic Newtonian potentials; disks around black holes and some other special solutions (Stephani *et al.* 2003, Bonnor 1992, Bičák 2000, Semerák *et al.* 2002).

There are solutions of the Einstein–Maxwell equations representing external fields of masses endowed with electric charges, magnetic dipole moments, etc. (Stephani *et al.* 2003). Best known is the Kerr–Newman metric characterized by parameters $M$, $a$, and charge $Q$. For $M^2 \geq a^2 + Q^2$ it describes a charged, rotating black hole. Owing to the rotation, the charged black hole produces also a magnetic field of a dipole type. All the black hole solutions can be generalized to include a nonvanishing $\Lambda$ (for various applications, see Semerák *et al.* 2002)). Other generalizations incorporate the so-called Newman–Unti–Tamburino (NUT) parameter (corresponding to a "gravomagnetic monopole") or an "external" magnetic/electric field or a parameter leading to "uniform" acceleration (see Stephani *et al.* (2003) and Bičák (2000)). Much interest has recently been paid to black hole (and other) solutions with various types of gauge fields and to multidimensional solutions. References Frolov and Novikov (1998) and Ortín (2004) are two examples of good reviews.

## Radiative Solutions

### Plane Waves and Their Collisions

The best-known class are "plane-fronted gravitational waves with parallel rays" (pp-waves) which are defined by the condition that the spacetime admits a covariantly constant null vector field $k^\alpha$: $k_{\alpha;\beta} = 0$. In suitable null coordinates $u$, $v$ such that $k_\alpha = u_{,\alpha}$, $k^\alpha = (\partial/\partial v)^\alpha$, and complex coordinate $\zeta$ which spans the wave 2-surfaces $u = \text{const.}$, $v = \text{const.}$ with Euclidean geometry, the metric reads

$$g = 2d\zeta d\bar{\zeta} - 2du dv - 2H(u, \zeta, \bar{\zeta}) du^2 \qquad [14]$$

$H(u, \zeta, \bar{\zeta})$ is a real function. The vacuum EEs imply $H_{,\zeta\bar{\zeta}} = 0$ so that $2H = f(u, \zeta) + \bar{f}(u, \bar{\zeta})$, $f$ is an arbitrary function of $u$, analytic in $\zeta$. The Weyl tensor satisfies eqns [5] – the field is of type $N$ as is the field of plane electromagnetic waves. In the null tetrad $\{k^\alpha, l^\alpha, m^\alpha(\text{complex})\}$ with $l^\alpha k_\alpha = -1, m^\alpha \bar{m}_\alpha = 1$, all other products vanishing, the only nonzero projection of the Weyl tensor, $\Psi = C_{\alpha\beta\gamma\delta} l^\alpha \bar{m}^\beta l^\gamma m^\delta = H_{,\bar{\zeta}\bar{\zeta}}$, describes the transverse component of a wave propagating in the $k^\alpha$ direction. Writing $\Psi = \mathcal{A}e^{i\Theta}$, the real $\mathcal{A} > 0$ is the amplitude of the wave, $\Theta$ describes polarization. Waves with $\Theta = \text{const.}$ are called linearly polarized. Considering their effect on test particles, one finds that plane waves are transverse.

The simplest waves are homogeneous in the sense that $\Psi$ is constant along the wave surfaces. One gets $f(u, \zeta) = (1/2)\mathcal{A}(u)e^{i\Theta(u)}\zeta^2$. Instructive are "sandwich waves," for example, waves with a "square profile": $\mathcal{A} = 0$ for $u < 0$ and $u > a^2$, $\mathcal{A} = a^{-2} = \text{const.}$ for $0 \leq u \leq a^2$. This example demonstrates, within exact theory, that the waves travel with the speed of light, produce relative accelerations of test particles, focus astigmatically generally propagating parallel rays, etc. The focusing effects have a remarkable consequence: there exists no global spacelike hypersurface on which initial data could be specified – plane wave spacetimes contain no global Cauchy hypersurface.

"Impulsive" plane waves can be generated by boosting a "particle" at rest to the velocity of light by an appropriate limiting procedure. The ultrarelativistic limit of, for example, the Schwarzschild metric (the so-called Aichelburg–Sexl solution) can be employed as a "limiting incoming state" in black hole encounters (cf. monograph by d'Eath (1996)). Plane-fronted waves have been used in quantum field theory. For a review of exact impulsive waves, see Semerák *et al.* (2002).

A collision of plane waves represents an exceptional situation of nonlinear wave interactions which can be analyzed exactly. **Figure 3** illustrates a typical case in which the collision produces a spacelike singularity. The initial-value problem with data given at $v = 0$ and $u = 0$ can be formulated in terms of the equivalent matrix Riemann–Hilbert problem (*see* Riemann–Hilbert Problem); it is related to the hyperbolic counterpart of the Ernst equation [13]. For reviews, see Griffiths (1991), Stephani *et al.* (2003), and Bičák (2000).

### Cylindrical Waves

Discovered by G Beck in 1925 and known today as the Einstein–Rosen waves (1937), these vacuum solutions helped to clarify a number of issues, such as energy loss due to the waves, asymptotic structure of radiative

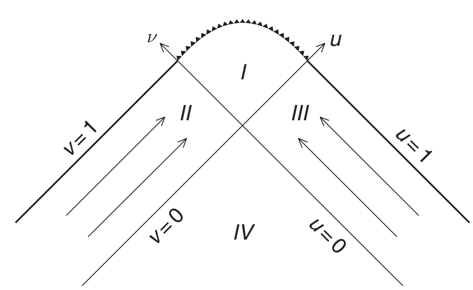



**Figure 3** A spacetime diagram indicating a collision of two plane-fronted gravitational waves which come from regions *II* and *III*, collide in region *I*, and produce a spacelike singularity. Region *IV* is flat. Reproduced from Bičák J (2000) Selected solutions of Einstein's field equations: their role in general relativity and astrophysics. In: Schmidt BG (ed.) *Einstein's Field Equations and their Physical Implications*, Lecture Notes in Physics, vol. 540, pp. 1–126. Heildelberg: Springer, with permission from Springer-Verlag.

spacetimes, dispersion of waves, quasilocal mass–energy, cosmic censorship conjecture, or quantum gravity in the context of midisuperspaces (see Bičák (2000) and Belinski and Verdaguer (2001)).

In the metric

$$g = e^{2(\gamma-\psi)}(-dt^2 + d\rho^2) + e^{2\psi}dz^2 + \rho^2 e^{-2\psi}d\varphi^2 \quad [15]$$

$\psi(t, \rho)$ satisfies the flat-space wave equation and $\gamma(\rho, t)$ is given in terms of $\psi$ by quadratures. Admitting a "cross term" $\sim \omega(t, \rho)\, dz\, d\phi$, one acquires a second degree of freedom (a second polarization) which makes all field equations nonlinear.

### Boost-Rotation Symmetric Spacetimes

These are the only explicit solutions available which are radiative and represent the fields of finite sources. Figure 4 shows two particles uniformly accelerated in opposite directions. In the space diagram (left), the "string" connecting the particles is the "cause" of the acceleration. In "Cartesian-type" coordinates and the $z$-axis chosen as the symmetry axis, the boost Killing vector has a flat-space form, $\zeta = z(\partial/\partial t) + t(\partial/\partial z)$, the same is true for the axial Killing vector. The metric contains two functions of variables $\rho^2 \equiv x^2 + y^2$ and $\beta^2 \equiv z^2 - t^2$. One satisfies the flat-space wave equation, the other is determined by quadratures.

The unique role of these solutions is exhibited by the theorem which states that in axially symmetric, locally asymptotically flat spacetimes, in the sense that a null infinity (*see* Asymptotic Structure and Conformal Infinity) exists but not necessarily globally, the only additional symmetry that does not exclude gravitational radiation is the boost symmetry. Various radiation characteristics can be expressed explicitly in these spacetimes. They have been used as tests in numerical relativity and approximation methods. The best-known example is the C-metric (representing accelerating black holes, in general charged and rotating, and admitting $\Lambda$), see Bonnor *et al.* (1994), Bičák (2000), Stephani *et al.* (2003), and Semerák *et al.* (2002).

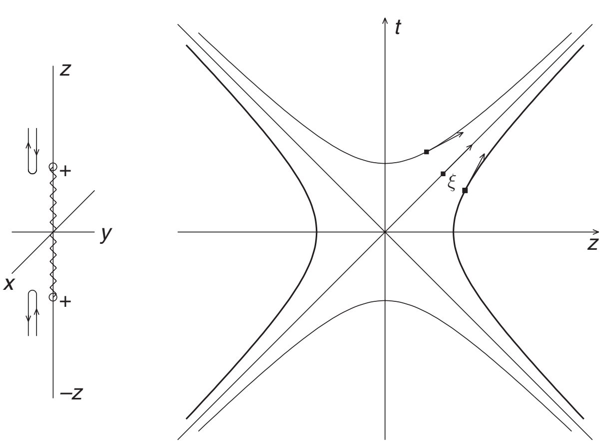



**Figure 4** Two particles uniformly accelerated in opposite directions. Orbits of the boost Killing vector (thinner hyperbolas) are spacelike in the region $t^2 > z^2$. Reproduced from Bičák J (2000) Selected solutions of Einstein's field equations: their role in general relativity and astrophysics. In: Schmidt BG (ed.) *Einstein's Field Equations and their Physical Implications*, Lecture Notes in Physics, vol. 540, pp. 1–126. Heildelberg: Springer, with permission from Springer-Verlag.

### Robinson–Trautman Solutions

These solutions are algebraically special but in general they do not possess any symmetry. They are governed by a function $P(u, \zeta, \bar{\zeta})$ ($u$ is the retarded time, $\zeta$ a complex spatial coordinate) which satisfies a fourth-order nonlinear parabolic differential equation. Studies by Chruściel and others have shown that RT solutions of Petrov type II exist globally for all positive "times" $u$ and converge asymptotically to a Schwarzschild metric, though the extension across the "Schwarzschild-like" horizon can only be made with a finite degree of smoothness. Generalization to the cases with $\Lambda > 0$ gives explicit models supporting the cosmic no-hair conjecture (an exponentially fast approach to the dS spacetime) under the presence of gravitational waves. See Bonnor *et al.* (1994), Bičák (2000), and Stephani *et al.* (2003).

## Material Sources

Finding physically sound material sources in an analytic form even for some simple vacuum metrics remains an open problem. Nevertheless, there are solutions representing regions of spacetimes filled with matter which are of considerable interest.

One of the simplest solutions, the spherically symmetric Schwarzschild interior solution with incompressible fluid as its source, represents "a star" of uniform density, $\rho = \rho_0 = \text{const.}$:

$$g = -\left[\frac{3}{2}\sqrt{1 - AR^2} - \frac{1}{2}\sqrt{1 - Ar^2}\right]^2 dt^2 + \frac{dr^2}{1 - Ar^2} + r^2(d\theta^2 + \sin^2\theta d\varphi^2) \quad [16]$$

$A = 8\pi\rho_0/3 = \text{const.}$, $R$ is the radius of the star.

The equation of hydrostatic equilibrium yields pressure inside the star:

$$8\pi p = 2A\frac{\sqrt{1 - Ar^2} - \sqrt{1 - AR^2}}{3\sqrt{1 - AR^2} - \sqrt{1 - Ar^2}} \quad [17]$$

Solution [16] can be matched at $r = R$, where $p = 0$, to the exterior vacuum Schwarzschild solution [10] if the Schwarzschild mass $M = (1/2)AR^3$. Although "incompressible fluid" implies an infinite speed of sound, the above solution provides an instructive

model of relativistic hydrostatics. A Newtonian star of uniform density can have an arbitrarily large radius $R = \sqrt{3p_c/2\pi\rho_0^2}$ and mass $M = (p_c/\rho_0^2)\ \sqrt{6p_c/\pi}$, $p_c$ is the central pressure. However, [17] implies that (1) $M$ and $R$ satisfy the inequality $2M/R \leq 8/9$, (2) equality is reached as $p_c$-becomes infinite and $R$ and $M$ attain their limiting values $R_{\lim} = (3\pi\rho_0)^{-1/2} = (9/4)M_{\lim}$. For a density typical in neutron stars, $\rho_0 = 10^{15}$ g cm$^{-3}$, we get $M_{\lim} \doteq 3.96 M_\odot$ ($M_\odot$ solar mass) – even this simple model shows that in Einstein's theory neutron stars can only be a few solar masses. In addition, one can prove that the "Buchdahl's inequality" $2M/R \leq 8/9$ is valid for an arbitrary equation of state $p = p(\rho)$. Only a limited mass can thus be contained within a given radius in general relativity. The gravitational redshift $z = (1 - 2M/R)^{-1/2} - 1$ from the surface of a static star cannot be higher than 2.

Many other explicit static perfect fluid solutions are known (we refer to Stephani *et al.* (2003) for a list), however, none of them can be considered as really "physical." Recently, the dynamical systems approach to relativistic spherically symmetric static perfect fluid models was developed by Uggla and others which gives qualititative characteristics of masses and radii.

The most significant nonstatic spacetime describing a bounded region of matter and its external field is undoubtedly the Oppenheimer–Snyder model of "gravitational collapse of a spherical star" of uniform density and zero pressure (a "ball of dust"). The model does not represent any new (local) solution: the interior of the star is described by a part of a dust-filled FLRW universe (cf. [8]), the external region by the Schwarzschild vacuum metric (cf. eqn [10], **Figure 1**).

Since Vaidya's discovery of a "radiating Schwarzschild metric," null dust ("pure radiation field") has been widely used as a simple matter source. Its energy–momentum tensor, $T_{\alpha\beta} = \varrho k_\alpha k_\beta$, where $k_\alpha k^\alpha = 0$, may be interpreted as an incoherent superposition of waves with random phases and polarizations moving in a single direction, or as "lightlike particles" (photons, neutrinos, gravitons) that move along $k^\alpha$. The "Vaidya metric" describing spherical implosion of null dust implies that in case of a "gentle" inflow of the dust, a naked singularity forms. This is relevant in the context of the cosmic censorship conjecture (cf., e.g., Joshi (1993)).

## Cosmological Models

There exist important generalizations of the standard FLRW models other than the above-mentioned Bianchi models, particularly those that maintain spherical symmetry but do not require homogeneity. The best known are the Lemaître–Tolman–Bondi models of inhomogeneous universes of pure dust, the density of which may vary (Krasiński 1997).

Other explicit cosmological models of principal interest involve, for example, the Gödel universe – a homogeneous, stationary spacetime with $\Lambda < 0$ and incoherent rotating matter in which there exist closed timelike curves through every point; the Kantowski–Sachs solutions – possessing homogeneous spacelike hypersurfaces but (in contrast to the Bianchi models) admitting no simply transitive $G_3$; and vacuum Gowdy models ("generalized Einstein–Rosen waves") admitting $G_2$ with compact 2-tori as its group orbits and representing cosmological models closed by gravitational waves. See Cosmology: Mathematical Aspects and references Stephani *et al.* (2003), Belinski and Verdaguer (2001), Bičák (2000), Hawking and Ellis (1973), Krasiński (1997) and Wainwright and Ellis (1997).

*See also:* AdS/CFT Correspondence; Asymptotic Structure and Conformal Infinity; Cosmology: Mathematical Aspects; Dirac Fields in Gravitation and Nonabelian Gauge Theory; Einstein Manifolds; Einstein's Equations with Matter; General Relativity: Experimental Tests; General Relativity: Overview; Hamiltonian Reduction of Einstein's Equations; Integrable Systems: Overview; Newtonian Limit of General Relativity; Pseudo-Riemannian Nilpotent Lie Groups; Reimann–Hilbert Problem; Spacetime Topology, Causal Structure and Singularities; Spinors and Spin Coefficients; Stability of Minkowski Space; Stationary Black Holes; Twistor Theory: Some Applications.

## Further Reading


Anderson MT (2000) On the structure of solutions to the static vacuum Einstein equations. *Annales Henri Poincaré* 1: 995–1042.

Belinski V and Verdaguer E (2001) *Gravitational Solitons.* Cambridge: Cambridge University Press.

Bičák J (2000) Selected solutions of Einstein's field equations: their role in general relativity and astrophysics. In: Schmidt BG (ed.) *Einstein's Field Equations and Their Physical Implications*, Lecture Notes in Physics, vol. 540, pp. 1–126, (see also gr-qc/0004016). Heidelberg: Springer.

Bonnor WB (1992) Physical interpretation of vacuum solutions of Einstein's equations. Part I. Time-independent solutions. *General Relativity and Gravitation* 24: 551–574.

Bonnor WB, Griffiths JB, and MacCallum MAH (1994) Physical interpretation of vacuum solutions of Einstein's equations. Part II. Time-dependent solutions. *General Relativity and Gravitation* 26: 687–729.

Dafermos M (2005) The interior of charged black holes and the problem of uniqueness in general relativity. *Communications on Pure and Applied Mathematics* LVIII: 445–504.

D'Eath PD (1996) *Black Holes: Gravitational Interactions.* Oxford: Clarendon Press.

Frolov VP and Novikov ID (1998) *Black Hole Physics.* Dordrecht: Kluwer Academic.

Griffiths JB (1991) *Colliding Plane Waves in General Relativity.* Oxford: Oxford University Press.

Hall GS (2004) *Symmetries and Curvature Structure in General Relativity*. Singapore: World Scientific.
Hawking SW and Ellis GFR (1973) *The Large Scale Structure of Space-Time*. Cambridge: Cambridge University Press.
Joshi PS (1993) *Global Aspects in Gravitation and Cosmology*. Oxford: Clarendon.
Krasiński A (1997) *Inhomogeneous Cosmological Models*. Cambridge: Cambridge University Press.
Misner C, Thorne KS, and Wheeler JA (1973) *Gravitation*. San Francisco: WH Freeman.
Ortín T (2004) *Gravity and Strings*. Cambridge: Cambridge University Press.
Semerák O, Podolský J, and Žofka M (eds.) (2002) *Gravitation: Following the Prague Inpiration*. Singapore: World Scientific.
Stephani H, Kramer D, MacCallum MAH, Hoenselaers C, and Herlt E (2003) *Exact Solutions of Einstein's Field Equations – Second Edition*. Cambridge: Cambridge University Press.
Wainwright J and Ellis GFR (eds.) (1997) *Dynamical Systems in Cosmology*. Cambridge: Cambridge University Press.
Wald RM (1984) *General Relativity*. Chicago: The University of Chicago Press.

# Einstein Equations: Initial Value Formulation

**J Isenberg**, University of Oregon, Eugene, OR, USA



## Introduction

Einstein's theory of gravity models a gravitating physical system $\mathcal{S}$ using a spacetime $(M^4, g, \psi)$ which satisfies the Einstein field equations

$$G_{\mu\nu}(g) = \kappa T_{\mu\nu}(g, \psi) \quad [1]$$

$$\mathcal{F}(g, \psi) = 0 \quad [2]$$

Here, $M^4$ is a four-dimensional spacetime manifold, $g$ is a Lorentz signature metric on $M$, $\psi$ represents the nongravitational ("matter") fields of interest, $G_{\mu\nu} := R_{\mu\nu} - (1/2)g_{\mu\nu}R$ is the Einstein curvature tensor, $\kappa$ is a constant, $T_{\mu\nu}$ is the stress–energy tensor for the field $\psi$, and $\mathcal{F} = 0$ represents the nongravitational field equations (e.g., $\nabla_\mu F^\mu_\nu = 0$ for the Einstein–Maxwell theory).

By far the most widely used way to obtain and to study spacetime solutions $(M^4, g, \psi)$ of equations [1]–[2] is via the initial-value (or Cauchy) formulation. The idea is as follows:

1. One chooses a set of initial data $\mathcal{D}$ which consists of geometric as well as matter information on a spacelike slice of $M^4$. This data must satisfy a system of constraint equations, which comprise a portion of the field equations [1]–[2], and are analogous to the Maxwell constraint equation $\nabla \cdot E = 0$.
2. One fixes a time and coordinate choice to be used in evolving the fields into the spacetime (e.g., maximal time slicing and zero shift). This choice should result in a fixed set of evolution equations for the data.
3. Using the evolution equations, one evolves the data into the future and the past. From the evolved data, one constructs the spacetime solution $(M^4, g, \psi)$.

Why is this procedure so popular? First, because we have known for over 50 years that at least for a short time, it works. That is, as shown by Choquet-Bruhat (Foures-Bruhat 1952), the Cauchy formulation is well posed. Second, because it fits with the way we like to model physical systems. That is, we first specify what the system is like now, and we then use the equations to determine the behavior of the system as it evolves into the future (or the past). Third, because the formulation is eminently amenable to numerical treatment. Indeed, virtually all numerical simulations of colliding black hole systems as well as of most other relativistic astrophysical systems are done using some version of the initial-value formulation. Finally, because the initial-value formulation casts the Einstein equations into a form which is readily accessible to many of the tools of geometric analysis. Questions such as cosmic censorship are turned into conjectures which can be analyzed and proved mathematically, and the proofs of both the positivity of mass and the Penrose mass inequality rely on an initial-value interpretation.

There are of course drawbacks to the Cauchy formulation. Foremost, Einstein's theory of general relativity is inherently a spacetime-covariant theory; why break spacetime apart into space plus time when covariance has played such a key role in the theory's success? As well, we have learned over and over again that null cones and null hypersurfaces play a major role in general relativity; the initial-value formulation is not especially good at handling them. These drawbacks show that there are analyses in general relativity for which the initial-value formulation may not be well suited. However, there is a preponderance of applications for which this formulation is an invaluable tool, as evidenced by its ubiquitous use.

A complete treatment of the initial-value formulation for Einstein's equations would include discussion of each of the following topics:

1. A statement and proof of well-posedness theorems, including a discussion of the regularity of the data needed for such results.
2. A space + time decomposition of the fields, and a formal derivation of the Einstein constraint equations and the Einstein evolution equations.

3. An outline of the Hamiltonian version of the initial-value formulation.
4. A listing of those choices of field variables and gauge choices for which the system is manifestly hyperbolic.
5. A description of the known methods for finding and parametrizing solutions of the Einstein constraint equations.
6. A comparison of the virtues and drawbacks of various choices of time foliation and coordinate threading.
7. A compendium of results concerning long-time behavior of solutions.
8. An account of the difficulties which arise in attempts to construct solutions numerically using the Cauchy formulation.
9. A recounting of cases in which the initial-value formulation has been used to model physically interesting systems.
10. A note regarding the extent to which the initial-value formulation (and the various aspects of it just enumerated) generalize to dimensions other than $3+1$ (three space and one time).
11. A determination of which nongravitational fields may be coupled to Einstein's theory in such a way that the resulting coupled theory admits an initial-value formulation.

We do not have the space here for such a complete treatment. So we choose to focus on those topics directly related to the Einstein constraint equations. Generalizing a bit to the Einstein–Maxwell theory (thereby including representative nongravitational fields), we first carry out the space plus time "$3+1$" decomposition of the gravitational and electromagnetic fields. Then, applying the Gauss–Codazzi–Mainardi equations to the spacetime curvature, we turn the spacetime-covariant Einstein–Maxwell equations into a set of constraint equations restricting the choice of initial data together with a set of evolution equations developing the data in time. Next, we discuss the most widely used approach for obtaining sets of initial data which satisfy the constraint equations: the conformal method. We include in this discussion an account of some of what is known about the extent to which the equations which are produced by the conformal method admit solutions in various situations (e.g., working on a closed manifold, or working with asymptotically Euclidean data). We then discuss alternate procedures which have been used to obtain and analyze solutions of the constraints, including the conformal thin sandwich approach, the quasispherical method, and various gluing procedures. Finally, we make concluding remarks. For more details on some of the topics discussed here, and for treatment of some of the other topics listed above, see the recent review paper of Bartnik and Isenberg (2004).

## Space + Time Field Decomposition and Derivation of the Constraint Equations

To understand what sort of initial data one needs to choose in order to construct a spacetime via the initial-value formulation, it is useful to consider a spacetime $(M^4, g)$ which satisfies the Einstein (–Maxwell) field equations and contains a Cauchy surface $i_0 : \Sigma^3 \to M^4$. We note that the existence of a Cauchy surface in $(M^4, g, A)$ is not automatic; if one exists, the spacetime is said to be (by definition) "globally hyperbolic."[1]

Among its other properties, a Cauchy surface is a spacelike embedded submanifold of a Lorentz geometry. It immediately follows that the spacetime $(M^4, g, A)$ induces on $\Sigma^3$ a Riemannian metric $\gamma$, a timelike normal vector field $e_\perp$, an intrinsic ($\gamma$-compatible) covariant derivative $\nabla$, and a symmetric "extrinsic curvature" tensor field $K$ (second fundamental form). It also follows that certain components of the spacetime curvature tensor can be written in terms of these Cauchy surface quantities $(\gamma, e_\perp, \nabla, K)$ along with other geometric quantities related to them, such as the spatial curvature $R$ corresponding to the induced covariant derivative $\nabla$ (Gauss–Codazzi equations).

To complete the curvature $3+1$ decomposition (i.e., to carry it out for all components of the spacetime curvature), we need not just one Cauchy surface, but rather a full local foliation $i_t : \Sigma^3 \to M^4$ of the spacetime by such submanifolds. This foliation allows one to define $e_\perp$ as a smooth vector field on an open neighborhood of the Cauchy surface $i_0(\Sigma^3)$ in $M^4$. It also results in a threading of spacetime by a congruence of timelike paths (see Figure 1). This threading may be viewed as a spacetime-filling family of observers. It also defines for the spacetime a set of coordinates relative to which one can measure and calculate the dynamics of the spacetime geometry.

It is useful for later purposes to note that at each spacetime point $p \in \Sigma_t \subset M^4$ (Here $\Sigma_t := i_t(\Sigma^3)$.) the vector $\partial/\partial t$ tangent to the threading path through $p$ may be decomposed as

$$\frac{\partial}{\partial t} = N e_\perp + X \qquad [3]$$

[1]The Taub–NUT spacetime is an example of a spacetime which is *not* globally hyperbolic.

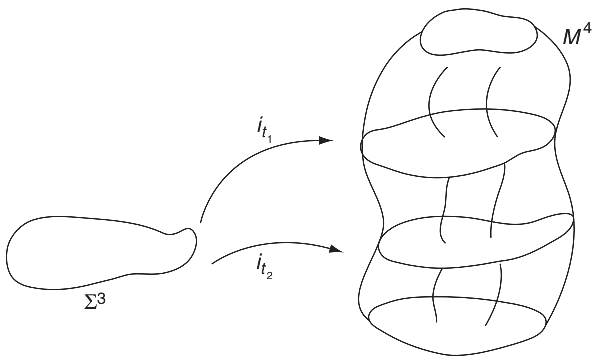



**Figure 1** 3 + 1 Foliation and threading of spacetime.

with the "shift vector" $X$ tangent to the surface ($X \in T_p\Sigma_t$), and with the "lapse" $N$ a scalar (see Figure 2). Using these quantities, we can write the spacetime metric in the form

$$\begin{aligned} g &= \gamma - \theta^\perp \otimes \theta^\perp \\ &= \gamma_{ab}(\mathrm{d}x^a + X^a \mathrm{d}t)(\mathrm{d}x^b + X^b \mathrm{d}t) - N^2 \mathrm{d}t^2 \end{aligned} \quad [4]$$

where $\theta^\perp$ is the unit length timelike 1-form which annihilates all vectors tangent to the hypersurfaces of the foliation.

Relying on the following 3 + 1 decomposition of the spacetime-covariant derivative ${}^4\nabla$ (Here $\{\partial_a\}$ is a coordinate basis for the vectors tangent to the hypersurfaces of the foliation; $\{\partial_a, e_\perp\}$ constitutes a basis for the full set of spacetime vectors at $p$.):

$${}^4\nabla_{\partial_a}\partial_b = \nabla_{\partial_a}\partial_b - K_{ab}e_\perp \quad [5]$$

$${}^4\nabla_{\partial_a}e_\perp = -K_a^m\partial_m \quad [6]$$

$${}^4\nabla_{e_\perp}\partial_b = -K_b^m\partial_m + [e_\perp, \partial_b] + \frac{\partial_b N}{N}e_\perp \quad [7]$$

$${}^4\nabla_{e_\perp}e_\perp = \gamma^{mn}\frac{\partial_m N}{N}\partial_n \quad [8]$$

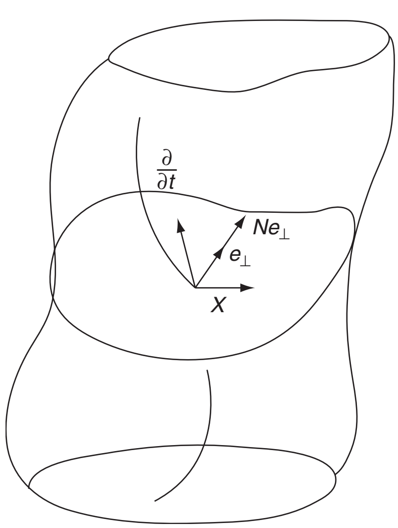



**Figure 2** Decomposition of the time evolution vector field $\partial/\partial t$.

one readily derives the (Gauss–Codazzi–Mainardi) 3 + 1 decomposition of the curvature: [2]

$${}^4R^e_{abc} = R^e_{abc} + K_{ac}K^e_b - K_{ab}K^e_c \quad [9]$$

$${}^4R^\perp_{abc} = \nabla_{\partial_c}K_{ab} - \nabla_{\partial_b}K_{ac} \quad [10]$$

$${}^4R^\perp_{a\perp b} = -\mathcal{L}_{e_\perp}K_{ab} - K_{am}K^m_b + \frac{\nabla_{\partial_a}\nabla_{\partial_b}N}{N} \quad [11]$$

where $\mathcal{L}$ denotes the surface-projected Lie derivative.

Since we are interested here in the 3 + 1 formulation of the Einstein–Maxwell system, we need a 3 + 1 decomposition for the electromagnetic as well as the gravitational field. The spacetime 1-form "vector potential" ${}^4A$ pulls back on each Cauchy surface $\Sigma_t$ to a spatial 1-form $A$. One may then write

$${}^4A = A + \mu\theta^\perp = A_b\mathrm{d}x^b + (N\mu + A_bX^b)\,\mathrm{d}t \quad [12]$$

for a scalar $\mu$. Based on this decomposition, one has the following 3 + 1 decomposition for the electromagnetic 2-form $F$:

$${}^4F_{\perp a} = \gamma_{ac}E^c \quad [13]$$

$${}^4F_{ab} = \nabla_{\partial_a}A_b - \nabla_{\partial_b}A_a \quad [14]$$

where $E^c$ is the electric vector field.

We may now use all of these decomposition formulas to write out the 14 field equations for the Einstein–Maxwell theory

$${}^4G_{\alpha\beta} = F^\mu_\alpha F_{\beta\mu} - \tfrac{1}{4}g_{\alpha\beta}F^{\mu\nu}F_{\mu\nu} \quad [15]$$

$${}^4\nabla_\mu\left({}^4F^\mu_\alpha\right) = 0 \quad [16]$$

in terms of the spatial fields $(\gamma, K, N, X; A, E, \mu)$ and their derivatives. We obtain

$$R - K^{mn}K_{mn} + (\mathrm{tr}\,K)^2 = \tfrac{1}{2}E^mE_m + \tfrac{1}{2}B^mB_m \quad [17]$$

$$\nabla_m K^m_a - \nabla_{\partial_a}(\mathrm{tr}\,K) = \epsilon_{amn}E^mB^n \quad [18]$$

$$\begin{aligned} \mathcal{L}_{e_\perp}K_{ab} &= R_{ab} - 2K^m_aK_{mb} + (\mathrm{tr}\,K)K_{ab} \\ &\quad + E_aE_b + B_aB_b - \frac{\nabla_{\partial_a}\nabla_{\partial_b}N}{N} \end{aligned} \quad [19]$$

$$\nabla_{\partial_m}E^m = 0 \quad [20]$$

$$\mathcal{L}_{e_\perp}E^a = \epsilon^{amn}\nabla_{\partial_m}B_n \quad [21]$$

where $\epsilon_{abc}$ is the alternating Levi-Civita symbol (component representation of the Hodge dual), and where we have used $B_a := \epsilon_a^{mn}(\nabla_{\partial_m}A_n - \nabla_{\partial_n}A_m)$ as a convenient shorthand.

---

[2]Here and throughout this article, we use the Misner–Thorne–Wheeler (MTW) (Misner *et al.* 1973) conventions for the definition of the Riemann curvature, for the signature $-+++$ of the metric, for the index labels (Greek indices run over $\{0, 1, 2, 3\}$ while Latin indices run over $\{1, 2, 3\}$), etc.

It is immediately evident that nine of these equations ([19] and [21]) involve time derivatives of the spatial fields, while five of them ([17], [18], and [20]) do not. Thus, we may split the field equations of the Einstein–Maxwell theory into two sets: (1) the constraint equations [17], [18], and [20], which restrict our choice of the Einstein–Maxwell initial data $(\gamma, K, A, E)$; and (2) the evolution equations, which describe how to evolve the data $(\gamma, K, A, E)$ in time, presuming that one has also prescribed (freely!) the "atlas fields" $(N, X, \mu)$.[3] We note that the complete system of evolution equations for the Einstein–Maxwell field equations includes equations which are based on the definitions of $K$ and $E$. Written in terms of (surface-projected) Lie derivatives along $\partial/\partial_t$, the full system takes the form

$$\mathcal{L}_{\frac{\partial}{\partial_t}}\gamma_{ab} = -2NK_{ab} + \mathcal{L}_X\gamma_{ab} \quad [22]$$

$$\mathcal{L}_{\frac{\partial}{\partial_t}}K_{ab} = N\big(R_{ab} - 2K_a^m K_{mb} + K_m^m K_{ab} + E_a E_b + B_a B_b\big) - \nabla_{\partial_a}\nabla_{\partial_b}N + \mathcal{L}_X K_{ab} \quad [23]$$

$$\mathcal{L}_{\frac{\partial}{\partial_t}}A_a = N(E_a + \nabla_{\partial_a}\mu) + \mathcal{L}_X A_a \quad [24]$$

$$\mathcal{L}_{\frac{\partial}{\partial_t}}E^a = N\epsilon^{amn}\nabla_{\partial_m}B_n + \mathcal{L}_X E^a \quad [25]$$

As noted earlier, well-posedness theorems[4] guarantee that initial data satisfying the constraint equations [17], [18], and [20] on a manifold $\Sigma^3$ can always at least locally be evolved into a spacetime solution $(\Sigma^3 \times I, g, {}^4A)$ (for $I$ some interval in $R^1$) of the Einstein–Maxwell equations. We now turn our attention to the issue of finding sets of data which do satisfy the constraints.

## The Conformal Method

We seek to find sets of data $(\gamma, K, A, E)$ on a manifold $\Sigma^3$ which satisfy the constraint equations

$$R - K^{mn}K_{mn} + (\mathrm{tr}\,K)^2 = \tfrac{1}{2}E^m E_m + \tfrac{1}{2}B^m B_m \quad [26]$$

$$\nabla_m K_a^m - \nabla_a(\mathrm{tr}\,K) = \epsilon_{amn}E^m B^n \quad [27]$$

$$\nabla_m E^m = 0 \quad [28]$$

[3]The collective name "atlas field" for the lapse $N$, the shift $X$, the electric potential $\mu$, and other such fields which are neither constrained by the constraint equations nor evolved by the evolution equations, derives from their role in controlling the evolution of coordinate charts and bundle atlases in the course of the construction of spacetime solutions of relativistic field equations like the Einstein Maxwell system.

[4]While the work cited earlier (Foures-Bruhat 1952) proves well posedness for the vacuum Einstein equations only, the extension to the Einstein–Maxwell system is straightforward

(Here and below, for convenience, we replace $\nabla_{\partial_a}$ by $\nabla_a$.) This is an underdetermined problem, with five equations to be solved for 18 functions.

The idea of the conformal method is to divide the initial data on $\Sigma^3$ into two sets – the "free (conformal) data," and the "determined data" – in such a way that, for a given choice of the free data, the constraint equations become a determined elliptic partial differential equation (PDE) system, to be solved for the determined data. There are a number of ways to do this; we focus here on one of them – the "semidecoupling split" or "method $A$." After describing this version of the conformal method, and discussing what one can do with it, we note some of its drawbacks and then later (in the next section) consider some alternatives. (See Choquet-Bruhat and York (1980) and Bartnik and Isenberg (2004) for a more complete discussion of these alternatives.)

For the Einstein–Maxwell theory, the split of the initial data is as follows:

**Free ("conformal") data**

- $\lambda_{ij}$ – a Riemannian metric, specified up to conformal factor;
- $\sigma_{ij}$ – a divergence-free[5] $(\nabla^i\sigma_{ij} = 0)$, tracefree $(\lambda^{ij}\sigma_{ij} = 0)$; symmetric tensor;
- $\tau$ – a scalar field;
- $\alpha_a$ – a 1-form;
- $\mathcal{E}^b$ – a divergence-free vector field;

**Determined data**

- $\phi$ – a positive-definite scalar field;
- $W^i$ – a vector field;
- $\xi$ – a scalar field.

For a given choice of the free data, the five equations to be solved for the five functions of the determined data take the form

$$\Delta\xi = 0 \quad [29]$$

$$\nabla_m(LW)_a^m = \tfrac{2}{3}\phi^6\nabla_{\partial_a}\tau + \epsilon_{amn}\mathcal{E}^m\beta^n \quad [30]$$

$$\Delta\phi = \tfrac{1}{8}R\phi - \tfrac{1}{8}(\sigma^{mn} + LW^{mn})(\sigma_{mn} + LW_{mn})\phi^{-7} + \tfrac{1}{16}(\mathcal{E}^m\mathcal{E}_m + \beta^m\beta_m)\phi^{-3} + \tfrac{1}{12}\tau^2(\phi^5) \quad [31]$$

where the Laplacian $\Delta$ and the scalar curvature $R$ are based on the $\lambda_{ab}$-compatible covariant derivative $\nabla_i$, where $L$ is the corresponding conformal Killing operator, defined by

$$(LW)_{ab} := \nabla_a W_b + \nabla_b W_a - \tfrac{2}{3}\lambda_{ab}\nabla_m W^m \quad [32]$$

[5]In the free data, the divergence-free condition is defined using the Levi-Civita-covariant derivative compatible with the conformal metric $\lambda_{ij}$.

and where $\beta_a := \epsilon_a^{mn}(\nabla_{\partial_m}\alpha_n - \nabla_{\partial_n}\alpha_m)$. Presuming that for the chosen free data one can indeed solve equations [29]–[31] for $\xi, \phi$, and $W$, then the initial data $(\gamma, K, A, E)$ constructed via the formulas

$$\gamma_{ab} = \phi^4 \lambda_{ab} \qquad [33]$$

$$K_{ab} = \phi^{-2}(\sigma_{ab} + LW_{ab}) + \tfrac{1}{3}\phi^4\lambda_{ab}\tau \qquad [34]$$

$$A_b = \alpha_b \qquad [35]$$

$$E^a = \phi^{-6}(\mathcal{E}_a + \nabla_{\partial a}\xi) \qquad [36]$$

satisfy the Einstein–Maxwell constraint equations [26]–[28].

Before discussing the extent to which one can solve equations [29]–[31] and consequently use the conformal method to generate solutions, we wish to comment on how these equations are derived. Three formulas are key to this derivation. The first is the formula for the scalar curvature of the metric $\gamma_{ab} = \phi^4\lambda_{ab}$, expressed in terms of the scalar curvature for $\lambda_{ab}$ and derivatives of $\phi$:

$$R(\gamma) = \phi^{-4}R(\lambda) - 8\Delta_\lambda\phi \qquad [37]$$

We note that if we were to use a different power of $\phi$ as the conformal factor multiplying $\lambda_{ab}$, then this formula would involve squares of first derivatives of $\phi$ as well. The second key formula relates the divergence of a traceless symmetric tensor $\rho_{ab}$ with respect to the covariant derivatives $\nabla_{(\gamma)}$ and $\nabla_{(\lambda)}$ compatible with conformally related metrics. One obtains

$$\nabla^m_{(\gamma)}\rho_{mb} = \phi^{-2}\nabla^m_{(\lambda)}(\phi^2\rho_{mb}) \qquad [38]$$

The third key formula does the same thing for a vector field $\zeta^a$:

$$\nabla_{(\gamma)m}\zeta^m = \phi^{-6}\nabla^m_{(\lambda)}(\phi^6\zeta_m) \qquad [39]$$

In addition to helping us derive equations [29]–[31] from the substitution of formulas [33]–[36] into [26]–[28], these key formulas indicate to some extent how the choice of the explicit decomposition of the initial data into free and determined data is made (see Isenberg, Maxwell, and Pollack for further elaboration).

It is easy to see that there are some choices of the free data for which [29]–[31] do not admit any solutions. Let us choose, for example, $\Sigma^3$ to be the 3-sphere, and let us set $\lambda$ to be the round sphere metric, $\sigma$ to be zero everywhere, $\tau$ to be unity everywhere, and both $\alpha$ and $\mathcal{E}$ to vanish everywhere. We then readily determine that eqn [29] requires that $\xi$ be constant and that eqn [30] requires that $LW_{ab}$ be zero. The remaining equation [31] now takes the form $\Delta\phi = (1/8)R\phi + (1/12)\phi^5$. Since the right-hand side of this equation is positive definite (recall the requirement that $\phi > 0$), it follows from the maximum principle on closed (compact without boundary) manifolds that there is no solution.

In light of this example, one would like to know exactly for which sets of free data eqns [29]–[31] can be solved, and for which sets they cannot. Since one readily determines that every set of initial data which satisfies the Einstein–Maxwell constraints [26]–[28] can be obtained via the conformal method, such a classification effectively provides a parametrization of the space of solutions of the constraints.[6]

What we know and do not know about classifying free data for the solubility of eqns [29]–[31] is largely determined by whether or not the function $\tau$ is chosen to be constant on $\Sigma^3$. If $\tau$ is chosen to be constant, then eqns [29]–[31] effectively decouple, and the classification is essentially completely known. Sets of initial data generated from free data with constant $\tau$ are called "constant mean curvature" (CMC) sets, since the mean curvature of the initial slice embedded in its spacetime development is given by $\tau$. We also know a considerable amount about the classification if $|\nabla\tau|$ is sufficiently small ("near CMC"), while virtually nothing is known for the general non-CMC case.

A full account of the classification results known to date is beyond the scope of this article. Indeed, such an account must separately deal with a number of alternatives regarding manifold and asymptotic conditions (data on a closed manifold; asymptotically Euclidean data; asymptotically hyperbolic data; data on an incomplete manifold with boundaries) and regularity (analytic data, smooth data, $C^k$ data, or data contained in various Hölder or Sobolev spaces), among other things. We will, however, now summarize some of the results; see, for example, Bartnik and Isenberg (2004) or Choquet-Bruhat for more complete surveys.

### CMC Data on Closed Manifolds

Generalizing the $S^3$ example given above, we note that for any set of free data $(\Sigma^3, \lambda_{ab}, \sigma_{ab}, \tau, \alpha_a, \mathcal{E}^b)$ with constant $\tau$ and with no conformal Killing fields, eqn [29] is easily solved for $\xi$, and then eqn [30] takes the form

$$\nabla_m(LW)^m_a = \epsilon_{amn}\mathcal{E}^m\beta^n \qquad [40]$$

[6] Of course, in claiming that appropriate sets of the free data parametrize the space of solutions of the constraints, one needs to determine if inequivalent sets of free data are mapped to the same set of solutions. We discuss this below.

which is a linear elliptic PDE for $W_m$ with invertible operator.[7] This equation admits a unique solution, and then the problem of solving the constraints reduces to the analysis of the "Lichnerowicz equation" [31].

To determine if this equation admits a solution for the given set of free data, we use the following classification criteria: (1) The metric is labeled positive $\mathcal{Y}^{+}(\Sigma^3)$, zero $\mathcal{Y}^{0}(\Sigma^3)$, or negative $\mathcal{Y}^{-}(\Sigma^3)$ Yamabe class depending upon whether the metric $\lambda_{ab}$ on $\Sigma^3$ can be conformally deformed so that its scalar curvature is everywhere positive, everywhere zero, or everywhere negative.[8] (2) The $(\sigma_{ab}, \alpha_a, \mathcal{E}^b)$ portion of the data is labeled either $\equiv$ or $\not\equiv$, depending upon whether the quantity $\sigma_{mn}\sigma^{mn} + \mathcal{E}^m\mathcal{E}_m + \beta^m\beta_m$ is identically zero, or not. (3) The mean curvature $\tau$ is labeled "max" or "nonmax" depending upon whether the constant $\tau$ is zero or not. In terms of these criteria, we have 12 classes of free data, and one can prove (Choquet-Bruhat and York 1980, Isenberg 1995) the following:

- Solutions exist for the classes $(\mathcal{Y}^{+}, \not\equiv, \text{max})$, $(\mathcal{Y}^{+}, \not\equiv, \text{nonmax})$, $(\mathcal{Y}^{0}, \equiv, \text{max})$, $(\mathcal{Y}^{0}, \not\equiv, \text{max})$, $(\mathcal{Y}^{-}, \equiv, \text{nonmax})$, $(\mathcal{Y}^{-}, \not\equiv, \text{nonmax})$ and
- Solutions do not exist for the classes $(\mathcal{Y}^{+}, \equiv, \text{max})$, $(\mathcal{Y}^{+}, \equiv, \text{nonmax})$, $(\mathcal{Y}^{0}, \equiv, \text{nonmax})$, $(\mathcal{Y}^{0}, \not\equiv \text{max})$, $(\mathcal{Y}^{-}, \equiv, \text{max})$, $(\mathcal{Y}^{-}, \not\equiv, \text{max})$.

This classification is exhaustive, in the sense that every set of CMC data on a closed manifold fits neatly into exactly one of the classes. We note that the proofs of existence of solutions can generally be done using the sub–super solution technique, while the nonexistence results follow from application of the maximum principle.

### Maximal Asymptotically Euclidean Data

Just as is the case for data on a closed manifold, the constraint equations [29] and [30] decouple from the Lichnerowicz equation [31] for asymptotically Euclidean data with constant $\tau$. We note that $\tau \neq 0$ is inconsistent with the data being asymptotically Euclidean, so we restrict to the maximal case, $\tau = 0$.

The criterion for solubility of the constraints in conformal form for maximal asymptotically Euclidean free data is quite a bit simpler to state than that for CMC data on a closed manifold. It involves the metric $\lambda$ only; the rest of the free data is irrelevant. Specifically, as shown by Brill and Cantor (with a correction by Maxwell (2005)), a solution exists if and only if for every nonvanishing, compactly supported, smooth function $f$ on $\Sigma^3$, we have

$$\inf_{\{f \neq 0\}} \frac{\int_M (|\nabla f|^2 + Rf^2)\sqrt{\det \lambda}}{\|f\|^2_{L^2}} > 0 \qquad [41]$$

## Alternative Methods for Finding Solutions to the Constraint Equations

While the conformal method has proved to be a very useful tool for generating and analyzing solutions of the Einstein constraint equations, it does have some minor drawbacks: (1) The free data is remote from the physical data, since the conformal factor can vastly change the physical scale on different regions of space. (2) While casting the constraints into a determined PDE form has the advantage of producing PDEs of a relatively familiar (elliptic) form, one does give up certain flexibilities inherent in an underdetermined set of PDEs. (We expand upon this point below in the course of discussing gluing.). (3) In choosing a set of free data, one does have to first project out a divergence-free vector field ($\mathcal{E}$) and a divergence-free tracefree tensor field ($\sigma$). (4) While the choice of CMC free data for the conformal method is conformally covariant in the sense that conformally related sets of CMC free data $(\Sigma^3, \lambda_{ab}, \sigma_{ab}, \tau, \alpha_a, \mathcal{E}^b)$ and $(\Sigma^3, \theta^4\lambda_{ab}, \theta^{-2}\sigma_{ab}, \tau, \alpha_a, \theta^{-6}\mathcal{E}^b)$ produce the same physical solution to the constraints, this is not the case for non-CMC free data.

### Conformal Thin Sandwich

The last two of these problems can be removed by modifying the conformal method in a way which York (1999) has called the "conformal thin sandwich" (CTS) approach. The basic idea of the CTS approach is the same as that of the conformal method. However, CTS free data sets are larger – the divergence-free tracefree symmetric tensor field $\sigma$ is replaced by a tracefree symmetric tensor field $U$,

---

[7] A metric $\lambda$ has a conformal Killing field if the equation $LY = 0$ has a nontrivial solution $Y$. Geometrically, the existence of a conformal Killing field $Y$ indicates that the flow of $(\Sigma^3, \lambda_{ab})$ along $Y$ is a conformal isometry. While free data with nonvanishing conformal Killing fields can be handled, for convenience we shall stick to data without them here.

[8] Work on the Yamabe problem (Aubin 1998) shows that every Riemannian metric on a closed manifold is contained in one and only one of these classes. In fact, the Yamabe theorem (Schoen 1984) shows that every metric can be conformally deformed so that its scalar curvature is $+1, 0$, or $-1$, but this result is not needed for the analysis of the constraint equations.

and an extra scalar field $\eta$ is added – and after solving the CTS constraint equations

$$\Delta\xi = 0 \quad [42]$$

$$\nabla_m((2\eta)^{-1}(LX))_a^m = \tfrac{2}{3}\Phi^6\nabla_a\tau + \epsilon_{amn}\mathcal{E}^m\beta^n + \nabla_m\left((2\eta)^{-1}U_a^m\right) \quad [43]$$

$$\Delta\Phi = \tfrac{1}{8}R\Phi - \tfrac{1}{8}(U^{mn} + LY^{mn})(U_{mn} + LY_{mn})\Phi^{-7} + \tfrac{1}{16}(\mathcal{E}^m\mathcal{E}_m + \beta^m\beta_m)\Phi^{-3} + \tfrac{1}{12}\tau^2\Phi^5 \quad [44]$$

for the vector field $Y$ and the conformal factor $\Phi$, one obtains not just the full set of physical initial data satisfying the constraint equations [26]–[28]

$$\gamma_{ab} = \phi^4\lambda_{ab} \quad [45]$$

$$K_{ab} = \Phi^{-2}(-U_{ab} + LY_{ab}) + \tfrac{1}{3}\Phi^4\lambda_{ab}\tau \quad [46]$$

$$A_b = \alpha_b \quad [47]$$

$$E^a = \Phi^{-6}(\mathcal{E}_a + \nabla_a\xi) \quad [48]$$

but also the lapse $N$ and shift $X$

$$N = \Phi^6\eta \quad [49]$$

$$X^a = Y^a \quad [50]$$

Clearly, in using the CTS approach, one need not project out a divergence-free part of a symmetric tracefree tensor. One also readily checks that the CTS method is conformally covariant in the sense discussed above: the physical data generated from CTS free data $(\lambda_{ab}, U_{ab}, \tau, \eta, \alpha_a, \mathcal{E}^b)$ and from data $(\theta^4\lambda_{ab}, \theta^{-2}U_{ab}, \tau, \theta^6\eta, \alpha_a, \theta^{-6}\mathcal{E}^b)$ are the same. Furthermore, since the mathematical form of eqns [42]–[44] is very similar to that of [29]–[31], the solvability results for the conformal method can be essentially carried over to the CTS approach.

There is, however, one troubling feature of the CTS approach. The problem arises if we seek CMC initial data with the lapse function chosen so that the evolving data continue to have CMC (such a gauge choice is often used in numerical relativity). In the case of the conformal method, after solving [29]–[31] to obtain initial data $(\gamma_{ab}, K_{ab}, A_a, E^b)$ which satisfies the constraints, one achieves this by proceeding to solve a linear homogeneous elliptic PDE for the lapse function. One easily verifies that solutions to this extra equation always exist. By contrast, in the CTS approach, the extra equation takes the form

$$\Delta(\Phi^7\eta) = \tfrac{1}{8}\Phi^7\eta R + \tfrac{5}{2}(\Phi\eta)^{-1}(U - LX)^2 + \tfrac{1}{16}(\Phi\eta)^{-1}(\mathcal{E}^2 + \beta^2) + \Phi^5 Y^m\nabla_{\partial_m}\tau - \Phi^5 \quad [51]$$

which is coupled to the system [42]–[44]. The coupling is fairly intricate; hence little is known about the existence of solutions to the system, and it has been seen that there are problems with uniqueness. Such problems of course do not arise if one makes no attempt to preserve CMC.

### The Quasispherical Ansatz and Parabolic Methods

Applying either the conformal method or the CTS approach to the constraint equations results in systems of elliptic equations. Another approach, pioneered by Bartnik (1993), produces instead parabolic equations. In the simplest version of this approach, known as the "quasispherical ansatz," one works on a manifold $\Sigma^3 = R^3\backslash B_3$, where $B_3$ is a 3-ball; one presumes that there exist coordinates $(r, \theta, \phi)$ on $\Sigma^3$ in terms of which the metric takes the "quasispherical" form

$$\gamma_{QS} = u^2\mathrm{d}r^2 + (r\mathrm{d}\theta + \beta^\theta\mathrm{d}r)^2 + (r\sin\theta\,\mathrm{d}\phi + \beta^\phi\mathrm{d}r)^2 \quad [52]$$

for functions $u(r,\theta,\phi)$, $\beta^\theta(r,\theta,\phi)$, $\beta^\phi(r,\theta,\phi)$, and then one attempts to satisfy the time-symmetric constraint $R_{(\gamma_{QS})} = 0$ on $\Sigma^3$.[9] Calculating the scalar curvature for the metric in this form, one finds that the equation $R_{(\gamma_{QS})} = 0$ can be written as

$$(r\partial_r - \beta^\theta\partial_\theta - \beta^\phi\partial_\phi)u - u^2\Delta u = Q(u, \beta^\theta, \beta^\phi, r, \theta, \phi) \quad [53]$$

where $Q$ is a polynomial in the positive function $u$.

One can now show that if one specifies $\beta^\theta$ and $\beta^\phi$ everywhere on $\Sigma^3$ (subject to an upper bound on the divergence of the vector field $(\beta^\theta, \beta^\phi)$), and if one specifies regular initial data for $u$ on the inner boundary of $\Sigma^3$, then one has a well-posed initial-value problem (in terms of the "evolution" coordinate $r$) for the parabolic PDE [53]. Ideally, one can use this approach to extend solutions of the time-symmetric constraints from an isolated region (corresponding to $B_3$) out to spatial infinity.

The basic quasispherical ansatz approach just outlined can be generalized significantly (Sharples 2001, Bartnik and Isenberg 2004) to allow for more general spatial metrics, and to allow nonzero $K_{ab}, A_c$, and $E^b$. It has been an especially valuable tool for the study of mass in asymptotically Euclidean data sets. It does not, however, purport to construct general solutions of the constraint equations.

[9] This version of the constraints is called "time symmetric" since one is solving the full set of constraints with $K_{ab}$ assumed to be zero. Data with $K_{ab} = 0$ is time symmetric.

### Gluing Solutions of the Constraint Equations

Starting around the year 2000, a number of new "gluing" procedures have been developed for constructing and studying solutions of the constraint equations. Unlike the conformal method, the CTS method, and the quasispherical ansatz, all of which construct solutions from scratch, the gluing procedures construct new solutions from given ones. This feature, and the considerable flexibility of the procedures, has resulted in a wealth of applications already in the short five-year history of gluing in general relativity.

One of the gluing approaches, developed by Corvino (2000) and Corvino and Schoen (preprint) (see also Chruściel and Delay (2002)), allows one to choose a compact region $\Omega$ in almost any smooth, asymptotically Euclidean vacuum solution of the constraints, and from this produce a new smooth solution which is completely unchanged in the region $\Omega$ and is identical to Schwarzschild or Kerr outside some larger region. In proving this result, one exploits the underdetermined character of the constraint equations: such a construction could not be carried out if the constraints were a determined PDE system.[10]

The other main gluing approach, developed first by Isenberg *et al.* (2001), and then further developed with Chruściel (Chruściel *et al.* 2005) and with Maxwell (Isenberg *et al.* 2005), starts with a pair of solutions of the (vacuum) constraints $(\Sigma_1^3, \gamma_1, K_1)$ and $(\Sigma_2^3, \gamma_2, K_2)$ together with a choice of a pair of points $p_1 \in \Sigma_1^3$, $p_2 \in \Sigma_2^3$, one from each solution. From these solutions, this gluing procedure produces a new set of initial data $(\Sigma^3_{(1-2)}, \gamma_{(1-2)}, K_{(1-2)})$ with the following properties: (1) $\Sigma_{(1-2)}$ is diffeomorphic to the connected sum $\Sigma_1^3 \# \Sigma_2^3$; (2) $(\Sigma^3_{(1-2)}, \gamma_{(1-2)}, K_{(1-2)})$ is a solution of the constraints everywhere on $\Sigma^3_{(1-2)}$; (3) On that portion of $\Sigma^3_{(1-2)}$ which corresponds to $\Sigma_1^3 \backslash \{\text{ball around } p_1\}$, the data $(\gamma_{(1-2)}, K_{(1-2)})$ is isomorphic to $(\gamma_1, K_1)$, with a corresponding property holding on that portion of $\Sigma_2^3$ which corresponds to $\Sigma_2^3 \backslash \{\text{ball around } p_2\}$ (see **Figure 3**).[11]

This connected sum gluing can be carried out for very general sets of initial data. The sets can be asymptotically Euclidean, asymptotically hyperbolic, specified on a closed manifold, or indeed anything else. The only condition that the data sets must satisfy is that, in sufficiently small neighborhoods of each of the points at which the gluing is to be done, there do not exist nontrivial solutions $\xi$ to the equation $D\Theta^*_{(\gamma,K)}\zeta = 0$, where $D\Theta^*_{(\gamma,K)}$ is the operator obtained by taking the adjoint of the linearized constraint operator.[12] In work by Beig, Chruściel, and Schoen, it is shown that this condition (sometimes referred to as "No KIDs," meaning "no (localized) Killing initial data)" is indeed generically satisfied.

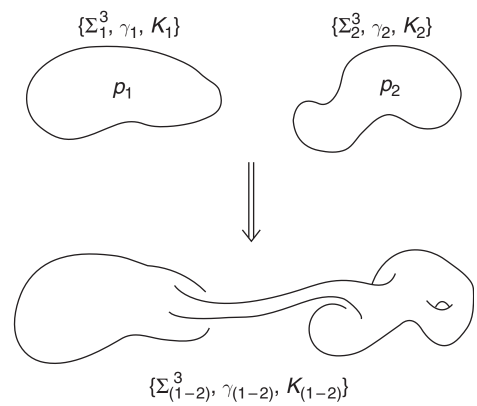


**Figure 3** Connected sum gluing.

While a discussion of the proof that connected sum gluing can be carried out to this degree of generality is beyond the scope of this paper (see Chruściel *et al.* (2005), along with references cited therein for details of the proof), we note three features of it: first, the proof is constructive in the sense that it outlines a systematic, step-by-step mathematical procedure for doing the gluing. In principle, one should be able to carry out the gluing procedure numerically. Second, connected sum gluing relies primarily on the conformal method, but it also uses a nonconformal deformation at the end (dependent on the techniques of Corvino and Schoen, and of Chruściel and Delay), so as to guarantee that the glued data is not just very close to the given data on regions away from the bridge, but is indeed identical to it. Third, while Corvino–Schoen gluing has not yet been proved to work for solutions of the constraints with source fields, connected sum gluing (up to the last step, which relies on Corvino–Schoen) has been shown to work for most matter source fields of interest (Isenberg *et al.*). It has also been shown to work for general dimensions greater than or equal to three.

---

[10]Hence if one tries to do Corvino Schoen-type gluing using a fixed conformal geometry, the gluing fails because the determined elliptic system satisfies the unique continuation property.

[11]The connected sum of the two manifolds (see property (1)) is constructed as follows: first we remove a ball from each of the manifolds $\Sigma_1^3$ and $\Sigma_2^3$. We then use a cylindrical bridge $S^2 \times I$ (where $I$ is an interval in $R^1$) to connect the resulting $S^2$ boundaries on each manifold

[12]When a solution to this equation does exist on some region $\Lambda \in \Sigma^3$, it follows from the work of Moncrief that the spacetime development of the data on $\Lambda$ admits a nontrivial isometry.

While gluing is not an efficient tool for studying the complete set of solutions to the constraints, it has proved to be very valuable for a number of applications. We note a few here.

1. *Spacetimes with regular asymptotic structure.* Until recently, it was not known whether there is a large class of solutions which admit the conformal compactification and consequent asymptotically simple structure at null and space-like infinity characteristic of the Minkowski and Schwarzschild spacetimes. Using Corvino–Schoen gluing, together with Friedrich's analyses of spacetime asymptotic structures and an argument of Chruściel and Delay (2002), one produces such a class of solutions.
2. *Multi-black hole data sets.* Given an asymptotically Euclidean solution of the constraints, connected sum gluing allows a sequence of (almost) flat space initial data sets to be glued to it. The bridges that result from this gluing each contain a minimal surface, and consequently an apparent horizon. With a bit of care, one can do this in such a way that indeed the event horizons which appear in the development of this glued data are disjoint, and therefore indicative of independent black holes.
3. *Adding a black hole to a cosmological spacetime.* Although there is no clear established definition for a black hole in a spatially compact solution of Einstein's equations, one can glue an asymptotically Euclidean solution of the constraints to a solution on a compact manifold, in such a way that there is an apparent horizon on the bridge. Studying the nature of these solutions of the constraints, and their evolution, could be useful in trying to understand what one might mean by a black hole in a cosmological spacetime.
4. *Adding a wormhole to your spacetime.* While we have discussed connected sum gluing as a procedure which builds solutions of the constraints with a bridge connecting two points on different manifolds, it can also be used to build a solution with a bridge connecting a pair of points on the same manifold. This allows one to do the following: if one has a globally hyperbolic spacetime solution of Einstein's equations, one can choose a Cauchy surface for that solution, choose a pair of points on that Cauchy surface, and glue the solution to itself via a bridge from one of these points to the other. If one now evolves this glued-together initial data into a spacetime, it will likely become singular very quickly because of the collapse of the bridge. Until the singularity develops, however, the solution is essentially as it was before the gluing, with the addition of an effective wormhole. Hence, this procedure can be used to glue a wormhole onto a generic spacetime solution.
5. *Removing topological obstructions for constraint solutions.* We know that every closed three-dimensional manifold $M^3$ admits a solution of the vacuum constraint equations. To show this, we use the fact that $M^3$ always admits a metric $\Gamma$ of constant negative scalar curvature. One easily verifies that the data $(\gamma = \Gamma, K = \Gamma)$ is a CMC solution. Combining this result with connected sum gluing, one can show that for every closed $\Sigma^3$, the manifold $\Sigma^3 \setminus \{p\}$ admits both an asymptotically Euclidean and an asymptotically hyperbolic solution of the vacuum constraint equations.
6. *Proving the existence of vacuum solutions on closed manifolds with no CMC Cauchy surface.* Based on the work of Bartnik (1988) one can show that if one has a set of initial data on the manifold $T^3 \# T^3$ with the metric components symmetric across a central sphere and the components of $K$ skew symmetric across that same central sphere, then the spacetime development of that data does not admit a CMC Cauchy surface. Using connected sum gluing, one can show that indeed initial data sets of this sort exist (Chruściel *et al.* 2005).

## Conclusion

Much is known about the Einstein constraint equations and those sets of initial data which satisfy them. We know how to use the conformal method or the CTS approach to construct (and parametrize in terms of free data) the CMC and near CMC sets of data which solve the constraints, with or without matter fields present. We know how to use the quasispherical approach to explore extensions of solutions of the constraint equations from compact regions. We know how to use gluing techniques to produce new solutions of both physical and mathematical interest from old ones, and we know how to use gluing as a tool for proving such results as the existence of vacuum spacetimes with no CMC Cauchy surfaces.

There is much that is not yet known as well. Very little is known about solutions of the constraint equations which have neither CMC nor near CMC. It is not known how to systematically extend solutions of the constraints from a compact region to all of $R^3$ in such a way that the extension is asymptotically Euclidean (unless we know *a priori* that such an extension exists). Very little is known regarding how to control the constraints during the course of

numerical evolution of solutions.[13] Most importantly, we do not yet know how to systematically find solutions of the constraint equations which serve as physically realistic model initial data sets for studying astrophysical and cosmological systems of interest.

Many of these questions concerning the Einstein constraints and their solutions are fairly daunting. However, in view of the rapid progress in our understanding during the last few years, and in view of the pressing need to further develop the initial-value formulation as a tool for studying general relativity and gravitational physics, we are optimistic that this progress will continue, and we will soon have answers to a number of these questions.

## Acknowledgments

The work of J Isenberg is supported by the NSF under Grant PHY-035 4659.

*See also:* Asymptotic Structure and Conformal Infinity; Computational Methods in General Relativity: The Theory; Dirac Fields in Gravitation and Nonabelian Gauge Theory; Einstein Manifolds; Einstein's Equations with Matter; General Relativity: Overview; Geometric Analysis and General Relativity; Hamiltonian Reduction of Einstein's Equations; Spacetime Topology, Causal Structure and Singularities; Stationary Black Holes; Symmetric Hyperbolic Systems and Shock Waves.

## Further Reading

Aubin T (1998) *Some Nonlinear Problems in Riemannian Geomertry*. Springer.

Bartnik R (1988) Remarks on cosmological spacetimes and constant mean curvature surfaces. *Communications in Mathematical Physics* 117: 615–624.

Bartnik R (1993) Quasi-spherical metrics and prescribed scalar curvature. *Journal of Differential Geometry* 37: 31–71.

Bartnik R and Isenberg J (2004) The constraint equations. In: Chruściel PT and Friedrich H (eds.) *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, pp. 1–39. Basel: Birkhäuser.

Choquet-Bruhat Y *General Relativity* (to be published).

Choquet-Bruhat Y and York J (1980) The Cauchy problem. In: Held A (ed.) *General Relativity and Gravitation – The Einstein Centenary*, pp. 99–160. Plenum.

Chruściel P and Delay E (2002) Existence of non-trivial, vacuum, asymptotically simple spacetimes. *Classical and Quantum Gravity* 19: L71–L79.

Chruściel P and Delay E (2003) On mapping properties of the general relativistic constraints operator in weighted function spaces, with applications. *Mémoires de la Société Mathématique de France* 93: 1–103.

Chruściel P, Isenberg J, and Pollack D (2005) Initial data engineering. *Communications in Mathematical Physics* 257: 29–42 (gr-qc/0403066).

Corvino J (2000) Scalar curvature deformation and a gluing construction for the Einstein constraint equations. *Communications in Mathematical Physics* 214: 137–189.

Corvino J and Schoen R On the asymptotics for the vacuum constraint equations. *Preprint*, gr-qc/0301071 (to appear *Journal of Differential Geometry*).

Foures-Bruhat Y (1952) Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires. *Acta Mathematicae* 88: 141–225.

Isenberg J (1995) Constant mean curvature solutions of the Einstein constraint equations on closed manifolds. *Classical and Quantum Gravity* 12: 2249–2274.

Isenberg J, Maxwell D, and Pollack D, A gluing construction for non-vacuum solutions of the Einstein constraint equations. *Advances in Theoretical and Mathmatical Physics* (to be published).

Isenberg J, Mazzeo R, and Pollack D (2001) Gluing and wormholes for the Einstein constraint equations. *Communications in Mathematical Physics* 231: 529–568.

Maxwell D (2005) Rough solutions of the Einstein constraints on compact manifolds. *Journal of Hyperbolic Differential Equations* 2: 521–546.

Misner C, Thorne K, and Wheeler JA (1973) *Gravitation.* Chicago: Freeman.

Schoen R (1984) Conformal deformation of a Riemannian metric to constant scalar curvature. *Journal of Differential Geometry* 20: 479–495.

Sharples J (2001) *Spacetime Initial Data and Quasi-Spherical Coordinates*. Ph.D. thesis, University of Canberra.

York JW (1999) Conformal "Thin-Sandwich" data for the initial-value problem of general relativity. *Physical Review Letters* 82: 1350–1353.

[13]If the constraints are satisfied by an initial data set and if this data set is evolved completely accurately, then the constraints remain satisfied for all time. However, during the course of a numerical evolution, there are inevitable numerical inaccuracies which result in the constraints not being exactly zero. In practice, during the majority of such numerical simulations to date, the constraints have been seen to increase very rapidly in time, calling into question the reliability of the simulation.

# Einstein Manifolds

**A S Dancer**, University of Oxford, Oxford, UK



## Introduction

The Einstein condition on a manifold $M$ with metric $g$ says that the Ricci curvature should be proportional to the metric. Of course, this condition originally appeared in relativity, but it is of tremendous interest from the point of view of pure mathematics. Demanding a metric of constant sectional curvature is a very strong condition, while metrics of constant scalar curvature always occur. The Einstein property, which is essentially a constant-Ricci-curvature condition, occupies an intermediate position between these conditions, and it is still not clear exactly how strong it is. In

dimensions higher than four, it is still unknown whether there are obstructions to a manifold admitting an Einstein metric.

The study of Einstein manifolds is a vast and rapidly expanding area, and this article can merely touch on some points of particular interest. The focus of the article is very much on the Riemannian rather than Lorentzian case (see, e.g., Hawking and Ellis (1973) or the articles by Christodoulou and Tod in LeBrun and Wang (1999) for a discussion of the Lorentzian case in general relativity). For further reading, the books of Besse (1987) and LeBrun and Wang (1999) are strongly recommended.

## Basic Properties

Let $(M, g)$ be a (pseudo)-Riemannian manifold. There is a unique connection $\nabla$, the Levi-Civita connection of $g$, with the following properties:

1. the torsion $T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y]$ vanishes and
2. $\nabla g = 0$

We can now form the Riemann curvature tensor of $g$:

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z$$

This is a type (3,1) tensor. There is one nontrivial contraction we can perform to obtain a (2, 0) tensor, that is, the Ricci curvature

$$\mathrm{Ric}(X, Y) = \mathrm{tr}(Z \mapsto R(X, Z)Y)$$

We may perform a further contraction and obtain the scalar curvature $s = \mathrm{tr}_g \mathrm{Ric}$.

The Ricci curvature is a symmetric tensor of the same type as the metric, so we can make the following definition:

**Definition 1** A metric $g$ is Einstein if

$$\mathrm{Ric} = \Lambda g \qquad [1]$$

for some constant $\Lambda$.

In this article, we shall take $g$ to be a Riemannian (positive-definite) metric.

**Remark 1** In dimension higher than 2, we do not have to put in the assumption that $\Lambda$ is constant by hand. For, taking the divergence of [1] gives $(1/2)\mathrm{d}s = \mathrm{d}\Lambda$, while taking instead the trace gives $s = n\Lambda$, so if $n \neq 2$, we see $\mathrm{d}\Lambda = 0$.

**Remark 2** In dimension 2 and 3, the Einstein condition is equivalent to constant curvature. The only complete Einstein manifolds in these dimensions are therefore the model spaces $S^n, \mathbb{R}^n$ and hyperbolic space, and quotients of these by discrete groups of isometries.

**Remark 3** As noticed by Hilbert, the Einstein equations admit a variational interpretation. They are the variational equations for the total scalar curvature functional

$$g \mapsto \int_M s_g \, \mathrm{d}\mu_g$$

restricted to the space of volume 1 metrics (here $\mathrm{d}\mu_g$ denotes the volume form defined by $g$).

## Obstructions

The most fundamental question, we can ask is: *Given a smooth manifold M does it support an Einstein metric?*

One is also interested in the question of uniqueness of such a metric, or more generally of describing the moduli space of such metrics.

In this section we discuss obstructions to existence. In dimension 2 Remark 2 shows that any compact manifold admits an Einstein metric, while in dimension 3 the only possibilities are space forms. In particular, there is no Einstein metric on $S^1 \times S^2$.

The picture is much less clear in higher dimensions. If $\Lambda \geq 0$, one obtains some elementary obstructions just by considering the sign of the Ricci curvature:

1. If $M$ supports a complete Einstein metric with $\Lambda > 0$, then by Myers's theorem $M$ is compact and $\pi_1(M)$ is finite. Also there are obstructions coming from the positivity of the scalar curvature (e.g., if $M$ is spin and $4m$-dimensional, then the $\hat{A}$ genus vanishes).
2. If $M$ supports a complete Ricci-flat metric, then every finitely generated subgroup of $\pi_1(M)$ has polynomial growth.

However, if $\dim M \geq 5$, there is, at the time of writing, no known obstruction to $M$ supporting an Einstein metric of negative Einstein constant.

In the borderline dimension 4, Hitchin and Thorpe observed that the Einstein condition put topological constraints on the manifold. for, we have the following expressions for the Euler characteristic $\chi$ and signature $\tau$ in terms of the curvature tensor:

$$\tau = \frac{1}{12\pi^2} \int_M |W_+|^2 - |W_-|^2 \, \mathrm{d}\mu_g$$

$$\chi = \frac{1}{8\pi^2} \int_M |W_+|^2 + |W_-|^2 - |\mathrm{Ric}_0|^2 + \frac{s^2}{24} \, \mathrm{d}\mu_g$$

where $W_+$ and $W_-$ are the self-dual and anti-self-dual parts of the Weyl tensor, $s$ is the scalar curvature, and $\mathrm{Ric}_0$ is the trace-free part of the Ricci tensor.

The Einstein condition is just $\mathrm{Ric}_0 = 0$, so we immediately obtain the following inequality.

**Theorem** (Hitchin 1974). *A compact four-dimensional Einstein manifold satisfies the inequality*

$$|\tau| \leq \tfrac{2}{3}\chi$$

Note that equality is obtained if and only if $g$ is Ricci-flat and (anti)-self-dual, which is equivalent to locally hyper-Kähler for some orientation. The only examples are the flat torus, the K3 surface with the Yau metric (now $\tau = 16$ and $\chi = 24$), and two quotients of K3.

Since the mid-1990s, LeBrun (2003) has obtained a series of results which sharpen the Hitchin–Thorpe inequality by obtaining estimates on the Weyl and scalar curvature terms. These estimates are obtained by using Seiberg–Witten theory, the general theme being that nonemptiness of the Seiberg–Witten moduli space gives lower bounds on the curvature terms. LeBrun shows there are infinitely many compact smooth simply connected 4-manifolds that satisfy the Hitchin–Thorpe inequality but nonetheless do not admit Einstein metrics.

## Uniqueness and Moduli

In Yang–Mills theory, there is a highly developed theory of moduli spaces of instantons, including formulas for the dimension. The situation for Einstein metrics is far less well understood. The relevant moduli space here is the set of Einstein metrics modulo the action of the diffeomorphism group, but there are very few manifolds for which the moduli space has been determined. In dimension 2, of course, this is essentially the subject of the Teichmuller theory.

One example where the moduli space is understood is the K3 surface. As explained above, the Hitchin–Thorpe argument shows that any Einstein metric is hyper-Kähler, and the moduli space of such structures on K3 is understood as an open set in a certain noncompact symmetric space.

Some uniqueness results have been obtained in four dimensions. LeBrun used Seiberg–Witten techniques to show that the Einstein metric on a compact quotient of the complex hyperbolic plane $\mathbb{C}H^2$ is unique up to homotheties and diffeomorphisms. The analogous result for compact quotients of real hyperbolic 4-space was obtained using entropy methods by Besson, Courtois, and Gallot. It is still unknown, however, whether nonstandard Einstein metrics can exist on $S^4$.

In higher dimensions, very little is known. One can, by analogy with the theory of instantons, consider the linearization of the Einstein equations together with a further linear equation expressing orthogonality to the orbits of the diffeomorphism group. This gives a notion of formal tangent space to the Einstein moduli space. However, Koiso has shown that formal tangent vectors need not integrate to a curve of Einstein metrics. The structure of the moduli space (dimension, possible singularities) remains quite mysterious in general. It is known from the Wang–Ziller torus bundle examples that the moduli space can have infinitely many components.

## Special Holonomy

Berger classified the possible holonomy groups of simply connected, irreducible, nonsymmetric $n$-dimensional Riemannian manifolds. The generic case is that of holonomy $SO(n)$, and there are six other possibilities, each of which corresponds to some special geometry. Interestingly, four of these are automatically Ricci-flat, while a fifth is Einstein with $\Lambda \neq 0$. The remaining example, that of Kähler geometry, is not automatically Einstein, but the Einstein equations with the additional Kähler assumption reduce to a *scalar* Monge–Ampere equation and are therefore simpler than the general Einstein system.

For further reading in this section, see the articles by Boyer–Galicki, Joyce, Salamon, Tian, Yau and the author in part I of LeBrun and Wang (1999), and also the book of Joyce (2000). For the Kähler case, see also Tian (2000).

### **Kähler Manifolds** (Holonomy $U(n/2)$, $SU(n/2)$)

A Kähler manifold $(M, g)$ admits a covariant constant complex structure $I$, and associated Kähler 2-form $\omega$ defined by $\omega(X, Y) = g(IX, Y)$. The Ricci form $\rho$ is defined by $\rho(X, Y) = \mathrm{Ric}(IX, Y)$, so the Einstein condition for a Kähler manifold becomes

$$\rho = \Lambda\omega$$

On a Kähler manifold, $\rho$ is the curvature of the canonical bundle, so $[\rho/2\pi]$ is a representative for the cohomology class $c_1(M)$.

We see that a *necessary* condition for a complex manifold $(M, I)$ to admit a Kähler–Einstein metric is that $c_1$ has a definite sign. We consider, in turn, the three cases:

$c_1 < 0$

In this case, we have:

**Theorem** (Aubin, Yau). *Let $(M, I)$ be a compact complex manifold with $c_1 < 0$. Then $(M, I)$ admits a Kähler–Einstein metric with $\Lambda < 0$. The metric is unique up to homothety.*

$c_1 = 0$

This is a special case of the Calabi conjecture, proved by Yau.

**Theorem** (Yau). *Let $M$ be a compact Kähler manifold with Kähler form $\omega$. For any closed real form $\rho$ of type $(1,1)$ with $[\rho/2\pi] = c_1(M)$, there exists a unique Kähler metric with Kähler form cohomologous to $\omega$ and Ricci form equal to $\rho$.*

*In particular, if $M$ is a compact Kähler manifold with $c_1 = 0$, there exists a Ricci-flat Kähler metric on $M$.*

Ricci-flat Kähler metrics are called Calabi–Yau metrics, and are exactly the metrics with holonomy in $SU(n/2)$. They admit two parallel spinors and are of great interest to string theorists, because in some string theories spacetime is expected to be a product of the four-dimensional macroscopic factor with a compact Calabi–Yau manifold of complex dimension 3.

Yau's theorem provides many examples of Calabi–Yau spaces. For example, we can take a nonsingular complex submanifold defined as a complete intersection by the vanishing of $r$ polynomials of degree $d_1, \ldots, d_r$ in $\mathbb{CP}^n$. Now, $M$ has complex dimension $n - r$ and $c_1 = 0$ if and only if $n + 1 = \sum_{i=1}^{r} d_i$. We obtain examples of complex dimension 2 by considering a quartic in $\mathbb{CP}^3$, the intersection of a quadric and a cubic in $\mathbb{CP}^4$, or the intersection of three quadrics in $\mathbb{CP}^5$; these all give examples of K3 surfaces. A famous example of a Calabi–Yau manifold of complex dimension 3 is given by the quintic in $\mathbb{CP}^4$. This technique can be extended, for example, by considering complete intersections in weighted projective space or constructing Calabi–Yau desingularizations of singular spaces.

$c_1 > 0$

This case is the most complicated and, at the time of writing, is not yet fully understood. It is known that not every compact manifold with $c_1 > 0$ supports a Kähler–Einstein metric.

An early result of Matsushima was that the identity component of the automorphism group of a Kähler–Einstein space with $c_1 > 0$ must be reductive. This shows, for example, that the blow-up of $\mathbb{CP}^2$ at one or two points does not admit a Kähler–Einstein metric, despite having $c_1 > 0$. (The one-point blow-up does admit a Hermitian–Einstein metric due to Page.) A second obstruction is the Futaki invariant, a character of the Lie algebra of the automorphism group. This character vanishes if there is a Kähler–Einstein metric.

Both the above obstructions depend on having a nontrivial algebra of holomorphic automorphisms of $M$. More recently, Tian has discovered further obstructions (in complex dimension 3 or higher) which can be present even if the automorphism algebra is trivial.

However, for compact complex *surfaces* with $c_1 > 0$, Tian has proved that vanishing of the Futaki invariant is sufficient. In particular, the blow-up of $\mathbb{CP}^2$ at $k$ points in general position, where $3 \leq k \leq 8$, admits a Kähler–Einstein metric (note that $c_1^2 = 9 - k$ so if $k > 8$ then $c_1$ is no longer definite).

LeBrun–Catanese and Kotschick used these results to give an example of a topological 4-manifold carrying Einstein metrics of different signs. A deformation of the Barlow surface (a surface of general type) has $c_1 < 0$ and hence carries an Einstein metric with $\Lambda < 0$. But this space is homeomorphic (though not diffeomorphic) to the blow-up of $\mathbb{CP}^2$ at eight points, which carries an Einstein metric with $\Lambda > 0$. One may use this example to construct higher-dimensional examples of *diffeomorphic* manifolds carrying Einstein metrics of opposite sign.

### Hyper-Kähler Manifolds (Holonomy Sp($n$/4))

These are always Ricci-flat. They have a triple $(I, J, K)$ of covariant constant complex structures, satisfying the quaternionic multiplication relations $IJ = K = -JI$, etc., and defining Kähler forms $\omega_I, \omega_J, \omega_K$. Hyper-Kähler manifolds of dimension $n = 4N$ have $N + 1$ parallel spinors.

The most effective way of producing complete hyper-Kähler metrics has been the hyper-Kähler quotient construction (Hitchin *et al.* 1987), which was motivated by the Marsden–Weinstein quotient in symplectic geometry. Let $G$ be a group acting freely on a hyper-Kähler manifold $(M, g, I, J, K)$ preserving the hyper-Kähler structure. Subject to mild assumptions, we obtain a $G$-equivariant moment map $\mu : M \to \mathfrak{g}^* \otimes \mathbb{R}^3$, satisfying

$$d\mu_X(Y) = (\omega_I(X, Y), \omega_J(X, Y), \omega_K(X, Y))$$

Now the quotient $\mu^{-1}(0)/G$ is a hyper-Kähler manifold of dimension $\dim M - 4 \dim G$.

The power of this construction comes from the fact that even if $M$ is just flat quaternionic space, one can obtain highly nontrivial quotients by suitable choice of group $G$ (e.g., the asymptotically locally Euclidean four-dimensional examples of Kronheimer, which include as a subcase the multi-instanton metrics of Gibbons and Hawking).

Many examples of interest in mathematical physics may be obtained by taking hyper-Kähler quotients of an infinite-dimensional space of connections and Higgs fields (Hitchin 1987). Examples include moduli spaces of instantons over a hyper-Kähler base, moduli spaces of monopoles on $\mathbb{R}^3$, and moduli spaces of Higgs pairs over a Riemann surface.

The hyper-Kähler manifolds produced so far by the quotient construction have all been noncompact. Examples of compact hyper-Kähler manifolds are rarer but some are known. Beauville has produced examples in all dimensions as desingularizations of symmetric products of the basic four-dimensional compact examples (K3 and the 4-torus).

Further material for this section may be found, for example, in Hitchin (1992) and in the chapter by the author on hyper-Kähler manifolds in LeBrun and Wang (1999).

### Quaternionic Kähler Manifolds (Holonomy Sp($n$/4)) Sp(1))

These are always Einstein with nonzero Einstein constant. Instead of globally defined parallel complex structures as in the hyper-Kähler case, we have a sub-bundle $\mathcal{G}$ of End($TM$) with fiber isomorphic to the imaginary quaternions, parallel with respect to the Levi-Civita connection. Thus, we have locally defined almost-complex structures $I$, $J$, $K$, satisfying the quaternionic multiplication relations, such that covariant differentiation of one of $I$, $J$, $K$ gives a linear combination of the other two. In particular, note that quaternionic Kähler manifolds are not Kähler.

If the Einstein constant $\Lambda$ is positive, the only known complete examples are symmetric, the so-called compact Wolf spaces, which are in one-to-one correspondence with the compact simple Lie groups. It is conjectured that these are the only examples with $\Lambda > 0$, and some results in this direction have been established (e.g., it is known if $\dim M \leq 12$). It is also known that for fixed dimension, there are only finitely many types of compact quaternionic Kähler manifold with $\Lambda > 0$.

Many *orbifold* examples, however, are known to exist, for example, via the Galicki–Lawson quaternionic Kähler quotient construction.

If $\Lambda < 0$, more complete examples are known. In addition to the noncompact duals of the Wolf spaces, there are homogeneous, nonsymmetric examples due to Alekseevski, and infinite-dimensional families of inhomogeneous examples constructed via twistor methods by LeBrun (see also Biquard (2000)).

### Exceptional Holonomy ($G_2$ or Spin(7))

Such metrics exist in dimension 7 or 8, respectively. They are always Ricci-flat and admit a parallel spinor. Local examples were constructed by Bryant using Cartan–Kähler theory, and some explicit complete noncompact examples were produced by Salamon and Bryant using a cohomogeneity-1 construction. More complicated explicit noncompact examples have recently been produced by several authors (see Cvetič *et al*. (2003) for a survey). Compact examples were produced using analytical methods by Joyce, and later by Kovalev. Joyce starts with a flat singular metric on quotients of the seven- or eight-dimensional torus and constructs an approximate solution to the special holonomy condition on a resolution of this singular space. Then an analytic argument is used to show that an exact nearby solution exists.

For further reading, consult Joyce (2000) as well as the article by Joyce in LeBrun and Wang (1999).

There are also some interesting examples of Einstein metrics which, although not of special holonomy themselves, are closely related to special holonomy geometries. In recent years, these have yielded many new examples of compact Einstein manifolds in the work of Boyer, Mann, Galicki, Kollar, Rees, Piccinni, and Nakamaye.

### Einstein–Sasaki Structures

There are several different ways of defining these, but the simplest is to say that $(M, g)$ is Einstein–Sasaki if the cone $(\mathbb{R} \times M, \mathrm{d}t^2 + t^2 g)$ is Ricci-flat Kähler. Also, an Einstein–Sasaki manifold has a circle action with quotient a Kähler–Einstein orbifold. Existence theorems for such orbifold metrics have led to many examples of Einstein–Sasaki metrics, including families on odd-dimensional spheres.

### 3-Sasakian Structures

Again, we can define these in terms of cones; $(M, g)$ has a 3-Sasakian structure if the cone over it is hyper-Kähler. The basic example is $S^{4n+3}$ with associated cone $\mathbb{H}^n - \{0\}$. A 3-Sasakian manifold is always Einstein with positive Einstein constant.

The hyper-Kähler quotient construction induces a 3-Sasakian quotient, and many examples of compact 3-Sasakian manifolds have been produced as 3-Sasakian quotients of $S^{4n+3}$. In particular, there are examples in dimension 7 with arbitrarily large second Betti number, showing that one cannot, in general, expect compactness/finiteness results for Einstein moduli spaces without further assumptions.

## Homogeneous Examples

Another strategy to study the Einstein equations is to reduce the difficulty of the problem by imposing symmetries. More precisely, we consider Einstein manifolds $(M, g)$ with an isometric action of a Lie group $G$. In general, the Einstein equations with this symmetry will now involve $r$ independent variables where $r$ is the dimension of the stratified space $M/G$. We call $r$ the *cohomogeneity* of the manifold.

In this section, we consider the situation where $(M, g)$ is *homogeneous,* that is, when the action of $G$ is transitive so $r = 0$. The Einstein equations now reduce to a system of *algebraic* equations.

We may now write $M = G/K$, where $K$ is the stabilizer of a point of $M$. We choose an $\mathrm{Ad}_K$-invariant vector space complement $\mathfrak{p}$ to $\mathfrak{k}$ in $\mathfrak{g}$, and identify $\mathfrak{p}$ with the tangent space to $G/K$ at the identity coset. The key point is that $G$-invariant metrics on $M = G/K$ may now be identified with $\mathrm{Ad}_K$-invariant inner products on $\mathfrak{p}$, which may, in turn, be studied by looking at the decomposition of $\mathfrak{p}$ into irreducible representations of $K$.

In the special case when $G/K$ is isotropy irreducible (i.e., $\mathfrak{p}$ is an irreducible representation of $K$), both the metric $g$ and its Ricci tensor are proportional by Schur's lemma, and hence $g$ is automatically Einstein. Isotropy-irreducible homogeneous spaces have been classified by Kramer, Manturov, Wolf, and Wang–Ziller.

In the general case, the Einstein equations become a system of polynomial equations. Determining whether this system has a real positive solution is, in general, a highly nontrivial problem. However, the situation of homogeneous metrics is one area in which the variational formulation of the Einstein equations has proved highly successful.

We are now considering the scalar curvature functional on the finite-dimensional space of unit $G$-invariant metrics on $G/K$. The behavior of the scalar curvature functional is related to the structure of the lattice of intermediate subalgebras between the Lie algebras of $K$ and $G$.

An early result along these lines (Wang and Ziller 1986) is that if $K$ is maximal in $G$ (compact), then $G/K$ admits a $G$-invariant Einstein metric. The idea of the proof is to show that maximality of $K$ forces the scalar curvature functional on the space of volume-1 homogeneous metrics to be both bounded above and proper, and therefore to have a maximum.

These ideas have been greatly extended by Böhm, Wang, and Ziller. Given a compact connected homogeneous space $G/K$, they define a graph whose vertices are $\mathrm{Ad}(K)$-invariant subalgebras strictly intermediate between $\mathfrak{g}$ and $\mathfrak{k}$. The edges correspond to inclusions between subalgebras. A component of the graph is called toral if all subalgebras $\mathfrak{h}$ in this component are such that the identity component of $H/K$ is abelian. They now show that if the graph has at least two nontoral components, then $G/K$ admits a $G$-invariant Einstein metric. The Einstein metrics in the theorem are produced by a mountain pass argument and may have co-index 1, contrasting with the maxima of the earlier theorem.

Further advances in this direction have recently been made by Böhm. He associates to $G/K$ a simplicial complex, and shows that nonzero homology groups of the complex imply the existence of higher co-index Einstein metrics.

One can also study homogeneous noncompact Einstein spaces with $\Lambda < 0$. It is conjectured by Alekseevski that for all such examples $K$ is a maximal compact subgroup of $G$. The reader is referred to Heber (1998) for further information on the noncompact case.

The above results give some powerful existence results for Einstein metrics. However, there are examples known of homogeneous spaces $G/K$ which admit no $G$-invariant Einstein metric (Wang and Ziller 1986). One such example is $\mathrm{SU}(4)/\mathrm{SU}(2)$, where $\mathrm{SU}(2)$ is a maximal subgroup of $\mathrm{Sp}(2) \subset \mathrm{SU}(4)$.

Techniques similar to those in the homogeneous case have been used to construct Einstein metrics on total spaces of certain bundles, via Riemannian submersions. Some highlights are Jensen's exotic Einstein metrics on $(4n + 3)$-dimensional spheres, and the Wang–Ziller metrics on total spaces of torus bundles over products of Kähler–Einstein manifolds. The latter construction gives examples of spaces admitting volume-1 Einstein metrics with infinitely many Einstein constants $\Lambda$.

## Examples of Higher Cohomogeneity

One can also look for Einstein metrics of higher cohomogeneity. Most progress has been made in the cohomogeneity-1 case, that is, where the principal orbit $G/K$ of the action has real codimension one in

$M$ (see Eschenburg and Wang (2000) for background on such metrics). On the open dense set in $M$ which is the union of the principal orbits, we may write the metric as

$$dt^2 + g_t$$

where $g_t$ is a $t$-dependent homogeneous metric on $G/K$. The Einstein equations are now a system of ordinary differential equations in $t$.

One may also add a special orbit $G/H$ at one or both ends of the interval over which $t$ ranges. This will impose boundary conditions on the ODEs. For the manifold structure to extend smoothly over the special orbit, $H/K$ must be a sphere. Notice that if $\Lambda > 0$, then to obtain a complete metric $M$ must be compact, so we must add two special orbits. If $\Lambda \leq 0$ and the metric is irreducible, then a Bochner argument tells us that $M$ is noncompact. In the Ricci-flat case, the Cheeger–Gromoll theorem tells us that to obtain a complete irreducible metric, we must have exactly one special orbit, so $M$ is topologically the total space of a vector bundle over the special orbit. In fact, most of the known examples even with $\Lambda < 0$ have a special orbit too.

The system of ODEs we obtain is still highly nonlinear and difficult to analyze in general. However, there are certain situations in which the equations, or a subsystem, can be solved in closed form. If we take $G/K$ to be a principal circle bundle over a Hermitian symmetric space, Bérard Bergery (1982) showed that the resulting Einstein equations are solvable. (His work was inspired by the earlier example of Page, which corresponds to the case when $G/K = U(2)/U(1)$, a circle bundle over $\mathbb{CP}^1$.) In fact, Bérard Bergery's construction works in greater generality as we obtain the same equations if $G/K$ is replaced by any Riemannian submersion with circle fibers over a positive Kähler–Einstein space. This illustrates a general principle that systems arising as cohomogeneity-1 Einstein equations also typically arise from certain bundle ansätze without homogeneity assumptions.

Wang and Wang generalized this construction to be the case when the hypersurface in $M$ is a Riemannian submersion with circle fibers over a product of an arbitrary number of Kähler–Einstein factors. Other solvable Einstein systems have been studied by, for example, Wang and Dancer.

It may also be possible in certain situations to get existence results without an explicit solution. This observation underlies the important work of Böhm (1998). He constructs cohomogeneity-1 Einstein metrics on certain manifolds with dimension between 5 and 9, including all the spheres in this range of dimensions. The equations are not now solved in closed form, but it is possible to get a qualitative understanding of the flow and to show that certain trajectories will give metrics on the desired compact manifolds.

Böhm has also shown, in an analogous result to the homogeneous case, that there are examples of manifolds with a cohomogeneity-1 $G$-action which do not support any $G$-invariant Einstein metric.

So far, not much is known about Einstein metrics of higher cohomogeneity. An exception is the situation of self-dual Einstein metrics in dimension 4, where the self-dual condition greatly simplifies the resulting equations. Calderbank, Pedersen, and Singer have achieved a good understanding of such metrics with $T^2$ symmetry, including construction of such metrics on Hirzebruch–Jung resolutions of cyclic quotient singularities.

## Analytical Methods

So far there is no really general analytical method for proving existence of global Riemannian Einstein metrics (although, of course, such techniques do exist in more restrictive situations of special holonomy).

Although the Einstein equations admit a variational formulation, this has (except for homogeneous metrics) not yielded general existence results. Note that the Wang–Ziller torus bundle examples at the end of the section "Homogeneous examples" show that the Palais–Smale condition does not hold in full generality.

One early suggestion was to adopt a minimax procedure. In each conformal class $[g]$, one looks for a minimizer of the volume-normalized scalar curvature. Such a minimizer always exists. One then takes the supremum over all conformal classes. The resulting supremum of the functional is called the Yamabe invariant $Y(M)$ of the manifold $M$. If a maximizer $g$ exists, and $Y(M) \leq 0$, then $g$ is Einstein.

However, striking work of Petean shows that this procedure must fail to produce an Einstein metric in many cases. He proves that if $\dim M \geq 5$ and $M$ is simply connected, then the Yamabe invariant is nonnegative. So, for such an $M$, any Einstein metric produced will have $\Lambda \geq 0$, and we know that this puts constraints on the topology of $M$.

Another possible technique is to use the Hamilton Ricci flow. If this converges as $t \to \infty$, the limiting metric is Einstein. However, it seems hard in higher dimensions to get control over the flow. In particular, the Wang–Ziller example in the section "Homogeneous examples" of a homogeneous space with no invariant Einstein metric shows that the flow may fail to converge (the Hamilton flow preserves the property of $G$-invariance).

Graham–Lee and Biquard have used analytical methods to produce Einstein deformations of hyperbolic space (real, complex, quaternionic, or Cayley). The idea is to show that a sufficiently small deformation of the conformal infinity of hyperbolic space can be extended to a deformation of the hyperbolic metric.

Recently, Anderson has shown the existence of Einstein metrics with $\Lambda < 0$ on a large class of manifolds obtained by Dehn filling from hyperbolic manifolds with toral ends. The strategy is to glue on to the hyperbolic metric copies of a simple explicit asymptotically hyperbolic metric, and to show that the resulting metric can be perturbed to an exact solution of the Einstein equations.

*See also:* Einstein Equations: Exact Solutions; Einstein Equations: Initial Value Formulation; Hamiltonian Reduction of Einstein's Equations; Several Complex Variables: Compact Manifolds; Singularities of the Ricci Flow.

## Further Reading


Bérard Bergery L (1982) Sur des nouvelles variétés riemanniennes d'Einstein. Publications de l'Institut Elie Cartan (Nancy), vol. 6, pp. 1–60.

Besse A (1987) *Einstein Manifolds*. Berlin: Springer.

Biquard O (2000) Métriques d'Einstein asymptotiquement symétriques. *Astérisque* 265.

Böhm C (1998) Inhomogeneous Einstein manifolds on low-dimensional spheres and other low-dimensional spaces. *Inventiones Mathematicae* 134: 145–176.

Cvetič M, Gibbons GW, Lü H, and Pope CN (2003) Special holonomy spaces and M-theory. In: *Unity from Duality, Gravity, Gauge Theory and Strings (Les Houches, 2001)*, 523–545. NATO Advanced Study Institute, EDP Sci, Les Ulis.

Eschenburg J and Wang M (2000) The initial value problem for cohomogeneity one Einstein metrics. *Journal of Geometrical Analysis* 10: 109–137.

Hawking SW and Ellis GFR (1973) *The Large-Scale Structure of Space-Time*. Cambridge: Cambridge University Press.

Heber J (1998) Noncompact homogeneous Einstein spaces. *Inventiones Mathematicae* 133: 279–352.

Hitchin NJ (1974) Compact four-dimensional Einstein manifolds. *Journal of Differential Geometry* 9: 435–441.

Hitchin NJ (1987) *Monopoles, Minimal Surfaces and Algebraic Curves*. Les presses de l'Université de Montreal.

Hitchin NJ (1992) Hyper-Kähler manifolds. *Asterisque* 206: 137–166.

Hitchin NJ, Karlhede A, Lindström U, and Roček M (1987) Hyper-Kähler metrics and supersymmetry. *Communications in Mathematical Physics* 108: 535–589.

Joyce D (2000) *Compact Manifolds with Special Holonomy*. Oxford: Oxford University Press.

LeBrun C (2003) Einstein metrics, four-manifolds and differential topology. In: Yau S-T (ed.) *Surveys in Differential Geometry*, vol. VIII, pp. 235–255. Somerville, MA: International Press.

LeBrun C and Wang M (eds.) (1999) *Essays on Einstein Manifolds*, vol. VI, *Surveys in Differential Geometry*. Boston, MA: International Press.

Page D (1979) A compact rotating gravitational instanton. *Physics Letters* 79B: 235–238.

Tian G (2000) *Canonical Metrics in Kähler Geometry*, Lectures in Mathematics, ETH Zurich. Birkhäuser.

Wang M and Ziller W (1986) Existence and nonexistence of homogeneous Einstein metrics. *Invent. Math* 84: 177–194.

Wang M and Ziller W (1990) Einstein metrics on principal torus bundles. *Journal of Differential Geometry* 31: 215–248.

Yau S-T (1978) On the Ricci curvature of a compact Kähler manifold and the complex Monge–Ampere equations I. *Communications in Pure and Applied Mathematics* 31: 339–411.


# Einstein–Cartan Theory

**A Trautman**, Warsaw University, Warsaw, Poland



## Introduction

### Notation

Standard notation and terminology of differential geometry and general relativity are used in this article. All considerations are local, so that the four-dimensional spacetime $M$ is assumed to be a smooth manifold diffeomorphic to $\mathbb{R}^4$. It is endowed with a metric tensor $g$ of signature $(1,3)$ and a linear connection defining the covariant differentiation of tensor fields. Greek indices range from 0 to 3 and refer to spacetime. Given a field of frames $(e_\mu)$ on $M$, and the dual field of coframes $(\theta^\mu)$, one can write the metric tensor as $g = g_{\mu\nu}\theta^\mu\theta^\nu$, where $g_{\mu\nu} = g(e_\mu, e_\nu)$ and Einstein's summation convention is assumed to hold. Tensor indices are lowered with $g_{\mu\nu}$ and raised with its inverse $g^{\mu\nu}$. General-relativistic units are used, so that both Newton's constant of gravitation and the speed of light are 1. This implies $\hbar = l^2$, where $l \approx 10^{-33}$ cm is the Planck length. Both mass and energy are measured in centimeters.

### Historical Remarks

The Einstein–Cartan theory (ECT) of gravity is a modification of general relativity theory (GRT), allowing spacetime to have torsion, in addition to curvature, and relating torsion to the density of intrinsic angular momentum. This modification was put forward in 1922 by Élie Cartan, before the discovery of spin. Cartan was influenced by the work of the Cosserat brothers (1909), who considered besides an (asymmetric) force stress tensor also a moments stress tensor in a suitably

generalized continuous medium. Work done in the 1950s by physicists (Kondo, Bilby, Kröner, and other authors) established the role played by torsion in the continuum theory of crystal dislocations. A recent review (Ruggiero and Tartaglia 2003) describes the links between ECT and the classical theory of defects in an elastic medium.

Cartan assumed the linear connection to be metric and derived, from a variational principle, a set of gravitational field equations. He required, without justification, that the covariant divergence of the energy–momentum tensor be zero; this led to an algebraic constraint equation, bilinear in curvature and torsion, severely restricting the geometry. This misguided observation has probably discouraged Cartan from pursuing his theory. It is now known that conservation laws in relativistic theories of gravitation follow from the Bianchi identities and, in the presence of torsion, the divergence of the energy–momentum tensor need not vanish. Torsion is implicit in the 1928 Einstein theory of gravitation with teleparallelism. For a long time, Cartan's modified theory of gravity, presented in his rather abstruse notation, unfamiliar to physicists, did not attract any attention. In the late 1950s, the theory of gravitation with spin and torsion was independently rediscovered by Sciama and Kibble. The role of Cartan was recognized soon afterward and ECT became the subject of much research; see Hehl *et al.* (1976) for a review and an extensive bibliography. In the 1970s, it was recognized that ECT can be incorporated within supergravity. In fact, simple supergravity is equivalent to ECT with a massless, anticommuting Rarita–Schwinger field as the source. Choquet–Bruhat considered a generalization of ECT to higher dimensions and showed that the Cauchy problem for the coupled system of Einstein–Cartan and Dirac equations is well posed. Penrose (1982) has shown that torsion appears in a natural way when spinors are allowed to be rescaled by a complex conformal factor. ECT has been generalized by allowing nonmetric linear connections and additional currents, associated with dilation and shear, as sources of such a "metric-affine theory of gravity" (Hehl *et al.* 1995).

### Physical Motivation

Recall that, in special relativity theory (SRT), the underlying Minkowski spacetime admits, as its group of automorphisms, the full Poincaré group, consisting of translations and Lorentz transformations. It follows from the first Noether theorem that classical, special-relativistic field equations, derived from a variational principle, give rise to conservation laws of energy–momentum and angular momentum. Using Cartesian coordinates $(x^\mu)$, abbreviating $\partial\varphi/\partial x^\rho$ to $\varphi_{,\rho}$ and denoting by $t^{\mu\nu}$ and $s^{\mu\nu\rho} = -s^{\nu\mu\rho}$ the tensors of energy–momentum and of intrinsic angular momentum (spin), respectively, one can write the conservation laws in the form

$$t^{\mu\nu}{}_{,\nu} = 0 \qquad [1]$$

and

$$(x^\mu t^{\nu\rho} - x^\nu t^{\mu\rho} + s^{\mu\nu\rho})_{,\rho} = 0 \qquad [2]$$

In the presence of spin, the tensor $t^{\mu\nu}$ need not be symmetric,

$$t^{\mu\nu} - t^{\nu\mu} = s^{\mu\nu\rho}{}_{,\rho}$$

Belinfante and Rosenfeld have shown that the tensor

$$T^{\mu\nu} = t^{\mu\nu} + \tfrac{1}{2}(s^{\nu\mu\rho} + s^{\nu\rho\mu} + s^{\mu\rho\nu})_{,\rho}$$

is symmetric and its divergence vanishes.

In quantum theory, the irreducible, unitary representations of the Poincaré group correspond to elementary systems such as stable particles; these representations are labeled by the mass and spin.

In Einstein's GRT, the spacetime $M$ is curved; the Lorentz group – but not the Poincaré group – appears as the structure group acting on orthonormal frames in the tangent spaces of $M$. The energy–momentum tensor $T$ appearing on the right-hand side of the Einstein equation is necessarily symmetric. In GRT there is no room for translations and the tensors $t$ and $s$.

By introducing torsion and relating it to $s$, Cartan restored the role of the Poincaré group in relativistic gravity: this group acts on the affine frames in the tangent spaces of $M$. Curvature and torsion are the surface densities of Lorentz transformations and translations, respectively. In a space with torsion, the Ricci tensor need not be symmetric so that an asymmetric energy–momentum tensor can appear on the right-hand side of the Einstein equation.

## Geometric Preliminaries

### Tensor-Valued Differential Forms

It is convenient to follow Cartan in describing geometric objects as tensor-valued differential forms. To define them, consider a homomorphism $\sigma : \mathrm{GL}_4(\mathbb{R}) \to \mathrm{GL}_N(\mathbb{R})$ and an element $A = (A^\mu_\nu)$ of $\mathrm{End}\,\mathbb{R}^4$, the Lie algebra of $\mathrm{GL}_4(\mathbb{R})$. The derived representation of Lie algebras is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\sigma(\exp At)|_{t=0} = \sigma^\nu_\mu A^\mu_\nu$$

If $(e_a)$ is a frame in $\mathbb{R}^N$, then $\sigma^\nu_\mu(e_a)=\sigma^{b\nu}_{a\mu}e_b$, where $a,b=1,\ldots,N$.

A map $a=(a^\mu{}_\nu):M\to \mathrm{GL}_4(\mathbb{R})$ transforms fields of frames so that

$$e'_\mu = e_\nu a^\nu_\mu \quad \text{and} \quad \theta^\nu = a^\nu_\mu \theta'^\mu \qquad [3]$$

A differential form $\varphi$ on $M$, with values in $\mathbb{R}^N$, is said to be of type $\sigma$ if, under changes of frames, it transforms so that $\varphi'=\sigma(a^{-1})\varphi$. For example, $\theta=(\theta^\mu)$ is a 1-form of type id. If now $A=(A^\mu_\nu):M\to \mathrm{End}\,\mathbb{R}^4$, then one puts $a(t)=\exp tA:M\to \mathrm{GL}_4(\mathbb{R})$ and defines the variations induced by an infinitesimal change of frames,

$$\delta\theta = \frac{\mathrm{d}}{\mathrm{d}t}(a(t)^{-1}\theta)|_{t=0} = -A\theta$$
$$\delta\varphi = \frac{\mathrm{d}}{\mathrm{d}t}(\sigma(a(t)^{-1})\varphi)|_{t=0} = -\sigma^\nu_\mu A^\mu_\nu \varphi \qquad [4]$$

## Hodge Duals

Since $M$ is diffeomorphic to $\mathbb{R}^4$, one can choose an orientation on $M$ and restrict the frames to agree with that orientation so that only transformations with values in $\mathrm{GL}^+_4(\mathbb{R})$ are allowed. The metric then defines the Hodge dual of differential forms. Put $\theta_\mu=g_{\mu\nu}\theta^\nu$. The forms $\eta,\eta_\mu,\eta_{\mu\nu},\eta_{\mu\nu\rho}$, and $\eta_{\mu\nu\rho\sigma}$ are defined to be the duals of $1,\theta_\mu,\theta_\mu\wedge\theta_\nu,\theta_\mu\wedge\theta_\nu\wedge\theta_\rho$, and $\theta_\mu\wedge\theta_\nu\wedge\theta_\rho\wedge\theta_\sigma$, respectively. The 4-form $\eta$ is the volume element; for a holonomic coframe $\theta^\mu=\mathrm{d}x^\mu$, it is given by $\sqrt{-\det(g_{\mu\nu})}\mathrm{d}x^0\wedge\mathrm{d}x^1\wedge\mathrm{d}x^2\wedge\mathrm{d}x^3$. In SRT, in Cartesian coordinates, one can define the tensor-valued 3-forms

$$t^\mu = t^{\mu\nu}\eta_\nu \quad \text{and} \quad s^{\mu\nu} = s^{\mu\nu\rho}\eta_\rho \qquad [5]$$

so that eqns [1] and [2] become

$$\mathrm{d}t^\mu = 0 \quad \text{and} \quad \mathrm{d}j^{\mu\nu} = 0$$

where

$$j^{\mu\nu} = x^\mu t^\nu - x^\nu t^\mu + s^{\mu\nu} \qquad [6]$$

For an isolated system, the 3-forms $t^\mu$ and $j^{\mu\nu}$, integrated over the 3-space $x^0=\text{const.}$, give the system's total energy–momentum vector and angular momentum bivector, respectively.

## Linear Connection, Its Curvature and Torsion

A linear connection on $M$ is represented, with respect to the field of frames, by the field of 1-forms

$$\omega^\mu_\nu = \Gamma^\mu_{\rho\nu}\theta^\rho$$

so that the covariant derivative of $e_\nu$ in the direction of $e_\mu$ is $\nabla_\mu e_\nu=\Gamma^\rho_{\mu\nu}e_\rho$. Under a change of frames [3], the connection forms transform as follows:

$$a^\mu_\rho \omega'^\rho_\nu = \omega^\mu_\rho a^\rho_\nu + \mathrm{d}a^\mu_\nu$$

If $\varphi=\varphi^a e_a$ is a $k$-form of type $\sigma$, then its covariant exterior derivative

$$D\varphi^a = \mathrm{d}\varphi^a + \sigma^{a\mu}_{b\nu}\omega^\nu_\mu\wedge\varphi^b$$

is a $(k+1)$-form of the same type. For a 0-form one has $D\varphi^a=\theta^\mu\nabla_\mu\varphi^a$. The infinitesimal change of $\omega$, defined similarly as in [4], is $\delta\omega^\mu_\nu=DA^\mu_\nu$. The 2-form of curvature $\Omega=(\Omega^\mu{}_\nu)$, where

$$\Omega^\mu_\nu = \mathrm{d}\omega^\mu_\nu + \omega^\mu_\rho\wedge\omega^\rho_\nu$$

is of type ad: it transforms with the adjoint representation of $\mathrm{GL}_4(\mathbb{R})$ in $\mathrm{End}\,\mathbb{R}^4$. The 2-form of torsion $\Theta=(\Theta^\mu)$, where

$$\Theta^\mu = \mathrm{d}\theta^\mu + \omega^\mu_\nu\wedge\theta^\nu$$

is of type id. These forms satisfy the Bianchi identities

$$D\Omega^\mu{}_\nu = 0 \quad \text{and} \quad D\Theta^\mu = \Omega^\mu{}_\nu\wedge\theta^\nu$$

For a differential form $\varphi$ of type $\sigma$, the following identity holds:

$$D^2\varphi^a = \sigma^{a\nu}_{b\mu}\Omega^\mu{}_\nu\wedge\varphi^b \qquad [7]$$

The tensors of curvature and torsion are given by

$$\Omega^\mu{}_\nu = \tfrac{1}{2}R^\mu{}_{\nu\rho\sigma}\theta^\rho\wedge\theta^\sigma$$

and

$$\Theta^\mu = \tfrac{1}{2}Q^\mu{}_{\rho\sigma}\theta^\rho\wedge\theta^\sigma$$

respectively. With respect to a holonomic frame, $\mathrm{d}\theta^\mu=0$, one has

$$Q^\mu{}_{\rho\sigma} = \Gamma^\mu_{\rho\sigma} - \Gamma^\mu_{\sigma\rho}$$

In SRT, the Cartesian coordinates define a radius-vector field $X^\mu=-x^\mu$, pointing towards the origin of the coordinate system. The differential equation it satisfies generalizes to a manifold with a linear connection:

$$DX^\mu + \theta^\mu = 0 \qquad [8]$$

By virtue of [7], the integrability condition of [8] is

$$\Omega^\mu{}_\nu X^\nu + \Theta^\mu = 0$$

Integration of [8] along a curve defines the Cartan displacement of $X$; if this is done along a small closed circuit spanned by the bivector $\Delta f$, then the radius vector changes by about

$$\Delta X^\mu = \tfrac{1}{2}(R^\mu{}_{\nu\rho\sigma}X^\nu + Q^\mu{}_{\rho\sigma})\Delta f^{\rho\sigma}$$

This holonomy theorem – rather imprecisely formulated here – shows that torsion bears to translations a relation similar to that of curvature to linear homogeneous transformations.

In a space with torsion, it matters whether one considers the potential of the electromagnetic field to be a scalar-valued 1-form $\varphi$ or a covector-valued 0-form $(\varphi_\mu)$. The first choice leads to a field $\mathrm{d}\varphi$ that is invariant with respect to the gauge transformation $\varphi \mapsto \varphi + \mathrm{d}\chi$. The second gives $\frac{1}{2}(\nabla_\mu\varphi_\nu - \nabla_\nu\varphi_\mu)\theta^\mu \wedge \theta^\nu = (D\varphi_\mu) \wedge \theta^\mu = \mathrm{d}\varphi - \varphi_\mu\Theta^\mu$, a gauge-dependent field.

### Metric-Affine Geometry

A metric-affine space $(M, g, \omega)$ is defined to have a metric and a linear connection that need not depend on each other. The metric alone determines the torsion-free Levi-Civita connection $\overset{\circ}{\omega}$ characterized by

$$\mathrm{d}\theta^\mu + \overset{\circ}{\omega}{}^\mu_\nu \wedge \theta^\nu = 0 \quad \text{and} \quad \overset{\circ}{D} g_{\mu\nu} = 0$$

Its curvature is

$$\overset{\circ}{\Omega}{}^\mu_\nu = \mathrm{d}\overset{\circ}{\omega}{}^\mu_\nu + \overset{\circ}{\omega}{}^\mu_\rho \wedge \overset{\circ}{\omega}{}^\rho_\nu$$

The 1-form of type ad,

$$\kappa^\mu{}_\nu = \omega^\mu_\nu - \overset{\circ}{\omega}{}^\mu_\nu \qquad [9]$$

determines the torsion of $\omega$ and the covariant derivative of $g$,

$$\Theta^\mu = \kappa^\mu{}_\nu \wedge \theta^\nu, \qquad Dg_{\mu\nu} = -\kappa_{\mu\nu} - \kappa_{\nu\mu}$$

The curvature of $\omega$ can be written as

$$\Omega^\mu{}_\nu = \overset{\circ}{\Omega}{}^\mu_\nu + \overset{\circ}{D}\kappa^\mu{}_\nu + \kappa^\mu{}_\rho \wedge \kappa^\rho{}_\nu \qquad [10]$$

The transposed connection $\tilde{\omega}$ is defined by

$$\tilde{\omega}^\mu_\nu = \omega^\mu_\nu + Q^\mu{}_{\nu\rho}\theta^\rho$$

so that, with respect to a holonomic frame, one has $\tilde{\Gamma}^\mu_{\nu\rho} = \Gamma^\mu_{\rho\nu}$. The torsion of $\tilde{\omega}$ is opposed to that of $\omega$.

### Riemann–Cartan Geometry

A Riemann–Cartan space is a metric-affine space with a connection that is metric,

$$Dg_{\mu\nu} = 0 \qquad [11]$$

The metricity condition implies that $\kappa_{\mu\nu} + \kappa_{\nu\mu} = 0$ and $\Omega_{\mu\nu} + \Omega_{\nu\mu} = 0$. In a Riemann–Cartan space, the connection is determined by its torsion $Q$ and the metric tensor. Let $Q_{\rho\mu\nu} = g_{\rho\sigma}Q^\sigma{}_{\mu\nu}$; then

$$\kappa_{\mu\nu} = \tfrac{1}{2}(Q_{\mu\sigma\nu} + Q_{\nu\mu\sigma} + Q_{\sigma\mu\nu})\theta^\sigma \qquad [12]$$

The transposed connection of a Riemann–Cartan space is metric if and only if the tensor $Q_{\rho\mu\nu}$ is completely antisymmetric. Let $\tilde{\nabla}$ denote the covariant derivative with respect to $\tilde{\omega}$. By definition, a symmetry of a Riemann–Cartan space is a diffeomorphism of $M$ preserving both $g$ and $\omega$. The one-parameter group of local transformations of $M$, generated by the vector field $v$, consists of symmetries of $(M, g, \omega)$ if and only if

$$\tilde{\nabla}^\mu v^\nu + \tilde{\nabla}^\nu v^\mu = 0 \qquad [13]$$

and

$$D\tilde{\nabla}_\nu v^\mu + R^\mu{}_{\nu\rho\sigma}v^\rho\theta^\sigma = 0 \qquad [14]$$

In a Riemannian space, the connections $\omega$ and $\tilde{\omega}$ coincide and [14] is a consequence of the Killing equation [13]. The metricity condition implies

$$D\eta_{\mu\nu\rho} = \eta_{\mu\nu\rho\sigma}\Theta^\sigma \qquad [15]$$

## The Einstein–Cartan Theory of Gravitation

### An Identity Resulting from Local Invariance

Let $(M, g, \omega)$ be a metric-affine spacetime. Consider a Lagrangian $L$ which is an invariant 4-form on $M$; it depends on $g, \theta, \omega, \varphi$, and the first derivatives of $\varphi = \varphi^a e_a$. The general variation of the Lagrangian is

$$\delta L = L_a \wedge \delta\varphi^a + \tfrac{1}{2}\tau^{\mu\nu}\delta g_{\mu\nu} + \delta\theta^\mu \wedge t_\mu$$
$$- \tfrac{1}{2}\delta\omega^\mu_\nu \wedge s^\nu{}_\mu + \text{an exact form} \qquad [16]$$

so that $L_a = 0$ is the Euler–Lagrange equation for $\varphi$. If the changes of the functions $g, \theta, \omega$, and $\varphi$ are induced by an infinitesimal change of the frames [4], then $\delta L = 0$ and [16] gives the identity

$$g_{\mu\rho}\tau^{\rho\nu} - \theta^\nu \wedge t_\mu + \tfrac{1}{2}Ds^\nu{}_\mu - \sigma^{b\nu}_{a\mu}L_a \wedge \varphi^b = 0$$

It follows from the identity that the two sets of Euler–Lagrange equations obtained by varying $L$ with respect to the triples $(\varphi, \theta, \omega)$ and $(\varphi, g, \omega)$ are equivalent. In the sequel, the first triple is chosen to derive the field equations.

### Projective Transformations and the Metricity Condition

Still under the assumption that $(M, g, \omega)$ is a metric-affine spacetime, consider the 4-form

$$8\pi K = \tfrac{1}{2}g^{\nu\rho}\eta_{\mu\rho} \wedge \Omega^\mu{}_\nu \qquad [17]$$

which is equal to $\eta R$, where $R = g^{\mu\nu}R_{\mu\nu}$ is the Ricci scalar; the Ricci tensor $R_{\mu\nu} = R^\rho_{\mu\rho\nu}$ is, in general, asymmetric. The form [17] is invariant with

respect to projective transformations of the connection,

$$\omega^{\mu}_{\nu} \mapsto \omega^{\mu}_{\nu} + \delta^{\mu}_{\nu}\lambda \qquad [18]$$

where $\lambda$ is an arbitrary 1-form. Projectively related connections have the same (unparametrized) geodesics. If the total Lagrangian for gravitation interacting with the matter field $\varphi$ is $K + L$, then the field equations, obtained by varying it with respect to $\varphi, \theta$, and $\omega$ are: $L_a = 0$,

$$\tfrac{1}{2} g^{\rho\sigma}\eta_{\mu\nu\rho} \wedge \Omega^{\nu}{}_{\sigma} = -8\pi t_{\mu} \qquad [19]$$

and

$$D(g^{\mu\rho}\eta_{\rho\nu}) = 8\pi s^{\mu}{}_{\nu} \qquad [20]$$

respectively. Put $s_{\mu\nu} = g_{\mu\rho}s^{\rho}{}_{\nu}$. If

$$s_{\mu\nu} + s_{\nu\mu} = 0 \qquad [21]$$

then $s^{\nu}{}_{\nu} = 0$ and $L$ is also invariant with respect to [18]. One shows that, if [21] holds, then, among the projectively related connections satisfying [20], there is precisely one that is metric. To implement properly the metricity condition in the variational principle, one can use the Palatini approach with constraints (Kopczyński 1975). Alternatively, following Hehl, one can use [9] and [12] to eliminate $\omega$ and obtain a Lagrangian depending on $\varphi, \theta$, and the tensor of torsion.

## The Sciama–Kibble Field Equations

From now on the metricity condition [11] is assumed, so that [21] holds and the Cartan field equation [20] is

$$\eta_{\mu\nu\rho} \wedge \Theta^{\rho} = 8\pi s_{\mu\nu} \qquad [22]$$

Introducing the asymmetric energy–momentum tensor $t_{\mu\nu}$ and the spin density tensor $s_{\mu\nu\rho} = g_{\rho\sigma}s^{\sigma}{}_{\mu\nu}$ similarly as in [5], one can write the Einstein–Cartan equations [19] and [22] in the form given by Sciama and Kibble,

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu}R = 8\pi t_{\mu\nu} \qquad [23]$$

$$Q^{\rho}{}_{\mu\nu} + \delta^{\rho}_{\mu}Q^{\sigma}{}_{\nu\sigma} - \delta^{\rho}_{\nu}Q^{\sigma}{}_{\mu\sigma} = 8\pi s^{\rho}{}_{\mu\nu} \qquad [24]$$

Equation [24] can be solved to give

$$Q^{\rho}{}_{\mu\nu} = 8\pi(s^{\rho}{}_{\mu\nu} + \tfrac{1}{2}\delta^{\rho}_{\mu}s^{\sigma}{}_{\nu\sigma} + \tfrac{1}{2}\delta^{\rho}_{\nu}s^{\sigma}{}_{\sigma\mu}) \qquad [25]$$

Therefore, torsion vanishes in the absence of spin and then [23] is the classical Einstein field equation. In particular, there is no difference between the Einstein and Einstein–Cartan theories in empty space. Since practically all tests of relativistic gravity are based on consideration of Einstein's equations in empty space, there is no difference, in this respect, between the Einstein and the Einstein–Cartan theories: the latter is as viable as the former.

In any case, the consideration of torsion amounts to a slight change of the energy–momentum tensor that can be also obtained by the introduction of a new term in the Lagrangian. This observation was made in 1950 by Weyl in the context of the Dirac equation.

In Einstein's theory, one can also satisfactorily describe spinning matter without introducing torsion (Bailey and Israel 1975).

## Consequences of the Bianchi Identities: Conservation Laws

Computing the covariant exterior derivatives of both sides of the Einstein–Cartan equations, using [15] and the Bianchi identities, one obtains

$$8\pi Dt_{\mu} = \tfrac{1}{2}\eta_{\mu\nu\rho\sigma}\Theta^{\nu} \wedge \Omega^{\rho\sigma} \qquad [26]$$

and

$$8\pi Ds_{\mu\nu} = \eta_{\nu\sigma} \wedge \Omega^{\sigma}{}_{\mu} - \eta_{\mu\sigma} \wedge \Omega^{\sigma}{}_{\nu} \qquad [27]$$

Cartan required the right-hand side of [26] to vanish. If, instead, one uses the field equations [19] and [22] to evaluate the right-hand sides of [26] and [27], one obtains

$$Dt_{\mu} = Q^{\rho}{}_{\mu\nu}\theta^{\nu} \wedge t_{\rho} - \tfrac{1}{2}R^{\rho}{}_{\sigma\mu\nu}\theta^{\nu} \wedge s^{\sigma}{}_{\rho} \qquad [28]$$

and

$$Ds_{\mu\nu} = \theta_{\nu} \wedge t_{\mu} - \theta_{\mu} \wedge t_{\nu} \qquad [29]$$

Let $v$ be a vector field generating a group of symmetries of the Riemann–Cartan space $(M, g, \omega)$ so that eqns [13] and [14] hold. Equations [28] and [29] then imply that the 3-form

$$j = v^{\mu}t_{\mu} + \tfrac{1}{2}\tilde{\nabla}^{\nu}v^{\mu}s_{\mu\nu}$$

is closed, $dj = 0$. In particular, in the limit of SRT, in Cartesian coordinates $x^{\mu}$, to a constant vector field $v$ there corresponds the projection, onto $v$, of the energy–momentum density. If $A^{\mu\nu}$ is a constant bivector, then $v^{\mu} = A^{\mu}{}_{\nu}x^{\nu}$ gives $j = j^{\mu\nu}A_{\mu\nu}$, where $j^{\mu\nu}$ is as in [6].

## Spinning Fluid and the Generalized Mathisson–Papapetrou Equation of Motion

As in classical general relativity, the right-hand sides of the Einstein–Cartan equations need not necessarily be derived from a variational principle; they may be determined by phenomenological

considerations. For example, following Weyssenhoff, consider a spinning fluid characterized by

$$t^{\mu\nu} = P^{\mu}u^{\nu} \quad \text{and} \quad s^{\mu\nu\rho} = S^{\mu\nu}u^{\rho}$$

where $S^{\mu\nu} + S^{\nu\mu} = 0$ and $u$ is the unit, timelike velocity field. Let $U = u^{\mu}\eta_{\mu}$ so that

$$t_{\mu} = P_{\mu}U \quad \text{and} \quad s_{\mu\nu} = S_{\mu\nu}U$$

Define the particle derivative of a tensor field $\varphi^{a}$ in the direction of $u$ by

$$\dot{\varphi}^{a}\eta = D(\varphi^{a}U)$$

For a scalar field $\varphi$, the equation $\dot{\varphi} = 0$ is equivalent to the conservation law $d(\varphi U) = 0$. Define $\rho = g_{\mu\nu}P^{\mu}u^{\nu}$, then [29] gives an equation of motion of spin

$$\dot{S}_{\mu\nu} = u_{\nu}P_{\mu} - u_{\mu}P_{\nu}$$

so that

$$P_{\mu} = \rho u_{\mu} + \dot{S}_{\mu\nu}u^{\nu}$$

From [28] one obtains the equation of translatory motion,

$$\dot{P}_{\mu} = (Q^{\rho}{}_{\mu\nu}P_{\rho} - \tfrac{1}{2}R^{\rho\sigma}{}_{\mu\nu}S_{\rho\sigma})u^{\nu}$$

which is a generalization to the ECT of the Mathisson–Papapetrou equation for point particles with an intrinsic angular momentum.

### From ECT to GRT: The Effective Energy–Momentum Tensor

Inside spinning matter, one can use [12] and [25] to eliminate torsion and replace the Sciama–Kibble system by a single Einstein equation with an effective energy–momentum tensor on the right-hand side. Using the split [10], one can write [23] as

$$\overset{\circ}{R}_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}\overset{\circ}{R} = 8\pi T^{\text{eff}}_{\mu\nu} \qquad [30]$$

Here $\overset{\circ}{R}_{\mu\nu}$ and $\overset{\circ}{R}$ are, respectively, the Ricci tensor and scalar formed from $g$. The term in [10] that is quadratic in $\kappa$ contributes to $T^{\text{eff}}$ an expression quadratic in the components of the tensor $s_{\mu\nu\rho}$ so that, neglecting indices, one can symbolically write

$$T^{\text{eff}} = T + s^{2} \qquad [31]$$

The symmetric tensor $T$ is the sum of $t$ and a term coming from $\overset{\circ}{D}\kappa^{\mu}_{\nu}$ in [10]:

$$T^{\mu\nu} = t^{\mu\nu} + \tfrac{1}{2}\overset{\circ}{\nabla}_{\rho}(s^{\nu\mu\rho} + s^{\nu\rho\mu} + s^{\mu\rho\nu}) \qquad [32]$$

It is remarkable that the Belinfante–Rosenfeld symmetrization of the canonical energy–momentum tensor appears as a natural consequence of ECT. From the physical point of view, the second term on the right-hand side of [31], can be thought of as providing a spin–spin contact interaction, reminiscent of the one appearing in the Fermi theory of weak interactions.

It is clear from eqns [30]–[32] that whenever terms quadratic in spin can be neglected – in particular, in the linear approximation – ECT is equivalent to GRT. To obtain essentially new effects, the density of spin squared should be comparable to the density of mass. For example, to achieve this, a nucleon of mass $m$ should be squeezed so that its radius $r_{\text{Cart}}$ be such that

$$\left(\frac{l^{2}}{r^{3}_{\text{Cart}}}\right)^{2} \approx \frac{m}{r^{3}_{\text{Cart}}}$$

Introducing the Compton wavelength $r_{\text{Compt}} = l^{2}/m \approx 10^{-13}$ cm, one can write

$$r_{\text{Cart}} \approx (l^{2}r_{\text{Compt}})^{1/3}$$

The "Cartan radius" of the nucleon, $r_{\text{Cart}} \approx 10^{-26}$ cm, so small when compared to its physical radius under normal conditions, is much larger than the Planck length. Curiously enough, the energy $l^{2}/r_{\text{Cart}}$ is of the order of the energy at which, according to some estimates, the grand unification of interactions is presumed to occur.

### Cosmology with Spin and Torsion

In the presence of spinning matter, $T^{\text{eff}}$ need not satisfy the positive-energy conditions, even if $T$ does. Therefore, the classical singularity theorems of Penrose and Hawking can be overcome here. In ECT, there are simple cosmological solutions without singularities. The simplest such solution, found in 1973 by Kopczyński, is as follows. Consider a universe filled with a spinning dust such that $P^{\mu} = \rho u^{\mu}, u^{\mu} = \delta^{\mu}_{0}, S_{23} = \sigma$, and $S_{\mu\nu} = 0$ for $\mu + \nu \neq 5$, and both $\rho$ and $\sigma$ are functions of $t = x^{0}$ alone. These assumptions are compatible with the Robertson–Walker line element $dt^{2} - \mathcal{R}(t)^{2}(dx^{2} + dy^{2} + dz^{2})$, where $(x, y, z) = (x^{1}, x^{2}, x^{3})$ and torsion is determined from [25]. The Einstein equation [23] reduces to the modified Friedmann equation,

$$\tfrac{1}{2}\dot{\mathcal{R}}^{2} - M\mathcal{R}^{-1} + \tfrac{3}{2}S^{2}\mathcal{R}^{-4} = 0 \qquad [33]$$

supplemented by the conservation laws of mass and spin,

$$M = \tfrac{4}{3}\pi\rho\mathcal{R}^{3} = \text{const.}, \qquad S = \tfrac{4}{3}\pi\sigma\mathcal{R}^{3} = \text{const.}$$

The last term on the left-hand side of [33] plays the role of a repulsive potential, effective at small values of $\mathcal{R}$; it prevents the solution from vanishing. It should be

noted, however, that even a very small amount of shear in $u$ results in a term counteracting the repulsive potential due to spin. Neglecting shear and making the (unrealistic) assumption that matter in the universe at $t = 0$ consists of $\sim 10^{80}$ nucleons of mass $m$ with aligned spins, one obtains the estimate $\mathcal{R}(0) \approx 1\,\mathrm{cm}$ and a density of the order of $m^2/l^4$, very large, but much smaller than the Planck density $1/l^2$.

Tafel (1975) found large classes of cosmological solutions with a spinning fluid, admitting a group of symmetries transitive on the hypersurfaces of constant time. The models corresponding to symmetries of Bianchi types I, $\mathrm{VII}_0$, and V are nonsingular, provided that the influence of spin exceeds that of shear.

## Summary

ECT is a viable theory of gravitation that differs very slightly from the Einstein theory; the effects of spin and torsion can be significant only at densities of matter that are very high, but nevertheless much smaller than the Planck density at which quantum gravitational effects are believed to dominate. It is possible that ECT will prove to be a better classical limit of a future quantum theory of gravitation than the theory without torsion.

*See also:* Cosmology: Mathematical Aspects; General Relativity: Overview.

## Further Reading

Arkuszewski W, Kopczyński W, and Ponomariev VN (1974) On the linearized Einstein–Cartan theory. *Annales de l'Institut Henri Poincaré* 21: 89–95.

Bailey I and Israel W (1975) Lagrangian dynamics of spinning particles and polarized media in general relativity. *Communications in Mathematical Physics* 42: 65–82.

Cartan É (1923, 1924, 1925) Sur les variétés à connexion affine et la théorie de la relativité généralisée. Part I: *Annales de l'École Normale Superiéure* 40: 325–412 and *ibid*. 41: 1–25; Part II: *ibid*. 42: 17–88; English transl. by A Magnon and A Ashtekar, *On manifolds with an affine connection and the theory of general relativity*. Napoli: Bibliopolis (1986).

Cosserat EF (1909) *Théorie des corps déformables*. Paris: Hermann.

Hammond RT (2002) Torsion gravity. *Reports of Progress in Physics* 65: 599–649.

Hehl FW, von der Heyde P, Kerlick GD, and Nester JM (1976) General relativity with spin and torsion: foundations and prospects. *Reviews of Modern Physics* 48: 393–416.

Hehl FW, McCrea JD, Mielke EW, and Ne'eman Y (1995) Metric-affine gauge theory of gravity: field equations, Noether identities, world spinors, and breaking of dilation invariance. *Physics Reports* 258: 1–171.

Kibble TWB (1961) Lorentz invariance and the gravitational field. *Journal of Mathematical Physics* 2: 212–221.

Kopczyński W (1975) The Palatini principle with constraints. *Bulletin de l'Académie Polonaise des Sciences, Série des Sciences Mathématiques, Astronomiques et Physiques* 23: 467–473.

Mathisson M (1937) Neue Mechanik materieller Systeme. *Acta Physica Polonica* 6: 163–200.

Penrose R (1983) Spinors and torsion in general relativity. *Foundations of Physics* 13: 325–339.

Ruggiero ML and Tartaglia A (2003) Einstein–Cartan theory as a theory of defects in space–time. *American Journal of Physics* 71: 1303–1313.

Sciama DW (1962) On the analogy between charge and spin in general relativity. In: (volume dedicated to Infeld L) *Recent Developments in General Relativity*, pp. 415–439. Oxford: Pergamon and Warszawa: PWN.

Tafel J (1975) A class of cosmological models with torsion and spin. *Acta Physica Polonica B* 6: 537–554.

Trautman A (1973) On the structure of the Einstein–Cartan equations. *Symposia Mathematica* 12: 139–162.

Van Nieuwenhuizen P (1981) Supergravity. *Physics Reports* 68: 189–398.

# Einstein's Equations with Matter

**Y Choquet-Bruhat**, Université P.-M. Curie, Paris VI, Paris, France



## Introduction

Newton's theory of gravity with absolute time and Euclidean 3-space connects the gravitational potential $U$ with its source, the density of matter $r$, by the Poisson equation

$$\Delta U = -4\pi\kappa r$$

where $\Delta$ is the Laplace operator and $\kappa$ is the gravitational constant. The trajectories of massive test particles are the flow lines of the gradient of $U$.

Newton's theory has proven to be very accurate in the laboratory as well as in the solar system (except for a small discrepancy with the observed value of Mercury perihelion). Newton's theory together with special relativity, the equivalence principle, and ideas of Mach, have been an inspiration for Einstein to uncover the equations which must be satisfied by the geometry of spacetime. They link the curvature of the spacetime metric with a phenomenological symmetric 2-tensor $T$, which must represent the energy, momentum, and stresses of all the sources, by the equality:

$$S(g) \equiv \mathrm{Ricci}(g) - \tfrac{1}{2}\mathrm{gR(g)} = 8\pi\kappa T$$

where Ricci($g$) is the Ricci tensor of the spacetime metric $g$ and $R(g)$ its scalar curvature. The symmetric 2-tensor $S(g)$ is called the Einstein tensor. The

Bianchi identities, due to the invariance of curvature by isometries of $g$, imply that the divergence of the Einstein tensor is identically zero: the Einstein equations imply therefore the vanishing of the divergence of the source tensor $T$. The equations so obtained generalize in a relativistic context the conservation laws of Newtonian mechanics. In local spacetime coordinates $x^\alpha$, the Einstein equations and conservation laws read

$$S_{\alpha\beta} \equiv R_{\alpha\beta} - \tfrac{1}{2}g_{\alpha\beta}R = 8\pi\kappa T_{\alpha\beta}, \qquad \nabla_\alpha T^{\alpha\beta} \equiv 0$$

where $\nabla$ denotes the covariant derivative in the metric $g$.

The gravitational constant $\kappa$ is inspired by the Newtonian equation relating the potential $U$ with the density of matter. This equation can be obtained as an approximation of Einstein's equations with matter in the case of low velocities of matter and weak gravitational fields. The Newton's equation of motion of test particles is also an approximation of Einstein's geodesic motion of such particles which can be deduced from Einstein's equations themselves. However, if one wants to remain in the framework of the general relativity theory, it is these Einstein's equations which define the mass of a body, there is no comparison possible with some fixed given mass. As length had the dimension of time already in special relativity, now mass is found to have dimension of length. We write the equations in geometrical units, where $8\pi\kappa = 1$, keeping in mind the corresponding change to usual laboratory units only in specific applications. In geometrical units the mass of the Earth is of the order of the centimeter. The most precise measures of $\kappa$ are still made using Newton type experiments, giving $\kappa = 6.67259 \times 10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}}$.

In the case of electromagnetic (or classical Yang–Mills) field sources, the stress energy tensor in special relativity is the well-known Maxwell tensor $\tau$ (or its generalizations), whose divergence vanishes when the field satisfies the Maxwell (or Yang–Mills) equations in vacuum. The expression of this tensor in a curved spacetime can be trivially deduced from its Minkowskian form. Its expression can also be deduced from the Lagrangian, and the vanishing of its divergence results from the invariance of this Lagrangian under isometries of the metric. It is the natural source of Einstein equations coupled with these fields. In the case of matter, the construction of a stress energy tensor is already delicate even in special relativity.

The simplest models of sources with well-understood properties – kinetic matter and perfect fluids – are reviewed in this article. Physical situations difficult to model, even in special relativity, dissipative fluids and elasticity, are mentioned. The extension to electrically, or classical Yang–Mills–Higgs, charged matter, offers no conceptual difficulty, but interesting new situations.

## Fluid Sources

A fluid source in a domain of a spacetime $(V, g)$ is such that there exists, in this domain, a unit timelike vector field $u$, satisfying $g(u,u) \equiv g_{\alpha\beta}u^\alpha u^\beta = -1$, whose trajectories are the flow lines of matter. A moving Lorentzian orthonormal frame is called a proper frame if its timelike vector is $u$. Since the Einstein gravitational potentials reduce at a point in a Lorentzian orthonormal frame to Minkowskian values, one admits that the spacetime symmetric 2-tensor $T$, which embodies the density of stress, energy, and momentum of a given type of matter, in a proper frame takes the expression it would have in special relativity and inertial coordinates. The expression of $T$ in a general frame results from its tensorial character and the equivalence principle. The problem is to find a good expression of $T$ in special relativity.

### Case of Dust (Incoherent Matter)

In a proper frame there is neither momentum nor stresses. Therefore, the stress energy tensor reads in a general frame, with $r$ a scalar function representing the matter density:

$$T = ru \otimes u, \quad \text{i.e., } T_{\alpha\beta} = ru_\alpha u_\beta$$

Using the property $g(u,u) = -1$, the conservation laws imply the vanishing of the divergence of the matter flow $ru$, that is, the continuity equation (conservation of matter)

$$\nabla_\alpha(ru^\alpha) = 0$$

and the motion of the particles along geodesics of the metric:

$$u^\alpha \nabla_\alpha u^\beta = 0$$

Similar equations are obtained for a null dust model where $g(u,u) = 0$.

### Perfect Fluid

**Euler equations** In Newtonian mechanics, a continuous matter flow is characterized by its mass density and flow velocity. The equations are a continuity equation (conservation of matter) and equations of motion resulting from Newton's law, which link the acceleration vector and the space

divergence of the stress symmetric 2-tensor whose contraction with the normal to a small 2-surface gives the force applied to it. A fluid is called perfect if the pressure it applies to a small surface element with normal $n$ is independent of $n$. Its stress tensor $t$, symmetric 2-tensor on Euclidean space, is then invariant by rotations. By generalization, a relativistic fluid is called perfect if its stress energy tensor has the following form:

$$T_{\alpha\beta} = \mu u_\alpha u_\beta + p(g_{\alpha\beta} + u_\alpha u_\beta)$$

Then in a proper frame, where $g$ takes the Minkowskian values and the only nonvanishing component of $u$ is along the time axis and equal to 1, the projection of $T$ on space is the Newtonian stress tensor with pressure $p$, while $\mu$, the projection of $T$ on the time axis, is the fluid energy density. There is no momentum density in the proper frame. The conservation laws, also called Euler equations, are shown to split, as in the case of dust, into a continuity equation

$$\nabla_\alpha[(\mu + p)u^\alpha] - u^\alpha \partial_\alpha p = 0$$

and equations of motion

$$(\mu + p)u^\alpha \nabla_\alpha u^\beta + (g^{\alpha\beta} + u^\alpha u^\beta)\partial_\alpha p = 0$$

In relativity, where mass and energy are equivalent, the continuity equation is no more a conservation law.

**Equations of state** As in Newtonian mechanics, the Euler equations must be completed by a relation, called equation of state, depending on the physical properties of the fluid. In general in addition to mechanics, thermodynamic properties must be considered. In relativity, they are borrowed from classical thermodynamics formulated in a spacetime context.

In the simplest cases one introduces a conserved rest mass density $r$ (or particle number density for particles with rest mass zero), satisfying the equation

$$\nabla_\alpha P^\alpha = 0 \quad \text{with} \quad P^\alpha \equiv r u^\alpha$$

This $r$ differs from the density of energy $\mu$. One sets $\mu = r(1 + \varepsilon)$ and calls $\varepsilon$ the internal specific energy. The first law of (reversible) thermodynamics is extended to relativistic perfect fluids by the identity

$$\Theta \, \mathrm{d}S \equiv \mathrm{d}\varepsilon + p\mathrm{d}(r^{-1})$$

which defines both the absolute temperature $\Theta$ and the differential of the specific entropy $S$. Modulo the continuity equation and the thermodynamic identity, the matter conservation is equivalent to the conservation of entropy along the flow lines:

$$\nabla_\alpha(rSu^\alpha) = 0 \quad \text{hence} \quad u^\alpha \partial_\alpha S = 0$$

The scalars $p$, $\mu$,$S$, $r$ are not independent. Simple situations can be modeled by an "equation of state" linking these quantities. In astrophysics, one is inspired by what is known from classical fluids, with additional relativistic considerations. General relativity plays a role in the case of strong gravitational field.

Very cold matter and nuclear matter are barotropic fluids; they obey an equation of state of the form $p = p(\mu)$.

When the energy $\mu$ is largely dominated by the radiation energy, the fluid is called ultrarelativistic. The Stefan–Boltzmann laws give $\mu = KT^4$ and $p = (1/3)KT^4$, hence $p = (1/3)\mu$; the stress energy tensor is traceless.

In white dwarves, the fluid is considered as polytropic: it obeys an equation of state of the form $p = f(S)r^\gamma$. If only the internal energy $\varepsilon$ and pressure $p$ are dominated by radiation, then $\varepsilon = Kr^{-1}T^4$ and $p = (1/3)KT^4$, hence $p = (1/3)r\varepsilon$. The use of the thermodynamic identity leads to $\gamma = 4/3$, $p = (K/3)(3S/4K)^{4/3}r^{4/3}$, with $\mu = 3p + r$.

For most other stars, the physical situation is too complex to be modeled by a simple equation; only tables of numerical values may be available.

In cosmology, there is little physical information about the fluid which is to represent the energy content of the universe. It is assumed that in the early universe of the big-bang models, at very high temperature, the fluid was ultrarelativistic. At later times, it is generally assumed, for simplicity, that there is an equation of state linear and independent of entropy, $p = (\gamma - 1)\mu$. In order that the speed of sound waves be not greater than the speed of light, one assumes that $1 \leq \gamma \leq 2$; $\gamma = 1$ corresponds to dust, $\gamma = 2$ to a stiff (see below) fluid.

Recent confrontations of theory and observations seem to imply the existence of a new, not directly seen, type of matter, called "dark matter."

**Wave fronts and propagation speeds** The wave fronts of a differential system are the submanifolds of spacetime whose normals $n$ annul the characteristic determinant. Discontinuities propagate along wave fronts. For a hyperbolic system, the wave fronts determine the domain of dependence of a solution. For a perfect fluid, they are found to be

1. the matter wave fronts, generated by the flow lines, such that $u^\alpha n_\alpha = 0$ and

2. the sound wave fronts, whose normals satisfy the equation

$$D \equiv (p'_\mu - 1)(u^\alpha n_\alpha)^2 + p'_\mu n^\alpha n_\alpha = 0$$

in a proper frame at a point of spacetime $u^\alpha = \delta_0^\alpha$, $g_{\alpha\beta} = \eta_{\alpha\beta}$; this equation states that the slope of the spacetime normal to the wave front can be written as

$$\left(\frac{\Sigma(n_i)^2}{n_0^2}\right)^{1/2} = \frac{1}{\sqrt{p'_\mu}}$$

The sound propagation speed is the inverse of this slope, that is, $v = \sqrt{p'_\mu}$. It is less than the speed of light, as expected from a relativistic theory, if $p'_\mu \leq 1$. The limiting case where these speeds are equal is called incompressible or stiff fluid.

**Hyperbolicity, existence, and uniqueness theorem** The characteristics of the perfect fluid equations are real, but the apparent multiplicity of the matter wave fronts poses a problem for the hyperbolicity of the relativistic Euler equations, even in a given background metric. However, Choquet-Bruhat has proven that this system is a hyperbolic Leray system as well as its coupling with the Einstein equations, for instance, in wave gauge. The following theorem can then be proved using the general theorem on hyperbolic systems and an extension of the method used for Einstein's equations in vacuum.

**Theorem** *Let $(M,\bar{g},K)$ be an initial data set for the Einstein equations and $(\bar{u},\bar{\mu},\bar{S})$ be Cauchy data in a local Sobolev space $H_s^{\text{loc}}$, $s \geq 3$, on the 3-manifold M for a perfect fluid with a smooth equation of state. Suppose $\bar{\mu} > 0$ and $p'_{\bar{\mu}} \leq 1$. There exists a globally hyperbolic spacetime of maximal extension solution of the Einstein equations with source such as perfect fluid taking these Cauchy data. Such a spacetime and fluid flow are smooth for smooth initial data. They are unique, up to spacetime isometries.*

The Euler equations have also been written as a first-order symmetric hyperbolic system by Boillat, Ruggeri, and Strumia using general methods relying on the existence of a convex functional, and directly by Rendall, who pointed out the difficulty of modeling the general motion of isolated fluid bodies, because of the assumption $\bar{\mu} > 0$. He constructed some solutions without this assumption where the boundaries are freely falling. The general problem of determining the evolution of boundaries appears everywhere in general relativity, and in classical mechanics.

**Global problems** The spacetimes obtained above are, in general, incomplete: even in Minkowski spacetime, the Euler equations do not in general have solutions that are global in time. Shocks appear in relativistic perfect fluids as in classical ones. Global existence results have been obtained for four-dimensional ultrarelativistic fluids (limited data), and in the case of 1-space dimension. A detailed study of the global behavior of spherically symmetric solutions of the Einstein–Euler equations with equation of state admitting a phase transition from zero pressure to stiff fluid has been done by Christodoulou.

### Dissipative Fluids

A general fluid stress energy tensor is with $u$, a unit vector whose trajectories are the flow lines:

$$T^{\alpha\beta} = \mu u^\alpha u^\beta + q^\alpha u^\beta + q^\beta u^\alpha + Q^{\alpha\beta}$$
$$\text{with } q^\alpha u_\alpha = 0, \quad Q^{\alpha\beta} u_\alpha = 0$$

$\mu = T^{\alpha\beta} u_\alpha u_\beta$ is the energy density, which must satisfy $\mu \geq 0$, $Q$ is a space tensor representing the stresses, orthogonal to $u$ and $q$ is a space vector considered as a heat flow. The fundamental equations are still $\nabla_\alpha T^{\alpha\beta} = 0$, but they must be implemented by constitutive equations for $q$ and $Q$ which do not have simple satisfactory answer in a relativistic context. The transfer of results from classical mechanics on viscous fluids or on heat transfer leads to propagation speeds greater than the speed of light. It should be remarked that these classical equations are obtained as governing asymptotic states; thus, the parabolic character of their relativistic version does not contradict relativistic causality. However, it would be interesting to obtain, for dissipative relativistic fluids, hyperbolic dissipative equations. Various systems have been proposed, in particular, by Marle by using an approximation near equilibrium of a solution of the relativistic Boltzmann equation. A promising system, also inspired from kinetic theory, is the "extended thermodynamics" of Müller and Ruggeri which takes as 14 fundamental unknowns, the vector $P = ru$ and the tensor $T$, satisfying the conservation laws. These equations are supplemented by equations linking a totally symmetric 3-tensor $A$ with a symmetric 2-tensor $I$ by equations of the form

$$\nabla_\alpha A^{\alpha\beta\gamma} = I^{\beta\gamma} \qquad [1]$$

$A$ and $I$ are functions of $P$ and $T$ depending on the model and called constitutive equations. The system is shown to be symmetric hyperbolic under the existence of a convex entropy function, property which holds under appropriate physical assumptions.

Reasonable equations have been proposed and studied for several constituent fluids and superfluids.

### Charged Fluids

The stress energy tensor of a charged fluid with electric (or Yang–Mills) charge is generally the sum of the stress energy tensor of the fluid and of the Maxwell (or Yang–Mills) field. This tensor is conserved modulo the Maxwell (or Yang–Mills) equations with source the electric current, and the Euler equations completed by the Lorentz force. The corresponding Einstein–Maxwell perfect fluid system is well posed in the case of zero or infinite conductivity (magnetohydrodynamics). A subtlety appears in the case of finite conductivity: the system is still well posed, but for a restricted (Gevrey) class of $C^\infty$ fields.

## Kinetic Models

### Distribution Function and Moments

A general relativistic kinetic theory can be formulated without appeal to classical mechanics or special relativity. The matter is composed of particles whose size is negligible in the considered scale: rarefied gases in the laboratory, galaxies or even clusters of galaxies at the cosmological scale. The number of particles is so great and their motion so chaotic that the state of the matter can be described by a "one-particle distribution function," a positive scalar function on the tangent bundle to the spacetime $(x,p) \mapsto f(x,p)$, which gives the mean number of particles with momentum $p$ present at the point $x$ of spacetime.

The first moment of $f$ is a causal vector field $P$ defined by the integral over the space $\mathcal{P}_x$ of momenta at $x$, with $\omega_p$ a volume element in that space:

$$P(x) =: \int_{\mathcal{P}_x} pf(x,p)\omega_p$$

Out of the first moment, one extracts a scalar $r \geq 0$, interpreted as the square of a proper mass density given by $r^2 =: -g(P, P)$ and, if $r > 0$, a unit vector $u = r^{-1}P$ interpreted as the macroscopic flow velocity.

The second moment of the distribution function $f$ is the symmetric 2-tensor on spacetime given by

$$T(x) =: \int_{\mathcal{P}_x} f(x,p)p \otimes p\omega_p$$

It is interpreted as the stress energy tensor of the distribution $f$. Higher moments are defined similarly.

### Liouville–Vlasov Equation

When the gas is so rarefied that the particle trajectories do not cross, then in the absence of nongravitational forces, these trajectories are geodesics of $g$, orbits in $TV$ of the vector field $X = (p^\alpha,\ Q^\alpha \equiv -\Gamma^\alpha_{\lambda\mu} p^\lambda p^m)$ with $\Gamma^\alpha_{\lambda\mu}$, the Christoffel symbols of $g$.

In a collisionless model, the physical law of conservation of particles imposes the conservation of $f$ along the trajectories of $X$, that is, the Liouville–Vlasov equation

$$\mathcal{L}_X f \equiv p^\alpha \frac{\partial f}{\partial x^\alpha} + Q^\alpha \frac{\partial f}{\partial p^\alpha} = 0$$

**Conservation laws** If $f$ satisfies the Vlasov equation, then all moments satisfy a conservation law, in particular,

$$\nabla_\alpha P^\alpha = 0 \text{ and } \nabla_\alpha T^{\alpha\beta} = 0$$

equations which make the Einstein–Vlasov system consistent.

The theory extends without problem to particles having the same rest mass $m$, because the scalar $g(p, p) = -m^2$ is constant on a geodesic.

**Cauchy problem** The Einstein–Vlasov system is an integro-differential system for $g$ and $f$ on a manifold $V = M \times R$. The Cauchy data for the spacetime metric $g$ on $M_0 = M \times \{0\}$ is, as usual, a pair $(\bar{g}, K)$, implemented with gauge initial data which complete the definition of Cauchy data for a well-posed hyperbolic system in the chosen gauge. The Cauchy data for $f$ are a function $\bar{f}$ on the bundle $P_{M_0}$. It has been proved long ago that there exists a solution, geometrically unique, in a neighborhood of $M_0$ if the data are in Sobolev spaces, weighted by a power of $p^0$ in the case of $\bar{f}$.

Since the Vlasov matter model, solution of a linear equation for given $g$, has no singularity by itself, the Einstein–Vlasov system is a good candidate for solutions that are global in time. This global existence has been proved by Rein and Rendall in the case of small data, asymptotically flat with spherical symmetry or plane symmetry, or with hyperbolic symmetry and compact space. Global existence without these symmetries is an open problem.

### Boltzmann Equation

When the particles undergo collisions, their trajectories in phase space are no more connected integral curves of the vector field $X$, that is, their moment

undergoes a jump with the crossing of another trajectory. In the Boltzmann model, the derivative $\mathcal{L}_X f$ is equal to the so-called collision operator, $\mathcal{I}f$:

$$(\mathcal{L}_X f)(x,p) = (\mathcal{I}f)(x,p)$$

where $\mathcal{I}f$ is an integral operator linked with the probability that two particles of momentum, respectively, $p'$ and $q'$, collide at $x$ and give, after the shock, two particles of momentum $p$ and $q$. For "elastic" shocks, the total momentum is conserved, that is, $p'$ and $q'$ lie in the submanifold $\Sigma_{pq} =: \{p' + q' = p + q\}$, with volume element $\xi'$ and

$$(\mathcal{I}f)(x,p) \equiv \int_{\mathcal{P}_x}\int_{\Sigma_{pq}} [f(x,p')f(x,q') - f(x,p)f(x,q)]A(x,p,q,p',q')\xi' \wedge \omega_q$$

The function $A(x, p, q, p', q')$ is called the shock cross section; it is a phenomenological quantity. No explicit expression is known for it in relativity. A generally admitted property is the reversibility of elastic shocks, $A(x, p, q, p', q') = A(x, p', q', p, q)$. It can be proved that under this hypothesis, the first and second moment of $f$ are conserved as in the collisionless case, making the Einstein–Boltzmann system consistent. Existence of solutions (that are local in time) of the Cauchy problem for this system has long been known. No global existence for the coupled system is known yet.

One defines, in a relativistic context, an entropy flux vector $H$ which is proved to satisfy an $H$-theorem, that is, $\nabla_\alpha H^\alpha \geq 0$. In an expanding universe, for instance, Robertson Walker, where $H$ depends only on time and an entropy density is defined by $H^0$, one finds that a decrease in entropy is linked with the expansion of the universe, thus permitting its ever-increasing organization from an initial anisotropy of $f$ in momentum space.

## Other Matter Sources

### Elastic Media

There are no solids in general relativity; in special relativity rigid motions are already very restricted. A theory of elastic deformations can only be defined relatively to some *a priori* given state of matter whose perturbations will satisfy laws analogous to the classical laws. Various such theories have been proposed through geometric considerations, extending methods of classical elasticity; they have been used to predict the possible signals from bar detectors of gravitational waves, or the motions in the crust of neutron stars. A general theory constructed by Lagrangian formalism has recently been developed.

### Spinor Sources

A symmetric stress energy tensor can be associated to classical spinors of spin 1/2, leading to a well-posed Einstein–Dirac system. The theories of supergravity couple the Einstein–Cartan equations with anticommuting spin 3/2 sources.

*See also:* Boltzmann Equation (Classical and Quantum); Einstein Equations: Exact Solutions; Einstein Equations: Initial Value Formulation; General Relativity: Overview; Geometric Analysis and General Relativity; Kinetic Equations; Spinors and Spin Coefficients.

## Further Reading


Anile M (1989) *Relativistic Fluids and Magneto Fluids*. Cambridge: Cambridge University Press.

Bancel D and Choquet-Bruhat Y (1973) Existence, uniqueness and local stability for the Einstein–Maxwell–Boltzmann system. *Communications in Mathematical Physics* 33: 76–83.

Beig R and Schmidt B (2002) Relativistic elasticity, gr-qc 0211054.

Carter B and Quintana H (1972) Foundations of general relativistic high pressure elasticity theory. *Proceedings of the Royal Society A* 331: 57–83.

Choquet-Bruhat Y (1958) Théorèmes d'existence en mécanique des fluides relativistes. *Bull. Soc. Math. de France* 86: 155–175.

Choquet-Bruhat Y (1966) Etude des équations des fluides relativistes inductifs et conducteurs. *Communications in Mathematical Physics* 3: 334–357.

Choquet-Bruhat Y (1987) Spin $\frac{1}{2}$ fields in arbitrary dimensions and the Einstein Cartan theory. In: Rindler W and Trautman A (eds.) *Gravitation and Geometry*, pp. 83–106. Bibliopolis.

Choquet-Bruhat Y and Lamoureux-Brousse L (1973) Sur les équations de l'élasticité relativiste. *Comptes Rendus Hebdomadalrun dun de l' Academic des Sciences, Paris A* 276: 1317–1321.

Choquet-Bruhat Y. General Relativity and Einstein's equations (in preparation).

Christodoulou D (1995) Self gravitating fluids: a two-phase model. *Archives for Rational and Mechanical Analysis* 130: 343–400 and subsequent papers.

Ehlers J (1969) General relativity and kinetic theory. corso XLVII Scuola Internazionale Enrico Fermi.

Friedrich H (1998) Evolution equations for gravitating fluid bodies in general relativity. *Physical Review D* 57: 2317–2322.

Geroch R and Lindblom L (1991) Causal theories of relativistic fluids. *Annals of Physics* 207: 394–416.

Israel W (1987) Covariant fluid mechanics and thermodynamics, an introduction. In: Anile M and Choquet-Bruhat Y (eds.) *Relativistic Fluid Dynamics*, LNM 1385. Springer.

Lichnerowicz A (1967) *Relativistic Hydrodynamics and Magnetohydrodynamics*. Benjamin.

Marle C (1969) Sur l'établissement des équations de l'hydrodynamique des fluides relativistes dissipatifs. *Ann Inst. H. Poincaré* 10: 67–194.

Müller I and Ruggeri T (1999) *Rational extended thermodynamics*. Springer.

Rein G and Rendall AD (1992) Global existence of solutions of the sphericall symmetric Vlasov–Einstein system with small initial data. *Communications in Mathematical Physics* 150: 561–583.

Rendall AD (1992) The initial value problem for a class of general relativistic fluid bodies. *Journal of Mathematical Physics* 33: 1047–1053.

Taub AH (1957) Relativistic hydrodynamics III. *Journal of Mathematics* 1: 370–400.

# Electric–Magnetic Duality

**Tsou Sheung Tsun**, University of Oxford, Oxford, UK



## Introduction

Classical electromagnetism is described by Maxwell's equations, which, in 3-vector notation and corresponding respectively to the laws of Coulomb, Ampère, Gauss, and Faraday, are given by eqns [1a]–[1d]:

$$\operatorname{div} \boldsymbol{E} = \rho \tag{1a}$$

$$\operatorname{curl} \boldsymbol{B} - \frac{\partial \boldsymbol{E}}{\partial t} = \boldsymbol{J} \tag{1b}$$

$$\operatorname{div} \boldsymbol{B} = 0 \tag{1c}$$

$$\operatorname{curl} \boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} = 0 \tag{1d}$$

Equivalently, in covariant 4-vector notation, these correspond to eqns [2a] and [2b]:

$$\partial_\nu F^{\mu\nu} = -j^\mu \tag{2a}$$

$$\partial^*{}_\nu F^{\mu\nu} = 0 \tag{2b}$$

In eqns [1], $\boldsymbol{E}$ and $\boldsymbol{B}$ are the electric and magnetic fields, respectively, $\rho$ is the electric charge density, and $\boldsymbol{J}$ is the electric current. In eqns [2], $F_{\mu\nu}$ is the field tensor, ${}^*F_{\mu\nu}$ the dual field tensor, and $j^\mu$ is the 4-current, related to the previous vector quantities by the following relations:

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & -B_3 & B_2 \\ -E_2 & B_3 & 0 & -B_1 \\ -E_3 & -B_2 & B_1 & 0 \end{pmatrix}$$

$${}^*F_{\mu\nu} = \begin{pmatrix} 0 & B_1 & B_2 & B_3 \\ -B_1 & 0 & E_3 & -E_2 \\ -B_2 & -E_3 & 0 & E_1 \\ -B_3 & E_2 & -E_1 & 0 \end{pmatrix}$$

$$j^\mu = (\rho, \boldsymbol{J})$$

Throughout this article, we shall denote the three spatial indices by lower-case Latin letters such as $i$, $j$, while Greek indices such as $\mu$, $\nu$ denote spacetime indices running through $0, 1, 2, 3$. The Einstein summation convention is used, whereby repeated indices are summed. Spacetime indices are raised and lowered by the (flat) Minkowski metric $g_{\mu\nu} = \operatorname{diag}(1, -1, -1, -1)$. We also use units conventional in particle physics, in which the reduced Planck constant $\hbar$ and the speed of light $c$ are both set to 1.

In terms of the totally skew symmetric symbol $\varepsilon_{\mu\nu\rho\sigma}$ (with $\varepsilon_{0123} = 1$), the two field tensors are related by eqn [3]:

$${}^*F_{\mu\nu} = -\tfrac{1}{2}\varepsilon_{\mu\nu\rho\sigma}F^{\rho\sigma} \tag{3}$$

We say that ${}^*F_{\mu\nu}$ is the dual of $F_{\mu\nu}$, and eqn [3] is indeed a duality relation because eqn [4] holds, which means that up to a sign, $F_{\mu\nu}$ and ${}^*F_{\mu\nu}$ are duals of each other:

$${}^*({}^*F) = -F \tag{4}$$

This duality is in fact the Hodge duality between $p$-forms and $(n-p)$-forms in an $n$-dimensional space. In our particular case, $p = 2$ and $n = 4$, so that both $F$ and its dual are 2-forms. The minus sign in eqn [4] comes about because of the Lorentzian (or pseudo-Riemannian) signature of Minkowski spacetime.

The physical significance of this duality is that such a symmetry interchanges electric and magnetic fields (again up to sign) (eqn [5]), as can be seen from the matrix representation of $F_{\mu\nu}$ and ${}^*F_{\mu\nu}$ above:

$${}^*\!: \boldsymbol{E} \mapsto \boldsymbol{B},\ \boldsymbol{B} \mapsto -\boldsymbol{E} \tag{5}$$

Now in the absence of electric charges and currents, one sees immediately that Maxwell's equations [1] or [2] are dual symmetric. This means that, *in vacuo*, whether we call an electromagnetic field electric or magnetic is a matter of convention. As far as the dynamics is concerned, there is no distinction.

On the other hand, eqns [1] and [2] as presented, that is, in the presence of matter, are manifestly not dual symmetric. The underlying reason for this asymmetry has been much studied both in physics and in mathematics. One of the two questions that this article addresses is precisely this. Following on this, we shall see what happens if we try somehow to restore this dual symmetry even in the presence of matter.

The second question that we wish to discuss is a generalization of this duality. Electromagnetism is a gauge theory, in which the gauge group is the abelian circle group U(1), representing the phase of wave-functions in quantum mechanics. A physically relevant generalization, in which the abelian U(1) is replaced by a nonabelian group (e.g., SU(2), SU(3))

is called Yang–Mills theory (Yang and Mills 1954), which is the theoretical basis of all modern particle physics. We shall show in this article how the concept of electric–magnetic duality can be generalized in the context of Yang–Mills theory.

## Gauge Invariance, Sources, and Monopoles

Electric–magnetic duality, whether in the well-known abelian case or in the still somewhat open nonabelian case, is intimately connected with gauge invariance, sources, and monopoles, and also the dynamics as embodied in the gauge action. These questions in turn find their natural setting in differential geometry, particularly the geometry of fibre bundles.

Although classical electrodynamics can be fully described by the field tensor $F_{\mu\nu}$, one needs to introduce the electromagnetic (or gauge) potential $A_\mu$ if one considers quantum mechanics, as has been beautifully demonstrated by the Bohm–Aharonov experiment. The two quantities are related by eqn [6]:

$$F_{\mu\nu}(x) = \partial_\nu A_\mu(x) - \partial_\mu A_\nu(x) \qquad [6]$$

The fact that the phase of a wave function $\psi(x)$ (e.g., of the electron) is not a measurable quantity (although relative phases of course are) implies that we are free to make the following transformation:

$$\psi(x) \mapsto \mathrm{e}^{\mathrm{i}e\Lambda(x)}\psi(x) \qquad [7]$$

This in turn implies an unobservable transformation [8] on the gauge potential, where $\Lambda(x)$ is a real-valued function on spacetime:

$$A_\mu(x) \mapsto A_\mu(x) + \partial_\mu\Lambda(x) \qquad [8]$$

This invariance is called gauge invariance. Since in this abelian case $F_{\mu\nu}$ is gauge invariant, so are the Maxwell equations, for which we shall take from now on the covariant form [2]. Inasmuch as the Maxwell equations dictate the dynamics of electromagnetism, gauge invariance is an intrinsic ingredient even in the classical theory.

In Yang–Mills theory, the U(1) phase $\mathrm{e}^{\mathrm{i}e\Lambda(x)}$ is replaced by an element $S(x)$ of a nonabelian group $G$, so that eqns [7], [8], and [6] become, respectively, eqns [9], [10], and [11]:

$$\psi(x) \mapsto S(x)\psi(x) \qquad [9]$$

$$A_\mu(x) \mapsto S(x)A_\mu(x)S^{-1}(x) - \left(\frac{i}{g}\right)\partial_\mu S(x)S^{-1}(x) \qquad [10]$$

$$F_{\mu\nu}(x) = \partial_\nu A_\mu(x) - \partial_\mu A_\nu(x) + \mathrm{i}g[A_\mu(x), A_\nu(x)] \qquad [11]$$

Here the electric coupling $e$ is replaced by a general gauge coupling $g$. The quantities $A_\mu$ and $F_{\mu\nu}$ now take values in the Lie algebra of the Lie group $G$ and the bracket is the Lie bracket. The wave function $\psi(x)$ takes values in a vector space on which an appropriate representation of $G$ acts. Notice that now the field tensor $F_{\mu\nu}$ is no longer invariant, but only covariant:

$$F_{\mu\nu}(x) \mapsto S(x)F_{\mu\nu}(x)S^{-1}(x) \qquad [12]$$

Next we consider the charges of gauge theory. For the moment, we wish to distinguish between two types of charges: sources and monopoles. These are defined with respect to the gauge field, which in turn is derivable from the gauge potential.

Source charges are those charges that give rise to a nonvanishing divergence of the field. For example, the electric current $j$ due to the presence of the electric charge $e$ occurs on the right-hand side of the first Maxwell equation, and is given in the quantum case by eqn [13], where $\gamma^\mu$ is a Dirac gamma matrix, identifiable as a basis element of the Clifford algebra over spacetime:

$$j^\mu = e\bar{\psi}\gamma^\mu\psi \qquad [13]$$

In the Yang–Mills case, the first Maxwell equation is replaced by the Yang–Mills equation

$$D_\nu F^{\mu\nu} = -j^\mu, \qquad j^\mu = g\bar{\psi}\gamma^\mu\psi \qquad [14]$$

We define the covariant derivative $D$ as in

$$D_\mu F^{\mu\nu} = \partial_\mu F^{\mu\nu} - \mathrm{i}g[A_\mu, F^{\mu\nu}] \qquad [15]$$

Monopole charges, on the other hand, are topological obstructions specified geometrically by nontrivial $G$-bundles over every 2-sphere $S^2$ surrounding the charge. They are classified by elements of $\pi_1(G)$, the fundamental group of $G$. They are typified by the (abelian) magnetic monopole as first discussed by Dirac in 1931.

Let us go into a little more detail about the Dirac magnetic monopole. If the field tensor $F_{\mu\nu}$ does come from a gauge potential $A_\mu$ as in eqn [6], then simple algebra will tell us that this implies $\partial_\nu {}^*F^{\mu\nu} = 0$ as in eqn [2]. Hence, we conclude the following:

$$\exists \text{ monopole} \implies A_\mu \text{ cannot be well defined everywhere}$$

The result is actually stronger. Suppose there exists a magnetic monopole at a certain point in spacetime, and, without loss of generality, we shall consider a static monopole. If we surround this point by a (spatial) 2-sphere $\Sigma$, then the magnetic flux out of the sphere is given by

$$\iint_\Sigma \boldsymbol{B}\cdot\mathrm{d}\boldsymbol{\sigma} = \iint_{\Sigma^\mathrm{N}} \boldsymbol{B}\cdot\mathrm{d}\boldsymbol{\sigma} + \iint_{\Sigma^\mathrm{S}} \boldsymbol{B}\cdot\mathrm{d}\boldsymbol{\sigma} \qquad [16]$$

Here $\Sigma^{\mathrm{N}}$ and $\Sigma^{\mathrm{S}}$ are the northern and southern hemispheres overlapping on the equator $S$. By Stokes' theorem, since $F_{\mu\nu}$ has no components $F_{0i} = E_i$, we have

$$\iint_{\Sigma^{\mathrm{N}}} \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{\sigma} = \oint_S \boldsymbol{A} \cdot \mathrm{d}\boldsymbol{s} \qquad [17a]$$

$$\iint_{\Sigma^{\mathrm{S}}} \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{\sigma} = \oint_{-S} \boldsymbol{A} \cdot \mathrm{d}\boldsymbol{s} \qquad [17b]$$

In eqn [17b], $-S$ means the equator with the opposite orientation. Hence, $\oint_S + \oint_{-S} = 0$. But this contradicts the assumption that there exists a magnetic monopole at the center of the sphere. Hence, we see that if a monopole exists, then $A_\mu$ will have at least a string of singularities leading out of it. This is the famous Dirac string.

The more mathematically elegant way to describe this is that the principal bundle corresponding to electromagnetism with a magnetic monopole is nontrivial, so that the gauge potential $A_\mu$ has to be patched (i.e., related by transition functions in the overlap). Consider the example of a static monopole of magnetic charge $\tilde{e}$. For any (spatial) sphere $S_r$ of radius $r$ surrounding the monopole, we cover it with two patches N, S as follows:

$$\begin{aligned} (\mathrm{N}){:}\quad & 0 \le \theta < \pi,\ 0 \le \phi \le 2\pi \\ (\mathrm{S}){:}\quad & 0 < \theta \le \pi,\ 0 \le \phi \le 2\pi \end{aligned}$$

In each patch we define the following:

$$\begin{aligned} A_1^{(\mathrm{N})} &= \frac{\tilde{e}y}{4\pi r(r+z)} \\ A_2^{(\mathrm{N})} &= -\frac{\tilde{e}x}{4\pi r(r+z)} \\ A_3^{(\mathrm{N})} &= 0 \\ A_1^{(\mathrm{S})} &= -\frac{\tilde{e}y}{4\pi r(r-z)} \\ A_2^{(\mathrm{S})} &= \frac{\tilde{e}x}{4\pi r(r-z)} \\ A_3^{(\mathrm{S})} &= 0 \end{aligned}$$

In the overlap (containing the equator), $A^{(\mathrm{N})}$ and $A^{(\mathrm{S})}$ are related by a gauge transformation:

$$\begin{aligned} A_i^{(\mathrm{N})} - A_i^{(\mathrm{S})} &= \partial_i \Lambda \\ \Lambda &= \left(\frac{\tilde{e}}{2\pi}\right) \tan^{-1}\left(\frac{y}{x}\right) = \frac{\tilde{e}\phi}{2\pi} \end{aligned} \qquad [18]$$

Notice that $A_i^{(\mathrm{N})}$ has a line of singularity along the negative $z$-axis (which is the Dirac string in this case); similarly for $A_i^{(\mathrm{S})}$ along the positive $z$-axis.

Furthermore, the corresponding field strength is given by

$$\boldsymbol{E} = 0 \qquad [19a]$$

$$\boldsymbol{B} = \frac{\tilde{e}\boldsymbol{r}}{4\pi r^3} \qquad [19b]$$

If we now evaluate the "magnetic flux" out of $S_r$, we have

$$\iint_{S_r} \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{\sigma} = \oint_{\text{Equator}} \left(A_\mu^{(\mathrm{N})} - A_\mu^{(\mathrm{S})}\right) \mathrm{d}x^\mu = \tilde{e} \qquad [20]$$

In other words, in the presence of a magnetic monopole, the second half of Maxwell's equations is modified according to eqn [21], with $\tilde{j}^\mu$ given by eqn [22].

$$\left.\begin{aligned} \operatorname{div} \boldsymbol{B} &= \tilde{\rho} \\ \operatorname{curl} \boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} &= \tilde{\boldsymbol{J}} \end{aligned}\right\}, \qquad \partial_\nu {}^*F^{\mu\nu} = -\tilde{j}^\mu \qquad [21]$$

$$\tilde{j}^\mu = \tilde{e}\bar{\psi}\gamma^\mu\psi \qquad [22]$$

Furthermore, the form of eqn [21] tells us that a monopole of the $F_{\mu\nu}$ field can also be considered as a source of the ${}^*F_{\mu\nu}$ field. The two descriptions are equivalent.

How are the charges $e$ and $\tilde{e}$ related? The gauge transformation $S = \mathrm{e}^{\mathrm{i}e\Lambda}$ relating $A_\mu^{(\mathrm{N})}$ and $A_\mu^{(\mathrm{S})}$ must be well defined; that is, if one goes round the equator once, $\phi = 0 \to 2\pi$, one should get the same $S$. This gives

$$e\tilde{e} = 2\pi n, \qquad n \in \mathbb{Z} \qquad [23]$$

In particular, the unit electric and magnetic charges are related by eqn [24], which is Dirac's quantization condition,

$$e\tilde{e} = 2\pi \qquad [24]$$

So, in principle, just as in the electric case, where we could have charges $e, 2e, \ldots,$ here we could also have magnetic charges of $\tilde{e}, 2\tilde{e}, \ldots$ . In other words, both charges are quantized.

Another way to look at this is to consider the classification of principal bundles over $S^2$. The reason for these topological 2-spheres is that we are interested in enclosing a point charge. For a nontrivial bundle, the patching is given by a function $S$ defined in the overlap (the equator), in other words, a map $S^1 \to \mathrm{U}(1)$. What this amounts to is a closed curve in the circle group U(1). Now, curves that can be continuously deformed into one another cannot give

distinct fibre bundles, so that one sees easily that there exists a one-to-one correspondence:

$$\{\text{principal U(1) bundles over } S^2\}$$
$$\updownarrow$$
$$\{\text{homotopy classes of closed curves in U(1)}\}$$

This last is $\pi_1(\text{U}(1)) \cong \mathbb{Z}$. Hence, we recover Dirac's quantization condition.

So, for electromagnetism, there are two equivalent ways of defining the magnetic charge, as a source or as a monopole:

1. $\partial_\nu {}^*F^{\mu\nu} = -\tilde{j}^\mu \propto n\tilde{e} \neq 0$.
2. An element of $\pi_1(\text{U}(1)) \cong \mathbb{Z}$.

The same goes for the electric charge. We also note that both definitions give us the fact that these charges are discrete (quantized) and conserved (invariant under continuous deformations).

We now want to apply similar considerations to the magnetic charges in the nonabelian case. For several (subtle) reasons the obvious expression $D_\nu {}^*F^{\mu\nu} \stackrel{?}{=} -\tilde{j}^\mu$ as a source (see **Table 1**) does not work. The quickest way to say this is that ${}^*F^{\mu\nu}$ in general has no corresponding potential $\tilde{A}_\mu$ and so is not a gauge field. Moreover, in contrast to the abelian case, the field tensor does not fully specify the physical field configuration, as demonstrated by Wu and Yang. We shall come back to this later.

But we have just seen that in the abelian case there is another equivalent definition, which is that a magnetic monopole is given by the gauge configuration corresponding to a nontrivial U(1) bundle over $S^2$. This can be generalized to the nonabelian case without any problem. Moreover, this definition automatically guarantees that a nonabelian monopole charge is quantized and conserved. This is the way monopoles are defined above.

Arguments similar to the abelian case easily yield the nonabelian analog of the Dirac quantization condition, eqn [25], the difference between the two cases being only a matter of conventional normalization.

$$g\tilde{g} = 4\pi \qquad [25]$$

**Table 1** Definitions of charges

| | *Sources* | *Monopoles* |
|---|---|---|
| Abelian | $\partial_\nu F^{\mu\nu} = -j^\mu$ | $\partial_\nu {}^*F^{\mu\nu} = -\tilde{j}^\mu$ |
| Nonabelian | $D_\nu F^{\mu\nu} = -j^\mu$ | ? |

## Abelian Duality and the Wu–Yang Criterion

We saw above the well-known fact that classical Maxwell theory is invariant under the duality operator. By this we mean that at any point in spacetime free of electric and magnetic charges we have the two dual symmetric Maxwell equations:

$$\partial_\nu {}^*F^{\mu\nu} = 0 \quad [\text{d}F = 0] \qquad [26]$$

$$\partial_\nu F^{\mu\nu} = 0 \quad [\text{d}^*F = 0] \qquad [27]$$

Displayed in square brackets are the equivalent equations in the language of differential forms. Then by the Poincaré lemma we deduce immediately the existence of potentials $A$ and $\tilde{A}$ such that eqns [28] and [29] hold:

$$F_{\mu\nu}(x) = \partial_\nu A_\mu(x) - \partial_\mu A_\nu(x) \qquad [F = \text{d}A] \qquad [28]$$

$${}^*F_{\mu\nu}(x) = \partial_\nu \tilde{A}_\mu(x) - \partial_\mu \tilde{A}_\nu(x) \qquad [{}^*F = \text{d}\tilde{A}] \qquad [29]$$

The two potentials transform independently under independent gauge transformations $\Lambda$ and $\tilde{\Lambda}$:

$$A_\mu(x) \mapsto A_\mu(x) + \partial_\mu \Lambda(x) \qquad [30]$$

$$\tilde{A}_\mu(x) \mapsto \tilde{A}_\mu(x) + \partial_\mu \tilde{\Lambda}(x) \qquad [31]$$

This means that the full symmetry of this theory is doubled to $\text{U}(1) \times \tilde{\text{U}}(1)$, where the tilde on the second circle group indicates that it is the symmetry of the dual potential $\tilde{A}$. It is important to note that the physical degrees of freedom remain the same. This is clear because $F$ and ${}^*F$ are related by an algebraic equation [3]. As a consequence, the physical theory is the same: the doubled gauge symmetry is there all the time but is just not so readily detected.

As mentioned in the Introduction, this dual symmetry means that what we call "electric" or "magnetic" is entirely a matter of choice.

In the presence of electric charges, the Maxwell equations usually appear as

$$\partial_\nu {}^*F^{\mu\nu} = 0 \qquad [32]$$

$$\partial_\nu F^{\mu\nu} = -j^\mu \qquad [33]$$

The apparent asymmetry in these equations comes from the experimental fact that there is only one type of charges observed in nature, which we choose to regard as a source of the field $F$ (or, equivalently but unconventionally, as a monopole of the field ${}^*F$). But as we see by dualizing eqns [32] and [33], that is, by interchanging the role of electricity and magnetism in relation to $F$, we could equally have thought of these instead as source charges of

the field $^*F$ (or, similarly to the above, as monopoles of $F$):

$$\partial_\nu {}^*F^{\mu\nu} = -\tilde{j}^\mu \qquad [34]$$

$$\partial_\nu F^{\mu\nu} = 0 \qquad [35]$$

If both electric and magnetic charges existed in nature, then we would have the dual symmetric pair:

$$\partial_\nu {}^*F^{\mu\nu} = -\tilde{j}^\mu \qquad [36]$$

$$\partial_\nu F^{\mu\nu} = -j^\mu \qquad [37]$$

This duality in fact goes much deeper, as can be seen if we use the Wu–Yang criterion to derive the Maxwell equations, although we should note that what we present here is not the textbook derivation of the Maxwell equations from an action, but we conisder this method to be much more intrinsic and geometric. Consider first pure electromagnetism. The free Maxwell action is given by

$$\mathcal{A}_F^0 = -\frac{1}{4}\int F_{\mu\nu}F^{\mu\nu} \qquad [38]$$

The true variables of the (quantum) theory are the $A_\mu$, so in eqn [38] we should put in a constraint to say that $F_{\mu\nu}$ is the curl of $A_\mu$ [28]. This can be viewed as a topological constraint, because it is precisely equivalent to [26]. Using the method of Lagrange multipliers, we form the constrained action

$$\mathcal{A} = \mathcal{A}_F^0 + \int \lambda_\mu(\partial_\nu {}^*F^{\mu\nu}) \qquad [39]$$

We can now vary this with respect to $F_{\mu\nu}$, obtaining eqn [40], which implies [27]:

$$F^{\mu\nu} = 2\varepsilon^{\mu\nu\rho\sigma}\partial_\rho\lambda_\sigma \qquad [40]$$

Moreover, the Lagrange multiplier $\lambda$ is exactly the dual potential $\tilde{A}$.

This derivation is entirely dual symmetric, since we can equally well use [27] as constraint for the action $\mathcal{A}_F^0$, now considered as a functional of $^*F^{\mu\nu}$ (eqn [41]), and obtain [26] as the equation of motion:

$$\mathcal{A}_F^0 = \frac{1}{4}\int {}^*F_{\mu\nu}{}^*F^{\mu\nu} \qquad [41]$$

This method applies to the interaction of charges and fields as well. In this case we start with the free field plus free particle action (eqn [42]), where we assume the free particle $m$ to satisfy the Dirac equation,

$$\mathcal{A}^0 = \mathcal{A}_F^0 + \int \bar{\psi}(\mathrm{i}\partial_\mu\gamma^\mu - m)\psi \qquad [42]$$

To fix ideas, let us regard this particle carrying an electric charge $e$ as a monopole of the potential $\tilde{A}_\mu$. Then the constraint we put in is [33], giving

$$\mathcal{A}' = \mathcal{A}^0 + \int \tilde{\lambda}_\mu(\partial_\nu F^{\mu\nu} + j^\mu) \qquad [43]$$

Variation with respect to $^*F$ gives eqn [32], and varying with respect to $\bar{\psi}$ gives

$$(\mathrm{i}\partial_\mu\gamma^\mu - m)\psi = -eA_\mu\gamma^\mu\psi \qquad [44]$$

So, the complete set of equations for a Dirac particle carrying an electric charge $e$ in an electromagnetic field is [32], [33], and [44]. The duals of these equations will describe the dynamics of a Dirac magnetic monopole in an electromagnetic field.

We see from this that the Wu–Yang criterion actually gives us an intuitively clear picture of interactions. The assertion that there is a monopole at a certain spacetime point $x$ means that the gauge field on a 2-sphere surrounding $x$ has to have a certain topological configuration (e.g., giving a nontrivial bundle of a particular class), and if the monopole moves to another point then the gauge field will have to rearrange itself so as to maintain the same topological configuration around the new point. There is thus naturally a coupling between the gauge field and the position of the monopole, or, in physical language, a topologically induced interaction between the field and the charge (Wu and Yang, 1976). Furthermore, this treatment of interaction between field and matter is entirely dual symmetric.

As a side remark, consider that although the action $\mathcal{A}_F^0$ is not immediately identifiable as geometric in nature, the Wu–Yang criterion, by putting the topological constraint and the equation of motion on equal (or dual) footing, suggests that in fact it is geometric in a subtle manner not yet fully understood. Moreover, as pointed out, eqn [40] says that the dual potential is given by the Lagrange multiplier of the constrained action.

## Nonabelian Duality Using Loop Variables

The next natural step is to generalize this duality to the nonabelian Yang–Mills case. Although there is no difficulty in defining $^*F^{\mu\nu}$, which is again given by [3], we immediately come to difficulties in the relation between field and potential; for example, as in eqn [11],

$$F_{\mu\nu}(x) = \partial_\nu A_\mu(x) - \partial_\mu A_\nu(x) + \mathrm{i}g[A_\mu(x), A_\nu(x)]$$

First of all, despite appearances the Yang–Mills equation [45] (in the free-field case) and the Bianchi

identity [46] are not dual-symmetric, because the correct dual of the Yang–Mills equation ought to be given by eqn [47], where $\tilde{D}_\nu$ is the covariant derivative corresponding to a dual potential:

$$D_\nu F^{\mu\nu} = 0 \qquad [45]$$

$$D_\nu {}^*F^{\mu\nu} = 0 \qquad [46]$$

$$\tilde{D}_\nu {}^*F^{\mu\nu} = 0 \qquad [47]$$

Secondly, the Yang–Mills equation, unlike its abelian counterpart [27], says nothing about whether the 2-form ${}^*F$ is closed or not. Nor is the relation [11] about exactness at all. In other words, the Yang–Mills equation does not guarantee the existence of a dual potential, in contrast to the Maxwell case. In fact, Gu and Yang have constructed a counterexample. Because the true variables of a gauge theory are the potentials and not the fields, this means that Yang–Mills theory is not symmetric under the Hodge star operation [3].

Nevertheless, electric–magnetic duality is a very useful physical concept, so one may wish to seek a more general duality transform ($\tilde{\ }$), satisfying the following properties:

1. $(\quad)^{\tilde{}\tilde{}} = \pm(\quad)$.
2. Electric field $F_{\mu\nu} \overset{\sim}{\longleftrightarrow}$ magnetic field $\tilde{F}_{\mu\nu}$.
3. Both $A_\mu$ and $\tilde{A}_\mu$ exist as potentials (away from charges).
4. Magnetic charges are monopoles of $A_\mu$, and electric charges are monopoles of $\tilde{A}_\mu$.
5. $\tilde{\ }$ reduces to $^*$ in the abelian case.

One way to do this is to study the Wu–Yang criterion more closely. This reveals the concept of charges as topological constraints to be crucial even in the pure field case, as can be seen in **Figure 1**. The point to stress is that, in the above abelian case, the condition for the absence of a topological charge (a monopole) exactly removes the redundancy of the variables $F_{\mu\nu}$, and hence recovers the potential $A_\mu$.

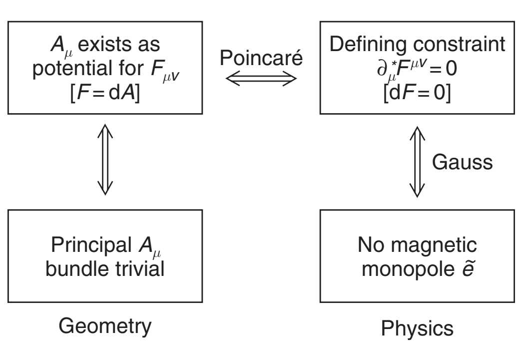



**Figure 1**

Now the nonabelian monopole charge was defined topologically as an element of $\pi_1(G)$, and this definition also holds in the abelian case of U(1), with $\pi_1(\text{U}(1)) = \mathbb{Z}$. So the first task is to write down a condition for the absence of a nonabelian monopole.

To fix ideas, let us consider the group SO(3), whose monopole charges are elements of $\mathbb{Z}_2$, which can be denoted by a sign $\pm$. The vacuum, charge $(+)$ (that is, no monopole) is represented by a closed curve in the group manifold of even winding number, and the monopole charge $(-)$ by a closed curve of odd winding number. It is more convenient, however, to work in SU(2), which is the double cover of SO(3) and which has the topology of $S^3$, as sometimes it is useful to identify the fundamental group of SO(3) with the center of SU(2) and hence consider the monopole charge as an element of this center. There the charge $(+)$ is represented by a closed curve, and the charge $(-)$ by a curve that winds an odd number of "half-times" round the sphere $S^3$. Since these charges are defined by closed curves, it is reasonable to try to write the constraint in terms of loop variables. The treatment presented below is not as rigorous as some others, but the latter are not so well adapted to the problem in hand. Furthermore, it is important to emphasize that this approach aims to generalize electric–magnetic duality to Yang–Mills theory in direct and close analogy to duality in electromagnetism, without any further symmetries with which it may be expedient to enrich the theory. Other approaches are referred to in the next section.

Consider the gauge-invariant Dirac phase factor (or holonomy) $\Phi(C)$ of a loop $C$, which can be written symbollically as a path-ordered exponential:

$$\Phi[\xi] = P_s \exp ig \int_0^{2\pi} ds\, A_\mu(\xi(s))\dot{\xi}^\mu(s) \qquad [48]$$

In eqn [48], we parametrize the loop $C$ as is eqn [49] and a dot denotes differentiation with respect to the parameter $s$.

$$C: \quad \{\xi^\mu(s):\ s = 0 \to 2\pi, \xi(0) = \xi(2\pi) = \xi_0\} \qquad [49]$$

We thus regard loop variables in general as functionals of continuous piecewise smooth functions $\xi$ of $s$. In this way, loop derivatives and loop integrals are just functional derivatives and functional integrals. This means that loop derivatives $\delta_\mu(s)$ are defined by a regularization procedure approximating delta functions with finite bump functions and then taking limits in a definite order. For functional integrals, there exist various regularization procedures, which are treated elsewhere in this Encyclopedia.

Polyakov (1980) introduces the logarithmic loop derivative of $\Phi[\xi]$:

$$F_\mu[\xi|s] = \frac{i}{g}\Phi^{-1}[\xi]\delta_\mu(s)\Phi[\xi] \qquad [50]$$

This acts as a kind of "connection" in loop space since it tells us how the phase of $\Phi[\xi]$ changes from one loop to a neighbouring loop. One can go a step further and define its "curvature" in direct analogy with $F_{\mu\nu}(x)$ by

$$G_{\mu\nu}[\xi|s] = \delta_\nu(s)F_\mu[\xi|s] - \delta_\mu(s)F_\nu[\xi|s] + ig[F_\mu[\xi|s], F_\nu[\xi|s]] \qquad [51]$$

It can be shown that by using the $F_\mu[\xi|s]$ we can rewrite the Yang–Mills action as eqn [52], where the normalization factor $\bar{N}$ is an infinite constant:

$$\mathcal{A}_F^0 = -\frac{1}{4\pi\bar{N}}\int \delta\xi \int_0^{2\pi} \mathrm{d}s\, \mathrm{tr}\{F_\mu[\xi|s]F^\mu[\xi|s]\}|\dot{\xi}(s)|^{-2} \qquad [52]$$

However, the true variables of the theory are still the $A_\mu$. They represent 4 functions of a real variable, whereas the loop connections represent 4 functionals of the real function $\xi(s)$. Just as in the case of the $F_{\mu\nu}$, these $F_\mu[\xi|s]$ have to be constrained so as to recover $A_\mu$, but this time much more severely.

It turns out that, in pure Yang–Mills theory, the constraint that says there are no monopoles ([53]) also removes the redundancy of the loop variables, exactly as in the abelian case,

$$G_{\mu\nu}[\xi|s] = 0 \qquad [53]$$

That this condition is necessary is easy to see by simple algebra. The proof of the converse of this "extended Poincaré lemma" is fairly lengthy. Granted this, we can now apply the Wu–Yang criterion to the action [52] and derive the Polyakov equation [54], which is the loop version of the Yang–Mills equation:

$$\delta_\mu(s)F^\mu[\xi|s] = 0 \qquad [54]$$

In the presence of a monopole charge (–), the constraint [53] will have a nonzero right-hand side,

$$G_{\mu\nu}[\xi|s] = -J_{\mu\nu}[\xi|s] \qquad [55]$$

The loop current $J_{\mu\nu}[\xi|s]$ can be written down explicitly. However, its global form is much easier to understand. Recall that $F^\mu[\xi|s]$ can be thought of as a loop connection, for which we can form its "holonomy." This is defined for a closed (spatial) surface $\Sigma$ (enclosing the monopole), parametrized by a family of closed curves $\xi_t(s), t = 0 \to 2\pi$. The "holonomy" $\Theta_\Sigma$ is then the total change in phase of $\Phi[\xi_t]$ as $t \to 2\pi$, and thus equals the charge (–).

To formulate an electric–magnetic duality that is applicable to nonabelian theory, one defines yet another set of loop variables. Instead of the Dirac phase factor $\Phi[\xi]$ for a complete curve [48], we consider the parallel phase transport for part of a curve from $s_1$ to $s_2$:

$$\Phi_\xi(s_2, s_1) = P_s \exp ig\int_{s_1}^{s_2} \mathrm{d}s A_\mu(\xi(s))\dot{\xi}^\mu(s) \qquad [56]$$

Then the new variables are defined by [57].

$$E_\mu[\xi|s] = \Phi_\xi(s,0)F_\mu[\xi|s]\Phi_\xi^{-1}(s,0) \qquad [57]$$

These are not gauge invariant like $F_\mu[\xi|s]$ and may not be as useful in general, but seem more convenient for dealing with duality.

Using these variables, we now define their dual $\tilde{E}_\mu[\eta|t]$ according to

$$\omega^{-1}(\eta(t))\tilde{E}_\mu[\eta|t]\omega(\eta(t)) = -\frac{2}{\bar{N}}\varepsilon_{\mu\nu\rho\sigma}\dot{\eta}^\nu(t)\int \delta\xi\, \mathrm{d}s E^\rho[\xi|s]\dot{\xi}^\sigma(s)\dot{\xi}^{-2}(s) \times \delta(\xi(s) - \eta(t)) \qquad [58]$$

In eqn [58], $\omega(x)$ is a (local) rotation matrix transforming from the frame in which the orientation in internal symmetry space of the fields $E_\mu[\xi|s]$ are measured to the frame in which the dual fields $\tilde{E}_\nu[\eta|t]$ are measured. It can be shown that this dual transform satisfies all five of the required conditions listed earlier.

Electric–magnetic duality in Yang–Mills theory is now fully reestablished using this generalized duality. We have the dual pairs of equations [59]–[60] and [61]–[62]:

$$\delta_\nu E_\mu - \delta_\mu E_\nu = 0 \qquad [59]$$

$$\delta^\mu E_\mu = 0 \qquad [60]$$

$$\delta^\mu \tilde{E}_\mu = 0 \qquad [61]$$

$$\delta_\nu \tilde{E}_\mu - \delta_\mu \tilde{E}_\nu = 0 \qquad [62]$$

Equation [59] guarantees that the potential $A$ exists, and so is equivalent to [53], and hence is the nonabelian analog of [26]; while equation [60] is equivalent to the Polyakov version of Yang–Mills equation [54], and hence is the nonabelian analog of [27]. Equation [61] is equivalent by duality to [59] and is the dual Yang–Mills equation. Similarly equation [62] is equivalent to [60], and guarantees the existence of the dual potential $\tilde{A}$.

The treatment of charges using the Wu–Yang criterion also follows the abelian case, and will not be further elaborated here. For this and further details, the reader is referred to the orginal papers (Chan and Tsou 1993, 1999).

Also, just as in the abelian case, the gauge symmetry is doubled: from the group $G$ we deduce that the full gauge symmetry is in fact $G \times \tilde{G}$, but that the physical degrees of freedom remain the same.

The above exposition establishes electric–magnetic duality in Yang–Mills theory only for classical fields. A hint that this duality persists at the quantum level comes from the work of 't Hooft (1978) on confinement. There he introduces two loop quantities $A(C)$ and $B(C)$ that are operators in the Hilbert space of quantum states satisfying the commutation relation [63] for an SU($N$) gauge theory, where $n$ is the linking number between the two (spatial) loops $C$ and $C'$:

$$A(C)B(C') = B(C')A(C)\exp(2\pi i n/N) \qquad [63]$$

The order or Wilson operator is given explicitly by $A(C) = \operatorname{tr}\Phi(C)$. These two operators play dual roles in the sense of electric–magnetic duality:

- $A(C)$ measures the magnetic flux through $C$ and creates electric flux along $C$.
- $B(C)$ measures the electric flux through $C$ and creates magnetic flux along $C$.

By defining the disorder operator $B(C)$ as the Wilson operator corresponding to the dual potential $\tilde{A}$ obtained above, one can prove the commutation relation [63], thus showing that these classical fields, when promoted to operators, retain their duality relation. Furthermore, there is a remarkable relation between the two (abstractly identical) gauge groups, in that if one is confined then the dual must be broken (that is, in the Higgs phase). This result is known as 't Hooft's theorem.

The doubling of gauge symmetry, together with 't Hooft's theorem, has been applied to the confined colour group SU(3) of quantum chromodynamics (QCD), in the Dualized Standard Model, to solve the puzzle of the existence of exactly three generations of fermions, with good observational support, by identifying the (necessarily broken) dual SU(3) with the generation symmetry (Chan and Tsou, 2002).

## Other Treatments of Nonabelian Duality

Since Yang–Mills theory is not symmetric under the Hodge $^*$-operation, there are several routes one can take to generalize the concept of electric–magnetic duality to the nonabelian case. What was presented in the last section is a modification of the $^*$-operation so as to restore this symmetry for Yang–Mills theory, keeping to the original gauge structure as much as possible. However, Yang–Mills theory as used today in particle and field theories are usually embedded in theories with more structures. In the simplest case we have the Standard Model of Particle Physics, which describes all of particle interactions (except gravity) and which has the gauge group usually written as $SU(3) \times SU(2) \times U(1)$, corresponding to the SU(3) of strong interaction and $SU(2) \times U(1)$ of electroweak interaction. [Strictly speaking, it is $(SU(3) \times SU(2) \times U(1))/\mathbb{Z}_6$, if we have the standard particle spectrum.] However, the former group is confined and the latter broken. The breaking is usually effected by introducing scalar fields called Higgs fields into the theory.

Besides the experimentally well-tested Standard Model, there are many theoretically popular models of gauge theory in which supersymmetry is postulated, thereby introducing extra symmetries into the theory. Many of these are remnants of string theory, and are usually envisaged as gauge theories in a spacetime dimension higher than 4.

Because of the extra structures and increased symmetries in these theories, there is quite a proliferation of concepts of duality, which could all be thought of as generalizations of abelian electric–magnetic duality (Schwarz, 1997). They come under the names of Seiberg–Witten duality, S-duality, T-duality, mirror symmetry, and so on. All these other aspects of duality have their own entries in this Encyclopedia.

*See also:* AdS/CFT Correspondence; Duality in Topological Quantum Field Theory; Four-Manifold Invariants and Physics; Large-*N* Dualities; Measure on Loop Spaces; Mirror Symmetry: a Geometric Survey; Nonperturbative and Topological Aspects of Gauge Theory; Seiberg–Witten theory; Standard Model of Particle Physics.

## Further Reading

Chan Hong-Mo and Tsou Sheung Tsun (1993) *Some Elementary Gauge Theory Concepts*. Singapore: World Scientific.

Chan Hong-Mo and Tsou Sheung Tsun (1999) Nonabelian generalization of electric–magnetic duality – a brief review. *International Journal of Modern Physics* A14: 2139–2172.

Chan Hong-Mo and Tsou Sheung Tsun (2002) Fermion generations and mixing from dualized standard model. *Acta Physica Polonica* B33(12): 4041–4100.

Polyakov AM (1980) Gauge fields as rings of glue. *Nuclear Physics* B164: 171–188.

Schwarz John H (1997) Lectures on superstring and M theory dualities. *Nuclear Physics Proceedings Supplements* 55B: 1–32.

't Hooft G (1978) On the phase transition towards permanent quark confinement. *Nuclear Physics* B138: 1–25.

Wu Tai Tsun and Yang Chen Ning (1976) Dirac monopoles without strings: classical Lagrangian theory. *Physical Review* D14: 437–445.

Yang Chen Ning and Mills RL (1954) Conservation of isotopic spin and isotopic gauge invariance. *Physical Review* 96: 191–195.

# Electroweak Theory

**K Konishi**, Università di Pisa, Pisa, Italy



## Introduction

The discovery of the electroweak theory crowned long years of investigation on weak interactions. The key earlier developments included Fermi's phenomenological four-fermion interactions for the $\beta$-decay, discovery of parity violation and establishment of $V - A$ structure of the weak currents, the Feynman–Gell–Mann conserved vector current (CVC) hypothesis, current algebra and its beautiful applications in the 1960s, Cabibbo mixing and lepton–hadron universality, and finally, the proposal of intermediate vector bosons (IVBs) to mitigate the high-energy behavior of the pointlike Fermi's interaction theory.

It turned out that the scattering amplitudes in IVB theory still generally violated unitarity, due to the massive vector boson propagator,

$$\frac{-g^{\mu\nu} + q^{\mu}q^{\nu}/M^2}{q^2 - M^2 + \mathrm{i}\epsilon}$$

The electroweak theory, known as Glashow–Weinberg–Salam (GWS) theory (Weinberg 1967, Salam 1968, Taylor 1976), was born through the attempts to make the hypothesis of IVBs for the weak interactions such that it is consistent with unitarity.

The GWS theory contains, and is in a sense a generalization of, quantum electrodynamics (QED) which was earlier successfully established as the quantum theory of electromagnetism in interaction with matter. GWS theory describes the weak and electromagnetic interactions in a single, unified gauge theory with gauge group

$$\mathrm{SU_L}(2) \times \mathrm{U}(1) \qquad [1]$$

Part of this gauge symmetry is realized in the so-called "spontaneously broken" mode; only a $\mathrm{U_{EM}}(1) \subset \mathrm{SU_L}(2) \times \mathrm{U}(1)$ subgroup, corresponding to the usual local gauge symmetry of the electromagnetism, remains manifest at low energies, with a massless gauge boson (photon). The other three gauge bosons $W^{\pm}, Z$, are massive, with masses $\sim$80.4 and 91.2 GeV, respectively.

The theory is renormalizable, as conjectured by S Weinberg and by A Salam, and subsequently proved by G 't Hooft (1971), and makes well-defined predictions order by order in perturbation theory.

Since the experimental observation of neutral currents (a characteristic feature of the Weinberg–Salam theory which predicts an extra, neutral massive vector boson, $Z$, as compared to the naive IVB hypothesis) at Gargamelle bubble chamber at CERN (1973), the theory has passed a large number of experimental tests. The first basic confirmation also included the discovery of various new particles required by the theory: the charm quark (SLAC, BNL, 1974), the bottom quark (Fermilab, 1977), and the tau ($\tau$) lepton (SLAC, 1975). The heaviest top quark, having mass about two hundred times that of the proton, was found later (Fermilab, 1995). The direct observation of W and Z vector bosons was first made by UA1 and UA2 experiments at CERN (1983).

The GWS theory is today one of the most precise and successful theories in physics. Even more important, perhaps, together with quantum chromodynamics (QCD), which is a SU(3) (color) gauge theory describing the strong interactions (which bind quarks into protons and neutrons, and the latter two into atomic nuclei), it describes correctly – within the present experimental and theoretical uncertainties – all the presently known fundamental forces in Nature, except gravity. The $\mathrm{SU}(3)_{\mathrm{QCD}} \times (\mathrm{SU_L}(2) \times \mathrm{U}(1))_{\mathrm{GWS}}$ theory is known as the standard model (SM).

Both the electroweak (GSW) theory and QCD are gauge theories with a nonabelian (noncommutative) gauge group. This type of theories, known as Yang–Mills theories, can be constructed by generalizing the well-known gauge principle of QED to more general group transformations. It is a truly remarkable fact that all of the fundamental forces known today (apart from gravity) are described by Yang–Mills theories, and in this sense a very nontrivial unification can be said to underlie the basic laws of Nature (G 't Hooft).

There are further deep and remarkable conditions (anomaly cancellations), satisfied by the structure of the theory and by the charges of experimentally known spin-1/2 elementary particles (see Tables 1 and 2), which guarantees the consistency of the theory as a quantum theory.

It should be mentioned, however, that the recent discovery of neutrino oscillations (SuperKamiokande (1998), SNO, KamLAND, K2K experiments), which proved the neutrinos to possess nonvanishing masses, clearly indicates that the standard GWS theory must be extended, in an as yet unknown way.

**Table 1** Quarks and their charges

| *Quarks* | $SU_L(2)$ | $U_Y(1)$ | $U_{EM}(1)$ |
|---|---|---|---|
| $\begin{pmatrix} u_L \\ d'_L \end{pmatrix}, \begin{pmatrix} c_L \\ s'_L \end{pmatrix}, \begin{pmatrix} t_L \\ b'_L \end{pmatrix}$ | $\underline{2}$ | $\frac{1}{3}$ | $\begin{pmatrix} \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$ |
| $u_R, c_R, t_R$ | $\underline{1}$ | $\frac{4}{3}$ | $\frac{2}{3}$ |
| $d_R, s_R, b_R$ | $\underline{1}$ | $-\frac{2}{3}$ | $-\frac{1}{3}$ |

The primes indicate that the mass eigenstates are different from the states transforming as multiplets of $SU_L(2) \times U_Y(1)$. They are linearly related by CKM mixing matrix.

**Table 2** Leptons and their charges

| *Leptons* | $SU_L(2)$ | $U_Y(1)$ | $U_{EM}(1)$ |
|---|---|---|---|
| $\begin{pmatrix} \nu'_{eL} \\ e_L \end{pmatrix}, \begin{pmatrix} \nu'_{\mu L} \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu'_{\tau L} \\ \tau_L \end{pmatrix}$ | $\underline{2}$ | $-1$ | $\begin{pmatrix} 0 \\ -1 \end{pmatrix}$ |
| $e_R, \mu_R, \tau_R$ | $\underline{1}$ | $-2$ | $-1$ |

The primes indicate again that the mass eigenstates are in different from the states transforming as multiplets of $SU_L(2) \times U_Y(1)$, as required by the observed neutrino oscillations.

The following is a brief summary of the GWS theory, its characteristic features, its implications to the symmetries of Nature, the status of the precision tests, and its possible extensions.

## GWS Theory

All the presently known elementary particles (except for the gauge bosons $W^{\pm}, Z, \gamma$, the gluons, the graviton, possibly right-handed neutrinos) are listed in Tables 1–3 together with their charges with respect to the $SU_L(2) \times U(1)$ gauge group.

A doublet of Higgs scalar particles is included even though the physical component (which should appear as an ordinary scalar particle) has not yet been experimentally observed.

The Lagrangian is given by

$$\mathcal{L} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{quarks}} + \mathcal{L}_{\text{leptons}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}} + \mathcal{L}_{\text{g.f.}} + \mathcal{L}_{\text{ghosts}}$$

The gauge kinetic terms are

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4}\sum_{a=1}^{3} F^a_{\mu\nu}F^{a\mu\nu} - \frac{1}{4}G_{\mu\nu}G^{\mu\nu}$$

**Table 3** Higgs doublet scalars and their charges

| *Higgs doublet* | $SU_L(2)$ | $U_Y(1)$ | $U_{EM}(1)$ |
|---|---|---|---|
| $\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$ | $\underline{2}$ | 1 | $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ |

where

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g\epsilon^{abc}A^b_\mu A^c_\nu$$
$$G_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$$

are $SU_L(2) \times U(1)$ gauge field tensors; $\mathcal{L}_{\text{g.f.}}$ and $\mathcal{L}_{\text{FP}}$ are the so-called gauge-fixing term and Faddeev–Popov ghost term, needed to define the gauge-boson propagators appropriately and to eliminate certain unphysical contributions. The gauge invariance of the theory is ensured by a set of identities (A Slavnov, J C Taylor). The quark kinetic terms have the form

$$\mathcal{L}_{\text{quarks}} = \sum_{\text{quarks}} \bar{\psi} i\gamma^\mu \mathcal{D}_\mu \psi$$

where $\mathcal{D}_\mu$ are appropriate covariant derivatives,

$$\mathcal{D}_\mu q_L = \left(\partial_\mu - \frac{ig}{2}\tau \cdot A_\mu - \frac{ig'}{6}B_\mu\right)q_L$$

for the left-handed quark doublets,

$$\mathcal{D}_\mu u_R = \left(\partial_\mu - \frac{2ig'}{3}B_\mu\right)u_R$$
$$\mathcal{D}_\mu d_R = \left(\partial_\mu + \frac{ig'}{3}B_\mu\right)d_R$$

and similarly for other "up" quarks $c_R$ (charm) and $t_R$ (top), and "down" quarks, $s_R$ (strange), and $b_R$ (bottom). Analogously, the lepton kinetic terms are given by

$$\mathcal{L}_{\text{leptons}} = \sum_{i=1}^{3} \bar{\psi}^i i\gamma^\mu \mathcal{D}_\mu \psi^i = \sum_{i=1}^{3} \bar{\psi}^i_L i\gamma^\mu \left(\partial_\mu - ig\frac{\tau^a A^a_\mu}{2} + \frac{ig'}{2}B_\mu\right)\psi^i_L + \sum_{i=1}^{3} \bar{\psi}^i_R i\gamma^\mu (\partial_\mu + ig' B_\mu)\psi^i_R$$

where $\psi^i (i = 1, 2, 3)$ indicate the $e, \mu, \tau$ lepton families; finally, the parts involving the Higgs fields are

$$\mathcal{L}_{\text{Higgs}} = \mathcal{D}_\mu \phi * \mathcal{D}^\mu \phi + V(\phi, \phi^\dagger)$$
$$V(\phi, \phi^\dagger) = -\mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2$$

and

$$\mathcal{L}_{\text{Yukawa}} = \sum_{i,j=1}^{3}\left[g^{ij}_d \bar{q}^i_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} d^j_R + g^{ij}_u \bar{q}^i_L \begin{pmatrix} \phi^{0*} \\ -\phi^- \end{pmatrix} u^j_R\right] + \text{h.c.} + \sum_{i,j=1}^{3}\left[g^i_e \bar{\psi}^i_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \psi^i_R\right] + \text{h.c.} \quad [2]$$

For $\mu^2 < 0$, the Higgs potential has a minimum at

$$\langle \phi^\dagger \phi \rangle = \langle |\phi^+|^2 + |\phi^0|^2 \rangle = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2} \neq 0$$

By choosing conveniently the direction of the Higgs field, its vacuum expectation value (VEV) is expressed as

$$\left\langle \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \right\rangle = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}, \quad v \equiv \sqrt{-\frac{\mu^2}{\lambda}} \qquad [3]$$

The physical properties of Higgs and gauge bosons are best seen by choosing the so-called unitary gauge,

$$\Phi(x) = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = e^{i\zeta^a(x)\tau^a/v} \begin{pmatrix} 0 \\ (v + \eta(x))/\sqrt{2} \end{pmatrix} \equiv U(\zeta)\Phi'(x)$$

$$\psi_L = U(\zeta)\psi'_L, \qquad \psi_R = \psi'_R$$

$$A_\mu = U(\zeta)\left(A'_\mu + \frac{i}{g}\partial_\mu\right)U^{-1}(\zeta), \qquad A_\mu \equiv \frac{\tau^a A^a_\mu}{2}$$

and expressing everything in terms of primed variables. It is easy to see that

1. There is one physical scalar (Higgs) particle with mass,

$$m_\eta = \sqrt{-2\mu^2} \qquad [4]$$

2. The Higgs kinetic term $(\mathcal{D}\phi'^\dagger)(\mathcal{D}\phi')$ produces the gauge-boson masses

$$M^2_{W^\pm} = \frac{g^2 v^2}{4}, \qquad M^2_Z = \frac{v^2}{4}(g^2 + g'^2) \qquad [5]$$

3. The physical gauge bosons are the charged $W^\pm$, and two neutral vector bosons described by the fields

$$Z_\mu = \cos\theta_W A_{3\mu} - \sin\theta_W B_\mu,$$
$$A_\mu = \sin\theta_W A_{3\mu} + \cos\theta_W B_\mu$$

where the mixing angle

$$\theta_W = \tan^{-1}\frac{g'}{g} \quad \left(\sin\theta_W = \frac{g'}{\sqrt{g^2 + g'^2}}\right)$$

is known as the Weinberg angle. The massless $A_\mu$ field describes the photon.

## Fermi Interactions and Neutral Currents

The fermions interact with gauge bosons through the charge and neutral currents

$$\mathcal{L} = \frac{g}{2}(J_{-\mu}W^\mu_+ + J_{+\mu}W^\mu_-) + \mathcal{L}^{\text{n.c.}} \qquad [6]$$

$$\mathcal{L}^{\text{n.c.}} = gJ^3_\mu A^{3\mu} + \frac{g'}{2}J^Y_\mu B^\mu$$
$$= eJ^\mu_{\text{em}}A_\mu + \frac{g}{\cos\theta_W}J^0_\mu Z^\mu \qquad [7]$$

where

$$J_{+\mu} = \sum \bar{\psi}_L \gamma_\mu \tau^+ \psi_L$$
$$= \frac{1}{2}\sum \bar{\psi}\gamma_\mu \tau^+(1 - \gamma_5)\psi$$
$$\equiv \frac{1}{2}J^{V-A}_{+\mu} \qquad [8]$$

corresponds to the standard charged current, and

$$J^0_\mu = J^3_\mu - \sin^2\theta_W J^{\text{em}}_\mu \qquad [9]$$

is the neutral current to which the $Z$ boson is coupled ($J^3_\mu = (1/2)\sum \bar{\psi}_L\gamma_\mu\tau^3\psi_L$ and $J^{\text{em}}_\mu$ is the electromagnetic current). The model thus predicts the existence of neutral current processes, mediated by the $Z$ boson, such as $\nu_\mu e \to \nu_\mu e$ or $\bar{\nu}_\mu e \to \bar{\nu}_\mu e$, with cross section of the same order of that for the charged current process, $\bar{\nu}_e e \to \bar{\nu}_e e$, but with a characteristic L–R asymmetric couplings depending on the Weinberg angle. By eqn [9] appropriate ratios of cross sections, such as $\sigma(\nu_\mu e \to \nu_\mu e)/\sigma(\bar{\nu}_\mu e \to \bar{\nu}_\mu e)$, can be used to measure $\sin^2\theta_W$.

The exchange of heavy $W$ bosons generates an effective current–current interaction at low energies:

$$\mathcal{L}^{\text{c.c}}_{\text{eff}} = -\frac{g^2}{2M^2_W}J_{-\mu}J^\mu_+$$

the well-known Fermi–Feynman–Gell-Mann Lagrangian $-\frac{G_F}{\sqrt{2}}J^\dagger_{V-A\mu}J^\mu_{V-A}$, with

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8M^2_W}$$

This means that the Higgs VEV must be taken to be

$$v = 2^{-1/4}G_F^{-1/2} \simeq 246\ \text{GeV} \qquad [10]$$

## Masses

It is remarkable that "all" known masses of the elementary particles – except perhaps those of the neutrino masses – are generated in GWS theory through the spontaneous breakdown of $SU_L(2) \times U(1)$ symmetry, through the Higgs VEV (eqns [3] and [10]). The boson masses are given by [4] and [5]. Note that the relation

$$\rho = \frac{M^2_W}{M^2_Z\cos^2\theta_W} = 1 + O(\alpha)$$

reflects an accidental SO(3) symmetry present (note the SO(4) symmetry of the Higgs potential in the limit $\alpha \to 0$, before the spontaneous breaking) in the model, called custodial symmetry. This is a characteristic, model-dependent feature of the minimal model, not

necessarily required by the gauge symmetry. This relation is well met experimentally, although a quantitative discussion requires the choice of the renormalization scheme (including the definition of $\sin\theta_W$ itself) and check of consistency with various other data.

The fermions get mass through the Yukawa interactions (eqn [2]); the fermion masses are arbitrary parameters of the model and cannot be predicted within the GWS theory. An important feature of this mechanism is that the coupling of the physical Higgs particle to each fermion is proportional to the mass of the latter. This should give a clear, unambiguous experimental signature for the Higgs scalar of the minimal GWS model.

The recent discovery of nonvanishing neutrino masses requires the theory to be extended. Actually, there is a natural way to incorporate such masses in the standard GWS model, by a minimal extension. As the right-handed neutrinos, if they exist, are entirely neutral with respect to the $SU_L(2)\times U(1)$ gauge symmetry, they do not need its breaking to have mass. In other words, $\nu_R$ may get Majorana masses, $\sim M_R\nu_R\nu_R$, by some yet unknown mechanism, much larger than those of other fermions (such a mechanism is quite naturally present in some grand unified models). If now the Yukawa couplings are introduced as for the quarks and for the down leptons, then the Dirac mass terms result upon condensation of the Higgs field, and the neutrino mass matrix would take the form, for one flavor (in the space of $(\nu_L,\bar{\nu}_R)$):

$$\begin{pmatrix} 0 & m_D \\ m_D & M_R \end{pmatrix} \qquad [11]$$

**Table 4** Quark masses

| $u$ (MeV) | $c$ (GeV) | $t$ (GeV) | $d$ (MeV) | $s$ (MeV) | $b$ (GeV) |
|---|---|---|---|---|---|
| 1.5–4 | 1.15–1.35 | $174.3\pm5.1$ | 4–8 | 80–130 | 4.1–4.4 |

**Table 5** Leptons masses

| $\nu_e$(eV) | $\nu_\mu$(MeV) | $\nu_\tau$(MeV) |
|---|---|---|
| $<3$ | $<0.19$ | $<18.2$ |
| $e$ (MeV) | $\mu$ (MeV) | $\tau$ (MeV) |
| $0.510\,998\,92\pm4\times10^{-8}$ | $105.658369\pm9\times10^{-6}$ | $1776.99\pm0.26$ |

**Table 6** Gauge-boson masses

| *Photon* | *Gluons* | $W^\pm$(GeV) | $Z$ (GeV) |
|---|---|---|---|
| 0 | 0 | $80.425\pm0.038$ | $91.1876\pm0.0021$ |

If the Dirac masses are assumed to be of the same order of those of the quarks and if the right-handed Majorana masses $M_R$ are far larger, for example, of the order of the grand unified scale, $O(10^{16}\,\text{GeV})$, then diagonalization of the mass matrix would give, for the physical masses of the left-handed neutrinos, $\sim m_D^2/M_R \ll m_D$, much smaller than other fermion masses, quite naturally ("see-saw" mechanism).

**CKM Quark Mixing** As there is *a priori* no reason why the weak-interaction eigenstates should be equal to the mass eigenstates, the Yukawa couplings in eqn [2] are in general nondiagonal matrices in the flavor. Suppose that the the weak base for the quarks is given in terms of the mass eigenstates (in which quark masses are made diagonal), by unitary transformations

$$u_{Li}=\sum_j V^{\text{up}}_{ij}\tilde{u}_{Lj}, \qquad d_{Li}=\sum_j V^{\text{down}}_{ij}\tilde{d}_{Lj}$$

then the interaction terms with $W^\pm$ bosons [6] can be cast in the form (Kobayashi and Maskawa 1972)

$$\begin{aligned}\mathcal{L}^{W\text{-exc}} &= \bar{u}^i_L\gamma^\mu W^+_\mu U^{(\text{CKM})}_{ij} d^j_L \\ &\quad + \bar{d}^k_L\gamma^\mu W^-_\mu U^{(\text{CKM})\dagger}_{k\ell} u^\ell_L \end{aligned} \qquad [12]$$

where $U^{\text{CKM}}_{ij}\equiv(V^{\text{up}\dagger}\cdot V^{\text{down}})_{ij}$ is called Cabibbo–Kobayashi–Maskawa (CKM) matrix. It can be parametrized in terms of three Euler angles and one phase

$$U=\begin{pmatrix} U_{ud} & U_{us} & U_{ub} \\ U_{cd} & U_{cs} & U_{cb} \\ U_{td} & U_{ts} & U_{tb}\end{pmatrix}$$
$$=\begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta_{13}} \\ -s_{12}c_{23}-c_{12}s_{23}s_{13}e^{i\delta_{13}} & c_{12}c_{23}-s_{12}s_{23}s_{13}e^{i\delta_{13}} & s_{23}c_{13} \\ s_{12}s_{23}-c_{12}c_{23}s_{13}e^{i\delta_{13}} & c_{12}s_{23}-s_{12}c_{23}s_{13}e^{i\delta_{13}} & c_{23}c_{13}\end{pmatrix} \qquad [13]$$

where $c_{12}=\cos\theta_{12}, s_{23}=\sin\theta_{23}$, etc. The requirement that charge–current weak processes are all described by these matrix elements, satisfying the unitarity relation,

$$\sum_\ell U^{\text{CKM}}_{i\ell}U^{\text{CKM}\dagger}_{\ell k}=\delta_{ik} \qquad [14]$$

gives a very stringent test for the validity of the model.

### *CP* Violation

*CP* (product of charge conjugation and parity transformation) invariance is an approximate symmetry of Nature. Although it is known to be broken by very tiny amounts only, the exact extent and the nature of *CP* violation can have far-reaching consequences.

*CP* violation has first been discovered by Cronin and Fitch (BNL, 1964) in the *K*-meson system; more precise information on the nature of *CP* violation from the neutral kaon decays has been obtained more recently (2000) in NA48 (CERN) and KTeV (Fermilab) experiments. *CP* violation has been established in the *B*-meson systems as well, very recently (2002), by Babar experiments at SLAC and Belle experiments at KEK.

Through the so-called *CPT* theorem, *CP* invariance (or violation) is closely related to the *T* (time-reversal invariance) symmetry. Also, *CP* noninvariance is one of the conditions needed in the cosmological baryon number generation (baryogenesis).

In the GWS theory, with three families of quark flavors (six quarks), there is just one source of *CP* violation: the phase $\delta_{13}$ appearing in the CKM matrix (eqn [13]). For $\delta \neq 0, \pi$, *W*-exchange interactions [12] induce *CP* violation. The earlier and more recent experimental data on $K^0 - \bar{K}^0$ mixing and $K_{L,S}$ decay data appear to be compatible with the CKM mechanism for *CP* violation, but a quantitative comparison with the SM remains somewhat hindered by the difficulty of estimating certain strong interaction effects. The recent confirmation of *CP* violation in *B* systems is made in the context of a global fit with the SM predictions such as the "unitarity triangle" relations, for example,

$$1 + \frac{U_{ud}\,U^*_{ub}}{U_{cd}\,U^*_{cb}} + \frac{U_{td}\,U^*_{tb}}{U_{cd}\,U^*_{cb}} = 0 \qquad [15]$$

(eqn [14]), and by combining data from kaon deays, charmed meson decays, *B* meson decay and mixings, etc., and is a part of direct tests of the GWS model, with nonvanishing *CP* violation CKM

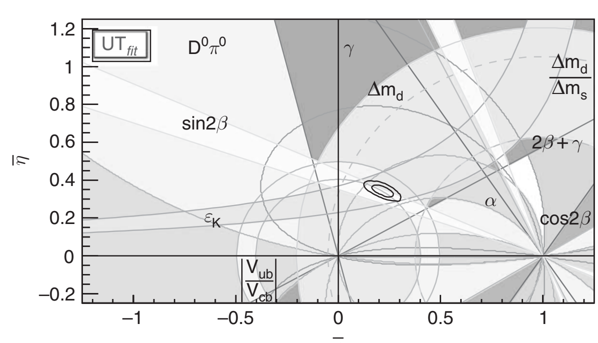



**Figure 1** Unitarity triangle test (Eq. (15)). The small ellipses represent 68% and 95% probability zones for the apex corresponding to $U_{ud}U^*_{ub}/U_{cd}U^*_{cb}$. Reproduced from M. Bona *et al.* (2005) The 2004 UTfit collaboration report on the status of the unitarity triangle in the standard model. Journal of High Energy Physics. 0507: 028–059 (hep-ph/0501199), with permission from IoP Publishing Ltd and the UTfit collaboration.

phase (eqn [16] and **Figure 1**). Recent evidence for nonzero neutrino masses and mixings opens the way to possible *CP* violation in the leptonic processes as well.

Finally, within the SM including strong interactions, there is one more source of *CP* violation: the so-called $\theta$ (vacuum) parameter of QCD.

### *B* and *L* Nonconservation

Another set of approximate symmetries in Nature are the baryon and lepton number conservations. In the electroweak theory, these global symmetries are exact to all orders of perturbation theory. Nonperturbative effects (a sort of barrier penetration in gauge field space) however violate both *B* and *L*; the combination *B–L* is conserved even nonperturbatively though. The nonperturbative electroweak baryon number violation is an extremely tiny effect, the amplitude being proportional to the typical tunneling factor $e^{-2\pi/\alpha}$, but the process is unsuppressed at finite temperatures as might have been experienced by the universe at some early stage after big bang.

*B* or *L* nonconservation can also arise naturally at high energy scales, if the electroweak theory is embedded as the low-energy approximation in a grand unified model. The experimental lower limit of proton lifetime, $\tau_P \geq 10^{32}$ years, from Kamiokande experiments, however severely restricts acceptable models of this type (the simplest SU(5) model is already ruled out).

On the other hand, cosmological baryogenesis requires sufficient amount of baryon number violation, at least in some stage of cosmological expansion. Detailed analyses suggest that the standard electroweak transition might not in itself explain the baryon number $n_P/n_\gamma \sim 10^{-10}$ observed in the present universe. Recent observations of neutrino oscillations suggest the right-handed Majorana-type neutrino masses to be present, which violate the lepton number *L*. In such a case it might be possible that the correct amount of baryon number excess would be generated, through the leptogenesis.

### Global Fit

Various relations exist at the tree level among the masses, scattering cross sections, decay rates, various asymmetries, etc., which can be read off or calculated from the formulas given earlier. These quantities receive corrections at higher orders, and the experimental checks of these modified relations provide precision tests of the model on the one hand, and possibly a hint for new physics, if there is any discrepancy with the prediction. Very often the amplitudes of interest

receive important contributions due to strong interactions, which are difficult to estimate.

The basic parameters of the model, apart from the Higgs mass, and fermion masses and mixing parameters, can be taken to be (1) the fine structure constant, $\alpha = 1/137.035\,999\,11(46)$; (2) the Fermi constant $G_F = 1.166\,37 \times 10^{-5}\,\text{GeV}^{-2}$ (which can be determined from the muon lifetime), and the Z-boson mass, $M_Z = 91.1876 \pm 0.0021$ GeV (observed directly at LEP). $M_W$ and $\sin^2\theta_W$ are then calculable numbers, in terms of these quantities, and depending on $m_t$ (measured independently by CDF and DØ experiments at Fermilab) and on the unknown $M_H$.

Such precision tests of the GWS model are being made, combining the analyses of various decay rates and asymmetries in $B$-meson systems at $B$ factories and in colliders, production and decays of $Z$ and $W$ bosons, elastic $\nu e$ or $\bar{\nu} e$ scatterings, elastic $\nu p$ or $\bar{\nu} p$ scatterings, deep inelastic lepton nucleon (or deuteron) scatterings, the muon anomalous magnetic moment, atomic parity violation experiments, etc.

An overall fit to the data gives an excellent agreement, with the input parameters

$$M_H = 113^{+56}_{-40}\,\text{GeV}, \quad m_t = 176.9 \pm 4.0\,\text{GeV}, \quad \alpha_s(M_Z) = 0.1213 \pm 0.0018$$

For instance (in GeV),

$$M_W = 80.390 \pm 0.018 \quad \text{vs.} \quad 80.412 \pm 0.042 \; \text{(exp. value (LEP))}$$
$$\Gamma_Z = 2.4972 \pm \pm 0.0012 \quad \text{vs.} \quad 2.4952 \pm 0.0023 \; \text{(exp. value)}$$

For $\sin^2\theta_W$ (defined in the so-called $\overline{MS}$ scheme) all data give consistently the value

$$\sin^2\theta_W = 0.231\,20 \pm 0.000\,15$$

(a slightly larger value is reported by an $\nu N$ experiment at Fermilab).

The unitarity-triangle tests of the SM and determination of CKM matrix have already been mentioned. The results of global fit can be summarized in Figure 1, and by the angles

$$\begin{aligned} s_{12} &= 0.2243 \pm 0.0016 \\ s_{23} &= 0.0413 \pm 0.0015 \\ s_{13} &= 0.037 \pm 0.0005 \\ \delta_{13} &= 60^\circ \pm 14^\circ \end{aligned} \qquad [16]$$

For the muon anomalous gyromagnetic ratio $(g-2)$, the experimental data

$$a_\mu^{\text{exp}} = \frac{g_\mu - 2}{2} = (1.116\,5920(37) \pm 0.78) \times 10^{-9}$$

is to be compared with the theoretical prediction

$$a_\mu^{\text{th}} = (1.1165918(83) \pm 0.49) \times 10^{-9}$$

which is slightly smaller $(1.9\sigma)$, where the largest theoretical uncertainty comes from the two-loop hadronic contribution $a_\mu^{\text{had}} \simeq (69.63 \pm 0.72) \times 10^{-9}$ (the QED corrections to $O(\alpha^5)$ are included).

For further details of the analyses and the present status of experimental tests of the electroweak theory, see the reviews by J Erler and P Langacker, and by F J Gilman *et al.*, cited in "Further reading" (most of numbers cited here come from these two reviews).

## Need for Extension of the Model

In spite of such an impressive experimental confirmation, there are reasons to believe that the electroweak theory, in its standard minimal form, is not a complete story. As already mentioned, neutrino oscillations, predicted earlier by Pontecorvo, have recently been experimentally confirmed, giving uncontroversial evidence for nonvanishing neutrino masses and their mixing. This is a clear signal that the theory must be extended. If the mass is instead taken in the form of eqn [11] but with three neutrinos families, the diagonalization in general yields a mixing for the light neutrinos, as for the quarks. Some of the experimental data on the neutrinos are summarized in Table 7.

In addition, the Higgs sector of the theory (the part of the interactions responsible for spontaneous breaking $SU_L(2) \times U(1) \to U_{EM}(1)$) is still largely untested. The theory predicts a physical scalar particle, the Higgs particle, of unknown mass. The present-day expectation for its mass, which combines the experimental lower limit and an indirect upper limit following from the analysis of various radiative corrections, is

$$114\,(\text{GeV}) < m_H < 250\,(\text{GeV})$$

This particle should be observable either in the Tevatron at Fermilab or in the coming LHC

**Table 7** Neutrino mass square differences and mixing

| $\nu_e$ | | $\nu_\mu$ | | $\nu_\tau$ |
|---|---|---|---|---|
| $\Delta_{12} m^2$ | = | (6 – 9) | $\times$ | $10^{-5}\,\text{eV}^2$ |
| $\Delta_{23} m^2$ | = | (1 – 3) | $\times$ | $10^{-3}\,\text{eV}^2$ |

Solar neutrinos and reactor (SNO, SuperKamiokande, KamLAND) experiments give the first results. Atmospheric neutrino data and the long baseline experiment (SuperKamiokande, K2K) provide the second. The mixing angle relevant to the solar and reactor neutrino oscillation is large, $\tan^2\theta_{12} \sim 0.40^{+0.10}_{-0.07}$, while the one related to the atmospheric neutrino data is maximal, $\sin^2 2\theta_{23} \sim 1$. Cosmological considerations give $\sum m_{\nu_i} < O(1\,\text{eV})$.

experiments at CERN; negative results would force upon us a substantial modification of the electroweak theory.

Last, but not least, there are a few theoretical motivations for an extension of the model to be considered necessary. First, the structure of the GWS theory is not entirely determined by the gauge principle. The form of the Higgs self-interactions, as well as their number and the Yukawa couplings of the Higgs scalar to the fermions, are unconstrained by any principle, and the particular, minimal form assumed by Weinberg and Salam is yet to be confirmed experimentally.

In addition, the theory is not really a unified gauge theory: $SU_L(2)$ and U(1) gauge couplings are distinct. One possibility is that the $SU(3)_{QCD} \times SU_L(2) \times U(1)$ theory of the SM is actually a low-energy manifestation of a truly unified gauge theory – grand unified theory (GUT) – defined at some higher mass scale. The simplest version of GUT models based on SU(5) or SO(10) gauge groups has however a difficulty with the proton decay rates, and with the coupling-constant unification itself. Supersymmetric GUTs appear to be more acceptable both from the coupling-constant unification and from the proton lifetime constraints.

A more subtle, but perhaps more severe theoretical problem, is the so-called naturalness problem. At the quantum level, due to the quadratic divergences in the scalar mass, the structure of the theory turns out to be quite peculiar. If the ultraviolet cutoff of the theory is taken to be the Planck mass scale, $\Lambda_{UV} \sim m_{Pl} \sim 10^{19}$ GeV, at which gravity becomes strongly coupled, the theory at $\Lambda_{UV}$ would have to possess parameters which are fine-tuned with an excessive precision. The problem is known also as a "hierarchy" problem.

A way to avoid having such a difficulty is to introduce supersymmetry. In a supersymmetric version of the standard theory – in fact, there are phenomenologically well-acceptable models such as the minimal supersymmetric standard model (MSSM) – this problems is absent due to the cancellation of bosonic and fermionic loop contributions typical of supersymmetric theories. As a result, the properties of the theory at low energies are much less sensitive to those of the theory at the Planck mass scale. Experiments at LHC (expected to be performed after 2008, CERN) should be able to produce a whole set of new particles associated with supersymmetry, if this is a part of the physical law beyond TeV energies.

At a deeper level, however, the hierarchy problem in a more general sense persists, even in supersymmetric models: why the masses of the order of $O(100\ \text{GeV})$ should appear at all in a theory with a natural cutoff of the order of the Planck mass? Furthermore, if the masses of the neutrinos turn out to be of the order of $O(10^{-3}\text{–}10^{0})$ eV, we are left with the problem of understanding the large disparities among the quark and lepton masses, spanning the range of more than 13 orders of magnitudes: another "hierarchy" problem.

It is also possible that the spacetime the physical world lives in is actually higher dimensional: the usual four-dimensional Minkowski spacetime times either compactified or uncompactified "extra dimensions." In theories of this type, some of the difficulties mentioned above might find a natural solution. It is yet to be seen whether a consistent theory of this type can be constructed that correctly account for the properties of the universe we inhabit.

## Bibliographic Notes

A short but comprehensive introduction to the Weinberg–Salam theory is found in Taylor (1976); see also Abers and Lee (1973) and 't Hooft and Veltman (1973).

The reprint collection edited by Taylor (2001) contains many of fundamental papers, e.g., on Yang–Mills theories (by C N Yang, R L Mills, R Shaw), on spontaneous symmetry breaking and its application to gauge theories (by Y Nambu, J Schwinger, P W Anderson, P W Higgs, F Englert, R Brout, T W B Kibble) and on renormalization of Yang–Mills theories and application to the electroweak theory (L D Faddeev, V N Popov, G 't Hooft).

For up-to-date review on precision tests of the GWS theory, and details of the analyses, see Erler and Langacker (2004), and references cited therein.

For a recent review on neutrino experiments, see Shirai (2005). For theory on neutrinos, see Fukugita and Yanagida (2003).

For the unitarity triangle test of the GWS model and determination of CKM matrix elements, see results from the CKM fitter Group (Bret *et al.* 2005), the UTfit Collaboration (Bona *et al.* 2005), and a review by Gilman *et al.* (2005), and references cited therein.

*See also:* Abelian and Nonabelian Gauge Theories using Differential Forms; Current Algebra; Effective Field Theories; Finite Group Symmetry Breaking; Noncommutative Tori, Yang–Mills, and String Theory; Quantum Chromodynamics; Quantum Electrodynamics and Its Precision Tests; Quantum Field Theory: A Brief Introduction; Renormalization: General Theory; Standard Model of Particle Physics; Symmetries and Conservation Laws; Symmetry and Symmetry Breaking in Dynamical Systems; Symmetry Breaking in Field Theory.

## Further Reading

Abers ES and Lee BW (1973) Gauge theories. *Physics reports* C9: 1–141.
Bona M, Ciuchina M, and Franco E *et al.* (2005) The 2004 UTfit collaboration report on the status of the unitarity triangle in the standard model. *Journal of High Energy Physics* 0507: 028–059 (hep-ph/0501199).
Charles J, Hocker A, Lacker H *et al.* (2005) CP violation and the CKM matrix: assesing the impact of the asymmetric B factories. *European Physical Journal* C 41: 1–131 (hep-ph/0406184).
Erler J and Langacker P (2004) Electroweak Model and constraints on new physics. In: Eidelman S, Hayes KG, Olive KA *et al.* (ed.) *Review of Particle Physics*, Particle Data Group, Physics Letters B vol. 592, pp. 1–1109, (http://pdglbl.gov/2005).
Fukugita M and Yanagida T (2003) *Physics of Neutrinos and Applications to Astrophysics*. Berlin: Springer.
Gilman FJ, Kleinknecht K, and Renk B The Cabibbo–Kobayashi–Maskawa quark-mixing matrix. In Eidelman S, Hayes KG, Olive KA *et al.* (eds.) PDG *Review of Particle Physics*, Particle Data Group, Physics Letters B, vol. 592, pp. 1–1109 (http://pdg.lbl.gov/2005).
Glashow SL, Iliopoulos J, and Maiani L (1970) Weak interactions with Lepton–Hadron symmetry. *Physical Review D* 2: 1285–1292.
Kobayashi M and Maskawa T (1972) *CP* violation in renormalizable theory of weak interaction. *Progress in Theoretical Physics* 49: 652–657.
Salam A (1968) Weak and electromagnetic interactions. In: Svartholm N (ed.) *Elementary Particle Theory*, p. 367. , Lerum, Sweden: Almqvist Forlag AB.
Shirai J (2005) Neutrino experiments: review of recent results. *Nuclear Physics B (Proc. Suppl.)* 144: 286–296.
Taylor JC (1976) *Gauge Theories of Weak Interactions*. Cambridge: Cambridge University Press.
Taylor JC (ed.) (2001) *Gauge Theories in the Twentieth Century*. London: Imperial College Press.
't Hooft G (1971) Renormalizable Lagrangians for massive Yang–Mills fields. *Nuclear Physics* B35: 167–188.
't Hooft G and Veltman M (1973) Diagrammar. *CERN Yellow Report* 73–9.
Weinberg S (1967) Theory of Leptons. *Physical Review Letters* 19: 1264–1266.

# Elliptic Differential Equations: Linear Theory

**C Amrouche**, Université de Pau et des Pays de l'Adour, Pau, France
**M Krbec and Š Nečasová**, Academy of Sciences, Prague, Czech Republic
**B Lucquin-Desreux**, Université P.-M. Curie, Paris VI, Paris, France



## Introduction

### Motivation: A Model Problem

Many physical problems can be modeled by partial differential equations. Let us consider, for example, the case of an elastic membrane $\Omega$, with fixed boundary $\Gamma$, subject to pressure forces $f$. The vertical membrane displacement is represented by a real-valued function $u$, which solves the equation

$$-\Delta u(x) = f(x), \quad x = (x_1, x_2) \in \Omega \qquad [1]$$

where the Laplace operator $\Delta$ is defined, in two dimensions, by

$$\Delta u = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2}$$

As the membrane is glued to the curve $\Gamma$, $u$ satisfies the condition

$$u(x) = 0, \quad x \in \Gamma \qquad [2]$$

The system [1]–[2] is the homogeneous Dirichlet problem for the Laplace operator. It enters the more general framework of (linear) elliptic boundary value problems, which consist of a (linear) partial differential equation (in the example above, of order two: the highest order in the derivatives) inside an open set $\Omega$ of the whole space $\mathbb{R}^N$, satisfying some "elliptic" property, completed by (linear) conditions on the boundary $\Gamma$ of $\Omega$, called "boundary conditions." In the sequel, we only consider the linear case.

Our aim is to answer the following questions: does this problem admit a solution? in which space? is this solution unique? does it depend continuously on the given data $f$? In case of positive answers, we say that the problem is "well posed" in the Hadamard sense. But other questions can also be raised, such as the sign of the solution, for example, or its regularity. We give a full survey of linear elliptic problems in a bounded or in an exterior domain with a sufficiently smooth boundary and in the whole space. In the general theory of the elliptic problem, we consider only smooth coefficients. We survey the standard theory, which can be found in the several well-known monographs of the 1960s. The new trends in the investigation of the elliptic problems is to consider more general domains with nonsmooth boundaries and nonsmooth coefficients. On the other hand, the regularity results for elliptic systems have not been improved during last 30 years. New trends also require employment of more general function spaces and more general functional background.

The number of references (see "Further reading" section) is strictly limited here; we list only some of the most important publications. The basic facts can usually be found in more places and sometimes we do not mention the particular reference. Among the

very basic references are Friedman (1969), Gilbarg and Trudinger (1977), Dautray and Lions (1988), Hörmander (1964), Ladyzhenskaya and Uraltseva (1968), Lions and Magenes (1968), Renardy and Rogers (1992), and Weinberger (1965); of course, there are many others.

### The Method

To answer the above questions, we generally use, for such elliptic problems, an approach based on what is called a "variational formulation" (see the section "Variational approach"): the boundary-value problem is first transformed into a variational problem of lower order, which is solved in a Hilbertian frame with help of the Lax–Milgram theorem (based on the representation theorem). All questions are then solved (e.g., existence, uniqueness, continuity in terms of the data, regularity). But this variational formalism does not necessary allow to treat all the situations and it is limited to the Hilbertian case. Other strategies can then be developed, based on *a priori* estimates and duality arguments for the existence problem, or maximum principle for the question of unicity. Without forgetting the particular cases where an explicit Green kernel is computable (e.g., the Laplacian operator in the whole space case).

Moreover, the study of linear elliptic equations is directly linked to the background of function spaces. It is the reason why we first deal with Sobolev spaces – both of the integer and fractional order and we survey their basic properties, imbedding and trace theorems. We pay attention to the Riesz and Bessel potentials and we define weighted Sobolev spaces important in the context of unbounded opens. Second, we present the variational approach and the Lax–Milgram theorem as a key point to solve a large class of boundary-value problems. We give examples: the Dirichlet and Neumann problems for the Poisson equation, the Newton problem for more general second-order operators; we also investigate mixed boundary conditions and present an example of a problem of fourth order. Then, we briefly present the arguments for studying general elliptic problems and concentrate on second-order elliptic problems; we recall the weak and strong maximum principle, formulate the Fredholm alternative and tackle the regularity questions. Moreover, we are interested in the existence and uniqueness of solution of the Laplace equation in the whole space and in exterior opens. Finally, we present some particular examples arising from physical problems, either in fluid mechanics (the Stokes system) or in elasticity.

## Sobolev and Other Types of Spaces

Throughout, $\Omega \subset \mathbb{R}^N$ will generally be an open subset of the $N$-dimensional Euclidean space $\mathbb{R}^N$. A domain will be an open and connected subset of $\mathbb{R}^N$. We shall use standard notations for the spaces $L^p(\Omega)$, $C^\infty(\Omega)$, etc., and their norms. Let us agree that $C^{k,r}(\Omega)$, $k \in \mathbb{N}$, $r \in (0,1)$, denote the space of functions $f$ in $C^k(\Omega)$, whose derivatives $D^\alpha f$, $\alpha = (\alpha_1, \ldots, \alpha_N) \in \mathbb{N}^N$, of order $|\alpha| = \sum_{i=1}^N \alpha_i = k$ are all $r$-Hölder continuous. In the notations for some of these spaces, by $\bar{\Omega}$ we mean that the functions have the corresponding property on $\Omega$ and that they can be continuously extended to $\bar{\Omega}$.

Let us recall several fundamental concepts. The space $\mathcal{D}(\Omega)$ of the test functions in $\Omega$ consists of all infinitely differentiable $\varphi$ with a compact support in $\Omega$. A locally convex topology can be introduced here. The elements of the dual space $\mathcal{D}'(\Omega)$ are called the distributions. If $f \in L^1_{\text{loc}}(\Omega)$ (i.e., $f \in L^1(K)$ for all compact subsets $K$ of $\Omega$), then $f$ is a regular distribution; the duality is represented by $\int_\Omega f(x)\varphi(x)\,dx$. If $f \in \mathcal{D}'(\Omega)$, we define the distributional or the weak derivative $D^\alpha$ of $f$ as the distribution $\varphi \mapsto (-1)^{|\alpha|}\langle f, D^\alpha \varphi\rangle$. Plainly, if $f \in L^1_{\text{loc}}$ has "classical" partial derivatives in $L^1_{\text{loc}}$, then it coincides with the corresponding weak derivative.

If $\Omega = \mathbb{R}^N$, it is sometimes more suitable to work with the tempered distributions. The role of $\mathcal{D}(\Omega)$ is played by the space $\mathcal{S}(\mathbb{R}^N)$ of $C^\infty$-functions with finite pseudonorms $\sup |D^\alpha f(x)|(1+|x|)^k, |\alpha|, k = 0, 1, 2, \ldots$. Recall that the Fourier transform $\mathcal{F}$ maps $\mathcal{S}(\mathbb{R}^N)$ into itself and the same is true for the space of the tempered distributions $\mathcal{S}'(\mathbb{R}^N)$.

### Sobolev Spaces of Positive Order

The Sobolev space $W^{k,p}(\Omega)$, $1 \le p \le \infty$, $k \in \mathbb{N}$, is the space of all $f \in L^p(\Omega)$ whose weak derivatives up to order $k$ are regular distributions belonging to $L^p(\Omega)$; in $W^{k,p}(\Omega)$ we introduce the norm

$$\|f\|_{W^{k,p}(\Omega)} = \left( \sum_{|\alpha| \le k} \int_\Omega |D^\alpha f(x)|^p \,dx \right)^{1/p} \qquad [3]$$

when $p < \infty$ and $\max_{|\alpha| \le k} \sup \operatorname{ess}_{x \in \Omega} |D^\alpha f(x)|$ if $p = \infty$. The space $W^{k,p}(\Omega)$ is a Banach space, separable for $p < \infty$ and reflexive for $1 < p < \infty$; it is a Hilbert space for $p = 2$, more simply denoted $H^m(\Omega)$. In the following, we shall consider only the range $p \in (1, \infty)$.

The link with the classical derivatives is given by this well-known fact: a function $f$ belongs to $W^{1,p}(\Omega)$ if and only if it is a.e. equal to a function $\tilde{u}$, absolutely continuous on almost all line segments in $\Omega$ parallel to the coordinate axes, whose (classical) derivatives belong to $L^p(\Omega)$ (the Beppo–Levi theorem).

For $1 < p < \infty$ and noninteger $s > 0$ the Sobolev space $W^{s,p}(\Omega)$ of order $s$ is defined as the space of all $f$ with the finite norm

$$\|f\|_{W^{s,p}(\Omega)} = \left( \|f\|^p_{W_p^{[s]}(\Omega)} + \sum_{|\alpha|=[s]} \int_\Omega \int_\Omega \frac{|D^\alpha f(x) - D^\alpha f(y)|^p}{|x-y|^{N+p(s-[s])}} \right)^{1/p}$$

where $[s]$ is the integer part of $s$ (for details, see, e.g., Adams and Fournier (2003) and Ziemer (1989)).

### Imbedding Theorems

One of the most useful and important features of the functions in Sobolev spaces is an improvement of their integrability properties and the compactness of various imbeddings. Theorems of this type were first proved by Sobolev and Kondrashev. Let us agree that the symbols $\hookrightarrow$ and $\hookrightarrow\hookrightarrow$ stand for an imbedding and for a compact imbedding, respectively.

**Theorem 1** *Let $\Omega$ be a Lipschitz open. Then*

(i) *If $sp < N$, then $W^{s,p}(\Omega) \hookrightarrow L^{p^*}(\Omega)$ with $p^* = Np/(N - ps)$ (the Sobolev exponent). If $|\Omega| < \infty$, then the target space is any $L^r(\Omega)$ with $0 < r \le p^*$.*
*If $\Omega$ is bounded, then $W^{s,p}(\Omega) \hookrightarrow\hookrightarrow L^q(\Omega)$ for all $1 \le q < p^*$.*

(ii) *If $sp > N$, then $W^{j+s,p}(\Omega) \hookrightarrow C^j(\Omega)$ for $j = 0,1,\ldots$. If $\Omega$ has the Lipschitz boundary, then $W^{j+s,p}(\Omega) \hookrightarrow C^{j,\mu}(\bar\Omega)$ for $j = 0,1,\ldots$ and $\mu = s - N/p$.*
*If $sp > N$, then $W^{j+s,p}(\Omega) \hookrightarrow\hookrightarrow C^j(\Omega)$, $j = 0,1,\ldots$ and $W^{j+s,p}(\Omega) \hookrightarrow\hookrightarrow W^j_q(\Omega)$ for all $1 \le q \le \infty$. If, moreover, $\Omega$ has the Lipschitz boundary, then the target space can be replaced by $C^{j,\mu}(\bar\Omega)$ provided $sp > N > (s-1)p$ and $0 < \mu < s - N/p$.*

Note that if the imbedding $W^{s,p}(\Omega) \hookrightarrow L^q(\Omega)$ is compact for some $q \ge p$, then $|\Omega| < \infty$. Moreover, if $\limsup_{r\to\infty} |\{x \in \Omega; r \le |x| < r+1\}| > 0$, then $W^{s,p}(\Omega) \hookrightarrow L^q(\Omega)$ cannot be compact.

### Traces and Sobolev Spaces of Negative Order

Let $s > 0$ and let $\Omega$ be, for simplicity, a bounded open subset of $\mathbb{R}^N$ with boundary $\Gamma$ of class $C^{[s],1}$. Then with the help of local coordinates, we can define Sobolev spaces $W^{s,p}(\Gamma)$ (also denoted $H^s(\Gamma)$ for $p = 2$) on $\Gamma = \partial\Omega$ (see, e.g., Nečas (1967) and Adams and Fournier (2003) for details). If $f \in C(\bar\Omega)$, then $f_{|\Gamma}$ has sense. Introducing the space $\mathcal{D}(\bar\Omega)$ of restrictions in $\Omega$ of functions in $\mathcal{D}(\mathbb{R}^N)$, one can show that if $f \in \mathcal{D}(\bar\Omega)$, we have $\|f_{|\Gamma}\|_{W^{1-1/p,p}(\Gamma)} \le C\|f\|_{W^{1,p}(\Omega)}$ so that, in view of the density of $\mathcal{D}(\bar\Omega)$ in $W^{1,p}(\Omega)$, the restriction of $f$ to $\Gamma$ can be uniquely extended to the whole $W^{1,p}(\Omega)$. The result is the bounded trace operator $\gamma_0 : W^{1,p}(\Omega) \to W^{1-1/p,p}(\Gamma)$. Moreover, every $g \in W^{1-1/p,p}(\Gamma)$ can be extended to a (nonunique) function $f \in W^{1,p}(\Omega)$ and this extension operator is bounded with respect to the corresponding norms.

More generally, let us suppose $\Gamma$ is of class $C^{k-1,1}$ and define the operator $\mathrm{Tr}_n$ for any $f \in \mathcal{D}(\bar\Omega)$ by $\mathrm{Tr}_n f = (\gamma_0 f, \gamma_1 f, \ldots, \gamma_{k-1} f)$, where

$$\gamma_j f(x) = \frac{\partial^j f}{\partial n^j}(x) = \sum_{|\alpha|=j} \frac{j!}{\alpha!}(\partial^\alpha f(x)/\partial x^\alpha) n^\alpha, \quad x \in \Gamma$$

is the $j$th-order derivative of $f$ with respect to the outer normal $n$ at $x \in \Gamma$; by density, this operator can be uniquely extended to a continuous linear mapping defined on the space $W^{k,p}(\Omega)$; moreover, $\gamma_0(W^{k,p}(\Omega)) = W^{k-1/p,p}(\Gamma)$.

The kernel of this mapping is the space $\mathring{W}^{k,p}(\Omega)$ (denoted by $H_0^k(\Omega)$ for $p = 2$), where $\mathring{W}^{s,p}(\Omega)$ is defined as the closure of $\mathcal{D}(\Omega)$ in $W^{s,p}(\Omega)$ ($s > 0$). For $1 < p < \infty$, the following holds: $\mathring{W}^{s,p}(\mathbb{R}^N) = W^{s,p}(\mathbb{R}^N)$, $\mathring{W}^{s,p}(\Omega) = W^{s,p}(\Omega)$ provided $0 < s \le 1$ $p$. If $s < 0$, then the space $W^{s,p}(\Omega)$ is defined as the dual to $\mathring{W}^{-s,p'}(\Omega)$, where $p' = p/(p-1)$ (see, e.g., Triebel (1978, 2001)). Observe that, for an arbitrary $\Omega$, a function $f \in W^{1,p}(\Omega)$ has the zero trace if and only if $f(x)/\mathrm{dist}(x,\Gamma)$ belongs to $L^p(\Omega)$.

For $p = 2$, we simply denote by $H^{-k}(\Omega)$ the dual space of $H_0^k(\Omega)$. In the case of bounded opens, we recall the following useful Poincaré–Friedrichs inequality (for simplicity, we state it here in the Hilbert frame):

**Theorem 2** *Let $\Omega$ be bounded (at least in one direction of the space). Then there exists a positive constant $C_P(\Omega)$ such that*

$$\|v\|_{L_2(\Omega)} \le C_P(\Omega)\|\nabla v\|_{[L_2(\Omega)]^N} \quad \text{for all } v \in H_0^1(\Omega) \qquad [4]$$

### The Whole-Space Case: Riesz and Bessel Potentials

The Riesz potentials $\mathcal{I}_\alpha$ naturally occur when one defines the formal powers of the Laplace operator $\Delta$. Namely, if $f \in \mathcal{S}(\mathbb{R}^N)$ and $\alpha > 0$, then

$$\mathcal{F}\left[(-\Delta)^{\alpha/2} f\right](\xi) = |\xi|^\alpha \mathcal{F}f(\xi).$$

This can be taken formally as a definition of the Riesz potential $\mathcal{I}_\alpha$ on $\mathcal{S}'(\mathbb{R}^N)$,

$$\mathcal{I}_\alpha f(.) = \mathcal{F}^{-1}[|\xi|^{-\alpha}\mathcal{F}f(\xi)](.)$$

for any $\alpha \in \mathbb{R}$. If $0 < \alpha < N$, then $I_\alpha f(x) = (I_\alpha * f)(x)$, where $I_\alpha$ is the inverse Fourier transform of $|\xi|^{-\alpha}$,

$$I_\alpha(x) = C_\alpha |x|^{\alpha-N}$$
$$C_\alpha = \Gamma((N-\alpha)/2)\left(\pi^{N/2}2^\alpha\Gamma(\alpha/2)\right)^{-1}$$

where $\Gamma$ is the Gamma function and $I_\alpha$ is the Riesz kernel. The following formula is also true:

$$I_\alpha(x) = C_\alpha \int_0^\infty t^{(\alpha-N)/2} e^{-\pi|x|^2/t} \frac{dt}{t}$$

Recall that every $f \in \mathcal{S}(\mathbb{R}^N)$ can be represented as the Riesz potential $\mathcal{I}_\alpha g$ of a suitable function $g \in \mathcal{S}(\mathbb{R}^N)$, namely $g = (-\Delta)^{\alpha/2} f$; we get the representation formula

$$f(x) = \mathcal{I}_\alpha g(x) = C_\alpha \int_{\mathbb{R}^N} \frac{g(y)}{|x-y|^{N-\alpha}} dy$$

The standard density argument implies then an appropriate statement for functions in $W^{k,p}(\mathbb{R}^N)$ with an integer $k$ and for the Bessel potential spaces $\mathcal{H}^{\alpha,p}(\mathbb{R}^N)$ – see below for their definition. The original Sobolev imbedding theorem comes from the combination of this representation and the basic continuity property of $I_\alpha, \alpha p < N$,

$$I_\alpha: L^p(\mathbb{R}^N) \to L^q(\mathbb{R}^N), \quad \frac{1}{q} = \frac{1}{p} - \frac{\alpha}{N}$$

To get an isomorphic representation of a Bessel potential space (of a Sobolev space with positive integer smoothness in particular) it is more convenient to consider the Bessel potentials (of order $\alpha \in \mathbb{R}$),

$$\mathcal{G}_\alpha f(x) = (G_\alpha * f)(x) = \mathcal{F}^{-1}\left([1+|\xi|^2]^{-\alpha/2}\mathcal{F}f(\xi)\right)(x)$$

(with a slight abuse of the notations); the following formula for the Bessel kernel $G_\alpha$ is well known:

$$G_\alpha(x) = c_\alpha^{-1} \int_0^\infty t^{(\alpha-N)/2} e^{-(\pi|x|^2/t)-(t/4\pi)} \frac{dt}{t}$$

(cf. the analogous formula for $I_\alpha$), where $c_\alpha = (4\pi)^{\alpha/2}\Gamma(\alpha/2)$. The kernels $G_\alpha$ can alternatively be expressed with help of Bessel or Macdonald functions.

Now we can define the Bessel potential spaces. For $s \in \mathbb{R}$ and $1 < p < \infty$, let $\mathcal{H}^{s,p}(\mathbb{R}^N)$ be the space of all $f \in \mathcal{S}'(\mathbb{R}^N)$ with the finite norm

$$\|f\|_{\mathcal{H}^{s,p}(\mathbb{R}^N)} = \left(\int_{\mathbb{R}^N} \mathcal{F}^{-1}\left((1+|\xi|^2)^{s/2}\mathcal{F}f(\xi)\right)^p d\xi\right)^{1/p}$$

In other words, the spaces $\mathcal{H}^{s,p}(\mathbb{R}^N)$ are isomorphic copies of $L^p(\mathbb{R}^N)$.

For $k = 0, 1, 2, \ldots$, plainly $\mathcal{H}^{k,2}(\mathbb{R}^N) = W^{k,2}(\mathbb{R}^N)$ by virtue of the Plancherel theorem. But it is true also for integer $s$ and general $1 < p < \infty$ (see, e.g., Triebel (1978)).

**Remark 3** Much more comprehensive theory of general Besov and Lizorkin–Triebel spaces in $\mathbb{R}^N$ has been established in the last decades, relying on the the Littlewood–Paley theory. Spaces on opens can be defined as restrictions of functions in the corresponding space on the whole $\mathbb{R}^N$, allowing to derive their properties from those valid for functions on $\mathbb{R}^N$. The justification for that are extension theorems. In particular, there exists a universal extension operator for the Lipschitz open, working for all the spaces mentioned up to now. We refer to Triebel (1978, 2001).

### Unbounded Opens and Weighted Spaces

The study of the elliptic problems in unbounded opens is usually carried out with use of suitable Sobolev weighted space. The Poisson equation

$$-\Delta u = f \quad \text{in } \mathbb{R}^N, \ N \geq 2 \qquad [5]$$

is the typical example; the Poincaré inequality [4] is not true here and it is suitable to introduce Sobolev spaces with weights.

Let $m \in \mathbb{N}$, $1 < p < \infty$, $\alpha \in \mathbb{R}$, $k = m - N/p - \alpha$ if $N/p + \alpha \in \{1, \ldots, m\}$ and $k = -1$ elsewhere. For an open $\Omega \subset \mathbb{R}^N$, we define

$$W_\alpha^{m,p}(\Omega) = \Big\{v \in \mathcal{D}'(\Omega), 0 \leq |\lambda| \leq k, \ \rho^{\alpha-m-|\lambda|}(\log\rho)^{-1}D^\lambda u \in L^p(\Omega), \ k+1 \leq |\lambda| \leq m, \ \rho^{\alpha-m+|\lambda|}D^\lambda u \in L^p(\Omega)\Big\}$$

where $\rho(x) = (1+|x|^2)^{1/2}$. Note that $W_\alpha^{m,p}$ is a reflexive Banach space for the norm $\|.\|_{W_\alpha^{m,p}}$ defined by

$$\|u\|^p_{W_\alpha^{m,p}} = \sum_{0\leq|\lambda|\leq k} \|\rho^{\alpha-m+|\alpha|}(\log\rho)^{-1}D^\lambda u\|^p_{L^p(\Omega)} + \sum_{k+1\leq|\lambda|\leq m} \|\rho^{\alpha-m+|\alpha|}D^\lambda u\|^p_{L^p(\Omega)}$$

We also introduce the following seminorm:

$$|u|_{W_\alpha^{m,p}} = \left(\sum_{|\lambda|=m} \|\rho^\alpha D^\lambda u\|^p_{L^p(\Omega)}\right)^{1/p}$$

Let

$$\mathring{W}_\alpha^{m,p}(\Omega) = \{v \in W_\alpha^{m,p}; \gamma_0(v) = \cdots \gamma_{m-1}(v) = 0\}$$

If $\Omega$ is a Lipschitz domain, then $\overset{\circ}{W}{}^{m,p}_{\alpha}(\Omega)$ is the closure of $\mathcal{D}(\Omega)$ in $W^{m,p}_{\alpha}(\Omega)$, while $\mathcal{D}(\bar{\Omega})$ is dense in $W^{m,p}_{\alpha}(\Omega)$. We denote by $W^{-m,p'}_{-\alpha}(\Omega)$ the dual of $\overset{\circ}{W}{}^{m,p}_{\alpha}(\Omega)$ $(p' = p/(p-1))$. We note that these spaces also contain polynomials,

$$P_j \subset W^{m,p}_{\alpha}(\Omega)$$
$$\Leftrightarrow \begin{cases} j = \left[m - \dfrac{N}{p} - \alpha\right] & \text{if } \dfrac{N}{p} + \alpha \notin Z \\ j = m - \dfrac{N}{p} - \alpha & \text{elsewhere} \end{cases}$$

where $[s]$ is the integer part of $s$ and $P_{[s]} = \{0\}$ if $[s] < 0$. The fundamental property of functions belonging to these spaces is that they satisfy the Poincaré weighted inequality. An open $\Omega$ is an exterior domain if it is the complement of a closure of a bounded domain in $R^N$.

**Theorem 4** *Suppose that $\Omega$ is an exterior domain or $\Omega = \mathbb{R}^N_+$ or $\Omega = \mathbb{R}^N$. Then*

(i) *the seminorm $|\cdot|_{W^{m,p}_{\alpha}(\Omega)}$ is a norm on $W^{m,p}_{\alpha}(\Omega)/P_j$, equivalent to the quotient norm with $j' = \min(m-1, j)$;*
(ii) *the seminorm $|\cdot|_{W^{m,p}_{\alpha}(\Omega)}$ is equivalent to the full norm on $\overset{\circ}{W}{}^{m,p}_{\alpha}(\Omega)$.*

## Variational Approach

Let us first describe the method on the model problem [1]–[2], supposing $f \in L^2(\Omega)$ and $\Omega$ bounded. We first suppose that this problem admits a sufficiently smooth function $u$. Let $v$ be any arbitrary (smooth) function; we multiply eqn [1] by $v(x)$ and integrate with respect to $x$ over $\Omega$; this gives

$$\int_{\Omega} -(\Delta u v)(x)\mathrm{d}x = \int_{\Omega} (f v)(x)\mathrm{d}x$$

Using the following Green's formula ($\mathrm{d}\sigma(x)$ denotes the measure on $\Gamma = \partial\Omega$ and $\partial u(x)/\partial n = \nabla u(x) \cdot n(x)$, where $n(x)$ is the unit normal at point $x$ of $\Gamma$ oriented towards the exterior of $\Omega$):

$$\int_{\Omega} (\Delta u v)(x)\mathrm{d}x = -\int_{\Omega} (\nabla u \cdot \nabla v)(x)\mathrm{d}x + \int_{\Gamma} \left(\frac{\partial u}{\partial n} v\right)(\sigma)\mathrm{d}\sigma \qquad [6]$$

we get, since $v|_{\Gamma} = 0$: $\mathcal{A}(u,v) = L(v)$, where we have set

$$\mathcal{A}(u,v) = \int_{\Omega} \nabla u(x) \cdot \nabla v(x)\mathrm{d}x$$
$$L(v) = \int_{\Omega} f(x) v(x)\mathrm{d}x \qquad [7]$$

The idea is to study in fact this new problem (showing first its equivalence with the boundary-value problem), noting that it makes sense for far less regular functions $u$, $v$ (and also $f$), in fact $u, v \in H^1_0(\Omega)$ (and $f \in H^{-1}(\Omega)$).

### The Lax–Milgram Theorem

The general form of a variational problem is

$$\text{to find } u \in V \text{ such that } \mathcal{A}(u,v) = L(v) \text{ for all } v \in V \qquad [8]$$

where $V$ is a Hilbert space, $\mathcal{A}$ a bilinear continuous form defined on $V \times V$ and $L$ a linear continuous form defined on $V$. We say, moreover, that $\mathcal{A}$ is V-elliptic if there exists a positive constant $\alpha$ such that

$$\mathcal{A}(u,u) \geq \alpha \|u\|^2_V \quad \text{for all } u \in V \qquad [9]$$

The following theorem is due to Lax and Milgram.

**Theorem 5** *Let V be a Hilbert space. We suppose that $\mathcal{A}$ is a bilinear continuous form on $V \times V$ which is V-elliptic and that L is a linear continuous form on V. Then the variational problem [8] has a unique solution u on V. Moreover, if $\mathcal{A}$ is symmetric, u is characterized as the minimum value on V of the quadratic functional E defined by*

$$\text{for all } v \in V, \; E(v) = \tfrac{1}{2}\mathcal{A}(v,v) - L(v) \qquad [10]$$

**Remark 6**

(i) We have the following "energy estimate": $\|u\|_V \leq \frac{1}{\alpha}\|L\|_{V'}$ where $V'$ is the dual space to $V$. In the particular case of our model problem, this inequality shows the continuity of the solution $u \in H^1_0(\Omega)$ with respect to the data $f \in L^2(\Omega)$ (that can be weakened by choosing $f \in H^{-1}(\Omega)$).

(ii) Theorem 5 can be extended to sesquilinear continuous forms $\mathcal{A}$ defined on $V \times V$; such a form is called $V$-elliptic if there exists a positive constant $\alpha$ such that

$$\operatorname{Re}\mathcal{A}(u,u) \geq \alpha\|u\|^2_V \quad \text{for all } u \in V \qquad [11]$$

(iii) Denoting by $A$ the linear operator defined on the space $V$ by $\mathcal{A}(u,v) = \langle Au, v\rangle_{V',V}$, for all $v \in V$, the Lax–Milgram theorem shows that $A$ is an isomorphism from $V$ onto its dual space $V'$, and the problem [8] is equivalent to solving the equation $Au = L$.

(iv) Let us make some remarks concerning the numerical aspects. First, this variational formulation is the starting point of the well-known finite element method: the idea is to compute a solution of an approximate variational problem stated on a finite subspace of $V$ (leading to the resolution of a linear

system), with a precise control of the error with the exact solution $u$. Second, the equivalence with a minimization problem allows the use of other numerical algorithms.

Let us now present some classical examples of second-order elliptic problems than can be solved with help of the variational theory.

### The Dirichlet Problem for the Poisson Equation

We consider the problem on a bounded Lipschitz open $\Omega \subset \mathbb{R}^N$,

$$\begin{aligned} -\Delta u &= f \\ u &= u_0 \quad \text{on } \Gamma = \partial\Omega \end{aligned} \qquad [12]$$

with $u_0 \in H^{1/2}(\Gamma)$, so that there exists $U_0 \in H^1(\Omega)$ satisfying $\gamma_0(U_0) = u_0$. The variational formulation of problem [12] is

$$\begin{aligned} &\text{to find } u \in U_0 + H_0^1(\Omega) \text{ such that} \\ &\text{for all } v \in H_0^1(\Omega),\ \mathcal{A}(u,v) = L(v) \end{aligned} \qquad [13]$$

with $\mathcal{A}$ given by [7] and a more general $L$ with $f \in H^{-1}(\Omega)$, defined by

$$L(v) = \langle f, v\rangle_{H^{-1}(\Omega),H_0^1(\Omega)} \qquad [14]$$

The existence and uniqueness of a solution of [13] follows from Theorem 5 (and Poincaré inequality [4]). Conversely, thanks to the density of $\mathcal{D}(\Omega)$ in $H_0^1(\Omega)$, we can show that $u$ satisfies [12]. More precisely, we get:

**Theorem 7** *Let us suppose $f \in H^{-1}(\Omega)$ and $u_0 \in H^{1/2}(\Gamma)$; let $U_0 \in H^1(\Omega)$ satisfy $\gamma_0(U_0) = u_0$. Then the boundary-value problem [12] has a unique solution $u$ such that $u - U_0 \in H_0^1(\Omega)$. This is also the unique solution of the variational problem [13]. Moreover, there exists a positive constant $C = C(\Omega)$ such that*

$$\|u\|_{H^1(\Omega)} \le C\Big(\|f\|_{H^{-1}(\Omega)} + \|u_0\|_{H^{1/2}(\Gamma)}\Big) \qquad [15]$$

*which shows that $u$ depends continuously on the data $f$ and $u_0$.*

Moreover, using techniques of Nirenberg's differential quotients, we have the following regularity result (see, e.g., Grisvard (1980)):

**Theorem 8** *Let us suppose that $\Omega$ is a bounded open subset of $\mathbb{R}^N$ with a boundary of class $\mathcal{C}^{1,1}$ and let $f \in L^2(\Omega), u_0 \in H^{3/2}(\Gamma)$. Then $u \in H^2(\Omega)$ and each equation in [12] is satisfied almost everywhere (on $\Omega$ for the first one and on $\Gamma$ for the boundary condition). Moreover, there exists a positive constant $C = C(\Omega)$ such that*

$$\|u\|_{H^2(\Omega)} \le C[\|f\|_{L^2(\Omega)} + \|g\|_{H^{3/2}(\Gamma)}] \qquad [16]$$

By induction, if the data are more regular, that is, $f \in H^k(\Omega)$ and $u_0 \in H^{k+3/2}(\Gamma)$ (with $k \in \mathbb{N}$), and if $\Gamma$ is of class $\mathcal{C}^{k+1,1}$, we get $u \in H^{k+2}(\Omega)$.

**Remark 9** Let us point out the importance of the open geometry. For example, if $\Omega$ is a bounded plane polygon, one can find $u \in H_0^1(\Omega)$ with $\Delta u \in \mathcal{C}^\infty(\bar{\Omega})$, such that $u \notin H^{1+\pi/w}(\Omega)$, where $w$ is the biggest value of the interior angles of the polygon. In particular, if the polygon is not convex, the solution of the Dirichlet problem [12] cannot be in $H^2(\Omega)$.

### The Neumann Problem for the Poisson Equation

We consider the problem ($n$ is the unit outer normal on $\Gamma$)

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega \\ \frac{\partial u}{\partial n} &= h \quad \text{on } \Gamma \end{aligned} \qquad [17]$$

Setting $E(\Delta) = \{v \in H^1(\Omega);\ \Delta v \in L^2(\Omega)\}$, the space $\mathcal{D}(\bar{\Omega})$ is a dense subspace, and we have the following Green formula for all $u \in E(\Delta)$ and $v \in H^1(\Omega)$:

$$\begin{aligned} &\int_\Omega \Delta u(x)v(x)\mathrm{d}x \\ &= -\int_\Omega \nabla u(x)\cdot\nabla v(x)\mathrm{d}x + \left\langle \frac{\partial u}{\partial n}, \gamma_0 v\right\rangle_{H^{-1/2}(\Gamma),H^{1/2}(\Gamma)} \end{aligned}$$

If $u \in H^1(\Omega)$ satisfies [17] with $f \in L^2(\Omega)$ and $h \in H^{-1/2}(\Gamma)$, then for any function $v \in H^1(\Omega)$, we have, by virtue of the above Green formula,

$$\begin{aligned} \mathcal{A}(u,v) &= \tilde{L}(v) \\ \tilde{L}v &= \int_\Omega (fv)(x)\mathrm{d}x + \langle h, \gamma_0 v\rangle_{H^{-1/2}(\Gamma),H^{1/2}(\Gamma)} \end{aligned}$$

But, here the form $\mathcal{A}$ is not $H^1(\Omega)$-elliptic; in fact, one can check that, if problem [17] has a solution, then we have necessarily (take $v = 1$ above)

$$\int_\Omega f(x)\,\mathrm{d}x + \langle h, 1\rangle_{H^{-1/2}(\Gamma),H^{1/2}(\Gamma)} = 0 \qquad [18]$$

Moreover, we note that if $u$ is a solution, then $u + C$, where $C$ is an arbitrary constant, is also a solution. So the variational problem is not well posed on $H^1(\Omega)$. It can, however, be solved in the quotient space $H^1(\Omega)/\mathbb{R}$, which is a Hilbert space for the quotient norm

$$\|\dot{v}\|_{H^1(\Omega)/\mathbb{R}} = \inf_{k\in\mathbb{R}} \|v + k\|_{H^1(\Omega)} \qquad [19]$$

but also for the seminorm $v \mapsto |v|_{H^1(\Omega)} = \sqrt{\mathcal{A}(v,v)}$, which is an equivalent norm on this quotient space (see Nečas (1967)).

Then, supposing that the data $f$ and $h$ satisfy the "compatibility condition" [18], we can apply the Lax–Milgram theorem to the variational problem

$$\begin{aligned}&\text{to find } \dot{u} \in V \text{ such that}\\ &\mathcal{A}(\dot{u}, \dot{v}) = \tilde{L}(\dot{v}) \text{ for all } \dot{v} \in V\end{aligned} \qquad [20]$$

with $V = H^1(\Omega)/\mathbb{R}$. We get the following result (see, e.g., Nečas (1967):

**Theorem 10** *Let us suppose that $\Omega$ is connected and that the data $f \in L^2(\Omega)$ and $h \in H^{-1/2}(\Gamma)$ satisfy* [18]. *Then the variational problem* [20] *has a unique solution $\dot{u}$ in the space $H^1(\Omega)/\mathbb{R}$ and this solution is continuous with respect to the data, that is, there exists a positive constant $C = C(\Omega)$ such that*

$$|u|_{H^1(\Omega)} \le C\Big(\|f\|_{L^2(\Omega)} + \|h\|_{H^{-1/2}(\Gamma)}\Big)$$
*for all* $u \in \dot{u}$

Moreover, if $\Gamma$ is of class $\mathcal{C}^{1,1}$ and if the data satisfy $f \in L^2(\Omega), g \in H^{1/2}(\Gamma)$, then every $u \in \dot{u}$ is such that $u \in H^2(\Omega)$ and it satisfies each equation in [17] almost everywhere.

### Problem with Mixed Boundary Conditions

Here we consider more general boundary conditions: the Dirichlet conditions on a closed subset $\Gamma_1$ of $\Gamma = \partial\Omega$, and the Neumann, or more generally the "Robin", conditions on the other part $\Gamma_2 = \Gamma - \Gamma_1$. We seek $u$ such that ($f \in L^2(\Omega)$, $h \in L^2(\Gamma_2)$, $a \in L^\infty(\Gamma_2)$)

$$\begin{aligned}-\Delta u &= f \quad \text{in } \Omega\\ u &= 0 \quad \text{on } \Gamma_1\\ au + \frac{\partial u}{\partial n} &= h \quad \text{on } \Gamma_2\end{aligned} \qquad [21]$$

Let $V = \{v \in H^1(\Omega); \gamma_0 v = 0 \text{ on } \Gamma_1\}$. Then [8] is the variational formulation of this problem with

1. $\mathcal{A}(u,v) = \int_\Omega \nabla u(x) \cdot \nabla v(x)\,dx + \int_{\Gamma_2} (a\gamma_0 u \gamma_0 v)(\sigma)d\sigma$;
2. $L(v) = \int_\Omega f(x)v(x)dx + \int_{\Gamma_2} (h\gamma_0 v)(\sigma)d\sigma$.

Supposing, for example, $a \ge 0$, we get a unique solution $u \in V$ for this variational problem by virtue of the Lax–Milgram theorem. Moreover, if $u \in H^2(\Omega)$, then $u$ is the unique solution in $H^2(\Omega) \cap V$ of the problem [21].

### The Newton Problem for More General Operators

Let $\Omega$ be a bounded open subset of $\mathbb{R}^n$. We now consider more general second-order operators of the form $v \mapsto -\nabla.(M\nabla v) + b \cdot \nabla v + cv$, where $b \in [W^{1,\infty}(\Omega)]^N$, $c \in L^\infty(\Omega)$, $M$ is an $N \times N$ square matrix with entries $M_{ij}$, and $\nabla \cdot (M\nabla v)$ stands for

$$\sum_{i,j=1}^{N} \frac{\partial}{\partial x_i}\left[M_{ij}\frac{\partial u}{\partial x_j}\right]$$

We also assume that there is a positive constant $\alpha_M$ such that

$$\sum_{i,j=1}^{N} M_{ij}(x)\xi_i\xi_j \ge \alpha_M \sum_{i=1}^{N} \xi_i^2$$
$$\text{for a.e. } x \in \Omega \text{ and } \xi = (\xi_1, \ldots, \xi_N) \in \mathbb{R}^N$$

For given data $f \in L^2(\Omega), h \in L^2(\Gamma)$, we look for a solution $u$ of the problem

$$\begin{aligned}-\nabla \cdot (M\nabla u) + b \cdot \nabla u + cu &= f \quad \text{in } \Omega\\ au + n \cdot (M\nabla u) &= h \quad \text{on } \Gamma\end{aligned} \qquad [22]$$

We assume that $a \in L^\infty(\Gamma)$. The variational formulation of this problem is still [8], with $V = H^1(\Omega)$ and

$$\begin{aligned}\mathcal{A}(u,v) = &\int_\Omega M\nabla u \cdot \nabla v dx\\ &+ \int_\Omega [b \cdot \nabla u + cu]v dx + \int_\Gamma a\gamma_0 u \gamma_0 v d\sigma\end{aligned} \qquad [23]$$

$$L(v) = \int_\Omega f(x)v(x)dx + \int_\Gamma (h\gamma_0 v)(\sigma)\,d\sigma \qquad [24]$$

If the conditions

$$\begin{aligned}&c - \tfrac{1}{2}\nabla \cdot b \ge C_0 \ge 0 \quad \text{a.e. on } \Omega\\ &a + \tfrac{1}{2} b \cdot \nu \ge C_1 \ge 0 \quad \text{a.e. on } \Gamma\end{aligned}$$

are fulfilled, with $(C_0, C_1) \ne (0,0)$, then the bilinear form $\mathcal{A}$ is $V$-elliptic and the Lax–Milgram theorem applies.

### A Biharmonic Problem

We consider the Dirichlet problem for the operator of fourth order: ($c \in L^\infty(\Omega)$):

$$\Delta^2 u + cu = f \quad \text{in } \Omega \qquad [25]$$

$$u = u_0 \text{ on } \Gamma, \ \frac{\partial u}{\partial n} = h \text{ on } \Gamma \qquad [26]$$

**Theorem 11** *Let us suppose that $\Omega$ has a boundary of class $\mathcal{C}^{1,1}$ and that the data satisfy $f \in H^{-2}(\Omega)$, $u_0 \in H^{3/2}(\Gamma)$, $h \in H^{1/2}(\Gamma)$. Let $U_0 \in H^2(\Omega)$ be such that $\gamma_0(U_0) = u_0$, $\gamma_1(U_0) = h$. Then, if $c \ge 0$ a.e. in $\Omega$, the boundary value problem* [25]–[26] *has a unique*

*solution u such that* $u - U_0 \in H_0^2(\Omega)$, *and u is also the unique solution of the variational problem*

$$\text{to find } u \in U_0 + H_0^2(\Omega) \text{ such that } \mathcal{A}(u,v) = l(v) \text{ for all } v \in H_0^2(\Omega) \qquad [27]$$

*where* $l(v) = \langle f, v\rangle_{H^{-2}(\Omega), H_0^2(\Omega)}$ *and*

$$\mathcal{A}(u,v) = \int_\Omega \Delta u(x)\Delta v(x)\mathrm{d}x + \int_\Omega (cuv)(x)\mathrm{d}x \qquad [28]$$

*Moreover, there exists a positive constant* $C = C(\Omega)$ *such that*

$$\|u\|_{H^2(\Omega)} \le C\,[\|f\|_{H^{-2}(\Omega)} + \|u_0\|_{H^{3/2}(\Gamma)} + \|h\|_{H^{1/2}(\Gamma)}] \qquad [29]$$

*which shows that u depends continuously upon the data f*, $u_0$, *and h*.

**Remark 12** The Hilbert space choice $V$ is of crucial importance for the $V$-ellipticity. In fact, let us consider for example the problem [25], with

$$\Delta u = 0 \text{ on } \Gamma, \quad \frac{\partial \Delta u}{\partial n} = 0 \text{ on } \Gamma \qquad [30]$$

In fact, the associated bilinear form is not $V$-elliptic for $V = H^2(\Omega)$ but it is $V$-elliptic for $V = \{v \in L^2\,(\Omega);\ \Delta v \in L^2\Omega)\}$.

## General Elliptic Problems

Here $\Omega$ will be a bounded and sufficiently regular open subset of $\mathbb{R}^N$. Let us consider a general linear differential operator of the form

$$A(x,D)u = \sum_{|\mu| \le l} a_\mu(x) D^\mu u, \quad a_\mu(x) \in \mathbb{C} \qquad [31]$$

Setting $A_0(x,\xi) = \sum_{|\mu|=l} a_\mu(x)\xi^\mu$, we say that the operator $A$ is elliptic at a point $x$ if $A_0(x,\xi) \ne 0$ for all $\xi \in \mathbb{R}^N - \{0\}$. One can show that, if $N \ge 3$, $l$ is even, that is, $l = 2m$; the same result holds for $N = 2$ if the coefficients $a_\mu$ are real. Moreover, for $N \ge 3$, every elliptic operator is properly elliptic, in the following sense: for any independent vectors $\xi$, $\xi'$ in $\mathbb{R}^N$, the polynomial $\tau \mapsto A_0(.,\xi + \tau\xi')$ has $m$ roots with positive imaginary part.

The aim here is to study boundary-value problems of the following type:

$$Au = f \quad \text{in } \Omega \qquad [32]$$

$$B_j u = g_j \quad \text{on } \Gamma, \quad j = 0, \ldots, m-1, \qquad [33]$$

where $A$ is properly elliptic on $\bar{\Omega}$, with sufficiently regular coefficients, and the operators $B_j$ are boundary operators, of order $m_j \le 2m-1$, that must satisfy some compatibility conditions with respect to the operator $A$ (see Renardy and Rogers (1992) for details; these conditions were introduced by Agmon, Douglis, and Nirenberg). For example, $A = (-1)^m\Delta^m$ and $B_j = \partial^j/\partial n^j$ is a convenient choice.

In order to show that problem [32]–[33] has a solution $u \in H^{2m+r}(\Omega)$ ($r \in \mathbb{N}$), the idea is to show that the operator $\mathcal{P}$ defined by $u \mapsto \mathcal{P}(u) = (Au, B_0u, \ldots, B_{m-1}u)$ is an index operator from $H^{2m+r}(\Omega)$ into $G = H^r(\Omega) \times \Pi_{j=0}^{m-1} H^{2m+r-m_j-1/2}(\Gamma)$ and to express the compatibility conditions through the adjoint problem.

We recall that a linear continuous operator $\mathcal{P}$ is an index operator if

(1) $\dim \operatorname{Ker} \mathcal{P} < \infty$, and $\operatorname{Im} \mathcal{P}$ closed;
(2) $\operatorname{codim} \operatorname{Im} \mathcal{P} < \infty$.

Then the index $\chi(\mathcal{P})$ is given by $\chi(\mathcal{P}) = \dim \operatorname{Ker} \mathcal{P} - \operatorname{codim} \operatorname{Im} \mathcal{P}$. We recall the following Peetre's theorem:

**Theorem 13** *Let E, F, and G be three reflexive Banach spaces such that* $E \hookrightarrow\hookrightarrow F$, *and* $\mathcal{P}$ *a linear continuous operator from E to G. Then condition (1) is equivalent to: "there exists* $C \ge 0$, *such that for all* $u \in E$, *we have* $\|u\|_E \le C(\|\mathcal{P}u\|_G + \|u\|_F)$."

Applying this theorem to our problem [32]–[33], condition (1) results from *a priori* estimates of the following type:

$$\|u\|_{H^{2m+r}(\Omega)} \le C\left(\|\mathcal{P}u\|_G + \|u\|_{H^{2m+r-1}(\Omega)}\right)$$

and condition (2) by similar *a priori* estimates for the dual problem.

## Second-Order Elliptic Problems

We consider a second-order differential operator of the "divergence form"

$$Au = -\sum_{i,j=1}^N (a^{ij}(x)u_{x_i})_{x_j} + \sum_{i=1}^N b^i(x)u_{x_i} + c(x)u \qquad [34]$$

with given coefficient functions $a^{ij}, b^i, c\,(i,j = 1, \ldots, N)$, and where we have used the notation $u_{x_i} = \frac{\partial u}{\partial x_i}$. Such operators are said uniformly strongly elliptic in $\Omega$ if there exists $\alpha > 0$ such that

$$\sum_{|i|=|j|=1} a^{ij}(x)\xi^i\xi^j \ge \alpha|\xi|^2 \quad \text{for all } x \in \Omega,\ \xi \in \mathbb{R}^N$$

**Remark 14** There exist elliptic problems for which the associated variational problem does not necessarily satisfy the ellipticity condition. Let us consider

the following example, due to Seeley: let $\Omega = \{(r,\theta) \in (\pi, 2\pi) \times [0, 2\pi]\}$ and

$$A = -\left(e^{i\theta}\frac{\partial}{\partial\theta}\right)^2 - e^{2i\theta}\left(1 + \frac{\partial^2}{\partial r^2}\right)$$

One can check that, for all $\lambda \in \mathbb{C}$, the problem $Au + \lambda u = f$ in $\Omega$ and $u = 0$ on $\Gamma$ admits nonzero solutions $u$ which are given by (with $\mu$ such that $\mu^2 = \lambda$) $u = \sin r \cos(\mu e^{-i\theta})$ and $u = \sin r \sin(\mu e^{-i\theta})$ for $\lambda \neq 0$; $u = \sin r$ and $u = \sin\theta\, e^{-i\theta}$ for $\lambda = 0$.

Most of the results concerning existence, unicity, and regularity for second-order elliptic problems can be established thanks to a maximum principle. There exist different types of maximum principles, which we now present.

### Maximum Principle

**Theorem 15** (Weak maximum principle). *Let $A$ be a uniformly strongly elliptic operator of the form [34] in a bounded open $\Omega \subset \mathbb{R}^N$, with $a^{ij}, b^i, c \in L^\infty(\Omega)$ and $c \geq 0$. Let $u \in C^2(\Omega) \cap C(\bar{\Omega})$ and*

$$Au \geq 0 \ [\textit{resp. } Au \leq 0] \quad \textit{in } \Omega$$

*Then*

$$\inf_{\Omega} u \geq \inf_{\partial\Omega} u^- \ [\textit{resp. } \sup_{\Omega} u \leq \sup_{\partial\Omega} u^+]$$

*where $u^+ = \max(u, 0)$ and $u^- = -\min(u, 0)$. If $c = 0$ in $\Omega$, one can replace $u^-$ [resp. $u^+$] by $u$.*

**Theorem 16** (Strong principle maximum). *Under the assumptions of the above theorem, if $u$ is not a constant function in $C^2(\Omega) \cap C(\bar{\Omega})$ such that $Au \geq 0$ [resp. $Au \leq 0$], then $\inf_\Omega u < u(x)$ [resp. $\sup_\Omega u > u(x)$], for all $x \in \Omega$.*

**Remark 17** These two maximum principles can be adapted to elliptic operators in nondivergence form, that is,

$$Au = -\sum_{i,j=1}^{N} a^{ij}(x) u_{x_i x_j} + \sum_{i=1}^{N} b^i(x) u_{x_i} + c(x) u \qquad [35]$$

### Fredholm Alternative

We now present some existence results which are based on the Fredholm alternative rather than on the variational method.

Let us consider two Hilbert spaces $V$ and $H$, where $V$ is a dense subspace of $H$ and $V \hookrightarrow\hookrightarrow H$. Denoting by $V'$ the dual space of $V$, and identifying $H$ with its dual space, we have the following imbeddings: $V \hookrightarrow H \hookrightarrow V'$. Let $\mathcal{A}$ be a sesquilinear form on $V \times V$, $V$-coercive with respect to $H$, that is, there exist $\lambda_0 \in \mathbb{R}$ and $\alpha > 0$ such that

$$\mathrm{Re}(\mathcal{A}(v,v)) + \lambda_0 \|v\|_H^2 \geq \alpha \|v\|_V^2 \quad \text{for all } v \in V$$

Denoting by $A$ the operator associated with the bilinear form $\mathcal{A}$ (see Remark 6(iii)), the equation $Au = f$ is equivalent to $u - \lambda_0 Tu = g$, with $T = (A + \lambda_0 \mathrm{Id})^{-1}$ and $g = Tf$. Note that $T$ is an isomorphism from $H$ onto $D(A) = \{u \in H; Au \in H\}$).

The operator $T: H \to H$ is compact and, thanks to the Fredholm alternative, there are two situations:

1. either $\mathrm{Ker}\, A = 0$ and $A$ is an isomorphism from $D(A)$ onto $H$;
2. or $\mathrm{Ker}\, A \neq 0$; then $\mathrm{Ker}\, A$ is of finite dimension, and the problem $Au = f$ with $f \in H$ admits a solution if and only if $f \in \mathrm{Im}\, A = [\mathrm{Ker}(A^*)]^\perp$.

We now give another example in a non-Hilbertian frame. Let us consider the problem (Grisvard 1980): $Au = f$ in $\Omega$ and $Bu = g$ on $\Gamma$, where $\Gamma$ is of class $C^{1,1}$, $A$, which is defined by [34], is uniformly strongly elliptic with $a^{ij} = a^{ji} \in C^{0,1}(\bar{\Omega})$, $b^i, c \in L^\infty(\Omega)$, and $Bu = \gamma_0(u)$ or $Bu = \gamma_1(u)$. One can show that the operator $u \mapsto (Au, Bu)$ is a Fredholm operator of index zero from $W^{2,p}(\Omega)$ in $L^p(\Omega) \times W^{2-d-1/p,p}(\Gamma)$ (with $d = 0$ if $Bu = \gamma_0(u)$ and $d = 1$ if $Bu = \gamma_1(u)$).

### Regularity

Assume that $\Omega$ is a bounded open. Suppose that $u \in H_0^1(\Omega)$ is a weak solution of the equation

$$\begin{aligned} Au &= f \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \Gamma \end{aligned} \qquad [36]$$

where $A$ has the divergence form [34]. We now address the question whether $u$ is in fact smooth: this is the regularity problem for weak solutions.

**Theorem 18** ($H^2$-regularity). *Let $\Omega$ be open, of class $C^{1,1}$, $a^{ij} \in C^1(\bar{\Omega})$, $b^i, c \in L^\infty(\Omega)$, $f \in L^2(\Omega)$. Suppose, furthermore, that $u \in H^1(\Omega)$ is a weak solution of [36]. Then $u \in H^2(\Omega)$ and we have the estimate*

$$\|u\|_{H^2(\Omega)} \leq C(\|f\|_{L^2(\Omega)} + \|u\|_{L^2(\Omega)})$$

*where the constant $C$ depends only on $\Omega$ and on the coefficients of $A$.*

**Theorem 19** (Higher regularity). *Let $m$ be a nonnegative integer, $\Omega$ be open, of class $C^{m+1,1}$ and assume that $a^{ij} \in C^{m+1}(\bar{\Omega})$, $b^i$, $c \in C^{m+1}(\bar{\Omega})$, $f \in H^m(\Omega)$. Suppose, furthermore, that $u \in H^1(\Omega)$ is a weak solution of [36]. Then $u \in H^{m+2}(\Omega)$ and*

$$\|u\|_{H^{m+2}(\Omega)} \leq C(\|f\|_{H^m(\Omega)} + \|u\|_{L^2(\Omega)})$$

*where the constant C depends only on $\Omega$ and on the coefficients of A. In particular, if $m > N/2$, then $u \in C^2(\bar{\Omega})$. Moreover, if $\Omega$ is of $C^\infty$ class and $f \in C^\infty(\bar{\Omega})$, $a^{ij} \in C^\infty(\bar{\Omega})$, $b^i, c \in C^\infty(\bar{\Omega})$, then $u \in C^\infty(\bar{\Omega})$.*

**Remark 20**

(i) If $u \in H_0^1(\Omega)$ is the unique solution of [36], one can omit the $L^2$-norm of $u$ in the right-hand side of the above estimate.
(ii) Moreover, let us suppose the coefficients $a^{ij}, b^i$ and $c$ are all $C^\infty$ and $f \in C^\infty(\Omega)$; then, if $u \in H^1(\Omega)$ satisfies $Au = f, u \in C^\infty(\Omega)$; this is due to the "hypoellipticity" property satisfied by the operator $A$.

We have a similar result in the $L^p$ frame (Grisvard 1980):

**Theorem 21** ($W^{2,p}$-regularity). *Let $\Omega$ be open, of class $C^{1,1}$, $a^{ij} \in C^1(\bar{\Omega})$, $b^i$, $c \in L^\infty(\Omega)$. Suppose, furthermore, that $b^i = 0$, $1 \le i \le N$ and $c \ge 0$ a.e. Then for every $f \in L^p(\Omega)$ there exists a unique solution $u \in W^{2,p}(\Omega)$ of [36].*

## Unbounded Open

### The Whole Space

Note in passing that we shall work with the weighted Sobolev spaces $W_\alpha^{m,p}(\Omega)$ defined in the subsection "Unbounded opens and weighted spaces."

**Theorem 22** *The following claims hold true:*

(i) *Let $f \in W_0^{-1,p}(\mathbb{R}^N)$ satisfy the compatibility condition*

$$\langle f, 1\rangle_{W_0^{-1,p}(\mathbb{R}^N)\times W_0^{1,p'}(\mathbb{R}^N)} = 0 \quad \textit{if } p' \ge N$$

*Then the problem [5] has a solution $u \in W_0^{1,p}(\mathbb{R}^N)$, which is unique up to an element in $\mathcal{P}_{[1-N/p]}$ and satisfies the estimate*

$$\|u\|_{W_0^{1,p}(\mathbb{R}^N)/\mathcal{P}_{[1-N/p]}} \le C\|f\|_{W_0^{-1,p}(\mathbb{R}^N)}$$

*Moreover, if $1 < p < N$, then $u = E * f$.*

(ii) *If $f \in L^p(\mathbb{R}^N)$, then the problem [5] has a solution $u \in W_0^{2,p}(\mathbb{R}^N)$, which is unique up to an element in $\mathcal{P}_{[2-N/p]}$ and if $1 < p < N/2$, then $u = E * f$.*

The Calderón–Zygmund inequality

$$\left\|\frac{\partial^2 \varphi}{\partial x_i \partial x_j}\right\|_{L^p(\mathbb{R}^N)} \le C(N,p)\|\Delta\varphi\|_{L^p(\mathbb{R}^N)}$$
$$\varphi \in \mathcal{D}(\mathbb{R}^N)$$

and Theorem 4 are crucial for establishing Theorem 22.

Further, point (i) means that the Riesz potential of second order satisfies

$$I_2\colon W_0^{-1,p}(\mathbb{R}^N) \perp \mathcal{P}_{[1-N/p']} \to W_0^{1,p}(\mathbb{R}^N)/\mathcal{P}_{[1-\frac{N}{p'}]}$$

(where the initial space is the orthogonal complement of $\mathcal{P}_{[1-N/p']}$ in $W_0^{-1,p}(\mathbb{R}^N)$) and it is an isomorphism.

Note that here

$$W_0^{1,p}(\mathbb{R}^N) = \{v \in L^{p^*}(\mathbb{R}^N); \nabla v \in L^p(\mathbb{R}^N)\}$$

for $1 < p < N$ and $1/p^* = 1/p - 1/N$. And for $1 < r < N/2$, we also have the continuity property

$$I_2\colon L^r(\mathbb{R}^N) \to L^q(\mathbb{R}^N), \text{ for } \frac{1}{q} = \frac{1}{r} - \frac{2}{N}$$

**Remark 23** The problem

$$u - \Delta u = f \quad \text{in } \mathbb{R}^N \qquad [37]$$

is of a completely different nature than the problem [5]. The class of function spaces appropriate for the problem [37] are the classical Sobolev spaces. With the help of the Calderón–Zygmund theory, one can prove that if $f \in L^p(\mathbb{R}^N)$, then the unique solution of [37] belongs to $W^{2,p}(\mathbb{R}^N)$ and can be represented as the Bessel potential of second order (see Stein (1970)): $u = G * f$, where $G$ is the appropriate Bessel kernel, that is, $G$, for which $\widehat{G}(\xi) \sim (1 + |\xi|^2)^{-1/2}$. Recall that in particular $G(x) \sim |x|^{-1}e^{-|x|}$ for $N = 3$. In the Hilbert case, $f \in L^2(\mathbb{R}^N)$, we get

$$(1 + |\xi|^2)\widehat{u} \in L^2(\mathbb{R}^N)$$

which, by Plancherel's theorem, implies that $u \in H^2(\mathbb{R}^N)$. For $f \in W^{-1,p}(\mathbb{R}^N)$, the problem [37] has a unique solution $u \in W^{1,p}(\mathbb{R}^N)$ satisfying the estimate

$$\|u\|_{W^{1,p}(\mathbb{R}^N)} \le C(p,n)\|f\|_{W^{-1,p}(\mathbb{R}^N)}$$

### Exterior Domain

We consider the problem in an exterior domain with the Dirichlet boundary condition

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega \\ u &= g \quad \text{on } \Gamma = \partial\Omega \end{aligned} \qquad [38]$$

where $f \in W_0^{-1,p}(\Omega)$ and $g \in W^{1-1/p,p}(\partial\Omega)$. Invoking the results for $\mathbb{R}^N$ and bounded domains, one can prove the existence of a solution $u \in W_0^{1,p}(\Omega)$ which is unique up to an element of the kernel $A_0^p(\Omega) = \{z \in$

$W_0^{1,p}(\Omega); \Delta z = 0\}$ provided that $f$ satisfies the compatibility condition

$$\langle f, \varphi \rangle = \left\langle g, \frac{\partial \varphi}{\partial n} \right\rangle \quad \text{for all } \varphi \in A_0^{p'}(\Omega)$$

The kernel can be characterized in the following way: it is reduced to $\{0\}$ if $p=2$ or $p<N$ and if not, then

$$A_0^p(\Omega) = \{C(\lambda - 1); C \in \mathbb{R}\} \quad \text{if } p \geq N \geq 3$$

where $\lambda$ is (unique) solution in $W_0^{1,2}(\Omega) \cap W_0^{1,p}(\Omega)$ of the problem $\Delta\lambda = 0$ in $\Omega$ and $\lambda = 1$ on $\partial\Omega$, and

$$A_0^p(\Omega) = \{C(\mu - u_0);\ C \in \mathbb{R}\} \quad \text{if } p > N = 2$$

where $u_0(x) = (2\pi|\Gamma|)^{-1} \int_\Gamma \log|y - x|\, \mathrm{d}\sigma_y$ and $\mu$ is the only solution in $W_0^{1,2}(\Omega) \cap W_0^{1,p}(\Omega)$ of the problem $\Delta\mu = 0$ in $\Omega$ and $\mu = u_0$ on $\Gamma$.

**Remark 24** Similar results exist for the Neumann problem in an exterior domain (see Amrouche *et al.* (1997)). The framework of the spaces $W_\alpha^{m,p}(\mathbb{R}_+^N)$ for the Dirichlet problem in $\mathbb{R}_+^N$ was also considered in the literature. For a more general theory see Kozlov and Maz'ya (1999).

## Elliptic Systems

### The Stokes System

The Stokes problem is a classical example in the fluid mechanics. This system models the slow motion with the field of the velocity $\boldsymbol{u}$ and the pressure $\pi$, satisfying

$$(\mathrm{S}) \qquad \begin{aligned} -\nu\Delta\boldsymbol{u} + \nabla\pi &= \boldsymbol{f} \quad \text{in} \quad \Omega \\ \operatorname{div}\boldsymbol{u} &= h \quad \text{in} \quad \Omega \\ \boldsymbol{u} &= \boldsymbol{g} \quad \text{on} \quad \Gamma = \partial\Omega \end{aligned}$$

where $\nu > 0$ denotes the viscosity, $\boldsymbol{f}$ is an exterior force, $\boldsymbol{g}$ is the velocity of the fluid on the domain boundary, and $h$ measures the compressibility of the fluids (if $h=0$, it is an incompressible fluid). The functions $h$ and $\boldsymbol{g}$ must satisfy the compatibility condition

$$\int_\Omega h(x)\,\mathrm{d}x = \int_\Gamma \boldsymbol{g} \cdot \boldsymbol{n}\,\mathrm{d}\sigma \qquad [39]$$

**Theorem 25** *Let $\Omega$ be a Lipschitz bounded domain in $\mathbb{R}^N$, $N \geq 2$. Let $\boldsymbol{f} \in H^{-1}(\Omega)^N$, $h \in L^2(\Omega)$, and $g \in H^{1/2}(\Gamma)^N$ satisfy [39]. Then the problem* (S) *has a unique solution $(\boldsymbol{u}, \pi) \in H^1(\Omega)^N \times L^2(\Omega)/\mathbb{R}$ satisfying the* a priori *estimate*

$$\begin{aligned} &\|\boldsymbol{u}\|_{H^1(\Omega)} + \|\pi\|_{L^2(\Omega)/\mathbb{R}} \\ &\quad \leq C\left(\|\boldsymbol{f}\|_{H^{-1}(\Omega)} + \|h\|_{L^2(\Omega)} + \|g\|_{H^{1/2}(\Gamma)}\right) \end{aligned}$$

In order to prove Theorem 25, one can start with a homogeneous problem. The procedure of finding $\boldsymbol{u}$ is a simple application of the Lax–Milgram theorem. Application of de Rham's theorem gives the pressure $\pi$. We introduce the space

$$\mathcal{V} = \{\boldsymbol{v} \in \mathcal{D}(\Omega)^N;\ \operatorname{div}\boldsymbol{v} = 0\}$$

and define $\boldsymbol{F} \in H^{-1}(\Omega)^N$ by

$$\langle \boldsymbol{F}, \boldsymbol{v} \rangle_{H^{-1} \times \overset{\circ}{H}{}^1} = 0 \quad \text{for all } \boldsymbol{v} \in \mathcal{V}$$

Moreover, there exists $\pi \in L^2(\Omega)$, unique up to an additive constant, and such that $\boldsymbol{F} = \nabla\pi$. The problem (S), which we transform to the homogeneous case ($h=0, g=0$), can be formulated on an abstract level. Let $X$ and $M$ be two real Hilbert spaces and consider the following variational problem: Given $\boldsymbol{L} \in \boldsymbol{X}'$ and $X \in M'$, find $(\boldsymbol{u}, \pi) \in X \times M$ such that

$$\begin{aligned} \mathcal{A}(\boldsymbol{u}, \boldsymbol{v}) + B[\boldsymbol{v}, \pi] &= \boldsymbol{L}(\boldsymbol{v}), \quad \boldsymbol{v} \in \boldsymbol{X} \\ B[\boldsymbol{u}, q] &= X(q), \quad q \in M \end{aligned} \qquad [40]$$

where the bilinear forms $\mathcal{A}$, $B$ and the linear form $\boldsymbol{L}$ are defined by

$$\mathcal{A}(\boldsymbol{u}, \boldsymbol{v}) = \int_\Omega \nabla\boldsymbol{u} \cdot \nabla\boldsymbol{v}$$

$$B[\boldsymbol{v}, q] = -\int_\Omega [q\nabla \cdot \boldsymbol{v}]$$

$$L(\boldsymbol{v}) = \int_\Omega \boldsymbol{f} \cdot \boldsymbol{v}$$

**Theorem 26** *If the bilinear form $\mathcal{A}$ is coercive in the space*

$$V = \{\boldsymbol{v} \in X;\ B[\boldsymbol{v}, q] = 0\} \quad \text{for all } q \in M$$

*that is, if there exists $\alpha > 0$ such that*

$$\mathcal{A}(\boldsymbol{v}, \boldsymbol{v}) \geq \alpha\|\boldsymbol{v}\|_X^2, \quad \boldsymbol{v} \in V$$

*then the problem [40] has a unique solution $(\boldsymbol{u}, \pi)$ if and only if the bilinear form $\mathcal{B}$ satisfies the* "inf–sup" *condition*:

*there exists $\beta > 0$ such that*

$$\inf_{q \in M} \sup_{v \in X} \frac{\mathcal{B}(\boldsymbol{v}, q)}{\|\boldsymbol{v}\|_X \|q\|_M} \geq \beta$$

As for the Dirichlet problem, the regularity result is the following:

**Theorem 27** *Let $\Omega$ be a bounded domain in $\mathbb{R}^N$, of the class $C^{m+1,1}$ if $m \in \mathbb{N}$ and $C^{1,1}$ if $m = -1$. Let $f \in W^{m,p}(\Omega)^N, h \in W^{m+1,p}(\Omega)$ and $\boldsymbol{g} \in W^{m+2-1/p,p}(\Gamma)^N$ satisfy condition [39]. Then the problem* (S) *has a unique solution $(\boldsymbol{u}, \pi) \in W^{m+2,p}(\Omega)^N \times W^{m+1,p}(\Omega)/\mathbb{R}$.*

**Remark 28** It is possible to solve (S) under weaker assumption, for instance, if $f \in W^{-1/p}(\Omega')$, $h=0$ and $g \in W^{-1/p,p}(\Gamma)^N$. We can prove that then $(u,\pi) \in L^p(\Omega)^N \times W^{-1,p}(\Omega)$.

### The Linearized Elasticity

The equations governing the displacement $u=(u_1,u_2,u_3)$ of a three-dimensional structure subjected to an external force field $f$ are written as ($\Omega$ is a bounded open subset of $\mathbb{R}^3$ and $\Gamma=\partial\Omega$)

$$-\mu\Delta u - (\lambda+\mu)\nabla(\nabla\cdot u) = f \quad \text{in } \Omega$$
$$u = 0 \quad \text{on } \Gamma_0$$
$$\sum_{j=1}^{3} \sigma_{ij}(u)\nu_j = g_i \quad \text{on } \Gamma_1 = \Gamma - \Gamma_0$$

where $\lambda > 0$ and $\mu > 0$ are two material characteristic constants, called the Lamé coefficients, and ($v=(v_1,v_2,v_3)$)

$$\sigma_{ij}(v) = \sigma_{ji}(v) = \lambda\delta_{ij}\sum_{k=1}^{3}\varepsilon_{kk}(v) + 2\mu\varepsilon_{ij}(v) \qquad [41]$$
$$\text{with } \varepsilon_{ij}(v) = \varepsilon_{ji}(v) = \tfrac{1}{2}(\partial_j v_i + \partial_i v_j)$$

where $\delta_{ij}$ denotes the Kronecker symbol, that is, $\delta_{ij}=1$, for $i=j$ and $\delta_{ij}=0$, for $i\neq j$. These equations describe the equilibrium of an elastic homogeneous isotropic body that cannot move along $\Gamma_0$; along $\Gamma_1$, surface forces of density $g=(g_1,g_2,g_3)$ are given. The case $\Gamma_1=\emptyset$ physically corresponds to clamped structures. The matrix with entries $\varepsilon_{ij}(u)$ is the linearized strain tensor while $\sigma_{ij}(u)$ represents the linearized stress tensor; the relationship [41] between these tensors is known as Hooke's law. We refer for example to Ciarlet and Lions (1991) and Nečas and Hlaváček (1981) (and references therein) for most of the results stated in this paragraph. The variational formulation of this problem is

$$\text{to find } u \in V \text{ such that}$$
$$\mathcal{A}(u,v) = L(v) \text{ for all } v \in V \qquad [42]$$

where the bilinear form $\mathcal{A}$ and the linear form $L$ are given by

$$\mathcal{A}(u,v) = \int_\Omega [\lambda(\nabla\cdot u)(\nabla\cdot v) + 2\mu\sum_{i,j=1}^{3}\varepsilon_{ij}(u)\varepsilon_{ij}(v)](x)dx, \qquad [43a]$$

$$L(v) = \int_\Omega f(x)\cdot v(x)dx + \int_{\Gamma_1} g(\sigma)\cdot v(\sigma)d\sigma \qquad [43b]$$

The functional space $V$ is defined as

$$V = \{v=(v_1,v_2,v_3) \in [H^1(\Omega)]^3; \ \gamma_0 v_i = 0 \text{ on } \Gamma_0,\ 1\le i\le 3\}$$

To prove the ellipticity of $\mathcal{A}$, one needs the following Korn inequality: There exists a positive constant $C(\Omega)$ such that, for all $v=(v_1,v_2,v_3) \in [H^1(\Omega)]^3$, we have

$$\|v\|_{1,\Omega} \le C(\Omega)\left[\sum_{i,j=1}^{3}\|\varepsilon_{ij}(v)\|^2_{L^2(\Omega)} + \sum_{i=1}^{3}\|v_i\|^2_{L^2(\Omega)}\right]^{1/2} \qquad [44]$$

The following result holds true:

**Theorem 29** *Let $\Omega$ be a bounded open in $\mathbb{R}^3$ with a Lipschitz boundary, and let $\Gamma_0$ be a measurable subset of $\Gamma$, whose measure (with respect to the surface measure $d\Gamma(x)$) is positive. Then the mapping*

$$v \mapsto \left[\sum_{i,j=1}^{3}\|\varepsilon_{ij}(v)\|^2_{L^2(\Omega)}\right]^{1/2}$$

*is a norm on $V$, equivalent to the usual norm $\|.\|_{1,\Omega}$.*

As a consequence, we get:

**Theorem 30** *Under the above assumptions, there exists a unique $u \in V$ solving the variational problem [42]–[43]. This solution is also the unique one which minimizes the energy functional*

$$E(v) = \frac{1}{2}\int_\Omega\left[\lambda(\nabla\cdot v)^2 + 2\mu\sum_{i,j=1}^{3}[\varepsilon_{ij}(v)]^2\right](x)\,dx - \left[\int_\Omega f(x)\cdot v(x)\,dx + \int_{\Gamma_1} g(\sigma)\cdot v(\sigma)\,d\sigma\right]$$

*over the space $V$.*

## Acknowledgmnt

M Krbec and Š Nečasová were supported by the Institutional Research Plan No. AV0Z10190503 and by the Grant Agency of the Academy of Sciences No. IAA10019505.

*See also:* Evolution Equations: Linear and Nonlinear; $\Gamma$-Convergence and Homogenization; Image Processing: Mathematics; Inequalities in Sobolev Spaces; Partial Differential Equations: Some Examples; Schrödinger Operators; Separation of Variables for Differential Equations; Viscous Incompressible Fluids: Mathematical Theory.

## Further Reading

Adams RA and Fournier JJF (2003) *Sobolev Spaces*, 2nd edn. Amsterdam: Academic Press.

Agmon S (1965) *Lectures on Elliptic Boundary Value Problems.* Princeton: D. van Nostrand.

Amrouche C, Girault V, and Giroire J (1997) Dirichlet and Neumann exterior problems for the *n*-dimensional Laplace operator. An approach in weighted Sobolev spaces. *Journal de Mathématiques Pures et Appliquées IX. Série* 76: 55–81.
Ciarlet PG and Lions JL (1991) *Handbook of Numerical Analysis, Vol. 2, Finite Element Methods*. Amsterdam: North-Holland.
Dautray R and Lions JL (1988) *Analyse mathématique et calcul numérique pour les sciences et les techniques*. Paris: Masson.
Friedman A (1969) *Partial Differential Equations*. New York: Holt, Rinehart and Winston.
Gilbarg D and Trudinger N (1977) *Elliptic Partial Differential Equations of Second Order*. Berlin: Springer.
Grisvard P (1980) *Boundary Value Problems in Non-Smooth Domains*, Lecture Notes 19. University of Maryland, Department of Mathematics.
Hörmander L (1964) *Linear Partial Differential Operators*. Berlin: Springer.
Kozlov V and Maz'ya V (1999) *Differential Equations with Operators Coefficients with Applications to Boundary Value Problems for Partial Differential Equations*. Berlin: Springer.
Ladyzhenskaya O and Uraltseva N (1968) *Linear and Quasilinear Elliptic Equations*. New York: Academic Press.
Lions J-L and Magenes E (1968) *Problèmes aux limites non homogènes et applications*, vol. 1. Paris: Dunod.
Nečas J (1967) *Les méthodes directes en équations elliptiques*. Prague: Academia.
Nečas J and Hlaváček I (1981) *Mathematical Theory of Elastic and Elastoplastic Bodies: An Introduction*. Amsterdam: Elsevier.
Renardy M and Rogers RC (1992) *An Introduction to Partial Differential Equations*. New York: Springer.
Stein EM (1970) *Singular Integrals and Differentiability Properties of Functions*. Princeton, NJ: Princeton University Press.
Triebel H (1978) *Interpolation Theory, Function Spaces, Differential Operators*, VEB Deutsch. Verl. Wissenschaften, Berlin. Sec. rev. enl. ed. (1995): Barth, Leipzig.
Triebel H (2001) *The Structure of Functions*. Basel: Birkhäuser.
Weinberger HF (1965) *A First Course in Partial Differential Equations*. New York: Wiley.
Ziemer WP (1989) *Weakly Differentiable Functions*. New York: Springer.

# Entanglement

**R F Werner**, Technische Universität Braunschweig, Braunschweig, Germany



## Introduction

Entanglement is a type of correlation between subsystems, which cannot be explained by the action of a classical random generator. It is a key notion of quantum information theory and corresponds closely to the possibility of channels which transmit quantum information, and cannot be simulated by classical channels. In this article, we consider the development of the concept, and its qualitative aspects. The quantitative aspects are treated in a separate article (*see* Entanglement Measures).

## Historical Development

The first realization that quantum mechanics comes with new, and perhaps rather strange, correlations came in the famous 1935 paper by Einstein, Podolsky, and Rosen (EPR) (Einstein *et al.* 1935), in which they set up a paradox showing that the statistics of certain quantum states could not be realized by assigning wave functions to subsystems. It was in response to this paper that Schrödinger (1935), in the same year, coined the term "entanglement," as well as its German equivalent "*Verschränkung*." The subject lay dormant for a long time, since Bohr, in his reply, completely ignored the entanglement theme, and there was a widespread reluctance in the physics community to consider problems of interpretation. The leaf turned slowly with Bohm's reduced model of the EPR paradox using spins rather than continuous variables, and decisively with Bell's 1964 strengthening of the paradox (Bell 1964). He showed that not only wave functions assigned to individual systems failed to describe the correlations predicted by quantum mechanics, but any set of classical parameters assigned to the subsystems. This eliminated all reference to a possibly dubious quantum ontology and all reference to the quantum formalism from the argument. Bell derived a set of inequalities from the assumption that each subsystem could be described in terms of classical variables, and that these (possibly hidden) variables would not be changed by the mere choice of a measurement for the distant correlated system. The only relation to quantum mechanics was then the simple quantum calculation showing, in certain situations, such as the state described by EPR, quantum mechanics predicted a violation of Bell's inequalities. This immediately suggested an experiment, and although it was difficult at first to find an efficient source of suitably quantum-correlated pairs of particles, the experiments that have been made since then have supported the quantum-mechanical result beyond reasonable doubt. This came too late for Einstein, whose research program in quantum mechanics had been precisely to build a "local hidden-variable theory" of the type seen in contradiction with Bell's inequality. But at least the EPR paper had finally received the response it deserved.

In Schrödinger's work, entanglement was a purely qualitative term for the strange way the subsystems

seemed to be intertwined as soon as one insisted on discussing their individual properties. After Bell's work, the favored mathematical definition of entanglement would probably have been the existence of measurements on the subsystems, such that Bell's inequality (or some generalization derived on the same assumptions) is violated. However, around 1983 another notion of (the lack of) entanglement was independently proposed by Primas (1983) and Werner (1983). According to this definition, a quantum state $\rho$ is called unentangled if it can be written as

$$\rho = \sum_{\alpha} p_\alpha \rho_\alpha^1 \otimes \rho_\alpha^2 \qquad [1]$$

where the $\rho_\alpha^i$ are arbitrary states of the subsystems ($i = 1, 2$), which depend on a "hidden variable" $\alpha$, drawn by a classical random generator with probabilities $p_\alpha$. Such states are now called separable, which is a bit awkward, since the notion is typically applied to systems which are widely separated. However, the term is so firmly established that it is hopeless to try to improve on it.

In any case, it was shown by Werner (1989) that there are nonseparable states, which nevertheless satisfy Bell's inequalities and all its generalizations. The next step was the observation by Popescu (1994) that entanglement could be distilled: this is a process by which some number of moderately entangled pair states is converted to a smaller number of highly entangled states, using only local quantum operations, and classical communication between the parties. For some time it seemed that this might close the gap, that is, that the failure of separability might be equivalent to "distillability" (i.e., the existence of a distillation procedure producing arbitrarily highly entangled states from many copies of the given one). However, this turned out to be false, as shown by the Horodecki family in 1998 (Horodecki *et al.* 1998), by explicitly exhibiting bound entangled, that is, nonseparable, but also not distillable states. In 2003 Oppenheim and the Horodeckis introduced a further distinction, namely whether it is possible to extract a secret key from copies of a given quantum state by local quantum operations and public classical communication (Horodecki *et al.* 2005). This task had hitherto been viewed as an application of entanglement distillation, but it turned out that secret key can be distilled from some bound entangled (but never from separable) states.

For the entanglement theory of multipartite states, that is, states on systems composed of three or more parts, between which no quantum interaction takes place, one key observation is that new entanglement properties must be expected with any increase of the number of parties. As shown by Bennett *et al.* (1999), there are states of three parties which cannot be written in the three-party analog of [1], but are nevertheless separable for all three splits of the system into one vs. two subsystems.

The crucial advance of entanglement theory, however, lies not so much in the distinctions outlined above, but in the quantitative turn of the theory. With the discovery of the teleportation and dense coding processes (Bennett and Wiesner 1992, Bennett *et al.* 1993), entanglement changed its role from a property of counterintuitive contortedness to a resource, which is used up in teleportation and similar processes. Distillation is then seen as a method to upgrade a given source to a new source of highly entangled states suitable for this purpose, and it is not just the possibility of doing this, but the rate of this conversion, which becomes the focus of the investigation. All the tasks in which entanglement appears suggest quantitative measures of entanglement. In addition, there are many entanglement measures, which appear natural from a mathematical point of view, or are introduced simply because they can be estimated relatively easily and in turn give bounds on other entanglement measures of interest. The current situation is that there is no shortage of entanglement measures in the literature, but it is not yet clear which ones will be of interest in the long run. Some of these measures are described in Entanglement Measures.

The current state of entanglement theory is marked firstly by some long-standing open problems in the basic bipartite theory on the one hand (additivity of the entanglement of formation, the existence of NPT bound entangled states, and more recently the existence of entangled states with vanishing key rate). Secondly, there is significant effort to try to compute some of the entanglement measures, at least for simple subclasses of states. This is so difficult, because many definitions involve an optimization over operations on an asymptotically large system. Thirdly, there is a new trend in multipartite entanglement theory, namely looking specifically at entanglement in lattice structures such as spin systems of harmonic-oscillator lattices. Here one can expect very fruitful interaction with the statistical mechanics and solid-state physics in the near future.

## Qualitative Entanglement Theory

### Setup

Throughout this section, we will consider density operators on a Hilbert space split in some fixed way into a tensor product of a Hilbert space $\mathcal{H}_A$ for

Alice's system and a Hilbert space $\mathcal{H}_B$ for Bob's system, that is, $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. For simplicity, we will mostly consider finite-dimensional spaces, and if a dimension parameter $d < \infty$ appears, it is understood that $d = \dim \mathcal{H}_A = \dim \mathcal{H}_B$. By $\mathcal{B}(\mathcal{H})$ we will denote the set of bounded operators on a Hilbert space, and by $\mathcal{B}_*(\mathcal{H})$ the set of trace-class operators. We distinguish these even in the finite-dimensional case, because of their different norms. By $\mathfrak{S}$ we will denote the state space of the combined system, that is, the set of positive elements of $\mathcal{B}_*(\mathcal{H})$ with trace 1.

For such a density operator $\rho = \rho^{AB}$ we denote by $\rho^A$ and $\rho^B$ the restrictions to the subsystems, defined by the partial trace over the other system, or by $\mathrm{tr}(\rho^A F) = \mathrm{tr}(\rho^{AB}(F \otimes \mathbb{1}))$. We denote by $\Theta$ the operation of matrix transposition, and by $\mathrm{id} \otimes \Theta$ the partial transposition, applied only to the second tensor factor. Since transposition is not completely positive (*see* Channels in Quantum Information Theory) partial transposition may take positive operators to non-positive operators. The relative entropy (*see* Entropy and Quantitative Transversality) of two density operators $\rho, \sigma$ will be used with the convention $S(\rho||\sigma) = \mathrm{tr}\,\rho(\log \rho - \log \sigma)$.

### Witnesses and the Criterion of Positivity of Partial Transpose

A state $\rho$ is called separable iff it is of the form [1], and entangled otherwise. The set of separable states $\mathfrak{C}$ is a convex subset of the set $\mathfrak{S}$ of all states. Its extreme points are obvious from the representation [1], namely the pure product states $\rho = |\phi_A \otimes \phi_B\rangle\langle\phi_A \otimes \phi_B|$. Since $\mathfrak{C}$, like $\mathfrak{S}$, is a convex set in $(d^4 - 1)$ dimensions, Caratheodory's theorem asserts that the sum can be taken to be a decomposition into $d^4$ such terms. For a given $\rho$, deciding whether it is separable or entangled, hence, involves a nonlinear search problem in roughly $4d^5$ real parameters, namely the vector components of the $\phi_A, \phi_B$ appearing in the sum.

Dually, the convex set $\mathfrak{C}$ can be described by a set of linear inequalities. Here is a simple way of generating such inequalities: let $T : \mathcal{B}_*(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$ be a positive linear map, that is, a map taking positive matrices to positive matrices. Then for $\rho_A, \rho_B \geq 0$ the expression $\mathrm{tr}(\rho_A T(\rho_B))$ is positive. It is also bilinear, so we can find a Hermitian operator $T^\natural \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ such that

$$\mathrm{tr}(\rho_A T(\rho_B)) = \mathrm{tr}((\rho_A \otimes \rho_B) T^\natural)$$

Since the left-hand side is positive, we see by taking convex combinations that $\mathrm{tr}(\rho T^\natural) \geq 0$ for all separable states $\rho$. Hence, if we find a state with a negative expectation of $T^\natural$, we can be sure it is entangled. Therefore, such operators $T^\natural$ are called entanglement witnesses. This is often a useful criterion, especially when one has some additional information about the state, allowing for an intelligent choice of witness. It is known from the theory of ordered vector spaces and their tensor products that the set of witnesses constructed above is complete. Hence, in principle, checking all such witnesses provides a necessary and sufficient criterion for entanglement. However, in practice this remains a difficult task, because the extreme points of the set of positive maps are only known for some low dimensions.

By restricting $T$ to completely positive maps, we get a useful necessary criterion. It can be seen that it is equivalent to

$$(\mathrm{id} \otimes \Theta)(\rho) \geq 0$$

that is, to the positivity of the partial transpose (PPT). States with this property are called "PPT states" in current jargon.

### Pure States, Purification

For pure states, that is, for the extreme points of $\mathfrak{S}$, separability is trivial to decide: since for pure states the sum [1] can only be a single term, a pure state is separable iff it factorizes.

A useful observation is that, for pure states $\rho = |\Phi\rangle\langle\Phi|$, all information about entanglement is contained in the spectrum of the reduced states. Consider a vector $\Phi \in \mathcal{H}_A \otimes \mathcal{H}_B$ of the form

$$\Phi = \sum_\alpha \sqrt{r_\alpha}\, \phi_\alpha^A \otimes \phi_\alpha^B \qquad [2]$$

where $\phi_\alpha^A \in \mathcal{H}_A$ and $\phi_\alpha^B \in \mathcal{H}_B$ are orthonormal systems, $r_\alpha > 0$, and $\sum_\alpha r_\alpha = 1$. Then it is easy to check that $\rho^A = \sum_\alpha r_\alpha |\phi_\alpha^A\rangle\langle\phi_\alpha^A|$ is the spectral resolution of the restriction. Conversely, by diagonalizing the restriction of a general unit vector $\Phi$, we find a biorthogonal decomposition of the from [2], also known as the Schmidt decomposition. The Schmidt spectrum $\{r_1, \ldots, r_d\}$ hence classifies vectors up to local basis changes in $\mathcal{H}_A$ and $\mathcal{H}_B$.

Since any $\rho^A$ can appear in this construction, we see that any mixed state can be considered as the restriction of a pure state, which is essentially unique, namely up to the choice of basis in the purifying system B, and up to perhaps adding or deleting some irrelevant dimensions in $\mathcal{H}_B$. The resulting vector $\Phi$ is known as the purification of $\rho_A$.

The extreme cases of [2] are pure product states on the one hand, and vectors, for which $\rho^A = \mathbb{1}/d$ is the totally chaotic state. These are known as maximally entangled and embody, in the most extreme way, the observation that in quantum

mechanics, as opposed to classical probability, the restriction of a pure state may be mixed.

Let us fix a maximally entangled vector $\Omega$, and the matching Schmidt bases, so that

$$\Omega = \frac{1}{\sqrt{d}} \sum_k |kk\rangle \qquad [3]$$

where we have used the simplified ket notation, in which only the basis label is written. Then, an arbitrary vector can be written as $\Phi = (X \otimes \mathbb{1})\Omega = (\mathbb{1} \otimes X^{\mathrm{T}})\Omega$, where $X^{\mathrm{T}}$ denotes the matrix transpose of $X$. Clearly, this vector is again maximally entangled iff $X$ is unitary. Hence, the set of maximally entangled vectors is a single orbit under unilateral unitary transformations, and we even have the choice to which side we apply the unitaries.

### Teleportation

Suppose we have an orthonormal basis of maximally entangled vectors $\Phi_\alpha \in \mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$. By the remarks above, this is equivalent to choosing unitaries $U_\alpha, \alpha = 1, \ldots, d^2$ such that $\Phi_\alpha = (U_\alpha \otimes \mathbb{1})\,\Omega$, and $\mathrm{tr}(U_\alpha^* U_\beta) = d\delta_{\alpha\beta}$. For example, a finite Weyl system constitutes such a system of unitaries, which shows that we can find realizations in any dimension $d$.

Suppose that Alice and Bob each own part of a system prepared in the state $\Omega$ then they can transmit perfectly the state of a $d$-dimensional system, using only classical communication. Classical communication by itself would never suffice to transmit quantum information, and the entangled resource $\Omega$ by itself does not allow the transmission of any signal. But the combination of these resources does the trick: Alice measures the observable associated with the basis $\Phi_\alpha$ on the combined system formed by the unknown input and her part of the entangled pair. The result $\alpha$ is then transmitted to Bob, who performs a $U_\alpha$-rotation on his part of the entangled pair, producing the output state of the teleportation. One can show by direct calculation that this is exactly equal to the input state.

Note that the resource $\Omega$ is destroyed in this process, so that for every transmission we need a fresh entangled pair. Less than maximally entangled states instead of $\Omega$ lead to less-than-perfect transmission, which can be extended to quantitative relations between entanglement and channel capacity.

## Special Systems

### Qubits

For qubit pairs, there is a special basis of maximally entangled vectors, which has some amazing properties. It consists of the vectors $\Phi_0 = \Omega$, and $\Phi_k = \mathrm{i}(\sigma_k \otimes \mathbb{1})\Omega$, where $\sigma_k, k = 1, 2, 3$, denotes the Pauli matrices. Then a vector is maximally entangled iff its components are real in this basis, up to a common phase. A unitary matrix of determinant 1 factorizes into $U_1 \otimes U_2$ iff its matrix elements are real, up to a common phase.

For qubit pairs, and also for dimensions $2 \otimes 3$, the partial transposition criterion for entanglement is necessary and sufficient, as shown by Woronowicz and the Horodecki family.

### Orthogonally Invariant States

A state $\rho$ on $\mathbb{C}^d \otimes \mathbb{C}^d$ is called orthogonally invariant if, for any orthogonal matrix $U$ (with respect to some fixed product basis) $[\rho, U \otimes U] = 0$. This leaves a three-dimensional space of operators, spanned by the identity, the permutation $F = \sum_{i,j} |ij\rangle\langle ji|$, and its partial transpose $\widehat{F} = \sum_{i,j} |ii\rangle\langle jj|$, which is $d$ times the projection onto the maximally entangled vector $\Omega$. Figure 1 shows the plane of Hermitian operators $\rho$ with the described symmetry and $\mathrm{tr}\,\rho = 1$. Convenient coordinates are $\mathrm{tr}\,\rho F$ and $\mathrm{tr}\,\rho \widehat{F}$. Note that these are defined for any density operator, and are also invariant under the "twirl" operation $\rho \mapsto \int \mathrm{d}U (U \otimes U)\rho(U \otimes U)^*$, using the Haar measure $\mathrm{d}U$, which projects onto the orthogonally invariant states. Hence, the diagram provides a section as well as a projection of the state space. The intersection of the positive operators with those having positive partial transpose is the set of PPT states, which in this case coincides with the separable states. The thin lines correspond to states of higher symmetry, namely on the one hand the "isotropic states" commuting with $U \otimes \bar{U}$, with $\bar{U}$ the complex conjugate of $U$, and the "Werner states" commuting with all unitaries $U \otimes U$. Their intersection point is the normalized trace.

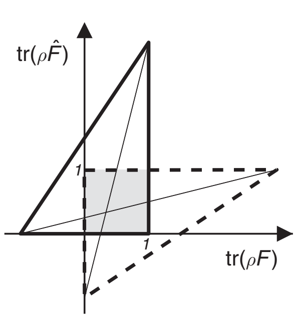



**Figure 1** The plane of orthogonally invariant unit trace Hermitian operators of a $3 \otimes 3$-system. The upright triangle gives the positive operators, and the dashed one those with positive partial transpose. The shaded area gives the PPT states.

### Gaussians

In general, the entanglement in systems with infinite-dimensional Hilbert spaces is more difficult to analyze. However, if the system is characterized by variables satisfying canonical commutation relations, like positions and momenta, or the components of the free quantum electromagnetic field, there is a special class of states, which is again characterized by low-dimensional matrices. This allows the discussion of entanglement questions, in a way largely parallel to the finite-dimensional theory.

Let $R_1, \ldots, R_{2f}$ denote the canonical operators, where $f$ is the number of degrees of freedom. The commutation relations can be summarized as $\mathrm{i}[R_\mu, R_\nu] = \sigma_{\mu\nu}\mathbb{1}$, where $\sigma$ is the symplectic matrix. Operators $R_\mu$ have a common set of analytic vectors, and generate the unitary Weyl operators $W(a) = \exp(\mathrm{i}a^\mu R_\mu)$, which describe the phase space displacements. Gaussian states are those making $a \mapsto \mathrm{tr}\,\rho W(\rho)$ a Gaussian function or, equivalently, those with Gaussian Wigner function. Up to a gloal displacement, they are completely characterized by the covariance matrix

$$\gamma_{\mu\nu} = \mathrm{tr}\,\rho(R_\mu R_\nu + R_\nu R_\mu) \qquad [4]$$

The only constraint for a real symmetric matrix to be a covariance matrix of a quantum state is that $\gamma + \mathrm{i}\sigma$ is a positive semidefinite matrix, which is a version of the uncertainty relations.

Now for entanglement theory, we take some of the degrees of freedom as Alice's and some as Bob's. Separability can be characterized in terms of $\gamma$, namely by the condition that $\gamma \geq \gamma'$, where $\gamma'$ is the covariance matrix of a Gaussian product state. Similarly, partial transposition can be implemented as an operation on covariance matrices, which allows a simple verification of the PPT condition. It turns out that as long as one partner has only a single degree of freedom, the PPT condition is necessary and sufficient for separability, but this fails for larger systems.

The pure Gaussian states allow a normal form with respect to local symplectic transformations analogous to the Schmidt decomposition. For the minimal case of one degree of freedom on either side, one obtains a one-parameter family of "two mode squeezed states." Its limit for infinite squeezing parameter is the state used by EPR (Einstein *et al.* 1935), which, however, makes rigorous mathematical sense only as a singular state, that is, a linear functional on $\mathcal{B}(\mathcal{H})$, which can no longer be represented as the trace with a density operator.

### Multipartite Stars

A key feature of entanglement in a multipartite system is usually referred to as "monogamy": when Alice shares a highly entangled state with Bob, her system cannot also be highly entangled with Bill. More formally, suppose that a multipartite state for systems $\mathrm{A}, \mathrm{B}_1, \ldots, \mathrm{B}_n$ is given, such that the restriction to each pair $\mathrm{AB}_k$ is the same bipartite state $\rho$. Then as $n$ becomes larger, the existence of such a star-shaped extension constrains $\rho$ to become less and less entangled. In fact, as $n \to \infty$, this condition is equivalent to the separability of $\rho$.

## Open Problems

Recall from the introduction the following chain of inclusions:

$$\begin{aligned} \text{separable states} &\subset \text{states with vanishing key rate} \\ &\subset \text{PPT state} \\ &\subset \text{undistillable states} \\ &\subset \text{all states} \end{aligned}$$

The second and fourth inclusions are strict, but for the first and third one might have equality, for all we know. Especially for the third inclusion, this is a long-standing problem.

Finally, we would like to point out that qualitative and conceptual aspects of entanglement are surveyed by Bub (2001), Popescu and Rohrlich (1998), and Horodecki *et al.* (2001). For quantitative aspects *see* Entanglement Measures.

*See also:* Capacities Enhanced by Entanglement; Capacity for Quantum Information; Channels in Quantum Information Theory; Entanglement Measures; Entropy and Quantitative Transversality.

## Further Reading

Bell JS (1964) *Physics* 1: 195.

Bennett CH and Wiesner S (1992) Communication via one- and two-particle operators on Einstein–Podolsky–Rosen states. *Physical Review Letters* 69: 28812884.

Bennett CH, Brassard G, Crépeau C *et al.* (1993) Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels. *Physical Review Letters* 70: 1895–1899.

Bennett CH, *et al.* (1999) Unextendible product bases and bound entanglement. *Physical Review Letters* 82: 5385.

Bub J (2001) *Quantum Entanglement and Information*, The Stanford Encyclopedia of Philosophy, URL = http://plato.stanford.edu/entries/qt-entangle/.

Einstein A, Podolsky B, and Rosen N (1935) Can quantum-mechanical description of physical reality be considered complete? *Physical Review* 47: 777–780.

Horodecki K, Horodecki M, Horodecki P, and Oppenheim J (2005) Secure key from bound entanglement. *Physical Review Letters* 94: 160502.

Horodecki M, Horodecki P, and Horodecki R (1998) Mixed-state entanglement and distillation: is there a "bound" entanglement in nature? *Physical Review Letters* 80: 5239–5242.

Horodecki M, Horodecki P, and Horodecki R (2001) Mixed-state entanglement and quantum communication. In: Alber G *et al.* (eds.) *Quantum Information*, Springer Tracts in Modern Physics, vol. 173.

Popescu S (1994) Teleportation versus Bell's inequalities. What is nonlocality? *Physical Review Letters*, 72: 797.

Popescu S and Rohrlich D (1998) The joy of entanglement. In: Lo H-K, Popescu S, and Spiller T (eds.) *Introduction to Quantum Computation and Information*. Singapore: World Scientific.

Primas H (1983) Verschränkte Systeme und Quantenmechanik. In: Kanitscheider B (ed.) *Moderne Naturphilosophie*. Würzburg: Königshausen+Neumann.

Schrödinger E (1935) Die gegenwärtige Situation in der Quantenmechanik. *Naturwiss* 23: 807–812, 823–828, 844–849.

See the server for open problems at URL = http://www.imaph.tu-bs.de/qi/problems.

Werner R (1983) Bell's inequalities and the reduction of statistical theories. In: Balzer W, Pearce DA, and Schmidt H-J (eds.) *Reduction in Science*, vol. 175, Synthese Library, (Reidel 1984).

Werner RF (1989) Quantum states with Einstein–Rosen–Podolsky correlations admitting a hidden-variable model. *Physical Review A* 40: 4277–4281.

# Entanglement Measures

**R F Werner**, Institut für Mathematische Physik, Braunschweig, Germany



## Introduction

Entanglement, or quantum correlation, is one of the central concepts in quantum information theory. Its theory can be roughly separated into three parts. The first is *qualitative*, that is, it addresses the question "Is this state entangled or not?" The second, *comparative* part asks "Is this state more entangled than that state?," and finally the *quantitative* theory asks "How entangled is this state?," and gives its answers in the form of entanglement measures assigning a number to every state. Quantitative questions come up naturally whenever entanglement is used as a resource for tasks of quantum information processing. For example, entangled states are in a way the fuel for the processes of teleportation and dense coding: in each transmission step a maximally entangled pair system is required, and cannot be used for a further transmission. The process also works with less than maximally entangled states, but then it also becomes less efficient. Since entangled states created in the laboratory typically have imperfections, it becomes important to understand the rates at which imperfectly entangled states may be distilled to maximally entangled ones, and this rate is a direct measure of the usefulness of the given state for many purposes. The quantitative, task related turn is a new development in the study of the foundations of quantum mechanics. It has been imported from classical information theory, where this way of thinking has been standard for a long time. The combination makes the particular flavor of quantum information theory.

In this article we consider the comparative and quantitative aspects of entanglement. The historical aspects and qualitative theory are treated in a separate article (*see* Entanglement), to which we refer for basic notions and notations. The example of teleportation suggests close links between quantitative entanglement theory and the theory of capacity Bennett *et al.* (1996), which is the transfer rate of quantum information through a given channel. These connections are described in Quantum Channels: Classical Capacity.

We follow the notations of the basic article on entanglement (*see* Entanglement). In particular, $\Theta$ denotes the transpose operation, and $(\mathrm{id}\otimes\Theta)$ the partial transpose. A state is called "PPT" if its partial transpose is positive. The two physicists operating the laboratories in which the two parts of a bipartite system are kept are called Alice and Bob, as usual. The restriction of a state $\rho$ to Alice's subsystem is denoted by $\rho^{\mathrm{A}}$.

## Comparative Entanglement and Protocols

### Protocols

In this section we introduce relations of the kind "state $\rho_1$ is more entangled than $\rho_2$." We take this to mean that $\rho_2$ can be obtained by applying to $\rho_1$ some operations which "cannot create entanglement." The definition of a class of operations of which this can be claimed then defines the comparison. It turns out that there are different choices for the class of such operations, depending on the resources available for the transformation steps. The class of operations is usually referred to as a *protocol*.

Certainly local operations performed separately by Alice and Bob cannot increase entanglement. Alice and Bob might have to make some choices, and even if they make these according to a

prearranged scheme, by using a shared table of random numbers, entanglement will not be generated. In this restrictive protocol, which we abbreviate by LO, for "local operations," no communication is allowed. It is clear that by just discarding the initial state, and preparing a new one, based on the random instruction allows Alice and Bob to make any separable state, so these states come out as the "least entangled" ones for this and any richer protocol.

Next we might allow classical communication from Alice to Bob. That is, Bob's decision to perform some operation in his laboratory is allowed to depend on measuring results obtained by Alice in an earlier stage. Of course, Alice is not allowed to send quantum systems, since in this case she might just send a particle entangled to one of her own, and *any* state could be generated. This protocol is referred to as "local operations and one-way classical communication" (LOWC). Obviously, we might also allow Bob to talk back, arriving at "local operations and classical communication" (LOCC). This is the protocol underlying most of the work in entanglement theory.

The drawback of the LOCC protocol is that its operations are extremely difficult to characterize: an LOCC operation can take many rounds, and there is no way to simplify a general operation to some kind of standard form. This is the main reason why other protocols have been considered. For example, it is obvious that an LOCC operation can be written as a sum of tensor products of local operations, in a form reminiscent of the definition of separability. However, such "separable superoperators" may fall outside LOCC. Another property easily checked for all LOCC operations is that PPT states go into PPT states. The protocol "PPT-preserving operations" (PPTP) can also be characterized as the set of channels $T$ for which $(\mathrm{id}\otimes\Theta)T(\mathrm{id}\otimes\Theta)$ is positive (although not necessarily completely positive). This condition is relatively easy to handle mathematically, so that the best way to show that some $\rho_1$ cannot be converted to $\rho_2$ by LOCC is often to show that this transition is impossible under PPTP. The drawback of the PPTP protocol is that it may create some entanglement after all, namely arbitrary PPT states. So it properly belongs to a modified entanglement theory in which separability is replaced by the PPT condition.

### Converting Pure States and Majorization

The entanglement ordering is exactly known for pure states due to a famous theorem by Nielsen (1999): a pure state $\rho_1$ is *more entangled* than a pure state $\rho_2$ under the LOCC protocol iff the restriction $\rho_1^{\mathrm{A}}$ is *more mixed* than the restriction $\rho_2^{\mathrm{A}}$ in the sense of majorization of spectra (i.e., for every $k$ the sum of the $k$ largest eigenvalues of $\rho_1^{\mathrm{A}}$ is less than the corresponding sum for $\rho_2^{\mathrm{A}}$). Equivalently, there is a doubly stochastic channel (completely positive linear map preserving both the identity and the trace functional) taking $\rho_2^{\mathrm{A}}$ to $\rho_1^{\mathrm{A}}$.

An interesting aspect of this theory is the phenomenon of *catalysis*: It may happen that although $\rho_1$ cannot be converted by LOCC to $\rho_2, \rho_1\otimes\sigma$ can be converted to $\rho_2\otimes\sigma$. The "catalyst" $\sigma$ is a resource borrowed at the beginning of the transformation, and is returned unchanged afterwards. The order relation allowing such catalysts is yet to be fully characterized.

### Asymptotic Conversion

In many applications we are not interested in exact conversion of one state to another, but are quite satisfied if the transformation can be done with a small controlled error. In particular, when we ask for the *achievable conversion rate* between many copies of the states involved, we allow small errors, but require the errors to go to zero. Given any protocol, and states $\rho_1, \rho_2$, we say that $\rho_1$ can be converted to $\rho_2$ with rate $r$ if, for all sufficiently large $n$, there is a channel of the protocol, which takes $n$ copies of $\rho_1$, that is, the state $\rho_1^{\otimes n}$, to a state $\rho'$ which approximates roughly $m\approx rn$ copies of $\rho_2$, in the sense that $m\geq rn$, and the trace norm $\|\rho'-\rho_2^{\otimes m}\|$ goes to zero.

Of course, one is usually interested in the supremum of the achievable conversion rates, which we call simply the maximal conversion rate. In particular, when $\rho_2$ is the maximally entangled pure state of a qubit pair (usually called the "singlet"), the maximal rate is called the *distillable entanglement* $E_{\mathrm{D}}(\rho_1)$. In the other direction, when $\rho_1$ is the singlet, we call the inverse of the maximal conversion rate the *entanglement cost* $E_{\mathrm{C}}(\rho_2)$. These are two of the key entanglement measures to be discussed below.

In general, $E_{\mathrm{D}}(\rho)<E_{\mathrm{C}}(\rho)$, so the asymptotic conversion between different states is usually not reversible. However, this is the case for pure states, and one finds

$$E_{\mathrm{D}}(\rho)=E_{\mathrm{C}}(\rho)=S(\rho^{\mathrm{A}}) \qquad [1]$$

where $S(\rho)=-\mathrm{tr}\,\rho\log_2(\rho)$ denotes the von Neumann entropy (*see* Entropy and Quantitative Transversality) based on the binary logarithm.

Since one can do the conversion between different pure states via singlets, it is clear that the maximal conversion rate from a pure state $\rho_1$ to a pure state $\rho_2$

equals $S(\rho_1^A)/S(\rho_2^A)$. Hence, in contrast to the ordering given by Nielsen's theorem, all pure states are interconvertible, and the ordering is described by a single number. For this simplification, the allowance of small errors is crucial. Without asymptotically small but nonzero errors, it would also be impossible to obtain singlets from any generic mixed state.

## Entanglement Measures

### Properties of Interest

We now consider more systematically functions $E: \mathfrak{S} \to \mathbb{R}$ defined on the states spaces of arbitrary bipartite quantum systems. When can we regard this as a measure of entanglement? The minimal requirements are that $E(\rho) \geq 0$ for all $\rho$, and $E(\rho) = 0$ for separable states. Since the choice of local bases should be irrelevant, we will require $E((U_A \otimes U_B)\rho(U_A \otimes U_B)^*) = E(\rho)$ for unitaries $U_A, U_B$. We also *normalize* all entanglement measures so that $E(\sigma) = 1$, when $\sigma$ is the maximally entangled state of a pair of qubits. Beyond that, consider the following:

1. $\boldsymbol{V}$ (*Convexity* $E(\sum_\alpha p_\alpha \rho_\alpha) \leq \sum_\alpha p_\alpha E(\rho_\alpha)$) Starting from any $E$, possibly defined only on a subset containing the pure states, we can enforce this property by taking the *convex hull* (or "roof") co$E$, defined as the largest convex function, which is $\leq E$ wherever it is defined.
2. $\boldsymbol{M}$ (*Monotonicity*) Suppose that some LOCC protocol applied to $\rho$ returns some classical parameter $\alpha$ with probability $p_\alpha$, and in that case a bipartite state $\rho_\alpha$. Then $\sum_\alpha p_\alpha E(\rho_\alpha) \leq E(\rho)$.
3. $\boldsymbol{A}^-$ (*Subadditivity* $E(\rho_1 \otimes \rho_2) \leq E(\rho_1) + E(\rho_2)$) In this and the following, the tensor products of bipartite states are to be reordered from $A_1B_1A_2B_2$ to $(A_1A_2)(B_1B_2)$, so the separation into Alice's and Bob's subsystems is respected.
4. $\boldsymbol{A}^+$ (*Superadditivity* $E(\rho_1 \otimes \rho_2) \geq E(\rho_1) + E(\rho_2)$)
5. $\boldsymbol{A}^{++}$ (*Strong superadditivity* $E(\rho_{12}) \geq E(\rho_1) + E(\rho_2)$) Here $\rho_i$ denotes the restriction of a general state $\rho_{12}$ to the $i$th subsystem.
6. $\boldsymbol{A}^\infty$ (*Weak additivity* $E(\rho^{\otimes n}) = nE(\rho)$) This can be enforced by *regularization*, going from $E$ to

$$E^\infty(\rho) = \lim_n \frac{1}{n} E(\rho^{\otimes n})$$

   Note that this is implied by *additivity*, which is the conjunction of $\boldsymbol{A}^+$ and $\boldsymbol{A}^-$.
7. $\boldsymbol{C}$ (*Continuity*) Here it is crucial to postulate the right kind of dimensional dependence. A good choice is to demand that $|E(\rho_1) - E(\rho_2)| \leq \log d f(\|\rho_1 - \rho_2\|)$, where $f$ is some function with $\lim_{t\to 0} f(t) = 0$.
8. $\boldsymbol{L}$ (*Lockability*) A property related to, but not equal to, discontinuity: a measure is called lockable, if the loss (i.e., the tracing out) of a single qubit by Alice or Bob can make $E(\rho)$ drop by an arbitrarily large amount.

### The Collection of Entanglement Measures

The following are the main entanglement measures discussed in the literature. Note that all measures defined by conversion rates in principle depend on the protocol used. Unless otherwise stated, we will only consider LOCC. For every function we list in brackets the properties which are known.

1. $E_F$ (*Entanglement of formation* $[\boldsymbol{V}, \boldsymbol{M}, \boldsymbol{A}^-, \boldsymbol{C}, \boldsymbol{L}]$) This is defined as the convex hull of the entanglement of pure states given by eqn [1]. For qubit pairs, there is a closed formula due to Wootters (1998), orthogonally invariant states (Vollbrecht and Werner 2001) (see 00510), and permutation symmetric 2-mode Gaussians. One of the big open questions is whether $E_F$ is additive. This is equivalent to $E_F$ satisfying $\boldsymbol{A}^{++}$, and also to the additivity of Holevo's $\chi$-capacity of quantum channels (see Quantum Channels: Classical Capacity).
2. $E_C$ (*Entanglement cost* $[\boldsymbol{V}, \boldsymbol{M}, \boldsymbol{A}^-, \boldsymbol{A}^\infty, \boldsymbol{C}, \boldsymbol{L}]$) This was already defined in the section "Asymptotic conversion." It has been shown to be equal to the regularization of $E_F$, that is, $E_C = E_F^\infty$. If $E_F$ would turn out to be additive, we would thus have $E_C = E_F$.
3. $E_D$ (*Distillable entanglement* $[\boldsymbol{M}, \boldsymbol{A}^{++}, \boldsymbol{A}^\infty]$) Again, see the section "Asymptotic conversion." This is one of the important measures from the practical point of view, but notoriously difficult to compute explicitly. Convexity of $E_D$ is an open problem related to the existence of bound entangled, but not PPT states.
4. $E_\rightarrow$ (*One-way distillable entanglement* $[\boldsymbol{M}, \boldsymbol{A}^{++}, \boldsymbol{A}^\infty]$) Same as $E_D$, but restricting to the LOWC protocol. Obviously, $E_\rightarrow(\rho) \leq E_D(\rho)$. There are examples of proper inequality "$<$" (Bennett *et al.* 1996). $E_\rightarrow$ is more directly linked to quantum capacity than $E_D$, which in turn corresponds to the quantum capacity, allowing classical backwards communication as a resource.
5. $E_N$ (*Logarithmic negativity* $[\boldsymbol{M}, \boldsymbol{A}^-, \boldsymbol{A}^+, \boldsymbol{L}]$) This is a quantitative companion of the PPT criterion: one sets $E_N(\rho) = \log_2 \|(\mathrm{id} \otimes \Theta)(\rho)\|$, where the norm is the trace norm. For PPT states, $\rho$ this is equal to the trace, and $E_N(\rho) = 0$. If the partial transpose has negative eigenvalues, the sum of their absolute values is $>1$, and $E_N(\rho) > 0$. $E_N$ is an easily computed upper bound to $E_D$, but gives the wrong value for nonmaximally entangled pure states.

6. $E_R$ (*Relative entropy of entanglement* [$V, M, A^-$]) This measure (Vedral *et al.* 1997) is motivated geometrically: it is simply the relative entropy distance of $\rho$ to the separable subset: $E_R(\rho) = \inf_\sigma S(\rho\|\sigma)$, where $\sigma$ ranges over all separable states. $E_R$ is an upper bound to $E_D$. However, it can be improved by taking the distance to the PPT states rather than the separable states, and by combining with $E_N$, in the following way:
7. $E_B$ (*The Rains bound* [$V, M, A^-, C$]) Following Rains (2001), we set

$$E_B(\rho) = \inf_\sigma (S(\rho\|\sigma) + E_N(\sigma))$$

   where the infimum is over all states $\sigma$. This is still an upper bound to $E_D$, although clearly smaller than both $E_R$ (take only separable $\sigma$) and $E_N$ (take $\sigma = \rho$). No example of $E_D(\rho) < E_R(\rho)$ is known, but any bound entangled non-PPT state would be such an example.
8. $E_S$ (*Squashed entanglement* [$V, M, A^-, A^{++}, C, L$]) This measure, introduced by Christandl and Winter (2004), amazingly has all the good properties, but is as difficult to compute as any of the other measures. $E_S(\rho^{AB})$ is the infimum over the entropy combination

$$S(\rho^{AC}) + S(\rho^{BC}) - S(\rho^{ABC}) - S(\rho^{C})$$

   over all extensions $\rho^{ABC}$ of the given state $\rho^{AB}$ to a system enlarged by a part C, where the density operators in the above expression are the restrictions of $\rho^{ABC}$ to the subsystems indicated.
9. $E_K$ (*Key rate* [$V, M, A^-$]) The bit rate at which secret key can be generated is certainly larger than $E_D$, since distillation is one way to do it. It is, in general, strictly larger, since there are undistillable states with positive key rate.
10. $E_C$ (*Concurrence* [$V$]) This measure was originally only defined for qubit pairs, as a step in Wootter's (1998) formula for $E_F$ in this case. It has an extension to arbitrary dimensions (Rungta *et al.* 2001), namely the convex hull of the function $c(|\psi\rangle\langle\psi|) = \sqrt{2(1 - \mathrm{tr}(\rho^2))}$, where $\rho = |\psi\rangle\langle\psi|^A$ is the reduced density operator. Both upper and lower bounds exist in the literature. The main interest in this measure stems from the fact that it has interesting extensions to the multipartite case.

To conclude, we would like to point out that many of the themes discussed in this article were set by Bennett *et al.* (1996); their article is worth reading even today. Good review articles covering entanglement measures, with more complete references, are Plenio and Virmani (2005), Bruß (2002), and Donald *et al.* (2002).

*See also:* Entanglement; Entropy and Quantitative Transversality; Quantum Channels: Classical Capacity.

## Further Reading


Bennett CH, DiVincenzo DP, Smolin JA, and Wootters WK (1996) Mixed-state entanglement and quantum error correction. *Physical Review A* 54: 3824–3851.

Bruß D (2002) Characterizing entanglement. *Journal of Mathematical Physics* 43: 4237.

Christandl M and Winter A (2004) Squashed entanglement – an additive entanglement measure. *Journal of Mathematical Physics* 45: 829–840.

Donald MJ, Horodecki M, and Rudolph O (2002) The uniqueness theorem for entanglement measures. *Journal of Mathematical Physics* 43: 4252–4272 and quant-ph/0105017.

Horodecki M, Horodecki P, and Horodecki R (2000) Unified approach to quantum capacities: towards quantum noisy coding theorem. *Physical Review Letters* 85: 433.

Nielsen M (1999) Conditions for a class of entanglement transformations. *Physical Review Letters* 83: 436–439.

Plenio MB and Virmani S (2005) An introduction to entanglement measures, quant-ph/0504163.

Rains E (2001) *IEEE Trans. Inf. Theory* 47: 2921–2933.

Rungta P *et al.* (2001) Universal state inversion and concurrence in arbitrary dimensions. *Physical Review A* 64: 042315.

Vedral V, Plenio MB, Rippin MA, and Knight PL (1997) Quantifying entanglement. *Physical Review Letters* 78: 2275.

Vollbrecht KGH and Werner RF (2001) Entanglement measures under symmetry. *Physical Review A* 64: 062307.

Wootters WK (1998) Entanglement of formation of an arbitrary state of two qubits. *Physical Review Letters* 80: 2245–2248.

# Entropy and Quantitative Transversality

**G Comte**, Université de Nice Sophia Antipolis, Nice, France



## Introduction

A mathematical law for a physical phenomenon, describing the variation of a value $y(\in \mathbb{R})$ in terms of parameters $x_i \in \mathbb{R}$, $i \in \{1, \ldots, n\}$, is usually given:

1. in the simplest cases (and hence in exceptional cases), by an explicit functional equation $y = F(x_1, \ldots, x_n)$, or
2. by an implicit equation $G(y, x_1, \ldots, x_n) = 0$, or
3. more generally, by a partial differentiable equation,

$$H\left(y, \frac{\partial^{|\alpha_1|}y}{\partial x_{i_1} \cdots \partial x_{i_{\alpha_1}}}, \ldots, \frac{\partial^{|\alpha_k|}y}{\partial x_{j_1} \cdots \partial x_{j_{\alpha_k}}}, x_1, \ldots, x_n\right) = 0 + \text{initial values}$$

In the first case, the exact equation $y = F(x_1, \ldots, x_n)$ fully describes the behavior of $y$ as $(x_1, \ldots, x_n)$ vary, but in practice this information is too substantive: using the Taylor formula, knowledge of the value $y^0$ at some point $(x_1^0, \ldots, x_n^0)$ and of the value of

$$\nabla F_{(x_1^0,\ldots,x_n^0)} = \left(\frac{\partial F}{\partial x_1}, \ldots, \frac{\partial F}{\partial x_n}\right)(x_1^0, \ldots, x_n^0)$$

is enough to predict, with controlled accuracy, by linear approximation, the behavior of $y$ for parameters $(x_1, \ldots, x_n)$ close to $(x_1^0, \ldots, x_n^0)$.

In the case (2), both the parameters $(x_1, \ldots, x_n)$ and the value $y$ belong to the set $M = \{(y, x_1, \ldots, x_n) \in \mathbb{R}^{n+1}; G(y, x_1, \ldots, x_n) = 0\}$, and we would like to know whether or not this set may be (at least locally around one of its point $(y^0, x_1^0, \ldots, x_n^0)$) a graph of some function $(x_1, \ldots, x_n) \mapsto y = F(x_1, \ldots, x_n)$, as in the case (1). Using the implicit function theorem, we may try to reduce our equation to the explicit equation of (1), and then perform a linear approximation involving $\nabla F_{(x_1^0,\ldots,x_n^0)}$. Assuming that *a priori* we know a value $y^0$ such that for $(x_1^0, \ldots, x_n^0)$, $(y^0, x_1^0, \ldots, x_n^0) \in M$, this reduction is possible, locally around $(y^0, x_1^0, \ldots, x_n^0)$, under the condition that

$$\frac{\partial G}{\partial y}(y^0, x_1^0, \ldots, x_n^0) \neq 0$$

In this situation

$$\nabla F_{(x_1^0,\ldots,x_n^0)} = -\left(\frac{\partial G}{\partial x_1}, \ldots, \frac{\partial G}{\partial x_n}\right) \times (y^0, x_1^0, \ldots, x_n^0) \Big/ \frac{\partial G}{\partial y}(y^0, x_1^0, \ldots, x_n^0)$$

Now, as it is normally the case, when they come from observation, the variables $x_1, \ldots, x_n$ are known with an estimate and one sees that the larger

$$\left|\left(\frac{\partial G}{\partial x_1}, \ldots, \frac{\partial G}{\partial x_n}\right)(y^0, x_1^0, \ldots, x_n^0) \Big/ \frac{\partial G}{\partial y}(y^0, x_1^0, \ldots, x_n^0)\right|$$

is, the worse the estimate on $y$ near $y^0$.

Furthermore, assuming that $M$ is locally a graph of a function $(x_1, \ldots, x_n) \mapsto y = F(x_1, \ldots, x_n)$, for a given $(x_1, \ldots, x_n)$, the exact expression of $y = F(x_1, \ldots, x_n)$ and consequently the exact value of $\nabla F_{(x_1,\ldots,x_n)}$ is not possible to obtain; we have to approach it using an algorithm (classically the Newton algorithm), and closer

$$\frac{\partial G}{\partial y}(y^0, x_1^0, \ldots, x_n^0)$$

is to 0, the more such an algorithm is unstable.

Finally, in the case (3), skipping technical details, we encounter the same type of difficulties: we have to avoid small values for some gradient functions at a given point, in order to obtain, locally at some point $(x_1^0, \ldots, x_n^0)$, in a stable way, reliable information on $y$ in terms of $(x_1, \ldots, x_n)$.

To sum up, the prediction of a physical phenomenon by a mathematical law greatly depends not only on the noncancellation of some gradient functions, but, as we deal with approximations and algorithms, on how different those gradient functions are from zero.

This principle, of course, extends directly to applied problems (see the last of our examples in the final section): being close to singular values essentially means that the control (e.g., of the positions of some device by a manipulator) is poor.

The geometric counterpart of this analytic phenomenmon is called "transversality," the condition for some function $G$ to have a nonzero partial derivative

$$\frac{\partial G}{\partial y}(y^0, x_1^0, \ldots, x_n^0)$$

is equivalent to the condition

$$\nabla G_{(y^0, x_1^0,\ldots,x_n^0)} \oplus Ox_1 \cdots x_n = \mathbb{R}^{n+1}$$

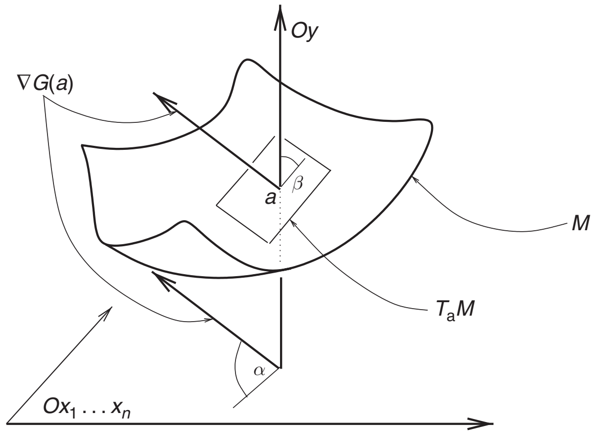



**Figure 1** Transversality of the manifold *M* and *Oy*.

or to the condition

$$T_{(y^0, x_1^0, \ldots, x_n^0)} M \oplus Oy = \mathbb{R}^{n+1}$$

where $T_a M$ is the tangent space of $M$ at $a \in M$.

We say that $\nabla G_{(y^0, x_1^0, \ldots, x_n^0)}$ is transverse to the space of parameters $Ox_1 \cdots x_n$ at $(y^0, x_1^0, \ldots, x_n^0)$, or that $M$ is transverse to $Oy$ at $(y^0, x_1^0, \ldots, x_n^0)$.

For some quantity $\epsilon > 0$, the condition that

$$\left| \left( \frac{\partial G}{\partial x_1}, \ldots, \frac{\partial G}{\partial x_n} \right) (y^0, x_1^0, \ldots, x_n^0) \Big/ \frac{\partial G}{\partial y} (y^0, x_1^0, \ldots, x_n^0) \right| \geq \frac{1}{\epsilon}$$

means that the angle $\alpha = (\widehat{\nabla G_{(y^0, x_1^0, \ldots, x_n^0)}, Ox_1 \cdots x_n})$ or the angle $\beta = (\widehat{T_{(y^0, x_1^0, \ldots, x_n^0)} M, Oy})$ is smaller than $\epsilon$ (see Figure 1).

Our purpose in the sequel is to indicate how we can quantify the situations described above (the defect of transversality), in order to generically or almost generically avoid them with quantified accuracy.

## Quantifying Transversality

Given two submanifolds $M$ and $N$ of the Euclidean space $\mathbb{R}^n$, we can measure the transversality defect of $(M, N)$ at $x \in \mathbb{R}^n$ with a differential criterion, both analytical and geometric.

Let us first introduce some notations. For a given linear map $L : \mathbb{R}^n \to \mathbb{R}^p$, the image by $L$ of the unit ball of $\mathbb{R}^n$ is an $r$-dimensional ellipsoid in $\mathbb{R}^p$ with semi-axes denoted as $l_1(L) \geq \cdots \geq l_r(L)$, where $r$ is the rank of $L$. For $r < p$, we denote $l_{r+1}(L) = 0, \ldots, l_p(L) = 0$.

Now, let $x \in M \cap N$; let $\pi : \mathbb{R}^n \to T_x N^\perp$ be the projection onto the orthogonal space of $T_x N$, $p = n - \dim(N)$ and $\pi_{|M}$ the restriction of $\pi$ to $M$.

### Definitions

We say that $(M, N)$ is transverse at $x$, and we denote it by $M \pitchfork_x N$, if and only if $\pi_{|M}$ is a submersion at $x$, that is, $D\pi_{|M(x)} : T_x M \to T_x N^\perp$ is onto.

For a given $\Lambda = (\epsilon_1, \ldots, \epsilon_p)$, $\epsilon_1 \geq \cdots \geq \epsilon_p$, we say that $(M, N)$ is $\Lambda$-nontransverse at $x$, and we denote it by $M \not\pitchfork_x^\Lambda N$, if and only if $l_i(D\pi_{|M(x)}) \leq \epsilon_i, \forall\, i \in \{1, \ldots, p\}$.

With these notations, we have: $M \not\pitchfork_x N$ (*i.e.*, $(M, N)$ nontransverse at $x$) if and only if $x \notin M \cap N$ or $M \not\pitchfork_x^\Lambda N$, for some $\Lambda$ with $\epsilon_p = 0$, and the more $(M, N)$ is $\Lambda$-nontransverse, with $\Lambda$ close to $(\epsilon_1, \ldots, \epsilon_{p-1}, 0)$, the less the manifolds $M$ and $N$ seem transverse at $x \in M \cap N$ (see Figure 2).

The final step in our formalism to give a convenient quantitative approach of transversality is the following: let $X$, $Y$ be two (real) Riemannian manifolds, $f : X \to Y$ a (smooth) mapping, $N \subset Y$ a submanifold of $Y$ with codimension $p$ in $N$, $y \in N$, and $\Phi : \mathcal{O} \to \mathbb{R}^p$ a submersion, where $\mathcal{O}$ is an open neighborhood of $x$ in $Y$, such that $\Phi^{-1}(\{0\}) = N \cap \mathcal{O}$. Then we say that $(f, N)$ is transverse at $x$, and we denote it by $f \pitchfork_x N$, if and only if $f \circ \Phi$ is submersive in $x$.

For a given $\Lambda = (\epsilon_1, \ldots, \epsilon_p)$, $\epsilon_1 \geq \cdots \geq \epsilon_p$, we say that $(f, N)$ is $(\Phi, \Lambda)$-nontransverse at $x$, and we denote it by $f \not\pitchfork_x^{(\Phi,\Lambda)} N$, if and only if $l_i(D[f \circ \Phi]_{(x)}) \leq \epsilon_i, \forall\, i \in \{1, \ldots, p\}$.

Clearly, we recognize the definition of transversality and of $\Lambda$-nontransversality of two submanifolds $M, N$ of $\mathbb{R}^n$ by letting $f : M \to \mathbb{R}^n$ be the inclusion and $\Phi = \pi_{|M}$ (for more details on transversality and stability, see, e.g., Golubitski and Guillemin (1973)).

With the definitions and notations above, our general problem may be posed as follows:

> For a $C^k$-regular $(k \in \mathbb{N} \cup \{\infty\})$ mapping $f : \mathbb{R}^n \to \mathbb{R}^p$ and a given $\Lambda = (\epsilon_1, \ldots, \epsilon_p)$, how large is the set $\Delta(f, B_r, \Lambda) = f(\Sigma(f, B_r, \Lambda))$, where $\Sigma(f, B_r, \Lambda) = \{x \in B_r \subset \mathbb{R}^n;\ l_i(Df_{(x)}) \leq \epsilon_i,\ \forall i \in \{1, \ldots, p\}\}$ and $B_r$ is a ball of radius $r$ in $\mathbb{R}^n$?

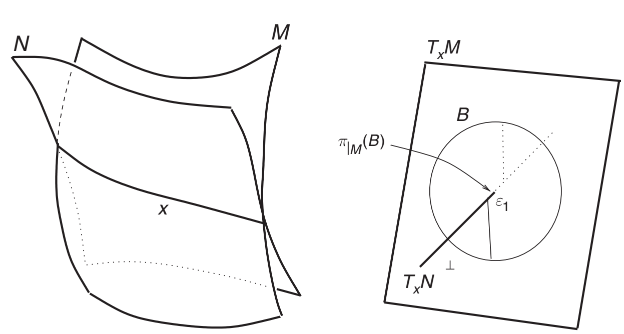



**Figure 2** Almost-nontransversality of *M* and *N*.

The "bad" set $\Delta(f, B_r, \Lambda)$ is called the set of $\Lambda$-almost critical values of $f$ (restricted to $B_r$). Our purpose is to show that one can control its size in terms of $k$ and $\Lambda$. However, before explicitly stating quantitative results, let us precise what we understand by "big set" or by "size of a set."

## Measure and Dimensions

We have a very natural way to measure a subset $A$ of a metric space. To do this, we consider $\alpha \geq 0$ a real number and we denote

$$\mathcal{A}_\nu = \left\{ (D_i)_{i \in \mathbb{N}}; A \subset \bigcup_{i \in \mathbb{N}} D_i \text{ and } |D_i| \leq \nu \right\}$$

where $|D_i|$ is the diameter of $D_i$,

$$\mathcal{H}^\alpha_\nu(A) = \inf \left\{ \sum_{i \in \mathbb{N}} |D_i|^\alpha; (D_i)_{i \in \mathbb{N}} \in \mathcal{A}_\nu \right\}$$

and

$$\mathcal{H}^\alpha(A) = \lim_{\nu \to 0} \mathcal{H}^\alpha_\nu(A) \in \mathbb{R} \cap \{\infty\}$$

$\mathcal{H}^\alpha(A)$ is called the $\alpha$-dimensional Hausdorff measure of $A$. It appears that when $\mathcal{H}^\alpha(A) \neq \infty$, $\mathcal{H}^{\alpha'}(A) = 0$ for $\alpha' > \alpha$, and when $\mathcal{H}^\alpha(A) \neq 0$, $\mathcal{H}^{\alpha'}(A) = \infty$ for $\alpha' < \alpha$. This gives rise to the following definition of the Hausdorff dimension of $A$:

$$\begin{aligned} \dim_{\mathcal{H}}(A) &= \inf\{\alpha; \mathcal{H}^\alpha(A) = 0\} \\ &= \sup\{\alpha; \mathcal{H}^\alpha(A) = \infty\} \end{aligned}$$

The Hausdorff dimension generalizes the classical notions of dimension, for instance, when $A$ is a subset of $\mathbb{R}^n, \dim_{\mathcal{H}}(A) \leq n$, a $d$-dimensional manifold has Hausdorff dimension $d$, and $\mathcal{H}^n(A)$ is the same as the Lebesgue measure $\mathcal{L}_n$ of $A$ (for a very large class of subset $A$, which we do not describe here. For more details on geometric measure theory, see Falconer (1986) and Federer (1969)).

Another convenient notion of dimension is the (metric) entropy dimension. Let us briefly define it. For a bounded subset $A$ in some metric space and a real number $\alpha > 0$, we denote $M(\alpha, A)$ the minimal number of closed balls of radius $\leq \alpha$, covering $A$. $H_\alpha(A) = \log_2(M(\alpha, A))$ is called the $\alpha$-entropy of the set $A$. This terminology was introduced in Kolmogorov and Tihomirov (1961) and reflects the fact that $H_\alpha(A)$ is the amount of information needed to digitally memorize $A$ with accuracy $\alpha$. The entropy dimension of $A$, $\dim_e(A)$, is the order of $M(\alpha, A)$ as $\alpha \to 0$. Precisely,

$$\begin{aligned} \dim_e(A) &= \limsup_{\alpha \to 0} \frac{\log(M(\alpha, A))}{\log(1/\alpha)} \\ &= \inf\{\delta; M(\alpha, A) \leq (1/\alpha)^\delta, \\ &\quad \text{for sufficiently small } \alpha\} \end{aligned}$$

We clearly have

$$\dim_{\mathcal{H}}(A) \leq \dim_e(A)$$

For any bounded set $A$ in $\mathbb{R}^n$, we can bound $M(\alpha, A)$ from above by a polynomial in $1/\alpha$ (see Ivanov (1975) and Yomdin and Comte (2004)):

$$M(\alpha, A) \leq c(n) \sum_{i=0}^{n} V_i(A)(1/\alpha)^i$$

where $c(n)$ only depends on $n$ and $V_i(A)$ (the $i$th variation of the set $A$) is the mean value, with respect to $P$ (for a suitable measure), of the number of connected components of $A \cap P$, with $P$ an affine $(n-i)$-dimensional space of $\mathbb{R}^n$.

Since for $A$ contained in a $d$-dimensional manifold, $V_i(A) = 0$ for $i > d$, we deduce from this inequality that in this case $M(\alpha, A)$ is bounded from above by a polynomial of degree $\leq d$ in $1/\alpha$.

Our goal is to explain that we can be more precise than this general inequality when $A$ is a set of critical or almost-critical values of a $\mathcal{C}^k$ mapping.

## Transversality Is a Generic Situation

The results in this section concern critical values, and not almost-critical values. They show that a "generic" point of the target space is not a critical value, and the more regular, the mapping the smaller the set of critical values. Such theorems relating the regularity of a mapping and the size of its critical values are called Morse–Sard type theorems (see Sard (1942, 1958, 1965)). The simplest theorem in this direction is the following:

**Theorem 1** ($\mathcal{C}^\infty$ Morse–Sard theorem) (Morse 1939, Sard 1942, Holm 1987). *Let $f: \mathbb{R}^n \to \mathbb{R}^p$ be a $\mathcal{C}^\infty$-regular mapping. Then $\mathcal{H}^p(\Delta(f, B_r)) = 0$, where $\Delta(f, B_r) = f(\Sigma(\mathrm{f}, \mathrm{B_r}))$ and $\Sigma(f, B_r)$ is the set of points $x \in B_r$ where* $\mathrm{rank}(Df_{(x)}) < p$.

The set $\Delta(f, B_r)$ is the image, under $f$, of the points of the ball $B_r$ in the source space at which $f$ is not submersive, that is, the set of critical values of $f$. Consequently, the Morse–Sard theorem ensures that for almost all points $y$ in the target space, $f^{-1}(\{y\})$ is either empty or a smooth submanifold of the source space of dimension $n - p$.

Note that $\Delta(f, B_r) = \Delta(f, B_r, \Lambda)$ for some convenient $\Lambda = (\epsilon_1, \ldots, \epsilon_p)$ with $\epsilon_p = 0$, because $x \mapsto l_i(Df_{(x)})$ is bounded on $B_r$, for all $i \in \{1, \ldots, p\}$.

Now, we can concentrate our attention on more singular points than the critical ones, those at which the rank $\rho$ of $f$ is prescribed. Let us denote such points by $\Delta^\rho(f, B_r)$, for $\rho < p$. By definition, $\Delta^\rho(f, B_r) = f(\Sigma^\rho(f, B_r))$, where $\Sigma^\rho(f, B_r) = \{x \in B_r \subset R^n; \mathrm{rank}(Df_{(x)}) \leq \rho\}$. With these notations, the result for rank-$r$ critical values is the following:

**Theorem 2** ($\mathcal{C}^k$ Morse–Sard theorem for rank-$r$ critical values) (Federer 1969). *Let $f: \mathbb{R}^n \to \mathbb{R}^p$ be a $\mathcal{C}^k$-regular mapping. Then $\mathcal{H}^{\rho+(n-\rho)/k}(\Delta^\rho(f, B_r)) = 0$. In particular,*

$$\dim_{\mathcal{H}}(\Delta^\rho(f, B_r)) \leq \rho + \frac{n-\rho}{k}$$

One can produce examples showing that the bound of Theorem 2 is the sharpest one (see Comte (1996), Whitney (1935), Grinberg (1985), and Yomdin and Comte (2004)).

We note that Theorem 1 is a corollary of Theorem 2 (just replace $k$ by $\infty$ and $\rho$ by $p-1$ in Theorem 2). This result tells nothing about the entropy dimension of $\Delta^\rho(f, B_r)$; in the next section, we will bound the growth of entropy of almost-critical values.

## Almost-Transversality Is Almost Generic

In this section, $f: \mathbb{R}^n \to \mathbb{R}^p$ is a $\mathcal{C}^k$ mapping. We denote by $K$ a Lipschitz constant of $D^{k-1}f$ on $B_r$ and by $R_k(f)$ the quantity $(K/(k-1)!) \cdot r^k$. We have:

**Theorem 3** ($\mathcal{C}^k$ quantitative Morse–Sard theorem) (Yomdin 1983 Yomdin and Comte 2004). *Let $f: \mathbb{R}^n \to \mathbb{R}^p$ be a $\mathcal{C}^k$ mapping, $\Lambda = (\epsilon_1, \ldots, \epsilon_p)$, $\epsilon_1 \geq \cdots \geq \epsilon_p$, and let us denote $\epsilon_0 = 1$. We have (for $\alpha \leq R_k(f)$):*

$$M(\alpha, \Delta(f, B_r, \Lambda)) \leq C \cdot \sum_{i=0}^{p} \epsilon_0 \cdots \epsilon_i \left(\frac{r}{\alpha}\right)^i \left(\frac{R_k(f)}{\alpha}\right)^{(n-i)/k}$$

*where $C$ is a constant depending only on $n$, $p$, and $k$.*

As a corollary, one can bound the entropy dimension of $\Delta^\rho(f, B_r)$ by $\rho + (n-\rho)/k$, and hence its Hausdorff dimension, again finding Theorem 2: we just have to put $\epsilon_{\rho+1} = 0$ and $\epsilon_1, \ldots, \epsilon_\rho$ large enough, that is, $\epsilon_i \geq \lambda_i(Df_{(x)})$, for all $x \in B_r$, in Theorem 3, to obtain:

**Theorem 4** ($\mathcal{C}^k$ entropy Morse–Sard theorem) (Yomdin 1983 Yomdin and Comte 2004). *Let $f: \mathbb{R}^n \to \mathbb{R}^p$ be a $\mathcal{C}^k$ mapping, let us denote $\epsilon_0 = 1$ and $\epsilon_i = \sup\{\lambda_i(Df_{(x)}); x \in B_r\}$, for $i \in \{1, \ldots, \rho\}$. We have (for $\alpha \leq R_k(f)$):*

$$M(\alpha, \Delta^\rho(f, B_r)) \leq C \cdot \sum_{i=0}^{\rho} \epsilon_0 \cdots \epsilon_i \left(\frac{r}{\alpha}\right)^i \left(\frac{R_k(f)}{\alpha}\right)^{(n-i)/k}$$

*where $C$ is a constant depending only on $n$, $p$, and $k$. In particular,*

$$\dim_{\mathcal{H}}(\Delta^\rho(f, B_r)) \leq \dim_e(\Delta^\rho(f, B_r)) \leq \rho + \frac{n-\rho}{k}$$

Again we have examples showing that this bound is sharp (see Yomdin and Comte 2004).

Furthermore, the mapping $f$ in Theorems 2–4 may be of real differentiability class (Hölder smoothness class $\mathcal{C}^k$), with the same conclusions in these theorems. That is, $k$ may be a real number written as $k = p + \beta$ with $\beta \in [0, 1]$, $p \in \mathbb{N}\backslash\{0\}$, and $f$ is $\mathcal{C}^k$ means that $f$ is $p$ times differentiable and there exists a constant $C > 0$ such that for all $x, y \in B_r, \|D^p f_{(x)} - D^p f_{(y)}\| \leq C \cdot \|x - y\|^\beta$ (see Yomdin and Comte (2004)).

## Examples

Let us denote by $A$ the set of real polynomial mappings of degree $d$ and of the following type:

$$x \mapsto Q(a, x) = 1 + \sum_{j=1}^{d} a_j x^j$$

with $a = (a_1, \ldots, a_d)$ and $\|a\| \leq 1$ (where $\|\|$ is the Euclidean norm of $\mathbb{R}^d$). We identify the set $A$ with $B^d(0,1) = \{a \in \mathbb{R}^d; \|a\| \leq 1\}$.

We want to bound the $\alpha$-entropy of the set of such polynomials for which the real roots are multiple or almost multiple.

We denote by $V$ the set $V = \{(a, x) \in \mathbb{R}^{d+1}; Q(a, x) = 0\}$. At points $(a, x)$ of $V$ with $\nabla Q_{(a,x)} \neq 0$, $V$ is a $\mathcal{C}^\infty$ manifold of codimension 1 of $\mathbb{R}^{d+1}$. We denote by $V^{\mathrm{reg}} = \{(a, x) \in V; \nabla Q_{(a,x)} \neq 0\}$ and by $V^{\mathrm{sing}} = \{(a, x) \in V; \nabla Q_{(a,x)} = 0\} = V \setminus V^{\mathrm{reg}}$. By Whitney (1957), $V^{\mathrm{sing}}$ is a union of smooth manifolds of dimension $\leq d-1$.

A root $x$ of a polynomial $Q(a, \cdot)$ is multiple if and only if

$$Q(a, x) = \frac{\partial Q}{\partial x}(a, x) = 0$$

Consequently, the set $A^\Sigma$ of polynomials of $A$ with multiple roots is $\pi(V^{\mathrm{sing}}) \cup \Delta(\pi_{|V^{\mathrm{reg}}})$, where $\pi: \mathbb{R}^{d+1} \to \mathbb{R}^d$ is the standard projection $\pi(a, x) = a$, and $\Delta(\pi_{|V^{\mathrm{reg}}})$ is the set $\{(a, x) \in V^{\mathrm{reg}}; Ox \subset T_{(a,x)} V^{\mathrm{reg}}\}$ of critical values of $\pi_{|V^{\mathrm{reg}}}$. By Sard's theorem

(Theorem 2), $\dim_{\mathcal{H}}(\Delta(\pi_{|V^{\mathrm{reg}}})) \leq d-1$. Since $\dim_{\mathcal{H}}(\pi(V^{\mathrm{sing}})) \leq d-1$, we obtain: $\dim_{\mathcal{H}}(A^{\Sigma}) \leq d-1$: thus, having distinct roots is a generic property.

Let, as above, $\Lambda = (\epsilon_1, \ldots, \epsilon_d)$ with $\epsilon_1 \geq \cdots \geq \epsilon_d$ and $\epsilon_0 = 1$. A root $x$ of a polynomial $Q(a,\cdot) \in A$ is said to be $\Lambda$-almost multiple if and only if $Q(a,x) = 0$ and $V \not\pitchfork_x^{\Lambda} Ox$, that is, $(a,x) \in V^{\mathrm{sing}}$ or $\sin(T_{(a,x)}V^{\mathrm{reg}}, Ox) \leq \epsilon_d$. This condition only concerns $\epsilon_d$ and we can take $\epsilon_1 = \cdots = \epsilon_{d-1} = 1$. We denote $A^{\Sigma,\Lambda}$ to be the set of polynomials of $A$ with (at least) a $\Lambda$-almost multiple root. By Theorem 3,

$$M(\alpha, A^{\Sigma,\Lambda} \backslash \pi(V^{\mathrm{sing}})) \leq C \cdot \left[\sum_{i=0}^{d-1}\left(\frac{1}{\alpha}\right)^i + \epsilon_d \cdot \left(\frac{1}{\alpha}\right)^d\right]$$

But $\pi(V^{\mathrm{sing}})$ being a finite union of manifolds of dimension at most $d-1$, we finally obtain

$$M(\alpha, A^{\Sigma,\Lambda}) \leq C' \cdot \left[\sum_{i=0}^{d-1}\left(\frac{1}{\alpha}\right)^i + \epsilon_d \cdot \left(\frac{1}{\alpha}\right)^d\right]$$

Thus, having no $\Lambda$-almost multiple root is $\Lambda$-almost a generic property. In Figure 3, we represent $V$ for $d = 3$ and $a_3 = 1$,

$$W = \left\{(a,x) \in \mathbb{R}^{d+1};\ \frac{\partial Q}{\partial x}(a,x) = 0\right\}$$

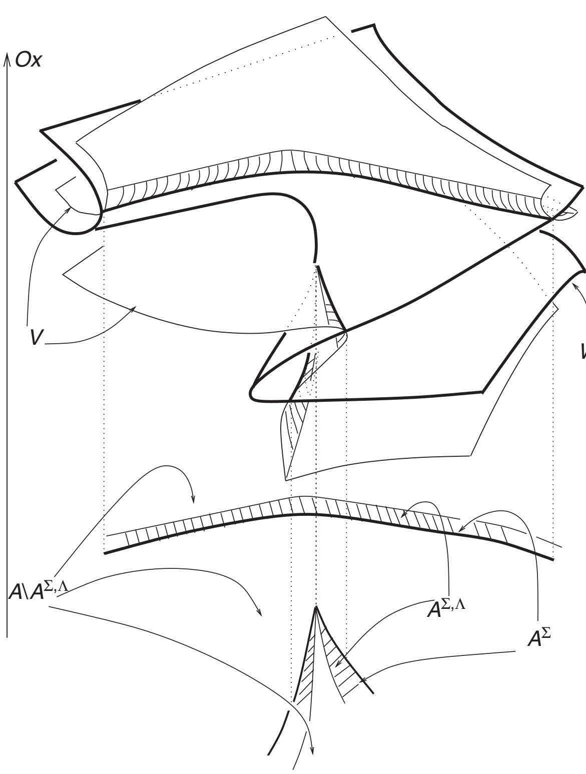



**Figure 3** The space of polynomials of type $1 + a_1x + a_2x + x^3$ with almost-multiple roots.

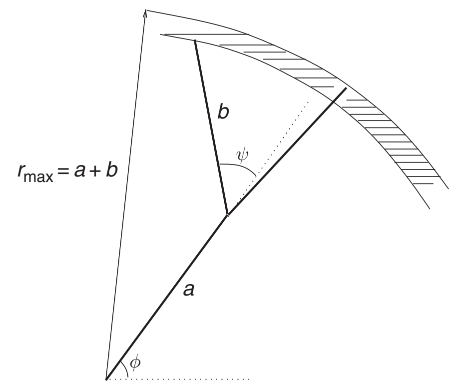



**Figure 4** Almost-critical points of the distance function of $P$ to the origin.

The next example comes from robotics: let us consider a planar robotic manipulator consisting of two jointed bars of length $a$ and $b$, as presented in Figure 4. We may parametrize the positions of the endpoint $P$ of this device by the angles $\phi$ and $\psi$ (see Figure 4). Now the distance $r$ from the origin to $P$ is $r^2 = \|P\|^2 = a^2 + b^2 + 2ab\cos(\psi)$. The critical points of $r$ are given by

$$\frac{\mathrm{d}r}{\mathrm{d}\psi}(\psi) = -2ab\sin(\psi) = 0$$

and correspond to the circle $\psi = 0$. The critical value of $r$ is $a+b$. Near these critical positions, the control of $r$ with respect to $\psi$ is poor; we would like to avoid those near-critical values. Given $\epsilon > 0$, the condition

$$\left|\frac{\mathrm{d}r}{\mathrm{d}\psi}(\psi)\right| \leq \epsilon$$

implies $|\psi| \leq \arcsin(\epsilon/2ab)$, and the $\epsilon$-near-critical values of $r$ are

$$r_{\max}^2 - r^2 \leq 2ab[1 - \cos(\arcsin(\epsilon/2ab))]$$

where $r_{\max}$ is $a+b$; thus, they are contained in an interval of length $\leq c \cdot \epsilon^2/(4ab \cdot r_{\max})$, and $M(\alpha, \Delta(r,\epsilon)) \leq c \cdot \epsilon^2/(4ab \cdot r_{\max} \cdot \alpha)$ (Theorem 3 gives $M(\alpha, \Delta(r,\epsilon)) \leq C(1 + \epsilon/\alpha)$.

*See also:* Entanglement; Entanglement Measures; Quantum Entropy; Singularity and Bifurcation Theory.

## Further Reading


Comte G (1996) Sur les hypothèses du théorème de Sard pour le lieu critique de rang 0. *C. R. Academy of Science Paris* 323(I): 143–146.

de Araujo Moreira CGT (2001) Hausdorff measures and the Morse–Sard theorem. *Publication Mathemàtiques* 45(1): 149–162.

Falconer KJ (1986) *The Geometry of Fractal Sets*, Cambridge Tracts in Mathematics, vol. 85. Cambridge: Cambridge University Press.
Federer H (1969) *Geometric Measure Theory*, Grundlehren Math. Wiss., vol. 153. Springer.
Golubitski V and Guillemin V (1973) *Stable Mappings and Their Singularities*, Graduate Texts in Mathematics, vol. 14. Springer.
Grinberg EL (1985) On the smoothness hypothesis in Sard's theorem. *American Mathematical Monthly* 92(10): 733–734.
Holm P (1987) The theorem of Brown and Sard. *L'enseignement Mathématiques* 33: 199–202.
Ivanov LD (1975) In: Vituskin AG (ed.) *Variazii množhestv i funktsii (Russian) (Variations of Sets and Functions)*. Moscow: Nauka.
Kolmogorov AN and Tihomirov VM (1961) $\varepsilon$-Entropy and $\varepsilon$-capacity of sets in functional space. *American Mathematical Society Translations* 17: 277–364.
Morse AP (1939) The behavior of a function on its critical set. *Annals of Mathematics* 40(1): 62–70.
Sard A (1942) The measure of the critical values of differentiable maps. *Bulletin of the American Mathematical Society* 48: 883–890.
Sard A (1958) Images of critical sets. *Annals of Mathematics* 68: 247–259.
Sard A (1965) Hausdorff measure of critical images on Banach manifolds. *American Journal of Mathematics* 87: 158–174.
Whitney H (1935) A function not constant on a connected set of critical points. *Duke Mathematical Journal* 1: 514–517.
Whitney H (1957) Elementary structure of real algebraic varieties. *Annals of Mathematics* 66: 545–556.
Yomdin Y (1983) The geometry of critical and near-critical values of differentiable mappings. *Mathematische Annalen* 264(4): 495–515.
Yomdin Y and Comte G (2004) *Tame Geometry with Application in Smooth Analysis*, Lecture Notes in Mathematics, vol. 1834. Berlin: Springer.

# Equivariant Cohomology and the Cartan Model

**E Meinrenken**, University of Toronto, Toronto, ON, Canada



## Introduction

If a compact Lie group $G$ acts on a manifold $M$, the space $M/G$ of orbits of the action is usually a singular space. Nonetheless, it is often possible to develop a "differential geometry" of the orbit space in terms of appropriately defined equivariant objects on $M$. This article is mostly concerned with "differential forms on $M/G$." A first idea would be to work with the complex of "basic" forms on $M$, but for many purposes this complex turns out to be too small. A much more useful complex of equivariant differential forms on $M$ was introduced by Cartan (1950). In retrospect, Cartan's approach presented a differential form model for the equivariant cohomology of $M$, as defined by A Borel (1960). Borel's construction replaces the quotient $M/G$ by a better-behaved (but usually infinite-dimensional) homotopy quotient $M_G$, and Cartan's complex should be viewed as a model for forms on $M_G$.

One of the features of equivariant cohomology are the localization formulas for the integrals of equivariant cocycles. The first instance of such an integration formula was the "exact stationary phase formula," discovered by Duistermaat and Heckman. This formula was quickly recognized by Berline and Vergne (1983) and Atiyah and Bott (1984), as a localization principle in equivariant cohomology. Today, equivariant localization is a basic tool in mathematical physics, with numerous applications.

This article begins with Borel's topological definition of equivariant cohomology, then proceeds to describe H Cartan's more algebraic approach, and concludes with a discussion of localization principles.

As additional references for the material covered here, we particularly recommend books by Berline, Getzler, and Vergne (1992) and Guillemin and Sternberg (1999).

## Borel's Model of $H_G(M)$

Let $G$ be a topological group. A $G$-space is a topological space $M$ on which $G$ acts by transformations $g \mapsto a_g$, in such a way that the action map

$$a : G \times M \to M \qquad [1]$$

is continuous. An important special case of $G$-spaces are principal $G$-bundles $E \to B$, that is, $G$-spaces locally isomorphic to products $U \times G$.

**Definition 1** A classifying bundle for $G$ is a principal $G$-bundle $EG \to BG$, with the following universal property: for any principal $G$-bundle $E \to B$, there is a map $f : B \to BG$, unique up to homotopy, such that $E$ is isomorphic to the pullback bundle $f^*EG$. The map $f$ is known as a "classifying map" of the principal bundle.

To be precise, the base spaces of the principal bundles considered here must satisfy some technical condition. For a careful discussion, see Husemoller (1994). Classifying bundles exist for all $G$ (by a construction due to Milnor (1956)), and are unique up to $G$-homotopy equivalence.

It is a basic fact that principal $G$-bundles with contractible total space are classifying bundles.

**Examples 2**

(i) The bundle $\mathbb{R} \to \mathbb{R}/\mathbb{Z} = S^1$ is a classifying bundle for $G = \mathbb{Z}$.

(ii) Let $\mathcal{H}$ be a separable complex Hilbert space, $\dim \mathcal{H} = \infty$. It is known that unit sphere $S(\mathcal{H})$ is contractible. It is thus a classifying U(1)-bundle, with the projective space $P(\mathcal{H})$ as base. More generally, the Stiefel manifold $\mathrm{St}(k, \mathcal{H})$ of unitary $k$-frames is a classifying U($k$)-bundle, with base the Grassmann manifold $\mathrm{Gr}(k, \mathcal{H})$ of $k$-planes.

(iii) Any compact Lie group $G$ arises as a closed subgroup of U($k$), for $k$ sufficiently large. Hence, the Stiefel manifold $\mathrm{St}(k, \mathcal{H})$ also serves as a model for $EG$.

(iv) The based loop group $G = L_0K$ of a connected Lie group $K$ acts by gauge transformations on the space of connections $\mathcal{A}(S^1) = \Omega^1(S^1, \mathfrak{k})$. This is a classifying bundle for $L_0K$, with base $K$. The quotient map takes a connection to its holonomy.

For any commutative ring $R$ (e.g., $\mathbb{Z}, \mathbb{R}, \mathbb{Z}_2$), let $H(\cdot\,; R)$ denote the (singular) cohomology with coefficients in $R$. Recall that $H(\cdot\,; R)$ is a graded commutative ring under cup product.

**Definition 3** The equivariant cohomology $H_G(M) = H_G(M; R)$ of a $G$-space $M$ is the cohomology ring of its homotopy quotient $M_G = EG \times_G M$:

$$H_G(M; R) = H(M_G; R) \qquad [2]$$

Equivariant cohomology is a contravariant functor from the category of $G$-spaces to the category of $R$-modules. The $G$-map $M \to \mathrm{pt}$ induces an algebra homomorphism from $H_G(\mathrm{pt}) = H(BG)$ to $H_G(M)$. In this way, $H_G(M)$ is a module over the ring $H(BG)$.

**Example 4** (Principal $G$-bundles). Suppose $E \to B$ is a principal $G$-bundle. The homotopy quotient $E_G$ may be viewed as a bundle $E \times_G EG$ over $B$. Since the fiber is contractible, there is a homotopy equivalence

$$E_G \simeq B \qquad [3]$$

and therefore $H_G(E) = H(B)$.

**Example 5** (Homogeneous spaces). If $K$ is a closed subgroup of a Lie group $G$, the space $EG$ may be viewed as a model for $EK$, with $BK = EG/K = EG \times_K (G/K)$. Hence,

$$H_G(G/K) = H(BK) \qquad [4]$$

Let us briefly describe two of the main techniques for computing $H_G(M)$.

1. *Leray spectral sequences*. If $R$ is a field, the equivariant cohomology may be computed as the $E_\infty$ term of the spectral sequence for the fibration $M_G \to BG$. If $BG$ is simply connected (as is the case for all compact connected Lie groups), the $E_2$-term of the spectral sequence reads

$$E_2^{p,q} = H^p(BG) \otimes H^q(M) \qquad [5]$$

2. *Mayer–Vietoris sequences*. If $M = U_1 \cup U_2$ is a union of two $G$-invariant open subsets, there is a long exact sequence

$$\cdots \to H_G^k(M) \to H_G^k(U_1) \oplus H_G^k(U_2) \to$$
$$\to H_G^k(U_1 \cap U_2) \to H_G^{k+1}(M) \to \cdots$$

More generally, associated to any $G$-invariant open cover, there is a spectral sequence converging to $H_G(M)$.

**Example 6** Consider the standard U(1)-action on $S^2$ by rotations. Cover $S^2$ by two open sets $U_\pm$, given as the complement of the south pole and north pole, respectively. Since $U_+ \cap U_-$ retracts onto the equatorial circle, on which U(1) acts freely, its equivariant cohomology vanishes except in degree 0. On the other hand, $U_\pm$ retract onto the poles $p_\pm$. Hence, by the Mayer–Vietoris sequence the map $H^k_{\mathrm{U}(1)}(S^2) \cong H^k_{\mathrm{U}(1)}(p_+) \oplus H^k_{\mathrm{U}(1)}(p_-)$ given by pullback to the fixed points is an isomorphism for $k > 0$. Since the pullback map is a ring homomorphism, we conclude that $H_{\mathrm{U}(1)}(S^2; R)$ is the commutative ring generated by two elements $x_\pm$ of degree 2, subject to a single relation $x_+x_- = 0$.

## $\mathfrak{g}$-Differential Algebras

Let $G$ be a Lie group, with Lie algebra $\mathfrak{g}$. A $G$-manifold is a manifold $M$ together with a $G$-action such that the action map [1] is smooth. We would like to introduce the concept of equivariant differential forms on $M$. This complex should play the role of differential forms on the infinite-dimensional space $M_G$. In Cartan's approach, the starting point is an algebraic model for the differential forms on the classifying bundle $EG$.

The algebraic machinery will only depend on the infinitesimal action of $G$. It is therefore convenient to introduce the following concept.

**Definition 7** Let $\mathfrak{g}$ be a finite-dimensional Lie algebra. A $\mathfrak{g}$-manifold is a manifold $M$, together with a Lie algebra homomorphism $a: \mathfrak{g} \to \mathfrak{X}(M), \xi \mapsto a_\xi$ into the Lie algebra of vector fields on $M$, such that the map $\mathfrak{g} \times M \to TM, (\xi, m) \mapsto a_\xi(m)$ is smooth.

Any $G$-manifold $M$ becomes a $\mathfrak{g}$-manifold by taking $a_\xi$ to be the generating vector field

$$a_\xi(m) := \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} a_{\exp(-t\xi)}(m) \qquad [6]$$

Conversely, if $G$ is simply connected, and $M$ is a $\mathfrak{g}$-manifold for which all of the vector fields $a_\xi$ are complete, the $\mathfrak{g}$-action integrates uniquely to an action of the group $G$.

The de Rham algebra $(\Omega(M), \mathrm{d})$ of differential forms on a $\mathfrak{g}$-manifold $M$ carries graded derivations $L_\xi = L(a_\xi)$ (Lie derivatives, degree 0) and $\iota_\xi = \iota(a_\xi)$ (contractions, degree –1). One has the following graded commutation relations:

$$[\mathrm{d}, \mathrm{d}] = 0, \qquad [L_\xi, \mathrm{d}] = 0, \qquad [\iota_\xi, \mathrm{d}] = L_\xi \qquad [7]$$

$$[\iota_\xi, \iota_\eta] = 0, \quad [L_\xi, L_\eta] = L_{[\xi,\eta]_\mathfrak{g}}, \quad [L_\xi, \iota_\eta] = \iota_{[\xi,\eta]_\mathfrak{g}} \qquad [8]$$

More generally, the following definitions are introduced.

**Definition 8** A $\mathfrak{g}$-differential algebra ($\mathfrak{g}$-da) is a commutative graded algebra $A = \bigoplus_{n=0}^{\infty} A^n$, equipped with graded derivations $\mathrm{d}, L_\xi, \iota_\xi$ of degrees $1, 0, -1$ (where $L_\xi, \iota_\xi$ depend linearly on $\xi \in \mathfrak{g}$), satisfying the graded commutation relations [7] and [8].

**Definition 9** For any $\mathfrak{g}$-da $A$, one defines the horizontal subalgebra $\mathcal{A}_{\mathrm{hor}} = \bigcap_\xi \ker(\iota_\xi)$, the invariant subalgebra $\mathcal{A}^\mathfrak{g} = \bigcap_\xi \ker(L_\xi)$, and the basic subalgebra $\mathcal{A}_{\mathrm{basic}} = \mathcal{A}_{\mathrm{hor}} \cap \mathcal{A}^\mathfrak{g}$.

Note that the basic subalgebra is a differential subcomplex of $A$.

**Definition 10** A connection on a $\mathfrak{g}$-da is an invariant element $\theta \in A^1 \otimes \mathfrak{g}$, with the property $\iota_\xi \theta = \xi$. The curvature of a connection is the element $F^\theta \in A^2 \otimes \mathfrak{g}$ given as $F^\theta = \mathrm{d}\theta + (1/2)[\theta, \theta]_\mathfrak{g}$.

$\mathfrak{g}$-da's $\mathcal{A}$ admitting connections are the algebraic counterparts of (smooth) principal bundles, with $\mathcal{A}_{\mathrm{basic}}$ playing the role of the base of the principal bundle.

## Weil Algebra

The Weil algebra $W\mathfrak{g}$ is the algebraic analog to the classifying bundle $EG$. Similar to $EG$, it may be characterized by a universal property:

**Theorem 11** *There exists a* $\mathfrak{g}$-da $W\mathfrak{g}$ *with connection* $\theta_W$, *having the following universal property*: *if* $\mathcal{A}$ *is a* $\mathfrak{g}$-da *with connection* $\theta$, *there is a unique algebra homomorphism* $c : W\mathfrak{g} \to \mathcal{A}$ taking $\theta_W$ to $\theta$.

Clearly, the universal property characterizes $W\mathfrak{g}$ up to a unique isomorphism. To get an explicit construction, choose a basis $\{e_a\}$ of $\mathfrak{g}$, with dual basis $\{e^a\}$ of $\mathfrak{g}^*$. Let $y^a \in \wedge^1 \mathfrak{g}^*$ be the corresponding generators of the exterior algebra, and $v^a \in S^1 \mathfrak{g}^*$ the generators of the symmetric algebra. Let

$$W^n \mathfrak{g} = \bigoplus_{2i+j=n} S^i \mathfrak{g}^* \otimes \wedge^j \mathfrak{g}^* \qquad [9]$$

carry the differential

$$\mathrm{d}y^a = v^a + \tfrac{1}{2} f^a_{bc} y^b y^c \qquad [10]$$

$$\mathrm{d}v^a = -f^a_{bc} v^b y^c \qquad [11]$$

where $f^a_{bc} = \langle e^a, [e_b, e_c]_\mathfrak{g} \rangle$ are the structure constants of $\mathfrak{g}$. Define the contractions $\iota_a = \iota_{e_a}$ by

$$\iota_a y^b = \delta^b_a, \qquad \iota_a v^b = 0 \qquad [12]$$

and let $L_a = [\mathrm{d}, \iota_a]$. Then $L_a$ are the generators for the adjoint action on $W\mathfrak{g}$. The element $\theta_W = y^a \otimes e_a \in W^1\mathfrak{g} \otimes \mathfrak{g}$ is a connection on $W\mathfrak{g}$. Notice that we could also use $y^a$ and $\mathrm{d}y^a$ as generators of $W\mathfrak{g}$. This identifies $W\mathfrak{g}$ with the Koszul algebra, and implies:

**Theorem 12** $W\mathfrak{g}$ *is acyclic*, *that is*, *the inclusion* $\mathbb{R} \to W\mathfrak{g}$ *is a homotopy equivalence.*

Acyclicity of $W\mathfrak{g}$ corresponds to the contractibility of the total space of $EG$.

The basic subalgebra of $W\mathfrak{g}$ is equal to $(S\mathfrak{g}^*)^\mathfrak{g}$, and the differential restricts to zero on this subalgebra, since d changes parity. Hence, if $\mathcal{A}$ is a $\mathfrak{g}$-da with connection, the characteristic homomorphism $c : W\mathfrak{g} \to \mathcal{A}$ induces an algebra homomorphism, $(S\mathfrak{g}^*)^\mathfrak{g} \to H(\mathcal{A}_{\mathrm{basic}})$. This homomorphism is independent of $\theta$:

**Theorem 13** *Suppose* $\theta_0, \theta_1$ *are two connections on a* $\mathfrak{g}$-da $\mathcal{A}$. *Then their characteristic homomorphisms* $c_0, c_1 : W\mathfrak{g} \to \mathcal{A}$ are $\mathfrak{g}$-*homotopic. That is*, *there is a chain homotopy intertwining contractions and Lie derivatives*.

**Remark 14** One obtains other interesting examples of $\mathfrak{g}$-da's if one drops the commutativity assumption from the definition. For instance, suppose $\mathfrak{g}$ carries an invariant scalar product. Let $\mathrm{Cl}(\mathfrak{g})$ be the corresponding Clifford algebra, and $\mathrm{U}(\mathfrak{g})$ the enveloping algebra. The noncommutative Weil algebra (introduced by Alekseev and Meinrenken 2002)

$$\mathcal{W}\mathfrak{g} = U\mathfrak{g} \otimes \mathrm{Cl}(\mathfrak{g}) \qquad [13]$$

is a (noncommutative) $\mathfrak{g}$-da, with the derivations d, $L_a, \iota_a$ defined on generators by the same formulas as for $W\mathfrak{g}$.

## Equivariant Cohomology of $\mathfrak{g}$-da's

In analogy to $H_G(M) := H(M_G)$, we now declare:

**Definition 15** The equivariant cohomology algebra of a $\mathfrak{g}$-da $\mathcal{A}$ is the cohomology of the differential algebra $\mathcal{A}_\mathfrak{g} := (W\mathfrak{g} \otimes \mathcal{A})_{\text{basic}}$:

$$H_\mathfrak{g}(\mathcal{A}) := H(\mathcal{A}_\mathfrak{g}) \qquad [14]$$

The equivariant cohomology $H_\mathfrak{g}(\mathcal{A})$ has functorial properties parallel to those of $H_G(M)$. In particular, $H_\mathfrak{g}(A)$ is a module over

$$H_\mathfrak{g}(\{0\}) = H((W\mathfrak{g})_{\text{basic}}) = (S\mathfrak{g}^*)^\mathfrak{g} \qquad [15]$$

**Theorem 16** *Suppose $\mathcal{A}$ is a $\mathfrak{g}$-da with connection $\theta$, and let $c: W\mathfrak{g} \to \mathcal{A}$ be the characteristic homomorphism. Then*

$$W\mathfrak{g} \otimes \mathcal{A} \to \mathcal{A}, \qquad w \otimes x \mapsto c(w)\, x \qquad [16]$$

*is a $\mathfrak{g}$-homotopy equivalence, with $\mathfrak{g}$-homotopy inverse the inclusion*

$$\mathcal{A} \to W\mathfrak{g} \otimes \mathcal{A}, \qquad x \mapsto 1 \otimes x \qquad [17]$$

*In particular, there is a canonical isomorphism*

$$H(\mathcal{A}_{\text{basic}}) \cong H_\mathfrak{g}(\mathcal{A}) \qquad [18]$$

**Proof** By Theorem 13, the automorphism $w \otimes x \mapsto 1 \otimes c(w)\, x$ of $W\mathfrak{g} \otimes \mathcal{A}$ is $\mathfrak{g}$-homotopic to the identity map.

□

The above definition of the complex $\mathcal{A}_\mathfrak{g}$ is often referred to as the Weil model of equivariant cohomology, while the term Cartan model is reserved for a slightly different description of $\mathcal{A}_\mathfrak{g}$. Identify the space $(S\mathfrak{g}^* \otimes \mathcal{A})^\mathfrak{g}$ with the algebra of equivariant $\mathcal{A}$-valued polynomial functions $\alpha: \mathfrak{g} \to \mathcal{A}$. Define a differential $\mathrm{d}_\mathfrak{g}$ on this space by setting

$$(\mathrm{d}_\mathfrak{g}\alpha)(\xi) = \mathrm{d}(\alpha(\xi)) - \iota_\xi \alpha(\xi) \qquad [19]$$

**Theorem 17** (H Cartan). *The natural projection $W\mathfrak{g} \otimes \mathcal{A} \to S\mathfrak{g}^* \otimes \mathcal{A}$ restricts to an isomorphism of differential algebras, $\mathcal{A}_\mathfrak{g} \cong (S\mathfrak{g}^* \otimes \mathcal{A})^\mathfrak{g}$.*

Suppose $\mathcal{A}$ carries a connection $\theta$. The $\mathfrak{g}$-homotopy equivalence [16] induces a homotopy equivalence $\mathcal{A}_\mathfrak{g} \to \mathcal{A}_{\text{basic}}$ of the basic subcomplexes. By explicit calculation, the corresponding map for the Cartan model is given by

$$(S\mathfrak{g}^* \otimes \mathcal{A})^\mathfrak{g} \to \mathcal{A}_{\text{basic}}, \qquad \alpha \mapsto P^\theta_{\text{hor}}(\alpha(F^\theta)) \qquad [20]$$

Here $\alpha(F^\theta) \in \mathcal{A}^\mathfrak{g}$ is the result of substituting the curvature of $\theta$, and $P_{\text{hor}}: \mathcal{A} \to \mathcal{A}_{\text{hor}}$ is horizontal projection. On elements of $(S\mathfrak{g}^*)^\mathfrak{g} \subset (S\mathfrak{g}^* \otimes \mathcal{A})^\mathfrak{g}$, the map [20] specializes to the Chern–Weil homomorphism.

There is an algebraic counterpart of the Leray spectral sequence: introduce a filtration

$$F^p \mathcal{A}_\mathfrak{g}^{p+q} := \bigoplus_{2i \geq p} (S^i \mathfrak{g}^* \otimes \mathcal{A}^q)^\mathfrak{g} \qquad [21]$$

Since second term in the equivariant differential [19] raises the filtration degree by 2, it follows that

$$E_2^{p,q} = (S^{p/2}\mathfrak{g}^*)^\mathfrak{g} \otimes H^q(\mathcal{A}) \qquad [22]$$

for $p$ even, $E_2^{p,q} = 0$ for $p$ odd. In fortunate cases, the spectral sequence collapses at the $E_2$-stage (see below).

## Equivariant de Rham Theory

We will now restrict ourselves to the case that $\mathcal{A} = \Omega(M)$ is the algebra of differential forms on a $G$-manifold, where $G$ is compact and connected.

**Theorem 18** (Equivariant de Rham theorem). *Suppose $G$ is a compact, connected Lie group, and that $M$ is a $G$-manifold. Then there is a canonical isomorphism*

$$H_G(M; \mathbb{R}) \cong H_\mathfrak{g}(\Omega(M)) \qquad [23]$$

*where the left-hand side is the equivariant cohomology as defined by the Borel construction.*

Motivated by this result, the notation can be changed slightly; write

$$\Omega_G(M) = (S\mathfrak{g}^* \otimes \Omega(M))^G \qquad [24]$$

for the Cartan complex of equivariant differential forms, and $\mathrm{d}_G$ for the equivariant differential [19].

**Remark 19** Theorem 18 fails, in general, for noncompact Lie groups $G$. A differential form model for the noncompact case was developed by Getzler (1990).

**Example 20** Let $(M, \omega)$ be a symplectic manifold, and $a: G \to \mathrm{Diff}(M)$ a Hamiltonian group action. That is, $a$ preserves the symplectic form, $a_g^*\omega = \omega$, and there exists an equivariant moment map $\Phi: M \to \mathfrak{g}^*$ such that $\iota_\xi \omega + \mathrm{d}\langle \Phi, \xi \rangle = 0$. Then the equivariant symplectic form $\omega_G(\xi) := \omega + \langle \Phi, \xi \rangle$ is equivariantly closed.

**Example 21** Let $G$ be a Lie group, and denote, respectively, by

$$\theta^{\mathrm{L}} = g^{-1}\mathrm{d}g \text{ and } \theta^{\mathrm{R}} = \mathrm{d}g g^{-1} \qquad [25]$$

the left- and right-invariant Maurer–Cartan forms. Suppose $\mathfrak{g} = \mathrm{Lie}(G)$ carries an invariant scalar product "·", and consider the closed 3-form

$$\phi = \frac{1}{12}\,\theta^L \cdot [\theta^L, \theta^L] \qquad [26]$$

Then

$$\phi_G(\xi) = \phi + \frac{1}{2}(\theta^L + \theta^R) \cdot \xi \qquad [27]$$

is a closed equivariant extension for the conjugation action of $G$. More generally, transgression gives explicit differential forms $\phi_j$ generating the cohomology ring $H(G) = (\wedge\mathfrak{g}^*)^G$. Closed equivariant extensions of these forms were obtained by Jeffrey (1995), using a construction of Bott–Shulman.

A $G$-manifold is called equivariantly formal if

$$H_G(M) = (S\mathfrak{g}^*)^G \otimes H(M) \qquad [28]$$

as an $(S\mathfrak{g}^*)^G$-module. Equivalently, this is the condition that the spectral sequence [22] for $H_G(M)$ collapses at the $E_2$-term. $M$ is equivariantly formal under any of the following conditions: (1) $H^q(M) = 0$ for $q$ odd, (2) the map $H_G(M) \to H(M)$ is onto, (3) M admits a $G$-invariant Morse function with only even indices, and (4) $M$ is a symplectic manifold and the $G$-action is Hamiltonian. (The last fact is a theorem due to Ginzburg and Kirwan.

**Example 22** The conjugation action of a compact Lie group is equivariantly formal, by criterion [2]. In this case, eqn [28] is an isomorphism of algebras.

It is important to note that eqn [28] is not an algebra isomorphism, in general. Already the rotation action of $G = \mathrm{U}(1)$ on $M = S^2$, discussed in Example 6, provides a counter-example.

**Theorem 23** (Injectivity). *Suppose $T$ is a compact torus, and $M$ is $T$-equivariantly formal. Then the pullback map $H_T(M) \to H_T(M^T)$ to the fixed point set is injective.*

Since the pullback map to the fixed point set is an algebra homomorphism, one can sometimes use this result to determine the algebra structure on $H_T(M)$: let $\alpha_r \in H(M)$ be generators of the ordinary cohomology algebra, and let $(\alpha_r)_T$ be equivariant extensions. Denote by $x_r \in H_T(M^T)$ the pullbacks of $(\alpha_r)_T$ to the fixed point set, and let $y_i$ be a basis of $\mathfrak{t}^*$, viewed as elements of $S\mathfrak{t}^* \subset H_T(M^T)$. Then $H_T(M)$ is isomorphic to the subalgebra of $H_T(M^T)$ generated by the $x_r$ and $y_j$.

The case of nonabelian compact groups $G$ may be reduced to maximal torus $T$ using the following result. Observe that for any $G$-manifold $M$, there is a natural action of the Weyl group $W = N(T)/T$ on $H_T(M)$.

**Theorem 24** *The natural restriction map*

$$H_G(M; \mathbb{R}) \to H_T(M; \mathbb{R})^W \qquad [29]$$

*onto the Weyl group invariants is an algebra isomorphism.*

**Remark 25** The Cartan complex [24] may be viewed as a small model for the differential forms on the infinite-dimensional space $M_G$. In the noncommutative case, there exists an even "smaller" Cartan model, with underlying complex $(S\mathfrak{g}^*)^G \otimes \Omega(M)^G$, involving only invariant differential forms on $M$ (see Alekseev and Meinrenken (2005) and Goresky, Kottwitz, and MacPherson (1998)).

## Equivariant Characteristic Forms

Let $G$ be a compact Lie group, and $E \to B$ a principal $G$-bundle with connection $\theta \in \Omega^1(E) \otimes \mathfrak{g}$. Suppose the principal $G$-action commutes with the action of a compact Lie group $K$ on $E$, and that $\theta$ is $K$-invariant. The $K$-equivariant curvature of $\theta$ is defined as follows:

$$F_K^\theta = \mathrm{d}_K\theta + \frac{1}{2}[\theta, \theta] \in \Omega_K^2(E) \otimes \mathfrak{g}$$

By the equivariant version of eqn [20], there is a canonical chain map

$$\Omega_{K\times G}(E) \to \Omega_K(B) \qquad [30]$$

defined by substituting the $K$-equivariant curvature for the $\mathfrak{g}$-variable, followed by horizontal projection with respect to $\theta$. The Cartan map [30] is homotopy inverse to the pullback map from $\Omega_K(B)$ to $\Omega_{K\times G}(B)$.

**Example 26** The complex $\Omega_{K\times G}(E)$ contains a subcomplex $(S\mathfrak{g}^*)^G$. The restriction of eqn [30] is the equivariant Chern–Weil map

$$(S\mathfrak{g}^*)^G \to \Omega_K(B) \qquad [31]$$

Forms in the image of eqn [31] are equivariantly closed; they are called the $K$-equivariant characteristic forms of $E$.

**Example 27** Similarly, if $\mathcal{V} \to B$ is a $K$-equivariant vector bundle with structure group $G \subset \mathrm{GL}(k)$, one defines the $K$-equivariant characteristic forms of $\mathcal{V}$ to be those of the corresponding bundle of $G$-frames in $\mathcal{V}$.

For instance, suppose $\mathcal{V}$ is an oriented $K$-equivariant vector bundle of even rank $k$, with an invariant metric and compatible connection. The Pfaffian defines an invariant polynomial on $\mathfrak{so}(k)$:

$$\zeta \mapsto \det{}^{1/2}(\zeta/2\pi) \qquad [32]$$

(equal to 0 if $k$ is odd). The $K$-equivariant characteristic form of degree $k$ on $B$ determined by eqn [32] is known as the equivariant Euler form

$$\mathrm{Eul}_K(\mathcal{V}) \in \Omega_K^k(B) \qquad [33]$$

Similarly, one defines equivariant Pontrjagin forms of $\mathcal{V}$, and (for Hermitian vector bundles) equivariant Chern forms.

**Example 28** Suppose $G$ is a maximal rank subgroup of the compact Lie group $K$. The bundle $K \to K/G$ admits a unique $K$-invariant connection. Hence, one obtains a canonical chain map $(S\mathfrak{g}^*)^G \to \Omega_K(K/G)$, realizing the isomorphism $H_K(K/G) \cong (S\mathfrak{g}^*)^G$. In particular, any $G$-invariant element of $\mathfrak{g}^*$ defines a closed $K$-equivariant 2-form on $K/G$. For instance, symplectic forms on coadjoint orbits are obtained in this way.

Suppose $M$ is a $G$-manifold, and let $Q = E \times_G M$ be the associated bundle. For any $K$-invariant connection on $E$, one obtains a chain map

$$\Omega_G(M) \to \Omega_{K\times G}(E \times M) \to \Omega_K(Q) \qquad [34]$$

by composing the pullback to $E \times M$ with the Cartan map for the principal bundle $E \times M \to Q$.

**Example 29** Suppose $(M, \omega)$ is a Hamiltonian $G$-manifold, with moment map $\Phi : M \to \mathfrak{g}^*$. The image of $\omega_G = \omega + \Phi$ under the map [34] defines a closed $K$-equivariant 2-form on $Q$. This construction is of importance in symplectic geometry, where it arises in the context of Sternberg's minimal coupling.

## Equivariant Thom Forms

Let $\pi : \mathcal{V} \to B$ be a $G$-equivariant oriented real vector bundle of rank $k$ over a compact base $B$. There is a canonical chain map, called fiber integration

$$\pi_* : \Omega^\bullet(\mathcal{V})_{\mathrm{cp}} \to \Omega^{\bullet-k}(B) \qquad [35]$$

where the subscript indicates "compact support." It is characterized by the following properties:

(1) for a form of degree $k$, the value of its fiber integral at $x \in B$ is equal to the integral over the fiber $\mathcal{V}_x$, and
(2)

$$\pi_*(\alpha \wedge \pi^*\beta) = \pi_*\alpha \wedge \beta \qquad [36]$$

for all $\alpha \in \Omega(\mathcal{V})_{cp}$ and $\beta \in \Omega(B)$. Fiber integration extends to $G$-equivariant differential forms, and commutes with the equivariant differential.

**Theorem 30** (Equivariant Thom isomorphism). *Fiber integration defines an isomorphism,*

$$H_G^{\bullet+k}(\mathcal{V})_{cp} \to H_G^\bullet(B) \qquad [37]$$

An equivariant Thom form for a $G$-vector bundle is a cocycle $\mathrm{Th}_G(\mathcal{V}) \in \Omega_G^k(\mathcal{V})_{cp}$, with the property,

$$\pi_*\mathrm{Th}_G(\mathcal{V}) = 1 \qquad [38]$$

Given $\mathrm{Th}_G(\mathcal{V})$, the inverse to eqn [37] is realized on the level of differential forms as

$$\Omega_G^\bullet(B) \to \Omega_G^{\bullet+k}(E), \qquad \alpha \mapsto \mathrm{Th}_G(\mathcal{V}) \wedge \pi^*\alpha \qquad [39]$$

A beautiful "universal" construction of Thom forms was obtained by Mathai and Quillen (1986). Using eqn [34], it suffices to describe an $\mathrm{SO}(k)$-equivariant Thom form for the trivial bundle $\mathbb{R}^k \to \{0\}$. Using multi-index notation for ordered subsets $I \subset \{1, \ldots, k\}$,

$$\mathrm{Th}_{\mathrm{SO}(k)}(\mathbb{R}^k)(\zeta) = \frac{\mathrm{e}^{-\|x\|^2}}{\pi^{k/2}} \sum_I \epsilon_I \det{}^{1/2}\left(\frac{\zeta_I}{2}\right)(\mathrm{d}x)^{I^c} \qquad [40]$$

Here the sum is over all subsets $I$ with $|I|$ even, and $I^c$ is the complement of $I$. The matrix $\zeta_I$ is obtained from $\zeta$ by deleting all rows and columns that are not in $I$, and $\det^{1/2}$ is defined as a Pfaffian. Finally, $\epsilon_I$ is the sign of the shuffle permutation defined by $I$, that is, $(\mathrm{d}x)^I(\mathrm{d}x)^{I^c} = \epsilon_I\,\mathrm{d}x_1 \cdots \mathrm{d}x_k$. As shown by Mathai and Quillen, the form [40] is equivariantly closed, and clearly eqn [38] holds since the top degree part is just a Gaussian. If $k$ is even, the Mathai–Quillen formula can also be written, on the open dense where $\zeta \in \mathfrak{so}(k)$ is invertible, as

$$\mathrm{Th}_{\mathrm{SO}(k)}(\mathbb{R}^k)(\zeta) = \det{}^{1/2}\left(\frac{\zeta}{2\pi}\right)\mathrm{e}^{-\|x\|^2 - \langle \mathrm{d}x, \zeta^{-1}(\mathrm{d}x)\rangle} \qquad [41]$$

The form $\mathrm{Th}_{\mathrm{SO}(k)}(\mathbb{R}^k)$ given by these formulas does not have compact support, but is rapidly decreasing at infinity. One obtains a compactly supported Thom form, by applying an $SO(k)$-equivariant diffeomorphism from $\mathbb{R}^k$ onto some open ball of finite radius.

Note that the pullback of eqn [40] to the origin is equal to $\det^{1/2}(\zeta/2\pi)$ (equal to 0 if $k$ is odd). This implies:

**Theorem 31** *Let $\iota : B \to \mathcal{V}$ denote the inclusion of the zero section. Then*

$$\iota^*\mathrm{Th}_G(\mathcal{V}) = \mathrm{Eul}_G(\mathcal{V}) \qquad [42]$$

*where $\mathrm{Eul}_G(\mathcal{V}) \in \Omega_G^k(B)$ is the equivariant Euler form.*

Suppose, $M$ is a $G$-manifold, and $S$ a closed $G$-invariant submanifold with oriented normal

bundle $\nu_S$. Choose a $G$-equivariant tubular neighborhood embedding

$$\nu_S \to U \subset M \qquad [43]$$

and let $\mathrm{PD}_G(S) \in \Omega_G(M)_{cp}$ be the image of $\mathrm{Th}_G(\mathcal{V})$ under this embedding. The form $\mathrm{PD}_G(S)$ has the property

$$\int_M \mathrm{PD}_G(S) \wedge \alpha = \int_S \iota_S^* \alpha \qquad [44]$$

for all closed equivariant forms $\alpha \in \Omega_G(M)$. It is called an "equivariant Poincaré dual" of $S$. By construction, the pullback to $S$ is the equivariant Euler form:

$$\iota_S^* \mathrm{PD}_G(S) = \mathrm{Eul}_G(\nu_S) \qquad [45]$$

Equivariant Poincaré duality takes transversal intersections of $G$-manifolds to wedge products, similar to the nonequivariant case.

**Remark 32** In general, the $(S\mathfrak{g}^*)^G$-submodule generated by Poincaré duals of $G$-invariant submanifolds is strictly smaller than $H_G(M)$. In this sense, the terminology "duality" is misleading.

## Localization Theorem

In this section, $T$ will denote a torus. Suppose $M$ is a compact oriented $T$-manifold. For any component $F$ of the fixed point set of $T$, the action of $T$ on $\nu_F$ fixes only the zero section $F$. This implies that the normal bundle $\nu_F$ has even rank and is orientable. Fix an orientation, and give $F$ the induced orientation.

Since $T$ is compact, the list of stabilizer groups of points in $M$ is finite. Call $\xi \in \mathfrak{t}$ generic if it is not in the Lie algebra of any of these stabilizers, other than $T$ itself. In this case, value $\mathrm{Eul}_T(\nu_F, \xi)$ of the equivariant Euler form is invertible as an element of $\Omega(F)$.

**Theorem 33** (Integration formula). *Suppose $M$ is a compact oriented $T$-manifold, where $T$ is a torus. Let $\alpha \in \Omega_T(M)$ be a closed equivariant form, and let $\xi \in \mathfrak{t}$ be generic. Then*

$$\int_M \alpha(\xi) = \sum_F \int_F \frac{\iota_F^* \alpha(\xi)}{\mathrm{Eul}_T(\nu_F, \xi)} \qquad [46]$$

*where the sum is over the connected components of the fixed point set.*

Rather than fixing $\xi$, one can also view eqn (46) as an equality of rational functions of $\xi \in \mathfrak{t}$.

**Remark 34** The integration formula was obtained by Berline and Vergne (1983), based on ideas of Bott (1967). The topological counterpart, as a "localization principle," was proved independently by Atiyah and Bott (1984). More abstract versions of the localization theorem in equivariant cohomology had been proved earlier by Borel, Chiang–Skjelbred and others.

**Remark 35** If $\alpha = \mathrm{PD}_T(F) \wedge \beta$, where $\beta$ is equivariantly closed, the integration formula is immediate from the property [44] of Poincaré duals. The essence of the proof is to reduce to this case.

**Remark 36** The localization contributions are particularly nice if $F = \{p\}$ is isolated (which can only happen if $\dim M$ is even). In this case, $\iota_F^* \alpha(\xi)$ is simply the value of the function $\alpha_{[0]}(\xi)$ at $p$. For the Euler form, one has

$$\mathrm{Eul}(\nu_F, \xi) = (-1)^{\dim M/2} \prod \langle \mu_j(p), \xi \rangle \qquad [47]$$

where $\mu_j(p) \in \mathfrak{t}^*$ are the (real) weights of the action on the tangent space $T_p M$. (Here we have chosen an isomorphism $T_p M \cong \mathbb{C}^l$ compatible with the orientation.) Hence, if all fixed points are isolated,

$$\int_M \alpha(\xi) = (-1)^{\dim M/2} \sum_p \frac{\alpha_{[0]}(\xi)(p)}{\prod_j \langle \mu_j(p), \xi \rangle} \qquad [48]$$

**Example 37** Let $M$ be a compact oriented manifold, and $e(M) = \int_M \mathrm{Eul}(TM)$ its Euler characteristic. Suppose a torus $T$ acts on $M$. Then

$$e(M) = \sum_F e(F) \qquad [49]$$

where the sum is over the fixed point set of $T$. This follows from the integral of the equivariant Euler form $\alpha(\xi) = \mathrm{Eul}_T(M, \xi)$, by letting $\xi \to 0$ in the localization formula. In particular, if $M$ admits a circle action with isolated fixed points, the number of fixed points is equal to the Euler characteristic.

In a similar fashion, the localization formula gives interesting expressions for other characteristic numbers of manifolds and vector bundles, in the presence of a circle action. Some of these formulas were discovered prior to the localization formula, see in particular Bott (1967).

**Example 38** In this example, we show that for a simply connected, simple Lie group $G$ the 3-form $\phi \in \Omega^3(G)$ defined in eqn [26] is integral, provided "·" is taken to be the basic inner product (for which the length squared of the short coroots equals 2). Since any such $G$ is known to contain an SU(2) subgroup, it suffices to prove this for $G = \mathrm{SU}(2)$. Consider the conjugation action of the maximal torus $T \cong \mathrm{U}(1)$, consisting of diagonal matrices. The fixed point set for this action is $T$ itself. The normal

bundle $\nu_F$ is trivial, with $T$ acting on the fiber $\mathfrak{g}/\mathfrak{t}$ by the negative root $-\alpha$. Hence, $\mathrm{Eul}(\nu_F, \xi) = \langle \alpha, \xi \rangle$. Let $\breve{\alpha} \in \mathfrak{t}$ be the coroot, defined by $\langle \alpha, \breve{\alpha} \rangle = 2$. By definition, $\langle \alpha, \breve{\alpha} \rangle = 2$. Let us integrate the $T$-equivariant extension $\phi_T(\xi)$ (cf. [27]). Its pullback to $T$ is $\theta^T \cdot \xi$, where $\theta^T \in \Omega(T, \mathfrak{t})$ is the Maurer–Cartan form. The integral of $\theta^T$ is a generator of the integral lattice, that is, it equals $\breve{\alpha}$. Thus,

$$\int_{\mathrm{SU}(2)} \phi_T(\xi) = \frac{\int_T \theta^T \cdot \xi}{\langle \alpha, \xi \rangle} = \frac{\breve{\alpha} \cdot \xi}{\langle \alpha, \xi \rangle} = 1 \qquad [50]$$

## Duistermaat–Heckman Formulas

In this section, we discuss the Duistermaat–Heckman formula, for the case of isolated fixed points. Let $T$ be a torus, and $(M, \omega)$ a compact Hamiltonian $T$-space, with moment map $\Phi : M \to \mathfrak{t}^*$. Denote by $\omega_T = \omega + \Phi$ the equivariant extension of $\omega$. Assuming isolated fixed points, the localization formula gives, for all integers $k \geq 0$,

$$\int_M (\omega + \langle \Phi, \xi \rangle)^k = (-1)^n \sum_p \frac{\langle \Phi(p), \xi \rangle^k}{\prod_j \langle \mu_j(p), \xi \rangle} \qquad [51]$$

where $n = (1/2) \dim M$. Note that both sides are homogeneous of degree $k - n$ in $\xi$, but the terms on the right-hand side are only rational functions while the left-hand side is a polynomial. For $k = n$, both sides are independent of $\xi$, and compute the integral $\int_M \omega^n$. For $k < n$, the integral [51] is zero, and the cancellation of the terms on the right-hand side gives identities among the weights $\mu_j(p)$. Equation [51] also implies

$$\int_M \mathrm{e}^{\omega + \langle \Phi, \xi \rangle} = (-1)^n \sum_p \frac{\mathrm{e}^{\langle \Phi(p), \xi \rangle}}{\prod_j \langle \mu_j(p), \xi \rangle} \qquad [52]$$

Assume, in particular, that $T = \mathrm{U}(1)$, and let $\xi = t\xi_0$, where $\xi_0$ is the generator of the integral lattice in $\mathfrak{t}$. Identify $\mathfrak{t} \cong \mathbb{R}$ in such a way that $\xi_0$ corresponds to $1 \in \mathbb{R}$. Then $H = \langle \Phi, \xi_0 \rangle$ is a Hamiltonian function with periodic flow. Write $a_j(p) = \langle \mu_j(p), \xi_0 \rangle \in \mathbb{Z}$. Then eqn [52] reads

$$\int_M \mathrm{e}^{tH} \frac{\omega^n}{n!} = \frac{(-1)^n}{t^n} \sum_p \frac{\mathrm{e}^{tH(p)}}{\prod_j a_j(p)} \qquad [53]$$

The right-hand side of eqn [53] is the leading term for the stationary phase approximation of the integral on the left. For this reason, eqn [52] is known as the Duistermaat–Heckman exact stationary phase theorem.

Formula [52] has the following consequence for the push-forward of the Liouville measure under the moment map, the so-called Duistermaat–Heckman measure $H_*(\omega^n/n!)$. Let $\Theta$ be the Heaviside measure (i.e., the characteristic measure of the positive real axis).

**Theorem 39** (Duistermaat–Heckman). *The push-forward $H_*(\omega^n/n!)$ is piecewise polynomial measure of degree $n-1$, with singularities at the set of all $H(p)$ for fixed points $p$ of the action. One has the formula*

$$H_*\left(\frac{\omega^n}{n!}\right) = \sum_p \frac{(\lambda - H(p))^{n-1}}{\prod_j a_j(p)} \Theta(\lambda - H(p)) \qquad [54]$$

**Proof** It is enough to show that the Laplace transforms of the two sides are equal. Multiplying by $\mathrm{e}^{t\lambda}$ and integrating over $\lambda$ (take $t < 0$ to ensure convergence of the integral), the resulting identity is just eqn [53]. □

**Remark 40** The theorem generalizes to Hamiltonian actions of higher-rank tori, and also to nonisolated fixed points. See the paper by Guillemin, Lerman, and Sternberg (1988) for a detailed discussion of this formula and of its "quantum analog."

## Equivariant Index Theory

By definition, the Cartan model consists of equivariant forms $\alpha(\xi)$ with polynomial dependence on the equivariant parameter $\xi$. However, the integration formula holds in much greater generality. For instance, one may consider generalized Cartan complexes (Kumar and Vergne 1993). Here the parameter $\xi$ varies in some invariant open subset of $\mathfrak{g}$, and the polynomial dependence is replaced by smooth dependence. The use of these more general complexes in equivariant index theory was pioneered by Berline and Vergne (1992).

Assume that $M$ is an even-dimensional, compact oriented Riemannian manifold, equipped with a Spin-c structure. According to the Atiyah–Singer theorem, the index of the corresponding Dirac operator $D$ is given by the formula

$$\mathrm{ind}(D) = \int_M \hat{A}(M) \mathrm{e}^{c/2} \qquad [55]$$

Here $c$ is the curvature 2-form of the complex line bundle associated to the Spin-c structure, and $\hat{A}(M)$ is the $\hat{A}$-form. Recall that $\hat{A}(M)$ is obtained by substituting the curvature form in the formal power series expansion of the function $\hat{A}(x) = \det^{1/2}((x/2)/\sinh(x/2))$ on $\mathfrak{so}(n)$.

Suppose now that a compact, connected Lie group $G$ acts on $M$ by isometries, and that the action lifts to the Spin-c bundle. Replacing curvatures with equivariant curvatures, one defines the equivariant form $\hat{A}(M)(\xi)$ and the form $c(\xi)$. Note that $\hat{A}(\xi)$ is only

defined for $\xi$ in a sufficiently small neighborhood of 0, since the function $\hat{A}(x)$ is not analytic for all $x$.

The $G$-index of the equivariant Spin-c Dirac operator is a virtual character $g \mapsto \text{ind}(D)(g)$ of the group $G$. For $g = \exp \xi$ sufficiently small, it is given by the formula

$$\text{ind}(D)(\exp \xi) = \int_M \hat{A}(M)(\xi) e^{c(\xi)/2} \qquad [56]$$

For $\xi$ sufficiently small, the fixed point set of $g$ coincides with the set of zeroes of the vector field $a_\xi$. The localization formula reproduces the Atiyah–Segal formula for $\text{ind}(D)(g)$, as an integral over $M^g$.

Berline and Vergne (1996) gave similar formulas for the equivariant index of any $G$-equivariant elliptic operator, and more generally for operators that are transversally elliptic in the sense of Atiyah.

*See also:* Cohomology Theories; Compact Groups and Their Representations; Hamiltonian Group Actions; *K*-theory; Lie Groups: General Theory; Mathai–Quillen Formalism; Path-Integrals in Noncommutative Geometry; Stationary Phase Approximation.

## Further Reading

Alekseev A and Meinrenken E (2000) The non-commutative Weil algebra. *Inventiones Mathematical* 139: 135–172.
Alekseev A and Meinrenken E (2006) *Equivariant Cohomology and the Maurer–Cartan Equation*.
Atiyah MF and Bott R (1984) The moment map and equivariant cohomology. *Topology* 23(1): 1–28.
Berline N, Getzler E, and Vergne M (1992) *Heat Kernels and Dirac Operators*. Grundlehren der Mathematischen Wissenschaften, vol. 298. Berlin: Springer.
Berline N and Vergne M (1983) Z'ero d'un champ de vecteurs et classes caractéristiques équivariantes. *Duke Mathematics Journal* 50: 539–549.
Berline N and Vergne M (1996) L'indice équivariant des opérateurs transversalement elliptiques. *Inventiones Mathematical* 124(1–3): 51–101.
Borel A (1960) *Seminar on Transformation Groups* (with contributions by Bredon G, Floyd EE, Montgomery D, and Palais R), *Annals of Mathematics Studies*, No. 46. Princeton: Princeton University Press.
Bott R (1967) Vector fields and characteristic numbers. *Michigan Mathematical Journal* 14: 231–244.
Cartan H (1950) *La transgression dans un groupe de Lie et dans un fibré principal*, Colloque de topologie (espaces fibrés) (Bruxelles), Centre belge de recherches mathématiques, Georges Thone, Liège, Masson et Cie., Paris, pp. 73–81.
Cartan H (1950) *Notions d'algèbre différentielle*; *application aux groupes de Lie et aux variétés où opère un groupe de Lie.*, Colloque de topologie (espaces fibrés) (Bruxelles), Georges Thone, Liège, Masson et Cie., Paris.
Duistermaat JJ and Heckman GJ (1982) On the variation in the cohomology of the symplectic form of the reduced phase space. *Inventiones Mathematical* 69: 259–268.
Getzler E (1994) The equivariant Chern character for non-compact Lie groups. *Advances in Mathematics* 109(1): 88–107.
Goresky M, Kottwitz R, and MacPherson R (1998) Equivariant cohomology, Koszul duality, and the localization theorem. *Inventiones Mathematical* 131(1): 25–83.
Guillemin V, Lerman E, and Sternberg S (1988) On the Kostant multiplicity formula. *Journal of Geometry and Physics* 5(4): 721–750.
Guillemin V and Sternberg S (1999) *Supersymmetry and Equivariant de Rham Theory*. Springer.
Husemoller D (1994) *Fibre Bundles*, Graduate Texts in Mathematics. 3rd edn., vol. 20. New York: Springer.
Jeffrey L (1995) Group cohomology construction of the cohomology of moduli spaces of flat connections on 2-manifolds. *Duke Mathematical Journal* 77: 407–429.
Kumar S and Vergne M (1993) Equivariant cohomology with generalized coefficients. *Astérisque* 215: 109–204.
Mathai V and Quillen D (1986) Superconnections, Thom classes, and equivariant differential forms. *Topology* 25: 85–106.
Milnor J (1956) Construction of universal bundles. II. *Annals of Mathematics* 63(2): 430–436.

# Ergodic Theory

**M Yuri**, Hokkaido University, Sapporo, Japan



## Introduction

The ergodic theory was developed from the following Poincaré's work, which served as the starting point in the measure theory of dynamical systems in the sense of the study of the properties of motions that take place at "almost all" initial states of a system: let $(X, \mathcal{B}, \mu)$ be a probability space and a transformation $T: X \to X$ preserve $\mu$ (i.e., $\mu(T^{-1}A) = \mu(A)$ for any $A \in \mathcal{B}$). If $\mu(A) > 0$, then for almost all points $x \in A$ the orbit $\{T^n x\}_{n \geq 0}$ returns to $A$ infinitely more often (the Poincaré–Caratheodory recurrence theorem). The main theme of the ergodic theory is to know whether averages of quantities generated in a stationary manner converge. In the classical situation the stationary is described by a measure-preserving transformation $T$, and one considers averages taken along a sequence $f, fT, fT^2, \ldots$ for integrable $f$. This corresponds to the probabilistic concept of stationarity. Hence, traditionally, the ergodic theory is the qualitative study of iterates of an individual transformation, of one parameter flow of transformations (such as that obtained from the solution of an autonomous ordinary differential equation). We should note that an important purpose behind this theory is to verify significant facts from a statistical point of view (e.g., the law of large numbers, convergence to limit distributions). The oldest branch

of this theory is the study of ergodic theorems. It was started in 1931 by Birkhoff (1931) and von Neumann (1932), having its origins in statistical mechanics. More specifically, the central notion is that of ergodicity, which is intended to capture the idea that a flow is "random" or "chaotic." In dealing with the motion of molecules, Boltzmann and Gibbs made such hypotheses from the beginning. One of the earliest precise definitions of randomness of a dynamical system was "minimality": the orbit of almost every point is dense. In order to describe such phenomena in measure-theoretical setting, von Neumann and Birkhoff required the stronger assumption of ergodicity as follows. Let $(X, \mathcal{B}, \mu)$ be a measure space and $F_t$ a measurable flow on $X$. We call $F_t$ ergodic if the only invariant measurable sets are $\emptyset$ or all of $X$. Here, the invariance of the set $A$ means that $F_t(A) = A$ for all $t \in \boldsymbol{R}$ and we agree to write $A = B$ if $A$ and $B$ differ by a null set with respect to $\mu$. Note that ergodicity implies minimality if we are on a second countable Borel space. A function $f : X \to \boldsymbol{R}$ will be called a "constant of the motion" iff $f \circ F_t = f$ a.e. for each $t \in \boldsymbol{R}$. Then we see that a flow $F_t$ on $X$ is ergodic iff the only constants of the motion are constant a.e. In case of a measurable transformation $T$ on $X$, the invariance of the set $A$ means that $T^{-1}A = A$, and the measurable function $f$ is called invariant if $f \circ T = f$ a.e. Then we call $T$ ergodic provided if $A$ is invariant then either $\mu(A) = 0$ or $\mu(A) = 1$; equivalently, any invariant function is constant a.e. (Cornfeld *et al.* 1982). The most basic example where ergodicity can be verified is the following: if $M$ is a compact Riemannian and has negative sectional curvatures at each point, then the geodesic flow on each sphere bundle is ergodic (Hopf–Hadamard). In general, verifying ergodicity can still be very difficult. In the Hamiltonian case, the first step is to pass to an energy surface. For example, Sinai (1970) shows that one has ergodicity on an energy surface of a classical model for molecular motion, that is, a collection of hard spheres in a box.

## Ergodic Theorems

Koopman (1931) published the following significant observation: if $T$ is an invertible measure-preserving transformation of a measure space $(X, \mathcal{B}, \mu)$, then the operator $U$, defined on $L^2(X, \mathcal{B}, \mu)$ by $Uf(x) := f(Tx)$, is unitary. Thus, the association of $U$ with $T$ replaces a nonlinear finite-dimensional problem with a linear infinite-dimensional one. Then von Neumann (1932) showed an intimate connection between measure-preserving transformations and unitary operators (the mean ergodic theorem): let $U$ be a unitary operator on a Hilbert space $\mathcal{H}$. Denote by $P$ the orthogonal projection onto the subspace $\mathcal{H}_0 := \{f \in \mathcal{H} | Uf = f\}$. For any $f \in \mathcal{H}$, one has

$$\lim_{N\to\infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} U^n f - Pf \right\|_{\mathcal{H}} = 0$$

As a corollary, one can show that if $T : X \to X$ is an ergodic measure-preserving transformation on a probability space $(X, \mathcal{B}, \mu)$ then, for any $f \in L^1(X, \mathcal{B}, \mu)$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) = \frac{1}{\mu(X)} \int_X f \, d\mu$$

in $L^1$-norm. We also know that $T$ is ergodic if and only if $U$ has 1 as a simple eigenvalue. In the case of a continuous invertible process, the setting is the following. Let $M$ be a manifold and $\Omega$ a volume on $M$, with $\mu_\Omega$ the corresponding measure. If $F_t$ is a volume-preserving flow on $M$, then $F_t$ induces a linear one-parameter group of isometries on $\mathcal{H} = L^2(M, \mu_\Omega)$ by $U_t(f) = f \circ F_{-t}$. Then $U_t$ has 1 as a simple eigenvalue for all $t$ if and only if $F_t$ is ergodic.

On the other hand, Birkhoff (1931) proved the following almost everywhere statement (the pointwise ergodic theorem): for any $f \in L^1(X, \mathcal{B}, \mu)$, there exists a function $\bar{f} \in L^1(X, \mathcal{B}, \mu)$ such that for $\mu$-a.e. $x$, $\bar{f}T(x) = \bar{f}(x)$ and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) = \bar{f}(x)$$

In particular, if $T$ is ergodic then $\mu$-a.e. $x, \bar{f}(x) = \int_X f d\mu$. Thus, the Birkhoff theorem allows one to prove the ergodic hypothesis by Boltzmann–Gibbs, that is, the space average of an observable function coincides with its time averages almost everywhere, and guarantees the existence, for almost everywhere, of the mean number of occurrences in any measurable set. On the other hand, physical meanings of the mean ergodic theorem can be explained as follows. We now turn to one-parameter flow of transformations. In order to study continuous averages

$$\frac{1}{t} \int_0^t f(F_s x) ds$$

fix some $s_0 \in \boldsymbol{R}$ and consider the averages of the form

$$\frac{1}{N} \sum_{n=0}^{N-1} f(T^n x)$$

where $T = F_{s_0}$. In reality, the measurements can be done only approximately at times $t = 0, 1, \ldots, N-1$, and it is natural to consider the perturbed averages

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^{n+\delta_n}x)$$

where $\{\delta_n\}_{n\in N}$ is an independent random sequence in a small interval $(-\epsilon, \epsilon)$. Assuming that $T = F_{s_0}$ is ergodic, we would like to know whether for large $N$, the averages

$$\frac{1}{N}\sum_{n=0}^{N-1} f(T^{n+\delta_n}x)$$

are close to

$$\int_X f(x)\mathrm{d}\mu(x)$$

The answer to this question is satisfactory if one is concerned with norm convergence (see, e.g., Bergelson *et al.* (1994)).

## Induced Transformations and Tower Constructions

Suppose $T$ is a measure-preserving transformation on a probability space $(X, \mathcal{B}, \mu)$ and $A \in \mathcal{B}$ with $\mu(A) > 0$. Let us transform $A$ into a space with normalized measure by choosing the $\sigma$-algebra $\mathcal{B}_A$ consisting of all subsets $E \subset A, E \in \mathcal{B}$ and setting $\mu_A(E) = \mu(A \cap E)/\mu(A)$. Let $R_A : A \to \mathbf{N} \cup \{\infty\}$ be the "first return function," that is, $R_A(x) := \inf\{n \in \mathbf{N} \mid T^n x \in A\}$. Then it follows from the Poincaré reccurrence theorem that $\mu_A(\{x \in A \mid R_A(x) < \infty\} = 1$. Define $T_A : \{x \in A \mid R_A(x) < \infty\} \to A$ by $T_A x := T^{R_A(x)}x$, which is called the "induced transformation" over $A$ (constructed from $T$). For each $n \in \mathbf{N}$ we define $A_n := \{x \in A \mid R_A(x) = n\}$. Then for every $E \in \mathcal{B}_A$ we see that $T_A{}^{-1}E = \bigcup_{n=1}^{\infty} T^{-n}(A_n \cap E)$. Hence, if $T$ is invertible, then we have immediately $\mu_A(T_A{}^{-1}E) = \mu_A(E)$; thus, $\mu_A$ is invariant under $T_A$. Even if $T$ is noninvertible, since for every $k \geq 1$ the equality,

$$\begin{aligned}\mu\left(\bigcup_{j=0}^{k-1} T^{-j}A^{\mathrm{c}} \cap T^{-k}(A \cap E)\right) &= \mu(A_{k+1} \cap T^{-(k+1)}E) \\ &\quad + \mu\left(\bigcup_{j=0}^{k} T^{-j}A^{\mathrm{c}} \cap T^{-(k+1)}(A \cap E)\right)\end{aligned}$$

holds, we have $\mu(E) = \sum_{k=1}^{\infty} \mu(A_k \cap T^{-k}E) = \mu(T_A{}^{-1}E)$, which allows us to see that $T_A$ preserves $\mu_A$. We note that for every $E \in \mathcal{B}_A$ with $\mu(E) > 0$, $\{n \in \mathbf{N} \mid T^n x \in E\} = \{n \in \mathbf{N} \mid T_A{}^n x \in E\}$. Therefore, for a.e. $x \in E$, $\sum_{n=1}^{\infty} 1_E T^n(x) = \sum_{n=1}^{\infty} 1_E T_A{}^n(x)$. This equality allows us to see that if $(T, \mu)$ is ergodic then $(T_A, \mu_A)$ is ergodic. Indeed, suppose $T_A{}^{-1}E = E$ and $\mu(A \cap E^{\mathrm{c}}) > 0$. Then for $x \in A \cap E^{\mathrm{c}}$, we have $\sum_{n=1}^{\infty} 1_E T_A{}^n(x) = 0$. On the other hand, as $E \subset \bigcup_{n=1}^{\infty} T^{-n}E \pmod{\mu}$, $\bigcup_{n=1}^{\infty} T^{-n}E = \bigcup_{n=0}^{\infty} T^{-n}E$ is a $T$-invariant set. Hence, ergodicity of $(T, \mu)$ allows us to see that $\bigcup_{n=1}^{\infty} T^{-n}E = X$ ($\mu$ mod 0), which implies $\sum_{n=1}^{\infty} 1_E T^n(x) = \infty$. In the case when $T$ is invertible, we can write $\int_A R_A \,\mathrm{d}\mu = \mu(\bigcup_{n\geq 0} T^n A)$, so that Kac's formula (Darling and Kac 1957):

$$\int_A R_A \,\mathrm{d}\mu_A = \mu(A)^{-1}$$

is valid when $\bigcup_{n\geq 0} T^n A = X \pmod{\mu}$. In particular, $\mu(\bigcup_{n\geq 0} T^n A) = 1$ if $T$ is ergodic. The key to establish the Kac formula is to show that $T^i A_k (0 \geq i \geq k-1, k \geq 1)$ are pairwise disjoint. This property holds when $T$ is invertible. On the other hand, in the case when $T$ is noninvertible, if $\bigcup_{n=0}^{\infty} T^{-n}A = X(\mu \bmod 0)$ then we can establish, for every $E \in \mathcal{B}$,

$$\mu(E) = \mu(A)\int_A \sum_{h=0}^{R_A(x)-1} 1_E T^h(x)\,\mathrm{d}\mu_A(x) \qquad [1]$$

by noting that the following equality holds for all $n \geq 1$:

$$\begin{aligned}\mu(E) = &\sum_{k=1}^{n} \mu|_A\left(A \cap \bigcap_{h=1}^{k} T^{-h}A^c \cap T^{-k}A\right) \\ &+ \mu|_A(A \cap E) + \mu\left(\left(\bigcup_{j=0}^{n} T^{-j}A\right)^{\mathrm{c}} \cap T^{-n}E\right)\end{aligned}$$

Then choosing $E = X$ allows one to establish the Kac formula. As we have observed in the above, the assumption that $\bigcup_{n=0}^{\infty} T^{-n}A = X(\mu \bmod 0)$ is automatically satisfied if $(T, \mu)$ is ergodic. Conversely, if $(T_A, \mu_A)$ is ergodic and $\bigcup_{n=1}^{\infty} T^{-n}A \pmod{\mu}$ holds, then $(T, \mu)$ is ergodic. We should remark that the formula [1] allows one to obtain a $T$-invariant measure when a $T_A$-invariant measure $\mu_A$ is obtained previously. Even if $R_A$ is nonintegrable, we may have a $\sigma$-finite infinite invariant measure. Then if $\mu_A$ is ergodic, $\mu$ obtained by [1] is still ergodic (i.e., $T^{-1}E = E$ implies that $\mu(E) = 0$ or $\mu(E^c) = 0$) under the assumption that $\bigcup_{n=1}^{\infty} T^{-n}A = X \pmod{\mu}$ (cf. Aaronson (1997)). In particular, the recent progress in the study of nonhyperbolic systems strongly depends on such constructions of induced maps over hyperbolic regions. More specifically, if one can find a subset $A$ over which the induced map possesses an

invariant measure satisfying nice statistical properties, then the formula [1] may give a $\sigma$-finite invariant measure $\mu$ for the original map $T$ which reflects the statistical properties of the induced system. The fundamental problem in the study of nonhyperbolic phenomena arising from complex systems is to clarify how to predict statistical properties of nonhyperbolic systems $(T, \mu)$ by using those of induced systems $(T_A, \mu_A)$ over hyperbolic regions. We should claim that induced maps are well defined over positive-measure sets with respect to a reference measure $\nu$ that is "conservative." Here conservativity of $(T, \nu)$ implies that there are no wandering sets of positive measure with respect to $\nu$. In many cases, the reference measures are physical measures (e.g., Lebesgue measures, conformal measures) which satisfy nonsingularity with respect to $T$. Here nonsingularity of $\nu$ means that $\nu T^{-1} \sim \nu$. Then as long as we obtain a $T_A$-invariant measure $\mu_A$ which is equivalent to $\nu | A$, the formula [1] may give us a $T$-invariant $\sigma$-finite measure which is equivalent to $\nu$.

At the end of this section, we will explain that the folmula [1] can be obtained via Rohlin tower (Kakutani's skyscraper) in the case when $T$ is invertible. This tower construction is a dual construction to the construction of induced transformations. Assuming that we are given an invertible transformation $T$ of the measure space $(X, \mathcal{B}, \mu)$, consider the measurable integer-valued positive function $f \in L^1(X, \mathcal{B}, \mu)$. By using this function, construct a new measure space $X^f$, whose points are of the form $(x, i)$, where $x \in X, 1 \le i \le f(x)$ and $i$ is an integer. The $\sigma$-algebra $\mathcal{B}^f$ of measurable sets in $X^f$ is constructed in an obvious way. The measure $\mu^f$ is defined as follows: for any subset of the form $(A, i), A \in \mathcal{B}$ we put

$$\mu^f((A, i)) := \frac{\mu(A)}{\int_X f \, d\mu}$$

Let

$$T^f(x, i) = \begin{cases} (x, i+1) & \text{if } i+1 \le f(x) \\ (Tx, 1) & \text{if } i+1 > f(x) \end{cases}$$

It is easy to see that $T^f$ preserves $\mu^f$. The space can naturally be visualized as a tower whose foundation is the space $X$ and which has $f(x)$ floors over the point $x \in X$. The space $X$ is identified with the set of points $(x, 1)$. We see that $T = (T^f)_X$ and the construction of $(X^f, T^f)$ is called the Rohlin tower over $X$. Let $T$ be an invertible measure-preserving transformation on a probability space $(X, \mathcal{B}, \mu)$ and $A \in \mathcal{B}$ with $\mu(A) > 0$. Suppose that $X = \bigcup_{n=0}^{\infty} T^n A \pmod{\mu}$. Then $T$ is represented as the Rohlin tower $(A^{R_A}, \mathcal{B}^{R_A}, (\mu_A)^{R_A})$ over $A$ as follows. We define $p : (A^{R_A}, \mathcal{B}^{R_A}, (\mu_A)^{R_A}) \to (X, \mathcal{B}, \mu)$ by $p(x, i) := T^i x$. Then $p$ is an isomorphism satisfying $p(T^{R_A})_A = Tp$ (almost everywhere). Moreover, we can verify that $(\mu_A)^{R_A} p^{-1} = \mu$ by assuming ergodicity of $\mu$. This is because $\forall E \in \mathcal{B}$ we have

$$\begin{aligned} &\left( \bigcup_{n=0}^{\infty} T^n A \right) \cap E \\ &= \bigcup_{n=1}^{\infty} \bigcup_{i=0}^{n-1} p((A_n \cap T^{-i} E) \times \{i\}) \end{aligned}$$

so that

$$\begin{aligned} (\mu_A)^{R_A}(p^{-1}E) &= \sum_{n=1}^{\infty} \sum_{i=0}^{n-1} \frac{\mu_A(A_n \cap T^{-i}E)}{\int_A R_A \, d\mu_A} \\ &= \mu(A) \int_A \sum_{h=0}^{R_A(x)-1} 1_E T^h(x) \, d\mu_A(x) \end{aligned}$$

On the other hand, in the case when $T$ is noninvertible, the formula [1] is not necessarily obtained by any tower construction, except in very special cases. For example, even if $T$ is not invertible, the tower construction is valid if $T|_A$ and $T|_{A^c}$ are one-to-one and $TA = X$.

## Convergence to Equilibrium States and Mixing Properties

Let $T : X \to X$ be a measure-preserving transformation on a probability space $(X, \mathcal{B}, \mu)$. We call $T$ to be "weak mixing" if for any $A, B \in \mathcal{B}$

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \mu(T^{-n} A \cap B) - \mu(A)\mu(B) \right| = 0$$

The weak-mixing property of $(T, \mu)$ can be represented by; $\forall f, g \in L^2(X, \mathcal{B}, \mu)$

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \left| \int_X (f T^n) g \, d\mu - \int_X f \, d\mu \int_X g \, d\mu \right| = 0$$

and this is equivalent to the ergodicity of $(T \times T, \mu \times \mu)$. Moreover, $(T, \mu)$ is weak mixing if and only if the unitary operator $U : \mathcal{H} \to \mathcal{H}$ defined by $Uf(x) = f(Tx)$ has no eigenfunctions that are not constants ($\mu$ mod 0). We say that the operator $U$ has continuous spectrum if there are no eigenvectors. If $\mathcal{H}$ is the closure of the linear span of the eigenvectors, then we say that the operator $U$ has pure point spectrum. The weak-mixing property of $(T, \mu)$ just implies that $U$ restricted on the orthonormal subspace of the subspace consisting of

constant functions has continuous spectrum. We recall that if $U$ has one as a simple eigenvalue then $T$ is ergodic. Additionally, if there are no other eigenvalues, then $T$ is weakly mixing. Hence, if $T$ is weak mixing, then it is necessarily ergodic. The next property corresponds to the term "relaxation" in physics literature which is used to describe processes under which the system passes to a certain stationary state independently of its original state. We call $T$ (strong) mixing if for any $A, B \in \mathcal{B}$

$$\lim_{n\to\infty} \mu(T^{-n}A \cap B) = \mu(A)\mu(B)$$

Then $(T, \mu)$ is (strong) mixing if and only if for any $f, g \in L^2(X, \mathcal{B}, \mu)$

$$\lim_{n\to\infty} \int_X (fT^n) g \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu \int_X g \, \mathrm{d}\mu$$

and mixing is necessarily weak mixing. Moreover, for any probability measure $\nu$ absolutely continuous with respect to $\mu$, one can show that $\lim_{n\to\infty} \nu(T^{-n}A) = \mu(A)$ for every $A \in \mathcal{B}$. Thus, any nonequilibrium distribution tends to an equilibrium one with time. The mixing property has a significant meaning from a physical point of view, as it implies decay of correlation of observable functions; moreover, limiting distributions of averaged observables are determined by the decay rates of correlation functions for many cases (e.g., hyperbolic systems). For any $f \in L^2(X, \mathcal{B}, \mu)$ we consider the scalar products $s_n = s_n(f) = (U^n f, f), n \geq 0$ and define $s_n := \bar{s}_{-n}$ for $n < 0$. The sequence $\{s_n\}_{n\in\mathbf{Z}}$ is positive definite and so by Bohner's theorem, we can write $s_n(f) = \int_0^1 \exp[2\pi i n\lambda] \, \mathrm{d}\sigma_f(\lambda)$, where $\sigma_f$ is a finite Borel measure on the unit circle $S^1$ and satisfies the condition that $\sigma_f(S^1) = \|f\|^2$. Such a measure is called a spectral measure of $f$. We see that $T$ is mixing iff for any $f \in L^2(X, \mathcal{B}, \mu)$ with $\int_X f \, \mathrm{d}\mu = 0$ the Fourier coefficients $\{s_n\}$ of the spectral measure $\sigma_f$ tend to zero as $|n| \to \infty$. Let $(X, \mathcal{B}, \mu)$ be isomorphic to $([0,1], \mathcal{B}_0, \lambda)$, where $\mathcal{B}_0$ is the Borel $\sigma$-algebra on [0, 1] and $\lambda$ is the normalized Lebesgue measure of [0, 1]. Then we call a measure-preserving transformation $T$ on $(X, \mathcal{B}, \mu)$ an exact endomorphism if $\bigcap_{n=0}^{\infty} T^{-n}\mathcal{B} = \{X, \emptyset\}(\mu \bmod 0)$. We can verify that an exact endomorphism is (strong) mixing (Rohlin 1964). Moreover, $\mu$ is exact if for any positive-measure set $A \in \mathcal{B}$ with $T^n A \in \mathcal{B}(\forall n \geq 0) \lim_{n\to\infty} \mu(T^n A) = 1$ holds. Let $T$ be a nonsingular transformation on $(X, \mathcal{B}, \nu)$, that is, $\nu T^{-1} \sim \nu$. Then we can define the transfer (Perron–Frobenius) operator $\mathcal{L}_\nu : L^1(X, \nu) \to L^1(X, \nu)$ by $\mathcal{L}_\nu f := \mathrm{d}(f\nu)T^{-1}/\mathrm{d}\nu$, which satisfies

$$\int_X (\mathcal{L}_\nu f) g \, \mathrm{d}\nu = \int_X f(gT) \mathrm{d}\nu \quad (\forall g \in L^\infty(X, \nu))$$

We say that a nonsingular measure $\nu$ is exact if $A \in \bigcap_{n=0}^{\infty} T^{-n}\mathcal{B}$ implies $\nu(A)\nu(A^c) = 0$. By Lin's theorem (Lin 1971) the exactness of $\nu$ can be described as follows; $\forall f \in L^1(X, \nu)$ with $\int_X f \, \mathrm{d}\nu = 0$, $\lim_{n\to\infty} \|\mathcal{L}_\nu{}^n f\|_1 = 0$. Let $\mu = h\nu$ be an exact $T$-invariant probability measure equivalent to $\nu$. Then the upper bounds of mixing rates of the exact measure $\mu = h\nu$ are determined by the speed of $L^1$-convergence of the iterated transfer operators $\{\mathcal{L}_\nu{}^n\}$. This is because $\mathcal{L}_\nu h = h$ and for every $f \in L^1(X, \nu)$ with $\int_X f \, \mathrm{d}\nu = 1$, $\lim_{n\to\infty} \|\mathcal{L}_\nu{}^n f - h\|_1 = 0$. Hence, the property $\mathcal{L}_\mu f = h^{-1}\mathcal{L}_\nu(hf)$ allows one to see that for every $f, g \in L^\infty(X, \mu)$ the correlation function

$$C_{f,g}(n) := \left| \int_X (fT^n) g \, \mathrm{d}\mu - \int_X f \, \mathrm{d}\mu \int_X g \, \mathrm{d}\mu \right|$$

is bounded from above by

$$\begin{aligned} &\|f\|_\infty \|\mathcal{L}_\mu^n g - \int_X g \, \mathrm{d}\mu\|_1 \\ &\quad = \|f\|_\infty \|h^{-1}\{\mathcal{L}_\nu{}^n(gh) - P(gh)\}\|_1 \end{aligned}$$

where $P : L^1(X, \nu) \to L^1(X, \nu)$ is a linear operator defined by $Pf := h \int_X f \, \mathrm{d}\nu$. The operator $P$ is the one-dimensional projection operator associated to the eigenvalue 1 (which is maximal in many cases) of $\mathcal{L}_\nu$ satisfying $P^2 = P$ and $P\mathcal{L}_\nu = \mathcal{L}_\nu P = P$. Moreover, since $\mathcal{L}_\nu{}^n - P = (\mathcal{L}_\nu - P)^n$, the exponential decay of mixing rates follows from the spectral gap of $\mathcal{L}_\nu$, that is, 1 is the simple isolated maximal eigenvalue of $\mathcal{L}_\nu$.

## Entropy and Reversibility

We recall one of the fundamental problems of ergodic theory, namely deciding when two automorphisms $T_1, T_2$ of probability spaces $(X_1, \mathcal{B}_1, \mu_1)$ and $(X_2, \mathcal{B}_2, \mu_2)$ are equivalent. The approach developed for this problem involved the study of spectral properties of the associated isometric operators $U_i : L^2(X_i, \mu_i) \to L^2(X_i, \mu_i)(i = 1, 2)$ and is based on the concept of the entropy of automorphism $T$, introduced by Kolmogorov (1958). The entropy is a non-negative number, which is the same for equivalent automorphisms. For example, the entropy of the Bernoulli shift $\sigma : \Pi_{n\in\mathbf{Z}}\{1, 2, \ldots, d\} \to \Pi_{n\in\mathbf{Z}}\{1, 2, \ldots, d\}$ with probability vector $(p_1, p_2, \ldots, p_d)$ is equal to $-\sum_{k=1}^{d} p_k \log p_k$. A remarkable theorem of Ornstein (1970) states that Bernoulli shifts with the same entropy are equivalent. On the other hand, Shannon (1948) introduced a notion of entropy in his work information theory, which is essentially the same as Kolmogorov's. Let $T : X \to X$ be a measure-preserving transformation on a probability space $(X, \mathcal{B}, \mu)$. We define the entropy of a measurable partition $\alpha$ of $X$ by

$H_\mu(\alpha) = -\sum_{A\in\mathcal{A}} \mu(A)\log\mu(A)$ and define the entropy of $T$ with respect to $\alpha$ by

$$h_\mu(T,\alpha) := \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\bigvee_{i=0}^{n-1}T^{-i}\alpha\right)$$

Then the (measure-theoretic) entropy of $T$ is defined by

$$h_\mu(T) = \sup_{\alpha:H_\mu(\alpha)<\infty} h_\mu(T,\alpha)$$

The next Abramov theorem gives an important method of practical computation: let $\{\alpha_n\}_{n\geq 1}$ be an increasing sequence of partitions with $H_\mu(\alpha_n) < \infty (\forall n \geq 1)$ and such that $\bigcup_{n\geq 1}\alpha_n$ generates the $\sigma$-algebra $\mathcal{B}$. Then $h_\mu(T) = \lim_{n\to\infty} h_\mu(T,\alpha_n)$. We say that a partition $\alpha$ is called a generator for a noninvertible measure-preserving transformation $T$ on a probability space $(X,\mathcal{B},\mu)$ if $\bigvee_{i=0}^{\infty}T^{-i}\alpha$ generates $\mathcal{B}$. If $T$ is invertible then a partition $\alpha$ is called a generator if $\bigvee_{i=-\infty}^{\infty}T^{-i}\alpha$ generates $\mathcal{B}$. In the case when $\alpha$ is a generator with $H_\mu(\alpha) < \infty$, by the Kolmogorov–Sinai theorem we have $h_\mu(T) = h_\mu(T,\alpha)$. Let $\alpha_n(x)$ denote an element of $\bigvee_{i=0}^{n-1}T^{-i}\alpha$ containg $x \in X$. By the Shannon–McMillan–Breiman theorem, if $T$ is a measure-preserving transformation of the probability space $(X,\mathcal{B},\mu)$ and $\alpha$ is a partition of $X$ with $H_\mu(\alpha) < \infty$, then $-(1/n)\mu(\alpha_n(x))$ converges $\mu$-a.e. and in $L^1(X,\mu)$ as $n\to\infty$. If $T$ is ergodic, then the limit coincides with $h_\mu(T,\alpha)$. Now we can apply these results to piecewise expanding transitive (countable) Markov transformations $T$ of $X \subset \boldsymbol{R}^d$. More specifically, let $\nu$ be the normalized Lebesgue measure of $X$. It is well known that under certain conditions there exists the unique ergodic invariant probability measure $\mu$ equivalent to $\nu$. Then we can establish the Rohlin's entropy formula (Rohlin 1964):

$$h_\mu(T) = \int_X \log|\det DT|\,\mathrm{d}\mu$$

under the assumptions that $H_\nu(\alpha) < \infty$ and $\log|\det DT| \in L^1(X,\nu)$. In particular, if $\alpha$ is a finite partition and $\phi = -\log|\det DT|$ is piecewise Hölder continuous, then the entropy formula just implies that $\mu$ is an equilibrium state for the potential $\phi$ in the following sense:

$$h_\mu(T) + \int_X \phi\,\mathrm{d}\mu = \sup\{h_m(T) + \int_X \phi\,\mathrm{d}m | m$$

is a $T$-invariant Borel probability measure on $X\}$, where the right-hand side is called the pressure for $\phi$ (Walters 1981).

We now turn our attention to results which relate entropy to Lyapunov exponent in the context of smooth invertible systems. Let $T$ be a diffeomorphism of a compact manifold $M$. We say that $x \in M$ is a regular point of $T$ if there exist numbers $\lambda_1(x) > \lambda_2(x) > \cdots > \lambda_d(x)$ and a decomposition $T_xM = E_1(x) + E_2(x) + \cdots + E_d(x)$ such that

$$\lim_{n\to\infty}\frac{1}{n}\log\|DT^n(x)u\| = \lambda_j(x)$$

for every $0 \neq u \in E_j(x)$ and every $1 \leq j \leq d$. Let $\Lambda$ be the set of regular points of $T$. Then we define a function

$$\chi(x) := \sum_{\lambda_j(x)\geq 0}\lambda_j(x)\dim E_j(x)$$

In the case when all Lyapunov exponents at $x$ are negative, we put $\chi(x) = 0$. Then for every $T$-invariant Borel probability measure $\mu$ on $(X,\mathcal{B})$, it holds that $h_\mu(T) \leq \int_X \chi\,\mathrm{d}\mu$ (Ruelle 1978). Moreover, the equality holds whenever $T$ is $C^1$-Hölder and $\mu$ is absolutely continuous with respect to the Lebesgue measure of $X$ (Pesin 1977). Let $T$ be a transitive $C^1$-Hölder Anosov diffeomorphism. $E^{\mathrm{s}}, E^{\mathrm{u}}$ denote the stable and unstable fiber bundles of $T$. Suppose that $\mu_+$ is the unique $T$-invariant probability measure which satisfies

$$\int_M f\,\mathrm{d}\mu_+ = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}fT^k(x)$$

for every continuous function $f : M \to \boldsymbol{R}$ and almost everywhere $x \in M$ with respect to the Lebesgue measure. The probability measure is the so-called Sinai–Ruelle–Bowen (SRB) measure. Then we have

$$h_{\mu_+}(T) = \int_M \log|\det DT(x)|_{E_x^u}|\mathrm{d}\mu_+(x)$$

On the other hand, we have

$$h_{\mu_+}(T) = \int_M \log|\det DT^{-1}(x)|_{E_x^s}|\mathrm{d}\mu_+(x) + \int_M \log|\det DT(x)|\mathrm{d}\mu_+(x)$$

We also define unti-SRB measure $\mu_-$ by replacing $T$ by $T^{-1}$. Then the SRB measure $\mu_+$ is absolutely continuous with respect to the Lebesgue measure of $M$ iff $\mu_+$ coincides with the unti-SRB measure $\mu_-$ (Bowen 1975). Hence, the SRB measure is absolutely continuous iff $\int_M \log|\det DT(x)|\mathrm{d}\mu_+(x) = 0$. This property is sometimes explained as "zero entropy production" and also as "reversibility" in the context of non-equilibrium statistical mechanics (Ruelle 1997).

*See also:* Chaos and Attractors; Determinantal Random Fields; Dissipative Dynamical Systems of Infinite Dimension; Dynamical Systems and Thermodynamics; Finitely Correlated States; Fourier Law; Fractal Dimensions in Dynamics; Homeomorphisms and Diffeomorphisms of the Circle; Hyperbolic Billiards; Hyperbolic dynamical Systems; Intermittency in

Turbulence; Large Deviations in Equilibrium Statistical Mechanics; Lyapunov Exponents and Strange Attractors; Nonequilibrium Statistical Mechanics: Interaction Between Theory and Numerical Simulations; Nonequilibrium Statistical Mechanics (Stationary): Overview; Phase Transitions in Continuous Systems; Polygonal Billiards; Regularization for Dynamical Zeta Functions; Singularity and Bifurcation Theory; von Neumann Algebras: Introduction, Modular Theory, and Classification Theory.

## Further Reading

Aaronson J (1997) *An Introduction to Infinite Ergodic Theory*. Mathematical Surveys and Monographs 50. Providence, RI: American Mathematical Society.

Bergelson V, Boshernitzan M, and Bourgain J (1994) Some results on non-linear recurrence. *Journal d'analyse mathematique* 62: 444–458.

Billingsley P (1965) *Ergodic Theory and Information*. New York: Wiley.

Birkhoff G (1931) Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences of the USA* 17: 656–660.

Bowen R (1975) Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms. *Springer Lecture Notes in Mathematics*, vol. 470. Berlin: Springer.

Bunimovich LA *et al.* (2000) Dynamical Systems, Ergodic Theory and Applications. *Encyclopedia of Mathematical Science*, vol. 100. Berlin: Springer.

Cornfeld IP, Fomin SV, and Sinai YaG (1982) Ergodic Theory. *Grundlehren der mathematischen Wissenschaften*, vol. 245. Berlin: Springer.

Darling DA and Kac M (1957) On occupation times for Markov processes. *Transactions of the American Mathematical Society* 84: 444–458.

Kolmogorov AN (1958) A new metric invariant of transitive dynamical systems and Lebesgue space automorphisms. *Doklady Akademii Nauk USSR* 119: 861–864.

Koopman BO (1931) Hamiltonian systems and transformations in Hilbert space. *Proceedings of the National Academy of Sciences of the USA* 17: 315–318.

Lin M (1971) Mixing for Markov operators. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete* 19: 231–234.

Ornstein D (1970) Bernoulli shifts with the same entropy are isomorphic. *Advances in Mathematics* 4: 337–352.

Pesin YB (1977) Characteristic Lyapunov exponents and smooth ergodic theory. *Russian Mathematical Surveys* 32: 54–114.

Rohlin VI (1964) Exact endomorphism of a Lebesgue space. *American Mathematical Society Translations* 39: 1–36.

Ruelle D (1978) An inequality for the entropy of differential maps. *Boletim da Sociedade Brasileira de Mathematica* 9: 83–87.

Ruelle D (1997) Entropy production in nonequilibrium statistical mechanics. *Communications in Mathematical Physics* 189: 365–371.

Shanonn C (1948) A mathematical theory of communications. *Bell System Technical Journal* 27: 379–423, 623–656.

Sinai YaG (1970) Dynamical systems with elastic reflections. *Russian Mathematical Surveys* 25: 137–189.

von Neumann J (1932) Proof of the quasi-ergodic hypothesis. *Proceedings of the National Academy of Sciences of the USA* 18: 70–82.

Walters P (1981) *An Introduction to Ergodic Theory*. GTM, Springer Verlag.

# Euclidean Field Theory

**F Guerra**, Università di Roma "La Sapienza," Rome, Italy



## Introduction

In this article, we consider Euclidean field theory as a formulation of quantum field theory which lives in some Euclidean space, and is expressed in probabilistic terms. Methods arising from Euclidean field theory have been introduced in a very successful way in the study of concrete models of constructive quantum field theory.

Euclidean field theory was initiated by Schwinger (1958) and Nakano (1959), who proposed to study the vacuum expectation values of field products analytically continued into the Euclidean region (Schwinger functions), where the first three (spatial) coordinates of a world point are real and the last one (time) is purely imaginary (Schwinger points). The possibility of introducing Schwinger functions, and their invariance under the Euclidean group are immediate consequences of the now classic formulation of quantum field theory in terms of vacuum expectation values given by Wightman (Streater and Wightman 1964). The convenience of dealing with the Euclidean group, with its positive-definite scalar product, instead of the Lorentz group, is evident, and has been exploited by several authors, in different contexts.

The next step was made by Symanzik (1966), who realized that Schwinger functions for boson fields have a remarkable positivity property, allowing to introduce Euclidean fields on their own sake. Symanzik also pointed out an analogy between Euclidean field theory and classical statistical mechanics, at least for some interactions (Symanzik 1969).

This analogy was successfully extended, with a different interpretation, to all boson interactions by Guerra *et al.* (1975), with the purpose of using rigorous results of modern statistical mechanics for

the study of constructive quantum field theory, within the program advocated by Wightman (1967), and further pursued by Glimm and Jaffe (see Glimm and Jaffe (1981) for an overall presentation).

The most dramatic advance of Euclidean theory was due to Nelson (1973a, b). He was able to isolate a crucial property of Euclidean fields (the Markov property) and gave a set of conditions for these fields, which allow us to derive all properties of relativistic quantum fields satisfying Wightman axioms. The Nelson theory is very deep and rich in new ideas. Even after so many years since the basic papers were published, we lack a complete understanding of the radical departure from the conventional theory afforded by Nelson's ideas, especially about their possible further developments.

By using the Nelson scheme, in particular a very peculiar symmetry property, it was very easy to prove (Guerra 1972) the convergence of the ground-state energy density, and the van Hove phenomenon in the infinite-volume limit for two-dimensional boson theories. A subsequent analysis (Guerra *et al.* 1972) gave other properties of the infinite-volume limit of the theory, and allowed a remarkable simplification in the proof of a very important regularity property for fields, previously established by Glimm and Jaffe.

Since then, all work on constructive quantum field theory has exploited in different ways ideas coming from Euclidean field theory. Moreover, a very important reconstruction theorem has been established by Osterwalder and Schrader (1973), allowing a reconstruction of relativistic quantum fields from the Euclidean Schwinger functions, and avoiding the previously mentioned Nelson reconstruction theorem, which is technically more difficult to handle.

This article is intended to be an introduction to the general structure of Euclidean quantum field theory, and to some of the applications to constructive quantum field theory. Our purpose is to show that, 50 years after its introduction, the Euclidean theory is still interesting, both from the point of view of technical applications and physical interpretation.

The article is organized as follows. In the next section, by considering simple systems made of a single spinless relativistic particle, we introduce the relevant structures in both Euclidean and Minkowski worlds. In particular, a kind of (pre)Markov property is introduced already at the one-particle level.

Next we present a description of the procedure of second quantization on the one-particle structure. The free Markov field is introduced, and its crucial Markov property explained. Following Nelson, we use probabilistic concepts and methods, whose relevance for constructive quantum field theory became immediately more and more apparent. The very structure of classical statistical mechanics for Euclidean fields is firmly based on these probabilistic methods. This is followed by an introduction of interaction, and we show the connection between the Markov theory and the Hamiltonian theory, for two-dimensional space-cutoff interacting scalar fields. In particular, we present the Feynman–Kac–Nelson formula that gives an explicit expression of the semigroup generated by the space-cutoff Hamiltonian in $\Phi o \kappa$ space. We also deal with some applications to constructive quantum field theory. This is followed by a short discussion about the physical interpretation of the theory. In particular, we discuss the Osterwalder–Schrader reconstruction theorem on Euclidean Schwinger functions, and the Nelson reconstruction theorem on Euclidean fields. For the sake of completeness, we sketch the main ideas of a proposal, advanced in Guerra and Ruggiero (1973), according to which the Euclidean field theory can be interpreted as a stochastic field theory in the physical Minkowski spacetime.

Our treatment will be as simple as possible, by relying on the basic structural properties, and by describing methods of presumably very long lasting power. The emphasis given to probabilistic methods, and to the statistical mechanics analogy, is a result of the historical development. Our opinion is that not all possibilities of Euclidean field theory have been fully exploited yet, both from technical and physical points of view.

## One-Particle Systems

A system made of only one relativistic scalar particle, of mass $m > 0$, has a quantum state space represented by the positive-frequency solutions of the Klein–Gordon equation. In momentum space, with points $p_\mu, \mu = 0, 1, 2, 3$, let us introduce the upper mass hyperboloid, characterized by the constraints $p^2 \equiv p_0^2 - \sum_{i=1}^{3} p_i^2 = m^2, p_0 \geq m$, and the relativistic invariant measure on it, formally given by $\mathrm{d}\mu(p) = \theta(p_0)\delta(p^2 - m^2)\,\mathrm{d}p$, where $\theta$ is the step function $\theta(x) = 1$ if $x \geq 0$, and $\theta(x) = 0$ otherwise, and $\mathrm{d}p$ is the four-dimensional Lebesgue measure. The Hilbert space of quantum states $F$ is given by the square-integrable functions on the mass hyperboloid equipped with the invariant measure $\mathrm{d}\mu(p)$. Since in some reference frame the mass hyperboloid is uniquely characterized by the space values of the momentum $\boldsymbol{p}$, with the energy given by $p_0 \equiv \omega(\boldsymbol{p}) = \sqrt{\boldsymbol{p}^2 + m^2}$, the Hilbert space $F$ of the states is, in fact, made of those complex-valued tempered distributions $f$ in the configuration space $R^3$ whose Fourier transforms, $\tilde{f}(\boldsymbol{p})$, are square-integrable functions in momentum space with respect to the image of the relativistic invariant measure $\mathrm{d}\boldsymbol{p}/2\omega(\boldsymbol{p})$, where

d$\boldsymbol{p}$ is the Lebesgue measure in momentum space. The scalar product on $F$ is defined by

$$\langle f,g\rangle_F=(2\pi)^3\int\tilde{f}^*(\boldsymbol{p})\tilde{g}(\boldsymbol{p})\frac{\mathrm{d}\boldsymbol{p}}{2\omega(\boldsymbol{p})}$$

where we have normalized the Fourier transform in such a way that

$$f(\boldsymbol{x})=\int\exp(\mathrm{i}\boldsymbol{p}.\boldsymbol{x})\tilde{f}(\boldsymbol{p})\,\mathrm{d}\boldsymbol{p}$$

$$\tilde{f}(\boldsymbol{p})=(2\pi)^{-3}\int\exp(-\mathrm{i}\boldsymbol{p}.\boldsymbol{x})\tilde{f}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

$$\int\exp(\mathrm{i}\boldsymbol{p}.\boldsymbol{x})\,\mathrm{d}\boldsymbol{p}=(2\pi)^3\delta(\boldsymbol{x})$$

The scalar product on $F$ can also be expressed in the form

$$\langle f,g\rangle_F=\iint f(\boldsymbol{x}')^*W(\boldsymbol{x}'-\boldsymbol{x})g(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}'\,\mathrm{d}\boldsymbol{x}$$

where we have introduced the two-point Wightman function at fixed time, defined by

$$W(\boldsymbol{x}'-\boldsymbol{x})=(2\pi)^{-3}\int\exp(\mathrm{i}\boldsymbol{p}.(\boldsymbol{x}'-\boldsymbol{x}))\frac{\mathrm{d}\boldsymbol{p}}{2\omega(\boldsymbol{p})}$$

A unitary irreducible representation of the Poincaré group can be defined on $F$ in the obvious way. In particular, the generators of space translations are given by multiplication by the components of $\boldsymbol{p}$ in momentum space, and the generator of time translations (the energy of the particle) is given by $\omega(\boldsymbol{p})$.

For the scalar product of time-evolved wave functions, we can write

$$\begin{aligned}&\langle\exp(-\mathrm{i}t')f,\exp(-\mathrm{i}t)g\rangle_F\\&=\iint f(\boldsymbol{x}')^*W(t'-t,\boldsymbol{x}'-\boldsymbol{x})g(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}'\mathrm{d}\boldsymbol{x}\end{aligned}$$

where we have introduced the two-point Wightman function, defined by

$$\begin{aligned}&W(t'-t,\boldsymbol{x}'-\boldsymbol{x})\\&=(2\pi)^{-3}\int\exp(-\mathrm{i}(t-t'))\exp(\mathrm{i}\boldsymbol{p}.(\boldsymbol{x}'-\boldsymbol{x}))\frac{\mathrm{d}\boldsymbol{p}}{2\omega(\boldsymbol{p})}\end{aligned}$$

To the physical single-particle system living in Minkowski spacetime, we associate a kind of mathematical image, living in Euclidean space, from which all properties of the physical system can be easily derived. We start from the two-point Schwinger function

$$S(x)=\frac{1}{(2\pi)^4}\int\frac{\exp(\mathrm{i}p\cdot x)}{\sqrt{p^2+m^2}}\,\mathrm{d}p$$

which is the analytic continuation of the previously given two-point Wightman function into the Schwinger points. Here $x,p\in R^4$, and $p\cdot x=\sum_{i=1}^4 x_ip_i$. Here d$p$ and d$x$ are the Lebesgue measures in the $R^4$ momentum and configuration spaces, respectively. The function $S(x)$ is positive and analytic for $x\neq 0$, decreases as $\exp(-m\|x\|)$ as $x\to\infty$, and satisfies the equation

$$(-\Delta+m^2)S(x)=\delta(x)$$

where $\Delta=\sum_{i=1}^4\partial^2/\partial x_i^2$ is the Laplacian in four dimensions.

The mathematical image we are looking for is described by the Hilbert space $N$ of those tempered distributions in four-dimensional configuration space $R^4$ whose Fourier transforms are square integrable with respect to the measure $\mathrm{d}p/\sqrt{p^2+m^2}$. The scalar product on $N$ is defined by

$$\langle f,g\rangle_N=(2\pi)^4\int\tilde{f}^*(p)\tilde{g}(p)\frac{\mathrm{d}p}{\sqrt{p^2+m^2}}$$

Four-dimensional Fourier transforms are normalized as follows:

$$f(x)=\int\exp(\mathrm{i}p.x)\tilde{f}(p)\,\mathrm{d}p$$

$$\tilde{f}(p)=(2\pi)^{-4}\int\exp(-\mathrm{i}p.x)\tilde{f}(x)\,\mathrm{d}x$$

$$\int\exp(\mathrm{i}p.x)\,\mathrm{d}p=(2\pi)^4\delta(x)$$

We also write

$$\begin{aligned}\langle f,g\rangle_N&=\iint f^*(x)S(x-y)g(y)\,\mathrm{d}x\,\mathrm{d}y\\&=\left\langle f,(-\Delta+m^2)^{-1}g\right\rangle\end{aligned}$$

where $\langle\,,\rangle$ is the ordinary Lebesgue product defined on Fourier transforms and, in momentum space, $(-\Delta+m^2)^{-1}$ amounts to a multiplication by $(p^2+m^2)^{-1}$. The Schwinger function $S(x-y)$ is formally the kernel of the operator $(-\Delta+m^2)^{-1}$. The Hilbert space $N$ is the carrier space of a unitary (nonirreducible) representation of the four-dimensional Euclidean group $E(4)$. In fact, let $(a,R)$ be an element of $E(4)$

$$\begin{aligned}(a,R):R^4&\to R^4\\x&\to Rx+a\end{aligned}$$

where $a\in R^4$, and $R$ is an orthogonal matrix, $RR^{\mathrm{T}}=R^{\mathrm{T}}R=1_4$. Then the transformation $u(a,R)$ defined by

$$\begin{aligned}&u(a,R):N\to N\\&\quad f(x)\to(u(a,R)f)(x)=f(R^{-1}(x-a))\end{aligned}$$

provides the representation. In particular, we consider the reflection $r_0$ with respect to the hyperplane

$x_4 = 0$, and the translations $u(t)$ in the $x_4$-direction. Then we have $r_0 u(t) r_0 = u(-t)$, and analogously for other hyperplanes.

Now we introduce a local structure on $N$ by considering, for any closed region $A$ of $R^4$, the subspace $N_A$ of $N$ made by distributions in $N$ with support on $A$. We call $e_A$ the orthogonal projection on $N_A$. It is obvious that if $A \in B$ then $N_A \in N_B$ and $e_A e_B = e_B e_A = e_A$. A kind of (pre)Markov property for one-particle systems is introduced as follows. Consider a closed three-dimensional piecewise smooth manifold $\sigma$, which divides $R^4$ in two closed regions $A$ and $B$, having $\sigma$ in common. Therefore, $\sigma \in A, \sigma \in B, A \cap B = \sigma, A \cup B = R^4$. Let $N_A, N_B, N_\sigma$, and $e_A, e_B, e_\sigma$ be the associated subspaces and projections, respectively. Then $N_\sigma \subset N_A, N_\sigma \subset N_B$, and $e_\sigma e_A = e_A e_\sigma = e_\sigma, e_\sigma e_B = e_B e_\sigma = e_\sigma$. It is very simple to prove the following:

**Theorem 1** *Let $e_A, e_B, e_\sigma$ be defined as above, then $e_A e_B = e_B e_A = e_\sigma$.*

Clearly, it is enough to show that for any $f \in N$ we have $e_A e_B f \in N_\sigma$. In that case, $e_\sigma e_A e_B f = e_A e_B f$, from which the theorem easily follows. Since $e_A e_B f$ has support on $A$, we must show that for any $C_0^\infty$ function $g$ with support on $A_\sigma$ we have $\langle g, e_A e_B f \rangle = 0$. Then $e_A e_B f$ has support on $\sigma$, and the proof is complete. Now we have

$$\begin{aligned}\langle g, e_A e_B f\rangle &= \langle(-\Delta + m^2)g, e_A e_B f\rangle_N \\ &= \langle e_A(-\Delta + m^2)g, e_B f\rangle_N \\ &= \langle(-\Delta + m^2)g, e_B f\rangle_N \\ &= \langle g, e_B f\rangle = 0\end{aligned}$$

where we have used the definition of $\langle\,\rangle_N$ in terms of $\langle\,\rangle$, the fact that $e_A(-\Delta + m^2)g = (-\Delta + m^2)g$, since $(-\Delta + m^2)g$ has support on $A_\sigma$, and the fact that $e_B f$ has support on $B$. This ends the proof of the (pre)Markov property for one-particle systems.

A very important role in the theory is played by subspaces of $N$ associated to hyperplanes in $R^4$. To fix ideas, consider the hyperplane $x_4 = 0$ and the associated subspace $N_0$. A tempered distribution in $N$ with support on $x_4 = 0$ has necessarily the form $(f \otimes \delta_0)(x) \equiv f(\mathbf{x})\delta(x_4)$, with $f \in F$. By using the basic magic formula, for $x \geq 0$ and $M > 0$,

$$\int_{-\infty}^{+\infty} \frac{\exp(ipx)}{p^2 + M^2}\,dp = \frac{\pi}{M}\exp(-Mx)$$

it is immediate to verify that $\|f \otimes \delta_0\|_N = \|f\|_F$. Therefore, we have an isomorphic and isometric identification of the two Hilbert spaces $F$ and $N_0$. Obviously, similar considerations hold for any hyperplane. In particular, we consider the hyperplanes $x_4 = t$ and the associated subspaces $N_t$. Let us introduce injection operators $j_t$ defined by

$$\begin{aligned} j_t : F &\to N \\ f &\to f \otimes \delta_t \end{aligned}$$

where $f$ is a generic element of $F$, with values $f(\mathbf{x})$, and $(f \otimes \delta_t)(x) = f(\mathbf{x})\delta(x_4 - t)$. It is immediate to verify the following properties for $j_t$ and its adjoint $j_t^*$: the range of $j_t$ is $N_t$; moreover, $j_t$ is an isometry, so that $j_t^* j_t = 1_F, j_t j_t^* = e_t$, where $1_F$ is the identity on $F$, and $e_t$ is the projection on $N_t$. Moreover, $e_t j_t = j_t$ and $j_t^* = j_t^* e_t$.

If we introduce translations $u(t)$ along the $x_4$-direction and the reflection $r_0$ with respect to $x_4 = 0$, then we also have the covariance property $u(t) j_s = j_{t+s}$, and the reflexivity property $r_0 j_0 = j_0$, $j_0^* r_0 = j_0^*$. The reflexivity property is very important. It tells us that $r_0$ leaves $N_0$ pointwise invariant, and it is an immediate consequence of the fact that $\delta(x_4) = \delta(-x_4)$.

Therefore, if we start from $N$ we can obtain $F$, by taking the projection $j_\pi$ with respect to some hyperplane $\pi$, in particular $x_4 = 0$. It is also obvious that we can induce on $F$ a representation of $E(3)$ by taking those elements of $E(4)$ that leave $\pi$ invariant.

Let us now see how we can define the Hamiltonian on $F$ starting from the properties of $N$. Since we are considering the simple case of the one-particle system, we could just perform the following construction explicitly by hand, through a simple application of the basic magic formula given earlier. But we prefer to follow a route that emphasizes Markov property and can be immediately generalized to more complicated cases.

Let us introduce the operator $p(t)$ on $F$ defined by the dilation $p(t) = j_0^* j_t = j_0^* u(t) j_0, t \geq 0$. Then we prove the following:

**Theorem 2** *The operator $p(t)$ is bounded and self-adjoint. The family $\{p(t)\}$, for $t \geq 0$, is a norm-continuous semigroup.*

**Proof** Boundedness and continuity are obvious. Self-adjointness is a consequence of reflexivity. In fact,

$$p^*(t) = j_0^* u(-t) j_0 = j_0^* r_0 u(t) r_0 j_0 = j_0^* u(t) j_0 = p(t)$$

The semigroup property is a consequence of the Markov property. In fact, let us introduce $N_+, N_0, N_-$ as subspaces of $N$ made by distributions with support in the regions $x_4 \geq 0, x_4 = 0, x_4 \leq 0$, respectively, and call $e_+, e_0, e_-$ the respective projections. By Markov property, we have $e_0 = e_- e_+$. Now write, for $s, t \geq 0$,

$$p(t)p(s) = j_0^* u(t) j_0 j_0^* u(s) j_0 = j_0^* u(t) e_0 u(s) j_0$$

If $e_0$ could be cancelled, then the semigroup property would follow from the group property of the translations $u(t)u(s) = u(t+s)$ (a miracle of the dilations!). For this, consider the matrix element

$$\langle f, p(t)p(s)g\rangle_F = \langle u(-t)j_0 f, e_0 u(s) j_0\, g\rangle_N$$

recall $e_0 = e_- e_+$, and use $u(s)j_0 g \in N_+$ and $u(-t)j_0 f \in N_-$.

Let us call $h$ the generator of $p(t)$, so that $p(t) = \exp(-th)$, for $t \geq 0$. By definition, $h$ is the Hamiltonian of the physical system. A simple explicit calculation shows that $h$ is just the energy $\omega$ introduced earlier. Starting from the representation of the Euclidean group $E(3)$ already given and from the Hamiltonian, we immediately get a representation of the full Poincaré group on $F$. Therefore, all physical properties of the one-particle system have been reconstructed from its Euclidean image on the Hilbert space $N$.

As a last remark of this section, let us note that we can consider the real Hilbert spaces $N_r$ and $F_r$, made of real elements (in configuration space) in $N$ and $F$. The operators $u(a,t), u(t), r_0, j_\pi, j_\pi^*, e_A$ are all reality preserving, that is, they map real spaces into real spaces.

This completes our discussion about the one-particle system. For more details we refer to Guerra *et al.* (1975) and Simon (1974). We have introduced the Euclidean image, discussed its main properties, and shown how we can derive all properties of the physical system from its Euclidean image. In the next sections, we will show how this kind of construction carries through the second-quantized case and the interacting case.

## Second Quantization and Free Fields

We begin this section with a short review about the procedure of second quantization based on probabilistic methods, by following mainly Nelson (1973b); see also Guerra *et al.* (1975) and Simon (1974). Probabilistic methods are particularly useful in the framework of the Euclidean theory.

Let $\mathcal{H}$ be a real Hilbert space with symmetric scalar product $\langle\,,\rangle$. Let $\phi(u)$ be the elements of a family of centered Gaussian random variables indexed by $u \in \mathcal{H}$, uniquely defined by the expectation values $E(\phi(u)) = 0, E(\phi(u)\phi(v)) = \langle u, v\rangle$. Since $\phi$ is Gaussian, we also have

$$E(\exp(\lambda\phi(u))) = \exp\left(\tfrac{1}{2}\lambda^2\langle u,u\rangle\right)$$

and

$$E(\phi(u_1)\phi(u_2)\cdots\phi(u_n)) = [u_1 u_2 \cdots u_n]$$

Here $[\ldots]$ is the Hafnian of elements $[u_i u_j] = \langle u_i, u_j\rangle$, defined to be zero for odd $n$, and for even $n$ given by the recursive formula

$$[u_1 u_2 \cdots u_n] = \sum_{i=2}^{n} [u_1 u_i][u_1 u_2 \cdots u_n]'$$

where in $[\ldots]'$ the terms $u_1$ and $u_i$ are suppressed. Hafnians, from the Latin name of Copenhagen, the first seat of the theoretical group of CERN, were introduced in quantum field theory by Caianiello (1973), as a useful tool when dealing with Bose statistics.

Let $(Q, \Sigma, \mu)$ be the underlying probability space where $\phi$ are defined as random variables. Here $Q$ is a compact space, $\Sigma$ a $\sigma$-algebra of subsets of $Q$, and $\mu$ a regular, countable additive probability measure on $\Sigma$, normalized to $\mu(Q) = \int_Q d\mu = 1$.

The fields $\phi(u)$ are represented by measurable functions on $Q$. The probability space is uniquely defined, but for trivial isomorphisms, if we assume that $\Sigma$ is the smallest $\sigma$-algebra with respect to which all fields $\phi(u)$, with $u \in \mathcal{H}$, are measurable. Since $\phi(u)$ are Gaussian, they are represented by $L^p(Q, \Sigma, \mu)$ functions, for any $p$ with $1 \leq p < \infty$, and the expectations will be given by

$$E(\phi(u_1)\phi(u_2)\cdots\phi(u_n)) = \int_Q \phi(u_1)\phi(u_2)\cdots\phi(u_n)\, d\mu$$

where, by a mild abuse of notation, $\phi(u_i)$ on the right-hand side denote the $Q$ space functions which represent the random variables $\phi(u_i)$. We call the complex Hilbert space $\mathcal{F} = \Gamma(\mathcal{H}) = L^2(Q, \Sigma, \mu)$ the Фок space constructed on $\mathcal{H}$, and the function $\Omega_0 \equiv 1$ on $Q$ the Фок vacuum.

In order to introduce the concept of second quantization of operators, we must introduce subspaces of $\mathcal{F}$ with a "fixed number of particles." Call $\mathcal{F}_{(0)} = \{\lambda\Omega_0\}$, where $\lambda$ is any complex number. Define $\mathcal{F}_{(\leq n)}$ as the subspace of $\mathcal{F}$ generated by complex linear combinations of monomials of the type $\phi(u_1)\cdots\phi(u_j)$, with $u_i \in \mathcal{H}$, and $j \leq n$. Then $\mathcal{F}_{(\leq n-1)}$ is a subspace of $\mathcal{F}_{(\leq n)}$. We define $\mathcal{F}_{(n)}$, the $n$-particle subspace, as the orthogonal complement of $\mathcal{F}_{(\leq n-1)}$ in $\mathcal{F}_{(\leq n)}$, so that

$$\mathcal{F}_{(\leq n)} = \mathcal{F}_{(n)} \oplus \mathcal{F}_{(\leq n-1)}$$

By construction, the $\mathcal{F}_{(n)}$ are orthogonal, and it is not difficult to verify that

$$\mathcal{F} = \bigoplus_{n=0}^{\infty} \mathcal{F}_{(n)}$$

Let us now introduce the Wick normal products by the definition

$$:\phi(u_1)\phi(u_2)\cdots\phi(u_n):=E_{(n)}\phi(u_1)\phi(u_2)\cdots\phi(u_n)$$

where $E_{(n)}$ is the projection on $\mathcal{F}_{(n)}$. It is not difficult to prove the usual Wick theorem (see, e.g., Guerra *et al.* (1975), and its inversion given by Caianiello (1973).

It is interesting to remark that, in the framework of the second quantization performed with probabilistic methods, it is not necessary to introduce creation and destruction operators as in the usual treatment. However, the two procedures are completely equivalent, as shown, for example, in Simon (1974).

Given an operator $A$ from the real Hilbert space $\mathcal{H}_1$ to the real Hilbert space $\mathcal{H}_2$, we define its second-quantized operator $\Gamma(A)$ through the following definitions:

$$\Gamma(A)\Omega_{01}=\Omega_{02}$$

$$\Gamma(A):\phi_1(u_1)\phi_1(u_2)\ldots\phi_1(u_n):\\ =:\phi_2(Au_1)\phi_2(Au_2)\cdots\phi_2(Au_n):$$

where we have introduced the probability spaces $Q_1$ and $Q_2$, their vacua $\Omega_{01}$ and $\Omega_{02}$, and the random variables $\phi_1$ and $\phi_2$, associated to $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. The following remarkable theorem by Nelson (1973b) gives a full characterization of $\Gamma(A)$, very useful in the applications.

**Theorem 3** *Let $A$ be a contraction from the real Hilbert space $\mathcal{H}_1$ to the real Hilbert space $\mathcal{H}_2$. Then $\Gamma(A)$ is an operator from $L^1_{(1)}$ to $L^1_{(2)}$ which is positivity preserving, $\Gamma(A)u\geq 0$ if $u\geq 0$, and such that $E(\Gamma(A)u)=E(u)$. Moreover, $\Gamma(A)$ is a contraction from $L^p_{(1)}$ to $L^p_{(2)}$ for any $p$, $1\leq p<\infty$. Finally, $\Gamma(A)$ is also a contraction from $L^p_{(1)}$ to $L^q_{(2)}$, with $q\geq p$, if $\|A\|^2\leq(p-1)/(q-1)$.*

We have indicated with $L^p_{(1)}, L^p_{(2)}$ the $L^p$ spaces associated to $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. This is the celebrated best hypercontractive estimate given by Nelson. For the proof, we refer to the original paper of Nelson (1973b); see also Simon (1974).

This completes our short review on the theory of second quantization based on probabilistic methods.

The usual time-zero quantum field $\bar{\phi}(u), u\in F_r$, in the Фок representation, can be obtained through second quantization starting from $F_r$. We call $(\bar{Q},\bar{\Sigma},\bar{\mu})$ the underlying probability space, and $\mathcal{F}=\Gamma(F_r)=L^2(\bar{Q},\bar{\Sigma},\bar{\mu})$ the Hilbert Фок space of the free physical particles.

Now we introduce the free Markov field $\phi(f), f\in N_r$, by taking $N_r$ as the starting point. We call $(Q,\Sigma,\mu)$ the associated probability space. We introduce the Hilbert space $\mathcal{N}=\Gamma(N_r)=L^2(Q,\Sigma,\mu)$, and the operators $U(a,R)=\Gamma(u(a,R))$, $R_0=\Gamma(r_0)$, $U(t)=\Gamma(u(t))$, $E_A=\Gamma(e_A)$, and so on, for which the previous Nelson theorem holds (take $\mathcal{H}_1=\mathcal{H}_2=N_r$).

Since in general $\Gamma(AB)=\Gamma(A)\Gamma(B)$, we have immediately the following expression of the Markov property $E_\sigma=E_AE_B$, where the closed regions $A,B,\sigma$ of the Euclidean space have the same properties as explained earlier in the proof of the (pre)Markov property for one-particle systems.

It is obvious that $E_A$ can also be understood as conditional expectation with respect to the sub-$\sigma$-algebra $\Sigma_A$ generated by the field $\phi(f)$ with $f\in N_r$ and the support of $f$ on $A$.

The relation, previously pointed out, between $N_t$ subspaces and $F$ are also valid for their real parts $N_{rt}$ and $F_r$. Therefore, they carry out through the second quantization procedure. We introduce $J_t=\Gamma(j_t)$ and $J_t^*=\Gamma(j_t^*)$; then the following properties hold. $J_t$ is an isometric injection of $L^p(\bar{Q},\bar{\Sigma},\bar{\mu})$ into $L^p(Q,\Sigma,\mu)$; the range of $J_t$ as an operator $L^2\to L^2$ is obviously $\mathcal{N}_t=\Gamma(N_{rt})$; moreover, $J_tJ_t^*=E_t$. The free Hamiltonian $H_0$ is given for $t\geq 0$ by

$$J_0^*J_t=\exp(-tH_0)=\Gamma(\exp(-t\omega))$$

Moreover, we have the covariance property $U(t)J_0=J_t$, and the reflexivity $R_0J_0=J_0, J_0^*R_0=J_0^*$.

These relations allow a very simple expression for the matrix elements of the Hamiltonian semigroup in terms of Markov quantities. In fact, for $u,v\in\mathcal{F}$ we have

$$\langle u,\exp(-tH_0)v\rangle=\int_Q (J_tu)^*J_0v\,\mathrm{d}\mu$$

In the next section, we will generalize this representation to the interacting case.

Finally, let us derive the hypercontractive property of the free Hamiltonian semigroup.

Since $\|\exp(-t\omega)\|\leq\exp(-tm)$, where $m$ is the mass of the particle, we have immediately, by a simple application of Nelson theorem,

$$\|\exp(-tH_0)\|_{p,q}\leq 1$$

provided $q-1\leq(p-1)\exp(2tm)$, where $\|\ldots\|_{p,q}$ denotes the norm of an operator from $L^p$ to $L^q$ spaces.

## Interacting Fields

The discussion of the previous sections was limited to free fields both in Minkowski and Euclidean spaces. Now we must introduce interaction in order to get nontrivial theories.

First, as a general motivation, we will proceed quite formally and then we will resort to precise statements.

Let us recall that in standard quantum field theory, for scalar self-coupled fields, the time-ordered products of quantum fields in Minkowski spacetime can be expressed formally through the formula

$$\frac{\langle T(\phi(x_1)\cdots\phi(x_n)\exp(\mathrm{i}\int\mathcal{L}\,\mathrm{d}x))\rangle}{\langle T\exp(\mathrm{i}\int\mathcal{L}\,\mathrm{d}x)\rangle}$$

where $T$ denotes time ordering, $\phi$ are free fields in Minkowski spacetime, $\mathcal{L}$ is the interaction Lagrangian, and $\langle\ldots\rangle$ are vacuum averages. As is well known, this expression can be put, for example, at the basis of perturbative expansions, giving rise to terms expressed through Feynman graphs. The appropriately chosen normalization provides automatic cancelation of the vacuum to vacuum graphs.

Now we can introduce a formal analytic continuation to the Schwinger points, as previously done for the one-particle system, and obtain the following expression for the analytic continuation of the field time-ordered products, now called Schwinger functions,

$$S(x_1,\ldots,x_n)=\frac{\langle\phi(x_1)\cdots\phi(x_n)\exp U\rangle}{\langle\exp U\rangle}$$

Here $x_1,\ldots,x_n$ denote points in Euclidean space, $\phi$ are the Euclidean fields introduced earlier. The chronological time ordering disappears, because the fields $\phi$ are commutative, and there is no distinguished "time" direction in Euclidean space. Here the symbol $\langle\ldots\rangle$ denotes the expectation values represented by $\int\ldots\mathrm{d}\mu$, as explained earlier, and $U$ is the Euclidean "action" of the system formally given by the integral on Euclidean space

$$U=-\int P(\phi(x))\,\mathrm{d}x$$

if the field self-interaction is produced by the polynomial $P$.

Therefore, these formal considerations suggest that the passage from the free Euclidean theory to the fully interacting one is obtained through a change of the free probability measure $\mathrm{d}\mu$ to the interacting measure

$$\exp U\,\mathrm{d}\mu\Big/\int_Q\exp U\,\mathrm{d}\mu$$

The analogy with classical statistical mechanics is evident. The expression exp $U$ acts as *Boltzmannfaktor*, and $Z=\int_Q\exp U\,\mathrm{d}\mu$ is the partition function.

Our task will be to make these statements precise from a mathematical point of view. We will be obliged to introduce cutoffs, and then be involved in their careful removal.

For the sake of convenience, we make the substantial simplification of considering only two-dimensional theories (one space, one time dimension in the Minkowski region) for which the well-known ultraviolet problem of quantum field theory gives no trouble. There is no difficulty in translating the contents of the previous sections to the two-dimensional case.

Let $P$ be a real polynomial, bounded below and normalized to $P(0)=0$. We introduce approximations $h$ to the Dirac $\delta$ function at the origin of the two-dimensional Euclidean space $R^2$, with $h\in N_{\mathrm{r}}$. Let $h_x$ be the translate of $h$ by $x$, with $x\in R^2$. The introduction of $h$, equivalent to some ultraviolet cutoff, is necessary, because local fields, of the formal type $\phi(x)$, have no rigorous meaning, and some smearing is necessary.

For some compact region $\Lambda$ in $R^2$, acting as space cutoff (infrared cutoff), introduce the $Q$ space function

$$U_\Lambda^{(h)}=-\int_\Lambda :P(\phi(h_x)):\,\mathrm{d}x$$

where $\mathrm{d}x$ is the Lebesgue measure in $R^2$. It is immediate to verify that $U_\Lambda^{(h)}$ is well defined, bounded below and belongs to $L^p(Q,\Sigma,\mu)$, for any $p$, $1\le p<\infty$. This is the infrared and ultraviolet cutoff action. Notice the presence of the Wick normal products in its definition. They provide a kind of automatic introduction of counterterms, in the framework of renormalization theory.

The following theorem allows us to remove the ultraviolet cutoff.

**Theorem 4** *Let $h\to\delta$, in the sense that the Fourier transforms $\tilde{h}$ are uniformly bounded and converge pointwise in momentum space to the Fourier transform of the $\delta$-function given by $(2\pi)^{-2}$. Then $U_\Lambda^{(h)}$ is $L^p$-convergent for any $p, 1\le p<\infty$, as $h\to\delta$. Call $U_\Lambda$ the $L^p$-limit, then $U_\Lambda$, $\exp U_\Lambda\in L^p(Q,\Sigma_\Lambda,\mu)$, for $1\le p<\infty$.*

The proof uses standard methods of probability theory, and originates from pioneering work of Nelson in (1966). It can be found for example in Guerra *et al.* (1975), and Simon (1974).

Since $U_\Lambda$ is defined with normal products, and the interaction polynomial $P$ is normalized to $P(0)=0$, an elementary application of Jensen inequality gives

$$\int_Q\exp U_\Lambda\,\mathrm{d}\mu\ge\exp.\int_Q U_\Lambda\,\mathrm{d}\mu=1$$

Therefore, we can rigorously define the new space cutoff measure in $Q$ space:

$$d\mu_\Lambda = \exp U_\Lambda \, d\mu \Big/ \int_Q \exp U_\Lambda \, d\mu$$

The space-cutoff interacting Euclidean theory is defined by the same fields on $Q$ space, but with a change in the measure and, therefore, in the expectation values. The correlations for the interacting fields $\bar{\phi}$ are the cutoff Schwinger functions

$$\begin{aligned} S_\Lambda(x_1, \ldots, x_n) &= \langle \bar{\phi}(x_1) \cdots \bar{\phi}(x_n) \rangle \\ &= Z_\Lambda^{-1} \langle \phi(x_1) \cdots \phi(x_n) \exp U_\Lambda \rangle \end{aligned}$$

where the partition function is

$$Z_\Lambda = \langle \exp U_\Lambda \rangle$$

We see that the analogy with statistical mechanics is complete here. Of course, the introduction of the space cutoff $\Lambda$ destroys translation invariance. The full Euclidean covariant theory must be recovered by taking the infinite-volume limit $\Lambda \to R^2$ on field correlations. For the removal of the space cutoff, all methods of statistical mechanics are available. In particular, correlation inequalities of ferromagnetic type can be easily exploited, as shown, for example, in Guerra *et al.* (1975) and Simon (1974).

We would like to conclude this section by giving the connection between the space-cutoff Euclidean theory and the space-cutoff Hamiltonian theory in the physical Фок space.

For $\ell \geq 0, t \geq 0$, consider the rectangle in $R^2$,

$$\Lambda(\ell, t) = \left\{ (x_1, x_2) \colon -\frac{\ell}{2} \leq x_1 \leq \frac{\ell}{2}, 0 \leq x_2 \leq t \right\}$$

and define the operator in the physical Фок space

$$P_\ell(t) = J_0^* \exp U_\Lambda(\ell, t) J_t$$

where $J_0$ and $J_t$ are injections relative to the lines $x_2 = 0$ and $x_2 = t$, respectively. Then the following theorem, largely due to Nelson, holds.

**Theorem 5** *The operator $P_\ell(t)$ is bounded and self-adjoint. The family $\{P_\ell(t)\}$, for $\ell$ fixed and $t \geq 0$, is a strongly continuous semigroup. Let $H_\ell$ be its lower bounded self-adjoint generator, so that $P_\ell(t) = \exp(-tH_\ell)$. On the physical Фок space, there is a core $\mathcal{D}$ for $H_\ell$ such that on $\mathcal{D}$ the equality $H_\ell = H_0 + V_\ell$ holds, where $H_0$ is the free Hamiltonian introduced earlier and $V_\ell$ is the volume-cutoff interaction given by*

$$V_\ell = \lim \int_{-\ell/2}^{\ell/2} : P(\bar{\phi}(h_{x_1})) : dx_1$$

*where $h_{x_1}$ are the translates of approximations to the $\delta$-function at the origin on the $x_1$-space, and the limit is taken in $L^p$, in analogy to what has been explained for the two-dimensional case in the definition of $U_\Lambda$.*

While we refer to Guerra *et al.* (1975) and Simon (1974) for a full proof, we mention here that boundedness is related to hypercontractivity of the free Hamiltonian, self-adjointness is a consequence of reflexivity, and the semigroup property follows from Markov property. This theorem is remarkable, because it expresses the cutoff interacting Hamiltonian semigroup in an explicit form in the Euclidean theory through probabilistic expectations. In fact, we have

$$\langle u, \exp(-tH_\ell)v \rangle = \int_Q (J_t u)^* J_0 v \exp U_\Lambda(\ell, t) \, d\mu$$

We could call this expression as the Feynman–Kac–Nelson formula, in fact it is nothing but a path integral expressed in stochastic terms, and adapted to the Hamiltonian semigroup.

By comparison with the analogous formula given for the free Hamiltonian semigroup, we see that the introduction of the interaction inserts the *Boltzmannfaktor* under the integral.

As an immediate consequence of the Feynman–Kac–Nelson formula, together with Euclidean covariance, we have the following astonishing Nelson symmetry:

$$\langle \Omega_0, \exp(-tH_\ell)\Omega_0 \rangle = \langle \Omega_0, \exp(-\ell H_t)\Omega_0 \rangle$$

which was at the basis of Guerra (1972) and Guerra *et al.* (1972), and played some role in showing the effectiveness of Euclidean methods in constructive quantum field theory.

It is easy to establish, through simple probabilistic reasoning, that $H_\ell$ has a unique ground state $\Omega_\ell$ of lowest energy $E_\ell$. For a convenient choice of normalization and phase factor, one has $\|\Omega_\ell\|_2 = 1$, and $\Omega_\ell > 0$ almost everywhere on $Q$ space (for bosonic systems, ground states have no nodes in configuration space!). Moreover, $\Omega_\ell \in L^p$, for any $1 \leq p < \infty$. If $\ell > 0$ and the interaction is not trivial, then $\Omega_\ell \neq \Omega_0, E_\ell < 0$, and $\|\Omega_\ell\|_1 < 1$. Obviously, $\|\exp(-tH_\ell)\|_{2,2} = \exp(-tE_\ell)$.

The general structure of Euclidean field theory, as explained in this section, has been at the basis of all applications in constructive quantum field theory. These applications include the proof of the existence of the infinite-volume limit, with the establishment of all Wightman axioms, for two- and three-dimensional theories. Moreover, the existence of phase transitions and symmetry breaking has been firmly established.

Extensions have also been given to theories involving Fermions, and to gauge field theory. Due to the scope of this review, limited to a description of the general structure of Euclidean field theory, we cannot give a detailed treatment of these applications. Therefore, we refer to recent general reviews on constructive quantum field theory for a complete description of all results (see, e.g., Jaffe (2000)). For recent applications of Euclidean field theory to quantum fields on curved spacetime manifolds we refer, for example, to Schlingemann (1999).

## The Physical Interpretation of Euclidean Field Theory

Euclidean field theory has been considered by most researchers as a very useful tool for the study of quantum field theory. In particular, it is quite easy, for example, to obtain the fully interacting Schwinger functions in the infinite-volume limit in two-dimensional spacetime. At this point, there arises the problem of connecting these Schwinger functions with observable physical quantities in Minkowski spacetime. A very deep result of Osterwalder and Schrader (1973) gives a very natural interpretation of the resulting limiting theory. In fact, the Euclidean theory, as has been shown earlier, arises from an analytic continuation from the physical Minkowski spacetime to the Schwinger points, through a kind of analytic continuation in time (also called Wick rotation, because Wick exploited this trick in the study of the Bethe–Salpeter equation). Therefore, having obtained the Schwinger functions for the full covariant theory, after all cutoff removal, it is very natural to try to reproduce the inverse analytic continuation in order to recover the Wightman functions in Minkowski spacetime. Therefore, Osterwalder and Schrader have been able to identify a set of conditions, quite easy to verify, wich allow us to recover Wightman functions from Schwinger functions. A key role in this reconstruction theorem is played by the so-called reflection positivity for Schwinger functions, a property quite easy to verify. In this way, a fully satisfactory solution for the physical interpretation of Euclidean field theory is achieved.

From a historical point of view, an alternate route is possible. In fact, at the beginning of the exploitation of Euclidean methods in constructive quantum field theory, Nelson was able to isolate a set of axioms for the Euclidean *fields* (Nelson 1973a), allowing the reconstruction of the physical theory. Of course, Nelson axioms are more difficult to verify, since they also involve properties of the Euclidean fields and not only of the Schwinger functions. However, it is still very interesting to investigate whether the Euclidean fields play only an auxiliary role in the construction of the physical content of relativistic theories, or if they have a more fundamental meaning.

From a physical point of view, the following considerations could also lead to further developments along this line. By its very structure, the Euclidean theory contains the fixed-time quantum correlations in the vacuum. In elementary quantum mechanics, it is possible to derive all physical content of the theory from the simple knowledge of the ground state wave function, including scattering data. Therefore, at least in principle, it should be possible to derive all physical content of the theory directly from the Euclidean theory, without any analytic continuation.

We conclude this short section on the physical interpretation of the Euclidean theory with a mention of a quite surprising result (Guerra and Ruggiero 1973) obtained by submitting classical field theory to the procedure of stochastic quantization in the sense of Nelson (1985). The procedure of stochastic quantization associates a stochastic process to each quantum state. In this case, in a fixed reference frame, the procedure of stochastic quantization, applied to interacting fields, produces, for the ground state, a process in the physical spacetime that has the same correlations as Euclidean field theory. This opens the way to a possible interpretation of Euclidean field theory directly in Minkowski spacetime. However, a consistent development along this line requires a new formulation of representations of the Poincaré group in the form of measure-preserving transformations in the probability space where the Euclidean fields are defined. This difficult task has not been accomplished as yet.

## Acknowledgments

Research connected with this work was supported in part by MIUR (Italian Minister of Instruction, University and Research), and by INFN (Italian National Institute for Nuclear Physics).

*See also:* Axiomatic Quantum Field Theory; Constructive Quantum Field Theory; Feynman Path Integrals; Functional Integration in Quantum Physics; High $T_c$ Superconductor Theory; Malliavin Calculus; Quantum Chromodynamics; Quantum Field Theory: A Brief Introduction; Quantum Fields with Indefinite Metric: Non-Trivial Models; Relativistic Wave Equations including Higher Spin Fields; Renormalization: General Theory; Two-dimensional Models.

## Further Reading


Caianiello ER (1973) *Combinatorics and Renormalization in Quantum Field Theory*. Reading, MA: Benjamin.

Glimm J and Jaffe A (1981) *Quantum Physics, A Functional Integral Point of View*. Berlin: Springer.

Guerra F (1972) Uniqueness of the vacuum energy density and van Hove phenomenon in the infinite volume limit for two-dimensional self-coupled Bose fields. *Physical Review Letters* 28: 1213.

Guerra F, Rosen L, and Simon B (1972) Nelson's symmetry and the infinite volume behavior of the vacuum in $P(\phi)_2$. *Communication in Mathematical Physics* 27: 10.

Guerra F, Rosen L, and Simon B (1975) The $P(\phi)_2$ Euclidean quantum field theory as classical statistical mechanics. *Annuals of Mathematics* 101: 111.

Guerra F and Ruggiero P (1973) New interpretation of the Euclidean Markov field in the framework of physical Minkowski space-time. *Physical Review Letters* 31: 1022.

Jaffe A (2000) *Constructive quantum field theory*, available on http://www.arthurjaffe.com.

Nakano T (1959) Quantum field theory in terms of Euclidean parameters. *Progress in Theoretical Physics* 21: 241.

Nelson E (1966) A quartic interaction in two dimensions. In: Goodman R and Segal I (eds.) *Conference on the Mathematical Theory of Elementary Particles*. Cambridge, MA: MIT Press.

Nelson E (1973a) Construction of quantum fields from Markoff fields. *Journal of Functional Analysis* 12: 97.

Nelson E (1973b) The free Markoff field. *Journal of Functional Analysis* 12: 211.

Nelson E (1985) *Quantum Fluctuations*. Princeton, NJ: Princeton University Press.

Osterwalder K and Schrader R (1973) Axioms for Euclidean Green's functions. *Communication in Mathematical Physics* 31: 83.

Schlingemann D (1999) *Euclidean field theory on a sphere*, available on http://arXiv.org.

Schwinger J (1958) On the Euclidean structure of relativistic field theory. *Proceedings of the National Academy Sciences* 44: 956.

Simon B (1974) *The $P(\phi)_2$ Euclidean (Quantum) Field Theory*. Princeton, NJ: Princeton University Press.

Streater R and Wightman AS (1964) *PCT, Spin and Statistics and All That*. New York: Benjamin.

Symanzik K (1966) Euclidean quantum field theory, I, Equations for a scalar model. *Journal of Mathematical Physics* 7: 510.

Symanzik K (1969) Euclidean quantum field theory. In: Jost R (ed.) *Local Quantum Theory*. New York: Academic Press.

Wightman AS (1967) An introduction to some aspects of the relativistic dynamics of quantized fields. In: Levy M (ed.) *1964 Cargése Summer School Lectures*. New York: Gordon and Breach.


# Evolution Equations: Linear and Nonlinear

**J Escher**, Universität Hannover, Hannover, Germany



## Introduction

In this article we present the semigroup approach to linear and nonlinear evolution equations in general Banach spaces. In the first part we introduce the general frame and we explain the cornerstones of the widely developed theory of linear evolution equations. Besides the classical approach to linear evolution equations based on $C_0$-semigroups, we also give a brief introduction to the more recent theory of maximal regularity. The entire linear theory is not only important on its own (which we prove by discussing applications to the heat equation, Schrödinger equation, wave equation, and Maxwell equations) but it is also the indispensable basis for the theory of nonlinear evolution equation, which we present in the second part.

## Linear Evolution Equations

Let $E_0$ be a Banach space, $T > 0$, and assume that $\mathcal{A} := \{A(t); t \in [0, T]\}$ is a family of closed linear operators in $E_0$. By this we mean that, given $t \in [0, T]$, there is a linear subspace $D(A(t))$ of $E_0$ and linear mapping $A(t) : D(A(t)) \subset E_0 \to E_0$ such that the graph $\{(x, A(t)x); x \in D(A(t))\}$ of $A(t)$ is a closed subspace of $E_0 \times E_0$. Given a mapping $f : [0, T] \to E_0$ and a vector $u_0 \in E_0$, we study the following initial-value problem for $(\mathcal{A}, f, u_0)$: find a function $u \in C^1((0, T], E_0)$ such that $u(t) \in D(A(t))$ for $t \in (0, T]$ and

$$u'(t) = A(t)u(t) + f(t), \quad t \in (0, T], \quad u(0) = u_0 \qquad [1]$$

Sometimes we call [1] also the Cauchy problem of the linear evolution equation $u'(t) = A(t)u(t) + f(t)$. In the following, we will specify different conditions on $(\mathcal{A}, f, u_0)$ which guarantee the well-posedness of [1], and we shall discuss several examples of equations of type [1] which are relevant in mathematical physics.

### Autonomous Homogeneous Equations

As in the case of ordinary differential equations in finite-dimensional spaces, it is convenient to consider first the autonomous version of [1], that is, we assume that $\mathcal{A}$ is trivial in the sense that $T = \infty$ and that $A(0) = A(t)$ for all $t \geq 0$. In order to simplify our notation, we set $A := A(0)$. We consider first the homogeneous problem

$$u'(t) = Au(t), \quad t \in (0, \infty), \quad u(0) = u_0 \qquad [2]$$

where $u_0 \in E_0$ is given. The question of the well-posedness of [2] is closely tied to the notion of a $C_0$-semigroup in $E_0$. Let $\mathcal{L}(E_0)$ denote the Banach space of all bounded linear operators on $E_0$, endowed with the usual operator norm. A one-parameter family $\mathcal{T} = \{T(t) \in \mathcal{L}(E_0); t \geq 0\}$ is called "$C_0$-semigroup" in $\mathcal{L}(E_0)$ iff

1. $T(0) = \mathrm{id}_{E_0}$ (normalization),
2. $T(s+t) = T(s)T(t)$ for all $s, t \geq 0$ (semigroup property), and
3. $\lim_{t\to 0} T(t)x = x$ for all $x \in E_0$ (strong continuity at 0).

Given a $C_0$-semigroup $\mathcal{T}$, we define its (infinitesimal) generator $B$ by setting

$$\mathrm{dom}(B) := \left\{ x \in E_0; \lim_{t\to 0} \frac{T(t)x - x}{t} \text{ exists in } E_0 \right\}$$

and by defining

$$Bx := \lim_{t\to 0} \frac{T(t)x - x}{t} \quad \text{for } x \in \mathrm{dom}(B)$$

This clearly defines a linear operator in $E_0$ and it is well known that $B$ is closed and densely defined. Moreover, we have

**Theorem 1** *Assume that $A : D(A) \subset E_0 \to E_0$ is the generator of a $C_0$-semigroup $\{T(t); t \geq 0\}$. Then, given $u_0 \in D(A)$, problem [2] possesses a unique solution $u$ in $C^1([0,\infty), E_0)$, which is given by $u(t) = T(t)u_0$.*

Under suitable additional assumptions it can be shown that the converse of Theorem 1 also holds true. However, we shall not go into these details but we prefer to present the following characterization of generators of $C_0$-semigroups:

**Theorem 2** (Hille–Yosida). *The operator $A : D(A) \subset E_0 \to E_0$ generates a $C_0$-semigroup iff it is closed, densely defined, and there exists $\omega$, $M \in \mathbb{R}$ such that the resolvent set $\rho(A)$ of $A$ contains the ray $(\omega, \infty)$ and such that $\|(\lambda - \omega)^n (\lambda - A)^{-n}\| \leq M$ for all $\lambda > \omega$ and all $n \in \mathbb{N}$.*

In applications, it is in general rather difficult to derive a uniform estimate of powers of the resolvent of an unbounded operator. Luckily, generators of $C_0$-semigroups of contractions (i.e., $\|T(t)\|_{\mathcal{L}(E_0)} \leq 1$ for all $t \geq 0$) can be characterized in a rather useful way. To formulate this result we call an operator $B : D(B) \subset E_0 \to E_0$ "dissipative" iff for any $x \in D(B)$ there is an $x' \in E_0'$ with $\langle x', x\rangle = \|x\|^2_{E_0} = \|x'\|^2_{E_0'}$ such that $\mathrm{Re}\langle x', Bx\rangle \leq 0$. Here $\langle\cdot,\cdot\rangle$ denotes the duality pairing between $E_0'$ and $E_0$. The operator $B$ is called "$m$-dissipative" if it is dissipative and $im(\lambda_0 - A) = E_0$ for some $\lambda_0 > 0$.

**Theorem 3** (Lumer–Phillips). *Let $A : D(A) \subset E_0 \to E_0$ be a closed and densely defined operator. Then $A$ generates a $C_0$-semigroup of contractions in $\mathcal{L}(E_0)$ iff $A$ is $m$-dissipative.*

Before we shall discuss examples of $C_0$-semigroups and their infinitesimal generators, let us introduce the following definition: given $\alpha \in (0, \pi]$, let $\Sigma_\alpha := \{z \in \mathbb{C}; |\arg(z)| < \alpha\}$ denote the sector in $\mathbb{C}$ of angle $2\alpha$. A family of operators $\mathcal{T} = \{T(z) \in \mathcal{L}(E_0); z \in \Sigma_\alpha\}$ is called a "holomorphic $C_0$-semigroup" in $\mathcal{L}(E_0)$ iff

1. $[z \mapsto T(z)] : \Sigma_\alpha \to \mathcal{L}(E_0)$ is holomorphic,
2. $T(0) = \mathrm{id}_{E_0}$ and $\lim_{z\to 0} T(z)x = x$ for all $x \in E_0$, and
3. $T(w+z) = T(w)T(z)$ for all $w, z \in \Sigma_\alpha$.

Generators of holomorphic $C_0$-semigroups can be characterized in the following way:

**Theorem 4** *A densely defined closed linear operator $A : D(A) \subset E_0 \to E_0$ generates a holomorphic $C_0$-semigroup iff there exist $M > 0$ and $\omega_0 \geq 0$ such that $\lambda \in \rho(A)$ and $\|\lambda(\lambda - A)^{-1}\| \leq M$ for all $\lambda \in \mathbb{C}$ with $\mathrm{Re}\,\lambda > \omega_0$.*

**Examples 5**

(i) *Self-adjoint generators*. Let $E_0$ be a Hilbert space and assume that $A$ is self-adjoint and that there exists an $\alpha_0 \in \mathbb{R}$ such that $A \leq \alpha_0$. Then $A$ generates a holomorphic $C_0$-semigroup $\{T(t); t \geq 0\}$. If $\{E_A(\lambda); \lambda \in \mathbb{R}\}$ denotes the spectral resolution of $A$, then $T(t) = \int_{\mathbb{R}} \exp(t\lambda)\, \mathrm{d}E_A(\lambda)$ for $t \geq 0$.

(ii) *Dissipative operators in Hilbert spaces*. Assume again that $E_0$ is a Hilbert space. Then, by Riesz' representation formula, an operator $A$ is dissipative iff $\mathrm{Re}(u|Au) \leq 0$ for all $u \in D(A)$.

(iii) *The heat semigroup*. Let $M$ be either a smooth compact closed Riemannian manifold or $\mathbb{R}^m$ with the Euclidean metric and write $\Delta$ for the Laplace–Beltrami operator on $M$. Then it is known that $\Delta \in \mathcal{L}(\mathcal{D}'(M))$, where $\mathcal{D}'(M)$ is the space of all distributions on $M$. Given $1 \leq p < \infty$, let

$$D(\Delta_p) := \{u \in L_p(M); \Delta u \in L_p(M)\}$$

and set $\Delta_p u = \Delta u$ for $u \in D(\Delta_p)$. Then $\Delta_p$ generates a holomorphic $C_0$-semigroup on $L_p(M)$, the so-called "diffusion" or "heat semigroup" on $M$. If $1 < p < \infty$, then it can be shown that $D(\Delta_p) = W_p^2(M)$, where $W_p^k(M)$ denotes the Sobolev space of order $k \in \mathbb{N}$, built over $L_p(M)$.

If $M = \mathbb{R}^m$ then the operators $T(t)$ of the semigroup generated by $\Delta_{\mathbb{R}^m}$ are given by

$$T(t)u(x) = \frac{1}{(4\pi t)^{m/2}} \int_{\mathbb{R}^m} \exp\left(\frac{-|x-y|^2}{4t}\right) u(y)\, \mathrm{d}y$$

for all $t > 0$ and almost all $x \in \mathbb{R}^m$.

Observe that the case $L_\infty(M)$ is excluded here. In fact, it is known that if a linear operator $A$ generates a $C_0$-semigroup on $L_\infty(M)$, then $A$ must be bounded. However, it can be shown that suitable realizations of the Laplace–Beltrami operator on spaces of continuous and Hölder continuous functions generate holomorphic semigroups. For more details on that topic the reader is referred to the "Further reading" section.

(iv) *Stone's theorem and the Schrödinger equation*. Let $E_0$ be a Hilbert space and assume that $A$ is self-adjoint. Then Theorem 3 and Remark (ii) imply that i$A$ generates a $C_0$-group $\{U(t); t \in \mathbb{R}\}$ of unitary operators. In fact, Stone's theorem ensures that every generator of a $C_0$-group of unitary operators is of the form i$A$ with a self-adjoint operator $A$. As an example of particular interest, let us consider the Schrödinger equation

$$\frac{1}{\mathrm{i}}\frac{\partial u}{\partial t} = \Delta u - Vu \qquad [3]$$

with a bounded potential $V: \mathbb{R}^m \to \mathbb{R}$. Letting $D(A) := H^2(\mathbb{R}^m)$ and $Au := \Delta u - Vu$, it follows that $A$ is self-adjoint in $L_2(\mathbb{R}^m)$. Hence, the evolution of [3] is governed by the group of unitary operators generated by i$A$. Of course, the assumption that $V$ be bounded is rather restrictive. In fact, there are numerous contributions which show that this assumption can be weakened considerably. Again reader is referred to the "Further reading" section for more details in this direction.

(v) *The wave equation*. Let us consider the following initial-value problem

$$\begin{aligned} \Box u(t,x) &= 0, \quad x \in \mathbb{R}^m, \quad t > 0 \\ u(0,x) &= \varphi_1(x), \quad \partial u/\partial t(0,x) \\ &= \varphi_2(x), \quad x \in \mathbb{R}^m \end{aligned} \qquad [4]$$

for the d'Alembert operator $\Box = \partial^2 u/\partial t^2 - \Delta_{\mathbb{R}^m}$ in $m+1$ dimensions. In order to associate with [4] a semigroup, let us formally re-express [4] as the following first-order system:

$$\frac{\mathrm{d}U}{\mathrm{d}t} = AU, \quad t > 0, \quad U(0) = \Phi$$

where

$$U = (u, u'), \quad A = \begin{pmatrix} 0 & \mathrm{id} \\ \Delta & 0 \end{pmatrix}, \ \Phi = (\varphi_1, \varphi_2)$$

Letting now $E_0 := H^1(\mathbb{R}^m) \times L_2(\mathbb{R}^m)$ and $D(A) := H^2(\mathbb{R}^m) \times H^1(\mathbb{R}^m)$, it can be shown that $A$ generates a $C_0$-group of linear operators in $\mathcal{L}(E_0)$. Hence, given any initial datum $(\varphi_1, \varphi_2) \in H^2(\mathbb{R}^m) \times H^1(\mathbb{R}^m)$, there exists a unique solution $u \in C^1([0,\infty), L_2(\mathbb{R}^m))$ to the initial-value problem [4]. It can be shown that this solution possesses the following additional regularity:

$$u \in C^2([0,\infty), L_2(\mathbb{R}^m)) \cap C([0,\infty), H^2(\mathbb{R}^m))$$

Hence, eqns [4] are satisfied for all $t \in [0, \infty)$ and for almost all $x \in \mathbb{R}^m$.

(vi) *Maxwell equations*. Let $\boldsymbol{E}$ and $\boldsymbol{H}$ denote the electric and magnetic field vector, respectively, $\varepsilon$ and $\mu$ the electrical permittivity and magnetic permeability, respectively, and consider the initial-value problem for Maxwell equations in vacuum and without charges and currents: given sufficiently smooth vector fields $(\boldsymbol{E}_0, \boldsymbol{H}_0)$ find a pair $(\boldsymbol{E}, \boldsymbol{H})$ such that

$$\begin{aligned} &\varepsilon\frac{\partial \boldsymbol{E}}{\partial t} - \operatorname{rot}\boldsymbol{H} = 0 \quad \text{in } (0,\infty)\times\mathbb{R}^3 \\ &\mu\frac{\partial \boldsymbol{H}}{\partial t} + \operatorname{rot}\boldsymbol{E} = 0 \quad \text{in } (0,\infty)\times\mathbb{R}^3 \\ &\boldsymbol{E}(0,\cdot) = \boldsymbol{E}_0, \quad \boldsymbol{H}(0,\cdot) = \boldsymbol{H}_0 \quad \text{in } \mathbb{R}^3 \end{aligned} \qquad [5]$$

We assume that $\varepsilon$ and $\mu$ belong to $L_\infty(\mathbb{R}^3, \mathcal{L}_{\mathrm{sym}}(\mathbb{R}^3))$ and are uniformly positive definite, that is, we assume that there are $\varepsilon_0 > 0$ and $\mu_0 > 0$ such that

$$(\varepsilon(x)y|y) \geq \varepsilon_0|y|^2, \quad (\mu(x)y|y) \geq \mu_0|y|^2$$

for all $x, y \in \mathbb{R}^3$. Based on these assumptions we endow the space $L_2(\mathbb{R}^3) \times L_2(\mathbb{R}^3)$ with the inner product

$$((u_1,\, u_2)|(v_1,\, v_2)) := (\varepsilon u_1|v_1)_{L_2} + (\mu u_2|v_2)_{L_2}$$

for $(u_1, u_2), (v_1, v_2) \in L_2(\mathbb{R}^3) \times L_2(\mathbb{R}^3)$, and call this Hilbert space $E_0$. We further set

$$E_1 := \{(u_1, u_2) \in E_0; (\operatorname{rot} u_1, \operatorname{rot} u_2) \in E_0\}$$

Finally, given $\boldsymbol{u} = (u_1, u_2) \in E_1$, let

$$Au := \left(\varepsilon^{-1}\operatorname{rot} u_2, -\mu^{-1}\operatorname{rot} u_1\right)$$

It can be shown that i$A$ is self-adjoint in $E_0$. Hence, Stone's theorem ensures that $A$ generates a $C_0$-group of unitary operators in $\mathcal{L}(E_0)$. Therefore, given $(\boldsymbol{E}_0, \boldsymbol{H}_0) \in E_1$, there exists a unique solution $(\boldsymbol{E}(\cdot), \boldsymbol{H}(\cdot))$ of [5]. For this solution, the energy functional

$$E(t) = \frac{1}{2}\int_{\mathbb{R}^3}\left[(\varepsilon\boldsymbol{E}(t)|\boldsymbol{E}(t))_{\mathbb{R}^3} + (\mu\boldsymbol{H}(t)|\boldsymbol{H}(t))_{\mathbb{R}^3}\right]\mathrm{d}x$$

is constant on $[0, \infty)$.

### Autonomous Inhomogeneous Equations

Next, we study problem [1] in the case $A(t) = A$ for all $t \in [0, T)$. Throughout this section we assume that the following minimal hypotheses

1. $A$ generates a $C_0$-semigroup in $\mathcal{L}(E_0)$,
2. $f \in L_1((0,T), E_0)$, and
3. $u_0 \in E_0$

are satisfied. Later on we shall discuss several more restrictive assumptions on $(A, f, u_0)$. A function $u:[0,T] \to E_0$ is called a "(classical) solution" of

$$u'(t) = Au(t) + f(t), \quad t \in (0,T], \quad u(0) = u_0 \qquad [6]$$

iff $u \in C([0,T], E_0) \cap C^1((0,T], E_0)$, $u(t) \in D(A)$ for all $t \in (0,T]$, and $u$ satisfies [6] pointwise on $[0,T]$. It can be shown that [6] has at most one solution. If it has a solution, this solution is represented by the following variation-of-constant-formula:

$$u(t) = T(t)u_0 + \int_0^t T(t-s)f(s)\,\mathrm{d}s, \quad t \in [0,T] \qquad [7]$$

where $\{T(t); t \geq 0\}$ denotes the semigroup generated by $A$. Observe that the function $u:[0,T] \to E_0$, defined by [7], is continuous, but in general not differentiable on $(0,T]$. For this reason one calls [7] the "mild solution" of [6].

It is not difficult to see that if $u_0 \in D(A)$ and $f \in C^1([0,T], E_0)$, then the mild solution is a classical solution, that is, [6] is uniquely solvable in the classical sense. In application to nonlinear problems, the assumption $f \in C^1([0,T], E_0)$ is often too restrictive. Fortunately, in the case of generators of holomorphic semigroups, this assumption on $f$ can be weakened in two different directions. Let $\|x\|_A := \|x\|_{E_0} + \|Ax\|_{E_0}$ denote the graph norm on $D(A)$. Then the closedness of $A$ implies that $(D(A); \|\cdot\|_A)$ is a Banach space. In the following, we call this Banach space $E_1$. Moreover, given $\alpha \in (0,1)$, we write $E_\alpha = (E_0, E_1)_\alpha$ for the complex interpolation space between $E_0$ and $E_1$. Then we have the following result.

**Theorem 6** *Let $A$ generate a holomorphic $C_0$-semigroup in $\mathcal{L}(E_0)$ and assume that there is a constant $\alpha \in (0,1)$ such that*

$$f \in C^\alpha([0,T], E_0) + C([0,T], E_\alpha)$$

*Then, given $u_0 \in E_0$, the Cauchy problem [6] possesses a unique classical solution. It is given by*

$$u(t) = T(t)u_0 + \int_0^t T(t-s)f(s)\,\mathrm{d}s, \quad t \in [0,T]$$

*where $\{T(t); t \geq 0\}$ stands for the semigroup generated by $A$.*

In the following, we discuss an alternative approach to the Cauchy problem [6], which is based on the so-called theory of maximal regularity. There are several different types of results on maximal regularity, which we cannot discuss in full detail here. We decided to give a brief introduction to the theory of the so-called "maximal $L_p$-regularity." For further results on maximal regularity, we again draw the reader's attention to the "Further reading" section.

The Banach space $E_0$ is called an unconditionality of martingale "differences" (UMD) space if the Hilbert transform is bounded on $L_q(\mathbb{R}, E_0)$ for some $q \in (1,\infty)$. It is known that Hilbert spaces, the Lebesgue spaces $L_p(X, \mathrm{d}\mu)$ with $1 < p < \infty$ and with a $\sigma$-finite measure space $(X,\mu)$, and closed subspaces of UMD spaces are UMD spaces. Furthermore, UMD spaces are without exception reflexive. Thus, the spaces $L_1(X, \mathrm{d}\mu)$, $L_\infty(X, \mathrm{d}\mu)$, and spaces of continuous or Hölder continuous functions are not UMD spaces.

Next, assume that $-A$ generates a holomorphic $C_0$-semigroup in $\mathcal{L}(E_0)$ and that $[0,\infty) \subset \rho(-A)$. Then, it is known that, given $z \in \mathbb{C}$, the fractional power $A^z$ of $A$ is a densely defined closed operator in $E_0$. We say that $A$ has bounded imaginary powers (BIP) of angle $\theta \geq 0$ if there exist positive constants $M$ and $\varepsilon$ such that

$$A^{it} \in \mathcal{L}(E_0) \quad \text{and} \quad \|A^{it}\|_{\mathcal{L}(E_0)} \leq M \exp(\theta|t|) \quad t \in (-\varepsilon, \varepsilon) \qquad [8]$$

In order to have a neat notation, we write $A \in \mathrm{BIP}(\theta)$ if [8] holds true.

**Remarks 7** In the following, we assume that $-A$ generates a holomorphic $C_0$-semigroup in $\mathcal{L}(E_0)$ and that $[0,\infty) \subset \rho(-A)$.

(i) If $\mathrm{Re}\, z < 0$, then $A^z$ is bounded on $E_0$.
(ii) There are several representation formulas for the fractional powers of $A$. Among them we picked the following: if $\mathrm{Re}\, z \in (-1,1)$ and $x \in D(A)$, then

$$A^z x = \frac{\sin(\pi z)}{\pi z} \int_0^\infty s^z (s+A)^{-2} Ax\,\mathrm{d}s$$

(iii) Assume that $E_0$ is a Hilbert space, that $A$ is self-adjoint, and that there is a positive constant $\alpha$ such that $A \geq \alpha$. Further, let $\{E_A(\lambda) \in \mathbb{R}\}$ be the spectral resolution of $A$; then

$$A^z := \int_0^\infty \lambda^z\,\mathrm{d}E_A(\lambda), \quad z \in \mathbb{C}$$

Moreover, $A \in \mathrm{BIP}(0)$.
(iv) Let again $E_0$ be a Hilbert space and assume that $-A$ is $m$-dissipative and satisfies $0 \in \rho(A)$. Then $A \in \mathrm{BIP}(\pi/2)$.

Given $p \in (1,\infty)$, Sobolev's embedding theorem ensures that $W_p^1((0,T), E_0)$ is continuously injected into $C([0,T], E_0)$. Consequently, given any function $u \in W_p^1((0,T), E_0)$ and $t \in [0,T]$, the pointwise

evaluation $u(t)$ is well defined. In particular, the trace at 0 with respect to time

$$\mathrm{tr}: W_p^1((0,T),E_0) \to E_0, \quad u \mapsto u(0)$$

is a well-defined and bounded linear operator. In order to formulate the next result, let $E_{s,p} = (E_0, E_1)_{s,p}$, with $p \in (1,\infty)$ and $s \in (0,1)$, denote the real interpolation space between the basic space $E_0$ and $E_1$, the domain $D(A)$ of $A$, endowed with the graph norm. Furthermore, we set

$$\mathbb{E}_0 := L_p((0,T),E_0)$$
$$\mathbb{E}_1 := L_p((0,T),E_1) \cap W_p^1((0,T),E_0)$$

and we write $\mathrm{Isom}(E,F)$ for the set of all topological isomorphisms mapping the Banach space $E$ onto the Banach space $F$.

**Theorem 8** (Dore and Venni). *Suppose that $E_0$ is a UMD space and that $A \in \mathrm{BIP}(\theta)$ for some $\theta \in [0,\pi/2)$. Then, given $p \in (1,\infty)$, we have*

$$(\partial_t + A, \mathrm{tr}) \in \mathrm{Isom}(\mathbb{E}_1, \mathbb{E}_0 \times E_{1-1/p,p})$$

*This means that, given $(f,u_0) \in L_p((0,T),E_0) \times E_{1-1/p,p}$, there exists a unique solution $u \in L_p((0,T),E_1) \cap W_p^1((0,T),E_0)$ of the Cauchy problem [6]. Moreover, $u$ depends continuously on $(f,u_0)$ and fulfills the following* a priori *estimate*:

$$\|u\|_{\mathbb{E}_1} \le c(\|f\|_{\mathbb{E}_0} + \|u_0\|_{E_{1-1/p,p}})$$

*where* $c := \|(\partial_t + A, \mathrm{tr})^{-1}\|_{\mathcal{L}(\mathbb{E}_0 \times E_{1-1/p,p}, \mathbb{E}_1)}$.

## Nonautonomous Equations of Hyperbolic Type

According to Theorem 1 and the corresponding remark, it is reasonable to impose in the study of the Cauchy problem [1] the minimal hypothesis that, given $s \in [0,T]$, each individual operator $A(s)$ be the generator of a $C_0$-semigroup $\{T_s(t); t \ge 0\}$ in $\mathcal{L}(E_0)$. If this semigroup is holomorphic, we call [1] of "parabolic type." Otherwise the evolution equation [1] is said to be of "hyperbolic type."

A family $\{A(t); t \in [0,T]\}$ of generators of $C_0$-semigroups in $\mathcal{L}(E_0)$ is called "stable" iff there exist positive constants $M$ and $\omega$ such that $(\omega,\infty) \subset \rho(A(t))$ for all $\in [0,T]$ and such that

$$\left\| \prod_{j=1}^{k} (\lambda - A(t_j))^{-1} \right\| \le M(\lambda - \omega)^{-k} \quad \text{for } \lambda > \omega$$

and every finite sequence $0 \le t_1 \le t_2 \le \cdots t_k \le T$ with $k \in \mathbb{N}$. Observe that the resolvent operators $(\lambda - A(t_j))^{-1}$ do not commute in general. Therefore, the order of the terms on the left-hand side of the above estimate has to be obeyed. Assume that $\mathcal{A} = \{A(t); t \in [0,T]\}$ is a family of $m$-dissipative operators. Then, $\mathcal{A}$ is stable, since any $m$-dissipative operator $B$ satisfies the estimate $\|(\lambda - B)^{-1}\| \le 1/\lambda$ for all $\lambda > 0$.

It turns out that the stability of a family of generators is not sufficient to construct a solution of [1] even in the case $f \equiv 0$. We also need a certain time regularity of the mapping $t \mapsto A(t)$. For this we say that the family $\{A(t); t \in [0,T]\}$ has a common domain $D$ iff $D$ is a dense subspace of $E_0$ such that $D(A(t)) = D$ for all $t \in [0,T]$. The family $\{A(t); t \in [0,T]\}$ is called "strongly differentiable" iff it has a common domain $D$ and, given $v \in D$, the function $t \mapsto A(t)v$ belongs to $C^1([0,T],E_0)$.

We are now prepared to formulate the following result.

**Theorem 9** (Kato). *Let $\{A(t); t \in [0,T]\}$ be a stable and strongly differentiable family of generators of $C_0$-semigroups with common domain $D$. If $f \in C^1([0,T],E_0)$ and $u_0 \in D$ then [1] possesses a unique classical solution.*

The above result is based on the construction of an evolution operator $U(t,s)$, which can be considered as the generalization of the notion of a $C_0$-semigroup for autonomous equations to the case evolution equations of the form

$$u'(t) = A(t)u(t), \quad t \in (s,T], \quad u(s) = v$$

for fixed $s \in [0,T)$. Once an evolution operator is available, the solution of [1] is given by

$$u(t) = U(t,0)u_0 + \int_0^t U(t,s)f(s)\,\mathrm{d}s, \quad t \in [0,T]$$

Of course, this generalizes [7] and if $A(t)$ is independent of $t$, then $U(t,s) = T(t-s)$, where $\{T(t); t \ge 0\}$ is the semigroup generated by $A(0)$.

Furthermore, there are several extensions of the Kato's result. Among them the most interesting contributions are concerned to weaken the time regularity of $f$ and to weaken the assumption that $\{A(t); t \in [0,T]\}$ be strongly differentiable. In particular, it is possible to study [1] for families without a common domain.

For the construction of evolution operators as well as generalizations of Theorem 9, the reader is again referred to the "Further reading" section.

## Nonautonomous Equations of Parabolic Type

Throughout this section we assume that $E_0$ and $E_1$ are Banach spaces such that $E_1$ is dense and continuously injected in $E_0$. In the study of parabolic evolution equations, the class of all operators in $\mathcal{L}(E_1,E_0)$, considered as unbounded operators in $E_0$ with common domain $E_1$, which generate holomorphic $C_0$-semigroups in $\mathcal{L}(E_0)$ has turned out to

be very useful. In the following, we call this class $\mathcal{H}(E_1, E_0)$. It is known that $A \in \mathcal{H}(E_1, E_0)$ iff there exist constants $\omega > 0$ and $\kappa \geq 1$ such that $\omega - A \in \mathrm{Isom}(E_1, E_0)$ and such that

$$\kappa^{-1} \leq \frac{\|(\lambda - A)x\|_0}{|\lambda|\|x\|_0 + \|x\|_1} \leq \kappa$$
$$x \in E_1 \backslash \{0\}, \quad \mathrm{Re}\,\lambda \geq \omega$$

where $\|\cdot\|_j$ denotes the norm of $E_j$. Using the above characterization, it can be shown that $\mathcal{H}(E_1, E_0)$ is an open subset of $\mathcal{L}(E_1, E_0)$. In the following, we always endow $\mathcal{H}(E_1, E_0)$ with the topology induced by the norm of $\mathcal{L}(E_1, E_0)$. As a consequence of this convention it is meaningful to consider, for example, continuous mappings from $[0, T]$ into $\mathcal{H}(E_1, E_0)$. Observe that if $A \in C([0, T], \mathcal{H}(E_1, E_0))$, then $\mathcal{A} = \{A(t); t \in [0, T]\}$ is a family of generators of holomorphic semigroups with the common domain $E_1$. Then we have the following result.

**Theorem 10** (Sobolevskii, Tanabe). *Assume that there is a* $\rho \in (0, 1)$ *such that*

$$(A, f) \in C^\rho([0, T], \mathcal{H}(E_1, E_0) \times E_0)$$

*Then, given* $u_0 \in E_0$, *the Cauchy problem* [1] *possesses a unique classical solution u. This solution has the additional regularity*

$$u \in C^\rho((0, T], E_1) \cap C^{1+\rho}((0, T], E_0)$$

*Finally, if* $u_0 \in E_1$, *then* $u \in C^1([0, T], E_0)$.

As in the hyperbolic case, the proof of Theorem 10 is based on the evolution operator $U(t, s)$ for the homogeneous problem, although the constructions of the corresponding evolution operators are completely different.

In addition, there are several extensions and generalizations of Theorem 10. In particular, the assumption that the family $\{A(t); t \in [0, T]\}$ possesses a common domain can be weakened considerably. Furthermore, it is possible to look at parabolic evolution equations in the so-called interpolation and extrapolation scales. This offers a great flexibility in the study of nonlinear problems. Further details in this direction can be found in the "Further reading" section.

## Nonlinear Evolution Equations

Let $E_0, E_1$ be Banach spaces such that $E_1$ is dense and continuously embedded in $E_0$. Assume further that $u_0 \in E_1$ and that we are given a nonlinear operator $F \in C([0, T] \times V, E_0)$, where $V$ is an open neighborhood of $u_0$ in $E_1$. In this section, we will discuss the well-posedness of the Cauchy problem for the following nonlinear evolution equation

$$u'(t) = F(t, u(t)), \quad t \in (0, T], \quad u(0) = u_0 \qquad [9]$$

in the Banach space $E_0$. We will always assume that the nonlinear operator $F$ either carries a quasilinear structure or is of fully nonlinear parabolic type. By a "quasilinear structure," we mean that there is mapping $A \in C([0, T] \times V, \mathcal{L}(E_1, E_0))$ and a suitable "lower-order term" $f \in C([0, T] \times V, E_0)$ such that

$$F(t, v) = A(t, v)v + f(t, v)$$
$$\text{for all } (t, v) \in [0, T] \times V$$

Problem [9] is of fully nonlinear parabolic type if $F \in C^1([0, T] \times V, E_0)$ and if the Fréchet derivative $D_2F(0, u_0)$ of $F$ with respect to $v$ at $(0, u_0)$ belongs to the class $\mathcal{H}(E_1, E_0)$.

### Quasilinear Evolution Equations of Hyperbolic Type

Assume that $E_0$ is a reflexive Banach space and let $u_0 \in V \subset E_1$ be chosen as above. We consider the following abstract quasilinear evolution equation of hyperbolic type:

$$\begin{aligned} &u'(t) = A(t, u(t))u(t) + f(t, u(t)), \ t \in (0, T] \\ &u(0) = u_0 \end{aligned} \qquad [10]$$

and assume that the following hypotheses are satisfied:

$(H_1)$ $A \in C([0, T] \times V, \mathcal{L}(E_1, E_0))$ is bounded on bounded subsets of $V$ and, given $(t, v) \in [0, T] \times V$, the operator $A(t, v)$ is $m$-dissipative and there is a constant $\mu_A$ such that

$$\|A(t, v) - A(t, w)\|_{\mathcal{L}(E_1, E_0)} \leq \mu_A \|v - w\|_{E_0}$$

for all $t \in [0, T]$ and all $v, w \in V$.

$(H_2)$ There is a $Q \in \mathrm{Isom}(E_1, E_0)$ such that $QA(t, v)Q^{-1} = A(t, v) + B(t, v)$, where $B(t, v) \in \mathcal{L}(E_0)$ is bounded, uniformly on bounded subsets of $V$. Moreover,

$$\|B(t, v) - B(t, w)\|_{\mathcal{L}(E_0)} \leq \mu_B \|v - w\|_{E_1}$$

for all $t \in [0, T]$ and all $v, w \in V$.

$(H_3)$ $f \in C([0, T] \times V, E_1)$ is bounded on bounded subsets of $V$ and there are $\mu_0$ and $\mu_1$ such that

$$\|f(t, v) - f(t, w)\|_{E_j} \leq \mu_j \|v - w\|_{E_j}$$
$$\text{for all } v, w \in V, j \in \{0, 1\}$$

Then we have the following result.

**Theorem 11** (Kato). *Assume that* $(H_1)$, $(H_2)$, *and* $(H_3)$ *are satisfied. Then there is a maximal*

$t^+ \in (0, T]$, *depending only on* $\|u_0\|_{E_1}$, *and a unique solution* $u$ *to*[10] *such that*

$$u = u(\cdot, u_0) \in C([0, t^+), V) \cap C^1([0, t^+), E_0)$$

*Moreover, the mapping* $u_0 \mapsto u(\cdot, u_0)$ *is continuous from* $V$ *to* $C([0, t^+), V) \cap C^1([0, t^+), E_0)$.

There are many applications of Theorem 11 to different concrete partial differential equations (PDEs), including symmetric hyperbolic first-order systems, the Korteweg–de Vries equation, nonlinear elastodynamics, quasilinear wave equations, Navier–Stokes and Euler equations, and coupled Maxwell–Dirac equations. We decided to explain in some detail an application to the so-called periodic Camassa–Holm equation:

$$\begin{aligned} &u_t - u_{xxt} + 3uu_x = 2u_x u_{xx} + uu_{xxx} \\ &t > 0,\ x \in \mathbb{S}^1 \end{aligned} \qquad [11]$$

where $\mathbb{S}^1$ stands for the unit circle. In the above model, the function $u$ is the height of a unilinear water wave over a flat bottom.

Set $X := L_2(\mathbb{S}^1), V := H^1(\mathbb{S}^1)$, and $Q := (I - \partial_x^2)^{1/2}$. With $y := u - u_{xx}$, eqn [11] can be re-expressed as

$$y_t + (Q^{-2})y_x = -2y(Q^{-2}y)_x \quad \text{in } L_2(\mathbb{S}^1)$$

which is of type [10] with

$$A(y) = (Q^{-2}y)\partial_x, \quad f(y) = -2y(Q^{-2}y)_x, \quad y \in V$$

where $\mathrm{dom}(A(y)) := \{v \in L_2(\mathbb{S}^1); (Q^{-2}y)v \in H^1(\mathbb{S}^1)\}$.

### Quasilinear Evolution Equations of Parabolic Type

Assume that $E_0$ and $E_1$ are Banach spaces such that $E_1$ is dense and continuously injected in $E_0$. Moreover, let $(\cdot, \cdot)_0$ for each $\theta \in (0, 1)$ be an admissible interpolation functor (e.g., the real or complex interpolation functor) and set $E_\theta := (E_0, E_1)_\theta$ for $\theta \in (0, 1)$. Given a subset $X \subset E_\theta$ for some $\theta \in (0, 1)$, we set $X_\eta := X \cap E_\eta$ for $\eta \in [0, 1]$, equipped with the topology induced by $E_\eta$. Finally, we write $C^{1-}(M, N)$ for the class of all locally Lipschitz continuous functions mapping the metric space $M$ into the metric space $N$.

**Theorem 12** (Amann). *Suppose that* $0 < \gamma \le \beta < \alpha < 1$, *that* $X_\beta$ *is open in* $E_\beta$, *and that*

$$(A, f) \in C^{1-}([0, T] \times X_\beta, \mathcal{H}(E_1, E_0) \times E_\gamma)$$

*Then, given* $u_0 \in X_\alpha$, *there exists a unique maximal* $t^+ \in (0, T]$, *such that the quasilinear parabolic Cauchy problem*

$$u'(t) = A(t, u(t))u(t) + f(t, u(t)), \quad t \in (0, T], u(0) = u_0$$

*possesses a unique classical solution*

$$u := u(\cdot, u_0) \in C([0, t^+), X_\alpha) \cap C^1((0, t^+), E_0)$$

*Assume* $A$ *and* $f$ *are independent of* $t$ *and let* $u(\cdot, u_0)$ *be the solution to corresponding autonomous problem*

$$u'(t) = A(u(t))u(t) + f(u(t)), \quad t \in (0, \infty), u(0) = u_0$$

*Then the mapping* $(t, u_0) \mapsto u(t, u_0)$ *is a semiflow on* $X_\alpha$.

Due to its clarity and flexibility, Theorem 12 has found a plethora of applications, which we cannot discuss in detail here. Let us at least mention the following: reaction–diffusion systems, population dynamics, phase transition models, flows through porous media, Stefan problems, and nonlinear and dynamic boundary conditions in boundary-value problems. In addition, many geometric evolution equations fall into the scope of Theorem 12. Consider, for example, the volume-preserving gradient flow of the area functional of a compact hypersurface $M$ in $\mathbb{R}^{m+1}$ with respect to $L_2(M)$ and $W_2^{-1}(M)$, respectively. These flows are known as the averaged mean curvature flow and the surface diffusion flow, respectively, and have been investigated on the basis of Theorem 12.

### Fully Nonlinear Evolution Equations of Parabolic Type

Based on the theory of maximal regularity for linear evolution equations, it is possible to investigate abstract fully nonlinear parabolic problems of type [9]. As there are different techniques of maximal regularity, there are also different approaches to [9]. We present here a result which uses maximal regularity properties in singular Hölder spaces $C_\beta^\beta$. Let $E_0$ and $E_1$ be Banach spaces such that $E_1$ is continuously embedded into $E_0$ (density of $E_1$ in $E_0$ is not needed here). As before, $V$ is an open subset of $E_1$ and $D_2F$ stands for the Fréchet derivative of $F(t, v)$ with respect to the second variable.

**Theorem 13** (Lunardi). *Assume that* $F \in C^2([0, T] \times V, E_0)$ *such that* $D_2F \in C^1([0, T] \times V, \mathcal{H}(E_1, E_0))$. *Then, given* $u_0 \in V$, *there is a maximal* $t^+ \in (0, T]$ *such that problem* [9] *has a solution* $u \in C([0, t^+), E_1) \cap C^1([0, t^+), E_0)$. *This solution is unique in the class*

$$\bigcup_{0<\beta<1} C_\beta^\beta((0, t^+ - \varepsilon], E_1) \cap C([0, t^+ - \varepsilon], E_1)$$

*for each* $\varepsilon \in (0, t^+)$.

Theorem 13 has important applications to problems for which the hypotheses of Theorem 12 (in particular the assumption on the quasilinear structure) are not

satisfied. We mention here fully nonlinear second-order boundary-value problems, Hele–Shaw models, models from combustion theory, and Bellman equations.

*See also:* Boltzmann Equation (Classical and Quantum); Breaking Water Waves; Dissipative Dynamical Systems of Infinite Dimension; Elliptic Differential Equations: Linear Theory; Ginzburg–Landau Equation; Image Processing: Mathematics; Incompressible Euler Equations: Mathematical Theory; Nonlinear Schrödinger Equations; Partial Differential Equations: Some Examples; Quantum Dynamical Semigroups; Relativistic Wave Equations Including Higher Spin Fields; Semilinear Wave Equations; Separation of Variables for Differential Equations; Singularities of the Ricci Flow; Symmetric Hyperbolic Systems and Shock Waves; Wave Equations and Diffraction.

## Further Reading


Amann H (1993) Nonhomogeneous linear and quasilinear elliptic and parabolic boundary value problems. *Teubner-Texte Math* 133: 9–126.

Amann H (1995) *Linear and Quasilinear Parabolic Problems*, vol. I. Basel: Birkhäuser.

Arendt W, Brézis H, and Pierre M (eds.) (2004) *Nonlinear Evolution Equations and Related Topics*. Basel: Birkhäuser.

Constantin A and Escher J (1998) Well-posedness, global existence, and blowup phenomena for a periodic quasi-linear hyperbolic equation. *Communications on Pure and Applied Mathematics* LI: 443–472.

Da Prato G and Grisvard P (1979) Equations d'évolution abstraites nonlinéaires de type parabolique. *Annali di Matematica Pura ed Applicata* 120: 329–396.

Dore G and Venni A (1987) On the closedness of the sum of two closed operators. *Mathematische Zeitschrift* 196: 189–201.

Escher J and Simonett G (1999) *Moving Surfaces and Abstract Evolution Equations*, Progress in Nonlinear Differential Equations and Their Application, vol. 35, pp. 183–212. Basel: Birkhäuser.

Fattorini HO (1983) *The Cauchy Problem*. Reading: Addison-Wesley.

Kato T (1993) *Abstract Evolution Equations, Linear and Quasilinear, Revisited*, Lecture Notes in Mathematics, vol. 1540, pp. 103–125. Berlin: Springer.

Lunardi A (1991) *Analytic Semigroups and Optimal Regularity in Parabolic Equations*. Basel: Birkhäuser.

Lunardi A (2004) *Nonlinear Parabolic Equations and Systems, Handbook of Differential Equations*, vol. 1, pp. 385–436. Amsterdam: North-Holland.

Pazy A (1983) *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Berlin: Springer.

Tanabe H (1979) *Equations of Evolutions*. London: Pitman.


# Exact Renormalization Group

**P K Mitter**, Université de Montpellier 2, Montpellier, France



## Introduction

The renormalization group (RG) in its modern form was invented by K G Wilson in the context of statistical mechanics and Euclidean quantum field theory (EQFT). It offers the deepest understanding of renormalization in quantum field theory (QFT) by connecting EQFT with the the theory of second-order phase transition and associated critical phenomena. Thermodynamic functions of many statistical mechanical models (the prototype being the Ising model in two or more dimensions) exhibit power-like singularities as the temperature approaches a critical value. One of the major triumphs of the Wilson RG was the prediction of the exponents (known as critical exponents) associated to these singularities. Wilson's fundamental contribution was to realize that many length scales begin to cooperate as one approaches criticality and that one should disentangle them and treat them one at a time. This leads to an iterative procedure known as the "renormalization group." Singularities and critical exponents then arise from a limiting process. Ultraviolet singularities of field theory can also be understood in the same way. Wilson reviews this (Wilson and Kogut 1974) and gives the historical genesis of his ideas (Wilson 1983).

The early work in the subject was heuristic, in the sense that clever but uncontrolled approximations were made to the exact equations often with much success. Subsequently, authors with mathematical bent began to use the underlying ideas to prove theorems. Benfatto, Cassandro, Gallavotti, Nicolo, Olivieri *et al.* pioneered the rigorous use of Wilson's renormalization group in the construction of super-renormalizable QFTs, (see Benfatto and Gallavotti (1995) and references therein). The subject saw further mathematical development in the work of Gawedzki and Kupiainen (1984, 1986) and that of Bałaban (1982), and references therein. Bałaban in a series of papers ending in Bałaban (1989) proved a basic result on the continuum limit of Wilson's lattice gauge theory. Brydges and Yau (1990) simplified the mathematical treatment of the renormalization group for a class of models and this has led to further systemization and simplification in the work of Brydges *et al.* (1998, 2003). Another method which has been intensely developed during the same historical period is based on phase cell expansions: Feldman, Magnen, Rivasseau, and Sénéor developed the early phase cell ideas of

Glimm and Jaffe and were able to prove independently many of the results cited earlier (see Rivasseau (1991) and references therein). Although these methods share many features of the Wilson RG, they are different in methodology and thus remain outside of the purview of the present exposition.

A somewhat different line of development has been the use of the RG to give simple proofs of perturbative renormalizability of various QFTs: Gallavotti and Nicolo, via iterative methods (see Benfatto and Gallavotti (1995) and references therein), and Polchinski (1984), who exploited a continuous version of the RG for which Wilson (1974) had derived a nonlinear differential equation. These early works were devoted to the standard $(\phi^4)_4$ scalar field theory, but subsequently Polchinski's work has been extended to a large class of models, including four-dimensional nonabelian gauge theories (*see* Kopper and Muller (2000) and references therein).

Finally, it should be mentioned that apart from QFT and statistical mechanics, the RG method has proved fruitful in other domains. An example is the study of interacting fermion systems in condensed matter physics (*see* Fermionic Systems and Renormalization: Statistical Mechanics and Condensed Matter). In the rest of this article, our focus will be on EQFT and statistical mechanics.

## The RG as a Discrete Semigroup

We will first define a discrete version of the RG and consider its continuous version later. As we will see, the RG is really a semigroup, so calling it a group is a misnomer.

Let $\phi$ be a Gaussian random field (see, e.g., Gelfand and Vilenkin (1964) for a discussion of random fields) in $\mathbf{R}^d$. Associated to it there is a positive-definite function which is identified as its covariance. In QFT one is interested in the covariance

$$E(\phi(x)\phi(y)) = \text{const.}\, |x-y|^{-2[\phi]} = \int_{\mathbf{R}^d} \mathrm{d}p\, \mathrm{e}^{ip.(x-y)} \frac{1}{|p|^{d-2[\phi]}} \qquad [1]$$

Here $[\phi] > 0$ is the (canonical) dimension of the field, which for the standard massless free field is $[\phi] = (d-2)/2$. The latter is positive for $d > 2$. However, other choices are possible but in EQFT they are restricted by the Osterwalder–Schrader positivity. It is assured if $[\phi] = (d-\alpha)/2$, with $0 < \alpha \leq 2$. If $\alpha < 2$, we get a generalized free field.

Observe that the covariance is singular for $x = y$ and this singularity is responsible for the ultraviolet divergences of QFT. This singularity has to be initially cut off and there are many ways to do this. A simple way is as follows. Let $u(x)$ be a smooth, rotationally invariant, positive-definite function of fast decrease. Examples of such functions are legion. Observe that

$$|x-y|^{-2\phi} = \text{const.}\int_0^\infty \frac{\mathrm{d}l}{l} l^{-2[\phi]}\, u\left(\frac{x-y}{l}\right) \qquad [2]$$

as can be seen by scaling in $l$. We define the unit ultraviolet cutoff covariance $C$ by cutting off at the lower end point of the $l$ integration (responsible for the singularity at $x = y$) at $l = 1$,

$$C(x-y) = \int_1^\infty \frac{\mathrm{d}l}{l} l^{-2[\phi]}\, u\left(\frac{x-y}{l}\right) \qquad [3]$$

$C(x-y)$ is positive-definite and everywhere smooth. Being positive-definite, it qualifies as the covariance of a Gaussian probability measure denoted $\mu_C$ on a function space $\Omega$ (which it is not necessary to specify any further). The covariance $C$ being smooth implies that the sample fields of the measure are $\mu_C$ almost everywhere sufficiently differentiable.

**Remark** Note that, more generally, we could have cut off the lower end point singularity in [1] at any $\epsilon > 0$. The $\epsilon$-cutoff covariance is related to the unit cutoff covariance by a scale transformation (defined below) and we will exploit this relation later.

Let $L > 1$ be any real number. We define a scale transformation $S_L$ on fields $\phi$ by

$$S_L\phi(x) = L^{-[\phi]}\,\phi\left(\frac{x}{L}\right) \qquad [4]$$

on covariances by

$$S_L C(x-y) = L^{-2[\phi]} C\left(\frac{x-y}{L}\right) \qquad [5]$$

and on functions of fields $F(\phi)$ by

$$S_L F(\phi) = F(S_L\phi) \qquad [6]$$

The scale transformations form a multiplicative group: $S_L^n = S_{L^n}$.

Now define a fluctuation covariance $\Gamma_L$:

$$\Gamma_L(x-y) = \int_1^L \frac{\mathrm{d}l}{l} l^{-2[\phi]}\, u\left(\frac{x-y}{l}\right) \qquad [7]$$

$\Gamma_L(x-y)$ is smooth, positive-definite and of fast decrease on scale $L$. It generates a key scaling decomposition

$$C(x-y) = \Gamma_L(x-y) + S_L C(x-y) \qquad [8]$$

Iterating this, we get

$$C(x-y) = \sum_{n=0}^{\infty} \Gamma_n(x-y) \qquad [9]$$

where

$$\Gamma_n(x-y) = S_{L^n}\Gamma_L(x-y) = L^{-2n[\phi]}\Gamma_L\left(\frac{x-y}{L^n}\right) \quad [10]$$

The functions $\Gamma_n(x-y)$ are of fast decrease on scale $L^{n+1}$.

Thus, [9] achieves the decomposition into a sum over increasing length scales as desired. Being positive definite, $\Gamma_n$ qualify as covariances of Gaussian probability measures, and therefore $\mu_C = \bigotimes_{n=0}^{\infty}\mu_{\Gamma_n}$. Correspondingly introduce a family of independent Gaussian random fields $\zeta_n$, called fluctuation fields, distributed according to $\mu_{\Gamma_n}$. Then

$$\phi = \sum_{n=0}^{\infty}\zeta_n \quad [11]$$

Note that the fluctuation fields $\zeta_n$ are slowly varying over length scales $L^n$. In fact, an easy estimate using a Tchebycheff inequality shows that, for any $\gamma > 0$,

$$|x-y| \le L^n \\ \Rightarrow \mu_C(|\zeta_n(x) - \zeta_n(y)| \ge \gamma) \le \text{const.}\,\gamma^{-2} \quad [12]$$

which reveals the slowly varying nature of $\zeta_n$ on scale $L^n$. Equation [11] is an example of a multiscale decomposition of a Gaussian random field.

The above implies that the $\mu_C$ integral of a function can be written as a multiple integral over the fields $\zeta_n$. We calculate it by integrating out the fluctuation fields $\zeta_n$ step by step, going from shorter to longer length scales. This can be accomplished by the iteration of a single transformation $T_L$, a renormalization group transformation, as follows. Let $F(\phi)$ be a function of fields. Then we define a RG transformation $F \to T_L F$ by

$$(T_L F)(\phi) = S_L\mu_{\Gamma_L} * F(\phi) \\ = \int d\mu_{\Gamma_L}(\zeta)\,F(\zeta + S_L\phi) \quad [13]$$

Thus the renormalization group transformation consists of a convolution with the fluctuation measure followed by a rescaling.

### Semigroup Property

The discrete RG transformations form a semigroup:

$$T_L T_{L^n} = T_{L^{n+1}} \quad \text{for all } n \ge 0 \quad [14]$$

To prove this, we must first see how scaling commutes with convolution with a measure. We have the property

$$\mu_{\Gamma_L} * S_L F = S_L\mu_{S_L\Gamma_L} * F \quad [15]$$

To see this, observe first that if $\zeta$ is a Gaussian random field distributed with covariance $\Gamma_L$ then the Gaussian field $S_L\zeta$ is distributed according to $S_L\Gamma_L$. This can be checked by computing the covariance of $S_L\zeta$. Now the left-hand side of [15] is just the integral of $F(S_L\zeta + S_L\phi)$ with respect to $d\mu_{\Gamma_L}(\zeta)$. By the previous observation, this is the integral of $F(\zeta + S_L\phi)$ with respect to $d\mu_{S_L\Gamma_L}(\zeta)$, and the latter is the right-hand side of [15]. Now we can check the semigroup property trivially:

$$\begin{aligned} T_L T_{L^n} F &= S_L\mu_{\Gamma_L} * S_{L^n}\mu_{\Gamma_{L^n}} * F \\ &= S_L S_{L^n}\mu_{S_{L^n}\Gamma_L} * \mu_{\Gamma_{L^n}} * F \\ &= S_{L^{n+1}}\mu_{\Gamma_{L^n}+S_{L^n}\Gamma_L} * F \\ &= S_{L^{n+1}}\mu_{\Gamma_{L^{n+1}}} * F \\ &= T_{L^{n+1}}F \end{aligned} \quad [16]$$

We have used the fact that $\Gamma_{L^n} + S_{L^n}\Gamma_L = \Gamma_{L^{n+1}}$. This is because $S_{L^n}\Gamma_L$ has the representation [7] with integration interval changed to $[L^n, L^{n+1}]$.

We note some properties of $T_L$. $T_L$ has an unique invariant measure, namely $\mu_C$: for any bounded function $F$,

$$\int d\mu_C\,T_L F = \int d\mu_C\,F \quad [17]$$

To understand [17], recall the earlier observation that if $\phi$ is distributed according to the covariance $C$, then $S_L\phi$ is distributed according to $S_L C$. By [8], $\Gamma_L + S_L C = C$. Therefore,

$$\begin{aligned} \int d\mu_C T_L F &= \int d\mu_C S_L\mu_{\Gamma_L} * F \\ &= \int d\mu_{S_L C}\mu_{\Gamma_L} * F \\ &= \int d\mu_C F \end{aligned} \quad [18]$$

The uniqueness of the invariant measure follows from the fact that the semigroup $T_L$ is realized by a convolution with a probability measure and, therefore, is positivity improving:

$$F \ge 0, \mu_C \text{ a.e.} \;\Rightarrow\; T_L F > 0, \mu_C \text{ a.e.}$$

Finally, note that $T_L$ is a contraction semigroup on $L^p(d\mu_C)$ for $1 \le p < \infty$. To see this, note that since $T_L$ is a convolution with a probability measure $T_L F = \mu_{S_{L^{-1}}\Gamma_L} * S_L F$, we have, via Hölder's inequality, $|T_L F|^p \le T_L|F|^p$. Then use the fact that $\mu_C$ is an invariant measure.

### Eigenfunctions

Let $:p_{n,m}:(\phi(x))$ be a $C$ Wick-ordered local monomial of $m$ fields with $n$ derivatives. Define

$$P_{n,m}(X) = \int_X dx : p_{n,m} :_C (x)$$

The functions $P_{n,m}(X)$ play the role of eigenfunctions of the RG transformation $T_L$ up to a scaling of volume:

$$T_L P_{n,m}(X) = L^{d-m[\phi]-n} P_{n,m}(L^{-1}X) \qquad [19]$$

Because of the scaling in volume, $P_{n,m}(X)$ are not true eigenfunctions. Nevertheless, they are very useful because they play an important role in the analysis of the evolution of the dynamical system which we will later associate with $T_L$. They are classified as expanding (relevant), contracting (irrelevant) or central (marginal), depending on whether the exponent of $L$ on the right-hand side of [19] is positive, negative, or zero, respectively. This depends, of course, on the space dimension $d$ and the field dimension $[\phi]$.

Gaussian measures are of limited interest. But we can create new measures by perturbing the Gaussian measure $\mu_C$ with local interactions. We cannot study directly the situation where the interactions are in infinite volume. Instead, we put them in a very large volume which will eventually go to infinity. We have a ratio of two length scales, one from the size of the diameter of the volume and the other from the ultraviolet cutoff in $\mu_C$, and this ratio is enormous. The RG is useful whenever there are two length scales whose ratio is very large. It permits us to do a scale-by-scale analysis and at each step the volume is reduced at the cost of changing the interactions. The largeness of the ratio is reflected in the large number of steps to be accomplished, this number tending eventually to infinity. This large number of steps has to be controlled mathematically.

### Perturbation of the Gaussian Measure

Let $\Lambda_N = [-L^N/2, L^N/2]^d \subset \mathbf{R}^d$ be a large cube in $\mathbf{R}^d$. For any $X \subset \Lambda_N$, let $V_0(X,\phi)$ be a local semibounded function where the fields are restricted to the set $X$. Here "local" means that if $X, Y$ are sets with disjoint interiors then $V_0(X \cup Y, \phi) = V_0(X) + V_0(Y)$. Consider the integral (known as the partition function in QFT and statistical mechanics)

$$Z(\Lambda_N) = \int \mathrm{d}\mu_C(\phi) z_0(\Lambda_N, \phi) \qquad [20]$$

where

$$z_0(X,\phi) = \mathrm{e}^{-V_0(X,\phi)} \qquad [21]$$

and

$$\mathrm{d}\mu^{(0)}(\Lambda_N,\phi) = \frac{1}{Z(\Lambda_N)} \mathrm{d}\mu_C(\phi) \mathrm{e}^{-V_0(\Lambda_N,\phi)} \qquad [22]$$

is the corresponding probability measure. $V_0$ is typically not quadratic in the fields and therefore leads to a non-Gaussian perturbation. For example,

$$V_0(X,\phi) = \int_X \mathrm{d}x (\xi |\nabla\phi(x)|^2 + g_0\phi^4(x) + \mu_0\phi^2(x)) \qquad [23]$$

where we take $g_0 > 0$. The integral [20] is well defined because the sample fields are smooth.

We now proceed to the scale-by-scale analysis mentioned earlier. Because $\mu_C$ is an invariant measure of $T_L$, we have the partition function $Z(\Lambda_N)$ in the volume $\Lambda_N$ as

$$Z(\Lambda_N) = \int \mathrm{d}\mu_C(\phi) z_0(\Lambda_N,\phi) = \int \mathrm{d}\mu_C(\phi) T_L z_0(\Lambda_N,\phi) \qquad [24]$$

The integrand on the right-hand side is a new function of fields which, because of the final scaling, live in the smaller volume $\Lambda_{N-1}$. This leads to the following definition:

$$z_1(\Lambda_{N-1},\phi) = T_L z_0(\Lambda_N,\phi) \qquad [25]$$

Because $V_0$ is local, $z_0$ has a factorization property for unions of sets with disjoint interiors. This is no longer the case for $z_1$. Wilson noted that, nevertheless, the integral is well approximated by an integrand which does, but the approximator has new coupling constants. The phrase "well approximated" is what all the rigorous work is about and this was not evident in the early Wilson era. The idea is to extract out a local part and also consider the remainder. The local part leads to a flow of coupling constants and the (unexponentiated) remainder is an irrelevant term. This operation and its mathematical control is an essential feature of RG analysis.

Iterating the above transformation, we get, for all $0 \le n \le N$,

$$z_{n+1}(\Lambda_{N-n-1},\phi) = T_L z_n(\Lambda_{N-n},\phi) \qquad [26]$$

After $N$ iterations, we get

$$Z(\Lambda_N) = \int \mathrm{d}\mu_C(\phi) z_N(\Lambda_0,\phi) \qquad [27]$$

where $\Lambda_0$ is the unit cube. To take the limit as $N \to \infty$, we have to control the infinite sequence of iterations. We cannot hope to control the infinite sequence at the level of the entire partition function. Instead, one chooses representative coordinates for which the infinite sequence has a chance of having a meaning. The coordinates are provided by the coupling constants of the extracted local part and the irrelevant terms (an

approximate calculation of the flow of coupling constants is given in the next section). The existence of a global trajectory for such coordinates helps us to control the limit for moments of the probability measure (correlation functions). The question of coordinates and the representation of the irrelevant terms will be taken up in the section "Rigorous RG analysis."

### Ultraviolet Cutoff Removal

The next issue is ultraviolet cutoff removal in field theory. This problem can be put into the earlier framework as follows. Let $\epsilon_N$ be a sequence of positive numbers which tend to 0 as $N \to \infty$. Following the remark after [3], we replace the unit cutoff covariance $C$ by the covariance $C_{\epsilon_N}$ defined by taking $\epsilon_N$ (instead of 1) as the lower end point in the integral [3]. Thus, $\epsilon_N$ acts as a short-distance or ultraviolet cutoff. It is easy to see that

$$C_{\epsilon_N}(x-y) = S_{\epsilon_N} C(x-y) \qquad [28]$$

Consider the partition function $Z_{\epsilon_N}(\Lambda)$ in a cube $\Lambda = [-R/2, R/2]^d$:

$$Z_{\epsilon_N}(\Lambda) = \int \mathrm{d}\mu_{C_{\epsilon_N}}(\phi) \mathrm{e}^{-V_0(\Lambda, \phi, \tilde{\xi}_N, \tilde{g}_N, \tilde{\mu}_N)} \qquad [29]$$

where $V_0$ is given by [23] with $g_0, \mu_0$ replaced by $\tilde{g}_N, \tilde{\mu}_N$, respectively. By dimensional analysis we can write

$$\begin{aligned} \tilde{\xi}_N &= \epsilon_N^{(2[\phi]-d+2)} \xi, \quad \tilde{g}_N = \epsilon_N^{(4[\phi]-d)} g, \\ \tilde{\mu}_N &= \epsilon_N^{(2[\phi]-d)} \mu \end{aligned} \qquad [30]$$

where $g, \xi, \mu$ are dimensionless parameters. Now $\phi$ distributed according to $C_{\epsilon_N}$ equals in distribution $S_{\epsilon_N}\phi$ distributed according to $C$. Therefore, choosing $\epsilon_N = L^{-N}$, we get

$$\begin{aligned} Z_{\epsilon_N}(\Lambda) &= \int \mathrm{d}\mu_C(\phi) \mathrm{e}^{-V_0(\Lambda, S_{\epsilon_N}\phi, \tilde{\xi}_N, \tilde{g}_N, \tilde{\mu}_N)} \\ &= \int \mathrm{d}\mu_C(\phi) \mathrm{e}^{-V_0(\Lambda_N, \phi, \xi g, \mu)} \end{aligned} \qquad [31]$$

where $\Lambda_N = [-L^N R/2, L^N R/2]^d$. Thus, the field theory problem of removing the ultraviolet cutoff, that is, taking the limit $\epsilon_N \to 0$, has been reduced to the study of a statistical mechanical model in a very large volume. The latter has to be analyzed via RG iterations as before.

### Critical Field Theories

As mentioned earlier, we have to study the flow of local interactions as well as that of irrelevant terms. Together they constitute the RG trajectory and we have to prove that it exists globally. In general, the trajectory will tend to explode after a large number of iterations due to growing relevant terms (characterized in terms of the expanding Wick monomials mentioned earlier). Wilson pointed out that the saving factor is to exploit fixed points and their invariant manifolds by tuning the initial interaction so that the RG has a global trajectory. This leads to the notion of a critical manifold which can be defined as follows. A fixed point will have contracting and/or marginal attractive directions besides the expanding ones. In the language of dynamical systems, the critical manifold is the stable or center stable manifold of the fixed point in question. This is determined by a detailed study of the discrete flow. In the examples above, it amounts to fixing the initial "mass" parameter $\mu_0 = \mu_c(g_0)$ with a suitable function $\mu_c$ such that the flow remains bounded in an invariant set. The critical manifold is then the graph of a function from the space of contracting and marginal variables to the space of $\mu$'s which remains invariant under the flow. Restricted to it the flow will now converge to a fixed point. All references to initial coupling constants have disappeared. The result is known as a critical theory.

Critical theories have been rigorously constructed in a number of cases. Take the standard $\phi^4$ in $d$ dimensions. Then $[\phi] = (d-2)/2$. For $d > 5$ the $\phi^4$ interaction is irrelevant and the Gaussian fixed point is attractive with one unstable direction (corresponding to $\mu$). In this case one can prove that the interactions converge exponentially fast to the Gaussian fixed point on the critical manifold. For $d = 4$ the interaction is marginal and the Gaussian fixed point attractive for $g > 0$. The critical theory has been constructed by Gawedzki and Kupainen (1984) starting with a sufficiently small coupling constant. The fixed point is Gaussian (interactions vanish in the limit) and the convergence rate is logarithmic. This is thus a mean-field theory with logarithmic corrections, as expected on heuristic grounds. The mathematical construction of the critical theory in $d = 3$ is an open problem. (It is expected to exist with a non-Gaussian fixed point, and this is indicated by the perturbative $\epsilon$ expansion of Wilson and Fisher in $4 - \epsilon$ dimensions.) However, the critical theory for $d = 3$ for $[\phi] = (3 - \epsilon)/4$ for $\epsilon > 0$ held very small has been rigorously constructed by Brydges *et al.* (2003). This theory has a nontrivial hyperbolic fixed point of $O(\epsilon)$. The stable manifold is constructed in a small neighborhood of the fixed point. Note that the covariance without cutoff is Osterwalder–Schrader positive and thus this is a candidate for a nontrivial EQFT. For $\epsilon = 1$ we have the standard situation in $d = 3$, and

this remains open, as mentioned earlier. A very simplified picture of the above is furnished by the perturbative computation in the next section.

### Unstable Fixed Points

We may attempt to construct field theories around unstable fixed points. In this case the initial parameters have to be adjusted as functions of the cutoff in such a way as to stabilize the flow in the neighborhood of the fixed point. This may be called a genuine renormalization. A famous example of this is pure Yang–Mills theory in $d=4$, where the Gaussian fixed point has only marginal unstable directions. Bałaban in a series of papers ending in Bałaban (1989) considered Wilson's lattice cutoff version of Yang–Mills theory in $d=4$ with initial coupling fixed by the two-loop asymptotic freedom formula. He proved, by lattice RG iterations, that in the weak-coupling regime the free energy per unit volume is bounded above and below by constants independent of the lattice spacing. Instability of the flow is expected to lead to mass generation for observables but this is a famous open problem. Another example is the standard nonlinear sigma model for $d=2$. Here too the flow is unstable around the Gaussian fixed point and we can set the initial coupling constant by the two-loop asymptotic freedom formula. Although much is known via approximation methods (as well as by methods based on integrable systems) this theory remains to be rigorously constructed as an EQFT.

Let us now consider a relatively simpler example, that of constructing a massive super-renormalizable scalar field theory. This has been studied in $d=3$, with $[\phi]=(d-2)/2=1/2$. We get $\xi=\tilde{\xi}, g=L^{-N}\tilde{g}, \mu=L^{-2N}\tilde{\mu}$, and $\tilde{g}$ is taken to be small. $\xi$ is marginal, whereas $g, \mu$ are relevant parameters and grow with the iterations. After $N$ iterations, they are brought up to $\tilde{g}, \tilde{\mu}$ together with remainders. This realizes the so-called massive continuum $\phi^4$ theory in $d=3$, and this has been mathematically controlled in the exact RG framework. This was proved by Brydges, Dimock, and Hurd and earlier by Benfatto, Cassandro, Gallavotti, and others, (see the references in Brydges *et al.* (1998) and Benfatto and Gallavotti (1995)).

## The Exact RG as a Continuous Semigroup

The discrete semigroup defined in [13] of the previous section has a natural continuous counterpart. Just take $L$ to be a continuous parameter, $L=e^t, t\geq 0$, and write by abuse of notation $T_t, S_t, \Gamma_t$ instead of $T_{e^t}$, etc. The continuous transformations $T_t$,

$$T_t F = S_t \mu_{\Gamma_t} * F \qquad [32]$$

give a semigroup

$$T_t T_s = T_{t+s} \qquad [33]$$

of contractions on $L^2(d\mu_C)$ with $\mu_C$ as invariant measure. One can show that $T_t$ is strongly continuous and, therefore, has a generator which we will call $\mathcal{L}$. This is defined by

$$\mathcal{L}F = \lim_{t\to 0^+} \frac{T_t - 1}{t} F \qquad [34]$$

whenever this limit exists. This restricts $F$ to a suitable subspace $\mathcal{D}(\mathcal{L}) \subset L^2(d\mu_C)$. $\mathcal{D}(\mathcal{L})$ contains, for example, polynomials in fields as well as twice-differentiable bounded cylindrical functions. The generator $\mathcal{L}$ can be easily computed. To state it, we need some definitions. Define $(D^n F)(\phi; f_1, \ldots, f_n)$ as the $n$th tangent map at $\phi$ along directions $f_1, \ldots, f_n$. The functional Laplacian $\Delta_{\dot{\Gamma}}$ is defined by

$$\Delta_{\dot{\Gamma}} F(\phi) = \int d\mu_{\dot{\Gamma}}(\zeta)(D^2F)(\phi;\zeta,\zeta) \qquad [35]$$

where $\dot{\Gamma}=u$. Define an infinitesimal dilatation operator

$$\mathcal{D}\phi(x) = x\cdot\nabla\phi(x) \qquad [36]$$

and a vector field $\mathcal{X}$,

$$\mathcal{X}F = -[\phi](DF)(\phi;\phi) - (DF)(\phi;\mathcal{D}\phi) \qquad [37]$$

Then, an easy computation gives

$$\mathcal{L} = \tfrac{1}{2}\Delta_{\dot{\Gamma}} + \mathcal{X} \qquad [38]$$

$T_t$ is a semigroup with $\mathcal{L}$ as generator. Therefore, $T_t = e^{t\mathcal{L}}$. Let $F_t(\phi)=T_tF(\phi)$. Then $F_t$ satisfies the linear PDE

$$\frac{\partial F_t}{\partial t} = \mathcal{L}F_t \qquad [39]$$

with the initial condition $F_0=F$. This evolution equation assumes a more familiar form if we write $F_t = e^{-V_t}$, $V_t$ being known as the effective potential. We get

$$\frac{\partial V_t}{\partial t} = \mathcal{L}V_t - \frac{1}{2}(V_t)_\phi\cdot(V_t)_\phi \qquad [40]$$

where

$$(V_t(\phi))_\phi\cdot(V_t(\phi))_\phi = \int d\mu_{\dot{\Gamma}}(\zeta)((DV_t)(\phi;\zeta))^2 \qquad [41]$$

and $V_0=V$. This infinite-dimensional nonlinear PDE is a version of Wilson's flow equation. Note that the linear semigroup $T_t$ acting on

functions induces a semigroup $\mathcal{R}_t$ acting nonlinearly on effective potentials giving a trajectory $V_t = \mathcal{R}_t V_0$.

Equations like the above are notoriously difficult to control rigorously, especially for large times. However, they may be solved in formal perturbation theory when the initial $V_0$ is small via the presence of small parameters. In particular, they give rise easily to perturbative flow equations for coupling constants. They can be obtained to any order but then there is the remainder. It is hard to control the remainder from the flow equation for effective potentials in bosonic field theories. They require other methods based on the discrete RG. Nevertheless, these approximate perturbative flows are very useful for getting a preliminary view of the flow. Moreover, their discrete versions figure as an input in further nonperturbative analysis.

### Perturbative Flow

It is instructive to see this in second-order perturbation theory. We will simplify by working in infinite volume (no infrared divergences can arise because $\dot{\Gamma}(x-y)$ is of fast decrease). Now suppose that we are in standard $\phi^4$ theory with $[\phi] = (d-2)/2$ and $d > 2$. We want to show that

$$V_t = \int \mathrm{d}x \Big( \xi_t : |\nabla\phi(x)|^2 : + g_t : \phi(x)^4 : + \mu_t : \phi(x)^2 : \Big) \qquad [42]$$

satisfies the flow equation in second order modulo irrelevant terms provided the parameters flow correctly. We will ignore field-independent terms. The Wick ordering is with respect to the covariance $C$ of the invariant measure. The reader will notice that we have ignored a $\phi^6$ term which is actually relevant in $d=3$ for the above choice of $[\phi]$. This is because we will only discuss the $d=3$ case for the model discussed at the end of this section and for this case the $\phi^6$ term is irrelevant. We will assume that $\xi_t, \mu_t$ are of order $O(g^2)$. Plug in the above in the flow equation. The quantity $\lambda_t^{n,m}{:}P_{n,m}{:}$ represents one of the terms above with $m$ fields and $n$ derivatives. Because $\mathcal{L}$ is the generator of the semigroup $T_t$ we have

$$\left(\frac{\partial}{\partial t} - \mathcal{L}\right)\lambda_t^{n,m} : P_{n,m} : = \left(\frac{\mathrm{d}\lambda_t^{n,m}}{\mathrm{d}t} - (d - m[\phi] - n)\lambda_t^{n,m}\right) : P_{n,m} : \qquad [43]$$

Next turn to the nonlinear term in the flow equation and insert the $\phi^4$ term (the others are already of order $O(g^2)$). This produces a double integral of $\dot{\Gamma}(x-y){:}\phi(x)^3{:}{:}\phi(y)^3{:}$, which after complete Wick ordering, gives

$$-\frac{g_t^2}{2} 16 \int \mathrm{d}x\,\mathrm{d}y\,\dot{\Gamma}(x-y)\Big( : \phi(x)^3\phi(y)^3 : + 9C(x-y) : \phi(x)^2\phi(y)^2 : + 36C(x-y) : \phi(x)\phi(y) : + 6C(x-y)^2 \Big) \qquad [44]$$

Consider the nonlocal $\phi^4$ term. We can localize it by writing

$$: \phi(x)^2\phi(y)^2 : = \tfrac{1}{2} : \Big( \phi(x)^4 + \phi(y)^4 - \big(\phi(x)^2 - \phi(y)^2\big)^2 \Big) : \qquad [45]$$

The local part gives a $\phi^4$ contribution and the last term above gives rise to an irrelevant contribution because it produces additional derivatives. The coefficients are well defined because $C, \dot{\Gamma}$ are smooth and $\dot{\Gamma}(x-y)$ is of fast decrease. Now the nonlocal $\phi^2$ term is similarly localized. It gives a relevant local $\phi^2$ contribution as well as a marginal $|\nabla\phi|^2$ contribution. Finally, the same principle applies to the nonlocal $\phi^6$ contribution and gives rise to further irrelevant terms. Then it is easy to see by matching that the flow equation is satisfied in second order up to irrelevant terms (these would have to be compensated by adding additional terms in $V_t$) provided

$$\begin{aligned} \frac{\mathrm{d}g_t}{\mathrm{d}t} &= (4-d)g_t - a g_t^2 + O\big(g_t^3\big) \\ \frac{\mathrm{d}\mu_t}{\mathrm{d}t} &= 2\mu_t - b g_t^2 + O\big(g_t^3\big) \\ \frac{\mathrm{d}\xi_t}{\mathrm{d}t} &= c g_t^2 + O\big(g_t^3\big) \end{aligned} \qquad [46]$$

where $a, b, c$ are positive constants. We see from the above formulas that, up to second order in $g^2$, as $t \to \infty, g_t \to 0$ for $d \geq 4$. In fact, for $d \geq 5$ the decay rate is $O(\mathrm{e}^{-t})$ and for $d=4$ the rate is $O(t^{-1})$. However, to see if $V_t$ converges, we also have to discuss the $\mu_t, \xi_t$ flows. It is clear that in general the $\mu_t$ flow will diverge. This is fixed by choosing the initial $\mu_0$ to be the bare critical mass. This is obtained by integrating up to time $t$ and then expressing $\mu_0$ as a function of the entire $g$ trajectory up to time $t$. Assume that $\mu_t$ is uniformly bounded and take $t \to \infty$. This gives the critical mass as

$$\mu_0 = b\int_0^\infty \mathrm{d}s\, \mathrm{e}^{-2s} g_s^2 = \mu_c(g_0) \qquad [47]$$

This integral converges for all cases discussed above. With this choice of $\mu_0$ we get

$$\mu_t = b\int_0^\infty \mathrm{d}s\, \mathrm{e}^{-2s} g_{s+t}^2 \qquad [48]$$

and this exists for all $t$ and converges as $t \to \infty$. Now consider the perturbative $\xi$ flow. It is easy to see from the above that for $d \geq 4, \xi_t$ converges as $t \to \infty$.

We have not discussed the $d=3$ case because the perturbative $g$ fixed point is of order $O(1)$. But suppose we take, in the $d=3$ case, $[\phi]=(3-\epsilon)/4$ with $\epsilon > 0$ held small as in Brydges *et al.* (2003). Then the above perturbative flow equations are easily modified (by taking account of [43]) and we get, to second order, an attractive fixed point $g_* = O(\epsilon)$ of the $g$ flow. The critical bare mass $\mu_0$ can be determined as before and the $\xi_t$ flow converges. The qualitative picture obtained above has a rigorous justification.

## Rigorous RG Analysis

We will give a brief introduction to rigorous RG analysis in the discrete setup in the section "The RG as a Discrete Semigroup" concentrating on the principal problems encountered and how one attempts to solve them. Our approach is borrowed from Brydges *et al.* (2003). It is a simplification of the methods initiated by Brydges and Yau in (1990) and developed further by Brydges *et al.* (1998). The reader will find other approaches to rigorous RG methods in the selected references, such as those of Bałaban, Gawedzki and Kupiainen, Gallavotti, and others. We will take as a concrete example the scalar field model introduced earlier.

At the core of the analysis is the choice of good coordinates for the partition function density, $z$, of the section "The RG as a Discrete Semigroup". This is provided by a polymer representation (defined below) which parametrizes $z$ by a couple $(V, K)$, where $V$ is a local potential and $K$ is a set function also depending on the fields. Then the RG transformation $T_L$ maps $(V, K)$ to a new $(V, K)$. $(V, K)$ remain good coordinates as the volume tends to $\infty$, whereas $z$(volume) diverges. There exist norms which are suited to the fixed-point analysis of $(V, K)$ to new $(V, K)$. Now comes the important point: $z$ does not uniquely specify the representation $(V, K)$. Therefore, we can take advantage of this nonuniqueness to keep $K$ small in norm and let most of the action of $T_L$ reside in $V$. This process is called extraction in Brydges *et al.* (2003). It makes sure that $K$ is an irrelevant term, whereas the local flow of $V$ gives rise to discrete flow equations in coupling constants. We will not discuss extraction any further. In the following, we introduce the polymer representation and explain how the RG transformation acts on it.

To proceed further, we first introduce a simplification in the setup used in the section "The RG as a Discrete Semigroup." Recall that the function $u$ introduced in [3] was smooth, positive definite, and of rapid decrease. We will simplify further by imposing the stronger property that it is actually of finite range: $u(x)=0$ for $|x| \geq 1$. We say that $u$ is of finite range 1. It is easy to construct such functions. For example, if $g$ is any smooth function of finite range 1/2, then $u = g * g$ is a smooth positive-definite function of finite range 1. This implies that the fluctuation covariance $\Gamma_L$ of [7] has finite range $L$. As a result, $\Gamma_n$ in [10] has finite range $L^{n+1}$ and the corresponding fluctuation fields $\zeta_n(x)$ and $\zeta_n(y)$ are independent when $|x-y| \geq L^{n+1}$.

### Polymer Representation

Pave $\mathbf{R}^d$ with closed cubes of side length 1 called 1-blocks or unit blocks denoted by $\Delta$, and suppose that $\Lambda$ is a large cube consisting of unit blocks. A connected polymer $X \subset \Lambda$ is a closed connected subset of these unit blocks. A polymer activity $K(X, \phi)$ is a map $X, \phi \to \mathbf{R}$ where the fields $\phi$ depend only on the points of $X$. We will set $K(X, \phi)=0$ if $X$ is not connected. A generic form of the partition function density $z(\Lambda, \phi)$ after a certain number of RG iterations is

$$z(\Lambda) = \sum_{N=0}^{\infty} \frac{1}{N!} \sum_{X_1, \ldots, X_N} e^{-V(X_c)} \prod_{j=1}^{N} K(X_j) \qquad [49]$$

Here $X_j \subset \Lambda$ are disjoint polymers, $X = \bigcup X_j$, and $X_c = \Lambda \backslash X$. $V$ is a local potential of the form [23] with parameters $\xi, g, \mu$. We have suppressed the $\phi$-dependence. Initially, the activities $K_j = 0$, but they will arise under RG iterations and the form [49] remains stable, as we will see. The partition function density is thus parametrized as a couple $(V, K)$.

### Norms for Polymer Activities

Polymer activities $K(X, \phi)$ are endowed with a norm $\|K(X)\|$, which must satisfy two properties:

$$\dot{X} \cap \dot{Y} = 0 \Rightarrow \|K_1(X) K_2(Y)\| \leq \|K_1(X)\| \, \|K_1(Y)\|$$
$$\|T_L K(X)\| \leq c^{|X|} \|K(X)\| \qquad [50]$$

where $\dot{X}$ is the interior of $X$ and $|X|$ is the number of blocks in $X$. $c$ is a constant of order $O(1)$. The norm measures (Fréchét) differentiability properties of the activity $K(X, \phi)$ with respect to the field $\phi$ as well as its admissible growth in $\phi$. The growth is admissible if it is $\mu_C$ integrable. The second property above ensures the stability of the norm under RG iteration. For a fixed polymer $X$, the norm is such that it gives rise to a Banach

space of activities $K(X)$. The final norm $\|\cdot\|_{\mathcal{A}}$ incorporates the previous one and washes out the set dependence,

$$\|K\|_{\mathcal{A}} = \sup_{\Delta} \sum_{X \supset \Delta} A(X) \, \|K(X)\| \qquad [51]$$

where $A(X) = L^{(d+2)|X|}$. This norm essentially ensures that large polymers have small activities. The details of the above norms can be found in Brydges *et al.* (2003).

The RG operation map $f$ is a composition of two maps. The RG iteration map $z \to T_L z$ induces a map $V \to \tilde{V}'_L$ and a nonlinear map $\tilde{T}_L : K \to \tilde{K} = \tilde{T}_L(K)$. We then compose this with a (nonlinear) extraction map $\mathcal{E}$ which takes out the expanding (relevant) parts of $\tilde{K} \to \mathcal{E}(\tilde{K}) = K'$ and compensates the local potential $\tilde{V}'_L \to V'$ such that $T_L z$ remains invariant. We denote by $f$ the composition of these two maps with

$$V \to V' = f_V(V, K), \quad K \to K' = f_K(V, K) \qquad [52]$$

### The Map $\tilde{T}_L$

Consider applying the RG map $T_L$ to [49]. The map consists of a convolution $\mu_{\Gamma_L} *$ followed by the rescaling $S_L$. In the integration over the fluctuation field $\zeta$, we will exploit the independence of $\zeta(x)$ and $\zeta(y)$ when $|x - y| \geq L$. To do this, we pave $\Lambda$ by closed blocks of side $L$, called $L$-blocks, so that each $L$-block is a union of 1-blocks. Let $\bar{X}^L$ be the $L$-closure of a set $X$, namely the smallest union of $L$-blocks containing $X$. The polymers will be combined into $L$-polymers which are, by definition, connected unions of $L$ blocks. The combination is performed in such a way that the new polymers are associated to independent functionals of $\zeta$.

Let $\tilde{V}(X, \phi))$, to be chosen later, be a local potential independent of $\zeta$. For a coupling constant sufficiently small, there is a bound

$$\|e^{-V(Y)}\| \leq 2^{|Y|} \qquad [53]$$

We assume that $\tilde{V}$ is so chosen that the same bound holds when $V$ is replaced by $\tilde{V}$. Define

$$P(\Delta, \zeta, \phi) = e^{-V(\Delta, \zeta+\phi)} - e^{-\tilde{V}(\Delta, \phi)} \qquad [54]$$

Then we have

$$\begin{aligned} e^{-V(X_c, \zeta+\phi)} &= e^{-V(\bar{X}_c, \zeta+\phi)} \\ &= \prod_{\Delta \subset \bar{X}_c} (e^{-\tilde{V}(\Delta, \phi)} + P(\Delta, \zeta, \phi)) \end{aligned} \qquad [55]$$

where $\bar{X}_c$ is the closure of $X_c$. Expand out the product and insert into the representation [49] for $z(\Lambda, \zeta + \phi))$. We then rewrite the resulting sum in terms of $L$-polymers. The sum splits into a sum over connected components. Define, for every connected $L$-polymer $Y$,

$$\begin{aligned} \mathcal{B}K(Y) = &\sum_{N+M \geq 1} \frac{1}{N!M!} \\ &\times \sum_{(X_j),(\Delta_i) \to Y} e^{-\tilde{V}(X_0)} \prod_{j=1}^{N} K(X_j) \prod_{i=1}^{M} P(\Delta_i) \end{aligned} \qquad [56]$$

where $X_0 = Y \backslash (\cup X_j) \cup (\cup \Delta_j)$ and the sum over the distinct $\Delta_i$, and disjoint 1-polymers $X_j$ is such that their $L$-closure is $Y$. Equation [49] now becomes

$$z(\Lambda) = \sum \frac{1}{N!} \sum_{Y_1, \ldots, Y_N} e^{-\tilde{V}(Y_c)} \prod_{j=1}^{N} \mathcal{B}K(Y_j) \qquad [57]$$

where the sum is over disjoint, connected closed $L$-polymers. We now perform the fluctuation integration over $\zeta$ followed by the rescaling. Now $\tilde{V}(Y_c)$ is independent of $\zeta$. The $\zeta$-integration sails through and then factorizes because the $Y_j$, being disjoint closed $L$-polymers, are separated from each other by a distance $\geq L$. The rescaling brings us back to 1-polymers and reduces the volume from $\Lambda$ to $L^{-1}\Lambda$. Therefore,

$$\begin{aligned} z'(L^{-1}\Lambda) &= T_L z(\Lambda) \\ &= \sum \frac{1}{N!} \sum_{X_1, \ldots, X_N} e^{-\tilde{V}_L(X_c)} \prod_{j=1}^{N} (T_L \mathcal{B}K)(X_j) \end{aligned} \qquad [58]$$

where the sum is over disjoint 1-polymers, $X_c = L^{-1}\Lambda \backslash X$. By definition $\tilde{V}_L(\Delta) = S_L \tilde{V}(L\Delta)$ and $(T_L \mathcal{B}K)(Z) = S_L \mu_{\Gamma_L} * \mathcal{B}K(LZ)$. This shows that the representation [49] is stable under iteration and, furthermore, gives us the map

$$\begin{aligned} &V \to \tilde{V}_L \\ &K \to \tilde{K} = \tilde{T}_L(K) = T_L \mathcal{B}K \end{aligned} \qquad [59]$$

The norm boundedness of $K$ implies that $\tilde{T}_L(K)$ is norm bounded. We see from the above that a variation in the choice of $\tilde{V}$ is reflected in the corresponding variation of $\tilde{K}$. The extraction map $\mathcal{E}$ now takes out from $\tilde{K}$ the expanding parts and then compensates it by a change of $\tilde{V}_L$ in such a way that the representation [58] is left invariant by the simultaneous replacement $\tilde{V}_L \to V', \tilde{K} \to K' = \mathcal{E}(\tilde{K})$. The extraction map is nonlinear. Its linearization is a subtraction operation and this dominates in norm the nonlinearities, (Brydges *et al.* 1998).

The map $V \to \tilde{V}_L \to V'$ leads to a discrete flow of the coupling constants in $V$. It is convenient to write $K = K^{\mathrm{pert}} + R$, where $R$ is the remainder. Then the coupling constant flow is a discrete version of the

continuous flows encountered in the last section, together with remainders which are controlled by the size of $R$. In addition, we have the flow of $K$. The discrete flow of the pair (coupling constants, $K$) can be studied in a Banach space norm. Once one proves that the nonlinear parts satisfy a Lipshitz property, the discrete flow can be analyzed by the methods of stable-manifold theory of dynamical systems in a Banach space context. The reader is referred to the article by Brydges *et al.* (2003) for details of the extraction map and the application of stable-manifold theory in the construction of a global RG trajectory.

## Further Topics

### Lattice RG Methods

Statistical mechanical systems are often defined on a lattice. Moreover, the lattice provides an ultraviolet cutoff for Euclidean field theory compatible with Osterwalder–Schrader positivity. The standard lattice RG is based on Kadanoff–Wilson block spins. Its mathematical theory and applications have been developed by Bałaban, and Gawedzki and Kupiainen (see Gawedzki and Kupiainen (1986) and references therein). This leads to multiscale decompositions of the Gaussian lattice field as a sum of independent fluctuation fields on increasing length scales. Brydges *et al.* (2004) have shown that standard Gaussian lattice fields have multiscale decompositions as a sum of independent fluctuation fields with the finite-range property introduced in the last section. This permits the development of rigorous lattice RG theory in the spirit of the continuum framework of the previous section.

### Fermionic Field Theories

Field theories of interacting fermions are often simpler to handle than bosonic field theories. Because of statistics, fermion fields are bounded and perturbation series converges in finite volume in the presence of an ultraviolet cutoff. The notion of studying the RG flow at the level of effective potentials makes sense. At any given scale, there is always an ultraviolet cutoff and the fluctuation covariance being of fast decrease provides an infrared cutoff. This is illustrated by the work of Gawedzki and Kupiainen (1985), who gave a nonperturbative construction in the weak effective coupling regime of the RG trajectory for the Gross–Neveu model in two dimensions. This is an example of a model with an unstable Gaussian fixed point where the initial coupling has to be adjusted as a function of the ultraviolet cutoff consistent with ultraviolet asymptotic freedom so as to stabilize the flow.

## Acknowledgments

The author is grateful to David Brydges for very helpful comments on a preliminary version of this article.

*See also:* Operator Product Expansion in Quantum Field Theory; Renormalization: Statistical Mechanics and Condensed Matter.

## Further Reading

Bałaban T (1982) $(Higgs)_{2,3}$ quantum fields in a finite volume II, an upper bound. *Communications in Mathematical Physics* 86: 555–594.

Bałaban T (1989) Large field renormalization. 2: Localization, exponentiation and bounds for the *R* operation. *Communications in Mathematical Physics* 122: 355–392.

Benfatto G and Gallavotti G (1995) *Renormalization Group*. Princeton, NJ: Princeton University Press.

Brydges DC and Yau HT (1990) Grad $\phi$ perturbations of massless Gaussian fields. *Communication in Mathematical Physics* 129: 351–392.

Brydges D, Dimock J, and Hurd T (1998) Estimates on renormalization group transformations. *Canadian Journal of Mathematics* 50(4): 756–793.

Brydges DC, Mitter PK, and Scoppola B (2003) Critical $(\Phi^4)_{3,\epsilon}$. *Communications in Mathematical Physics* 240: 281–327.

Brydges D, Guadagni G, and Mitter PK (2004) Finite range decomposition of Gaussian processes. *Journal of Statistical Physics* 115: 415–449.

Gawedzki K and Kupiainen A (1984) Massless lattice $\phi^4$ theory in four dimensions: a nonperturbative control of a renormalizable model. *Communications in Mathematical Physics* 99: 197–252.

Gawedzki K and Kupiainen A (1985) Gross–Neveu models through convergent perturbation expansions. *Communications in Mathematical Physics* 102: 1–30.

Gawedzki K and Kupiainen A (1986) Asymptotic freedom beyond perturbation theory. In: Osterwalder K and Stora R (eds.) *Critical Phenomena, Random Systems, Gauge Theories*. North-Holland: Les Houches.

Gelfand IM and Vilenkin NYa (1964) *Generalized Functions*, vol. 4. New York: Academic Press.

Kopper C and Muller V (2000) Renormalization proof for spontaneously broken Yang–Mills theory with flow equations. *Communications in Mathematical Physics* 209: 477–516.

Polchinski J (1984) Renormalization and effective Lagrangians. *Nuclear Physics B* 231: 269–295.

Rivasseau V (1991) *From Perturbative to Constructive Renormalization*. Princeton NJ: Princeton University Press.

Wilson KG (1983) Renormalization group and critical phenomena. *Reviews of Modern Physics* 55: 583–599.

Wilson KG and Kogut J (1974) Renormalization group and the $\epsilon$ expansion. *Physics Reports* 12C: 75–200.

# F

## Falicov–Kimball Model

**Ch Gruber**, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
**D Ueltschi**, University of Arizona, Tucson, AZ, USA



### A Brief History

The "Falicov–Kimball model" was first considered by Hubbard and Gutzwiller during 1963–65 as a simplification of the Hubbard model. In 1969, Falicov and Kimball introduced a model that included a few extra complications, in order to investigate metal–insulator phase transitions in rare-earth materials and transition-metal compounds (Falicov and Kimball 1969). Experimental data suggested that this transition is due to the interactions between electrons in two electronic states: nonlocalized states (itinerant electrons), and states that are localized around the sites corresponding to the metallic ions of the crystal (static electrons).

A tight-binding approximation leads to a model defined on a lattice (the crystal) and two species of particles are considered. The first species consists of spinless quantum fermions (we refer to them as "electrons"), and the second species consists of localized holes or electrons ("classical particles"). Electrons hop between nearest-neighbor sites but classical particles do not. Both species obey Fermi statistics (in particular, the Pauli exclusion principle prevents more than one particle of a given species to occupy the same site). Interactions are on-site and thus involve particles of different species; they can be repulsive or attractive.

The very simplicity of the model allows for a broad range of applications. It was studied in the context of mixed valence systems, binary alloys, and crystal formation. Adding a magnetic field yields the flux phase problem. The Falicov–Kimball model can also be viewed as the simplest model where quantum particles interact with classical fields.

The fifteen years following the introduction of the model saw studies based on approximate methods, such as Green's function techniques, that gave rise to a lot of confusion. A breakthrough occurred in 1986 when Brandt and Schmidt, and Kennedy and Lieb, proposed the first rigorous results. In particular, Kennedy and Lieb showed in their beautiful paper that the electrons create an effective interaction between the classical particles and that a phase transition takes place for any value of the coupling constant, provided the temperature is low enough.

Many studies by mathematical physicists followed and several results are presented in this short survey. Recent years have seen an increasing interest from condensed matter physicists. We encourage interested readers to consult the reviews by Freericks and Zlatić (2003), Gruber and Macris (1996), and Jędrzejewski and Lemánski (2001).

### Mathematical Setting

#### Definitions

Let $\Lambda \subset \mathbb{Z}^d$ denote a finite cubic box. The configuration space for the classical particles is

$$\Omega_\Lambda = \{0,1\}^\Lambda = \{\omega = (\omega_x) : x \in \Lambda, \text{ and } \omega_x = 0,1\}$$

where $\omega_x = 0$ or 1 denotes the absence or presence of a classical particle at the site $x$. The total number of classical particles is $N_c(\omega) = \sum_{x\in\Lambda} \omega_x$. The Hilbert space for the spinless quantum particles ("electrons") is the usual fermionic Fock space

$$\mathcal{F}_\Lambda = \bigoplus_{N=0}^{|\Lambda|} \mathcal{H}_{\Lambda,N}$$

where $\mathcal{H}_{\Lambda,N}$ is the Hilbert space of square summable, antisymmetric, complex functions $\Psi = \Psi(x_1, \ldots, x_N)$ of $N$ variables $x_i \in \Lambda$. Let $a_x^\dagger$ and $a_x$ denote the standard creation and annihilation operators of an electron at $x$; recall that they satisfy the anticommutation relations

$$\{a_x, a_y\} = 0, \qquad \{a_x^\dagger, a_y^\dagger\} = 0, \qquad \{a_x, a_y^\dagger\} = \delta_{xy}$$

The Hamiltonian for the Falicov–Kimball model is an operator on $\mathcal{F}_\Lambda$ that depends on the configurations of classical particles. Namely, for $\omega \in \Omega_\Lambda$, we define

$$H_\Lambda(\omega) = -\sum_{\substack{x,y\in\Lambda \\ |x-y|=1}} a_x^\dagger a_y - U \sum_{x\in\Lambda} \omega_x a_x^\dagger a_x$$

The first term represents the kinetic energy of the electrons. The second term represents the on-site attraction ($U > 0$) or repulsion ($U < 0$) between electrons and classical particles.

The Falicov–Kimball Hamiltonian can be written with the help of a one-body Hamiltonian $h_\Lambda$, which is an operator on the Hilbert space for a single electron $\ell^2(\Lambda)$. Indeed, we have

$$H_\Lambda(\omega) = \sum_{x,y\in\Lambda} h_{xy}(\omega) a_x^\dagger a_y$$

The matrix $h_\Lambda(\omega) = (h_{xy}(\omega))$ is the sum of a hopping matrix (adjacency matrix) $t_\Lambda$, and of a matrix $v_\Lambda(\omega)$ that represents an external potential due to the classical particles. Namely, we have

$$h_{xy}(\omega) = -t_{xy} - U\omega_x\delta_{xy}$$

where $t_{xy}$ is one if $x$ and $y$ are nearest neighbors, and is zero otherwise. The spectrum of $t_\Lambda$ lies in $(-2d, 2d)$, and the eigenvalues of $v_\Lambda(\omega)$ are $-U$ (with degeneracy $N_c(\omega)$) and 0 (with degeneracy $|\Lambda| - N_c(\omega)$). Denoting $\lambda_j(A)$ the eigenvalues of a matrix $A$, it follows from the minimax principle that

$$\lambda_j(A) - \|B\| \le \lambda_j(A+B) \le \lambda_j(A) + \|B\|$$

Let $\lambda_1(\omega) \le \lambda_2(\omega) \le \cdots \le \lambda_{|\Lambda|}(\omega)$ be the eigenvalues of $h_\Lambda(\omega)$. Choosing $A = v_\Lambda(\omega)$ and $B = t_\Lambda$ in the inequality above, we find that for $U > 0$,

$$-U - 2d < \lambda_j(\omega) < -U + 2d \quad \text{for } j = 1, \ldots, N_c(\omega)$$
$$-2d < \lambda_j(\omega) < 2d \quad \text{for } j = N_c(\omega) + 1, \ldots, |\Lambda|$$

In particular, for any configuration $\omega$ and any $\Lambda$,

$$\operatorname{Spec} h_\Lambda(\omega) \subset (-U - 2d, -U + 2d) \cup (-2d, 2d)$$

Thus, for $U > 4d$, the spectrum of $h_\Lambda(\omega)$ has the "universal" gap $(-U + 2d, -2d)$. A similar property holds for $U < -4d$.

## Canonical Ensemble

A fruitful approach towards understanding the behavior of the Falicov–Kimball model is to first fix the configuration of the classical particles, and then to introduce the ground-state energy $E_\Lambda(N_e, \omega)$ as the lowest eigenvalue of $H_\Lambda(\omega)$ in the subspace $\mathcal{H}_{\Lambda, N_e}$:

$$E_\Lambda(N_e, \omega) = \inf_{\Psi\in\mathcal{H}_{\Lambda,N_e}, \|\Psi\|=1} \langle\Psi|H_\Lambda(\omega)|\Psi\rangle = \sum_{j=1}^{N_e} \lambda_j(\omega)$$

A typical problem is to find the set of ground-state configurations, that is, the set of configurations that minimize $E_\Lambda(N_e, \omega)$ for given $N_e$ and $N_c = N_c(\omega)$.

In the case $U > 4d$ and $N_e = N_c(\omega)$, the ground-state energy $E_\Lambda(N_c(\omega), \omega)$ has a convergent expansion in powers of $U^{-1}$:

$$E_\Lambda(N_c(\omega), \omega) = -UN_c(\omega) + \sum_{k\ge2} \frac{1}{kU^{k-1}} \times \sum_{\substack{x_1,\ldots,x_k \in \Lambda \\ |x_i - x_{i-1}| = 1 \\ 0 < m(\{x_i\}) < k}} (-1)^{m(\{x_i\})} \binom{k-2}{m(\{x_i\}) - 1} \qquad [1]$$

where $m(x_1, \ldots, x_k)$ is the number of sites $x_i$ with $\omega_{x_i} = 0$. The last sum also includes the condition $|x_k - x_1| = 1$. Simple estimates show that the series is less than $(2d/(U - 4d))N_c(\omega)$. The lowest-order term is a nearest-neighbor interaction,

$$-\frac{1}{U} \sum_{\{x,y\}:|x-y|=1} \delta_{w_x, 1-w_y}$$

that favors pairs with different occupation numbers. Formula [1] is the starting point for most studies of the phase diagram for large $U$. A similar expansion holds for $U < -4d$ and $N_e = |\Lambda| - N_c(\omega)$.

A simple derivation of expansion [1] using Cauchy formula can be found in Gruber and Macris (1996). It can be extended to positive temperatures with the help of Lie–Schwinger series (Datta *et al.* 1999).

Phase diagrams are better discussed in the limit of infinite volumes where boundary effects can be discarded. Let $\Omega^{\text{per}}$ be the set of configurations on $\mathbb{Z}^d$ that are periodic in all $d$ directions, and $\Omega^{\text{per}}(\rho_c) \subset \Omega^{\text{per}}$ be the set of periodic configurations with density $\rho_c$. For $\omega \in \Omega^{\text{per}}$ and $\rho_e \in [0, 1]$, we introduce the energy per site in the infinite volume limit by

$$e(\rho_e, \omega) = \lim_{\Lambda\nearrow\mathbb{Z}^d} \frac{1}{|\Lambda|} E_\Lambda(N_e, \omega) \qquad [2]$$

Here, the limit is taken over any sequence of increasing cubes, and $N_e = \lfloor \rho_e|\Lambda| \rfloor$ is the integer part of $\rho_e|\Lambda|$. Existence of this limit follows from standard arguments.

In the case of the empty configuration $\omega_x \equiv 0$, we get the well-known energy per site of free lattice electrons: for $k \in [-\pi, \pi]^d$, let $\varepsilon(k) = -\sum_{\nu=1}^d \cos k_\nu$; then

$$e(\rho_e, \omega \equiv 0) = \frac{1}{(2\pi)^d} \int_{\varepsilon(k)<\varepsilon_F(\rho_e)} \varepsilon(k)\, dk$$

where $\varepsilon_F(\rho_e)$ is the Fermi energy, defined by

$$\rho_e = \frac{1}{(2\pi)^d} \int_{\varepsilon(k)<\varepsilon_F(\rho_e)} dk$$

The other simple situation is the full configuration $\omega_x \equiv 1$, whose energy is $e(\rho_e, \omega \equiv 1) = e(\rho_e, \omega \equiv 0) - U\rho_e$.

Let $e(\rho_e, \rho_c)$ denote the absolute ground-state energy density, namely,

$$e(\rho_e, \rho_c) = \inf_{\omega \in \Omega^{per}(\rho_c)} e(\rho_e, \omega)$$

Notice that $e(\rho_e, \omega)$ is convex in $\rho_e$, and that $e(\rho_e, \rho_c)$ is the convex envelope of $\{e(\rho_e, \omega) : \omega \in \Omega^{per}(\rho_c)\}$. It may be locally linear around some $(\rho_e, \rho_c)$. This is the case if the infimum is not realized by a periodic configuration. The nonperiodic ground states can be expressed as linear combinations of two or more periodic ground states ("mixtures"). That is, for $1 \leq i \leq n$ there are $\alpha_i \geq 0$ with $\sum_i \alpha_i = 1$, $\omega^{(i)} \in \Omega^{per}$, and $\rho_e^{(i)}$, such that

$$\rho_e = \sum_i \alpha_i \rho_e^{(i)}, \qquad \rho_c = \sum_i \alpha_i \rho_c(\omega^{(i)})$$

and

$$e(\rho_e, \rho_c) = \sum_i \alpha_i e(\rho_e^{(i)}, \omega^{(i)})$$

The simplest mixture is the "segregated state" for densities $\rho_e < \rho_c$: take $\omega^{(1)}$ to be the empty configuration, $\omega^{(2)}$ to be the full configuration, $\rho_e^{(1)} = 0$, $\rho_e^{(2)} = \rho_e/\rho_c$, and $\alpha_2 = 1 - \alpha_1 = \rho_c$.

If $d \geq 2$, a mixture between configurations $\omega^{(i)}$ can be realized as follows. First, partition $\mathbb{Z}^d$ into domains $D_1 \cup \cdots \cup D_n$ such that $|D_i|/|\Lambda| \to \alpha_i$ and $|\partial D_i|/|\Lambda| \to 0$ as $\Lambda \nearrow \mathbb{Z}^d$. Then, define a nonperiodic configuration $\omega$ by setting $\omega_x = \omega_x^{(i)}$ for $x \in D_i$ (see the illustration in Figure 1). The canonical energy can be computed from [2], and it is equal to

$$e(\rho_e, \omega) = \inf_{(\rho_e^{(i)}): \sum_i \alpha_i \rho_e^{(i)} = \rho_e} \sum_{i=1}^{n} \alpha_i e(\rho_e^{(i)}, \omega^{(i)})$$

Furthermore, the infimum is realized by densities $\rho_e^{(i)}$ such that there exists $\mu_e$ with $\rho_e(\mu_e, \omega^{(i)}) = \rho_e^{(i)}$ for all $i$ (see [4] below for the definition of $\rho_e(\mu_e, \omega)$).

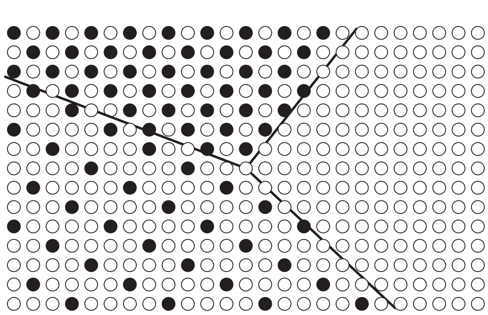


**Figure 1** A two-dimensional mixed configuration formed by periodic configurations of densities $0, 1/5$, and $1/2$.

We define the *canonical ground-state phase diagram* as the set of ground states $\omega$ (either a periodic configuration or a mixture) that minimize the ground-state energy for given densities $\rho_e, \rho_c$:

$$G_{can}(\rho_e, \rho_c) = \{\omega : e(\rho_e, \omega) = e(\rho_e, \rho_c) \text{ and } \rho_c(\omega) = \rho_c\}$$

### Grand-Canonical Ensemble

Properties of the system at finite temperatures are usually investigated within the grand-canonical formalism. The equilibrium state is characterized by an inverse temperature $\beta = 1/k_B T$, and by chemical potentials $\mu_e, \mu_c$, for the electrons and for the classical particles, respectively. In this formalism, the thermodynamic properties are derived from the partition functions

$$\begin{aligned} Z_\Lambda(\beta, \mu_e, \omega) &= \operatorname{tr}_{\mathcal{F}_\Lambda} e^{-\beta[H_\Lambda(\omega) - \mu_e N_\Lambda]} \\ Z_\Lambda(\beta, \mu_e, \mu_c) &= \sum_{\omega \in \Omega_\Lambda} e^{\beta \mu_c N_c(\omega)} Z_\Lambda(\beta, \mu_e, \omega) \end{aligned} \qquad [3]$$

Here, $N_\Lambda = \sum_{x \in \Lambda} a_x^\dagger a_x$ is the operator for the total number of electrons. We then define the free energy by

$$F_\Lambda(\beta, \mu_e, \mu_c) = -\frac{1}{\beta} \log Z_\Lambda(\beta, \mu_e, \mu_c)$$

The first partition function in [3] allows us to introduce an effective interaction for the classical particles, mediated by the electrons, by

$$F_\Lambda(\beta, \mu_e, \mu_c, \omega) = -\mu_c N_c(\omega) - \frac{1}{\beta} \log Z_\Lambda(\beta, \mu_e, \omega)$$

It depends on the inverse temperature $\beta$. Taking the limit of zero temperature gives the corresponding ground-state energy of the electrons in the classical configuration $\omega$:

$$\begin{aligned} E_\Lambda(\mu_e, \mu_c, \omega) &= \lim_{\beta \to \infty} F_\Lambda(\beta, \mu_e, \mu_c, \omega) \\ &= -\mu_c N_c(\omega) + \sum_{j: \lambda_j(\omega) < \mu_e} (\lambda_j(\omega) - \mu_e) \end{aligned}$$

Notice that $F_\Lambda$ and $E_\Lambda$ are strictly decreasing and concave in $\mu_e, \mu_c$ ($E_\Lambda$ is actually linear in $\mu_c$). We also define the energy density in the infinite volume limit by considering a sequence of increasing cubes. For $\omega \in \Omega^{per}$,

$$e(\mu_e, \mu_c, \omega) = \lim_{\Lambda \nearrow \mathbb{Z}^d} \frac{1}{|\Lambda|} E_\Lambda(\mu_e, \mu_c, \omega)$$

The corresponding electronic density is

$$\begin{aligned} \rho_e(\mu_e, \omega) &= \lim_{\Lambda \nearrow \mathbb{Z}^d} \frac{1}{|\Lambda|} \#\{j : \lambda_j(\omega) < \mu_e\} \\ &= -\frac{\partial}{\partial \mu_e} e(\mu_e, \mu_c, \omega) \end{aligned} \qquad [4]$$

and the density of classical particles is $\rho_c(\omega) = \lim_\Lambda N_c(\omega)/|\Lambda|$. One can check that canonical and grand-canonical energies are related by

$$e(\mu_e, \mu_c, \omega) = e(\rho_e(\mu_e, \omega), \omega) - \mu_e\rho_e(\mu_e, \omega) - \mu_c\rho_c(\omega) \qquad [5]$$

Given $(\mu_e, \mu_c)$, the ground-state energy density $e(\mu_e, \mu_c)$ is defined by

$$e(\mu_e, \mu_c) = \inf_{\omega\in\Omega^{\text{per}}} e(\mu_e, \mu_c, \omega)$$

The set of periodic ground-state configurations for given chemical potentials $\mu_e, \mu_c$ is the *grand-canonical ground-state phase diagram*:

$$G_{gc}(\mu_e, \mu_c) = \left\{\omega \in \Omega^{\text{per}} : e(\mu_e, \mu_c, \omega) = e(\mu_e, \mu_c)\right\}$$

It may happen that no periodic configuration minimizes $e(\mu_e, \mu_c, \omega)$ and that $G_{gc}(\mu_e, \mu_c) = \emptyset$. However, results suggest that $G_{gc}(\mu_e, \mu_c)$ is nonempty for almost all $\mu_e, \mu_c$.

The situation simplifies for $U > 4d$ and $\mu_e \in (-U + 2d, -2d)$. Since $\mu_e$ belongs to the gap of $h_\Lambda(\omega)$, we have $\rho_e(\mu_e, \omega) = \rho_c(\omega)$, and

$$e(\mu_e, \mu_c, \omega) = e(\rho_c(\omega), \omega) - (\mu_e + \mu_c)\rho_c(\omega)$$

Thus, $G_{gc}(\mu_e, \mu_c)$ is invariant along the line $\mu_e + \mu_c =$ const. (for $\mu_e$ in the gap).

### Symmetries of the Model

The Hamiltonian $H_\Lambda$ clearly has the symmetries of the lattice (for a box with periodic boundary conditions, there is invariance under translations, rotations by 90°, and reflections through an axis). More important, it also possesses particle–hole symmetries and these are useful since they allow us to restrict investigations to positive $U$ and to certain domains of densities or chemical potentials (see below).

- The classical particle–hole transformation $\omega_x \mapsto \overline{\omega}_x = 1 - \omega_x$ results in

$$H_\Lambda^U(\overline{\omega}) = H_\Lambda^{-U}(\omega) - UN_\Lambda$$

  and $N_c(\overline{\omega}) = |\Lambda| - N_c(\omega)$. It follows that $E_\Lambda^U(N_e, \overline{\omega}) = E_\Lambda^{-U}(N_e, \omega) - UN_e$, and

$$G_{\text{can}}^{-U}(\rho_e, \rho_c) = \left\{\overline{\omega} : \omega \in G_{\text{can}}^U(\rho_e, 1 - \rho_c)\right\}$$

$$G_{gc}^{-U}(\mu_e, \mu_c) = \left\{\overline{\omega} : \omega \in G_{gc}^U(\mu_e - U, -\mu_c)\right\}$$

- An electron–hole transformation can be defined via the unitary transformation $a_x \mapsto \varepsilon_x a_x^\dagger$ and $a_x^\dagger \mapsto \varepsilon_x a_x$, where $\varepsilon_x = 1$ on a sublattice, and $= -1$ on the other sublattice. Then,

$$H_\Lambda^U(\omega) \mapsto H_\Lambda^{-U}(\omega) - UN_c(\omega)$$

  and $N_\Lambda \mapsto |\Lambda| - N_\Lambda$. It follows that $E_\Lambda^U(|\Lambda| - N_e, \omega) = E_\Lambda^{-U}(N_e, \omega) - UN_c(\omega)$, and

$$G_{\text{can}}^{-U}(\rho_e, \rho_c) = G_{\text{can}}^U(1 - \rho_e, \rho_c)$$

$$G_{gc}^{-U}(\mu_e, \mu_c) = G_{gc}^U(-\mu_e, \mu_c - U)$$

- Finally, the particle–hole transformation for both the classical particles and the electrons gives

$$H_\Lambda^U(\overline{\omega}) \mapsto H_\Lambda^U(\omega) + UN_\Lambda + UN_c(\omega) - U|\Lambda|$$

  It follows that

$$E_\Lambda^U(|\Lambda| - N_e, \overline{\omega}) = E_\Lambda^U(N_e, \omega) + U(N_e + N_c(\omega) - |\Lambda|)$$

  and

$$G_{\text{can}}^U(\rho_e, \rho_c) = \left\{\overline{\omega} : \omega \in G_{\text{can}}^U(1 - \rho_e, 1 - \rho_c)\right\}$$

$$G_{gc}^U(\mu_e, \mu_c) = \left\{\overline{\omega} : \omega \in G_{gc}^U(-\mu_e - U, -\mu_c - U)\right\}$$

Any of the first two symmetries allow us to choose the sign of $U$. We assume from now on that $U \geq 0$. The third symmetry indicates that the phase diagrams have a point of central symmetry, given by $\rho_e = \rho_c = 1/2$ in the canonical ensemble and $\mu_e = \mu_c = -U/2$ in the grand-canonical ensemble. Consequently, it is enough to study densities satisfying $\rho_e \leq 1/2$ and chemical potentials satisfying $\mu_e \leq -U/2$.

These symmetries also have useful consequences at positive temperatures. In particular, both species of particles have average density 1/2 at $\mu_e = \mu_c = -U/2$, for all $\beta$.

## The Ground State – Arbitrary Dimensions

### The Segregated State

What follows is best understood in the limit $U \to \infty$ and when $\rho_e < \rho_c$. In this case, the electrons become localized in the domain $\mathcal{D}_\Lambda(\omega) = \{x \in \Lambda : w_x = 1\}$ and their energy per site is that of the full configuration, $e(\rho, \omega \equiv 1)$ (see the section "Canonical ensemble"), where $\rho = \rho_e/\rho_c$ is the effective electronic density. The presence of a boundary for $\mathcal{D}_\Lambda(\omega)$ raises the energy and the correction is roughly proportional to

$$B_\Lambda(\omega) = \#\left\{(x, y) : x \in \mathcal{D}_\Lambda(\omega) \text{ and } y \in \mathbb{Z}^d \setminus \mathcal{D}_\Lambda(\omega)\right\}$$

The following theorem was proposed by Freericks *et al.* (2002).

**Theorem 1**

(i) *Let* $\Lambda \subset \mathbb{Z}^d$ *be a finite box, and* $U > 4d$. *Then for all* $\omega \in \Omega_\Lambda$, *and all* $N_e \le N_c(\omega) = N_c$, *we have the following upper and lower bounds*:

$$\frac{1}{2d}\left| e\left(\frac{N_e}{N_c}, \omega \equiv 0\right)\right| B_\Lambda(\omega) \ge E_\Lambda(N_e, \omega)$$

$$- N_c e\left(\frac{N_e}{N_c}, \omega \equiv 1\right) \ge \left[ a\left(\frac{N_e}{N_c}\right) - \gamma(U)\right] B_\Lambda(\omega)$$

*Here,* $a(\rho) = a(1-\rho)$ *is strictly positive for* $0 < \rho < 1$. $\gamma(U)$ *behaves as* $8d^2/U$ *for large* $U$, *in the sense that* $U\gamma(U) \to 8d^2$ *as* $U \to \infty$.

(ii) *For any* $\rho_e \ne \rho_c$ *that differ from zero, the segregated state is the unique ground state if* $a(\rho_e/\rho_c) > \gamma(U)$, *that is, if* $U$ *is large enough.*

The proof of (i) is rather lengthy and we only show here that it implies (ii). Let $b(\omega) = \lim_\Lambda (B_\Lambda(\omega)/|\Lambda|)$, and notice that $b(\omega) = 0$ for the empty, the full, and the segregated configurations; $0 < b(\omega) < d$ for all other periodic configurations or mixtures. Recall that $\rho_c e(\rho_e/\rho_c, \omega \equiv 1)$ is the energy density of the segregated state. For all densities such that $a(\rho_e/\rho_c) > \gamma(U)$, and all configurations such that $\rho_c(\omega) = \rho_c$, we have

$$e(\rho_e, \omega) \ge \rho_c e\left(\frac{\rho_e}{\rho_c}, \omega \equiv 1\right)$$

and the inequality is strict for any periodic configuration. This shows that the segregated configuration is the unique ground state.

### General Properties of the Grand-Canonical Phase Diagram

We have already seen that the grand-canonical phase diagram is symmetric with respect to $(-U/2, -U/2)$. Other properties follow from concavity of $e(\mu_e, \mu_c)$.

Let $\omega \in G_{gc}(\mu_e, \mu_c) \backslash G_{gc}(\mu'_e, \mu'_c)$ and $\omega' \in G_{gc}(\mu'_e, \mu'_c) \backslash G_{gc}(\mu_e, \mu_c)$. Then,

(a) $\mu_e = \mu'_e$ and $\mu'_c > \mu_c$ imply $\rho_c(\omega') > \rho_c(\omega)$;
(b) $\mu_c = \mu'_c$ and $\mu'_e > \mu_e$ imply $\rho_e(\mu'_e, \omega') > \rho_e(\mu_e, \omega)$, and $\omega$ cannot be obtained by adding some classical particles to the configuration $\omega'$.

It follows from (b) that if $\omega \equiv 1 \in G_{gc}(\mu_e, \mu_c)$, then $\omega \equiv 1 \in G_{gc}(\mu'_e, \mu'_c)$ for all $\mu_e \ge \mu'_e$, $\mu_c \ge \mu'_c$. A similar property holds for the empty configuration. To establish these properties, we can start from

$$e(\mu_e, \mu_c, \omega') - e(\mu_e, \mu_c, \omega) > 0 > e(\mu'_e, \mu'_c, \omega') - e(\mu'_e, \mu'_c, \omega) \quad [6]$$

Since $e(\mu_e, \mu_c, \omega)$ is concave with respect to $\mu_e$ and linear with respect to $\mu_c$, we have

$$e(\mu'_e, \mu'_c, \omega) \le e(\mu_e, \mu_c, \omega) + (\mu_e - \mu'_e)\rho_e(\mu_e, \omega) + (\mu_c - \mu'_c)\rho_c(\omega) \quad [7]$$

Using this inequality for both terms on the right-hand side of [6], we obtain the inequality

$$(\mu'_e - \mu_e)[\rho_e(\mu'_e, \omega') - \rho_e(\mu_e, \omega)] + (\mu'_c - \mu_c)[\rho_c(\omega') - \rho_c(\omega)] \ge 0$$

which proves (a) and the first part of (b). The second part of (b) follows from

$$e(\mu_e, \mu_c, \omega) = -\int_{-\infty}^{\mu_e} d\mu\, \rho_e(\mu, \omega) - \mu_c \rho_c(\omega)$$

Indeed, the minimax principle implies that eigenvalues $\lambda_j(\omega)$ are decreasing with respect to $\omega$ (if $U \ge 0$), so that $\rho_e(\mu_e, \omega)$ is increasing (with respect to $\omega$). Then for any $\omega'' > \omega$ and $\mu'_e > \mu_e$,

$$e(\mu'_e, \mu_c, \omega'') - e(\mu'_e, \mu_c, \omega) > e(\mu_e, \mu_c, \omega'') - e(\mu_e, \mu_c, \omega)$$

and $\omega'' \notin G_{gc}(\mu_e, \mu_c)$ implies $\omega'' \notin G_{gc}(\mu'_e, \mu_c)$.

Next, we discuss domains in the plane of chemical potentials where the empty, full, and chessboard configurations have minimum energy (see, e.g., Gruber and Macris (1996), and references therein). One easily sees that $\omega \equiv 1$ is the unique ground-state configuration if $\mu_c > 0$, or if $\mu_e > 2d$ and $\mu_c > -U$. Similarly, $\omega \equiv 0$ is the unique ground state if $\mu_c < -U$, or if $\mu_e < -U - 2d$ and $\mu_c < 0$. For $U > 4d$, it follows from the expansion [1] that the full configuration is also ground state if $-U + 2d < \mu_e < -2d$ and $\mu_e + \mu_c + U > 4d/(U - 4d)$. These domains can be rigorously extended using energy estimates that involve correlation functions of classical particles. The results are illustrated in Figures 2 ($U < 4d$) and 3 ($U > 4d$).

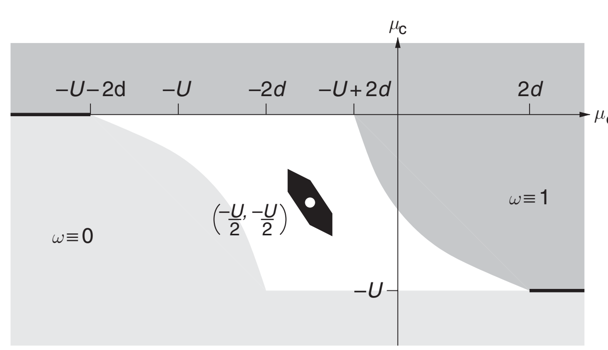



**Figure 2** Grand-canonical ground-state phase diagram for $U < 4d$. Domains for the empty, chessboard, and full configurations, are denoted in light gray, black, and dark gray, respectively.

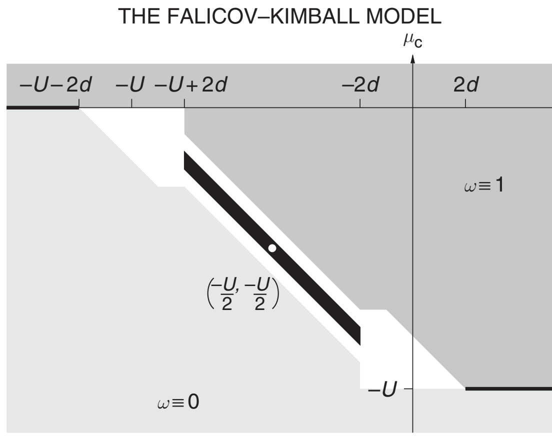



**Figure 3** Grand-canonical ground-state phase diagram for $U > 4d$. Domains for the empty, chessboard, and full configurations are denoted in light gray, black, and dark gray, respectively.

Finally, canonical and grand-canonical phase diagrams are related by the following properties:

(c) If $\omega \in G_{gc}(\mu_e, \mu_c)$, then $\omega \in G_{can}(\rho_e(\mu_e, \omega), \rho_c(\omega))$.
(d) More generally, suppose that $\omega^{(1)}, \ldots, \omega^{(n)} \in G_{gc}(\mu_e, \mu_c)$, and consider a mixture with coefficients $\alpha_1, \ldots, \alpha_n$. The mixture belongs to $G_{can}(\rho_e, \rho_c)$, with $\rho_e = \sum_i \alpha_i \rho_e(\mu_e, \omega^{(i)})$ and $\rho_c = \sum_i \alpha_i \rho_c(\omega^{(i)})$.

To establish (c), observe that any $\omega'$ satisfies $e(\mu_e, \mu_c, \omega') \geq e(\mu_e, \mu_c, \omega)$ if $\omega \in G_{gc}(\mu_e, \mu_c)$. Let $\rho_e = \rho_e(\mu_e, \omega)$ and $\rho_c = \rho_c(\omega)$, and let $\mu'_e$ be such that $\rho_e(\mu'_e, \omega') = \rho_e$. By eqns [5] and [7],

$$e(\rho_e(\mu'_e, \omega'), \omega') - \mu_e \rho_e(\mu'_e, \omega') - \mu_c \rho_c(\omega') \geq e(\rho_e(\mu_e, \omega), \omega) - \mu_e \rho_e(\mu_e, \omega) - \mu_c \rho_c(\omega)$$

Then, $e(\rho_e, \omega') \geq e(\rho_e, \omega)$ for any configuration $\omega'$ such that $\rho_c(\omega') = \rho_c$. Property (d) follows from (c) by a limiting argument, because a mixture can be approximated by a sequence of periodic configurations.

Next we describe further properties of the phase diagrams that are specific to dimensions 1 and 2.

## Ground-State Configurations – Dimension 1

A large number of investigations, either analytical or numerical, have been devoted to the study of the ground-state configurations in one dimension. One-dimensional results also serve as guide to higher dimensions. Recall that symmetries allow us to restrict to $U \geq 0$ and $\rho_e \leq 1/2$.

Most ground-state configurations that appear in the canonical phase diagram seem to be given by an intriguing formula, which we now describe. Let $\rho_e = p/q$ with $p$ relatively prime to $q$. Then corresponding periodic ground-state configurations have period $q$ and density $\rho_c = r/q$ ($r$ is an integer). The occupied sites in the cell $\{0, 1, \ldots, q-1\}$ are given by the solutions $k_0, \ldots, k_{r-1}$ of

$$(pk_j) = j \bmod q, \quad 0 \leq j \leq r-1 \qquad [8]$$

Note that the first classical particle is located at $k_0 = 0$, and $k_0, \ldots, k_{p-1}$ are *not* in increasing order. In order to discuss the solutions of [8], we introduce $\ell = \lfloor q/p \rfloor$ (the integer part of $q/p$), and we write

$$q = (\ell + 1)p - s \qquad [9]$$

where $1 \leq s \leq p-1$, and $s$ is relatively prime to $p$. Next, let $L(x)$ denote the distance between the particle at $x$ and the one immediately preceding it (to the left).

Let us observe that if $\rho_c = \rho_e$, that is, if $r = p$, then

(a) $L(k_j) = \ell$ for $0 \leq j \leq s-1$ and $k_j - \ell = k_{j+p-s}$.
(b) $L(k_j) = \ell + 1$ for $s \leq j \leq p-1$ and $k_j - (\ell+1) = k_{j-s}$.

Indeed, for $pk_j = j + nq$, eqn [9] implies

$$p(k_j - \ell) = j + (n-1)q + (p-s) = j + p - s \bmod q$$

and

$$p(k_j - \ell - 1) = j - s \bmod q$$

Therefore, $k_j - \ell$ is a solution of [8] if $j + p - s \leq p - 1$, while $k_j - (\ell + 1)$ is a solution of [8] if $j - s \geq 0$.

These two properties show that the configuration defined by [8] is such that $L(x) \in \{\ell, \ell+1\}$ for all occupied $x$. A periodic configuration such that all distances between consecutive particles are either $\ell$ or $\ell + 1$ is called *homogeneous*. Let $\omega$ be a homogeneous configuration with period $q$ and density $\rho_c = r/q$, and let $x_0 < \cdots < x_{p-1}$ be the occupied sites in $\{0, 1, \ldots, q-1\}$. We introduce the *derivative* $\omega'$ of $\omega$ as the periodic configuration with period $r$ defined by (see Figure 4)

$$\omega'_i = \begin{cases} 1 & \text{if } L(x_i) = \ell \\ 0 & \text{if } L(x_i) = \ell + 1 \end{cases}$$

A configuration is *most homogeneous* if it can be "differentiated" repeatedly until the empty or the full configuration is obtained.

Let $\omega$ be the homogeneous configuration from [8] and $\omega'$ be its derivative. Using the same arguments as for properties (a) and (b) above, and the fact that $s$ is relatively prime to $p$, we obtain

(c) Let $k'_0, \ldots, k'_{p-1}$ be the solutions of

$$(sk'_j) = j \bmod p$$

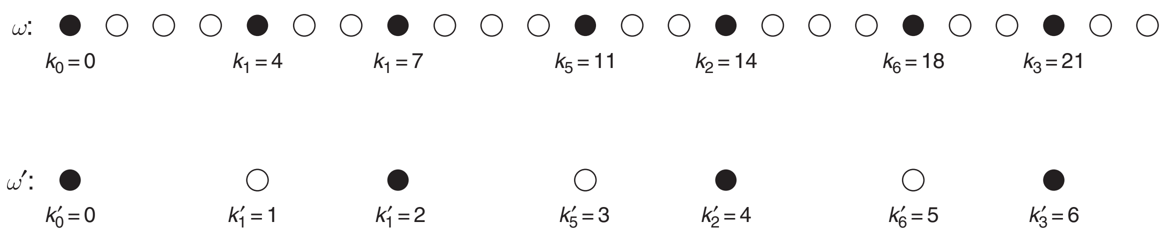



**Figure 4** The configuration $\omega$ given by the formula [8] with $q = 24$ and $p = 7$, and its derivative $\omega'$. Notice that $\ell = 3$ and $s = 4$.

Then $(k'_0, \ldots, k'_{p-1})$ is a permutation of $(0, 1, \ldots, p-1)$. Further, $k'_j - 1 = k'_{j+p-s}$ for $0 \le j \le s-1$, and $k'_j - 1 = k'_{j-s}$ for $s \le j \le p-1$.

Consider the periodic configuration with period $p$ where sites $k'_0, \ldots, k'_{s-1}$ are occupied and sites $k'_s, \ldots, k'_{p-1}$ are empty. Since $k'_0 = 0$, this configuration is precisely the derivative $\omega'$ of $\omega$. Iterating, these properties prove that the solutions of [8] are most homogeneous.

One of the most important results in one dimension is that only most homogeneous configurations are present in the canonical phase diagram, for $U$ large enough and for equal densities $\rho_e = \rho_c$.

**Theorem 2** *Suppose that $\rho_e = \rho_c = p/q$. There exists a constant $c$ such that for $U > c4^q$, the only ground-state configuration is the most homogeneous configuration, given by [8] (together with translations and reflections).*

This theorem was established using the expansion [1] of $E_\Lambda(N_e, \omega)$ in powers of $U^{-1}$. It suggests a devil's staircase structure with infinitely many domains. However, the number of domains for *fixed U* could still be finite. Results from Theorem 2 are illustrated in Figure 5. Notice that $\rho_e = \rho_c$ when $\mu_e$ is in the universal gap. These results have been extended

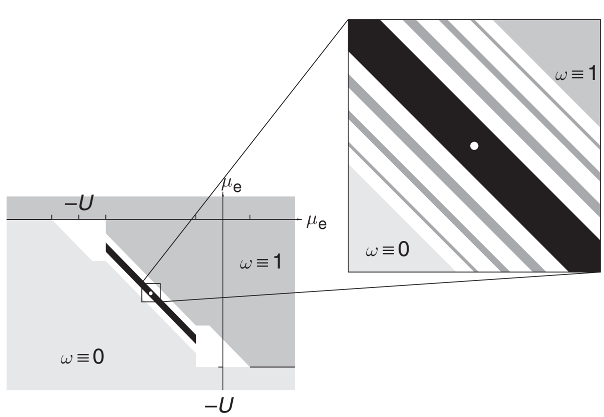



**Figure 5** Grand-canonical ground-state phase diagram in one dimension for $U > 4$ and $\mu_e$ in the universal gap. Chessboard configurations occur in the black domain. Dark gray oblique domains correspond to densities 1/5, 1/4, 1/3, 2/3, 3/4, 4/5. Total width of these domains is of order $U^{-1}$.

to positive temperatures by using "quantum Pirogov–Sinai theory" (Datta *et al.* 1999).

For small $U$, on the other hand, one can use a (nonrigorous) Wigner–Brillouin degenerate perturbation theory (a standard tool in band theory). Let $\rho_e = p/q$ with $p$ relatively prime to $q$, and $\omega$ be a periodic configuration with period $nq, n \in \mathbb{N}$. Then for $U$ small enough ($U \ll 1/q$), we obtain the following expansion for the ground-state energy (Freericks *et al.* 1996):

$$e(\rho_e, \omega) = -\frac{2}{\pi}\sin \pi\rho_e - U\rho_e\rho_c(\omega) - \frac{|\widehat{\omega}(\rho_e)|^2}{4\pi \sin \pi\rho_e} U^2 |\log U| + O(U^2) \qquad [10]$$

where $\widehat{\omega}(\rho_e)$ is the "structure factor" of the periodic configuration $\omega$, namely

$$\widehat{\omega}(\rho_e) = \frac{1}{nq}\sum_{j=0}^{nq-1} e^{-2\pi i \rho_e j} w_j$$

This expansion suggests that the ground-state configuration can be found by maximizing the structure factor. The following theorem holds independently of $U$.

**Theorem 3** *Let $\rho_e = p/q$. There exist $r_1 \ge q/4$ and $r_2 \le 3q/4$ such that the configurations maximizing the structure factor are given as follows:*

(i) *for $\rho_c = r/q$ with $r_1 \le r \le r_2$, use the formula* [8];
(ii) *for $\rho_c \in (r/q, (r+1/q))$ with $r_1 \le r \le r_2 - 1$, the configuration is a mixture of those for $\rho_c = r/q$ and $\rho_c = (r+1)/q$; and*
(iii) *for $\rho_c \in (0, r_1/q)$, the configurations are mixtures of $\omega \equiv 0$ and that for $\rho_c = r_1/q$. For $\rho_c \in (r_2/q, 1)$, the configurations are mixtures of $\omega \equiv 1$ and that for $\rho_c = r_2/q$.*

Some insight for low densities is provided by computing the energy of just one classical particle and one electron on the infinite line, and to compare it with two consecutive classical particles and two electrons. It turns out that the former is more favorable than the latter for $U > 2/\sqrt{3} \approx 1.15$, while "molecules" of two particles are forming

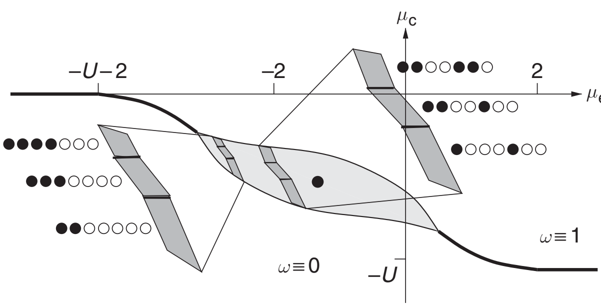



**Figure 6** Grand-canonical ground-state phase diagram for $U \approx 0.4$. Enlarged are domains for $\rho_e = 1/7$ and $2/7$, with the same densities $\rho_c = 2/7, 3/7, 4/7$.

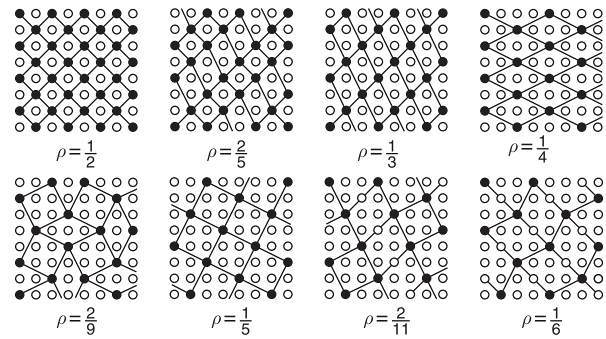



**Figure 7** Ground-state configurations for several densities. Occupied sites are denoted by black circles, empty sites by white circles. Lines are present only to clarify the patterns.

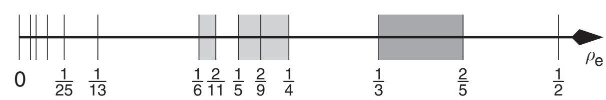



**Figure 8** Canonical ground-state phase diagram in two dimensions for $U > 8$.

when $U < 2/\sqrt{3}$. Smaller $U$ shows even bigger molecules for $\rho_c = n\rho_e$, and $n$-molecules are most homogeneously distributed according to the formula [8]. It should be stressed that the canonical ground state cannot be periodic if $U$ is small and $\rho_c \notin [1/4, 3/4]$, which is different from the case of large $U$.

Only numerical results are available for intermediate $U$. They suggest that configurations occurring in the phase diagram are essentially given by Theorem 3 (together with the segregated configuration). This is sketched in **Figure 6**, where bold coexistence lines for $\mu_e > -U - 2$ and $\mu_e < 2$ represent segregated states.

## Ground-State Configurations – Dimension 2

We discuss the canonical ensemble only, but many results extend to the grand-canonical ensemble. Recall that $G_{\text{can}}(1/2, 1/2)$ consists of the two chessboard configurations for any $U > 0$, and that segregation takes place when $\rho_e \neq \rho_c$, providing $U$ is large enough (Theorem 1). Other results deal with the case of equal densities, and for $U$ large enough (see Haller and Kennedy (2001), and references therein).

**Theorem 4** *Let* $\rho_e = \rho_c \equiv \rho \leq 1/2$.

(i) *If*

$$\rho \in \left\{ \frac{1}{2}, \frac{2}{5}, \frac{1}{3}, \frac{1}{4}, \frac{2}{9}, \frac{1}{5}, \frac{2}{11}, \frac{1}{6} \right\}$$

*then for $U$ large enough, the ground-state configurations are those displayed in* ***Figure*** *7. If $\rho = 1/(n^2 + (n+1)^2)$ with integer $n$, then for $U$ large enough (depending on $\rho$), the ground-state configurations are periodic.*

(ii) *If $\rho$ is a rational number between* 1/3 *and* 2/5, *then for $U$ large enough (depending on the denominator of $\rho$), the ground-state configurations are periodic. Further, the restriction to any horizontal line is a one-dimensional periodic configuration given by* [8], *and the configuration is constant in either the direction $\binom{1}{1}$ or $\binom{1}{-1}$.*

(iii) *Suppose that $U$ is large enough. If $\rho \in (1/6, 2/11)$, the ground-state configurations are mixtures of the configurations $\rho = 1/6$ and $\rho = 2/11$ of* ***Figure*** *7. If $\rho \in (1/5, 2/9)$, the ground-state configurations are mixtures of the configurations $\rho = 1/5$ and $\rho = 2/9$. If $\rho \in (2/9, 1/4)$, the ground-state configurations are mixtures of the configurations $\rho = 2/9$ and $\rho = 1/4$.*

The canonical phase diagram for $\rho_e = \rho_c$ is presented in **Figure 8**.

The situation for densities $\rho \leq 1/2$ that are not mentioned in Theorem 4 is unknown. All these periodic configurations are present in the grand-canonical phase diagram as well. Theorem 4(ii) suggests that the two-dimensional situation is similar to the one-dimensional one where a devil's staircase structure may occur. Let us stress that no periodic configurations occur for large $U$ and densities $\rho_e = \rho_c$ in the intervals (1/6, 2/11), (1/5, 2/9), and (2/9, 1/4). This resembles the one-dimensional situation, but for small $U$.

*See also:* Quantum Spin Systems; Quantum Statistical Mechanics: Overview; Fermionic Systems; Hubbard Model; Pirogov–Sinai Theory.

## Further Reading

Brandt U and Schmidt T (1986) Ground state properties of a spinless Falicov–Kimball model – additional features. *Zeitschrift für Physik B* 67: 43–51.
Datta N, Fernández R, and Fröhlich J (1999) Effective Hamiltonians and phase diagrams for tight-binding models. *Journal of Statistical Physics* 96: 545–611.
Falicov LM and Kimball JC (1969) Simple model for semiconductor–metal transitions: $SmB_6$ and transition-metal oxides. *Physical Review Letters* 22: 997–999.
Freericks JK, Gruber Ch, and Macris N (1996) Phase separation in the binary-alloy problem: the one-dimensional spinless Falicov–Kimball. *Physical Review B* 53: 16189–16196.
Freericks JK, Lieb EH, and Ueltschi D (2002) Segregation in the Falicov–Kimball model. *Communications in Mathematical Physics* 227: 243–279.
Freericks JK and Zlatić V (2003) Exact dynamical mean-field theory of the Falicov–Kimball model. *Reviews of Modern Physics* 75: 1333–1382.
Gruber Ch and Macris N (1996) The Falicov–Kimball model: a review of exact results and extensions. *Helvetica Physica Acta* 69: 850–907.
Haller K and Kennedy T (2001) Periodic ground states in the neutral Falicov–Kimball model in two dimensions. *Journal of Statistical Physics* 102: 15–34.
Jędrzejewski J and Lemański R (2001) Falicov–Kimball models of collective phenomena in solids (a concise guide). *Acta Physica Polonica* 32: 3243–3251.
Kennedy T and Lieb EH (1986) An itinerant electron model with crystalline or magnetic long range order. *Physica A* 138: 320–358.

# Fedosov Quantization

**N Neumaier**, Albert-Ludwigs-University in Freiburg, Freiburg, Germany



## Introduction

On the one hand, quantum mechanics and classical mechanics appear to be formulated within quite different mathematical frameworks, that is, in terms of Hilbert spaces and operators on Hilbert spaces on the quantum side and in terms of phase spaces, that is, symplectic, or more generally, Poisson manifolds, and functions on these phase spaces on the classical side. On the other hand, there is a strong structural similarity between the algebras of observable quantities in both theories which are associative $^*$-algebras over $\mathbb{C}$. In the classical case, the algebra is commutative the product being the pointwise product of functions on the phase space and is endowed with the additional structure of a Poisson bracket by means of which the dynamics of the system can be formulated. In the quantum case, the algebra is the noncommutative composition of operators on a Hilbert space and the dynamics is determined by the corresponding commutator. The difference between functions on a phase space and the operators on a Hilbert space constitutes the main difficulty for the passage from a classical theory to the corresponding quantum theory which would be desirable, since a formulation of the more fundamental but much less intuitive quantum theory is often impossible. Even the consideration of the classical limit leads to the same problem of comparing quite different mathematical objects. One possibility, which is the basic idea of deformation quantization, to avoid these problems is to pass from classical observables to quantum observables not by changing the underlying vector space, but only by deforming the algebraic structures namely the associative product and possibly the $^*$-involution.

This idea motivates the following definition of a star product by Bayen *et al.* (1978), which reassembles the minimal demands made on a suitable quantization:

**Definition 1** A star product on a Poisson manifold $(M,\Pi)$ is an associative $\mathbb{C}[[\nu]]$-bilinear product $\star$ on $\mathcal{C}^\infty(M)[[\nu]]$ such that – writing $f\star g=\sum_{r=0}^{\infty}\nu^r C_r(f,g)$ for $f,g\in\mathcal{C}^\infty(M)$ with $\mathbb{C}$-bilinear maps $C_r$ with values in $\mathcal{C}^\infty(M)$ – the following properties hold:

(i) $C_0(f,g)=fg$,
(ii) $C_1(f,g)-C_1(g,f)=\{f,g\}$, and
(iii) $1\star f=f=f\star 1$.

In case the $\mathbb{C}$-bilinear maps $C_r$ are differential operators, the star product is called differential. If $\overline{f\star g}=\overline{g}\star\overline{f}$, then $\star$ is called Hermitian.

The conditions (i) and (ii) express the correspondence principle in deformation quantization and in case the star product converges the formal parameter is to be identified with $i\hbar$, whence we set $\overline{\nu}=-\nu$ considering the formal parameter as purely imaginary. Since the Fedosov star products we are going to study in the sequel are differential, we shall drop stressing this property explicitly and refer to differential star products as star products, merely.

One main advantage of deformation quantization is that one has the following very general existence result:

**Theorem 1** *On every Poisson manifold* $(M,\Pi)$ *there exist (even differential) star products.*

This theorem was first shown by DeWilde and Lecomte (1983) for the symplectic case and

independently by Fedosov (1985) who gave a beautiful explicit construction using geometrical structures on $(M, \omega)$ to build a star product recursively. Omori *et al.* (1991) gave yet another existence proof of star products on a symplectic manifold $(M, \omega)$ that appears to combine the methods of DeWilde and Lecomte (1983) and Fedosov (1985). The general proof of existence on general Poisson manifolds is due to Kontsevich (2003) and is a consequence of Kontsevich's formality theorem.

If $S = \mathrm{id} + \sum_{r=1}^{\infty} \nu^r S_r$ is a formal series of differential operators on $\mathcal{C}^\infty(M)$ with $S_r 1 = 0$ for $r \geq 1$, then

$$f \star' g := S^{-1}((Sf) \star (Sg)) \qquad [1]$$

again defines a star product. Clearly, $\star'$ is Hermitian in case $\star$ is Hermitian and $\overline{Sf} = S\overline{f}$ for all $f \in \mathcal{C}^\infty(M)[[\nu]]$. The above observation of the shape of certain isomorphisms between star product algebras gives rise to the notion of equivalence of star products:

**Definition 2** Two star products $\star$ and $\star'$ on $(M, \Pi)$ are called equivalent in case there is a formal series $S = \mathrm{id} + \sum_{r=1}^{\infty} \nu^r S_r$ of differential operators on $\mathcal{C}^\infty(M)$ with $S_r 1 = 0$ for $r \geq 1$ such that eqn [1] is satisfied for all $f, g \in \mathcal{C}^\infty(M)[[\nu]]$.

The full classification of star products up to equivalence was first obtained in the symplectic case by Nest and Tsygan (1995) and independently by Deligne (1995) and Bertelson *et al.* (1997). The general Poisson case again follows from Kontsevich's formality theorem. In particular, in the symplectic case, star products are classified in a functorial way by the characteristic class

$$c : \star \mapsto c(\star) \in \frac{[\omega]}{\nu} + H^2_{\mathrm{deRham}}(M)[[\nu]] \qquad [2]$$

defined by Deligne that induces a bijection between the equivalence classes of star products and $[\omega]/\nu + H^2_{\mathrm{deRham}}(M)[[\nu]]$. Moreover, it has been shown (Bertelson *et al.* 1997, Deligne 1995, Nest and Tsygan 1995) that every star product on a symplectic manifold is equivalent to a Fedosov star product. This fact can also be seen as a direct consequence of the explicit computation of the characteristic class of a Fedosov star product (cf. Neumaier (2002)). The importance of Fedosov's construction for the general theory of deformation quantization in the symplectic case is also shown by the fact that in many proofs Fedosov's star products were used to have reference star products to compare with a given star product. Moreover, there is a great variety of modifications and generalizations of Fedosov's method and there are many examples where additional structures on the symplectic manifold suggest to look for star products adapted to them, where modified Fedosov constructions can be applied successfully.

## Fedosov Star Products on $(M, \omega)$

The attempt to construct a star product step by step in fact leads to a cohomological problem, where *a priori* an obstruction in the third Hochschild cohomology of $\mathcal{C}^\infty(M)$ occurs. This problem results from the demand for associativity which is the really most restricting condition on a star product. Therefore, additional arguments are necessary to show that these obstructions can be circumvented, since the concerning cohomology is isomorphic to $\Gamma^\infty(\bigwedge^3 TM)$ and hence, for $\dim(M) \geq 3$, is not trivial at all.

The basic strategy of Fedosov's construction to build in associativity of the resulting product is to begin with a "very large" associative algebra $(\mathcal{W} \otimes \Lambda, \circ)$, where $\circ$ mimicks the well-known Weyl–Moyal star product on a vector space with a constant symplectic Poisson tensor, and to specify a suitable subalgebra which is in bijection to $\mathcal{C}^\infty(M)[[\nu]]$. Pulling back the product to the subalgebra then clearly results in an associative product on $\mathcal{C}^\infty(M)[[\nu]]$, but as we shall see later on, one has to care for the bijection to be sufficiently nontrivial in order to obtain in fact a nontrivial deformation of the usual pointwise product on $\mathcal{C}^\infty(M)[[\nu]]$.

Defining

$$\mathcal{W} \otimes \Lambda := \left( \bigtimes_{s=0}^{\infty} \Gamma^\infty\left( \bigvee^s T^*M \otimes \bigwedge T^*M \right) \right)[[\nu]] \qquad [3]$$

$\mathcal{W} \otimes \Lambda$ becomes in a natural way an associative, supercommutative algebra using the symmetric $\vee$-product in the first factor and the antisymmetric $\wedge$-product in the second factor. This product is denoted by $\mu(a \otimes b) = ab$ for $a, b \in \mathcal{W} \otimes \Lambda$. By $\mathcal{W} \otimes \Lambda^k$ we denote the elements of antisymmetric degree $k$ and set $\mathcal{W} := \mathcal{W} \otimes \Lambda^0$. Besides this pointwise product, the Poisson tensor $\Pi$ corresponding to $\omega$ gives rise to another associative product $\circ$ on $\mathcal{W} \otimes \Lambda$ by

$$a \circ b = \mu\left( \exp\left( \frac{\nu}{2} \Pi^{ij} i_s(\partial_i) \otimes i_s(\partial_j) \right)(a \otimes b) \right) \qquad [4]$$

which is a deformation of $\mu$. Here $i_s(Y)$ denotes the symmetric insertion of a vector field $Y \in \Gamma^\infty(TM)$ and similarly $i_a(Y)$ shall be used to denote the antisymmetric insertion of a vector field. We set $\mathrm{ad}(a)b := [a, b]$, where the latter denotes the

$\deg_a$-graded supercommutator with respect to $\circ$. Denoting the obvious degree maps by $\deg_s, \deg_a$, and $\deg_\nu = \nu\partial_\nu$, one observes that they all are derivations with respect to $\mu$ but $\deg_s$ and $\deg_\nu$ fail to be derivations with respect to $\circ$. Instead, $\mathrm{Deg} := \deg_s + 2\deg_\nu$ is a derivation of $\circ$ and hence $(\mathcal{W} \otimes \Lambda, \circ)$ is formally Deg-graded and the corresponding degree is referred to as the total degree. Sometimes we write $\mathcal{W}_k \otimes \Lambda$ to denote the elements of total degree $\geq k$. The total degree can be used to define an ultrametric $d$ on $\mathcal{W} \otimes \Lambda$ and it is known that $(\mathcal{W} \otimes \Lambda, d)$ is complete, which implies that Banach's fixed-point theorem can be applied in this setting. This observation is important since all the proofs of existence and uniqueness of certain elements in $\mathcal{W} \otimes \Lambda$ we shall construct in the sequel can be reduced to the application of this theorem.

In local coordinates, we define the differential

$$\delta := (1 \otimes \mathrm{d}x^i) i_s(\partial_i) \qquad [5]$$

which satisfies $\delta^2 = 0$ and is a superderivation of $\circ$. Evaluated at a point $m \in M$, the product $a(m)b(m)$ of two elements $a, b \in \mathcal{W} \otimes \Lambda$ can be considered as the $\wedge$-product of two differential forms with polynomial coefficient functions on the vector space $T_m M$. Interpreted this way, the restriction of $\delta$ to the fiber at $m$ is nothing but the exterior derivative of differential forms with polynomial coefficients. Hence, it is clear that there is a homotopy operator $\delta^{-1}$ satisfying

$$\delta\delta^{-1} + \delta^{-1}\delta + \sigma = \mathrm{id} \qquad [6]$$

where $\sigma : \mathcal{W} \otimes \Lambda \to \mathcal{C}^\infty(M)[[\nu]]$ denotes the projection onto the part of symmetric and antisymmetric degree 0. With the above view of $\delta$, this is just the Poincaré Lemma for differential forms with polynomial coefficients, which says that all the cohomology spaces vanish except for the one of degree 0 and the cohomology in degree 0 is just given by the constant functions on the vector space $T_m M$. This means that the $\delta$-cohomology on $\mathcal{W} \otimes \Lambda$ is trivial except for the space of degree 0, which is given by the formal functions on $M$. For computational purposes, it is useful to have a concrete formula of the homotopy operator $\delta^{-1}$ which is given by

$$\delta^{-1} a := \begin{cases} \dfrac{1}{k+l}(\mathrm{d}x^i \otimes 1) i_a(\partial_i) a & \text{for } \deg_s a = ka, \\ \quad \deg_a a = la \text{ with } k + l \neq 0 \\ \quad 0 \quad \text{else} \end{cases} \qquad [7]$$

Now $\ker(\delta) \cap \mathcal{W} = \mathcal{C}^\infty(M)[[\nu]]$ and one might wonder whether this subalgebra of $(\mathcal{W} \otimes \Lambda, \circ)$ is already suitable to induce a deformed product on $\mathcal{C}^\infty(M)[[\nu]]$ by pulling back the product $\circ$ from $\mathcal{W} \otimes \Lambda$. Evidently, the answer to the question is negative since the resulting product just gives back the undeformed pointwise product of formal functions on $M$. Hence one has to find a less trivial superderivation of the product $\circ$ the kernel of which is still in bijection to $\mathcal{C}^\infty(M)[[\nu]]$. The essential new component of Fedosov's construction is a superderivation of $(\mathcal{W} \otimes \Lambda, \circ)$ that is not $\mathcal{C}^\infty(M)[[\nu]]$-linear and hence in a certain sense generates derivatives along the base manifold $M$. Using a torsion-free symplectic connection $\nabla$ on $M$ we define an endomorphism also denoted by $\nabla$ of $\mathcal{W} \otimes \Lambda$ by

$$\nabla := (1 \otimes \mathrm{d}x^i) \nabla_{\partial_i} \qquad [8]$$

which turns out to be a superderivation of $\circ$ due to the fact that $\nabla\omega = \nabla\Pi = 0$. The map $\nabla$ satisfies the identities

$$[\delta, \nabla] = 0, \quad \text{since the connection is torsion-free} \quad [9]$$

$$\nabla^2 = -\frac{1}{\nu}\mathrm{ad}(R),$$
$$\text{where } R := \tfrac{1}{4}\omega_{it} R^t_{jkl}\, \mathrm{d}x^i \vee \mathrm{d}x^j \otimes \mathrm{d}x^k \wedge \mathrm{d}x^l \in \mathcal{W} \otimes \Lambda^2 \quad [10]$$

involves the curvature of the connection. Moreover, we have

$$\delta R = 0 = \nabla R \qquad [11]$$

by the Bianchi identities.

Now one could consider the superderivation $-\delta + \nabla$ of $(\mathcal{W} \otimes \Lambda, \circ)$ and try to define a mapping $\tau$ from $\mathcal{C}^\infty(M)[[\nu]]$ to $\ker(-\delta + \nabla) \cap \mathcal{W}$ such that $\sigma(\tau(f)) = f$ for all $f \in \mathcal{C}^\infty(M)[[\nu]]$. But in case the curvature of the connection does not vanish, the necessary condition for the solvability of the equation $(-\delta + \nabla)\tau(f) = 0$ subject to the additional condition $\sigma(\tau(f)) = f$ is not satisfied. Only in case there is a torsion-free symplectic connection on $M$ with vanishing curvature, this procedure can be carried through and yields again the Weyl–Moyal star product since the fact that $\nabla$ is symplectic in this case implies that the components of the Poisson tensor are constant. However, in general, the kernel of $-\delta + \nabla$ does not have the desired properties to specify a suitable subalgebra of $(\mathcal{W} \otimes \Lambda, \circ)$ and one makes the ansatz

$$\mathfrak{D} = -\delta + \nabla - \frac{1}{\nu}\mathrm{ad}(r) \qquad [12]$$

with an element $r \in \mathcal{W}_3 \otimes \Lambda^1$ for a suitable superderivation. Now a direct computation yields that

$$\mathfrak{D}^2 = \frac{1}{\nu}\mathrm{ad}\left(\delta r - \nabla r + \frac{1}{\nu} r \circ r - R\right) \qquad [13]$$

which vanishes iff $\delta r - \nabla r + (1/\nu) r \circ r - R$ is a central element in $\mathcal{W}_2 \otimes \Lambda^2$. This is the case iff there is a formal series of 2-forms $\Omega \in \nu\Gamma^\infty(\bigwedge^2 T^*M)[[\nu]]$ with

$$\delta r - \nabla r + \frac{1}{\nu} r \circ r - R = 1 \otimes \Omega \qquad [14]$$

After these preparations, one is in the position to prove the following theorem:

**Theorem 2** (Fedosov 1994, theorem 3.2; Fedosov 1996, theorem 5.2.2). *For every formal series $\Omega \in \nu\Gamma^\infty(\bigwedge^2 T^*M)[[\nu]]$ of closed 2-forms there exists a unique element $r \in \mathcal{W}_3 \otimes \Lambda^1$ such that*

$$\delta r = \nabla r - \frac{1}{\nu} r \circ r + R + 1 \otimes \Omega \quad \text{and} \quad \delta^{-1} r = 0 \qquad [15]$$

*Moreover, $r$ satisfies*

$$r = \delta^{-1}\left(R + 1 \otimes \Omega + \nabla r - \frac{1}{\nu} r \circ r\right) \qquad [16]$$

*from which $r$ can be determined recursively. In this case the Fedosov derivation*

$$\mathfrak{D} := -\delta + \nabla - \frac{1}{\nu}\mathrm{ad}(r) \qquad [17]$$

*is a superderivation of antisymmetric degree 1 and has square zero*: $\mathfrak{D}^2 = 0$.

For obvious reasons Fedosov calls $\mathfrak{D}$ a flat or abelian connection for the bundle $\mathcal{W} \otimes \Lambda$ and $\omega + \Omega$ is referred to as the central Weyl curvature of the connection $\mathfrak{D}$. In some sense, the flatness property $\mathfrak{D}^2 = 0$ guarantees that there are sufficiently many flat sections. Before investigating the structure of $\ker(\mathfrak{D}) \cap \mathcal{W}$ we note that the $\mathfrak{D}$-cohomology is trivial on elements $a$ with positive antisymmetric degree since one has the following homotopy formula:

$$\mathfrak{D}\mathfrak{D}^{-1} a + \mathfrak{D}^{-1}\mathfrak{D} a = a$$

where

$$\mathfrak{D}^{-1} a := -\delta^{-1}\left(\frac{1}{\mathrm{id} - [\delta^{-1}, \nabla - (1/\nu)\mathrm{ad}(r)]} a\right) \qquad [18]$$

(cf. Fedosov (1996, theorem 5.2.5)). The reason for this fact, which is also the crucial point for the proof of Theorem 1, is the property of the $\delta$-cohomology to vanish except for the cohomology space of degree 0.

The next step in Fedosov's construction now consists in establishing a bijection between the flat sections $a \in \mathcal{W}$, that is, those elements of $\mathcal{W}$ with $\mathfrak{D} a = 0$, and $\mathcal{C}^\infty(M)[[\nu]]$.

**Theorem 3** (Fedosov 1994, theorem 3.3, Fedosov 1996, theorem 5.2.4). *Let $\mathfrak{D} = -\delta + \nabla - (1/\nu)\mathrm{ad}(r): \mathcal{W} \otimes \Lambda \to \mathcal{W} \otimes \Lambda$ be given as in [17] with $r$ as in [15].*

(i) *Then for any $f \in \mathcal{C}^\infty(M)[[\nu]]$ there exists a unique element $\tau(f) \in \ker(\mathfrak{D}) \cap \mathcal{W}$ such that*

$$\sigma(\tau(f)) = f \qquad [19]$$

*and $\tau : \mathcal{C}^\infty(M)[[\nu]] \to \ker(\mathfrak{D}) \cap \mathcal{W}$ is $\mathbb{C}[[\nu]]$-linear and referred to as the Fedosov–Taylor series corresponding to $\mathfrak{D}$.*

(ii) *In addition, $\tau(f)$ can be obtained recursively for $f \in \mathcal{C}^\infty(M)$ from*

$$\tau(f) = f + \delta^{-1}\left(\nabla \tau(f) - \frac{1}{\nu}\mathrm{ad}(r)\tau(f)\right) \qquad [20]$$

*Using $\mathfrak{D}^{-1}$ according to [18] one can also write*

$$\tau(f) = f - \mathfrak{D}^{-1}(1 \otimes \mathrm{d}f) \quad \text{for all } f \in \mathcal{C}^\infty(M)[[\nu]] \qquad [21]$$

(iii) *Since $\mathfrak{D}$ as constructed above is a $\circ$-superderivation, $\ker(\mathfrak{D}) \cap \mathcal{W}$ is a $\circ$-subalgebra and a new associative product $*$ for $\mathcal{C}^\infty(M)[[\nu]]$, which turns out to be a star product, is defined by pullback of $\circ$ via $\tau$:*

$$f * g := \sigma(\tau(f) \circ \tau(g)) \qquad [22]$$

In the following, we shall refer to the associative product $*$ defined above as the Fedosov star product corresponding to $(\nabla, \Omega)$. The choice of the formal series of closed 2-forms $\Omega$ in fact has a crucial effect on the equivalence class of the resulting star product, whereas the choice of the torsion-free symplectic connection, which in contrast to a Riemannian connection is not unique, does not affect this class. This observation has been the main step in all the proofs of the classification results in deformation quantization of symplectic manifolds. Another way to prove this fact is to compute the characteristic class $c(*)$ introduced by Deligne (1995) using the methods developed in Gutt and Rawnsley (1999) directly which yields:

**Theorem 4** (Neumaier 2002, theorem 2). *Deligne's characteristic class $c(*)$ of a Fedosov star product $*$ as constructed above is given by*

$$c(*) = \frac{1}{\nu}[\omega] + \frac{1}{\nu}[\Omega] \qquad [23]$$

The properties of $\Omega$ with respect to complex conjugation also decide on whether $*$ is Hermitian or not. In case $\Omega$ is real, that is, satisfies $\overline{\Omega} = \Omega$ it is easy to show – observing that $\overline{a \circ b} = (-1)^{kl}\overline{b} \circ \overline{a}$ for $a \in \mathcal{W} \otimes \Lambda^k, b \in \mathcal{W} \otimes \Lambda^l$ – that $\overline{r}$ solves the equations that uniquely determine $r$ and hence $\overline{r} = r$. But then $\mathfrak{D}$ commutes with complex conjugation and

therefore the unique characterization of the Fedosov–Taylor series yields $\overline{\tau(f)} = \tau(\bar{f})$ for all $f \in \mathcal{C}^\infty(M)[[\nu]]$, implying that $*$ is Hermitian.

## Derivations, Automorphisms, and Equivalence Transformations

Having defined the Fedosov star product $*$ corresponding to $(\nabla, \Omega)$, the next logical step is to investigate the structure of its derivations and automorphisms and to find out how they can be described in the framework of Fedosov's construction. In addition, one can ask for an explicit construction of equivalence transformations between two Fedosov star products $*$ and $*'$ obtained from $(\nabla, \Omega)$ and $(\nabla', \Omega')$ that exist according to Theorem 4 iff $[\Omega] = [\Omega']$.

Since the basic philosophy of Fedosov's construction is to consider suitable operations on the algebra $(\mathcal{W} \otimes \Lambda, \circ)$ in order to obtain induced mappings on the level of $(\mathcal{C}^\infty(M)[[\nu]], *)$, one may expect to be able to define derivations of $(\mathcal{C}^\infty(M)[[\nu]], *)$ by considering appropriate fiberwise quasi-inner derivations of $(\mathcal{W} \otimes \Lambda, \circ)$ of the shape

$$\mathcal{D}_h = -\frac{1}{\nu}\mathrm{ad}(h) \qquad [24]$$

where $h \in \mathcal{W}$ and without loss of generality we assume $\sigma(h) = 0$. Our aim is to define $\mathbb{C}[[\nu]]$-linear derivations of $*$ by $\mathcal{C}^\infty(M)[[\nu]] \ni f \mapsto \sigma(\mathcal{D}_h \tau(f))$, but for an arbitrary element $h \in \mathcal{W}$ with $\sigma(h) = 0$ this mapping fails to be a derivation as $\mathcal{D}_h$ does not map elements of $\ker(\mathfrak{D}) \cap \mathcal{W}$ to elements of $\ker(\mathfrak{D}) \cap \mathcal{W}$. In order to achieve this, the supercommutator of $\mathfrak{D}$ and $\mathcal{D}_h$ has to vanish. As $\mathfrak{D}$ is a $\mathbb{C}[[\nu]]$-linear $\circ$-superderivation, we obviously have

$$[\mathfrak{D}, \mathcal{D}_h] = -\frac{1}{\nu}\mathrm{ad}(\mathfrak{D}h) \qquad [25]$$

and hence obviously $\mathfrak{D}h$ must be central, that is, $\mathfrak{D}h$ has to be of the shape $1 \otimes B$ with $B \in \Gamma^\infty(T^*M)[[\nu]]$ to have $[\mathfrak{D}, \mathcal{D}_h] = 0$. From $\mathfrak{D}^2 = 0$, we get that the necessary condition for the solvability of the equation $\mathfrak{D}h = 1 \otimes B$ is the closedness of $B$ since $\mathfrak{D}(1 \otimes B) = 1 \otimes \mathrm{d}B$. But as the $\mathfrak{D}$-cohomology is trivial on elements with positive antisymmetric degree, this condition is also sufficient for the solvability of the equation $\mathfrak{D}h = 1 \otimes B$ and we get the following statement.

**Lemma 1** (Müller-Bahns and Neumaier 2004, lemma 2.1).

(i) *For all formal series $B \in \Gamma^\infty(T^*M)[[\nu]]$ of closed 1-forms on $M$ there is a uniquely determined element $h_B \in \mathcal{W}$ such that $\mathfrak{D}h_B = 1 \otimes B$ and $\sigma(h_B) = 0$. Moreover, $h_B$ is explicitly given by*

$$h_B = \mathfrak{D}^{-1}(1 \otimes B) \qquad [26]$$

(ii) *For all $B \in Z^1_{\mathrm{deRham}}(M)[[\nu]]$ the mapping $D_B : \mathcal{C}^\infty(M)[[\nu]] \to \mathcal{C}^\infty(M)[[\nu]]$, where*

$$D_B f := \sigma(\mathcal{D}_{h_B}\tau(f)) = \sigma\left(-\frac{1}{\nu}\mathrm{ad}(h_B)\tau(f)\right) \qquad [27]$$

*for $f \in \mathcal{C}^\infty(M)[[\nu]]$ defines a $\mathbb{C}[[\nu]]$-linear derivation of $*$ and hence this construction yields a mapping $Z^1_{\mathrm{deRham}}(M)[[\nu]] \ni B \mapsto D_B \in \mathsf{Der}_{\mathbb{C}[[\nu]]}(\mathcal{C}^\infty(M)[[\nu]], \star)$.*

Furthermore, one can show that one even obtains all $\mathbb{C}[[\nu]]$-linear derivations of $*$ by varying $B$ in the derivations $D_B$ constructed above.

**Proposition 1** (Müller-Bahns and Neumaier 2004, proposition 2.2). *The mapping*

$$Z^1_{\mathrm{deRham}}(M)[[\nu]] \ni B \mapsto D_B \in \mathsf{Der}_{\mathbb{C}[[\nu]]}(\mathcal{C}^\infty(M)[[\nu]], *)$$

*defined in Lemma 1 is a bijection. Moreover, $D_{\mathrm{d}f}$ is a quasi-inner derivation for all $f \in \mathcal{C}^\infty(M)[[\nu]]$, that is, $D_{\mathrm{d}f} = (1/\nu)\mathrm{ad}_*(f)$ and the induced mapping $[B] \mapsto [D_B]$ from $H^1_{\mathrm{deRham}}(M)[[\nu]]$ to $\mathsf{Der}_{\mathbb{C}[[\nu]]}(\mathcal{C}^\infty(M)[[\nu]], *)/\mathsf{Der}^{\mathrm{qi}}_{\mathbb{C}[[\nu]]}(\mathcal{C}^\infty(M)[[\nu]], *)$ the space of $\mathbb{C}[[\nu]]$-linear derivations of $*$ modulo the quasi-inner derivations, also is bijective.*

Actually, it is well known that for an arbitrary star product $\star$ on a symplectic manifold the space of $\mathbb{C}[[\nu]]$-linear derivations is in bijection with $Z^1_{\mathrm{deRham}}(M)[[\nu]]$ and that the quotient space of these derivations modulo the quasi-inner derivations is in bijection with $H^1_{\mathrm{deRham}}(M)[[\nu]]$ (cf. Bertelson *et al.* (1997), theorem 4.2), but the remarkable thing about Fedosov star products is that these bijections can be explicitly expressed in terms of $\mathfrak{D}$ resp. $\mathfrak{D}^{-1}$ in a very lucid way.

Now we turn to the consideration of $\mathbb{C}[[\nu]]$-linear automorphisms of $*$. For such automorphisms that start with id, which are also called self-equivalences, it is known (cf. Gutt and Rawnsley (1999), Proposition 3.3) for arbitrary star products $\star$ on $(M, \omega)$ that they are of the form

$$A = \exp(\nu D) \qquad [28]$$

with a $\mathbb{C}[[\nu]]$-linear derivation $D$ of $\star$. Therefore, the above result about the description of all the derivations of $*$ directly yields a complete description of all self-equivalences of $*$.

The description of $\mathbb{C}[[\nu]]$-linear automorphisms that are not self-equivalences of $*$ is slightly more involved and we first need some results about the concrete structure of the equivalence transformations between two Fedosov star products $*$ and $*'$. To compare two Fedosov star products obtained from different torsion-free symplectic connections

$\nabla$ and $\nabla'$ and different but cohomologous formal series of 2-forms $\Omega$ and $\Omega'$, one has to compare the corresponding Fedosov derivations $\mathfrak{D}$ and $\mathfrak{D}'$. First recall some well-known facts about torsion-free symplectic connections on $(M,\omega)$. Given two such connections $\nabla$ and $\nabla'$, it is obvious that $S^{\nabla-\nabla'}(X,Y) := \nabla_X Y - \nabla'_X Y$, where $X, Y \in \Gamma^\infty(TM)$ defines a symmetric tensor field $S^{\nabla-\nabla'} \in \Gamma^\infty(\bigvee^2 T^*M \otimes TM)$ on $M$. Defining $\sigma^{\nabla-\nabla'}(X,Y,Z) := \omega(S^{\nabla-\nabla'}(X,Y),Z)$ it is easy to see that $\sigma^{\nabla-\nabla'} \in \Gamma^\infty(\bigvee^3 T^*M)$ is a totally symmetric tensor field. Conversely, given an arbitrary element $\sigma \in \Gamma^\infty(\bigvee^3 T^*M)$ and a symplectic torsion-free connection $\nabla$ and defining $S^\sigma \in \Gamma^\infty(\bigvee^2 T^*M \otimes TM)$ by $\sigma(X,Y,Z) = \omega(S^\sigma(X,Y),Z)$, then $\nabla^\sigma$ defined by $\nabla^\sigma_X Y := \nabla_X Y - S^\sigma(X,Y)$ again is a torsion-free symplectic connection and all such connections can be obtained this way by varying $\sigma$. Using these relations, one can compare the corresponding mappings $\nabla$ and $\nabla'$ on $\mathcal{W}\otimes\Lambda$. With the notations from above we have

$$\nabla - \nabla' = -(\mathrm{d}x^j \otimes \mathrm{d}x^i) i_s(S^{\nabla-\nabla'}(\partial_i,\partial_j)) = \frac{1}{\nu}\mathrm{ad}(T^{\nabla-\nabla'}) \qquad [29]$$

where $T^{\nabla-\nabla'} \in \Gamma^\infty(\bigvee^2 T^*M \otimes T^*M) \subseteq \mathcal{W}\otimes\Lambda^1$ is defined by $T^{\nabla-\nabla'}(Z,Y;X) := \sigma^{\nabla-\nabla'}(X,Y,Z) = \omega(S^{\nabla-\nabla'}(X,Y),Z)$. Moreover, $T^{\nabla-\nabla'}$ satisfies the equations

$$\delta T^{\nabla-\nabla'} = 0 \qquad [30a]$$

and

$$\nabla T^{\nabla-\nabla'} = R' - R + \frac{1}{\nu} T^{\nabla-\nabla'} \circ T^{\nabla-\nabla'}$$
$$\nabla' T^{\nabla-\nabla'} = R' - R - \frac{1}{\nu} T^{\nabla-\nabla'} \circ T^{\nabla-\nabla'} \qquad [30b]$$

where $R = (1/4)\omega_{it} R^t_{jkl} \mathrm{d}x^i \vee \mathrm{d}x^j \otimes \mathrm{d}x^k \wedge \mathrm{d}x^l$ and $R' = (1/4)\omega_{it} R'^t_{jkl} \mathrm{d}x^i \vee \mathrm{d}x^j \otimes \mathrm{d}x^k \wedge \mathrm{d}x^l$ denote the corresponding elements of $\mathcal{W}\otimes\Lambda^2$ that are built from the curvature tensors of $\nabla$ and $\nabla'$.

Now we are in the position to compare two Fedosov derivations $\mathfrak{D}$ and $\mathfrak{D}'$ resp. the induced star products $*$ and $*'$ obtained from $(\nabla,\Omega)$ and $(\nabla',\Omega)$. The idea for the construction of an equivalence transformation from $*$ to $*'$ is to look for an automorphism $\mathcal{A}_h$ of $(\mathcal{W}\otimes\Lambda, \circ)$ of the form

$$\mathcal{A}_h = \exp\left(\frac{1}{\nu}\mathrm{ad}(h)\right) \quad \text{such that } \mathfrak{D}' = \mathcal{A}_h \mathfrak{D} (\mathcal{A}_h)^{-1} \qquad [31]$$

where $h$ is an element of $\mathcal{W}_3$ guaranteeing that $\mathcal{A}_h$ is well defined and without loss of generality is assumed to satisfy $\sigma(h)=0$. In case one can find such an element $h$ it is clear that $\mathcal{A}_h$ yields a bijection between $\ker(\mathfrak{D})\cap\mathcal{W}$ and $\ker(\mathfrak{D}')\cap\mathcal{W}$ and hence one would obtain an equivalence $S_h$ from $*$ to $*'$ defining

$$S_h f := \sigma(\mathcal{A}_h \tau(f)) = \sigma\left(\exp\left(\frac{1}{\nu}\mathrm{ad}(h)\right)\tau(f)\right) \qquad [32a]$$

with inverse

$$(S_h)^{-1} f = \sigma((\mathcal{A}_h)^{-1}\tau'(f)) = \sigma\left(\exp\left(-\frac{1}{\nu}\mathrm{ad}(h)\right)\tau'(f)\right) \qquad [32b]$$

A direct computation yields

$$\mathcal{A}_h \mathfrak{D}(\mathcal{A}_h)^{-1} = \mathfrak{D} - \frac{1}{\nu}\mathrm{ad}\left(\frac{\exp((1/\nu)\mathrm{ad}(h)) - \mathrm{id}}{(1/\nu)\mathrm{ad}(h)}(\mathfrak{D}h)\right)$$

which is equal to $\mathfrak{D}'$ iff $h$ has been chosen such that

$$T^{\nabla-\nabla'} + r' - r - \frac{\exp(\frac{1}{\nu}\mathrm{ad}(h)) - \mathrm{id}}{\frac{1}{\nu}\mathrm{ad}(h)}(\mathfrak{D}h) \in \mathcal{W}\otimes\Lambda^1 \qquad [33]$$

is a central element. Considering the total degree of the terms in this expression, this is the case iff there is a formal series of 1-forms $C \in \nu\Gamma^\infty(T^*M)[[\nu]]$ such that the expression in eqn [33] equals $1\otimes C$. Applying $\mathfrak{D}$ to this equation and using the equations that $r$ and $r'$ satisfy together with the relations [30] it is cumbersome but not difficult to show that necessarily $\Omega$ and $\Omega'$ have to be cohomologous:

$$\Omega - \Omega' = \mathrm{d}C \qquad [34]$$

with $C$ as above. Now, using [6] one can show that this condition is in fact sufficient and moreover one can even determine the element $h$ in question recursively:

**Theorem 5** (Fedosov 1994, theorem 4.3). *Two Fedosov star products $*$ and $*'$ obtained from $(\nabla,\Omega)$ and $(\nabla',\Omega')$ are equivalent iff $\Omega$ and $\Omega'$ are cohomologous. In case $C \in \nu\Gamma^\infty(T^*M)[[\nu]]$ satisfies $\Omega - \Omega' = \mathrm{d}C$ there is a uniquely determined element $h_C \in \mathcal{W}_3$ with $\sigma(h_C)=0$ such that*

$$T^{\nabla-\nabla'} + r' - r - \frac{\exp((1/\nu)\mathrm{ad}(h_C)) - \mathrm{id}}{(1/\nu)\mathrm{ad}(h_C)}(\mathfrak{D}h_C) = 1\otimes C$$

*Moreover, $h_C$ can be determined recursively from*

$$h_C = C\otimes 1 + \delta^{-1}\left(\nabla h_C - \frac{1}{\nu}\mathrm{ad}(r)h_C - \frac{(1/\nu)\mathrm{ad}(h_C)}{\exp((1/\nu)\mathrm{ad}(h_C)) - \mathrm{id}}(r' - r + T^{\nabla-\nabla'})\right) \qquad [35]$$

*and with the so-constructed $h_C$ one has $\mathfrak{D}' = \mathcal{A}_{h_C}\mathfrak{D}(\mathcal{A}_{h_C})^{-1}$ and thus $S_{h_C}$ according to eqn [32] defines an equivalence transformation from $*$ to $*'$.*

Evidently, in the above construction of the equivalence transformation $S_{h_C}$ there is some choice of the formal series of 1-forms $C$. Different possible choices $C$ and $\widetilde{C}$ differ by a formal series of closed 1-forms but choosing $\widetilde{C}$ instead of $C$ amounts to another equivalence transformation $S_{h_{\widetilde{C}}} = A' S_{h_C} A$ from $*$ to $*'$, where $A$ and $A'$ are certain self-equivalences of $*$ and $*'$, respectively. In case $\Omega$ and $\Omega'$ are real, we have seen that $*$ as well as $*'$ are Hermitian star products and it is easy to verify that choosing a formal series $C$ of 1-forms as above that is moreover real yields an element $h_C$ satisfying $\overline{h_C} = h_C$. But then it is evident that the resulting equivalence transformation is also compatible with complex conjugation, that is, $\overline{S_{h_C} f} = S_{h_C}\overline{f}$ for all $f \in \mathcal{C}^\infty(M)[[\nu]]$.

Now we are prepared to give a construction of all $\mathbb{C}[[\nu]]$-linear automorphisms of a Fedosov star product $*$. It is easy to show that any $\mathbb{C}[[\nu]]$-linear automorphism of a star product $\star$ on a symplectic manifold is the combination of the action of a symplectomorphism $\psi : M \to M$ and an equivalence between $\star$ and the pullback $\star'$ via $\psi^{-1}$ of $\star$, which is defined by $f \star' g = (\psi^{-1})^*((\psi^* f) \star (\psi^* g))$ (cf. Gutt and Rawnsley (1999 Proposition 9.4)). Since the characteristic class of $\star'$ is given by $c(\star') = (\psi^{-1})^* c(\star)$, the necessary and sufficient condition for a symplectomorphism $\psi$ to define a possible zeroth-order term of an automorphism is that $(\psi^{-1})^* c(\star) = c(\star)$ since $\star'$ and $\star$ have to be equivalent.

Within Fedosov's framework, it can be shown that the pullback $*'$ via a symplectomorphism $\psi^{-1}$ of $*$ is identical to the Fedosov star product obtained from $(\nabla' = (\psi^{-1})^* \nabla \psi^*, \Omega' = (\psi^{-1})^* \Omega)$, which just expresses the functoriality of Fedosov's construction. Together with Theorem 4 this particularly shows that $c(*') = (\psi^{-1})^* c(*)$, and therefore $*'$ is equivalent to $*$ iff $\Omega$ and $\Omega'$ differ by a formal series $\mathrm{d}C_\psi$ of exact 2-forms, where $C_\psi \in \nu\Gamma^\infty(T^*M)[[\nu]]$ clearly depends on $\psi$. But in this situation one can apply the construction of equivalence transformations between Fedosov star products given in Theorem 5 with $C$ replaced by $C_\psi$ and $\nabla', \Omega'$ as above yielding an equivalence $S_{h_\psi} := S_{h_{C_\psi}}$ from $*$ to $*'$. Finally, we therefore get that the combination

$$A_\psi := \psi^* S_{h_\psi} \qquad [36]$$

is a $\mathbb{C}[[\nu]]$-linear automorphism of $*$ and it is obvious from the above that every such automorphism can be obtained by considering all symplectomorphisms $\psi$ of $(M, \omega)$ satisfying $[(\psi^{-1})^*\Omega] = [\Omega]$ and composing the resulting $A_\psi$ according to [36] with all self-equivalences $A$ of $*$ according to [28].

## Adaptions, Modifications, and Generalizations

The geometrical construction of Fedosov has gone through many adaptions and modifications that are well suited to the particular geometry of the underlying symplectic manifold. Moreover, there are generalizations that go beyond the case of symplectic manifolds and others that yield more general deformations than star products. We just give a few important examples that stress the power and beauty of Fedosov's construction.

On a Kähler manifold, one can define the notion of star products with separation of variables (cf. Karabegov (1996) that are also called star products of Wick type (cf. Bordemann and Waldmann (1997) and Neumaier (2003). These are star products such that in local holomorphic coordinates the bidifferential operators $C_r$ are of the form

$$C_r(f, g) = \sum_{K, \overline{L}} C_r^{K;\overline{L}} \frac{\partial^{|K|} f}{\partial z^K} \frac{\partial^{|\overline{L}|} g}{\partial \overline{z}^L} \qquad [37]$$

with certain coefficient functions $C_r^{K;\overline{L}}$. These star products can be obtained by a modified Fedosov construction starting from the product $\circ_{\mathrm{Wick}}$ on $\mathcal{W} \otimes \Lambda$ given by

$$a \circ_{\mathrm{Wick}} b = \mu\left(\exp\left(\frac{2\nu}{\mathrm{i}} g^{k\bar{l}} i_s(\partial_{z_k}) \otimes i_s(\partial_{\bar{z}_l})\right)(a \otimes b)\right) \qquad [38]$$

where $g^{k\bar{l}}$ denotes the components of the inverse of the Kähler metric in local holomorphic coordinates. In the case of a Kähler manifold, there is a distinguished torsion-free symplectic connection namely the Kähler connection $\nabla$ that induces a superderivation of $\circ_{\mathrm{Wick}}$ in a way completely analogous to [8]. With these structures the Fedosov construction works for an arbitrary formal series $\Omega$ of closed 2-forms as before, but one can show that the resulting star product is of Wick type iff $\Omega$ is of type (1, 1) and one can even show that one obtains all star products of Wick type by varying $\Omega$ (cf. Neumaier (2003)).

In the case of an almost-Kähler manifold, one can consider a product $\circ'$ on $\mathcal{W} \otimes \Lambda$ similar to $\circ_{\mathrm{Wick}}$ which is adapted to the almost-complex structure (cf. Karabegov and Schlichenmaier (2001)). However, in this situation there is no torsion-free connection that yields a superderivation of this product but only a connection $\nabla'$ with torsion that defines such a superderivation. Nevertheless, one

can consider a generalized Fedosov construction. To this end, one shows that $[\delta, \nabla'] = (1/\nu)\mathrm{ad}'(T')$ with some $T' \in \mathcal{W} \otimes \Lambda^2$ that satisfies $\delta T' = 0$ and encodes the torsion of $\nabla'$ and $\delta R' = \nabla' T'$, where again $\nabla'^2 = -(1/\nu)\mathrm{ad}'(R')$ and $R'$, which depends on the curvature of $\nabla'$, satisfies $\nabla' R' = 0$. But then it is easy to show that there is a unique element $r' \in \mathcal{W}_2 \otimes \Lambda^1$ such that

$$\delta r' = \nabla' r' - \frac{1}{\nu} r' \circ' r' + T' + R' + 1 \otimes \Omega$$

and

$$\delta^{-1} r' = 0 \qquad [39]$$

with $\Omega$ as above, which can also be computed recursively. Clearly, $\mathfrak{D}' = -\delta + \nabla' - (1/\nu)\mathrm{ad}'(r')$ then is a suitable Fedosov derivation with square zero for $\circ'$ and one can proceed as described earlier to obtain a star product $*'$ adapted to the almost-complex structure.

On a cotangent bundle $\pi : T^*Q \to Q$, where $T^*Q$ is equipped with the canonical symplectic form $\omega_0 = -\mathrm{d}\theta_0$, one can consider (cf. Bordemann *et al.* (1998)) the following so-called standard ordered product $\circ_{\mathrm{std}}$ on $\mathcal{W} \otimes \Lambda$ given by

$$a \circ_{\mathrm{std}} b = \mu\Big(\exp\Big(-\nu i_{\mathrm{s}}(\partial_{p_i}) \otimes i_{\mathrm{s}}(\partial_{q^i} + p_l \pi^* \Gamma^l_{ik} \partial_{p_k})\Big) \times (a \otimes b)\Big) \qquad [40]$$

in local Darboux coordinates. Here $\Gamma^l_{ik}$ denotes the Christoffel symbols of a torsion-free connection $\nabla^Q$ on $Q$ in the chart of $Q$ corresponding to the bundle chart $(q, p)$ and it is straightforward to see that $\circ_{\mathrm{std}}$ does not depend on the chosen local coordinates and is associative. In the present situation, one can define a torsion-free symplectic connection $\nabla^{T^*Q}$ on $T^*Q$ solely in terms of $\nabla^Q$ but then the corresponding mapping $\nabla^{T^*Q}$ on $\mathcal{W} \otimes \Lambda$ again fails to be a superderivation of $\circ_{\mathrm{std}}$, whereas the combination $\nabla^{T^*Q} + \mathcal{B}$ with $\mathcal{B} = (\nu/3) p_l \pi^* R^l_{jik} (1 \otimes \mathrm{d}q^i) i_{\mathrm{s}}(\partial_{p_j}) i_{\mathrm{s}}(\partial_{p_k})$, where $R^l_{jik}$ denotes the components of the curvature tensor of $\nabla^Q$, turns out to be a suitable superderivation to start the Fedosov construction with $\circ_{\mathrm{std}}$. In fact, the square of $\nabla^{T^*Q} + \mathcal{B}$ turns out to equal the square of $\nabla^{T^*Q}$ and all the other preconditions of Fedosov's construction are easily verified just replacing $\nabla$ by $\nabla^{T^*Q} + \mathcal{B}$. The particular property of the resulting star product $*_{\mathrm{std}}$ for $\Omega = 0$ on $T^*Q$ is that it is a standard ordered star product, that is, for all $f \in \mathcal{C}^\infty(T^*Q)[[\nu]]$ and all $\chi \in \mathcal{C}^\infty(Q)[[\nu]]$ one has

$$\pi^* \chi *_{\mathrm{std}} f = \pi^* \chi f \qquad [41]$$

and hence $*_{\mathrm{std}}$ in a certain sense is adapted to the vertical polarization.

The methods mentioned so far can even be melted into a more general situation, where one considers a (complex) polarization on $(M, \omega)$ and looks for star products that are adapted to this polarization which are then called polarized deformation quantizations (cf. Donin (2003)). Here again a generalization of Fedosov's construction yields the existence and the classification of such particular star products.

Another recent generalization of Fedosov's construction that goes beyond the framework of smooth symplectic manifolds is that of the construction of star products on symplectic orbispaces (cf. Pflaum (2003)), which are stratified symplectic spaces. The main idea there is to consider Fedosov's construction in local orbicharts and to show that the changes of orbicharts induce isomorphisms between the locally defined deformation quantizations, implying that the locally defined products match together to define a global deformation quantization on the symplectic orbispace. To achieve this property, one has to adjust the local Fedosov constructions appropriately, that is, one has to use locally defined torsion-free symplectic connections and formal series of closed 2-forms that are related by the changes of the orbicharts.

Considering a vector bundle $E \to M$, the sections $\Gamma^\infty(E)$ are naturally a $\mathcal{C}^\infty(M)$-right module and a $\Gamma^\infty(\mathsf{End}(E))$-left module, and it is a natural question whether this bimodule structure can be deformed such that $\Gamma^\infty(E)[[\nu]]$ becomes a $(\mathcal{C}^\infty(M)[[\nu]], \star)$-right module and a $(\Gamma^\infty(\mathsf{End}(E))[[\nu]], \bigstar)$-left module, where $\bigstar$ is a deformation of the usual composition of elements of $\Gamma^\infty(\mathsf{End}(E))$. In order to construct such deformations, one can also adapt Fedosov's construction (cf. Waldmann (2002)) considering $\mathcal{W} \otimes \Lambda \otimes \mathcal{E} = (\mathsf{X}_{s=0}^\infty \Gamma^\infty(\bigvee^s T^*M \otimes \bigwedge T^*M \otimes E))[[\nu]]$ and $\mathcal{W} \otimes \Lambda \otimes \mathcal{E}nd(\mathcal{E}) = (\mathsf{X}_{s=0}^\infty \Gamma^\infty \ (\bigvee^s T^*M \otimes \bigwedge T^*M \otimes \mathsf{End}(E)))[[\nu]]$ and extending the product $\circ$ to these spaces in a natural way making $\mathcal{W} \otimes \Lambda \otimes \mathcal{E}$ a $(\mathcal{W} \otimes \Lambda \otimes \mathsf{End}(\mathcal{E}), \circ) - (\mathcal{W} \otimes \Lambda, \circ)$-bimodule. Furthermore, one has to consider a connection $\nabla^E$ that naturally induces a connection on $\mathsf{End}(E)$, and both have to be added to $\nabla$ to define the corresponding substitute of $\nabla$ on the respective space. Then the Fedosov construction with $\mathcal{W} \otimes \Lambda \otimes \mathcal{E}nd(\mathcal{E})$ can be considered yielding a Fedosov derivation $\mathfrak{D}^{\mathsf{End}(E)}$ with square zero, hence a Fedosov–Taylor series $\tau^{\mathsf{End}(E)}$ and an associative deformation $F \circledast G = \sigma(\tau^{\mathsf{End}(E)}(F) \circ \tau^{\mathsf{End}(E)}(G))$ of the usual composition of sections in the endomorphism bundle. Moreover, there is a map $\mathfrak{D}^E$ on $\mathcal{W} \otimes \Lambda \otimes \mathcal{E}$ that is a superderivation with respect to the bimodule multiplication $\circ$ along $\mathfrak{D}^{\mathsf{End}(E)}$ and $\mathfrak{D}$, respectively. This map also has square zero and the intersection of its kernel with the elements of antisymmetric degree is in bijection to $\Gamma^\infty(E)[[\nu]]$ via a natural generalization $\tau^E$ of the Fedosov–Taylor

series. Defining $F \odot s := \sigma(\tau^{\mathsf{End}(E)}(F) \circ \tau^E(s))$ and $s \cdot f := \sigma(\tau^E(s) \circ \tau(f)), \Gamma^\infty(E)[[\nu]]$ can be given the structure of a $(\Gamma^\infty(\mathsf{End}(E))[[\nu]], \circledast)$-$(\mathcal{C}^\infty(M)[[\nu]], *)$-bimodule which is indeed a deformation of the classical bimodule structure of $\Gamma^\infty(E)$. It is rather evident that the same procedure also works for other products on $\mathcal{W} \otimes \Lambda$ and the above generalizations, in particular for the product $\circ_{\mathrm{Wick}}$ on a Kähler manifold, where one can obtain $(\Gamma^\infty(\mathsf{End}(E))[[\nu]], \circledast_{\mathrm{Wick}})$-$(\mathcal{C}^\infty(M)\ [[\nu]], *_{\mathrm{Wick}})$-bimodules that are adapted to the complex structure in case the curvature endomorphism of the connection $\nabla^E$ is of type $(1,1)$. For example, this holds true for (anti-) holomorphic vector bundles endowed with a Hermitian fiber metric $h$ and the corresponding connection that is compatible with $h$ and the (anti-) holomorphic structure.

Finally, the proof of existence of deformation quantizations on arbitrary Poisson manifolds $(M, \Pi)$, that includes a concrete construction starting from Kontsevich's star product on the flat space $\mathbb{R}^n$ equipped with a Poisson tensor, given by Cattaneo *et al.* (2002) is similar in spirit to Fedosov's construction. There one constructs two bundles $J^\infty$ and $\mathcal{J}^\infty$ of associative algebras, where – as a bundle – $\mathcal{J}^\infty$ is isomorphic to $J^\infty[[\nu]]$ and $J^\infty$ is the bundle of infinite jets of smooth functions on $M$ which is equipped with the canonical flat connection $D_0$. The Poisson tensor gives rise to the structure of a Poisson algebra on each fiber of $J^\infty$ and the canonical map $\mathcal{C}^\infty(M) \to J^\infty$ yields a Poisson algebra isomorphism between $\mathcal{C}^\infty(M)$ and the Poisson algebra of $D_0$-flat sections in $J^\infty$. The second step in the construction consists in a deformation of this correspondence. Using the Kontsevich formula for $\mathbb{R}^n$, each fiber of $\mathcal{J}^\infty$ can be equipped with an associative product which is a deformation of the above product on the fibers of $J^\infty$ in the direction of the Poisson bracket induced by $\Pi$. Then analogously to Fedosov's construction, one constructs a compatible connection $D = D_0 + \nu D_1 + \nu^2 D_2 + \cdots$ which is a deformation of $D_0$. Here compatibility just means that $D$ is a derivation with respect to the above product on sections in $\mathcal{J}^\infty$ implying that the $D$-flat sections form a subalgebra. Moreover, one can achieve that this connection is flat and in this case $\mathcal{C}^\infty(M)[[\nu]]$ turns out to be in bijection to the $D$-flat sections in $\mathcal{J}^\infty$. For the proof of existence of $D$ and for its recursive determination using an adaption of Fedosov's method, again special cases of Kontsevich's formality theorem prove to be the crucial tools. Pulling back the above fiberwise product to $\mathcal{C}^\infty(M)[[\nu]]$ via this isomorphism, one then obtains a star product on $(M, \Pi)$. Since this isomorphism can be determined recursively, the star product can in principle be computed explicitly.

*See also:* Deformation Quantization; Deformation Quantization and Representation Theory; Deformation Theory; Deformations of the Poisson Bracket on a Symplectic Manifold; String Field Theory.

## Further Reading


Bayen F, Flato M, Frønsdal C, Lichnerowicz A, and Sternheimer D (1978) Deformation theory and quantization. *Annals of Physics* 111: 61–110 Part I.

Bayen F, Flato M, Frønsdal C, Lichnerowicz A, and Sternheimer D (1978) Deformation theory and quantization. *Annals of Physics* 111: 111–151 Part II.

Bertelson M, Cahen M, and Gutt S (1997) Equivalence of star products. *Classical and Quantum Gravity* 14: A93–A107.

Bordemann M and Waldmann S (1997) A Fedosov star product of the Wick type for Kahler manifolds. *Letters in Mathematical Physics* 41: 243–253.

Bordemann M, Neumaier N, and Waldmann S (1998) Homogeneous Fedosov star products on cotangent bundles. I. Weyl and standard ordering with differential operator representation. *Communications in Mathematical Physics* 198: 363–396.

Cattaneo AS, Felder G, and Tomassini L (2002) From local to global deformation quantization of Poisson manifolds. *Duke Mathematical Journal* 115: 329–352.

Deligne P (1995) Déformations de l'Algèbre des Fonctions d'une Variété Symplectique: Comparaison entre Fedosov et DeWilde, Lecomte. *Selecta Mathematical, New Series* 1(4): 667–697.

Donin J (2003) Classification of polarized deformation quantizations. *Journal of Geometry and Physics* 48: 546–579.

DeWilde M and Lecomte PBA (1983) Existence of star-products and of formal deformations of the Poisson Lie algebra of arbitrary symplectic manifolds. *Letters in Mathematical Physics* 7: 487–496.

Fedosov BV (1985) Formal Quantization. In: *Some Topics of Modern Mathematics and Their Applications to Problems of Mathematical Physics*, Moscow, pp. 129–136.

Fedosov BV (1986) Quantization and the index. *Soviet Physics–Doklady* 31(11): 877–878.

Fedosov BV (1994) A simple geometrical construction of deformation quantization. *Journal of Differential Geometry* 40: 213–238.

Fedosov BV (1996) *Deformation Quantization and Index Theory*. Berlin: Akademie Verlag.

Gutt S and Rawnsley J (1999) Equivalence of star products on a symplectic manifold; an introduction to Deligne's Čech cohomology classes. *Journal of Geometry and Physics* 29: 347–392.

Karabegov AV (1996) Deformation quantizations with separation of variables on a Kähler manifold. *Communications in Mathematical Physics* 180: 745–755.

Karabegov A and Schlichenmaier M (2001) Almost-Kähler deformation quantization. *Letters in Mathematical Physics* 57: 135–148.

Kontsevich M (2003) Deformation quantization of Poisson manifolds. *Letters in Mathematical Physics* 66: 157–216.

Müller-Bahns MF and Neumaier N (2004) Some remarks on $\mathfrak{g}$-invariant Fedosov star products and quantum momentum mappings. *Journal of Geometry and Physics* 50: 257–272.

Nest R and Tsygan B (1995) Algebraic index theorem. *Communications in Mathematical Physics* 172: 223–262.

Neumaier N (2002) Local $\nu$-Euler derivations and Deligne's characteristic class of Fedosov star products and star products

of special type. *Communications in Mathematical Physics* 230: 271–288.
Neumaier N (2003) Universality of Fedosov's construction for star products of Wick type on pseudo-Kähler manifolds. *Reports on Mathematical Physics* 52: 43–80.
Omori H, Maeda Y, and Yoshioka A (1991) Weyl manifolds and deformation quantization. *Advances in Mathematics* 85: 224–255.
Pflaum M (2003) On the deformation quantization of symplectic orbispaces. *Differential Geometry and Its Applications* 19: 343–368.
Waldmann S (2002) Morita equivalence of Fedosov star products and deformed Hermitian vector bundles. *Letters in Mathematical Physics* 60: 157–170.

**Feigenbaum Phenomenon** *see* Universality and Renormalization

# Fermionic Systems

**V Mastropietro**, Università di Roma "Tor Vergata", Rome, Italy



## Quantum Statistics

Quantum particles are described by a complex, square-integrable wave function $\Psi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ with $|\Psi|^2$ representing the probability density of finding $N$ particles at positions $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$, which will be assumed to be in a $d$-dimensional square box $V$ with side $L$ and periodic boundary conditions. If the $N$ particles are identical, $|\Psi|^2$ must be totally symmetric in the exchange of any pair of coordinates. Regarding the symmetry properties of $\Psi$ itself, it is an experimental fact (which finds its theoretical explanation in the context of relativistic quantum field theory) that only two possibilities can arise: either $\Psi$ is symmetric or it is antisymmetric, which means that $\Psi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = (-1)^P \Psi(\boldsymbol{x}_{P_1}, \ldots, \boldsymbol{x}_{P_N})$, where $P_1, \ldots, P_N$ is a permutation of $1, \ldots, N$, and $(-1)^P$ is the parity of the permutation. Particles described by a symmetric wave function are called bosons, while particles with an antisymmetric wave function are called fermions, after Bose and Fermi, who introduced these concepts. The fermionic wave function therefore vanishes if two coordinates are equal, a property called Pauli exclusion principle. Particles have an intrinsic quantized angular momentum called spin and particles with semi-integer spin are fermions, while particles with integer spin are bosons. Examples of fermions are electrons, protons, or neutrons, with spin $\sigma = \pm\hbar/2$, where $\hbar$ is the Plank constant; examples of bosons are phonons or mesons with integer spin.

The time evolution of a wave function is driven (through the Schrödinger equation) by the Hamiltonian operator, and the choice of such an operator is determined by the physical system we want to describe. One of the most important physical realizations of a fermionic system is given by the conduction electrons in solids with a crystalline structure (like metals). According to the classical theory of Drude, a crystal can be described as a lattice of atoms in which the valence electrons are lost by the atoms (which become ions) and move freely in the metal; they are responsible for the conduction properties of the crystal. However, if one assumes that the electrons are classical particles (in the sense that they obey the Newtonian mechanics), one obtains wrong predictions about the properties of crystals. One has to take into account that the conduction electrons are quantum particles and this provides us with a natural example of a fermionic system; the Hamiltonian can be taken as

$$H_N = \sum_{i=1}^{N} \left[ -\frac{\hbar^2 \partial^2_{\boldsymbol{x}_i}}{2m} + uc(\boldsymbol{x}_i) \right] + \sum_{i<j} \lambda v(\boldsymbol{x}_i - \boldsymbol{x}_j) \quad [1]$$

The first term represents the nonrelativistic kinetic energy of the electrons ($m$ is the mass), $uc(\boldsymbol{x})$ is a periodic potential due to the ions in the lattice ($c(\boldsymbol{x}) = c(\boldsymbol{x} + \boldsymbol{R})$ with $\boldsymbol{R} = (n_1 a_1, \ldots, n_d a_d), n_i \in Z$) and $\lambda v(\boldsymbol{x} - \boldsymbol{y})$ is a two-body interaction potential, which is modeled by a short-range potential to take into account, phenomenologically, the electrostatic screening. Finally, $\lambda$ and $u$ are couplings which measure the "strength" of the corresponding interaction. Much more complicated and "realistic" Hamiltonians could be considered; for instance, one can add an interaction with a stochastic field to take into account impurities in the lattice, or with a boson field to take into account the dynamics of the ions, and so on. Note also that one can study not only three-dimensional Fermi systems ($d = 3$), but also $d = 2$ or $d = 1$ systems; they can describe the conduction electrons of crystals that are anisotropic and should be considered as bidimensional or one-dimensional systems. We focus on the nonrelativistic

fermionic systems with Hamiltonian [1], which is a problem of great importance from both the conceptual and the applications point of view.

## Second-Quantization Formalism

The Hilbert space of states of a system of $N > 1$ fermions is the space $\mathcal{H}_N$ of all the complex square-integrable antisymmetric functions $\Psi(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$. Let $\{\phi_{\boldsymbol{k}}(\boldsymbol{x})\}_{\boldsymbol{k}\in R^d}$ be a basis for $\mathcal{H}_1$ (the one-particle Hilbert space of all the complex square-integrable functions $\Psi(\boldsymbol{x}_1)$), where $\boldsymbol{k}$ is an index called quantum number. Usually, the set of $\phi_{\boldsymbol{k}}(\boldsymbol{x})$ is chosen as the eigenfunctions of the single-particle Hamiltonian

$$-\frac{\hbar^2\partial_{\boldsymbol{x}}^2}{2m} + uc(\boldsymbol{x})$$

for instance, if $u = 0$ then

$$\phi_{\boldsymbol{k}}(\boldsymbol{x}) = \frac{1}{L^{d/2}}\mathrm{e}^{\mathrm{i}\boldsymbol{k}\boldsymbol{x}}$$

with $\hbar\boldsymbol{k}$ representing the momentum; due to periodic boundary conditions, $\boldsymbol{k}$ has the form $\boldsymbol{k} = (2\pi/L)\boldsymbol{n}$, $\boldsymbol{n} = n_1, \ldots, n_d$ with $n_i$ integer and $-[L/2] \le n_i \le [(L-1)/2]$. If we call $|\boldsymbol{k}_1, \ldots, \boldsymbol{k}_N\rangle$ the normalized antisymmetrization of $\phi_{\boldsymbol{k}_1}(\boldsymbol{x}_1)\phi_{\boldsymbol{k}_2}(\boldsymbol{x}_2)\cdots\phi_{\boldsymbol{k}_N}(\boldsymbol{x}_N)$ (Slater determinant), then the set of all possible $|\boldsymbol{k}_1, \ldots, \boldsymbol{k}_N\rangle$ is a basis for $\mathcal{H}_N$; $|\boldsymbol{k}_1, \ldots, \boldsymbol{k}_N\rangle$ describes a state in which the $N$ fermions have quantum numbers $\boldsymbol{k}_1, \ldots, \boldsymbol{k}_N$. One can introduce (Negele and Orland 1988, Berezin 1966) the creation or annihilation operators $a_{\boldsymbol{k}}^+, a_{\boldsymbol{k}}^-$: they are anticommuting operators,

$$\begin{aligned} \{a_{\boldsymbol{k}}^+, a_{\boldsymbol{k}'}^-\} &\equiv a_{\boldsymbol{k}}^+ a_{\boldsymbol{k}'}^- + a_{\boldsymbol{k}'}^- a_{\boldsymbol{k}}^+ = \delta_{\boldsymbol{k},\boldsymbol{k}'} \\ \{a_{\boldsymbol{k}}^+, a_{\boldsymbol{k}'}^+\} &= \{a_{\boldsymbol{k}}^-, a_{\boldsymbol{k}'}^-\} = 0 \end{aligned} \qquad [2]$$

such that $a_{\boldsymbol{k}}^+|\boldsymbol{k}_1, \ldots, \boldsymbol{k}_N\rangle = |\boldsymbol{k}, \boldsymbol{k}_1, \ldots, \boldsymbol{k}_N\rangle$ if $\boldsymbol{k} \neq \boldsymbol{k}_i$, $i = 1, \ldots, N$ and 0 otherwise; $a_{\boldsymbol{k}}^-$ is the adjoint of $a_{\boldsymbol{k}}^+$. The action of $a_{\boldsymbol{k}}^+$ is to create a particle with quantum number $\boldsymbol{k}$ if it is not present in the state, and to yield zero otherwise (according to the Pauli principle). The state $|0\rangle$ such that $a_{\boldsymbol{k}}^-|0\rangle = 0$ for all $\boldsymbol{k}$ is called the vacuum state and it represents a state with no particles. The Fock space is defined as the direct sum of the Hilbert spaces with any number of particles, and all the elements of the Fock space can be generated by linearly superposing products of creation operators acting over the vacuum state. We can extend such definitions by adding a label to such operators to take into account the spin of the particle; for example, $a_{\boldsymbol{k},\sigma}^\pm$ are creation or annihilation operators of a particle with spin $\sigma$ and position $\boldsymbol{k}$. In terms of $a_{\boldsymbol{x},\sigma}^+ = L^{-d/2}\sum_{\boldsymbol{k}}\phi_{\boldsymbol{k}}(\boldsymbol{x})a_{\boldsymbol{k},\sigma}^+$ and of its adjoint $a_{\boldsymbol{x},\sigma}^-$, the Hamiltonian can be written as

$$\begin{aligned} H = &\sum_\sigma\left[\int_V \mathrm{d}\boldsymbol{x}\, a_{\boldsymbol{x},\sigma}^+ \frac{-\hbar^2\partial_{\boldsymbol{x}}^2}{2m} a_{\boldsymbol{x},\sigma}^-\right. \\ &\left. + u\int_V \mathrm{d}\boldsymbol{x}\, c(\boldsymbol{x}) a_{\sigma,\boldsymbol{x}}^+ a_{\boldsymbol{x},\sigma}^-\right] \\ &+ \sum_{\sigma,\sigma'}\lambda\int_V \mathrm{d}\boldsymbol{x}\int_V \mathrm{d}\boldsymbol{y}\, v(\boldsymbol{x}-\boldsymbol{y}) a_{\boldsymbol{x},\sigma}^+ a_{\boldsymbol{x},\sigma}^- a_{\boldsymbol{y},\sigma'}^+ a_{\boldsymbol{y},\sigma'}^- \end{aligned} \qquad [3]$$

According to the postulates of quantum statistical mechanics, the grand canonical partition function is given by $Z = \mathrm{tr}\,\mathrm{e}^{-\beta(H-\mu N)}$, where $\beta = (\kappa T)^{-1}$, $\kappa$ is the Boltzmann constant, $T$ is the temperature, $\mu$ is the chemical potential, $N = \sum_\sigma\int \mathrm{d}\boldsymbol{x}\, a_{\boldsymbol{x},\sigma}^+ a_{\boldsymbol{x},\sigma}^-$, and tr is the trace operation over the Fock space. The thermodynamical average of an observable O is given by $\langle O\rangle = Z^{-1}\mathrm{tr}[\mathrm{e}^{-\beta(H-\mu N)}O]$. Given a fermionic system, one is often interested in its Schwinger functions defined as follows: if $\boldsymbol{x} = (\boldsymbol{x}, t)$ and $t_1 \ge t_2 \ge \cdots \ge t_s$, $s$ even, then

$$S(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_s) = \frac{\mathrm{tr}\,\mathrm{e}^{-(\beta-t_1)(H-\mu N)}\psi_{\boldsymbol{x}_1}^{\varepsilon_1}\mathrm{e}^{-(t_1-t_2)(H-\mu N)}\psi_{\boldsymbol{x}_2}^{\varepsilon_2}\cdots\mathrm{e}^{-t_s(H-\mu N)}}{\mathrm{tr}\,\mathrm{e}^{-\beta(H-\mu N)}} \qquad [4]$$

with $\varepsilon_i = \pm, -\beta/2 \le t_i \le \beta/2$; periodic and antiperiodic boundary conditions are, respectively, imposed over $x_i$ and $t_i$. From the knowledge of the Schwinger functions, one can compute all the thermodynamical properties of a system at equilibrium or close to equilibrium.

## The Free Fermi Gas

Computation of the physical observables corresponding to the complete Hamiltonian [3] is a very difficult task. The natural starting point consists in taking into account only the kinetic term by putting $\lambda = u = 0$ in [3], obtaining the free Fermi gas model. The resulting model is not trivial at all; its properties are radically different with respect to the ones of a gas of classical particles, and it is sufficient to understand many properties of matter (see, e.g., Mahan (1990)). If

$$\phi_{\boldsymbol{k}}(\boldsymbol{x}) = \frac{1}{L^{d/2}}\mathrm{e}^{\mathrm{i}\boldsymbol{k}\boldsymbol{x}}$$

then $|\boldsymbol{k}_1, \sigma_1, \ldots, \boldsymbol{k}_N, \sigma_N\rangle$ are eigenfunctions of $H$ with eigenvalue $\sum_{\boldsymbol{k},\sigma}\varepsilon(\boldsymbol{k})n_{\boldsymbol{k},\sigma}$, where $\varepsilon(\boldsymbol{k}) = \hbar^2|\boldsymbol{k}|^2/2m$ and $n_{\boldsymbol{k},\sigma} = 0{,}1$, the occupation number, is the eigenvalue $a_{\boldsymbol{k},\sigma}^+ a_{\boldsymbol{k},\sigma}^-$; $n_{\boldsymbol{k},\sigma} = 1$ if in the state there is a fermion with momentum $\boldsymbol{k}$ and spin $\sigma$, and it is zero otherwise. The eigenfunction $|\Omega\rangle$ of $H$ with lowest energy is called the ground state, and it determines the low-temperature properties of the system. In order to find the ground state $|\Omega\rangle$, one has

to minimize $\sum_{k,\sigma} \varepsilon(\boldsymbol{k}) n_{\boldsymbol{k},\sigma}$ with the constraint that $n_{\boldsymbol{k},\sigma}$ can take only the values 0 or 1 and $\sum_{\boldsymbol{k},\sigma} n_{\boldsymbol{k},\sigma} = N$; if there are many solutions to this problem, one says that the ground state is degenerate. An approximate solution is the following: if $d = 3$, one can consider a state such that $n_{\boldsymbol{k},\sigma} = 1$ if $\boldsymbol{k}$ is in a sphere of radius $k_{\mathrm{F}}$ and zero otherwise; since the number of momenta $\boldsymbol{k} = (2\pi/L)\boldsymbol{n}$ in the sphere is approximately given by

$$2\frac{4\pi k_{\mathrm{F}}^2}{3}\frac{L^3}{8\pi^2}$$

we can choose $k_{\mathrm{F}} = (3\pi^2\rho)^{1/3}$, with $\rho = NL^{-3}$. The state $\prod_{|\boldsymbol{k}| \le k_{\mathrm{F}}} a^+_{1/2,\boldsymbol{k}} a^+_{-1/2,\boldsymbol{k}} |0\rangle$ is not the true ground state when $N$,$L$ are finite, but it is a very good approximation of it and converges to it (in a suitable sense) in the limit $N$,$L \to \infty$, $\rho$ fixed. The boundary of the sphere with radius $k_{\mathrm{F}}$ in the space of momenta is called the Fermi surface, and it is a key notion in the theory of Fermi systems; if $d = 2$, it is replaced by a circle and in $d = 1$ by two points.

Coming to the thermodynamical properties, the partition function is given by

$$Z = \prod_{\boldsymbol{k}} \sum_{n_{\boldsymbol{k}}=0,1} \mathrm{e}^{-\beta(\varepsilon(\boldsymbol{k})-\mu)n_{\boldsymbol{k}}} = \prod_{\boldsymbol{k}} (1 + \mathrm{e}^{-\beta(\varepsilon_{\boldsymbol{k}}-\mu)})$$

and the specific heat by

$$C_v = -\frac{\partial}{\partial T}\frac{\partial}{\partial \beta}\log Z$$

One finds, by expressing $\mu$ in terms of $\beta$ through the relation $N = -\partial f/\partial\mu$, that if $d = 3$, in the $L \to \infty$ limit

$$C_v = \frac{\pi^2}{2}\rho\kappa\left(\frac{\kappa T}{\varepsilon_{\mathrm{F}}}\right) + O\left(\frac{\kappa T}{\varepsilon_{\mathrm{F}}}\right)^2$$

where $\varepsilon_{\mathrm{F}} = \hbar^2 k_{\mathrm{F}}^2/2m$. Early models for metals described the electrons as classical particles; however in such a case, a well-known result of classical statistical mechanics states that they should contribute to the specific heat by $\frac{3}{2}\rho\kappa$, while experimentally their contribution is much smaller. The solution of this puzzle was provided by the above formula for $C_v$; the classical value is in fact depressed by a factor

$$\frac{\pi^2}{3}\frac{\kappa T}{\varepsilon_{\mathrm{F}}}$$

which at room temperatures is $O(10^{-2})$, in agreement with experimental data. The average number of electrons with momentum $\hbar\boldsymbol{k}$ is given, in the infinite-volume limit, by

$$\left\langle a^+_{\boldsymbol{k},\sigma} a^-_{\boldsymbol{k},\sigma} \right\rangle = (1 + \mathrm{e}^{\beta(\varepsilon(\boldsymbol{k})-\mu)})^{-1}$$

At zero temperature, it reduces to $\theta(|\boldsymbol{k}| \le k_F)$, that is, it has a discontinuity at the Fermi surface, while at high temperatures it is very close to the Maxwell distribution $\simeq \mathrm{e}^{-\beta(\varepsilon(\boldsymbol{k})-\mu)}$.

Finally, in the free Fermi gas model, all Schwinger functions can be computed. One finds that, if, for instance, $\varepsilon_i = +$ for $i = 1, 2, \ldots, s/2$ and $\varepsilon_i = -$ otherwise, that the Schwinger function with $s \ge 4$ can be expressed as sum of products of the $s = 2$ Schwinger function (also called the propagator)

$$S(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_s) = \sum_{\pi} (-1)^{\pi} \prod_{i,j} S_0\left(\boldsymbol{x}_i - \boldsymbol{x}_{\pi(j)}\right) \qquad [5]$$

where $i = 1, \ldots, s/2$, $j = s/2 + 1, \ldots, s$, $\pi_j$ is a permutation of $j = s/2 + 1, \ldots, s$,$(-1)^{\pi}$ is the parity of this permutation, $\sum_{\pi}$ is the sum over all the possible permutations; such a formula is called the Wick rule. By an explicit computation, $S_0(\boldsymbol{x} - \boldsymbol{y})$ is given by

$$\begin{aligned} &\frac{2\pi}{\beta} \sum_{k_0 = 2\pi(n_0+1/2)\beta^{-1}} \left(\frac{2\pi}{L}\right)^d \sum_{\boldsymbol{k}=(2\pi/L)\boldsymbol{n}} \frac{\mathrm{e}^{\mathrm{i}\boldsymbol{k}(\boldsymbol{x}-\boldsymbol{y})}}{-\mathrm{i}k_0 + |\boldsymbol{k}|^2/2m - \mu} \\ &\quad \equiv \frac{2\pi}{\beta} \sum_{k_0 = 2\pi(n_0+1/2)\beta^{-1}} \left(\frac{2\pi}{L}\right)^d \sum_{\boldsymbol{k}=(2\pi/L)\boldsymbol{n}} \mathrm{e}^{\mathrm{i}\boldsymbol{k}(\boldsymbol{x}-\boldsymbol{y})} \hat{S}_0(\boldsymbol{k}) \end{aligned} \qquad [6]$$

where $\boldsymbol{k} = (k_0, \boldsymbol{k})$. In the limit $L, \beta \to \infty$, for large distances $S(\boldsymbol{x}, \boldsymbol{y})$ decays as a power law, $O(|\boldsymbol{x} - \boldsymbol{y}|^{-1})$ times an oscillating function of period $k_{\mathrm{F}}^{-1}$. Note that $S_0(\boldsymbol{k})$ in the limit $\beta, L \to \infty$ diverges for $k_0 = 0$ and $\varepsilon(\boldsymbol{k}) = \mu$, that is, at the Fermi surface ($\mu = \varepsilon_{\mathrm{F}}$ in the limit $\beta \to \infty$); when $\beta$ is finite, $S_0(\boldsymbol{k})$ is finite even for $L \to \infty$, that is, the finite temperature acts as an infrared cutoff.

## Fermions in an External Potential

The next step consists in adding an external periodic potential to the free Fermi gas model, taking into account the field generated by the ions of the lattice. We consider then [3] with $\lambda = 0$ and $u \ne 0$. As in the previous case, the eigenfunctions of the $N$-particle Hamiltonian can be computed and are expressed in terms of the single-particle eigenfunctions of $-\hbar^2\partial_{\boldsymbol{x}}^2/2m + uc(\boldsymbol{x})$; they are called Bloch waves and have the form

$$\phi_{\boldsymbol{k}}(\boldsymbol{x}) = \frac{1}{\sqrt{L^{d/2}}}\mathrm{e}^{\mathrm{i}\boldsymbol{k}\boldsymbol{x}} u_{\boldsymbol{k}}(\boldsymbol{x}), \quad u_{\boldsymbol{k}}(\boldsymbol{x}) = u_{\boldsymbol{k}}(\boldsymbol{x} + \boldsymbol{R})$$

$\boldsymbol{k}$, called the crystalline momentum, is conserved modulo $\boldsymbol{G}$, the vectors of the reciprocal lattice, defined as

$$G = 2\pi\left(\frac{n_1}{a_1}, \ldots, \frac{n_d}{a_d}\right)$$

The eigenvalue $\varepsilon(\boldsymbol{k})$ of $-\hbar^2\partial_{\boldsymbol{x}}^2/2m + uc(\boldsymbol{x})$ associated with a Bloch wave $\phi_{\boldsymbol{k}}(\boldsymbol{x})$ has some peculiar properties; in the $L \to \infty$ limit, one finds that $\varepsilon(\boldsymbol{k})$ is not a continuous function (unlike the $u = 0$ case) but it has gaps, that is, first-order discontinuities. For $d = 1$, by a convergent power-series expansion in $u$, one finds that $\varepsilon(\boldsymbol{k})$ is a continuous monotonically increasing function except at the points $\pm\pi n/a$, $n$ an integer; at these points $\varepsilon(\boldsymbol{R})$ is discontinuous and $\varepsilon((n\pi/a)^+) - \varepsilon((n\pi/a)^-) \equiv \Delta_n = u\hat{c}_n + O(u^2)$; the gaps divide $\varepsilon(\boldsymbol{k})$ into disconnected pieces called energy bands. Something similar happens in $d = 2, 3$, in which gaps open for $\boldsymbol{R}$ such that $\boldsymbol{G}^2 + 2\boldsymbol{k}\boldsymbol{G} = 0$.

Again, the eigenfunctions of $H$ are given by $|\boldsymbol{k}_1, \sigma_1, \ldots, \boldsymbol{k}_N, \sigma_N\rangle$ with eigenvalue $\sum_{\boldsymbol{k},\sigma} \varepsilon(\boldsymbol{k}) n_{\boldsymbol{k},\sigma}$, and the Fermi surface is still defined by the set $\boldsymbol{k}$ such that $\varepsilon(\boldsymbol{k}) = \varepsilon_{\mathrm{F}}$ with $\varepsilon_{\mathrm{F}}$ determined by the condition $\sum_{\boldsymbol{k}:\varepsilon(\boldsymbol{k})} \leq \varepsilon_F\ 1 = N$. However, in this case the Fermi surface is not anymore a sphere in $d = 3$, but it is in general a polyhedron of a very complex shape. The Schwinger functions are expressed by the Wick rule [5] in terms of the two-point Schwinger functions; they are given by [6] with $e^{i\boldsymbol{k}(x-y)}$ replaced by $\phi_{\boldsymbol{k}}(\boldsymbol{x})\phi_{\boldsymbol{k}}^*(\boldsymbol{y})$ and $|\boldsymbol{k}|^2/2m$ replaced by $\varepsilon(\boldsymbol{k})$. The asymptotic properties of the two-point Schwinger function are quite different with respect to the $u = 0$ case. This is easy to see if $d = 1$; in the limit $L, \beta \to \infty$, $S(\boldsymbol{k})$ is singular if $\mu$ does not belong to the interval $[\varepsilon((n\pi/a)^+), \varepsilon((n\pi/a)^-)]$, whereas it is finite if $\mu$ belongs to such an interval; in the first case, $S(\boldsymbol{x}, \boldsymbol{y})$ decays for large distances as $O(|\boldsymbol{x} - \boldsymbol{y}|^{-1})$, whereas in the second case it is $O(e^{-|\Delta_n||\boldsymbol{x}-\boldsymbol{y}|})$. This means that, depending on the number of particles (which essentially fixes $\mu$), the Schwinger function has a totally different asymptotic behavior. This fact has important consequences in many physical properties; for instance, the conductivity (which can be computed from the $s = 4$ Schwinger function) vanishes if $\mu$ belongs to the interval $[\varepsilon((n\pi/a)^+), \varepsilon((n\pi/a)^-)]$. Similar properties hold for $d = 2, 3$; hence, from the knowledge of the number of particles and the periodic potential generated by the ions, one can predict if the system is an insulator or a metal.

Note also that the conductivity is infinite in the infinite-volume and zero-temperature limit, when $\mu$ does not correspond to a gap; in other words, the electric current in a perfect crystal lattice is not subjected to any dissipation of energy. A finite resistivity is found only if one takes into account deviations from perfect periodicity. To simulate impurities in the lattice, one can add, according to Anderson, to the Hamiltonian an interaction term of the form $\alpha\phi_{\boldsymbol{x}}\psi_{\boldsymbol{x}}^+\psi_{\boldsymbol{x}}^-$, where $\phi_{\boldsymbol{x}}$ is a Gaussian stochastic field. A detailed mathematical investigation has been devoted to the properties of eigenfunctions of $-\hbar^2\partial_{\boldsymbol{x}}^2/2m + \alpha\phi_{\boldsymbol{x}}$, where $\phi_{\boldsymbol{x}}$ is a Gaussian field (see, e.g., Pastur and Figotin (1991)); it is found that if $\alpha$ is large enough in $d = 2, 3$ and for any $\alpha$ in $d = 1$, the single-particle eigenfunctions are exponentially localized, that is, they decay exponentially at large distances; this implies a finite conductivity. One can also add to the Hamiltonian a term $\beta\phi'_{\boldsymbol{x}}\psi_{\boldsymbol{x}}^+\psi_{\boldsymbol{x}}^-$, with $\phi'_{\boldsymbol{x}}$ a quasiperiodic function, in order to describe crystals in which the lattice develops a periodic distortion, with incommensurate period with respect to the lattice periodicity. For $d = 1$ and $\beta$ large, one again finds localized eigenfunctions, whereas for small $\beta$ there are extended states (see, e.g., Pastur and Figotin (1991)); such results are obtained with the Kolgomorov–Arnol'd–Moser (KAM) techniques.

## Interacting Systems

The analysis of noninteracting Fermi systems has been very successful in understanding qualitatively many features of crystals, but there are many properties (e.g., superconductivity or magnetism) which cannot be really explained without taking into account the interaction between fermions; however, the analysis becomes more involved. When there is no interaction, the properties of the many-body system can be understood in terms of the single-body properties; the eigenfunctions of the Hamiltonian are, in fact, obtained in terms of the single-particle eigenfunctions. This is not true when $\lambda \neq 0$ when a description of the system in terms of independent particles is impossible. In order to compute the interacting Schwinger functions, it is convenient to write them in terms of fermionic functional integrals (Berezin 1966). One introduces a set of anticommuting Grassmann variables $\psi_{\boldsymbol{k}}^+, \psi_{\boldsymbol{k}}^-$, $\boldsymbol{k} = (k_0, \boldsymbol{k})$; the Grassmann integration is defined by $\int d\psi_{\boldsymbol{k}}^\sigma \psi_{\boldsymbol{k}}^\sigma = 1$ and $\int d\psi_{\boldsymbol{k}}^\sigma = 0$, $\sigma = \pm$, and the integral of any analytic function of the Grassmann variables can be obtained by expanding it in Taylor series (which is a finite sum if suitable cutoffs are imposed and $L, \beta$ are finite) and using the above rules; finally,

$$\psi_{\boldsymbol{x}}^+ = \frac{1}{L^d\beta} \sum_{\boldsymbol{k}} \phi_{\boldsymbol{k}}(\boldsymbol{x}) e^{ik_0 t} \psi_{\boldsymbol{x}}^+$$

and $\psi_{\boldsymbol{x}}^-$ is defined in an analogous way. The Schwinger function can be written as a Grassmann integral as follows:

$$S(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N) = \frac{\partial^N}{\partial\phi_{\boldsymbol{x}_1}^{\varepsilon_i} \ldots \partial\phi_{\boldsymbol{x}_N}^{\varepsilon_i}} \log \int P(d\psi) e^{-\nu + \int d\boldsymbol{x} \phi_{\boldsymbol{x}}^\varepsilon \psi_{\boldsymbol{x}}^\varepsilon}\Big|_{\varphi=0} \qquad [7]$$

where $P(\mathrm{d}\psi)$ is the fermionic integration $[\prod_{\boldsymbol{k}} \mathrm{d}\psi^+_{\boldsymbol{k}} d\psi^-_{\boldsymbol{k}}] \exp[\sum_{\boldsymbol{k}} \psi^+_{\boldsymbol{k}}[-\mathrm{i}k_0 + \varepsilon(\boldsymbol{k}) - \mu]\psi^-_{\boldsymbol{k}}]$, while $\boldsymbol{y} = (\boldsymbol{y}, s)$, and

$$\nu = \lambda \sum_{\sigma,\sigma'} \int \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{y}\, v(\boldsymbol{x} - \boldsymbol{y}) \times \delta(t-s)\psi^+_{\boldsymbol{x},\sigma}\psi^-_{\boldsymbol{x},\sigma}\psi^+_{\boldsymbol{y},\sigma'}\psi^-_{\boldsymbol{y},\sigma'} \qquad [8]$$

The Grassmann integral of a monomial of Grassmann variables can be obtained by the Wick rule [5] with propagator the Fourier transform of $(-\mathrm{i}k_0 + \varepsilon(\boldsymbol{k}) - \mu)^{-1}$. As stated earlier, the propagator is finite at nonzero temperature, whereas if $\beta = \infty$, then it is singular when $\boldsymbol{k} = (k_0, \boldsymbol{k})$ is such that $k_0 = 0$ and $\varepsilon(\boldsymbol{k}) = \mu$.

One can write [7] as a series by Taylor-expanding the exponential and using the Wick rule; each order of the expansion can be represented as a sum of Feynman diagrams, very similar to the ones appearing in quantum field theory. We have then an algorithm to compute [7]; nevertheless, to extract information from such a series is quite difficult. One cannot really compute an infinite (in the $L = \infty$ limit) number of coefficients, so one is tempted, for small $\lambda$, to compute only the first few of them, neglecting the others. However, it appears that this approximation is generally not justified, and it leads to wrong results; the reason is that the Schwinger functions for $\lambda = 0$ or $\lambda \neq 0$ are not analytically close, or, in more physical terms, even if $\lambda$ is small, the physical behavior of the free and interacting theories can be quite different, especially at low temperatures. A number of very interesting concepts (e.g., spontaneous symmetry breaking or the mass generation phenomenon), or techniques (e.g., the renormalization group method, or the parquet or random phase approximation) have been introduced in the last 50 years to analyze [7], and indeed many results have been obtained which explain several physical properties of the matter, such as superconductivity or the Kondo effect (see, e.g., Anderson 1985, Abrikosov *et al.* 1965, Mahan 1990, Negele and Orland 1988, Pines 1961). Unfortunately, most of such results are not really mathematically consistent, and in many cases quantitative computations are impossible (in computations one generally neglects terms which, according to a heuristic physical intuition, are irrelevant, but no control of the error introduced by this approximation is attempted). In recent times, attempts towards a mathematical understanding of the functional integral [7] have started (see, e.g., Benfatto and Gallavotti (1995), and references therein); the methods rely on the mathematical implementation of Wilson's renormalization group methods via multiscale analysis (Gallavotti 1985). The necessity of a firmer mathematical basis was felt mainly under the pressure of the recent discovery of high-$T_\mathrm{c}$ superconductors whose behavior is still not understood in terms of the microsopic model [7]; this has forced reconsideration of the validity of the approximations usually made in the analysis of this model.

The behavior of [7] depends crucially on the temperature. At high temperatures, we can simply expand the exponential in [7] in a power series of $\lambda$, and find that each Feynman graph contributing to the $n$th perturbative order is bounded by $C^n_\beta|\lambda|^n$, with $C_\beta \leq C\beta^\gamma$ for some constants $C, \gamma$; this follows immediately by using the Wick rule and by remembering that the propagator is larger than $O(\beta^{-1})$. As the number of Feynman graphs contributing to order $n$ is $O(n!)$, a bound on each Feynman graph is not sufficient to prove the convergence of the series. To prove convergence, one has to take into account cancellations, due to the anticommutativity of fermionic variables. Such cancellations are proved via Gram's inequality for determinants and a bound $C^n_\beta\lambda^n$ can be obtained for the order $n$ (without factorials); hence, convergence follows for temperatures greater than $O(|\lambda|^\alpha)$ for some constant $\alpha > 0$. One finds that $S(\boldsymbol{k}) = S_0(\boldsymbol{k})(1 + A_\lambda(\boldsymbol{k}))$ with $|A_\lambda(\boldsymbol{k})| \leq C|\lambda|$, that is, the interaction has essentially no influence on the physical properties of the system at high temperatures.

## Landau Fermi Liquids

We consider next an intermediate region of temperatures, that is, $\mathrm{e}^{-a/|\lambda|} \leq T \leq |\lambda|^\alpha$ for some constants $a, \alpha$. In this region, the naive expansion in power series of $\lambda$ fails and other techniques, such as renormalization group, are necessary. Such a method allows us to perform a suitable resummation of the naive power series in $\lambda$, and one gets, for $\lambda$ small enough, $T \geq \mathrm{e}^{-a/|\lambda|}$ and $\varepsilon(\boldsymbol{k}) = |\boldsymbol{k}|^2/2m$,

$$\hat{S}(\boldsymbol{k}) = \frac{1}{Z(\lambda)} \frac{1 + A_\lambda(\boldsymbol{k})}{-\mathrm{i}k_0 + v_\mathrm{F}(\lambda)[|\boldsymbol{k}| - k_\mathrm{F}(\lambda)]} \qquad [9]$$

where $Z(\lambda) = 1 + z(\lambda), v_\mathrm{F}(\lambda) = \hbar k_\mathrm{F}/m + v(\lambda)$, and $k_\mathrm{F}(\lambda) = k_\mathrm{F} + \nu(\lambda)$, with $z(\lambda) = O(\lambda^2), \nu(\lambda) = O(\lambda)$, $v(\lambda) = O(\lambda^2)$, and $z(\lambda), \nu(\lambda), v_\mathrm{F}(\lambda)$ essentially temperature independent; moreover, $|A_\lambda(\boldsymbol{k})|$ is $O(\lambda)$. The above formula has been proved rigorously for $d = 2$ (see Rivasseou (1994), and references therein); for $d = 3$, it has been proved at the level of formal perturbation theory (Benfatto and Gallavotti 1995). The case $\varepsilon(\boldsymbol{k}) = |\boldsymbol{k}|^2/2m$ is quite special, as the shape

of the interacting Fermi surface is fixed by the rotation-invariant symmetry; it is necessarily circular ($d = 2$) or spherical ($d = 3$), whereas in general the interaction can also modify its shape. For $d = 2$, if the interacting Fermi surface is symmetric, smooth and convex, a formula like [9] still holds (with a function $k_{\rm F}(\lambda, \boldsymbol{k})$ replacing $k_{\rm F}(\lambda)$) up to exponentially small temperatures (see references in Gentile and Mastropietro (2001)).

It is apparent from [9] that one cannot derive such a formula from a power-series expansion in $\lambda$; by expanding [9] as a series in $\lambda$, one immediately finds that the $n$th term is $O(\lambda^n \beta^n)$, which means that the naive perturbative expansion cannot be convergent up to exponentially small temperatures. It can be derived only by selecting and resumming some special class of terms in the original expansion. A peculiar property of [9] is that the wave function renormalization $Z(\lambda)$ is essentially independent of the temperature. Such temperature independence is a consequence of cancellations in the perturbative series essentially due to the curvature of the Fermi surface. For $d = 1$, a formula similar to [9] is also valid; however, such cancellations are not present and one finds $Z(\lambda) = 1 + O(\lambda^2 \log \beta)$. Comparing $S(\boldsymbol{k})$ given by [9] with the Fourier transform $S_0(\boldsymbol{k})$ of [6], we note that the Schwinger function of the interacting system is still very similar to the Schwinger function of a free Fermi gas, with physical parameters (e.g., the Fermi momentum, the wave function renormalization, or the Fermi velocity) which are changed by the interaction. This property is quite remarkable: the eigenstates cannot be constructed when $\lambda = 0$ starting from the single-particle states but, nevertheless, the physical properties of the interacting system (which can be deduced from the Schwinger functions) are qualitatively very similar to the ones of the free Fermi gas, although with different parameters; this explains why the free Fermi gas model works so well to explain the properties of crystals, although one neglects the interactions between fermions which are, of course, quite relevant. A fermionic system with such a property is called a Landau Fermi liquid (see, e.g., Arbikosov *et al.* 1965, Mahan 1990, Pines 1961), after Landau, who postulated in the 1950s that interacting systems may evolve continuously from the free system in many cases.

It was generally accepted that metals in this range of temperatures were all Landau Fermi liquids (except one-dimensional systems). However, the experimental discovery of the high-$T_{\rm c}$ superconductors (see, e.g., Anderson (1997)) has changed this belief, as such metals in their normal state, that is, above $T_{\rm c}$ are not Landau Fermi liquids; their wave function renormalization behaves like $1 + O(\lambda^2 \log \beta)$ instead of $1 + O(\lambda^2)$ as in Landau Fermi liquid. This behavior has been called marginal-Fermi-liquid behavior and many attempts have been devoted to predict such behavior from [7]. In order to see deviations from Fermi liquid behavior, one could consider Fermi surfaces with flat or almost flat sides or corners (which are quite possible; e.g., in a square lattice with one conduction electron per atom, such as in the "half-filled Hubbard model").

Let us finally consider the last regime, that is, temperatures lower than $O(\mathrm{e}^{-a/|\lambda|})$. Except for very exceptional cases (e.g., asymmetric Fermi surfaces, i.e., such that $\varepsilon(\boldsymbol{k}) \neq \varepsilon(-\boldsymbol{k})$ except for a finite number of points, in which Fermi liquid behavior is found down to $T = 0$ (Feldman *et al.* 2002)), a strong deviation from Fermi liquid behavior is observed; the interacting Schwinger function is not similar to the free one and the physical properties in this regime are totally new.

## One-Dimensional Systems up to $T = 0$

The only case in which the Schwinger functions of the Hamiltonian [3] can be really computed down to $T = 0$ occurs for $d = 1$; in such a case, an expression like [9] is not valid anymore and the system is not a Fermi liquid. On the contrary, when $u = 0$ and for small repulsive $\lambda > 0$, one can prove, for spinning fermions (see Benfatto and Gallavotti (1995), Gentile and Mastropietro (2001) and references therein) that

$$\hat{S}(\boldsymbol{k}) = \frac{\left[k_0^2 + v_{\rm F}^2(\lambda)(|\boldsymbol{k}| - k_{\rm F}(\lambda))^2\right]^{\eta(\lambda)}}{-\mathrm{i}k_0 + v_{\rm F}(\lambda)[|\boldsymbol{k}| - k_{\rm F}(\lambda)]}[1 + A_\lambda(\boldsymbol{k})] \qquad [10]$$

where $k_{\rm F}(\lambda) = k_{\rm F} + O(\lambda)$ and $\eta(\lambda) = a\lambda^2 + O(\lambda^3)$ is a critical index. This means that the interaction changes qualitatively the nature of the singularity at the Fermi surface; $S(\boldsymbol{k})$ is still diverging at the Fermi surface but with an exponent which is no longer 1 but is $1 - 2\eta(\lambda)$, with $\eta(\lambda)$ a nonuniversal (i.e., $\lambda$-dependent) critical index. As a consequence, the physical properties are different with respect to the free Fermi gas; for instance, the occupation number $n_{\boldsymbol{k}}$ is not discontinuous at $k = \pm k_{\rm F}(\lambda)$ when $T = 0$. Nonuniversal critical indices appear in all the other response functions. Fermionic systems behaving in this way are called Luttinger liquids, as they behave like the exactly solvable Luttinger model describing relativistic spinless fermions with linear dispersion relation. The solvability of this model, due to Mattis and Lieb (1966), relies

on the possibility of mapping its Hamiltonian in a system of free bosons. Such a mapping is not possible for the Hamiltonian [3], which is not solvable; however, one can use renormalization group methods and suitable Ward identities to show that its behavior is similar to the Luttinger model (in a sense, one makes perturbation theory not around the free Fermi gas, but around the Luttinger model).

If we take into account the interaction with an external periodic potential with period $a$, that is, consider $u \neq 0$, we find that if $k_F \neq n\pi/a$, then the Schwinger function behaves essentially like [8]. On the contrary, in the filled-band case, $k_F = n\pi/a$, one finds that there is still an energy gap which becomes $O(u^{1+\eta_1})$ with $\eta_1 = O(\lambda)$; this means that the renormalization of the gap is described by a critical index; moreover, $S(\mathbf{x}) \simeq O(e^{-|u|^{1+\eta_1}|\mathbf{x}|})$. A similar behavior is also observed in the presence of quasi-periodic potential. In the attractive case, $\lambda < 0$, $u = 0$, the behavior is much less understood; it is believed that the interaction produces a gap $\Delta_\lambda$ in the spectrum which is nonanalytic in $\lambda$, and $S(\mathbf{x})$ shows an exponential decay rather than a power-law decay, and the interaction converts the system from a metal to an insulator.

Finally, it is remarkable that a large variety of models, like Heisenberg spin chains or bidimensional classical statistical mechanics models, such as the eight-vertex or the Ashkin–Teller model, can be mapped into interacting $d = 1$ fermionic systems, and consequently their critical behavior can be understood by using fermionic techniques (see Gentile and Mastropietro (2001), and references therein).

## Superconductors

The theory up to $T = 0$ for $d = 2, 3$ systems with dispersion relation $|\mathbf{k}|^2/2m$ is based only on approximate computations, predicting the phenomenon of superconductivity. According to the theory of Bardeen, Cooper, and Schrieffer (BCS theory), the interaction between fermions leads to the formation of a gap in the energy spectrum, below the critical temperature. There are many ways to derive the BCS theory. One is based on the fact that one verifies, by perturbative computations, that the effective interaction is stronger when the four momenta of the fermions are such that $\mathbf{k}_1 \simeq -\mathbf{k}_3$ and $\mathbf{k}_2 \simeq -\mathbf{k}_4$. This suggests, heuristically, to replace in [7] $\nu$ with

$$\nu_{BCS} = -\lambda \frac{1}{\beta L^{3d}} \sum_{\mathbf{k},\mathbf{k}'} \psi^+_{\mathbf{k},\sigma} \psi^+_{-\mathbf{k},-\sigma} \psi^-_{\mathbf{k}',-\sigma'} \psi^-_{-\mathbf{k}',\sigma'}$$

which is an interaction between pairs of electrons with opposite spin and momenta, which are called Cooper pairs. Replacing $\nu$ with $\nu_{BCS}$ has the great advantage that it makes the Schwinger functions exactly computable and explains the mechanism of superconductivity in many metals (but not in the recently discovered high-$T_c$ superconductors). On the other hand, proving that [7] with $\nu$ or $\nu_{BCS}$ has a similar behavior is still an important open problem. The two-point Schwinger function in the model with $\nu_{BCS}$ can be written, after the so-called Hubbard–Stratonovitch transformation, as

$$\hat{S}(k_0, \mathbf{k}) = -\beta L^d \frac{\int \frac{-ik_0 - \varepsilon(\mathbf{k}) + \mu}{k_0^2 + \varepsilon^2(\mathbf{k}) + \lambda u^2} e^{-\beta L^d \nu(u)} \, du}{\int e^{-\beta L^d \nu(u)} \, du} \qquad [11]$$

where $\nu(u)$ is a function with a global minimum in $u = 0$ for repulsive interactions $\lambda < 0$, whereas for $\lambda > 0$ and sufficiently small temperatures (for $T \leq T_c$, with $T_c = O(e^{-a/|\lambda|})$), it has the form of a double well with two minima at $u = \pm \Delta_\lambda$ with $\Delta_\lambda = O(e^{-a/|\lambda|})$; for $T$ greater than $T_c$, there is only a global minimum at $u = 0$. By the saddle-point theorem, we find, for $T \leq T_c$ and $\lambda < 0$,

$$\lim_{L \to \infty} S(\mathbf{k}) = \frac{-ik_0 - \varepsilon(\mathbf{k}) + \mu}{k_0^2 + (\varepsilon(\mathbf{k}) - \mu)^2 + \Delta_\lambda^2} \qquad [12]$$

The physical properties predicted by [12] are completely different with respect to the free case: the occupation number is continuous, there is an energy gap in the spectrum, the specific heat is $O(e^{-\Delta_\lambda T})$ and the phenomenon of superconductivity appears. The fact that the interaction generates a gap is called mass generation; a similar mechanism appears in particle theory.

## Conclusions

Many other physical phenomena, observed experimentally, can be essentially understood by studying fermionic systems, but a clear mathematical comprehension is still lacking. We mention: the Kondo effect, that is, the resistance minimum observed in some metals due to magnetic impurities; Mott transition, in which a strong interaction produces an insulating state in a system which should be conductors; antiferromagnetism; fractional quantum Hall effect, and many others. We can say that the situation in this area of study reminds one of the classical mechanics at the end of the nineteenth century; there is agreement on the models to consider, which are believed to be able to take into

account the marvelous properties of the matter experimentally found, but to extract information from them requires deeper and complex analytical and mathematical investigations.

*See also:* Falicov–Kimball Model; Fractional Quantum Hall Effect; Quantum Statistical Mechanics: Overview; Renormalization: Statistical Mechanics and Condensed Matter.

## Further Reading


Abrikosov AA, Gorkov LP, and Dzyaloshinskii IE (1965) *Methods of Quantum Field Theory in Statistical Physics*. New York: Dover.

Anderson PW (1985) *Basic Principles in Solid State Physics*. Menlo Park, CA: Benjamin Cummings.

Anderson PW (1997) *The Theory of Superconductivity in High $T_c$ Cuprates*. Princeton, NJ: Princeton University Press.

Bardeen J, Cooper LN, and Schrieffer JR (1957) Theory of superconductivity. *Physical Review* 108: 1175–1204.

Benfatto G and Gallavotti G (1995) *Renormalization Group*. Princeton: Princeton University Press.

Berezin FA (1966) *The Method of Second Quantization*. New York: Academic Press.

Bratelli O and Robinson D (1979) *Operator Algebras and Quantum Statistical Mechanics*. Berlin: Springer.

Feldman J, Knorrer H, and Trubowitz E (2002) *Fermionic Functional Integrals and Renormalization Group*, CRM Monograph Series, vol. 16. Providence: American Mathematical Society.

Gallavotti G (1985) Renormalization group and ultraviolet stability for scalar fields via renormalziation group methods. *Reviews of Modern Physics* 57: 471–562.

Gentile G and Mastropietro V (2001) Renormalization group for one dimensional fermions. A review of mathematical results. *Physics Reports* 352: 273–437.

Mahan GD (1990) *Many-Particle Physics*. New York: Plenum.

Mattis DC and Lieb E (1966) *Mathematical Physics in One Dimension*. New York: Academic Press.

Negele JW and Orland H (1988) *Quantum Many Particle Systems*. New York: Addison-Wesley.

Pastur L and Figotin A (1991) *Spectra of Random and Almost Periodic Operators*. Berlin: Springer.

Pines D (1961) *The Many Body Problem*. New York: Addison-Wesley.

Rivasseou V (1994) *From Perturbative to Constructive Renormalization*. Princeton, NJ: Princeton University Press.


# Feynman Path Integrals

**S Mazzucchi**, Università di Trento, Povo, Italy



## Introduction

In nonrelativistic quantum mechanics, the state of a $d$-dimensional particle is represented by a unitary vector $\psi$ in the complex separable Hilbert space $L^2(\mathbb{R}^d)$, the so-called "wave function," while its time evolution is described by the Schrödinger equation:

$$
\begin{aligned}
i\hbar \frac{\partial}{\partial t}\psi &= -\frac{\hbar^2}{2m}\Delta\psi + V\psi \\
\psi(0,x) &= \psi_0(x)
\end{aligned}
\qquad [1]
$$

where $\hbar$ is the reduced Planck constant, $m > 0$ is the mass of the particle, and $F = -\nabla V$ is an external force.

In 1942 R P Feynman, following a suggestion by Dirac, proposed an alternative (Lagrangian) formulation of quantum mechanics, and a heuristic but very suggestive representation for the solution of eqn [1]. According to Feynman, the wave function of the system at time $t$ evaluated at the point $x \in \mathbb{R}^d$ is given as an "integral over histories," or as an integral over all possible paths $\gamma$ in the configuration space of the system with finite energy passing at the point $x$ at time $t$:

$$
\psi(t,x) = \left( \int_{\{\gamma|\gamma(t)=x\}} e^{(i/\hbar)S_t^\circ(\gamma)} D\gamma \right)^{-1} \times \int_{\{\gamma|\gamma(t)=x\}} e^{(i/\hbar)S_t(\gamma)} \psi_0(\gamma(0)) D\gamma \qquad [2]
$$

$S_t(\gamma)$ is the classical action of the system evaluated along the path $\gamma$

$$
S_t(\gamma) \equiv S_t^\circ(\gamma) - \int_0^t V(\gamma(s))\,ds \qquad [3]
$$

$$
S_t^\circ(\gamma) \equiv \frac{m}{2}\int_0^t |\dot\gamma(s)|^2\,ds \qquad [4]
$$

$D\gamma$ is a heuristic Lebesgue "flat" measure on the space of paths and $(\int_{\{\gamma|\gamma(t)=x\}} e^{(i/\hbar)} S_t^\circ(\gamma) D\gamma)^{-1}$ is a normalization constant.

Some time later, Feynman himself extended formula [2] to more general quantum systems, including the case of quantum fields.

The Feynman path-integral formulation of quantum mechanics is particularly suggestive, as it provides a spacetime visualization of quantum dynamics, reintroducing in quantum mechanics the concept of trajectory (which was banned in the "orthodox interpretation" of the theory) and creating a connection between the classical description of

the physical world and the quantum one. Indeed, it provides a quantization method, allowing, at least heuristically, to associate a quantum evolution to each classical Lagrangian. Moreover, the application of the stationary-phase method for oscillatory integrals allows the study of the semiclassical limit of the Schrödinger equation, that is, the study of the detailed behavior of the solution when the Planck constant is regarded as a parameter converging to 0. Indeed, when $\hbar$ is small, the integrand in [2] is strongly oscillating and the main contributions to the integral should come from those paths $\gamma$ that make stationary the phase function $S(\gamma)$. These, by Hamilton's least action principle, are exactly the classical orbits of the system.

Feynman path integrals allow also a heuristic calculus in path space, leading to variational calculations of quantities of physical and mathematical interest. An interesting application can be found in topological field theories, as, for instance, Chern–Simons models. In this case, heuristic calculations based on the Feynman path-integral formulation of the theory, where the integration is performed on a space of geometrical objects, lead to the computation of topological invariants.

Even if from a physical point of view, formula [2] is a source of important results, from a mathematical point of view, it lacks rigor: indeed, neither the "infinite-dimensional Lebesgue measure," nor the normalization constant in front of the integral is well defined. In this article, we shall describe the main approaches to the rigorous mathematical realization of Feynman path integrals, as well as their most important applications.

## Possible Mathematical Definitions of Feynman's Measure

In the rigorous mathematical definition of Feynman's complex measure

$$\mu_F := \left( \int_{\{\gamma | \gamma(t)=x\}} \mathrm{e}^{(\mathrm{i}/\hbar)S_t^\circ(\gamma)} D\gamma \right)^{-1} \mathrm{e}^{(\mathrm{i}/\hbar)S_t^\circ(\gamma)} D\gamma \qquad [5]$$

one has to face mainly two problems. First of all, the integral is defined on a space of paths, that is, on an infinite-dimensional space. The implementation of an integration theory is nontrivial: for instance, it is well known that a Lebesgue-type measure cannot be defined on infinite-dimensional Hilbert spaces. Indeed, the assumption of the existence of a $\sigma$-additive measure $\mu$ which is invariant under rotations and translations and assigns a positive finite measure to all bounded open sets leads to a contradiction. In fact, by taking an orthonormal system $\{e_i\}_{i\in\mathbb{N}}$ in an infinite-dimensional Hilbert space $\mathcal{H}$ and by considering the open balls $B_i = \{x \in \mathcal{H}, \|x - e_i\| < 1/2\}$, one has that they are pairwise disjoint and their union is contained in the open ball $B(0,2) = \{x \in \mathcal{H}, \|x\| < 2\}$. By the Euclidean invariance of the Lebesgue-type measure $\mu$, one can deduce that $\mu(B_i) = a, 0 < a < \infty$, for all $i \in \mathbb{N}$. By the $\sigma$-additivity, one has

$$\mu(B(0,2)) \geq \mu(\cup_i B_i) = \sum_i \mu(B_i) = \infty$$

but, on the other hand, $\mu(B(0,2))$ should be finite as $B(0,2)$ is bounded. As a consequence, we can also deduce that the term $D\gamma$ in [2] does not make sense.

The second problem is the fact that the exponent in the density $\mathrm{e}^{(\mathrm{i}/\hbar)S_t^\circ(\gamma)}$ is imaginary, so that the exponential oscillates. Even in finite dimensions, integrals of the form $\int_{\mathbb{R}^N} \mathrm{e}^{\mathrm{i}\Phi(x)} f(x)\,\mathrm{d}x$, with $\Phi, f : \mathbb{R}^N \to \mathbb{R}$ are continuous functions and $f$ is not summable, have to be suitably defined, in order to exploit the cancelations in the integral due to the oscillatory behavior of the exponential.

The study of the rigorous foundation of Feynman path integrals began in the 1960s, when Cameron proved that Feynman's heuristic complex measure [5] cannot be realized as a complex bounded variation $\sigma$-additive measure, even on very nice subsets of the space $(\mathbb{R}^d)^{[0,t]}$ of paths, contrary to the case of complex measures on $\mathbb{R}^n$ of the form $\mathrm{e}^{(i/2)|x|^2}\,\mathrm{d}x$. In other words, it is not possible to implement an integration theory in the traditional (Lebesgue) sense. As a consequence, mathematicians tried to realize [5] as a linear continuous functional on a sufficiently rich Banach algebra of functions, inspired by the fact that a bounded measure can be regarded as a continuous functional on the space of bounded continuous functions. In order to mirror the features of the heuristic Feynman's measure, such a functional should have some properties:

1. it should behave in a simple way under "translations and rotations in path space," as $D\gamma$ denotes a "flat" measure;
2. it should satisfy a Fubini-type theorem, concerning iterated integrations in path space (allowing the construction, in physical applications, of a one-parameter group of unitary operators);
3. it should be approximable by finite-dimensional oscillatory integrals, allowing a sequential approach in the spirit of Feynman's original work; and
4. it should be sufficiently flexible to allow a rigorous mathematical implementation of an infinite-dimensional version of the stationary-phase

method and the corresponding study of the semiclassical limit of quantum mechanics.

Nowadays, several implementations of this program can be found in the literature of physics and mathematics, for instance, by means of analytic continuation of Wiener integrals, or as an infinite-dimensional distribution in the framework of Hida calculus, or via "complex Poisson measures," or via nonstandard analysis, or as an infinite-dimensional oscillatory integral. The last of these methods is particularly interesting as it allows the systematic implementation of an infinite-dimensional version of the stationary-phase method, which can be applied to the study of the semiclassical limit of the solution of the Schrödinger equation [1].

## Analytic Continuation

In one of the first approaches in the definition of Feynman path integrals, formula [2] was realized as the analytic continuation in a suitable complex parameter of a (nonoscillatory) Gaussian integral on the space of paths.

In 1949, inspired by Feynman's work, M Kac observed that by considering the heat equation

$$-\frac{\partial}{\partial t}u = -\frac{1}{2m}\Delta u + V(x)u \qquad u(0,x) = \psi_0(x) \tag{6}$$

instead of the Schrödinger equation [1] and by replacing the oscillatory term $e^{(i/\hbar)S_0(\gamma)}$ in Feynman complex measure with the fast decreasing one $e^{-(1/\hbar)S_0(\gamma)}$, it is possible to give a well-defined mathematical meaning to Feynman's heuristic formula [2] in terms of a well-defined integral on the space of continuous paths $W_{t,x} = \{w \in C(0,t;\mathbb{R}^d) : w(0) = x\}$ with respect to the Wiener Gaussian measure $P_{t,x}$:

$$u(t,x) = \int_{W_{t,x}} e^{-\int_0^t V(\sqrt{1/m}\,w(\tau))\mathrm{d}\tau} \times \psi_0(\sqrt{1/m}\,w(t))\,\mathrm{d}P_{t,x}(w) \tag{7}$$

The path-integral representation [7] for the solution of the heat equation [6] is called Feynman–Kac formula.

The underlying idea of the analytic continuation approach comes from the fact that by introducing in [6] a suitable parameter $\lambda$, proportional, for instance, to the time $t$ as in the case $\lambda = \lambda_1$,

$$-\lambda_1\hbar\frac{\partial}{\partial t}u = -\frac{1}{2m}\hbar^2\Delta u + V(x)u$$

$$u(t,x) = \int_{W_{t,x}} e^{-(1/\lambda_1\hbar)\int_0^t V(\sqrt{\hbar/(m\lambda_1)}\,w(\tau))\,\mathrm{d}\tau} \times \psi_0\left(\sqrt{\hbar/(m\lambda_1)}\,w(t)\right)\mathrm{d}P_{t,x}(w)$$

or to the Planck constant, as in the case $\lambda = \lambda_2$,

$$\lambda_2\frac{\partial}{\partial t}u = \frac{1}{2m}\lambda_2^2\Delta u + V(x)u$$

$$u(t,x) = \int_{W_{t,x}} e^{(1/\lambda_2)\int_0^t V(\sqrt{\lambda_2/m}\,w(\tau))\,\mathrm{d}\tau} \times \psi_0(\sqrt{\lambda_2/m}\,w(t))\,\mathrm{d}P_{t,x}(w)$$

or to the mass, as in the case $\lambda = \lambda_3$,

$$\frac{\partial}{\partial t}u = \frac{1}{2\lambda_3}\Delta u - iV(x)u$$

$$u(t,x) = \int_{W_{t,x}} e^{-i\int_0^t V\left(\sqrt{1/\lambda_3}\,w(\tau)\right)\mathrm{d}\tau} \times \psi_0(\sqrt{1/\lambda_3}\,w(t))\,\mathrm{d}P_{t,x}(w)$$

and by allowing $\lambda$ to assume complex values, then one gets, at least heuristically, Schrödinger equation and its solution by substituting, respectively, $\lambda_1 = -i$, $\lambda_2 = i\hbar$, or $\lambda_3 = -im$. These procedures can be made completely rigorous under suitable conditions on the potential $V$ and initial datum $\psi_0$.

## The Approach via Fourier Transform

This approach has its roots in a couple of papers by K Ito in the 1960s and was extensively developed by S Albeverio and R Høegh-Krohn in the 1970s. The main idea is the definition of oscillatory integrals with quadratic phase function on a real separable Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$, the Fresnel integrals,

$$\widetilde{\int_{\mathcal{H}}} e^{(i/2\hbar)\|x\|^2} f(x)\,\mathrm{d}x \tag{8}$$

as the distributional pairing between $e^{(i/2\hbar)\|x\|^2}$ and a complex-valued function $f$ belonging to the space $\mathcal{F}(\mathcal{H})$ of functions that are Fourier transforms of complex bounded variation measures on $\mathcal{H}$, that is,

$$f = \hat{\mu}_f, \qquad f(x) = \int_{\mathcal{H}} e^{i\langle x,y\rangle}\,\mathrm{d}\mu_f(y)$$

$\mathcal{F}(\mathcal{H})$ is a Banach algebra, where the product is the pointwise one and the identity is the function $f(x) = 1\ \forall x \in \mathcal{H}$. The norm of an element $f$ is the total variation of the corresponding measure $\mu_f$, that is, $\|\mu_f\| = \sup\sum_i |\mu_f(E_i)|$, where the supremum is taken over all sequences $\{E_i\}$ of pairwise-disjoint Borel subsets of $\mathcal{H}$, such that $\cup_i E_i = \mathcal{H}$.

Given a function $f \in \mathcal{F}(\mathcal{H})$, $f = \hat{\mu}_f$, its Fresnel integral is defined by the Parseval formula:

$$\widetilde{\int_{\mathcal{H}}} e^{(i/2\hbar)\|x\|^2} f(x)\,\mathrm{d}x := \int_{\mathcal{H}} e^{-(i\hbar/2)\|x\|^2}\,\mathrm{d}\mu_f(x) \tag{9}$$

where the right-hand side is a well-defined absolutely convergent integral with respect to a $\sigma$-additive measure on $\mathcal{H}$.

It is important to recall that this approach provides the implementation of a method of stationary phase for the expansion of the integral in powers of the small parameter $\hbar$ occurring in the integrand. We postpone the discussion of these results, as well as the application to the solution of the Schrödinger equation, to the next section where a generalization of the present approach is described.

### Infinite-Dimensional Oscillatory Integrals

The main idea of this approach is the extension of the definition of oscillatory integrals with quadratic phase function [8] to infinite-dimensional Hilbert spaces by means of a twofold limiting procedure.

The study of integrals of the form

$$I(\hbar) := \int_{\mathbb{R}^N} e^{(i/\hbar)\Phi(x)} f(x)\, dx \qquad [10]$$

where $\Phi(x): \mathbb{R}^N \to \mathbb{R}$ is the phase function and $f: \mathbb{R}^N \to \mathbb{C}$ a complex-valued continuous function, is a classical topic, largely developed in connection with various problems in mathematics (such as the theory of pseudodifferential operators) and physics (such as optics). Particular effort has been devoted to the study of the detailed behavior of the above integral in the limit of "strong oscillations," that is, when $\hbar \to 0$, by means of the method of stationary phase.

Thanks to the cancellations due to the oscillatory term $e^{(i/2\hbar)\Phi(x)}$, the integral can still be defined, even if the function $f$ is not summable, as the limit of a sequence of regularized, hence absolutely convergent, integrals. According to a Hörmander's proposal, the oscillatory integral of a function $f: \mathbb{R}^N \to \mathbb{C}$ is well defined if, for each test function $\phi \in \mathcal{S}(\mathbb{R}^N)$, such that $\phi(0) = 1$, the limit

$$\lim_{\epsilon \to 0} \int_{\mathbb{R}^N} e^{(i/2\hbar)\Phi(x)} \phi(\epsilon x) f(x)\, dx$$

exists and is independent of $\phi$.

This definition has been generalized in the 1980s by D Elworthy and A Truman to the case where the underlying space $\mathbb{R}^N$ is replaced by a real separable infinite-dimensional Hilbert space $(\mathcal{H}, \langle \cdot , \cdot \rangle)$, under the assumption that the phase function is quadratic, that is, $\Phi(x) = \|x\|^2/2$. The "infinite-dimensional oscillatory integral"

$$\widetilde{\int_{\mathcal{H}}^{\circ}} e^{(i/2\hbar)\|x\|^2} f(x)\, dx$$

is defined as the limit of a sequence of finite-dimensional approximations. More precisely, a function $f: \mathcal{H} \to \mathbb{C}$ is "integrable" if, for each increasing sequence $\{P_n\}_{n \in \mathbb{N}}$ of finite-dimensional projector operators in $\mathcal{H}$ converging strongly to the identity operator as $n \to \infty$, the limit

$$\lim_{n \to \infty} \left( \int_{P_n \mathcal{H}} e^{(i/2\hbar)\|P_n x\|^2}\, dP_n x \right)^{-1} \times \int_{P_n \mathcal{H}} e^{(i/2\hbar)\|P_n x\|^2} f(P_n x)\, dP_n x \qquad [11]$$

exists and is independent of the sequence $\{P_n\}_{n \in \mathbb{N}}$. In this case, the limit is denoted by

$$\widetilde{\int_{\mathcal{H}}^{\circ}} e^{(i/2\hbar)\|x\|^2} f(x)\, dx$$

The description of the largest class of integrable functions is still an open problem, even in finite dimension, but it is possible to find some interesting subsets of it. In particular, any function belonging to $\mathcal{F}(\mathcal{H})$, the Banach algebra considered in the approach by Fourier transform, is integrable. Indeed, by assuming that the function $f$ in [11] is of the type

$$f(x) = e^{-(i/\hbar)\langle x, Lx \rangle} g(x)$$

where $L: \mathcal{H} \to \mathcal{H}$ is a linear self-adjoint trace-class operator on $\mathcal{H}$ such that $(I - L)$ is invertible and $g \in \mathcal{F}(\mathcal{H})$, that is, $g(x) = \int_{\mathcal{H}} e^{\langle x, y \rangle} d\mu_g(y)$, then it is possible to prove that $f$ is integrable in the sense of definition [11] and the corresponding infinite-dimensional oscillatory integral can be explicitly computed in terms of a well-defined integral with respect to a bounded variation measure $\mu_f$ by means of the following Parseval's type equality:

$$\widetilde{\int_{\mathcal{H}}^{\circ}} e^{(i/2\hbar)\|x\|^2} e^{-(i/\hbar)\langle x, Lx \rangle} g(x) dx = \det(I - L)^{-1/2} \int_{\mathcal{H}} e^{-(i\hbar/2)\langle x, (I-L)^{-1} x \rangle} d\mu_f(x) \qquad [12]$$

$\det(I - L)$ being the Fredholm determinant of the operator $I - L$, that is, the product of its eigenvalues, counted with their multiplicity. If $L = 0$, then we obtain eqn [9], so that we can look at the infinite-dimensional oscillatory integrals approach as a generalization of the Fourier transform approach, since it allows at least in principle to integrate a class of function larger than $\mathcal{F}(\mathcal{H})$. In fact, recently this feature has been used by S Albeverio and S Mazzucchi in the proof of a Parseval's type equality similar to [12] for infinite-dimensional oscillatory integrals with polynomially growing phase functions.

Feynman's heuristic formula [2] for the representation of the solution of the Schrödinger equation [1] can

be realized as an infinite-dimensional oscillatory integral on the Hilbert space $\mathcal{H}_t$ of absolutely continuous paths $\gamma : [0, t] \in \mathbb{R}^d$ with fixed endpoint $\gamma(t) = 0$ and finite kinetic energy $\int_0^t \dot{\gamma}^2(\tau)\mathrm{d}\tau < \infty$, endowed with the inner product $\langle \gamma_1, \gamma_2 \rangle = \int_0^t \dot{\gamma}_1(\tau)\dot{\gamma}_2(\tau)\mathrm{d}\tau$. One has to take an initial datum $\psi_0 \in L^2(\mathbb{R}^d)$ that is the Fourier transform of a complex bounded variation measure on $\mathbb{R}^d$, that is, $\psi_0(x) = \int_{\mathbb{R}_d} \mathrm{e}^{\mathrm{i}k\cdot x}\mathrm{d}\mu_0(k)$. Moreover, one has to assume that the potential $V$ in [1] is the sum of a harmonic oscillator part plus a bounded perturbation $V_1$ that is the Fourier transform of a complex bounded variation measure $\mu_v$ on $\mathbb{R}^d$:

$$V(x) = \tfrac{1}{2}x\Omega^2 x + V_1(x)$$
$$V_1(x) = \int_{\mathbb{R}_d} \mathrm{e}^{\mathrm{i}k\cdot x}\mathrm{d}\mu_v(k)$$

($\Omega^2$ being a symmetric positive $d \times d$ matrix).

In this case, it is possible to prove that the linear operator $L$ on $\mathcal{H}_t$ defined by

$$(\gamma, L\gamma) \equiv \int_0^t \gamma(\tau)\Omega^2\gamma(\tau)\mathrm{d}\tau$$

is self-adjoint and trace class, and $(I - L)$ is invertible. Moreover, by considering the function $v : \mathcal{H}_t \to \mathbb{C}$

$$v(\gamma) \equiv \int_0^t V_1(\gamma(\tau) + x)\mathrm{d}\tau + 2x\Omega^2 \int_0^t \gamma(\tau)\mathrm{d}\tau, \quad \gamma \in \mathcal{H}_t$$

it is possible to prove that the function $f : \mathcal{H}_t \to \mathbb{C}$ given by

$$f(\gamma) = \mathrm{e}^{-(\mathrm{i}/\hbar)v(\gamma)}\psi_0(\gamma(0) + x)$$

is the Fourier transform of a complex bounded variation measure $\mu_f$ on $\mathcal{H}_t$ and the infinite-dimensional Fresnel integral of the function $g(\gamma) = \mathrm{e}^{-(\mathrm{i}/2\hbar)(\gamma, L\gamma)}f(\gamma)$, that is,:

$$\int_{\gamma(t)=0} \mathrm{e}^{(\mathrm{i}/2\hbar)\int_0^t \dot{\gamma}^2(\tau)\mathrm{d}\tau}\mathrm{e}^{-(\mathrm{i}/\hbar)\int_0^t V(\gamma(\tau)+x)\mathrm{d}\tau}\psi_0(\gamma(0) + x)\mathrm{d}\gamma$$
$$= \widetilde{\int_{\mathcal{H}_t}^0} \mathrm{e}^{(\mathrm{i}/2\hbar)(\gamma,(I-L)\gamma)}\mathrm{e}^{-(\mathrm{i}/\hbar)v(\gamma)}\psi_0(\gamma(0) + x)\mathrm{d}\gamma \qquad [13]$$

is well defined and it is equal to

$$\det(I - L)^{-1/2} \int_{\mathcal{H}_t} \mathrm{e}^{-(\mathrm{i}\hbar/2)(\gamma,(I-L)^{-1}\gamma)}\,\mathrm{d}\mu_f(\gamma)$$

Moreover, it is a representation of the solution of equation [1] evaluated at $x \in \mathbb{R}^d$ at time $t$. Recently, solutions of the Schrödinger equation with quartic anharmonic potential via infinite-dimensional oscillatory integrals have been provided by S Albeverio and S Mazzucchi using a combination of Parseval formula and a new analytic method (the inclusion of such potentials had been a stumbling block for many years).

In this framework, it is possible to implement an infinite-dimensional version of the stationary-phase method and study the asymptotic behavior of the oscillatory integrals in the limit $\hbar \to 0$.

The method of stationary phase was originally proposed by Stokes, who noted that when $\hbar \to 0$ the oscillatory integral [10] is $O(\hbar^n)$ for any $n \in \mathbb{N}$, provided that there are no critical points of the phase function $\Phi$ in the support of the function $f$. As a consequence, one can deduce that the leading contribution to the integral [10] should come from a neighborhood of those points $c \in \mathbb{R}^N$, such that $\nabla\Phi(c) = 0$. More precisely, by assuming that the set $C$ of critical points is finite, that is, $C = \{c_1, \ldots, c_k\}$ and that every critical point is nondegenerate, that is, $\det D^2\Phi(c_i) \neq 0\, \forall c_i \in C$, then one has

$$I(\hbar) \sim \sum_{c_i \in C} \mathrm{e}^{(\mathrm{i}/\hbar)\Phi(c_i)}I_i^*(\hbar) \qquad [14]$$

where $I_i^* : \mathbb{R} \to \mathbb{C}$ are $C^\infty$ functions of $\mathbb{R}$, such that

$$I_i^*(0) = f(c_i)(2\pi\mathrm{i}\hbar)^{N/2}(\det D^2\Phi(c_i))^{-1/2}$$

If some critical point is degenerate, the situation is more complicated: one has to take into account the type of degeneracy and apply the theory of unfoldings of singularities.

These results can be generalized to infinite-dimensional oscillatory integrals of the form

$$I(\hbar) = \widetilde{\int_{\mathcal{H}}} \mathrm{e}^{(\mathrm{i}/2\hbar)\langle x,(I-L)x\rangle}\mathrm{e}^{-(\mathrm{i}/\hbar)v(x)}g(x)\mathrm{d}x \qquad [15]$$

with $v(x) = \int_{\mathcal{H}} \mathrm{e}^{\mathrm{i}\langle x,y\rangle}\mathrm{d}\mu(y)$, $g(x) = \int_{\mathcal{H}} \mathrm{e}^{\mathrm{i}\langle x,y\rangle}\mathrm{d}\nu(y)$, $\mu, \nu$ being complex bounded variation measures on $\mathcal{H}$ satisfying suitable assumptions and $L : \mathcal{H} \to \mathcal{H}$ is a self-adjoint and trace-class linear operator, such that $(I - L)$ is invertible. Under suitable growth condition on the moments of the measures $\mu, \nu$ and by assuming that the phase function $\Phi(x) = \langle x, (I - L)x\rangle - v(x)$ has a finite number of nondegenerate critical points $c_1, \ldots, c_s$, it is possible to prove that the integral $I(\hbar)$ in [15] is equal to

$$I(\hbar) = \sum_{k=1}^{s} \mathrm{e}^{(\mathrm{i}/\hbar)\Phi(c_k)}I_k^*(\hbar) + I_0(\hbar)$$

for some $C^\infty$ functions $I_k^*$ satisfying:

$$I_k^*(0) = [\det(I - L - D^2V(c_k))]^{-1/2}g(c_k)$$
$$k = 1, \ldots, s$$
$$I_0^{(j)}(0) = 0, \quad j = 0, 1, 2, \ldots$$

Moreover, under some additional smallness assumptions on $v$, it has been proved that the phase function $\Phi$ has a unique stationary point $c$ and as $\hbar \to 0$

$$I(\hbar) \sim e^{(i/\hbar)\Phi(c)} I^*(\hbar)$$

for some $C^\infty$ function $I^*$. Each term of the asymptotic expansion in powers of $\hbar$ of the function $I^*$ can be explicitly computed, and it is possible to prove that such an asymptotic expansion is Borel-summable and determines $I^*$ uniquely.

The application of these results to the infinite-dimensional oscillatory integral representation [13] for the solution of the Schrödinger equation allows the study of its semiclassical limit. One has to consider a potential $V$ that is the Fourier transform of a complex bounded variation measure $\mu$ on $(\mathbb{R}^d)$, such that $\int_{\mathbb{R}^d} e^{|\beta|\epsilon} d|\mu|(\beta) < \infty$ for some $\epsilon > 0$, and a particular form for the initial wave function $\psi_0(x) = e^{(i/\hbar)\phi(x)}\chi(x)$, where $\phi$ is real and $\phi, \chi \in C_0^\infty(\mathbb{R}^d)$ are independent of $\hbar$. This initial datum corresponds to an initial particle distribution $\rho_0(x) = |\chi|^2(x)$ and to a limiting value of the probability current $J_{\hbar=0} = \nabla\phi(x)\rho_0(x)/m$, giving an initial particle flux associated to the velocity field $\nabla\phi(x)/m$. One also has to assume that the Lagrange manifold $L_f \equiv (y, -\nabla f)$ intersects transversally the subset $\Lambda_V$ of the phase space made of all points $(y, p)$ such that $p$ is the momentum at $y$ of a classical particle that starts at time zero from $x$, moves under the action of $V$, and ends at $y$ at time $t$. In this case, the Feynman path integral [13] has an asymptotic expansion in powers of $\hbar$ for $\hbar \to 0$, whose leading term is the sum of the values of the function

$$\left| \det\left( \left( \frac{\partial \bar{\gamma}_k^{(j)}}{\partial y_l^{(j)}} (y^{(j)}, t) \right) \right) \right|^{-1/2} e^{-(i/2)\pi m^{(j)}} e^{-(i/\hbar)S} e^{-(i/\hbar)\phi} \chi$$

taken at the points $y^{(j)}$ such that a classical particle starting at $y^{(j)}$ at time zero with momentum $\nabla\phi(y^{(j)})$ is at $x$ at time $t$. $S$ is the classical action along this classical path $\bar{\gamma}^{(j)}$ and $m^{(j)}$ is the Maslov index of the path $\bar{\gamma}^{(j)}$, that is, $m^{(j)}$ is the number of zeros of

$$\det\left( \left( \frac{\partial \bar{\gamma}_k^{(j)}}{\partial y_l^{(j)}} (y^{(j)}, \tau) \right) \right)$$

as $\tau$ varies on the interval $(0, t)$.

### White-Noise Calculus

The leading idea of the present approach, which was originally proposed by C DeWitt-Morette and P Krée and presently realized in the framework of white-noise calculus by T Hida, L Streit, and many other authors, is the realization of the Feynman integrand $e^{(i/\hbar)S_t^\circ(\gamma)}$ as an infinite-dimensional distribution. This idea is similar to the one of the approach via Fourier transform, where the expression $(2\pi i)^{-d/2} \int_{\mathbb{R}^d} e^{(i/2)(x,x)} f(x) dx$ is realized as a distributional pairing between $e^{(i/2)(x,x)}/(2\pi i)^{d/2}$ and the function $f \in \mathcal{F}(\mathbb{R}^d)$ by means of the Parseval-type equality [9] and generalized to infinite-dimensional spaces. In white-noise calculus, the pairing is realized in a different measure space. Indeed, by manipulating the integrand in

$$(2\pi i)^{-d/2} \int_{\mathbb{R}^d} e^{(i/2)(x,x)} f(x) dx$$

one has

$$\int_{\mathbb{R}^d} \frac{e^{(i/2)(x,x)}}{(2\pi i)^{d/2}} f(x) dx = \int_{\mathbb{R}^d} \frac{e^{(i/2)(x,x)+(1/2)(x,x)}}{i^{d/2}} f(x) \frac{e^{-(1/2)(x,x)}}{(2\pi)^{d/2}} dx \qquad [16]$$

where the latter line can be interpreted as the distributional pairing of

$$\frac{e^{(i/2)(x,x)+(1/2)(x,x)}}{i^{d/2}}$$

and $f$ not with respect to Lebesgue measure but rather with respect to the standard Gaussian measure

$$\frac{e^{-(1/2)(x,x)}}{(2\pi)^{d/2}} dx$$

on $\mathbb{R}^d$. The RHS of [16] can be generalized to the case in which $\mathbb{R}^d$ is replaced by a path space, thanks to the fact that on infinite-dimensional spaces, even if Lebesgue measure is meaningless, Gaussian measures are well defined and can be used as reference measures. The detailed realization of this idea as well as its application to the mathematical realization of the Feynman integrand are rather technical and we certainly do not provide details here. We recall that this approach has been successfully applied to the rigorous realization of Feynman path-integral formulation of Chern–Simons models.

### Other Possible Approaches

Another possible mathematical definition of Feynman path integrals is based on Poisson measures. It was originally proposed by A M Chebotarev and V P Maslov and further developed by several authors such as S Albeverio, Ph Blanchard, Ph Combe, R Høegh-Krohn, M Sirugue, and V Kolokol'tsov. It can be applied to "phase-space integrals," to the Dirac equation and in particular algebraic settings, as well as to the Schrödinger

equation, with potentials of the same type "Fourier transform of bounded measure" discussed in the subsection "Infinite-dimensional oscillatory integrals."

Another possible definition of Feynman path integrals is based on a "time-slicing" approximation and a limiting procedure, rather closed to Feynman's original work based on Trotter product formula. The "sequential approach" was proposed originally by A Truman and further extensively developed by D Fujiwara and N Kumano-go. The paths $\gamma$ in formula [2] are approximated by piecewise linear paths and the Feynman path integral is correspondingly approximated by a finite-dimensional integral. In particular, D Fujiwara and N Kumano-go proved that the integrals defined in this way have some important properties, such as invariance under translations and orthogonal transformations. It is also possible to interchange the order of integration with Riemann–Stieltjies integrals and study the semiclassical approximation.

Finally, it is worthwhile to recall a very interesting and intuitive approach to the Feynman integration which is based on nonstandard analysis. It was introduced by S Albeverio, J E Fenstad, R Høegh-Krohn, and T Linstrøm in the 1980s, but it has not been systematically developed yet.

## Abbreviations

| | |
|---|---|
| $D\gamma$ | Heuristic Lebesgue-type measure on the space of paths |
| $P_{t,x}$ | Wiener Gaussian measure on $W_{t,x}$ |
| $S_t$ | Action functional |
| $S_t^\circ$ | Action functional for the free particle |
| $V$ | Potential |
| $W_{t,x}$ | Space of continuous paths with fixed initial point $W_{t,x} = \{w \in C(0,t;\mathbb{R}^d) : w(0) = x\}$ |
| $\hbar$ | Reduced Planck constant |
| $\Phi$ | Phase function |
| $\gamma$ | Path, $\gamma : [0,t] \to \mathbb{R}^d$ |
| $\hat{\mu}$ | Fourier transform of the measure $\mu$ |
| $\psi$ | Wave function, solution of the Schrödinger equation |
| $\mathcal{H}$ | Hilbert space |
| $\widetilde{\int}_{\mathcal{H}}$ | Fresnel integral on the Hilbert space $\mathcal{H}$ |
| $\widetilde{\int^\circ}_{\mathcal{H}}$ | Infinite-dimensional oscillatory integral on the Hilbert space $\mathcal{H}$ |
| $\langle , \rangle$ | inner product |
| $\| \ \|$ | norm |

*See also:* Chern–Simons Models: Rigorous Results; Euclidean Field Theory; Functional Integration in Quantum Physics; Path Integrals in Noncommutative Geometry; Quillen Determinant; Singularity and Bifurcation Theory; Stationary Phase Approximation.

## Further Reading


Albeverio S (1997) Wiener and Feynman – path integrals and their applications. *Proceedings of the Norbert Wiener Centenary Congress 1994* (East Lansing, MI, 1994), 163–194, Proc. Sympos. Appl. Math., 52. Providence, RI: American Mathematical Society.

Albeverio S and Høegh-Krohn R (1977) Oscillatory integrals and the method of stationary phase in infinitely many dimensions, with applications to the classical limit of quantum mechanics. *Inventiones Mathematicae* 40(1): 59–106.

Albeverio S and Høegh-Krohn R (2005) *Mathematical Theory of Feynman Path Integrals*, 2nd edn., with Mazzucchi S Lecture Notes in Mathematics, vol. 523. Berlin: Springer.

Elworthy D and Truman A (1984) Feynman maps, Cameron–Martin formulae and anharmonic oscillators. *Annales de l'Institut Henri Poincare Physique Theorique* 41(2): 115–142.

Hida T, Kuo HH, Potthoff J, and Streit L (1995) *White Noise*. Dordrecht: Kluwer.

Johnson GW and Lapidus ML (2000) *The Feynman Integral and Feynman's Operational Calculus*. New York: Oxford University Press.

Special issue of *Journal of Mathematical Physics* on functional integration, vol. 36, no. 5 (1995).


# Finite-Dimensional Algebras and Quivers

**A Savage**, University of Toronto, Toronto, ON, Canada



## Introduction

Algebras and their representations are ubiquitous in mathematics. It turns out that representations of finite-dimensional algebras are intimately related to quivers, which are simply oriented graphs. Quivers arise naturally in many areas of mathematics, including representation theory, algebraic and differential geometry, Kac–Moody algebras, and quantum groups. In this article, we give a brief overview of some of these topics. We start by giving the basic definitions of associative algebras and their representations. We then introduce quivers and their representation theory, mentioning the connection to the representation theory of associative algebras. We also discuss in some detail the relationship between quivers and the theory of Lie algebras.

## Associative Algebras

An "algebra" is a vector space $A$ over a field $k$ equipped with a multiplication which is distributive and such that

$$a(xy)=(ax)y=x(ay), \quad \forall a\in k,\ x,y\in A$$

When we wish to make the field explicit, we call $A$ a $k$-algebra. An algebra is "associative" if $(xy)z=x(yz)$ for all $x,y,z\in A$. $A$ has a "unit," or "multiplicative identity," if it contains an element $\mathbf{1}_A$ such that $\mathbf{1}_A x = x\mathbf{1}_A = x$ for all $x\in A$. From now on, we will assume all algebras are associative with unit. $A$ is said to be "commutative" if $xy=yx$ for all $x,y\in A$ and finite dimensional if the underlying vector space of $A$ is finite dimensional.

A vector subspace $I$ of $A$ is called a "left (resp. right) ideal" if $xy\in I$ for all $x\in A, y\in I$ (resp. $x\in I, y\in A$). If $I$ is both a right and a left ideal, it is called a two-sided ideal of $A$. If $I$ is a two-sided ideal of $A$, then the factor space $A/I$ is again an algebra.

An algebra homomorphism is a linear map $f: A_1\to A_2$ between two algebras such that

$$f(\mathbf{1}_{A_1})=\mathbf{1}_{A_2}$$
$$f(xy)=f(x)f(y), \quad \forall x,y\in A$$

A representation of an algebra $A$ is an algebra homomorphism $\rho: A\to \mathrm{End}_k(V)$ for a $k$-vector space $V$. Here $\mathrm{End}_k(V)$ is the space of endomorphisms of the vector space $V$ with multiplication given by composition. Given a representation of an algebra $A$ on a vector space $V$, we may view $V$ as an $A$-module with the action of $A$ on $V$ given by

$$a\cdot v=\rho(a)v, \quad a\in A,\ v\in V$$

A morphism $\psi: V\to W$ of two $A$-modules (or equivalently, representations of $A$) is a linear map commuting with the action of $A$. That is, it is a linear map satisfying

$$a\cdot\psi(v)=\psi(a\cdot v), \quad \forall a\in A,\ v\in V$$

Let $G$ be a commutative monoid (a set with an associative multiplication and a unit element). A $G$-graded $k$-algebra is a $k$-algebra which can be expressed as a direct sum $A=\oplus_{g\in G}A_g$ such that $aA_g\subset A_g$ for all $a\in k$ and $A_{g_1}A_{g_2}\subset A_{g_1+g_2}$ for all $g_1,g_2\in G$. A morphism $\psi: A\to B$ of $G$-graded algebras is a $k$-algebra morphism respecting the grading, that is, satisfying $\psi(A_g)\subset B_g$ for all $g\in G$.

## Quivers and Path Algebras

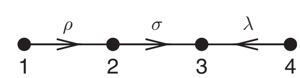
A "quiver" is simply an oriented graph. More precisely, a quiver is a pair $Q=(Q_0,Q_1)$ where $Q_0$ is a finite set of vertices and $Q_1$ is a finite set of arrows (oriented edges) between them. For $a\in Q_1$, we let $h(a)$ denote the "head" of $a$ and $t(a)$ denote the "tail" of $a$. A path in $Q$ is a sequence $x=\rho_1\rho_2\ldots\rho_m$ of arrows such that $h(\rho_{i+1})=t(\rho_i)$ for $1\le i\le m-1$. We let $t(x)=t(\rho_m)$ and $h(x)=h(\rho_1)$ denote the initial and final vertices of the path $x$. For each vertex $i\in Q_0$, we let $e_i$ denote the trivial path which starts and ends at the vertex $i$.

Fix a field $k$. The path algebra $kQ$ associated to a quiver $Q$ is the $k$-algebra whose underlying vector space has basis the set of paths in $Q$, and with the product of paths given by concatenation. Thus, if $x=\rho_1\ldots\rho_m$ and $y=\sigma_1\ldots\sigma_n$ are two paths, then $xy=\rho_1\ldots\rho_m\sigma_1\ldots\sigma_n$ if $h(y)=t(x)$ and $xy=0$ otherwise. We also have

$$e_ie_j=\begin{cases} e_i & \text{if } i=j\\ 0 & \text{if } i\neq j\end{cases}$$
$$e_ix=\begin{cases} x & \text{if } h(x)=i\\ 0 & \text{if } h(x)\neq i\end{cases}$$
$$xe_i=\begin{cases} x & \text{if } t(x)=i\\ 0 & \text{if } t(x)\neq i\end{cases}$$

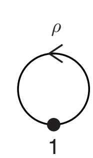
for $x\in kQ$. This multiplication is associative. Note that $e_iA$ and $Ae_i$ have bases given by the set of paths ending and starting at $i$, respectively. The path algebra has a unit given by $\sum_{i\in Q_0}e_i$.

**Example 1** Let $Q$ be the following quiver:




then $kQ$ has a basis given by the set of paths $\{e_1,e_2,e_3,e_4,\rho,\sigma,\lambda,\sigma\rho\}$. Some sample products are $\rho\sigma=0, \lambda\lambda=0, \lambda\sigma=0, e_3\sigma=\sigma e_2=\sigma, e_2\sigma=0$.

**Example 2** Let $Q$ be the following quiver (the so-called "Jordan quiver").




Then $kQ\cong k[t]$, the algebra of polynomials in one variable.

Note that the path algebra $kQ$ is finite dimensional if and only if $Q$ has no oriented cycles (paths with the same head and tail vertex).

**Example 3** Let $Q$ be the following quiver:

Then for every $1 \leq i \leq j \leq n$, there is a unique path from $i$ to $j$. Let $f : kQ \to M_n(k)$ be the linear map from the path algebra to the $n \times n$ matrices with entries in the field $k$ that sends the unique path from $i$ to $j$ to the matrix $E_{ji}$ with $(j, i)$ entry 1 and all other entries zero. Then one can show that $f$ is an isomorphism onto the algebra of lower triangular matrices.

## Representations of Quivers

Fix a field $k$. A representation of a quiver $Q$ is an assignment of a vector space to each vertex and to each arrow a linear map between the vector spaces assigned to its tail and head. More precisely, a representation $V$ of $Q$ is a collection

$$\{V_i | i \in Q_0\}$$

of finite-dimensional $k$-vector spaces together with a collection

$$\{V_\rho : V_{t(\rho)} \to V_{h(\rho)} | \rho \in Q_1\}$$

of $k$-linear maps. Note that a representation $V$ of a quiver $Q$ is equivalent to a representation of the path algebra $kQ$. The dimension of $V$ is the map $d_V : Q_0 \to \mathbb{Z}_{\geq 0}$ given by $d_V(i) = \dim V_i$ for $i \in Q_0$.

If $V$ and $W$ are two representations of a quiver $Q$, then a morphism $\psi : V \to W$ is a collection of $k$-linear maps

$$\{\psi_i : V_i \to W_i | i \in Q_0\}$$

such that

$$W_\rho \psi_{t(\rho)} = \psi_{h(\rho)} V_\rho, \quad \forall \rho \in Q_1$$

**Proposition 1** *Let $A$ be a finite-dimensional $k$-algebra. Then the category of representations of $A$ is equivalent to the category of representations of the algebra $kQ/I$ for some quiver $Q$ and some two-sided ideal $I$ of $kQ$.*

It is for this reason that the study of finite-dimensional associative algebras is intimately related to the study of quivers.

We define the direct sum $V \oplus W$ of two representations $V$ and $W$ of a quiver $Q$ by

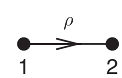
$$(V \oplus W)_i = V_i \oplus W_i, \quad i \in Q_0$$

and $(V \oplus W)_\rho : V_{t(\rho)} \oplus W_{t(\rho)} \to V_{h(\rho)} \oplus W_{h(\rho)}$ by

$$(V \oplus W)_\rho((v, w)) = (V_\rho(v), W_\rho(w))$$

for $v \in V_{t(\rho)}, w \in W_{t(\rho)}, \rho \in Q_1$. A representation $V$ is "trivial" if $V_i = 0$ for all $i \in Q_0$ and "simple" if its only subrepresentations are the zero representation and $V$ itself. We say that $V$ is "decomposable" if it is isomorphic to $W \oplus U$ for some nontrivial representations $W$ and $U$. Otherwise, we call $V$ "indecomposable." Every representation of a quiver has a decomposition into indecomposable representations that is unique up to isomorphism and permutation of the components. Thus, to classify all representations of a quiver, it suffices to classify the indecomposable representations.

**Example 4** Let $Q$ be the following quiver:




Then $Q$ has three indecomposable representations $U$, $V$, and $W$ given by:

$$\begin{aligned} U_1 &= k, \quad & U_2 &= 0, \quad & U_\rho &= 0 \\ V_1 &= 0, \quad & V_2 &= k, \quad & V_\rho &= 0 \\ W_1 &= k, \quad & W_2 &= k, \quad & W_\rho &= 1 \end{aligned}$$

Then any representation $Z$ of $Q$ is isomorphic to

$$Z \cong U^{d_1 - r} \oplus V^{d_2 - r} \oplus W^r$$

where $d_1 = \dim Z_1, d_2 = \dim Z_2, r = \text{rank } Z_\rho$.

**Example 5** Let $Q$ be the Jordan quiver. Then representations $V$ of $Q$ are classified up to isomorphism by the Jordan normal form of $V_\rho$ where $\rho$ is the single arrow of the quiver. Indecomposable representations correspond to single Jordan blocks. These are parametrized by a discrete parameter $n$ (the size of the block) and a continuous parameter $\lambda$ (the eigenvalue of the block).

A quiver is said to be of "finite type" if it has only finitely many indecomposable representations (up to isomorphism). If a quiver has infinitely many isomorphism classes but they can be split into families, each parametrized by a single continuous parameter, then we say the quiver is of "tame" (or "affine") type. If a quiver is of neither finite nor tame type, it is of "wild type." It turns out that there is a rather remarkable relationship between the classification of quivers and their representations and the theory of Kac–Moody algebras.

The "Euler form" or "Ringel form" of a quiver $Q$ is defined to be the asymmetric bilinear form on $\mathbb{Z}^{Q_0}$ given by

$$\langle \alpha, \beta \rangle = \sum_{i \in Q_0} \alpha(i)\beta(i) - \sum_{\rho \in Q_1} \alpha(t(\rho))\beta(h(\rho))$$

In the standard coordinate basis of $\mathbb{Z}^{Q_0}$, the Euler form is represented by the matrix $E = (a_{ij})$ where

$$a_{ij} = \delta_{ij} - \#\{\rho \in Q_1 \mid t(\rho) = i, h(\rho) = j\}$$

Here $\delta_{ij}$ is the Kronecker delta symbol. We define the "Cartan form" of the quiver $Q$ to be the symmetric bilinear form given by

$$(\alpha, \beta) = \langle \alpha, \beta \rangle + \langle \beta, \alpha \rangle$$

Note that the Cartan form is independent of the orientation of the arrows in $Q$. In the standard coordinate basis of $\mathbb{Z}^{Q_0}$, the Cartan form is represented by the Cartan matrix $C = (c_{ij})$ where $c_{ij} = a_{ij} + a_{ji}$.

**Example 6** For the quiver in Example 1, the Euler matrix is

$$E = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

and the Cartan matrix is

$$C = \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}$$

The "Tits form" $q$ of a quiver $Q$ is defined by

$$q(\alpha) = \langle \alpha, \alpha \rangle = \tfrac{1}{2}(\alpha, \alpha)$$

It is known that the number of continuous parameters describing representations of dimension $\alpha$ for $\alpha \neq 0$ is greater than or equal to $1 - q(\alpha)$.

Let $\mathfrak{g}$ be the Kac–Moody algebra associated to the Cartan matrix of a quiver $Q$. By forgetting the orientation of the arrows of $Q$, we obtain the underlying (undirected) graph. This is the Dynkin graph of $\mathfrak{g}$. Associated to $\mathfrak{g}$ is a root system and a set of simple roots $\{\alpha_i \mid i \in Q_0\}$ indexed by the vertices of the Dynkin graph.

**Theorem 1** (Gabriel's theorem).

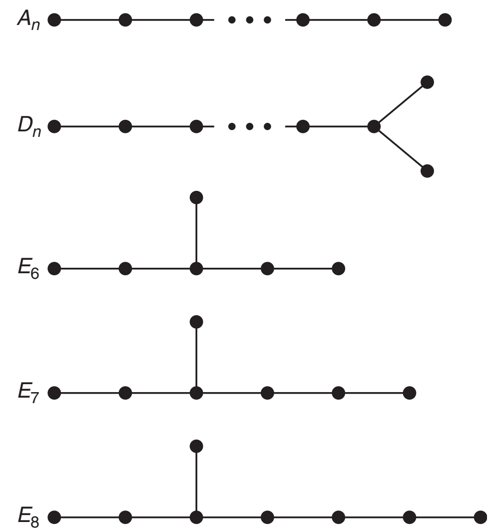
(i) *A quiver is of finite type if and only if the underlying graph is a union of Dynkin graphs of type A, D, or E.*

(ii) *A quiver is of tame type if and only if the underlying graph is a union of Dynkin graphs of type A, D, or E and extended Dynkin graphs of type* $\widehat{A}, \widehat{D}$, *or* $\widehat{E}$ (*with at least one extended Dynkin graph*).

(iii) *The isomorphism classes of indecomposable representations of a quiver Q of finite type are in one-to-one correspondence with the positive roots of the root system associated to the underlying graph of Q. The correspondence is given by*

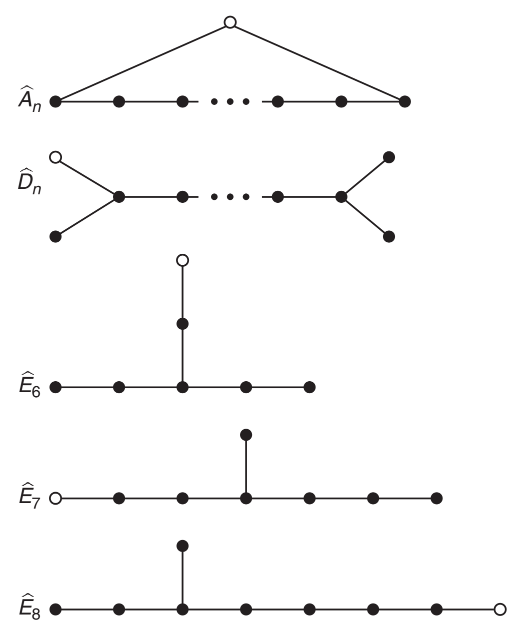
$$V \mapsto \sum_{i \in Q_0} d_V(i)\alpha_i$$

The Dynkin graphs of type $A$, $D$, and $E$ are as follows.




Here the subscript indicates the number of vertices in the graph.

The extended Dynkin graphs of type $\widehat{A}, \widehat{D}$, and $\widehat{E}$ are as follows.




Here we have used an open dot to denote the vertex that was added to the corresponding Dynkin graph of type $A$, $D$, or $E$.

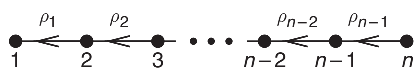
**Theorem 2** (Kac's theorem). *Let Q be an arbitrary quiver. The dimension vectors of indecomposable representations of Q correspond to positive roots*

*of the root system associated to the underlying graph of $Q$ (and are thus independent of the orientation of the arrows of $Q$). The correspondence is given by*

$$d_V \mapsto \sum_{i \in Q_0} d_V(i)\alpha_i$$

Note that in Kac's Theorem, it is not asserted that the isomorphism classes are in one-to-one correspondence with the roots as in the finite case considered in Gabriel's theorem. It turns out that in the general case, dimension vectors for which there is exactly one isomorphism class correspond to real roots while imaginary roots correspond to dimension vectors for which there are families of representations.

**Example 7** Let $Q$ be the quiver of type $A_n$, oriented as follows.




It is known that the set of positive roots of the simple Lie algebra of type $A_n$ is

$$\left\{ \sum_{i=j}^{l} \alpha_i \,\middle|\, 1 \le j \le l \le n \right\} \sqcup \{0\}$$

The zero root corresponds to the trivial representation. The root $\sum_{i=j}^{l} \alpha_i$ for some $1 \le j \le l \le n$ corresponds to the unique (up to isomorphism) representation $V$ with

$$V_i = \begin{cases} k & \textit{if } j \le i \le l \\ 0 & \textit{otherwise} \end{cases}$$

and

$$V_{\rho_i} = \begin{cases} 1 & \textit{if } j \le i \le l-1 \\ 0 & \textit{otherwise} \end{cases}$$

**Example 8** Let $Q$ be the quiver of type $\widehat{A}_n$, with all arrows oriented in the same direction (for instance, counter-clockwise). The positive root $\sum_{i=0}^{n} \alpha_i$ (where $\{0, 1, 2, \ldots, n\}$ are the vertices of the quiver) is imaginary. There is a one-parameter family of isomorphism classes of indecomposable representations where the maps assigned to each arrow are nonzero. The parameter is the composition of the maps around the loop.

If a quiver $Q$ has no oriented cycles, then the only simple $kQ$-modules are the modules $S^i$ for $i \in Q_0$ where

$$S^i_j = \begin{cases} k & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

and $S^i_\rho = 0$ for all $\rho \in Q_1$.

## Ringel–Hall Algebras

Let $k$ be the finite field $\mathbb{F}_q$ with $q$ elements and let $Q$ be a quiver with no oriented cycles. Let $\mathcal{P}$ be the set of all isomorphism classes of $kQ$-modules which are finite as sets (since $k$ is finite dimensional, these are just the quiver representations we considered above). Let $A$ be a commutative integral domain containing $\mathbb{Z}$ and elements $v, v^{-1}$ such that $v^2 = q$. The Ringel–Hall algebra $H = H_{A,v}(kQ)$ is the free $A$-module with basis $\{[V]\}$ indexed by the isomorphism classes of representations of the quiver $Q$, with an $A$-bilinear multiplication defined by

$$[V^1] \cdot [V^2] = v^{\langle \dim V^1, \dim V^2 \rangle} \sum_V g^V_{V^1,V^2}[V]$$

Here $\langle \dim V^1, \dim V^2 \rangle$ is the Euler form and $g^V_{V^1,V^2}$ is the number of submodules $W$ of $V$ such that $V/W \cong V^1$ and $W \cong V^2$. $H$ is an associative $\mathbb{Z}^{Q_0}_{\ge 0}$-graded algebra, with identity element $[0]$, the isomorphism class of the trivial representation. The grading $H = \oplus_\alpha H_\alpha$ is given by letting $H_\alpha$ be the $A$-span of the set of isomorphism classes $[V]$ such that $\dim V = \alpha$.

Let $C = C_{A,v}(kQ)$ be the $A$-subalgebra of $H$ generated by the isomorphism classes $[S^i]$ of the simple $kQ$-modules. $C$ is called the "composition algebra." If the underlying graph of $Q$ is of finite type, then $C = H$.

Now let $\mathcal{K}$ be a set of finite fields $k$ such that the set $\{|k| \,|\, k \in \mathcal{K}\}$ is infinite. Let $A$ be an integral domain containing $\mathbb{Q}$ and, for each $k \in \mathcal{K}$, an element $v_k$ such that $v_k^2 = |k|$. For each $k \in \mathcal{K}$, we have the corresponding composition algebra $C_k$, generated by the elements $[{}^kS^i]$ (here we make the field $k$ explicit). Now let $C$ be the subring of $\prod_{k \in \mathcal{K}} C_k$ generated by $\mathbb{Q}$ and the elements

$$\begin{aligned} t &= (t_k)_{k \in \mathcal{K}}, & t_k &= v_k \\ t^{-1} &= (t_k^{-1})_{k \in \mathcal{K}}, & t_k^{-1} &= (v^k)^{-1} \\ u^i &= (u^i_k)_{k \in \mathcal{K}}, & u^i_k &= [{}^kS^i],\ i \in Q_0 \end{aligned}$$

Now, $t$ lies in the center of $C$ and if $p(t) = 0$ for some polynomial $p$, then $p$ must be the zero polynomial since the set of $v_k$ is infinite. Thus, we may think of $C$ as the $A$-algebra generated by the $u^i, i \in Q_0$, with $A = \mathbb{Q}[t, t^{-1}]$ and $t$ an indeterminate. Let $C^* = \mathbb{Q}(t) \otimes_A C$. We call $C^*$ the "generic composition algebra."

Let $\mathfrak{g}$ be the Kac–Moody algebra associated to the Cartan matrix of the quiver $Q$ and let $U$ be the quantum group associated by Drinfeld and Jimbo to $\mathfrak{g}$. It has a triangular decomposition $U = U^- \otimes U^0 \otimes U^+$.

Specifically, $U^+$ is the $\mathbb{Q}(t)$-algebra with generators $E_i, i \in Q_0$ and relations

$$\sum_{p=0}^{1-c_{ij}} (-1)^p \begin{bmatrix} 1-c_{ij} \\ p \end{bmatrix} E_i^p E_j E_i^{1-c_{ij}-p}, \quad i \neq j$$

where $c_{ij}$ are the entries of the Cartan matrix and

$$\begin{bmatrix} m \\ p \end{bmatrix} = \frac{[m]!}{[p]![m-p]!}$$

$$[n] = \frac{t^n - t^{-n}}{t - t^{-1}}, \quad [n]! = [1][2]\ldots[n]$$

**Theorem 3** *There is a* $\mathbb{Q}(t)$*-algebra isomorphism* $C^* \to U^+$ *sending* $u_i \mapsto E_i$ *for all* $i \in Q_0$.

The proof of Theorem 3 is due to Ringel in the case that the underlying graph of $Q$ is of finite or affine type. The more general case presented here is due to Green.

All of the Kac–Moody algebras considered so far have been simply-laced. That is, their Cartan matrices are symmetric. There is a way to deal with non-simply-laced Kac–Moody algebras using species. We will not treat this subject in this article.

## Quiver Varieties

One can use varieties associated to quivers to yield a geometric realization of the upper half of the universal enveloping algebra of a Kac–Moody algebra $\mathfrak{g}$ and its irreducible highest-weight representations.

### Lusztig's Quiver Varieties

We first introduce the quiver varieties, first defined by Lusztig, which yield a geometric realization of the upper half $\mathcal{U}^+$ of the universal enveloping algebra of a simply laced Kac–Moody algebra $\mathfrak{g}$. Let $Q = (Q_0, Q_1)$ be the quiver whose vertices $Q_0$ are the vertices of the Dynkin diagram of $\mathfrak{g}$ and whose set of arrows $Q_1$ consists of all the edges of the Dynkin diagram with both orientations. By definition, $\mathcal{U}^+$ is the $\mathbb{Q}$-algebra defined by generators $e_i, i \in Q_0$, subject to the Serre relations

$$\sum_{p=0}^{1-c_{ij}} (-1)^p \begin{pmatrix} 1-c_{ij} \\ p \end{pmatrix} e_i^p e_j e^{1-c_{ij}-p} = 0$$

for all $i \neq j$ in $Q_0$, where $c_{ij}$ are the entries of the Cartan matrix associated to $Q$. For any $\nu = \sum_{i \in Q_0} \nu_i i, \nu_i \in \mathbb{N}$, let $\mathcal{U}_\nu^+$ be the subspace of $\mathcal{U}^+$ spanned by the monomials $e_{i_1} e_{i_2} \ldots e_{i_n}$ for various sequences $i_1, i_2, \ldots, i_n$ in which $i$ appears $\nu_i$ times for each $i \in Q_0$. Thus, $\mathcal{U}^+ = \oplus_\nu \mathcal{U}_\nu^+$. Let $\mathcal{U}_\mathbb{Z}^+$ be the subring of $\mathcal{U}^+$ generated by the elements $e_i^p/p!$ for $i \in Q_0, p \in \mathbb{N}$. Then $\mathcal{U}_\mathbb{Z}^+ = \oplus_\nu \mathcal{U}_{\mathbb{Z},\nu}^+$ where $\mathcal{U}_{\mathbb{Z},\nu}^+ = \mathcal{U}_\mathbb{Z}^+ \cap \mathcal{U}_\nu^+$.

We define the involution $\bar{\ }: Q_1 \to Q_1$ to be the function which takes $\rho \in Q_1$ to the element of $Q_1$ consisting of the same edge with opposite orientation. An orientation of our graph/quiver is a choice of a subset $\Omega \subset Q_1$ such that $\Omega \cup \bar{\Omega} = Q_1$ and $\Omega \cap \bar{\Omega} = \emptyset$.

Let $\mathcal{V}$ be the category of finite-dimensional $Q_0$-graded vector spaces $\boldsymbol{V} = \oplus_{i \in Q_0} \boldsymbol{V}_i$ over $\mathbb{C}$ with morphisms being linear maps respecting the grading. Then $\boldsymbol{V} \in \mathcal{V}$ shall denote that $\boldsymbol{V}$ is an object of $\mathcal{V}$. The dimension of $\boldsymbol{V} \in \mathcal{V}$ is given by $\boldsymbol{v} = \dim \boldsymbol{V} = (\dim \boldsymbol{V}_0, \ldots, \dim \boldsymbol{V}_n)$.

Given $\boldsymbol{V} \in \mathcal{V}$, let $E_V$ be the space of representations of $Q$ with underlying vector space $\boldsymbol{V}$. That is,

$$\boldsymbol{E}_V = \bigoplus_{\rho \in Q_1} \mathrm{Hom}(\boldsymbol{V}_{t(\rho)}, \boldsymbol{V}_{h(\rho)})$$

For any subset $Q_1'$ of $Q_1$, let $\boldsymbol{E}_{V,Q_1'}$ be the subspace of $\boldsymbol{E}_V$ consisting of all vectors $x = (x_\rho)$ such that $x_\rho = 0$ whenever $\rho \notin Q_1'$. The algebraic group $G_V = \prod_i \mathrm{Aut}(\boldsymbol{V}_i)$ acts on $\boldsymbol{E}_V$ and $\boldsymbol{E}_{V,Q_1'}$ by

$$(g, x) = ((g_i), (x_\rho)) \mapsto gx$$
$$= (x_\rho') = (g_{h(\rho)} x_\rho g_{t(\rho)}^{-1})$$

Define the function $\varepsilon : Q_1 \to \{-1, 1\}$ by $\varepsilon(\rho) = 1$ for all $\rho \in \Omega$ and $\varepsilon(\rho) = -1$ for all $\rho \in \bar{\Omega}$. Let $\langle \cdot, \cdot \rangle$ be the nondegenerate, $G_V$-invariant, symplectic form on $\boldsymbol{E}_V$ with values in $\mathbb{C}$ defined by

$$\langle x, y \rangle = \sum_{\rho \in Q_1} \varepsilon(\rho) \mathrm{tr}(x_\rho y_{\bar{\rho}})$$

Note that $\boldsymbol{E}_V$ can be considered as the cotangent space of $\boldsymbol{E}_{V,\Omega}$ under this form.

The moment map associated to the $G_V$-action on the symplectic vector space $\boldsymbol{E}_V$ is the map $\psi : \boldsymbol{E}_V \to \boldsymbol{gl}_V = \prod_i \mathrm{End}\boldsymbol{V}_i$, the Lie algebra of $\mathrm{GL}_V$, with $i$-component $\psi_i : \boldsymbol{E}_V \to \mathrm{End}\boldsymbol{V}_i$ given by

$$\psi_i(x) = \sum_{\rho \in Q_1, h(\rho) = i} \varepsilon(\rho) x_\rho x_{\bar{\rho}}$$

**Definition 1** An element $x \in \boldsymbol{E}_V$ is said to be nilpotent if there exists an $N \geq 1$ such that for any sequence $\rho_1, \rho_2, \ldots, \rho_N$ in $H$ satisfying $t(\rho_1) = h(\rho_2)$, $t(\rho_2) = h(\rho_3), \ldots, t(\rho_{N-1}) = h(\rho_N)$, the composition $x_{\rho_1} x_{\rho_2} \ldots x_{\rho_N} : \boldsymbol{V}_{t(\rho_N)} \to \boldsymbol{V}_{h(\rho_1)}$ is zero.

**Definition 2** Let $\Lambda_V$ be the set of all nilpotent elements $x \in \boldsymbol{E}_V$ such that $\psi_i(x) = 0$ for all $i \in I$.

A subset of an algebraic variety is said to be "constructible" if it is obtained from subvarieties from a finite number of the usual set-theoretic operations. A function $f : A \to \mathbb{Q}$ on an algebraic variety $A$ is said to be a constructible function if $f^{-1}(a)$ is a constructible set for all $a \in \mathbb{Q}$ and is empty for all but finitely many $a$. Let $M(\Lambda_V)$ denote the $\mathbb{Q}$-vector space of all constructible functions on $\Lambda_V$. Let $\widetilde{M}(\Lambda_V)$ denote the $\mathbb{Q}$-subspace of $M(\Lambda_V)$ consisting of those functions that are constant on any $G_V$-orbit in $\Lambda_V$.

Let $\boldsymbol{V}, \boldsymbol{V}', \boldsymbol{V}'' \in \mathcal{V}$ such that $\dim \boldsymbol{V} = \dim \boldsymbol{V}' + \dim \boldsymbol{V}''$. Now, suppose that $\boldsymbol{S}$ is an $I$-graded subspace of $\boldsymbol{V}$. For $x \in \Lambda_V$ we say that $\boldsymbol{S}$ is $x$-stable if $x(\boldsymbol{S}) \subset \boldsymbol{S}$. Let $\Lambda_{V;V',V''}$ be the variety consisting of all pairs $(x, \boldsymbol{S})$ where $x \in \Lambda_V$ and $\boldsymbol{S}$ is an $I$-graded $x$-stable subspace of $\boldsymbol{V}$ such that $\dim \boldsymbol{S} = \dim \boldsymbol{V}''$. Now, if we fix some isomorphisms $\boldsymbol{V}/\boldsymbol{S} \cong \boldsymbol{V}', \boldsymbol{S} \cong \boldsymbol{V}''$, then $x$ induces elements $x' \in \Lambda_{V'}$ and $x'' \in \Lambda_{V''}$. We then have the maps

$$\Lambda_{V'} \times \Lambda_{V''} \xleftarrow{p_1} \Lambda_{V;V',V''} \xrightarrow{p_2} \Lambda_V$$

where $p_1(x, \boldsymbol{S}) = (x', x''), p_2(x, \boldsymbol{S}) = x$.

For a holomorphic map $\pi$ between complex varieties $A$ and $B$, let $\pi_!$ denote the map between the spaces of constructible functions on $A$ and $B$ given by

$$(\pi_! f)(y) = \sum_{a \in \mathbb{Q}} a \chi(\pi^{-1}(y) \cap f^{-1}(a))$$

Let $\pi^*$ be the pullback map from functions on $B$ to functions on $A$ acting as $\pi^* f(y) = f(\pi(y))$. We then define a map

$$\widetilde{M}(\Lambda_{V'}) \times \widetilde{M}(\Lambda_{V''}) \to \widetilde{M}(\Lambda_V) \qquad [1]$$

by $(f', f'') \mapsto f' * f''$ where

$$f' * f'' = (p_2)_! p_1^*(f' \times f'')$$

Here $f' \times f'' \in \widetilde{M}(\Lambda_{V'} \times \Lambda_{V''})$ is defined by $(f' \times f'')(x', x'') = f'(x') f''(x'')$. The map [1] is bilinear and defines an associative $\mathbb{Q}$-algebra structure on $\oplus_\nu \widetilde{M}(\Lambda_{V^\nu})$ where $\boldsymbol{V}^\nu$ is the object of $\mathcal{V}$ defined by $\boldsymbol{V}_i^\nu = \mathbb{C}^{\nu_i}$.

There is a unique algebra homomorphism $\kappa : \mathcal{U}^+ \to \oplus_\nu \widetilde{M}(\Lambda_{V^\nu})$ such that $\kappa(e_i)$ is the function on the point $\Lambda_{V^i}$ with value 1. Then $\kappa$ restricts to a map $\kappa_\nu : \mathcal{U}_\nu^+ \to \widetilde{M}(\Lambda_{V^\nu})$. It can be shown that $\kappa_{pi}(e_i^p / p!)$ is the function 1 on the point $\Lambda_{V^{pi}}$ for $i \in Q_0, p \in \mathbb{Z}_{\geq 0}$.

Let $\widetilde{M}_{\mathbb{Z}}(\Lambda_V)$ be the set of all functions in $\widetilde{M}(\Lambda_V)$ that take on only integer values. One can show that if $f' \in \widetilde{M}_{\mathbb{Z}}(\Lambda_{V'})$ and $f'' \in \widetilde{M}_{\mathbb{Z}}(\Lambda_{V''})$, then $f' * f'' \in \widetilde{M}_{\mathbb{Z}}(\Lambda_V)$ in the setup of [1]. Thus $\kappa_\nu(\mathcal{U}_{\mathbb{Z},\nu}^+) \subseteq \widetilde{M}_{\mathbb{Z}}(\Lambda_{V^\nu})$.

Let $\mathrm{Irr}\Lambda_V$ denote the set of irreducible components of $\Lambda_V$. The following proposition was conjectured by Lusztig and proved by him in the affine (and finite) case. The general case was proved by Kashiwara and Saito.

**Proposition 2** *For any* $\nu \in (\mathbb{Z}_{\geq 0})^{Q_0}$, *we have* $\dim \mathcal{U}_\nu^+ = \#\mathrm{Irr}\Lambda_{V^\nu}$.

We then have the following important result due to Lusztig.

**Theorem 4** *Let* $\nu \in (\mathbb{Z}_{\geq 0})^{Q_0}$. *Then,*

(i) *For any* $Z \in \mathrm{Irr}\Lambda_{V^\nu}$, *there exists a unique* $f_Z \in \kappa_\nu(\mathcal{U}_{\mathbb{Z},\nu}^+)$ *such that* $f_Z$ *is equal to 1 on an open dense subset of* $Z$ *and equal to zero on an open dense subset of* $Z' \in \mathrm{Irr}\Lambda_{V^\nu}$ *for all* $Z' \neq Z$.
(ii) $\{f_Z \mid Z \in \mathrm{Irr}\Lambda_{V^\nu}\}$ *is a* $\mathbb{Q}$-*basis of* $\kappa_\nu(\mathcal{U}_\nu^+)$.
(iii) $\kappa_\nu : \mathcal{U}_\nu^+ \to \kappa_\nu(\mathcal{U}_\nu^+)$ *is an isomorphism.*
(iv) *Define* $[Z] \in \mathcal{U}_\nu^+$ *by* $\kappa_\nu([Z]) = f_Z$. *Then* $B_\nu = \{[Z] \times \mid Z \in \mathrm{Irr}\Lambda_{V^\nu}\}$ *is a* $\mathbb{Q}$-*basis of* $\mathcal{U}_\nu^+$.
(v) $\kappa_\nu(\mathcal{U}_{\mathbb{Z},\nu}^+) = \kappa_\nu(\mathcal{U}_\nu^+) \cap \widetilde{M}_{\mathbb{Z}}(\Lambda_{V^\nu})$.
(vi) $B_\nu$ *is a* $\mathbb{Z}$-*basis of* $\mathcal{U}_{\mathbb{Z},\nu}^+$.

From this theorem, we see that $B = \sqcup_\nu B_\nu$ is a $\mathbb{Q}$-basis of $\mathcal{U}^+$, which is called the "semicanonical basis." This basis has many remarkable properties. One of these properties is as follows. Via the algebra involution of the entire universal enveloping algebra $\mathcal{U}$ of $\mathfrak{g}$ given on the Chevalley generators by $e_i \mapsto f_i, f_i \mapsto e_i$ and $h \mapsto -h$ for $h$ in the Cartan subalgebra of $\mathfrak{g}$, one obtains from the results of this section a semicanonical basis of $\mathcal{U}^-$, the lower half of the universal enveloping algebra of $\mathfrak{g}$. For any irreducible highest-weight integrable representation $V$ of $\mathcal{U}$ (or, equivalently, $\mathfrak{g}$), let $v \in V$ be a nonzero highest-weight vector. Then the set

$$\{bv | b \in B, bv \neq 0\}$$

is a $\mathbb{Q}$-basis of $V$, called the semicanonical basis of $V$. Thus, the semicanonical basis of $\mathcal{U}^-$ is simultaneously compatible with all irreducible highest-weight integrable modules. There is also a way to define the semicanonical basis of a representation directly in a geometric way. This is the subject of the next subsection.

One can also obtain a geometric realization of the upper part $U^+$ of the quantum group in a similar manner using perverse sheaves instead of constructible functions. This construction yields the canonical basis of the associated quantum group (a $q$-deformation of the universal enveloping algebra) which also has many remarkable properties and is closely related to the theory of crystal bases.

### Nakajima's Quiver Varieties

We introduce here a description of the quiver varieties first presented by Nakajima. They yield a geometric realization of the irreducible highest-weight representations of simply-laced Kac–Moody algebras. The construction was motivated by the work of Kronheimer and Nakajima on solutions to the anti-self-dual Yang–Mills equations on ALE gravitational instantons (*see* Instantons: Topological Aspects).

**Definition 3** For $v, w \in \mathbb{Z}^I_{\geq 0}$, choose I-graded vector spaces $V$ and $W$ of graded dimensions $v$ and $w$, respectively. Then define

$$\Lambda \equiv \Lambda(v, w) = \Lambda_V \times \bigoplus_{i \in I} \mathrm{Hom}(V_i, W_i)$$

**Definition 4** Let $\Lambda^{st} = \Lambda(v, w)^{st}$ be the set of all $(x, t) \in \Lambda(v, w)$ satisfying the following condition: if $S = (S_i)$ with $S_i \subset V_i$ is $x$-stable and $t_i(S_i) = 0$ for $i \in I$, then $S_i = 0$ for $i \in I$.

The group $G_V$ acts on $\Lambda(v, w)$ via

$$(g, (x, t)) \mapsto ((g_{h(\rho)} x_\rho g^{-1}_{t(\rho)}), (t_i g_i^{-1}))$$

and the stabilizer of any point of $\Lambda(v, w)^{\mathrm{st}}$ in $G_V$ is trivial. We then make the following definition.

**Definition 5** Let $\mathcal{L} \equiv \mathcal{L}(v, w) = \Lambda(v, w)^{st}/G_V$.

We should note that while the above definition and other constructions in this article are algebraic, there are also more geometric ways of looking at quiver varieties. In particular, the space

$$M(v, w) = \left( \bigoplus_{\rho \in Q_1} \mathrm{Hom}(V_{t(\rho)}, V_{h(\rho)}) \right) \oplus \left( \bigoplus_{i \in I} \mathrm{Hom}(W_i, V_i) \oplus \mathrm{Hom}(V_i, W_i) \right)$$

has a natural hyper-Kähler metric and one can consider a hyper-Kähler quotient by the group $\prod U(V_i)$. The variety $\mathcal{L}(v, w)$ is a Lagrangian subvariety of (and is homotopic to) this hyper-Kähler quotient. In the case $\mathfrak{g} = \mathfrak{sl}_n$, the varieties involved are closely related to flag varieties.

Let $w, v, v', v'' \in \mathbb{Z}^I_{\geq 0}$ be such that $v = v' + v''$. Consider the maps

$$\Lambda(v'', 0) \times \Lambda(v', w) \xleftarrow{p_1} \tilde{F}(v, w; v'') \xrightarrow{p_2} F(v, w; v'') \xrightarrow{p_3} \Lambda(v, w) \qquad [2]$$

where the notation is as follows. A point of $F(v, w; v'')$ is a point $(x, t) \in \Lambda(v, w)$ together with an $I$-graded, $x$-stable subspace $S$ of $V$ such that $\dim S = v' = v - v''$. A point of $\tilde{F}(v, w; v'')$ is a point $(x, t, S)$ of $F(v, w; v'')$ together with a collection of isomorphisms $R'_i : V'_i \cong S_i$ and $R''_i : V''_i \cong V_i/S_i$ for each $i \in I$. Then we define $p_2(x, t, S, R', R'') = (x, t, S)$, $p_3(x, t, S) = (x, t)$ and $p_1(x, t, S, R', R'') = (x'', x', t')$ where $x'', x', t'$ are determined by

$$R'_{h(\rho)} x'_\rho = x_\rho R'_{t(\rho)} : V'_{t(\rho)} \to S_{h(\rho)}$$
$$t'_i = t_i R'_i : V'_i \to W_i$$
$$R''_{h(\rho)} x''_\rho = x_\rho R''_{t(\rho)} : V''_{t(\rho)} \to V_{h(\rho)}/S_{h(\rho)}$$

It follows that $x'$ and $x''$ are nilpotent.

**Lemma 1** *One has*

$$(p_3 \circ p_2)^{-1}(\Lambda(v, w)^{st}) \subset p_1^{-1}(\Lambda(v'', 0) \times \Lambda(v', w)^{st})$$

Thus, we can restrict [2] to $\Lambda^{\mathrm{st}}$, forget the $\Lambda(v'', 0)$-factor and consider the quotient by $G_V$ and $G_{V'}$. This yields the diagram

$$\mathcal{L}(v', w) \xleftarrow{\pi_1} \mathcal{F}(v, w; v - v') \xrightarrow{\pi_2} \mathcal{L}(v, w) \qquad [3]$$

where

$$\mathcal{F}(v, w, v - v') \stackrel{\mathrm{def}}{=} \{(x, t, S) \in F(v, w; v - v') | (x, t) \in \Lambda(v, w)^{\mathrm{st}}\}/G_V$$

Let $M(\mathcal{L}(v, w))$ be the vector space of all constructible functions on $\mathcal{L}(v, w)$. Then define maps

$$h_i : M(\mathcal{L}(v, w)) \to M(\mathcal{L}(v, w))$$
$$e_i : M(\mathcal{L}(v, w)) \to M(\mathcal{L}(v - e^i, w))$$
$$f_i : M(\mathcal{L}(v - e^i, w)) \to M(\mathcal{L}(v, w))$$

by

$$h_i f = u_i f$$
$$e_i f = (\pi_1)_!(\pi_2^* f)$$
$$f_i g = (\pi_2)_!(\pi_1^* g)$$

Here

$$u = {}^t(u_0, \ldots, u_n) = w - Cv$$

where $C$ is the Cartan matrix of $\mathfrak{g}$ and we are using diagram 3 with $v' = v - e^i$ where $e^i$ is the vector whose components are given by $e^i_j = \delta_{ij}$.

Now let $\varphi$ be the constant function on $\mathcal{L}(0, w)$ with value 1. Let $L(w)$ be the vector space of functions generated by acting on $\varphi$ with all possible combinations of the operators $f_i$. Then let $L(v, w) = M(\mathcal{L}(v, w)) \cap L(w)$.

**Proposition 3** *The operators $e_i, f_i, h_i$ on $L(w)$ provide it with the structure of the irreducible highest-weight*

*integrable representation of* $\mathfrak{g}$ *with highest weight* $\sum_{i\in Q_0} \boldsymbol{w}_i\omega_i$. *Each summand of the decomposition* $L(\boldsymbol{w}) = \bigoplus_{\boldsymbol{v}} L(\boldsymbol{v},\boldsymbol{w})$ *is a weight space with weight* $\sum_{i\in Q_0} \boldsymbol{w}_i\omega_i - \boldsymbol{v}_i\alpha_i$. *Here the* $\omega_i$ *and* $\alpha_i$ *are the fundamental weights and simple roots of* $\mathfrak{g}$, *respectively.*

Let $Z \in \mathrm{Irr}\mathcal{L}(\boldsymbol{v},\boldsymbol{w})$ and define a linear map $T_Z : L(\boldsymbol{v},\boldsymbol{w}) \to \mathbb{C}$ that associates to a constructible function $f \in L(\boldsymbol{v},\boldsymbol{w})$ the (constant) value of $f$ on a suitable open dense subset of $Z$. The fact that $L(\boldsymbol{v},\boldsymbol{w})$ is finite dimensional allows us to take such an open set on which any $f \in L(\boldsymbol{v},\boldsymbol{w})$ is constant. So we have a linear map

$$\Phi : L(\boldsymbol{v},\boldsymbol{w}) \to \mathbb{C}^{\mathrm{Irr}\mathcal{L}(\boldsymbol{v},\boldsymbol{w})}$$

Then we have the following proposition.

**Proposition 4** *The map* $\Phi$ *is an isomorphism*; *for any* $Z \in \mathrm{Irr}\mathcal{L}(\boldsymbol{v},\boldsymbol{w})$, *there is a unique function* $g_Z \in L(\boldsymbol{v},\boldsymbol{w})$ *such that for some open dense subset* $O$ *of* $Z$ *we have* $g_Z|_O = 1$ *and for some closed* $G_V$*-invariant subset* $K \subset \mathcal{L}(\boldsymbol{v},\boldsymbol{w})$ *of dimension* $< \dim \mathcal{L}(\boldsymbol{v},\boldsymbol{w})$ *we have* $g_Z = 0$ *outside* $Z \cup K$. *The functions* $g_Z$ *for* $Z \in \mathrm{Irr}\Lambda(\boldsymbol{v},\boldsymbol{w})$ *form a basis of* $L(\boldsymbol{v},\boldsymbol{w})$.

## Additional Topics

To conclude, we have given here a brief overview of some topics related to finite-dimensional algebras and quivers. There is much more to be found in the literature. For basics on associative algebras and their representations, the reader may consult introductory texts on abstract algebra such as Lang (2002). For further results (and their proofs) on Ringel–Hall algebras see the papers of Ringel (1990a, b, 1993, 1995, 1996) and of Green (1995) and the references cited therein. The reader interested in species, which extend many of these results to non-simply-laced Lie algebras, should consult Dlab and Ringel (1976).

The book by Lusztig (1993) covers the quiver varieties of Lusztig and canonical bases. Canonical bases are closely related to crystal bases and crystal graphs (see Hong and Kang (2002) for an overview of these topics). In fact, the set of irreducible components of the quiver varieties of Lusztig and Nakajima can be endowed with the structure of a crystal graph in a purely geometric way (see Kashiwara and Saito (1997) and Saito (2002)). Many results on Nakajima's quiver varieties can be found in the original papers (Nakajima 1994, 1998). The overview article (Nakajima 1996) is also useful.

Quiver varieties can also be used to give geometric realizations of tensor products of representations (see Malkin (2002, 2003), Nakajima (2001), and Savage (2003)) and finite-dimensional representations of quantum affine Lie algebras (see Nakajima (2001)). This is just a select few of the many applications of quiver varieties. Much more can be found in the literature.

## Acknowledgments

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

*See also:* Instantons: Topological Aspects.

## Further Reading

Dlab V and Ringel CM (1976) Indecomposable representations of graphs and algebras. *Memoirs of the American Mathematical Society* 6(173): v+57.

Green JA (1995) Hall algebras, hereditary algebras and quantum groups. *Inventiones Mathematicae* 120(2): 361–377.

Hong J and Kang S-J (2002) *Introduction to Quantum Groups and Crystal Bases*, Graduate Studies in Mathematics, vol. 42. Providence, RI: American Mathematical Society.

Kashiwara M and Saito Y (1997) Geometric construction of crystal bases. *Duke Mathematical Journal* 89(1): 9–36.

Lang S (2002) *Algebra*, Graduate Texts in Mathematics, vol. 211. New York: Springer.

Lusztig G (1993) *Introduction to Quantum Groups*, Progress in Mathematics, vol. 110. Boston MA: Birkhäuser.

Malkin A (2002) Tensor product varieties and crystals: GL case. *Transactions of the American Mathematical Society* 354(2): 675–704 (electronic).

Malkin A (2003) Tensor product varieties and crystals: the *ADE* case. *Duke Mathematical Journal* 116(3): 477–524.

Nakajima H (1994) Instantons on ALE spaces, quiver varieties, and Kac–Moody algebras. *Duke Mathematical Journal* 76(2): 365–416.

Nakajima H (1996) Varieties associated with quivers. In: Bautista R, Martinez-Villa R, and de la Peña JA (eds.) *Representation Theory of Algebras and Related Topics (Mexico City, 1994)*, CMS Conf. Proc., vol. 19, pp. 139–157. Providence, RI: American Mathematical Society.

Nakajima H (1998) Quiver varieties and Kac–Moody algebras. *Duke Mathematical Journal* 91(3): 515–560.

Nakajima H (2001a) Quiver varieties and finite-dimensional representations of quantum affine algebras. *Journal of the American Mathematical Society* 14(1): 145–238 (electronic).

Nakajima H (2001b) Quiver varieties and tensor products. *Inventiones Mathematicae* 146(2): 399–449.

Ringel CM (1990a) Hall algebras. In: Balcerzyk S, Józefiak T, Krempa J, Simson D, and Vogel W (eds.) *Topics in Algebra, Part 1 (Warsaw, 1988)*, Banach Center Publ., vol. 26, pp. 433–447. Warsaw: PWN.

Ringel CM (1990b) Hall algebras and quantum groups. *Inventiones Mathematicae* 101(3): 583–591.

Ringel CM (1993) Hall algebras revisited. In: Joseph A and Shnider S (eds.) *Quantum Deformations of Algebras and Their Representations (Ramat-Gan, 1991/1992; Rehovot, 1991/1992)*, Israel Math. Conf. Proc., vol. 7, pp. 171–176. Ramat Gan: Bar-Ilan Univ.

Ringel CM (1995) The Hall algebra approach to quantum groups. In: Gomez-Mont X, Montejano L, de la Peña JA, and Seade J (eds.) *XI Latin American School of Mathematics (Spanish) (Mexico City, 1993)*, Aportaciones Mat. Comun., vol. 15, pp. 85–114. México: Soc. Mat. Mexicana.

Ringel CM (1996) Green's theorem on Hall algebras. In: Bautista R, Martinez-Villa R, and de la Peña JA (eds.) *Representation Theory of Algebras and Related Topics (Mexico City, 1994)*, CMS Conf. Proc., vol. 19, pp. 185–245. Providence, RI: American Mathematical Society.

Saito Y (2002) Crystal bases and quiver varieties. *Mathematische Annalen* 324(4): 675–688.

Savage A (2003) The tensor product of representations of $U_q(\mathfrak{sl}_2)$ via quivers. *Advances in Mathematics* 177(2): 297–340.

# Finite Group Symmetry Breaking

**G Gaeta**, Università di Milano, Milan, Italy



## Introduction

It is a commonplace situation that symmetric laws of Nature give rise to physical states which are not symmetric. States related by symmetry operations are equivalent, but still nature selects one of them.

As an example, consider a ferromagnetic system of interacting spins with no external magnetic field. The "up" and "down" states are equivalent, but one of the two is chosen: the interaction makes states with agreeing spin orientation (and therefore macroscopic magnetization) energetically preferred, and fluctuations will decide which state is actually chosen by a given sample.

Finite group symmetry is also commonplace in physics, in particular through crystallographic groups occurring in condensed matter physics – but also through the inversions ($C, P, T$ and their combinations) occurring in high-energy physics and field theory.

The breaking of finite group symmetry has thus been thoroughly studied, and general approaches exist to investigate it in mathematically precise terms with physical counterparts. In particular, a widely applicable approach is provided by the Landau theory of phase transitions – whose mathematical counterpart resides in the realm of equivariant singularity and bifurcation theory. In Landau theory, the state of a system is described by a finite-dimensional variable (the "order parameter"), and physical states correspond to minima of a potential, invariant under a group.

In this article we describe the basics of symmetry breaking analysis for systems described by a symmetric polynomial; in particular, we discuss generic symmetry breakings, that is, those determined by the symmetry properties themselves and independent of the details of the polynomial describing a concrete system. We also discuss how the plethora of invariant polynomials can be to some extent reduced by means of changes of coordinates, that is, how one can reduce to consider certain types of polynomials with no loss of generality. Finally, we will give some indications on extension of this theory, that is, on how one deals with symmetry breakings for more general groups and/or more general physical systems.

## Basic Notions

### Finite Groups

A finite group $(G, \circ)$ is a finite set $G$ of elements $\{g_0, \ldots, g_N\}$ equipped with a composition law $\circ$, and such that the following conditions hold:

1. for all $g, h \in G$ the composition $g \circ h$ belongs to $G$, that is, $g \circ h \in G$;
2. the composition is associative, that is, $(g \circ h) \circ k = g \circ (h \circ k)$ for all $g, h, k \in G$;
3. there is an element in $G$ – which we will denote as $e$ – which is the identity for the action of $\circ$ on $G$, that is, $e \circ g = g = g \circ e$ for all $g \in G$; and
4. for each $g \in G$ there is an element $g^{-1}$ which is the inverse of $g$, that is, $g^{-1} \circ g = e = g \circ g^{-1}$.

In the following, we omit the symbol $\circ$, that is, we write $gh$ to mean $g \circ h$. Similarly, we usually write simply $G$ for the group, rather than $(G, \circ)$.

Given a subset $H \subseteq G$, this is a subgroup of $(G, \circ)$ if $(H, \circ)$ satisfies the group axioms (1)–(4) above. Note that this implies that $e \in H$ whenever $H$ is a subgroup, and $\{e\}$ is a subgroup. Subgroups not coinciding with the whole $G$ and with $\{e\}$ are said to be "proper."

Given two elements $g, h$ we say that $ghg^{-1}$ is the conjugate of $h$ by $g$. The conjugate of a subgroup $H \subseteq G$ by $g \in G$ is the subgroup of elements conjugated to elements of $H, gHg^{-1} = \{(ghg^{-1}), h \in H\}$.

### Group Action

In physics, one is usually interested in a realization of an abstract group as a group of transformations in some set $X$; in physical applications, this is usually a (possibly, function) space or a manifold, and we refer to elements of $X$ as "points." That is, there is a map $\rho: G \mapsto \mathrm{End}(X)$ from $G$ to the group of endomorphisms of $X$, such to preserve the composition law:

$$\rho(g) \cdot \rho(h) = \rho(g \circ h) \quad \forall g, h \in G$$

In this case, we say that we have a "representation" of the abstract group $G$ acting in the "carrier" space or manifold $X$; we also say that $X$ is a $G$-space or $G$-manifold. We often denote by the same letter the abstract element and its representation, that is, write simply $g$ for $\rho(g)$ and $G$ for $\rho(G)$. (In many physically relevant cases, but not necessarily, $X$ has a linear structure and we consider linear endomorphisms. In this case, we sometimes write $T_g$ for the linear operator representing $g$.)

If $x \in X$ is a point in $X$, the $G$-orbit $G(x)$ is the set of points to which $x$ is mapped under $G$, that is,

$$G(x) = \{y \in X: \; y = gx, g \in G\} \subseteq X$$

Belonging to the same orbit is obviously an equivalence relation, and partitions $X$ into equivalence classes. The "orbit space" for the $G$ action on $X$, also denoted as $\Omega = X/G$, is the set of these equivalence classes. It corresponds, in physical terms, to considering $X$ modulo identification of elements related by the group action.

For any point $x \in X$, the "isotropy (sub)group" $G_x$ is the set of elements leaving $x$ fixed,

$$G_x = \{g \in G: \; gx = x\} \subseteq G$$

Points on the same $G$-orbit have conjugated isotropy subgroups: indeed, $y = gx$ implies immediately that $G_y = gG_xg^{-1}$.

When a topology is defined on $X$, the problem arises if the $G$-action preserves it; if this is the case, we say that the $G$-action is "regular." In the case of a compact Lie group (and *a fortiori* for a finite group) we are guaranteed the action is regular. (A physically relevant example of nonregular action is provided by the irrational flow on a torus. In this case $G = R$, realized as the time $t$ irrational flow on the torus $X = \boldsymbol{T}^k$.)

### Spontaneous Symmetry Breaking

Let us now consider the case of physical systems whose state is described by a point $x$ in the $G$-space or $G$-manifold $X$, with $G$ a group acting by smooth mappings $g: X \to X$. In physical problems, $G$ quite often acts by linear and orthogonal transformations. (If this is not the case, the Palais–Mostow theorem guarantees that, for suitable groups (including in particular the finite ones) we can reduce to this case upon embedding $X$ into a suitably larger carrier space $Y$.)

Usually, $G$ represents physical equivalence of states, and $G$-orbits are collections of physically equivalent states. A point which is $G$-invariant, that is, such that $G_x = G$, is called "symmetric" for short.

Let $\Phi$ be a scalar function (potential) defined on $X, \Phi: X \to \boldsymbol{R}$, possibly depending on some parameter $\mu$, such that the physical state corresponds to critical points – usually the (local) minima – of $\Phi$.

A concrete example is provided by the case where $\Phi$ is the Gibbs free energy; more generally, this is the framework met in the Landau theory of phase transitions (Landau 1937, Landau and Lifshitz 1958).

We are interested in the case where $\Phi$ is invariant under the group action, or briefly $G$-invariant, that is, where

$$\Phi(gx) = \Phi(x) \quad \forall x \in X, \; \forall g \in G \qquad [1]$$

A critical point $x$ such that $G_x = G$ is a symmetrical critical point. If $G_x$ is strictly smaller than $G$, then $x$ is a symmetry-breaking critical point.

If a physical system corresponds to a nonsymmetric critical point, we have a spontaneous symmetry breaking: albeit the physical laws (the potential function $\Phi$) are symmetric, the physical state (the critical point for $\Phi$) breaks the symmetry and chooses one of the $G$-equivalent critical points.

It follows from [1] that the gradient of $\Phi$ is covariant under $G$. If $y = g(x)$, then the differential $(Dg)$ of the map $g: X \to X$ is a linear map between the corresponding tangent spaces, $(Dg): T_xX \to T_yX$. The covariance amounts, with $\eta$ the Riemannian metric in $X$, to $(\eta^{ij}\partial_j\Phi)(gx) = [(Dg)^i_k \eta^{km}\partial_m\Phi](x)$; this is also written compactly, with obvious notation, as

$$(\nabla\Phi)(gx) = (Dg)[(\nabla\Phi)(x)] \qquad [2]$$

(in the case of euclidean spaces $(\eta = \delta)$ and linear actions described by matrices $T_g$, the covariance condition reduces to $(\nabla\Phi)^i(T_gx) = (T_g)^i_j \; [(\nabla\Phi)^j(x)]$). As $(Dg)$ is a linear map, $(\nabla\Phi)(x) = 0$ implies the vanishing of $\nabla\Phi$ at all points on the $G$-orbit of $x$.

We conclude that critical points of a $G$-invariant potential come in $G$-orbits: if $x$ is a critical point for

$\Phi$, then each $y \in G(x)$ is also a critical point for $\Phi$. We speak therefore of critical orbits for $\Phi$.

It is thus possible (thanks to the regularity of the $G$-action), and actually convenient, to study spontaneous symmetry breaking in the orbit space $\Omega = X/G$ rather than in the carrier manifold $X$ (Michel 1971).

If $G$ describes physical equivalence, physical states whose symmetries are $G$-conjugated should be seen as physically equivalent. An equivalence class of isotropy types under conjugation will be said to be a symmetry type. We are thus interested, given a $G$-invariant polynomial $\Phi$, to know the symmetry types of its critical points. We denote symmetry types as $[H] = \{gHg^{-1}\}$, and say that $[H] < [K]$ if a group conjugated to $H$ is strictly contained in a group conjugated to $K$.

As we have seen, points on the same $G$-orbit have the same symmetry type. On the other hand, points on different $G$-orbits can have the same isotropy type (e.g., for the standard action of $O(n)$ in $\boldsymbol{R}^n$, all collinear nonzero points will have the same isotropy subgroup but will lie on distinct group orbits).

## G-Invariant Polynomials

Consider a finite group $G$ acting in $X$. (Many of the notions and results mentioned in this section have a much wider range of applicability.) We look at the ring of $G$-invariant scalar polynomials in $x^1, \ldots, x^n$.

By the Hilbert basis theorem, there is a set $\{J_1(x), \ldots, J_k(x)\}$ of $G$-invariant homogeneous polynomials of degrees $\{d_1, \ldots, d_k\}$ such that any $G$-invariant polynomial $\Phi(x)$ can be written as a polynomial in the $\{J_1, \ldots, J_k\}$, that is,

$$\Phi(x) = \Psi[J_1(x), \ldots, J_k(x)] \qquad [3]$$

with $\Psi$ a polynomial. (A similar theorem holds for smooth functions.)

The algebra of $G$-invariant polynomials is finitely generated, that is, we can choose $k$ finite. When the $J_a$ are chosen so that none of them can be written as a polynomial of the others and $r$ has the smallest possible value (this value depends on $G$), we say that they are a minimal integrity basis (MIB). (Note that some of the $J_a$ could be written as nonpolynomial functions of the others, and the $J_\alpha$ could satisfy polynomial relations. For example, consider the group $Z_2$ acting in $\boldsymbol{R}^2$ via $g: (x, y) \to (-x, -y)$; an MIB is made of $J_1(x, y) = x^2, J_2(x, y) = y^2$, and $J_3(x, y) = xy$. None of these can be written as a polynomial function of the others, but $J_1 J_2 = J_3^2$.) In this case, we say that the $\{J_a\}$ are a set of basic invariants for $G$. There is obviously some arbitrarness in the choice of the $J_a$ in an MIB, but the degrees $\{d_1, \ldots, d_k\}$ of $\{J_1, \ldots, J_k\}$ are fixed by $G$. (In mathematical terms, they are determined through the Poincaré series of the graded algebra $P_G$ of $G$-invariant polynomials.)

We will henceforth assume that we have chosen an MIB, with elements $\{J_1, \ldots, J_k\}$ of degrees $\{d_1, \ldots, d_k\}$ in $x$, say with $d_1 \leq d_2 \leq \cdots \leq d_k$.

When the elements of an MIB for $G$ are algebraically independent, we say that the MIB is regular; if $G$ admits a regular MIB we say that $G$ is coregular.

An algebraic relation between elements $J_\alpha$ of the MIB is said to be a relation of the first kind. The algebraic relations among the $J$ are a set of polynomials in $\{J_1, \ldots, J_r\}$, which are identically zero when seen as polynomials in $x$. If there are algebraic relations among these, they are called relations of the second kind, and so on. A theorem by Hilbert guarantees that the chain of relations has finite maximal length. (This is the homological dimension of the graded algebra $P_G$ mentioned above.)

In the following, we will consider a matrix built with the gradients of basic invariants, the $\mathcal{P}$-matrix (Sartori). This is defined as

$$\mathcal{P}_{ih}(x) := \langle \nabla J_i(x), \nabla J_h(x) \rangle \qquad [4]$$

with $\langle .\,, . \rangle$ the scalar product in $T_*X$.

The gradient of an invariant is necessarily a covariant quantity; the scalar product of two covariant quantities is an invariant one, and thus can be expressed again in terms of the basic invariants. Thus, the $\mathcal{P}$-matrix can always be written in terms of the basic invariants themselves.

## Geometry of Group Action

The use of an MIB allows to introduce a map $\boldsymbol{J}: x \to \{J_1(x), \ldots, J_k(x)\}$ from $X$ to a subset $P$ of $\boldsymbol{R}^k$. If the MIB is regular, $P = \boldsymbol{R}^k$, while if the $J_i$ satisfy some relation then $P \subset \boldsymbol{R}^k$ is the submanifold satisfying the corresponding relations. The manifold $P$ is isomorphic to the orbit space $\Omega = X/G$ (the isomorphism being realized by the $\boldsymbol{J}$ map) and provides a more convenient framework to study $\Omega$.

As mentioned above, on physical terms we are mainly interested in the orbit space up to equivalence of symmetry type. The set of points in $X$ (of orbits in $\Omega$) with the same symmetry type will be called a $G$-stratum in $X$ (a $G$-stratum in $\Omega$); the $G$-stratum of the point $x$ will be denoted as $\sigma(x) \subset X$ (the $G$-stratum of the orbit $\omega$ as $\Sigma(\omega) \subset \Omega$). (The notion of stratum was introduced by Whitney in topology; a stratified manifold is a set which can be decomposed as the disjoint union of smooth

manifolds of different dimensions, the topological (or Whitney) strata: $M = \bigcup M^k$, with $M^k \subset \partial M^j$ for all $k < j$.)

It results that the $G$-stratification is compatible with the topological stratification. Indeed, $P$ is a semialgebraic (i.e., it is defined by algebraic equalities and inequalities) stratified manifold in $\boldsymbol{R}^k$; the image of any $G$-stratum in $\Omega$ belongs to a single topological stratum in $P$, and topological strata in $P$ are the union of images of $G$-strata in $\Omega$.

Moreover, the subgroup relations correspond to bordering relations between $G$-strata: if $[G_x] < [G_y]$, then $\sigma(y) \in \partial\sigma(x)$ and (with $\omega_x$ the orbit of $x$) $\Sigma(\omega_y) \in \partial\Sigma(\omega_x)$.

There is a stratum, called the principal stratum $\sigma_0$, which corresponds to minimal isotropy, open and dense in $X$; similarly, the principal stratum $\Sigma_0$ is open and dense in $\Omega$.

## Landau Polynomial

In the Landau (1937) theory of phase transitions, the state of the system under study is described by a $G$-invariant polynomial $\Phi : X \to \boldsymbol{R}$ having a critical point in the origin, with at least some of its coefficients – in particular those controlling the stability of the zero critical point – depending on external control parameters (usually, $X = \boldsymbol{R}^n$ and $G \subseteq \mathrm{O}(n)$; in particular, in solid-state physics $G$ is a crystallographic group). This should be chosen as the most general $G$-invariant polynomial of the lowest degree $\ell$ sufficient to ensure termodynamic stability; in mathematical terms, this amounts to the requirement that there is some open set $\mathcal{B}$ containing the origin and such that – for all values of the control parameters – $\nabla\Phi$ points inwards at all points of $\partial\mathcal{B}$ (i.e., $\mathcal{B}$ is invariant under the gradient flow of $\Phi$). If the polynomials in the MIB are of degree $d_1 \leq d_2 \cdots \leq d_r$, then usually $\ell = 2d_r$.

The $G$-invariance of $\Phi$ and the results recalled above mean that we can always write it in terms of the polynomials in an MIB for $G$ as in [3], $\Phi(x) = \Psi[\boldsymbol{J}(x)]$.

The discussion of previous sections shows that we can study symmetry breakings for $\Phi : X \to \boldsymbol{R}$ by studying critical points of $\Psi : P \to \boldsymbol{R}$; in other words, Landau theory can be worked out in the G-orbit space $\Omega := M/G$. The polynomial $\Psi$ – providing a representation of the Landau polynomial in the orbit space – will also be called Landau–Michel polynomial. (Louis Michel (1923–1999) pioneered the use of orbit space techniques in physics and nonlinear dynamics, originally motivated by the study of hadronic interactions.)

In this way, the evaluation of the map $\Phi : X \to \boldsymbol{R}$ is, in principle, substituted by evaluation of two maps, $\boldsymbol{J} : X \to P$ and $\Psi : P \to \boldsymbol{R}$. However, if, as in Landau theory, we have to consider the most general $G$-invariant polynomial on $X$, we can just consider the most general polynomial on $P$.

## Critical Points of the Landau Polynomial and Geometry of Orbit Space

The $G$-invariance has consequences on the critical points of $\Phi$. We have already seen one such consequence: critical points come in $G$-orbits.

However, this is not all. Indeed, $G$-invariance enforces the presence of a certain set $\chi(G) \in X$ of critical points, and conversely if we look for points which are critical under any $G$-invariant potential, these are precisely the points in $\chi(G)$; the critical points on $\chi(G)$ correspond to critical orbits which we call principal critical orbits.

The set $\chi(G)$ can be determined on the basis of the geometry of the $G$-action. (A trivial example is provided by $X = \boldsymbol{R}$ and $G = Z_2$ acting via $g : x \to -x$; any even function has a critical point in zero, and albeit even functions can, and in general will, have nonzero critical points, this is the only critical point common to all the even functions.) Indeed (Michel 1971): *an orbit $\omega$ is a principal critical orbit if and only if it is isolated in its stratum*.

For the linear orthogonal group actions in $\boldsymbol{R}^n$ often occurring in physics, no nonzero point or orbit can be isolated in its stratum. However, we can quotient out the radial degeneracy and work on $X = S^{n-1} \subset \boldsymbol{R}^n$. In this case, a $G$-orbit $\omega_1$ in $S^{n-1}$ which is isolated in its stratum corresponds to a one-dimensional family $\{\omega_r\}$ of $G$-orbits in $\boldsymbol{R}^n$ (call $X_0$ the corresponding submanifold in $X$); the gradient of $\Phi$ at $x \in X_0$ points along $T_xX_0$. We can thus reduce to consider the restriction $\Phi_0$ of the potential $\Phi$ to $X_0$. (See also the reduction lemma of Golubitsky and Stewart in this context.)

Correspondingly, if $P_0 \subset P$ is the submanifold in $P$ image of $X_0$, that is, $P_0 = \boldsymbol{J}(X_0)$, we can reduce to consider the restriction $\Psi_0$ of $\Psi$ to $P_0$.

As these become one-dimensional problems, general results are available. In particular, one can provide general conditions ensuring the existence of one-dimensional branches of symmetry-breaking solutions bifurcating from zero along any such $X_0$ or $P_0$; this is also known as the equivariant branching lemma of Cicogna and Vanderbauwhede.

## Reduction of the Landau Potential

In realistic problems, $\Phi$ quickly becomes extremely complicated, that is, it includes a high number of terms and therefore of coefficients. A thorough study of different symmetry-breaking patterns, that is, of the symmetry type of minima of $\Phi$ for different values of these coefficients and of the external control parameter, is in this case a prohibitive task. It is possible to reduce the generality of the Landau polynomial with no loss of generality for the corresponding physical problem. Indeed, a change of coordinates in the $X$ space will produce a formally different – but obviously equivalent – Landau polynomial; it is convenient to use coordinates in which the Landau polynomial is simpler.

A systematic and algorithmic reduction procedure – based on perturbative expansion near the origin – is well known in dynamical systems theory (Poincaré–Birkhoff normal forms), and can be adapted to the reduction of Landau polynomials. (An alternative and more general – but also much more demanding – approach is provided by the spectral sequence approach, also originating in normal-form theory.)

We work near the origin, so that we can assume $X = \boldsymbol{R}^n$ (with metric $\eta$), and for simplicity we also take the case where $G$ acts via a linear representation $T_g$. We consider changes of coordinates of the (Poincaré) form

$$x^i = y^i + h^i(y) \qquad [5]$$

generated by a $G$-invariant function $H$: $h^i(y) = \eta^{ij}$ $(\partial H(y)/\partial y^j)$; this guarantees that [5] preserves the $G$-invariance of $\Phi$. The action of [5] on $\Phi$ can be read from its action on the basic invariants $J_a$. It results

$$\begin{aligned} J_a(x) &= J_a(y) + (\delta J_a)(y) \\ \delta J_a &:= \mathcal{P}_{ab}(\partial H/\partial J_b) \end{aligned} \qquad [6]$$

Let us now consider the reduction of an invariant polynomial $\Phi(x) = \Psi(J)$. We write $D_\alpha := \partial/\partial J_\alpha$, and understand that summation over repeated indices is implied. In general,

$$\begin{aligned} \Psi(J) &\to \Psi(J + \delta J) \\ &= \Psi(J) + \sum_{\alpha=1}^{r} \frac{\partial \Psi(J)}{\partial J_\alpha} \delta J_\alpha + \cdots \end{aligned}$$

where the ellipsis means higher-order terms.

Disregarding higher-order terms and using [6] and [4], we get

$$\delta\Psi = \frac{\partial \Psi}{\partial J_\alpha} \mathcal{P}_{\alpha\beta} \frac{\partial H}{\partial J_\beta} \equiv (D_\alpha \Psi)\mathcal{P}_{\alpha\beta}(D_\beta H) \qquad [7]$$

We expand $\Phi$ as a sum of homogeneous polynomials, and write $\Phi(x) = \sum_{k=0}^{\ell} \Phi_k(x)$, where $\Phi_k(ax) = a^{k+1}\Phi_k(x)$. Also, write $\Psi = \sum_k \Psi_k$, where $\Phi_k(x) := \Psi_k[J(x)]$.

It results that under a change of coordinates [5] generated by $H = H_m$ homogeneous of degree $m+1$, the terms $\Psi_k$ with $k \le m$ are not changed, while the terms $\Psi_{m+p}$ change according to

$$\begin{aligned} \Psi_{m+p} &\to \Psi_{m+p} \\ &= \Psi_{m+p} + (D_\alpha \Psi_p)\mathcal{P}_{\alpha\beta}(D_\beta H_m) + \cdots \end{aligned} \qquad [8]$$

We can then operate sequentially with $H_m$ of degree $3, 4, \ldots$; at each stage (generator $H_m$), we are not affecting the terms $\Psi_k$ with $k \le m$. Moreover, we can just consider [8], as higher-order terms are generic and will be taken care of in subsequent steps. (This procedure requires to determine suitable generating functions $H_m$; these are obtained as solutions to homological equations.)

In the above, we disregarded the dependence on the control parameters, such as temperature, pressure, magnetic field, etc; that is, we implicitly considered fixed values for these. However, they have to change for a phase transition to take place. If we consider a full range of values – including in particular the critical ones – for the control parameters, say $\lambda \in \Lambda$, we should take care that the concerned quantities and operators are nonsingular uniformly in $\Lambda$.

This leads to reduction criteria for the Landau and Landau–Michel polynomials (Gufan). Define, for $i = 1, \ldots, k$ the quantities $U_i(J_1, \ldots, J_k) := (\partial F/\partial J_s)\mathcal{P}_{si}$.

### Reduction Criterion

For $\Phi(x) = \Psi(J_1, \ldots, J_k) : \boldsymbol{R}^n \to \boldsymbol{R}$ a $G$-invariant potential depending on physical parameters $\lambda \in \Lambda$, there is a sequence of Poincaré changes of coordinates such that $\Phi$ is expressed in the new coordinates $y$ as $\hat{\Phi}(y) = \hat{\Psi}(J)$, where terms which can be written (up to higher-order terms) uniformly in $\Lambda$ as $\sum_{\alpha=1}^{k} Q_\alpha(J_1, \ldots, J_k)$ $U_\alpha(J_1, \ldots, J_k)$, with $Q_\alpha$ polynomials in $J_1, \ldots, J_k$ satisfying the compatibility condition $(\partial Q_\beta/\partial J_\alpha) = (\partial Q_\alpha/\partial J_\beta)$, are not present in $\hat{\Psi}$.

## Nonstationary and Nonvariational Problems

So far we have considered stationary physical states. In some cases, one is not satisfied with such a description, and wants to study time evolution. A model framework for this is provided by the Ginzburg–Landau equation

$$\dot{x} = f(x) \qquad [9]$$

where $f = \eta(\nabla\Phi) : X \to TX$ (see above for notation). In this case, $G$-invariance of $\Phi$ implies equivariance

of [9]. More generally, we can consider [9] for an equivariant smooth $f$ (not necessarily a gradient), that is, $f^i(gx) = (Dg)^i_j f^j(x)$.

In this case, one shows that

$$f(x) \in T_x\sigma(x) \qquad [10]$$

so that closures of $G$-strata are dynamically invariant, and the dynamics can be reduced to them. This is of special interest for the "most singular" strata, that is, those of lower dimension. The reduction lemma and the equivariant branching lemma mentioned above also hold (and were originally formulated) in this context.

The relation [10] also implies that one can project the dynamics [9] in $X$ to a smooth dynamics $\dot{p} = F(p)$ in the orbit space; this satisfies $F[\boldsymbol{J}(x)] = (D\boldsymbol{J})[f(x)]$. In the gradient case, this (together with initial conditions) embodies the full dynamics in $X$, while in the generic case one loses all information about motions along group orbits (note that these correspond to phonon modes).

An orbit $\omega$ isolated in its stratum is still an orbit of fixed points for any $G$-equivariant dynamics in $X$ in the gradient case, while in the generic case it corresponds to a fixed point for $F$ and to relative equilibria (dynamical orbits which belong to a single group orbit) in $X$. In this case, time averages of physical quantities can be $G$-invariant for nontrivial relative equilibria.

## Extensions and Physical Applications

We have discussed finite group symmetry breaking and focused on polynomial potentials (which can be thought of as Taylor expansions around critical points). For nonfinite groups, and in particular noncompact ones, the situation can be considerably more complicated.

1. An extension of the theory sketched here is provided by Palais' theory, and in particular by his "symmetric criticality principle," which applies in Hilbert or Banach spaces of sections of a fiber bundle satisfying certain conditions. This is especially relevant in connection with field theory and gauge groups.
2. We focused on the situation discussed in classical physics. Finite group symmetry breaking is of course also relevant in quantum mechanics; this is discussed, for example, in the classical books by Weyl (1931) and Wigner (1959), and in the review by Michel *et al.* (2004).
3. One speaks of "explicit symmetry breaking" when a nonsymmetric perturbation is introduced in a symmetric problem. In the Hamiltonian case (or in the Lagrangian one for Noether symmetries), Hamiltonian symmetries correspond to conserved quantities, and nonsymmetric perturbations make these become approximate constants of motion.
4. The symmetry of differential equations – as well as symmetric and symmetry-breaking solutions for symmetric equations – can be studied in general mathematical terms (see, e.g., Olver (1986)).
5. Physical applications of the theory discussed here abound in the literature, in particular through the Landau theory of phase transitions. A number of these, together with a deeper discussion of the underlying theory, is given in the monumental review paper by Michel *et al.* (2004).

*See also:* Central Manifolds, Normal Forms; Compact Groups and Their Representations; Electroweak Theory; Finite Group Symmetry Breaking; Phase Transitions in Continuous Systems; Quasiperiodic Systems; Symmetry and Symmetry Breaking in Dynamical Systems; Symmetry Breaking in Field Theory.

## Further Reading


Abud M and Sartori G (1983) The geometry of spontaneous symmetry breaking. *Annals of Physics (NY)* 150: 307–372.

Gaeta G (1990) Bifurcation and symmetry breaking. *Physics Reports* 189: 1–87.

Gaeta G (2004) Lie–Poincaré transformations and a reduction criterion in Landau theory. *Annals of Physics (NY)* 312: 511–540.

Golubitsky M, Stewart I, and Schaeffer DG (1988) *Singularities and Groups in Bifurcation Theory*. New York: Springer.

Landau LD (1937) On the theory of phase transitions I & II. *Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki* 7: 19 and 627.

Landau LD and Lifshitz EM (1958) *Statistical Physics*. London: Pergamon.

Michel L (1971) Points critiques des fonctions invariantes sur une G-varieté. *Comptes Rendus de l'Academie des Sciences de Paris* 272: 433–436.

Michel L (1980) Symmetry, defects, and broken symmetry. *Reviews of Modern Physics* 52: 617–650.

Michel L, Kim JS, Zak J, and Zhilinskii B (2004) Symmetry, invariants, topology. *Physics Reports* 341: 1–395.

Olver PJ (1986) *Applications of Lie Groups to Differential Equations*. Berlin: Springer.

Palais RS (1979) The principle of symmetric criticality. *Communications in Mathematical Physics* 69: 19–30.

Sartori G (2002) Geometric invariant theory in a model-independent analysis of spontaneous symmetry and supersymmetry breaking. *Acta Applicandae Mathematicae* 70: 183–207.

Toledano JC and Toledano P (1987) *The Landau Theory of Phase Transitions*. Singapore: World Scientific.

Weyl H (1931) *The Theory of Groups and Quantum Mechanics*. New York: Dover.

Wigner EP (1959) *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra*. New York: Academic Press.

# Finite Weyl Systems

**D-M Schlingemann**, Technical University of Braunschweig, Braunschweig, Germany



## Introduction

Finite Weyl systems have their applications in various branches of quantum information theory. They are helpful to tame the growth of complexity for a large class of quantum systems: a key discrepancy between classical and quantum systems is the difference in the growth of complexity as one goes to larger and larger systems. This is encountered by simulating a quantum spin system on a computer, for example, with the aim to determine the ground state of a solid-state model of magnetism. For a model of $N$ classical spins, this involves checking the energy for $2^N$ different configurations, but for a model with quantum spins it requires the solution of an eigenvalue equation in a Hilbert space of dimension $2^N$, which is a vastly more difficult problem for large $N$. For a three-dimensional lattice, three sites each way ($N = 27$), this is a problem in $10^8$ dimensions, and lattice size 4 leads to utterly untractable $10^{19}$ dimensions.

It is therefore highly desirable to find ways of treating at least some aspects of large, complex quantum systems without actually having to write out state vectors component by component. States which are invariant under a suitable discrete abelian symmetry group satisfy this condition. They can be characterized by simple combinatorial data, which do not grow exponentially with the system size $N$. At the same time, the class of these so-called stabilizer states is sufficiently complex to capture some of the key features needed for computation, especially the quantum correlation (entanglement) between subsystems. They have also been shown to be sufficient to generate large quantum error correcting codes.

A further motivation for finite Weyl systems is directly based on constructing quantum error correcting codes from classical coding procedures (*see* Quantum Error Correction and Fault Tolerance). The "quantization" technique which is used there naturally leads to the structure of finite Weyl systems.

Finite Weyl systems precisely represent quantum versions of discrete abelian symmetry groups. It is a standard procedure to build the quantum version of a symmetry group by an appropriate central extension, or equivalently, to study all its projective representations: the composition of two symmetry transformations is only preserved up to a phase on the representation Hilbert space. The unitary operators which represent the symmetry transformations are called Weyl operators.

The simplest and most prominent example for a finite Weyl system is given by the three Pauli matrices and the identity. These four unitary operators build a projective representation of the symmetry group of binary vectors $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$, where the group law is the addition modulo two. The null-vector $(0,0)$ corresponds to the identity, the vector $(0,1)$ is assigned to $X$, $(1,0)$ corresponds to $Z$, and $(1,1)$ is mapped to $\mathrm{i}Y$. It is not difficult to verify that the product of two Pauli operators preserves the addition of binary vectors up to a phase.

Discrete Weyl systems are deeply related to symplectic geometry for vector spaces over finite fields. The additive structure of the vector space is the underlying abelian symmetry group. The exchange of two Weyl operators within a product produces a phase that is the exponential of an antisymmetric bilinear form, as it is explained in the next section. For irreducible Weyl systems, this antisymmetric form must be symplectic because the Weyl operators generate a full matrix algebra. In particular, this requires that the dimension of the underlying vector space is even. The Pauli matrices are also an example for this more special structure: the binary vectors $(p,q)_{p,q=0,1}$ are a two-dimensional vector space over the field with two elements $\{0,1\}$. The commutation relations for Pauli operators imply that the symplectic form can be evaluated for two binary vectors $(p,q),(p',q')$ according to $pq' - qp' \bmod 2$. It is apparent to interpret the binary vectors $(p,q)$ as points in a discrete phase space, where the first entry corresponds to the momentum and the second to the position. In view of this, discrete Weyl systems serve as a finite-dimensional analog of the canonical commutation relations.

For the generic situation in quantum information theory, an irreducible Weyl system is represented on the Hilbert space describing a system of several single particles. Stabilizer states are left unchanged under the action of a so-called isotropic subgroup which consists of mutually commuting Weyl operators: this kind of invariance is precisely the type of constraint that reduces the complexity for the parametrization of the state. For an efficient description of such states, there are combinatorial techniques available e.g., graph theory.

Operations that preserve the class of stabilizer states (for a particular symmetry group) must be covariant with respect to this symmetry. These operations are called Clifford channels which have far-reaching applications in the theory of quantum error correction. They also allow to take classical coding procedures and turn them into quantum codes: on the classical level, the encoding operation acts on classical phase space as a linear map (additive code). Up to a choice of phases, this induces a quantum channel that preserves the structure of Weyl systems. These codes are called stabilizer codes and have been investigated by many authors (Calderbank *et al.* 1997, Cleve and Gottesman 1996, 1997) (*see* Quantum Error Correction and Fault Tolerance). In particular, the first quantum error correcting codes belong to this class.

This article is organized as follows. In the next section, the basic mathematical notions are provided, like projective representations, Weyl systems, and irreducibility. Moreover, statements on the main structure of Weyl systems are presented. Next, the notion of Weyl covariant channels (Clifford channels) is introduced and their basic properties are stated. In particular, stronger results for the reversible case are given. The relation between symplectic geometry and reversible Clifford operations on finite Weyl systems is explained. Results on the general structure of stabilizer states and stabilizer codes are given in the penultimate section. Finally, the representation of stabilizer codes in terms of graphs is described.

## Finite Weyl Systems

A projective representation of a group $\Xi$ assigns to each group element $\xi$ a unitary operator $\boldsymbol{w}(\xi)$ on a Hilbert space $\mathcal{H}$ such that the group law is preserved up to a phase, that is, the relation

$$\boldsymbol{w}(\xi_1+\xi_2)=f(\xi_1,\xi_2)\boldsymbol{w}(\xi_1)\boldsymbol{w}(\xi_2) \qquad [1]$$

is fulfilled for a phase-valued function $f$ on $\Xi^2$. In the following, we denote a projective representation by a triple $(\boldsymbol{w},f,\mathcal{H})$. A finite Weyl system is a projective representation of a finite abelian group. The operators $\boldsymbol{w}(\xi)$ are called Weyl operators and the function $f$ is called the factor system. We refer to the work by Zmud (1971, 1972) for an analysis of projective representations for general abelian groups.

The Weyl algebra $\mathfrak{A}(\boldsymbol{w},f,\mathcal{H})$ associated with a Weyl system $(\boldsymbol{w},f,\mathcal{H})$ is the smallest norm-closed subalgebra in the space of bounded operators $\mathcal{B}(\mathcal{H})$ which contains all Weyl operators. If the Weyl algebra coincides with the algebra of all bounded operators, then the Weyl system is called irreducible. This is equivalent to the fact that each operator that commutes with all Weyl operators must be a multiple of the identity.

In order to analyze the properties of factor systems systematically, we introduce here a few pieces of the cohomology theory of groups. For each positive integer $k=1,2,3,\ldots$ we introduce the abelian group $C^k(\Xi)$ of $k$-cochains which consists of all phase-valued functions on $\Xi^k$. The product and the inverse of $k$-cochains is defined pointwise. Factor systems are special 2-cochains. Namely, if we consider a Weyl system $(\boldsymbol{w},f,\mathcal{H})$, then associativity implies that the so-called 2-cocycle condition,

$$f(\xi_1+\xi_2,\xi_3)f(\xi_2,\xi_3)^{-1}f(\xi_1,\xi_2+\xi_3)^{-1}f(\xi_1,\xi_2)=1 \qquad [2]$$

holds. This property can also be expressed by a coboundary map $\mathfrak{d}$ which is a group homomorphism from $k$-cochains to $(k+1)$-cochains. We consider here the action of the coboundary map on a 1-cochain $\varphi$ and a 2-cochain $f$:

$$(\mathfrak{d}\varphi)(\xi_1,\xi_2):=\varphi(\xi_1+\xi_2)\varphi(\xi_1)^{-1}\varphi(\xi_2)^{-1} \qquad [3]$$

$$(\mathfrak{d}f)(\xi_1,\xi_2,\xi_3):=f(\xi_1+\xi_2,\xi_3)f(\xi_2,\xi_3)^{-1}\times f(\xi_1,\xi_2+\xi_3)^{-1}f(\xi_1,\xi_2) \qquad [4]$$

The group of 2-cocycles $Z^2(\Xi)$ consists of all 2-cochains $f$ with $\mathfrak{d}f=1$ and the group of all 2-coboundaries $B^2(\Xi)$ contains all 2-cochains of the form $f=\mathfrak{d}\varphi$. The 2-fold concatenation of the coboundary map is the trivial homomorphism $\mathfrak{d}\circ\mathfrak{d}=1$, which implies that each 2-coboundary is a 2-cocycle. The converse is in general not the case and the 2-cohomology group $H^2(\Xi):=Z^2(\Xi)/B^2(\Xi)$ is nontrivial.

The Zmud (1971, 1972) analysis shows that the set of Weyl systems are characterized by elements of the 2-cohomology $H^2(\Xi)$. The multiplication of a Weyl system $(\boldsymbol{w},f,\mathcal{H})$ by a 1-cochain $\varphi$ yields a new family of Weyl operators $(\varphi\boldsymbol{w})(\xi)=\varphi(\xi)\boldsymbol{w}(\xi)$. The 2-cocycle $f$ is altered by the multiplication of the 2-coboundary $\mathfrak{d}\varphi$ and the new Weyl system is given by $(\varphi\boldsymbol{w},\mathfrak{d}\varphi f,\mathcal{H})$. This kind of transformation does not change the cohomology class of the factor system and the corresponding Weyl algebras coincide: $\mathfrak{A}(\boldsymbol{w},f,\mathcal{H})=\mathfrak{A}(\varphi\boldsymbol{w},\mathfrak{d}\varphi f,\mathcal{H})$. Thus, the fundamental properties of a Weyl system only depend on the cohomology class of the factor system. In particular, if the factor system $f=\mathfrak{d}\varphi$ is a 2-coboundary, then we can trivialize the Weyl system $(\boldsymbol{w},\mathfrak{d}\varphi,\mathcal{H})$ by multiplying the inverse 1-cochain $\varphi^{-1}$ and we obtain a true unitary representation $(\varphi^{-1}\boldsymbol{w},1,\mathcal{H})$. The corresponding Weyl algebra $\mathfrak{A}(\boldsymbol{w},\mathfrak{d}\varphi,\mathcal{H})$ is abelian. The relation between cohomology and Weyl systems

can be made even more precise by the following theorem:

**Theorem 1** (Zmud 1971, 1972). *$\theta$ is the group homomorphism on 2-cochains that exchanges the variables*: $(\theta f)(\xi_1, \xi_2) = f(\xi_2, \xi_1)$.

(i) *The antisymmetric part $f^{-1}(\theta f)$ of a factor system (2-cocycle) is an antisymmetric bicharacter, that is, a group homomorphism in both arguments keeping the other variable fixed.*
(ii) *Each symmetric 2-cocycle $f = \theta f$ is a 2-coboundary $f = \eth\varphi$.*
(iii) *The group of antisymmetric bicharacters on $\Xi$ is isomorphic to the 2-cohomology group $H^2(\Xi)$. For each antisymmetric bicharacter $\sigma$ the corresponding 2-cohomology class is uniquely determined by $\sigma = f^{-1}(\theta f)$ for some representative $f \in Z^2(\Xi)$.*

**Example 2** The following Weyl system describes $n$-quantum digits (in short qudits). The system's Hilbert space is spanned by orthonormal vectors $|a\rangle = |a_1, a_2, \ldots, a_n\rangle$ which are labeled by vectors $a$ of the additive group $\mathbb{F}^n$, where $\mathbb{F} = \mathbb{Z}_d$ is the cyclic field of prime order. A projective representation $(\boldsymbol{w}, \chi, \mathbb{C}^{d^n})$ of the additive group $\mathbb{F}^{2n}$ is given by

$$\boldsymbol{w}(p,q)|a\rangle := \mathrm{e}^{(2\pi\mathrm{i}/d)p^{\mathrm{t}}a}|a+q\rangle \quad [5]$$

where $p^{\mathrm{t}}$ is the transposed vector. The factor system $\chi$ assigns to each pair $(p,q),(p',q')$ the phase

$$\chi(p,q|p',q') := \mathrm{e}^{(2\pi\mathrm{i}/d)p^{\mathrm{t}}q'} \quad [6]$$

The finite vector space $\mathbb{F}^{2n}$ is interpreted as finite phase space with a multiplicative symplectic form $\sigma$. It assigns to a pair of vectors $(p,\, q)$, $(a,\, b)$ the phase

$$\sigma(p,q|a,b) := \mathrm{e}^{(2\pi\mathrm{i}/d)(p^{\mathrm{t}}b - a^{\mathrm{t}}q)} \quad [7]$$

The commutation relation for Weyl operators comprise the symplectic form:

$$\boldsymbol{w}(p,q)\boldsymbol{w}(a,b) = \sigma(a,b|p,q)\boldsymbol{w}(a,b)\boldsymbol{w}(p,q) \quad [8]$$

The $d^{2n}$ Weyl operators $\boldsymbol{w}(p,q)$ are a basis of the algebra of all operators acting on the Hilbert space $\mathbb{C}^{d^n}$, hence $(\boldsymbol{w}, \chi, \mathbb{C}^{d^n})$ is irreducible. In particular, this Weyl system is a nice error basis in the sense of (Klappenecker and Roetteler 2002, 2005). Namely, the Weyl operators form a projective representation, on the one hand, and a unitary basis (Werner 2001) on the other.

For $d = 2$ and $n = 4$, we obtain a system of four qubits and the Weyl operators are tensor products of four Pauli matrices including the identity. For instance, the Weyl operator of the binary vector $(p,q) = (0011, 1010)$ can be expressed in terms of Pauli matrices (see Introduction) as follows:

$$\begin{aligned}\boldsymbol{w}(0011,1010) &= \boldsymbol{w}(0,1) \otimes \mathbb{1} \otimes \boldsymbol{w}(1,1) \otimes \boldsymbol{w}(1,0)\\ &= \boldsymbol{i}X \otimes \mathbb{1} \otimes Y \otimes Z\end{aligned} \quad [9]$$

## Clifford Channels

Weyl systems can be seen as quantized symmetries corresponding to finite abelian groups. In the Heisenberg picture the symmetry transformations act on operators $A \in \mathcal{B}(\mathcal{H})$ of the observable algebra by automorphisms (reversible quantum channels):

$$\mathrm{Ad}[\boldsymbol{w}(\xi)](A) := \boldsymbol{w}(\xi)A\boldsymbol{w}(\xi)^* \quad [10]$$

Since a projective representation preserves the group law up to a phase, the corresponding automorphisms preserve the group law:

$$\mathrm{Ad}[\boldsymbol{w}(\xi)] \circ \mathrm{Ad}[\boldsymbol{w}(\eta)] = \mathrm{Ad}[\boldsymbol{w}(\xi+\eta)] \quad [11]$$

A quantum channel $T$ is called a Clifford channel if it is covariant with respect to Weyl systems $(\boldsymbol{w}_1, f_1, \mathcal{H}_1)$ and $(\boldsymbol{w}_2, f_2, \mathcal{H}_2)$, that is, the intertwiner relation

$$T \circ \mathrm{Ad}[\boldsymbol{w}_2(\xi)] = \mathrm{Ad}[\boldsymbol{w}_1(\xi)] \circ T \quad [12]$$

holds. It is required that the antisymmetric part of the factor systems $f_1$ and $f_2$ coincide, that is, $\sigma = f_1^{-1}\theta f_1 = f_2^{-1}\theta f_2$. We call $(\boldsymbol{w}_1, f_1, \mathcal{H}_1)$ the input and $(\boldsymbol{w}_2, f_2, \mathcal{H}_2)$ the output system. We refer to the article by Scutaru (1979), which is concerned with the general properties of covariant channels.

It is a natural question to ask how Clifford channels act on Weyl operators. As shown by Holevo (n.d.), a Clifford channel maps Weyl operators of the output system to multiples of a Weyl operators of the input system, provided the input system is irreducible.

**Theorem 3** (Holevo (n.d.)). *Let $T$ be a Clifford channel such that the input system $(\boldsymbol{w}_1, f_1, \mathcal{H}_1)$ is irreducible. Then there exists a function $\varphi : \Xi \to \mathbb{C}$ such that*

$$T(\boldsymbol{w}_2(\xi)) = \varphi(\xi)\boldsymbol{w}_1(\xi) \quad [13]$$

*holds for all $\xi \in \Xi$. The function $\varphi$ is of positive type, that is, for all complex functions $f$ on $\Xi$ the inequality*

$$0 \le \sum_{\xi,\eta\in\Xi} \varphi(\xi-\eta)\overline{f(\xi)}f(\eta) \quad [14]$$

*holds. Conversely, if the factor systems $f_1 = f_2$ coincide, then a well-defined channel is determined*

*by [13] for any function $\varphi$ of positive type with $\varphi(0)=1$.*

We apply Theorem 3 to a reversible Clifford channel T. Each output Weyl operator $\boldsymbol{w}_2(\xi)$ is mapped to a multiple of an input Weyl operator

$$T(\boldsymbol{w}_2(\xi)) = \varphi(\xi)\boldsymbol{w}_1(\xi) \qquad [15]$$

where $\varphi$ is phase-valued (a 1-cochain) according to the reversibility of $T$. We focus now on the converse problem: construct all reversible Clifford channels for irreducible Weyl systems that have a common antisymmetric part of the factor system. The following theorem gives a useful characterization of reversible Clifford channels.

**Theorem 4** (Schlingemann and Werner 2001). *If $(\boldsymbol{w}_1, f_1, \mathcal{H}_1)$ and $(\boldsymbol{w}_2, f_2, \mathcal{H}_2)$ are irreducible Weyl systems with $f_1^{-1}(\theta f_1) = f_2^{-1}(\theta f_2)$, then there exists a 1-cochain $\varphi$ with coboundary $\eth\varphi = f_1^{-1} f_2$, and a reversible Clifford channel $T_\varphi$ is determined by*

$$T_\varphi(\boldsymbol{w}_2(\xi)) = \varphi(\xi)\boldsymbol{w}_1(\xi) \qquad [16]$$

*If $\tau$ is a 1-cochain that also satisfies $\eth\tau = f_1^{-1} f_2$, then there exists $\eta \in \Xi$ such that*

$$\tau(\xi) = \sigma(\eta|\xi)\varphi(\xi) \qquad [17]$$

$$T_\tau = \mathrm{Ad}[\boldsymbol{w}_1(\eta)] \circ T_\varphi = T_\varphi \circ \mathrm{Ad}[\boldsymbol{w}_2(\eta)] \qquad [18]$$

*holds. In other words, two irreducible Weyl systems determine a reversible Clifford channel up to a "phase space translation $\eta$."*

We consider the Weyl system $(\boldsymbol{w}, f, \mathcal{H})$ over a discrete phase space $\mathbb{F}^{2n}$, where $\mathbb{F}$ is a finite field of prime order. The group of symplectic transformations $\mathrm{Sp}(n, \mathbb{F})$ consists of all $\mathbb{F}$-linear maps s on the phase space $\mathbb{F}^{2n}$ that preserve the symplectic form $\sigma = f^{-1}\theta f$. A further Weyl system $(\boldsymbol{w} \circ \boldsymbol{s}, f \circ \boldsymbol{s}, \mathcal{H})$ is obtained for each symplectic transformation $\boldsymbol{s}$. Here the factor system $f \circ \boldsymbol{s}$ is defined according to $(f \circ \boldsymbol{s})(\xi, \eta) := f(\boldsymbol{s}\xi, \boldsymbol{s}\eta)$ and the corresponding Weyl operators are $(\boldsymbol{w} \circ \boldsymbol{s})(\xi) = \boldsymbol{w}(\boldsymbol{s}\xi)$. Obviously, the antisymmetric part of the factor system $f \circ \boldsymbol{s}$ is the symplectic form $\sigma \circ \boldsymbol{s} = \sigma$. The following statement is a direct consequence of Theorem 4.

**Corollary 5** *For each symplectic transformation $\boldsymbol{s} \in \mathrm{Sp}(n, \mathbb{F})$ there exists a 1-cochain $\varphi$ with coboundary $\eth\varphi = f^{-1}(f \circ \boldsymbol{s})$ and the corresponding reversible Clifford channel $T_{[\varphi, s]}$ is given by*

$$T_{[\varphi, s]}(\boldsymbol{w}(\xi)) = \varphi(\xi)\boldsymbol{w}(\boldsymbol{s}\xi) \qquad [19]$$

*with $\xi, \eta \in \mathbb{F}^{2n}$.*

**Example 6** We consider a finite field $\mathbb{F}$. To a symmetric matrix $\Gamma \in \mathrm{M}_n(\mathbb{F})$ we associate the symplectic transformation on $\mathbb{F}^{2n}$ that maps a phase space vector $(p, q)$ to $(p - \Gamma q, q)$. This shear transformation is viewed as one elementary step of a discrete dynamics. The quantized version of this dynamics is given by the unitary multiplication operator

$$\boldsymbol{u}(\Gamma)|q\rangle = \zeta_d^{q^{\mathrm{t}}\Gamma q}|q\rangle \qquad [20]$$

with the root of unity $\zeta_d = \exp(\mathrm{i}\pi(d+1)/d)$ for $d \neq 2$ and $\zeta_2 = \mathrm{i}$. The unitary operator $\boldsymbol{u}(\Gamma)$ implements a reversible Clifford operation for the symplectic transformation $(p, q) \mapsto (p - \Gamma q, q)$ since the relation

$$\boldsymbol{u}(\Gamma)\boldsymbol{w}(p, q)\boldsymbol{u}(\Gamma)^* = \zeta_d^{q^{\mathrm{t}}\Gamma q}\boldsymbol{w}(p - \Gamma q, q) \qquad [21]$$

holds. The symmetric matrix $\Gamma$ describes a pattern of two-qudit interactions. This can be visualized by a graph $\Gamma$ whose vertices are the positions $x, y = 1, \ldots, n$. Two vertices $x$, $y$ are connected by an edge if the matrix element $\Gamma_y^x \neq 0$ is nonvanishing. The value of the matrix element $\Gamma_y^x$ is interpreted as the strength of the interaction.

**Example 7** The second type of symplectic transformations, which is relevant here, is determined by an invertible matrix $C \in \mathrm{M}_n(\mathbb{F})$. It induces a symplectic transformation which maps the vector $(p, q)$ to $(Cq, -\tilde{C}p)$, where $\tilde{C}$ is the inverse of the transpose of $C$. This is implemented by a unitary transformation $F_{[C]}$. It is called the Fourier transform associated with the invertible matrix $C$:

$$F_{[C]}|p\rangle = \frac{1}{\sqrt{d}^n}\sum_{q \in \mathbb{F}^n} \mathrm{e}^{(2\pi\mathrm{i}/d)p^{\mathrm{t}}Cq}|q\rangle \qquad [22]$$

By construction, the relation

$$F_{[C]}\boldsymbol{w}(p, q)F_{[C]}^* = \mathrm{e}^{(2\pi\mathrm{i}/d)p^{\mathrm{t}}q}\boldsymbol{w}(Cq, -\tilde{C}p) \qquad [23]$$

follows. If $C = \mathrm{diag}(c_1, \ldots, c_n)$ is a diagonal matrix, then $F_{[C]}$ is a local unitary transformation. In fact, the Fourier transform is a tensor product

$$F_{[C]} = F_{[c_1]} \otimes F_{[c_2]} \otimes \cdots \otimes F_{[c_n]} \qquad [24]$$

with $c_x \in \mathbb{F}\backslash 0$, where the tensor product structure is determined by $|q\rangle = |q_1\rangle \otimes \cdots \otimes |q_n\rangle$.

## The Stabilizer Formalism

This section is dedicated to the stabilizer formalism, which has widely been discussed in the literature (Calderbank *et al.* 1997, Gottesman 1996, 1997). We investigated here stabilizer codes from a point of view of symmetries and show how they can be characterized by Clifford channels. We verify that

stabilizer codes are specific Clifford channels in the sense described in the last section. To begin with, we consider an irreducible Weyl system $(\boldsymbol{w}, f, \mathcal{H})$ of an even-dimensional $\mathbb{F}$-vector space $\Xi$ such that the antisymmetric part of the factor system $\sigma := f^{-1}\theta f$ is a symplectic form on $\Xi$. Furthermore, we need to introduce the following notions:

The symplectic complement of a subspace $\mathcal{Q} \subset \Xi$ is the subspace

$$\mathcal{Q}^{\sigma} = \{\xi \in \Xi | \sigma(\xi|q) = 1 \, \forall q \in \mathcal{Q}\} \qquad [25]$$

Furthermore, a subspace $\mathcal{Q}$ of $\Xi$ is isotropic if it is contained in its symplectic complement $\mathcal{Q}^{\sigma} \supset \mathcal{Q}$. In other words, for all pairs of vectors $q, q' \in \mathcal{Q}$ we have $\sigma(q\,|\,q') = 1$.

We consider an isotropic subspace $\mathcal{Q}$ and we denote by $(\boldsymbol{w}|_{\mathcal{Q}}, f|_{\mathcal{Q}}, \mathcal{H})$ the corresponding restriction of the Weyl system $(\boldsymbol{w}, f, \mathcal{H})$. Since $\mathcal{Q}$ is isotropic, it follows that the restriction $f|_{\mathcal{Q}}$ is symmetric. Hence, the Weyl algebra for the restricted system $\mathfrak{A}_{\mathcal{Q}} := \mathfrak{A}(\boldsymbol{w}|_{\mathcal{Q}}, f|_{\mathcal{Q}}, \mathcal{H})$ is an abelian subalgebra of $\mathcal{B}(\mathcal{H})$. As a consequence, all the operators in $\mathfrak{A}_{\mathcal{Q}}$ can be diagonalized simultaneously. To obtain the joint spectral resolution for all operators in $A$, we employ some facts from the theory of finite dimensional abelian C*-algebras:

1. $\mathfrak{A}_{\mathcal{Q}}$ is a finite-dimensional abelian C*-algebra and can be identified with the algebra of complex functions $\mathcal{C}(\mathcal{Q}^{\wedge})$ on a finite set $\mathcal{Q}^{\wedge}$.
2. Each element $\varpi \in \mathcal{Q}^{\wedge}$ is a character (pure state), that is, a linear functional such that $\varpi(AB) = \varpi(A)\varpi(B)$ and $\varpi(A^*) = \overline{\varpi(A)}$.
3. For each operator $A \in \mathfrak{A}_{\mathcal{Q}}$ there exists precisely one function $f_A$ on $\mathcal{Q}^{\wedge}$ which is uniquely determined by $\varpi(A) = f_A(\varpi)$. The isomorphism $A \to f_A$ is called the Gelfand isomorphism.
4. A character $\varpi \in \mathcal{Q}^{\wedge}$ is an irreducible representation of $\mathfrak{A}_{\mathcal{Q}}$ and there is a unique projection $\boldsymbol{e}_{\varpi}$ onto the subspace in $\mathcal{H}$ which carries this irreducible representation.

From these facts we derive a joint spectral resolution for all operators in $\mathfrak{A}_{\mathcal{Q}}$. Namely, each $A \in \mathfrak{A}_{\mathcal{Q}}$ can be written as

$$A = \sum_{\varpi \in \mathcal{Q}^{\wedge}} \boldsymbol{e}_{\varpi}\, \varpi(A) \qquad [26]$$

We are now prepared to introduce the notion of stabilizer codes in accordance with Calderbank *et al.* (1997) and Gottesman (1996, 1997): Let $\mathcal{Q}$ be an isotropic subspace in $\Xi$ and let $\varpi \in \mathcal{Q}^{\wedge}$ be a character of $\mathfrak{A}_{\mathcal{Q}}$. The projection $\boldsymbol{e}_{\varpi}$ is called a stabilizer code. The abelian group that is generated by the Weyl operators $\boldsymbol{w}(q), q \in \mathcal{Q}$, is called stabilizer group. The abelian C*-algebra $\mathfrak{A}_{\mathcal{Q}}$ is called stabilizer algebra. According to the following theorem, each stabilizer code is uniquely associated with a Clifford channel:

**Theorem 8** (Schlingemann 2002, 2004). *Let $\mathcal{Q}$ be an isotropic subspace of $\Xi$ and let $\boldsymbol{e}_{\varpi}$ be the stabilizer code of a character $\varpi$. Then there exists a unique Clifford channel $\boldsymbol{E}_{\varpi}$ with input system $(\boldsymbol{w}_{\varpi}, f_{\varpi}, \mathcal{H}_{\varpi})$ and output system $(\boldsymbol{w}|_{\mathcal{Q}^{\sigma}}, f|_{\mathcal{Q}^{\sigma}}, \mathcal{H})$ such that the following is true:*

(i) *For each $\xi \in \Xi$ the identity*

$$\boldsymbol{E}_{\varpi}(\boldsymbol{w}(\xi)) = \delta_{\mathcal{Q}^{\sigma}}(\xi)\boldsymbol{w}_{\varpi}(\xi) \qquad [27]$$

*is fulfilled.*

(ii) *Let $\boldsymbol{v}_{\varpi} : \mathcal{H}_{\varpi} \to \mathcal{H}$ be the isometry which embeds $\mathcal{H}_{\varpi}$ into $\mathcal{H}$, then*

$$\boldsymbol{E}_{\varpi}(A) = \boldsymbol{v}_{\varpi}^* A \boldsymbol{v}_{\varpi} \qquad [28]$$

*holds for all $A \in \mathcal{B}(\mathcal{H})$.*

(iii) *The channel $\boldsymbol{E}_{\varpi}$ is invariant under translations in the isotropic subspace $\mathcal{Q}$, that is, the identity*

$$\boldsymbol{E}_{\varpi} \circ \mathrm{Ad}[\boldsymbol{w}(q)] = \boldsymbol{E}_{\varpi} \qquad [29]$$

*holds for all $q \in \mathcal{Q}$.*

Stabilizer codes for maximally isotropic subspaces $\mathcal{Q} = \mathcal{Q}^{\sigma}$ are special, since the projection $\boldsymbol{e}_{\varpi}$ onto the eigenspace of the character $\varpi$ is one-dimensional. Thus, $\boldsymbol{e}_{\varpi}$ is the density matrix of a pure state which is called stabilizer state. In view of Theorem 8, the expectation value of a Weyl operators $\boldsymbol{w}(\xi)$ is given by

$$\mathrm{tr}(\boldsymbol{e}_{\varpi}\boldsymbol{w}(\xi)) = \varpi(\boldsymbol{w}(\xi))\delta_{\mathcal{Q}}(\xi) \qquad [30]$$

## Representation by Graphs

As described in the previous section (Theorem 8), each stabilizer codes is a pure Clifford channel which is completely determined by an isotropic subspace and a character of the corresponding stabilizer algebra. A constructive characterization of isotropic subspaces can be given in terms of graphs, as it has been shown in Schlingemann (2002, 2004). The complete description of a stabilizer code requires in addition the choice of a character of the stabilizer algebra. Both data, the isotropic subspace and the character, can be encoded in a single graph $\Lambda$. The set of vertices `N` is partitioned into four different types, the input vertices `I`, the output vertices `J`, the measurement vertices `K` and the syndrome vertices `L` (see **Figure 1**). The edges of the graph are undirected, and a pair of vertices can be connected by at most $d-1$ edges, where self-links are also allowed. The adjacency matrix (also

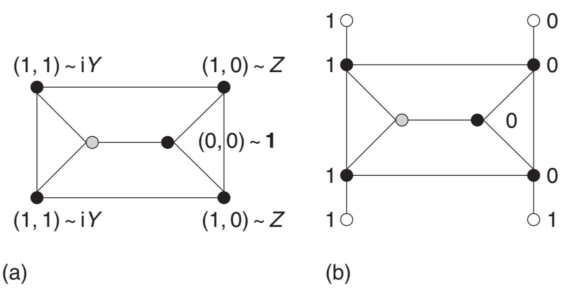



**Figure 1** (a) A graphical representation of a Weyl operator $-Y \otimes Y \otimes Z \otimes Z \otimes \mathbb{1}$ of the stabilizer algebra of a quantum error correcting code, encoding one qubit into five (see 00273). The input vertex is gray, the output vertices are black. Each binary vector represents a Pauli matrix sitting at a tensor position of the output system. (b) The expectation values which are products over all edges, where to each edge with labels $q, q'$ the value $(-1)^{qq'}$ is assigned. The character corresponds to the syndrome configuration (1110) (blanc vertices).

denoted by $\Lambda$) is a symmetric matrix with entries $\Lambda^{\mathtt{x}}_{\mathtt{y}} = 0, 1, \ldots, d-1$ according to the number of edges between $\mathtt{x}$ and $\mathtt{y}$. Thus, the adjacency matrix can be seen as a linear operator on $\mathbb{F}^{\mathtt{N}}$ with cyclic field $\mathbb{F} = \mathbb{Z}_d$. Each subset $\mathtt{A} \subset \mathtt{N}$ corresponds to a linear projection onto the subspace $\mathbb{F}^{\mathtt{A}} \subset \mathbb{F}^{\mathtt{N}}$, which we denote by $\pi_{\mathtt{A}}$. For a convenient description we introduce the following notation: the union of two sets of vertices is written without the symbol $\cup$, that is, instead of $\mathtt{I} \cup \mathtt{J}$ we write $\mathtt{IJ}$.

**Theorem 9** (Schlingemann 2002, 2004). *Let $\mathcal{Q} \subset \mathbb{F}^{\mathtt{J}} \oplus \mathbb{F}^{\mathtt{J}}$ be an isotropic subspace and let $\varpi$ be a character of the stabilizer algebra $\mathfrak{A}_{\mathcal{Q}}$. Then there exists a graph $\Lambda$ with input vertices* $\mathtt{I}$, *output vertices* $\mathtt{J}$, *measurement vertices* $\mathtt{K}$ *and syndrome vertices* $\mathtt{L}$ *such that the following holds*:

(i) *The linear operator $\pi_{\mathtt{JK}} \Lambda \pi_{\mathtt{IKL}}$ is invertible.*
(ii) *The isotropic subspace $\mathcal{Q}$ consists of the vectors $(\pi_{\mathtt{J}} \Lambda \pi_{\mathtt{JK}} q, \pi_{\mathtt{J}} q)$ with $q \in \ker(\pi_{\mathtt{IK}} \Lambda \pi_{\mathtt{JK}})$.*
(iii) *There is a unique vector $a$ in the syndrome subspace $\mathbb{F}^{\mathtt{L}}$ such that the expectation values of the character $\varpi$ are given by*

$$\varpi(\boldsymbol{w}(\pi_{\mathtt{J}} \Lambda \pi_{\mathtt{JK}} q, \pi_{\mathtt{J}} q)) = \zeta_d^{(q+a)^t \Lambda (q+a)} \qquad [31]$$

*with $q \in \ker(\pi_{\mathtt{IK}} \Lambda \pi_{\mathtt{JK}})$.*

Theorems 8 and 9 provide different useful characterizations of stabilizer codes, namely in terms of eigenspaces, Clifford channels, and graphs.

- The original definition of stabilizer codes in terms of eigenspaces goes back to Calderbank, Gottesman, Rains, Shor, and Sloane (see, e.g., Calderbank *et al.* (1997), Gottesman (1996, 1997). They have developed an approach to derive quantum codes from classical binary codes.
- Stabilizer codes can also be characterized by specific Clifford channels (see Theorem 8). The condition for a channel to be a stabilizer code is the covariance with respect to a subgroup of phase space translations. This reflects stabilizer codes in terms of symmetries.
- Theorem 9 yields a characterization of stabilizer codes in terms of graphs providing an explicit expression for the isotropic subspace and the character of the stabilizer code. This graphical representation provides a suggestive encoding of various properties like error-correcting capabilities, multipartite entanglement, the effects of specific local operations. In fact, as it has been shown in Briegel and Raussendorf (2001), Dür *et al.* (2003), and Hein *et al.* (2004) that the entanglement present in a graph state can be derived from its shape.

*See also:* Capacities Enhanced by Entanglement; Quantum Error Correction and Fault Tolerance.

## Further Reading


Briegel H-J and Raussendorf R (2001) Persistent entanglement in arrays of interacting particles. *Physical Review Letters* 86: 910.

Calderbank AR, Rains EM, Shor PW, and Sloane NJA (1997) Quantum error correction and orthogonal geometry. *Physical Review Letters* 78: 405–408.

Cleve R and Gottesman D (1997) Efficient computations of encodings for quantum error correction. *Physical Review A* 56: 76–83.

Dür W, Aschauer H, and Briegel H-J (2003) Multiparticle entanglement purification for graph states. *Physical Review Letters* 91: 107903.

Gottesman D (1996) Class of quantum error-correcting codes saturating the quantum Hamming bound. *Physical Review A* 54: 1862.

Gottesman D (1997) *Stabilizer Codes and Quantum Error Correction*. Ph.D. thesis, CalTec.

Hein M, Eisert J, Dür W, and Briegel H-J (2004) Multi-party entanglement in graph states. *Physical Review A* 69: 062311.

Holevo AS (n.d.) Remarks on the classical capacity of quantum channel, quant-ph/0212025.

Holevo AS (2004) Additivity conjecture and covariant channels. In Proc. Conference Foundations of Quantum Information. Camerino.

Klappenecker A and Roetteler M (2002) Beyond stabilizer codes I: Nice error bases. *IEEE Transactions on Information Theory* 48(8): 2392.

Klappenecker A and Roetteler M (2005) On the monomiality of nice error bases. *IEEE Transactions on Information Theory* 51(3): 1.

Schlingemann D-M (2000) Stabilizer codes can be realized as graph codes. *Quant. Inf. Comp.* 2(4): 307–323.

Schlingemann D-M (2004) Cluster states, graphs and algorithms. *Quant. Inf. Comp.* 4(4): 287–324.

Schlingemann D-M and Werner RF (2001) Quantum error-correcting codes associated with graphs. *Physical Review A* 65: 012308.

Schlingemann D-M and Werner RF (2005) On the structure of Clifford quantum cellular automata. *Preprint* (in preparation).
Scutaru H (1979) Some remarks on covariant completely positive linear maps. *Rep. Math. Phys.* 16: 79–87.
Werner RF (2001) All teleportation and dense coding schemes. *Journal of Physics A* 35: 7081.
Zmud EM (1971) Symplectic geometries over finite abelian groups. *Mathematics of the USSR Sbornik* 15: 7–29.
Zmud E (1972) Symplectic geometry and projective representations of finite abelian groups. *Mathematics of the USSR Sbornik* 16: 1–16.

# Finitely Correlated States

**R F Werner**, Technische Universität Braunschweig, Braunschweig, Germany



## Introduction

A typical problem of quantum statistical mechanics is to compute equilibrium states of quantum dynamical systems. However, there is a strange difficulty inherent in this task, which is to describe the solution: if we try to describe the quantum state by specifying all matrix elements of all local density operators, we have a job which grows exponentially with the system size. This approach is obviously out of the question for the large systems statistical mechanics is interested in. Luckily, in practice nobody really wants to see all those numbers anyway, and one is content with determining a few correlation functions, or other easily parametrized characteristics of the state. But for computing a state in the first place, we cannot restrict the state description to a such parameters. So the problem there is again: how can we efficiently parametrize the states of interest?

In this article we collect some results on a particular way of addressing this problem. It originated in the early 1990s (Fannes *et al.* 1992b) in ideas for quantizing the notion of Markov chains (Accardi and Frigerio 1983). Recently, there has been a new surge of interest in such ideas, because they turned out to be very useful for numerical work on quantum spin chains.

Its typical feature is that one does not directly describe expectation values of the state, but instead generates the state from a description of its correlations between neighboring sites. In the language of quantum information theory, it could be said that the method focuses on the entanglement between different parts of the system.

## The Basic Construction

### Notation

We consider a quantum spin chain, that is, a system of infinitely subsystems, labeled by the integers, each of which is a quantum-mechanical $d$-level system. Let us denote the observable algebra at site $x \in \mathbb{Z}$ by $\mathcal{A}_x$. Each $\mathcal{A}_x$ is hence isomorphic to the $d \times d$ matrices. The observables of the whole (infinite) system lie in the infinite tensor product $\mathcal{A}_{\mathbb{Z}} = \bigotimes_{x \in \mathbb{Z}} \mathcal{A}_x$. This is defined as a quasilocal algebra (Bratteli and Robinson 1987, 1997), which is to say that it is the algebra generated by all finite tensor products of elements of the $\mathcal{A}_x$, say $\bigotimes_{x \in \Lambda} A_x$ with $A_x \in \mathcal{A}_x$ and $\Lambda$ finite. Such an element is said to be localized in $\Lambda$, and we denote by $\mathcal{A}_\Lambda$ the corresponding algebra. For $\Lambda_1 \subset \Lambda_2$, we identify $\mathcal{A}_{\Lambda_1}$ with a subalgebra of $\mathcal{A}_{\Lambda_2}$, by tensoring with the identity operator on all sites in $\Lambda_2 \backslash \Lambda_1$. $\mathcal{A}_{\mathbb{Z}}$ is the completion of the union of all $\mathcal{A}_\Lambda$, with $\Lambda$ finite, under the C$^*$-norm.

A state $\omega$ on $\mathcal{A}_{\mathbb{Z}}$ is uniquely specified by its expectations on the subalgebras $\mathcal{A}_\Lambda$. Since these are finite-dimensional matrix algebras, we can write $\omega(A) = \mathrm{tr}(\rho_\Lambda A)$ for $A \in \mathcal{A}_\Lambda$, with a "local density operator" $\rho_\Lambda$. The system of local density operators must be consistent with respect to restrictions (partial traces).

So far we have not used the structure of the underlying lattice $\mathbb{Z}$ in any way. This enters via the translation automorphisms $\tau_n$ of $\mathcal{A}_{\mathbb{Z}}$, which identify $\mathcal{A}_x$ with $\mathcal{A}_{x+n}$. A state is called translationally invariant, if $\omega \circ \tau_n = \omega$. The translationally invariant states form a weakly compact convex subset of the state space of $\mathcal{A}_{\mathbb{Z}}$, whose extreme points are called ergodic states.

### How to Generate Correlations

Correlations between parts of a systems typically have their origin in an interaction in the past. Even if the subsystems are dynamically separated later on, the correlation persists, and one can take this as a motivation to model correlations from two ingredients: a simplified prototype of a correlated system, and some evolution taking the parts of the simplified system to the parts of the given system. Let us consider a composite system, whose parts have observable algebras $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively, so that the whole system has algebra $\mathcal{A}_1 \otimes \mathcal{A}_2$. We can build a state $\omega$ on this system from a simpler one,

say a state $\eta$ on some $\mathcal{B}_1 \otimes \mathcal{B}_2$, and two completely positive unit preserving maps $T_i : \mathcal{A}_i \to \mathcal{B}_i$ such that

$$\omega(A_1 \otimes A_2) = \eta(T_1(A_1) \otimes T_2(B_2))$$

Some features of $\eta$ are inherited by $\omega$. For example, when $\eta$ is separable (a convex combination of products), which is always the case if either $\mathcal{B}_1$ or $\mathcal{B}_2$ is classical (i.e., an abelian algebra), then the same holds for $\omega$. Hence, if we want to describe quantum correlated "entangled" states, we have to build the correlations on an entangled state $\eta$. Similarly, the "size" of the model system $\mathcal{B}_1 \otimes \mathcal{B}_2$ limits the strength of correlations in $\omega$. As for every correlated state, we can look at the linear functionals on $\mathcal{A}_2$, which are of the form $A \mapsto \omega(A_1 \otimes A)$ with fixed $A_1 \in \mathcal{A}_2$. The dimension of the space of such functionals might be called the correlation dimension of $\omega$. This dimension is 1 for product states, and can clearly not increase by passing from $\eta$ to $\omega$. Hence, it is bounded by the dimensions of $\mathcal{B}_1$ and $\mathcal{B}_2$, even if $\mathcal{A}_1$ and $\mathcal{A}_2$ are infinite dimensional. "Finite correlation" in the sense of the title of this article refers to the finiteness of the correlation dimension between the two halves of a spin chain.

## The VBS Construction, and Matrix Product States

The so-called valence bond solid (VBS) states on a chain are constructed by applying these ideas to the correlations across every link of a spin chain. Let us introduce a correlated model state $\eta_x$ on some algebra $\mathcal{B}_x^- \otimes \mathcal{B}_x^+$ for every bond $(x, x+1)$. Then the state at site $x$ is a function of contributions from both bonds connecting it, and we express this by a completely positive map $T_x : \mathcal{A}_x \to \mathcal{B}_{x-1}^+ \otimes \mathcal{B}_x^-$. Then an observable $A_1 \otimes \cdots \otimes A_L$ on a chain piece of length $L$ is first mapped by $\bigotimes_{x=1}^L T_x$ to an element of $\mathcal{B}_0^+ \otimes \mathcal{B}_1^- \otimes \cdots \otimes \mathcal{B}_{L-1}^+ \otimes \mathcal{B}_L^-$. Evaluating with the states $\eta_1 \otimes \cdots \otimes \eta_{L-1}$, we are left with an element of $\mathcal{B}_0^+ \otimes \mathcal{B}_L^-$, which we can evaluate with yet another state $\eta_{0L}$ describing the boundary conditions for the construction (see Figure 1).

Clearly, if we take the algebras $\mathcal{B}_x^\pm$ large enough, and the model states $\eta_x$ sufficiently highly entangled, we can generate every state on the finite chain. However, we can get an interesting class of states, even for fixed finite dimensions of the $\mathcal{B}_x^\pm$. By restricting this correlation dimension, we can set a level of complexity for the state description. We can then try to handle a given physical problem first with simple states of low correlation dimension, and increase this parameter only as needed. A typical problem here is to determine the ground state of a finite-range Hamiltonian. We can then optimize each $T_x$ and $\eta_x$ separately, minimizing the ground

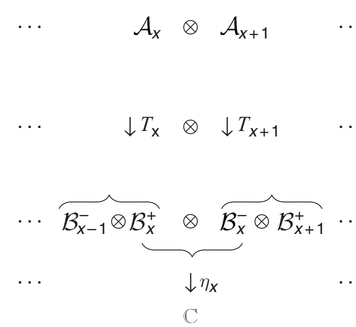


**Figure 1**

state energy with all other elements fixed. This is a semidefinite programming problem, for which very efficient methods are known. The global minimization is then done by letting the optimization site $x$ sweep over the whole chain as often as needed.

In a ground-state problem one is looking for a pure state, and it is therefore sufficient to choose both the model states $\eta_x$ and the operations $T_x$ to pure, that is, without decomposition into sums of similar objects. The scheme is thus run at the vector level rather than the operator level: we take the algebras $\mathcal{B}_x^+ = \mathcal{B}_x^-$ as the operators on a Hilbert space $\mathcal{K}_x$, and $\eta_x = (\dim \mathcal{K}_x)^{-1} |\Omega_x\rangle\langle\Omega_x|$ with the (unnormalized) maximally entangled vector

$$\Omega_x = \sum_j |j\rangle \otimes |j\rangle \in \mathcal{K}_x \otimes \mathcal{K}_x \qquad [1]$$

The maps $T_x$ will be implemented by a single operator $V_x : \mathcal{K}_{x-1} \otimes \mathcal{K}_x \to \mathcal{H}$ as $T_x(A) = V_x^* A V_x$. Then the vectors $\Psi \in \mathcal{H}^{\otimes L}$ contributing to the state on the chain of length $L$ are of the form

$$\begin{aligned}\Psi &= V_1 \otimes \cdots V_L \left(|j_0\rangle \otimes \Omega^{\otimes L} \otimes |i_L\rangle\right) \\ &= \sum_{j_0, j_1, \ldots, j_L} (V_1 \otimes \cdots V_L) |j_0, j_1, j_1, \ldots, j_{L-1}, j_L\rangle\end{aligned}$$

where $j_0, j_L$ are labels for bases in $\mathcal{K}_0$ and $\mathcal{K}_L$, describing the possible choices at the boundary, and we have used the special form of $\Omega$. We write out the operators $V_x$ in components, so that

$$V_x |jj'\rangle = \sum_\mu |\mu\rangle V_{x;jj'}^\mu$$

with suitable $\dim \mathcal{K}_{x-1} \times \dim \mathcal{K}_x$ dimensional matrices $V_x^\mu$, in terms of which the above expression can be interpreted as a matrix product. The components of $\Psi$ in a product basis $\{|\mu\rangle\}$ become

$$\langle \mu_1, \ldots, \mu_L | \Psi \rangle = \langle j_0 | V_1^{\mu_1} V_2^{\mu_2} \cdots V_L^{\mu_L} | j_L \rangle \qquad [2]$$

Due to this form the states generated in this way have also been called "matrix product states" (Klümper 1991). If one wants to consider periodic

boundary conditions, the indices $j_0$ and $j_L$ can also be contracted, and the expression becomes a trace. For some simulations it is also convenient to choose $\sim\mathcal{K}_0 = \dim \mathcal{K}_L = 1$, so there is only one matrix element to be considered.

The scheme for getting ground-state vectors described here is essentially the same as the density matrix renormalization group method (Verstraete *et al.* 2004). However, the version given here appears to be more transparent, more flexible, and in some cases (e.g., periodic boundary conditions) vastly more efficient. However, it may be too early for such judgment, since this is very much work in progress (Verstraete *et al.* 2005).

In the sequel, we will focus not so much on the numerical aspects, but on the possibility this construction offers to explicitly construct nontrivial translationally invariant states on the infinite chain. Numerically, even in a translation invariant situation the matrices $V_x^\mu$ obtained by optimization may turn out to depend on $x$ (Wolf, Private Communication), that is, one has to admit the possibility of a spontaneous symmetry breaking. However, for the construction of states on the infinite chain we will simply fix all $V_x$ to be equal. In some sense this turns the matrix product into a matrix power, which could be analyzed by methods familiar from the transfer matrix formalism of statistical mechanics. In eqn [2] this does not work, because of the $\mu$-dependence of the matrices involved. Nevertheless, a slight reorganization of the construction will lead to a transfer-matrix-like formalism.

### The Evolution Operator Construction

Fixing all $T_x$ to be the same in Figure 1 still does not fix the state uniquely, since both in the mixed state version and in the pure state version of the construction some boundary information enters, as well. This boundary information then has to be chosen in such a way that a consistent family of local density operators is generated. It turns out that by rearranging the construction a little bit one can trivially solve one boundary condition, and reduce the other to finding a fixed point of a linear operator. This rearrangement was first carried out in Fannes *et al.* (1992b), where the term "finitely correlated state" was also coined.

The basic element of the VBS construction was the operators $T : \mathcal{A} \to \mathcal{B}^+ \otimes \mathcal{B}^-$ (here already taken independent of $x$). This is specified by $\dim \mathcal{A} \cdot \dim \mathcal{B}^+ \cdot \dim \mathcal{B}^-$ matrix elements. However, assuming we can identify the algebras $\mathcal{B}^\pm$, we can also consider these matrix elements as those of an "evolution operator" $\mathbb{E} : \mathcal{A} \otimes \mathcal{B} \to \mathcal{B}$. This operator is once again taken to be completely positive and unit preserving. We introduce its $n$th iterate $\mathbb{E}^{(n)} : \mathcal{A}^{\otimes n} \otimes \mathcal{B} \to \mathcal{B}$ by the recursion

$$\mathbb{E}^{(1)} = \mathbb{E}, \qquad \mathbb{E}^{(n+1)} = \mathbb{E}(\mathrm{id}_{\mathcal{A}} \otimes \mathbb{E}^{(n)}) \qquad [3]$$

Clearly, these operators are again completely positive and unit preserving. Another way to express this iteration is to look at $\mathbb{E}$ as a family of maps on $\mathcal{B}$, parametrized by $A \in \mathcal{A}$: We set $\mathbb{E}_A(B) = \mathbb{E}(A \otimes B)$, and find

$$\mathbb{E}^{(n)}(A_1 \otimes \cdots \otimes A_n \otimes B) = \mathbb{E}_{A_1} \cdots \mathbb{E}_{A_1}(B) \qquad [4]$$

An important special role is played by the operator $\widehat{\mathbb{E}} = \mathbb{E}_{\mathbb{1}}$, which is again completely positive and unit preserving.

Now given any state $\eta$ on $\mathcal{B}$, we get a state $\omega_n$ on $\mathcal{A}^{\otimes n}$, by setting

$$\omega_n(A_1 \otimes \cdots A_n) = \eta\Big(\mathbb{E}^{(n)}(A_1 \otimes \cdots \otimes A_n \otimes \mathbb{1})\Big) \qquad [5]$$

Since $\widehat{\mathbb{E}}(\mathbb{1}) = \mathbb{1}$, this family of states is consistent with respect to increasing $n$, by adding sites on the right, that is, $\omega_{n+1}(A \otimes \mathbb{1}) = \omega_n(A)$. In other words, the family $\omega_n$ defines a state on the infinite right half-chain. This state can be extended to the full chain, as a translationally invariant state if and only if consistency also holds for adding sites on the left, that is, if $\omega_{n+1}(\mathbb{1} \otimes A) = \omega_n(A)$ for all $A \in \mathcal{A}^{\otimes n}$. For this we need a condition on the state $\eta$: it must be invariant under the map $\widehat{\mathbb{E}}$ (i.e., $\eta(\widehat{\mathbb{E}}(B)) = \eta(B)$ for all $B \in \mathcal{B}$). This is the only requirement, and we call $\omega$ the state $\mathcal{A}_{\mathbb{Z}}$ generated by $\mathbb{E}$ and $\eta$. Note that since $\widehat{\mathbb{E}}$ has the invariant vector $\mathbb{1}$, its transpose also has an invariant vector, which can also be chosen as a state. We will often look at unique invariant state, in which case we can call $\omega$ the state generated by $\mathbb{E}$, without having to mention $\eta$.

The valence bond picture was very much suggested by trying to describe correlations in a spatially distributed quantum system (the chain). The construction given here is perhaps more readily suggested by a process in time, rather than space. In fact, the paper by Fannes *et al.* (1992b) was partly motivated by an attempt to define a quantum analog of Markov processes (Accardi and Frigerio 1983). In fact, we can think of the construction as a general form for a repeated measurement in quantum theory. The object on which the measurements are performed has observable algebra $\mathcal{B}$, whereas $\mathcal{A}$ describes the successive outputs. Choosing $\mathcal{A}$ to be classical (abelian) we would find in $\omega$ the joint probability distribution of the sequence of measured values, when the initial state of the object is $\eta$ (not necessarily invariant). Allowing nonabelian $\mathcal{A}$ would

then correspond to a family of delayed choice experiments: while $\mathbb{E}$ describes the interaction of the system with the measurement apparatus (including the overall state change $\widehat{\mathbb{E}}$), we are still free to make correlated and even entangled measurements on the successive output systems. This interpretation suggests many extensions, in particular, to continuous time (where the case of abelian outputs is discussed extensively in the classic book by Davies 1976), or to cases allowing an external quantum input in each step, in which case we are looking at a quantum channel with memory $\mathcal{B}$ (Kretschmann and Werner 2005).

In spite of the different natural interpretations, however, the constructions in this and previous paragraphs give exactly the same class of translationally invariant states on the chain, as was shown in (Fannes *et al.* 1992b).

## Ergodic Decomposition

A state on $\mathcal{A}_{\mathbb{Z}}$ is called ergodic if it is an extreme point of the compact convex subset of translationally invariant states. Often in statistical mechanics, one finds states which may be ergodic, but nevertheless contain a breaking of translation symmetry. Such states can be decomposed into periodic states, that is, states which are invariant with respect to some power of the shift. In general, new decompositions may become possible for any period. If no decomposition into periodic states is possible, the state is called completely ergodic.

In this section we consider the question of how to decompose a finitely correlated state into ergodic components, using a well-established connection between ergodicity and clustering properties (Bratteli and Robinson 1987, 1997), that is, the decay of correlation functions.

Correlation functions are very easily evaluated for finitely correlated states: let $A_\pm$ be two observables localized on $n_\pm$ sites, and suppose that these sites are separated by $L$ sites. Then eqn [5] gives

$$\omega\left(A_- \otimes \mathbb{1}^{\otimes L} \otimes A_+\right) = \eta\left(\mathbb{E}_{A_-}^{(n_-)}\, \widehat{\mathbb{E}}^L\, \mathbb{E}_{A_-}^{(n_+)}(\mathbb{1})\right) \qquad [6]$$

The $L$-dependence of this operator is clearly governed by the matrix powers of $\widehat{\mathbb{E}}$. By assumption this operator always has the eigenvalue 1, because $\widehat{\mathbb{E}}(\mathbb{1}) = \mathbb{1}$, and has norm $\leq 1$, because it is also completely positive. The spectrum is hence contained in the unit circle. Each eigenvalue with modulus $<1$ thus contributes exponentially decaying terms to the correlation function [6]. From eigenvalues of modulus 1, which make up the so-called peripheral spectrum, we may get constant or periodic contributions. This distinction is directly reflected in the ergodic properties (Fannes *et al.* 1992b):

- When the eigenvalue 1 is simple, there is a unique invariant state $\eta$, and $\lim_n n^{-1}\sum_{k=0}^{n-1}\widehat{\mathbb{E}}^k(B) = \eta(B)\mathbb{1}$. This implies, by [6] and (Bratteli and Robinson 1987, 1997, theorem 4.3.22), that $\omega$ is ergodic.
- When the eigenvalue 1 is simple, the peripheral spectrum consists precisely of the $p$th roots of unity for some $p \geq 1$. The state $\omega$ is then the equal-weight convex combination of $p$ periodic states with period $p$, which are translates of each other.
- In particular (i.e., for $p = 1$), a peripheral spectrum consisting only of the simple eigenvalue 1 implies that $\omega$ is exponentially clustering in the sense that

$$\begin{aligned}&\left|\omega\left(A_- \otimes \mathbb{1}^{\otimes L} \otimes A_+\right) - \omega(A_-)\omega(A_+)\right| \\ &\quad \leq \mathrm{poly}(L)\, r^L \|A_-\|\|A_+\|\end{aligned} \qquad [7]$$

  where $r$ is the largest modulus of eigenvalues other than 1, and poly is polynomial obtained from the Jordan normal form of $\widehat{\mathbb{E}}$. By the previous item, the state $\omega$ is then completely ergodic.
- Conversely, if a state is finitely correlated, and is ergodic (resp. completely ergodic), it has a representation such that 1 is a simple eigenvalue (resp. the peripheral spectrum is trivial).

## Purity

### Pure States

As in the case of the VBS construction, there is a version of the evolution operator construction, which is especially suited to produce pure states. Pure states are those which cannot be decomposed into a weighted sum of other states. For a translationally invariant state, this is a much stronger property than ergodicity and even complete ergodicity: not only the decomposition into periodic states is impossible, but any decomposition whatsoever. Nevertheless, this is what one expects from a ground state of translationally invariant interaction.

From the formula [5] it is clear that if we decompose the $\mathbb{E}$-operator entering for a site $x$ into a sum two completely positive terms, we will have decomposed $\omega$ into two positive terms. These might still be equal, but it is certainly suggestive to look at states generated with an $\mathbb{E}$, which cannot be decomposed nontrivially into a sum of other

completely positive maps. Such maps are called pure, and are characterized by the form

$$\mathbb{E}(A \otimes B) = V^*(A \otimes B)V \\ V : \mathbb{C}^k \to \mathbb{C}^d \otimes \mathbb{C}^k \text{ is isometric} \qquad [8]$$

and $\mathcal{A}$ and $\mathcal{B}$ are the algebras of $d \times d$ and $k \times k$ matrices, respectively. Finitely correlated states generated from such a pure evolution operator are called purely generated. These are the candidates for pure finitely generated states.

The form of a pure map is reminiscent of the Stinespring dilation of a general completely positive map: for a general $\mathbb{E}$, we can set

$$\mathbb{E}(A \otimes B) = V^*(\mathbb{1}_{\widetilde{\mathcal{A}}} \otimes A \otimes B)V \qquad [9]$$

where $\widetilde{\mathcal{A}}$ is some auxiliary matrix algebra. Since the invariance condition for $\eta$ does not involve the $\mathcal{A}$ algebras, we get a purely generated state $\widetilde{\omega}$ with one-site algebra $(\widetilde{\mathcal{A}} \otimes \mathcal{A})$, whose restriction to the original chain is $\omega$. Hence, purely generated states are the prototypes from which all other finitely correlated are obtained by sitewise restriction.

But are such states pure? Since $\widehat{\mathbb{E}}$ need not have a trivial peripheral spectrum, the previous section tells us that a purely generated state may have a nontrivial decomposition into other, perhaps periodic states. But this is the only restriction we have to make. Indeed, the following statements about a finitely correlated state $\omega$ are equivalent (Fannes *et al.* 1994, theorem 1.5):

- $\omega$ is pure;
- $\omega$ is purely generated, and the operator $\widehat{\mathbb{E}}$ has trivial peripheral spectrum;
- the mean entropy of $\omega$ vanishes, and $\omega$ is clustering, that is, [7] holds; and
- $\mathbb{E}$ has the form [8], and no subalgebra of $\mathcal{B}$, which contains $\mathbb{1}$, is invariant under all operators $\mathbb{E}_A$.

### The Asymptotic Form of the Local Support

Let us now fix an isometry $V$, such that $\widehat{\mathbb{E}}$ has trivial peripheral spectrum, and let $\rho$ denote the unique invariant state of $\widehat{\mathbb{E}}$. Then the vectors $\Psi \in \mathcal{H}^{\otimes n}$ in the support of $\omega$ are of the form [2] and depend, apart from the fixed choice of the $V_x^\mu \equiv V^\mu$, on the boundary indices $j_0, j_n = 1, \ldots, k$. We can consider this as a map $\Gamma_n$ from $k \times k$ matrices to $\mathcal{H}^{\otimes n}$:

$$\langle \mu_1, \ldots, \mu_n \,|\, \Gamma_n(B) \rangle = \mathrm{tr}(B V^{\mu_1} V^{\mu_2} \cdots V^{\mu_n}) \qquad [10]$$

and denote the range of $\Gamma_n$ by $\mathcal{G}_n$. Then $\mathcal{G}_n$ is at most $k^2$-dimensional. Moreover, this family of subspaces is nested, that is, $\mathcal{G}_{n+m} \subseteq \mathcal{G}_n \otimes \mathcal{H}^{\otimes n}$ and $\mathcal{G}_{n+m} \subset \mathcal{H}^{\otimes m} \otimes \mathcal{G}_m$. Using that $\widehat{\mathbb{E}}(B)^n \to \rho(B)\mathbb{1}$ converges exponentially fast, we also find that $\Gamma_n$ is asymptotically an isometry between $\mathcal{G}_n$, and the Hilbert space of $k \times k$ matrices with scalar product $\langle B, C \rangle_\rho = \mathrm{tr}(\rho B^* C)$. Hence, all the spaces $\mathcal{G}_n$ are asymptotically identified, even though they are contained in each other. This "self-similarity" is the source of many further properties. For example, for any density matrix $\widetilde{\rho}$ on $\ell + m + r$ sites supported by $\mathcal{G}_{\ell+m+r}$, and any observable $A$, localized on $m$ sites in the middle of this interval (with $\ell$ to the left and $r$ to the right), we get the expectation $\mathrm{tr}(\widetilde{\rho}A) \approx \omega(A)$, up to exponentially small terms depending only on $\ell$ and $r$.

### Ground States and Gaps

Suppose we fix some interval length $\ell$, and let $h$ be the projection onto the complement of $\mathcal{G}_\ell$ in $\mathcal{H}^{\otimes \ell}$. We now consider $h$ as the interaction term of a lattice interaction, that is, we consider the formal Hamiltonian

$$H = \sum_x \tau_x(h) \qquad [11]$$

Then in the finitely correlated state $\omega$, each term in this sum has expectation zero, which is the absolute minimum for such expectations, because $h \geq 0$. In this sense $\omega$ is the ground state for this Hamiltonian. Usually, ground states are not characterized in this way: one can only require that the average energy is minimized with respect to all translationally invariant states (Bratteli and Robinson 1987, 1997, theorem 6.2.58). Hence, one can usually perturb a ground state locally such that some terms in [11] have less than average expectation, at the expense of others. For $\omega$ this is clearly impossible. Moreover, any state $\omega'$ with $\omega'(\tau_x(h)) = 0$ for all $x$ must coincide with $\omega$, even if we do not impose translation invariance. This follows from the previous section: the local density operators of $\omega'$ must all be supported in $\mathcal{G}_n$ by the nesting property; hence, if we compare density operators on intervals of length $\ell + m + r$ on observables localized on the middle $m$ sites, we get $\omega'(A) \approx \omega(A)$, up to errors exponentially small in $\ell$ and $r$.

The Hamiltonian [11] involves an infinite sum, which can be mathematically understood as a quadratic form in the GNS-representation associated with $\omega$ (Bratteli and Robinson 1987, 1997). This is the Hilbert space spanned by vectors written as $A\Omega$, with the scalar products $\langle A\Omega, B\omega \rangle = \omega(A^*B)$, for local operators $A, B$. The ground-state property then implies $H\Omega = 0$, and $H \geq 0$, because $h \geq 0$. It can be seen that $H$ generates a well-defined dynamics, and is essentially self-adjoint on the domain of such vectors. Thus also the spectrum of $H$ is a well-defined concept. This suggests a strengthening of the

ground-state property: not only is $\Omega$ the unique eigenvector of $H$ for eigenvalue 0, but there is a gap $\gamma > 0$ between zero and the next eigenvalue. This property is of considerable interest for models in solid-state physics and statistical mechanics. It was shown for all ergodic pure finitely correlated states in (Fannes *et al.* 1992b).

## Density

### Density of Finitely Correlated Pure States

The natural topology in which to consider the approximation between states on the chain is the weak topology. A sequence $\omega_n$ converges weakly to $\omega$ if for all local $A$ the expectations converge, that is, $\omega_n(A) \to \omega(A)$.

Let us start from an arbitrary translationally invariant state $\omega$, and see how we can approximate it. First, we can split the chains into intervals of length $L$, and replace $\omega$ by the tensor product of the restrictions of $\omega$ to each of these intervals. This state is not translationally invariant, so we average it over the $L$ translations, and call the resulting state $\omega_L$. Consider a local observable $A$, whose localization region has length $R$. Then for $L - R$ out of the $L$ translates contributing to $\omega_L$ the expectation will be the same as for $\omega$, and we get

$$\omega_L(A) = \left(1 - \frac{R}{L}\right)\omega(A) + \frac{R}{L}\widetilde{\omega}(A),$$

where the error term $\widetilde{\omega}$ is again a state. Hence, $\omega_L$ converges weakly to $\omega$ as $L \to \infty$. One can show easily that $\omega_L$ is finitely correlated, with an algebra $\mathcal{B}$ essentially equal to $\mathcal{A}^{\otimes L}$. Hence, the finitely correlated states are weakly dense in the set of translationally invariant states.

We can make the approximating states purer by a very simple trick. In the previous construction we always take two intervals together, and replace the tensor product of the two restrictions by a purification, that is, by a pure state on an interval of length $2L$, whose restrictions to the two length-$L$ subintervals coincide with $\omega$. We average this over $2L$ translates, and call the result $\eta_L$. The estimates showing that $\eta_L \to \omega$ weakly are exactly the same as before. Moreover, one can show (Fannes *et al.* 1992a) that $\eta_L$ is purely generated.

Being defined as a convex combination of other states, $\eta_L$ is not pure, and the peripheral spectrum of $\widehat{\mathbb{E}}$ will contain all the $2L$th roots of unity. However, we can use that such a rich peripheral spectrum is not generic for $\widehat{\mathbb{E}}$ constructed from an isometry $V$. Therefore, if we choose an isometry $V_\varepsilon$ close to the isometry $V$ generating $\eta_L$, we obtain a purely generated state $\eta_L^\varepsilon$ with trivial peripheral spectrum. Since the expression for expectations of such states depends continuously on the generating isometry, we have that $\eta_L^\varepsilon \to \eta_L$ as $\varepsilon \to 0$. But we know from the previous section that such states are pure. Hence, the pure finitely correlated states are weakly dense in the set of all translationally invariant states (Fannes *et al.* 1992a).

This has implications for the geometry of the compact convex set of translationally invariant states, which are rather counter-intuitive for the intuitions trained on finite-dimensional convex bodies. To begin with, the extreme points (the ergodic states) are dense in the whole body. This is not such a rare occurrence in infinite-dimensional convex sets, and is shared, for example, by the set of operators $F$ with $0 \le F \le \mathbb{1}$ on an infinite-dimensional Hilbert space (Davies 1976). Together with the property that the translationally invariant states form a simplex, it actually fixes the structure of this compact convex set to be the so-called Poulsen simplex. This was known also without looking at finitely correlated states. The rather surprising result of the above density argument is that even the small subclass of states which are extremal, not only in the translationally invariant subset but even in the whole state space, is still dense.

### Finitely Correlated Pure States with Bounded Memory Dimension

It is clear in the above construction that the dimension of the algebra $\mathcal{B}$ goes to infinity for an approximating sequence. How many states can we get with a fixed memory algebra $\mathcal{B}$? The dimension of this manifold can be estimated easily from the number of parameters needed to describe the map $\mathbb{E}$, and this dimension is certainly small compared to the dimension of the state space of the length $L$ piece of the chain as $L \to \infty$. However, since this is an infinite set, and not a linear subspace, we do not get an immediate bound on the dimension of the linear span of these states. What we want to show in this section is that the space of finitely correlated states with fixed $\mathcal{B}$ nevertheless generates a low-dimensional subspace of states on any large interval of the chain. To this end we will have to exhibit many observables $A$, localized on $L$ sites, whose expectation is the same for all finitely correlated states with given $\mathcal{B}$.

Let us look first at the case of purely generated states, or rather at the vectors $\Psi \in \mathcal{H}^{\otimes L}$, which can be written in the form [2], which in the translation invariant case becomes

$$\langle \mu_1, \ldots, \mu_L \mid \Psi \rangle = \langle j_0 | V^{\mu_1} V^{\mu_2} \cdots V^{\mu_L} | j_L \rangle \qquad [12]$$

for some collection $V^1, \ldots, V^d$ of $k \times k$ matrices, and some basis labels $j_0, j_L \in \{1, \ldots, k\}$. The span of all such vectors will be denoted by $\mathcal{V}_L(k, d)$, and we would like to analyze the growth of dim $\mathcal{V}_L(k, d)$, as $L \to \infty$. Now a vector with components $a(\mu_1, \ldots, \mu_L)$ lies in the orthogonal complement of $\mathcal{V}_L(k, d)$ if and only if

$$\sum_{\mu_1,\ldots,\mu_L} a(\mu_1, \ldots, \mu_L) V^{\mu_1} V^{\mu_2} \cdots V^{\mu_L} = 0$$

for any collection of matrices $V^\mu$. In other words, this expression, considered as a noncommutative polynomial in $d$ variables, is a polynomial identity for $k \times k$ matrices. The simplest such identity, for $k = 2$, $d = 3, L = 5$, is $[A, [B, C]^2] = 0$. (For the proof observe that $[B, C]$ is traceless, so its square is a multiple of the identity by the Cayley–Hamilton theorem.) This identity alone implies the existence of many more identities. For example, we can substitute higher-order polynomials for $A, B, C$, and multiply the identity with arbitrary polynomial from the right or form the left. There is a well-developed theory for such identities, called the theory of polynomial identity (PI) rings. In that context, the precise growth we are looking for has been worked out (Drensky 1998):

$$\lim_{L\to\infty} \frac{\log \dim \mathcal{V}_L(k, d)}{\log L} = (d-1)k^2 + 1 \qquad [13]$$

Thus, the dim $\mathcal{V}_L(k, d)$ only grows like a polynomial in $L$, of known degree, and the joint support of all purely generated finitely correlated state is exponentially small compared to $\mathcal{H}^{\otimes L}$.

We can apply the same idea to the set of all finitely correlated states with $\mathcal{B}$ equal to the $k \times k$ matrices. The joint support in this case is the full space, since the trace state on the chain, which is a product state generated with $k = 1$, already has full support. However, it is still true all but a polynomial number of expectation values of $\omega$ are already fixed by specifying $k$. Indeed, formula [5] for a general state is precisely of the form [12], with the difference that the arguments $A$ replace $\mu$, and the matrices $\mathbb{E}_A$ are now operators on the $k^2$-dimensional space $\mathcal{B}$. If we only want an upper bound, we can ignore subtlatties coming from Hermiticity and normalization constraints on $\mathbb{E}$, and we get that the dimension of all finitely correlated states generated from the $k \times k$ matrices, restricted to a subchain of length $L$, grows at most like $L^\alpha$, with $\alpha \leq (d^2 - 1)k^2 + 1$.

*See also:* Ergodic Theory; Quantum Spin Systems; Quantum Statistical Mechanics: Overview.

## Further Reading


Accardi L and Frigerio A (1983) Markovian cocycles. *Proceedings of the Royal Irish Academy* 83A: 251–263.

Bratteli O and Robinson DW (1987, 1997) *Operator Algebras and Quantum Statistical Mechanics I, II*, 2nd edn. Springer.

Davies EB (1976) *Quantum Theory of Open Systems*. Academic Press.

Drensky V (1998) Gelfand–Kirillov dimension of PI-algebras. In: *Methods in Ring Theory*, (Levico Terme, 1997), Lecture Notes in Pure and Appl. Math., vol. 198, pp. 97–113. New York: Dekker.

Fannes M, Nachtergaele B, and Werner RF (1992a) Abundance of translation invariant pure states on quantum spin chains. *Lett. Mathematical Physics* 25: 249–258.

Fannes M, Nachtergaele B, and Werner RF (1992b) Finitely correlated states on quantum spin chains. *Communications in Mathematical Physics* 144: 443–490.

Fannes M, Nachtergaele B, and Werner RF (1994) Finitely correlated pure states. *Journal of Functional Analysis* 120: 511–534.

Klümper A, Schadschneider A, and Zittartz J (1991) *Journal of Physics A* 24: L955.

Kretschmann D and Werner RF (2005) *Quantum channels with memory*, quant-ph/0502106.

Verstraete F, Porras D, and Cirac JI (2004) DMRG and periodic boundary conditions: a quantum information perspective. *Physical Review Letters* 93: 227205.

Verstraete F, Weichselbaum A, Schollwöck U, Cirac JI, and von Delft J (2005) *Variational matrix product state approach to quantum impurity models*, cond-mat/0504305.

Wolf M private communication.


# Finite-Type Invariants


**D Bar-Natan**, University of Toronto, Toronto, ON, Canada




## Introduction

Knots belong to sailors and climbers and upon further reflection, perhaps also to geometers, topologists, or combinatorialists. Surprisingly, throughout the 1980s, it became apparent that knots are also closely related to several other branches of mathematics in general and mathematical physics in particular. Many of these connections (though not all!) factor through the notion of "finite-type invariants" (aka "Vassiliev" or "Goussarov–Vassiliev" invariants) (Goussarov 1991, 1993, Vassiliev 1990, 1992, Birman-Lin 1993, Kontsevich 1993, Bar-Natan 1995).

Let $V$ be an arbitrary invariant of oriented knots in oriented space with values in some abelian group $A$. Extend $V$ to be an invariant of 1-singular knots, knots that may have a single singularity that locally looks like a double point $\times$, using the formula

$$V(\times) = V(\overcrossing) - V(\undercrossing) \qquad [1]$$

Further extend $V$ to the set $\mathcal{K}^m$ of $m$-singular knots (knots with $m$ double points) by repeatedly using [1].

**Definition 1** We say that $V$ is of type $m$ if its extension $V|_{\mathcal{K}^{m+1}}$ to $(m+1)$-singular knots vanishes identically. We also say that $V$ is of finite type if it is of type $m$ for some $m$.

Repeated differences are similar to repeated derivatives; hence, it is fair to think of the definition of $V|_{\mathcal{K}^m}$ as repeated differentiation. With this in mind, the above definition imitates the definition of polynomials of degree $m$. Hence, finite-type invariants can be considered as "polynomials" on the space of knots.

As described in the section "Basic facts", finite-type invariants are plenty and powerful and they carry a rich algebraic structure and are deeply related to Lie algebras. There are several constructions for a "universal finite-type invariant" and those are related to conformal field theory, the Chern–Simons–Witten topological quantum field theory, and Drinfel'd's theory of associators and quasi-Hopf algebras (see the section "The proofs of the fundamental theorem"). Finite-type invariants have been studied extensively (see the section "Some further directions") and generalized in several directions (see the section "Beyond knots"). But the first question on finite-type invariants remains unanswered:

**Problem 2** *Honest polynomials are dense in the space of functions. Are finite-type invariants dense within the space of all knot invariants? Do they separate knots?*

In a similar way, one may define finite-type invariants of framed knots (and ask the same questions).

## Basic Facts

### Classical Knot Polynomials

The first (nontrivial!) thing to notice is that there are plenty of finite-type invariants and they are at least as powerful as all the standard knot polynomials combined (finite-type invariants are like polynomials on the space of knots; the standard phrase "knot polynomials" refers to a different thing – knot invariants with polynomial values):

**Theorem 3** (Bar-Natan 1995, Birman-Lin 1993). *Let $J(K)(q)$ be the Jones polynomial of a knot $K$ (it is a Laurent polynomial in a variable $q$). Consider the power series expansion $J(K)(e^x) = \sum_{m=0}^{\infty} V_m(K)x^m$. Then each coefficient $V_m(K)$ is a finite-type knot invariant (thus, the Jones polynomial can be reconstructed from finite-type information).*

A similar theorem holds for the Alexander–Conway, HOMFLY-PT, and Kauffman polynomials (Bar-Natan 1995), and indeed, for arbitrary Reshetikhin–Turaev invariants (Reshetikhin and Turaev 1990, Lin 1991), although it is still unknown if the signature of a knot can be expressed in terms of its finite-type invariants.

### Chord Diagrams and the Fundamental Theorem

The top derivatives of a multivariable polynomial form a system of constants which determine the polynomial up to polynomials of lower degree. Likewise the $m$th derivative $V^{(m)} := V(\times \overset{m}{\cdots} \times)$ of a type $m$ invariant $V$ is a constant (for $V(\times \overset{m}{\cdots} \overcrossing\times) - V(\times \overset{m}{\cdots} \undercrossing\times) = V(\times \overset{m+1}{\cdots} \times) = 0$ so $V^{(m)}$ is blind to 3D topology), and likewise $V^{(m)}$ determines $V$ up to invariants of lower type. Hence, a primary tool in the study of finite-type invariants is the study of the "top derivative" $V^{(m)}$, also known as "the weight system of $V$."

Blind to 3D topology, $V^{(m)}$ only sees the combinatorics of the circle that parametrizes an $m$-singular knot. On this circle, there are $m$ pairs of points that are pairwise identified in the image; one indicates those by drawing a circle with $m$ chords marked (an "$m$-chord diagram") (see Figure 1).

**Definition 4** Let $\mathcal{D}_m$ denote the space of all formal linear combinations with rational coefficients of $m$-chord diagrams. Let $\mathcal{A}_m^r$ be the quotient of $\mathcal{D}_m$ by all $4T$ and $FI$ relations as drawn in Figure 2 (full details are given in, e.g., Bar-Natan (1995)), and let $\hat{\mathcal{A}}^r$ be the graded completion of $\mathcal{A} := \bigoplus_m \mathcal{A}_m^r$. Let $\mathcal{A}_m, \mathcal{A}$, and $\hat{\mathcal{A}}$ be the same as $\mathcal{A}_m^r, \mathcal{A}^r$, and $\hat{\mathcal{A}}^r$ but without imposing the $FI$ relations.

**Theorem 5** (The fundamental theorem)

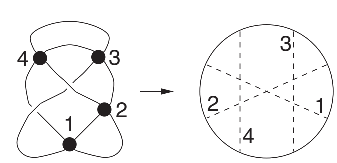



**Figure 1** A 4-singular knot and its corresponding chord diagram.

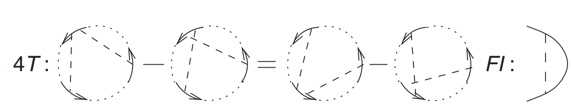



**Figure 2** The 4T and FI relations.

- (*Easy part*). *If V is a rational valued type m-invariant then* $V^{(m)}$ *defines a linear functional on* $\mathcal{A}^r_m$. *If in addition* $V^{(m)} \equiv 0$, *then V is of type* $m-1$.
- (*Hard part*). *For any linear functional W on* $\mathcal{A}^r_m$, *there is a rational valued type m invariant V so that* $V^{(m)} = W$.

Thus, to a large extent, the study of finite-type invariants is reduced to the finite (though super-exponential in $m$) algebraic study of $\mathcal{A}^r_m$. A similar theorem reduces the study of finite-type invariants of framed knots to the study of $\mathcal{A}_m$.

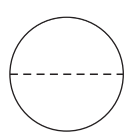
### The Structure of $\mathcal{A}$

Knots can be multiplied (the "connected sum" operation) and knot invariants can be multiplied. This structure interacts well with finite-type invariants and induces the following structure on $\mathcal{A}^r$ and $\mathcal{A}$:

**Theorem 6** (Kontsevich 1993, Bar-Natan 1995, Willerton 1996, Chmutov *et al.* 1994). $\mathcal{A}^r$ *and* $\mathcal{A}$ *are commutative and cocommutative graded bialgebras* (*i.e.*, *each carries a commutative product and a compatible cocommutative coproduct*). *Thus*, *both* $\mathcal{A}^r$ *and* $\mathcal{A}$ *are graded polynomial algebras over their spaces of primitives*, $\mathcal{P}^r = \oplus_m \mathcal{P}^r_m$ *and* $\mathcal{P} = \oplus_m \mathcal{P}_m$.

Framed knots differ from knots only by a single integer parameter (the "self-linking," itself a type 1 invariant). Thus, $\mathcal{P}^r$ and $\mathcal{P}$ are also closely related.

**Theorem 7** (Bar-Natan 1995). $\mathcal{P} = \mathcal{P}^r \oplus \langle \theta \rangle$, *where* $\theta$ *is the unique 1-chord diagram*:



### Bounds and Computational Results

Table 1 shows the number of type $m$-invariants of knots and framed knots modulo type $m-1$ invariants ($\dim \mathcal{A}^r_m$ and $\dim \mathcal{A}_m$) and the number of multiplicative generators of the algebra $\mathcal{A}$ in degree $m$ ($\dim \mathcal{P}_m$) for $m \le 12$. Some further tabulated results are in Bar-Natan (1996).

Little is known about these dimensions for large $m$. There is an explicit conjecture in Broadhurst (1997), but no progress has been made in the direction of proving or disproving it. The best asymptotic bounds available are the following.

**Theorem 8** *For large m*, $\dim \mathcal{P}_m > e^{c\sqrt{m}}$ (*for any fixed* $c < \pi\sqrt{2/3}$) *and* $\dim \mathcal{A}_m < 6^m m! \sqrt{m}/\pi^{2m}$ (Stoimenow 1998, Zagier 2001).

### Jacobi Diagrams and the Relation with Lie Algebras

Much of the richness of finite-type invariants stems from their relationship with Lie algebras. Theorem 9 below suggests this relationship on an abstract level, Theorem 10 makes that relationship concrete, and Theorem 12 makes it a bit deeper.

**Theorem 9** (Bar-Natan 1995). *The algebra* $\mathcal{A}$ *is isomorphic to the algebra* $\mathcal{A}^t$ *generated by "Jacobi diagrams in a circle"* (*chord diagrams that are also allowed to have oriented internal trivalent vertices*) *modulo the AS, STU, and IHX relations* (see Figure 3).

Thinking of trivalent vertices as graphical analogs of the Lie bracket, the AS relation becomes the anti-commutativity of the bracket, STU becomes the equation $[x, y] = xy - yx$, and IHX becomes the Jacobi identity. This analogy is made concrete within the proof of the following:

**Theorem 10** (Bar-Natan 1995). *Given a finite-dimensional metrized Lie algebra* $\mathfrak{g}$ (*e.g.*, *any semi-simple Lie algebra*), *there is a map* $\mathcal{T}_{\mathfrak{g}} : \mathcal{A} \to \mathcal{U}(\mathfrak{g})^{\mathfrak{g}}$ *defined on* $\mathcal{A}$ *and taking values in the invariant part* $\mathcal{U}(\mathfrak{g})^{\mathfrak{g}}$ *of the universal enveloping algebra* $\mathcal{U}(\mathfrak{g})$ *of* $\mathfrak{g}$. *Given also a finite-dimensional representation R of* $\mathfrak{g}$ *there is a linear functional* $W_{\mathfrak{g},R} : \mathcal{A} \to \mathbb{Q}$.

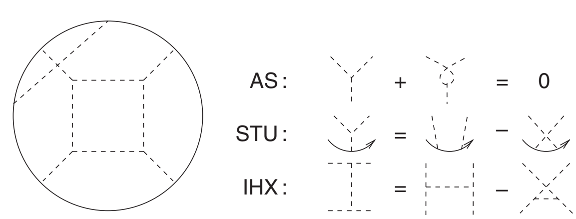



**Figure 3** A Jacobi diagram in a circle and the AS, STU, and IHX relations.

**Table 1** Some dimensions of spaces of finite type invariants

| $m$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\dim \mathcal{A}^r_m$ | 1 | 0 | 1 | 1 | 3 | 4 | 9 | 14 | 27 | 44 | 80 | 132 | 232 |
| $\dim \mathcal{A}_m$ | 1 | 1 | 2 | 3 | 6 | 10 | 19 | 33 | 60 | 104 | 184 | 316 | 548 |
| $\dim \mathcal{P}_m$ | 0 | 1 | 1 | 1 | 2 | 3 | 5 | 8 | 12 | 18 | 27 | 39 | 55 |

Source: Bar-Natan (1995); Kneissler (1997).

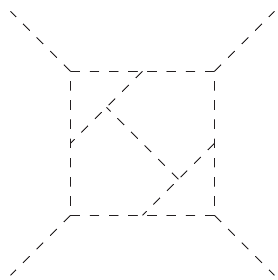


**Figure 4** A free Jacobi diagram.

The last assertion along with Theorem 5 show that associated with any $\mathfrak{g}$, $R$, and $m$, there is a weight system and hence a knot invariant. Thus, knots are unexpectedly linked with Lie algebras.

The hope (Bar-Natan 1995) that all finite-type invariants arise in this way was dashed by Vogel (1997, 1999) and Lieberum (1999). But finite-type invariants that do not arise in this way remain rare and not well understood.

The Poincaré–Birkhoff–Witt (PBW) theorem of the theory of Lie algebras says that the obvious "symmetrization" map $\chi_{\mathfrak{g}}: \mathcal{S}(\mathfrak{g}) \to \mathcal{U}(\mathfrak{g})$ from the symmetric algebra $\mathcal{S}(\mathfrak{g})$ of a Lie algebra $\mathfrak{g}$ to its universal enveloping algebra $\mathcal{U}(\mathfrak{g})$ is a $\mathfrak{g}$-module isomorphism. The following definition and theorem form a diagrammatic counterpart of this theorem:

**Definition 11** Let $\mathcal{B}$ be the space of formal linear combinations of "free Jacobi diagrams" (Jacobi diagrams as before, but with unmarked univalent ends ("legs") replacing the circle; see an example in Figure 4), modulo the AS and IHX relations of before. Let $\chi: \mathcal{B} \to \mathcal{A}$ be the symmetrization map which maps a $k$-legged free Jacobi diagram to the average of the $k$! ways of planting these legs along a circle.

**Theorem 12** (Diagrammatic PBW; Kontsevich 1993, Bar-Natan 1995). *$\chi$ is an isomorphism of vector spaces. Furthermore, fixing a metrized $\mathfrak{g}$ there is a commutative square as in* Figure 5.

Note that $\mathcal{B}$ can be graded (by half the number of vertices in a Jacobi diagram) and that $\chi$ respects degrees so it extends to an isomorphism $\chi: \hat{\mathcal{B}} \to \hat{\mathcal{A}}$ of graded completions.

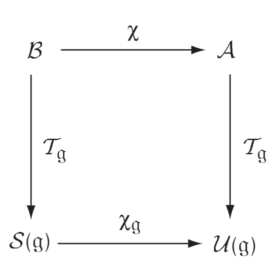



**Figure 5** The diagrammatic PBW isomorphism and its classical counterpart.

## Proofs of the Fundamental Theorem

The heart of all known proofs of Theorem 5 is always a construction of a "universal finite-type invariant" (see below); it is simple to show that the existence of a universal finite-type invariant is equivalent to Theorem 5.

**Definition 13** A universal finite-type invariant is a map $Z: \{\text{knots}\} \to \hat{\mathcal{A}}^r$ whose extension to singular knots satisfies $Z(K) = D + (\text{higher degrees})$ whenever a singular knot $K$ and a chord diagram $D$ are related as discussed before.

### The Kontsevich Integral

The first construction of a universal finite-type invariant was given by Kontsevich (1993) (see also Bar-Natan (1995) and Chmutov and Duzhin (2001)). It is known as "the Kontsevich integral" and up to a normalization factor it is given by

$$Z_1(K) = \sum_{m=0}^{\infty} \frac{1}{(2\pi i)^m} \int\!\!\!\!\!\!\sum_{\substack{t_1<\cdots<t_m \\ P=\{(z_i,z_i')\}}} (-1)^{\#P_\downarrow} D_P \bigwedge_{i=1}^{m} \frac{dz_i - dz_i'}{z_i - z_i'}$$

where the relationship between the knot $K$, the pairing $P$, the real variables $t_i$, the complex variables $z_i$ and $z_i'$, and the chord diagram $D_P$ is summarized in Figure 6 (the symbol $\int\!\!\!\!\!\!\sum$ means "sum over all discrete variables and integrate over all continuous variables.")

The Kontsevich integral arises from studying the holonomy of the Knizhnik–Zamolodchikov equation of conformal field theory (Knizhnik and Zamolodchikov 1984). When evaluating $Z_1$, one encounters multiple $\zeta$-numbers (Le-Murakami 1995) in a substantial way, and the proof that the end result is rational is quite involved (Le-Murakami 1996) and relies on deep results about associators and quasitriangular quasi-Hopf algebras (Drinfel'd 1990, 1991). Employing the same techniques, in Le-Murakami

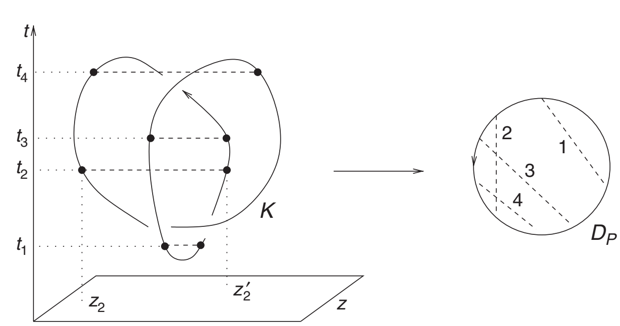



**Figure 6** The key ingredients of the Kontsevich integral.

(1996), it is also shown that the composition of $W_{\mathfrak{g},R} \circ Z_1$ precisely reproduces the Reshetikhin–Turaev invariants (Reshetikhin and Turaev 1990).

### Perturbative Chern–Simons–Witten Theory and Configuration Space Integrals

Historically, the first approach to the construction of a universal finite-type invariant was to use perturbation theory with the Chern–Simons–Witten topological quantum field theory; this is also how the relationship with Lie algebras first arose. But taming the integrals involved turned out to be difficult and working constructions using this approach appeared only a bit later.

In short, one writes a perturbative expansion for the large $k$ asymptotics of the Chern–Simons–Witten path integral for some metrized Lie algebra $\mathfrak{g}$ with a Wilson loop in some representation $R$ of $\mathfrak{g}$,

$$\int_{\mathfrak{g}\text{-connections}} \mathcal{D}A \, \mathrm{tr}_R \, \mathrm{hol}_K(A) \times \exp\left[\frac{\mathrm{i}k}{4\pi}\int_{\mathbb{R}^3} \mathrm{tr}\left(A \wedge \mathrm{d}A + \frac{2}{3} A \wedge A \wedge A\right)\right]$$

The result is of the form

$$\sum_{D:\ \text{Feynman diagram}} W_{\mathfrak{g}(D),R} \sum\!\!\!\!\!\!\int \mathcal{E}(D)$$

where $\mathcal{E}(D)$ is a very messy integral expression and the diagrams $D$ as well as the weights $W_{\mathfrak{g}(D),R}$ were already discussed before. Replacing $W_{\mathfrak{g}(D),R}$ by simply $D$ in the above formula, we get an expression with values in $\hat{\mathcal{A}}$:

$$Z_2(K) := \sum_D D \sum\!\!\!\!\!\!\int \mathcal{E}(D) \in \hat{\mathcal{A}}$$

For formal reasons $Z_2(K)$ ought to be a universal finite-type invariant, and after much work taming the $\mathcal{E}(D)$ factors and after multiplying by a further framing-dependent renormalization term $Z^{\text{anomaly}}$, the result is indeed a universal finite-type invariant.

Upon further inspection, the $\mathcal{E}(D)$ factors can be reinterpreted as integrals of certain spherical volume forms on certain (compactified) configuration spaces (Bott–Taubes 1994). These integrals can be further interpreted as counting certain "tinker toy constructions" built on top of $K$ (Thurston 1995). The latter viewpoint makes the construction of $Z_2$ visually appealing (Bar-Natan 2000), but there is no satisfactory write-up of this perspective yet.

We note that the precise form of the renormalization term $Z^{\text{anomaly}}$ remains an open problem. An appealing conjecture is that $Z^{\text{anomaly}} = \exp(1/2)\ominus$. If this is true then $Z_2 = Z_1$ (Poirier 1999); but the conjecture is only verified up to degree 6 (Lescop 2001) (there is also an unconfirmed verification to all orders (Yang 1997)).

The most important open problem about perturbative Chern–Simons–Witten theory is not directly about finite-type invariants, but it is nevertheless worthwhile to recall it here:

**Problem 14** *Does the perturbative expansion of the Chern–Simons–Witten theory converge* (*or is asymptotic to*) *the exact solution due to Witten (1989) and Reshetikhin and Turaev* (1990) *when the parameter k converges to infinity?*

### Associators and Trivalent Graphs

There is also an entirely algebraic approach for the construction of a universal finite-type invariant $Z_3$. The idea is to find some algebraic context within which knot theory is finitely presented – that is, presented by finitely many generators subject to finitely many relations. If the algebraic context at hand is compatible with the definitions of finite-type invariants and of chord diagrams, one may hope to define $Z_3$ by defining it on the generators in such a way that the relations are satisfied. Thus, the problem of defining $Z_3$ is reduced to finding finitely many elements of $\mathcal{A}$-like spaces which solve certain finitely many equations.

A concrete realization of this idea is in Le-Murakami (1996) and Bar-Natan (1997) (following ideas from Drinfel'd (1990, 1991) on quasitriangular quasi-Hopf algebras). The relevant "algebraic context" is a category with certain extra operations, and within it, knot theory is generated by just two elements, the braiding ⤫ and the re-association ⟋. Thus, to define $Z_3$ it is enough to find $R = Z_3(⤫)$ and "an associator" $\Phi = Z_3(⟋)$ which satisfy certain normalization conditions as well as the pentagon and hexagon equations:

$$\Phi^{123} \cdot (1\Delta 1)(\Phi) \cdot \Phi^{234} = (\Delta 11)(\Phi) \cdot (11\Delta)(\Phi)$$

$$(\Delta 1)(R^{\pm}) = \Phi^{123}(R^{\pm})^{23}(\Phi^{-1})^{132}(R^{\pm})^{13}\Phi^{312}$$

As it turns out, the solution for $R$ is easy and nearly canonical. But finding an associator $\Phi$ is rather difficult. There is a closed-form integral expression $\Phi^{KZ}$ due to Drinfel'd (1990) but one encounters the same not-too-well-understood multiple $\zeta$ numbers. There is a rather complicated iterative procedure for finding an associator (Drienfel'd 1991, Bar-Natan 1998). On a computer it had been used to find an associator up to degree 7. There is also closed-form associator that works only with the Lie superalgebra gl(1|1) (Lieberum 2002). But it remains an open

problem to find a closed-form formula for a rational associator (existence by Drienfel'd (1991) and Bar- Natan (1998)).

On the positive side, we should note that the end result, the invariant $Z_3$, is independent of the choice of $\Phi$ and that $Z_3 = Z_1$.

There is an alternative (more symmetric and intrinsically three dimensional, but less well-documented) description of the theory of associators in terms of knotted trivalent graphs (Bar-Natan and Thurston). There ought to be a perturbative invariant associated with knotted trivalent graphs in the spirit of the last subsection and such an invariant should lead to a simple proof that $Z_2 = Z_3 = Z_1$. But the $\mathcal{E}(D)$ factors remain untamed in this case.

#### Step-by-Step Integration

The last approach for proving the fundamental theorem is the most natural and historically the first. But here it is last because it is yet to lead to an actual proof. A weight system $W : \mathcal{A}^r_m \to \mathbb{Q}$ is an invariant of $m$-singular knots. We want to show that it is the $m$th derivative of an invariant $V$ of nonsingular knots. It is natural to try to integrate $W$ step by step, first finding an invariant $V^{m-1}$ of $(m-1)$-singular knots whose derivative in the sense of [1] is $W$, then an invariant $V^{m-2}$ of $(m-2)$-singular knots whose derivative is $V^{m-1}$, and so on all the way up to an invariant $V^0 = V$ whose $m$th derivative will then be $W$. If proven, the following conjecture would imply that such an inductive procedure can be made to work:

**Conjecture 15** (*Hutchings 1998*). *If $V^r$ is a once-integrable invariant of $r$-singular knots, then it is also twice integrable. That is, if there is an invariant $V^{r-1}$ of $(r-1)$-singular knots whose derivative is $V^r$, then there is an invariant $V^{r-2}$ of $(r-2)$-singular knots whose second derivative is $V^r$.*

Hutchings (1998) reduced this conjecture to a certain appealing topological statement and further to a certain combinatorial-algebraic statement about the vanishing of a certain homology group $H^1$ which is probably related to Kontsevich's graph homology complex (Kontsevich 1994) (Kontsevich's $H^0$ is $\mathcal{A}$, so this is all in the spirit of many deformation theory problems where $H^0$ enumerates infinitesimal deformations and $H^1$ is the obstruction to globalization). Hutchings (1998) was also able to prove the vanishing of $H^1$ (and hence reprove the fundamental theorem) in the simpler case of braids. But no further progress has been made along these lines since then.

## Some Further Directions

We would like to touch upon a number of significant further directions in the theory of finite-type invariants and describe each of those only briefly; the reader is referred to the "Further reading" section for more information.

#### The Original "Vassiliev" Perspective

V A Vassiliev came to the study of finite-type knot invariants by studying the infinite-dimensional space of all immersions of a circle into $\mathbb{R}^3$ and the topology of the "discriminant," the locus of all singular immersions within the latter space (Vassiliev 1990, 1992). Vassiliev studied the topology of the complement of the discriminant (the space of embeddings) using a certain spectral sequence and found that certain terms in it correspond to finite-type invariants. This later got related to the Goodwillie calculus and back to the configuration spaces discussed in the last section. See Volic (2004).

#### Interdependent Modifications

The standard definition of finite-type invariants is based on modifying a knot by replacing over (or under) crossings with under (or over) crossings. Goussarov (1998) generalized this by allowing arbitrary modifications done to a knot – just take any segment of the knot and move it anywhere else in space. The resulting new "finite-type" theory turns out to be equivalent to the old one though with a factor of 2 applied to the grading (so an "old" type $m$ invariant is a "new" type $2m$ invariant and vice versa). (see also Bar-Natan (2001) and Conant (2003)).

#### $n$-Equivalence, Commutators, and Claspers

While little is known about the overall power of finite-type invariants, much is known about the power of type $n$-invariants for any given $n$. Goussarov (1993) defined the notion of $n$-equivalence: two knots are said to be "$n$-equivalent" if all their type $n$-invariants are the same. This equivalence relation is well understood both in terms of commutator subgroups of the pure braid group (Stanford 1998, Ng and Stanford 1999) and in terms of Habiro's calculus of surgery over "claspers" (Habiro 2000) (the latter calculus also gives a topological explanation for the appearance of Jacobi diagrams). In particular, already Goussarov (1993) shows that the set of equivalence classes of knots modulo $n$-equivalence is a finitely generated abelian group $G_n$ under the operation of connected sum, and the rank of that group is equal to the dimension of the space of type $n$-invariants.

Ng (1998) has shown that ribbon knots generate an index 2 subgroup of $G_n$.

### Polynomiality and Gauss Sums

Goussarov (1998) (see also Goussarov–Polyak–Viro (2001)) found an intriguing way to compute finite-type invariants from a Gauss diagram presentation of a knot, showing in particular that finite-type invariants grow as polynomials in the number of crossings $n$ and can be computed in polynomial time in $n$ (though actual computer programs are still missing!).

Gauss diagrams are obtained from knot diagrams in much of the same way as Chord diagrams are obtained from singular knots, except all crossings are counted and not just the double points, and certain over/under and sign information is associated with each crossing/chord so that the knot diagram can be recovered from its Gauss diagram. In the example below (**Figure 8**), we also dashed a subdiagram of the Gauss diagram equivalent to the chord diagram shown in **Figure 7**.

If $G$ is a Gauss diagram and $D$ is a chord diagram, then let $\langle D, G\rangle$ be the number of subdiagrams of $G$ equivalent to $D$, counted with appropriate signs (to be precise, we also need to base the diagrams involved and count subdiagrams that respect the basing).

**Theorem 16** (Goussarov 1998, Goussarov *et al.* 2000). *If $V$ is a type $m$ invariant, then there are finitely many (based) chord diagrams $D_i$ with at most $m$ chords and rational numbers $\alpha_i$ so that $V(K) = \sum_i \alpha_i \langle D_i, G\rangle$ whenever $G$ is a Gauss diagram representing a knot $K$.*

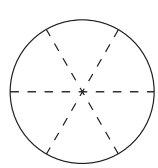


**Figure 7** A chord diagram.

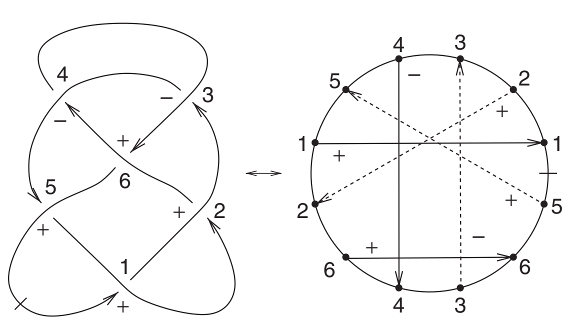



**Figure 8** A knot and its Gauss diagram.

### Computing the Kontsevich Integral

While the Kontsevich integral $Z_1$ is a cornerstone of the theory of finite-type invariants, it has been computed for surprisingly few knots. Even for the unknot, the result is nontrivial:

**Theorem 17** ("Wheels," Bar-Natan *et al.* 2000, 2003). *The framed Kontsevich integral of the unknot, $Z_1^F(\bigcirc)$, expressed in terms of diagrams in $\hat{\mathcal{B}}$, is given by $\Omega = \exp_{\cup}\sum_{n=1}^{\infty} b_{2n}\omega_{2n}$, where the "modified Bernoulli numbers" $b_{2n}$ are defined by the power series expansion $\sum_{n=0}^{\infty} b_{2n}x^{2n} = (1/2)\log(\sinh x/2)/(x/2)$, the "2n-wheel" $\omega_{2n}$ is the free Jacobi diagram made of a 2n-gon with 2n legs (so, e.g., $\omega_6 =$ ), and where* $\exp_{\cup}$ *means "exponential in the disjoint union sense."*

Closed-form formulas have also been given for the Kontsevich integral of framed unknots, the Hopf link and Hopf chains.

Theorem 17 has a companion that utilizes the same element $\Omega$, the "wheeling" theorem (Bar-Natan *et al.* 2000, 2003). The wheeling theorem "upgrades" the vector space isomorphism $\chi: \mathcal{B} \to \mathcal{A}$ to an algebra isomorphism and is related to the Duflo isomorphism of the theory of Lie algebras. It is amusing to note that the wheeling theorem (and hence Duflo's theorem in the metrized case) follows using finite-type techniques from the "$1 + 1 = 2$ on an abacus" identity (**Figure 9**).

### Taming the Kontsevich Integral

While explicit calculations are rare, there is a nice structure theorem for the values of the Kontsevich integral, saying that for a knot $K$ and up to any fixed number of loops in the Jacobi diagrams, $\chi^{-1}Z_1(K)$ can be described by finitely many rational functions (with denominators powers of the Alexander polynomial) which dictate the placement of the legs. This structure theorem was conjectured in Rozansky (2003), proven in Kricker (2000), and partially generalized to links in Garoufalidis and Kricker (2004).

### The Rozansky–Witten Theory

One way to construct linear functionals on $\mathcal{A}$ (and hence finite-type invariants) is using Lie algebras and representations as discussed earlier; much of our insight about $\mathcal{A}$ comes this way. But there is another construction for such functionals (and hence invariants), due to Rozansky and Witten (1997), using contractions of curvature tensors on hyper-Kähler

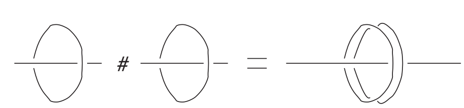



**Figure 9** A knot theoretic $1 + 1 = 2$.

manifolds. Very little is known about the Rozansky–Witten approach; in particular, it is not known if it is stronger or weaker than the Lie algebraic approach. For an application of the Rozansky–Witten theory back to hyper-Kähler geometry check Hitchin and Sawon (2001), and for a unification of the Rozansky–Witten approach with the Lie algebraic approach (albeit at a categorical level) check Roberts and Willeton (in preparation).

### The Melvin–Morton Conjecture and the Volume Conjecture

The Melvin–Morton conjecture (stated Melvin and Morton (1995), proven Bar-Natan and Garoufalidis (1996)) says that the Alexander polynomial can be read off certain coefficients of the colored Jones polynomial. The Kashaev–Murakami–Murakami volume conjecture (stated Kashaev (1997) and J Murakami and H Murakami (2001), unproven) says that a certain asymptotic growth rate of the colored Jones polynomial is the hyperbolic volume of the knot complement.

Both conjectures are not directly about finite-type invariants but both have ramifications to the theory of finite-type invariants. The Melvin–Morton conjecture was first proven using finite-type invariants and several later proofs and generalizations (see (Bar-Natan)) also involve finite-type invariants. The volume conjecture would imply, in particular, that the hyperbolic volume of a knot complement can be read from that knot's finite-type invariants, and hence finite-type invariants would be at least as strong as the volume invariant.

A particularly noteworthy result and direction for further research is Gukov's (preprint) recent unification of these two conjectures under the Chern–Simons umbrella (along with some relations to three-dimensional quantum gravity).

## Beyond Knots

For lack of space, we have restricted ourselves here to a discussion of finite-type invariants of knots. But the basic "differentiation" idea of the first section calls for generalization, and indeed it has been generalized extensively. We will only make a few quick comments.

Finite-type invariants of homotopy links (links where each component is allowed to move across itself freely) and of braids are extremely well behaved. They separate, they all come from Lie algebraic constructions and in the case of braids, step-by-step integration as discussed previously works (for homotopy links the issue was not studied).

Finite-type invariants of 3-manifolds and especially of integral and rational homology spheres have been studied extensively and the picture is nearly a complete parallel of the picture for knots. There are several competing definitions of finite-type invariants, and they all agree up to regrading. There are weight systems and they are linear functionals on a space $\mathcal{A}(\emptyset)$ which is a close cousin of $\mathcal{A}$ and $\mathcal{B}$ and is related to Lie algebras and hyper-Kähler manifolds in a similar way. There is a notion of a "universal" invariant, and there are several constructions; they all agree or are conjectured to agree, and they are related to the Chern–Simons–Witten theory.

Finite-type invariants were studied for several other types of topological objects, including knots within other manifolds, higher-dimensional knots, virtual knots, plane curves and doodles and more (see Bar-Natan).

## Acknowledgmnts

This work was partially supported by NSERC grant RGPIN 262178. Electronic versions: http://www.math.toronto.edu/~drorbn/papers/EMP/ and arXiv:math.GT/0408182. Copyright is retained by the author. The author wishes to thank O Dasbach and A Savage for comments and corrections, especially pertaining to Problem 2.

*See also:* Finite-Type Invariants of 3-Manifolds; Knot Invariants and Quantum Gravity; Kontsevich Integral; Mathematical Knot Theory; von Neumann Algebras: Introduction, Modular Theory, and Classification Theory.

## Further Reading

The first reference is an extensive bibliography of finite invariants, listing more than 550 references. To save space, references for this article are only cited above stating the authors' full last name(s) and approximate year of publication.

Bar-Natan D Bibliography of Vassiliev invariants, http://www.math.toronto.edu/~drorbn/VasBib.

Bar-Natan D (1998) On associators and the Grothendieck–Teichmuller group I. *Selecta Mathematica* 4: 183–212.

Bar-Natan D and Thurston D Algebraic Structures on knotted objects and universal finite type invariants (in prepartion), http://www.math.toronto.edu/~drorbn/papers/AlgebraicStructures.

Drinfel'd VG (1990) Quasi-Hopf algebras. *Leningrad Mathematical Journal* 1: 1419–1457.

Drinfel'd VG (1991) On quasitriangular Quasi-Hopf algebras and a group closely connected with Gal($\mathbb{Q}/\mathbb{Q}$). *Leningrad Mathematical Journal* 2: 829–860.

Gukov S Three-dimensional quantum gravity, Chern–Simons theory, and the A-polynomial. *Preprint*, arXiv:hep-th/0306165.

Hitchin N and Sawon J (2001) Curvature and characteristic numbers of hyper-Kähler manifolds. *Duke Mathematical Journal* 106–3: 599–615, arXiv:math.DG/9908114.

Kashaev RM (1997) The hyperbolic volume of knots from quantum dilogarithm. *Letters in Mathematical Physics* 39: 269–275, arXiv:q-alg/9601025.

Knizhnik VG and Zamolodchikov AB (1984) Current algebra and Wess–Zumino model in two dimensions. *Nuclear Physics B* 247: 83–103.
Murakami J and Murakami H (2001) The colored Jones polynomials and the simplicial volume of a knot. *Acta Mathematica* 186(1): 85–104, arXiv:math.GT/9905075.
Reshetikhin Yu N and Turaev VG (1990) Ribbon graphs and their invariants derived from quantum groups. *Communications in Mathematical Physics* 127: 1–26.
Witten E (1989) Quantum field theory and the Jones polynomial. *Communications in Mathematical Physics* 121: 351–399.

# Finite-Type Invariants of 3-Manifolds

**T T Q Lê**, Georgia Institute of Technology, Atlanta, GA, USA



## Introduction

### Physics Background and Motivation

Suppose $G$ is a semisimple compact Lie group and $M$ a closed oriented 3-manifold. Witten (1989) defined quantum invariants by the path integral over all $G$-connections $A$:

$$Z(M, G; k) := \int \exp(\sqrt{-1}k\, \mathrm{CS}(A))\mathcal{D}A$$

where $k$ is an integer and $\mathrm{CS}(A)$ is the Chern–Simons functional,

$$\mathrm{CS}(A) = \frac{1}{4\pi}\int_M \mathrm{tr}\left(A \wedge \mathrm{d}A + \frac{2}{3}A^3\right)$$

The path integral is not mathematically rigorous. According to the stationary-phase approximation in quantum field theory, in the limit $k \to \infty$ the path integral decomposes as a sum of contributions from the flat connections:

$$Z(M, G; k) \sim \sum_{\text{flat connections } f} Z^{(f)}(M, G; k) \quad \text{as } k \to \infty$$

Each contribution is $\exp(2\pi\sqrt{-1}\,k\,\mathrm{CS}(f))$ times a power series in $1/k$. The contribution from the trivial connection is important, especially for rational homology 3-spheres, and the coefficients of the powers $(1/k)^n$, calculated using $(n+1)$-loop Feynman diagrams by quantum field theory techniques, are known as perturbative invariants.

### Mathematical Theories

A mathematically rigorous theory of quantum invariants $Z(M, G\,; k)$ was pioneered by Reshetikhin and Turaev in 1990 (see Turaev (1994)). A number-theoretical expansion of the quantum invariants into power series that should correspond to the perturbative invariants was given by Ohtsuki (in the case of $\mathrm{sl}_2$, and general simple Lie algebras by the author) in 1994. This led him to introducing finite-type invariant (FTI) theory for 3-manifolds. A universal perturbative invariant was constructed by Le–Murakami–Ohtsuki (LMO) in 1995; it is universal for both finite-type invariants and quantum invariants, at least for homology 3-spheres. Rozansky in 1996 defined perturbative invariants using Gaussian integral, very close in the spirit to the original physics point of view. Later Habiro (for $\mathrm{sl}_2$ and Habiro and the author for all simple Lie algebras) found a finer expansion of quantum invariants, known as the cyclotomic expansion, but no physics origin is known for the cyclotomic expansion. The cyclotomic expansion helps to show that the LMO invariant dominates all quantum invariants for homology 3-spheres.

The purpose of this article is to give an overview of the mathematical theory of finite-type and perturbative invariants of 3-manifolds.

### Conventions and Notations

All vector spaces are assumed to be over the ground field $\mathbb{Q}$ of rational numbers, unless otherwise stated. For a graded space $A$, let $\mathrm{Gr}_n A$ be the subspace of grading $n$ and $\mathrm{Gr}_{\leq n}A$ the subspace of grading $\leq n$. For $x \in A$, let $\mathrm{Gr}_n x$ and $\mathrm{Gr}_{\leq n}x$ be the projections of $x$ onto, respectively, $\mathrm{Gr}_n A$ and $\mathrm{Gr}_{\leq n}A$.

All 3-manifolds are supposed to be closed and oriented. A 3-manifold $M$ is an integral homology 3-sphere ($\mathbb{Z}$HS) if $H_1(M, \mathbb{Z}) = 0$; it is a rational homology 3-sphere ($\mathbb{Q}$HS) if $H_1(M, \mathbb{Q}) = 0$. For a framed link $L$ in a 3-manifold $M$ denote $M_L$ the 3-manifold obtained from $M$ by surgery along $L$ (see e.g., Turaev (1994)).

## Finite-Type Invariants

After its introduction by Ohtsuki in 1994, the theory of FTIs of 3-manifolds has been developed rapidly by many authors. Later Goussarov and Habiro independently introduced clasper calculus, or $Y$-surgery, which provides a powerful geometric technique and deep insight in the theory. $Y$-surgery, corresponding to the commutator in group theory, naturally gives rise to 3-valent graphs.

### Generality on FTIs

**Decreasing filtration** In a theory of FTIs, one considers a class of objects, and a "good" decreasing filtration $\mathcal{F}_0 \supset \mathcal{F}_1 \supset \mathcal{F}_2 \supset \cdots$ on the vector space $\mathcal{F} = \mathcal{F}_0$ spanned by these objects. An invariant of the objects with values in a vector space is of order less than or equal to $n$ if its restriction to $\mathcal{F}_{n+1}$ is 0; it is of finite type if it is of order $\le n$ for some $n$. An invariant has order $n$ if it is of order $\le n$ but not $\le n-1$. Good here means at least the space of FTI of each order is finite dimensional. It is desirable to have an algorithm of polynomial time to calculate every FTI. In addition, one wants the set of FTIs to separate the objects (completeness).

The space of invariants of order $\le n$ can be identified with the dual space of $\mathcal{F}_0/\mathcal{F}_{n+1}$; its subspace $\mathcal{F}_n/\mathcal{F}_{n+1}$ is isomorphic to the space of invariants of order $\le n$ modulo the space of invariants of order $\le n-1$. Informally, one can say that $\mathcal{F}_n/\mathcal{F}_{n+1}$ is more or less the set of invariants of order $n$.

**Elementary moves, the knot case** Usually the filtrations are defined using "independent elementary moves." For the class of knots the elementary move is given by crossing change. Any two knots can be connected by a finite sequence of such moves. The idea is if $K, K' \in \mathcal{F}_n$, the $n$th term of the filtration, then $K - K' \in \mathcal{F}_{n+1}$, where $K'$ is obtained from $K$ by an elementary move. Formal definition is as follows. Suppose $S$ is a set of double points of a knot diagram $D$. Let

$$[D, S] = \sum_{S' \subset S} (-1)^{\#S'} D_{S'}$$

where the sum is over all subsets $S'$ of $S$, including the empty set, $D_{S'}$ is the knot obtained by changing the crossing at every point in $S'$, and $\#S'$ is the number of elements of $S'$. Then $\mathcal{F}_n$ is the vector space spanned by all elements of the form $[D, S]$ with $\#S = n$. For the knot case, the Kontsevich integral is an invariant that is universal for all FTIs (see Bar-Natan (1995)).

### Ohtsuki's Definition of FTIs for $\mathbb{Z}$HS

An elementary move here is a surgery along a knot: $M \to M_K$, where $K$ is a framed knot in a $\mathbb{Z}$HS $M$. A collection of moves corresponds to surgery on a framed link. To always remain in the class of $\mathbb{Z}$HS we need to restrict ourselves to unit-framed and algebraically split links, that is, framed links in $\mathbb{Z}$HS each component of which has framing $\pm 1$ and the linking number of every two components is 0. It is easy to prove that a link $L$ in a $\mathbb{Z}$HS $M$ is unit-framed and algebraically split if and only if $M_{L'}$ is a $\mathbb{Z}$HS for every sublink $L'$ of $L$. For a unit-framed, algebraically split link $L$ in a $\mathbb{Z}$HS $M$ define

$$[M, L] = \sum_{L' \subset L} (-1)^{|\#L'|} M_{L'}$$

which is an element in the vector space $\mathcal{M}$ freely spanned by $\mathbb{Z}$HS.

For a non-negative integer $n$ let $\mathcal{F}_n^{\mathrm{AS}}$ be the subspace of $\mathcal{M}$ spanned by $[M, L]$ with $\#L = n$. Then the descending filtration $\mathcal{M} = \mathcal{F}_0^{\mathrm{AS}} \supset \mathcal{F}_1^{\mathrm{AS}} \supset \mathcal{F}_2^{\mathrm{AS}} \supset \cdots$ defines a theory of FTIs on the class of $\mathbb{Z}$HS.

**Theorem 1**

(i) (Ohtsuki) *The dimension of* $\mathcal{F}_n(\mathcal{M})$ *is finite for every* $n$.
(ii) (Garoufalidis–Ohtsuki) *One has* $\mathcal{F}_{3n+1}(\mathcal{M}) = \mathcal{F}_{3n+2}(\mathcal{M}) = \mathcal{F}_{3n+3}(\mathcal{M})$.

The orders of FTIs in this theory are multiples of 3. The first nontrivial invariant, which is the only (up to scalar) invariant of degree 3, is the Casson invariant.

### The Goussarov–Habiro Definition

**$Y$-surgery or clasper surgery** Consider the standard $Y$-graph $Y$ and a small neighborhood $N(Y)$ of it in the standard $\mathbb{R}^2$ (see Figure 1). Denote by $L(Y)$ the six-component framed link diagram in $N(Y) \subset \mathbb{R}^3$, each component of which has framing 0 in $\mathbb{R}^3$ (see Figure 1).

A framed $Y$-graph $C$ in a 3-manifold $M$ is the image of an embedding of $N(Y)$ into $M$. The surgery of $M$ along the image of the six-component link $L(Y)$ is called a $Y$-surgery along $C$, denoted by $M_C$. If one of the leaves bounds a disk in $M$ whose interior is disjoint from the graph, then $M_C$ is homeomorphic to $M$.

Matveev in 1987 proved that two 3-manifolds $M$ and $M'$ are related by a finite sequence of $Y$-surgeries if and only if there is an isomorphism from $H_1(M, \mathbb{Z})$ onto $H_1(M', \mathbb{Z})$ preserving the

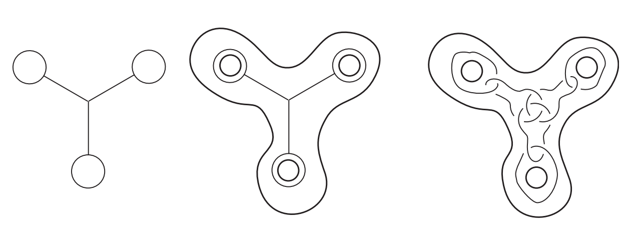



**Figure 1** Y-graph.

linking form on the torsion group. It is natural to partition the class of 3-manifolds into subclasses of the same $H_1$ and the same linking form.

**Goussarov–Habiro filtrations** For a 3-manifold $M$ denote by $\mathcal{M}(M)$ the vector space spanned by all 3-manifolds with $H_1$ and linking form the same as those of $M$. Define, for a set $S$ of $Y$-graphs in $M$, $[M,S] = \sum_{S' \subset S} (-1)^{\#S'} M_{S'}$, and $\mathcal{F}_n^Y \mathcal{M}(M)$ the vector space spanned by all $[N,S]$ such that $N$ is in $\mathcal{M}(M)$ and $\#S = n$. The following theorem of Goussarov and Habiro (Goussarov 1999, Garoufalidis *et al.* 2001, Habiro 2000) shows that the FTI theory based on $Y$-surgery is the same as the one of Ohtsuki in the case of $\mathbb{Z}$HS.

**Theorem 2** *For the case* $\mathcal{M} = \mathcal{M}(S^3)$, *one has* $\mathcal{F}_{2n-1}^Y = \mathcal{F}_{2n}^Y = \mathcal{F}_{3n}^{AS}$.

### The Fundamental Theorem of FTIs of $\mathbb{Z}$HS

**Jacobi diagrams** A closed Jacobi diagram is a vertex-oriented trivalent graph, that is, a graph for which the degree of each vertex is equal to 3 and a cyclic order of the three half-edges at every vertex is fixed. Here, multiple edges and self-loops are allowed. In pictures, the orientation at a vertex is the clockwise orientation, unless otherwise stated. The "degree" of Jacobi diagram is half the number of its vertices.

Let $\mathrm{Gr}_n\mathcal{A}(\emptyset), n \geq 0$, be the vector space spanned by all closed Jacobi diagrams of degree $n$, modulo the antisymmetry (AS) and Jacobi (IHX) relations (see Figure 2).

**The universal weight map $W$** Suppose $D$ is a closed Jacobi diagram of degree $n$. Embedding $D$ into $\mathbb{R}^3 \subset S^3$ arbitrarily and then projecting down onto $\mathbb{R}^2$ in general position, one can describe $D$ by a diagram, with over/under-crossing information at every double point just as in the case of a link diagram. We can assume that the orientation at every vertex of $D$ is given by a clockwise cyclic order. From the image of $D$, construct a set $G$ of $2n$ $Y$-graphs as in

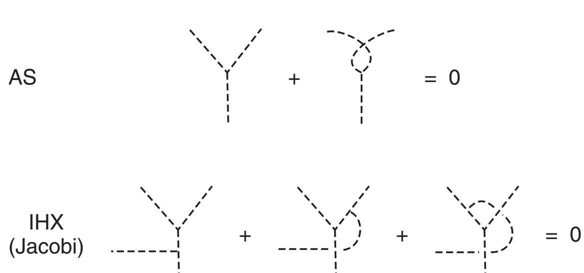



**Figure 2** The AS and IHX relations.

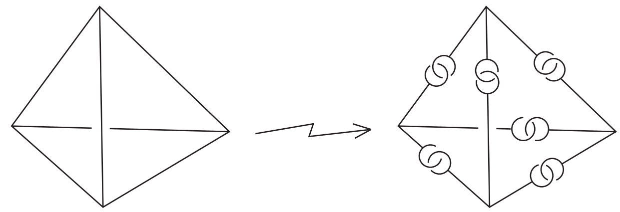


**Figure 3** The weight map.

Figure 3. Here only the cores of a $Y$-graph are drawn, with the convention that each framed $Y$-graph is a small neighborhood of its core in $\mathbb{R}^2$.

If $G'$ is a proper subset of $G$, then in $G'$ there is a $Y$-graph, one of the leaves of which bounds a disk, hence $S^3_{G'} = S^3$. Thus, $W(D) := [S^3, G] = S^3_G - S^3$. By definition, $W(D) \in \mathcal{F}_{2n}^Y$; it might depend on the embedding of $D$ into $\mathbb{R}^3$, but one can show that $W(D)$ is well defined in $\mathcal{F}_{2n}^Y/\mathcal{F}_{2n+1}^Y$. The map $W$ was first constructed by Garoufalidis and Ohtsuki in the framework of $\mathcal{F}^{AS}$.

#### Fundamental theorem

**Theorem 3** (Lê *et al.* 1998, Lê 1997). *The map $W$ descends to a well-defined linear map* $W : \mathrm{Gr}_n\mathcal{A}(\emptyset) \to \mathcal{F}_{2n}^Y/\mathcal{F}_{2n+1}^Y$ *and moreover, is an isomorphism between the vector spaces* $\mathrm{Gr}_n\mathcal{A}(\emptyset)$ *and* $\mathcal{F}_{2n}^Y/\mathcal{F}_{2n+1}^Y$, *for* $\mathcal{M} = \mathcal{M}(S^3)$.

The theorem essentially says that the set of invariants of degree $2n$ is dual to the space of closed Jacobi diagram $\mathrm{Gr}_n\mathcal{A}(\emptyset)$. The proof is based on the LMO invariant (see the next section).

A $\mathbb{Q}$-valued invariant $I$ of order $\leq 2n$ restricts to a linear map from $\mathcal{F}_{2n}/\mathcal{F}_{2n+1}$ to $\mathbb{Q}$. The composition of $I$ and $W$ is a functional on $\mathrm{Gr}_n\mathcal{A}(\emptyset)$ called the "weight system" of $I$. The theorem shows that every linear functional on $\mathrm{Gr}_{\leq n}\mathcal{A}(\emptyset)$ is the weight of an invariant of order $\leq 2n$.

**Relation to knot invariants** Under the map that sends an (unframed) knot $K \subset S^3$ to the $\mathbb{Z}$HS obtained by surgery along $K$ with framing 1, an invariant of degree $\leq 2n$ (in the $\mathcal{F}^Y$ theory) of $\mathbb{Z}$HS pulls back to an invariant of order $\leq 2n$ of knots. This was conjectured by Garoufalidis and proved by Habegger.

**Other classes of rational homology 3-spheres** Actually, the theorem was first proved in the framework of $\mathcal{F}^{AS}$. Clasper surgery theory allows Habiro (2000) to generalize the fundamental theorem to $\mathbb{Q}$HS: for $M$ a $\mathbb{Q}$HS, the universal weight map $W : \mathrm{Gr}_n \mathcal{A}(\emptyset) \to \mathcal{F}_{2n}\mathcal{M}(M)/\mathcal{F}_{2n+1}\mathcal{M}(M)$, defined similarly as

in the case of $\mathbb{Z}$HS, is an isomorphism, and $\mathcal{F}_{2n-1}\mathcal{M}(M) = \mathcal{F}_{2n}\mathcal{M}(M)$.

**Other filtrations and approaches** Other equivalent filtrations were introduced (and compared) by Garoufalidis, Garoufalidis and Levine (1997), and Garoufalidis–Goussarov–Polyak (2001). Of importance is the one using subgroups of mapping class groups in Garoufalidis and Levine (1997). A theory of $n$-equivalence was constructed by Goussarov and Habiro that encompasses many geometric aspects of FTIs of 3-manifolds (Habiro 2000, Goussarov 1999). Cochran and Melvin (2000) extended the original Ohtsuki definition to manifolds with homology, using algebraically split links, but the filtrations are different from those of Goussarov–Habiro.

## The Le–Murakami–Ohtsuki Invariant

### Jacobi Diagrams

An open Jacobi diagram is a vertex-oriented uni–trivalent graph, that is, a graph with univalent and trivalent vertices together with a cyclic ordering of the edges incident to the trivalent vertices. A univalent vertex is also called "a leg." The degree of an open Jacobi diagram is half the number of vertices (trivalent and univalent). A Jacobi diagram based on $X$, a compact oriented 1-manifold, is a graph $D$ together with a decomposition $D = X \cup \Gamma$, such that $D$ is the result of gluing all the legs of an open Jacobi diagram $\Gamma$ to distinct interior points of $X$. The degree of $D$, by definition, is the degree of $\Gamma$. In Figure 4 $X$ is depicted by bold lines. Let $\mathcal{A}^f(X)$ be the space of Jacobi diagrams based on $X$ modulo the usual antisymmetry, Jacobi and the new STU relations. The completion of $\mathcal{A}^f(X)$ with respect to degree is denoted by $\mathcal{A}(X)$.

When $X$ is a set of $m$-ordered oriented intervals, denote $\mathcal{A}(X)$ by $\mathcal{P}_m$, which has a natural algebra structure where the product $DD'$ of two Jacobi diagrams is defined by stacking $D$ on top of $D'$ (concatenating the corresponding oriented intervals). When $X$ is a set of $m$-ordered oriented circles, denote $\mathcal{A}(X)$ by $\mathcal{A}_m$. By identifying the two endpoints of each interval, one gets a map $pr : \mathcal{P}_m \to \mathcal{A}_m$, which is an isomorphism if $m = 1$ (see Bar-Natan (1995)).

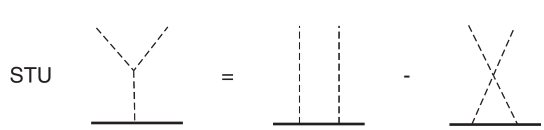



**Figure 4** The STU relation.

For $x \in \mathcal{A}_m$ and $y \in \mathcal{A}_1$, the connected sum is defined by $x\#_m y := pr((pr^{-1}x)(pr^{-1}y)^{\otimes m})$, where $(pr^{-1}y)^{\otimes m}$ is the element in $\mathcal{P}_m$ with $pr^{-1}y$ on each oriented interval.

**Symmetrization maps** Let $\mathcal{B}_m$ be the vector space spanned by open Jacobi diagrams whose legs are labeled by elements of $\{1, 2, \ldots, m\}$, modulo the antisymmetry and Jacobi relations. One can define an analog of the Poincare–Birkhoff–Witt isomorphism $\chi : \mathcal{B}_m \to \mathcal{P}_m$ as follows. For a diagram $D$, $\chi(D)$ is obtained by taking the average over all possible ways of ordering the legs labeled by $j$ and attaching them to the $j$th oriented interval. It is known that $\chi$ is a vector space isomorphism (Bar-Natan (1995)).

### The Framed Kontsevich Integral of Links

For an $m$-component framed link $L \subset \mathbb{R}^3$, the (framed version of the) Kontsevich integral $Z(L)$ is an invariant taking values in $\mathcal{A}_m$ (see, e.g., Ohtsuki (2002)). Let $\nu := Z(K)$, when $K$ is the unknot with framing 0, and $\check{Z}(L) := Z(L)\#_m\nu$. An explicit formula for $\nu$ is given in Bar-Natan *et al.* (2003).

### Removing Solid Loops: The Maps $\iota_n$

Suppose $x \in \mathcal{B}_m$ is an open Jacobi diagram with legs labeled by $\{1, \ldots, m\}$. If the number of vertices of any label is different from $2n$, or if the degree of $D > (m+1)n$, we set $\iota_n(D) = 0$. Otherwise, partitioning the $2n$ vertices of each label into $n$ pairs and identifying points in each pair, from $x$ we get a trivalent graph which may contain some isolated loops (no vertices) and which depends on the partition. Replacing each isolated loop by a factor $-2n$, and summing up over all partitions, we get $\iota_n(D) \in \mathrm{Gr}_{\leq n}\mathcal{A}(\emptyset)$.

For $x \in \mathcal{A}_m$, choose $y \in \mathcal{P}_m$ such that $pr(y) = x$. Using the isomorphism $\chi$ we pull back $\chi^{-1}y \in \mathcal{B}_m$. Define $\iota_n(x) := \iota_n(\chi^{-1}y)$. One can prove that $\iota_n(x)$ does not depend on the choice of the preimage $y$ of $x$. Note that $\iota_n$ lowers the degree by $nm$.

### Definition of the Le–Murakami–Ohtsuki Invariant $Z^{\mathrm{LMO}}$

In $\mathcal{A}(\emptyset) := \prod_{n=0}^{\infty} \mathrm{Gr}_n\mathcal{A}(\emptyset)$ let the product of two Jacobi diagrams be their disjoint union. In addition, define the coproduct $\Delta(D) = 1 \otimes D + D \otimes 1$ for $D$ a connected Jacobi diagram. Then $\mathcal{A}(\emptyset)$ is a commutative cocommutative graded Hopf algebra.

For the unknot $U_\pm$ with framing $\pm 1$, one has $\iota_n(\check{Z}(U_\pm)) = (\mp 1)^n + (\text{terms of degree} \geq 1)$; hence, their inverses exist. Suppose the linking matrix of an

oriented framed link $L \subset \mathbb{R}^3$ has $\sigma_+$ positive eigenvalues and $\sigma_-$ negative eigenvalues. Define

$$\Omega_n(L) = \frac{\iota_n(\check{Z}(L))}{(\iota_n(\check{Z}(U_+)))^{\sigma_+}(\iota_n(\check{Z}(U_-)))^{\sigma_-}} \in \mathrm{Grad}_{\leq n}(\mathcal{A}(\emptyset)) \qquad [1]$$

**Theorem 4** (Lê *et al.* 1998). *$\Omega_n(L)$ is an invariant of the 3-manifold $M = S^3_L$.*

We can combine all the $\Omega_n$ to get a better invariant:

$$Z^{\mathrm{LMO}}(M) := 1 + \mathrm{Grad}_1(\Omega_1(M)) + \cdots + \mathrm{Grad}_n(\Omega_n(M)) + \cdots \in \mathcal{A}(\emptyset)$$

For $M$ a $\mathbb{Q}$HS, we also define

$$\hat{Z}^{\mathrm{LMO}}(M) := 1 + \frac{\mathrm{Grad}_1(\Omega_1(M))}{d(M)} + \cdots + \frac{\mathrm{Grad}_n(\Omega_n(M))}{d(M)^n} + \cdots$$

where $d(M)$ is the cardinality of $H_1(M, \mathbb{Z})$.

**Proposition 1** (Lê *et al.* 1998). *Both $Z^{\mathrm{LMO}}(M)$ and $\hat{Z}^{\mathrm{LMO}}(M)$ (when defined) are group-like elements, that is,*

$$\Delta(Z^{\mathrm{LMO}}(M)) = Z^{\mathrm{LMO}}(M) \otimes Z^{\mathrm{LMO}}(M)$$
$$\Delta(\hat{Z}^{\mathrm{LMO}}(M)) = \hat{Z}^{\mathrm{LMO}}(M) \otimes \hat{Z}^{\mathrm{LMO}}(M)$$

*Moreover, $\hat{Z}^{\mathrm{LMO}}(M_1 \# M_2) = \hat{Z}^{\mathrm{LMO}}(M_1) \times \hat{Z}^{\mathrm{LMO}}(M_2)$.*

### Universality Properties of the LMO Invariant

Let us restrict ourselves to the case of $\mathbb{Z}$HS.

**Theorem 5** (Lê 1997). *The less than or equal to $n$ degree part $\mathrm{Gr}_{\leq n} Z^{\mathrm{LMO}}$ is an invariant of degree $2n$. Any invariant of degree $\leq 2n$ is a composition $w(\mathrm{Gr}_{\leq n} Z^{\mathrm{LMO}})$, where $w : \mathrm{Gr}_{\leq n}\mathcal{A}(\emptyset) \to \mathbb{Q}$ is a linear map.*

Clasper calculus (or $Y$-surgery) theory allows Habiro to extend the theorem to rational homology 3-spheres.

### The Arhus Integral

The Arhus integral (ca. 1998) of Bar-Natan, Garoufalidis, Rozansky and Thurston, based on a theory of formal integration, calculates the LMO invariant of rational homology 3-spheres. The formal integration theory has a conceptual flavor and helps to relate the LMO invariant to perturbative expansions of quantum invariants. We give here the definition for the case when one does surgery on a knot $K$ with nonzero framing $b$. The link case is similar (see Bar-Natan *et al.* (2002a, b)).

When $K$ is a knot, $\check{Z}(K)$ is an element of $\mathcal{A}_1 \equiv \mathcal{P}_1 \equiv \mathcal{B}_1$. Note that $\mathcal{B}_1$ is an algebra where the product is the disjoint union $\sqcup$. Since the framing is $b$, one has

$$\check{Z}(K) = \exp_{\sqcup}(b\, w_1/2) \sqcup Y$$

where $w_1$ is the "dashed interval" (the only connected open Jacobi diagram without trivalent vertex), and $Y$ is an element in $\mathcal{B}$ every term of which must have at least one trivalent vertex. For uni–trivalent graphs $C, D \in \mathcal{B}_1$ let

$$\langle C, D\rangle = \begin{cases} 0 & \text{if the numbers of legs of } C, D \text{ are different} \\ \text{sum of all ways to glue legs of } C \text{ and } D \text{ together} \end{cases}$$

One defines $\int^{FG} \check{Z}(K) := \langle \exp_{\sqcup}(-w_1/2b), Y\rangle$. Then

$$\int^{FG} \check{Z}(K) = \sum_{n=0}^{\infty} \frac{\mathrm{Gr}_n(\iota_n \check{Z}(K))}{(-b)^n}$$

Hence,

$$\hat{Z}^{\mathrm{LMO}}(S^3_K) = \frac{\int^{FG} \check{Z}(K)}{\int^{FG} \check{Z}(U_{\mathrm{sign}(b)})}$$

### Other Approaches

Another construction of a universal perturbative invariant based on integrations over configuration spaces, closer to the original physics approach but harder to calculate because of the lack of a surgery formula, was developed by Axelrod and Singer, Kontsevich, Bott and Cattaneo, Kuperberg and Thurston (see Axelred and Singer (1992), Bott and Cattaneo (1998)).

## Quantum Invariants and Perturbative Expansion

Fix a simple (complex) Lie algebra $\mathfrak{g}$ of finite dimension. Using the quantized enveloping algebra of $\mathfrak{g}$ one can define quantum link and 3-manifold invariants. We recall here the definition, adapted for the case of roots lattice (projective group case).

Here our $q$ is equal to $q^2$ in the text book (Jantzen 1995). Fix a root system of $\mathfrak{g}$. Let $X, X_+, Y$ denote respectively the weight lattice, the set of dominant weights, and the root lattice. We normalize the invariant scalar product in the real vector space of the weight lattice so that the length of any short root is $\sqrt{2}$.

### Quantum Link Invariants

Suppose $L$ is a framed oriented link with $m$-ordered components, then the quantum invariant $J_L(\lambda_1, \ldots, \lambda_m)$ is a Laurent polynomial in $q^{1/2D}$, where $\lambda_1, \ldots, \lambda_m$ are dominant weights, standing for the simple $\mathfrak{g}$-modules of highest weights $\lambda_1, \ldots, \lambda_m$, and $D$ is the determinant of the Cartan matrix of $\mathfrak{g}$ (see, e.g., Turaev (1994) and Lê (1996)). The Jones polynomial is the case when $\mathfrak{g} = \mathrm{sl}_2$ and all the $\lambda_i$'s are the highest weights of the fundamental representation. For the unknot $U$ with zero framing, one has (here $\rho$ is the half-sum of all positive roots)

$$J_U(\lambda) = \prod_{\text{positive roots } \alpha} \frac{q^{(\lambda+\rho|\alpha)/2} - q^{-(\lambda+\rho|\alpha)/2}}{q^{(\rho|\alpha)/2} - q^{-(\rho|\alpha)/2}}$$

We will also use another normalization of the quantum invariant:

$$Q_L(\lambda_1, \ldots, \lambda_m) := J_L(\lambda_1, \ldots, \lambda_m) \times \prod_{j=1}^{m} J_U(\lambda_j)$$

This definition is good only for $\lambda_j \in X_+$. Note that each $\lambda \in X$ is either fixed by an element of the Weyl group under the dot action (see Humphreys (1978)) or can be moved to $X_+$ by the dot action. We define $Q_L(\lambda_1, \ldots, \lambda_m)$ for arbitrary $\lambda_j \in X$ by requiring that $Q_L(\lambda_1, \ldots, \lambda_m) = 0$ if one of the $\lambda_j$'s is fixed by an element of the Weyl group, and that $Q_L(\lambda_1, \ldots, \lambda_m)$ is component-wise invariant under the dot action of the Weyl group, that is, for every $w_1, \ldots, w_m$ in the Weyl group,

$$Q_L(w_1 \cdot \lambda_1, \ldots, w_m \cdot \lambda_m) = Q_L(\lambda_1 \ldots, \lambda_m)$$

**Proposition 2** (Lê 1996). *Suppose $\lambda_1, \ldots, \lambda_m$ are in the root lattice $Y$.*

(i) *(Integrality) Then $Q_L(\lambda_1, \ldots, \lambda_m) \in \mathbb{Z}[q^{\pm 1}]$, (no fractional power).*
(ii) *(Periodicity) When $q$ is an $r$th root of 1, then $Q_L(\lambda_1, \ldots, \lambda_m)$ is invariant under the action of the lattice group $rY$, that is, for $y_1, \ldots, y_m \in Y$, $Q_L(\lambda_1, \ldots, \lambda_m) = Q_L(\lambda_1 + ry_1, \ldots, \lambda_m + ry_m)$.*

### Quantum 3-Manifold Invariants

Although the infinite sum $\sum_{\lambda_j \in Y} Q_L(\lambda_1, \ldots, \lambda_m)$ does not have a meaning, heuristic ideas show that it is invariant under the second Kirby move, and hence almost defines a 3-manifold invariant. The problem is to regularize the infinite sum. One solution is based on the fact that at $r$th roots of unity, $Q_L(\lambda_1, \ldots, \lambda_m)$ is periodic, so we should use the sum with $\lambda_j$'s run over a fundamental set $P_r$ of the action of $rY$, where

$$P_r := \{x = c_1\alpha_1 + \cdots + c_\ell\alpha_\ell \mid 0 \le c_1, \ldots, c_\ell < r\}$$

Here $\alpha_1, \ldots, \alpha_\ell$ are basis roots. For a root $\xi$ of unity of order $r$, let

$$F_L(\xi) = \sum_{\lambda_j \in (P_r \cap Y)} Q_L(\lambda_1, \ldots, \lambda_m)|_{q=\xi}$$

If $F_{U_\pm}(\xi) \neq 0$, define

$$\tau_L(\xi) := \frac{F_L(\xi)}{(F_{U_+}(\xi))^{\sigma_+}(F_{U_-}(\xi))^{\sigma_-}}$$

Recall that $D$ is the determinant of the Cartan matrix. Let $d$ be the maximum of the absolute values of entries of the Cartan matrix outside the diagonal.

**Theorem 6** (Lê 2003)

(i) *If the order $r$ of $\xi$ is coprime with d$D$, then $F_{U_\pm}(\xi) \neq 0$.*
(ii) *If $F_{U_\pm}(\xi) \neq 0$ then $\tau_M^{P\mathfrak{g}}(\xi) := \tau_L(\xi)$ is an invariant of the 3-manifold $M = S^3_L$.*

**Remark 1** The version presented here corresponds to projective groups. It was defined by Kirby and Melvin for $\mathrm{sl}_2$, Kohno and Takata for $\mathrm{sl}_n$, and by Lê (2003) for arbitrary simple Lie algebra. When $r$ is coprime with d$D$, there is also an associated modular category that generates a topological quantum field theory. In most texts in literature, say Kirillov (1996) and Turaev (1994), another version $\tau^{\mathfrak{g}}$ was defined. The reason we choose $\tau^{P\mathfrak{g}}$ is: it has nice integrality and eventually perturbative expansion. For relations between the version $\tau^{P\mathfrak{g}}$ and the usual $\tau^{\mathfrak{g}}$, see Lê (2003).

**Examples** When $M$ is the Poincaré sphere and $\mathfrak{g} = \mathrm{sl}_2$,

$$\tau_M^{Psl_2}(q) = \frac{1}{1-q}\sum_{n=0}^{\infty} q^n(1-q^{n+1}) \times (1-q^{n+2}) \ldots (1-q^{2n+1})$$

Here $q$ is a root of unity, and the sum is easily seen to be finite.

**Integrality** The following theorem was proved for $\mathfrak{g} = \mathrm{sl}_2$ by Murakami (1995) and for $\mathfrak{g} = \mathrm{sl}_n$ by Takata–Yokota and Masbaum–Wenzl (using ideas of J Roberts) and for arbitrary simple Lie algebras by Lê (2003).

**Theorem 7** *Suppose the order $r$ of $\xi$ is a prime big enough, then $\tau_M^{P\mathfrak{g}}(\xi)$ is in $\mathbb{Z}[\xi] = \mathbb{Z}[\exp(2\pi i/r)]$.*

### Perturbative Expansion

Unlike the link case, quantum 3-manifold invariants can be defined only at certain roots of unity. In general, there is no analytic extension of the function $\tau_M^{P\mathfrak{g}}$ around $q=1$. In perturbative theory, we want to expand the function $\tau_M^{\mathfrak{g}}$ around $q=1$ into power series. For $\mathbb{Q}$HS, Ohtsuki (for $\mathfrak{g}=\mathrm{sl}_2$) and then the present author (for all other simple Lie algebras) showed that there is a number-theoretical expansion of $\tau_M^{P\mathfrak{g}}$ around $q=1$ in the following sense.

Suppose $r$ is a big enough prime, and $\xi=\exp(2\pi i/r)$. By the integrality (Theorem 7),

$$\tau_M^{P\mathfrak{g}}(\xi)\in\mathbb{Z}[\xi]=\mathbb{Z}[q]/(1+q+q^2+\cdots+q^{r-1})$$

Choose a representative $f(q)\in\mathbb{Z}[q]$ of $\tau_M^{P\mathfrak{g}}(\xi)$. Formally substitute $q=(q-1)+1$ in $f(q)$:

$$f(q)=c_{r,0}+c_{r,1}(q-1)+\cdots+c_{r,n-2}(q-1)^{n-2}$$

The integers $c_{r,n}$ depend on $r$ and the representative $f(q)$. It is easy to see that $c_{r,n}$ (mod $r$) does not depend on the representative $f(q)$ and hence is an invariant of $\mathbb{Q}$HS. The dependence on $r$ is a big drawback. The theorem below says that there is a rational number $c_n$, not depending on $r$, such that $c_{r,n}$ (mod $r$) is the reduction of either $c_n$ or $-c_n$ modulo $r$, for sufficiently large prime $r$. It is easy to see that if such $c_n$ exists, it must be unique. Let $s$ be the number of positive roots of $\mathfrak{g}$. Recall that $\ell$ is the rank of $\mathfrak{g}$.

**Theorem 8** *For every $\mathbb{Q}$HS $M$, there is a sequence of numbers*

$$c_n\in\mathbb{Z}\left[\frac{1}{(2n+2s)!|H_1(M,\mathbb{Z})|}\right]$$

*such that for sufficiently large prime $r$*

$$c_{r,n}\equiv\left(\frac{|H_1(M,\mathbb{Z})|}{r}\right)^{\ell}c_n\ (\mathrm{mod}\ r)$$

*where*

$$\left(\frac{|H_1(M,\mathbb{Z})|}{r}\right)=\pm1$$

*is the Legendre symbol. Moreover, $c_n$ is an invariant of order $\le 2n$.*

The series $\mathrm{t}_M^{P\mathfrak{g}}(q-1):=\sum_{n=0}^{\infty}c_n(q-1)^n$, called the Ohtsuki series, can be considered as the perturbative expansion of the function $\tau_M^{P\mathfrak{g}}$ at $q=1$. For actual calculation of $\mathrm{t}_M^{P\mathfrak{g}}(q-1)$, see Lê (2003), Ohtsuki (2002), and Rozansky (1997).

**Recovery from the LMO invariant** It is known that for any metrized Lie algebra $\mathfrak{g}$, there is a linear map $W_{\mathfrak{g}}:\mathrm{Gr}_n\mathcal{A}(\emptyset)\to\mathbb{Q}$ (see Bar-Natan (1995)).

**Theorem 9** *One has*

$$\sum_{n=0}^{\infty}W_{\mathfrak{g}}(\mathrm{Gr}_nZ^{\mathrm{LMO}})\,h^n=\mathrm{t}_M^{P\mathfrak{g}}(q-1)|_{q=e^h}$$

This shows that the Ohtsuki series $\mathrm{t}_M^{P\mathfrak{g}}(q-1)$ can be recovered from, and hence totally determined by, the LMO invariant. The theorem was proved by Ohtsuki for $\mathrm{sl}_2$. For other simple Lie algebras, the theorem follows from the Arhus integral (see Bar-Natan *et al.* (2002a, b) and Ohtsuki (2002)).

### Rozansky's Gaussian Integral

Rozansky (1997) gave a definition of the Ohtsuki series using formal Gaussian integral in the important work. The work is only for $\mathrm{sl}_2$, but can be generalized to other Lie algebras; it is closer to the original physics ideas of perturbative invariants.

## Cyclotomic Expansion

### The Habiro Ring

Let us define the Habiro ring $\widehat{\mathbb{Z}[q]}$ by

$$\widehat{\mathbb{Z}[q]}:=\lim_{\leftarrow n}\mathbb{Z}[q]/((1-q)(1-q^2)\ldots(1-q^n))$$

Habiro (2002) called it the cyclotomic completion of $\mathbb{Z}[q]$. Formally, $\widehat{\mathbb{Z}[q]}$ is the set of all series of the form

$$f(q)=\sum_{n=0}^{\infty}f_n(q)(1-q)(1-q^2)\ldots(1-q^n)$$
$$f_n(q)\in\mathbb{Z}[q]$$

Suppose $U$ is the set of roots of 1. If $\xi\in U$ then $(1-\xi)(1-\xi^2)\cdots(1-\xi^n)=0$ if $n$ is big enough; hence, one can define $f(\xi)$ for $f\in\widehat{\mathbb{Z}[q]}$. One can consider every $f\in\widehat{\mathbb{Z}[q]}$ as a function with domain $U$. Note that $f(\xi)\in\mathbb{Z}[\xi]$ is always an algebraic integer. It turns out that $\widehat{\mathbb{Z}[q]}$ has remarkable properties, and plays an important role in quantum topology.

Note that the formal derivative of $(1-q)(1-q^2)\ldots(1-q^n)$ is divisible by $(1-q)(1-q^2)\ldots(1-q^k)$ with $k>(n-1)/2$. This means every element $f\in\widehat{\mathbb{Z}[q]}$ has a derivative $f'\in\widehat{\mathbb{Z}[q]}$, and hence derivatives of all orders in $\widehat{\mathbb{Z}[q]}$. One can then associate to $f\in\widehat{\mathbb{Z}[q]}$ its Taylor series at a root $\xi$ of 1:

$$T_\xi(f):=\sum_{n=0}^{\infty}\frac{f^{(n)}(\xi)}{n!}(q-\xi)^n$$

which can also be obtained by noticing that $(1-q)(1-q^2)\ldots(1-q^n)$ is divisible by $(q-\xi)^k$ if $n$ is bigger than $k$ times the order of $\xi$. Thus, one has a map $T_\xi:\widehat{\mathbb{Z}[q]}\to\mathbb{Z}[\xi][[q-\xi]]$.

**Theorem 10** (Habiro 2004)

(i) *For each root of unity* $\xi$, *the map* $T_\xi$ *is injective, that is, a function in* $\widehat{\mathbb{Z}[q]}$ *is determined by its Taylor expansion at a point in the domain U.*
(ii) *if* $f(\xi) = g(\xi)$ *at infinitely many roots* $\xi$ *of prime power orders, then* $f = g$ *in* $\widehat{\mathbb{Z}[q]}$.

One important consequence is that $\widehat{\mathbb{Z}[q]}$ is an integral domain, since we have the embedding $T_1 : \widehat{\mathbb{Z}[q]} \hookrightarrow \mathbb{Z}[[q-1]]$.

In general the Taylor series $T_1 f$ has 0 convergence radius. However, one can speak about $p$-adic convergence to $f(\xi)$ in the following sense. Suppose the order $r$ of $\xi$ is a power of prime, $r = p^k$. Then it is known that $(\xi - 1)^n$ is divisible by $p^m$ if $n > mk$. Hence, $T_1 f(\xi)$ converges in the $p$-adic topology, and it can be easily shown that the limit is exactly $f(\xi)$.

The above properties suggest considering $\widehat{\mathbb{Z}[q]}$ as a class of "analytic functions" with domain $U$.

#### Quantum Invariants as an Element of $\widehat{\mathbb{Z}[q]}$

It was proved, by Habiro for $sl_2$ and by Habiro with the present author for general simple Lie algebras, that quantum invariants of $\mathbb{Z}$HSs belong to $\widehat{\mathbb{Z}[q]}$ and thus have remarkable integrality properties:

**Theorem 11**

(i) *For every* $\mathbb{Z}$HS *M, there is an invariant* $I_M^{\mathfrak{g}} \in \widehat{\mathbb{Z}[q]}$ *such that if* $\xi$ *is a root of unity for which the quantum invariant* $\tau_M^{P\mathfrak{g}}(\xi)$ *can be defined, then* $I_M^{\mathfrak{g}}(\xi) = \tau_M^{P\mathfrak{g}}(\xi)$.
(ii) *The Ohtsuki series is equal to the Taylor series of* $I_M^{\mathfrak{g}}$ *at 1.*

**Corollary 1** *Suppose M is a* $\mathbb{Z}$HS.

(i) *For every root of unity* $\xi$, *the quantum invariant at* $\xi$ *is an algebraic integer,* $\tau_M^{\mathfrak{g}}(\xi) \in \mathbb{Z}[\xi]$. (*No restriction on the order of* $\xi$ *is required.*)
(ii) *The Ohtsuki series* $t_M^{P\mathfrak{g}}(q-1)$ *has integer coefficients. If* $\xi$ *is a root of order* $r = p^k$, *where p is prime, then the Ohtsuki series at* $\xi$ *converges p-adically to the quantum invariant at* $\xi$.
(iii) *The quantum invariant* $\tau_M^{P\mathfrak{g}}$ *is determined by values at infinitely many roots of prime power orders and also determined by its Ohtsuki series.*
(iv) *The* LMO *invariant totally determines the quantum invariants* $\tau_M^{P\mathfrak{g}}$.

Part (ii) was conjectured by R Lawrence for $sl_2$ and first proved by Rozansky (also for $sl_2$). Part (iv) follows from the fact that the LMO invariant determines the Ohtsuki series; it exhibits another universality property of the LMO invariant.

*See also:* Finite-Type Invariants; Knot Invariants and Quantum Gravity; Lie Groups: General Theory; Quantum 3-Manifold Invariants.

## Further Reading


Axelrod S and Singer IM (1992) Chern–Simons perturbation theory. In: *Proceedings of the XXth International Conference on Differential Geometric Methods in Theoretical Physics*, New York, 1991, vols. 1 and 2, pp. 3–45. River Edge, NJ: World Scientific.

Bar-Natan D (1995) On the Vassiliev knot invariants. *Topology* 34: 423–472.

Bar-Natan D, Garoufalidis S, Rozansky L, and Thurston DP (2002a) The Arhus integral of rational homology 3-spheres. *Selecta Mathematica (NS)* 8: 315–339.

Bar-Natan D, Garoufalidis S, Rozansky L, and Thruston DP (2002b) The Arhus integral of rational homology 3-spheres. *Selecta Mathematica (NS)* 8: 341–371.

Bar-Natan D, Garoufalidis S, Rozansky L, and Thurston DP (2004) The Arhus integral of rational homology 3-spheres. *Selecta Mathematica (NS)* 10: 305–324.

Bar-Natan D, Lê TTQ, and Thurston DP (2003) Two applications of elementary knot theory to Lie algebras and Vassiliev invariants. *Geometry and Topology* 7: 1–31 (electronic).

Bott R and Cattaneo AS (1998) Integral invariants of 3-manifolds. *Journal of Differential Geometry* 48: 91–133.

Cochran TD and Melvin P (2000) Finite type invariants of 3-manifolds. *Inventiones Mathematicae* 140: 45–100.

Garoufalidis S and Levine J (1997) Finite type 3-manifold invariants, the mapping class group and blinks. *Journal of Differential Geometry* 47: 257–320.

Garoufalidis S, Goussarov M, and Polyak M (2001) Calculus of clovers and finite type invariants of 3-manifolds. *Geometry and Topology* 5: 75–108 (electronic).

Goussarov M (1999) Finite type invariants and n-equivalence of 3-manifolds. *Comptes Rendus des Seances de l'Academie des Sciénces. Série I. Mathématique* 329: 517–522.

Habiro K (2000) Claspers and finite type invariants of links. *Geometry and Topology* 4: 1–83 (electronic).

Habiro K (2002) On the quantum $sl_2$ invariants of knots and integral homology spheres. In: *Invariants of Knots and 3-Manifolds (Kyoto, 2001)*, Geometry and Topology Monogram, (electronic), vol. 4, pp. 55–68. Coventry: Geometry and Topology Publisher.

Habiro K (2004) Cyclotomic completions of polynomial rings. *Publications of the Research Institute for Mathematical Sciences, Kyoto University* 40: 1127–1146.

Humphreys J (1978) *Introduction to Lie Algebras and Representation Theory*. Graduate Texts in Mathematics, vol. 6. Berlin: Springer.

Jantzen JC (1995) *Lecture on Quantum Groups*. Graduate Studies in Mathematics, vol. 6. Providence, RI: American Mathematical Society.

Kirillov A (1996) On an inner product in modular categories. *Journal of American Mathematical Society* 9: 1135–1169.

Lê TTQ (1997) In: Buchtaber and Novikov S (eds.) *An Invariant of Integral Homology 3-Spheres which is Universal for all Finite Type Invariants*. AMS Translation Series 2, vol. 179, pp. 75–100. Providence, RI: American Mathematical Society.

Lê TTQ (2000) Integrality and symmetry of quantum link invariants. *Duke Mathematical Journal* 102: 273–306.

Lê TTQ (2003) Quantum invariants of 3-manifolds: integrality, splitting, and perturbative expansion. *Topology and Its Applications* 127: 125–152.

Lê TTQ, Murakami J, and Ohtsuki T (1998) On a universal perturbative invariant of 3-manifolds. *Topology* 37: 539–574.
Murakami H (1995) Quantum SO(3)-invariants dominate the SU(2)-invariant of Casson and Walker. *Math Proceedings Cambridge Philosophical Society* 117: 237–249.
Ohtsuki T (2002) *Quantum Invariants. A Study of Knots, 3-Manifolds and their Sets*. Series on Knots and Everything, vol. 29. River Edge, NJ: World Scientific.
Rozansky L (1997) The trivial connection contribution to Witten's invariant and finite type invariants of rational homology spheres. *Communications in Mathematical Physics* 183(1): 23–54.
Rozansky L (1998) On p-adic propreties of the Witten–Reshetikhin–Turaev invariant. *Preprint math.QA/9806075*.
Turaev VG (1994) *Quantum Invariants of Knots and 3-Manifolds*. de Gruyter Studies in Mathematics, vol. 18. Berlin: Walter de Gruyter.
Witten E (1989) Quantum field theory and the Jones polynomial. *Communications in Mathematical Physics* 121: 360–379.

# Floer Homology

**P B Kronheimer**, Harvard University, Cambridge, MA, USA



## Introduction

Morse theory allows one to reconstruct the homology of a compact manifold $B$ from data obtained from the gradient flow of a function $f: B \to \mathbb{R}$, the Morse function. The term "Floer homology" is used to describe homology groups that arise from carrying out the same construction, but in a setting where the space $B$ is replaced by an infinite-dimensional manifold (a space of maps, or a space of configurations for a gauge theory), and where the gradient trajectories of the Morse function correspond to solutions of an elliptic differential equation. There are two important types of such homology theories that have been extensively developed, and the study of both was initiated in the 1980s by Andreas Floer. In the first type, the elliptic equation that arises is a Cauchy–Riemann equation, whose solutions are pseudoholomorphic maps from a two-dimensional domain into a symplectic manifold. In the second type, the elliptic equation is an equation of gauge theory on a 4-manifold: either the anti-self-dual Yang–Mills equations or the Seiberg–Witten equations. Important antecedents of Floer's work included work of Conley, Zehnder, and others on the symplectic fixed-point problem, and Witten's ideas about Morse theory.

This article describes the background material from Morse theory before discussing Floer homology of Cauchy–Riemann type and its application to the Arnol'd conjecture in symplectic topology. Floer homology in the context of four-dimensional gauge theories is discussed more briefly.

## Morse Theory

Let $B$ be a smooth, compact manifold and $f: B \to \mathbb{R}$ a smooth function. A critical point $p$ of $f$ is said to be nondegenerate if the Hessian of $f$ is a nonsingular operator on $T_pB$. The function $f$ is a Morse function if all its critical points are nondegenerate. In the presence of a Riemannian metric $g$ on $B$, the derivative $\mathrm{d}f$ becomes a vector field, the gradient $\nabla f$, and we can consider the downward gradient-flow equation for a path $x(s)$ in $B$:

$$\frac{\mathrm{d}x}{\mathrm{d}s} = -\nabla f(x)$$

If $p$ and $q$ are nondegenerate critical points, let us write $M(p,q)$ for the space of solutions $x(s)$ satisfying

$$\lim_{s\to-\infty} x(s) = p$$
$$\lim_{s\to+\infty} x(s) = q$$

To understand the structure of $M(p,q)$, consider the linearization of the gradient-flow equation at a solution $x \in M(p,q)$. This is a linear equation for a vector field $X$ along the path $x$ in $B$, and takes the form

$$\nabla_{\partial/\partial s} X = -\nabla\nabla f(X) \qquad [1]$$

where $\nabla\nabla f$ is the covariant derivative of the gradient $\nabla f$, an operator on tangent vectors. Let $\epsilon_x$ be the dimension of the space of solutions $X$ to this linear equation, with the boundary conditions $\lim_{s\to\pm\infty} X(s) = 0$, and let $\epsilon'_x$ be the dimension of the space of solutions to the adjoint equation

$$\nabla_{\partial/\partial s} X = +\nabla\nabla f(X)$$

We say that the trajectory $x$ is "regular" if $\epsilon'_x = 0$. In this case, the trajectory space $M(p,q)$ has the structure of smooth manifold near $x$: its dimension is $\epsilon_x$ and its tangent space is the space of solutions $X$ to [1]. The gradient flow is said to be Morse–Smale if all trajectories between critical points are regular. If $f$ is any Morse function, one can always choose the metric $g$ so that the corresponding flow is Morse–Smale. (It is also the case that one can leave $g$ fixed and perturb $f$ to achieve the same effect.)

In the Morse–Smale case, each $M(p,q)$ is a smooth manifold. The dimension of $M(p,q)$ in the neighborhood of a trajectory $x$ depends only on $p$ and $q$, not otherwise on $x$. Indeed, even without the regularity condition, the index of eqn [1], namely the difference $\epsilon_x - \epsilon'_x$, is given by

$$\epsilon_x - \epsilon'_x = \mathrm{index}(p) - \mathrm{index}(q)$$

where $\mathrm{index}(p)$ denotes the number of negative eigenvalues (counting multiplicity) of the Hessian at $p$. In the Morse–Smale case therefore, the dimension of $M(p,q)$ is given by $\mathrm{index}(p) - \mathrm{index}(q)$. If $x(s)$ is a solution of the gradient-flow equation, then so is the reparametrized trajectory $x(s+c)$; and this is different from $x(s)$ as long as $p \neq q$. Let us denote by $\breve{M}(p,q)$ the quotient of $M(p,q)$ by the action of $\mathbb{R}$ given by these reparametrizations. We have

$$\dim \breve{M}(p,q) = \mathrm{index}(p) - \mathrm{index}(q) - 1 \quad (p \neq q)$$

as long as the trajectory space is nonempty.

Let $\mathbb{F}_2$ denote the field with two elements. The Morse complex of a Morse–Smale gradient flow, with coefficients in $\mathbb{F}_2$, is defined as follows. For each $i$, let $C_i(f)$ be the finite-dimensional vector space over $\mathbb{F}_2$ having a basis

$$e_{p_1}, \ldots, e_{p_{r_i}}$$

indexed by the critical points $p_1, \ldots, p_{r_i}$ with index $i$. For each pair of critical points $p$ and $q$ with indices $i$ and $i-1$ respectively, let $\delta_{pq} \in \mathbb{F}_2$ denote the number of points in the zero-dimensional manifold $\breve{M}(p,q)$, counted mod 2:

$$\delta_{pq} = \#\breve{M}(p,q) \pmod 2$$

The Morse–Smale condition ensures that the zero-dimensional space $\breve{M}(p,q)$ is finite, so this definition is satisfactory. Define a differential

$$\delta : C_i(f) \to C_{i-1}(f)$$

by

$$\delta(e_p) = \sum_{\mathrm{index}(q)=i-1} \delta_{pq} e_q$$

The first important fact is that $\delta$ really is a differential: as long as the flow is Morse–Smale, we have

$$\text{the composite } \delta \circ \delta : C_i(f) \to C_{i-2}(f) \text{ is zero} \qquad [2]$$

We can therefore construct the homology of the complex $(C_*(f), \delta)$. This is the Morse homology:

$$H_i(f) = \frac{\ker(\delta : C_i(f) \to C_{i-1}(f))}{\mathrm{im}(\delta : C_{i+1}(f) \to C_i(f))} \qquad [3]$$

The proof of [2] is as follows. Suppose that $p$ has index $i$ and $r$ is a critical point with index $i-2$, and consider $\breve{M}(p,r)$, which has dimension 1. The key step is to understand that $\breve{M}(p,r)$ is noncompact, and that its ends correspond to "broken trajectories": pairs $(x_1, x_2)$ (modulo reparametrization), where $x_1$ is a gradient trajectory from $p$ to some $q$ of index $i-1$, and $x_2$ is a trajectory from $q$ to $r$. The number of ends is thus $\sum_q \delta_{qr}\delta_{pq}$. Since the number of ends of a 1-manifold is even, this sum is zero in $\mathbb{F}_2$. This sum is also the matrix entry of $\delta \circ \delta$ from $e_p$ to $e_r$; so $\delta \circ \delta = 0$.

The main result about Morse homology in finite dimensions is the following:

**Theorem 1** *The Morse homology $H_i(f)$ is isomorphic to the ordinary homology of the compact manifold $B$ with coefficients $\mathbb{F}_2$: the group $H_i(B; \mathbb{F}_2)$.*

This result can be proved by first showing that $H_i(f)$ depends only on $B$, not on the choice of $f$ or the metric. (This step can be accomplished by examining a nonautonomous flow of the form $\mathrm{d}x/\mathrm{d}s = -\nabla f(s,x)$.) Then one can examine the Morse complex in the case of a self-indexing Morse function (where the value of $f$ at the critical points is a monotone-increasing function of their index). In the self-indexing case, the unstable manifolds of the critical points give rise to a cell decomposition of the manifold $B$, and the Morse complex is easily identified with the cellular chain complex for this cell decomposition.

The sum of the dimensions of the Morse homology groups cannot be larger than the sum of the dimensions of the chain groups $C_i(f)$, which is the total number of critical points. The above theorem therefore implies the following basic version of the "Morse inequalities":

**Corollary 2** *The number of critical points of a Morse function $f: B \to \mathbb{R}$ cannot be less than $\sum_i \dim H_i(B; \mathbb{F}_2)$.*

The Morse complex can be refined in various ways. For example, one can use integer coefficients in place of coefficients $\mathbb{F}_2$ by taking account of orientations of the spaces of trajectories. One can also introduce Morse theory with coefficients in a local system, and in both these cases a version of the above theorem continues to hold. One can also study the Morse complex of a multivalued Morse function: that is, one can start with closed 1-form $\alpha$ on $B$, with nontrivial periods, and study the flow generated by the corresponding vector field $-g^{-1}\alpha$. Such a theory was developed by Novikov.

The Morse complex can be generalized in a different direction, replacing $f$ by a functional

related to a geometric problem. The canonical example of this (and one of the very few cases in which the theory works as in the finite-dimensional case) is the case that $B = LW$ is the space of loops $u : S^1 \to W$ in a Riemannian manifold $W$ and $f$ is the "energy function," $f_E(u) = \int (du/dt)^2\, dt$. If the Morse–Smale condition holds, then the Morse homology $H_i(f_E)$ computes the homology of $LW$, as expected. Critical points of $f_E$ are geodesics, and the relationship between geodesics and the topology of $LW$, for which Corollary 2 provides a prototype, is an idea with many applications.

For the energy functional, the downward gradient-flow equation is a parabolic equation (the ordinary heat equation if the target space is Euclidean), and a solution to the flow exists for each choice of initial condition. Floer homology can be loosely characterized as the Morse theory of certain variational problems for which the gradient-flow equation is not parabolic, but elliptic of first order: the important models are the Cauchy–Riemann equation in dimension 2, the anti-self-dual Yang–Mills equations in dimension 4, or the closely related Seiberg–Witten equations. For an elliptic equation, one does not expect to solve the Cauchy problem with arbitrary initial condition; so with Floer homology, one is studying a functional for which the gradient flow is not everywhere defined. However, to define the Morse complex, the important thing is only that we have a good understanding of the trajectory spaces $M(p, q)$, which will now be solution spaces for an elliptic problem of geometric origin. The proof of Theorem 1 depends very much on the fact that the flow is everywhere defined: this theorem will therefore fail for the Morse complexes arising in Floer theory, and one must look elsewhere for a means to compute the Morse homology groups.

Before discussing Floer homology in more specific terms, we shall describe the problem in symplectic geometry that motivated its development.

## The Arnol'd Conjecture

A symplectic manifold of dimension $2n$ is a smooth manifold $W$ equipped with a 2-form $\omega$ which is closed and nondegenerate. On a symplectic manifold, one can associate to each smooth function $H : W \to \mathbb{R}$ a vector field $X_H$ on $W$: the vector field is characterized by the property that

$$\omega(X_H, V) = dH(V)$$

for all vector fields $V$. In this situation, one refers to $H$ as the Hamiltonian and $X_H$ as the corresponding Hamiltonian vector field. If $W$ is compact, or if $X_H$ is otherwise complete, then this vector field generates a flow $\phi_t : W \to W (t \in \mathbb{R})$. We also wish to consider the case that $H$ is time dependent: we suppose that $H_t : W \to \mathbb{R}$ is a Hamiltonian which varies smoothly with $t \in \mathbb{R}$ and is periodic, in that $H_{t+1} = H_t$. In this case, there is a time-dependent Hamiltonian vector field $X_t$, and we can consider the flow $\phi_t$ that it generates: so for $x \in W$, the path $\phi_t(x)$ will be the solution to

$$\frac{d}{dt}\phi_t(x) = X_t(x) \qquad [4]$$

with initial condition $\phi_0(x) = x$. The Arnol'd conjecture, in one formulation, concerns the 1-periodic solutions to this equation, or equivalently the fixed points of $\phi_1 : W \to W$. A fixed point $x$ with $\phi_1(x) = x$ is called nondegenerate if $d\phi_1 : T_xX \to T_xX$ does not have 1 as an eigenvalue. With this understood, one version of the conjecture states:

**Conjecture 3** Suppose $W$ is compact and let $H_t$ be any 1-periodic, time-dependent Hamiltonian. If the fixed points of $\phi_1$ are all nondegenerate, then the number of fixed points is not less than the sum of the Betti numbers of the manifold $W$.

There is another, more general version of this conjecture. Let $L \subset W$ be a closed Lagrangian submanifold: that is, an $n$-dimensional submanifold such that the restriction of $\omega$ to $L$ as a 2-form is identically zero. Let $L' \subset W$ be another Lagrangian, obtained from $L$ by a Hamiltonian isotopy: that is, $L'$ is $\phi_1(L)$, for some flow $\phi_t$ generated by a time-dependent Hamiltonian $H_t$ as above.

**Question 4** If $L$ and $L'$ intersect transversely, is it always true that the number of intersection points of $L$ and $L'$ is at least the sum of the Betti numbers of the manifold $L$:

$$\#(L \cap L') \geq \sum_i \operatorname{rank} H_i(L)?$$

This is phrased as a question rather than a conjecture, because the answer is certainly "no" in some cases. For example, $L$ might be a circle contained in a small disk in a symplectic 2-manifold, in which case there is no reason why $\phi_1$ should not move the disk to be completely disjoint from itself. Nevertheless, with extra hypotheses, it is known that the answer is often "yes."

We can exhibit Conjecture 3 as a special case of Question 4, as follows. Given a symplectic manifold $(V, \omega)$, we can form the product $W = V \times V$, with the symplectic form $\omega_W = -p_1^*\omega + p_2^*\omega$, where the $p_i$ are the two projections. The result of this definition is

that the diagonal in $V \times V$ is a Lagrangian submanifold,

$$L \subset W = V \times V$$

for this symplectic form. Let $H_t$ be a time-dependent Hamiltonian on $V$, and let $\phi_t : V \to V$ be the flow. Then $H_t \circ p_2$ is a time-dependent Hamiltonian generating a flow on $W$. For the flow on $W$, the image $L'$ of the diagonal $L \subset W$ at time 1 is the graph of $\phi_1 : V \to V$. Thus, $(L \cap L')$ can be identified with the set of fixed points of $\phi_1$ in $V$, and an affirmative answer to Question 4 for $L \subset W$ implies Conjecture 3 for $V$.

Conjecture 3 and Question 4 can both be extended to the case of isolated degenerate fixed points of $\phi_1$ for Conjecture 3, or to the case of isolated, nontransverse intersections for Question 4. For example, one can ask whether, in the nontransverse case, the sum of the intersection multiplicities can ever be less than the sum of the Betti numbers.

## Morse Theory and the Arnol'd Conjecture

The Arnol'd conjecture, and the related Question 4, can both be studied by reformulating them as questions about the number of critical points of a carefully chosen functional.

We begin with the situation addressed by Conjecture 3. For simplicity, we suppose that $\pi_2(W)$ is zero. Let $\mathcal{B}$ be the space of smooth, null-homotopic loops in $W$:

$$\mathcal{B} = \{u : S^1 \to W | u \text{ is smooth and null homotopic}\}$$

This is a smooth, infinite-dimensional manifold. There is a natural functional $f_0 : \mathcal{B} \to \mathbb{R}$, the symplectic action, defined as

$$f_0(u) = \int_{D^2} v^*(\omega)$$

where $v : D^2 \to W$ is any extension of the map $u : S^1 \to W$. The extension $v$ exists because $u$ is null homotopic, and the value of $f_0$ is independent of the choice of $v$ because $\pi_2(W) = 0$. This functional can be modified in the presence of a periodic Hamiltonian. Introduce a coordinate $t$ on $S^1$ with period 1, and so regard $u$ as a periodic function of $t$. Write the Hamiltonian as $H_t$ as before, and define

$$f(u) = f_0(u) + \int_0^1 H_t(u(t))\,\mathrm{d}t$$

To compute the first variation of $f$, consider a one-parameter family of loops $u_s(t) = u(s,t)$ parametrized by $s \in \mathbb{R}$. We compute

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}s} f(u_s) &= \int_0^1 \omega\left(\frac{\partial u}{\partial s}, \frac{\partial u}{\partial t}\right) \mathrm{d}t + \int_0^1 \mathrm{d}H_t\left(\frac{\partial u}{\partial s}\right) \mathrm{d}t \\ &= \int_0^1 \omega\left(\frac{\partial u}{\partial s}, \frac{\partial u}{\partial t} - X_t(u)\right) \mathrm{d}t \end{aligned}$$

using the relationship between $\mathrm{d}H_t$ and $X_t$. Thus, a loop $u \in \mathcal{B}$ is a critical point of $f : \mathcal{B} \to \mathbb{R}$ if and only if it is a solution of the equation

$$\frac{\mathrm{d}u}{\mathrm{d}t} = X_t(u(t)) \qquad [5]$$

This means that there is a one-to-one correspondence between these critical points and certain 1-periodic solutions of eqn [4]: these in turn correspond to fixed points $p$ of $\phi_1$ with the additional property that the path $\phi_t(p)$ from $p$ to $p$ is null homotopic.

To consider the formal gradient flow of the functional $f$, on must introduce a metric on $\mathcal{B}$. A Riemannian metric $g$ on the symplectic manifold $(W, \omega)$ is compatible with $\omega$ if there is an almost-complex structure $J : TW \to TW$ such that $\omega(X, Y) = g(JX, Y)$ for all tangent vectors $X$ and $Y$ at any point of $W$. Let $g_t$ be a 1-periodic family of compatible Riemannian metrics on $W$. Using these, on can define an inner product on the tangent bundle of $\mathcal{B}$ by the formula

$$\langle U, V \rangle = \int_0^1 g_t(U(t), V(t))\,\mathrm{d}t$$

in which $U$ and $V$ are tangent vectors at $u \in \mathcal{B}$, regarded as vector fields along the loop $u$ in $W$. We can rewrite the above formula for the variation of $f$ in terms of this inner product:

$$\left\langle \frac{\partial u}{\partial s}, J_t\left(\frac{\partial u}{\partial t} - X_t(u)\right)\right\rangle$$

where $J_t$ is the almost-complex structure corresponding to $g_t$. Formally then, a one-parameter family of loops $u(s,t)$ is a solution of the downward gradient-flow equations for the functional $f$ with respect to this metric, if $u$ satisfies the differential equation

$$\frac{\partial u}{\partial s} + J_t\left(\frac{\partial u}{\partial t} - X_t(u)\right) = 0 \qquad [6]$$

In the absence of the term $X_t$, and with $W$ replaced by $\mathbb{C}^n$ with the standard $J$, this equation becomes the Cauchy–Riemann equation $\mathrm{d}u/\mathrm{d}\bar{z} = 0$, for a function $u$ of the complex variable $z = s + \mathrm{i}t$, periodic in $t$.

Let us now suppose we are in the situation of Conjecture 3, so $W$ is closed, and the fixed points of $\phi_1$ are nondegenerate. As we have seen, each fixed point $p$ of $\phi_1$ corresponds to a 1-periodic solution $u_p$ of eqn [5], a critical point of $f$. For each pair of fixed points $p$ and $q$, introduce $M(p,q)$ as the space of solutions of the formal gradient-flow equations of $f$, running from $p$ to $q$: that is, $M(p,q)$ is the space of maps $u:\mathbb{R}\times S^1\to W$ satisfying eqn [6], with

$$\lim_{s\to-\infty} u(s,t) = u_p(t)$$
$$\lim_{s\to+\infty} u(s,t) = u_q(t)$$

With these definitions in place, one can follow the same sequence of steps that we outlined previously in the context of finite-dimensional Morse theory, to construct the Morse complex. First, if $u$ belongs to $M(p,q)$, we can consider the linearization at $u$ of eqns [6], to obtain the counterpart of eqn [1]. These are linear equations for a vector field $U(s,t)$ along $u$ in $W$, and take the form

$$\nabla_{\partial/\partial s}U + J_t\nabla_{\partial/\partial t}U + h(U) = 0 \qquad [7]$$

where $h$ is a linear operator of order zero. Let $\epsilon_u$ denote the dimension of the space solutions $U$ which decay at $s=\pm\infty$, and let $\epsilon_u'$ denote the dimension of the space of solutions of the formal adjoint equation. Elliptic theory for the Cauchy–Riemann equation, and the nondegeneracy condition for $u_p$ and $u_q$, mean that the operator that appears on the left-hand side of the equation is Fredholm: so both $\epsilon_u$ and $\epsilon_u'$ are finite, and the index $\epsilon_u-\epsilon_u'$ is deformation invariant. This index depends only on $p$ and $q$: we give it a name,

$$\epsilon_u - \epsilon_u' = \mathrm{index}(p,q)$$

As before, $u$ is said to be regular if $\epsilon_u'$ is zero. For suitable choice of the almost-complex structures $J_t$ (or equivalently the metrics $g_t$), the Morse–Smale condition will hold: that is, the trajectories in all spaces $M(p,q)$ are regular. In this case, each $M(p,q)$ is a smooth manifold and has dimension index$(p,q)$ if it is nonempty.

The "relative index" index$(p,q)$ plays the role of the difference of the Morse indices in the finite-dimensional case. It can be defined whether or not $M(p,q)$ is empty by considering an equation such as [7] along an arbitrary path $u(s,t)$. In general, there is no natural way to define the "index" of $p$: if we wish, we can select one fixed point $p_0$ and declare it to have index zero; we can then define index$(p)$ as index$(p,p_0)$. Alternatively, we can regard the critical points as indexed by an affine copy of $\mathbb{Z}$ (without a preferred zero).

Imitating the construction of the Morse complex, we define a vector space $\mathrm{CF}_*$ over $\mathbb{F}_2$ as having a basis consisting of elements $e_p$ indexed by the fixed points $p$. We then define $\delta:\mathrm{CF}_*\to\mathrm{CF}_*$ by

$$\delta e_p = \sum_{\mathrm{index}(p,q)=1} \delta_{pq} e_q$$

where $\delta_{pq}$ is defined by counting points in $\breve{M}(p,q)$ as before. The vector space $\mathrm{CF}_*$ is $\mathbb{Z}$-graded if we make a choice of critical point $p_0$ to have index zero; otherwise, $\mathrm{CF}_*$ has an "affine" $\mathbb{Z}$-grading. The map $\delta$ maps $\mathrm{CF}_i$ into $\mathrm{CF}_{i-1}$.

To show that $\delta$ is well defined, and to show that $\delta\circ\delta=0$, one must show that the zero-dimensional spaces $\breve{M}(p,q)$ are compact, and that the ends of the one-dimensional spaces $\breve{M}(p,r)$ correspond bijectively to broken trajectories, as in the finite-dimensional case. Both of these desired properties hold, under the Morse–Smale conditions; but this is a very special feature of the specific problem. Without the hypothesis that $\pi_2(W)$ is zero, additional noncompactness can arise from the following "bubbling" phenomenon. There could be a sequence of solutions $u^i\in M(p,q)$ to eqns [6], and a point $(s_0,t_0)$ in $\mathbb{R}\times S^1$, such that for suitable constants $\epsilon_i$ converging to zero, the rescaled solutions

$$\tilde{u}^i(\sigma,\tau) = u^i(s_0+\epsilon_i\sigma, t_0+\epsilon_i\tau)$$

converge on compact subsets of the plane $\mathbb{R}^2$ to a nonconstant pseudoholomorphic map $\tilde{u}:\mathbb{CP}^1\to W$, or more precisely a solution of the equation

$$\frac{\partial\tilde{u}}{\partial\sigma} + J_{t_0}\left(\frac{\partial\tilde{u}}{\partial\tau}\right) = 0$$

(In the original coordinates, the derivatives of the $u^i$ would grow like $1/\epsilon_i$ near $(s_0,t_0)$.) A pseudoholomorphic sphere always has nontrivial homology class (and therefore nontrivial homotopy class); so this sort of noncompactness does not occur when $\pi_2(W)=0$.

Granted the compactness results, the proof that $\delta\circ\delta=0$ runs as before, and we can construct a Floer homology group,

$$\mathrm{HF}_* = \ker(\delta)/\mathrm{im}(\delta)$$

Unlike the Morse homology of the energy functional, the Floer homology does not yield the ordinary homology of $\mathcal{B}$. To compute it, one first shows that it depends only on the symplectic manifold $(W,\omega)$, not on the choice of Hamiltonian $H_t$ or metrics $g_t$: this step is similar to the proof that the finite-dimensional Morse homology $H_*(f)$ does not depend on the Morse function. Once one has

established this independence, $HF_*$ can be computed by examining a special case. Floer did this by taking the Hamiltonian to be independent of $t$ and equal to a small negative multiple $-\eta h$ of a fixed Morse function $h : W \to \mathbb{R}$ on the symplectic manifold. If the multiple $\eta \in \mathbb{R}$ is small enough, the only fixed points of $\phi_1$ are the stationary points of the flow, and these are exactly the critical points of $h$. Furthermore the only index-1 solutions of eqn [6] for small $\eta$ are the solutions $u(s,t)$ with no $t$ dependence; and these are the solutions of $du/ds = -\eta\nabla h$, the downward gradient flow of $h$, scaled by $\eta$. In this case therefore, the Floer complex $CF_*$ is precisely the Morse complex $C_*(h)$ of the Morse function $h$, and Theorem 1 yields:

**Theorem 5** *For a periodic, time-dependent Hamiltonian $H_t$ on a closed symplectic manifold $(W, \omega)$ with $\pi_2(W) = 0$, the Floer homology $HF_*$ is isomorphic to the ordinary homology of $W$ with $\mathbb{F}_2$ coefficients, $H_*(W; \mathbb{F}_2)$.*

Because the generators of $CF_*$ correspond to fixed points $p$ of $\phi_1$ such that the path $\phi_t(p)$ is null homotopic, the number of these fixed points is not less than the dimension of $HF_*$, and therefore not less than $\sum_i \dim H_i(W; \mathbb{F}_2)$ because of the above result. The sum of the mod 2 Betti numbers is at least as large as the sum of the ordinary Betti numbers (the dimensions of the rational homology groups); so one deduces, following Floer,

**Corollary 6** *The Arnol'd conjecture (Conjecture 3) holds for symplectic manifolds $(W, \omega)$ satisfying the additional condition $\pi_2(W) = 0$.*

Orientations can be introduced rather as in the case of finite-dimensional Morse theory, allowing one to define Floer groups with arbitrary coefficients.

The Arnol'd conjecture is now known to hold in complete generality, without the hypothesis on $\pi_2$. The proof has been achieved by successive extensions of the Floer homology technique. When $\pi_2(W)$ is nonzero, the space $\mathcal{B}$ is not simply connected. The first complication that arises is that the symplectic action functional $f_0$, and therefore $f$ also, is multivalued. This is not an obstacle initially, because $\nabla f$ is still well defined, and the spaces $M(p,q)$ of gradient trajectories can still be assumed to satisfy the Morse–Smale condition: this is the type of Morse theory considered by Novikov, as mentioned above. Because $\pi_1(\mathcal{B})$ is nontrivial, $M(p,q)$ is a union of parts $M_z(p,q)$, one for each homotopy class of paths from $p$ to $q$. For each homotopy class $z$, we have the index $\text{index}_z(p,q)$, which is the dimension of $M_z(p,q)$.

The spaces $M_z(p,q)$ may now have additional noncompactness, due to the presence of pseudo-holomorphic spheres $\tilde{u} : \mathbb{CP}^1 \to W$. The simplest manifestation is when a sequence $u^i$ in $M_z(p,q)$ "bubbles off" a single such sphere at a point $(s_0, t_0)$, and converges elsewhere to a smooth trajectory $u'$ in $M_{z'}(p,q)$, belonging to a different homotopy class. Let $\sigma$ be the homology class of the sphere $\tilde{u}$. Because the sphere has positive area, the pairing of $\sigma$ with the de Rham class $[\omega]$ is positive: $\langle[\omega], \sigma\rangle > 0$. The indices are related by

$$\text{index}_{z'}(p,q) = \text{index}_z(p,q) - 2\langle c_1(W), \sigma\rangle$$

where $c_1(W) \in H^2(W; \mathbb{Z})$ is the first Chern class of a compatible almost-complex structure. The symplectic manifold is said to be "monotone" if, in real cohomology, $c_1(W)$ is a positive multiple of $[\omega]$. In the monotone case, we always have $\text{index}_{z'}(p,q) < \text{index}_z(p,q)$, and no bubbling off can occur for trajectory spaces $M_z(p,q)$ of index 2 or less: the above formula either makes $M_{z'}(p,q)$ a space of negative dimension (in which case it is empty) or a zero-dimensional space (in which case one has to exploit an additional transversality argument, to show that the holomorphic spheres belonging to classes $\sigma$ with $\langle c_1(W), \sigma\rangle = 1$ cannot intersect one of the loops $u_p$ in $W$). Since the construction of $HF_*$ involves only the trajectories of indices 1 and 2, the construction goes through with minor changes. Because $\text{index}_z(p,q)$ depends on the path $z$, the group $HF_*$ will no longer be $\mathbb{Z}$-graded: the grading is defined only modulo $2d$, where $d$ is the smallest nonzero value of $\langle c_1(W), \sigma\rangle$ for spherical classes $\sigma$.

In the case that $W$ is not monotone, additional techniques are needed to deal with the essential noncompactness of the trajectory spaces. These techniques involve (amongst other things) multi-valued perturbations on orbifolds – a strategy that requires the use of rational coefficients in order to perform the necessary averaging. For this reason, in the monotone case, the Arnol'd conjecture is known to hold only in its original form: with the ordinary (rational) Betti numbers.

To address Question 4 for Lagrangian intersections, a closely related Floer homology theory is used. Assume $L$ is connected, and introduce the space of smooth paths joining $L$ to $L'$:

$$\begin{aligned}&\Omega(W; L, L')\\&\quad = \{u : [0,1] \to W \mid u(0) \in L, u(1) \in L'\}\end{aligned}$$

Fix a point $x_0$ in $L$, and let $u_0$ be the path $u_0(t) = \phi_t(x_0)$. Let $\mathcal{B}$ be the connected component

of $\Omega(W; L, L')$ containing $u_0$. On $\mathcal{B}$ we have a symplectic action functional, defined as

$$f(u) = \int_{[0,1]\times[0,1]} v^*(\omega)$$

where $v : [0,1] \times [0,1] \to W$ is a path in $\mathcal{B}$ with $v(0,t) = u_0(t)$ and $v(1,t) = u(t)$. The symplectic action is single valued if $\pi_2(W, L)$ is trivial (even though this condition does not guarantee that $\mathcal{B}$ is simply connected). The critical points of $f$ correspond to constant paths whose image in $W$ is an intersection point of $L$ and $L'$ (though not all such constant paths belong to the connected component $\mathcal{B}$). If we fix a one-parameter family of compatible metrics $g_t$ and almost-complex structures $J_t$ on $W$, then we can consider the downward gradient trajectories of the functional. These are maps

$$u : \mathbb{R} \times [0,1] \to W$$

satisfying the Cauchy–Riemann equation

$$\frac{\partial u}{\partial s} + J_t\left(\frac{\partial u}{\partial t}\right) = 0$$

with boundary conditions $u(s,0) \in L$ and $u(s,1) \in L'$. With coefficients $\mathbb{F}_2$, a Morse complex can be constructed much as in the case just considered. If $\pi_2(W,L)$ is trivial, then the Floer homology group $\mathrm{HF}_*$ obtained as the homology of this Morse complex is isomorphic to $H_*(L; \mathbb{F}_2)$; and as a corollary, Question 4 has an affirmative answer in this case.

Without the hypothesis that $\pi_2(W,L)$ is trivial, one does not expect an affirmative answer to Question 4 in all cases. There is a "monotone" case, in which $\mathrm{HF}_*$ can always be defined; but it is not always isomorphic to $H_*(L; \mathbb{F}_2)$: instead, there is a spectral sequence relating the two. In the general case, there is once again the need to use rational coefficients in place of mod 2 coefficients, in order to deal with the orbifold nature of the trajectory spaces that appear. This raises the question of orientability for the trajectory spaces. In contrast to the Morse theory for Hamiltonian diffeomorphisms, there is an obstruction to orientability, involving spin structures on $L$ and $W$. Even when the trajectory spaces are orientable, there are further obstructions to the existence of a Morse differential satisfying $\delta \circ \delta = 0$. The theory of these obstructions is developed in Fukaya *et al.* (2000). There are still open questions in this area.

## Instanton Floer Homology

A "Floer homology theory" for 3-manifolds should assign to each 3-manifold $Y$ (satisfying perhaps some additional topological requirements) a group, say $\mathrm{HF}(Y)$. Furthermore, given a four-dimensional cobordism $W$ from $Y_1$ to $Y_2$, the theory should provide a corresponding homomorphism of groups, from $\mathrm{HF}(Y_1)$ to $\mathrm{HF}(Y_2)$. These homomorphisms should satisfy the natural composition law for composite cobordisms. One can formulate this by considering the category in which an object is a closed, connected, oriented 3-manifold $Y$, and in which the morphisms from $Y_1$ to $Y_2$ are the oriented four-dimensional cobordisms, considered up to diffeomorphism. A Floer homology theory is then a functor from this category (perhaps with some additional decorations or restrictions) to the category of groups. Such a functor was constructed by Floer (1988a), at least for the full subcategory of homology 3-spheres (manifolds $Y$ with $H_1 \times (Y; \mathbb{Z}) = 0$). We outline the construction.

Let $P \to Y$ be a principal SU(2) bundle (necessarily trivial). Let $\mathcal{A}$ denote the space of SU(2) connections in the bundle $P$, and let $A_0$ be any chosen basepoint in $\mathcal{A}$. Any other $A \in \mathcal{A}$ can be written as $A_0 + a$, for some 1-form $a$ with values in the adjoint bundle $\mathrm{ad}(P)$ whose fiber is the Lie algebra $\mathfrak{su}(2)$. So $\mathcal{A}$ is an affine space,

$$\mathcal{A} = A_0 + \Omega^1(Y; \mathrm{ad}(P))$$

and we can identify the tangent space $T_A\mathcal{A}$ at any $A$ with $\Omega^1(Y; \mathrm{ad}(P))$. The Chern–Simons functional is a smooth function

$$\mathrm{CS} : \mathcal{A} \to \mathbb{R}$$

depending on our choice of a reference connection $A_0$. It can be defined by stating that its derivative at $A \in \mathcal{A}$ is the linear map $T_A\mathcal{A} \to \mathbb{R}$ given by

$$a \mapsto -\int_Y \mathrm{tr}(a \wedge F_A)$$

where $F_A$ denotes the curvature of $A$, as an $\mathrm{ad}(P)$-valued 2-form on $Y$, and tr denotes the trace of a matrix-valued 3-form. If we equip $Y$ with a Riemannian metric, then we have the $L^2$ inner product on $\Omega^1(Y; \mathrm{ad}(P))$, with respect to which we can consider the gradient of CS. The formal downward gradient-flow equation on $\mathcal{A}$ is then

$$(\mathrm{d}/\mathrm{d}s)A = - * F_A \qquad [8]$$

where $*$ is the Hodge star on $Y$. If $A(s)$ is a solution defined on an interval $[s_1, s_2]$, then we can form the corresponding four-dimensional connection $\boldsymbol{A}$ on $[s_1, s_2] \times Y$, and eqn [8] implies that $\boldsymbol{A}$ is a solution of the anti-self-dual Yang–Mills equation, $F^+_{\boldsymbol{A}} = 0$. Here $F^+_{\boldsymbol{A}}$ is the self-dual part of the curvature 2-form on the cylinder. The critical points of CS are the flat connections on $Y$, with $F_A = 0$.

Let $\mathcal{G}$ denote the gauge group, by which we mean the group of automorphisms of $P$. When a trivialization of $P$ is chosen, $\mathcal{G}$ becomes the group of smooth maps $g: Y \to SU(2)$. A connection $A \in \mathcal{A}$ is irreducible if its stabilizer in $\mathcal{G}$ consists only of the constant gauge transformations $\pm 1$. The functional CS is invariant only under the identity component of $\mathcal{G}$: it descends to a function $CS: \mathcal{A}/\mathcal{G} \to \mathbb{R}/(4\pi^2\mathbb{Z})$. If we choose a basepoint in $Y$, then the gauge-equivalence classes of flat connections in $\mathcal{A}$ are in one-to-one correspondence with conjugacy classes of representations,

$$\rho : \pi_1(Y) \to SU(2)$$

Given representations $\rho$ and $\sigma$, we write $M(\rho, \sigma)$ for the quotient by $\mathcal{G}$ of the space of trajectories $A(s)$ which satisfy the gradient-flow equation [8] and which are asymptotic to flat connections belonging to the classes $\rho$ and $\sigma$ as $s \to \pm\infty$. There is a purely four-dimensional interpretation of $M(\rho, \sigma)$: it can be identified with the moduli space of solutions $\boldsymbol{A}$ to the anti-self-dual Yang–Mills equation, or "instantons," on $\mathbb{R} \times Y$, satisfying the same asymptotic conditions.

One defines the "instanton Floer homology" of $Y$, roughly speaking, as the Morse homology arising from the functional CS. In the case that $Y$ is a homology 3-sphere, Floer defined $I_*(Y)$ as the homology $H_*(C, \delta)$ of a complex $C$ whose generators correspond to the irreducible representations $\rho$, and whose differential $\delta$ is defined in terms of the one-dimensional components of the moduli spaces $M(\rho, \sigma)$. To carry out the construction of $I_*(Y)$, it is necessary to perturb the functional CS to achieve a Morse–Smale condition: this is done by adding a function $f: \mathcal{A} \to \mathbb{R}$ defined in terms of the holonomy of connections along families of loops in $Y$. The group $\mathcal{G}$ is not connected, and for given $\rho$ and $\sigma$, the moduli space $M(\rho, \sigma)$ has components differing in dimension by multiples of 8. For this reason, $I_*(Y)$ is a $\mathbb{Z}/8$-graded homology theory. It is a topological invariant of $Y$, and is functorial for cobordisms, in the manner outlined at the beginning of this section.

Various extensions have been made, to allow the definition of $I_*(Y)$ for 3-manifolds with nontrivial $H_1$, and to incorporate the reducible representations. Although there have been some successes (Donaldson 2002), a completely satisfactory general theory has not been constructed. The main difficulties stem from the noncompactness of the instanton moduli spaces (a bubbling phenomenon) and the interaction of this bubbling with the reducible solutions.

The instanton Floer theory for 3-manifolds is closely tied up with Donaldson's polynomial invariants of closed 4-manifolds, which are also defined using the anti-self-dual Yang–Mills equations.

## Seiberg–Witten Floer Homology

Seiberg–Witten Floer homology can be defined in a manner very similar to the instanton case. Again, we start with a Riemannian 3-manifold $Y$, equipped now with a spin$^c$ structure $\mathfrak{s}$: a rank-2 Hermitian vector bundle $S \to Y$ together with a Clifford multiplication $\rho: \Lambda^*(Y) \to \mathrm{End}(S)$. The configuration space $\mathcal{C}$ is defined as the space of pairs $(A, \Phi)$, where $A$ is a spin$^c$ connection and $\Phi$ is a section of $S$. In place of the Chern–Simons functional considered above, we have the Chern–Simons–Dirac functional $CSD: \mathcal{C} \to \mathbb{R}$ defined by

$$CSD(A, \Phi) = \frac{1}{4} CS(\mathrm{tr}(A)) + \frac{1}{2} \int_Y \langle \Phi, D_A \Phi \rangle \, d\mu$$

where $\mathrm{tr}(A)$ denotes the connection induced by $A$ on the line bundle $\Lambda^2 S$ and $D_A$ is the Dirac operator for the connection $A$. The functional is invariant again under the identity component of the gauge group $\mathcal{G}$, which this time is the group of maps $g: Y \to S^1$, acting as automorphisms of $S$. The critical points are the solutions $(A, \Phi)$ to the three-dimensional "Seiberg–Witten equations,"

$$\tfrac{1}{2}\rho(F_{\mathrm{tr}(A)}) - (\Phi\Phi^*)_0 = 0$$
$$D_A \Phi = 0$$

in which the subscript 0 denotes the traceless part of the endomorphism. If $\alpha$ and $\beta$ are gauge-equivalence classes of critical points, then we write $M(\alpha, \beta)$ for the quotient by $\mathcal{G}$ of the space of gradient trajectories from $\alpha$ to $\beta$.

As in the instanton case, $M(\alpha, \beta)$ has a four-dimensional interpretation: it is the quotient by the four-dimensional gauge group of a space of solutions $(\boldsymbol{A}, \boldsymbol{\Phi})$ on $\mathbb{R} \times Y$ to the four-dimensional Seiberg–Witten equations:

$$\tfrac{1}{2}\rho\left(F^+_{\mathrm{tr}(\boldsymbol{A})}\right) - (\boldsymbol{\Phi}\boldsymbol{\Phi}^*)_0 = 0$$
$$D^+_{\boldsymbol{A}} \boldsymbol{\Phi} = 0$$

Here $\boldsymbol{\Phi}$ is a section of the summand $S^+$ of the four-dimensional spin$^c$ bundle $S = S^+ \oplus S^-$, and $D^+_{\boldsymbol{A}}: \Gamma(S^+) \to \Gamma(S^-)$ is the four-dimensional Dirac operator.

The action of the gauge group on $\mathcal{C}$ is free except at configurations with $\Phi = 0$. These reducible configurations have an $S^1$ stabilizer. Reducible critical points of CSD correspond to flat connections in the line bundle $\Lambda^2 S$. We can now distinguish two cases, according to whether $c_1(S)$ is a torsion class or not.

If $c_1(S)$ is not a torsion class, then there are no flat connections in $\Lambda^2 S$, so all critical points are irreducible. In this case, there is a straightforward Floer-type Morse theory for the functional CSD on

the space $\mathcal{C}/\mathcal{G}$: for generators of our complex we take the gauge-equivalence classes of critical points, and we use the one-dimensional trajectory spaces $M(\alpha, \beta)$ to define the boundary map. The resulting Morse homology group is denoted $\mathrm{HM}_*(Y, \mathfrak{s})$. It has a canonical $\mathbb{Z}/2$-grading, and is a topological invariant of $Y$ and its spin$^c$ structure.

If $c_1(S)$ is torsion, the theory is more complex. There will be reducible critical points, and one cannot exclude these from the Morse complex and still obtain a topological invariant of $Y$. One may incorporate the reducible critical points in two different ways, that are in a sense dual to one another; and there is a third homology theory that one can define, using the reducibles alone. Thus, one can construct three Floer groups associated to $Y$ with the spin$^c$ structure $\mathfrak{s}$. The resulting theory closely resembles the Heegaard Floer homology that is described next.

## Heegaard Floer Homology and Other Floer Theories

Heegaard Floer homology is a Floer homology theory for 3-manifolds that is formally similar to Seiberg–Witten Floer homology, and conjecturally isomorphic to it. Unlike the instanton and Seiberg–Witten theories, its construction, due to Ozsváth and Szabó, does not use gauge theory. Instead, one begins with a decomposition of the 3-manifold into two handlebodies with common boundary $\Sigma$, and one studies a symplectic manifold $s^g\Sigma$, the configuration space of $g$-tuples of points on $\Sigma$, where $g$ denotes the genus. The Heegaard Floer groups are then defined by a variant of the construction used for Lagrangian intersections (see the section "Morse theory and the Arnol'd conjecture"), applied to a particular pair of Lagrangian tori in $s^g\Sigma$.

As in the case of Seiberg–Witten theory, Heegaard Floer homology assigns to each oriented 3-manifold $Y$ three different Floer groups, $\mathrm{HF}^+(Y), \mathrm{HF}^-(Y)$, and $\mathrm{HF}^\infty(Y)$, related by a long exact sequence:

$$\cdots \to \mathrm{HF}^+(Y) \to \mathrm{HF}^-(Y) \to \mathrm{HF}^\infty(Y) \to \mathrm{HF}^+(Y) \to \cdots$$

The first two groups are dual, in that there is a nondegenerate pairing between $\mathrm{HF}^+(Y)$ and $\mathrm{HF}^-(-Y)$, where $-Y$ denotes the same 3-manifold with opposite orientation. If $W$ is an oriented four-dimensional cobordism from $Y_1$ to $Y_2$, then there are associated functorial maps

$$\mathcal{F}^+(W) : \mathrm{HF}^+(Y_1) \to \mathrm{HF}^+(Y_2)$$
$$\mathcal{F}^-(W) : \mathrm{HF}^-(Y_1) \to \mathrm{HF}^-(Y_2)$$
$$\mathcal{F}^\infty(W) : \mathrm{HF}^\infty(Y_1) \to \mathrm{HF}^\infty(Y_2)$$

In addition, if the intersection form of $W$ is not negative semidefinite, there is a map

$$\mathcal{F}(W) : \mathrm{HF}^-(Y_1) \to \mathrm{HF}^+(Y_2)$$

As a special case, one can start with a closed 4-manifold $X$, and consider the cobordism $W$ from $S^3$ to $S^3$ obtained from $X$ by removing two 4-balls. In this case, the map

$$\mathcal{F}(W) : \mathrm{HF}^-(S^3) \to \mathrm{HF}^+(S^3)$$

encodes a diffeomorphism invariant of the original 4-manifold $X$. This invariant is conjectured to be equivalent to the Seiberg–Witten invariants of $X$.

Heegaard Floer homology, and its cousin Seiberg–Witten Floer homology, have been applied successfully to settle long-standing problems in topology, particularly questions related to surgery on knots. An example of such an application is the theorem of Kronheimer *et al.* that one cannot obtain the projective space $\mathbb{RP}^3$ by surgery on a nontrivial knot in the 3-sphere.

In these and other applications of both Heegaard and Seiberg–Witten Floer homology, two key properties of the homology groups play an important part. The first is a nonvanishing theorem, which shows, for example, that these Floer groups can distinguish $S^1 \times S^2$ from any other manifold with the same homology. The second is a long exact sequence, which relates the Floer groups of the manifolds obtained by three different surgeries on a knot. The latter property is shared by the instanton Floer groups, as was shown by Floer (Braam and Donaldson 1995).

Other Floer-type theories have been considered, not all of which arise from a gradient flow, but in which the boundary map of the complex is obtained by counting solutions to a geometric differential equation. At the time of writing, Floer homology is an area of very active development.

*See also:* Four-Manifold Invariants and Physics; Gauge Theoretic Invariants of 4-Manifolds; Gauge Theory: Mathematical Applications; Knot Homologies; Ljusternik–Schnirelman Theory; Minimax Principle in the Calculus of Variations; Moduli Spaces: An Introduction; Seiberg–Witten Theory; Topological Quantum Field Theory: Overview.

## Further Reading

Braam PJ and Donaldson SK (1995) Floer's work on instanton homology, knots and surgery. In: *The Floer Memorial Volume*, Progr. Math., vol. 133, pp. 195–256. Basel: Birkhäuser.

Conley C and Zehnder E (1986) A global fixed point theorem for symplectic maps and subharmonic solutions of Hamiltonian equations on tori. In: *Nonlinear Functional Analysis and Its Applications, Part 1 (Berkeley, Calif., 1983)*, Proc. Sympos.

Pure Math., vol. 45, pp. 283–299. Providence, RI: American Mathematical Society.
Donaldson SK (2002) *Floer Homology Groups in Yang–Mills Theory*, Cambridge Tracts in Mathematics, vol. 147. (With the assistance of Furuta M and Kotschick D.), Cambridge: Cambridge University Press.
Floer A (1988a) An instanton-invariant for 3-manifolds. *Communications in Mathematical Physics* 118(2): 215–240.
Floer A (1988b) Morse theory for Lagrangian intersections. *Journal of Differential Geometry* 28(3): 513–547.
Floer A (1989a) Symplectic fixed points and holomorphic spheres. *Communications in Mathematical Physics* 120(4): 575–611.
Floer A (1989b) Witten's complex and infinite-dimensional Morse theory. *Journal of Differential Geometry* 30(1): 207–221.
Fukaya K, Oh Y-G, Ohta H, and Ono K (2000) *Lagrangian intersection Floer theory – anomaly and obstruction*. Preprint.
Kronheimer PB and Mrowka TS Floer homology of Seiberg–Witten monopoles (to appear).
Kronheimer PB, Mrowka TS, Ozsváth P and Szabó Z Monopoles and lens space surgeries. *Annals of Mathematics* (to appear).
Milnor J (1963) Morse Theory. Based on lecture notes by Spivak M and Wells R, Annals of Mathematics Studies, vol. 51. Princeton, NJ: Princeton University Press.
Ozsváth P and Szabó Z Holomorphic triangles and invariants for smooth four-manifolds. *Advances in Mathematics* (to appear).
Ozsváth P and Szabó Z (2004) Holomorphic disks and topological invariants for closed three-manifolds. *Annals of Mathematics (2)* 159(3): 1027–1158.
Poincaré H (1993) *New Methods of Celestial Mechanics*, vol. 1. History of Modern Physics and Astronomy, vol. 13. Periodic and asymptotic solutions, Translated from the French, Revised reprint of the 1967 English translation, With end-notes by V. I. Arnol'd, Edited and with an introduction by Daniel L. Goroff. New York: American Institute of Physics.
Witten E (1982) Supersymmetry and Morse theory. *Journal of Differential Geometry* 17(4): 661–692.

# Fluid Mechanics: Numerical Methods

**J-L Guermond**, Université de Paris Sud, Orsay, France



The objective of this article is to give an overview of some advanced numerical methods commonly used in fluid mechanics. The focus is set primarily on finite-element methods and finite-volume methods.

## Fluid Mechanics Models

Let $\Omega$ be a domain in $\mathbb{R}^d (d=2,3)$ with boundary $\partial\Omega$ and outer unit normal $\boldsymbol{n}$. $\Omega$ is assumed to be occupied by a fluid. The basic equations governing fluid flows are derived from three conservation principles: *conservation of mass*, *momentum*, *and energy*. Denoting the density by $\rho$, the velocity by $\boldsymbol{u}$, and the mass specific internal energy by $e_i$, these equations are

$$\partial_t \rho + \nabla\cdot(\rho\boldsymbol{u}) = 0 \qquad [1]$$

$$\partial_t(\rho\boldsymbol{u}) + \nabla\cdot(\rho\boldsymbol{u}\otimes\boldsymbol{u}) = \nabla\cdot\boldsymbol{\sigma} + \rho\boldsymbol{f} \qquad [2]$$

$$\partial_t(\rho e_i) + \nabla\cdot(\rho\boldsymbol{u}e_i) = \boldsymbol{\sigma}:\boldsymbol{\varepsilon} + q_{\mathrm{T}} - \nabla\cdot\boldsymbol{j}_T \qquad [3]$$

where $\boldsymbol{\sigma}$ is the *stress tensor*, $\boldsymbol{\varepsilon} = (1/2)(\nabla\boldsymbol{u} + \nabla\boldsymbol{u})^{\mathrm{T}}$ is the *strain tensor*, $\boldsymbol{f}$ is a body force per unit mass (gravity is a typical example), $q_T$ is a volume source (it may model chemical reactions, Joule effects, radioactive decay, etc.), and $\boldsymbol{j}_T$ is the heat flux. In addition to the above three fundamental conservation equations, one may also have to add $L$ equations that account for the conservation of other quantities, say $\phi_\ell, 1 \le \ell \le L$. These quantities may, for example, be the concentration of constituents in an alloy, the turbulent kinetic energy, the mass fractions of various chemical species by unit volume, etc. All these conservation equations take the following form:

$$\partial_t(\rho\phi_\ell) + \nabla\cdot(\rho\boldsymbol{u}\phi_\ell) = q_{\phi_\ell} - \nabla\cdot\boldsymbol{j}_{\phi_\ell}, \quad 1 \le \ell \le L \qquad [4]$$

Henceforth, the index $\ell$ is dropped to alleviate the notation.

The above set of equations must be supplemented with initial and boundary conditions. Typical initial conditions are $\rho_{|t=0} = \rho_0$, $\boldsymbol{u}_{|t=0} = \boldsymbol{u}_0$, and $\phi_{|t=0} = \phi_0$. Boundary conditions are usually classified into two types: the *essential boundary conditions* and the *natural boundary conditions*. Natural conditions impose fluxes at the boundary. Typical examples are

$$(\boldsymbol{\sigma}\cdot\boldsymbol{n} + \mathcal{R}\cdot\boldsymbol{u})_{|\partial\Omega} = \boldsymbol{a_u}$$

$$(\boldsymbol{j}_T\cdot\boldsymbol{n} + r_T e_i)_{|\partial\Omega} = a_T$$

and

$$(\boldsymbol{j}_\phi\cdot\boldsymbol{n} + r_\phi\phi)_{|\partial\Omega} = a_\phi$$

The quantities $\mathcal{R}$, $r_T$, $r_\phi$, $\boldsymbol{a_u}$, $a_T$, $a_\phi$ are given. Essential boundary conditions consist of enforcing boundary values on the dependent variables. One typical example is the so-called *no-slip boundary condition*: $\boldsymbol{u}_{|\partial\Omega} = 0$.

The above system of conservation laws is closed by adding three constitutive equations whose purpose is to relate each field $\boldsymbol{\sigma}$, $\boldsymbol{j}_T$, and $\boldsymbol{j}_\phi$ to the fields $\rho$, $\boldsymbol{u}$, and $\phi$. They account for microscopic properties of the fluid and thus must be frame-independent. Depending on the constitutive equations and

adequate hypotheses on time and space scales, various models are obtained. An important class of fluid model is one for which the stress tensor is a linear function of the strain tensor, yielding the so-called Newtonian fluid model:

$$\boldsymbol{\sigma} = (-p + \lambda \nabla \cdot \boldsymbol{u})\boldsymbol{I} + 2\mu\boldsymbol{\varepsilon} \qquad [5]$$

Here $p$ is the pressure, $\boldsymbol{I}$ is the identity matrix, and $\lambda$ and $\mu$ are viscosity coefficients. Still assuming linearity, common models for heat and solute fluxes consist of assuming

$$\boldsymbol{j}_T = -\kappa \nabla T, \qquad \boldsymbol{j}_\phi = -D\nabla\phi \qquad [6]$$

where $T$ is the temperature. These are the so-called Fourier's law and Fick's law, respectively.

Having introduced two new quantities, namely the pressure $p$ and the temperature $T$, two new scalar relations are needed to close the system. These are the state equations. One admissible assumption consists of setting $\rho = \rho(p, T)$. Another usual additional hypothesis consists of assuming that the variations in the internal energy are proportional to those in the temperature, that is, $\partial e_i = c_P \partial T$.

Let us now simplify the above models by assuming that $\rho$ is constant. Then, mass conservation implies that the flow is incompressible, that is, $\nabla \cdot \boldsymbol{u} = 0$. Let us further assume that neither $\lambda, \mu$, nor $p$ depend on $e_i$. Then, upon abusing the notation and still denoting by $p$ the ratio $p/\rho$, the above set of assumptions yields the so-called incompressible Navier–Stokes equations:

$$\nabla \cdot \boldsymbol{u} = 0 \qquad [7]$$

$$\partial_t \boldsymbol{u} + \boldsymbol{u} \cdot \nabla \boldsymbol{u} - \nu \Delta \boldsymbol{u} + \nabla p = \boldsymbol{f} \qquad [8]$$

As a result, the mass and momentum conservation equations are independent of that of the energy and those of the solutes:

$$\rho c_P(\partial_t T + \boldsymbol{u} \cdot \nabla T) - \nabla \cdot (\kappa \nabla T) = 2\mu\boldsymbol{\varepsilon} : \boldsymbol{\varepsilon} + q_T \qquad [9]$$

$$\partial_t \phi + \boldsymbol{u} \cdot \nabla \phi - \frac{1}{\rho} \nabla \cdot (D \nabla \phi) = \frac{1}{\rho} q_\phi \qquad [10]$$

Another model allowing for a weak dependency of $\rho$ on the temperature, while still enforcing incompressibility, consists of setting $\rho = \rho_0(1 - \beta(T - T_0))$. If buoyancy effects induced by gravity are important, it is then possible to account for them by setting $\boldsymbol{f} = \rho_0 g(1 - \beta(T - T_0))$, where $g$ is the gravitational acceleration, yielding the so-called Boussinesq model.

Variations on these themes are numerous and a wide range of fluids can be modeled by using nonlinear constitutive laws and nonlinear state laws. For the purpose of numerical simulations, however, it is important to focus on simplified models.

## The Building Blocks

From the above considerations we now extract a small set of elementary problems which constitute the building blocks of most numerical methods in fluid mechanics.

### Elliptic Equations

By taking the divergence of the momentum equation [8] and assuming $\boldsymbol{u}$ to be known and renaming $p$ to $\phi$, one obtains the Poisson equation

$$-\Delta \phi = f \qquad [11]$$

where $f$ is a given source term. This equation plays a key role in the computation of the pressure when solving the Navier–Stokes equations; see [54b]. Assuming that adequate boundary conditions are enforced, this model equation is the prototype for the class of the so-called elliptic equations. A simple generalization of the Poisson equation consists of the advection–diffusion equation

$$\boldsymbol{u} \cdot \nabla \phi - \nabla \cdot (\kappa \nabla \phi) = f \qquad [12]$$

where $\kappa > 0$. Admissible boundary conditions are $(\kappa \partial_n \phi + r\phi)_{|\partial\Omega} = a$, $r \geq 0$, or $\phi_{|\partial\Omega} = a$. This type of equation is obtained by neglecting the time derivative in the heat equation [9] or in the solute conservation equation [10]. Mathematically speaking, [12] is also elliptic since its properties (in particular, the way the boundary conditions must be enforced) are controlled by the second-order derivatives. For the sake of simplicity, assume that $\boldsymbol{u} = 0$ in the above equation and that the boundary condition is $\phi_{|\partial\Omega} = 0$, then it is possible to show that $\phi$ solves [12] if and only if $\phi$ minimizes the functional

$$\mathcal{J}(\psi) = \int_\Omega (|\nabla \psi|^2 - f\psi)\, \mathrm{d}\boldsymbol{x}$$

where $|\cdot|$ is the Euclidean norm and $\psi$ spans

$$H = \left\{ \psi; \int_\Omega |\nabla \psi|^2 \, \mathrm{d}\boldsymbol{x} < \infty; \psi_{|\partial\Omega} = 0 \right\} \qquad [13]$$

Writing the first-order optimality condition for this optimization problem yields

$$\int_\Omega \nabla \phi \cdot \nabla \psi = \int_\Omega \boldsymbol{f} \psi$$

for all $\psi \in H$. This is the so-called variational formulation of [12]. When $\boldsymbol{u}$ is not zero, no variational principle holds but a similar way to

reformulate [12] consists of multiplying the equation by arbitrary functions in $H$ and integrating by parts the second-order term to give

$$\int_\Omega (\boldsymbol{u}\cdot\nabla\phi)\psi + \kappa\nabla\phi\cdot\nabla\psi = \int_\Omega f\psi, \quad \forall\psi\in H \qquad [14]$$

This is the so-called weak formulation of [12]. Weak and variational formulations are the starting point for finite-element approximations.

### Stokes Equations

Another elementary building block is deduced from [8] by assuming that the time derivative and the nonlinear term are both small. The corresponding model is the so-called Stokes equations,

$$-\nu\Delta\boldsymbol{u} + \nabla p = \boldsymbol{f} \qquad [15]$$

$$\nabla\cdot\boldsymbol{u} = 0 \qquad [16]$$

Assume for the sake of simplicity that the no-slip boundary condition is enforced: $\boldsymbol{u}_{|\partial\Omega}=\boldsymbol{0}$. Introduce the Lagrangian functional

$$\mathcal{L}(\boldsymbol{v},q) = \int_\Omega (\nabla\boldsymbol{u}:\nabla\boldsymbol{v} - q\nabla\cdot\boldsymbol{v} - \boldsymbol{f}\cdot\boldsymbol{v})\,\mathrm{d}\boldsymbol{x}$$

Set

$$\boldsymbol{X} = \left\{\boldsymbol{v};\int_\Omega |\nabla\boldsymbol{v}|^2\,\mathrm{d}\boldsymbol{x} < \infty; \boldsymbol{v}_{|\partial\Omega} = 0\right\}$$
$$M = \left\{q;\int_\Omega q^2\,\mathrm{d}\boldsymbol{x} < \infty\right\}$$

Then, the pair $(\boldsymbol{u},p)\in\boldsymbol{X}\times M$ solves the Stokes equations if and only if it is a saddle point of $\mathcal{L}$, that is,

$$\mathcal{L}(\boldsymbol{u},q) \le \mathcal{L}(\boldsymbol{u},p) \le \mathcal{L}(\boldsymbol{v},p), \quad \forall(\boldsymbol{v},q)\in\boldsymbol{X}\times M \qquad [17]$$

In other words, the pressure $p$ is the Lagrange multiplier of the incompressibility constraint $\nabla\cdot\boldsymbol{u}=0$. Realizing this fact helps to understand the nature of the Stokes equations, specially when it comes to constructing discrete approximations. A variational formulation of the Stokes equations is obtained by writing the first-order optimality condition, namely:

$$\int_\Omega (\nu\nabla\boldsymbol{u}:\nabla\boldsymbol{v} - p\nabla\cdot\boldsymbol{v} - \boldsymbol{f}\cdot\boldsymbol{v})\,\mathrm{d}\boldsymbol{x} = 0 \quad \forall\boldsymbol{v}\in\boldsymbol{X}$$
$$\int_\Omega q\nabla\cdot\boldsymbol{u}\,\mathrm{d}\boldsymbol{x} = 0 \quad \forall q\in M$$

When the nonlinear term is not zero in the momentum equation, or when this term is linearized, there is no saddle point, but a weak formulation is obtained by multiplying the momentum equation by arbitrary functions $\boldsymbol{v}$ in $\boldsymbol{X}$ and integrating by parts the Laplacian, and by multiplying the mass equation by arbitrary functions $q$ in $M$:

$$\int_\Omega ((\boldsymbol{u}\cdot\nabla\boldsymbol{u})\cdot\boldsymbol{v} + \nu\nabla\boldsymbol{u}:\nabla\boldsymbol{v} - p\nabla\cdot\boldsymbol{v})\,\mathrm{d}\boldsymbol{x} = \int_\Omega \boldsymbol{f}\cdot\boldsymbol{v} \qquad [18]$$

$$\int_\Omega q\nabla\boldsymbol{u} = 0 \qquad [19]$$

### Parabolic Equations

The class of elliptic equations generalizes to that of the parabolic equations when time is accounted for:

$$\partial_t\phi + \boldsymbol{u}\cdot\nabla\phi - \nabla\cdot(\kappa\nabla\phi) = f, \quad \phi_{|t=0} = \phi_0 \qquad [20]$$

Fundamentally, this equation has many similarities with the elliptic equation

$$\alpha\phi + \boldsymbol{u}\cdot\nabla\phi - \nabla\cdot(\kappa\nabla\phi) = f \qquad [21]$$

where $\alpha > 0$. In particular, the set of boundary conditions that are admissible for [20] and [21] are identical, that is, it is legitimate to enforce $(\kappa\partial_n\phi + r\phi)_{|\partial\Omega} = a, r\ge 0$, or $\phi_{|\partial\Omega} = a$. Moreover, solving [21] is always a building block of any algorithm solving [20]. The important fact to remember here is that if a good approximation technique for solving [21] is at hand, then extending it to solve [20] is usually straightforward.

### Hyperbolic Equations

When $\kappa/UL\to 0$, where $U$ is the reference velocity scale and $L$ is the reference length scale, [20] degenerates into the so-called transport equation

$$\partial_t\phi + \boldsymbol{u}\cdot\nabla\phi = f \qquad [22]$$

This is the prototypical example for the class of hyperbolic equations. For this equation to be well-posed, it is necessary to enforce an initial condition $\phi_{|t=0} = \phi_0$ and an inflow boundary condition, that is, $\phi_{|\partial\Omega^-} = a$, where $\partial\Omega^- = \{\boldsymbol{x}\in\partial\Omega; (\boldsymbol{u}\cdot\boldsymbol{n})(\boldsymbol{x}) < 0\}$ is the so-called *inflow boundary* of the domain. To better understand the nature of this equation, introduce the characteristic lines $\boldsymbol{X}(x,s;t)$ of $\boldsymbol{u}(\boldsymbol{x},t)$ defined as follows:

$$\begin{aligned} \mathrm{d}_t\boldsymbol{X}(x,s;t) &= \boldsymbol{u}(\boldsymbol{X}(x,s;t),t) \\ \boldsymbol{X}(x,s;s) &= \boldsymbol{x} \end{aligned} \qquad [23]$$

If $\boldsymbol{u}$ is continuous with respect to $t$ and Lipschitz with respect to $\boldsymbol{x}$, this ordinary differential equation has a unique solution. Furthermore, [22] becomes

$$\mathrm{d}_t[\phi(\boldsymbol{X}(\boldsymbol{x},s;t),t)] = f(\boldsymbol{X}(\boldsymbol{x},s;t),t) \qquad [24]$$

Then

$$\phi(\boldsymbol{x},t) = \phi_0(\boldsymbol{X}(\boldsymbol{x},t;0)) + \int_0^t f(\boldsymbol{X}(\boldsymbol{x},t;\tau),\tau)\,\mathrm{d}\tau$$

provided $\boldsymbol{X}(\boldsymbol{x},t;\tau) \in \Omega$ for all $\tau \in [0,t]$. This shows that the concept of characteristic curves is important to construct an approximation to [22].

## Meshes

The starting point of every approximation technique for solving any of the above model problems consists of defining a mesh of $\Omega$ on which the approximate solution is defined. To avoid having to account for curved boundaries, let us assume that the domain $\Omega$ is a two-dimensional polygon (resp. three-dimensional polyhedron). A mesh of $\Omega$, say $\mathcal{T}_h$, is a partition of $\Omega$ into small cells, hereafter assumed to be simple convex polygons in two dimensions (resp. polyhedrons in three dimensions), say triangles or quadrangles (resp. tetrahedrons or cuboids). Moreover, this partition is usually assumed to be such that if two different cells have a nonempty intersection, then the intersection is a vertex, or an entire edge, or an entire face. The left panel of Figure 1 shows a mesh satisfying the above requirement. The mesh in the right panel is not admissible.

## Finite Elements: Interpolation

The finite-element method is foremost an interpolation technique. The goal of this section is to illustrate this idea by giving examples.

Let $\mathcal{T}_h = \{K_m\}_{1\le m\le N_{\mathrm{el}}}$ be a mesh composed of $N_{\mathrm{el}}$ simplices, that is, triangles in two dimensions or tetrahedrons in three dimensions. Consider the following vector spaces of functions:

$$V_h = \{v_h \in \mathcal{C}^0(\bar{\Omega}); v_{h|K_m} \in \mathbb{P}_k, 1 \le m \le N_{\mathrm{el}}\} \quad [25]$$

where $\mathbb{P}_k$ denotes the space of polynomials of global degree at most $k$. $V_h$ is called a *finite-element approximation space*. We now construct a basis for $V_h$.

Given a simplex $K_m$ in $\mathbb{R}^d$, let $\boldsymbol{v}_n$ be a vertex of $K_m$, let $F_n$ be the face of $K_m$ opposite to $\boldsymbol{v}_n$, and

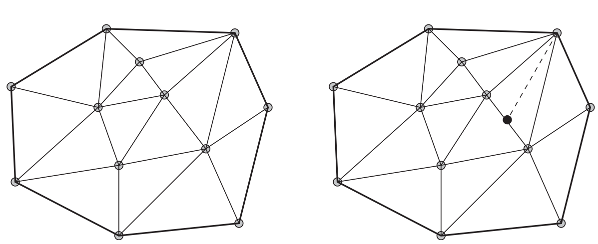


**Figure 1** Admissible (left) and nonadmissible (right) meshes.

define $\boldsymbol{n}_n$ to be the outward normal to $F_n, 1 \le n \le d+1$. Define the *barycentric coordinates*

$$\lambda_n(\boldsymbol{x}) = 1 - \frac{(\boldsymbol{x}-\boldsymbol{v}_n)\cdot \boldsymbol{n}_n}{(\boldsymbol{v}_l-\boldsymbol{v}_n)\cdot\boldsymbol{n}_n}, \quad 1 \le n \le d+1 \quad [26]$$

where $\boldsymbol{v}_l$ is an arbitrary vertex in $F_n$ (the definition of $\lambda_n$ is clearly independent of $\boldsymbol{v}_l$ provided $\boldsymbol{v}_l$ belongs to $F_n$). The barycentric coordinate $\lambda_n$ is an affine function; it is equal to 1 at $\boldsymbol{v}_n$ and vanishes on $F_n$; its level sets are hyperplanes parallel to $F_n$. The barycenter of $K_m$ has barycentric coordinates

$$\left(\frac{1}{d+1},\ldots,\frac{1}{d+1}\right)$$

The barycentric coordinates satisfy the following properties: for all $\boldsymbol{x} \in K_m, 0 \le \lambda_n(\boldsymbol{x}) \le 1$, and for all $\boldsymbol{x} \in \mathbb{R}^d$,

$$\sum_{n=1}^{d+1} \lambda_n(\boldsymbol{x}) = 1 \quad \text{and} \quad \sum_{n=1}^{d+1} \lambda_n(\boldsymbol{x})(\boldsymbol{x}-\boldsymbol{v}_n) = 0$$

Consider the set of nodes $\{\boldsymbol{a}_{n,m}\}_{1\le n\le n_{\mathrm{sh}}}$ of $K_m$ with barycentric coordinates

$$\left(\frac{i_0}{k},\ldots,\frac{i_d}{k}\right), \quad 0 \le i_0,\ldots,\, i_d \le k,\; i_0+\cdots+i_d = k$$

These points are called the *Lagrange nodes* of $K_m$. It is clear that there are $n_{\mathrm{sh}} = (1/2)(k+1)(k+2)$ of these points in two dimensions and $n_{\mathrm{sh}} = (1/6)(k+1)(k+2)(k+3)$ in three dimensions. It is remarkable that $n_{\mathrm{sh}} = \dim \mathbb{P}_k$.

Let $\{\boldsymbol{b}_1,\ldots,\boldsymbol{b}_N\} = \bigcup_{K_m\in\mathcal{T}_h}\{\boldsymbol{a}_{1,m},\ldots,\boldsymbol{a}_{n_{\mathrm{sh}},m}\}$ be the set of all the Lagrange nodes in the mesh. For $K_m \in \mathcal{T}_h$ and $n \in \{1,\ldots,n_{\mathrm{sh}}\}$, let $j(n,m) \in \{1,\ldots,N\}$ be the integer such that $\boldsymbol{a}_{n,m} = \boldsymbol{b}_{j(n,m)}$; $j(n,m)$ is the global index of the Lagrange node $\boldsymbol{a}_{n,m}$. Let $\{\varphi_1,\ldots,\varphi_N\}$ be the set of functions in $V_h$ defined by $\varphi_i(\boldsymbol{b}_j) = \delta_{ij}$, then it can be shown that

$$\{\varphi_1,\ldots,\,\varphi_N\} \text{ is a basis for } V_h \quad [27]$$

The functions $\varphi_i$ are called *global shape functions*. An important property of global shape functions is that their supports are small sets of cells. More precisely, let $i \in \{1,\ldots,N\}$ and let $\mathcal{V}_i = \{m; \exists n; i = j(n,m)\}$ be the set of cell indices to which the node $\boldsymbol{b}_i$ belongs, then the support of $\varphi_i$ is $\bigcup_{m\in\mathcal{V}_i} K_m$. For $k=1$, it is clear that $\varphi_{i\,|\,K_m} = \lambda_n$ for all $m \in \mathcal{V}_i$ and all $n$ such that $i = j(n,m)$, and $\varphi_{i\,|\,K_m} = 0$ otherwise. The graph of such a shape function in two dimensions is shown in the left panel of Figure 2. For $k=2$, enumerate from 1 to $d+1$ the vertices of $K_m$, and enumerate from $d+2$ to $n_{\mathrm{sh}}$ the Lagrange nodes located at the midedges. For a midedge node of index $d+2 \le n \le n_{\mathrm{sh}}$, let $b(n), e(n) \in \{1,\ldots, d+1\}$ be the two indices of the two Lagrange

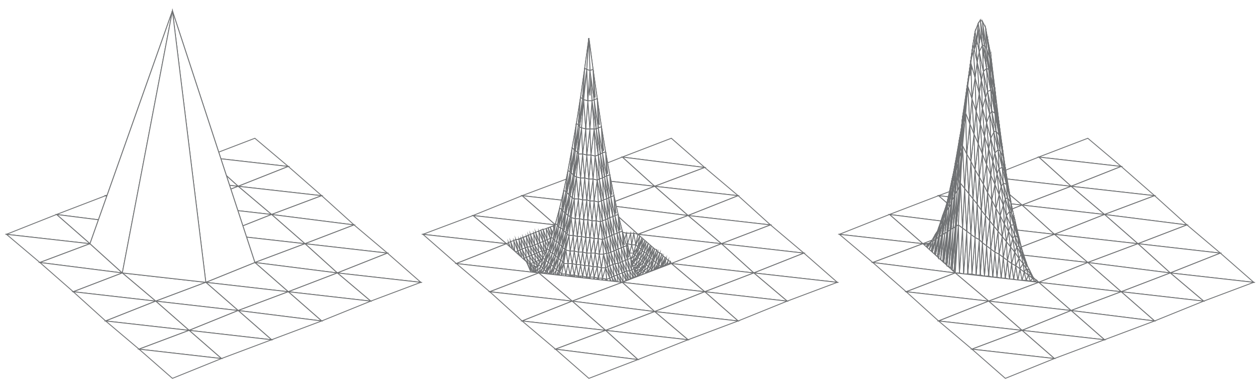


**Figure 2** Two-dimensional Lagrange shape functions: piecewise $\mathbb{P}_1$ (left) and piecewise $\mathbb{P}_2$ (center and right).

nodes at the extremities of the edge in question. Then, the restriction to $K_m$ of a $\mathbb{P}_2$ shape function $\varphi_i$ is

$$\varphi_{i|K_m} = \begin{cases} \lambda_n(2\lambda_n - 1), & \text{if } 1 \le n \le d+1 \\ 4\lambda_{b(n)}\lambda_{e(n)}, & \text{if } d+2 \le n \le n_{\text{sh}} \end{cases} \quad [28]$$

Figure 2 shows the graph of two $\mathbb{P}_2$ shape functions in two dimensions.

Once the space $V_h$ is introduced, it is natural to define the interpolation operator

$$\Pi_h : \mathcal{C}^0(\bar{\Omega}) \ni v \longmapsto \sum_{i=1}^{N} v(\boldsymbol{b}_i)\varphi_i \in V_h \quad [29]$$

This operator is such that for all continuous functions $v$, the restriction of $\Pi_h(v)$ to each mesh cell is a polynomial in $\mathbb{P}_k$ and $\Pi_h(v)$ takes the same values as $v$ at the Lagrange nodes. Moreover, setting $h = \max_{K_m \in \mathcal{T}_h} \operatorname{diam}(K_m)$, and defining

$$\|r\|_{L^p} = \left( \int_\Omega |r|^p \mathrm{d}\boldsymbol{x} \right)^{1/p} \quad \text{for } 1 \le p < \infty$$

the following approximation holds:

$$\|v - \Pi_h(v)\|_{L^p} + h\|\nabla(v - \Pi_h(v))\|_{L^p} \le ch^{k+1}\|v\|_{\mathcal{C}^{k+1}(\bar{\Omega})} \quad [30]$$

where $c$ is a constant that depends on the quality of the mesh. More precisely, for $K_m \in \mathcal{T}_h$, let $\rho_{K_m}$ be the diameter of the largest ball that can be inscribed into $K_m$ and let $h_{K_m}$ be the diameter of $K_m$. Then, $c$ depends on $\sigma = \max_{K_m \in \mathcal{T}_h} h_{K_m}/\rho_{K_m}$. Hence, for the mesh to have good interpolation properties, it is recommended that the cells be not too flat. Families of meshes for which $\sigma$ is bounded uniformly with respect to $h$ as $h \to 0$ are said to be *shape-regular families*.

The above example of finite-element approximation space generalizes easily to meshes composed of quadrangles or cuboids. In this case, the shape functions are piecewise polynomials of partial degree at most $k$. These spaces are usually referred to as $\mathbb{Q}_k$ approximation spaces.

## Finite Elements: Approximation

We show in this section how finite-element approximation spaces can be used to approximate some model problems exhibited in the section "Building blocks."

### Advection–Diffusion

Consider the model problem [21] supplemented with the boundary condition $(\kappa\partial_n\phi + r\phi)_{|\partial\Omega} = g$. Assume $\kappa > 0, \alpha + (1/2)\nabla \cdot \boldsymbol{u} \ge 0$, and $r \ge 0$. Define

$$a(\phi, \psi) = \int_\Omega ((\alpha\phi + \beta \cdot \nabla\phi)\psi + \kappa\nabla\phi \cdot \nabla\psi)\mathrm{d}\boldsymbol{x} + \int_{\partial\Omega} r\phi\psi\mathrm{d}s$$

Then, the weak formulation of [21] is: seek $\phi \in H$ ($H$ defined in [13]) such that for all $\psi \in H$

$$a(\phi, \psi) = \int_\Omega f\psi\mathrm{d}\boldsymbol{x} + \int_{\partial\Omega} g\psi\mathrm{d}s \quad [31]$$

Using the approximation space $V_h$ defined in [25] together with the basis defined in [27], we seek an approximate solution to the above problem in the form $\phi_h = \sum_{i=1}^N U_i\varphi_i \in V_h$. Then, a simple way of approximating [31] consists of seeking $U = (U_1, \ldots, U_N)^{\mathrm{T}} \in \mathbb{R}^N$ such that for all $1 \le i \le N$

$$a(\phi_h, \varphi_i) = \int_\Omega f\varphi_i\mathrm{d}\boldsymbol{x} + \int_{\partial\Omega} g\varphi_i\mathrm{d}s \quad [32]$$

This problem finally amounts to solving the following linear system:

$$\mathcal{A}U = F \quad [33]$$

where $\mathcal{A}_{ij} = a(\varphi_j, \varphi_i)$ and

$$F_i = \int_\Omega f\varphi_i \mathrm{d}\boldsymbol{x} + \int_{\partial\Omega} g\varphi_i \mathrm{d}s$$

The above approximation technique is usually referred to as *the Galerkin method*. The following error estimate can be proved:

$$\|\phi - \phi_h\|_{L^p} + h\|\nabla(\phi - \phi_h)\|_{L^p} \leq ch^{k+1}\|\phi\|_{C^{k+1}(\bar{\Omega})} \quad [34]$$

where, in addition to depending on the shape regularity of the mesh, the constant $c$ also depends on $\kappa$, $\alpha$, and $\beta$.

### Stokes Equations

The line of thought developed above can be used to approximate the Navier–Stokes problem [15]–[16]. Let us assume that the nonlinear term $\boldsymbol{u} \cdot \nabla\boldsymbol{u}$ is linearized in the form $\boldsymbol{v} \cdot \nabla\boldsymbol{u}$, where $\boldsymbol{v}$ is known. Let $\mathcal{T}_h$ be a mesh of $\Omega$, and assume that finite-element approximation spaces have been constructed to approximate the velocity and the pressure, say $\boldsymbol{X}_h$ and $M_h$. Assume for the sake of simplicity that $\boldsymbol{X}_h \subset \boldsymbol{X}$ and $M_h \subset M$. Assume that bases for $\boldsymbol{X}_h$ and $M_h$ are at hand, say $\{\varphi_1, \ldots, \varphi_{N_u}\}$ and $\{\psi_1, \ldots, \psi_{N_p}\}$, respectively. Set

$$a(\boldsymbol{u}, \boldsymbol{\varphi}) = \int_\Omega ((\boldsymbol{v} \cdot \nabla\boldsymbol{u}) \cdot \boldsymbol{\varphi} + \nu\nabla\boldsymbol{u} : \nabla\boldsymbol{\varphi} \mathrm{d}\boldsymbol{x}$$

and

$$b(\boldsymbol{v}, \psi) = -\int_\Omega \psi\nabla \cdot \boldsymbol{v} \mathrm{d}\boldsymbol{x}$$

Then, we seek an approximate velocity $\boldsymbol{u}_h = \sum_{i=1}^{N_u} U_i\boldsymbol{\varphi}_i$ and an approximate pressure $p_h = \sum_{k=1}^{N_p} P_k\psi_k$ such that for all $i \in \{1, \ldots, N_u\}$ and all $k \in \{1, \ldots, N_p\}$ the following holds:

$$a(\boldsymbol{u}_h, \boldsymbol{\varphi}_i) + b(\boldsymbol{\varphi}_i, p_h) = \int_\Omega \boldsymbol{f} \cdot \boldsymbol{\varphi}_i \mathrm{d}\boldsymbol{x} \quad [35]$$

$$b(\boldsymbol{u}_h, \psi_k) = 0 \quad [36]$$

Define the matrix $\mathcal{A} \in \mathbb{R}^{N_u, N_u}$ such that $\mathcal{A}_{ij} = a(\boldsymbol{\varphi}_j, \boldsymbol{\varphi}_i)$. Define the matrix $\mathcal{B} \in \mathbb{R}^{N_p, N_u}$ such that $\mathcal{B}_{ki} = b(\boldsymbol{\varphi}_i, \psi_k)$. Then, the above problem can be recast into the following partitioned linear system:

$$\begin{bmatrix} \mathcal{A} & \mathcal{B}^{\mathrm{T}} \\ \mathcal{B} & 0 \end{bmatrix} \begin{bmatrix} U \\ P \end{bmatrix} = \begin{bmatrix} F \\ 0 \end{bmatrix} \quad [37]$$

where the vector $F \in \mathbb{R}^{N_u}$ is such that $F_i = \int_\Omega \boldsymbol{f} \cdot \boldsymbol{\varphi}_i$.

An important aspect of the above approximation technique is that, for the linear system to be invertible, the matrix $\mathcal{B}^{\mathrm{T}}$ must have full row rank (i.e., $\mathcal{B}$ has full column rank). This amounts to

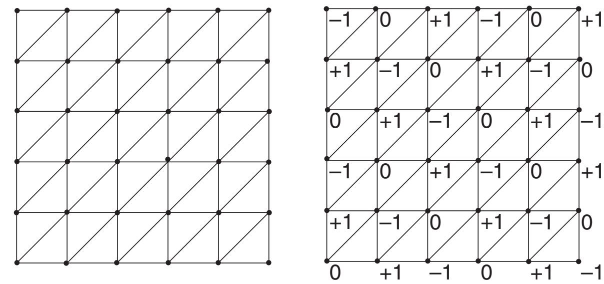



**Figure 3** The $\mathbb{P}_1/\mathbb{P}_1$ finite element: the mesh (left); one pressure spurious mode (right).

$$\exists\beta_h > 0, \quad \inf_{q_h \in M_h} \sup_{\boldsymbol{v}_h \in \boldsymbol{X}_h} \frac{\int_\Omega q_h \nabla \cdot \boldsymbol{v}_h \, \mathrm{d}\boldsymbol{x}}{\|\boldsymbol{v}_h\|_X \|q_h\|_M} \geq \beta_h \quad [38]$$

where

$$\|\boldsymbol{v}_h\|_X^2 = \int_\Omega |\nabla\boldsymbol{v}_h|^2 \mathrm{d}\boldsymbol{x}, \qquad \|q_h\|_M^2 = \int_\Omega q_h^2 \, \mathrm{d}\boldsymbol{x}$$

This nontrivial condition is called the *Ladyženskaja–Babuška–Brezzi condition* (LBB) in the literature. For instance, if $\mathbb{P}_1$ finite elements are used to approximate both the velocity and the pressure, the above condition does not hold, since there are nonzero pressure fields $q_h$ in $M_h$ such that $\int_\Omega q_h \nabla \cdot \boldsymbol{v}_h \mathrm{d}\boldsymbol{x} = 0$ for all $\boldsymbol{v}_h$ in $\boldsymbol{X}_h$. Such fields are called *spurious pressure modes*. An example is shown in Figure 3. The spurious function alternatively takes the values $-1, 0$, and $+1$ at the vertices of the mesh so that its mean value on each cell is zero.

Couples of finite-element spaces satisfying the LBB condition are numerous. For instance, assuming $k \geq 2$, using $\mathbb{P}_k$ finite elements to approximate the velocity and $\mathbb{P}_{k-1}$ finite elements to approximate the pressure is acceptable. Likewise, using $\mathbb{Q}_k$ elements for the velocity and $\mathbb{Q}_{k-1}$ elements for the pressure on meshes composed of quadrangles or cuboids is admissible.

Approximation techniques for which the pressure and the velocity degrees of freedom are not associated with the same nodes are usually called *staggered approximations*. Staggering pressure and velocity unknowns is common in solution methods for the incompressible Stokes and Navier–Stokes equations; see also the subsection "Stokes equations."

## Finite Volumes: Principles

The finite-volume method is an approximation technique whose primary goal is to approximate conservation equations, whether time dependent or

not. Given a mesh, say $\mathcal{T}_h = \{K_m\}_{1\leq m\leq N_{\text{el}}}$, and a conservation equation

$$\alpha\partial_t\phi + \nabla\cdot \boldsymbol{F}(\phi, \nabla\phi, \boldsymbol{x}, t) = f \qquad [39]$$

($\alpha = 0$ if the problem is time independent and $\alpha = 1$ otherwise), the main idea underlying every finite-volume method is to represent the approximate solution by its mean values over the mesh cells $(\phi_{K_1}, \ldots, \phi_{K_{N_{\text{el}}}})^{\text{T}} \in \mathbb{R}^{N_{\text{el}}}$ and to test the conservation equation by the characteristic functions of the mesh cells $\{1_{K_1}, \ldots, 1_{K_{N_{\text{el}}}}\}$. For each cell $K_m \in \mathcal{T}_h$, denote by $\boldsymbol{n}_{K_m}$ the outward unit normal vector and denote by $\mathcal{F}_m$ the set of the faces of $K_m$. The finite-volume approximation to [39] consists of seeking $(\phi_{K_1}, \ldots, \phi_{K_{N_{\text{el}}}})^{\text{T}} \in \mathbb{R}^{N_{\text{el}}}$ such that the function $\phi_h = \sum_{m=1}^{N_{\text{el}}} \phi_{K_m} 1_{K_1 m}$ satisfies the following: for all $1 \leq m \leq N_{\text{el}}$

$$|K_m|\alpha d_t\phi_{K_m}(t) + \sum_{\sigma\in\mathcal{F}_m} F_h^{m,\sigma}(\phi_h, \nabla_h\phi_h, t) = \int_K f\,\mathrm{d}\boldsymbol{x} \qquad [40]$$

where

$$|K_m| = \int_K \mathrm{d}\boldsymbol{x}$$

$\nabla_h\phi_h$ is an approximation of $\nabla\phi$, and $F_h^{m,\sigma}$ is an approximation of

$$\int_\sigma \boldsymbol{F}(\phi, \nabla\phi, \boldsymbol{x}, t)\cdot \boldsymbol{n}_{K_m}\mathrm{d}\boldsymbol{\sigma}$$

The precise definition of the so-called approximate flux $F_h^{m,\sigma}$ depends on the nature of the problem (e.g., elliptic, parabolic, hyperbolic, saddle point) and the desired accuracy. In general, the approximate fluxes are required to satisfy the following two important properties:

1. Conservativity: for $K_m, K_l \in \mathcal{T}_h$ such that $\sigma = K_m \cap K_l, F_h^{m,\sigma} = -F_h^{l,\sigma}$.
2. Consistency: let $\psi$ be the solution to [39], and set

$$\psi_h = \frac{1_{K_1}}{|K_1|}\int_{K_1}\psi\mathrm{d}\boldsymbol{x} + \cdots + \frac{1_{K_{N_{\text{el}}}}}{|K_{N_{\text{el}}}|}\int_{K_{N_{\text{el}}}}\psi\mathrm{d}\boldsymbol{x}$$

then

$$F_h^{m,\sigma}(\psi_h, \nabla_h\psi_h, t) \to \int_\sigma \boldsymbol{F}(\psi, \nabla\psi, \boldsymbol{x}, t)\cdot\boldsymbol{n}\,\mathrm{d}\boldsymbol{\sigma} \text{ as } h\to 0$$

The quantity

$$\left|F_h^{m,\sigma}(\psi_h, \nabla_h\psi_h, t) - \int_\sigma \boldsymbol{F}(\psi, \nabla\psi, \boldsymbol{x}, t)\cdot\boldsymbol{n}\,\mathrm{d}\boldsymbol{\sigma}\right|$$

is called the consistency error.

Note that [40] is a system of ordinary differential equations. This system is usually discretized in time by using standard time-marching techniques such as explicit Euler, Runge–Kutta, etc.

The discretization technique described above is sometimes referred to as *cell-centered finite-volume method*. Another method, called *vertex-centered finite volume* method, consists of using the characteristic functions associated with the vertices of the mesh instead of those associated with the cells.

## Finite Volumes: Examples

In this section we illustrate the ideas introduced above. Three examples are developed: the Poisson equation, the transport equation, and the Stokes equations.

### Poisson Problem

Consider the Poisson equation [11] equipped with the boundary condition $\partial_n\phi_{|\partial\Omega} = a$. To avoid technical details, assume that $\Omega = [0, 1]^d$. Let $\mathcal{K}_h$ be a mesh of $\Omega$ composed of rectangles (or cuboids in three dimensions).

The flux function is $\boldsymbol{F}(\phi, \nabla\phi, \boldsymbol{x}) = -\nabla\phi$; hence, $F_h^{m,\sigma}$ must be a consistent conservative approximation of $-\int_\sigma \boldsymbol{n}_{K_m}\cdot\nabla\phi\,\mathrm{d}\boldsymbol{\sigma}$. Let $\sigma$ be an interior face of the mesh and let $K_m, K_l$ be the two cells such that $\sigma = K_m \cap K_l$. Let $\boldsymbol{x}_{K_m}, \boldsymbol{x}_{K_l}$ be the barycenters of $K_m$ and $K_l$, respectively. Then, an admissible formula for the approximate flux is

$$F_h^{m,\sigma} = -\frac{|\sigma|}{|\boldsymbol{x}_{K_m} - \boldsymbol{x}_{K_l}|}(\phi_{K_l} - \phi_{K_m}) \qquad [41]$$

where $|\sigma| = \int_\sigma \mathrm{d}\boldsymbol{\sigma}$. The consistency error is $\mathcal{O}(h)$ in general, and is $\mathcal{O}(h^2)$ if the mesh is composed of identical cuboids. The conservativity is evident. If $\sigma$ is part of $\partial\Omega$, an admissible formula for the approximate flux is $F_h^{m,\sigma} = -\int_\sigma a\,\mathrm{d}\boldsymbol{\sigma}$. Then, upon defining $\mathcal{F}^i_{K_m} = \mathcal{F}_{K_m}\backslash\partial\Omega$ and $\mathcal{F}^\partial_{K_m} = \mathcal{F}_{K_m}\cap\partial\Omega$, the finite-volume approximation of the Poisson problem is: seek $\phi_h \in \mathbb{R}^{N_{\text{el}}}$ such that for all $1 \leq m \leq N_{\text{el}}$

$$\sum_{\sigma\in\mathcal{F}^i_{K_m}} F_h^{m,\sigma} = \int_{K_m} f\,\mathrm{d}\boldsymbol{x} + \sum_{\sigma\in\mathcal{F}^\partial_{K_m}}\int_\sigma a\,\mathrm{d}\boldsymbol{\sigma} \qquad [42]$$

### Transport Equation

Consider the transport equation

$$\partial_t\phi + \nabla\cdot(\boldsymbol{u}\phi) = f \qquad [43]$$

$$\phi_{|t=0} = \phi_0, \quad \phi_{|\partial\Omega^-} = a \qquad [44]$$

where $\boldsymbol{u}(\boldsymbol{x}, t)$ is a given field in $\mathcal{C}^1(\bar{\Omega}\times[0, T])$. Let $\mathcal{T}_h$ be a mesh of $\Omega$. For the sake of simplicity, let us use the explicit Euler time-stepping to approximate [40].

Let $N$ be positive integer, set $\Delta t = T/N$, set $t^n = n\Delta t$ for $0 \le n \le N$, and partition $[0, T]$ as follows:

$$[0, T] = \bigcup_{n=0}^{N-1} [t^n, t^{n+1}]$$

Denote by $\phi_h^n \in \mathbb{R}^{N_{\mathrm{el}}}$ the finite-volume approximation of $\phi_h(t^n)$. Then, [40] is approximated as follows:

$$\frac{|K_m|}{\Delta t}(\phi_{K_m}^{n+1} - \phi_{K_m}^n) + \sum_{\sigma \in \mathcal{F}_m} F_h^{m,\sigma}(\phi_h, \nabla_h \phi_h, t^n) = \int_K f(\boldsymbol{x}, t^n)\, \mathrm{d}\boldsymbol{x} \qquad [45]$$

where $\phi_{K_m}^0 = \int_{K_m} \phi_0\, \mathrm{d}\boldsymbol{x}$. The approximate flux $F_h^{m,\sigma}$ must be a consistent conservative approximation of $\int_\sigma (\boldsymbol{u} \cdot \boldsymbol{n}_{K_m})\phi\, \mathrm{d}\boldsymbol{\sigma}$. Let $\sigma$ be a face of the mesh and let $K_m, K_l$ be the two cells such that $\sigma = K_m \cap K_l$ (note that if $\sigma$ is on $\partial\Omega, \sigma$ belongs to one cell only and we set $K_m = K_l$). If $\sigma$ is on $\partial\Omega^-$, set

$$F_h^{m,\sigma} = \int_\sigma (\boldsymbol{u} \cdot \boldsymbol{n}_{K_m}) a\, \mathrm{d}\boldsymbol{\sigma} \qquad [46]$$

If $\sigma$ is not on $\partial\Omega^-$, set $u_{m,\sigma}^n = \int_\sigma (\boldsymbol{u} \cdot \boldsymbol{n}_{K_m})\mathrm{d}\boldsymbol{\sigma}$ and define

$$F_h^{m,\sigma} = \begin{cases} \phi_{K_m}^n u_{m,\sigma}^n & \text{if } u_{m,\sigma}^n \ge 0 \\ \phi_{K_l}^n u_{m,\sigma}^n & \text{if } u_{m,\sigma}^n < 0 \end{cases} \qquad [47]$$

The above choice for the approximate flux is usually called the *upwind flux*. It is consistent with the analysis that has been done for [22], that is, information flows along the characteristic lines of the field $\boldsymbol{u}$; see [24]. In other words, the updating of $\phi_{k_m}^{n+1}$ must be done by using the approximate values $\phi_h^n$ coming from the cells that are upstream the flow field.

An important feature of the above approximation technique is that it is $L^\infty$-stable, in the sense that

$$\max_{0 \le n \le N, 1 \le m \le N_{\mathrm{el}}} |\phi_{K_m}^n| \le c(u_0, f)$$

if the two mesh parameters $\Delta t$ and $h$ satisfy the so-called *Courant–Friedrichs–Levy* (CFL) condition $\|\boldsymbol{u}\|_{L^\infty} \Delta t / h \le c(\sigma)$, where $c(\sigma)$ is a constant that depends on the mesh regularity parameter $\sigma = \max_{K_m \in \mathcal{T}_h} h_{K_m}/\rho_{K_m}$. In one dimension, $c(\sigma) = 1$.

### Stokes Equations

To finish this short review of finite-volume methods, we turn our attention to the Stokes problem (15)–(16) equipped with the homogeneous Dirichlet boundary condition $\boldsymbol{u}_{|\partial\Omega} = \boldsymbol{0}$.

Let $\mathcal{T}_h$ be a mesh of $\Omega$ composed of triangles (or tetrahedrons). All the angles in the triangulation are assumed to be acute so that, for all $K \in \mathcal{T}_h$, the intersection of the orthogonal bisectors of the sides of $K$, say $\boldsymbol{x}_K$, is in $K$. We propose a finite-volume approximation for the velocity and a finite-element approximation for the pressure. Let $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_d\}$ be a Cartesian basis for $\mathbb{R}^d$. Set $\mathbf{1}_{K_m}^k = 1_{K_m}\boldsymbol{e}_k$ for all $1 \le m \le N_{\mathrm{el}}$ and $1 \le k \le d$; then define

$$\boldsymbol{X}_h = \mathrm{span}\left\{\mathbf{1}_{K_1}^1, \ldots, \mathbf{1}_{K_1}^d, \ldots, \mathbf{1}_{K_{N_{\mathrm{el}}}}^1, \ldots, \mathbf{1}_{K_{N_{\mathrm{el}}}}^d\right\}$$

Let $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{N_v}\}$ be the vertices of the mesh, and let $\{\varphi_1, \ldots, \varphi_{N_v}\}$ be the associated piecewise linear global shape functions. Then, set (see the section "Finite elements: interpolation")

$$N_h = \mathrm{span}\{\varphi_1, \ldots, \varphi_{N_v}\}$$
$$M_h = \{q \in N_h; \int_\Omega q\, \mathrm{d}\boldsymbol{x} = 0\}$$

The approximate problem consists of seeking $(\boldsymbol{u}_{K_1}, \ldots, \boldsymbol{u}_{K_{N_{\mathrm{el}}}}) \in \mathbb{R}^{dN_{\mathrm{el}}}$ and $p_h \in M_h$ such that for all $1 \le m \le N_{\mathrm{el}}, 1 \le k \le d$, and all $1 \le i \le N_v$,

$$\sum_{\sigma \in \mathcal{F}_m} \mathbf{1}_{K_m}^k \cdot \boldsymbol{F}_h^{m,\sigma} + c\left(\mathbf{1}_{K_m}^k, p_h\right) = \int_{K_m} \mathbf{1}_{K_m}^k \cdot \boldsymbol{f} \mathrm{d}\boldsymbol{x} \qquad [48]$$

$$c(\boldsymbol{u}_{K_m}, \varphi_i) = 0 \qquad [49]$$

where

$$c(\boldsymbol{v}_{K_m}, p_h) = \int_{K_m} \boldsymbol{v}_{K_m} \cdot \nabla p_h \mathrm{d}\boldsymbol{x}$$

Moreover,

$$\boldsymbol{F}_h^{m,\sigma} = \begin{cases} \dfrac{\nu|\sigma|}{|\boldsymbol{x}_m - \boldsymbol{x}_l|}(\boldsymbol{u}_{K_m} - \boldsymbol{u}_{K_l}) & \text{if } \sigma = K_m \cap K_l \\ \dfrac{\nu|\sigma|}{d(\boldsymbol{x}_m, \sigma)} \boldsymbol{u}_{K_m} & \text{if } \sigma = K_m \cap \partial\Omega \end{cases}$$

where $d(\boldsymbol{x}_{K_m}, \sigma)$ is the Euclidean distance between $\boldsymbol{x}_{K_m}$ and $\sigma$. This formulation yields a linear system with the same structure as in [37]. Note in particular that

$$\sup_{\boldsymbol{v}_h \in \boldsymbol{X}_h} \frac{c(\boldsymbol{v}_h, p_h)}{\|\boldsymbol{v}_h\|_{L^\infty}} = \|\nabla p_h\|_{L^1} \qquad [50]$$

Since the mean value of $p_h$ is zero, $\|\nabla p_h\|_{L^1}$ is a norm on $M_h$. As a result, an inequality similar to [38] holds. This inequality is a key step to proving that the linear system is wellposed and the approximate solution converges to the exact solution of (15)–(16).

## Projection Methods for Navier–Stokes

In this section we focus on the time approximation of the Navier–Stokes problem:

$$\partial_t \boldsymbol{u} - \nu \Delta \boldsymbol{u} + \boldsymbol{u} \cdot \nabla \boldsymbol{u} + \nabla p = \boldsymbol{f} \qquad [51a]$$

$$\nabla \cdot \boldsymbol{u} = 0 \qquad [51b]$$

$$\boldsymbol{u}_{|\partial\Omega} = 0 \qquad [51c]$$

$$\boldsymbol{u}_{|t=0} = \boldsymbol{u}_0 \qquad [51d]$$

where $\boldsymbol{f}$ is a body force and $\boldsymbol{u}_0$ is a solenoidal velocity field. There are numerous ways to discretize this problem in time, but, undoubtedly, one of the most popular strategies is to use projection methods, sometimes also referred to as *Chorin–Temam methods*.

A projection method is a fractional-step time-marching technique. It is a predictor–corrector strategy aiming at uncoupling viscous diffusion and incompressibility effects. One time step is composed of three substeps: in the first substep, the pressure is made explicit and a provisional velocity field is computed using the momentum equation; in the second substep, the provisional velocity field is projected onto the space of incompressible (solenoidal) vector fields; in the third substep, the pressure is updated.

Let $q > 0$ be an integer and approximate the time derivative of $\boldsymbol{u}$ using a backward difference formula of order $q$. To this end, introduce a positive integer $N$, set $\Delta t = T/N$, set $t^n = n\Delta t$ for $0 \leq n \leq N$, and consider a partitioning of the time interval in the form

$$[0, T] = \bigcup_{n=0}^{N-1} [t^n, t^{n+1}]$$

For all sequences $\boldsymbol{v}_{\Delta t} = (\boldsymbol{v}^0, \boldsymbol{v}^1, \ldots, \boldsymbol{v}^N)$, set

$$D^{(q)}\boldsymbol{v}^{n+1} = \beta_q \boldsymbol{v}^{n+1} - \sum_{j=0}^{q-1} \beta_j \boldsymbol{v}^{n-j} \qquad [52]$$

where $q - 1 \leq n \leq N - 1$. The coefficients $\beta_j$ are such that

$$\frac{1}{\Delta t}(\beta_q \boldsymbol{u}(t^{n+1}) - \sum_{j=0}^{q-1} \beta_j \boldsymbol{u}(t^{n-j}))$$

is a $q$th-order backward difference formula approximating $\partial_t \boldsymbol{u}(t^{n+1})$. For instance,

$$D^{(1)}\boldsymbol{v}^{n+1} = \boldsymbol{v}^{n+1} - \boldsymbol{v}^n$$

$$D^{(2)}\boldsymbol{v}^{n+1} = \tfrac{3}{2}\boldsymbol{v}^{n+1} - 2\boldsymbol{v}^n + \tfrac{1}{2}\boldsymbol{v}^{n-1}$$

Furthermore, for all sequences $\phi_{\Delta t} = (\phi^0, \phi^1, \ldots, \phi^N)$, define

$$\phi^{\star, n+1} = \sum_{j=0}^{q-1} \gamma_j \phi^{n-j} \qquad [53]$$

so that $\sum_{j=0}^{q-1} \gamma_j p(t^{n-j})$ is a $(q-1)$th-order extrapolation of $p(t^{n+1})$. For instance, $p^{\star,n+1} = 0$ for $q = 1, p^{\star,n+1} = p^n$ for $q = 2$, and $p^{\star,n+1} = 2p^n - p^{n-1}$ for $q = 3$. Finally, denote by $(\boldsymbol{u} \cdot \nabla \boldsymbol{u})^{\star,n+1}$ a $q$th-order extrapolation of $(\boldsymbol{u} \cdot \nabla \boldsymbol{u})(t^{n+1})$. For instance,

$$(\boldsymbol{u} \cdot \nabla \boldsymbol{u})^{\star,n+1} = \begin{cases} \boldsymbol{u}^n \cdot \nabla \boldsymbol{u}^n & \text{for } q = 1 \\ 2\boldsymbol{u}^n \cdot \nabla \boldsymbol{u}^n - \boldsymbol{u}^{n-1} \cdot \nabla \boldsymbol{u}^{n-1} & \text{if } q = 2 \end{cases}$$

A general projection algorithm is as follows. Set $\tilde{\boldsymbol{u}}^0 = \boldsymbol{u}_0$ and $\phi^l = 0$ for $0 \leq l \leq q - 1$. If $q > 1$; assume that $\tilde{\boldsymbol{u}}^1, \ldots, \tilde{\boldsymbol{u}}^{q-1}, p^{\star,q}$ and $(\boldsymbol{u} \cdot \nabla \boldsymbol{u})^{\star,q}$ have been initialized properly. For $n \geq q - 1$, seek $\tilde{\boldsymbol{u}}^{n+1}$ such that $\tilde{\boldsymbol{u}}^{n+1}_{|\partial\Omega} = 0$ and

$$\frac{D^{(q)}}{\Delta t}\tilde{\boldsymbol{u}}^{n+1} - \nu \Delta \tilde{\boldsymbol{u}}^{n+1} + \nabla\left(p^{\star,n+1} + \sum_{j=0}^{q-1} \frac{\beta_j}{\Delta t} \phi^{n-j}\right) = \boldsymbol{S}^{n+1} \qquad [54a]$$

where $\boldsymbol{S}^{n+1} = \boldsymbol{f}(t^{n+1}) - (\boldsymbol{u} \cdot \nabla \boldsymbol{u})^{\star,n+1}$. Then solve

$$\Delta \phi^{n+1} = \nabla \cdot \tilde{\boldsymbol{u}}^{n+1}, \qquad \partial_n \phi^{n+1}_{|\partial\Omega} = 0 \qquad [54b]$$

Finally, update the pressure as follows:

$$p^{n+1} = \frac{\beta_q}{\Delta t} \phi^{n+1} + p^{\star,n+1} - \nu \nabla \cdot \tilde{\boldsymbol{u}}^{n+1} \qquad [54c]$$

The algorithm [54a–c] is known in the literature as the rotational form of the pressure-correction method. Upon denoting $u_{\Delta t} = (\boldsymbol{u}(t^0), \ldots, \boldsymbol{u}(t^N))$ and $p_{\Delta t} = (p(t^0), \ldots, p(t^N))$, the above algorithm has been proved to yield the following error estimates:

$$\|u_{\Delta t} - \tilde{\boldsymbol{u}}_{\Delta t}\|_{\ell^2(L^2)} \leq c\Delta t^2$$

$$\|\nabla(u_{\Delta t} - \tilde{\boldsymbol{u}}_{\Delta t})\|_{\ell^2(L^2)} + \|p_{\Delta t} - p_{\Delta t}\|_{\ell^2(L^2)} \leq c\Delta t^{3/2}$$

where $\|\phi_{\Delta t}\|^2_{\ell^2(L^2)} = \Delta t \sum_{n=0}^{N} \int_\Omega |\phi^n|^2 \, d\boldsymbol{x}$.

A simple strategy to initialize the algorithm consists of using $D^{(1)}\boldsymbol{u}^1$ at the first step in [54a]; then using $D^{(2)}\boldsymbol{u}^2$ at the second step, and proceeding likewise until $\tilde{\boldsymbol{u}}^1, \ldots, \tilde{\boldsymbol{u}}^{q-1}$ have all been computed.

At the present time, projection methods count among the few methods that are capable of solving the time-dependent incompressible Navier–Stokes equations in three dimensions on fine meshes within reasonable

computation times. The reason for this success is that the unsplit strategy, which consists of solving

$$\frac{D^{(q)}}{\Delta t}\boldsymbol{u}^{n+1} - \nu\Delta\boldsymbol{u}^{n+1} + \nabla p^{n+1} = \boldsymbol{S}^{n+1} \qquad [55a]$$

$$\nabla \cdot \boldsymbol{u}^{n+1} = 0, \qquad \boldsymbol{u}^{n+1}_{|\partial\Omega} = 0 \qquad [55b]$$

yields a linear system similar to [37], which usually takes far more time to solve than sequentially solving [54a] and [54b]. It is commonly reported in the literature that the ratio of the CPU time for solving [55a]–[55b] to that for solving [54a–c] ranges between 10 to 30.

*See also:* Compressible Flows: Mathematical Theory; Computational Methods in General Relativity: The Theory; Geophysical Dynamics; Image Processing: Mathematics; Incompressible Euler Equations: Mathematical Theory; Interfaces and Multicomponent Fluids; Magnetohydrodynamics; Newtonian Fluids and Thermohydraulics; Non-Newtonian Fluids; Partial Differential Equations: Some Examples; Variational Methods in Turbulence.

## Further Reading


Doering CR and Gibbon JD (1995) *Applied Analysis of the Navier–Stokes Equations*, Cambridge Texts in Applied Mathematics. Cambridge: Cambridge University Press.

Ern A and Guermond J-L (2004) *Theory and Practice of Finite Elements*. Springer Series in Applied Mathematical Sciences, vol. 159. New York: Springer-Verlag.

Eymard R, Gallouët T, and Herbin R (2000) Finite volume methods. In: Ciarlet PG and Lions JL (eds.) *Handbook of Numerical Analysis*, vol. VII , pp. 713–1020. Amsterdam: North-Holland.

Karniadakis GE and Sherwin SJ (1999) *Spectral/hp Element Methods for CFD*, Numerical Mathematics and Scientific Computation. New York: Oxford University Press.

Rappaz M, Bellet M, and Deville M (2003) *Numerical Modeling in Material Science and Engineering*. Springer Series in Computational Mathematics, vol. 32. Berlin: Springer Verlag.

Temam R (1984) *Navier–Stokes Equations. Theory and Numerical Analysis*. Studies in Mathematics and its Applications, vol. 2. Amsterdam: North-Holland.

Toro EF (1997) *Riemann Solvers and Numerical Methods for Fluid Dynamics. A Practical Introduction*. Berlin: Springer.

Wesseling P (2001) *Principles of Computational Fluid Dynamics*. Springer Series in Computational Mathematics, vol. 29. Berlin: Springer.


# Fourier Law


**F Bonetto**, Georgia Institute of Technology, Atlanta, GA, USA

**L Rey-Bellet**, University of Massachusetts, Amherst, MA, USA




## Introduction

In the famous 1822 treatise by Jean Baptiste Joseph Fourier, *Théorie analytique de la chaleur*, the *Discours préliminaire* opens with: "Primary causes are unknown to us; but are subject to simple and constant laws, which may be discovered by observation, the study of them being the subject of natural philosophy. Heat, like gravity, penetrates every substance of the universe, its rays occupy all parts of space. The object of our work is to set forth the mathematical laws which this element obeys. The theory of heat will hereafter form one of the most important branches of general physics." After a brief discussion of rational mechanics, he continues with the sentence: "But whatever may be the range of mechanical theories, they do not apply to the effects of heat. These make up a special order of phenomena, which cannot be explained by the principles of motion and equilibria." Fourier goes on with a thorough description of the phenomenology of heat transport and the derivation of the partial differential equation describing heat transport: the heat equation. A large part of the treatise is then devoted to solving the heat equation for various geometries and boundary conditions. Fourier's treatise marks the birth of Fourier analysis. After Boltzmann, Gibbs, and Maxwell and the invention of statistical mechanics in the decades after Fourier's work, we believe that Fourier was wrong and that, in principle, heat transport can and should be explained "by the principles of motion and equilibria," that is, within the formalism of statistical mechanics. But well over a century after the foundations of statistical mechanics were laid down, we still lack a mathematically reasonable derivation of Fourier's law from first principles. Fourier's law describes the macroscopic transport properties of heat, that is, energy, in nonequilibrium systems. Similar laws are valid for the transport of other locally conserved quantities, for example, charge, particle density, momentum, etc. We will not discuss these laws here, except to point out that in none of these cases macroscopic transport laws have been derived from microscopic dynamics. As Peierls once put it: "It seems there is no problem in modern physics for which there are on record as many false starts, and as many theories which overlook some essential feature, as in the problem of the thermal conductivity of [electrically] non-conducting crystals."

## Macroscopic Law

Consider a macroscopic system characterized at some initial time, say $t = 0$, by a nonuniform

temperature profile $T_0(\boldsymbol{r})$. This temperature profile will generate a heat, that is, energy current $\boldsymbol{J}(\boldsymbol{r})$. Due to energy conservation and basic thermodynamics:

$$c_v(T)\frac{\partial}{\partial t}T(\boldsymbol{r},t) = -\nabla\cdot\boldsymbol{J} \qquad [1]$$

where $c_v(T)$ is the specific heat per unit volume. On the other hand, we know that if the temperature profile is uniform, that is, if $T_0(\boldsymbol{r}) \equiv T_0$, there is no current in the system. It is then natural to assume that, for small temperature gradients, the current is given by

$$\boldsymbol{J}(\boldsymbol{r}) = -\kappa(T(\boldsymbol{r}))\nabla T(\boldsymbol{r}) \qquad [2]$$

where $\kappa(T)$ is the conductivity. Here we have assumed that there is no mass flow or other mode of energy transport besides heat conduction (we also ignore, for simplicity, any variations in density or pressure). Equation [2] is normally called as Fourier's law. Putting together eqns [1] and [2], we get the heat equation:

$$c_v(T)\frac{\partial}{\partial t}T(\boldsymbol{r},t) = \nabla\cdot[\kappa(T)\nabla T] \qquad [3]$$

This equation must be completed with suitable boundary conditions. Let us consider two distinct situations in which the heat equation is observed to hold experimentally with high precision:

1. An isolated macroscopic system, for example, a fluid or solid in a domain $\Lambda$ surrounded by effectively adiabatic walls. In this case, eqn [3] is to be solved subject to the initial condition $T(\boldsymbol{r},0) = T_0(\boldsymbol{r})$ and no heat flux across the boundary of $\Lambda$ (denoted by $\partial\Lambda$), that is, $\boldsymbol{n}(\boldsymbol{r})\cdot\nabla T(\boldsymbol{r}) = 0$ if $\boldsymbol{r}\in\partial\Lambda$ with $\boldsymbol{n}$ the normal vector to $\partial\Lambda$ at $\boldsymbol{r}$. As $t\to\infty$, the system reaches a stationary state characterized by a uniform temperature $\bar{T}$ determined by the constancy of the total energy.
2. A system in contact with heat reservoirs. Each reservoir $\alpha$ fixes the temperature of some portion $(\partial\Lambda)_\alpha$ of the boundary $\partial\Lambda$. The rest of the boundary is insulated. When the system reaches a stationary state (again assuming no matter flow), its temperature will be given by the solution of eqn [3] with the left-hand side set equal to zero,

   $$\nabla\cdot\tilde{J}(\boldsymbol{r}) = \nabla\cdot(\kappa\nabla\tilde{T}(\boldsymbol{r})) = 0 \qquad [4]$$

   subject to the boundary condition $\tilde{T}(\boldsymbol{r}) = T_\alpha$ for $\boldsymbol{r}\in(\partial\Lambda)_\alpha$ and no flux across the rest of the boundary.

The simplest geometry for a conducting system is that of a cylindrical slab of height $h$ and cross-sectional area $A$. It can be either a cylindrical container filled with a fluid or a piece of crystalline solid. In both cases, one keeps the lateral surface of the cylinder insulated. If the top and the bottom of the cylinder are also insulated we are in case (1). If one keeps the top and the bottom in contact with thermostats at temperatures $T_\mathrm{h}$ and $T_\mathrm{b}$, respectively, this is (for a fluid) the usual setup for a Benard experiment. To avoid convection, one has to make $T_\mathrm{h} > T_\mathrm{b}$ or keep $|T_\mathrm{h} - T_\mathrm{b}|$ small. Assuming uniformity in the direction perpendicular to the vertical $x$-axis one has, in the stationary state, a temperature profile $\tilde{T}(x)$ with $\tilde{T}(0) = T_\mathrm{b}, \tilde{T}(h) = T_\mathrm{h}$ and $\kappa(\tilde{T})\mathrm{d}\tilde{T}/\mathrm{d}x = \text{const.}$ for $x\in(0,h)$.

In deriving the heat equation, we have implicitly assumed that the system is described fully by specifying its temperature $T(\boldsymbol{r},t)$ everywhere in $\Lambda$. What this means on the microscopic level is that we imagine the system to be in local thermal equilibrium (LTE). Heuristically, we might think of the system as being divided up (mentally) into many little cubes, each large enough to contain very many atoms yet small enough on the macroscopic scale to be accurately described, at a specified time $t$, as a system in equilibrium at temperature $T(\boldsymbol{r}_i,t)$, where $\boldsymbol{r}_i$ is the center of the $i$th cube. For slow variation in space and time, we can then use a continuous description $T(\boldsymbol{r},t)$. The theory of the heat equation is very developed and, together with its generalizations, plays a central role in modern analysis. In particular, one can consider more general boundary conditions. Here we are interested in the derivation of eqn [2] from first principles. This clearly presupposes, as a first fundamental step, a precise definition of the concept of LTE and its justification within the law of mechanics.

## Empirical Argument

A theory of heat conduction has as a goal the computation of the conductivity $\kappa(T)$ for realistic models, or, at the very least, the derivation of behavior of $\kappa(T)$ as a function of $T$. The early analysis was based on "kinetic theory." Its application to heat conduction goes back to the works of Clausius, Maxwell, and Boltzmann, who obtained a theoretical expression for the heat conductivity of gases, $\kappa\sim\sqrt{T}$, independent of the gas density. This agrees with experiment (when the density is not too high) and was a major early achievement of the atomic theory of matter.

### Heat Conduction in Gases

Clausius and Maxwell used the concept of a "mean free path" $\lambda$: the average distance a particle (atom or

molecule) travels between collisions in a gas with particle density $\rho$. Straightforward analysis gives $\lambda \sim 1/\rho\pi\sigma^2$, where $\sigma$ an "effective" hard-core diameter of a particle. They considered a gas with temperature gradient in the $x$-direction and assumed that the gas is (approximately) in local equilibrium with density $\rho$ and temperature $T(x)$. Between collisions, a particle moves a distance $\lambda$ carrying a kinetic energy proportional to $T(x)$ from $x$ to $x + \lambda/\sqrt{3}$, while in the opposite direction the amount carried is proportional to $T(x + \lambda\sqrt{3})$. Taking into account the fact that the speed is proportional to $\sqrt{T}$ the amount of energy $J$ transported per unit area and time across a plane perpendicular to the $x$-axis is approximately

$$J \sim \rho\sqrt{T}\Big[T(x) - T(x + \lambda\sqrt{3})\Big] \sim -\sigma^{-2}\sqrt{T}\frac{\mathrm{d}T}{\mathrm{d}x} \qquad [5]$$

and so $\kappa \sim \sqrt{T}$ independent of $\rho$, in agreement with experiment. It was clear to the founding fathers that starting with a local equilibrium situation the process described above will produce, as time goes on, a deviation from LTE. They reasoned, however, that this deviation from local equilibrium will be small when $(\lambda/T)\mathrm{d}T/\mathrm{d}x \ll 1$, the regime in which Fourier's law is expected to hold, and the above calculation should yield, up to some factor of order unity, the right heat conductivity. To have a more precise theory, one can describe the state of the gas through the probability distribution $f(\boldsymbol{r}, \boldsymbol{p}, t)$ of finding a particle in the volume element $\mathrm{d}\boldsymbol{r}\,\mathrm{d}\boldsymbol{p}$ around the phase space point $(\boldsymbol{r}, \boldsymbol{p})$. Here LTE means that

$$f(\boldsymbol{r}, \boldsymbol{p}, t) \simeq \exp\left(-\frac{p^2}{2mkT(\boldsymbol{r})}\right)$$

where $m$ is the mass of the particles. If one computes the heat flux at a point $\boldsymbol{r}$ by averaging the microscopic energy current at $\boldsymbol{r}$, $j = \rho\boldsymbol{v}(1/2mv^2)$, over $f(\boldsymbol{r}, \boldsymbol{p}, t)$ then it is only the deviation from local equilibrium which makes a contribution. The result however is essentially the same as eqn [5]. This was shown by Boltzmann, who derived an accurate formula for $\kappa$ in gases by using the Boltzmann equation. If one takes $\kappa$ from experiment, the above analysis yields a value for $\sigma$, the effective size of an atom or molecule, which turns out to be close to other determinations of the characteristic size of an atom. This gave an evidence for the reality of atoms and the molecular theory of heat.

### Heat Conduction in Insulating Crystals

In (electrically) conducting solids, heat is mainly transported by the conduction electron. In this case, one can adapt the theory discussed in the previous section. In (electrically) insulating solids, on the other hand, heat is transmitted through the vibrations of the lattice. In order to use the concepts of kinetic theory, it is useful to picture a solid as a gas of phonons which can store and transmit heat. A perfectly harmonic crystal, due to the fact that phonons do not interact, has an infinite thermal conductivity: in the language of kinetic theory, the mean free path $\lambda$ is infinite. In a real crystal, the anharmonic forces produce interactions between the phonons and therefore a finite mean free path. Another source of finite thermal conductivity may be the lattice imperfections and impurities which scatter the phonons. Debye devised a kind of kinetic theory for phonons in order to describe thermal conductivity. One assumes that a small gradient of temperature is imposed and that the collisions between phonons maintain local equilibrium. An elementary argument gives a thermal conductivity analogous to eqn [5] obtained in the last subsection for gases (remembering, however, that the density of phonons is itself a function of $T$)

$$\kappa \sim c_v c^2 \tau \qquad [6]$$

where, with respect to eqn [5], $\rho$ has been replaced by $c_v$, the specific heat of phonons, $\sqrt{T}$ by $c$, the (mean) velocity of the phonons, and $\lambda$ by $c\tau$, where $\tau$ is the effective mean free time between phonon collisions. The thermal conductivity depends on the temperature via $\tau$, and a more refined theory is needed to account for this dependence. This was done by Peierls via a Boltzmann equation for the phonons. In collisions among phonons, the momentum of phonons is conserved only modulo a vector of the reciprocal lattice. One calls "normal processes" those where the phonon momentum is conserved and "Umklap processes" those where the initial and final momenta differ by a nonzero reciprocal lattice vector. Peierls' theory may be summarized (very roughly) as follows: in the absence of Umklap processes, the mean free path, and thus the thermal conductivity of an insulating solid, is infinite. A success of Peierls' theory is to describe correctly the temperature dependence of the thermal conductivity. Furthermore, on the basis of this theory, one does not expect a finite thermal conductivity in one-dimensional monoatomic lattices with pair interactions. This seems so far to be a correct prediction, at least in the numerous numerical results performed on various models.

## Statistical Mechanics Paradigm: Rigorous Analysis

In a rigorous approach to the above arguments, we have to first formulate precisely the problem on a

mathematical level. It is natural to adapt the standard formalism of statistical mechanics to our situation. To this end, we assume that our system is described by the positions $Q$ and momenta $P$ of a (very large) number of particles, $N$, with $Q = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N) \in \Lambda^N, \Lambda \subset \mathbb{R}^d$, and $P = (\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N) \in \mathbb{R}^{dN}$. The dynamics (in the bulk) is given by a Hamiltonian function $H(Q, P)$. A state of the system is a probability measure $\mu(P, Q)$ on phase space. As usual in statistical mechanics, the value of an observable $f(P, Q)$ will be given by the expected value of $f$ with respect to the measure $\mu$. In the case of a fluid contained in a region $\Lambda$, we can assume that the Hamiltonian has the form

$$H(P, Q) = \sum_{i=1}^{N} \left[ \frac{\boldsymbol{p}_i^2}{2m} + \sum_{j \neq i} \phi(\boldsymbol{q}_j - \boldsymbol{q}_i) + u(\boldsymbol{q}_i) \right] = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i^2}{2m} + \mathcal{V}(Q) \qquad [7]$$

where $\phi(\boldsymbol{q})$ is some short-range interparticle potential and $u(\boldsymbol{q}_i)$ an external potential (*e.g.*, the interaction of the particle with fixed obstacles such as a conduction electron interacting with the fixed crystalline ions). If we want to describe the case in which the temperature at the boundary is kept different in different regions $\partial\Lambda_\alpha$, we have to properly define the dynamics at the boundary of the system. A possibility is to use "Maxwell boundary conditions": when a particle hits the wall in $\partial\Lambda_\alpha$, it gets reflected and re-emerges with a distribution of velocities

$$f_\alpha(\mathrm{d}\boldsymbol{v}) = \frac{m^2}{2\pi(kT_\alpha)^2} |v_x| \exp\left[ -\frac{m\boldsymbol{v}^2}{2kT_\alpha} \right] \mathrm{d}\boldsymbol{v} \qquad [8]$$

Several other ways to impose boundary conditions have been considered in the literature. The notion of LTE can be made precise here in the so-called hydrodynamic scaling limit (HSL), where the ratio of microscopic to macroscopic scales goes to zero. The macroscopic coordinates $\boldsymbol{r}$ and $t$ are related to the microscopic ones $\boldsymbol{q}$ and $\tau$, by $\boldsymbol{r} = \epsilon\boldsymbol{q}$ and $t = \epsilon^\alpha \tau$, that is, if $\Lambda$ is a cube of macroscopic sides $l$, then its sides, now measured in microscopic length units, are of length $L = \epsilon^{-1} l$. We then suppose that at $t = 0$ our system of $N = \rho L^d$ particles is described by an equilibrium Gibbs measure with a temperature $T(\boldsymbol{r}) = T(\epsilon\boldsymbol{q})$: roughly speaking, the phase-space ensemble density has the form

$$\mu_0(P, Q) \sim \exp\left\{ -\sum_{i=1}^{N} \beta_0(\epsilon\boldsymbol{q}_i) \times \left[ \frac{\boldsymbol{p}_i^2}{2m} + \sum_{j \neq i} \phi(\boldsymbol{q}_j - \boldsymbol{q}_i) + u(\boldsymbol{q}_i) \right] \right\} \qquad [9]$$

where $\beta_0^{-1}(\boldsymbol{r}) = T_0(\boldsymbol{r})$. In the limit $\epsilon \to 0, \rho$ fixed, the system at $t = 0$ will be macroscopically in LTE with a local temperature $T_0(\boldsymbol{r})$ (as already noted, here we suppress the variation in the particle density $n(\boldsymbol{r})$). We are interested in the behavior of a macroscopic system, for which $\epsilon \ll 1$, at macroscopic times $t \geq 0$, corresponding to microscopic times $\tau = \epsilon^{-\alpha} t, \alpha = 2$ for heat conduction or other diffusive behavior. The implicit assumption then made in the macroscopic description given earlier is that, since the variations in $T_0(\boldsymbol{r})$ are of order $\epsilon$ on a microscopic scale, then for $\epsilon \ll 1$, the system will, also at time $t$, be in a state very close to LTE, with a temperature $T(\boldsymbol{r}, t)$ that evolves in time according to Fourier's law, eqn [1]. From a mathematical point of view, the difficult problem is to prove that the system stays in LTE for $t > 0$ when the dynamics are given by a Hamiltonian time evolution. This requires proving that the macroscopic system has some very strong ergodic properties, for example, that the only time-invariant measures locally absolutely continuous with respect to the Lebesgue measure are, for infinitely extended spatially uniform systems, of the Gibbs type. This has only been proved so far for systems evolving via stochastic dynamics (e.g., interacting Brownian particles or lattice gases). For such stochastic systems, one can sometimes prove the hydrodynamical limit and derive macroscopic transport equations for the particle or energy density and thus verify the validity of Fourier law. Another possibility, as we already saw, is to use the Boltzmann equation. Using ideas of hydrodynamical space and time scaling described earlier, it is possible to derive a controlled expansion for the solution of the stationary Boltzmann equation describing the steady state of a gas coupled to temperature reservoirs at the top and bottom. One then shows that for $\epsilon \ll 1$, $\epsilon$ being now the ratio $\lambda/L$, the Boltzmann equation for $f$ in the slab has a time-independent solution which is close to a local Maxwellian, corresponding to LTE (apart from boundary layer terms) with a local temperature and density given by the solution of the Navier–Stokes equations which incorporates Fourier's law as expressed in eqn [2]. The main mathematical problem is in controlling the remainder in an asymptotic expansion of $f$ in power of $\epsilon$. This requires that the macroscopic temperature gradient, that is, $|T_1 - T_2|/h$, where $h = \epsilon L$ is the thickness of the slab on the macroscopic scale, be small. Even if this apparently technical problem could be overcome, we would still be left with the question of justifying the Boltzmann equation for such steady states and, of course, it would not tell us anything

about dense fluids or crystals. In fact, the Boltzmann equation itself is really closer to a macroscopic than to a microscopic description. It is obtained in a well-defined kinetic scaling limit in which, in addition to rescaling space and time, the particle density goes to zero, that is, $\lambda \gg \sigma$.

A simplified model of a crystal is characterized by the fact that all atoms oscillate around given equilibrium positions. The equilibrium positions can be thought of as the points of a regular lattice in $\mathbb{R}^d$, say $\mathbb{Z}^d$. Although $d = 3$ is the physical situation, one can also be interested in the case $d = 1, 2$. In this situation, $\Lambda \subset \mathbb{Z}^d$ with cardinality $N$, and each atom is identified by its position $\boldsymbol{x}_i = \boldsymbol{i} + \boldsymbol{q}_i$, where $\boldsymbol{i} \in \Lambda$ and $\boldsymbol{q}_i \in \boldsymbol{R}^d$ is the displacement of the particle at lattice site $\boldsymbol{i}$ from this equilibrium position. Since interatomic forces in real solids have short range, it is reasonable to assume that the atoms interact only with their nearest neighbors via a potential that depends only on the relative distance with respect to the equilibrium distance. Accordingly, the Hamiltonians that we consider have the general form

$$H(P, Q) = \sum_{i \in \Lambda} \frac{\boldsymbol{p}_i^2}{2m} + \sum_{|i-j|=1} V(\boldsymbol{q}_i - \boldsymbol{q}_j) + \sum_i U_i(\boldsymbol{q}_i) = \sum_{i \in \Lambda} \frac{\boldsymbol{p}_i^2}{2m} + \mathcal{V}(Q) \qquad [10]$$

where $P = (\boldsymbol{p}_i)_{i \in \Lambda}$ and analogously for $Q$. We shall further assume that as $|\boldsymbol{q}| \to \infty$ so do $U_i(\boldsymbol{q})$ and $V(\boldsymbol{q})$. The addition of $U_i(\boldsymbol{q})$ pins down the crystal and ensures that $\exp[-\beta H(P, Q)]$ is integrable with respect to $\mathrm{d}P\mathrm{d}Q$, and thus the corresponding Gibbs measure is well defined. In this case, in order to fix the temperature at the boundary, one can add a Langevin term to the equation of particles on the boundaries, that is, if $\boldsymbol{i} \in \partial\Lambda_\alpha$ the equation for the particle is

$$\dot{\boldsymbol{p}}_i = -\partial_{q_i} H(P, Q) - \lambda \boldsymbol{p}_i + \sqrt{\lambda T_\alpha}\, \dot{w}_i \qquad [11]$$

where $\dot{w}_i$ is a standard white noise. Other thermostatting mechanisms can be considered. In this case we can also define LTE using eqn [9] but we run into the same difficulties described above – although the problem is somehow simpler due to the presence of the lattice structure and the fact that the particles oscillate close to their equilibrium points. We can obtain Fourier's law only by adding stochastic terms, for example, terms like eqn [11], to the equation of motion of every particle and assuming that $U(\boldsymbol{q})$ and $V(\boldsymbol{q})$ are harmonic. These added noises can be thought of as an effective description of the chaotic motion generated by the anharmonic terms in $U(\boldsymbol{q})$ and $V(\boldsymbol{q})$.

Just how far we are from establishing rigorously the Fourier law is clear from our very limited mathematical understanding of the stationary nonequilibrium state (SNS) of mechanical systems whose ends are, as in the example of the Benard problem, kept at fixed temperatures $T_1$ and $T_2$. Various models have been considered, for example, models with Hamiltonian [10] coupled at the boundaries with heat reservoirs described by eqns [11]. The best mathematical results one can prove are: the existence and uniqueness of SNS; the existence of a stationary nontrivial heat flow; properties of the fluctuations of the heat flow in the SNS; the central-limit theorem type fluctuations (related to Kubo formula and Onsager relations; and large-deviation type fluctuations related to the Gallavotti–Cohen fluctuation theorem). What is missing is information on how the relevant quantities depend on the size of the system, $N$. In this context, the heat conductivity can be defined precisely without invoking LTE. To do this, we let $\tilde{J}$ be the expectation value in the SNS of the energy or heat current flowing from reservoir 1 to reservoir 2. We then define the conductivity $\kappa_L$ as $\tilde{J}/(A\delta T/L)$, where $\delta T/L = (T_1 - T_2)/L$ is the effective temperature gradient for a cylinder of microscopic length $L$ and uniform cross section $A$, and $\kappa(T)$ is the limit of $\kappa_L$ when $\delta T \to 0 (T_1 = T_2 = T)$ and $L \to \infty$. The existence of such a limit with $\kappa$ positive and finite is what one would like to prove.

*See also:* Dynamical Systems and Thermodynamics; Ergodic Theory; Interacting Particle Systems and Hydrodynamic Equations; Kinetic Equations; Nonequilibrium Statistical Mechanics: Dynamical Systems Approach; Nonequilibrium Statistical Mechanics: Interaction Between Theory and Numerical Simulations.

## Further Reading

Ashcroft NW and Mermin ND (1988) *Solid State Physics.* Philadelphia: Saunders College.

Berman R (1976) *Thermal Conduction in Solids.* Oxford: Clarendon.

Bonetto F, Lebowitz JL, and Rey-Bellet L (2000) Fourier's law: a challenge to theorists. In: Fokas A, Grigoryan A, Kibble T, and Zegarlinsky B (eds.) *Mathematical Physics 2000*, pp. 128–150. London: Imperial College Press.

Brush S (1994) *The Kind of Motion We Call Heat.* Amsterdam: Elsevier.

Debye P (1914) *Vorträge über die Kinetische Theorie der Wärme.* Leipzig: Teubner.

Fourier JBJ (1822) *Théorie Analytique de la Chaleur*, Paris: Firmin Didot. (English translation in 1878 by Alexander Freeman. New York: Dover, Reprinted in 2003).

Kipnis C and Landim C (1999) *Scaling Limits of Interacting Particle Systems.* Berlin: Springer.

Lepri S, Livi R, and Politi A (2003) Thermal conduction in classical low-dimensional lattices. *Physics Reports* 377(1): 1–80.
Peierls RE (1955) *Quantum Theory of Solids*. Oxford: Clarendon.
Peierls RE (1960) Quantum theory of solids. In: Saint-Aubin Y and Vinet L (eds.) *Theoretical Physics in the Twentieth Century*, pp. 140–160. New York: Interscience.
Rey-Bellet L (2003) Statistical mechanics of anharmonic lattices. In: *Advances in Differential Equations and Mathematical Physics*, (Birmingham, AL, 2002), Contemporary Mathematics,, vol. 327, pp. 283–298. Providence, RI: American Mathematical Society.
Spohn H (1991) *Large Scale Dynamics of Interacting Particle*. Berlin: Springer.

# Fourier–Mukai Transform in String Theory

**B Andreas**, Humboldt-Universität zu Berlin, Berlin, Germany



## Introduction

The Fourier–Mukai transform has been introduced in the study of abelian varieties by Mukai and can be thought of as a nontrivial algebro-geometric analog of the Fourier transform. Since its original introduction, the Fourier–Mukai transform turned out to be a useful tool for studying various aspects of sheaves on varieties and their moduli spaces, and as a natural consequence, to learn about the varieties themselves. Various links between geometry and derived categories have been uncovered; for instance, Bondal and Orlov proved that Fano varieties, and certain varieties of general type, can be reconstructed from their derived categories. Moreover, Orlov proved a derived version of the Torelli theorem for K3 surfaces and also a structure theorem for derived categories of abelian varieties. Later, Kawamata gave evidence to the conjecture that two birational smooth projective varieties with trivial canonical sheaves have equivalent derived categories, which has been proved by Bridgeland in dimension 3.

The Fourier–Mukai transform also enters into string theory. The most prominent example is Kontsevich's homological mirror-symmetry conjecture. The conjecture predicts (for mirror dual pairs of Calabi–Yau manifolds) an equivalence between the bounded derived category of coherent sheaves and the Fukaya category. The conjecture implies a correspondence between certain self-equivalences (given by Fourier–Mukai transforms) of the derived category and symplectic self-equivalences of the mirror manifold.

Besides their importance for geometrical aspects of mirror symmetry, the Fourier–Mukai transforms have also been important for heterotic string compactifications. The motivation for this came from the conjectured correspondence between the heterotic string and F-theory, which both rely on elliptically fibered Calabi–Yau manifolds. To give evidence for this correspondence, an explicit description of stable holomorphic vector bundles was necessary and inspired a series of publications by Friedman, Morgan, and Witten. Their bundle construction relies on two geometrical objects: a hypersurface in the Calabi–Yau manifold together with a line bundle on it; more precisely, they construct vector bundles using a relative Fourier–Mukai transform.

Various aspects and refinements of this construction have been studied by now. For instance, a physical way to understand the bundle construction can be given using the fact that holomorphic vector bundles can be viewed as D-branes and that D-branes can be mapped under $T$-duality to new D-branes (of different dimensions).

We survey aspects of the Fourier–Mukai transform, its relative version and outline the bundle construction of Friedman, Morgan, and Witten. The construction has led to many new insights, for instance, the presence of 5-branes in heterotic string vacua has been understood. The construction also inspired a tremendous amount of work towards a heterotic string phenomenology on elliptic Calabi–Yau manifolds. For the many topics omitted the reader should consult the "Further reading" section.

## The Fourier–Mukai Transforms

Every object $E$ of the derived category on the product $X \times Y$ of two smooth algebraic varieties $X$ and $Y$ gives rise to a functor $\Phi^E$ from the bounded derived category $D(X)$ of coherent sheaves on $X$ to the similar category on $Y$:

$$\Phi^E : D(X) \to D(Y)$$
$$F \mapsto \Phi^E(F) = R\hat{\pi}_*(\pi^* F \otimes E)$$

where $\pi, \hat{\pi}$ are the projections from $X \times Y$ to $X$ and $Y$, respectively, and $\otimes$ denotes the derived tensor product. $\Phi^E(F)$ is called Fourier–Mukai transform with kernel $E \in D(X \times Y)$ (in analogy

with the definition of an integral transform with kernel). Note that given a Fourier–Mukai functor $\Phi^E$, $\Phi^E(F)$ is in general a complex having homology in several degrees even if $F$ is a sheaf. Furthermore, a result by Orlov states that if $X$ and $Y$ are smooth projective varieties then any fully faithful functor $D(X) \to D(Y)$ is a Fourier–Mukai functor.

In analogy with the Fourier transform, there is a kind of "convolution product" giving the composition of two such functors. More precisely, given smooth algebraic varieties $X$, $Y$, $Z$, and elements $E \in D(X \times Y)$ and $G \in D(Y \times Z)$, we can define $G \circ E \in D(X \times Z)$ by

$$G \circ E = R\pi_{XZ_*}(\pi^*_{XY}E \otimes \pi^*_{YZ}G)$$

where $\pi_{XY}, \pi_{YZ}, \pi_{XZ}$ are the projections from $X \times Y \times Z$ to the pairwise products giving a natural isomorphism of functors

$$\Phi^G \circ \Phi^E = \Phi^{G \circ E}$$

Another analogy with the Fourier transform can be drawn. For this, assume that we have sheaves $F$ and $G$ which only have one nonvanishing Fourier–Mukai transform, the $i$th one $\Phi^i(F)$ (where $\Phi^i$: $D(X) \to \mathrm{Coh}(Y), F \mapsto \mathcal{H}^i(\Phi^E(F))$; cf. remarks below) in the case of $F$, and the $j$th one $\Phi^j(G)$ in the case of $G$. Given such sheaves, there is the Parseval formula

$$\mathrm{Ext}^h_X(F, G) = \mathrm{Ext}^{h+i-j}_Y(\Phi^i(F), \Phi^j(G))$$

which gives a correspondence between the extensions of $F, G$ and the extensions of their Fourier–Mukai transforms. This formula can be considered as the analog of the Parseval formula for the ordinary Fourier transform for functions on a torus.

The Parseval formula can be proved using two facts. First, for arbitrary coherent sheaves $E, G$ the Ext groups can be computed in terms of the derived category, namely

$$\mathrm{Ext}^i(E, G) = \mathrm{Hom}_{D(X)}(E, G[i])$$

Second, the Fourier–Mukai transforms of $F$ and $G$ in the derived category $D(X)$ are given by $\Phi(F) = \Phi^i(F)[-i]$ and $\Phi(G) = \Phi^j(G)[-j]$. Since the Fourier–Mukai transform is an equivalence of categories, we have

$$\mathrm{Hom}_{D(X)}(F, G[i]) = \mathrm{Hom}_{D(X)}(\Phi^i(F), \Phi^j(G)[i - j + h])$$

implying the Parseval formula.

A first simple example of a Fourier–Mukai functor can be given: let $F$ be the complex in $D(X \times X)$ defined by the structure sheaf $\mathcal{O}_\Delta$ of the diagonal $\Delta \subset X \times X$. Then it is easy to check that $\Phi^F$: $D(X) \to D(X)$ is isomorphic to the identity functor on $D(X)$. Moreover, if we shift degrees by $n$ taking $F = \mathcal{O}_\Delta[n]$ (a complex with only the sheaf $\mathcal{O}_\Delta$ placed in degree $n$), then $\Phi^F: D(X) \to D(X)$ is the degree shifting functor $\mathcal{G} \mapsto \mathcal{G}[n]$.

As we will be interested in relative Fourier–Mukai transforms for elliptic fibrations, let us consider the case of a Fourier– Mukai transform on an elliptic curve: consider an elliptic curve $E$ with a fixed origin $p_0$ and identify $E$ with $\hat{E} = \mathrm{Pic}^0(E)$ via $f: E \to \hat{E}, x \mapsto \mathcal{O}_E(x - p_0)$. As kernel we take the normalized Poincaré line bundle $\mathcal{P} := \mathcal{O}_{E \times E}(\Delta - \{p_0\} \times E - E \times \{p_0\})$. The restriction of $\mathcal{P}$ to $p_0 \times E$ or $E \times p_0$ is isomorphic to the trivial line bundle $\mathcal{O}$. $\mathcal{P}$ has the universal property which can be expressed by $\Phi^{\mathcal{P}}(k(x)) = f(x)$, where $k(x)$ is the sheaf supported at a point $x \in E$; in particular, $\Phi^{\mathcal{P}}(k(p_0)) = \mathcal{O}_E$ and $\Phi^{\mathcal{P}}(\mathcal{O}_E) = k(p_0)[-1]$, where $\mathcal{O}_E$ is the structure sheaf of $E$.

### Relative Fourier–Mukai Transforms for Elliptic Fibrations

It is often convenient to study problems for families rather than for single varieties. The main advantage of the relative setting is that base-change properties (or parameter dependencies) are better encoded into the problem. We can do that for Fourier–Mukai functors as well. To this end, we consider two morphisms $p: X \to B, \hat{p}: \widehat{X} \to B$ of algebraic varieties. We will assume that the morphisms are flat and so give nice families of algebraic varieties. We shall define relative Fourier–Mukai functors in this setting by means of a "kernel" $E$ in the derived category $D(X \times_B \widehat{X})$.

Let us make the relative setting explicit for elliptic fibrations: an elliptic fibration is a proper flat morphism $p: X \to B$ of schemes whose fibers are Gorenstein curves of arithmetic genus 1. We also assume that $p$ has a section $\sigma: B \hookrightarrow X$ taking values in the smooth locus $X' \to B$ of $p$. The generic fibres are then smooth elliptic curves, whereas some singular fibers are allowed. If the base $B$ is a smooth curve, elliptic fibrations were studied and classified by Kodaira, who described all the types of singular fibers that may occur, the so-called Kodaira curves. When the base is a smooth surface, more complicated configuration of singular curves can occur and have indeed been studied by Miranda.

First let us fix notation and setup. We denote by $\sigma = \sigma(B)$ the image of the section, by $X_t$ the fiber of $p$ over $t \in B$ (we assume, in what follows, $B$ is either a smooth curve or surface) and by $i_t: X_t \hookrightarrow X$ the inclusion. Furthermore, $\omega_{X/B}$ is the relative dualizing

sheaf and $\omega = R^1 p_* \mathcal{O}_X \xrightarrow{\sim} (p_* \omega_{X/B})^*$, where the isomorphism is Grothendieck–Serre duality for $p$. The sheaf $\mathcal{L} = p_* \omega_{X/B}$ is a line bundle whose first Chern class we denote by $K = c_1(\mathcal{L})$. The adjunction formula for $\sigma \hookrightarrow X$ gives that $\sigma^2 = -\sigma \cdot p^* K$ as cycles on $X$. Moreover, we will consider elliptic fibrations with a section whose fibers are all geometrically integral. This means that the fibration is isomorphic with its Weierstrass model.

From Kodaira's classification of possible singular fibers one finds that the components of reducible fibers of $p$ which do not meet $\sigma$ form rational double point configurations disjoint from $\sigma$. Let $X \to \bar{X}$ be the result of contracting these configurations and let $\bar{p} : \bar{X} \to B$ be the induced map. Then all fibers of $\bar{p}$ are irreducible with at worst nodes or cusps as singularities. In this case, one refers to $\bar{X}$ as the Weierstrass model of $X$.

The Weierstrass model can be constructed as follows: the divisor $3\sigma$ is relatively ample and, if $\mathcal{E} = p_* \mathcal{O}_X(3\sigma) \xrightarrow{\sim} \mathcal{O}_B \oplus \omega^{\otimes 2} \oplus \omega^{\otimes 3}$ and $\bar{p} : P = \mathbb{P}(\mathcal{E}^*) \to B$ is the associated projective bundle, there is a projective morphism $j : X \to P$ such that $j(X) = \bar{X}$.

Now special fibers of $X \to B$ can have at most one singular point, either a cusp or a simple node. Thus, in this case $3\sigma$ is relatively very ample and gives rise to a closed immersion $j : X \hookrightarrow P$ such that $j^* \mathcal{O}_P(1) = \mathcal{O}_X(3\sigma)$, where $j$ is locally a complete intersection whose normal sheaf is $\mathcal{N}(X/P) \xrightarrow{\sim} \pi^* \omega^{-\otimes 6} \otimes \mathcal{O}_X(9\sigma)$. This follows by relative duality since $\omega_{P/B} = \bigwedge \Omega_{P/B} \xrightarrow{\sim} \bar{\pi}^* \omega^{\otimes 5}(-3)$, due to the Euler exact sequence

$$0 \to \Omega_{P/B} \to \bar{\pi}^* \mathcal{E}(-1) \to \mathcal{O}_P \to 0$$

The morphism $p : X \to B$ is then a local complete intersection morphism (cf. Fulton (1984)) and has a virtual relative tangent bundle $T_{X/B} = [j^* T_{P/B}] - [\mathcal{N}_{X/P}]$ in the $K$-group $K^\bullet(X)$. The Todd class of $T_{X/B}$ is given by

$$\begin{aligned}\mathrm{Td}(T_{X/B}) = 1 - \tfrac{1}{2} p^{-1} K + \tfrac{1}{12}(12\sigma \cdot p^{-1} K + 13 p^{-1} K^2) \\ - \tfrac{1}{2} \sigma \cdot p^{-1} K^2 + \text{terms of higher degree}\end{aligned}$$

Now if $\hat{p} : \hat{X} \to B$ denotes the dual elliptic fibration, defined as the relative moduli space of torsion-free rank-1 sheaves of relative degree 0, it is known that for $t \in B$ there is an isomorphism $\hat{X}_t \cong X_t$ between the fibers of both fibrations. Since we assume that the original fibration $p : X \to B$ has a section $\sigma$, then $p$ and $\hat{p}$ are globally isomorphic; hereafter we identify $X \cong \widehat{X}$, where $\hat{X}$ denotes the compactified relative Jacobian of $X$.

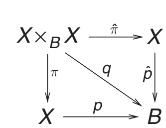
Note that $\hat{X}$ is the scheme representing the functor which, to any scheme morphism $\phi : S \to B$, associates the space of equivalence classes of $S$-flat sheaves on $p_s : X \times_B S \to S$, whose restrictions to the fibers of $\phi$ are torsion-free (the usual definition of "torsion free" is only for integral varieties, i.e., varieties whose local rings have no zero-divisors. In this case, a sheaf $M$ is torsion free if for any open subset $U$, any nonzero section $m$ of $M$ on $U$ and any nonzero section $a$ of the relevant functions sheaf, one has $a \cdot m \neq 0$. When the variety is not integral (it is reducible, or nonreduced) this definition has no real meaning, then what substitutes the notion of "torsion free" is the Simpson definition of "pure of maximal dimension": a sheaf $M$ is "torsion free" in this sense if the support of any of its subsheaves is the whole variety (cf. Huybrechts and Lehn (1997)), of rank 1 and degree 0; two such sheaves $\mathcal{F}, \mathcal{F}'$ are considered to be equivalent if $\mathcal{F}' \cong \mathcal{F} \otimes p_s^* \mathcal{L}$ for a line bundle $\mathcal{L}$ on $S$ (cf. Altman and Kleiman (1980); note the Altman–Kleiman compactification of the relative Jacobian applies to our situation since we consider elliptic fibrations with integral fibers). Moreover, the natural morphism $X \to \widehat{X}, x \mapsto \mathcal{I}_x \otimes \mathcal{O}_{X_t}(\sigma(t))$ is an isomorphism (of $B$-schemes); here $\mathcal{I}_x$ is the ideal sheaf of the point $x$ in $X_t$.

Note also that if $\pi : Y \to X_t$ is the normalization of one of our fibers $X_t$ and $z$ is the exceptional divisor (the pre-image of the singular point $x$) then $\pi_*(\mathcal{O}_Y(-z))$ is the maximal ideal of $x$.

The variety $\widehat{X}$ is a fine moduli space. This means that there exists a coherent sheaf $\mathcal{P}$ on $X \times_B \widehat{X}$ flat over $\widehat{X}$, whose restrictions to the fibers of $\hat{p}$ are torsion free, and of rank 1 and degree 0. The sheaf $\mathcal{P}$ is defined, up to tensor product, by the pullback of a line bundle on $\widehat{X}$, and is called the universal Poincaré sheaf, which we will normalize by letting $\mathcal{P}_{|\sigma \times_B \widehat{X}} \simeq \mathcal{O}_X$. We shall henceforth assume that $\mathcal{P}$ is normalized in this way, so that

$$\mathcal{P} = \mathcal{I}_\Delta \otimes \pi^* \mathcal{O}_X(\sigma) \otimes \hat{\pi}^* \mathcal{O}_X(\sigma) \otimes q^* \omega^{-1}$$

where $\pi, \hat{\pi}$ and $q = p \circ \pi = \hat{p} \circ \hat{\pi}$ refer to the diagram




and $\mathcal{I}_\Delta$ is the ideal sheaf of the diagonal immersion $X \hookrightarrow X \times_B X$.

Starting with the diagram and with the kernel given by the normalized relative universal Poincaré sheaf $\mathcal{P}$ on the fibered product $X \times_B X$, we define the relative Fourier–Mukai transform as

$$\begin{aligned}\Phi = \Phi^{\mathcal{P}} : D(X) \to D(X) \\ F \mapsto \Phi(F) = R\hat{\pi}_*(\pi^* F \otimes \mathcal{P})\end{aligned}$$

Note that $\Phi(F)$ can be generalized if we allow changes in the base space $B$, that is, we consider base-change morphisms $g: S \to B$.

We close this section with some remarks:

- An important feature of Fourier–Mukai functors is that they are exact as functors of triangulated categories. In more familiar terms, we can say that for any exact sequence $0 \to \mathcal{N} \to \mathcal{F} \to \mathcal{G} \to 0$ of coherent sheaves in $X$, we obtain an exact sequence

  $$\cdots \to \Phi^{i-1}(\mathcal{G}) \to \Phi^i(\mathcal{N}) \to \Phi^i(\mathcal{F}) \to \Phi^i(\mathcal{G}) \to \Phi^{i+1}(\mathcal{N}) \to \cdots$$

  where we have written $\Phi = \Phi^E$ and $\Phi^i(F) = \mathcal{H}^i(\Phi(F))$ denotes the $i$th cohomology sheaves of the complexes $\Phi(F)$.

  Given a Fourier–Mukai functor $\Phi^E$, a complex $F$ in $D(X)$ satisfies the $\mathrm{WIT}_i$ condition (or is $\mathrm{WIT}_i$) if there is a coherent sheaf $\mathcal{G}$ on $\widehat{X}$ such that $\Phi^E(F) \simeq \mathcal{G}[i]$ in $D(\widehat{X})$, where $\mathcal{G}[i]$ is the associated complex concentrated in degree $i$. Furthermore, we say that $F$ satisfies the $\mathrm{IT}_i$ condition if, in addition, $\mathcal{G}$ is locally free.

  When the kernel $E$ is simply a sheaf $\mathcal{Q}$ on $X \times \widehat{X}$ flat over $\widehat{X}$, the cohomology and base-change theorem (cf. Hartshorne (1977)) allows one to show that a coherent sheaf $\mathcal{F}$ on $X$ is $\mathrm{IT}_i$ if and only if $H^j(X, \mathcal{F} \otimes \mathcal{Q}_\xi) = 0$ for all $\xi \in \widehat{X}$ and for all $j \neq i$, where $\mathcal{Q}_\xi$ denotes the restriction of $\mathcal{Q}$ to $X \times \{\xi\}$ and $\mathcal{F}$ is $\mathrm{WIT}_0$ if and only if it is $\mathrm{IT}_0$.

  The acronym "IT" stands for "index theorem," while "W" stands for "weak." This terminology comes from Nahm transforms for connections on tori in complex differential geometry.
- The Parseval formula for the relative Fourier–Mukai transform has been proved by Mukai in his original Fourier–Mukai transform for abelian varieties and can be extended to any situation in which a Fourier–Mukai transform is fully faithful.
- For physical applications, it is often convenient to work in cohomology $H^*(X, \mathbb{Q})$. The passage from $D(X)$ to $H^*(X, \mathbb{Q})$ can be described as follows. We first send a complex $Z \in D(X)$ to its natural class in the $K$-group; we then make use of the fact that the Chern character ch maps $K(X) \to \mathrm{CH}^*(X) \otimes \mathbb{Q}$ and finally we apply the cycle map to $H^*(X, \mathbb{Q})$. This passage (by abuse of notation) is often denoted by $\mathrm{ch}: D(X) \to H^{\mathrm{even}}(X, \mathbb{Q})$, it commutes with pullbacks and transforms tensor products into dot products. Moreover, if we substitute the Mukai vector $v(Z) = \mathrm{ch}(Z)\sqrt{\mathrm{Td}(X)}$ for the Chern character $\mathrm{ch}(Z)$ then we find the commutative diagram

  $$\begin{array}{ccc} D(X) & \xrightarrow{\Phi^E} & D(Y) \\ \downarrow{\scriptstyle v} & & \downarrow{\scriptstyle v} \\ H^*(X, \mathbb{Q}) & \xrightarrow{\Phi^{v(E)}} & H^*(Y, \mathbb{Q}) \end{array}$$

  This can be shown using the Grothendieck–Riemann–Roch theorem and the fact that the power series defining the Todd class starts with constant term 1 and thus is invertible.

## Vector Bundles for Heterotic Strings

A compactification of the ten-dimensional heterotic string is given by a holomorphic, stable $G$-bundle $V$ (with $G$ some Lie group specified below) over a Calabi–Yau manifold $X$. The Calabi–Yau condition, the holomorphy and stability of $V$ are a direct consequence of the required supersymmetry in the uncompactified spacetime. We assume that the underlying ten-dimensional space $M_{10}$ is decomposed as $M_{10} = M_4 \times X$, where $M_4$ (the uncompactified spacetime) denotes the four-dimensional Minkowski space and $X$ a six-dimensional compact space given by a Calabi–Yau 3-fold. To be more precise: supersymmetry requires that the connection $A$ on $V$ satisfies

$$F_A^{2,0} = F_A^{0,2} = 0, \qquad F^{1,1} \wedge J^2 = 0$$

where $J$ denotes a Kähler form of $X$. It follows that the connection has to be a holomorphic connection on a holomorphic vector bundle and, in addition, satisfies the Donaldson–Uhlenbeck–Yau equation, which has a unique solution if and only if the vector bundle is polystable.

In addition to $X$ and $V$, we have to specify a $B$-field on $X$ of field strength $H$. In order to get an anomaly-free theory, the Lie group $G$ is fixed to be either $E_8 \times E_8$ or $\mathrm{Spin}(32)/\mathbb{Z}_2$ or one of their subgroups and $H$ must satisfy the identity

$$\mathrm{d}H = \mathrm{tr}\, R \wedge R - \mathrm{Tr}\, F \wedge F$$

where $R$ and $F$ are, respectively, the associated curvature forms of the spin connection on $X$ and the gauge connection on $V$. Also tr refers to the trace of the composite endomorphism of the tangent bundle to $X$ and Tr denotes the trace in the adjoint representation of $G$. For any closed four-dimensional submanifold $X_4$ of the ten-dimensional spacetime $M_{10}$, the 4-form $\mathrm{tr}\, R \wedge R - \mathrm{Tr}\, F \wedge F$ must have trivial cohomology. Thus, a necessary topological condition $V$ has to satisfy is $\mathrm{ch}_2(TX) = \mathrm{ch}_2(V)$, which simplifies to $c_2(TX) = c_2(V)$ for Calabi–Yau manifolds, $V$ being an $\mathrm{SU}(n)$ vector bundle.

A physical interpretation of the third Chern class can be given as a result of the decomposition of the ten-dimensional spacetime into a four-dimensional flat Minkowski space and $X$. The decomposition of the corresponding ten-dimensional Dirac operator with values in $V$ shows that massless four-dimensional fermions are in one-to-one correspondence with zero modes of the Dirac operator $D_V$ on $X$. The index of $D_V$ can be effectively computed using the Hirzebruch–Riemann–Roch theorem and is given by

$$\mathrm{index}(D) = \int_X \mathrm{Td}(X)\mathrm{ch}(V) = \frac{1}{2}\int_X c_3(V)$$

equivalently, we can write the index as $\mathrm{index}(D) = \sum_{i=0}^{3}(-1)^k \dim H^k(X, V)$. For stable vector bundles, we have $H^0(X, V) = H^3(X, V) = 0$ and so the index computes the net number of fermion generations $N_{\mathrm{gen}}$ in the respective model.

Now it has been observed that the inclusion of background 5-branes changes the anomaly constraint. Various 5-brane solutions of the heterotic string equations of motion have been discussed in the gauge 5-brane, the symmetric 5-brane, and the neutral 5-brane. It has been shown that the gauge and symmetric 5-brane solutions involve finite-size instantons of an unbroken nonabelian gauge group. In contrast, the neutral 5-branes can be interpreted as zero-size instantons of the SO(32) heterotic string. The magnetic 5-brane contributes a source term to the Bianchi identity for the 3-form $H$,

$$\mathrm{d}H = \mathrm{tr}\, R \wedge R - \mathrm{Tr}\, F \wedge F + n_5 \sum_{\text{five-branes}} \delta_5^{(4)}$$

and integration over a 4-cycle in $X$ gives the anomaly constraint

$$c_2(TX) = c_2(V) + [W]$$

The new term $\delta_5^{(4)}$ is a current that integrates to 1 in the direction transverse to a single 5-brane whose class is denoted by $[W]$. The class $[W]$ is the Poincaré dual of an integer sum of all these sources and thus $[W]$ should be an integral class, representing a class in $H_2(X, \mathbb{Z})$. $[W]$ can be further specified taking by into account that supersymmetry requires that 5-branes are wrapped on holomorphic curves and thus $[W]$ must correspond to the homology class of holomorphic curves. This fact constrains $[W]$ to be an algebraic class. Further, algebraic classes include negative classes; however, these lead to negative magnetic charges, which are unphysical, and so they have to be excluded. This constrains $[W]$ to be an effective class. Thus, for a given Calabi–Yau 3-fold $X$ the effectivity of $[W]$ constrains the choice of vector bundles $V$.

The study of the correspondence between the heterotic string (on an elliptic Calabi–Yau 3-fold) and $F$-theory (on an elliptic Calabi–Yau fourfold) has led Friedman, Morgan, and Witten to introduce a new class of vector bundles which satisfy the anomaly constraint with $[W]$ nonzero. As a result, they prove that the number obtained by integration of $[W]$ over the elliptic fibers of the Calabi–Yau 3–fold agrees with the number of 3-branes given by the Euler characteristic of the Calabi–Yau fourfold divided by 24.

### Fourier–Mukai Transforms and Spectral Covers

Let us now describe how the construction of vector bundles out of spectral data (first considered in Hitchin and Beauville, Narasimhan, and Ramanan) can be easily described in the case of elliptic fibrations by means of the relative Fourier–Mukai transform. This construction was widely exploited by Friedman, Morgan, and Witten to construct stable vector bundles on elliptic Calabi–Yau three-folds $X$, which we will summarize now.

If $V \to X$ is a vector bundle of rank $n$ which is semistable and of degree 0 on each fibre $\mathfrak{f}$ of $X \to B$, then its Fourier–Mukai transform $\Phi^1(V)$ is a torsion sheaf of pure dimension 2 on $X$. The support of $\Phi^1(V)$ is a surface $i : C \hookrightarrow X$, which is finite of degree $n$ over $B$. Moreover, $\Phi^1(V)$ is of rank 1 on $C$ and, if $C$ is smooth, then $\Phi^1(V) = i_* L$ is just the extension by zero of some line bundle $L \in \mathrm{Pic}(C)$. Conversely, given a sheaf $\mathcal{G} \to X$ of pure dimension 2 which is flat over $B$, then $\Phi(\mathcal{G})$ is a vector bundle on $X$ of rank equal to the degree of $\mathrm{supp}(\mathcal{G})$ over $B$.

This correspondence between vector bundles on $X$ and sheaves on $X$ supported on finite covers of $B$ is known as the spectral cover construction. The torsion sheaf $\mathcal{G}$ is called the spectral sheaf (or line bundle) and the surface $C = \mathrm{supp}(\mathcal{G})$ is called the spectral cover.

For the description of vector bundles on elliptic Calabi–Yau 3-folds $X$ it is appropriate to take $i_* L$ with Chern characters given by ($\eta_E, \eta \in H^2(B, \mathbb{Q})$ and $a_E, s_E \in \mathbb{Z}$)

$$\mathrm{ch}_0(i_* L) = 0, \qquad \mathrm{ch}_1(i_* L) = n\sigma + \pi^*\eta$$
$$\mathrm{ch}_2(i_* L) = \sigma\pi^*\eta_E + a_E \mathfrak{f}, \qquad \mathrm{ch}_3(i_* L) = s_E$$

The characteristic classes of the rank-$n$ vector bundle $V$ can be obtained if we apply the Grothendieck–Riemann–Roch theorem to the projection $\pi$:

$$\mathrm{ch}(V) = \pi_*[\hat{\pi}^*(\mathrm{ch}(i_* L))\, \mathrm{ch}(\mathcal{P})\mathrm{Td}(T_{X/B})]$$

where $\mathrm{Td}(T_{X/B})$ as given above.

To make sure that the construction leads to SU($n$) vector bundles we set $\eta_E = (1/2)nc_1$ giving $c_1(V) = 0$ and the remaining Chern classes are given by

$$c_2(V) = \pi^*(\eta)\sigma + \pi^*(\varpi), \qquad c_3(V) = -2\gamma_{|s}$$

where

$$\varpi = \tfrac{1}{24}c_1(B)^2(n^3 - n) + \tfrac{1}{2}(\lambda^2 - \tfrac{1}{4})n\eta(\eta - nc_1(B))$$

and $\gamma \in H^{1,1}(C, \mathbb{Z})$ is some cohomology class satisfying $\pi_{C*}\gamma = 0 \in H^{1,1}(B, \mathbb{Z})$. The general solution for $\gamma$ has been derived by Friedman, Morgan, and Witten and is given by $\gamma = \lambda(n\sigma_{|C} - \pi_C^*\eta + n\pi_C^* c_1(B))$ and $\gamma_{|s} = -\lambda\pi^*\eta(\pi^*\eta - n\pi^* c_1(B))\sigma$ with $S = C \cap \sigma$. The parameter $\lambda$ has to be determined such that $c_1(L)$ is an integer class. If $n$ is even, $\lambda = m (m \in \mathbb{Z})$ and in addition we must impose $\eta = c_1(B)$ modulo 2. If $n$ is odd, $\lambda = m + 1/2$.

It remains to discuss the stability of $V$. The stability depends on the properties of the defining data $C$ and $L$. If $C$ is irreducible and $L$ a line bundle over $C$ then $V$ will be a vector bundle stable with respect to the polarization

$$J = \epsilon J_0 + \pi^* H_B, \quad \epsilon > 0$$

if $\epsilon$ is sufficiently small. This has been proved by Friedman, Morgan, and Witten under the additional assumption that the restriction of $V$ to the generic fiber is regular and semistable. Here $J_0$ refers to some arbitrary Kähler class on $X$ and $H_B$ a Kähler class on the base $B$. It implies that the bundle $V$ can be taken to be stable with respect to $J$ while keeping the volume of the fiber $\mathfrak{f}$ of $X$ arbitrarily small compared to the volumes of effective curves associated with the base. That $J$ is actually a good polarization can be seen by assuming $\epsilon = 0$. Now we observe that $\pi^* H_B$ is not a Kähler class on $X$ since its integral is non-negative on each effective curve $C$ in $X$; however, there is one curve, the fiber $\mathfrak{f}$, where the integral vanishes. This means that $\pi^* H_B$ is on the boundary of the Kähler cone and, to make $V$ stable, we have to move slightly into the interior of the Kähler cone, that is, into the chamber which is closest to the boundary point $\pi^* H_B$. Also we note that although $\pi^* H_B$ is in the boundary of the Kähler cone, we can still define the slope $\mu_{\pi^* H_B}(V)$ with respect to it. Since $(\pi^* H_B)^2$ is some positive multiple of the class of the fiber $\mathfrak{f}$, semistability with respect to $\pi^* H_B$ is implied by the semistability of the restrictions $V|_{\mathfrak{f}}$ to the fibers. Assume that $V$ is not stable with respect to $J$, then there is a destabilizing sub-bundle $V' \subset V$ with $\mu_J(V') \geq \mu_J(V)$. But semistability along the fibers says that $\mu_{\pi^* H_B}(V') \leq \mu_{\pi^* H_B}(V)$. If we had equality, it would follow that $V'$ arises by the spectral construction from a proper subvariety of the spectral cover of $V$, contradicting the assumption that this cover is irreducible. So we must have a strict inequality $\mu_{\pi^* H_B}(V') < \mu_{\pi^* H_B}(V)$. Now taking $\epsilon$ small enough, we can ensure that $\mu_J(V') < \mu_J(V)$, thus $V'$ cannot destabilize $V$.

## D-Branes and Homological Mirror Symmetry

Kontsevich proposed a homological mirror symmetry for a pair $(X, Y)$ of mirror dual Calabi–Yau manifolds; it is conjectured that there exists a categorical equivalence between the bounded derived category $D(X)$ and Fukaya's $A_\infty$ category $\mathcal{F}(Y)$, which is defined by using the symplectic structure on $Y$. A Lagrangian submanifold with a flat bundle gives an object of $\mathcal{F}(Y)$. If we consider a locally trivial family of symplectic manifolds $Y$ (i.e., the symplectic form is locally constant as we vary $Y$ in the family) the object of $\mathcal{F}(Y)$ undergoes monodromy transformations going round a loop in the base. On the other hand, the object of $D(X)$ is a complex of coherent sheaves on $X$ and under the categorical equivalence between $D(X)$ and $\mathcal{F}(Y)$ the monodromy (of 3-cycles) is mapped to certain self-equivalences in $D(X)$.

Since all elements in $D(X)$ can be represented by suitable complexes of vector bundles on $X$, we can consider the topological $K$-group and the image $K_{\text{hol}}(X)$ of $D(X)$. The Fourier–Mukai transform $\Phi^{\mathcal{E}}: D(X) \to D(X)$ induces then a corresponding automorphism $K_{\text{hol}}(X) \to K_{\text{hol}}(X)$ and also an automorphism on $H^{\text{even}}(X, \mathbb{Q})$ if we use the Chern character ring homomorphism $\text{ch}: K(X) \to H^{\text{even}}(X, \mathbb{Q})$, as described above. With this in mind, we can introduce various kernels and their associated monodromy transformations.

For instance, let $D$ be the associated divisor defining the large-radius limit in the Kähler moduli space and consider the kernel $\mathcal{O}_\Delta(D)$, with $\Delta$ being the diagonal in $X \times X$. The corresponding Fourier–Mukai transform acts on an object $G \in D(X)$ as twisting by a line bundle, that is, $G \mapsto G \otimes \mathcal{O}(D)$. This automorphism is then identified with the monodromy about the large complex structure limit point (LCSL point) in the complex structure moduli space.

Furthermore, if we consider the kernel given by the ideal sheaf $\mathcal{I}_\Delta$ on $\Delta$, we find that the action of $\Phi^{\mathcal{I}_\Delta}$ on $H^{\text{even}}(X)$ can be expressed by taking the Chern character ring homomorphism:

$$\begin{aligned}\text{ch}(\Phi^{\mathcal{I}_\Delta}(G)) &= \text{ch}_0(\Phi^{\mathcal{O}_{X\times X}}(G)) - \text{ch}(G) \\ &= \left(\int \text{ch}(G) \cdot \text{Td}(X)\right) - \text{ch}(G)\end{aligned}$$

Kontsevich proposed that this automorphism should reproduce the monodromy about the principal component of the discriminant of the mirror family $Y$. At the principal component we have vanishing $S^3$ cycles (and the conifold singularity), thus the action of this monodromy on cohomology may be identified with the Picard–Lefschetz formula.

Now for a given pair of mirror dual Calabi–Yau 3-folds, it is generally assumed that $A$-type and $B$-type D-branes exchange under mirror symmetry. For such a pair, Kontsevich's correspondence between automorphisms of $D(X)$ and monodromies of 3-cycles can then be tested. More specifically, a comparison relies on the identification of two central charges associated to D-brane configurations on both sides of the mirror pair.

For this, we first have to specify a basis for the 3-cycles $\Sigma_i \in H^3(Y, \mathbb{Z})$ such that the intersection form takes the canonical form $\Sigma_i \cdot \Sigma_j = \delta_{j, i+b_{2,1}+1} = \eta_{i,j}$ for $i = 0, \ldots, b_{2,1}$. It follows that a 3-brane wrapped about the cycle $\Sigma = \sum_i n_i \Sigma^i$ has an (electric, magnetic) charge vector $\boldsymbol{n} = (n_i)$. The periods of the holomorphic 3-form $\Omega$ are then given by

$$\Pi_i = \int_{\Sigma_i} \Omega$$

and can be used to provide projective coordinates on the complex structure moduli space. If we choose a symplectic basis $(A_i, B_j)$ of $H_2(Y, \mathbb{Z})$ then the $A_i$ periods serve as projective coordinates and the $B_j$ periods satisfy the relations $\Pi^j = \eta_{i,j} \partial \mathcal{F} / \partial \Pi^i$, where $\mathcal{F}$ is the prepotential which has, near the large-radius limit, the asymptotic form (as analyzed by Candelas, Klemm, Theisen, Yau, and Hosono, cf. "Further reading"):

$$\mathcal{F} = \frac{1}{6} \sum_{abc} k_{abc} t_a t_b t_c + \frac{1}{2} \sum_{ab} c_{ab} t_a t_b - \sum_a \frac{c_2(X) J_a}{24} t_a + \frac{\zeta(3)}{2(2\pi i)^3} \chi(X) + \text{const.}$$

where $\chi(X)$ is the Euler characteristic of $X, c_{ab}$ are rational constants (with $c_{ab} = c_{ba}$) reflecting an $\mathrm{Sp}(2h^{11} + 2)$ ambiguity, and $k_{abc}$ is the classical triple intersection number given by

$$k_{abc} = \int_X J_a \wedge J_b \wedge J_c$$

The periods determine the central charge $Z(\boldsymbol{n})$ of a 3-brane wrapped about the cycle $\Sigma = \sum_i n_i [\Sigma_i]$:

$$Z(\boldsymbol{n}) = \int_\Sigma \Omega = \sum_i n_i \Pi_i$$

On the other hand, the central charge associated with an object $E$ of $D(X)$ is given by

$$Z(E) = -\int_X \mathrm{e}^{-t_a J_a} \operatorname{ch}(E) \left( 1 + \frac{c_2(X)}{24} \right)$$

Now, physically it is assumed that the two central charges are to be identified under mirror symmetry. If we compare the two central charges $Z(\boldsymbol{n})$ and $Z(E)$, then we obtain a map relating the Chern characters $\operatorname{ch}(E)$ of $E$ to the D-brane charges $\boldsymbol{n}$. If we insert the expressions for $\operatorname{ch}(E)$ in $\operatorname{ch}(\Phi^{\mathcal{I}_\Delta}(E))$, it yields a linear transformation acting on $\boldsymbol{n}$, such that $n_6 \to n_6 + n_3$, which agrees with the monodromy transformation about the conifold locus.

Similarly, the monodromy transformation about the LCSL point corresponding to automorphisms $[E] \to [E \otimes \mathcal{O}_X(D)]$ can be made explicit.

Using the central charge identification, the automorphism/monodromy correspondence has been made explicit for various dual pairs of mirror Calabi–Yau 3-folds (given as hypersurfaces in weighted projective spaces). This identification provides evidence for Kontsevich's proposal of homological mirror symmetry.

*See also:* Derived Categories; Mirror Symmetry: A Geometric Survey.

## Further Reading


Altman AB and Kleiman SL (1980) Compactifying the Picard scheme. *Advances in Mathematics* 35: 50–112.

Bartocci C, Bruzzo U, Hernández Ruipérez D, and Jardim M (2006) Nahm and Fourier–Mukai transforms in geometry and mathematical physics. *Progress in Mathematical Physics* (to appear).

Bondal AI and Orlov DO (2001) Reconstruction of a variety from the derived category and groups of autoequivalences. *Compositio Mathematica* 125: 327–344.

Beauville A, Narasimhan MS, and Ramanan S (1989) Spectral curves and the generalised theta divisor. *J. Reine Angew. Math.* 398: 169–179.

Callan CG Jr., Harvey JA, and Strominger A (1991) Worldbrane actions for string solitons. *Nuclear Physics B* 367: 60–82.

Candelas P, Font A, Katz S, and Morrison DR (1994) Mirror symmetry for two-parameter models. II. *Nuclear Physics B* 429: 626–674.

Donagi RY (1997) Principal bundles on elliptic fibrations. *Asian Journal of Mathematics* 1: 214–223 (alg-geom/9702002).

Donagi RY (1998) Taniguchi lecture on principal bundles on elliptic fibrations, hep-th/9802094.

Friedman R, Morgan JW, and Witten E (1997) Vector bundles and *F* theory. *Communications in Mathematical Physics* 187: 679–743.

Fulton W (1984) *Intersection Theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 2. Berlin: Springer.

Hartshorne R (1977) *Algebraic Geometry*. Berlin: Springer.

Huybrechts D (2006) Fourier–Mukai transforms in algebraic geometry. Oxford: Oxford University Press (to appear).
Huybrechts D and Lehn M (1997) The geometry of moduli spaces of sheaves. Braunschweig/Wiesbaden: Vieweg & Sohn.
Hitchin NJ (1987) The self-duality equations on a Riemann surface. *Proceedings of London Mathematical Society* 55(3): 59–126.
Hosono S, Klemm A, Theisen S, and Yau S-T (1995) Mirror symmetry, mirror map and applications to complete intersection Calabi–Yau spaces. *Nuclear Physics B* 433: 501–552.
Hosono S (1998) GKZ systems, Gröbner fans, and moduli spaces of Calabi–Yau hypersurfaces. In: *Topological Field Theory, Primitive Forms and Related Topics (Kyoto, 1996)*, Progr. Math.,, vol. 160, pp. 239–265. Boston: Birkhäuser.
Kawamata Y (2002) D-equivalence and K-equivalence. *Journal of Differential Geometry* 61: 147–171.
Kodaira K (1963a) On compact analytic surfaces. II, III. *Annals of Mathematics* 77(2): 563–626.
Kodaira K (1963b) On compact analytic surfaces. I, III. *Annals of Mathematics* 78: 1–40.
Kontsevich M (1995) Homological algebra of mirror symmetry. In: *Proceedings of the International Congress of Mathematicians, vol. 1, 2 (Zürich, 1994)*, pp. 120–139. Basel: Birkhäuser.
Miranda R (1983) Smooth models for elliptic threefolds. In: *The Birational Geometry of Degenerations (Cambridge, MA, 1981), Progr. Math*., vol. 29, pp. 85–133. Boston: Birkhäuser.
Mukai S (1981) Duality between $D(X)$ and $D(\hat{X})$ with its application to Picard sheaves. *Nagoya Math. J.* 81: 153–175.
Orlov DO (1997) Equivalences of derived categories and K3 surfaces. *Journal of Mathematical Sciences (New York)* 84: 1361–1381 (Algebraic geometry, 7).
Witten E (1986) New issues in manifolds of SU(3) holonomy. *Nuclear Physics B* 268: 79–112.

# Four-Manifold Invariants and Physics

**C Nash**, National University of Ireland, Maynooth, Ireland



## Introduction

Manifolds of dimension 4 play a distinguished role in physics and have done so ever since special and general relativity ushered in the celebrated four-dimensional spacetime. It is also the case that manifolds of dimension 4 play a distinguished role in mathematics: many generalities about manifolds of a general dimension do not apply in dimension 4; there are also phenomena in dimension 4 with no counterpart in other dimensions.

This article describes some of the more important physical and mathematical properties of dimension 4. We begin with an account of some topological and geometric properties for manifolds in general, but avoiding dimension 4, and then embark on the dimension 4 discussion. The references at the end will serve to take the reader further into the subject.

## Topological, Piecewise-Linear, and Differentiable Structures for Manifolds

In dealing with topological spaces which are manifolds, one distinguishes three types of manifolds $M$: topological, piecewise-linear, and differentiable (also called smooth). It is possible to describe the more important differences between these three types using topological techniques.

Consider then a manifold $M$ of dimension $n$; $M$ will always be assumed to be compact, connected and closed unless we indicate the contrary. The type of $M$ is determined by examining whether the transition functions $g_{\alpha\beta}$ are homeomorphisms, (invertible) piecewise-linear maps, or diffeomorphisms. Now, since the transition functions are maps from one subset of $\boldsymbol{R}^n$ to another, we introduce the groups $TOP_n$, $PL_n$, and $DIFF_n$ which are all the homeomorphisms, piecewise-linear maps, and diffeomorphisms of $\boldsymbol{R}^n$, respectively. We are naturally led to the three sets of inclusions:

$$\begin{array}{ccccccccc} TOP_1 & \subset & TOP_2 & \subset & \cdots & \subset & TOP_n & \subset & \cdots \\ PL_1 & \subset & PL_2 & \subset & \cdots & \subset & PL_n & \subset & \cdots \\ DIFF_1 & \subset & DIFF_2 & \subset & \cdots & \subset & DIFF_n & \subset & \cdots \end{array} \quad [1]$$

For each of the three sets of inclusions we pass to the direct limit and construct the three limiting groups

$$TOP, \quad PL, \quad DIFF \quad [2]$$

With these three groups are associated the classifying spaces $BTOP$, $BPL$ and $BDIFF$. The transition functions $g_{\alpha\beta}$ are those of the tangent bundle to $M$; and there are three possible tangent bundles depending on the type of $M$ and we denote these tangent bundles by $TM_{TOP}, TM_{PL}$, and $TM_{DIFF}$ in an obvious notation. Then to determine the tangent bundles $TM_{TOP}$, $TM_{PL}$, and $TM_{DIFF}$ one simply selects an element of the homotopy classes

$$[M, BTOP], \quad [M, BPL], \text{ and } [M, BDIFF] \quad [3]$$

respectively.

Given this threefold hierarchy of manifold structures one wishes to know when one can straighten out a topological manifold to make it piecewise linear; and also, when can one *smooth* a piecewise-linear manifold to make it differentiable?

If $\dim M \geq 5$ of $M$ these two questions can be formulated as lifting problems.

### TOP versus PL for $\dim M \neq 4$

Taking the first of them, so that we are comparing piecewise-linear and topological structures on $M$, one can check $BPL$ fibers over $BTOP$ with fiber $TOP/PL$ yielding

$$\begin{array}{ccc} TOP/PL & \rightarrow & BPL \\ & & \downarrow \pi \\ & & BTOP \end{array} \qquad [4]$$

A method for *straightening out* a PL manifold is now apparent: now a topological manifold is a choice of map $\alpha : M \rightarrow BTOP$, and a *factorization* of $\alpha$ through $BPL$ will give $M$ a PL structure. We show this below

$$\begin{array}{ccc} & BPL & \\ {\scriptstyle\beta}\nearrow & \downarrow \pi & \alpha = \pi \circ \beta \\ M \xrightarrow{\alpha} & BTOP & \end{array} \qquad [5]$$

The existence of the map $\beta : M \rightarrow BPL$ satisfying $\alpha = \pi \circ \beta$ provides $M$ with a PL structure and is a *lifting* of the map $\alpha$ from the base $BTOP$ to the total space $BPL$.

This lifting method, for passing from TOP structures to PL structures, does work, *provided* $\dim M \geq 5$, since we have the stability result that

$$\frac{TOP_n}{PL_n} \simeq \frac{TOP}{PL}, \quad n \geq 5 \qquad [6]$$

For the map $\beta$ to exist the obstructions to the lifting which are cohomology classes of the form

$$H^{k+1}(M; \pi_k(TOP/PL)) \qquad [7]$$

must vanish. However, Kirby and Siebenmann have shown that

$$TOP/PL \simeq K(\mathbf{Z}_2, 3) \qquad [8]$$

where $K(\mathbf{Z}_2, 3)$ is *Eilenberg–Mac Lane space* so that its sole nonvanishing homotopy group is in dimension 3 giving us

$$\pi_n(TOP/PL) = \begin{cases} \mathbf{Z}_2 & \text{if } n = 3 \\ 0 & \text{otherwise} \end{cases} \qquad [9]$$

Any obstruction to $\beta$'s existence is a class $e(M)$, say, in

$$H^4(M; \mathbf{Z}_2) \quad \dim M \geq 5 \qquad [10]$$

When $e(M)$ *vanishes*, the map $\beta$ exists and furnishes $M$ with a PL structure; if $e(M) = 0$ it is natural to go on to ask how many (homotopy classes of) such $\beta$'s exist? Standard obstruction theory says the relevant homotopy classes are just the whole cohomology group

$$H^k(M; \pi_k(TOP/PL)) \qquad [11]$$

which, since $k = 3$, is just

$$H^3(M; \mathbf{Z}_2) \qquad [12]$$

So, for $\dim M \geq 5$, we see that when a closed topological manifold $M$ acquires a PL structure by the lifting process just described, then the possible distinct PL structures are isomorphic to

$$H^3(M; \mathbf{Z}_2) \qquad [13]$$

which is not zero in general.

Finally, if $\dim M \leq 3$, then the notions PL and TOP coincide, so we are left with the case $\dim M = 4$ which we shall come to below. Now we wish to describe the next step in the sequence TOP, PL, DIFF which is the smoothing problem.

### PL versus DIFF for $\dim M \neq 4$

Similar ideas are used to address the question of smoothing a piecewise-linear manifold – however, the results are different. Let us assume that $M$ is a closed PL manifold with $\dim M \geq 5$. This time the fibration is

$$\begin{array}{ccc} PL/DIFF & \rightarrow & BDIFF \\ & & \downarrow \pi \\ & & BPL \end{array} \qquad [14]$$

The smoothing of a piecewise-linear $M$ can also be handled with obstruction theory and leads us immediately to the consideration of the homotopy groups $\pi_n(PL/DIFF)$. This time the nontrivial homotopy groups of the fiber are much more numerous than in the piecewise-linear case. In fact one has

$$\pi_n(PL/DIFF) = \begin{cases} 0 & \text{if } n \leq 6 \\ \mathbf{Z}_{28} & \text{if } n = 7 \\ \mathbf{Z}_2 & \text{if } n = 8 \\ \vdots & \vdots \\ \mathbf{Z}_{992} & \text{if } n = 11 \\ \vdots & \vdots \end{cases} \qquad [15]$$

The obstructions to passing from a PL to a DIFF structure on $M$ now lie in

$$H^{k+1}(M; \pi_k(PL/DIFF)) \qquad [16]$$

and the number of distinct liftings comprises the cohomology group

$$H^k(M; \pi_k(PL/DIFF)) \qquad [17]$$

As an illustration of all this, consider the case $M = S^7$; then the first nontriviality occurs when $n = 7$ and so the obstruction to smoothing $S^7$ lies in

$$H^8(S^7; \pi_7(PL/DIFF)) \qquad [18]$$

which is of course zero – this means that $S^7$ can be smoothed, a fact which we know from first principles. However, by the obstruction theory introduced above, the resulting smooth structures are isomorphic to

$$H^7(S^7; \pi_7(PL/DIFF)) = H^7(S^7; \mathbf{Z}_{28}) = \mathbf{Z}_{28} \qquad [19]$$

Hence, we have the celebrated result of Milnor and Kervaire and Milnor that $S^7$ has 28 distinct differentiable structures, 27 of which correspond to what are known as *exotic* spheres.

Lastly, if $\dim M \leq 3$, then PL and DIFF coincide – this leaves us with the case of greatest interest namely $\dim M = 4$.

## The Strange Case of Four Dimensions

In four dimensions there are phenomena which have no counterpart in any other dimension. First of all, there are topological 4-manifolds which have no smooth structure, though if they have a PL structure, then they possess a unique smooth structure. Second, the impediment to the existence of a smooth structure is of a completely different type to that met in the standard obstruction theory – it is not the pullback of an element in the cohomology of a classifying space, that is, it is not a characteristic class. Also the four-dimensional story is far from completely known. Nevertheless, there are some very striking results dating from the early 1980s onwards.

We begin by disposing of the difference between PL and DIFF structures: our earlier results together with the vanishing statement

$$\pi_n(PL/DIFF) = 0, \quad n \leq 6 \qquad [20]$$

mean that every PL 4-manifold possesses a unique DIFF structure. Thus, we can take the crucial difference to be between DIFF and TOP.

In Freedman (1982) all, simply connected, *topological* 4-manifolds were classified by their intersection form $q$.

We recall that $q$ is a quadratic form constructed from the cohomology of $M$ as follows: take two elements $\alpha$ and $\beta$ of $H^2(M; \mathbf{Z})$ and form their cup product $\alpha \cup \beta \in H^4(M; \mathbf{Z})$; then we define $q(\alpha, \beta)$ by

$$q(\alpha, \beta) = (\alpha \cup \beta)[M] \qquad [21]$$

where $(\alpha \cup \beta)[M]$ denotes the integer obtained by evaluating $\alpha \cup \beta$ on the generating cycle $[M]$ of the top homology group $H_4(M; \mathbf{Z})$ of $M$. Poincaré duality ensures that such a form is always nondegenerate over $\mathbf{Z}$ and so has $\det q = \mp 1$; $q$ is then called unimodular. Also we refer to $q$, as "even" if all its diagonal entries are even, and as "odd" otherwise.

Freedman's work yields the following:

**Theorem** (Freedman). *A simply connected 4-manifold M with even intersection form q belongs to a unique homeomorphism class, while if q is odd there are precisely two nonhomeomorphic manifolds M with q as their intersection form.*

This is a very powerful result – the intersection form $q$ very nearly determines the homeomorphism class of a simply connected $M$, and actually only fails to do so in the odd case where there are still just two possibilities. Further, *every* unimodular quadratic form occurs as the intersection form of some manifold.

As an illustration of the impressive nature of Freedman's work, choose $M$ to be the sphere $S^4$, since $H^2(S^4; \mathbf{Z})$ is trivial, then $q$ is the zero quadratic form and is of course even; we write this as $q = \emptyset$. Now recollect that the *Poincaré conjecture* in four dimensions is the statement that any homotopy 4-sphere, $S^4_h$ say, is actually *homeomorphic* to $S^4$. Well, since $H^2(S^4_h; \mathbf{Z})$ is also trivial then any $S^4_h$ also has intersection form $q = \emptyset$. Applying Freedman's theorem to $S^4_h$ immediately asserts that $S^4_h$ belongs to a unique homeomorphism class which must be that of $S^4$ thereby establishing the Poincaré conjecture.

Freedman's result combined with a much earlier result of Rohlin (1952) also gives us an example of a nonsmoothable 4-manifold: Rohlin's theorem asserts that given a smooth, simply connected, 4-manifold with even intersection form $q$, then the signature – the signature of $q$ being defined to be the difference between the number of positive and negative eigenvalues of $q$ – $\sigma(q)$ of $q$ is divisible by 16.

Now write

$$q = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 2 \end{pmatrix} = E_8 \qquad [22]$$

($E_8$ is actually the Cartan matrix for the exceptional Lie algebra $e_8$), then, by inspection, $q$ is even, and by calculation, it has signature 8. By Freedman's theorem there is a single, simply connected, 4-manifold with intersection form $q = E_8$. However, by

Rohlin's theorem, it cannot be smoothed since its signature is 8.

The next breakthrough was due to Donaldson (1983). Donaldson's theorem is applicable to definite forms $q$, which by appropriate choice of orientation on $M$ we can take to be *positive* definite. One has:

**Theorem** (Donaldson). *A simply connected, smooth 4-manifold, with positive-definite intersection form $q$ is always diagonalizable over the integers to* $q = \mathrm{diag}(1, \ldots, 1)$.

One can immediately deduce that no, simply connected, 4-manifold for which $q$ is even and positive definite can be smoothed!

For example, the manifold with $q = E_8 \oplus E_8$ has signature 16 (by Rohlin's theorem). But since $E_8$ is even, then so is $E_8 \oplus E_8$ and so Donaldson's theorem forbids such a manifold from existing smoothly.

In fact, in contrast to Freedman's theorem, which allows *all* unimodular quadratic forms to occur as the intersection form of some topological manifold, Donaldson's theorem says that in the positive-definite, smooth, case only *one* quadratic form is allowed, namely $I$.

Donaldson's work makes contact with physics because it uses the *Yang–Mills equations* as we now outline.

Let $A$ be a connection on a principal SU(2) bundle over a simply connected 4-manifold $M$ with positive-definite intersection form. If the curvature 2-form of $A$ is $F$, then $F$ has an $L^2$ norm which is the Euclidean Yang–Mills action $S$. One has

$$S = \|F\|^2 = -\int_M \mathrm{tr}(F \wedge *F) \qquad [23]$$

where $*F$ is the usual dual 2-form to $F$. The minima of the action $S$ are given by those $A$, called instantons, which satisfy the famous self-duality equations

$$F = *F \qquad [24]$$

Given one instanton $A$ which minimizes $S$ one can perturb about $A$ in an attempt to find more instantons. This process is successful and the space of all instantons can be fitted together to form a global moduli space of finite dimension. For the instanton which provides the absolute minimum of $S$, the moduli space $\mathcal{M}$ is a noncompact space of dimension 5.

We can now summarize the logic that is used to prove Donaldson's theorem: there are very strong relationships between $M$ and the moduli space $\mathcal{M}$; for example, let $q$ be regarded as an $n \times n$ matrix with precisely $p$ unit eigenvalues (clearly $p \le n$ and Donaldson's theorem is just the statement that $p = n$), then $\mathcal{M}$ has precisely $p$ singularities which look like cones on the space $CP^2$. These combine to produce the result that the 4-manifold $M$ has the same topological signature Sign($M$) as $p$ copies of $CP^2$; and so they have signature $a - b$, where $a$ of the $CP^2$'s are oriented as usual and $b$ have the opposite orientation. Thus,

$$\mathrm{Sign}\,(M) = a - b \qquad [25]$$

Now by definition, $\mathrm{Sign}\,(M)$ is the signature $\sigma(q)$ of the intersection form $q$ of $M$. But, by assumption, $q$ is positive definite $n \times n$ so $\sigma(q) = n = \mathrm{Sign}\,(M)$. Hence,

$$n = a - b \qquad [26]$$

However, $a + b = p$ and $p \le n$ so we can say that

$$n = a - b, \qquad p = a + b \le n \qquad [27]$$

but one always has $a + b \ge a - b$ so we have

$$n \le p \le n \Rightarrow p = n \qquad [28]$$

which is Donaldson's theorem.

## Donaldson's Polynomial Invariants

Donaldson extended his work by introducing polynomial invariants also derived from Yang–Mills theory and to discuss them we must introduce some notation.

Let $M$ be a smooth, simply connected, orientable Riemannian 4-manifold without boundary and $A$ be an SU(2) connection which is anti-self-dual so that

$$F = - * F \qquad [29]$$

Then the space of all gauge-inequivalent solutions to this anti-self-duality equation – the moduli space $\mathcal{M}_k$ – has a dimension given by the integer

$$\dim \mathcal{M}_k = 8k - 3(1 + b_2^+) \qquad [30]$$

Here *$k$ is the instanton number* which gives the topological type of the solution $A$. The instanton number is minus the second Chern class $c_2(F) \in H^2(M; \mathbf{Z})$ of the bundle on which the $A$ is defined. This means that we have

$$k = -c_2(F)[M] = \frac{1}{8\pi^2} \int_M \mathrm{tr}(F \wedge F) \in \mathbf{Z} \qquad [31]$$

The number $b_2^+$ is defined to be the rank of the positive part of the *intersection form $q$ of $M$*.

A Donaldson invariant $q^M_{d,r}$ is a symmetric integer polynomial of degree $d$ in the 2-homology $H_2(M;\mathbf{Z})$ of $M$

$$q^M_{d,r} : \underbrace{H_2(M) \times \cdots \times H_2(M)}_{d \text{ factors}} \longrightarrow \mathbf{Z} \qquad [32]$$

Given a certain map $m_i$,

$$m_i : H_i(M) \to H^{4-i}(\mathcal{M}_k) \qquad [33]$$

if $\alpha \in H_2(M)$ and $*$ represents a point in $M$, we define $q^M_{d,r}(\alpha)$ by writing

$$q^M_{d,r}(\alpha) = m_2^d(\alpha) m_0^r(*)[\mathcal{M}_k] \qquad [34]$$

The evaluation of $[\mathcal{M}_k]$ on the RHS of the above equation means that

$$2d + 4r = \dim \mathcal{M}_k \qquad [35]$$

so that $\mathcal{M}_k$ is even dimensional; this is achieved by requiring $b_2^+$ to be odd.

Now the Donaldson invariants $q^M_{d,r}$ are *differential* topological invariants rather than topological invariants but they are difficult to calculate as they require detailed knowledge of the instanton moduli space $\mathcal{M}_k$. However they are nontrivial and their values are known for a number of 4-manifolds $M$. For example, if $M$ is a complex algebraic surface, a positivity argument shows that they are nonzero when $d$ is large enough. Conversely, if $M$ can be written as the connected sum

$$M = M_1 \# M_2$$

where both $M_1$ and $M_2$ have $b_2^+ > 0$, then they all vanish.

## Topological Quantum Field Theories

Turning now to physics, it is time to point out that the $q^M_{d,r}$ can also be obtained, Witten (1988), as the correlation functions of twisted $N=2$ supersymmetric topological quantum field theory.

The action $S$ for this theory is given by

$$\begin{aligned} S = \int_M \mathrm{d}^4x \sqrt{g}\, \mathrm{tr} \Big\{ & \frac{1}{4} F_{\mu\nu}F^{\mu\nu} + \frac{1}{4} F^*_{\mu\nu}F^{\mu\nu} \\ & + \frac{1}{2}\phi D_\mu D^\mu \lambda + \mathrm{i} D_\mu \psi_\nu \chi^{\mu\nu} - \mathrm{i}\eta D_\mu \psi^\mu \\ & - \frac{\mathrm{i}}{8}\phi[\chi_{\mu\nu}, \chi^{\mu\nu}] - \frac{\mathrm{i}}{2}\lambda[\psi_\mu, \psi^\mu] \\ & - \frac{\mathrm{i}}{2}\phi[\eta, \eta] - \frac{1}{8}[\phi, \lambda]^2 \Big\} \end{aligned} \qquad [36]$$

where $F_{\mu\nu}$ is the curvature of a connection $A_\mu$ and $(\phi, \lambda, \eta, \psi_\mu, \chi_{\mu\nu})$ are a collection of fields introduced in order to construct the right supersymmetric theory; $\phi$ and $\lambda$ are both spinless while the multiplet $(\psi_\mu, \chi_{\mu\nu})$ contains the components of a 0-form, a 1-form, and a self-dual 2-form, respectively.

The significance of this choice of multiplet is that the instanton deformation complex used to calculate $\dim \mathcal{M}_k$ contains precisely these fields.

Even though $S$ contains a metric, its correlation functions are *independent* of the metric $g$ so that $S$ can still be regarded as a topological quantum field theory. This is because both $S$ and its associated energy momentum tensor $T \equiv (\delta S/\delta g)$ can be written as BRST commutators $S = \{Q, V\}, T = \{Q, V'\}$ for suitable $V$ and $V'$.

With this theory, it is possible to show that the correlation functions are *independent* of the gauge coupling and hence we can evaluate them in a small coupling limit. In this limit, the functional integrals are dominated by the classical minima of $S$, which for $A_\mu$ are just the instantons

$$F_{\mu\nu} = -F^*_{\mu\nu} \qquad [37]$$

We also need $\phi$ and $\lambda$ to vanish for irreducible connections. If we expand all the fields around the minima up to quadratic terms and do the resulting Gaussian integrals, the correlation functions may be formally evaluated.

A general correlation function of this theory is given by

$$< P >= \int \mathcal{DF} \exp[-S]\, P(\mathcal{F}) \qquad [38]$$

where $\mathcal{F}$ denotes the collection of fields present in $S$ and $P(\mathcal{F})$ is some polynomial in the fields.

$S$ has been constructed so that the zero modes in the expansion about the minima are the tangents to the moduli space $\mathcal{M}_k$. This suggests doing the $\mathcal{DF}$ integration as follows: express the integral as an integral over modes, then integrate out all the nonzero modes first leaving a *finite-dimensional* integration over the compactified moduli space $\overline{\mathcal{M}}_k$. The Gaussian integration over the nonzero modes is a boson–fermion ratio of determinants, which supersymmetry constrains to be $\mp 1$, bosonic and fermionic eigenvalues being equal in pairs.

This amounts to writing

$$< P >= \int_{\overline{\mathcal{M}}_k} P_n \qquad [39]$$

where $P_n$ denotes some $n$-form over $\overline{\mathcal{M}}_k$ and $n = \dim \overline{\mathcal{M}}_k$. If the original polynomial $P(\mathcal{F})$ is judiciously chosen, then calculation of $< P >$ reproduces evaluation of the Donaldson polynomials $q^M_{d,r}$.

## The Seiberg–Witten Equations

The Seiberg–Witten equations constitute another breakthrough in the work on the topology of 4-manifolds, since they greatly simplify the calculation of the data supplied by the Donaldson polynomial invariants. We shall discuss this later below but turn now to the equations themselves.

If we choose an oriented, compact, closed, Riemannian manifold $M$, then the data we need for the Seiberg–Witten equations are a connection $A$ on a line bundle $L$ over $M$ and a "local spinor" field $\psi$. The Seiberg–Witten equations are then

$$\partial\!\!\!/_A \psi = 0, \qquad F^+ = -\tfrac{1}{2}\bar{\psi}\Gamma\psi \qquad [40]$$

where $\partial\!\!\!/_A$ is the Dirac operator and $\Gamma$ is made from the gamma matrices $\Gamma_i$ according to $\Gamma = (1/2)[\Gamma_i, \Gamma_j]\mathrm{d}x^i \wedge \mathrm{d}x^j$.

We call $\psi$ a local spinor because global spinors may not exist on $M$; however, in dimension 4, orientability guarantees that a $\mathrm{spin}_c$ structure exists on $M$ (a choice of $\mathrm{spin}_c$ structure on $M$ is an extra piece of data in the Seiberg–Witten case); $\psi$ is then the appropriate section for the $\mathrm{spin}_c$ bundle and behaves locally like a spinor coupled to the U(1) connection $A$. Let $\mathrm{Spin}_c(M)$ denote the set of isomorphism classes of $\mathrm{spin}_c$ structures on $M$ then, for the case $b_2^+ > 1$ – the case $b_2^+ = 1$ has some technicalities – the Seiberg–Witten invariants determine a map SW of the form

$$\mathrm{SW} : \mathrm{Spin}_c(M) \longrightarrow \mathbf{Z} \qquad [41]$$

We emphasize that $A$ is just a U(1) abelian connection and so $F = \mathrm{d}A$, with $F^+$ denoting the self-dual part of $F$.

We shall now have a look at an example of a new result obtained directly from the Seiberg–Witten equations. The equations clearly provide the absolute minima for the action

$$S = \int_M \left\{ |\partial\!\!\!/_A\psi|^2 + \tfrac{1}{2}|F^+ + \tfrac{1}{2}\bar{\psi}\Gamma\psi|^2 \right\} \qquad [42]$$

If we use a Weitzenböck formula to relate the Laplacian $\nabla_A^*\nabla_A$ to $\partial\!\!\!/_A^*\partial\!\!\!/_A$ plus curvature terms, we find that $S$ satisfies

$$\int_M \left\{ |\partial\!\!\!/_A\psi|^2 + \tfrac{1}{2}|F^+ + \tfrac{1}{2}\bar{\psi}\Gamma\psi|^2 \right\} = \int_M \left\{ |\nabla_A\psi|^2 + \tfrac{1}{2}|F^+|^2 + \tfrac{1}{8}|\psi|^4 + \tfrac{1}{4}R|\psi|^2 \right\} \qquad [43]$$

$$= \int_M \left\{ |\nabla_A\psi|^2 + \tfrac{1}{4}|F|^2 + \tfrac{1}{8}|\psi|^4 + \tfrac{1}{4}R|\psi|^2 \right\} + \pi^2 c_1^2(L) \qquad [44]$$

where $R$ is the scalar curvature of $M$ and $c_1(L)$ is the Chern class of $L$.

We notice that the action now looks like one for monopoles. But now suppose that $R$ is positive and that the pair $(A, \psi)$ is a solution to the Seiberg–Witten equations, then the left-hand side (LHS) of this last expression is zero and all the integrands on the RHS are positive so the solution must obey $\psi = 0$ and $F^+ = 0$. A technical point is that if $M$ has $b_2^+ > 1$, then a perturbation of the metric can preserve the positivity of $R$ but perturb $F^+ = 0$ to be simply $F = 0$ rendering the connection $A$ flat. Hence, in these circumstances, the solution $(A, \psi)$ is the trivial one. This means that we have a new kind of vanishing theorem in four dimensions.

**Theorem** (Witten 1994). *No 4-manifold with $b_2^+ > 1$ and nontrivial solution to the Seiberg–Witten equations admits a metric of positive scalar curvature.*

Now, for technical reasons, we assume that the $q_{d,r}^M$ have the property that

$$q_{d,r+2}^M = 4q_{d,r}^M \qquad [45]$$

A simply connected $M$ with this property is called of *simple type*. We also define $\tilde{q}_d^M$ by writing

$$\tilde{q}_d^M = \begin{cases} q_{d,0}^M, & \text{if } d = (b_2^+ + 1) \bmod 2 \\ \tfrac{1}{2}q_{d,1}^M, & \text{if } d = b_2^+ \bmod 2 \end{cases} \qquad [46]$$

The generating function $G_M(\alpha)$ is now given by

$$G_M(\alpha) = \sum_{d=0}^{\infty} \frac{1}{d!}\tilde{q}_d^M(\alpha) \qquad [47]$$

According to Kronheimer and Mrowka (1994), $G_M(\alpha)$ can be expressed in terms of a *finite* number of classes (known as *basic classes*) $\kappa_i(\kappa_i \in H^2(M))$ with rational coefficients $a_i$ (the Seiberg–Witten invariants) resulting in the formula

$$G_M(\alpha) = \exp[\alpha \cdot \alpha/2] \sum_i a_i \exp[\kappa_i \cdot \alpha] \qquad [48]$$

Hence, for $M$ of simple type, the polynomial invariants are determined by a (finite) number of basic classes and the Seiberg–Witten invariants.

Returning now to the physics we find that the quantum field theory approach to the polynomial invariants relates them to properties of the moduli space for the Seiberg–Witten equations rather than to properties of the instanton moduli space $\mathcal{M}_k$.

The moduli space for the Seiberg–Witten equations, unlike the instanton case, is *compact* and generically has dimension

$$\frac{c_1^2(L) - 2\chi(M) - 3\sigma(M)}{4} \qquad [49]$$

$\chi(M)$ and $\sigma(M)$ being the Euler characteristic and signature of $M$, respectively. When

$$c_1^2(L) = 2\chi(M) + 3\sigma(M) \quad [50]$$

we get a zero-dimensional moduli space consisting of a finite collection of points

$$\{P_1, \ldots, P_N\} \quad [51]$$

Now each point $P_i$ has a sign $\epsilon_i = \mp 1$ associated with it coming from the sign of the determinant of elliptic operator whose index gave the dimension of the moduli space. The sum of these signs is an integer topological invariant denoted by $n_L$, that is,

$$n_L = \sum_{j=1}^{N} \epsilon_i \quad [52]$$

Returning now to our formula for $G_M(\alpha)$, one finds that

$$G_M(\alpha) = 2^{p(M)} \exp[\alpha \cdot \alpha/2] \sum_T n_L \exp[c_1(L) \cdot \alpha] \quad [53]$$

$$p(M) = 1 + \tfrac{1}{4}(7\chi(M) + 11\sigma(M)) \quad [54]$$

and the sum over $L$ on the RHS of the formula is over line bundles $L$ that satisfy

$$c_1^2(L) = 2\chi(M) + 3\sigma(M) \quad [55]$$

that is, it is a sum over $L$ with zero-dimensional Seiberg–Witten moduli spaces.

Comparison of the two formulas for $G_M(\alpha)$ – the first mathematical in origin and the second physical – allows one to identify the Seiberg–Witten invariants $a_i$ and the Kronheimer–Mrowka basic classes $\kappa_i$ as the $c_1(L)$'s.

The results described thus far are for *simply connected* 4-manifolds but this condition is not obligatory for and there is also a theory in the non-simply-connected case (Mariño and Moore 1999).

The physics underlying these topological results is of great importance since many of the ideas originate there. It is known that the computation of the Donaldson invariants there uses the fact that the $N = 2$ gauge theory is asymptotically free. This means that the ultraviolet limit being one of weak coupling is tractable. However, the less tractable infrared or strong-coupling limit would do just as well to calculate the Donaldson invariants since these latter are metric independent.

In Seiberg and Witten's work, this infrared behavior is actually determined and it is found that, in the strong-coupling infrared limit, the theory is equivalent to a *weakly coupled* theory of abelian fields and monopoles. There is also a duality between the original theory and the theory with monopoles which is expressed by the fact that the (abelian) gauge group of the monopole theory is the dual of the maximal torus of the group of the nonabelian theory.

We recall that the Yang–Mills gauge group in this discussion is SU(2). Seiberg and Witten's results mean the replacement of SU(2) instantons used to compute the Donaldson invariants by the counting of U(1) monopoles. This calculation of the non-abelian Donaldson data by abelian Seiberg–Witten data theory is much like the representation theory of a nonabelian Lie group $G$ where everything is determined by an abelian object: the maximal torus.

The theory considered by Seiberg and Witten possesses a collection of quantum vacua labeled by a complex parameter $u$ which turns out to parametrize a family of elliptic curves. A central part is played by a function $\tau(u)$ on which there is a modular action of $SL(2, \mathbf{Z})$. The successful determination of the infrared limit involves an electric–magnetic duality and the whole matter is of very considerable independent interest for quantum field theory, quark confinement, and string theory in general.

## Seiberg–Witten Theory and Exotic Structures on 4-Manifolds

We saw earlier that, when $\dim M \neq 4$, a manifold may possess a finite number of differentiable structures, $S^7$ having 28 distinct smooth structures. However, in dimension 4, Seiberg–Witten theory has been used to show that there are many 4-manifolds with a countable infinity of smooth structures. We just mention two: the $K3$ surface has infinitely many smooth structures as does the manifold $\mathbf{CP}^2 \# 5\overline{\mathbf{CP}}^2$. This is another instance of how dimension 4 differs from all other dimensions. This infinite variety of exotic smooth structures in four dimensions is also of great interest to physics.

An outstanding four-dimensional matter still is the *smooth Poincaré conjecture* which asks whether a smooth 4-manifold $M$ homotopic to $S^4$ is *diffeomorphic* to $S^4$? Such an $M$ is certainly *homeomorphic* to $S^4$ because this is the standard Poincaré conjecture proved by Freedman and, if the answer to this question is yes then $S^4$ would be an example of a 4-manifold with no exotic smooth structures. There is at present no consensus on the answer to this question.

## Exotic Structures on Open 4-Manifolds

If $M$ is an open manifold, that is, a noncompact manifold without boundary, and $M = \mathbf{R}^n$ then, for

$n \neq 4$, there is only one smooth structure; but for $n = 4$, there are exotic differentiable structures on $\boldsymbol{R}^4$. In fact, Gompf showed that there is a *continuum* of exotic differentiable structures that can be placed on $\boldsymbol{R}^4$.

### Symplectic and Kähler 4-Manifolds

Many 4-manifolds are symplectic, and symplectic manifolds are central in physics; there are many results obtained using Seiberg–Witten theory concerning the topology and geometry of symplectic manifolds. The exotic $K3$ structures referred to above are all symplectic and so there is no shortage of symplectic structures even within one homeomorphism class. Taubes obtained far-reaching new results for symplectic 4-manifolds including establishing an equivalence between the Seiberg–Witten invariants in the symplectic case and the Gromov invariants.

Kähler manifolds possess, simultaneously, compatible, Riemannian, symplectic and complex structures and, beginning with Witten's work, there are many results to be found for Kähler 4-manifolds using Seiberg–Witten techniques.

## 4-Manifolds with Boundary

There is a very important extension of the Donaldson–Seiberg–Witten theory to 4-manifolds $M$ with boundary $\partial M = N$. When $\partial M \neq \phi$, the Donaldson invariants are not numerical invariants but take values in HF($N$) where HF($N$) denotes what is called the Floer homology of the 3-manifold $N$. Topological quantum field theory is the ideal setting for this theory since it naturally treats manifolds with boundaries. The Floer homology groups HF($N$) act as Hilbert spaces for the quantum fields defined on the boundary. There is now a full interplay of 4-manifold theory and 3-manifold theory as well as Yang–Mills theory in three and four dimensions. This interplay is often realized by taking two 4-manifolds $M_1$ and $M_2$ with the same boundary $N$ and joining them along $N$ to obtain a *closed 4-manifold* $M$ so that

$$M = M_1 \cup_N M_2 \qquad [56]$$

Given a 3-manifold $N$, and an SU(2) connection $A$, Floer studied the critical points of the Chern–Simons function $f(A)$ defined by

$$f(A) = \frac{1}{8\pi^2}\int_N \operatorname{tr}\left(A \wedge \mathrm{d}A + \tfrac{2}{3}A \wedge A \wedge A\right) \qquad [57]$$

where $f(A)$ is regarded as a function on the infinite-dimensional space $A$ of connections. The function $f(A)$ changes by an integer under a gauge transformation and so descends to a single-valued gauge-invariant function on the space of gauge orbits $\mathcal{A}/\mathcal{G}$ if one considers $\exp(2\pi k i f(A))$ where $k \in \boldsymbol{Z}$ ($\mathcal{G}$ being the group of gauge transformations). Morse theory applied to this infinite-dimensional setting gives an infinite Morse index to each critical point, a pathology which is avoided by only defining the difference of the index between two critical points using spectral flow. The critical points correspond, via gradient flow and a consideration of the instanton equations

$$F = \pm * F \qquad [58]$$

on the 4-manifold $N \times \boldsymbol{R}$, to the flat connections on the 3-manifold $N$. The latter are identifiable as the set of (equivalence classes) of representations of the fundamental group $\pi_1(M)$ in the gauge group SU(2), that is, with

$$\operatorname{Hom}(\pi_1(N), \mathrm{SU}(2))/\mathrm{Ad}\ \mathrm{SU}(2) \qquad [59]$$

For the Seiberg–Witten formulation, let $\hat{A}$ denote a connection on the 3-manifold $N$ with curvature $F(\hat{A})$. Then the Chern–Simons function $f(A)$ is replaced by the abelian Chern–Simons function together with a quadratic fermion term resulting in the function $f^{\mathrm{SW}}(\hat{A})$, defined by

$$f^{\mathrm{SW}}(\hat{A}) = \int_N \left\{\phi \not{D}_{\hat{A}}\phi + \hat{A} \wedge F(\hat{A})\right\} \qquad [60]$$

where $\not{D}_{\hat{A}}$ denotes the self-adjoint Dirac operator in three dimensions acting on a spinor $\phi$ on $N$; because of the presence of the Chern–Simons function $f^{\mathrm{SW}}(\hat{A})$ is only defined up to a multiple of $8\pi^2$ in a manner similar to the case for $f(A)$. Gradient flow together with the Seiberg–Witten equations on the 4-manifold $N \times \boldsymbol{R}$ result in critical points corresponding to the solutions to

$$\not{D}_{\hat{A}}\phi = 0, \qquad F(\hat{A}) = -\tfrac{1}{2}\phi\Gamma\phi \qquad [61]$$

which is a three-dimensional version of the Seiberg–Witten equations.

The critical point theory of these two functions $f(A)$ and $f^{\mathrm{SW}}(\hat{A})$ permit the construction of the instanton Floer homology groups $\mathrm{HF}^{\mathrm{inst}}(N)$ and $\mathrm{HF}^{\mathrm{SW}}(N)$, respectively. In fact, there are several kinds of Floer homology: Lagrangian Floer homology, instanton Floer homology, Heegard–Floer homology, Seiberg–Witten–Floer homology and conjectures concerning their relations to one another.

There are still many unanswered questions of joint interest to mathematicians and physicists in the entire area of 4-manifold theory.

*See also:* Electric–Magnetic Duality; Gauge Theoretic Invariants of 4-Manifolds; Floer Homology; Topological Quantum Field Theory: Overview.

## Further Reading


Atiyah MF (1989) Topological quantum field theories. *Institut des Hautes Etudes Scientifiques Publications Mathematiques* 68: 175–186.
Atiyah MF (1995) Floer homology. In: Hofer H, Weinstein A, Taubes C, and Zehnder E (eds.) *Floer Memorial Volume*. Boston: Birkhäuser.
Donaldson SK (1984) An application of gauge theory to four dimensional topology. *Journal of Differential Geometry* 96: 387–407.
Donaldson SK (1996) The Seiberg–Witten equations and 4-manifold topology. *Bulletin of the American Mathematical Society* 33: 45–70.
Donaldson SK (2002) *Floer Homology Groups in Yang–Mills Theory*. Cambridge: Cambridge University Press.
Donaldson SK and Kronheimer PB (1990) *The Geometry of Four Manifolds*. Oxford: Oxford University Press.
Fintushel R and Stern RJ (1998) Knots, links and 4-manifolds. *Inventiones Mathematicae* 134: 363–400.
Fintushel R and Stern RJ (2004) Double node neighborhoods and families of simply connected 4-manifolds with $b^+ = 1$, arXiv: math.GT/0412126.
Freedman MH (1982) The topology of 4-dimensional manifolds. *Journal of Differential Geometry* 17: 357–453.
Gadgil S (2004) Open manifolds, Ozsvath–Szabo invariants and exotic $\boldsymbol{R}^4$s, arXiv:math.GT/0408379.
Gompf R (1985) An infinite set of exotic $\boldsymbol{R}^4$'s. *Journal of Differential Geometry* 21: 283–300.
Gromov M (1985) Pseudo-holomorphic curves in symplectic manifolds. *Inventiones Math*. 82: 307–347.
Kronheimer PB and Mrowka TS (1994) Recurrence relations and asymptotics for four manifold invariants. *Bulletin of the American Mathematical Society* 30: 215–221.
Mandelbaum R (1980) Four-dimensional topology: an introduction. *Bulletin of the American Mathematical Society* 2: 1–159.
Mariño M and Moore G (1999) Donaldson invariants for non-simply connected manifolds. *Communications in Mathematical Physics* 203: 249–267.
Morgan J and Mrowka T (1992) A note on Donaldson's polynomial invariants. *International Mathematics Research Notices* 10: 223–230.
Nash C (1992) *Differential Topology and Quantum Field Theory*. London–San Diego: Academic Press.
Park J, Stipsicz AI, and Szabo Z (2004) Exotic smooth structures on $\boldsymbol{CP}^2\#5\overline{\boldsymbol{CP}}^2$, arXiv:math.GT/041221.
Rohlin VA (1952) New results in the theory of 4 dimensional manifolds. *Doklady Akademii Nauk SSSR* 84: 221–224.
Seiberg N and Witten E (1994a) Electric–magnetic duality, monopole condensation, and confinement in $N = 2$ supersymmetric Yang–Mills theory. *Nuclear Physics B* 426: 19–52.
Seiberg N and Witten E (1994b) Electric-magnetic duality, monopole condensation, and confinement in $N = 2$ supersymmetric Yang–Mills theory – erratum. *Nuclear Physics B* 430: 485–486.
Seiberg N and Witten E (1994c) Monopoles, duality and chiral symmetry breaking in $N = 2$ supersymmetric QCD. *Nuclear Physics B* 431: 484–550.
Taubes CH (1995a) More constraints on symplectic forms from Seiberg–Witten invariants. *Mathematical Research Letters* 2: 9–13.
Taubes CH (1995b) The Seiberg–Witten and Gromov invariants. *Mathematical Research Letters* 2: 221–238.
Witten E (1988) Topological quantum field theory. *Communications in Mathematical Physics* 117: 353–386.
Witten E (1994) Monopoles and four manifolds. *Mathematical Research Letters* 1: 769–796.


# Fractal Dimensions in Dynamics

**V Županović and D Žubrinić**, University of Zagreb, Zagreb, Croatia



## Introduction

Since the 1970s, dimension theory for dynamics has evolved into an independent field of mathematics. Its main goal is to measure complexity of invariant sets and measures using fractal dimensions. The history of fractal dimensions is closely related to the names of H Minkowski (Minkowski content, 1903), H Hausdorff (Hausdorff dimension, 1919), G Bouligand (Bouligand dimension, 1928), L S Pontryagin and L G Schnirelmann (metric order, 1932), P Moran (Moran geometric constructions, 1946), A S Besicovitch and S J Taylor (Besicovitch–Taylor index, 1954), A Rényi (Rényi spectrum for dimensions, 1957), A N Kolmogorov and V M Tihomirov (metric dimension, Kolmogorov complexity, 1959), Ya G Sinai, D Ruelle, R Bowen (thermodynamic formalism, Bowen's equation, 1972, 1973, 1979), B Mandelbrot (fractals and multifractals, 1974), J L Kaplan and J A Yorke (Lyapunov dimension, 1979), J E Hutchinson (fractals and self-similarity, 1981), C Tricot, D Sullivan (packing dimension, 1982, 1984), H G E Hentschel and I Procaccia (Hentschel–Procaccia spectrum for dimensions, 1983), Ya Pesin (Carathéodory–Pesin dimension, 1988), M Lapidus and M van Frankenhuysen (complex dimensions for fractal strings, 2000), etc. Fractal dimensions enable us to have a better insight into the dynamics appearing in various problems in physics, engineering, chemistry, medicine, geology, meteorology, ecology, economics, computer science, image processing, and, of course, in many branches of mathematics. Concentrating on box and Hausdorff dimensions only, we describe basic methods of fractal analysis in dynamics, sketch their applications, and indicate some trends in this rapidly growing field.

## Fractal Dimensions

### Box Dimensions

Let $A$ be a bounded set in $\mathbb{R}^N$, and let $d(x,A)$ be Euclidean distance from $x$ to $A$. The Minkowski sausage of radius $\varepsilon$ around $A$ (a term coined by B Mandelbrot) is defined as $\varepsilon$-neighborhood of $A$, that is, $A_\varepsilon := \{y \in \mathbb{R}^N\colon d(y,A) < \varepsilon\}$. By the upper s-dimensional Minkowski content of $A$, $s \geq 0$, we mean

$$\mathcal{M}^{*s}(A) := \overline{\lim_{\varepsilon\to 0}} \frac{|A_\varepsilon|}{\varepsilon^{N-s}} \in [0,\infty]$$

Here $|\cdot|$ denotes $N$-dimensional Lebesgue measure. The corresponding upper box dimension is defined by

$$\overline{\dim}_B A := \inf\{s \geq 0\colon\ \mathcal{M}^{*s}(A) = 0\}$$

The lower s-dimensional Minkowski content $\mathcal{M}_*^s(A)$ and the corresponding lower box dimension $\underline{\dim}_B A$ are defined analogously. The name of box dimension stems from the following: if we have an $\varepsilon$-grid in $\mathbb{R}^N$ composed of closed $N$-dimensional boxes with side $\varepsilon$, and if $N(A,\varepsilon)$ is the number of boxes of the grid intersecting $A$, then

$$\overline{\dim}_B A = \overline{\lim_{\varepsilon\to 0}} \frac{\log N(A,\varepsilon)}{\log(1/\varepsilon)}$$

and analogously for $\underline{\dim}_B A$. It suffices to take any geometric subsequence $\varepsilon_k = b^{-k}$ in the limit, where $b > 1$ (H Furstenberg, 1970). There are many other names for the upper box dimension appearing in the literature, like the Cantor–Minkowski order, Minkowski dimension, Bouligand dimension, Borel logarithmic rarefaction, Besicovitch–Taylor index, entropy dimension, Kolmogorov dimension, fractal dimension, capacity dimension, and limit capacity. If $A$ is such that $\underline{\dim}_B A = \overline{\dim}_B A$, the common value is denoted by $d := \dim_B A$, and we call it the box dimension of $A$. If, in addition to this, both $\mathcal{M}_*^d(A)$ and $\mathcal{M}^{*d}(A)$ are in $(0,\infty)$, we say that $A$ is Minkowski nondegenerate. If, moreover, $\mathcal{M}_*^d(A) = \mathcal{M}^{*d}(A) =: \mathcal{M}^d(A) \in (0,\infty)$, then $A$ is said to be Minkowski measurable.

Assume that $A$ is such that $d := \dim_B A$ and $\mathcal{M}^d(A)$ exist. Then the value of $\mathcal{M}^d(A)^{-1}$ is called the lacunarity of $A$ (B Mandelbrot, 1982). A bounded set $A \subset \mathbb{R}^N$ is said to be porous (A Denjoy, 1920) if there exist $\alpha > 0$ and $\delta > 0$ such that for every $x \in A$ and $r \in (0,\delta)$ there is $y \in \mathbb{R}^N$ such that the open ball $B_{\alpha r}(y)$ is contained in $B_r(x) \setminus A$. If $A$ is porous then it is easy to see that $\overline{\dim}_B A < N$ (O Martio and M Vuorinen, 1987, A Salli, 1991).

We proceed with two examples. Let $A := C^{(a)}$, $a \in (0,1/2)$, be the Cantor set obtained from $[0,1]$ by consecutive deletion of $2^k$ middle open intervals of length $a^k(1-2a)$ in step $k \in \mathbb{N} \cup \{0\}$. Then $\dim_B A = (\log 2)/(\log(1/a))$ (G Bouligand, 1928), and $A$ is nondegenerate, but not Minkowski measurable (Lapidus and Pomerance, 1993). For the spiral $\Gamma$ of focus type defined by $r = m\varphi^{-\alpha}$ in polar coordinates, where $\alpha \in (0,1)$ and $m > 0$ are fixed, $\varphi \geq \varphi_1 > 0$, we have $\dim_B \Gamma = 2/(1+\alpha)$ (Y Dupain, M Mendés-France, C Tricot, 1983). It is Minkowski measurable (Žubrinić and Županović, 2005), and the larger $m$, the smaller the lacunarity; see Figure 1.

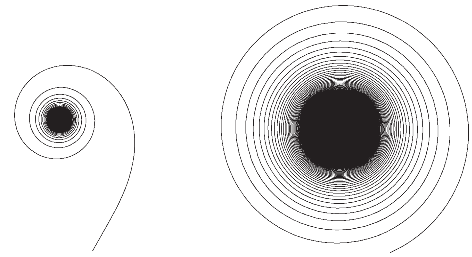


**Figure 1** Spirals of equal box dimensions (4/3) and different lacunarities (0.43 and 0.05).

### Hausdorff Dimension

For a given subset $A$ of $\mathbb{R}^N$ (not necessarily bounded) and $s \geq 0$ we define $\mathcal{H}^s(A) := \lim_{\varepsilon\to 0} \inf\{\sum_{i=1}^{\infty} r_i^s\} \in [0,\infty]$, where the infimum is taken over all finite or countable coverings of $A$ by open balls of radii $r_i \leq \varepsilon$. The value of $\mathcal{H}^s(A)$ is called $s$-dimensional Hausdorff outer measure of $A$. The Hausdorff dimension of $A$, sometimes called the Hausdorff–Besicovitch dimension, is defined by

$$\dim_H A := \inf\{s \geq 0\colon\ \mathcal{H}^s(A) = 0\}$$

If $A$ is bounded then $\dim_H A \leq \underline{\dim}_B A \leq \overline{\dim}_B A \leq N$.

We say that $A$ is Hausdorff nondegenerate (or $d$-set) if $\mathcal{H}^d(A) \in (0,\infty)$ for some $d \geq 0$. Cantor sets share this property, and $\dim_H C^{(a)} = (\log 2)/(\log(1/a))$, where $a \in (0,1/2)$ (Hausdorff, 1919).

### Gauge Functions

The notions of Minkowski contents and Hausdorff measure can be generalized using gauge functions $h : [0,\varepsilon_0) \to \mathbb{R}$ that are assumed to be continuous, increasing, and $h(0) = 0$. For example,

$$\mathcal{M}^{*h}(A) := \overline{\lim_{\varepsilon\to 0}} \frac{|A_\varepsilon|}{\varepsilon^N} h(\varepsilon)$$

and similarly for $\mathcal{M}_*^h(A)$ (M Lapidus and C He, 1997), while for $\mathcal{H}^h(A)$ it suffices to change $r_i^s$ with $h(r_i)$ in the above definition of the Hausdorff outer measure (Besicovitch, 1934). Gauge functions are used for sets that are Minkowski or Hausdorff

degenerate. The aim, if possible, is to find an explicit gauge function so that the corresponding generalized Minkowski contents or Hausdorff measure of $A$ be nondegenerate.

## Methods of Fractal Analysis in Dynamics

### Thermodynamic Formalism

Thermodynamic formalism has been developed by Sinai (1972), Ruelle (1973), and Bowen (1975), using methods of statistical mechanics in order to study dynamics and to find dimensions of various fractal sets. We first describe a "dictionary" for explicit geometric constructions of Cantor-like sets. Let $X_p$ be the set of all sequences $\boldsymbol{i}=(i_1,i_2,\ldots)$ of elements $i_k$ from a given set of $p$ symbols, say $\{1,2,\ldots,p\}$. We endow $X_p$ with the metric $d(\boldsymbol{i},\boldsymbol{j}):=\sum_k 2^{-k}|i_k-j_k|$ and introduce the one-sided shift operator (or left shift) $\sigma: X_p\to X_p$ defined by $(\sigma(\boldsymbol{i}))_n=i_{n+1}$, that is, $\sigma(i_1,i_2,i_3,\ldots)=(i_2,i_3,i_4,\ldots)$. A set $Q\subseteq X_p$ is called the symbolic dynamics if it is compact and $\sigma$-invariant, that is, $\sigma(Q)\subseteq Q$. Hence, $(Q,\sigma)$ is a symbolic dynamical system. Denote $\boldsymbol{i}[n]:=(i_1,\ldots,i_n)$. Given a continuous function $\varphi: Q\to\mathbb{R}$, let us define the topological pressure of $\varphi$ with respect to $\sigma$ by

$$P(\varphi):=\lim_{n\to\infty}\frac{1}{n}\log\sum_{\{\boldsymbol{i}[n]:\boldsymbol{i}\in Q\}}E(\boldsymbol{i}[n])$$

$$E(\boldsymbol{i}[n]):=\exp\left(\sup_{\{\boldsymbol{j}\in Q:\ \boldsymbol{j}[n]=\boldsymbol{i}[n]\}}\sum_{k=0}^{n-1}\varphi(\sigma^k(\boldsymbol{j}))\right)$$

The topological entropy of $\sigma|Q$ is defined by $h(\sigma|Q):=P(0)$, that is,

$$h(\sigma|Q)=\lim_{n\to\infty}\frac{1}{n}\log\#\{\boldsymbol{i}[n]:\ \boldsymbol{i}\in Q\}$$

where # denotes the cardinal number of a set. The above function $\varphi_n:=\sum_{k=0}^{n-1}\varphi\circ\sigma^k$ has the property $\varphi_{n+m}=\varphi_n+\varphi_m\circ\sigma^n$, and therefore we speak about additive thermodynamic formalism. Topological pressure was introduced by D Ruelle (1973) and extended by P Walters (1976). Bowen's equation (1979) has a very important role in the computation of the Hausdorff dimension of various sets. For the unknown $s\in\mathbb{R}$, and with a suitably chosen function $\varphi$, this equation reads

$$P(s\varphi)=0$$

### Geometric Constructions

A geometric construction $(Q,\Delta)$ in $\mathbb{R}^m$ indexed by symbolic dynamics $Q$ is a family $\Delta$ of compact sets $\Delta_{\boldsymbol{i}[n]}\subset\mathbb{R}^m, \boldsymbol{i}\in Q, n\in\mathbb{N}$, such that $\operatorname{diam}\Delta_{\boldsymbol{i}[n]}\to 0$ as $n\to\infty$, $\Delta_{\boldsymbol{i}[n+1]}\subseteq\Delta_{\boldsymbol{i}[n]}$, $\Delta_{\boldsymbol{i}[n]}=\overline{\operatorname{int}\Delta_{\boldsymbol{i}[n]}}$ for every $\boldsymbol{i}\in Q$ and all $n$, and $\operatorname{int}\Delta_{\boldsymbol{i}[n]}\cap\operatorname{int}\Delta_{\boldsymbol{j}[n]}=\emptyset$ whenever $\boldsymbol{i}[n]\neq\boldsymbol{j}[n]$ (Moran's open set condition). This family induces the Cantor-like set

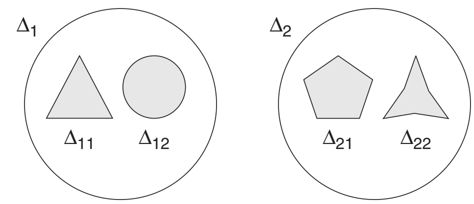



**Figure 2** Cantor-like set.

$$F:=\bigcap_{n=1}^{\infty}\left(\bigcup_{\boldsymbol{i}\in Q}\Delta_{\boldsymbol{i}[n]}\right)$$

(see Figure 2). The mapping $h: Q\to F$ defined by $h(\boldsymbol{i}):=\cap_{n=1}^{\infty}\Delta_{\boldsymbol{i}[n]}$ is called the coding map of $F$. The above geometric construction includes well-known iterated function systems of similarities as a special case. If $\lambda_1,\ldots,\lambda_p$ are given numbers in (0, 1), and $\Delta_{\boldsymbol{i}[n]}$ are balls of radii $r_{\boldsymbol{i}[n]}:=\lambda_{i_1}\ldots\lambda_{i_n}$, then $s:=\dim_H F$ is the unique solution of Bowen's equation $P(s\varphi)=0$, where $\varphi$ is defined by $\varphi(\boldsymbol{i}):=\log\lambda_{i_1}$ (Ya Pesin and H Weiss, 1996). In this case Bowen's equation is equivalent to Moran's equation (1946),

$$\sum_{k=1}^{p}{\lambda_k}^s=1$$

This result has been generalized by L Barreira (1996) using the Carathéodory–Pesin construction (1988). Let us illustrate Barreira's theory of nonadditive thermodynamic formalism with a special case. Assume that $(Q,\Delta)$ is a geometric construction for which the sets $\Delta_{\boldsymbol{i}[n]}$ are balls, and let there exist $\delta>0$ such that $r_{\boldsymbol{i}[n+1]}\geq\delta\cdot r_{\boldsymbol{i}[n]}$ and $r_{\boldsymbol{i}[n+m]}\leq r_{\boldsymbol{i}[n]}r_{\sigma^n(\boldsymbol{i})[m]}$ for all $\boldsymbol{i}\in Q, n, m\in\mathbb{N}$. Then $\dim_H F=\dim_B F=s$, where $s$ is the unique real number such that

$$\lim_{n\to\infty}\frac{1}{n}\log\sum_{\{\boldsymbol{i}[n]:\boldsymbol{i}\in Q\}}{r_{\boldsymbol{i}[n]}}^s=0 \qquad [1]$$

This is a special case of Barreira's extension of Bowen's equation to nonadditive thermodynamic formalism. Moran's equation can be deduced from [1] by defining $r_{\boldsymbol{i}[n]}:=\lambda_{i_1}\ldots\lambda_{i_n}$, where $\boldsymbol{i}=(i_1,i_2,\ldots)$, and $\lambda_1,\ldots,\lambda_p\in(0,1)$ are given numbers. Pesin and Weiss (1996) showed that Moran's open set condition can be weakened so that partial intersections of interiors of pairs of basic sets in the family $\Delta$ are allowed. Thermodynamic formalism has been used to study the Hausdorff dimension of Julia sets

(Ruelle, 1982), horseshoes (H McCluskey and A Manning, 1983), etc.

An important example of symbolic dynamics is the topological Markov chain $X_A$ generated by a $p \times p$ matrix $A$ with entries $a_{ij} \in \{0, 1\}$:

$$X_A := \{\boldsymbol{i} = (i_1, i_2, \ldots) \in X_p\colon\ a_{i_k i_{k+1}} = 1 \text{ for all } k \in \mathbb{N}\}$$

It is a compact, $\sigma$-invariant subset of $X_p$. The map $\sigma | X_A$ is called the subshift of finite type (Bowen, 1975). A construction of Cantor-like set $F$ using dynamics $Q = X_p$ is called a simple geometric construction, while a geometric construction is said to be a Markov geometric construction if $Q = X_A$. If $F$ is obtained by a Markov geometric construction such that all $\Delta_{\boldsymbol{i}[n]}$ are balls of radii $r_{\boldsymbol{i}[n]} := \lambda_{i_1} \ldots \lambda_{i_n}$, where $\lambda_{i_j} \in (0,1), i_j \in \{1, \ldots, p\}$, then $\dim_{\mathrm{B}} F = \dim_{\mathrm{H}} F = s$, where $s$ is the unique solution of equation $\rho(AM_s) = 1$. Here $M_s := \mathrm{diag}({\lambda_1}^s, \ldots, {\lambda_p}^s)$ and $\rho(AM_s)$ is the spectral radius of the matrix $AM_s$. This and more general results have been obtained by Pesin and Weiss (1996).

Any Cantor-like set $F$ obtained via iterated function system of similarities satisfying Moran's open set condition is Hausdorff nondegenerate (Moran, 1946). If $F$ is of nonlattice type, that is, the set $\{\log \lambda_1, \ldots, \log \lambda_p\}$ is not contained in $r \cdot \mathbb{Z}$ for any $r > 0$, then $F$ is Minkowski measurable (D Gatzouras, 1999).

### Hyperbolic Measures

Let $X$ be a complete metric space and assume that $f : X \to X$ is continuous. Let $\mu$ be an $f$-invariant Borel probability measure on $X$ (i.e., $\mu(f^{-1}(A)) = \mu(A)$ for measurable sets $A$) with a compact support. The Hausdorff dimension of $\mu$, and the lower and upper box dimensions of $\mu$ (L-S Young, 1982) are defined by

$$\dim_{\mathrm{H}} \mu := \inf\{\dim_{\mathrm{H}} Z : Z \subseteq X, \mu(Z) = 1\}$$
$$\underline{\dim}_{\mathrm{B}} \mu := \lim_{\delta \to 0} \inf\{\underline{\dim}_{\mathrm{B}} Z : Z \subseteq X, \mu(Z) \geq 1 - \delta\}$$
$$\overline{\dim}_{\mathrm{B}} \mu := \lim_{\delta \to 0} \inf\{\overline{\dim}_{\mathrm{B}} Z : Z \subseteq X, \mu(Z) \geq 1 - \delta\}$$

It is natural to introduce the lower and upper pointwise dimensions of $\mu$ at $x \in X$ by

$$\underline{d}_\mu(x) := \varliminf_{r \to 0} \frac{\log \mu(B_r(x))}{\log r}$$

and similarly $\overline{d}_\mu(x)$. It has been shown by Young (1982) that if $X$ has finite topological dimension and if $\mu$ is exact dimensional, that is, $\underline{d}_\mu(x) = \overline{d}_\mu(x) =: d$ for $\mu$-a.e. $x \in X$, then

$$\dim_{\mathrm{H}} \mu = \dim_{\mathrm{B}} \mu = d$$

She also proved that hyperbolic measures (ergodic measures with nonzero Lyapunov exponents), invariant under a $C^{1+\alpha}$-diffeomorphism, $\alpha > 0$, are exact dimensional. F Ledrappier (1986) derived exact dimensionality for hyperbolic Bowen–Ruelle–Sinai measures. This result was extended by Ya Pesin and Ch Yue (1996) to hyperbolic measures with semilocal product structure. J-P Eckmann and D Ruelle (1985) conjectured that the exact dimensionality holds for general hyperbolic measures, and this was proved by Barreira, Pesin, and Schmeling (1996). More precisely, if $f$ is a $C^{1+\alpha}$-diffeomorphism on a smooth Riemann manifold $X$ without boundary, and if $\mu$ is $f$-invariant, compactly supported Borel probability measure, then its hyperbolicity implies that

$$\underline{d}_\mu(x) = \overline{d}_\mu(x) = d^s_\mu(x) + d^u_\mu(x)$$

for $\mu$-a.e. $x \in X$, where $d^s_\mu(x)$ and $d^u_\mu(x)$ are stable and unstable pointwise dimensions of $\mu$ at $x$ introduced by Ledrappier and Young (1985).

### Multifractal Analysis of Functions and Measures

Invariant sets of many dynamical systems are not self-similar. Roughly speaking, the aim of multifractal analysis is to make a decomposition of the invariant set with respect to desired fractal properties and then to study a fractal dimension of each set of the decomposition. Some dynamical systems have invariant sets equal to graphs of Hölderian functions $f : \mathbb{R}^N \to \mathbb{R}$, so that wavelet methods can be used. One of the goals of multifractal analysis of functions is to study the spectrum of singularities of $f$ defined by

$$d_f(\alpha) := \dim_{\mathrm{H}} H_\alpha(f)$$

introduced by U Frisch and G Parisi (1985) in the context of fully developed turbulence. Here $H_\alpha(f)$ is the set of points at which the corresponding pointwise Hölder exponent of $f$ is equal to $\alpha \geq 0$. If the function $f$ is self-similar then $d_f(\alpha)$ is real analytic and strictly concave (first increasing and then decreasing) on an explicit interval $(\underline{a}, \overline{a})$ (S Jaffard, 1997). It is natural to consider the set $C_{\alpha,\beta}(f)$ of points $x_0$ called chirps of order $(\alpha, \beta)$ (Y Meyer 1996), at which $f$ behaves roughly like $|x - x_0|^\alpha \sin(1/|x - x_0|^\beta), \beta > 0$. The function $D_f(\alpha, \beta) := \dim_H C_{\alpha,\beta}(f)$ is called the chirp spectrum of $f$ (S Jaffard 2000). Wavelet methods have found applications in the study of evolution equations and in modeling and detection of chirps in turbulent flows (S Jaffard, Y Meyer, R D Robert, 2001).

Basic ideas of multifractal analysis have been introduced by physicists T Halsey, M H Jensen,

L P Kadanoff, I Procaccia, and B I Shraiman (1988). In applications it often deals with an invariant ergodic probability measure associated with the dynamical system considered. Multifractal analysis of a Borel finite measure $\mu$ defined on $\mathbb{R}^N$ consists in the study of the function

$$d_\mu(\alpha) := \dim_{\mathrm{H}} K_\alpha(\mu), \quad \alpha \geq 0$$

called the spectrum of pointwise dimensions of $\mu$. Here $K_\alpha(\mu)$ is the set of points where the pointwise dimension of $\mu$ is equal to $\alpha$:

$$K_\alpha(\mu) := \{x \in \mathbb{R}^N \colon \; \underline{d}_\mu(x) = \overline{d}_\mu(x) = \alpha\}$$

It is also of interest to study the Hausdorff dimension of irregular set $K(\mu) := \{x \in \mathbb{R}^N \colon \underline{d}_\mu(x) < \overline{d}_\mu(x)\}$. These sets are pairwise disjoint and constitute a multifractal decomposition of $\mathbb{R}^N$, that is,

$$\mathbb{R}^N = K(\mu) \cup (\cup_{\alpha \in \mathbb{R}} K_\alpha(\mu))$$

The function $d_\mu(\alpha)$ provides an important information about the complexity of multifractal decomposition. In many situations, there is an open interval $(\underline{\alpha}, \overline{\alpha})$ on which the function $d_\mu(\alpha)$ is analytic and strictly concave (first increasing and then decreasing), and equal to the Legendre transform of an explicit convex function. We thus obtain an uncountable family of sets $K_\alpha(\mu)$ with positive Hausdorff dimension, which shows enormous complexity of the multifractal decomposition of $\mathbb{R}^N$. These and related questions have been studied by L Olsen (1995), K Falconer (1996), Pesin and Weiss (1996), Barreira and Schmeling (2000), and many other authors.

### Local Lyapunov Dimension

Let $\Omega$ be an open set in $\mathbb{R}^N$ and let $f : \Omega \to \mathbb{R}^N$ be a $C^1$-map. To any fixed $x \in \Omega$ we assign $N$ singular values $a_1 \geq a_2 \geq \cdots \geq a_N \geq 0$ of $f$, defined as square roots of eigenvalues of the matrix $f'(x)^\top \cdot f'(x)$, where $f'(x)$ is the Jacobian of $f$ at $x$, and $f'(x)^\top$ its transpose. The local Lyapunov dimension of $f$ at $x$ is defined by

$$\dim_{\mathrm{L}}(f, x) := j + s$$

where $j$ is the largest integer in $[0, N]$ such that $a_1 \cdots a_j \geq 1$ (if there is no such $j$ we let $j = 0$), and $s \in [0, 1)$ is the unique solution of $a_1 \cdots a_j a_{j+1}^s = 1$ (except for $j = N$, when we define $s = 0$). This definition, due to B R Hunt (1996), is close to that of Kaplan and Yorke (1979). The Jacobian $f'(x)$ contracts $k$-dimensional volumes (that is, $a_1 \cdots a_k < 1$) if and only if $\dim_{\mathrm{L}}(f, x) < k$. In this case, we say that $f$ is $k$-contracting at $x$. Furthermore, the function $x \mapsto \dim_{\mathrm{L}}(f, x)$ is upper-semicontinuous, so that for any compact subset $A$ of $\Omega$ the Lyapunov dimension of $f$ on A,

$$\dim_{\mathrm{L}}(f, A) := \max_{x \in A} \dim_{\mathrm{L}}(f, x)$$

is well defined. Yu S Ilyashenko conjectured that if $f$ locally contracts $k$-dimensional volumes then the upper box dimension of any compact invariant set is $< k$. Hunt (1996) proved that if $A$ is a compact, strictly invariant set of $f$ (i.e., $f(A) = A$) then

$$\overline{\dim}_{\mathrm{B}} A \leq \dim_{\mathrm{L}}(f, A) \qquad [2]$$

This is an improvement of $\dim_H A \leq \dim_{\mathrm{L}}(f, A)$ obtained by A Douady and J Oesterlé (1980), and independently by Ilyashenko (1982). M A Blinchevskaya and Yu S Ilyashenko (1999) proved that if $A$ is any attractor of a smooth map in a Hilbert space that contracts $k$-dimensional volumes then $\overline{\dim}_{\mathrm{B}} A \leq k$. See [3] below.

A continuous variant of this method is used in order to obtain estimates of fractal dimensions of global attractors of dynamical systems $(X, S)$ on a Hilbert space $X$. Here $S(t), t \geq 0$, is a semigroup of continuous operators on $X$, that is, $S(t+s) = S(t)S(s)$ and $S(0) = I$. A set $A$ in $X$ is called a global attractor of dynamical system if it is compact, attracting (i.e., for any bounded set $B$ and $\varepsilon > 0$ there exists $t_0$ such that for $t \geq t_0$ we have $S(t)B \subseteq A_\varepsilon$), and $A$ is strictly invariant (i.e., $S(t)A = A$ for all $t \geq 0$).

## Applications in Dynamics

### Logistic Map

M Feigenbaum, a mathematical physicist, introduced and studied the dynamics of the logistic map $f_\lambda : [0,1] \to [0,1], f_\lambda(x) := \lambda x(1-x), \lambda \in (0, 4]$. Taking $\lambda = \lambda_\infty \approx 3.570$ the corresponding invariant set $A \subset [0, 1]$ (i.e., $S_1(A) \cup S_2(A) = A$, where $S_i$ are two branches of $f_\lambda^{-1}$) has both Hausdorff and box dimensions equal to $\approx 0.538$ (P Grassberger 1981, P Grassberger and I Procaccia, 1983). The set $A$ has Cantor-like structure, but is not self-similar. Its multifractal properties have been studied by U Frisch, K Khanin, and T Matsumoto (2004).

### Smale Horseshoe

In the early 1960s S Smale defined his famous horseshoe map and showed that it has a strange invariant set resulting in chaotic dynamics. The notion of strange attractor was introduced in 1971 by Ruelle and Takens in their study of turbulence. Let $S$ be a square in the plane and let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be a map transforming $S$ as indicated in Figure 3, such that on both components of $S \cap f^{-1}(S)$ the map $f$ is

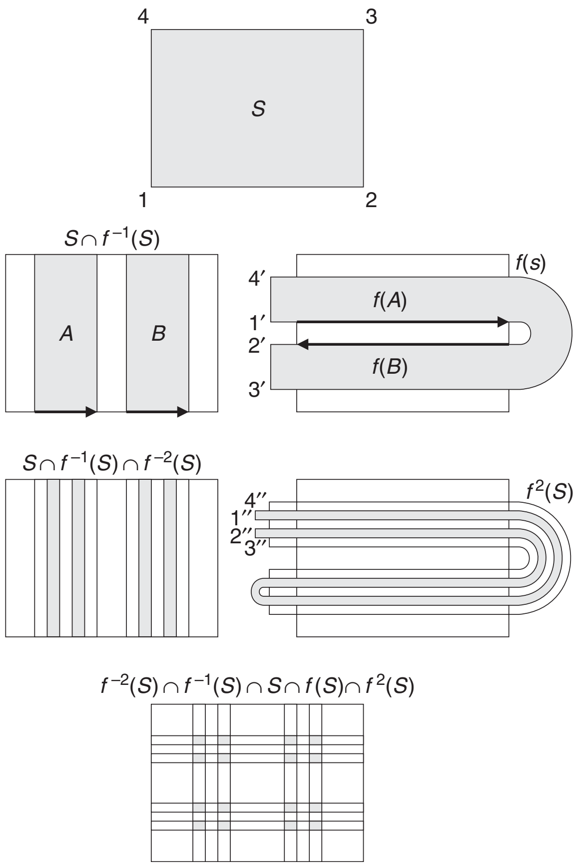



**Figure 3** The Smale horseshoe.

affine and preserves both horizontal and vertical directions, and such that points 1, 2, 3, and 4 are mapped to $1', 2', 3'$, and $4'$. Iterating $f$ we get backward invariant set $\Lambda_- := \cap_{j=0}^{\infty} f^{-j}(S)$, forward invariant set $\Lambda_+ := \cap_{j=0}^{\infty} f^{j}(S)$, and invariant set (horseshoe) $\Lambda_f := \Lambda_+ \cap \Lambda_-$. These sets have the Cantor set structure. More precisely, assuming that the contraction parameter of $f$ in vertical direction is $a \in (0, 1/2)$, and the expansion parameter in horizontal direction is $b > 2$, then $\Lambda_+ = [0,1] \times C^{(a)}$, where $C^{(a)}$ is the Cantor set, $\Lambda_- = C^{(1/b)} \times [0,1]$, and $\Lambda_f = C^{(1/b)} \times C^{(a)}$, so that $\dim_B \Lambda_+ = \dim_H \Lambda_+ = 1 + (\log 2)/(\log(1/a))$ and

$$\dim_B \Lambda_f = \dim_H \Lambda_f = \frac{\log 2}{\log b} + \frac{\log 2}{\log(1/a)}$$

This is a special case of a general result about horseshoes in $\mathbb{R}^2$ (not necessarily affine), due to McCluskey and Manning (1983), stated in terms of the pressure function. Analogous result as above can be obtained for Smale solenoids. In $\mathbb{R}^3$ it is possible to construct affine horseshoes $\Lambda_f$ such that $\dim_H \Lambda_f < \dim_B \Lambda_f$ (M Pollicott and H Weiss, 1994).

Smale discovered a connection between homoclinic orbits and the horseshoe map. It has been noticed that fractal dimensions have important role in the study of homoclinic bifurcations of nonconservative dynamical systems. Since the 1970s the relationship between invariants of hyperbolic sets and the typical dynamics appearing in the unfolding of a homoclinic tangency by a parametrized family of surface diffeomorphisms has been studied by J Newhouse, J Palis, F Takens, J-C Yoccoz, C G Moreira and M Viana. The main result is that if the Hausdorff dimension of the hyperbolic set involved in the tangency is $<1$ then the parameter set where the hyperbolicity prevails has full Lebesgue density. If the Hausdorff dimension is $>1$, then hyperbolicity is not prevalent. This result and its proof were inspired by previous work of J M Marstrand (1954) about arithmetic differences of Cantor sets on the real line. According to the result by Moreira, Palis, and Viana (2001) the paradigm "hyperbolicity prevails if and only if the Hausdorff dimension is $<1$" extends to homoclinic bifurcations in any dimension.

Using methods of thermodynamic formalism McCluskey and Manning (1983) proved that if $f$ is the above horseshoe map, then there exists a $C^1$-neighborhood $U$ of $f$ such that the mapping $f \mapsto \dim_H \Lambda_f$ is continuous. Continuity of box and Hausdorff dimensions for horseshoes has been studied also by Takens, Palis, and Viana (1988).

## Lorenz Attractor

E N Lorenz (1963), a meteorologist and student of G Birkhoff, showed by numerical experiments that for certain values of positive parameters $\sigma, r, b$, the quadratic system

$$\dot{x} = \sigma(y - x), \qquad \dot{y} = rx - y - xz, \qquad \dot{z} = xy - bz$$

has the global attractor $A$, for example, for $\sigma = 10$, $r = 28, b = 8/3$. In this case $\dim_B A \approx 2.06$, which is a numerical result (Grassberger and Procaccia, 1983). Using the analysis of local Lyapunov dimension along the flow in $A$, G A Leonov (2001) showed that if $\sigma + 1 \geq b \geq 2$ and $r\sigma^2(4 - b) + 2\sigma(b - 1) \times (2\sigma - 3b) > b(b - 1)^2$ then

$$\overline{\dim}_B A \leq 3 - \frac{2(\sigma + b + 1)}{\sigma + 1 + \sqrt{(\sigma - 1)^2 + 4r\sigma}}$$

## Hénon Attractor

M Hénon (1976), a theoretical astronomer, discovered the map $f : \mathbb{R}^2 \to \mathbb{R}^2, f(x, y) := (a + by - x^2, x)$, capturing several essential properties of the Lorenz system. In the case of $a = 1.4$ and $b = 0.3$, Hunt (1996) derived from [2] that for any compact, strictly $f$-invariant set $A$ in the trapping region $[-1.8, 1.8]^2$

there holds $\overline{\dim}_B A < 1.5$. Numerical experiments show that $\dim_B A \approx 1.28$ (Grassberger, 1983). Assuming $a > 0, b \in (0, 1)$, and $P_\pm(x_\pm, x_\pm) \in A$, where $P_\pm$ are fixed points of $f$, Leonov (2001) obtained that

$$\overline{\dim}_B A \le 1 + \frac{1}{1 - \ln b / \ln(\sqrt{x_-^2 + b} - x_-)}$$

Here

$$x_\pm := \frac{1}{2}\left[b - 1 \pm \sqrt{(b-1)^2 + 4a}\right]$$

The proof is based on the study of local Lyapunov dimension of $f$ and its iterates on $A$.

### Embedology

The physical relevance of box dimensions in the study of attractors is related to the problem of finding the smallest possible dimension $n$ sufficient to "embed" an attractor into $\mathbb{R}^n$. If $A \subset \mathbb{R}^k$ is a compact set and if $n > 2\underline{\dim}_B A$, then almost every map from $\mathbb{R}^k$ into $\mathbb{R}^n$, in the sense of prevalence, is one-to-one on $A$ and, moreover, it is an embedding on smooth manifolds contained in $A$ (T Sauer, J A Yorke, and M Casdagli, 1991). If $A$ is a strange attractor then the same is true for almost every delay-coordinate map from $\mathbb{R}^k$ to $\mathbb{R}^n$. This improves an earlier result by H Whitney (1936) and F Takens (Takens' embedology, 1981). The above notion of prevalence means the following: a property holds almost everywhere in the sense of prevalence if it holds on a subset $S$ of the space $V := C^1(\mathbb{R}^k, \mathbb{R}^n)$ for which there exists a finite-dimensional subspace $E \subset V$ (probe space) such that for each $v \in V$ we have that $v + e \in S$ for Lebesgue a.e. $e \in E$.

### Julia and Mandelbrot Sets

M Shishikura (1998) proved that the boundary of the Mandelbrot set $M$ generated by $f_c(z) := z^2 + c$ has the Hausdorff dimension equal to 2, thus answering positively to the conjecture by B Mandelbrot, J Milnor, and other mathematicians. Also for Julia sets there holds $\dim_H J(f_c) = 2$ for generic $c$ in $M$ (i.e., on the set of second Baire category). The proof is based on the study of the bifurcation of parabolic periodic points. Also, each baby Mandelbrot set sitting inside of $M$ has the boundary of Hausdorff dimension 2 (L Tan, 1998). Shishikura's results hold for more general functions $f(z) := z^d + c$, where $d \ge 2$.

For Julia sets $J(f_c)$ generated by $f_c(z) := z^2 + c$ there holds $d(c) := \dim_H J(f_c) = 1 + |c|^2/(4 \log 2) + o(|c|^2)$ for $c \to 0$. This and more general results have been obtained by Ruelle (1982). He also proved that the function $d(c)$ when restricted to the interval $[0, \infty)$ is real analytic in $[0, 1/4) \cup (1/4, \infty)$. Furthermore, it is left continuous at 1/4 (O Bodart and M Zinsmeister, 1996), but not continuous (A Douady, P Sentenac, and M Zinsmeister, 1997). Discontinuity of this map is related to the phenomenon of parabolic implosion at $c = 1/4$. The derivative $d'(c)$ tends to $+\infty$ from the left at $c = 1/4$ like $(1/4 - c)^{d(1/4) - 3/2}$ (G Havard and M Zinsmeister, 2000). Here $d(1/4) \approx 1.07$, which is a numerical result. Analysis of dimensions is based on methods of thermodynamic formalism.

C McMullen (1998) showed that if $\theta$ is an irrational number of bounded type (i.e., its continued fractional expansion $[a_1, a_2, \ldots]$ is such that the sequence $(a_i)$ is bounded from above) and $f(z) := z^2 + e^{2\pi\theta i} z$, then the Julia set $J(f)$ is porous. In particular, $\overline{\dim}_B J(f) < 2$. Y C Yin (2000) showed that if all critical points in $J(f)$ of a rational map $f : \overline{\mathbb{C}} \to \overline{\mathbb{C}}$ are nonrecurrent (a point is nonrecurrent if it is not contained in its $\omega$-limit set) then $J(f)$ is porous, hence $\overline{\dim}_B J(f) < 2$. Urbański and Przytycki (2001) described more general rational maps such that $\overline{\dim}_B J(f) < 2$.

### Spiral Trajectories

A standard planar model where the Hopf–Takens bifurcation occurs is $\dot{r} = r(r^{2l} + \sum_{i=0}^{l-1} a_i r^{2i}), \dot{\varphi} = 1$, where $l \in \mathbb{N}$. If $\Gamma$ is a spiral tending to the limit cycle $r = a$ of multiplicity $m$ (i.e., $r = a$ is a zero of order $m$ of the right-hand side of the first equation in the system) then $\dim_B \Gamma = 2 - 1/m$. Furthermore, for $m > 1$ the spiral is Minkowski measurable (Žubrinić and Županović, 2005). For $m = 1$ the spiral is Minkowski nondegenerate with respect to the gauge function $h(\varepsilon) := \varepsilon(\log(1/\varepsilon))^{-1}$.

### Infinite-Dimensional Dynamical Systems

In many situations the dynamics of the global attractor $A$ of the flow corresponding to an autonomous Navier–Stokes system is finite-dimensional (Ladyzhenskaya, 1972). This means that there exists a positive integer $N$ such that any trajectory in $A$ is completely determined by its orthogonal projection onto an $N$-dimensional subspace of a Hilbert space $X$. The aim is to find estimates of box and Hausdorff dimensions of the global attractor, in order to understand some of the basic and challenging problems of turbulence theory. If $A$ is a subset of a Hilbert space $X$, its Hausdorff dimension is defined analogously as for $A \subset \mathbb{R}^N$. The definition of the upper box dimension can be extended from $A \subset \mathbb{R}^N$ to

$$\overline{\dim}_B A := \overline{\lim}_{\varepsilon \to 0} \frac{\log m(A, \varepsilon)}{\log(1/\varepsilon)} \qquad [3]$$

where $m(A, \varepsilon)$ is the minimal number of balls sufficient to cover a given compact set $A \subset X$. The value of $\log m(A, \varepsilon)$ is called $\varepsilon$-entropy of $A$.

Foiaş and Temam (1979), Ladyzhenskaya (1982), A V Babin and M I Vishik (1982), Ruelle (1983), and E Lieb (1984) were among the first who obtained explicit upper bounds of Hausdorff and box dimensions of attractors of infinite-dimensional systems. For global attractors $A$ associated with some classes of two-dimensional Navier–Stokes equations with nonhomogeneous boundary conditions it can be shown that $\overline{\dim}_B A \le c_1 G + c_2 Re^{3/2}$, where $G$ is the Grashof number, $Re$ is the Reynolds number, and $c_i$ are positive constants (R M Brown, P A Perry, and Z Shen, 2000). V V Chepyzhov and A A Ilyin (2004) obtained that $\overline{\dim}_B A \le (1/\sqrt{2}\pi)(\lambda_1 |\Omega|)^{1/2} G$ for equations with homogeneous boundary conditions, where $\Omega \subset \mathbb{R}^2$ is a bounded domain, and $\lambda_1$ is the first eigenvalue of $-\Delta$. In the case of periodic boundary conditions Constantin, Foiaş, and Temam (1988) proved that $\overline{\dim}_B A \le c_1 G^{2/3}(1 + \log G)^{1/3}$, while for a special class of external forces there holds $\dim_H A \ge c_2 G^{2/3}$ (V X Liu, 1993). Let us mention an open problem by V I Arnol'd: is it true that the Hausdorff dimension of any attracting set of the Navier–Stokes equation on two-dimensional torus is growing with the Reynolds number?

In their study of partial regularity of solutions of three-dimensional Navier–Stokes equations, L Caffarelli, R Kohn, and L Nirenberg (1982) proved that the one-dimensional Hausdorff measure in space and time (defined by parabolic cylinders) of the singular set of any "suitable" weak solution is equal to zero. A weak solution is said to be singular at a point $(x_0, t_0)$ if it is essentially unbounded in any of its neighborhoods. Dimensions of attractors of many other classes of partial differential equations (PDEs) have been studied, like for reaction–diffusion systems, wave equations with dissipation, complex Ginzburg–Landau equations, etc. Related questions for nonautonomous PDEs have been considered by V V Chepyzhov and M I Vishik since 1992.

### Probability

Important examples of trajectories appearing in physics are provided by Brownian motions. Brownian motions $\omega$ in $\mathbb{R}^N, N \ge 2$, have paths $\omega([0,1])$ of Hausdorff dimension 2 with probability 1, and they are almost surely Hausdorff degenerate, since $\mathcal{H}^2(\omega([0,1])) = 0$ for a.e. $\omega$ (S J Taylor, 1953). Defining gauge functions $h(\varepsilon) := \varepsilon^2 \log(1/\varepsilon) \times \log\log\log(1/\varepsilon)$ when $N = 2$, and $h(\varepsilon) := \varepsilon^2 \log(1/\varepsilon)$ when $N \ge 3$, there holds $\mathcal{H}^h(\omega([0,1])) \in (0, \infty)$ for a.e. $\omega$ (D Ray, 1963, S J Taylor, 1964). If $N = 1$ then a.e. $\omega$ has the box and Hausdorff dimensions of the graph of $\omega|_{[0,1]}$ equal to 3/2 (Taylor, 1953), and for the gauge function $h(\varepsilon) := \varepsilon^{3/2} \log\log(1/\varepsilon)$ the corresponding generalized Hausdorff measure is nondegenerate. In the case of $N \ge 2$ we have the uniform dimension doubling property (R Kaufman, 1969). This means that for a.e. Brownian motion $\omega$ there holds $\dim_H \omega(A) = 2 \dim_H A$ for all subsets $A \subset [0, \infty)$. There are also results concerning almost sure Hausdorff dimension of double, triple, and multiple points of a Brownian motion and of more general Lévy stable processes.

Fractal dimensions also appear in the study of stochastic differential equations, like

$$dx_t = X_0(x_t)\,dt + \sum_{k=1}^{d} X_k(x_t)\,d\theta_k(t), \quad x_0 = x \in \mathbb{R}^N$$

The stochastic flow $(x_t)_{t \ge 0}$ in $\mathbb{R}^N$ is driven by a Brownian motion $(\theta(t))_{t \ge 0}$ in $\mathbb{R}^d$. Let us assume that $X_k, k = 0, \dots, d$, are $C^\infty$-smooth $T$-periodic divergence-free vector fields on $\mathbb{R}^N$. Then for almost every realization of the Brownian motion $(\theta(t))_{t \ge 0}$, the set of initial points $x$ generating the flow $(x_t)_{t \ge 0}$ with linear escape to infinity (i.e., $\underline{\lim}_{t \to \infty}(|x_t|/t) > 0$) is dense and of full Hausdorff dimension $N$ (D Dolgopyat, V Kaloshin, and L Koralov, 2002).

### Other Directions

There are many other fractal dimensions important for dynamics, like the Rényi spectrum for dimensions, correlation dimension, information dimension, Hentschel–Procaccia spectrum for dimensions, packing dimension, and effective fractal dimension. Relations between dimension, entropy, Lyapunov exponents, Gibbs measures, and multifractal rigidity have been investigated by Pesin, Weiss, Barreira, Schmeling, etc. Fractal dimensions are used to study dynamics appearing in Kleinian groups (D Sullivan, C J Bishop, P W Jones, C McMullen, B O Stratmann, etc.), quasiconformal mappings and quasiconformal groups (F W Gehring, J Väisäla, K Astala, C J Bishop, P Tukia, J W Anderson, P Bonfert-Taylor, E C Taylor, etc.), graph directed Markov systems (R D Mauldin, M Urbański, etc.), random walks on fractal graphs (J Kigami, A Telcs, etc.), billiards (H Masur, Y Cheung, P Bálint, S Tabachnikov, N Chernov, D Szász, I P Tóth, etc.), quantum dynamics (J-M Barbaroux, J-M Combes, H Schulz-Baldes, I Guarneri, etc.), quantum gravity (M Aizenman, A Aharony, M E Cates, T A Witten, G F Lawler, B Duplantier, etc.), harmonic analysis (R S Strichartz, Z M Balogh, J T Tyson, etc.),

number theory (L Barreira, M Pollicott, H Weiss, B Stratmann, B Saussol, etc.), Markov processes (R M Blumenthal, R Getoor, S J Taylor, S Jaffard, C Tricot, Y Peres, Y Xiao, etc.), and theoretical computer science (B Ya Ryabko, L Staiger, J H Lutz, E Mayordomo, etc.), and so on.

*See also:* Bifurcations of Periodic Orbits; Chaos and Attractors; Dissipative Dynamical Systems of Infinite Dimension; Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Ergodic Theory; Generic Properties of Dynamical Systems; Holomorphic Dynamics; Homoclinic Phenomena; Hyperbolic Dynamical Systems; Image Processing: Mathematics; Lyapunov Exponents and Strange Attractors; Partial Differential Equations: Some Examples; Polygonal Billiards; Quantum Ergodicity and Mixing of Eigenfunctions; Stochastic Differential Equations; Synchronization of Chaos; Universality and Renormalization; Wavelets: Applications; Wavelets: Mathematical Theory.

## Further Reading

Barreira L (2002) Hyperbolicity and recurrence in dynamical systems: a survey of recent results. *Resenhas* 5(3): 171–230.

Chepyzhov VV and Vishik MI (2002) *Attractors for Equations of Mathematical Physics*, Colloquium Publications, vol. 49. Providence, RI: American Mathematical Society.

Chueshov ID (2002) *Introduction to the Theory of Infinite-Dimensional Dissipative systems*, ACTA Scientific Publ. House. Kharkiv.

Falconer K (1990) *Fractal Geometry*. Chichester: Wiley.

Falconer K (1997) *Techniques in Fractal Geometry*. Chichester: Wiley.

Ladyzhenskaya OA (1991) *Attractors for Semigroups of Evolution Equations*. Cambridge: Cambridge University Press.

Lapidus M and van Frankenhuijsen M (eds.) (2004) *Fractal Geometry and Applications: A Jubilee of Benoît Mandelbrot*, Proc. Sympos. Pure Math., vol. 72, Parts 1 and 2, Providence, RI: American Mathematical Society.

Mandelbrot B and Frame M (2002) Fractals. In: *Encyclopedia of Physical Science and Technology*, 3rd edn., vol. 6, pp. 185–207. San Diego, CA: Academic Press.

Mauldin RD and Urbański M (2003) *Graph Directed Markov Systems: Geometry and Dynamics of Limit Sets*, Cambridge Tracts in Mathematics, 148. Cambridge: Cambridge University Press.

Palis J and Takens F (1993) *Hyperbolicity & Sensitive Chaotic Dynamics at Homoclinic Bifurcations*, Cambridge Studies in Advanced Mathematics, 35. Cambridge: Cambridge University Press.

Pesin Ya (1997) *Dimension Theory in Dynamical Systems: Contemporary Views and Applications*, Chicago Lecture Notes in Mathematics. Chicago, IL: University of Chicago Press.

Schmeling J and Weiss H (2001) An overview of the dimension theory of dynamical systems. In: Katok A, de la Llave R, Pesin Ya, and Weiss H (eds.) *Smooth Ergodic Theory and Its Applications*, (Seattle, 1999), Proceedings of Symposia in Pure Mathematics, vol. 69, pp. 429–488. Providence, RI: American Mathematical Society.

Tan L (ed.) (2000) *The Mandelbrot Set, Theme and Variations*. London Mathematical Society Lecture Note Series, 274. Cambridge: Cambridge University Press.

Temam R (1997) *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*, 2nd edn. Applied Mathematical Sciences, 68. New York: Springer.

Zinsmeister M (2000) *Thermodynamic Formalism and Holomorphic Dynamical Systems*, SMF/AMS Texts and Monographs 2. Providence, RI: American Mathematical Society, Société Mathématique de France, Paris.

# Fractional Quantum Hall Effect

**J K Jain**, The Pennsylvania State University, University Park, PA, USA



## Introduction

Interacting particles sometimes collectively behave in ways that take us by complete surprise. In a superfluid $^4$He atoms flow without viscosity, and in a superconductor electrons flow without resistance. Such behaviors announce emergent structures and principles which have often found applications in other areas. This article concerns the surprising collective effects that occur when electrons are confined in two dimensions and subjected to a strong transverse magnetic field. At low temperatures, the Hall resistance (defined below) exhibits plateaus on which it is precisely quantized at

$$R_H = \frac{h}{fe^2} \qquad [1]$$

where $h$ and $e$ are fundamental constants and $f$ is a plateau-specific rational fraction. This phenomenon is known as the "fractional quantum Hall effect" (FQHE), or, after its discoverers, the "Tsui–Stormer–Gossard" (TSG) effect. The underlying state provides a new paradigm for collective behavior in nature, and is understood in terms of a new class of quasiparticles known as "composite fermions," which are topological bound states of electrons and quantized vortices. This article will outline the basics of the experimental phenomenology and our theoretical understanding of this effect.

## The Hall Effect

The Ohm's law, $I = V/R$, tells us that the current through a resistor is proportional to the applied voltage. The local form of the law is

$$\boldsymbol{J} = \sigma \boldsymbol{E} \qquad [2]$$

where $\sigma$ is the conductivity, and $\boldsymbol{J} = q\rho \boldsymbol{v}$ is the current density for particles of charge $q$ and density $\rho$ moving with a velocity $\boldsymbol{v}$.

In 1879, E H Hall discovered that in the presence of a crossed electric and magnetic fields ($\boldsymbol{E}$ and $\boldsymbol{B}$), the current flows in a direction "perpendicular" to the plane containing the two fields. Alternatively, the passage of current induces a voltage perpendicular to the direction of the current flow. This is known as the Hall effect (see **Figure 1**). The phenomenon has a classical origin. A consequence of the Lorentz force law of electrodynamics,

$$\boldsymbol{F} = q\left(\boldsymbol{E} + \frac{1}{c}\boldsymbol{v} \times \boldsymbol{B}\right) \qquad [3]$$

which gives the force on a particle of charge $q$ moving with a velocity $\boldsymbol{v}$, is that for crossed electric and magnetic fields the particle drifts in the direction $\boldsymbol{E} \times \boldsymbol{B}$ with a velocity $v = cE/B$. The current density is therefore given by $\boldsymbol{J} = q\rho\boldsymbol{v}$, where $\rho$ is the (three-dimensional) density of particles. That produces the Hall resistivity

$$\rho_{\mathrm{H}} = \frac{E_y}{J_x} = \frac{B}{\rho q c} \qquad [4]$$

## The von Klitzing Effect

Molecular beam epitaxy allows controllable layer by layer growth in which one type of semiconductor, say GaAs, can be grown on top of another, say $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$, to produce an atomically sharp interface. By appropriately doping such structures, electrons can be captured at the interface, thus producing a two-dimensional electron system (2DES). We note that these are three-dimensional electrons confined to move in two dimensions. The interaction has the standard Coulomb form $V(r) = e^2/\epsilon r$, where $\epsilon$ is the dielectric constant of the host material. (In a hypothetical world which has only two space dimensions, the interaction would be logarithmic.)

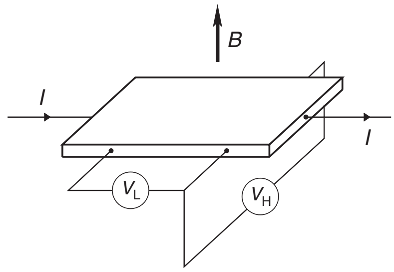



**Figure 1** Schematics of magnetotransport measurement. $I$, $V_{\mathrm{L}}$, and $V_{\mathrm{H}}$ are the current, longitudinal voltage, and the Hall voltage, respectively. The longitudinal and Hall resistances are defined as $R_{\mathrm{L}} \equiv V_{\mathrm{L}}/I$ and $R_{\mathrm{H}} \equiv V_{\mathrm{H}}/I$.

The "integral quantum Hall effect" (IQHE) or the "von Klitzing effect" was discovered unexpectedly by von Klitzing and collaborators in 1980, in their study of Hall effect in a 2DES. In two dimensions, one defines the Hall resistance as

$$R_{\mathrm{H}} = \frac{V_{\mathrm{H}}}{I} \qquad [5]$$

which, from classical electrodynamics, is expected to be proportional to the magnetic field $B$. That is indeed the case at small magnetic fields. At sufficiently high $B$, however, quantum mechanical effects appear in a dramatic manner. The essential observations are as follows.

1. When plotted as a function of the magnetic field $B$, the Hall resistance exhibits numerous plateaus. On any given plateau, $R_{\mathrm{H}}$ is precisely quantized with values given by

$$R_{\mathrm{H}} = \frac{h}{ne^2} \qquad [6]$$

where $n$ is an integer (hence the name "integral quantum Hall effect"). The plateau occurs in the vicinity of $\nu \equiv Be/\rho hc = n$, where $\nu$ is the "filling factor" (defined below).

2. In the plateau region, the longitudinal resistance exhibits an Arrhenius behavior:

$$R_{\mathrm{L}} \sim \exp\left(-\frac{\Delta}{2k_{\mathrm{B}}T}\right) \qquad [7]$$

This gives a filling-factor dependent energy scale $\Delta$, which indicates the presence of a gap in the excitation spectrum. $R_{\mathrm{L}}$ vanishes in the limit $T \to 0$.

The absolute accuracy of the quantization has been established to a few parts in $10^8$ for $1\sigma$ uncertainty, and the relative accuracy to a few parts in $10^{10}$. There is presently no known "intrinsic" correction to the quantization. Perhaps, the most remarkable aspect of the effect is its universality. It is independent of the sample type, geometry, various material parameters (the band mass of the electron or the dielectric constant of the semiconductor), and disorder. The combination $h/e^2$ also occurs in the definition of the fine

structure constant $\alpha = e^2/\hbar c$, the value of which is approximately 1/137. The Hall effect measurements in dirty, solid state systems thus provide one of the most accurate values for $\alpha$. Finally, the lack of resistance at $T = 0$ is to be contrasted with ordinary metals, for which the resistance at $T \to 0$, called the residual resistance, is finite and proportional to disorder.

## The TSG Effect

The next revolution occurred in 1982 with the discovery of the TSG effect, that is, plateaus on which the Hall resistance is quantized at values given by eqn [1] (see Figure 2). The observation of the $R_H = h/fe^2$ plateau is often referred to as the observation of the fraction $f$. Improvement of experimental conditions has led to the observation of a large number of fractions over the years, revealing the richness of the TSG effect. At the time of the writing of this article, the number of observed fractions is more than 50 if one counts only fractions below unity. As in the von Klitzing effect, the longitudinal resistance exhibits an Arrhenius behavior, vanishing in the limit $T \to 0$.

## Landau Levels

The Hamiltonian for a nonrelativistic electron moving in two space dimensions in a perpendicular magnetic field is given by

$$H = \frac{1}{2m_b}\left(p + \frac{eA}{c}\right)^2 \quad [8]$$

Here, $m_b$ is the electron's band mass and $-e$ its charge. For a uniform magnetic field, the vector potential $A$ satisfies

$$\nabla \times A = B\hat{z} \quad [9]$$

Because $A$ is a linear function of the spatial coordinates, it follows that $H$ is a generalized two-dimensional harmonic oscillator Hamiltonian which is quadratic in both the spatial coordinates and in the canonical momentum $p = -i\hbar\nabla$, and therefore can be diagonalized exactly.

A convenient gauge choice is the symmetric gauge:

$$A = \frac{B \times r}{2} = \frac{B}{2}(-y, x, 0) \quad [10]$$

With the magnetic length $\ell = \sqrt{\hbar c/eB}$ and the cyclotron energy $\hbar\omega_c = \hbar eB/m_b c$ chosen as the

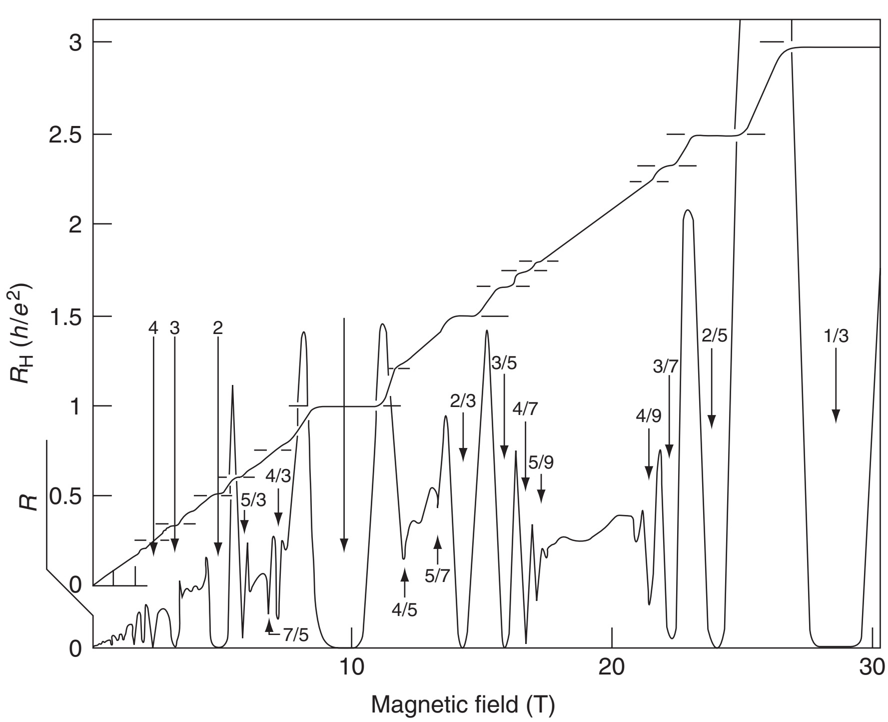



**Figure 2** The TSG effect. The Hall resistance ($R_H$) exhibits many precisely quantized plateaus, concurrent with minima in the longitudinal resistance ($R$). Reproduced with permission from *Perspectives in Quantum Hall Effects*; HL Stormer and DC Tsui; SD Sarma and A Pinczuk (eds.); Copyright © 1997, Wiley. Reprinted with permission of John Wiley & Sons, Inc.

units for length and energy, the Hamiltonian can be expressed as

$$H = \frac{1}{2}\left[\left(-\mathrm{i}\frac{\partial}{\partial x} - \frac{y}{2}\right)^2 + \left(-\mathrm{i}\frac{\partial}{\partial y} + \frac{x}{2}\right)^2\right] \qquad [11]$$

Choosing as independent variables

$$z \equiv x - \mathrm{i}y, \qquad \bar{z} \equiv x + \mathrm{i}y \qquad [12]$$

we get

$$H = \frac{1}{2}\left(-4\frac{\partial^2}{\partial z \partial \bar{z}} + \frac{1}{4}z\bar{z} - z\frac{\partial}{\partial z} + \bar{z}\frac{\partial}{\partial \bar{z}}\right) \qquad [13]$$

Now define the following sets of ladder operators:

$$b = \frac{1}{\sqrt{2}}\left(\frac{\bar{z}}{2} + 2\frac{\partial}{\partial z}\right) \qquad [14]$$

$$b^\dagger = \frac{1}{\sqrt{2}}\left(\frac{z}{2} - 2\frac{\partial}{\partial \bar{z}}\right) \qquad [15]$$

$$a^\dagger = \frac{1}{\sqrt{2}}\left(\frac{\bar{z}}{2} - 2\frac{\partial}{\partial z}\right) \qquad [16]$$

$$a = \frac{1}{\sqrt{2}}\left(\frac{z}{2} + 2\frac{\partial}{\partial \bar{z}}\right) \qquad [17]$$

which have the property that

$$[a, a^\dagger] = 1, \qquad [b, b^\dagger] = 1 \qquad [18]$$

and all the other commutators are zero. In terms of these operators, the Hamiltonian can be written as

$$H = a^\dagger a + \tfrac{1}{2} \qquad [19]$$

The eigenvalue of $a^\dagger a$ is an integer, $n$, called the Landau level (LL) index. The $z$-component of the canonical angular momentum operator, the only relevant component for the two-dimensional problem, is defined as

$$L_z = -\mathrm{i}\frac{\partial}{\partial \theta} = \bar{z}\frac{\partial}{\partial \bar{z}} - z\frac{\partial}{\partial z} = a^\dagger a - b^\dagger b \qquad [20]$$

Exploiting the property $[H, L_z] = 0$, the eigenfunctions will be chosen to diagonalize $H$ and $L_z$ simultaneously. The eigenvalue of $L_z$ will be denoted by $-m$. The analogy to the Harmonic oscillator problem immediately gives the solution

$$H|m, n\rangle = E_n|m, n\rangle \qquad [21]$$

where

$$E_n = \left(n + \frac{1}{2}\right) \qquad [22]$$

and

$$|m, n\rangle = \frac{(b^\dagger)^{m+n}}{\sqrt{(m+n)!}}\frac{(a^\dagger)^n}{\sqrt{n!}}|0, 0\rangle \qquad [23]$$

where $m = -n, -n+1, \ldots$. The single-particle orbital at the bottom of the two ladders defined by the two sets of raising and lowering operators is

$$\langle \boldsymbol{r}|0, 0\rangle \equiv \eta_{0,0}(\boldsymbol{r}) = \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-z\bar{z}/4} \qquad [24]$$

which satisfies

$$a|0, 0\rangle = b|0, 0\rangle = 0 \qquad [25]$$

The single-particle states are particularly simple in the lowest Landau level ($n = 0$):

$$\eta_{0,m}(\boldsymbol{r}) = \langle \boldsymbol{r}|0, m\rangle = \frac{z^m\,\mathrm{e}^{-z\bar{z}/4\ell^2}}{\sqrt{2\pi\ell^2 2^m m!}} \qquad [26]$$

Aside from the ubiquitous Gaussian factor, a general state in the lowest Landau level is given by a polynomial of $z$; it does not involve any $\bar{z}$. In other words, apart from the Gaussian factor, the lowest Landau level wave functions are analytic functions of $z$.

### Landau Level Degeneracy

The state $\eta_{0,m}(\boldsymbol{r})$ is peaked strongly at $r = \sqrt{2m}\,\ell$. Neglecting order-1 effects, there are $m$ states in the lowest Landau level in a disk of radius $r = \sqrt{2m}\,\ell$, giving a degeneracy of $(2\pi\ell^2)^{-1}$ per unit area per Landau level. (The same degeneracy is obtained for higher Landau levels as well.) It is equal to $B/\phi_0$, where $\phi_0 = hc/e$ is called the flux quantum, that is, there is one state per flux quantum in each Landau level.

### Filling Factor

The number of filled Landau levels, called the filling factor, is given by

$$\nu = \rho 2\pi\ell^2 = \frac{\rho\phi_0}{B} \qquad [27]$$

## The Origin of Plateaus

The von Klitzing effect can be explained in terms of a model which neglects the interactions between electrons. It occurs because the ground state at an integral filling is unique and nondegenerate, separated from excitations by a gap. Laughlin (1981) showed that the disorder-induced Anderson localization also plays a crucial role in the establishment of the Hall plateaus. To see this, imagine changing the filling away from an integer by adding some

electrons or holes. In a perfect system, the additional particles would also be free to carry current, but in the actual, disordered sample, they are immobilized by impurities (which create localized states in the energy gap), and do not contribute to transport. The transport properties therefore remain unaffected as the filling factor is varied slightly away from an integer, and the system continues to behave as though it had filled shells.

## The Lowest Landau Level Problem

The TSG effect arises due to interelectron interaction. We wish to obtain solutions for the Schrödinger equation

$$H\Psi = E\Psi \qquad [28]$$

at an arbitrary filling $\nu$, where

$$H = \sum_j \frac{1}{2m_{\rm b}}\left[\frac{\hbar}{\rm i}\nabla_j + \frac{e}{c}\boldsymbol{A}(\boldsymbol{r}_j)\right]^2 + \frac{e^2}{\epsilon}\sum_{j<k}\frac{1}{|\boldsymbol{r}_j - \boldsymbol{r}_k|} \qquad [29]$$

The first term on the right-hand side is the kinetic energy in the presence of a constant external magnetic field $\boldsymbol{B} = \nabla \times \boldsymbol{A}$, and the second term is the Coulomb interaction energy. (The Zeeman energy is not included explicitly because we consider, for now, magnetic fields that are sufficiently high that only fully spin-polarized states are relevant.) It is convenient to consider the limit $(e^2/\epsilon\ell)/(\hbar\omega_{\rm c}) \to 0$, when the Coulomb interaction is so weak that it is not able to cause Landau level mixing, so electrons can be taken to be within the lowest Landau level. The kinetic energy then is an irrelevant constant which can be thrown away, and the Hamiltonian reduces to

$$H = \mathcal{P}_{\rm LLL}\frac{e^2}{\epsilon}\sum_{j<k}\frac{1}{|\boldsymbol{r}_j - \boldsymbol{r}_k|}\mathcal{P}_{\rm LLL} \qquad [30]$$

"which must be solved with the lowest LL restriction," as explicitly indicated by the lowest LL projection operator $\mathcal{P}_{\rm LLL}$. The problem is thus mathematically well defined, but requires degenerate perturbation theory in an enormously large Hilbert space, with $\binom{N/\nu}{N}$ many particle basis vectors. The usual perturbative techniques are not useful due to the absence of a small parameter in the problem; $e^2/\epsilon\ell$ merely sets the energy scale in the lowest Landau level.

## Composite-Fermion Theory

Inspired by the qualitative similarity between the integral and the fractional Hall effects, the composite-fermion (CF) theory (Jain 1989) postulates that the eigenfunctions of interacting electrons at filling factor $\nu$, $\Psi_\nu$, are related to the (known) eigenfunctions of *non*interacting electrons at filling factor $\nu^*$, $\Phi_{\nu^*}$, according to

$$\Psi_\nu = \mathcal{P}_{\rm LLL}\Phi_{\nu^*}\prod_{j<k}(z_j - z_k)^{2p} \qquad [31]$$

where $\mathcal{P}_{\rm LLL}$ denotes projection of the wave function on its right into the lowest Landau level. The filling factors are related by

$$\nu = \frac{\nu^*}{2p\nu^* + 1} \qquad [32]$$

which can be seen as follows: the largest power of $z_1$ in $\Phi_{\nu^*}$ (neglecting order-one corrections) is $N/\nu^*$, as follows from the definition of the filling factor. The largest power of $z_1$ on the right-hand side is therefore $pN(N-1) + N/\nu^*$. This is the number of flux quanta penetrating the "sample." Dividing it by $N$ and taking the limit $N \to \infty$ gives the inverse of the filling factor $\nu^{-1}$. These wave functions are now known to capture the correct nonperturbative physics of the TSG effect (see below), and also to provide extremely accurate representations for the actual correlated ground states and their excitations. They recover Laughlin's 1983 wave function for the ground state at $\nu = 1/(2p+1)$, while clarifying that it is a part of a much bigger conceptual structure.

### Physical Interpretation

The crucial property of the wave function in eqn [31] is that the complex Jastrow factor $\prod_{j<k}(z_j - z_k)^{2p}$ binds $2p$ vortices on each electron. More precisely, each electron sees $2p$ vortices on every other electron, in that a complete loop of an electron around any other electron produces a phase of $2\pi \times 2p$. The bound state is interpreted as a particle, called the "composite fermion." Because the vortex is a topological object, so is the composite fermion. The vorticity $2p$ is quantized to be an even integer, as required by the single-valuedness and antisymmetry requirements of quantum mechanics, which will be seen to lie at the root of the exact quantization of the Hall resistance.

When composite fermions move about, they experience, in addition to the Aharonov–Bohm (AB) phase, also the Berry phases coming from vortices on other composite fermions. Imagine taking a composite fermion in a closed loop enclosing an area $A$. The phase associated with that loop is given by

$$\Phi^* = -2\pi\frac{BA}{\phi_0} + 2\pi 2pN_{\rm enc} \qquad [33]$$

where $N_{\rm enc}$ is the number of composite fermions inside the loop. The first term is the familiar AB phase due to a charge going around in a loop. The second is the Berry phase due to $2p$ vortices going around $N_{\rm enc}$ particles, with each particle producing a phase of $2\pi$. Replacing $N_{\rm enc}$ by its average value $\rho A$ shows that, on average, $\Phi^*$ is equal to the AB phase from an "effective" magnetic field

$$B^* = B - 2p\rho\phi_0 \qquad [34]$$

The composite fermions thus experience an effective magnetic field $B^*$ which is much smaller than the external, applied field $B$. That lies at the heart of the phenomenology of this lowest Landau level liquid. One treats composite fermions as noninteracting in the simplest approximation. They form their own Landau-like levels in $B^*$. Their filling factor is defined as $\nu^* = \rho\phi_0/B^*$, with which eqn [34] becomes equivalent to eqn [32]. The effective field $B^*$ can be antiparallel to $B$, in which case $\nu^* = \rho\phi_0/B^*$ is formally negative. For negative values of $\nu^*$,$\Phi_{\nu^*}$ in eqn [31] is defined as $\Phi_{-|\nu^*|} = [\Phi_{|\nu^*|}]^*$, because complex conjugation is equivalent to switching the direction of the magnetic field.

## Fermion Chern–Simons Theory

Lopez and Fradkin (1991) developed a field-theoretic formulation of composite fermions through a singular gauge transformation defined by

$$\Psi = \prod_{j<k}\left(\frac{z_j - z_k}{|z_j - z_k|}\right)^{2p}\Psi' \qquad [35]$$

under which the eigenvalue problem of eqn [29] transforms into

$$H'\Psi' = E\Psi' \qquad [36]$$

$$H' = \frac{1}{2m_{\rm b}}\sum_i\left(\boldsymbol{p}_i + \frac{e}{c}\boldsymbol{A}(\boldsymbol{r}_i) - \frac{e}{c}\boldsymbol{a}(\boldsymbol{r}_i)\right)^2 + V \qquad [37]$$

$$\boldsymbol{a}(\boldsymbol{r}_i) = \frac{2p}{2\pi}\phi_0\sum_j{}' \nabla_i\phi_{ij} \qquad [38]$$

where

$$\phi_{jk} = \mathrm{i}\ln\frac{z_j - z_k}{|z_j - z_k|}$$

is the relative angle between the particles $j$ and $k$. The magnetic field corresponding to $\boldsymbol{a}(\boldsymbol{r}_i)$ is given by

$$\boldsymbol{b}_i = \nabla_i \times \boldsymbol{a}(\boldsymbol{r}_i) = 2p\phi_0\sum_l{}' \delta^2(\boldsymbol{r}_i - \boldsymbol{r}_l) \qquad [39]$$

The above transformation thus amounts to attaching a point flux of strength $-2p\phi_0$ to each electron, which is how the composite fermion is modeled in this approach. (A flux quantum is topologically equivalent to a vortex.) This definition is reminiscent of the treatments of particles obeying fractional statistics ("anyons") introduced by Leinaas and Myrheim (1977) and Wilczek (1982); an anyon is modeled as an electron bound to a point flux of magnitude $\alpha\phi_0$, where $\alpha$ determines the winding statistics.

It is not possible to proceed further without making approximations. The usual approach is to make a "mean-field" approximation, which amounts to spreading the point flux on each electron into a uniform magnetic field. Formally, one writes

$$\boldsymbol{A} - \boldsymbol{a} \equiv \boldsymbol{A}^* + \delta\boldsymbol{A} \qquad [40]$$

$$\nabla \times \boldsymbol{A}^* = B^*\hat{z} \qquad [41]$$

The transformed Hamiltonian is written as

$$H' = \frac{1}{2m_{\rm b}}\sum_i\left(\boldsymbol{p}_i + \frac{e}{c}\boldsymbol{A}^*(\boldsymbol{r}_i)\right)^2 + V + V' = H_0' + V + V' \qquad [42]$$

$V$ is the Coulomb interaction and $V'$ denotes the terms containing $\delta\boldsymbol{A}$. The solution to $H_0'$ is trivial, describing free fermions in an effective magnetic field $B^*$. We have thus decomposed the Hamiltonian into a part $H_0'$, which can be solved exactly, and the rest, $V + V'$, which is to be treated perturbatively.

Lopez and Fradkin recast the problem in the language of functional integrals, which is suitable for studying corrections to the mean-field theory. One writes the zero-temperature quantum partition function

$$\mathcal{Z} = \int \mathcal{D}\psi\mathcal{D}\psi^*\mathcal{D}\boldsymbol{a}\exp\left(\frac{\mathrm{i}}{\hbar}\mathcal{S}\right) \qquad [43]$$

$$\mathcal{S} = \int \mathrm{d}^2\boldsymbol{r}\int \mathrm{d}t\mathcal{L} \qquad [44]$$

$$\mathcal{L} = \psi^*(\mathrm{i}\partial_t - a_0)\psi + \frac{1}{2m_{\rm b}}\left|\left(-\mathrm{i}\hbar\nabla + \frac{e}{c}\boldsymbol{A} - \frac{e}{c}\boldsymbol{a}\right)\psi\right|^2 + \frac{1}{2p\phi_0}a_0\nabla\times\boldsymbol{a} + \int \mathrm{d}^2\boldsymbol{r}'\rho(\boldsymbol{r})V(\boldsymbol{r} - \boldsymbol{r}')\rho(\boldsymbol{r}') \qquad [45]$$

where $\psi$ and $\psi^*$ are anticommuting Grassmann variables. The flux attachment is introduced through a Lagrange multiplier $a_0$; because $a_0$ enters linearly in the action, it can be integrated out to produce a delta function that imposes the constraint

$$\nabla \times \boldsymbol{a}(\boldsymbol{r}) = 2p\phi_0\rho(\boldsymbol{r}) = 2p\phi_0\psi^*(\boldsymbol{r})\psi(\boldsymbol{r}) \qquad [46]$$

This formalism is closely related to the topological Chern–Simons (CS) field theory. Recall that the CS Lagrangian has the form

$$\mathcal{L}_{\mathrm{CS}} \sim \epsilon^{\mu\nu\lambda} A_\mu F_{\nu\lambda} = 2\epsilon^{\mu\nu\lambda} A_\mu \partial_\nu A_\lambda \qquad [47]$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, and $\epsilon^{\mu\nu\lambda}$ is the antisymmetric Levy-Civita tensor, with $\epsilon^{012} = 1$. The index takes values $\mu = 0, 1, 2$, the first being the time component and the remaining space components. The CS action is invariant, up to surface terms, under a gauge transformation, because the change in $\mathcal{L}_{\mathrm{CS}}$ under a functional variation $\delta A_\mu = \partial_\mu \Lambda$ is a total derivative.

Zhang *et al.* (1989) noted that the term proportional to $a_0 \nabla \times \boldsymbol{a}$ in eqn [45], which enforces flux attachment, is precisely equal to the CS Langrangian in the Coulomb gauge. Write

$$\begin{aligned}\mathcal{L}_{\mathrm{CS}} &= \frac{1}{4p\phi_0} \epsilon^{\mu\nu\lambda} a_\mu \partial_\nu a_\lambda \\ &= \frac{1}{2p\phi_0} \epsilon^{ij} a_0 \partial_i a_j - \frac{1}{4p\phi_0} \epsilon^{ij} a_i \partial_0 a_j\end{aligned} \qquad [48]$$

where $i, j$ represent the spatial components $(i, j = 1, 2)$, and the time components have been displayed explicitly in the second step ($\partial_0 = \partial_t$). The first term on the right-hand side of eqn [48] is identical to the third term on the right-hand side of eqn [45]. In the Fourier space the last term is proportional to

$$\epsilon^{ij} a_i(\boldsymbol{q}, \omega)(-\mathrm{i}\omega) a_j(-\boldsymbol{q}, -\omega) \qquad [49]$$

By choosing the $x$-axis along $\boldsymbol{q}$, the Coulomb gauge condition $\boldsymbol{q}.\boldsymbol{a} = 0$ implies $a_2(\boldsymbol{q}, \omega) = 0$, guaranteeing that the last term in eqn [48] is identically zero.

The constraint of eqn [46] is used to eliminate the two factors of density in the last term of eqn [45]. The action is then quadratic in the fermion field, which can be integrated out. Various response functions can be expressed as correlation functions of the vector potential field and their averages over the CS field configurations are evaluated perturbatively by standard diagrammatic methods.

The fermion CS theory is believed to capture the topological properties of composite fermions, but has not lent itself, because of the lack of a small parameter, to quantitative calculations. It is not known what classes of Feynman diagrams will need to be summed to eliminate the electron mass $m_b$ (which is not a parameter of the lowest Landau problem – see eqn [30]) in the fermion CS approach). Halperin *et al.* (1993) proceeded by replacing $m_b$ by an adjustable parameter $m^*$, interpreted as the composite-fermion mass. Murthy and Shankar (1997) proposed to separate out the inter- and intra-Landau level degrees of freedom by making a sequence of further transformations.

## Consequences

### Fractional Quantum Hall Effect

The CF theory provides a simple understanding of why gaps open up at "fractional" fillings, which happens at those fillings $\nu = f$ for which composite fermions fill integral numbers of CF Landau levels. That results in Hall plateaus at $R_{\mathrm{H}} = h/fe^2$ in the presence of disorder. The fractional QHE is thus understood as the integral QHE for composite fermions.

### Sequences of Fractions

The integral fillings of composite fermions correspond to fractional fillings of electrons given by

$$\nu = \frac{|n|}{2p|n| \pm 1} \qquad [50]$$

which are precisely the observed fractions. Some of these are:

$$f = \frac{|n|}{2|n| + 1} = \frac{1}{3}, \frac{2}{5}, \ldots, \frac{10}{21} \qquad [51]$$

$$f = \frac{|n|}{2|n| - 1} = \frac{2}{3}, \frac{3}{5}, \ldots, \frac{10}{19} \qquad [52]$$

$$f = \frac{|n|}{4|n| + 1} = \frac{1}{5}, \frac{2}{9}, \ldots, \frac{6}{25} \qquad [53]$$

$$f = \frac{|n|}{4|n| - 1} = \frac{2}{7}, \frac{2}{5}, \ldots, \frac{6}{23} \qquad [54]$$

Particle–hole symmetry in the lowest Landau level also implies fractions $1 - f$. The fractions appear in the form of sequences because they are all derived from the sequence of integers. The Hall quantization is exact because the right-hand side of eqn [50] is made up of whole numbers and therefore is not susceptible to small perturbations in the Hamiltonian. The CF theory unifies the FQHEs and IQHEs.

### Fermi Sea at Half Filling

Equation [50] is consistent with the fact that only odd-denominator fractions have been observed in the lowest Landau level (i.e., with $f < 1$). Halperin *et al.* (1993) and Kalmeyer and Zhang (1992) proposed that at the simplest even-denominator fraction, namely $\nu = 1/2$, composite fermions form a Fermi sea. This was motivated by the fact that the effective magnetic field is $B^* = 0$ at $\nu = 1/2$. A number of experiments have directly measured the Fermi sea of composite fermions. The TSG effect with $f = 1/2$ is absent because the Fermi sea has gapless excitations.

### Effective Magnetic Field

For small values of $B^*$ (i.e., in the vicinity of $\nu = 1/2$), the cyclotron radius of composite fermions can be very large compared to the radius of the cyclotron orbit of a classical electron in $B$. Direct measurements of the cyclotron orbit in several geometric experiments have confirmed that the charge carriers experience a magnetic field $B^*$ rather than $B$.

### Fractional Charge

Laughlin (1983) showed that the presence of a gap at a fractional filling implies the existence of fractionally charged excitations. He obtained an excitation through the adiabatic insertion of a point flux quantum at, say the origin, which can be gauged away at the end leaving behind an exact excited state. The Faraday's law implies that the azimuthal component of the induced electric field is $E_\phi = -(2\pi r)^{-1}\, d\phi/dt$. The current density then is $j_r = \sigma_H E_\phi$, where $\sigma_H = fe^2/h$ is the Hall conductivity. The charge leaving the area defined by a circle of radius $r$ per unit time is $2\pi r j_r$. The total charge leaving this area in the adiabatic process then is

$$Q = \int 2\pi r j_r \, dt = -\sigma_H \phi_0 = -fe \qquad [55]$$

The charge excess associated with the excitation is therefore $fe$. It is in general not an elementary excitation. For $f = n/(2pn \pm 1)$, it can be shown to be a collection of $n$ elementary excitations, giving a charge of $e^* = e/(2pn \pm 1)$ for a single elementary excitation.

### Microscopic Tests

Exact solutions of the Schrödinger equation can be obtained, for a finite number of particles, by a brute-force diagonalization of the Hamiltonian in the lowest LL subspace, which enables a rigorous and nontrivial testing of the CF theory. **Figure 3** shows some typical comparisons, which help test both the qualitative and the quantitative aspects of the CF theory in a model-independent manner. The low-energy spectrum of interacting electrons at $B$ is explicitly seen to have a one-to-one correspondence to that of weakly interacting electrons at $B^*$. Furthermore, there is a remarkably good quantitative agreement. The predicted energies agree with the exact energies to better than 0.05%, and the overlaps between the wave functions of eqn [31] with the exact eigenfunctions are close to 100%. Such comparisons are even more convincing in light of the fact that the wave functions of eqn [31] do not contain any adjustable parameters for the states at $\nu$ in eqn [50], because the ground state wave function and its low-energy excitations at $\nu^* = n$ are unique and fully known: the former is the Slater determinant corresponding to $n$ filled

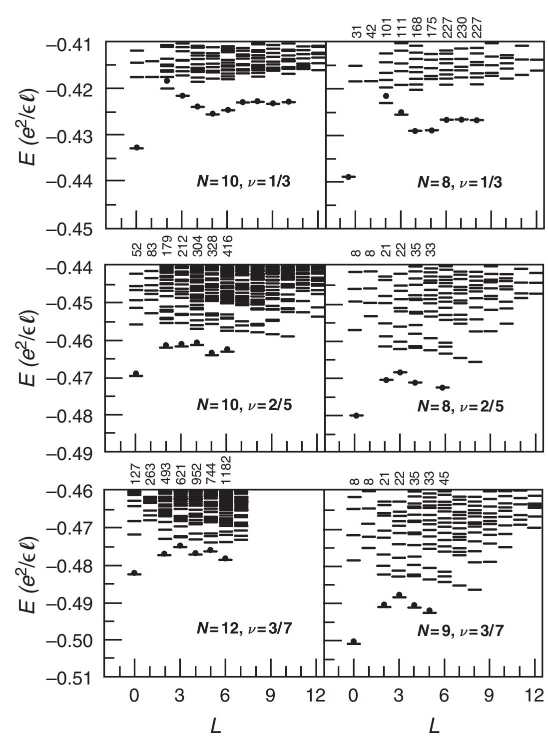



**Figure 3** Exact spectra (dashes) for several particle numbers at $\nu = 1/3$, 2/5, and 3/7. Dots show the CF prediction for the energy, obtained with no adjustable parameters. The electrons are taken to be confined on the surface of a sphere in the presence of a radial magnetic field; $L$ is the total orbital angular momentum, and each dash represents a multiplet of $2L + 1$ degenerate states. The number on top is the dimension of the Fock space in the corresponding $L$ sector. Reproduced from Jain JK (2000) The composite fermion: a quantum particle and its quantum fluids. *Physics Today* 39(4): 39–42, with permission from American Institute of Physics.

Landau levels, and the latter are the excitons. The predicted energies are calculated by determining the expectation values of the full Hamiltonian of eqn [30] with respect to the wave functions in eqn [31].

## More Physics

### Spin

At small Zeeman energies, partially spin-polarized or spin-unpolarized FQHE states become possible. The TSG effect with spin is well described by a generalization of the CF theory. The observed fractions are still given by eqn [50], but with

$$n = n_{\uparrow} + n_{\downarrow} \qquad [56]$$

where $n_{\uparrow}$ is the number of occupied spin-up Landau-like CF bands and $n_{\downarrow}$ is the number of occupied spin-down Landau-like CF bands. There are in general several states with different spin polarizations possible at any given fraction. The observed quantum phase transitions as a function of the Zeeman energy, which can be changed by increasing the parallel component of the magnetic field, are consistent with this picture. Direct measurements of the spin polarization further confirm this, but also see evidence for certain additional fragile states, which are presumably caused by the residual interaction between composite fermions.

### Bilayers

It has been proposed that for two parallel 2DES planes at small separations and at total filling $\nu = 1$, neutral interlayer excitons (each exciton made up of an electron in one layer and a hole in the other) undergo Bose–Einstein condensation, producing a true off-diagonal long-range order. Tunneling and transport experiments by Eisenstein and collaborators provide evidence for nontrivial behavior under such conditions. The resistivity in the antisymmetric channel is very small but does not vanish.

### Pairing

An even-denominator fraction $f = 5/2$ has been observed. Writing $5/2 = 2 + 1/2$ and noting that the lowest LL contributes 2 (counting the spin degree of freedom), $\nu = 5/2$ corresponds to a filling of 1/2 in the second Landau level. The most promising scenario for the explanation of the 5/2 effect is that composite fermions form a $p$-wave paired state, which opens up a gap to excitations. This state is believed to be well described by a Pfaffian wave function proposed by Moore and Read (1991)

$$\Psi^{\mathrm{Pf}}_{1/2} = \mathrm{Pf}\left(\frac{1}{z_i - z_j}\right) \times \prod_{i<j} (z_i - z_j)^2 \exp\left(-\frac{1}{4}\sum_k |z_k|^2\right) \qquad [57]$$

The Pfaffian of an antisymmetric matrix $M$ is defined, apart from an overall factor, as

$$\mathrm{Pf}(M_{ij}) = A(M_{12}M_{34}\ldots M_{N-1,N}) \qquad [58]$$

where $A$ is the antisymmetrization operator. The Bardeen–Cooper–Schrieffer wave function

$$\Psi_{\mathrm{BCS}} = A[\phi_0(\boldsymbol{r}_1, \boldsymbol{r}_2)\phi_0(\boldsymbol{r}_3, \boldsymbol{r}_4)\ldots\phi_0(\boldsymbol{r}_{N-1}, \boldsymbol{r}_N)] \qquad [59]$$

has the same form as the Pfaffian in eqn [58]. Hence, Pf $1/(z_i - z_j)$ describes a $p$-wave pairing of electrons, and $\Psi^{\mathrm{Pf}}_{1/2}$ is interpreted as a paired state of composite fermions carrying two vortices.

### FQHE of Composite Fermions

Recently, some fractions other than those in eqn [50] have been observed, for example, $f = 4/11$ and $f = 5/13$. These are understood as the delicate "fractional" QHE of composite fermions at $\nu^* = 1 + 1/3$ and $\nu^* = 1 + 2/3$.

### TSG Effect in Higher Landau Levels

The short-range part of the Coulomb interaction is less effective in higher Landau levels because of the greater spread of the electron wave function. As a result, composite fermions are less stable, often losing to charge density wave states. A few fractions have been observed in the second Landau level (1/3, 2/3, 2/5, 1/2) and one (1/3) in the third.

### Edge states

There is a gap to excitations in the bulk at the magic fillings of eqn [50], but there is no gap at the edge of the sample. The dynamics of the low-energy edge excitations is formally equivalent to that of a chiral one-dimensional Tomonaga–Luttinger liquid. Wen (1991) argued that the exponent characterizing the long-distance behavior of this liquid is quantized, fully determined by the filling factor of the bulk state. Experimental studies of the tunneling of an external electron into the edge of an FQHE system provide evidence for a nontrivial

Tomonaga–Luttinger liquid but do not find the predicted universal value for the edge exponent.

## Acknowledgments

The author is grateful to the US National Science Foundation for financial support under grant no. DMR-0240458.

*See also:* Abelian Higgs Vortices; Aharonov–Bohm Effect; Chern–Simons Models: Rigorous Results; Fermionic Systems; Geometric Phases; Quantum Hall Effect; Quantum Phase Transitions; Quantum Statistical Mechanics: Overview.

## Further Reading

Chang AM (2003) Chiral Luttinger liquids at the fractional quantum Hall edge. *Reviews of Modern Physics* 75: 1449–1505.
Das Sarma S and Pinczuk A (eds.) (1997) *Perspectives in Quantum Hall Effects*. New York: Wiley.
Eisenstein JP and MacDonald AH (2004) Bose–Einstein condensation of excitons in bilayer electron systems. *Nature* 432: 691–694.
Giuliani GF and Vignale G (2005) *Quantum Theory of the Electron Liquid*. Cambridge: Cambridge University Press.
Halperin BI (2003) Composite fermions and the Fermion–Chern–Simons theory. *Physica E* 20: 71–78.
Halperin BI, Lee PA, and Read N (1993) Theory of the half-filled Landau level. *Physical Review B* 47: 7312–7343.
Heinonen O (ed.) (1998) *Composite Fermions*. New York: World Scientific.
Jain JK (1989) Composite-fermion approach for the fractional quantum Hall effect. *Physical Review Letters* 63: 199–202.
Jain JK (2000) The composite fermion: a quantum particle and its quantum fluids. *Physics Today* 53(4): 39–42.
Jain JK (2003) The role of analogy in unraveling the fractional quantum Hall effect mystery. *Physica E* 20: 79–88.
Kalmeyer V and Zhang SC (1992) Metallic phase of the quantum Hall system at even-denominator filling fractions. *Physical Review B* 46: 9889–9892.
Klitzing Kv (1986) Nobel lecture: The quantized Hall effect. *Reviews of Modern Physics* 58: 519–531.
Laughlin RB (1981) Quantized Hall conductivity in two dimensions. *Physical Review B* 23: 5632–5634.
Laughlin RB (1983) Anomalous quantum Hall effect: an incompressible quantum fluid with fractionally charged excitations. *Physical Review Letters* 50: 1395–1398.
Laughlin RB (1999) Nobel lecture: Fractional quantization. *Reviews of Modern Physics* 71: 863–874.
Leinaas JM and Myrheim J (1977) On the theory of identical particles. *Nuovo Cimento B* 37: 1–23.
Lopez A and Fradkin E (1991) Fractional quantum Hall effect and Chern–Simons gauge theories. *Physical Review B* 44: 5246–5262.
Moore G and Read N (1991) Nonabelions in the fractional quantum Hall effect. *Nuclear Physics B* 360: 362–396.
Murthy G and Shankar R (2003) Hamiltonian theories of the fractional quantum Hall effect. *Reviews of Modern Physics* 75: 1101–1158.
Stormer HL (1999) Nobel lecture: the fractional quantum Hall effect. *Reviews of Modern Physics* 71: 875–889.
Stormer HL, Tsui DC, and Gossard AC (1999) The fractional quantum Hall effect. *Reviews of Modern Physics* 71: S298–S305.
Tsui DC (1999) Nobel lecture: interplay of disorder and interaction in two-dimensional electron gas in intense magnetic fields. *Reviews of Modern Physics* 71: 891–895.
Wen XG (1992) Theory of the edge states in fractional quantum Hall effects. *International Journal of Modern Physics B* 6: 1711–1762.
Wilczek F (1982) Quantum mechanics of fractional-spin particles. *Physical Review Letters* 49: 957–959.
Zhang SC, Hansson H, and Kivelson S (1989) Effective-field-theory model for the fractional quantum Hall effect. *Physical Review Letters* 62: 82–85.

# Free Interfaces and Free Discontinuities: Variational Problems

**G Buttazzo**, Università di Pisa, Pisa, Italy



## Introduction

In several models coming from very different applications, one needs to describe physical phenomena where the state function may present some regions of discontinuity. We may think, for instance, of problems arising in fracture mechanics, where the function which describes the displacement of the body has a jump along the fracture, phase transitions, or also of problems of image reconstruction, where the function that describes a picture (the intensity of black, e.g., in black-and-white pictures) has naturally some discontinuities along the profiles of the objects.

The Sobolev space analysis is then no longer appropriate for this kind of problem, since Sobolev functions cannot have jump discontinuities along hypersurfaces, as, on the contrary, is required by the models above. For a rigorous presentation of variational problems involving functions with discontinuities, the essential tool is the space, BV, of functions with bounded variation. The first ideas about this space were developed by De Giorgi in the 1950s, in order to provide a variational framework to

study the problems of minimal surfaces, and several monographs are now available on the subject. We quote, for instance, the classical volumes of Evans and Gariepy (1992), Federer (1969), Giusti (1984), Massari and Miranda (1984), Ziemer (1989), and the recent book by Ambrosio *et al.* (2000), where a systematic presentation is given, also in view of the applications mentioned above.

## The Space BV

Consider a generic open subset $\Omega$ of $\mathbb{R}^N$, which, for simplicity, we take bounded and with a Lipschitz boundary. In the following, we denote by $\mathcal{L}^N(E)$, or simply $|E|$, the Lebesgue measure of $E$ in $\mathbb{R}^N$, while $\mathcal{H}^k$ denotes the $k$-dimensional Hausdorff measure.

**Definition 1** We say that a function $u \in L^1(\Omega)$ is a function of bounded variation in $\Omega$ if its distributional gradient Du is an $\mathbb{R}^N$-valued finite Borel measure on $\Omega$. In other words, we have

$$\int_\Omega u D_i \phi \, dx = -\int_\Omega \phi \, dD_i u \qquad \forall \phi \in C_c^\infty(\Omega), \quad \forall i = 1, \ldots, N \tag{1}$$

where $D_i u$ are finite Borel measures. The space of all functions of bounded variation in $\Omega$ is denoted by $BV(\Omega)$.

The space $BV(\Omega)$ is clearly a vector space and, with the norm

$$\|u\|_{BV(\Omega)} = \|u\|_{L^1(\Omega)} + |Du|(\Omega) \tag{2}$$

it becomes a Banach space. The total variation $|Du|(\Omega)$ appearing above is intended as

$$|Du|(\Omega) = \sup\left\{\sum_{i=1}^N \int_\Omega \phi_i \, dD_i u : \phi \in C_c^\infty(\Omega; \mathbb{R}^N), \ |\phi| \le 1\right\} = \sup\left\{-\int_\Omega u \, \mathrm{div}\phi \, dx : \phi \in C_c^\infty(\Omega; \mathbb{R}^N), \ |\phi| \le 1\right\}$$

and is sometimes indicated by $\int_\Omega |Du|$. The space $BV_{loc}(\Omega)$ is defined in a similar way, requiring that $u \in BV(\Omega')$ for every $\Omega' \subset\subset \Omega$.

From the point of view of functional analysis, the space $BV(\Omega)$ does not verify the nice properties of Sobolev spaces. In particular,

- the Banach space $BV(\Omega)$ is not separable;
- the Banach space $BV(\Omega)$ is not reflexive; and
- the class of smooth functions is not dense in $BV(\Omega)$ for the norm [2].

The above issues motivate why the norm [2] is not very helpful in the study of variational problems involving the space $BV(\Omega)$. On the contrary, the weak* convergence defined below is much more suitable to treat minimization problems for integral functionals.

**Definition 2** We say that a sequence $(u_n)$ weakly* converges in $BV(\Omega)$ to a function $u \in BV(\Omega)$ if $u_n \to u$ strongly in $L^1(\Omega)$ and $Du_n \to Du$ in the weak* convergence of measures.

The weak* convergence on $BV(\Omega)$ satisfies the following properties:

- *Compactness* Every bounded sequence in $BV(\Omega)$ for the norm [2] admits a weakly* convergent subsequence.
- *Lower-semicontinuity* The norm [2] is sequentially lower-semicontinuous with respect to the weak* convergence.
- *Density* Every function $u \in BV(\Omega)$ can be approximated, in the weak* convergence, by a sequence $(u_n)$ of smooth functions.

The density property above can be actually made stronger: in fact, the approximation of $(u_n)$ to $u$ holds in the sense that

$$\begin{cases} u_n \to u \text{ strongly in } L^1(\Omega) \\ Du_n \to Du \text{ weakly}^* \text{ as measures} \\ |Du_n|(\Omega) \to |Du|(\Omega) \end{cases}$$

Further properties of the space $BV(\Omega)$ concern the embeddings into Lebesgue spaces, traces, and Poincaré-type inequalities. More precisely, we have:

- *Embeddings* The space $BV(\Omega)$ is embedded continuously into $L^{N/(N-1)}(\Omega)$ and compactly into $L^p(\Omega)$ for every $p < N/(N-1)$.
- *Traces* Every function $u \in BV(\Omega)$ has a boundary trace which belongs to $L^1(\partial\Omega)$, and the trace operator from $BV(\Omega)$ into $L^1(\partial\Omega)$ is continuous.
- *Poincaré inequalities* There exist suitable constants $c_1$ and $c_2$ such that for every $u \in BV(\Omega)$

$$\int_\Omega |u| \, dx \le c_1 \left[|Du|(\Omega) + \int_{\partial\Omega} |u| \, d\mathcal{H}^{N-1}\right]$$

$$\int_\Omega |u - u_\Omega| \, dx \le c_2 |Du|(\Omega)$$

$$\left(\text{where } u_\Omega = \frac{1}{|\Omega|}\int_\Omega u \, dx\right)$$

## Sets of Finite Perimeter

An important class of functions with bounded variation are those that can be written as $1_E$, the characteristic function of a set $E$, taking the value 1 on $E$ and 0 elsewhere. This is the natural class where many phase-transition problems with sharp interfaces may be framed.

**Definition 3** For a measurable set $E \subset \mathbb{R}^N$ the perimeter of $E$ in $\Omega$ is defined as

$$\mathrm{Per}(E, \Omega) = |D1_E|(\Omega)$$

The equality above is intended as $\mathrm{Per}(E, \Omega) = +\infty$ whenever $1_E \notin \mathrm{BV}(\Omega)$. If $\mathrm{Per}(E, \Omega) < +\infty$ then the set $E$ is called a set of finite perimeter in $\Omega$.

Note that by the compactness property above for BV functions, a family of characteristic functions of sets with finite perimeter in a bounded open set $\Omega$ with equibounded perimeter is weakly$^*$-precompact, and its limit is of the same form.

For a set $E$ of finite perimeter in $\Omega$, we may define the inner normal versor and the reduced boundary as follows.

**Definition 4** Let $E$ be a set of finite perimeter in $\Omega$. We call reduced boundary $\partial^* E$ the set of all points $x \in \Omega \cap \mathrm{spt}|D1_E|$ such that the limit

$$\nu_E(x) = \lim_{r \to 0} \frac{D1_E(B_r(x))}{|D1_E|(B_r(x))}$$

exists and satisfies $|\nu_E(x)| = 1$. The vector $\nu_E(x)$ is called the generalized inner normal versor to $E$.

In order to link the measure-theoretical objects introduced above with some structure property of sets of finite perimeter, we introduce, for every $t \in [0, 1]$ and every measurable set $E \subset \mathbb{R}^N$, the set $E^t$ defined by

$$E^t = \left\{ x \in \mathbb{R}^N \colon \lim_{r \to 0} \frac{|E \cap B_r(x)|}{|B_r(x)|} = t \right\} \qquad [3]$$

For instance, if $E$ is a smooth domain of $\mathbb{R}^N$, $E^1$ is the interior part of $E$, $E^0$ is its exterior part, while $E^{1/2}$ is the boundary $\partial E$.

The main properties of the reduced boundary and of the generalized inner normal versor are stated in the following result.

**Theorem 5** *Let E be a set of finite perimeter in* $\Omega$. *Then its reduced boundary* $\partial^* E$ *coincides* $\mathcal{H}^{N-1}$*-a.e. with the set* $E^{1/2}$ *introduced in Definition 3, and we have the equality*

$$\mathrm{Per}(E, \Omega) = \mathcal{H}^{N-1}(\Omega \cap \partial^* E) = \mathcal{H}^{N-1}(\Omega \cap E^{1/2})$$

*Moreover, the generalized inner normal versor* $\nu_E(x)$ *exists for* $\mathcal{H}^{N-1}$*-a.e.* $x \in \partial^* E$, *and we have*

$$D1_E = \nu_E(x) \mathcal{H}^{N-1} \llcorner \partial^* E$$

Note that the lower-semicontinuity of $|D1_E|(\Omega)$ entails the lower-semicontinuity of $E \mapsto \mathcal{H}^{N-1}(\Omega \cap \partial^* E)$ with respect to the weak$^*$-convergence of $1_E$. As a consequence, we may apply the direct methods of the calculus of variations to obtain, for example, existence of minimizers of

$$\min \left\{ \mathrm{Per}(E, \mathbb{R}^N) - \int_E g \, \mathrm{d}x \right\}$$

that are sets with prescribed mean curvature $g$. This lower-semicontinuity property can be further generalized, for example, as in the following result for anisotropic perimeters.

**Theorem 6** *Let* $\varphi : S^{N-1} \to \mathbb{R}$ *be a Borel function. The energy*

$$\int_{\Omega \cap \partial^* E} \varphi(\nu_E) \, \mathrm{d}\mathcal{H}^{N-1}$$

*is lower-semicontinuous with respect to the weak*$^*$*-convergence of* $1_E$ *in* $\mathrm{BV}(\Omega)$ *if and only if the positively one-homogeneous extension of* $\varphi$ *from* $S^{N-1}$ *to* $\mathbb{R}^N$ *is convex.*

This result immediately implies the existence of solutions of isovolumetric problems of the form

$$\min \left\{ \int_{\partial^* E} \varphi(\nu_E) \, \mathrm{d}\mathcal{H}^{N-1} \colon \ |E| = c \right\}$$

whose solutions are obtained by suitably scaling the Wulff shape of $\varphi$.

## The Structure of BV Functions

The simplest situation occurs when $N = 1$ and so $\Omega$ is an interval of the real line. In this case, decomposing the derivative $u'$ into positive and negative parts, and taking their primitives, we obtain that $u \in \mathrm{BV}(\Omega)$ if and only if $u$ is the sum of two bounded monotone functions (one increasing and one decreasing). Therefore, in the one-dimensional case, the BV functions share all the properties of monotone functions.

The situation is more delicate when $N > 1$, for which we need the notion of approximate limit.

**Definition 7** Let $u \in \mathrm{BV}(\Omega)$. We say that $u$ has the approximate limit $z$ at $x$ if

$$\lim_{r \to 0} \frac{1}{|B_r(x)|} \int_{B_r(x)} |u(y) - z| \, \mathrm{d}y = 0$$

The set where no approximate limit exists is called the approximate discontinuity set, and is denoted by $S_u$. In a similar way, when $x \in S_u$ we may define the approximate values $z^+$ and $z^-$, by requiring that

$$\lim_{r\to 0} \frac{1}{|B_r^+(x,\nu)|} \int_{B_r^+(x,\nu)} |u(y) - z^+|\,\mathrm{d}y = 0$$
$$\lim_{r\to 0} \frac{1}{|B_r^-(x,\nu)|} \int_{B_r^-(x,\nu)} |u(y) - z^-|\,\mathrm{d}y = 0$$

where

$$B_r^+(x,\nu) = \{y \in B_r(x):\ (y-x)\cdot\nu > 0\}$$
$$B_r^-(x,\nu) = \{y \in B_r(x):\ (y-x)\cdot\nu < 0\}$$

Analogous definitions can be given in the vector-valued case, when $u \in \mathrm{BV}(\Omega;\mathbb{R}^m)$.

The triplet $(z^+, z^-, \nu)$ in Definition 7 is unique up to interchanging $z^+$ with $z^-$ and changing sign to $\nu$, and is denoted by $(u^+(x), u^-(x), \nu_u(x))$.

We are now in a position to describe the structure of the measure $Du$ when $u \in \mathrm{BV}(\Omega)$, or more generally $u \in \mathrm{BV}(\Omega;\mathbb{R}^m)$. We first apply the Radon–Nikodym theorem to $Du$ and we decompose it into absolutely continuous and singular parts: $Du = (Du)^{\mathrm{a}} + (Du)^{\mathrm{s}}$. We denote by $\nabla u$ the density of the absolutely continuous part, so that we have

$$Du = \nabla u \cdot \mathcal{L}^N + (Du)^{\mathrm{s}}$$

The singular part $(Du)^{\mathrm{s}}$ can be further decomposed into an $(N-1)$-dimensional part, concentrated on the approximate discontinuity set $S_u$, and the remaining part, which vanishes on all sets with finite $\mathcal{H}^{N-1}$ measure. More precisely, if $u \in \mathrm{BV}(\Omega;\mathbb{R}^m)$, we have

$$Du = \nabla u \cdot \mathcal{L}^N + (u^+(x) - u^-(x)) \otimes \nu_u(x) \cdot \mathcal{H}^{N-1} \llcorner S_u + (Du)^{\mathrm{c}} \qquad [4]$$

the three terms on the right-hand side are mutually singular and are, respectively, called the absolutely continuous part, the jump part, and the Cantor part of the gradient measure $Du$.

In the vector-valued case, $Du$ is an $m \times N$ matrix of finite Borel measures, $\nabla u$ is an $m \times N$ matrix of functions in $L^1(\Omega)$, and the jump term in [4] is an $(N-1)$-dimensional measure of rank 1. The structure of the Cantor part $(Du)^{\mathrm{c}}$ is described by the Alberti's rank-1 theorem (see Alberti (1993)).

**Theorem 8** *For every $u \in \mathrm{BV}(\Omega;\mathbb{R}^m)$ the Cantor part $(Du)^{\mathrm{c}}$ is a measure with values in the $m \times N$ matrices of rank 1.*

## Convex Functionals on BV

Many problems of the calculus of variations deal with the minimization of energies of the form

$$F(u) = \int_\Omega f(x, u, Du)\,\mathrm{d}x \qquad [5]$$

The direct methods to obtain the existence of at least a minimizer require some coercivity hypotheses on $F$, as well as its lower-semicontinuity. This last issue, already rather delicate when working in Sobolev spaces (see, e.g., Buttazzo (1989) and Dacorogna (1989)), presents additional difficulties when the unknown function $u$ varies in the space $\mathrm{BV}(\Omega)$, due to the fact that $Du$ is a measure, and the precise meaning of the integral in [5] has to be clarified.

In this section, we limit ourselves to consider the simpler situation of convex functionals, and we also assume that the integrand $f(x, u, Du)$ depends only on $x$ and $Du$. It is then convenient to study the problem in the framework of functionals defined on the space of finite Borel vector measures $\mathcal{M}(\Omega;\mathbb{R}^k)$. Let $f: \mathbb{R}^N \times \mathbb{R}^k \to [0, +\infty]$ be a Borel function such that

- $f$ is lower-semicontinuous, and
- $f(x,\cdot)$ is convex for every $x \in \mathbb{R}^N$.

We denote by $f^\infty(x,z)$ the recession function associated with $f$, given by

$$f^\infty(x,z) = \lim_{t\to+\infty} \frac{f(x, z_0 + tz)}{t}$$

where $z_0$ is any point in $\mathbb{R}^k$ such that $f(x,z_0) < +\infty$ (in fact, the definition above is independent of the choice of $z_0$). Then we may consider the functional

$$F(\lambda) = \int_\Omega f(x, \lambda^{\mathrm{a}}(x))\,\mathrm{d}x + \int_\Omega f^\infty\left(x, \frac{\mathrm{d}\lambda^{\mathrm{s}}}{\mathrm{d}|\lambda^{\mathrm{s}}|}\right)\mathrm{d}|\lambda^{\mathrm{s}}| \quad [6]$$

where $\lambda = \lambda^{\mathrm{a}} \cdot \mathrm{d}x + \lambda^{\mathrm{s}}$ is the Lebesgue–Nikodym decomposition of $\lambda$ into absolutely continuous and singular parts, and the notation $\mathrm{d}\lambda^{\mathrm{s}}/\mathrm{d}|\lambda^{\mathrm{s}}|$ stands for the density of $\lambda^{\mathrm{s}}$ with respect to its total variation $|\lambda^{\mathrm{s}}|$. For simplicity, the last term on the right-hand side of [6] is often denoted by $\int_\Omega f^\infty(x,\lambda^{\mathrm{s}})$.

For the functional $F$, the following lower-semicontinuity result holds (see, e.g., Buttazzo (1989)).

**Theorem 9** *Under the assumptions above the functional [6] is sequentially lower-semicontinuous for the weak* convergence on $\mathcal{M}(\Omega;\mathbb{R}^k)$. Moreover, if*

$$f(x,z) \geq c_0|z| - a(x) \quad \text{with } c_0 > 0 \text{ and } a \in L^1(\Omega) \qquad [7]$$

*then the functional F turns out to be coercive for the same topology.*

From Theorem 9 we deduce immediately a lower-semicontinuity result for functionals defined on $BV(\Omega; \mathbb{R}^m)$.

**Corollary 10** *Under the assumptions above on the integrand f(with k = mN) the functional defined on* $BV(\Omega; \mathbb{R}^m)$ *by*

$$F(u) = \int_\Omega f(x, (Du)^a)\,dx + \int_\Omega f^\infty\left(x, \frac{d(Du)^s}{d|Du|^s}\right) d|Du|^s \qquad [8]$$

*is sequentially lower-semicontinuous for the weak* convergence. Moreover, under the assumption [7] the functional F is coercive with respect to the same topology.*

For some extensions of the result above to the case when $f(x, \cdot)$ is quasiconvex (in the vector-valued situation $m > 1$), we refer the interested reader to Fonseca and Müller (1992) and references therein.

Fixing boundary data is another difference between variational problems on Sobolev spaces and on BV spaces. Due to the fact that the class $\{u \in BV(\Omega): u = u_0 \text{ on } \partial\Omega\}$ is not weakly* closed, to set in a correct way a minimum problem of Dirichlet type on $BV(\Omega)$ with datum $u_0 \in BV(\mathbb{R}^N)$ it is convenient to consider a larger domain $\Omega' \supset\supset \Omega$ and for every $u \in BV(\Omega)$ the extended function

$$\tilde{u} = \begin{cases} u & \text{on } \Omega \\ u_0 & \text{on } \Omega' \setminus \Omega \end{cases}$$

whose distributional gradient is

$$D\tilde{u} = Du \llcorner \Omega + Du_0 \llcorner \Omega' \setminus \bar{\Omega} + (u_0 - u)\nu_\Omega \mathcal{H}^{N-1} \llcorner \partial\Omega$$

$\nu_\Omega$ being the exterior normal versor to $\Omega$. We have then the following functional on $BV(\Omega')$:

$$\begin{aligned}\tilde{F}(\tilde{u}) &= \int_{\Omega'} f(x, (D\tilde{u})^a)\,dx + \int_{\Omega'} f^\infty(x, (D\tilde{u})^s) \\ &= \int_\Omega f(x, (Du)^a)\,dx + \int_{\Omega'\setminus\Omega} f(x, (Du_0)^a)\,dx \\ &\quad + \int_\Omega f^\infty(x, (Du)^s) + \int_{\Omega'\setminus\bar{\Omega}} f^\infty(x, (Du_0)^s) \\ &\quad + \int_{\partial\Omega} f^\infty(x, (u_0 - u)\nu_\Omega)\,d\mathcal{H}^{N-1}\end{aligned}$$

If we drop the constant term

$$\int_{\Omega'\setminus\Omega} f(x, (Du_0)^a)\,dx + \int_{\Omega'\setminus\bar{\Omega}} f^\infty(x, (Du_0)^s)$$

irrelevant for the minimization, we end up with the functional

$$F_{u_0}(u) = F(u) + \int_{\partial\Omega} f^\infty(x, (u_0 - u)\nu_\Omega)\,d\mathcal{H}^{N-1}$$

where $F$ is as in [8]. The Dirichlet problem we consider is then

$$\min\left\{F(u) + \int_{\partial\Omega} f^\infty(x, (u_0 - u)\nu_\Omega) d\mathcal{H}^{N-1}: u \in BV(\Omega)\right\} \qquad [9]$$

For instance, if $f(z) = |z|$, problem [9] becomes

$$\min\left\{\int_\Omega |Du| + \int_{\partial\Omega} |u - u_0|\,d\mathcal{H}^{N-1}: u \in BV(\Omega)\right\}$$

Under the assumptions considered, the problem above admits a solution $u \in BV(\Omega)$, but in general we do not have $u = u_0$ on $\partial\Omega$ in the sense of BV traces.

## Nonconvex Functionals on BV

In order to introduce the class of nonconvex functionals on $BV(\Omega)$, let us denote $v = Du$ so that every functional $\Phi(v)$ provides an energy $F(u)$. If we work in the setting of Sobolev spaces, we have $u \in W^{1,p}(\Omega)$ $(p \geq 1)$, which implies $v \in L^p(\Omega; \mathbb{R}^N)$; now, it happens that in this case all "interesting" functionals $\Phi$ are convex. More precisely, it can be proved that a functional $\Phi : L^p(\Omega; \mathbb{R}^N) \to [0, +\infty]$, which is

- sequentially lower-semicontinuous for the weak convergence of $L^p(\Omega; \mathbb{R}^N)$, and
- local on $L^p(\Omega; \mathbb{R}^N)$ in the sense that $\Phi(v + w) = \Phi(v) + \Phi(w)$ whenever $v \cdot w \equiv 0$ in $\Omega$,

has to be necessarily convex, and of the form

$$\Phi(v) = \int_\Omega \phi(x, v(x))\,dx$$

for a suitable integrand $\phi$ such that $\phi(x, \cdot)$ is convex. Then the energies $F(u)$ defined on Sobolev spaces and obtained by a functional $\Phi(v)$ through the identification $v = Du$ are necessarily convex. This is no longer true if $\Phi$ is defined on the space $\mathcal{M}(\Omega; \mathbb{R}^N)$ of measures, and hence $F$ is defined on $BV(\Omega)$. The first example of a nonconvex functional $\Phi$ on $\mathcal{M}(\Omega; \mathbb{R}^N)$ in the literature comes from the so-called Mumford–Shah model for computer vision (see below) and is given by

$$\Phi(\lambda) = \int_\Omega |\lambda^a(x)|^2 dx + \#(A_\lambda)$$

where $\lambda^a$ is the absolutely continuous part of $\lambda, A_\lambda$ is the set of atoms of $\lambda$, and $\#$ is the counting measure. The functional $\Phi$ is set equal to $+\infty$ on all measures $\lambda$ whose singular part $\lambda^s$ is nonatomic. A general representation result (see Bouchitte and Buttazzo (1992) and references therein) establishes that a functional $\Phi : \mathcal{M}(\Omega; \mathbb{R}^N) \to [0, +\infty]$, which is

- sequentially lower-semicontinuous for the weak$^*$ convergence of $\mathcal{M}(\Omega; \mathbb{R}^N)$, and
- local on $\mathcal{M}(\Omega; \mathbb{R}^N)$ in the sense that $\Phi(\lambda + \nu) = \Phi(\lambda) + \Phi(\nu)$ whenever $\lambda$ and $\nu$ are mutually singular in $\Omega$,

has to be of the form

$$\Phi(\lambda) = \int_\Omega \phi(x, \lambda^a)\, d\mu + \int_\Omega \phi^\infty(x, \lambda^c) + \int_\Omega \psi(x, \lambda^\#(x))\, d\#$$

where $\mu$ is a non-negative measure, $\lambda = \lambda^a \cdot dx + \lambda^c + \lambda^\#$ is the decomposition of $\lambda$ into absolutely continuous, Cantor, and atomic parts, $\phi(x, v)$ is an integrand convex in $v$, and $\phi^\infty$ is its recession function. The novelty is now represented by the integrand $\psi(x, v)$ which has to be subadditive in $v$ and satisfying the compatibility condition

$$\lim_{t \to +\infty} \frac{\phi(x, tv)}{t} = \lim_{t \to 0^+} \frac{\psi(x, tv)}{t}$$

When $\phi$ has a superlinear growth the condition above gives that the slope of $\psi(x, \cdot)$ at the origin has to be infinite. For instance, in the Mumford–Shah case we have

$$\phi(x, v) = |v|^2, \quad \psi(x, v) = \begin{cases} 1 & \text{if } v \neq 0 \\ 0 & \text{if } v = 0 \end{cases} \qquad [10]$$

Coming back to the case $u \in \mathrm{BV}(\Omega)$, we have the decomposition (see [4]):

$$Du = \nabla u \cdot \mathcal{L}^N + (Du)^c + [u]\nu_u(x) \cdot \mathcal{H}^{N-1} \llcorner S_u$$

where we considered, for simplicity, only the scalar case $m = 1$ and denoted by $[u]$ the jump $u^+ - u^-$. We have then the functional

$$F(u) = \int_\Omega \phi(x, \nabla u)\, dx + \int_\Omega \phi^\infty(x, (Du)^c) + \int_{S_u} \psi(x, [u]\nu_u)\, d\mathcal{H}^{N-1}$$

For instance, in the homogeneous–isotropic case, when $\phi(x, v)$ and $\psi(x, v)$ are independent of $x$ and depend only on $|v|$, the formula above reduces to

$$F(u) = \int_\Omega \phi(|\nabla u|) dx + \beta |Du|^c(\Omega) + \int_{S_u} \psi(|[u]|) d\mathcal{H}^{N-1} \qquad [11]$$

where $\beta, \phi, \psi$ satisfy the compatibility condition

$$\beta = \phi^\infty(1) = \lim_{t \to 0^+} \frac{\psi(t)}{t} \qquad [12]$$

In the original Mumford–Shah model for computer vision, $\Omega$ is a rectangle of the plane, $u_0 : \Omega \to [0, 1]$ represents the gray level of a picture, $c_1$ and $c_2$ are positive scale and contrast parameters, and the variational problem under consideration is

$$\min\left\{ \int_\Omega |\nabla u|^2 dx + c_1 \int_\Omega |u - u_0|^2 dx + c_2 \mathcal{H}^{N-1}(S_u) \colon (Du)^c \equiv 0 \right\} \qquad [13]$$

The solution $u$ then represents the reconstructed image, whose contours are given by the jump set $S_u$. We refer to Giorgi and Ambrosio (1988) and to the book by Morel and Solimini (1995) for further details about this model.

Analogously, in the case of the study of fractures of an elastic membrane, a problem similar to [13] provides the vertical displacement $u$ of the membrane, together with its fracture set $S_u$. We refer to some recent papers (see Dal Maso and Toader (2002) and Francroft and Marigo (1998), and references therein) for a more detailed description of fracture mechanics problems, even in the more delicate vectorial setting of elasticity.

Using the functional $F$ in [11] we have the generalized Mumford–Shah problem,

$$\min\left\{ F(u) + c_1 \int_\Omega |u - u_0|^2 dx : u \in \mathrm{BV}(\Omega) \right\}$$

where $\phi$ is convex, $\psi$ is subadditive, and the compatibility condition [12] is fulfilled.

If we set $K = S_u$ and assume that it is closed, the Mumford–Shah problem can be rewritten as

$$\min\left\{ \int_{\Omega \setminus K} |\nabla u|^2 dx + c_1 \int_{\Omega \setminus K} |u - u_0|^2 dx + c_2 \mathcal{H}^{N-1}(K \cap \Omega) : K \subset \bar{\Omega} \text{ closed}, u \in H^1(\Omega \setminus K) \right\}$$

and this justifies the name "free discontinuity problems," which is often used in this setting.

The regularity properties of optimal pairs $(u, K)$ are far from being fully understood; some partial results are available but the Mumford–Shah conjecture:

- in the case $N = 2$ for an optimal pair $(u, K)$ the set $K$ is locally the finite union of $C^{1,1}$ arcs

remains still open. We refer to Ambrosio *et al.* (2000) for a list of the regularity results on the problem above that are known thus for.

## Further Reading


Alberti G (1993) Rank one properties for derivatives of functions with bounded variation. *Proc. Roy. Soc. Edinburgh A* 123: 239–274.

Ambrosio L, Fusco N, and Pallara D (2000) *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Mathematical Monographs. Oxford: Clarendon.

Bouchitte G and Buttazzo G (1992) Integral representation of nonconvex functionals defined on measures. *Ann. Inst. H. Poincaré Anal. Non Linéaire* 9: 101–117.

Buttazzo G (1989) Semicontinuity, *Relaxation and Integral Representation in the Calculus of Variations*. Pitman Res. Notes Math. Ser., vol. 207. Harlow: Longman.

Dacorogna B (1989) *Direct Methods in the Calculus of Variations*. Appl. Math. Sciences. vol. 78. Berlin: Springer.

Dal Maso G and Toader R (2002) A model for the quasi-static growth of brittle fractures: existence and approximation results. *Arch. Ration. Mech. Anal.* 162: 101–135.

De Giorgi E and Ambrosio L (1988) New functionals in the calculus of variations. *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur.* 82(2): 199–210.

Evans LC and Gariepy RF (1992) *Measure Theory and Fine Properties of Functions*. Studies in Advanced Math Ann. Harbor: CRC Press.

Federer H (1969) *Geometric Measure Theory*. Berlin: Springer.

Fonseca I and Müller S (1992) Quasi-convex integrands and lower semicontinuity in $L^1$. *SIAM J. Math. Anal.* 23: 1081–1098.

Francfort G and Marigo J-J (1998) Revisiting brittle fracture as an energy minimization problem. *J. Mech. Phys. Solids* 46: 1319–1342.

Giusti E (1984) *Minimal Surfaces and Functions of Bounded Variation*. Boston: Birkhäuser.

Massari U and Miranda M (1984) *Minimal Surfaces of Codimension One*. Amsterdam: North-Holland.

Morel J-M and Solimini S (1995) *Variational Methods in Image Segmentation*. Progress in Nonlinear Differential Equations and their Applications, vol. 14. Boston: Birkhäuser.

Ziemer WP (1989) *Weakly Differentiable Functions*. Berlin: Springer.


# Free Probability Theory


**D-V Voiculescu**, University of California at Berkeley, Berkeley, CA, USA




## Introduction

Free probability is a probability theory adapted to quantities with the highest degree of noncommutativity. A basic feature of this is that the definition of independence is modified in such a way that the freely independent random variables will not commute in general. The exploration of this notion of independence, which was initially motivated by questions about operator algebras (Voiculescu 1985), has produced a theory that runs parallel to an unexpectedly large part of classical probability theory. The applications of the theory have also gone into unexpected directions, once it turned out that the large-$N$ limit of systems of random matrices is a key asymptotic model in the theory (Voiculescu 1991). There are several signs like the connections to large $N$ for random matrices and to the combinatorics of noncrossing partitions (Speicher 1998) (which correspond to certain planar diagrams), that perhaps these connections may go even further towards the large-$N$ limit of models in gauge theory.

In this article the noncommutative probability and the random matrix angle will be emphasized and very little will be said about the operator algebras and the combinatorics. After discussing free independence and models based on free products of groups and creation and annihilation operators on the Boltzmann full Fock space, we continue with the semicircle law, which is the substitute for the Gauss law in this context, and with the nonlinear free harmonic analysis arising from addition and multiplication of free random variables.

We then devote two longer sections to the asymptotic free independence of large random matrices and to free entropy, the free probability analog of Shannon's information-theoretic entropy for continuous random variables.

## Freeness of Noncommutative Random Variables

Classical probability deals with expectation values of numerical random variables, that is, with

numerical functions on a space of events and with their integrals with respect to a probability measure on the space of events. In noncommutative probability, the random variables, like quantum-mechanical quantities, are elements of a noncommutative algebra $A$ over $\mathbb{C}$, with unit $1 \in A$, which is endowed with a linear expectation functional $\varphi : A \to \mathbb{C}$, so that $\varphi(1) = 1$. Frequently, $A$ is a $*$-algebra of operators on some Hilbert space $\mathcal{H}$ and $\varphi(T) = \langle T\xi, \xi\rangle$ for some unit vector $\xi \in \mathcal{H}$. We call $(A,\varphi)$ a noncommutative probability space and the elements $a \in A$, noncommutative random variables. In this section we shall discuss the basics around the notion of freeness (Voiculescu 1985), which plays the role of independence in free probability.

If $\alpha = (a_i)_{i\in I} \subset A$ is a family of noncommutative random variables, the role of joint distribution is played by the collection of noncommutative moments $\varphi(a_{i_1} \dots a_{i_n})$. This can also be extended by linearity to a distribution functional $\Phi_\alpha : \mathbb{C}\langle X_i \mid i \in I\rangle \to \mathbb{C}$, where $\mathbb{C}\langle X_i \mid i \in I\rangle$ is the ring of polynomials in noncommutative indeterminates $X_i (i \in I)$ and

$$\Phi_\alpha(P(X_i|i \in I)) = \varphi(P(a_i|i \in I))$$

If $A$ is a $C^*$-algebra of operators on $\mathcal{H}$, $a = a^* \in A$ and $\varphi(\,\cdot\,) = \langle \cdot\,\xi, \xi\rangle$, the distribution of $a$ can also be identified with the probability measure $\mu_a$ on $\mathbb{R}$

$$\mu_a(\omega) = \langle E(\omega; a)\xi, \xi\rangle$$

where $E(\,\cdot\,; a)$ is the spectral measure of $a$. Indeed, then

$$\Phi_a(P(X)) = \int P(t)\mathrm{d}\mu_a(t)$$

A family $(A_i)_{i\in I} \subset A, 1 \in A_i$ of subalgebras is "free" (which is short for freely independent) if

$$\varphi(a_1 \dots a_n) = 0$$

whenever $a_j \in A_{i_j}, 1 \le j \le n$, $i_j \ne i_{j+1}$ and $\varphi(a_j) = 0$. (Here it is only required that consecutive $a_j$'s be in different $A_i$'s. Thus, we may have $i_1 = i_3$, provided $i_1 \ne i_2$.)

A family of sets of random variables $(\omega_i)_{i\in I}$, $\omega_i \subset A$ is free if the algebras $A_i$ generated by $1 \cup \{\omega_i\}$ are free in $(A,\varphi)$.

Except for rather trivial situations, free random variables in $(A,\varphi)$ do not commute.

Note also that, as in the case of classical independence, if $(\omega_i)_{i\in I}$ are disjoint freely independent sets of random variables, then, if the distributions $\Phi_{\omega_i}(i \in I)$ are given, the distribution $\Phi_\omega$ of $\omega = \bigcup_{i\in I} \omega_i$ is completely determined.

**Example 1** Let the group $G$ be the free product of its subgroups $(G_i)_{i\in I}$, that is, $G$ is generated by these subgroups and there is no nontrivial relation among elements of different $G_i$'s. Further, let $\lambda$ be the regular representation $\lambda(g)e_h = e_{gh}$ of $G$ on the Hilbert space with orthonormal basis $(e_g)_{g\in G}$. Then, with respect to the expectation functional $\tau(T) = \langle Te_e, e_e\rangle$ on operators on $l^2(G)$, the sets $(\lambda(G_i))_{i\in I}$ are freely independent.

**Example 2** If $\mathcal{H}$ is a complex Hilbert, let $\mathcal{TH} = \bigoplus_{k\ge 0} \mathcal{H}^{\otimes k}$ denote the full Boltzmann Fock space, with vacuum vector 1 so that $\mathcal{H}^{\otimes 0} = \mathbb{C}1$. If $h \in \mathcal{H}$ and $\xi \in \mathcal{TH}$, let $l(h)\xi = h \otimes \xi$ denote the left creation operator and $\varphi(X) = \langle X1, 1\rangle$ the vacuum expectation. Then, if the $\mathcal{H}_i (i \in I)$ are pairwise orthogonal subspaces in $\mathcal{H}$, the $*$-subalgebras of operators generated by $l(\mathcal{H}_i) \cup l^*(\mathcal{H}_i)$, indexed by $i \in I$, are freely independent with respect to $\varphi$.

## Free Independence with Amalgamation over a Subalgebra

The classical notion of conditional independence also has a free counterpart based on the notion of free independence with amalgamation over a subalgebra. This subject is technically more complicated and we will only aim at giving an idea about what kind of concepts are involved.

In the classical context, if $(X, \Sigma, \mu)$ is a probability space with a $\sigma$-algebra $\Sigma$, then the conditional independence with respect to a $\sigma$-subalgebra of events, $\Sigma_0 \subset \Sigma$, amounts to replacing in the definition of independence the expectation functional (which is the integral with respect to $\mu$) by the conditional expectation functional $L^\infty(X, \Sigma, \mu) \xrightarrow{E} L^\infty(X, \Sigma_0, \mu(\Sigma_0))$.

In free probability, one considers an extension of the theory, from the $(A,\varphi)$ framework to an $(A,\Phi, B)$ framework (Voiculescu 1995), where $A$ is an algebra with unit over $\mathbb{C}, B \ni 1$ is a subalgebra, and $\Phi : A \to B$ is $B$–$B$-bilinear and $\Phi|_B = \mathrm{id}_B$. Then the definition of $B$-freeness (or free independence with amalgamation over $B$) of a family of subalgebras $(A_i)_{i\in I}, B \subset A_i \subset A$ requires that

$$\Phi(a_1 \dots a_n) = 0$$

whenever $a_j \in A_{i_j}, i_j \ne i_{j+1}(1 \le j \le n)$, and $\Phi(a_j) = 0$.

In the case of a unital $*$-algebra of bounded operators $M$ with an expectation functional $\tau(\,\cdot\,) = \langle \cdot\,\xi, \xi\rangle$ which is tracial (i.e., $\tau([m_1, m_2]) = 0$ if $m_1, m_2 \in M$) and given a subalgebra $1 \in N \subset M$, as in the classical theory, there is a certain canonical construction in operator algebra theory of a "conditional expectation" $\Phi : \bar{M} \to \bar{N}$, where $\bar{M}, \bar{N}$ are

algebras of operators obtained as completion-separates from $M$ and $N$. With this construction, in the trace-state setting there is complete analogy with the classical notion of conditional independence.

Several other constructions of free probability have been extended to the $(A, \Phi, B)$ $B$-valued context.

A group-theoretic example similar to Example 1 can be constructed from a group $G$ which is a free product with amalgamation over a subgroup $H \subset G$ of subgroups $H \subset G_i \subset Gi \in I$. Then $A$ is the algebra constructed from the left-regular representation of $G$, whereas $B$ is an algebra constructed from the left-regular representation of $H$.

## The Semicircle Law

In free probability the semicircle law appears as the limit law in the free central limit theorem (Voiculescu 1985). Here is a weak, rather algebraic, version of this fact:

If $(a_n)_{n \in \mathbb{N}}$ are freely independent in $(A, \varphi)$ and satisfy the conditions that

$$\varphi(a_n) = 0 (n \in \mathbb{N})$$
$$\lim_{N \to \infty} N^{-1} \sum_{1 \le n \le N} \varphi\left(a_n^2\right) = 1$$
$$\sup_{n \in \mathbb{N}} \left| \varphi\left(a_n^k\right) \right| = C_k < \infty (k \in \mathbb{N})$$

then, if $S_N = N^{-1/2} \sum_{1 \le n \le N} a_n$, we have the convergence of moments of the distribution of $S_N$ to the semicircle distribution

$$\lim_{N \to \infty} \varphi(S_N^k) = (2\sigma)^{-1} \int_{-2}^{2} t^k (4 - t^2)^{1/2} \, \mathrm{d}t$$

Thus, the semicircle law, given by the density $(2\pi)^{-1}(4 - t^2)^{1/2}$ on $[-2, 2]$ is the free analog of the (0,1) Gauss law.

Two coincidences involving the semicircle law should be noted.

The field operators $s(h) = 2^{-1}(l(h) + l(h)^*)$ on the Boltzmann Fock space (Example 2) have semicircle distributions with respect to the vacuum expectation $\epsilon(\cdot) = \langle \cdot 1, 1 \rangle$. It turns out that this goes farther: if $\mathcal{H} = \mathcal{H}_{\mathbb{R}} \otimes_{\mathbb{R}} \mathbb{C}$ is the complexification of a real Hilbert space, then the map $\mathcal{H}_{\mathbb{R}} \ni h \to s(h)$ is the analog in free probability of the Gaussian process over the Hilbert space $\mathcal{H}_{\mathbb{R}}$ (Voiculescu 1985). It is often called the semicircular process over $\mathcal{H}_{\mathbb{R}}$. This points to an important connection of free probability to the full Boltzmann statistics.

The other coincidence is that the semicircle law is well known as the Wigner limit distribution of eigenvalues of large Gaussian random matrices. As we shall see, this is a clue to a deep connection of free probability to the large-$N$ limit of random matrices (Voiculescu 1991).

## Free Convolution Operations

In classical probability theory, the distribution of the sum of two independent random variables is computed by the convolution product of their distributions. This has a free probability analog. If $a$,$b$ are free random variables in $(A, \varphi)$ with distributions $\mu_a, \mu_b : \mathbb{C}[X] \to \mathbb{C}$, then the joint distribution $\mu_{\{a,b\}}$ is completely determined by $\mu_a, \mu_b$ and in particular $\mu_{a+b}$, the distribution of $a + b$, also depends only on $\mu_a, \mu_b$. It follows that there is an additive free convolution operation $\boxplus$ on distributions so that $\mu_a \boxplus \mu_b = \mu_{a+b}$ whenever $a, b$ are free (Voiculescu 1985). The same can be done with multiplication replacing addition, and this defines the multiplicative free convolution operation $\boxtimes$ by the equation $\mu_a \boxtimes \mu_b = \mu_{ab}$, when $a, b$ are free (Voiculescu 1985). A slightly surprising feature of $\boxtimes$ is that in spite of noncommutativity of $a$ and $b$, the multiplicative operation $\boxtimes$ turns out to be commutative, which of course is obvious for $\boxplus$.

In the classical context, convolutions are bilinear operations which can be computed using integrals. The free convolutions are quite nonlinear and their computation is via another route, which can also be explained by a classical analogy. Classically, the logarithm of the Fourier transform linearizes convolution, that is,

$$\log \mathcal{F}(\mu * \nu) = \log \mathcal{F}(\mu) + \log \mathcal{F}(\nu)$$

and we may compute $\mu * \nu$ as the $(\log \mathcal{F})^{-1}$ of $\log \mathcal{F}(\mu) + \log \mathcal{F}(\nu)$. The linearizing transform for $\boxplus$ is the $R$-transform (Voiculescu 1986), which is obtained by the following procedure.

If $\mu : \mathbb{C}[X] \to \mathbb{C}$ is a distribution, let $G_\mu(z) = z^{-1} + \sum_{n \ge 1} \mu(X^n) z^{-n-1}$, which, in case $\mu$ is a compactly supported probability measure on $R$, is the Laurent series at $\infty$ of the Cauchy transform

$$\int \frac{\mathrm{d}\mu(t)}{t - z}$$

From this, one obtains, by inversion at $\infty$, the series $K_\mu$, so that $G_\mu(K_\mu(z)) = z$ and one defines $R_\mu(z) = K_\mu(z) - z^{-1}$, which is a power series in $z$. Then

$$R_{\mu \boxplus \nu} = R_\mu + R_\nu$$

In case the distribution corresponds to a measure, the formal inversion amounts to inverting an analytic function.

For the multiplicative operation $\boxplus$, it is more convenient to describe an analog of the Mellin transform, that is, no logarithm will be taken. This is the $S$-transform (Voiculescu 1991), obtained as follows.

If $\mu : \mathbb{C}[X] \to \mathbb{C}$ is a distribution with $\mu(X) \neq 0$, one forms $\psi_\mu(z) = \sum_{n\geq 1} \mu(X^n) z^n$ and its inverse $\chi_\mu$ so that $\psi_\mu(\chi_\mu(z)) = z$. Then

$$S_\mu(z) = z^{-1}(1+z)\chi_\mu(z)$$

has the property that

$$S_{\mu \boxtimes \nu} = S_\mu S_\nu$$

The free central limit theorem can be easily proved using the $R$-transform. Another easy application of the $R$-transform is to find the free analog of the Poisson law, that is,

$$\lim_{n\to\infty} ((1-a/n)\delta_0 + a/n\delta_1)^{\boxplus}$$

where $a > 0$. The free Poisson law is

$$\mu = \begin{cases} (1-a)\delta_0 + \nu & \text{if } 0 \leq a \leq 1 \\ \nu & \text{if } a > 1 \end{cases}$$

where $\nu$ has support $[(1-a^{1/2})^2, (1+a^{1/2})^2]$ and density $(2\pi t)^{-1}(4a - (t-(1+a))^2)^{1/2}$. This distribution is well known in random matrix theory as the Marchenko–Pastur distribution, again a coincidence pointing to a random matrix theory connection.

Because probability measures on $\mathbb{R}$ are distributions of self-adjoint operators and a sum of self-adjoint operators is again such an operator, the additive free convolution $\boxplus$ yields an operation on probability measures on $\mathbb{R}$. Similarly, it can be shown that $\boxtimes$ gives rise to operations on probability measures on $\{z \in \mathbb{C} \mid |z| = 1\}$ and on probability measures on $[0,\infty)$.

With the $R$-transform machinery at hand, the free analogs of many of the classical results around addition of independent random variables have been developed (we recommend Voiculescu (1998c) for a survey of these developments). This includes the classification of infinitely divisible laws (Levy–Khintchine type theorem), classification of stable laws, domains of attraction, and convolution semigroups. Note that the free laws are rather different from the classical ones, but the classification results are quite parallel, that is, the indexing parameters are almost the same. The situation is similar in the multiplicative context. As in the classical case, these results about laws yield in particular processes with independent increments, which in the free framework are free increments.

As in the classical setting, also in the free setting, convolution semigroups are connected to differential equations. In the additive free case, a semigroup is a family $(\mu_t)_{t\geq 0}$ of probability measures on $\mathbb{R}$, so that $\mu_{t+s} = \mu_t \boxplus \mu_s$. If $G(t,z)$ is the Cauchy transform of $\mu_t$ (which is an analytic function on the half-plane $\operatorname{Im} z > 0$), the equation (Voiculescu 1986) is a semilinear complex PDE:

$$\frac{\partial G}{\partial t} + R_{\mu_1}(G)\frac{\partial G}{\partial z} = 0$$

where $R_{\mu_1}$ is the $R$-transform of $\mu_1$. In particular, when $\mu_1$ is the semicircle law, $R_{\mu_1}(z) = \alpha z$ $\alpha > 0$ and the PDE is a complex Burgers equation in the upper half-plane.

## Noncrossing Partitions

The series expansion of the $R$-transform

$$R_\mu(z) = \sum_{n\geq 0} R_n(\mu) z^n$$

has as coefficients polynomials $R_n(\mu)$ in the moments $\mu(X^k)$. More precisely, assigning to $\mu(X^k)$ a degree $k, R_n(\mu)$ is a polynomial of degree $n$ and $R_n(\mu) - \mu(X^n) =$ polynomial in $\mu(X^k)$ with $k < n$. The linearization property of the $R$-transform implies that

$$R_n(\mu \boxplus \nu) = R_n(\mu) + R_n(\nu)$$

For classical convolution, polynomials with similar properties satisfying

$$C_n(\mu * \nu) = C_n(\mu) + C_n(\nu)$$

are called cummulants and satisfy

$$\log \mu(e^{zX}) = \sum_{n\geq 1} C_m(\mu) z^n$$

There are combinatorial formulas involving the lattice of all partitions of the set $\{1, \ldots, n\}$ which give the classical cummulants. For free cummulants, like $R_n(\mu)$ and generalizations of these, there are similar formulas provided the lattice of all partitions is replaced by the lattice NC($n$) of noncrossing partitions (Speicher 1998). A partition $\pi = (V_1, \ldots, V_m)$ of $\{1, \ldots, n\}$ is noncrossing if there are no $a < b < c < d$ so that $\{a, c\} \subset V_k, \{b, d\} \subset V_l$ and $k \neq l$.

More generally, a family $R^{(n)}(a_1, \ldots, a_n)$ of free cummulants, where $a_1, \ldots, a_n$ are in some $(A,\varphi)$, is defined recursively as follows (Speicher 1998). For n = 1, one has $R^{(1)}(a) = \varphi(a)$. If $\pi = (V_1, \ldots, V_m) \in$ NC($n$), where $V_k = \{i(1,k) < \cdots < i(n_k, k)\}$, we define

$$R[\pi](a_1, \ldots, a_n) = \prod_{1\leq k\leq m} R^{(|V_k|)}(a_{i(1,k)}, \ldots, a_{i(n_k,k)})$$

The recurrence relation for cummulants is then

$$\varphi(a_1 \dots a_n) = \sum_{\pi \in \mathrm{NC}(n)} R[\pi](a_1, \dots, a_n)$$

Note that the right-hand side involves only $R^{(k)}$'s with $k \leq n$ and that actually $R^{(n)}$ appears only in and is equal to $R[(\{1, \dots, n\})](a_1, \dots, a_n)$ (the coarsest partition).

A key property of $R^{(n)}(a_1, \dots, a_n)$ is that if $\{1, \dots, n\} = \alpha \amalg \beta$ and $(a_k)_{k \in \alpha}, (a_l)_{l \in \beta}$ are freely independent, then $R^{(n)}(a_1, \dots, a_n) = 0$.

If $\mu$ is the distribution of $a \in (A, \varphi)$, then the cummulants $R_n(\mu)$ are given by

$$R_n(\mu) = R^{(n)}(a, \dots, a)$$

The noncrossing condition on partitions corresponds to a planarity requirement for diagrams and as such is very suggestive of connections to planar diagrams occurring in the constant term of large-$N$ expansions from random matrix theory and more generally gauge theory.

For more details on the subject of noncrossing partitions, we refer the reader to the memoir by Speicher (1998).

## Asymptotic Freeness of Random Matrices

The explanation for the coincidences between certain laws in free probability and in random matrix theory is that freeness occurs asymptotically among random matrices in the large-$N$ limit (Voiculescu 1991).

Random matrices can be put in a noncommutative probability framework $(\mathcal{A}_N, \varphi_N)$, where $\mathcal{A}_N = L^{\infty - 0}(\Omega, \mathcal{M}_N; \mathrm{d}\sigma)$ (the $N \times N$ complex matrix-valued functions on the probability space $(\Omega, \mathrm{d}\sigma)$ which are $p$-integrable for all $p \in [1,\infty)$) and the expectation functional is

$$\varphi_N(X) = N^{-1} \int_\Omega \mathrm{tr}\, X(\omega) \mathrm{d}\sigma(\omega)$$

The basic example is provided by an $n$-tuple of Gaussian random matrices (Voiculescu 1991). Let

$$T_j^{(N)} = \left(a_{p,q;j}^{(N)}\right)_{1 \leq p, q \leq N} \in N, \quad 1 \leq j \leq n$$

where $a_{p,q;j}^{(N)} = a_{q,p;j}^{(N)}$ and the $a_{p,q;j}^{(N)} 1 \leq p \leq q \leq N$, $1 \leq j \leq n$ are $(0, N^{-1})$-Gaussian and independent. Then $(T_j^{(N)})_{1 \leq j \leq n}$ as $N \to \infty$ converges in noncommutative distribution to the freely independent $n$-tuple $(l(e_j) + l^*(e_j))_{1 \leq j \leq n}$ in the Boltzmann Fock space context of Example 2 for an orthonormal system $e_1, \dots, e_n \in \mathcal{H}$, that is, convergence of moments:

$$\lim_{N \to \infty} \varphi_N \left(T_{i_1}^{(N)} \cdots T_{i_k}^{(N)}\right) = \langle (l(e_{i_1}) + l^*(e_{i_1})) \cdots (l(e_{i_k}) + l^*(e_{i_k}))1, 1 \rangle$$

In particular, the limit variables $(l(e_j) + l^*(e_j))_{1 \leq j \leq n}$ are free.

More generally, asymptotic freeness of variables or sets of variables in $(\mathcal{A}_N, \varphi_N)$ can be defined without the existence of a limit distribution, that is, by requiring only that the freeness relations among noncommutative moments hold asymptotically as $N \to \infty$.

Note that in these random matrix questions, the joint classical distribution of an $n$-tuple of random matrices $(X_1^{(N)}, \dots, X_n^{(N)})$ in $\mathcal{A}_N$ is a probability measure on $(\mathcal{M}_N)^n$ which contains more information than the collection of noncommutative moments, which is the distribution of the noncommutative variables in $(\mathcal{A}_N, \varphi_N)$. In particular, for one random matrix the classical distribution gives the joint distribution of all entries, whereas the noncommutative distribution gives information only about the distribution of eigenvalues.

From the Gaussian $n$-tuple using operator techniques much more general asymptotic freeness results have been obtained. For instance (Voiculesu 1998b):

Let $(X_1^{(N)}, \dots, X_m^{(N)}, Y_1^{(N)}, \dots, Y_n^{(N)})$ be $(m+n)$-tuples of self-adjoint $N \times N$ random matrices with classical joint distribution $\mu_N$ on $(\mathcal{M}_N^{\mathrm{sa}})^{m+n}$. Assume that $\mu_N$ is invariant under the action of the unitary group $\mathrm{U}(N)$ which takes $(X_1, \dots, X_m, Y_1, \dots, Y_n)$ into $(X_1, \dots, X_m, \mathrm{U}Y_1\mathrm{U}^*, \dots, \mathrm{U}Y_n\mathrm{U}^*)$ and assume that there is a bound $R$ on the operator norms $\|X_j^{(N)}\|$ and $\|Y_j^{(N)}\|$ independent of $N$. Then the sets $\{X_1^{(N)}, \dots, X_m^{(N)}\}$ and $\{Y_1^{(N)}, \dots, Y_n^{(N)}\}$ are asymptotically free as $N \to \infty$.

Note that the uniform bound on the operator norms can be easily replaced by weaker conditions.

Once we know that certain random matrices are asymptotically free and that the large-$N$ limit in noncommutative distribution exists, the results of free probability apply. For instance, if $X^{(N)}$ and $Y^{(N)}$ are asymptotically free and have limit distributions $\mu$ and $\nu$, then the limit distribution of $X^{(N)} + Y^{(N)}$ and of $X^{(N)}Y^{(N)}$ are the free convolutions $\mu \boxplus \nu$ and, respectively, $\mu \boxtimes \nu$.

Free probability techniques have also been successful in dealing with other questions about the asymptotic behavior of random matrices.

If $T_1^{(N)}, \dots, T_n^{(N)}$ is an $n$-tuple of i.i.d. Hermitian Gaussian random, then the uniform operator norms

of polynomials in noncommutative indeterminates have the property that

$$\lim_{N\to\infty} \|P(T_1^{(N)}, \ldots, T_n^{(N)})\| = \|P(l(e_1) + l(e_1)^*, \ldots, l(e_n) + l(e_n)^*)\|$$

almost surely (Haagerup and Thorbjoernsen).

This result is a far-reaching generalization of the results about largest eigenvalues of one Gaussian random matrix. The use of operator-valued free random variables (with respect to certain subalgebra) was an essential ingredient in the proof. Also, in another direction, freeness of operator-valued free random variables was used to obtain a free probability treatment of Gaussian random band matrices and generalizations of these (Shlyakhtenko 1996).

Finally, quite recently, extensions of the free probability framework have appeared which are adapted to the study of fluctuations of systems of random matrices in the large-$N$ limit.

## Free Entropy

There are free probability analogs also for information-theoretic quantities (Voiculescu 1994, 1998a).

Let $(f_1, \ldots, f_n)$ be an $n$-tuple of classical numerical random variables the joint distribution of which has density $p(t_1, \ldots, t_n)$ with respect to the $n$-dimensional Lebesgue measure $\lambda_n$ on $\mathbb{R}^n$. The entropy quantity associated by Shannon to $(f_1, \ldots, f_n)$ is

$$H(f_1, \ldots, f_n) = -\int_{\mathbb{R}^n} p \log p \, d\lambda_n$$

The free analog of $H(f_1, \ldots, f_n)$ is the free entropy quantity $\chi(X_1, \ldots, X_n)$. Here $X_j = X_j^*$, $1 \le j \le n$, are noncommutative self-adjoint random variables in $(M,\tau)$, where $M$ is a $*$-algebra of bounded operators on a Hilbert space $\mathcal{H}$. The expectation functional in addition to the positivity properties, equivalent to the requirement that it can be defined by a unit vector $\tau(\cdot) = \langle \cdot\xi, \xi\rangle$, also has the property of a trace $\tau(XY) = \tau(YX)$ for all $X, Y \in M$. For instance, the noncommutative random variables arising from the large-$N$ limit of $n$-tuples of self-adjoint random matrices live in noncommutative probability frameworks $(M,\tau)$ of this kind.

There are two approaches to defining free entropy and, since there are only partial results about the equivalence of these approaches, the quantities obtained are denoted by $\chi(X_1, \ldots, X_n)$ (Voiculescu 1994) and $\chi^*(X_1, \ldots, X_n)$ (Voiculescu 1998a). The quantity $\chi$ is often referred to as the "microstates free entropy," its definition being inspired by the Boltzmann formula $S = k \log W$, whereas the other entropy, sometimes called "microstates-free free entropy," is obtained via a free probability analog of the Fisher information (Voiculescu 1998a).

The microstates used to define $\chi$ are matricial and the reason why this choice produced a quantity with the right behavior with respect to free independence can be found in the asymptotic freeness properties of random matrices.

Given $X_j = X_j^* \in M, 1 \le j \le n$ and $m \in \mathbb{N}, k \in \mathbb{N}$, $\epsilon > 0$ the microstates $\Gamma(X_1, \ldots, X_n; m, k, \epsilon)$ are $n$-tuples $(A_1, \ldots, A_n)$ of self-adjoint $k \times k$ matrices, such that, for noncommutative moments of order up to $m$, we have

$$|k^{-1}\mathrm{tr}_k(A_{i_1} \ldots A_{i_p}) - \tau(X_{i_1} \ldots X_{i_p})| < \epsilon$$

where $1 \le p \le m, 1 \le i_j \le n, 1 \le j \le p$.

One obtains $\chi(X_1, \ldots, X_n)$ by taking the infimum over $\epsilon > 0$ and $m \in \mathbb{N}$ of

$$\limsup_{k\to\infty}\left(k^{-2} \log \mathrm{vol}\, \Gamma(\ldots) + \frac{n}{2}\log k\right)$$

where vol is the volume on $(\mathcal{M}_k^{\mathrm{sa}})^n$ corresponding to the Hilbert–Schmidt norm Hilbert space structure (Voiculescu 1994).

When $n = 1$, there is a simple formula for $\chi(X)$. If $\mu$ is the probability measure on $\mathbb{R}$ which represents the distribution of $X = X^* \in M$ with respect to the expectation $\tau$, then

$$\chi(X) = \iint \log|s - t| d\mu(s) d\mu(t) + C$$

where the exact value of the constant $C$ is $3/4 + 1/2 \log 2\pi$.

For $n > 1$ there is no simple formula for $\chi(X_1, \ldots, X_n)$, but there are several properties which provide a better understanding of this quantity.

If $X_j$ are such that $\chi(X_j) > -\infty$, then

$$\chi(X_1, \ldots, X_n) = \chi(X_1) + \cdots + \chi(X_n)$$

if and only if $X_1, \ldots, X_n$ are freely independent in $(M, \tau)$. Clearly, this property of $\chi$ with respect to free independence is analogous to the property of $H(f_1, \ldots, f_n)$ with respect to classical independence.

Further, if $F_1, \ldots, F_n$ are power series in $n$ noncommuting indeterminates, there is a change-of-variable formula

$$\chi(F_1(X_1, \ldots, X_n), \ldots, F_n(X_1, \ldots, X_n)) = \log|\det|(\mathcal{J}(F)) + \chi(X_1, \ldots, X_n)$$

involving the Kadison–Fuglede positive determinant $|\det|$ and a certain noncommutative Jacobian $\mathcal{J}(F), F = (F_1, \ldots, F_n)$ defined in $\mathcal{M}_n \otimes M \otimes M^{\mathrm{op}}$, where $M^{\mathrm{op}}$ is the opposite algebra of $M$. (For

definitions and the many technical conditions under which this formula holds, see Voiculescu (1994).)

The free entropy $\chi$ also satisfies semicontinuity, subadditivity, and a semicircular bound (analogous to the classical Gaussian bound) properties.

An unexpected feature of $\chi$ is a degeneration of convexity. If the trace state $\tau$ is a convex combination $\tau = \theta\tau' + (1-\theta)\tau''$, where $\tau', \tau''$ are trace states and where $\tau' \neq \tau''$ on the algebra generated by $X_1, \ldots, X_n$, and $n > 1$, then

$$\chi(X_1, \ldots, X_n) = -\infty$$

(for a reference consult the survey Voiculescu (2002)).

With the free entropy at hand, an important variational problem can be formulated for the noncommutative distribution of an $n$-tuple of self-adjoint noncommutative random variables $T_1, \ldots, T_n$ in the tracial context. The quantity to be maximized is

$$\chi(T_1, \ldots, T_n) - \tau(P(T_1, \ldots, T_n))$$

where $P$ is a given self-adjoint polynomial in noncommutative indeterminates (see Voiculescu (2002) for comments on this problem). If $n = 1$, this is a classical problem for the logarithmic energy

$$\iint \log|s-t|\mathrm{d}\mu(s)\mathrm{d}\mu(t) - \int P(t)\mathrm{d}\mu(t)$$

where $\mu$ is a probability measure on $\mathbb{R}$.

To explain the second approach, based on Fisher information, we begin by recalling some facts about Fisher information in the classical context.

If $f$ is a numerical random variable with distribution given by the density $p(t)$ on $\mathbb{R}$, then

$$\text{Fisher}(f) = \int \left(\frac{-p'}{p}\right)^2 p\,\mathrm{d}t = \left\| \left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^* 1 \right\|_{L^2(R, p\,\mathrm{d}t)}$$

Here $\mathrm{d}/\mathrm{d}t$ is the differential operator defined on test functions in $L^2(R, p\,\mathrm{d}t)$. Then

$$\frac{p'}{p} = -\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^* 1$$

The classical connection to entropy is that the Fisher information is a derivative of the entropy when the variable becomes the starting point of a Brownian motion. This can be written as

$$\text{Fisher}(f) = \frac{\mathrm{d}}{\mathrm{d}t} H(f + t^{1/2}\mathfrak{g})|_{t=0}$$

where $\mathfrak{g}$ and $f$ are independent and $\mathfrak{g}$ is $(0,1)$ Gaussian.

The several-variables version is treated by using partial derivatives.

The analog in free probability of the Fisher information (Voiculescu 1998a) is obtained by using the free difference quotient derivations, which are the appropriate derivations in this maximally noncommutative setting. On the polynomials in $n$ noncommutative indeterminates, the $k$th partial free difference quotient

$$\partial_k : \mathbb{C}\langle X_1, \ldots, X_n\rangle \to \mathbb{C}\langle X_1, \ldots, X_n\rangle^{\otimes 2}$$

is defined on noncommutative monomials by the formula

$$\partial_k X_{i_1} \cdots X_{i_p} = \sum_{\{j | i_j = k\}} X_{i_1} \cdots X_{i_{j-1}} \otimes X_{i_{j+1}} \cdots X_{i_p}$$

If $X_j = X_j^*, 1 \leq j \leq n$, are noncommutative random variables in $(M, \tau)$, which do not satisfy any nontrivial algebraic relations, to simplify matters we can assume that $M$ is generated by $X_1, \ldots, X_n$ and identify $M$ with $\mathbb{C}\langle X_1, \ldots, X_n\rangle$. The trace state $\tau$ gives rise to a scalar product $\langle m_1, m_2\rangle = \tau(m_2^* m_1)$ on $M$. Let $L^2(M, \tau)$ denote the Hilbert space obtained from $M$. Then, skipping some technicalities, $\partial_k$ will give rise to a densely defined operator of $L^2(M, \tau)$ into $L^2(M, \tau) \otimes L^2(M, \tau)$. If $1 \otimes 1$ is in the domain of the adjoints $\partial_k^*$, the free Fisher information of the $n$-tuple $X_1, \ldots, X_n$ is defined to be

$$\Phi^*(X_1, \ldots, X_n) = \sum_{1 \leq k \leq n} \|\partial_k^*(1 \otimes 1)\|^2_{L^2(M, \tau)}$$

In case $1 \otimes 1$ is not in the domain of some $\partial_k^*$, the free Fisher information is given the value $\infty$.

The "microstates-free free entropy" $\chi^*$ is then defined by

$$\chi^*(X_1, \ldots, X_n) = \frac{n}{2}\log 2\pi e + \int_0^\infty \left(\frac{n}{1+t} - \Phi^* \times (X_1 + t^{1/2}S_1, \ldots, X_n + t^{1/2}S_n)\right)\mathrm{d}t$$

where $S_1, \ldots, S_n$ are $(0,1)$-semicircular and freely independent and also freely independent of $\{X_1, \ldots, X_n\}$.

For $n = 1$ it is known that $\chi^*(X) = \chi(X)$ and the free Fisher information is

$$\Phi^*(X) = \frac{2\pi^2}{3}\int p^3(t)\mathrm{d}t$$

if $p(t)$ is the density with respect to the Lebesgue measure of the distribution of $X$. The computation of $\partial^* 1 \otimes 1$ is possible in the one-variable case and up to a factor the result is $(Hp)(X)$, where $Hp$ is the Hilbert transform of $p$.

Several of the classical inequalities for the Fisher information have free probability analogs (Voiculescu 1998a) (Cramer–Rao inequality, Stam inequality, information-log-Sobolev inequality, and others).

For $n > 1$ only $\chi \leq \chi^*$, the easier of the inequalities among $\chi$ and $\chi^*$, has been established (Biane *et al.* 2003). This result was obtained based on an important connection of $\chi$ and $\chi^*$ to large deviations. The deviations studied are for the noncommutative distributions of $n$-tuples of matrices in the case of an $n$-tuple of Gaussian random matrices. In this context $\chi$ is related to the quantity to be estimated and $\chi^*$ is related to the rate function.

For more details on free entropy, the reader is referred to the survey articles by Voiculescu (1998c, 2002).

## Concluding Comments

For more details, additional results, and bibliography, we refer the reader to the expositions in Voiculescu (1998c), Voiculescu *et al.* (1992) and Speicher (1998). To get even more detail, the reader may consult, besides the original papers of the present author, those of P Biane, R Speicher, D Shlyakhtenko, K J Dykema, A Nica, U Haagerup, H Bercovici, L Ge, F Radulescu, A Guionnet, T Cabanal–Duvillard, M Anshelevich, to name a few of the main contributors.

Also, via random matrices, there are connections to physics models (especially large-$N$ 2D Yang–Mills QCD) in work of I M Singer, M Douglas, D Gross–R Gopakumar, P Zinn–Justin. In a loose sense, one may view the noncrossing partitions combinatorics as related to the work on planar diagrams and the large-$N$ limit of t'Hooft and Brezin–Itzykson–Parisi–Zuber in the 1970s.

*See also:* Large Deviations in Equilibrium Statistical Mechanics; Large-*N* and Topological Strings; Random Matrix Theory in Physics.

## Further Reading

Biane P, Capitaine M, and Guionnet A (2003) Large deviation bounds for matrix Brownian motion. *Inventiones Mathematicae* 152: 433–459.

Haagerup U and Thorbjoernsen S (2005) A new application of random matrices: $\mathrm{Ext}(C^*_r(F_2))$ is not a group. *Annals of Mathematics* 162(2).

Shlyakhtenko D (1996) Random Gaussian band matrices and freeness with amalgamation. *International Mathematics Research Notices* 20: 1013–1025.

Speicher R (1998) Combinatorial theory of the free product with amalgamation and operator-valued free probability. *Memoirs of the American Mathematical Society* 627.

Voiculescu D (1985) Symmetries of some reduced free product $C^*$-algebras. In: Araki H, Moore CC, Stratila S-V, and Voiculescu DV (eds.) *Operator Algebras and Their Connections with Topology and Ergodic Theory*, Lecture Notes in Mathematics,, vol. 1132, pp. 556–588. Springer.

Voiculescu D (1986) Addition of certain non-commuting random variables. *Journal of Functional Analysis* 66: 323–346.

Voiculescu D (1987) Multiplication of certain non-commuting random variables. *Journal of Operator Theory* 18: 223–235.

Voiculescu D (1991) Limit laws for random matrices and free products. *Inventiones Mathematicae* 104: 201–220.

Voiculescu D (1994) The analogues of entropy and of Fisher's information measure in free probability theory II. *Inventiones Mathematicae* 118: 411–440.

Voiculescu D (1995) Operations on certain noncommuting operator-valued random variables. *Asterisque* 232: 243–275.

Voiculescu D (1998a) The analogues of entropy and of Fisher's information measure in free probability theory V: Noncommutative Hilbert transforms. *Inventiones Mathematicae* 132: 182–227.

Voiculescu D (1998b) A strengthened asymptotic freeness result for random matrices with applications to free entropy. *International Mathematics Research Notices* 1: 41–63.

Voiculescu D (1998c) Lectures on free probability theory. In: *Lectures on Probability Theory and Statistics*, Ecole d'Ete de Probabilites de Saint-Flour XXVIII–1998, Lecture Notes in Mathematics,, vol. 1738, pp. 280–349. Berlin: Springer.

Voiculescu D (2002) Free entropy. *Bulletin of London Mathematical Society* 34: 257–278.

Voiculescu D, Dykema KJ, and Nica A (1992) *Free Random Variables*, CRM Monograph Series, vol. 1. Providence: American Mathematical Society.

**Frobenius Manifolds** *see* WDVV Equations and Frobenius Manifolds

# Functional Equations and Integrable Systems

**H W Braden**, University of Edinburgh, Edinburgh, UK



## Introduction

Functional equations have a long and interesting history in connection with mathematical physics and touch upon many branches of mathematics. They have arisen in the context of both classical and quantum completely integrable systems in several different ways and we shall survey some of these.

In the great majority of cases functional equations appear in the integrable system setting as the result of an ansatz: a particular form of a solution is either guessed or postulated, the consistency of which yields a functional equation. What the ansatz is for can vary significantly. As outlined below, amongst others, one may postulate algebraic structures in the form of the existence of a Lax pair or of conserved quantities; in the quantum setting, one may postulate properties of a ground-state wave function or the ring of commuting differential operators. Appearing in this way, functional equations are really just another of the (significant) tools-of-the-trade for constructing and discovering new integrable systems. However, as one surveys both the functional equations and the functions they describe one sees certain common features. The functions are most frequently associated with an elliptic curve, a genus-1 abelian variety. One can seek to associate these to another fundamental ingredient of modern integrable systems, the Baker–Akhiezer function. Indeed, very few of the ansätze made directly suggest that the systems being constructed will be completely integrable. This very desirable property usually is a bonus of the construction and hints of more fundamental connections. Another fundamental connection we shall mention is that with topology. The phase space of a completely integrable system is rather special, admitting (generically) a foliation by tori. The functional equations we encounter often also characterize the Hirzebruch genera associated with the index theorems of known elliptic operators. These are typically evaluated by Atiyah–Bott fixed-point theorems for circle actions on the manifold. A general understanding of the various interconnections has yet to be achieved.

To bring to focus our discussion we shall concentrate on functional equations arising from studying systems with an arbitrary number of particles ($n$ below). In principle, there could be many different interactions between the particles and symmetry will be used to limit these. The use of symmetry is a key ingredient, often implicit, in the various ansätze we shall describe. For simplicity, we shall most often focus on the situation where the particles are identical. In algebraic terms, we focus on the symmetric group $S_n$ and root systems of type $a_n$; generalizations frequently exist for other root systems and Weyl groups and we shall simply note this at the outset.

## Lax Pairs

The modern approach to integrable systems is to utilize a Lax pair, that is, a pair of matrices $L, M$ such that the zero curvature condition $\dot{L} = [L, M]$ is equivalent to the equations of motion. By construction, Lax pairs produce the conserved quantities tr $L^k$. To establish integrability, one must further show both that there are enough functionally independent conserved quantities and that these are in involution. ($R$-matrices are the additional ingredient of the modern approach to establishing involutivity.) Lax pairs can fail on both counts, and so the construction of a Lax pair is but the first step in establishing a system to be completely integrable. The great merit of the modern approach is that it provides a unified framework for treating the many disparate completely integrable systems known. Unfortunately the construction of a Lax pair is often far from straightforward and typically hides the "clever tricks" frequently employed in establishing integrability. In the present context, we shall outline how functional equations have been used to construct Lax pairs. The paradigm for this approach is the Calogero–Moser system.

Beginning with the ansatz (for $n \times n$ matrices)

$$L_{jk} = p_j \delta_{jk} + g(1 - \delta_{jk}) A(q_j - q_k)$$

$$M_{jk} = g \left[ \delta_{jk} \sum_{l \neq j} B(q_j - q_l) - (1 - \delta_{jk}) C(q_j - q_k) \right]$$

one finds $\dot{L} = [L, M]$ yields the equations of motion for the Hamiltonian system ($n \geq 3$)

$$H = \frac{1}{2} \sum_j p_j^2 + g^2 \sum_{j<k} U(q_j - q_k) \qquad [1]$$

$$U(x) = A(x)A(-x) + \text{const.}$$

provided $C(x) = -A'(x)$, and that $A(x)$ and $B(x)$ satisfy the functional equation

$$A(x+y)[B(x) - B(y)] = A(x)A'(y) - A(y)A'(x) \qquad [2]$$

This is a particular example of a more general functional equation whose solution will be described

below. For the present, we simply note that for this system the corresponding potential is the Weierstrass $\wp$-function, $A(x)A(-x) = \wp(\nu) - \wp(x)$, and the resulting Hamiltonian system [1] is known as the Calogero–Moser system. It is completely integrable though, as already remarked, the ansatz did not necessitate this. The Lax pair presented here and the reduction of its consistency to a functional equation and algebraic constraints follows Calogero (1976) in which he discovered the elliptic generalization of the model he had introduced in 1975.

A different ansatz for a Lax pair is

$$L_{jk} = \dot{q}_j \delta_{jk} + (1 - \delta_{jk})\sqrt{\dot{q}_j \dot{q}_k} A(q_j - q_k)$$

$$M_{jk} = \delta_{jk} \sum_{l \neq j} \dot{q}_l B(q_j - q_l) + (1 - \delta_{jk})\sqrt{\dot{q}_j \dot{q}_k} C(q_j - q_k)$$

Now the consistency of the Lax pair yields equations of motion of the form

$$\ddot{q}_j = \sum_{k \neq j} \dot{q}_j \dot{q}_k V(q_j - q_k)$$

$$V(x) = \begin{vmatrix} A(x) & A(-x) \\ C(x) & C(-x) \end{vmatrix} = -V(-x)$$

provided $B(x) = B(-x)$, $C(x) = A'(x) - A(x)G(x)$, where we have defined $G(x) = B(x) + (1/2)V(x)$, and the functions satisfy the functional equation

$$A(x+y) = A(x)A(y) + \frac{\begin{vmatrix} A(x) & A(y) \\ A'(x) & A'(y) \end{vmatrix}}{G(x) - G(y)} = \frac{\begin{vmatrix} A(x) & A(y) \\ C(x) & C(y) \end{vmatrix}}{G(x) - G(y)} \qquad [3]$$

Again we shall briefly defer describing the solution of this equation and simply note that the general solution for $V(x)$ is again given in terms of the Weierstrass $\wp$-function $V(x) = \wp'(x)/(\wp(\nu) - \wp(x))$ and that the equations of motion follow from the Hamiltonian

$$H = \sum_j e^{p_j} \prod_{k \neq j} \sqrt{\wp(\nu) - \wp(q_j - q_k)}$$

This is known as the Ruijsenaars–Schneider model and it too is completely integrable. The Lax pair here was constructed by Bruschi and Calogero.

In the two examples of Lax pairs just presented, each particle interacts with every other pairwise. By modifying the ansatz, it is possible to construct models that interact with just their nearest neighbors (which include the Toda systems). More generally, an ansatz exists for a Lax pair associated with equations of motion of the form

$$\ddot{q}_j = \sum_{k \neq j} (a + b\dot{q}_j)(a + b\dot{q}_k) V_{jk}(q_j - q_k) \qquad [4]$$

which unifies, for example, the Calogero–Moser, Ruijsenaars–Schneider, and Toda systems. The functional equations now encountered are typically (and whenever $b \neq 0$) of the form

$$\phi_1(x+y) = \frac{\begin{vmatrix} \phi_2(x) & \phi_2(y) \\ \phi_3(x) & \phi_3(y) \end{vmatrix}}{\begin{vmatrix} \phi_4(x) & \phi_4(y) \\ \phi_5(x) & \phi_5(y) \end{vmatrix}} \qquad [5]$$

This functional equation, for five *a priori* unknown functions, includes [2] and [3] as special cases.

The general analytic solution of [5] is, up to symmetries, given by

$$\phi_1(x) = \frac{\Phi(x;\nu_1)}{\Phi(x;\nu_2)}, \qquad \begin{pmatrix} \phi_2(x) \\ \phi_3(x) \end{pmatrix} = \begin{pmatrix} \Phi(x;\nu_1) \\ \Phi'(x;\nu_1) \end{pmatrix}$$

$$\begin{pmatrix} \phi_4(x) \\ \phi_5(x) \end{pmatrix} = \begin{pmatrix} \Phi(x;\nu_2) \\ \Phi'(x;\nu_2) \end{pmatrix}$$

where

$$\Phi(x;\nu) \equiv \frac{\sigma(\nu - x)}{\sigma(\nu)\sigma(x)} e^{\zeta(\nu)x} \qquad [6]$$

Here, $\zeta(x) = \sigma(x)'/\sigma(x)$ is the Weierstrass $\zeta$-function. The solution of [2] arises as the $\nu_2 \to 0$ limit of [5].

The proof of the general solution just stated is in fact constructive (Braden and Buchstaber 1997). The parameters appearing in the solution are determined as follows. Suppose $x_0$ is a generic point for [5]. Then (for $k = 1, 2$), we have that

$$\partial_y \ln \left| \begin{vmatrix} \phi_{2k}(x+x_0) & \phi_{2k}(y+x_0) \\ \phi_{2k+1}(x+x_0) & \phi_{2k+1}(y+x_0) \end{vmatrix} \right|_{y=0} = \zeta(\nu_k) - \zeta(x) - \zeta(\nu_k - x) - \lambda_k = -\frac{1}{x} - \lambda_k + \sum_{l=0} F_l \frac{x^{l+1}}{(l+1)!} \qquad [7]$$

The Laurent expansion determines the parameters $g_1, g_2$ (which are the same for both $k = 1, 2$) characterizing the elliptic functions of [6] by

$$g_2 = \tfrac{5}{3}(F_2 + 6F_0^2), \qquad g_3 = 6F_0^3 - F_1^2 + \tfrac{5}{3}F_0 F_2$$

and the parameters $\nu_k$ via $F_0 = -\wp(\nu_k)$. Here, $\wp(x) = -\zeta'(x)$ is the Weierstrass elliptic $\wp$-function

with periods $2\omega, 2\omega'$ that satisfies the differential equation

$$\wp'(x)^2 = 4\wp(x)^3 - g_2\wp(x) - g_3$$

The constructive nature of the solutions of [5] means that it is straightforward to construct solutions to various specializations of the equation such as

$$\phi_1(x+y) = \phi_4(x)\phi_5(y) + \phi_4(y)\phi_5(x)$$

(obtained by requiring $\phi_2(x) = \phi_4^2(x)$ and $\phi_3(x) = \phi_5^2(x)$). More complicated functional equations such as

$$\begin{aligned}\Psi_1(x+y) = &\Psi_2(x+y)\phi_2(x)\phi_3(y)\\ &+ \Psi_3(x+y)\phi_4(x)\phi_5(y)\end{aligned} \qquad [8]$$

may be solved using the solutions of [5].

Finally, let us note that the general system [4] may lead to functional equations not just of the form [5], for example,

$$\begin{aligned}\phi_6(x+y) = &\phi_1(x+y)(\phi_4(x) - \phi_5(y))\\ &+ \begin{vmatrix}\phi_2(x) & \phi_3(y)\\ \phi_2'(x) & \phi_3'(y)\end{vmatrix}\end{aligned} \qquad [9]$$

The general analytic solution to [9] has yet to be determined although particular solutions are known.

As a final example of a functional equation coming from an ansatz for a Lax pair, consider

$$L_{jk} = \sqrt{p_j p_k}A(q_j - q_k), \qquad M_{jk} = \sqrt{p_j p_k}C(q_j - q_k)$$

where we now assume $A(0)$ and $C(0)$ regular. Then the consistency of this Lax pair corresponds to the equations of motion for the Hamiltonian

$$H = \sum_{j,k}^{n} p_j p_k f(q_j - q_k) \qquad [10]$$

provided $f$ is even and the functional equation

$$\begin{aligned}&2A'(x+y)[f(x) - f(y)]\\ &\quad - A(x+y)[f'(x) - f'(y)] = \begin{vmatrix}A(x) & A(y)\\ C(x) & C(y)\end{vmatrix}\end{aligned} \qquad [11]$$

is satisfied. The Hamiltonian system [10] corresponds to geodesic motion. Nonanalytic solutions are known to the functional equation [11].

## An Algebraic Ansatz: Conserved Quantities

Another way in which functional equations may appear is by making an ansatz for an additional conserved quantity beyond the Hamiltonian. For two and three particles on the line, Hietarinta derived functional equations by seeking a second quartic or cubic integral (respectively). Here, a key ingredient is the assumption of a further invariant polynomial in the momenta. Polynomial invariance, together with symmetry, is quite constraining. Consider

**Theorem 1** *Let $H$ and $P$ be the (natural) Hamiltonian and center of mass momentum*

$$H = \frac{1}{2}\sum_{i=1}^{n} p_i^2 + V, \qquad P = \sum_{i=1}^{n} p_i$$

*Denote by $Q$ an independent third-order quantity*

$$Q = \sum_{i=1}^{n} p_i^3 + \frac{1}{6}\sum_{i\neq j\neq k} d_{ijk}p_i p_j p_k + \sum_{i\neq j} d_{ij}p_i^2 p_j + \frac{1}{2}\sum_{ij} a_{ij}p_i p_j + \sum_i b_i p_i + c$$

*If these are $S_n$-invariant and Poisson-commute,*

$$\{P, H\} = \{P, Q\} = \{Q, H\} = 0$$

*then*

$$V = \frac{1}{6}\sum_{i\neq j}\wp(q_i - q_j) + \text{const.}$$

*and we have the Calogero–Moser system.*

Here, the symmetric group invariance means that for any coefficient $\alpha_{ij}(q_1, q_2, \ldots, q_n)$ in the expansions above, we have $\alpha_{\sigma(i)\sigma(j)}(q_{\sigma(1)}, q_{\sigma(2)}, \ldots, q_{\sigma(n)})$ for all $\sigma \in S_n$. In particular, $V(q_1, q_2, \ldots, q_n) = V(q_{\sigma(1)}, q_{\sigma(2)}, \ldots, q_{\sigma(n)})$ for all $\sigma \in S_n$. We remark that had we begun with particles of possibly different particle masses, $H = (1/2)\ \sum_{i=1}^{n} m_i p_i^2 + V$; the effect of $S_n$-invariance is such as to require these masses to be the same. Thus, we are assuming the $S_n$-invariant Hamiltonian of the theorem. Finally, by "an independent third-order quantity" $Q$, we mean one functionally independent of $H$ and $P$ and for which one cannot obtain an invariant of lower degree by subtracting multiples of $P^3$ and $PH$. We are not dealing with quadratic conserved quantities here.

The assumed polynomial behavior of the conserved quantities means that when calculating Poisson brackets, the coefficients of independent monomials must vanish. This, together with symmetry, leads to the functional equation

$$\begin{vmatrix}1 & 1 & 1\\ F(x) & F(y) & F(z)\\ F'(x) & F'(y) & F'(z)\end{vmatrix} = 0, \quad x+y+z=0 \qquad [12]$$

The result follows in light of

**Theorem 2** *Let f be a three-times differentiable function satisfying the functional equation [12]. Up to the manifest invariance*

$$F(x) \to \alpha F(\delta x) + \beta$$

*the solutions of [12] are one of* $F(x) = \wp(x + d)$, $F(x) = e^x$ *or* $F(x) = x$. *Here,* $\wp$ *is the Weierstrass* $\wp$*-function and 3d is a lattice point of the* $\wp$*-function.*

Again we note that the ansatz *per se* has not established complete integrability: the ansatz leads us to the Calogero–Moser model whose complete integrability must be established by other means. This result may be interpreted as a rigidity theorem for the $a_n$ Calogero–Moser system and in part explains this models' ubiquity: demanding a cubic invariant together with $S_n$-invariance necessitates the model. A natural generalization is to replace the $S_n$-invariance with the invariance of a general Weyl group $W$ and make connection with the Calogero–Moser models associated to other root systems (Perelomov 1990).

We shall encounter the functional equation [12] again in this survey and now note that this may be generalized to

$$\begin{vmatrix} 1 & 1 & 1 \\ F(x) & G(y) & H(z) \\ F'(x) & G'(y) & H'(z) \end{vmatrix} = 0, \quad x + y + z = 0 \quad [13]$$

If $F, G$, and $H$ are three-times differentiable functions satisfying the functional equation [13], then, up to the manifest invariance,

$$F(x) \to \alpha F(\delta x + \gamma_1) + \beta$$

$$G(x) \to \alpha G(\delta x + \gamma_2) + \beta$$

$$H(x) \to \alpha H(\delta x + \gamma_3) + \beta$$

where $\gamma_1 + \gamma_2 + \gamma_3 = 0$, the nonconstant solutions of [13] are given by $F(x) = G(x) = H(x) = e^x$, $x$, or $\wp(x)$. If (say) $H(z)$ is a constant then either

1. one of the functions $F(x)$ or $G(y)$ is the same constant as $H(z)$, in which case the remaining function is arbitrary, or
2. $F(x) = G(x) = e^x$.

We remark that in fact the exponential and linear function solutions satisfy [12] and [13] without the constraint $x + y + z = 0$. Further, the theorems immediately give the general analytic solutions to the same functional equations viewed as functions of a complex variable, showing that the solutions are in fact meromorphic. These theorems were established in Braden and Byatt-Smith (1999) where earlier results are described.

## Quantum Calogero–Moser Systems

Quite a bit is known about the quantum generalizations of the Calogero–Moser system. The polynomial and Weyl group $W$-invariance of the classical conserved quantities is replaced by a commutative ring $\mathcal{R}$ of $W$-invariant, holomorphic, differential operators, whose highest-order terms generate $W$-invariant differential operators with constant coefficients. The Poisson bracket is then replaced by a commutator of operators. When this is done functional equations again ensue and one finds that the potential term for the Laplacian $\mathcal{H}$ (the quantum Hamiltonian) has Calogero–Moser potential appropriate to $W$ (Oshima and Sekiguchi 1995). In this setting, it is known that the commutativity of just a few low-order elements of $\mathcal{R}$ dictate the form of the potential and the commuting algebra (at least for the classical root systems). In particular, Theorem 1 above is the classical analog for the $a_n$ root system of a quantum result where a functional equation equivalent to [12] was obtained by requiring the commutativity of certain linear, quadratic, and cubic holomorphic differential operators. Taniguchi's results (Taniguchi 1997) are also indicative of the rigidity of these quantum models: if $\mathcal{H}$ is the quantum Hamiltonian just discussed, and $\mathcal{Q}_{1,2}$ are holomorphic (but not *a priori* $W$-invariant), differential operators of appropriate degrees for which $[\mathcal{Q}_{1,2}, \mathcal{H}] = 0$, then $\mathcal{Q}_{1,2} \in \mathcal{R}$ and consequently $[\mathcal{Q}_1, \mathcal{Q}_2] = 0$.

## An Algebraic Ansatz: The Poincaré Algebra

We have earlier encountered the Ruijssenaars–Schneider models when considering functional equations ensuing from ansatz for Lax pairs. These models were however discovered by another route (Ruijsenaars and Schneider 1986) in the course of investigating mechanical models obeying the Poincaré algebra

$$\{H, B\} = P, \qquad \{P, B\} = H, \qquad \{H, P\} = 0 \quad [14]$$

Here, $H$ will be the Hamiltonian of the system generating time translations, $P$ is a space-translation

generator, and $B$ the generator of boosts. Ruijsenaars and Schneider began with the ansatz

$$H = \sum_{j=1}^{n} \cosh p_j \prod_{k \neq j} f(x_j - x_k)$$

$$P = \sum_{j=1}^{n} \sinh p_j \prod_{k \neq j} f(x_j - x_k)$$

$$B = \sum_{j=1}^{n} x_j$$

With this ansatz and the canonical Poisson bracket $\{p_i, x_j\} = \delta_{ij}$, the first two Poisson brackets of [14] involving the boost operator $B$ are automatically satisfied. The remaining Poisson bracket is then

$$\begin{aligned}\{H, P\} = &-\sum_{j=1}^{n} \partial_j \prod_{k \neq j} f^2(x_j - x_k) \\ &-\frac{1}{2}\sum_{j \neq k} \cosh(p_j - p_k) \prod_{l \neq j} f(x_j - x_l) \\ &\times \prod_{m \neq k} f(x_k - x_m)\big(\partial_j \ln f(x_k - x_j) \\ &+ \partial_k \ln f(x_j - x_k)\big)\end{aligned}$$

and for the independent terms proportional to $\cosh(p_j - p_k)$ to vanish we require that $f'(x)/f(x)$ be odd. This entails that $f(x)$ is either even or odd (Ruijsenaars and Schneider assumed the function even) and in either case $F(x) = f^2(x)$ is even. Supposing that $f(x)$ is so constrained, then the final Poisson bracket is equivalent to the functional equation

$$\{H, P\} = 0 \Longleftrightarrow \sum_{j=1}^{n} \partial_j \prod_{k \neq j} f^2(x_j - x_k) = 0 \qquad [15]$$

For $n = 3$, eqn [15] takes precisely the form [12] with $F(x) = f^2(x)$. From Theorem 2, the even solutions to this have the form $F(x) = \wp(x) + c$. This was found by Ruijsenaars and Schneider who further showed this function satisfies [15] for all $n$. The general solution to [15] has recently been established.

**Theorem 3** (Byatt-Smith and Braden 2003). *The general even solution of [15] amongst the class of meromorphic functions whose only singularities on the real axis are either a double pole at the origin, or double poles at $np$ ($p$ real, $n \in \mathbb{Z}$) is:*

(i) *for all odd $n$ given by the solution of Ruijsenaars and Schneider while*

(ii) *for even $n \geq 4$, there are in addition to the Ruijsenaars–Schneider solutions the following:*

$$F_i(z) = \sqrt{(\wp(z) - e_j)(\wp(z) - e_k)} \qquad [16]$$

*where $i, j, k$ are a cyclic permutation of 1, 2, 3.*

These functions have simple expressions in terms of Weierstrass elliptic functions, theta functions, and the Jacobi elliptic functions (Whittaker and Watson 1927). For example,

$$\begin{aligned}F_1(z) &= \sqrt{(\wp(z) - e_2)(\wp(z) - e_3)} = \frac{\sigma_2(z)\sigma_3(z)}{\sigma^2(z)} \\ &= \frac{\theta_3(v)\theta_4(v)}{\theta_1^2(v)} \frac{\theta_1'^2(0)}{4\omega^2\theta_3(0)\theta_4(0)} = b\frac{\mathrm{dn}(u)}{\mathrm{sn}^2(u)}\end{aligned}$$

where

$$\sigma_\alpha(z) = \frac{\sigma(z + \omega_\alpha)}{\sigma(\omega_\alpha)} \mathrm{e}^{-z\zeta(\omega_\alpha)}$$

$$u = \sqrt{e_1 - e_3}z$$

$v = z/2\omega, b = e_1 - e_3$ with $\omega_1 = \omega$, $\omega_2 = -\omega - \omega'$, and $\omega_3 = \omega'$. For appropriate ranges of $z$ the solutions are real. Their degenerations yield all the even solutions with only a double pole at $x = 0$ on the real axis. These degenerations may in fact coincide with the degenerations of the Ruijsenaars–Schneider solution.

Thus far, complete integrability has not been mentioned. The models discovered by Ruijsenaars and Schneider not only exhibited an action of the Poincaré algebra but were completely integrable as well. In particular, Ruijsenaars and Schneider demonstrated the Poisson commutativity for their solutions of the light-cone quantities

$$S_{\pm k} = \sum_{\substack{I \subseteq \{1,2,\ldots,n\} \\ |I| = k}} \exp\left(\pm \sum_{i \in I} p_i\right) \prod_{\substack{i \in I \\ j \notin I}} f(x_i - x_j) \qquad [17]$$

Then, $H = (S_1 + S_{-1})/2$ and $P = (S_1 - S_{-1})/2$. (Note the even/oddness of the functions $f(x)$ means that there really are only $n$ functionally independent quantities.) It is an open problem whether the new solutions [16] of Theorem 3 yield integrable systems. We know that these new solutions do not always yield Poisson commuting quantities using the ansatz of Ruijsenaars and Schneider, but as yet one cannot rule out other Poisson commuting conserved quantities.

## Quantum Ruijsenaars–Schneider Models

Ruijsenaars later investigated the quantum version of the classical models he and Schneider introduced.

From the outset, he sought operator analogs of the light-cone quantities [17]. He showed that (for $k = 1, \ldots, n$)

$$\hat{S}_k = \sum_{\substack{I \subseteq \{1,2,\ldots,n\} \\ |I|=k}} \prod_{\substack{i \in I \\ j \notin I}} h(x_j - x_i)^{1/2} \times \exp\left(-\sqrt{-1}\beta \sum_{i \in I} \partial_i\right) \prod_{\substack{i \in I \\ j \notin I}} h(x_i - x_j)^{1/2}$$

pairwise commute if and only if

$$\sum_{\substack{I \subseteq \{1,2,\ldots,n\} \\ |I|=k}} \left( \prod_{\substack{i \in I \\ j \notin I}} h(x_j - x_i) h(x_i - x_j - i\beta) - \prod_{\substack{i \in I \\ j \notin I}} h(x_i - x_j) h(x_j - x_i - i\beta) \right) = 0 \qquad [18]$$

held for all $k$ and $n \geq 1$. Here, $\beta$ is an arbitrary positive number and the sum is over all subsets with $k$ elements. Observe that upon dividing [18] by $\beta$ and letting $\beta \to 0$ this yields [15] with $F(x) = h(x)h(-x)$ when $k = 1$.

Ruijsenaars found a solution to [18] which has subsequently been shown to be unique. The general solution of the functional equation [18] analytic in a neighborhood of the real axis with either a simple pole at the origin or an array of such poles at $np$ on the real axis ($n \in \mathbb{Z}$) is given by

$$h(x) = b \frac{\sigma(x + \nu)}{\sigma(x)\,\sigma(\nu)} \, e^{\alpha x} \qquad [19]$$

This solution is related to the earlier Ruijsenaars–Schneider solution via

$$\frac{\sigma(x + \nu)\sigma(x - \nu)}{\sigma^2(x)\sigma^2(\nu)} = \wp(\nu) - \wp(x)$$

## Geometric Ansatz

We have already encountered the Hamiltonian system [10] corresponding to geodesic motion while discussing Lax pairs. We shall now consider various ansätze with a geometric flavor and their attendant functional equations.

It is known that the Ruijsenaars–Schneider model has the Calogero–Moser system as a scaling limit. Other scaling limits also exist for the Ruijsenaars–Schneider model. In particular, we may consider one in which the Poincaré algebra scales to either the Galilean algebra or a central extension of the Galilean algebra.

Similar to our analysis of the Poincaré algebra, we find that the functions

$$H = \frac{1}{2} \sum_{j=1}^{n} p_j^2 \prod_{k \neq j} f(x_j - x_k),$$

$$P = \sum_{j=1}^{n} p_j \prod_{k \neq j} f(x_j - x_k), \qquad B = \sum_{j=1}^{n} x_j$$

obey the algebra

$$\{H, B\} = P, \qquad \{P, B\} = \lambda, \qquad \{H, P\} = 0 \qquad [20]$$

if and only if $f(x)$ is either an even or odd function satisfying

$$\sum_{j=1}^{n} \prod_{k \neq j} f(x_j - x_k) = \lambda \qquad [21]$$

where $\lambda$ is a constant. When $\lambda = 0$ this is the Galilean algebra, while $\lambda \neq 0$ is a central extension of the Galilean algebra. Again we are encountering models of the form $H = (1/2)\sum_{j=1}^{n} g^{jj} p_j^2$ and so dealing with diagonal metrics. We note that if [21] holds for $n = 3$ then it holds for all $n$; and if it holds for $n = 4$ then it holds for all "even" $n$. This type of behavior was already encountered in Theorem 3.

Some particular solutions of [21] are known although the general solution is not known as yet. The odd functions $f(x) = 1/x$ ($\lambda = 0$), $\coth(x)$ ($\lambda = 1$ for $n$ odd and $\lambda = 0$ for $n$ even), $\sqrt{\wp(x) - e_\alpha}$ ($\lambda = 0$) yield solutions for example. Interestingly, in the case of an even number of particles, particular cases of the elliptic Ruijsenaars–Schneider model are in this list.

Diagonal metrics arise in many settings in integrable systems. By taking the ansatz

$$ds^2 = \sum_{i=1}^{n} \left( \prod_{j \neq i} \Psi(x^i - x^j) \right) (dx^i)^2$$

we may construct and solve a functional equation to show that the potentially nonvanishing curvature components $R^i_{jik}, R^i_{jjk} (k \neq i, j)$, and $R^i_{jij}$ have

1. $R^i_{jik} = R^i_{jjk} = 0$ $(k \neq i, j)$ if and only if $\Psi(x) = \alpha(e^{2bx} - 1)^a$ or $\alpha x^a$. We may set $\alpha = 1$ by rescaling $x$.
2. $R^i_{jij} = (-1)^n b^2$ when $\Psi(x) = (e^{2bx} - 1)$.
3. $R^i_{jij} = 0$ when $\Psi(x) = x$.

Thus, $\Psi(x) = x$ yields a solution of the Lamé equations. These metrics are of Stäckel form. The rational degenerations of the Galilean models above are given by this theorem. They may be understood as a parabolic limit of Jacobi elliptic coordinates.

Similar techniques may be applied to the more general metric

$$\mathrm{d}s^2 = \sum_{i=1}^{n}\left(\chi_i(x^i)\prod_{j\neq i}\Psi(x^i - x^j)\right)(\mathrm{d}x^i)^2$$

to show that $R^i_{jik} = R^i_{jjk} = 0 (k \neq i, j)$ if and only if $\Psi(x) = \alpha(\mathrm{e}^{2bx} - 1)^a$ or $\alpha x^a$.

## Ground-State Factorization

Some years ago, Sutherland and Calogero considered the problem as to when the ground-state wave function of a one-dimensional $n$-body Schrödinger equation with pairwise interactions would factorize. Thus, the problem is to determine those potentials $v(x)$ for which

$$\left\{-\frac{\hbar^2}{2}\sum_{i=1}^{n}\partial_i^2 + \frac{1}{2}\sum_{i\neq j}v(x_i - x_j) - E\right\} \times \bar{\Psi}(x_1, x_2 \ldots x_n) = 0 \qquad [22]$$

and where

$$\bar{\Psi}(x_1, x_2 \ldots x_n) = \prod_{i<j}\psi(x_i - x_j)$$

It is convenient to set

$$\psi(x_i - x_j) = \exp\left(\frac{1}{\hbar}\int^{x_i - x_j} f(x)\,\mathrm{d}x\right)$$

Substitution now shows

$$\frac{\hbar^2}{2}\sum_{i=1}^{n}\partial_i^2\bar{\Psi} = \left\{\hbar\sum_{i<j}f'(x_i - x_j) + \sum_{i<j}f(x_i - x_j)^2 + \sum_{i<j<k}\left[f(x_i - x_j)f(x_i - x_k) - f(x_i - x_j)f(x_j - x_k) + f(x_j - x_k)f(x_i - x_k)\right]\right\}\bar{\Psi}$$

Comparison with [22] shows that this may be expressed in terms of two-body potentials if and only if we have the functional equation

$$f(a)f(-b) - f(a)f(c) + f(c)f(-b) = G_1(a) + G_1(c) + G_2(-b), \quad a + b + c = 0 \qquad [23]$$

Now [23] is not quite the functional equation studied by Sutherland and Calogero. On physical grounds, Sutherland implicitly, and Calogero explicitly, made the "assumption" that $f$ is an odd function. This ensured that the potential was even and so bounded from below; equally it may be imposed so that $\psi(x_i - x_j) = \psi(x_j - x_i)$ and the ground state describes bosons. With this assumption, one arrives at the functional equation of Sutherland:

$$f(a)f(b) + f(b)f(c) + f(c)f(a) = G(a) + G(b) + G(c) \qquad [24]$$

Actually the assumption of $f$ being odd is unnecessary. One can show that there is a bijection between analytic solutions of [23] and analytic solutions of [24] for which $f'(x)$ is even. Upon requiring a potential of the stated form then necessitates $f$ being odd. Whatever, we arrive at the functional equation [24]. This is connected with [12] by

**Lemma 4** *If $a + b + c = 0$, then*

$$\begin{vmatrix} f''(a) & f''(b) & f''(c) \\ f'(a) & f'(b) & f'(c) \\ 1 & 1 & 1 \end{vmatrix} = 0 \qquad [25]$$

$$\Longleftrightarrow (f(a) + f(b) + f(c))^2 = F(a) + F(b) + F(c) \quad [26]$$

$$\Longleftrightarrow f(a)f(b) + f(b)f(c) + f(c)f(a) = G(a) + G(b) + G(c) \qquad [27]$$

Now, we may use Theorem 2 to determine those potentials with factorizable ground-state wave functions. We remark that the $\delta$-function potential $a\delta(x)$ of many-body quantum mechanics on the line, which also has a factorizable ground-state wave function, can be viewed as the $\alpha \to 0$ limit of $-b/\alpha\sinh^2(-x/\alpha + \pi\mathrm{i}/3)$ with $\pi a\alpha = 6b$. Thus, all of the known quantum mechanical problems with factorizable ground-state wave function are included in [12].

## Baker–Akiezer Functions

Baker–Akiezer functions are one of the foundations of the algebro-geometric or finite-gap integration of integrable systems. These functions may be viewed as an extension of the exponential function to curves of arbitrary genus $g$. They have essential singularities at various points on the curve and a prescribed asymptotic expansion at these points. The functions may be described in terms of theta functions on the Jacobian of the curve, and suitable meromorphic differentials on the curve. The functions [6] and [19] may be viewed as the Baker–Akhiezer function for a genus-1 curve. Now, just as the exponential function satisfies Cauchy's functional equation one may ask what functional equations (if any) characterize the

Baker–Akiezer function. This is an area of research still ongoing. Theta functions of a general abelian variety are known to satisfy addition formulas with $N=2^g$ terms. It appears Baker–Akiezer functions satisfy a similar functional equation with far fewer terms. Such a characterization of Baker–Akiezer functions, if found, will provide an analogous answer to that of the Riemann–Schottky problem which seeks to describe the Jacobians of curves amongst general abelian varieties.

The functional equations [5], and after suitably symmetrizing [9], are particular cases of the functional equation

$$\sum_{i=0}^{N} \phi_{3i}(x+y)\begin{vmatrix} \phi_{3i+1}(x) & \phi_{3i+1}(y) \\ \phi_{3i+2}(x) & \phi_{3i+2}(y) \end{vmatrix} = 0 \qquad [28]$$

with $N=1$ in the former case and $N=2$ in the latter. In the case $\phi_{3i+2}=\phi'_{3i+1}$, these may be viewed as differentiated forms of

$$\sum_{i=0}^{N} \phi_{3i}(x+y)\phi_{3i+1}(x)\phi_{3i+1}(y) = 1 \qquad [29]$$

For $N=0$, this is Cauchy's equation characterizing the exponential function and for $N=2$ it is equivalent to [8]. For $N=1$ and $N=2$, Buchstaber and Krichever have shown that "all" the solutions to this equation are the Baker–Akhiezer functions corresponding to algebraic curves of genus 1 and 2, respectively. In general, the Baker–Akhiezer functions for a genus-$g$ curve are known to satisfy [29] for $N=g$. Thus, many of the equations we have encountered are related to Baker–Akhiezer functions. Dubrovin, Fokas, and Santini have shown that Baker–Akhiezer functions for a genus-$g$ curve are related to the functional equation

$$\frac{q(x,y)q(y,z)}{q(x,z)} = r(x,y) - r(z,y) + \sum_{k=1}^{g} s_k(y)p_k(x,z)$$

Multivariable generalizations of [29] have been sought as a means of characterizing Baker–Akhiezer functions but such a characterization remains unproved as yet.

## Topology

Several of the functional equations we have encountered also arise in topology, where the German and Russian schools have powerfully applied functional equations to formal group laws and genera. It is still unclear whether these common threads form part of a greater fabric. A genus is a ring homomorphism

$$\varphi : \Omega \otimes \mathbb{Q} \to R, \quad \varphi(1) = 1$$

where $\Omega$ is the cobordism ring and $R$ an integral domain over $\mathbb{Q}$. To each even power series $Q(x)$ with $Q(0)=1$, one can associate a genus $\varphi_Q$ and vice versa (Hirzebruch *et al.* 1992). Defining the odd power series $f(x)=x/Q(x)$ with first term 1 and coefficients in $R$, the inverse function $g=f^{-1}$ is such that

$$g'(y) = \sum_{n=0}^{\infty} \varphi_Q(\mathbb{CP}^n)y^n$$

The genus corresponding to $Q(x)=x/\tanh(x)$ is known as the $L$-genus; it takes the value 1 on every even complex projective space. The genus corresponding to $Q(x)=(x/2)/\sinh(x/2)$ is known as the $\hat{A}$-genus. The so-called (string-inspired) Witten or elliptic genus corresponds to $Q(x)=x/\sigma(x)$. Certain genera may be associated with the index of natural differential operators on the manifold. Thus, the signature of $M$, sign($M$), is given in terms of the de Rham differential $d$ and its adjoint $d^*$,

$$\mathrm{ind}(d+d^*) = \mathrm{sign}(M) = \left(\prod_{j=0}^{2n} \frac{x_j}{\tanh(x_j)}\right)[M]$$

with variants for the $\hat{A}$-genus and elliptic genus. Further, when a compact topological group acts on the manifold, Atiyah and Bott showed how these indices may be determined from the fixed point sets of the action.

Now, functional equations arise naturally in this context when seeking genera with special properties. Novikov's school has shown, for example, that the genera associated with the index theorems of known elliptic operators arise as solutions of functional equations which are particular examples of [5]. Similarly, one may seek the following property of a genus $\varphi$: for the fiber bundle $p: E \xrightarrow{F} B$ with smooth fiber and base, one has that

$$\varphi(E) = \varphi(F)\cdot\varphi(B)$$

Such a genus is said to be strictly multiplicative. It may be shown that a genus is strictly multiplicative in bundles with fiber $\mathbb{CP}^{n-1}$ if and only if

$$\sum_{j=1}^{n}\prod_{k\neq j}\frac{1}{f(x_j-x_k)} = \lambda \qquad [30]$$

which is essentially [21]. Following the remarks of that equation, a genus $\varphi$ is strictly multiplicative for all fiber bundles with fibers $\mathbb{CP}^{n-1}$ if and only if it is strictly multiplicative for all fiber bundles with fiber $\mathbb{CP}^2$, in which case the genus is the $L$-genus. If, on the other hand, we only demand strict multiplicativity for all fiber bundles with fibers $\mathbb{CP}^{2k-1}$, then this is equivalent to requiring it to hold for all fiber bundles with fiber $\mathbb{CP}^3$, in which case the genus is an elliptic

genus. That the same functional equations arise in both the integrable systems and topological settings may reflect something deeper. String theory physics, for example, allows some topology changes such as flops, and physical quantities such as the partition function should reflect this invariance; invariance under classical flops characterizes the elliptic genus. In addition, connections have been made between the complex cobordism ring and conformal field theory.

## Other Areas

The constraints placed on this review have meant that several further applications of functional equations and integrable systems can only be noted. Using an ansatz together with functional equations, Wojciechowski gives an analog of the Bäcklund transformation for integrable many-body systems. Similarly, Inozemtsev constructs generalizations of the Calogero–Moser models, while this route was used to construct new solutions to the Witten–Dijkgraaf–Verlinde–Verlinde (WDVV) equations by Braden, Marshakov, Mironov, and Morozov. In the quantum regime, Gutkin derived and solved several functional relations by requiring a nondiffractive potential, while functional equations have been used to construct $R$-operators, solutions of the quantum Yang–Baxter equation on a function space.

*See also:* Calogero–Moser–Sutherland Systems of Nonrelativistic and Relativistic Type; Classical $r$-matrices, Lie Bialgebras, and Poisson Lie Groups; Cohomology Theories; Eigenfunctions of Quantum Completely Integrable Systems; Integrability and Quantum Field Theory; Integrable Systems and Algebraic Geometry; Integrable Systems: Overview; Lie Groups: General Theory; Quantum Calogero–Moser Systems; Toda Lattices; WDVV Equations and Frobenius Manifolds.

## Further Reading


Aczel J (1996) *Lectures on Functional Equations and Their Applications*. New York: Academic Press.

Braden HW (2001) Rigidity, functional equations and the Calogero–Moser model. *Journal of Physics A* 34: 2197–2204.

Braden HW and Buchstaber VM (1997) The general analytic solution of a functional equation of addition type. *SIAM Journal on Mathematical Analysis* 28: 903–923.

Braden HW and Byatt-Smith JGB (1999) On a functional differential equation of determinantal type. *Bulletin of the London Mathematical Society* 31: 463–470.

Bruschi M and Calogero F (1987) The Lax representation for an integrable class of relativistic dynamical systems. *Communications in Mathematical Physics* 109: 481–492.

Buchstaber VM and Krichever IM (1993) Vector addition theorems and Baker–Akhiezer functions. *Teoreticheskaya i Matematicheskaya Fizika* 94: 200–212.

Byatt-Smith JGB and Braden HW (2003) Functional equations and Poincaré invariant mechanical systems. *SIAM Journal on Mathematical Analysis* 34: 736–758.

Calogero F (1975) Exactly solvable one-dimensional many-body problems. *Lettere al Nuovo Cimento* 13(2): 411–416.

Calogero F (1976) On a functional equation connected with integrable many-body problems. *Lettere al Nuovo Cimento* 16: 77–80.

Calogero F (2001) *Classical Many-Body Problems Amenable to Exact Treatments*, Lecture Notes in Physics, New Series m: Monographs, vol. 66. Berlin: Springer.

Dubrovin BA, Fokas AS, and Santini PM (1994) Integrable functional equations and algebraic geometry. *Duke Mathematical Journal* 76: 645–668.

Hietarinta J (1987) Direct methods for the search of the second invariant. *Physics Reports* 147: 87–154.

Hirzebruch F, Berger T, and Jung R (1992) *Manifolds and Modular Forms*. Aspects of Mathematics, vol. E20. Braunschweig: Vieweg.

Oshima T and Sekiguchi H (1995) Commuting families of differential operators invariant under the action of a Weyl group. *Journal of Mathematical Sciences, The University of Tokyo* 2: 1–75.

Perelomov AM (1990) *Integrable Systems of Classical Mechanics and Lie Algebras,* vol. I. Basel: Birkhäuser.

Ruijsenaars SNM (1987) Complete integrability of relativistic Calogero–Moser systems and elliptic function identities. *Communications in Mathematical Physics* 110: 191–213.

Ruijsenaars SNM and Schneider H (1986) A new class of integrable systems and its relation to solitons. *Annals of Physics (NY)* 170: 370–405.

Sutherland B (1975) Exact ground-state wave function for a one-dimensional plasma. *Physical Review Letters* 34: 1083–1085.

Taniguchi K (1997) On uniqueness of commutative rings of Weyl group invariant differential operators. *Publications of the Research Institute for Mathematical Sciences, Kyoto University* 33: 257–276.

Whittaker ET and Watson GN (1927) *A Course of Modern Analysis*. Cambridge: Cambridge University Press.

# Functional Integration in Quantum Physics

**C DeWitt-Morette**, The University of Texas at Austin, Austin, TX, USA



## The Domain of Integration

Functional integration is integration over function spaces, that is, the variable of integration is a function $f$ with values in a $D$-dimensional manifold:

$$f : U \longrightarrow \mathbb{M}^D \tag{1}$$

Generically a space of functions is an infinite-dimensional space. Our understanding of infinite-dimensional spaces has progressed significantly during the twentieth century, and we can formulate functional integration in its proper setting.

Let $\mathcal{F}$ be the domain of integration, and $f \in \mathcal{F}$ the variable of integration. If the domain of $f$ is a subset $U$ of $\mathbb{R}$, the functional integral is called a path integral; if $U$ is of dimension higher than 1 (e.g., spacetime), $\mathcal{F}$ is often called a space of histories.

The information necessary for defining a domain of integration includes

- the domain and the range of the variable of integration $f$,
- the analytical properties of $f$, and
- possibly additional information, such as requirements on the values of $f$ on its boundary.

### Examples of variables of integration *f* in [1]

The domain $U$ of $f$ may be a time interval, a scale range, or any parameter. The range $\mathbb{M}^D$ of $f$ may be a group manifold, a Riemannian manifold, a symplectic manifold, a multiply-connected space, etc., or simply $\mathbb{R}^D$. The domain of integration $\mathcal{F}$ may be a space of pointed paths, for example,

$$x := \boldsymbol{T} \rightarrow \mathbb{M}^D, \quad \boldsymbol{T} = [t_a, t_b] \tag{2}$$

$$x(t_b) = x_b \in \mathbb{M}^D \quad \text{for all } x \in \mathcal{F}$$

The paths $x$ may be continuous (e.g., Brownian paths), or may have square integrable derivatives; $\mathcal{F}$ is then an $L^{2,1}$ space (e.g., quantum physics).

$$\int_T \mathrm{d}t\, |\dot{x}(t)|^2 < \infty, \qquad x \in \mathcal{F} \tag{3}$$

Given a domain of integration $\mathcal{F}$, one needs to select a volume element appropriate to $\mathcal{F}$. This is a challenge which has been met in a number of cases (Cartier and DeWitt-Morette 2006). Examples are given below. Given a volume element, one can then characterize the functionals $F$ on $\mathcal{F}$ integrable with respect to the chosen volume element.

## Two Basic Techniques

The two most useful techniques for computing integrals are change of variable of integration and integration by parts. They follow from fundamental properties that apply to functional integrals as well as to ordinary integrals. Let us recall them in the context of ordinary integrals.

Let $f$ and $g$ be functions on $\mathbb{R}$ of compact support. Let $I$ stand for integration

$$I(f) = \int_{\mathbb{R}} \mathrm{d}x\, f(x), \quad x \in \mathbb{R}$$

and $D$ for derivation of $f$ with respect to $x$,

$$(Df)(x) = \frac{\mathrm{d}}{\mathrm{d}x} f(x)$$

The fundamental rule

$$DI = 0 \Longrightarrow 0 = \frac{\mathrm{d}}{\mathrm{d}x} \int_{\mathbb{R}} \mathrm{d}x\, f(x) \tag{4}$$

The functional $I(f)$ is invariant under a change of variable of integration.

Another fundamental rule is $ID = 0$:

$$\begin{aligned} ID = 0 \Longrightarrow 0 &= \int \mathrm{d}(f(x)g(x)) \\ &= \int \mathrm{d}f(x) \cdot g(x) + \int \mathrm{d}g(x) \cdot f(x) \end{aligned} \tag{5}$$

The fundamental rules [4] and [5] apply to functional integration. The derivation $D$ can be either a functional derivative or a Lie derivative defined as follows. Let $\mathbb{K}$ be the reals $\mathbb{R}$ or the complex $\mathbb{C}$, let $f$ be a differentiable functional on a Banach space $\mathbb{X}$

$$f : U \subset \mathbb{X} \longrightarrow \mathbb{K} \tag{6}$$

The functional derivative $Df|_{x_0}$ of $f$ at $x_0$ is defined by the equation

$$f(x_0 + h) - f(x_0) = Df|_{x_0} h + R(h) \tag{7}$$

with the norm $\|R(h)\|$ of order less than the norm $\|h\|$.

The Lie derivative $\mathcal{L}_V$ along the vector field $V$ is conceptually intuitive and of practical interest: an infinite-dimensional space $\mathbb{X}$ of paths $x$ is not an intuitive concept, but a one-parameter family of

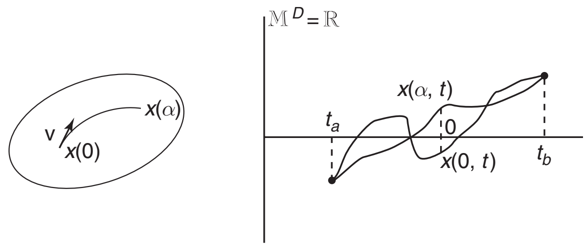



**Figure 1** A one-parameter family of paths with fixed endpoints.

paths $\{x(\alpha)\} \in \mathbb{X}$, with $\alpha \in [0,1]$, is a convenient tool for dealing with $\mathbb{X}$:

$$\begin{aligned} x(\alpha) &: T \to \mathbb{M}^D \\ x(\alpha, t) &:= (x(\alpha))(t) \in \mathbb{M}^D \end{aligned} \qquad [8]$$

Set $x(0) = x_0$. A differentiable family $\{x(\alpha)\}$ defines a vector field $V(x_0)$ along the path $x_0$ (see Figure 1):

$$\begin{aligned} V(x_0) &:= \frac{\mathrm{d}}{\mathrm{d}\alpha} x(\alpha)\Big|_{\alpha=0} \\ V(x_0(t)) &= \frac{\partial}{\partial \alpha} x(\alpha, t)\Big|_{\alpha=0} \end{aligned} \qquad [9]$$

The functional vector field $V$ on the tangent bundle of $\mathbb{X}$ defines a group of transformations on $\mathbb{X}$, and a Lie derivative $\mathcal{L}_V$ of tensor fields on $\mathbb{X}$.

The Lie derivative $\mathcal{L}_V$ obeys the Cartan (Elie and Henri) equation

$$\mathcal{L}_V = d\imath_V + \imath_V d \qquad [10]$$

where $d$ is the exterior differential and $\imath_V$ is the interior product, defined as usual on Banach spaces.

**Remark** (Berezin integrals). To show the power of the rules [4] and [5], we can mention that they provide Berezin rules of integration over Grassmann variables (Cartier and DeWitt-Morette 2006).

## Path Integrals and Quantum Dynamics

The history of path integrals in quantum physics did not begin with the definitions of domain of integration, volume elements, etc. It began with the Ph.D. thesis of R P Feynman in 1942. Feynman expressed the time evolution of a system as the limit $N = \infty$ of the following $N$-tuple integral:

$$\begin{aligned} (x_t|x_a) = \lim_{N=\infty} \int\!\!\int \cdots \int (x_t|x_N)\mathrm{d}x_N(x_N|x_{N-1}) \\ \times \mathrm{d}x_{N-1} \cdots (x_2|x_1)\,\mathrm{d}x_1\,(x_1|x_0) \end{aligned} \qquad [11]$$

where the time interval $T = [t, 0]$ has been replaced by $N$ of its points $\{t_i\}, 1 \le i \le N$:

$$t_0 < t_1 < \cdots < t_N < t \qquad [12]$$

and the path $x : T \to \mathbb{R}^D$ is replaced by $N$ of its values

$$x_i := x(t_i) \qquad [13]$$

Dirac (1933) had shown that $(x_t|x_0)$ defines the exponential of a quantum function $\mathcal{S}_Q$, by

$$\exp\big(\mathrm{i}\mathcal{S}_Q(x_t, x_0, t)/\hbar\big) := (x_t|x_0) \qquad [14]$$

such that the real part of $\mathcal{S}_Q$ is the classical action function (a.k.a. Hamilton's principal function; further studies have shown that the correct statement is: the real part is the classical action, up to order $\hbar$), and the imaginary part of $\mathcal{S}_Q$ is of order $\hbar$, the normalized Planck constant

$$\hbar = h/2\pi \qquad [15]$$

Feynman remarked that for a system with Lagrangian $L$ the short-time probability amplitude $(x_{t+\delta t}|x_t)$ is "often equal to

$$A^{-1} \exp\Big(\mathrm{i}\,\delta t\, L\Big(\frac{x_{t+\delta t} - x_t}{\delta t}, x_{t+\delta t}\Big)\Big/\hbar\Big) \qquad [16]$$

within a normalizing constant A as the limit $\delta t$ approaches zero." The absolute value of $A$ can be obtained from a unitary requirement (Morette 1951).

Feynman expressed the finite probability amplitude as a path integral, limit of the discretized expression [11]

$$(x_t|x_0) = \int \mathcal{D}x \exp(\mathrm{i}\mathcal{S}(x)/\hbar) \qquad [17]$$

where $S(x)$ is the action functional

$$\mathcal{S}(x) = \int_{t_0}^{t} \mathrm{d}sL(\dot{x}(s), x(s)) \qquad [18]$$

The undefined symbol $\mathcal{D}x$ is a "volume element" on the space of paths, corresponding to the infinite product of the normalization constant $A^{-1}$.

The issues raised by the path integral [17] are

- the definition of the volume element $\mathcal{D}x$; and
- a method for computing [17] for a given action functional $\mathcal{S}$.

The explicit calculation of the limit [11] of an $N$-tuple integral when $N = \infty$ is a Herculean task of very limited use. But two other methods of wide applications, leaving the volume element $\mathcal{D}x$ as a heuristic symbol, have vindicated the power of functional integration: the diagram technique and the semiclassical expansions.

Feynman devised practical rules for computing asymptotic expansions of path integrals, order by order in perturbation theory. The rules are depicted by graphs, known as the Feynman diagrams ('t Hooft and Veltmann 1973). Feynman's first explicit nontrivial calculation was the Lamb shift. It earned him the Nobel prize in 1965 (Feynman 1966). The diagram technique is widely used in quantum mechanics and quantum field theory. The time ordering provided by the time parameter in quantum mechanics becomes, in quantum field theory, a chronological ordering dictated by light cones.

Another explicit calculation of a path integral [17] uses the Taylor expansion of the action functional $\mathcal{S}(x)$ around one of its values. It is known as the background method (DeWitt 2004). It is called a semiclassical WKB approximation when one expands around an extremum $\mathcal{S}(x_{\rm cl})$ where $x_{\rm cl}$ is a solution of the Euler–Lagrange equation $\mathcal{S}'(x_{\rm cl})=0$ (Wenzel 1926, Kramers 1927, Brillouin 1926).

Introduced in 1951 (Morette (1951)) semiclassical approximations are now the subject of a rich literature reviewed briefly below.

## Gaussian Volume Elements

A lesson from Gaussians on $\mathbb{R}^D$ suggests a definition of volume elements on infinite-dimensional Banach spaces $\mathbb{X}$. Let

$$I_D(a) := \int_{\mathbb{R}^D} dx \exp\left(-\frac{\pi}{a}|x|^2\right) \quad \text{for } a>0 \qquad [19]$$

$$dx = dx^1 \cdots dx^D \text{ and } |x|^2 = \sum_{j=1}^{D} (x^j)^2 = \delta_{ij}x^i x^j$$

An elementary calculation gives $I_D(a)=a^{D/2}$. Therefore, when $D=\infty$,

$$I_\infty(a) = \begin{cases} 0 & \text{if } 0<a<1 \\ 1 & \text{if } a=1 \\ \infty & \text{if } 1<a \end{cases} \qquad [20]$$

This is clearly an unsatisfactory situation, but it can be corrected by introducing a dimensionless volume element:

$$\mathcal{D}_a x := \frac{1}{a^{D/2}} dx^1 \cdots dx^D \qquad [21]$$

The volume element $\mathcal{D}_a x$ can be defined by the integral

$$\int_{\mathbb{R}^D} \mathcal{D}_a x \exp\left(-\frac{\pi}{a}|x|^2 - 2\pi i\langle x', x\rangle\right) := \exp\left(-a\pi|x'|^2\right) \qquad [22]$$

where $x'$ is in the dual $\mathbb{R}_D$ of $\mathbb{R}^D$. Equation [22] suggests the following generalization of Gaussians on $\mathbb{R}^D$ to Gaussians on a Banach space $\mathbb{X}$:

$$\int_X \mathcal{D}_{s,Q}x \exp\left(-\frac{\pi}{s}Q(x)\right)\exp(-2\pi i\langle x',x\rangle) := \exp(-s\pi W(x')) \qquad [23]$$

where $s\in\{1,i\}$, $Q(x)$ is a quadratic form on $\mathbb{X}$ (see condition on $Q$ below). $W(x')$ is a quadratic form on the dual $\mathbb{X}'$ of $\mathbb{X}$, inverse of $Q(x)$ in the following sense. Set

$$Q(x)=\langle Dx,x\rangle \text{ and } W(x')=\langle x',Gx'\rangle \qquad [24]$$

where $\langle\,,\rangle$ is a duality product, for example, the product of $x\in\mathbb{X}$ and $Dx\in\mathbb{X}'$; then

$$DG = 1_{X'}, \qquad GD = 1_{\mathbb{X}} \qquad [25]$$

Equation [23] defines a Gaussian volume element $d\Gamma$ by its Fourier transform

$$\mathcal{F}\Gamma_{s,Q}(x') := \int_X d\Gamma_{s,Q}(x)\exp(-2\pi i\langle x',x\rangle) := \exp(-s\pi W(x')) \qquad [26]$$

where the Gaussian volume element

$$d\Gamma_{s,Q}(x) \underset{=}{\int} \mathcal{D}_{s,Q}(x)\exp\left(-\frac{\pi}{s}Q(x)\right) \qquad [27]$$

This is a qualified equality valid upon integration.

The definition of the Gaussian volume element by its Fourier transform $\mathcal{F}\Gamma$ is valid for $s=1$ (Wiener integral) when $Q(x)>0$; it is valid for $s=i$ (Feynman integral) when $\mathrm{Re}\,Q(x)>0$.

**Remark** Volume elements were introduced with the notation such as $dx$; later they were identified with forms such as $\omega=dx$. In [26] we omit d on the left-hand side (LHS) for visual clarity.

**Example** (diagram expansion). The following integrals follow readily (Cartier and DeWitt-Morette 2006) from the definition [26]. Let $x'$ be in the dual $\mathbb{X}'$ of $\mathbb{X}$,

$$\int_X d\Gamma_{s,Q}(x)\langle x',x\rangle^{2n+1} = 0 \qquad [28]$$

$$\int_X d\Gamma_{s,Q}(x)\langle x',x\rangle^{2n} = \frac{2n!}{2^n n!}\left(\frac{s}{2\pi}\right)^n W(x')^n \qquad [29]$$

$$\int_X d\Gamma_{s,Q}(x)\langle x'_1,x\rangle\cdots\langle x'_{2n},x\rangle = \left(\frac{s}{2\pi}\right)^n \sum W(x'_{i_1},x'_{i_2})\cdots W(x'_{i_{2n-1}},x'_{i_{2n}}) \qquad [30]$$

where $\sum$ is a sum without repetitions of identical terms.

For instance when $n=1$, eqn [30] reads

$$\int_X d\Gamma_{s,Q}(x)\langle x'_1, x\rangle\langle x'_2, x\rangle = \frac{s}{2\pi} W(x'_1, x'_2) \qquad [31]$$

$W(x'_1, x'_2)$ is called the two-point function (a.k.a. the propagator). In a diagram it stands for a line from $x'_1$ to $x'_2$.

Feynman diagrams represent Gaussian integrals of polynomials.

For instance when $n=2$, the diagram representation of [30] is the sum of three terms,

$$W(x'_1, x'_2)W(x'_3, x'_4) + W(x'_1, x'_3)W(x'_2, x'_4) + W(x'_1, x'_4)W(x'_2, x'_3)$$

**Example** (Linear maps). Linear maps on $\mathbb{R}^D$ are limited to $L: x \to Ax$, where $A$ is a $D \times D$ constant matrix. Linear maps on a Banach space $\mathbb{X}$ offer many possibilities:

(i) Projections. For example, let $x: \boldsymbol{T} \to \mathbb{R}$ and

$$L : x \in \mathbb{X} \longrightarrow \{x(t_1), x(t_2), \ldots, x(t_n)\} \in \mathbb{R}^n) \qquad [32]$$

This projection is a discretization of the path, useful in particular in numerical calculations of path integrals. Equation [32] is unambiguous, whereas the limit of the discretized expression [11] is ill-defined.

(ii) Liouville decomposition. For example, let $D$ be a second-order differential operator on a space of paths $x: [t_a, t_b] \to \mathbb{M}^D$ vanishing on the boundary, $x(t_a)=0, x(t_b)=0$. Let $\{\varphi_k\}$ be a complete, orthogonal set of eigenfunctions of $D$, then the decomposition of $x$ into the basis $\{\varphi_k\}$,

$$x^\alpha(t) = \sum_{k=1}^{\infty} u^\alpha \psi_k^\alpha(t) \qquad [33]$$

is a linear map

$$L : x \in \mathbb{X} \longrightarrow \{u^1, \ldots, u^\infty\} \in \mathbb{R}^\infty$$

It is useful in particular for diagonalizing (see, e.g., [107]) the Green function $G$ of $D$ [25] (a.k.a. the covariance in a Gaussian integral [24], or the two-point function in [31]).

(iii) Volterra maps. For example, let $L: \mathbb{X} \to \mathbb{Y}$ by

$$y(t) = \int_T ds\, \theta(t-s)x(s)$$

$$\theta(t-s) = 1 \quad \text{for } s < t, 0 \text{ otherwise} \qquad [34]$$

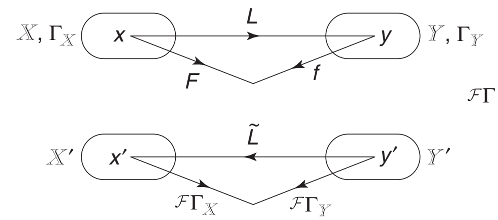



**Figure 2** Linear maps. (Published with permission by Elsevier, North Holland.)

Let $\mathbb{X}$ be the space of square-integrable functions on $T$ and $\mathbb{Y}$ be an $L^{2,1}$ space (square-integrable function for which the first derivative is also square integrable) then $L$ maps the canonical quadratic form on $\mathbb{X}$ into the canonical quadratic form on $\mathbb{Y}$, hence the canonical Gaussian on $\mathbb{X}$ into the canonical Gaussian on $\mathbb{Y}$. The identity mapping $i$ from $\mathbb{Y}$ into the space $\mathcal{C}$ of continuous functions maps the canonical Gaussian on $\mathbb{Y}$ into the Wiener Gaussian on $\mathcal{C}$ (DeWitt-Morette *et al.* 1979).

The linear maps [32]–[34] and their obvious generalizations have been used for computing explicitly many functional integrals (see **Figure 2**). The basic formula reads

$$\int_X d\Gamma_X(x)F(x) = \int_Y d\Gamma_Y(y)f(y), \quad F = f \circ L \qquad [35]$$

where the Fourier transform $\mathcal{F}\Gamma$ is given by

$$\mathcal{F}\Gamma_Y = \mathcal{F}\Gamma_X \circ \tilde{L} \qquad [36]$$

$\tilde{L}$ is the transpose of the linear map $L$ defined by

$$\langle \tilde{L}y', x\rangle = \langle y', Lx\rangle \qquad [37]$$

Computing $\mathcal{F}\Gamma_Y$ does not require any calculation. It can be read off eqn [36]. Computing $d\Gamma_Y$ is easy in a number of cases such as the following:

1. $\mathbb{Y}$ is finite-dimensional. In other words [35] is a cylindrical integral. Then

$$d\Gamma_Y(y) = dy^1 \cdots dy^D \left(\det Q_{ij}\right)^{1/2} \times \exp\left(-\frac{\pi}{s}Q(y)\right) \qquad [38]$$

where $Q(y)$ is an abbreviation of

$$Q_Y(y) = Q_{Yij}\, y^i y^j \qquad [39]$$

its inverse$W_{Y'}(y')$ in the sense of [24]–[25] is

$$W_{Y'}(y') = W^{ij}_{Y'} y'_i y'_j \qquad [40]$$

that is given by [36]:

$$W_{Y'}(y') = W_{X'} \circ \tilde{L} \qquad [41]$$

$W_{X'}$ is the quadratic form defining $\Gamma_X$ by [26]. When $D$ is small, say less than 4, and this is not an unusual situation, it is easy to compute [38].

**Example** (Wiener Gaussians and Brownian motion). The Wiener Gaussian on the space $\mathcal{P}_a\mathbb{R}$ of pointed paths $x: T \to \mathbb{R}$, $T = [t_a, t_b], x(t_a) = 0$ is defined by its variance [26]:

$$W(x') := \int_T \mathrm{d}x'(t') \int_T \mathrm{d}x'(s) \inf(t, s) \qquad [42]$$

Let $\mathbb{Y}$ be the Wiener differential space consisting of the differences of two consecutive values of $x$ on the $n$-discretized time interval. The space $\mathbb{Y}$ is finite dimensional,

$$L: \mathbb{X} \to \mathbb{Y}$$
$$\text{by} \quad y^j = x(t_{j+1}) - x(t_j) = \langle \delta_{t_{j+1}} - \delta_{t_j}, x \rangle$$

It follows from [37] that

$$\tilde{L}y' = \sum_j y'_j (\delta_{t_{j+1}} - \delta_{t_j})$$

and

$$\mathrm{d}\Gamma_Y(\Delta x) = \mathrm{d}y^1 \cdots \mathrm{d}y^n \frac{1}{\prod_{j=1}^n (s\Delta t_j)^{1/2}} \times \exp\left( -\frac{\pi}{s} \sum_j \frac{(\Delta x_j)^2}{\Delta t_j} \right) \qquad [43]$$

where $\Delta t_j := t_{j+1} - t_j$ and $\Delta x_j := x(t_{j+1}) - x(t_j)$.

When $s = 1$ the Gaussian $\Gamma_Y$ defines the distribution of a Brownian path. The Gaussian $\Gamma_X$ of covariance $\inf(t, s)$ is the Wiener measure. □

2. In semiclassical approximations, $Q_X$ is the Hessian (second variation) of an action functional $\mathcal{S}$:

$$Q_X(h) = \frac{\mathrm{d}^2}{\mathrm{d}\alpha^2} \mathcal{S}(x(\alpha)) \bigg|_{\alpha=0}$$
$$\text{with } h = \frac{\partial}{\partial \alpha} x(\alpha) \bigg|_{\alpha=0} \qquad [44]$$

where $\{x(\alpha)\}$ is a one-parameter family of paths [8]. The Jacobi field technology (the Jacobi operator is defined by [103]; a Jacobi field is a solution of [102]) yields the inverse of $W_{Y'}$ and its determinants (Cartier and DeWitt-Morette 2006); they have been worked out for a variety of boundary conditions on classical paths.

## Volume Elements Other than Gaussian

The definition [26]–[27] of Gaussian volume elements is a particular case of volume elements on a Banach space $\Phi$ defined by

$$\int_\Phi \mathcal{D}_{\Theta, Z}\varphi \cdot \Theta(\varphi, J) := Z(J) \qquad [45]$$

for $\varphi$ in $\Phi$, and $J$ in the dual $\Phi'$ of $\Phi$. The volume element $\mathcal{D}_{\Theta, Z}$ is defined by two continuous bounded functionals

$$\Theta: \Phi \times \Phi' \longrightarrow \mathbb{C} \quad \text{and} \quad Z: \Phi' \longrightarrow \mathbb{C} \qquad [46]$$

In quantum field theory, $\varphi$ is a field and $J$ is a source. The functional $Z(J)$ is then the Schwinger generating functional for the $n$-point functions. An axiomatic and applications of functional integrals on $\Phi$ with volume elements $\mathcal{D}_{\Theta, Z}$ can be found in (Cartier and DeWitt-Morette (1993)).

**Example** (Poisson volume elements) (Cartier and DeWitt-Morette (2006) and Collins (1997)). A Poisson random variable is a random variable $N$ taking values in the set $\mathbb{N}$ of non-negative integers such that the probability $p_n$ that $N = n$ is

$$p_n := \Pr(N = n) := \exp(-\lambda) \frac{\lambda^n}{n!}, \quad \lambda \geq 0 \qquad [47]$$

Thanks to the normalizing constant $\exp(-\lambda)$, $\sum_{n=0}^\infty p_n = 1$. The parameter $\lambda$ is the mean value of $N$:

$$\langle N \rangle = \lambda \qquad [48]$$

A record of fortuitous events occurring at random times $t_0 < T_1 < T_2 \cdots$ can consist either of the number $N(t)$ of events occurring at times less than or equal to $t$, or of the waiting times

$$W_k = T_k - T_{k-1} \qquad [49]$$

between two consecutive events.

When the waiting times are stochastically independent and when

$$\Pr(t < W_k < t + \mathrm{d}t) = p_a(t)\,\mathrm{d}t \qquad [50]$$

$$p_a(t) = a \exp(-at), \quad t > 0 \qquad [51]$$

the record is a Poisson random variable. It is related to the number of events $N(t)$ as follows.

Let $\boldsymbol{T}$ be a finite time interval $[t', t'']$, and

$$N_T = N(t'') - N(t') \qquad [52]$$

the number of events during $\boldsymbol{T}$. The random variable $N_T$ follows a Poisson law [47] with mean value

$$\lambda_a(\boldsymbol{T}) = a(t'' - t') \qquad [53]$$

For mutually disjoint time intervals $\boldsymbol{T}^{(1)}, \boldsymbol{T}^{(2)}, \ldots$ the random variables $N_{T^{(1)}}, N_{T^{(2)}}, \ldots$ are stochastically independent.

Whereas the parameter $\lambda$ must be real non-negative, the parameter $a$ can be pure imaginary; therefore, Poisson processes defined by waiting

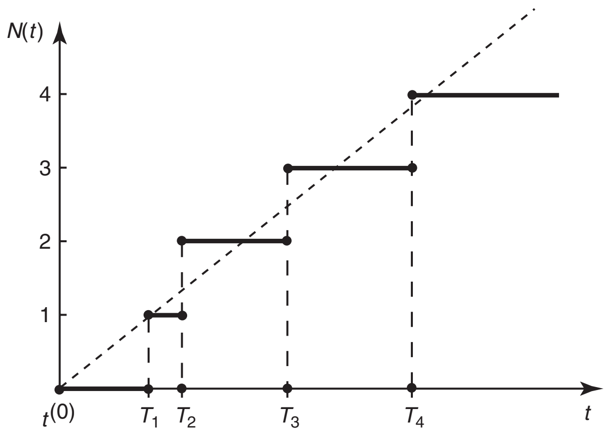



**Figure 3** A Poisson path in $\mathbb{X}_4$.

times can be used in quantum physics as well as in probability. When $a$ is real, it is called the decay constant because its physical dimension is $[\text{time}]^{-1}$.

A Poisson path $x \in \mathbb{X}_n$ is characterized by $n$ jumps and the jump times during a given time interval $T = [t_a, t_b]$ (**Figure 3** illustrates a Poisson path in $\mathbb{X}_4$). The space $\mathbb{X}$ of Poisson paths is the union of all $\mathbb{X}_n$:

$$\mathbb{X} = \cup \mathbb{X}_n \qquad [54]$$

One can define a volume element $\mathcal{D}_{a,T}$ on $\mathbb{X}$ by its Fourier transform:

$$\int_X \mathcal{D}_{a,T} x \cdot \exp(\mathrm{i}\langle x, f\rangle) := \exp\left(\int_T \mathrm{d}t\, a\, \mathrm{e}^{\mathrm{i}f(t)}\right) \qquad [55]$$

Here a path $x \in \mathbb{X}_n$, characterized by $n$ jump times $T_1, \ldots, T_n$, is represented by the sum

$$\delta_{T_1} + \cdots + \delta_{T_n}$$

Hence

$$\langle x, f\rangle = f(T_1) + \cdots + f(T_n) \qquad [56]$$

The dimensionless volume element on $\boldsymbol{T}$ is

$$\mathrm{d}v(t) = a\ \mathrm{d}t$$

Therefore,

$$\mathrm{vol}(T) = aT, \quad T = t_b - T_a$$

$$\mathrm{vol}(\mathbb{X}_n) = a^n T^n / n! \qquad [57]$$

$$\mathrm{vol}(\mathbb{X}) = \exp(\mathrm{vol}(\boldsymbol{T})) \qquad [58]$$

and it makes sense to write formally

$$\mathbb{X} = \exp \boldsymbol{T}$$

It can be proved that the volume element $\mathcal{D}_{a,T}x$ is a measure, in the technical sense of the word (Cartier and DeWitt-Morette 2006).

Functional integration on spaces of Poisson paths have been used extensively in solutions of Klein–Gordon equations, the telegrapher equation and the Dirac equation (Cartier and DeWitt-Morette 2006).

Other volume elements of interest in quantum physics include (LaChapelle 2004):

- gamma volume elements, which are to gamma probability distributions what Gaussian volume elements are to Gaussian probability distributions; and
- Hermite volume elements convenient for integrating Wick-ordered polynomials.

A Dirac "$\delta$-function" is formally the limit of a Gaussian integral. Formally, one can introduce a Dirac functional volume element as the limit of a Gaussian volume element.

## The Koszul Formula

There are several roadblocks on the road from finite to infinite-dimensional spaces. For instance, a volume in a $D$-dimensional space is a top-differential form, that is, a $D$-form. There is no top-form in an infinite-dimensional space – neither on Grassmann manifolds since Grassmann forms are totally symmetric tensors. A $D$-form in $\mathbb{R}^D$ has only one strict component and is equivalent to a scalar density of weight 1, but scalar densities of weight 1 do not form an algebra.

For these reasons, volume elements have so far been defined by integrals [26], [27], [43], and [55]. Short of giving an explicit expression for their differential forms, one can require them to satisfy the Koszul formula

$$\mathcal{L}_X \omega = \mathrm{Div}(X)\omega \qquad [59]$$

where $\omega$ is a volume element on a Banach space $\mathbb{X}$, $X$ a vector field generating a group of transformations on $\mathbb{X}$, $\mathcal{L}_X$ the Lie derivative defined by $X$, and $\mathrm{Div}(X)$ the standard generalization of div(ergence) on finite-dimensional spaces (see, e.g., Cartier and DeWitt-Morette (2006) for the explicit expression of divergences on Riemannian, symplectic, Grassmann manifolds). The Koszul formula dictates how a volume element changes under a group of transformations.

It often happens that an object cannot be defined *per se*, but that it is sufficient to define its variation. For example, one does not define potentials, but potential differences; the ratio of infinite-dimensional determinants can be defined without defining each determinant; the work of Wiener on "differential-spaces," which is a landmark in functional integration, is based

on differences between two consecutive values of a function, etc. Similarly, the Koszul formula does not define $\omega$ but gives its variation $\mathcal{L}_X\omega$.

## The Operator Formalism of Quantum Physics

Functional integrals can be used to represent operator matrix elements, and solutions of the Schrödinger equation.

1. Matrix elements of operators on Hilbert spaces. Symbolically,

$$\langle\beta|\exp(-\mathrm{i}Ht/\hbar)|\alpha\rangle = \int_{\mathbb{X}_{\alpha\beta}} \mathcal{D}x\exp(\mathrm{i}\mathcal{S}(x)/\hbar) \qquad [60]$$

   The domain of integration $\mathbb{X}_{\alpha\beta}$ is a space of paths $x$ on $[t, 0]$ satisfying initial conditions that characterize the quantum state $\alpha$, and final conditions that characterize the quantum state $\beta$. The action functional $\mathcal{S}$ yields the Hamiltonian $H$.

   A key property of path integrals is their representations of matrix elements of time-ordered operators. The path parameter (time, scale, or any other parameter) provides the operator ordering [11]. A simple example is the two-point function of the Wiener measure [42]:

$$\int_{\mathbb{X}} \mathrm{d}\Gamma(x)\, x(t)x(s) = \inf(t, s) \qquad [61]$$

   The function integral orders the time, that is, the argument of the variable of integration. In quantum field theory, time ordering becomes a chronological ordering dictated by light cones.
2. Schrödinger equation and other parabolic equations (Cartier and DeWitt-Morette 2006).

The following theorem provides the mathematical underpinning for a great variety of functional integrals. It also provides a construction of functional integrals, which begins with the symmetries of a given physical system rather than its action functional. The theorem consists of two parts: the definition of a functional integral, and the partial differential equation satisfied by the value of the functional integral, as a function of a set of parameters.

Given a manifold $\mathbb{M}$, consider the contractible space $\mathcal{P}_0\mathbb{M}$ of pointed $L^{2,1}$ paths over $\boldsymbol{T}=[t_a, t_b]$:

$$x : T \to \mathbb{M}, \quad \text{e.g., } x(t_b) = x_b, \quad \text{i.e., } x \in \mathcal{P}_0\mathbb{M} \qquad [62]$$

Given $D+1$ vector field $Y$, $\{X_{(\alpha)}\}$, generators of group of transformations on $\mathbb{M}$, define a map

$$P : \mathcal{P}_0\mathbb{R}^D \to \mathcal{P}_b\mathbb{M} \quad \text{by } z \to x \qquad [63]$$

explicitly

$$\mathrm{d}x(t, z) = X_{(\alpha)}(x(t, z))\mathrm{d}z^\alpha + Y(x(t, z))\mathrm{d}t \qquad [64]$$

$$x(t_b, z) = x_b, \qquad z(t_b) = 0 \qquad [65]$$

In general, the vector fields do not commute and the solution of [64]–[65] is of the form

$$x(t, z) = x_b \cdot \sum(t, z) \qquad [66]$$

where $\Sigma(t, z)$ is an element of a group of right actions on $\mathbb{M}$, defined by the $D+1$ generators $Y$, $\{X_{(\alpha)}\}$:

$$x_b \cdot \sum(t + t', z \times z') = x_b \cdot \sum(t, z) \cdot \sum(t', z')$$

The path $z$ defined on $[t_a, t]$ is followed by the path $z'$ on $[t, t']$.

Consider the following functional integral over $\mathcal{P}_0\mathbb{R}^D$ of a functional of paths on $\mathcal{P}_b\mathbb{M}$:

$$(U_T\phi)(x_b) := \int_{\mathcal{P}_0\mathbb{R}^D} \mathcal{D}_{s,Q}z \exp\Big(-\frac{\pi}{s}Q(z)\Big) \times \phi\Big(x_b \cdot \sum(t, z)\Big) \qquad [67]$$

where

$$Q(z) = \int_T \mathrm{d}t\; h_{\alpha\beta}\dot{z}^\alpha(t)\dot{z}^\beta(t) \qquad [68]$$

The functional $(U_T\phi)$ at $x_b$ is a function $\Psi(T, x_b)$. It is a solution of the generalized Schrödinger equation,

$$\frac{\partial\Psi}{\partial T} = \frac{s}{\pi}h^{\alpha\beta}\mathcal{L}_{X_{(\alpha)}}\mathcal{L}_{X_{(\beta)}}\Psi + \mathcal{L}_Y\Psi \qquad [69]$$

This equation is valid on manifolds $\mathbb{M}$ (e.g., frame bundles, U($N$) bundles, multiply connected spaces, symplectic manifold phase space) in arbitrary systems of coordinates.

**Example** (Polar coordinates on $\mathbb{R}^D$). Let us abbreviate $z^\alpha(t)$ to $z^\alpha, x^1(t)$ to $r$, and $x^2(t)$ to $\theta$. It follows from

$$z^1 = r\cos\theta, \qquad z^2 = r\sin\theta \qquad [70]$$

that

$$\begin{cases} \mathrm{d}r = \cos\theta\cdot\mathrm{d}z^1 + \sin\theta\cdot\mathrm{d}z^2 \\ \quad =: X^1_{(1)}\mathrm{d}z^1 + X^1_{(2)}\mathrm{d}z^2 \\ \mathrm{d}\theta = -\dfrac{\sin\theta}{r}\mathrm{d}z^1 + \dfrac{\cos\theta}{r}\mathrm{d}z^2 \\ \quad =: X^2_{(1)}\mathrm{d}z^1 + X^2_{(2)}\mathrm{d}z^2 \end{cases} \qquad [71]$$

The dynamical vector fields are, therefore,

$$X_{(1)} = \cos\theta \frac{\partial}{\partial r} - \frac{\sin\theta}{r}\frac{\partial}{\partial\theta} \qquad [72]$$

$$X_{(2)} = \sin\theta \frac{\partial}{\partial r} + \frac{\cos\theta}{r}\frac{\partial}{\partial\theta} \qquad [73]$$

Here $h^{\alpha\beta} = \delta^{\alpha\beta}$ and eqn [69] reads

$$\frac{\partial\Psi}{\partial t} = \frac{s}{4\pi}\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2} + \frac{1}{r}\frac{\partial}{\partial r}\right)\Psi \qquad [74]$$

This example is trivial because $x(t,z)$ is not a functional of $z$ but a function of $z(t)$ given by [70]. In the following example, $x(t,z)$ is a functional of $z$.

**Example** (Paths with values on a Riemannian manifold $(\mathbb{M}^D, g)$). Consider the frame bundle over $\mathbb{M}^D$ and a connection $\sigma$ defining the horizontal lift $\dot\rho(t)$ of a vector $\dot x(t)$,

$$\dot\rho(t) = \sigma(\rho(t)) \cdot \dot x(t) \qquad [75]$$

In order to bring eqn [75] in the form [64], we think of a frame $u(t)$ as a linear map from $\mathbb{R}^D$ into the tangent space $T_{x(t)}\mathbb{M}^D$:

$$u(t) : \mathbb{R}^D \to T_{x(t)}\mathbb{M}^D \qquad [76]$$

Let

$$\dot z(t) := u(t)^{-1}\dot x(t) \qquad [77]$$

Choose a basis $\{e_{(A)}\}$ in $\mathbb{R}^D$ and $\{e_{(\alpha)}\}$ in $T_{x(t)}\mathbb{M}^D$ such that

$$\dot z(t) = \dot z^A(t) e_{(A)} = u(t)^{-1}(\dot x^\alpha(t) e_{(\alpha)}) \qquad [78]$$

Insert $u(t) \circ u(t)^{-1}$ into [75], then

$$\dot\rho(t) = X_{(A)}(\rho(t))\, \dot z^A(t) \qquad [79]$$

where the dynamical vector fields are

$$X_{(A)}(\rho(t)) = (\sigma(\rho(t)) \circ u(t)) \cdot e_{(A)} \qquad [80]$$

The construction [64]–[69] gives a parabolic equation on the bundle. If the connection $\sigma$ is the metric connection, then the parabolic equation on the bundle gives, by projection on the base space, the parabolic equation with the Laplace–Beltrami operator. Explicitly, the projection on the base space of [67] is

$$\psi(t_b, x_b) := \int_{\mathcal{P}_b\mathbb{R}^D} \mathcal{D}_{s,Q}(z) \exp\left(-\frac{\pi}{s}Q(z)\right) \times \phi((\mathrm{Dev}\, z)(t_a)) \qquad [81]$$

where Dev is the Cartan development map, namely the bijection, defined by [82], from the space of pointed paths $z$ on $T_b\mathbb{M}^D$ (identified to $\mathbb{R}^D$ via the frame $u_B$) into the space of pointed paths $x$ on $\mathbb{M}^D$ (paths such that $x(t_b) = x_b$):

$$(\Pi \circ \rho)(t) =: (\mathrm{Dev}\, z)(t) \qquad [82]$$

$\Pi$ is the projection on the base space. The path integral [81] is the solution of the equations

$$\frac{\partial}{\partial t_b}\psi(t_b, x_b) = \frac{s}{4\pi}\Delta\psi(t_b, x_b) \qquad [83]$$

$$\psi(t_a, x) = \phi(x) \qquad [84]$$

where $\Delta$ is the Laplace–Beltrami operator on $(\mathbb{M}^D, g)$,

$$\Delta = g^{ij} D_i D_j \qquad [85]$$

and $D_i$ is the covariant derivative defined by the Riemann connection $\sigma$.

## Semiclassical Expansions

Classical mechanics is a limit of quantum mechanics; therefore, it is natural to expand the action functional $S$ of a given system around, or near, its classical value – namely its minimum $\mathcal{S}(q)$, where $q$ is a solution of the Euler–Lagrange equation,

$$\mathcal{S}'(q) = 0 \qquad [86]$$

Set

$$\mathcal{S}(x) = \mathcal{S}(q) + \mathcal{S}'(q)\cdot\xi + \frac{1}{2!}\mathcal{S}''(q)\cdot\xi\xi + \frac{1}{3!}\mathcal{S}'''(q)\cdot\xi\xi\xi + \cdots \qquad [87]$$

where $x \in \mathbb{X}$ is a path

$$x : \boldsymbol{T} \to \mathbb{M}^D$$

and $\xi, \eta \in T_q\mathbb{X}$ is a vector field at $q \in \mathbb{X}$. The second variation of $\mathcal{S}$ is called its Hessian

$$\mathcal{S}''(q)\xi\eta =: \mathrm{Hess}(q; \xi, \eta) \qquad [88]$$

The arena of semiclassical expansions of a functional integral schematically written as

$$I = \int_{X_{a,b}} \mathcal{D}x \exp(\mathrm{i}\mathcal{S}(x)/\hbar)\cdot\phi((x(t_a))) \qquad [89]$$

consists of the intersection $U_{a,b}$ of two spaces $\mathbb{X}_{a,b} \subset \mathbb{X}$ the space of paths satisfying $D$ initial conditions $(a)$ and $D$ final conditions $(b)$, and $U^{2D}(\mathcal{S})$ the space of critical points of $\mathcal{S}$

$$q \in U^{2D}(\mathcal{S}), \quad \mathcal{S}'(q) = 0 \qquad [90]$$

$$U_{a,b} := \mathbb{X}_{a,b} \cap U^{2D}(\mathcal{S}) \qquad [91]$$

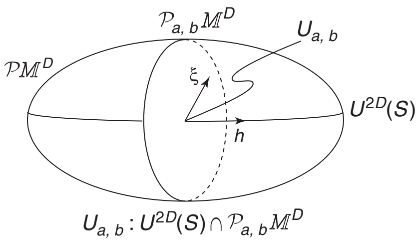



**Figure 4** Intersection of the space $\mathcal{P}_{a,b}\mathbb{M}^D$ (abbreviated to $\mathbb{X}_{a,b}$) of paths on $\mathbb{M}^D$ with fixed points, and the 2$D$-dimensional space $U^{2D}(\mathcal{S})$ of critical points of the system $\mathcal{S}$. (Adapted from a Plenum Press publication with permission by Springer-Verlag.)

The nature of the intersection $U_{a,b}$ determines the behavior of the system $\mathcal{S}$. Figure 4 shows the intersection of the space $\mathbb{X}_{a,b}$ of paths on $\mathbb{M}^D$ with fixed points. It also shows the space $U^{2D}(\mathcal{S})$ of critical points of $\mathcal{S}$.

We consider first the case in which $U_{a,b}$ consists of a single point $q$, or several isolated points $q_{(i)}$. The semiclassical expansion consists in dropping the terms beyond the Hessian:

$$I_{\text{WKB}} := \int_{\mathbb{X}_b} \mathcal{D}\xi \exp\left(\frac{2\pi i}{h}\left(\mathcal{S}(q) + \frac{1}{2}\mathcal{S}''(q)\cdot\xi\xi\right)\right) \times \phi(x(t_a)) \qquad [92]$$

where the initial wave function $\phi$ accounts for the $D$ initial conditions of the system, and $\mathbb{X}_b$ is the space of pointed paths

$$x(t_b) = x_b, \text{ and } \xi(t_b) = 0 \quad \text{for every } x \in \mathbb{X}_b \qquad [93]$$

### WKB Approximations

The integral $I_{\text{WKB}}$ is the Gaussian defined by the Hessian. Explicit calculations of $I_{\text{WKB}}$ exploit the power of Jacobi fields of $\mathcal{S}$ at $q$.

**Example** (Momentum-to-position transitions) (e.g., Cartier and DeWitt-Morette 2006). We have

$$I_{\text{WKB}}(x_b, t_b; p_a, t_a) = \exp\frac{2\pi i}{h} S(q(t_b), p(t_a)) \times \left(\det\frac{\partial^2 S}{\partial q^i(t_b)\partial p_j(t_a)}\right)^{1/2} \qquad [94]$$

where $S$ is the action function (a.k.a. Hamilton's principal function)

$$S(q(t_b), p(t_a)) = \mathcal{S}(q) + \langle p_a, x(t_a)\rangle \qquad [95]$$

where the classical path $q$ is characterized by its initial momentum $p_a$ and its final position $x_b$. The proof of [94] rests on the following property of quadratic forms $Q$. Let $L : \mathbb{X} \to \mathbb{Y}$ linearly and

$$Q_X = Q_Y \circ L \qquad [96]$$

According to the notations used in [26], [27],

$$\int_X \mathcal{D}_X(x) \exp\left(-\frac{\pi}{s} Q_X(x)\right) = 1 \qquad [97]$$

According to [35], [27],

$$1 = \int_X \mathcal{D}_X(x) \exp\left(-\frac{\pi}{s} Q_X(x)\right) = \int_Y \mathcal{D}_Y(Lx) \exp\left(-\frac{\pi}{s} Q_Y(Lx)\right) \qquad [98]$$

$$= |\det L| \int \mathcal{D}_Y(x) \exp\left(-\frac{\pi}{s} Q_X(x)\right) \qquad [99]$$

If $s = 1$, that is, if $Q_X$ and $Q_Y$ are positive definite, then

$$\int_X \mathcal{D}_Y(x) \exp(-\pi Q_X(x)) == \det(Q_X/Q_Y)^{-1/2} \qquad [100]$$

If $s = i$, that is, for Feynman integrals

$$\int_X \mathcal{D}_Y(x) \exp(-\pi Q_X(x)) = |\det(Q_X/Q_Y)|^{-1/2} i^{\text{Ind}(Q_X/Q_Y)} \qquad [101]$$

where "Ind($Q_X/Q_Y$)" is the ratio of the numbers of negative eigenvalues of $Q_X$ and $Q_Y$ respectively, and $i = \sqrt{-1} = e^{i\pi/2}$.

Equation [100] is a key equation for semiclassical expansions where it is convenient to break up the second variation $\mathcal{S}''(q)\xi\xi$ into two quadratic forms:

$$\mathcal{S}''(q)\xi\xi = Q_0(\xi) + Q(\xi) \qquad [102]$$

where $Q_0$ is the kinetic energy. The quadratic form $Q_0$ is a convenient Gaussian volume element for computing [92]. Moreover, splitting the Hessian into $Q_0 + Q$ corresponds to splitting the system into a "free" system and a perturbation.

In eqns [100] and [101] the determinant of the ratios of the infinite-dimensional quadratic forms $Q_X/Q_Y$ have been shown (Cartier and DeWitt-Morette 2006) to be a finite-dimensional determinant, thanks to Jacobi field technology.

### Degenerate Hessians; Beyond WKB

When $U_{a,b}$ consists of isolated points, the Hessian is not degenerate, and the semiclassical expansion is usually called the (strict) WKB approximation. When the Hessian is degenerate,

$$\mathcal{S}''(q)\xi\xi = 0 \quad \text{for} \quad \xi \neq 0 \qquad [103]$$

there is at least one nonzero Jacobi field $h$ along $q$,

$$\mathcal{S}''(q)h = 0, \quad h \in T_q U^{2D}(\mathcal{S}) \qquad [104]$$

with $D$ vanishing initial conditions ($a$) and $D$ vanishing final conditions ($b$). Equation [104] is the defining equation of Jacobi fields. The vanishing boundary conditions imply that $h \in T_q\mathbb{X}_{a,b}$ as well as being a Jacobi field.

For understanding the intersections $U_{a,b}$ when the Hessian is degenerate, one can construct the following basis for the intersecting tangent spaces $T_qU^{2D}(\mathcal{S})$ and $T_q\mathbb{X}_{a,b}$:

- Basis for $T_qU^{2D}(\mathcal{S})$: a complete set (if it exists) of linearly independent Jacobi fields. It can be constructed by varying the $2D$ conditions ($a$), ($b$) satisfied by $q \in \mathbb{X}_{a,b}$.
- Basis for $T_q\mathbb{X}_{a,b}$: a complete set of orthonormal eigenvectors $\{\Psi_k\}$ of the Jacobi operator $\mathcal{J}(q)$ defined by the Hessian

$$\mathcal{S}''(q) \cdot \xi\xi =: \langle\langle \mathcal{J}(q), \xi\rangle, \xi\rangle \qquad [105]$$

$$\mathcal{J}(q)\Psi_k = \alpha_k\Psi_k, \quad k \in \{0, 1, \ldots\} \qquad [106]$$

The basis $\{\Psi_k\}$ diagonalizes the Hessian. When the Hessian is degenerate, there is at least one eigenvector of $\mathcal{J}(q)$ with zero eigenvalue.

1. The intersection $U_{a,b}$ is of dimension $l > 0$. Let $\{u^k\}$ be the coordinates of $\xi$ in the $\{\Psi_k\}$ basis of $T_q\mathbb{X}_{a,b}$. Then the diagonalized Hessian is

$$\mathcal{S}''(q) \cdot \xi\xi = \sum_{k=0}^{\infty} \alpha_k (u^k)^2 \qquad [107]$$

There are $l$ zero eigenvalues $\{\alpha_k\}$ when the system of Euler–Lagrange equations decouples (possibly after a change of variable in $\mathbb{X}_{a,b}$) into two sets: $l$ constraint equations, and $D - l$ equations determining $D - l$ coordinates $\{q^A\}$ of $q$. Say $l = 1$, for simplicity. Then

$$\mathcal{S}(x) = \mathcal{S}(q) + c_0u^0 + \frac{1}{2}\sum_{k=1}^{\infty} \alpha_k (u^k)^2 + \mathcal{O}(|u|^3) \qquad [108]$$

where

$$c_0 = \int_T \mathrm{d}t \frac{\delta\mathcal{S}}{\delta q^j(t)} \Psi_0^j(t) \qquad [109]$$

The change of variable $\xi \to \{u^k\}$ is a linear change of variable of type [33]. The integral [92] decomposes into the product of an ordinary integral over $u^0$ and a Gaussian functional integral defined by a nondegenerate quadratic form. The integral over $u^0$ yields a Dirac $\delta$-function, $\delta(c_0/h)$. The propagator vanishes unless the conservation law $c_0 = 0$ is satisfied.

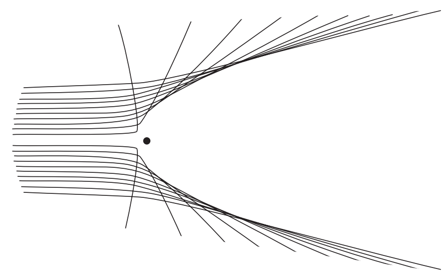


**Figure 5** A flow of particles scattered by a repulsive Coulomb potential. (Reprinted from *Physical Review D* with permission by the American Physical Society.)

Conservation laws appear in the classical limit of quantum physics. The quantum system may have less symmetry than its classical limit.

2. The intersection $U_{a,b}$ is a multiple root of the Euler–Lagrange equation. The flow of classical solutions has an envelope, known as a caustic. Caustics abound in physics: the soap bubble problem, scattering of particles by a repulsive Coulomb potential (see **Figure 5**), rainbow scattering from a source at infinity, glory scattering etc. (Cartier and DeWitt-Morette 2006).

Let us consider a specific example for simplicity. For instance, the scattering of particles of given momenta $p_a$ by a repulsive Coulomb potential. Let $q$ and $q^\Delta$ be two solutions of the Euler–Lagrange equation with slightly different boundary conditions at $t_b$. Compute $I(x_b^\Delta, t_b; p_a, t_a)$ by expanding the action functional not around $q^\Delta$ but around $q$. The path $q^\Delta$ is not in $\mathbb{X}_{a,b}$ and the expansion of the action functional has to be carried up to and including the third variation. As before, let $\{u^k\}$ be the coordinates of $\xi$ in the base $\{\Psi_k\}, k \in \{0, 1 \ldots\}$. The integral over $u^0$ is an Airy integral

$$\nu^{-1/3}\mathrm{Ai}\left(\nu^{-1/3}c\right) \int_{\mathbb{R}} \mathrm{d}u^0 \exp\left(\mathrm{i}\left(cu^0 + \frac{\nu}{3}\left(u^0\right)^3\right)\right) \qquad [110]$$

where

$$\nu = \frac{\pi}{\hbar}\int_T \mathrm{d}r \int_T \mathrm{d}s \int_T \mathrm{d}t \frac{\delta^3\mathcal{S}}{\delta q^\alpha(r)\delta q^\beta(s)\delta q^\gamma(t)} \times \Psi_0^\alpha(r)\Psi_0^\beta(s)\Psi_0^\gamma(t) \qquad [111]$$

$$c = -\frac{2\pi}{h}\int_T \mathrm{d}t \frac{\delta\mathcal{S}}{\delta q(t)} \cdot \Psi_0(t)\left(x_b^\Delta - x_b\right) \qquad [112]$$

The leading contribution of the Airy function when $\hbar$ tends to zero can be computed by the stationary phase method. When $x_b^\Delta$ is in the "illuminated" region, the probability amplitude $I(x_b^\Delta, t_b; p_a, t_a)$ oscillates rapidly as $\hbar$ tends to zero. When $x_b^\Delta$ is in the "dark" region, the probability amplitude decays exponentially. Quantum mechanics softens up the caustics.

The two kinds of degeneracies described in sections (1) and (2) may occur simultaneously. This happens, for instance, in glory scattering for which the cross section, to leading terms in the semiclassical expansions, has been obtained by functional integration in closed form in terms of Bessel functions (Cartier and DeWitt-Morette 2005).

3. The intersection $U_{a,b}$ is the empty set. There is no classical solution corresponding to the quantum transition. This phenomenon, called "tunneling" or "barrier penetration," is a rich chapter of quantum physics which can be found in most of the books listed under "Further reading."

## A Multipurpose Tool

Functional integration provides insight and techniques to quantum physics not available from the operator formalism. Just as an example, one can quote the section "Beyond WKB" which has often been dismissed in the operator formalism by stating that "WKB breaks down" in such cases.

The power of functional integration stems from the power of infinite-dimensional spaces. For instance, compare the Lagrangian of a system with its action functional

$$\mathcal{S}(x) = \int_{\boldsymbol{T}} \mathrm{d}t\, L(\dot{x}(t), x(t)), \quad x \in \mathbb{X}_{a,b}$$
$$x : \boldsymbol{T} \to \mathbb{M}^D, \qquad \mathcal{S} : \mathbb{X}_{a,b} \to \mathbb{R} \qquad [113]$$

A classical solution $q$ of the system can be defined either by a solution of the Euler–Lagrange equation, together with the boundary conditions dictated by $q \in \mathbb{X}_{a,b}$ or by an extremum of the action functional, $\mathcal{S}'(q) = 0$. The path $q$ is a significant point in $\mathbb{X}_{a,b}$ but it is not isolated and the Hessian $\mathcal{S}''(q)$ gives much information on $q$, such as conservation laws, caustics, tunneling.

A list of applications is beyond the scope of this article. We treat only two applications, then give in the "Further reading" section a short list of books that develop such applications as polarons, phase transitions, properties of quantum gases, scattering processes, many-body theory of bosons and fermions, knot invariants, quantum crystals, quantum field theory, anomalies, etc.

### The Homotopy Theorem for Paths Taking Their Values in a Multiply-Connected Space

The space $\mathbb{X}_{a,b}$ of paths $x$

$$x : \boldsymbol{T} \to \mathbb{M}^D, \quad x \in \mathbb{X}_{a,b}$$

probes the global properties of their ranges $\mathbb{M}^D$. When $\mathbb{M}^D$ is multiply connected, $\mathbb{X}_{a,b}$ is the sum of distinct homotopy classes of paths. The integral over $\mathbb{X}_{a,b}$ is a linear combination of integrals over each homotopy class of paths. The coefficients of this linear combinations are provided by the homotopy theorem.

The principle of superposition of quantum states requires the probability amplitude for a given transition to be a linear combination of probability amplitudes. It follows that the absolute value of the probability amplitude for a transition from the state $a$ at $t_a$ to the state $b$ at $t_b$ has the form

$$|K(b, t_b; a, t_a)| = \left| \sum_\alpha \chi(\alpha) K^\alpha(b, t_b; a, t_a) \right| \qquad [114]$$

where $K^\alpha$ is the interval over paths in the same homotopy class. The homotopy theorem (Laidlaw and Morette-DeWitt 1971) and (Schulman 1971) in Cartier and DeWitt-Morette (2006)) states that the set $\{\chi(\alpha)\}$ forms a representation of the fundamental group of the multiply connected space $\mathbb{M}^D$. One cannot label a homotopy class by an element of the fundamental group unless one has chosen a point $c \in \mathbb{M}^D$ and a homotopy class for paths going from $c$ to $a$ and for paths going from $c$ to $b$ – in brief, unless one has chosen a homotopy mesh on $\mathbb{M}^D$. The fundamental group based at $c$ is isomorphic to the fundamental group based at any other point of $\mathbb{M}^D$ but not canonically so. Therefore, eqn [114] is only an equality between absolute values of probability amplitudes. The proof of the homotopy theorem consists in requiring [114] to be independent of the chosen homotopy mesh.

### Application: Systems of *n*-Indistinguishable Particles in $\mathbb{R}^D$

In order that there be a one-to-one correspondence between the system and its configuration space,

$$x : \boldsymbol{T} \to \mathbb{R}^{D \otimes n}/S_n =: \mathbb{R}^{D,n}$$

where $S_n$ is the symmetric group for $n$ permutations; the coincidence points in $\mathbb{R}^{D,n}$ are excluded so that $S_n$ acts effectively on $\mathbb{R}^{D,n}$. Note that $\mathbb{R}^{1,n}$ is not connected, but $\mathbb{R}^{2,n}$ is multiply connected. When $D \geq 3$, $\mathbb{R}^{D,n}$ is simply connected and the fundamental group on $\mathbb{R}^{D,n}$ is isomorphic to $S_n$.

There are only two scalar unitary representations of $S_n$:

$$\chi^B : \alpha \in S_n \to 1 \qquad \text{for all permutations } \alpha$$
$$\chi^F : \alpha \in S_n \to \begin{cases} 1 & \text{for even permutations} \\ -1 & \text{for odd permutations} \end{cases}$$

Therefore, in $\mathbb{R}^3$ there are two different propagators of indistinguishable particles:

$$K^{\text{bose}} = \sum_{\alpha} \chi^B(\alpha) K^{\alpha} \qquad [115]$$

is a symmetric propagator

$$K^{\text{fermi}} = \sum_{\alpha} \chi^F(\alpha) K^{\alpha} \qquad [116]$$

is an antisymmetric propagator.

The arguments leading to the existence of (scalar) bosons and fermions in $\mathbb{R}^3$ fails in $\mathbb{R}^2$. Statistics cannot be assigned to particles in $\mathbb{R}^2$; particles "without" statistics have been called anyons.

### Application: a Spinning Top

Schulman's analysis of the Schrödinger equation for a spinning top (Schulmann 1968) motivated the formulation of the homotopy theorem. Therefore, Schulman's results can easily be formulated as an application of [114].

### Application: Instantons (DeWitt 2004)

The homotopy theorem reformulated for functional integrals applies to the total $\langle \text{out} | \text{in} \rangle$ amplitude of instantons in Minkowski spacetime.

### Scaling Properties of Gaussians

We rewrite the definition [26] of Gaussian volume elements as

$$\int_X d\Gamma_G(x) \exp(-2\pi i \langle x', x \rangle) := \exp(-\pi i W(x')) \qquad [117]$$

where the covariance $G$ is defined by the variance $W$,

$$W(x') = \langle x', Gx' \rangle$$

In quantum field theory the definition [26] reads

$$\int_\Phi d\Gamma_G(\varphi) \exp(-2\pi i \langle J, \varphi \rangle) := \exp(-\pi i W(J)) \qquad [118]$$

where $\varphi$ is a field on spacetime (Minkowski, or Euclidean) and $J$ is called the source. A Gaussian $\Gamma_G$ can be decomposed into the convolution of any number of Gaussians. For example, if

$$W = W_1 + W_2 \longrightarrow G = G_1 + G_2 \qquad [119]$$

then

$$\Gamma_G = \Gamma_{G_1} * \Gamma_{G_2} \qquad [120]$$

Explicitly, in QFT

$$\int_\Phi d\Gamma_G(\varphi) \exp(-2\pi i \langle J, \varphi \rangle) = \int_\Phi d\Gamma_{G_2}(\varphi_2) \int d\Gamma_{G_1}(\varphi_1) \times \exp(-2\pi i \langle J, \varphi_1 + \varphi_2 \rangle) \qquad [121]$$

where

$$\varphi = \varphi_1 + \varphi_2 \qquad [122]$$

The additive property [119] makes it possible to express a covariance $G$ as an integral over an independent scale variable.

Let $\lambda \in [0, \infty]$ be an independent scale variable. (some authors use $\lambda \in [1, \infty[$ and $\lambda^{-1} \in [0, 1[$). A scale variable has no physical dimension:

$$[\lambda] = 0 \qquad [123]$$

The scaling operator $S_\lambda$ acting on a function $f$ of length dimension $[f]$ is by definition

$$S_\lambda f(x) := \lambda^{[f]} f(x/\lambda) \qquad [124]$$

the scaling of an interval $[a, b[$ is given by $S_\lambda[a, b[ = \{s/\lambda | s \in [a, b[\}$, that is,

$$S_\lambda[a, b[ = [a/\lambda, b/\lambda[ \qquad [125]$$

The scaling of a functional $F$ is

$$(S_\lambda F)(\varphi) = F(S_\lambda \varphi) \qquad [126]$$

In order to decompose a covariance into an integral of scale-dependent contributions we note that a covariance $G$ is a two-point function [31]. In quantum field theory [118], the engineering length dimension of $G$ is twice the field dimension

$$[G] = 2[\varphi] \qquad [127]$$

Let $x, y \in$ spacetime and $G$ be a Laplacian Green function. One can introduce a scaled (truncated) Green function

$$G_{[l_0, l[}(x, y) := \int_{l_0}^{l} d^\times s\, S_{s/l_0} u(|x - y|) \qquad [128]$$

where

$$[l] = 1, \quad [s] = 1, \quad d^\times s = ds/s$$
$$l_0 \leq l, \quad [u] = [G] \qquad [129]$$

such that

$$\lim_{l_0 = 0, l = \infty} G_{[l_0, l[}(x, y) = G(x, y) \qquad [130]$$

**Example** $G(x,y)=c_D/|x-y|^{D-2}$; then the only requirement on the function $u$ in [128] is

$$\int_0^\infty \mathrm{d}^\times r r^{-2[\varphi]} u(r) = c_D, \quad [r]=1 \qquad [131]$$

All objects defined by the scaled covariance [128] are labeled with the interval $[l_0, l[$. For instance, a Gaussian volume element $\Gamma_{G_{[l_0,l[}}$ is abbreviated to $\Gamma_{[l_0,l[}$.

### A Coarse-Graining Operator

The following coarse-graining operator has been used for constructing a parabolic semigroup equation in the scaling variable (Brydges *et al.* 1998):

$$P_l F := S_{l/l_0} \cdot \Gamma_{[l_0,l[} * F \qquad [132]$$

where the convolution product is by definition

$$(\Gamma_{[l_0,l[} * F)(\varphi) = \int_\Phi \mathrm{d}\Gamma_{[l_0,l[}(\psi) F(\varphi + \psi)$$

The coarse-graining operator $P_l$ rescales the convolution of a Gaussian volume element $\Gamma_{[l_0,l[}$ so that all volume elements entering the construction of the semigroup renormalization equation are scale independent.

Some properties of the coarse-graining operator:

- $P_{l_2} P_{l_1} = P_{l_2 l_1 / l_0}$.
- The scaled eigenfunctions of the coarse-graining operator are Wick-ordered monomials (Wurm and Berg 2002)

$$P_l : \varphi^n(x) :_{[l_0,\infty[} = \left(\frac{l}{l_0}\right)^{n[\varphi]} : \varphi^n\left(\frac{l_0}{l} x\right) :_{[l_0,\infty[} \qquad [133]$$

  Note that $P_l$ preserves the scale range.
- Let $H$ be the generator of the coarse-graining operator

$$H := \frac{\partial^\times}{\partial l} P_l \Big|_{l=l_0}, \quad \frac{\partial^\times}{\partial l} = l \frac{\partial}{\partial l} \qquad [134]$$

  The semigroup renormalization equation (a.k.a. the flow equation)

$$\frac{\partial^\times}{\partial l} P_l F(\varphi) = H P_l F(\varphi)$$
$$P_{l_0} F(\varphi) = F(\varphi) \qquad [135]$$

Brydges *et al.* have applied the coarse-graining operator to the quantum field theory known as "$\lambda\varphi^4$" (more precisely the Wick-ordered Lagrangian of $\lambda\varphi^4$). The flow equation [135] plays the role of the "$\beta$-function" equation in perturbative quantum field theory.

### Functional Integrals in Quantum Field Theory

Functional integrals in quantum field theory have been modeled to some extent on path integrals in quantum mechanics: mutatis mutandis, the definition [23] of Gaussian volume elements, the diagram expansion [30], the property [36] of linear maps, semiclassical expansions [87], the homotopy theorem [114], and the scaling eqns [135] apply to functional integrals in quantum field theory. The time ordering encoded in a path integral becomes a chronological ordering dictated by light cones in functional integrals of fields on Minkowski fields.

The fundamental difference between quantum mechanics (systems with a finite number of degrees of freedom) and quantum field theory (systems with an infinite number of degrees of freedom) can be said to be "radiative corrections." In quantum field theory, the concept of "particle" is intrinsically associated to the concept of "field." A particle is affected by its field. Its mass and charge are modified by the surrounding fields, namely its own and other fields interacting with it. One speaks of "bare mass" and "renormalized mass" when the bare mass is renormalized by surrounding fields. Computing radiative corrections is a delicate procedure because the Green functions $G$ defined by [25] are singular. Regularization techniques have been developed for handling singular Green functions.

Particles in quantum mechanics are simply particles, and bosons and fermions can be treated separately. Not so in quantum field theory. Therefore, the configuration space in quantum field theory is a supermanifold. For functional integrals in this theory, we refer the reader to the "Further reading" section, in particular to the book of A Das for an introduction, to the book of B DeWitt for an in-depth study, and to the book of K Fujikawa and H Suzuki for applications to quantum anomalies.

## Concluding Remarks

The key issue in functional integration is the domain of integration, that is, a function space. This infinite-dimensional space, say $\mathbb{X}$, cannot be considered as the limit $n=\infty$ of $\mathbb{R}^n$.

Concepts of $\mathbb{R}^D$ stated without reference to $D$ are likely to be meaningful on $\mathbb{X}$. Other approaches which have been used for exploring $\mathbb{X}$ are

- projective system of finite-dimensional spaces coherently defined on $\mathbb{X}$ (DeWitt-Morette *et al.* 1979),
- one-parameter curves on $\mathbb{X}$ (**Figure 1**), and
- projecting $\mathbb{X}$ on finite-dimensional spaces (cylindrical integrals).

Functional integration has advanced our understanding of infinite-dimensional spaces, and like all good mathematical tools, it improves with usage.

*See also:* BRST Quantization; Euclidean Field Theory; Feynman Path Integrals; Infinite-Dimensional Hamiltonian Systems; Knot Theory and Physics; Malliavin Calculus; Path Integrals in Noncommutative Geometry; Quantum Mechanics: Foundations; Stationary Phase Approximation; Topological Sigma Models.

## Bibliography/Further Reading

The book (Cartier and DeWitt-Morette 2006) was nearing completion while this article was being written; many results and brief comments in this article are fully developed in the book. The book (Grosche and Steiner 1998) includes a bibliography with 945 references.


Albeverio SA and Hoegh-Krohn RJ, *Mathematical Theory of Feynman Path Integrals*, Lecture Notes in Mathematics, vol. 523. Berlin: Springer.

Brillouin ML (1926) La méchanique ondulatoire de Schrödinger; une méthode générale de resolution par approximations successives. *Comptes Rendus Acad. Sci.* 183: 24–44. Remarques sur la Mécanique Ondulatoire. *Journal of Physics Radium* 7: 353–368.

Brydges DC, Dimock J, and Hurd TR (1998) Estimates on renormalization group transformations. *Can. J. Math.* 50: 756–793. A non-gaussian fixed point for $O^4$ in 4-dimensions. *Communications in Mathematical Physics* 198: 111–156.

Cartier P and DeWitt-Morette C (2006) *Functional Integration; Action and Symmetries*. Cambridge: Cambridge University Press.

Chaichian M and Demichev A (2001) *Path Integrals in Physics*, vol. I and II. Bristol, UK: Institute of Physics.

Collins D (1997) Functional integration over complex Poisson processes, an appendix to a rigorous mathematical foundation of functional integration by Cartier P and Dewitt-Morette. In: Dewitt-Morette C, Cartier P, and Folacci A (eds.) *Functional Integration: Basic and Applications*, pp. 42–50. New York: Plenum.

Das A (1993) *Field Theory, a Path Integral Approach*. Singapore: World Scientific.

DeWitt B (2004) *The Global Approach to Quantum Field Theory*. Oxford: Oxford University Press.

DeWitt-Morette C, Maheswari A, and Nelson B (1979) Path integration in non-relativistic quantum mechanics. *Physics Reports* 50: 255–372.

Dirac PAM (1933) The Lagrangian in quantum mechanics. *Physics Zeitschrift Sowjetunion* 3: 64–72. Reprinted in *Quantum Electrodynamics*. Edited by J Schwinger. Dover, New York, 1958, 312–320.

Elworthy KD (1982) *Stochastic Differential Equations on Manifolds*. Cambridge: Cambridge University Press.

Feynman RP (1966) The development of the space-time view of quantum electrodynamics, Nobel Prize Award Address, Stockholm, December 11, 1965. *Lex Prix Nobel en 1965*. Reprinted in *Physics Today*. 19(8): 31–44. Stockholm: Nobel Foundation.

Feynman RP and Hibbs AR (1961) *Quantum Mechanics and Path Integrals*. New York: McGraw-Hill.

Fujikawa K and Suzuki H (2004) *Path Integrals and Quantum Anomalies* (Originally published in Japanese in 2001 by Iwanami Shoten Pub. Tokyo.). Oxford: Oxford University Press.

Glimm J and Jaffe A (1987) *Quantum Physics: A Functional Integral Point of View*. New York: Springer.

Grosche G and Steiner F (1998) Handbook of Feynman Path Integrals. *Springer Tracts in Modern Physics*. Berlin: Springer.

Johnson GW and Lapidus ML (2000) *The Feynman Integral and Feynman's Operational Calculus*. Oxford: Oxford University Press.

Kleinert H (2004) *Path Integrals in Quantum Mechanics, Statistics, and Polymer Physics*, 3rd edn. Singapore: World Scientific.

Kramers HA (1927) Wellenmechanik und halbzahlige Quantisierung. *Zeitschrift Physics* 39: 828–840.

LaChapelle J (2004) Path integral solution of linear second order partial differential equations, I. The general construction; II. Elliptic, parabolic, and hyperbolic cases. *Annals in Physics* 314: 262–424.

Laidlaw MGG and DeWitt CM (1971) Feynman functional integrals for systems of indistinguishable particles. *Physical Review D* 3: 1375–1378.

Marinov MS (1980) Path integrals in quantum theory: an outlook of basic concepts. *Physics Reports* 60: 1.

Morette C (1951) On the definition and approximation of Feynman's path integrals. *Physical Review* 81: 848–852.

Roepstorff G (1994) *Path Integral Approach to Quantum Physics: An Introduction* (Original German edition: Pfadintegrale in der Quantenphysik, Braunschweig: Vieweg 1991.). Berlin: Springer.

Schulman LS (1968) A path integral for spin. *Physical Review* 176: 1558–1569.

Schulman LS (1971) Approximate topologies. *Journal of Mathematical Physics* 12: 304–308.

Schulman LS (1981) *Techniques and Applications of Path Integration*. New York: Wiley.

Simon B (1979) *Functional Integration and Quantum Physics*. New York: Academic Press.

t'Hooft G and Veltmann HJG (1973) Diagrammer. *CERN Preprint* 73–9: 1–114.

Wenzel G (1926) Eine Verallgemeinerung der Quantembedingungen für die Zwecke der Wellenmechanik. *Zeitschrift Physics* 38: 518–529.

Wurm A and Berg M (2002) Wick calculus. ArXiv:Physics/0212061.

Zinn-Justin J (2003) *Intégrale de chemin en mécanique quantique: Introduction* (Path Integrals in Quantum Mechanics, Oxford University Press, 2004). Paris: EDP Sciences, Les Ulis et CNRS Editions.

# G

## Γ-Convergence and Homogenization

**G Dal Maso**, SISSA, Trieste, Italy



### Introduction

Several asymptotic problems in the calculus of variations lead to the following question: given a sequence $\mathcal{F}_k$ of functionals, defined on a suitable function space, does there exist a functional $\mathcal{F}$ such that the solutions of the minimum problems for $\mathcal{F}_k$ converge to the solutions of the corresponding minimum problems for $\mathcal{F}$? Γ-convergence, introduced by Ennio De Giorgi and his collaborators in 1975, and developed as a powerful tool to attack a wide range of applied problems, provides a unified answer to this kind of question.

### Definition and Main Properties

Let $\mathcal{U}$ be a topological space with a countable base and let $\mathcal{F}_k$ be a sequence of functions defined on $\mathcal{U}$ with values in the extended real line $\overline{\boldsymbol{R}} := \boldsymbol{R} \cup \{-\infty, +\infty\}$. We say that $\mathcal{F}_k$ Γ-converges to a function $\mathcal{F}:\mathcal{U} \to \overline{\boldsymbol{R}}$, or that $\mathcal{F}$ is the Γ-limit of $\mathcal{F}_k$, if for every $u \in \mathcal{U}$ the following conditions are satisfied:

1. For every sequence $u_k$ converging to $u$ in $\mathcal{U}$ we have
$$\mathcal{F}(u) \le \liminf_{k\to\infty} \mathcal{F}_k(u_k)$$
2. There exists a sequence $u_k$ converging to $u$ in $\mathcal{U}$ such that
$$\mathcal{F}(u) = \lim_{k\to\infty} \mathcal{F}_k(u_k)$$

Property (1) appears to be a variant of the usual definition of lower semicontinuity. Property (2) requires the existence, for every $u \in \mathcal{U}$, of a "recovery sequence," which provides an approximation of the value of $\mathcal{F}$ at $u$ by means of values attained by $\mathcal{F}_k$ near $u$.

It follows immediately from the definition that, if $\mathcal{F}_k$ Γ-converges to $\mathcal{F}$, then $\mathcal{F}_k + \mathcal{G}$ Γ-converges to $\mathcal{F} + \mathcal{G}$ for every continuous function $\mathcal{G}:\mathcal{U} \to \boldsymbol{R}$.

The first general property of Γ-limits is lower semicontinuity: if $\mathcal{F}_k$ Γ-converges to $\mathcal{F}$, then $\mathcal{F}$ is lower semicontinuous on $\mathcal{U}$; that is,
$$\mathcal{F}(u) \le \liminf_{k\to\infty} \mathcal{F}(u_k)$$
for every $u \in \mathcal{U}$ and for every sequence $u_k$ converging to $u$ in $\mathcal{U}$.

Another important property of Γ-convergence is compactness: every sequence $\mathcal{F}_k$ has a Γ-convergent subsequence.

For every $k$ assume that the function $\mathcal{F}_k$ has a minimum point $u_k$. The following property is the link between Γ-convergence and convergence of minimizers: if $\mathcal{F}_k$ Γ-converges to $\mathcal{F}$ and $u_k$ converges to $u$, then $u$ is a minimum point of $\mathcal{F}$ and $\mathcal{F}_k(u_k)$ converges to $\mathcal{F}(u)$, hence
$$\min_{v\in\mathcal{U}} \mathcal{F}(v) = \lim_{k\to\infty} \min_{v\in\mathcal{U}} \mathcal{F}_k(v) \qquad [1]$$

Under suitable coerciveness assumptions, the convergence of $u_k$ is obtained by a compactness argument. We recall that a sequence of functions $\mathcal{F}_k$ is said to be equicoercive if for every $t \in \boldsymbol{R}$ there exists a compact set $\mathcal{K}_t$ (independent of $k$) such that
$$\{u \in \mathcal{U}: \mathcal{F}_k(u) \le t\} \subset \mathcal{K}_t \qquad [2]$$
for every $k$.

If $\mathcal{F}_k$ is equicoercive and Γ-converges to $\mathcal{F}$, the previous result implies that [1] holds. If, in addition, $\mathcal{F}$ is not identically $+\infty$, then the sequence $u_k$ of minimizers considered above has a subsequence $u_{k_j}$ which converges to a minimizer $u$ of $\mathcal{F}$. The whole sequence $u_k$ converges to $u$ whenever $\mathcal{F}$ has a unique minimizer $u$.

In many applications to the calculus of variations, $\mathcal{U}$ is the Lebesgue space $L^p(\Omega;\boldsymbol{R}^m)$, with $\Omega$ a bounded open subset of $\boldsymbol{R}^n$ and $1 \le p < +\infty$, but the effective domains of the functionals $\mathcal{F}_k$, defined as $\{u \in \mathcal{U}: \mathcal{F}_k(u) \in \boldsymbol{R}\}$, are often contained in the Sobolev space $W^{1,p}(\Omega;\boldsymbol{R}^m)$, composed of all functions $u \in L^p(\Omega;\boldsymbol{R}^m)$ whose distributional gradient

$\nabla u$ belongs to $L^p(\Omega;\boldsymbol{R}^{m\times n})$. When one considers homogeneous Dirichlet boundary conditions, the effective domains of the functionals $\mathcal{F}_k$ are often contained in the smaller Sobolev space $W_0^{1,p}(\Omega;\boldsymbol{R}^m)$, composed of all functions of $W^{1,p}(\Omega;\boldsymbol{R}^m)$ which vanish on the boundary $\partial\Omega$, technically defined as the closure of $C_0^\infty(\Omega;\boldsymbol{R}^m)$ in $W^{1,p}(\Omega;\boldsymbol{R}^m)$.

In this case, the equicoerciveness condition [2] can be obtained by using Rellich's theorem, which asserts that the natural embedding of $W_0^{1,p}(\Omega;\boldsymbol{R}^m)$ into $L^p(\Omega;\boldsymbol{R}^m)$ is compact. Therefore, a sequence of functionals $\mathcal{F}_k$ defined on $L^p(\Omega;\boldsymbol{R}^m)$ is equicoercive if there exists a constant $\alpha > 0$ such that

$$\mathcal{F}_k(u) \geq \alpha \int_\Omega |\nabla u|^p \mathrm{d}x$$

for every $u \in W_0^{1,p}(\Omega;\boldsymbol{R}^m)$, while $\mathcal{F}_k(u) = +\infty$ for every $u \notin W_0^{1,p}(\Omega;\boldsymbol{R}^m)$.

## Homogenization Problems

Many problems for composite materials (fibered or stratified materials, porous media, materials with many small holes or fissures, etc.) lead to the study of mathematical models with many interacting scales, which may differ by several orders of magnitude. From a microscopic viewpoint, the systems considered are highly inhomogeneous. Typically, in such composite materials, the physical parameters (such as electric and thermal conductivity, elasticity coefficients, etc.) are discontinuous and oscillate between the different values characterizing each component.

When these components are intimately mixed, these parameters oscillate very rapidly and the microscopic structure becomes more and more complex. On the other hand, the material becomes quite simple from a macroscopic point of view, and it tends to behave like an ideal homogeneous material, called "homogenized material." The purpose of the mathematical theory of homogenization is to describe this limit process when the parameters which describe the fineness of the microscopic structure tend to zero.

Homogenization problems are often treated by studying the partial differential equations that govern the physical properties under investigation. Due to the small scale of the microscopic structure, these equations contain some small parameters. The mathematical problem consists then in the study of the limit of the solutions of these equations when the parameters tend to zero. Γ-convergence is a very useful tool to obtain homogenization results for systems governed by variational principles, which are the only ones described in this article.

Let $Q := (-1/2, 1/2)^n$ be the open unit cube in $\boldsymbol{R}^n$ centered at 0. We say that a function $u$ defined on $\boldsymbol{R}^n$ is $Q$-periodic if, for every $z \in \boldsymbol{R}^n$ with integer coordinates, we have $u(x+z) = u(x)$ for every $x \in \boldsymbol{R}^n$.

Let $f : \boldsymbol{R}^n \times \boldsymbol{R}^{m\times n} \to [0,+\infty)$ be a function such that $x \mapsto f(x,\xi)$ is measurable and $Q$-periodic on $\boldsymbol{R}^n$ for every $\xi \in \boldsymbol{R}^{m\times n}$ and $\xi \mapsto f(x,\xi)$ is convex on $\boldsymbol{R}^{m\times n}$ for every $x \in \boldsymbol{R}^n$. Given a bounded open set $\Omega \subset \boldsymbol{R}^n$ and a constant $p > 1$, let $\mathcal{F}_\varepsilon : L^p(\Omega;\boldsymbol{R}^m) \to [0,+\infty]$ be the family of functionals defined by

$$\mathcal{F}_\varepsilon(u) := \begin{cases} \int_\Omega f(x/\varepsilon, \nabla u)\,\mathrm{d}x & \text{if } u \in W_0^{1,p}(\Omega;\boldsymbol{R}^m) \\ +\infty & \text{otherwise} \end{cases}$$

In the applications to composite materials, the functional $\mathcal{F}_\varepsilon$ represents the energy of the portion of the material occupying the domain $\Omega$. The fact that the energy density depends on $x/\varepsilon$ reflects the $\varepsilon$-periodic structure of the material, which implies that the energy density oscillates faster and faster as $\varepsilon \to 0$.

Assume that there exist two constants $\beta \geq \alpha > 0$ such that

$$\alpha|\xi|^p \leq f(x,\xi) \leq \beta(1+|\xi|^p) \qquad [3]$$

for every $x \in \Omega$ and every $\xi \in \boldsymbol{R}^{m\times n}$. Then for every sequence $\varepsilon_k \to 0$ the functionals $\mathcal{F}_{\varepsilon_k}$ Γ-converge to the functional $\mathcal{F}_{\text{hom}} : L^p(\Omega;\boldsymbol{R}^m) \to [0,+\infty]$ defined by

$$\mathcal{F}_{\text{hom}}(u) := \begin{cases} \int_\Omega f_{\text{hom}}(\nabla u)\,\mathrm{d}x & \text{if } u \in W_0^{1,p}(\Omega;\boldsymbol{R}^m) \\ +\infty & \text{otherwise} \end{cases} \qquad [4]$$

The integrand $f_{\text{hom}} : \boldsymbol{R}^{m\times n} \to [0,+\infty)$ is obtained by solving the cell problem

$$f_{\text{hom}}(\xi) := \min_{w \in W^{1,p}_{\text{per}}(Q;\boldsymbol{R}^m)} \int_Q f(x, \xi + \nabla w)\,\mathrm{d}x \qquad [5]$$

where $W^{1,p}_{\text{per}}(Q;\boldsymbol{R}^m)$ denotes the space of functions $w \in W^{1,p}_{\text{loc}}(\boldsymbol{R}^n;\boldsymbol{R}^m)$ which are $Q$-periodic.

The function $f_{\text{hom}}$ is always convex and satisfies [3]. If it is strictly convex, the basic properties of Γ-convergence imply that for every $g \in L^q(\Omega;\boldsymbol{R}^m)$, with $1/p + 1/q = 1$, the solutions $u_\varepsilon$ of the minimum problems

$$\min_{v \in W_0^{1,p}(\Omega;\boldsymbol{R}^m)} \int_\Omega \left[ f\left(\frac{x}{\varepsilon}, \nabla v\right) - g(x)v \right] \mathrm{d}x \qquad [6]$$

converge in $L^p(\Omega;\boldsymbol{R}^m)$, as $\varepsilon \to 0$, to the solution $u$ of the minimum problem

$$\min_{v \in W_0^{1,p}(\Omega;\boldsymbol{R}^m)} \int_\Omega [f_{\text{hom}}(\nabla v) - g(x)v]\,\mathrm{d}x \qquad [7]$$

Similar results can be proved for nonhomogeneous Dirichlet boundary conditions, as well as for Neumann boundary conditions.

In the special case $m=1$, $p=2$, and

$$f(x,\xi) = \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x)\xi_j\,\xi_i \qquad [8]$$

with $a_{ij}(x)$ $Q$-periodic, the function $f_{\text{hom}}$ takes the form

$$f_{\text{hom}}(\xi) = \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}^{\text{hom}}\xi_j\,\xi_i$$

for suitable constant coefficients $a_{ij}^{\text{hom}}$.

By considering the Euler equations of the problems [6] and [7] in this special case, from the previous result we obtain the homogenization theorem for symmetric elliptic operators in divergence form, which asserts that for every $g \in L^2(\Omega)$ the solutions $u_\varepsilon$ of the Dirichlet problems

$$-\sum_{i,j=1}^{n} D_i\Big(a_{ij}\Big(\frac{x}{\varepsilon}\Big)D_j u_\varepsilon(x)\Big) = g(x) \quad \text{on } \Omega$$
$$u_\varepsilon(x) = 0 \quad \text{on } \partial\Omega$$

converge in $L^2(\Omega)$ to the solution $u$ of the Dirichlet problem

$$-\sum_{i,j=1}^{n} a_{ij}^{\text{hom}}\, D_i D_j u(x) = g(x) \quad \text{on } \Omega$$
$$u(x) = 0 \quad \text{on } \partial\Omega$$

An extensive literature is devoted to precise estimates of the homogenized coefficients $a_{ij}^{\text{hom}}$, depending on various structure conditions on the periodic coefficients $a_{ij}(x)$. Some of these estimates are based on a clever use of the variational formula [5].

Explicit formulas for $a_{ij}^{\text{hom}}$ are known in the case of layered materials, which correspond to the case where $\boldsymbol{R}^n$ is periodically partitioned into parallel layers on which the coefficients $a_{ij}(x)$ take constant values.

Easy examples show that, even if the composite material is isotropic at a microscopic layer (i.e., $a_{ij}(x) = a(x)\delta_{ij}$ for some scalar function $a(x)$), the homogenized material can be anisotropic (i.e., $a_{ij}^{\text{hom}} \neq a\delta_{ij}$), due to the anisotropy of the periodic function $a(x)$, which describes the microscopic distribution of the different components of the composite material.

In the vector case $m > 1$, the convexity hypothesis on $\xi \mapsto f(x,\xi)$ is not satisfied by the most interesting functionals related to nonlinear elasticity. If $\xi \mapsto f(x,\xi)$ is not convex, one can still prove that $\mathcal{F}_{\varepsilon_k}$ Γ-converges to a functional $\mathcal{F}_{\text{hom}}: L^p(\Omega;\boldsymbol{R}^m) \to [0,+\infty]$ of the form [4], but this time $f_{\text{hom}}: \boldsymbol{R}^{m\times n} \to [0,+\infty)$ cannot be obtained by solving a problem in the unit cell. Instead, it is given by the asymptotic formula

$$f_{\text{hom}}(\xi) := \lim_{R\to\infty} \frac{1}{R^n} \min_{w \in W_0^{1,p}(Q_R;\boldsymbol{R}^m)} \int_{Q_R} f(x, \xi + \nabla w)\,dx$$

where $Q_R := (-R/2, R/2)^n$ is the open cube of side $R$ centered at 0. Similar formulas can be obtained for quasiperiodic integrands $f$ and for stochastic homogenization problems.

In the nonperiodic case one can prove that, if $g_\varepsilon: \boldsymbol{R}^n \times \boldsymbol{R}^{m\times n} \to [0,+\infty)$ are arbitrary Borel functions satisfying [3], with constants independent of $\varepsilon$, and $\mathcal{G}_\varepsilon : L^p(\Omega;\boldsymbol{R}^m) \to [0,+\infty]$ are defined by

$$\mathcal{G}_\varepsilon(u) := \begin{cases} \int_\Omega g_\varepsilon(x,\nabla u)\,dx & \text{if } u \in W_0^{1,p}(\Omega;\boldsymbol{R}^m) \\ +\infty & \text{otherwise} \end{cases}$$

then there exists a sequence $\varepsilon_k \to 0$ such that the functionals $\mathcal{G}_{\varepsilon_k}$ Γ-converge to a functional $\mathcal{G}$ of the form

$$\mathcal{G}(u) := \begin{cases} \int_\Omega g(x,\nabla u)\,dx & \text{if } u \in W_0^{1,p}(\Omega;\boldsymbol{R}^m) \\ +\infty & \text{otherwise} \end{cases}$$

with $g$ satisfying [3].

In this case, no easy formula provides the integrand $g(x,\xi)$ in terms of simple operations on the integrands $g_{\varepsilon_k}(x,\xi)$. The indirect connection between these integrands can be obtained by introducing the functions $M_\varepsilon(x,\xi,\rho)$ defined, for $x \in \Omega$, $\xi \in \boldsymbol{R}^{m\times n}$, and $0 < \rho < \operatorname{dist}(x,\partial\Omega)$, by

$$M_\varepsilon(x,\xi,\rho) := \min_{w \in W_0^{1,p}(B(x,\rho))} \int_{B(x,\rho)} g_\varepsilon(y, \xi + \nabla w)\,dy$$

where $B(x,\rho)$ is the open ball with center $x$ and radius $\rho$. These functions describe the local behavior of the integrands $g_\varepsilon$ in some special minimum problems. The sequence $\mathcal{G}_{\varepsilon_k}$ Γ-converges to $\mathcal{G}$ if and only if

$$g(x,\xi) = \liminf_{\rho\to 0}\,\liminf_{k\to\infty} \frac{M_{\varepsilon_k}(x,\xi,\rho)}{|B(x,\rho)|} = \limsup_{\rho\to 0}\,\limsup_{k\to\infty} \frac{M_{\varepsilon_k}(x,\xi,\rho)}{|B(x,\rho)|}$$

for almost every $x \in \Omega$ and every $\xi \in \boldsymbol{R}^{m\times n}$.

Similar results have also been proved for integral functionals of the form

$$\mathcal{G}_\varepsilon(u) := \begin{cases} \int_\Omega g_\varepsilon(x,u,\nabla u)\,dx & \text{if } u \in W_0^{1,p}(\Omega;\boldsymbol{R}^m) \\ +\infty & \text{otherwise} \end{cases}$$

under suitable structure conditions for the integrands $g_\varepsilon$.

## Perforated Domains

In some homogenization problems, the integrand is fixed, but the domain depends on a small parameter $\varepsilon$ and its boundary becomes more and more fragmented as $\varepsilon \to 0$. A typical example is given by periodically perforated domains with small holes. Given a bounded open set $\Omega \subset \boldsymbol{R}^n$ and a compact set $K \subset Q$, both with smooth boundaries, for every $\varepsilon > 0$ we consider the perforated sets

$$\Omega_\varepsilon := \Omega \setminus \bigcup_{z \in Z_\Omega^\varepsilon} (\varepsilon z + \varepsilon K) \qquad [9]$$

where $Z_\Omega^\varepsilon$ is the set of vectors $z \in \boldsymbol{R}^n$ with integer coordinates such that $\varepsilon z + \varepsilon Q \subset \Omega$.

Given $g \in L^2(\Omega)$, let $\mathcal{F}_\varepsilon : L^2(\Omega) \to [0, +\infty]$ be the functionals defined by

$$\mathcal{F}_\varepsilon(u) := \begin{cases} \int_{\Omega_\varepsilon} \left[\frac{1}{2}|\nabla u|^2 - gu\right] \mathrm{d}x & \text{if } u \in W_0^{1,2}(\Omega) \\ +\infty & \text{otherwise} \end{cases} \qquad [10]$$

Minimizing [10] is equivalent to solving the mixed problems

$$\begin{aligned} -\Delta u_\varepsilon &= g && \text{on } \Omega_\varepsilon \\ u_\varepsilon &= 0 && \text{on } \partial\Omega \\ \frac{\partial u_\varepsilon}{\partial \nu} &= 0 && \text{on } \partial\Omega_\varepsilon \setminus \partial\Omega \end{aligned} \qquad [11]$$

The homogenization formula [5] is still valid, with minor modifications. It leads to a matrix of coefficients $a_{ij}^{\text{hom}}$ such that

$$\sum_{i,j=1}^n a_{ij}^{\text{hom}} \xi_j \xi_i := \min_{w \in W_{\text{per}}^{1,2}(Q)} \int_{Q\setminus K} |\xi + \nabla w|^2 \, \mathrm{d}x$$

for every $\xi \in \boldsymbol{R}^n$. For every sequence $\varepsilon_k \to 0$ the Γ-limit of the functionals $\mathcal{F}_{\varepsilon_k}$ is the functional $\mathcal{F} : L^2(\Omega) \to [0, +\infty]$ defined by

$$\mathcal{F}(u) := \begin{cases} \int_\Omega \left[\frac{1}{2}\sum_{i,j=1}^n a_{ij}^{\text{hom}} \, D_j u \, D_i u - mgu\right] \mathrm{d}x \\ \qquad \text{if } u \in W_0^{1,2}(\Omega) \\ +\infty \qquad \text{otherwise} \end{cases}$$

where $m := |Q \setminus K|$ is the volume fraction of the sets $\Omega_\varepsilon$.

Since a slight modification of the functionals $\mathcal{F}_\varepsilon$ satisfies an equicoerciveness condition, it follows from the basic properties of Γ-convergence that the solutions $u_\varepsilon$ of the mixed problems [11] in the perforated domains [9], extended to the holes so that $u_\varepsilon$ are harmonic on $\Omega \setminus \bar{\Omega}_\varepsilon$ and $u_\varepsilon \in W_0^{1,2}(\Omega)$, converge in $L^2(\Omega)$ to the solution $u$ of the Dirichlet problem

$$\begin{aligned} -\sum_{i,j=1}^n a_{ij}^{\text{hom}} D_i D_j u &= mg && \text{on } \Omega \\ u &= 0 && \text{on } \partial\Omega \end{aligned}$$

Therefore, the asymptotic effect of the small holes with Neumann boundary condition is a change in the coefficients of the elliptic equation.

In the case of Dirichlet boundary conditions, it is interesting to consider perforated domains with holes of a different size, namely

$$\Omega_\varepsilon := \Omega \setminus \bigcup_{z \in Z_\Omega^\varepsilon} (\varepsilon z + \varepsilon^{n/(n-2)} K) \qquad [12]$$

with $\varepsilon^{n/(n-2)}$ replaced by $\exp(-1/\varepsilon^2)$ if $n = 2$, while the case $n = 1$ gives only trivial results.

Given $g \in L^2(\Omega)$, let $\mathcal{G}_\varepsilon : L^2(\Omega) \to [0, +\infty]$ be the functionals defined by

$$\mathcal{G}_\varepsilon(u) := \begin{cases} \int_{\Omega_\varepsilon} \left[\frac{1}{2}|\nabla u|^2 - gu\right] \mathrm{d}x & \text{if } u \in W_0^{1,2}(\Omega_\varepsilon) \\ +\infty & \text{otherwise} \end{cases} \qquad [13]$$

Minimizing [13] is equivalent to solving the Dirichlet problems

$$\begin{cases} -\Delta u_\varepsilon = g & \text{on } \Omega_\varepsilon \\ u_\varepsilon = 0 & \text{on } \partial\Omega_\varepsilon \end{cases} \qquad [14]$$

For every sequence $\varepsilon_k \to 0$ the Γ-limit of the functionals $\mathcal{G}_{\varepsilon_k}$ is the functional $\mathcal{G} : L^2(\Omega) \to [0, +\infty]$ defined by

$$\mathcal{G}(u) := \begin{cases} \int_\Omega \left[\frac{1}{2}|\nabla u|^2 + \frac{c}{2}u^2 - gu\right] \mathrm{d}x & \text{if } u \in W_0^{1,2}(\Omega) \\ +\infty & \text{otherwise} \end{cases}$$

where, for $n \geq 3$,

$$c := \operatorname{cap}(K) := \inf_{\substack{w \in C_c^\infty(\boldsymbol{R}^n) \\ w = 1 \text{ on } K}} \int_{\boldsymbol{R}^n} |\nabla w|^2 \mathrm{d}x$$

Since a slight modification of the functionals $\mathcal{G}_\varepsilon$ satisfies an equicoerciveness condition, it follows from the basic properties of Γ-convergence that the solutions $u_\varepsilon$ of the Dirichlet problems [14] in the perforated domains [12], extended as zero on $\Omega \setminus \Omega_\varepsilon$, converge in $L^2(\Omega)$ to the solution $u$ if the Dirichlet problem

$$\begin{aligned} -\Delta u + cu &= g && \text{on } \Omega \\ u &= 0 && \text{on } \partial\Omega \end{aligned} \qquad [15]$$

In the electrostatic interpretation of these problems, the boundary $\partial\Omega_\varepsilon$ is a conductor kept at potential

zero. The extra term $cu$ in [15] is due to the electric charges induced on $\partial\Omega_\varepsilon$ by the charge distribution $g$.

These results on Dirichlet and Neumann boundary conditions have been extended to more general functionals and also to a wide class of nonperiodic distributions of small holes.

## Dimension Reduction Problems

In the study of thin elastic structures, like plates, membranes, rods, and strings, it is customary to approximate the mechanical behavior of a thin three-dimensional body by an effective theory for two- or one-dimensional elastic bodies. Γ-convergence provides a useful tool for a rigorous deduction of the lower-dimensional theory.

Let us focus on the derivation of plate theory from three-dimensional finite elasticity. The reference configuration of the thin three-dimensional elastic body is a cylinder of the form

$$\Omega_\varepsilon := S \times \left(-\frac{\varepsilon}{2}, \frac{\varepsilon}{2}\right)$$

where $\varepsilon > 0$ and $S$ is a bounded open subset of $\boldsymbol{R}^2$ with smooth boundary. We assume that the body is hyperelastic, with stored elastic energy

$$\int_{\Omega_\varepsilon} W(\nabla u)\, \mathrm{d}x$$

where $u : \Omega_\varepsilon \to \boldsymbol{R}^3$ is the deformation. The energy density $W : \boldsymbol{R}^{3\times 3} \to [0, +\infty]$, depending on the material, is continuous and frame indifferent; that is, $W(QF) = W(F)$ for every rotation $Q$ and every $F \in \boldsymbol{R}^{3\times 3}$, where $QF$ denotes the usual product of $3\times 3$ matrices. We assume that $W$ vanishes on the set SO(3) of rotations, is of class $C^2$ in a neighborhood SO(3), and satisfies the inequality

$$W(F) \geq \alpha \operatorname{dist}^2(F, \mathrm{SO}(3)) \quad \text{for every } F \in \boldsymbol{R}^{3\times 3} \qquad [16]$$

with a constant $\alpha > 0$.

Plate theory is obtained in the limit as $\varepsilon \to 0$ when the densities of the volume forces applied to the body have the form $\varepsilon^2 f(x_1, x_2)$, with $f \in L^2(S; \boldsymbol{R}^3)$. We assume that $f$ is balanced; that is,

$$\int_{\Omega_\varepsilon} f\, \mathrm{d}x = 0, \qquad \int_{\Omega_\varepsilon} x \wedge f\, \mathrm{d}x = 0$$

Stable equilibria are then obtained by minimizing the functionals

$$\int_{\Omega_\varepsilon} \left[W(\nabla u) - \varepsilon^2 f \cdot u\right] \mathrm{d}x \qquad [17]$$

on $W^{1,2}(\Omega_\varepsilon; \boldsymbol{R}^3)$.

To study the behavior of [17] as $\varepsilon \to 0$, it is convenient to change variables, so that the scaled deformations $v(x_1, x_2, x_3) := u(x_1, x_2, \varepsilon x_3)$ are defined on the same domain

$$\Omega := S \times \left(-\frac{1}{2}, \frac{1}{2}\right)$$

The scaled energy density $W_\varepsilon : \boldsymbol{R}^{3\times 3} \to [0, +\infty]$ is then defined as

$$W_\varepsilon(F_1|F_2|F_3) := W\left(F_1|F_2|\frac{1}{\varepsilon}F_3\right)$$

where $(F_1|F_2|F_3)$ denotes the $3\times 3$ matrix with columns $F_1, F_2$, and $F_3$. This implies that

$$\int_{\Omega_\varepsilon} \left[W(\nabla u) - \varepsilon^2 f \cdot u\right] \mathrm{d}x = \varepsilon \int_{\Omega} \left[W_\varepsilon(\nabla v) - \varepsilon^2 f \cdot v\right] \mathrm{d}x$$

The asymptotic behavior of the minimizers of these functionals can be obtained from the knowledge of the Γ-limit of the functionals $\mathcal{F}_\varepsilon : L^2(\Omega; \boldsymbol{R}^3) \to [0, +\infty]$ defined by

$$\mathcal{F}_\varepsilon(v) := \begin{cases} \dfrac{1}{\varepsilon^2} \displaystyle\int_\Omega W_\varepsilon(\nabla v)\, \mathrm{d}x & \text{if } v \in W^{1,2}(\Omega; \boldsymbol{R}^3) \\ +\infty & \text{otherwise} \end{cases}$$

Let us fix a sequence $\varepsilon_k \to 0$. The Γ-limit of $\mathcal{F}_{\varepsilon_k}$ turns out to be finite on the set $\Sigma(S; \boldsymbol{R}^3)$ of all isometric embeddings of $S$ into $\boldsymbol{R}^3$ of class $W^{2,2}$; that is, $v \in \Sigma(S; \boldsymbol{R}^3)$ if and only if $v \in W^{2,2}(S; \boldsymbol{R}^3)$ and $(\nabla v)^T \nabla v = I$ a.e. on $S$. The elements of $\Sigma(S; \boldsymbol{R}^3)$ will be often regarded as maps from $\Omega$ into $\boldsymbol{R}^3$, independent of $x_3$.

To describe the Γ-limit, we introduce the quadratic form $Q_3$ defined on $\boldsymbol{R}^{3\times 3}$ by

$$Q_3(F) := \tfrac{1}{2} D^2 W(I)[F, F]$$

which is the density of the linearized energy for the three-dimensional problem, and the quadratic form $Q_2$ defined on the space of symmetric $2 \times 2$ matrices by

$$Q_2\begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} := \min_{(b_1, b_2, b_3) \in \boldsymbol{R}^3} Q_3 \begin{pmatrix} a_{11} & a_{12} & b_1 \\ a_{12} & a_{22} & b_2 \\ b_1 & b_2 & b_3 \end{pmatrix}$$

The Γ-limit of $\mathcal{F}_{\varepsilon_k}$ is the functional $\mathcal{F} : L^2(\Omega; \boldsymbol{R}^3) \to [0, +\infty]$ defined by

$$\mathcal{F}(v) := \begin{cases} \frac{1}{12} \int_\Omega Q_2(A)\, \mathrm{d}x & \text{if } v \in \Sigma(S; \boldsymbol{R}^3) \\ +\infty & \text{otherwise} \end{cases}$$

where $A(x_1, x_2)$ denotes the second fundamental form of $v$; that is,

$$A_{ij} := -D_i D_j v \cdot \nu \quad [18]$$

with normal vector $\nu := D_1 v \wedge D_2 v$.

The equicoerciveness of the functionals $\mathcal{F}_\varepsilon$ in $L^2(\Omega; \mathbf{R}^3)$ is not trivial for this problem: it follows from [16] through a very deep geometric rigidity estimate which generalizes Korn's inequality (see Friesecke *et al.* (2002)). The basic properties of Γ-convergence imply that

$$\min_{u \in W^{1,2}(\Omega_\varepsilon; \mathbf{R}^3)} \int_{\Omega_\varepsilon} \left[W(\nabla u) - \varepsilon^2 f \cdot u\right] dx = \varepsilon^3 \min_{v \in \Sigma(S; \mathbf{R}^3)} \int_S \left[\frac{1}{12} Q_2(A) - f \cdot v\right] dx' + o(\varepsilon^3)$$

with $x' := (x_1, x_2)$ and $A$ defined by [18].

For every $\varepsilon > 0$ let $u_\varepsilon$ be a minimizer of [17] and let $v_\varepsilon(x_1, x_2, x_3) := u_\varepsilon(x_1, x_2, \varepsilon x_3)$. Then the basic properties of Γ-convergence imply that there exists a sequence $\varepsilon_k \to 0$ such that $v_{\varepsilon_k}(x_1, x_2, x_3)$ converges in $L^2(\Omega; \mathbf{R}^3)$ to a solution $v(x_1, x_2)$ of the minimum problem

$$\min_{v \in \Sigma(S; \mathbf{R}^3)} \int_S \left[\frac{1}{12} Q_2(A) - f \cdot v\right] dx' \quad [19]$$

These results provide a sound mathematical justification of the reduced two-dimensional theory of plates based on the minimum problem [19].

Similar results have been proved for shells, membranes, rods, and strings.

## Phase Transition Problems

The Cahn–Hilliard gradient theory of phase transitions deals with a fluid with mass $m$, under isothermal conditions, confined in a bounded open subset $\Omega$ of $\mathbf{R}^n$ with smooth boundary, whose Gibbs free energy, per unit volume, is a prescribed function $W$ of the density distribution $u$. Given a small parameter $\varepsilon > 0$, the energy functional $\mathcal{F}_\varepsilon : L^1(\Omega) \to [0, +\infty]$ has the form

$$\mathcal{F}_\varepsilon(u) := \begin{cases} \int_\Omega \left[W(u) + \varepsilon^2 |\nabla u|^2\right] dx & \text{if } u \in A(m) \\ +\infty & \text{otherwise} \end{cases} \quad [20]$$

where $A(m)$ is the set of all functions $u \in W^{1,2}(\Omega)$ with $\int_\Omega u = m$.

We assume that $W : \mathbf{R} \to [0, +\infty)$ is continuous and that there exist $\alpha, \beta \in \mathbf{R}$, with $\alpha|\Omega| < m < \beta|\Omega|$, such that $W(t) = 0$ if and only $t = \alpha$ or $t = \beta$. Moreover, we assume that $W(t) \to +\infty$ as $t \to \pm\infty$. In the minimization of $\mathcal{F}_\varepsilon$, the Gibbs free energy $W(u)$ favors the functions whose values are close to $\alpha$ and $\beta$, which represent the pure phases, while the gradient term penalizes the transitions between different phases.

It is easy to see that for every sequence $\varepsilon_k \to 0$ the sequence $\mathcal{F}_{\varepsilon_k}$ Γ-converges to the functional $\mathcal{F} : L^1(\Omega) \to [0, +\infty]$ defined by

$$\mathcal{F}(u) := \begin{cases} \int_\Omega W(u)\, dx & \text{if } \int_\Omega u = m \\ +\infty & \text{otherwise} \end{cases}$$

The set $M(\alpha, \beta, m)$ of minimum points of $\mathcal{F}$ is composed of all measurable functions $u$ on $\Omega$ which take only the values $\alpha$ and $\beta$ (on $E_\alpha$ and $E_\beta$, respectively), and satisfy the mass constraint $\alpha|E_\alpha| + \beta|E_\beta| = m$, which is equivalent to

$$|E_\alpha| = \frac{\beta|\Omega| - m}{\beta - \alpha} \quad [21]$$

From the basic properties of Γ-convergence, we deduce that

$$\min_{u \in A(m)} \int_\Omega \left[W(u) + \varepsilon^2 |\nabla u|^2\right] dx \to 0 \quad [22]$$

and that there exists a sequence $\varepsilon_k \to 0$ such that the minimizers $u_{\varepsilon_k}$ of $\mathcal{F}_{\varepsilon_k}$ converge in $L^1(\Omega)$ to a function $u$ which takes only the values $\alpha$ and $\beta$ and satisfies [21].

This result can be improved by considering the rescaled functionals

$$\mathcal{G}_\varepsilon(u) := \frac{1}{\varepsilon} \mathcal{F}_\varepsilon(u) \quad [23]$$

where $\mathcal{F}_\varepsilon$ is defined by [20]. Then for every sequence $\varepsilon_k \to 0$ the sequence $\mathcal{G}_{\varepsilon_k}$ Γ-converges to the functional $\mathcal{G} : L^1(\Omega) \to [0, +\infty]$ defined by

$$\mathcal{G}(u) := \begin{cases} 2cP(E_\alpha, \Omega) & \text{if } u \in M(\alpha, \beta, m) \\ +\infty & \text{otherwise} \end{cases}$$

where

$$c := \int_\alpha^\beta \sqrt{W(t)}\, dt$$

and

$$P(E, \Omega) := \sup\left\{\int_E \operatorname{div} \varphi\, dx : \varphi \in C_c^1(\Omega; \mathbf{R}^n), |\varphi| \le 1\right\}$$

is the Caccioppoli–De Giorgi perimeter of $E$ in $\Omega$, which coincides with the $(n-1)$-dimensional measure of $\Omega \cap \partial E$ when $E$ is smooth enough.

Note that the effective domain $A(m)$ of the functionals $\mathcal{G}_\varepsilon$ is disjoint from the effective domain of the limit functional $\mathcal{G}$, which is the set of all functions $u \in M(\alpha, \beta, m)$ with $P(E_\alpha, \Omega) < +\infty$.

As the functionals [20] and [23] have the same minimizers, we deduce that there exists a sequence $\varepsilon_k \to 0$ such that the minimizers $u_{\varepsilon_k}$ of $\mathcal{F}_{\varepsilon_k}$ converge in $L^1(\Omega)$ to a function $u$ which takes only the values $\alpha$ and $\beta$, satisfies [21], and fulfills the minimal interface criterion

$$P(E_\alpha, \Omega) \le P(E, \Omega)$$

for every measurable set $E \subset \Omega$ with $|E| = |E_\alpha|$. Moreover, [22] can be improved, and we obtain

$$\min_{u \in W^{1,2}(\Omega)} \mathcal{F}_\varepsilon(u) = \varepsilon 2cP(E_\alpha, \Omega) + o(\varepsilon)$$

Similar results have been proved when the term $|\nabla u|^2$ in [20] is replaced by a general quadratic form like [8], which leads to an anisotropic notion of perimeter.

## Free-Discontinuity Problems

Free-discontinuity problems are minimum problems for functionals composed of two terms of different nature: a bulk energy, typically given by a volume integral depending on the gradient of an unknown function $u$; and a surface energy, given by an integral on the unknown discontinuity surface of $u$. These problems arise in many different fields of science and technology, such as liquid crystals, fracture mechanics, and computer vision.

The prototype of free-discontinuity problems is the minimum problem proposed by David Mumford and Jayant Shah:

$$\min_{(u,K)\in\mathcal{A}} \left\{ \int_{\Omega\setminus K} |\nabla u|^2 \mathrm{d}x + \mathcal{H}^{n-1}(K \cap \Omega) + \int_{\Omega\setminus K} |u-g|^2 \mathrm{d}x \right\} \qquad [24]$$

where $\Omega$ is a bounded open subset of $\boldsymbol{R}^n$, $\mathcal{H}^{n-1}$ denotes the $(n-1)$-dimensional Hausdorff measure, $g \in L^\infty(\Omega)$, and $\mathcal{A}$ is the set of all pairs $(u, K)$ with $K$ compact, $K \subset \boldsymbol{R}^n$, and $u \in C^1(\Omega \setminus K)$.

In the applications to image segmentation problems the dimension $n$ is 2 and the function $g$ represents the grey level of an image. Given a solution $(u, K)$ of the minimum problem [24], the set $K$ is interpreted as the set of the relevant boundaries of the objects in the image, while $u$ provides a smoothed version of the image. The first term in [24] has a regularizing effect, the purpose of the second term is to avoid over-segmentation, while the last term, called "fidelity term," forces $u$ to be close to $g$. Of course, in the applications these terms are multiplied by different coefficients, whose relative values are very important for image segmentation problems, since they determine the strength of the effect of each term. However, the mathematical analysis of the problem can be easily reduced to the case where all coefficients are equal to 1.

To solve [24], it is convenient to introduce a weak formulation of the problem based on the space $\mathrm{GSBV}(\Omega)$ of generalized special functions with bounded variation (see Ambrosio *et al.* (2000)). Without entering into details, here it is enough to say that every $u \in \mathrm{GSBV}(\Omega)$ has, at almost every point, an approximate gradient $\nabla u$ in the sense of geometric measure theory. This is a measurable map from $\Omega$ into $\boldsymbol{R}^n$ which coincides with the usual gradient in the sense of distributions on every open subset $U$ of $\Omega$ such that $u \in W^{1,1}(U)$.

The functional $\mathcal{F} : L^1(\Omega) \to [0, +\infty]$ used for the weak formulation of [24] is defined by

$$\mathcal{F}(u) := \begin{cases} \int_\Omega |\nabla u|^2 \mathrm{d}x + \mathcal{H}^{n-1}(J_u) & \text{if } u \in GSBV(\Omega) \\ +\infty & \text{otherwise} \end{cases} \qquad [25]$$

where $J_u$ is the jump set of $u$, defined in a measure-theoretical way as the set of points $x \in \Omega$ such that

$$\limsup_{\rho\to 0} \frac{1}{|B(x,\rho)|} \int_{B(x,\rho)} |u(y) - a| \, \mathrm{d}y > 0$$

for every $a \in \boldsymbol{R}$.

For every $g \in L^\infty(\Omega)$, the functional

$$\mathcal{F}(u) + \int_\Omega |u-g|^2 \mathrm{d}x$$

is lower semicontinuous and coercive on $L^1(\Omega)$; therefore, the minimum problem

$$\min_{u\in L^1(\Omega)} \left\{ \mathcal{F}(u) + \int_\Omega |u-g|^2 \mathrm{d}x \right\} \qquad [26]$$

has a solution. The connection with the Mumford–Shah problem is given by the following regularity result, proved by Ennio De Giorgi and his collaborators: if $u$ is a solution of [26] and $\overline{J_u}$ is its closure, then $\mathcal{H}^{n-1}(\Omega \cap (\overline{J_u} \setminus J_u)) = 0$, $u \in C^1(\Omega \setminus \overline{J_u})$, and $(u, \overline{J_u})$ is a solution of [24].

Since the numerical treatment of [24] and [26] is quite difficult, Γ-convergence has been used to approximate [26] by means of minimum problems for integral functionals, whose minimizers can be obtained by standard numerical techniques.

Let us consider the nonlocal functionals $\mathcal{F}_\varepsilon : L^1(\Omega) \to [0, +\infty]$ defined by

$$\mathcal{F}_\varepsilon(u) := \begin{cases} \dfrac{1}{\varepsilon} \displaystyle\int_\Omega f\Big(\varepsilon \, \mathrm{Av}(|\nabla u|^2, x, \varepsilon)\Big) \mathrm{d}x & \text{if } u \in W^{1,2}(\Omega) \\ +\infty & \text{otherwise} \end{cases}$$

where

$$\mathrm{Av}(|\nabla u|^2, x, \varepsilon) := \frac{1}{|B(x,\varepsilon)\cap\Omega|}\int_{B(x,\varepsilon)\cap\Omega} |\nabla u(y)|^2 \mathrm{d}y$$

and $f:[0,+\infty)\to[0,+\infty)$ is any increasing continuous function with $f(0)=0, f'(0)=1$, and $f(t)\to 1/2$ as $t\to+\infty$. Then for every sequence $\varepsilon_k\to 0$ the sequence $\mathcal{F}_{\varepsilon_k}$ Γ-converges to $\mathcal{F}$.

Given $g\in L^\infty(\Omega)$, for every $\varepsilon>0$ let $u_\varepsilon$ be a solution of the minimum problem

$$\min_{u\in W^{1,2}(\Omega)}\left\{\frac{1}{\varepsilon}\int_\Omega f\Big(\varepsilon\,\mathrm{Av}(|\nabla u|^2, x,\varepsilon)\Big)\,\mathrm{d}x + \int_\Omega |u-g|^2\mathrm{d}x\right\}$$

From the basic properties of Γ-convergence it follows that there exists a sequence $\varepsilon_k\to 0$ such that $u_{\varepsilon_k}$ converges in $L^1(\Omega)$ to a solution $u$ of [26], so that $(u,\overline{J_u})$ is a solution of [24].

Other approximations by nonlocal functionals use finite differences instead of averages of gradients.

A different approximation can be obtained by using the local functionals $\mathcal{G}_\varepsilon:(L^1(\Omega))^2\to[0,+\infty]$ defined by

$$\mathcal{G}_\varepsilon(u,v) := \begin{cases} \displaystyle\int_\Omega\left[g_{\eta_\varepsilon}(v)|\nabla u|^2 + \frac{\varepsilon}{2}|\nabla v|^2 + \frac{1}{2\varepsilon}h(v)\right]\mathrm{d}x \\ \qquad \text{if } (u,v)\in(W^{1,2}(\Omega))^2 \\ +\infty \quad \text{otherwise}\end{cases}$$

where $g_{\eta_\varepsilon}(t) := \eta_\varepsilon + t^2, 0<\eta_\varepsilon << \varepsilon$, and $h(t) := (1-t)^2$ for $0\le t\le 1$, while $h(t) := +\infty$ otherwise. Let $\mathcal{G}:(L^1(\Omega))^2\to[0,+\infty]$ be the functional defined by

$$\mathcal{G}(u,v) := \begin{cases}\mathcal{F}(u) & \text{if } v=1 \text{ a.e. on } \Omega \\ +\infty & \text{otherwise}\end{cases}$$

where $\mathcal{F}$ is defined [25]. Then for every sequence $\varepsilon_k\to 0$ the sequence $\mathcal{G}_{\varepsilon_k}$ Γ-converges to $\mathcal{G}$.

Given $g\in L^\infty(\Omega)$, for every $\varepsilon>0$ let $(u_\varepsilon, v_\varepsilon)$ be a solution of the minimum problem

$$\min_{(u,v)\in(W^{1,2}(\Omega))^2}\int_\Omega\left[g_{\eta_\varepsilon}(v)|\nabla u|^2 + \frac{\varepsilon}{2}|\nabla v|^2 + \frac{1}{2\varepsilon}h(v) + |u-g|^2\right]\mathrm{d}x \qquad [27]$$

From the basic properties of Γ-convergence it follows that there exists a sequence $\varepsilon_k\to 0$ such that $u_{\varepsilon_k}$ converges in $L^1(\Omega)$ to a solution $u$ of [26], so that $(u,\overline{J_u})$ is a solution of [24].

The approximation of the solutions of [24] based on [27] has been used to construct numerical algorithms for image segmentation.

Free discontinuity problems similar to [24] appear in the mathematical treatment of Griffith's model in fracture mechanics. In this case, $u$ is a vector-valued function, which represents the deformation of an elastic body, the first term in [24] is replaced by a more general integral functional which represents the energy stored in the elastic region $\Omega\backslash K$, while the second term is interpreted as the energy dissipated to produce the crack $K$. An approximation based on minimum problems similar to [27] has been used to construct numerical algorithms to study the process of crack growth in brittle materials.

An important research line, connected with these problems, has been developed in the last years to derive the macroscopic theories of fracture mechanics from the microscopic theories of interatomic interactions. Using Γ-convergence, some theories expressed in the language of continuum mechanics can be obtained as limits of discrete variational models on lattices, as the distance between neighboring points tends to zero.

*See also:* Convex Analysis and Duality Methods; Elliptic Differential Equations: Linear Theory; Free Interfaces and Free Discontinuities: Variational Problems; Geometric Measure Theory; Image Processing: Mathematics; Variational Techniques for Ginzburg–Landau Energies; Variational Techniques for Microstructures.

## Further Reading

Allaire G (2002) *Shape Optimization by the Homogenization Method*. Berlin: Springer.

Ambrosio L, Fusco N, and Pallara D (2000) *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford: Oxford University Press.

Bakhvalov N and Panasenko G (1989) *Homogenisation: Averaging Processes in Periodic Media. Mathematical Problems in the Mechanics of Composite Materials* (translated from the Russian by D. Leĭtes). Dordrecht: Kluwer.

Bensoussan A, Lions JL, and Papanicolaou GC (1978) *Asymptotic Analysis for Periodic Structures*. Amsterdam: North-Holland.

Braides A (1998) Approximation of Free-Discontinuity Problems, *Lecture Notes in Mathematics*, vol. 1694. Berlin: Springer.

Braides A (2002) *Γ-Convergence for Beginners*. Oxford: Oxford University Press.

Braides A and Defranceschi A (1998) *Homogenization of Multiple Integrals*. Oxford: Oxford University Press.

Cherkaev A and Kohn RV (eds.) (1997) *Topics in the Mathematical Modelling of Composite Materials*. Boston: Birkhäuser.

Christensen RM (1979) *Mechanics of Composite Materials*. New York: Wiley.

Cioranescu D and Donato P (1999) *An Introduction to Homogenization*. New York: The Clarendon Press, Oxford University Press.

Dal Maso G (1993) *An Introduction to Γ-Convergence*. Boston: Birkhäuser.

Friesecke G, James RD, and Müller S (2002) A theorem on geometric rigidity and the derivation of nonlinear plate theory

from three-dimensional elasticity. *Communications on Pure and Applied Mathematics* 55: 1461–1506.
Jikov VV, Kozlov SM, and Oleĭnik OA (1994) *Homogenization of Differential Operators and Integral Functionals* (translated from the Russian by G. A. Yosifian). Berlin: Springer.
Milton GW (2002) *The Theory of Composites*. Cambridge: Cambridge University Press.
Oleĭnik OA, Shamaev AS, and Yosifian GA (1992) *Mathematical Problems in Elasticity and Homogenization*. Amsterdam: North-Holland.
Panasenko G (2005) *Multi-scale Modelling for Structures and Composites*. Dordrecht: Springer.
Sanchez-Palencia E (1980) Nonhomogeneous Media and Vibration Theory, *Lecture Notes in Physics*, vol. 127. Berlin: Springer.

# Gauge Theoretic Invariants of 4-Manifolds

**S Bauer**, Universität Bielefeld, Bielefeld, Germany



## Introduction

Poincaré duality is fundamental in the study of manifolds. In the case of an orientable closed manifold $X$, this duality appears as an isomorphism

$$\psi : H^k(X;\mathbb{Z}) \to H_{n-k}(X;\mathbb{Z})$$

between integral cohomology and homology. The map $\psi$ is defined by cap product with a chosen orientation class. This article focuses on dimension $n = 4$, where Poincaré duality induces a bilinear form $Q$ on $H_2(X;\mathbb{Z})$ by use of the Kronecker pairing

$$Q(\xi, \xi') = \langle \psi^{-1}(\xi), \xi' \rangle \in \mathbb{Z}$$

One of the outstanding achievements of modern topology, the classification of simply connected topological 4-manifolds by Freedman (1982), can be phrased in terms of the intersection pairing $Q$. Indeed, two simply connected differentiable 4-manifolds $X$ and $X'$ are orientation preservingly homeomorphic if and only if the associated pairings $Q$ and $Q'$ are equivalent. Freedman's classification scheme has been extended to also cover a wide range of fundamental groups, resulting in a fair understanding of topological 4-manifolds (Freedman and Quinn 1990).

When it comes to differentiable 4-manifolds, the situation changes drastically. On the one hand, there is an abundance of topological 4-manifolds which do not admit a differentiable structure at all. On the other hand, there also are topological 4-manifolds supporting infinitely many distinct differentiable structures. A classification of differentiable 4-manifolds up to differentiable equivalence seems out of reach of current technology, even in the most simple cases.

The discrepancy between topological and differentiable 4-manifolds was uncovered by gauge-theoretic methods, applying the concepts of instantons and of monopoles. In order to study these, one has to equip a 4-manifold both with a Riemannian metric and some additional structure: a Hermitian rank-2 bundle in the case of instantons and a spin$^c$-structure in the case of monopoles. Given such data, instantons and monopoles arise as solutions to partial differential equations the gauge equivalence classes of which form finite-dimensional moduli spaces. As it turns out, these moduli spaces encode significant information about the differentiable structures of the underlying 4-manifolds.

A decoding of such information contained in the instanton moduli and in the monopole moduli is achieved through Donaldson invariants and Seiberg–Witten invariants, respectively. This article outlines these theories from a mathematical point of view.

## Instantons and Donaldson Invariants

Let $X$ denote a closed, connected, oriented differentiable Riemannian 4-manifold. We will consider a principal bundle $P$ over $X$ with fiber a compact Lie group $G$ with Lie algebra $\mathfrak{g}$. Connections on $P$ form an infinite-dimensional affine space $\mathcal{A}(P) = A_0 + \Omega^1(X;\mathfrak{g}_P)$ modeled on the vector space of 1-forms with values in the adjoint bundle

$$\mathfrak{g}_P = P \times_{\mathrm{Ad}(G)} \mathfrak{g}$$

The curvature $F_A \in \Omega^2(X, \mathfrak{g}_P)$ of a connection $A$ is a $\mathfrak{g}_P$-valued 2-form satisfying the Bianchi identity $D_A F_A = 0$. The group $\mathcal{G}$ of principal bundle automorphisms of $P$ acts in a natural way on the space of connections with quotient space

$$\mathcal{B}(P) = \mathcal{A}(P)/\mathcal{G}$$

The Yang–Mills functional

$$\mathrm{YM} : \mathcal{A}(P) \to \mathbb{R}_{\geq 0}$$

associates to a connection $A$ the norm square

$$\|F_A\|^2 = -\int_X \mathrm{tr}(F_A \wedge *F_A)$$

of its curvature. Here $*$ denotes the Hodge star operator defined by the metric on $X$ and the orientation. The metric $-\mathrm{tr}: \mathfrak{g} \otimes \mathfrak{g} \to \mathbb{R}$ is $\mathrm{Ad}(G)$-invariant and hence YM is invariant under the

action of $\mathcal{G}$. In particular, the Yang–Mills functional descends to a function on the space $\mathcal{B}(P)$ of a gauge equivalence class of connections.

The Euler–Lagrange equations for the critical points of YM, called Yang–Mills equations, are of the form

$$D_A(*F_A) = 0$$

and can be derived easily from the formula

$$F_{A+a} = F_A + D_A(a) + [a \wedge a]$$

Satisfying the equations

$$D_A(*F_A) = 0 \quad \text{and} \quad D_A(F_A) = 0$$

a Yang–Mills connection is characterized by the fact that it is harmonic with respect to its own Laplacian.

The bundle $\wedge^2 T^*X$ of 2-forms on $X$ decomposes into $(\pm 1)$-eigenbundles of the Hodge operator. This orthogonal splitting leads to a decomposition of curvature forms

$$F_A = F_A^+ + F_A^-$$

into self-dual and anti-self-dual components. The differential form $-(1/4\pi^2)\mathrm{tr}(F_A \wedge F_A)$ represents a characteristic class of the principal bundle $P$. In particular, the integral

$$\begin{aligned}\kappa(P) &= -\int_X \mathrm{tr}(F_A \wedge F_A) \\ &= \|F_A^+\|^2 - \|F_A^-\|^2\end{aligned}$$

is independent of the connection A. The Yang–Mills functional therefore is bounded

$$\mathrm{YM}(P) \geq |\kappa(P)|$$

and attains this minimum at connections $A$ which satisfy the equation

$$*F_A = \pm F_A$$

Such connections are either self-dual, anti-self-dual or both, that is, flat, depending on whether $\kappa(P)$ is negative, positive, or zero. The moduli space of instantons on $P$ is the subset of minima of the Yang–Mills functional

$$M(P) = \mathrm{YM}^{-1}(|\kappa(P)|) \subset \mathcal{B}(P)$$

The moduli space thus consists of gauge equivalence classes of connections which are either self-dual or anti-self-dual. Donaldson theory indeed considers anti-self-dual connections on principal bundles with structure group $\mathrm{PU}(2) = \mathrm{SO}(3)$.

The Hodge $*$ operator induces a decomposition of the second cohomology

$$H^2(X) = H^2_+(X) \oplus H^2_-(X)$$

into $(\pm 1)$-eigenspaces of dimension $b^+$ and $b^-$. Unless specified differently, cohomology groups are meant with real coefficients. In order to simplify the exposition, we will assume $X$ to be simply connected. The Donaldson invariants then are defined if $b^+$ is odd and greater than 1.

A "homology orientation" consists of an orientation of $H^2_+(X)$ and an integral homology class $c \in H_2(X; \mathbb{Z})$. The Donaldson invariant $D_{X,c} = D_c$ is defined after fixing such a homology orientation. It is a linear function

$$D_c : A(X) \to \mathbb{R}$$

where $A(X)$ is the graded algebra

$$A(X) = \mathrm{Sym}_*(H_0(X) \oplus H_2(X))$$

in which $H_i(X)$ has degree $(1/2)(4 - i)$. The significance of $D_c$ is its functoriality

$$D_{X', f(c)}(f(\alpha)) = D_{X,c}(\alpha)$$

under diffeomorphisms $f : X \to X'$ which preserve both orientation and homology orientation. Switching the orientation of $H^2_+(X; \mathbb{R})$ reverses the sign of $D_c$. Similarly,

$$D_{c'} = (-1)^{((c-c')/2)^2} D_c$$

if $c - c' \in 2H_2(X, \mathbb{Z}) \subset H_2(X; \mathbb{Z})$.

The construction of this invariant makes use of the following facts:

1. An SO(3) principal bundle $P$ over $X$ is determined by its first Pontrjagin number $p_1(P)$ and its Stiefel–Whitney class $w_2(P) \in H^2(X; \mathbb{Z}/2)$. As $X$ is simply connected, this Stiefel–Whitney class admits integer lifts. Let $c$ be such a lift and let $c^2$ be shorthand for the intersection pairing $Q(c, c)$. A pair $(p_1, w_2)$ is realized by a principal bundle provided it satisfies the relation $p_1 \equiv c^2$ modulo 4.

2. If $b^+$ is nonvanishing, then for generic metrics on $X$, the moduli space $M(P)$ is a manifold of dimension

$$-2p_1(P) - 3(1 + b^+)$$

This follows from a transversality theorem whose main ingredient in the Sard–Smale theorem. The dimension is computed by use of the Atiyah–Singer index theorem: to an anti-self-dual connection $A$ on $P$ there is an associated elliptic complex

$$\begin{aligned}0 \to \Omega^0(X; \mathfrak{g}_P) &\xrightarrow{D_A} \Omega^1(X; \mathfrak{g}_P) \\ &\xrightarrow{D_A^+} \Omega^2_+(X; \mathfrak{g}_P) \to 0\end{aligned}$$

where $\Omega^i(X; \mathfrak{g}_P)$ denotes $\mathfrak{g}_P$-valued $i$-forms on $X$. This complex describes the tangential structure of

the moduli space at the equivalence class of $A$. The space $\Omega^1(X;\mathfrak{g}_P)$ is the tangent space of $\mathcal{A}(P)$ at $A$, $\Omega^0(X;\mathfrak{g}_P)$ is the tangent space of the group $\mathcal{G}$ at the identity, and $D_A$ is the differential of the orbit map. The differential operator $D_A^+$ is the linearization of the anti-self-duality map

$$a \longmapsto F_{A+a}^+ = D_A^+(a) + [a \wedge a]^+$$

3. The moduli space $M(P)$ can be oriented if it is a manifold. The orientation depends on an orientation of $H_+^2(X)$ and on a U(2)-principal bundle which has $P$ as its PU(2)-quotient bundle. It is determined by an integer lift of $w_2(P)$. The elliptic complex above then can be compared with a corresponding elliptic complex where the differentials are given by a complex Dirac operator. This leads to an almost-complex structure on the tangent space for each point in the moduli space and in particular to an orientation on the moduli space itself.

4. Over the product $M(P)\times X$ there is a universal PU(2)-bundle $\mathbb{P}$ with first Pontrjagin class $p_1(\mathbb{P})$. Taking slant product with the class $-(1/4)p_1(\mathbb{P})$ results in a homomorphism

$$\mu : H_i(X) \to H^{4-i}(M(P))$$

5. The moduli space $M(P)$ in general is noncompact. There is an Uhlenbeck compactification $\overline{M(P)}$ describing "ideal instantons." Such an ideal instanton consists of an element $(x_1,\ldots,x_n)\in \mathrm{Sym}_n(X)$ and an anti-self- dual connection $A'$ on the principal bundle $P'$ on $X$ with $w_2(P')=w_2(P)$ for which the equality

$$p_1(P') - p_1(P) = 4n$$

of Pontrjagin numbers holds. Uhlenbeck's compactness theorem describes what happens if a sequence of anti-self-dual connections has no convergent subsequence: after passing to a subsequence, the sequence converges to an anti-self-dual connection on the restriction of $P$ to $X\backslash\{x_1,\ldots,x_n\}$. This limit connection extends to a connection $A'$ on the principal bundle $P'$. The functions $|F_{A_n}|^2$ on $X$ converge to the measure

$$|F_{A'}|^2 + \sum_{i=1}^{n} 8\pi^2\delta_{x_i}$$

The compactification $\overline{M(P)}$ is a stratified space and not usually a manifold. If $w_2(P)\neq 0$, then the singular set of codimension at least 2 and thus the space $\overline{M(X)}$ carries a fundamental class. In the case $w_2(P)=0$, such a fundamental class in general can only be defined if $-p_1(P) > 4 + 3b^+$. In practice, this problem can be circumvented by blowing up $X$ and considering bundles with $w_2(P)\neq 0$ over the connected sum $X\#\overline{\mathbb{C}P}^2$. Note that the complex projective plane $\mathbb{C}P^2$ as a complex manifold carries a natural orientation. The notation $\overline{\mathbb{C}P}^2$ indicates a reversed orientation.

6. The classes $\mu(\alpha)\in H^2(M(P))$ for $\alpha \in H^2(X)$ extend over the compactification. The same holds for the class $\mu(x)$, where $x\in H_0(X;\mathbb{Z})$ is the generator corresponding to the orientation, as long as $w_2(P)\neq 0$. Otherwise, there are certain dimension restrictions. However, the same blow-up trick as mentioned above allows to handle the case $w_2(P)=0$ as well.

Now fix an element $c\in H_2(X;\mathbb{Z})$ and let

$$M_c = \bigsqcup_{d\geq 0} \overline{M(P_{c,d})}$$

denote the disjoint union of all moduli spaces of anti-self-dual connections on principal PU(2)-bundles $P_{c,d}$ whose second Stiefel–Whitney class is Poincaré-dual to $c$ modulo 2 and whose Pontrjagin number equals $-d-(3/2)(b^+ + 1)$.

Our assumption of $b^+$ being odd corresponds to the fact that the dimension $2d$ of the moduli space $M(c,d)$ is even and congruent to $-c^2 + (1/2)(1+b^+)$ modulo 4. Neglecting the difficulties in the case $w_2(P)=0$ mentioned above, we may use the cup product on $H^*(M_c)$ to extend $\mu$ to an algebra homomorphism

$$\mu : A(X) \to H^*(M_c)$$

The Donaldson invariant $D_c$ is nonzero only on elements $z$ of $A(X)$ whose total degree $d$ is congruent to $-c^2 + (1/2)(1+b^+)$ modulo 4. For such an element it is defined by

$$D_c(z) = \langle \mu(z), \overline{M(P_{c,d})}\rangle = \int_{\overline{M(P_{c,d})}} \mu(z)$$

The Donaldson series $\mathbb{D}_c$ is defined as a formal power series

$$\mathbb{D}_c(\alpha) = D_c(\exp(\hat{\alpha})) = \sum_{d=0}^{\infty} \frac{D_c(\hat{\alpha}^d)}{d!}$$

for $\alpha\in H_2(X)$ and $\hat{\alpha} = (1+(x/2))\alpha$.

## Computations and Structure Theorems

The first results about these invariants are due to S Donaldson. He proved both a vanishing and a nonvanishing theorem (Donaldson and Kronheimer 1990):

**Theorem 1** *If both $b^+(X)>0$ and $b^+(Y)>0$, then all Donaldson invariants vanish for the connected sum $X\#Y$.*

**Theorem 2** *If c represents a divisor on a complex algebraic surface X and $\alpha$ represents an ample divisor, then*

$$D_c(\alpha^r) \neq 0 \quad \text{for } r \gg 0$$

The second theorem is a consequence of the fact that in the case of an algebraic surface the instanton moduli can be described in algebraic geometric terms: the moduli space $\overline{M(P_{c,d})}$ associated to the metric induced from the Fubini–Study metric on $\mathbb{C}P^n$ by an embedding $X \hookrightarrow \mathbb{C}P^n$ carries the structure of a projective variety. This variety is reduced and of complex dimension $d$, as soon as $d$ is large enough. Furthermore, $\mu(d)$ is the first Chern class of an ample line bundle.

The translation of instanton moduli into algebraic geometry uses two steps: suppose the first Chern class of a U($r$)-principal bundle $P$ on a Kähler surface is also the first Chern class of a holomorphic line bundle. Then the absolute minima of the Yang–Mills functional are achieved by Hermite–Einstein connections. These are connections for which the Ricci curvature is a constant multiple of the identity. The second step, the translation from differential geometry into algebraic geometry, is called the Kobayashi–Hitchin correspondence, which again was proved by Donaldson.

The Donaldson invariants have been computed for a number of 4-manifolds. A simply connected 4-manifold is said to have simple type, if the relation

$$D_c(x^2 z) = 4D_c(z)$$

is satisfied by its Donaldson invariant for all $z \in A(X)$ and $c \in H_2(X;\mathbb{Z})$. It is known that this simple type condition holds for many 4-manifolds. Indeed, it is an open question whether there are 4-manifolds which are not of simple type. For manifolds of simple type the Donaldson series $\mathbb{D}_c$ completely determines the Donaldson invariant $D_c$. A main result is due to Kronheimer and Mrowka (1995):

**Theorem 3** *Let X be a simply connected 4-manifold of simple type. Then, there exist finitely many basic classes $\kappa_1, \ldots, \kappa_n \in H_2(X;\mathbb{Z})$ such that*

$$\mathbb{D}_c = \exp(Q/2) \sum_{i=1}^{n} (-1)^{(c^2-\kappa_i c)/2} a_i \exp(\kappa_i)$$

*as analytic functions on $H_2(X)$. The numbers $a_i$ are rational and each basic class $\kappa_i$ is characteristic, that is, it satisfies $\alpha^2 \equiv Q(\alpha, \kappa_i)$ modulo 2 for all $\alpha \in H_2(X;\mathbb{Z})$. The homology class $\kappa_i$ in this formula acts on an arbitrary homology class by intersection.*

The geometric significance of the basic classes is underlined by the following theorem (Kronheimer and Mrowka 1995):

**Theorem 4** *If $\alpha \in H_2(X:\mathbb{Z})$ is represented by an embedded surface of genus g with self-intersection $\alpha^2 \geq 2$, then for each basic class $\kappa$ the following adjunction inequality is satisfied:*

$$2g - 2 \geq \alpha^2 + |Q(\kappa, \alpha)|$$

There are many 4-manifolds for which the Donaldson series have been computed (Friedman and Morgan 1997). The basic classes for complete intersections, for example, are the canonical divisor and its negative. Another example is given by elliptic surfaces. Let $E(n;p,q)$ be a minimal elliptic surface, that is, a holomorphic surface admitting a holomorphic map to $\mathbb{C}P^1$ with generic fiber $f$ an elliptic curve. For any numbers $n, p$, and $q$ with $p < q$ coprime, there exists such a simply connected elliptic surface with Euler characteristic $12n$ and two multiple fibers of multiplicity $p$ and $q$, respectively. The Donaldson series of $E(n;p,q)$ for $c = 0$ then is given by

$$\mathbb{D} = \exp\left(\frac{Q}{2}\right) \frac{\sinh^n(f)}{\sinh(f/p)\sinh(f/q)}$$

Another important formula relates the Donaldson series $\mathbb{D}$ a manifold $X$ of simple type and the Donaldson series $\hat{\mathbb{D}}$ of the blow-up $X \# \overline{\mathbb{C}P^2}$:

$$\hat{\mathbb{D}}_c = \mathbb{D}_c \cdot \exp(-e^2/2)\cosh(e)$$
$$\hat{\mathbb{D}}_{c+e} = -\mathbb{D}_c \cdot \exp(-e^2/2)\sinh(e)$$

Here $e \in H_2(\overline{\mathbb{C}P^2};\mathbb{Z})$ denotes a generator. Indeed, a more general blow-up formula is known which relates the Donaldson invariants for $X$ and its blow-up even in case $X$ is not of simple type. This formula, due to Fintushel and Stern (1996), involves Weierstraß sigma-functions.

The instanton moduli space carries nontrivial information about 4-manifolds even in the case $b^+(X) \leq 1$. However, one has to deal with singularities in the moduli space. Let us first consider the case $b^+(X) = 0$. If the intersection form on $X$ is negative definite, the instanton moduli spaces in general are bound to have singularities. Indeed, Donaldson examined the case with the Pontrjagin number $p_1(P) = -4$ and $w_2(P) = 0$. In this case, the moduli space for a generic metric on $X$ will be an orientable smooth manifold except at isolated singular points. The singularities are cones over $\overline{\mathbb{C}P^2}$ and they correspond to reducible connections, that is, reductions of the structure group of $P$ to U(1). These reductions are in bijective correspondence to pairs $\pm\alpha \in H_2(X;\mathbb{Z})$ with $\alpha^2 = -1$. The Uhlenbeck compactification of the moduli space thus leads to an oriented cobordism between $X$ and the disjoint union $\sqcup_\alpha \overline{\mathbb{C}P^2}$ over all pairs $\pm\alpha$ in $H_2(X;\mathbb{Z})$ of square –1. As the signature of a

manifold is an invariant of oriented cobordism, there have to be $b^-$ many pairs $\pm\alpha$ of square (–1) in $H_2(X;\mathbb{Z})$ and, in particular, the intersection form $Q$ is represented by the negative of the identity matrix (Donaldson 1983):

**Theorem 5** *The intersection form on a differentiable manifold with negative-definite intersection form is diagonal.*

Indeed, from rank 8 on there are lots of definite unimodular forms which are not diagonal. By Freedman's (1982) classification, any unimodular form is realized as the intersection form of a simply connected topological manifold. This theorem shows that most of these manifolds do not support differentiable structures.

The case $b^+(X) = 1$ is also interesting. Here, the moduli space is a smooth manifold for a generic metric, giving rise to Donaldson invariants. However, over a smooth path of metrics, there is in general no smooth cobordism of moduli spaces. So the invariants depend on the chosen metric. The singularities in the cobordisms again correspond to classes in $H_2(X;\mathbb{Z})$ with negative square. An analysis of these singularities leads to wall-crossing formulas describing how different choices of the metric do affect Donaldson invariants. The case of $\mathbb{C}P^2$ is special, as there are no elements of negative square in $H_2(\mathbb{C}P^2;\mathbb{Z})$. The Donaldson invariants for $\mathbb{C}P^2$ as well as the wall-crossing formulas turn out to be closely related to modular forms (Göttsche 2000).

## Monopoles and Seiberg–Witten Invariants

A spin$^c$-structure on an oriented Riemannian 4-manifold is a Spin$^c$(4)-principal bundle $P$ projecting to the orthonormal tangent frame bundle $\overline{P}$ over $X$ through the group homomorphism Spin$^c(4) \to SO(4)$ with kernel U(1). The group $H^2(X;\mathbb{Z})$ acts freely and transitively on the set of all spin$^c$-structures. A spin$^c$-connection is a lift to $P$ of the Levi-Civita connection on $\overline{P}$. Fixing a background spin$^c$-connection $A_0$, the monopole map

$$\mu : (A, \phi) \longmapsto \left(\mathcal{D}_A\phi, F_A^+ - \phi\phi_\circ^*, d^*a\right)$$

is defined (Witten 1994) for spin$^c$-connections $A \in A_0 + \Omega^1(X; \mathrm{i}\mathbb{R})$ and positive spinors $\phi$. Here, $\mathcal{D}_A$ denotes the complex Dirac operator associated to $A$ and $d^*a$ for $a \in \Omega^1(X;\mathrm{i}\mathbb{R})$ is the adjoint of the de Rham differential on forms. The section $\phi\phi_\circ^*$ of the traceless endomorphism bundle of positive spinors is viewed as a self-dual 2-form on $X$.

In case the first Betti number vanishes, this map – after suitable Sobolev completion – becomes a map between Hilbert spaces $\mu : \mathcal{A} \to \mathcal{C}$ which is a compact deformation of a linear Fredholm map. The Weitzenböck formula can be used to show that preimages under $\mu$ of bounded sets in $\mathcal{C}$ are bounded in $\mathcal{A}$. Furthermore, $\mu$ is U(1)-equivarant, where U(1) acts by complex multiplication on spinors and trivially on forms. If $b_1(X) > 0$, the monopole map is a map between Hilbert space bundles over the torus $H^1(X)/H^1(X;\mathbb{Z})$. These properties of the monopole map allow for an interpretation in terms of stable homotopy (Bauer 2004):

**Theorem 6** *If the first Betti number of X vanishes, then $\mu$ defines an element*

$$[\mu] \in \pi_i^{U(1)}(S^0)$$

*in an equivariant stable homotopy group of spheres. The index $i = \mathrm{ind}\,\mathcal{D}_A - H_+^2(X)$ as an element of the real representation ring* RO(U(1)) *is determind by the analytic index of the linearization of $\mu$.*

In the case $b^+(X) > 1$, these equivariant stable homotopy groups can be identified with nonequivariant stable cohomotopy groups $\pi_{st}^{b^+-1}(\mathbb{C}P^{d-1})$. Here, $d$ denotes the index of the complex Dirac operator ind $\mathcal{D}_A$. Fixing an orientation of $H_+^2(X)$ results in a Hurewicz homomorphism

$$h : \pi_{st}^{b^+-1}(\mathbb{C}P^{d-1}) \to H^{b^+-1}(\mathbb{C}P^{d-1};\mathbb{Z})$$

If $b^+(X)$ is odd, the image

$$h([\mu]) = SW(X)t^{(b^+-1)/2}$$

is an integer multiple of a power of the generator $t \in H^2(\mathbb{C}P^{d-1};\mathbb{Z})$. This integer SW($X$) is known as the Seiberg–Witten invariant (Witten 1994).

This invariant alternatively can be defined by considering the moduli space $M(a) = \mu^{-1}(a)$. Assuming $b^+ > 0$, this is a smooth oriented manifold with a free U(1)-action for generic $a \in \Omega^1(X; i\mathbb{R})$. The Seiberg–Witten invariant is the characteristic number obtainable by these data. In general, the stable homotopy invariant $[\mu]$ encodes global information about the monopole map, which cannot be recovered by only considering the moduli space. In case the spin$^c$-structure is associated to an almost-complex structure, however, there is a fortunate coincidence: the Hurewicz homomorphism in this case is an isomorphism. So for almost-complex spin$^c$-structures, the invariants $[\mu]$ and SW carry the same information.

The Seiberg–Witten invariants turn out to be directly computable for Kähler manifolds and to some degree also for symplectic manifolds (Taubes 1994). Indeed,

the following theorem follows from arguments of Witten and of Taubes:

**Theorem 7** *Let X be a 4-manifold with* $b^+ > 1$ *and* $b_1 = 0$ *which can be equipped with a Kähler or a symplectic structure. If* $[\mu]$ *is nonvanishing for a spin*$^c$*-structure on X, then the spin*$^c$*-structure is associated to an almost-complex structure. For the canonical spin*$^c$*-structure on X the Seiberg–Witten invariant is* $\pm 1$.

Seiberg–Witten invariants and Donaldson invariants are closely related: Witten gave physical arguments that an equality of the form

$$\mathbb{D} = 2^k \exp(Q/2) \cdot \sum_\alpha \mathrm{SW}(\alpha) \exp(\alpha)$$

should hold for the Donaldson series for $c = 0$ of a simply connected manifold of simple type. Here, $\alpha \in H^2(X;\mathbb{Z})$ denotes the first Chern class of the complex determinant line bundle. This first Chern class characterizes spin$^c$-structures in the simply connected case. The number $k$ is related to the signature $\sigma$ and the Euler characteristic $\chi$ of the manifold $X$ by the formula

$$4k = 11\sigma + 7\chi + 2$$

A mathematical proof of this formula is known in special cases (Feehan and Leness 2003).

As is the case for Donaldson invariants, the Seiberg–Witten invariants vanish for connected sums $X\#Y$ if both $b^+(X) > 0$ and $b^+(Y) > 0$ holds. This is not the case for the stable homotopy refinement as follows from the following theorem (Bauer 2004).

**Theorem 8** *For a connected sum* $X\#Y$ *of 4-manifolds the stable equivariant homotopy invariants are related by smash product*

$$[\mu_{X\#Y}] = [\mu_X] \wedge [\mu_Y]$$

As an example application, consider connected sums of elliptic surfaces of the form $E(2n; p, q)$. Now suppose $X$ and $X'$ are each connected sums of at most four copies of such elliptic surfaces. Then $X$ and $X'$ are diffeomorphic if and only if the summands were already diffeomorphic. This contrasts to the fact that the connected sum $E(2n; p, q)\#\mathbb{C}P^2$ is diffeomorphic to a connected sum of $4n - 1$ copies of $\mathbb{C}P^2$ and $20n - 1$ copies of $\overline{\mathbb{C}P}^2$, independently of $p$ and $q$.

As a final application, we consider the case of spin manifolds. If the manifold $X$ is spin, then the intersection form $Q$ is even, that is, $Q(\alpha, \alpha) = 0$ mod 2 for $\alpha \in H_2(X, \mathbb{Z})$. According to Rochlin's theorem, the signature of a spin 4-manifold is divisible by 16. The monopole map $\mu$ for the spin structure admits additional symmetry. It is Pin(2)-equivariant. The nonabelian group Pin(2) appears as the normalizer of the maximal torus SU(2). Methods from equivariant $K$-theory lead to Furuta's (2001) theorem:

**Theorem 9** *Let X be a spin 4-manifold. Then*

$$\chi(X) > \tfrac{5}{4}|\sigma(X)|$$

## Manifolds with Boundary

Both Donaldson invariants and Seiberg–Witten invariants to some extent satisfy formal properties which fit into a general conceptual framework known as "topological quantum field theories (TQFTs)." Such a TQFT in $3+1$ dimensions is a functor on the cobordism category of oriented 3-manifolds to the category of, say, vector spaces over a ground field: it assigns to an oriented 3-manifold $Y$ a vector space $h(Y)$. To a disjoint union it assigns

$$h(Y_1 \sqcup Y_2) = h(Y_1) \otimes h(Y_2)$$

Reversing orientation corresponds to dualizing

$$h(\overline{Y}) = h(Y)^*$$

Viewing a four-dimensional manifold $X$ with boundary $\partial X = \overline{Y}_1 \sqcup Y_2$ formally as a morphism from $Y_1$ to $Y_2$, this functor associates to $X$ a homomorphism

$$\mathcal{H}(X) : h(Y_1) \to h(Y_2)$$

that is, an element $\mathcal{H}(X) \in h(\overline{Y}_1 \sqcup Y_2)$. The most important feature is the composition law

$$\mathcal{H}(X_1 \cup_Y X_2) = \mathcal{H}(X_2) \circ \mathcal{H}(X_1)$$

So if a cobordism $X$ from $Y_1$ to $Y_2$ can be decomposed as a cobordism $X_1$ from $Y_1$ to an intermediate submanifold $Y$ and a cobordism $X_2$ from $Y$ to $Y_2$, then the homomorphism $\mathcal{H}(X)$ can be computed from $\mathcal{H}(X_1)$ and $\mathcal{H}(X_2)$ as their composition.

Donaldson invariants and Seiberg–Witten invariants fit neatly into the framework of a TQFT if one restricts to 3-manifolds which are disjoint unions of homology 3-spheres. In both the instanton and the monopole case, the vector spaces $h(Y)$ are Floer homology groups. The construction of Floer homology carries the Morse theory description of the homology of a finite-dimensional manifold over to an infinite-dimensional setting. In the instanton case, one considers the Chern–Simons function

$$\mathrm{CS}(a) = -\frac{1}{8\pi^2}\int_Y \mathrm{tr}\left(a \wedge \mathrm{d}a + \frac{2}{3} a \wedge a \wedge a\right)$$

This function is defined on the space of gauge equivalence classes of SU(2)-connections on $Y$. Note that for a homology 3-sphere, any SU(2) or PU(2)

principal bundle over $Y$ is trival. Choosing a trivialization, a connection becomes identified with a Lie-algebra-valued 1-form $a$. Critical points for the Chern–Simons functional lead to generaters in a chain complex the homology of which then gives the Floer groups. Such critical points correspond to flat connections on $Y$. The Floer homology groups $\mathrm{HF}_*(Y)$ are $\mathbb{Z}/8$-graded in the SU(2) case and $\mathbb{Z}/4$-graded in the SO(3) case. If $X$ is a 4-manifold with $b_1(X)=0$ and $b^+(X)>1$ and such that the boundary $\partial X$ is a disjoint union of homology 3-spheres, then the Donaldson invariants are linear maps

$$D_c : A(X) \longrightarrow HF_*(\partial X)$$

These invariants satisfy a composition law on the subring of $A(X)$ generated by two-dimensional homology classes (Donaldson 2002).

In the monopole case, one considers a Chern–Simons–Dirac functional

$$\mathrm{CSD}(a,\psi) = \frac{1}{2}\left(\int_Y \langle\psi, \mathcal{D}_a\psi\rangle \mathrm{dvol} - \int_Y a\wedge \mathrm{d}a\right)$$

and obtains integer graded Floer homology groups. Details and proofs of the relevant composition laws are announced.

*See also:* Floer Homology; Four-Manifold Invariants and Physics; Gauge Theory: Mathematical Applications; Instantons: Topological Aspects; Moduli Spaces: An Introduction; Several Complex Variables: Basic Geometric Theory; Topological Quantum Field Theory: Overview.

## Further Reading


Bauer S (2004) Refined Seiberg–Witten invariants. In: Donaldson, *et al.* (eds.) *Different Faces of Geometry*. New York, NY: Kluwer Academic/Plenum Publishers.

Donaldson SK (1983) An application of gauge theory to four-dimensional topology. *Journal of Differential Geometry* 18: 279–315.

Donaldson SK (2002) *Floer Homology Groups in Yang–Mills Theory*. Cambridge: Cambridge University Press.

Donaldson SK and Kronheimer PB (1990) *The Geometry of Four-Manifolds*. Oxford: Oxford University Press.

Feehan PMN and Leness TG (2003) On Donaldson and Seiberg–Witten invariants. In: *Topology and Geometry of Manifolds*, Proc. Sympos. Pure Math. (Athens, GA, 2001), vol. 71. pp. 237–248. Providence, RI: American Mathematical Society.

Fintushel R and Stern R (1996) The blow up formula for Donaldson invariants. *Annals of Mathematics (2)* 143: 529–546.

Freedman MH (1982) The topology of 4-dimensional manifolds. *Journal of Differential Geometry* 17: 357–453.

Freedman MH and Quinn F (1990) *Topology of 4-Manifolds*, Princeton Math. Ser., vol. 39. Princeton, NJ: Princeton University Press.

Friedman R and Morgan JW (eds.) (1997) *Gauge Theory and the Topology of Four-Manifolds*, IAS/Park City Math. Ser., vol. 4. Providence, RI: American Mathematical Society.

Furuta M (2001) Monopole equation and the 11/8-conjecture. *Mathematical Research Letter* 8: 363–400.

Göttsche L (2000) Donaldson invariants in algebraic geometry. School on Algebraic Geometry (Trieste 1999), 101–134. ICTP Lecture Notes **1**, Trieste, 2000.

Kronheimer PB and Mrowka T (1995) Embedded surfaces and the structure of Donaldson's polynomial invariants. *Journal of Differential Geometry* 41: 573–735.

Taubes CH (1994) The Seiberg–Witten invariants and sympletic forms. *Mathematical Research Letter* 1: 809–822.

Witten E (1994) Monopoles and four-manifolds. *Mathematics Research Letter* 1: 769–796.


# Gauge Theories from Strings

**P Di Vecchia**, Nordita, Copenhagen, Denmark



## Introduction

One of the most exciting properties of string theory, which led ten years ago to the formulation of the M theory as the unique theory unifying all interactions, has been the discovery that type II theories, besides a perturbative spectrum consisting of closed-string excitations, contain also a nonperturbative one consisting of "solitonic" $p$-dimensional objects called D$p$ branes. They are characterized by two important properties. They are coupled to closed-string states as the graviton, the dilaton, and the R–R $(p+1)$-form potential, and are described by a classical solution of the low-energy string effective action. Their dynamics is, on the other hand, described by open strings having the endpoints attached to their world volume and therefore satisfying Dirichlet boundary conditions in the directions transverse to their world volume. This is the reason why they are called D (Dirichlet) branes. Since the lightest open-string excitation corresponds to a gauge field, they have a gauge theory living on their world volume. This twofold description of D-branes has opened the way to study both the perturbative and nonperturbative properties of the gauge theory living on their world volume from their dynamics in terms of closed strings. With the addition of the decoupling limit, these two properties have led to the Maldacena (1998) conjecture of the equivalence between the maximally supersymmetric and conformal $\mathcal{N}=4$ super Yang–Mills and type IIB string theory on $\mathrm{AdS}_5 \times S^5$.

They have also been successfully applied to less supersymmetric and nonconformal gauge theories

that live on the world volume of fractional and wrapped branes. For general reviews of various approaches see Bertolini *et al.* (2000), Herzog *et al.* (2001), Bertolini (2003), Bigazzi *et al.* (2002), and Di Vecchia and Liccardo (2003). Also in these cases, one has constructed a classical solution of the supergravity equations of motion corresponding to these more sophisticated branes. These equations contain not only the supergravity fields present in the bulk ten-dimensional action but also boundary terms corresponding to the location of the branes. It turns out that in general the classical solution develops a naked singularity of the repulson type at short distances from the branes. This means that at short distances, it does not provide a reliable description of the branes. In the case of $\mathcal{N}=2$ supersymmetry, this can be explicitly seen because of the appearance of an enhançon located at distances slightly higher than the naked singularity (Johnson *et al.* 2000). The enhançon radius corresponds, in supergravity, to the distance where a brane probe becomes tensionless, and, in the gauge theory living on the branes, to the dynamically generated scale $\Lambda_{\mathrm{QCD}}$. Then, since short distances in supergravity correspond to large distances in the gauge theory, as implied by holography, the presence of the enhançon and of the naked singularity does not allow to get any information on the nonperturbative large-distance behavior of the gauge theory living on the D-branes. Above the radius of the enhançon, instead, the classical solution provides a good description of the branes and therefore it can be used to get information on the perturbative behavior of the gauge theory. This shows that, if we want to use the D-branes for studying the nonperturbative properties of the gauge theory living on their world volume, we must construct a classical solution that has no naked singularity at short distances in supergravity. We will see in a specific example that it will be possible to deform the classical solution, eliminating the naked singularity, and use it to describe nonperturbative properties as the gaugino condensate.

In this article, we review some of the results obtained by using fractional D3 branes of some orbifold and D5 branes wrapped on 2-cycles of some Calabi–Yau manifold. The analysis of the supersymmetric gauge theories living on the world volume of these D-branes will be based on the gauge/gravity relations that relate the gauge coupling constant and the $\theta$-angle to the supergravity fields (see, e.g., reference Di Vecchia *et al.* (2005) for a derivation of them):

$$\frac{4\pi}{g_{\mathrm{YM}}^2} = \frac{1}{g_s(2\pi\sqrt{\alpha'})^2}\int_{\mathcal{C}_2} \mathrm{d}^2\xi \mathrm{e}^{-\phi}\sqrt{\det(G_{AB}+B_{AB})} \qquad [1]$$

and

$$\theta_{\mathrm{YM}} = \frac{1}{2\pi\alpha' g_s}\int_{\mathcal{C}_2}(C_2 + C_0 B_2) \qquad [2]$$

where $\mathcal{C}_2$ is the 2-cycle where the branes are wrapped.

In the next section, we will describe the case of the fractional D3 branes of the orbifold $\mathbb{C}^2/\mathbb{Z}_2$ and show that the classical solution corresponding to a system of $N$ D3 and $M$ D7 branes reproduces the perturbative behavior of $\mathcal{N}=2$ super-QCD. Then, we will consider D5 branes wrapped on 2-cycles of a Calabi–Yau manifold described by the Maldacena–Núñez classical solution (Maldacena and Núñez 2001, Chamseddine and Volkov 1997) and show that in this case we are able to reproduce the phenomenon of gaugino condensate and to construct the complete $\beta$-function of $\mathcal{N}=1$ super Yang–Mills.

## Fractional D3 Branes of the Orbifold $\mathbb{C}^2/\mathbb{Z}_2$ and $\mathcal{N}=2$ Super-QCD

In this section, we consider fractional D3 and D7 branes of the noncompact orbifold $\mathbb{C}^2/\mathbb{Z}_2$ in order to study the properties of $\mathcal{N}=2$ super-QCD. We group the coordinates of the directions $(x^4,\ldots,x^9)$ transverse to the world volume of the D3 brane where the gauge theory lives, into three complex quantities: $z_1 = x^4 + \mathrm{i}x^5$, $z_2 = x^6 + \mathrm{i}x^7$, $z_3 = x^8 + \mathrm{i}x^9$. The nontrivial generator $h$ of $Z_2$ acts as $z_2 \to -z_2$, $z_3 \to -z_3$, leaving $z_1$ invariant. This orbifold has one fixed point, located at $z_2 = z_3 = 0$ and corresponding to a vanishing 2-cycle. Fractional D3 branes are D5 branes wrapped on the vanishing 2-cycle and therefore are, unlike bulk branes, stuck at the orbifold fixed point. By considering $N$ fractional D3 and $M$ fractional D7 branes of the orbifold $\mathbb{C}^2/\mathbb{Z}_2$, we are able to study $\mathcal{N}=2$ super-QCD with $M$ hypermultiplets. In order to do that, we need to determine the classical solution corresponding to the previous brane configuration. For the case of the orbifold $\mathbb{C}^2/Z_2$, the complete classical solution is found in Bertolini *et al.* (2002b); see also references therein and Bertolini *et al.* (2000) for a review on fractional branes. In the following, we write it explicitly for a system of $N$ fractional D3 branes with their world volume along the directions $x^0, x^1, x^2$, and $x^3$ and $M$ fractional D7 branes containing the D3 branes in their world volume and having the remaining four world-volume directions along the orbifolded ones. The metric,

the 5-form field strength, the axion, and the dilaton are given by

$$ds^2 = H^{-1/2}\eta_{\alpha\beta}\,dx^\alpha\,dx^\beta + H^{1/2}\left(\delta_{\ell m}\,dx^\ell\,dx^m + e^{-\phi}\delta_{ij}\,dx^i\,dx^j\right) \quad [3]$$

$$\widetilde{F}_{(5)} = d\left(H^{-1}\,dx^0 \wedge \cdots \wedge dx^3\right) + {}^*d\left(H^{-1}\,dx^0 \wedge \cdots \wedge dx^3\right) \quad [4]$$

$$\tau \equiv C_0 + ie^{-\phi} = i\left(1 - \frac{Mg_s}{2\pi}\log\frac{z}{\epsilon}\right)$$
$$z \equiv x^4 + ix^5 = ye^{i\theta} \quad [5]$$

where the self-dual field strength $\tilde{F}_{(5)}$ is given in terms of the NS–NS and R–R 2-forms $B_2$ and $C_2$ and of the 4-form potential $C_4$ by $\tilde{F}_{(5)} = dC_4 + C_2 \wedge dB_2$. The warp factor $H$ is a function of the coordinates $(x^4, \ldots, x^9)$ and $\epsilon$ is an infrared cutoff. We denote by $\alpha$ and $\beta$ the four directions corresponding to the world volume of the fractional D3 brane, by $\ell$ and $m$ those along the four orbifolded directions $x^6$, $x^7$, $x^8$, and $x^9$, and by $i$ and $j$ the directions $x^4$ and $x^5$ that are transverse to both the D3 and the D7 branes. The twisted fields are instead given by $B_2 = \omega_2 b, C_2 = \omega_2 c$ where $\omega_2$ is the volume form of the vanishing 2-cycle and

$$be^{-\phi} = \frac{(2\pi\sqrt{\alpha'})^2}{2}\left[1 + \frac{2N - M}{\pi}g_s\log\frac{y}{\epsilon}\right]$$
$$c + C_0 b = -2\pi\alpha'\theta g_s(2N - M) \quad [6]$$

The expression of $H$ (Kirsch and Vaman 2005) shows that the previous solution has a naked singularity of the repulson type at short distances. On the other hand, if we use a brane probe approaching from infinity the stack of branes, described by the previous classical solution, it can also be seen that the tension of the probe vanishes at a distance that is larger than that of the naked singularity. The point where the probe brane becomes tensionless is called "enhançon" (Johnson *et al.* 2000) and at this point the classical solution does not describe anymore the stack of fractional branes.

Let us now use the gauge/gravity relations given in the introduction, to determine the coupling constants of the world-volume theory from the supergravity solution. In the case of fractional D3 branes of the orbifold $\mathbb{C}^2/\mathbb{Z}_2$, that is characterized by one single vanishing 2-cycle $\mathcal{C}_2$, the gauge coupling constant given in eqn [1] reduces to

$$\frac{1}{g_{YM}^2} = \frac{1}{4\pi g_s(2\pi\sqrt{\alpha'})^2}\int_{\mathcal{C}_2} e^{-\phi}B_2 \quad [7]$$

By inserting the classical solution in eqns [7] and [2], we get the following expressions for the gauge coupling constant and the $\theta_{YM}$ angle (Bertolini *et al.* 2002b):

$$\frac{1}{g_{YM}^2} = \frac{1}{8\pi g_s} + \frac{2N - M}{16\pi^2}\log\frac{y^2}{\epsilon^2}$$
$$\theta_{YM} = -\theta(2N - M) \quad [8]$$

Notice that the gauge coupling constant appearing in the previous equation is the "bare" gauge coupling constant computed at the scale $m \sim y/\alpha'$, while the square of the bare gauge coupling constant computed at the cutoff $\Lambda \sim \epsilon/\alpha'$ is equal to $8\pi g_s$.

In the case of an $\mathcal{N} = 2$ supersymmetric gauge theory, the gauge multiplet contains a complex scalar field $\Psi$ that corresponds to the complex coordinate $z$ transverse to both the world volume of the D3 brane and the four orbifolded directions: $\Psi \sim z/2\pi\alpha'$. This is another example of holographic identification between a quantity, $\Psi$, peculiar of the gauge theory living on the fractional D3 branes and another one, the coordinate $z$, peculiar of supergravity. It allows one to obtain the gauge theory anomalies from the supergravity background. In fact, since we know how the scale and U(1) transformations act on $\Psi$, from the previous gauge/gravity relation we can deduce how they act on $z$, namely

$$\Psi \to se^{2i\alpha}\Psi \Longleftrightarrow z \to se^{2i\alpha}z \Longrightarrow y \to sy$$
$$\theta \to \theta + 2\alpha \quad [9]$$

Those transformations do not leave invariant the supergravity background in eqn [6] and when we use them in eqns [7] and [2], they generate the anomalies of the gauge theory living on the fractional D3 branes. In fact, by acting with those transformations in eqns [8], we get

$$\frac{1}{g_{YM}^2} \to \frac{1}{g_{YM}^2} + \frac{2N - M}{8\pi^2}\log s$$
$$\theta_{YM} \to \theta_{YM} - 2\alpha(2N - M) \quad [10]$$

The first equation generates the $\beta$-function of $\mathcal{N} = 2$ super-QCD with $M$ hypermultiplets given by

$$\beta(g_{YM}) = -\frac{2N - M}{16\pi^2}g_{YM}^3 \quad [11]$$

while the second one reproduces the chiral U(1) anomaly (Klebanov *et al.* 2002, Bertolini *et al.* 2002a). In particular, if we choose $\alpha = 2\pi/(2(2N - M))$, then $\theta_{YM}$ is shifted by a factor $2\pi$. But since $\theta_{YM}$ is periodic of $2\pi$, this means that the subgroup $Z_{2(2N-M)}$ is not anomalous in perfect agreement with the gauge theory results.

## Wrapped D5 Branes and $\mathcal{N}=1$ Super Yang–Mills

In this section, we will consider the classical solution corresponding to $N$ D5 branes wrapped on a 2-cycle of a noncompact Calabi–Yau space and we use it to study the properties of the gauge theory living on their world volume that can be shown to be $\mathcal{N}=1$ super Yang–Mills.

We start by writing the classical solution found in Maldacena and Núñez (2001) and Chamseddine and Volkov (1997). It has a nontrivial metric:

$$ds^2_{10} = e^{\Phi}\left[dx^2_{1,3} + \frac{e^{2h}}{\lambda^2}\left(d\tilde{\theta}^2 + \sin^2\tilde{\theta}\, d\tilde{\varphi}^2\right)\right] + \frac{e^{\Phi}}{\lambda^2}\left[d\rho^2 + \sum_{a=1}^{3}(\sigma^a - \lambda A^a)^2\right] \qquad [12]$$

a 2-form R–R potential

$$C^{(2)} = \frac{1}{4\lambda^2}\left[(\psi+\psi_0)\left(\sin\theta'\, d\theta'\wedge d\phi - \sin\tilde{\theta}\, d\tilde{\theta}\wedge d\tilde{\varphi}\right) - \cos\theta'\cos\tilde{\theta}\, d\phi\wedge d\tilde{\varphi}\right] + \frac{a}{2\lambda^2}\left[d\tilde{\theta}\wedge\sigma^1 - \sin\tilde{\theta}\, d\tilde{\varphi}\wedge\sigma^2\right] \qquad [13]$$

and a dilaton

$$e^{2\Phi} = \frac{\sinh 2\rho}{2e^h} \qquad [14]$$

where

$$\begin{aligned} e^{2h} &= \rho\coth 2\rho - \frac{\rho^2}{\sinh^2 2\rho} - \frac{1}{4} \\ e^{2k} &= e^h\frac{\sinh 2\rho}{2} \\ a &= \frac{2\rho}{\sinh 2\rho} \end{aligned} \qquad [15]$$

and

$$\begin{aligned} A^1 &= -\frac{1}{2\lambda}a(r)\, d\tilde{\theta} \\ A^2 &= \frac{1}{2\lambda}a(r)\sin\tilde{\theta}\, d\tilde{\varphi} \\ A^3 &= -\frac{1}{2\lambda}\cos\tilde{\theta}\, d\tilde{\varphi} \end{aligned} \qquad [16]$$

with $\rho \equiv \lambda r$ and $\lambda^{-2} = N g_s \alpha'$. The left-invariant 1-forms of $S^3$ are

$$\begin{aligned} \sigma^1 &= \frac{1}{2}\left[\cos\psi\, d\theta' + \sin\theta'\sin\psi\, d\phi\right] \\ \sigma^2 &= -\frac{1}{2}\left[\sin\psi\, d\theta' - \sin\theta'\cos\psi\, d\phi\right] \\ \sigma^3 &= \frac{1}{2}\left[d\psi + \cos\theta'\, d\phi\right] \end{aligned} \qquad [17]$$

with $0 \le \theta' \le \pi, 0 \le \phi \le 2\pi$, and $0 \le \psi \le 4\pi$. The variables $\tilde{\theta}$ and $\tilde{\varphi}$ describe a two-dimensional sphere and vary in the range $0 \le \tilde{\theta} \le \pi$ and $0 \le \tilde{\varphi} \le 2\pi$. Before proceeding, here we want to stress the fact that the presence of the function $a(\rho) \neq 0$ makes the solution regular everywhere. This will allow us to use it later on to describe the nonperturbative gaugino condensate property of $\mathcal{N}=1$ super Yang–Mills.

We can now use the previous solution for computing the running coupling constant and the $\theta$ parameter of $\mathcal{N}=1$ super Yang–Mills (see Di Vecchia *et al.* (2002), Bertolini and Merlatti (2003), and Mück (2003) reviewed in Bertolini (2003), Di Vecchia and Liccardo (2003), and Imeroni (2003)). In order to do that, we have to fix the cycle on which to perform the integrals in eqns [1] and [2]. It turns out that this 2-cycle is specified by

$$\tilde{\theta} = \theta' \cdot \tilde{\varphi} = -\phi, \quad \psi = 0 \qquad [18]$$

keeping $\rho$ fixed. If we now compute the gauge couplings on the previous cycle with $B_2 = C_0 = 0$, we get

$$\frac{4\pi^2}{N g^2_{YM}} = \rho\coth 2\rho + \tfrac{1}{2}a(\rho)\cos\psi \qquad [19]$$

and

$$\theta_{YM} = \frac{1}{2\pi g_s\alpha'}\int_{S^2} C_2 = -N(\psi + a(\rho)\sin\psi + \psi_0) \qquad [20]$$

where we have kept $\psi \neq 0$ for reasons that will become clear in a moment. Equation [19] shows that the coupling constant is running as a function of the distance $\rho$ from the branes. In order to obtain the correct running of the gauge theory, we have to find a relation between $\rho$ and the renormalization group scale $\mu$. This can be obtained with the following considerations. If we look at the previous solution, it is easy to see that the metric in eqn [12] is invariant under the following transformations:

$$\begin{aligned} \psi &\to \psi + 2\pi \quad \text{if } a \neq 0 \\ \psi &\to \psi + 2\epsilon \quad \text{if } a = 0 \end{aligned} \qquad [21]$$

where $\epsilon$ is an arbitrary constant. On the other hand, $C_2$ is not invariant under the previous transformations, but its flux, that is exactly equal to $\theta_{YM}$ in eqn [20], changes by an integer multiple of $2\pi$:

$$\theta_{YM} = \frac{1}{2\pi\alpha' g_s}\int_{C_2} C_2 \to \theta_{YM} + \begin{cases} -2\pi N, & \textit{if } a \neq 0 \\ -2N\epsilon, & \textit{if } a = 0, \ \epsilon = \frac{\pi k}{N} \end{cases} \qquad [22]$$

But since the physics does not change when $\theta_{YM} \to \theta_{YM} + 2\pi$, one gets that the transformation in eqn [22] is an invariance. Notice that also eqn [19] for

the gauge coupling constant is invariant under the transformation in eqn [21]. The previous considerations show that the classical solution and also the gauge couplings are invariant under the $Z_2$ transformation if $a \neq 0$, while this symmetry becomes $Z_{2N}$ if $a$ is taken to be zero. As a consequence, since in the ultraviolet $a(\rho)$ is exponentially small, we can neglect it and we have a $Z_{2N}$ symmetry, while in the infrared we cannot neglect $a(\rho)$ anymore and we have only a $Z_2$ symmetry left. This fits very well with the fact that $\mathcal{N}=1$ super Yang–Mills has a nonzero gaugino condensate $<\lambda\lambda>$ that is responsible for the breaking of $Z_{2N}$ into $Z_2$. Therefore, it is natural to identify the gaugino condensate precisely with the function $a(\rho) \neq 0$ that makes the classical solution regular also at short distances in supergravity (Di Vecchia *et al.* 2002, Apreda *et al.* 2002):

$$<\lambda\lambda> \sim \Lambda^3 = \mu^3 a(\rho) \qquad [23]$$

This provides the relation between the renormalization group scale $\mu$ and the supergravity spacetime parameter $\rho$. In the ultraviolet (large $\rho$) $a(\rho)$ is exponentially suppressed and in eqns [19] and [20] we can neglect it obtaining

$$\frac{4\pi^2}{Ng_{\mathrm{YM}}^2} = \rho \coth 2\rho \qquad [24]$$
$$\theta_{\mathrm{YM}} = -N(\psi + \psi_0)$$

The chiral anomaly can be obtained by performing the transformation $\psi \to \psi + 2\epsilon$ and getting

$$\theta_{\mathrm{YM}} \to \theta_{\mathrm{YM}} - 2N\epsilon \qquad [25]$$

This implies that the $Z_{2N}$ transformations corresponding to $\epsilon = \pi k/N$ are symmetries because they shift $\theta_{\mathrm{YM}}$ by multiples of $2\pi$.

In general, however, eqns [19] and [20] are only invariant under the $Z_2$ subgroup of $Z_{2N}$ corresponding to the transformation

$$\psi \to \psi + 2\pi \qquad [26]$$

that changes $\theta_{\mathrm{YM}}$ in eqn [20] as follows:

$$\theta_{\mathrm{YM}} \to \theta_{\mathrm{YM}} - 2N\pi \qquad [27]$$

leaving invariant the gaugino condensate:

$$<\lambda^2> = \mu^3 \frac{16\pi^2}{3Ng_{\mathrm{YM}}^2} \mathrm{e}^{-8\pi^2/Ng_{\mathrm{YM}}^2} \mathrm{e}^{\mathrm{i}\theta_{\mathrm{YM}}/N} \qquad [28]$$

Therefore, the chiral anomaly and the breaking of $Z_{2N}$ to $Z_2$ are encoded in eqns [19] and [20]. Finally, if we put $\psi = 0$ in eqn [19], we get

$$\frac{4\pi^2}{Ng_{\mathrm{YM}}^2} = \rho \coth 2\rho - \tfrac{1}{2}a(\rho) = \rho \tanh \rho \qquad [29]$$

This equation taken together with eqn [23] allows us to determine the running coupling constant as a function of $\mu$. From it, we get (Di Vecchia *et al.* 2002, Di Vecchia and Liccardo 2003) the Novikov–Shifman–Vainshtein–Zacharov (NSVZ) $\beta$-function plus nonperturbative corrections due to fractional instantons:

$$\beta(g_{\mathrm{YM}}) = -\frac{3Ng_{\mathrm{YM}}^3}{16\pi^2} \frac{1 + \dfrac{4\pi^2}{Ng_{\mathrm{YM}}^2}\sinh^{-2}\dfrac{4\pi^2}{Ng_{\mathrm{YM}}^2}}{1 - \dfrac{Ng_{\mathrm{YM}}^2}{8\pi^2} + \dfrac{1}{2}\sinh^{-2}\dfrac{4\pi^2}{Ng_{\mathrm{YM}}^2}} \qquad [30]$$

where in the ultraviolet we have approximated $\rho$ with $4\pi^2/(Ng_{\mathrm{YM}}^2) \coth 4\pi^2/(Ng_{\mathrm{YM}}^2)$.

*See also*: AdS/CFT Correspondence; Anomalies; BF Theories; Brane Construction of Gauge Theories; Gauge Theory: Mathematical Applications; Noncommutative Geometry from Strings; Nonperturbative and Topological Aspects of Gauge Theory; Perturbation Theory and its Techniques; Seiberg–Witten Theory; Superstring Theories.

## Further Reading

Apreda R, Bigazzi F, Cotrone AL, Petrini M, and Zaffaroni A (2002) Some comments on $\mathcal{N}=1$ gauge theories from wrapped branes. *Physics Letters B* 536: 161–168 (hep-th/0112236).

Bertolini M (2003) Four lectures on the gauge/gravity correspondence. *International Journal of Modern Physics* A18: 5647–5712 (hep-th/0303160).

Bertolini M, Di Vecchia P, Frau M, Lerda A, and Marotta R (2002a) More anomalies from fractional branes. *Physics Letters B* 540: 104–110 (hep-th/0202195).

Bertolini M, Di Vecchia P, Frau M, Lerda A, and Marotta R (2002b) $N=2$ gauge theories on systems of fractional D3/D7 branes. *Nuclear Physics B* 621: 157–178 (hep-th/0107057).

Bertolini M, Di Vecchia P, and Marotta R (2000) $N=2$ four-dimensional gauge theories from fractional branes. In: Olshanetsky M and Vainshtein A (eds.) *Multiple Facets of Quantization and Supersymmetry*, pp. 730–773. (hep-th/0112195). Singapore: World Scientific.

Bertolini M and Merlatti P (2003) A note on the dual of $\mathcal{N}=1$ super Yang–Mills theory. *Physics Letters B* 556: 80–86 (hep-th/0211142).

Bigazzi F, Cotrone AL, Petrini M, and Zaffaroni A (2002) Supergravity duals of supersymmetric four dimensional gauge theories. *Rivista del Nuovo Cimento* 25N12: 1–70 (hep-th/0303191).

Chamseddine AH and Volkov MS (1997) Non-abelian BPS monopoles in $\mathcal{N}=4$ gauged supergravity. *Physical Review Letters* 79: 3343–3346 (hep-th/9707176).

Di Vecchia P, Lerda A, and Merlatti P (2002) $\mathcal{N}=1$ and $\mathcal{N}=2$ super Yang–Mills theories from wrapped branes. *Nuclear Physics B* 646: 43–68 (hep-th/0205204).

Di Vecchia P and Liccardo A (2003) Gauge theories from D branes, Lectures given at the School "Frontiers in Number Theory, Physics and Geometry," Les Houches, hep-th/0307104.

Di Vecchia P, Liccardo A, Marotta R, and Pezzella F (2005) On the gauge/gravity correspondence and the open/closed string duality. *International Journal of Modern Physics A* 20: 4699–4796. hep-th/0503156.

Herzog CP, Klebanov IR, and Ouyang P (2001) D-branes on the conifold and $\mathcal{N}=1$ gauge/gravity dualities. Based on I.R.K.'s Lectures at the Les Houches Summer School Session 76, "Gravity, Gauge Theories, and Strings," hep-th/0205100.

Imeroni E (2003) *Studies of the Gauge/String Theory Correspondence*. Ph.D. thesis, hep-th/0312070.

Johnson CV, Peet AW, and Polchinski J (2000) Gauge theory and the excision of repulson singularities. *Physical Review D* 61: 086001 (hep-th/9911161).

Kirsch I and Vaman D (2005) The D3/D7 background and flavour dependence of Regge trajectories, hep-th/0505164.

Klebanov IR, Ouyang P, and Witten E (2002) A gravity dual of the chiral anomaly. *Physical Review D* 65: 105007 (hep-th/0202056).

Maldacena J (1998) The large $N$ limit of superconformal field theories and supergravity. *Advances in Theoretical and Mathematical Physics* 2: 231–252 (hep-th/9711200).

Maldacena J and Núñez C (2001) Towards the large $N$ limit of $N=1$ super Yang Mills. *Physical Review Letters* 86: 588–591 (hep-th/0008001).

Mück W (2003) Perturbative and nonperturbative aspects of pure $\mathcal{N}=1$ super Yang–Mills theory from wrapped branes. *Journal of High Energy Physics* 0302: 013 (hep-th/0301171).

# Gauge Theory: Mathematical Applications

**S K Donaldson**, Imperial College, London, UK



## Introduction

This article surveys some developments in pure mathematics which have, to varying degrees, grown out of the ideas of gauge theory in mathematical physics. The realization that the gauge fields of particle physics and the connections of differential geometry are one and the same has had wide-ranging consequences, at different levels. Most directly, it has led mathematicians to work on new kinds of questions, often shedding light later on well-established problems. Less directly, various fundamental ideas and techniques, notably the need to work with the infinite-dimensional gauge symmetry group, have found a place in the general world-view of many mathematicians, influencing developments in other fields. Still less directly, the work in this area – between geometry and mathematical physics – has been a prime example of the interaction between these fields which has been so fruitful since the 1970s.

The body of this article is divided into three sections: roughly corresponding to analysis, geometry, and topology. However, the different topics come together in many different ways: indeed the existence of these links between the topics is one of the most attractive features of the area.

### Gauge Transformations

For a review of the usual foundational material on connections, curvature, and related differential geometric constructions, the reader is referred to standard texts. We will, however, briefly recall the notions of gauge transformations and gauge fixing. The simplest case is that of abelian gauge theory – connections on a U(1)-bundle, say over $\boldsymbol{R}^3$. In that case the connection form, representing the connection in a local trivialization, is a pure imaginary 1-form $A$, which can also be identified with a vector field $\boldsymbol{A}$. The curvature of the connection is the 2-form $\mathrm{d}A$. Changing the local trivialization by a U(1)-valued function $g=\mathrm{e}^{\mathrm{i}\chi}$ changes the connection form to

$$\tilde{A} = A - \mathrm{d}gg^{-1} = A - \mathrm{i}\mathrm{d}\chi$$

The forms $A, \tilde{A}$ are two representations of the same geometric object: just as the same metric can be represented by different expressions in different coordinate systems. One may want to fix this choice of representation, usually by choosing $A$ to satisfy the Coulomb gauge condition $\mathrm{d}^*A=0$ (equivalently $\operatorname{div}\boldsymbol{A}=0$), supplemented by appropriate boundary conditions. Here we are using the standard Euclidean metric on $\boldsymbol{R}^3$. (Throughout this article we will work with positive-definite metrics, regardless of the fact that – at least at the classical level – the Lorentzian signature may have more obvious bearing on physics.) Arranging this choice of gauge involves solving a linear partial differential equation (PDE) for $\chi$.

The case of a general structure group $G$ is not much different. The connection form $A$ now takes values in the Lie algebra of $G$ and the curvature is given by the expression

$$F = \mathrm{d}A + \tfrac{1}{2}[A, A]$$

The change of bundle trivialization is given by a $G$-valued function and the resulting change in the connection form is

$$\tilde{A} = gAg^{-1} - \mathrm{d}gg^{-1}$$

(Our notation here assumes that $G$ is a matrix group, but this is not important.) Again, we can seek to impose the Coulomb gauge condition $\mathrm{d}^*A=0$, but now we cannot linearize this equation as before.

We can carry the same ideas over to a global problem, working on a $G$-bundle $P$ over a general

Riemannian manifold $M$. The space of connections on $P$ is an affine space $\mathcal{A}$: any two connections differ by a bundle-valued 1-form. Now the gauge group $\mathcal{G}$ of automorphisms of $P$ acts on $\mathcal{A}$ and, again, two connections in the same orbit of this action represent essentially the same geometric object. Thus, in a sense we would really like to work on the quotient space $\mathcal{A}/\mathcal{G}$. Working locally in the space of connections, near to some $A_0$, this is quite straightforward. We represent the nearby connections as $A_0 + a$, where $a$ satisfies the analog of the coulomb condition

$$d_A^* a = 0$$

Under suitable hypotheses, this condition picks out a unique representative of each nearby orbit. However, this gauge-fixing condition need not single out a unique representative if we are far away from $A_0$: indeed, the space $\mathcal{A}/\mathcal{G}$ typically has, unlike $\mathcal{A}$, a complicated topology which means that it is impossible to find any such global gauge-fixing condition. As noted above, this is one of the distinctive features of gauge theory. The gauge group $\mathcal{G}$ is an infinite-dimensional group, but one of a comparatively straightforward kind – much less complicated than the diffeomorphism groups relevant in Riemannian geometry for example. One could argue that one of the most important influences of gauge theory has been to accustom mathematicians to working with infinite-dimensional symmetry groups in a comparatively simple setting.

## Analysis and Variational Methods

### The Yang–Mills Functional

A primary object brought to mathematicians attention by physics is the Yang–Mills functional

$$\mathrm{YM}(A) = \int_M |F_A|^2 \, d\mu$$

Clearly, $\mathrm{YM}(A)$ is non-negative and vanishes if and only if the connection is flat: it is broadly analogous to functionals such as the area functional in minimal submanifold theory, or the energy functional for maps. As such, one can fit into a general framework associated with such functionals. The Euler–Lagrange equations are the Yang–Mills equations

$$d_A^* F_A = 0$$

For any solution (a Yang–Mills connection), there is a "Jacobi operator" $H_A$ such that the second variation is given by

$$\mathrm{YM}(A + ta) = \mathrm{YM}(A) + t^2 \langle H_A a, a \rangle + O(t^3)$$

The omnipresent phenomenon of gauge invariance means that Yang–Mills connections are never isolated, since we can always generate an infinite-dimensional family by gauge transformations. Thus, as explained in the last section, one imposes the gauge-fixing condition $d_A^* a = 0$. Then the operator $H_A$ can be written as

$$H_A a = \Delta_A a + [F_A, a]$$

where $\Delta_A$ is the bundle-valued "Hodge Laplacian" $d_A d_A^* + d_A^* d_A$ and the expression $[F_A, a]$ combines the bracket in the Lie algebra with the action of $\Lambda^2$ on $\Lambda^1$. This is a self-adjoint elliptic operator and, if $M$ is compact, the span of the negative eigenspaces is finite dimensional, the dimension being defined to be the index of the Yang–Mills connection $A$.

In this general setting, a natural aspiration is to construct a "Morse theory" for the functional. Such a theory should relate the topology of the ambient space to the critical points and their indices. In the simplest case, one could hope to show that for any bundle $P$ there is a Yang–Mills connection with index 0, giving a minimum of the functional. More generally, the relevant ambient space here is the quotient $\mathcal{A}/\mathcal{G}$ and one might hope that the rich topology of this is reflected in the solutions to the Yang–Mills equations.

### Uhlenbeck's Theorem

The essential foundation needed to underpin such a "direct method" in the calculus of variations is an appropriate compactness theorem. Here the dimension of the base manifold $M$ enters in a crucial way. Very roughly, when a connection is represented locally in a Coulomb gauge, the Yang–Mills action combines the $L^2$-norm of the derivative of the connection form $A$ with the $L^2$-norm of the quadratic term $[A, A]$. The latter can be estimated by the $L^4$-norm of $A$. If $\dim M \leq 4$, then the Sobolev inequalities allow the $L^4$-norm of $A$ to be controlled by the $L^2$-norm of its derivative, but this is definitely not true in higher dimensions. Thus, $\dim M = 4$ is the "critical dimension" for this variational problem. This is related to the fact that the Yang–Mills equations (and Yang–Mills functional) are conformally invariant in four dimensions. For any nontrivial Yang–Mills connection over the 4-sphere, one generates a one-parameter family of Yang–Mills connections, on which the functional takes the same value, by applying conformal transformations corresponding to dilations of $\boldsymbol{R}^4$. In such a family of connections the integrand $|F_A|^2$ – the "curvature density" – converges to a $\delta$-function

at the origin. More generally, one can encounter sequences of connections over 4-manifolds for which YM is bounded but which do not converge, the Yang–Mills density converging to $\delta$-functions. There is a detailed analogy with the theory of the harmonic maps energy functional, where the relevant critical dimension (for the domain of the map) is 2.

The result of Uhlenbeck (1982), which makes these ideas precise, considers connections over a ball $B^n \subset \mathbf{R}^n$. If the exponent $p \geq 2n$, then there are positive constants $\epsilon(p,n), C(p,n) > 0$ such that any connection with $\|F\|_{L^p(B^n)} \leq \epsilon$ can be represented in Coulomb gauge over the ball, by a connection form which satisfies the condition $d^*A = 0$, together with certain boundary conditions, and

$$\|A\|_{L_1^p} \leq C\|F\|_{L^p}$$

In this Coulomb gauge, the Yang–Mills equations are elliptic and it follows readily that, in this setting, if the connection $A$ is Yang–Mills one can obtain estimates on all derivatives of $A$.

### Instantons in Four Dimensions

This result of Uhlenbeck gives the analytical basis for the direct method of the calculus of variations for the Yang–Mills functional over base manifolds $M$ of dimension $\leq 3$. For example, any bundle over such a manifold must admit a Yang–Mills connection, minimizing the functional. Such a statement is definitely false in dimensions $\geq 5$. For example, an early result of Bourguignon and Lawson (1981) and Simons asserts that there is no minimizing connection on any bundle over $S^n$ for $n \geq 5$. The proof exploits the action of the conformal transformations of the sphere. In the critical dimension 4, the situation is much more complicated. In four dimensions, there are the renowned "instanton" solutions of the Yang–Mills equation. Recall that if $M$ is an oriented 4-manifold the Hodge $*$-operation is an involution of $\Lambda^2 T^*M$ which decomposes the two forms into self-dual and anti-self-dual parts, $\Lambda^2 T^*M = \Lambda^+ \oplus \Lambda^-$. The curvature of a connection can then be written as

$$F_A = F_A^+ + F_A^-$$

and a connection is a self-dual (respectively anti-self-dual) instanton if $F_A^-$ (respectively $F_A^+$) is 0. The Yang–Mills functional is

$$\mathrm{YM}(A) = \|F_A^+\|^2 + \|F_A^-\|^2$$

while the difference $\|F_A^+\|^2 - \|F_A^-\|^2$ is a topological invariant $\kappa(P)$ of the bundle $P$, obtained by evaluating a four-dimensional characteristic class on $[M]$. Depending on the sign of $\kappa(P)$, the self-dual or anti-self-dual connections (if any exist) minimize the Yang–Mills functional among all connections on $P$. These instanton solutions of the Yang–Mills equations are analogous to the holomorphic maps from a Riemann surface to a Kähler manifold, which minimize the harmonic maps energy functional in their homotopy class.

### Moduli Spaces

The instanton solutions typically occur in "moduli spaces." To fix ideas, let us consider bundles with structure group SU(2), in which case $\kappa(P) = -8\pi^2 c_2(P)$. For each $k > 0$, we have a moduli space $\mathcal{M}_k$ of anti-self-dual instantons on a bundle $P_k \to M^4$, with $c_2(P_k) = k$. It is a manifold of dimension $8k - 3$. The general goal of the calculus of variations in this setting is to relate three things:

1. the topology of the space $\mathcal{A}/\mathcal{G}$ of equivalence classes of connections on $P_k$;
2. the topology of the moduli space $\mathcal{M}_k$ of instantons; and
3. the existence and indices of other, nonminimal, solutions to the Yang–Mills equations on $P_k$.

In this direction, a very influential conjecture was made by Atiyah and Jones (1978). They considered the case when $M = S^4$ and, to avoid certain technicalities, work with spaces of "framed" connections, dividing by the restricted group $\mathcal{G}_0$ of gauge transformations equal to the identity at infinity. Then, for any $k$, the quotient $\mathcal{A}/\mathcal{G}_0$ is homotopy equivalent to the third loop space $\Omega^3 S^3$ of based maps from the 3-sphere to itself. The corresponding "framed" moduli space $\tilde{\mathcal{M}}_k$ is a manifold of dimension $8k$ (a bundle over $\mathcal{M}_k$ with fiber SO(3)). Atiyah and Jones conjectured that the inclusion $\tilde{\mathcal{M}}_k \to \mathcal{A}/\mathcal{G}_0$ induces an isomorphism of homotopy groups $\pi_l$ in a range of dimensions $l \leq l(k)$, where $l(k)$ increases with $k$. This would be consistent with what one might hope to prove by the calculus of variations if there were no other Yang–Mills solutions, or if the indices of such solutions increased with $k$.

The first result along these lines was due to Bourguignon and Lawson (1981), who showed that the instanton solutions are the only local minima of the Yang–Mills functional over the 4-sphere. Subsequently, Taubes (1983) showed that the index of an non-instanton Yang–Mills connection $P_k$ is at least $k+1$. Taubes' proof used ideas related to the action of the quaternions and the hyper-Kähler structure on the $\tilde{\mathcal{M}}_k$ (see the section on hyper-Kähler quotients). Contrary to some expectations, it was shown by

Sibner *et al.* (1989) that nonminimal solutions do exist; some later constructions were very explicit (Sadun and Segert 1992). Taubes' index bound gave ground for hope that an analytical proof of the Atiyah–Jones conjecture might be possible, but this is not at all straightforward. The problem is that in the critical dimension 4 a mini–max sequence for the Yang–Mills functional in a given homotopy class may diverge, with curvature densities converging to sums of $\delta$-functions as outlined above. This is related to the fact that the $\tilde{\mathcal{M}}_k$ are not compact. In a series of papers culminating in a framework for Morse theory for Yang–Mills functional, Taubes (1998) succeeded in proving a partial version of the Atiyah–Jones conjecture, together with similar results for general base manifolds $M^4$. Taubes showed that, if the homotopy groups of the moduli spaces stabilize as $k \to \infty$, then the limit must be that predicted by Atiyah and Jones. Related analytical techniques were developed for other variational problems at the critical dimension involving "critical points at infinity." The full Atiyah–Jones conjecture was established by Boyer *et al.* but using geometrical techniques: the "explicit" description of the moduli spaces obtained from the Atiyah–Drinfeld–Hitchin–Manin (ADHM) construction (see below). A different geometrical proof was given by Kirwan (1994), together with generalizations to other gauge groups.

There was a parallel story for the solutions of the Bogomolony equation over $\boldsymbol{R}^3$, which we will not recount in detail. Here the base dimension is below the critical case but the analytical difficulty arises from the noncompactness of $\boldsymbol{R}^3$. Taubes succeeded in overcoming this difficulty and obtained relations between the topology of the moduli space, the appropriate configuration space and the higher critical points. Again, these higher critical points exist but their index grows with the numerical parameter corresponding to $k$. At about the same time, Donaldson (1984) showed that the moduli spaces could be identified with spaces of rational maps (subsequently extended to other gauge groups). The analog of the Atiyah–Jones conjecture is a result on the topology of spaces of rational maps proved earlier by Segal, which had been one of the motivations for Atiyah and Jones.

### Higher Dimensions

While the scope for variational methods in Yang–Mills theory in higher dimensions is very limited, there are useful analytical results about solutions of the Yang–Mills equations. An important monotonicity result was obtained by Price (1983). For simplicity, consider a Yang–Mills connection over the unit ball $B^n \subset \boldsymbol{R}^n$. Then Price showed that the normalized energy

$$E(A, B(r)) = \frac{1}{r^{n-4}} \int_{|x| \leq r} |F|^2 \, \mathrm{d}\mu$$

decreases with $r$. Nakajima (1988) and Uhlenbeck used this monotonicity to show that for each $n$ there is an $\epsilon$ such that if $A$ is a Yang–Mills connection over a ball with $E(A, B(r)) \leq \epsilon$ then all derivatives of $A$, in a suitable gauge, can be controlled by $E(A, B(r))$. Tian (2000) showed that if $A_i$ is a sequence of Yang–Mills connections over a compact manifold $M$ with bounded Yang–Mills functional, then there is a subsequence which converges away from a set $Z$ of Haussdorf codimension at least 4 (extending the case of points in a 4-manifold). Moreover, the singular set $Z$ is a minimal subvariety, in a suitably generalized sense.

In higher dimensions, important examples of Yang–Mills connections arise within the framework of "calibrated geometry." Here, we consider a Riemannian $n$-manifold $M$ with a covariant constant calibrating form $\Omega \in \Omega_M^{n-4}$. There is then an analog of the instanton equation

$$F_A = \pm * (\Omega \wedge F_A)$$

whose solutions minimize the Yang–Mills functional. This includes the Hermitian Yang–Mills equation over a Kähler manifold (see the section on moment maps) and also certain equations over manifolds with special holonomy groups (Donaldson and Thomas 1998). For these "higher-dimensional instantons," Tian shows that the singular sets $Z$ that arise are calibrated varieties.

### Gluing Techniques

Another set of ideas from PDEs and analysis which has had great impact in gauge theory involves the construction of solutions to appropriate equations by the following general scheme:

1. constructing an "approximate solution," formed from some standard models using cutoff functions;
2. showing that the approximate solution can be deformed to a true solution by means of an implicit function theorem.

The heart of the second step usually consists of estimates for the relevant linear differential operator. Of course, the success of this strategy depends on the particular features of the problem. This approach, due largely to Taubes, has been particularly effective in finding solutions to the first-order instanton equations and their relatives. (The applicability of the approach is connected to the fact that

such solutions typically occur in moduli spaces and one can often "see" local coordinates in the moduli space by varying the parameters in the approximate solution.) Taubes applied this approach to the Bogomolny monopole equation over $\boldsymbol{R}^3$ (Jaffe and Taubes 1980) and to construct instantons over general 4-manifolds (Taubes 1982). In the latter case, the approximate solutions are obtained by transplanting standard solutions over $\boldsymbol{R}^4$ – with curvature density concentrated in a small ball – to small balls on the 4-manifold, glued to the trivial flat connection over the remainder of the manifold. These types of techniques have now become a fairly standard part of the armory of many differential geometers, working both within gauge theory and other fields. An example of a problem where similar ideas have been used is Joyce's construction of constant of manifolds with exceptional holonomy groups (Joyce 1996). (Of course, it is likely that similar techniques have been developed over the years in many other areas, but Taubes' work in gauge theory has done a great deal to bring them into prominence.)

## Geometry: Integrability and Moduli Spaces

### The Ward Correspondence

Suppose that $S$ is a complex surface and $\omega$ is the 2-form corresponding to a Hermitian metric on $S$. Then $S$ is an oriented Riemannian 4-manifold and $\omega$ is a self-dual form. The orthogonal complement of $\omega$ in $\Lambda^+$ can be identified with the real parts of forms of type $(0,2)$. Hence, if $A$ is an anti-self-dual instanton connection on a principle $\mathrm{U}(r)$-bundle over $S$ the $(0,2)$ part of the curvature of $A$ vanishes. This is the integrability condition for the $\bar{\partial}$-operator defined by the connection, acting on sections of the associated vector bundle $E \to S$. Thus, in the presence of the connection, the bundle $E$ is naturally a holomorphic bundle over $S$.

The Ward correspondence (Ward 1877) builds on this idea to give a complete translation of the instanton equations over certain Riemannian 4-manifolds into holomorphic geometry. In the simplest case, let $A$ be an instanton on a bundle over $\boldsymbol{R}^4$. Then, for any choice of a linear complex structure on $\boldsymbol{R}^4$, compatible with the metric, $A$ defines a holomorphic structure. The choices of such a complex structure are parametrized by a 2-sphere; in fact, the unit sphere in $\Lambda^+(\boldsymbol{R}^4)$. So, for any $\lambda \in S^2$ we have a complex surface $S_\lambda$ and a holomorphic bundle over $S_\lambda$. These data can be viewed in the following way. We consider the projection $\pi : \boldsymbol{R}^4 \times S^2 \to \boldsymbol{R}^4$ and the pull-back $\pi^*(E)$ to $\boldsymbol{R}^4 \times S^2$. This pullback bundle has a connection which defines a holomorphic structure along each fiber $S_\lambda \subset \boldsymbol{R}^4 \times S^2$ of the other projection. The product $\boldsymbol{R}^4 \times S^2$ is the twistor space of $\boldsymbol{R}^4$ and it is in a natural way a three-dimensional complex manifold. It can be identified with the complement of a line $L_\infty$ in $\mathbf{CP}^3$ where the projection $\boldsymbol{R}^4 \times S^2 \to S^2$ becomes the fibration of $\mathbf{CP}^3 \setminus L_\infty$ by the complex planes through $L_\infty$. One can see then that $\pi^*(E)$ is naturally a holomorphic bundle over $\mathbf{CP}^3 \setminus L_\infty$. The construction extends to the conformal compactification $S^4$ of $\boldsymbol{R}^4$. If $S^4$ is viewed as the quaternionic projective line $\mathbf{HP}^1$ and we identify $\boldsymbol{H}^2$ with $\mathbf{C}^4$ in the standard way, we get a natural map $\pi : \mathbf{CP}^3 \to \mathbf{HP}^1$. Then $\mathbf{CP}^3$ is the twistor space of $S^4$ and an anti-self-dual instanton on a bundle $E$ over $S^4$ induces a holomorphic structure on the bundle $\pi^*(E)$ over $\mathbf{CP}^3$.

In general, the twistor space $Z$ of an oriented Riemannian 4-manifold $M$ is defined to be the unit sphere bundle in $\Lambda^+_M$. This has a natural almost-complex structure which is integrable if and only if the self-dual part of the Weyl curvature of $M$ vanishes (Atiyah *et al.* 1978). The antipodal map on the 2-sphere induces an antiholomorphic involution of $Z$. In such a case, an anti-self-dual instanton over $M$ lifts to a holomorphic bundle over $Z$. Conversely, a holomorphic bundle over $Z$ which is holomorphically trivial over the fibers of the fibration $Z \to M$ (projective lines in $Z$), and which satisfies a certain reality condition with respect to the antipodal map, arises from a unitary instanton over $M$. This is the Ward correspondence, part of Penrose's twistor theory.

### The ADHM Construction

The problem of describing all solutions to the Yang–Mills instanton equation over $S^4$ is thus reduced to a problem in algebraic geometry, of classifying certain holomorphic vector bundles. This was solved by Atiyah *et al.* (1978). The resulting ADHM construction reduces the problem to certain matrix equations. The equations can be reduced to the following form. For a bundle Chern class $k$ and rank $r$, we require a pair of $k \times k$ matrices $\alpha_1, \alpha_2$, a $k \times r$ matrix $a$, and an $r \times k$ matrix $b$. Then the equations are

$$\begin{aligned} [\alpha_1, \alpha_2] &= ab \\ [\alpha_1^*, \alpha_1] + [\alpha_2^*, \alpha_2] &= aa^* - b^*b \end{aligned} \qquad [1]$$

We also require certain open, nondegeneracy conditions. Given such matrix data, a holomorphic

bundle over $\mathbf{CP}^3$ is constructed via a "monad": a pair of bundle maps over $\mathbf{CP}^3$

$$\boldsymbol{C}^k \otimes \mathcal{O}(-1)@>D_1>>\boldsymbol{C}^k \oplus \boldsymbol{C}^k \oplus \boldsymbol{C}^r@>D_2>>\boldsymbol{C}^k \otimes \mathcal{O}(1)$$

with $D_2 D_1 = 0$. That is, the rank-$r$ holomorphic bundle we construct is $\operatorname{Ker} D_2/\operatorname{Im} D_1$. The bundle maps $D_1, D_2$ are obtained from the matrix data in a straightforward way, in suitable coordinates. It is this matrix description which was used by Boyer *et al.* to prove the Atiyah–Jones conjecture on the topology of the moduli spaces of instantons. The only other case when the twistor space of a compact 4-manifold is an algebraic variety is the complex projective plane, with the nonstandard orientation. An analog of the ADHM description in this case was given by Buchdahl (1986).

### Integrable Systems

The Ward correspondence can be viewed in the general framework of integrable systems. Working with the standard complex structure on $\boldsymbol{R}^4$, the integrability condition for the $\bar{\partial}$-operator takes the shape

$$[\nabla_1 + \mathrm{i}\nabla_2, \nabla_3 + \mathrm{i}\nabla_4] = 0$$

where $\nabla_i$ are the components of the covariant derivative in the coordinate directions. So, the instanton equation can be viewed as a family of such commutator equations parametrized by $\lambda \in S^2$. One obtains many reductions of the instanton equation by imposing suitable symmetries. Solutions invariant under translation in one variable correspond to the Bogomolny "monopole equation" (Jaffe and Taubes 1980). Solutions invariant under three translations correspond to solutions of Nahm's equations,

$$\frac{\mathrm{d}T_i}{\mathrm{d}t} = \epsilon_{ijk}[T_j, T_k]$$

for matrix-valued functions $T_1, T_2, T_3$ of one variable $t$. Nahm (1982) and Hitchin (1983) developed an analog of the ADHM construction relating these two equations. This is now seen as a part of a general "Fourier–Mukai–Nahm transform" (Donaldson and Kronheimer 1990). The instanton equations for connections invariant under two translations, Hitchin's equations (Hitchin 1983), are locally equivalent to the harmonic map equation for a surface into the symmetric space dual to the structure group. Changing the signature of the metric on $\boldsymbol{R}^4$ to (2, 2), one gets the harmonic mapping equations into Lie groups (Hitchin 1990). More complicated reductions yield almost all the known examples of integrable PDEs as special forms of the instanton equations (Mason and Woodhouse 1996).

### Moment Maps: the Kobayashi–Hitchin Conjecture

Let $\Sigma$ be a compact Riemann surface. The Jacobian of $\Sigma$ is the complex torus $H^1(\Sigma, \mathcal{O})/H^1(\Sigma, \boldsymbol{Z})$: it parametrizes holomorphic line bundles of degree 0 over $\Sigma$. The Hodge theory (which was, of course, developed long before Hodge in this case) shows that the Jacobian can also be identified with the torus $H^1(\Sigma, \boldsymbol{R})/H^1(\Sigma, \boldsymbol{Z})$ which parametrizes flat U(1)-connections. That is, any holomorphic line bundle of degree 0 admits a unique compatible flat unitary connection.

The generalization of these ideas to bundles of higher rank began with Weil. He observed that any holomorphic vector bundle of degree 0 admits a flat connection, not necessarily unitary. Narasimhan and Seshadri (1965) showed that (in the case of degree 0) the existence of a flat, irreducible, unitary connection was equivalent to an algebro-geometric condition of stability which had been introduced shortly before by Mumford, for quite different purposes. Mumford introduced the stability condition in order to construct separated moduli spaces of holomorphic bundles – generalizing the Jacobian – as part of his general geometric invariant theory. For bundles of nonzero degree, the discussion is slightly modified by the use of projectively flat unitary connections. The result of Narasimhan and Seshadri asserts that there are two different descriptions of the same moduli space $\mathcal{M}^{d,r}$(Sigma): either as parametrizing certain irreducible projectively flat unitary connections (representations of $\pi_1(\Sigma)$), or parametrizing stable holomorphic bundles of degree $d$ and rank $r$. While Narasimhan and Seshadri probably did not view the ideas in these terms, another formulation of their result is that a certain nonlinear PDE for a Hermitian metric on a holomorphic bundle – analogous to the Laplace equation in the abelian case – has a solution when the bundle is stable.

Atiyah and Bott (1982) cast these results in the framework of gauge theory. (The Yang–Mills equations in two dimensions essentially reduce to the condition that the connection be flat, so they are rather trivial locally but have interesting global structure.) They made the important observation that the curvature of a connection furnishes a map

$$F : \mathcal{A} \to \mathrm{Lie}(\mathcal{G})^*$$

which is an equivariant moment map for the action of the gauge group on $\mathcal{A}$. Here the symplectic form on the affine space $\mathcal{A}$ and the map from the adjoint

bundle-valued 2-forms to the dual of the Lie algebra of $\mathcal{G}$ are both given by integration of products of forms. From this point of view, the Narasimhan–Seshadri result is an infinite-dimensional example of a general principle relating symplectic and complex quotients. At about the same time, Hitchin and Kobayashi independently proposed an extension of these ideas to higher dimensions. Let $E$ be a holomorphic bundle over a complex manifold $V$. Any compatible unitary connection on $E$ has curvature $F$ of type (1,1). Let $\omega$ be the (1,1)-form corresponding to a fixed Hermitian metric on $V$. The Hermitian Yang–Mills equation is the equation

$$F \cdot \omega = \mu 1_E$$

where $\mu$ is a constant (determined by the topological invariant $c_1(E)$). The Kobayashi–Hitchin conjecture is that, when $\omega$ is Kähler, this equation has an irreducible solution if and only if $E$ is a stable bundle in the sense of Mumford. Just as in the Riemann surface case, this equation can be viewed as a nonlinear second-order PDE of Laplace type for a metric on $E$. The moment map picture of Atiyah and Bott also extends to this higher-dimensional version. In the case when $V$ has complex dimension 2 (and $\mu$ is zero), the Hermitian Yang–Mills connections are exactly the anti-self-dual instantons, so the conjecture asserts that the moduli spaces of instantons can be identified with certain moduli spaces of stable holomorphic bundles.

The Kobayashi–Hitchin conjecture was proved in the most general form by Uhlenbeck and Yau (1986), and in the case of algebraic manifolds in Donaldson (1987). The proofs in Donaldson (1985, 1987) developed some extra structure surrounding these equations, connected with the moment map point of view. The equations can be obtained as the Euler–Lagrange equations for a nonlocal functional, related to the renormalized determinants of Quillen and Bismut. The results have been extended to non-Kähler manifolds and certain noncompact manifolds. There are also many extensions to equations for systems of data comprising a bundle with additional structure such as a holomorphic section or Higgs' field (Bradlow *et al.* 1995), or a parabolic structure along a divisor. Hitchin's equations (Hitchin 1987) are a particularly rich example.

## Topology of Moduli Spaces

The moduli spaces $\mathcal{M}_{r,d}(\Sigma)$ of stable holomorphic bundles/projectively flat unitary connections over Riemann surfaces $\Sigma$ have been studied intensively from many points of view. They have natural Kähler structures: the complex structure being visible in the holomorphic bundles guise and the symplectic form as the "Marsden–Weinstein quotient" in the unitary connections guise. In the case when $r$ and $d$ are coprime, they are compact manifolds with complicated topologies. There is an important basic construction for producing cohomology classes over these (and other) moduli spaces. One takes a universal bundle $U$ over the product $\mathcal{M} \times \Sigma$ with Chern classes

$$c_i(U) \in H^{2i}(\mathcal{M} \times \Sigma)$$

Then, for any class $\alpha \in H_p(\Sigma)$, we get a cohomology class $c_i(U)/\alpha \in H^{2i-p}(\mathcal{M})$. Thus, if $R_\Sigma$ is the graded ring freely generated by such classes, we have a homomorphism $\nu : R_\Sigma \to H^*(\mathcal{M})$. The questions about the topology of the moduli spaces which have been studied include:

1. finding the Betti numbers of the moduli space $\mathcal{M}$;
2. identifying the kernel of $\nu$;
3. giving an explicit system of generators and relations for the ring $H^*(\mathcal{M})$;
4. identifying the Pontrayagin and Chern classes of $\mathcal{M}$ within $H^*(\mathcal{M})$; and
5. evaluating the pairings

$$\int_{\mathcal{M}} \nu(W)$$

for elements $W$ of the appropriate degree in $R$.

All of these questions have now been solved quite satisfactorily. In early work, Newstead (1967) found the Betti numbers in the rank-2 case. The main aim of Atiyah and Bott was to apply the ideas of Morse theory to the Yang–Mills functional over a Riemann surface and they were able to reproduce Newstead's results in this way and extend them to higher rank. They also showed that the map $\nu$ is a surjection, so the universal bundle construction gives a system of generators for the cohomology. Newstead made conjectures on the vanishing of the Pontrayagin and Chern classes above a certain range which were established by Kirwan and extended to higher rank by Earl and Kirwan (1999). Knowing that $R_\Sigma$ maps on to $H^*(M)$, a full set of relations can (by Poincaré duality) be deduced in principle from a knowledge of the integral pairings in (5) above, but this is not very explicit. A solution to (5) in the case of rank 2 was found by Thaddeus (1992). He used results from the Verlinde theory (see section on 3-manifolds below) and the Riemann–Roch formula. Another point of view was developed by Witten (1991), who showed that the volume of the moduli space was related to the theory of torsion in algebraic topology and satisfied simple gluing axioms. These different

points of view are compared in Donaldson (1993). Using a nonrigorous localization principle in infinite dimensions, Witten (1992) wrote down a general formula for the pairings (5) in any rank, and this was established rigorously by Jeffrey and Kirwan, using a finite-dimensional version of the same localization method. A very simple and explicit set of generators and relations for the cohomology (in the rank-2 case) was given by King and Newstead (1998). Finally, the quantum cohomology of the moduli space, in the rank-2 case, was identified explicitly by Munoz (1999).

### Hyper-Kähler Quotients

Much of this story about the structure of moduli spaces extends to higher dimensions and to the moduli spaces of connections and Higgs fields. A particularly notable extension of the ideas involves hyper-Kähler structures. Let $M$ be a hyper-Kähler 4-manifold, so there are three covariant-constant self-dual forms $\omega_1, \omega_2, \omega_3$ on $M$. These correspond to three complex structures $I_1, I_2, I_3$ obeying the algebra of the quaternions. If we single out one structure, say $I_1$, the instantons on $M$ can be viewed as holomorphic bundles with respect to $I_1$ satisfying the moment map condition (Hermitian Yang–Mills equation) defined by the form $\omega_1$. Taking a different complex structure interchanges the role of the moment map and integrability conditions. This can be put in a general framework of hyper-Kähler quotients due to Hitchin *et al.* (1987). Suppose initially that $M$ is compact (so either a K3 surface or a torus). Then the $\omega_i$ components of the curvature define three maps

$$F_i : \mathcal{A} \to \mathrm{Lie}(\mathcal{G})^*$$

The structures on $M$ make $\mathcal{A}$ into a flat hyper-Kähler manifold and the three maps $F_i$ are the moment maps for the gauge group action with respect to the three symplectic forms on $\mathcal{A}$. In this situation, it is a general fact that the hyper-Kähler quotient – the quotient by $\mathcal{G}$ of the common zero set of the three moment maps – has a natural hyper-Kähler structure. This hyper-Kähler quotient is just the moduli space of instantons over $M$. In the case when $M$ is the noncompact manifold $\boldsymbol{R}^4$, the same ideas apply except that one has to work with the based gauge group $\mathcal{G}_0$. The conclusion is that the framed moduli spaces $\tilde{\mathcal{M}}$ of instantons over $\boldsymbol{R}^4$ are naturally hyper-Kähler manifolds. One can also see this hyper-Kähler structure through the ADHM matrix description. A variant of these matrix equations was used by Kronheimer to construct "gravitational instantons." The same ideas also apply to the moduli spaces of monopoles, where the hyper-Kähler metric, in the simplest case, was studied by Atiyah and Hitchin (1989).

## Low-Dimensional Topology

### Instantons and 4-Manifolds

Gauge theory has had unexpected applications in low-dimensional topology, particularly the topology of smooth 4-manifolds. The first work in this direction, in the early 1980s, involved the Yang–Mills instantons. The main issue in 4-manifold theory at that time was the correspondence between the diffeomorphism classification of simply connected 4-manifolds and the classification up to homotopy. The latter is determined by the intersection form, a unimodular quadratic form on the second integral homology group (i.e., a symmetric matrix with integral entries and determinant $\pm 1$, determined up to integral change of basis). The only known restriction was that Rohlin's theorem, which asserts that if the form is even the signature must be divisible by 16. The achievement of the first phase of the theory was to show that

1. There are unimodular forms which satisfy the hypotheses of Rohlin's theorem but which do not appear as the intersection forms of smooth 4-manifolds. In fact, no nonstandard definite form, such as a sum of copies of the $E_8$ matrix, can arise in this way.
2. There are simply connected smooth 4-manifolds which have isomorphic intersection forms, and hence are homotopy equivalent, but which are not diffeomorphic.

These results stand in contrast to the homeomorphism classification which was obtained by Freedman shortly before and which is almost the same as the homotopy classification.

The original proof of item (1) above argued with the moduli space $\mathcal{M}$ of anti-self-dual instantons SU(2) instantons on a bundle with $c_2 = 1$ over a simply connected Riemannian 4-manifold $M$ with a negative-definite intersection form (Donaldson 1983). In the model case when $M$ is the 4-sphere the moduli space $\mathcal{M}$ can be identified explicitly with the open 5-ball. Thus the 4-sphere arises as the natural boundary of the moduli space. A sequence of points in the moduli space converging to a boundary point corresponds to a sequence of connections with curvature densities converging to a $\delta$-function, as described earlier. One shows that in the general case (under our hypotheses on the 4-manifold $M$) the moduli space $\mathcal{M}$ has a similar behavior, it contains a collar $M \times (0, \delta)$

formed by instantons made using Taubes' gluing construction, described previously. The complement of this collar is compact. In the interior of the moduli space, there are a finite number of special points corresponding to U(1)-reductions of the bundle $P$. This is the way in which the moduli space "sees" the integral structure of the intersection form since such reductions correspond to integral homology classes with self-intersection –1. Neighborhoods of these special points are modeled on quotients $\mathbf{C}^3/\mathrm{U}(1)$; that is, cones on copies of $\mathbf{CP}^2$. The upshot is that (for generic Riemannian metrics on $M$) the moduli space gives a cobordism from the manifold $M$ to a set of copies of $\mathbf{CP}^2$ which can be counted in terms of the intersection form, and the result follows easily from standard topology. More sophisticated versions of the argument extended the results to rule out some indefinite intersection forms.

On the other hand, the original proofs of item (2) used "invariants" defined by instanton moduli spaces (Donaldson 1990). The general scheme exploits the same construction outlined in the previous section. We suppose that $M$ is a simply connected 4-manifold with $b^+(M) = 1 + 2p$, where $p > 0$ is an integer. (Here $b^+(M)$ is, as usual, the number of positive eigenvalues of the intersection matrix.) Ignoring some technical restrictions, there is a map

$$\nu : R_M \to H^*(\mathcal{M}_k)$$

where $R_M$ is a graded ring freely generated by the homology (below the top dimension) of the 4-manifold $M$ and $\mathcal{M}_k$ is the moduli space of anti-self-dual SU(2)-instantons on a bundle with $c_2 = k > 0$. For an element $W$ in $R_M$ of the appropriate degree, one obtains a number by evaluating, or integrating, $\nu(W)$ on $\mathcal{M}_k$. The main technical difficulty here is that the moduli space $\mathcal{M}_k$ is rarely compact, so one needs to make sense of this "evaluation." With all the appropriate technicalities in place, these invariants could be shown to distinguish various homotopy-equivalent, homeomorphic 4-manifolds. All these early developments are described in detail in the book by Donaldson and Kronheimer (1990).

### Basic Classes

Until the early 1990s, these instanton invariants could only be calculated in isolated favorable cases (although the calculations which were made, through the work of many mathematicians, led to a large number of further results about 4-manifold topology). Deeper understanding of their structure came with the work of Kronheimer and Mrowka. This work was, in large part, motivated by a natural question in geometric topology. Any homology class $\alpha \in H_2(M; \mathbf{Z})$ can be represented by an embedded, connected, smooth surface. One can define an integer $\underline{g}(\alpha)$ to be the minimal genus of such a representative. The problem is to find $\underline{g}(\alpha)$, or at least bounds on it. A well-known conjecture, ascribed to Thom, was that when $M$ is the complex projective plane the minimal genus is realized by a complex curve; that is,

$$\underline{g}(\pm dH) = \tfrac{1}{2}(d-1)(d-2)$$

where $H$ is the standard generator of $H_2(\mathbf{CP}^2)$ and $d \geq 1$.

The new geometrical idea introduced by Kronheimer and Mrowka was to study instantons over a 4-manifold $M$ with singularities along a surface $\Sigma \subset M$. For such connections, there is a real parameter: the limit of the trace of the holonomy around small circles linking the surface. By varying this parameter, they were able to interpolate between moduli spaces of nonsingular instantons on different bundles over $M$ and obtain relations between the different invariants. They also found that if the genus of $\Sigma$ is suitably small then some of the invariants are forced to vanish, thus, conversely, getting information about $\underline{g}$ for 4-manifolds with nontrivial invariants. For example, they showed that if $M$ is a K3 surface then $\underline{g}(\alpha) = (1/2)(\alpha \cdot \alpha + 2)$.

The structural results of Kronheimer and Mrowka (1995) introduced the notion of a 4-manifold of "simple type." Write the invariant defined above by the moduli space $\mathcal{M}_k$ as $I_k : R_M \to \mathbf{Q}$. Then $I_k$ vanishes except on terms of degree $2d(k)$, where $d(k) = 4k - 3(1+p)$. We can put all these together to define $\underline{I} = \sum I_k : R_M \to \mathbf{Q}$. The ring $R_M$ is a polynomial ring generated by classes $\alpha \in H_2(M)$, which have degree 2 in $R_M$, and a class $X$ of degree 4 in $R_M$, corresponding to the generator of $H_0(M)$. The 4-manifold is of simple type if

$$\underline{I}(X^2 W) = 4\underline{I}(W)$$

for all $W \in R_M$. Under this condition, Kronheimer and Mrowka showed that all the invariants are determined by a finite set of "basic" classes $K_1, \ldots, K_s \in H_2(M)$ and rational numbers $\beta_1, \ldots, \beta_s$. To express the relation, they form a generating function

$$\mathcal{D}_M(\alpha) = \underline{I}(\mathrm{e}^{\alpha}) + \underline{I}\left(\frac{X}{2}\mathrm{e}^{\alpha}\right)$$

This is *a priori* a formal power series in $H^2(M)$ but *a posteriori* the series converges and can be regarded

as a function on $H_2(M)$. Kronheimer and Mrowka's result is that

$$\mathcal{D}_M(\alpha) = \exp\left(\frac{\alpha \cdot \alpha}{2}\right) \sum_{r=1}^{s} \beta_r e^{K_r \cdot \alpha}$$

It is not known whether all simply connected 4-manifolds are of simple type, but Kronheimer and Mrowka were able to show that this is the case for a multitude of examples. They also introduced a weaker notion of "finite type," and this condition was shown to hold in general by Munoz and Froyshov. The overall result of this work of Kronheimer and Mrowka was to make the calculation of the instanton invariants for many familiar 4-manifolds a comparatively straightforward matter.

### 3-Manifolds: Casson's Invariant

Gauge theory has also entered into 3-manifold topology. In 1985, Casson introduced a new integer-valued invariant of oriented homology 3-spheres which "counts" the set $Z$ of equivalence classes of irreducible flat SU(2)-connections, or equivalently irreducible representations $\pi_1(Y) \to$ SU(2). Casson's approach (Akbulut and McCarthy 1990) was to use a Heegard splitting of a 3-manifold $Y$ into two handle bodies $Y^+, Y^-$ with a surface $\Sigma$ as common boundary. Then $\pi_1(\Sigma)$ maps onto $\pi_1(Y)$ and a flat SU(2) connection on $Y$ is determined by its restriction to $\Sigma$. Let $M_\Sigma$ be the moduli space of irreducible flat connections over $\Sigma$ (as discussed in the last section) and let $L^\pm \subset M_\Sigma$ be the subsets which extend over $Y^\pm$. Then $L^\pm$ are submanifolds of half the dimension of $\mathcal{M}_\Sigma$ and the set $Z$ can be identified with the intersection $L^+ \cap L^-$. The Casson invariant is one-half the algebraic intersection number of $L^+$ and $L^-$. Casson showed that this is independent of the Heegard splitting (and is also, in fact, an integer, although this is not obvious). He showed that when $Y$ is changed by Dehn surgery along a knot, the invariant changes by a term computed from the Alexander polynomial of the knot. This makes the Casson invariant computable in examples. (For a discussion of Casson's formula see Donaldson (1999).) Taubes showed that the Casson invariant could also be obtained in a more differential-geometric fashion, analogous to the instanton invariants of 4-manifolds (Taubes 1990).

### 3-Manifolds: Floer Theory

Independently, at about the same time, Floer (1989) introduced more sophisticated invariants – the Floer homology groups – of homology 3-spheres, using gauge theory. This development ran parallel to his introduction of similar ideas in symplectic geometry. Suppose, for simplicity, that the set $Z$ of equivalence classes of irreducible flat connections is finite. For pairs $\rho_-, \rho_+$ in $Z$, Floer considered the instantons on the tube $Y \times \boldsymbol{R}$ asymptotic to $\rho^\pm$ at $\pm\infty$. There is an infinite set of moduli spaces of such instantons, labeled by a relative Chern class, but the dimensions of these moduli spaces agree modulo 8. This gives a relative index $\delta(\rho_-, \rho_+) \in \boldsymbol{Z}/8$. If $\delta(\rho_-, \rho_+) = 1$ there is a moduli space of dimension 1 (possibly empty), but the translations of the tube act on this moduli space and, dividing by translations, we get a finite set. The number of points in this set, counted with suitable signs, gives an integer $n(\rho_-, \rho_+)$. Then, Floer considers the free abelian groups

$$C_* = \bigoplus_{\rho \in Z} \boldsymbol{Z}\langle\rho\rangle$$

generated by the set $Z$ and a map $\partial : C_* \to C_*$ defined by

$$\partial(\langle\rho_-\rangle) = \sum n(\rho_-, \rho_+)\langle\rho_+\rangle$$

Here the sum runs over the $\rho_+$ with $\delta(\rho_-, \rho_+) = 1$. Floer showed that $\partial^2 = 0$ and the homology $HF_*(Y) = \ker\partial/\mathrm{Im}\,\partial$ is independent of the metric on $Y$ (and various other choices made in implementing the construction in detail). The chain complex $C_*$ and hence the Floer homology can be graded by $\boldsymbol{Z}/8$, using the relative index, so the upshot is to define 8 abelian groups $HF_i(Y)$: invariants of the 3-manifold $Y$. The Casson invariant appears now as the Euler characteristic of the Floer homology. There has been extensive work on extending these ideas to other 3-manifolds (not homology spheres) and gauge groups, but this line of research does not yet seem to have reached a clear-cut conclusion.

Part of the motivation for Floer's work came from Morse theory, and particularly the approach to this theory expounded by Witten (1982). The Chern–Simons functional is a map

$$CS : \mathcal{A}/\mathcal{G} \to \boldsymbol{R}/\boldsymbol{Z}$$

from the space of SU(2)-connections over $Y$. Explicitly, in a trivialization of the bundle

$$\mathrm{CS}(A) = \int_Y A \wedge dA + \tfrac{3}{2} A \wedge A \wedge A$$

It appears as a boundary term in the Chern–Weil theory for the second Chern class, in a similar way as holonomy appears as a boundary term in the Gauss–Bonnet theorem. The set $Z$ can be identified with the critical points of CS and the instantons on the tube as integral curves of the gradient vector

field of CS. Floer's definition mimics the definition of homology in ordinary Morse theory, taking Witten's point of view. It can be regarded formally as the "middle-dimensional" homology of the infinite-dimensional space $\mathcal{A}/\mathcal{G}$. See Atiyah (1988) and Cohen *et al.* (1995) for discussions of these ideas.

The Floer theory interacts with 4-manifold invariants, making up a structure approximating to a $(3+1)$-dimensional topological field theory (Atiyah 1988). Roughly, the numerical invariants of closed 4-manifolds generalize to invariants for a 4-manifold $M$ with boundary $Y$ taking values in the Floer homology of $Y$. If two such manifolds are glued along a common boundary, the invariants of the result are obtained by a pairing in the Floer groups. There are, however, at the moment, some substantial technical restrictions on this picture. This theory, as well as Floer's original construction, is developed in detail by Donaldson (2002). At the time of writing, the Floer homology groups are still difficult to compute in examples. One important tool is a surgery-exact sequence found by Floer (Braam and Donaldson 1995), related to Casson's surgery formula.

### 3-Manifolds: Jones–Witten Theory

There is another, quite different, way in which ideas from gauge theory have entered 3-manifold topology. This is the Jones–Witten theory of knot and 3-manifold invariants. This theory falls outside the main line of this article, but we will say a little about it since it draws on many of the ideas we have discussed. The goal of the theory is to construct a family of $(2+1)$-dimensional topological field theories indexed by an integer $k$, assigning complex vector space $H_k(\Sigma)$ to a surface $\Sigma$ and an invariant in $H_k(\partial Y)$ to a 3-manifold-with-boundary $Y$. If $\partial Y$ is empty, the vector space $H_k(\partial Y)$ is taken to be $\mathbf{C}$, so one seeks numerical invariants of closed 3-manifolds. Witten's (1989) idea is that these invariants of closed 3-manifolds are Feynmann integrals

$$\int_{\mathcal{A}/\mathcal{G}} e^{i2\pi k \mathrm{CS}(A)} \mathcal{D}A$$

This functional integral is probably a schematic rather than a rigorous notion. The data associated with surfaces can, however, be defined rigorously. If we fix a complex structure $I$ on $\Sigma$, we can define a vector space $H_k(\Sigma, I)$ to be

$$H_k(\Sigma, I) = H^0(\mathcal{M}(\Sigma); L^k)$$

where $\mathcal{M}(\Sigma)$ is the moduli space of stable holomorphic bundles/flat unitary connections over $\Sigma$ and $L$ is a certain holomorphic line bundle over $\mathcal{M}(\Sigma)$. These are the spaces of "conformal blocks" whose dimension is given by the Verlinde formulas. Recall that $\mathcal{M}(\Sigma)$, as a symplectic manifold, is canonically associated with the surface $\Sigma$, without any choice of complex structure. The Hilbert spaces $H_k(\Sigma, I)$ can be regarded as the quantization of this symplectic manifold, in the general framework of geometric quantization: the inverse of $k$ plays the role of Planck's constant. What is not obvious is that this quantization is independent of the complex structure chosen on the Riemann surface: that is, that there is a natural identification of the vector spaces (or at least the associated projective spaces) formed by using different complex structures. This was established rigorously by Hitchin (1990) and Axelrod *et al.* (1991), who constructed a projectively flat connection on the bundle of spaces $H_k(\Sigma, I)$ over the space of complex structures $I$ on $\Sigma$. At a formal level, these constructions are derived from the construction of the metaplectic representation of a linear symplectic group, since the $\mathcal{M}_\Sigma$ are symplectic quotients of an affine symplectic space.

The Jones–Witten invariants have been rigorously established by indirect means, but it seems that there is still work to be done in developing Witten's point of view. If $Y^+$ is a 3-manifold with boundary, one would like to have a geometric definition of a vector in $H_k(\partial Y^+)$. This should be the quantized version of the submanifold $L^+$ (which is Lagrangian in $\mathcal{M}_\Sigma$) entering into the Casson theory.

### Seiberg–Witten Invariants

The instanton invariants of a 4-manifold can be regarded as the integrals of certain natural differential forms over the moduli spaces of instantons. Witten (1988) showed that these invariants could be obtained as functional integrals, involving a variant of the Feynman integral, over the space of connections and certain auxiliary fields (insofar as this latter integral is defined at all). A geometric explanation of Witten's construction was given by Atiyah and Jeffrey (1990). Developing this point of view, Witten made a series of predictions about the instanton invariants, many of which were subsequently verified by other means. This line of work culminated in 1994 where, applying developments in supersymmetric Yang–Mills QFT, Seiberg and Witten introduced a new system of invariants and a precise prediction as to how these should be related to the earlier ones.

The Seiberg–Witten invariants (Witten 1994) are associated with a $\mathrm{Spin}^c$ structure on a 4-manifold $M$. If $M$ is simply connected this is specified by a class $K \in H^2(M; \mathbf{Z})$ lifting $w_2(M)$. One has spin bundles

$S^+, S^- \to M$ with $c_1(S^\pm) = K$. The Seiberg–Witten equation is for a spinor field $\phi$ – a section of $S^+$ and a connection $A$ on the complex line bundle $\Lambda^2 S^+$. This gives a connection on $S^+$ and hence a Dirac operator

$$D_A : \Gamma(S^+) \to \Gamma(S^-)$$

The Seiberg–Witten equations are

$$D_A\phi = 0, \qquad F_A^+ = \sigma(\phi)$$

where $\sigma : S^+ \to \Lambda^+$ is a certain natural quadratic map. The crucial differential-geometric feature of these equations arises from the Weitzenbock formula

$$D_A^* D_A \phi = \nabla_A^* \nabla_A \phi + \frac{R}{4}\phi + \rho(F^+)\phi$$

where $R$ is the scalar curvature and $\rho$ is a natural map from $\Lambda^+$ to the endomorphisms of $S^+$. Then $\rho$ is adjoint to $\sigma$ and

$$\langle \rho(\sigma(\phi))\phi, \phi\rangle = |\phi|^4$$

It follows easily from this that the moduli space of solutions to the Seiberg–Witten equation is compact. The most important invariants arise when $K$ is chosen so that

$$K \cdot K = 2\chi(M) + 3\,\mathrm{sign}(M)$$

where $\chi(M)$ is the Euler characteristic and sign$(M)$ is the signature. (This is just the condition for $K$ to correspond to an almost-complex structure on $M$.) In this case, the moduli space of solutions is zero dimensional (after generic perturbation) and the Seiberg–Witten invariant SW$(K)$ is the number of points in the moduli space, counted with suitable signs.

Witten's conjecture relating the invariants, in its simplest form, is that when $M$ has simple type the classes $K$ for which SW$(K)$ is nonzero are exactly the basic classes $K_r$ of Kronheimer and Mrowka and that

$$\beta_r = 2^{C(M)}\mathrm{SW}(K_r)$$

where $C(M) = 2 + (1/4)(7\chi(M) + 11\,\mathrm{sign}(M))$. This asserts that the two sets of invariants contain exactly the same information about the 4-manifold.

The evidence for this conjecture, via calculations of examples, is very strong. A somewhat weaker statement has been proved rigorously by Feehan and Leness (2003). They use an approach suggested by Pidstragatch and Tyurin, studying moduli spaces of solutions to a nonabelian version of the Seiberg–Witten equations. These contain both the instanton and abelian Seiberg–Witten moduli spaces, and the strategy is to relate the topology of these two sets by standard localization arguments. (This approach is related to ideas introduced by Thaddeus (1994) in the case of bundles over Riemann surfaces.) The serious technical difficulty in this approach stems from the lack of compactness of the nonabelian moduli spaces. The more general versions of Witten's conjecture (Moore and Witten 1997) (e.g., when $b^+(M) = 1$) contain very complicated formulas, involving modular forms, which presumably arise as contributions from the compactification of the moduli spaces.

## Applications

Regardless of the connection with the instanton theory, one can go ahead directly to apply the Seiberg–Witten invariants to 4-manifold topology, and this has been the main direction of research since the 1990s. The features of the Seiberg–Witten theory which have led to the most prominent developments are the following.

1. The reduction of the equations to two dimensions is very easy to understand. This has led to proofs of the Thom conjecture and wide-ranging generalizations (Ozsvath and Szabo 2000).
2. The Weitzenboch formula implies that, if $M$ has positive scalar curvature, then solutions to the Seiberg–Witten equations must have $\phi = 0$. This has led to important interactions with four-dimensional Riemannian geometry (Lebrun 1996).
3. In the case when $M$ is a symplectic manifold, there is a natural deformation of the Seiberg–Witten equations, discovered by Taubes (1996), who used it to show that the Seiberg–Witten invariants of $M$ are nontrivial. More generally, Taubes showed that for large values of the deformation parameter the solutions of the deformed equation localize around surfaces in the 4-manifold and used this to relate the Seiberg–Witten invariants to the Gromov theory of pseudoholomorphic curves. These results of Taubes have completely transformed the subject of four-dimensional symplectic geometry.

Bauer and Furuta (2004) have combined the Seiberg–Witten theory with more sophisticated algebraic topology to obtain further results about 4-manifolds. They consider the map from the space of connections and spinor fields defined by the formulas on the left-hand side of the equations. The general idea is to obtain invariants from the homotopy class of this map, under a suitable notion of homotopy. A technical complication arises from the gauge group action, but this can be reduced to the action of a single U(1). Ignoring this issue, Bauer and Furuta have obtained invariants in the stable homotopy groups $\lim_{N\to\infty} \pi_{N+r}(S^N)$, which reduce to the ordinary numerical invariants when $r = 1$. Using these invariants, they obtain results about connected sums of

4-manifolds, for which the ordinary invariants are trivial. Using refined cobordism invariants ideas, Furuta made great progress towards resolving the question of which intersection forms arise from smooth, simply connected 4-manifolds. A well-known conjecture is that, if such a manifold is spin, then the second Betti number satisfies

$$b_2(M) \geq \tfrac{11}{8}|\mathrm{sign}(M)|$$

Furuta (2001) proved that $b_2(M) \geq (10/8)|\mathrm{sign}(M)| + 2$.

An important and very recent achievement, bringing together many different lines of work, is the proof of "Property P" in 3-manifold topology by Kronheimer and Mrowka (2004). This asserts that one cannot obtain a homotopy sphere (counter-example to the Poincaré conjecture) by +1-surgery along a nontrivial knot in $S^3$. The proof uses work of Gabai and Eliashberg to show that the manifold obtained by 0-framed surgery is embedded in a symplectic 4-manifold; Taubes' results to show that the Seiberg–Witten invariants of this 4-manifold are nontrivial; Feehan and Leness' partial proof of Witten's conjecture to show that the same is true for the instanton invariants; and the gluing rule and Floer's exact sequence to show that the Floer homology of the +1-surgered manifold is nontrivial. It follows then from the definition of Floer homology that the fundamental group of this manifold is not trivial; in fact, it must have an irreducible representation in SU(2).

*See also:* Cotangent Bundle Reduction; Floer Homology; Gauge Theories from Strings; Gauge Theoretic Invariants of 4-Manifolds; Instantons: Topological Aspects; Knot Homologies; Moduli Spaces: An Introduction; Nonperturbative and Topological Aspects of Gauge Theory; Seiberg–Witten Theory; Topological Quantum Field Theory: Overview; Variational Techniques for Ginzburg–Landau Energies.

## Further Reading


Akbulut S and McCarthy J (1990) *Casson's Invariant for Homology 3-Spheres*. Princeton, NJ: Princeton University Press.

Atiyah (1988) New invariants for 3 and 4 dimensional manifolds. In: *The Mathematical Heritage of Hermann Weyl*, Proceedings of Symposia in Pure Mathematics, vol. 48, pp. 285–299. American Mathematical Society.

Atiyah MF (1988) *Topological Quantum Field Theories,* vol. 68, pp. 135–186. Math. Publ. IHES.

Atiyah MF and Bott R (1982) The Yang–Mills equations over Riemann surfaces. *Philosophical Transactions of the Royal Society of London, Series A* 308: 523–615.

Atiyah MF, Drinfeld V, Hitchin NJ, and Manin YuI (1978) Construction of instantons. *Physics Letters A* 65: 185–187.

Atiyah MF and Hitchin NJ (1989) *The Geometry and Dynamics of Magnetic Monopoles*. Princeton, NJ: Princeton University Press.

Atiyah MF, Hitchin NJ, and Singer IM (1978) Self-duality in four-dimensional Riemannian geometry. *Proceedings of the Royal Society of London, Series A* 362: 425–461.

Atiyah MF and Jeffrey L (1990) Topological Lagrangians and cohomology. *Journal of Geometry and Physics* 7: 119–136.

Atiyah MF and Jones JDS (1978) Topological aspects of Yang–Mills theory. *Communications in Mathematical Physics* 61: 97–118.

Axelrod S, Della Pietra S, and Witten E (1991) Geometric quantisation of Chern–Simons gauge theories. *Journal of Differential Geometry* 33: 787–902.

Bauer S and Furuta M, A stable cohomotopy refinement of the Seiberg–Witten invariants, I. *Inventiones Mathematicae* 155: 1–19.

Bourguignon J-P and Lawson HB (1981) Stability and isolation phenomena for Yang–Mills fields. *Communications in Mathematical Physics* 79: 189–230.

Boyer C, Mann B, Hurtubise J, and Milgram R (1993) The topology of instanton moduli spaces. I: the Atiyah–Jones conjecture. *Annals of Mathematics* 137: 561–609.

Braam PJ and Donaldson SK (1995) Floer's work on instanton homology, knots and surgery. In: Hofer *et al.* (eds.) *The Floer Memorial Volume*, vol. 133, Progress in Mathematics. Basle: Birkhäuser.

Bradlow S, Daskalopoulos G, Garcia-Prada O, and Wentworth R (1995) Stable augmented bundles over Riemann surfaces. In: Hitchin *et al.* (eds.) *Vector Bundles in Algebraic Geometry*, pp. 15–67. Cambridge: Cambridge University Press.

Buchdahl NP (1986) Instantons on $\mathbf{CP}^2$. *Journal of Differential Geometry* 24: 19–52.

Cohen RL, Jones JDS, and Segal GB (1995) Floer's infinite-dimensional Morse theory and homotopy theory. In: Hofer *et al.* (eds.) *The Floer Memorial Volume*, pp. 297–325. Basle: Birkhäuser.

Donaldson SK (1983) An application of gauge theory to four-dimensional topology. *Journal of Differential Geometry* 18: 279–315.

Donaldson SK (1984) Nahm's equations and the classification of monopoles. *Communications in Mathematical Physics* 96: 387–407.

Donaldson SK (1985) Anti-self-dual Yang–Mills connections on complex algrebraic surfaces and stable vector bundles. *Proceedings of the London Mathematical Society* 3: 1–26.

Donaldson SK (1987) Infinite determinants, stable bundles and curvature. *Duke Mathematical Journal* 54: 231–247.

Donaldson SK (1990) Polynomial invariants of smooth four-manifolds. *Topology* 29: 257–315.

Donaldson SK (1993) Gluing techniques in the cohomology of moduli spaces. In: Goldberg and Phillips (eds.) *Topological Methods in Modern Mathematics*, pp. 137–170. Houston: Publish or Perish.

Donaldson SK (1999) *Topological Field Theories and Formulae of Casson and Meng-Taubes, Geometry and Topology Monographs,* vol. 2, pp. 87–102.

Donaldson SK (2002) *Floer Homology Groups in Yang–Mills Theory*. Cambridge: Cambridge University Press.

Donaldson SK and Kronheimer PB (1990) *The geometry of four-manifolds*. Oxford: Oxford University Press.

Donaldson SK and Thomas R (1998) Gauge theory in higher dimensions. In: Hugget *et al.* (eds.) *The Geometric Universe*, pp. 31–47. Oxford: Oxford University Press.

Earl R and Kirwan FC (1999) Pontrayagin rings of moduli spaces of arbitrary rank bundles over Riemann surfaces. *Journal of London Mathematical Society* 60: 835–846.

Feehan PMN and Leness TG (2003) A general SO(3)-monopole cobordism formula relating Donaldson and Seiberg–Witten invariants, arXiv:math.DG/0203047.

Floer A (1989) An instanton invariant for 3-manifolds. *Communications in Mathematical Physics* 118: 215–240.

Furuta M (2001) Monopole equation and the 11/8 conjecture. *Mathematical Research Letters* 8: 279–291.

Hitchin NJ (1983) On the construction of monopoles. *Communications in Mathematical Physics* 89: 145–190.

Hitchin NJ (1987) The self-duality equations over Riemann surfaces. *Proceedings of the London Mathematical Society* 55: 59–126.
Hitchin NJ (1990) Flat connections and geometric quantisation. *Communications in Mathematical Physics* 131: 347–380.
Hitchin NJ (1990) Harmonic maps from the 2-torus to the 3-sphere. *Journal of Differential Geometry* 31: 627–710.
Hitchin NJ, Karhlede A, Lindström U, and Roceck M (1987) Hyper-Kähler metrics and supersymmetry. *Communications in Mathematical Physics* 108: 535–589.
Jaffe A and Taubes CH (1980) *Vortices and Monopoles*, Progress in Physics, vol. 2. Basle: Birkhäuser.
Jeffrey L and Kirwan FC (1998) Intersection theory on moduili spaces of holomorphic bundles of arbitrary rank on a Riemann surface. *Annals of Mathematics* 148: 109–196.
Joyce D (1996) Compact Riemannian 7-manifolds with holonomy $G_2$, I. *Journal of Differential Geometry* 43: 291–328.
King AD and Newstead PE (1998) On the cohomology ring of the moduli space of stable rank 2 bundles over a curve. *Topology* 37: 407–418.
Kirwan FC (1994) Geometric invariant theory and the Atiyah–Jones conjecture. In: Laudel and Jahren (eds.) *The Sophus Lie Memorial Volume Pages*, pp. 161–186. Oslo: Scandinavian University Press.
Kronheimer PB and Mrowka TS (1995) embedded surfaces and the structure of Donaldson's polynomial invariants. *Journal of Differential Geometry* 41: 573–734.
Kronheimer PB and Mrowka TS, Witten's conjecture and property P. *Geometry and Topology* 8: 295–300.
Lebrun C (1996) Four-manifolds without Einstein metrics. *Mathematical Research Letters* 3: 133–147.
Mason L and Woodhouse N (1996) *Integrability, Self-duality and Twistor Theory. Advances in Theoretical and Mathematical Physics*. Oxford: Oxford University Press.
Moore G and Witten E (1997) Integration over the u-plane in Donaldson Theory. *Adv. Theor. Mathematical Phys* 1: 298–387.
Munoz V (1999) Quantum cohomology of the moduli space of stable bundles over a Riemann surface. *Duke Mathematics Journal* 98: 525–540.
Nahm W (1982) The construction of all self-dual monopoles by the ADHM method. In: *Monopoles in Quantum Field Theory*, pp. 87–94. Singapore: World Scientific.
Nakajima H (1988) Compactness of the moduli space of Yang–Mills connections in higher dimensions. *Journal of the Mathematical Society of Japan* 40: 383–392.
Narasimhan MS and Seshadri CS (1965) Stable and unitary vector bundles on compact Riemann surfaces. *Annals of Mathematics* 65: 540–567.
Newstead PE (1967) Topology of some spaces of stable bundles. *Topology* 6: 241–262.
Ozsvath P and Szabo Z (2000) The sympletic Thom conjecture. *Annals of Mathematics* 151: 93–124.
Price P (1983) A monotonicity formula for Yang–Mills fields. *Manuscripta Mathematica* 43: 131–166.
Quillen D (1985) Determinants of Cauchy–Riemann operators over a Riemann surface. *Functional Analysis and its Applications* 14: 31–34.
Sadun L and Segert J (1992) Non self-dual Yang–Mills connections with quadropole symmetry. *Communications in Mathematical Physics* 145: 363–391.
Sibner L, Sibner R, and Uhlenbeck K (1989) Solutions to the Yang–Mills equations that are not self-dual. *Proceedings of the National Academy of Sciences of the USA* 86: 8610–8613.
Taubes CH (1982) Self-dual Yang–Mills connections over non self-dual 4-manifolds. *Journal of Differential Geometry* 17: 139–170.
Taubes CH (1983) Stability in Yang–Mills theories. *Communications in Mathematical Physics* 91: 235–263.
Taubes CH (1990) Casson's invariant and gauge theory. *Journal of Differential Geometry* 31: 363–430.
Taubes CH (1998) A framework for Morse Theory for the Yang–Mills functional. *Inventiones Mathematical* 94: 327–402.
Taubes CH (1996) SW ⇒ Gr: from the Seiberg–Witten equations to pseudo-holomorphic curves. *Journal of American Mathematical Society* 9: 819–918.
Thaddeus M (1992) Conformal Field Theory and the cohomology of moduli spaces of stable bundles. *Journal of Differential Geometry* 35: 131–149.
Thaddeues M (1994) Stable pairs, linear systems and the Verlinde formulae. *Inventiones Mathematicae* 117: 317–353.
Tian G (2000) gauge theory and calibrated geometry, I. *Annals of Mathematics* 151: 193–268.
Uhlenbeck KK (1982) Connections with $L^p$ bounds on curvature. *Communications in Mathematical Physics* 83: 11–29.
Uhlenbeck KK and Yau S-T (1986) On the existence of hermitian Yang–Mills connections on stable bundles over compact hyper-Kähler manifolds. *Communications on Pure and Applied Mathematics* 39: 257–293.
Ward RS (1877) On self-dual gauge fields. *Physics Letters A* 61: 81–82.
Witten E (1982) Supersymmetry and Morse Theory. *Journal of Differential Geometry* 16: 661–692.
Witten E (1988) Topological quantum field theory. *Communications in Mathematical Physics* 117: 353–386.
Witten E (1989) Quantum field theory and the Jones polynomial. *Communications in Mathematical Physics* 121: 351–399.
Witten E (1991) On quantum gauge theory in two dimensions. *Communications in Mathematical Physics* 141: 153–209.
Witten E (1992) Two dimensional quantum gauge theory revisited. *Journal of Geometry and Physics* 9: 303–368.
Witten E (1994) Monopoles and four-manifolds. *Mathematical Research Letters* 1: 769–796.

# General Relativity: Experimental Tests

**C M Will**, Washington University, St. Louis, MO, USA



## Introduction

Einstein's general theory of relativity has become the foundation for our understanding of the gravitational interaction. Four decades of high-precision experiments have verified the theory with ever-increasing precision, with no confirmed evidence of a deviation from its predictions. The theory is now the standard framework for much of astronomy, with its searches for black holes, neutron stars, gravitational waves, and the origin and fate of the universe.

Yet modern developments in particle theory suggest that it may not be the entire story, and that

modification of the basic theory may be required at some level. String theory generally predicts a proliferation of gravity-like fields that could result in alterations of general relativity (GR) reminiscent of the Brans–Dicke theory of the 1960s. In the presence of extra dimensions, the gravity of the four-dimensional "brane" of a higher-dimensional world could be somewhat different from a pure four-dimensional GR. However, any theoretical speculation along these lines must still abide by the best current empirical bounds. This article will review experimental tests of GR and the theoretical implications of the results.

## The Einstein Equivalence Principle

The Einstein equivalence principle is a modern generalization of Einstein's 1907 idea of an equivalence between gravity and acceleration, or between free fall and an absence of gravity. It states that: (1) test bodies fall with the same acceleration independently of their internal structure or composition (weak equivalence principle, or WEP); (2) the outcome of any local nongravitational experiment is independent of the velocity of the freely falling reference frame in which it is performed (local Lorentz invariance, or LLI); and (3) the outcome of any local nongravitational experiment is independent of where and when in the universe it is performed (local position invariance, or LPI).

This principle is fundamental to gravitational theory, for it is possible to argue that, if EEP is valid, then gravitation and geometry are synonymous. In other words, gravity must be described by a "metric theory of gravity," in which (1) spacetime is endowed with a symmetric metric, (2) the trajectories of freely falling bodies are geodesics of that metric, and (3) in local freely falling reference frames, the nongravitational laws of physics are those written in the language of special relativity (see Will (1993) for further details).

GR is a metric theory of gravity, but so are many others, including the scalar–tensor theory of Brans and Dicke and many of its modern descendents, some of which are inspired by string theory.

### Tests of the Weak Equivalence Principle

To test the WEP, one compares the acceleration of two laboratory-sized bodies of different composition in an external gravitational field. Although legend suggests that Galileo may have demonstrated this principle to his students at the Leaning Tower of Pisa, and Newton tested it by means of pendulum experiments, the first true high-precision experiments were done at the end of the nineteenth century by the Hungarian physicist Baron Roland von Eötvös and colleagues.

Eötvös employed a torsion balance, in which (schematically) two bodies of different composition are suspended at the ends of a rod that is supported horizontally by a fine wire or fiber. One then looks for a difference in the horizontal accelerations of the two bodies as revealed by a slight rotation of the rod. The source of the horizontal gravitational force could be the Sun, a large mass in or near the laboratory, or, as Eötvös recognized, the Earth itself. A measurement or limit on the fractional difference in acceleration between two bodies yields a quantity $\eta \equiv 2|a_1 - a_2|/|a_1 + a_2|$, called the "Eötvös ratio." Eötvös' experiments showed that $\eta$ was smaller than a few parts in $10^9$, and later classic experiments in the 1960s and 1970s by Dicke and Braginsky improved the bounds by several orders of magnitude. Additional experiments were carried out during the 1980s as part of a search for a putative "fifth force," that was motivated in part by a re-analysis of Eötvös' original data.

The best limit on $\eta$ currently comes from experiments carried out during the 1985–2000 period at the University of Washington (called the "Eöt-Wash" experiments), which used a sophisticated torsion balance tray to compare the accelerations of bodies of different composition toward the Earth, the Sun, and the galaxy. Another strong bound comes from ongoing laser ranging to reflectors deposited on the Moon during the Apollo program in the 1970s (lunar laser ranging, LLR), which routinely determines the Earth–Moon distance to millimeter accuracies. The data may be used to check the equality of acceleration of the Earth and Moon toward the Sun. The results from laboratory and LLR experiments are (Will 2001):

$$\eta_{\text{Eöt-Wash}} < 4 \times 10^{-13}, \qquad \eta_{\text{LLR}} < 5 \times 10^{-13} \qquad [1]$$

LLR also shows that gravitational binding energy falls with the same acceleration as ordinary matter to $1.3 \times 10^{-3}$ (test of the Nordtvedt effect – see the section "Bounds on the PPN parameters" and Table 1).

Many of the high-precision, low-noise methods that were developed for tests of WEP have been adapted to laboratory tests of the inverse-square law of Newtonian gravitation at millimeter scales and below. The goal of these experiments is to search for additional gravitational interactions involving massive particles or for the presence of large extra dimensions. The challenge of these experiments is to distinguish gravitation-like interactions from electromagnetic and quantum-mechanical effects. No deviations from

**Table 1** Current limits on the PPN parameters

| *Parameter* | *Effect* | *Limit* | *Remarks* |
|---|---|---|---|
| $\gamma - 1$ | (i) Shapiro delay | $2.3 \times 10^{-5}$ | Cassini tracking |
| | (ii) Light deflection | $4 \times 10^{-4}$ | VLBI |
| $\beta - 1$ | (i) Perihelion shift | $3 \times 10^{-3}$ | $J_2 = 10^{-7}$ from helioseismology |
| | (ii) Nordtvedt effect | $2.3 \times 10^{-4}$ | LLR plus bounds on other parameters |
| $\xi$ | Anisotropy in Newton's $G$ | $10^{-3}$ | Gravimeter bounds on anomalous Earth tides |
| $\alpha_1$ | Orbit polarization for moving systems | $10^{-4}$ | Lunar laser ranging |
| $\alpha_2$ | Anomalous spin precession for moving bodies | $4 \times 10^{-7}$ | Alignment of solar axis relative to ecliptic |
| $\alpha_3$ | Anomalous self-acceleration for spinning moving bodies | $2 \times 10^{-20}$ | Pulsar spindown timing data |
| $\eta$[a] | Nordtvedt effect | $9 \times 10^{-4}$ | Lunar laser ranging |
| $\zeta_1$ | | $2 \times 10^{-2}$ | Combined PPN bounds |
| $\zeta_2$ | Anomalous self-acceleration for binary systems | $4 \times 10^{-5}$ | Timing data for PSR 1913 + 16 |
| $\zeta_3$ | Violation of Newton's third law | $10^{-8}$ | Lunar laser ranging |
| $\zeta_4$ | | | Not independent |

[a]Here $\eta = 4\beta - \gamma - 3 - 10\xi/3 - \alpha_1 + 2\alpha_2/3 - 2\zeta_1/3 - \zeta_2/3$.

Newton's inverse-square law have been found to date at distances between 10 μm and 10 mm.

### Tests of Local Lorentz Invariance

Although special relativity itself never benefited from the kind of "crucial" experiments, such as the perihelion advance of Mercury and the deflection of light, that contributed so much to the initial acceptance of GR and to the fame of Einstein, the steady accumulation of experimental support, together with the successful integration of special relativity into quantum mechanics, led to its being accepted by mainstream physicists by the late 1920s, ultimately to become part of the standard toolkit of every working physicist.

But in recent years new experiments have placed very tight bounds on any violations of the Lorentz invariance, which underlies special relativity. A simple way of interpreting this new class of experiments is to suppose that a coupling of some external gravitation-like field (not the metric) to the electromagnetic interactions results in an effective change in the speed of electromagnetic radiation, $c$, relative to the limiting speed of material test particles, $c_0$; in other words, $c \neq c_0$. It can be shown that such a Lorentz-noninvariant electromagnetic interaction would cause shifts in the energy levels of atoms and nuclei that depend on the orientation of the quantization axis of the state relative to our velocity relative to the rest of the universe, and on the quantum numbers of the state, resulting in orientation dependences of the fundamental frequencies of such atomic clocks. The magnitude of these "clock anisotropies" would be proportional to $\delta \equiv |(c_0/c)^2 - 1|$, which vanishes if Lorentz invariance holds (see Will (1993) and Haugan and Will (1987) for details).

The earliest clock anisotropy experiments were carried out around 1960 independently by Hughes and Drever, although their original motivation was somewhat different. Dramatic improvements were made in the 1980s using laser-cooled trapped atoms and ions. This technique made it possible to reduce the broadening of resonance lines caused by collisions, leading to the impressive bound $|\delta| > 10^{-21}$ (Will 2001).

Other recent tests of Lorentz invariance violation include comparisons of resonant cavities with atomic clocks, tests of dispersion and birefringence in the propagation of high-energy photons from astrophysical sources, threshold effects in elementary particle collisions, and anomalies in neutrino oscillations. Mattingly (2005) gives a thorough and up-to-date review of both the theoretical frameworks for studying these effects and the experimental results.

### Tests of Local Position Invariance

LPI requires, among other things, that the internal binding energies of atoms and nuclei be independent of location in space and time, when measured against some standard atom. This means that a comparison of the rates of two different kinds of atomic clocks should be independent of location or epoch, and that the frequency shift between two identical clocks at different locations is simply a consequence of the apparent Doppler shift between a pair of inertial frames momentarily comoving with the clocks at the moments of emission and reception, respectively. The relevant parameter $\alpha$ appears in the formula for the frequency shift,

$$\Delta f/f = (1 + \alpha)\Delta\Phi/c^2 \quad [2]$$

where $\Phi$ is the Newtonian gravitational potential. If LPI holds, $\alpha = 0$. An early test of this was the Pound–Rebka experiment of 1960, which measured the frequency shift of gamma rays from radioactive iron nuclei in a tower at Harvard University. The best bounds come from a 1976 experiment in which a hydrogen maser atomic clock was launched to 10 000 km altitude on a Scout rocket and its frequency compared via telemetry with an identical clock on the ground, and a 1993 experiment in which two different kinds of atomic clocks were intercompared as a function of the varying solar gravitational field as seen on Earth (a "null" redshift experiment). The results are (Will 2001):

$$\alpha_{\text{Maser}} < 2 \times 10^{-4}, \qquad \alpha_{\text{Null}} < 10^{-3} \qquad [3]$$

Recent "clock comparison" tests of LPI include experiments done at the National Institute of Standards and Technology (NIST) in Boulder and at the Observatory of Paris, to look for cosmological variations in clock rates. The NIST experiment compared laser-cooled mercury ions with neutral cesium atoms over a two-year period, while the Paris experiment compared laser-cooled cesium and rubidium atomic fountains over five years; the results showed that the fine-structure constant is constant in time to a part in $10^{15}$ per year. A better bound of $6 \times 10^{-17}\,\text{yr}^{-1}$ comes from analysis of fission yields of the Oklo natural reactor, which occurred in Africa two billion years ago.

## Solar-System Tests

### The Parametrized Post-Newtonian Framework

It was once customary to discuss experimental tests of GR in terms of the "three classical tests," the gravitational redshift (which is really a test of the EEP, not of GR itself; see the section on tests of LPI), the perihelion advance of Mercury (the first success of the theory), and the deflection of light (whose measurement in 1919 made Einstein a celebrity). However, the proliferation of additional experimental tests and of well-motivated alternative metric theories of gravity made it desirable to develop a more general theoretical framework for analyzing both experiments and theories. This "parametrized post-Newtonian (PPN) framework" dates back to Eddington in 1922, but was fully developed by Nordtvedt and Will in the period 1968–72 (see Will (1993) for details).

When attention is confined to metric theories of gravity and, further, the focus is on the slow-motion, weak-field limit appropriate to the solar system and similar systems, it turns out that, in a broad class of metric theories, only the numerical values of a set of coefficients in the spacetime metric vary from theory to theory. The resulting PPN framework contains ten parameters: $\gamma$, related to the amount of spatial curvature generated by mass; $\beta$, related to the degree of nonlinearity in the gravitational field; $\xi$, $\alpha_1$, $\alpha_2$, and $\alpha_3$, which determine whether the theory violates LPI or LLI in gravitational experiments; and $\zeta_1$, $\zeta_2$, $\zeta_3$, and $\zeta_4$, which describe whether the theory has appropriate momentum conservation laws. In GR, $\gamma = 1$, $\beta = 1$, and the remaining parameters all vanish. In the scalar–tensor theory of Brans–Dicke, $\gamma = (1 + \omega_{\text{BD}})/(2 + \omega_{\text{BD}})$, where $\omega_{\text{BD}}$ is an adjustable parameter.

A number of well-known relativistic effects can be expressed in terms of these PPN parameters:

**Deflection of light**

$$\Delta\theta = \left(\frac{1+\gamma}{2}\right)\frac{4GM}{dc^2} = \left(\frac{1+\gamma}{2}\right) \times 1.7505\,\frac{R_\odot}{d}\ \text{arcsec} \qquad [4]$$

where $d$ is the distance of closest approach of a ray of light to a body of mass $M$, and where the second line is the deflection by the Sun, with radius $R_\odot$.

**Shapiro time delay**

$$\Delta t = \left(\frac{1+\gamma}{2}\right)\frac{4GM}{c^3}\ln\left[\frac{(r_1 + \boldsymbol{x}_1 \cdot \boldsymbol{n})(r_2 - \boldsymbol{x}_2 \cdot \boldsymbol{n})}{d^2}\right] \qquad [5]$$

where $\Delta t$ is the excess travel time of a round-trip electromagnetic tracking signal, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are the locations relative to the body of mass $M$ of the emitter and receiver of the round-trip signal ($r_1$ and $r_2$ are the respective distances), and $\boldsymbol{n}$ is the direction of the outgoing tracking signal.

**Perihelion advance**

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = \left(\frac{2+2\gamma-\beta}{3}\right)\frac{GM}{Pa(1-e^2)c^2} = \left(\frac{2+2\gamma-\beta}{3}\right) \times 42.98\ \text{arcsec}/100\,\text{yr} \qquad [6]$$

where $P$, $a$, and $e$ are the period, semimajor axis, and eccentricity of the planet's orbit, respectively; the second line is the value for Mercury.

**Nordtvedt effect**

$$\frac{m_{\text{G}} - m_{\text{I}}}{m_{\text{I}}} = \left(4\beta - \gamma - 3 - \tfrac{10}{3}\xi - \alpha_1 + \tfrac{2}{3}\alpha_2 - \tfrac{2}{3}\zeta_1 - \tfrac{1}{3}\zeta_2\right)\frac{|E_{\text{g}}|}{m_{\text{I}}c^2} \qquad [7]$$

where $m_{\text{G}}$ and $m_{\text{I}}$ are, respectively, the gravitational and inertial masses of a body such as the Earth or

Moon, and $E_g$ is its gravitational binding energy. A nonzero Nordtvedt effect would cause the Earth and Moon to fall with a different acceleration toward the Sun. In GR, this effect vanishes.

### Precession of a gyroscope

$$\frac{dS}{dt} = (\Omega_{FD} + \Omega_{Geo}) \times S$$
$$\Omega_{FD} = -\frac{1}{2}\left(1+\gamma+\frac{\alpha_1}{4}\right)\frac{G}{r^3c^2}(J - 3nn \cdot J)$$
$$= \frac{1}{2}\left(1+\gamma+\frac{\alpha_1}{4}\right) \times 0.041 \,\text{arcsec yr}^{-1}$$
$$\Omega_{Geo} = -\frac{1}{2}(1+2\gamma)v \times \frac{Gmn}{r^2c^2}$$
$$= \frac{1}{3}(1+2\gamma) \times 6.6 \,\text{arcsec yr}^{-1} \qquad [8]$$

where $S$ is the spin of the gyroscope, and $\Omega_{FD}$ and $\Omega_{Geo}$ are, respectively, the precession angular velocities caused by the dragging of inertial frames (Lense–Thirring effect) and by the geodetic effect, a combination of Thomas precession and precession induced by spatial curvature; $J$ is the angular momentum of the Earth, and $v$, $n$, and $r$ are, respectively, the velocity, direction, and distance of the gyroscope. The second line in each case is the corresponding value for a gyroscope in polar Earth orbit at about 650 km altitude (Gravity Probe B).

## Bounds on the PPN Parameters

Four decades of high-precision experiments, ranging from the standard light-deflection and perihelion-shift tests, to LLR, planetary and satellite tracking tests of the Shapiro time delay, and geophysical and astronomical observations, have placed bounds on the PPN parameters that are consistent with GR. The current bounds are summarized in Table 1 (Will 2001).

To illustrate the dramatic progress of experimental gravity since the dawn of Einstein's theory, Figure 1 shows a history of results for $(1+\gamma)/2$, from the 1919 solar eclipse measurements of Eddington and his colleagues (which made Einstein a celebrity), to modern-day measurements using very long baseline radio interferometry (VLBI), advanced radar tracking of spacecraft, and the astrometry satellite Hipparcos. The most recent results include a 2003 measurement of the Shapiro delay, performed by tracking the "Cassini" spacecraft on its way to Saturn, and a 2004 measurement of the bending of light via analysis of VLBI data on 541 quasars and compact radio galaxies distributed over the entire sky.

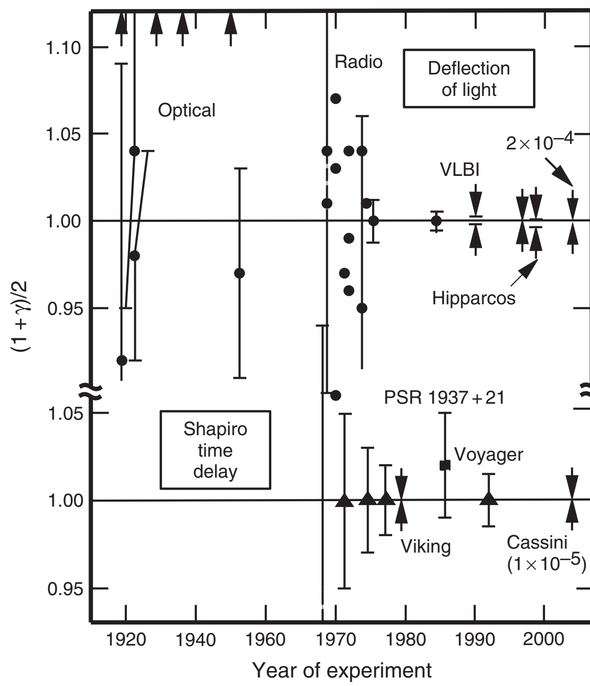



**Figure 1** Measurements of the coefficient $(1+\gamma)/2$ from observations of the deflection of light and of the Shapiro delay in propagation of radio signals near the Sun. The GR prediction is unity. "Optical" denotes measurements of stellar deflection made during solar eclipse, and "Radio" denotes interferometric measurements of radio-wave deflection. "Hipparcos" denotes the European optical astrometry satellite. Arrows denote values well off the chart from one of the 1919 eclipse expeditions and from others through 1947. Shapiro delay measurements using the Cassini spacecraft on its way to Saturn yielded tests at the 0.001% level, and light deflection measurements using VLBI have reached 0.02%.

The perihelion advance of Mercury, the first of Einstein's successes, is now known to agree with observation to a few parts in $10^3$. During the 1960s there was controversy about this test when reports of an excess solar oblateness implied an unacceptably large Newtonian contribution to the perihelion advance. However, it is now known from helioseismology, the study of short-period vibrations of the Sun, that the oblateness is of the order of a part in $10^7$, as expected from standard solar models, much too small to affect Mercury's orbit, within the observational errors.

## Gravity Probe B

The NASA Relativity Mission called Gravity Probe B (GPB) recently completed its mission to measure the Lense–Thirring and geodetic precessions of gyroscopes in Earth's orbit. Launched on 20 April 2004 for a 16-month mission, it consisted of four spherical rotors coated with a thin layer of superconducting niobium, spinning at 70–100 Hz, in a spacecraft filled with liquid helium, containing a telescope continuously pointed

toward a distant guide star (IM Pegasi). Superconducting current loops encircling each rotor were designed to measure the change in direction of the rotors by detecting the change in magnetic flux through the loop generated by the London magnetic moment of the spinning superconducting film. The spacecraft was in a polar orbit at 650 km altitude. The primary science goal of GPB was a 1% measurement of the 41 marcsec $yr^{-1}$ frame dragging or Lense–Thirring effect caused by the rotation of the Earth; its secondary goal was to measure to six parts in $10^5$ the larger 6.6 arcsec $yr^{-1}$ geodetic precession caused by space curvature.

## The Binary Pulsar

The binary pulsar PSR 1913 + 16, discovered in 1974, provided important new tests of GR. The pulsar, with a pulse period of 59 ms, was observed to be in orbit about an unseen companion (now generally thought to be a dead pulsar), with a period of $\sim$8 h. Through precise timing of apparent variations in the pulsar "clock" caused by the Doppler effect, the important orbital parameters of the system could be measured with exquisite precision. These included nonrelativistic "Keplerian" parameters, such as the eccentricity $e$, and the orbital period (at a chosen epoch) $P_b$, as well as a set of relativistic "post-Keplerian" (PK) parameters. The first PK parameter, $\langle\dot{\omega}\rangle$, is the mean rate of advance of periastron, the analog of Mercury's perihelion shift. The second, denoted $\gamma'$, is the effect of special relativistic time dilation and the gravitational redshift on the observed phase or arrival time of pulses, resulting from the pulsar's orbital motion and the gravitational potential of its companion. The third, $\dot{P}_b$, is the rate of decrease of the orbital period; this is taken to be the result of gravitational radiation damping (apart from a small correction due to the acceleration of the system in our rotating galaxy). Two other parameters, $s$ and $r$, are related to the Shapiro time delay of the pulsar signal if the orbital inclination is such that the signal passes in the vicinity of the companion; $s$ is a direct measure of the orbital inclination sin $i$. According to GR, the first three PK effects depend only on $e$ and $P_b$, which are known, and on the two stellar masses, which are unknown. By combining the observations of PSR 1913 + 16 (see Table 2) with the GR predictions, one obtains both a measurement of the two masses and a test of GR, since the system is overdetermined. The results are

$$m_1 = 1.4414 \pm 0.0002 M_\odot, \qquad m_2 = 1.3867 \pm 0.0002 M_\odot$$
$$\dot{P}_b^{GR}/\dot{P}_b^{OBS} = 1.0013 \pm 0.0021 \qquad [9]$$

Other relativistic binary pulsars may provide even more stringent tests. These include the relativistic neutron star/white dwarf binary pulsar J1141-6545, with a 0.19 day orbital period, which may ultimately lead to a very strong bound on the phenomenon of dipole gravitational radiation, predicted by many alternative theories of gravity, but not by GR; and the remarkable "double pulsar" J0737-3039, a binary system with two detected pulsars, in a 0.10 day orbit seen almost edge on and a periastron advance of 17° per year. For further discussion of binary pulsar tests, see Stairs (2003).

## Gravitational-Wave Tests

The detection of gravitational radiation by either laser interferometers or resonant cryogenic bars will usher in a new era of gravitational-wave astronomy (Barish and Weiss 1999). Furthermore, it will yield new and interesting tests of GR in its radiative regime (Will 1999).

GR predicts that gravitational waves possess only two polarization modes independently of the source; they are transverse to the direction of propagation and quadrupolar in their effect on a detector. Other theories of gravity may predict up to four additional modes of polarization. A suitable array of gravitational antennas could delineate or limit the number of modes present in a given wave. If distinct evidence were found of any mode other than the two transverse quadrupolar modes of GR, the result would be disastrous for the theory.

**Table 2** Parameters of the binary pulsars PSR 1913 + 16 and J0737-3039

| *Parameter* | *Symbol* | *Value*[a] *in PSR*1913 + 16 | *Value*[a] *in* J0737-3039 |
|---|---|---|---|
| *Keplerian parameters* | | | |
| Eccentricity | $e$ | 0.6171338(4) | 0.087779(5) |
| Orbital period | $P_b$ (day) | 0.322997448930(4) | 0.102251563(1) |
| *Post-Keplerian parameters* | | | |
| Periastron advance | $\langle\dot{\omega}\rangle$(°$yr^{-1}$) | 4.226595(5) | 16.90(1) |
| Redshift/time dilation | $\gamma'$ (ms) | 4.2919(8) | 0.382(5) |
| Orbital period derivative | $\dot{P}_b(10^{-12})$ | −2.4184(9) | |
| Shapiro delay ( sin $i$) | $s$ | | 0.9995(4) |

[a]Numbers in parentheses denote errors in last digit.

According to GR, gravitational waves propagate with the same speed, $c$, as light. In other theories, the speed could differ from $c$ because of coupling of gravitation to "background" gravitational fields, or propagation of the waves into additional spatial dimensions. Another way in which the speed of gravitational waves could differ from $c$ is if gravitation were propagated by a massive field (a massive graviton), in which case $v_g$ would be given by, in a local inertial frame,

$$\frac{v_g^2}{c^2} = 1 - \frac{m_g^2 c^4}{E^2} \approx 1 - \frac{c^2}{f^2 \lambda_g^2} \qquad [10]$$

where $m_g$, $E$, and $f$ are the graviton rest mass, energy, and frequency, respectively, and $\lambda_g = h/m_g c$ is the graviton Compton wavelength (it is assumed that $\lambda_g \gg c/f$).

The most obvious way to measure the speed of gravitational waves is to compare the arrival times of a gravitational wave and an electromagnetic wave from the same event (e.g., a supernova). For a source at a distance of 600 million light years (a typical distance for the currently operational detectors), and a difference in times on the order of seconds, the bound on the difference $|1 - v_g/c|$ could be as small as a part in $10^{17}$. It is worth noting that a 2002 report that the speed of gravity had been measured by studying light from a quasar as it propagated past Jupiter was fundamentally flawed. That particular measurement was not sensitive to the speed of gravity.

## Conclusions

The past four decades have witnessed a systematic, high-precision experimental verification of Einstein's theories. Relativity has passed every test with flying colors. A central theme of future work will be to test strong-field gravity in the vicinity of black holes and neutron stars, and to see how well GR works on cosmological scales. Gamma-ray, X-ray, microwave, infrared, neutrino, and gravitational-wave astronomy will all play a critical role in probing these largely unexplored aspects of GR.

GR is now the "standard model" of gravity. But, as in particle physics, there may be a world beyond the standard model. Quantum gravity, strings, and branes may lead to testable effects beyond Einstein's GR. Searches for such effects using laboratory experiments, particle accelerators, space instrumentation, and cosmological observations are likely to continue for some time to come.

*See also:* Cosmology: Mathematical Aspects; Einstein Equations: Exact Solutions; General Relativity: Overview; Geometric Flows and the Penrose Inequality; Gravitational Lensing; Gravitational Waves; Standard Model of Particle Physics.

## Further Reading

Barish BC and Weiss R (1999) LIGO and the detection of gravitational waves. *Physics Today* 52: 44.

Haugan MP and Will CM (1987) Modern tests of special relativity. *Physics Today* 40: 69.

Mattingly D (2005) *Modern tests of Lorentz invariance. Living Reviews in Relativity* 8: 5 (online article cited on 1 April 2005, http://www.livingreviews.org/lrr-2005-5).

Stairs IH (2003) *Testing general relativity with pulsar timing. Living Reviews in Relativity* 6: 5 (online article cited on 1 April 2005, http://www.livingreviews.org/lrr-2003-5).

Will CM (1993) *Theory and Experiment in Gravitational Physics.* Cambridge: Cambridge University Press.

Will CM (1999) Gravitational radiation and the validity of general relativity. *Physics Today* 52: 38.

Will CM (2001) *The confrontation between general relativity and experiment. Living Reviews in Relativity* 4: 4 (online article cited on 1 April 2005, http://www.livingreviews.org/lrr-2001-4).

# General Relativity: Overview

**R Penrose**, University of Oxford, Oxford, UK



## The Principle of Equivalence

The special theory of relativity is founded on two basic principles: that the laws of physics should be independent of the uniform motion of an inertial frame of reference, and that the speed of light should have the same constant value in any such frame. In the years between 1905 and 1915, Einstein pondered deeply on what was, to him, a profound enigma, which was the issue of why these laws retain their proper form only in the case of an inertial frame. In special relativity, as had been the case in the earlier dynamics of Galilei–Newton, the laws indeed retain their basic form only when the reference frame is unaccelerated (which includes it being nonrotating). It demonstrated a particular prescience on the part of Einstein that he should have demanded the seemingly impossible requirement that the very same dynamical laws should hold also in an accelerating (or even rotating) reference frame. The key realization came to him late in 1907, when sitting in his chair in the Bern

patent office he had the "happiest thought" in his life, namely that if a person were to fall freely in a gravitational field, then he would not notice that field at all while falling. The physical point at issue is Galileo's early insight (itself having roots even earlier from Simon Stevin in 1586 or Ioannes Philiponos in the fifth or sixth century) that the acceleration induced by gravity is independent of the body upon which it acts. Accordingly, if two neighboring bodies are accelerated together in the same gravitational field, then the motion of one body, in the (nonrotating) reference frame of the other, will be as though there were no gravitational field at all. To put this another way, the effect of a gravitational force is just like that of an accelerating reference system, and can be eliminated by free fall. This is now known as the "principle of equivalence."

It should be made clear that this is a particular feature of only the gravitational field. From the perspective of Newtonian dynamics, it is a consequence of the seemingly accidental fact that the concept of (passive) "mass" $m$ that features in Newton's law of gravitational attraction, where the attractive force due to the gravitational field of another body, of mass $M$, has the form

$$\frac{GmM}{r^2}$$

is the same as – or, at least, proportional to – the inertial mass $m$ of the body which is being acted upon. Thus, the impedance to acceleration of a body and the strength of the attractive force on that body are, in the case of gravity (and only in the case of gravity), in proportion to one another, so that the acceleration of a body in a gravitational field is independent of its mass (or, indeed, of any other localized magnitude) possessed by it. (The fact that the active gravitational mass, here given by the quantity $M$, is also in proportion to its own passive gravitational mass – from Newton's third law – may be regarded as a feature of the general Lagrangian/Hamiltonian framework of physics. But see Bondi (1957).) Other forces of nature do not have this property. For example, the electrostatic force on a charged body, by an electric field, acts in proportion to the electric charge on that body, whereas, the impedance to acceleration is still the inertial mass of that body, so the acceleration induced depends on the charge-to-mass ratio. Accordingly, it is the gravitational field alone which is equivalent to an acceleration.

Einstein's fundamental idea, therefore, was to take the view that the "relativity principle" could as well be applied to accelerating reference frames as to inertial ones, where the same physical laws would apply in each, but where now the perceived gravitational field would be different in the two frames. In accordance with this perspective, Einstein found it necessary to adopt a different viewpoint from the Newtonian one, both with regard to the notion of "gravitational force" and to the very notion of an "inertial frame." According to the Newtonian perspective, it would be appropriate to describe the action of the Earth's gravitational field, near some specific place on the Earth's surface, in terms of a "Newtonian inertial frame" in which the Earth is "fixed" (here we ignore the Earth's rotation and the Earth's motion about the Sun), and we consider that there is a constant gravitational field of force (directed towards the Earth's center). But the Einsteinian perspective is to regard that frame as noninertial where, instead, it would be a frame which falls freely in the Earth's (Newtonian) gravitational field that would be regarded as a suitable "Einsteinian inertial frame." Generally, to be inertial in Einstein's sense, the frame would refer to free fall under gravity, so that the Newtonian field of gravitational force would appear to have disappeared – in accordance with his "happiest thought" that Einstein had had in the Bern patent office. We see that the concept of a gravitational field must also be changed in the passage from Newton's to Einstein's viewpoint. For in Newton's picture we indeed have a "gravitational force" directed towards the ground with a magnitude of $gm$, where $m$ is the mass of the body being acted upon and $g$ is the "acceleration due to gravity" at the Earth's surface, whereas in Einstein's picture we have specifically eliminated this "gravitational force" by the choice of "Einsteinian inertial frame."

It might at first seem puzzling that the gravitational field has appeared to have been removed altogether by this device, and it is natural to wonder how gravitational effects can have any physical role to play at all from this point of view! However, this would be to go too far, as the Newtonian gravitational field may vary from place to place – as it does, indeed, in the case of the Earth's field, since it is directed towards the Earth's center, which is a different spatial direction at different places on the Earth's surface. Our considerations up to this point really refer only to a small neighborhood of a point. One might well take the view that a "frame" ought really to describe things also at widely separated places at once, and the considerations of the paragraphs above do not really take this into consideration.

## The Tidal Effect

To proceed further, it will be helpful to consider an astronaut A in free fall, high above the Earth's surface. Let us first adopt a Newtonian perspective.

We shall be concerned only with the instantaneous accelerations due to gravity in the neighborhood of A, so it will be immaterial whether we regard the astronaut as falling to the ground or – more comfortably! – in orbit about the Earth.

Let us imagine that the astronaut is initially surrounded, nearby, by a sphere of particles, with A at the centre, which are taken to be initially at rest with respect to A (see Figure 1). To a first approximation, all the particles will share the same acceleration as the astronaut, so they will seem to the astronaut to hover motionless all around. But now let us be a little more precise about the accelerations. Those particles which are initially located in a vertical line from A, that is, either directly below A, at B, or directly above A, at T, will have, like A, an acceleration which is in the direction AO, where O is the Earth's center. But for the bottom point B, the acceleration will be slightly greater than that at A, and for the top point T, the acceleration will be slightly less than the acceleration at A, because of the slightly differing distances from O. Thus, relative to A, both will initially accelerate away from A. With regard to particles in the sphere which are initially in a circle in the horizontal plane through A, the direction to O will now be somewhat inwards, so that the particles at these points $H_i$ will accelerate, relative to A, slightly inwards. Accordingly, the entire sphere of particles will begin to get distorted into a prolate spheroid (elongated ellipsoid of revolution). This is referred to as the tidal distortion, for the good reason that it is precisely the same physical effect which is responsible for the tides in the Earth's oceans, where for this illustration we are to think of the Earth's center as being at A, the Moon (or Sun) to be situated at O, and the sphere of particles to represent the surface of the water of the Earth's oceans.

It is not hard to calculate (reverting, now, to our original picture) that, as a reflection of Newton's inverse-square law of gravitational attraction, the amount of (small) outward vertical displacement from A (at B and T) will be twice the inward horizontal displacement (over the circle of points $H_i$); accordingly, the sphere will initially be distorted into an ellipsoid of the same volume. This depends upon there being no gravitating matter inside the sphere. The presence of such matter would contribute a volume-reducing effect in proportion to the total mass surrounded. (An extreme case illustrating this would occur if we take our sphere of particles to surround the entire Earth, where the volume-reducing effect would be manifest in the accelerations towards the ground at all points of the surrounding sphere.)

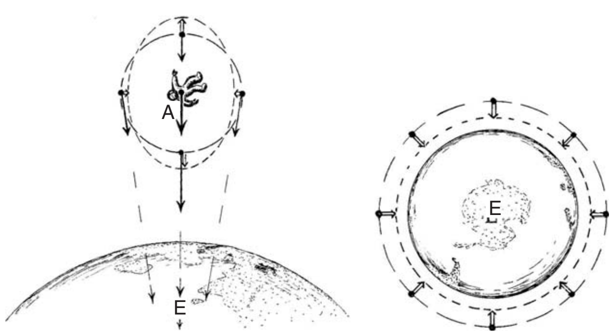



**Figure 1** (a) Tidal effect. The astronaut A surrounded by a sphere of nearby particles initially at rest with respect to A. In Newtonian terms, they have an acceleration towards the Earth's center E, varying slightly in direction and magnitude (single-shafted arrows). By subtracting A's acceleration from each, we obtain the accelerations relative to A (double-shafted arrows); this relative acceleration is slightly inward for those particles displaced horizontally from A, but slightly outward for those displaced vertically from A. Accordingly, the sphere becomes distorted into a (prolate) ellipsoid of revolution, with symmetry axis in the direction AE. The initial distortion preserves volume. (b) Now move A to the Earth's center E and the sphere of particles to surround E just above the atmosphere. The acceleration (relative to A = E) is inward all around the sphere, with an initial volume reduction acceleration $4\pi GM$, where $M$ is the total mass surrounded. Reproduced with permission from Penrose R (2004) *The Road to Reality: A Complete Guide to the Laws of the Universe*. London: Jonathan Cape.

## Gravity as Curved Spacetime

It is appropriate to take a spacetime view of these phenomena (Figure 2). The distortions that we have been considering are, in fact, direct manifestations

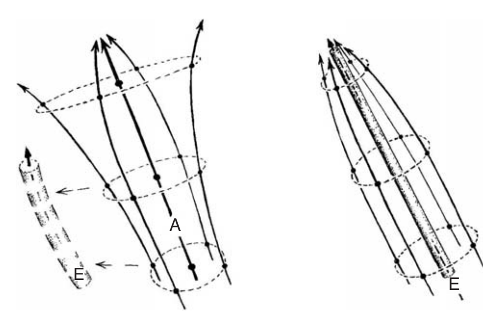
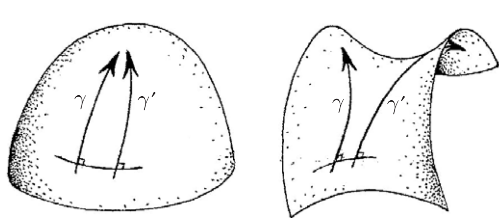



**Figure 2** Spacetime versions of Figure 1 in terms of the relative distortion of neighboring geodesics. (a) Geodesic deviation in empty space (basically Weyl curvature) as seen in the world lines of A and surrounding particles (one spatial dimension suppressed), as might be induced from the gravitational field of a nearby body E. (b) The corresponding inward acceleration (basically Ricci curvature) due to the mass density within the bundle of geodesics. Reproduced with permission from Penrose R (2004) *The Road to Reality: A Complete Guide to the Laws of the Universe*. London: Jonathan Cape.

of spacetime curvature, according to Einstein's viewpoint. We are to think of the world line of a particle, falling freely under gravity (Einsteinian inertial motion), as described as some kind of geodesic in spacetime. We shall be coming to this more completely shortly, but for the moment it will be helpful to picture the behavior of geodesics within an ordinary curved 2-surface S (Figure 3). If S has positive (Gaussian) curvature, then there will be a tendency for geodesics on S to bend towards each other, so that a pair of infinitesimally separated geodesics which are initially parallel will begin to get closer together as we move along them; if S has negative (Gaussian) curvature, then there will be a corresponding tendency for geodesics on S to bend away from each other. This is what happens in two dimensions, where the intrinsic curvature at a point is given by a single number. However, we are now concerned with a four-dimensional space, where the notion of curvature requires many more components. We see in **Figure 2** that we are indeed to expect mixtures of convergence and divergence of geodesics, which suggests that there are both positive and negative curvature components involved, the positive curvature being in the horizontally displaced directions from A and the negative curvature in the vertically displaced directions. In a curved space of dimension 4, as is the case for a curved spacetime, we can expect 20 independent components of curvature at each point altogether. In the present situation, the others would be called into play when differing velocities of A are considered.

Let us see how we are to accommodate the above considerations within the standard framework of differential geometry. So far, we have not really deviated from Newtonian theory, even though we have been considering "geodesics" in a four-dimensional spacetime. In fact, it is perfectly legitimate to view Newtonian theory in this way (*see* Newtonian




(a) (b)

**Figure 3** Geodesic deviation when M is a 2-surface (a) of positive (Gaussian) curvature, when the geodesics $\gamma, \gamma''$ bend towards each other, and (b) of negative curvature, when they bend apart. Reproduced with permission from Penrose R (2004) *The Road to Reality: A Complete Guide to the Laws of the Universe*. London: Jonathan Cape.

Limit of General Relativity), although the 4-geometry description is somewhat more complicated than one might wish. This is due to the fact that the infinite speed at which gravitation is taken to act in Newtonian theory demands that the "metric" of Newtonian spacetime is degenerate. (In effect, one would have a degenerate "dual metric" $G^{ab}$, of matrix rank 3, which plays a role in defining spatial displacements and a very degenerate "metric" $G_{ab}$, of matrix rank 1, which defines temporal differences, where $G^{ab}G_{bc} = 0$; *see* Newtonian Limit of General Relativity.) Accordingly, there is no unique notion of "geodesic" defined by the metric in Newtonian theory.

It is striking that although the insights provided by the principle of equivalence are to some considerable extent independent of special relativity (since we see from the paragraphs prior to the preceding one that a curved-spacetime-geometry view of gravity is natural in the light of the equivalence principle alone), it is the nondegenerate metric $g_{ab}$, (and its inverse $g^{ab}$) that special relativity gives us locally, which leads to an elegant spacetime theory of gravity. Although the metric $g_{ab}$ is Lorentzian (with preferred choice of signature $+---$ here) rather than positive definite, so that the spacetime is not strictly a Riemannian one, the change of signature makes little difference to the local formalism. In particular, the fact that the metric defines a unique (torsion-free) connection preserving it is unaffected by the signature. This connection is the one defined by Christoffel's symbols

$$\Gamma_{ac}{}^{b} = \tfrac{1}{2} g^{db}(\partial_c g_{da} + \partial_a g_{cd} - \partial_d g_{ca})$$

where $\partial_a$ stands for coordinate derivative $\partial/\partial x^a$, so that the covariant derivative of a vector $V^a$ is given by

$$\nabla_a V^b = \partial_a V^b + V^c \Gamma_{ac}{}^{b}$$

(Here the standard "physicist's conventions" are being used, whereby notation such as "$g_{ab}$" and "$V^a$" can be used interchangeably either for the sets of components of the metric tensor $\boldsymbol{g}$ and the vector $\boldsymbol{V}$, respectively, or alternatively for the entire geometrical metric tensor $\boldsymbol{g}$ or vector $\boldsymbol{V}$, in each case; moreover, the summation convention is being assumed, or this can alternatively be understood in terms of abstract indices. (For the abstract-index notation for tensors, see Penrose and Rindler (1984), especially Chapters 2 and 4. Sign and index-ordering conventions used here follow those given in that book. Many other authors use conventions which differ from these in various, usually minor respects.))

## Physical Interpretation of the Metric

Some words of clarification are needed, as to the meaning of the metric tensor $g_{ab}$ in relativity theory. In the early discussions by Einstein and others, the spacetime metric tended to be interpreted in terms of little "rulers" placed on a curved manifold. Although this is natural in the Riemannian (positive-definite) case, it is not quite so appropriate for the Lorentzian geometry of spacetime manifolds. An ordinary physical ruler has a spacetime description as a timelike strip, and it does not naturally express the spatial separation between two spacelike-separated events. In order for a ruler to measure such a spacelike separation, it would be necessary for the two events to be simultaneous in the ruler's rest frame, and for this to be assured, some further mechanism would be needed, such as Einstein's procedure for ensuring simultaneity by the use of light signals from the two events to be received simultaneously at their midpoint on the ruler. Clearly this complicates the issue, and it turns out to be much preferable to concentrate on temporal displacements rather than spatial ones.

The idea that spacetime geometry should really be regarded as "chronometry," in this way, has been stressed by a number of distinguished expositors of relativity theory, most notably John L Synge (1956, 1960) and Hermann Bondi (1961, 1964, 1967). Where needed, spatial displacements can then be defined by the use of temporal ones together with light signals. This has the additional advantage that in modern technology, the measurement of (proper) time far surpasses that of distance in accuracy, to the extent that the meter is now defined simply by the requirement that there are exactly 299792458 of them in a light-second! The proper time interval between two nearby events is, indeed, measured by a clock which encounters both events, moving inertially between the two, and very precise atomic and nuclear clocks are now a common feature of current technology. The physical role of the metric $g_{ab}$ is most clearly seen in the formula

$$\tau = \int_p^q (g_{ab}\,\mathrm{d}x^a\,\mathrm{d}x^b)$$

which measures the (proper) time interval $\tau$ between an event $p$ and a later event $q$ on its world line, the integral being taken along this curve, and where now that curve need not be a geodesic, so that accelerating (noninertial) motion of the clock is allowed. The metric (with choice of signature $+---$ so that it is the timelike displacements that are directly provided as real numbers) is very precisely specified by this physical requirement, and this tells us that the pseudo-Riemannian (Lorentzian) structure of spacetime is far from being an arbitrary construction, but is given to us by Nature with enormous precision. (Some theorists prefer to use the alternative spacetime signature $-+++$, because this more directly relates to familiar Newtonian concepts, these being normally described in spatial terms. The difference is essentially just a notational one, however. It may be remarked that the 2-spinor formalism (*see* Spinors and Spin Coefficients) fits in much more readily with the $+---$ signature being used here.) It may be noted, also that this time measure is ultimately fixed by quantum principles and the masses of the elementary ingredients involved (e.g., particle masses) via the Einstein and Planck relations $E=mc^2$ and $E=h\nu$, so that there is a natural frequency associated with a given mass, via $\nu=mc^2/h$ ($c$ being the speed of light and $h$ being Planck's constant).

## Riemann Curvature and Geodesic Deviation

The unique torsion-free (Christoffel–Levi-Civita) connection $\nabla_a$ is, via this physically determined metric, also fixed accordingly by these physical considerations, as is the notion of a geodesic, and therefore so also is the curvature. The 20-independent-component Riemann curvature tensor $R_{abcd}$ may be defined by

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)V^d = R_{abc}{}^d V^c$$

with normal index-raising/lowering conventions, so that $R_{abcd}=R^e_{abc}g_{ed}$, etc., and we have the standard classical formula

$$R_{abc}{}^d = \partial_a\Gamma_{cb}{}^d - \partial_b\Gamma_{ca}{}^d + \Gamma_{cb}{}^e\Gamma_{ea}{}^d - \Gamma_{ca}{}^e\Gamma_{eb}{}^d$$

The symmetries $R_{abcd}=R_{cdab}=-R_{bacd}$, $R_{abcd}+R_{bcad}+R_{cabd}=0$ reduce the number of independent components of $R_{abcd}$ to 20 (from a potential $4^4=64$). Of these, 10 are locally fixed by the kind of physical requirement indicated above, that in order to express something that agrees closely with Newton's inverse-square law we require that there should be a net inward curving of free world lines (the timelike geodesics that represent local inertial motions, or "free fall" under gravity). Let us see how this requirement is satisfied in Einstein's general relativity.

What we find, from Newton's theory, is that a system of test particles which, at some initial time constitutes a closed 2-surface at rest surrounding some gravitating matter, will begin to accelerate in such a way that the volume surrounded is initially reduced in proportion to the total mass surrounded.

This volume reduction is a direct consequence of Poisson's equation $\nabla^2\Phi = -4\pi\rho$ ($\Phi$ being the gravitational potential and $\rho$ the mass density) and of Newton's second law, which tell us that the second time derivative of the free-fall volume of our initially stationary closed surface of test particles is indeed $-4\pi GM$, where $M$ is the total gravitating mass surrounded (and $G$ is Newton's constant, as above). In Einstein's theory, we can basically carry this over to our four-dimensional Lorentzian spacetime. We do, however, find that such a general statement as this does not exactly hold. Instead of referring to 3-volumes of any size, we must restrict attention to infinitesimal volumes.

The basic mathematical tool is the equation of "geodesic deviation," namely the "Jacobi equation":

$$\mathrm{D}^2 u^d = R_{abc}{}^d t^a u^b t^c$$

where D describes "propagation derivative"

$$\mathrm{D} = t^a \nabla_a$$

along a timelike geodesic $\gamma$, where $t^a$ is a unit timelike tangent vector to $\gamma$ (so $t_a t^a = g_{ab} t^a t^b = 1$) which is (consequently) parallel-propagated along $\gamma$,

$$\mathrm{D}t^a = 0$$

(When acting on a scalar quantity defined along $\gamma$, we can read "D" as "$\mathrm{d}/\mathrm{d}\tau$," where $\tau$ measures proper time along $\gamma$.) The vector $u^a$ is what is called a connecting vector between the geodesic $\gamma$ and some "neighboring geodesic" $\gamma'$. We think of the vector $u^a$ as "connecting" a point $p$ on $\gamma$ to some neighboring point $p'$ on $\gamma'$, where it is usual to take $u^a$ to be orthogonal to $t^a$ (i.e., $u_a t^a = 0$). The derivative $\mathrm{D}u^a$ measures the rate of change of $u^a$, as $p$ and $p'$ move together into the future along $\gamma$. Mathematically, we express this as the vanishing of the Lie derivative of $u^a$ with respect to $t^a$ (with $t_a$ extended to a unit vector field which is tangent both to $\gamma$ and to $\gamma'$). By taking three independent vectors $u^a$ at $p$, we can form a spatial 3-volume element $W$ and investigate how this propagates along $\gamma$. We find

$$\mathrm{D}^2 W = W R_{ab} t^a t^b$$

where the Ricci tensor $R_{ab} (= R_{ba})$ is here defined by

$$R_{ab} = R_{acb}{}^c$$

## The Einstein Field Equations

In view of what has been said above, with regard to the way that the acceleration of volume behaves in Newtonian theory, it would be natural to "identify" $R_{ab} t^a t^b$ with ($-4\pi G\times$) the (active gravitational) mass density, with respect to the time direction $t^a$. In (special) relativity theory we expect to identify mass density with $c^{-2}\times$ energy density (by $E = mc^2$) and to take energy density as just one component (the time–time component) of a symmetric tensor $T_{ab}$, called the "energy tensor," and for simplicity we now take $c = 1$. The tensor quantity $T_{ab}$ is to incorporate the contributions to the local mass/energy density of all particles and fields other than gravity itself. Since we would require this to work for all choices of time-direction $t^a$, it would be natural, accordingly, to make the identification

$$R_{ab} = -4\pi G T_{ab}$$

Indeed, this was Einstein's initial choice for a gravitational field equation. However, this will actually not do, as Einstein later realized. The trouble comes from the Bianchi identity

$$\nabla_a R_{bcde} + \nabla_b R_{cade} + \nabla_c R_{abde} = 0$$

from which we deduce

$$\nabla^a \left(R_{ab} - \tfrac{1}{2} R g_{ab}\right) = 0$$

where

$$R = R_a{}^a$$

This causes trouble in connection with the standard requirement on the energy tensor, that it satisfy the local "conservation law"

$$\nabla^a T_{ab} = 0$$

The latter equation is an essential requirement in special relativity, since it expresses the conservation of energy and momentum for fields in flat spacetime. In standard Minkowski coordinates, each of $T_{a0}, T_{a1}, T_{a2}, T_{a3}$ satisfies an equation just like the $\nabla^a J_a = 0$ of the charge–current vector $J_a$ of Maxwell's theory of electromagnetism, with now $\nabla_a = \partial_a = \partial/\partial x^a$, which expresses global conservation of charge. Similarly to the way that $J_a$ encapsulates density and flux of electric charge, $T_{a0}$ encapsulates density and flux of energy, and $T_{a1}, T_{a2}, T_{a3}$ encapsulate the same for the three components of momentum. So the equation $\nabla^a T_{ab} = 0$ is essential in special relativity, for similarly expressing global conservation of energy and momentum. We find (referring to a local inertial frame) that, when we pass to general relativity, this equation should still hold, with $\nabla_a$ now standing for covariant derivative. But the initially proposed field equation $R_{ab} = -4\pi G T_{ab}$ would now give us $\nabla^a R_{ab} = 0$, which combined with the geometrically necessary $\nabla^a (R_{ab} - (\frac{1}{2}) R g_{ab}) = 0$, tells us that $R$ is constant. In turn this implies the physically unacceptable requirement that $T = T_a^a$ is constant (since we have $R = -4\pi GT$).

Einstein eventually became convinced (by 1915) of the modified field equations

$$R_{ab} - \tfrac{1}{2}Rg_{ab} = -8\pi G T_{ab}$$

(the "8" rather than "4" being now needed to fit in with the Newtonian limit) and it is these that are now commonly referred to as "Einstein's field equations." (Some authors prefer to use the singular form "field equation," especially if the formula is to be read as an abstract-index expression rather than a family of component equations, since the tensors involved are really single entities.) It may be noted that the formula can be rewritten as

$$R_{ab} = -8\pi G\left(T_{ab} - \tfrac{1}{2}Tg_{ab}\right)$$

from which we deduce that in Einstein's theory the source of gravity is not simply the mass (or equivalently energy) density, but there is an additional contribution from the pressure (momentum flux, i.e., space–space components of $T_{ab}$). This can have significant implications for the instability of very large and massive stars in highly relativistic regimes, where increases in pressure can, paradoxically, actually increase the tendency for a star to collapse, owing to its contribution to the attractive effect of its gravity.

In 1917, Einstein put forward a slight modification of his field equations – basically the only modification that can be made without fundamentally changing the foundations of his theory – by introducing the very tiny cosmological constant $\Lambda$. The modified equations are

$$R_{ab} - \tfrac{1}{2}Rg_{ab} + \Lambda g_{ab} = -8\pi G T_{ab}$$

and the source of gravity, or active gravitational mass is now

$$\rho + P_1 + P_2 + P_3 - \frac{\Lambda}{4\pi G}$$

where (with respect to a local Lorentzian orthonormal frame, units being chosen so that $c = 1$) $\rho = T_{00}$ is the mass/energy density and $P_1 = T_{11}, P_2 = T_{22}, P_3 = T_{33}$ are the principal pressures. The $\Lambda$-term, for positive $\Lambda$, provides a repulsive contribution to the gravitational effect, but it is extremely tiny (and totally ignorable) on all ordinary scales, beginning to show itself only at the most vast of observed cosmological distances (since the effect of $\Lambda$ adds up relentlessly at larger and larger distances). Einstein originally introduced the term in order to have the possibility of a static universe, where the attractive gravitational effect of the totality of ordinary matter would be balanced, overall, by $\Lambda$. But the discovery of the expansion of the universe (by Hubble and others) led Einstein to abandon the cosmological term. However, since 1998 (initially from the supernova observations of Brian Schmidt and Robert Kirschner, and Saul Perlmutter, *see* Perlmutter *et al.* (1998)), cosmological evidence has mounted in favor of the presence of a very small positive $\Lambda$-term, which has resulted in the expansion of the universe beginning to accelerate. While the presence of Einstein's constant $\Lambda$-term is consistent with observations, and remains the simplest explanation of this observed acceleration, many cosmologists prefer to allow for what would amount to a "varying $\Lambda$," and refer to it as "dark energy."

## Energy Conservation and Related Matters

One of the features of Einstein's general relativity theory that had been deeply puzzling to a good many of Einstein's contemporaries, and which may be said to be still not fully resolved, even today, is "energy conservation," in the presence of a dynamical gravitational field. We have noted that the energy tensor $T_{ab}$ is to incorporate the contributions of all particles and fields other than gravity. But what about gravity itself? There are many physical situations in which energy can be transferred back and forth from gravitational systems to nongravitational ones (most strikingly in the example of the emission of gravitational waves; *see* Gravitational Waves). The conservation of energy would make no sense without an understanding of how energy can be stored in a gravitational field. At first sight we seem to see no role for a gravitational contribution to energy in Einstein's theory, since the conservation law $\nabla^a T_{ab} = 0$ seems to be a self-contained expression of energy conservation with no direct contribution from the gravitational field in the tensor $T_{ab}$. However, this is illusory, since the formulation of a global conservation law from the local covariant expression $\nabla^a T_{ab} = 0$ does not work in curved spacetime (basically because, unlike the charge–current quantity $J_a$ of Maxwell's electrodynamical theory, the extra index on $T_{ab}$ prevents it from being regarded as a 1-form). We may take the view that the energy of gravitation enters nonlocally into the equation, so that the failure of $T_{ab}$ to provide a global conservation law on its own is an expression of the gravitational contributions of energy not being taken into account. This is no doubt a correct attitude to take, but it is a difficult one to express comprehensively in a mathematical form. Einstein himself provided a partial understanding, but at the expense of introducing concepts known as "pseudotensors" whose meaning was too tied up with arbitrary choices of coordinate systems to provide an overall picture. In modern approaches, the most clear-cut results come from the study of asymptotically flat or asymptotically de Sitter spacetimes (de Sitter space being the empty universe which takes over the role of Minkowski space when

there is a positive cosmological constant $\Lambda$; *see* Cosmology: Mathematical Aspects).

The important role of the "Weyl conformal tensor"

$$C_{abcd} = R_{abcd} - \tfrac{1}{2}(R_{ac}g_{bd} - R_{bc}g_{ad} + R_{bd}g_{ac} - R_{ad}g_{bc}) + \tfrac{1}{6}R(g_{ac}g_{bd} - g_{bc}g_{ad})$$

should also be pointed out. This tensor retains all the symmetries of the full Riemann tensor, but has the Ricci tensor contribution removed, so that all its contractions vanish, as is exemplified by

$$C_{abc}{}^{a} = 0$$

It describes the conformal part of the curvature, that is, that part that survives under conformal rescalings of the metric;

$$g_{ab} \mapsto \Omega^2 g_{ab}$$

where $\Omega$ is a smooth (positive) function of position. The tensor $C_{abc}{}^{d}$ is itself invariant under these conformal rescalings. This has importance in the asymptotic analysis of gravitational fields (*see* Asymptotic Structure and Conformal Infinity). We may take the view that $C_{abcd}$ describes the degrees of freedom in the free gravitational field, whereas $R_{ab}$ contains the information of the sources of gravity. This is analogous to the Maxwell tensor $F_{ab}$ describing the degrees of freedom in the free electromagnetic field, whereas $J_a$ contains the information of the sources of electromagnetism.

From the observational point of view, general relativity stands in excellent shape, with full agreement with all known relevant data, starting with the anomalous perihelion advance of the planet Mercury observed by LeVerrier in the mid-nineteenth century, through clock-slowing, light-bending (lensing) and time-delay effects, and the necessary corrections to GPS positioning systems, to the precise orbiting of double neutron-star systems, with energy loss due to the emission of gravitational waves. The effects of gravitational lensing now play vital roles in modern cosmology. To get some idea of the precision in Einstein's theory, we may take note of the fact that the double neutron-star system PSR 1913+16 has been observed for some 30 years, and the agreement between observation and theory overall is to about one part in $10^{14}$.

*See also:* Asymptotic Structure and Conformal Infinity; Canonical General Relativity; Computational Methods in General Relativity: the Theory; Cosmology: Mathematical Aspects; Einstein Equations: Exact Solutions; Einstein Equations: Initial Value Formulation; Einstein–Cartan Theory; Einstein's Equations with Matter; General Relativity: Experimental Tests; Geometric Flows and the Penrose Inequality; Gravitational Lensing; Gravitational Waves; Hamiltonian Reduction of Einstein's Equations; Lorentzian Geometry; Newtonian Limit of General Relativity; Noncommutative geometry and the Standard Model; Spacetime Topology, Causal Structure and Singularities; Spinors and Spin Coefficients; Symmetries and Conservation Laws; Twistor Theory: Some Applications [in Integrable Systems, Complex Geometry and String Theory].

## Further Reading

Bondi H (1957) Negative mass in general relativity. *Reviews of Modern Physics* 29: 423 (*Mathematical Reviews* 19: 814).
Bondi H (1961) *Cosmology*. Cambridge: Cambridge University Press.
Bondi H (1964) *Relativity and Common Sense*. London: Heinemann.
Hartle JB (2003) *Gravity: An Introduction to Einstein's General Relativity*. San Francisco: Addison Wesley.
Penrose R (2004) *The Road to Reality: A Complete Guide to the Laws of the Universe*. London: Jonathan Cape.
Penrose R and Rindler W (1984) *Spinors and Space-Time, Vol. 1: Two-Spinor Calculus and Relativistic Fields*. Cambridge: Cambridge University Press.
Perlmutter S *et al.* (1998) Cosmology from type Ia supernovae. *Bulletin of the American Astronomical Society* 29 (astro-ph/9812473).
Rindler W (2001) *Relativity: Special, General, and Cosmological*. Oxford: Oxford University Press.
Synge JL (1960) *Relativity: The General Theory*. Amsterdam: North-Holland.
Wald RM (1984) *General Relativity*. Chicago: University of Chicago Press.

# Generic Properties of Dynamical Systems

**C Bonatti**, Université de Bourgogne, Dijon, France



## Introduction

The state of a concrete system (from physics, chemistry, ecology, or other sciences) is described using (finitely many, say $n$) observable quantities (e.g., positions and velocities for mechanical systems, population densities for echological systems, etc.). Hence, the state of a system may be represented as a point $x$ in a geometrical space $\mathbb{R}^n$. In many cases, the quantities describing the state are related, so that the phase space (space of all possible states) is a submanifold $M \subset \mathbb{R}^n$. The time evolution of the system is represented by a curve $x_t, t \in \mathbb{R}$ drawn on the phase space $M$, or by a sequence $x_n \in M$, $n \in \mathbb{Z}$, if we consider discrete time (i.e., every day at the same time, or every January 1st).

Believing in determinism, and if the system is isolated from external influences, the state $x_0$ of the system at the present time determines its evolution. For continuous-time systems, the infinitesimal

evolution is given by a differential equation or vector field $dx/dt = X(x)$; the vector $X(x)$ represents velocity and direction of the evolution. For a discrete-time system, the evolution rule is a function $F: M \to M$; if $x$ is the state at time $t$, then $F(x)$ is the state at the time $t+1$. The evolution of the system, starting at the initial data $x_0$, is described by the orbit of $x_0$, that is, the sequence $\{(x_n)_{n\in\mathbb{Z}} \mid x_{n+1} = F(x_n)\}$ (discrete time) or the maximal solution $x_t$ of the differential equation $ax/dt = X(x)$ (continuous time).

**General problem** *Knowing the initial data and the infinitesimal evolution rule, what can we tell about the long-time evolution of the system?*

The dynamics of a dynamical system (differential equation or function) is the behavior of the orbits, when the time tends to infinity. The aim of "dynamical systems" is to produce a general procedure for describing the dynamics of any system. For example, Conley's theory presented in the next section organizes the global dymamics of a general system using regions concentrating the orbit accumulation and recurrence and splits these regions in elementary pieces: the chain recurrence classes.

We focus our study on $C^r$-diffeomorphisms $F$ (i.e., $F$ and $F^{-1}$ are $r$ times continuously derivable) on a compact smooth manifold $M$ (most of the notions and results presented here also hold for vector fields). Even for very regular systems ($F$ algebraic) of a low-dimensional space ($\dim(M) = 2$), the dynamics may be chaotic and very unstable: one cannot hope for a precise description of all systems. Furthermore, neither the initial data of a concrete system nor the infinitesimal-evolution rule are known exactly: fragile properties describe the evolution of the theoretical model, and not of the real system. For these reasons, we are mostly interested in properties that are persistent, in some sense, by small perturbations of the dynamical system. The notion of small perturbations of the system requires a topology on the space $\mathrm{Diff}^r(M)$ of $C^r$-diffeomorphisms: two diffeomorphisms are close for the $C^r$-topology if all their partial derivatives of order $\leq r$ are close at each point of $M$. Endowed with this topology, $\mathrm{Diff}^r(M)$ is a complete metric space.

The open and dense subsets of $\mathrm{Diff}^r(M)$ provide the natural topological notion of "almost all" $F$. Genericity is a weaker notion: by Baire's theorem, if $\mathcal{O}_i, i \in \mathbb{N}$, are dense and open subsets, the intersection $\bigcap_{i\in\mathbb{N}} \mathcal{O}_i$ is a dense subset. A subset is called residual if it contains such a countable intersection of dense open subsets. A property $\mathcal{P}$ is generic if it is verified on a residual subset. By a practical abuse of language, one says:

"$C^r$-generic diffeomorphisms verify $\mathcal{P}$"

A countable intersection of residual sets is a residual set. Hence, if $\{\mathcal{P}_i\}, i \in \mathbb{N}$, is a countable family of generic properties, generic diffeomorphisms verify simultanuously all the properties $\mathcal{P}_i$.

A property $\mathcal{P}$ is $C^r$-robust if the set of diffeomorphisms verifying $\mathcal{P}$ is open in $\mathrm{Diff}^r(M)$. A property $\mathcal{P}$ is locally generic if there is an (nonempty) open set $\mathcal{O}$ on which it is generic, that is, there is residual set $\mathcal{R}$ such that $\mathcal{P}$ is verified on $\mathcal{R} \cap \mathcal{O}$.

The properties of generic dynamical systems depend mostly on the dimension of the manifold $M$ and of the $C^r$-topology considered, $r \in \mathbb{N} \cup \{+\infty\}$ (an important problem is that $C^r$-generic diffeomorphisms are not $C^{r+1}$ ):

- On very low dimensional spaces (diffeomorphisms of the circle and vector fields on compact surfaces) the dynamics of generic systems (indeed in a open and dense subset of systems) is very simple (called Morse–Smale) and well understood; see the subsection "Generic properties of the low-dimensional systems."
- In higher dimensions, for $C^r$-topology, $r > 1$, one has generic and locally generic properties related to the periodic orbits, like the Kupka–Smale property (see the subsection "Kupka–Smale theorem") and the Newhouse phenomenon (see the subsection "Local $C^2$-genericity of wild behavior for surface diffeomorphisms"). However, we still do not know if the dynamics of $C^r$-generic diffeomorphisms is well approached by their periodic orbits, so that one is still far from a global understanding of $C^r$-generic dynamics.
- For the $C^1$-topology, perturbation lemmas show that the global dynamics is very well approximated by periodic orbits (see the section "$C^1$-generic systems: global dynamics and periodic orbits"). One then divides generic systems in "tame" systems, with a global dynamics analoguous to hyperbolic dynamics, and "wild" systems, which present infinitely many dynamically independent regions. The notion of dominated splitting (see the section "Hyperbolic properties of $C^1$-generic diffeormorphisms") seems to play an important role in this division.

## Results on General Systems

### Notions of Recurrence

Some regions of $M$ are considered as the heart of the dynamics:

- $\mathrm{Per}(F)$ denotes the set of periodic points $x \in M$ of $F$, that is, $F^n(x) = x$ for some $n > 0$.
- A point $x$ is recurrent if its orbit comes back arbitrarily close to $x$, infinitely many times. $\mathrm{Rec}(F)$ denotes the set of recurrent points.
- The limit set $\mathrm{Lim}(F)$ is the union of all the accumulation points of all the orbits of $F$.

- A point $x$ is "wandering" if it admits a neighborhood $U_x \subset M$ disjoint from all its iterates $F^n(U_x), n > 0$. The nonwandering set $\Omega(F)$ is the set of the nonwandering points.
- $\mathcal{R}(F)$ is the set of chain recurrent points, that is, points $x \in M$ which look like periodic points if we allow small mistakes at each iteration: for any $\varepsilon > 0$, there is a sequence $x = x_0, x_1, \ldots, x_k = x$ where $d(f(x_i), x_{i+1}) < \varepsilon$ (such a sequence is an $\varepsilon$-pseudo-orbit).

A periodic point is recurrent, a recurrent point is a limit point, a limit point is nonwandering, and a nonwandering point is chain recurrent:

$$\mathrm{Per}(F) \subset \mathrm{Rec}(F) \subset \mathrm{Lim}(F) \subset \Omega(F) \subset \mathcal{R}(F)$$

All these sets are invariant under $F$, and $\Omega(F)$ and $\mathcal{R}(F)$ are compact subsets of $M$. There are diffeomorphisms $F$ for which the closures of these sets are distinct:

- A rotation $x \mapsto x + \alpha$ with irrational angle $\alpha \in \mathbb{R}\backslash\mathbb{Q}$ on the circle $S^1 = \mathbb{R}/\mathbb{Z}$ has no periodic points but every point is recurrent.
- The map $x \mapsto x + (1/4\pi)(1 + \cos(2\pi x))$ induces on the circle $S^1$ a diffeomorphism $F$ having a unique fixed point at $x = 1/2$; one verifies that $\Omega(F) = \{1/2\}$ and $\mathcal{R}(F)$ is the whole circle $S^1$.

An invariant compact set $K \subset M$ is transitive if there is $x \in K$ whose forward orbit is dense in $K$. Generic points $x \in K$ have their forward and backward orbits dense in $K$: in this sense, transitive sets are dynamically indecomposable.

#### Conley's Theory: Pairs Attractor/Repeller and Chain Recurrence Classes

A trapping region $U \subset M$ is a compact set whose image $F(U)$ is contained in the interior of $U$. By definition, the intersection $A = \bigcap_{n \geq 0} F^n(U)$ is an attractor of $F$: any orbit in $U$ "goes to $A$." Denote by $V$ the complement of the interior of $U$: it is a trapping region for $F^{-1}$ and the intersection $R = \bigcap_{n \geq 0} F^{-n}(V)$ is a repeller. Each orbit either is contained in $A \cup R$, or "goes from the repeller to the attractor." More precisely, there is a smooth function $\psi: M \to [0, 1]$ (called Lyapunov function) equal to 1 on $R$ and 0 on $A$, and strictly decreasing on the other orbits:

$$\psi(F(x)) < \psi(x) \quad \text{for } x \notin A \cup R$$

So, the chain recurrent set is contained in $A \cup R$. Any compact set contained in $U$ and containing the interior of $F(U)$ is a trapping region inducing the same attracter and repeller pair $(A,R)$; hence, the set of attracter/repeller pairs is countable. We denote by $(A_i, R_i, \psi_i), i \in \mathbb{N}$, the family of these pairs endowed with an associated Lyapunov function. Conley (1978) proved that

$$\mathcal{R}(F) = \bigcap_{i \in \mathbb{N}} (A_i \cup R_i)$$

This induces a natural partition of $\mathcal{R}(F)$ in equivalence classes: $x \sim y$ if $x \in A_i \Leftrightarrow y \in A_i$. Conley proved that $x \sim y$ iff, for any $\varepsilon > 0$, there are $\varepsilon$-pseudo orbits from $x$ to $y$ and vice versa. The equivalence classes for $\sim$ are called chain recurrence classes.

Now, considering an average of the Lyapunov functions $\psi_i$ one gets the following result: there is a continuous function $\varphi: M \to \mathbb{R}$ with the following properties:

- $\varphi(F(x)) \leq \varphi(x)$ for every $x \in M$, (i.e., $\varphi$ is a Lyapunov function);
- $\varphi(F(x)) = \varphi(x) \Leftrightarrow x \in \mathcal{R}(F)$;
- for $x, y \in \mathcal{R}(F)$, $\quad \varphi(x) = \varphi(y) \Leftrightarrow x \sim y$; and
- the image $\varphi(\mathcal{R}(F)$ is a compact subset of $\mathbb{R}$ with empty interior.

This result is called the "fundamental theorem of dynamical systems" by several authors (see Robinson (1999)).

Any orbit is $\varphi$-decreasing from a chain recurrence class to another chain reccurence class (the global dynamics of $F$ looks like the dynamics of the gradient flow of a function $\phi$, the chain recurrence classes supplying the singularities of $\phi$). However, this description of the dynamics may be very rough: if $F$ preserves the volume, Poincaré's recurrence theorem implies that $\Omega(F) = \mathcal{R}(F) = M$; the whole $M$ is the unique chain recurrence class and the function $\varphi$ of Conley's theorem is constant.

Conley's theory provides a general procedure for describing the global topological dynamics of a system: one has to characterize the chain recurrence classes, the dynamics in restriction to each class, the stable set of each class (i.e., the set of points whose positive orbits goes to the class), and the relative positions of these stable sets.

## Hyperbolicity

Smale's hyperbolic theory is the first attempt to give a global vision of almost all dynamical systems. In this section we give a very quick overview of this theory. For further details, *see* Hyperbolic Dynamical Systems.

#### Hyperbolic Periodic Orbits

A fixed point $x$ of $F$ is hyperbolic if the derivative $DF(x)$ has no (neither real nor complex) eigenvalue with modulus equal to 1. The tangent space at $x$

splits as $T_xM = E^s \oplus E^u$, where $E^s$ and $E^u$ are the $DF(x)$-invariant spaces corresponding to the eigenvalues of moduli $<1$ and $>1$, respectively. There are $C^r$-injectively immersed $F$-invariant submanifolds $W^s(x)$ and $W^u(x)$ tangent at $x$ to $E^s$ and $E^u$; the stable manifold $W^s(x)$ is the set of points $y$ whose forward orbit goes to $x$. The implicit-function theorem implies that a hyperbolic fixed point $x$ varies (locally) continuously with $F$; (compact parts of) the stable and unstable manifolds vary continuously for the $C^r$-topology when $F$ varies with the $C^r$-topology.

A periodic point $x$ of period $n$ is hyperbolic if it is a hyperbolic fixed point of $F^n$ and its invariant manifolds are the corresponding invariant manifolds for $F^n$. The stable and unstable manifold of the orbit of $x$, $W^s_{orb}(x)$ and $W^u_{orb}(x)$, are the unions of the invariant manifolds of the points in the orbit.

### Homoclinic Classes

Distinct stable manifolds are always disjoint; however, stable and unstable manifolds may intersect. At the end of the nineteenth century, Poincaré noted that the existence of transverse homoclinic orbits, that is, transverse intersection of $W^s_{orb}(x)$ with $W^u_{orb}(x)$ (other than the orbit of $x$), implies a very rich dynamical behavior: indeed, Birkhoff proved that any transverse homoclinic point is accumulated by a sequence of periodic orbits (see Figure 1). The homoclinic class $H(x)$ of a periodic orbit is the closure of the transverse homoclinic point associated to $x$:

$$H(p) = \overline{W^s_{orb}(x) \pitchfork W^u_{orb}(x)}$$

There is an equivalent definition of the homoclinic class of $x$: we say that two hyperbolic periodic points $x$ and $y$ are homoclinically related if $W^s_{orb}(x)$ and $W^u_{orb}(x)$ intersect transversally $W^u_{orb}(y)$ and $W^s_{orb}(y)$, respectively; this defines an equivalence relation in $\mathrm{Per}_{hyp}(F)$ and the homoclinic classes are the closure of the equivalence classes.

The homoclinic classes are transitive invariant compact sets canonically associated to the periodic orbits. However, for general systems, homoclinic classes are not necessarily disjoint.

For more details, *see* Homoclinic Phenomena.

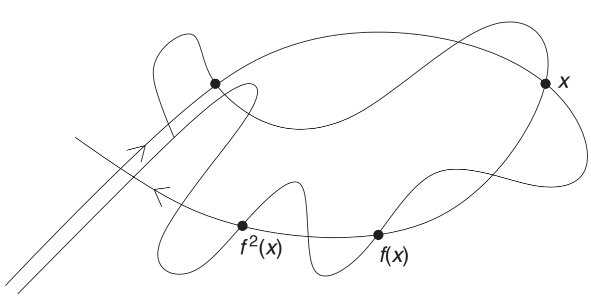



**Figure 1** A transverse homoclinic orbit.

### Smale's Hyperbolic Theory

A diffeomorphism $F$ is *Morse–Smale* if $\Omega(F) = \mathrm{Per}(F)$ is finite and hyperbolic, and if $W^s(x)$ is tranverse to $W^u(y)$ for any $x, y \in \mathrm{Per}(F)$. Morse–Smale diffeomorphisms have a very simple dynamics, similar to the one of the gradient flow of a Morse function; apart from periodic points and invariant manifolds of periodic saddles, each orbit goes from a source to a sink (hyperbolic periodic repellers and attractors). Furthermore, Morse–Smale diffeomorphisms are $C^1$-*structurally stable*, that is, any diffeomorphism $C^1$-close to $F$ is conjugated to $F$ by a homeomorphism: the topological dynamics of $F$ remains unchanged by small $C^1$-perturbation. Morse–Smale vector fields were known (Andronov and Pontryagin, 1937) to characterize the structural stability of vector fields on the sphere $S^2$. However, a diffeomorphism having transverse homoclinic intersections is robustly not Morse–Smale, so that Morse–Smale diffeomorphisms are not $C^r$-dense, on any compact manifold of dimension $\geq 2$. In the early 1960s, Smale generalized the notion of hyperbolicity for nonperiodic sets in order to get a model for homoclinic orbits. The goal of the theory was to cover a whole dense open set of all dynamical systems.

An invariant compact set $K$ is hyperbolic if the tangent space $TM|_K$ of $M$ over $K$ splits as the direct sum $TM_K = E^s \oplus E^u$ of two $DF$-invariant vector bundles, where the vectors in $E^s$ and $E^u$ are uniformly contracted and expanded, respectively, by $F^n$, for some $n > 0$. Hyperbolic sets persist under small $C^1$-perturbations of the dynamics: any diffeomorphism $G$ which is $C^1$-close enough to $F$ admits a hyperbolic compact set $K_G$ close to $K$ and the restrictions of $F$ and $G$ to $K$ and $K_G$ are conjugated by a homeomorphism close to the identity. Hyperbolic compact sets have well-defined invariant (stable and unstable) manifolds, tangent (at the points of $K$) to $E^s$ and $E^u$ and the (local) invariant manifolds of $K_G$ vary locally continuously with $G$.

The existence of hyperbolic sets is very common: if $y$ is a transverse homoclinic point associated to a hyperbolic periodic point $x$, then there is a transitive hyperbolic set containing $x$ and $y$.

Diffeomorphisms for which $\mathcal{R}(F)$ is hyperbolic are now well understood: the chain recurrence classes are homoclinic classes, finitely many, and transitive, and admit a combinatorical model (subshift of finite type). Some of them are

attractors or repellers, and the basins of the attractors cover a dense open subset of $M$. If, furthermore, all the stable and unstable manifolds of points in $\mathcal{R}(f)$ are transverse, the diffeomorphism is $C^1$-structurally stable (Robbin 1971, Robinson 1976); indeed, this condition, called "axiom A + strong transversality," is equivalent to the $C^1$-structural stability (Mañé 1988).

In 1970, Abraham and Smale built examples of robustly non-axiom A diffeomorphisms, when $\dim M \geq 3$: the dream of a global understanding of dynamical systems was postponed. However, hyperbolicity remains a key tool in the study of dynamical systems, even for nonhyperbolic systems.

## $C^r$-Generic Systems

### Kupka–Smale Theorem

Thom's transversality theorem asserts that two submanifolds can always be put in tranverse position by a $C^r$-small perturbations. Hence, for $F$ in an open and dense subset of $\mathrm{Diff}^r(M), r \geq 1$, the graph of $F$ in $M \times M$ is transverse to the diagonal $\Delta = \{(x,x), x \in M\}$: $F$ has finitely many fixed points $x_i$, depending locally continuously on $F$, and 1 is not an eigenvalue of the differential $DF(x_i)$. Small local perturbations in the neighborhood of the $x_i$ avoid eigenvalue of modulus equal to 1: one gets a dense and open subset $\mathcal{O}_1^r$ of $\mathrm{Diff}^r(M)$ such that every fixed point is hyperbolic. This argument, adapted for periodic points, provides a dense and open set $\mathcal{O}_n^r \subset \mathrm{Diff}^r(M)$, such that every periodic point of period $n$ is hyperbolic. Now $\bigcap_{n\in\mathbb{N}} \mathcal{O}_n^r$ is a residual subset of $\mathrm{Diff}^r(M)$, for which every periodic point is hyperbolic.

Similarly, the set of diffeomorphisms $F \in \bigcap_{i=0}^n \mathcal{O}_i^r(M)$ such that all the disks of size $n$, of invariant manifolds of periodic points of period less that $n$, are pairwise transverse, is open and dense. One gets the Kupka–Smale theorem (see Palis and de Melo (1982) for a detailed exposition): *for $C^r$-generic diffeomorphisms $F \in \mathrm{Diff}^r(M)$, every periodic orbit is hyperbolic and $W^s(x)$ is transverse to $W^u(y)$ for $x, y \in Per(F)$.*

### Generic Properties of Low-Dimensional Systems

Poincaré–Denjoy theory describes the topological dynamics of all diffeomorphisms of the circle $S^1$ (*see* Homeomorphisms and Diffeomorphisms of the Circle). Diffeomorphisms in an open and dense subset of $\mathrm{Diff}_+^r(S^1)$ have a nonempty finite set of periodic orbits, all hyperbolic, and alternately attracting (sink) or repelling (source). The orbit of a nonperiodic point comes from a source and goes to a sink. Two $C^r$-generic diffeomorphisms of $S^1$ are conjugated iff they have same rotation number and same number of periodic points.

This simple behavior has been generalized in 1962 by Peixoto for vector fields on compact orientable surfaces $S$. Vector fields $X$ in a $C^r$-dense and open subset are Morse–Smale, hence structurally stable (see Palis and de Melo (1982) for a detailed proof). Peixoto gives a complete classification of these vector fields, up to topological equivalence.

Peixoto's argument uses the fact that the return maps of the vector field on transverse sections are increasing functions: this helped control the effect on the dynamics of small "monotonous" perturbations, and allowed him to destroy any nontrivial recurrences. Peixoto's result remains true on nonorientable surfaces for the $C^1$-topology but remains an open question for $r > 1$: is the set of Morse–Smale vector fields $C^2$-dense, for $S$ nonorientable closed surface?

### Local $C^2$-Genericity of Wild Behavior for Surface Diffeomorphisms

The generic systems we have seen above have a very simple dynamics, simpler than the general systems. This is not always the case. In the 1970s, Newhouse exhibited a $C^2$-open set $\mathcal{O} \subset \mathrm{Diff}^2(S^2)$ (where $S^2$ denotes the two-dimensional sphere), such that $C^2$-generic diffeomorphisms $F \in \mathcal{O}$ have infinitely many hyperbolic periodic sinks. In fact, $C^2$-generic diffeomorphisms in $\mathcal{O}$ present many other pathological properties: for instance, it has been recently noted that they have uncountably many chain recurrence classes without periodic orbits. Densely (but not generically) in $\mathcal{O}$, they present many other phenomena, such as strange (Henon-like) attractors (*see* Lyapunov Exponents and Strange Attractors).

This phenomenon appears each time that a diffeomorphism $F_0$ admits a hyperbolic periodic point $x$ whose invariant manifolds $W^s(x)$ and $W^u(x)$ are tangent at some point $p \in W^s(x) \cap W^u(x)$ ($p$ is a homoclinic tangency associated to $x$). Homoclinic tangencies appear locally as a codimension-1 submanifold of $\mathrm{Diff}^2(S^2)$; they are such a simple phenomenon that they appear in very natural contexts. When a small perturbation transforms the tangency into tranverse intersections, a new hyperbolic set $K$ with very large fractal dimensions is created. The local stable and unstable manifolds of $K$, each homeomorphic to the product of a Cantor set by a segment, present tangencies in a $C^2$-robust way, that is, for $F$ in some $C^2$-open set $\mathcal{O}$ (see **Figure 2**). As a consequence, for a $C^2$-dense subset of $\mathcal{O}$, the

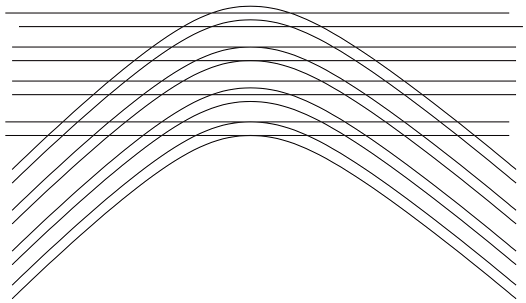


**Figure 2** Robust tangencies.

invariant manifolds of the point $x$ present some tangency (this is not generic, by Kupka–Smale theorem). If the Jacobian of $F$ at $x$ is $<1$, each tangency allows to create one more sink, by an arbitrarily small perturbation. Hence, the sets of diffeomorphisms having more than $n$ hyperbolic sinks are dense open subsets of $\mathcal{O}$, and the intersection of all these dense open subsets is the announced residual set. See Palis and Takens (1993) for details on this deep argument.

## $C^1$-Generic Systems: Global Dynamics and Periodic Orbits

See Bonatti *et al.* (2004), Chapter 10 and Appendix A, for a more detailed exposition and precise references.

### Perturbations of Orbits: Closing and Connecting Lemmas

In 1968, Pugh proved the following Lemma.

**Closing lemma** *If $x$ is a nonwandering point of a diffeomorphism $F$, then there are diffeomorphisms $G$ arbitrarily $C^1$-close to $F$, such that $x$ is periodic for $G$.*

Consider a segment $x_0, \ldots, x_n = F^n(x_0)$ of orbit such that $x_n$ is very close to $x_0 = x$; one would like to take $G$ close to $F$ such that $G(x_n) = x_0$, and $G(x_i) = F(x_i) = x_{i+1}$ for $i \neq n$. This idea works for the $C^0$-topology (so that the $C^0$-closing lemma is easy). However, if one wants $G$ $\varepsilon$-$C^1$-close to $F$, one needs that the points $x_i, i \in \{1, \ldots, n-1\}$, remain at distance $d(x_i, x_0)$ greater than $C(d(x_n, x_0)/\varepsilon)$, where $C$ bounds $\|Df\|$ on $M$. If $C/\varepsilon$ is very large, such a segment of orbit does not exist. Pugh solved this difficulty in two steps: the perturbation is first spread along a segment of orbit of $x$ in order to decrease this constant; then a subsegment $y_0, \ldots, y_k$ of $x_0, \ldots, x_n$ is selected, verifying the geometrical condition.

For the $C^2$ topology, the distances $d(x_i, x_0)$ need to remain greater than $\sqrt{d(x_n, x_0)/\varepsilon} \gg d(x_n, x_0)$. This new difficulty is why the $C^2$-closing lemma remains an open question.

Pugh's argument does not suffice to create homoclinic point for a periodic orbit whose unstable manifold accumulates on the stable one. In 1998, Hayashi solved this problem proving the

**Connecting lemma** (Hayashi 1997) *Let $y$ and $z$ be two points such that the forward orbit of $y$ and the backward orbit of $z$ accumulate on the same nonperiodic point $x$. Fix some $\varepsilon > 0$. There is $N > 0$ and a $\varepsilon$-$C^1$-perturbation $G$ of $F$ such that $G^n(y) = z$ for some $n > 0$, and $G \equiv F$ out of an arbitrary small neighborhood of $\{x, F(x), \ldots, F^N(x)\}$.*

Using Hayashi's arguments, we (with Crovisier) proved the following lemma:

**Connecting lemma for pseudo-orbits** (Bonatti and Crovisier 2004) *Assume that all periodic orbits of $F$ are hyperbolic; consider $x, y \in M$ such that, for any $\varepsilon > 0$, there are $\varepsilon$-pseudo-orbits joining $x$ to $y$; then there are arbitrarily small $C^1$-perturbations of $F$ for which the positive orbit of $x$ passes through $y$.*

### Densities of Periodic Orbits

As a consequence of the perturbations lemma above, we (Bonatti and Crovisier 2004) proved that for $F$ $C^1$-generic,

$$\mathcal{R}(F) = \Omega(F) = \overline{\mathrm{Per}_{\mathrm{hyp}}(F)}$$

where $\overline{\mathrm{Per}_{\mathrm{hyp}}(F)}$ denotes the closure of the set of hyperbolic periodic points.

For this, consider the map $\Psi: F \mapsto \Psi(F) = \overline{\mathrm{Per}_{\mathrm{hyp}}(F)}$ defined on $\mathrm{Diff}^1(M)$ and with value in $\mathcal{K}(M)$, space of all compact subsets of $M$, endowed with the Hausdorff topology. $\mathrm{Per}_{\mathrm{hyp}}(F)$ may be approximated by a finite set of hyperbolic periodic points, and this set varies continuously with $F$; so $\mathrm{Per}_{\mathrm{hyp}}(F)$ varies lower-semicontinuously with $F$: for $G$ very close to $F$, $\mathrm{Per}_{\mathrm{hyp}}(G)$ cannot be very much smaller than $\mathrm{Per}_{\mathrm{hyp}}(F)$. As a consequence, a result from general topology asserts that, for $C^1$-generic $F$, the map $\Psi$ is continuous at $F$. On the other hand, $C^1$-generic diffeomorphisms are Kupka–Smale, so that the connecting lemma for pseudo-orbits may apply: if $x \in \mathcal{R}(F), x$ can be turned into a hyperbolic periodic point by a $C^1$-small perturbation of $F$. So, if $x \notin \overline{\mathrm{Per}_{\mathrm{hyp}}(F)}, F$ is not a continuity point of $\Psi$, leading to a contradiction.

Furthermore, Crovisier proved the following result: "for $C^1$-generic diffeomorphisms, each chain recurrence class is the limit, for the Hausdorff distance, of a sequence of periodic orbits."

This good approximation of the global dynamics by the periodic orbits will now allow us to better understand the chain recurrence classes of $C^1$-generic diffeomorphisms.

### Chain Recurrence Classes/Homoclinic Classes of $C^1$-Generic Systems

Tranverse intersections of invariant manifolds of hyperbolic orbits are robust and vary locally continuously with the diffeomorphisms $F$. So, the homoclinic class $H(x)$ of a periodic point $x$ varies lower-semicontinuously with $F$ (on the open set where the continuation of $x$ is defined). As a consequence, for $C^r$-generic diffeomorphisms ($r \geq 1$), each homoclinic class varies continuously with $F$. Using the connecting lemma, Arnaud (2001) proved the following result: "for Kupka–Smale diffeomorphisms, if the closures $\overline{W^u_{orb}}(x)$ and $\overline{W^s_{orb}}(x)$ have some intersection point $z$, then a $C^1$-pertubation of $F$ creates a tranverse intersection of $\overline{W^u_{orb}}(x)$ and $\overline{W^s_{orb}}(x)$ at $z$." So, if $z \notin H(x)$, then $F$ is not a continuity point of the function $F \mapsto H(x, F)$. Hence, for $C^1$-generic diffeomorphisms $F$ and for every periodic point $x$,

$$H(x) = \overline{W^u_{orb}}(x) \cap \overline{W^s_{orb}}(x)$$

In the same way, $\overline{W^u_{orb}}(x)$ and $\overline{W^s_{orb}}(x)$ vary locally lower-semicontinuously with $F$ so that, for $F$ $C^r$-generic, the closures of the invariant manifolds of each periodic point vary locally continuously. For Kupka–Smale diffeomorphisms, the connecting lemma for pseudo-orbits implies: "if $z$ is a point in the chain recurrence class of a periodic point $x$, then a $C^1$-small perturbation of $F$ puts $z$ on the unstable manifold of $x$"; so, if $z \notin \overline{W^u_{orb}}(x)$, then $F$ is not a continuity point of the function $F \mapsto \overline{W^u_{orb}}(x, F)$. Hence, for $C^1$-generic diffeomorphisms $F$ and for every periodic point $x$, the chain recurrence class of $x$ is contained in $\overline{W^u_{orb}}(x) \cap \overline{W^s_{orb}}(x)$, and, therefore, coincides with the homoclinic class of $x$. This argument proves:

> For a $C^1$-generic diffeomorphism $F$, each homoclinic class $H(x)$ is a chain recurrence class of $F$ (of Conley's theory): a chain recurrence class containing a periodic point $x$ coincides with the homoclinic class $H(x)$. In particular, two homoclinic classes are either disjoint or equal.

### Tame and Wild Systems

For generic diffeomorphisms, the number $N(F) \in \mathbb{N} \cup \{\infty\}$ of homoclinic classes varies lower-semicontinuously with $F$. One deduces that $N(F)$ is locally constant on a residual subset of $\text{Diff}^1(M)$ (Abdenur 2003).

A local version (in the neighborhood of a chain recurrence class) of this argument shows that, for $C^1$-generic diffeomorphisms, any isolated chain recurrence classe $C$ is robustly isolated: for any diffeomorphism $G$, $C^1$-close enough to $F$, the intersection of $\mathcal{R}(G)$ with a small neighborhood of $C$ is a unique chain recurrence class $C_G$ close to $C$.

One says that a diffeomorphism is "tame" if each chain recurrence class is robustly isolated. We denote by $\mathcal{T}(M) \subset \text{Diff}^1(M)$ the ($C^1$-open) set of tame diffeomorphisms and by $\mathcal{W}(M)$ the complement of the closure of $\mathcal{T}(M)$. $C^1$-generic diffeomorphisms in $\mathcal{W}(M)$ have infinitely many disjoint homoclinic classes, and are called "wild" diffeomorphisms.

Generic tame diffeomorphisms have a global dynamics analogous to hyperbolic systems: the chain recurrence set admits a partition into finitely many homoclinic classes varying continuously with the dynamics. Every point belongs to the stable set of one of these classes. Some of the homoclinic classes are (transitive) topological attractors, and the union of the basins covers a dense open subset of $M$, and the basins vary continuously with $F$ (Carballo Morales 2003). It remains to get a good description of the dynamics in the homoclinic classes, and particularly in the attractors. As we shall see in the next section, tame behavior requires some kind of weak hyperbolicity. Indeed, in dimension 2, tame diffeomorphisms satisfy axiom A and the noncycle condition.

As of now, very little is known about wild systems. One knows some semilocal mechanisms generating locally $C^1$-generic wild dynamics, therefore proving their existence on any manifold with dimension $\dim(M) \geq 3$ (the existence of wild diffeomorphisms in dimension 2, for the $C^1$-topology, remains an open problem). Some of the known examples exhibit a universal dynamics: they admit infinitely many disjoint periodic disks such that, up to renormalization, the return maps on these disks induce a dense subset of diffeomorphisms of the disk. Hence, these locally generic diffeomorphisms present infinitely many times any robust property of diffeomorphisms of the disk.

### Ergodic Properties

A point $x$ is well closable if, for any $\varepsilon > 0$ there is $G$ $\varepsilon$-$C^1$-close to $F$ such that $x$ is periodic for $G$ and $d(F^i(x), G^i(x)) < \varepsilon$ for $i \in \{0, \ldots, p\}$, $p$ being the period of $x$. As an important refinement of Pugh's closing lemma, Mañé proved the following lemma:

**Ergodic closing lemma** *For any F-invariant probability, almost every point is well closable.*

As a consequence, “for $C^1$-generic diffeomorphisms, any ergodic measure $\mu$ is the weak limit of a sequence of Dirac measures on periodic orbits, which converges also in the Hausdorff distance to the support of $\mu$.”

It remains an open problem to know if, for $C^1$-generic diffeomorphisms, the ergodic measures supported in a homoclinic class are approached by periodic orbits in this homoclinic class.

### Conservative Systems

The connecting lemma for pseudo-orbits has been adapted for volume preserving and symplectic diffeomorphisms, replacing the condition on the periodic orbits by another generic condition on the eigenvalues. As a consequence, one gets: “$C^1$-generic volume-preserving or symplectic diffeomorphisms are transitive, and $M$ is a unique homoclinic class.”

Notice that the KAM theory implies that this result is wrong for $C^4$-generic diffeomorphisms, the persistence of invariant tori allowing to break robustly the transitivity.

The Oxtoby–Ulam (1941) theorem asserts that $C^0$-generic volume-preserving homeomorphisms are ergodic. The ergodicity of $C^1$-generic volume-preserving diffeomorphisms remains an open question.

## Hyperbolic Properties of $C^1$-Generic Diffeomorphisms

For a more detailed exposition of hyperbolic properties of $C^1$-generic diffeomorphisms, the reader is referred to Bonatti *et al.* (2004, chapter 7 and appendix B).

### Perturbations of Products of Matrices

The $C^1$-topology enables us to do small perturbations of the differential $DF$ at a point $x$ without perturbing either $F(x)$ or $F$ out of an arbitrarily small neighborhood of $x$. Hence, one can perturb the differential of $F$ along a periodic orbit, without changing this periodic orbit (Frank’s lemma). When $x$ is a periodic point of period $n$, the differential of $F^n$ at $x$ is fundamental for knowing the local behavior of the dynamics. This differential is (up to a choice of local coordinates) a product of the matrices $DF(x_i)$, where $x_i = F^i(x)$. So, the control of the dynamical effect of local perturbations along a periodic orbit comes from a problem of linear algebra: “consider a product $\mathcal{A} = A_n \circ A_{n-1} \circ \cdots \circ A_1$ of $n \gg 0$ bounded linear ismorphisms of $\mathbb{R}^d$; how do the eigenvalues and the eigenspaces of $\mathcal{A}$ vary under small perturbations of the $A_i$?”

A partial answer to this general problem uses the notion of dominated splitting. Let $X \subset M$ be an $F$-invariant set such that the tangent space of $M$ at the points $x \in X$ admits a $DF$-invariant splitting $T_x(M) = E_1(x) \oplus \cdots E_k(x)$, the dimensions $\dim(E_i(x))$ being independent of $x$. This splitting is dominated if the vectors in $E_{i+1}$ are uniformly more expanded than the vectors in $E_i$: there exists $\ell > 0$ such that, for any $x \in X$, any $i \in \{1, \ldots, k-1\}$ and any unit vectors $u \in E_i(x)$ and $v \in E_{i+1}(x)$, one has

$$\|DF^\ell(u)\| < \tfrac{1}{2}\|DF^\ell(v)\|$$

Dominated splittings are always continuous, extend to the closure of $X$, and persist and vary continuously under $C^1$-perturbation of $F$.

### Dominated Splittings versus Wild Behavior

Let $\{\gamma_i\}$ be a set of hyperbolic periodic orbits. On $X = \bigcup \gamma_i$ one considers the natural splitting $TM|_X = E^s \oplus E^u$ induced by the hyperbolicity of the $\gamma_i$. Mañé (1982) proved: “if there is a $C^1$-neighborhood of $F$ on which each $\gamma_i$ remains hyperbolic, then the splitting $TM|_X = E^s \oplus E^u$ is dominated.”

A generalization of Mañé’s result shows: “if a homoclinic class $H(x)$ has no dominated splitting, then for any $\varepsilon > 0$ there is a periodic orbit $\gamma$ in $H(x)$ whose derivative at the period can be turned into an homothety, by an $\varepsilon$-small perturbation of the derivative of $F$ along the points of $\gamma$”; in particular, this periodic orbit can be turned into a sink or a source. As a consequence, one gets: “for $C^1$-generic diffeomorphisms $F$, any homoclinic class either has a dominated splitting or is contained in the closure of the (infinite) set of sinks and sources.”

This argument has been used in two directions:

- Tame systems must satisfy some hyperbolicity. In fact, using the ergodic closing lemma, one proves that the homoclinic classes $H(x)$ of tame diffeomorphisms are volume hyperbolic, that is, there is a dominated splitting $TM = E_1 \oplus \cdots \oplus E_k$ over $H(x)$ such that $DF$ contracts uniformly the volume in $E_1$ and expands uniformly the volume in $E_k$.
- If $F$ admits a homoclinic class $H(x)$ which is robustly without dominated splittings, then generic diffeomorphisms in the neighborhood of $F$ are wild: at this time this is the unique known way to get wild systems.

*See also:* Cellular Automata; Chaos and Attractors; Fractal Dimensions in Dynamics; Homeomorphisms and Diffeomorphisms of the Circle; Homoclinic Phenomena; Hyperbolic Dynamical Systems; Lyapunov Exponents and Strange Attractors; Polygonal Billiards; Singularity and Bifurcation Theory; Synchronization of Chaos.

## Further Reading


Abdenur F (2003) Generic robustness of spectral decompositions. *Annales Scientifiques de l'Ecole Normale Superieure IV* 36(2): 213–224.

Andronov A and Pontryagin L (1937) Systèmes grossiers. *Dokl. Akad. Nauk. USSR* 14: 247–251.

Arnaud M-C (2001) Création de connexions en topologie $C^1$. *Ergodic Theory and Dynamical Systems* 21(2): 339–381.

Bonatti C and Crovisier S (2004) Récurrence et généricité. *Inventiones Mathematical* 158(1): 33–104 (French) (see also the short English version: (2003) Recurrence and genericity. *Comptes Rendus Mathematique de l'Academie des Sciences Paris* 336 (10): 839–844).

Bonatti C, Dìaz LJ, and Viana M (2004) Dynamics Beyond Uniform Hyperbolicity. *Encyclopedia of Mathematical Sciences*, vol. 102. Berlin: Springer.

Carballo CM and Morales CA (2003) Homoclinic classes and finitude of attractors for vector fields on *n*-manifolds. *Bulletins of the London Mathematical Society* 35(1): 85–91.

Hayashi S (1997) Connecting invariant manifolds and the solution of the $C^1$ stability and $\Omega$-stability conjectures for flows. *Annals of Mathematics* 145: 81–137.

Mañé R (1988) A proof of the $C^1$ stability conjecture. *Inst. Hautes Études Sci. Publ. Math.* 66: 161–210.

Palis J and de Melo W (1982) *Gemetric Theory of Dynamical Systems, An Introduction*. New York–Berlin: Springer.

Palis J and Takens F (1993) *Hyperbolicity and Sensitive-Chaotic Dynamics at Homoclinic Bifurcation*. Cambridge: Cambridge University Press.

Robbin JW (1971) A structural stability theorem. *Annals of Mathematics* 94(2): 447–493.

Robinson C (1974) Structural stability of vector fields. *Annals of Mathematics* 99(2): 154–175 (errata. *Annals of Mathematics* 101(2): 368 (1975)).

Robinson C (1976) Structural stability of $C^1$ diffeomorphisms. *Journal of Differential Equations* 22(1): 28–73.

Robinson C (1999) Dynamical Systems. *Studies in Advanced Mathematics*. Boca Raton, FL: CRC Press.


# Geometric Analysis and General Relativity

**L Andersson**, University of Miami, Coral Gables, FL, USA and Albert Einstein Institute, Potsdam, Germany



Geometric analysis can be said to originate in the nineteenth century work of Weierstrass, Riemann, Schwarz, and others on minimal surfaces, a problem whose history can be traced at least as far back as the work of Meusnier and Lagrange in the eighteenth century. The experiments performed by Plateau in the mid-19th century, on soap films spanning wire contours, served as an important inspiration for this work, and led to the formulation of the Plateau problem, which concerns the existence and regularity of area-minimizing surfaces in $\mathbb{R}^3$ spanning a given boundary contour. The Plateau problem for area-minimizing disks spanning a curve in $\mathbb{R}^3$ was solved by J Douglas (who shared the first Fields medal with Lars V Ahlfors) and T Rado in the 1930s. Generalizations of Plateau's problem have been an important driving force behind the development of modern geometric analysis. Geometric analysis can be viewed broadly as the study of partial differential equations arising in geometry, and includes many areas of the calculus of variations, as well as the theory of geometric evolution equations. The Einstein equation, which is the central object of general relativity, is one of the most widely studied geometric partial differential equations, and plays an important role in its Riemannian as well as in its Lorentzian form, the Lorentzian being most relevant for general relativity.

The Einstein equation is the Euler–Lagrange equation of a Lagrangian with gauge symmetry and thus in the Lorentzian case it, like the Yang–Mills equation, can be viewed as a system of evolution equations with constraints. After imposing suitable gauge conditions, the Einstein equation becomes a hyperbolic system, in particular using spacetime harmonic coordinates (also known as wave coordinates), the Einstein equation becomes a quasilinear system of wave equations. The constraint equations implied by the Einstein equations can be viewed as a system of elliptic equations in terms of suitably chosen variables. Thus, the Einstein equation leads to both elliptic and hyperbolic problems, arising from the constraint equations and the Cauchy problem, respectively. The groundwork for the mathematical study of the Einstein equation and the global nature of spacetimes was laid by, among others, Choquet-Bruhat, who proved local well-posedness for the Cauchy problem, Lichnerowicz, and later York who provided the basic ideas for the analysis of the constraint equations, and Leray who formalized the notion of global hyperbolicity, which is essential for the global study of spacetimes. An important framework for the mathematical study of the Einstein equations has been provided by the singularity theorems of Penrose and Hawking, as well as the cosmic censorship conjectures of Penrose.

Techniques and ideas from geometric analysis have played, and continue to play, a central role in recent mathematical progress on the problems posed by general relativity. Among the main results are the

proof of the positive mass theorem using the minimal surface technique of Schoen and Yau, and the spinor-based approach of Witten, as well as the proofs of the (Riemannian) Penrose inequality by Huisken and Illmanen, and Bray. The proof of the Yamabe theorem by Schoen has played an important role as a basis for constructing Cauchy data using the conformal method.

The results just mentioned are all essentially Riemannian in nature, and do not involve study of the Cauchy problem for the Einstein equations. There has been great progress recently concerning global results on the Cauchy problem for the Einstein equations, and the cosmic censorship conjectures of Penrose. The results available so far are either small data results (among these the nonlinear stability of Minkowski space proved by Christodoulou and Klainerman) or assume additional symmetries, such as the recent proof by Ringström of strong cosmic censorship for the class of Gowdy spacetimes. However, recent progress concerning quasilinear wave equations and the geometry of spacetimes with low regularity due to, among others, Klainerman and Rodnianski, and Tataru and Smith, appears to show the way towards an improved understanding of the Cauchy problem for the Einstein equations.

Since the constraint equations, the Penrose inequality and the Cauchy problem are discussed in separate articles, the focus of this article will be on the role in general relativity of "critical" and other geometrically defined submanifolds and foliations, such as minimal surfaces, marginally trapped surfaces, constant mean curvature hypersurfaces and null hypersurfaces. In this context it would be natural also to discuss geometrically defined flows such as mean curvature flows, inverse mean curvature flow, and Ricci flow. However, this article restricts the discussion to mean curvature flows, since the inverse mean curvature flow appears naturally in the context of the Penrose inequality and the Ricci flow has so far mainly served as a source of inspiration for research on the Einstein equations rather than an important tool. Other topics which would fit well under the heading "General relativity and geometric analysis" are spin geometry (the Witten proof of the Positive mass theorem), the Yamabe theorem and related results concerning the Einstein constraint equations, gluing and other techniques of "spacetime engineering." These are all discussed in other articles. Some techniques which have only recently come into use and for which applications in general relativity have not been much explored, such as Cheeger–Gromov compactness, are not discussed.

## Minimal and Related Surfaces

Consider a hypersurface $N$ in Euclidean space $\mathbb{R}^n$ which is a graph $x_n = u(x_1, \ldots, x_{n-1})$ with respect to the function $u$. The area of $N$ is given by $\mathcal{A}(N) = \int \sqrt{1 + |Du|^2}\, dx^1 \cdots dx^{n-1}$. $N$ is stationary with respect to $\mathcal{A}$ if $u$ satisfies the equation

$$\sum_i D_i \left( \frac{D_i u}{\sqrt{1 + |Du|^2}} \right) = 0 \qquad [1]$$

A hypersurface $N$ defined as a graph of $u$ solving [1] minimizes area with respect to compactly supported deformations, and hence is called a minimal surface. For $n \leq 7$, a solution to eqn [1] defined on all of $\mathbb{R}^{n-1}$ must be an affine function. This fact is known as a Bernstein principle. Equation [1], and more generally, the prescribed mean curvature equation which will be discussed below, is a quasilinear, uniformly elliptic second-order equation. The book by Gilbarg and Trudinger (1983) is an excellent general reference for such equations.

The theory of rectifiable currents, developed by Federer and Fleming, is a basic tool in the modern approach to the Plateau problem and related variational problems. A rectifiable current is a countable union of Lipschitz submanifolds, counted with integer multiplicity, and satisfying certain regularity conditions. Hausdorff measure gives a notion of area for these objects. One may therefore approach the study of minimal surfaces via rectifiable currents which are stationary with respect to variations of area. Suitable generalizations of familiar notions from smooth differential geometry such as tangent plane, normal vector, extrinsic curvature can be introduced. The book by Federer (1969) is a classic treatise on the subject. Further information concerning minimal surfaces and related variational problems can be found in Lawson, Jr. (1980) and Simon (1997). Note, however, that unless otherwise stated, all fields and manifolds considered in this article are assumed to be smooth. For the Plateau problem in a Riemannian ambient space, we have the following existence and regularity result.

**Theorem 1** (Existence of embedded solutions for Plateau problem). *Let $M$ be a complete Riemannian manifold of dimension $n \leq 7$ and let $\Gamma$ be a compact $(n-2)$-dimensional submanifold in $M$ which bounds. Then there is an $(n-1)$-dimensional area-minimizing hypersurface $N$ with $\Gamma$ as its boundary. $N$ is a smooth, embedded manifold in its interior.*

If the dimension of the ambient space is $>7$, solutions to the Plateau problem will in general have a singular set of dimension $n-8$. Let $N$ be an oriented hypersurface of a Riemannian manifold $M$

with covariant derivative $D$. Let $\eta$ be the unit normal of $N$ and define the second fundamental form and mean curvature of $N$ by $A_{ij} = \langle D_{e_i}\eta, e_j\rangle$ and $H = \mathrm{tr}A$. Define the action functional $\mathcal{E}(N) = \mathcal{A}(N) - \int_{M;N} H_0$, where $H_0$ is a function defined on $M$, and $\int_{M;N}$ denotes the integral over the volume bounded by $N$ in $M$. The problem of minimizing $\mathcal{E}$ is a useful generalization of the minimization problem for $\mathcal{A}$.

**Theorem 2** (Existence of minimizers in homology). *Let M be a compact Riemannian manifold of dimension* $\leq 7$, *and let* $\alpha$ *be an integral homology class on M of codimension* 1. *Then there is a smooth minimizer for* $\mathcal{E}$ *representing* $[\alpha]$.

Again, in higher dimensions, the minimizers will in general have singularities. The general form of this result deals with elliptic functionals. For surfaces in 3-manifolds, the problem of minimizing area within homotopy classes has been studied. Results in this direction played a central role in the approach of Schoen and Yau to manifolds with non-negative scalar curvature.

If $M$ is not compact, it is in general necessary to use barriers to control the minimizers, or consider some version of the Plateau problem. Barriers can be used due to the strong maximum principle, which holds for the mean curvature operator since it is quasilinear elliptic. Consider two hypersurfaces $N_1, N_2$ which intersect at a point $p$ and assume that $N_1$ lies on one side of $N_2$ with the normal pointing towards $N_1$. If the mean curvatures $H_1, H_2$ of the hypersurfaces, defined with respect to consistently oriented normals, satisfy $H_1 \leq \lambda \leq H_2$ for some constant $\lambda$, then $N_1$ and $N_2$ coincide near $p$ and have mean curvatures equal to $\lambda$. This result requires only mild regularity conditions on the hypersurfaces. Generalizations hold also for the case of spacelike or null hypersurfaces in a Lorentzian ambient space, see Andersson *et al.* (1998) and Galloway (2000).

Let $\phi$ be a smooth compactly supported function on $N$. The variation $\mathcal{E}' = \delta_{\phi\eta}\mathcal{E}$ of $\mathcal{E}$ under a deformation $\phi\eta$ is

$$\mathcal{E}' = \int_N \phi(H - H_0)$$

Thus, $N$ is stationary with respect to $\mathcal{E}$ if and only if $N$ solves the prescribed mean curvature equation $H(x) = H_0(x)$ for $x \in N$. Supposing that $N$ is stationary and $H_0$ is constant, the second variation $\mathcal{E}'' = \delta_{\phi\eta}\mathcal{E}'$ of $\mathcal{E}$ is of the form

$$\mathcal{E}'' = \int_N \phi(J\phi)$$

where $J$ is the second-variation operator, a second-order elliptic operator. A calculation, using the Gauss equation and the second-variation equation shows

$$J\phi = -\Delta_N\phi - \tfrac{1}{2}[(\mathrm{Scal}_M - \mathrm{Scal}_N) + H^2 + |A|^2]\phi \quad [2]$$

where $\Delta_N, \mathrm{Scal}_M, \mathrm{Scal}_N$ denote the Laplace–Beltrami operator of $N$, and the scalar curvatures of $M$ and $N$, respectively. If $J$ is positive semidefinite, $N$ is called stable.

To set the context where we will apply the above, let $(M, g_{ij})$ be a connected, asymptotically Euclidean three-dimensional Riemannian manifold with covariant derivative, and let $k_{ij}$ be a symmetric tensor on $M$. Suppose $(M, g_{ij}, K_{ij})$ is imbedded isometrically as a spacelike hypersurface in a spacetime $(V, \gamma_{\alpha\beta})$ with $g_{ij}, K_{ij}$ the first and second fundamental forms induced on $M$ from $V$, in particular $K_{ij} = \langle D_{e_i}T, e_j\rangle$ where $T$ is the timelike normal of $M$ in the ambient spacetime $V$, and $D$ is the ambient covariant derivative. We will refer to $(M, g_{ij}, K_{ij})$ as a Cauchy data set for the Einstein equations. Although many of the results which will be discussed below generalize to the case of a nonzero cosmological constant $\Lambda$, we will discuss only the case $\Lambda = 0$ in this article. $G_{\alpha\beta} = \mathrm{Ric}_{V\alpha\beta} - (1/2)\mathrm{Scal}_V\gamma_{\alpha\beta}$ be the Einstein tensor of $V$, and let $\rho = G_{\alpha\beta}T^\alpha T^\beta, \mu_j = G_{j\alpha}T^\alpha$. Then the fields $(g_{ij}, K_{ij})$ satisfy the Einstein constraint equations

$$R + \mathrm{tr}\ K^2 - |K|^2 = 2\rho \quad [3]$$

$$\nabla_j \mathrm{tr}\ K - \nabla^i K_{ij} = \mu_j \quad [4]$$

We assume that the dominant energy condition (DEC)

$$\rho \geq \left(\sum_i \mu_i\mu^i\right)^{1/2} \quad [5]$$

holds. We will sometimes make use of the null energy condition (NEC), $G_{\alpha\beta}L^\alpha L^\beta \geq 0$ for null vectors $L$, and the strong energy condition (SEC), $\mathrm{Ric}_{V\alpha\beta}v^\alpha v^\beta \geq 0$ for causal vectors $v$. $M$ will be assumed to satisfy the fall-off conditions

$$g_{ij} = \left(1 + \frac{2m}{r}\right)\delta_{ij} + O(1/r^2) \quad [6a]$$

$$K_{ij} = O(1/r^2) \quad [6b]$$

as well as suitable conditions for the fall-off of derivatives of $g_{ij}, K_{ij}$. Here $m$ is the ADM (Arnowitt, Deser, Misner) mass of $(M, g_{ij}, K_{ij})$.

## Minimal Surfaces and Positive Mass

Perhaps the most important application of the theory of minimal surfaces in general relativity is in the

Schoen–Yau proof of the positive-mass theorem, which states that $m \geq 0$, and $m = 0$ only if $(M, g, K)$ can be embedded as a hypersurface in Minkowski space. Consider an asymptotically Euclidean manifold $(M, g)$ with $g$ satisfying [6a] and with non-negative scalar curvature. By using Jang's equation, see below, the general situation is reduced to the case of a time symmetric data set, with $K = 0$. In this case, the DEC implies that $(M, g)$ has non-negative scalar curvature.

Assuming $m < 0$ one may, after applying a conformal deformation, assume that $\mathrm{Scal}_M > 0$ in the complement of a compact set. Due to the asymptotic conditions, level sets for sufficiently large values of one of the coordinate functions, say $x^3$, can be used as barriers for minimal surfaces in $M$. By solving a sequence of Plateau problems with boundaries tending to infinity, a stable entire minimal surface $N$ homeomorphic to the plane is constructed. Stability implies using [2],

$$\int_N \left( \frac{1}{2}\, \mathrm{Scal}_M - \kappa + \frac{1}{2}|A|^2 \right) \leq 0$$

where $\kappa = (1/2)\mathrm{Scal}_N$ is the Gauss curvature of $N$. Since by construction $\mathrm{Scal}_M \geq 0, \mathrm{Scal}_M > 0$ outside a compact set, this gives $\int_N \kappa > 0$. Next, one uses the identity, related to the Cohn–Vossen inequality

$$\int_N \kappa = 2\pi - \lim_i \frac{L_i^2}{2A_i}$$

where $A_i, L_i$ are the area and circumference of a sequence of large discs. Estimates using the fact that $M$ is asymptotically Euclidean show that $\lim_i (L_i^2/2A_i) \geq 2\pi$ which gives a contradiction and shows that the minimal surface constructed cannot exist. It follows that $m \geq 0$. It remains to show that the case $m = 0$ is rigid. To do this proves that for an asymptotically Euclidean metric with non-negative scalar curvature, which is positive near infinity, there is a conformally related metric with vanishing scalar curvature and strictly smaller mass. Applying this argument in case $m = 0$ gives a contradiction to the fact that $m \geq 0$. Therefore, $m = 0$ only if the scalar curvature vanishes identically. Suppose now that $(M, g)$ has vanishing scalar curvature but nonvanishing Ricci curvature $\mathrm{Ric}_M$. Then using a deformation of $g$ in the direction of $\mathrm{Ric}_M$, one constructs a metric close to $g$ with negative mass, which leads to a contradiction.

This technique generalizes to Cauchy surfaces of dimension $n \leq 7$. The proof involves induction on dimension. For $n > 7$ minimal hypersurfaces are singular in general and this approach runs into problems. The Witten proof using spinor techniques does not suffer from this limitation but instead requires that $M$ be spin.

## Marginally Trapped Surfaces

Consider a Cauchy data set $(M, g_{ij}, K_{ij})$ as above and let $N$ be a compact surface in $M$ with normal $\eta$, second fundamental form $A$ and mean curvature $H$. Then considering $N$ as a surface in an ambient Lorentzian space $V$ containing $M$, $N$ has two null normal fields which after a rescaling can be taken to be $L_\pm = T \pm \eta$. Here, $T$ is the future-directed time-like unit normal of $M$ in $V$. The null mean curvatures (or null expansions) corresponding to $L_\pm$ can be defined in terms of the variation of the area element $\mu_N$ of $N$ as $\delta_{L_\pm}\mu_N = \theta_\pm \mu_N$ or

$$\theta_\pm = \mathrm{tr}_N K \pm H$$

where $\mathrm{tr}_N K$ denotes the trace of the projection of $K_{ij}$ to $N$. Suppose $L_+$ is the outgoing null normal. $N$ is called outer trapped (marginally trapped, untrapped) if $\theta_+ < 0 (\theta_+ = 0, \theta_+ > 0)$. An asymptotically flat spacetime which contains a trapped surface with $\theta_- < 0, \theta_+ < 0$ is causally incomplete. In the following we will for simplicity drop the word outer from our terminology.

Consider a Cauchy surface $M$. The boundary of the region in $M$ containing trapped surfaces is, if it is sufficiently smooth, a marginally trapped surface. The equation $\theta_+ = 0$ is an equation analogous to the prescribed curvature equation, in particular it is a quasilinear elliptic equation of second order. Marginally trapped surfaces are not variational in the same sense as minimal surfaces. Nevertheless, they are stationary with respect to variations of area within the outgoing light cone. The second variation of area along the outgoing null cone is given, in view of the Raychaudhuri equation, by

$$\delta_{\phi L_+}\theta_+ = -(G_{++} + |\sigma_+|^2)\phi \qquad [7]$$

for a function $\phi$ on $N$. Here $G_{++} = G_{\alpha\beta}L_+^\alpha L_+^\beta$, and $\sigma_+$ denotes the shear of $N$ with respect to $L_+$, that is, the tracefree part of the null second fundamental form with respect to $L_+$. Equation [7] shows that the stability operator in the direction $L_+$ is not elliptic.

In the case of time-symmetric data, $K_{ij} = 0$, the DEC implies $\mathrm{Scal}_M \geq 0$ and marginally trapped surfaces are simply minimal surfaces. A stable compact minimal 2-surface $N$ in a 3-manifold $M$ with non-negative scalar curvature must satisfy

$$2\pi\chi(N) = \int \kappa \geq \frac{1}{2}\int_N \mathrm{Scal}_M + |A|^2 \geq 0$$

and hence by the Gauss–Bonnet theorem, $N$ is diffeomorphic to a sphere or a torus. In case $N$ is a stable minimal torus, the induced geometry is flat and the ambient curvature vanishes at $N$. If, in addition, $N$ minimizes, then $M$ is flat.

For a compact marginally trapped surface $N$ in $M$, analogous results can be proved by studying the stability operator defined with respect to the direction $\eta$. Let $J$ be the operator defined in terms of a variation of $\theta_+$ by $J\phi = \delta_{\phi\eta}\theta_+$. Then

$$J\phi = -\Delta_N\phi + 2s^A D_A\phi + \left(\frac{1}{2}\,\mathrm{Scal}_N - s_A s^A + D_A s^A - \frac{1}{2}|\sigma_+|^2 - G_{+-}\right)\phi$$

Here, $s_A = -(1/2)\langle L_-, D_A L_+\rangle$ and $G_{+-}$ is the Einstein tensor evaluated on $L_+, L_-$. We may call $N$ stable if the real part of the spectrum of $J$ is non-negative. A sufficient condition for $N$ to be stable is that $N$ is locally outermost. This can be formulated, for example, by requiring that a neighborhood of $N$ in $M$ contains no trapped surfaces exterior to $N$. In this case, assuming that the DEC holds, $N$ is a sphere or a torus, and if the real part of the spectrum of $J$ is positive then $N$ is a sphere. If $N$ is a torus, then the ambient curvature and shear vanishes at $N$, $s_A$ is a gradient, and $N$ is flat. One expects that in addition, global rigidity should hold, in analogy with the minimal surface case. This is an open problem. If $N$ satisfies the stronger condition of strict stability, which corresponds to the spectrum of $J$ having positive real part, then $N$ is in the interior of a hypersurface $H$ of the ambient spacetime, with the property that it is foliated by marginally trapped surfaces (Andersson *et al.* 2005). If the NEC holds and $N$ has nonvanishing shear, then $H$ is spacelike at $N$. A hypersurface $H$ with these properties is known as a dynamical horizon.

## Jang's Equation

Consider a Cauchy data set $(M, g_{ij}, K_{ij})$. Extend $K_{ij}$ to a tensor field on $M \times \mathbb{R}$, constant in the vertical direction. Then the equation for a graph

$$N = \{(x, t) \in M \times \mathbb{R}, \quad t = f(x)\}$$

such that $N$ has mean curvature equal to the trace of the projection of $K_{ij}$ to $N$ with respect to the induced metric on $N$, is given by

$$\sum_{i,j}\left(K^{ij} - \frac{\nabla^i\nabla^j f}{(1+|\nabla f|^2)^{1/2}}\right)\left(g_{ij} - \frac{\nabla_i f\nabla_j f}{1+|\nabla f|^2}\right) = 0 \qquad [8]$$

an equation closely related to the equation $\theta_+ = 0$. Equation [8] was introduced by P S Jang (Jang 1978) as part of an attempt to generalize the inverse mean curvature flow method of Geroch from time-symmetric to general Cauchy data.

Existence and regularity for Jang's equation were proved by Schoen and Yau (1981) and used to generalize their proof of the positive-mass theorem from the case of maximal slices to the general case. The solution to Jang's equation is constructed as the limit of the solution to a sequence of regularized problems. The limit consists of a collection $N$ of submanifolds of $M \times \mathbb{R}$. In particular, component near infinity is a graph and has the same mass as $M$. $N$ may contain vertical components which project onto marginally trapped surfaces in $M$, and in fact these constitute the only possibilities for blow-up of the sequence of graphs used to construct $N$. If the DEC is valid, the metric on N has non-negative scalar curvature in the weak sense that

$$\int_N \mathrm{Scal}_N\phi^2 + 2|\nabla\phi|^2 > 0$$

for smooth compactly supported functions $\phi$. If the DEC holds strictly, the strict inequality holds and in this case the metric on $N$ is conformal to a metric with vanishing scalar curvature.

Jang's equation can be applied to prove existence of marginally trapped surfaces, given barriers. Let $(M, g_{ij}, K_{ij})$ be a Cauchy data set containing two compact surfaces $N_1, N_2$ which together bound a compact region $M'$ in $M$. Suppose the surfaces $N_1$ and $N_2$ have $\theta_+ < 0$ on $N_1$ and $\theta_+ > 0$ on $N_2$. Schoen recently proved the following result.

**Theorem 3** (Existence of marginally trapped surfaces). *Let $M', N_1, N_2$ be as above. Then there is a finite collection of compact, marginally trapped surfaces $\{\Sigma_a\}$ contained in the interior of $M'$, such that $\cup\Sigma_a$ is homologous to $N_1$.* If the DEC holds, then $\Sigma_a$ is a collection of spheres and tori.

The proof proceeds by solving a sequence of Dirichlet boundary-value problems for Jang's equation with boundary value on $N_1, N_2$ tending to $-\infty$ and $\infty$, respectively. The assumption on $\theta_+$ is used to show the existence of barriers for Jang's equation. Let $f_k$ be the sequence of solutions to the Dirichlet problems. Jang's equation is invariant under renormalization $f_k \to f_k + c_k$ for some sequence $c_k$ of real numbers. A Harnack inequality for the gradient of the solutions to Jang's equation is used to show that the sequence of solutions $f_k$, possibly after a renormalization, has a subsequence converging to a vertical submanifold of $M' \times \mathbb{R}$, which projects to a collection $\Sigma_a$ of marginally trapped surfaces. By construction, the zero sets of the $f_k$ are homologous to $N_1$ and $N_2$. The estimates on the sequence $\{f_k\}$ show that this holds also in the limit $k \to \infty$. The statement about the topology of the $\Sigma_a$ follows by showing, using the above-mentioned inequality for $\mathrm{Scal}_N$, that if DEC holds, the total Gauss curvature of each surface $\Sigma_a$ is non-negative.

## Center of mass

Since by the positive-mass theorem $m > 0$ unless the ambient spacetime is flat, it makes sense to consider the problem of finding an appropriate notion of center of mass. This problem was solved by Huisken and Yau who showed that under the asymptotic conditions [6] the isoperimetric problem has a unique solution if one considers sufficiently large spheres.

**Theorem 4** (Huisken and Yau 1996). *There is an $H_0 > 0$ and a compact region $B_{H_0}$ such that for each $H \in (0, H_0)$ there is a unique constant mean curvature sphere $S_H$ with mean curvature $H$ contained in $M \backslash B_{H_0}$. The spheres form a foliation.*

The proof involves a study of the evolution equation

$$\frac{\mathrm{d}x}{\mathrm{d}s} = (H - \bar{H})\eta \qquad [9]$$

where $\bar{H}$ is the average mean curvature. This is the gradient flow for the isoperimetric problem of minimizing area keeping the enclosed volume constant. The solutions in Euclidean space are standard spheres. Equation [9] defines a parabolic system, in particular we have

$$\frac{\mathrm{d}}{\mathrm{d}s} H = \Delta H + (\mathrm{Ric}(\eta, \eta) + |A|^2)(H - \bar{H})$$

It follows from the fall-off conditions [6] that the foliation of spheres constructed in Theorem 4 are untrapped surfaces. They can therefore be used as outer barriers in the existence result for marginally trapped surfaces, (Theorem 3).

The mean curvature flow for a spatial hypersurface in a Lorentz manifold is also parabolic. This flow has been applied to construct constant mean curvature Cauchy hypersurfaces in spacetimes.

## Maximal and Related Surfaces

Let $N$ be the hypersurface $x_0 = u(x_1, \ldots, x_n)$ in Minkowski space $\mathbb{R}^{1+n}$ with line element $-\mathrm{d}x_0^2 + \mathrm{d}x_1^2 + \cdots + \mathrm{d}x_n^2$. Assume $|\nabla u| < 1$ so that $N$ is spacelike. Then $N$ is stationary with respect to variations of area if $u$ solves the equation

$$\sum_i \nabla_i \left( \frac{\nabla_i u}{\sqrt{1 - |\nabla u|^2}} \right) = 0 \qquad [10]$$

$N$ maximizes area with respect to compactly supported variations, and hence is called a maximal surface. As in the case of the minimal surface equation, eqn [10] and more generally the Lorentzian prescribed mean curvature equation, is quasilinear elliptic, but it is not uniformly elliptic, which makes the regularity theory more subtle.

A Bernstein principle analogous to the one for the minimal surface equation holds for the maximal surface equation [10]. Suppose that $u$ is a solution to [10] which is defined on all of $\mathbb{R}^n$. Then $u$ is an affine function (Cheng and Yau 1976). An important tool used in the proof is a Bochner type identity, originally due to Calabi, for the norm of the second fundamental form. For a hypersurface in a flat ambient space, the Codazzi equation states $\nabla_i A_{jk} - \nabla_j A_{ik} = 0$. This gives the identity

$$\Delta A_{ij} = \nabla_i \nabla_j H + A_{km} R^{m\ k}_{\ i\ j} + A_{mi} \mathrm{Ric}^m_j \qquad [11]$$

The curvature terms can be rewritten in terms of $A_{ij}$ if the ambient space is flat. Using [11] to compute $\Delta |A|^2$ gives an expression which is quadratic in $\nabla A$, and fourth order in $|A|$, and which allows one to perform maximum principle estimates on $|A|$. Generalizations of this technique for hypersurfaces in general ambient spaces play an important role in the proof of regularity of minimal surfaces, and in the proof of existence for Jang's equation as well as in the analysis of the mean curvature flow used to prove existence of round spheres. The generalization of eqn [11] is known as a Simons identity.

For the case of maximal hypersurfaces of Minkowski space, it follows from further maximum principle estimates that a maximal hypersurface of Minkowski space is convex, in particular, it has nonpositive Ricci curvature. Generalizations of this technique allow one to analyze entire constant mean curvature hypersurfaces of Minkowski space.

Consider a globally hyperbolic Lorentzian manifold $(V, \gamma)$. A $C^0$ hypersurface is said to be weakly spacelike if timelike curves intersect it in at most one point. Call a codimension-2 submanifold $\Gamma \subset V$ a weakly spacelike boundary if it bounds a weakly spacelike hypersurface $N_0$.

**Theorem 5** (Existence for Plateau problem for maximal surfaces (Bartnik 1988)). *Let $V$ be a globally hyperbolic spacetime and assume that the causal structure of $V$ is such that the domain of dependence of any compact domain in $V$ is compact. Given a weakly spacelike boundary $\Gamma$ in $V$, there is a weakly spacelike maximal hypersurface $N$ with $\Gamma$ as its boundary. $N$ is smooth except possibly on null geodesics connecting points of $\Gamma$.*

Here, maximal hypersurface is understood in a weak sense, referring to stationarity with respect to variations. Due to the nonuniform ellipticity for the maximal surface equation, the interior regularity

which holds for minimal surfaces fails to hold in general for the maximal surface equation.

A time-oriented spacetime is said to have a crushing singularity to the past (future) if there is a sequence $\Sigma_n$ of Cauchy surfaces so that the mean curvature function $H_n$ of $\Sigma_n$ diverges uniformly to $-\infty(\infty)$.

**Theorem 6** (Gerhardt 1983). *Suppose that* $(V, \gamma)$ *is globally hyperbolic with compact Cauchy surfaces and satisfies the SEC. Then if* $(V, \gamma)$ *has crushing singularities to the past and future it is globally foliated by constant mean curvature hypersurfaces. The mean curvature* $\tau$ *of these Cauchy surfaces is a global time function.*

The proof involves an application of results from geometric measure theory to an action $\mathcal{E}$ of the form discussed earlier. A barrier argument is used to control the maximizers. Bartnik (1984, theorem 4.1) gave a direct proof of existence of a constant mean curvature (CMC) hypersurface, given barriers. If the spacetime $(V, \gamma)$ is symmetric, so that a compact Lie group acts on $V$ by isometries, then CMC hypersurfaces in $V$ inherit the symmetry. Theorem 6 gives a condition under which a spacetime is globally foliated by CMC hypersurfaces. In general, if the SEC holds in a spatially compact spacetime, then for each $\tau \neq 0$, there is at most one constant mean curvature Cauchy surface with mean curvature $\tau$. In case $V$ is vacuum, $\mathrm{Ric}_V = 0$, and $3+1$ dimensional, then each point $x \in V$ is on at most one hypersurface of constant mean curvature unless $V$ is flat and splits as a metric product.

There are vacuum spacetimes with compact Cauchy surface which contain no CMC hypersurface (Chrusciel *et al.* 2004). The proof is carried out by constructing Cauchy data, using a gluing argument, on the connected sum of two tori, such that the resulting Cauchy data set $(M, g_{ij}, K_{ij})$ has an involution which reverses the sign of $K_{ij}$. The involution extends to the maximal vacuum development $V$ of the Cauchy data set. Existence of a CMC surface in $V$ gives, in view of the involution, barriers which allow one to construct a maximal Cauchy surface homeomorphic to $M$. This leads to a contradiction, since the connected sum of two tori does not carry a metric of positive scalar curvature, and therefore, in view of the constraint equations, cannot be imbedded as a maximal Cauchy surface in a vacuum spacetime. The maximal vacuum development $V$ is causally geodesically incomplete. However, in view of the existence proof for CMC Cauchy surfaces (cf. Theorem 6), these spacetimes cannot have a crushing singularity. It would be interesting to settle the open question whether there are stable examples of this type.

In the case of a spacetime $V$ which has an expanding end, one does not expect in general that the spacetime is globally foliated by CMC hypersurfaces even if $V$ is vacuum and contains a CMC Cauchy surface. This expectation is based on the phenomenon known as the collapse of the lapse; for example, the Schwarzschild spacetime does not contain a global foliation by maximal Cauchy surfaces (Beig and Murchadha 1998). However, no counterexample is known in the spatially compact case. In spite of these caveats, many examples of spacetimes with global CMC foliations are known, and the CMC condition, or more generally prescribed mean curvature, is an important gauge condition for general relativity.

Some examples of situations where global constant or prescribed mean curvature foliations are known to exist in vacuum or with some types of matter are spatially homogeneous spacetimes, and spacetimes with two commuting Killing fields. Small data global existence for the Einstein equations with CMC time gauge have been proved for spacetimes with one Killing field, with Cauchy surface a circle bundle over a surface of genus $>1$, by Choquet-Bruhat and Moncrief. Further, for $(3+1)$-dimensional spacetimes with Cauchy surface admitting a hyperbolic metric, small data global existence in the expanding direction has been proved by Andersson and Moncrief. See Andersson (2004) and Rendall (2002) for surveys on the Cauchy problem in general relativity.

## Null Hypersurfaces

Consider an asymptotically flat spacetime containing a black hole, that is, a region $B$ such that future causal curves starting in $B$ cannot reach observers at infinity. The boundary of the trapped region is called the event horizon $\mathcal{H}$. This is a null hypersurface, which under reasonable conditions on causality has null generators which are complete to the future. Due to the completeness, assuming that $\mathcal{H}$ is smooth, one can use the Raychaudhuri equation [7] to show that the null expansion $\theta_+$ of a spatial cross section of $\mathcal{H}$ must satisfy $\theta_+ \geq 0$, and hence that the area of cross sections of $\mathcal{H}$ grows monotonously to the future. A related statement is that null generators can enter $\mathcal{H}$ but may not leave it. This was first proved by Hawking for the case of smooth horizons, using essentially the Raychaudhuri equation. In general $\mathcal{H}$ can fail to be smooth. However, from the definition of $\mathcal{H}$ as the boundary of the trapped region it follows that it has support hypersurfaces, which are past light cones. This property allows one to prove that $\mathcal{H}$ is Lipschitz and hence smooth almost everywhere. At smooth points of $\mathcal{H}$, the calculations in the proof of

Hawking apply, and the monotonicity of the area of cross sections follows.

**Theorem 7** (Area theorem (Chrusciel *et al.* 2001)). *Let $\mathcal{H}$ be a black hole event horizon in a smooth spacetime $(M, g)$. Suppose that the generators are future complete and the NEC holds on $\mathcal{H}$. Let $S_a, a = 1, 2$, be two spacelike cross sections of $\mathcal{H}$ and suppose that $S_2$ is to the future of $S_1$. Then $\mathcal{A}(S_2) \geq \mathcal{A}(S_1)$.*

The eikonal equation $\nabla^\alpha u \nabla_\alpha u = 0$ plays a central role in geometric optics. Level sets of a solution $u$ are null hypersurfaces which correspond to wave fronts. Much of the recent progress on rough solutions to the Cauchy problem for quasilinear wave equations is based on understanding the influence of the geometry of these wave fronts on the evolution of high-frequency modes ‘in the background spacetime. In this analysis many objects familiar from general relativity, such as the structure equations for null hypersurfaces, the Raychaudhuri equation, and the Bianchi identities play an important role, together with novel techniques of geometric analysis used to control the geometry of cross sections of the wave fronts and to estimate the connection coefficients in a rough spacetime geometry. These techniques show great promise and can be expected to have a significant impact on our understanding of the Einstein equations and general relativity.

## Acknowledgments

The author is grateful to, among others, Greg Galloway, Gerhard Huisken, Jim Isenberg, Piotr Chrusciel, and Dan Pollack for helpful discussions concerning the topics covered in this article. The work of the author has been supported in part by the NSF under contract number DMS 0407732 with the University of Miami.

*See also:* Computational Methods in General Relativity: The Theory; Einstein Equations: Initial Value Formulation; Einstein’s Equations with Matter; Geometric Flows and the Penrose Inequality; Hamiltonian Reduction of Einstein’s Equations; Holomorphic Dynamics; Lorentzian Geometry; Minimal Submanifolds; Mirror Symmetry: A Geometric Survey; Spacetime Topology, Causal Structure and Singularities; Stability of Minkowski Space.

## Further Reading


Andersson L (2004) The global existence problem in general relativity. In: Chrusciel P and Friedrich H (eds.) *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, pp. 71–120. Basel: Birkhauser.

Andersson L, Galloway GJ, and Howard R (1998) A strong maximum principle for weak solutions of quasi-linear elliptic equations with applications to Lorentzian and Riemannian geometry. *Communications on Pure and Applied Mathematics* 51(6): 581–624.

Andersson L, Mars M, and Simon W (2005) Local existence of dynamical and trapping horizons. *Physical Review Letters* 95: 111102.

Bartnik R (1984) Existence of maximal surfaces in asymptotically flat spacetimes. *Communications in Mathematical Physics* 94(2): 155–175.

Bartnik R (1988) Regularity of variational maximal surfaces. *Acta Mathematica* 161(3–4): 145–181.

Beig R and Murchadha NÓ (1998) Late time behavior of the maximal slicing of the Schwarzschild black hole. *Physical Review D* 57(8): 4728–4737.

Cai M and Galloway GJ (2000) Rigidity of area minimizing tori in 3-manifolds of nonnegative scalar curvature. *Communications in Analysis and Geometry* 8(3): 565–573.

Shiu Yuch Cheng and Shing Tung Yau (1976) Maximal space-like hypersurfaces in the Lorentz–Minkowski spaces. *Annals of Mathematics* (2) 104(3): 407–419.

Chrusciel PT, Delay E, Galloway GJ, and Howard R (2001) Regularity of horizons and the area theorem. *Ann. Henri Poincaré* 2(1): 109–178.

Chrusciel PT, Isenberg J, and Pollack D (2004) Gluing initial data sets for general relativity. *Physical Review Letters* 93: 081101.

Federer H (1969) Geometric measure theory. *Die Grundlehren der mathematischen Wissenschaften, Band 153*. New York: Springer.

Galloway GJ (2000) Maximum principles for null hypersurfaces and null splitting theorems. *Annales Henri Poincaré* 1(3): 543–567.

Gerhardt C (1983) H-surfaces in Lorentzian manifolds. *Communications in Mathematical Physics* 89(4): 523–553.

Gilbarg D and Trudinger NS (1983) Elliptic partial differential equations of second order. In: *Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences)*, 2nd edn., vol. 224. Berlin: Springer.

Huisken G and Yau ST (1996) Definition of center of mass for isolated physical systems and unique foliations by stable spheres with constant mean curvature. *Inventiones Mathematicae* 124(1–3): 281–311.

Jang PS (1978) On the positivity of energy in general relativity. *Journal of Mathematical Physics* 19(5): 1152–1155.

Lawson HB Jr. (1980) *Lectures on Minimal Submanifolds*, vol. 1. Mathematics Lecture Series 9 (2nd edn). Wilmington, DE: Publish or Perish Inc.

Rendall AD (2002) Theorems on existence and global dynamics for the Einstein equations. *Living Reviews in Relativity* 5: 62 pp. (electronic).

Schoen R and Yau ST (1981) Proof of the positive mass theorem II. *Communications in Mathematical Physics* 79(2): 231–260.

Simon L (1997) The minimal surface equation. In: *Geometry, V, Encyclopaedia of Mathematical Science* vol. 90, pp. 239–272. Berlin: Springer.

# Geometric Flows and the Penrose Inequality

**H Bray**, Duke University, Durham, NC, USA



This article was originally published as "Black holes, geometric flows, and the Penrose inequality in general relativity." Notices of the *American Mathematical Society*, 49(2002), 1372–1381.

## Introduction

In a paper, R Penrose (1973) made a physical argument that the total mass of a spacetime which contains black holes with event horizons of total area $A$ should be at least $\sqrt{A/16\pi}$. An important special case of this physical statement translates into a very beautiful mathematical inequality in Riemannian geometry known as the Riemannian Penrose inequality. The Riemannian Penrose inequality was first proved by Huisken and Ilmanen (1997) for a single black hole and then by the author in 1999 for any number of black holes. The two approaches use two different geometric flow techniques. The most general version of the Penrose inequality is still open.

A natural interpretation of the Penrose inequality is that the mass contributed by a collection of black holes is (at least) $\sqrt{A/16\pi}$. More generally, the question "How much matter is in a given region of a spacetime?" is still very much an open problem. (Christodoulou and Yau 1988). In this paper, we will discuss some of the qualitative aspects of mass in general relativity, look at examples which are informative, and describe the two very geometric proofs of the Riemannian Penrose inequality.

### Total Mass in General Relativity

Two notions of mass which are well understood in general relativity are local energy density at a point and the total mass of an asymptotically flat spacetime. However, defining the mass of a region larger than a point but smaller than the entire universe is not very well understood at all.

Suppose $(M^3, g)$ is a Riemannian 3-manifold isometrically embedded in a $(3+1)$-dimensional Lorentzian spacetime $N^4$. Suppose that $M^3$ has zero-second fundamental form in the spacetime. This is a simplifying assumption which allows us to think of $(M^3, g)$ as a "$t=0$" slice of the spacetime. (Recall that the second fundamental form is a measure of how much $M^3$ curves inside $N^4$. $M^3$ is also sometimes called "totally geodesic" since geodesics of $N^4$ which are tangent to $M^3$ at a point stay inside $M^3$ forever.) The Penrose inequality (which allows for $M^3$ to have general second fundamental form) is known as the Riemannian Penrose inequality when the second fundamental form is set to zero.

We also want to only consider $(M^3, g)$ that are asymptotically flat at infinity, which means that for some compact set $K$, the "end" $M^3 \backslash K$ is diffeomorphic to $\boldsymbol{R}^3 \backslash B_1(0)$, where the metric $g$ is asymptotically approaching (with certain decay conditions) the standard flat metric $\delta_{ij}$ on $\boldsymbol{R}^3$ at infinity. The simplest example of an asymptotically flat manifold is $(\boldsymbol{R}^3, \delta_{ij})$ itself. Other good examples are the conformal metrics $(\boldsymbol{R}^3, u(x)^4\delta_{ij})$, where $u(x)$ approaches a constant sufficiently rapidly at infinity. (Also, sometimes it is convenient to allow $(M^3, g)$ to have multiple asymptotically flat ends, in which case each connected component of $M^3 \backslash K$ must have the property described above.) A qualitative picture of an asympotically flat 3-manifold is shown in Figure 1.

The purpose of these assumptions on the asymptotic behavior of $(M^3, g)$ at infinity is that they imply the existence of the limit

$$m = \frac{1}{16\pi} \lim_{\sigma\to\infty} \int_{S_\sigma} \sum_{i,j} (g_{ij,i}\nu_j - g_{ii,j}\nu_j)\, \mathrm{d}\mu$$

where $S_\sigma$ is the coordinate sphere of radius $\sigma, \nu$ is the unit normal to $S_\sigma$, and $\mathrm{d}\mu$ is the area element of $S_\sigma$ in the coordinate chart. The quantity $m$ is called the "total mass" (or ADM mass) of $(M^3, g)$ and does not depend on the choice of asymptotically flat coordinate chart.

The above equation is where many people would stop reading an article like this. But before you do, we will promise not to use this definition of the total mass in this paper. In fact, it turns out that total mass can be quite well understood with an example. Going back to the example $(\boldsymbol{R}^3, u(x)^4\delta_{ij})$, if we suppose that $u(x) > 0$ has the asymptotics at infinity

$$u(x) = a + b/|x| + \mathcal{O}(1/|x|^2) \qquad [1]$$

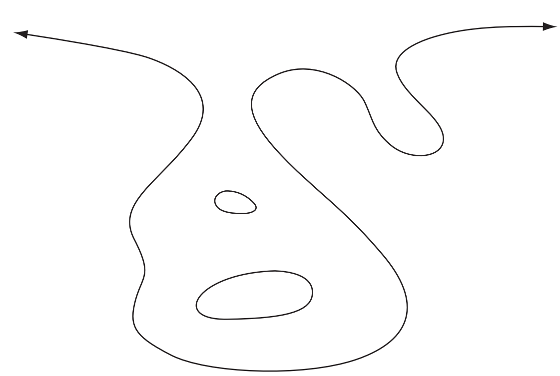


**Figure 1** A qualitative picture of an asymptotically flat 3-manifold.

(and derivatives of the $\mathcal{O}(1/|x|^2)$ term are $\mathcal{O}(1/|x|^3)$), then the total mass of $(M^3, g)$ is

$$m = 2ab \qquad [2]$$

Furthermore, suppose $(M^3, g)$ is any metric whose "end" is isometric to $(\boldsymbol{R}^3 \backslash K, u(x)^4 \delta_{ij})$, where $u(x)$ is harmonic in the coordinate chart of the end $(\boldsymbol{R}^3 \backslash K, \delta_{ij})$ and goes to a constant at infinity. Then expanding $u(x)$ in terms of spherical harmonics demonstrates that $u(x)$ satisfies condition [1]. We will call these Riemannian manifolds $(M^3, g)$ "harmonically flat at infinity," and we note that the total mass of these manifolds is also given by eqn [2].

A very nice lemma by Schoen and Yau is that, given any $\epsilon > 0$, it is always possible to perturb an asymptotically flat manifold to become harmonically flat at infinity such that the total mass changes less than $\epsilon$ and the metric changes less than $\epsilon$ pointwise, all while maintaining non-negative scalar curvature. Hence, it happens that to prove the theorems in this paper, we only need to consider harmonically flat manifolds! Thus, we can use eqn [2] as our definition of total mass. As an example, note that $(\boldsymbol{R}^3, \delta_{ij})$ has zero total mass. Also, note that, qualitatively, the total mass of an asymptotically flat or harmonically flat manifold is the $1/r$ rate at which the metric becomes flat at infinity.

### The Phenomenon of Gravitational Attraction

What do the above definitions of total mass have to do with anything physical? That is, if the total mass is the $1/r$ rate at which the metric becomes flat at infinity, what does this have to do with our real-world intuitive idea of mass?

The answer to this question is very nice. Given a Schwarzschild spacetime metric

$$\left(\boldsymbol{R}^4, \left(1 + \frac{m}{2|x|}\right)^4 (\mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}x_3^2) - \left(\frac{1 - m/2|x|}{1 + m/2|x|}\right)^2 \mathrm{d}t^2\right)$$

$|x| > m/2$, for example, note that the $t = 0$ slice (which has zero-second fundamental form) is the spacelike Schwarzschild metric

$$\left(\boldsymbol{R}^3 \setminus B_{m/2}(0), \left(1 + \frac{m}{2|x|}\right)^4 \delta_{ij}\right)$$

(discussed more later). Note that according to eqn [2], the parameter $m$ is in fact the total mass of this 3-manifold.

On the other hand, suppose we were to release a small test particle, initially at rest, a large distance $r$ from the center of the Schwarzschild spacetime. If this particle is not acted upon by external forces, then it should follow a geodesic in the spacetime. It turns out that with respect to the asymptotically flat coordinate chart, these geodesics "accelerate" towards the middle of the Schwarzschild metric proportional to $m/r^2$ (in the limit as $r$ goes to infinity). Thus, our Newtonian notion of mass also suggests that the total mass of the spacetime is $m$.

### Local Energy Density

Another quantification of mass which is well understood is local energy density. In fact, in this setting, the local energy density at each point is

$$\mu = \frac{1}{16\pi} R$$

where $R$ is the scalar curvature of the 3-manifold (which has zero-second fundamental form in the spacetime) at each point. Note that $(\boldsymbol{R}^3, \delta_{ij})$ has zero energy density at each point as well as zero total mass. This is appropriate since $(\boldsymbol{R}^3, \delta_{ij})$ is in fact a "$t = 0$" slice of Minkowski spacetime, which represents a vacuum. Classically, physicists consider $\mu \geq 0$ to be a physical assumption. Hence, from this point on, we will not only assume that $(M^3, g)$ is asymptotically flat, but also that it has non-negative scalar curvature,

$$R \geq 0$$

This notion of energy density also helps us understand total mass better. After all, we can take any asymptotically flat manifold and then change the metric to be perfectly flat outside a large compact set, thereby giving the new metric zero total mass. However, if we introduce the physical condition that both metrics have non-negative scalar curvature, then it is a beautiful theorem that this is in fact not possible, unless the original metric was already $(\boldsymbol{R}^3, \delta_{ij})$! (This theorem is actually a corollary to the positive mass theorem discussed below.) Thus, the curvature obstruction of having non-negative scalar curvature at each point is a very interesting condition.

Also, notice the indirect connection between the total mass and local energy density. At this point, there does not seem to be much of a connection at all. The total mass is the $1/r$ rate at which the metric becomes flat at infinity, and local energy density is the scalar curvature at each point. Furthermore, if a metric is changed in a compact set, local energy density is changed, but the total mass is unaffected.

The reason for this is that the total mass is "not" the integral of the local energy density over the manifold. In fact, this integral fails to take potential energy into account (which would be expected to

contribute a negative energy) as well as gravitational energy. Hence, it is not initially clear what we should expect the relationship between total mass and local energy density to be, so let us begin with an example.

### Example Using Superharmonic Functions in $\boldsymbol{R}^3$

Once again, let us return to the $(\boldsymbol{R}^3, u(x)^4\delta_{ij})$ example. The formula for the scalar curvature is

$$R = -8u(x)^{-5}\Delta u(x)$$

Hence, since the physical assumption of non-negative energy density implies non-negative scalar curvature, we see that $u(x) > 0$ must be superharmonic ($\Delta u \leq 0$). For simplicity, let us also assume that $u(x)$ is harmonic outside a bounded set so that we can expand $u(x)$ at infinity using spherical harmonics. Hence, $u(x)$ has the asymptotics of eqn [1]. By the maximum principle, it follows that the minimum value for $u(x)$ must be $a$, referring to eqn [1]. Hence, $b \geq 0$, which implies that $m \geq 0$! Thus, we see that the assumption of non-negative energy density at each point of $(\boldsymbol{R}^3, u(x)^4\delta_{ij})$ implies that the total mass is also non-negative, which is what one would hope.

### The Positive Mass Theorem

Why would one hope this? What would be the difference if the total mass were negative? This would mean that a gravitational system of positive energy density could collectively act as a net negative total mass. This phenomenon has not been observed experimentally, and so it is not a property that we would hope to find in general relativity.

More generally, suppose we have any asymptotically flat manifold with non-negative scalar curvature, is it true that the total mass is also non-negative? The answer is yes, and this fact is know as the positive mass theorem, first proved by Schoen and Yau (1979) using minimal surface techniques and then by Witten (1981) using spinors. In the zero-second fundamental form case, the positive mass theorem is known as the Riemannian positive mass theorem and is stated below.

**Theorem 1** (Schoen, Yau). *Let $(M^3, g)$ be any asymptotically flat, complete Riemannian manifold with non-negative scalar curvature. Then the total mass $m \geq 0$, with equality if and only if $(M^3, g)$ is isometric to $(\boldsymbol{R}^3, \delta)$.*

### Gravitational Energy

The previous example neglects to illustrate some of the subtleties of the positive mass theorem. For example, it is easy to construct asymptotically flat manifolds $(M^3, g)$ (not conformal to $\boldsymbol{R}^3$) which have zero scalar curvature everywhere and yet have "nonzero" total mass. By the positive mass theorem, the mass of these manifolds is positive. Physically, this corresponds to a spacetime with zero energy density everywhere which still has positive total mass. From where did this mass come? How can a vacuum have positive total mass?

Physicists refer to this extra energy as gravitational energy. There is no known local definition of the energy density of a gravitational field, and presumably such a definition does not exist. The curious phenomenon, then, is that for some reason, gravitational energy always makes a non-negative contribution to the total mass of the system.

### Black Holes

Another very interesting and natural phenomenon in general relativity is the existence of black holes. Instead of thinking of black holes as singularities in a spacetime, we will think of black holes in terms of their horizons. For example, suppose we are exploring the universe in a spacecraft capable of traveling at any speed less than the speed of light. If we are investigating a black hole, we would want to make sure that we don't get too close and get trapped by the "gravitational forces" of the black hole. In fact, we could imagine a "sphere of no return" beyond which it is impossible to escape from the black hole. This is called the event horizon of a black hole.

However, one limitation of the notion of an event horizon is that it is very hard to determine its location. One way is to let daredevil spacecraft see how close they can get to the black hole and still escape from it eventually. The only problem with this approach (besides the cost in spacecraft) is that it is hard to know when to stop waiting for a daredevil spacecraft to return. Even if it has been 50 years, it could be that this particular daredevil was not trapped by the black hole but got so close that it will take it 1000 or more years to return. Thus, to define the location of an event horizon even mathematically, we need to know the entire evolution of the spacetime. Hence, event horizons can not be computed based only on the local geometry of the spacetime.

This problem is solved (at least for the mathematician) with the notion of apparent horizons of black holes. Given a surface in a spacetime, suppose that it emits an outward shell of light. If the surface area of this shell of light is decreasing everywhere on the surface, then this is called a trapped surface. The outermost boundary of these trapped surfaces is called the apparent horizon of the black hole. Apparent horizons can be computed based on their

local geometry, and an apparent horizon always implies the existence of an event horizon outside of it (Hawking and Ellis 1973).

Now let us return to the case we are considering in this paper where $(M^3, g)$ is a "$t = 0$" slice of a spacetime with zero-second fundamental form. Then it is a very nice geometric fact that apparent horizons of black holes intersected with $M^3$ correspond to the connected components of the outermost minimal surface $\Sigma_0$ of $(M^3, g)$.

All of the surfaces we are considering in this paper will be required to be smooth boundaries of open bounded regions, so that outermost is well defined with respect to a chosen end of the manifold. A minimal surface in $(M^3, g)$ is a surface which is a critical point of the area function with respect to any smooth variation of the surface. The first variational calculation implies that minimal surfaces have zero mean curvature. The surface $\Sigma_0$ of $(M^3, g)$ is defined as the boundary of the union of the open regions bounded by all of the minimal surfaces in $(M^3, g)$. It turns out that $\Sigma_0$ also has to be a minimal surface, so we call $\Sigma_0$ the "outermost minimal surface." A qualitative sketch of an outermost minimal surface of a 3-manifold is shown in **Figure 2**.

We will also define a surface to be "(strictly) outer minimizing" if every surface which encloses it has (strictly) greater area. Note that outermost minimal surfaces are strictly outer minimizing. Also, we define a "horizon" in our context to be any minimal surface which is the boundary of a bounded open region.

It also follows from a stability argument (using the Gauss–Bonnet theorem interestingly) that each component of an outermost minimal surface (in a 3-manifold with non-negative scalar curvature) must have the topology of a sphere. Furthermore, there is a physical argument, based on Penrose (1973), which suggests that the mass contributed by the black holes (thought of as the connected components of $\Sigma_0$) should be defined to be $\sqrt{A_0/16\pi}$,

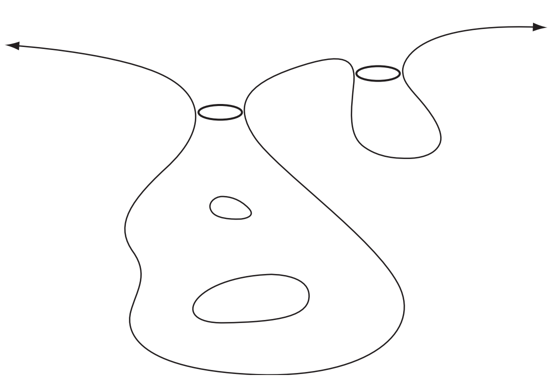


**Figure 2** A qualitative sketch of an outermost minimal surface of a 3-manifold.

where $A_0$ is the area of $\Sigma_0$. Hence, the physical argument that the total mass should be greater than or equal to the mass contributed by the black holes yields the following geometric statement.

**The Riemannian Penrose Inequality** Let $(M^3, g)$ be a complete, smooth, 3-manifold with non-negative scalar curvature which is harmonically flat at infinity with total mass $m$ and which has an outermost minimal surface $\Sigma_0$ of area $A_0$. Then,

$$m \geq \sqrt{\frac{A_0}{16\pi}} \qquad [3]$$

with equality if and only if $(M^3, g)$ is isometric to the Schwarzschild metric

$$\left(\mathbf{R}^3 \backslash \{0\}, \left(1 + \frac{m}{2|x|}\right)^4 \delta_{ij}\right)$$

outside their respective outermost minimal surfaces.

The above statement has been proved by the present author, and Huisken and Ilmanen proved it when $A_0$ is defined instead to be the area of the largest connected component of $\Sigma_0$. We will discuss both approaches in this paper, which are very different, although they both involve flowing surfaces and/or metrics.

We also clarify that the above statement is with respect to a chosen end of $(M^3, g)$, since both the total mass and the definition of outermost refer to a particular end. In fact, nothing very important is gained by considering manifolds with more than one end, since extra ends can always be compactified by connect summing them (around a neighborhood of infinity) with large spheres while still preserving non-negative scalar curvature, for example. Hence, we will typically consider manifolds with just one end. In the case that the manifold has multiple ends, we will require every surface (which could have multiple connected components) in this paper to enclose all of the ends of the manifold except the chosen end.

## The Schwarzschild Metric

The Schwarzschild metric

$$\left(\mathbf{R}^3 \setminus \{0\}, \left(1 + \frac{m}{2|x|}\right)^4 \delta_{ij}\right)$$

referred to in the above statement of the Riemannian Penrose inequality, is a particularly important example to consider, and corresponds to a zero-second fundamental form, spacelike slice of the usual $(3+1)$-dimensional Schwarzschild metric (which represents a spherically symmetric static black hole in vacuum). The three-dimensional

Schwarzschild metrics have total mass $m > 0$ and are characterized by being the only spherically symmetric, geodesically complete, zero scalar curvature 3-metrics, other than $(\mathbf{R}^3, \delta_{ij})$. They can also be embedded in four-dimensional Euclidean space $(x, y, z, w)$ as the set of points satisfying

$$|(x, y, z)| = (w^2/8m) + 2m$$

which is a parabola rotated around an $S^2$. This last picture allows us to see that the Schwarzschild metric, which has two ends, has a $Z_2$ symmetry which fixes the sphere with $w = 0$ and $|(x, y, z)| = 2m$, which is clearly minimal. Furthermore, the area of this sphere is $4\pi(2m)^2$, giving equality in the Riemannian Penrose inequality.

## A Brief History of the Problem

The Riemannian Penrose inequality has a rich history spanning nearly three decades and has motivated much interesting mathematics and physics. In 1973, R Penrose in effect conjectured an even more general version of inequality [3] using a very clever physical argument, which we will not have room to repeat here (Penrose 1973). His observation was that a counterexample to inequality [3] would yield Cauchy data for solving the Einstein equations, the solution to which would likely violate the cosmic censor conjecture (which says that singularities generically do not form in a spacetime unless they are inside a black hole).

Jang and Wald (1977), extending ideas of Geroch, gave a heuristic proof of inequality [3] by defining a flow of 2-surfaces in $(M^3, g)$ in which the surfaces flow in the outward normal direction at a rate equal to the inverse of their mean curvatures at each point. The Hawking mass of a surface (which is supposed to estimate the total amount of energy inside the surface) is defined to be

$$m_{\text{Hawking}}(\Sigma) = \sqrt{\frac{|\Sigma|}{16\pi}}\left(1 - \frac{1}{16\pi}\int_\Sigma H^2\right)$$

(where $|\Sigma|$ is the area of $\Sigma$ and $H$ is the mean curvature of $\Sigma$ in $(M^3, g)$) and, amazingly, is nondecreasing under this "inverse mean curvature flow." This is seen by the fact that under inverse mean curvature flow, it follows from the Gauss equation and the second variation formula that

$$\frac{\mathrm{d}}{\mathrm{d}t} m_{\text{Hawking}}(\Sigma) = \sqrt{\frac{|\Sigma|}{16\pi}}\left[\frac{1}{2} + \frac{1}{16\pi}\int_\Sigma 2\frac{|\nabla_\Sigma H|^2}{H^2} + R - 2K + \frac{1}{2}(\lambda_1 - \lambda_2)^2\right]$$

when the flow is smooth, where $R$ is the scalar curvature of $(M^3, g)$, $K$ is the Gauss curvature of the surface $\Sigma$, and $\lambda_1$ and $\lambda_2$ are the eigenvalues of the second fundamental form of $\Sigma$, or principle curvatures. Hence,

$$R \geq 0$$

and

$$\int_\Sigma K \leq 4\pi \qquad [4]$$

(which is true for any connected surface by the Gauss–Bonnet theorem) imply

$$\frac{\mathrm{d}}{\mathrm{d}t} m_{\text{Hawking}}(\Sigma) \geq 0 \qquad [5]$$

Furthermore,

$$m_{\text{Hawking}}(\Sigma_0) = \sqrt{\frac{|\Sigma_0|}{16\pi}}$$

since $\Sigma_0$ is a minimal surface and has zero mean curvature. In addition, the Hawking mass of sufficiently round spheres at infinity in the asymptotically flat end of $(M^3, g)$ approaches the total mass $m$. Hence, if inverse mean curvature flow beginning with $\Sigma_0$ eventually flows to sufficiently round spheres at infinity, inequality [3] follows from inequality [5].

As noted by Jang and Wald, this argument only works when inverse mean curvature flow exists and is smooth, which is generally not expected to be the case. In fact, it is not hard to construct manifolds which do not admit a smooth inverse mean curvature flow. The problem is that if the mean curvature of the evolving surface becomes zero or is negative, it is not clear how to define the flow.

For 20 years, this heuristic argument lay dormant until the work of Huisken and Ilmanen in 1997. With a very clever new approach, Huisken and Ilmanen discovered how to reformulate inverse mean curvature flow using an energy minimization principle in such a way that the new generalized inverse mean curvature flow always exists. The added twist is that the surface sometimes jumps outward. However, when the flow is smooth, it equals the original inverse mean curvature flow, and the Hawking mass is still monotone. Hence, as will be described in the next section, their new flow produced the first complete proof of inequality [3] for a single black hole.

Coincidentally, the author found another proof of inequality [3], submitted in 1999, which works for any number of black holes. The approach involves flowing the original metric to a Schwarzschild metric (outside the horizon) in such a way that the area of the outermost minimal surface does not change and the

total mass is nonincreasing. Then, since the Schwarzschild metric gives equality in inequality [3], the inequality follows for the original metric.

Fortunately, the flow of metrics which is defined is relatively simple, and in fact stays inside the conformal class of the original metric. The outermost minimal surface flows outwards in this conformal flow of metrics, and encloses any compact set (and hence all of the topology of the original metric) in a finite amount of time. Furthermore, this conformal flow of metrics preserves nonnegative scalar curvature. We will describe this approach later in the paper.

Other contributions on the Penrose conjecture have also been made by Herzlich using the Dirac operator which Witten used to prove the positive mass theorem, by Gibbons in the special case of collapsing shells, by Tod, by Bartnik for quasi-spherical metrics, and by the present author using isoperimetric surfaces. There is also some interesting work of Ludvigsen and Vickers using spinors and Bergqvist, both concerning the Penrose inequality for null slices of a spacetime.

## Inverse Mean Curvature Flow

Geometrically, Huisken and Ilmanen's idea can be described as follows. Let $\Sigma(t)$ be the surface resulting from inverse mean curvature flow for time $t$ beginning with the minimal surface $\Sigma_0$. Define $\bar{\Sigma}(t)$ to be the outermost minimal area enclosure of $\Sigma(t)$. Typically, $\Sigma(t) = \bar{\Sigma}(t)$ in the flow, but in the case that the two surfaces are not equal, immediately replace $\Sigma(t)$ with $\bar{\Sigma}(t)$ and then continue flowing by inverse mean curvature.

An immediate consequence of this modified flow is that the mean curvature of $\bar{\Sigma}(t)$ is always non-negative by the first variation formula, since otherwise $\bar{\Sigma}(t)$ would be enclosed by a surface with less area. This is because if we flow a surface $\Sigma$ in the outward direction with speed $\eta$, the first variation of the area is $\int_\Sigma H\eta$, where $H$ is the mean curvature of $\Sigma$.

Furthermore, by stability, it follows that in the regions where $\bar{\Sigma}(t)$ has zero mean curvature, it is always possible to flow the surface out slightly to have positive mean curvature, allowing inverse mean curvature flow to be defined, at least heuristically at this point.

Furthermore, the Hawking mass is still monotone under this new modified flow. Notice that when $\Sigma(t)$ jumps outwards to $\bar{\Sigma}(t)$,

$$\int_{\bar{\Sigma}(t)} H^2 \leq \int_{\Sigma(t)} H^2$$

since $\bar{\Sigma}(t)$ has zero mean curvature where the two surfaces do not touch. Furthermore,

$$|\bar{\Sigma}(t)| = |\Sigma(t)|$$

since (this is a neat argument) $|\bar{\Sigma}(t)| \leq |\Sigma(t)|$ (since $\bar{\Sigma}(t)$ is a minimal area enclosure of $\Sigma(t)$) and we cannot have $|\bar{\Sigma}(t)| < |\Sigma(t)|$ since $\Sigma(t)$ would have jumped outwards at some earlier time. This is only a heuristic argument, but we can then see that the Hawking mass is nondecreasing during a jump by the above two equations.

This new flow can be rigorously defined, always exists, and the Hawking mass is monotone. Huisken and Ilmanen define $\Sigma(t)$ to be the level sets of a scalar valued function $u(x)$ defined on $(M^3, g)$ such that $u(x) = 0$ on the original surface $\Sigma_0$ and satisfies

$$\operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) = |\nabla u| \qquad [6]$$

in an appropriate weak sense. Since the left-hand side of the above equation is the mean curvature of the level sets of $u(x)$ and the right-hand side is the reciprocal of the flow rate, the above equation implies inverse mean curvature flow for the level sets of $u(x)$ when $|\nabla u(x)| \neq 0$.

Huisken and Ilmanen use an energy minimization principle to define weak solutions to eqn [6]. Equation [6] is said to be weakly satisfied in $\Omega$ by the locally Lipschitz function $u$ if for all locally Lipschitz $v$ with $\{v \neq u\} \subset\subset \Omega$,

$$J_u(u) \leq J_u(v)$$

where

$$J_u(v) := \int_\Omega |\nabla v| + v|\nabla u|$$

It can then be seen that the Euler–Lagrange equation of the above energy functional yields eqn [6].

In order to prove that a solution $u$ exists to the above two equations, Huisken and Ilmanen regularize the degenerate elliptic equation 6 to the elliptic equation

$$\operatorname{div}\left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \epsilon^2}}\right) = \sqrt{|\nabla u|^2 + \epsilon^2}$$

Solutions to the above equation are then shown to exist using the existence of a subsolution, and then taking the limit as $\epsilon$ goes to zero yields a weak solution to eqn [6]. There are many details which we are skipping here, but these are the main ideas.

As it turns out, weak solutions $u(x)$ to eqn [6] often have flat regions where $u(x)$ equals a constant. Hence, the level sets $\Sigma(t)$ of $u(x)$ will be

discontinuous in $t$ in this case, which corresponds to the "jumping out" phenomenon referred to at the beginning of this section.

We also note that since the Hawking mass of the level sets of $u(x)$ is monotone, this inverse mean curvature flow technique not only proves the Riemannian Penrose inequality, but also gives a new proof of the positive mass theorem in dimension 3. This is seen by letting the initial surface be a very small, round sphere (which will have approximately zero Hawking mass) and then flowing by inverse mean curvature, thereby proving $m \geq 0$.

The Huisken and Ilmanen inverse mean curvature flow also seems ideally suited for proving Penrose inequalities for 3-manifolds which have $R \geq -6$ and which are asymptotically hyperbolic. This situation occurs if $(M^3, g)$ is chosen to be a constant mean curvature slice of the spacetime or if the spacetime is defined to solve the Einstein equation with nonzero cosmological constant. In these cases, there exists a modified Hawking mass which in monotone under inverse mean curvature flow which is the usual Hawking mass plus $4(|\Sigma|/16\pi)^{3/2}$. However, because the monotonicity of the Hawking mass relies on the Gauss–Bonnet theorem, these arguments do not work in higher dimensions, at least so far. Also, because of the need for eqn [4], inverse mean curvature flow only proves the Riemannian Penrose inequality for a single black hole. In the next section, we present a technique which proves the Riemannian Penrose inequality for any number of black holes, and which can likely be generalized to higher dimensions.

## The Conformal Flow of Metrics

Given any initial Riemannian manifold $(M^3, g_0)$ which has non-negative scalar curvature and which is harmonically flat at infinity, we will define a continuous, one-parameter family of metrics $(M^3, g_t)$, $0 \leq t < \infty$. This family of metrics will converge to a three-dimensional Schwarzschild metric and will have other special properties which will allow us to prove the Riemannian Penrose inequality for the original metric $(M^3, g_0)$.

In particular, let $\Sigma_0$ be the outermost minimal surface of $(M^3, g_0)$ with area $A_0$. Then, we will also define a family of surfaces $\Sigma(t)$ with $\Sigma(0) = \Sigma_0$ such that $\Sigma(t)$ is minimal in $(M^3, g_t)$. This is natural since as the metric $g_t$ changes, we expect that the location of the horizon $\Sigma(t)$ will also change. Then, the interesting quantities to keep track of in this flow are $A(t)$, the total area of the horizon $\Sigma(t)$ in $(M^3, g_t)$, and $m(t)$, the total mass of $(M^3, g_t)$ in the chosen end.

In addition to all of the metrics $g_t$ having non-negative scalar curvature, we will also have the very nice properties that

$$A'(t) = 0$$
$$m'(t) \leq 0$$

for all $t \geq 0$. Then, since $(M^3, g_t)$ converges to a Schwarzschild metric (in an appropriate sense) which gives equality in the Riemannian Penrose inequality as described in the introduction,

$$m(0) \geq m(\infty) = \sqrt{\frac{A(\infty)}{16\pi}} = \sqrt{\frac{A(0)}{16\pi}} \qquad [7]$$

which proves the Riemannian Penrose inequality for the original metric $(M^3, g_0)$. The hard part, then, is to find a flow of metrics which preserves non-negative scalar curvature and the area of the horizon, decreases total mass, and converges to a Schwarzschild metric as $t$ goes to infinity.

### The Definition of the Flow

In fact, the metrics $g_t$ will all be conformal to $g_0$. This conformal flow of metrics can be thought of as the solution to a first-order ODE in $t$ defined by eqns [8]–[11]. Let

$$g_t = u_t(x)^4 g_0 \qquad [8]$$

and $u_0(x) \equiv 1$. Given the metric $g_t$, define

$$\Sigma(t) = \text{the outermost minimal area enclosure of } \Sigma_0 \text{ in } (M^3, g_t) \qquad [9]$$

where $\Sigma_0$ is the original outer minimizing horizon in $(M^3, g_0)$. In the cases in which we are interested, $\Sigma(t)$ will not touch $\Sigma_0$, from which it follows that $\Sigma(t)$ is actually a strictly outer minimizing horizon of $(M^3, g_t)$. Then given the horizon $\Sigma(t)$, define $v_t(x)$ such that

$$\begin{cases} \Delta_{g_0} v_t(x) \equiv 0 & \text{outside } \Sigma(t) \\ v_t(x) = 0 & \text{on } \Sigma(t) \\ \lim_{x\to\infty} v_t(x) = -\mathrm{e}^{-t} & \end{cases} \qquad [10]$$

and $v_t(x) \equiv 0$ inside $\Sigma(t)$. Finally, given $v_t(x)$, define

$$u_t(x) = 1 + \int_0^t v_s(x)\,\mathrm{d}s \qquad [11]$$

so that $u_t(x)$ is continuous in $t$ and has $u_0(x) \equiv 1$.

Note that eqn [11] implies that the first-order rate of change of $u_t(x)$ is given by $v_t(x)$. Hence, the first-order rate of change of $g_t$ is a function of itself, $g_0$, and $v_t(x)$ which is a function of $g_0$, $t$, and $\Sigma(t)$ which is in turn a function of $g_t$ and $\Sigma_0$. Thus, the first-order rate of change of $g_t$ is a function of $t, g_t, g_0$, and $\Sigma_0$.

**Theorem 2** *Taken together, eqns [8]–[11] define a first-order ODE in t for $u_t(x)$ which has a solution which is Lipschitz in the t variable, $C^1$ in the x variable everywhere, and smooth in the x variable outside $\Sigma(t)$. Furthermore, $\Sigma(t)$ is a smooth, strictly outer minimizing horizon in $(M^3, g_t)$ for all $t \geq 0$, and $\Sigma(t_2)$ encloses but does not touch $\Sigma(t_1)$ for all $t_2 > t_1 \geq 0$.*

Since $v_t(x)$ is a superharmonic function in $(M^3, g_0)$ (harmonic everywhere except on $\Sigma(t)$, where it is weakly superharmonic), it follows that $u_t(x)$ is superharmonic as well. Thus, from eqn [11] we see that $\lim_{x\to\infty} u_t(x) = \mathrm{e}^{-t}$ and consequently that $u_t(x) > 0$ for all $t$ by the maximum principle. Then, since

$$R(g_t) = u_t(x)^{-5}(-8\Delta_{g_0} + R(g_0))u_t(x) \qquad [12]$$

it follows that $(M^3, g_t)$ is an asymptotically flat manifold with non-negative scalar curvature.

Even so, it still may not seem like $g_t$ is particularly naturally defined since the rate of change of $g_t$ appears to depend on $t$ and the original metric $g_0$ in eqn [10]. We would prefer a flow where the rate of change of $g_t$ can be defined purely as a function of $g_t$ (and $\Sigma_0$ perhaps), and interestingly enough this actually does turn out to be the case! The present author has proved this very important fact and defined a new equivalence class of metrics called the harmonic conformal class. Then, once we decide to find a flow of metrics which stays inside the harmonic conformal class of the original metric (outside the horizon) and keeps the area of the horizon $\Sigma(t)$ constant, then we are basically forced to choose the particular conformal flow of metrics defined above.

**Theorem 3** *The function $A(t)$ is constant in t and $m(t)$ is nonincreasing in t, for all $t \geq 0$.*

The fact that $A'(t) = 0$ follows from the fact that to first order the metric is not changing on $\Sigma(t)$ (since $v_t(x) = 0$ there) and from the fact that to first order the area of $\Sigma(t)$ does not change as it moves outward since $\Sigma(t)$ is a critical point for area in $(M^3, g_t)$. Hence, the interesting part of Theorem 3 is proving that $m'(t) \leq 0$. Curiously, this follows from a nice trick using the Riemannian positive mass theorem, which we describe later.

Another important aspect of this conformal flow of the metric is that outside the horizon $\Sigma(t)$, the manifold $(M^3, g_t)$ becomes more and more spherically symmetric and "approaches" a Schwarzschild manifold $(\boldsymbol{R}^3\backslash\{0\}, s)$ in the limit as $t$ goes to $\infty$. More precisely,

**Theorem 4** *For sufficiently large t, there exists a diffeomorphism $\phi_t$ between $(M^3, g_t)$ outside the horizon $\Sigma(t)$ and a fixed Schwarzschild manifold $(\boldsymbol{R}^3\backslash\{0\}, s)$ outside its horizon. Furthermore, for all $\epsilon > 0$, there exists a T such that for all $t > T$, the metrics $g_t$ and $\phi_t^*(s)$ (when determining the lengths of unit vectors of $(M^3, g_t)$) are within $\epsilon$ of each other and the total masses of the 2-manifolds are within $\epsilon$ of each other. Hence,*

$$\lim_{t\to\infty} \frac{m(t)}{\sqrt{A(t)}} = \sqrt{\frac{1}{16\pi}}$$

Theorem 4 is not that surprising really although a careful proof is reasonably long. However, if one is willing to believe that the flow of metrics converges to a spherically symmetric metric outside the horizon, then Theorem 4 follows from two facts. The first fact is that the scalar curvature of $(M^3, g_t)$ eventually becomes identically zero outside the horizon $\Sigma(t)$ (assuming $(M^3, g_0)$ is harmonically flat). This follows from the facts that $\Sigma(t)$ encloses any compact set in a finite amount of time, that harmonically flat manifolds have zero scalar curvature outside a compact set, that $u_t(x)$ is harmonic outside $\Sigma(t)$, and eqn [12]. The second fact is that the Schwarzschild metrics are the only complete, spherically symmetric 3-manifolds with zero scalar curvature (except for the flat metric on $R^3$).

The Riemannian Penrose inequality, inequality [3], then follows from eqn [7] using Theorems 2–4, for harmonically flat manifolds. Since asymptotically flat manifolds can be approximated arbitrarily well by harmonically flat manifolds while changing the relevant quantities arbitrarily little, the asymptotically flat case also follows. Finally, the case of equality of the Penrose inequality follows from a more careful analysis of these same arguments.

### Qualitative Discussion

Figures 3 and 4 are meant to help illustrate some of the properties of the conformal flow of the metric. Figure 3 is the original metric which has a strictly outer minimizing horizon $\Sigma_0$. As $t$ increases, $\Sigma(t)$ moves outwards, but never inwards. In Figure 4, we can observe one of the consequences of the fact that $A(t) = A_0$ is constant in $t$. Since the metric is not changing inside $\Sigma(t)$, all of the horizons $\Sigma(s)$, $0 \leq s \leq t$

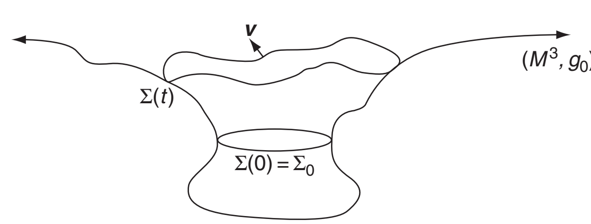



**Figure 3** Original metric having a strictly outer minimizing horizon $\Sigma_0$.

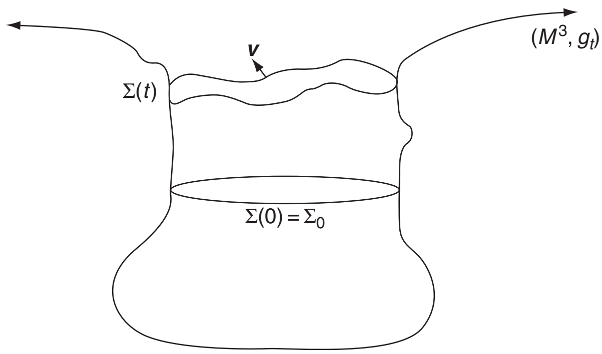



**Figure 4** Metric after time *t*.

have area $A_0$ in $(M^3, g_t)$. Hence, inside $\Sigma(t)$, the manifold $(M^3, g_t)$ becomes cylinder-like in the sense that it is laminated (i.e., foliated but with some gaps allowed) by all of the previous horizons which all have the same area $A_0$ with respect to the metric $g_t$.

Now let us suppose that the original horizon $\Sigma_0$ of $(M^3, g)$ had two components, for example. Then each of the components of the horizon will move outwards as $t$ increases, and at some point before they touch they will suddenly jump outwards to form a horizon with a single component enclosing the previous horizon with two components. Even horizons with only one component will sometimes jump outwards, but no more than a countable number of times. It is interesting that this phenomenon of surfaces jumping is also found in the Huisken–Ilmanen approach to the Penrose conjecture using their generalized $1/H$ flow.

### Proof that $m'(t) \leq 0$

The most surprising aspect of the flow defined earlier is that $m'(t) \leq 0$. As mentioned in that section, this important fact follows from a nice trick using the Riemannian positive mass theorem.

The first step is to realize that while the rate of change of $g_t$ appears to depend on $t$ and $g_0$, this is in fact an illusion. As described in detail by Bray, the rate of change of $g_t$ can be described purely in terms of $g_t$ (and $\Sigma_0$). It is also true that the rate of change of $g_t$ depends only on $g_t$ and $\Sigma(t)$. Hence, there is no special value of $t$, so proving $m'(t) \leq 0$ is equivalent to proving $m'(0) \leq 0$. Thus, without loss of generality, we take $t = 0$ for convenience.

Now expand the harmonic function $v_0(x)$, defined in eqn [10], using spherical harmonics at infinity, to get

$$v_0(x) = -1 + \frac{c}{|x|} + \mathcal{O}\left(\frac{1}{|x|^2}\right) \qquad [13]$$

for some constant $c$. Since the rate of change of the metric $g_t$ at $t = 0$ is given by $v_0(x)$ and since the total mass $m(t)$ depends on the $1/r$ rate at which the metric $g_t$ becomes flat at infinity (see eqn [2]), it is not surprising that direct calculation gives us that

$$m'(0) = 2(c - m(0))$$

Hence, to show that $m'(0) \leq 0$, we need to show that

$$c \leq m(0) \qquad [14]$$

In fact, counterexamples to eqn [14] can be found if we remove either of the requirements that $\Sigma(0)$ (which is used in the definition of $v_0(x)$) be a minimal surface or that $(M^3, g_0)$ have non-negative scalar curvature. Hence, we quickly see that eqn [14] is a fairly deep conjecture which says something quite interesting about manifold with non-negative scalar curvature. Well, the Riemannian positive mass theorem is also a deep conjecture which says something quite interesting about manifolds with non-negative scalar curvature. Hence, it is natural to try to use the Riemannian positive mass theorem to prove eqn [14].

Thus, we want to create a manifold whose total mass depends on $c$ from eqn [13]. The idea is to use a reflection trick similar to one used by Bunting and Masood-ul-Alam (1987) for another purpose. First, remove the region of $M^3$ inside $\Sigma(0)$ and then reflect the remainder of $(M^3, g_0)$ through $\Sigma(0)$. Define the resulting Riemannian manifold to be $(\bar{M}^3, \bar{g}_0)$ which has two asymptotically flat ends since $(M^3, g_0)$ has exactly one asymptotically flat end not contained by $\Sigma(0)$. Note that $(\bar{M}^3, \bar{g}_0)$ has non-negative scalar curvature everywhere except on $\Sigma(0)$ where the metric has corners. In fact, the fact that $\Sigma(0)$ has zero mean curvature (since it is a minimal surface) implies that $(\bar{M}^3, \bar{g}_0)$ has "distributional" non-negative scalar curvature everywhere, even on $\Sigma(0)$. This notion is made rigorous by Bray. Thus, we have used the fact that $\Sigma(0)$ is minimal in a critical way.

Recall from eqn [10] that $v_0(x)$ was defined to be the harmonic function equal to zero on $\Sigma(0)$ which goes to –1 at infinity. We want to reflect $v_0(x)$ to be defined on all of $(\bar{M}^3, \bar{g}_0)$. The trick here is to define $v_0(x)$ on $(\bar{M}^3, \bar{g}_0)$ to be the harmonic function which goes to –1 at infinity in the original end and goes to 1 at infinity in the reflect end. By symmetry, $v_0(x)$ equals 0 on $\Sigma(0)$ and so agrees with its original definition on $(M^3, g_0)$.

The next step is to compactify one end of $(\bar{M}^3, \bar{g}_0)$. By the maximum principle, we know that $v_0(x) > -1$ and $c > 0$, so the new Riemannian manifold $(\bar{M}^3, (v_0(x) + 1)^4 \bar{g}_0)$ does the job quite nicely and compactifies the original end to a point. In fact, the compactified point at infinity and the metric there

can be filled in smoothly (using the fact that $(M^3, g_0)$ is harmonically flat). It then follows from eqn [12] that this new compactified manifold has non-negative scalar curvature since $v_0(x) + 1$ is harmonic.

The last step is simply to apply the Riemannian positive mass theorem to $(\bar{M}^3, (v_0(x) + 1)^4 \bar{g}_0)$. It is not surprising that the total mass $\tilde{m}(0)$ of this manifold involves $c$, but it is quite lucky that direct calculation yields

$$\tilde{m}(0) = -4(c - m(0))$$

which must be positive by the Riemannian positive mass theorem. Thus, we have that

$$m'(0) = 2(c - m(0)) = -\tfrac{1}{2}\tilde{m}(0) \leq 0$$

## Open Questions and Applications

Now that the Riemannian Penrose conjecture has been proved, what are the next interesting directions? What applications can be found? Is this subject only of physical interest, or are there possibly broader applications to other problems in mathematics?

Clearly, the most natural open problem is to find a way to prove the general Penrose inequality in which $M^3$ is allowed to have any second fundamental form in the spacetime. There is good reason to think that this may follow from the Riemannian Penrose inequality, although this is a bit delicate. On the other hand, the general positive mass theorem followed from the Riemannian positive mass theorem as was originally shown by Schoen and Yau using an idea due to Jang. For physicists, this problem is definitely a top priority since most spacetimes do not even admit zero-second fundamental form spacelike slices.

Another interesting question is to ask these same questions in higher dimensions. The author is currently working on a paper to prove the Riemannian Penrose inequality in dimensions $<8$. Dimension 8 and higher are harder because of the surprising fact that minimal hypersurfaces (and hence apparent horizons of black holes) can have codimension 7 singularities (points where the hypersurface is not smooth). This curious technicality is also the reason that the positive mass conjecture is still open in dimensions 8 and higher for manifolds which are not spin.

Naturally, it is harder to tell what the applications of these techniques might be to other problems, but already there have been some. One application is to the famous Yamabe problem: given a compact 3-manifold $M^3$, define $E(g) = \int_{M^3} R_g \, dV_g$ where $g$ is scaled so that the total volume of $(M^3, g)$ is 1, $R_g$ is the scalar curvature at each point, and $dV_g$ is the volume form. An idea due to Yamabe was to try to construct canonical metrics on $M^3$ by finding critical points of this energy functional on the space of metrics. Define $C(g)$ to be the infimum of $E(\bar{g})$ over all metrics $\bar{g}$ conformal to $g$. Then the (topological) Yamabe invariant of $M^3$, denoted here as $Y(M^3)$, is defined to be the supremum of $C(g)$ over all metrics $g$. $Y(S^3) = 6 \cdot (2\pi^2)^{2/3} \equiv Y_1$ is known to be the largest possible value for Yamabe invariants of 3-manifolds. It is also known that $Y(T^3) = 0$ and $Y(S^2 \times S^1) = Y_1 = Y(S^2 \tilde{\times} S^1)$, where $S^2 \tilde{\times} S^1$ is the nonorientable $S^2$ bundle over $S^1$.

The author, working with Andre Neves on a problem suggested by Richard Schoen, recently was able to compute the Yamabe invariant of $RP^3$ using inverse mean curvature flow techniques and found that $Y(RP^3) = Y_1/2^{2/3} \equiv Y_2$. A corollary is $Y(RP^2 \times S^1) = Y_2$ as well. These techniques also yield the surprisingly strong result that the only prime 3-manifolds with Yamabe invariant larger than $RP^3$ are $S^3, S^2 \times S^1$, and $S^2 \tilde{\times} S^1$. The Poincare conjecture for 3-manifolds with Yamabe invariant greater than $RP^3$ is therefore a corollary. Furthermore, the problem of classifying 3-manifolds is known to reduce to the problem of classifying prime 3-manifolds. The Yamabe approach then would be to make a list of prime 3-manifolds ordered by Y. The first five prime 3-manifolds on this list are therefore $S^3, S^2 \times S^1, S^2 \tilde{\times} S^1, RP^3$, and $RP^2 \times S^1$.

## Acknowledgments

This research was supported in part by NSF grants #DMS-9706006 and #DMS-9971960.

*See also:* Black Hole Mechanics; General Relativity: Experimental Tests; General Relativity: Overview; Geometric Analysis and General Relativity; Holomorphic Dynamics; Mirror Symmetry: A Geometric Survey; Spinors and Spin Coefficients; Stationary Black Holes.

## Further Reading

Bray HL (1997) *The Penrose Inequality in General Relativity and Volume Comparison Theorems Involving Scalar Curvature*. Thesis, Stanford University.

Bray HL (2001) Proof of the Riemannian Penrose inequality using the positive mass theorem. *Journal of Differential Geometry*. 59: 177–267.

Bray HL and Neves A (2004) Classification of Prime 3-Manifolds with Yamabe Invariant Greater than $RP^3$. *Annals of Mathematics* 159: 407–424.

Bray HL and Schoen RM (1999) Recent proofs of the Riemannian Penrose conjecture. In: Yau S-T (ed.) *Current Developments in Mathematics*.

Bunting and Masood-ul-Alam (1987) Non-existence of multiple black holes in asymptotically euclidean static vacuum space-time. *General Relativity and Gravitation* 19(2).

Christodoulou D and Yau S-T (1988) Some remarks on the quasi-local mass. *Contemporary Mathematics* 71: 9–14.

Hawking SW and Ellis GFR (1973) *The Large-Scale Structure of Space-Time*.Cambridge: Cambridge University Press.
Huisken G and Ilmanen T (2001) The inverse mean curvature flow and the Riemannian Penrose inequality. *Journal of Differential Geometry* 59: 353–437.
Huisken G and Ilmanen T (1997a) *A Note on Inverse Mean Curvature Flow*. Proceedings of the Workshop on Nonlinear Partial Differential Equations, *Saitama University*, *Sept. 1997* (available from Saitama University).
Huisken G and Ilmanen T (1997b) The Riemannian Penrose inequality. *International Mathematics Research Notices* 20: 1045–1058.
Jang PS and Wald RM (1977) The positive energy conjecture and the cosmic censor hypothesis. *Journal of Mathematical Physics* 18: 41–44.
Parker TH and Taubes CH (1982) On Witten's proof of the positive energy theorem. *Communications in Mathematical Physics* 84: 223–238.
Penrose R (1973) Naked singularities. *Annals of the New York Academy of Sciences* 224: 125–134.
Schoen R and Yau S-T (1979a) On the proof of the positive mass conjecture in general relativity. *Communications in Mathematical Physics* 65: 45–76.
Schoen R and Yau S-T (1979b) Positivity of the total mass of a general space-time. *Physical Review Letters* 43: 1457–1459.
Schoen R and Yau S-T (1981) Proof of the positive mass theorem II. *Communications in Mathematical Physics* 79: 231–260.
Witten E (1981) A new proof of the positive energy theorem. *Communications in Mathematical Physics* 80: 381–402.

# Geometric Measure Theory

**G Alberti**, Università di Pisa, Pisa, Italy



## Introduction

The aim of these pages is to give a brief, self-contained introduction to that part of geometric measure theory which is more directly related to the calculus of variations, namely the theory of currents and its applications to the solution of Plateau problem. (The theory of finite-perimeter sets, which is closely related to currents and to the Plateau problem, is treated in the article Free Interfaces and Free Discontinuities: Variational Problems in the Encyclopedia.)

Named after the Belgian physicist J A F Plateau (1801–1883), this problem was originally formulated as follows: find the surface of minimal area spanning a given curve in the space. Nowadays, it is mostly intended in the sense of developing a mathematical framework where the existence of $k$-dimensional surfaces of minimal volume that span a prescribed boundary can be rigorously proved. Indeed, several solutions have been proposed in the last century, none of which is completely satisfactory.

One difficulty is that the infimum of the area among all smooth surfaces with a certain boundary may not be attained. More precisely, it may happen that all minimizing sequences (i.e., sequences of smooth surfaces whose area approaches the infimum) converge to a singular surface. Therefore, one is forced to consider a larger class of admissible surfaces than just smooth ones (in fact, one might want to do this also for modeling reasons – this is indeed the case with soap films, soap bubbles, and other capillarity problems). But what does it mean that a set "spans" a given curve? and what should we intend by area of a set which is not a smooth surface?

The theory of integral currents developed by Federer and Fleming (1960) provides a class of generalized (oriented) surfaces with well-defined notions of boundary and area (called mass) where the existence of minimizers can be proved by direct methods. More precisely, this class is large enough to have good compactness properties with respect to a topology that makes the mass a lower-semicontinuous functional. This approach turned out to be quite powerful and flexible, and in the last decades the theory of currents has found applications in several different areas, from dynamical systems (in particular, Mather theory) to the theory of foliations, to optimal transport problems.

## Hausdorff Measures, Dimension, and Rectifiability

The volume of a smooth $d$-dimensional surface in $\mathbb{R}^n$ is usually defined using parametrizations by subsets of $\mathbb{R}^d$. The notion of Hausdorff measure allows to compute the $d$-dimensional volume using coverings instead of parametrizations, and, what is more important, applies to all sets in $\mathbb{R}^n$, and makes sense even if $d$ is not an integer. Attached to Hausdorff measure is the notion of Hausdorff dimension. Again, it can be defined for all sets in $\mathbb{R}^n$ and is not necessarily an integer. The last fundamental notion is rectifiability: $k$-rectifiable sets can be roughly understood as the largest class of $k$-dimensional sets for which it is still possible to define a $k$-dimensional tangent bundle, even if only in a very weak sense. They are essential to the construction of integral currents.

### Hausdorff Measure

Let $d \geq 0$ be a positive real number. Given a set $E$ in $\mathbb{R}^n$, for every $\delta > 0$ we set

$$\mathscr{H}^d_\delta(E) := \frac{\omega_d}{2^d} \inf \left\{ \sum_j (\operatorname{diam}(E_j))^d \right\} \qquad [1]$$

where $\omega_d$ is the $d$-dimensional volume of the unit ball in $\mathbb{R}^d$ whenever $d$ is an integer (there is no canonical choice for $\omega_d$ when $d$ is not an integer; a convenient one is $\omega_d = 2^d$), and the infimum is taken over all countable families of sets $\{E_j\}$ that cover $E$ and whose diameters satisfy $\operatorname{diam}(E_j) \leq \delta$. The $d$-dimensional Hausdorff measure of $E$ is

$$\mathscr{H}^d(E) := \lim_{\delta \to 0} \mathscr{H}^d_\delta(E) \qquad [2]$$

(the limit exists because $\mathscr{H}^d_\delta(E)$ is decreasing in $\delta$).

#### Remarks

(i) $\mathscr{H}^d$ is called $d$-dimensional because of its scaling behavior: if $E_\lambda$ is a copy of $E$ scaled homothetically by a factor $\lambda$, then

$$\mathscr{H}^d(E_\lambda) = \lambda^d \mathscr{H}^d(E)$$

Thus, $\mathscr{H}^1$ scales like the length, $\mathscr{H}^2$ scales like the area, and so on.

(ii) The measure $\mathscr{H}^d$ is clearly invariant under rigid motions (translations and rotations). This implies that $\mathscr{H}^d$ agrees on $\mathbb{R}^d$ with the Lebesgue measure up to some constant factor; the renormalization constant $\omega_d/2^d$ in [1] makes this factor equal to 1. Thus, $\mathscr{H}^d(E)$ agrees with the usual $d$-dimensional volume for every set $E$ in $\mathbb{R}^d$, and the area formula shows that the same is true if $E$ is (a subset of) a $d$-dimensional surface of class $C^1$ in $\mathbb{R}^n$.

(iii) Besides the Hausdorff measure, there are several other, less popular notions of $d$-dimensional measure: all of them are invariant under rigid motion, scale in the expected way, and agree with $\mathscr{H}^d$ for sets contained in $\mathbb{R}^d$ or in a $d$-dimensional surface of class $C^1$, and yet they differ for other sets (for further details, see Federer (1996, section 2.10)).

(iv) The definition of $\mathscr{H}^d(E)$ uses only the notion of diameter, and therefore makes sense when $E$ is a subset of an arbitrary metric space. Note that $\mathscr{H}^d(E)$ depends only on the restriction of the metric to $E$, and not on the ambient space.

(v) The measure $\mathscr{H}^d$ is countably additive on the $\sigma$-algebra of Borel sets in $\mathbb{R}^n$, but not on all sets; to avoid pathological situations, we shall always assume that sets and maps are Borel measurable.

### Hausdorff Dimension

According to intuition, the length of a surface should be infinite, while the area of a curve should be null. These are indeed particular cases of the following implications:

$$\mathscr{H}^d(E) > 0 \;\Rightarrow\; \mathscr{H}^{d'}(E) = \infty \quad \text{for } d' < d$$
$$\mathscr{H}^d(E) < \infty \;\Rightarrow\; \mathscr{H}^{d'}(E) = 0 \quad \text{for } d' > d$$

Hence, the infimum of all $d$ such that $\mathscr{H}^d(E) = 0$ and the supremum of all $d$ such that $\mathscr{H}^d(E) = \infty$ coincide. This number is called Hausdorff dimension of $E$, and denoted by $\dim_{\mathrm{H}}(E)$. For surface of class $C^1$, the notion of Hausdorff dimension agrees with the usual one. Example of sets with nonintegral dimension are described in the next subsection.

#### Remarks

(i) Note that $\mathscr{H}^d(E)$ may be 0 or $\infty$ even for $d = \dim_{\mathrm{H}}(E)$.

(ii) The Hausdorff dimension of a set $E$ is strictly related to the metric on $E$, and not just to the topology. Indeed, it is preserved under diffeomorphisms but not under homeomorphisms, and it does not always agree with the topological dimension. For instance, the Hausdorff dimension of the graph of a continuous function $f: \mathbb{R} \to \mathbb{R}$ can be any number between 1 and 2 (included).

(iii) For nonsmooth sets, the Hausdorff dimension does not always conform to intuition: for example, the dimension of a Cartesian product $E \times F$ of compact sets does not agree in general with the sum of the dimensions of $E$ and $F$.

(iv) There are many other notions of dimension besides Hausdorff and topological ones. Among these, packing dimension and box-counting dimension have interesting applications (see Falconer (2003, chapters 3 and 4)).

### Self-Similar Fractals

Interesting examples of sets with nonintegral dimension are self-similar fractals. We present here a simplified version of a construction due to Hutchinson (Falconer 2003, chapter 9). Let $\{\Psi_i\}$ be a finite set of similitudes of $\mathbb{R}^n$ with scaling factor $\lambda_i < 1$, and assume that there exists a bounded open set $V$ such that the sets $V_i := \Psi_i(V)$ are pairwise disjoint and contained in $V$. The self-similar fractal associated with the system $\{\Psi_i\}$ is the compact set $C$ that satisfies

$$C = \bigcup_i \Psi_i(C) \qquad [3]$$

The term "self-similar" follows by the fact that $C$ can be written as a union of scaled copies of itself.

The existence (and uniqueness) of such a $C$ follows from a standard fixed-point argument applied to the map $C \mapsto \bigcup \Psi_i(C)$. The dimension $d$ of $C$ is the unique solution of the equation

$$\sum_i \lambda_i^d = 1 \qquad [4]$$

Formula [4] can be easily justified: if the sets $\Psi_i(C)$ are disjoint – and the assumption on $V$ implies that this almost the case – then [3] implies $\mathscr{H}^d(C) = \Sigma \mathscr{H}^d(\Psi_i(C)) = \Sigma \lambda_i^d \mathscr{H}^d(C)$, and therefore $\mathscr{H}^d(C)$ can be positive and finite if and only if $d$ satisfies [4].

An example of this construction is the usual Cantor set in $\mathbb{R}$, which is given by the similitudes

$$\Psi_1(x) := \tfrac{1}{3}x \text{ and } \Psi_2(x) := \tfrac{2}{3} + \tfrac{1}{3}x$$

By [4], its dimension is $d = \log 2/\log 3$. Other examples are described in **Figures 1–3**.

### Rectifiable Sets

Given an integer $k = 1, \ldots, n$, we say that a set $E$ in $\mathbb{R}^n$ is $k$-rectifiable if it can be covered by a countable family of sets $\{S_j\}$ such that $S_0$ is $\mathscr{H}^k$-negligible (i.e., $\mathscr{H}^k(S_0) = 0$) and $S_j$ is a $k$-dimensional surface of class $C^1$ for $j = 1, 2, \ldots$ Note that $\dim_H(E) \le k$ because each $S_j$ has dimension $k$.

A $k$-rectifiable set $E$ bears little resemblance to smooth surfaces (it can be everywhere dense!), but it still admits a suitably weak notion of tangent bundle.

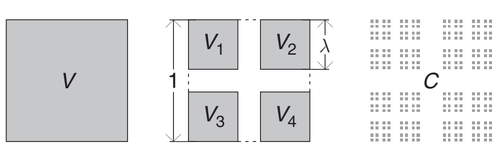



**Figure 1** The maps $\Psi_i, i = 1, \ldots, 4$, take the square $V$ into the squares $V_i$ at the corners of $V$. The scaling factor is $\lambda$ for all $i$, hence $\dim_H(C) = \log 4/(-\log \lambda)$. Note that $\dim_H(C)$ can be any number between 0 and 2, including 1.

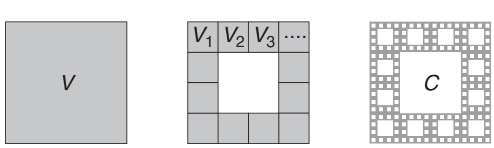



**Figure 2** A self-similar fractal with more complicated topology. The scaling factor is 1/4 for all twelve similitudes, hence $\dim_H(C) = \log 12/\log 4$.

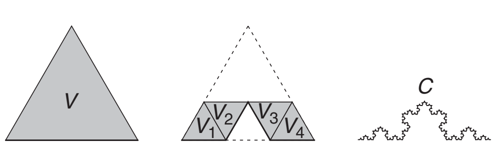



**Figure 3** The von Koch curve (or snowflake). The scaling factor is 1/3 for all four similitudes, hence $\dim_H(C) = \log 4/\log 3$.

More precisely, it is possible to associate with every $x \in E$ a $k$-dimensional subspace of $\mathbb{R}^n$, denoted by $\mathrm{Tan}(E, x)$, so that for every $k$-dimensional surface $S$ of class $C^1$ in $\mathbb{R}^n$ there holds

$$\mathrm{Tan}(E, x) = \mathrm{Tan}(S, x) \quad \text{for } \mathscr{H}^k\text{-a.e. } x \in E \cap S \qquad [5]$$

where $\mathrm{Tan}(S, x)$ is the tangent space to $S$ at $x$ according to the usual definition.

It is not difficult to see that $\mathrm{Tan}(E, x)$ is uniquely determined by [5] up to an $\mathscr{H}^k$-negligible amount of points $x \in E$, and if $E$ is a surface of class $C^1$, then it agrees with the usual tangent space for $\mathscr{H}^k$-almost all points of $E$.

**Remarks**

(i) In the original definition of rectifiability, the sets $S_j$ with $j > 0$ are Lipschitz images of $\mathbb{R}^k$, that is, $S_j := f_j(\mathbb{R}^k)$, where $f_j : \mathbb{R}^k \to \mathbb{R}^n$ is a Lipschitz map. It can be shown that this definition is equivalent to the one above.

(ii) The construction of the tangent bundle is straightforward: Let $\{S_j\}$ be a covering of $E$ as earlier, and set $\mathrm{Tan}(E, x) := \mathrm{Tan}(S_j, x)$, where $j$ is the smallest positive integer such that $x \in S_j$. Then [5] is an immediate corollary of the following lemma: if $S$ and $S'$ are $k$-dimensional surfaces of class $C^1$ in $\mathbb{R}^n$, then $\mathrm{Tan}(S, x) = \mathrm{Tan}(S', x)$ for $\mathscr{H}^k$-almost every $x \in S \cap S'$.

(iii) A set $E$ in $\mathbb{R}^n$ is called purely $k$-unrectifiable if it contains no $k$-rectifiable subset with positive $k$-dimensional measure, or, equivalently, if $\mathscr{H}^k(E \cap S) = 0$ for every $k$-dimensional surface $S$ of class $C^1$. For instance, every product $E := E_1 \times E_2$, where $E_1$ and $E_2$ are $\mathscr{H}^1$-negligible sets in $\mathbb{R}$ is a purely 1-unrectifiable set in $\mathbb{R}^2$ (it suffices to show that $\mathscr{H}^1(E \cap S) = 0$ whenever $S$ is the graph of a function $f : \mathbb{R} \to \mathbb{R}$ of class $C^1$, and this follows by the usual formula for the length of the graph). Note that the Hausdorff dimension of such product sets can be any number between 0 and 2, hence rectifiability is not related to dimension. The self-similar fractals described in **Figures 1 and 3** are both purely 1-unrectifiable.

### Rectifiable Sets with Finite Measure

If $E$ is a $k$-rectifiable set with finite (or locally finite) $k$-dimensional measure, then $\mathrm{Tan}(E, x)$ can be related to the behavior of $E$ close to the point $x$.

Let $B(x, r)$ be the open ball in $\mathbb{R}^n$ with center $x$ and radius $r$, and let $C(x, T, a)$ be the cone with center $x$, axis $T$ – a $k$-dimensional subspace of $\mathbb{R}^n$ – and amplitude $\alpha = \arcsin a$, that is,

$$C(x, T, a) := \{x' \in \mathbb{R}^n\colon\ \mathrm{dist}(x' - x, T) \le a|x' - x|\}$$

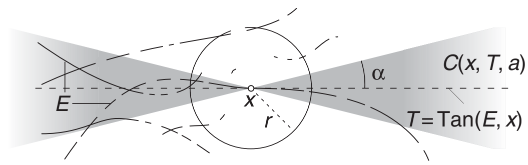



**Figure 4** A rectifiable set $E$ close to a point $x$ of approximate tangency. The part of $E$ contained in the ball $B(x, r)$ but not in the cone $C(x, T, a)$ is not empty, but only small in measure.

For $\mathscr{H}^k$-almost every $x \in E$, the measure of $E \cap B(x, r)$ is asymptotically equivalent, as $r \to 0$, to the measure of a flat disk of radius $r$, that is,

$$\mathscr{H}^k(E \cap B(x, r)) \sim \omega_k r^k$$

Moreover, the part of $E$ contained in $B(x, r)$ is mostly located close to the tangent plane $\mathrm{Tan}(E, x)$, that is,

$$\mathscr{H}^k(E \cap B(x, r) \cap C(x, \mathrm{Tan}(E, x), a)) \sim \omega_k r^k \quad \text{for every } a > 0$$

When this condition holds, $\mathrm{Tan}(E, x)$ is called the approximate tangent space to $E$ at $x$ (see Figure 4).

### The Area Formula

The area formula allows to compute the measure $\mathscr{H}^k(\Phi(E))$ of the image of a set $E$ in $\mathbb{R}^k$ as the integral over $E$ of a suitably defined Jacobian determinant of $\Phi$. When $\Phi$ is injective and takes values in $\mathbb{R}^k$, we recover the usual change of variable formula for multiple integrals.

We consider first the linear case. If $L$ is a linear map from $\mathbb{R}^k$ to $\mathbb{R}^m$ with $m \geq k$, the volume ratio $\rho := \mathscr{H}^k(L(E))/\mathscr{H}^k(E)$ does not depend on $E$, and agrees with $|\det(PL)|$, where $P$ is any linear isometry from the image of $L$ into $\mathbb{R}^k$, and $\det(PL)$ is the determinant of the $k \times k$ matrix associated with $PL$. The volume ratio $\rho$ can be computed using one of the following identities:

$$\rho = \sqrt{\det(L^* L)} = \sqrt{\sum (\det M)^2} \qquad [6]$$

where $L^*$ is the adjoint of $L$ (thus, $L^* L$ is a linear map from $\mathbb{R}^k$ into $\mathbb{R}^k$), and the sum in the last term is taken over all $k \times k$ minors $M$ of the matrix associated with $L$.

Let $\Phi : \mathbb{R}^k \to \mathbb{R}^m$ be a map of class $C^1$ with $m \geq k$, and $E$ a set in $\mathbb{R}^k$. Then

$$\int_{\Phi(E)} \#(\Phi^{-1}(y) \cap E) \, \mathrm{d}\mathscr{H}^k(y) = \int_E J(x) \, \mathrm{d}\mathscr{H}^k(x) \qquad [7]$$

where $\#A$ stands for the number of elements of $A$, and the Jacobian $J$ is

$$J(x) := \sqrt{\det(\nabla\Phi(x)^* \nabla\Phi(x))} \qquad [8]$$

Note that the left-hand side of [7] is $\mathscr{H}^k(\Phi(E))$ when $\Phi$ is injective.

**Remark**

Formula [7] holds even if $E$ is a $k$-rectifiable set in $\mathbb{R}^n$. In this case, the gradient $\nabla\Phi(x)$ in [8] should be replaced by the tangential derivative of $\Phi$ at $x$ (viewed as a linear map from $\mathrm{Tan}(E, x)$ into $\mathbb{R}^m$). No version of formula [7] is available when $E$ is not rectifiable.

## Vectors, Covectors, and Differential Forms

In this section, we review some basic notions of multilinear algebra. We have chosen a definition of $k$-vectors and $k$-covectors in $\mathbb{R}^n$, and of the corresponding exterior products, which is quite convenient for computations, even though not as satisfactory from the formal viewpoint. The main drawback is that it depends on the choice of a standard basis of $\mathbb{R}^n$, and therefore cannot be used to define forms (and currents) when the ambient space is a general manifold.

### *k*-Vectors and Exterior Product

Let $\{e_1, \ldots, e_n\}$ be the standard basis of $\mathbb{R}^n$. Given an integer $k \leq n$, $I(n, k)$ is the set of all multi-indices $\boldsymbol{i} = (i_1, \ldots, i_k)$ with $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, and for every $\boldsymbol{i} \in I(n, k)$ we introduce the expression

$$e_{\boldsymbol{i}} = e_{i_1} \wedge e_{i_2} \wedge \cdots \wedge e_{i_k}$$

A $k$-vector in $\mathbb{R}^n$ is any formal linear combination $\Sigma \alpha_{\boldsymbol{i}} e_{\boldsymbol{i}}$ with $\alpha_{\boldsymbol{i}} \in \mathbb{R}$ for every $\boldsymbol{i} \in I(n, k)$. The space of $k$-vectors is denoted by $\wedge_k(\mathbb{R}^n)$; in particular, $\wedge_1(\mathbb{R}^n) = \mathbb{R}^n$. For reasons of formal convenience, we set $\wedge_0(\mathbb{R}^n) := \mathbb{R}$ and $\wedge_k(\mathbb{R}^n) := \{0\}$ for $k > n$.

We denote by $|\cdot|$ the Euclidean norm on $\wedge_k(\mathbb{R}^n)$.

The exterior product $v \wedge w \in \wedge_{k+h}(\mathbb{R}^n)$ is defined for every $v \in \wedge_k(\mathbb{R}^n)$ and $w \in \wedge_h(\mathbb{R}^n)$, and is completely determined by the following properties: (1) associativity, (2) linearity in both arguments, and (3) $e_i \wedge e_j = -e_j \wedge e_i$ for every $i \neq j$ and $e_i \wedge e_i = 0$ for every $i$.

### Simple Vectors and Orientation

A simple $k$-vector is any $v$ in $\wedge_k(\mathbb{R}^n)$ that can be written as a product of 1-vectors, that is,

$$v = v_1 \wedge v_2 \wedge \cdots \wedge v_k$$

It can be shown that $v$ is null if and only if the vectors $\{v_i\}$ are linearly dependent. If $v$ is not null,

then it is uniquely determined by the following objects: (1) the $k$-dimensional space $M$ spanned by $\{v_i\}$; (2) the orientation of $M$ associated with the basis $\{v_i\}$; (3) the euclidean norm $|v|$. In particular, $M$ does not depend on the choice of the vectors $v_i$. Note that $|v|$ is equal to the $k$-dimensional volume of the parallelogram spanned by $\{v_i\}$.

Hence, the map $v \mapsto M$ is a one-to-one correspondence between the class of simple $k$-vectors with norm $|v| = 1$ and the Grassmann manifold of oriented k-dimensional subspaces of $\mathbb{R}^n$.

This remark paves the way to the following definition: if $S$ is a $k$-dimensional surface of class $C^1$ in $\mathbb{R}^n$, possibly with boundary, an orientation of $S$ is a continuous map $\tau_S : S \to \wedge_k(\mathbb{R}^n)$ such that $\tau_S(x)$ is a simple $k$-vector with norm 1 that spans $\mathrm{Tan}(S, x)$ for every $x$. With every orientation of $S$ (if any exists) is canonically associated the orientation of the boundary $\partial S$ that satisfies

$$\tau_S(x) = \eta(x) \wedge \tau_{\partial S}(x) \quad \text{for every } x \in \partial S \qquad [9]$$

where $\eta(x)$ is the inner normal to $\partial S$ at $x$.

### *k*-Covectors

The standard basis of the dual of $\mathbb{R}^n$ is $\{dx_1, \ldots, dx_n\}$, where $dx_i : \mathbb{R}^n \to \mathbb{R}$ is the linear functional that takes every $x = (x_1, \ldots, x_n)$ into the $i$th component $x_i$. For every $\boldsymbol{i} \in I(n, k)$ we set

$$dx_{\boldsymbol{i}} = dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_k}$$

and the space $\wedge^k(\mathbb{R}^n)$ of $k$-covectors consists of all formal linear combinations $\Sigma \alpha_{\boldsymbol{i}} dx_{\boldsymbol{i}}$. The exterior product of covectors is defined as that for vectors. The space $\wedge^k(\mathbb{R}^n)$ is dual to $\wedge_k(\mathbb{R}^n)$ via the duality pairing $\langle\, ;\, \rangle$ defined by the relations $\langle dx_{\boldsymbol{i}}; e_{\boldsymbol{j}} \rangle := \delta_{\boldsymbol{ij}}$ (that is, 1 if $\boldsymbol{i} = \boldsymbol{j}$ and 0 otherwise).

### Differential Forms and Stokes Theorem

A differential form of order $k$ on $\mathbb{R}^n$ is a map $\omega : \mathbb{R}^n \to \wedge^k(\mathbb{R}^n)$. Using the canonical basis of $\wedge^k(\mathbb{R}^n)$, we can write $\omega$ as

$$\omega(x) = \sum_{\boldsymbol{i} \in I(n,k)} \omega_{\boldsymbol{i}}(x) dx_{\boldsymbol{i}}$$

where the coordinates $\omega_{\boldsymbol{i}}$ are real functions on $\mathbb{R}^n$. The exterior derivative of a $k$-form $\omega$ of class $C^1$ is the $(k+1)$-form

$$d\omega(x) := \sum_{\boldsymbol{i} \in I(n,k)} d\omega_{\boldsymbol{i}}(x) \wedge dx_{\boldsymbol{i}}$$

where, for every scalar function $f$, $df$ is the 1-form

$$df(x) := \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x) dx_i$$

If $S$ is a $k$-dimensional oriented surface, the integral of a $k$-form $\omega$ on $S$ is naturally defined by

$$\int_S \omega := \int_S \langle \omega(x); \tau_S(x) \rangle \, \mathrm{d}\mathscr{H}^k(x)$$

Stokes theorem states that for every $(k-1)$-form $\omega$ of class $C^1$ there holds

$$\int_{\partial S} \omega = \int_S \mathrm{d}\omega \qquad [10]$$

provided that $\partial S$ is endowed with the orientation $\tau_{\partial S}$ that satisfies [9].

## Currents

The definition of $k$-dimensional currents closely resembles that of distributions: they are the dual of smooth $k$-forms with compact support. Since every oriented $k$-dimensional surface defines by integration a linear functional on forms, currents can be regarded as generalized oriented surfaces. As every distribution admits a derivative, so every current admits a boundary. Indeed, many other basic notions of homology theory can be naturally extended to currents – this was actually one of the motivations behind the introduction of currents, due to de Rham.

For the applications to variational problems, smaller classes of currents are usually considered; the most relevant to the Plateau problem is that of integral currents. Note that the definitions of the spaces of normal, rectifiable, and integral currents and the symbols used to denote them vary, sometimes more than slightly, depending on the author.

### Currents, Boundary, and Mass

Let $n, k$ be integers with $n \geq k$. The space of $k$-dimensional currents on $\mathbb{R}^n$, denoted by $\mathscr{D}_k(\mathbb{R}^n)$, is the dual of the space $\mathscr{D}^k(\mathbb{R}^n)$ of smooth $k$-forms with compact support in $\mathbb{R}^n$. For $k \geq 1$, the boundary of a $k$-current $T$ is the $(k-1)$-current $\partial T$ defined by

$$\langle \partial T; \omega \rangle := \langle T; \mathrm{d}\omega \rangle \quad \text{for every } \omega \in \mathscr{D}^{k-1}(\mathbb{R}^n) \qquad [11]$$

while the boundary of a 0-current is set equal to 0. The mass of $T$ is the number

$$\mathbb{M}(T) := \sup\Big\{ \langle T; \omega \rangle : \ \omega \in \mathscr{D}^k(\mathbb{R}^n), |\omega| \leq 1 \Big\} \qquad [12]$$

Fundamental examples of $k$-currents are oriented $k$-dimensional surfaces: with each oriented surface $S$ of class $C^1$ is canonically associated the current $\langle T; \mathrm{d}\omega \rangle := \int_S \omega$ (in fact, $S$ is completely determined by the action on forms, i.e., by the associated

current). By Stokes theorem, the boundary of $T$ is the current associated with the boundary of $S$; thus, the notion of boundary for currents is compatible with the classical one for oriented surfaces. A simple computation shows that $\mathbb{M}(T) = \mathscr{H}^k(S)$; therefore, the mass provides a natural extension of the notion of $k$-dimensional volume to $k$-currents.

### Remarks

(i) Not all $k$-currents look like $k$-dimensional surfaces. For example, every $k$-vectorfield $v : \mathbb{R}^n \to \wedge_k(\mathbb{R}^n)$ defines by duality the $k$-current

$$\langle T;\omega\rangle := \int \langle \omega(x); v(x)\rangle \,\mathrm{d}\mathscr{H}^n(x)$$

The mass of $T$ is $\int |v| \mathrm{d}\mathscr{H}^n$, and the boundary is represented by a similar integral formula involving the partial derivatives of $v$ (in particular, for 1-vectorfields, the boundary is the 0-current associated with the divergence of $v$). Note that the dimension of such $T$ is $k$ because $k$-vectorfields act on $k$-forms, and there is no relation with the dimension of the support of $T$, which is $n$.

(ii) To be precise, $\mathscr{D}^k(\mathbb{R}^n)$ is a locally convex topological vector space, and $\mathscr{D}_k(\mathbb{R}^n)$ is its topological dual. As such, $\mathscr{D}_k(\mathbb{R}^n)$ is endowed with a dual (or weak$^*$) topology. We say that a sequence of $k$-currents $(T_j)$ converge to $T$ if they converge in the dual topology, that is,

$$\langle T_j;\omega\rangle \to \langle T;\omega\rangle \quad \text{for every } \omega \in \mathscr{D}^k(\mathbb{R}^n) \qquad [13]$$

Recalling the definition of mass, it is easy to show that it is lower-semicontinuous with respect the dual topology, and in particular

$$\liminf \mathbb{M}(T_j) \geq \mathbb{M}(T) \qquad [14]$$

### Currents with Finite Mass

By definition, a $k$-current $T$ with finite mass is a linear functional on $k$-forms which is bounded with respect to the supremum norm, and by Riesz theorem it can be represented as a bounded measure with values in $\wedge_k(\mathbb{R}^n)$. In other words, there exist a finite positive measure $\mu$ on $\mathbb{R}^n$ and a density function $\tau : \mathbb{R}^n \to \wedge_k(\mathbb{R}^n)$ such that $|\tau(x)| = 1$ for every $x$ and

$$\langle T;\omega\rangle = \int \langle \omega(x); \tau(x)\rangle \,\mathrm{d}\mu(x)$$

The fact that currents are the dual of a separable space yields the following compactness result: a sequence of $k$-currents $(T_j)$ with uniformly bounded masses $\mathbb{M}(T_j)$ admits a subsequence that converges to a current with finite mass.

### Normal Currents

A $k$-current $T$ is called normal if both $T$ and $\partial T$ have finite mass. The compactness result stated in the previous paragraph implies the following compactness theorem for normal currents: a sequence of normal currents $(T_j)$ with $\mathbb{M}(T_j)$ and $\mathbb{M}(\partial T_j)$ uniformly bounded admits a subsequence that converges to a normal current.

### Rectifiable Currents

A $k$-current $T$ is called rectifiable if it can represented as

$$\langle T;\omega\rangle = \int_E \langle \omega(x); \tau(x)\rangle \theta(x) \,\mathrm{d}\mathscr{H}^k(x)$$

where $E$ is a $k$-rectifiable set $E$, $\tau$ is an orientation of $E$ – that is, $\tau(x)$ is a simple unit $k$-vector that spans $\mathrm{Tan}(E,x)$ for $\mathscr{H}^k$-almost every $x \in E$ – and $\theta$ is a real function such that $\int_E |\theta| \mathrm{d}\mathscr{H}^k$ is finite, called multiplicity. Such $T$ is denoted by $T = [E, \tau, \theta]$. In particular, a rectifiable 0-current can be written as $\langle T;\omega\rangle = \Sigma \theta_i \omega(x_i)$, where $E = \{x_i\}$ is a countable set in $\mathbb{R}^n$ and $\{\theta_i\}$ is a sequence of real numbers with $\Sigma|\theta_i| < +\infty$.

### Integral Currents

If $T$ is a rectifiable current and the multiplicity $\theta$ takes integral values, $T$ is called an integer multiplicity rectifiable current. If both $T$ and $\partial T$ are integer multiplicity rectifiable currents, then $T$ is an integral current.

The first nontrivial result is the boundary rectifiability theorem: if $T$ is an integer multiplicity rectifiable current and $\partial T$ has finite mass, then $\partial T$ is an integer multiplicity rectifiable current, too, and therefore $T$ is an integral current.

The second fundamental result is the compactness theorem for integral currents: a sequence of integral currents $(T_j)$ with $\mathbb{M}(T_j)$ and $\mathbb{M}(\partial T_j)$ uniformly bounded admits a subsequence that converges to an integral current.

### Remarks

(i) The point of the compactness theorem for integral currents is not the existence of a converging subsequence – that being already established by the compactness theorem for normal currents – but the fact that the limit is an integral current. In fact, this result is often referred to as a "closure theorem" rather than a "compactness theorem."

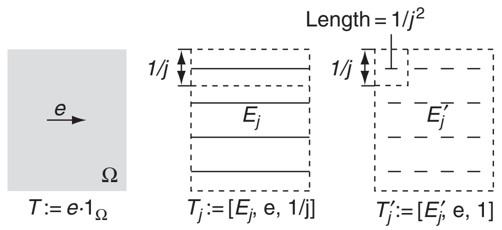



**Figure 5** $T$ is the normal 1-current on $\mathbb{R}^2$ associated with the vectorfield equal to the unit vector $e$ on the unit square $\Omega$, and equal to 0 outside. $T_j$ are the rectifiable currents associated with the sets $E_j$ (middle) and the constant multiplicity $1/j$, and then $\mathbb{M}(T_j) = 1, \mathbb{M}(\partial T_j) = 2$. $T'_j$ are the integral currents associated with the sets $E'_j$ (left) and the constant multiplicity 1, and then $\mathbb{M}(T'_j) = 1, \mathbb{M}(\partial T'_j) = 2j^2$. Both $(T_j)$ and $(T'_j)$ converge to $T$.

(ii) The following observations may clarify the role of assumptions in the compactness theorem: (1) a sequence of integral currents $(T_j)$ with $\mathbb{M}(T_j)$ uniformly bounded – but not $\mathbb{M}(\partial T_j)$ – may converge to any current with finite mass, not necessarily a rectifiable one.
(2) A sequence of rectifiable currents $(T_j)$ with rectifiable boundaries and $\mathbb{M}(T_j), \mathbb{M}(\partial T_j)$ uniformly bounded may converge to any normal current, not necessarily a rectifiable one. Examples of both situations are described in Figure 5.

### Application to the Plateau Problem

The compactness result for integral currents implies the existence of currents with minimal mass: if $\Gamma$ is the boundary of an integral $k$-current in $\mathbb{R}^n, 1 \le k \le n$, then there exists a current $T$ of minimal mass among those that satisfy $\partial T = \Gamma$.

The proof of this existence result is a typical example of the direct method: let $m$ be the infimum of $\mathbb{M}(T)$ among all integral currents with boundary $\Gamma$, and let $(T_j)$ be a minimizing sequence (i.e., a sequence of integral currents with boundary $\Gamma$ such that $\mathbb{M}(T_j)$ converges to $m$). Since $\mathbb{M}(T_j)$ is bounded and $\mathbb{M}(\partial T_j) = \mathbb{M}(\Gamma)$ is constant, we can apply the compactness theorem for integral currents and extract a subsequence of $(T_j)$ that converges to an integral current $T$. By the continuity of the boundary operator, $\partial T = \lim \partial T_j = \Gamma$, and by the semicontinuity of the mass $\mathbb{M}(T) \le \lim \mathbb{M}(T_j) = m$ (cf. [14]). Thus, $T$ is the desired minimal current.

### Remarks

(i) Every integral $(k-1)$-current $\Gamma$ with null boundary and compact support in $\mathbb{R}^n$ is the boundary of an integral current, and therefore is an admissible datum for the previous existence result.

(ii) A mass-minimizing integral current $T$ is more regular than a general integral current. For $k = n-1$, there exists a closed singular set $S$ with $\dim_{\mathrm{H}}(S) \le k-7$ such that $T$ agrees with a smooth surface in the complement of $S$ and of the support of the boundary. In particular, $T$ is smooth away from the boundary for $n \le 7$. For general $k$, it can only be proved that $\dim_{\mathrm{H}}(S) \le k-2$ Both results are optimal: in $\mathbb{R}^4 \times \mathbb{R}^4$, the minimal 7-current with boundary $\Gamma := \{|x| = |y| = 1\}$ – a product of two 3-spheres – is the cone $T := \{|x| = |y| \le 1\}$, and is singular at the origin. In $\mathbb{R}^2 \times \mathbb{R}^2$, the minimal 2-current with boundary $\Gamma := \{x = 0, |y| = 1\} \cup \{y = 0, |x| = 1\}$ – a union of two disjoint circles – is the union of the disks $\{x = 0, |y| \le 1\} \cup \{y = 0, |x| \le 1\}$, and is singular at the origin.

(iii) In certain cases, the mass-minimizing current $T$ may not agree with the solution of the Plateau problem suggested by intuition. The first reason is that currents do not include nonorientable surfaces, which sometimes may be more convenient (Figure 6). Another reason is that the mass of an integral current $T$ associated with a $k$-rectifiable set $E$ does not agree with the measure $\mathscr{H}^k(E)$ – called size of $T$ – because multiplicity must be taken into account, and for certain $\Gamma$ the mass-minimizing current may be not size-minimizing (Figure 7). Unfortunately, proving the existence of size-minimizing currents is much more complicated, due to lack of suitable compactness theorems.

(iv) For $k = 2$, the classical approach to the Plateau problem consists in parametrizing surfaces in $\mathbb{R}^n$ by maps $f$ from a given two-dimensional domain $D$ into $\mathbb{R}^n$, and looking for minimizers of the area functional

$$\int_D \sqrt{\det(\nabla f^* \nabla f)}$$

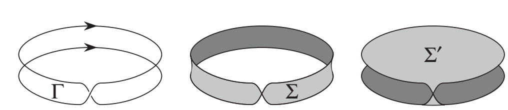



**Figure 6** The surface with minimal area spanning the (oriented) curve $\Gamma$ is the Möbius strip $\Sigma$. However, $\Sigma$ is not orientable, and cannot be viewed as a current. The mass-minimizing current with boundary $\Gamma$ is $\Sigma'$.

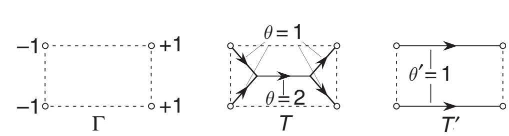



**Figure 7** The boundary $\Gamma$ is a 0-current associated with four oriented points. The size (length) of $T$ is smaller than that of $T'$. However, $\partial T = \Gamma$ implies that the multiplicity of $T$ must be 2 on the central segment and 1 on the others; thus the mass of $T$ is larger than its size. The size-minimizing current with boundary $\Gamma$ is $T$, while the mass-minimizing one is $T'$.

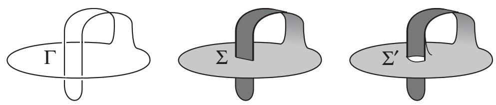



**Figure 8** The surface $\Sigma$ minimizes the area among surfaces parametrized by the disk with boundary $\Gamma$. The mass-minimizing current $\Sigma'$ can only be parametrized by a disk with a handle. Note that $\Sigma$ is a singular surface, while $\Sigma'$ is not.

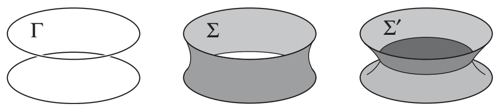



**Figure 9** Two possible soap films spanning the wire $\Gamma$: unlike $\Sigma$, $\Sigma'$ cannot be viewed as a current with multiplicity 1 and boundary $\Gamma$.

(recall the area formula, discussed earlier) under the constraint $f(\partial D) = \Gamma$. In this framework, the choice of the domain $D$ prescribes the topological type of admissible surfaces, and therefore the minimizer may differ substantially from the mass-minimizing current with boundary $\Gamma$ (Figure 8).

(v) For some modeling problems, for instance, those related to soap films and soap bubbles, currents do not provide the right framework (Figure 9). A possible alternative are integral varifolds (cf. Almgren 2001). However, it should be pointed out that this framework does not allow for "easy" application of the direct method, and the existence of minimal varifolds is in general quite difficult to prove.

### Miscellaneous Results and Useful Tools

(i) An important issue, related to the use of currents for solving variational problems, concerns the extent to which integral currents can be approximated by regular objects. For many reasons, the "right" regular class to consider are not smooth surfaces, but integral polyhedral currents, that is, linear combinations with integral coefficients of oriented simplexes. The following approximation theorem holds: for every integral current $T$ in $\mathbb{R}^n$ there exists a sequence of integral polyhedral currents $(T_j)$ such that

$$T_j \to T, \ \partial T_j \to \partial T$$
$$\mathbb{M}(T_j) \to \mathbb{M}(T), \mathbb{M}(\partial T_j) \to \mathbb{M}(\partial T)$$

The proof is based on a quite useful tool, called polyhedral deformation.

(ii) Many geometric operations for surfaces have an equivalent for currents. For instance, it is possible to define the image of a current in $\mathbb{R}^n$ via a smooth proper map $f: \mathbb{R}^n \to \mathbb{R}^m$. Indeed, with every $k$-form $\omega$ on $\mathbb{R}^m$ is canonically associated a $k$-form $f^{\#}\omega$ on $\mathbb{R}^n$, called pullback of $\omega$ according to $f$. The adjoint of the pullback is an operator, called push-forward, that takes every $k$-current $T$ in $\mathbb{R}^n$ into a $k$-current $f_{\#}T$ in $\mathbb{R}^m$. If $T$ is the rectifiable current associated with a rectifiable set $E$ and a multiplicity $\theta$, the push-forward $f_{\#}T$ is the rectifiable current associated with $f(E)$ – and a multiplicity $\theta'(y)$ which is computed by adding up with the right sign all $\theta(x)$ with $x \in f^{-1}(y)$. As one might expect, the boundary of the push-forward is the push-forward of the boundary.

(iii) In general, it is not possible to give a meaning to the intersection of two currents, and not even of a current and a smooth surface. However, it is possible to define the intersection of a normal $k$-current $T$ and a level surface $f^{-1}(y)$ of a smooth map $f: \mathbb{R}^n \to \mathbb{R}^h$ (with $k \le h \le n$) for almost every $y$, resulting in a current $T_y$ with the expected dimension $h - k$. This operation is called slicing.

(iv) When working with currents, a quite useful notion is that of flat norm:

$$\mathbb{F}(T) := \inf\{\mathbb{M}(R) + \mathbb{M}(S): \ T = R + \partial S\}$$

where $T$ and $R$ are $k$-currents, and $S$ is a $(k+1)$-current. The relevance of this notion lies in the fact that a sequence $(T_j)$ that converges with respect to the flat norm converges also in the dual topology, and the converse holds if the masses $\mathbb{M}(T_j)$ and $\mathbb{M}(\partial T_j)$ are uniformly bounded. Hence, the flat norm metrizes the dual topology of currents (at least on sets of currents where the mass and the mass of the boundary are bounded).

Since $\mathbb{F}(T)$ can be explicitly estimated from above, it can be quite useful in proving that a sequence of currents converges to a certain limit. Finally, the flat norm gives a (geometrically significant) measure of how far apart two currents are: for instance, given the 0-currents $\delta_x$ and $\delta_y$ (the Dirac masses at $x$ and $y$, respectively), then $\mathbb{F}(\delta_x - \delta_y)$ is exactly the distance between $x$ and $y$.

*See also:* Free Interfaces and Free Discontinuities: Variational Problems; Γ-Convergence and Homogenization; Geometric Phases; Image Processing: Mathematics; Minimal Submanifolds; Mirror Symmetry: A Geometric Survey; Moduli Spaces: An Introduction.

## Further Reading

Almgren FJ Jr. (2001) *Plateau's Problem: An Invitation to Varifold Geometry*, Revised Edition, Student Mathematical Library, vol. 13. Providence: American Mathematical Society.

Falconer KJ (2003) *Fractal Geometry. Mathematical Foundations and Applications*, 2nd edn. Hoboken, NJ: Wiley.

Federer H (1969) *Geometric Measure Theory,* Grundlehren der mathematischen Wissenschaften, vol. 153. Berlin: Springer. (Reprinted in the series Classics in Mathematics. Springer, Berlin, 1996).

Federer H and Fleming WH (1960) Normal and integral currents. *Annals of Mathematics* 72: 458–520.
Mattila P (1995) *Geometry of Sets and Measures in Euclidean Spaces, Fractals and Rectifiability*, Cambridge Studies in Advanced Mathematics, vol. 44. Cambridge: Cambridge University Press.
Morgan F (2000) *Geometric Measure Theory. A Beginner's Guide*, 3rd edn. San Diego: Academic Press.
Simon L (1983) *Lectures on Geometric Measure Theory*. Proceedings of the Centre for Mathematical Analysis, 3. Australian National University, Centre for Mathematical Analysis, Canberra 1983.

# Geometric Phases

**P Lévay**, Budapest University of Technology and Economics, Budapest, Hungary



## Introduction

We invite the reader to perform the following simple experiment. Put your arm out in front of you keeping your thumb pointing up perpendicular to your arm. Move your arm up over your head, then bring it down to your side, and at last bring the arm back in front of you again. In this experiment an object (your thumb) was taken along a closed path traced by another object (your arm) in a way that a simple local law of transport was applied. In this case the local law consisted of two ingredients: (1) preserve the orthogonality of your thumb with respect to your arm and (2) do not rotate the thumb about its instantaneous axis (i.e., your arm). Performing the experiment in this way, you will manage to avoid rotations of your thumb locally; however, in the end you will experience a rotation of $90^\circ$ globally.

The experiment above can be regarded as the archetypical example of the phenomenon called anholonomy by physicists and holonomy by mathematicians. In this article, we consider the manifestation of this phenomenon in the realm of quantum theory. The objects to be transported along closed paths in suitable manifolds will be wave functions representing quantum systems. After applying local laws dictated by inputs coming from physics, one ends up with a new wave function that has picked up a complex phase factor. Phases of this kind are called geometric phases, with the famous Berry phase being a special case.

## The Space of Rays

Let us consider a quantum system with physical states represented by elements $|\psi\rangle$ of some Hilbert space $\mathcal{H}$ with scalar product $\langle|\rangle:\mathcal{H}\times\mathcal{H}\to \mathbf{C}$. For simplicity, we assume that $\mathcal{H}$ is finite dimensional, $\mathcal{H}\simeq \mathbf{C}^{n+1}$ with $n\geq 1$. The infinite-dimensional case can be studied by taking the inductive limit $n\to\infty$. Let us denote the complex amplitudes characterizing the state $|\psi\rangle$ by $Z^\alpha, \alpha=0,1,\ldots,n$. For a normalized state,

$$\|\psi\|^2=\langle\psi|\psi\rangle\equiv\delta_{\alpha\beta}\bar{Z}^\alpha Z^\beta\equiv\bar{Z}_\alpha Z^\alpha=1 \qquad [1]$$

where summation over repeated indices is understood, indices raised and lowered by $\delta^{\alpha\beta}$ and $\delta_{\alpha\beta}$, respectively, and the overbar refers to complex conjugation. A normalized state lies on the unit sphere $\mathcal{S}\simeq S^{2n+1}$ in $\mathbf{C}^{n+1}$. Two nonzero states $|\psi\rangle$ and $|\varphi\rangle$ are equivalent, $|\psi\rangle\sim|\varphi\rangle$, iff they are related as $|\psi\rangle=\lambda|\varphi\rangle$ for some nonzero complex number $\lambda$. For equivalent states, physically meaningful quantities such as

$$\frac{\langle\psi|\boldsymbol{A}|\psi\rangle}{\langle\psi|\psi\rangle},\qquad \frac{|\langle\psi|\varphi\rangle|^2}{\|\psi\|^2\|\varphi\|^2} \qquad [2]$$

(mean value of a physical quantity represented by a Hermitian operator $A$, transition probability from a physical state represented by $|\psi\rangle$ to one represented by $|\varphi\rangle$) are invariant. Hence, the real space of states representing the physical states of a quantum system unambiguously is the set of equivalence classes $\mathcal{P}\equiv\mathcal{H}/\sim$. $\mathcal{P}$ is called the "space of rays." For $\mathcal{H}\simeq\mathbf{C}^{n+1}$, we have $\mathcal{P}\simeq \boldsymbol{CP}^n$, where $\boldsymbol{CP}^n$ is the $n$-dimensional complex projective space. For normalized states, $|\psi\rangle$ and $|\varphi\rangle$ are equivalent iff $|\psi\rangle=\lambda|\varphi\rangle$, where $|\lambda|=1$, that is, $\lambda\in \mathrm{U}(1)$. Thus, two normalized states are equivalent iff they differ merely in a complex phase. It is well known that $\mathcal{S}$ can be regarded as the total space of a principal bundle over $\mathcal{P}$ with structure group U(1). This means that we have the projection

$$\pi:|\psi\rangle\in\mathcal{S}\subset\mathcal{H}\to|\psi\rangle\langle\psi|\in\mathcal{P} \qquad [3]$$

where the rank-1 projector $|\psi\rangle\langle\psi|$ represents the equivalence class of $|\psi\rangle$. Since we will use this bundle frequently in this article, we call it $\eta_1$ (the meaning of the subscript 1 will be clarified later). Then, we have

$$\eta_1:\mathrm{U}(1)\hookrightarrow\mathcal{S}\xrightarrow{\pi}\mathcal{P} \qquad [4]$$

For $Z^0\neq 0$ the space of rays $\mathcal{P}$ can be given local coordinates

$$w^j\equiv Z^j/Z^0,\quad j=1,\ldots,n \qquad [5]$$

The $w^j$ are inhomogeneous coordinates for $\boldsymbol{CP}^n$ on the coordinate patch $\mathcal{U}_0$ defined by the condition $Z^0 \neq 0$.

$\mathcal{P}$ is a compact complex manifold with a natural Riemannian metric $g$. This metric $g$ is induced from the scalar product on $\mathcal{H}$. Let us consider the construction of $g$ by using the physical input provided by the invariance of the transition probability of [1]. For this we define a distance between $|\psi\rangle\langle\psi|$ and $|\varphi\rangle\langle\varphi|$ in $\mathcal{P}$ as follows:

$$\cos^2(\delta(\psi,\varphi)/2) \equiv \frac{|\langle\psi|\varphi\rangle|^2}{\|\psi\|^2\|\varphi\|^2} \qquad [6]$$

This definition makes sense since, due to the Cauchy–Schwartz inequality, the right-hand side of [6] is non-negative and $\leq 1$. It is equal to 1 iff $|\psi\rangle$ is a nonzero complex multiple of $|\varphi\rangle$, that is, iff they define the same point in $\mathcal{P}$. Hence in this case, $\delta(\psi,\varphi) = 0$ as expected.

Suppose now that $|\psi\rangle$ and $|\varphi\rangle$ are separated by an infinitesimal distance $\mathrm{d}s \equiv \delta(\psi,\varphi)$. Putting this into the definition [6], using the local coordinates $w^j$ of [5] for $|\psi\rangle$ and $w^j + \mathrm{d}w^j$ for $|\varphi\rangle$ after expanding both sides using Taylor series, one gets

$$\mathrm{d}s^2 = 4g_{j\bar{k}}\,\mathrm{d}w^j\,\mathrm{d}w^{\bar{k}}, \quad j,\bar{k} = 1,2,\ldots,n \qquad [7]$$

where

$$g_{j\bar{k}} \equiv \frac{(1+\bar{w}_l w^l)\delta_{jk} - \bar{w}_j w_k}{(1+\bar{w}_m w^m)^2} \qquad [8]$$

with $\mathrm{d}w^{\bar{k}} \equiv \mathrm{d}\bar{w}^k$. The line element [7] defines the Fubini–Study metric for $\mathcal{P}$.

## The Pancharatnam Connection

Having defined the basic entity, the space of rays $\mathcal{P}$, and the principal U(1) bundle $\eta_1$, now we define a connection giving rise to a local law of parallel transport. This approach gives rise to a very general definition of the geometric phase. In the mathematical literature, the connection defined below is called the "canonical connection" on the principal bundle. However, since the motivation is coming from physics, we are going to rediscover this construction using merely physical information provided by quantum theory alone.

The information needed is an adaptation of Pancharatnam's study of polarized light to quantum mechanics. Let us consider two normalized states $|\psi\rangle$ and $|\varphi\rangle$. When these states belong to the same ray, then we have $|\psi\rangle = \mathrm{e}^{\mathrm{i}\phi}|\varphi\rangle$ for some phase factor $\mathrm{e}^{\mathrm{i}\phi}$; hence, the phase difference between them can be defined to be just $\phi$. How to define the phase difference between $|\psi\rangle$ and $|\varphi\rangle$ (not orthogonal) when these states belong to different rays? To compare the phases of nonorthogonal states belonging to different rays, Pancharatnam employed the following simple rule: two states are "in phase" iff their interference is maximal. In order to find the state $|\varphi\rangle \equiv \mathrm{e}^{\mathrm{i}\phi}|\varphi'\rangle$ from the ray spanned by the representative $|\varphi'\rangle$ which is "in phase" with $|\psi\rangle$, we have to find a $\phi$ modulo $2\pi$ for which the interference term in

$$\|\psi + \mathrm{e}^{\mathrm{i}\phi}\varphi'\|^2 = 2(1 + \mathrm{Re}(\mathrm{e}^{\mathrm{i}\phi}\langle\psi|\varphi'\rangle)) \qquad [9]$$

is maximal. Obviously the interference is maximal iff $\mathrm{e}^{\mathrm{i}\phi}\langle\psi|\varphi'\rangle$ is a real positive number, that is,

$$\mathrm{e}^{\mathrm{i}\phi} = \frac{\langle\varphi'|\psi\rangle}{|\langle\varphi'|\psi\rangle|}, \quad |\varphi\rangle = |\varphi'\rangle\frac{\langle\varphi'|\psi\rangle}{|\langle\varphi'|\psi\rangle|} \qquad [10]$$

Hence for the state $|\varphi\rangle$ "in phase" with $|\psi\rangle$, one has

$$\langle\psi|\varphi\rangle = |\langle\psi|\varphi'\rangle| \in \boldsymbol{R}^+ \qquad [11]$$

When such $|\psi\rangle$ and $|\varphi\rangle \equiv |\psi + \mathrm{d}\psi\rangle$ are infinitesimally separated, from [11] it follows that

$$\mathrm{Im}\langle\psi|\mathrm{d}\psi\rangle = \frac{1}{2\mathrm{i}}\left(\bar{Z}_\alpha\,\mathrm{d}Z^\alpha - \mathrm{d}\bar{Z}_\alpha Z^\alpha\right) = 0 \qquad [12]$$

where $\bar{Z}_\alpha Z^\alpha = \bar{Z}_0 Z^0(1+\bar{w}_j w^j) = 1$ due to normalization. Writing $Z^0 \equiv |Z^0|\mathrm{e}^{\mathrm{i}\Phi}$ using [5], one obtains

$$\mathrm{Im}\langle\psi|\mathrm{d}\psi\rangle = \mathrm{d}\Phi + A = 0, \quad A \equiv \mathrm{Im}\frac{\bar{w}_j\mathrm{d}w^j}{1+\bar{w}_k w^k} \qquad [13]$$

In order to clarify the meaning of the 1-form $A$, notice that the choice

$$|\psi'\rangle \equiv \frac{1}{\sqrt{1+\bar{w}_k w^k}}\begin{pmatrix}1\\ w^j\end{pmatrix} \qquad [14]$$

defines a local section of the bundle $\eta_1$. In terms of this section, the state $|\psi\rangle$ can be expressed as

$$|\psi\rangle = \begin{pmatrix}Z^0\\ Z^j\end{pmatrix} = |Z^0|\mathrm{e}^{\mathrm{i}\Phi}\begin{pmatrix}1\\ w^j\end{pmatrix} = \mathrm{e}^{\mathrm{i}\Phi}|\psi'\rangle \qquad [15]$$

For a path $w^j(t)$ lying entirely in $\mathcal{U}_0 \subset \mathcal{P}$, $|\psi(t)\rangle = \mathrm{e}^{\mathrm{i}\Phi(t)}|\psi'(t)\rangle$ defines a path in $\mathcal{S}$ with a $\Phi(t)$ satisfying the equation $\dot{\Phi} + A = 0$. For a closed path $C$, the equation above defines a (generically) open path $\Gamma$ projecting onto $C$ by the projection $\pi$. It must be clear by now that the process described is the one of parallel transports with respect to a connection with a connection 1-form $\omega$. The pull-back of $\omega$ with respect to the local section in [14] is the 1-form (U(1) gauge field) $A$ in [13]. The curve $\Gamma$ corresponding to $|\psi(t)\rangle$ is the horizontal lift of $C$ in $\mathcal{P}$. The U(1) phase

$$\mathrm{e}^{\mathrm{i}\Phi[C]} \equiv \mathrm{e}^{-\mathrm{i}\oint_C A} \qquad [16]$$

is the holonomy of the connection. We call this connection the "Pancharatnam connection," and its holonomy for a closed path in the space of rays is the geometric phase acquired by the wave function. Now the question of fundamental importance is: how to realize closed paths in $\mathcal{P}$ physically? This question is addressed in the following sections.

## Quantum Jumps

We have seen that physical states of a quantum system are represented by the space of rays $\mathcal{P}$ and normalized states used as representatives for such states form the total space $\mathcal{S}$ of a principal U(1) bundle $\eta_1$ over $\mathcal{P}$. Moreover, in the previous section we have realized that the physical notions of transition probability, and quantum interference naturally lead to the introduction of a Riemannian metric $g$ and an abelian U(1) gauge field $A$ living on $\mathcal{P}$.

An interesting result based on the connection between $g$ and $A$ concerns a nice geometric description of a special type of quantum evolution consisting of a sequence of "quantum jumps."

Consider two nonorthogonal rays $|A\rangle\langle A|$ and $|B\rangle\langle B|$ in $\mathcal{P}$. Let us suppose that the system's normalized wave function initially is $|A\rangle \in \mathcal{S}$, and measure by the "polarizer" $|B\rangle\langle B|$. Then the result of this filtering measurement is $|B\rangle\langle B|A\rangle$, or after projecting back to the set of normalized states we have the "quantum jump"

$$|A\rangle \to |B\rangle \frac{\langle B|A\rangle}{|\langle B|A\rangle|} \qquad [17]$$

Now we have the following theorem:

**Theorem** *The* [17] *jump can be recovered by parallel transporting the normalized state* $|A\rangle$ *according to the Pancharatnam connection along the shortest geodesic* (*with respect to the* [8] *metric*), *connecting* $|A\rangle\langle A|$ *and* $|B\rangle\langle B|$ *in* $\mathcal{P}$.

Let us now consider a cyclic series of filtering measurements with projectors $|A_a\rangle\langle A_a|$, $a = 1, 2, \ldots, N+1$, where $|A_1\rangle\langle A_1| = |A_{N+1}\rangle\langle A_{N+1}|$. Prepare the system in the state $|A_1\rangle \in \mathcal{S}$, and then subject it to the sequence of filtering measurements. Then according to the theorem, the phase

$$\mathrm{e}^{\mathrm{i}\Phi} = \frac{\langle A_1|A_N\rangle\langle A_N|A_{N-1}\rangle\cdots\langle A_2|A_1\rangle}{|\langle A_1|A_N\rangle\langle A_N|A_{N-1}\rangle\cdots\langle A_2|A_1\rangle|} \qquad [18]$$

picked up by the state is equal to the one obtained by parallel transporting $|A_1\rangle$ along a geodesic polygon consisting of the shorter arcs connecting the projectors $|A_a\rangle\langle A_a|$ and $|A_{a+1}\rangle\langle A_{a+1}|$ with $a = 1, 2, \ldots, N$. It is important to realize that this filtering measurement process is not a unitary one; hence, unitarity is not essential for the geometric phase to appear.

In this section we have managed to obtain closed paths in the form of geodesic polygons in $\mathcal{P}$ via the physical process of subjecting the initial state $|A_1\rangle$ to a sequence of filtering measurements. It is clear that for any type of evolution, the geodesics of the Fubini-study metric play a fundamental role since any smooth closed curve in $\mathcal{P}$ can be approximated by geodesic polygons.

Nonunitary evolution provided by the quantum measuring process is only half of the story. In the next section, we start describing closed paths in $\mathcal{P}$ arising also from unitary evolutions generated by parameter-dependent Hamiltonians, the original context where geometric phases were discovered.

## Unitary Evolutions

### Adiabatic Evolution

Suppose that the evolution of our quantum system with $\mathcal{H} \simeq \mathbf{C}^{n+1}$ is generated by a Hermitian Hamiltonian matrix depending on a set of external parameters $x^\mu, \mu = 1, 2, \ldots, M$. Here we assume that the $x^\mu$ are local coordinates on some coordinate patch $\mathcal{V}$ of a smooth $M$-dimensional manifold $\mathcal{M}$. We lable the eigenvalues of $H(x)$ by the numbers $r = 0, 1, 2, \ldots, n$, and assume that the $r$th eigenvalue $E_r(x)$ is nondegenerate:

$$H(x)|r,x\rangle = E_r(x)|r,x\rangle, \quad r = 0, 1, 2, \ldots, n \qquad [19]$$

We assume that $H(x), E_r(x), |r,x\rangle$ are smooth functions of $x$. The rank-1 spectral projectors

$$P_r(x) \equiv |r,x\rangle\langle r,x|, \quad r = 0, 1, 2, \ldots, n \qquad [20]$$

for each $r$ define a map $f_r : \mathcal{M} \to \mathcal{P}$:

$$f_r : x \in \mathcal{V} \subset \mathcal{M} \mapsto P_r(x) \in \mathcal{P} \qquad [21]$$

Recall now that we have the bundle $\eta_1$ over $\mathcal{P}$, at our disposal, and we can pull back $\eta_1$ using the map $f_r$ to construct a new bundle $\xi_1^r$ over the parameter space $\mathcal{M}$. Moreover, we can define a connection on $\xi_1^r$ by pulling back the canonical (Pancharatnam) connection of $\eta_1$. The resulting bundle $\xi_1^r$ is called the Berry–Simon bundle over the parameter space $\mathcal{M}$. Explicitly,

$$\xi_1^r : \mathrm{U}(1) \hookrightarrow \xi_1^r \xrightarrow{\pi_\xi} \mathcal{M} \qquad [22]$$

The states $|r,x\rangle$ of [19] define a local section of $\xi_1^r$. Supressing the index $r$, the relationship between $\eta_1$

and $\xi_1$ can be summarized by the following diagram:

$$\begin{array}{ccc} \xi_1 & \xleftarrow{f^*} & \eta_1 \\ \pi_\xi \big\downarrow & & \pi_\eta \big\downarrow \\ \mathcal{M} & \xrightarrow{f} & \mathcal{P} \end{array} \qquad [23]$$

Here $f^*$ denotes the pullback map, and we have $\xi_1 \equiv f^*(\eta_1)$. (We have denoted the total space $\mathcal{S}$ as $\eta_1$.)

The local section of $\xi_1$ arising as the pullback of [14] an $\eta_1$ is given by

$$|r, x\rangle = \frac{1}{\sqrt{1 + \bar{w}^k(x) w_k(x)}} \begin{pmatrix} 1 \\ w^j(x) \end{pmatrix}, \quad x \in \mathcal{V} \subset \mathcal{M} \qquad [24]$$

with $j = 1, 2, \ldots, n$. The pullback of the Pancharatnam connection $\omega$ on $\eta_1$ is $f^*(\omega)$. We can further pull back $f^*(\omega)$ to $\mathcal{V} \subset \mathcal{M}$ with respect to the local section of [24] to obtain a gauge field living on the parameter space. This gauge field is called the "Berry gauge field" and the corresponding connection is the Berry connection. Thus,

$$\mathcal{A} = f^*(A) = \mathcal{A}_\mu(x) \mathrm{d}x^\mu = (A_j \partial_\mu w^j + A_{\bar{j}} \partial_\mu w^{\bar{j}}) \mathrm{d}x^\mu \qquad [25]$$

here $\partial_\mu \equiv \partial / \partial x^\mu$ and $A$ is given by [13]. When we have a closed curve $\mathcal{C}$ in $\mathcal{M}$, then $f \circ \mathcal{C}$ defines a closed curve C in $\mathcal{P}$. We already know that the holonomy for C in $\mathcal{P}$ can be written in the [16] form; hence,

$$\Phi_B = -\oint_{f \circ \mathcal{C}} A = -\oint_{\mathcal{C}} f^*(A) = -\oint_{\mathcal{C}} \mathcal{A} \qquad [26]$$

This formula states that there is a geometric phase picked up by the eigenstates of a parameter-dependent Hermitian Hamiltonian when we change the parameters along a closed curve. Our formula shows that the geometric phase can be calculated using either the canonical connection on $\eta_1$ or the Berry connection on $\xi_1$.

Let us then change the parameters $x^\mu$ adiabatically. The closed path in parameter space then defines Hamiltonians satisfying $H(x(T)) = H(x(0))$ for some $T \in \boldsymbol{R}^+$. Moreover, there is also the associated closed curve $P_r(x(T)) = P_r(x(0))$ in $\mathcal{P}$. The quantum adiabatic theorem states that if we prepare a state $|\Psi(0)\rangle \equiv |r, x(0)\rangle$ at $t = 0$, which is an eigenstate of the instantaneous Hamiltonian $H(x(0))$, then after changing the parameters infinitely slowly, the time evolution generated by the time-dependent Schrödinger equation

$$\mathrm{i}\hbar \frac{\mathrm{d}}{\mathrm{d}t} |\Psi(t)\rangle = H(t) |\Psi(t)\rangle \qquad [27]$$

takes the form

$$|\Psi(t)\rangle = |r, x(t)\rangle \mathrm{e}^{\mathrm{i}\Lambda_r(t)} \qquad [28]$$

after time $t$, which belongs to the same eigensubspace. The point is that the theorem holds only for cases when the kinetic energy associated with the slow change in the external parameters is much smaller than the energy separation between $E_r(x)$ and $E_{r'}(x)$ for all $x \in \mathcal{M}$. Under this assumption, transitions between adjacent levels are prohibited during evolution. Notice that the adiabatic theorem clearly breaks down in the vicinity of level crossings where the gap is comparable with the magnitude of the kinetic energy of the external parameters.

However, if one takes it for granted that the projector $P_r(t) \equiv P_r(x(t))$ for some $r$ satisfies the Schrödinger–von Neumann equation

$$\mathrm{i}\hbar \frac{\mathrm{d}}{\mathrm{d}t} P_r(t) = [H(t), P_r(t)] \qquad [29]$$

by virtue of [19], we get zero for the right-hand side. This means that $P_r(t)$ is constant; hence, the curve in $\mathcal{P}$ degenerates to a point. The upshot of this is that exact adiabatic cyclic evolutions do not exist. It can be shown, however, that under certain conditions one can find an initial state $|\Psi(0)\rangle \neq |r, x(0)\rangle$ that is "close enough" to $P_r(x(t)) = |r, x(t)\rangle\langle r, x(t)|$. Then, we can say that the projector analog of [28] only approximately holds

$$|\Psi(t)\rangle\langle\Psi(t)| \simeq |r, x(t)\rangle\langle r, x(t)| \qquad [30]$$

This means that the use of the bundle picture for the generation of closed curves for $\mathcal{P}$ via the adiabatic evolution can merely be used as an approximation.

## Berry's Phase

The straightforward calculation after substituting [28] into [27] shows that

$$\exp(\mathrm{i}\Lambda_r(T)) \exp\left(-\frac{\mathrm{i}}{\hbar} \int_0^T E_r(t) \mathrm{d}t\right) \exp\left(-\mathrm{i} \oint_{\mathcal{C}} \mathcal{A}^{(r)}\right) \qquad [31]$$

where $\mathcal{C}$ is a closed curve lying entirely in $\mathcal{V} \subset \mathcal{M}$. The first phase factor is the dynamical and the second is the celebrated Berry phase. Notice that the index $r$ labeling the eigensubspace in question

should now be included in the definition of $\mathcal{A}$ (see eqn [25]).

As an explicit example, let us take the Hamiltonian

$$H(\boldsymbol{X}(t)) = -\omega_0 \boldsymbol{J}\boldsymbol{X}(t), \qquad \omega_0 \equiv \frac{Bge}{2mc}, \\ \boldsymbol{X} \in \boldsymbol{R}^3, \quad |\boldsymbol{X}| = 1 \qquad [32]$$

where $e$, $m$, and $g$ are the charge, mass, and Landé factor of a particle, $c$ is the speed of light, and $B$ is the (constant) magnitude of an applied magnetic field. The three components of $\boldsymbol{J}$ are $(2J+1)\times(2J+1)$-dimensional spin matrices satisfying $\boldsymbol{J}\times\boldsymbol{J} = \mathrm{i}\hbar\boldsymbol{J}$. The Hamiltonian (eqn [32]) describes a spin $J$ particle moving in a magnetic field with slowly varying direction. It is obvious that the parameter space is a 2-sphere. Introducing polar coordinates $0 \leq \theta < \pi$, $0 \leq \chi < 2\pi$ for the patch $\mathcal{V}$ of $S^2$ excluding the south pole, we have $x^1 \equiv \theta$, $x^2 \equiv \chi$.

As an illustration, let us consider the spin 1/2 case. Then $H$ can be expressed in terms of the $2\times 2$ Pauli matrices. The eigenvalues are $E_0 = -\omega_0\hbar/2$ and $E_1 = \omega_0\hbar/2$ $(r = 0, 1)$. For the ground state, the mapping $f_0$ of [21] from $\mathcal{V} \subset \mathcal{M} \simeq S^2$ to $\mathcal{P} \simeq \boldsymbol{CP}^1$ is given by

$$w(\theta,\chi) \equiv \tan\left(\frac{\theta}{2}\right)\mathrm{e}^{\mathrm{i}\chi} \qquad [33]$$

which is stereographic projection of $S^2$ from the south pole onto the complex plane corresponding to the coordinate patch $\mathcal{U}_0 \subset \boldsymbol{CP}^1$. Using [13] and [25], one can calculate the pullback gauge field and its curvature $\mathcal{F}^{(0)} \equiv \mathrm{d}\mathcal{A}^{(0)}$, where

$$\mathcal{A}^{(0)} = \frac{1}{2}(1-\cos\theta)\mathrm{d}\chi, \qquad \mathcal{F}^{(0)} = \frac{1}{2}\sin\theta\mathrm{d}\theta\wedge\mathrm{d}\chi \quad [34]$$

Notice that $\mathcal{F}^{(0)}$ is the field strength of a magnetic monopole of strength 1/2 living on $\mathcal{M}$. Using Stokes theorem, from [26] one can calculate Berry's phase

$$\Phi^{(0)}[\mathcal{C}] = -\oint_{\mathcal{C}} \mathcal{A}^{(0)} = -\int_S \mathcal{F}^{(0)} = -\frac{1}{2}\Omega[\mathcal{C}] \qquad [35]$$

where $S$ is the surface bounded by the loop $\mathcal{C}$ and $\Omega[\mathcal{C}]$ is the solid angle subtended by the curve $\mathcal{C}$ at $\boldsymbol{X} = \boldsymbol{0}$.

The above result can be generalized for arbitrary spin $J$. Then, we have the eigenvalues $E_r = -\omega_0\hbar(J-r)$, where $0 \leq r \leq 2J$. The final result in this case is

$$\Phi^{(r)}[\mathcal{C}] = -(J-r)\Omega[\mathcal{C}], \quad 0 \leq r \leq 2J \qquad [36]$$

## The Aharonov–Anandan Phase

We have seen that the quantum adiabatic theorem can only be used approximately for generating closed curves in $\mathcal{P}$. This section, describes as to how such curves can be generated exactly.

Let us consider the Schrödinger equation with a time-dependent Hamiltonian (eqn [27]). Then we call its solution $|\Psi(t)\rangle$ cyclic if the state of the system returns, after a period $T$, to its original state. This means that the projector $|\Psi(t)\rangle\langle\Psi(t)|$ traverses a closed path $C$ in $\mathcal{P}$. In order to realize this situation, we have to find solutions of [27] for which $|\Psi(T)\rangle = \mathrm{e}^{\mathrm{i}\Delta_\Psi}|\Psi(0)\rangle$ for some $\Delta_\Psi$.

Taking for granted the existence of such a solution, let us first explore its consequences. First, we remove the dynamical phase from the cyclic solution $|\Psi(t)\rangle$

$$|\psi(t)\rangle \equiv \exp\left(\frac{\mathrm{i}}{\hbar}\int_0^t \langle\Psi(t')|H(t)|\Psi(t')\rangle \mathrm{d}t'\right)|\Psi(t)\rangle \quad [37]$$

Then, $|\psi(t)\rangle$ satisfies [12], that is, it defines a unique horizontal lift of the closed curve $C$ in $\mathcal{P}$. Following the same steps as in section describing the Pancharatnam condition, we see that the phase

$$\Phi_{AA}[C] = -\oint_C A \\ = \Delta_\Psi + \frac{1}{\hbar}\int_0^T \langle\Psi(t)|H(t)|\Psi(t)\rangle\,\mathrm{d}t \quad [38]$$

is purely geometric in origin. It is called the Aharonov–Anandan (AA) phase.

Let us now turn back to the question of finding cyclic states satisfying $|\Psi(T)\rangle = \mathrm{e}^{\mathrm{i}\Delta_\Psi}|\Psi(0)\rangle$. One possible solution is as follows. Suppose that $H$ depends on time through some not necessarily slowly changing parameters $x$. Let us find a partner Hamiltonian $h$ for our $H$ by defining a smooth mapping $\sigma: \mathcal{M} \to \mathcal{M}$, such that

$$h(x) \equiv H(\sigma(x)), \quad x \in \mathcal{V} \subset \mathcal{M} \qquad [39]$$

For the special class we study here, the cyclic vectors are eigenvectors of $h(x)$. Hence, the projectors $p_r$ and $P_r$ of $h$ and $H$ are related as $p_r(x) = P_r(\sigma(x))$; this means that we have a map $g_r: \mathcal{M} \to \mathcal{P}$,

$$g_r \equiv f_r \circ \sigma : x \in \mathcal{V} \subset \mathcal{M} \to p_r(x) \in \mathcal{P} \qquad [40]$$

which associates with every $x$ an eigenstate of $h(x)$. Moreover, $g_r$ associates with a closed curve $\mathcal{C}$ in $\mathcal{M}$ a closed curve $C$ in $\mathcal{P}$. Notice that generically $[h(x), H(x)] \neq 0$; hence, cyclic states are not eigenstates of the instanteneous Hamiltonian.

It should be clear by now that we can repeat the construction as discussed in the adiabatic case with $g_r$ replacing $f_r$. In particular, we can construct a new bundle $\zeta_1$ over the parameter space via the usual

pullback procedure. More precisely, we have the corresponding diagram

$$\begin{array}{ccc} \zeta_1 & \xleftarrow{g^*} & \eta_1 \\ \pi_\zeta \downarrow & & \pi_\eta \downarrow \\ \mathcal{M} & \xrightarrow{g} & \mathcal{P} \end{array} \qquad [41]$$

The AA connection can be obtained by pulling back the Pancharatnam connection:

$$a \equiv g^*(A) = \sigma^* \circ f^*(A) = \sigma^*(\mathcal{A}) \qquad [42]$$

where the last equality relates the AA connection with the Berry connection. Now the AA phase is

$$\Phi_{AA} = -\oint_{g\circ\mathcal{C}} A = -\oint_{\mathcal{C}} g^*(A) = -\oint_{\mathcal{C}} a \qquad [43]$$

As an example, let us take the Hamiltonian [32] with the curve $\mathcal{C}$ on $\mathcal{M} \equiv S^2$:

$$\boldsymbol{X}(t) = (\sin\theta\cos(\chi+\omega t), \sin\theta\sin(\chi+\omega t), \cos\theta) \quad [44]$$

Here $\theta$ and $\chi$ are the polar coordinates of a fixed point in $S^2$ where the motion starts. The curve $\mathcal{C}$ is a circle of fixed latitude and is traversed with an arbitrary speed. This model can be solved exactly and it can be shown that the mapping $\sigma_s$: $S^2 \to S^2$ is given by

$$\sigma : (u, \chi) \mapsto \left( \frac{u - s}{\sqrt{s^2 - 2us + 1}}, \chi \right),$$

$$u \equiv \cos\theta, \quad s \equiv \frac{\omega}{\omega_0} \qquad [45]$$

One can prove that for $0 \leq s < 1, \sigma_s$ is a diffeomorphism. In the $s \to 0$ (the adiabatic) limit, the mapping $g_{r,s} \equiv f_{r,s} \circ \sigma_s$ is continuously deformed to $f_r$. Moreover, $h(x)$ as defined above commutes with the time evolution operator; hence, cyclic states are indeed eigenstates of $h(x)$.

Using [42], [43], and [45], the explicit form of $\sigma_s$, we get for the AA phase

$$\Phi_{AA}^{(\mathrm{r,s})}[\mathcal{C}] = -2\pi(J - r)\left(1 - \frac{u - s}{\sqrt{s^2 - 2us + 1}}\right) \qquad [46]$$

In the adiabatic limit, the result goes to $-2\pi(J - r)(1 - u)$ which is just $-(J - r)$ times the solid angle of the path of fixed latitude, as it has to be.

## Generalization

In the sequence of examples, we have shown that geometric phases are related to the geometric structures on the bundle $\eta_1$. The Berry and AA phases are special cases arising from Pancharatnam's phase via a pullback procedure with respect to suitable maps defined by the physical situation in question. Hence, the Pancharatnam connection in this sense is universal. The root of this universality rests in a deep theorem of mathematics concerning the existence of universal bundles and their universal connections. In order to elaborate the insight provided by this theorem into the geometry of quantum evolution, let us first make a further generalization.

In our study of time-dependent Hamiltonians we have assumed that the eigenvalues of [19] were nondegenerate. Let us now relax this assumption. Fix an integer $N \geq 1$, the degeneracy of the eigensubspace corresponding to the eigenvalue $E_r$. One can then form a U($N$) principal bundle $\xi_N$ over $\mathcal{M}$, furnished with a connection, that is a natural generalization of the Berry connection. The pullback of this connection to a patch of $\mathcal{M}$ is a U($N$)-valued gauge field and its holonomy along a loop in $\mathcal{M}$ gives rise to a U($N$) matrix generalization of the U(1) Berry phase.

The natural description of this connection and its AA analog is as follows. Take the complex Grassmannian $Gr(n+1, N)$ of $N$ planes in $\mathbf{C}^{n+1}$. Obviously, $Gr(n+1, 1) \equiv \mathcal{P}$. Each point of $Gr(n+1, N)$ corresponds to an $N$ plane through the origin represented by a rank-$N$ projector. This projector can be written in terms of $N$ orthonormal basis vectors in an infinite number of ways. This ambiguity of choosing orthonormal frames is captured by the U($N$) gauge symmetry, the analog of the U(1) (phase) ambiguity in defining a normalized state as the representative of the rank-1 projector. This bundle of frames is the Stiefel bundle $V(n+1, N)$ alternatively denoted by $\eta_N$. $V(n+1, N)$ is a principal U($N$) bundle over $Gr(n+1, N)$ equipped with a canonical connection $\omega_N$ which is the U($N$) analog of Pancharatnam's connection.

Now according to the powerful theorem of Narasimhan and Ramanan if we have a U($N$) bundle $\xi_N$ over the $M$-dimensional parameter space $\mathcal{M}$, then there exists an integer $n_0(N, M)$ such that for $n \leq n_0$ there exists a map $f$: $\mathcal{M} \to Gr(n+1, N)$ such that $\eta_N = f^*(V(n+1, N))$. Moreover, given any two such maps $f$ and $g$, the corresponding pullback bundles are isomorphic if and only if $f$ is homotopic to $g$.

For the examples of the sections "Berry's phase" and "The Aharonov–Anandan phase," we have $N=1, n=1$, and $M=2$. Since the maps $f_r$ and $g_{r,s}$ defined by the rank-1 spectral projectors of $H(x)$ and $h(x)$ for $0 \leq s < 1$ are homotopic, the corresponding pullback bundles $\xi_1$ and $\zeta_1$ are isomorphic. Moreover, the Berry and AA connections are the pullbacks of the universal connection on $V(n+1, 1) \equiv \eta_1$ which is just Pancharatnam's connection.

For the infinite-dimensional case, one can define $Gr(\infty, N)$ by taking the union of the natural inclusion maps of $Gr(n, N)$ into $Gr(n+1, N)$.

We denote this universal classifying bundle $V(\infty, N)$ as $\eta$. Then, we see that given an $N$-dimensional eigensubspace bundle over $\mathcal{M}$ and a map $f_r : x \in \mathcal{M} \mapsto P_r(x) \in Gr(\infty, N)$ defined by the physical situation, the geometry of evolving eigensubspaces can be understood in terms of the holonomy of the pullback of the universal connection on $\eta$.

## Conclusions

In this article, we elucidate the mathematical origin of geometric phases. We have seen that the key observation is the fact that the space of rays $\mathcal{P}$ represents unambiguously the physical states of a quantum system. The particular representatives of a class in $\mathcal{P}$ belonging to the usual Hilbert space $\mathcal{H}$ form (local) sections of a U(1) bundle $\eta_1$. Based on the physical notions of transition probability and interference, $\eta_1$ can be furnished with extra structures: the metric and the connection, the latter giving rise to a natural definition of parallel transport. We have seen that the geodesics of $\mathcal{P}$ with respect to the metric play a fundamental role in approximating evolutions of any kind, giving rise to a curve in $\mathcal{P}$.

The geometric structures of $\eta_1$ induce similar structures for pullback bundles. These bundles encapsulate the geometric details of time evolutions generated by Hamiltonians that depend on a set of parameters $x$ belonging to a manifold $\mathcal{M}$. It was shown that the famous examples of Berry and AA phases arise as an important special case in this formalism. A generalization of evolving $N$-dimensional subspaces based on the theory of universal connections can also be given. This shows that the basic structure responsible for the occurrence of anholonomy effects in evolving quantum systems is the universal bundle $\eta$ which is the bundle of subspaces of arbitrary dimension $N$ in a Hilbert space.

The important issue of applying the idea of anholonomy to physical problems has not been dealt with in this article. There are spectacular applications such as holonomic quantum computation, the gauge kinematics of deformable bodies, quantum Hall-effect, fractional spin and statistics. The interested reader should consult the vast literature on the subject or as a first glance, the book of Shapere and Wilczek (1989).

*See also:* Fractional Quantum Hall Effect; Geometric Measure Theory; Holomorphic Dynamics; Moduli Spaces: An Introduction.

## Further Reading

Aharonov Y and Anandan J (1987) Phase change during a cyclic evolution. *Physical Review Letters* 58: 1593–1596.
Benedict MG and Fehér LGy (1989) Quantum jumps, geodesics, and the topological phase. *Physical Review D* 39: 3194–3196.
Berry MV (1984) Quantal phase factors accompanying adiabatic changes. *Proceedings of the Royal Society A* 392: 45–57.
Berry MV (1987) The adiabatic phase and Pancharatnam's phase for polarized light. *Journal of Modern Optics* 34: 1401–1407.
Bohm A and Mostafazadeh A (1994) The relation between the Berry and the Anandan–Aharonov connections for U($N$) bundles. *Journal of Mathematical Physics* 35: 1463–1470.
Kobayashi S and Nomizu K (1969) *Foundations of Differential Geometry*, vol. 2. Berlin: Springer.
Lévay P (1990) Geometrical description of SU(2) geometric phases. *Physical Review A* 41: 2837–2840.
Narasimhan MS and Ramanan S (1961) Existence of universal connections. *American Journal of Mathematics* 83: 563–572.
Pancharatnam S (1956) Generalized theory of interference, and its applications. *The Proceedings of the Indian Academy of Sciences* XLIV. No 5. Sec. A: 247–262.
Samuel J and Bhandari R (1988) General setting for Berry's phase. *Physical Review Letters* 60: 2339–2342.
Shapere A and Wilczek F (eds.) (1989) *Geometric Phases in Physics*. Wiley.
Simon B (1983) Holonomy, the quantum adiabatic theorem, and Berry's phase. *Physical Review Letters* 51: 2167–2170.
Wilczek F and Zee A (1984) Appearance of gauge structure in simple dynamical systems. *Physical Review Letters* 52: 2111–2114.

# Geophysical Dynamics

**M B Ziane**, University of Southern California, Los Angeles, CA, USA



## Introduction

The equations of geophysical fluid dynamics are the equations governing the motion of the atmosphere and the ocean, and are derived from the conservation equations from physics, namely conservation of mass, momentum, energy, and some other components such as salt for the ocean, humidity (or chemical pollutants) for the atmosphere.

The first assumption used in any circulation model is the well-accepted Boussinesq approximation, that is, the density differences are neglected in the system except in the buoyancy term and in the equation of state. The resulting system is the so-called Boussinesq equations (Pedlosky 1987). Due to the extremely high accuracy of this approximation, these equations are considered as the basic equations

in geophysical dynamics. From the computational point of view, however, the Boussinesq equations are still not accessible.

Owing to the difference of sizes of the vertical and horizontal dimensions, both in the atmosphere and in the ocean (10–20 km versus several thousands of kilometers), the second approximation is based on the smallness of the vertical length scales with respect to the horizontal length scales, that is, oceans (and the atmosphere) compose very thin layers. The scale analysis ensures that the dominant forces in the vertical-momentum equation come from the pressure gradient and the gravity. This leads to the so-called hydrostatic approximation, which amounts to replacing the vertical component of the momentum equation by the hydrostatic balance equation, and hence leading to the well-accepted primitive equations (PEs) (Washington and Parkinson 1986). As far as we know, the primitive equations were first considered by L F Richardson (1922); when it appeared that they were still too complicated they were left out and, instead, attention was focused on even simpler models, the geostrophic and quasigeostrophic models, considered in the late 1940s by J von Neumann and his collaborators, in particular J G Charney. With the increase of computing power, interest eventually returned to the PEs, which are now the core of many global circulation models (GCMs) or ocean global circulation models (OGCMs), available at the National Center for Atmospheric Research (NCAR) and elsewhere. GCMs and OGCMs are very complex models which contain many components, but still, the PEs are the central component for the dynamics of the air or the water. Further approximations based on the fast rotation of the Earth implying the smallness of the Rossby number lead to the quasigeostrophic and geostrophic equations (Pedlosky 1987).

The mathematical study of the PEs was initiated by Lions, Temam, and Wang in the early 1990s. They produced a mathematical formulation of the PEs which resembles that of the Navier–Stokes due to Leray, and obtained the existence, for all time, of weak solutions (see Lions *et al.* 1992a, b, 1993, 1995). Further works conducted during the 1990s have improved and supplemented these early results bringing the mathematical theory of the PEs to that of the three-dimensional incompressible Navier–Stokes equations (Constantin and Foias 1998, Teman 2001). In summary, the following results are now available which will be presented in this article:

1. existence of weak solutions for all time;
2. existence of strong solutions in space dimension three, local in time;
3. existence and uniqueness of a strong solution in space dimension two, for all time; and
4. uniqueness of weak solutions in space dimension two.

## The PEs of the Ocean

The ocean is made up of a slightly compressible fluid subject to a Coriolis force. The full set of equations of the large-scale ocean are the following: the conservation of momentum equation, the continuity equation (conservation of mass), the thermodynamics equation, the equation of state and the equation of diffusion for the salinity $S$:

$$\rho\frac{\mathrm{d}\boldsymbol{V}_3}{\mathrm{d}t} + 2\rho\boldsymbol{\Omega}\times\boldsymbol{V}_3 + \nabla_3 p + \rho\boldsymbol{g} = D \qquad [1]$$

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} + \rho\,\mathrm{div}_3\boldsymbol{V}_3 = 0 \qquad [2]$$

$$\frac{\mathrm{d}T}{\mathrm{d}t} = Q_T \qquad [3]$$

$$\frac{\mathrm{d}S}{\mathrm{d}t} = Q_S \qquad [4]$$

$$\rho = f(T,S,p) \qquad [5]$$

Here $\boldsymbol{V}_3$ is the three-dimensional velocity vector, $\boldsymbol{V}_3 = (u,v,w)$, $\rho$, $p$, $T$ are respectively, the density, pressure, and temperature, and $S$ is the concentration of salinity; $\boldsymbol{g} = (0,0,g)$ is the gravity vector, $D$ the molecular dissipation, $Q_T$ and $Q_S$ are the heat and salinity diffusions, respectively.

**Remark 1** The equation of state for the oceans is derived on a phenomenological basis. Only empirical forms of the function $f(T,S,\rho)$ are known (see Washington and Parkinson (1986)). It is natural, however, to expect that $\rho$ decreases if $T$ increases and that $\rho$ increases if $S$ increases. The simplest law is

$$\rho = \rho_0(1 - \beta_T(T - T_\mathrm{r}) + \beta_S(S - S_\mathrm{r})) \qquad [6]$$

corresponding to a linearization around reference values $\rho_0$, $T_\mathrm{r}$, $S_\mathrm{r}$ of respectively, the density, temperature, and the salinity, $\beta_T$ and $\beta_S$ are positive expansion coefficients.

The Mach number for the flow in the ocean is not large and, therefore, as a starting point, we can make the so-called Boussinesq approximation in which the density is assumed constant, $\rho = \rho_0$, except in the buoyancy term and in the equation of state. This amounts to replacing [1], [2] by

$$\rho_0\frac{\mathrm{d}\boldsymbol{V}_3}{\mathrm{d}t} + 2\rho_0\Omega\times\boldsymbol{V}_3 + \nabla_3 p + \rho\boldsymbol{g} = D \qquad [7]$$

$$\mathrm{div}_3\boldsymbol{V}_3 = 0 \qquad [8]$$

Furthermore, since for large-scale ocean, the horizontal scale is much larger than the vertical one, a scale analysis (Pedlosky 1987) shows that $\partial p/\partial z$ and $\rho g$ are

the dominant terms in the vertical-momentum equation, leading to the hydrostatic approximation

$$\frac{\partial p}{\partial z} = -\rho g \qquad [9]$$

For mid-latitude regional studies, it is usual to consider the beta-plane approximation of the equations. Thus, we assume that the ocean fills a domain $\mathcal{M}_\varepsilon$ of $\mathbb{R}^3$. The top of the ocean is a domain $\Gamma_i$ included in the surface of the earth $S_a$ (sphere of radius $a$ centered at 0). The bottom $\Gamma_b$ of the ocean is defined by $(z = x_3 = r - a)$, $z = -\varepsilon h(\theta, \varphi)$, where $\varepsilon > 0$ is a positive parameter. It is introduced to take into consideration the smallness of the vertical scales compared to the horizontal scales. $h$ is a function of class $\mathcal{C}^2$ at least on $\bar{\Gamma}_i$; it is assumed also that $h$ is bounded from below, that is, $0 < \underline{h} \le h(\theta, \varphi) \le \underline{h}$, $(\theta, \varphi) \in \Gamma_i$. The lateral surface $\Gamma_l$ consists of the part of cylinder $\{(\theta, \varphi) \in \partial\Gamma_i, -\varepsilon h(\theta, \varphi) \le r \le 0\}$. The PEs of the ocean are given by

$$\frac{\partial \boldsymbol{v}}{\partial t} + \nabla_v \boldsymbol{v} + w\frac{\partial \boldsymbol{v}}{\partial z} + \frac{1}{\rho_0}\nabla p + 2\Omega \sin\theta \boldsymbol{k} \times \boldsymbol{v} - \mu_v \Delta \boldsymbol{v} - \nu_v \frac{\partial^2 \boldsymbol{v}}{\partial z^2} = F_v \qquad [10]$$

$$\frac{\partial p}{\partial z} = -\rho, \qquad \operatorname{div} \boldsymbol{v} + \frac{\partial w}{\partial z} = 0 \qquad [11]$$

$$\frac{\partial T}{\partial t} + \nabla_{\boldsymbol{v}} T + w\frac{\partial T}{\partial z} - \mu_T \Delta T - \nu_T \frac{\partial^2 T}{\partial z^2} = F_T \qquad [12]$$

$$\frac{\partial S}{\partial t} + \nabla_{\boldsymbol{v}} S + w\frac{\partial S}{\partial z} - \mu_S \Delta S - \nu_S \frac{\partial^2 S}{\partial z^2} = F_S \qquad [13]$$

$$\operatorname{div} \int_{-h}^{0} \boldsymbol{v}\, dz = 0 \qquad [14]$$

$$p = p_s + P, \qquad P = P(T, S) = g \int_z^0 p\, dz' \qquad [15]$$

$$\rho = \rho_0(1 - \beta_T(T - T_r) + \beta_S(S - S_r)) \qquad [16]$$

$$\int_{\mathcal{M}_\varepsilon} S\, d\mathcal{M}_\varepsilon = 0 \qquad [17]$$

where $\boldsymbol{v}$ is the horizontal velocity of the water, $w$ is the vertical velocity, and $T_r$, $S_r$ are averaged (or reference) values of $T$ and $S$. The diffusion coefficients $\mu_v$, $\mu_T$, $\mu_S$ and $\nu_v$, $\nu_T$, $\nu_S$ are different in the horizontal and vertical directions, accounting for some eddy diffusions in the sense of Smagorinsky (1962). Note that $F_v$, $F_T$, and $F_S$ correspond to volumic sources of horizontal momentum, heat, and salt, respectively.

### Boundary conditions

There are several sets of natural boundary conditions that one can associate to the PEs; for instance, the following:

**On the top of the ocean** $\Gamma_i(z = 0)$

$$\nu_v \frac{\partial \boldsymbol{v}}{\partial z} + \alpha_v(\boldsymbol{v} - \boldsymbol{v}_a) = \tau_v, \qquad w = 0$$
$$\nu_T \frac{\partial T}{\partial z} + \alpha_T(T - T_a) = 0, \qquad \frac{\partial S}{\partial z} = 0 \qquad [18]$$

**At the bottom of the ocean** $\Gamma_b(z = -h(\theta, \varphi))$

$$\boldsymbol{v} = 0, \; w = 0, \; \frac{\partial T}{\partial \boldsymbol{n}_T} = 0, \; \frac{\partial S}{\partial \boldsymbol{n}_S} = 0 \qquad [19]$$

**On the lateral boundary** $\Gamma_l = \{-h(\theta, \varphi) < z < 0, (\theta, \varphi) \in \partial\Gamma_i\}$

$$\boldsymbol{v} = 0, \; w = 0, \; \frac{\partial T}{\partial \boldsymbol{n}_T} = 0, \; \frac{\partial S}{\partial \boldsymbol{n}_S} = 0 \qquad [20]$$

Here $\boldsymbol{n} = (n_H, n_z)$ is the unit outward normal on $\partial\mathcal{M}_\varepsilon$ decomposed into its horizontal and vertical components; the conormal derivatives $\partial/\partial \boldsymbol{n}_T$ and $\partial/\partial \boldsymbol{n}_S$ are those associated with the linear (temperature and salinity) operators,

$$\frac{\partial}{\partial \boldsymbol{n}_T} = \mu_T n_H \cdot \nabla + \nu_T n_z \frac{\partial}{\partial z}$$
$$\frac{\partial}{\partial \boldsymbol{n}_S} = \mu_S n_H \cdot \nabla + \nu_S n_z \frac{\partial}{\partial z} \qquad [21]$$

Equations [10]–[17] with boundary conditions [18]–[20] are supplemented with the initial conditions

$$\boldsymbol{v}|_{t=0} = \boldsymbol{v}_0, \; T|_{t=0} = T_0, \; S|_{t=0} = S_0 \qquad [22]$$

where $\boldsymbol{v}_0, T_0, S_0$ are given initial data.

Following the work of Lions *et al.* (1992a, b, 1993, 1995) (see also Temam and Ziane (2004)), we introduce the following function spaces $V = V_1 \times V_2 \times V_3$, $H = H_1 \times H_2 \times H_3$, where

$$V_1 = \left\{\boldsymbol{v} \in H^1(\mathcal{M})^2, \operatorname{div} \int_{-h}^{0} \boldsymbol{v} dz = 0, \; \boldsymbol{v} = 0 \text{ on } \Gamma_b \cup \Gamma_l\right\}$$

$$V_2 = H^1(\mathcal{M})$$

$$V_3 = \dot{H}^1(\mathcal{M}) = \left\{S \in H^1(\mathcal{M}), \int_{\mathcal{M}} S d\mathcal{M} = 0\right\}$$

$$H_1 = \left\{\boldsymbol{v} \in L^2(\mathcal{M})^2, \operatorname{div} \int_{-h}^{0} \boldsymbol{v}\, dz = 0, \; n_H \cdot \int_{-h}^{0} \boldsymbol{v}\, dz = 0 \text{ on } \partial\Gamma_i \text{ (i.e., on } \Gamma_l)\right\}$$

$$H_2 = L^2(\mathcal{M})$$

$$H_3 = \dot{L}^2(\mathcal{M}) = \left\{S \in L^2(\mathcal{M}), \int_{\mathcal{M}} S d\mathcal{M} = 0\right\}$$

The global existence of weak solutions is established in Lions *et al.* (1992b), using the Galerkin method and assuming the $H^2$-regularity of the GFD–Stokes problem, which was established in Ziane

(1995). A more general global existence result based on the method of finite differences in time and independent of the $H^2$-regularity is established in Temam and Ziane (2004), which we state here.

**Theorem 2** *Given $t_1 > 0$, $U_0$ in $H$, and $F = (F_v, F_T, F_S)$ in $L^2(0, t_1; H)$; $g = g_v$, $g_T$ is given in $L^2(0, t_1; (L^2(\Gamma_i)^3)$. Then there exists*

$$U \in L^\infty(0, t_1; H) \cap L^2(0, t_1; V) \qquad [23]$$

*which is a weak solution of* [10]–[17] and [18]–[20], [22]; *furthermore, $U$ is weakly continuous from $[0, t_1]$ into $H$.*

### Strong Solutions

The local existence and uniqueness of strong solutions of the primitive equations of the ocean relies on the $H^2$-regularity of the stationary linear primitive equations associated to [10]–[17]:

$$\frac{1}{\rho_0}\nabla p + 2\Omega \sin\theta \boldsymbol{k} \times \boldsymbol{v} - \mu_v \Delta \boldsymbol{v} - \nu_v \frac{\partial^2 \boldsymbol{v}}{\partial z^2} = F_v$$
$$\int_{-h}^{0} \operatorname{div} \boldsymbol{v}\, \mathrm{d}z = 0 \qquad [24]$$

$$-\mu_T \Delta T - \nu_T \frac{\partial^2 T}{\partial z^2} = F_T$$
$$-\mu_S \Delta S - \nu_S \frac{\partial^2 S}{\partial z^2} = F_S \qquad [25]$$

$$p = p_s + P, \quad P = P(T, S) = g\int_z^0 p\, \mathrm{d}z' \qquad [26]$$

with boundary conditions [18]–[20]. Here $F_v, F_T, F_S$ are independent of time. We have the following $H^2$-regularity of solutions (Ziane 1995, Hu *et al.* 2002, Temam and Ziane 2004).

**Theorem 3** *Assume that $h$ is in $\mathcal{C}^4(\bar{\Gamma}_i), h \ge \underline{h} > 0$, $F_v, F_T, F_S \in (\mathbb{L}^2(\mathcal{M}_\varepsilon))^4$ and $g_v = \tau_v + \alpha_v \boldsymbol{v}_a$, $g_T = \alpha_a T_a \in (\mathbb{H}_0^1(\Gamma_i))^4$. Let $(v, T, S; p) \in (\mathbb{H}^1(M_\varepsilon))^4 \times L^2(\Gamma_i)$ be a weak solution of* [24]–[26]. *Then*

$$(v, p) \in \left(\mathbb{H}^2(\mathcal{M}_\varepsilon)\right)^2 \times H^1(\mathcal{M}_\varepsilon)$$
$$(T, S) \in \left(\mathbb{H}^2(\mathcal{M}_\varepsilon)\right)^2 \qquad [27]$$

*Moreover, the following inequalities hold:*

$$|v|^2_{\mathbb{H}^2(\mathcal{M}_\varepsilon)} + \varepsilon |p|^2_{H^1(\Gamma_i)} \le C\Big[|F_v|^2_\varepsilon + |g_v|^2_{L^2(\Gamma_i)} + \varepsilon|\nabla g_v|^2_{L^2(\Gamma_i)}\Big]$$

$$|T|^2_{\mathbb{H}^2(\mathcal{M}_\varepsilon)} \le C\Big[|F_T|^2 + |g_T|^2_{L^2(\Gamma_i)} + \varepsilon|\nabla g_T|^2_{L^2(\Gamma_i)}\Big]$$
$$|S|^2_{\mathbb{H}^2(\mathcal{M}_\varepsilon)} \le C|F_S|^2$$

*where $C$ is a positive constant independent of $\varepsilon$.*

We now turn our attention to the nonlinear time-dependent PEs. The local-in-time existence and uniqueness of strong solutions is obtained in Temam and Ziane (2004); see also Hu *et al.* (2003) and Guillén-González *et al.* (2001). The proof is more involved than that of the three-dimensional Navier–Stokes equations. It consists of several steps. In the first step, one proves the global existence of strong solutions to the linearized time-dependent problem. In the second step, one uses the solution of the linearized equation in order to reduce the PEs to a nonlinear evolution equation with zero initial data and homogeneous boundary conditions. Finally, in the last step, one uses nonisotropic Sobolev inequalities together with Theorem 3. The local existence result is given by the following:

**Theorem 4** *Let $\varepsilon > 0$ be given. We assume that $\Gamma_i$ is of class $\mathcal{C}^3$ and that $h : \bar{\Gamma}_i \to \mathbb{R}_+$ is of class $\mathcal{C}^3$. We are given $U_0$ in $V$, $F = (F_v, F_T, F_S)$ in $L^2(0, t_1; H)$ with $\partial F/\partial t$ in $L^2(0, t_1; L^2(\mathcal{M}_\varepsilon)^4)$, and $g = (g_v, g_T)$ in $L^2(0, t_1; H_0^1(\Gamma_i)^3)$ with $\partial g/\partial t$ in $L^2(0, t_1; H_0^1(\Gamma_i)^3)$. Then there exists $t_* > 0, t_* = t_*(\|U_0\|)$, and there exists a unique solution $U = U(t) = (\boldsymbol{v}(t), T(t), S(t))$ of the PEs* [10]–[17], [18]–[20], *and* [22] *such that*

$$U \in C([0, t_*]; V) \cap L^2(0, t_*, H^2(\mathcal{M}_\varepsilon)^4) \qquad [28]$$

## The PEs of the Atmosphere

In this section we briefly describe the PEs of the atmosphere, for which all the mathematical results obtained for the PEs of the ocean are valid. We start from the conservations equations similar to [1]–[5]; in fact [1] and [2] are the same; the equation of energy conservation (temperature) is slightly different from [3] because of the compressibility of air; the state equation is that of perfect gas instead of [5]; finally, instead of the concentration of salt in the water, we consider the amount of water in air, $q$. Hence, we have

$$\rho \frac{\mathrm{d}\boldsymbol{V}_3}{\mathrm{d}t} + 2\rho\boldsymbol{\Omega} \times \boldsymbol{V}_3 + \nabla_3 p + \rho \boldsymbol{g} = D \qquad [29]$$

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} + \rho \operatorname{div}_3 \boldsymbol{V}_3 = 0 \qquad [30]$$

$$hc_p \frac{\mathrm{d}T}{\mathrm{d}t} - \frac{RT}{p}\frac{\mathrm{d}p}{\mathrm{d}t} = Q_T$$
$$\frac{\mathrm{d}q}{\mathrm{d}t} = 0, \qquad p = RpT \qquad [31]$$

Here $c_p > 0$ is the specific heat of air at constant pressure, and $R$ is the specific gas constant for the air. Proceeding as in the PEs of the ocean, we decompose $\boldsymbol{V}_3$ into its horizontal and vertical components, $\boldsymbol{V}_3 = \boldsymbol{v} + w$; then we use the hydrostatic approximation, replacing the equation of

conservation of vertical momentum by the hydrostatic equation [9]. We find

$$\frac{\partial v}{\partial t} + \nabla_v \boldsymbol{v} + w\frac{\partial \boldsymbol{v}}{\partial z} + \frac{1}{\nabla \rho_0} p + 2\Omega \sin\theta \times \boldsymbol{v} - \mu_v \Delta \boldsymbol{v} - \nu_v \frac{\partial^2 \boldsymbol{v}}{\partial z^2} = 0 \quad [32]$$

$$\frac{\partial p}{\partial z} = -\rho g \quad [33]$$

$$\frac{\partial T}{\partial t} + \nabla_v T + w\frac{\partial T}{\partial z} - \mu_T \Delta T - \nu_T \frac{\partial^2 T}{\partial z^2} - \frac{RT}{p}\frac{\mathrm{d}p}{\mathrm{d}t} = Q_T \quad [34]$$

$$\frac{\partial q}{\partial t} + \nabla_v q + w\frac{\partial q}{\partial z} - \mu_q \Delta q - \nu_q \frac{\partial^2 q}{\partial z^2} = 0 \quad [35]$$

$$p = RpT \quad [36]$$

The right-hand side of [34], represents the solar heating.

### Change of Vertical Coordinate

Since $\rho$ does not vanish, the hydrostatic equation [33] implies that $p$ is a strictly decreasing function of $z$, and we are thus allowed to use $p$ as the vertical coordinate; hence in spherical geometry the independent variables are now $\varphi, \theta, p$, and $t$. By an abuse of notation, we still denote by $\boldsymbol{v}$, $p$, $T$, $q$, $\rho$ these functions expressed in the $\varphi$, $\theta$, $p$, $t$ variables. We denote by $\omega$ the vertical component of the wind in the new variables, and one can show that the PEs of the atmosphere become

$$\frac{\partial \boldsymbol{v}}{\partial t} + \nabla_v \boldsymbol{v} + \omega\frac{\partial \boldsymbol{v}}{\partial p} + 2\Omega \sin\theta k \times \boldsymbol{v} + \nabla\Phi - L_v \boldsymbol{v} = F_v \quad [37]$$

$$\frac{\partial \Phi}{\partial p} + \frac{R}{p} T = 0 \quad [38]$$

$$\operatorname{div} \boldsymbol{v} + \frac{\partial \mathrm{w}}{\partial}\mathrm{z} = 0 \quad [39]$$

$$\frac{\partial T}{\partial t} + \nabla_v T + \omega\frac{\partial T}{\partial p} - \frac{RT}{p}\omega - L_T = F_T \quad [40]$$

$$\frac{\partial q}{\partial t} + \nabla_v q + \omega\frac{\partial q}{\partial z} - L_q q = F_q \quad [41]$$

$$p = R\rho T \quad [42]$$

We have denoted by $\Phi = gz$ the geopotential ($z$ is now function of $\varphi, \theta, p, t$); $L_v, L_T, L_q$ are the Laplace operators, with suitable eddy viscosity coefficients, expressed in the $\varphi$, $\theta$, $p$ variables. Hence, for example,

$$L_v \boldsymbol{v} = \mu_v \Delta \boldsymbol{v} + \nu_v \frac{\partial}{\partial p}\left[\left(\frac{gp}{R\bar{T}}\right)^2 \frac{\partial \boldsymbol{v}}{\partial p}\right] \quad [43]$$

with similar expressions for $L_T$ and $L_q$. Note that $F_T$ corresponds to the heating of the Sun, whereas $F_v$ and $F_q$ (which vanish in reality) are added here for mathematical generality. The change of variable gives, for $\partial^2 \boldsymbol{v}/\partial z^2$, a term different from the coefficient of $\nu_v$. The expression above is simplified for of this coefficient; the simplification is legitimate because $\nu_v$ is a very small coefficient (in particular, $T$ has been replaced by $\bar{T}$ (known) average value of the temperature).

### Pseudogeometrical Domain

For physical and mathematical reasons, we do not allow the pressure to go to zero, and assume that $p \geq p_0$, with $p_0 > 0$ "small." Physically, in the very high atmosphere ($p$ very small), the air is ionized and the equations above are not valid anymore. The pressure is then restricted to an interval $p_0 < p < p_1$, where $p_1$ is a value of the pressure smaller in average than the pressure on Earth, so that the isobar $p = p_1$ is slightly above the Earth and the isobar $p = p_0$ is an isobar high in the sky. We study the motion of the air between these two isobars.

For the whole atmosphere, the boundary of this domain

$$\mathcal{M} = \{(\varphi, \theta, p), p_0 < p < p_1\}$$

consists first of an upper part $\Gamma_{\mathrm{u}}, p = p_0$; the lower part $p = p_1$ is divided into two parts $\Gamma_i$ the part of $p = p_1$ at the interface with the ocean, and $\Gamma_e$ the part of $p = p_1$ above the earth.

### Boundary Conditions

Typically, the boundary conditions are as follows:

**On the top of the atmosphere** $\Gamma_{\mathrm{u}}(p = p_0)$

$$\frac{\partial \boldsymbol{v}}{\partial p} = 0, \ \omega = 0, \ \frac{\partial T}{\partial p} = 0, \ \frac{\partial q}{\partial p} = 0 \quad [44]$$

## Acknowledgments

This work was partially supported by the National Science Foundation under the grant NSF-DMS-0204863.

*See also:* Boundary Control Method and Inverse Problems of Wave Propagation; Compressible Flows: Mathematical Theory; Fluid Mechanics: Numerical Methods; Turbulence Theories.

## Further Reading


Bresch D, Kazhikov A, and Lemoine J (2004/05) On the two-dimensional hydrostatic Navier–Stokes equations. *SIAM Journal of Mathematical Analysis* 36(3): 796–814.

Constantin P and Foias C (1998) *Navier–Stokes Equations.* Chicago: University of Chicago Press.

Guillén-González F, Masmoudi N, and Rodríguez-Bellido MA (2001) Anisotropic estimates and strong solutions of the primitive equations. *Differential and Integral Equations* 14(11): 1381–1408.

Hu C, Temam R, and Ziane M (2002) Regularity results for linear elliptic problems related to the primitive equations. *Chinese Annals of Mathematics Series B* 23(2): 277–292.

Hu C, Temam R, and Ziane M (2003) The primitive equations of the large scale ocean under the small depth hypothesis. *Discrete and Continuous Dynamical Systems Series A* 9(1): 97–131.

Lions JL, Temam R, and Wang S (1992a) New formulations of the primitive equations of the atmosphere and applications. *Nonlinearity* 5: 237–288.

Lions JL, Temam R, and Wang S (1992b) On the equations of the large-scale ocean. *Nonlinearity* 5: 1007–1053.

Lions JL, Temam R, and Wang S (1993) Models of the coupled atmosphere and ocean (CAO I). In: Oden JT (ed.) *Computational Mechanics Advances*, vol. 1: 5–54. North-Holland: Elsevier Science.

Lions JL, Teman R, and Wang S (xxxx) Numerical analysis of the coupled models of atmosphere and ocean. (CAO II). In: Oden JT (ed.) *Computational Mechanics Advances*, vol. 1: 55–119. North-Holland: Elsevier Science.

Lions JL, Temam R, and Wang S (1995) Mathematical study of the coupled models of atmosphere and ocean (CAO III). *Journal de Mathématiques Pures et Appliquées, Neuviéme Série* 74: 105–163.

Pedlosky J (1987) *Geophysical Fluid Dynamics*, 2nd edn. New York: Springer.

Schlichting H (1979) *Boundary Layer Theory*, 7th edn. New York: McGraw-Hill.

Smagorinsky J (1963) General circulation experiments with the primitive equations, I. The basic experiment. *Monthly Weather Review*. 91: 98–164.

Temam R (2001) *Navier–Stokes Equations, Studies in Mathematics and Its Applications, AMS-Chelsea Series.* Providence: American Mathematical Society.

Temam R and Ziane M (2004) Some mathematical problems in geophysical fluid dynamics. In: Friedlander S and Serre D (eds.) *Handbook of Mathematical Fluid Dynamics*, vol. III, pp. 535–657. Amsterdam: North-Holland.

Washington WM and Parkinson CL (1986) *An Introduction to Three Dimensional Climate Modeling.*Oxford: Oxford University Press.

Ziane M (1995) Regularity results for Stokes type systems. *Applicable Analysis* 58: 263–292.


# Gerbes in Quantum Field Theory

**J Mickelsson**, KTH Physics, Stockholm, Sweden



## Definitions and an Example

A gerbe can be viewed as a next step in a ladder of geometric and topological objects on a manifold which starts from ordinary complex-valued functions and in the second step of sections of complex line bundles.

It is useful to recall the construction of complex line bundles and their connections. Let $M$ be a smooth manifold and $\{U_\alpha\}$ an of open cover of $M$ which trivializes a line bundle $L$ over $M$. Topologically, up to equivalence, the line bundle is completely determined by its Chern class, which is a cohomology class $[c] \in H^2(M, \mathbb{Z})$. On each open set $U_\alpha$ we may write $2\pi i c = \mathrm{d}A_\alpha$, where $A_\alpha$ is a 1-form. On the overlaps $U_{\alpha\beta} = U_\alpha \cap U_\beta$ we can write

$$A_\alpha - A_\beta = f_{\alpha\beta}^{-1} \mathrm{d} f_{\alpha\beta} \qquad [1]$$

at least when $U_{\alpha\beta}$ is contractible, where $f_{\alpha\beta}$ is a circle-valued complex function on the overlap. The data $\{c, A_\alpha, f_{\alpha\beta}\}$ define what is known as a (representative of a) Deligne cohomology class on the open cover $\{U_\alpha\}$. The 1-forms $A_\alpha$ are the local potentials of the curvature form $2\pi i c$ and the $f_{\alpha\beta}$'s are the transition functions of the line bundle $L$. Each of these three different data defines separately the equivalence class of the line bundle but together they define the line bundle with a connection.

The essential thing here is that there is a bijection between the second integral cohomology of $M$ and the set of equivalence classes of complex line bundles over $M$. It is natural to ask whether there is a geometric realization of integral third (or higher) cohomology. In fact, gerbes provide such a realization. Here, we shall restrict to a smooth differential geometric approach which by no means is the most general possible, but it is sufficient for most applications to quantum field theory. However, there are examples of gerbes over orbifolds that do not need to come from finite group action on a manifold, which are not covered by the following definition.

For the examples in this article, it is sufficient to adapt the following definition. A gerbe over a manifold $M$ (without geometry) is simply a principal bundle $\pi: P \to M$ with fiber equal to $PU(H)$, the projective unitary group of a Hilbert space $H$. The Hilbert space may be either finite or infinite dimensional.

The quantum field theory applications discussed in this article are related to the chiral anomaly for

fermions in external fields. The link comes from the fact that the chiral symmetry breaking leads in the generic case to projective representations of the symmetry groups. For this reason, when modding out by the gauge or diffeomorphism symmetries, one is led to study bundles of projective Hilbert spaces. The anomaly is reflected as a nontrivial characteristic class of the projective bundle, known in mathematics literature as the Dixmier–Douady class.

In a suitable open cover, the bundle $P$ has a family of local trivializations with transition functions $g_{\alpha\beta}: U_{\alpha\beta} \to PU(H)$, with the usual cocycle property

$$g_{\alpha\beta} g_{\beta\gamma} g_{\gamma\alpha} = 1 \quad [2]$$

on triple overlaps. Assuming that the overlaps are contractible, we can choose lifts $\hat{g}_{\alpha\beta}: U_{\alpha\beta} \to U(H)$, to the unitary group of the Hilbert space. However,

$$\hat{g}_{\alpha\beta} \hat{g}_{\beta\gamma} \hat{g}_{\gamma\alpha} = f_{\alpha\beta\gamma} \quad [3]$$

where the $f$'s are circle-valued functions on triple overlaps. They satisfy automatically the cocycle property

$$f_{\alpha\beta\gamma} f^{-1}_{\alpha\beta\delta} f_{\alpha\gamma\delta} f^{-1}_{\beta\gamma\delta} = 1 \quad [4]$$

on quadruple overlaps. There is an important difference between the finite- and infinite-dimensional cases. In the finite-dimensional case, the circle bundle $U(H) \to U(H)/S^1 = PU(H)$ reduces to a bundle with fiber $\mathbb{Z}/N\mathbb{Z} = \mathbb{Z}_N$, where $N = \dim H$. This follows from $\mathrm{U}(N)/S^1 = \mathrm{SU}(N)/Z_N$ and the fact that $\mathrm{SU}(N)$ is a subgroup of $\mathrm{U}(N)$. For this reason one can choose the lifts $\hat{g}_{\alpha\beta}$ such that the functions $f_{\alpha\beta\gamma}$ take values in the finite subgroup $\mathbb{Z}_N \subset S^1$.

The functions $f_{\alpha\beta\gamma}$ define an element $a = \{a_{\alpha\beta\gamma\delta}\}$ in the Čech cohomology $H^3(\mathcal{U}, \mathbb{Z})$ by a choice of logarithms,

$$2\pi i a_{\alpha\beta\gamma\delta} = \log f_{\alpha\beta\gamma} - \log f_{\alpha\beta\delta} + \log f_{\alpha\gamma\delta} - \log f_{\beta\gamma\delta} \quad [5]$$

In the finite-dimensional case, the Čech cocycle is necessarily torsion, $Na = 0$, but not so if $H$ is infinite dimensional. In the finite-dimensional case (by passing to a good cover and using the Čech – de Rham equivalence over real or complex numbers), the class is third de Rham cohomology constructed from the transition functions is necessarily zero. Thus, in general one has to work with Čech cohomology to preserve torsion information. One can prove:

**Theorem** *The construction above is a one-to-one map between the set of equivalence classes of* $PU(H)$ *bundles over M and elements of* $H^3(M, \mathbb{Z})$.

The characteristic class in $H^3(M, \mathbb{Z})$ of a $PU(H)$ bundle is called the *Dixmier–Douady class*.

### First example

Let $M$ be an oriented Riemannian manifold and $FM$ its bundle of oriented orthonormal frames. The structure group of $FM$ is the rotation group $\mathrm{SO}(n)$ with $n = \dim M$. The spin bundle (when it exists) is a double covering $\mathrm{Spin}(M)$ of $FM$, with structure group $\mathrm{Spin}(n)$, a double cover of $\mathrm{SO}(n)$. Even when the spin bundle does not exist there is always the bundle $\mathrm{Cl}(M)$ of Clifford algebras over $M$. The fiber at $x \in M$ is the Clifford algebra defined by the metric $g_x$, that is, it is the complex $2^n$-dimensional algebra generated by the tangent vectors $v \in T_x(m)$ with the defining relations

$$\gamma(u)\gamma(v) + \gamma(v)\gamma(u) = 2g_x(u, v)$$

The Clifford algebra has a faithful representation in $N = 2^{[n/2]}$ dimensions ($[x]$ is the integral part of $x$) such that

$$\gamma(a \cdot u) = S(a)\gamma(u)S(a)^{-1}$$

where $S$ is an unitary representation of $\mathrm{Spin}(n)$ in $\mathbb{C}^N$. Since $\mathrm{Spin}(n)$ is a double cover of $\mathrm{SO}(n)$, the representation $S$ may be viewed as a projective representation of $\mathrm{SO}(n)$. Thus again, if the overlaps $U_{\alpha\beta}$ are contractible, we may choose a lift of the frame bundle transition functions $g_{\alpha\beta}$ to unitaries $\hat{g}_{\alpha\beta}$ in $H = \mathbb{C}^N$. In this case, the functions $f_{\alpha\beta\gamma}$ reduce to $\mathbb{Z}_2$-valued functions, and the obstruction to the lifting problem, which is the same as the obstruction to the existence of spin structure, is an element of $H^2(M, \mathbb{Z}_2)$, known as the second Stiefel–Whitney class $w_2$. The image of $w_2$ with respect to the Bockstein map (in this case, given by the formula [5]) gives a 2-torsion element in $H^3(M, \mathbb{Z})$, the Dixmier–Douady class.

Another way to think of a gerbe is the following (we shall see that this arises in a natural way in quantum field theory). There is a canonical complex line bundle $L$ over $PU(H)$, the associated line bundle to the circle bundle $S^1 \to U(H) \to PU(H)$. Pulling back $L$ by the local transition functions $g_{\alpha\beta} \to PU(H)$, we obtain a family of line bundles $L_{\alpha\beta}$ over the open sets $U_{\alpha\beta}$. By the cocycle property [2] we have natural isomorphisms

$$L_{\alpha\beta} \otimes L_{\beta\gamma} = L_{\alpha\gamma} \quad [6]$$

We can take this as a definition of a gerbe over $M$: a collection of line bundles over intersections of open sets in an open cover of $M$, satisfying the cocycle condition [6]. By [6] we have a trivialization

$$L_{\alpha\beta} \otimes L_{\beta\gamma} \otimes L_{\gamma\alpha} = f_{\alpha\beta\gamma} \cdot \mathbf{1} \quad [7]$$

where the $f$'s are circle-valued functions on the triple overlaps. By the theorem above, we conclude

that indeed the data in [6] define (an equivalence class of) a principal $PU(H)$ bundle.

If $L_{\alpha\beta}$ and $L'_{\alpha\beta}$ are two systems of local line bundles over the same cover, then the gerbes are equivalent if there is a system of line bundles $L_\alpha$ over open sets $U_\alpha$ such that

$$L'_{\alpha\beta} = L_{\alpha\beta} \otimes L^*_\alpha \otimes L_\beta \qquad [8]$$

on each $U_{\alpha\beta}$.

A gerbe may come equipped with geometry, encoded in a Deligne cohomology class with respect to a given open covering of $M$. The Deligne class is given by functions $f_{\alpha\beta\gamma}$, 1-forms $A_{\alpha\beta}$, 2-forms $F_\alpha$, and a global 3-form (the Dixmier–Douady class of the gerbe) $\Omega$, subject to the conditions

$$\begin{aligned} dF_\alpha &= 2\pi i\Omega \\ F_\alpha - F_\beta &= dA_{\alpha\beta} \\ A_{\alpha\beta} - A_{\alpha\gamma} + A_{\beta\gamma} &= f^{-1}_{\alpha\beta\gamma} df_{\alpha\beta\gamma} \end{aligned} \qquad [9]$$

## Gerbes from Canonical Quantization

Let $D_x$ be a family of self-adjoint Fredholm operators in a complex Hilbert space $H$ parametrized by $x \in M$. This situation arises in quantum field theory, for example, when $M$ is some space of external fields, coupled to Dirac operator $D$ on a compact manifold. The space $M$ might consist of gauge potentials (modulo gauge transformations) or $M$ might be the moduli space of Riemann metrics. In these examples, the essential spectrum of $D_x$ is both positive and negative and the family $D_x$ defines an element of $K^1(M)$. In fact, one of the definitions of $K^1(M)$ is that its elements are homotopy classes of maps from $M$ to the space $\mathcal{F}_*$ of self-adjoint Fredholm operators with both positive and negative essential spectrum. In physics applications, one deals most often with unbounded Hamiltonians, and the operator norm topology must be replaced by something else; popular choices are the Riesz topology defined by the map $F \mapsto F/(|F|+1)$ to bounded operators or the gap topology defined by graph metric.

The space $\mathcal{F}_*$ is homotopy equivalent to the group $G = U_1(H)$ of unitary operators $g$ in $H$ such that $g-1$ is a trace-class operator. This space is a classifying space for principal $U_{res}$ bundles, where $U_{res}$ is the group of unitary operators $g$ in a polarized complex Hilbert space $\mathcal{H} = \mathcal{H}_+ \oplus \mathcal{H}_-$ such that the off-diagonal blocks of $g$ are Hilbert–Schmidt operators. This is related to Bott periodicity. There is a natural principal bundle $P$ over $G = U_1(H)$ with fiber equal to the group $\Omega G$ of based loops in $G$. The total space $P$ consists of smooth paths $f(t)$ in $G$ starting from the neutral element such that $f^{-1}df$ is smooth and periodic. The projection $P \to G$ is the evaluation at the end point $f(1)$. The fiber is clearly $\Omega G$. By Bott periodicity, the homotopy groups of $\Omega G$ are shifted from those of $G$ by one dimension, that is,

$$\pi_n \Omega G = \pi_{n+1} G$$

The latter are zero in even dimensions and equal to $\mathbb{Z}$ in odd dimensions. On the other hand, it is known that the even homotopy groups of $U_{res}(\mathcal{H})$ are equal to $\mathbb{Z}$ and the odd ones vanish. In fact, with a little more effort, one can show that the embedding of $\Omega G$ to $U_{res}(\mathcal{H})$ is a homotopy equivalence, when $\mathcal{H} = L^2(S^1, H)$, the polarization being the splitting to nonnegative and negative Fourier modes and the action of $\Omega G$ is the pointwise multiplication on $H$-valued functions on the circle $S^1$.

Since $P$ is contractible, it is indeed the classifying bundle for $U_{res}$ bundles. Thus, we conclude that "$K^1(M)$ = the set of homotopy classes of maps $M \to G$ = the set of equivalence classes of $U_{res}$ bundles over $M$." The relevance of this fact in quantum field theory follows from the properties of representations of the algebra of canonical anticommutation relations ($CAR$). For any complex Hilbert space $H$, this algebra is the algebra generated by elements $a(v)$ and $a^*(v)$, with $v \in H$, subject to the relations

$$a^*(u)a(v) + a(v)a^*(u) = 2 < v, u >$$

where the Hilbert space inner product on the right-hand side is antilinear in the first argument, and all other anticommutators vanish. In addition, $a^*(u)$ is linear and $a(v)$ antilinear in its argument.

An irreducible Dirac representation of the CAR algebra is given by a polarization $H = H_+ \oplus H_-$. The representation is characterized by the existence of a vacuum vector $\psi$ in the fermionic Fock space $\mathcal{F}$ such that

$$a^*(u)\psi = 0 = a(v)\psi \quad \text{for } u \in H_-,\ v \in H_+ \qquad [10]$$

A theorem of D Shale and W F Stinespring says that two Dirac representations defined by a pair of polarizations $H_+, H'_+$ are equivalent if and only if there is $g \in U_{res}(H_+ \oplus H_-)$ such that $H'_+ = g \cdot H_+$. In addition, in order that a unitary transformation $g$ is implementable in the Fock space, that is, there is a unitary operator $\hat{g}$ in $\mathcal{F}$ such that

$$\hat{g} a^*(v) \hat{g}^{-1} = a^*(gv), \quad \forall v \in H \qquad [11]$$

and similarly for the $a(v)$'s, one must have $g \in U_{res}$ with respect to the polarization defining the vacuum vector. This condition is both necessary and sufficient.

The polarization of the one-particle Hilbert space comes normally from a spectral projection onto the positive-energy subspace of a Hamilton operator. In the background field problems one studies families of Hamilton operators $D_x$ and then one would like to construct a family of fermionic Fock spaces parametrized by $x \in M$. If none of the Hamilton operators has zero modes, this is unproblematic. However, the presence of zero modes makes it impossible to define the positive-energy subspace $H_+(x)$ as a continuous function of $x$. One way out of this is to weaken the condition for the polarization: each $x \in M$ defines a Grassmann manifold $\mathrm{Gr}_{\mathrm{res}}(x)$ consisting of all subspaces $W \subset H$ such that the projections onto $W$ and $H_+(x)$ differ by Hilbert–Schmidt operators. The definition of $\mathrm{Gr}_{\mathrm{res}}(x)$ is stable with respect to finite-rank perturbations of $D_x/|D_x|$. For example, when $D_x$ is a Dirac operator on a compact manifold then $(D_x - \lambda)/|D_x - \lambda|$ defines the same Grassmannian for all real numbers $\lambda$ because in each finite interval there are only a finite number of eigenvalues (with multiplicities) of $D_x$. From this follows that the Grassmannians form a locally trivial fiber bundle Gr over families of Dirac operators.

If the bundle Gr has a global section $x \mapsto W_x$ then we can define a bundle of Fock space representations for the CAR algebra over the parameter space $M$. However, there are important situations when no global sections exist. It is easier to explain the potential obstruction in terms of a principal $U_{\mathrm{res}}$ bundle $P$ such that Gr is an associated bundle to $P$.

The fiber of $P$ at $x \in M$ is the set of all unitaries $g$ in $H$ such that $g \cdot H_+ \in \mathrm{Gr}_x$ where $H = H_+ \oplus H_-$ is a fixed reference polarization. Then we have

$$\mathrm{Gr} = P \times_{U_{\mathrm{res}}} \mathrm{Gr}_{\mathrm{res}},$$

where the right action of $U_{\mathrm{res}} = U_{\mathrm{res}}(H_+ \oplus H_-)$ in the fibers of $P$ is the right multiplication on unitary operators and the left action on $\mathrm{Gr}_{\mathrm{res}}$ comes from the observation that $\mathrm{Gr}_{\mathrm{res}} = U_{\mathrm{res}}/(U_+ \times U_-)$, where $U_\pm$ are the diagonal block matrices in $U_{\mathrm{res}}$. By a result of N Kuiper, the subgroup $U_- \times U_-$ is contractible and so Gr has a global section if and only if $P$ is trivial.

Thus, when $P$ is trivial we can define the family of Dirac representations of the CAR algebra parametrized by $M$ such that in each of the Fock spaces we have a Dirac vacuum which, in a precise sense, is close to the vacuum defined by the energy polarization. However, the triviality of $P$ is not a necessary condition. Actually, what is needed is that $P$ has a prolongation to a bundle $\hat{P}$ with fiber $\hat{U}_{\mathrm{res}}$. The group $\hat{U}_{\mathrm{res}}$ is a central extension of $U_{\mathrm{res}}$ by the group $S^1$.

The Lie algebra $\hat{\boldsymbol{u}}_{\mathrm{res}}$ is as a vector space the direct sum $\boldsymbol{u}_{\mathrm{res}} \oplus \mathrm{i}\mathbb{R}$, with commutators

$$[X + \lambda, Y + \mu] = [X, Y] + c(X, Y) \qquad [12]$$

where $c$ is the Lie algebra cocycle

$$c(X, Y) = \tfrac{1}{4}\mathrm{tr}\,\epsilon[\epsilon, X][\epsilon, Y] \qquad [13]$$

Here $\epsilon$ is the grading operator with eigenvalues $\pm 1$ on $H_\pm$. The trace exists since the off diagonal blocks of $X, Y$ are Hilbert–Schmidt.

The group $\hat{U}_{\mathrm{res}}$ is a circle bundle over $U_{\mathrm{res}}$. The Chern class of the associated complex line bundle is the generator of $H^2(U_{\mathrm{res}}, \mathbb{Z})$ and is given explicitly at the identity element as the antisymmetric bilinear form $c/2\pi\mathrm{i}$ and at other points on the group manifold through left-translation of $c/2\pi\mathrm{i}$. If $P$ is trivial, then it has an obvious prolongation to the trivial bundle $M \times \hat{U}_{\mathrm{res}}$. In any case, if the prolongation exists we can define the bundle of Fock spaces carrying CAR representations as the associated bundle

$$\mathcal{F} = \hat{P} \times_{\hat{U}_{\mathrm{res}}} \mathcal{F}_0$$

where is $\mathcal{F}_0$ is the fixed Fock space defined by the same polarization $H = H_+ \oplus H_-$ used to define $U_{\mathrm{res}}$. By the Shale–Stinespring theorem, any $g \in U_{\mathrm{res}}$ has an implementation $\hat{g}$ in $\mathcal{F}_0$, but $\hat{g}$ is only defined up to phase, thus the central $S^1$ extension.

The action of the CAR algebra in the fibers is given as follows. For $x \in M$ choose any $\hat{g} \in \hat{P}_x$. Define

$$a^*(v) \cdot (\hat{g}, \psi) = (\hat{g}, a^*(g^{-1}v)\psi)$$

where $\psi \in \mathcal{F}_0$ and $v \in H$; similarly for the operators $a(v)$. It is easy to check that this definition passes to the equivalence classes in $\mathcal{F}$. Note that the representations in different fibers are in general inequivalent because the tranformation $g$ is not implementable in the Fock space $\mathcal{F}_0$.

The potential obstruction to the existence of the prolongation of $P$ is again a 3-cohomology class on the base. Choose a good cover of $M$. On the intersections $U_{\alpha\beta}$ of the open cover the transition functions $g_{\alpha\beta}$ of $P$ can be prolonged to functions $\hat{g}_{\alpha\beta} : U_{\alpha\beta} \to \hat{U}_{\mathrm{res}}$. We have

$$\hat{g}_{\alpha\beta}\hat{g}_{\beta\gamma}\hat{g}_{\gamma\alpha} = f_{\alpha\beta\gamma} \cdot 1 \qquad [14]$$

for functions $f_{\alpha\beta\gamma} : U_{\alpha\beta\gamma} \to S^1$, which by construction satisfy the cocycle property [4]. Since the cocycle is defined on a good cover, it defines an integral Čech cohomology class $\omega \in H^3(M, \mathbb{Z})$.

Let us return to the universal $U_{\mathrm{res}}$ bundle $P$ over $G = U_1(H)$. In this case the prolongation obstruction can be computed relatively easily. It turns out that

the 3-cohomology class is represented by the de Rham class which is the generator of $H^3(G, \mathbb{Z})$. Explicitly,

$$\omega = \frac{1}{24\pi^2} \mathrm{tr}\,(g^{-1}\mathrm{d}g)^3 \qquad [15]$$

Any principal $U_{\mathrm{res}}$ bundle over $M$ comes from a pullback of $P$ with respect to a map $f: M \to G$, so the Dixmier–Douady class in the general case is the pullback $f^*\omega$.

The line bundle construction of the gerbe over the parameter space $M$ for Dirac operators is given by the observation that the spectral subspaces $E_{\lambda\lambda'}(x)$ of $D_x$, corresponding to the open interval $]\lambda,\lambda'[$ in the real line, form finite-rank vector bundles over open sets $U_{\lambda\lambda'} = U_\lambda \cap U_{\lambda'}$. Here $U_\lambda$ is the set of points $x \in M$ such that $\lambda$ does not belong to the spectrum of $D_x$. Then we can define, as top exterior power,

$$L_{\lambda\lambda'} = \bigwedge^{\mathrm{top}}(E_{\lambda\lambda'})$$

as the complex vector bundle over $U_{\lambda\lambda'}$. It follows immediately from the definition that the cocycle property [6] is satisfied.

**Example 1** (Fermions on an interval). Let $K$ be a compact group and $\rho$ its unitary representation in a finite-dimensional vector space $V$. Let $H$ be the Hilbert space of square-integrable $V$-valued functions on the interval $[0, 2\pi]$ of the real axis. For each $g \in K$ let $\mathrm{Dom}_g \subset H$ be the dense subspace of smooth functions $\psi$ with the boundary condition $\psi(2\pi) = \rho(g)\psi(0)$. Denote by $D_g$ the operator $-\mathrm{i}\,\mathrm{d}/\mathrm{d}x$ on this domain. The spectrum of $D_g$ is a function of the eigenvalues $\lambda_k$ of $\rho(g)$, consisting of real numbers $n + \log(\lambda_k)/2\pi\mathrm{i}$ with $n \in \mathbb{Z}$. For this reason the splitting of the one-particle space $H$ to positive and negative modes of the operator $D_g$ is in general not continuous as function of the parameter $g$. This leads to the problems described above. However, the principal $U_{\mathrm{res}}$ bundle can be explicitly constructed. It is the pullback of the universal bundle $P$ with respect to the map $f: K \to G$ defined by the embedding $\rho(K) \subset G$ as $N \times N$ block matrices, $N = \dim V$. Thus, the Dixmier–Douady class in this example is

$$\omega = \frac{1}{24\pi^2} \mathrm{tr}\,(\rho(g)^{-1}\ \mathrm{d}\rho(g))^3 \qquad [16]$$

**Example 2** (Fermions on a circle). Let $H = L^2(S^1, V)$ and $D_A = -\mathrm{i}(\mathrm{d}/\mathrm{d}x + A)$ where $A$ is a smooth vector potential on the circle taking values in the Lie algebra $\boldsymbol{k}$ of $K$. In this case, the domain is fixed, consisting of smooth $V$-valued functions on the circle. The $\boldsymbol{k}$-valued function $A$ is represented as a multiplication operator through the representation $\rho$ of $K$. The parameter space $\mathcal{A}$ of smooth vector potentials is flat; thus, there cannot be any obstruction to the prolongation problem. However, in quantum field theory, one wants to pass to the moduli space $\mathcal{A}/\mathcal{G}$ of gauge potentials. Here $\mathcal{G}$ is the group of smooth based gauge transformations, that is, $\mathcal{G} = \Omega K$. Now the moduli space is the group of holonomies around the circle, $\mathcal{A}/\mathcal{G} = K$. Thus, we are in a similar situation as in Example 1. In fact, these examples are really two different realizations of the same family of self-adjoint Fredholm operators. The operator $D_A$ with $k = \mathrm{holonomy}(A)$ has exactly the same spectrum as $D_k$ in Example 1. For this reason, the Dixmier–Douady class on $K$ is the same as before.

The case of Dirac operators on the circle is simple because all the energy polarizations for different vector potentials are elements in a single Hilbert–Schmidt Grassmannian $\mathrm{Gr}(H_+ \oplus H_-)$, where we can take as the reference polarization the splitting to positive and negative Fourier modes. Using this polarization, the bundle of fermionic Fock spaces over $\mathcal{A}$ can be trivialized as $\mathcal{F} = \mathcal{A} \times \mathcal{F}_0$. However, the action of the gauge group $\mathcal{G}$ on $\mathcal{F}$ acquires a central extension $\hat{\mathcal{G}} \subset \widehat{LK}$, where $LK$ is the free loop group of $K$. The Lie algebra cocycle determining the central extension is

$$c(X, Y) = \frac{1}{2\pi\mathrm{i}} \int_{S^1} \mathrm{tr}_\rho X\ \mathrm{d}Y \qquad [17]$$

where $\mathrm{tr}_\rho$ is the trace in the representation $\rho$ of $K$. Because of the central extension, the quotient $\mathcal{F}/\hat{\mathcal{G}}$ defines only a projective vector bundle over $\mathcal{A}/\mathcal{G}$, the Dixmier–Douady class being given by [16].

In the Example 1 (and Example 2) above, the complex line bundles can be constructed quite explicitly. Let us study the case $K = \mathrm{SU}(n)$. Define $U_\lambda \subset K$ as the set of matrices $g$ such that $\lambda$ is not an eigenvalue of $g$. Select $n$ different points $\lambda_j$ on the unit circle such that their product is not equal to 1. We assume that the points are ordered counterclockwise on the circle. Then the sets $U_j = U_{\lambda_j}$ form an open cover of $\mathrm{SU}(n)$. On each $U_j$ we can choose a continuous branch of the logarithmic function $\log: U_j \to \boldsymbol{su}(\boldsymbol{n})$. The spectrum of the Dirac operator $D_g$ with the holonomy $g$ consists of the infinite set of numbers $\mathbb{Z} + \mathrm{Spec}(-\mathrm{i}\log(g))$. In particular, the numbers $\mathbb{Z} - \mathrm{i}\log\lambda_j$ do not belong to the spectrum of $D_g$. Choosing $\mu_k = -\mathrm{i}\log\lambda_k$ as an increasing sequence in the interval $[0, 2\pi]$, we can as well define $U_j = \{x \in M | \mu_j \notin \mathrm{Spec}(D_x)\}$. In any case, the

top exterior power of the spectral subspace $E_{\mu_j,\mu_k}(x)$ is given by zero Fourier modes consisting of the spectral subspace of the holonomy $g$ in the segment $[\lambda_j, \lambda_k]$ of the unit circle.

## Index Theory and Gerbes

Gauge and gravitational anomalies in quantum field theory can be computed by Atiyah–Singer index theory. The basic setup is as follows. On a compact even-dimensional spin manifold $S$ (without boundary) the Dirac operators coupled to vector potentials and metrics form a family of Fredholm operators. The parameter space is the set $\mathcal{A}$ of smooth vector potentials (gauge connections) in a vector bundle over $S$ and the set of smooth Riemann metrics on $S$. The family of Dirac operators is covariant with respect to gauge transformations and diffeomorphisms of $S$; thus, we may view the Dirac operators parametrized by the moduli space $\mathcal{A}/\mathcal{G}$ of gauge connections and the moduli space $\mathcal{M}/\mathrm{Diff}_0(S)$ of Riemann metrics. Again, in order that the moduli spaces are smooth manifolds, one has to restrict to the based gauge transformations, that is, those which are equal to the neutral element in a fixed base point in each connected component of $S$. Similarly, the Jacobian of a diffeomorphism is required to be equal to the identity matrix at the base points. Passing to the quotient modulo gauge transformations and diffeomorphims, we obtain a vector bundle over the space

$$S \times \mathcal{A}/\mathcal{G} \times \mathcal{M}/\mathrm{Diff}_0(S) \qquad [18]$$

Actually, we could as well consider a generalization in which the base space is a fibering over the moduli space with model fiber equal to $S$, but for simplicity we stick to [18].

According to the Atiyah–Singer index formula for families, the K-theory class of the family of Dirac operators acting on the smooth sections of the tensor product of the spin bundle and the vector bundle $V$ over [18] is given through the differential forms

$$\hat{A}(R) \wedge \mathrm{ch}(V)$$

where $\hat{A}(R)$ is the A-roof genus, a function of the Riemann curvature tensor $R$ associated with the Riemann metric,

$$\hat{A}(R) = \det{}^{1/2}\left(\frac{R/4\pi \mathrm{i}}{\sinh(R/4\pi \mathrm{i})}\right)$$

and $\mathrm{ch}(V)$ is the Chern character

$$\mathrm{ch}(V) = \mathrm{tr}\, \mathrm{e}^{F/2\pi \mathrm{i}}$$

where $F$ is the curvature tensor of a gauge connection. Here both $R$ and $F$ are forms on the infinite-dimensional base space [18]. After integrating over the fiber $S$,

$$\mathrm{Ind} = \int_S \hat{A}(R) \wedge \mathrm{ch}(V) \qquad [19]$$

we obtain a family of differential forms $\phi_{2k}$, one in each even dimension, on the moduli space.

The (cohomology classes of) forms $\phi_{2k}$ contain important topological information for the quantized Yang–Mills theory and for quantum gravity. The form $\phi_2$ describes potential chiral anomalies. The chiral anomaly is a manifestation of gauge or reparametrization symmetry breaking. If the class $[\phi_2]$ is nonzero, the quantum effective action cannot be viewed as a function on the moduli space. Instead, it becomes a section of a complex line bundle DET over the moduli space.

Since the Dirac operators are Fredholm (on compact manifolds), at a given point in the moduli space we can define the complex line

$$\mathrm{DET}_x = \bigwedge^{\mathrm{top}}(\ker D_x^+) \otimes \bigwedge^{\mathrm{top}}(\mathrm{coker}\, D_x^+) \qquad [20]$$

for the chiral Dirac operators $D_x^+$. In the even-dimensional case, the spin bundle is $\mathbb{Z}_2$ graded such that the grading operator $\Gamma$ anticommutes with $D_x$. Then $D_x^+ = P_- D_x P_+$, where $P_\pm = (1/2)(1 \pm \Gamma)$ are the chiral projections. $\wedge^{\mathrm{top}}$ means the operation on finite-dimensional vector spaces $W$ taking the exterior power of $W$ to dim $W$.

When the dimensions of the kernel and cokernel of $D_x$ are constant, eqn [20] defines a smooth complex line bundle over the moduli space. In the case of varying dimensions, a little extra work is needed to define the smooth structure.

The form $\phi_2$ is the Chern class of DET. So if DET is nontrivial, gauge covariant quantization of the family of Dirac operators is not possible.

One can also give a geometric and topological meaning to the chiral symmetry breaking in Hamiltonian quantization, and this leads us back to gerbes on the moduli space. Here we have to use an odd version of the index formula [19]. Assuming that the physical spacetime is even dimensional, at a fixed time the space is an odd-dimensional manifold $S$. We still assume that $S$ is compact. In this case, the integration in [19] is over odd-dimensional fibers and, therefore, the formula produces a sequence of odd forms on the moduli space.

The first of the odd forms $\phi_1$ gives the spectral flow of a one-parameter family of operators $D_{x(s)}$. Its integral along the path $x(t)$, after a correction by the difference of the eta invariant at the end points

of the path, in the moduli space, gives twice the difference of positive eigenvalues crossing over to the negative side of the spectrum minus the flow of eigenvalues in the opposite direction. The second term $\phi_3$ is the Dixmier–Douady class of the projective bundle of Fock spaces over the moduli space. In Examples 1 and 2, the index theory calculation gives exactly the form [16] on $K$.

**Example** Consider Dirac operators on the three-dimensional sphere $S^3$ coupled to vector potentials. Any vector bundle on $S^3$ is trivial, so let $V = S^3 \times \mathbb{C}^N$. Take SU($N$) as the gauge group and let $\mathcal{A}$ be the space of 1-forms on $S^3$ taking values in the Lie algebra $\boldsymbol{su}(N)$ of SU($N$). Fix a point $x_s$ on $S^3$, the "south pole," and let $\mathcal{G}$ be the group of gauge transformations based at $x_s$. That is, $\mathcal{G}$ consists of smooth functions $g: S^3 \to \mathrm{SU}(N)$ with $g(x_s) = 1$. In this case $\mathcal{A}/\mathcal{G}$ can be identified as $\mathrm{Map}(S^2, \mathrm{SU}(N))$ times a contractible space. This is because any point $x$ on the equator of $S^3$ determines a unique semicircle from the south pole to the north pole through $x$. The parallel transport along this path with respect to a vector potential $A \in \mathcal{A}$ defines an element $g'_A(x) \in \mathrm{SU}(N)$, using the fixed trivialization of $V$. Set $g_A(x) = g'_A(x) g'_A(x_0)^{-1}$, where $x_0$ is a fixed point on the equator. The element $g_A(x)$ then depends only on the gauge equivalence class $[A] \in \mathcal{A}/\mathcal{G}$. It is not difficult to show that the map $A \mapsto g_A$ is a homotopy equivalence from the moduli space of gauge potentials to the group $\mathcal{G}_2 = \mathrm{Map}_{x_0}(S^2, \mathrm{SU}(N))$, based at $x_0$. When $N > 2$, the cohomology $H^5(\mathrm{SU}(N), \mathbb{Z}) = \mathbb{Z}$ transgresses to the cohomology $H^3(\mathcal{G}_2, \mathbb{Z}) = \mathbb{Z}$. In particular, the generator

$$\omega_5 = \left(\frac{\mathrm{i}}{2\pi}\right)^3 \frac{2}{5!} \mathrm{tr}(g^{-1}\,\mathrm{d}g)^5$$

of $H^5(\mathrm{SU}(N), \mathbb{Z})$ gives the generator of $H^3(\mathcal{G}_2, \mathbb{Z})$ by contraction and integration,

$$\Omega = \int_{S^2} \omega_5$$

## Gauge Group Extensions

The new feature for gerbes associated with Dirac operators in higher than one dimension is that the gauge group, acting on the bundle of Fock spaces parametrized by vector potentials, is represented through an abelian extension. On the Lie algebra level this means that the Lie algebra extension is not given by a scalar cocycle $c$ as in the one-dimensional case but by a cocycle taking values in an abelian Lie algebra. In the case of Dirac operators coupled to vector potentials, the abelian Lie algebra consists of a certain class of complex functions on $\mathcal{A}$. The extension is then defined by the commutators

$$[(X,\alpha),(Y,\beta)] = ([X,Y], \mathcal{L}_X\beta - \mathcal{L}_Y\alpha + c(X,Y)) \quad [21]$$

where $\alpha, \beta$ are functions on $\mathcal{A}$ and $\mathcal{L}_X\beta$ denotes the Lie derivative of $\beta$ in the direction of the infinitesimal gauge transformation $X$. The 2-cocycle property of $c$ is expressed as

$$c([X,Y],Z) + \mathcal{L}_X c(Y,Z) + \text{cyclic permutations of } X, Y, Z = 0$$

In the case of Dirac operators on a 3-manifold $S$ the form $c$ is the Mickelsson–Faddeev cocycle

$$c(X,Y) = \frac{i}{12\pi^2} \int_S \mathrm{tr}_\rho A \wedge (\mathrm{d}X \wedge \mathrm{d}Y - \mathrm{d}Y \wedge \mathrm{d}X) \quad [22]$$

The corresponding gauge group extension is an extension of $\mathrm{Map}(S, G)$ by the normal subgroup $\mathrm{Map}(\mathcal{A}, S^1)$. As a topological space, the extension is the product

$$\mathrm{Map}(\mathcal{A}, S^1) \times_{S^1} P$$

where $P$ is a principal $S^1$ bundle over $\mathrm{Map}(S, G)$.

The Chern class $c_1$ of the bundle $P$ is again computed by transgression from $\omega_5$; this time

$$c_1 = \int_S \omega_5$$

In fact, we can think of the cocycle $c$ as a 2-form on the space of flat vector potentials $A = g^{-1}\mathrm{d}g$ with $g \in \mathrm{Map}(S^3, G)$. Then one can show that the cohomology classes $[c]$ and $[c_1]$ are equal.

As we have seen, the central extension of a loop group is the key to understanding the quantum field theory gerbe. Here is a brief description of it starting from the 3-form [16] on a compact Lie group $G$. First define a central extension $\mathrm{Map}(D, G) \times S^1$ of the group of smooth maps from the unit disk $D$ to $G$, with pointwise multiplication. The group multiplication is given as

$$(g, \lambda) \cdot (g', \lambda') = (gg', \lambda\lambda' \cdot \mathrm{e}^{2\pi \mathrm{i}\gamma(g,g')})$$

where

$$\gamma(g, g') = \frac{1}{8\pi^2} \int_D \mathrm{tr}_\rho\, g^{-1}\mathrm{d}g \wedge \mathrm{d}g' g'^{-1} \quad [23]$$

where the trace is computed in a fixed unitary representation $\rho$ of $G$. This group contains as a normal subgroup the group $N$ consisting of pairs $(g, \mathrm{e}^{2\pi \mathrm{i} C(g)})$ with

$$C(g) = \frac{1}{24\pi^2} \int_B \mathrm{tr}_\rho\, (g^{-1}\mathrm{d}g)^3 \quad [24]$$

Here $g(x) = 1$ on the boundary circle $S^1 = \partial D$, and thus can be viewed as a function $S^2 \to G$. The three-dimensional unit ball $B$ has $S^2$ as a boundary and $g$ is extended in an arbitrary way from the boundary to the ball $B$. The extension is possible since $\pi_2(G) = 0$ for any finite-dimensional Lie group. The value of $C(g)$ depends on the extension only modulo an integer and therefore $e^{2\pi i C(g)}$ is well defined.

The central extension is then defined as

$$\widehat{LG} = (\mathrm{Map}(D, G) \times S^1)/N$$

One can show easily that the Lie algebra of $\widehat{LG}$ is indeed given through the cocycle [17]. When $G = \mathrm{SU}(n)$ in the defining representation, this central extension is the basic extension: The cohomology class is the generator of $H^2(LG, \mathbb{Z})$. In general, to obtain the basic extension one has to correct [23] and [24] by a normalization factor.

This construction generalizes to the higher loop groups $\mathrm{Map}(S, G)$ for compact odd-dimensional manifolds $S$. For example, in the case of a 3-manifold, one starts from an extension of $\mathrm{Map}(D, G)$, where $D$ is a 4-manifold with boundary $S$. The extension is defined by a 2-cocycle $\gamma$, but now for given $g, g'$ the cocycle $\gamma$ is a real-valued function of a point $g_0 \in \mathrm{Map}(S, G)$, which is a certain differential polynomial in the Maurer–Cartan 1-forms $g_0^{-1} dg_0, g^{-1} dg, g^{-1} dg$. The normal subgroup $N$ is defined in a similar way; now $C(g)$ is the integral of the 5-form $\omega_5$ over a 5-manifold $B$ with boundary $\partial B$ identified as $D/\sim$, the equivalence shrinking the boundary of $D$ to one point. This gives the extension only over the connected component of identity in $\mathrm{Map}(S, G)$, but it can be generalized to the whole group. For example, when $S = S^3$ and $G$ is simple, the connected components are labeled by elements of the third homotopy group $\pi_3 G = \mathbb{Z}$.

In some cases, the de Rham cohomology class of the extension vanishes but the extension still contains interesting torsion information. In quantum field theory this comes from Hamiltonian formulation of global anomalies. A typical example of this phenomenon is the Witten SU(2) anomaly in four spacetime dimensions. In the Hamiltonian formulation, we take $S^3$ as the physical space, the gauge group $G = \mathrm{SU}(2)$. In this case, the second cohomology of $\mathrm{Map}(S^3, G)$ becomes pure torsion, related to the fact that the 5-form $\omega_5$ on SU(2) vanishes for dimensional reasons. Here the homotopy group $\pi_4(G) = \mathbb{Z}_2$ leads the nontrivial fundamental group $\mathbb{Z}_2$ in each connected component of $\mathrm{Map}(S^3, G)$. Using this fact, one can show that there is a nontrivial $\mathbb{Z}_2$ extension of the group $\mathrm{Map}(S^3, G)$.

*See also:* Anomalies; Bosons and Fermions in External Fields; Characteristic Classes; Dirac Operator and Dirac Field; Index Theorems; *K*-Theory.

## Further Reading


Araki H (1987) Bogoliubov automorphisms and Fock representations of canonical anticommutation relations. *Contemporary Mathematics*, vol. 62. Providence: American Mathematical Society.

Berline N, Getzler E, and Vergne M (1992) *Heat Kernels and Dirac Operators*, Die Grundlehren der mathematischen Wissenschaften, vol. 298. Berlin: Springer.

Brylinski J-L *Loop Spaces, Characteristic Classes, and Geometric Quantization,* Progress in Mathematics, vol. 107. Boston: Birkhäuser.

Carey AL, Mickelsson J, and Murray M (2000) Bundle gerbes applied to quantum field theory. *Reviews in Mathematical Physics* 12: 65–90.

Gawedzki K and Reis N (2002) WZW branes and gerbes. *Reviews in Mathematical Physics* 14: 1281–1334.

Giraud J (1971) *Cohomologie Non Abelienne,* (French) Die Grundlehren der Mathematischen Wissenschaften, vol. 179. Berlin: Springer.

Hitchin N (2001) Lectures on special Lagrangian submanifolds. In: *Winter School on Mirror Symmetry, Vector Bundles and Lagrangian Submanifolds (Cambridge, MA, 1999)*, AMS/IP Stud. Adv. Math., vol. 23, pp. 151–182. Providence, RI: American Mathematical Society.

Mickelsson J (1989) *Current Algebras and Groups*. London: Plenum.

Treiman SB, Jackiw R, Zumino B, and Witten E (1985) *Current Algebra and Anomalies*, Princeton Series in Physics. Princeton, NJ: Princeton University Press.

# Ginzburg–Landau Equation

**Y Morita**, Ryukoku University, Otsu, Japan



## Introduction

In the Ginzburg–Landau theory of superconductivity, a complex order parameter $\Psi$ characterizes a macroscopic/mesoscopic superconducting state in a bulk superconductor. The square of the magnitude $|\Psi|^2$ expresses the density of superconducting electrons and $\Psi$ is regarded as a macroscopic wave function. With a magnetic vector potential $\mathbb{A}$ and the order parameter $\Psi$, the Helmholtz free energy density in a superconducting material near the critical temperature is given by

$$F = F_n + \alpha|\Psi|^2 + \frac{\beta}{2}|\Psi|^4 + \frac{1}{2m_s}\left|\left(-i\hbar\nabla - \frac{e_s}{c}\mathbb{A}\right)\Psi\right|^2 + \frac{|\mathbb{H}|^2}{8\pi}$$

where $F_n$ denotes the energy density of the normal state, $c$ is the light speed, $\mathbb{H} = \mathrm{curl}\ \mathbb{A}$, and $m_s$ and $e_s$ are mass and charge of a superconducting electron, respectively. The parameters $\alpha$ and $\beta$ depend on temperature and are determined by the material. Moreover, below the critical temperature $T_c$, $\alpha = \alpha(T)$ and $\beta = \beta(T)$ take negative and positive values, respectively. In the presence of an applied magnetic field $\mathbb{H}_{ap}$, we have to consider the Gibbs free energy density, $G = F - \mathbb{H}\cdot\mathbb{H}_{ap}/4\pi$.

Introduce the following physical parameters:

$$\Psi_0 = \sqrt{-\alpha/\beta}, \qquad H_c = \sqrt{4\pi\alpha^2/\beta}$$
$$\lambda = \sqrt{-\beta m_s c^2/4\pi\alpha e_s^2}, \qquad \xi = \sqrt{-\hbar^2/2m_s\alpha} \qquad [1]$$
$$\kappa = \lambda/\xi$$

The value $\Psi_0^2$ implies the equilibrium density and $H_c$ is the thermodynamic critical field, which is obtained by equating $G = F_n - |\mathbb{H}_{ap}|^2/8\pi$ (for the normal state $\Psi = 0$, $\mathbb{H} = \mathbb{H}_{ap}$) with $G = F_n - \alpha^2/2\beta$ (for the perfect superconductivity $|\Psi|^2 = \Psi_0^2$, $\mathbb{A} = 0$). The parameters $\lambda$ and $\xi$ stand for penetration depth and coherence length, respectively. The ratio $\kappa$ of these characteristic lengths is called the Ginzburg–Landau parameter, which determines the type of superconducting material: type I for $\kappa < 1/\sqrt{2}$ and type II for $\kappa > 1/\sqrt{2}$.

We use the nondimensional variables $x', \Psi', \mathbb{A}'$, $\mathbb{H}_{ap}'$, and $\tilde{G}$:

$$x = \lambda x', \qquad \Psi = \Psi_0\Psi'$$
$$\mathbb{A} = \sqrt{2}H_c\xi\mathbb{A}' \ (\mathbb{H}' = \mathrm{curl}'\mathbb{A}'),$$
$$\mathbb{H}_{ap} = \sqrt{2}H_c\mathbb{H}_{ap}'/\kappa \qquad [2]$$
$$F = F_n + (\tilde{G}/\kappa^2 - 1/2 + 2\mathbb{H}'\cdot\mathbb{H}_{ap}'/\kappa^2 - |\mathbb{H}_{ap}'|^2/\kappa^2)H_c^2/4\pi$$

Dropping the primes after the change of variables and integrating $\tilde{G}$ over a domain $\Omega \subset \mathbb{R}^n (n = 2, 3)$, which is occupied by a superconducting sample, yields a functional of $\Psi$ and $\mathbb{A}$, called the Ginzburg–Landau energy in a nondimensional form,

$$E(\Psi, \mathbb{A}) = \int_\Omega \left\{ |(\nabla - i\mathbb{A})\Psi|^2 + \frac{\kappa^2}{2}(1 - |\Psi|^2)^2 + |\mathrm{curl}\ \mathbb{A} - \mathbb{H}_{ap}|^2 \right\} dx \qquad [3]$$

The Ginzburg–Landau equations are the Euler–Lagrange equations of this energy, which are given by

$$(\nabla - i\mathbb{A})^2\Psi = \kappa^2(|\Psi|^2 - 1) \quad \text{in } \Omega \qquad [4]$$

$$\mathrm{curl}^2\ \mathbb{A} = \mathbb{J} + \mathrm{curl}\ \mathbb{H}_{ap} \quad \text{in } \Omega \qquad [5]$$

where

$$\mathbb{J} := \frac{1}{2i}(\Psi^*\nabla\Psi - \Psi\nabla\Psi^*) - |\Psi|^2\mathbb{A} \qquad [6]$$

$\Psi^*$ stands for the complex conjugate of $\Psi$. In a two-dimensional domain $\Omega$, the differential operator "curl" acts on $\mathbb{A} = (A_1, A_2) : \mathbb{R}^2 \to \mathbb{R}^2$ such that

$$\mathrm{curl}\ \mathbb{A} = \partial_{x_1}A_2 - \partial_{x_2}A_1$$
$$\mathrm{curl}\ H = (\partial_{x_2}H, -\partial_{x_1}H)$$
$$H := \mathrm{curl}\ \mathbb{A}$$

and $\mathbb{H}_{ap}$ is replaced by a scalar-valued function. Note that $\mathbb{J}$ represents a supercurrent in the material. Every critical point of the energy is obtained by solving the Ginzburg–Landau equations with appropriate boundary conditions and, thus, a physical state in the superconducting sample is realized by a solution of the equations. A minimizer of [3] is a solution of [4]–[5] that minimizes the energy [3] in an appropriate function space, whereas a local minimizer is a solution minimizing the energy locally in the space. A solution is called a stable solution if it is a local minimizer of the energy. A physically stable phenomenon could be realized by a minimizer or at least a local minimizer.

The Ginzburg–Landau energy and the equations are gauge invariant under the transformation

$$(\Psi, \mathbb{A}) \longmapsto (\Psi e^{i\chi}, \mathbb{A} + \nabla\chi) \quad [7]$$

for a smooth scalar function $\chi(x)$. Therefore, we can identify two solutions which have the correspondence through the transformation [7]. The following London (Coulomb) gauge is often chosen:

$$\operatorname{div} \mathbb{A} = 0 \quad \text{in } \Omega \quad [8]$$

(with a boundary condition if necessary).

Let $(\Psi, \mathbb{A})$ be a smooth solution of [4]–[5]. In a region for $|\Psi(x)| > 0$, the expression $\Psi = w(x) \exp(i\theta(x))$ $(w = |\Psi(x)|)$ leads to

$$\nabla^2 w = |\nabla\theta - \mathbb{A}|^2 w + \kappa^2(w^2 - 1)w \quad [9]$$

$$\operatorname{div}(w^2(\nabla\theta - \mathbb{A})) = 0 \quad [10]$$

$$\operatorname{curl}^2 \mathbb{A} = \mathbb{J} = w^2(\nabla\theta - \mathbb{A}) \quad [11]$$

where the gauge [8] is fixed and curl $\mathbb{H}_{ap} = 0$ is assumed. Let $S$ be a surface in $\Omega$ bounded by a closed curve $\partial S$. Suppose $w(x) > 0$ on $\partial S$. Then from [11],

$$\begin{aligned}\Phi &:= \int_{\partial S} (\mathbb{J}/w^2 + \mathbb{A}) \cdot d\mathbf{s} \\ &= \int_{\partial S} \frac{1}{w^2} \mathbb{J} \cdot d\mathbf{s} + \int_S \operatorname{curl} \mathbb{A} \cdot d\mathbf{S} \\ &= \int_{\partial S} \nabla\theta \cdot d\mathbf{s} = 2d\pi \end{aligned} \quad [12]$$

where $d$ is an integer; in fact, $d = \deg(\Psi, \partial S)$ is the winding number of $\Psi(\partial S)$ in the complex plane. Thus, the identity [12] relates the magnetic field to a topological degree of the order parameter. The quantity $\Phi$, multiplied by an appropriate constant, is called the fluxoid. A connected component of vanishing points of $\Psi$ generally has codimension 2 in the domain, and it is called a vortex.

From the expression [9], the asymptotic behavior $w \to 1$ as $\kappa \to \infty$ is expected under a suitable condition. Then, by [11], $\mathbb{H} = \operatorname{curl} \mathbb{A}$ enjoys the property $\operatorname{curl}^2 \mathbb{H} + \operatorname{curl} \mathbb{H} = 0$, which is known as the London equation. However, this is valid for $|\Psi| > 0$. Otherwise, a singularity appears around zeros of $\Psi$.

There are several characteristic phenomena observed in a bulk superconductor. Typical phenomena are: perfect conductivity (persistent current), perfect diamagnetism (Meissner effect), nucleation of superconductivity, and vortices (quantization of a penetrating magnetic field). These phenomena can be expressed by solutions of the Ginzburg–Landau equations in various settings.

## Ginzburg–Landau Equations in $\mathbb{R}^2$

A standard model of the Ginzburg–Landau energy is considered in the whole space $\mathbb{R}^2$. Let $\mathbb{A} = (A_1, A_2)$ and assume $\mathbb{H}_{ap} = 0$ in [3]. Consider then the energy functional

$$\begin{aligned}\mathcal{E}(\Psi, A) = \int_{\mathbb{R}^2} &|D_A\Psi|^2 + |\operatorname{curl} \mathbb{A}|^2 \\ &+ \frac{\kappa^2}{2}(1 - |\Psi|^2)^2 \, dx \end{aligned} \quad [13]$$

where $D_A := \nabla - i\mathbb{A}$. Then the Ginzburg–Landau equations are

$$D_A^2 \Psi = \kappa^2(|\Psi|^2 - 1)\Psi \quad \text{in } \mathbb{R}^2 \quad [14]$$

$$\operatorname{curl}^2 \mathbb{A} = \operatorname{Im}(\Psi^* D_A \Psi) \quad \text{in } \mathbb{R}^2 \quad [15]$$

In the gauge theory, this model can be regarded as a two-dimensional abelian (U(1)) Higgs model. In that context, $\Psi$ is a scalar (Higgs) field, $\mathbb{A}$ is a connection on the U(1) bundle $\mathbb{R}^2 \times U(1)$, and $D_A$ is the covariant derivative.

Equations [14]–[15] are useful in observing quantization of the magnetic field, although it is an ideal model for superconductivity. By the natural condition that the right-hand side of [13] is finite, we may assume that $|D_A\Psi|, |\operatorname{curl} \mathbb{A}| \to 0$ and $|\Psi| \to 1$ as $|x| \to \infty$. From [12], the flux quantization follows:

$$\int_{\mathbb{R}^2} \operatorname{curl} \mathbb{A} \, dx = 2d\pi \quad [16]$$

If $\Psi$ has a finite number of zeros $\{a_j\}_{j=1}^N$, [16] implies

$$\int_{\mathbb{R}^2} \operatorname{curl} \mathbb{A} \, dx = 2\pi \sum_{j=1}^N \deg(\Psi, \partial B(a_j, \rho))$$

for a small positive number $\rho$, where $B(a_j, \rho)$ stands for the disk with the center $a_j$ and the radius $\rho$. A zero of $\Psi$ represents a vortex, at which the magnetic field is quantized, and a supercurrent moves around the field.

To characterize the configuration analytically, we find a solution $(\Psi, \mathbb{A})$ expressed by the polar coordinate in the form

$$\Psi = f(r) \exp(id\theta), \qquad \mathbb{A}(r) = \alpha(r)(-\sin\theta, \cos\theta)$$

Substituting these into [14]–[15], one obtains

$$\frac{1}{r}(rf')' - \left(\frac{d}{r} - \alpha\right)^2 f = \kappa^2(f^2 - 1)f$$

$$\left(\frac{1}{r}(r\alpha)'\right)' = f^2\left(\alpha - \frac{d}{r}\right)$$

$(' = d/dr)$ with the boundary conditions

$$f(0) = 0, \quad f(\infty) = 1, \quad \alpha(\infty) = 0$$

This system of the equations has a solution for $\kappa > 0$. In addition to these types of solutions, when $\kappa = 1/\sqrt{2}$, a special transformation reduces the system of [14]–[15] to a scalar nonlinear equation with a singular term. Then, it is proved that for an arbitrary $d \in \mathbb{Z}$, under the constraint of [16] there exists a minimizer of [13] with zeros of prescribed points $\{a_j\}_{j=1}^{|d|}$ (Jaffe and Taubes 1980).

## Solutions for Persistent Current

A current flowing in a superconducting ring with no decay even in the absence of an applied magnetic field is called a persistent current. Assume that a superconducting sample $\Omega$ in $\mathbb{R}^3$ is surrounded by vacuum and adopt the energy functional as

$$\mathcal{E}(\Psi, \mathbb{A}) = \int_\Omega |D_A\Psi|^2 + \frac{\kappa^2}{2}(1 - |\Psi|^2)^2 \, dx + \int_{\mathbb{R}^3} |\text{curl}\, \mathbb{A}|^2 \quad [17]$$

Although the functional [17] is minimized by a trivial solution $(\Psi, \mathbb{A}) = (\exp(ic), 0)(c \in \mathbb{R})$, which is the case for perfect diamagnetism, this is not the solution describing a persistent current since $\mathbb{J} = 0$ everywhere. We have to look for a nontrivial solution that locally minimizes the energy, that is, a local minimizer of [17]. To characterize a solution representing the persistent current, we define a mapping from $\Omega$ to $S^1 \subset \mathbb{C}$ by $x \in \Omega \to \Psi(x)/|\Psi(x)|$ for a solution $(\Psi, \mathbb{A})$ of the corresponding Ginzburg–Landau equations to [17]. Consider a domain having infinitely many homotopy classes in the space of continuous functions $C^0(\bar{\Omega}, S^1)$ (e.g., a solid torus). If $(\Psi, \mathbb{A})$ is a local minimizer and $\Psi/|\Psi|$ is not homotopic to a constant map of $C^0(\bar{\Omega}, S^1)$, then it is a solution describing a persistent current. The existence of such a solution has been established mathematically for large $\kappa$ (Jimbo and Morita 1996, Rubinstein and Sternberg 1996).

## Configuration of Solutions under an Applied Magnetic Field

In the presence of an applied magnetic field, according to the magnitude of the field, a sample exhibits the transition from the superconducting state to the normal state and vice versa. This transition can be considered mathematically as a bifurcation of solutions to the Ginzburg–Landau equations with a parameter measuring the magnitude of the applied magnetic field. In fact, let $\mathbb{H}_{ap}$ be an applied magnetic field perpendicular to the horizontal plane and assume that it is constant along the vertical axis, that is, $\mathbb{H}_{ap} = (0, 0, H_a)$. Then a rich bifurcation structure is suggested by numerical and analytical studies in the parameter space of $(H_a, \kappa)$. Mathematical developments for variational methods and nonlinear analysis reveal the configuration of the solutions and provide rigorous estimates for critical fields in a parameter regime for a two-dimensional model, predicted by physicists.

Throughout this section, we consider the Ginzburg–Landau model in an infinite cylinder $\Omega = D \times \mathbb{R}$ $(D \subset \mathbb{R}^2)$ with a constant applied magnetic field $\mathbb{H}_{ap} = H_a \boldsymbol{e}_3 = (0, 0, H_a), H_a > 0$. Assuming the uniformity along the vertical axis, we may write $\mathbb{A} = (A_1, A_2)$ and $H = \text{curl}\,\mathbb{A} = \partial_{x_1}A_2 - \partial_{x_2}A_1$ as in the previous section. Then the Ginzburg–Landau energy on $D$ is

$$\mathcal{E}(\Psi, \mathbb{A}) = \int_D \left\{ |D_A\Psi|^2 + \frac{\kappa^2}{2}(1 - |\Psi|^2)^2 + |\text{curl}\,\mathbb{A} - H_a|^2 \right\} dx \quad [18]$$

With the London gauge

$$\text{div}\,\mathbb{A} = 0 \quad \text{in } D, \qquad \mathbb{A} \cdot \boldsymbol{n} = 0 \quad \text{on } \partial D$$

the Ginzburg–Landau equations in the present setting are written as

$$D_A^2\Psi = (|\Psi|^2 - 1)\Psi \quad \text{in } D \quad [19]$$

$$-\nabla^2\mathbb{A} = \text{Im}(\Psi^* D_A\Psi) \quad \text{in } D \quad [20]$$

$$\boldsymbol{n} \cdot \nabla\Psi = 0 \quad \text{on } \partial D \quad [21]$$

$$\text{curl}\,\mathbb{A} = \text{H}_a \quad \text{on } \partial \text{D} \quad [22]$$

where $\boldsymbol{n}$ denotes the outer unit normal.

### Meissner Solutions

As seen in the case of no applied magnetic field, the trivial solution $(\Psi, \mathbb{A}) = (\exp(ic), 0)$ is a minimizer of [18]. This solution expresses no magnetic field in the sample. In a superconducting sample, the diamagnetism holds even in the presence of an applied magnetic field if the field is weak. Namely, the sample is shielded so that penetration of the field is only allowed near the surface of the sample. This phenomenon is called the Meissner effect. A solution expressing Meissner effect is called a Meissner solution. Mathematically, it is understood that as $H_a$ increases, such a Meissner solution continues from the trivial solution. Then the solution preserves the configuration $0 < |\Psi(x)| < 1$. A study of the asymptotic behavior of the Meissner solution as $\kappa$ tends to $\infty$ shows that the Meissner solution is a

minimizer up to $H_a = O(\log \kappa)$ for sufficiently large $\kappa$ (Serfaty 1999).

### Nucleation of Superconductivity

In an experiment, the Meissner state breaks down by a stronger applied magnetic field. Then the sample turns to be the normal state (in a type I conductor) or it allows a mixed state of superconductivity and normal state (in a type II conductor). In the former case, the critical magnitude of the field is denoted by $H_c$, which corresponds to the one of [1], while it is denoted by $H_{c_1}$ in the latter case. Moreover, the mixed state eventually breaks down to be normal state by further increasing the applied field up to another critical field $H_{c_2}$. To characterize these two types mathematically, we consider a transition from the normal state to the superconducting state by reducing the magnitude of the field.

Let $\mathbb{A}_{ap}$ satisfy curl $\mathbb{A}_{ap} = H_a (x \in D)$ and $\mathbb{A}_{ap} \cdot \boldsymbol{n} = 0 (x \in \partial D)$. Then eqns [19]–[22] have a trivial solution $(\Psi, \mathbb{A}) = (0, \mathbb{A}_{ap})$, which stands for the normal state. Consider the second variation of the energy functional [18] at this trivial solution

$$\frac{1}{2}\frac{d^2}{ds^2}E(s\psi, \mathbb{A}_{ap} + s\mathbb{B})\bigg|_{s=0} = \int_D |(\nabla - i\mathbb{A}_{ap})\psi|^2 - \kappa^2|\psi|^2 + |\text{curl}\,\mathbb{B}|^2 dx$$

If the minimum of this second variation for nonzero $(\psi, \mathbb{B})$ is positive (or negative), then the trivial solution is stable (or unstable). The minimum gives the least eigenvalue of the linearized problem of [19]–[20] around the trivial solution. Seeking such a least eigenvalue $\mu$ is reduced to studying an eigenvalue problem of the Schrödinger operator $L[\psi] := -(\nabla - i\mathbb{A}_{ap})^2\psi$.

If the domain $D$ is the whole space $\mathbb{R}^2$, it is proved that $\mu = H_a$. Back to the original variable of [2], we can define a critical field $H_{c_2} = \sqrt{2}H_c\kappa$; $\kappa = 1/\sqrt{2}$ separates a class of superconductors into type I by $\kappa < 1/\sqrt{2}$ $(H_{c_2} < H_c)$ and type II by $\kappa > 1/\sqrt{2}$ $(H_{c_2} > H_c)$.

In the bounded domain $D$, however, the critical field at which superconductivity nucleates in the interior of a sample is larger than $H_{c_2}$ (it is denoted by $H_{c_3}$), since the eigenvalue problem of $L$ is considered in the domain with the Neumann boundary condition. A study of the least eigenvalue $\mu$ shows that the critical field has the asymptotics as

$$H_{c_3}/\sqrt{2}H_c = \frac{\kappa}{\beta} + O(1), \quad \kappa \to \infty$$

where $0 < \beta < 1$. If the applied field is very close to $H_{c_3}$ and $\kappa$ is sufficiently large, the amplitude of the eigenfunction associated with the least eigenvalue of $L$ (with the Neumann boundary condition) is very small except for a $1/\kappa$ neighborhood of the boundary. This implies that the nucleation of superconductivity takes place at the boundary. This phenomenon is called surface nucleation (Del Pino *et al.* 2000, Lu and Pan 1999).

### Solutions of Vortices

In a type II superconductor, it is well known that there exists a mixed state of superconductivity and normal state in a parameter regime $H_{c_1} < H_a < H_{c_2}$. In the mixed state, the magnetic field penetrating in the sample is quantized such that it delivers a finite number of lines or curves in the sample. This configuration (called vortex) is characterized by zero sets of the order parameter of the Ginzburg–Landau equations. In a two-dimensional domain, isolating vanishing points of the order parameter are called vortices. Thus, it is quite an interesting problem how such a vortex configuration can be described mathematically by a minimizer of the energy functional. In the section "Ginzburg–Landau equations in $\mathbb{R}^2$," a specific configuration for vortex solutions is stated under very special conditions, $\kappa = 1/\sqrt{2}$, on the whole space and no applied magnetic field. However, this result is not generalized in the present setting.

A standard approach to a solution with the vortex configuration is using a bifurcation analysis near the critical field $H_{c_2}$ (or $H_{c_3}$) by expanding a solution and the difference $H_a - H_{c_2}$ in a small parameter. Then the leading term is given by an eigenfunction of the least eigenvalue of the Schrödinger operator coming from the linearization. Under the doubly periodic conditions in the whole space $\mathbb{R}^2$, the spatial pattern of vortices, called Abrikosov's vortex lattice, is studied by a local bifurcation theory.

However, this kind of bifurcation analysis only works near the critical field and the trivial solution $(\Psi, \mathbb{A}) = (0, \mathbb{A}_{ap})$, which implies that only a small-amplitude solution can be found. To realize a sharp configuration of vortices, we need to consider a parameter regime far from the bifurcation point. As a matter of fact, mathematical and numerical studies for sufficiently large $\kappa$ exhibit nice configurations of vortex solutions. In this case, in a neighborhood of each vortex, with radius $O(1/\kappa)$, a sharp layer arises, and there exists a solution with multivortices in an appropriate parameter region for $H_a$. In addition, as $H_a$ increases (up to $H_{c_2}$), the number of vortices also increases. This implies that the minimizer of the energy functional [18] admits a larger number of zeros for a higher magnitude of applied magnetic field. However, it is a puzzle since

a solution with a smaller number of vortices seems to have less energy. Thus, there is some balance mechanism between contributions of the vortices and the applied magnetic field to the total energy.

Mathematically, it is possible to estimate $\mathcal{E}(\Psi, \mathbb{A})$ for the vortex solution to [19]–[22] as follows: consider a family of square tiles $K_j$ with side-length $\rho$ which are periodically arranged over the whole space. Assume each square in the domain $D$ has a single vortex. For an appropriate test function, the energy over $K_j$ is estimated as $O(\log(\kappa\rho))$. Since the number of vortices in the domain is $O(|D|/\rho^2)$ ($|D|$: the measure of $D$), we obtain an upper bound $O((|D|/\rho^2)\log(\kappa\rho))$. This bound is less than $\mathcal{E}(0, \mathbb{A}_{\mathrm{ap}}) = |D|\kappa^2/2$ for $H_{\mathrm{a}}/\kappa^2 = o(1)$ and $\rho = 1/\sqrt{H_{\mathrm{a}}}$. Although in a general case it is difficult to estimate the energy of the minimizer from below, the leading order can be precisely determined in some range of the interval $(H_{c_1}, H_{c2})$ if $\kappa$ is sufficiently large (Sandier and Serfaty 2000).

## A Simplified Model

Since the Ginzburg–Landau equations [4]–[5] are coupled equations for $\Psi$ and $\mathbb{A}$, we often encounter mathematical difficulty in realizing a solution with the configuration shown by a numerical experiment. To look at a specific configuration, we may use a simpler model equation. A typical simplification is to neglect the magnetic field, which leads to the equation for the order parameter $\psi$:

$$\nabla^2\psi + \kappa^2(1 - |\psi|^2)\psi = 0 \quad \text{in } \Omega \qquad [23]$$

This equation is also called the Ginzburg–Landau equation and it is the Euler–Lagrange equation of the energy

$$G(\psi) = \int_\Omega |\nabla\psi|^2 + \frac{\kappa^2}{2}(1 - |\psi|^2)^2\, \mathrm{d}x \qquad [24]$$

in an appropriate function space. Under no constraint, a constant solution with $|\psi| = 1$ is a minimizer. If a domain is topologically nontrivial, eqn [23] also allows local minimizers of [24] for large $\kappa$ as seen in the section "Solutions for Persistent Current."

On the other hand, [23] in a simply connected domain $D \subset \mathbb{R}^2$ with a boundary condition $\psi = g(x)(x \in \partial D)$ is used for a study of a vortex solution for large $\kappa$. Let $\epsilon = 1/\kappa$. Under the constraint $\deg(g, \partial D) = d$, a minimizer $\psi_\epsilon$ must have at least $|d|$ zeros. The leading order of the energy around each vortex is estimated as $2\pi\log(1/\epsilon)$. The result of Bethuel *et al.* (1994) describes the energy for a minimizer

$$G(\psi_\epsilon) = 2\pi|d|\,\log(1/\epsilon) + \gamma + W(a_1^\epsilon, \ldots, a_{|d|}^\epsilon) + o(1)$$

where $\{a_j^\epsilon\}$ are zeros of $\psi_\epsilon$ and $\gamma$ is a universal constant. The function $W$ is explicitly given as

$$W(a_1, \ldots, a_{|d|}) = 2\pi \sum_{1 \le j,k \le |d|, j \ne k} \log|a_j - a_k| + R$$

where $R$ is derived from a Green function satisfying some boundary condition depending on $g$. Moreover, as $\epsilon \to 0$, the zeros converge to a minimizer of $W$, which implies that the asymptotic position of every zero (vortex) is determined by the explicit function $W$. The first term of $W$ shows that vortices with the same sign of the degree are repulsive to one another and the optimal arrangement of vortices never allows the superposition of multivortices. Although the boundary condition is rather artificial, their mathematical formulation promoted the development of variational methods applied to the Ginzburg–Landau equation.

## Time-Dependent Ginzburg–Landau Equations

The Ginzburg–Landau equations in the preceding sections are static models. We consider time evolution models called the time-dependent Ginzburg–Landau equations. The evolution equations serve various numerical simulations exhibiting dynamical properties of solutions. They also provide mathematical problems on global time behaviors of solutions, stability of stationary solutions, dynamical laws of vortices, etc. The Ginzburg–Landau energy is denoted by $\mathcal{E}(u), u = (\Psi, \mathbb{A})$. The simplest model for the time-dependent problem is the gradient flow for $\mathcal{E}(u)$

$$\partial_t u = -\frac{\delta\mathcal{E}}{\delta u}$$

where $\delta\mathcal{E}/\delta u$ is the first variation of the energy. A more standard evolution equation in a nondimensional form is given by

$$(\partial_t + \mathrm{i}\phi)\Psi - D_A^2\Psi = \kappa^2(1 - |\Psi|^2)\Psi \qquad [25]$$

$$\eta(\partial_t\mathbb{A} + \nabla\phi) + \mathrm{curl}^2\,\mathbb{A} = \mathrm{Im}(\Psi^* D_A\Psi) + \mathrm{curl}\,\mathbb{H}_{\mathrm{ap}} \qquad [26]$$

where $\phi(x, t)$ is the electric (scalar) potential and $\eta$ is a positive parameter with a physical quantity. In fact, this equation was derived by Gor'kov and Eliashberg from the Bardeen, Cooper, and Schrieffer (BCS) theory.

The system of the equations [25]–[26] is invariant under the following time-dependent gauge transformation:

$$(\Psi, \phi, \mathbb{A}) \longmapsto (\Psi\,\exp(\mathrm{i}\chi), \phi - \partial_t\chi, \mathbb{A} + \nabla\chi)$$

The equations in the bounded domain $D \subset \mathbb{R}^2$ are considered subject to boundary and initial conditions

$$\begin{aligned} D_A\Psi \cdot \boldsymbol{n} &= 0 && \text{on } \partial\Omega \times (0,T) \\ \operatorname{curl} \mathbb{A} &= H_{\mathrm{a}} && \text{on } \partial\Omega \times (0,T) \\ \Psi(x,0) &= \Psi_0 && \text{in } \Omega \\ \mathbb{A}(x,0) &= \mathbb{A}_0(x) && \text{in } \Omega \end{aligned} \qquad [27]$$

Then, besides the Coulomb gauge [8], we can choose the Lorentz gauge as follows:

$$\begin{aligned} \operatorname{div} \mathbb{A} + \phi &= 0 \quad \text{in } D, \qquad \int_D \phi \, \mathrm{d}x = 0 \\ \mathbb{A} \cdot \boldsymbol{n} &= 0 \quad \text{on } \partial D \end{aligned}$$

For a smooth solution $u(x,t)$ to [25]–[26] with [27],

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(u) = -2\int_\Omega |(\partial_t + \mathrm{i}\phi)\Psi|^2 + \eta|\partial_t\mathbb{A} + \nabla\phi|^2 \, \mathrm{d}x \le 0$$

holds if $\mathbb{H}_{\mathrm{ap}}$ is time independent. This is also true in the case of the whole space $\mathbb{R}^2$ with a condition for the asymptotic behavior as $|x| \to \infty$.

Suppose that a domain $\Omega \subset \mathbb{R}^3$ is occupied by a superconducting sample and it is surrounded by a medium (or vacuum). Then the electromagnetic behavior in the outside domain, caused by the induced magnetic field of a supercurrent in $\Omega$ and an applied magnetic field, should be expressed by the Maxwell equations. With the electric field $\mathbb{E} = -(\mu\partial_t\mathbb{A} + \nabla\phi)$, we obtain

$$-\nu\partial_t\mathbb{E} - \sigma\mathbb{E} + \operatorname{curl}^2 \mathbb{A} = \operatorname{curl} \mathbb{H}_{\mathrm{ap}} \quad \text{in } \mathbb{R}^3 \setminus \bar{\Omega}$$

where $\mu$, $\nu$, and $\sigma$ are physical parameters (e.g., $\sigma = 0$ in the vacuum). To match the inside and the outside of $\Omega$, appropriate boundary conditions are required.

From a point of the gauge theory as in the section "Ginzburg–Landau equations in $\mathbb{R}^2$," the following time-dependent equations in the whole space are also considered:

$$\begin{aligned} (\partial_t + \mathrm{i}\phi)^2\Psi - D_A^2\Psi &= \kappa^2(1 - |\Psi|^2)\Psi \\ -\partial_t\mathbb{E} + \operatorname{curl}^2 \mathbb{A} &= \operatorname{Im}(\Psi^* D_A\Psi) \\ -\nabla \cdot \mathbb{E} &= \operatorname{Im}(\Psi^*(\partial_t + \mathrm{i}\phi)\Psi) \end{aligned}$$

## Other Topics

In realistic problems, a superconducting sample contains impurities. This inhomogeneity is usually expressed by putting a variable coefficient into the Ginzburg–Landau energy and the equations. Such a model with a variable coefficient is useful in studies for pinning of vortices, Josephson effect through an inhomogeneous media, etc. A model in a thin film with variable thickness is also described by the Ginzburg–Landau equations with a variable coefficient. Since the Ginzburg–Landau equations (or a modified model) can be considered in various settings, more applications to realistic problems would be treated by the development of nonlinear analysis.

*See also:* Abelian Higgs Vortices; Bifurcation Theory; Evolution Equations: Linear and Nonlinear; High $T_c$ Superconductor Theory; Image Processing: Mathematics; Integrable Systems: Overview; Interacting Stochastic Particle Systems; Ljusternik–Schnirelman Theory; Nonlinear Schrödinger Equations; Quantum Phase Transitions; Variational Techniques for Ginzburg–Landau Energies.

## Further Reading


Bethuel F, Brezis H, and Hélein F (1994) *Ginzburg–Landau Vortices*. Boston: Birkhäuser.

Chapman SJ, Howison SD, and Ockendon JR (1992) Macroscopic models for superconductivity. *SIAM Review* 34: 529–560.

Del Pino M, Felmer PL, and Sternberg P (2000) Boundary concentration for the eigenvalue problems related to the onset of superconductivity. *Communications in Mathematical Physics* 210: 413–446.

Du Q, Gunzberger MD, and Peterson JS (1992) Analysis and approximation of the Ginzburg–Landau model of superconductivity. *SIAM Review* 34: 54–81.

Helffer B and Morame A (2001) Magnetic bottles in connection with superconductivity. *Journal of Functional Analysis* 185: 604–680.

Hoffmann K-H and Tang Q (2001) *Ginzburg–Landau Phase Transition Theory and Superconductivity*. Basel: Birkhäuser.

Jaffe A and Taubes C (1980) *Vortices and Monopoles*. Boston: Birkhäuser.

Jimbo S and Morita Y (1996) Ginzburg–Landau equations and stable solutions in a rotational domain. *SIAM Journal of Mathematical Analysis* 27: 1360–1385.

Lin FH and Du Q (1997) Ginzburg–Landau vortices: dynamics, pinning and hysteresis. *SIAM Journal of Mathematical Analysis* 28: 1265–1293.

Lu K and Pan X-B (1999) Estimates of the upper critical field for the Ginzburg–Landau equations of superconductivity. *Physica D* 127: 73–104.

Rubinstein J and Sternberg P (1996) Homotopy classification of minimizers of the Ginzburg–Landau energy and the existence of permanent currents. *Communications in Mathematical Physics* 179: 257–263.

Sandier E and Serfaty S (2000) On the energy of type-II superconductors in the mixed phase. *Reviews in Mathematical Physics* 12: 1219–1257.

Serfaty S (1999) Stable configurations in superconductivity: uniqueness, multiplicity, and vortex-nucleation. *Archives for Rational and Mechanical Analysis* 149: 329–365.

Tinkham M (1996) *Introduction to Superconductivity*, 2nd edn. New York: McGraw-Hill.

# Glassy Disordered Systems: Dynamical Evolution

**S Franz**, The Abdus Salam ICTP, Trieste, Italy



## Introduction

Many macroscopic systems if left to evolve in isolation or in contact with a bath, are able to relax, after a finite time, to history-independent equilibrium states characterized by time-independent values of the state variables and time-translation invariance correlations. In glassy systems, the relaxation time becomes so large that equilibrium behavior is never observed. On short timescales, the microscopic degrees of freedom appear to be frozen in far-from-equilibrium disordered states. On longer timescales slow, history-dependent, off-equilibrium relaxation phenomena become detectable.

The list of physical systems falling in disordered glassy states at low temperature is long, just to mention a few examples one can cite the canonical case of simple and complex liquid systems undergoing a glass transition, polymeric glasses, dipolar glasses, spin glasses, charge density wave systems, vortex systems in type II superconductors, and many other systems.

Experimental and theoretical research has pointed out the existence of dynamical scaling laws characterizing the off-equilibrium evolution of glassy systems. These laws, in turn, reflect the statistical properties of the regions of configuration space explored during relaxation.

The goal of a theory of glassy systems is the comprehension of the mechanisms that lead to the growth of relaxation time and the nature of the scaling laws in off-equilibrium relaxation. A well-developed description of glassy phenomena is provided by mean-field theory based on spin glass models, which gives a coherent framework that is able to describe the dynamics of glassy systems and provides a statistical interpretation of glassy relaxation. Despite important limitations of the mean-field description for finite-dimensional systems, it allows precise discussions of general concepts such as effective temperatures and configurational entropies that have been successfully applied to the description of glassy systems.

In the following, examples of two different ways of freezing will be discussed: spin glasses, where disorder is built in the random nature of the coupling between the dynamical variables, and structural glasses, where the disordered nature of the frozen state has a self-induced character. These systems are examples of two different ways of freezing.

## A Glimpse of Freezing Phenomenology

### Spin Glasses

The archetypical example of systems undergoing the complex dynamical phenomena described in this article is the case of spin glasses (Fischer and Hertz 1991, Young 1997). Spin glass materials are magnetic systems where the magnetic atoms occupy random position in lattices formed by nonmagnetic matrices fixed at the moment of the preparation of the material. The exchange interaction between the spin of the magnetic impurities in these materials is an oscillating function, taking positive and negative values according to the distance between the atoms.

Spin glass models (*see* Spin Glasses, Mean Field Spin Glasses and Neural Networks, and Short-Range Spin Glasses: The Metastate Approach) are defined by giving the form of the exchange Hamiltonian, describing the interaction between the spins $S_i$ of the magnetic atoms. In the presence of an external magnetic field $h$, the exchange Hamiltonian can be written as

$$H = -\sum_{i,j\in\Lambda} J_{ij} S_i \cdot S_j - h \sum_{i\in\Lambda} S_i \qquad [1]$$

The spin variable can have classical or quantum nature. This article will be limited to the physics of classical systems. The most common choice in models is to use Ising variables $S_i = \pm 1$. The couplings $J_{ij}$, which in real material depend on the distance, are most commonly chosen to be independent random variables with a distribution with support on both positive and negative values. Most commonly, one considers either a symmetric bimodal distribution on $\{-1, 1\}$ or a symmetric Gaussian. The sums are restricted to lattices $\Lambda$ of various types. The most common choices are $\Lambda = Z^d$ for the Edwards–Anderson model, the complete graph $\Lambda = \{(i,j) | i < j; i,j = 1, \ldots, N\}$ for the Sherrington–Kirkpatrick (SK) model, and the Erdos–Renyi random graph for the Viana–Bray (VB) model.

The presence of interactions of both signs induces frustration in the system: the impossibility of minimizing all the terms of the Hamiltonian at the same time. One then has a complex energy landscape, where relaxation to equilibrium is hampered by barriers of energetic and entropic nature.

Spin glass materials, which have a paramagnetic behavior at high temperature, show glassy behavior at low temperature, where magnetic degrees of freedom appear to be frozen for long times in apparently random directions. There is quite a

general consensus, based on the analysis of the experimental data and the numerical simulations, that in three dimensions and in the absence of a magnetic field, the two regimes are separated by a thermodynamic phase transition at a temperature $T_c$ where the magnetic response $\chi$ exhibits a cusp (see Figure 1). By linear response, $\chi$ is related to the equilibrium spin correlation function

$$\chi = \frac{1}{K_B TN} \sum_i \left( \langle S_i^2 \rangle - \langle S_i \rangle^2 \right)$$

having denoted by $\langle \cdot \rangle$ the Boltzmann–Gibbs average. A cusp in $\chi$ indicates a second-order transition where the so-called Edwards–Anderson parameter $q = (1/N) \sum_i \langle S_i \rangle^2$ becomes different from zero, indicating freezing of the spins in random directions. In the presence of a magnetic field, although the low-temperature phenomenology is similar to the one at zero field, the thermodynamic nature of the freezing transition is more controversial. Theoretically, mean-field theory, based on the SK model, predicts a phase transition with a cusp in the susceptibility both in the absence and in the presence of a magnetic field. Unfortunately, no firm theoretical result is available on the existence and the nature of phase transitions in finite-dimensional spin glass models which is a completely open problem.

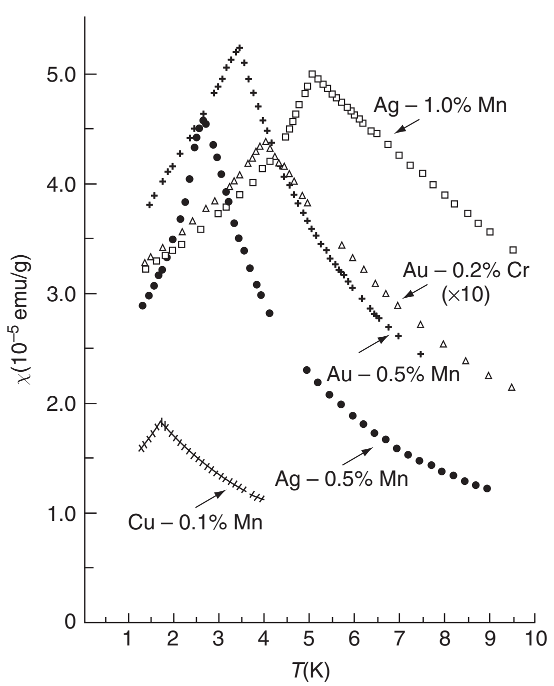



**Figure 1** Magnetic susceptibility as a function of temperature in spin glass materials. Reproduced from Fischer KH and Hertz JA (1991) *Spin Glasses*. Cambridge, UK: Cambridge University Press.

### Structural Glasses

Analogous freezing of dynamical variables is observed in a variety of systems. Some of them share with the spin glasses the presence of quenched disorder; in many others, this feature is absent. This is the case of structural glasses (Debenedetti 1996).

Many liquids under fast enough cooling, instead of crystallizing, as dictated by equilibrium thermodynamics, form glasses. Simple liquids can be modeled as classical systems of particles with pairwise interactions. In the simplest example of a monoatomic liquid, the potential energy of a configuration is then written as

$$V(r_1, \ldots, r_N) = \sum_{i<j} \phi(r_i - r_j) \qquad [2]$$

In the case of atomic mixtures, the potential $\phi$ acquires a dependence on the species of the interacting atoms.

Liquids can be characterized as good or bad glass formers depending on the facility by which they form glasses. In good glass formers, in order to avoid crystallization, it is in general sufficient to cross the region around the liquid–crystal transition point fast enough, so that the systems can set in a supercooled liquid metastable equilibrium. On lowering the temperature, the supercooled liquid becomes denser and more viscous while the relaxation time of the system, related to the viscosity through the Maxwell relation $\tau = \eta/G$ ($G$ is the instantaneous shear modulus of the liquid), undergoes a rapid growth. One defines a conventional glass transition temperature $T_g$ as the point where $\eta$ takes the solid-like value $\eta = 10^{13} Poise$, corresponding to a relaxation time $\tau \approx 100$ s. After that point, the system falls out of equilibrium; under usual experimental conditions, it does not have the time to adjust to external solicitations and behaves mechanically like a solid. The glass transition temperature is then characterized as the point where the liquid goes out of equilibrium, the relaxation time becomes larger than the external timescale and the positions of the atoms appear as frozen on that scale.

A great effort has been devoted to understand the behavior of the temperature dependence of the relaxation time and the nature of the dynamical processes in supercooled liquids. In deeply supercooled liquids, the empirical behavior of the relaxation time ranges from the Arrhenius form for "strong glasses" $\tau \sim \exp(\Delta/T)$ to the Vogel–Fulcher form $\tau(T) \sim \exp(D/(T - T_0))$ for "fragile glasses." The Vogel–Fulcher law predicts a finite-temperature divergence of the relaxation time at the temperature $T_0$. Unfortunately, in typical cases, the $T_0$ results are

estimated to be 10–15% lower then $T_g$ so that it is not possible to verify the law close enough to $T_0$ to support the divergent behavior.

As a consequence of freezing, one observes important qualitative changes in the behavior of thermodynamic quantities similar to those encountered in equilibrium phase transitions. In a narrow interval around $T_g$, specific heat and compressibility undergo jumps from liquid-like values to much lower solid-like values.

## Aging and Slow Dynamics

While the crudest picture of the glass transition describes freezing as complete structural arrest, both for the cases where the glass transition is a genuine off-equilibrium phenomenon, as in structural glasses, and in the case where it has a thermodynamical character as in spin glasses, the study of dynamical quantities reveals the existence of persisting, history-dependent, slow relaxation processes in the frozen phase (Norblad and Svendlidh 1997). This is the phenomenon of aging, which is a constitutive feature of the glassy state. Its theoretical analysis occupies a central theoretical role in the comprehension of the way glassy systems explore configuration space. A first characterization of relaxation is given by the behavior of "one-time quantities" like internal energy, density, etc., which slowly evolve in the course of time towards values corresponding to states of lower free energy. More interesting is the behavior of "two-time quantities," time-dependent correlation functions and responses, which reveal the deep off-equilibrium nature of glassy relaxation. In experimental, numerical, and theoretical studies, a special position is occupied by the linear response function. Using the language of magnetic systems, apt to the spin glasses, one considers the response of the magnetization to an applied magnetic field. To deal with other systems, different conjugated couples of variables are considered and simple changes of language are needed. Linear perturbations allow to reveal the dynamics of the systems without affecting its evolution. Denoting by $M(t)$ the magnetization at a time $t$ and by $h(t')$ the magnetic field at time $t'$, the instantaneous linear response function is defined as

$$R(t,t') = \left(\frac{\delta M(t)}{\delta h(t')}\right)_{h=0} \qquad [3]$$

Measures of the time integral of $R(t,t')$ are commonly performed to reveal the presence of aging in glassy systems. Aging is usually studied observing the dynamics that follows a rapid quench from high temperature, at an instant that marks the origin of time. One can reconstruct the response function measuring the zero-field-cooled (ZFC) magnetization as the response to a magnetic field acting from a waiting time $t_w$ to the measuring time $t$,

$$\chi_{ZFC}(t,t_w) = \int_{t_w}^{t} dt'\, R(t,t') \qquad [4]$$

or its complement, the thermoremanent magnetization (TRM) corresponding to the response to a magnetic field acting from the time of the quench up to $t_w$

$$\chi_{TRM}(t,t_w) = \int_{0}^{t_w} dt'\, R(t,t') \qquad [5]$$

In Figure 2, the behavior of the susceptibility $\chi_{ZFC}$ is shown as a function of $t - t_w$ in a typical example of aging experiment at low temperature. Out-of-equilibrium behavior is manifest in the dependence of the curves on the waiting time $t_w$. The relaxation appears slower and slower for larger waiting times, and the $t_w$ dependence does not disappear even for very large times. Two nontrivial dynamical regimes can be identified: a first regime for small $t - t_w$, that is, $t - t_w << t_w$ where the relaxation is independent of $t_w$ and a second regime roughly valid for $t - t_w \sim t_w$ where time-translation invariance is manifestly violated. The analysis of experimental and simulation data shows a scale-invariant behavior according to which curves corresponding to different waiting times can be superimposed rescaling the time difference $t - t_w$ with a suitable $t_w$-dependent relaxation time $\tau(t_w)$. This is a growing function of $t_w$ which seems to diverge for large $t_w$. Up to the waiting times where it has been possible to test the relation, $\tau(t_w)$ behaves as a power $\tau(t_w) \sim t_w^{\alpha}$ where in different materials and models, $\alpha = 0.8$–$0.9$.

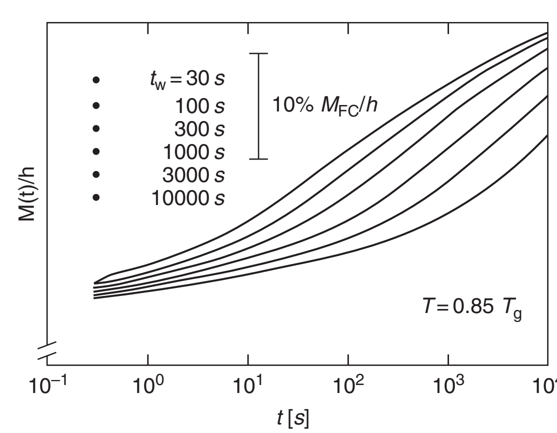



**Figure 2** ZFC magnetization in an aging experiment. The curves, from bottom to top, correspond to increasing waiting times. Reproduced from Norblad P and Svendlidh P (1997) Experiments in spin glasses. In: Young AP (ed.) *Spin Glasses and Random Fields*. Singapore: World Scientific, with permission from World Scientific Publishing Co. Pte Ltd.

Many efforts have been devoted to the comprehension of the scaling laws in aging (Bouchaud *et al.*). Among the theories and models of aging that have been proposed, one can cite the phenomenological model known as "trap model," developed by Bouchaud and collaborators that assimilates aging to a random walk between "traps" characterized by a broad distribution of trapping times. Suitable choices of the trapping-time distribution allow to derive scaling laws similar to the ones characteristic of aging systems. A different theory, the "droplet model" for spin glasses assimilates aging phenomena to the competition between slowly growing domains of equilibrium phases, in analogy with the dynamics of phase separation in first-order phase transitions. The approach that has led to the most detailed and spectacular predictions has been the study of microscopic mean-field models.

## Mean-Field Models of Disordered Systems

Mean-field theory starts from the analysis of the relaxation dynamics of disordered systems with weak long-range forces (Bouchaud *et al.*). The reference model of spin glass mean-field theory is the so called $p$-spin model, which considers $N$ spins $S_i$ with random $p$-body interactions with each other and is described by the Hamiltonian

$$H_p(\mathbf{S}) = \sum_{i_1<\cdots<i_p}^{1,N} J_{i_1,\ldots,i_p} S_{i_1}\cdots S_{i_p} \qquad [6]$$

where the quenched coupling constants $J_{i_1,\ldots,i_p}$ are assumed to be i.i.d. Gaussian variables with zero average and $N$ dependent variance $E(J^2_{i_1,\ldots,i_p}) = p!/2N^{p-1}$. The case $p=2$ coincides with the SK model defined in the introduction. The reason for considering the $p$-spin generalization is that the order of the transition passes from the second one for $p=2$ to the first one for $p \geq 3$ and that this last case has been suggested to provide a mean-field limit for the structural glass transition. It is also useful to define Hamiltonians

$$H[S] = \sum_{p\geq 1} a_p H_p[S] \qquad [7]$$

that mix $p$-spin Hamiltonians for different $p$. These are random Gaussian functions of the spin variables, with covariance induced by the coupling distribution

$$E[H(\mathbf{S})H(\mathbf{S}')] = Nf(q(\mathbf{S},\mathbf{S}')) = \frac{N}{2}\sum_{p\geq 1} a_p^2 q(\mathbf{S},\mathbf{S}')^p \qquad [8]$$

where the function

$$q(\mathbf{S},\mathbf{S}') = \frac{1}{N}\sum_{i=1}^{N} S_i S_i'$$

is the overlap between configurations. A crucial hypothesis in the study of relaxation in spin systems is that any local spin update rule verifying the detailed balance condition with respect to the Boltzmann–Gibbs measure gives rise to the same long-time properties. In this perspective, in Monte Carlo simulations, it is convenient to use Ising spins with Metropolis or Glauber dynamics. Much theoretical progress has been achieved considering spherical models where the spin variables are real numbers subject to a global spherical constraint $\sum_i S_i^2 = N$ and evolve according to the following Langevin dynamics:

$$\frac{\mathrm{d}S_i(t)}{\mathrm{d}t} = -\frac{\partial H(\mathbf{S}(t))}{\partial S_i} - \mu(t)S_i(t) + \eta_i(t) \qquad [9]$$

where $\mu(t)$ is a time-dependent multiplier that at each instant of time insures that the spherical constraint is respected, and $\eta_i(t)$ is a thermal white noise with variance

$$E(\eta_i(t)\eta_j(s)) = 2\,T\delta_{ij}\delta(t-s) \qquad [10]$$

In order to model the quench from high temperature performed in experiments, the initial conditions are randomly chosen with uniform probability. To describe long but finite-time dynamics, it is necessary to consider the limit of large volume $N\to\infty$ for finite time, which is the only case where one can have infinite relaxation times. Application of functional Martin–Siggia–Rose techniques has allowed the derivation of closed integro-differential equations for the spin autocorrelation function

$$C(t,t') = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\langle S_i(t)S_i(t')\rangle$$

and the response to an impulsive external field

$$h_i(t), R(t,t') = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left\langle\frac{\delta S_i(t)}{\delta h_i(t')}\right\rangle$$

where the average has to be intended on quenched disordered couplings, initial conditions, and realization of thermal noise. Unfortunately, in the case $p=2$ relevant for spin glass phenomenology, the spherical constraints reduce the model to a linear system where different eigenmodes of the interaction matrix $J_{ij}$ evolve independently. This oversimplification renders the model similar to systems apt to describe phase separation rather then freezing phenomena. Many of the glassy features of the SK model however are captured by a mixture of $p=2$

with $p=4$ Hamiltonians and $f(q)=(1/2)(q^2+aq^4)$. For the general Hamiltonian [7] one gets the coupled equations

$$\frac{\partial C(t,t')}{\partial t} = -\mu(t)C(t,t') + \int_0^t dt'' f''(C(t,t''))R(t,t'')C(t',t'') + \int_{t'}^t dt'' f'(C(t,t''))R(t'',t')$$

$$\frac{\partial R(t,t')}{\partial t} = -\mu(t)R(t,t') + \int_{t'}^t dt'' f''(C(t,t''))R(t,t'')R(t'',t') \quad [11]$$

$\mu(t)$ is a multiplier that at each instant of time insures the spherical constraint $C(t,t)=1$, and is determined by

$$\mu(t) = \int_0^t dt'' C(t,t'')R(t,t'') + T \quad [12]$$

In the next sections, we will discuss how these equations describe dynamical freezing at low temperature. The gross features are determined by the form of the function $f(q)$. Two main behaviors can be identified:

1. Systems of type I. This behavior is found if $f'''(q)/(f''(q))^{3/2}$ is a monotonically decreasing function of $q$. To this family belongs the pure spherical $p$-spin model for $p \geq 3$, and one finds a dynamical transition not corresponding to a point of singularity in the free energy where the Edwards–Anderson parameter jumps discontinuously to a nonzero value. Models of this family have been proposed as appropriate mean-field limits for structural glass behavior.
2. Systems of type II. This behavior is found if $f'''(q)/(f''(q))^{3/2}$ is a monotonically increasing function of $q$. This family mimics the behavior of the SK model. An example of function $f$ verifying the condition for type II behavior is $f(q)=1/2(q^2+aq^4)$ for sufficiently small but positive values of $a$. In this case, the dynamical transition is found at a point of second-order singularity of the free energy and the Edwards–Anderson parameter is continuous at the transition. Models of this family provide a mean-field limit for spin glass type behavior.

### Equilibrium Dynamics at High Temperature

At high temperature, after a finite transient, eqn [11] describes equilibrium behavior. In these conditions, time-translation invariance holds $C(t,t')=C(t-t')$, $R(t,t')=R(t-t')$ while the Lagrange multiplier $\mu$ becomes time independent. In addition, correlation and response are related by the fluctuation–dissipation theorem (FDT) relation

$$R(t) = -\frac{1}{T}\frac{dC(t)}{dt} \quad [13]$$

Ergodic behavior is manifest in the fact that the dynamics decorrelate completely; $\lim_{t\to\infty} C(t)=0$. Then from [11] one gets the equilibrium equation:

$$\frac{dC(t)}{dt} = -TC(t) - \frac{1}{T}\int_0^t ds f'(C(t-s))\frac{dC(s)}{ds} \quad [14]$$

It is worth noticing that this equation, apart from an irrelevant inertial term, coincides for type I systems with the schematic mode-coupling theory (MCT) equation which has been successfully used to describe moderate supercooled liquids (Goetze 1989). In the context of liquid theory, mode-coupling equations stem from an approximate treatment for the dynamical evolution of the density–density space and time-dependent correlation function. The schematic MCT equations consider an equation for a single mode, neglecting any space dependence of the correlator.

Both in type I and in type II systems, eqn [14] displays a dynamical transition at a finite temperature $T_c$ where the relaxation time diverges as a power law $\tau \sim |T-T_c|^c$ and the asymptotic value of the correlation acquires a nonzero value.

This behavior in type I systems represents a failure of MCT to describe the temperature dependence of the relaxation time in supercooled liquids, which, as previously observed, empirically follows the Vogel–Fulcher law. The MCT temperature is interpreted as a singularity which is avoided in supercooled liquids, thanks to relaxation mechanisms specific of short-range systems. It has been noticed that this singularity at $T_c$ can be associated to the growth of spatial heterogeneities and dynamical correlations, as exemplified in the behavior of the four-point function

$$\chi_4(t) = \frac{1}{N}\sum_{i,j}\langle S_i(t)S_j(t)S_i(0)S_j(0)\rangle$$

and its associate correlation length (Franz and Parisi 2000, Biroli and Bouchaud 2004).

### Off-Equilibrium Dynamics Below $T_c$: Aging and Slow Dynamics

**Type I systems** Below the transition temperature $T_c$ slow dynamics and aging set in. In 1993, Cugliandolo and Kurchan found a long-time

solution to the equations of motion [11] for type I systems describing an asymptotic off-equilibrium state that follows from high-temperature quench. Soon after, type II systems were also analyzed (Bouchaud *et al.*).

The equations can be analyzed in the limit in which both times tend to infinity $t, t' \to \infty$. In this regime all "one-time quantities," that is, state functions like energy, magnetization, etc., reach asymptotic time-independent limit. Though the decay to the asymptotic value cannot read directly from the analysis of the equations in that limit, numerical and theoretical evidence suggests that the final values are approached as power laws in time.

The study of correlation and response functions displays an asymptotic scaling behavior similar to the one observed in glassy systems in laboratory and numerical experiments.

Two different interesting regimes are found, first of all there is a stationary regime: the limit $t, t_w \to \infty$ is performed keeping the difference $t - t_w = s$ finite. In this regime, equilibrium behavior is observed, with correlation and response related by the FDT relation $R_{st}(s) = -\beta \partial C_{st}(s)/\partial s$. The stationary regime is followed by an aging regime, where correlations decay below the value $q_{EA} = \lim_{s \to \infty} C_{st}(s)$ down to zero. One of the most striking features of aging evolution is that the system – though at a decreasing speed – constantly move far apart from any visited region of configuration space. The decay of correlations is nonstationary and takes place on a timescale $\tau(t_w)$ diverging for large $t_w$. While the theory can infer the existence of the timescale $\tau(t_w)$, its precise form remains undetermined. This is a consequence of an asymptotic invariance under monotonous time reparametrizations $t \to g(t)$ appearing for large times. Coherently with nonstationary behavior, other equilibrium properties break down in the aging regime. Correlation and response which do not verify the FDT are rather asymptotically related by a generalized form of the fluctuation–dissipation relation

$$R_{ag}(t, t_w) = \frac{X}{T} \frac{\partial C_{ag}((t - t_w)/\tau(t_w))}{\partial t_w} \qquad [15]$$

This relation, despite predicting the vanishing of the instantaneous response, implies a finite contribution of the aging dynamics to the value of the integrated ZFC and TRM responses. The constant $X$, called fluctuation–dissipation ratio (FDR), is a temperature-dependent factor monotonically varying between the values 1 and 0 as the temperature is decreased from $T_c$ down to zero. Violations of the FDT have to be expected in any off-equilibrium regime; however, a constant ratio between response and derivative of the correlation is very nongeneric. It is of great theoretical importance that the same constant that governs the FDR among spin autocorrelation and magnetic response, also appears in the relation of any other conceivable couple of correlation and conjugated response in the system. Slow dynamics can be interpreted as motion between finite-life metastable states with well-defined free energy $f$ and exponential multiplicity $\exp(N\Sigma(f))$. The FDR verifies the generalized thermodynamic relation

$$\frac{\partial \Sigma}{\partial f} = \frac{X}{T} \qquad [16]$$

This relation is in turn intimately related to the possibility of considering the ratio $T_{eff} = T/X$ as an effective temperature, that governs the heat exchanges among slow degrees of freedom (Cugliandolo *et al.* 1997). Slow degrees of freedom do not exchange heat with the fast ones, but they are in equilibrium between themselves at the temperature $T_{eff}$. The validity of relation [16] has been put at the basis of a detailed statistical description of the glassy state (Franz and Virasoro 2000, Biroli and Kurchan 2001, Nieuwenhuizen 2000) which assumes that metastable states with equal free energy are encountered with equal probability during the descent to equilibrium. Modified thermodynamic relations follow, that condensate all the dependence on the thermal history in the value of the effective temperature. Given the interest of a thermodynamic description of the glassy state, many numerical studies have addressed the problem of the identification and determination of effective temperatures from the fluctuation–dissipation relations, and its relation with configurational entropy. In **Figure 3** the result of a numerical study on a realistic system is presented, verifying relation [15]. Experimental verifications are at the moment starting and new results are waited in the future.

**Type II systems** In these systems the dynamic transition occurs at the point of thermodynamic singularity, where the Edwards–Anderson parameter becomes nonzero in a second-order fashion. The magnetic susceptibility exhibits a cusp singularity similar to the one found in spin glass materials. Differently from type I systems, one-time quantities tend to their equilibrium values for long times. The off-equilibrium nature of the relaxation shows up in the behavior of correlations and responses, which display aging behavior.

Their behavior generalizes the one found in type I systems, with a more complex pattern of violation

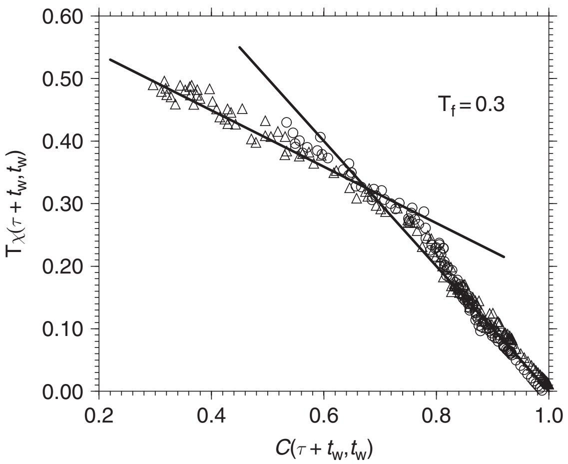



**Figure 3** Fluctuation–dissipation plot; $\chi_{ZFC}(t, t_w)$ vs. $C(t, t_w)$ in a model of Lennard-Jones glass for different values of the waiting time $t_w$. The slope of the curves is equal to the finite-time FDR divided by the temperature. One observes the characteristic shape of type-I systems with an FDR equal to 1 in the stationary regime for high correlations and equal to a constant smaller then 1 in the low-correlations aging regime. Reproduced from Kob W and Barrat J-L (1999) *Europhysics Letters* 46: 637, with permission from EDP Sciences.

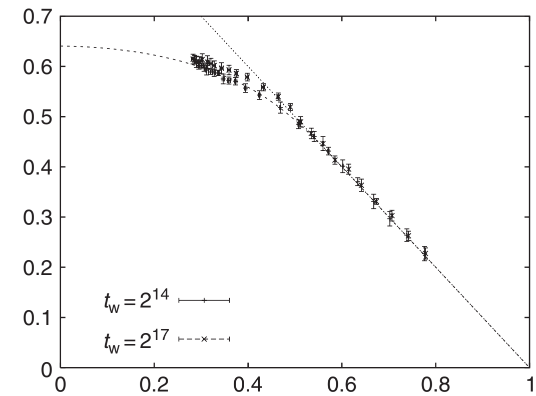



**Figure 4** Fluctuation–dissipation plot in a three-dimensional spin glass at low temperature. As predicted for type-II systems, the FDR is an increasing function of the correlation, constant only in the stationary part of the relaxation. Reproduced from Marinari E *et al.* (1998) Violation of the fluctuation dissipation theorem in finite dimensional spin glasses. *Journal of Physics A: Mathematical and General* 31: 2611, with permission from Institute of Physics Publishing Ltd.

of time-translation invariance and FDT. Also in this case a short-time equilibrium behavior can be identified where the correlation decreases from 1 to $q_{EA}$ and a long-time inhomogeneous aging behavior where correlations decrease to zero. Differently from type I system it is impossible to characterize aging through a unique timescale $\tau(t_w)$. One finds instead a continuum of timescales hierarchically organized. The analysis of the equations at the reparametrization-invariant level reveals the existence of a continuum of separate timescales $\tau(t_w, q)$ associated to each value of $C(t, t_w) = q < q_{EA}$ and that $\lim_{t_w \to \infty} \tau(t_w, q)/\tau(t_w, q') = 0$ for $q > q'$, meaning that for finite $t_w$ the time to decay to $q'$ is much larger than the time to decay to $q$. For large times, $1 << t_1 << t_2 << t_3$, the correlations verify the ultrametric property $C(t_3, t_1) = \min[C(t_3, t_2), C(t_2, t_1)]$. To each timescale corresponds in this case a different effective temperature, and correlation and response are related by the equation

$$\beta X(q) = \lim_{\substack{t, t_w \to \infty \\ C(t,t') = q}} \frac{R(t, t')}{\partial C(t, t')/\partial t'} \qquad [17]$$

where the function $X(q)$ is an increasing function of $q$ with the properties of a cumulative probability distribution. In fact it can be seen (Franz *et al.* 1999) that this is related to the Parisi overlap probability function describing the correlations among ergodic components at equilibrium, in a generalization of relation [14]. **Figure 4** shows the result of a numerical experiment in a three-dimensional spin glass, where $X(q)$ is not piecewise constant.

The ideas presented in this article, fruits of mean-field theory of disordered systems, are objects of intense debate in their application to the physics of short-range systems. Many of the relations derived have stimulated a lot of numerical, experimental, and theoretical work. Some of the predictions of the theory are very well verified in many short-range glassy systems, at least on the accessible timescales. Notably, the violations of FDR, and the possibility to associate the values of the FDR to effective temperatures is very well verified both in structural glass models, and in finite-range spin glasses. Since finite aging times imply finite length scales over which the dynamic variables can exhibit correlated behavior, this indicates that the mean-field theory is at least good at describing glassy phenomena on a local scale. The question if the mean-field theory also gives a good description on the infinite time limit and the anomalous response persists forever is at present an open theoretical problem. It relates to the possibility of having mean-field type of equilibrium ergodicity breaking, which is an open question, object of active research.

*See also:* Interacting Stochastic Particle Systems; Short-Range Spin Glasses: The Metastate Approach; Spin Glasses.

## Further Reading


Barrat J-L and Kob W (1999) *Europhysics Letters* 46: 637.

Barrat J-L and Kob W Fluctuation–dissipation ratio in an aging Lennard–Jones glass.

Biroli G and Bouchaud J-P (2004) Diverging length scale and upper critical dimension in the mode-coupling theory of the glass transition. *Europhysics Letters* 67: 21.

Biroli G and Kurchan J (2001) Metastable states in glassy systems. *Physical Review E* 64: 16101.

Bouchaud J-P, Cugliandolo L, Kurchan J, and Mezard M Out of equilibrium dynamics of spin-glasses and other glassy systems. In: Young AP (ed.) *Spin Glasses and Random Fields*. Singapore: World Scientific.

Cugliandolo LF, Kurchan J, and Peliti L (1997) Energy flow, partial equilibration and effective temperatures in systems with slow dynamics. *Physical Review E* 55: 3898.

Debenedetti P (1996) *Metastable Liquids, Concepts and Principles*. New Jersey: Princeton University Press.

Fischer KH and Hertz JA (1991) *Spin Glasses*. Cambridge: Cambridge University Press.

Franz S, Mezard M, Parisi G, and Peliti L (1999) The response of glassy systems to random perturbations: a bridge between equilibrium and off-equilibrium. *Journal of Statistical Physics* 97: 459.

Franz S and Parisi G (2000) On nonlinear susceptibility in supercooled liquids. In: Franz S, Glotzer SC, and Sastry S (eds.) *Proceedings of the Conference "Unifying Concepts in Glass Physics"*, Trieste, 5–18 September, 1999. *Journal of Physics: Condensed Matter* 12: 6335.

Franz S and Virasoro M (2000) Quasi-equilibrium interpretation of aging dynamics. *Journal of Physics A* 33: 891.

Goetze W (1991) Aspects of structural glass transitions. In: Hansen JP, Levesque D, and Zinn-Justin J (eds.) *Liquid, Freezing and the Glass Transition*. Amsterdam: Les Houches, North-Holland.

Marinari E *et al.* (1998) Violation of the fluctuation dissipation theorem in finite dimensional spin glasses. *Journal of Physics A: Mathematical and General* 31: 2611.

Nieuwenhuizen ThM (2000) In: Franz S, Glotzer SC, and Sastry S (eds.) *Thermodynamic Picture of the Glassy State*, Trieste, 5–18 September, 1999. *Journal of Physics: Condensed Matter* 12: 6335.

Norblad P and Svendlidh P (1997) Experiments in spin glasses. In: Young AP (ed.) *Spin Glasses and Random Fields*. Singapore: World Scientific.

Young AP (ed.) (1997) *Spin Glasses and Random Fields*. Singapore: World Scientific.

# Graded Poisson Algebras

**A S Cattaneo**, Universität Zürich, Zürich, Switzerland
**D Fiorenza and R Longoni**, Università di Roma "La Sapienza," Rome, Italy



## Definitions

### Graded Vector Spaces

By a $\mathbb{Z}$-graded vector space (or simply, graded vector space) we mean a direct sum $A = \oplus_{i\in\mathbb{Z}} A_i$ of vector spaces over a field $k$ of characteristic zero. The $A_i$ are called the components of $A$ of degree $i$ and the degree of a homogeneous element $a \in A$ is denoted by $|a|$. We also denote by $A[n]$ the graded vector space with degree shifted by $n$, namely $A[n] = \oplus_{i\in\mathbb{Z}}(A[n])_i$ with $(A[n])_i = A_{i+n}$. The tensor product of two graded vector spaces $A$ and $B$ is again a graded vector space whose degree $r$ component is given by $(A\otimes B)_r = \oplus_{p+q=r} A_p \otimes B_q$.

The symmetric and exterior algebras of a graded vector space $A$ are defined, respectively, as $S(A) = T(A)/I_S$ and $\bigwedge(A) = T(A)/I_\wedge$, where $T(A) = \oplus_{n\geq 0} A^{\otimes n}$ is the tensor algebra of $A$ and $I_S$ (resp. $I_\wedge$) is the two-sided ideal generated by elements of the form $a\otimes b - (-1)^{|a|\,|b|} b\otimes a$ (resp. $a\otimes b + (-1)^{|a|\,|b|}\ b\otimes a$), with $a$ and $b$ homogeneous elements of $A$. The images of $A^{\otimes n}$ in $S(A)$ and $\bigwedge(A)$ are denoted by $S^n(A)$ and $\bigwedge^n(A)$, respectively. Notice that there is a canonical decalage isomorphism $S^n(A[1]) \simeq \bigwedge^n(A)[n]$.

### Graded Algebras and Graded Lie Algebras

We say that $A$ is a graded algebra (of degree zero) if $A$ is a graded vector space endowed with a degree zero bilinear associative product: $A\otimes A \to A$. A graded algebra is graded commutative if the product satisfies the condition

$$a\cdot b = (-1)^{|a|\,|b|} b\cdot a$$

for any two homogeneous elements $a, b\in A$ of degree $|a|$ and $|b|$, respectively.

A graded Lie algebra of degree $n$ is a graded vector space $A$ endowed with a graded Lie bracket on $A[n]$. Such a bracket can be seen as a degree $-n$ Lie bracket on $A$, that is, as a bilinear operation $\{\cdot,\cdot\}: A\otimes A\to A[-n]$ satisfying graded antisymmetry and graded Jacobi relations:

$$\{a,b\} = -(-1)^{(|a|+n)(|b|+n)}\{b,a\}$$

$$\{a,\{b,c\}\} = \{\{a,b\},c\} + (-1)^{(|a|+n)(|b|+n)}\{a\{b,c\}\}$$

### Graded Poisson Algebras

We can now define the main object of interest of this note:

**Definition 1** A graded Poisson algebra of degree $n$, or $n$-Poisson algebra, is a triple $(A,\cdot,\{,\})$ consisting of a graded vector space $A = \oplus_{i\in\mathbb{Z}} A_i$ endowed with a degree zero graded commutative product and with

a degree $-n$ Lie bracket. The bracket is required to be a biderivation of the product, namely:

$$\{a, b \cdot c\} = \{a, b\} \cdot c + (-1)^{|b|(|a|+n)} b \cdot \{a, c\}$$

*Notation*. Graded Poisson algebras of degree zero are called Poisson algebras, while for $n = 1$ one speaks of Gerstenhaber (1963) algebras or of Schouten algebras.

Sometimes a $\mathbb{Z}_2$-grading is used instead of a $\mathbb{Z}$-grading. In this case, one just speaks of even and odd Poisson algebras.

**Example 1** Any graded commutative algebra can be seen as a Poisson algebra with the trivial Lie structure, and any graded Lie algebra can be seen as a Poisson algebra with the trivial product.

**Example 2** The most classical example of a Poisson algebra (already considered by Poisson himself) is the algebra of smooth functions on $\mathbb{R}^{2n}$ endowed with usual multiplication and with the Poisson bracket $\{f, g\} = \partial_{q^i} f \partial_{p_i} g - \partial_{q^i} g \partial_{p_i} f$, where the $p_i$'s and the $q^i$'s, for $i = 1, \ldots, n$, are coordinates on $\mathbb{R}^{2n}$. The bivector field $\partial_{q^i} \wedge \partial_{p_i}$ is induced by the symplectic form $\omega = \mathrm{d}p_i \wedge \mathrm{d}q^i$. An immediate generalization of this example is the algebra of smooth functions on a symplectic manifold $(\mathbb{R}^{2n}, \omega)$ with the Poisson bracket $\{f, g\} = \omega^{ij} \partial_i f \partial_j g$, where $\omega^{ij} \partial_i \wedge \partial_j$ is the bivector field defined by the inverse of the symplectic form $\omega = \omega_{ij} \mathrm{d}x^i \wedge \mathrm{d}x^j$; viz. $\omega_{ij} \omega^{jk} = \delta_i^k$.

A further generalization is when the bracket on $C^\infty(\mathbb{R}^m)$ is defined by $\{f, g\} = \alpha^{ij} \partial_i f \partial_j g$, with the matrix function $\alpha$ not necessarily nondegenerate. The bracket is Poisson if and only if $\alpha$ is skewsymmetric and satisfies

$$\alpha^{ij} \partial_i \alpha^{kl} + \alpha^{il} \partial_i \alpha^{jk} + \alpha^{ik} \partial_i \alpha^{lj} = 0$$

An example of this, already considered by Lie (1894), is $\alpha^{ij}(x) = f_k^{ij} x^k$, where the $f_k^{ij}$'s are the structure constants of some Lie algebra.

**Example 3** Example 2 can be generalized to any symplectic manifold $(M, \omega)$. To every function $h \in C^\infty(M)$ one associates the Hamiltonian vector field $X_h$ which is the unique vector field satisfying $i_{X_h} \omega = \mathrm{d}h$. The Poisson bracket of two functions $f$ and $g$ is then defined by

$$\{f, g\} = i_{X_f} i_{X_g} \omega$$

In local coordinates, the corresponding Poisson bivector field is related to the symplectic form as in Example 2.

A generalization is the algebra of smooth functions on a manifold $M$ with bracket $\{f, g\} = \langle \alpha | \mathrm{d}f \wedge \mathrm{d}g \rangle$, where $\alpha$ is a bivector field (i.e., a section of $\bigwedge^2 TM$) such that $\{\alpha, \alpha\}_{SN} = 0$, where $\{\cdot, \cdot\}_{\mathrm{SN}}$ is the Schouten–Nijenhuis bracket (see the first subsection in the next section for details, and Example 2 for the local coordinate expression). Such a bivector field is called a Poisson bivector field and the manifold $M$ is called a Poisson manifold. Observe that a Poisson algebra structure on the algebra of smooth functions on a smooth manifold is necessarily defined this way. In the symplectic case, the bivector field corresponding to the Poisson bracket is the inverse of the symplectic form (regarded as a bundle map $TM \to T^*M$).

The linear case described at the end of Example 2 corresponds to $M = \mathfrak{g}^*$ where $\mathfrak{g}$ is a (finite-dimensional) Lie algebra. The Lie bracket $\bigwedge^2 \mathfrak{g} \to \mathfrak{g}$ is regarded as an element of $\mathfrak{g} \otimes \bigwedge^2 \mathfrak{g}^* \subset \Gamma(\bigwedge^2 T\mathfrak{g}^*)$ and reinterpreted as a Poisson bivector field on $\mathfrak{g}$. The Poisson algebra structure restricted to polynomial functions is described at the beginning of the next section.

## Batalin–Vilkovisky Algebras

When $n$ is odd, a generator for the bracket of an $n$-Poisson algebra $A$ is a degree $-n$ linear map from $A$ to itself,

$$\Delta : A \to A[-n]$$

such that

$$\Delta(a \cdot b) = \Delta(a) \cdot b + (-1)^{|a|} a \cdot \Delta(b) + (-1)^{|a|} \{a, b\}$$

A generator $\Delta$ is called exact if and only if it satisfies the condition $\Delta^2 = 0$, and in this case $\Delta$ becomes a derivation of the bracket:

$$\Delta(\{a, b\}) = \{\Delta(a), b\} + (-1)^{|a|+1} \{a, \Delta(b)\}$$

**Remark 1** Notice that not every odd Poisson algebra $A$ admits a generator. For instance, a nontrivial odd Lie algebra seen as an odd Poisson algebra with trivial multiplication admits no generator. Moreover, even if a generator $\Delta$ for an odd Poisson algebra exists, it is far from being unique. In fact, all different generators are obtained by adding to $\Delta$ a derivation of $A$ of degree $-n$.

**Definition 2** An $n$-Poisson algebra $A$ is called an $n$-Batalin–Vilkovisky algebra, if it is endowed with an exact generator.

*Notation*. When $n = 1$ it is customary to speak of Batalin–Vilkovisky algebras, or simply BV algebras (*see* Batalin–Vilkovisky Quantization; see also Batalin and Vilkovisky (1963), Getzler (1994), and Koszul (1985)).

There exists a characterization of $n$-Batalin–Vilkovisky algebras in terms of only the product and the generator (Getzler 1994, Koszul (1985). Suppose in fact that a graded vector space $A$ is endowed with a degree zero graded commutative product and a linear map $\Delta : A \to A[-n]$ such that $\Delta^2 = 0$, satisfying the following "seven-term" relation:

$$\begin{aligned}\Delta(a\cdot b\cdot c) + \Delta(a)\cdot b\cdot c + (-1)^{|a|}a\cdot\Delta(b)\cdot c\\ + (-1)^{|a|+|b|}a\cdot b\cdot\Delta(c)\\ = \Delta(a\cdot b)\cdot c + (-1)^{|a|}a\cdot\Delta(b\cdot c)\\ + (-1)^{(|a|+1)|b|}b\cdot\Delta(a\cdot c)\end{aligned}$$

In other words, $\Delta$ is a derivation of order 2.

Then, if we define the bilinear operation $\{,\} : A\otimes A \to A[-n]$ by

$$\{a,b\} = (-1)^{|a|}\Big(\Delta(a\cdot b) - \Delta(a)\cdot b - (-1)^{|a|}a\cdot\Delta(b)\Big)$$

we have that the quadruple $(A,\cdot,\{,\},\Delta)$ is an $n$-Batalin–Vilkovisky algebra. Conversely, one easily checks that the product and the generator of an $n$-Batalin–Vilkovisky algebra satisfy the above "seven-term" relation.

## Examples

### Schouten–Nijenhuis Bracket

Suppose $\mathfrak{g}$ is a graded Lie algebra of degree zero. Then $A = S(\mathfrak{g}[n])$ is a $(-n)$-Poisson algebra with its natural multiplication (the one induced from the tensor algebra $T(A)$) and a degree $-n$ bracket defined as follows (Koszul 1985, Krasil'shchik 1988): the bracket on $S^1(\mathfrak{g}[n]) = \mathfrak{g}[n]$ is defined as the suspension of the bracket on $\mathfrak{g}$, while on $S^k(\mathfrak{g}[n])$, for $k > 1$, the bracket, often called the Schouten–Nijenhuis bracket, is defined inductively by forcing the Leibniz rule

$$\{a, b\cdot c\} = \{a,b\}\cdot c + (-1)^{|b|(|a|+n)}b\cdot\{a,c\}$$

Moreover, when $n$ is odd, there exists a generator defined as

$$\Delta(a_1\cdot a_2\cdots a_k) = \sum_{i<j}(-1)^{\epsilon}\{a_i,a_j\}\cdot a_1\cdots\widehat{a_i}\cdots\widehat{a_j}\cdots a_k$$

where $a_1,\ldots,a_k \in \mathfrak{g}$ and $\epsilon = |a_i| + (|a_i|+1)(|a_1|+\cdots+|a_{i-1}|+i-1) + (|a_j|+1)(|a_1|+\cdots+\widehat{|a_i|}+\cdots+|a_{j-1}|+j-2)$. An easy check shows that $\Delta^2 = 0$, thus $S(\mathfrak{g}[n])$ is an $n$-Batalin–Vilkovisky algebra for every odd $n \in \mathbb{N}$. For $n = -1$ the $\Delta$-cohomology on $\bigwedge\mathfrak{g}$ is the usual Cartan–Chevalley–Eilenberg cohomology.

In particular, one can consider the Lie algebra $\mathfrak{g} = \mathrm{Der}(B) = \oplus_{j\in\mathbb{Z}}\mathrm{Der}^j(B)$ of derivations of a graded commutative algebra $B$. More explicitly, $\mathrm{Der}^j(B)$ consists of linear maps $\phi : B \to B$ of degree $j$ such that $\phi(ab) = \phi(a)b + (-1)^{j|a|}a\phi(b)$ and the bracket is $\{\phi,\psi\} = \phi\circ\psi - (-1)^{|\phi||\psi|}\psi\circ\phi$. The space of multiderivations $S(\mathrm{Der}(B)[-1])$, endowed with the Schouten–Nijenhuis bracket, is a Gerstenhaber algebra.

We can further specialize to the case when $B$ is the algebra $\mathcal{C}^\infty(M)$ of smooth functions on a smooth manifold $M$; then $\mathfrak{X}(M) = \mathrm{Der}(\mathcal{C}^\infty(M))$ is the space of vector fields on $M$ and $\mathcal{V}(M) = S(\mathfrak{X}(M)[-1])$ is the space of multivector fields on $M$. It is a classical result by Koszul (1985) that there is a bijective correspondence between generators for $\mathcal{V}(M)$ and connections on the highest exterior power $\bigwedge^{\dim M}TM$ of the tangent bundle of $M$. Moreover, flat connections correspond to generators which square to zero.

### Lie Algebroids

A Lie algebroid $E$ over a smooth manifold $M$ is a vector bundle $E$ over $M$ together with a Lie algebra structure (over $\mathbb{R}$) on the space $\Gamma(E)$ of smooth sections of $E$, and a bundle map $\rho : E \to TM$, called the anchor, extended to a map between sections of these bundles, such that

$$\{X, fY\} = f\{X,Y\} + (\rho(X)f)Y$$

for any smooth sections $X$ and $Y$ of $E$ and any smooth function $f$ on $M$. In particular, the anchor map induces a morphism of Lie algebras $\rho_* : \Gamma(E) \to \mathfrak{X}(M)$, namely $\rho_*(\{X,Y\}) = \{\rho_*(X), \rho_*(Y)\}$.

The link between Lie algebroids and Gerstenhaber algebras is given by the following Proposition (Kosmann-Schwarzbach and Monterde 2002, Xu 1999):

**Proposition 1** *Given a vector bundle $E$ over $M$, there exists a one-to-one correspondence between Gerstenhaber algebra structures on $A = \Gamma(\bigwedge(E))$ and Lie algebroid structures on $E$.*

The key of the proposition is that one can extend the Lie algebroid bracket to a unique graded antisymmetric bracket on $\Gamma(\bigwedge(E))$ such that $\{X,f\} = \rho(X)f$ for $X \in \Gamma(\bigwedge^1(E))$ and $f \in \Gamma(\bigwedge^0(E))$, and that for $Q \in \Gamma(\bigwedge^{q+1}(E))$, $\{Q,\cdot\}$ is a derivation of $\Gamma(\bigwedge(E))$ of degree $q$.

**Example 4** A finite-dimensional Lie algebra $\mathfrak{g}$ can be seen as a Lie algebroid over a trivial base manifold. The corresponding Gerstenhaber algebra is the one of last subsection.

**Example 5** The tangent bundle $TM$ of a smooth manifold $M$ is a Lie algebroid with anchor map given by the identity and algebroid Lie bracket given by the usual Lie bracket on vector fields. In this case, we recover the Gerstenhaber algebra of multivector fields on $M$ described in the last subsection.

**Example 6** If $M$ is a Poisson manifold with Poisson bivector field $\alpha$, then the cotangent bundle $T^*M$ inherits a natural Lie algebroid structure where the anchor map $\alpha^\# : T_p^*M \to T_pM$ at the point $p \in M$ is given by $\alpha^\#(\xi)(\eta) = \alpha(\xi, \eta)$, with $\xi, \eta \in T_p^*M$, and the Lie bracket of the 1-forms $\omega_1$ and $\omega_2$ is given by

$$\{\omega_1, \omega_2\} = L_{\alpha^\#(\omega_1)}\omega_2 - L_{\alpha^\#(\omega_2)}\omega_1 - \mathrm{d}\alpha(\omega_1, \omega_2)$$

The associated Gerstenhaber algebra is the de Rham algebra of differential forms endowed with the bracket defined by Koszul (1985). As shown in Kosmann-Schwarzbach (1995), $\Gamma(\bigwedge(T^*M))$ is indeed a BV algebra with an exact generator $\Delta = [\mathrm{d}, i_\alpha]$ given by the commutator of the contraction $i_\alpha$ with the Poisson bivector $\alpha$ and the de Rham differential d. Similar results hold if $M$ is a Jacobi manifold.

It is natural to ask what additional structure on a Lie algebroid $E$ makes the Gerstenhaber algebra $\Gamma(\bigwedge(E))$ into a BV algebra. The answer is given by the following result, which is proved in Xu (1999).

**Proposition 2** *Given a Lie algebroid E, there is a one-to-one correspondence between generators for the Gerstenhaber algebra $\Gamma(\bigwedge(E))$ and E-connections on $\bigwedge^{\mathrm{rk}\,E}E$ (where rk E denotes the rank of the vector bundle E). Exact generators correspond to flat E-connections , and in particular, since flat E-connections always exist, $\Gamma(\bigwedge(E))$ is always a BV algebra.*

### Lie Algebroid Cohomology

A Lie algebroid structure on $E \to M$ defines a differential $\delta$ on $\Gamma(\bigwedge E^*)$ by

$$\delta f := \rho^* \mathrm{d}f, \qquad f \in C^\infty(M) = \Gamma(\Lambda^0 E^*)$$

and

$$\langle \delta\alpha, X \wedge Y\rangle := \langle \delta\langle\alpha, X\rangle, Y\rangle - \langle \delta\langle\alpha, Y\rangle, X\rangle \\ - \langle \alpha, \{X, Y\}\rangle X, Y \in \Gamma(E), \alpha \in \Gamma(E^*)$$

where $\rho^* : \Omega^1(M) \to \Gamma(E^*)$ is the transpose of $\rho_* : \Gamma(E) \to \mathfrak{X}(M)$ and $\langle\,,\rangle$ is the canonical pairing of sections of $E^*$ and $E$. On $\Gamma(\bigwedge^n E^*)$, with $n \geq 2$, the differential $\delta$ is defined by forcing the Leibniz rule.

In Example 4 we get the Cartan–Chevalley–Eilenberg differential on $\bigwedge \mathfrak{g}^*$; in Example 5 we recover the de Rham differential on $\Omega^*(M) = \Gamma(\bigwedge T^*M)$, while in Example 6 the differential on $\mathcal{V}(M) = \Gamma(\bigwedge TM)$ is $\{\alpha, \}_{\mathrm{SN}}$.

### Lie–Rinehart Algebras

The algebraic generalization of a Lie algebroid is a Lie–Rinehart algebra. Recall that given a commutative associative algebra $B$ (over some ring $R$) and a $B$-module $\mathfrak{g}$, then a Lie–Rinehart algebra structure on $(B, \mathfrak{g})$ is a Lie algebra structure (over $R$) on $\mathfrak{g}$ and an action of $\mathfrak{g}$ on the left on $B$ by derivations, satisfying the following compatibility conditions:

$$\{\gamma, a\sigma\} = \gamma(a)\delta + a\{\gamma, \sigma\} \\ (a\gamma)(b) = a(\gamma(b))$$

for every $a, b \in B$ and $\gamma, \sigma \in \mathfrak{g}$.

The Lie–Rinehart structures on the pair $(B, \mathfrak{g})$ bijectively correspond to the Gerstenhaber algebra structures on the exterior algebra $\bigwedge_B(\mathfrak{g})$ of $\mathfrak{g}$ in the category of $B$-modules. When $\mathfrak{g}$ is of finite rank over $B$, generators for these structures are in turn in bijective correspondence with $(B, \mathfrak{g})$-connections on $\bigwedge_B^{\mathrm{rk}} B^{\mathfrak{g}}\mathfrak{g}$, and flat connections correspond to exact generators. For additional discussions, see Gerstenhaber and Schack (1992) and Huebschmann (1998).

Lie algebroids are Lie–Rinehart algebras in the smooth setting. Namely, if $E \to M$ is a Lie algebroid, then the pair $(\mathcal{C}^\infty(M), \Gamma(E))$ is a Lie–Rinehart algebra (with action induced by the anchor and the given Lie bracket).

### Lie–Rinehart Cohomology

Lie algebroid cohomology may be generalized to every Lie–Rinehart algebra $(B, \mathfrak{g})$. Namely, on the complex $\mathrm{Alt}_B(\mathfrak{g}, B)$ of alternating multilinear functions on $\mathfrak{g}$ with values in $B$, one can define a differential $\delta$ by the rules

$$\langle \delta a, \gamma\rangle = \gamma(a), \qquad a \in B = \mathrm{Alt}_B^0(\mathfrak{g}, B), \quad \gamma \in \mathfrak{g} \\ \langle \delta a, \gamma \wedge \sigma\rangle = \langle \delta\langle a, \gamma\rangle, \sigma\rangle - \langle \delta\langle a, \sigma\rangle, \gamma\rangle - \langle a, \{\gamma, \sigma\}\rangle \\ \gamma, \sigma \in \mathfrak{g}, \quad a \in \mathrm{Alt}_B^1(\mathfrak{g}, B)$$

and forcing the Leibniz rule on elements of $\mathrm{Alt}_B^n(\mathfrak{g}, B)$, $n \geq 2$.

### Hochschild Cohomology

Let $A$ be an associative algebra with product $\mu$, and consider the Hochschild cochain complex $\mathsf{Hoch}(A) = \prod_{n\geq 0} \mathrm{Hom}(A^{\otimes n}, A)[-n+1]$. There are two basic operations between two elements $f \in \mathrm{Hom}(A^{\otimes k}, A)[-k+1]$ and $g \in \mathrm{Hom}(A^{\otimes l}, A)[-l+1]$, namely a degree zero product

$$f \cup g(a_1 \otimes \cdots \otimes a_{k+l}) \\ = (-1)^{kl} f(a_1 \otimes \cdots \otimes a_l) \cdot g(a_{l+1} \otimes \cdots \otimes a_{k+l})$$

and a degree $-1$ bracket $\{f,g\}=f\circ g-(-1)^{(k-1)(l-1)}g\circ f$, where

$$f\circ g(a_1\otimes\cdots\otimes a_{k+l-1})\\=\sum_{i=1}^{k-1}(-1)^{i(l-1)}f(a_1\otimes\cdots\otimes a_i\otimes g(a_{i+1}\\\otimes\cdots\otimes a_{i_l})\otimes\cdots\otimes a_{k+l-1})$$

It is well known from Gerstenhaber (1963) that the cohomology $\mathsf{HHoch}(A)$ of the Hochschild complex with respect to the differential $d_{\mathsf{Hoch}}=\{\mu,\cdot\}$ has the structure of a Gerstenhaber algebra. More generally, there is a Gerstenhaber algebra structure on Hochschild cohomology of differential graded associative algebras (Loday 1998).

### Graded Symplectic Manifolds

The construction of Example 3 can be extended to graded symplectic manifolds (*see* Supermanifolds; see also Alexandrov *et al.* (1997), Getzler (1994), and Schwarz (1993)). Recall that a symplectic structure of degree $n$ on a graded manifold $N$ is a closed nondegenerate 2-form $\omega$ such that $L_E\omega=n\omega$ where $L_E$ is the Lie derivative with respect to the Euler field of $N$ (see Roytenberg (2002) for details). Let us denote by $X_h$ the vector field associated to the function $h\in\mathcal{C}^\infty(N)$ by the formula $i_{X_h}\omega=dh$. Then the bracket

$$\{f,g\}=i_{X_f}i_{X_g}\omega$$

gives $\mathcal{C}^\infty(N)$ the structure of a graded Poisson algebra of degree $n$.

If the symplectic form has odd degree and the graded manifold has a volume form, then it is possible to construct an exact generator defined by

$$\Delta(f)=\tfrac{1}{2}\mathrm{div}(X_f)$$

where div is the divergence operator associated to the given volume form (Getzler 1994, Kosmann-Schwazbach and Monterde 2002).

An explicit characterization of graded symplectic manifolds has been given in Roytenberg (2002). In particular, it is proved there that every symplectic form of degree $n$ with $n\geq 1$ is necessarily exact. More precisely, one has $\omega=d(i_E\omega/n)$.

### Shifted Cotangent Bundles

The main examples of graded symplectic manifolds are given by shifted cotangent bundles. If $N$ is a graded manifold then the shifted cotangent bundle $T^*[n]N$ is the graded manifold obtained by shifting by $n$ the degrees of the fibers of the cotangent bundle of $N$. This graded manifold possesses a nondegenerate closed 2-form of degree $n$, which can be expressed in local coordinates as

$$\omega=\sum_i dx^i\wedge dx_i^\dagger$$

where $\{x^i\}$ are local coordinates on $N$ and $\{x_i^\dagger\}$ are coordinate functions on the fibers of $T^*[n]N$. In local coordinates, the bracket between two homogeneous functions $f$ and $g$ is given by

$$\{f,g\}=-(-1)^{|x_i^\dagger||f|}\frac{\partial f}{\partial x_i^\dagger}\frac{\partial g}{\partial x^i}\\-(-1)^{(|f|+n)(|g|+n)+|x_i^\dagger||g|}\frac{\partial g}{\partial x_i^\dagger}\frac{\partial f}{\partial x^i}$$

If in addition the graded manifold $N$ is orientable, then $T^*[n]N$ has a volume form too; when $n$ is odd, the exact generator $\Delta(f)=(1/2)\mathrm{div}X_f$ is written in local coordinates as

$$\Delta=\frac{\partial}{\partial x_i^\dagger}\frac{\partial}{\partial x^i}$$

In the case $n=1$, we have a natural identification between functions on $T^*[1]N$ and multivector fields $\mathcal{V}(N)$ on $N$, and we recover again the Gerstenhaber algebra of the subsection "Schouten–Nijenhuis bracket." Moreover, it is easy to see that, under the above identification, $\Delta$ applied to a vector field of $N$ is the usual divergence operator.

### Examples from Algebraic Topology

For any $n>1$, the homology of the $n$-fold loop space $\Omega^n(M)$ of a topological space $M$ has the structure of an $(n-1)$-Poisson algebra (May 1972). In particular, the homology of the double loop space $\Omega^2(M)$ is a Gerstenhaber algebra, and has an exact generator defined using the natural circle action on this space (Getzler 1994). The homology of the free loop space $\mathcal{L}(M)$ of a closed oriented manifold $M$ is also a BV algebra when endowed with the "Chas–Sullivan intersection product" and with a generator defined again using the natural circle action on the free loop space (Cohen and Jones 2002).

## Applications

### BRST Quantization in the Hamiltonian Formalism

The BRST procedure is a method for quantizing classical mechanical systems or classical field theories in the presence of symmetries (*see* BRST Quantization). The starting point is a symplectic manifold $M$ (the "phase space"), a function $H$ (the "Hamiltonian" of the system) governing the evolution of the system, and the "constraints" given by

several functions $g_i$ which commute with $H$ and among each other up to a $\mathcal{C}^\infty(M)$-linear combination of the $g_i$'s.

Then the dynamics is constrained on the locus $V$ of common zeros of the $g_i$'s. When $V$ is a submanifold, the $g_i$'s are a set of generators for the ideal $I$ of functions vanishing on $V$. Observe that $I$ is closed under the Poisson bracket. Functions in $I$ are called "first class constraints." The Hamiltonian vector fields of first-class constraints, which by construction tangential to $V$, are the "symmetries" of the system.

When $V$ is smooth, then it is a coisotropic submanifold of $M$ and the Hamiltonian vector fields determined by the constraints give a foliation $\mathcal{F}$ of $V$. In the nicest case $V$ is a principal bundle with $\mathcal{F}$ its vertical foliation and the algebra of functions $\mathcal{C}^\infty(V/\mathcal{F})$ on the "reduced phase space" (*see* Poisson Reduction, and Symmetry and Symplectic Reduction) $V/\mathcal{F}$ is identified with the $I$-invariant subalgebra of $\mathcal{C}^\infty(M)/I$.

From a physical point of view, the points of $V/\mathcal{F}$ are the interesting states at a classical level, and a quantization of this system means a quantization of $\mathcal{C}^\infty(V/\mathcal{F})$. The BRST procedure gives a method of quantizing $\mathcal{C}^\infty(V/\mathcal{F})$ starting from the (known) quantization of $\mathcal{C}^\infty(M)$. Notice that these notions immediately generalize to graded symplectic manifolds.

From an algebraic point of view, one starts with a graded Poisson algebra $P$ and a multiplicative ideal $I$ which is closed under the Poisson bracket. The algebra of functions on the "reduced phase space" is replaced by $(P/I)^I$, the $I$-invariant subalgebra of $P/I$. This subalgebra inherits a Poisson bracket even if $P/I$ does not. Moreover, the pair $(B,\mathfrak{g})=(P/I,I/I^2)$ inherits a graded Lie–Rinehart structure. The "Rinehart complex" $\mathrm{Alt}_{P/I}(I/I^2,P/I)$ of alternating multilinear functions on $I/I^2$ with values in $P/I$, endowed with the differential described in the subsection "Lie-Rinehart cohomology," plays the role of the de Rham complex of vertical forms on $V$ with respect to the foliation $\mathcal{F}$ determined by the constraints.

In case $V$ is a smooth submanifold, we also have the following geometric interpretation: let $N^*V$ denote the conormal bundle of $V$ (i.e., the annihilator of $TV$ in $T_VP$). This is a Lie subalgebroid of $T^*P$ if and only if $V$ is coisotropic. Since we may identify $I/I^2$ with sections of $N^*C$ (by the de Rham differential), $(P/I,I/I^2)$ is the corresponding Lie–Rinehart pair. The Rinehart complex is then the corresponding Lie algebroid complex $\Gamma(\bigwedge(N^*V)^*)$ with differential described in the subsection "Lie algebroid cohomology." The image of the anchor map $N^*V\to TV$ is the distribution determining $\mathcal{F}$, so by duality we get an injective chain map from the vertical de Rham complex to the Rinehart complex.

The main point of the BRST procedure is to define a chain complex $C^\bullet=\bigwedge(\Psi^*\oplus\Psi)\otimes P$, where $\Psi$ is a graded vector space, with a coboundary operator $D$ (the "BRST operator"), and a quasi-isomorphism (i.e., a chain map that induces an isomorphism in cohomology)

$$\pi:(C^\bullet,D)\to\left(\mathrm{Alt}_{P/I}(I^2/I,P/I),\mathrm{d}\right)$$

This means in particular that the zeroth cohomology $H^0_D(C)$ gives the algebra $(P/I)^I$ of functions on the "reduced phase space." Observe that there is a natural symmetric inner product on $\Psi^*\oplus\Psi$ given by the evaluation of $\Psi^*$ on $\Psi$. This inner product, as an element of $S^2(\Psi\oplus\Psi^*)\simeq S^2(\Psi)\oplus(\Psi\otimes\Psi^*)\oplus S^2(\Psi^*)$, is concentrated in the component $\Psi\otimes\Psi^*$, and so it defines an element in $\bigwedge^2(\Psi[1]\oplus\Psi^*[-1])\simeq S^2(\Psi)[2]\oplus(\Psi\otimes\Psi^*)\oplus S^2(\Psi^*)[-2]$, that is, a degree zero bivector field on $\Psi[1]\oplus\Psi^*[-1]$. It is easy to see that this bivector field induces a degree zero Poisson structure on $S(\Psi^*[-1]\oplus\Psi[1])$. From another viewpoint this is the Poisson structure corresponding to the canonical symplectic structure on $T^*\Psi[1]$. Finally, we have that $S(\Psi^*[-1]\oplus\Psi[1])\otimes P$ is a degree zero Poisson algebra. Note that the superalgebra underlying the graded algebra $S(\Psi^*[-1]\oplus\Psi[1])\otimes P$ is canonically isomorphic to the complex $C^\bullet=\bigwedge(\Psi^*\oplus\Psi)\otimes P$. When $P=\mathcal{C}^\infty(M)$, we can think of $S(\Psi^*[-1]\oplus\Psi[1])\otimes\mathcal{C}^\infty(M)$ as the algebra of functions on the graded symplectic manifold $N=(\Psi[1]\oplus\Psi^*[-1])\times M$ (the "extended phase space"). In physical language, coordinate functions on $\Psi[1]$ are called "ghost fields" while coordinate functions on $\Psi^*[-1]$ are called "ghost momenta" or, by some authors, "antighost fields" (not to be confused with the antighosts of the Lagrangian functional-integral approach to quantization).

Suppose now that there exists an element $\Theta\in S(\Psi^*[-1]\oplus\Psi[1])\otimes P$ such that $\{\Theta,\cdot\}=D$, that one can extend the "known" quantization of $P$ to a quantization of $S(\Psi^*[-1]\oplus\Psi[1])\otimes P$ as operators on some (graded) Hilbert space $\mathcal{T}$ and that the operator $Q$ which quantizes $\Theta$ has square zero. Then one can consider the "true space of physical states" $H^0_Q(\mathcal{T})$ on which the $\mathrm{ad}_Q$-cohomology of operators will act. This provides one with a quantization of $(P/I)^I$.

For further details on this procedure, and in particular for the construction of $D$, we refer to Henneaux and Teitelboim (1992), Kostant and Sternberg (1987), and Stasheff (1997), and references therein. Observe that some authors refer to this method as BVF (Batalin–Vilkovisky–Fradkin) and reserve the name BRST for the case when the $g_i$'s are the components of an equivariant moment map.

For a generalization to graded manifolds different from $(\Psi[1] \oplus \Psi^*[-1]) \times M$ we refer to Roytenberg (2002). There it is proved that the element $\Theta$ exists if the graded symplectic form has degree different from –1.

#### BV Quantization in the Lagrangian Formalism

The BV formalism (*see* Batalin–Vilkovisky Quantization; see also Batalin and Vilkovisky (1983) and Henneaux and Teitelboim (1992)) is a procedure for the quantization of physical systems with symmetries in the Lagrangian formalism. As a first step, the "configuration space" $M$ of the system is augmented by the introduction of "ghosts." If $G$ is the group of symmetries, this means that one has to consider the graded manifold $W = \mathfrak{g}[1] \times M$. The second step is to double this space by introducing "antifields for fields and ghosts," namely one has to consider the "extended configuration space" $T^*[-1]W$, whose space of functions is a BV algebra (see the subsection "Shifted cotangent bundle." The algebra of "observables" is by definition the cohomology $H^*_\Delta(\mathcal{C}^\infty(T^*[-1]W))$ with respect to the exact generator $\Delta$.

## Related Topics

### AKSZ

The graded manifold $T^*[-1]W$ considered above is a particular example of a $QP$-manifold, that is, of a graded manifold $M$ endowed with an integrable (i.e., self-commuting) vector field $Q$ of degree 1 and a graded $Q$-invariant symplectic structure $P$. In quantization of classical mechanical theories, the graded symplectic manifold of interest is the space of fields and antifields with symplectic form of degree 1, while $Q$ is the Hamiltonian vector field defined by the action functional $S$; the integrability of $Q$ is equivalent to the classical master equation $\{S, S\} = 0$ for the action functional. Quantization of the theory is then reduced to the computation of the functional integral $\int_{\mathcal{L}} \exp(\mathrm{i}S/\hbar)$, where $\mathcal{L}$ is a Lagrangian submanifold of $M$. This functional integral actually depends only on the homology class of the Lagrangian. Locally, a $QP$ manifold is a shifted cotangent bundle $T^*[-1]N$ and a Lagrangian submanifold is the graph of an exact 1-form. In the notations of the subsection "Shifted cotangent bundle," a Lagrangian submanifold $\mathcal{L}$ is therefore locally defined by equations $x_i^\dagger = \partial\Phi/\partial x^i$, and the function $\Phi$ is called a gauge-fixing fermion. The action functional of interest is then the gauge-fixed action $S|_{\mathcal{L}} = S(x^i, \partial\Phi/\partial x^i)$.

The language of $QP$ manifolds has powerful applications to sigma models (*see* Topological Sigma Models): if $\Sigma$ is a finite-dimensional graded manifold equipped with a volume element, and $M$ is a $QP$ manifold, then the graded manifold $C^\infty(\Sigma, M)$ of smooth maps from $\Sigma$ to $M$ has a natural structure of $QP$ manifold which describes some field theory if one arranges for the symplectic structure to be of degree 1. As an illustrative example, if $\Sigma = T[1]X$, for a compact oriented three-dimensional smooth manifold $X$, and $M = \mathfrak{g}[1]$, where $\mathfrak{g}$ is the Lie algebra of a compact Lie group, the $QP$ manifold $C^\infty(\Sigma, M)$ is relevant to Chern–Simons theory on $X$. Similarly, if $\Sigma = T[1]X$, for a compact oriented two-dimensional smooth manifold $X$ and $M = T[1]N$ is the shifted tangent bundle of a symplectic manifold, then the $QP$ structure on $C^\infty(\Sigma, M)$ is related to the A-model with target $N$; if the symplectic manifold $N$ is of the form $N = T^*K$ for a complex manifold $K$, then one can endow $C^\infty(\Sigma, M)$ with a complex $QP$ manifold structure, which is related to the B-model with target $K$; this shows that, in some sense, the B-model can be obtained from the A-model by "analytic continuation" (Alexandrov *et al.* 1997). If $\Sigma = T[1]X$, for a compact oriented two-dimensional smooth manifold $X$ and $M = T^*[1]N$ with canonical symplectic structure, then the $QP$ structure on $C^\infty(\Sigma, M)$ is related to the Poisson sigma model ($QP$ structures on $T^*[1]N$ with canonical symplectic structure are in one-to-one correspondence with Poisson structures on $N$). The study of $QP$ manifolds is sometimes referred to as "the AKSZ formalism". In Roytenberg (2002) $QP$ manifolds with symplectic structure of degree 2 are studied and shown to be in one-to-one correspondence with Courant algebroids.

### Graded Poisson Algebras from Cohomology of $P_\infty$

The Poisson bracket on a Poisson manifold can be derived from the Poisson bivector field $\alpha$ using the Schouten–Nienhuis bracket as follows:

$$\{f, g\} = \{\{\alpha, f\}_{\mathrm{SN}}, g\}_{\mathrm{SN}}$$

This may be generalized to the case of a graded manifold $M$ endowed with a multivector field $\alpha$ of total degree 2 (i.e., $\alpha = \sum_{i=0}^\infty \alpha_i$, where $\alpha_i$ is an $i$-vector field of degree $2 - i$) satisfying the equation $\{\alpha, \alpha\}_{SN} = 0$. One then has the derived multibrackets

$$\begin{aligned} &\lambda_i : A^{\otimes i} \to A \\ &\lambda_i(a_1, \ldots, a_i) \\ &\quad := \{\{\ldots\{\{\alpha_i, a_1\}_{\mathrm{SN}}, a_2\}_{\mathrm{SN}} \ldots\}_{\mathrm{SN}}, a_i\}_{\mathrm{SN}} \end{aligned}$$

with $A = \mathcal{C}^\infty(M)$. Observe that $\lambda_i$ is a multiderivation of degree $2 - i$. The operations $\lambda_i$ define the structure of an $L_\infty$-algebra on $A$. Such a structure is called a $P_\infty$-algebra ($P$ for Poisson) since the $\lambda_i$'s are multiderivations. If $\lambda_0 = \alpha_0$ vanishes, then $\lambda_1$ is a differential, and the $\lambda_1$-cohomology inherits a graded Poisson algebra structure. This structure can be used

to describe the deformation quantization of coisotropic submanifolds and to describe their deformation theory.

## Acknowledgments

The authors thank Johannes Huebschmann, Yvette Koszmann-Schwarzbach, Dmitry Roytenberg, and Marco Zambon for very useful comments and suggestions.

A S C acknowledges partial support of SNF Grant No. 200020-107444/1.

*See also:* Batalin–Vilkovisky Quantization; BRST Quantization; Poisson Reduction; Supermanifolds; Symmetry and Symplectic Reduction; Topological Sigma Models.

## Further Reading


Alexandrov M, Schwarz A, Zaboronsky O, and Kontsevich M (1997) The geometry of the master equation and topological quantum field theory. *International Journal of Modern Physics A* 12(7): 1405–1429.

Batalin IA and Vilkovisky GA (1981) Gauge algebra and quantization. *Physics Letters B* 102(1): 27–31.

Batalin IA and Vilkovisky GA (1983) Quantization of gauge theories with linearly dependent generators. *Physical Review D (3)* 28(10): 2567–2582.

Cohen RL and Jones JDS (2002) A homotopy theoretic realization of string topology. *Mathematische Annalen* 324(4): 773–798.

Gerstenhaber M (1963) The cohomology structure of an associative ring. *Annals of Mathematics* 78: 267–288.

Gerstenhaber M and Schack SD (1992) Algebras, bialgebras, quantum groups, and algebraic deformations. In: *Deformation Theory and Quantum Groups with Applications to Mathematical Physics (Amherst, MA, 1990)*, Contemp. Math., vol. 134, pp. 51–92. Providence, RI: American Mathematical Society.

Getzler E (1994) Batalin–Vilkovisky algebras and two-dimensional topological field theories. *Communications in Mathematical Physics* 159(2): 265–285.

Henneaux M and Teitelboim C (1992) *Quantization of Gauge Systems*. Princeton, NJ: Princeton University Press.

Huebschmann J (1998) Lie–Rinehart algebras, Gerstenhaber algebras and Batalin–Vilkovisky algebras. *Annales de l'Institut Fourier (Grenoble)* 48(2): 425–440.

Kosmann-Schwarzbach Y (1995) Exact Gerstenhaber algebras and Lie bialgebroids. *Acta Applicandae Mathematicae* 41(1–3): 153–165.

Kosmann-Schwarzbach Y and Monterde J (2002) Divergence operators and odd Poisson brackets. *Annales de l'Institut Fourier (Grenoble)* 52(2): 419–456.

Kostant B and Sternberg S (1987) Symplectic reduction, BRS cohomology, and infinite-dimensional Clifford algebras. *Annals of Physics* 176(1): 49–113.

Koszul J-L (1985) Crochet de Schouten-Nijenhuis et cohomologie. *Astérisque* (Numero Hors Serie) 257–271.

Krasil'shchik IS (1988) Schouten bracket and canonical algebras. In: *Global Analysis – Studies and Applications, III*, Lecture Notes in Math., vol. 1334, pp. 79–110. Berlin: Springer.

Lie S (1890) *Theorie der Transformationsgruppen*. Leipzig: Teubner.

Loday J-L (1998) *Cyclic Homology*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] vol. 301. Berlin: Springer.

May JP (1972) *The Geometry of Iterated Loop Spaces*. Berlin: Springer.

Roytenberg D (2002) On the structure of graded symplectic supermanifolds and Courant algebroids. In: *Quantization, Poisson Brackets and Beyond (Manchester, 2001)*, Contemp. Math., vol. 315, pp. 169–185. Providence, RI: American Mathematical Society.

Schwarz A (1993) Geometry of Batalin–Vilkovisky quantization. *Communications in Mathematical Physics* 155(2): 249–260.

Stasheff J (1997) Homological reduction of constrained Poisson algebras. *Journal of Differential Geometry* 45(1): 221–240.

Xu P (1999) Gerstenhaber algebras and BV-algebras in Poisson geometry. *Communications in Mathematical Physics* 200(3): 545–560.


# Gravitational Lensing

**J Wambsganss**, Universität Heidelberg, Heidelberg, Germany



## Introduction

Einstein's theory of general relativity states that gravity attracts light. The deflection angle of a light ray by an object with mass $m$ was predicted to be

$$\tilde{\alpha} = \frac{4Gm}{c^2 r} \qquad [1]$$

where $c$ and $G$ are the velocity of light and the gravitational constant, respectively, and $r$ is the impact parameter. The quantitative measurement of this light deflection at the solar limb during the solar eclipse in 1919 with

$$\tilde{\alpha} = \frac{4GM_\odot}{c^2 R_\odot} \approx 1.74 \text{ arcsec} \qquad [2]$$

(here $m$ is replaced by the solar mass $M_\odot$ and the impact parameter is the solar radius $R_\odot$) confirmed Einstein's theory.

In the decades following this measurement, various aspects of the gravitational lens effect were explored theoretically, which include (1) the possibility of multiple or ring-like images of background sources, (2) the use of lensing as a gravitational telescope on very faint and distant objects, and (3) the possibility of determining Hubble's constant with lensing. Only relatively recently – after the

discovery of the first doubly imaged quasar in 1979 – gravitational lensing became an observational science. Today gravitational lensing is a booming part of astrophysics.

Lensing has established itself as a very useful astrophysical tool with some remarkable successes: with the discovery of multiply-imaged quasars, giant luminous arcs, Einstein rings, quasar and galactic microlensing significant new results in areas as different as cosmology, physics of quasars, and galaxy structure could be reached. In this article, only the aspects of "strong lensing" can be treated. More detailed studies on strong and weak lensing can be found in the "Further reading" section.

## Basics of Gravitational Lensing

The path, the size, and the cross section of a light bundle propagating through spacetime in principle are affected by all the matter between the light source and the observer. For most practical purposes, we can assume that the lensing action is dominated by a single matter inhomogeneity at some location between source and observer. This is usually called the "thin-lens approximation": all the action of deflection is thought to take place at a single distance. This approach is valid only if the relative velocities of lens, source, and observer are small compared to the velocity of light ($v \ll c$) and if the Newtonian potential is small ($|\Phi| \ll c^2$). These two assumptions are justified in all astronomical cases of interest. The size of a galaxy, for example, is of order 50 kpc, even a cluster of galaxies is not much larger than 1 Mpc. This "lens thickness" is small compared to the typical distances of the order of few Gpc between observer and lens or lens and background quasar/galaxy, respectively. We assume that the underlying spacetime is well described by a perturbed Friedmann–Robertson–Walker metric:

$$ds^2 = \left(1 + \frac{2\Phi}{c^2}\right)c^2 dt^2 - a^2(t)\left(1 - \frac{2\Phi}{c^2}\right)d\sigma^2 \qquad [3]$$

A detailed description of optics in curved spacetimes and a derivation of the lens equation from Einstein's field equations can be found in Schneider *et al.* (1992, chapters 3 and 4).

### Lens Equation

The basic setup for such a simplified gravitational lens scenario involving a point source and a point lens is displayed in Figure 1. The three ingredients in such a lensing situation are the source S, the lens L, and the observer O. Light rays emitted from the source are deflected by the lens. For a point-like lens, there will always be (at least) two images $S_1$ and $S_2$ of the source. With external shear – due to the tidal field of objects outside but near the light bundles – there can be more images. The observer sees the images in directions corresponding to the tangents to the real incoming light paths.

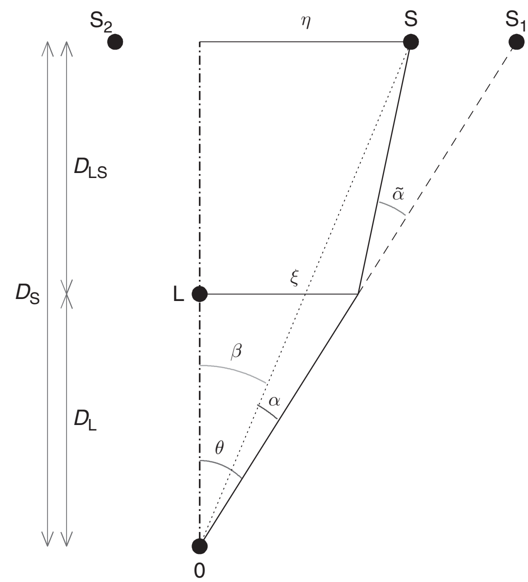



**Figure 1** The relation between the various angles and distances involved in the lensing setup can be derived for the case $\tilde{\alpha} \ll 1$ and formulated in the lens equation [6].

In Figure 1, the corresponding angles and angular diameter distances $D_L, D_S, D_{LS}$ are indicated. (In cosmology, the various methods to define distance diverge. The relevant distances for gravitational lensing are the angular diameter distances.) In the thin-lens approximation, the hyperbolic paths are approximated by their asymptotes. In the circular-symmetric case, the deflection angle is given as

$$\tilde{\alpha}(\xi) = \frac{4GM(\xi)}{c^2}\frac{1}{\xi} \qquad [4]$$

where $M(\xi)$ is the mass inside a radius $\xi$. In this depiction, the origin is chosen at the observer. From the diagram, it can be seen that the following relation holds:

$$\theta D_S = \beta D_S + \tilde{\alpha} D_{LS} \qquad [5]$$

(for $\theta, \beta, \tilde{\alpha} \ll 1$; this condition is fulfilled in practically all astrophysically relevant situations). With the definition of the reduced deflection angle as $\alpha(\theta) = (D_{LS}/D_S)\tilde{\alpha}(\theta)$, this can be expressed as

$$\beta = \theta - \alpha(\theta) \qquad [6]$$

This relation between the positions of images and source can easily be derived for a nonsymmetric mass distribution as well. In that case, all angles are

vector valued. The two-dimensional lens equation then reads

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta}) \qquad [7]$$

### Einstein Radius

For a point lens of mass $M$, the deflection angle is given by eqn [4]. Plugging this deflection angle into eqn [6] and using the relation $\xi = D_L\theta$ (cf. Figure 1), one obtains

$$\beta(\theta) = \theta - \frac{D_{LS}}{D_L D_S}\frac{4GM}{c^2\theta} \qquad [8]$$

For the special case in which the source lies exactly behind the lens ($\beta = 0$), due to the symmetry, a ring-like image occurs whose angular radius is called Einstein radius $\theta_E$:

$$\theta_E = \sqrt{\frac{4GM}{c^2}\frac{D_{LS}}{D_L D_S}} \qquad [9]$$

The Einstein radius defines the angular scale for a lens situation. For a massive galaxy with a mass of $M = 10^{12}M_\odot$ at a redshift of $z_L = 0.5$ and a source at redshift $z_S = 2.0$ (we used here $H = 50\,\mathrm{km\ s^{-1}\ Mpc^{-1}}$ as the value of the Hubble constant and an Einstein–de Sitter universe), the Einstein radius is

$$\theta_E \approx 1.8\sqrt{\frac{M}{10^{12}M_\odot}}\,\mathrm{arcsec} \qquad [10]$$

(note that for cosmological distances, in general, $D_{LS} \neq D_S - D_L$!). For a galactic microlensing scenario in which stars in the disk of the Milky Way act as lenses for stars close to its center, the scale defined by the Einstein radius is

$$\theta_E \approx 0.5\sqrt{\frac{M}{M_\odot}}\,\mathrm{marcsec} \qquad [11]$$

An application and some illustrations of the point lens case can be found in the section on galactic microlensing.

### Critical Surface Mass Density

In the more general case of a three-dimensional mass distribution of an extended lens, the density $\rho(\boldsymbol{r})$ can be projected along the line of sight onto the lens plane to obtain the two-dimensional surface mass density distribution $\Sigma(\boldsymbol{\xi})$ as

$$\Sigma(\boldsymbol{\xi}) = \int_0^{D_S} \rho(\boldsymbol{r})\,\mathrm{d}z \qquad [12]$$

Here $\boldsymbol{r}$ is a three-dimensional vector in space, and $\boldsymbol{\xi}$ is a two-dimensional vector in the lens plane. The two-dimensional deflection angle $\tilde{\boldsymbol{\alpha}}$ is then given as the sum over all mass elements in the lens plane:

$$\tilde{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \frac{4G}{c^2}\int \frac{(\boldsymbol{\xi} - \boldsymbol{\xi}')\Sigma(\boldsymbol{\xi}')}{|\boldsymbol{\xi} - \boldsymbol{\xi}'|^2}\,\mathrm{d}^2\xi' \qquad [13]$$

For a finite circle with constant surface mass density $\Sigma$, the deflection angle can be written as

$$\alpha(\xi) = \frac{D_{LS}}{D_S}\frac{4G}{c^2}\frac{\Sigma\pi\xi^2}{\xi} \qquad [14]$$

With $\xi = D_L\theta$ this simplifies to

$$\alpha(\theta) = \frac{4\pi G\Sigma}{c^2}\frac{D_L D_{LS}}{D_S}\theta \qquad [15]$$

With the definition of the critical surface mass density $\Sigma_{crit}$ as

$$\Sigma_{crit} = \frac{c^2}{4\pi G}\frac{D_S}{D_L D_{LS}} \qquad [16]$$

the deflection angle for a such a mass distribution can be expressed as

$$\tilde{\alpha}(\theta) = \frac{\Sigma}{\Sigma_{crit}}\theta \qquad [17]$$

The critical surface mass density can be visualized as the lens mass $M$ "smeared out" over the area of the Einstein ring: $\Sigma_{crit} = M/(R_E^2\pi)$, where $R_E = \theta_E D_L$. The value of the critical surface mass density is roughly $\Sigma_{crit} \approx 0.8\,\mathrm{g\ cm^{-2}}$ for lens and source redshifts of $z_L = 0.5$ and $z_S = 2.0$, respectively. For an arbitrary mass distribution, the condition $\Sigma > \Sigma_{crit}$ at any point is sufficient to produce multiple images.

### Image Positions and Magnifications

The lens equation [6] can be re-formulated in the case of a single-point lens:

$$\beta = \theta - \frac{\theta_E^2}{\theta} \qquad [18]$$

Solving this for the image positions $\theta$, one finds that an isolated point source always produces two images of a background source. The positions of the images are given by the two solutions:

$$\theta_{1,2} = \frac{1}{2}\left(\beta \pm \sqrt{\beta^2 + 4\theta_E^2}\right) \qquad [19]$$

The magnification of an image is defined by the ratio between the solid angles of the image and the source, since the surface brightness is conserved. Hence, the magnification $\mu$ is given as

$$\mu = \frac{\theta}{\beta}\frac{\mathrm{d}\theta}{\mathrm{d}\beta} \qquad [20]$$

In the symmetric case above, the image magnification can be written as (by using the lens equation)

$$\mu_{1,2} = \left(1 - \left[\frac{\theta_E}{\theta_{1,2}}\right]^4\right)^{-1} = \frac{u^2+2}{2u\sqrt{u^2+4}} \pm \frac{1}{2} \qquad [21]$$

Here we defined $u$ as the "impact parameter," the angular separation between lens and source in units of the Einstein radius: $u = \beta/\theta_E$. The magnification of one image (the one inside the Einstein radius) is negative. This means it has negative parity: it is mirror-inverted. For $\beta \to 0$ the magnification diverges. In the limit of geometrical optics, the Einstein ring of a point source has infinite magnification! (Due to the fact that physical objects have a finite size, and also because at some limit wave optics has to be applied, in reality the magnification stays finite.) The sum of the absolute values of the two image magnifications is the measurable total magnification $\mu$:

$$\mu = |\mu_1| + |\mu_2| = \frac{u^2+2}{u\sqrt{u^2+4}} \qquad [22]$$

Note that this value is (always) larger than 1! (This does not violate energy conservation, since this is the magnification relative to an "empty" universe and not relative to a "smoothed out" universe. This issue is treated in detail in Schneider *et al.* (1992, chapter 4.5).) The "sum" of the two image magnifications is unity:

$$\mu_1 + \mu_2 = 1 \qquad [23]$$

### (Non)Singular Isothermal Sphere

A popular model for galaxy lenses is the singular isothermal sphere with a three-dimensional density distribution of

$$\rho(r) = \frac{\sigma_v^2}{2\pi G}\frac{1}{r^2} \qquad [24]$$

where $\sigma_v$ is the one-dimensional velocity dispersion. Projecting the matter on a plane, one obtains the circularly symmetric surface mass distribution

$$\Sigma(\xi) = \frac{\sigma_v^2}{2G}\frac{1}{\xi} \qquad [25]$$

With $M(\xi) = \int_0^\xi \Sigma(\xi')2\pi\xi' d\xi'$ plugged into eqn [4], one obtains the deflection angle for an isothermal sphere, which is a constant (i.e., independent of the impact parameter $\xi$):

$$\tilde{\alpha}(\xi) = 4\pi\frac{\sigma_v^2}{c^2} \qquad [26]$$

In practical units for the velocity dispersion of a galaxy, this can be expressed as

$$\tilde{\alpha}(\xi) = 1.15\left(\frac{\sigma_v}{200\,\mathrm{km\,s^{-1}}}\right)^2 \mathrm{arcsec} \qquad [27]$$

Two generalizations of this isothermal model are commonly used: models with finite cores are more realistic for (spiral) galaxies. In this case, the deflection angle is modified to (core radius $\xi_c$):

$$\tilde{\alpha}(\xi) = 4\pi\frac{\sigma_v^2}{c^2}\frac{\xi}{(\xi_c^2+\xi^2)^{1/2}} \qquad [28]$$

Furthermore, a realistic galaxy lens usually is not perfectly symmetric but is slightly elliptical. Depending on whether one wants an elliptical mass distribution or an elliptical potential, various formalisms are in use.

### Lens Mapping

In the vicinity of an arbitrary point, the lens mapping as shown in eqn [7] can be described by its Jacobian matrix $\mathcal{A}$:

$$\mathcal{A} = \frac{\partial\boldsymbol{\beta}}{\partial\boldsymbol{\theta}} = \left(\delta_{ij} - \frac{\partial\alpha_i(\boldsymbol{\theta})}{\partial\theta_j}\right) = \left(\delta_{ij} - \frac{\partial^2\psi(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right) \qquad [29]$$

Here we made use of the fact that the deflection angle can be expressed as the gradient of an effective two-dimensional scalar potential $\psi$: $\boldsymbol{\nabla}_\theta\psi = \boldsymbol{\alpha}$, where

$$\psi(\boldsymbol{\theta}) = \frac{D_{LS}}{D_L D_S}\frac{2}{c^2}\int \Phi(\boldsymbol{r})\,dz \qquad [30]$$

and $\Phi(\boldsymbol{r})$ is the Newtonian potential of the lens. The determinant of the Jacobian $\mathcal{A}$ is the inverse of the magnification:

$$\mu = \frac{1}{\det \mathcal{A}} \qquad [31]$$

Defining

$$\psi_{ij} = \frac{\partial^2\psi}{\partial\theta_i\partial\theta_j} \qquad [32]$$

the Laplacian of the effective potential $\psi$ is twice the convergence:

$$\psi_{11} + \psi_{22} = 2\kappa = \mathrm{tr}\,\psi_{ij} \qquad [33]$$

With the definitions of the components of the external shear $\gamma$,

$$\gamma_1(\boldsymbol{\theta}) = \tfrac{1}{2}(\psi_{11} - \psi_{22}) = \gamma(\boldsymbol{\theta})\cos[2\varphi(\boldsymbol{\theta})] \qquad [34]$$

and

$$\gamma_2(\boldsymbol{\theta}) = \psi_{12} = \psi_{21} = \gamma(\boldsymbol{\theta})\sin[2\varphi(\boldsymbol{\theta})] \qquad [35]$$

(where the angle $\varphi$ reflects the direction of the shear-inducing tidal force relative to the coordinate system), the Jacobian matrix can be written as

$$\mathcal{A} = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix} = (1-\kappa)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \gamma \begin{pmatrix} \cos 2\varphi & \sin 2\varphi \\ \sin 2\varphi & -\cos 2\varphi \end{pmatrix} \quad [36]$$

The magnification can now be expressed as a function of the local convergence $\kappa$ and the local shear $\gamma$:

$$\mu = (\det \mathcal{A})^{-1} = \frac{1}{(1-\kappa)^2 - \gamma^2} \quad [37]$$

Locations at which $\det A = 0$ have formally infinite magnification. They are called "critical curves" in the lens plane. The corresponding locations in the source plane are the "caustics." For spherically symmetric mass distributions, the critical curves are circles. For a point lens, the caustic degenerates into a point. For elliptical lenses or spherically symmetric lenses plus external shear, the caustics consist of cusps and folds.

### Time Delay and Fermat's Theorem

The deflection angle is the gradient of an effective lensing potential $\psi$. Hence, the lens equation can be rewritten as

$$(\boldsymbol{\theta} - \boldsymbol{\beta}) - \boldsymbol{\nabla}_\theta \psi = 0 \quad [38]$$

or

$$\boldsymbol{\nabla}_\theta \left( \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\beta})^2 - \psi \right) = 0 \quad [39]$$

The term in brackets appears as well in the physical time delay function for gravitationally lensed images:

$$\tau(\boldsymbol{\theta}, \boldsymbol{\beta}) = \tau_{\text{geom}} + \tau_{\text{grav}} = \frac{1+z_L}{c} \frac{D_L D_S}{D_{LS}} \left( \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\beta})^2 - \psi(\theta) \right) \quad [40]$$

This time delay surface is a function of the image geometry $(\boldsymbol{\theta}, \boldsymbol{\beta})$, the gravitational potential $\psi$, and the distances $D_L, D_S$, and $D_{LS}$. The first part – the geometrical time delay $\tau_{\text{geom}}$ – reflects the extra path length compared to the direct line between observer and source. The second part – the gravitational time delay $\tau_{\text{grav}}$ – is the retardation due to gravitational potential of the lensing mass (known and confirmed as Shapiro delay in the solar system). From eqns [39] and [40], it follows that the gravitationally lensed images appear at locations that correspond to extrema in the light travel time, which reflects Fermat's principle in gravitational-lensing optics.

The (angular-diameter) distances that appear in eqn [40] depend on the value of the Hubble constant. Therefore, it is possible to determine the latter by measuring the time delay between different images and using a good model for the effective gravitational potential $\psi$ of the lens.

## Lensing Phenomena

Strong lensing phenomena involve multiple images, caustics, critical lines, usually a significant magnification, and large distortions if extended sources are involved. Below we discuss the most frequent strong lensing phenomena.

### Galactic Microlensing

The conceptually simplest strong lensing scenario is a foreground star acting as a lens on a background star. Since stars in the Milky Way move relative to each other, this can be observed as a time-variable situation: due to the relative motion between observer, lensing star, and source star, the projected impact parameter between lens and source changes with time and produces a time-dependent magnification. If the impact parameter is smaller than an Einstein radius ($u < 1$), then the magnification is $\mu_{\text{min}} > 1.34$ (cf. eqn [22]).

For an extended source, a sequence image configurations with decreasing impact parameter is illustrated in Figure 2 for five instants of time. The separation of the two images is of order-2 Einstein radii when they are of comparable magnification, which corresponds to only about 1 marcsec in a realistic situation in the Milky Way. Hence, the two images cannot be resolved individually; we can only observe the combined brightness of the image pair. This is illustrated in Figures 3 and 4, which show the relative tracks and the respective light curves for five values of the minimum impact parameter $u_{\text{min}}$.

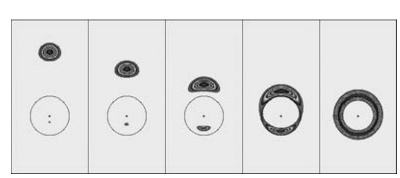


**Figure 2** Five snapshots of a gravitational lens situation with a point lens and an extended source: from left to right the alignment between lens and source gets better and better, until it is perfect in the rightmost panel. This results in the image of an "Einstein ring."

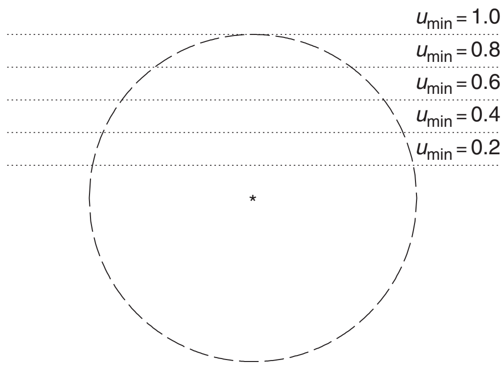



**Figure 3** Five relative tracks between background star and foreground lens (indicated as the central star) parametrized by the impact parameter $u_{min}$. The dashed line indicates the Einstein ring for the lens.

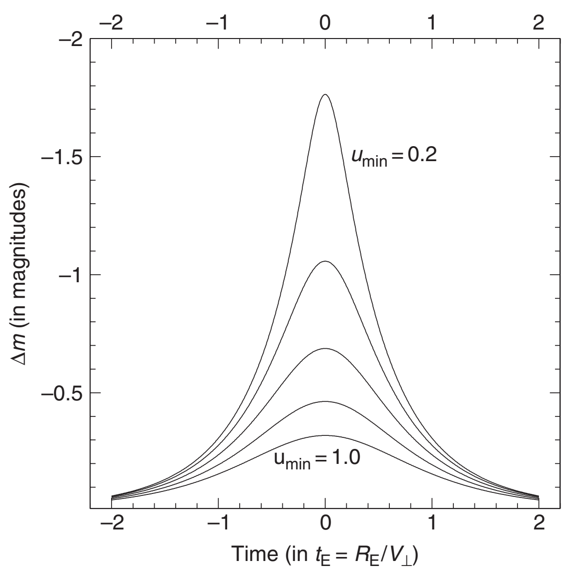



**Figure 4** Five microlensing light curves for the tracks indicated in **Figure 3**, parametrized by the impact parameter $u_{min}$. The vertical axis is the magnification in astronomical magnitudes relative to the unlensed case, the horizontal axis displays the time in "normalized" units.

Quantitatively, the total magnification $\mu = |\mu_1| + |\mu_2|$ of the two images (cf. eqn [22]) entirely depends on the impact parameter $u(t) = r(t)/R_E$ between the lensed star and the lensing object, measured in the lens plane (here $R_E$ is the Einstein radius of the lens, i.e., the radius at which a circular image appears for perfect alignment between source, lens, and observer, cf. Figure 2, rightmost panel):

$$\mu(u(t)) = \frac{u(t)^2 + 2}{u(t)\sqrt{(u(t)^2 + 4)}} \qquad [41]$$

The timescale of such a "microlensing event" is defined as the time it takes for the source to cross the Einstein radius. With realistic values for distances and relative velocity, this can be expressed as

$$t_0 = \frac{R_E}{v_\perp} \approx (0.214\,\mathrm{yr})\sqrt{\frac{M}{M_\odot}}\sqrt{\frac{D_L}{10\,\mathrm{kpc}}} \times \sqrt{1 - \frac{D_L}{D_S}}\left(\frac{v_\perp}{200\,\mathrm{km\,s^{-1}}}\right)^{-1} \qquad [42]$$

(here $v_\perp$ is the (relative) transverse velocity of the lens; we applied the simple relation $D_{LS} = D_S - D_L$, which is valid here).

Note that from eqn [42] it is obvious that it is not possible to determine the mass of the lens from one individual microlensing event. The duration of an event is determined by three unknown parameters: the mass of the lens $m$, the transverse velocity $v_\perp$, and the distance of the lens $D_L$ (assuming we know the distance to the source). It is impossible to disentangle these for individual events. Only with a model for the spatial and velocity distribution of the lensing stars in the Milky Way, one can obtain approximate information about the masses of the lensing objects.

In 1986, Bohdan Paczyński suggested to use this microlensing method as an observational test for potential dark matter candidates in the halo of the Milky Way. If the dark matter is in the form of astrophysical objects (such as brown dwarfs, neutron stars, black holes, sometimes called "MACHO" for MAssive Compact Halo Object), then they should occasionally act as lenses on stars in the neighboring galaxy Large Magellanic Cloud. It turned out that too few of such microlensing events were observed, in order to explain the dark matter this way.

However, this method produced more than 2000 microlensing events by ordinary stars in the direction to the center of the Milky Way. Two of these events provide convincing evidence for a planet accompanying the lensing star. It is likely that gravitational microlensing will provide a statistically very valuable sample of extrasolar planets, because in contrast to most other methods these planets are pre-selected by their host stars. Furthermore, microlensing is sensitive to masses as low as a few Earth masses.

### Multiply-Imaged Quasars

The first gravitationally lensed double quasar was discovered in 1979: two images of the same quasar, separated by about 6 arcsec. This led to the field of gravitational lensing as an observational science. By now, more than 120 multiply imaged quasars are known, mostly double and quadruple images. They

span image separations from 0.3 arcsec to almost 30 arcsec.

Gravitationally lensed quasar systems are studied individually in great detail to get a better understanding of both lens and source. The lens systems are analyzed statistically as well, in order to get information about the population of lenses (and quasars) in the universe, their distribution in distance (i.e., cosmic time) and mass, and hence about the cosmological model.

**Time delay and Hubble constant** As stated above, the signals from a gravitational lens system reach us with a certain "time delay" $\Delta t$, so that the measured fluxes as functions of time, $I_A(t)$ and $I_B(t)$, can be described as: $I_B(t) = \text{const.} \times I_A(t + \Delta t)$. Any intrinsic fluctuation of the quasar shows up in both images, in general with an overall offset in apparent magnitude and an offset in time.

Q0957 + 561 is the first lens system in which the time delay was firmly established:

$$\Delta t_{Q0957+561} = (417 \pm 3)\,\text{days} \qquad [43]$$

With a model of the lens system, the time delay can be used to determine the Hubble constant. (This can be seen very simply: imagine a certain lens situation like the one displayed in Figure 1. If now all length scales are reduced by a factor of 2 and at the same time all masses are reduced by a factor of 2, then for an observer, the angular configuration in the sky would appear exactly identical. But the total length of the light path is reduced by a factor of 2. Now, since the time delay between the two paths is the same fraction of the total lengths in either scenario, a measurement of this fractional length allows us to determine the total length, and hence the Hubble constant, the constant of proportionality between distance and redshift.) The resulting value of $H_0$ is

$$H_0 = (67 \pm 13)\,\text{km s}^{-1}\,\text{Mpc}^{-1} \qquad [44]$$

where the uncertainty comprises the 95% confidence level. To date, about a dozen quasar lens systems have measured time delays. The derived values of the Hubble constant are "lowish," if we assume the best astrophysical motivated lens models.

## Quasar Microlensing

Light bundles from "lensed" quasars are split by intervening galaxies. Usually the quasar light bundle passes through the galaxy and/or the galaxy halo. Galaxies consist at least partly of stars, and galaxy halos consist possibly of compact objects as well.

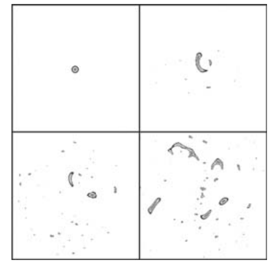


**Figure 5** "Microimages": the top left panel shows an assumed "unlensed" source profile of a quasar. The other three panels illustrate the microimage configuration as it would be produced by stellar objects in the foreground. The surface mass density of the lenses is 20% (top right), 50% (bottom left), and 80% (bottom right) of the critical density (cf. eqn [16]).

Each of these stars (or other compact objects like black holes, brown dwarfs, or planets) acts as a "compact lens" or "microlens" and produces at least one additional microimage of the source. This means the "macroimage" consists of many "microimages" (Figure 5). But because the image splitting is proportional to the square root of the lens mass, these microimages are only of order a microarcsecond apart and cannot be resolved. Various aspects of microlensing have been addressed after the first double quasar had been discovered.

The microlenses produce a complicated two-dimensional magnification distribution in the source plane. It consists of many caustics, locations that correspond to formally infinitely high magnification. An example for such a magnification pattern is shown in Figure 6. It is determined with the parameters of image A of the quadruple quasar Q2237 + 0305

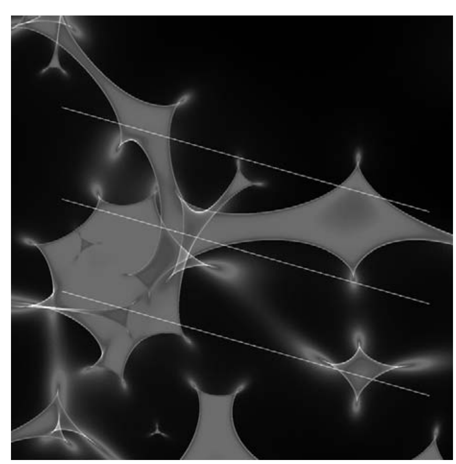


**Figure 6** Magnification pattern in the source plane, produced by a dense field of stars in the lensing galaxy. The gray scale reflects the magnification as a function of the quasar position. Light curves taken along the straight tracks are shown in Figure 7. The microlensing parameters were chosen according to a model for image A of the quadruple quasar *Q*2237 + 0305: $\kappa = 0.36$, $\gamma = 0.44$.

(surface mass density $\kappa = 0.36$; external shear $\gamma = 0.44$). Gray scale indicates the magnification.

Due to the relative motion between observer, lens, and source, the quasar changes its position relative to this arrangement of caustics, that is, the apparent brightness of the quasar changes with time. A one-dimensional cut through such a magnification pattern, convolved with a source profile of the quasar, results in a microlensed light curve. Examples for microlensed light curves taken along the straight lines in Figure 6 can be seen in Figure 7 for two different quasar sizes.

In particular when the quasar track crosses a caustic (the sharp lines in Figure 6 for which the magnification formally is infinite, because the determinant of the Jacobian disappears, cf. eqn [31]), a pair of highly magnified microimages appears newly or merges and disappears. Such a microlensing event can easily be detected as a strong peak in the light curve of the quasar image.

Microlens-induced fluctuations in the observed brightness of quasars contain information both about the light-emitting source (size of continuum region or broad line region of the quasar, brightness profile of quasar) and about the lensing objects (masses, density, transverse velocity). Hence, from a comparison between observed and simulated quasar microlensing, one can draw conclusions about the density and mass scale of the microlenses. So far the "best" example of a microlensed quasar is the quadruple quasar Q2237 + 0305.

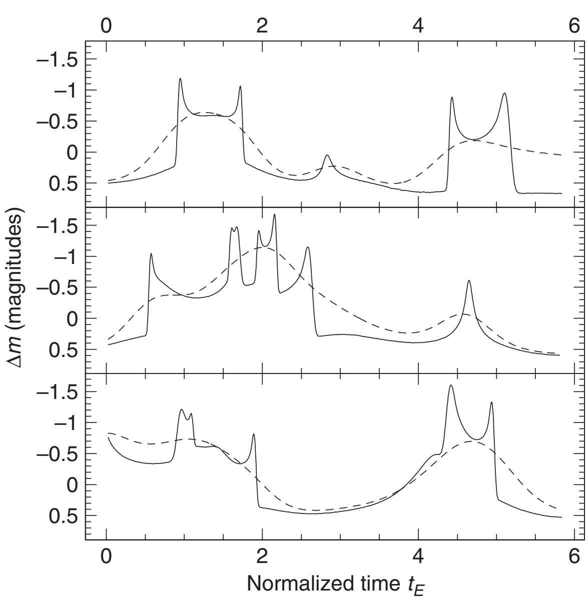



**Figure 7** Microlensing light curve for the straight lines in Figure 6. The solid and dashed lines indicate relatively small and large quasar sizes. The time axis is in units of Einstein radii divided by unit velocity.

### Einstein Rings

If a point source lies exactly behind a point lens, a ring-like image occurs. Theorists had recognized early on that such a symmetric lensing arrangement would result in a ring image, the so-called "Einstein ring." There are two necessary requirements for the observability of Einstein rings: the mass distribution of the lens needs to be approximately axially symmetric, as seen from the observer, and the source must lie exactly on top of the resulting degenerate pointlike caustic. Such a geometric arrangement is highly unlikely for pointlike sources. But astrophysical sources in the real universe have a finite extent, and it is enough if a part of the source covers the point caustic (or the complete astroid caustic in the case of a not quite axially symmetric mass distribution) in order to produce such an annular image.

In 1988, the first example of an "Einstein ring" was discovered. With high-resolution radio observations, the extended radio source MG1131 + 0456 turned out to be a ring with a diameter of about 1.75 arcsec. The source was identified as a radio lobe at a redshift of $z_S = 1.13$, whereas the lens is a galaxy at $z_L = 0.85$. By now more than a dozen cases have been found that qualify as Einstein rings. Their diameters vary between 0.33 and about 2 arcsec.

### Giant Luminous Arcs and Arclets

Fritz Zwicky had pointed out the potential use of galaxies and galaxy clusters as gravitational lenses in the 1930s. With background galaxies as sources, the apparent lensing consequences for them would be far more dramatic than for quasars: galaxies should be heavily deformed once they are strongly lensed. Rich clusters of galaxies at redshifts beyond $z \approx 0.2$ with masses of order $10^{14} M_\odot$ are very effective lenses if they are centrally concentrated. Their Einstein radii are of the order of 20 arcsec.

In 1986, the following gravitational lensing phenomenon was discovered: magnified, distorted, and strongly elongated images of background galaxies which happen to lie behind foreground clusters of galaxies, the so-called giant luminous arcs. The giant arcs can be exploited in two ways, as is typical for many lens phenomena. Firstly, they provide us with strongly magnified galaxies at (very) high redshifts. These galaxies would be too faint to be detected or analyzed in their unlensed state. Hence, with the lensing boost, we can study these galaxies in their early evolutionary stages, possibly as infant or protogalaxies, relatively shortly after the big bang. The other practical application of the arcs is to take them as tools to study the potential and mass distribution of the lensing galaxy cluster. In the simplest model of a spherically symmetric mass distribution for the cluster,

giant arcs form very close to the critical curve, which marks the Einstein ring. So with the redshifts of the cluster and the arc, it is easy to determine a rough estimate of the lensing mass by just determining the radius of curvature and interpreting it as the Einstein radius of the lens system.

#### Weak Lensing/Statistical Lensing/Cosmic Shear

In contrast to the phenomena that were mentioned here, "weak lensing" deals with effects of light deflection that cannot be measured individually, but rather in a statistical way only. No caustics, critical lines, or multiple images are involved. As was discussed above, "strong lensing" – usually defined as the regime that involves multiple images, high magnifications, and caustics in the source plane – is a rare phenomenon. Weak lensing on the other hand is much more common. In principle, weak lensing acts along each line of sight in the universe, since each light bundle's path is affected by matter inhomogeneities along or near its path. It is just a matter of how accurately we can measure. In recent years, many teams started impressive and ambitious observational programs to determine the slight distortion of tens of thousands of background galaxies by foreground galaxy clusters and/or by the large-scale structure in the universe, the so-called cosmic shear. It is beyond the scope of this article to discuss these applications of weak gravitational lensing. The interested reader is referred to the "Further reading" section, in particular to Bartelmann and Schneider (2001).

*See also:* Cosmology: Mathematical Aspects; General Relativity: Experimental Tests; General Relativity: Overview; Newtonian Limit of General Relativity; Singularity and Bifurcation Theory.

### Further Reading

Bartelmann M and Schneider P (2001) Weak gravitational lensing. *Physics Reports* 340: 291.
Bliokh PV and Minakov AA (1989) *Gravitational Lensing*, (in Russian). Kiev: Izdatel'stvo Naukova Dumka.
Kochanek CS, Schneider P, and Wambsganss J (2005) In: Meylan G, Jetzer P, and North P (eds.) *Gravitational Lensing: Strong, Weak & Micro*, Proceedings of the 33rd Saas-Fee Advanced Course. Berlin: Springer.
Mollerach S and Roulet E (2002) *Gravitational Lensing and Microlensing*. World Scientific.
Narayan R and Bartelmann M (1997) Lectures on gravitational lensing. In: Dekel A and Ostriker JP (eds.) *Proceedings of the 1995 Jerusalem Winter School*.
Paczyński B (1996) Gravitational microlensing in the local group. *Annual Review of Astronomy and Astrophysics* 34: 419.
Perlick V (2000) *Ray Optics, Fermat's Principle, and Applications to General Relativity*, Lecture Notes in Physics, vol. 61. Springer.
Perlick V (2004) *Gravitational Lensing from a Spacetime Perspective*, Living Reviews in Relativity, 9/2004 http://relativity.livingreviews.org
Petters AO, Levine H, and Wambsganss J (2001) *Singularity Theory and Gravitational Lensing (Birkhauser 2001)*, Progress in Mathematical Physics, vol. 21.
Schneider P, Ehlers J, and Falco EE (1992) *Gravitational Lenses*. Berlin: Springer.
Wambsganss J (1998) *Gravitational Lensing in Astronomy*, Living Reviews in Relativity, 12/1998 http://relativity.livingreviews.org

# Gravitational *N*-Body Problem (Classical)

**D C Heggie**, The University of Edinburgh, Edinburgh, UK



## Introduction

Let a number, $N$, of particles interact classically through Newton's laws of motion and Newton's inverse-square law of gravitation. Then the equations of motion are

$$\ddot{\boldsymbol{r}}_i = -G \sum_{j=1,\, j\neq i}^{j=N} m_j \frac{\boldsymbol{r}_i - \boldsymbol{r}_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j|^3} \qquad [1]$$

where $\boldsymbol{r}_i$ is the position vector of the $i$th particle relative to some inertial frame, $G$ is the universal constant of gravitation, and $m_i$ is the mass of the $i$th particle. These equations provide an approximate mathematical model with numerous applications in astrophysics, including the motion of the Moon and other bodies in the solar system (planets, asteroids, comets, and meteor particles); stars in stellar systems ranging from binary and other multiple stars to star clusters and galaxies; and the motion of dark-matter particles in cosmology. For $N=1$ and $N=2$, the equations can be solved analytically. The case $N=3$ provides one of the richest of all unsolved dynamical problems – the general three-body problem. For problems dominated by one massive body, as in many planetary problems, approximate methods based on perturbation expansions have been developed. In stellar dynamics, astrophysicists have developed numerous numerical and theoretical approaches to the problem for larger values of $N$, including treatments based on the Boltzmann equation and the Fokker–Planck equation; such $N$-body systems can also be modeled as self-gravitating

gases, and thermodynamic insights underpin much of our qualitative understanding.

## Few-Body Problems

### The Two-Body Problem

For $N = 2$, the relative motion of the two bodies can be reduced to the force-free motion of the center of mass and the problem of the relative motion. If $\boldsymbol{r} = \boldsymbol{r}_1 - \boldsymbol{r}_2$, then

$$\ddot{\boldsymbol{r}} = -G(m_1 + m_2)\frac{\boldsymbol{r}}{|\boldsymbol{r}|^3} \qquad [2]$$

often called the Kepler problem. It represents motion of a particle of unit mass under a central inverse-square force of attraction. Energy and angular momentum are constant, and the motion takes place in a plane passing through the origin. Using plane polar coordinates $(r, \theta)$ in this plane, the equations for the energy and angular momentum reduce to

$$E = \frac{1}{2}\left(\dot{r}^2 + \frac{L^2}{r^2}\right) - \frac{G(m_1 + m_2)}{r} \qquad [3]$$

$$L = r^2\dot{\theta} \qquad [4]$$

(Note that these are not the energy and angular momentum of the two-body problem, even in the barycentric frame of the center of mass; $E$ and $L$ must be multiplied by the reduced mass $m_1 m_2/(m_1 + m_2)$.) Using eqns [3] and [4], the problem is reduced to quadratures. The solution shows that the motion is on a conic section (ellipse, circle, straight line, parabola, or hyperbola), with the origin at one focus.

This reduction depends on the existence of integrals of the equations of motion, and these in turn depend on symmetries of the underlying Lagrangian or Hamiltonian. Indeed, eqns [1] yield ten first integrals: six yield the rectilinear motion of the center of mass, three the total angular momentum, and one the energy. Furthermore, eqn [2] may be transformed, via the Kustaanheimo–Stiefel (KS) transformation, to a four-dimensional simple harmonic oscillator. This reveals further symmetries, corresponding to further invariants: the three components of the Lenz vector. Another manifestation of the abundance of symmetries of the Kepler problem is the fact that there exist action-angle variables in which the Hamiltonian depends only on one action, that is, $H = H(L)$. Another application of the KS transformation is one that has practical importance: it removes the singularity of (i.e., regularizes) the Kepler problem at $r = 0$, which is troublesome numerically.

To illustrate the character of the KS transformation, we consider briefly the planar case, which can be handled with a complex variable obeying the equation of motion $\ddot{z} = -z/|z|^3$ (after scaling eqn (2)). By introducing the Levi-Civita transformation $z = Z^2$ and Sundman's transformation of the time, that is, $\mathrm{d}t/\mathrm{d}\tau = |z|$, the equation of motion transforms to $Z'' = hZ/2$, where $h = |\dot{z}|^2/2 - 1/|z|$ is the constant of energy. The KS transformation is a very similar exercise using quaternions.

### The Restricted Three-Body Problem

The simplest three-body problem is given by the motion of a test particle in the gravitational field of two particles, of positive mass $m_1$, $m_2$, in circular Keplerian motion. This is called the circular restricted three-body problem, and the two massive particles are referred to as primaries. In a rotating frame of reference, with origin at the center of mass of these two particles, which are at rest at positions $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, the equation of motion is

$$\ddot{\boldsymbol{r}} + 2\Omega \times \dot{\boldsymbol{r}} + \Omega \times (\Omega \times \boldsymbol{r}) = G\nabla\left(\frac{m_1}{|\boldsymbol{r} - \boldsymbol{r}_1|} + \frac{m_2}{|\boldsymbol{r} - \boldsymbol{r}_2|}\right) \qquad [5]$$

where $\boldsymbol{r}$ is the position of the massless particle and $\Omega$ is the angular velocity of the frame.

This problem has three degrees of freedom but only one known integral: it is the Hamiltonian in the rotating frame, and equivalent to the Jacobi integral, $J$. One consequence is that Liouville's theorem is not applicable, and more elaborate arguments are required to decide its integrability. Certainly, no general analytical solution is known.

There are five equilibrium solutions, discovered by Euler and Lagrange (see **Figure 1**). They lie at critical points of the effective potential in the

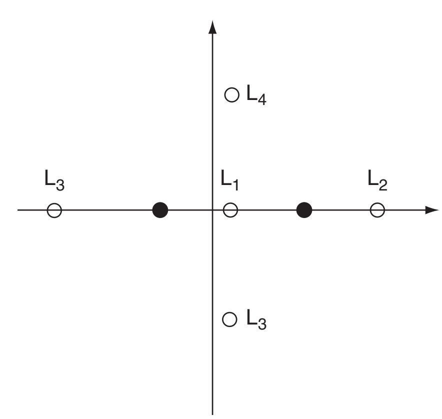



**Figure 1** The equilibrium solutions of the circular restricted three-body problem. A rotating frame of reference is chosen in which two particles are at rest on the *x*-axis. The massless particle is at equilibrium at each of the five points shown. Five similar configurations exist for the general three-body problem; these are the "central" configurations.

rotating frame, and demarcate possible regions of motion.

Throughout the twentieth century, much numerical effort was used in finding and classifying periodic orbits, and in determining their stability and bifurcations. For example, there are families of periodic orbits close to each primary; these are perturbed Kepler orbits, and are referred to as satellite motions. Other important families are the series of Liapounov orbits starting at the equilibrium points.

Some variants of the restricted three-body problem include the following:

1. The elliptic restricted three-body problem, in which the primaries move on an elliptic Keplerian orbit; in suitable coordinates the equation of motion closely resembles eqn [5], except for a factor on the right side which depends explicitly on the independent variable (transformed time); this system has no first integral.
2. Sitnikov's problem, which is a special case of the elliptic problem, in which $m_1 = m_2$, and the motion of the massless particle is confined to the axis of symmetry of the Keplerian motion; this is still nonintegrable, but simple enough to allow extensive analysis of such fundamental issues as integrability and stochasticity.
3. Hill's problem, which is a scaled version suitable for examining motions close to one primary; its importance in applications began with studies of the motion of the moon, and it remains vital for understanding the motion of asteroids.

## The General Three-Body Problem

**Exact solutions** When all three particles have nonzero masses, the equations of motion become

$$m_i \ddot{\boldsymbol{r}}_i = -\nabla_i W$$

where the potential energy is

$$W = -G \sum_{1 \le i < j \le 3} \frac{m_i m_j}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}$$

Then the exact solutions of Euler and Lagrange survive in the form of homographic solutions. In these solutions, the configuration remains geometrically similar, but may rotate and/or pulsate in the same way as in the two-body problem.

Let us represent the position vector $\boldsymbol{r}_i$ in the planar three-body problem by the complex number $z_i$. Then, it is easy to see that we have a solution of the form $z_i(t) = z(t) z_{0i}$, provided that

$$\ddot{z} = -C \frac{z}{|z|^3}$$

and

$$m_i C z_{0i} = \nabla_i W(z_{01}, z_{02}, z_{03})$$

for some constant $C$. Thus, $z(t)$ may take the form of any solution of the Kepler problem, while the complex numbers $z_{0i}$ must correspond to what is called a central configuration. These are in fact critical points of the scale-free function $W\sqrt{I}$, where $I$ (the "moment of inertia of the system") is given by $I = \sum_1^3 m_i r_i^2$; and $C = -W/I$.

The existence of other important classes of periodic solutions can be proved analytically, even though it is not possible to express the solution in closed form. Examples include hierarchical three-body systems, in which two masses $m_1$, $m_2$ exhibit nearly elliptic relative motion, while a third mass orbits the barycenter of $m_1$ and $m_2$ in another nearly elliptic orbit. In the mathematical literature, this is referred to as motion of elliptic–elliptic type. More surprisingly, the existence of a periodic solution in which the three bodies travel in succession along the same path, shaped like a figure 8 (cf. **Figure 2**), was established by Chenciner and Montgomery (2000), following its independent discovery by Moore using numerical methods. Another interesting periodic motion that was discovered numerically, by Schubart, is a solution of the collinear three-body problem, and so collisions are inevitable. In this motion, the body in the middle alternately encounters the other two bodies.

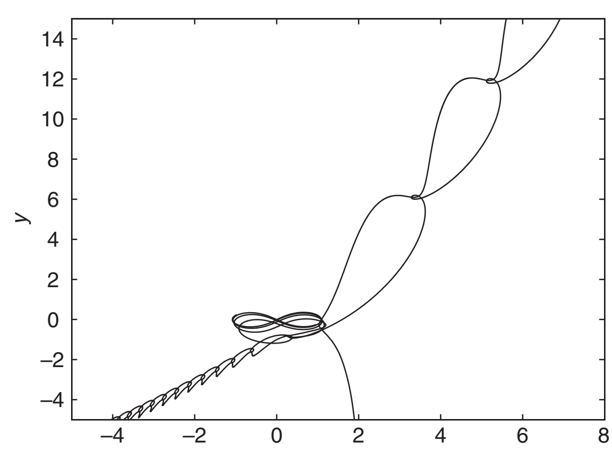



**Figure 2** A rare example of a scattering encounter between two binaries (which approach from upper right and lower left) which leads to a permanently bound triple system describing the "figure-8" periodic orbit. A fourth body escapes at the bottom. Note the differing scales on the two axes. (Reproduced with permission from Heggie DC (2000) A new outcome of binary-binary scattering. *Monthly Notices of the Royal Astronomical Society* 318(4): L61–L63; © Blackwell Publishing Ltd.)

**Singularities** As Schubart's solution illustrates, two-body encounters can occur in the three-body problem. Such singularities can be regularized just as in the pure two-body problem. Triple collisions cannot be regularized in general, and this singularity has been studied by the technique of "blowup." This has been worked out most thoroughly in the collinear three-body problem, which has only two degrees of freedom. The general idea is to transform to two variables, of which one (denoted by $r$, say) determines the scale of the system, while the other ($s$) determines the configuration (e.g., the ratio of separations of the three masses). By scaling the corresponding velocities and the time, one obtains a system of three equations of motion for $s$ and the two velocities which are perfectly regular in the limit $r \to 0$. In this limit, the energy integral restricts the solutions of the system to a manifold (called the collision manifold). Exactly the same manifold results for zero-energy solutions, which permits a simple visualization. Equilibria on the collision manifold correspond to the Lagrangian collinear solutions in which the system either expands to infinity or contracts to a three-body collision.

**Qualitative ideas** Reference has already been made to motion of elliptic–elliptic type. In a motion of elliptic–hyperbolic type, there is again an "inner" pair of bodies describing nearly Keplerian motion, while the relative motion of the third body is nearly hyperbolic. In applications, this is referred to as a kind of scattering encounter between a binary and a third body. When the encounter is sufficiently close, it is possible for one member of the binary to be exchanged with the third body. One of the major historical themes of the general three-body problem is the classification of connections between these different types of asymptotic motion. It is possible to show, for instance, that the measure of initial conditions of hyperbolic–elliptic type leading asymptotically to elliptic–elliptic motion (or any other type of permanently bound motion) is zero. Much of the study of such problems has been carried out numerically.

There are many ways in which the stability of three-body motions may be approached. One example is furnished by the central configurations already referred to. They can be used to establish sufficient conditions for ensuring that exchange is impossible, and similar conclusions.

A powerful tool for qualitative study of three-body motions is Lagrange's identity, which is now thought of as the reduction to three bodies of the virial theorem. Let the size of the system be characterized by the "moment of inertia" $I$. Then it is easy to show that

$$\frac{\mathrm{d}^2 I}{\mathrm{d}t^2} = 4T + 2W$$

where $T$, $W$ are, respectively, the kinetic and potential energies of the system. Usually, the barycentric frame is adopted. Since $E = T + V$ is constant and $T \geq 0$, it follows that the system is not bounded for all $t > 0$ unless $E < 0$.

**Perturbation theory** The question of the integrability of the general three-body problem has stimulated much research, including the famous study by Poincaré which established the nonexistence of integrals beyond the ten classical ones. Poincaré's work was an important landmark in the application to the three-body problem of perturbation methods. If one mass dominates, that is, $m_1 \gg m_2$ and $m_1 \gg m_3$, then the motion of $m_2$ and $m_3$ relative to $m_1$ is a mildly perturbed two-body motion, unless $m_2$ and $m_3$ are close together. Then it is beneficial to describe the motion of $m_2$ relative to $m_1$ by the parameters of Keplerian motion. These would be constant in the absence of $m_3$, and vary slowly because of the perturbation by $m_3$. This was the idea behind Lagrange's very general method of variation of parameters for solving systems of differential equations. Numerous methods were developed for the iterative solution of the resulting equations. In this way, the solution of such a three-body problem could be represented as a type of trigonometric series in which the arguments are the angle variables describing the two approximate Keplerian motions. These were of immense value in solving problems of celestial mechanics, that is, the study of the motions of planets, their satellites, comets, and asteroids.

A major step forward was the introduction of Hamiltonian methods. A three-body problem of the type considered here has a Hamiltonian of the form

$$H = H_1(L_1) + H_2(L_2) + R$$

where $H_i$, $i = 1, 2$, are the Hamiltonians describing the interaction between $m_i$ and $m_1$, and $R$ is the "disturbing function." It depends on all the variables, but is small compared with the $H_i$. Now perturbation theory reduces to the task of performing canonical transformations which simplify $R$ as much as possible.

Poincaré's major contribution in this area was to show that the series solutions produced by perturbation methods are not, in general, convergent, but

asymptotic. Thus, they were of practical rather than theoretical value. For example, nothing could be proved about the stability of the solar system using perturbation methods. It took the further analytic development of KAM theory to rescue this aspect of perturbation theory. This theory can be used to show that, provided that two of the three masses are sufficiently small, then for almost all initial conditions the motions remain close to Keplerian for all time. Unfortunately, now it is the practical aspect of the theory which is missing; though we have introduced this topic in the context of the three-body problem, it is extensible to any $N$-body system with $N-1$ small masses in nearly Keplerian motion about $m_1$, but to be applicable to the solar system the masses of the planets would have to be ridiculously small.

**Numerical methods** Numerical integrations of the three-body problem were first carried out near the beginning of the twentieth century, and are now commonplace. For typical scattering events, or other short-lived solutions, there is usually little need to go beyond common Runge–Kutta methods, provided that automatic step-size control is adopted. When close two-body approaches occur, some regularization based on the KS transformation is often exploited. In cases of prolonged elliptic–elliptic motion, an analytic approximation based on Keplerian motion may be adequate. Otherwise (as in problems of planetary motion, where the evolution takes place on an extremely long timescale), methods of very high order are often used. Symplectic methods, which have been developed in the context of Hamiltonian problems, are increasingly adopted for problems of this kind, as their long-term error behavior is generally much superior to that of methods which ignore the geometrical properties of the equations of motion.

### Four- and Five-Body Problems

Many of the foregoing remarks, on central configurations, numerical methods, KAM theory, etc., apply equally to few-body problems with $N > 3$. Of special interest from a theoretical point of view is the occurrence of a new kind of singularity, in which motions become unbounded in finite time. For $N = 4$, the known examples also require two-body collisions, but noncollision orbits exhibiting finite-time blowup are known for $N = 5$.

One of the practical (or, at least, astronomical) applications is again to scattering encounters, this time involving the approach of two binaries on a hyperbolic relative orbit. Numerical results show that a wide variety of outcomes is possible, including even the creation of the figure-8 periodic orbit of the three-body problem, while a fourth body escapes (**Figure 2**).

## Many-Body Problems

Many of the concepts already introduced, such as the virial theorem, apply equally well to the many-body classical gravitational problem. This section refers mainly to the new features which arise when $N$ is not small. In particular, statistical descriptions become central. The applications also have a different emphasis, moving from problems of planetary dynamics (celestial mechanics) to those of stellar dynamics. Typically, $N$ lies in the range $10^2$–$10^{12}$.

### Evolution of the Distribution Function

The most useful statistical description is obtained if the correlations we neglect and focus on the one-particle distribution function $f(\boldsymbol{r}, \boldsymbol{v}, t)$, which can be interpreted as the number density at time $t$ at the point in phase space corresponding to position $\boldsymbol{r}$ and velocity $\boldsymbol{v}$. Several processes contribute to the evolution of $f$.

**Collective effects** When the effects of near neighbors are neglected, the dynamics is described by the Vlasov–Poisson system

$$\frac{\partial f}{\partial t} + \boldsymbol{v} \cdot \frac{\partial f}{\partial \boldsymbol{r}} - \frac{\partial \phi(\boldsymbol{r}, t)}{\partial \boldsymbol{r}} \cdot \frac{\partial f}{\partial \boldsymbol{v}} = 0 \qquad [6]$$

$$\nabla^2 \phi = 4\pi G m \int f(\boldsymbol{r}, \boldsymbol{v}, t) \mathrm{d}^3 \boldsymbol{v} \qquad [7]$$

where $\phi$ is the gravitational potential and $m$ is the mass of each body. Obvious extensions are necessary if not all bodies have the same mass.

Solutions of eqn [6] may be found by the method of characteristics, which is most useful in cases where the equation of motion $\ddot{\boldsymbol{r}} = -\nabla\phi$ is integrable, for example, in stationary, spherical potentials. An example is the solution

$$f = |E|^{7/2} \qquad [8]$$

where $E$ is the specific energy of a body, that is, $E = v^2/2 + \phi$. This satisfies eqn [6] provided that $\phi$ is static. Equation [7] is satisfied provided that $\phi$ satisfies a case of the Lane–Emden equation, which is easy to solve in this case.

The solution just referred to is an example of an equilibrium solution. In an equilibrium solution, the virial theorem takes the form $4T + 2W = 0$, where

$T$, $W$ are appropriate mean-field approximations for the kinetic and potential energy, respectively. It follows that $E = -T$, where $E = T + V$ is the total energy. An increase in $E$ causes a decrease in $T$, which implies that a self-gravitating $N$-body system exhibits a negative specific heat.

There is little to choose between one equilibrium solution and another, except for their stability. In such an equilibrium, the bodies orbit within the potential on a timescale of the crossing time, which is conventionally defined to be

$$t_{\mathrm{cr}} = \frac{GM^{5/2}}{(2|E|)^{3/2}}$$

The most important evolutionary phenomenon of collisionless dynamics is violent relaxation. If $f$ is not time independent then $\phi$ is time dependent in general. Also, from the equation of motion of one body, $E$ varies according to $\mathrm{d}E/\mathrm{d}t = \partial\phi/\partial t$, and so energy is exchanged between bodies, which leads to an evolution of the distribution of energies. This process is known as violent relaxation.

Two other relaxation processes are of importance:

1. Relaxation is possible on each energy hypersurface, even in a static potential, if the potential is nonintegrable.
2. The range of collective phenomena becomes remarkably rich if the system exhibits ordered motions, as in rotating systems. Then an important role is played by resonant motions, especially resonances of low order. The corresponding theory lies at the basis of the theory of spiral structure in galaxies, for instance.

**Collisional effects** The approximations of collisionless stellar dynamics suppress two important processes:

1. The exponential divergence of stellar orbits, which takes place on a timescale of order $t_{\mathrm{cr}}$. Even in an integrable potential, therefore, $f$ evolves on each energy hypersurface.
2. Two-body relaxation. It operates on a timescale of order $(N/\ln N)t_{\mathrm{cr}}$, where $N$ is the number of particles. Although this two-body relaxation timescale, $t_{\mathrm{r}}$, is much longer than any other timescale we have considered, this process leads to evolution of $f(E)$, and it dominates the long-term evolution of large $N$-body systems. It is usually modeled by adding a collision term of Fokker–Planck type on the right-hand side of eqn [6].

In this case, the only equilibrium solutions in a steady potential are those in which $f(E) \propto \exp(-\beta E)$, where $\beta$ is a constant. Then eqn [7] becomes Liouville's equation, and for the case of spherical symmetry the relevant solutions are those corresponding to the isothermal sphere.

### Collisional Equilibrium

We consider the collisional evolution of an $N$-body system further in a later subsection and here develop fundamental ideas about the isothermal model. The isothermal model has infinite mass, and much has been learned by considering a model confined within an adiabatic boundary or enclosure. There is a series of such models, characterized by a single dimensionless parameter, which can be taken to be the ratio between the central density and the density at the boundary, $\rho_0/\rho_{\mathrm{e}}$ (**Figure 3**).

These models are extrema of the Boltzmann entropy $S = -k \int f \ln f \mathrm{d}^6\tau$, where $k$ is the Boltzmann constant, and the integration is taken over all available phase space. Their stability may be determined by evaluating the second variation of $S$. It is found that it is negative definite, so that $S$ is a local maximum and the configuration is stable, only if $\rho_0/\rho_{\mathrm{e}} < 709$ approximately. A physical explanation for this is the following. In the limit when $\rho_0/\rho_{\mathrm{e}} \simeq 1$, the self-gravity (which causes the spatial inhomogeneity) is weak, and the system behaves like an ordinary perfect gas. When $\rho_0/\rho_{\mathrm{e}} \gg 1$, however, the system is highly inhomogeneous, consisting of a core of low mass and high density surrounded by an extensive halo of high mass and low density. Consider a transfer of energy from the deep interior to the envelope. In the envelope, which is restrained by the enclosure, the additional energy causes a rise in temperature, but this is small, because of the very large mass of the halo. Extraction of energy from around the core, however, causes the bodies there to sink and accelerate, and, because of the negative

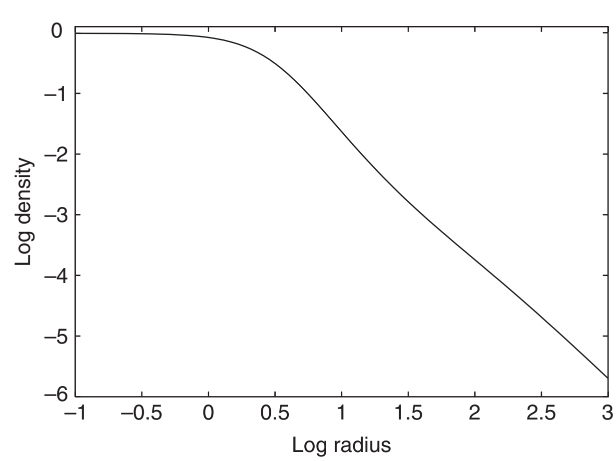



**Figure 3** The density profile of the nonsingular isothermal model, with conventional scalings.

specific heat of a self-gravitating system, they gain more kinetic energy than they lost in the original transfer. Now the system is hotter in the core than in the halo, and the transfer of energy from the interior to the exterior is self-sustaining, in a gravothermal runaway. The isothermal model with large density contrast is therefore unstable.

The negative specific heat, and the lack of an equilibrium which maximizes the entropy, are two examples of the anomalous thermodynamic behavior of the self-gravitating *N*-body problem. They are related to the long-range nature of the gravitational interaction, the importance of boundary terms, and the nonextensivity of the energy. Another consequence is the inequivalence of canonical and microcanonical ensembles.

### Numerical Methods

The foregoing considerations are difficult to extend to systems without a boundary, although they are a vital guide to the behavior even in this case. Our knowledge of such systems is due largely to numerical experiments, which fall into several classes:

1. Direct *N*-body calculations. These minimize the number of simplifying assumptions, but are expensive. Special-purpose hardware is readily available, which greatly accelerates the necessary calculations. Great care has to be taken in the treatment of few-body configurations, which otherwise consume almost all resources.
2. Hierarchical methods, including tree methods, which shorten the calculation of forces by grouping distant masses. They are mostly used for collisionless problems.
3. Grid-based methods, which are used for collisionless problems.
4. Fokker–Planck methods, which usually require a theoretical knowledge of the statistical effects of two-, three- and four-body interactions. Otherwise they can be very flexible, especially in the form of Monte Carlo codes.
5. Gas codes. The behavior of a self-gravitating system is simulated surprisingly well by modeling it as a self-gravitating perfect gas, rather like a star.

### Collisional Evolution

Consider an isolated *N*-body system, which is supposed initially to be given by a spherically symmetric equilibrium solution of eqns [6] and [7], such as eqn [8]. The temperature decreases with increasing radius, and a gravothermal runaway causes the "collapse" of the core, which reaches

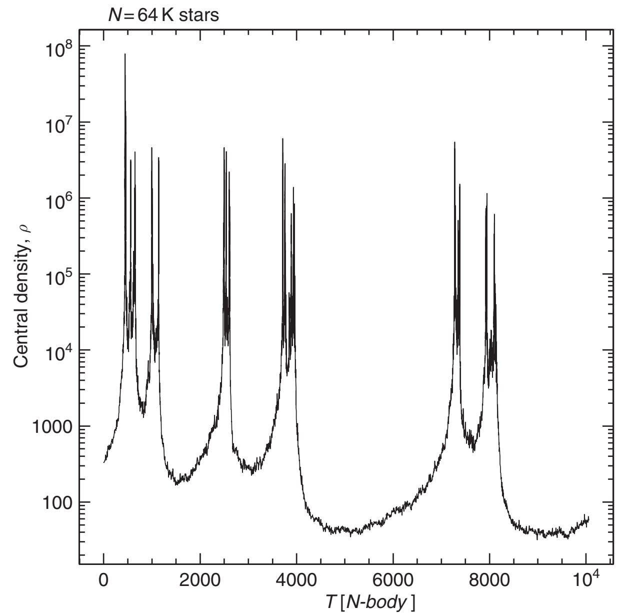



**Figure 4** Gravothermal oscillations in an *N*-body system with $N = 65\,536$. The central density is plotted as a function of time in units such that $t_{cr} = 2\sqrt{2}$. (Source: Baumgardt H, Hut P, and Makino J, with permission.)

extremely high density in finite time. (This collapse takes place on the long two-body relaxation timescale, and so it is not the rapid collapse, on a free-fall timescale, which the name rather suggests.)

At sufficiently high densities, the timescale of three-body reactions becomes competitive. These create bound pairs, the excess energy being removed by a third body. From the point of view of the one-particle distribution function, *f*, these reactions are exothermic, causing an expansion and cooling of the high-density central regions. This temperature inversion drives the gravothermal runaway in reverse, and the core expands, until contact with the cool envelope of the system restores a normal temperature profile. Core collapse resumes once more, and leads to a chaotic sequence of expansions and contractions, called gravothermal oscillations (Figure 4).

The monotonic addition of energy during the collapsed phases causes a secular expansion of the system, and a general increase in all timescales. In each relaxation time, a small fraction of the masses escape, and eventually (it is thought) the system consists of a dispersing collection of mutually unbound single masses, binaries, and (presumably) stable higher-order systems.

It is very remarkable that the long-term fate of the largest self-gravitating *N*-body system appears to be intimately linked with the three-body problem.

*See also:* Boltzmann Equation (Classical and Quantum); Chaos and Attractors; Dynamical Systems and Thermodynamics; KAM Theory and Celestial Mechanics; Lyapunov Exponents and Strange Attractors; Nonequilibrium Statistical Mechanics: Interaction between Theory and Numerical Simulations; Quantum *N*-Body Problem; Stability Problems in Celestial Mechanics; Stability Theory and KAM.

## Further Reading


Arnol'd VI, Kozlov VV, and Neishtadt AI (1993) *Mathematical Aspects of Classical and Celestial Mechanics*, 2nd edn. Berlin: Springer.

Barrow-Green J (1997) *Poincaré and the Three Body Problem*. Providence: American Mathematical Society.

Binney J and Tremaine S (1988) *Galactic Dynamics*. Princeton: Princeton University Press.

Chenciner A and Montgomery R (2000) A remarkable periodic solution of a three-body problem in the case of equal masses. *Annals of Mathematics* 152: 881–901.

Dauxois T *et al.* (eds.) (2003) *Dynamics and Thermodynamics of Systems with Long Range Interactions*. Berlin: Springer.

Diacu F (2000) A century-long loop. *Mathematical Intelligencer* 22(2): 19–25.

Heggie DC and Hut P (2003) *The Gravitational Million-Body Problem*. Cambridge: Cambridge University Press.

Marchal C (1990) *The Three-Body Problem*. Amsterdam: Elsevier.

Murray CD and Dermott SF (2000) *Solar System Dynamics*. Cambridge: Cambridge University Press.

Padmanabhan T (1990) Statistical mechanics of gravitating systems. *Physics Reports* 188: 285–362.

Siegel CL, Moser JK, and Kalme CI (1995) *Lectures on Celestial Mechanics*. Berlin: Springer.

Steves BJ and Maciejewski AJ (eds.) (2001) *The Restless Universe. Applications of Gravitational N-Body Dynamics to Planetary, Stellar and Galactic Systems*. Bristol: IoP and SUSSP.

Szebehely V (1967) *Theory of Orbits: The Restricted Problem of Three Bodies*. New York: Academic Press.


# Gravitational Waves

**G González and J Pullin**, Louisiana State University, Baton Rouge, LA, USA



In elementary physics presentations, one learns about electricity and magnetism, and also about gravity. There appear striking similarities between Newton's law of gravitational attraction and Coulomb's law of attraction between charges. There are also obvious differences, the most immediate one being that in gravitation all masses are positive and always attract each other, whereas in electromagnetism charges may attract or repel, depending on their signs. We also know today that Newton's theory of gravity is not considered an entirely correct description of the gravitational field, particularly when fields are time dependent and intense. The currently accepted theory of gravity is Einstein's theory of general relativity.

The similarity between electromagnetism and gravitation also holds to a certain extent when the fields depend on time. This is usually not discussed in elementary treatments since a full description of time-dependent gravitational fields requires the use of general relativity. It is true, however, that if the fields are weak, there exist several similarities between gravitation and electromagnetism. In particular, one can have waves in the gravitational field that are able to carry energy from a source to a receptor.

If one assumes that the metric of spacetime is close to the flat Minkowski metric $\eta_{\mu\nu}$, that is, $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ with $|h_{\mu\nu}| << 1$ in Cartesian coordinates, the Einstein equations of general relativity, expanding to linear order in $h_{\mu\nu}$, become

$$\begin{aligned} 0 =& R_{\mu\nu} - \tfrac{1}{2}\eta_{\mu\nu}R \\ =& \tfrac{1}{2}\left(\partial_\sigma\partial_\nu h^\sigma_{\mu\nu} + \partial_\sigma\partial_\mu h^\sigma_\nu - \partial_\mu\partial_\nu h \right. \\ & \left. - \Box h_{\mu\nu} - \eta_{\mu\nu}\partial_\mu\partial_\nu h^{\mu\nu} + \eta_{\mu\nu}\Box h\right) \end{aligned} \qquad [1]$$

These do not look like wave equations. However, if one chooses "harmonic coordinates," $\Box x^\mu = 0$, where $\Box$ is the d'Alembertian constructed with the full metric and then linearized, the vacuum Einstein equations become

$$\Box h_{\mu\nu} = 0 \qquad [2]$$

where $\Box$ is the d'Alembertian computed in the flat Minkowski metric.

Just as in electromagnetism the motion of charges produces waves, the motion of masses produces waves in the gravitational field. In the above wave equations, one would have nonzero right-hand sides if masses were present. In electromagnetism, the conservation of charge implies that the lowest order of "structure" a source must have to produce electromagnetic waves is that of a time-dependent dipole. In the gravitational field, the conservation of momentum implies that the lowest multipolar order of a source of gravitational waves must be a quadrupole. Moreover, gravity is a weaker force than electromagnetism when one considers usually available situations. One can exert forces of the orders of fractions of Newton with electromagnetic

charges easily collected in tabletop experiments. To produce similar amounts of gravitational force, one needs large quantities of mass. This last fact, coupled with the quadrupolar nature of the sources of gravitational waves, makes their production quite challenging in experimental terms. The luminosity of a gravitational wave source is given by the celebrated Einstein quadrupole formula,

$$L = \left(\frac{G}{5c^5}\right) \sum_{j=1,k=1}^{3} \left(\dddot{I}_{jk}\right)^2 \qquad [3]$$

where $G$ is Newton's gravitational constant, $c$ is the speed of light, and $\dddot{I}_{ij}$ is the third-order time derivative of the traceless part of the quadrupole mass moment of the source.

Gravity is, however, a dominant force if one considers the universe at large (say, at least planetary) scales. There one would expect gravitational waves to play some role in the dynamics of the systems. In such systems, the presence of gravitational waves has indeed been experimentally confirmed. We know of a system of two pulsars in mutual orbit, PSR1913 + 16, whose orbit has been tracked with enough accuracy via radioastronomy to make the influence of gravitational waves observable. The motion of the pulsars makes the system an emitter of gravitational waves. Since the waves carry away energy, the orbit of the system decreases in radius and the period of oscillation increases. The system has now been tracked for over 20 years, and the prediction of the emitted amount of energy in gravitational waves due to general relativity has been confirmed with a very significant degree of accuracy. Penrose was the first to notice that if one considers how accurately Newton's theory plus the corrections due to general relativity predict the positions of the pulsars in their orbit, this is in fact the most accurately verified physical prediction ever.

Technically, even the existence of gravitational waves at a conceptual mathematical level, was an open problem for many years. Since the correct description of the waves is through the general theory of relativity, a "gravitational wave" should really be viewed as a "ripple in spacetime." Disentangling if such ripples are a true physical effect or a time-dependent coordinate transformation that propagates – to use the words of Eddington – "with the speed of thought" took quite a bit of technical development within the general theory of relativity. It was only in the 1960s that a clear enough conceptual picture was developed to determine that gravitational waves were indeed a true physical phenomenon akin to electromagnetic waves. And in particular that one can unambiguously characterize them as transporting energy, momentum, and angular momentum from a source to an observer.

Gravitational waves are as difficult to detect as they are to produce. Since all masses fall in the same way in a gravitational field, one needs to couple to the gradients of the field to detect gravitational waves, which diminishes the efficiency. Attempting to produce gravitational waves via mechanical means in the lab (e.g., by rotating a bar of metal) produces too little luminosity, and in addition, the relatively low frequency implies that the wave zone is far away, which further decreases the chances for detection. Up to date, no one has succeeded in producing a Hertz-like experiment for gravitational waves and the jury is still out on the issue if future technologies (e.g., the use of superconductors to produce waves of gigahertz frequency) will ever allow such an experiment.

Efforts to attempt to detect gravitational waves produced by astrophysical phenomena started in the 1960s with pioneering work by Weber. The initially proposed technology for detection was the construction of large ($\sim 1$ ton) resonant bars. The idea was to use sensitive technology to measure the resonance of the bar as gravitational waves of astrophysical origin impact on it. Gravitational waves manifest themselves as a stretching and contraction of lengths. The contraction or stretching is proportional to the length of the object considered and is therefore characterized by a dimensionless number, the "strain" $\Delta L/L$ usually called "$h$." Conservative current estimates of possible astrophysical sources state that on Earth one should not expect strains larger than $10^{-22}$ for events that repeat more frequently than a few times every year. Detectors with bar technology are approaching their fundamental quantum limits with strains that appear to be too large for detection to be ensured. This led to the proposal of a new technology, the use of Michelson-type interferometers to detect the waves. Currently, several interferometric detectors are being built in the US, Europe, Japan, and Australia that expect to achieve enough sensitivity for detection within a few years. Contrary to the bars, which are quintessentially narrow-band detectors (most bars operate $\sim 900$ Hz with a bandwidth of $\sim 10$ Hz), interferometric detectors are broadband. Current detectors have a sensitivity curve limited by various sources of noise that make them suitable for detection within the 10 Hz–1 kHz band. The broadband nature of the detectors opens several opportunities for the use of data analysis techniques that can allow the detection of gravitational waves that have strains even lower than the noise of the detectors. Moreover, several of

the candidate events "evolve" in frequency as they emit gravitational waves (in the case of the binary pulsar, for instance, the frequency "sweeps up" as the system loses energy), and such evolution could be monitored with interferometric detectors. This would allow several insights into the physics of the observed systems.

An important limitation of any type of detector based on Earth is that the seismic noise increases quite significantly below 10 Hz. Even if seismic isolation allowed sensitivities below 10 Hz, gravity gradients due to Earth's seismic motion and due to clouds would limit ground-based detectors to 1 Hz and above. The frequency at which a system emits gravitational waves is inversely proportional to the system's mass (a simple way to see this is to realize that larger systems move proportionally slower to their size). However, larger systems generically have more mass and therefore consequently emit larger amounts of energy in gravitational waves. This suggests that setting up detectors in space, free of the constraints of seismic noise, would offer significant promise in detecting gravitational waves. Currently, there is a proposal for a space-borne gravitational wave detector consisting of three satellites in a solar orbit that trails that of Earth. Lasers would be sent between the satellites to track their relative positions, which will be separated by 5 million kilometers. Such a detector would be sensitive in frequencies of $10^{-4}$–$10^{-2}$ Hz. In such a frequency band, one expects that compact objects plunging into supermassive black holes and other sources will be readily available. Detection of gravitational waves on Earth is considered marginal, in the sense that conservative current estimates cannot guarantee that there will be enough events to make the detection successful at significant event rates. Conversely, for the detectors in space, detection should be guaranteed at high event rates.

Possible sources of gravitational waves to be detected by the Earth-based interferometric detectors are:

1. Binary systems of compact objects. As the system orbits, it emits gravitational waves, which makes the orbit shrink in size and the orbiting period shorter with the objects eventually merging together. Potential systems include black hole binaries, neutron star binaries and mixed black hole/neutron star binaries. As the system sweeps up in orbital speed towards the merger, so does the frequency of the gravitational waves emitted. For binaries of neutron stars, which usually have masses slightly larger than the mass of the Sun, the last few minutes of the binary inspiral will be detectable by the current generation of gravitational wave detectors, up to a distance of several mega-parsecs for the initial detectors, increasing to a few hundreds of mega-parsecs for improvements planned for the next few years. For black hole binaries, since the masses can be larger, one expects larger signal-to-noise ratio for the same distance or to be able to detect at larger distances.
2. Spinning neutron stars that develop "mountains" or other irregularities in the surface would produce gravitational waves of small amplitude but of a very regular periodic nature. This makes them prime candidates for data analysis techniques that could exhibit the presence of the wave even though it is weaker than the background noise of the interferometers. Integration periods of several months may be needed for detection, depending on the size of the asymmetries in the neutron stars.
3. Supernovas or other violent events are obviously possible sources of gravitational waves. However, the quadrupole nature of the waves requires the events to be asymmetric in order to produce gravitational waves. Current numerical models of supernovas are not accurate enough to predict in a clear way the level of asymmetry to make reliable predictions of how frequently and at what intensity could these types of sources be detected.
4. The primordial background of gravitational waves produced in the big bang is not expected to be detectable by the Earth-based detectors. The precise amplitude of the background is unknown, depending on details of cosmological models. The detectors are likely to be able to constrain some of the models that predict large amplitudes for the gravitational wave background.

For the space-based detectors, the situation is more favorable, since there exist sources of gravitational waves that are guaranteed to be detected. Potential sources of gravitational waves are:

1. *Merger of the supermassive black holes at the centers of two galaxies*. Given the large amounts of mass involved, they would be easily detected and very precise measurements of the system's parameters and of various general relativistic behaviors could be possible. Such systems should be detectable all across the universe, although it is not expected that such systems form for redshifts larger than 30.
2. *Inspiral of compact objects into the supermassive black holes at the centers of galaxies* (neutron

stars, white dwarfs, solar-sized black holes). These processes will allow the usage of gravitational waves to map precisely the gravitational field of the supermassive object.

3. *White-dwarf binaries and low-mass X-ray binaries*. There exist about a dozen such systems optically observable with gravitational wave frequencies above 0.1 mHz that the space-based detectors should be able to detect. There is likely to be a large population of other systems that are also detectable and are not optically visible. In fact, there may be so many of these sources that time resolution would be impossible, and they would form a random background.
4. *Collapse of supermassive stars*. The formation mechanism for the supermassive black holes in the centers of galaxies is still uncertain. One possibility is that they stem from the collapse of supermassive stars, and in that case a potentially significant emission of gravitational waves could take place.
5. *Primordial background of gravitational waves*. Unfortunately, the abundance of white-dwarf binaries as a source is expected to cloud the ability of the space detectors to observe primordial gravitational waves in an important portion of the spectrum of the instrument, although it appears possible at low frequencies, where it could compete with the bounds set by pulsar timing.

The current Earth-based gravitational wave projects include the LIGO project in the US, funded by the National Science Foundation and jointly operated by Caltech and MIT and a consortium of institutions known as the LIGO Science Collaboration. LIGO consists of two 4 km long Fabry–Perot recycled Michelson interferometers, one in Hanford, WA, and one in Livingston, LA. In Europe, the GEO600 project is a 600 m dual-recycled interferometer near Hanover in Germany and the Virgo project is a 3 km interferometer near Pisa in Italy operated by a French–Italian consortium with a similar optical configuration as LIGO. TAMA300 is a 300 m interferometer in Japan also with the same configuration as LIGO. When all these detectors are in operation, sources seen in coincidence could be localized by triangulation. TAMA is now operating close to design sensitivity, GEO600 and LIGO are likely to operate at design sensitivity in 2006, with VIRGO following close behind. The space-based interferometer project is called the LISA project and is planned as a joint NASA/ESA project. ESA has approved a launching date for 2015, but it is plausible that the mission could be launched at an earlier date.

A direct detection of gravitational waves would be a breakthrough in experimental science, as well as a confirmation of the dynamic nature of gravity in general relativity. Once the detection of gravitational waves becomes a routine matter, one can imagine a revolution in astronomy as one uses gravitational waves to "see" the universe. Since they are so hard to produce and interfere with, gravitational waves become an excellent type of "light" to look at the universe with. Gravitational waves will be produced by important concentrations of mass, correlating well with "interesting" astronomical processes, and is not expected to be affected by the presence of dust or other interfering objects that could easily obscure electromagnetic waves. In addition to this, one has several "standard candles" for gravitational waves (e.g., most neutron stars have masses that differ by a few percent from 1.4 solar masses). This could allow, for instance, to determine with a high degree of accuracy the Hubble constant. Gravitational waves will also provide insight into the nuclear equation of state that holds in the interior of compact objects like neutron stars. Contrary to ordinary electromagnetic radiation, which "decoupled" from matter only when the universe became cool enough after the big bang, gravitational waves could be used to probe the universe further into the past. The detection could also prove that gravitational waves travel at the speed of light, a prediction of general relativity and other theories.

An interesting observation is that most astrophysical objects that are quite visible in the electromagnetic spectrum are unlikely to be visible in terms of gravitational waves, and vice versa. This makes the information we will gather from gravitational wave astronomy complementary to what we learn from optical (electromagnetic) astronomy. Moreover, it should be noted that wavelengths of electromagnetic waves are typically very small compared to the size of the astronomical objects they depict. This is due to the fact that the waves are really not produced by the objects themselves but by atoms on the surface of the objects or in regions nearby, usually very hot and in gaseous form. In contrast, gravitational waves are produced by the bulk matter of astronomical objects and their wavelengths are expected to be long as compared to the objects that produce them. They are more akin to a sound than to light in this respect, another reason to suspect that the information we will get from them is unlike any information obtained electromagnetically.

Gravitational waves are likely to bring great surprises. Every time a new window has been opened on the universe – for instance, the use of radio waves – our view of the universe has been revolutionized. Given how differently they operate

at a detailed level with respect to radio waves, the surprises from gravitational waves used as tools to view the universe are potentially even greater.

*See also:* Asymptotic Structure and Conformal Infinity; Computational Methods in General Relativity: The Theory; General Relativity: Experimental Tests; General Relativity: Overview.

## Further Reading

Bartusiak M (2003) *Einstein's Unfinished Symphony: Listening to the Sounds of Space-Time*. New York: Berkley Publishing Group.
Cutler C and Thorne K (2002) An overview of gravitational wave sources, Arxiv.org:gr-qc/0204090.
Douglass D and Braginsky V (1979) Gravitational radiation experiments. In: Hawking S and Israel W (eds.) *General Relativity: An Einstein Centenary Survey*. Cambridge: Cambridge University Press.
Kennefick D (1997) Controversies in the history of the radiation reaction problem in general relativity, Arxiv.org:gr-qc/9704002.
Saulson P (1994) *Fundamentals of Interferometric Gravitational Wave Detectors*.Singapore: World Scientific.
Schutz B (1984) Gravitational waves on the back of the envelope. *American Journal of Physics* 52: 412.
Taylor J (1994) Nobel lecture: binary pulsars and relativistic gravity. *Reviews of Modern Physics* 66: 711.
Thorne K (1987) Gravitational radiation. In: Hawking S and Israel W (eds.) *300 Years of Gravitation*. Cambridge: Cambridge University Press.
Thorne K (1995a) Gravitational waves, Arxiv.org:gr-qc/9506086.
Thorne K (1995b) *Black Holes and Time Warps: Einstein's Outrageous Legacy*. New York: W W Norton.
Throne K (1997) Gravitational radiation: a new window on the universe, Arxiv.org:gr-qc/9704042.
The LIGO project web page, which includes links to all the other experimental efforts: http://www.ligo.caltech.edu.

# Growth Processes in Random Matrix Theory

**K Johansson**, Kungl Tekniska Högskolan, Stockholm, Sweden



## Introduction

Probability distributions coming from random matrix theory (RMT), RMT laws, occur in different contexts, notably in quantum physics and in number theory. RMT laws are also seen in certain local random growth models and related problems in discrete probability, random permutations, exclusion processes, and random tilings (dimer models). In these models limit laws for height/shape fluctuations are given by limit laws from RMT, in particular the largest eigenvalue or Tracy–Widom distributions. These models belong to the Kardar–Parisi–Zhang (KPZ) universality class. Models in this class have two universal exponents, 1/3 describing the interface fluctuations and 2/3 describing the correlations in the transversal direction. By a local random growth model, we mean a model where the random growth mechanism is local in that it does not depend on the global geometry as in diffusion limited aggregation (DLA). Typically there is also some smoothing mechanism. The connection with RMT can only be established for special exactly solvable models. Below we discuss a basic model based on a last-passage percolation problem, which translates into a polynuclear growth (PNG) process. Other models that can be treated are in a sense variations of this model. Point processes with determinantal correlation functions play a central role in RMT and in the analysis of the basic model and we start by discussing these. The basic model has several different interpretations that will be outlined. Another basic tool, which can be formulated in different ways, is the Robinson–Schensted–Knuth (RSK) correspondence well known in combinatorics. One approach is in terms of nonintersecting paths which translates into a multilayer PNG process. Limit theorems can be formulated for the height above a fixed location, and also for the whole height function in terms of the Airy process which extends the Hermitian Tracy–Widom distribution $F_2$. It is expected that several results should generalize to a broader class of models. There is a natural universality problem of extending the validity of the RMT laws.

### Determinantal Processes

Point processes with determinantal correlation functions play an important role in the exactly solvable models. We consider probability measures on $\Lambda^n, \Lambda \subseteq \mathbb{R}$, of the form

$$\frac{1}{Z_n}\det(\phi_i(x_j))_{i,j=1}^n \det(\psi_i(x_j))_{i,j=1}^n \, d^n\mu(x) \qquad [1]$$

which can be thought of as describing random points in $\Lambda$ at positions $x_1, \ldots, x_n$. Here, $\mu$ is a reference measure on $\Lambda$, for example, Lebesgue or counting measure, $Z_n$ a normalization constant, and $\phi_i, \psi_i$ given functions. A measure of this form has determinantal correlation functions in the sense that the density, with respect to $d^m\mu(y)$, of particles at $y_1, \ldots, y_m$ is

$$\rho(y_1, \ldots, y_m) = \det(K_n(y_i, y_j))_{i,j=1}^m \qquad [2]$$

There is an explicit formula for the correlation kernel $K_n$ in terms of the functions $\phi_i, \psi_i$.

The eigenvalue measures in the basic random matrix ensembles have the form

$$\frac{1}{Z_n}|\Delta_n(x)|^\beta \prod_{j=1}^n w(x_j)\, \mathrm{d}^n\mu(x) \qquad [3]$$

where $\Delta_n(x) = \det(x_j^{i-1})_{i,j=1}^n$ is Vandermonde's determinant, $x \in \Lambda^n$ and $x_1, \ldots, x_n$ are the eigenvalues. For the Gaussian unitary ensemble (GUE$_n$), $Z_n^{-1}\exp(-\operatorname{tr} M^2)\mathrm{d}M$ of $n \times n$ Hermitian matrices $M$, we have $\beta = 2, \Lambda = \mathbb{R}, w(x) = \exp(-x^2)$ and $\mu$ the Lebesgue measure. For the Laguerre unitary ensemble (LUE$_{n,\nu}$) of complex covariance matrices, $M^*M$, where $M$ is an $(n+\nu) \times n$-matrix with standard complex Gaussian elements, we have $w(x) = x^\nu \mathrm{e}^{-x}, \nu \geq 0, \beta = 2, \Lambda = [0, \infty)$ and $\mu$ the Lebesgue measure. The $\beta = 2$ case of [3] can be put into the form [1] and hence has determinantal correlation functions. In this case the correlation kernel can be expressed in terms of the normalized orthogonal polynomials $p_k(x)$ with respect to $w(x)\,\mathrm{d}\mu(x)$ on $\Lambda$. Because of this when $\beta = 2$ the ensemble [3] is referred to as an orthogonal polynomial ensemble (OPE). The kernel is given by

$$K_n(x,y) = \sum_{k=0}^{n-1} p_k(x)p_k(y)(w(x)w(y))^{1/2} \qquad [4]$$

A consequence of [2] is that the probability of finding no particle in a set $J \subseteq \Lambda$ is given by a Fredholm determinant,

$$\mathbb{P}[\text{no particle in } J] = \det(I - K_n)_{L^2(J,\mu)} \qquad [5]$$

In particular the distribution function $F(\xi)$ of the largest eigenvalue or rightmost particle $x_{\max} = \max_{1 \leq j \leq n} x_j$ in an OPE is given by [5] with $K_n$ as in [4] and $J = (\xi, \infty)$.

## A Basic Model

Let $(w(i,j))_{(i,j)\in\mathbb{Z}_+^2}$ be independent geometric random variables with parameter $a_ib_j$,

$$\mathbb{P}[w(i,j) = k] = (1 - a_ib_j)(a_ib_j)^k \qquad [6]$$

$k \geq 0$ and $0 \leq a_ib_j < 1$. As a limiting case we can obtain exponential random variables. Consider the last-passage time

$$G(M,N) = \max_\pi \sum_{(i,j)\in\pi} w(i,j) \qquad [7]$$

where the maximum is over all up/right paths $\pi$ from $(1,1)$ to $(M,N)$, that is, $\pi = \{(i_1,j_1), \ldots, (i_m,j_m)\}$ with $(i_{k+1},j_{k+1}) - (i_k,j_k) = (1,0)$ or $(0,1), (i_1,j_1) = (1,1)$ and $(i_m,j_m) = (M,N), m = M+N-1$. We can also think of this as a zero-temperature directed polymer, by thinking of the $w(i,j)$:s as (minus) energies and $\pi$ as random walk paths.

As will be explained in some more detail below, if the $w(i,j)$:s are exponential with mean 1 and $M \geq N$, then $G(M,N) = \lambda_{\max}$ in distribution, $M \geq N$, where $\lambda_{\max}$ is the largest eigenvalue in LUE$_{N,M-N}$. Hence, in this case $G(M,N)$ behaves exactly like a largest eigenvalue. If the $w(i,j)$:s are geometric with parameter $q$, then $G(M,N)$ has the same distribution as the rightmost particle in an OPE, namely [3], with $\beta = 2, w(x) = \binom{M-N+x}{x}q^x$ and $\mu$ the counting measure on $\Lambda = \mathbb{N}$, called the Meixner ensemble. Since in this case the relevant orthogonal polynomials are discrete the ensemble is referred to as a discrete OPE.

The random variables $\{G(M,N)\}_{(M,N)\in\mathbb{Z}_+^2}$ have two interpretations related to random growth. It follows from [7] that

$$G(M,N) = \max(G(M-1,N), G(M,N-1)) + w(M,N) \qquad [8]$$

This can be thought of as a growth rule. We change variables by letting $G(M,N) = h(M-N, M+N-1)$ and $w(M,N) = \omega(M-N, M+N-1)$ with $w(M,N) = 0$ if $(M,N) \notin \mathbb{Z}_+^2$. Then

$$h(x,t+1) = \max(h(x-1,t), h(x,t), h(x+1,t)) + \omega(x,t) \qquad [9]$$

$x \in \mathbb{Z}, t \in \mathbb{N}, h(x,0) \equiv 0$, and $\omega(x,t) = 0$ if $|x| \geq t$ or if $x - t$ is even. We can extend it to the whole real line by letting $h(x,t) = h([x],t)$. The growth rule [9] is a discrete polynuclear growth (PNG) model. Up-steps in the interface, $x \to h(x,t)$ (see the top curve in Figure 1), move at unit speed to the left and down-steps move at unit speed to the right and they merge at collision. On top of this smoothing mechanism, we have random deposition given by $\omega(x,t)$. Looking at the definition of $\omega$, we see that all deposition up to time $t$ is on top of a basic layer $(-t,t)$. The asymptotic shape will look like a droplet and this setting of PNG is called the droplet geometry. We see that height fluctuations are directly related to fluctuations of $G(M,N)$.

We get another growth model, the corner growth model, somewhat similar to the classical Eden growth model, by considering the random shape (see Figure 2),

$$\Omega(n) = \{(M,N) \in \mathbb{Z}_+^2; G(M,N) + M + N - 1 \leq n\} + [-1,0]^2 \qquad [10]$$

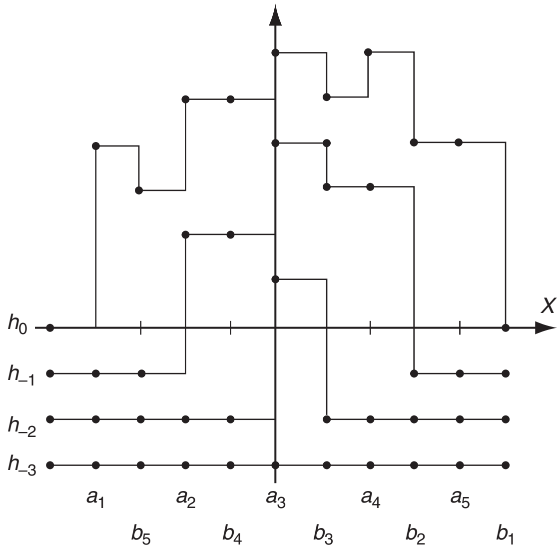



**Figure 1** Multilayer PNG model.

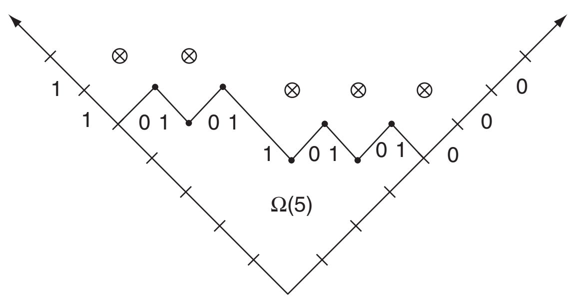



**Figure 2** Corner growth model at time $n=5$. The crosses are the possible growth sites.

The complement of this set in $\mathbb{Z}_+^2$ has a boundary $B(n)$ which we can think of as an interface. By [8] and the lack of memory property of the geometric/exponential distribution, the region $\Omega(n)$ grows by adding new squares independently at each corner of $B(n)$ with geometric/exponential waiting times (see Figure 2). If we look at $B(n)$ in a coordinate system with $M=N$ as vertical axis and write a 1 for every unit down-step on $B(n)$ and a 0 for every unit up-step (see Figure 2), the corner growth dynamics translates into the totally asymmetric simple exclusion process (TASEP), in discrete or continuous time, with initial configuration $\dots 1111000 \dots$ . As shown by Jockush, Propp, and Shor, $\Omega(n)$ also occurs in a uniform random domino tiling of a region called the Aztec diamond (see Figure 3). The shape $\Omega(n)$, when $q=1/2$, has the same law as the completely regular (frozen) North Polar Region (NPR) in the tiling and hence the boundary fluctuations of the NPR are related to the fluctuations of $G(M,N)$. The NPR in Figure 3 has the same shape as $\Omega(5)$ in Figure 2. This connects the models considered here with dimer or tiling problems in two-dimensional equilibrium statistical mechanics.

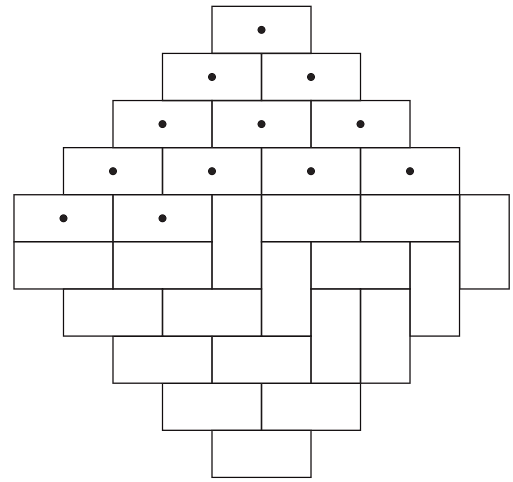


**Figure 3** Domino tiling of an Aztec diamond of size $n=5$. Dominos marked by dots form the NPR.

Consider a (Poissonized) random permutation $\sigma$ from $S_N$, where $N$ is a Poisson($\alpha$) random variable. Let $L(\alpha)$ denote the length of the longest increasing subsequence in $\sigma$, for example, $\sigma=316452$ has $L=3$. By thinking of the representation of a permutation by its permutation matrix, we see that $G(N,N)$ with $w(i,j)$ geometric with parameter $q=\alpha/N^2$ converges to $L(\alpha)$ in distribution as $N\to\infty$. We call this limit the Poisson limit. Taking this limit in the PNG process yields the Prähofer–Spohn continuous time PNG (cont-PNG) model, which is similar to the discrete PNG defined above but where all steps have unit size and we have continuous time dynamics with deposition events according to a two-dimensional spacetime Poisson process. The study of $L(\alpha)$, and its de-Poissonization when $N$ is nonrandom, is known as Ulam's problem in combinatorial probability.

## The RSK Correspondence

The mapping of the last-passage problem [7] into a determinantal process is based on the RSK correspondence. This correspondence maps the integer matrix $(w(i,j))_{1\le i,j\le M}$ bijectively to a pair of semi-standard Young tableaux $(P,Q)$ with common shape $\lambda$, which is a partition $\lambda=(\lambda_1,\lambda_2,\dots)$ of $\sum_{1\le i,j\le M} w(i,j)$. This map has the property that $G(M,N)=\lambda_1$, the length of the first row in the Young diagram. From the combinatorial definition of the Schur polynomials $s_\lambda$ it follows that the measure [6] on the integer matrix is mapped to a

probability measure on partitions, the Schur measure, given by

$$\mathbb{P}_{\text{Schur}}[\lambda] = \frac{1}{Z} s_\lambda(a_1, \ldots, a_M) s_\lambda(b_1, \ldots, b_M) \qquad [11]$$

This measure has determinantal correlation functions if we think of $x_i = \lambda_i - i$ as the positions of particles in $\mathbb{Z}$. If we use $x_i = \lambda_i + N - i$ as variables and specialize to $a_1 = \cdots = a_M = \sqrt{q}, b_1 = \cdots = b_N = \sqrt{q}$ and $b_j = 0$ for $j > N$ we get the Meixner ensemble. The case of exponential random variables, for example the relation to LUE discussed above, is obtained from the Meixner ensemble by taking an appropriate limit. In the Poisson limit we get the Poissonized Plancherel measure,

$$\mathbb{P}^{\alpha}_{\text{Plan}}[\lambda] = \sum_{N=0}^{\infty} \frac{e^{-\alpha/\alpha^n}}{N!} \qquad [12]$$

where $\mathbb{P}_{\text{Plan},N}[\lambda] = (\dim \lambda)^2/N!$ if $\lambda$ is a partition of $N$ and 0 otherwise. Here $\dim \lambda$ is the dimension of the irreducible representation of $S_N$ labeled by $\lambda$. In the work of Borodin and Olshanski in representation theory various measures on partitions with determinantal correlation functions occur naturally. Also Okounkov and co-workers have used the Plancherel and Schur measures in Gromov–Witten theory. The correlation kernel for the Plancherel measure represented as the point process $(x_i)_{i\geq 1}$ in $\mathbb{Z}$ with $x_i = \lambda_i - i$ has the correlation kernel, called the discrete Bessel kernel,

$$\begin{aligned} B^{\alpha}(x,y) = {} & \sqrt{\alpha}(x-y)^{-1} \\ & \times (J_x(2\sqrt{\alpha}) J_{y+1}(2\sqrt{\alpha}) \\ & - J_{x+1}(2\sqrt{\alpha}) J_y(2\sqrt{\alpha}) \end{aligned} \qquad [13]$$

where $J_n$ is the ordinary Bessel function. The random variable $L(\alpha)$ has the same distribution as $\max x_j + 1$. Hence, by [5],

$$\mathbb{P}[L(\alpha) \leq n] = \det(I - B^{\alpha})_{\ell^2(\{n, n+1, \ldots\})} \qquad [14]$$

The random variable $L(\alpha)$ also gives the height above the origin in cont-PNG.

There is a geometric interpretation of RSK going back to Viennot. The pair $(P, Q)$ is represented as a family of nonintersecting paths in a directed graph. These paths can be obtained by running a multilayer version of the PNG process where the size of collisions are deposited as growth in lower layers which evolve according to the same PNG dynamics. The information lost in the collisions is recorded in the lower layers. This can be done also for $\{w(i,j)\}_{i+j-1\leq t}$ and leads to a multilayer version of [9], $\{h_{-j}(x,t)\}_{j=1}^{\infty}$, where $h_{-j}(x,0) \equiv -j, h_{-j}(\pm t, t) = -j$, and $h(x,t) = h_0(x,t)$ is the top path (see Figure 1).

The Karlin–McGregor theorem or the Gessel–Viennot method say that the weight (probability) of a family of nonintersecting paths with fixed initial and final positions on a weighted directed acyclic graph is given by a determinant. It follows that the probability of a certain configuration $\{h_{-j}(0,t)\}_{j\geq 1}$ is given by a product of two determinants and hence has the form [1]. In Figure 1, the weights of the horizontal line segments will be 1, whereas each unit vertical step has weight $a_i$ or $b_j$ as indicated in the figure. This leads to [9] using the Jacobi–Trudi formula for the Schur polynomial.

## Limit Theorems

The existence of a limit shape in a model like [6] with $w(i,j)$ independent random variables, and in related problems follows by a subadditivity argument, although explicit shapes are only known in a few cases. The formalism described above makes it possible to get more detailed results about the fluctuations around the limit shape, like a central-limit theorem, but with a non-normal limit law. We know that $G(M,N)$ has the same distribution as $x_{\max} - N + 1$, where $x_{\max}$ is the rightmost particle in the Meixner ensemble. This, together with [4], [5], and an asymptotic analysis of the Meixner polynomials, gives

$$\mathbb{P}\left[G(M,N) \leq \omega(\gamma,q)N + \xi\sigma(\gamma,q)N^{1/3}\right] \to F_2(\xi) \qquad [15]$$

as $N \to \infty, M \to \infty, M/N \to \gamma \geq 1$, where

$$\omega(\gamma,q) = \frac{(1+\sqrt{q\gamma})^2}{1-q} - 1 \qquad [16]$$

and

$$\sigma(\gamma,q) = \frac{(q/\gamma)^{1/6}}{1-q} (\sqrt{\gamma} + \sqrt{q})^{2/3} (1+\sqrt{q\gamma})^{2/3} \qquad [17]$$

The limiting distribution function $F_2$ is the Tracy–Widom distribution given by

$$F_2(\xi) = \det(I - A)_{L^2(\xi,\infty)} \qquad [18]$$

where

$$A(x,y) = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\mathrm{Ai}(y)}{x-y} \qquad [19]$$

is the Airy kernel. It is also the limiting largest eigenvalue distribution for GUE$_n$,

$$\lim_{N\to\infty} \mathbb{P}\left[\frac{\sqrt{2n}\lambda^{(n)}_{\max} - 2n}{n^{1/3}} \leq \xi\right] = F_2(\xi) \qquad [20]$$

The function $F_2$ can also be expressed in terms of a Painlevé II function. The limit theorem [15]

translates into a fluctuation result for the height function in the corner growth and the PNG models, saying that the height fluctuations above a fixed location at time $t$ are of order $t^{1/3}$ and given by the $F_2$-distribution. Here we see the KPZ exponent 1/3.

For the length $L(\alpha)$ of the length of a longest increasing subsequence in a random permutation or the height above the origin in the cont-PNG, [14] and asymptotics for Bessel functions yield

$$\mathbb{P}[(L(\alpha) - 2\sqrt{\alpha})/\alpha^{1/6} \le \xi] \to F_2(\xi) \qquad [21]$$

as $\alpha \to \infty$. This result was first proved by Baik, Deift, and Johansson using a Toeplitz determinant formula (Gessel's formula) for the left-hand side of [14] and the Deift–Zhou nonlinear steepest descent method for oscillatory Riemann–Hilbert problems. The above limit theorems can be extended to limit theorems for the whole point process rescaled around the rightmost point. This results in a limiting determinantal point process given by the Airy kernel [19].

## The Airy Process

From the point of view of the growth processes, for example, the PNG process [7], it is natural to consider a scaling limit of the whole height function $x \to h(x,t)$ as $t \to \infty$. Looking at the height configuration in the multilayer growth process $x \to \{h_{-j}(x,t)\}_{j\ge 1}$ at different locations $x_1, \dots, x_r, x_{r+1}, \dots, x_{M-1}$ leads, via the Karlin–McGregor or Gessel–Viennot method, to probability measures of the form

$$\frac{1}{Z}\prod_{r=0}^{M-1} \det\left(\phi_{r,r+1}\left(y_i^r, y_j^{r+1}\right)\right)_{i,j=1}^{n} \qquad [22]$$

with $y^0$ and $y^M$ fixed configurations. Here, in the discrete PNG model, $\phi_{r,r+1}(x,y)$ is the transition probability (weight) to go from height $x$ to height $y$ between positions $x_r$ and $x_{r+1}$. This measure generalizes [1] and it also has determinantal correlation functions. Measures of this form also arise in multimatrix models and in Dyson's Brownian motion model, $t \to M(t), t \in \mathbb{R}$, for Hermitian matrices, which is a Gaussian multimatrix model. The elements of the time-dependent Hermitian matrix $M(t)$ evolve according to independent Ornstein–Uhlenbeck processes and we have the transition kernel

$$Z^{-1} \exp[-\mathrm{tr}(M(t) - qM(0))^2/(1-q^2)] \qquad [23]$$

where $q = \exp(-t)$. This process has GUE as its stationary distribution. The Harish–Chandra/Itzykson–Zuber integral can be used to show that the joint eigenvalue measure for $M(t_1), \dots, M(t_{M-1})$ has the form [22] and hence has determinantal correlation functions. The correlation kernel is the extended Hermite kernel,

$$K_n(\tau, x; \sigma, y) = \begin{cases} \displaystyle\sum_{k=1}^{\infty} \mathrm{e}^{-k(\tau-\sigma)} p_{n-k}(x) p_{n-k}(y)\, \mathrm{e}^{-\frac{1}{2}(x^2+y^2)} & \text{if } \tau \ge \sigma \\ \displaystyle -\sum_{k=-\infty}^{0} \mathrm{e}^{-k(\tau-\sigma)} p_{n-k}(x) p_{n-k}(y) \mathrm{e}^{-\frac{1}{2}(x^2+y^2)} & \text{if } \tau < \sigma \end{cases} \qquad [24]$$

with $p_k$ the normalized Hermite polynomials, $p_k \equiv 0$ if $k < 0$. Notice that this reduces to the Hermite kernel [4] when $\tau = \sigma$. This machinery can be used to show that the largest eigenvalue process $t \to \lambda_{\max}^{(n)}(t)$ induced by $M(t)$ converges in the sense of finite-dimensional distributions to a limiting process, the Airy process,

$$\left(\sqrt{2n}\lambda_{\max}^{n}(n^{-1/3}t) - 2n\right)/n^{1/3} \to \mathcal{A}(t) \qquad [25]$$

as $n \to \infty$. The Airy process $\mathcal{A}(t)$, which is a stationary process, can be viewed as the top curve of a multilayer process $t \to (\mathcal{A}_{-j}(t))_{j\ge 1}, \mathcal{A}(t) = \mathcal{A}_0(t)$ such that the point process $\{\mathcal{A}_{-j}(t_k)\}_{1\le k\le M, j\ge 0}$ has determinantal correlation functions with correlation kernel

$$A(\tau, \xi; \tau', \xi') = \begin{cases} \displaystyle\int_0^{\infty} \mathrm{e}^{-\lambda(\tau-\tau')} \mathrm{Ai}\,(\xi+\lambda) \mathrm{Ai}\,(\xi'+\lambda) \mathrm{d}\lambda & \text{if } \tau \ge \tau' \\ \displaystyle -\int_{-\infty}^{0} \mathrm{e}^{-\lambda(\tau-\tau')} \mathrm{Ai}\,(\xi+\lambda) \mathrm{Ai}\,(\xi'+\lambda)\, \mathrm{d}\lambda & \text{if } \tau < \tau' \end{cases} \qquad [26]$$

the extended Airy kernel, which reduces to the ordinary Airy kernel [19] when $\tau = \tau'$. The Airy process can be viewed as an extension of the Tracy–Widom distribution $F_2$. For the PNG model above, the multilayer process is described by an extended kernel, which in the cont-PNG is an extended version of the discrete Bessel kernel [13]. In a suitable scaling limit, this extended kernel converges to the extended Airy kernel. For the PNG process [7], this leads to the limit law

$$(dN^{1/3})^{-1}\left[h\left(2d^{-1}\frac{1+\sqrt{q}}{1-\sqrt{q}}N^{2/3}\tau, 2N-1\right) - \frac{2\sqrt{q}}{1-q}N\right] \to \mathcal{A}(\tau) \qquad [27]$$

as $N \to \infty$, where $d = (1-q)^{-1}(\sqrt{q})^{1/3}(1+\sqrt{q})^{1/3}$. Notice the exponent 2/3 which is the second KPZ exponent. This exponent can also be seen in the transversal fluctuations of the maximal paths in [6] for $G(N, N)$. These are superdiffusive, they have fluctuations of order $N^{2/3}$ around the diagonal, compared to $N^{1/2}$ for random walk paths between the same points. A fluctuation result like [27] can also be proved for the corner growth model and hence also for the Aztec diamond. The boundary of the NPR suitably rescaled converges to the Airy process.

## Variations

Above we discussed one possible geometry, the droplet, for the PNG process. If we start with $h(x, 0) \equiv 0$ and allow random depositions along the whole line, we get an interface that is macroscopically flat, and not curved as in the droplet case. In this case, the height fluctuations above a fixed location at time $t$ are again of size $t^{1/3}$ and described by the Gaussian Orthogonal Ensemble largest eigenvalue distribution. This law comes from the scaling limit of the rightmost particle in [3] with $\beta = 1, w(x) = \exp(-x^2), \Lambda = \mathbb{R}$, and $\mu$ the Lebesgue measure. In this case, the correlation functions are not determinants but rather pfaffians. The result for flat PNG follows from the Baik–Rains analysis of symmetrized last-passage or permutation problems. In the PNG model we can also consider an interface in equilibrium. This can be put into the last-passage percolation picture by suitable boundary conditions, different parameters for $w(i, j)$ when $i$ or $j$ equals 1 or extra Poisson points on the axes in the Poisson limit. Results by Baik and Rains show that in the cont-PNG in equilibrium the height fluctuations are given by a relative of the Tracy–Widom distribution, $F_0$. In these last two cases, the scaling limit of the whole height profile is not known.

The types of results discussed above can only be obtained for very special models. However, it is expected that many of the results (in particular the KPZ exponents 1/3 and 2/3, and also the fluctuation laws, including the Airy process) should generalize to many other models. The different interpretations of [6] mentioned above suggest different generalizations, various local growth models, directed polymers, asymmetric exclusion processes, and dimer/tiling problems. RMT laws are natural limit laws for which the domain of attraction is not understood.

*See also:* Combinatorics: Overview; Determinantal Random Fields; Dimer Problems; Integrable Systems in Random Matrix Theory; Random Walks in Random Environments; Random Matrix Theory in Physics; Random Partitions.

## Further Reading

Baik J and Rains E (2001) Symmetrized random permutations. In: Ruck M (ed.) *Random Matrix Models and Their Applications*, Math. Sci. Res. Publ., vol. 4, pp. 1–19. Cambridge: Cambridge University Press. (arXiv/math.CO/9910019).

Baik J, Deift PA, and Johansson K (1999) On the distribution of the length of the longest increasing subsequence in a random permutation. *Journal of American Mathematical Society* 12: 1119–1178.

Borodin A and Olshanski G (2001) In: *z-Measures on Partitions, Robinson–Schensted–Knuth Correspondence, and $\beta = 2$ Random Matrix Ensembles*, Math. sci. Res. Publ., vol. 4, pp. 71–94. Cambridge: Cambridge University Press. (arXiv/math.CO/9905189).

Ferrari PL, Prähofer M, and Spohn H (2003) Stochastic growth in one dimension and Gaussian multi-matrix models. *Proceedings of the International Congress of Mathematical Physics, Lissabon 2003* (arXiv/mathph/0310053).

Johansson K (2002) Toeplitz determinants, random growth and determinantal processes. *Proceedings of the ICM Beijing 2002*, vol. 3, pp. 53–62 (arXiv/math.PR/0304368).

Krug J and Spohn H (1992) Kinetic roughening of growing interfaces. In: Godrèche C (ed.) *Solids Far from Equilibrium: Growth, Morphology and Defects*, pp. 479–582. Cambridge: Cambridge University Press.

Mehta ML (1991) *Random Matrices*, 2nd edn. San Diego: Academic Press.

Okounkov A (2003) The uses of random partitions, arXiv/math-ph/0309015 *Proceedings of the International Congress of Mathematical Physics, Lissabon 2003* (arXiv/mathph/0309015).

Stanley RP (1999) *Enumerative Combinatorics*, vol. 2. Cambridge: Cambridge University Press.

Tracy C and Widom H (2002) Distribution functions for largest eigenvalues and their applications. *Proceedings of the ICM Beijing 2002*, vol. 1, pp. 587–596 (arXiv/math.PR/0210034).

# H

## Hamiltonian Fluid Dynamics

**P J Morrison**, University of Texas at Austin, Austin, TX, USA



### Introduction

The ideal fluid description is one in which viscosity or other phenomenological terms are neglected. Thus, as is the case for systems governed by Newton's second law without dissipation, such fluid descriptions possess Lagrangian and Hamiltonian descriptions. In fact, in the eighteenth century, Lagrange himself discussed what is in essence the action principle for the incompressible fluid. The subsequent history of action functional and Hamiltonian formulations of the ideal fluid is long and convoluted with contributions from Clebsch in the nineteenth century, and the likes of L Landau and V Arnol'd in the mid-twentieth century. In the early 1980s, there was a flurry of activity on the noncanonical Poisson bracket formulation, and this formulation is the focus of the present treatment, which is motivated by the work of the author, D Holm, J Marsden, T Ratiu, A Weinstein, and others.

### Noncanonical Hamiltonian Structure

The traditional arena for Hamiltonian dynamics is the cotangent bundle $\mathcal{M} := T^*\mathcal{Q}$, the phase space, which is naturally a symplectic manifold with a closed nondegenerate 2-form. In coordinates, the 2-form is given by $\omega_c = dq \wedge dp$, where $q$ denotes the configuration coordinate for the base space manifold $\mathcal{Q}$ and $p$ denotes the corresponding canonical momenta that arise from Legendre (convex) transformation. The 2-form $\omega_c$ provides a natural identification at a point $z = (q, p) \in \mathcal{M}$ of $T_z\mathcal{M}$ with $T_z^*\mathcal{M}$, and because of nondegeneracy its inverse, the cosymplectic form, provides the map $J_c : T_z^*\mathcal{M} \to T_z\mathcal{M}$. Thus, for a Hamiltonian $H : \mathcal{M} \to \mathbb{R}$ we have the Hamiltonian system of ordinary differential equations $\dot{z} = J_c\, dH$, which in canonical coordinates has the familiar form

$$\dot{q}^i = \partial H/\partial p_i, \qquad \dot{p}_i = -\partial H/\partial q^i \qquad [1]$$

with $i = 1, 2, \ldots, N$, where $N$ is the number of degrees of freedom.

Hamilton's equations can also be written in terms of the Poisson bracket $[f, g] := \omega_c(J_c\, df, J_c\, dg)$, where $f, g : \mathcal{M} \to \mathbb{R}$ are smooth phase-space functions. In terms of $z = (q, p)$, Hamilton's equations are

$$\dot{z}^\alpha = J_c^{\alpha\beta} \frac{\partial H}{\partial z^\beta} = [z^\alpha, H] \qquad [2]$$

where the Poisson bracket is

$$[f, g] = \frac{\partial f}{\partial z^\alpha} J_c^{\alpha\beta} \frac{\partial g}{\partial z^\beta} \qquad [3]$$

with

$$(J_c^{\alpha\beta}) = \begin{pmatrix} 0_N & I_N \\ -I_N & 0_N \end{pmatrix} \qquad [4]$$

Note, repeated indices are to be summed with $\alpha, \beta = 1, 2, \ldots, 2N$. In [4], $0_N$ is an $N \times N$ matrix of zeros and $I_N$ is the $N \times N$ unit matrix.

#### Noncanonical Poisson Brackets

The canonical Poisson bracket description of [2]–[4] suggests a generalization, with antecedents to S Lie and others, that was termed noncanonical Hamiltonian form in the fluid mechanics context by P Morrison and J Greene (1980):

*A system has noncanonical Hamiltonian form if it can be written as $\dot{z} = [z, H]$, where the noncanonical Poisson bracket $[\,,]$ is a Lie product for a realization of a Lie enveloping algebra on phase-space functions.*

Recall a Lie enveloping algebra $\mathfrak{a}$ is a Lie algebra, with the usual product $[\,,]$ that is bilinear, antisymmetric, and satisfies the Jacobi identity, which in addition has a product $\mathfrak{a} \times \mathfrak{a} \to \mathfrak{a}$ that satisfies the Leibniz identity $[fg, h] = f[g, h] + [f, h]g$ for all $f, g, h \in \mathfrak{a}$.

The geometric description of noncanonical Hamiltonian form has evolved into a structure called the Poisson manifold, a differential manifold $\mathcal{Z}$ endowed with the binary bracket operation $[\,,]$ defined on smooth functions, say, $f, g : \mathcal{Z} \to \mathbb{R}$. Poisson manifolds differ from symplectic manifolds

because the nondegeneracy condition is removed. In coordinates, [ , ] is given by

$$[f, g] = \frac{\partial f}{\partial z^\alpha} J^{\alpha\beta} \frac{\partial g}{\partial z^\beta}, \quad \alpha, \beta = 1, 2, \ldots, M \qquad [5]$$

where $M = \dim \mathcal{Z}$. Note that $J$ need not have the form of [3], may depend upon the coordinate $z$, and may have vanishing determinant. Bilinearity, $[f, g] = -[g, f]$ for all $f, g$, and the Jacobi identity, $[f, [g, h]] + [g, [h, f]] + [h, [f, g]] \equiv 0$, for all $f, g, h$, imply that the cosymplectic matrix satisfies $J^{\alpha\beta} = -J^{\beta\alpha}$ and

$$J^{\alpha\delta} \frac{\partial J^{\beta\gamma}}{\partial z^\delta} + J^{\beta\delta} \frac{\partial J^{\gamma\alpha}}{\partial z^\delta} + J^{\gamma\delta} \frac{\partial J^{\alpha\beta}}{\partial z^\delta} \equiv 0 \qquad [6]$$

respectively, for $\alpha, \beta, \gamma, \delta = 1, 2, \ldots, M$.

The local structure of $\mathcal{Z}$ is elucidated by the Darboux–Lie theorem, which states that in a neighborhood of a point $z \in \mathcal{Z}$, for which rank $J = M$, there exist coordinates in which $J$ has the following form:

$$(J) = \begin{pmatrix} 0_N & I_N & 0 \\ -I_N & 0_N & 0 \\ 0 & 0 & 0_{M-2N} \end{pmatrix} \qquad [7]$$

From [7] it is clear that in the right coordinates, the system looks like a canonical $N$-degree-of-freedom Hamiltonian system with some extraneous coordinates, $M - 2N$ in fact. Through any point of the $M$-dimensional phase space $\mathcal{Z}$, there exists a local foliation by symplectic leaves of dimension $2N$.

A consequence of the degeneracy is that there exists a special class of invariants called Casimir invariants that is built into the phase space. Since the rank of $J$ is $2N$, there exist possibly $M - 2N$ independent null eigenvectors. A consequence of the Darboux–Lie theorem is that the independent null eigenvectors exist and, moreover, the null space can in fact be spanned by the gradients of the Casimir invariants, which satisfy $J^{\alpha\beta} \partial C^{(a)} / \partial z^\beta = 0$, where $a = 1, 2, 3, \ldots, M - 2N$. That the Casimir invariants are constants of motion follows from

$$\dot{C}^{(a)} = \frac{\partial C^{(a)}}{\partial z^\alpha} J^{\alpha\beta} \frac{\partial H}{\partial z^\beta} = 0 \qquad [8]$$

Thus, Casimir invariants are constants of motion for any Hamiltonian. The symplectic leaves of dimension $2N$ are the intersections of the $M - 2N$ surfaces defined by $C^{(a)} = \text{constant}$. Dynamics generated by any $H$ that begins on a particular symplectic leaf remains there. The structure of Poisson manifolds has now been widely studied, but we will not pursue this further here.

Let us turn to infinite-dimensional systems, field theories such as those that govern ideal fluids, where the governing equations are partial differential equations. Although the level of rigor does not match that achieved for the finite systems described above, formally one can parody most of the steps and, consequently, the finite theory provides cogent imagery and serves as a beacon for shedding light. In infinite dimensions, an analog of [5] is given by

$$\{F, G\} = \int_\Omega \mathrm{d}\mu \, \frac{\delta F}{\delta \psi^i} \mathcal{J}^{ij} \frac{\delta G}{\delta \psi^j} =: \left\langle \frac{\delta F}{\delta \psi}, \mathcal{J} \frac{\delta G}{\delta \psi} \right\rangle \qquad [9]$$

where $F$ and $G$ are functionals of the functions $\psi^i(\mu, t)$, which are functions of $\mu = (\mu_1, \ldots, \mu_n)$, independent variables of some kind, $\delta F / \delta \psi^i$ denotes the functional (variational) derivative, and $\langle \, , \rangle$ is a pairing between a vector (function) space and its dual. The $\psi^i, i = 1, \ldots, n$, are $n$ field components, and now $\mathcal{J}$ is a cosymplectic operator. To be noncanonically Hamiltonian requires antisymmetry, $\{F, G\} = -\{G, F\}$, and the Jacobi identity, $\{F, \{G, H\}\} + \{G, \{H, F\}\} + \{H, \{F, G\}\} \equiv 0$, for all functionals $F$, $G$, and $H$. Antisymmetry requires $\mathcal{J}$ to be skew-symmetric, that is, $\langle f, \mathcal{J} g \rangle = \langle \mathcal{J}^\dagger f, g \rangle = -\langle g, \mathcal{J} f \rangle$. The Jacobi identity for infinite-dimensional systems has a condition analogous to [6]; it can be shown that one need only consider variations of $\mathcal{J}$ when calculating, for example, $\{F, \{G, H\}\}$.

## Lie–Poisson Brackets

As noted in the Introduction, the usual variables of fluid mechanics are not a set of canonical variables, and, consequently, the Hamiltonian description in terms of these variables is noncanonical. There is a special general form that the Poisson bracket takes for equations that describe media in terms of Eulerian-like variables, the so-called Lie–Poisson brackets, a special form of noncanonical Poisson bracket. Lie–Poisson brackets describe essentially every fundamental equation that describes classical media. In addition to the equations for the ideal fluid, they describe Liouville's equation for the dynamics of the phase-space density of a collection of particles, the various hierarchy of kinetic theory, the Vlasov equation of plasma physics, and various approximations thereof, and magnetized and other more complicated fluids.

Both finite- and infinite-dimensional Lie–Poisson brackets are intimately associated with a Lie group $\mathfrak{G}$. We use the pairing between a vector space and its dual, $\langle \, , \rangle$, where the second slot is reserved for elements of the Lie algebra $\mathfrak{g}$ of $\mathfrak{G}$ and the first slot for elements of its dual $\mathfrak{g}^*$. Thus, $\langle \, , \rangle : \mathfrak{g}^* \times \mathfrak{g} \to \mathbb{R}$. In terms of the pairing, noncanonical Lie–Poisson brackets have the following compact form:

$$\{F, G\} = \langle \chi, [F_\chi, G_\chi] \rangle \qquad [10]$$

where we suppose the dynamical variable $\chi \in \mathfrak{g}^*$, $[\,,\,]$ is the Lie algebra product, which takes $\mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$, and we have introduced the shorthand $F_\chi := \delta F/\delta\chi$. The quantities $F_\chi$ and $G_\chi$ are, of course, in $\mathfrak{g}$. We refer to $\{\,,\,\}$ as the "outer" bracket of the realization enveloping algebra and $[\,,\,]$ as the "inner" bracket of the Lie algebra $\mathfrak{g}$. The binary operator $[\,,\,]^\dagger$ is defined as follows:

$$\langle \chi, [f, g] \rangle =: \left\langle [\chi, g]^\dagger, f \right\rangle \qquad [11]$$

where evidently $\chi \in \mathfrak{g}^*, g, f \in \mathfrak{g}$, and $[\,,\,]^\dagger : \mathfrak{g}^* \times \mathfrak{g} \to \mathfrak{g}^*$. The operator $[\,,\,]^\dagger$, which defines the coadjoint orbit, is necessary for obtaining the equations of motion from a Lie–Poisson bracket.

For finite-dimensional systems, the group $\mathfrak{G}$ must be a finite-parameter Lie group, the variable $\psi$ corresponds to $w$, and the cosymplectic form in coordinates is given by $J_{ab} = c^c_{ab} w_c$, where the $c^c_{ab}$ are the structure constants for the Lie algebra $\mathfrak{g}$, which satisfy

$$\begin{aligned} c^c_{ab} &= -c^c_{ba} \\ c^e_{ab} c^d_{ec} + c^e_{bc} c^d_{ea} &+ c^e_{ca} c^d_{eb} = 0 \end{aligned} \qquad [12]$$

relations that imply [10] satisfies the antisymmetry condition and the Jacobi identity.

For infinite-dimensional systems, the group $\mathfrak{G}$ must be an infinite-parameter Lie group and the cosymplectic operator has the form $\mathcal{J}_{ij} = \mathcal{C}^k_{ij} \chi_k$, where $\mathcal{C}^k_{ij}$ are structure operators. The meaning of these structure operators will be clarified when we consider brackets for fluid mechanics.

## The Fluid State

Fluid mechanics has a long history, and thus it comes as no surprise that the fluid state has been described in many ways. Because the Hamiltonian structure depends on the state variables, some of these ways are described below, beginning with Lagrangian variable description.

### Lagrangian Variables

The description of a fluid that is most like that of particle mechanics occurs in terms of variables usually referred to as Lagrangian variables. This description dates to the eighteenth century. The idea behind the use of these variables is a simple one: if a fluid is described as a continuum collection of fluid particles, also called fluid parcels or elements, then its motion is governed by an equation that is an infinite-dimensional version of Newton's second law and, consequently, as we will see, both the Hamiltonian and the Lagrangian descriptions are infinite degree-of-freedom generalizations of those of ordinary particle mechanics.

The position of a fluid element, referred to a fixed rectangular coordinate systems, is given by $q = q(a, t)$, where $q = (q_1, q_2, q_3)$ and $a = (a_1, a_2, a_3)$ is a continuum label that replaces the index $i$ of [1]. In practice, the label can be any quantity that identifies a fluid particle, but it is often taken to be the position of the fluid particle at time $t = 0$ in rectangular coordinates. The quantities $q^i(a, t)$ are coordinates for the configuration space $\mathcal{Q}$, which is in fact a function space because in addition to the three indices "$i$" there is the continuum label $a$. We assume that $a$ varies over a fixed domain, $\Omega \subset \mathbb{R}^3$, which is completely filled with fluid, and that the function $q : \Omega \to \Omega$ is one-to-one and onto. We will assume that as many derivatives of $q$ with respect to $a$ as needed exist, but we will not say more about $\mathcal{Q}$; in fact, not much is known about the solution function space for the 3D fluid equations in Lagrangian variables. Often in the Hamiltonian context the functions $q = q(a, t)$ are assumed to be diffeomorphisms and their collection is referred to as the diffeomorphism group.

In the sequel several manipulations are needed and so we record here some identities for later use. Viewing the map $a \mapsto q$ at fixed $t$ as a coordinate change, the Jacobian matrix $\partial q^k/\partial a^i =: q^k_{,i}$ has an inverse given by

$$\frac{\partial q^k}{\partial a^j} \frac{A^i_k}{\mathfrak{J}} = \delta^i_j \qquad [13]$$

where $A^i_k$ is the cofactor of $q^k_{,i}$ and $\mathfrak{J}$ is its determinant. A convenient expression for $A^i_k$ is given by

$$A^i_k = \frac{1}{2} \epsilon_{kjl} \epsilon^{imn} \frac{\partial q^j}{\partial a^m} \frac{\partial q^l}{\partial a^n} \qquad [14]$$

where $\epsilon_{ijk} (= \epsilon^{ijk})$ is the skew-symmetric tensor (density). Evidently, $\partial\mathfrak{J}/\partial q^i_{,j} = A^j_i$ follows from [13].

### Eulerian Variables

In the Lagrangian variable description, one picks out a particular particle, labeled by $a$, and keeps track in time $t$ of where it goes. However, in the Eulerian variable description, one stays at a spatial observation point $r = (x_1, x_2, x_3) \in \Omega$ and monitors the nature of the fluid at $r$ at time $t$.

The most important Eulerian variable is the Eulerian velocity field $v(r, t)$. This quantity is the velocity of the particular fluid element that is located at the spatial point $r$ at time $t$. The label of that particular fluid element is given by $a = q^{-1}(r, t)$, and so

$$v(r, t) = \dot{q}(a, t)|_{a = q^{-1}(r,t)} := \dot{q} \circ q^{-1}(r, t) \qquad [15]$$

where $\cdot$ denotes differentiation with respect to time at fixed label $a$. Attached to a fluid element is a certain amount of mass described by a density

function $\rho_0(a)$. As the fluid moves so that $a \mapsto q$, the volume of an infinitesimal region will change, but its mass must remain fixed. The statement of local mass conservation is $\rho d^3r = \rho_0 d^3a$, where $d^3a$ is an initial infinitesimal volume element that maps to $d^3q$ at time $t$, and $d^3r = \mathfrak{J}\, d^3a$. (When integrating over $\Omega$ we will replace $d^3q$ by $d^3r$.) Thus, we obtain

$$\rho(r,t) = \left.\frac{\rho_0(a)}{\mathfrak{J}(a,t)}\right|_{a=q^{-1}(r,t)} = \frac{\rho_0}{\mathfrak{J}} \circ q^{-1}(r,t) \qquad [16]$$

where recall the Jacobian $\mathfrak{J} = \det(q^i{}_{,j})$. Besides the density, for the ideal fluid, one attaches an entropy per unit mass, $s = s_0(a)$, to a fluid element, and this quantity remains fixed in time. In the Eulerian description this gives rise to the entropy field

$$s(r,t) = s_0(a)|_{a=q^{-1}(r,t)} = s_0 \circ q^{-1}(r,t) \qquad [17]$$

One could attach other scalar, vector, etc., quantities to the fluid element, but we will not pursue this. In the usual ideal fluid closure only the above variables are considered.

Equations [15]–[17] express the Euler–Lagrange map. There is a natural representation of this map in terms of the Eulerian density variables, $M := \rho v, \rho$, and $\sigma := \rho s$, the momentum, mass, and entropy densities, respectively, which, as will be seen, are variables in which the noncanonical Poisson bracket has Lie–Poisson from.

### Other Variables

Fluid mechanics is rife with variables that have been used for its description. For example, Euler, Monge, Clebsch, and others introduced potential representations, of varying generality, for the Eulerian velocity field, an example being

$$v(r,t) = \alpha \nabla \beta + \nabla \phi \qquad [18]$$

where the three components of $v$ are replaced by the functions $\alpha, \beta$, and $\phi$, all of which depend on $(r,t)$.

Often reduced variables that are tailored to specific ideal flows with less generality than those described by $\rho, s$, and $v$ are considered. Examples include incompressible flow with $\nabla \cdot v = 0$, vortex dynamics, including contour dynamics and point vortex dynamics, flow governed by the shallow-water equations, quasigeostrophy, etc. The Hamiltonian structure in terms of these reduced variables derives from that of the parent model in terms of Lagrangian variables. Specific variables may embody constraints, and understanding these constraints, although tractable, can be a cause of confusion. Pursuing this further is beyond the scope here.

## Hamilton's Principle for Fluid

Lagrange, in his famous work of 1788, Mécanique Analytique, produced in essence a variational principle for incompressible fluid flow in terms of Lagrangian variables. The generalization to compressible flow awaited the discovery of thermodynamics, and that is what we describe here. In traditional mechanics nomenclature, this variational principle is an infinite-dimensional generalization of what is known variously as the action principle, the principle of least action, or Hamilton's principle, whereby one constructs, on physical grounds, a Lagrangian function on $T\mathcal{Q}$ used in the action principle, where $\mathcal{Q}$ is the function space of the $q(a,t)$.

Construction of the Lagrangian requires identification of the potential energy, and this requires thermodynamics, because potential energy is stored in terms of pressure and temperature. A basic assumption of the fluid approximation is that of local thermodynamic equilibrium. In the energy representation of thermodynamics, the extensive energy is treated as a function of the entropy and the volume. For a fluid, it is convenient to consider the energy per unit mass, denoted by $U$, to be a function of the entropy per unit mass, $s$, and the mass density, $\rho$, a measure of the volume. The intensive quantities, pressure and temperature, are given by $T = \partial U/\partial s$ and $p = \rho^2 \partial U/\partial \rho$. Choices for $U$ produce equations of state. For barotropic or isentropic flow, $U$ depends only on $\rho$. For an ideal monatomic gas $U(\rho,s) = c\rho^{\gamma-1}\exp(\alpha s)$, where $c, \gamma$, and $\alpha$ are constants. The function $U$ could also depend on additional scalar quantities, such as a quantity known as spice that has been considered in oceanography.

Conventional thermodynamic variables can be viewed as Eulerian variables with a static velocity field. Thus, we write $U(\rho,s)$, where $\rho$ and $s$ are spatially independent or, if the system has only locally relaxed, these variables can be functions of $r$. For the ideal fluid, each fluid element can be viewed as a self-contained isentropic thermodynamic system that moves with the fluid. Thus, the total fluid potential energy functional is given by $V[q] = \int_\Omega d^3a \rho_0 U (s_0, \rho_0/\mathfrak{J})$, which is a functional of $q$ that depends only upon $\mathfrak{J}$ and hence only upon $\partial q/\partial a$.

The next item required for constructing Hamilton's principle is the kinetic energy functional, which is given by $T[q,\dot{q}] = \int_\Omega d^3a \rho_0 \dot{q}^2/2$, where $\dot{q}^2 := \eta_{ij}\dot{q}^i\dot{q}^j$, with the Cartesian metric $\eta_{ij} := \delta_{ij}$. This metric and its inverse can be used to raise and lower indices.

The Lagrangian functional is $L[q,\dot{q}] := T - V$, where $L[q,\dot{q}] = \int_\Omega d^3a \mathcal{L}(q,\dot{q},\partial q/\partial a)$ and $\mathcal{L}$ is the

Lagrangian density, in terms of which the action functional of Hamilton's principle is given by

$$S[q] = \int_{t_0}^{t_1} \mathrm{d}t L[q, \dot{q}] = \int_{t_0}^{t_1} \mathrm{d}t \int_{\Omega} \mathrm{d}^3 a \left[ \frac{1}{2} \rho_0 \dot{q}^2 - \rho_0 U \right] \qquad [19]$$

The end conditions for Hamilton's principle for the fluid are the same as those of mechanics, that is, $\delta q(a, t_0) = \delta q(a, t_1) = 0$. The nonpenetration condition, $\delta q \cdot \hat{n} = 0$ on $\partial\Omega$, where $\hat{n}$ is a unit normal vector is also assumed. Other boundary conditions, such as periodic and free boundary conditions, are also possibilities. Hamilton's principle amounts to $\delta S/\delta q(a, t) = 0$, which, with the end and boundary conditions, implies the following equations of motion:

$$\rho_0 \ddot{q}_i + A_i^j \frac{\partial}{\partial a^j} \left( \frac{\rho_0^2}{\mathfrak{J}^2} \frac{\partial U}{\partial \rho} \right) = 0 \qquad [20]$$

Here we have used $\partial A_i^j / \partial a^j = 0$, which can be seen using [14]. Equation [20] amounts to Newton's second law for the ideal fluid, which is made clearer by using the following useful identity:

$$\frac{\partial}{\partial q^k} = \frac{1}{\mathfrak{J}} A_k^i \frac{\partial}{\partial a^i} \qquad [21]$$

Alternatively, upon using [13], [20] is sometimes written in the form

$$\rho_0 \ddot{q}_j \frac{\partial q^j}{\partial a^i} + \mathfrak{J} \frac{\partial}{\partial a^i} \left( \frac{\rho_0^2}{\mathfrak{J}^2} \frac{\partial U}{\partial \rho} \right) = 0 \qquad [22]$$

The Eulerian variable force law follows from [20] upon using [21]:

$$\rho \left( \frac{\partial v}{\partial t} + v \cdot \nabla v \right) = -\nabla p \qquad [23]$$

where $v = v(r, t)$. The remaining Eulerian equations of mass conservation and entropy advection follow from the constraints that $s_0$ and $\rho_0$ are constant on fluid elements. Time differentiation and the transformations of [16] and [17] yield

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0 \qquad [24]$$

$$\frac{\partial s}{\partial t} + v \cdot \nabla s = 0 \qquad [25]$$

Equations [23]–[25] together with a given function $U(\rho, s)$ and the relation $p = \rho^2 \partial U/\partial \rho$ constitute the Eulerian description.

Variational principles similar to that described above exist for essentially all ideal fluid models, including incompressible flow, magnetohydrodynamics, the two-fluid equations of plasma physics, etc.

## Eulerian Action Principles

Some early researchers sought variational principles that directly produce the ideal fluid equations in Eulerian form. Because the Eulerian form of the equations does not treat the fluid as a collection of particles, the resulting action principles possess a certain awkwardness. Below, we describe three approaches to such action principles.

**Clebsch action** The action principle for electromagnetism proceeds by introducing the 4-vector potential. In a similar way, the Clebsch action principle anticipates this idea by using a potential representation of the velocity field, an example being that of [18].

Although compressible flow with an arbitrary equation of state can be treated in full generality, for simplicity and variety we will restrict to incompressible flow and set $\nabla \cdot v \equiv 0$. This constraint is enforced by requiring $\phi$ to be dependent on $\alpha$ and $\beta$ according to $\phi[\alpha, \beta] := -\Delta^{-1}(\alpha \nabla \beta)$, where $\Delta^{-1}$ is the inverse Laplacian. The Clebsch action is then written as follows:

$$S_{\mathrm{C}}[\alpha, \beta] := \int_{t_0}^{t_1} \mathrm{d}t \int_{\Omega} \mathrm{d}^3 r \left[ \beta \alpha_t - \frac{1}{2} v^2 \right] \qquad [26]$$

where the subscript $t$ denotes differentiation at fixed $r$, we have set $\rho \equiv 1$, and $v$ is a shorthand for the expression of [18] with $\phi = \phi[\alpha, \beta]$. The form of $S_{\mathrm{C}}$ is that of the phase-space action that produces Hamilton's equations upon independent variation of the configuration space coordinate and its conjugate momentum, which are here $\alpha$ and $\beta$, respectively. Thus, we require $\delta\alpha(r, t_0) = \delta\alpha(r, t_1) = 0$, but no condition is needed for $\delta\beta$ at $t_{0,1}$. We also require $\hat{n} \cdot v = 0$ on $\partial\Omega$. The variations $\delta S_{\mathrm{C}}/\delta\beta = 0$ and $\delta S_{\mathrm{C}}/\delta\alpha = 0$ imply

$$\alpha_t = \frac{\delta H}{\delta \beta} = -v \cdot \nabla \alpha$$
$$\beta_t = -\frac{\delta H}{\delta \alpha} = -v \cdot \nabla \beta = 0 \qquad [27]$$

an infinite-dimensional version of [1] with $H := \int_{\Omega} \mathrm{d}^3 r\, v^2/2$. Evidently, both $\alpha$ and $\beta$ are advected by the flow.

Because the vorticity, $\zeta := \nabla \times v = \nabla\alpha \times \nabla\beta$, knowledge of $\alpha$ and $\beta$ determines $\zeta$ and one can invert the curl operator to obtain $v$ in the usual way. The intersection of level sets of $\alpha$ and $\beta$ define vortex lines, and, evidently, these quantities, like the entropy for compressible dynamics, are constant on fluid elements. It is not difficult to show that the advection of $\alpha$ and $\beta$ implies the correct dynamical equation for incompressible $v$.

**Herivel–Lin action** The Herivel–Lin action incorporates [24] and [25] as constraints with Lagrange

multipliers, $\varphi$ and $\rho\beta$. (Here $\beta$ is not the Clebsch $\beta$ and the factor of $\rho$ is included for convenience.) It was discovered early on that these constraints were not enough to achieve complete generality and so a new one, known as the Lin constraint, was added. The Lin constraint corresponds to constancy of the fluid particle label. One defines an Eulerian label field by setting $q(a,t) = r$ and solving for the label $a = q^{-1}(r,t) =: a(r,t)$. Conservation of particle identity is thus given by $a_t + v \cdot \nabla a = 0$, and this constraint is associated with a Lagrange multiplier $\gamma = (\gamma_1, \gamma_2, \gamma_3)$. The Herivel–Lin action is thus given by

$$S_{\mathrm{HL}}[v,\rho,s,a;\varphi,\beta,\gamma] = \int_{t_0}^{t_1} \mathrm{d}t \int_\Omega \mathrm{d}^3 r \bigg( \frac{1}{2}\rho v^2 - \rho U(\rho,s) + \varphi[\rho_t + \nabla\cdot(\rho v)] - \rho\beta[s_t + v\cdot\nabla s] - \rho\gamma\cdot[a_t + v\cdot\nabla a] \bigg) \quad [28]$$

Variation of [28] with respect to the Lagrange multipliers just reproduces the constraints; however, variation with respect to $v$, $\rho$, $s$, and $a$ produces equations that imply [23]. Moreover, every flow can be shown to be an extremal of $S_{\mathrm{HL}}$.

**Euler–Poincaré–Hamel action** Another approach is to use directly constrained variations. The essential idea is to only consider Eulerian variable variations that are induced by underlying Lagrangian variable variations $\delta q$, the so-called dynamically accessible variations. Explicitly, a basic Eulerian variation $\eta = (\eta_1, \eta_2, \eta_3)$ is given by $\eta(r,t) = \delta q(a,t)|_{a=q^{-1}(r,t)}$. In terms of this quantity, the dynamically accessible variations of the Eulerian velocity field, density, and entropy are given, respectively, by $\delta v = \eta_t + v\cdot\nabla\eta - \eta\cdot\nabla v$, $\delta\rho = -\nabla\cdot(\rho\eta)$, and $\delta s = -\eta\cdot\nabla s$. Upon inserting them into the variation of

$$S_{\mathrm{EPH}}[\eta] = \int_{t_0}^{t_1} \mathrm{d}t \int_\Omega \mathrm{d}^3 r \left[\frac{1}{2}\rho v^2 - \rho U(\rho,s)\right] \quad [29]$$

and integrating by parts gives

$$\delta S_{\mathrm{EPH}} = \int_{t_0}^{t_1} \mathrm{d}t \int_\Omega \mathrm{d}^3 r[\ldots]\cdot\eta = 0$$

where $[\ldots]$ is equivalent to [23]. Thus, assuming $\eta$ is arbitrary, we obtain directly the equation of motion.

There is a version of this kind of constrained variational principle for all ideal fluid and plasma equations. Also, it possesses a geometric interpretation. In a more practical vein, constrained variations can be used to derive reduced models, and dynamically accessible variations can also be used for stability calculations. Exploring these ideas is outside the present scope.

## Fluid Hamiltonian Description

Having described variational principles, we turn to the associated canonical and noncanonical Hamiltonian descriptions.

### Canonical Description

Because the action of [19] is of standard form, it is convex in $\dot{q}$ and the Legendre transform follows easily: the canonical momentum density is $\pi_i(a,t) := \delta L/\delta\dot{q}^i(a) = \rho_0\dot{q}_i$ and $H[q,\pi] = \int_\Omega \mathrm{d}^3 a[\pi\cdot\dot{q} - \mathcal{L}] = \int_\Omega \mathrm{d}^3 a[\pi^2/(2\rho_0) + \rho_0 U]$. Hamilton's equations are then

$$\dot{q}^i = \frac{\delta H}{\delta \pi_i} = \{q^i, H\}, \qquad \dot{\pi}_i = -\frac{\delta H}{\delta q^i} = \{\pi_i, H\} \quad [30]$$

an infinite-dimensional version of [1], with the canonical Poisson bracket

$$\{F,G\} = \int_\Omega \left[\frac{\delta F}{\delta q}\cdot\frac{\delta G}{\delta \pi} - \frac{\delta G}{\delta q}\cdot\frac{\delta F}{\delta \pi}\right]\mathrm{d}^3 a \quad [31]$$

(Note, $\delta q^i(a)/\delta q^j(a') = \delta^i_j\delta(a-a')$, a relation analogous to $\partial q^i/\partial q^i = \delta^j_i$ for finite systems.)

### Reduction to Noncanonical Poisson Brackets

Reduction is a procedure for reducing the size of a Hamiltonian system. Given constants of motion in involution, that is, with pairwise vanishing Poisson brackets, the dimension of a Hamiltonian system can be reduced by 2 for each such constant of motion. However, when constants do not commute, the situation is more complicated and one must invoke a theory due to Lie, Poincaré, Cartan, and others. Associated with invariants are symmetries, and so a complete discussion of this theory requires examination of symmetry groups and associated geometry. For the ideal fluid, the map from the Lagrangian to the Eulerian descriptions is an example of reduction, whereby the Poisson bracket of [31] is mapped into a noncanonical Poisson bracket. En route to describing this example, a brief discussion of reduction of finite systems is considered first.

### Reduction of Finite-Dimensional Systems

Consider a canonical system with the phase space $\mathcal{M}$, a $2N$-dimensional symplectic manifold. In a coordinate patch with coordinates $z = (q,p)$ the system has the canonical description of [2]–[4]. Suppose we have a map $P: \mathcal{M} \to \mathfrak{m}^*$, where $\mathfrak{m}^*$ is some $M < 2N$-dimensional space described by coordinates $w = (w_1, w_2, \ldots, w_M)$. In coordinates, this map is represented in terms of functions $w_a = w_a(z)$,

with $a = 1, 2, \ldots, M$, which, because $M < 2N$, is always noninvertible. Suppose $f, g : \mathcal{M} \to \mathbb{R}$ obtain their $z$-dependence through the functions $w$, that is, $f(z) = \bar{f}(w(z)) = \bar{f} \circ w$. Making use of the chain rule yields

$$[f, g] = \frac{\partial \bar{f}}{\partial w_a} J_{ab} \frac{\partial \bar{g}}{\partial w_b} \qquad [32]$$

where the quantity

$$J_{ab} := \frac{\partial w_a}{\partial z^\alpha} J_c^{\alpha\beta} \frac{\partial w_b}{\partial z^\beta} \qquad [33]$$

is in general a function of $z$. However, it is possible that $J_{ab}$ may only depend on $w$. When this happens, we have a reduction of the phase space $\mathcal{M}$.

If the original dynamics of interest has the Hamiltonian vector field generated by $H(z)$, and if it is possible that $H(z)$ can be expressed solely in terms of the $w$'s, that is, $H(z) = \bar{H}(w)$, then the system has been reduced. Clearly, this is a statement of symmetry, since the function $H(z)$ in reality depends on a fewer number of variables, the $w$'s.

A beautiful form of reduction occurs when the map $P$ has a special form $w_a = L_a^i(q) p_i$, where the quantity $L$ is associated with a symmetry group. An identity for what is required of $L_a^i$ in order for the transformed bracket to be expressible in terms of the $w$'s can be worked out, but this is explained in terms of Lie groups. If the space $\mathfrak{m}$ is a Lie algebra $\mathfrak{g}$, then the functions $\bar{f}, \bar{g}$ are real-valued functions on $\mathfrak{g}^*$ that can be extended by left or right translation to functions $\bar{f}, \bar{g}$ on $T^*\mathfrak{G}$. Thus, $f$ restricted to $T^*\mathfrak{G}$ at the identity, $T_e^*\mathfrak{G} = \mathfrak{g}^*$, is $\bar{f}$. Because $T^*\mathfrak{G}$ is a cotangent bundle, it carries the canonical Poisson bracket and we get a natural map $P$, called a momentum map, into the dual of a Lie algebra. This geometrical description of obtaining brackets on $\mathfrak{g}^*$ from brackets on $T^*\mathfrak{G}$ is a case of Marsden–Weinstein reduction. In the early 1980s, these authors and others developed the geometrical interpretation of the noncanonical Poisson brackets for the ideal fluid.

### Ideal Fluid Noncanonical Poisson Brackets

The Euler–Lagrange map of the fluid is of the form of the map $P$ above. It maps the canonical bracket of [31] into a noncanonical Poisson bracket. If we use the Eulerian variables $M := \rho v, \rho$, and $\sigma := \rho s$, then the resulting noncanonical bracket is of Lie–Poisson form. To effect this map, one must vary [15]–[17] to relate functional derivatives with respect to $q$ and $\pi$ to those with respect to $M, \rho$, and $\sigma$. This amounts to working out the chain rule for functionals. Upon doing this, one obtains the following noncanonical bracket:

$$\begin{aligned}\{F, G\} = -\int_\Omega \Bigg[ & M_i \left( \frac{\delta F}{\delta M_j} \frac{\partial}{\partial x^j} \frac{\delta G}{\delta M_i} - \frac{\delta G}{\delta M_j} \frac{\partial}{\partial x^j} \frac{\delta F}{\delta M_i} \right) \\ & + \rho \left( \frac{\delta F}{\delta M} \cdot \nabla \frac{\delta G}{\delta \rho} - \frac{\delta G}{\delta M} \cdot \nabla \frac{\delta F}{\delta \rho} \right) \\ & + \sigma \left( \frac{\delta F}{\delta M} \cdot \nabla \frac{\delta G}{\delta \sigma} - \frac{\delta G}{\delta M} \cdot \nabla \frac{\delta F}{\delta \sigma} \right) \Bigg] \mathrm{d}^3 r \end{aligned} \qquad [34]$$

This bracket, together with the Hamiltonian $\bar{H}[M, \rho, \sigma] = \int_\Omega \mathrm{d}^3 r [M^2/(2\rho) + \rho U(\rho, \sigma/\rho)]$ generates the ideal fluid equations. This Hamiltonian follows from $\bar{H}[M, \rho, \sigma] := H[q, \pi]$ with $H[q, \pi] = \int_\Omega \mathrm{d}^3 a\, [\pi^2/(2\rho_0) + \rho_0 U]$. The bracket of [34] is clearly seen to be linear in the variables $M, \rho$, and $\sigma$, and the form of the cosymplectic operator and structure operators $C_{ij}^k$ can be obtained by integration by parts. The Lie group in this case can be seen to be an extension by semidirect product of the diffeomorphism group.

An alternative form of the noncanonical Poisson bracket is given in terms of the variables $v, \rho$, and $s$. Upon changing to these coordinates, the noncanonical Poisson bracket transforms into

$$\begin{aligned}\{F, G\} = -\int_\Omega \Bigg[ & \left( \frac{\delta F}{\delta \rho} \nabla \cdot \frac{\delta G}{\delta v} - \frac{\delta G}{\delta \rho} \nabla \cdot \frac{\delta F}{\delta v} \right) \\ & + \left( \frac{\nabla \times v}{\rho} \cdot \frac{\delta G}{\delta v} \times \frac{\delta F}{\delta v} \right) \\ & + \frac{\nabla s}{\rho} \cdot \left( \frac{\delta F}{\delta s} \frac{\delta G}{\delta v} - \frac{\delta G}{\delta s} \frac{\delta F}{\delta v} \right) \Bigg] \mathrm{d}^3 r \end{aligned} \qquad [35]$$

which, with the Hamiltonian $H[v, \rho, s] = \int_\Omega \mathrm{d}^3 r\, [\rho v^2/2 + \rho U(\rho, s)]$, produces the Eulerian fluid equations of [23]–[25] directly as $v_t = \{v, H\}, \rho_t = \{\rho, H\}$, and $s_t = \{s, H\}$, respectively. Observe that in these variables, the bracket is no longer of Lie–Poisson form.

## Conclusion

In a general sense, Hamiltonian dynamics is about coordinate changes, and it is clear from the above that there is no shortage of coordinates for describing the ideal fluid. The most intuitive form of fluid equations (at present) is the Eulerian form, and this possesses a noncanonical Hamiltonian description. Other noncanonical variables are also used for both less and more general fluid systems than those described above. Vortex dynamics, shallow-water theory, and other equations of geophysical fluid dynamics are possibilities, as well as equations from plasma physics and other disciplines. The general story for these systems is much the same as above, although in some descriptions constraints are involved and they can complicate matters.

There are various motivations for pursuing an understanding of the Hamiltonian structure of fluids, but ultimately these motivations are the same as those for investigating the Hamiltonian dynamics of particle and other finite degree-of-freedom systems. Hamiltonian theory serves as an organizing framework, one that can be used for the derivation and approximation of systems. If one understands something about a particular Hamiltonian system, then often it can be said to be true of a general class of Hamiltonian systems. By now, many applications have been worked out, some of which can be accessed from the literature cited below.

*See also:* Adiabatic Piston; Adiabatic Piston; Bi-Hamiltonian Methods in Soliton Theory; Bi-Hamiltonian Methods in Soliton Theory; Classical *r*-Matrices, Lie Bialgebras, and Poisson Lie Groups; Contact Manifolds; Contact Manifolds; Hamiltonian Group Actions; Infinite-Dimensional Hamiltonian Systems; Korteweg–de Vries Equation and other Modulation Equations; Korteweg–de Vries Equation and Other Modulation Equations; Stochastic Hydrodynamics; Stochastic Hydrodynamics.

## Further Reading


Holm DD, Marsden JE, and Ratiu TS (1998) The Euler–Poincaré equations and semidirect products with applications to continuum theories. *Advance in Mathematics* 137: 1–81.

Kuznetzov EA and Zakharov VE (1997) Hamiltonian formalism for nonlinear waves. *Usp. Fiz. Nauk* 167: 1137–1168.

Morrison PJ (1982) Poisson brackets for fluids and plasmas. In: Tabor M and Treve Y (eds.) *Mathematical Methods in Hydrodynamics and Integrability in Dynamical Systems*, American Institute of Physics Conference Proceedings, No. 88. New York: American Institute of Physics.

Morrison PJ (1998) Hamiltonian description of the ideal fluid. *Reviews in Modern Physics* 70: 467–521.

Morrison PJ (2005) Hamiltonian and action principle formulations of plasma physics. *Physics of Plasmas* 12: 058102–1–13.

Morrison PJ and Greene JM (1980) Noncanonical Hamiltonian density formulation of hydrodynamics and ideal magnetohydrodynamics. *Physical Review Letters* 45: 790–793; (E) (1982) 48: 569.

Salmon R (1988) Hamiltonian fluid mechanics. *Annual Reviews of Fluid Mechanics* 20: 225–256.

Serrin J (1959) Mathematical principles of classical fluid mechanics. In: Flügge S (ed.) *Handbuch der Physik,* vol. VIII. Part 1, pp. 125–263. Berlin: Springer.

Shepherd TG (1990) Symmetries, conservation laws, and Hamiltonian structure in geophysical fluid dynamics. *Advances in Geophysics* 32: 287–338.


# Hamiltonian Group Actions

**L C Jeffrey**, University of Toronto, Toronto, ON, Canada



## Introduction

The idea of a Hamiltonian flow on a symplectic manifold has its roots in Hamilton's equations, which govern the trajectory of a particle in phase space (the space parametrizing coordinates and momenta of a classical particle). A fundamental idea in theoretical physics (Noether's theorem) is that to every symmetry in a physical system (such as a group action), there is an associated conserved quantity: invariance under translation corresponds to conservation of linear momentum, invariance under rotation corresponds to conservation of angular momentum and so on, and these momenta are functions on the phase space. The mathematical formulation of this idea is the idea of the moment map associated to a group action on a symplectic manifold; the group action is obtained from the Hamiltonian flow of the moment map.

This article will describe some basic features of moment maps associated to Hamiltonian group actions, and some recent results about the geometry and topology of symplectic manifolds which have such group actions. We first define Hamiltonian group actions and list some of their properties. Next we give the definition of the symplectic quotient, which is a means of dividing out the symmetry to form a new symplectic manifold. We also explain some properties of the quotient construction. The convexity theorem and the moment polytope are outlined and toric manifolds (a particular type of symplectic manifold with a Hamiltonian torus action of maximal dimension) are defined. Finally, we list some properties of cohomology rings of symplectic quotients.

Two standard references on this material are the books of Cannas da Silva (2001) and McDuff and Salamon (1995). An authoritative and comprehensive reference is the monograph by Guillemin, Ginzburg and Karshon (2002).

## Hamiltonian Group Actions

Let $(M, \omega)$ be a symplectic manifold. The Hamiltonian vector field $\xi_H$ generated by a function $H$ is defined by

$$\omega_m(\xi_H, Y) = \mathrm{d}H_m(Y)$$

for any $Y \in T_mM$. If $X \in \boldsymbol{g} \mapsto X^\#$ are the vector fields on $M$ generated by the symplectic action of a compact Lie group $G$ with Lie algebra $\boldsymbol{g}$, then the moment map $\mu : M \to \boldsymbol{g}^*$ is defined by two properties:

1. $d\mu_m(Y)(X) = \omega_m(X^\#, Y)$ for any $Y \in T_mM$: in other words the function $\mu_X : M \to \mathbb{R}$ defined by
$$\mu_X(m) \stackrel{\text{def}}{=} \mu(m)(X)$$
is the Hamiltonian function generating the vector field $X^\#$.
2. $\mu : M \to \boldsymbol{g}^*$ is equivariant (where $G$ acts on $\boldsymbol{g}^*$ by the coadjoint action).

**Remark 1** In this article, we shall only consider actions of compact connected Lie groups, although the definition of Hamiltonian group action may be extended to noncompact groups. In particular, unless otherwise specified the term "torus" refers to the compact torus $T \cong \mathrm{U}(1)^n$.

**Remark 2** (Existence and uniqueness of moment maps). One sees that $\mathcal{L}_{X^\#}\omega = d(\iota_{X^\#}\omega)$, so that $\iota_{X^\#}\omega$ is closed. The moment map $\mu_X$ exists if and only if $\iota_{X^\#}\omega$ is also "exact." The moment map need not always exist: for example, if $S^1$ acts on $T^2$ by
$$e^{iX} : (e^{i\theta_1}, e^{i\theta_2}) \mapsto (e^{i(\theta_1+X)}, e^{i\theta_2})$$
we see that for the standard symplectic form $\omega = d\theta_1 \wedge d\theta_2$ we have $\iota_{X^\#}\omega = d\theta_2$. Since $\theta_2$ is only defined mod $2\pi$ we see that the moment map does not exist as a map into $\mathbb{R}$. Conditions guaranteeing the existence of a moment map (other than $M$ being simply connected) include the hypothesis that $G$ is semisimple (Guillemin and Sternberg (1990, theorem 26.1)); conditions on the existence and uniqueness of the moment map can be formulated in terms of Lie algebra cohomology (see Guillemin and Sternberg (1990)). The obstruction to the existence of the moment map for a symplectic action of $G$ is an element of $H^1(\boldsymbol{g})$; the obstruction to uniqueness of the moment map is an element of $H^2(\boldsymbol{g})$, where $\boldsymbol{g}$ is the Lie algebra of $G$. See Guillemin and Sternberg (1990, proposition 24.1).

## Basic Properties of Moment Maps

**Proposition 1** (Guillemin–Sternberg (1982, 1984))
$$\mathrm{Im}(d\mu_m)^\perp = \mathrm{Lie}(\mathrm{Stab}(m))$$
*where $\perp$ denotes the annihilator under the canonical pairing $\boldsymbol{g}^* \otimes \boldsymbol{g} \to \mathbb{R}$.*

**Proof** We have
$$\omega(Y_m^\#, Z) = d\mu_Y(Z) = \langle Y, d\mu_m(Z)\rangle$$
for all $Z \in T_mM$. Thus, $Y$ annihilates all $\xi \in \mathrm{Im}(d\mu_m)$ if and only if $Y \in \mathrm{Lie}(\mathrm{Stab}(m))$.

**Corollary 1** *Zero is a regular value of $\mu$ if and only if $\mathrm{Stab}(m)$ is finite for all $m \in \mu^{-1}(0)$. In this situation, $\mu^{-1}(0)$ is a manifold and the stabilizer of the action at any point in $\mu^{-1}(0)$ is finite.*

**Example 1** Let $T$ be a torus acting on $M$ and let $F \subset M^T$ be a component of the fixed-point set. Then for any $f \in F$, we have $d\mu_f = 0$, so $\mu(F)$ is a point.

**Proposition 2**

(i) *If $H \subset G$ are two groups acting in a Hamiltonian fashion on a symplectic manifold M, then $\mu_H = \pi \circ \mu_G$ where $\pi : \boldsymbol{g}^* \to \boldsymbol{h}^*$ is the projection map. In other words, if $X \in \boldsymbol{h}$, then $\mu_H(m)(X) = \mu_G(m)(X)$ for any $m \in M$. One example that frequently arises is the case when H = T is a maximal torus of a compact Lie group G.*

(ii) *More generally if $f : H \to G$ is a Lie group homomorphism, and the two groups G and H act in a Hamiltonian fashion on a symplectic manifold M, in such a way that the action is compatible with the homomorphism f, then $\mu_H = f^* \circ \mu_G$ where $f^* : \boldsymbol{g} \to \boldsymbol{h}^*$ is induced from the homomorphism f. (The case* (i) *is the special case where f is the inclusion map.)*

(iii) *If two symplectic manifolds $M_1$ and $M_2$ are acted on in a Hamiltonian fashion by a group G with moment maps $\mu_1$ and $\mu_2$, then the moment map for the diagonal action of G on $M_1 \times M_2$ with the product symplectic structure is $\mu_1 + \mu_2$.*

**Example 2** The standard symplectic form on $S^2$ is $\omega = -d\cos\theta \wedge d\phi = -dz \wedge d\phi$ (where $\theta$ is the polar angle, $\phi$ is the azimuthal angle, and $z$ is the height function). The associated moment map for the action of U(1) on $S^2$ by rotation about the $z$ axis is $\mu(z, \phi) = z$.

**Example 3** If $\mathbb{R}^2 = \mathbb{C}$ has the symplectic structure $\omega = dx \wedge dy$, the moment map for the standard action of U(1) on $\mathbb{R}^2$ with multiplicity $m \in \mathbb{Z}$, in other words the action
$$u \in U(1) : z \in \mathbb{C} \mapsto u^m z$$
is $\mu(x, y) = -m(x^2 + y^2)/2$.

**Example 4** Suppose a torus $T$ acts on $\mathbb{C}$ preserving the standard symplectic structure, and suppose the action factors through a homomorphism $\mathcal{B} : T \to \mathrm{U}(1)$ which can be written as
$$\mathcal{B}(\exp_T X) = \exp_{\mathrm{U}(1)}(\beta(X))$$

in terms of a linear map $\beta \in \boldsymbol{t}^*$ that maps the integer lattice of $\boldsymbol{t}$ into $\mathbb{Z}$ (in other words, a weight) and the exponential maps

$$\exp_T : \boldsymbol{t} \to T$$

and

$$\exp_{\mathrm{U}(1)} : \mathbb{R} \to U(1)$$

(the latter being normalized as $\exp_{\mathrm{U}(1)}(t) = \mathrm{e}^{2\pi \mathrm{i} t}$). Then, by Proposition 2(ii) and Example 3 we see that the moment map for the action of $T$ on $\mathbb{C}$ is

$$\mu(z) = -\tfrac{1}{2}\beta |z|^2$$

It follows that if $T$ acts on $\mathbb{C}^n$ via a collection of weights $\beta_1, \ldots, \beta_n \in \boldsymbol{t}^*$, then the moment map is

$$\mu(z_1, \ldots, z_n) = -\frac{1}{2}\sum_{j=1}^{n} |z_j|^2 \beta_j$$

and the image of the moment map is the cone in $\boldsymbol{t}^*$ spanned by $\{\beta_1, \ldots, \beta_n\}$.

## The Symplectic Quotient

Since the moment map $\mu$ is equivariant, we may form the symplectic quotient (or Marsden–Weinstein reduction)

$$M_{\mathrm{red}} = M_0 = \mu^{-1}(0)/G$$

The symplectic structure on $M$ descends to give a symplectic structure on $M_0$. Corollary 1 implies that if 0 is a regular value of $\mu$, then $M_0$ is an orbifold.

**Remark 3** Another way to formulate Corollary 1 is that if $G$ acts freely, 0 is a regular value of $\mu$ so $\mu^{-1}(0)$ is a manifold with a free $G$ action, and hence $\mu^{-1}(0)/G$ is also a manifold. If the $G$ action is only locally free, then $\mu^{-1}(0)$ is still a manifold, but the quotient $\mu^{-1}(0)/G$ is only an orbifold.

**Remark 4** The definition of orbifold is due to Satake (1957); an alternate formulation is given in the paper by Henriques and Metzler (2004) and references cited there.

If $T$ is a torus, then the equivariance condition on the moment map reduces to invariance, so we may form the reduced space $M_t = \mu^{-1}(t)/T$ for any regular value $t \in \boldsymbol{t}^*$ of the moment map $\mu$; the space $M_t$ is a symplectic orbifold for any regular value $t$ of $\mu$.

**Example 5** Let U(1) act diagonally on $\mathbb{C}^n$ equipped with the standard symplectic structure

$$\omega = \frac{\mathrm{i}}{2}\sum_{j=1}^{n} \mathrm{d}z_j \wedge \mathrm{d}\bar{z}_j = \sum_{j=1}^{n} \mathrm{d}x_j \wedge \mathrm{d}y_j \qquad [1]$$

where $z_j = x_j + \mathrm{i}y_j$. The moment map for this action is

$$\mu(z_1, \ldots, z_n) = -\frac{1}{2}\sum_{j=1}^{n} |z_j|^2$$

so the symplectic quotient $\mu^{-1}(-1/2)/\mathrm{U}(1)$ is complex projective space

$$S^{2n-1}/\mathrm{U}(1) \cong \mathbb{C}P^{n-1}$$

More generally we may consider the reduced space $M_\lambda = \mu^{-1}(\mathcal{O}_\lambda)/G$ when $\mathcal{O}_\lambda$ is the orbit in $\boldsymbol{g}^*$ through $\lambda \in \boldsymbol{g}^*$ (coadjoint orbit). All such orbits may be parametrized by $\lambda \in \boldsymbol{t}^*_+$, where $\boldsymbol{t}^*_+$ is a chosen positive Weyl chamber in $\boldsymbol{t}^*$.

**Example 6** Let U($n$) act on $\mathbb{C}^n$ in the standard way, where $\mathbb{C}^n$ is equipped with the standard symplectic structure [1]. The moment map for this action is

$$\mu(z_1, \ldots, z_n)_{jk} = \frac{\mathrm{i}}{2} z_j \bar{z}_k \qquad [2]$$

which is the $(j, k)$ element of a matrix in the Lie algebra of U($n$). The standard symplectic form on $\mathbb{C}^n$ descends under reduction to the standard symplectic form on $\mathbb{C}P^{n-1}$ (which corresponds to the Fubini–Study metric).

**Example 7** (Coadjoint orbits). Let $\lambda \in \boldsymbol{g}^*$. We define a symplectic structure $\omega_\lambda$ on the coadjoint orbit $\mathcal{O}_\lambda$ (in terms of the vector fields $X^\#, Y^\#$ generated by the action of $X, Y \in \boldsymbol{g}$) by $\omega_\lambda(X^\#_\lambda, Y^\#_\lambda) = -\lambda([X, Y])$ at the point $\lambda \in \mathcal{O}_\lambda$ (and everywhere else on the orbit by equivariance). The moment map for the action of $G$ on $\mathcal{O}_\lambda$ with respect to this symplectic structure is the inclusion of $\mathcal{O}_\lambda$ in $\boldsymbol{g}^*$. (The symplectic structure on the orbit was found by Kirillov and Kostant; see, for instance, Berline *et al.* (1992, section 7.5).

**Example 8** (The shifting trick). Define a symplectic structure $\Omega$ on $M \times \mathcal{O}_\lambda$ by

$$\Omega = \omega^M - \omega_\lambda$$

Then for the moment map with respect to the induced action of $G$ on $M \times \mathcal{O}_\lambda$ we have

$$M_\lambda \cong (M \times \mathcal{O}_\lambda)_0$$

**Corollary 2** *Combining Example 6 with Proposition 2*(ii) *we see that for any linear action of a group G on* $\mathbb{C}P^{n-1}$ (*i.e., an action factoring through a representation* $G \to \mathrm{U}(n)$*, or in other words an action descending from a linear action on* $\mathbb{C}^n$) *the moment map factors as*

$$\mu = \pi \circ \hat{\mu}$$

*where* $\hat{\mu} : \mathbb{C}P^{n-1} \to \boldsymbol{u}(n)^*$ *is given in* [3] *below, and* $\pi : \boldsymbol{u}(\boldsymbol{n})^* \to \boldsymbol{g}^*$ *is the projection map.*

In particular, one often requires for a projective manifold $M$ (i.e., a compact complex manifold with an embedding into $\mathbb{C}P^{n-1}$) that the action of $G$ extends to a linear action on $\mathbb{C}P^{n-1}$. Thus, moment maps for such linear actions are given by [3] composed with $\pi$ and with the embedding of $M$ into $\mathbb{C}P^{n-1}$ (*see* also Cotangent Bundle Reduction, Poisson Reduction, Symmetry and Symplectic Reduction).

## Reduction in Stages

Suppose a compact Lie group $G$ acts in a Hamiltonian fashion on a symplectic manifold $M$, and $H$ is a normal subgroup of $G$. (For example, this hypothesis is satisfied if both $H$ and $G$ are tori.) Suppose also that 0 is a regular value for $\mu_H$ and $\mu_G$. Then the symplectic quotient $\mu_H^{-1}(0)/H$ is acted on naturally by the quotient group $G/H$, and this action is Hamiltonian; furthermore, the symplectic quotient of $\mu_H^{-1}(0)/H$ by $G/H$ is naturally isomorphic to $\mu_G^{-1}(0)/G$. (This result is known as "reduction in stages.")

Let $M$ be a symplectic manifold equipped with the Hamiltonian action of a torus $T$. Let $H \subset T$ be a Lie subgroup of $T$ (so $H$ is a torus whose dimension is smaller than the dimension of $T$). Let $\mu_T : M \to \mathrm{Lie}(T)^*$ and $\mu_H : M \to \mathrm{Lie}(H)^*$ be the moment maps: recall that $\mu_H = \pi_H \circ \mu_T$, where $\pi_H : \mathrm{Lie}(T)^* \to \mathrm{Lie}(H)^*$ is the standard projection.

For any $\eta \in \mathrm{Lie}(H)^*$ we may form the reduced space $M_\eta = \phi_H^{-1}(\eta)/H$. This is equipped with a Hamiltonian action of $T/H$.

**Example 9** Let $\mathrm{U}(n)$ act on $\mathbb{C}^n$ in the standard way. This action descends to an action on $\mathbb{C}P^{n-1}$, which is the symplectic quotient of $\mathbb{C}^n$ under the action of the diagonal $\mathrm{U}(1)$ subgroup of $\mathrm{U}(n)$. Hence, the moment map $\hat{\mu}$ for the action of $\mathrm{U}(n)$ on $\mathbb{C}P^{n-1}$ is given by the formula

$$\hat{\mu}([z_1, \ldots, z_n])_{jk} = \frac{\mathrm{i}}{2} \frac{z_j \bar{z}_k}{\sum_{\ell=1}^n |z_\ell|^2} \qquad [3]$$

which comes from the moment map [2] for the action of $\mathrm{U}(n)$ on $\mathbb{C}^n$.

## The Normal Form Theorem

There is a neighborhood of $\mu^{-1}(0)$ on which the symplectic form is given in a standard way related to the symplectic form $\omega_0$ on $M_{\mathrm{red}}$ (see, e.g., Guillemin and Sternberg (1990, sections 39–41)).

**Proposition 3** (Normal form theorem). *Assume 0 is a regular value of* $\mu$ (*so that* $\mu^{-1}(0)$ *is a smooth manifold and G acts on* $\mu^{-1}(0)$ *with finite stabilizers*). *Then there is a neighborhood* $U \cong \mu^{-1}(0) \times \{z \in \boldsymbol{g}^*, |z| \le h\} \subseteq \mu^{-1}(0) \times \boldsymbol{g}^*$ *of* $\mu^{-1}(0)$ *on which the symplectic form is given as follows. Let* $P \stackrel{\mathrm{def}}{=} \mu^{-1}(0) \stackrel{q}{\to} M_{\mathrm{red}}$ *be the orbifold principal G-bundle given by the projection map* $q : \mu^{-1}(0) \to \mu^{-1}(0)/G$, *and let* $\theta \in \Omega^1(P) \otimes \boldsymbol{g}$ *be a connection for it. Let* $\omega_0$ *denote the induced symplectic form on* $M_{\mathrm{red}}$, *in other words* $q^*\omega_0 = i_0^*\omega$. *Then if we define a 1-form* $\tau$ *on* $U \subset P \times \boldsymbol{g}^*$ *by* $\tau_{p,z} = z(\theta)$ (*for* $p \in P$ *and* $z \in \boldsymbol{g}^*$), *the symplectic form on U is given by*

$$\omega = q^*\omega_0 + \mathrm{d}\tau \qquad [4]$$

*Further, the moment map on U is given by* $\mu(p, z) = z$.

**Corollary 3** *Let t be a regular value for the moment map for the Hamiltonian action of a torus T on a symplectic manifold M. Then in a neighborhood of t, all symplectic quotients* $M_t$ *are diffeomorphic to* $M_{t_0}$ *by a diffeomorphism under which* $\omega_t = \omega_{t_0} + (t - t_0, \mathrm{d}\theta)$ *where* $\theta \in \Omega^1(\mu^{-1}(t_0)) \otimes \boldsymbol{t}$ *is a connection for the action of T on* $\mu^{-1}(t_0)$.

**Corollary 4** *Suppose G acts in a Hamiltonian fashion on a symplectic manifold M, and suppose 0 is a regular value for the moment map* $\mu$. *Then the reduced space* $M_\lambda = \mu^{-1}(\mathcal{O}_\lambda)/G$ *at the orbit* $\mathcal{O}_\lambda$ *fibers over* $M_0 = \mu^{-1}(0)/G$ *with fiber the orbit* $\mathcal{O}_\lambda$; *furthermore, if* $\pi : M_\lambda \to M_0$ *is the projection map, then the symplectic form* $\omega_\lambda$ *on* $\mu^{-1}(\mathcal{O}_\lambda)/G$ *is given as* $\omega_\lambda = \pi^*\omega_0 + \Omega_\lambda$, *where* $\omega_0$ *is the symplectic form on* $M_0$ *and* $\Omega_\lambda$ *restricts to the standard Kirillov–Kostant symplectic form on the fiber.*

## Convexity Theorems

**Theorem 1** (Atiyah (1982); Guillemin–Sternberg (1982 and 1984)). *Suppose M is a connected compact symplectic manifold equipped with a Hamiltonian action of a torus T. Then the image* $\mu(M)$ *is a convex polytope, the convex hull of* $\{\mu(F)\}$, *where F are the components of the fixed-point set of T in M.*

**Example 10** Consider the orbits $\mathcal{O}_t$ of $\mathrm{SU}(2)$ in $\mathbf{su}(2) \cong \mathbb{R}^3$ through $t \in \mathbb{R}^+$. The image of the moment map for the action of the maximal torus $T \cong \mathrm{U}(1)$ is the interval $[-t, t]$.

**Example 11** When $\mathcal{O}_t$ is the coadjoint orbit (through $t \in \boldsymbol{t}^*$) for a compact Lie group $G$ with maximal torus $T$, the image $\mu_T(\mathcal{O}_t)$ of the moment map $\mu_T$ for the action of the maximal torus $T$ is the convex hull $\text{Conv}\{wt : w \in W\}$, where $W$ is the Weyl group.

The convexity theorem above can be generalized to actions of nonabelian groups. If $M$ is a connected compact symplectic manifold equipped with a Hamiltonian action of a compact Lie group $G$ with maximal torus $T$ and positive Weyl chamber $\boldsymbol{t}_+$, then the intersection of the image $\mu(M)$ of the moment map with the positive Weyl chamber $\boldsymbol{t}_+$ (in other words, a fundamental domain for the action of the Weyl group on $\boldsymbol{t}$) is a convex polytope. This result is due to Kirwan (1984b) and for Kähler manifolds to Guillemin and Sternberg (1982 and 1984).

The proofs of Atiyah and Guillemin–Sternberg are based on Morse theory applied to the moment map. A key ingredient in the proofs is to establish that the fibers of the moment map are connected.

### The Moment Polytope

Given a compact symplectic manifold $M$ equipped with the Hamiltonian action of a torus $T$, we see that there is an associated polytope $P$, the "moment polytope." The fibers of the moment map $\mu$ are preserved by the action of $T$, so the value of $\mu$ parametrizes a family $\{M_t\}$ of symplectic quotients. By Theorem 1 the moment polytope is the convex hull of the images of the fixed-point set under the moment map.

By Proposition 1, we see that the moment polytope is decomposed according to the stabilizers of points in the preimage, and the critical values of the moment map are the images $\mu_T(W_j)$ of the fixed-point sets $W_j$ of one-parameter subgroups $S_j$ of $T$. These critical values form hyperplanes ("walls") which subdivide the moment polytope: the complement of the walls is a collection of open regions consisting of regular values of the moment map.

**Example 12** The group SU(3) has maximal torus $T \cong \text{U}(1)^2$. We identify $\boldsymbol{g}^*$ with $\boldsymbol{g}$ via the bi-invariant inner product (i.e., the Killing form) on $\boldsymbol{g}$, and thus identify $\boldsymbol{t}^*$ with $\boldsymbol{t}$. For $\lambda \in \boldsymbol{t}$, the Weyl group images of $\lambda$ are the six vertices of a hexagon: the "walls" in the moment polytope for the action of $T$ on the coadjoint orbit $\mathcal{O}_\lambda$ arising from the action of $G$ on $\boldsymbol{g}^*$ through $\lambda \in \boldsymbol{t}^*$ are the edges of the hexagon (exterior walls) and the three lines connecting opposite vertices (interior walls).

## Toric Manifolds

**Definition 1** *A toric manifold is a compact symplectic manifold $M$ of dimension $2n$ equipped with the effective Hamiltonian action of a torus $T$ of dimension $n$.*

**Example 13** Complex projective space $\mathbb{C}P^n$ with the obvious Hamiltonian action of $\text{U}(1)^n \subset \text{U}(1)^{n+1}$ is a toric manifold.

**Example 14** A special case of Example 13 is the 2-sphere $S^2 \cong \mathbb{C}P^1$ (with the action of U(1) given by rotation around one axis). The 2-sphere is a toric manifold.

### Elementary Properties of Toric Manifolds

If $M$ is a toric manifold, the fiber of the moment map for the action of $T$ is an orbit of the action. Hence, the symplectic quotient $M_t$ at any value $t \in \boldsymbol{t}^*$ is a point (if it is nonempty).

The regular values of $\mu$ are the interior points of the moment polytope $P$. All points in the preimage $\mu^{-1}(\partial P)$ are fixed points of some one-parameter subgroup of $T$. Points in the interior of a face $P_j$ of dimension $j$ are fixed by a subtorus of $T$ of dimension $n - j$. Hence, each fiber of $\mu$ over a point in $P_j$ is a quotient torus of dimension $j$. In particular, the vertices of the polytope are the images of the components of the fixed-point set of the whole torus $T$, and the inverse image of a vertex is contained in the fixed-point set of $T$.

The push-forward function $\mu_*(\omega^n/n!)$ under the moment map is just the characteristic function of the moment polytope.

### Delzant's Theorem

In fact, toric manifolds are characterized by their moment polytopes. A theorem of Delzant (1988) says that any polytope $P$ satisfying appropriate hypotheses (a simple polytope) is the moment polytope for some toric manifold; furthermore, if two toric manifolds acted on effectively by a torus $T$ have the same moment polytope, then they are $T$-equivariantly symplectomorphic. The first statement is proved by constructing a toric manifold which has the polytope $P$ as its moment polytope; if $P$ has $d$ faces of codimension 1, one constructs the toric manifold $M$ as a symplectic quotient of a vector space $V \cong \mathbb{C}^d$ by the linear action of a torus $T' \cong \text{U}(1)^{d-n}$. The torus $T \cong \text{U}(1)^n$ acting on $M$ is then obtained by reduction in stages, as the quotient of $\text{U}(1)^d$ by $T'$.

The construction of a toric manifold whose moment polytope is a given simple polytope is

given in Guillemin (1994, chapter 1). The second statement (namely that toric manifolds are classified by their moment polytopes) is proved in Delzant (1988).

**Example 15** The moment polytope for the action of $U(1)^n$ on $\mathbb{C}P^n$ is the $n$-simplex. This action descends from the action of $U(1)^{n+1}$ on $\mathbb{C}^{n+1}$, using reduction in stages: recall from Example 5 that we constructed $\mathbb{C}P^n$ as the symplectic quotient of $\mathbb{C}^{n+1}$ by the standard action of U(1).

## Cohomology Rings of Symplectic Quotients

For material on the equivariant cohomology of symplectic manifolds equipped with Hamiltonian group actions and the relation to the fixed-point set, we refer to Equivalent Cohomology and the Cartan Model. As in that reference we shall describe the equivariant cohomology of a Hamiltonian $G$-manifold using the Cartan model.

Two fundamental results of Kirwan give complementary descriptions of the equivariant cohomology of a symplectic manifold.

### Kirwan Injectivity

Kirwan's first theorem is the injectivity theorem:

**Theorem 2** (Injectivity theorem). *If $T$ is a compact torus and $M$ is a Hamiltonian $T$-space, then the direct sum of restriction maps to all components of the fixed-point set*

$$\oplus_F : H_T^*(M) \to H^*(F) \otimes S(\boldsymbol{t}^*)$$

*is injective.*

The proof appears in Kirwan (1984a); this material is treated in Equivalent Cohomology and the Cartan Model (theorem 6.6).

### Kirwan Surjectivity

Let $G$ be a complex torus, and let 0 be a regular value of the moment map $\mu$. Suppose $M$ is a compact symplectic manifold equipped with a Hamiltonian action of a compact Lie group $G$. There is a natural map $\kappa : H_G^*(M) \to H^*(M_{\text{red}})$ defined by

$$\kappa : H_G^*(M) \mapsto H_G^*(\mu^{-1}(0)) \cong H^*(M_{\text{red}})$$

(where the first map is the restriction map and the second is the identification of $H_G^*(Z)$ with $H^*(Z/G)$ when $G$ acts locally freely on $Z$ and the cohomology is taken with rational coefficients). The map $\kappa$ is obviously a ring homomorphism. Kirwan's second theorem treats the image of $\kappa$.

**Theorem 3** (Surjectivity theorem). *Under the above hypotheses, the map $\kappa$ is surjective.*

The proof of this theorem (Kirwan (1984a, 5.4 and 8.10); see also Kirwan (1992, section 6)) uses the Morse theory of the "Yang–Mills function" $|\mu|^2 : M \to \mathbb{R}$ to define an equivariant stratification of $M$ by strata $S_\beta$ which flow under the gradient flow of $|\mu|^2$ to a critical set $C_\beta$ of $|\mu|^2$. One shows that the function $|\mu|^2$ is equivariantly perfect (i.e., that the Thom–Gysin (long) exact sequence in equivariant cohomology decomposes into short exact sequences, so that one may build up the cohomology as

$$H_G^*(M) \cong H_G^*(\mu^{-1}(0)) \oplus \bigoplus_{\beta \neq 0} H_G^*(S_\beta)$$

Here, the stratification by $S_\beta$ has a partial order $>$; thus, one may define an open dense set $U_\beta = M - \cup_{\gamma > \beta} S_\gamma$ which includes the open dense stratum $S_b$ of points that flow into $\mu^{-1}(0)$ (note $S_b$ retracts onto $\mu^{-1}(b)$). The equivariant Thom–Gysin sequence is

$$\cdots \to H_G^{n-2d(\beta)}(S_\beta) \xrightarrow{i_{\beta *}} H_G^n(U_\beta) \to H_G^n(U_\beta - S_\beta) \to \cdots$$

To show that the Thom–Gysin sequence splits into short exact sequences, it suffices to know that the maps $(i_\beta)_*$ are injective. Since $i_\beta^*(i_\beta)_*$ is multiplication by the equivariant Euler class $e_\beta$ of the normal bundle to $S_\beta$, injectivity follows because this equivariant Euler class is not a zero divisor (see Kirwan (1984a, 5.4) for the proof).

Because $\kappa$ is a surjective ring homomorphism, it follows that

$$H_G^*(M_{\text{red}}) \cong H_T^*(M)/\text{Ker}(\kappa)$$

The above theorem is also valid when $G$ is the complexification of a compact semisimple Lie group. In this case, one must reduce at 0 (because of the condition that the moment map is equivariant, since $b = 0$ is the only value which is invariant under the coadjoint action). The case of reducing at coadjoint orbits can be treated using the proof for the case of reducing at 0 via the shifting trick (Example 8).

Several recent articles (Jeffrey and Kirwan (1995, 1997), Tolman and Weitsman 2003) compute Ker($\kappa$). Some articles compute Ker($\kappa$) in specific examples, notably the action of $S^1$ on products of two-dimensional spheres of general radii.

### The Residue Formula

One approach to identifying Ker($\kappa$) is the "residue formula," Jeffrey and Kirwan (1995), theorem 8.1:

**Theorem 4** (Jeffrey and Kirwan (1995), *corrected as in* Jeffrey and Kirwan (1997)).

*Let* $\eta \in H^*_G(M)$ *induce* $\eta_0 \in H^*(M_{\text{red}})$. *Then we have*

$$\int_{M_{\text{red}}} \kappa(\eta) e^{i\omega_{\text{red}}} = n_0 C^G \text{Res}\left(\mathcal{D}^2(X) \sum_{F \in \mathcal{F}} h_F^\eta(X)[dX]\right) \qquad [5]$$

*where* $n_0$ *is the order of the stabilizer in* $G$ *of a generic element of* $\mu^{-1}(0)$, *and the constant* $C^G$ *is defined by*

$$C^G = \frac{(-1)^{s+n_+}}{|W|\text{vol}(T)} \qquad [6]$$

*We have introduced* $s = \dim G$ *and* $l = \dim T$; *here* $n_+ = (s-l)/2$ *is the number of positive roots. Also,* $\mathcal{F}$ *denotes the set of components of the fixed-point set of* $T$, *and if* $F$ *is one of these components, then the meromorphic function* $h_F^\eta$ *on* $\boldsymbol{t} \otimes \mathbb{C}$ *is defined by*

$$h_F^\eta(X) = e^{i\mu(F)(X)} \int_F \frac{i_F^*\eta(X) e^{i\omega}}{e_F(X)} \qquad [7]$$

*and the polynomial* $\mathcal{D} : \boldsymbol{t} \to \mathbb{R}$ *is defined by* $\mathcal{D}(X) = \prod_{\gamma > 0} \gamma(X)$, *where* $\gamma$ *runs over the positive roots of* $G$.

The residue map Res is defined on (a subspace of) the meromorphic differential forms on $\boldsymbol{t} \otimes \mathbb{C}$: its definition depends on some choices, but the sum of the residues over all $F \in \mathcal{F}$ is independent of these choices. When $T = U(1)$, we define the residue on meromorphic functions of the form $e^{i\lambda X}/X^N$ when $\lambda \neq 0$ (for $N \in \mathbb{Z}$) by

$$\text{Res}\left(\frac{e^{i\lambda X}}{X^N}\right) = \text{Res}_{X=0} \frac{e^{i\lambda X}}{X^N}, \quad \text{if } \lambda > 0$$
$$= 0, \quad \text{if } \lambda < 0$$

More generally, the residue is specified by certain axioms (see Jeffrey and Kirwan (1995, proposition 8.11)), and may be defined as a sum of iterated multivariable residues $\text{Res}_{X_1 = \lambda_1} \ldots \text{Res}_{X_l = \lambda_l}$ for a suitably chosen basis of $\boldsymbol{t}$ yielding coordinates $X_1, \ldots, X_l$ (see Jeffrey and Kirwan (1997)).

### The Tolman–Weitsman Theorem

The Tolman and Weitsman (2002) theorem is as follows:

**Theorem 5** *We have*

$$\text{Ker}(\kappa) = \sum_S (K_-^S \oplus K_+^S) \qquad [8]$$

*Here,* $S$ *is a generic circle subgroup of* $T$ *and* $K_-^S$ *(resp.* $K_+^S$*) denote the set of all equivariant cohomology classes* $\eta$ *whose restriction to* $\mathcal{F}_-^S$ *(resp.* $\mathcal{F}_+^S$*) is zero. Here,*

$$\mathcal{F}_\pm = \{F \in \mathcal{F} : \pm\mu_S(F) > 0\}$$

*where* $\mu_S$ *is the component of the moment map in the direction of the Lie algebra of* $S$.

For more information, *see* Intersection Theory, Moduli Spaces: An Introduction, and Equivariant Cohomology and the Cartan Model.

## Acknowledgments

The author acknowledges support from NSERC.

*See also:* Cotangent Bundle Reduction; Equivariant Cohomology and the Cartan Model; Hamiltonian Fluid Dynamics; Intersection Theory; Moduli Spaces: An Introduction; Poisson Reduction; Stationary Phase Approximation; Symmetry and Symplectic Reduction.

## Further Reading


Atiyah MF (1982) Convexity and commuting Hamiltonians. *Bulletin of the London Mathematical Society* 14: 1–15.

Berline N, Getzler E, and Vergne M (1992) *Heat Kernels and Dirac Operators*, Grundlehren, vol. 298. Berlin–Heidelberg: Springer.

Cannas da Silva A (2001) *Lectures on Symplectic Geometry*, Lecture Notes in Mathematics, vol. 1764. Berlin: Springer.

Delzant T (1988) Hamiltoniens périodiques et images convexes de l'application moment. *Bulletin de la Société Mathématique de la France* 116: 315–339.

Guillemin V (1994) *Moment Maps and Combinatorial Invariants of Hamiltonian $T^n$-Spaces*, Progress in Mathematics, vol. 122. Boston: Birkhäuser.

Guillemin V, Ginzburg V, and Karshon Y (2002) *Moment Maps, Cobordisms, and Hamiltonian Group Actions*, Mathematical Surveys and Monographs, vol. 98. Providence, RI: American Mathematical Society.

Guillemin V and Kalkman J (1996) The Jeffrey–Kirwan localization theorem and residue operations in equivariant cohomology. *Journal für die Reine und Angewandte Mathematik* 470: 123–142.

Guillemin V and Sternberg S (1982) Convexity properties of the moment mapping I and II. *Inventiones Mathematical* 67: 491–513.

Guillemin V and Sternberg S (1984) Convexity properties of the moment mapping I and II. *Inventiones Mathematical* 77: 533–546.

Guillemin V and Sternberg S (1990) *Symplectic Techniques in Physics*. Cambridge: Cambridge University Press.

Henriques A and Metzler D (2004) Presentations of noneffective orbifolds. *Transactions of the American Mathematical Society* 356: 2481–2499.

Jeffrey LC and Kirwan FC (1995) Localization for nonabelian group actions. *Topology* 34: 291–327.

Jeffrey LC and Kirwan FC (1997) Localization and the quantization conjecture. *Topology* 36: 647–693.

Jeffrey LC and Kirwan FC (1998) Intersection theory on moduli spaces of holomorphic bundles of arbitrary rank on a Riemann surface. *Annals of Mathematics* 148: 109–196.

Kalkman J (1995) Cohomology rings of symplectic quotients. *Journal für die Reine und Angewandte Mathematik* 458: 37–52.

Kirwan F (1984a) *Cohomology of Quotients in Symplectic and Algebraic Geometry*. Princeton, NJ: Princeton University Press.
Kirwan F (1984b) Convexity properties of the moment mapping III. *Inventiones Mathematical* 77: 547–552.
Kirwan F (1992) The cohomology rings of moduli spaces of vector bundles over Riemann surfaces. *Journal of the American Mathematical Society* 5: 853–906.
McDuff D and Salamon D (1995) *Introduction to Symplectic Topology*. Oxford: Oxford University Press.
Satake I (1957) The Gauss–Bonnet theorem for V-manifolds. *Journal of the Mathematical Society of Japan* 9: 464–492.
Tolman S and Weitsman J (2003) The cohomology rings of symplectic quotients. *Communications in Analysis and Geometry* 11: 751–773.

# Hamiltonian Reduction of Einstein's Equations

**A E Fischer**, University of California, Santa Cruz, CA, USA
**V Moncrief**, Yale University, New Haven, CT, USA



## Introduction

In general relativity there are several levels within the framework of symplectic reduction of Einstein's equations at which one could attempt to define a Hamiltonian for the gravitational dynamics of a spatially closed universe. At the most basic unreduced level, this Hamiltonian is simply a linear function of the Einstein constraints and thus vanishes for any solution of the field equations. At the other extreme, at the deepest fully reduced level, one affects a transformation to a complete set of new canonical variables, the so-called "observables," which Poisson-commute with all of the constraints. At this level, the relevant Hamiltonian vanishes identically since each of the new canonical variables is a constant of the motion.

There is, however, an intermediate level wherein, after making a suitable choice of coordinate gauge and imposing the constraint equations, one can define a nonvanishing Hamiltonian that generates the gauge-fixed and constrained evolution equations and whose global infimum as a function on the relevant reduced phase space has direct topological significance. For the large class of manifolds on which this Hamiltonian can be defined, it has the attractive feature of globally monotonically decaying in the direction of cosmological expansion and thus evolves in such a way so as to seek and, in certain cases at least, to asymptotically attain its infimum value in the limit of this expansion. This Hamiltonian provides in these cases a weak Lyapunov function for the dynamics that can be used to partially control its global behavior. Since understanding the global behavior of solutions to Einstein's equations and its dependence upon the spatial topology is one of the central open problems in classical general relativity, the mathematical properties of this quantity are worthy of study.

Further information and details regarding the authors' work discussed in this article can be found in Fischer and Moncrief (2000, 2002a, b) and in the references therein.

## Topological Background

Einstein's field equations are nonvacuous and compatible with the introduction of material sources in $(n+1)$ dimensions for all $n \geq 2$, the case of most physical interest being of course $n=3$. For the field equations to be deterministic in a classical sense, that is, for the Cauchy problem to be well-posed, it is essential that they be formulated on a manifold that is globally hyperbolic and, in particular, has a product topology $M \times \boldsymbol{R}$ (roughly, space $\times$ time $=$ spacetime) where $M$ is a smooth ($C^\infty$) connected manifold of dimension $n$ and $\boldsymbol{R}$ is the real line. For the case of spatially closed universes of interest here, $M$ should be closed, that is, compact and without boundary. To simplify the analysis further, we also assume that $M$ is oriented, that is, orientable and an orientation has been chosen. Thus, unless stated otherwise, throughout this article $M$ will denote a smooth closed connected oriented $n$-manifold, $n \geq 2$, and all maps will be smooth.

Let "$\approx$" denote the diffeomorphic equivalence relation between smooth manifolds. Let $S^n$ denote the unit $n$-sphere in Euclidean $(n+1)$ space $\boldsymbol{R}^{n+1}, n \geq 1$. An $n$-manifold $M$ is trivial if $M \approx S^n$ and nontrivial if $M \not\approx S^n$.

The connected sum $M\#N$ of two closed connected oriented $n$-manifolds $M$ and $N$ is constructed by removing the interior of an embedded closed $n$-ball in $M$ and $N$, respectively, and then identifying the resulting $S^{n-1}$-boundary components by an orientation-reversing diffeomorphism of the $(n-1)$ spheres. The resulting manifold is smooth, connected, closed, and orientable, and is naturally oriented by the orientations on $M$ and $N$. Up to orientation-preserving diffeomorphism, this construction is independent of the choice of the embeddings of the $n$-balls and of the choice of the orientation-reversing diffeomorphism used to join the manifolds together.

Let $M$ be a nontrivial closed connected oriented $n$-manifold. Then $M$ is prime if $M \approx M_1 \# M_2$ implies that either $M_1 \approx S^n$ or $M_2 \approx S^n$ (but not both since we are assuming that $M$ is nontrivial). $M$ is a composite if $M$ can be written as a nontrivial connected sum, that is, if $M \approx M_1 \# M_2$ where both $M_1 \not\approx S^n$ and $M_2 \not\approx S^n$.

Note that with this definition, $S^n$ itself is not prime. This is analogous to the fact that for the positive integers, the unit 1 is not prime.

Now let $M$ be a connected $n$-manifold without boundary (not necessarily compact or orientable) and let $\boldsymbol{\pi}$ be a group. Then $M$ is a $K(\boldsymbol{\pi}, 1)$-manifold if $M$ is an Eilenberg–MacLane space, that is, if its first homotopy group (or fundamental group) $\pi_1(M) = \boldsymbol{\pi}$ and if all of its higher homotopy groups are trivial, that is, $\pi_i(M) = 0$ for $i > 1$ (equivalently, the universal covering space $\widetilde{M}$ of $M$ is contractible). Since the higher homotopy groups $\pi_i(M)$, $i > 1$, can be interpreted as the homotopy classes of continuous maps $S^i \to M$, each such map must be homotopic to a constant map. Thus a $K(\boldsymbol{\pi}, 1)$-manifold is said to be aspherical. Moreover, at the level of homotopy, all of the information about the topology of $M$ is contained in $\pi_1(M) = \boldsymbol{\pi}$. Thus, in particular, if $f$ is a map between connected aspherical manifolds that induces an isomorphism on their fundamental groups, then $f$ is a homotopy equivalence. Consequently, any two connected aspherical manifolds are homotopy equivalent if and only if their fundamental groups are isomorphic.

It is useful to define a connected $n$-manifold $M$ to be hyperbolizable if there exists a complete Riemannian metric $g$ on $M$ with constant negative sectional curvature, $K(g) = \text{constant} < 0$. We introduce this terminology to emphasize the underlying topology of manifolds that can support hyperbolic metrics rather than the geometry of such metrics. Similarly, $M$ is of flat type if $M$ admits a complete flat Riemannian metric $g, K(g) = 0$, and $M$ is of spherical type if $M$ admits a complete Riemannian metric $g$ on $M$ with constant positive sectional curvature, $K(g) = \text{constant} > 0$. In this latter case, by the Bonnet–Myers theorem, $M$ is necessarily compact and if $n$ is odd, then by Synge's theorem, $M$ is necessarily orientable. In fact, all such manifolds have been classified. As an important example, we note that a connected 3-manifold $M$ is of spherical type if and only if it is diffeomorphic to a spherical space form $S^3/\boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma}$ is a finite subgroup of $\mathbf{SO}(4)$ acting freely and orthogonally, that is, isometrically, on $S^3$.

Within the class of $K(\boldsymbol{\pi}, 1)$-manifolds are all flat-type and hyperbolizable $n$-manifolds, since any such manifold is isometrically covered by $\boldsymbol{R}^n$ in the flat case and homothetically covered by $\boldsymbol{H}^n$ in the hyperbolic case, where $\boldsymbol{H}^n$ is the standard single-sheeted spacelike hyperboloid with constant sectional curvature $K = -1$ embedded in $(n+1)$-Minkowski space $\boldsymbol{R}_1^{n+1}$.

We now return to our standard assumptions on $M$, so that $M$ is connected, closed, and oriented. For $n = 2$, these assumptions restrict the possibilities to $S^2, \boldsymbol{T}^2$, and the orientable higher genus surfaces $\boldsymbol{\Sigma}_p^2 = \boldsymbol{T}^2 \# \boldsymbol{T}^2 \# \cdots \# \boldsymbol{T}^2$ ($p$ factors) consisting of the connected sum of $p$ copies of $\boldsymbol{T}^2, p \geq 2$. However, from the point of view of $(2+1)$ gravity, unless one includes material sources or a cosmological constant, the spherical case is vacuous in that there are no vacuum solutions of the field equations on $S^2 \times \boldsymbol{R}$. The torus case is nonvacuous but the solutions, the so-called flat Kazner spacetimes, can all be found by elementary means. Thus only the case of genus $p \geq 2$ surfaces presents problems of interest.

For $n = 3$, although not essential for the program of reduction, it is convenient to assume the elliptization conjecture of 3-manifold topology. This conjecture asserts that a closed connected 3-manifold $M$ with finite fundamental group $\pi_1(M)$ must be diffeomorphic to a spherical space form $S^3/\boldsymbol{\Gamma}$, where, in such a quotient, $\boldsymbol{\Gamma}$ will always be a finite subgroup of $\mathbf{SO}(4)$ acting freely and orthogonally on $S^3$ and thus $\boldsymbol{\Gamma}$ is isomorphic to $\pi_1(M)$.

The simply connected case is the Poincaré conjecture. The full elliptization conjecture is equivalent to the Poincaré conjecture and a conjecture asserting that the only free actions of finite groups on $S^3$ are equivalent to the standard orthogonal ones. The elliptization conjecture is part of Thurston's geometrization program (Thurston 1997). For background information regarding 3-manifold topology, see Hempel (1976) and Jaco (1980).

Under the assumption of the elliptization conjecture, the Kneser–Milnor prime decomposition theorem asserts that if $M$ is nontrivial, then up to order, $M$ is uniquely diffeomorphic to a finite connected sum of the following form:

$$M \approx \left( \underbrace{S^3/\boldsymbol{\Gamma}_1 \# \cdots \# S^3/\boldsymbol{\Gamma}_k}_{k \text{ spherical factors}} \right) \# \left( \underbrace{(S^1 \times S^2) \# \cdots \# (S^1 \times S^2)}_{l \text{ wormholes (or handles)}} \right) \# \left( \underbrace{K(\boldsymbol{\pi}_1, 1) \# \cdots \# K(\boldsymbol{\pi}_m, 1)}_{m \text{ aspherical factors}} \right) \qquad [1]$$

where $k$, $l$, and $m$ are integers $\geq 0, k + l + m \geq 1$, and if either $k$, $l$, or $m$ is 0, terms of that type do not appear. Moreover, if $k \geq 1$, then each $\boldsymbol{\Gamma}_i, 1 \leq i \leq k$, is a finite nontrivial ($\boldsymbol{\Gamma}_i \neq \{I\}$) subgroup of $\mathbf{SO}(4)$

acting freely and orthogonally on $S^3$, and if $m \geq 1$, then each aspherical factor is a $K(\boldsymbol{\pi}_j, 1)$-manifold, $1 \leq j \leq m$, and thus is universally covered by a contractible manifold.

We remark that although in general a contractible 3-manifold need not be $\boldsymbol{R}^3$, conjecturally the universal covering manifold of a $K(\boldsymbol{\pi}, 1)$ 3-manifold is diffeomorphic to $\boldsymbol{R}^3$.

In 3-manifold topology, a concept closely related to that of a prime manifold is that of an irreducible manifold. A closed 3-manifold $M$ is irreducible if every embedded 2-sphere in $M$ is the boundary of an embedded closed 3-ball.

An embedded 2-sphere that does not bound such a 3-ball is essential. Thus in the prime decomposition [1] above, $M$ is decomposed along essential 2-spheres. For this reason, the prime decomposition is sometimes referred to as the sphere decomposition.

With the exception of $S^3$ which is irreducible but not prime (by definition of prime) and $S^1 \times S^2$ which is prime but not irreducible, a closed oriented 3-manifold is prime if and only if it is irreducible. We also remark that the Poincaré conjecture, when taken in the form that there do not exist any fake 3-cells, is equivalent to every $K(\boldsymbol{\pi}, 1)$ 3-manifold being irreducible. Thus in this article, since we are assuming the elliptization conjecture and hence the Poincaré conjecture, every $K(\boldsymbol{\pi}, 1)$ 3-manifold will automatically be irreducible.

Examples of the kinds of $K(\boldsymbol{\pi}, 1)$-factors that can occur in the decomposition [1] are as follows (we will explain the Seifert and graph designations below):

1. **Non-Seifert manifolds**. Closed oriented hyperbolizable manifolds diffeomorphic to $\boldsymbol{H}^3/\boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma}$ is a discrete torsion-free (i.e., no nontrivial element has finite order) co-compact subgroup of the Lie group $\mathrm{Isom}^+(\boldsymbol{H}^3)$ of orientation-preserving isometries of $\boldsymbol{H}^3$ which is Lie-group isomorphic to the proper orthochronous Lorentz group $\mathbf{SO}^{\uparrow}(1,3)$.
2. **Seifert manifolds**. $\boldsymbol{T}^3$ and five other 3-manifolds of flat type which are finitely covered by $\boldsymbol{T}^3$. Noting that $\boldsymbol{\Sigma}_1^2 = \boldsymbol{T}^2$, we remark that the product manifold $S^1 \times \boldsymbol{\Sigma}_1^2 = S^1 \times T^2 = \boldsymbol{T}^3$ is included in this class.
3. **Seifert manifolds.** Product manifolds $S^1 \times \boldsymbol{\Sigma}_p^2, p \geq 2$.
4. **Seifert manifolds**. Nontrivial circle bundles over $\boldsymbol{\Sigma}_p^2, p \geq 1$.
5. **Graph manifolds**. Any 3-manifold which fibers nontrivially over a circle with fiber $\boldsymbol{\Sigma}_p^2, p \geq 1$. Any such manifold is obtained by identifying the boundary components of $[0,1] \times \boldsymbol{\Sigma}_p^2$ with an orientation-reversing diffeomorphism of $\boldsymbol{\Sigma}_p^2$.

Since the handle $\boldsymbol{S}^1 \times \boldsymbol{S}^2$ and spherical manifolds $\boldsymbol{S}^3/\boldsymbol{\Gamma}$ are well understood, under the assumption of the elliptization conjecture the task of 3-manifold topology now reduces to understanding the topology of the (automatically irreducible) $K(\boldsymbol{\pi}, 1)$-factors that can occur in the prime decomposition [1]. Since essential 2-spheres have already been used to decompose $M$ into its prime components, the idea now is to use the next simplest 2-manifold, the 2-torus, to probe the irreducible $K(\boldsymbol{\pi}, 1)$-factors.

Let $i: \boldsymbol{T}^2 \to M$ be an embedding of $\boldsymbol{T}^2$ into a closed oriented 3-manifold $M$. Then the embedded torus $i(\boldsymbol{T}^2)$, identified with $\boldsymbol{T}^2$, is incompressible if the induced mapping of fundamental groups $i_*: \pi_1(\boldsymbol{T}^2) \to \pi_1(M)$ is injective. Thus noncontractible loops in $\boldsymbol{T}^2$ remain noncontractible when $\boldsymbol{T}^2$ is embedded in $M$, or, in other words, the ambient manifold $M$ does not fill in any homotopy hole that exists in $\boldsymbol{T}^2$ when standing alone.

A closed oriented 3-manifold $M$ is a Seifert-fibered space, or a Seifert manifold, if $M$ admits a foliation by circles. For example, if $\boldsymbol{S}^1$ acts freely on $M$, then $M$ is the total space of an $\boldsymbol{S}^1$-bundle over a surface $M/\boldsymbol{S}^1$ and $M$ is a Seifert-fibered space (see examples 2, 3, and 4 above). More generally, if $\boldsymbol{S}^1$ acts without fixed points (locally free), then $M$ is a Seifert-fibered space, and in either case the fibers of $M$ are the orbits of the $\boldsymbol{S}^1$-action.

All spherical 3-manifolds are Seifert fibered with base $\boldsymbol{S}^2$. Also, the product manifold $\boldsymbol{S}^1 \times \boldsymbol{S}^2$ is Seifert fibered, as are all manifolds finitely covered by $\boldsymbol{T}^3$, and thus all 3-manifolds of flat type are Seifert fibered. The only nontrivial connected sum that is a Seifert-fibered space is $\boldsymbol{P}^3 \# \boldsymbol{P}^3$. No hyperbolizable manifold is Seifert fibered. Thus the remaining Seifert manifolds are among the nonhyperbolizable nonflat type $K(\boldsymbol{\pi}, 1)$-manifolds (i.e., those for which $M$ does not admit either a hyperbolic or a flat Riemannian metric).

A generalization of Seifert-fibered spaces are the graph manifolds. A closed oriented 3-manifold $M$ is a graph manifold if there exists a finite collection $\{\boldsymbol{T}_i^2\}$ of disjoint embedded incompressible tori $\boldsymbol{T}_i^2 \subset M$ such that each component $M_j$ of $M \backslash \cup \boldsymbol{T}_i^2$ is a Seifert-fibered space. Thus a graph manifold is a union of Seifert-fibered spaces glued together by toral automorphisms along toral boundary components. The collection of tori may be empty so that, in particular, a Seifert-fibered manifold is a graph manifold.

We remark that the manifolds described by example 5 above are graph manifolds. We also remark that graph manifolds are closed under connected sums so that a graph manifold may be a composite. This contrasts with the situation for Seifert spaces which, with the exception of $\boldsymbol{P}^3 \# \boldsymbol{P}^3$, are not composites.

Conjecturally, the most general $K(\boldsymbol{\pi}, 1)$-manifold, not included in the list above, consists of "gluing together" across disjoint embedded incompressible

tori a finite collection of finite-volume-type hyperbolizable manifolds, that is, noncompact manifolds that admit a finite-volume complete hyperbolic metric, together with a possibly empty finite collection of irreducible graph manifolds with toral boundaries. Thus, overall, in this picture, to decompose an arbitrary closed oriented 3-manifold $M$ into its elementary constituents, one first cuts along essential 2-spheres to break $M$ down into its prime factors, that is, the nontrivial spherical $S^3/\Gamma$-factors, the wormhole $(S^1 \times S^2)$-factors, and the aspherical $K(\pi, 1)$-factors, as given by [1]. Then one cuts each nonelementary $K(\boldsymbol{\pi}, 1)$-factor along incompressible tori to separate these factors into their final finite-volume-type hyperbolizable and irreducible graph manifold components. The graph manifold components can then be further broken down along incompressible tori into Seifert-fibered pieces, finally yielding the toral decomposition of Jaco, Shalen, and Johannson (see Anderson (1997), Jaco (1980), and the end of the section "The Reduced Hamiltonian" for further details).

The Thurston (1997) geometrization program, which implies that every closed oriented 3-manifold has the structure described by the above prime (or spherical) and toroidal decomposition, has been the subject of recent work by G Perelman (see Anderson (2003) and the references therein) who has argued that it can be proved by an enhancement of the Ricci flow program of R Hamilton (see the collected papers edited by Cao *et al*. (2003)). Without entering into the technical issues surrounding the completeness of Perelman's proof, one can simply limit one's attention to 3-manifolds of the above type. If geometrization is correct, then no 3-manifolds of interest have been excluded.

Returning to the general case of $n$-manifolds, in the program of Hamiltonian reduction of Einstein's equations, an important consideration is under what topological conditions on $M$ can the conformal classes of $M$ be uniquely represented by a given metric in each class. To analyze that question, we introduce the concept of the Yamabe type of a manifold.

Let $M$ be a connected closed oriented $n$-manifold, $n \geq 3$. There is no topological obstruction to the existence of Riemannian metrics with constant negative scalar curvature, so all such manifolds admit a Riemannian metric $g$ such that $R(g) = -1$. However, there are topological obstructions for zero scalar curvature and positive constant scalar curvature metrics on $M$. To help categorize these topological obstructions, we introduce the following terminology:

1. $M$ is of *positive Yamabe type* if $M$ admits a Riemannian metric $g_1$ with scalar curvature $R(g_1) = 1$;
2. $M$ is of *zero Yamabe type* if $M$ admits a Riemannian metric $g_0$ with $R(g_0) = 0$, but no Riemannian metric $g$ with $R(g) = 1$; and
3. $M$ is of *negative Yamabe type* if $M$ admits no Riemannian metric $g$ with $R(g) = 0$.

The definition of Yamabe type partitions the class of connected closed oriented $n$-manifolds, $n \geq 3$, into three classes that are mutually exclusive and exhaustive. The following rather complete topological information regarding 3-manifolds of negative Yamabe type is known.

Let $M$ be a connected closed oriented 3-manifold. Assume that the Poincaré conjecture is true. Then $M$ is of negative Yamabe type if and only if $M$ satisfies one of the following three mutually exclusive conditions:

1. $M$ is hyperbolizable (and thus is a $K(\boldsymbol{\pi}, 1)$-manifold; see example 1 of $K(\boldsymbol{\pi}, 1)$-manifolds);
2. $M$ is a nonhyperbolizable nonflat type $K(\boldsymbol{\pi}, 1)$-manifold (see examples 3, 4, and 5 of $K(\boldsymbol{\pi}, 1)$-manifolds);
3. $M$ has a nontrivial connected sum decomposition (i.e., $M$ is a composite) in which at least one factor is a $K(\boldsymbol{\pi}, 1)$-manifold; that is, $M \approx M' \# K(\boldsymbol{\pi}, 1)$, where $M' \not\approx S^3$. In this case the $K(\boldsymbol{\pi}, 1)$-factor may be either of flat type or hyperbolizable.

We remark that (1) is the vast class of closed oriented hyperbolizable 3-manifolds. We also remark that the six closed orientable 3-manifolds of flat type, although $K(\boldsymbol{\pi}, 1)$-manifolds, are excluded from (2) as they are not of negative Yamabe type (they are of zero Yamabe type). Lastly we remark that if $M$ is of negative Yamabe type and Seifert fibered, then $M$ must be of type (2) (see remarks on Seifert-fibered spaces above).

In any dimension $n \geq 3$, a manifold $M$ of negative Yamabe type has the property that it admits no Riemannian metric $g$ having scalar curvature $R(g) \geq 0$ everywhere on $M$, or, in other words, every Riemannian metric on $M$ has scalar curvature which is negative somewhere. For such a manifold $M$, Yamabe's theorem asserts that each Riemannian metric $g$ on $M$ is uniquely globally conformal to a metric $\gamma$ with scalar curvature $R(\gamma) = -1$ (see also [21]). Thus one can represent the conformal classes of Riemannian metrics on $M$ in a suitable function space setting by an infinite-dimensional submanifold

$$\mathcal{M}_{-1} = \mathcal{M}_{-1}(M) = \{\gamma \in \mathcal{M} \mid R(\gamma) = -1\} \quad [2]$$

of the space $\mathcal{M} = \mathcal{M}(M) = \text{Riem}(M)$ of Riemannian metrics on $M$ (see Fischer and Marsden (1975) for details). For this reason, we refer to metrics $\gamma$ in $\mathcal{M}_{-1}$ as conformal metrics.

The quotient of $\mathcal{M}_{-1}$ by the natural action of $\mathcal{D}_0 = \mathcal{D}_0(M) = \mathrm{Diff}_0(M)$, the connected component of the identity of the diffeomorphism group $\mathcal{D} = \mathcal{D}(M) = \mathrm{Diff}(M)$ of $M$, defines an orbit space (not necessarily a manifold) $\mathcal{T} = \mathcal{T}(M)$,

$$\mathcal{T} = \frac{\mathcal{M}_{-1}}{\mathcal{D}_0} \qquad [3]$$

which, when $M$ is of negative Yamabe type, we define as the Teichmüller space of conformal structures on $M$.

In two dimensions in the case of a higher genus manifold $\mathbf{\Sigma}_p^2, p \geq 2$, this construction leads precisely to the conventional Teichmüller space, as discussed by Fischer and Tromba (1984). In this case the resulting Teichmüller space

$$\mathcal{T}_p = \mathcal{T}(\mathbf{\Sigma}_p^2) = \frac{\mathcal{M}_{-1}(\mathbf{\Sigma}_p^2)}{\mathcal{D}_0(\mathbf{\Sigma}_p^2)} \approx \boldsymbol{R}^{6p-6} \qquad [4]$$

is then a manifold diffeomorphic to $\boldsymbol{R}^{6p-6}$, which then plays the role of the natural reduced configuration space for the Einstein equations in $(2+1)$ dimensions. Moreover, these constructions can be carried out globally using known global cross sections for the $\mathcal{D}_0(\mathbf{\Sigma}_p^2)$ action on $\mathcal{M}_{-1}(\mathbf{\Sigma}_p^2)$. These global cross sections can then be used to provide an explicit model for the Teichmüller space $\mathcal{T}_p$ as a finite-dimensional subspace of $\mathcal{M}_{-1}(\mathbf{\Sigma}_p^2)$.

For $n = 3, \mathcal{T} = \mathcal{T}(M)$ plays the analogous role for the reduced field equations in $(3+1)$ dimensions. Moreover, for many 3-manifolds it is possible to show that $\mathcal{T}$ is itself an infinite-dimensional contractible manifold, rather than something more general such as an orbifold or a stratified union of manifolds. For technical simplicity, we shall assume throughout this article that $\mathcal{T}$ is a manifold. Our results remain valid in the more general case but in that case one must work on stratified spaces (see Fischer (1970) for results on the structure of orbit spaces when they are not manifolds).

For higher-dimensional manifolds there is no analog of the Thurston geometrization program. Indeed, it is known that the set of closed $n$-manifolds for $n \geq 4$ is so rich that no purely algebraic classification is possible. Nevertheless, for manifolds of negative Yamabe type, every Riemannian metric $g$ is still uniquely conformal to a metric $\gamma \in \mathcal{M}_{-1}$ so that the orbit space $\mathcal{T} = \mathcal{M}_{-1}/\mathcal{D}_0$ still represents the Teichmüller space of conformal equivalence classes on $M$. However, in these higher-dimensional cases, very little is known about the structure of $\mathcal{T}$.

## The Field Equations

Relative to a global time coordinate $t = x^0$ and local spatial coordinates $(x^1, \ldots, x^n)$ on a connected closed oriented $n$-manifold $M$, one can express the line element of an arbitrary $(n+1)$-Lorentzian metric with signature $(-+\cdots+)$ ($n$ positive signs) in the form

$$\begin{aligned} \mathrm{d}s^2 &= {}^{(n+1)}g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \\ &= -N^2\mathrm{d}t^2 + g_{ij}(\mathrm{d}x^i + X^i\mathrm{d}t)(\mathrm{d}x^j + X^j\mathrm{d}t) \end{aligned} \qquad [5]$$

where ${}^{(n+1)}g_{\mu\nu}$ denotes the components of the space-time metric, $0 \leq \mu, \nu \leq n$, where the Riemannian metric $g$ with components $g_{ij}$ is the first fundamental form induced on each $t =$ constant hypersurface, where the time-dependent positive function $N = N(x,t) > 0$ is referred to as the lapse function, and where the time-dependent spatial vector field $X = X(x,t)$ with components $X^i = {}^{(n+1)}g_{0j}\, g^{ij}$, where $g^{ij}$ denotes the inverse of the spatial metric $g_{ij}$, is referred to as the shift vector field.

Let $\ell$ denote the dimension *length*. In this article we use the convention that the spatial coordinates $(x^1, \ldots, x^n)$ are always dimensionless, but the time coordinate $t$ may have a dimension (see [19] and [36]). Since the line element $\mathrm{d}s^2$ [5] has dimension $\ell^2$ and the spatial coordinates are dimensionless, the physical spatial metric coefficients $g_{ij}$ also have dimension $\ell^2$. If the time coordinate $t$ has a dimension, then the dimension of the lapse function $N$ is such that the quantity $N\mathrm{d}t$ has dimension $\ell$ and the dimension of the shift vector field $X$ is such that the quantity $X\mathrm{d}t$ is dimensionless.

We now briefly consider the canonical formulation of Einstein's equations. For more information regarding this formulation, see Arnowitt, Deser, and Misner (1962) (ADM) or Fischer and Marsden (1972) for a global perspective. We remark that the canonical formulation of gravity itself is local and is valid for any spatial topology of $M$. However, as we shall see, Hamiltonian reduction of gravity along the lines described in this article requires the topological restriction that $M$ be of negative Yamabe type.

The standard definition of the second fundamental form $k$, or extrinsic curvature, induced on a $t =$ constant hypersurface leads to the coordinate formula

$$k_{ij} = -\frac{1}{2N}\left(\frac{\partial g_{ij}}{\partial t} - X_{i|j} - X_{j|i}\right) \qquad [6]$$

where the vertical bar signifies covariant differentiation with respect to the spatial metric $g$ and spatial indices are raised and lowered using this metric. The

natural momentum variable conjugate to $g$ turns out to be the 2-contravariant symmetric tensor density $\pi$ (that is, $\pi$ is a relative tensor of weight 1) whose components in a positively oriented local coordinate chart $(x^1, \ldots, x^n)$, that is, in a chart in the orientation atlas of $M$, are given by

$$\pi^{ij} = -\sqrt{\det g_{kl}}\left(k^{ij} - (\mathrm{tr}_g k)g^{ij}\right) \qquad [7]$$

where $k^{ij} = g^{ik}g^{jl}k_{kl}$ is the contravariant form of $k$, and where

$$\tau = \tau(g, k) = \mathrm{tr}_g k = g^{ij}k_{ij} \qquad [8]$$

is the trace of the second fundamental form, or the mean (extrinsic) curvature. From the coordinate formula [6] for the extrinsic curvature, we see that the components $k_{ij}$ have dimension $\ell^{-1}\ell^2 = \ell$ and thus the mean curvature $\tau = \mathrm{tr}_g k = g^{ij}k_{ij}$ has the dimension $\ell^{-2}\ell = \ell^{-1}$.

Let $\sqrt{\det g}$ denote the (global) scalar density and $\mathrm{d}\mu_g$ denote the (global) Riemannian measure on $M$ determined by the Riemannian metric $g$ (note that here $d$ is not the exterior derivative). Similarly, let $\mu_g$ denote the volume element, a nonvanishing $n$-form on $M$, determined by $g$ and the orientation on $M$. In a positively oriented local coordinate chart $(x^i) = (x^1, \ldots, x^n)$ on $M$, $(\sqrt{\det g})_{(x^i)} = \sqrt{\det g_{ij}}$, $(\mathrm{d}\mu_g)_{(x^i)} = \sqrt{\det g_{ij}}\,\mathrm{d}x^1\mathrm{d}x^2\cdots\mathrm{d}x^n = \sqrt{\det g_{ij}}\,\mathrm{d}^n x$, where $\mathrm{d}^n x = \mathrm{d}x^1\mathrm{d}x^2\cdots\mathrm{d}x^n$ is the Lebesque measure in $\boldsymbol{R}^n$, and $(\mu_g)_{(x^i)} = \sqrt{\det g_{ij}}\,\mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \cdots \wedge \mathrm{d}x^n$. We adopt the convention of suppressing the coordinate-chart designation $(x^i)$ so that one can, for example, write with some ambiguity $\sqrt{\det g} = (\sqrt{\det g})_{(x^i)} = \sqrt{\det g_{ij}}$.

We let

$$\mathrm{vol}(M, g) = \int_M \mu_g = \int_M \mathrm{d}\mu_g = \int_M \sqrt{\det g}\,\mathrm{d}^n x \qquad [9]$$

denote the volume of the Riemannian manifold $(M, g)$, given by either the integral of the volume $n$-form $\mu_g$ or the Riemannian measure $\mathrm{d}\mu_g$ over $M$, which is given in the last integral in its coordinate form using the suppressed coordinate-chart convention adopted above. As expected, the spatial physical volume has dimension $(\ell^2)^{n/2} = \ell^n$.

We shall refer to the canonical variables $(g_{ij}, \pi^{ij})$ as the physical variables, in contrast to the reduced or conformal variables $(\gamma_{ij}, (p^{\mathrm{TT}})^{ij})$ to be introduced later.

Note that the mean curvature $\tau = \mathrm{tr}_g k$ is a scalar function on $M$ whereas $\mathrm{tr}_g\pi$ is a scalar density on $M$. Taking the trace of [7] expresses the mean curvature in terms of the canonical variables $(g, \pi)$,

$$\tau = \tau(g, \pi) = \mathrm{tr}_g k = \frac{1}{(n-1)\sqrt{\det g}}\mathrm{tr}_g \pi \qquad [10]$$

Using [10], eqn [7] can be inverted to give $k$ in terms of $g$ and $\pi$,

$$k_{ij} = -\frac{1}{\sqrt{\det g}}\left(\pi_{ij} - \frac{1}{(n-1)}(\mathrm{tr}_g\pi)g_{ij}\right) \qquad [11]$$

and then combined with [6] to give the kinematical equation

$$\frac{\partial g_{ij}}{\partial t} = \frac{2N}{\sqrt{\det g}}\left(\pi_{ij} - \frac{1}{(n-1)}(\mathrm{tr}_g\pi)g_{ij}\right) + X_{i|j} + X_{j|i} \qquad [12]$$

In terms of the canonical variables $(g, \pi)$, a Hamiltonian form for the action for Einstein's vacuum field equations can be expressed as

$$I_{\mathrm{ADM}}(g, \pi) = \int_I \mathrm{d}t \int_M \left(\pi^{ij}\frac{\partial g_{ij}}{\partial t} - N\mathcal{H}(g, \pi) - X^i \mathcal{J}_i(g, \pi)\right)\mathrm{d}^n x \qquad [13]$$

where $I = [t_0, t_1] \subset \boldsymbol{R}$ is a closed interval and where the Hamiltonian (scalar) density $\mathcal{H}(g, \pi)$ and the momentum (1-form) density $\mathcal{J}(g, \pi)$ are given by

$$\mathcal{H}(g, \pi) = \frac{1}{\sqrt{\det g}}\left(\pi \cdot \pi - \frac{1}{n-1}(\mathrm{tr}_g\pi)^2\right) - \sqrt{\det g}\,R(g) \qquad [14]$$

$$= \frac{1}{\sqrt{\det g}}\left(g_{ij}g_{kl}\pi^{ik}\pi^{jl} - \frac{1}{n-1}(g_{ij}\pi^{ij})^2\right) - \sqrt{\det g}\,R(g) \qquad [15]$$

$$\mathcal{J}_i(g, \pi) = 2(\delta_g \pi)_i = -2g_{ij}\pi^{jk}{}_{|k} \qquad [16]$$

where $\pi \cdot \pi$ is the $g$-metric contraction of $\pi$ with itself, and where, as above, $R(g)$ is the scalar curvature of the spatial metric. We also note that each of the three terms in the integrand of [13] are global scalar densities and thus can be integrated over $M$ without any further involvement of the metric $g$.

Variation of $I_{\mathrm{ADM}}$ with respect to the lapse function and shift vector field yields the constraint equations

$$\mathcal{H}(g, \pi) = 0 \qquad [17]$$

$$\mathcal{J}_i(g, \pi) = 0 \qquad [18]$$

which comprise that subset of the empty space $(n+1)$-Einstein field equations corresponding to the normal–normal and normal–tangential projections of the Einstein tensor relative to a $t = $ constant initial hypersurface. Variation of $I_{\mathrm{ADM}}$ with respect to $\pi^{ij}$ reproduces the kinematical equation [12], whereas

variation of $I_{ADM}$ with respect to $g_{ij}$ generates the complementary tangential–tangential projections of Einstein's equations.

There are no evolution or constraint equations for either the lapse function $N$ or the shift vector field $X$ and therefore these quantities must be fixed by either externally imposed or implicitly defined gauge conditions. A convenient choice, for which a local existence and well-posedness theorem for the corresponding field equations can be established in any dimension $n \geq 2$, is given indirectly by imposing constancy of the mean curvature and a spatial harmonic gauge condition on each $t = \text{constant}$ slice (see Andersson and Moncrief (2003, 2004)). These constant mean curvature spatial harmonic (CMCSH) gauge conditions are given, respectively, by the equations

$$t = \tau \qquad [19]$$

$$g^{ij}(\Gamma^k_{ij}(g) - \Gamma^k_{ij}(\hat{g})) = 0 \qquad [20]$$

where from [10], $\tau$ is a function of the canonical variables $(g, \pi)$ and where $\hat{g}$ is some convenient fixed spatial reference metric (or background metric) on $M$. The latter condition corresponds to the requirement that the identity map between the Riemannian manifolds $(M, g)$ and $(M, \hat{g})$ be harmonic. Neither of these conditions involves the lapse function or shift vector field directly but their preservation in time implemented by the demand that the time derivatives of the given conditions be enforced leads immediately to a linear elliptic system for $(N, X^i)$ which determines these variables. The foregoing formalism is easily extended to the nonvacuum field equations in the presence of suitable material sources whose field equations are amenable to a constrained Hamiltonian treatment. To simplify the analysis, such sources will be ignored in the present discussion.

For the special case of Einstein gravity in $(2+1)$ dimensions, there is an elegant, alternative, triad-based formulation of the action functional as an $\mathrm{Isom}(\boldsymbol{R}^3_1)$-invariant gauge-theoretic Chern–Simons action, where $\mathrm{Isom}(\boldsymbol{R}^3_1)$ denotes the full isometry group, or the Poincaré group (= the inhomogeneous Lorentz group), of $(2+1)$-Minkowski space $\boldsymbol{R}^3_1$. For nondegenerate triads the resulting field equations for this alternative formulation can easily be shown to be equivalent to those of the conventional formalism when the latter is re-expressed in terms of triads but the new formulation allows for meaningful field equations in the case of degenerate triads as well and thus suggests a potentially interesting generalization of the theory (see Carlip (1998) for details).

In any dimension $n \geq 2$, there is a well-known technique, pioneered by Lichnerowicz (1955), for solving the constraint equations on a constant mean curvature (CMC) hypersurface (see Choquet-Bruhat and York (1980) and Isenberg (1995)). Of major importance for the treatment of Hamiltonian reduction is that if $n = 2$ and $M = \boldsymbol{\Sigma}^2_p, p \geq 2$, or if $n \geq 3$ and $M$ is of negative Yamabe type, then every Riemannian metric $g$ on $M$ is uniquely globally pointwise conformal to a metric $\gamma$ which satisfies $R(\gamma) = -1$ (see remark above [2]). Thus, from now on, we assume this topological condition on $M$. In this case, every Riemannian metric $g$ on $M$ can be uniquely expressed as

$$g = \begin{cases} e^{2\varphi}\gamma & \text{if } n = 2 \text{ and } M = \boldsymbol{\Sigma}^2_p,\ p \geq 2 \\ \varphi^{4/(n-2)}\gamma & \text{if } n \geq 3 \text{ and } M \text{ is of negative Yamabe type} \end{cases} \qquad [21]$$

with the conformal metric $\gamma$ normalized so that $R(\gamma) = -1$ and with the specific form of the coefficient conformal factor being chosen to simplify calculations involving the curvature tensors. In the case $n \geq 3, \varphi$ is positive and thus the space of all Riemannian metrics on $M$ is parametrized by $\mathcal{M}_{-1}$ and the space of scalar functions $\varphi > 0$ on $M$. The function $\varphi$ is then determined by solving the Hamiltonian constraint [17] (see also the remark before [33]).

In the given CMC slicing and imposing the vacuum field equations, since by the momentum constraint $\pi$ must have zero divergence (see [16] and [18]), one finds that $\pi^{ij}$ must be expressible in the form

$$\pi^{ij} = \left(\pi^{\mathrm{TT}}\right)^{ij} + \frac{1}{n}(\mathrm{tr}_g\pi)g^{ij} \qquad [22]$$

where $\pi^{\mathrm{TT}}$ is transverse (i.e., divergence-free) and traceless with respect to $g$. In the nonvacuum case, $\pi^{ij}$ picks up an additional summand determined by the sources in the modified momentum constraint [18].

Substitution of the foregoing decompositions of $(g_{ij}, \pi^{ij})$ into the Hamiltonian constraint leads to a nonlinear elliptic equation for $\varphi$ which, under the conditions assumed here, determines this function uniquely, provided $\tau \neq 0$. No solutions exist for $\tau = 0$ (equivalently, $\mathrm{tr}_g\pi = 0$) since from [14], [17], and [22], the Hamiltonian constraint would then immediately imply that

$$R(g) = \frac{1}{\det g}(\pi \cdot \pi) = \frac{1}{\det g}\left(\pi^{\mathrm{TT}} \cdot \pi^{\mathrm{TT}}\right) \geq 0 \qquad [23]$$

everywhere on $M$, which is not possible for a manifold $M$ of negative Yamabe type. Instantaneous vanishing of the mean curvature, the defining property of a maximal hypersurface, would correspond to a moment at which an expanding universe

ceases to expand or a collapsing universe ceases to collapse. From [23], such behavior is topologically excluded here by the requirement that $M$ be of negative Yamabe type (see also the discussion after [36]).

In the unreduced formalism of $I_{\mathrm{ADM}}$, the role of a super-Hamiltonian is played by the functional

$$H_{\mathrm{super}}(g,\pi) = \int_M (N\mathcal{H}(g,\pi) + X^i \mathcal{J}_i(g,\pi))\,\mathrm{d}^n x \qquad [24]$$

which evidently vanishes whenever the constraints are satisfied. To achieve a fully reduced formulation wherein again the effective Hamiltonian would vanish, one could endeavor to solve the associated Hamilton–Jacobi equations

$$\mathcal{H}(g_{ij}, \delta S/\delta g_{ij}) = 0 \qquad [25]$$

$$\mathcal{J}_k(g_{ij}, \delta S/\delta g_{ij}) = 0 \qquad [26]$$

for a real-valued functional $S = S(g, \alpha^A)$ of the metric $g$ and a set of additional independent parameters $\alpha^A$. A complete solution $S(g_{ij}, \alpha^A)$ would be one for which an arbitrary solution $(g_{ij}, \pi^{ij})$ of the constraints could be realized as $(g_{ij}, \delta S/\delta g_{ij})$ for a suitable (unique) choice of the $\alpha^A$. A complementary set of reduced canonical variables $\beta_A$ (the momenta conjugate to the $\alpha_A$'s) could then be defined by $\beta_A = \delta S/\delta\alpha^A$ and one could in principle solve the equations

$$\pi^{ij} = \frac{\delta S}{\delta g_{ij}} \qquad [27]$$

$$\beta_A = \frac{\delta S}{\delta \alpha^A} \qquad [28]$$

for $(\alpha^A, \beta_A)$ as functionals of the canonical variables $(g_{ij}, \pi^{ij})$. This procedure, if it could be carried out, would ensure that these functionals $(\alpha^A(g,\pi), \beta_A(g,\pi))$ Poisson-commute with all of the constraints and hence are conserved for an arbitrary slicing of spacetime. Conversely, if a suitable set of gauge conditions such as the CMCSH conditions were imposed, one could in principle solve for the remaining independent canonical variables as functionals of the $(\alpha^A, \beta_A)$ and an internal variable, such as the mean curvature $\tau$, which plays the role of time, and hence solve the field equations for $(g_{ij}, \pi^{ij})$ in the chosen gauge.

This proposal is purely heuristic in $(3+1)$ and higher dimensions in that there is no known procedure for finding the needed complete solution of the Hamilton–Jacobi equations in these cases. However, by exploiting the Chern–Simons analogy discussed earlier in this section, a complete solution can be found in $(2+1)$ dimensions and the corresponding complete set of "observables" $(\alpha^A, \beta_A)$ identified. The latter are equivalent, up to a diffeomorphism of the associated reduced phase space, to a complete set of traces of holonomies of the flat $\mathrm{Isom}(\boldsymbol{R}^3_1)$-connections defined in this Chern–Simons formulation (see Carlip (1998) for more details).

## The Reduced Hamiltonian

We continue with the assumption that $M$ is a connected closed oriented $n$-manifold, with either $n=2$ and $M = \boldsymbol{\Sigma}^2_p, p \geq 2$, or $n \geq 3$ and $M$ of negative Yamabe type. We now define the reduced phase space as the set of conformal variables given by

$$P_{\mathrm{reduced}} = \{(\gamma, p^{\mathrm{TT}}) \mid \gamma \in \mathcal{M}_{-1} \text{ and } p^{\mathrm{TT}} \text{ is a 2-contravariant symmetric tensor density that is transverse and traceless with respect to } \gamma\} \qquad [29]$$

We remark that the fully reduced phase space is given by $P_{\mathrm{reduced}}/\mathcal{D}_0$, where $\mathcal{D}_0$ is the group of diffeomorphisms of $M$ isotopic to the identity. However, here, for clarity of exposition, we work on $P_{\mathrm{reduced}}$ rather than the fully reduced phase space.

Given a scalar function $\varphi$, with $\varphi > 0$ if $n \geq 3$, the physical variables $(g, \pi^{\mathrm{TT}})$ are related to the conformal variables $(\gamma, p^{\mathrm{TT}})$ by

$$(g, \pi^{\mathrm{TT}}) = \begin{cases} (\mathrm{e}^{2\varphi}\gamma, \mathrm{e}^{-2\varphi}p^{\mathrm{TT}}) & \text{if } n = 2 \\ (\varphi^{4/(n-2)}\gamma, \varphi^{-4/(n-2)}p^{\mathrm{TT}}) & \text{if } n \geq 3 \end{cases} \qquad [30]$$

We adopt the convention that raising and lowering of indices on either momentum variable $\pi^{\mathrm{TT}}$ or $p^{\mathrm{TT}}$ will be with respect to its own conjugate metric, either $g$ or $\gamma$, respectively. With this convention, the mixed forms of $\pi^{\mathrm{TT}}$ and $p^{\mathrm{TT}}$ are equal, since for $n \geq 3$,

$$(\pi^{\mathrm{TT}})^i{}_j = g_{jl}\pi^{\mathrm{TT}il} = \varphi^{4/(n-2)}\gamma_{jl}\varphi^{-4/(n-2)}p^{\mathrm{TT}il} = \gamma_{jl}p^{\mathrm{TT}il} = (p^{\mathrm{TT}})^i{}_j \qquad [31]$$

(and similarly for the $n=2$ case). Thus the squared norms of $p^{\mathrm{TT}}$ and $\pi^{\mathrm{TT}}$ are equal,

$$p^{\mathrm{TT}} \cdot p^{\mathrm{TT}} = \gamma_{ik}\gamma_{jl}p^{\mathrm{TT}ij}p^{\mathrm{TT}kl} = g_{ik}g_{jl}\pi^{\mathrm{TT}ij}\pi^{\mathrm{TT}kl} = \pi^{\mathrm{TT}} \cdot \pi^{\mathrm{TT}} \qquad [32]$$

where in the first term the center dot is $\gamma$-metric contraction and in the last term the center dot is $g$-metric contraction.

The uniquely determined scalar factor $\varphi$ relating the physical metric $g$ to the conformal metric $\gamma$ is obtained by solving the Hamiltonian constraint

equation [17]. In the special case that $p^{\mathrm{TT}}=0$ (or equivalently, from [30], that $\pi^{\mathrm{TT}}=0$), $\varphi$ is constant and is given in the $n \geq 3$ case by

$$\varphi=\left(\frac{n}{(n-1)\tau^2}\right)^{(n-2)/4} \qquad [33]$$

Thus in this case

$$\gamma=\varphi^{-4/(n-2)}g=\frac{(n-1)}{n}\tau^2 g \qquad [34]$$

In particular, since $\tau$ has the dimension $\ell^{-1}$ (see the remark after [8]) and the components $g_{ij}$ have the dimension $\ell^2$, we see from this formula that the conformal metric $\gamma_{ij}$ is dimensionless. Although $\varphi$ is not constant in the general case when $p^{\mathrm{TT}} \neq 0$, its dimension, as in [33], is still $\ell^{(n-2)/2}$ and thus the components $\gamma_{ij}$ are still dimensionless in the general case. Since in the conventions used in this article, the spatial coordinates are dimensionless, the volume $\mathrm{vol}(M,\gamma)$ of the Riemannian manifold $(M,\gamma)$, as well as all curvature tensors of $\gamma$, are also dimensionless. Having a dimensionless conformal metric $\gamma$ with a dimensionless volume has its advantages over the physical metric $g$ with dimension $\ell^2$ inasmuch, as we shall see below, an infimum of the volume of the conformal metric is related to a dimensionless topological invariant of $M$ (see [48] and the remark thereafter).

If one now uses the conformal variables given by [30] and the decomposition [22] in the ADM action given by [13], one finds the reduced action to be

$$I_{\mathrm{reduced}}=\int_I \mathrm{d}t\int_M\left(p^{\mathrm{TT}ij}\frac{\partial\gamma_{ij}}{\partial t}-\frac{2(n-1)}{n}\frac{\partial\tau}{\partial t}\sqrt{\det g}+\frac{2}{n}\frac{\partial \mathrm{tr}_g\pi}{\partial t}\right)\mathrm{d}^n x \qquad [35]$$

In this expression one can discard the final time derivative which contributes only a boundary integral and so does not contribute to the equations of motion. Moreover, the conformal metric $\gamma_{ij}$ is constrained to lie in the intersection of $\mathcal{M}_{-1}$ and a slice for the action of $\mathcal{D}_0$ on $\mathcal{M}_{-1}$. This space can be regarded as a local chart for the reduced configuration space $\mathcal{T}=\mathcal{M}_{-1}/\mathcal{D}_0$, under the technical assumption that $\mathcal{T}$ is a manifold. Thus, taken together, the conformal variables $(\gamma_{ij}, p^{\mathrm{TT}ij})$ can be viewed as local canonical coordinates for the cotangent bundle $T^*\mathcal{T}$ of Teichmüller space $\mathcal{T}$, where $T^*\mathcal{T}$ now plays the role of the reduced phase space.

For $n=2$, these constructions can be carried out globally for the Teichmüller space $\mathcal{T}_p$ of an arbitrary closed oriented surface $\boldsymbol{\Sigma}_p^2, p \geq 2$ (see the remarks after [4]). Using these global constructions, the reduced phase space $T^*\mathcal{T}_p$ for the $(2+1)$-reduced Einstein equations can be modeled explicitly.

Having restricted the slices to be CMC, one need only choose the relationship between the time coordinate and the CMC $\tau$ in order to fix a corresponding reduced Hamiltonian. The most natural choice of time coordinate from the present point of view is to take

$$t=t(\tau)=\frac{2}{n(-\tau)^{n-1}} \qquad [36]$$

Note that this choice of time coordinate, although also denoted $t$, is no longer dimensionless but has dimension $\ell^{n-1}$.

This choice of time coordinate is motivated by three considerations. Firstly, we remark that since $\tau=0$ is excluded in the setting used in this article (see [23] and the discussion after), $\tau$ can range in either the domain $\boldsymbol{R}^-=(-\infty,0)$ or $\boldsymbol{R}^+=(0,\infty)$. The usual convention on the sign of $k$, as adopted here, is that the sign of $k$ is negative when the tips of the normals on a spacelike hypersurface are further apart than their bases, as for example in the expansion of a model universe, in which case $\tau=\mathrm{tr}_g k<0$. Thus, with this convention, $\tau$ in the range $\boldsymbol{R}^-$ corresponds to an expanding universe and $\tau$ in the range $\boldsymbol{R}^+$ corresponds to a collapsing one in the future direction of increasing $t$. Thus for manifolds of negative Yamabe type that we consider here, the expected maximal range of the CMC $\tau$ is $\boldsymbol{R}^-$ for which $\tau\to-\infty$ corresponds to a "crushing singular" big bang of vanishing spatial volume and $\tau\to 0^-$ corresponds to the limit of infinite volume expansion. Then, with the time function given by [36], the coordinate time $t$ ranges in the interval $\boldsymbol{R}^+$, vanishes at the big bang, and tends to positive infinity in the limit of infinite cosmological expansion.

We remark that to prove that a solution determined by Cauchy data prescribed at some initial coordinate time $t_0 \in \boldsymbol{R}^+$ actually exhausts the range $\boldsymbol{R}^+$ is a difficult global existence problem that is not dealt with here. Nevertheless, one of the main motivations for this work is the hope that Hamiltonian reduction will lead to advances in the study of the global existence question for Einstein's equations.

We also remark that with the choice of temporal gauge function given by [36] and with $\tau$ in its natural range $\boldsymbol{R}^-$,

$$\frac{\mathrm{d}\tau}{dt}=\frac{n}{2(n-1)}(-\tau)^n>0 \qquad [37]$$

so that this temporal coordinate choice preserves the time orientation of the flow for all $n \geq 2$.

Secondly, with this choice of temporal gauge, the reduced action given by [35] simplifies to

$$I_{\text{reduced}} = \int_I \mathrm{d}t \int_M \left( p^{\mathrm{TT}ij} \frac{\partial \gamma_{ij}}{\partial t} - (-\tau)^n \sqrt{\det g} \right) \mathrm{d}^n x \qquad [38]$$

from which one can read off an effective reduced Hamiltonian density,

$$\mathcal{H}_{\text{reduced}}(\tau, \gamma, p^{\mathrm{TT}}) = (-\tau)^n \sqrt{\det g} \qquad [39]$$

and an effective reduced Hamiltonian,

$$\begin{aligned} H_{\text{reduced}}(\tau, \gamma, p^{\mathrm{TT}}) &= \int_M (-\tau)^n \sqrt{\det g}\, \mathrm{d}^n x \\ &= (-\tau)^n \int_M \mathrm{d}\mu_g \\ &= (-\tau)^n \mathrm{vol}(M, g) \end{aligned} \qquad [40]$$

where $\mathrm{vol}(M,g) = \int_M \mathrm{d}\mu_g$ is the volume of the Riemannian manifold $(M, g)$. Thus in terms of the physical variables $(g_{ij}, \pi^{ij})$, the reduced Hamiltonian $H_{\text{reduced}}$ at "time" $\tau$ is simply the volume of the CMC slice with mean curvature $\tau$ rescaled by the factor $(-\tau)^n$. With this reduced Hamiltonian density, the reduced action [38] takes the canonical form

$$I_{\text{reduced}} = \int_I \mathrm{d}t \int_M \left( p^{\mathrm{TT}ij} \frac{\partial \gamma_{ij}}{\partial t} - \mathcal{H}_{\text{reduced}} \right) \mathrm{d}^n x \qquad [41]$$

As the third consideration for the given choice of the time function, we note that rescaling the physical volume $\mathrm{vol}(M, g)$ by the factor $(-\tau)^n$ yields a dimensionless quantity. Indeed, as we have seen, the spatial physical volume has the dimension $\ell^n$ and the constant mean curvature $\tau$ has the dimension $\ell^{-1}$, so that the reduced Hamiltonian $(-\tau)^n \mathrm{vol}(M, g)$ is dimensionless.

The main advantage of having a dimensionless reduced Hamiltonian is that only such a reduced Hamiltonian can have a topological significance, and indeed, the infimum of $H_{\text{reduced}}$ is closely related to a dimensionless topological invariant of $M$ (see the remarks after [48]).

In terms of the conformal variables $(\gamma, p^{\mathrm{TT}})$, the reduced Hamiltonian is found from [21] and [40] to be given for $n \geq 3$ by

$$\begin{aligned} H_{\text{reduced}}(\tau, \gamma, p^{\mathrm{TT}}) &= (-\tau)^n \int_M \sqrt{\det g}\, d^n x \\ &= (-\tau)^n \int_M \sqrt{\det(\varphi^{4/(n-2)}\gamma)}\, \mathrm{d}^n x \\ &= (-\tau)^n \int_M (\varphi^{4/(n-2)})^{n/2} \sqrt{\det \gamma}\, d^n x \\ &= (-\tau)^n \int_M \varphi^{2n/(n-2)}\, \mathrm{d}\mu_\gamma \end{aligned} \qquad [42]$$

where $\mathrm{d}\mu_\gamma$ is the Riemannian measure on $M$ determined by $\gamma$ (locally, $\mathrm{d}\mu_\gamma = \sqrt{\det \gamma}\, \mathrm{d}^n x$) and $\varphi = \varphi(\tau, \gamma, p^{\mathrm{TT}})$ is the conformal factor which, through the solution of the Hamiltonian constraint [17], is expressed as a function of the "time" $\tau$ and the independent conformal (or canonical) variables $(\gamma, p^{\mathrm{TT}})$.

In the special case $n = 2, M = \boldsymbol{\Sigma}_p^2, p \geq 2$, a simple formula for $H_{\text{reduced}}$ can be derived. In terms of the conformal variables $(\gamma, p^{\mathrm{TT}})$, we find from [40], [10], [14], [17], [21], [22], and [32] that

$$\begin{aligned} &H_{\text{reduced}}(\tau, \gamma, p^{\mathrm{TT}}) \\ &= \int_{\Sigma_p^2} (-\tau)^2 \mathrm{d}\mu_g = 2 \int_{\Sigma_p^2} \left( (\det g)^{-1} (\pi^{\mathrm{TT}} \cdot \pi^{\mathrm{TT}}) - R(g) \right) \mathrm{d}\mu_g \\ &= 2 \int_{\Sigma_p^2} (\det(\mathrm{e}^{2\varphi}\gamma))^{-1} (p^{\mathrm{TT}} \cdot p^{\mathrm{TT}})\, \mathrm{d}\mu_{(\mathrm{e}^{2\varphi}\gamma)} - 2 \int_{\Sigma_p^2} R(g)\, \mathrm{d}\mu_g \\ &= 2 \int_{\Sigma_p^2} (\mathrm{e}^{2\varphi})^{-2} (\det \gamma)^{-1} (p^{\mathrm{TT}} \cdot p^{\mathrm{TT}})\, \mathrm{e}^{2\varphi} \mathrm{d}\mu_\gamma - 8\pi \chi(\boldsymbol{\Sigma}_p^2) \\ &= 2 \int_{\Sigma_p^2} \mathrm{e}^{-2\varphi} (\det \gamma)^{-1} (p^{\mathrm{TT}} \cdot p^{\mathrm{TT}}) \mathrm{d}\mu_\gamma + 16\pi(p-1) \end{aligned} \qquad [43]$$

where $\varphi = \varphi(\tau, \gamma, p^{\mathrm{TT}}), \chi(\boldsymbol{\Sigma}_p^2) = 2(1-p)$ is the Euler characteristic of the genus $p$ surface $\boldsymbol{\Sigma}_p^2$, and where we have used the Gauss–Bonnet theorem

$$\int_{\Sigma_p^2} R(g)\, \mathrm{d}\mu_g = 4\pi \chi(\boldsymbol{\Sigma}_p^2) = 8\pi(1-p) \qquad [44]$$

Since

$$\begin{aligned} &H_{\text{reduced}}(\tau, \gamma, p^{\mathrm{TT}}) \\ &\quad = 2 \int_{\Sigma_p^2} \mathrm{e}^{-2\varphi} (\det \gamma)^{-1} (p^{\mathrm{TT}} \cdot p^{\mathrm{TT}}) \mathrm{d}\mu_\gamma \\ &\qquad + 16\pi(p-1) \geq 16\pi(p-1) \end{aligned} \qquad [45]$$

the infimum of $H_{\text{reduced}}$ is attained precisely when $p^{\mathrm{TT}} = 0$ and this infimum coincides with the topological invariant $-8\pi\chi(\boldsymbol{\Sigma}_p^2) = 16\pi(p-1)$, which characterizes the surface $\boldsymbol{\Sigma}_p^2$ (see also [51] below). As we shall see shortly, an analogous result holds for $n \geq 3$.

A straightforward but lengthy calculation, which is valid in arbitrary dimensions, shows that the reduced Hamiltonian is strictly monotonically decreasing in the direction of cosmological expansion except for a family of continuously self-similar spacetimes for which this Hamiltonian is constant (Fischer and Moncrief 2002b). The latter solutions exist if and only if $M$ admits a Riemannian metric $\gamma \in \mathcal{M}_{-1}$ which is an Einstein metric, that is, for which the Ricci tensor satisfies $\mathrm{Ric}(\gamma) = -(1/n)\gamma$. Using the mean curvature as a convenient time coordinate, that is, temporarily taking $t = \tau$, the

corresponding self-similar vacuum spacetime metrics then have the line element

$$ds^2 = -\left(\frac{n}{\tau^2}\right)^2 d\tau^2 + \frac{n}{(n-1)}\frac{1}{\tau^2}\gamma_{ij}dx^i\,dx^j \qquad [46]$$

In the case that $n = 3$, the Einstein metric $\gamma$ is actually hyperbolic with constant sectional curvature $K(\gamma) = -1/6$ and Ricci curvature $\mathrm{Ric}(\gamma) = -(1/3)\gamma$. Although the conformal variables $(\gamma, p^{\mathrm{TT}}) = (\gamma, 0)$ are static in this model, the physical variables $(g, \pi)$ are not. In this case, the resulting spacetimes (which depend on the underlying topology of $M$) have expanding closed hyperbolic spacelike hypersurfaces where the physical volume $\mathrm{vol}(M, g)$ "starts" at zero at the big bang and expands to infinity in the forward time direction, as befits a universe endlessly expanding from the big bang. Such a universe is depicted in Figure 1, where the genus-2 surface is used to represent a generic closed hyperbolic 3-manifold. The Bianchi and Thurston types of this model are discussed in the next section.

The line element [46] is locally isometric to the vacuum Friedmann–Lemaître–Robertson–Walker (FLRW) $k = -1$ spacetime, which is well known to be flat. Although these spatially compactified models are technically not classical FLRW spacetimes since the expanding compact hypersurfaces are not homogeneous (and thus not isotropic), they are Lorentz-covered by the FLRW $k = -1$ spacetime and thus are locally isometric to this classical spacetime.

The same result leading to [46] holds even if matter sources are allowed, provided they satisfy a suitable energy condition, in which case the corresponding reduced Hamiltonian will only be stationary in the vacuum limit and then only when the metric is of the above type; otherwise it monotonically decays. This result even has a quasilocal

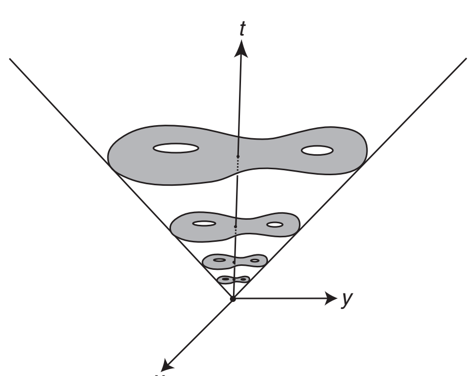



**Figure 1** Expansion of the physical universe in the Bianchi V, Thurston type $H^3$, spatially compactified FLRW flat spacetime cosmology.

generalization expressible in terms of the corresponding quasilocal reduced Hamiltonian defined for an arbitrary domain $D_\tau$ within the CMC slice $\tau = \text{constant}$ by restricting $H_{\text{reduced}}$ in [42] to the domain $D_\tau$, so that for $n \geq 3$,

$$\begin{aligned} H_{D_\tau}(\tau, \gamma, p^{\mathrm{TT}}) &= (-\tau)^n \int_{D_\tau} d\mu_g \\ &= (-\tau)^n \int_{D_\tau} \varphi^{2n/(n-2)}\, d\mu_\gamma \qquad [47] \end{aligned}$$

If $D_\tau$ is determined from its specification on some initial slice $\tau = \tau_0$, by letting the domain flow along the normal trajectories of the CMC foliation, one can then verify that $H_{D_\tau}$ is monotonically decreasing except for the vacuum solutions of self-similar type described above, in which case $H_{D_\tau}$ is constant. This result is independent of the initial domain chosen.

We remark that one cannot use the quasilocal Hamiltonian to get equations of motion (even quasilocally) since the full true Hamiltonian is nonlocal and so one gets contributions from the whole manifold.

Since the reduced Hamiltonian $H_{\text{reduced}}$ as well as its quasilocal variant $H_{D_\tau}$ is monotonically decreasing for generic solutions of Einstein's equations, it is natural to ask what its infimum is and whether this infimum is ever attained, at least asymptotically, by solutions of the field equations. The infimum of the reduced Hamiltonian for $n \geq 3$ and for a spatial manifold $M$ of negative Yamabe type can be characterized in terms of a certain topological invariant of $M$ called the sigma constant $\sigma(M)$ of $M$. For manifolds of negative Yamabe type, this quantity can be defined in terms of the infimum of the volume of all metrics which range over the space of conformal metrics $\mathcal{M}_{-1}$. The precise definition leads to the formula

$$\sigma(M) = -\left(\inf_{\gamma \in \mathcal{M}_{-1}} \mathrm{vol}(M, \gamma)\right)^{2/n} \qquad [48]$$

Interestingly, this equation defines the topological invariant $\sigma(M)$ by a purely geometrical equation involving the volume functional restricted to $\mathcal{M}_{-1}$. We also remark that [48] is a dimensionless equation, the left-hand side being dimensionless since it is a topological invariant of $M$ and the right-hand side being dimensionless since the conformal metric and its volume are dimensionless (see the remarks after [34]).

Although the $\sigma$-constant can be defined for all Yamabe types, [48] holds only for manifolds of negative Yamabe type. From this equation, one can conclude that for such manifolds

$$\sigma(M) \leq 0 \qquad [49]$$

One can relate the foregoing to the reduced Hamiltonian by showing that the infimum of $H_{\text{reduced}}$ defined for arbitrary $\tau < 0$ as a functional on the reduced phase space

$$T^*\mathcal{T} = T^*\left(\frac{\mathcal{M}_{-1}}{\mathcal{D}_0}\right) \qquad [50]$$

is given by

$$\inf_{(\gamma, p^{\text{TT}}) \in T^*\mathcal{T}} H_{\text{reduced}}(\tau, \gamma, p^{\text{TT}}) = \begin{cases} -8\pi\chi(\boldsymbol{\Sigma}_p^2) = 16\pi(p-1) & \text{if } n=2 \text{ and } p \geq 2 \\ \left(\frac{n}{n-1}(-\sigma(M))\right)^{n/2} & \text{if } n \geq 3 \end{cases} \qquad [51]$$

where for $n \geq 3, M$ is of negative Yamabe type and thus $\sigma(M) \leq 0$ (see [49]).

One proves this result by first showing that within an arbitrary fiber of the cotangent bundle $T^*(\mathcal{M}_{-1}/\mathcal{D}_0)$, one minimizes $H_{\text{reduced}}$ by setting the fiber variable $p^{\text{TT}}$ to zero. In this case, the solution for the conformal factor $\varphi$ reduces to a spatial constant which is a function of $\tau$ alone (see [33]), and thus the formula for $H_{\text{reduced}}$ given in [42] reduces to

$$H_{\text{reduced}}(\tau, \gamma, 0) = \left(\frac{n}{n-1}\right)^{n/2} \text{vol}(M, \gamma) \qquad [52]$$

The infimum over all conformal metrics $\gamma \in \mathcal{M}_{-1}$ of this latter functional yields the $\sigma$-constant as outlined above. If matter sources obeying a suitable energy condition are allowed, the argument goes through in much the same way with the additional implication that the infimum is achieved only for a vacuum solution so that in fact the matter must be "turned off."

Thus, as a consequence of the above analysis, one has

$$\begin{aligned} &H_{\text{reduced}}(\tau, \gamma, p^{\text{TT}}) \\ &\geq H_{\text{reduced}}(\tau, \gamma, 0) = \left(\frac{n}{n-1}\right)^{n/2} \text{vol}(M, \gamma) \\ &\geq \left(\frac{n}{n-1}\right)^{n/2} \inf_{\gamma' \in \mathcal{M}_{-1}} \text{vol}(M, \gamma') \\ &= \left(\frac{n}{n-1}(-\sigma(M))\right)^{n/2} \end{aligned} \qquad [53]$$

where the last equality follows by inverting [48] to give

$$\inf_{\gamma \in \mathcal{M}_{-1}} \text{vol}(M, \gamma) = (-\sigma(M))^{n/2} \qquad [54]$$

Moreover, if $\gamma \in \mathcal{M}_{-1}$ actually achieves the $\sigma$-constant, that is, if $\text{vol}(M, \gamma) = (-\sigma(M))^{n/2}$ (and not just asymptotically approaches it as a curve or sequence), then $\gamma$ must be an Einstein metric with

$$\text{Ric}(\gamma) = -\frac{1}{n}\gamma \qquad [55]$$

If, additionally, $n = 3$, then $\gamma$ must be hyperbolic (with constant sectional curvature $K(\gamma) = -1/6$).

Although Thurston's conjectures do not refer to the $\sigma$-constant, Anderson (1997) has been able to reformulate and somewhat refine the Thurston geometrization conjectures for 3-manifolds of arbitrary Yamabe type in terms of conjectured properties of the $\sigma$-constant. Additionally, if Perelman's results are technically complete, they would provide a proof of Anderson's conjectures as well as those of Thurston's (see Anderson (2003)).

The conjectured behavior for a sequence of conformal metrics $\{\gamma_i\}, \gamma_i \in \mathcal{M}_{-1}, i = 1, 2, \ldots,$ which seeks to minimize the volume of a stand-alone $K(\boldsymbol{\pi}, 1)$ 3-manifold $M$ of negative Yamabe type can be described as follows:

1. If $M$ is hyperbolizable, then $\sigma(M) < 0$ is attained by a hyperbolic metric $\gamma_h \in \mathcal{M}_{-1}$, unique up to diffeomorphism, and the sequence of conformal metrics $\{\gamma_i\}$ converges to this metric in a suitable function space topology.
2. If $M$ is a pure graph manifold, then $\sigma(M) = 0$ and the sequence $\{\gamma_i\}$ of conformal metrics "volume collapses" $M$ with bounded curvature. Typically this occurs through collapse of circular or toroidal fibers in the associated circle or 2-torus bundle structure (see examples 3, 4, and 5 in the section "Topological Background" and see also the penultimate section). The six manifolds of flat type are not included here as they are of zero Yamabe type.
3. If $M$ is a generic $K(\boldsymbol{\pi}, 1)$-manifold (not of type 1 or 2 above), then $M$ can be decomposed along incompressible tori into its final finite-volume-type hyperbolizable and (possibly empty set of) graph-manifold pieces. In this case, $\sigma(M) < 0$ and the sequence $\{\gamma_i\}$ of conformal metrics collapses the graph-manifold components and converges to finite-volume complete hyperbolic metrics on the hyperbolizable components (normalized to have $R(\gamma) = -1$) yielding a $\sigma$-constant that is entirely determined by the volumes of these final hyperbolic components (see the final section).

We shall return to this conjectured characterization of sequences of conformal metrics in the next two sections.

## Reduction of Bianchi Models and Conformal Volume Collapse

For manifolds of negative Yamabe type, the strict monotonic decay of $H_{\text{reduced}}$ in the direction of cosmological expansion along nonconstant integral curves of the reduced Einstein equations suggests

that the reduced Hamiltonian is seeking to achieve its infimum $\inf H_{\text{reduced}} = ((n/(n-1))(-\sigma(M)))^{n/2}$. But does this ever happen? Does the reduced Einstein flow of the conformal geometry asymptotically approach $\inf H_{\text{reduced}}$ in the limit of infinite cosmological expansion?

To answer this question, one can consider for $n = 3$ known locally homogeneous vacuum solutions of Einstein's equations which spatially compactify to manifolds of negative Yamabe type. Applying the theory of Hamiltonian reduction to these classical models, one can show that the reduced Hamiltonian behaves as expected under the reduced Einstein flow defined by these models. Since these models existed long before this theory, it is somewhat satisfying to see that they can be interpreted in terms of Hamiltonian reduction and how, with this interpretation, new properties of these classical solutions can be found.

Since $H_{\text{reduced}}$ is a strictly monotonically decreasing function along nonconstant integral curves of the reduced Einstein flow, it is expected that under certain conditions, the reduced Hamiltonian is monotonically seeking to decay to its infimum. Thus, it is of interest to look at Hamiltonian reduction under the consequence of the following two assumptions:

1. The reduced Einstein field equations give rise to the existence of a positive semiglobal nonconstant solution $(\gamma(t), p^{\text{TT}}(t))$ defined for all $t \in (0, \infty)$ (or equivalently, for all $\tau \in (-\infty, 0)$;
2. The reduced Hamiltonian strictly monotonically decays to its infimum along nonconstant integral curves,

$$H_{\text{reduced}}(\tau(t), \gamma(t), p^{\text{TT}}(t)) \longrightarrow \inf H_{\text{reduced}} \quad \text{as } t \longrightarrow \infty \qquad [56]$$

From [40] and [51], in terms of the physical variables $(g, \pi)$ (or $(g, k)$), [56] can be written equivalently as

$$-\tau^3 \operatorname{vol}(M,g) = -(\operatorname{tr}_g k)^3 \operatorname{vol}(M,g) \longrightarrow \left(\frac{3}{2}(-\sigma(M))\right)^{3/2} \quad \text{as } t \longrightarrow \infty \qquad [57]$$

As a consequence of these assumptions, it follows from [53] that the conformal volume $\operatorname{vol}(M,\gamma)$ must also decay to its infimum [54] (although not necessarily monotonically),

$$\operatorname{vol}(M, \gamma(t)) \longrightarrow \inf_{\gamma \in \mathcal{M}_{-1}} \operatorname{vol}(M, \gamma) = (-\sigma(M))^{3/2} \quad \text{as } t \longrightarrow \infty \qquad [58]$$

Now suppose that $\sigma(M) = 0$. A large class of manifolds for which this is true are the graph manifolds (and thus also the Seifert manifolds) of negative Yamabe type since $\sigma(M) \geq 0$ for graph manifolds in general and since $\sigma(M) \leq 0$ for manifolds of negative Yamabe type. In this case the curve $\gamma(t) \in \mathcal{M}_{-1}$ of conformal metrics must necessarily (conformally) volume collapse $M$ in the direction of cosmological expansion,

$$\operatorname{vol}(M, \gamma(t)) \longrightarrow (-\sigma(M))^{3/2} = 0 \quad \text{as } t \longrightarrow \infty \qquad [59]$$

Consequently, the curve of conformal metrics $\gamma(t)$ must undergo some form of degeneration as its volume collapses. The details of this metric degeneration are of importance and are discussed below.

Not all locally homogeneous vacuum Bianchi models admit spatially compact quotients. Fortunately, the general theory of which Bianchi models admit spatially compact quotients has been worked out in detail by Tanimoto, Koike, and Hosoya (see Tanimoto *et al.* (1997) and the references therein). These Bianchi models together with their corresponding Thurston classification and typical examples of their closed quotient manifolds are listed in **Table 1**, where "K–S" indicates "Kantowski–Sachs," "P," "Z," and "N" denote manifolds of Yamabe type positive, zero, and negative, respectively (see the section "Topological Background"), "Seifert" means Seifert fibered, "Hyper" means hyperbolizable, "?" indicates "unknown, but conjectured to be so," and "manifold collapse" denotes the type of collapse that the conformal manifold $(M, \gamma(t))$ goes through as the conformal volume $\operatorname{vol}(M, \gamma(t))$ collapses. We also remark that all of the manifolds

**Table 1** Bianchi, Thurston, and Yamabe type of a connected closed oriented irreducible 3-manifold

| *Bianchi type* | *Thurston type* | *Typical examples* | *Yamabe type* | *σ-constant* | *Manifold structure* | *Manifold collapse* |
|---|---|---|---|---|---|---|
| K–S | $S^2 \times R$ | $S^2 \times S^1$ | $P$ | $>0$ | Seifert | |
| IX | $S^3$ | Nontrivial $S^1$-bundles over $S^2$ | $P$ | $>0$ | Seifert | |
| I | $R^3$ | $T^3$ | $Z$ | 0 | Seifert | |
| II | Nil | Nontrivial $S^1$-bundles over $T^2$ | $N$ | 0 | Seifert | Total |
| III | $H^2 \times R$ | $\Sigma_p^2 \times S^1, p \geq 2$ | $N$ | 0 | Seifert | Pancake |
| VIII | $\widetilde{SL}(2, R)$ | Nontrivial $S^1$-bundles over $\Sigma_p^2$ | $N$ | 0 | Seifert | Pancake |
| $VI_0$ | Sol | Nontrivial $T^2$-bundles over $S^1$ | $N$ | 0 | Graph | Barrel |
| V, $VII_h$ | $H^3$ | Closed hyperbolizable manifolds | $N$ | $<0$ ? | Hyper | None |

listed in the "Typical examples" column are irreducible with the exception of $S^1 \times S^2$, which is prime but not irreducible. Also, in this column, $p \geq 2$.

In this table, the eight Thurston types are grouped into three sets according to their Yamabe type. The first set of such Bianchi models are those that spatially compactify to yield 3-manifolds of positive Yamabe type which allow metrics with positive constant scalar curvature, for example, Bianchi IX models defined over spherical space forms. The second set (consisting of one type) yields manifolds of zero Yamabe type which allow zero scalar curvature metrics but not constant positive scalar curvature metrics, for example, Bianchi I models defined over $T^3$ or one of the other five manifolds of flat type finitely covered by $T^3$. The third set (the last five entries in Table 1 and the set of most interest in this article) yields manifolds of negative Yamabe type which do not allow metrics with zero scalar curvature.

These latter models are the five Bianchi models of types II, III, VIII, $\mathrm{VI}_0$, and V (and in part $\mathrm{VII}_h$), which in turn correspond in Thurston's classification to manifolds of type Nil, $H^2 \times R, \widetilde{\mathrm{SL}(2, R)}$, Sol, and $H^3$, respectively.

In the first three cases, the models of Bianchi type II, III, and VIII compactify to a nontrivial $S^1$-bundle over $T^2$ or to a trivial or nontrivial $S^1$-bundle over $\Sigma_p^2, p \geq 2$, respectively. Each of these spaces is Seifert fibered. In the fourth case, the model of Bianchi type $\mathrm{VI}_0$ compactifies to a nontrivial $T^2$-bundle over $S^1$ which is an irreducible graph manifold. Since each of these manifolds is also of negative Yamabe type, in each of these four cases, as discussed in the beginning of this section, $\sigma(M) = 0$. In the fifth case, we consider vacuum Bianchi V metrics as well as a special case of Bianchi type $\mathrm{VII_h}$ which compactify to an arbitrary closed oriented hyperbolizable manifold $M$.

For these latter five Bianchi models that spatially compactify to manifolds of negative Yamabe type, one can consider the classical solutions from the point of view of Hamiltonian reduction. The starting point for this point of view is to use explicitly known vacuum metrics for the simplest "standard" metric forms, given, for example, in Wainwright and Ellis (1997). One need not consider all such possible spatially compact quotients, even though that would appear to be quite feasible, but one need only consider some representative examples for each of the Bianchi types listed.

It can be shown by explicit calculation, using the known solutions, that in the four nonhyperbolizable cases where $\sigma(M) = 0$, each of the classical Bianchi solutions gives rise to the existence of a positive semiglobal nonconstant solution to the reduced Einstein field equations and that along this solution, the reduced Hamiltonian asymptotically approaches 0 under the reduced Einstein flow, thereby confirming the expectation that the reduced Hamiltonian asymptotically approaches its infimum $((3/2)(-\sigma(M)))^{3/2} = 0$. Thus in these cases the reduced Einstein flow conformally volume-collapses the 3-manifold.

The explicit calculations also show the details of this collapse. In the second and third models of Bianchi type III, Thurston type $H^2 \times R$, and Bianchi type VIII, Thurston type $\widetilde{\mathrm{SL}(2, R)}$, respectively, the conformal metric degenerates along embedded circular fibers and this metric degeneration causes $M$ to collapse to its base manifold $\Sigma_p^2$, $p \geq 2$. Since the collapse is along one-dimensional fibers and since the two-dimensional base manifold $\Sigma_p^2$ does not collapse, we refer to this type of collapse as pancake collapse (see Figure 2).

In the fourth model of Bianchi type $\mathrm{VI}_0$, Thurston type Sol, the conformal metric degenerates along embedded $T^2$-fibers and this metric degeneration causes $M$ to collapse to its base manifold $S^1$. Since the collapse is along two-dimensional fibers and since the one-dimensional base manifold $S^1$ does not collapse, we refer to this type of collapse as barrel collapse (see Figure 3).

In the first model of Bianchi type II, Thurston type Nil, as in the second and third models, the conformal metric degenerates along embedded circular fibers. Additionally, not only do the circular fibers collapse but simultaneously the flat quotient 2-torus base manifold $T^2 \simeq M/S^1$ of $M$ modulo its circular fibers also collapses. Thus the metric degeneration collapses $M$ to a point, exhibiting a

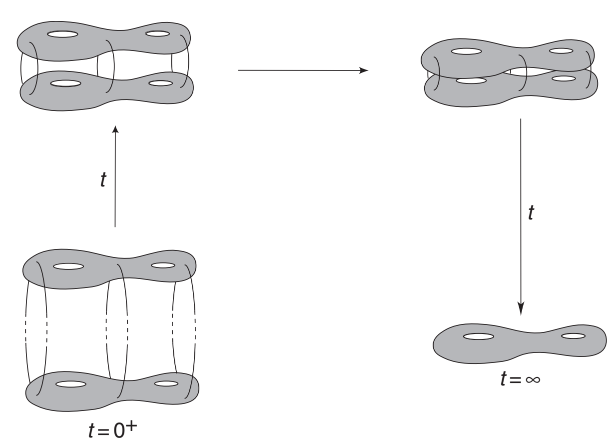



**Figure 2** Bianchi III, Thurston type $H^2 \times R, M = \Sigma_p^2 \times S^1$, pancake collapses to $\Sigma_p^2, p = 2$. The conformal geometry starts with an infinite $S^1$-fiber at the big bang ($t = 0^+$) and pancake collapses with bounded curvature to $\Sigma_2^2$ at infinite cosmological expansion ($t \rightarrow \infty$).

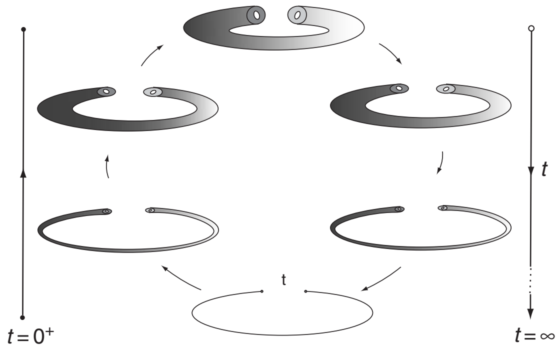



**Figure 3** Bianchi $VI_0$, Thurston type Sol, nontrivial $T^2$-bundle over $S^1$, barrel collapses to $S^1$. The conformal geometry evolves from a base manifold $S^1$ at the big bang ($t=0^+$). Instantaneously after the big bang, flat $T^2$-fibers bloom out of the collapsed $S^1$ state. The conformal metric then expands to a maximum volume and then barrel collapses with bounded curvature back to the base manifold $S^1$ at infinite physical cosmological expansion ($t \to \infty$). The two facial 2-tori are flat and are glued together by an orientation-reversing toral automorphism so as to give a nontrivial $T^2$-bundle over $S^1$. The gray-scale density grading along the tube also indicates the nontriviality of the bundle.

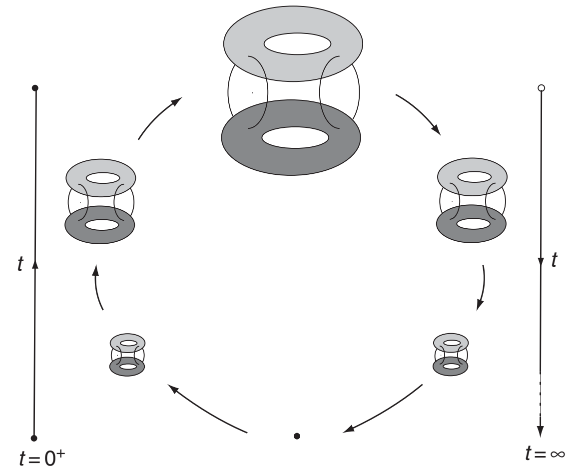



**Figure 4** Bianchi II, Thurston type Nil, nontrivial $S^1$-bundle over $T^2$, totally collapses to a point. The conformal geometry evolves from a point at the big bang ($t=0^+$). Instantaneously after the big bang, the full 3-manifold blooms out from that point. The conformal geometry then evolves to a metric of maximum volume and then totally collapses with bounded curvature back to a point at infinite physical cosmological expansion ($t \to \infty$). The two 2-tori, represented here by doughnuts, are flat and are glued together by an orientation-reversing toral automorphism so as to give a nontrivial $S^1$-bundle over $T^2$.

case of total collapse. Thus these model universes provide examples of nonflat almost-flat manifolds that exhibit total collapse with bounded curvature. Since the conformal geometries of these model universes collapse to a point, they aptly deserve their name Nil (see Figure 4).

Remarkably, in each of these four cases of collapse, the collapse occurs with bounded curvature, precisely as occurs in the totally different setting of the Cheeger–Gromov theory of collapsing Riemannian manifolds, recognized many years ago to be of importance in the understanding of the behavior of sequences of metrics with uniform curvature bound (see Gromov (1999) for references and Anderson (2004) for other applications of Cheeger–Gromov theory to general relativity). What is somewhat remarkable is that the above cosmological models were constructed completely independently of that setting and thus provide naturally occurring cosmological models whose closed spatial hypersurfaces undergo conformal volume collapse and metric degeneration exactly as occurs in the theory of collapsing Riemannian manifolds.

Of course, this volume collapse and metric degeneration only occur as described in the conformal variables. The physical variables behave differently. Indeed, in contrast to the conformal volume which collapses to zero in the first four cases and is constant in the hyperbolizable case (see below), the volume of the physical metric in all five cases goes to infinity since the flow is temporally oriented in the direction of infinite cosmological expansion.

In the fifth case where $M$ is hyperbolizable, $\sigma(M)$ is conjectured to be negative and to be determined by the hyperbolic volume, $\sigma(M) = -(\mathrm{vol}(M, \gamma_h))^{2/3}$, of the hyperbolic conformal metric $\gamma_h$ normalized so that $R(\gamma_h) = -1$. In this case, $\gamma_h$ together with $p^{\mathrm{TT}} = 0$ is a fixed point for the reduced Einstein flow so that trivially the conformal volume does not collapse. Moreover, if $\sigma(M)$ is determined by the volume of $\gamma_h$, then the constant reduced Hamiltonian also trivially achieves its infimum $H_{\mathrm{reduced}}(\tau, \gamma_h, 0) = ((3/2)\ (-\sigma(M)))^{3/2} = (3/2)^{3/2}$ (vol $(M, \gamma_h)$), again confirming the expectation for the behavior of $H_{\mathrm{reduced}}$ on these Bianchi models.

Note that for this static case, the physical variables behave as described after [46] and as shown in Figure 1. Also note that in contrast to Figures 2–4 where the conformal geometry is depicted, Figure 1 depicts the physical geometry.

Overall, in all five cases, subject in the hyperbolizable case to a hyperbolic metric realizing the $\sigma$-constant, the reduced Hamiltonian asymptotically approaches its $\sigma$-constant infimum along the flow lines of the reduced Einstein system. In doing so, the volumes of the conformal metrics either go to zero (in the first four cases) or to the hyperbolic volume (in the hyperbolic case). In all five cases, the curvature of the conformal metrics is uniformly bounded.

Because the reduced Einstein field equations behave as expected for the Bianchi models that we have considered with spatially compactified manifolds being either Seifert fibered, graph, or hyperbolizable, it seems plausible that for a more complicated starting manifold $M$, the reduced Einstein flow may induce a decomposition of $M$ into geometric pieces. Indeed, Anderson's conjectures (Anderson 1997) predict how a sequence of geometries with bounded curvature approaching $\sigma(M)$ degenerate. Assuming these conjectures, the asymptotic behavior of large classes of Einstein spacetimes may perhaps be characterized rather explicitly in terms of the geometrization program of 3-manifolds (see the next section).

Conversely, it is conceivable that the damped hyperbolic system of equations defined by the reduced Einstein flow (with its strictly monotonically decreasing reduced Hamiltonian on nonconstant curves) could be used to try to establish some form of the geometrization conjectures for 3-manifolds, much like the parabolic system of equations defined by Ricci flow is currently being used. If such a program were to be successful, it would amount to a spectacular consequence of Einstein's equations, implying as it does that geometrization may actually occur in nature.

## Possible Cosmological Applications of the Reduced Hamiltonian

Astronomical observations strongly support the view that in a sufficiently coarse-grained sense, the universe is homogeneous and isotropic. Furthermore, it is expanding at such a rate, relative to its observable energy density, that it will continue to expand forever. The simplest cosmological model consistent with these properties and which has a vacuum limit is the $k=-1$ FLRW model. Spatially compactified variants of this model are still locally homogeneous and isotropic even though they are no longer globally so (see the discussion after [46]). Evidence for one or another of the infinitely many compactifications possible could be sought in patterns of fluctuations of the cosmic microwave background radiation and the detection of such patterns could be strong evidence for a spatially closed universe.

However, is one really justified in extrapolating local observations of that portion of the universe visible to astronomers to a conclusion about its global topology? Could it be instead that there is a dynamical reason, provided by Einstein's equations, for the observed fact that the universe seems to be locally homogeneous and isotropic and in such a state as to continue expanding forever?

Suppose for the sake of argument that the universe has a more complicated topology, such as that of one of the generic $K(\boldsymbol{\pi},1)$-manifolds which does not admit a locally homogeneous and isotropic metric even though its hyperbolizable components would each individually do so. A plausible scenario suggested by the results in this article is that under the Einstein evolution, the reduced Hamiltonian given by [40] consisting of the rescaled spatial volume becomes asymptotically dominated in the future direction of cosmological expansion by the contribution of the hyperbolizable components. On each of these components, the limiting conformal metric approaches local homogeneity and isotropy with the relative contribution of the graph-manifold constituents, if any are present, collapsing asymptotically to a negligible fraction of the whole. The idea is that if structure formation develops sufficiently late in the evolution of such a universe, then it should occur, with overwhelmingly high probability, in those regions which dominate the conformal volume and admit an asymptotically locally homogeneous and isotropic metric of constant negative curvature, locally indistinguishable from a $k=-1$ FLRW model.

One can speculate still further and imagine what happens if the spatial topology is not of prime type but rather consists of a connected sum of several $K(\boldsymbol{\pi},1)$'s together perhaps with nontrivial spherical manifolds $S^3/\Gamma$ and handles $S^1\times S^2$. Here it seems conceivable, especially in view of the expected tendency of spherical manifolds to "recollapse," that the evolving universe would develop pinch-off singularities along the essential 2-spheres that separate the individual prime factors. Such singularities might occur in finite time between connected sums of spherical recollapsing factors or in infinite time between connected sums of $K(\boldsymbol{\pi},1)$-factors. Similar patterns of singularity formation are seen to occur in Ricci flow and must be treated in the resolution of the 3-manifold geometrization program.

Of course there is no proof of such behavior for the full $(3+1)$-dimensional Einstein gravity but for the model problem of Einstein's theory in $(2+1)$ dimensions, something close to a proof of the analogous conjecture is already at hand. In the vacuum case, which can be described rather explicitly, one can construct the generic solution for a higher genus surface topology by cutting open the corresponding $k=-1$ FLRW model and gluing in the so-called Kazner wedges. These wedges play the role of the graph-manifold constituents of a generic $K(\boldsymbol{\pi},1)$-manifold in three dimensions and evolve anisotropically. However, it is known rigorously in this case that the rescaled spatial area $H_{\mathrm{reduced}}(\tau,\gamma,p^{\mathrm{TT}})=(-\tau)^2\mathrm{Area}(\Sigma_p^2,g)$ is asymptotically exhausted by the FLRW components with the contribution from the flat Kazner anisotropic pieces shrinking to zero in this

limit. If certain types of matter sources are included, for example, those analogous to terms which result from Kaluza–Klein reduction of vacuum gravity in $(3+1)$-dimensions, then a similar result can be proved at least for sufficiently small but fully nonlinear perturbations away from the vacuum backgrounds (see Choquet-Bruhat (2004)).

In fully general $(3+1)$-dimensional gravity, there are few known topologically general results beyond those mentioned earlier and the problem is complicated by the presence of gravitational waves (which are absent in $(2+1)$ dimensions) and the fact that on such more general manifolds, there are no known "background" solutions to perturb about. However, for the special case of (future) vacuum evolution on a pure closed hyperbolizable manifold, one can show that if the initial data is sufficiently close to that of an FLRW model, then the fully nonlinear gravitational perturbations eventually die out leaving a locally homogeneous and isotropic model in the asymptotic limit (see Andersson and Moncrief (2004)). It seems likely that this result can be generalized to allow for the inclusion of various types of matter sources as in the $(2+1)$-dimensional case.

## Acknowledgments

This research was supported in part by NSF grant PHY-0354391 to Yale University. We would also like to thank the Institut des Hautes Études Scientifiques, the Max-Planck-Institut für Gravitationsphysik, Albert-Einstein-Institut, and The Erwin Schrödinger International Institute for Mathematical Physics for their hospitality and support for several periods during which this research was carried out.

*See also:* Computational Methods in General Relativity: The Theory; Einstein Equations: Exact Solutions; Einstein Equations: Initial Value Formulation; Einstein Manifolds; General Relativity: Overview; Geometric Analysis and General Relativity.

## Further Reading


Anderson M (1997) Scalar curvature and geometrization conjectures for 3-manifolds. In: Grove K and Petersen P (eds.) *Comparison Geometry*. Mathematical Science Research Institute Publication Cambridge: Cambridge University Press.

Anderson M (2003) Remarks on Perelman's papers, http://www.math.sunysb.edu/anderson/papers.html.

Anderson M (2004) Cheeger–Gromov theory and applications to general relativity. In: Chruściel PT and Friedrich H (eds.) *The Einstein Equations and the Large Scale Behavior of Gravitational Fields (Cargese 2002)*, pp. 347–377. Basel: Birkhäuser.

Andersson L and Moncrief V (2003) Elliptic–hyperbolic systems and the Einstein equations. *Annales Henri Poincaré* 4: 1–34.

Andersson L and Moncrief V (2004) Future complete vacuum spacetimes. In: Chruściel PT and Friedrich H (eds.) *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, pp. 299–330. Basel: Birkhäuser.

Arnowitt R, Deser S, and Misner C (1962) The dynamics of general relativity. In: Witten L (ed.) *Gravitation: An Introduction to Current Research*. New York: Wiley.

Cao H, Chow B, Chu S, and Yau ST (eds.) (2003) *Collected Papers on Ricci Flow*, Series in Geometry and Topology, vol. 37. Somerville, MA: International Press.

Carlip S (1998) *Quantum Gravity in* $2+1$ *Dimensions*. Cambridge: Cambridge University Press.

Choquet-Bruhat Y (2004) Future complete U(1) symmetric Einsteinian spacetimes, the unpolarized case. In: Chruściel PT and Friedrich H (eds.) *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, pp. 251–298. Basel: Birkhäuser.

Choquet-Bruhat Y and York J (1980) The Cauchy problem. In: Held A (ed.) *General Relativity and Gravitation*, vol. 1, New York: Plenum.

Fischer A (1970) The theory of superspace. In: Carmeli M, Fickler S, and Witten L (eds.) *Relativity*. New York: Plenum.

Fischer A and Marsden J (1972) The Einstein equations of evolution – a geometrical approach. *Journal of Mathematical Physics* 28: 1–38.

Fischer A and Marsden J (1975) Deformations of the scalar curvature. *Duke Mathematical Journal* 42: 519–547.

Fischer A and Moncrief V (2000) The reduced Hamiltonian of general relativity and the $\sigma$-constant of conformal geometry. In: Cotsakis S and Gibbons GW (eds.) *Proceedings of the 2nd Samos Meeting on Cosmology, Geometry and Relativity, Mathematical and Quantum Aspects of Relativity and Cosmology*, Samos, August 31–September 4, 1998, *Lecture Notes in Physics*, vol. 537, pp. 70–101. Berlin: Springer.

Fischer A and Moncrief V (2002a) Conformal volume collapse of 3-manifolds and the reduced Einstein flow. In: Holmes P, Newton PK, and Weinstein A (eds.) *Geometry, Mechanics, and Dynamics: Volume in Honor of the 60th Birthday of J. E. Marsden*, pp. 463–522. New York: Springer.

Fischer A and Moncrief V (2002b) Hamiltonian reduction and perturbations of continuously self-similar $(n+1)$-dimensional Einstein vacuum spacetimes. *Classical Quantum Gravity* 19: 5557–5589.

Fischer A and Tromba A (1984) On a purely "Riemannian" proof of the structure and dimension of the unramified moduli space of a compact Riemann surface. *Mathematische Annalen* 267: 311–345.

Gromov M (1999) *Metric Structures for Riemannian and Non-Riemannian Spaces*. Boston: Birkhäuser.

Hempel J (1976) 3-Manifolds, *Annals of Mathematical Studies*, Number 86. Princeton: Princeton University Press.

Isenberg J (1995) Constant mean curvature solutions of the Einstein constraint equations on closed manifolds. *Classical Quantum Gravity* 12: 2249–2274.

Jaco W (1980) Lectures on three-manifold topology. In *Conference Board of the Mathematical Sciences, number 43, Regional Conference Series in Mathematics*. Providence, RI: American Mathematical Society.

Lichnerowicz A (1955) *Théories Relativistes de la Gravitation et de Electromagnetisme*. Paris: Masson.

Tanimoto M, Koike T, and Hosoya A (1997) Hamiltonian structures for compact homogeneous universes. *Journal of Mathematical Physics* 38: 6560–6577.

Thurston W (1997) *Three-Dimensional Geometry and Topology* vol. 1. Princeton: Princeton University Press.

Wainwright J and Ellis GFR (eds.) (1997) *Dynamical Systems in Cosmology*. Cambridge: Cambridge University Press.

# Hamiltonian Systems: Obstructions to Integrability

**M Irigoyen**, Université P.-M. Curie, Paris VI, Paris, France



## Introduction

In the study of differential systems, and particularly of Hamiltonian differential equations, a fundamental problem is the question of their integrability. Because there are different definitions of this notion, a system which is integrable according to one definition can be nonintegrable according to another one. The notion of integrability is connected to the existence of a sufficiently large number of first integrals, which are linked to conservation laws. For a real analytic Hamiltonian system with $n$ degrees of freedom, the "complete integrability" means the existence of $n$ first integrals, which are functionally independent, and "in involution," in the entire phase space. These integrals can be functions of class $C^r$ ($r$ finite), $C^\infty$, or analytic.

For the classical problems of Hamiltonian mechanics which are integrable, their first integrals can be continued into the complex domain of the variables, as one-valued holomorphic, or meromorphic, functions of complex time. This fact leads to the concept of "complex integrability" of a system. Note that a real Hamiltonian system which is integrable may be nonintegrable in the complex domain, if the real first integrals cannot be continued as one-valued holomorphic functions of the complex time.

Generally, the branching of solutions of a system, as functions of complex time, is an obstruction to the existence of one-valued first integrals. To study this problem, one can, following Poincaré, expand the solutions in convergent series of a small parameter: this is the base of "perturbation methods," and the main fact is that a small perturbation of an integrable Hamiltonian system generally destroys its integrability. Another method of proving nonintegrability consists of studying the linearized equations along a particular solution. This last direction has been exploited recently, in particular, through methods based on algebraic results inspired by differential Galois theory.

## Hamiltonian Systems and Mechanics

Let us consider a conservative holonomic real dynamical system with $n$ degrees of freedom: the positions of this system are points of an $n$-dimensional real manifold $N$ (the state space or configuration space) with local coordinates $x_1, x_2, \ldots, x_n$. If the velocities are denoted by $\dot{x}_i = \mathrm{d}x_i/\mathrm{d}t$, we consider the Lagrangian function $L$ associated to this system:

$$L(x_1, \ldots, x_n, \dot{x}_1, \ldots, \dot{x}_n) = T(x_1, \ldots, x_n, \dot{x}_1, \ldots, \dot{x}_n) + V(x_1, \ldots, x_n)$$

where $\dot{x} = (\dot{x}_1, \ldots, \dot{x}_n)$ is a tangent vector to the manifold $N$ at the point $x = (x_1, \ldots, x_n)$. The kinetic energy $T(x, \dot{x})$ is a positive-definite quadratic form in $\dot{x}_1, \ldots, \dot{x}_n$, and $V(x)$ is the potential energy, whose gradient determines the forces acting on the system.

The motions $(x_1(t), x_2(t), \ldots, x_n(t))$ of the system on the manifold $N$ are the extremals of the action integral: $\int_{t_1}^{t_2} L(x_1, \ldots, x_n, \dot{x}_1, \ldots, \dot{x}_n)\,\mathrm{d}t$ ("principle of stationary action of Hamilton") and they are the solutions of the Euler–Lagrange system, which consists in $n$ differential equations of second order for the coordinates $x_1, x_2, \ldots, x_n$ (Whittaker 1904):

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}_i}\right) - \frac{\partial L}{\partial x_i} = 0, \quad 1 \leq i \leq n$$

This system can be written in the Hamiltonian form: the Lagrangian $L$ is a function defined on the tangent bundle $TN$ of the state space $N$, with local coordinates $x_1, \ldots, x_n, \dot{x}_1, \ldots, \dot{x}_n$ (i.e., an element of $TN$ consists in a point $\boldsymbol{x}$ of $N$, joint with a tangent vector to $N$ at $\boldsymbol{x}$). Now, we consider the cotangent bundle $T^*N$: an element of $T^*N$ consists in a point $\boldsymbol{x}$ of $N$ joint with a cotangent vector to $N$ at $\boldsymbol{x}$, that is, a linear form defined in the tangent space to $N$ at $\boldsymbol{x}$. In local coordinates, the components of this linear form are $y_1, \ldots, y_n$, defined by: $y_i = \partial L/\partial \dot{x}_i$ ; $y_1, \ldots, y_n$ are called the generalized momenta, or impulsions. $x_i$ and $y_i$ are called conjugate canonical variables.

The mapping from $TN$ to $T^*N$ thus defined is the Legendre transformation (Abraham and Marsden 1967). Through it, the Euler–Lagrange equations become a system of $2n$ differential equations of first order:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \frac{\partial H}{\partial y_i}, \quad \frac{\mathrm{d}y_i}{\mathrm{d}t} = -\frac{\partial H}{\partial x_i}, \qquad 1 \leq i \leq n$$

where $H(x_1, \ldots, x_n, y_1, \ldots, y_n) = T(x_1, \ldots, x_n, y_1, \ldots, y_n) - V(x_1, \ldots, x_n)$.

$H$ is the Hamiltonian function of this system. The solutions of these differential equations are curves on the $2n$-dimensional manifold $T^*N$, whose projections in the $n$-dimensional state manifold $N$ coincide with the solutions of the Lagrangian system. $T^*N$ is

called the phase space of the system. The second members of the differential system define a vector field in the phase space.

Let $M = T^*N$. On this $2n$-dimensional manifold, consider the standard symplectic form $\Omega = \sum_{i=1}^{n} \mathrm{d}y_i \wedge \mathrm{d}x_i$. If $f$ and $g$ are $C^\infty$-functions on $M$, we define their Poisson bracket $\{f, g\}$ in local coordinates by

$$\{f, g\} = \sum_{i=1}^{n} \left( \frac{\partial f}{\partial x_i} \frac{\partial g}{\partial y_i} - \frac{\partial f}{\partial y_i} \frac{\partial g}{\partial x_i} \right)$$

It defines the space $C^\infty(M)$ as a Lie algebra over $R$.

Then, if $H \in C^\infty(M)$ is the Hamiltonian function associated to a system, the corresponding Hamiltonian equations can be written as the following $2n$ "canonical equations" (Arnol'd 1976):

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = \{x_i, H\} = \frac{\partial H}{\partial y_i}, \quad \frac{\mathrm{d}y_i}{\mathrm{d}t} = \{y_i, H\} = -\frac{\partial H}{\partial x_i}, \quad 1 \le i \le n \qquad [1]$$

A function $F \in C^\infty(M)$ is a (first) integral of eqns [1] if it is constant along any solution of [1], that is, if it verifies: $\{F, H\} = 0$. Thus, a first integral is a quantity which is preserved along a solution ("conservation law"). In particular, $H$ itself is a first integral of the eqns [1]. It represents the "total energy" of the system.

1. The simplest example of Hamiltonian system is the harmonic oscillator defined by the one degree of freedom Hamiltonian:

$$H(x, y) = \tfrac{1}{2} y^2 + \tfrac{1}{2} x^2$$

It possesses the energy integral $H$. Thus, the trajectories in the phase space $R^2$ (phase plane) are given by $x^2 + y^2 = 2h$, which are concentric circles if the constant energy verifies $h \ge 0$. The phase space $R^2$ is foliated by these circles. The system is said to be "integrable." Obviously, it is also possible to construct Hamiltonian systems with $n$ degrees of freedom ($n \succ 1$), by coupling $n$ harmonic oscillators, with a Hamiltonian defined by

$$H(x_1, \ldots, x_n, y_1, \ldots, y_n) = \frac{1}{2} \sum_{i=1}^{n} {y_i}^2 + \sum_{i=1}^{n} a_i {x_i}^2$$

with $n$ constant coefficients $a_i \succ 0$.

2. Another example of Hamiltonian system with one degree of freedom is the simple mathematical pendulum. The state coordinate is the angle $\theta$ of the pendulum with the vertical axis, defined modulo $2\pi$. The phase space is: $M = S^1 \times R$ ($x = \theta (\mathrm{mod}\, 2\pi) \in S^1, y \in R$),that is, a cylinder. The Hamiltonian function is: $H(x, y) = (1/2)y^2 - \cos x$; $H$ is a first integral of the differential equations, the system is integrable and the trajectories on the cylinder $S^1 \times R$ are defined by $(1/2)y^2 - \cos x = h$. According to the constant value of $h$ on each phase curve, the solutions are periodic oscillations of the pendulum (if $h \prec h_0$), periodic solutions of rotation where the angle varies monotonically with time (if $h \succ h_0$), two equilibria (one stable, one unstable) and solutions which "begin" when $t \to -\infty$ at the unstable equilibrium and "finish" when $t \to +\infty$ at the same point (if $h = h_0$): the corresponding phase curves are called "separatrices."

3. The system of Hénon–Heiles (Hénon and Heiles 1964) is a system with two degrees of freedom. The phase space is $R^2 \times R^2$ and the Hamiltonian is defined by

$$H(x_1, x_2, y_1, y_2) = \tfrac{1}{2}({y_1}^2 + {y_2}^2) + \tfrac{1}{2}({x_1}^2 + {x_2}^2) + {x_1}^2 x_2 - \lambda {x_2}^3$$

where $\lambda$ is a real constant. This system is "integrable" for some isolated values of the parameter $\lambda$ (Ziglin 1983) and "nonintegrable" otherwise. Of course, it is necessary to define the integrability of a Hamiltonian system, although according to Poincaré: "A system of differential equations is only more or less integrable."

## Integrability of Hamiltonian Systems

Generally, if a differential system is of order $p$, it is necessary to know $p$ first integrals to integrate it. But if the system is Hamiltonian of order $2n$, only $n$ first integrals are sufficient to integrate it "by quadratures," that is, by "algebraic" operations such as integrations and inverting of functions. The reason is that the existence of one first integral allows us to reduce the order of the system by two: a system of order $2n$ with one first integral can be reduced to order $2n - 2$.

**Theorem of Liouville** (see Arnol'd (1976)). *Suppose that $F_1, F_2, \ldots, F_n \in C^\infty(M)$ are $n$ first integrals of the Hamiltonian system [1] which are "in involution," that is, such that: $\{F_i, F_j\} = 0, \forall i, j$, and suppose that they are functionally independent, that is, the $n$ differentials, $\mathrm{d}F_i$, are linearly independent at each point of the level set $M_f$ defined by*

$$M_f = \{(x_1, \ldots, x_n, y_1, \ldots, y_n) \in M : F_i(x_1, \ldots, y_1, \ldots) = f_i, i = 1, 2, \ldots, n\}$$

*Then*

(i) *the set $M_f$ is a manifold which is invariant along the solutions of the system [1];*

(ii) *if $M_f$ is compact and connected, it is diffeomorphic to an $n$-dimensional torus*

$$T^n = S^1 \times S^1 \times \cdots \times S^1 = \{(\varphi_1, \ldots, \varphi_n) : \varphi_i \in R/2\pi Z\};$$

(iii) *the Hamiltonian flow on each torus* $M_f$ *is linear and "quasiperiodic" with frequencies* $\omega_i$ *defined by* $\mathrm{d}\varphi_i/\mathrm{d}t = \omega_i(f_1, f_2, \ldots, f_n)$; *and*
(iv) *the Hamiltonian equations are integrable by quadratures.*

If a Hamiltonian verifies the assumptions of the theorem of Liouville, one can prove that it exists, locally, canonical coordinates $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_n)$ and $\boldsymbol{I} = (I_1, \ldots, I_n)$ such that the Hamiltonian function depends only on the variables $I_i$. Then

$$\frac{\mathrm{d}\varphi_i}{\mathrm{d}t} = \frac{\partial H}{\partial I_i}$$
$$\frac{\mathrm{d}I_i}{\mathrm{d}t} = -\frac{\partial H}{\partial \varphi_i} = 0$$

These equations are immediately integrated as follows:

$I_i = \text{constant}$, and $\varphi_i = \omega_i \cdot t + \varphi_i(0)$, with

$$\omega_i(I_1, I_2, \ldots, I_n) = \left.\frac{\partial H}{\partial I_i}\right|_{I=cst}$$

Such local coordinates $(\varphi_i, I_i)$ are called "action-angle" variables. They were defined for the first time by Delaunay and they play an important part in the theory of perturbations.

**Remark** An invariant torus $T^n$ of the theorem of Liouville is characterized by the constant values of the actions $I_i$, which determine the frequencies $\omega_i$ on it. Such a torus is said to be nonresonant if the relation between the frequencies $\omega_i$: $\sum_{i=1}^{n} k_i \omega_i = 0$ (where $k_1, \ldots, k_n$ are integers) implies that $k_i = 0, \forall i$. The frequencies $\omega_i$ are then rationally independent. If a torus is nonresonant, the phase trajectories are dense everywhere and the motion is quasiperiodic on it.

A torus is said to be resonant, if the frequencies $\omega_i$ are rationally dependent: they verify a relation $\sum_{i=1}^{n} k_i \cdot \omega_i = 0$, with $(k_1, \ldots, k_n) \neq (0, \ldots, 0)$. Then the phase trajectories are not dense on the torus; they belong to tori of lower dimension.

A consequence of the theorem of Liouville is that, if a two-degree-of-freedom Hamiltonian system possesses one first integral $F$ (in addition to $H$, and independent of $H$), it is integrable because $F$ is necessarily in involution with $H$: $\{F, H\} = 0$.

An example of system with three degrees of freedom which is integrable is the Lagrangian symmetric top with one fixed point (there exists a cylindric symmetry for the inertia momenta and the center of mass is on the symmetry axis). This system possesses three first integrals that are in involution and independent: $H$, and the angular momenta $M_z$ and $M_3$, which correspond to the (constant) frequencies of precession and nutation of the top. The level sets $M_f$ are here tori of dimension 3, which are indexed by the three frequencies (or by the constant values of the three integrals).

There are other integrable cases for this problem of a rigid body with a fixed point (see Kozlov (1983)): the Euler's case (when the fixed point is the center of mass); the Kowalevskaya's case (in which the inertia momenta verify two relations and the third coordinate of the center of mass vanishes – see Kowalevski (1889)); and the Goryachev–Chaplygin's case, which is integrable only on a single integral level.

A fundamental and classical example of integrable Hamiltonian system is the Kepler's problem: the motion of a ponctual mass in the gravitational (Newtonian) field of a center, for instance, a planet in the field of attraction of the Sun.

Another example is the problem of two fixed centers: an infinitesimal mass in the field of two centers, problem which was integrated by Lagrange (Lagrange, 1810).

## Isolated Periodic Orbits and Nonintegrability

We consider a real Hamiltonian system with $n$ degrees of freedom and we suppose that there exists a particular $T$-periodic solution $\Gamma_T$ (which is not an equilibrium). Along $\Gamma_T$, we consider the linearized equations deduced from the Hamiltonian system. They can be decoupled into the tangential equation (one degree of freedom) which possesses the first integral $\mathrm{d}H$ and the normal variational system which can be written as

$$\frac{\mathrm{d}\xi}{\mathrm{d}t} = J \cdot K(\Gamma_T(t)) \cdot \xi \qquad [2]$$

where

$$J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$$

is the standard symplectic matrix of order $2(n-1)$ and $K(\Gamma_T(t))$ is a $T$-periodic matrix depending on the solution $\Gamma_T$.

The solutions of the linear system [2] form a vector space. As a definition, the monodromy matrix $M(T)$ expresses how fundamental solutions of the linear system [2] are transformed after one period $T$, that is, along the periodic closed orbit $\Gamma_T$:

$$\xi(t+T) = M(T) \cdot \xi(t)$$

Poincaré showed that if one of the eigenvalues of $M(T)$ is different from 1, then the periodic solution $\Gamma_T$ is isolated. Furthermore, if the number of first

integrals of the Hamiltonian system, independent along $\Gamma_T$, is equal to $k$, then at least $2k$ eigenvalues of $M(T)$ are equal to 1.

**Theorem** (Poincaré 1892). *If the Hamiltonian system possesses n integrals in involution, and independent along a periodic solution $\Gamma_T$, then $\Gamma_T$ is nonisolated.*

Then, if the Hamiltonian system possesses a dense set of isolated periodic orbits, it cannot have $n$ integrals in involution and independent in an open domain.

## Nearly Integrable Hamiltonian Systems, Theorem of Poincaré

Consider the Hamiltonian system with $n$ degrees of freedom, depending on a small real parameter $\varepsilon \in (-\varepsilon_0, +\varepsilon_0)$, defined by the analytic function $H$:

$$H(\boldsymbol{\varphi}, \boldsymbol{I}, \varepsilon) = H_0(\boldsymbol{I}) + \varepsilon \cdot H_1(\boldsymbol{\varphi}, \boldsymbol{I}) \qquad [3]$$

where $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_n) \in T^n, \boldsymbol{I} = (I_1, \ldots, I_n) \in R^n$, and where $H_1$ is periodic in the angles $\varphi_i$.

This system is called "nearly integrable" because when $\varepsilon = 0$, the "unperturbed system" $H_0$ is integrable in the action-angle variables $\boldsymbol{\varphi}, \boldsymbol{I}$:

$$H(\boldsymbol{\varphi}, \boldsymbol{I}, 0) = H_0(\boldsymbol{I})$$

then

$$\frac{\mathrm{d}\boldsymbol{I}}{\mathrm{d}t} = 0, \quad \frac{\mathrm{d}\boldsymbol{\varphi}}{\mathrm{d}t} = \frac{\partial H_0}{\partial \boldsymbol{I}} = \boldsymbol{\omega}(\boldsymbol{I})$$

system which can be integrated by quadratures:

$$\boldsymbol{I} = \boldsymbol{I}^0 \quad \text{and} \quad \boldsymbol{\varphi} = \boldsymbol{\varphi}^0 + \boldsymbol{\omega}(\boldsymbol{I}^0) \cdot t$$

According to the theorem of Liouville, the motion of the unperturbed problem takes place on $n$-dimensional tori $(S^1)^n$ in the phase space. On these invariant tori, indexed by the actions $\boldsymbol{I}$, the motion is generally quasiperiodic (if the frequencies $\boldsymbol{\omega}(\boldsymbol{I})$ are rationally independent).

We are now interested in studying the perturbed system [3] with $\varepsilon \neq 0$, and its integrability which is, according to Poincaré (1892), "the fundamental problem of dynamics." This problem of nearly integrable Hamiltonian systems is directly inspired by celestial mechanics where the motions in the solar system are, in a first approximation, described by the (integrable) Kepler's problem. In particular, the "restricted three-body problem" is the study of the motion of a planet in the gravitational field of the Sun, with the perturbative attraction of Jupiter. It is also the problem of the Moon in the field of the Earth, with the perturbative attraction of the Sun (Poincaré 1892).

**Theorem of Poincaré** (Poincaré 1892). *Assume that, in the Hamiltonian function* [3]:

(i) (*nondegeneracy condition*) *the unperturbed Hamiltonian $H_0$ is nondegenerate, that is,*

$$\det\left|\frac{\partial^2 H_0}{\partial I_i \partial I_j}\right| = \det\left|\frac{\partial \omega_i}{\partial I_j}\right| \neq 0$$

*in an open domain of the phase space;*

(ii) (*genericity condition*) *no coefficient $h_{\boldsymbol{k}}(\boldsymbol{I})$ in the Fourier expansion of $H_1$ with*

$$H_1(\boldsymbol{\varphi}, \boldsymbol{I}) = \sum_{\boldsymbol{k} \in Z^n} h_{\boldsymbol{k}}(\boldsymbol{I}) \cdot \mathrm{e}^{\mathrm{i}\langle \boldsymbol{k}, \boldsymbol{\varphi}\rangle}$$

*does identically vanish in the nonresonant domain $G \in R^n$ of the actions defined by*

$$G = \left\{ \boldsymbol{I} \in R^n : \sum_{i=1}^n k_i \cdot \omega_i(\boldsymbol{I}) = 0, \right.$$
$$\left. \text{iff } (k_1, \ldots, k_n) = (0, \ldots, 0) \right\}$$

*then, there is no analytic first integral $F(\boldsymbol{\varphi}, \boldsymbol{I}, \varepsilon)$ independent of the Hamiltonian function H.*

Thus, a perturbation of a nondegenerate integrable Hamiltonian system is generically nonintegrable.

When one wants to apply this theorem to celestial mechanics, a peculiarity is that the unperturbed problem corresponds to the Keplerian system, which is degenerate, and this is a specific difficulty of these systems.

## Splitting of Separatrices and Nonintegrability

Consider a Hamiltonian system with $n = 2$ (degrees of freedom) defined as in eqn [3] by a perturbation of an integrable Hamiltonian:

$$H(\varphi_1, \varphi_2, I_1, I_2, \varepsilon) = H_0(I_1, I_2) + \varepsilon \cdot H_1(\varphi_1, \varphi_2, I_1, I_2) \qquad [4]$$

The unperturbed problem is integrable and its four-dimensional phase space is foliated by two-dimensional invariant tori $T^2 : \boldsymbol{I} = \text{constant}$. If $H_0$ is nondegenerate, the nonresonant tori are dense and the resonant tori also are dense in the phase space.

According to Kolmogorov's theorem and the Kolomogorov–Arnol'd–Moser (KAM) theory (Arnol'd 1985), the majority of the nonresonant tori of the unperturbed problem $H_0$ are preserved in the full problem [4]: they are slightly deformed, and are invariant in the perturbed

system. The resonant tori of $H_0$ are destroyed in the perturbed problem.

Now we consider, in the phase space, a transverse surface $S$ to the invariant tori $T^2$ of the perturbed system. A trajectory of the system generated by [4], which crosses $S$ through a point $w_0$, will cross $S$ again, for the first time, through a point $w_1$: this defines the "first return map" or "Poincaré's map" $R: w_0 \mapsto R(w_0) = w_1$. $S$ is called a Poincaré's section. If $w_0$ belongs to a preserved invariant torus of the perturbed system, the successive points $w_0, w_1 = R(w_0)$, $w_2 = R(w_1)$, $w_3 = R(w_2), \dots$ belong to the intersection of this torus with $S$; thus, they belong to a curve diffeomorphic to a circle, which is an invariant curve of the map $R$. If $w_0$ does not belong to a preserved invariant torus of [4], the sequence of points $w_0, w_1, w_2, \dots$ through the Poincaré's map belongs to a curve much more complicated than a curve diffeomorphic to a circle (Poincaré 1890, Arnol'd 1985) and the "chaotic" behavior of this sequence is the mark of the nonintegrability of the system [4].

The best way of numerically showing the "evidence" of nonintegrability is to study the example of a system with "one and a half" degree of freedom, that is, a system with one degree of freedom whose Hamiltonian depends on time: $H(\varphi, I, t)$. An example of such a system is the problem of a mathematical pendulum whose length $l$ performs periodic oscillations, defined by the Hamiltonian function

$$H(\varphi, p, t) = \frac{p^2}{2} - \omega^2(1 + \varepsilon \cdot f(t)) \cdot \cos\varphi \qquad [5]$$

where $\varphi \in S^1$, $p \in R$, and $f$ is periodic of period $T$.

The unperturbed system ($\varepsilon = 0$) is integrable (one degree of freedom with a Hamiltonian independent of $t$):

$$H_0(\varphi, p) = \frac{p^2}{2} - \omega^2 \cdot \cos\varphi$$

The phase portrait of this problem is similar to the one of the simple mathematical pendulum of constant length: on the cylinder $S^1 \times R$ there are two equilibria (stable and unstable) and separatrices "beginning" and "finishing" at the hyperbolic point $\chi$. The invariant stable and unstable manifolds associated to $\chi$ and represented by these separatrices were called by Poincaré as "homoclinic" trajectories, because each of them, drawn on the phase cylinder, joins equilibrium $\chi$ to itself.

If $\varepsilon \neq 0$, we define a Poincaré section of the perturbed system [4] in the following way: from an initial point $w_0(\varphi_0, p_0, t_0)$, we consider the successive planes perpendicular to the $t$-axis in the "extended" phase space $\{(\varphi, p, t)\}$, defined by: $t_0, t_1 = t_0 + T$, $t_2 = t_0 + 2T$, $t_3 = t_0 + 3T, \dots$ and we look at the successive intersections of the orbit of $w_0$ with these planes: $w_0, w_1, w_2, \dots$. If we identify all the successive planes and if we draw on the same picture, the points $w_0, w_1, w_2, \dots$, we obtain a phase portrait in which the equilibria of the unperturbed problem are present, but the separatrix which "leaves" the point $\chi$ is not confounded with the separatrix which "ends" at $\chi$, as in the unperturbed problem: the two invariant curves are transversal to each other: they "split" and have an infinite number of intersections. This splitting is the traduction of the nonintegrability of the perturbed system [5].

A method to detect this splitting of separatrices consists in computing the Melnikov's function which gives a measure of the angle between the separatrices at their first intersection when they split.

Many concrete Hamiltonian systems have been studied by this method and numerical investigations on the splitting have permitted detection of their nonintegrability.

## Topological Obstructions to Integrability

We are interested in a natural mechanical system with two degrees of freedom and we suppose that the state space $N$ is a real analytic surface which is compact and orientable. Then, $N$ consists of a two-dimensional sphere with $k$ handles (or a torus with $k$ holes). The number $k$ is a topological invariant of the surface and is called the genus of $N$.

Let $H$ be the Hamiltonian function associated to this problem. The Hamiltonian system possesses the first integral $H$. It is completely integrable if and only if another analytic integral $F$ exists, functionally independent of $H$. In this case, the state space $N$ belongs necessarily to a very restrictive class of surfaces.

**Theorem** (Kozlov 1983). *If the genus $k$ of the state manifold $N$ is not equal to 0 or 1 (i.e., if $N$ is neither diffeomorphic to the sphere $S^2$ nor to the torus $T^2$), then the Hamiltonian system generated by $H$ does not possess a first integral, analytic on $T^*N$ and functionally independent of the energy integral $H$.*

Note that this theorem does not apply to first integrals which are $C^\infty$ only, and examples can be given which illustrate this case (Kozlov 1983, 1989).

For systems with more than two degrees of freedom, an open question is to know whether the complete integrability imposes restrictions to the topology of the state manifold $N$.

## Singular Point Analysis, Branching of Solutions and Ziglin's Theory

If we look at the classical Hamiltonian problems which have been integrated, their first integrals are real functions which can be continued in the complex domain as one-valued holomorphic or meromorphic functions of the complex time $t$ (polynomials, rational functions, etc.). This fact leads to the concept of "complex integrability." But the nonintegrability of a complex Hamiltonian system does not imply the nonintegrability of its restriction to the real domain: it may happen that a real analytic first integral does not possess a continuation in the complex domain as a meromorphic function.

Adopting this point of view, S Kowalevskaya (Kowalevski 1889) studied the problem of a top rotating around a fixed point, and she discovered a new case of integrability for this classical problem of Hamiltonian mechanics. She searched for conditions on the parameters such that the movable singularities of the solutions in the complex plane of time are poles (as a definition, a singularity is movable if its location in the complex domain depends on the initial conditions). Such differential systems are said to be of "Painlevé's type." In this case, the solutions are single valued in the complex $t$-plane and there is no branching of these solutions. The leading idea is the following: a first integral must be constant along a solution, and an eventual branching would change its value along a loop around a singularity in the complex $t$-plane. However, finite branchings of solutions can be compatible with integrability.

The main tool in this analysis is the calculation of the "Kowalevski's exponents" which determine the eventual branching of a solution around a singularity.

In spite of the efficiency of the Painlevé analysis for the search of integrable (or nonintegrable) systems, the relation between the analytic properties of their solutions (Painlevé) and their integrability in the sense of Liouville remains mysterious. The most fundamental result obtained in this field is a theorem of Adler and van Moerbeke which proves that, if a system has the Painlevé property and if it is integrable in the sense of Arnol'd–Liouville, then it is algebraically integrable (Adler and van Moerbeke 1989).

The discovery of S Kowalevskaya inspired Ziglin, who related the existence of meromorphic first integrals for a Hamiltonian system, to the properties of the linearized equations along a particular periodic solution of this system, especially to the monodromy group associated to this linear system. Ziglin used the constraints imposed to this monodromy group by the existence of first integrals.

Let us consider a Hamiltonian system defined on a complex analytic symplectic manifold of dimension $2n$, and suppose that there exists a family of periodic solutions $\Gamma$. The linearized equations deduced from the Hamiltonian system along $\Gamma$ are decoupled into tangential and normal equations. We are interested by the normal equations, which are linear with periodic coefficients.

**Ziglin's Theorem** (Ziglin 1983). *Assume that a Hamiltonian system has a family of particular solutions $\Gamma_h$ (which are not equilibria) parametrized by periodic functions of the complex time and depending analytically on a real parameter $h \in (h_1, h_2)$. Let $G$ be the monodromy group of the normal variational equation associated to the solution $\Gamma_h$. A monodromy matrix $g \in G$ is said to be nonresonant if every eigenvalue of $g$ is different from a root of unity. If the Hamiltonian system has a meromorphic integral $F$, functionally independent of the Hamiltonian $H$ in a neighborhood of $\Gamma_h$, and if the monodromy group $G$ contains a nonresonant element $g_1$, then for any $g_2 \in G$, the commutator $g^* = g_2^{-1} \cdot g_1^{-1} \cdot g_2 \cdot g_1$ satisfies either $g^* = \mathrm{Id}$ or $g^* = (g_1)^2$.*

As a corollary of this theorem, we have sufficient conditions of nonintegrability: if the necessary conditions of integrability of Ziglin are not satisfied by a Hamiltonian system, it is not analytically integrable. For instance, this will happen if we can find two nonresonant monodromy matrices $g_1$ and $g_2$ which do not commute. If the periodic solution $\Gamma_h$ has two complex periods, the monodromy group $G$ has generators $g_1$ and $g_2$, respectively associated to each of these periods and their commutativity can be sometimes studied.

These sufficient conditions of nonintegrability were studied for particular Hamiltonian systems, first by Ziglin himself.

Several concrete systems with two degrees of freedom were proved to be nonintegrable by Ito, Yoshida, Churchill, Rod, and many other mathematicians who applied this "Ziglin's method": for instance, the Hénon–Heiles system, the Yang–Mills system, and Hamiltonian systems with a homogeneous potential.

## Nonintegrability and Differential Galois Theory (Morales-Ruiz 1999)

Recently, the integrability of Hamiltonian systems was studied with algebraic tools from the differential Galois theory, applied to linear differential

systems. As in Ziglin's theory, we consider a particular solution $\Gamma$ (not necessarily periodic) of a differential system generated by an analytic Hamiltonian $H$ with $n$ degrees of freedom, and the (linear) variational equations along $\Gamma$. The idea is that, if the Hamiltonian system is integrable, we can assume that the linearized equations along $\Gamma$ must also have a "regular behavior." If the Hamiltonian system is integrable, it will also be the case for the variational equations.

The normal variational system (of order $2n-2$) can be written as

$$\frac{d\xi}{dt} = J \cdot K(\Gamma(t)) \cdot \xi \qquad [6]$$

with

$$J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$$

$K(\Gamma(t))$ is a matrix depending on the particular solution $\Gamma$.

We have to define the "Galois group" of the linear equation [6]. Recall that in the classical Galois theory of algebraic equations, the Galois group is defined by the automorphisms which map roots onto roots of the equation. In an analogous way, in the differential Galois theory, we consider the maps which send a fundamental solution of eqn [6] on a fundamental solution. In order to define the Galois group $G$ associated to [6], we consider a differential field $K$ of functions over $C$ (i.e., a field of functions equipped with a derivation). The field of constants of $K$ is $C$; it is the subfield of $K$ whose elements have a derivative equal to zero. We denote by $K\langle\xi, \eta, \ldots\rangle$ the differential field extension obtained from $K$ by the adjunction of the functions $\xi, \eta, \ldots$. If $(\varphi, \psi)$ is a fundamental system of solutions of eqn [6], then $L = K\langle\varphi, \psi\rangle$ is the smallest differential field extension which contains all the solutions of [6]. The field of constants of $L$ is the same as the one of $K$, that is, $C$. By definition, $L$ is a Picard–Vessiot extension of $K$.

The differential Galois group of $L$ is defined as the group of the automorphisms $\gamma$ of $L$ (that map a solution of [6] onto a solution) leaving the field of constants fixed. Given a fundamental system of solutions $(\varphi, \psi)$, we can associate to each automorphism $\gamma$ the matrix $M$ such that $(\gamma(\varphi), \gamma(\psi)) = (\varphi, \psi).M$. By definition, the set of these matrices $M$ is the Galois group $G$ of eqn [6]. It is a linear algebraic group (because, the matrices $M$ being symplectic, their coefficients verify polynomial equations) and a subgroup of the linear group of matrices GL($C$). We note that, for a given linear system, the monodromy group is contained in the Galois group and both are subgroups of the symplectic group Sp($C$).

In the Galois group $G$ of eqn [6], we consider $G^0$, the connected component of the identity. The integrability of the initial Hamiltonian system is connected to the integrability of the variational equation [6] and, through it, to the properties of its Galois group:

**Theorem of Morales and Ramis** (Morales-Ruiz 1999). *If an analytic Hamiltonian system is completely integrable, then the Galois group associated to the variational equation along a particular solution $\Gamma$ is such that its connected component of identity $G^0$ is Abelian.*

Thus, if a Hamiltonian system is such that $G^0$ is not Abelian, there cannot exist a complete set of first integrals in involution in a neighborhood of the particular solution $\Gamma$ and the system is not integrable.

In the concrete applications of this theory, an algorithm of Kovacic allows us to determine the Galois group explicitly. By this method, several Hamiltonian systems were proved to be nonintegrable: for instance, systems of points on a line with a potential in $1/r^2$, studied by Julliard-Tosel (1998), but also ancient proofs of nonintegrability of homogeneous potentials, which were improved by Yoshida and Umeno, thanks to the theorem of Morales–Ramis.

*See also:* Billiards in Bounded Convex Domains; Infinite-Dimensional Hamiltonian Systems; Integrable Systems: Overview; Peakons; Separatrix Splitting.

## Further Reading


Abraham R and Marsden J (1967) In: Benjamin WA (ed.) *Foundations of Mechanics*. Reading, MA: Benjamin.

Adler M and van Moerbeke P (1989) *Inventiones Mathematicae* 97(1): 3–51.

Arnol'd VI (1976) In: Mir (ed.) *Méthodes mathématiques de la Mécanique classique*. Moscow: Editions Mir.

Arnol'd VI (1985) *Dynamical Systems III, Encyclopedia of Mathematical Sciences*. Berlin: Springer.

Birkhoff D (1927) Dynamical Systems. *American Mathematical Society Colloquium Publication*, 9.

Hénon M and Heiles C (1964) *Astronomical Journal* 69: 73–79.

Julliard-Tosel E (1998) *Comptes-Rendus de l'Academie des Sciences, Paris t.* 327: (Série I) 387–389.

Kowalevski S (1889) *Acta Mathematica* 12: 177–232.

Kozlov VV (1983) *Russian Mathematical Surveys* 38-1: 1–76.

Kozlov VV (1989) *Sovietic Science Review, Mathematical Physics*. 8: 1–81.

Lagrange JL (1810) *Mécanique analytique*. Œuvres complétes, Gauthier-Villars.

Morales-Ruiz JJ (1999) *Differential Galois Theory and Non-Integrability*. Basel: Birkhaüser-Verlag.

Poincaré H (1890) *Acta Mathematica* 13: 1–270.

Poincaré H (1892) Les Méthodes Nouvelles de la Mécanique Céleste. tome I, A. Blanchard.

Whittaker ET (1904) *A Treatise on the Analytical Dynamics*. Cambridge: Cambridge University Press.

Ziglin SL (1983) *Functional Analysis and Applications* 16: 181–189; 17: 6–17.

# Hamiltonian Systems: Stability and Instability Theory

**P Bernard**, Université de Paris Dauphine, Paris, France



The solar system has long appeared to astronomers and mathematicians as a model of stability. On the other hand, statistical mechanics relies on the assumption that large assemblies of particles form highly unstable systems (at the microscopic scale). Yet all these physical situations are described, at least to a certain degree of approximation, by Hamiltonian systems.

One may hope that Hamiltonian systems can be classified in two different categories, stable and unstable ones. However, the situation is much more complicated and both stable and unstable behaviors cohabit in typical systems. Even our examples are not perfect paradigms of stability and instability. Indeed, it is now clear from numerical as well as theoretical points of views that some instability is present over long timescales in the solar systems, so that for example future collisions between planets cannot be completely ruled out in view of our present understanding. On the other hand, unexpected patterns of stability have been discovered in systems involving a large number of particles.

Understanding the impact of stable and unstable effects in Hamiltonian systems has been considered ever since Poincaré as one of the most important questions in dynamical systems. In this article, we will discuss model Hamiltonian systems of the form

$$H_\epsilon(q,p) = h(p) + \epsilon G_\epsilon(q,p)$$

where $(q,p) \in \mathbb{T}^d \times U$, with $U$ a bounded open subset of $\mathbb{R}^n$. Recall that the equations of motion are

$$\dot{q}(t) = \partial_p h(p) + \epsilon \partial_p G_\epsilon(q,p) \quad [1]$$

$$\dot{p}(t) = -\epsilon \partial_q G_\epsilon(q,p) \quad [2]$$

The textbook by Arnol'd (1964) is a good general introduction on Hamiltonian systems. We will always denote by $\omega(p)$ the frequency map $\partial_p h(p)$, which plays a crucial role. Here, as is obvious in [2], the action variables $p$ are preserved under the evolution in the unperturbed case $\epsilon = 0$. We will try to explain what is known on the evolution of these action variables for the perturbed system. As we will see, in many situations, these variables are extremely stable. For example, KAM theorem implies that, for a positive measure of initial conditions $(q_0, p_0)$ the trajectory $(q(t), p(t))$ satisfies $\|p(t) - p(0)\| \le C\epsilon$ for all times. Examples show that some initial conditions may lead to unstable trajectories, that is, trajectories such that $\|p(t) - p(0)\| \ge 1/C$ for some $t$ (depending on $\epsilon$) and some fixed constant $C$ independent of $\epsilon$. However, this is, as we will see, possible only for very large time $t$ (meaning that $t$ as a function of $\epsilon$ has to go to infinity very quickly when $\epsilon \longrightarrow 0$). The main questions here are to understand in what situation instability is or is not possible, and what kind of evolutions can have the action variable $p$. Another important question is to estimate the speed (as a function of the parameter $\epsilon$) of the evolutions of $p$.

### A Convention

We assume, unless otherwise stated, that the Hamiltonians are real analytic. The norm $|H|$ of the Hamiltonian $H$ is the uniform norm of its holomorphic extension to a certain complex strip. We do not specify the width of this strip. Whenever we consider a family $H_\epsilon, F_\epsilon \ldots$ of Hamiltonians, we mean that the norm $|H_\epsilon|$ is bounded when $\epsilon \longrightarrow 0$.

## Averaging and Exponential Stability

The first observation concerning the action variables is that they should evolve at a speed of the order of $\epsilon$. However, averaging effects occur. More precisely, in the equation $\dot{p}(t) = -\epsilon \partial_q H_\epsilon(q(t), p(t))$, the variable $q(t)$ is moving fast compared to $p(t)$. If the evolution of $q(t)$ nicely fills the torus $\mathbb{T}^n$, it is tempting to think that the averaged equation

$$\dot{\bar{p}}(t) = -\epsilon \bar{V}_\epsilon(\bar{p}(t))$$

should approximate accurately the actual behavior of $p(t)$, where

$$\bar{V}_\epsilon(p) := \int_{\mathbb{T}^d} \partial_q H(q,p)\, \mathrm{d}q$$

We have $\bar{V} \equiv 0$, which leads us to think that the evolution should consist mainly of oscillations of small amplitude with no large evolution. This reasoning is limited by the presence of resonances.

### Frequencies

A frequency $\omega \in \mathbb{R}^d$ is said to be resonant if there exists $k \in \mathbb{Z}^d_*(= \mathbb{Z}^d - \{0\})$ such that $\langle k, \omega \rangle = 0$. The resonance module of $\omega$,

$$Z(\omega) = \{k \in \mathbb{Z}^d / \langle k, \omega \rangle = 0\}$$

is a subgroup $\mathbb{Z}^d$; we denote by $R(\omega)$ the vector space generated by $Z(\omega)$ in $\mathbb{R}^d$. The order of resonance $r(\omega)$ is the dimension of $R(\omega)$. The main examples of resonances of order $r$ are the

frequencies $\omega = (\omega_1, 0)$, where $\omega_1 \in \mathbb{R}^{d-r}$ is nonresonant. This example is universal. Indeed, if $\omega$ is a resonant frequency, then there exists a matrix $A \in \mathrm{Gl}_d(\mathbb{Z})$ such that $A\omega = (\omega_1, 0)$, where $\omega_1 \in \mathbb{R}^{d-r}$ is not resonant. The matrix $A$ can be seen as a diffeomorphism of $\mathbb{T}^d$, which transports the constant vector field $\omega$ to the constant vector field $A\omega = (\omega_1, 0)$. It is useful to distinguish, among nonresonant frequencies, some which are sufficiently nonresonant. A frequency $\omega \in \mathbb{R}^d$ is called Diophantine if there exist real constants $\gamma > 0$ and $\tau \geq d$ such that

$$|\langle \omega, k \rangle| \geq \gamma \|k\|^{1-\tau}$$

for each $k \in \mathbb{Z}^d_*$. Finally, a frequency is called resonant Diophantine if there exists a matrix $A \in \mathrm{Gl}_n(\mathbb{Z})$ such that $\omega = A(\omega_1, 0)$, where $\omega_1 \in \mathbb{R}^{d_1}$ is a Diophantine frequency.

### Symplectic Diffeomorphisms and Normal Forms

An efficient mathematical method to take averaging effects into account is the use of normal forms. Normal form theory consists in finding new coordinates in which the fast angles have been eliminated from the equations up to a small remainder. This is done exploiting the existence of a large group of diffeomorphisms preserving the Hamiltonian structure of equations, called symplectic diffeomorphisms or canonical transformations. We refer the reader to standard textbooks for these notions, for example to Arnol'd (1964). An important point is that a symplectic diffeomorphism $\phi$ sends the trajectories of the Hamiltonian $H \circ \phi$ to the trajectories of the Hamiltonian $H$. A Hamiltonian $N(q, p)$ is said to be in $R$-normal form, where $R$ is a linear subspace of $\mathbb{R}^n$, if $\partial_q N \in R$ for each $(q, p)$. Let us give an illustrative result, taken from Lochak *et al.* (2003). Note that this result is not sufficient to obtain uniform stability estimates, as in Nekhoroshev theorem below. More precise normal form results are given in Nekhoroshev (1977) and Pöschel (1993).

### Normal Form Theorem

*Let $\omega_0 = \omega(p_0)$ be a given Diophantine or resonant-Diophantine frequency. Let us denote $B_r(p_0)$ the open ball of radius $r$ in $\mathbb{R}^d$ centered at $p_0$. There exists a constant $a$ which depends only on $\omega$, and constants $\epsilon_0 > 0$ and $C > 0$ such that the following holds: for each $\epsilon < \epsilon_0$, there exists an analytic symplectic embedding $\phi_\epsilon : \mathbb{T}^d \times B_{r(\epsilon)} \longrightarrow \mathbb{T}^d \times U$, which is $\epsilon$-close to identity and such that*

$$H_\epsilon \circ \phi_\epsilon(q, p) = h(p) + \epsilon N_\epsilon(q, p) + \mu(\epsilon) F_\epsilon(q, p)$$

*where $N$ is in $R(\omega_0)$-normal form, $r(\epsilon) \geq \sqrt{\epsilon}$, and $\mu(\epsilon) \leq e^{-C\epsilon^{-a}}$.*

This means that the motions with resonant initial conditions are confined, up to small oscillations, in the associated affine plane $p(0) + R(\omega(p(0))$ until they live in the domain of the normal form, or until time $\mu^{-1}(\epsilon)$.

### Geometry of Resonances

In view of the normal form theorem, we are led to consider the curves $P(\theta) : \mathbb{R} \longrightarrow \mathbb{R}^d$ which satisfy

$$P(\theta') - P(\theta) \in R(\omega(P(\theta)))$$

for each $\theta$ and $\theta'$. Indeed, it appears that these curves are the ones the action variables can follow on timescales not involving the remainders of the normal forms. Note that here the parameter $\theta$ is not the physical time. Assuming that $P(\theta)$ is such a curve, we can define the affine space

$$R := P(0) + \cap_{\theta \in \mathbb{R}} R(\omega(P(\theta)))$$

We then have $P(\theta) \in R$ for each $\theta$. In addition, each point $P(\theta), \theta \in \mathbb{R}$, is a critical point of the restriction $h_{|R}$ of the unperturbed Hamiltonian $h$ to the affine space $R$. It follows that the curve $P(\theta)$ has to be constant if the unperturbed Hamiltonian satisfies the following hypothesis.

### Nekhoroshev Steepness

*We say that the unperturbed Hamiltonian $h$ is steep if, for each affine subspace $\Lambda$ in $\mathbb{R}^d$, the restriction $h_{|\Lambda}$ has only isolated critical points.*

This formulation, due to Niederman, is much simpler than the equivalent one first given by Nekhoroshev. It turns out that this condition, which was made natural by our heuristic explanation, implies stability over exponential timescales for all initial conditions (see Nekhoroshev (1977)). We first need another condition.

### Kolmogorov Nondegeneracy

*We say that the unperturbed Hamiltonian $h$ is nondegenerate in the sense of Kolmogorov if it has nondegenerate Hessian at each point, or equivalently if the frequency map $p \longmapsto \omega(p)$ is an immersion.*

### Nekhoroshev Stability Theorem

*Assume that the unperturbed Hamiltonian does not have critical points ($\omega(p)$ does not vanish), satisfies Nekhoroshev steepness and Kolmogorov nondegeneracy conditions. Then there exists constants $a > 0$ and $b > 0$, which depend only on $h$, and constants $\epsilon_0 > 0$ and $C > 0$ such that the following holds: for $\epsilon < \epsilon_0$, each trajectory $(q_\epsilon(t), p_\epsilon(t))$ satisfies the estimate*

$$\|p_\epsilon(t) - p_\epsilon(0)\| \leq C\epsilon^b$$

*for all t such that* $|t| \leq e^{C\epsilon^{-a}}$.

### Herman's Example

In order to illustrate the necessity of the condition of steepness, let us consider the Hamiltonian

$$H_\epsilon(q_1, q_2, p_1, p_2) = p_1 p_2 + \epsilon V(q_1)$$

with $V : \mathbb{T} \longrightarrow \mathbb{R}$. The associated equations are

$$\dot{p}_2 = 0, \quad \dot{p}_1 = -V', \quad \dot{q}_1 = p_2, \quad \dot{q}_2 = p_1$$

The trajectories whose initial conditions are subjected to $p_2(0) = 0$ and $V'(q_1(0)) \neq 0$ satisfy

$$p_1(t) = p_1(0) - t\epsilon V'(q_1(0))$$
$$p_2(t) = 0, \qquad q_1(t) = q_1(0)$$

We see an evolution at speed $\epsilon$ of the action variable $p_1$ contradicting the conclusion of Nekhoroshev theorem. In this example, we have $R(\omega(p(t))) = \mathbb{R} \times \{0\}$, and $h_{|\mathbb{R} \times \{0\}} \equiv 0$, so that the curve

$$P(\theta) = (\theta, 0)$$

is indeed a curve of critical points of $h_{|\mathbb{R} \times \{0\}}$.

### Genericity of Steepness

The condition of steepness is frequently satisfied. In order to be more precise, we mention that, for $N \in \mathbb{N}$ large enough (how large depends on the dimension $d$), steepness is a generic condition in the finite-dimensional space of polynomials of degree less than $N$. Note in contrast that a quadratic Hamiltonian is steep if and only if it is positive definite. Finally, it is important to mention that convex Hamiltonians $h$ with positive-definite Hessian are steep. More generally, quasiconvex Hamiltonians are steep. A function $h : U \longrightarrow \mathbb{R}$ is said to be quasiconvex if, at each point, the restriction of its Hessian to the kernel of its differential is positive definite.

### The Quasiconvex Case

It is interesting to be more precise about the values of $a$ and $b$ in Nekhoroshev theorem. We shall do so in the quasiconvex case, which is the most stable case, and where much more is known. If $h$ is quasiconvex, one can take

$$a = b = \frac{1}{2d}$$

as was proved by Lochak (1992). It is a question of active present research whether these exponents are optimal. It now appears that this is almost so, and that the optimal exponent $a$ should not be larger than $1/2(d-3)$. That this exponent deteriorates as the dimension increases is of course very natural in the perspective of statistical mechanics. As a matter of fact, not only the exponent $a$ but also the threshold $\epsilon_0$ of validity of Nekhoroshev theorem deteriorates with the dimension, as was noticed in Bourgain and Kaloshin.

Another important fact was proved in Lochak (1992): in these expressions, the important value of $d$ is not the total number of degrees of freedom, but the number of active degrees of freedom. More precisely, resonant initial conditions are more stable than generic ones. If $r$ is the order of resonance of a given initial condition, then the number $d - r$ of fast angles can be substituted to the total number of degrees of freedom for the computation of the stability exponent. This phenomenon may account for the surprising stability obtained numerically by Fermi, Pasta, and Ulam.

## Permanent Stability

Many initial conditions satisfy more than exponential stability: they are permanently stable.

### Kolmogorov Theorem

*Assume that h satisfies Kolmogorov nondegeneracy condition* ("Kolmogorov nondegeneracy"). *Then for each open subset* $V \subset \mathbb{R}^d$ such that $\bar{V} \subset U$, *there exists* $\epsilon_0 > 0$ *such that, for each* $\epsilon < \epsilon_0$, *there exists*

- *a smooth symplectic embedding* $\phi_\epsilon : \mathbb{T}^d \times V \longrightarrow \mathbb{T}^d \times U$, *which is* $\epsilon$*-close to the identity*,
- *a compact subset* $F_\epsilon$ *of* $V$, *whose relative measure in* $V$ *is converging to 1 as* $\epsilon \longrightarrow 0$,

*such that the Hamiltonian system* $H_\epsilon \circ \phi_\epsilon$ *preserves the torus* $\mathbb{T}^d \times \{p\}$ *for each* $p \in F_\epsilon$.

The union

$$\mathcal{F}_\epsilon = \phi_\epsilon(\mathbb{T}^d \times F_\epsilon)$$

of all the invariant tori has positive measure. Its complement is usually an open dense subset of $\mathbb{T}^d \times U$. All the orbits starting in this invariant set obviously undergo oscillations of amplitude of the order of $\epsilon$ for all times. It is worth mentioning that some energy surfaces may not intersect the invariant set $\mathcal{F}_\epsilon$. This is illustrated in example, i.e., "Herman's example," where the surface of zero energy does not contain invariant tori. The following condition guarantees the existence of invariant tori on each energy surface.

#### Arnol'd Nondegeneracy

*The Hamiltonian $h$ is said to be nondegenerate in the sense of Arnol'd if it does not have critical points and if the map*

$$p \longmapsto \frac{\omega(p)}{\|\omega(p)\|}$$

*is a local diffeomorphism between each level set of $h$ and $S^{d-1}$. This is equivalent to say that the function $(\lambda, p) \in \mathbb{R} \times U \longmapsto \lambda h(p)$ has nondegenerate Hessian at each point of the form $(1, p)$.*

#### Arnol'd Theorem

If $h$ satisfies Arnol'd nondegeneracy condition, then the relative measure of the set $\mathcal{F}_\epsilon$ of invariant tori is converging to 1 in each energy surface.

This theorem prevents ergodicity of the perturbed systems for the canonical invariant measure on its energy surface. This may be considered as a very disappointing result for statistical mechanics, whose mathematical foundation has often been considered to be the Boltzmann hypothesis of ergodicity. However, statistical mechanics is first of all a question of letting $d$ go to infinity, and ergodicity might not be such a crucial hypothesis (see Khinchin).

When $d = 2$, the Arnol'd theorem has particularly strong consequences. Indeed, in this case, the invariant tori cut the energy surfaces in small connected components. The motion is then confined in these connected components. As a consequence, we obtain permanent stability for all initial conditions.

In higher dimensions however, the complement of $\mathcal{F}_\epsilon$ in each energy shell is usually a dense, connected open set. There may exist orbits wandering in this large connected set, although the speed of evolution of these orbits is limited by Nekhoroshev theory. Understanding the dynamics in this open set is a very important and difficult question. It is the subject of the next section.

#### Relaxed Assumption

For many applications, such as celestial mechanics, the nondegeneracy conditions of Arnol'd or Kolmogorov are not satisfied, or difficult to check. However, the existence of invariant tori has been proved under much milder assumptions. As a rule, invariant tori exist in the perturbed systems if the frequency map $p \longmapsto \omega(p)$ stably contains Diophantine vectors in its image.

## The Mechanism of Arnol'd

Understanding instability is the subject of intense present research. General methods of construction of interesting orbits as well as clever classes of examples are being developed. These methods are exploring the limits of stability theory. Here we shall only describe the fundamental ideas of Arnol'd (see Arnol'd 1964), where most of the present activity finds its roots. Although these ideas have some ambition of universality, they are best presented, like in Arnol'd (1964), on an example. We consider the quasiconvex Hamiltonian

$$\begin{aligned} H(q_1, q_2, q_3, p_1, p_2, p_3) \\ = (p_1^2 + p_2^2)/2 - p_3 + \epsilon \cos 2\pi q_2 \\ + \mu(\cos 2\pi q_2)(\cos 2\pi q_1 + \cos 2\pi q_3) \end{aligned}$$

As we have seen, this system is typical of the kind of Hamiltonians one gets after reduction to resonant normal form. However, it is illuminating to consider $\mu$ not as a function of $\epsilon$ but as an independent parameter. This is an idea of Poincaré then followed by Arnol'd. We shall expose the main steps of the proof of the following result.

#### Theorem

*Let us fix numbers $0 < A < B$. For each $\epsilon > 0$, there exists a number $\mu_0(\epsilon)$ such that, when $0 < \mu < \mu_0(\epsilon)$, there exists a trajectory*

$$(q_1(t), q_2(t), p_1(t), p_2(t))$$

*and a time $T > 0$ (which depends on $\epsilon$ and $\mu$) such that*

$$p_1(0) \leq A, \quad p_1(T) \geq B$$

#### The Truncated System

Let us begin with some remarks about the truncated Hamiltonian obtained when $\mu = 0$:

$$\begin{aligned} H_0(q, p) &= H_1(q_1, q_3, p_1, p_3) + H_2(q_2, p_2) \\ &= p_1^2/2 - p_3 + p_2^2/2 + \epsilon \cos 2\pi q_2 \end{aligned}$$

This system is the uncoupled product of $H_1$ and of the pendulum described by $H_2$. The variable $p_1$ is constant along motion; hence, the theorem can not hold for $\mu = 0$.

Recall that the point $q_2 = 0, p_2 = 0$ is a hyperbolic fixed point of the pendulum $H_2(q_2, p_2) = p_2^2/2 + \epsilon \cos 2\pi q_2$. The stable and unstable manifolds of this integrable system coincide; they form the energy level $H_2 = \epsilon$. As a consequence, in the product system of Hamiltonian $H_0 = H_1 + H_2$, there exists,

in the zero energy level, a one-parameter family $T_\omega$ of invariant tori of dimension 2:

$$T_\omega = \{p_1 = \omega, p_3 = \omega^2/2 + \epsilon, q_2 = 0, p_2 = 0\} \subset \mathbb{T}^3 \times \mathbb{R}^3$$

Each of these tori is hyperbolic in the sense that it has a stable manifold of dimension 3 and an unstable manifold of dimension 3, which are nothing but the liftings of the stable and unstable manifolds of the hyperbolic fixed point of $H_2$. Notice that these manifolds do not intersect transversally along $T_\omega$.

When $\mu \neq 0$, the perturbation is chosen in such a way that the tori $T_\omega$ are left invariant by the Hamiltonian flow.

### Splitting

*For* $0 < \mu < \mu_0(\epsilon)$, *the invariant tori* $T_\omega$ *still have stable and unstable manifolds of dimension 3. These stable and unstable manifolds intersect transversally in the energy surface*, *along an orbit which is homoclinic to the torus.*

The first point is that the tori remain hyperbolic, and that the stable and unstable manifolds are deformed, but not destroyed by the additional term. This results from the observation that the manifold $M$ formed by the union of the invariant tori is normally hyperbolic in its energy surface. Note that this step does not require exponential smallness of $\mu$.

It is then a very general result that the stable and unstable manifolds have nonempty intersection. It is a global property, which can be established by variational methods, and which still does not rest on exponential smallness of $\mu$.

The key point, where exponential smallness is required, is transversality. Since transversality is a generic phenomenon, one may think that this step is not so crucial. And indeed, it is very likely that the statement remains true for most values of $\mu \in ]0, \epsilon]$ (and not only for $\mu \leq \mu_0(\epsilon)$). However, there are two important issues here. First, transversality is difficult to establish on explicit examples. Second, it is useful for many further discussions to obtain some quantitative estimates.

Indeed, we can associate to the intersection between the stable and unstable manifolds a quantity, the splitting, which in a sense measures transversality. Discussions on such a definition are available in Lochak *et al.* (2003). Using methods of Poincaré and Melnikov, Arnol'd showed that this splitting can be estimated, for sufficiently small $\epsilon$, by

$$\alpha \geq \mu e^{-C/\sqrt{\epsilon}} + O(\mu^2) \qquad [3]$$

This implies non-nullity of the splitting, hence transversality, for small $\mu$.

### Transition Chain

We have established the existence, when $\mu > 0$ is small enough, of a family $T_\omega$ of hyperbolic invariant tori such that the stable manifold $W_\omega^+$ and the unstable manifold $W_\omega^-$ intersect transversally along a homoclinic orbit (but not along $T_\omega$!) for each $\omega$.

A stability argument shows that the stable manifold $W_\omega^+$ of the torus $T_\omega$ intersects transversally the stable manifold $W_{\omega_0}^-$ of the torus $T_{\omega_0}$ when $\omega$ is close enough to $\omega_0$. How close directly depends on the size of the splitting. We obtain heteroclinic orbits between tori close to each other.

Given two values $\omega$ and $\omega'$, we can find a sequence $\omega_i, 1 \leq i \leq N$, such that $\omega_0 = \omega, \omega_N = \omega'$, and $W_i^-$ intersects transversally $W_{i+1}^+$ for all $i$. The associated family $T_{\omega_i}$ of tori is called a transition chain.

The left step consists in proving that some orbits shadow the transition chain. Arnol'd solved this step by a very simple topological argument which, however, does not provide any estimate on the time $T$. He proves the existence of an orbit joining any neighborhood of $T_\omega$ to any neighborhood of $T_{\omega'}$. This ends the proof of the main theorem, since we can chose $\omega$ and $\omega'$ such that $\omega < A < B < \omega'$.

The dynamics associated to hyperbolic tori and transition chains have later been studied more carefully. It particular, a $\lambda$-lemma can be proved in this context, which allows us to conclude that, in a transition chain, the unstable manifold $W_0^-$ of the first torus intersects transversally the stable manifold of the last torus $W_N^+$. These detailed studies also allow us to relate the speed of diffusion to the splitting of the invariant manifolds.

### Diffusion Speed

It is interesting to estimate the speed of evolution of the variable $p_1$, or in other words the time $T$ in the statement. It follows from Nekhoroshev theory that this time $T$ has to be exponentially large as a function of $\epsilon$. In fact, it is possible to prove, either by recent developments on the ideas of Arnol'd exposed above, or more easily by variational methods, (Bessi 1996) that

$$T \leq \frac{e^{C/\sqrt{\epsilon}}}{-\mu \log \mu}$$

for $\mu \leq \mu_0(\epsilon)$. This time is of course highly related to the estimate [3] of the splitting. In addition, Ugo Bessi proved that one can take $\mu_0(\epsilon) = e^{-C/\sqrt{\epsilon}}$. Plugging this value of $\mu$ in the estimate of $T$, we get the estimate $T \leq e^{C/\sqrt{\epsilon}}$ as a function of the only parameter $\epsilon$.

Considering the fact that the orbit we have described goes close to double resonances, this is the best estimate one may hope for in view of the improved Nekhoroshev stability estimates at resonances.

The idea is now well spread that the time of diffusion is exponentially large. However, we point out that, if it is indeed exponentially small as a function of the parameter $\epsilon$, it is only polynomially small as a function of the second parameter $\mu$, as was first understood by P Lochak and proved in Bernard (1996) using the variational method of U Bessi.

## Conclusion

The theories of instability are developing in several directions. One of them is to try to understand the limits of stability, and to test to what extent the stability results obtained so far are optimal. This aspect has quickly developed recently, for example, the optimal stability exponent $a$ for convex systems is almost known. Another direction is to try to give a description of unstable orbits in typical systems. This remains a widely open question.

Let us finally mention that the application of the theories we have presented to concrete systems is very difficult. One of the reasons is that the estimates of the threshold $\epsilon_0$ of validity of Nekhoroshev and KAM theorems that can (painfully) be obtained by inspection in the proofs are very bad, and it is much too bad, for example, to think about applications to the solar systems with the physical values of the parameters.

*See also:* Averaging Methods; Hyperbolic Billiards; KAM Theory and Celestial Mechanics; Separatrix Splitting; Stability Problems in Celestial Mechanics; Stability Theory and KAM; Weakly Coupled Oscillators.

## Further Reading


Arnol'd VI (1964) Instability of dynamical systems with several degrees of freedom. *Soviet Mathematical Doklady* 5: 581–585.

Arnol'd VI *Mathematical Methods of Classical Mechanics*. Springer.

Bernard P (1996) Perturbation d'un hamiltonien partiellement hyperbolique. *Comptes Rendus de l'Academie des Sciences, Serie I*, 323.

Bessi U (1996) An approach to Arnold's diffusion through the calculus of variations. *Nonlinear Analysis* 26(6): 1115–1135.

Bourgain J and Kaloshin V (2005) On diffusion in high-dimensional Hamiltonian systems. *Journal of Functional Analysis* 229(1): 1–61.

Khinchin AI *Mathematical Foundations of Statistical Mechanics*. Dover.

Lochak P (1992) Canonical perturbation theory via simultaneous approximation. *Russian Math. Surveys* 47: 57–133.

Lochak P, Marco JP, and Sauzin D (2003) On the splitting of invariant manifolds in multidimensional near-integrable Hamiltonian systems. *Mem. Amer. Math. Soc.* 163(775).

Nekhoroshev NN (1977) An exponential estimate for the time of stability of nearly integrable Hamiltonian systems. *Russian Math. Surveys* 32: 1–65.

Niedreman L Hamiltonian stability and subanalytic geometry. *Ann. Inst. Fourier*. (in press).

Pöschel J (1993) Nekhoroshev estimates for quasi-convex Hamiltonian systems. *Math. Z.* 213: 187–216.


# Hamilton–Jacobi Equations and Dynamical Systems: Variational Aspects

**A Siconolfi**, Università di Roma "La Sapienza", Rome, Italy



## Overview

Given a continuous Hamiltonian $H(x,p)$ defined on the cotangent bundle of a compact boundaryless manifold, where $x$ and $p$ are the state and the momentum variable, respectively, and satisfying suitable convexity and coercivity assumptions, we consider the family of Hamilton–Jacobi equations

$$H(x, D\phi) = a \qquad [1]$$

with $a$ a real parameter. If, in addition, $H$ is assumed to be smooth, we also consider the Hamilton's equations

$$\dot{\xi} = H_p(\xi, \eta), \qquad \dot{\eta} = -H_x(\xi, \eta) \qquad [2]$$

whose analysis is related to the variational problem of minimizing the action functional

$$\int_I L(\xi, \dot{\xi})\, dt \qquad [3]$$

among all Lipschitz–continuous or, equivalently, continuous piecewise $C^1$ curves defined on $I$ with fixed end points. Here $I$ is a compact interval and $L$, the Lagrangian, is the Fenchel transform of $H$. A "conjugate" flow, named after Euler–Lagrange, is

also defined on the tangent bundle of the underlying manifold.

A connection between [1] and [2] is provided by the classical Hamilton–Jacobi method, which shows that the graph of the differential of any regular, say $C^1$, global solution to [1] is an invariant subset for the Hamiltonian flow. The drawback of this approach is that such regular solutions do not exist in general, even for very regular Hamiltonians.

However, for any continuous Hamiltonian a distinguished value of the parameter $a$ can be detected, denoted by $c$ and qualified, from now on, as critical, for which there are a.e. subsolutions of the corresponding Hamilton–Jacobi equations enjoying some extremality properties. Note that such functions can be equivalently defined as weak solutions, in the viscosity sense, of [1] with $a = c$, or as fixed points of the associated Lax–Oleinik semigroup (see Fathi (to appear)). We do not give these interpretations here to avoid any technicalities.

Even if they are just Lipschitz–continuous on the whole underlying manifold, these extremal subsolutions become of class $C^1$, when restricted on a special compact subset, the same for any of them, say $\mathcal{A}$, and the corresponding differentials coincide on $\mathcal{A}$. More generally, all critical subsolutions, that is, the a.e. subsolutions to [1] with $a = c$, are continuously differentiable on $\mathcal{A}$. This regularity property holds if $H$ is at least locally Lipschitz–continuous in both variables. When, in addition, the Hamiltonian is smooth, so that the Hamiltonian flow is defined, the graph of this common differential defined in $\mathcal{A}$, denoted by $\tilde{\mathcal{A}}$, is an invariant set for the flow, and is foliated by integral curves of [1] possessing some global minimizing properties with respect to the action functional.

The aim of this presentation is to give an explanation of the previously described phenomena occurring at the critical level, and of some related facts, using tools and arguments as simply as possible. We propose a metric approach to the subject and consider as central in our analysis a family of distances, denoted by $S_a$, for any $a \geq c$. We emphasize that such distances can be defined for only continuous Hamiltonians, and the qualitative analysis of the critical subsolutions has an interest independent from the dynamical applications. Indeed, it can be used in other contexts such as in homogenization problems, and the large-time behavior of the viscosity solutions to the time-dependent equation $u_t + H(x, Du) = 0$.

The discovery of the critical value has a history that reflects the dual character of the topics, which has a dynamical as well as a partial differential equation (PDE) interest.

It was probably Ricardo Mañé who first focused his attention on it, at the beginning of the 1980s, in connection with the analysis of integral curves of the Euler–Lagrange flow with some global minimizing properties. The set, previously denoted by $\tilde{A}$, has been found and analyzed by Serge Aubry, in a purely dynamical way, as the union of the supports of such minimizing curves. On the other hand, John Mather (1986) independently defined, in a more general framework, a set, contained in the Aubry set, through a weak approach that utilizes minimal probability measures invariant with respect to the Euler–Lagrange flow. The Mather set is actually the closure of the union of the supports of such measures. We will follow the approach of Aubry (see Fathi (2005b)), and will not introduce the Mather's measures.

In the viscosity solution theory, the critical value has instead been introduced in a famous unpublished paper of P L Lions, S R S Varadhan, and G Papanicolaou (1987), in connection with some periodic homogenization problems for Hamilton–Jacobi equations. It is worth noticing that they consider continuous Hamiltonian, defined on the flat $N$-dimensional torus, without any convexity assumption.

They define the critical value, and show the existence of viscosity solutions to the critical equation by means of an ergodic approximation, that is, by considering the equation $\varepsilon u + H(x, Du) = 0$ and then passing to the limit for $\varepsilon \to 0$. The critical viscosity solutions are used as correctors in the homogenization. They do not perform any qualitative analysis, and if such analysis can be done, and something similar to the Aubry–Mather sets exists for nonconvex Hamiltonian this is still an important open problem.

The two pieces of the picture were pasted together by Fathi (1996) with his weak KAM theory (see Contreras and Iturriaga (1999) and Fathi (2005a) for a general treatment, where the relevance of the extremal critical subsolutions has first been recognized for the analysis of the dynamics, and the Aubry–Mather sets have been characterized as a regularity set for such subsolutions, as described above). Evans and co-workers have been presently using more general PDE methods in weak KAM theory to address some integrability issues and to find a quantum analog (see Evans and Gomes (2001, 2002) and Evans (2004)).

## Critical Value and Extremal Subsolutions

We consider the family of Hamilton–Jacobi equations [1] defined, for simplicity, on the flat torus $\mathbb{T}^N = \mathbb{R}^N/\mathbb{Z}^N$, endowed with the flat Riemannian

metric induced by the Euclidean metric on $\mathbb{R}^N$. The tangent, as well as the cotangent bundle of $\mathbb{T}^N$ will be identified with $\mathbb{T}^N \times \mathbb{R}^N$. All the results discussed in the remainder of the paper are still true in any compact boundaryless manifold, and some of them also hold in noncompact manifolds. We require $H$ to be continuous in both variables, to satisfy the coercivity assumption,

$$\{(y,p)\colon H(y,p) \le a\} \text{ is compact for any } a$$

and the following (strict) quasiconvexity conditions for any $x \in \mathbb{T}^N$, $a \in \mathbb{R}$:

$$\{p\colon H(y,p) \le a\} \text{ is strictly convex}$$
$$\partial\{p\colon H(y,p) \le a\} = \{p\colon H(y,p) = a\}$$

where $\partial$, in the above formula, indicates the boundary. We denote by $\mathcal{S}_a$ the (possibly empty) set of the Lipschitz–continuous a.e. subsolutions to [1]. They will be called in the sequel, for short, just subsolutions. Due to the convex character of the Hamiltonian and its continuity, the property of being a subsolution, for some function $u$, can be equivalently expressed by requiring the inequality $H(x,p) \le a$ to hold for any $x \in \mathbb{T}^N$ and any $p$ in the (Clarke) generalized gradient $\partial u(x)$, defined by

$$\partial u(x) = \mathrm{co}\{p = \lim_i Du(x_i)\colon x_i \text{ differentiability point of } u, \lim_i x_i = x\}$$

where co indicates the convex hull. Note that if this set of weak derivatives reduces to a singleton at some $x$, then the function $u$ is strictly differentiable at $x$, i.e., it is differentiable and $Du$ is continuous at $x$.

By a strict subsolution to [1] we mean a Lipschitz–continuous function $w$ with $\mathrm{ess\,sup}_{\mathbb{T}^N} H \times (x, Dw(x)) < a$. The property of being a (strict) subsolution is not affected by addition of constants. Moreover, the pointwise supremum (resp. infimum) of any class of equibounded subsolution to [1] is itself a subsolution, and $\mathcal{S}_a$ is stable with respect to the uniform convergence in $\mathbb{T}^N$.

The purpose of this section is to show that there is a unique value $c$ (the critical value) for which the corresponding equation

$$H(x, Du) = c \qquad [4]$$

possesses subsolutions enjoying some extremality properties. We, more precisely, call a subsolution $u \in \mathcal{S}_a$ maximal (resp. minimal) if for any open subset $\Omega$ of $\mathbb{T}^N$ and any Lipschitz–continuous function $\phi$ with

$$u = \phi \quad \text{on } \partial\Omega \quad \text{and} \quad \mathrm{ess\,sup}_\Omega H(x, D\phi(x)) < a \qquad [5]$$

one has $u \ge \phi$ (resp. $u \le \phi$) in $\Omega$.

Any maximal (resp. minimal) subsolution $u$ is actually an a.e. solution of [1]. If, in fact, $H(x_0, Du(x_0)) < a$ for some differentiability point $x_0$ of $u$, then the function $\phi(x) = u(x_0) + Du(x_0)(x - x_0) - \varepsilon|x - x_0| + \varepsilon$ (resp. $\phi(x) = u(x_0) + Du(x_0)(x - x_0) + \varepsilon|x - x_0| - \varepsilon$) should satisfy [5] for a suitable choice of $\varepsilon > 0$ and of a neighborhood $\Omega$ of $x_0$, and so should violate the maximality (resp. minimality) condition for $u$.

The previous argument can be easily adapted to show something more general: if $u$ is a maximal (resp. minimal) subsolution then no subtangents (resp. supertangents) to $u$ at any $y \in \mathbb{T}^N$ can be local strict subsolutions at $y$, that is, strict subsolutions in some neighborhood of $y$.

The subtangency (resp. supertangency) condition of a function $\phi$ to $u$ at a point $x_0$ means that $x_0$ is a local minimizer (resp. maximizer) of $u - \phi$. We denote by $D^-u(x_0)$ (resp. $D^+u(x_0)$) the sets made up by the differentials of the $C^1$-subtangent (resp. supertangent) to $u$ at $x_0$. They are (possibly empty) closed convex subsets of $\partial u(x_0)$. It is apparent that if $D^+u(x_0) \ne \emptyset \ne D^-u(x_0)$ then $u$ is differentiable at $x_0$ and $D^+u(x_0) = D^-u(x_0) = \{Du(x_0)\}$.

It is an immediate consequence of the previous fact that no extremal subsolutions can exist in $\mathcal{S}_a$, whenever [1] admits a strict subsolution, say $\phi$, since there are global minimizers and maximizers of $u - \phi$, for any $u \in \mathcal{S}_a$, because of the compactness of $\mathbb{T}^N$. The function $\phi$ is then subtangent and supertangent, respectively, to $u$ at such points.

The unique value we can look at for finding extremal subsolutions is therefore

$$c = \inf\{a \in \mathbb{R}\colon \mathcal{S}_a \ne \emptyset\} \qquad [6]$$

The set on the right-hand side of [6] is nonempty since the null function belongs to $\mathcal{S}_a$ when $a > \max_{\mathbb{T}^N} H(x,0)$, and bounded from below by $\min_{\mathbb{T}^N} H(x,0)$. The value $c$ is consequently well defined by [6].

Moreover, any sequence $u_n \in \mathcal{S}_{a_n}$, with $a_n$ decreasing and convergent to $c$, is equi-Lipschitz–continuous because of the coercivity of $H$, and equibounded, up to addition of suitable constants. It is therefore uniformly convergent, up to a subsequence, to some $u$, which belongs to $\mathcal{S}_{a_n}$, for any $n$, since these classes are stable for the uniform convergence. This implies that $u$ is a subsolution to [4], so that $\mathcal{S}_c \ne \emptyset$. The critical value $c$ is then characterized by the property that the corresponding eqn [4] admits subsolutions but not strict subsolutions. Our aim is to show that extremal subsolutions do exist for the critical eqn [4].

For any supercritical value $a$, that is, $a \geq c$, we can define the functional nonsymmetric semidistance:

$$S_a(y,x) = \sup\{u(x) - u(y)\colon u \in \mathcal{S}_a\} \\ = \sup\{u(x)\colon u \in \mathcal{S}_a, u(y) = 0\}$$

for any $x$, $y$ in $\mathbb{T}^N$. It is immediate that $S_a$ satisfies the triangle inequality and $S_a(y,y) = 0$ for any $y$. But it fails, in general, to be symmetric and positive if $x \neq y$. We will nevertheless call it a distance, in the sequel, to ease terminology. The function $x \mapsto S_a(y,x)$ is itself a subsolution to [1], for any $y$, being the pointwise supremum of a family of equibounded subsolutions. Taking into account the inequality

$$u(x) - u(y) \geq -S_a(x,y)$$

which holds for any $u \in \mathcal{S}_a$, and the fact that it becomes an equality by setting $u = S_a(x,\cdot)$, we also get

$$-S_a(x,y) = \inf\{u(x) - u(y)\colon u \in \mathcal{S}_a\} \\ = \inf\{u(x)\colon u \in \mathcal{S}_a, u(y) = 0\}$$

and $-S_a(\cdot,y)$ is, as well, a subsolution to [1]. Note that

$$S_a(x,y) + S_a(y,x) \geq 0 \quad \text{for any } y, x \qquad [7]$$

The interest of introducing the distance $S_a$ in the present context is that, for any $a \geq c$ and $y \in \mathbb{T}^N$, the function $x \mapsto S_a(y,x)$ (resp. $x \mapsto -S_a(x,y)$) satisfies the maximality (resp. minimality) condition for subsolutions of [1] in any open set not containing $y$. If, by contradiction, the maximality property of $S_a(y,\cdot)$ were violated in some open set $\Omega$ with $y \notin \Omega$ by a $\phi$ satisfying [5] then one could make the set $\{x\colon \phi(x) > S_a(y,x)\}$ nonempty and compactly contained in $\Omega$, by adding a suitable constant. Hence, the formula

$$u = \begin{cases} \max\{\phi, S_a(y,\cdot)\} & \text{in } \Omega \\ S_a(y,\cdot) & \text{otherwise} \end{cases} \qquad [8]$$

could provide a subsolution to [1] with $u(y) = S_a(y,y) = 0$ and $u > S_a(y,\cdot)$ at some point of $\Omega$, which is in contrast with the very definition of $S_a$. One can similarly prove the minimality condition for $-S_a(\cdot,y)$.

We now focus our attention on the critical case. We derive from the previous considerations that if a maximal subsolution to [4] does not exist then, for any $y$, we can find a neighborhood $\Omega'_y$ of $y$ where $S_c(y,\cdot)$ fails to be maximal. We can thus construct, through a formula like [8], a $u_y \in \mathcal{S}_c$ with

$$\text{ess sup}_{\Omega_y} H(\cdot, Du_y(\cdot)) < c \qquad [9]$$

in some neighborhood $\Omega_y$ of $y$ contained in $\Omega'_y$. Thanks to the compactness of $\mathbb{T}^N$, we can extract from $\{\Omega_y\}$ a finite subcover $\{\Omega_{y_i}\}, i = 1, \ldots, m$, for some $m \in \mathbb{N}$, and define

$$u = \sum_i \lambda_i u_{y_i}$$

where $\lambda_i$ are positive constants with $\sum_1^m \lambda_i = 1$. The convex character of the Hamiltonian and [9] imply that $u$ is a strict critical subsolution, which cannot be. We therefore conclude that there is a nonempty subset of $y$, denoted henceforth by $\mathcal{A}$, for which $S_c(y,\cdot)$ is indeed a maximal critical subsolution. It can also be proved, by exploiting some stability properties of the maximal subsolutions, that $\mathcal{A}$ is closed. Similarly, $-S_a(\cdot,y)$ must be a minimal critical subsolution for some $y$. We denote by $\bar{\mathcal{A}}$ the closed set made up by such points.

The previous covering argument shows that if $y \notin \mathcal{A}$ (resp. $y \notin \bar{\mathcal{A}}$) then there is a local strict critical subsolution at $y$. The converse is also true: let in fact $\phi$ be such a strict subsolution satisfying $\phi(y) = S_c(y,y) = 0$; then $\phi$ is subtangent to $S_c(y,\cdot)$ (resp. supertangent to $-S_c(\cdot,y)$) at $y$, by the very definition of the distance $S_c$. This shows that $S_c(y,\cdot)$ (resp. $-S_c(\cdot,y)$) is not a maximal (resp. minimal) critical subsolution, and so $y \notin \mathcal{A}$ (resp. $y \notin \bar{\mathcal{A}}$). Since the previous characterization holds for both $\mathcal{A}$ and $\bar{\mathcal{A}}$, it follows that $\mathcal{A} = \bar{\mathcal{A}}$. This set is a generalization of the (projected) Aubry set. We will come back on this point later on.

We also see from the covering argument that there is a critical subsolution $\phi$, which is strict outside $\mathcal{A}$, that is, such that $\text{ess sup}_\Omega H(x, D\phi(x)) < c$ for any open set $\Omega$ compactly contained in $\mathbb{T}^N \backslash \mathcal{A}$.

This implies that any $y$ such that $\{p\colon H(y,p) \leq c\}$ has empty interior, belongs to $\mathcal{A}$. The empty interior condition in fact implies, thanks to the strict quasiconvexity of $H$, that the sublevel set reduces to a singleton, say $\{p_0\}$. We know that $\partial u(y) \subset \{p\colon H(y,p) \leq c\}$, for any $u \in \mathcal{S}_c$; therefore, $\partial u(y)$ is a singleton and so any critical subsolution $u$ is strictly differentiable at $y$ with $H(y, Du(y)) = H(y, p_0) = c$. Hence, there cannot be critical subsolutions which are strict around $y$.

The previously described points will be called, in the sequel, equilibria, and the (possibly empty) closed set made up by them will be denoted by $\mathcal{E}$. The reason of this terminology will be explained later. The differentiability property of the critical subsolutions at equilibria, can be extended, quite surprisingly, to any point of $\mathcal{A}$, under more stringent assumptions on $H$. We will discuss this issue in the next section.

## Qualitative Properties of Generalized Aubry Set

We introduce some dynamical aspects in the picture by showing that the distances $S_a$, defined in the previous section for any $a \geq c$, are actually of length type, in the sense that $S_a(y,x)$ equals, for any pair $y$, $x$, the infimum of the intrinsic length of absolutely continuous, or equivalently Lipschitz-continuous, curves joining $y$ to $x$. By intrinsic length, we mean the total variation of $S_a$ on the curve. It will be denoted by $\ell_a$, while $\ell$ will indicate the natural (i.e., Euclidean) length.

For this purpose, we proceed to give a line-integral representation formula of $S_a$. To start with, we consider a $C^1$ subsolution $u$ to [1], some $x$, $y$ $\mathbb{T}^N$ and a (Lipschitz-continuous) curve $\xi$, defined in some compact interval $I$, joining $y$ to $x$. We have

$$u(x) - u(y) = \int_I Du\,\dot{\xi}\,\mathrm{d}t \leq \int_I \sigma_a(\xi,\dot{\xi})\,\mathrm{d}t \qquad [10]$$

where, for any $(x,v) \in \mathbb{T}^N \times \mathbb{R}^N, \sigma_a(x,v) := \max_{p\in Z_a(x)} pv$ and

$$Z_a(x) := \{p\colon H(x,p) \leq a\}$$

Inequality [10] also holds for a Lipschitz-continuous subsolution to [1] through suitable replacement of the differential by the generalized gradient. The set-valued map $Z_a$ is compact convex valued, by the coercivity and quasiconvexity assumptions on $H$, and continuous with respect to the Hausdorff metric. The function $\sigma_a$ is accordingly continuous in the first variable, and convex and positively homogeneous in the second, being a support function. This implies in particular that the integral on the right-hand side of [10] is invariant under change of parameter preserving the orientation. We derive, from [10],

$$S_a(y,x) \leq \inf\left\{ \int_0^1 \sigma_a(\xi,\dot{\xi})\,\mathrm{d}t\colon \xi \text{ defined in } [0,1] \text{ and joining } y \text{ to } x \right\} \qquad [11]$$

for any $y, x$. We denote by $\overline{S}_a(y,x)$ the quantity on the right-hand side of [11]. It is immediate that the triangle inequality holds for $\overline{S}_a$. The function $u := \overline{S}_a(y,\cdot)$ is, moreover, Lipschitz-continuous since $\sigma_a(x,v)/|v|$ is bounded from above in $\mathbb{T}^N \times (\mathbb{R}^N \backslash \{0\})$ because of the coercivity of $H$. Given $v \in \mathbb{R}^N$, we exploit the definition of $\overline{S}_a$, the continuity of $\sigma_a$, and the triangle inequality for $\overline{S}_a$, to get at any differentiability point $x_0$ of $u$,

$$\begin{aligned} Du(x_0)v &= \lim_{h\to 0^+} \frac{u(x_0 - hv) - u(x_0)}{h} \\ &\leq \limsup_{h\to 0^+} \frac{\overline{S}_a(x_0 - hv, x_0)}{h} \\ &\leq \lim_{h\to 0^+} \frac{1}{h}\int_0^1 \sigma_a(x_0 - hvt, hv)\,\mathrm{d}t \\ &= \lim_{h\to 0^+} \int_0^1 \sigma_a(x_0 - hvt, v)\,\mathrm{d}t \\ &= \sigma_a(x_0, v) \end{aligned}$$

This implies by Hahn–Banach theorem that $Du(x_0) \in Z_a(x)$ or, in other terms, that $u = \overline{S}_a(y,\cdot) \in \mathcal{S}_a$. We then derive, from [11] and the very definition of $S_a$,

$$S_a(y,x) = \inf\left\{ \int_0^1 \sigma(\xi,\dot{\xi})\,\mathrm{d}t\colon \xi \text{ defined in } [0,1] \text{ with } \xi(0) = y,\ \xi(1) = x \right\}$$

Taking into account that the integral functional appearing in the previous formula is lower semicontinuous for the uniform convergence of equi-Lipschitz-continuous sequence of curves, by standard variational results, we in turn infer that it equals the intrinsic length $\ell_a$. Mathematically,

$$\ell_a(\xi) = \int_I \sigma_a(\xi,\dot{\xi})\,\mathrm{d}t$$

for any compact interval $I$ and any curve $\xi$ defined in $I$.

Since $S_a$ is just a semidistance, we do not have any *a priori* information on the sign of $\ell_a$; however, by [10], the intrinsic length of any cycle must be nonnegative. Furthermore, while $|\ell_a(\xi)|$ must be small for any curve $\xi$ with small natural length, by the coercivity condition on $H$, no converse estimates hold, in general. If $a > c$, some information in this direction can be gathered by taking a strict subsolution $\phi$ to [1], that it can be assumed smooth, up to regularization by mollification, then $D\phi(x)v \leq \sigma_a(x,v) - \rho|v|$ for any $(x,v) \in \mathbb{T}^N \times \mathbb{R}^N$, and some $\rho > 0$, and consequently

$$\ell_a(\xi) \geq \int_I (\sigma_a(\xi,\dot{\xi}) - D\phi(\xi)\dot{\xi})\,\mathrm{d}t + \phi(x) - \phi(y) \geq \rho\ell(\xi) - S_a(x,y) \qquad [12]$$

for any pair $y$, $x$ and any curve $\xi$, defined in some interval $I$, joining $y$ to $x$. The previous formula says, in particular, that when $|x-y|$ is small then any curve whose intrinsic length approximates $S_a(y,x)$ must have small natural length. The previous

argument cannot be extended to the critical case. This gap suggests the next definition. The main purpose for introducing it is to get a metric characterization of the Aubry set $\mathcal{A}$.

We say that $S_c$ is localizable at some $y$ if for every $\varepsilon > 0$ there is $0 < \delta_\varepsilon < \varepsilon$ such that

$$S_c(y,x) = \inf\{\ell_c(\xi)\colon \xi \text{ joins } y \text{ to } x \text{ and } \ell(\xi) < \varepsilon\} \quad [13]$$

whenever $|x-y| < \delta_\varepsilon$. If $y \notin \mathcal{A}$, we adapt the argument previously used in the strict subcritical case to get that $S_c$ is indeed localizable at $y$. In this case we have, in fact, at our disposal a critical subsolution, say $\phi$, which is strict in some neighborhood $\Omega$ of $y$, thanks to the characterization of the Aubry set given in the previous section.

We assume, to simplify, $\phi$ to be $C^1$; under the natural condition of Lipschitz-continuity, generalized gradients should be used in place of differentials. We have $D\phi(x)v \le \sigma(x,v) - \rho|v|$ for any $x \in \Omega$, any $v \in \mathbb{R}^N$, and some $\rho > 0$, and $D\phi(x)v \le \sigma(x,v)$, for any $x$, $v$. Exploiting these inequalities, we obtain an estimate analogous to [12] for curves starting from $y$, which allows us to prove [13].

Conversely, let $y \notin \mathcal{E}$ be a point where $S_c$ is localizable. We claim that $Z_c(y) \subset D^- u(y)$, where $u := S_c(y,\cdot)$. It is enough to show that any $p_0$ in the interior of $Z_c(y)$ belongs to $D^-u(y)$, since $D^-u(y)$ is closed. Note that the interior of $Z_c(y)$ is nonempty since we are assuming that $y$ is not an equilibrium. Such a $p_0$ belongs to the interior of $Z_c(x)$ for $x$ sufficiently close to $y$, thanks to the continuity of $Z_c$; consequently, $p(x-y) < \ell_c(\xi)$ for any $x$ close to $y$ and any curve $\xi$ joining $y$ to $x$ with $\ell(\xi)$ sufficiently small. Taking into account [13], we then deduce

$$p(x-y) \le S_c(y,x) \quad \text{for } x \text{ close to } y$$

and so the linear function $\phi(x) := p_0(x-y)$ is subtangent to $u$ at $y$. This in turn implies that $y$ is out of $\mathcal{A}$ since $\phi$ is a local strict critical subsolution at $y$, and so $S_c(y,\cdot)$ cannot be a maximal subsolution by the characterization given in the previous section.

The fact that $S_c$ is not localizable at any point of $y \in \mathcal{A}\backslash\mathcal{E}$ leads to the announced metric characterization of $\mathcal{A}$. If $y$ is such a point, there is an $\varepsilon > 0$, a point $x$, with $|x-y|$, and so $|S_c(y,x)|$, as small as desired, and a curve $\xi$ joining $y$ to $x$ with $\ell_c(\xi) \sim S_c(y,x)$ and $\ell(\xi) > \varepsilon$. We construct a cycle $\gamma$, passing through $y$, by juxtaposition of $\xi$ and the Euclidean segment joining $x$ to $y$. We obtain, in this way, a sequence of cycles $\gamma_n$, passing through $y$, with length $\ell_c(\gamma_n) \to 0$ and $\ell(\gamma_n) \ge \varepsilon$, for any $n$.

The same result can also be obtained for $y \in \mathcal{E}$. In this case we select $\varepsilon > 0$ and $v_0 \in \mathbb{R}^N$ with $\sigma_c(y, v_0) = 0$, and denote by $B_n$ a sequence of Euclidean balls, centered at $y$, satisfying $\sigma_c(\cdot, v_0) < 1/n$ in $B_n$. We construct a sequence of cycles, passing through $y$, by going up and down on the line $\{y + sv\}$ in such a way that $\gamma_n(t) \in B_n$, for every $t$, and $\varepsilon < \ell(\gamma_n) < 2\varepsilon$; therefore $0 \le \ell_c(\gamma_n) < 2\varepsilon/n$.

Conversely, such a sequence of cycles cannot exist at any $y \notin \mathcal{A}$ because $S_c$ is localizable at $y$.

We emphasize that the previous definition of $\mathcal{A}$ through cycles and the fact that $S_c$ is not localizable at any point $y \in \mathcal{A}$ with $\mathrm{int} Z_c(y) \neq \emptyset$ shows that, apart for the special case of equilibria, the property of being a point of $\mathcal{A}$ is definitively not of local nature.

As pointed out already, if $y \notin \mathcal{A}$, and so $S_c$ is localizable at $y$, then $Z_c(y) \subset D^-u(y)$, where $u := S_c(y,\cdot)$; on the other hand, we know that $D^-u(y) \subset \partial u(y)$ and $\partial u(y) \subset Z_c(y)$, where the latter inclusion holds since $u$ is a critical subsolution. We then derive

$$D^-u(y) = \partial u(y) = Z_c(y)$$

We interpret these inequalities as a convexity–type property, or, to use a more appropriate terminology, a semiconvexity property of the distance function $S_c(y,\cdot)$ at $y$. The same property holds for the Euclidean distance function $|x|$ at 0.

A contrasting phenomenon takes place if $y \in \mathcal{A}$, namely $S_c(y,\cdot)$ is semiconcave at $y$, which means that $D^+u(y) = \partial u(y)$. This is more complicated to prove (see Fathi 2005b), and requires, in addition, $H$ to be strictly convex in $p$ and locally Lipschitz-continuous in $(x,p)$. Under these assumptions one can, more generally, show that $S_c(y,\cdot)$ is semiconcave in $\mathbb{T}^N$, if $y \in \mathcal{A}$, while it is semiconcave in $\mathbb{T}^N\backslash\{y\}$ and semiconvex in $y$, if $y \notin \mathcal{A}$. Some important consequences can be deduced.

First, thanks to the semiconcavity property there are $C^1$ supertangents to $u := S_c(y,\cdot)$ at $y$, whenever $y \in \mathcal{A}$. Such a function, say $\phi$, is also supertangent to $-S_c(\cdot,y)$, which is a minimal critical subsolution, at the same point. We know from the previous section that no supertangents to $-S_c(\cdot,y)$ at $y$ can be strict critical subsolution locally at $y$, and so $H(y, D\phi(y)) = c$. This implies that $D^+u(y)$ is contained in the boundary of $Z_c(y)$. We then see, taking into account that $D^+u(y)$ is convex and $Z_c(y)$ strictly convex, that $D^+u(y)$ reduces to a singleton, and so, by the semiconcavity property, $\partial u(y)$ reduces to a singleton. Therefore, $S_c(y,\cdot)$ is strictly differentiable at $y$, for any $y \in \mathcal{A}$. One can similarly show that $-S_c(\cdot,y)$ is strictly differentiable at $y$.

Second, given $y \in \mathcal{A}$ and a critical subsolution $w$, which can be assumed, up to addition of a constant, to vanish at $y$, we see that $S_c(y,\cdot)$ (resp. $-S_c(\cdot,y)$) is supertangent (resp. subtangent) at $y$ because of its extremality properties. Since both these

super(sub)-tangents are differentiable, by the previous point, we deduce that $w$ itself is differentiable at $y$. Moreover, the differentials at $y$ of all three functions under consideration, namely $S_c(y,\cdot)$, $-S_c(\cdot,y)$, and $w$, coincide. In particular, $H(y,Dw(y))=c$, and $y\mapsto Dw(y)$ is continuous on $\mathcal{A}$, since $S_c(y,\cdot)$ has been proved to be strictly differentiable at $y$, whenever $y\in\mathcal{A}$. Any critical subsolution, restricted to $\mathcal{A}$, is consequently a continuously differentiable solution to [4].

Summing up, we have discovered (under the assumption of strict convexity and Lipschitz-continuity for $H$) that every critical subsolution is differentiable on $\mathcal{A}$, and the differential on $\mathcal{A}$ is the same for every critical subsolution. A continuous map $G:\mathcal{A}\to\mathbb{R}^N$ is then defined by taking $G(y)$ equal to the common differential of any critical subsolution at $y$. We denote by $\tilde{\mathcal{A}}$ the graph of $G$, which is a subset of the cotangent bundle of $\mathbb{T}^N$, identified with $\mathbb{T}^N\times\mathbb{R}^N$.

As we have already pointed out, the existence of a $C^\infty$ subsolution to [1] is obvious when $a>c$, and such a subsolution can be obtained through a suitable regularization by mollification of any strict subsolution. The same construction cannot be performed at the critical level, since no strict critical subsolutions are available to start the regularization procedure. We can nevertheless show the existence of $C^1$ critical subsolutions by exploiting the information gathered on the Aubry set. We start by considering a countable locally finite open cover of $\mathbb{T}^N\backslash\mathcal{A},\{\Omega_i\}$; we know from the previous section that there is a critical subsolution, say $w_i$, which is strict on $\Omega_i$, for any $i$. Loosely speaking, we have some space, also in this case, for regularizing $w_i$ in such a way that the regularized function is still a critical subsolution, at least on $\Omega_i$.

We can glue together, with some precautions, these regularized local critical subsolutions through a $C^\infty$ partition of the unity, to produce a critical subsolution which is $C^\infty$ outside $\mathcal{A}$. Using the fact that any critical subsolution is differentiable on $\mathcal{A}$, we can further adjust the previous construction so that the critical subsolution is $C^1$ on the whole $\mathbb{T}^N$. We state this result in the following way: if the equation [1] has a subsolution then it also has a $C^1$ subsolution. It is worth noticing that it holds even if the underlying manifold is noncompact (see Fathi (2004, 2005b)).

## The Intrinsic Lengths and the Action Functional

Here we assume $H$ to satisfy all the usual assumptions in order to define the Hamilton's equations [2] and to have the completeness of the associated Hamiltonian flow. Namely, we require $H$ to be $C^2$ in both variables, $C^2$-strictly convex, that is, $H_{pp}>0$ in $\mathbb{T}^N\times\mathbb{R}^N$ and superlinear, in the sense that

$$\lim_{|p|\to+\infty}\frac{H(x,p)}{|p|}=+\infty\quad\text{uniformly in } x$$

We define the Lagrangian $L$ as the Fenchel transform of $H$. It takes finite values thanks to the superlinearity condition, and, in addition, inherits, from $H, C^2$ regularity, $C^2$-strict convexity and superlinearity. In our setting, the Fenchel transform is involutive.

We call a vector $v_0$ and a covector $p_0$ conjugate at a point $x$ if $v_0=H_p(x,p_0)$, and so $L(x,v_0)=p_0v_0-H(x,p_0)$. This also implies the relations $p_0=L_v(x,v_0)$ and $H(x,p_0)=p_0v_0-L(x,v_0)$. If $H(x,p_0)=a$, for some $a$, then $p_0v_0=\sigma_a(x,v_0)$, and $p_0$ is the unique element of $Z_a(x_0)$ for which such a relation holds. Since the function $y\mapsto p_0v_0-H(y,p_0)$ is subtangent to $L(\cdot,v_0)$ at $x$, we see that $L_x(x,v_0)=-H_x(x,p_0)$.

We introduce, for any (Lipschitz-continuous) curve $\xi$ defined in $[a,b]$, for some $a<b$, the action functional $A(\xi)$ through

$$A(\xi)=\int_I L(\xi,\dot\xi)\,\mathrm{d}t$$

We say that the curve $\xi$ is a minimizer of the action if $A(\xi)\le A(\gamma)$ for any $\gamma$ defined $[a,b]$ and with the same end points of $\xi$. It is a classical result in calculus of variations that any of such minimizers $\xi$ is of class $C^2$ and satisfies the Euler–Lagrange equation

$$\frac{\mathrm{d}}{\mathrm{d}t}L_v(\xi,\dot\xi)=L_x(\xi,\dot\xi)\quad\text{in } ]a,b[$$

Consequently, $\xi$ and the conjugate curve $\eta=L_v(\xi,\dot\xi)$ satisfy the Hamilton's equations [2]. Note that all the integral curves of [2] lie in a fixed level of the Hamiltonian, which is compact by the superlinearity condition. The corresponding Hamiltonian flow is consequently complete.

We show that if $x_0\in\mathcal{E}$, and $Z_c(x_0)=\{G(x_0)\}$, then $(x_0,G(x_0))$ is a steady state of the Hamiltonian flow. In this case, in fact, $c=\min_p H(x_0,p)$ and so $L(x_0,0)=-c$ and $H_p(x_0,G(x_0))=0$, or equivalently $G(x_0)$ and 0 are conjugate at $x_0$. Taking into account that $c$ is the critical value, we have that

$$L(x,0)=-\min_p H(x,p)\ge -c\quad\text{for any } x\in\mathbb{T}^N$$

so that $x_0$ is a minimizer of $x\mapsto L(x,0)$ and $L_x(x_0,0)=-H_p(x_0,G(x_0)=0$. It is easy to see that,

conversely, if $(x_0, p_0)$ is a steady state of the Hamiltonian flow and $H(x_0, p_0) = c$ then $x_0 \in \mathcal{E}$ and $p_0 = G(x_0)$.

We want to establish a relation between $A(\cdot)$ and the length functionals $\ell_a$ defined in the previous section for $a \geq c$. This will allow, among other things, to show that the Aubry set $\tilde{\mathcal{A}}$ is invariant for the Hamiltonian flow and to analyze the properties of the integral curves lying on it. To this aim, we consider the minimal geodesics for $S_a, a \geq c$, that is, the curves, defined on compact intervals, whose intrinsic lengths $\ell_a$ equal the distance $S_a$ between their end points.

If $a > c$, we claim that, given any pair of points in $\mathbb{T}^N$, there is a minimal geodesics joining them. Recalling the formula [12], whose validity depends on the fact that in the strict supercritical case there is a smooth strict subsolution to [1], we have

$$\ell_a(\xi) \to +\infty \quad \text{whenever} \quad \ell(\xi) \to +\infty$$

The claim is then proved by using the Ascoli theorem and the lower-semicontinuity property of $\ell_a$. In the critical case, given $y \notin \mathcal{A}$, we can use the same argument to deduce the existence of minimizing geodesics for $S_c$ between $y$ and any point $x$ sufficiently close to $y$ (in the Euclidean sense). This comes from the fact that $S_c$ is localizable at $y$, and so any sequence of curves $\xi_n$ with $\ell_c(\xi_n) \to S_c(y, x)$ has bounded natural length. For a general pair of points, we will show, on the contrary, that existence of a minimal geodesic is not guaranteed in the critical case.

We consider a minimizing geodesics $\xi$ for $S_a$ between a pair of points $y$ and $x$. We assume $a > c$ or $a = c$ and $\xi \cap \mathcal{E} = \emptyset$. We want to show that $\xi$ is a minimizing curve for the action, up to a change of parameter. We choose the new parameter in such a way that

$$L(\tilde{\xi}, \dot{\tilde{\xi}}) + a = \sigma_a(\tilde{\xi}, \dot{\tilde{\xi}}) \qquad [14]$$

where we have denoted by $\tilde{\xi}$ the reparametrized curve. Since $\tilde{\xi}$ stays away from $\mathcal{E}$, the velocities $|\dot{\tilde{\xi}}|$ are bounded from below by a positive constant and so the domain of definition of $\tilde{\xi}$, denoted by $[0, T]$, is a compact interval. Note that $\ell_a(\xi) = \ell_a(\tilde{\xi})$, since the intrinsic length is invariant under change of parameter. We take into account that $\xi$ is a minimal geodesic and the inequality $L(x, v) + a \geq \sigma_a(x, v)$, which holds for any $x$, $v$, to get

$$A(\tilde{\xi}) = \ell_a(\xi) - aT \leq \ell_a(\gamma) - aT \leq A(\gamma)$$

for any $\gamma$ defined in $[0, T]$ with $\gamma(0) = y, \gamma(T) = x$. This proves the announced minimality property of $\tilde{\xi}$.

Furthermore, we show that the function $u := S_a(y, \cdot)$ is strictly differentiable at $\tilde{\xi}(s)$, for $s \in ]0, T[$, and

$$Du(\tilde{\xi}) = L_v(\tilde{\xi}, \dot{\tilde{\xi}}) =: \eta \qquad [15]$$

in $[0, T]$. Hence, $(\tilde{\xi}, Du(\tilde{\xi}))$ is a solution of the Hamilton's equations in $]0, T[$. To see this, we start from the relations

$$\int_0^t \frac{\mathrm{d}}{\mathrm{d}s} u(\tilde{\xi}(s))\, \mathrm{d}s = u(\tilde{\xi}(t)) = \int_0^t \sigma_c(\tilde{\xi}, \dot{\tilde{\xi}})\, \mathrm{d}s = \int_0^t \eta \dot{\tilde{\xi}}\, \mathrm{d}s \qquad [16]$$

which hold in $[0, T]$ because $u$ is Lipschitz-continuous, $\tilde{\xi}$ is a minimizing geodesic, and $\eta(s)$ is conjugate to $\dot{\tilde{\xi}}(s)$ at $\tilde{\xi}(s)$ for any $s \in [0, T]$. We know that

$$\frac{\mathrm{d}}{\mathrm{d}s} u(\tilde{\xi}(s)) = p\dot{\tilde{\xi}}(s)$$
$$\text{for a.e. } s \text{ and some } p \in \partial u(\tilde{\xi}(s))$$

We have that $p \in Z_c(\tilde{\xi}(s))$, since $u$ is a critical subsolution, and so

$$p\dot{\tilde{\xi}}(s) \leq \sigma_c(\tilde{\xi}(s), \dot{\tilde{\xi}}(s)) = \eta(s)\dot{\tilde{\xi}}(s))$$

We see, in the light of [16], that equality must hold in the previous formula, for a.e. $s$. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}s} u(\tilde{\xi}(s)) = \eta(s)\, \dot{\tilde{\xi}}(s) \quad \text{for a.e. } s \qquad [17]$$

we derive from the fact that the function $\eta(\cdot)\dot{\tilde{\xi}}(\cdot)$ is continuous that $(u(\tilde{\xi}(\cdot))$ is actually continuously differentiable in $]0, T[$ and that [17] holds for any $s$. We finally exploit that $u$ is semiconcave in $\mathbb{T}^N \backslash \{y\}$, as pointed out in the previous section, and so $D^+u(\tilde{\xi}(s)) = \partial u(\tilde{\xi}(s))$, for any $s$. If $\phi$ is a $C^1$-supertangent to $u$ at $\tilde{\xi}(s)$ then

$$D\phi(\tilde{\xi}(s))\dot{\tilde{\xi}}(s)) = \frac{\mathrm{d}}{\mathrm{d}s} u(\tilde{\xi}(s))$$

accordingly,

$$p\dot{\tilde{\xi}}(s) = \eta(s)\tilde{\xi}(s) \quad \text{for any } s \text{ and } p \in \partial u(\tilde{\xi}(s))$$

Since $\partial u(\tilde{\xi}(s)) \subset Z_c(\tilde{\xi}(s))$, this implies that $\partial u(\tilde{\xi}(s)) = \{\eta(s)\}$. This actually gives the strict differentiability function $u$ at $\tilde{\xi}(s)$, and $Du(\tilde{\xi}(s)) = \eta(s)$ for any $s$.

The same argument works, with some adjustment, also when $a = c$ and $\xi \cap \mathcal{E} \neq \emptyset$. If, for instance, $y \notin \mathcal{E}, t_0 = \min\{t\colon \xi(t) \in \mathcal{E}\}$, then by reparametrizing $\xi$ in $[0, t_0]$, as indicated in [14], we get a curve $\tilde{\xi}$ defined in $[0, +\infty[$ which is a minimizer of the action functional in any compact interval contained in $[0, +\infty[$. Moreover, $u := S_c(y, \cdot)$ is strictly

differentiable in $]0,+\infty[$ and $(\tilde{\xi}, Du(\tilde{\xi}))$ is a solution of the Hamilton's equations.

We proceed to investigate the properties of the Hamiltonian flow on $\mathcal{A}$. We take a $y_0$ in $\mathcal{A}\backslash\mathcal{E}$, and consider a sequence $\xi_n$ of cycles passing through $y_0$ with $\ell_c(\xi_n)) \to 0, \ell(\xi_n) \geq 2\delta$, for some positive $\delta$. Such a sequence does exist in view of the characterization of $\mathcal{A}$ through cycles given in the previous section. Moreover, we assume that the $\xi_n$ are parametrized by the natural arc length in $[-T_n, T_n]$, for some $T_n \geq \delta$, and satisfy $\xi_n(0) = y_0$ for any $n$. There is then defined a uniform limit curve $\gamma$ in $[-\delta, \delta]$, up to a subsequence, thanks to the Ascoli theorem.

The idea is to construct a new sequence of cycles $\gamma_n$ by replacing the portion of the $\xi_n$ between $-\delta$ and $\delta$ by $\gamma$, and pasting this new piece with the remainder of $\xi_n$ through Euclidean segments at the end points. The $\gamma_n$ are still of infinitesimal intrinsic length $\ell_c$, which shows, in particular, that $\gamma$ is contained in $\mathcal{A}$. By exploiting that $S_c$ is a length distance, that the $\gamma_n$ are cycles, and the formula [7], with $a = c$, we get

$$\begin{aligned}\ell_c(\gamma_n) &\geq S_c(\gamma(-\delta), \gamma(\delta)) + S_c(\gamma(\delta), \gamma(-\delta)) \\ &\geq 0\end{aligned}$$

for any $n$, and we at last derive

$$\ell_c(\gamma) = S_c(\gamma(\delta), \gamma(-\delta)) = -\,S_c(\gamma(-\delta), \gamma(\delta))$$

Note that the second equality is actually redundant. By reparametrizing $\gamma$, as in [14], with $a = c$, in some open interval containing 0 as interior point and contained in $[-\delta, \delta]$, we get a curve contained in $\mathcal{A}\backslash\mathcal{E}$, denoted by $\xi$, defined on some open interval $I$ and satisfying

$$\begin{aligned}A(\xi|_{[s,t]}) + c(t-s)) &= \ell_c(\xi|_{[s,t]}) \\ &= -S_c(\xi(t), \xi(s)) \quad \text{for any } t > s \qquad [18]\end{aligned}$$

This, in particular, shows that $\xi$ is a minimizer of the action functional in any $[s,t] \subset I$. If we denote, as usual, by $\eta$ the curve conjugate to $\dot{\xi}$, we have, arguing as above, that $\eta(t)$ is the differential of the function $S_c(\xi(s), \cdot)$ at $\xi(t)$, but, since the differentials of all critical subsolutions coincide on $\mathcal{A}$, we finally get that $\eta(t) = G(\xi(t)$ for every $t \in I$. Therefore, $(\xi, G(\xi))$ is a solution of the Hamilton's equation in $I$ and is contained in $\tilde{\mathcal{A}}$. The same properties can be extended on the whole $\mathbb{R}$.

Taking into account that if $y \in \mathcal{E}$ then $(y, G(y))$ is a steady state of the Hamiltonian flow, we in the end see that $\tilde{\mathcal{A}}$ is foliated by integral curves of the Hamiltonian flow $(\xi, G(\xi))$, with $\xi$ enjoying the variational property [18]. This is indeed a characterization since if, conversely, a curve $\xi$ satisfies [18] then it must be contained in $\mathcal{A}$.

As an application, we finally show that there cannot be minimal geodesics, for the critical metric $S_c$, joining a point of $\mathcal{A}$, say $y$, to some $x \notin \mathcal{A}$, at least when $\mathcal{E} = \emptyset$. If such a geodesic, say $\xi$, exists, and is defined in $[0, T]$, for some $T > 0$, then $(\xi, Du(\xi))$ is a solution of the Hamilton's equations, up to a change of parameter, where $u := S(y, \cdot)$, satisfying the initial conditions $\xi(0) = y_0, \eta(0) = \lim_{t\to 0^+} Du(\xi(t))$.

The last relation tells us that $\eta(0) \in \partial u(y)$ and, since $u$ is differentiable at $y \in \mathcal{A}$ with $Du(y) = G(y)$, we conclude that $\eta(0) = G(y)$. Therefore, $(\xi, Du(\xi))$ is a part of the integral curve of the Hamiltonian flow starting at $(y, G(y))$ that we know, by the above reasoning, to be contained in $\tilde{\mathcal{A}}$, which is in contradiction with $\xi(T) = x \notin \mathcal{A}$.

*See also:* Control Problems in Mathematical Physics; Dynamical Systems in Mathematical Physics: An Illustratrion form Water Waves; KAM Theory and Celestial Mechanics; Minimax Principle in the Calculus of Variations; Optimal Transportation; Stability Theory and KAM.

## Further Reading


Arnold VI, Kozlov VV, and Neishtadt AI (1988) *Mathematical Aspects of Classical and Celestial Mechanics*, Encyclopedia of Mathematical Sciences. Dynamical Systems III. New York: Springer.

Contreras G and Iturriaga R (1999) *Global Minimizers of Autonomous Lagrangians*, 22nd Brazilian Mathematics Colloquium. IMPA: Rio de Janeiro.

Evans LC (2004) A survey of partial differential methods in weak KAM theory. *Communications on Pure and Applied Mathematics* 57: 445–480.

Evans LC and Gomes D (2001) Effective Hamiltonians and averaging for Hamilton dynamics I. *Arch. Rat. Mech. Analysis* 157: 1–33.

Evans LC and Gomes D (2002) Effective Hamiltonians and averaging for Hamilton dynamics II. *Arch. Rat. Mech. Analysis* 161: 271–305.

Fathi A *Weak Kam Theorem in Lagrangian Dynamics*. Cambridge: Cambridge University Press (to appear).

Fathi A and Siconolfi A (2004) Existence of $C^1$ critical subsolutions of the Hamilton–Jacobi equations. *Inventiones Mathematicae* 155(2): 363–388.

Fathi A and Siconolfi A (2005) PDE aspects of Aubry–Mather theory for quasiconvex Hamiltonians. *Calculus of Variations and Partial Differential Equations* 22(1): 185–228.

Forni G and Mather J (1994) In: Graffi S (ed.) Action Minimizing Orbits in Hamiltonian Systems (transition to *Chaos in Classical and Quantum Mechanics*), Lecture Notes in Mathematics, 1589. Springer.

Weinan E (1999) Aubry–Mather theory and periodic solutions of the forced Burgers equation. *Communications on Pure and Applied Mathematics* 52: 811–828.

**Hard Hexagon Model** *see* Eight Vertex and Hard Hexagon Models

# High $T_c$ Superconductor Theory

**S-C Zhang**, Stanford University, Stanford, CA, USA



## Introduction

The phenomenon of superconductivity is one of the most profound manifestations of quantum mechanics in the macroscopic world. The celebrated Bardeen–Cooper–Schrieffer (BCS) theory (Bardeen *et al.* 1957) of superconductivity (SC) provides a basic theoretical framework to understand this remarkable phenomenon in terms of the pairing of electrons with opposite spin and momenta to form a collective condensate state. This theory does not only quantitatively explain the experimental data of conventional superconductors, the basic concepts developed from this theory, including the concept of spontaneous broken symmetry, the Nambu–Goldstone modes and the Anderson–Higgs mechanism provide the essential building blocks for the unified theory of fundamental forces. The discovery of high-temperature superconductivity (HTSC) in the copper oxide material poses a profound challenge to theoretically understand the phenomenon of superconductivity in the extreme limit of strong correlations. While the basic idea of electron pairing in the BCS theory carries over to the HTSC, other aspects like the weak coupling mean field approximation and the phonon mediated pairing mechanism may not apply without modifications. Therefore, HTSC system provides an exciting opportunity to develop new theoretical frameworks and concepts for strongly correlated electronic systems.

To date, a number of different HTSC materials have been discovered. The most studied ones include the hole-doped $La_{2-x}Sr_xCuO_{4+\delta}$ (LSCO), $YBa_2Cu_3O_{6+\delta}$ (YBCO), $Bi_2Sr_2CaCu_2O_{8+\delta}$ (BSCO), $Tl_2Ba_2CuO_{6+\delta}$ (TBCO) materials and the electron-doped $Nd_{2-x}Ce_xCuO_4$ (NCCO) material. All these materials have a two-dimensional (2D) $CuO_2$ plane, and have an antiferromagnetic (AF) insulating phase at half-filling. The magnetic properties of this insulating phase is well approximated by the antiferromagnetic Heisenberg model with spin $S = 1/2$ and an AF exchange constant $J \sim 100\,\text{meV}$. The Neel temperature for the 3D AF ordering is approximately given by $T_N \sim 300 \sim 500\,\text{K}$. The HTSC material can be doped either by holes or by electrons. In the doping range of $5\% \lesssim x \lesssim 15\%$, there is an SC phase with a dom-like shape in the temperature versus doping plane. The maximal SC transition temperature $T_c$ is of the order of 100 K. The generic phase diagram of HTSC is shown in Figure 1.

One of the main questions concerning the HTSC phase diagram is the transition region between the AF and the SC phases. Partly because of the complicated material chemistry in this regime, there is no universal agreement among different experiments. Different experiments indicate several different possibilities, including phase separation with an inhomogeneous density distribution, uniform coexistence phase between AF and SC and periodically ordered spin and charge distributions in the form of stripes or checkerboards.

The phase diagram of the HTSC cuprates also contains a regime with anomalous behaviors conventionally called the pseudogap phase. This region of the phase diagram is indicated by the dashed lines in Figure 1. In conventional superconductors, a pairing gap opens up at $T_c$. In a large class of HTSC cuprates, however, an electronic gap starts to open up at a temperature much higher than $T_c$. Many experiments indicate that the pseudogap "phase" is

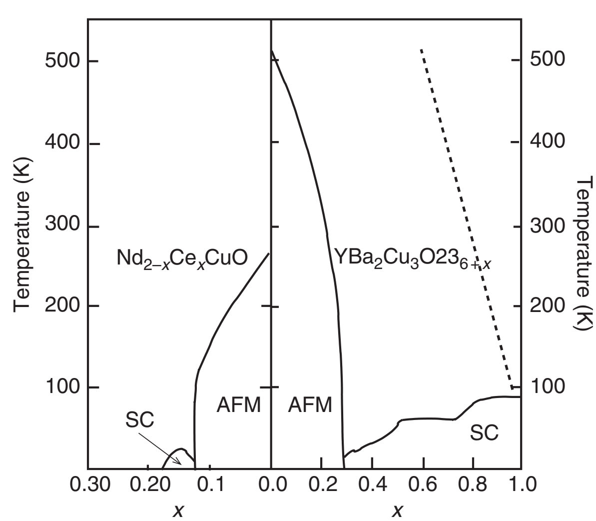



**Figure 1** Phase diagram of the of the NCCO and the YBCO superconductors.

not a true thermodynamical phase, but rather the precursor towards a crossover behavior.

The SC phase of the HTSC has a number of striking properties not shared by conventional superconductors. First of all, phase-sensitive experiments indicate that the SC phase for most of the cuprates has *d* wave like pairing symmetry. This is also supported by the photoemission experiments which show the existence of the nodal points in the quasiparticle gap. Neutron scattering experiments find a new type of collective mode, carrying spin 1, lattice momentum close to $(\pi, \pi)$, and a resolution-limited sharp resonance energy around 20–40 meV. Most remarkably, this resonance mode appears only below $T_c$ of the optimally doped cuprates. Another property uniquely different from the conventional superconductors is the vortex state. Most HTSCs are type II superconductors where the magnetic field can penetrate into the SC state in the form of a vortex lattice, where the SC order is destroyed at the center of the vortex core. In conventional superconductors, the vortex core is filled by the normal metallic electrons. However, a number of different experimental probes, including neutron scattering, muon spin resonance ($\mu$sR), and nuclear magnetic resonance (NMR), have shown that the vortex cores in the HTSC cuprates are antiferromagnetic, rather than normal metallic. This phenomenon has been observed in almost all HTSC materials, including LSCO, YBCO, TBCO, and NSCO, making it one of the most universal properties of the HTSC cuprates.

The HTSC materials also have highly unusual transport properties. While conventional metals have a $T^2$ dependence of resistivity, in accordance with the predictions of the Fermi liquid theory, the HTSC materials have a linear $T$ dependence of resistivity near optimal doping. This linear $T$ dependence extends over a wide temperature window, and seems to be universal among most of the cuprates. When the underdoped or sometimes optimally doped SC state is destroyed by applying a high magnetic field, the "normal state" is not a conventional conducting state, but exhibits insulator-like behavior, at least along the *c*-axis. This phenomenon may be related to the insulating AF vortices mentioned in the previous paragraph.

The discovery of HTSC has greatly stimulated the theoretical understanding of superconductivity in strongly correlated systems. There are a number of promising approaches, partially reviewed in Dagotto (1994), Imada *et al.* (1998), and Orenstein and Millis (2000), but an universally accepted theory has not yet emerged. This article focuses on a particular theory, which unifies the AF and the SC phases of the HTSC cuprates based on an approximate SO(5) symmetry (Zhang 1997). The SO(5) theory draws its inspirations from the successful application of symmetry concepts in theoretical physics. All fundamental laws of Nature are statements about symmetry. Conservation of energy, momentum, and charge are direct consequences of global symmetries. The form of fundamental interactions is dictated by local gauge symmetries. Symmetry unifies apparently different physical phenomena into a common framework. For example, electricity and magnetism were discovered independently, and viewed as completely different phenomena before the nineteenth century. Maxwell's theory, and the underlying relativistic symmetry between space and time, unify the electric field $\boldsymbol{E}$ and the magnetic field $\boldsymbol{B}$ into a common electromagnetic field tensor $F_{\mu\nu}$. This unification shows that electricity and magnetism share a common microscopic origin, and can be transformed into each other by going to different inertial frames. As discussed previously, the two robust and universal ordered phases of the HTSC are the AF and the SC phases. The central question of HTSC concerns the transition from one phase to the other as the doping level is varied. The SO(5) theory unifies the 3D AF order parameter $(N_x, N_y, N_z)$ and the 2D SC order parameter $(\mathrm{Re}\Delta, \mathrm{Im}\Delta)$ into a single, 5D order parameter called "superspin," in a way similar to the unification of electricity and magnetism in Maxwell's theory:

$$F_{\mu\nu} = \begin{pmatrix} 0 & & & \\ E_x & 0 & & \\ E_y & B_z & 0 & \\ E_z & -B_y & B_x & 0 \end{pmatrix} \Leftrightarrow n_a = \begin{pmatrix} \mathrm{Re}\,\Delta \\ N_x \\ N_y \\ N_z \\ \mathrm{Im}\,\Delta \end{pmatrix} \qquad [1]$$

This unification relies on the postulate that a common microscopic interaction is responsible for both AF and SC in the HTSC cuprates and related materials. A well-defined SO(5) transformation rotates one form of the order into another. Within this framework, the mysterious transition from the AF and the SC as a function of doping is explained in terms of a rotation in the 5D order parameters space. Symmetry principles are not only fundamental and beautiful, they are also practically useful in extracting information from a strongly interacting system, which can be tested quantitatively. The approximate SO(5) symmetry between the AF and the SC phases has many direct consequences, which can be, and some of them have been, tested both numerically and experimentally.

The commonly used microscopic model of the HTSC materials is the repulsive Hubbard model, which describes the electronic degrees of freedom in

the $CuO_2$ plane. Its low-energy limit, the $t-J$ model is defined by

$$H = -t \sum_{\langle x,x' \rangle} (c_\sigma^\dagger(x) c_\sigma(x') + \text{h.c.}) + J \sum_{\langle x,x' \rangle} S(x) \cdot S(x') \qquad [2]$$

where the term $t$ describes the hopping of an electron with spin $\sigma$ from a site $x$ to its nearest neighbor $x'$, with double occupancy removed, and the $J$ terms describe the nearest-neighbor exchange of its spin $S$. The main merit of these models does not lie in the microscopic accuracy and realism, but rather in the conceptual simplicity. However, despite their simplicity, these models are still very difficult to solve, and their phase diagrams cannot be compared directly with experiments. The idea of the SO(5) theory is to derive an effective quantum Hamiltonian on a coarse-grained lattice, which contains only the superspin degrees of freedom. The resulting SO(5) quantum nonlinear $\sigma$-model is much simpler to solve using the standard field theoretical techniques, and the resulting phase diagram can be compared directly with experiments.

## SO(4) Symmetry of the Hubbard Model

Before presenting the full SO(5) theory, let us first discuss a much simpler toy model, namely the negative $U$ Hubbard model, which has an SC ground state with $s$-wave pairing. However, it also has a charge-density-wave (CDW) ground state at half-filling. The competition between CDW and the SC states is similar to the competition between AF and SC states in the HTSC cuprates. In the negative $U$ Hubbard model, the CDW/SC competition can be accurately described by a hidden symmetry, namely the SO(4) symmetry of the Hubbard model.

The Hubbard model is defined by the Hamiltonian

$$H = -t \sum_{\langle x,x' \rangle} (c_\sigma^\dagger(x) c_\sigma(x') + \text{h.c.}) + U \sum_x \left(n_\uparrow(x) - \frac{1}{2}\right)\left(n_\downarrow(x) - \frac{1}{2}\right) - \mu \sum_x n_\sigma(x) \qquad [3]$$

where $c_\sigma(x)$ is the fermion operator and $n_\sigma(x) = c_\sigma^\dagger(x) c_\sigma(x)$ is the electron density operator at site $x$ with spin $\sigma$, $t$, $U$, and $\mu$ are the hopping, interaction, and the chemical potential parameters, respectively. The Hubbard model has a pseudospin SU(2) symmetry generated by the operators

$$\eta^- = \sum_x (-)^x c_\uparrow(x) c_\downarrow(x), \qquad \eta^+ = (\eta^-)^\dagger$$
$$\eta^z = \frac{1}{2} \sum_\sigma \left(n_\sigma(x) - \frac{1}{2}\right), \qquad [\eta^\alpha, \eta^\beta] = \mathrm{i}\epsilon_{\alpha\beta\gamma}\eta^\gamma \qquad [4]$$

where $\eta^\pm = \eta^x \pm \mathrm{i}\eta^y$ and $\alpha = x, y, z$. The model is defined on any bipartite lattice, and the lattice function $(-)^x$ takes the value 1 on even sublattice and $-1$ on odd sublattice. These operators commute with the Hubbard Hamiltonian at half-filling when $\mu = 0$, that is, $[H, \eta^\alpha] = 0$; therefore, they form the symmetry generators of the model (Yang and Zhang 1990). Combined with the standard SU(2) spin rotational symmetry, the Hubbard model enjoys an $SO(4) = SU(2) \otimes SU(2)/Z_2$ symmetry. This symmetry has important consequences in the phase diagram and the collective modes in the system. In particular, it implies that the SC and CDW orders are degenerate at half-filling. The SC and the CDW order parameters are defined by

$$\Delta^- = \sum_x c_\uparrow(x) c_\downarrow(x), \qquad \Delta^+ = (\Delta^-)^\dagger$$
$$\Delta^z = \frac{1}{2} \sum_{x\sigma} (-1)^x n_\sigma(x), \qquad [\eta^\alpha, \Delta^\beta] = \mathrm{i}\epsilon_{\alpha\beta\gamma}\Delta^\gamma \qquad [5]$$

where $\Delta^\pm = \Delta^x \pm \mathrm{i}\Delta^y$. The last equation of [5] shows that the $\eta$ operators perform the rotation between the SC and CDW order parameters. Thus, $\eta^\alpha$ is the pseudospin generator and $\Delta^\alpha$ is the pseudospin order parameter. Just like the total spin and the Neel order parameter in the AF Heisenberg model, they are canonically conjugate variables. Since $[H, \eta^\alpha] = 0$ at $\mu = 0$, this exact pseudospin symmetry implies the degeneracy of SC and CDW orders at half-filling.

The phase diagram of the $U < 0$ Hubbard model is identical to the phase diagram of the AF Heisenberg model in a uniform magnetic field. If the AF order parameter originally points along the $z$-direction, a magnetic field applied along the $z$-direction causes the AF order parameter to flop into the $xy$-plane. This transition is called the spin-flop transition, and is depicted in Figures 2a and 2c. The chemical potential $\mu$ in the negative $U$ Hubbard model plays a role similar to the magnetic field in the AF Heisenberg model. It transforms a CDW state at half-filling to an SC state away from half-filling, as depicted in Figures 2a and 2c.

In the low-energy sector, both the AF Heisenberg model in a magnetic field and the negative-$U$ Hubbard model with a chemical potential can be described by the SO(3) nonlinear $\sigma$-model, which is defined by the following Lagrangian density (in

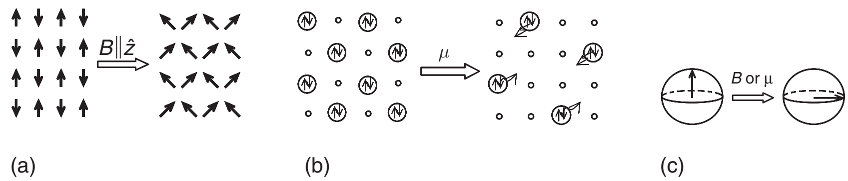



**Figure 2** The spin-flop transition. (a) The spin-flop transition of the AF Heisenberg model. When a uniform magnetic field is applied along the direction of the AF moments, there is no net gain of the Zeeman energy. Therefore, after a critical value of the magnetic field, the AF spin component flops into the *xy*-plane, while a uniform spin component aligns in the direction of applied magnetic field. (b) The Mott insulator to superfluid transition of the hardcore boson model or the $U < 0$ Hubbard model. At half-filling, one possible state is the CDW state of ordered boson pairs. Upon doping, the pairs become mobile and form the superfluid state. (c) Both transitions can be described by the spin or the pseudospin flop in the SO(3) nonlinear $\sigma$-model, induced either by the magnetic field or by the chemical potential.

imaginary time coordinates) for a unit vector field $n_\alpha$ with $n_\alpha^2 = 1$:

$$\mathcal{L} = \frac{\chi}{2}\omega_{\alpha\beta}^2 + \frac{\rho}{2}(\partial_i n_\alpha)^2 + V(n)$$
$$\omega_{\alpha\beta} = n_\alpha(\partial_t n_\beta - \mathrm{i}B_{\beta\gamma}n_\gamma) - (\alpha \to \beta) \qquad [6]$$

where the magnetic field, or equivalently the chemical potential, is given by $B_\alpha = (1/2)\epsilon_{\alpha\beta\gamma}B_{\beta\gamma}$. $\chi$ and $\rho$ are the susceptibility and stiffness parameters, and $V(n)$ is the anisotropy potential, which can be taken as $V(n) = -(g/2)n_z^2$. Exact SO(3) symmetry is obtained when $g = B_\alpha = 0$. $g > 0$ corresponds to easy axis anisotropy, while $g < 0$ corresponds to easy plane anisotropy. In the case of $g > 0$, there is a phase transition as a function of $B_z$ with $B_x = B_y = 0$. To see this, let us expand out the first term in [6]. The time-independent part contributes to an effective potential

$$V_{\text{eff}} = V(n) - \frac{\chi B^2}{2}(n_x^2 + n_y^2)$$

from which we see that there is a phase transition at $B_{c1} = \sqrt{g/\chi}$. For $B < B_{c1}$, the system is in the Ising phase, while for $B > B_{c1}$, the system is in the XY phase. Therefore, tuning $B$ for a fixed $g > 0$ leads to the spin-flop transition. In $D = 2$, both the XY and the Ising phase can have a finite-temperature phase transition into the disordered state. However, because of the Mermin–Wagner theorem, a finite-temperature phase transition is forbidden at the point $B = g = 0$, where the system has an enhanced SO(3) symmetry. This SO(3) symmetric point leads to a large regime below the mean field transition temperature where the fluctuation dominates. This large fluctuation regime can be identified as the pseudogap behavior.

The pseudospin SU(2) symmetry of the negative-$U$ Hubbard model has another important consequence. Away from half-filling, the $\eta$ operators no longer commute with the Hamiltonian, but they are eigenoperators of the Hamiltonian, in the sense that

$$[H, \eta^\pm] = \mp 2\mu\eta^\pm \qquad [7]$$

This means that the $\eta$ operators create well-defined collective modes with energy $2\mu$. Since they carry charge $\pm 2$, they usually do not couple to any physical probes. However, in an SC state, the SC order parameter mixes the $\eta$ operators with the CDW operator $\Delta^z$, via eqn [5]. From this reasoning, a pseudo-Goldstone mode was predicted to exist in the density response function at wave vector $(\pi, \pi)$ and energy $2\mu$, which appears only below the SC transition temperature $T_c$.

## Unification of Antiferromagnetism and Superconductivity through the SO(5) Theory

### Order Parameters and SO(5) Group Properties

The negative $U$ Hubbard model and the SO(3) nonlinear $\sigma$-models discussed in the previous section give a nice description of the quantum phase transition from the Mott insulating phase with CDW order to the SC phase. On the other hand, these simple models do not have enough complexity to describe the AF insulator at half-filling and the SC order away from half-filling. Therefore, a natural step is to generalize these models so that the Mott insulating phase with the scalar CDW order parameter is replaced by a Mott insulating phase with the vector AF order parameter. The pseudospin SO(3) symmetry group considered previously arises from the combination of one real scalar component of the CDW order parameter with one complex, or two real components of the SC order parameter. After replacing the scalar CDW order parameter by the three components of the AF order parameter, and combining with the two components of the SC order parameters, we are naturally led to consider a five-component order parameter vector, and the SO(5) symmetry group which transforms it.

It is simplest to define the concept of the SO(5) symmetry generator and order parameter on two sites with fermion operators $c_\sigma$ and $d_\sigma$, respectively,

where $\sigma = 1, 2$ is the usual spin index. The AF order parameter operator can be naturally defined in terms of the difference between the spins of the $c$ and $d$ fermions as follows:

$$N^\alpha = \tfrac{1}{2}(c^\dagger \tau^\alpha c - d^\dagger \tau^\alpha d), \qquad n_2 \equiv N_1$$
$$n_3 \equiv N_2, \qquad n_4 \equiv N_3 \qquad [8]$$

where $\tau^\alpha$ are the Pauli matrices. In view of the strong on-site repulsion in the cuprate problem, the SC order parameter should be naturally defined on a bond connecting the $c$ and $d$ fermions, explicitly given by

$$\Delta^\dagger = \frac{-\mathrm{i}}{2} c^\dagger \tau^y d^\dagger = \frac{1}{2}(-c_\uparrow^\dagger d_\downarrow^\dagger + c_\downarrow^\dagger d_\uparrow^\dagger),$$
$$n_1 \equiv \frac{(\Delta^\dagger + \Delta)}{2}, \qquad n_5 \equiv \frac{(\Delta^\dagger - \Delta)}{2i} \qquad [9]$$

We can group these five components together to form a single vector $n_a = (n_1, n_2, n_3, n_4, n_5)$ called the superspin, since it contains both superconducting and antiferromagnetic spin components. The individual components of the superspin are explicitly defined in the last parts of eqns [8] and [9].

The concept of the superspin is only useful if there is a natural symmetry group acting on it. In this case, since the order parameter is 5D, it is natural to consider the most general rotation in the 5D order parameter space spanned by $n_a$. In 3D, three Euler angles specify a general rotation. In higher dimensions, a rotation is specified by selecting a plane and the angle of rotation within this plane. Since there are $n(n-1)/2$ independent planes in $n$ dimensions, the group SO($n$) is generated by $n(n-1)/2$ elements, specified in general by antisymmetric matrices $L_{ab} = -L_{ba}$, with $a = 1, \ldots, n$. In particular, the SO(5) group has ten generators. The total spin and the total charge operator

$$S_\alpha = \tfrac{1}{2}(c^\dagger \tau_\alpha c + d^\dagger \tau_\alpha d)$$
$$Q = \tfrac{1}{2}(c^\dagger c + d^\dagger d - 2) \qquad [10]$$

perform the function of rotating the AF and SC order parameters within each subspace. In addition, there are six so-called $\pi$ operators, defined by

$$\pi_\alpha^\dagger = -\tfrac{1}{2} c^\dagger \tau_\alpha \tau_y d^\dagger, \quad \pi_\alpha = (\pi_\alpha^\dagger)^\dagger \qquad [11]$$

which perform the rotation from AF to SC and vice versa. These infinitesimal rotations are defined by the commutation relations

$$[\pi_\alpha^\dagger, N_\beta] = \mathrm{i}\delta_{\alpha\beta}\Delta^\dagger, \qquad [\pi_\alpha^\dagger, \Delta] = \mathrm{i}N_\alpha \qquad [12]$$

The ten operators, the total spin $S_\alpha$, the total charge $Q$, and the six $\pi$ operators form the ten generators of the SO(5) group.

The superspin order parameter $n_a$, the associated SO(5) generators $L_{ab}$, and their commutation relations can be expressed compactly and elegantly in terms of the SO(5) spinor and the five Dirac $\Gamma$ matrices. The four-component SO(5) spinor is defined by

$$\Psi_\mu = \begin{pmatrix} c_\sigma \\ d_\sigma^\dagger \end{pmatrix} \qquad [13]$$

They satisfy the usual anticommutation relations

$$\{\Psi_\mu^\dagger, \Psi_\nu\} = \delta_{\mu\nu}, \qquad \{\Psi_\mu, \Psi_\nu\} = \{\Psi_\mu^\dagger, \Psi_\nu^\dagger\} = 0 \qquad [14]$$

Using the $\Psi$ spinor and the five Dirac $\Gamma$ matrices, we can express $n_a$ and $L_{ab}$ as

$$n_a = \tfrac{1}{2}\Psi_\mu^\dagger \Gamma_{\mu\nu}^a \Psi_\nu, \qquad L_{ab} = -\tfrac{1}{2}\Psi_\mu^\dagger \Gamma_{\mu\nu}^{ab}\Psi_\nu \qquad [15]$$

The $L_{ab}$ operators satisfy the commutation relation

$$[L_{ab}, L_{cd}] = -\mathrm{i}(\delta_{ac}L_{bd} + \delta_{bd}L_{ac} - \delta_{ad}L_{bc} - \delta_{cc}L_{ad}) \qquad [16]$$

The $n_a$ and the $\Psi_\mu$ operators form the vector and the spinor representations of the SO(5) group, satisfying the following equations:

$$[L_{ab}, n_c] = -\mathrm{i}(\delta_{ac}n_b - \delta_{bc}n_a) \qquad [17]$$

and

$$\left[L_{ab}, \Psi_\mu\right] = -\tfrac{1}{2}\Gamma_{\mu\nu}^{ab}\Psi_\nu \qquad [18]$$

If we arrange the ten operators $S_\alpha, Q$, and $\pi_\alpha$ into $L_{ab}$'s by the following matrix form:

$$L_{ab} = \begin{pmatrix} 0 & & & & \\ \pi_x^\dagger + \pi_x & 0 & & & \\ \pi_y^\dagger + \pi_y & -S_z & 0 & & \\ \pi_z^\dagger + \pi_z & S_y & -S_x & 0 & \\ Q & \frac{1}{i}(\pi_x^\dagger - \pi_x) & \frac{1}{i}(\pi_y^\dagger - \pi_y) & \frac{1}{i}(\pi_z^\dagger - \pi_z) & 0 \end{pmatrix} \qquad [19]$$

and group $n_a$ as in eqns [8] and [9], we see that eqns [16] and [17] compactly reproduce all the commutation relations worked out previously. These equations show that $L_{ab}$ and $n_a$ are the symmetry generators and the order parameter vectors of the SO(5) theory.

Having introduced the concept of local symmetry generators and order-parameter-based sites in real space, we now proceed to discuss definitions of these operators in momentum space. The AF and dSC order parameters can be naturally expressed in terms of the microscopic fermion operators as

$$N^\alpha = \sum_p c_{p+\Pi}^\dagger \tau^\alpha c_p, \quad \Delta^\dagger = \frac{-\mathrm{i}}{2}\sum_p d(p) c_p^\dagger \tau^y c_{-p}^\dagger$$
$$d(p) \equiv \cos p_x - \cos p_y \qquad [20]$$

where $\Pi \equiv (\pi, \pi)$ and $d(p)$ is the form factor for $d$-wave pairing in 2D. They can be combined into the five-component superspin vector $n_a$ by using the

**Table 1** Quantum number of the AF and the dSC order parameters, and the $\pi$ operator, which rotates the AF and the dSC order parameters into each other

| | *Charge* | *Spin* | *Momentum* | *Internal angular momentum* |
|---|---|---|---|---|
| $\Delta, \Delta^\dagger$ or $n_1, n_5$ | $\pm 2$ | 0 | 0 | d-Wave |
| $N^\alpha$ or $n_{2,3,4}$ | 0 | 1 | $(\pi, \pi)$ | s-Wave |
| $\pi_\alpha, \pi_\alpha^\dagger$ | $\pm 2$ | 1 | $(\pi, \pi)$ | d-Wave |

same convention as before. The total spin and total charge operator are defined microscopically as

$$S_\alpha = \sum_p c_p^\dagger \tau^\alpha c_p, \qquad Q = \frac{1}{2}\sum_p (c_p^\dagger c_p - 1) \qquad [21]$$

and the $\pi$-operators can be defined as

$$\pi_\alpha^\dagger = \sum_p g(p) c_{p+\Pi}^\dagger \tau^\alpha \tau^y c_{-p}^\dagger \qquad [22]$$

The form factor $g(p)$ needs to be chosen appropriately to satisfy the SO(5) commutation relation [16], and this requirement determines $g(p) = \mathrm{sgn}(d(p))$.

The SO(5) symmetry generators perform the most general rotation among the five-order parameters. It is easy to see that the quantum number of the $\pi$ operators exactly patches up the difference in quantum numbers between the AF and the dSC order parameters, according to Table 1.

### The SO(5) quantum nonlinear $\sigma$-model

In the previous section we presented the concept of the local SO(5) order parameters and symmetry generators. These relationships are purely kinematic, and do not refer to any particular Hamiltonian. One can in fact construct microscopic models with exact SO(5) symmetry out of these operators. A large class of models, however, may not have SO(5) symmetry at the microscopic level, but their long-distance, low-energy properties may be described in terms of an effective SO(5) model. In the previous section, we have seen that many different microscopic models indeed all have the SO(3) nonlinear $\sigma$-model as their universal low-energy description. Similarly, we present the SO(5) quantum nonlinear $\sigma$-model as a general theory of AF and dSC in the HTSC.

From eqn [17] and the discussions in the previous subsection, we see that $L_{ab}$ and $n_a$ are conjugate degrees of freedom, very much similar to $[q, p] = i\hbar$ in quantum mechanics. This suggests that we can construct a Hamiltonian from these conjugate degrees of freedom. The Hamiltonian of the SO(5) quantum nonlinear $\sigma$-model takes the following form:

$$H = \frac{1}{2\chi}\sum_x L_{ab}^2(x) + \frac{\rho}{2}\sum_{<x,x'>} n_a(x) n_a(x') + \sum_x B_{ab}(x) L_{ab}(x) + \sum_x V(n(x)) \qquad [23]$$

where the $n_a$ vector field is subjected to the constraint

$$n_a^2 = 1 \qquad [24]$$

This Hamiltonian is quantized by the canonical commutation relations [16] and [17]. Here, the first term is the kinetic energy of the SO(5) rotors, where $\chi$ has the physical interpretation of the moment of inertia of the SO(5) rotors. The second term describes the coupling of the SO(5) rotors on different sites, through the generalized stiffness $\rho$. The third term introduces the coupling of external fields to the symmetry generators, while the $V(n)$ can include anisotropic terms to break the SO(5) symmetry to the SO(3) × U(1) symmetry. The SO(5) quantum nonlinear $\sigma$-model is a natural combination of the SO(3) nonlinear $\sigma$-model describing the AF Heisenberg model and the quantum XY model describing the SC to insulator transition. If we restrict to the values $a = 2, 3, 4$, then the first two terms describe the symmetric Heisenberg model, the third term describes easy plane or easy axis anisotropy of the Neel vector, while the last term represents the coupling to the uniform external magnetic field. On the other hand, for $a = 1, 5$, the first term describes Coulomb or capacitance energy, the second term is the Josephson coupling energy, while the last term describes coupling to external chemical potential.

The first two terms of the SO(5) model describe the competition between the quantum disorder and classical order. In the ordered state, the last two terms describe the competition between the AF and the SC order. Let us first consider the quantum competition. The first term prefers sharp eigenstates of the angular momentum. At an isolated site, $C \equiv \sum L_{ab}^2$ is the Casimir operator of the SO(5) group, in the sense that it commutes with all the SO(5) generators. The eigenvalues of this operator can be determined completely from group theory; they are 0, 4, 6, and 10, respectively, for the 1D SO(5) singlet, 5D SO(5) vector, 10D antisymmetric tensor, and 14D symmetric, traceless tensors. Therefore, we see that this term always prefers a quantum-disordered SO(5) singlet ground state, which is a total spin singlet. This ground state is separated from the first excited state, the fivefold

SO(5) vector state with an energy gap of $2/\chi$. This gap will be reduced, when the different SO(5) rotors are coupled to each other by the second term. This term represents the effect of stiffness, which prefers a fixed direction of the $n_a$ vector, rather than a fixed angular momentum. This competition is an appropriate generalization of the competition between the number sharp and phase sharp states in a superconductor and the competition between the classical Neel state and the bond or plaquette singlet state in the Heisenberg AF. The quantum phase transition occurs near $\chi\rho \simeq 1$.

In the classically ordered state, the last two anisotropy terms compete to select a ground state. To simplify the discussion, we can first consider the following simple form of the static anisotropy potential:

$$V(n) = -g(n_2^2 + n_3^2 + n_4^2) \qquad [25]$$

At the particle-hole symmetric point with vanishing chemical potential $B_{15} = \mu = 0$, the AF ground state is selected by $g > 0$, while the SC ground state is selected by $g < 0$ coupled with the constraint $n_a^2 = 1$. $g = 0$ is the quantum phase transition point separating the two ordered phases.

However, it is unlikely that the HTSC cuprates can be close to this quantum phase transition point. In fact, we expect the anisotropy term $g$ to be large and positive, so that the AF phase is strongly favored over the SC phase at half-filling. However, the chemical potential term has the opposite, competing effect favoring SC. To see this, we transform the Hamiltonian into the Lagrangian density (in imaginary time coordinates) in the continuum limit:

$$\mathcal{L} = \frac{\chi}{2}\omega_{ab}(x,t)^2 + \frac{\rho}{2}(\partial_k n^a(x,t))^2 + V(n(x,t)) \qquad [26]$$

where

$$\omega_{ab} = n_a(\partial_t n_b - \mathrm{i}B_{bc}n_c) - (a \to b) \qquad [27]$$

is the angular velocity. We see that the chemical potential enters the Lagrangian as a gauge coupling in the time direction. Expanding the time derivative term, we obtain an effective potential

$$V_{\text{eff}}(n) = V(n) - \frac{(2\mu)^2\chi}{2}(n_1^2 + n_5^2) \qquad [28]$$

from which we see that the $V$ term competes with the chemical potential term. For $\mu < \mu_c = \sqrt{g/\chi}$, the AF ground state is selected, while for $\mu > \mu_c$, the SC ground state is realized. At the transition point, even though each term strongly breaks SO(5) symmetry, the combined term gives an effective static potential which is SO(5) symmetric, as we can see from [28]. Even though the static potential is SO(5) symmetric, the full quantum dynamics is not. This can be most easily seen from the time-dependent term in the Lagrangian. When we expand out the square, the term quadratic in $\mu$ enters the effective static potential in eqn [28]. However, there is also a time-dependent term linear in $\mu$. This term breaks the particle-hole symmetry, and it dominates over the second-order time derivative term in the $n_1$ and $n_5$ variables. In the absence of an external magnetic field, only second-order time derivative terms of $n_{2,3,4}$ enter the Lagrangian. Therefore, while the chemical potential term compensates the anisotropy potential in eqn [28] to arrive at an SO(5) symmetric static potential, its time-dependent part breaks the full quantum SO(5) symmetry. This observation leads to the concept of the projected, or static SO(5) symmetry (Zhang *et al.* 1999). A model with projected or static SO(5) symmetry is described by a quantum effective Lagrangian of the form

$$\mathcal{L} = \frac{\chi}{2}\sum_{a=2,3,4}(\partial_t n_a)^2 - \chi\mu(n_1\partial_t n_5 - n_5\partial_t n_1) - V_{\text{eff}}(n) \qquad [29]$$

where the static potential $V_{\text{eff}}$ is SO(5) symmetric, but the time-dependent part contains a first-order time derivative term in $n_1$ and $n_5$.

The SO(5) quantum nonlinear $\sigma$-model is constructed from two canonically conjugate field operators $L_{ab}$ and $n_a$. In fact, there is a kinematic constraint among these field operators:

$$L_{ab}n_c + L_{bc}n_a + L_{ca}n_b = 0 \qquad [30]$$

This identity is valid for any triples $a$, $b$, and $c$, and can be easily proved by expressing $L_{ab} = n_a p_b - n_b p_a$, where $p_a$ is the conjugate momentum of $n_a$. Geometrically, this identity expresses the fact that $L_{ab}$ generates a rotation of the $n_a$ vector. The infinitesimal rotation vector lies on the tangent plane of the four sphere $S^4$, and is therefore orthogonal to the $n_a$ vector itself.

In a large class of materials, including the high-$T_c$ cuprates, the organic superconductors, and the heavy fermion compounds, the AF and SC phases occur in close proximity to each other. The SO(5) theory is developed based on the assumption that these two phases share a common microscopic origin and should be treated on an equal footing. The SO(5) theory gives a coherent description of the rich global phase diagram of the high-$T_c$ cuprates and its low-energy dynamics through a simple symmetry principle and a unified effective model based on a single quantum Hamiltonian. A number of theoretical predictions, including the intensity dependence of the neutron resonance

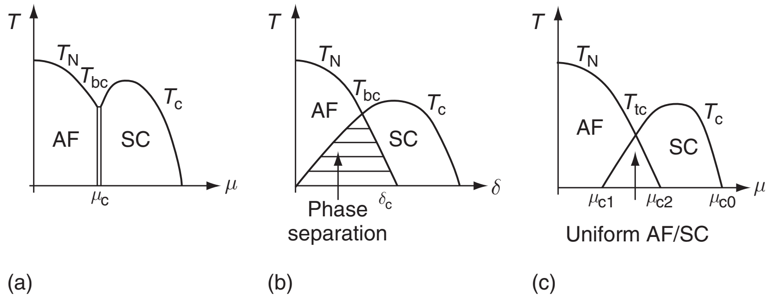



**Figure 3** The finite-temperature phase diagram of the SO(5) model in the temperature ($T$) versus chemical potential ($\mu$) plane. (a) and (b) are two different representations of the same phase diagram, corresponding to a direct first-order phase transition between AF and SC, as a function of the chemical potential and doping, respectively. (c) corresponds to two second-order phase transitions with a uniform AF/SC mix phase in between. The AF and the SC transition temperatures $T_N$ and $T_c$ merge into a bicritical $T_{bc}$ or a tetra-critical point $T_{tc}$. Both possibilities are allowed theoretically; it is up to experiments to determine which one is actually realized in the high-$T_c$ cuprates.

mode, the AF vortex state, and the mixed phase of AF and SC, have been verified experimentally (Figure 3). The theory also sheds light on the microscopic mechanism of superconductivity and quantitatively correlates the AF exchange energy with the condensation energy of superconductivity. However, the theory is still incomplete in many ways and lacks full quantitative predictive power. While the role of fermions is well understood within the exact SO(5) models, their roles in the effective SO(5) models are still not fully worked out. As a result, the theory has not made many predictions concerning the transport properties of these materials.

*See also:* Abelian Higgs Vortices; Effective Field Theories; Euclidean Field Theory; Ginzburg–Landau Equation; Hubbard Model; Quantum Phase Transitions; Quantum Spin Systems; Quantum Statistical Mechanics: Overview; Renormalization: General Theory; Renormalization: Statistical Mechanics and Condensed Matter; Superfluids; Symmetry Classes in Random Matrix Theory; Variational Techniques for Ginzburg–Landau Energies.

## Further Reading


Bardeen J, Cooper LN, and Schrieffer JR (1957) *Physical Review* 108(5): 1175.

Dagotto E (1994) *Reviews of Modern Physics* 66(3): 763.

Imada M, Fujimori A, and Tokura Y (1998) *Reviews of Modern Physics* 70(4): 1039.

Orenstein J and Millis AJ (2000) *Science* 288(5465): 468.

Yang CN and Zhang SC (1990) *Modern Physics Letters B* 4(6–7): 759.

Zhang SC (1997) *Science* 275(5303): 1089.

Zhang SC, Hu JP, Arrigoni E, Hanke W, and Auerbach A (1999) *Physical Review B* 60(18): 13070.


# Holomorphic Dynamics

**M Lyubich**, University of Toronto, Toronto, ON, Canada and Stony Brook University, NY, USA



## Introduction

### Subject

*Holomorphic dynamics* (in a narrow sense) is a theory of iterates of rational endomorphisms of the Riemann sphere $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. The goal is to understand the *phase portrait* of this dynamical system, that is, the structure of its trajectories, and the dependence of the phase portrait on *parameters* (coefficients of $f$).

Holomorphic dynamics in a broader sense would include the theory of analytic transformations, local and global, in dimension 1 and higher, as well as the theory of groups and pseudogroups of analytic transformations, which would cover theory of Kleinian groups and holomorphic foliations. However, we will mostly focus on holomorphic dynamics in the narrow sense.

### Brief History

Local dynamical theory of analytic maps was laid down in the late nineteenth and early twentieth century by Königs, Schröder, Böttcher, and Leau. Global theory of iterates of rational maps was founded by Fatou and Julia in comprehensive memoires of 1918–19. The theory had been

developed very little since then until early 1980s when it exploded with new methods, ideas, and computer images. Particularly influential were the works of D Sullivan who introduced ideas of *quasiconformal deformations* into the field, of A Douady and J Hubbard who gave a comprehensive combinatorial description of the Mandelbrot set, and W Thurston who linked holomorphic dynamics to three-dimensional hyperbolic geometry bringing to the field ideas of geometrization and rigidity. As a result, profound rigidity conjectures were formulated. Renormalization ideas introduced to the theory later on led to a significant progress towards these conjectures (*see* Universality and Renormalization).

Another source of ideas came from ergodic theory and the general theory of dynamical systems, particularly hyperbolic dynamics and thermodynamical formalism. They led to constructions of natural geometric measures on the Julia sets that helped to penetrate into their fractal nature.

### General Terminology and Notations

$\mathbb{N} = \{1, 2, \ldots\}$ is the set of natural numbers; $\mathbb{D}$ is the unit disk; $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}; \mathbb{T} = \partial\mathbb{D}$.

A *topological disk* is a simply connected domain in $\mathbb{C}$. A *topological annulus* is a doubly connected domain in $\mathbb{C}$ (i.e., a domain homeomorphic to a round annulus). A *Cantor set* is a totally disconnect compact subset of $\mathbb{R}^n$ without isolated points.

Given a map $f: X \to X, f^n$ will stand for its $n$-fold iterate. The semigroup of iterates form a *dynamical system with discrete time*. An *orbit* or *trajectory* of a point $z$ is $\mathrm{orb}_f(z) = (f^n z)_{n=0}^{\infty}$.

A subset $Y \subset \hat{\mathbb{C}}$ is called *invariant* if $f(Y) \subset Y$ and *completely invariant* if also $f^{-1}(Y) \subset Y$.

A point $\alpha \in \hat{\mathbb{C}}$ is called *periodic* if $f^p \alpha = \alpha$ for some natural $p$. The smallest such $p$ is called the *period* of $\alpha$. If $p = 1$, then $\alpha$ is called a *fixed point*. The orbit of a periodic point is also called a *cycle*.

Two maps $f: X \to X$ and $g: Y \to Y$ on topological spaces $X$ and $Y$ are called *topologically conjugate* if there exists a homeomorphism $h: X \to Y$ such that $h \circ f = g \circ h$. If $h$ has better regularity properties, for example, it is quasiconformal/conformal/affine, then $f$ and $g$ are called *quasiconformally/conformally/affinely conjugate*.

Let $f(z) = P(z)/Q(z)$ be a rational function viewed as a map $\hat{\mathbb{C}} \to \hat{\mathbb{C}}$. Its *topological degree* $\deg f = \# f^{-1}(z)$, $z \in \hat{\mathbb{C}}$, (where the preimages of $z$ are counted with multiplicity), is equal to the *algebraic degree* $\max(\deg P, \deg Q)$. The dynamics of $f$ is very simple in degree 1, so in what follows we assume that $\deg f \geq 2$.

Let $C_f = \{c : Df(c) = 0\}$ stand for the set of *critical points* of $f$, and $V_f = f(C_f)$ be the set of *critical values*. A rational function of degree $d$ has $2d - 2$ critical points counted with multiplicity. Moreover,

$$C_{f^n} = \bigcup_{k=0}^{n-1} f^{-k}(C_f), \qquad V_{f^n} = \bigcup_{k=1}^{n} f^k(C_f)$$

The latter formula explains why the behavior of the critical orbits crucially influences the global dynamics of $f$. The set $O_f = \cup V_{f^n}$ is called *postcritical*.

## Basic Dynamical Theory

### Local Theory

The local theory describes the dynamics of an analytic map $f: z \mapsto \lambda z + \sum_{n=2} a_n z^n$ near its fixed point 0. The derivative $\lambda = f'(0)$ is called the multiplier of 0. The fixed point is called *attracting*, *repelling*, or *neutral*, depending on whether $|\lambda| < 1$, $|\lambda| > 1$, or $|\lambda| = 1$. It is called *superattracting* if $\lambda = 0$.

In case when 0 is an attracting (but not superattracting) or repelling fixed point, the map is *linearizable*, that is, it is conformally conjugate to its linear part $z \mapsto \lambda z$; thus, there is a local conformal solution of the *Schröder equation* $\phi(fz) = \lambda\phi(z)$. This solution is also called the *linearizing coordinate* near 0.

In the superattracting case, the map is conformally conjugate to the map $z \mapsto z^d$, where $a_d z^d$ is the first nonvanishing term in the local expansion of $f$. Thus, in this case there is a local conformal solution of the *Böttcher equation* $\phi(fz) = \phi(z)^d$. It is also called the *Böttcher coordinate* near 0.

The situation in the neutral case (when $\lambda = e^{2\pi i\theta}$, $\theta \in \mathbb{R}/\mathbb{Z}$) depends in a delicate way on the arithmetic properties of the *rotation number* $\theta$. If $\theta = q/p$ is rational, the fixed point 0 is called *parabolic*. The local dynamics is then described in terms of the *Leau–Fatou flower* consisting of attracting petals alternating with repelling petals. In each petal, the map is conformally conjugate to the translation $z \mapsto z + 1$. The quotients of the petals by dynamics are conformally equivalent to the cylinder $\mathbb{C}/\langle z \mapsto z + 1 \rangle$. They are called (*attracting/repelling*) *Ecalle–Voronin cylinders*.

In the *irrational* case, when $\theta \in \mathbb{R} \backslash \mathbb{Q}$, the map can be either linearizable or not. Accordingly, 0 is called a *Siegel* or a *Cremer* fixed point. If the multiplier is *Diophantine* (i.e., there exist $C > 0$ and $\alpha > 2$ such that for all rational numbers $q/p$, we have: $|\theta - q/p| \geq Cp^{\alpha}$), then 0 is linearizable (Siegel 1942).

Notice that almost all numbers are Diophantine. A sharper arithmetic condition for linearizability in terms of the continuous fraction expansion for $\theta$ was given by Bruno (1965). In the quadratic case, $z \mapsto e^{2\pi i\theta} z + z^2$, this condition was proved to be sharp (Yoccoz 1988).

### Fatou and Julia Sets

From now on, $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ is a rational endomorphism of the Riemann sphere. The theory starts with the splitting of the sphere into two subsets now called Fatou and Julia sets based on the notion of a normal family in the sense of Montel. A family $(\phi_\alpha : S \to \hat{\mathbb{C}})$ of meromorphic functions on some Riemann surface $S$ is called *normal* if it is precompact in the open–closed topology. The *Fatou set* $F(f)$ is the maximal open subset of $\hat{\mathbb{C}}$ on which the family of iterates $(f^n)_{n=0}^{\infty}$ is normal. The *Julia set* $J(f)$ is the complement of the Fatou set. Both sets are completely invariant. The Julia set is always nonempty, and is either nowhere dense or coincides with the whole sphere. The trajectories on the Fatou set are *Lyapunov stable* (if $z$ is close to $z_0 \in F(f)$, then $\mathrm{orb}_f(z)$ is uniformly close to $\mathrm{orb}_f(z_0)$), while the dynamics on the Julia set is "chaotic."

If $f$ is a polynomial, then the Fatou and Julia sets can be defined in a more concrete way as follows. In this case, $\infty$ is a superattracting fixed point for $f$. Let us consider its *basin of attraction*,

$$D_f(\infty) = \{z : f^n z \to \infty \text{ as } z \to \infty\}$$

Its complement, $K(f)$, is called the *filled Julia set*. Then,

$$J(f) = \partial K(f) = \partial D_f(\infty)$$

### Periodic Points

Let $\alpha$ be a periodic point of $f$ of period $p$. As a fixed point of $f^p$, it is subject of the local theory. Thus, it (and its cycle) is classified as *attracting*, *repelling*, etc., according to the properties of the *multiplier* $\lambda = (f^p)'(\alpha)$ (that can be calculated in *any* local chart near $\alpha$).

The *basin of attraction* $D_f(\boldsymbol{\alpha})$ of an attracting cycle $\boldsymbol{\alpha} = (f^n \alpha)_{n=0}^{p-1}$ is the set $\{z : f^n z \to \boldsymbol{\alpha} \text{ as } n \to \infty\}$. The *immediate basin of attraction* $D_f^*(\boldsymbol{\alpha})$ is the union of components of $D_f(\boldsymbol{\alpha})$ containing the points of $\boldsymbol{\alpha}$.

**Theorem 1** (Fatou–Julia). *The immediate basin of any attracting cycle contains a critical point.* (*Note that a superattracting cycle actually contains some critical point.*)

It follows that *a rational function of degree d has at most* $2d-2$ *attracting cycles. A polynomial of degree d has at most* $d-1$ *attracting cycles in* $\mathbb{C}$.

Attracting cycles belong to the Fatou set, while repelling cycles lie on the Julia set. Parabolic and Cremer points lie on the Julia set, while Siegel points belong to the Fatou set. The *basin of attraction* of a parabolic cycle $\boldsymbol{\alpha}$ is defined as

$$D_f(\boldsymbol{\alpha}) = \{z : f^n z \to \boldsymbol{\alpha} \text{ as } n \to \infty\} \setminus \bigcup_{n=0}^{\infty} f^{-n}(\boldsymbol{\alpha})$$

It is the union of some components of the Fatou set. The union of the components of $D_f(\boldsymbol{\alpha})$ containing the petals of the Leau–Fatou flower is called the *immediate basin of attraction* $D_f^*(\boldsymbol{\alpha})$ of $\boldsymbol{\alpha}$. As in the attracting case, *the immediate basin* $D_f^*(\boldsymbol{\alpha})$ *of a parabolic cycle contains a critical point of f*.

Components of the Fatou set containing Siegel periodic points are called *Siegel disks*. If $D$ is a Siegel disk of period $p$, then $f^p|D$ is conformally conjugate to the irrational rotation $z \mapsto e^{2\pi i\theta} z$ of the unit disk.

**Theorem 2** (Shishikura 1987). *A rational function of degree d has at most* $2d-2$ *nonrepelling cycles.*

The proof of this result uses the methods of *quasiconformal surgery*.

### Examples

For $f : z \mapsto z^d, d \geq 2$, the Julia set $J(f)$ is the unit circle. Moreover, $D_f(\infty) = \hat{\mathbb{C}} \setminus \bar{\mathbb{D}}$, while $\mathbb{D}$ is the basin of attraction of the superattracting fixed point 0.

For maps $f_\varepsilon : z \mapsto z^2 + \varepsilon$ with sufficiently small $\varepsilon \neq 0$, the Julia set $J(f)$ is a *nowhere-differentiable* Jordan curve (see Figure 1). The domain bounded by this curve is the basin of attraction of an attracting fixed point $\alpha_\varepsilon$.

The filled Julia set of the map $f : z \mapsto z^2 - 1$ called the *basilica* is depicted in black in Figure 2. The

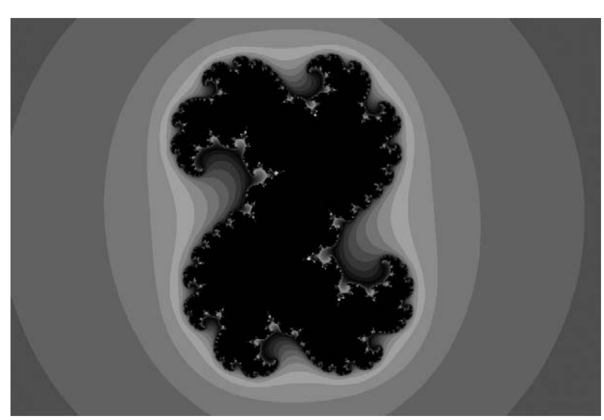


**Figure 1** Nowhere-differentiable Jordan curve.

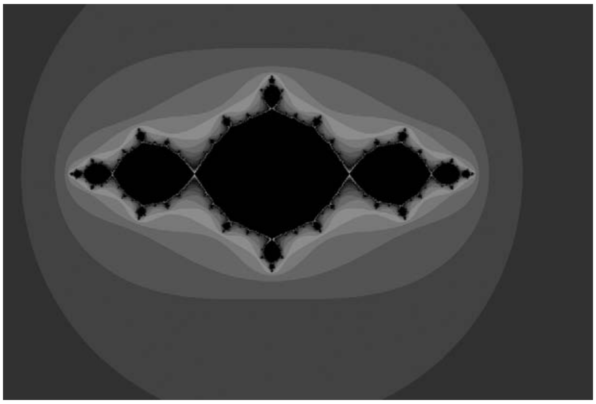


**Figure 2** Basilica.

interior of the basilica is the basin of the super-attracting cycle $\alpha = (0, 1)$ of period 2.

For the map $f: z \mapsto z^2 - 2$, the Julia set is the interval $[-2, 2]$. It is affinely conjugate to the Chebyshev quadratic polynomial $\mathrm{Ch}_2 : z \mapsto 2z^2 - 1$. More generally, for a Chebyshev polynomial $\mathrm{Ch}_d$ of any degree $d$, the Julia set is the interval. (By definition, the *Chebyshev polynomial* $\mathrm{Ch}_d$ is the solution of the functional equation $\cos dz = \mathrm{Ch}_d(\cos z)$.)

For quadratic maps $f_c : z \mapsto z^2 + c$ with $c < -2$, the Julia set is a *Cantor set* on $\mathbb{R}$. For maps $f_c$ with $c > 1/4$, the Julia set is a Cantor set that does not meet $\mathbb{R}$. For $c \in (-2, 1/4]$, the Julia set contains an invariant interval on $\mathbb{R}$, but is not contained in $\mathbb{R}$.

For $f: z \mapsto z^2 + i$, the Julia set is a "dendrite" (see Figure 3).

For $c \approx 0.12 + 0.74i$, the map $f: z \mapsto z^2 + c$ has an attracting cycle of period 3. Its Julia set in known as the *Douady rabbit* (Figure 4).

## No Wandering Domains Theorem and Dynamics on the Fatou Set

A component $D$ of the Fatou set is called *wandering* if $f^n(D) \cap f^m(D) = \emptyset$ for all natural $n > m$.

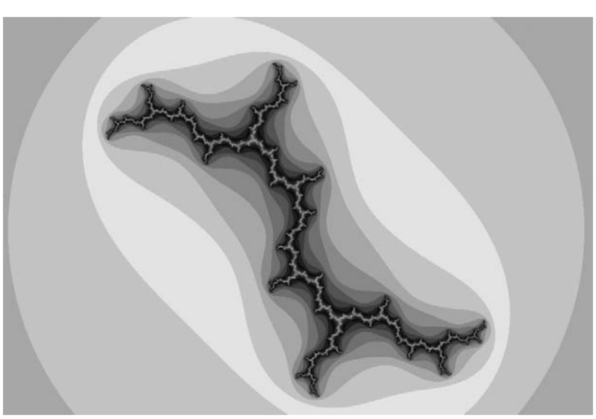


**Figure 3** Dendrite.

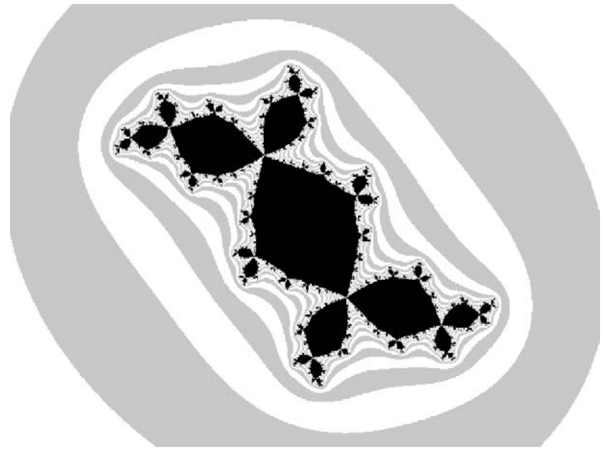


**Figure 4** Douady rabbit.

**Theorem 3** (Sullivan 1982). *Rational functions do not have wandering domains.*

This theorem is analogous to Ahlfors theorem in the theory of Kleinian groups. Its proof introduced to holomorphic dynamics the methods of *quasiconformal deformations* that has become the basic tool of the subject.

The "no wandering domains theorem" has completed the picture of dynamics on the Fatou set. Namely, for any $z \in F(f)$, one of the following three events may happen:

- $z$ belongs to the basin of attraction of some attracting cycle;
- $z$ belongs to the basin of attraction of some parabolic cycle; and
- for some $n$, $f^n z$ belongs to a rotation domain.

Here a *rotation domain* is either a Siegel disk, or a *Herman ring*, that is, a topological annulus $A$ such that $f^p(A) = A$ for some $p \in \mathbb{N}$ and $f^p \,|\, A$ is conformally equivalent to an irrational rotation $z \mapsto e^{2\pi i} z$ of a round annulus $\{z : 1 < |z| < R\}$. Note that Herman rings cannot occur for polynomial maps.

## More Properties of the Julia Set

There are two more useful characterizations of the Julia set:

- If $z$ is not an attracting periodic point and does not belong to a rotation domain, then *the set of accumulation points of the full preimages $f^{-n}z$ is equal to $J(f)$.*
- *The Julia set is the closure of the set of repelling periodic points.*

In the polynomial case, the Julia set $J(f)$ (and the filled Julia set $K(f)$) *is connected if and only if the critical points do not escape to* $\infty$ (in other words,

$C_f \subset K(f)$). In the quadratic case, the *Basic Dichotomy* holds: *the Julia set (and the filled Julia set) is either connected or a Cantor set*.

### Böttcher Coordinate

Let $f = z^d + a_1 z^{d-1} + \cdots + a_d$ be a monic polynomial of degree $d \geq 2$. Then $\infty$ is a superattracting fixed point, and hence there is a univalent function $B(z) = B_f(z)$ near $\infty$ satisfying the Böttcher equation $B(fz) = B(z)^d$ (the *Böttcher coordinate* near $\infty$). Moreover, $B(z) \sim z$ as $z \to \infty$ since $f$ is monic.

If $J(f)$ is connected, $B(z)$ can be analytically extended to the whole basin of $\infty$, and provides us with the Riemann map $\mathbb{C} \setminus K(f) \to \mathbb{C} \setminus \bar{\mathbb{D}}$. Otherwise, $B(z)$ extends to a conformal map from some invariant domain $\Omega_f$ whose boundary contains a critical point onto $\mathbb{C} \setminus \bar{\mathbb{D}}_R$, where $R = R_f > 1$.

The $B$-preimage of a straight ray $\{re^{2\pi i\theta}: 0 < r < \infty\}$ is called the *external ray* $\mathcal{R}_\theta$ of angle $\theta$. The $B$-preimage of a round circle $\{re^{2\pi i\theta}: 0 \leq \theta < 1\}$ is called the *equipotential* $\mathcal{E}_t$ *of level* $t = \log r$. External rays and equipotentials form two orthogonal $f$-invariant foliations. We let $\mathcal{R}_\theta(t) = \mathcal{R}_\theta \cap \mathcal{E}_t$.

### Combinatorial Equivalence

Assume now that $J(f)$ is connected. One says that an external ray $\mathcal{R}_\theta$ *lands* at some point $z \in J(f)$ if $\mathcal{R}_\theta(t) \to z$ as $t \to 0$. *Any external ray of rational angle* $\theta = q/p$ *with odd* $p$ *lands at some repelling or parabolic periodic point of period dividing* $p$ (Douady and Hubbard 1982). Vice versa, *any repelling or parabolic point is a landing point of at least one rational ray as above* (Douady 1990s).

Let us consider the following equivalence relation on the set of rational numbers with odd denominators: two such numbers $\theta$ and $\theta'$ are equivalent if the corresponding rays $\mathcal{R}_\theta$ and $\mathcal{R}_{\theta'}$ land at the *same* point $z \in J(f)$. Two polynomials $f$ and $\tilde{f}$ with connected Julia set are called *combinatorially equivalent* if the corresponding equivalence relations coincide. Notice that topologically equivalent polynomials are combinatorially equivalent.

## Parameter Phenomena

### Spaces of Rational Functions

Let $\mathrm{Rat}_d$ stand for the space of rational functions of degree $d$. As an open subset of the complex projective space $\mathbb{CP}^{2d+1}$, it is endowed with the natural topology and complex structure.

### Hyperbolic Maps and Fatou's Conjecture

Hyperbolic maps form an important and best-understood class of rational maps (compare with Hyperbolic Dynamical Systems). A rational map $f$ is called *hyperbolic* if one of the following equivalent conditions holds:

- All critical points of $f$ converge to attracting cycles;
- The map is expanding on the Julia set:

$$|Df^n(z)| \geq C\lambda^n, \quad z \in J(f)$$

where $C > 0$, $\lambda > 1$.

For instance, the maps $z \mapsto z^2 + \varepsilon$ for small $\varepsilon$, $z \mapsto z^2 - 1$, and $z \mapsto z^2 + c$ for $c \in \mathbb{R} \setminus [-2, 1/4]$ are hyperbolic. It is easy to see from the first definition that hyperbolicity is a *stable* property, that is, the set of hyperbolic maps is open in the space $\mathrm{Rat}_d$ of rational maps of degree $d$. One of the central open problems in holomorphic dynamics is to prove that this set is also dense. This problem is known as *Fatou's conjecture*.

### Postcritically Finite Maps and Thurston's Theory

A rational map is called *postcritically finite* if the orbits of all critical points are finite. In this case, any critical point $c$ is either a superattracting periodic point, or a repelling *preperiodic* point (i.e., $f^n c$ is a repelling periodic point for some $n$). *If all critical points of f are preperiodic, then* $J(f) = \hat{\mathbb{C}}$.

Important examples of postcritically finite maps with $J(f) = \hat{\mathbb{C}}$ come from the theory of elliptic functions. Namely, let $\mathcal{P}_\tau : \mathbb{C}/\Gamma_\tau \to \hat{\mathbb{C}}$ be the Weierstrass $\mathcal{P}$-function, where $\Gamma_\tau$ is the lattice in $\mathbb{C}$ generated by 1 and $\tau$, $\mathrm{Im}\,\tau > 0$. It satisfies the functional equation $\mathcal{P}_\tau(nz) = f_{\tau,n}(\mathcal{P}(z))$, where $f_{\tau,n}$ is a rational function. These functions called *Lattés examples* possess the desired properties. (For some special lattices, $n$ can be selected complex: the corresponding maps are also called Lattés.)

More generally, one can consider postcritically finite topological branched coverings $f : S^2 \to S^2$. Two such maps, $f$ and $g$, are called *Thurston combinatorially equivalent* if there exist homeomorphisms $h, h' : (S^2, O_f) \to (S^2, O_g)$ homotopic $\mathrm{rel}\, O_f$ (and hence coinciding on $O_f$) such that $h' \circ f = h \circ g$.

A combinatorial class is called *realizable* if it contains a rational function. Thurston (1982) gave a combinatorial criterion for a combinatorial class to be realizable. If it is realizable, then *the realization is unique*, except for Lattés examples (*Thurston's Rigidity Theorem*).

## Structural Stability and Holomorphic Motions

A map $f \in \mathrm{Rat}_d$ is called *J-stable* if for any maps $g \in \mathrm{Rat}_d$ sufficiently close to $f$, the maps $f \,|\, J(f)$ and $g \,|\, J(g)$ are topologically conjugate, and moreover, the conjugacy $h_g : J(f) \to J(g)$ is close to id. Thus, the Julia set $J(f)$ *moves continuously* over the set of $J$-stable maps. The following result proves a weak version of Fatou's conjecture:

**Theorem 4** (Lyubich and Mañé-Sad-Sullivan 1983). *The set of J-stable maps is open and dense in* $\mathrm{Rat}_d$. *Moreover, the set of unstable maps is the closure of maps that have a parabolic periodic point.*

A map $f \in \mathrm{Rat}_d$ is called *structurally stable* if for any maps $g \in \mathrm{Rat}_d$ sufficiently close to $f$, the maps $f$ and $g$ are topologically conjugate on the whole sphere, and moreover, the conjugacy $h_g : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ is close to id. *The set of structurally stable maps is also open and dense in* $\mathrm{Rat}_d$ (Mañé-Sad-Sullivan).

The proofs make use of the theory of holomorphic motions developed for this purpose but having much broader range of applications in dynamics and analysis. Let $X$ be a subset of $\hat{\mathbb{C}}$, and let $h_\lambda : X \to \hat{\mathbb{C}}$ be a family of injections depending on parameter $\lambda \in \Lambda$ in some complex manifold with a marked point $\lambda_*$. Assume that $h_{\lambda_*} = \mathrm{id}$ and that the functions $\lambda \mapsto h_\lambda(z)$ are holomorphic in $\lambda$ for any $z \in X$. Such a family of injections is called a *holomorphic motion*.

A holomorphic motion of any set $X$ over $\Lambda$ *extends to a holomorphic motion of the whole sphere* $\hat{\mathbb{C}}$ over some smaller manifold $\Lambda' \subset \Lambda$ (Bers–Royden, Sullivan–Thurston 1986). If $h_\lambda$ is a holomorphic motion of an open subset of the sphere, then *the maps* $h_\lambda$ *are quasiconformal* (Mañè-Sad-Sullivan). These statements are usually referred to as the *λ-lemma*.

If $\Lambda = \mathbb{D}$, then the holomorphic motion of a set $X \subset \hat{\mathbb{C}}$ extends to a holomorphic motion of $\hat{\mathbb{C}}$ over the *whole disk* $\mathbb{D}$ (Slodkowsky 1991).

## Fundamental Conjectures

The above rigidity and stability results led to the following profound conjectures:

**QC Rigidity Conjecture** *If two rational maps are topologically conjugate, then they are quasiconformally conjugate.*

Let us consider the real projective tangent bundle PT over $\hat{\mathbb{C}}$, with a natural action of the map $f$. *A measurable invariant line field on the Julia set* is an invariant measurable section $X \to \mathrm{PT}$ over an invariant set $X \subset J(f)$ of positive Lebesgue measure. In other words, it is a family of tangent lines $L_z \subset T_z$, $z \in X$, such that $Df(L_z) = L_{fz}$. Note that such a field can exist only if $J(f)$ has positive Lebesgue measure.

**No Invariant Line Fields Conjecture** *Let us consider two rational maps, $f$ and $\tilde{f}$, that are not Lattés examples. If they are quasiconformally conjugate and the conjugacy is conformal on the Fatou set, then they are conjugate by a Möbius transformation. Equivalently, if $f$ is not Lattés, then there are no measurable invariant line fields on $J(f)$.*

This conjecture would imply Fatou's conjecture.

## Mandelbrot Set

Let us consider the quadratic family $f_c : z \mapsto z^2 + c$. (Note that any quadratic polynomial is affinely conjugate to a unique map $f_c$.) The *Mandelbrot set* classifies parameters $c$ according to the Basic Dichotomy of the subsection "More properties of the Julia set":

$$M = \{c : J(f_c) \text{ is connected}\} = \{c : f_c^n(0) \mapsto \infty\}$$

Note that $\phi_n(c) \equiv f_c^n(0)$ is a polynomial in $c$ of degree $2^{n-1}$, and these polynomials satisfy a recursive relation $\phi_{n+1} = \phi_n^2 + c$. Moreover, $M = \{c : |\phi_n(c)| \le 2,\ n \in \mathbb{Z}_+\}$, which gives an easy way to make a computer image of $M$ (see Figure 5).

A distinguished curve seen at the picture of $M$ is the *main cardioid* $\mathcal{C} = \{c = \mathrm{e}^{2\pi i\theta} - \mathrm{e}^{4\pi i\theta}/4\}$, $\theta \in \mathbb{R}/\mathbb{Z}$. For such a $c = c(\theta) \in \mathcal{C}$, the map $f_c$ has a neutral fixed point $\alpha_c$ with rotation number $\theta$. For $c$ inside the domain $H_0$ bounded by $\mathcal{C}, f_c$ has an attracting fixed point $\alpha_c$, and the Julia set $J(f_c)$ is a Jordan curve (see Figure 1).

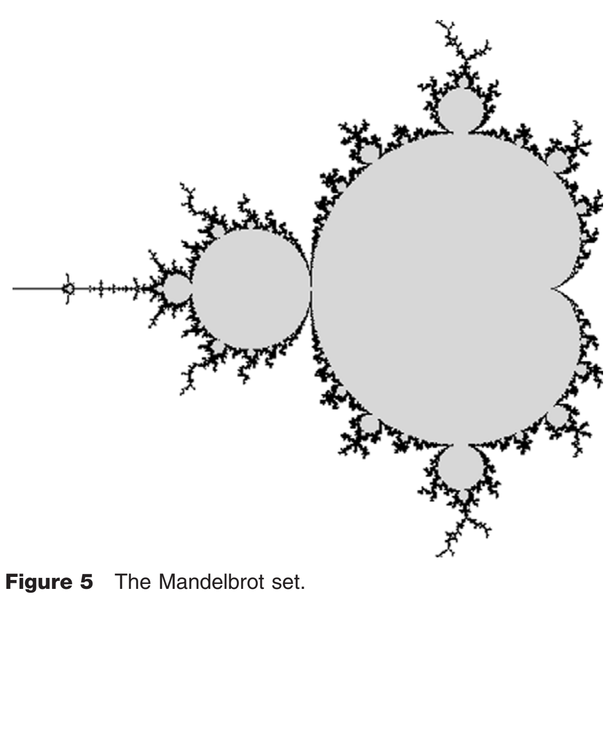


**Figure 5** The Mandelbrot set.

At the cusp $c = 1/4 = c(0)$ of the main cardioid, the map $f_c$ has a parabolic point with multiplier 1. This point is also called the *root* of $\mathcal{C}$. Other parabolic points $c = c(q/p)$ on $\mathcal{C}$ are *bifurcation points*: if one crosses $\mathcal{C}$ transversally at $c$, then the fixed point $\alpha_c$ "gives birth" to an attracting cycle of period $p$. This cycle preserves its "attractiveness" within some component $H_{q/p}$ of int $M$ attached to $\mathcal{C}$.

On the boundary of $H_{q/p}$, the above attracting cycle becomes neutral, and similar bifurcations happen as one crosses this boundary transversally, etc. In this way we obtain *cascades* of bifurcations and associated necklaces of components of int $M$. The most famous one is the cascade of *doubling bifurcations* that occur along the real slice of $M$.

Components of int $M$ that occur in these bifurcation cascades give examples of hyperbolic components of int $M$. More generally, a component $H$ of int $M$ is called *hyperbolic of period p* if the maps $f_c, c \in H$, have an attracting cycle of period $p$. Many other hyperbolic components become visible if one begins to zoom-in into the Mandelbrot set. Some of them are *satellite*, that is, they are born as above by bifurcation from other hyperbolic components. Others are *primitive*. They can be easily distinguished geometrically: primitive components have a cusp at their root, while satellite components are bounded by smooth curves.

Given a hyperbolic component $H$, let us consider the multiplier $\lambda(c), c \in H$, of the corresponding attracting cycle, as a function of $c \in H$. *The function $\lambda$ univalently maps H onto the unit disk $\mathbb{D}$* (Douady and Hubbard 1982). Thus, there is a single parameter $c_0 \in H$ for which $\lambda(c_0) = 0$, so that $f_{c_0}$ has a *superattracting cycle*. This parameter is called the *center* of $H$.

Nonhyperbolic components of int $M$ are called *queer*. *Conjecturally, there are no queer components*. This conjecture is equivalent to Fatou's conjecture for the quadratic family.

*The boundary of M coincides with the set of J-unstable quadratic maps* (see the subsection "Structural stability and holomorphic motions").

### Connectivity and Local Connectivity

**Theorem 5** (Douady and Hubbard 1982). *The Mandelbrot set is connected.*

The proof provides an explicit uniformization $R_M : \mathbb{C} \setminus M \to \mathbb{C} \setminus \bar{\mathbb{D}}$. Namely, let $B_c : \Omega_c \to \mathbb{C} \setminus \mathbb{D}_{R_c}$, $c \in \mathbb{C} \setminus M$, be the Böttcher coordinate near $\infty$. Then $R_M(c) = B_c(c)$. This remarkable formula explains the *phase-parameter similarity* between the Mandelbrot set near a *parameter* $c \in M$ and the corresponding Julia set $J(f_c)$ near the *critical value c*.

The following is the most prominent open problem in holomorphic dynamics:

**MLC Conjecture** *The Mandelbrot set is locally connected.*

If this is the case, then the inverse map $R_M^{-1}$ extends to the unit circle $\mathbb{T}$, and the Mandelbrot set can be represented as the quotient of $\mathbb{T}$ modulo certain equivalence relation that can be explicitly described. Thus, we would have *an explicit topological model for the Mandelbrot set* (Douady and Hubbard, Thurston).

The MLC conjecture is equivalent to the following conjecture:

**Combinatorial Rigidity Conjecture** *If two quadratic maps $f_c$ and $f_{c'}$ with all periodic points repelling are combinatorially equivalent, then $c = c'$.*

In turn, this conjecture would imply, in the quadratic case, the above fundamental conjectures. For a progress towards the MLC conjecture (*see* Universality in Mathematical Physics).

### Parabolic Implosion

Parabolic maps $f_{c_0} : z \mapsto z^2 + c_0$ are unstable in a dramatic way. In particular, the Julia set $J(f_c)$ *does not depend continuously on c* near $c_0$. Instead, $J(f_c)$ tends to fill in a good part of int $J(f_{c_0})$. This phenomenon called *parabolic implosion* has been explored by Douady, Lavaurs, Shishikura, and many others.

## Geometric Aspects

### Area

One of the basic problems in holomorphic dynamics is *whether a Julia set that does not coincide with $\hat{\mathbb{C}}$ can have positive area*. It would give an example of "observable chaos" that occurs on a topologically small set. It is also related to the No Invariant Line Fields Conjecture.

Maps with strong hyperbolic properties have zero area Julia set. A rational map $f$ is called *Collet–Eckmann* if there exist constants $C > 0$ and $\lambda > 1$ such that:

$$|Df^n(fc)| \geq C\lambda^n, \quad n \in \mathbb{N}$$

for all critical points $c$. *If f is a Collet–Eckmann map with $J(f) \neq \hat{\mathbb{C}}$, then* area $J(f) = 0$ (Przytycki and Rohde 1998) (*see* Universality and Renormalization for more examples). On the other hand, A Douady has set up a compelling program of constructing a Cremer quadratic polynomial $f : z \mapsto e^{2\pi i\theta} z + z^2$ whose Julia set would have *positive area*. Buff and

Cheritat have recently announced that they have completed the program, thus constructing the first example of a Julia set of positive area. (It makes use of a renormalization theorem for parabolic implosion recently announced by Shishikura.)

In the parameter plane, it would be interesting to know *whether the boundary of the Mandelbrot set has zero area*.

#### Hausdorff Dimension

Hausdorff dimension (HD) gives us a further refinement of fractal sets of zero area. Any Julia set has positive HD. *If $f$ is a polynomial with connected Julia set, then* $\mathrm{HD}(J(f)) > 1$ *unless $f$ is affinely conjugate to $z \mapsto z^d$ or a Chebyshev polynomial* (Zdunik 1990). If $f$ is a Collet–Eckmann map with $J(f) \neq \hat{\mathbb{C}}$, then $\mathrm{HD}\, J(f) < 2$ (Przytycki–Rohde 1998). On the other hand, in the quadratic case $f_c : z \mapsto z^2 + c, \mathrm{HD}(J(f_c)) = 2$ *for a generic parameter* $c \in \partial M$. The corresponding parameter result is that $\mathrm{HD}(\partial M) = 2$ (Shishikura 1998). It is based on the parabolic implosion phenomenon.

#### Conformal Measure

Let $\delta \geq 0$. A Borel measure $\mu$ on $\hat{\mathbb{C}}$ is called *$\delta$-conformal* if

$$\mu(f\,X) = \int_X |Df|^\delta \, \mathrm{d}\mu$$

for any measurable set $X$ such that $f \,|\, X$ is injective.

**Theorem 6** (Sullivan 1983). *Any rational map $f$ has a $\delta$-conformal measure with $\delta \in (0, 2]$ supported on $J(f)$.*

This is a dynamical measure that captures well geometric properties of $J(f)$. For instance, for Collet–Eckmann maps, $\delta = \mathrm{HD}(J(f))$, and $\mu$ is equivalent to the Hausdorff measure on $J(f)$ in dimension $\delta$.

The *hyperbolic dimension*, $\mathrm{HD}_{\mathrm{hyp}}$ of $J(f)$ is the supremum of $\mathrm{HD}(X)$ over all compact invariant hyperbolic subsets of $J(f)$. Denker and Urbanski (1991) proved that $\mathrm{HD}_{\mathrm{hyp}}(J(f))$ *is equal to the smallest exponent $\delta$ of all $\delta$-conformal measures supported on $J(f)$* (*see* Universality and Renormalization).

#### Measure of Maximal Entropy

An $f$-invariant measure $\mu$ is called *balanced* if $\mu(f\,X) = d\mu(X)$ for any measurable set $X$ such that $f \,|\, X$ is injective (where $d = \deg f$).

**Theorem 7** (Brolin 1965, Lyubich 1982). *Any rational map $f$ has a unique balanced measure $\mu$. Moreover, preimages of any point $z$ except at most two are equidistributed with respect to $\mu$ (meaning that the probability measures $\mu_{n,z}$ that assign mass 1 to every $f^n$-preimage of $z$ converge weakly to $\mu$ as $n \to \infty$).*

For polynomials, the balanced measure coincides with the *harmonic measure* on $J(f)$ (Brolin). (The latter is the charge distribution on the conductor $J(f)$ generated by the unit charge placed at $\infty$.) In general, the balanced measure is the *unique measure of maximal entropy* of $f$, and moreover, *periodic points are equidistributed with respect to $\mu$* (Lyubich).

Measure of maximal entropy is supported on a relatively small measurable set: *its HD is strictly less than* $\mathrm{HD}(J(f))$, unless $f$ is conformally equivalent to $z \mapsto z^d$, a Chebyshev polynomial, or a Lattés example (Zdunik 1990). In the polynomial case, it is supported on a set of HD at most 1 (Manning 1984).

In complex analysis, there has been an extensive study of fractal properties of harmonic measures, providing insights at the balanced measure $\mu$ and the other way around (Carleson, Makarov, Jones, Binder, Smirnov, ...)

*See also:* Fractal Dimensions in Dynamics; Geometric Analysis and General Relativity; Geometric Flows and the Penrose Inequality; Geometric Phases; Polygonal Billiards; Renormalization: General Theory; Renormalization: Statistical Mechanics and Condensed Matter; Universality and Renormalization.

## Further Reading

Space limitations prohibit the inclusion of references to all the results quoted in this article. Where an author is quoted in the text, the reader can find the corresponding reference in one of the books listed in this section and in the MatSciNet.


Buff X and Hubbard J *Dynamics in One Complex Variable*. Ithaca, NY: Matrix Editions (to appear).

Carleson L and Gamelin W (1993) *Complex Dynamics*. Berlin: Springer.

Eremenko A and Lyubich M (1990) The dynamics of analytical transformations. *Leningrad Mathematical Journal* 1: 563–633.

Milnor J (1999) *Dynamics in One Complex Variable: Introductory Lectures*. Braunschweig: Vieweg and Sohn.

Pilgrim KM (2003) *Combinatorics of Complex Dynamical Systems*, Lecture Notes in Mathematics no 1827. Berlin–Heidelberg–New York: Springer-Verlag.

Przytycki F and Urbanski M, *Fractals in the Plane – the Ergodic Theory Methods*. (www.math.unt.edu/uranski/book1.html) Cambridge: Cambridge University Press (to appear).

Tan Lei (ed.) (2000) *The Mandelbrot Set, Theme and Variations*, London Math. Soc. Lecture Note Ser., vol. 274. Cambridge: Cambridge University Press.

# Holonomic Quantum Fields

**J Palmer**, University of Arizona, Tucson, AZ, USA



## Introduction

The term, holonomic field, was coined by Sato, Miwa, and Jimbo (SMJ) in 1978 and the subject was investigated by them in a series of five long papers and many shorter notes in the period from 1978–81 (Sato *et al.* 1979a, 1979b, 1979c, 1980, Tracy and Widom 1994). The term refers to a special class of two-dimensional interacting quantum field theories whose $n$ point correlations can be expressed in terms of the solution to a holonomic system of differential equations. A holonomic system is an overdetermined system of differential equations with only a finite-dimensional family of solutions. There is a sense in which these interacting systems with infinitely many degrees of freedom have a finite-dimensional substrate (at the level of $n$ point functions for fixed $n$). After developing their theory, SMJ realized that such quantum fields made an earlier appearance in work of Thirring and Federbush. The models considered by Thirring and Federbush are self-interacting fermionic systems whose nonlinear classical field equations have solutions that are an explicit nonlinear transformation of solutions to the free field equations. This inspired the idea of trying to study these models by "quantizing" the nonlinear transformation. Expressions were obtained for the correlations and S-matrix but the connection with deformation theory was not understood until the SMJ work.

In what follows we will sketch the SMJ theory and discuss some of its offshoots. There is one circumstance that it might help the reader to be aware of even though it will be mostly glossed over. Quantum fields in one space and one time dimension have correlations which transform under the symmetries of spacetime with metric signature (1,1). Since the work of Osterwalder and Schrader, it is customary to pass back and forth between this Minkowski regime and the Schwinger functions obtained by analytically continuing the $n$ point functions to pure imaginary values for the time variable where they possess the rotational symmetries associated with a positive-definite metric. The Ising model, which we take up next, is naturally considered in the Euclidean domain where the correlations have an interpretation in statistical mechanics as the expected value of a product of random variables. Ultimately, the SMJ deformation analysis is done in the Euclidean domain.

## The Two-Dimensional Ising Model

The SMJ theory was inspired by, and provides an attractive setting for, an earlier result of Wu, McCoy, Tracy, and Baruch (WMTB), concerning the spin–spin scaling functions of the two-dimensional Ising model (Wu *et al.* 1976). Since the Ising model is the example with the most direct significance for physics, we will take some time to explain the WMTB result and to sketch the way in which it fits into the SMJ theory.

The Ising model is a statistical model of magnetism on a lattice that incorporates ferromagnetic interactions of nearest-neighbor spins. In the 1920s, Ising solved the model for the one-dimensional lattice and showed that there was no phase transition in the infinite volume limit. Interest in the two-dimensional model intensified dramatically following Onsager's calculation of the specific heat in the infinite volume limit (see Palmer and Tracy (1981) and references within). His formula for the specific heat was the first instance of a thermodynamic quantity in a nearest-neighbor model which exhibits the sort of discontinuity in temperature dependence expected at a phase transition. For many years, the Ising model served as a testbed for the now accepted notion that the infinite volume limit of Gibbsian statistical mechanics provides a suitable setting for the study of phase transitions.

A configuration for the Ising model on a finite subset, $\Lambda$, of the integer lattice, $\mathbf{Z}^2$, is a map $\mathcal{C}: \Lambda \to \{+1, -1\}$, which assigns to each site on the lattice either an up spin $(+1)$ or a down spin $(-1)$. The energy function of the Ising model, $E_\Lambda(\sigma)$, is defined by

$$E_\Lambda(\sigma) = -J \sum_{\langle i,j \rangle \in \Lambda} \sigma(i)\sigma(j)$$

for $J > 0$ and a spin configuration $\sigma$ is a sum over pairs of nearest-neighbor sites $i, j$ in $\Lambda$ (boundary terms require special consideration). This energy function tends to favor spin configurations, $\sigma$, in which the nearest-neighbor spins are aligned in the sense that the Boltzmann weight, $e^{-E_\Lambda(\sigma)/kT}$, is larger for such configurations. In the Gibbs ensemble, which is expected to describe systems in equilibrium at temperature $T$, the configuration $\sigma$ occurs with a probability proportional to the Boltzmann weight. The factor $k$ which appears is a conversion factor between thermal and kinetic energy called the Boltzmann constant. It is clear from the formula for the Boltzmann weight that small temperatures (near 0) tend to accentuate the difference in

statistical weights assigned to configurations with different energies, and large temperatures tend to wash out the difference in statistical weights associated to configurations with different energies.

Remarkably, there is a sharp critical temperature $0 < T_c < \infty$ so that for $T < T_c$ the propensity for order built into the energy triumphs in the infinite volume limit $\Lambda \uparrow \boldsymbol{Z}^2$, and for $T > T_c$ the randomness or disorder associated with high temperatures governs the infinite volume behavior. More specifically, if $T < T_c$ and the infinite volume limit is taken with plus spins assigned to the boundary of $\Lambda$, the system exhibits a residual magnetism (there is a positive expected value, $\langle\sigma\rangle$, for the spin per site). This infinite volume plus state is the quintessential example of symmetry breaking – the spin flip symmetry possessed by the bulk energy is broken below $T_c$ in the thermodynamic limit. For $T > T_c$, the spin per site is 0 no matter what boundary conditions are imposed on the infinite volume limit.

Pure equilibrium states both above and below $T_c$ exhibit clustering in the thermodynamic limit (uniqueness for the ground state in field theory). This is the tendency of spin variables $\sigma(a)$ and $\sigma(b)$ at sites $a, b \in \boldsymbol{Z}^2$ to become statistically independent as the distance $|a-b|$ tends to $\infty$. In such a pure state the two-point function, which is the expected value of the product of spin variables, $\langle\sigma(a)\sigma(b)\rangle$, will tend to the square $\langle\sigma\rangle^2$ both below $(\langle\sigma\rangle \neq 0)$ and above $(\langle\sigma\rangle = 0)$ the critical temperature $T_c$ as $|a-b| \to \infty$. To leading order, this clustering takes place at an exponential rate, $e^{-|a-b|/\xi(T)}$, for a function $\xi(T)$ called the correlation length. The correlation length $\xi(T) \to \infty$ as $T \to T_c$. The scaling limit (from below $T_c$) of the spin–spin correlation is the leading-order correction to the clustering behavior of the correlations when these correlations are examined at the scale of the correlation length. It is the limit

$$\langle\sigma(a)\sigma(b)\rangle = \lim_{T\uparrow T_c} \frac{\langle\sigma(\xi(T)a)\sigma(\xi(T)b)\rangle_T}{\langle\sigma\rangle_T^2}$$

where the correlations on the right-hand side are thermodynamic correlations on the lattice at temperature $T$. Since $\langle\sigma\rangle_T$ tends to 0 as $T \to T_c$, the normalization by $\langle\sigma\rangle_T^{-1}$ on the right produces an "infinite wave function renormalization" in the limit.

Equivalently, one may think of this continuum limit being achieved by letting the lattice spacing shrink to 0 as $T$ approaches $T_c$ so that the correlation length stays fixed on the new scale. The scaling limit from above $T_c$ turns out to be different from the scaling limit from below $T_c$ and since $\langle\sigma\rangle_T = 0$ for $T > T_c$, it is defined by a different wave function renormalization. The resulting asymptotics are expected to capture quite a lot about what is interesting in the behavior of the correlations near the phase transition. In the late 1970s, the scaling behavior in this model was also a prototype for the emerging connection between renormalization group ideas in quantum field theory and statistical mechanics.

Wu *et al.* (1976) showed that the two-point scaling function, $\langle\sigma(0)\sigma(x)\rangle$, is a function of $r = |x|$ and can be written as

$$\langle\sigma(0)\sigma(x)\rangle = \begin{cases} \cosh(\psi/2) \\ \quad \times \exp\dfrac{1}{4}\displaystyle\int_r^\infty \left(\sinh^2\psi - \left(\dfrac{d\psi}{dt}\right)^2\right) s\, ds \\ \quad \times (T \uparrow T_c) \\ \sinh(\psi/2) \\ \quad \times \exp\dfrac{1}{4}\displaystyle\int_r^\infty \left(\sinh^2\psi - \left(\dfrac{d\psi}{dt}\right)^2\right) s\, ds \\ \quad \times (T \downarrow T_c) \end{cases} \qquad [1]$$

where $\psi = \psi(r)$ satisfies the differential equation,

$$\frac{d}{dr}\left(r\frac{d\psi}{dr}\right) = \frac{r}{2}\sinh(2\psi)$$

The substitution $\eta = e^{-\psi}$ transforms this differential equation into a Painlevé equation of the third kind. This was used by McCoy, Tracy, and Wu (see Palmer and Tracy (1981) and references within) to study the short-distance behavior, $r \to 0$, of the scaling functions – behavior which is far from manifest in the infinite series expansions obtained for the scaling functions.

## Deformation Theory

Sato, Miwa, and Jimbo showed that there was a class of quantum field theories that included the scaling limits of the Ising model which have the property that the $n$-point correlations are "tau functions" for monodromy-preserving deformations of the Dirac equation in two dimensions. The two-dimensional (Euclidean) Dirac operator is

$$\mathcal{D} = \begin{bmatrix} m & -2\partial \\ -2\bar{\partial} & m \end{bmatrix}$$

with

$$\partial = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right), \qquad \bar{\partial} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)$$

the usual complex derivatives acting on smooth functions on $\boldsymbol{R}^2$. The monodromy-preserving deformations mentioned above are families of (multivalued) solutions $w(x)$ to

$$\mathcal{D}w = 0 \qquad [2]$$

which are branched at points $a_j \in \boldsymbol{R}^2 (j = 1, 2, \ldots, n)$ and change by a factor $e^{2\pi i \ell_j}$ as $x$ makes a small circuit about $a_j$. SMJ (1979b) show that the $L^2(\boldsymbol{R}^2)$ (square-integrable) solutions $w(x)$ of the Dirac equation with this prescribed branching behavior comprise an $n$-dimensional subspace of $L^2(\boldsymbol{R}^2)$. The constants $e^{2\pi i \ell_j}$ are called the monodromy of the solutions and the term "deformation" in the description refers to the fact that the monodromy constants do not change as the branch points $a_j$ are varied. SMJ show that it is possible to choose a basis $w_j(x,a) = w_j(x, a_1, \ldots, a_n)$ $(j = 1, \ldots, n)$ so that the vector

$$W(x,a) := [w_1(x,a), w_2(x,a), \ldots, w_n(x,a)]$$

becomes a section for a flat (Dirac compatible) connection in the $(x, a)$ variables. That is,

$$d_{x,a} W = \Omega(x,a) W$$

where $d_{x,a}$ is the exterior derivative in the $x$ and $a$ variables and $\Omega$ is a matrix-valued 1-form that satisfies the zero curvature condition,

$$d_{x,a}\Omega = \Omega \wedge \Omega$$

They also introduced the notion of a tau function, $\tau(a)$, for such deformations. The logarithmic derivative $d_a \log \tau(a) = \omega$, where $\omega$ is a 1-form on $\boldsymbol{R}^2 \backslash \{a_1, a_2, \ldots, a_n\}$ expressed in terms of the matrix elements of $\Omega$. The 1-form $\omega$ introduced by SMJ is shown to be closed when $\Omega$ satisfies the zero curvature condition above. The scaling limit of the Ising model is related to the situation for monodromy multipliers –1 and when the scaling limits of correlations are identified as suitable $\tau$-functions in this case, the WMTB result emerges when the nonlinear zero curvature condition is identified with a Painlevé equation.

The connection between the deformation theory and quantum field theory is developed in the computationally intensive paper SMJ (1980). Extensive use is made of local operator product expansions, analytic continuation, and formal series expansions that are infinite-dimensional analogs of Wick-type theorems for finite-dimensional spin representations (developed by SMJ (1978)). One can get a feeling for the source of the connection by recalling that in one of the "exact solutions" of the two-dimensional Ising model the spin operators, $\sigma(a)$, are identified as elements of an infinite spin representation of the orthogonal group and are characterized by their linear action on Fermionic variables (Palmer and Tracy 1981). In the physics literature, the $\sigma(a)$ are referred to as Bogoliubov transformations. In the scaling limit the associated representation space is the home to a free Fermi field $\psi(x)$, an operator-valued solution to the Dirac equation. Of course, $\psi(x)$ has components $\psi_j(x)$ but for simplicity we will suppress such details in the mostly schematic discussion that follows. For coincident second coordinates $x_2 = a_2$ the Fermi field $\psi(x)$ and $\sigma(a)$ satisfy the commutation relation

$$\sigma(a)\psi(x) = -\mathrm{sgn}(x_1 - a_1)\psi(x)\sigma(a) \qquad [3]$$

which is a surviving remnant of the linear action of $\sigma(a)$ on lattice fermions. In the transfer matrix formalism, which is natural for statistical mechanics, translation in the "space" variable $x_1$ is unitary, but translation in the "time" variable, $x_2$, is governed by the transfer matrix, the generator of a contractive semigroup. Because of this, the quantities that are well behaved in this formalism are "time-ordered vacuum expectations"; these involve only "positive" powers of the transfer matrix. Let $\mathcal{T}$ denote the "time"-ordering operator; a sequence of operators depending on coordinates in $\boldsymbol{R}^2$ is reordered following $\mathcal{T}$ so that the second coordinates appear in increasing order from left to right. Sign changes are incorporated whenever it is necessary to exchange Fermi type operators like $\psi^*(x)$ and $\psi(y)$ to put them in the correct order. In the Euclidean setting (pure imaginary time) it is well known that

$$G(x,y) = \langle \mathcal{T}\psi^*(x)\psi(y)\rangle$$

is a Green function for the Dirac operator $\mathcal{D}$ (the distribution kernel for $\mathcal{D}^{-1}$).

This observation and [3] suggests that the hybrid vacuum expectation

$$G(x,y;a) = \frac{\langle \mathcal{T}\psi^*(x)\psi(y)\sigma(a_1)\cdots\sigma(a_n)\rangle}{\langle \mathcal{T}\sigma(a_1)\cdots\sigma(a_n)\rangle}$$

should be the Green function for a Dirac operator with a domain containing "functions" branched at the points $a_j$ having "monodromy" –1 there. It is possible to recast the SMJ analysis so that a Dirac operator, $\mathcal{D}(a)$, on a suitable vector bundle with base $\boldsymbol{R}^2 \backslash \{a_1, \ldots, a_n\}$ becomes the central player (see Palmer *et al.* (1994) and references therein). The data for the vector bundle includes the factors $e^{2\pi i \ell_j}$ incorporated in transition functions for the bundle. The $\tau$-function becomes an infinite determinant (or Pfaffian in the Ising case)

$$\tau(a) = \det \mathcal{D}(a) \qquad [4]$$

in the Segal–Wilson sense (see Palmer *et al.* (1994) and references therein). The Green function $G(x,y;a)$ has a finite-rank derivative,

$$d_a G(x,y;a) = \sum_j r_j(x,a)s_j(y,a)\,\mathrm{d}a_j + u_j(x,a)v_j(y,a)\,\mathrm{d}\bar{a}_j \qquad [5]$$

which is the key result in this version of the SMJ analysis (this observation appears in SMJ (1980) but does not have a central role there). The "wave functions" $r$, $s$, $u$, and $v$ are closely related to the $L^2$ wave functions $w_j$ described above. Equation [5] is both the source of the deformation equations for $r$, $s$, $u$, and $v$ which arise from $d_a^2 G = 0$ coupled with the rotational and translational symmetries of the Green function, and also of the expression for $d_a \log \tau(a)$ in terms of data associated with the deformation theory. A "transfer matrix" calculation of the determinant allows one to make the connection with the scaling limits of lattice fields including the Ising model (see Palmer *et al.* (1994) and references therein).

The short-distance behavior of the two-point function for the Ising model scaling functions has been rigorously calculated by Tracy and later by Tracy and Widom (see Harnad and Its (2002) and references therein). A less detailed analysis of the short-distance behavior of the $n$ point functions that uses the deformation analysis of the correlations in a crucial way can be found in Palmer (2000).

## The Riemann–Hilbert Problem

In SMJ (1979b), a "massless" version of holonomic fields is developed. This concerns monodromy-preserving deformations of the Cauchy–Riemann operator $\bar{\partial}$. The techniques used to study this lead back to the Riemann–Hilbert problem – the problem of determining a linear differential equation in the complex plane with rational coefficients and prescribed monodromy at the poles of the coefficients. More specifically, suppose one is given $n$ distinct points $\{a_1,\ldots,a_n\}$ in $\boldsymbol{P}^1$, the Riemann sphere, and a base point $a_0$ distinct from the $a_j, j \neq 0$. Let $\gamma_j$ denote a simple closed curve based at $a_0$ which winds counterclockwise once around $a_j$ but has winding number 0 for the other points $a_k, k \neq j$. Choose $n$ invertible $p \times p$ matrices $M_j$ which satisfy the single condition $M_1 M_2 \cdots M_n = I$. Then, the homotopy classes of the curves $\gamma_j$ are the generators for the fundamental group of the punctured sphere $\boldsymbol{P}^1 \backslash \{a_1,\ldots,a_n\}$ with base point $a_0$ and the map which sends $\gamma_j \to M_j$ determines a representation of the fundamental group. One version of the Riemann–Hilbert problem is to find $p \times p$ complex matrices $A_j$ for $j = 1,\ldots,n$ so that the linear differential equation

$$\frac{\mathrm{d}Y}{\mathrm{d}z} = \sum_j \frac{A_j}{z - a_j} Y \qquad [6]$$

has monodromy representation given by $\gamma_j \to M_j$. This means that the fundamental solution $Y(z)$ defined in a neighborhood of $z = a_0$ and normalized so the $Y(a_0) = I$ (the identity) will become the fundamental solution $Y(z)M_j^{-1}$ after analytic continuation around the curve $\gamma_j$. This form of the problem does not always have a solution but when it does, it is interesting to consider deformations $a \to A_j(a)$ that preserve the monodromy multipliers $M_j$. Such monodromy-preserving deformations were first considered by Schlesinger in 1912 (see Palmer and Tracy (1981) and references therein) and he discovered that the coefficients $A_j$ must satisfy nonlinear differential equations that, for $a_0 = \infty$, can be written as

$$d_a A_j = \sum_{k \neq j} \frac{A_k - A_j}{a_k - a_j} d(a_k - a_j)$$

SMJ introduced $\tau$-functions associated with these deformations and they gave these $\tau$-functions a quantum field theory interpretation as $n$ point functions. Eventually this theory was extended to include the Birkhoff generalization of the Riemann–Hilbert problem, a generalization which incorporates the additional information needed to fix local holomorphic equivalence at higher-order poles (formal asymptotics and Stokes' multipliers) (Jimbo and Miwa 1981, Sato *et al.* 1978). Roughly speaking, the problem is to reconstruct a global connection with specified singularities from its local holomorphic equivalence data and its global monodromy representation. Thinking of the differential equation [6] as a holomorphic connection proved very helpful in a geometric reworking of the SMJ analysis given by Malgrange (1983a, 1983b) who showed that the zeros of the $\tau$-function occurred at points where a suitably defined Riemann–Hilbert problem fails to have a solution (see also Palmer (1999) references within). The mathematical significance of massless holonomic quantum fields as (quantized) singular elements of a gauge group is apparent from the SMJ work and later work of Miwa but the possibility of interesting physics in these models does not seem to have been much investigated at this time. These quantum fields are also conformal fields; however, a comprehensive integration into the highly developed formalism of conformal field theory on compact Riemann surfaces has not currently been developed

(an analog of [5] should survive on compact Riemann surfaces but the deformation analysis of the correlations is likely limited to symmetric spaces).

## Further Developments

This work on massless holonomic fields and the connection with the Riemann–Hilbert problem is doubtless the aspect of holonomic fields with the most "spin offs" in the mathematics and physics literature. These include an analysis of the delta-function gas done by Jimbo, Miwa, Mori, and Sato in 1981, random matrix models first looked at by Jimbo, Miwa, Mori, and Sato and later systematically investigated by Tracy and Widom (1994), the deformations of line bundles on Riemann surfaces that led to KdV in the work of Segal and Wilson (1985), which emerged from work of Sato, Miwa, Jimbo and collaborators, the analysis of Painlevé equations starting with work of McCoy, Tracy and Wu (see Palmer and Tracy (1981) and references within) and more systematically developed by Its and Novokshenov (1986), and the revival of interest in monodromy-preserving deformations (Harnad and Its 2002).

Holonomic fields are related to free fields in a well-understood way and it is natural to study them in situations where free fields make sense. In particular, they are an interesting testbed for the nonperturbative investigation of the influence of geometry (or curvature) on quantum fields. In Palmer *et al.* (1994), the deformation analysis of $\tau$-functions for holonomic fields is carried out for the Poincaré disk. The two-point functions are shown to be expressible in terms of solutions to the family of Painlevé VI equations. A quantum field theory interpretation of these $\tau$-functions is given by Doyon and there are natural analogs of the scaling limit of the Ising model on the Poincaré disk as well. The role of "spacetime" symmetries in the deformation theory suggests that such analysis will be limited to symmetric spaces. In addition to the plane and the Poincaré disk, the cylinder, the sphere, and the torus round out the possibilities in two dimensions. Lisovyy has recently worked out the analysis for the cylinder, which is important for the study of thermodynamic correlations. It should be possible to recast the analysis of the continuum Ising model on the torus (Zuber and Itzykson 1977) in deformation theoretic terms. It does not appear that the holonomic fields associated with the Dirac operator for the constant curvature metric on the 2-sphere have been studied yet.

*See also:* Deformation Theory; Integrable Systems: Overview; Isomonodromic Deformations; Riemann–Hilbert Problem; Two-Dimensional Ising Model.

## Further Reading


Harnad J and Its A (2002) *CRM Workshop: Isomonodromic Deformations and Applications in Physics*, vol. 31 American Mathematical Society.

Its AR and Novokshenov VYu (1986) *The Isomonodromic Deformation Method in the Theory of Painlevé Equations*, Lecture Notes in Mathematics, vol. 1191. Berlin: Springer.

Jimbo M and Miwa T (1981) Monodromy preserving deformations of linear ordinary differential equations with rational coefficients II. *Physica D* 2: 407–448.

Jimbo M, Miwa T, and Ueno K (1981) Monodromy preserving deformations of linear ordinary differential equations with rational coefficients I. *Physica D* 2: 306–352.

Malgrange B (1983a) *Sur les déformations isomonodromiques I, Singularités régulières*, Progress in Mathematics vol. 37, pp. 401–426. Boston: Birkhäuser.

Malgrange B (1983b) *Sur les déformations isomonodromiques II, Singularités irrégulières*, Progress in Mathematics vol. 37, pp. 427–438. Boston: Birkhäuser.

Palmer J (1999) Zeros of the Jimbo, Miwa, Ueno tau function. *Journal of Mathematical Physics* 40: 6638–6681.

Palmer J (2000) Short distance asymptotics of Ising correlations. *Journal of Mathematical Physics* 7: 1–50.

Palmer J, Beatty M, and Tracy C (1994) Tau functions for the Dirac operator on the Poincaré disk. *Communications in Mathematical Physics* 165: 97–173.

Palmer J and Tracy CA (1981) Two dimensional Ising correlations: Convergence of the scaling limit. *Advances in Applied Mathematics* 2: 329–388.

Sato M, Miwa T, and Jimbo M (1978) Holonomic quantum fields I. *Publications of the Research Institute for Mathematical Sciences, Kyoto University* 14: 223–267.

Sato M, Miwa T, and Jimbo M (1979a) Holonomic quantum fields II. *Publications of the Research Institute for Mathematical Sciences, Kyoto University* 15: 201–278.

Sato M, Miwa T, and Jimbo M (1979b) Holonomic quantum fields III. *Publications of the Research Institute for Mathematical Sciences, Kyoto University* 15: 577–629.

Sato M, Miwa T, and Jimbo M (1979c) Holonomic quantum fields IV. *Publications of the Research Institute for Mathematical Sciences, Kyoto University* 15: 871–972.

Sato M, Miwa T, and Jimbo M (1980) Holonomic quantum fields V. *Publications of the Research Institute for Mathematical Sciences, Kyoto University* 16: 531–584.

Tracy CA and Widom H (1994) Fredholm determinants, differential equations and matrix models. *Communications in Mathematical Physics* 163: 33–72.

Wu TT, McCoy B, Tracy CA, and Barouch E (1976) Spin–spin correlation functions for the two dimensional Ising model: exact theory in the scaling region. *Physical Review B* 13: 316–374.

Wilson G and Segal G (1985) Loop groups and equations of KdV type. *Publications of Mathématiques IHES* 61: 5–65.

Zuber JB and Itzykson C (1977) Quantum field theory and the two dimensional Ising model. *Physical Review D* 15: 2875–2884.

# Homeomorphisms and Diffeomorphisms of the Circle

**A Zumpano and A Sarmiento**, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil



## Introduction

In this article we consider the following question: which homeomorphisms of the circle transport one given class of continuous functions into another? The allowed classes of functions are Banach spaces contained in $C(\mathbb{T})$, the space of continuous functions on the unit circle $\mathbb{T}$, and will be defined by the properties of the Fourier series of the functions. Next, we will develop the theory of Poincaré–Denjoy which describes some basic geometric properties about diffeomorphisms of the circle such as existence and properties of the rotation number, classifications of possible orbits of diffeomorphisms, and Denjoy counterexample.

A homeomorphism of the circle is regarded here as a change of variables for periodic functions. So, it will be our major concern to describe the changes of variables that do not affect "too much" the behavior of the Fourier series of the functions in the given class.

We say that a function $h:\mathbb{R}\to\mathbb{R}$ is a homeomorphism of the circle $\mathbb{T}=\{(x,y)\in\mathbb{R}^2 : x^2+y^2=1\}$, if $h$ itself is a homeomorphism such that $h(t+2\pi)=h(t)\pm 2\pi$ for all $t\in\mathbb{R}$. It is clear that such function $h$ induces a unique homeomorphism $\widetilde{h}:\mathbb{T}\to\mathbb{T}$ that makes the following diagram commutative:

$$\begin{array}{ccc} \mathbb{T} & \xrightarrow{\widetilde{h}} & \mathbb{T} \\ e^{it}\downarrow & & \downarrow e^{it}, \\ \mathbb{R} & \xrightarrow[h]{} & \mathbb{R} \end{array} \quad \text{i.e.,}\ \widetilde{h}(e^{it})=e^{ih(t)}$$

In the same way, we identify functions $\widetilde{\psi}:\mathbb{T}\mapsto\mathbb{C}$ with $2\pi$-periodic functions $\psi:\mathbb{R}\mapsto\mathbb{C}$.

Let $\mathcal{U}(\mathbb{T})$ be the space of all continuous functions on $\mathbb{T}$ that have uniformly convergent Fourier series, and let $\mathcal{A}(\mathbb{T})$ be the space of all continuous functions on $\mathbb{T}$ with absolute convergent Fourier series.

In 1953, Beurling and Helson proved an important result about the homeomorphisms that preserve the space $\mathcal{A}(\mathbb{T})$: they are rotations and symmetries, that is, if $f\circ h\in\mathcal{A}(\mathbb{T})$ for all $f\in\mathcal{A}(\mathbb{T})$, then the homeomorphism $h$ must have the form $h(t)=t+\alpha$ or $h(t)=-t+\alpha$. It is quite obvious that rotations and symmetries preserve $\mathcal{A}(\mathbb{T})$, since the Fourier coefficients of $f\circ h$ and $f$ have the same modulus, but to prove the converse is very hard. So, homeomorphisms that preserve $\mathcal{A}(\mathbb{T})$ are a very restrict class.

A wider class is obtained when we transport $\mathcal{A}(\mathbb{T})$ into $\mathcal{U}(\mathbb{T})$, that is, $f\circ h\in\mathcal{U}(\mathbb{T})$ for all $f\in\mathcal{A}(\mathbb{T})$. The major object of this article is to study such changes of variables.

We say that a homeomorphism of the circle $h$ is of finite type, if there is an integer $\nu$, satisfying $3\le\nu<\infty$, such that $h$ is of class $C^\nu$ and

$$|h''(t)|+|h^{(3)}(t)|+\cdots+|h^{(\nu)}(t)|\neq 0, \quad \text{for all } t\in\mathbb{R}$$

In the realm of Fourier analysis, the most important and general result about homeomorphisms of the circle is due to R Kaufman, who showed in 1974 that a finite-type homeomorphism $h$ transports $\mathcal{A}(\mathbb{T})$ into $\mathcal{U}(\mathbb{T})$. We shall analyze in detail such seminal result.

## Homeomorphism of the Circle of Finite Type

In this section we prove the theorem of R Kaufman mentioned before, which means that it is sufficient for a homeomorphism of the circle $h$ to have a certain amount of curvature in order to transport $\mathcal{A}(\mathbb{T})$ into $\mathcal{U}(\mathbb{T})$. We present a simple proof of this fact, based on a result due to Stein and Wainger.

If $f:\mathbb{T}\to\mathbb{C}$ is a continuous function and if

$$\hat{f}_n=\frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}\,dt, \quad n\in\mathbb{Z}$$

denote the Fourier coefficients of $f$, then $f\in\mathcal{A}(\mathbb{T})$ if and only if

$$\sum_{n\in\mathbb{Z}}|\hat{f}_n|=\lim_{N\to\infty}\sum_{-N}^{N}|\hat{f}_n|<\infty$$

Of course, $\mathcal{A}(\mathbb{T})$ is a Banach space with the norm

$$\|f\|_{\mathcal{A}(\mathbb{T})}=\sum_{n\in\mathbb{N}}|\hat{f}_n|$$

The space $\mathcal{U}(\mathbb{T})$ is defined as the space of all continuous functions $f:\mathbb{T}\to\mathbb{C}$ such that

$$\sum_{-N}^{N}\hat{f}_n e^{int}\to f(t), \quad \text{when } N\to\infty, \quad \text{for all } t\in[-\pi,\pi]$$

uniformly on $\mathbb{T}$, that is, $\mathcal{U}(\mathbb{T})$ is the space of continuous functions from $\mathbb{T}$ to $\mathbb{C}$ that are the uniform limit of their Fourier partial sums

$$S_N(f,t)=\sum_{-N}^{N}\hat{f}_n e^{int}$$

Hence, under the natural norm given by

$$\|f\|_{\mathcal{U}(\mathbb{T})} = \sup\{|S_N(f,t)|\colon N \in \mathbb{N} = \{0,1,\ldots\} \text{ and } t \in [-\pi,\pi]\}$$

the space $\mathcal{U}(\mathbb{T})$ is a Banach space.

We shall prove:

**Theorem 1** (Kaufman 1974). *Let $h$ be a homeomorphism of the circle of class $C^\nu$, with $\nu \geq 3$. Suppose that*

$$|h''(t)| + |h^{(3)}(t)| + \cdots + |h^{(\nu)}(t)| \neq 0, \quad \textit{for all } t \in \mathbb{R}$$

*Then, $h$ transports $\mathcal{A}(\mathbb{T})$ into $\mathcal{U}(\mathbb{T})$, that is, $f \circ h \in \mathcal{U}(\mathbb{T})$, whenever $f \in \mathcal{A}(\mathbb{T})$.*

It follows from the theorem that an analytic homeomorphism of the circle transports $\mathcal{A}(\mathbb{T})$ into $\mathcal{U}(\mathbb{T})$. To see this, suppose that $h$ is not of finite type. Then, for each $n \geq 3$, there exists $t_n \in [-\pi,\pi]$ such that $h^{(j)}(t_n) = 0$ for all $j \in \{2,\ldots,n\}$. Since $\{t_n\}$ has a convergent subsequence, there exists $t \in [-\pi,\pi]$ such that $h^{(j)}(t) = 0$ for all $j \geq 2$. This implies that $h''$ must be a constant function and, therefore, $h(t) = \pm t + \alpha$. Since we know that this kind of homeomorphism preserves $\mathcal{A}(\mathbb{T})$, we are done.

One can ask why to demand $\nu \geq 3$. The answer is easy. Since $h(t+2\pi) = h(t) \pm 2\pi$ for all $t \in \mathbb{R}$, it follows that $h'(t+2\pi) = h'(t)$ for all $t \in \mathbb{R}$, that is, $h'$ is a periodic function of period $2\pi$. So, it will always exist a point $t \in (\pi,\pi)$ such that $h''(t) = 0$.

We can also infer from the theorem that a $C^\infty$ homeomorphism of the circle that has no flat point, that is, no point $t$ such that $h^{(j)}(t) = 0$ for all $j \geq 2$, transports $\mathcal{A}(\mathbb{T})$ into $\mathcal{U}(\mathbb{T})$. This is obvious, because the negation of being of finite type implies the existence of a flat point. It is not true, however, that every $C^\infty$ homeomorphism of the circle transports $\mathcal{A}(\mathbb{T})$ into $\mathcal{U}(\mathbb{T})$.

The proof of the theorem is based on the two lemmas that follows. The first lemma was obtained by Stein and Wainger, who proved it in a more general setting in 1965, although that proof was only published five years later. The second lemma was proved by R Kaufman in 1974.

**Lemma 2** (Stein and Wainger 1970). *Let $p(t)$ be a real polynomial of degree $d$. Then*

$$\left|\int_{-r}^{r} e^{ip(t)}\frac{dt}{t}\right| \leq 6(2^{d+1}) - 2d - 10 = 2d + 2\sum_{k=0}^{d}[3(2^k) - 2]$$

*for all $r > 0$.*

**Lemma 3** (Kaufman 1974). *Let $f$ be a real function of class $C^k$ on the interval $[-r,r]$, with $k \geq 2$. Suppose that $1 \leq |f^{(k)}(t)| \leq b$ for all $t \in [-r,r]$. Then*

$$\left|\int_{-r}^{r} e^{if(t)}\frac{dt}{t}\right| \leq C(k,b)$$

*where $C(k,b)$ is a constant that depends only on $k$ and $b$.*

We shall see that Lemma 3 can be proved from Lemma 2 in a quite simple way. The proof given by R Kaufman for Lemma 3 does not make use of Lemma 2 at all. Also, it is not difficult to see that Lemma 2 follows from Lemma 3, if we consider $d \geq 2$. So, they are indeed equivalent results.

Before getting into the proof of these two lemmas, let us state a result which is the primary tool in dealing with oscillatory integrals as those in the lemmas.

**Lemma 4** (Van der Corput lemma). *Let $f$ be real valued and smooth in $[a,b]$, with $0 < a < b$. Suppose that $|f^{(k)}(t)| \geq \lambda > 0$ for all $t \in [a,b]$. Then*

$$\left|\int_{a}^{b} e^{if(t)}\frac{dt}{t}\right| \leq [3(2^k) - 2]\frac{\lambda^{-1/k}}{a}$$

*holds if*

*(i)* $k \geq 2$, *and*
*(ii)* $k = 1$ *and $f'(t)$ is monotonic.*

Now, let us prove the two lemmas and Theorem 1.

**Proof of Lemma 2** The proof is by induction on the degree of the polynomial. Suppose that $p(t)$ is a polynomial of degree 0, that is, $p(t)$ is a constant function. In this case the result is trivial, since the integral is equal to zero.

By induction, assume that the statement is true for polynomials of degree less than or equal to $d$. Let

$$p(t) = a_{d+1}t^{d+1} + a_d t^d + \cdots + a_1 t + a_0, \quad a_{d+1} \neq 0$$

Make the change of variables $t = |a_{d+1}|^{-1/(d+1)}s$. Then we have

$$\int_{-r}^{r} e^{ip(t)}\frac{dt}{t} = \int_{-\sigma}^{\sigma} e^{i(q(t) \pm t^{d+1})}\frac{dt}{t}$$

where $\sigma = |a_{d+1}|^{1/(d+1)} r$ and $q(t)$ is a polynomial of degree at most equal to $d$. Suppose $\sigma > 1$. Then

$$\begin{aligned}\left|\int_{-\sigma}^{\sigma} e^{i(q(t)\pm t^{d+1})}\frac{dt}{t}\right| &\le \left|\int_{1}^{\sigma} e^{i(q(-t)\pm(-t)^{d+1})}\frac{dt}{t}\right| \\ &+ \left|\int_{1}^{\sigma} e^{i(q(t)\pm t^{d+1})}\frac{dt}{t}\right| \\ &+ \left|\int_{-1}^{1} e^{i(q(t)\pm t^{d+1})}\frac{dt}{t}\right| \\ &\le I + II + III\end{aligned}$$

By Van der Corput lemma, $I \le [3(2^{d+1}) - 2]$ and $II \le [3(2^{d+1}) - 2]$, so $I + II \le 6(2^{d+1}) - 4$. Now

$$\begin{aligned}III &\le \left|\int_{-1}^{1}\left[e^{i(q(t)\pm t^{d+1})} - e^{iq(t)}\right]\frac{dt}{t}\right| + \left|\int_{-1}^{1} e^{iq(t)}\frac{dt}{t}\right| \\ &\le \int_{-1}^{1}|t|^d\,dt + 6(2^{d+1}) - 2d - 10 \\ &\le 2 + 6(2^{d+1}) - 2d - 10\end{aligned}$$

since the degree of $q(t)$ is at most equal to $d$. So

$$\begin{aligned}&I + II + III \\ &\quad\le 6(2^{d+1}) - 4 + 2 + 6(2^{d+1}) - 2d - 10 \\ &\quad= 6(2^{d+1}) - 2(d+1) - 10\end{aligned}$$

On the other hand, if $\sigma \le 1$, then

$$\begin{aligned}&\left|\int_{-\sigma}^{\sigma} e^{i(q(t)\pm t^{d+1})}\frac{dt}{t}\right| \\ &\quad\le \left|\int_{-\sigma}^{\sigma}\left[e^{i(q(t)\pm t^{d+1})} - e^{iq(t)}\right]\frac{dt}{t}\right| + \left|\int_{-\sigma}^{\sigma} e^{iq(t)}\frac{dt}{t}\right| \\ &\quad\le 2 + 6(2^{d+1}) - 2d - 10 \\ &\quad\le 2 + 6(2^{d+1}) - 2d - 10 + 6(2^{d+1}) - 4 \\ &\quad\le 6(2^{d+1}) - 2(d+1) - 10\end{aligned}$$

and the proof is completed. □

**Proof of Lemma 3** Assume first that $r > 1$. Then

$$\begin{aligned}&\left|\int_{-r}^{r} e^{if(t)}\frac{dt}{t}\right| \\ &\quad\le \left|\int_{1}^{r} e^{if(-t)}\frac{dt}{t}\right| + \left|\int_{1}^{r} e^{if(t)}\frac{dt}{t}\right| + \left|\int_{-1}^{1} e^{if(t)}\frac{dt}{t}\right| \\ &\quad= I + II + III\end{aligned}$$

Since $|f^{(k)}(t)| \ge 1$ and $k \ge 2$, then by Van der Corput lemma, $I \le [3(2^k) - 2]$ and $II \le [3(2^k) - 2]$. (Note that we have to assume $k \ge 2$ in order to apply Van der Corput lemma, since we know nothing about the monotonicity of $f'(t)$.)

To evaluate $III$, we proceed as following:

$$\begin{aligned}f(t) &= f(0) + f'(0)t + \cdots + \frac{f^{(k-1)}(0)}{(k-1)!}t^{k-1} + \frac{f^{(k)}(\sigma_t t)}{k!}t^k \\ &= p(t) + f^{(k)}(\sigma_t t)\frac{t^k}{k!}\end{aligned}$$

where the number $\sigma_t$ depends on $t$ and $0 < \sigma_t < 1$. So

$$\begin{aligned}&\left|\int_{-1}^{1} e^{if(t)}\frac{dt}{t}\right| \\ &\quad\le \left|\int_{-1}^{1}\left[e^{i(p(t)+f^{(k)}(\sigma_t t)\frac{t^k}{k!})} - e^{ip(t)}\right]\frac{dt}{t}\right| + \left|\int_{-1}^{1} e^{ip(t)}\frac{dt}{t}\right| \\ &\quad\le b\frac{2}{k!} + 6(2^k) - 2(k-1) - 10\end{aligned}$$

by Lemma 2, since $p(t)$ is a polynomial of degree at most equal to $k-1$.

On the other hand, if $r \le 1$, it also follows from Lemma 2 that

$$\begin{aligned}&\left|\int_{-r}^{r} e^{if(t)}\frac{dt}{t}\right| \\ &\quad\le \left|\int_{-r}^{r}\left[e^{i(p(t)+f^{(k)}(\sigma_t t)\frac{t^k}{k!}} - e^{ip(t)}\right]\frac{dt}{t}\right| + \left|\int_{-r}^{r} e^{ip(t)}\frac{dt}{t}\right| \\ &\quad\le b\frac{2}{k!} + 6(2^k) - 2(k-1) - 10\end{aligned}$$

Hence

$$\left|\int_{-r}^{r} e^{if(t)}\frac{dt}{t}\right| \le C(k, b)$$

and we are done. □

**Proof of Theorem 1** Let $h$ be a homeomorphism of the circle satisfying the hypotheses of the theorem.

We claim: there exists $\delta > 0$ such that, for all $x \in [-\pi, \pi]$, there is $k$ depending on $x$, with $2 \le k \le \nu$, such that $|h^{(k)}(t + x)| \ge \delta$ for all $t$ satisfying $|t| \le \delta$.

The proof of the claim is simple: suppose that there is no such $\delta$. Then, for each $n \in \mathbb{N}$ and each $k$ with $2 \le k \le \nu$, there exist $x_n \in [-\pi, \pi]$ and $t_{kn}$ such that $|t_{nk}| \le 1/n$ and $|h^{(k)}(t_{nk} + x_n)| < 1/n$. Taking a subsequence if necessary, we have $x_n \to x \in [-\pi, \pi]$. Also, $t_{kn} \to 0$ when $n \to \infty$ for all such $k$. So, $h^{(k)}(t_{kn} + x_n) \to h^{(k)}(x)$ when $n \to \infty$. Since $|h^{(k)}(t_{kn} + x_n)| < 1/n$, we conclude that $h^{(k)}(x) = 0$ for all $k$ with $2 \le k \le \nu$, thus reaching a contradiction.

Now, let $f \in \mathcal{A}(\mathbb{T})$. So, $\sum_{-\infty}^{\infty}|\hat{f}_n| < \infty$, thus implying that

$$f(t) = \sum_{-\infty}^{\infty}\hat{f}_n e^{int} = \lim_{N\to\infty}\sum_{-N}^{N}\hat{f}_n e^{int}$$

Hence

$$f(h(t)) = \sum_{-\infty}^{\infty} \hat{f}_n e^{inh(t)} = \lim_{N\to\infty} \sum_{-N}^{N} \hat{f}_n e^{inh(t)}$$

Put $g_N(t) = \sum_{n=-N}^{N} \hat{f}_n e^{inh(t)}$. Since $g_N$ is smooth, we have $g_N \in \mathcal{U}(\mathbb{T})$ for all $N \in \mathbb{N}$. If $g(t)$ stands for $f(h(t))$, then $g_N \to g$ uniformly, since $f \in \mathcal{A}(\mathbb{T})$. Thus, it suffices to prove that $g \in \mathcal{U}(\mathbb{T})$. This happens if and only if $S_m(g,x) = \sum_{k=-m}^{m} \hat{g}_k e^{ikx}$ converges uniformly to $g$ as $m \to \infty$, that is, given $\epsilon > 0$, there exists $m_0 \in \mathbb{N}$ such that $|S_m(g,x) - g(x)| < \epsilon$ for all $m > m_0$ and $x \in [-\pi, \pi]$.

We have

$$\begin{aligned}&|S_m(g,x) - g(x)|\\&\quad \le |g_N(x) - g(x)| + |S_m(g_N,x) - g_N(x)|\\&\qquad + |S_m(g_N,x) - S_m(g,x)|\end{aligned}$$

for all $m, n \in \mathbb{N}$. Since $g_N \to g$ uniformly and $g_N \in \mathcal{U}(\mathbb{T})$, the last inequality shows that we need to demonstrate that, for each $\epsilon > 0$, there exists $N_0 \in \mathbb{N}$ such that

$$\begin{aligned}&|S_m(g_N,x) - S_m(g,x)| < \epsilon\\&\quad \forall\, N > N_0, x \in [-\pi,\pi] \quad \text{and} \quad m \in \mathbb{N}\end{aligned}$$

thus proving that $S_m(g_N,x) \to S_m(g,x)$ uniformly in $x$ and $m$ when $N \to \infty$.

But, if $K > N \in \mathbb{N}$, we have

$$\begin{aligned}&|S_m(g_N,x) - S_m(g_K,x)|\\&= \left|\frac{1}{2\pi}\int_{-\pi}^{\pi}(g_N(t+x) - g_K(t+x))D_m(t)\,dt\right|\\&= \left|\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\sum_{n=-N}^{N}\hat{f}_n e^{inh(t+x)} - \sum_{n=-K}^{K}\hat{f}_n e^{inh(t+x)}\right)\right.\\&\qquad \left.\times D_m(t)\,dt\right|\\&\le \left|\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\sum_{K\ge|n|\ge N}\hat{f}_n e^{inh(t+x)}\right)D_m(t)\,dt\right|\\&\le \frac{1}{2\pi}\sum_{K\ge|n|\ge N}|\hat{f}_n|\left|\int_{-\pi}^{\pi} e^{inh(t+x)}D_m(t)\,dt\right|\end{aligned}$$

where

$$D_m(t) = \sum_{k=-m}^{m} e^{ikt} = \frac{\sin(m+(1/2))t}{\sin(t/2)}$$

is the Dirichlet kernel.

Hence, we are done if we show that

$$\left|\int_{-\pi}^{\pi} e^{inh(t+x)}D_m(t)\right| \le C \qquad [1]$$

where C is a constant that does not depend on $m$, $n$, and $x$.

To prove that the oscillatory integral above is bounded, we make use of Lemma 3. We have that

$$D_m(t) = \frac{2\sin(mt)}{t} + O(1)$$

on any compact subset of $(-2\pi, 2\pi)$, that is,

$$\begin{aligned}&\left|\frac{\sin(m+(1/2))t}{\sin(t/2)} - \frac{2\sin(mt)}{t}\right|\\&\quad \le \left|\frac{t\cos(t/2) - 2\sin(t/2)}{t\sin(t/2)}\right| + 1 \le C^*\end{aligned}$$

where the constant $C^*$ does not depend on $m$, on any compact subset of $(-2\pi, 2\pi)$.

In order to prove [1], consider $x \in [-\pi, \pi]$. We have already proved that there exists $k$ (depending on $x$), with $2 \le k \le \nu$, such that $|h^{(k)}(t+x)| \ge \delta > 0$ for all $t$ such that $|t| \le \delta$. Therefore,

$$\begin{aligned}&\left|\int_{-\pi}^{\pi} e^{inh(t+x)}\frac{\sin(mt)}{t}\,dt\right| \le \left|\int_{-\delta}^{\delta} e^{inh(t+x)}\frac{\sin(mt)}{t}\,dt\right|\\&\quad + 2\log\left(\frac{\pi}{\delta}\right)\end{aligned}$$

We can assume that $n$ is a positive integer: if $n$ is negative, we take complex conjugate; and if $n = 0$, the integral is trivially bounded, as we see by integration by parts or by Van der Corput lemma. (Indeed, we do not need to worry about $n = 0$, since it is necessary to bound the integral only for large $n$.)

So, assuming that $n$ is a positive integer, we change variables: define $t = rs$, where $r = n^{-1/k}\delta^{-1/k}$. Since $\sin(mt) = (e^{imt} - e^{-imt})/(2i)$, we have

$$\begin{aligned}&\left|\int_{-\delta}^{\delta} e^{inh(t+x)}\frac{\sin(mt)}{t}\,dt\right|\\&\quad \le \left|\int_{-\delta}^{\delta} e^{i[nh(t+x)+mt]}\frac{dt}{t}\right| + \left|\int_{-\delta}^{\delta} e^{i[nh(t+x)-mt]}\frac{dt}{t}\right|\\&\quad = \left|\int_{-\delta/r}^{\delta/r} e^{i[nh(rs+x)+mrs]}\frac{ds}{s}\right|\\&\qquad + \left|\int_{-\delta/r}^{\delta/r} e^{i[nh(rs+x)-mrs]}\frac{ds}{s}\right|\end{aligned}$$

Put $\phi(t) = nh(rt+x) + mrt$ and $\psi(t) = nh(rt+x) - mrt$. We have $\phi^{(k)}(t) = nr^k h^{(k)}(rt+x)$ and $\psi^{(k)}(t) = nr^k h^{(k)}(rt+x)$. But, since $nr^k = 1/\delta$, we conclude that

$$\begin{aligned}&\left|\phi^{(k)}(t)\right| = \left|\psi^{(k)}(t)\right|\\&\quad = \frac{1}{\delta}\left|h^{(k)}(rt+x)\right| \ge 1, \quad \forall\, t \in \left[-\frac{\delta}{r}, \frac{\delta}{r}\right]\end{aligned}$$

Also,

$$\begin{aligned}|\phi^{(k)}(t)| = |\psi^{(k)}(t)| &\le b_k \\ &= \frac{1}{\delta}\max\left\{|h^{(k)}(s)|\colon -2\pi \le s \le 2\pi\right\}\end{aligned}$$

for all $t \in [-\delta/r, \delta/r]$. Therefore, by Lemma 3, we get

$$\begin{aligned}\left|\int_{-\delta/r}^{\delta/r} e^{i\phi(t)} \frac{dt}{t}\right| &\le C(k, b_k) \\ &\le \max\{C(j, b_j)\colon\ 2 \le j \le \nu\}\end{aligned}$$

and

$$\begin{aligned}\left|\int_{-\delta/r}^{\delta/r} e^{i\psi(t)} \frac{dt}{t}\right| &\le C(k, b_k) \\ &\le \max\{C(j, b_j)\colon\ 2 \le j \le \nu\}\end{aligned}$$

This concludes the proof. □

## Diffeomorphisms of the Circle

In this section we study the circle diffeomorphisms. This theory goes back to Poincaré (1885), who studied circle diffeomorphisms to decide when differential equations on the torus have periodic orbits of a specified type. For this he introduced the rotation number as an important dynamical invariant, which later turned out to be very fruitful in the theory of dynamical systems, and proved that a diffeomorphism with an irrational rotation number is combinatorially equivalent to a rotation with the same rotation number.

Denjoy (1932) constructed examples of diffeomorphisms of class $C^1$ with irrational rotation number having wandering intervals, in opposition to early ideas of Poincaré. It was necessary to assume that a diffeomorphism without periodic points is more smooth, in fact $C^2$, to prove that it is topologically conjugate to the rotation.

### The Poincaré Rotation Number

Let $\widetilde{h}: \mathbb{T} \to \mathbb{T}$ be an orientation-preserving homeomorphism. Given such a map, there is a (nonunique) map $h: \mathbb{R} \to \mathbb{R}$, which is called a lift of $\widetilde{h}$, such that $\widetilde{h} \circ p = p \circ h$, where $p: \mathbb{R} \to \mathbb{T}$ is covering map $p(t) = e^{2\pi i t}$.

A lift, $h$, of $\widetilde{h}$ satisfies:

1. $h$ is monotonically increasing, that is, $h(t_1) \le h(t_2)$ if $t_1 < t_2$.
2. $h(t+1) = h(t) + 1$ for all $t \in \mathbb{R}$, so $(h - \mathrm{id})$ has period 1.
3. If $h_1$ $h_2$ are two lifts of $\widetilde{h}$, then there is an integer $k$ such that $h_2(t) = h_1(t) + k$ for all $t \in \mathbb{R}$.

These conditions immediately yield the following: the transformation $h^k := h \circ \cdots \circ h$ is monotonically increasing and $h^k(t+r) = h^k(t) + r, t \in \mathbb{R}, k \in \mathbb{N}, r \in \mathbb{Z}$.

The rotation number gives an asymptotic indication (i.e., in the limit) of the average amount of rotation of a point along an orbit. We start by defining, for a lift $h$ of $\widetilde{h}$, the number

$$\rho_0(h, t) = \lim_{k\to\infty} \frac{h^k(t) - t}{k}$$

This limit exists and does not depend on the choice of the point $t \in \mathbb{R}$; so, we denote it by $\rho_0(h)$. If $h_1$ $h_2$ are two lifts of $\widetilde{h}$, then $\rho_0(h_1, t) - \rho_0(h_2, t)$ is an integer, so

$$\rho(\widetilde{h}) := \rho_0(h, t) \bmod 1$$

is well defined. The number $\rho(\widetilde{h}) \in [0, 1)$ is called the rotation number of $\widetilde{h}$, and depends continuously on $\widetilde{h}$. For detailed proof, see Katok and Hasselblatt (1995) or Robinson (1999).

**Theorem 5** *The rotation number $\rho(\widetilde{h})$ is rational if and only if $\widetilde{h}$ has a periodic point, this is, there exist $z_0 \in S^1$ and $k \in \mathbb{N}$ such that $\widetilde{h}^k(z_0) = z_0$.*

**Proof** Take a lift $h$ of $\widetilde{h}$ such that $h(0) \in [0, 1)$. Suppose that $\rho(\widetilde{h}) = q/m$.

If $\widetilde{h}$ has no fixed point. Then $h(t) - t \in \mathbb{R} \setminus \mathbb{Z}$ for all $t \in \mathbb{R}$, since $h(t) - t \in \mathbb{Z}$ implies that $p(t)$ is a point fixed for $\widetilde{h}$. In particular, $h(t) - t \ne q$ for all $t \in \mathbb{R}$, since $h - \mathrm{id}$ is continuous and periodic, there exist real numbers $a > 0$ such that $h(t) - t < q - a$ for all $t \in \mathbb{R}$. Then

$$\begin{aligned}&h^{km}(t) - h^{(k-1)m}(t) \\ &\quad = h^m[h^{(k-1)m}(t)] - [h^{(k-1)m}(t)] \\ &\quad < q - a, \quad \forall k \in \mathbb{N} \\ &\Downarrow \\ &h^{km}(t) - t \\ &\quad = \{h^m[h^{(k-1)m}(t)] - [h^{(k-1)m}(t)]\} \\ &\qquad + \{h^m[h^{(k-2)m}(t)] - [h^{(k-2)m}(t)]\} \\ &\qquad + \{h^m[h^{(k-3)m}(t)] - [h^{(k-3)m}(t)]\} + \cdots \\ &\qquad + \{h^m(t) - t\} < k(q - a)\end{aligned}$$

So

$$\begin{aligned}\frac{q}{m} = \rho(\widetilde{h}) &= \lim_{k\to\infty} \frac{h^{mk}(t) - t}{mk} \\ &\le \frac{1}{m} \lim_{k\to\infty} \frac{k(q-a)}{k} = \frac{q-a}{m}\end{aligned}$$

proving the claim by contraposition.

To see the converse, assume that there exists a periodic point $t_0 \in \mathbb{R}$, that is, there are $m, q \in \mathbb{Z}$ such that $h^m(t_0) = t_0 + q$ then

$$h^{km}(t_0) = t_0 + kq$$

$$\Rightarrow \rho(\widetilde{h}) = \lim_{k\to\infty} \frac{h^{mk}(t_0) - t_0}{mk} = \frac{q}{m} \qquad \square$$

**Corollary 6** *A homeomorphism $\widetilde{h}:\mathbb{T}\to\mathbb{T}$ does not have periodic points if and only if the rotation number $\rho(\widetilde{h})$ is irrational.*

Let $R_\lambda$ be defined on $\mathbb{T}$ by $R_\lambda(e^{2\pi i t}) = e^{2\pi i(t+\lambda)}$. This map is called a rigid rotation of angle $\lambda$ and it is easy to see that $h_\lambda(t) = t + \lambda$ is lift of $R_\lambda$ and that $\rho(R_\lambda) = \rho(h_\lambda) = \lambda \bmod 1$.

In this example we can see the connection between the rationality of the rotation number and the existence of a periodic orbit. Assume $\lambda = m/q$ is rational. Then $h_\lambda^q(t) = t + q\lambda = t + m$. Therefore, every point is periodic with period $q$. Now, assume that $\lambda$ is irrational. Since $h_\lambda^n(t) = t + n\lambda$ for all $n$, then $R_\lambda$ has no periodic points. In this case, show that every point in $\mathbb{T}$ has a dense orbit.

Now, again let $\widetilde{h}:\mathbb{T}\to\mathbb{T}$ be any orientation-preserving homeomorphism.

**Lemma 7** *If the rotation number of $\widetilde{h}$ is rational, then all periodic orbits have the same period.*

**Proof** If $\rho(\widetilde{h}) = m/q$ with $m, q \in \mathbb{Z}$ relatively prime, then we need to show that for any periodic point $z_0 = p(t)$ (where $p(t) = e^{2\pi i t}$ is a covering space projection of $\mathbb{T}$) there is a lift $h$ of $\widetilde{h}$ such that $h(0) \in [0, 1)$ for which $h^q(t) = t + m$. If $z_0$ is periodic point, then $h^r(t) = t + s$ for some $r, s \in \mathbb{Z}$ and

$$\frac{m}{q} = \rho(\widetilde{h}) = \lim_{n\to\infty} \frac{h^{rn}(t) - t}{nr} = \lim \frac{ns}{nr} = \frac{s}{r}$$

So that $s = km$ and $r = kq$. Then by monotonicity of $h$, we have that $h^q(t) = t + m$ as claimed. $\square$

### The Poincaré Denjoy Theory

A homeomorphism of the circle with rational rotation number has all its orbits asymptotic to periodic ones and this, together with Theorem 5, yields a complete classifications of the possible asymptotic behavior when the rotation number is rational. This motivates the study of the asymptotic behavior of orbits of homeomorphisms with irrational rotation number.

The $\omega$-limit set of a point $z_0 \in \mathbb{T}$ with respect to $\widetilde{h}$ is the set $\omega(z_0) = \{z \in \mathbb{T};\ \widetilde{h}^{n_k}(z_0) \to z$ as $n_k \to \infty$, for same sequence $\{n_k\}_{k=1}^{\infty}\}$. The $\alpha$-limit set $\alpha(z_0)$ of an arbitrary point $z_0 \in \mathbb{T}$ is defined similarly (with $n_k \to -\infty$ instead $n_k \to +\infty$).

Any orbit of a rotation $R_\lambda$ with irrational $\lambda$ is dense in $\mathbb{T}$, that is, $\omega(z_0) = \alpha(z_0) = \mathbb{T}$ for all $z_0 \in \mathbb{T}$.

**Theorem 8** (Poincaré 1885). *Let $\widetilde{h}:\mathbb{T}\to\mathbb{T}$ be an orientation-preserving homeomorphism with irrational rotation number. Then the $\omega$-limit set is independent of $x$ and is either $\mathbb{T}$ or perfect and nowhere dense.*

The preceding proposition says that maps with irrational rotation number have either all orbits dense or all orbits asymptotic to a Cantor set.

We say that two maps $f, g:\mathbb{T}\to\mathbb{T}$ are topologically conjugate if there exists a homeomorphism $h:\mathbb{T}\to\mathbb{T}$ such that $h \circ f = g \circ h$. This implies that $h \circ f^n = g^n \circ h$ for every integer $n$. Hence, the conjugacy $h$ maps orbits of $f$ into orbits of $g$. If a monotone map $l:\mathbb{T}\to\mathbb{T}$ satisfies $l \circ f = g \circ l$ but is not a necessarily homeomorphism, we only have that inverse image of each point is either a point or a closed interval. We say that $l$ is a semiconjugacy between $f$ and $g$; this case $l$ maps orbits or pack of orbits of $f$ into orbits of $g$.

**Theorem 9** (Denjoy 1932). *Let $\widetilde{f}:\mathbb{T}\to\mathbb{T}$ be an orientation-preserving diffeomorphism of class $C^2$, with irrational rotation number ($\rho(\widetilde{f}) = \lambda$). Then $\widetilde{f}$ is topologically conjugate to the rigid rotation $R_\lambda$.*

Note that in spite of the hypothesis of $\widetilde{f}$ being $C^2$, we obtain only a continuous conjugacy. It took almost 50 years until Michael Herman (1979) was able to solve the more difficult problem of obtaining a smooth conjugacy for rotation number satisfying extra arithmetic conditions.

If $\widetilde{f}$ is a circle homeomorphism which does not have periodic points, then there exists a semiconjugacy $h$ between $\widetilde{f}$ and a rotation $R_\lambda$. If $h$ is not a conjugacy, then there exists a point $x$ of the circle whose inverse image by $h$ is an interval $J$. Since $h \circ \widetilde{f} = R_\lambda \circ h$, we have that $h(\widetilde{f}^n(J)) = R_\lambda^n(x)$. It follows that the intervals of the family $\{J, f(J), f^2(J), \ldots\}$ are pairwise disjoint, and the $\omega$-limit set of $J$ does not reduce to a periodic orbit. We say that $J$ is a wandering interval of the map $\widetilde{f}$. Thus, $C^2$-differentiability implies that $\widetilde{f}$ does not have a wandering interval. For details of the proof of Theorem 9, see Melo and Strien (1993).

### The Denjoy Example

Denjoy also proved the following result, which shows that the hypothesis of class $C^2$ is essential.

**Theorem 10** (Denjoy 1932). *For any irrational number $\lambda \in [0, 1)$, there exists a $C^1$-circle diffeomorphism $f$ which has a wandering interval, and rotation number equal to $\lambda$.*

**Proof** The construction of a diffeomorphism with wandering interval will be done in the following manner. Given an irrational rotation $R_\lambda(e^{2\pi i t}) = e^{2\pi i(t+\lambda)}$, cut the circle $\mathbb{T}$ at all the points of an orbit $\{z_n = R_\lambda^n(e^{2\pi i t_0});\ n \in \mathbb{Z}\}$ of $R_\lambda$. In each cut insert a segment $J_n$ of length $l_n$ where $\sum_{n=-\infty}^{\infty} l_n = 1$. We obtain in this manner a new circle longer than the first. The open intervals correspond to the gaps of the Cantor set.

In order to construct $f$ formally. Let $l_n$ be a sequence of positive real numbers with $n \in \mathbb{Z}$ satisfying

(i) $\lim_{n\to\pm\infty}(l_{n+1}/l_n) = 1$
(ii) $\sum_{n=-\infty}^{\infty} l_n = 1$
(iii) $l_n > l_{n+1}$ for $n \geq 0$
(iv) $l_n < l_{n+1}$ for $n < 0$ and
(v) $3l_{n+1} - l_n > 0$ for $n \geq 0$

For example

$$l_n = T(|n|+2)^{-1}(|n|+3)^{-1}$$

where

$$T^{-1} = \sum_{n=-\infty}^{\infty}(|n|+2)^{-1}(|n|+3)^{-1}$$

Let $J_n$ be a closed interval of length $l_n$. We place these intervals on the circle in the same order as the order of the orbit $R_\lambda^n(0)$. So to place an interval $J_n$, consider the sum of the lengths of the intervals $J_i$ where $R_\lambda^i(0)$ is between $R_\lambda^n(0)$ and 0. This determines the placement of $J_n$.

The next step is to define $f$ on the union of the $J_n$. It is necessary and sufficient for $f'(t) = 1$ on the endpoint in order for the map to have a continuous derivative when it is extended to the closure. Assume $J_n = [a_n, b_n]$, so $l_n = b_n - a_n$. The integral

$$\int_{a_n}^{b_n}(b_n - t)(t - a_n)\,dt = \frac{l_n^3}{6}$$

so

$$\frac{6(l_{n+1} - l_n)}{l_n^3}\int_{a_n}^{b_n}(b_n - t)(t - a_n)\,dt = l_{n+1} - l_n$$

Therefore, if we define $f$ for $x \in J_n$ by

$$f(x) = a_{n+1} + \int_{a_n}^{x}\left[1 + \frac{6(l_{n+1} - l_n)}{l_n^3}(b_n - t)(t - a_n)\right]dt$$

then $f(b_n) = a_{n+1} + l_n + l_{n+1} - l_n = b_{n+1}$. Also, $f$ is differentiable on $J_n$ with

$$f'(x) = 1 + \frac{6(l_{n+1} - l_n)}{l_n^3}(b_n - x)(x - a_n)$$

Thus, $f'(a_n) = 1 = f'(b_n)$. Notice that for $n < 0$, $l_{n+1} - l_n > 0$, that

$$1 \leq f'(x) \leq 1 + \frac{6(l_{n+1} - l_n)}{l_n^3}\left(\frac{l_n}{2}\right)^2 = \frac{3l_{n+1} - l_n}{2l_n}$$

and $(3l_{n+1} - l_n)/(2l_n)$ goes to 1 as $n \to -\infty$. Similarly for $n \geq 0$ and $x \in J_n$,

$$1 \geq f'(x) \geq \frac{3l_{n+1} - l_n}{2l_n} > 0$$

so $f'(x)$ goes to 1 as $n \to +\infty$ uniformly for $x \in J_n$. From these facts, it follows that $f$ is uniformly $C^1$ on the union of the interiors of the $J_n$ and has a $C^1$ extension to all of $\mathbb{T}$.

Let $\Lambda = \mathbb{T} \setminus \cup_{n\in\mathbb{Z}}\operatorname{int}(J_n)$. This is a Cantor set. The orbit of a point $x \in \Lambda$ is dense in $\Lambda$ since it is like the orbit of 0 for $R_\lambda$. Thus, $\omega(x) = \Lambda$. If $x \in \operatorname{int}(J_n)$, then there is a smaller interval $I$ whose closure is contained in $\operatorname{int}(J_n)$. Since the interval $J_n$ never returns to $J_n$ but wanders among the other $J_k$, then $J_n$ is a wandering interval. □

## Further Results

In this section we shall state some additional results about homeomorphisms of the circle in the area of Fourier analysis.

The first result is a theorem of Pál (1914) and Bohr (1935): let $f: \mathbb{T} \to \mathbb{R}$ be a real continuous function; then, there exists a homeomorphism of the circle $h$ such that $f \circ h \in \mathcal{U}(\mathbb{T})$. The best proof of this theorem is due to Salem (1945). In 1978, Kahane and Katznelson showed that the result is still valid for $f: \mathbb{T} \to \mathbb{C}$ continuous.

A similar question was posed by Lusin: given a continuous function $f: \mathbb{T} \to \mathbb{R}$, is there a homeomorphism of the circle $h$ such that $f \circ h \in \mathcal{A}(\mathbb{T})$? The problem remained open until 1981, when Olevskii, Kahane, and Katznelson answered negatively the question: there exists a real (or complex) continuous function $f$ on the circle, such that, for all homeomorphism of the circle $h$, $f \circ h \notin \mathcal{A}(\mathbb{T})$.

It was proved by the author that there are $C^\infty$ homeomorphisms of the circle, not necessarily of finite type, that transport $\mathcal{A}(\mathbb{T})$ into $\mathcal{U}(\mathbb{T})$. It is a very technical work, published in 1998, and it gives a necessary and sufficient condition for a homeomorphism of the circle with a flat point to transport $\mathcal{A}(\mathbb{T})$ into $\mathcal{U}(\mathbb{T})$.

Finally, the Denjoy theorem (Theorem 9) is rather close to being optimal. The example constructed here can be improved by obtaining a circle diffeomorphism whose first derivatives have Hölder exponent arbitrarily close to 1 (see Katok and Hasselblatt (1995)). Recent

work has dealt with the existence of a differentiable conjugacy between a diffeomorphism $f$ with irrational rotation number $\lambda$ and $R_\lambda$. Arnol, Moser, and Herman have obtained results (see Melo and Strien (1993) for a discussion of this results and references).

## Acknowledgments

The author was supported in part by CNPq-Brazil. and was partially supported by FAPESP Grant # TEMÁTICO 03/03107-9.

*See also:* Chaos and Attractors; Ergodic Theory; Generic Properties of Dynamical Systems; Random Dynamical Systems; Wavelets: Mathematical Theory.

## Further Reading

Denjoy A (1932) Sur les courves définies par les équations differéntialles a la surface du tore. *Journal de Mathématiques Pures et Appliquées* 11(9): 333–375.

Herman M (1979) Sur les conjugation différentiable des difféomorphisms du cercle à des rotations. *Publ. Math. IHES* 49: 5–233.

Kahane JP (1970) *Séries de Fourier Absolument Convergentes,* p. 84. Berlin: Springer.

Kahane JP (1983) Quatre Leçons sur les Homéomorphismes du Circle et les Series de Fourier. In: *Topics in Modern Harmonic Analysis,* Proceedings of a Seminar Held in Torino and Milano, May–June 1982, vol. II, pp. 955–990. Roma.

Katok A and Hasselblatt B (1995) *Introduction to the Modern Theory of Dynamical Systems.* Cambridge: Cambridge University Press.

Katznelson Y (1976) *An Introduction to Harmonic Analysis*, 2nd edn, p. 217. New York: Dover.

Melo W and Strien S (1993) *One-Dimensional Dynamics*, ch. 1. Berlin: Springer.

Poincaré H (1881) Mémoire Sur les courves définies par les équations differéntialles I. *Journal of Mathématiques Pures et Appliquées* 3(7): 375–422.

Poincaré H (1882) Mémoire Sur les courves définies par les équations differéntialles II. *Journel de Math. Pure et Appl.* 8: 251–286.

Poincaré H (1885) Mémoire Sur les courves définies par les équations differéntialles III. *Journel de Math. Pure et Appl.* 4(series 1): 167–244.

Poincaré H (1886) Mémoire Sur les courves définies par les équations differéntialles IV. *Journel de Math. Pure et Appl.* 2: 151–217.

Robinson C (1962) *Dynamical Systems: Stability, Symbolic Dynamicas, and Chaos.* London: CRC Press.

Robinson C (1999) *Dynamical Systems: Stability, Symbolic Dynamics, and Chaos.* London: CRC Press.

Rudin W (1962) *Fourier Analysis on Groups*, ch. 4. New York: Wiley.

Stein EM (1986) In: Stein EM (ed.) *Oscillatory Integral and Fourier Analysis*, Beijing Lectures in Harmonic Analysis, pp. 307–355. Princeton: Princeton University Press.

Stein EM and Wainger S (1970) The estimation of an integral arising in multiplier transformation. *Studia Math.* 35: 101–104.

Zumpano A (1998) Infinite type homeomorphisms of the circle and convergence of Fourier series. *Transactions of the American Mathematical Society* 350(10): 4023–4040.

# Homoclinic Phenomena

**S E Newhouse**, Michigan State University, E. Lansing, MI, USA



## Introduction

Homoclinic orbits (or motions) were first defined by Poincaré in his treatise on the "restricted three-body problem." (Poincaré 1987) Further advances were made by Birkhoff (Birkhoff 1960) in the 1930s, and, by Smale in the 1960s. Since that time, they have been studied by many people and have been shown to be intimately related to our understanding of nonlinear dynamical systems. There are many systems which possess homoclinic orbits. In one striking example (as discussed in the book of Moser (1973), they can be used to account for the unbounded oscillatory motion discovered by Sitnikov in the three-body problem. They also commonly occur in two-dimensional mappings derived from periodically forced oscillations (e.g., see the book by Guckenheimer and Holmes (1983).

Roughly speaking, a homoclinic orbit is an orbit of a mapping or differential equation which is both forward and backward asymptotic to a periodic orbit which satisfies a certain nondegeneracy condition called "hyperbolicity." On its own, such an orbit is only of mild interest. However, these orbits induce quite interesting structures among nearby orbits, and this latter fact is responsible for the main importance of homoclinic orbits. In addition, when homoclinic orbits are created in a parametrized system, many interesting and unexpected phenomena arise.

In this article, we first describe the history and basic properties of homoclinic orbits. Next, we consider some simple polynomial diffeomorphisms of the plane (the so-called Hénon family) which exhibit homoclinic orbits. Subsequently, we discuss a general theorem due to Katok which gives sufficient conditions for the existence of such orbits. Finally, we briefly consider issues related to homoclinic bifurcations and some of their consequences.

## Homoclinic Orbits in Diffeomorphisms

Consider a discrete dynamical system given by a $C^r$ diffeomorphism $f: M \to M$ where $M$ is a $C^\infty$ manifold and $r$ is a positive integer. That is, $f$ is bijective and both $f$ and $f^{-1}$ are $r$-times continuously differentiable. Given a point $x \in M$, set $x_0 = x$. For non-negative integers $n$ we inductively define $x_{n+1} = f(x_n)$ and $x_{-n-1} = f^{-1}(x_{-n})$. We also write $f^n(x) = x_n$ for $n$ in the set $\mathbf{Z}$ of all integers. The "orbit" of $x$ is the set $O(x) = \{f^n(x): n \in Z\}$.

A "periodic point" $p$ of $f$ is a point such that there is a positive integer $N > 0$ such that $f^N(p) = p$. The least such number $\tau(p)$ is called the "period" of $p$. If $\tau(p) = 1$, we call $p$ a "fixed point." The periodic point $p$ with period $\tau$ is called called "hyperbolic" if all eigenvalues of the derivative $Df^\tau(p)$ at $p$ have absolute value different from 1. For convenience, we refer to the eigenvalues of $Df^\tau(p)$ as eigenvalues associated to $p$. If $p$ is a hyperbolic periodic point all of whose associated eigenvalues have norm less than one, we call $p$ a "sink" or "attracting periodic point." The opposite case in which all associated eigenvalues have norm larger than one is called a "source." A hyperbolic periodic point $p$ which is neither a source nor a sink is called a "saddle" or "hyperbolic saddle."

Given a saddle $p$ of period $\tau$, we consider the set $W^s(p) = W^s(p, f)$ of points $y \in M$ which are forward asymptotic to $p$ under the iterates $f^{n\tau}$. That is, the points $y \in M$ such that $f^{n\tau}(y) \to p$ as $n \to \infty$. This is called the "stable set" of $p$. Similarly, we consider the "unstable set" of $p$ which we may define as $W^u(p) = W^u(p, f) = W^s(p, f^{-1})$. The stable manifold theorem guarantees that $W^s(p)$ and $W^u(p)$ are injectively immersed submanifolds of $M$ whose dimensions add up to dim $M$. In these cases, they are called the stable and unstable manifolds of $p$, respectively. A point $q \in W^s(p) \bigcap W^u(p) \setminus \{p\}$ is called a "homoclinic point" of $p$ (or of the pair $(f, p)$). If the submanifolds $W^s(p)$ and $W^u(p)$ meet transversely at $q$, then $q$ is called a "transverse homoclinic point." Otherwise, $q$ is called a "homoclinic tangency."

In the special case when $M$ is a two-dimensional manifold, the stable and unstable manifolds of a saddle periodic point $p$ are injectively immersed curves in $M$. A transverse homoclinic point $q$ of $p$ is a point of intersection off $p$ where the curves are not tangent to each other. This is depicted in Figure 1 for the case of a saddle fixed point for the map $H(x, y) = (7 - x^2 - y, x)$, a member of the so-called Hénon family, which we will discuss later. The figure was made using the numerical package "Dynamics" which comes with the book by Nusse and Yorke (1998).

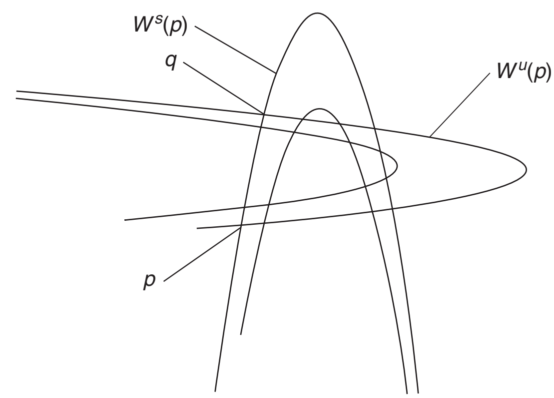



**Figure 1** Stable and unstable manifolds in the map $H(x, y) = (7 - x^2 - y, x)$ for the fixed point $p \approx (-3.83, -3.83)$.

One easily sees that every point in the orbit of a transverse homoclinic point $q$ of a hyperbolic saddle fixed point $p$ is again a transverse homoclinic point of $p$. Also, the curves $W^u(p)$ and $W^s(p)$ are invariant; that is, $f(W^u(p)) = W^u(p)$ and $f(W^s(p)) = W^s(p)$. This implies that the curves $W^u(p)$ and $W^s(p)$ extend, wind around, and accumulate on each other forming a complicated web.

Upon seeing this complicated structure in the restricted three-body problem, Poincaré very poetically wrote (p. 389, Poincaré 1987)

> Que l'on cherche à se représenter la figure formée par ces deux courbes et leurs intersections en nombre infini dont chacune correspond à une solution doublement asymptotique, ces intersections forment une sorte de treillis, de tissu, de réseau à mailles infiniment serrées; chacune des deux courbes ne doit jamais se recouper elle-même, mais elle doit se replier sur elle-même d'une manière trés complexe pour venir recouper une infinité de fois toutes les mailles du réseau.
> On sera frappé de la complexité de cette figure, que je ne cherche même pas à tracer. Rien n'est plus propre à nous donner une idée de la complication du problème des trois corps et en général de tous les problèmes de Dynamique où il n'y a pas d'intégrale uniforme ...

The next major advance concerning homoclinic orbits was made by Birkhoff (1960), who proved that in every neighborhood of a transverse homoclinic point of a surface diffeomorphism, one can find infinitely many distinct periodic points. Birkhoff also presented a symbolic description of the nearby orbits and noticed the analogy with Hadamard's description of geodesics on a surface. Birkhoff's analysis was generalized by Smale to arbitrary dimension, and, in addition, Smale gave a simpler analysis of the associated nearby orbits in terms of compact zero-dimensional

symbolic spaces which we now call “shift spaces” or “topological Markov chains.”

Once one knows that a diffeomorphism $f$ has a transverse homoclinic point for a saddle periodic point $p$, it is interesting to consider the closure of the orbits of all such homoclinic points. This turns out to be a closed invariant set containing a dense orbit and a countable dense set of periodic saddle points (Newhouse 1980). It is usually called a “homoclinic closure” or $h$-closure. These sets form the basis of chaotic or irregular motions in nonlinear systems.

## The Smale Horseshoe Map and Associated Symbolic System

To understand the geometric picture discovered by Smale, it is best to start with a concrete example of a diffeomorphism of the plane known as the “Smale horseshoe diffeomorphism.”

Given any homeomorphism $f : X \to X$ on a space $X$ and a subset $U \subset X$, let us define $I(f, U)$ to be the set of points $x \in X$ such that $f^n(x) \in U$ for every integer $n$. Thus, we have

$$I(f, U) = \bigcap_{n \in \mathbf{Z}} f^n(U)$$

We call $I(f, U)$ the invariant set of $f$ in $U$, or, alternatively, the invariant set of the pair $(f, U)$.

We now construct a special diffeomorphism $f$ of the Euclidean plane to itself in which $U = Q$ is the unit square and for which $I(f, U)$ has a very interesting structure. It is this map which is usually known as the Smale horseshoe map.

Let $Q = [0, 1] \times [0, 1]$ be the unit square in the plane $\boldsymbol{R}^2$. Let $0 < \alpha < 1/2$, and consider a diffeomorphism $f : \boldsymbol{R}^2 \to \boldsymbol{R}^2$ which is a composition of two diffeomorphisms $f = T_2 \circ T_1$ as follows. The map $T_1(x, y) = (\alpha^{-1}x, \alpha y)$ contracts vertically, expands horizontally, and maps $Q$ to the thin rectangle $Q_1 = \{(x, y) : 0 \le x \le \alpha^{-1}, 0 \le y \le \alpha\}$ which is short and wide. The map $T_2$ bends the right side of $Q_1$ up and around so that $T_2(Q_1) = f(Q)$ has the shape of a “horseshoe” or “rotated arch.” We arrange for $T_2$ to take the lower-right corner of $Q_1$ up to the upper-left corner of $Q$ in such a way that $f(Q)$ meets $Q$ in two full width subrectangles which we call $R_1$ and $R_2$. This can be done in such a way that the preimages $R_1^{-1} = T_1^{-1}(R_1)$ and $R_2^{-1} = T_1^{-1}(T_2^{-1}(R_2))$ are both full-height subrectangles of $Q$, and the restricted maps $f_1 \stackrel{\text{def}}{=} f \mid R_1^{-1}$ and $f_2 \stackrel{\text{def}}{=} f \mid R_2^{-1}$ are both affine. Thus, we arrange that $f_1$ is simply the restriction of $T_1$ to $R_1^{-1}$, and the map $f_2$ can be expressed in formulas as $f_2(x, y) = (-\alpha^{-1}x + \alpha^{-1}, -\alpha y + 1)$. This construction implies that $f$ will have the origin $p = (0, 0)$ as a

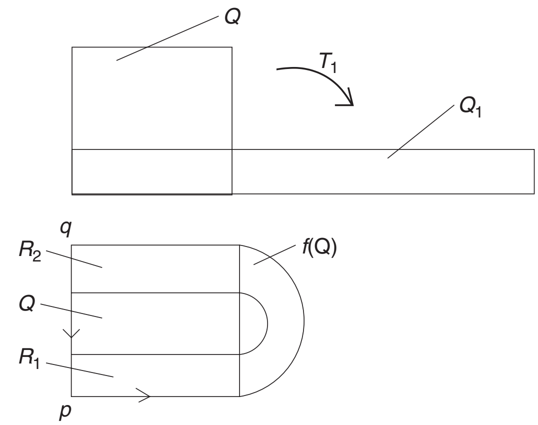



**Figure 2** The horseshoe map.

hyperbolic fixed point. We label the upper-left corner $(0, 1)$ of $Q$ with the letter $q$. It follows that the bottom and left edges of $Q$ will be in the unstable and stable manifolds of $p$, respectively, and we have indicated this in Figure 2 with small arrows.

The above construction gives us a diffeomorphism $f$ of the plane $\boldsymbol{R}^2$ such that $Q_1^+ \stackrel{\text{def}}{=} f(Q) \bigcap Q = R_1 \bigcup R_2$ is the union of two full-width subrectangles of $Q$. We wish to describe $I(f, Q)$. We begin with the sets $Q^+ = \bigcap_{n \ge 0} f^n(Q)$ and $Q^- = \bigcap_{n \ge 0} f^{-n}(Q)$. Thus, $Q^+$ is simply the set of points in $Q$ whose backward orbits stay in $Q$, and $Q^-$ is the set of points whose forward orbits stay in $Q$. For $i = 1, 2$, each rectangle $R_i$ is mapped to a thin horseshoe in $f(Q)$ which meets $Q$ in two full-width subrectangles. Combining these for $i = 1, 2$ gives four full-width rectangles as shaded in Figure 3. Thus, $Q \bigcap f(Q) \bigcap f^2(Q)$ consists of these four subrectangles. **Figure 3** shows the sets $f^2(Q), f^{-2}(Q)$ as well as the shaded rectangles we just mentioned.

Continuing in this way, one sees that, for each $n > 0$, the set $Q_n^+ = Q \bigcap f(Q) \bigcap \ldots \bigcap f^n(Q)$ consists of $2^n$ full-width subrectangles of $Q$, each with height

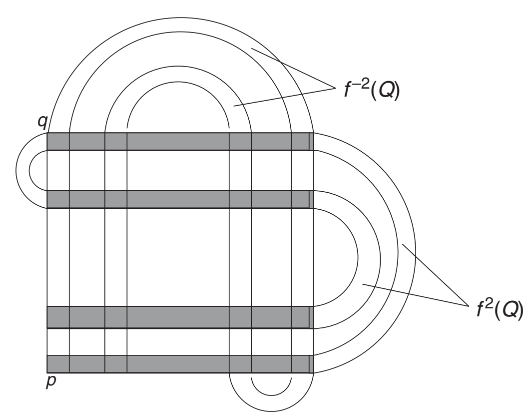



**Figure 3** The sets $f^2(Q)$ and $f^{-2}(Q)$ for the horseshoe map $f$.

$\alpha^n$. It follows that $Q^+ = \bigcap_n f^n(Q)$ is an interval times a Cantor set. Analogously, $Q^-$ is a Cantor set times an interval, and the set $I(f, Q)$ is a Cantor set in the plane. Let us recall the definition of a Cantor set $C$ in a metric space $X$. We first define a Cantor space $C$ to be a compact, perfect, totally disconnected metric space. That is, $C$ is a compact metric space, whose connected components are points such that every point $x$ in $C$ is a limit point of $C \setminus \{x\}$. A Cantor set $C$ in a metric space $X$ is a subset which is a Cantor space in the induced subspace (relative) topology.

The dynamics of $f$ on the invariant set $I(f, Q)$ can be conveniently described as follows.

Let $\Sigma_2 = \{1, 2\}^{\mathbf{Z}}$ be the set of doubly infinite sequences of 1's and 2's. Writing elements $\boldsymbol{a} \in \Sigma_2$ as $\boldsymbol{a} = (a_i) = (a_i)_{i \in Z}$, we define a metric $\rho$ on $\Sigma_2$ by

$$\rho(\boldsymbol{a}, \boldsymbol{b}) = \sum_{n \in \mathbf{Z}} \frac{1}{2^{|n|}} |a_i - b_i|$$

The pair $(\Sigma_2, \rho)$, then, is a Cantor space.

The "left-shift automorphism" on $\Sigma_2$ is the map $\sigma : \Sigma_2 \to \Sigma_2$ defined by $\sigma(\boldsymbol{a})_i = a_{i+1}$ for each $i \in \mathbf{Z}$. This is a homeomorphism from $\Sigma_2$ to itself. It has a dense orbit and a dense set of periodic points.

For a point $x \in I(f, Q)$, define an element $\phi(x) = \boldsymbol{a} = (a_i) \in \Sigma_2$ by $a_i = j$ if and only if $f^i(x) \in R_j$. It turns out that the map $\phi : I(f, Q) \to \Sigma_2$ is a homeomorphism such that $\sigma\phi = \phi f$.

In general, given two discrete dynamical systems $f : X \to X$, and $g : Y \to Y$, a homeomorphism $h : X \to Y$ such that $gh = hf$ is called a topological conjugacy from the pair $(f, X)$ to the pair $(g, Y)$. When such a conjugacy exists, the two systems have virtually the same dynamical properties.

In the present case, one sees that the dynamics of $f$ on $I(f, Q)$ is completely described by that of $\sigma$ on $\Sigma_2$.

It turns out the the Smale horseshoe map contains essentially all of the geometry necessary to describe the orbit structures near homoclinic orbits. To begin to see this, recall that the left and bottom boundaries of $Q$ were in the stable and unstable manifolds of $p$. Extending these curves as in Figure 4, one sees that the three corners of $Q$ different from $p$ are, in fact, all transverse homoclinic points of $p$.

It was a great discovery of Smale that, in the case of a general transverse homoclinic point, one sees the above geometric structure after taking some power $f^N$ of the diffeomorphism $f$. Thus, we have

**Theorem 1** (Smale). *Let $f : M \to M$ be a $C^1$ diffeomorphism of a manifold $M$ with a hyperbolic periodic point $p$ and a transverse homoclinic point $q$ of the pair $(f, p)$. Then, one can find a positive integer $N$ and a compact neighborhood $U$ of the points $p$ and $q$ such that the pair $(f^N, I(f^N, U))$ is topologically conjugate to the full 2-shift $(\sigma, \Sigma_2)$.*

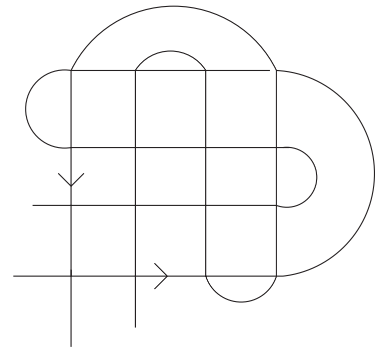


**Figure 4** Stable and unstable manifolds in the horseshoe map.

In modern language, we can assert that more is true. Let $\Lambda(f) = \bigcup_{0 \le j < N} f^j(I(f^N, U))$ be the $f$-orbit of the set $I(f^N, U)$. Then, $\Lambda(f)$ is a compact zero-dimensional hyperbolic basic set for $f$ with $V \stackrel{\text{def}}{=} \bigcup_{0 \le j < N} f^j(U)$ as an "adapted" or "isolating" neighborhood. This means that $\Lambda(f) = \bigcap_{n \in Z} f^n(V)$ is a compact, zero-dimensional hyperbolic set (see Robinson (1999) for definitions and related references) contained in the interior of $V$ and $f \mid \Lambda(f)$ has a dense orbit. If $g$ is $C^1$ near $f$, then $\Lambda(g) \stackrel{\text{def}}{=} \bigcap_{n \in Z} g^n(V)$ is a hyperbolic basic set for $g$ and the pairs $(f, \Lambda(f))$ and $(g, \Lambda(g))$ are topologically conjugate.

To get some appreciation for the magnitude of the contribution here, one might note the complicated arguments employed by Poincaré at the end of Poincaré (1987) to show that so-called heteroclinic points (intersections between stable and unstable manifolds of saddles with different orbits) existed. Birkhoff found a symbolic description (using infinitely many symbols) of the orbits near a transverse homoclinic orbit from which the existence of both infinitely many periodic and heteroclinic points is obvious. Smale extended the treatment of transverse homoclinic points to all dimensions, and found the symbolic description (using two symbols for some iterate of the map) given above. Moreover, Smale proved the "robustness" of these structures: they persist under small $C^1$ perturbations. Note that Poincaré's discovery of homoclinic points was in 1899, Birkhoff's results came in 1935, and Smale's results came in 1965. Thus, the above advances took over 65 years!

One can understand the geometry of Smale's construction fairly easily in the two-dimensional case. Let $q$ be the transverse homoclinic point of the saddle fixed point $p$ of the $C^r$ diffeomorphism $f$ on the plane $\mathbf{R}^2$. Given a small neighborhood $\tilde{U}$ of $p$, let

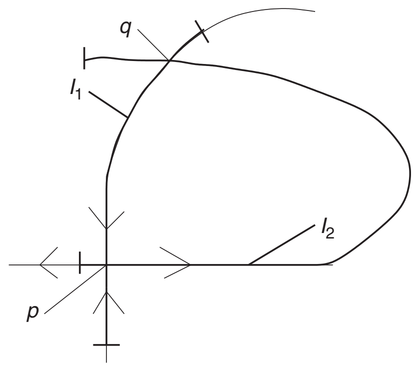



**Figure 5** The curves $I_1 \subset W^s(p)$ and $I_2 \subset W^u(p)$.

$W^s(p, \tilde{U})$ denote connected component of $W^s(p) \bigcap \tilde{U}$ containing $p$, and define $W^u(p, \tilde{U})$ similarly. We may choose $C^r$ coordinates $(x, y)$ so that in some small neighborhood $\tilde{U}$ of $p$, the point $p$ corresponds to $(0, 0)$, the set $W^u(p, \tilde{U})$ corresponds to $(y = 0)$, and the set $W^s(p, \tilde{U})$ corresponds to $(x = 0)$. We assume that $\tilde{U}$ is small enough that $f$ in $\tilde{U}$ is closely approximated by its derivative $Df_{(0,0)}$. Hence, $f$ nearly contracts vertical directions and expands horizontal directions in $\tilde{U}$.

Take compact arcs $I_1 \subset W^s(p)$ and $I_2 \subset W^u(p)$ both containing the points $p$ and $q$ as in Figure 5.

Let $D$ be a curvilinear rectangle which is a slight thickening of $I_1$. The forward iterates $f^i(D)$ will stay near $I_1$ for a while and then start to approach $I_2$. If we choose $D$ appropriately, we can arrange for some high iterate $f^N(D)$ to be a slight thickening of $I_2$ as in Figure 6. This looks geometrically like the horseshoe map. Let $A_1$ be the connected component of the intersection $D \bigcap f^N(D)$ containing $p$, and let $A_2$ be the connected component of the intersection $D \bigcap f^N(D)$ containing $q$. These sets (which are shaded in **Figure 6**) play the role of the rectangles $R_1$ and $R_2$, respectively, in the horseshoe construction. We use the set $A_1 \bigcup A_2$ for $U$ in Theorem 1.

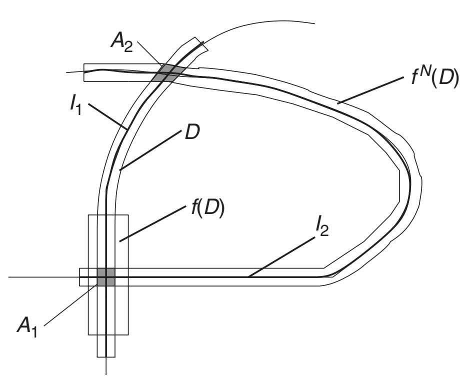



**Figure 6** The curvilinear rectangle $D$ and its $N$th iterate $f^N(D)$ are geometrically like the horseshoe map.

## The Hénon Family

To give explicit formulas for the horseshoe map above is somewhat tedious, and it is of interest to note that similar properties occur in maps with simple formulas. Indeed, such properties occur quite often in a well-known family of maps known as the "Hénon family." As we have mentioned, the map in Figure 1 provides an example.

One may simply define a Hénon map as a diffeomorphism $H = (H_1(x, y), H_2(x, y))$ with inverse $G(x, y) = (G_1(x, y), G_2(x, y))$ such that all the maps $F_i(x, y), G_i(x, y)$ are polynomials of degree at most two. It is known (see, e.g., Friedland and Milnor (1989)) that such maps $H$ have constant Jacobian determinant, and, up to affine conjugacy, may be represented in the form $H = H_{a,b}(x, y) = (a - x^2 - by, x)$ with $a$, $b$ constants and $b \neq 0$. This makes sense when all the terms are real or complex. In the real case, we speak of the real Hénon family and, in the complex case, we speak of the complex Hénon family.

The real Hénon family was first presented by the physicist M Hénon in 1976 as perhaps the simplest nonlinear diffeomorphism of the plane exhibiting a so-called "strange attractor." These mappings in the real and complex cases have been the focus of much attention. Our interest here is that, at least for certain parameters $a$, $b$, they provide concrete globally defined maps whose dynamics are analogous to that of the horseshoe diffeomorphism. In fact, Devaney and Nitecki (1979) proved (in the real case) that for fixed $b \neq 0$, there is a constant $a_0 > 0$ such that if $a > a_0$, then the set $B_{a,b}$ of bounded orbits of $H_{a,b}$ is a compact zero-dimensional set and the pair $(H_{a,b}, B_{a,b})$ is topologically conjugate to $(\sigma, \Sigma_2)$. In addition, it can be shown that the invariant set $B_{a,b}$ is a single hyperbolic $h$-closure. Analogous results are true for the complex Hénon family and proofs were originally given in the thesis of Ralph Oberste–Vorth (unpublished) under the supervision of John Hubbard at Cornell University. More recent proofs are in Newhouse (2004) and Hruska (2004). Many interesting results have been obtained for the complex Hénon map by Bedford and Smillie and Sibony and Fornaess (see the references in Hruska (2004).

## Homoclinic Points in Systems with Positive Topological Entropy

There is an invariant of topological conjugacy which is known as the topological entropy. In a certain sense, this gives a quantitative measurement of the amount of complicated or chaotic motion in the system.

Let $f: X \to X$ be a continuous self-map of the compact metric space $(X, d)$. For a positive integer $n > 0$, we define an $n$-orbit to be a finite sequence $O(x, n) = \{x, f(x), \ldots, f^{n-1}(x)\}$. Given a positive real number $\epsilon > 0$, we say that two $n$-orbits $O(x, n)$ and $O(y, n)$ are "$\epsilon$-distinguishable" if there is a $0 \leq j < n$ such that $d(f^j x, f^j y) > \epsilon$. Another way to look at this is the following. Define the so-called $d_n$-metric on $X$ by setting $d_n(x, y) = \max_{0 \leq j < n} d(f^j x, f^j y)$. Then, the two $n$-orbits $O(x, n), O(y, n)$ are $\epsilon$-distinguishable if and only if $d_n(x, y) > \epsilon$. It follows from compactness of $X$ and the uniform continuity of each of the maps $f^j, 0 \leq j < n$, that the number $r(n, \epsilon, f)$ of $\epsilon$-distinguishable $n$-orbits is finite for each given $\epsilon > 0$ and each positive integer $n$. We define the number

$$h(f) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log r(n, \epsilon, f)$$

This means that, for some sequence of integers $n_1 < n_2 < \ldots$, the map $f$ has roughly $e^{n_i h(f)}$ $\epsilon$-distinguishable $n_i$-orbits for $i$ large and $\epsilon$ small.

The number $h(f)$ is called the topological entropy of the map $f$. It may be infinite for homeomorphisms, but it is always finite for smooth maps on finite-dimensional manifolds. The number $h(f)$ has many nice properties. For instance, $h(f^N) = N h(f)$ for every positive integer $N$, and, if $f$ is a homeomorphism, then $h(f^{-1}) = h(f)$. Further, if $f$ and $g$ are topologically conjugate, then $h(f) = h(g)$. The so-called "variational principle for topological entropy" asserts that $h(f)$ is the supremum of the measure-theoretic entropies of the invariant probability measures for $f$. Our interest in this invariant here is the following theorem of Katok.

**Theorem 2**(Katok). *Let $f$ be a $C^2$ diffeomorphism of a compact two-dimensional manifold $M$ to itself with positive topological entropy. Then, $f$ has transverse homoclinic points.*

In fact, Katok extended this theorem (see the supplement in Hasselblatt and Katok (1995)) to show that, if $h(f) > 0$ and $\epsilon > 0$, then there is a compact zero-dimensional hyperbolic basic set $\Lambda$ for $h$ such that $h(f, \Lambda) > h(f) - \epsilon$. Thus, one can find nice invariant topologically transitive sets for $f$ (i.e., sets with dense orbits) on which the topological entropies of restriction of $f$ are arbitrarily close to that of $f$.

This theorem has the interesting consequence that the map $f \to h(f)$ is lower-semicontinuous on the space of $C^2$ diffeomorphisms of a surface. It was proved in Newhouse (1989) (and, independently by Yomdin (1987)) that the map $f \to h(f)$ is upper-semicontinuous on the space of $C^\infty$ diffeomorphisms of any compact manifold. Combining these results gives the theorem that the map $f \to h(f)$ is continuous on the space of $C^\infty$ diffeomorphisms on a compact surface, and that positivity of $h(f)$ implies the existence of transverse homoclinic points.

It is also worth noting that, for any continuous self-map $f: M \to M$ on a compact manifold $M$, one has the inequality $h(f) \geq \log |\mu|$ where $\mu$ is the eigenvalue of largest norm of the induced map $f_\star$ on the first real homology group (Manning 1975). Putting this together with Theorem 2 gives the fact that there are whole homotopy classes of diffeomorphisms on surfaces all of whose elements have transverse homoclinic points. For instance, consider a $2 \times 2$ matrix

$$L = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with integer entries, determinant 1, and eigenvalues $\lambda_1, \lambda_2$ with $0 < |\lambda_1| < 1 < |\lambda_2|$. Let $\tilde{L}: \boldsymbol{T}^2 \to \boldsymbol{T}^2$ be the induced diffeomorphism on the two-dimensional torus $\boldsymbol{T}^2$. This is an example of what is called an "Anosov" diffeomorphism. In this case the number $\mu$ above is simply $\lambda_2$, and this holds for any diffeomorphism $f$ of $\boldsymbol{T}^2$ which can be continuously deformed into $\tilde{L}$. Hence, any such $f$ must have transverse homoclinic points.

## Homoclinic Tangencies

Let $\{f_\lambda, \lambda \in [0,1]\}$ be a parametrized family of $C^r$ diffeomorphisms of the plane with $\lambda$ an external parameter. It frequently occurs that there is a hyperbolic saddle fixed point $p_\lambda$ for each parameter $\lambda$ moving continuously with $\lambda$ such that, at some value $\lambda_0$, a homoclinic tangency is created at a point $q_0$. This means that there are an $\epsilon > 0$, a small neighborhood $U$ of $q_0$, and curves $\gamma_\lambda^u \subset W^u(p_\lambda)$, $\gamma_\lambda^s \subset W^s(p_\lambda)$ such that $\gamma_\lambda^s \bigcap \gamma_\lambda^u = \emptyset$ for $\lambda_0 - \epsilon < \lambda < \lambda_0, \gamma_{\lambda_0}^s \bigcap \gamma_{\lambda_0}^u = \{q_0\}$, and $\gamma_\lambda^s \bigcap \gamma_\lambda^u$ consists of two distinct points for $\lambda_0 < \lambda < \lambda + \epsilon$. In most cases, the tangency of $\gamma_{\lambda_0}^u$ and $\gamma_{\lambda_0}^s$ at $q_0$ will be of the second order, and we will assume that occurs here. The geometry is as in **Figure 7**.

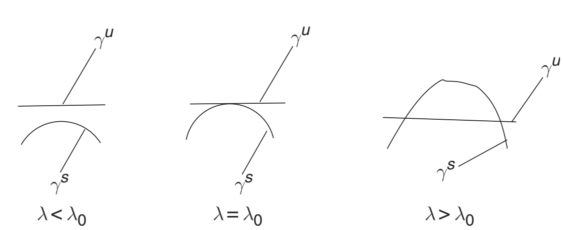



**Figure 7** Creation of a homoclinic tangency.

The creation of homoclinic tangencies is part of the general subject of "homoclinic bifurcations." A recent survey of this subject is in the book by Bonatti *et al.* (2005). Typical results are the following. If $p = p_{\lambda_0}$ is a saddle fixed point whose derivative is area-decreasing (i.e., $|\text{Det}(Df(p))| < 1$), then there are infinitely many parameters $\lambda$ near $\lambda_0$ for which each transverse homoclinic point of $p_\lambda$ is a limit of periodic sinks (asymptotically stable periodic orbits) (Newhouse 1979, Robinson 1983). In addition, so-called strange attractors and SRB measures appear (Mora and Viana 1993).

Finally, we mention that recently it has been shown that, generically in the $C^r$ topology for $r \geq 2$, homoclinic closures associated to a homoclinic tangency (in dimension 2) have maximal Hausdorff dimension (Theorem 1.6 in Downarowicz and Newhouse (2005)).

*See also:* Chaos and Attractors; Fractal Dimensions in Dynamics; Generic Properties of Dynamical Systems; Hyperbolic Dynamical Systems; Lyapunov Exponents and Strange Attractors; Saddle Point Problems; Singularity and Bifurcation Theory; Solitons and Other Extended Field Configurations.

## Further Reading

Birkhoff GD (1960) Nouvelles recherches sur les systèmes dynamiques. In: *George David Birkhoff, Collected Mathematical Papers, Vol. II*, pp. 620–631. New York: American Mathematical Society.

Bonatti C, Diaz L, and Viana M (2005) *Dynamics Beyond Uniform Hyperbolicity*, Encyclopedia of Mathematical Sciences, Subseries: Mathematical Physics III, vol. 102, Berlin–Heidelberg–New York: Springer.

Devaney R and Nitecki Z (1979) Shift automorphism in the henon mapping. *Communications in Mathematical Physics* 67: 137–148.

Downarowicz T and Newhouse S (2005) Symbolic extensions in smooth dynamical systems. *Inventiones Mathematicae* 160(3): 453–499.

Friedland S and Milnor J (1989) Dynamical properties of plane polynomial automorphisms. *Ergodic Theory and Dynamical Systems*. 9: 67–99.

Guckenheimer J and Holmes P (1983) *Nonlinear Oscillations, Dynamical Systems, and Bifurcation of Vector Fields*, Applied Mathematical Sciences, vol. 42, New York: Springer.

Hasselblatt B and Katok A (1995) *Introduction to the Modern Theory of Dynamical Systems*, Encyclopedia of Mathematics and Its Applications, vol. 54, Cambridge: Cambridge University Press.

Hruska SL (2004) A numerical method for proving hyperbolicity in complex hénon mappings (*http://arxiv.org*).

Manning A (1975) Topological entropy and the first homology group. In: Manning A (ed.) *Dynamical Systems – Warwick 1974*, Lecture Notes in Math. vol. 468, pp. 567–573. New York: Springer.

Moser J (1973) *Stable and Random Motions in Dynamical Systems*, Annals of Mathematical Studies, Number 77. Princeton: Princeton University Press.

Mora L and Viana M (1993) Abundance of strange attractors. *Acta Mathematica* 171: 1–71.

Newhouse S (1979) The abundance of wild hyperbolic sets and non-smooth stable sets for diffeomorphisms. *Publications Mathématiques de l'I.H.E.S.* 50: 101–151.

Newhouse S (1980) Lectures on dynamical systems. In: Coates J and Helgason S (eds.) *Dynamical Systems, CIME Lectures, Bressanone, Italy, June 1978*, Progress in Mathematics, vol. 8, pp. 1–114. Cambridge, MA: Birkhäuser.

Newhouse S (1989) Continuity properties of entropy. *Annals of Mathematics* 129: 215–235.

Newhouse S (2004) Cone-fields, domination, and hyperbolicity. In: Brin M, Hasselblatt B, and Pesin Y (eds.) *Modern Dynamical Systems and Applications*, pp. 419–432. Cambridge: Cambridge University Press.

Nusse HE and Yorke JA (1998) *Dynamics: Numerical Explorations*. Applied Mathematical Sciences, vol. 101, New York: Springer.

Poincaré H (1899) *Les methodes nouvelles de la Mecanique Celeste*–Tome 3. Guathier-Villars, Original publication (New Printing: Librairie Scientifique et Technique Albert Blanchard 9, Rue Medecin 75006, Paris, 1987).

Robinson C (1983) Homoclinic bifurcation to infinitely many sinks. *Communications in Mathematical Physics* 90: 433–459.

Robinson C (1999) *Dynamical Systems, Stability, Symbolic Dynamics, and Chaos*, Studies in Advanced Mathematics, 2nd edn. New York: CRC Press.

Yomdin Y (1987) Volume growth and entropy. *Israel Journal of Mathematics* 57: 285–300.

# Hopf Algebra Structure of Renormalizable Quantum Field Theory

**D Kreimer**, IHES, Bures-sur-Yvette, France



## Overview

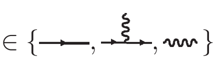
Renormalization theory is a venerable subject put to daily use in many branches of physics. Here, we focus on its applications in quantum field theory, where a standard perturbative approach is provided through an expansion in Feynman diagrams. Whilst the combinatorics of the Bogoliubov recursion, solved by suitable forest formulas, has been known for a long time, the subject regained interest on the conceptual side with the discovery of an underlying Hopf algebra structure behind these recursions.

Perturbative expansions in quantum field theory are organized in terms of one-particle irreducible (1PI) Feynman graphs. The goal is to calculate the corresponding 1PI Green functions order by order in the coupling constants of the theory, by applying Feynman rules to these 1PI graphs of a

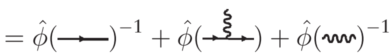
renormalizable theory under consideration. This allows one to disentangle the problem into an algebraic part and an analytic part.

For the algebraic part, one studies Feynman graphs as combinatorial objects which lead to the Lie and Hopf algebras discussed below. Feynman rules then assign analytic expressions to these graphs, with the analytic structure of finite renormalized quantum field theory largely dictated by the underlying algebra.

The objects of interest in quantum field theory are the 1PI Green functions. They are parametrized by the quantum numbers – masses, momenta, spin, and such – of the particles participating in the scattering process under consideration. We call a set of such quantum numbers an external leg structure $\underline{r}$. For example, the three terms in the Lagrangian of massless quantum electrodynamics correspond to

$\underline{r} \in$  [1]

Note that the Lagrangian $L$ of massless quantum electrodynamics is obtained accordingly as

$L =$ 

$$= \bar{\psi}\partial\psi + \bar{\psi}A\psi + \tfrac{1}{4}F^2 \qquad [2]$$

where $\hat{\phi}$ are coordinate space Feynman rules.

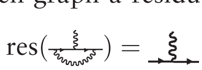
The renormalized 1PI Green function in momentum space, $G_{\mathrm{R}}^{\underline{r}}(\{g\};\{p\},\{m\};\mu)$, is obtained as the image under renormalized Feynman rules $\phi_{\mathrm{R}}$ applied to a series of graphs:

$$\Gamma^{\underline{r}} = 1 + \sum_{k=1}^{\infty} g^k c_k^{\underline{r}} \equiv 1 + \sum_{\mathrm{res}(\Gamma)=\underline{r}} g^{|\Gamma|}\frac{\Gamma}{\mathrm{Sym}(\Gamma)} \qquad [3]$$

Here $\underline{r}$ is a given such external leg structure, while $c_k^{\underline{r}}$ is the finite sum of 1PI graphs having $k$ loops,

$$c_k^{\underline{r}} = \sum_{\substack{\mathrm{res}(\Gamma)=\underline{r}\\ |\Gamma|=k}} \frac{\Gamma}{\mathrm{Sym}(\Gamma)} \qquad [4]$$

and $0 < g < 1$ is a coupling constant. The generalization to the case of several couplings $\{g\}$ and masses $\{m\}$ is straightforward. In the above, the sum is over all 1PI graphs with the same given external leg structure. We have denoted the map which assigns $\underline{r}$ to a given graph a residue, for example,

 [5]

The unrenormalized but regularized Feynman rules $\phi$ assign to a graph a function

$$\phi(\Gamma)(\{g\};\{p\},\{m\};\mu,z) = \int \prod_{v\in\Gamma^{[0]}} \delta^{(4)}\left(\sum_{f \text{ incident } v} k_f\right) \prod_{e\in\Gamma^{[1]}_{\mathrm{int}}} \mathrm{Prop}(k_e)\frac{\mathrm{d}^4 k_e}{4\pi^2} \qquad [6]$$

and formally the unrenormalized Green function

$$G_{\mathrm{u}}^{\underline{r}}(\{g\};\{p\},\{m\};\mu,z) = \phi(\Gamma^{\underline{r}})(\{g\};\{p\},\{m\};\mu,z) \qquad [7]$$

which is a function of a suitably chosen regulator $z$. Note that in [6] the four-dimensional Dirac-$\delta$ distribution guarantees momentum conservation at each vertex and restricts the number of four-dimensional integrations to the number of independent cycles in the graph. It is assumed that the reader is familiar with the readily established fact that these integrals suffer from UV singularities, which render the integration over the momenta in internal cycles ill-defined. We also remind the reader that the problem persists in coordinate space, where one confronts the continuation of products of distributions to regions of coinciding support. We restrict ourselves here to a discussion of the situation in momentum space and refer the reader to the literature for the situation in coordinate space.

Ignoring problems of convergence in the sum over all graphs, the problem of renormalization is to make sense of these functions term by term: We have to determine invertible series $Z^{\underline{r}}(\{g\},z)$ in the couplings $g$ such that the modified Lagrangian

$$\tilde{L} = \sum_{\underline{r}} Z^{\underline{r}}(\{g\},z)\,\hat{\phi}(\underline{r}) \qquad [8]$$

produces a perturbation series in graphs that allows for the removal of the regulator $z$.

This amounts to a transition from unrenormalized to renormalized Feynman rules $\phi \to \phi_{\mathrm{R}}$. Let us first describe how this transition is achieved using the Lie and Hopf algebra structure of the perturbative expansion, which is described in detail below:

- Decide on the free fields and local interactions of the theory, appropriately specifying quantum numbers (spin, mass, flavor, color, and such) of fields, restricting interactions so as to obtain a renormalizable theory.
- Consider the set of all 1PI graphs with edges corresponding to free-field propagators. Define vertices for local interactions. This allows one to construct a pre-Lie algebra of graph insertions. Antisymmetrize this pre-Lie product to get a Lie algebra $\mathcal{L}$ of graph insertions and define the Hopf algebra $\mathcal{H}$ which is dual to the enveloping algebra $\mathcal{U}(\mathcal{L})$ of this Lie algebra.
- Realize that the coproduct and antipode of this Hopf algebra give rise to the forest formula, which generates local counter-terms upon introducing a Rota–Baxter map, a renormalization scheme in physicists' parlance.

- Use the Hochschild cohomology of this Hopf algebra to show that one can absorb singularities in local counter-terms.
- Determine the corepresentations of this Hopf algebra to identify the sub-Hopf algebras corresponding to time-ordered products in physical fields. This is most easily achieved by rewriting the Dyson–Schwinger equations using Hochschild 1-cocycles.

The last point exhibits close connections, in particular, between the structure of gauge theories and the corepresentation theory of their perturbative Hopf algebras which we discuss below in brief.

This program can be carried out in coordinate space as well as momentum space renormalization. It has given a firm mathematical background to the process of renormalization, justifying the practice of quantum field theory. The notion of locality has achieved a precise formulation in terms of the Hochschild cohomology of the perturbation expansion. In momentum space, this approach emphasizes the connections to number theory, which emerge when one investigates the role of the Hopf algebra primitives, which in turn furnish the Hochschild 1-cocycles underlying locality.

The next sections describe the above setup in some detail.

## Lie and Hopf Algebras of Graphs

All algebras are supposed to be over some field $\mathbb{K}$ of characteristic zero, associative and unital, and similarly for coalgebras. The unit (and, by abuse of notation, also the unit map) will be denoted by $\mathbb{I}$, the counit map by $\bar{e}$. All algebra homomorphisms are supposed to be unital. A bialgebra $(A = \bigoplus_{i=0}^{\infty} A_i, m, \mathbb{I}, \Delta, \bar{e})$ is called graded connected if $A_i A_j \subset A_{i+j}$ and $\Delta(A_i) \subset \bigoplus_{j+k=i} A_j \otimes A_k$, and if $\Delta(\mathbb{I}) = \mathbb{I} \otimes \mathbb{I}$ and $A_0 = k\mathbb{I}, \bar{e}(\mathbb{I}) = 1 \in \mathbb{K}$ and $\bar{e} = 0$ on $\bigoplus_{i=1}^{\infty} A_i$. We call $\ker \bar{e}$ the augmentation ideal of $A$ and denote by $P$ the projection $A \to \ker \bar{e}$ onto the augmentation ideal, $P = \mathrm{id} - \mathbb{I}\bar{e}$. Furthermore, we use Sweedler's notation, $\Delta(h) = \sum h' \otimes h''$, for the coproduct. We define

$$\mathrm{Aug}^{(k)} = \Bigg( \underbrace{P \otimes \cdots \otimes P}_{k \text{ times}} \Bigg) \Delta^{k-1}, \qquad A \to \{\ker \bar{e}\}^{\otimes k} \tag{9}$$

as a map into the $k$-fold tensor product of the augmentation ideal. We let $A^{(k)} = \ker \mathrm{Aug}^{(k+1)} / \ker \mathrm{Aug}^{(k)}, \forall\, k \geq 1$. All bialgebras considered here are bigraded in the sense that

$$A = \bigoplus_{i=0}^{\infty} A_{(i)} = \bigoplus_{k=0}^{\infty} A^{(k)} \tag{10}$$

where $A_{(k)} \subset \oplus_{j=1}^{k} A^{(j)}$ for all $k \geq 1$. $A_{(0)} \simeq A^{(0)} \simeq \mathbb{K}$.

The first construction we have to study is the pre-Lie algebra structure of 1PI graphs.

### The Pre-Lie Structure

For each Feynman graph we have vertices as well as internal and external edges. External edges are edges that have an open end not connected to a vertex. They indicate the particles participating in the scattering amplitude under consideration and each such edge carries the quantum numbers of the corresponding free field. The internal edges and vertices form a graph in their own right. For an internal edge, both ends of the edge are connected to a vertex.

We consider 1PI Feynman graphs. A graph $\Gamma$ is 1PI if and only if all graphs, obtained by removal of any one of its internal edges, are still connected. Such 1PI graphs are naturally graded by their number of independent loops, the rank of their first homology group $H_{[1]}(\Gamma, \mathbb{Z})$. We write $|\Gamma|$ for this degree of a graph $\Gamma$. Note that $|\mathrm{res}(\Gamma)| = 0$, where we let $\mathrm{res}(\Gamma)$ be the graph obtained when all edges in $\Gamma^{[1]}_{\mathrm{int}}$ shrink to a point, as before. Note that the graph obtained in this manner consists of a single vertex, to which the edges $\Gamma^{[1]}_{\mathrm{ext}}$ are attached.

For a 1PI graph $\Gamma, \Gamma^{[0]}$ denotes its set of vertices and $\Gamma^{[1]} = \Gamma^{[1]}_{\mathrm{int}} \cup \Gamma^{[1]}_{\mathrm{ext}}$ its set of internal and external edges. In addition, let $\omega_{\underline{r}}$ be the number of spacetime derivatives appearing in the corresponding monomial in the Lagrangian.

Having specified free quantum fields and local interaction terms between them, one immediately obtains the set of 1PI graphs. One can then consider for a given external leg structure $\underline{r}$ the set of graphs with that external leg structure. For a renormalizable theory, we can define a superficial degree of divergence,

$$\omega = \sum_{\underline{r} \in \Gamma^{[1]}_{\mathrm{int}} \cup \Gamma^{[0]}} \omega_{\underline{r}} - 4|H_{[1]}(\Gamma, \mathbb{Z})| \tag{11}$$

for each such external leg structure: $\omega(\Gamma) = \omega(\Gamma')$ if $\mathrm{res}(\Gamma) = \mathrm{res}(\Gamma')$; all graphs with the same external leg structure have the same superficial degree of divergence, and only for a finite number of distinct external leg structures $\underline{r}$ will this degree indeed signify a divergence.

This leaves a finite number of external leg structures to be considered to which we restrict ourselves from now. Our first observation is that there is a natural pre-Lie algebra structure on 1PI graphs.

To this end, we define a bilinear operation

$$\Gamma_1 * \Gamma_2 = \sum_{\Gamma} n(\Gamma_1, \Gamma_2; \Gamma)\Gamma \qquad [12]$$

where the sum is over all 1PI graphs $\Gamma$. Here, $n(\Gamma_1, \Gamma_2; \Gamma)$ is a section coefficient which counts the number of ways in which a subgraph $\Gamma_2$ can be reduced to a point in $\Gamma$ such that $\Gamma_1$ is obtained. The above sum is evidently finite as long as $\Gamma_1$ and $\Gamma_2$ are finite graphs, and the graphs which contribute necessarily fulfill $|\Gamma| = |\Gamma_1| + |\Gamma_2|$ and $\mathrm{res}(\Gamma) = \mathrm{res}(\Gamma_1)$.

One then has the following theorem.

**Theorem 1** *The operation* $*$ *is pre-Lie*:

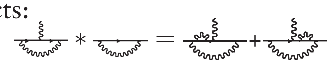
$$[\Gamma_1 * \Gamma_2] * \Gamma_3 - \Gamma_1 * [\Gamma_2 * \Gamma_3] = [\Gamma_1 * \Gamma_3] * \Gamma_2 - \Gamma_1 * [\Gamma_3 * \Gamma_2] \qquad [13]$$

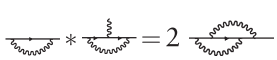
which is evident when one rewrites the $*$-product in suitable gluing operations.

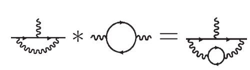
To understand this theorem, note that the equation claims that the lack of associativity in the bilinear operation $*$ is invariant under permutation of the elements indexed 2, 3. This suffices to show that the antisymmetrization of this map fulfills a Jacobi identity. Hence, we get a Lie algebra $\mathcal{L}$ by antisymmetrizing this operation:

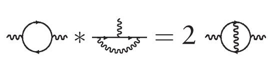
$$[\Gamma_1, \Gamma_2] = \Gamma_1 * \Gamma_2 - \Gamma_2 * \Gamma_1 \qquad [14]$$

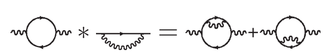
This Lie algebra is graded and of finite dimension in each degree. Let us look at a couple of examples for pre-Lie products:

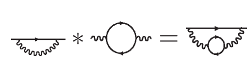
 [15]

 [16]

 [17]

 [18]

 [19]

 [20]

Together with $\mathcal{L}$ one is led to consider the dual of its universal enveloping algebra $\mathcal{U}(\mathcal{L})$ using the theorem of Milnor and Moore. For this we use the above grading by the loop number.

This universal enveloping algebra $\mathcal{U}(\mathcal{L})$ is built from the tensor algebra

$$T = \bigoplus_k T^k, \quad T^k = \underbrace{\mathcal{L} \otimes \cdots \otimes \mathcal{L}}_{k \text{ times}} \qquad [21]$$

by dividing out the ideal generated by the relations

$$a \otimes b - b \otimes a = [a, b] \in \mathcal{L} \qquad [22]$$

Note that in $\mathcal{U}(\mathcal{L})$ we have a natural concatenation product $m_*$. Furthermore, $\mathcal{U}(\mathcal{L})$ carries a natural Hopf algebra structure with this product. For that, the Lie algebra $\mathcal{L}$ furnishes the primitive elements:

$$\Delta_*(a) = a \otimes 1 + 1 \otimes a, \quad \forall\, a \in \mathcal{L} \qquad [23]$$

It is, by construction, a connected finitely graded Hopf algebra which is co-commutative but not commutative. We can then consider its graded dual, which will be a Hopf algebra $\mathcal{H}(m, \mathbb{I}, \Delta, \bar{e})$ that is commutative but not cocommutative. One finds it upon using a Kronecker pairing

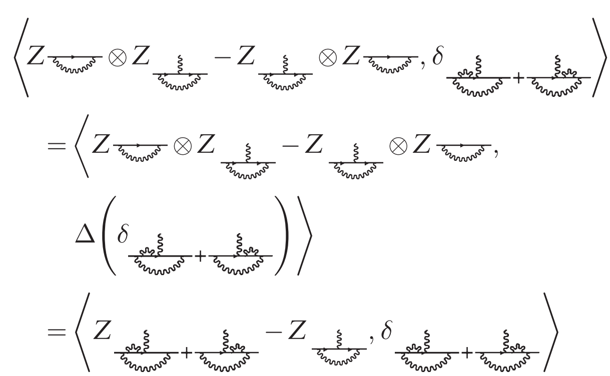
$$< Z_\Gamma, \delta_{\Gamma'} > = \begin{cases} 1, & \Gamma = \Gamma' \\ 0, & \text{else} \end{cases} \qquad [24]$$

The space of primitives of $\mathcal{U}(\mathcal{L})$ is in one-to-one correspondence with the set $\mathrm{Indec}(\mathcal{H})$ of indecomposables of $\mathcal{H}$, which is the linear span of its generators. One finds the following theorem.

**Theorem 2**

$$< Z_{\Gamma_1} \otimes Z_{\Gamma_2} - Z_{\Gamma_2} \otimes Z_{\Gamma_1},\ \delta_\Gamma > = < Z_{[\Gamma_2, \Gamma_1]}, \delta_\Gamma > \qquad [25]$$

For example, one finds



$= 2$ [26]

$\mathcal{H}$ is a graded commutative Hopf algebra which suffices to describe renormalization theory, as we see in the next section. We have formulated it for the superficially divergent 1PI graphs of the theory with the understanding that the residues of these graphs are in one-to-one correspondence with the terms in the Lagrangian of a given theory. Often, several terms in a Lagrangian correspond to graphs with the same number and type of external legs, but correspond to different form-factor projections of the graph. In such cases, the above approach can be easily adopted considering suitably colored or

labeled graphs. A similar remark applies if one desires to incorporate renormalization of superficially convergent Green functions, which requires nothing more than the consideration of an easily obtained semidirect product of the Lie algebra of superficially divergent graphs with the abelian Lie algebra of superficially convergent graphs.

#### The Principle of Multiplicative Subtraction

The above algebra structures are available once one has decided on the set of 1PI graphs of interest. We now use them toward the renormalization of any such chosen local quantum field theory.

From the above, 1PI graphs $\Gamma$ provide the linear generators $\delta_\Gamma$ of the Hopf algebra $\mathcal{H} = \oplus_{i=0}^{\infty} H_i$, where $\mathcal{H}_{\text{lin}} = \text{span}(\delta_\Gamma)$ and their disjoint union provides the commutative product.

Now let $\Gamma$ be a 1PI graph. We find the Hopf algebra $\mathcal{H}$ as described above to have a coproduct explicitly given as $\Delta : \mathcal{H} \to \mathcal{H} \otimes \mathcal{H}$:

$$\Delta(\Gamma) = \Gamma \otimes 1 + 1 \otimes \Gamma + \sum_{\gamma \subset \Gamma} \gamma \otimes \Gamma/\gamma \qquad [27]$$

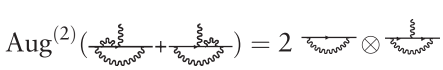
where the sum is over all unions of 1PI superficially divergent proper subgraphs, and we extend this definition to products of graphs so that we get a bialgebra.

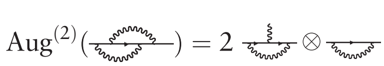
While the Lie bracket inserted graphs into each other, the coproduct disentangles them. It is this latter operation which is needed in renormalization theory: we have to render each subgraph finite before we can construct a local counter-term. That is precisely what the Hopf algebra structure maps do.

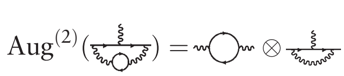
Having a coproduct, two further structure maps of $\mathcal{H}$ are immediate: the counit and the antipode. The counit $\bar{e}$ vanishes on any nontrivial Hopf algebra element, $\bar{e}(1) = 1, \bar{e}(X) = 0$. The antipode is

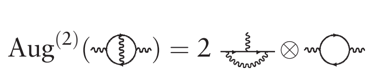
$$S(\Gamma) = -\Gamma - \sum_{\gamma \subset \Gamma} S(\gamma)\Gamma/\gamma \qquad [28]$$

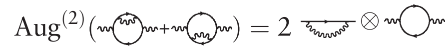
We can work out a few coproducts and antipodes as follows:

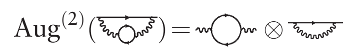

[29]


[30]

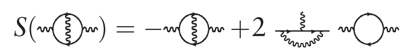

[31]


[32]


[33]


[34]

We give just one example for an antipode:


[35]

Note that for each term in the sum $\tilde{\Delta}(\Gamma) = \sum_i \Gamma'_{(i)} \otimes \Gamma''_{(i)}$, we have unique gluing data $G_i$ such that

$$\Gamma = \Gamma''_{(i)} \leftarrow_{G_i} \Gamma'_{(i)}, \quad \forall\, i \qquad [36]$$

These gluing data describe the necessary bijections to glue the components $\Gamma'_{(i)}$ back into $\Gamma''_{(i)}$ so as to obtain $\Gamma$: using them, we can reassemble the whole from its parts. Each possible gluing can be interpreted as a composition in the insertion operad of Feynman graphs.

We have by now obtained a Hopf algebra generated by combinatorial elements, 1PI Feynman graphs. Its existence is automatic from the above choices of interactions and free fields. What remains to be done is a structural analysis of these algebras for the renormalizable theories we are confronted with in four spacetime dimensions.

The assertion underlying perturbation theory is the fact that meaningful approximations to physical observable quantities can be found by evaluating these graphs using Feynman rules.

First, as disjoint scattering processes give rise to independent amplitudes, one is led to the study of characters of the Hopf algebra, maps $\phi : \mathcal{H} \to V$ such that $\phi \circ m = m_V(\phi \otimes \phi)$.

Such maps assign to any element in the Hopf algebra an element in a suitable target space $V$. The study of tree-level amplitudes in lowest-order perturbation theory justifies assigning to each edge a propagator and to each elementary scattering process a vertex, which define the Feynman rules $\phi(\text{res}(\Gamma))$ and the underlying Lagrangian, on the level of residues of these very graphs. Graphs are constructed from edges and vertices which are provided precisely by the residues of those divergent graphs, hence one is led to assign to each Feynman graph an evaluation in terms of an integral over the continuous quantum numbers assigned to edges or vertices, which leads to the familiar integrals over momenta in closed loops mentioned before.

Then, with the Feynman rules providing a canonical character $\phi$, we will have to make one further choice: a renormalization scheme. The need for such a choice is no surprise: after all we are eliminating short-distance singularities in the graphs,

which renders their remaining finite part ambiguous, albeit in a most interesting manner.

Hence, we choose a map $R : V \to V$, from which we obviously demand that it does not modify the UV-singular structure, and furthermore that it obeys

$$R(xy) + R(x)R(y) = R(R(x)y) + R(xR(y)) \quad [37]$$

which guarantees the multiplicativity of renormalization and is at the heart of the Birkhoff decomposition, which emerges below: it tells us that elements in $V$ split into two parallel subalgebras given by the image and kernel of $R$. Algebras for which such a map exists are known as Rota–Baxter algebras. The role Rota–Baxter algebras play for associative algebras is similar to the role Yang–Baxter algebras play for Lie algebras. The structure of these algebras allows one to connect renormalization theory to integrable systems. In addition, most of the results obtained initially for a specific renormalization scheme, such as minimal subtraction, can also be obtained, in general, upon a structural analysis of the corresponding Rota–Baxter algebras.

To see how all the above comes together in renormalization theory, we define a further character $S_R^\phi$ that deforms $\phi \circ S$ slightly and delivers the counter-term for $\Gamma$ in the renormalization scheme $R$:

$$\begin{aligned} S_R^\phi(\Gamma) &= -Rm_V(S_R^\phi \otimes \phi \circ P)\Delta \\ &= -R[\phi(\Gamma)] - R\left[\sum_{\gamma\subset\Gamma} S_R^\phi(\gamma)\phi(\Gamma/\gamma)\right] \end{aligned} \quad [38]$$

which should be compared with the undeformed

$$\begin{aligned} \phi \circ S &= m_V(S \circ \phi \otimes \phi \circ P)\Delta \\ &= -\phi(\Gamma) - \sum_{\gamma\subset\Gamma} \phi \circ S(\gamma)\phi(\Gamma/\gamma) \end{aligned} \quad [39]$$

The fact that $R$ is a Rota–Baxter map ensures that $S_R^\phi$ is an element of the character group $\mathcal{G}$ of the Hopf algebra, $S_R^\phi \in \mathrm{Spec}(\mathcal{G})$. Note that we have now determined the modified Lagrangian:

$$Z^{\underline{r}} = S_R^\phi(\Gamma^{\underline{r}}) \quad [40]$$

The classical results of renormalization theory follow immediately using this group structure: we obtain the renormalization of $\Gamma$ by the application of a renormalized character

$$S_R^\phi \star \phi(\Gamma) = m_V(S_R^\phi \otimes \phi)\Delta \quad [41]$$

and Bogoliubov's $\bar{R}$ operation as

$$\begin{aligned} \bar{R}(\Gamma) &= m_V(S_R^\phi \otimes \phi)(\mathrm{id} \otimes P)\Delta(\Gamma) \\ &= \phi(\Gamma) + \sum_{\gamma\subset\Gamma} S_R^\phi(\gamma)\phi(\Gamma/\gamma) \end{aligned} \quad [42]$$

so that

$$S_R^\phi \star \phi(\Gamma) = \bar{R}(\Gamma) + S_R^\phi(\Gamma) \quad [43]$$

Here, $S_R^\phi \star \phi$ is an element in the group of characters of the Hopf algebra, with the group law given by the convolution

$$\phi_1 \star \phi_2 = m_V \circ (\phi_1 \otimes \phi_2) \circ \Delta \quad [44]$$

so that the coproduct, counit, and coinverse (the antipode) give the product, unit, and inverse of this group, as befits a Hopf algebra. This Lie group has the previous Lie algebra $\mathcal{L}$ of graph insertions as its Lie algebra: $\mathcal{L}$ exponentiates to $\mathcal{G}$.

What we have achieved above is a local renormalization of quantum field theory. Let $M^{\underline{r}}$ be a monomial in the Lagrangian $L$ of degree $\omega_{\underline{r}}$:

$$M^{\underline{r}} = D_{\underline{r}}\{\phi\} \quad [45]$$

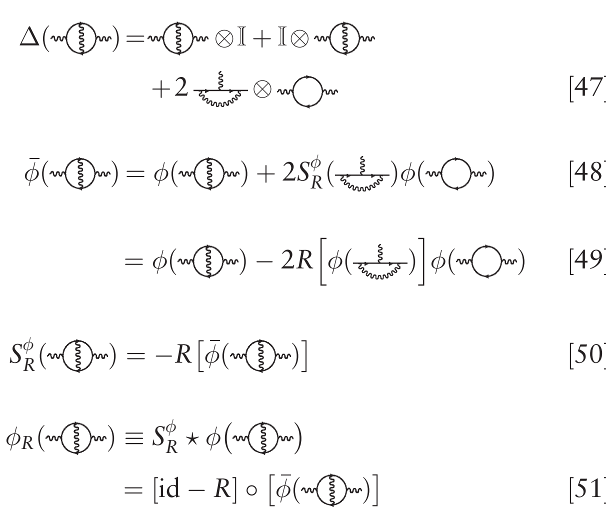
Then one can prove, using the Hochschild cohomology of $\mathcal{H}$:

**Theorem 3** (*Locality*)

$$Z^{\underline{r}} D_{\underline{r}}\{\phi\} = D_{\underline{r}} Z^{\underline{r}}\{\phi\} \quad [46]$$

*that is, renormalization commutes with infinitesimal spacetime variations of the fields.*

We can now work out the renormalization of a Feynman graph $\Gamma$:



The formulas [47]–[51] are given in their recursive form. Zimmermann's original forest formula solving this recursion is obtained when we trace our considerations back to the fact that the coproduct can be written in nonrecursive form as a sum over forests, and similarly for the antipode.

## Diffeomorphisms of Physical Parameters

In the above, we have effectively obtained a Birkhoff decomposition of the Feynman rules $\phi \in \mathrm{Spec}(\mathcal{G})$ into two characters – $\phi_+^R = S_R^\phi \star \phi \in \mathrm{Spec}(\mathcal{G})$ and $\phi_-^R = S_R^\phi \in \mathrm{Spec}(\mathcal{G})$ – for any Rota–Baxter map $R$. Thanks to Atkinson's theorem, this is possible for any renormalization scheme $R$. For the minimal subtraction scheme, it amounts to the decomposition of the Laurent series $\phi(\Gamma)(\epsilon)$, which has poles of finite order in the regulator $\epsilon$, into a part holomorphic at the origin and a part holomorphic at complex infinity. This has a particularly nice geometric interpretation upon considering the Birkhoff decomposition of a loop around the origin, providing the clutching data for the two half-spheres defined by that very loop.

Whilst in this manner a satisfying understanding of perturbative renormalization is obtained, the character group $\mathcal{G}$ remains rather poorly understood. On the other hand, renormalization can be captured by the study of diffeomorphisms of physical parameters as, by definition, the range of allowed modification in renormalization theory is determined by the variation of the coefficients of monomials $\hat{\phi}(\underline{r})$ of the underlying Lagrangian

$$L = \sum_{\underline{r}} Z^{\underline{r}} \hat{\phi}(\underline{r}) \qquad [52]$$

Thus, one desires to obtain the whole Birkhoff decomposition at the level of diffeomorphisms of the coupling constants.

The crucial step toward that goal is to realize the role of a standard quantum field-theoretic formula of the form

$$g_{\mathrm{new}} = g_{\mathrm{old}} Z^g \qquad [53]$$

where

$$Z^g = \frac{Z^v}{\prod_{e \in \mathrm{res}(v)^{[1]}_{\mathrm{ext}}} \sqrt{Z^e}} \qquad [54]$$

for some vertex $v$, which obtains the new coupling in terms of a diffeomorphism of the old. This formula provides, indeed, a Hopf algebra homomorphism from the Hopf algebra of diffeomorphisms to the Hopf algebra of Feynman graphs, regarding $Z^g$ (a series over counter-terms for all 1PI graphs with the external leg structure corresponding to the coupling $g$), in two different ways: it is, at the same time, a formal diffeomorphism in the coupling constant $g_{\mathrm{old}}$ and a formal series in Feynman graphs. As a consequence, there are two competing coproducts acting on $Z_g$. That both give the same result defines the required homomorphism, which transposes to a homomorphism from the largely unknown group of characters of $\mathcal{H}$ to the one-dimensional diffeomorphisms of this coupling.

In summary, one finds that a couple of basic facts enable one to make a transition from the abstract group of characters of a Hopf algebra of Feynman graphs (which, incidentally, equals the Lie group assigned to the Lie algebra with universal enveloping algebra the dual of this Hopf algebra) to the rather concrete group of diffeomorphisms of physical observables. These steps are given as follows:

- Recognize that $Z$ factors are given as counterterms over a formal series of graphs starting with 1, graded by powers of the coupling, hence invertible.
- Recognize the series $Z_g$ as a formal diffeomorphism, with Hopf algebra coefficients.
- Establish that the two competing Hopf algebra structures of diffeomorphisms and graphs are consistent in the sense of a Hopf algebra homomorphism.
- Show that this homomorphism transposes to a Lie algebra and hence Lie group homomorphism.

The effective coupling $g_{\mathrm{eff}}(\varepsilon)$ now allows for a Birkhoff decomposition in the space of formal diffeomorphisms.

**Theorem 4** *Let the unrenormalized effective coupling constant $g_{\mathrm{eff}}(\varepsilon)$ viewed as a formal power series in $g$ be considered as a loop of formal diffeomorphisms and let $g_{\mathrm{eff}}(\varepsilon) = (g_{\mathrm{eff}_-})^{-1}(\varepsilon)\, g_{\mathrm{eff}_+}(\varepsilon)$ be its Birkhoff decomposition in the group of formal diffeomorphisms. Then the loop $g_{\mathrm{eff}_-}(\varepsilon)$ is the bare coupling constant and $g_{\mathrm{eff}_+}(0)$ is the renormalized effective coupling.*

The above results hold as they stand for any massless theory which provides a single coupling constant. If there are multiple interaction terms in the Lagrangian, one finds similar results relating the group of characters of the corresponding Hopf algebra to the group of formal diffeomorphisms in the multidimensional space of coupling constants.

## The Role of Hochschild Cohomology

The Hochschild cohomology of the combinatorial Hopf algebras which we discuss here plays three major roles in quantum field theory:

1. it allows one to prove locality from the accompanying filtration by the augmentation degree coming from the kernels $\ker \mathrm{Aug}^{(k)}$;

2. it allows one to write the quantum equations of motion in terms of the Hopf algebra primitives, elements in $\mathcal{H}_{\text{lin}} \cap \{\ker \text{Aug}^{(2)}/\ker \text{Aug}^{(1)}\}$; and
3. it identifies the relevant sub-Hopf algebras formed by time-ordered products.

Before we discuss these properties, let us first introduce the relevant Hochschild cohomology.

### Hochschild Cohomology of Bialgebras

Let $(A, m, \mathbb{I}, \Delta, \epsilon)$ be a bialgebra, as before. We regard linear maps $L : A \to A^{\otimes n}$ as $n$-cochains and define a coboundary map $b, b^2 = 0$ by

$$bL := (\text{id} \otimes L) \circ \Delta + \sum_{i=1}^{n} (-1)^i \Delta_i \circ L + (-1)^{n+1} L \otimes \mathbb{I} \qquad [55]$$

where $\Delta_i$ denotes the coproduct applied to the $i$th factor in $A^{\otimes n}$, which defines the Hochschild cohomology of $A$.

For the case $n = 1$, for $L : A \to A$, [55] reduces to

$$bL = (\text{id} \otimes L) \circ \Delta - \Delta \circ L + L \otimes \mathbb{I} \qquad [56]$$

The category of objects $(A, C)$, which consists of a commutative bialgebra $A$ and a Hochschild 1-cocycle $C$ on $A$, has an initial object $(\mathcal{H}_{\text{rt}}, B_+)$, where $\mathcal{H}_{\text{rt}}$ is the Hopf algebra of (nonplanar) rooted trees, and the closed but nonexact 1-cocycle $B_+$ grafts a product of rooted trees together at a new root as described below.

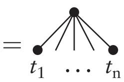
The higher ($n > 1$) Hochschild cohomology of $\mathcal{H}_{\text{rt}}$ vanishes, but in what follows, the closedness of $B_+$ will turn out to be crucial.

### The Hopf Algebra of Rooted Trees

A rooted tree is a simply connected contractible compact graph with a distinguished vertex, the root. A forest is a disjoint union of rooted trees. Isomorphisms of rooted trees or forests are isomorphisms of graphs preserving the distinguished vertex/vertices. Let $t$ be a rooted tree with root $o$. The choice of $o$ determines an orientation of the edges of $t$, away from the root, say. Forests are graded by the number of vertices they contain.

Let $\mathcal{H}_{\text{rt}}$ be the free commutative algebra generated by rooted trees. The commutative product in $\mathcal{H}_{\text{rt}}$ corresponds to the disjoint union of trees, such that monomials in $\mathcal{H}_{\text{rt}}$ are scalar multiples of forests. We demand that the linear operator $B_+$ on $\mathcal{H}_{\text{rt}}$, defined by

$$B_+(\mathbb{I}) = \bullet \qquad [57]$$

$B_+(t_1 \dots t_n) =$  [58]

is a Hochschild 1-cocycle, which makes $\mathcal{H}_{\text{rt}}$ a Hopf algebra. The resulting coproduct can be described as follows:

$$\Delta(t) = \mathbb{I} \otimes t + t \otimes \mathbb{I} + \sum_{adm\ c} P_c(t) \otimes R_c(t) \qquad [59]$$

where the sum goes over all admissible cuts of the tree $t$. Such a cut of $t$ is a nonempty set of edges of $t$ that are to be removed. The forest which is disconnected from the root upon removal of those edges is denoted by $P_c(t)$ and the part which remains connected to the root is denoted by $R_c(t)$. A cut $c(t)$ is admissible if, for each vertex $l$ of $t$, it contains at most one edge on the path from $l$ to the root.

This Hopf algebra of nonplanar rooted trees is the universal object after which all such commutative Hopf algebras $\mathcal{H}$ providing pairs $(\mathcal{H}, B)$, for $B$ a Hochschild 1-cocycle, are formed.

**Theorem 5** *The pair $(\mathcal{H}_{\text{rt}}, B_+)$, unique up to isomorphism, is universal among all such pairs. In other words, for any pair $(\mathcal{H}, B)$ where $\mathcal{H}$ is a commutative Hopf algebra and B a closed nonexact 1-cocycle, there exists a unique Hopf algebra morphism $\mathcal{H}_{\text{rt}} \xrightarrow{\rho} \mathcal{H}$ such that $B \circ \rho = \rho \circ B_+$.*

This theorem suggests that we investigate the Hochschild cohomology of the Hopf algebras of 1PI Feynman graphs. It clarifies the structure of 1PI Green functions.

### The Roles of Hochschild Cohomology

The Hochschild cohomology of the Hopf algebras of 1PI graphs sheds light on the structure of 1PI Green function in at least four different ways:

- it gives a coherent proof of locality of counterterms – the very fact that

$$[Z^{\underline{r}}, D_{\underline{r}}] = 0 \qquad [60]$$

  means that the coefficients in the Lagrangian remain independent of momenta, and hence the Lagrangian remains a polynomial expression in fields and their derivatives;
- the quantum equation of motions takes a very succinct form, identifying the Dyson kernels with the primitives of the Hopf algebra;
- sub-Hopf algebras emerge from the study of the Hochschild cohomology, which connects the representation theory of these Hopf algebras to the structure of theories with internal symmetries; and
- these Hopf algebras are intimately connected to the structure of transcendental functions, such as

the generalized polylogarithms, which play a prominent role these days ranging from applied particle physics to recent developments in mathematics.

To determine the Hochschild 1-cocycles of some Feynman graph Hopf algebra $\mathcal{H}$, one determines first the primitives graphs $\gamma$ of the Hopf algebra, which, by definition, fulfill the condition

$$\Delta(\gamma) = \gamma \otimes \mathbb{I} + \mathbb{I} \otimes \gamma \qquad [61]$$

Using the pre-Lie product above, one then determines the maps

$$B_+^\gamma : \mathcal{H} \to \mathcal{H}_{\text{lin}} \qquad [62]$$

such that

$$B_+^\gamma(h) = B_+^\gamma(h) \otimes \mathbb{I} + (\text{id} \otimes B_+^\gamma)\Delta(h) \qquad [63]$$

where $B_+^\gamma(h) = \sum_\Gamma n(\gamma, h, \Gamma)\Gamma$. The coefficients $n(\gamma, h, \Gamma)$ are closely related to the section coefficients noted earlier.

Using the definition of the Bogoliubov map $\bar{\phi}$, this immediately shows that

$$S_R^\phi(B_+^\gamma(h)) = \int D_\gamma \leftarrow_{G_i} \phi_{\text{R}}(h) \qquad [64]$$

which proves locality of counter-terms upon recognizing that $B_+^\gamma$ increases the augmentation degree. Here, the insertion of the functions for the subgraph is achieved using the relevant gluing data of [36].

To recover the quantum equation of motions from the Hochschild cohomology, one proves that

$$\Gamma^{\underline{r}} = 1 + \sum_\gamma \frac{g^\gamma}{\text{Sym}(\gamma)} B_+^\gamma(X_\gamma) \qquad [65]$$

where

$$X_\gamma = \prod_{e \in \gamma^{[1]}\text{int}} \prod_{v \in \gamma^{[0]}} \frac{\Gamma^{\underline{v}}}{\Gamma^{\underline{e}}} \qquad [66]$$

has the required solution. Upon application of the Feynman rules, the maps $B_+^\gamma$ turn into the integral kernels of the usual Dyson–Schwinger equations. This allows for new nonperturbative approaches which are a current theme of investigation.

Finally, we note that the 1-cocycles introduced above allow one to determine sub-Hopf algebras of the form

$$\Delta(c_n^r) = \sum P(\{c_j^s\}) \otimes c_j^r \qquad [67]$$

where the $c_j^s$ are defined in eqn [3]. These algebras do not necessitate the considerations of single Feynman graphs any longer, but allow one to establish renormalization directly for the sum of all graphs at a given loop order. Hence, they establish a Hopf algebra structure on time-ordered products in momentum space. For theories with internal symmetries, one expects and indeed finds that the existence of these subalgebras establishes relations between graphs that are same as the Slavnov–Taylor identities between the couplings in the Lagrangian.

## Outlook

Thanks to the Hopf and Lie algebra structures described above, quantum field theory has started to reveal its internal mathematical structure in recent years, which connects it to a motivic theory and arithmetic geometry. Conceptually, quantum field theory has been the most sophisticated means by which a physicist can describe the character of the physical law. We have slowly begun to understanding that, in its short–distance singularities, it encapsulates concepts of matching beauty. We can indeed expect local point-particle quantum field theory to remain a major topic of mathematical physics investigations in the foreseeable future.

*See also:* Bicrossproduct Hopf Algebras and Noncommutative Spacetime; Exact Renormalization Group; Hopf Algebras and q-Deformation Quantum Groups; Number Theory in Physics; Operads; Perturbation Theory and Its Techniques; Renormalization: General Theory; von Neumann Algebras: Introduction, Modular Theory, and Classification Theory.

## Further Reading


Bloch S, Esnault H, and Kreimer D (2005) *On Motives associated to graph polynomials.*

Connes A and Kreimer D (1998) Hopf algebras, renormalization and noncommutative geometry. *Communications in Mathematical Physics* 199: 203 (arXiv:hep-th/9808042).

Connes A and Kreimer D (2001) Renormalization in quantum field theory and the Riemann–Hilbert problem. II: The beta-function, diffeomorphisms and the renormalization group. *Communications in Mathematical Physics* 216: 215 (arXiv:hep-th/0003188).

Connes A and Marcolli M (2004) From physics to number theory via noncommutative geometry. II: Renormalization, the Riemann–Hilbert correspondence, and motivic Galois theory, arXiv:hep-th/0411114.

Kreimer D (2003) New mathematical structures in renormalizable quantum field theories. *Annals of Physics* 303: 179 (arXiv:hep-th/0211136).

Kreimer D (2004) The residues of quantum field theory: numbers we should know, arXiv:hep-th/0404090.

Kreimer D (2005) *Anatomy of a gauge theory.*

# Hopf Algebras and $q$-Deformation Quantum Groups

**S Majid**, Queen Mary, University of London,
London, UK



## Introduction

Quantum groups are a remarkable generalization of conventional groups using an algebraic language by now quite well known to mathematical physicists. This language is first and foremost the concept of a "Hopf algebra." In fact, the axioms of a Hopf algebra are so attractive from a mathematical point of view that they were proposed in the 1940s long before the advent of truly representative examples, which did not come until the 1980s (from mathematical physics). Until then, they were used mainly by mathematicians as a way for redoing group theory and Lie algebra theory in a more uniform way.

It is remarkable that at least three points of view lead to the same axioms of a Hopf algebra:

1. *Generalized symmetry* A generalization of a usual group algebra or enveloping algebra of a Lie algebra that can nevertheless act on other algebraic objects. The structure that controls this is the "coproduct" $\Delta: H \to H \otimes H$, while the group or Lie structure is encoded in the algebra $H$ which is typically not changed up to isomorphism. $\Delta$ allows $H$ to act on tensor products and this is needed to define what it means, for example, for a product $A \otimes A \to A$ of an algebra to be an intertwiner. The usual flip map between two representations $V \otimes W \to W \otimes V$ is not typically an intertwiner any more, instead that is provided by an $R$-matrix solving the Yang–Baxter equations (YBE).
2. *Noncommutative geometry* A generalization of the coordinate algebra of functions on a conventional group to allow noncommutative or "quantum" coordinate algebras. Here the group structure is encoded in a coproduct $\Delta: H \to H \otimes H$ in a way which would, in the case of functions on a group, be defined by the group product. It is typically not changed, the change being in the algebra.
3. *Duality* An object that admits observer–observed duality or Fourier transform. Such a duality is known for abelian groups, lost for nonabelian groups but re-emerges for Hopf algebras. If there is to be an algebra with product $H \otimes H \to H$, then there should also be a "coproduct" $\Delta: H \to H \otimes H$ to maintain the duality symmetry. Then a suitable dual space $H^*$ is also a Hopf algebra, with the roles of product and coproduct interchanged.

In line with these main ideas are three known classes of true quantum groups, and these remain the main types of example at the time of writing: the $q$-deformed enveloping algebras $U_q(g)$ of Drinfeld and Jimbo, their duals as quantizations of the Drinfeld–Sklyanin Poisson bracket on a simple Lie group (both of these arising from quantum inverse scattering but also in the case of $\mathbb{C}_q[SU_2]$ from $C^*$-algebras) and the bicrossproduct quantum groups based on Lie group factorizations (arising from ideas for Planck-scale physics and quantum gravity). The latter are self-dual and hence are both generalized symmetries and noncommutative or quantum geometries at the same time. The impact of such quantum groups has been very far reaching from a mathematician's point of view, spanning revolutions in the theory of knot and 3-manifold invariants, Poisson geometry, new directions in noncommutative geometry, to name some. In physics they are, at the time of writing, beginning seriously to be applied in a variety of contexts beyond the original ones, such as in bookkeeping overlapping divergences in general quantum field theories, quantum computing, and construction of anyons. This article will mention some of these, but just as groups have many different roles in physics, one can expect that quantum groups and variants of them can and will have diverse roles as well. What follows is a short overview.

## Hopf Algebras and First Examples

The general theory works over any field $k$ but (to be concrete) we write our examples over $\mathbb{C}$; one can also have examples over, say, the field $\mathbb{Z}_2$ of two elements. A Hopf algebra then is:

1. An algebra $H$ with unit which is also a "coalgebra" with counit, that is, there are maps $\Delta: H \to H \otimes H, \epsilon: H \to k$ obeying:

$$(\Delta \otimes \mathrm{id})\Delta = (\mathrm{id} \otimes \Delta)\Delta$$
$$(\epsilon \otimes \mathrm{id})\Delta = (\mathrm{id} \otimes \epsilon)\Delta = \mathrm{id}$$

2. $\Delta, \epsilon$ should be algebra homomorphisms.
3. There should be a map $S: H \to H$ called the antipode or "linearized inverse" obeying

$$\cdot(\mathrm{id} \otimes S)\Delta = \cdot(S \otimes \mathrm{id})\Delta = 1\epsilon$$

If the third axiom is not obeyed one has a "quantum semigroup" or "bialgebra." Note also that $S$ looks nothing like a usual inverse and it is not, yet it plays the same role. For example, we can define conjugation or the "adjoint action" of any Hopf algebra on itself by

$$\mathrm{Ad}_a(b) = \sum a_{(1)} b S a_{(2)}, \quad \Delta a = \sum a_{(1)} \otimes a_{(2)}$$

where we use here the "Sweedler notation" for $\Delta a$ a sum of unspecified pieces in $H \otimes H$. Moreover, if it exists, then $S$ is unique and (it can be shown) $S(ab) = (Sb)S(a)$ for all $a, b \in H$, just like an inverse.

The self-duality of these axioms is evident from the first one: a coalgebra is just an algebra with its product map $H \otimes H \to H$, unit element (viewed as a map $k \to H$ sending 1 to 1) and the associativity and unity axioms all written backwards. Meanwhile, the middle axiom means in explicit terms $\Delta(ab) = (\Delta a)(\Delta b), \epsilon(ab) = \epsilon(a)\epsilon(b)$ for all $a, b \in H$ and $\Delta(1) = 1 \otimes 1, \epsilon(1) = 1$. This may not look self-dual but it is equivalent to saying that the product and unit are coalgebra homomorphisms. Indeed, if one takes the trouble to write out all the axioms as commutative diagrams, the set of axioms is invariant under arrow reversal. Such arrow reversal can also be concretely implemented, for example, by taking adjoints. Thus, the coproduct dualizes to a map $(H \otimes H)^* \to H^*$ and since $H^* \otimes H^* \subseteq (H \otimes H)^*$ we have a product on the dual $H^*$. If the dual space is defined correctly, one also has a coproduct by dualizing the product, etc. One says that two Hopf algebras $H, H'$ are "in duality" if their maps are adjoint to each other in such a way.

The role of quantum groups as generalized symmetries is typified by the following examples. Thus, let $G$ be a group; then its group algebra $\mathbb{C}G$ defined as a vector space (written here over $\mathbb{C}$) with basis identified with $G$ and product given by the group product extended linearly, is a Hopf algebra with

$$\Delta g = g \otimes g, \ \epsilon g = 1, \ Sg = g^{-1}, \quad \forall g \in G$$

Likewise, if $\mathfrak{g}$ is a Lie algebra, then its universal enveloping algebra $U(\mathfrak{g})$ generated by $\mathfrak{g}$ is a Hopf algebra with

$$\Delta \xi = \xi \otimes 1 + 1 \otimes \xi, \quad \epsilon \xi = 0, \quad S\xi = -\xi, \quad \forall \xi \in \mathfrak{g}$$

The two examples are related if one informally allows exponentials, then $g = e^{\xi}$ has coproduct

$$\Delta e^{\xi} = e^{\Delta \xi} = e^{\xi \otimes 1 + 1 \otimes \xi} = e^{\xi} \otimes e^{\xi}$$

using axiom 2 and that $\xi \otimes 1, 1 \otimes \xi$ commute in the tensor product algebra.

The coproduct structures are therefore implicit already in Lie theory and group theory. As for any Hopf algebra $\Delta$, specifies how the algebra $H$ acts in a tensor product of two representations. For groups the tensor product is diagonal ($g$ acts on each copy), for Lie algebras it is additive (e.g., the addition of angular momenta). In general, the action of $a \in H$ is defined as the action of $\Delta a$ on the tensor product. This has far-reaching consequences. For example, for the product $A \otimes A \to A$ of an algebra to be covariant means that $H$ acting before and after the product map gives the same answer, similarly for the unit map where $k$ has the trivial representation afforded by $\epsilon$, that is,

$$h \triangleright (ab) = \sum (h_{(1)} \triangleright a)(h_{(2)} \triangleright b), \quad h \triangleright 1 = \epsilon(h)1$$

for all $a, b \in A$ and $h \in H$. What that means in the case of a group is therefore $g \triangleright (ab) = (g \triangleright a)(g \triangleright b)$ or $G$ acts by automorphisms. What it means for a Lie algebra is $\xi \triangleright (ab) = (\xi \triangleright a)b + a(\xi \triangleright b)$, that is, $\mathfrak{g}$ acts by derivations. This is how Hopf algebra theory unifies group theory and Lie algebra theory and potentially takes us beyond.

In another, dual, point of view, if $G$ is a group defined by polynomial equations in $\mathbb{C}^n$, then the Hilbert's "nullstellensatz" in algebraic geometry says that it corresponds algebraically to a commutative nilpotent-free algebra with $n$ generators, called its "coordinate algebra" $H = \mathbb{C}[G]$. The group product then corresponds to $\Delta$ making $\mathbb{C}[G]$ into a Hopf algebra. If one replaces $\mathbb{C}$ by any field, one has an algebraic group over the field. For example, the group $\mathrm{SL}_2(\mathbb{C}) \subset \mathbb{C}^4$ has coordinate algebra generated by four functions $a, b, c, d$ where $a$ at matrix $g \in \mathrm{SL}_2(\mathbb{C})$ has value $g_{11}$ the 1,1 entry of the matrix, similarly $b(g) = g_{12}$ etc. Then $\mathbb{C}[\mathrm{SL}_2]$ is the commutative algebra generated by $a, b, c, d$ with the relation $ad - bc = 1$. A little thought about matrix multiplication should convince the reader that

$$\Delta \begin{pmatrix} a & b \\ c & b \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \otimes \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

where we have written the operation on each generator as an array and where matrix multiplication is understood (so $\Delta a = a \otimes a + b \otimes c$, etc.). The counit and antipode are

$$\epsilon \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$S \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

One could also let $G$ be a finite group, in which case the algebra $\mathbb{C}(G)$ of (say complex-valued) functions on it is more obviously a Hopf algebra with

$$(\Delta a)(g, h) = a(gh), \ \epsilon(a) = a(1), \ (Sa)(g) = a(g^{-1})$$

for any function $a \in \mathbb{C}(G)$. Here we identify $\mathbb{C}(G) \otimes \mathbb{C}(G) = \mathbb{C}(G \times G)$ or functions in two variables on the group. These examples are dually paired with $U(\mathfrak{g})$ in the Lie case and $\mathbb{C}G$ in the finite case, respectively.

In such a coordinate algebra point of view, usual constructions in group theory appear expressed backwards with arrows reversed. So an action of the group appears for such a Hopf algebra $H$ as a "coaction" $\Delta_R : V \to V \otimes H$ (here a right coaction, one can similarly have $\Delta_L$ a left coaction). It obeys

$$(\Delta_R \otimes \mathrm{id})\Delta_R = (\mathrm{id} \otimes \Delta)\Delta_R, \qquad (\mathrm{id} \otimes \epsilon)\Delta_R = \mathrm{id}$$

which are the axioms of an algebra acting written backwards for the coalgebra of $H$ "coacting." An example is the right action of a group on itself which in the coordinate ring point of view is $\Delta_R = \Delta$, that is, the coproduct viewed as a right coaction. It is the algebra of $H$ that determines the tensor product of two coactions, so, for example, $A$ is a coaction algebra in this sense if $\Delta_R : A \to A \otimes H$ is a coalgebra and an algebra homomorphism. Similarly, in this coordinate point of view, an integral on the group means a map $\int : H \to k$ and right invariance translates into invariance under the right coaction, or

$$\left(\int \otimes \mathrm{id}\right)\Delta = 1 \int$$

There is a theorem that such an integration, if it exists, is unique up to scale. In the finite-dimensional case it always exists, for any field $k$. At least in this case, let $\exp = \sum_i e_i \otimes f^i$ for a basis $\{e_i\}$ of $H$ and $\{f^i\}$ a dual basis. Then an application of the integral is Fourier transform $H \to H^*$ defined by

$$\mathcal{F}(a) = \int \sum_i e_i a \otimes f^i$$

with properties that one would expect of Fourier transform. The inverse is given similarly the other way up to a normalization factor and using the antipode of $H$. This is one among the many results from the abstract theory of Hopf algebras, see Sweedler (1969) and Larson and Radford (1988) among others.

A given Hopf algebra $H$ does not know which point of view one is taking on it; the axioms of a Hopf algebra include and unify both enveloping and coordinate algebras. So an immediate consequence is that constructions which are usual in one point of view give new constructions when the wrong point of view is taken (put another way, the self-duality of the axioms means that any general theorem has a second theorem for free, given, if we keep the interpretation of $H$ fixed, by reversing all arrows in the original theorem and its proof). Even the elementary examples above are quite interesting for physics if taken "upside down" in this way. For example, if $G$ is nonabelian, then $\mathbb{C}G$ is noncommutative, so it cannot be functions on any actual group. But it is a Hopf algebra, so one could think of it as being like $\mathbb{C}(\hat{G})$, where $\hat{G}$ is not a group but a quantum group defined as $\mathbb{C}(\hat{G}) = \mathbb{C}G$. The latter is a well-defined Hopf algebra viewed the wrong way. So this is an application of noncommutative geometry to allow nonabelian Fourier transform $\mathcal{F} : \mathbb{C}(G) \to \mathbb{C}G$. Similarly, $U(\mathfrak{g})$ is noncommutative but one could view it upside down as a quantization of $\mathbb{C}[\mathfrak{g}^*] = S(\mathfrak{g})$ (the symmetric algebra on $\mathfrak{g}$). To do this let us scale the generators of $\mathfrak{g}$ so that the relations on $U(\mathfrak{g})$ have the form $\xi\eta - \eta\xi = \lambda[\xi, \eta]$ where $\lambda$ is a deformation parameter. Then the Poisson bracket that this algebra quantizes (deforms) is the Kirillov–Kostant one on $\mathfrak{g}^*$ where $\{\xi, \eta\} = [\xi, \eta]$. Here $\xi, \eta$ on the left-hand side are regarded as functions on $\mathfrak{g}^*$, while on the right-hand side we take their Lie bracket and then regard the result as a function on $\mathfrak{g}^*$. Examples which have been used successfully in physics include:

$$[t, x_i] = \mathrm{i}\lambda x_i \qquad \left(\text{bicrossproduct model } \mathbb{R}^{1,3}_\lambda\right)$$
$$[x_i, x_j] = \mathrm{i}2\lambda\epsilon_{ijk}x_k \qquad \left(\text{spin space model } \mathbb{R}^3_\lambda\right)$$

(summation understood over $k$). In both cases, we may develop geometry on these algebras using quantum group methods as if they were coordinates on a usual space (*see* Bicrossproduct Hopf Algebras and Noncommutative Spacetime). They are versions of $\mathbb{R}^n$ because the coproduct which expresses the addition law on the noncommutative space is the additive one according to the above. In the second case, setting the Casimir to the value for a spin $j$ is the quadratic relation of a "fuzzy sphere." As algebras, the latter are just the algebras of $(2j+1)\times(2j+1)$ matrices.

Going the other way, we can take a classical coordinate ring $\mathbb{C}[G]$ and regard it upside down as some kind of group or enveloping algebra but with a nonsymmetric $\Delta$. In the finite group case, an action of $\mathbb{C}(G)$ just means a $G$-grading. Here if an element $v$ of a vector space has $G$-valued degree $|v|$ then $a \triangleright v = a(|v|)v$ is the action of $a \in \mathbb{C}(G)$. Alternatively, this is the same thing as a right coaction of $\mathbb{C}G, \Delta_R v = v \otimes |v|$. Thus, the notion of group representation and group grading are also unified. This is familiar in physics for abelian groups (a $U(1)$ action is the same thing as a $\mathbb{Z}$-grading) but works fine using Hopf algebra methods for nonabelian groups and beyond.

Returning to axioms, if one wants to speak of real forms and unitary representations, this corresponds, for Hopf algebras, to $H$ a $*$-algebra over $\mathbb{C}$ with

$$\Delta^*(\ ) = \tau(*\otimes *)\Delta, \qquad * \circ S = S^{-1}\circ *$$

where $\tau$ (throughout this article) denotes transposition of tensor factors. This requires in particular that $S$ in invertible (which is not assumed for a general Hopf algebra though it does hold in the finite-dimensional case and in all examples of interest). Thus, $\mathbb{C}[\mathrm{SU}_2]$ denotes the above with a certain $*$ structure whereby the matrix of generators is unitary.

## *q*-Deformation Enveloping Algebras

For a genuinely representative example of a Hopf algebra, consider, $U_q(sl_2)$ defined with noncommutative generators and relations, coproduct etc.,

$$\begin{aligned} q^{h/2}x_\pm q^{-h/2} &= q^{\pm 1}x_\pm \\ [x_+, x_-] &= \frac{q^h - q^{-h}}{q - q^{-1}} \\ \Delta x_\pm &= x_\pm \otimes q^{h/2} + q^{-h/2}\otimes x_\pm \\ \Delta q^{h/2} &= q^{h/2}\otimes q^{h/2} \\ \epsilon x_\pm &= 0, \quad \epsilon q^{h/2} = 1 \\ S x_\pm &= -q^{\pm 1}x_\pm, \qquad S q^{h/2} = q^{-h/2} \end{aligned}$$

The actual generators here are $x_\pm, q^{\pm h/2}$ but the notation is intended to be suggestive: if $h$ existed and we took the limit $q \to 1$, we would have the usual enveloping algebra of the Lie algebra $sl_2$. The quantum group $U_q(su_2)$ is the same with the $*$-structure $h^* = h, x_\pm^* = x_\mp$ when $q$ is real (there are other possibilities).

Two words of warning here. Although some authors write $q = e^{\hbar/2}$, the parameter $q$ here has little to do with quantization. In fact, the cases of direct relevance to physics are $q^{2\pi i/(2+k)}$, where $k$ is the level of the Wess–Zumino–Witten (WZW) model in which this quantum group appears as a generalized symmetry. This quantum group also (first) appeared in the theory of exactly solvable lattice models, namely the Ising model with an applied external magnetic field: $q \neq 1$ is a measure of the resulting nonhomogeneity of the model. Its origins go further back to the algebraic Bethe ansatz and the emergence of the YBE in such models (Baxter 1982). The general $U_q(\mathfrak{g})$ emerged from this context in Drinfeld (1987) and Jimbo (1985) and the same remark applies (*see* Affine Quantum Groups; Yang–Baxter Equations).

The second warning is that at least informally (if one works with $H$ and allows formal power series etc.), the algebra here is isomorphic to usual $U(sl_2)$, that is, it looks deformed but the true deformation is not here but in the coproduct, which enters into the tensor product of representations. The latter are labeled as usual because the algebra is not really changed, for example, the unitary ones of $U_q(su_2)$ are labeled by spin. The spin-$\frac{1}{2}$ one even looks the same with $x_\pm, h$ represented by the standard Pauli matrices. Tensor products of representations start to look different but their multiplicities are the same as classically and if $V, W$ are representations then $V\otimes W \cong W \otimes V$. Because the coproduct above is not symmetric in its two factors, this isomorphism $\Psi_{V,W} = \tau \circ R_{V,W}$ has $R_{V,W}$ nontrivial. From the formulas given, the reader can compute that

$$R_{1/2,1/2} = q^{-1/2}\begin{pmatrix} q & 0 & 0 & 0 \\ 0 & 1 & q - q^{-1} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & q \end{pmatrix}$$

in a tensor product basis. For this particular quantum group, and others like it, one finds that these "$R$-matrices" obey the braid relations as a version of the YBE. As a result, they can and do lead to knot invariants; the one above leads to the Jones knot invariant as a polynomial in $q$. Briefly, one represents the knot on a plane, assigns $R$ or $R^{-1}$ to each braid crossing and takes a suitable trace (*see* The Jones Polynomial).

Since such features hold in any representation, these matrices are in fact representations of an invertible element $\mathcal{R} \in H \otimes H$ provided one allows $h$ as a generator and formal power series:

$$\mathcal{R} = q^{(h\otimes h)/2}\, e_{q^{-2}}^{(q-q^{-1})e\otimes f}; \; e = x_+ q^{h/2}, \; f = q^{-h/2}x_-$$

where

$$e_q(x) = \sum_{m=0}^{\infty}\frac{x^m}{[m]_q!}, \qquad [m]_q = \frac{1-q^m}{1-q}$$

are the $q$-exponential and $q$-integer, respectively. Their proper explanation is in the section "Braided groups and quantum planes." This $\mathcal{R}$ is called the "universal R-matrix" or quasitriangular structure and obeys

$$\tau\Delta = \mathcal{R}(\Delta\ )\mathcal{R}^{-1}$$
$$(\Delta\otimes \mathrm{id})\mathcal{R} = \mathcal{R}_{13}\mathcal{R}_{23}, \qquad (\mathrm{id}\otimes\Delta)\mathcal{R} = \mathcal{R}_{13}\mathcal{R}_{12}$$

and from the axioms of a Hopf algebra, one may deduce that the YBE

$$\mathcal{R}_{12}\mathcal{R}_{13}\mathcal{R}_{23} = \mathcal{R}_{23}\mathcal{R}_{13}\mathcal{R}_{12}$$

hold in the algebra. This induces the YBE for matrices $R_{V,W}$ in the representation $V\otimes W$. Such a

Hopf algebra is called "quasitriangular" and its representations form a braided category (*see* Braided and Modular Tensor Categories). Even if $\mathcal{R}$ for a quasitriangular Hopf algebra is defined by a power series, the $R_{V,W}$ in finite-dimensional representations are typically actual matrices.

Of considerable interest is the special case when $q$ is a primitive $n$th root of unity. In this case the quasitriangular Hopf algebra $u_q(sl_2)$ has the above generators but the additional relations

$$e^n = f^n = 0, \qquad g^n = 1; \quad g = q^h$$

which render the algebra generated by $e, f, g$ as $n^3$-dimensional. The algebra no longer has a matrix block decomposition (is not semisimple) and not all representations descend to it. For example, if $n$ is odd, then only representations of dimension $\leq n$ descend. Other than this, one has many of the features of a classical enveloping algebra now for this finite-dimensional object. There is evidence that such objects over $\mathbb{C}$ are intimately related to classical Lie algebras but over a finite field.

Finally, there is a similar theory of $U_q(\mathfrak{g})$ for all Lie algebras determined by symmetrizable Cartan matrices $\{a_{ij}\}$, including affine ones. Here $i, j \in I$ an indexing set and $a_{ij} = 2i \cdot j / i \cdot i \in \{0, -1, -2, \ldots\}$ for $i \neq j$, where $\cdot$ is a symmetric bilinear form on the root lattice $\mathbb{Z}[I]$ generated by $I$ with $i \cdot i$ a positive even integer. To be precise, one should also fix a "root datum" in the form of an inclusion $\mathbb{Z}[I] \subseteq X$ of the root lattice into a choice of character lattice $X$ and an inclusion $\mathbb{Z}[I] \subseteq Y$ of the coroot lattice (also labeled by $I$) into the cocharacter lattice $Y$ (the dual of $X$). Here the evaluation pairing is required to restrict to $\langle \check{i}, j \rangle = a_{ij}$ if $i, j \in I$ and $\check{i}$ is $i$ viewed in the cocharacter lattice $Y$. We let $q_i = q^{i \cdot i/2}$ and require $q_i^2 \neq 1$ for all $i$ (or one may consider $q$ as an indeterminate). We have generators $e_i, f^i$ for $i \in I$ and invertible $g_a$ for $a$ each generator of $Y$, and the relations

$$g_a e_i = q^{\langle a,i \rangle} e_i g_a, \qquad f^i g_a = q^{\langle a,i \rangle} g_a f^i$$

$$[e_i, f^j] = \frac{g_{\check{i}}^{i \cdot i/2} - g_{\check{i}}^{-i \cdot i/2}}{q_i - q_i^{-1}} \delta_i^j$$

$$\sum_{r=0}^{1-a_{ij}} (-1)^r \begin{pmatrix} 1 - a_{ij} \\ r \end{pmatrix}_{q_i} (e_i)^r e_j (e_i)^{1-a_{ij}-r} = 0$$

for all $i \neq j$ and an identical set for the $\{f^i\}$. The coalgebra and antipode are

$$\Delta e_i = e_i \otimes g_{\check{i}}^{i \cdot i/2} + 1 \otimes e_i$$

$$\Delta f^i = f^i \otimes 1 + g_{\check{i}}^{-i \cdot i/2} \otimes f^i$$

$$\Delta g_a = g_a \otimes g_a, \quad \epsilon(g_a) = 1, \quad \epsilon(e_i) = \epsilon(f^i) = 0$$

$$S g_a = g_a^{-1}, \quad S e_i = -e_i g_{\check{i}}^{-i \cdot i/2}, \quad S f^i = -g_{\check{i}}^{i \cdot i/2} f^i$$

The $q$-Serre relations are those above involving the $q$-binomial coefficients, defined now using the symmetric $q$-integers $(m)_q = (q^m - q^{-m})/(q - q^{-1})$. They have their true explanation as

$$\mathrm{Ad}_{(e_i)^{1-a_{ij}}}(e_j) = 0$$

where Ad is a braided group adjoint action in the sense of the section "Braided groups and quantum planes." Notice that while the root generators are modeled on the Lie algebra, the Cartan generators are modeled on the torus of an algebraic group, which contains global information. Thus, the more precise form of $U_q(sl_2)$ is the $e$, $f$, $g$ form with the generator $g = q^h$ as above, with $\mathbb{Z}[I] \subsetneq X$ and $\mathbb{Z}[I] = Y$. Meanwhile $U_q(psl_2)$ has the square root of this as generator (what we called $q^{h/2}$ before) with $\mathbb{Z}[I] = X$ and $\mathbb{Z}[I] \subsetneq Y$ where the strict inclusion has $\check{1} = 2$ in the lattice $\mathbb{Z}$. Note that, in the complex case, $SL_2$ has compact real form $SU_2$ while its quotient, $PSL_2$, has compact real form $SO_3$, so these are distinguished at the Hopf algebra level. In general, the root datum has an associated reductive algebraic group which is simply connected when $Y = \mathbb{Z}[I]$ and generated by its adjoint representation when $X = \mathbb{Z}[I]$. The complexified character lattice is a sublattice of the more familiar Lie algebra weight lattice and labels representations that extend to the (algebraic) group. Langlands duality interchanges the roles of $X, Y$. These subtleties are lost when we work over formal power series with $q = e^{\lambda/2}$ and Lie-algebra-like Cartan generators.

These objects are mathematically so interesting that some authors define "quantum groups" as nothing more than this particular extension of the theory of Lie algebras, Cartan matrices and root systems. Among the deepest theorems is the existence of the Lusztig–Kashiwara canonical basis which is obtained from $q = 0$ but valid also at $q = 1$ (i.e., for classical enveloping algebras) and which has the remarkable property of inducing bases coherently across highest-weight representations. From a physicist's point of view, however, there are many other Hopf algebras rather more closely connected with actual quantization. Most often, the terms quantum group and Hopf algebra are used interchangeably.

There is similarly a reduced version $u_q(\mathfrak{g})$. The simplest of all possible cases, even simpler than $u_q(sl_2)$, is for what one could call $u_q(1)$ with a single generator $g$ and

$$g^n = 1, \ \Delta g = g \otimes g, \ \epsilon g = 1, \ S g = g^{-1}$$

$$\mathcal{R}_q = \frac{1}{n} \sum_{a,b=0}^{n-1} q^{-ab} g^a \otimes g^b$$

where $q$ is a primitive $n$th root of unity. The Hopf algebra is the same as the group algebra $\mathbb{C}\mathbb{Z}_n = \mathbb{C}(\hat{\mathbb{Z}}_n)$ but the $\mathcal{R}$ is nontrivial. A representation means a $\hat{\mathbb{Z}}_n$-graded space, that is, graded into degrees $0, 1, \ldots, n-1$. The braiding matrices have the diagonal form $R_{V_a, W_b} = q^{ab}$ on components of degree $a$, $b$, respectively. The braided category generated in this case is the one where anyons live. From this point of view, $u_q(g)$ generate the category where nonabelian anyons live. Here $\mathcal{R}_{q^2}$ (in place of $q^{(h\otimes h)/2}$) along with an additional $e_{q^{-2}}$ factor as above gives the quasitriangular structure of $u_q(sl_2)$. The physical model here is the rational conformal field theory mentioned above with these anyons as particular bound states. There is a proposal to use them in the construction of quantum computers.

## q-Deformation Coordinate Algebras

From the coordinate algebra point of view, the corresponding deformation to the one in the last section is the Hopf algebra $\mathbb{C}_q[SL_2]$ with noncommuting generators and relations

$$\begin{aligned} ca &= qac, & ba &= qab \\ db &= qbd, & dc &= qcd \\ bc &= cb, & da - ad &= (q - q^{-1})bc \\ ad - q^{-1}bc &= 1 \end{aligned}$$

The coalgebra has the same matrix form on the generators as for $\mathbb{C}[SL_2]$ and the antipode and $*$-structure (for $\mathbb{C}_q[SU_2]$) are

$$S\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} d & -qb \\ -q^{-1}c & a \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix}^*$$

Its duality pairing with $U_q(sl_2)$ is afforded by the $2 \times 2$ Pauli-matrix representation of the latter. The $\mathbb{C}_q[SU_2]$ Hopf $*$-algebra may be completed to a $C^*$-algebra.

One similarly has $\mathbb{C}_q[G]$ for all semisimple Lie groups $G$ and their various real forms. From an axiomatic point of view, such quantum groups are "coquasitriangular" in the sense that there is a map $\mathcal{R}: H \otimes H \to k$ such that

$$\sum \mathcal{R}(a_{(1)} \otimes b_{(1)}) a_{(2)} b_{(2)} = \sum b_{(1)} a_{(1)} \mathcal{R}(a_{(2)} \otimes b_{(2)})$$

for all $a, b \in H$ and

$$\begin{aligned} \mathcal{R}(ab \otimes c) &= \sum \mathcal{R}(a \otimes c_{(1)}) \mathcal{R}(b \otimes c_{(2)}) \\ \mathcal{R}(a \otimes bc) &= \sum \mathcal{R}(a_{(1)} \otimes c) \mathcal{R}(a_{(2)} \otimes b) \end{aligned}$$

for all $a, b, c \in H$. We also require that $\mathcal{R}$ is invertible in a certain sense. These are just the arrow reversal of the axioms of a quasitriangular structure. In general, for the deformation of a linear algebraic group we will have some $n^2$ generators $t^i{}_j$, now taken to be noncommutative, and with a matrix form of coalgebra

$$\Delta t^i{}_j = t^i{}_k \otimes t^k{}_j, \qquad \epsilon t^i{}_j = \delta^i{}_j$$

For the compact real form we will have $St^i{}_j = t^{j*}{}_i$. Moreover, from the first of the above axioms we will have among the relations

$$R^i{}_a{}^k{}_b t^a{}_j t^b{}_l = t^k{}_b t^i{}_a R^a{}_j{}^b{}_l$$

where $R^i{}_j{}^k{}_l = \mathcal{R}(t^i{}_j \otimes t^k{}_l)$ is a matrix $R \in M_n \otimes M_n$ obeying the YBE. If we take only these quadratic relations, we have the "Faddier Reshetikhin Takhtajan (FRT) bialgebra" $A(R)$ and it can be shown (see Majid 1995) that $R$ extends to a coquasitriangular structure $\mathcal{R}$ on it. However, in our case we also have

$$\begin{aligned} R^{-1}{}^i{}_j{}^k{}_l &= \mathcal{R}\left(St^i{}_j \otimes t^k{}_l\right) \\ \tilde{R}^i{}_j{}^k{}_l &= \mathcal{R}\left(t^i{}_j \otimes St^k{}_l\right) \end{aligned}$$

where $\tilde{R} = ((R^{t_2})^{-1})^{t_2}$ ($t_2$ transposition in the second factor of $M_n$) is called the "second inverse" of $R$. With these additional matrices, one may define a $q$-determinant and antipode relations as well (Majid 1995). One may also generate a rigid braided monoidal category and reconstruct a Hopf algebra $\check{A}(R)$ from it. In this way, the $R$-matrix plays a role similar to that of the structure constants of a Lie algebra and can in principle define the quantum group coordinate algebra. Such $R$-matrices have been classified in low dimension and include multiparameter and other deformations of classical group coordinate algebra as well as other nonstandard quantum groups.

In the $\mathbb{C}_q[G]$ examples it is not the coalgebra which is essentially deformed but the algebra. We already see this above on the generators but the coproduct of a product of generators may look different. Nonetheless, one can identify the vector space that the products generate with that of $\mathbb{C}[G]$ and at least informally with respect to a deformation parameter express the product as a power series in the undeformed product (a $\bullet$-product deformation). For generic values, one still has a Peter–Weyl decomposition $\mathbb{C}_q[G] = \oplus (V \otimes V^*)$, where the sum is over irreducibles corepresentations, which can be identified with the classical representations of the algebraic group. One can make the same decomposition for $\mathbb{C}[G]$ and identify the matrix blocks $V \otimes V^*$ in order to find this $\bullet$-product. Also, since this is a flat deformation, it follows that the commutator at lowest order defines a Poisson bracket on $G$, given by

$$\{t^i{}_j, t^k{}_l\} = t^i{}_a t^k{}_b r^a{}_j{}^b{}_l - r^i{}_a{}^k{}_b t^a{}_j t^b{}_l$$

and this Poisson bracket is compatible with the group product $G\times G\to G$ as a Poisson map (because the Hopf algebra coproduct was an algebra map). Here $r$ is the first order part in the expansion of the $R$-matrix. A Lie group equipped with a Poisson bracket compatible in this way is called a "Poisson Lie group." On general functions its Poisson bivector is generated by the first order part $r\in\mathfrak{g}\otimes\mathfrak{g}$ in the expansion of $\mathcal{R}$ in the $q$-deformed enveloping algebra. In place of the YBE obeyed by $\mathcal{R}$, we have the "classical Yang–Baxter equations (CYBE),"

$$[r_{12},r_{23}]+[r_{12},r_{13}]+[r_{13},r_{23}]=0$$

In this way, one may characterize an "infinitesimal version" of $U_q(\mathfrak{g})$ as $(\mathfrak{g},r,\delta)$ where $\delta:\mathfrak{g}\to\mathfrak{g}\otimes\mathfrak{g}$ is the leading part of $\tau\Delta-\Delta$ and makes the triple into a quasitriangular "Lie Bialgebra" (*see* Classical $r$-Matrices, Lie Bialgebras, and Poisson Lie Groups).

Finally, returning to our example, when $q$ is an $n$th root of unity, one has the $q$-Frobenius Hopf algebra homomorphism

$$\mathbb{C}[SL_2]\hookrightarrow\mathbb{C}_q[SL_2]$$

$$\begin{pmatrix}a & b\\ c & d\end{pmatrix}\mapsto\begin{pmatrix}a^n & b^n\\ c^n & d^n\end{pmatrix}$$

that is, a classical copy sitting inside the quantum group. Quotienting by this means adding the relations

$$a^n=d^n=1,\qquad b^n=c^n=0$$

which gives the finite-dimensional reduced quantum group $\mathbb{C}_q^{\mathrm{red}}[SL_2]$. Similarly for other $\mathbb{C}_q^{\mathrm{red}}[G]$. These reduced quantum groups provide finite noncommutative geometries having the geometric flavor of the classical geometry but where geometry and physics (such as electromagnetic gauge theory modes) are fully computable.

## Self-Dual Quantum Groups

The arrow-reversibility of the axioms of a quantum group make it possible to search for self-dual quantum groups or for quantum groups which, if not self-dual, have a self-dual form. This leads to the bicrossproduct quantum groups coming from models of quantum gravity (Majid 1988) (*see* Bicrossproduct Hopf Algebras and Noncommutative Spacetime).

The context here is that of Figure 1 which shows how Hopf algebras relate to other objects and to duality in a representation-theoretic sense. Along the central axis, we have put self-dual categories or in physical terms categories admitting Fourier transform. This is clear for abelian Groups where the

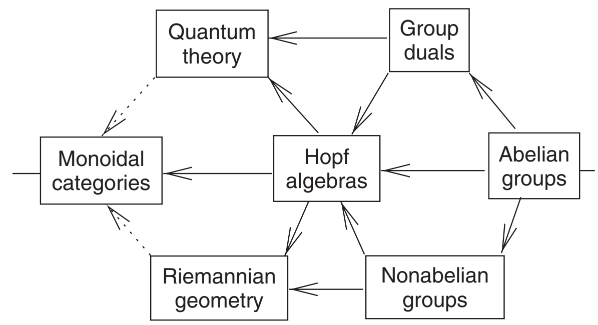



**Figure 1** Role of Hopf algebras along the self-dual axis.

dual $\hat{G}$ of an abelian group $G$ is also an abelian group. Below the axis, we have nonabelian groups which we view as toy models of geometries with curvature. Every compact Lie group, for example, has an associated Killing metric. Above the axis, a nonabelian group dual $\hat{G}$ means to construct unitary representations etc., which we view as toy models of quantum theory. We have seen that Hopf algebras are another self-dual category and provide a framework in which both groups and group duals can be unified (see the section "Hopf algebras and first examples"). Thus, $G$ can be viewed as a coordinate Hopf algebra $\mathbb{C}(G)$ or $\mathbb{C}[G]$ in the finite or Lie cases, and $\hat{G}$ as the dual Hopf algebra $\mathbb{C}G$ or $U(\mathfrak{g})$ as a definition of the coordinate algebra "$\mathbb{C}(\hat{G})$." Note that $\hat{G}$ is not merely the set of representations, as these alone are not enough to reconstruct the group (e.g., both $S^1$ and $SO_3$ have the same set). We see that Hopf algebras are a microcosm for the unification of quantum theory and gravity. Hopf algebra duality interchanges the role of position and momentum on the one hand and of quantum and gravitational effects on the other. A self-dual Hopf algebra has both aspects unified and interchanged by the self-duality.

One can also ask what the next most general self-dual category of objects is in which to look for more general unifications. One answer here is the category whose objects are themselves categories $\mathcal{C}$ equipped with a tensor product (a "monoidal category") and a monoidal functor to a fixed monoidal category $\mathcal{V}$. Motivated by the above, a theorem from the 1980s is that for any such $\mathcal{C}$ there is a dual $\mathcal{C}^0$ of "representations in $\mathcal{V}$" (Majid 1991a). The dotted arrows in Figure 1 indicate that this may be a setting for more ambitious models than those achieved by Hopf algebras alone. In fact, the $\mathcal{C}^0$ construction was one of the ingredients going into the invention of 2-categories a few years later. See also several articles on TQFT (such as Topological Quantum Field Theory: Overview; Axiomatic Approach to Topological Quantum Field Theory; Duality in Topological Quantum Field Theory).

The simplest self-dual quantum group is $\mathbb{C}[x]$ as the Hopf algebra of polynomial functions on a line with additive coproduct. This is dually paired with itself in the form of the enveloping algebra $U(gl_1) = \mathbb{C}[p]$ with pairing

$$\langle p^m, x^n \rangle = (-\mathrm{i})^n \delta_{m,n} n!$$

and similarly for higher-dimensional flat space. In the case of $\mathbb{C}[x]$, a basis is $x^n$ and from the above the dual basis is $(\mathrm{i}p)^n/n!$. Hence the canonical element is $\exp = \mathrm{e}^{\mathrm{i}x \otimes p}$ so that Hopf algebra Fourier transform on a suitable completion of these algebras reduces to usual Fourier transform.

A more nontrivial example (Majid 1988) is given by the "Planck-scale Hopf algebra" $\mathbb{C}[x] \blacktriangleright\!\!\triangleleft \mathbb{C}[p]$ which has algebra and coalgebra

$$[p, x] = \mathrm{i}\hbar(1 - \mathrm{e}^{-\gamma x}), \qquad \Delta x = x \otimes 1 + 1 \otimes x$$
$$\Delta p = p \otimes \mathrm{e}^{-\gamma x} + 1 \otimes p, \qquad \epsilon x = \epsilon p = 0$$
$$Sx = -x, \qquad Sp = -p\mathrm{e}^{\gamma x}$$

The actual generator here should be $\mathrm{e}^{\gamma x}$ rather than $x$ for an algebraic treatment (otherwise one should allow power series or use $C^*$-algebras). The dually paired Hopf algebra has the same form $\mathbb{C}[p] \triangleright\!\!\blacktriangleleft \mathbb{C}[x]$, with new parameters $\hbar' = 1/\hbar$ and $\gamma' = \hbar\gamma$ and quantum group Fourier transform connects the two. More details and the general construction of Hopf algebras $\mathbb{C}[M] \blacktriangleright\!\!\triangleleft U(\mathfrak{g})$ with dual $U(\mathfrak{m}) \triangleright\!\!\blacktriangleleft \mathbb{C}[G]$ are in the article on "bicrossproduct" Hopf algebras (*see* Bicrossproduct Hopf Algebras and Noncommutative Spacetime). These quantize particles in $M$ moving under momentum Lie group $G$ with Lie algebra $\mathfrak{g}$ and vice versa. The states of one (in a $C^*$-algebra context) lie in the algebra of observables of the other ("observable–state duality"). The data required are a matched pair of actions of $(G, M)$ on each other. Such equations correspond locally to a factorization of a larger group $G \bowtie M$ but typically have singularities and other features in keeping with a toy model of Einstein's equations.

There are, by the time of writing, many applications of bicrossproducts beyond the original one, including a Poincaré quantum group for the $\mathbb{R}^{1,3}_\lambda$ mentioned in the section "Hopf algebras and first examples," with links to Planck-scale physics. There is also a bicrossproduct quantum group $\mathbb{C}[G^*] \blacktriangleright\!\!\triangleleft U(\mathfrak{g})$ canonically associated to any simple Lie algebra $\mathfrak{g}$ and related to $T$-duality. The classical data here are Lie bialgebras and solutions of the CYBE as in the section "$q$-Deformation coordinate algebras," however there is no known relation with the $q$-deformation Hopf algebras themselves. Finite group bicrossproducts are also interesting and examples (but not with both actions nontrivial) were already in the works of GI Kac in the 1960s.

These constructions also work when the groups above are themselves Hopf algebras. For example, any finite-dimensional Hopf algebra $H$ has a "quantum double" $D(H) = H \bowtie H^{*\mathrm{op}}$, where the double cross product $\bowtie$ is by mutual coadjoint actions. The cross-relations between the two sub-Hopf algebras are

$$\sum \langle h_{(1)}, a_{(1)} \rangle h_{(2)} a_{(2)} = \sum a_{(1)} h_{(1)} \langle h_{(2)}, a_{(2)} \rangle$$

for $h \in H$ and $a \in H^*$. The construction is due to Drinfeld (1987) while the $\bowtie$ form is due to the author. Moreover, $D(H)$ is quasitriangular with $\mathcal{R} = \exp$, the canonical element used in the Fourier transform on $H$. Its representations consist of vector spaces where $H$ acts and at the same time $H^{*\mathrm{op}}$ acts or (which makes sense when $H$ is infinite dimensional) where $H$ coacts, in a compatible way. Such objects are called "crossed modules" because when $H = \mathbb{C}G$, one has exactly a linearization of the crossed $G$-sets of JC Whitehead. They are a special case of the $\mathcal{C}^0$ construction mentioned above.

Finally, one can also view the $q$-deformed linear spaces on which quantum groups such as $U_q(\mathfrak{g})$ act as self-dual Hopf algebras under an additive coproduct. However, this needs to be as braided groups or Hopf algebras with braid statistics, see the next section. The simplest example here is the "braided line" $B = \mathbb{C}[x]$ developed not as above but as a self-dual Hopf algebra with $q$-statistics. Its "bosonization" gives a self-dual Hopf algebra $U_q(b_+) \subset U_q(sl_2)$, and similarly for other $U_q(b_+) \subset U_q(\mathfrak{g})$. Perhaps more surprisingly, the quantum groups $U_q(\mathfrak{g})$ and $\mathbb{C}_q[G]$ also both have canonical braided group versions (a process called "transmutation") and as such they too are isomorphic. This isomorphism extends the linear isomorphism $\mathfrak{g} \to \mathfrak{g}^*$ afforded by the Killing form of any semisimple Lie algebra. In physical terms, what this means is that there is in $q$-deformed geometry just one self-dual object $B_q(G)$ with two different scaling limits

$$U(\mathfrak{g}) \longleftarrow B_q(G) \to \mathbb{C}[G]$$

as $q \to 1$, and the structure of which underlies the deeper structure of $U_q(\mathfrak{g})$ and $\mathbb{C}_q[G]$ as well.

## Braided Groups and Quantum Planes

A super quantum group or super-Hopf algebra is not a quantum group or Hopf algebra since the key homomorphism property of $\Delta : H \to H \otimes H$ is modified: one must use in the target $H \underline{\otimes} H$ the $\mathbb{Z}_2$-graded or super tensor product of super algebras. Here,

$$(a \otimes b)(c \otimes d) = (-1)^{|b||c|} ac \otimes bd$$

for elements of degree $|b|, |c|$. Super quantum groups $U_q(gl_{m|n})$ etc., have been constructed and have an analogous theory to the bosonic versions above. Super spaces in physics are associated to differential forms and in the same way a bicovariant exterior algebra on a quantum group $H$ is generally a super quantum group. Here the exterior algebra is generated on by 1-forms and the coproduct on 1-forms is

$$\underline{\Delta} = \Delta_{\mathrm{L}} + \Delta_{\mathrm{R}}$$

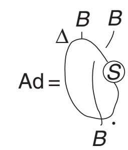
Here $\Delta_{\mathrm{L,R}}$ are the coactions of $H$ on 1-forms induced by the left and right coaction of $H$ on itself.

For a true understanding of quantum groups one must, however, go beyond such objects to "braided groups" or Hopf algebras with braid statistics (see Majid 1995). This theory was introduced by the author in the early 1990s as a more systematic method for $q$-deformation of structures in physics based on $q$-group covariance. We have seen that a quasitriangular quantum group, or any Hopf algebra through its double, generates a braided category with the flip map $\tau$ replaced by a braiding $\Psi_{V,W}$ between any two representations. Anything which is covariant under the quantum group means by definition that it lives in the braided category. Working with such "braided algebras" is similar to working with superalgebras except that one should use $\Psi$ in place of the graded transposition in any algebraic construction. In particular, two braided algebras have a natural "braided tensor product" also in the category. In concrete terms,

$$(a \otimes b)(c \otimes d) = a\Psi(b \otimes c)d$$

Then a Hopf algebra in the braided category or braided group is $B$, an algebra in the category along with a coalgebra and antipode, where $\underline{\Delta}: B \to B\underline{\otimes}B$ is an algebra homomorphism (*see* Braided and Modular Tensor Categories).

Next, we have mentioned in the section "*q*-Deformation enveloping algebras" that $q$-algebras generate topological invariants, but we now turn this on its head and use braid diagrams to do $q$-algebra. We write all operations as flowing down the page, any transpositions in the algebraic construction are expressed as a braid crossing $\Psi = \times$ or its inverse by the reversed braid crossing, and any other operations as nodes. Thus, a product is denoted $\curlyvee$ and a coproduct $\curlywedge$. Algebraic information "flows" along these "wires" much like the way that information flows along the wiring in a computer, except that under- and over-crossings represent distinct nontrivial operators. (In fact, one may formulate topological quantum computers exactly in this way.) In this notation, tensor products are denoted by juxtaposition and the trivial object in the category is omitted. In particular, one has the axioms and all general theorems of Hopf algebras at this diagrammatic level. For example, the adjoint action of any braided group $B$ on itself is (see Majid 1995)




In any concrete example, such diagrams turn into $R$-matrix formulas where $\Psi = \tau R$ as explained in the section "*q*-Deformation enveloping algebras."

A basic example of a braided group is the braided $q$-plane $\mathbb{C}_q^2$ with generators $x, y$ and relations $yx = qxy$. Its coproduct is the additive one $\underline{\Delta}x = x \otimes 1 + 1 \otimes x$ (and similarly for $y$) reflecting addition in the plane, but this is extended to products as a braided group with braiding $q^{1/2}R_{1/2,1/2}$ in terms of the $R$-matrix in the section "*q*-Deformation enveloping algebras." The extra factor here means that $\mathbb{C}_q^2$ lives in the braided category of representations of $U_q(gl_2) = \widetilde{U}_q(sl_2)$ (i.e., with an additional central $U_q(1)$ generator to provide the $q^{1/2}$). More precisely, the category is that of corepresentations of $\mathbb{C}_q[GL_2] = \widetilde{\mathbb{C}}_q[SL_2]$. The coaction in this case is

$$\Delta_{\mathrm{R}}(x\ y) = (x\ y) \otimes \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

where the additional central generator is encoded in the $q$-determinant (which is no longer set equal to 1). Notice that $q^{1/2}R_{1/2,1/2}$ has eigenvalues $q, -q^{-1}$ (one says that it is $q$-Hecke). Another braided group, associated now to the second eigenvalue is $\mathbb{C}_q^{0|2}$ with generators $\xi, \eta$ and relations $\eta\xi = -q^{-1}\xi\eta$, $\xi^2 = \eta^2 = 0$. It is the quadratic algebra dual of $\mathbb{C}_q^2$ (Manin 1988).

One has natural braided linear spaces for the whole family $\mathbb{C}_q[G]$, on which the latter coact after central extension. The general construction is as follows. If $V$ is an object in a braided category (e.g., the fundamental representation of a quantum group), let $T(V)$ be the tensor algebra generated by a basis $\{e_i\}$ of $V$ with no relations and the additive braided coproduct as above. Assume that $V$ has a dual $V^*$ in the category, and similarly form $T(V^*)$ with dual basis generators $\{f^i\}$. These two braided groups will be dually paired by extending the evaluation map to products, which takes the form of "braided integers" (see Majid 1995)

$$\langle f^{i_m} \cdots f^{i_1}, e_{j_1} \cdots e_{j_n} \rangle = \delta_{n,m}[n, \Psi]!^{i_1 \cdots i_n}_{j_1 \cdots j_n}$$

$$[n, \Psi] = \mathrm{id} + \Psi_{12} + \Psi_{12}\Psi_{23} + \cdots + \Psi_{12} \cdots \Psi_{n-1,n}$$

We now quotient by the kernels of this pairing to obtain $B(V), B(V^*)$ as two nondegenerately paired braided groups. This quotient generates all the relations, which are very often but not necessarily quadratic (in practice, one typically imposes only the quadratic relations to have braided groups with a possibly degenerate pairing). The construction is due to the author. Moreover, we can define partial derivatives on these braided groups by $\underline{\Delta} a = 1 \otimes a + e_i \otimes \partial^i a + \cdots$ for any $a$ in the algebra, that is, as an infinitesimal generator of translations under the braided group law; similarly exp, indefinite and Gaussian integration, Fourier transform, etc. The simplest example here is $B = \mathbb{C}[x]$ viewed not as a usual Hopf algebra but as a braided group in the category of $\mathbb{Z}$-graded spaces with $\Psi(x \otimes x) = qx \otimes x$. Also in this example the braided addition law on $\mathbb{C}[x]$ is

$$\underline{\Delta} x^n = \sum_{m=0}^{n} \begin{bmatrix} n \\ m \end{bmatrix}_q x^m \otimes x^{n-m}$$

defined by $[m]_q$, and the partial derivative defined by it is the Jackson (1908) $q$-derivative

$$\partial f(x) = \frac{f(x) - f(qx)}{x(1-q)}$$

while $\underline{\Delta} e_q(x) = e_q(x) \otimes e_q(x)$ if we allow power series. Such objects occur in the theory of $q$-special functions (*see* $q$-Special Functions).

Among deeper theorems (see Majid 1995, 2002), there is a triangular decomposition

$$U_q(\mathfrak{g}) = U_q(n_-) \rtimes T \ltimes U_q(n_+)$$

where $U_q(n_+)$ is a braided group and $U_q(n_-)$ is dually paired to its opposite. $T$ denotes the torus generators $\{g_a\}$ in the section "$q$-Deformation enveloping algebras." More generally, if $\mathfrak{g}_0 \subset \mathfrak{g}$ is a principal embedding of Lie algebras (given by an inclusion of Dynkin diagrams), then $U_q(\mathfrak{g}) = B^* \rtimes \widetilde{U_q(\mathfrak{g}_0)} \ltimes B^{\mathrm{op}}$ for some additive braided group of additional root generators and its dual. The general construction $B^* \rtimes H \ltimes B^{\mathrm{op}}$ here is "double bosonization" which associates to dual braided groups $B, B^*$ in the category of representations of some quasitriangular Hopf algebra $H$, a new quasitriangular Hopf algebra. The simplest example $B = \mathbb{C}[x]$ lives in the category of representations of $T = U_q(1)$ in an algebraic form. The dual is another braided line $\mathbb{C}[p]$ and $\mathbb{C}[p] \rtimes U_q(1) \ltimes \mathbb{C}[x]$ is a version of $U_q(sl_2)$. In this way, the braided line $\mathbb{C}[x]$ is at the root of all $q$-deformation quantum groups.

An earlier theorem is that for any braided group $B$ covariant under a (co)quasitriangular $H$, we have its 'bosonization' $B \rtimes H$. There is a similar "biproduct" if $B$ lives in the category of crossed modules for any Hopf algebra $H$. These have been extensively applied in physics notably in the construction of inhomogeneous quantum groups. Similar to $\mathbb{C}_q^2$ (but as a $*$-algebra), there is a natural self-dual $q$-Minkowski space $B = \mathbb{R}_q^{1,3}$ which is covariant under $U_q(\widetilde{so_{1,3}})$, and its bosonization is the $q$-Poincaré plus dilations group $\mathbb{R}_q^{1,3} \rtimes U_q(\widetilde{so_{1,3}})$. It is not possible to avoid the dilation here. The double-bosonization extends this to the $q$-conformal group $U_q(so_{2,4})$. The braided adjoint action becomes the action of conformal translations on $\mathbb{R}_q^{1,3}$. The construction of $q$-propagators and $q$-deformed physics on such $q$-Minkowski space was achieved in the mid 1990s as one of the main successes of the theory of braided groups.

This $\mathbb{R}_q^{1,3}$ can be given also as a matrix of generators, relations, $*$-structure and, a second braided coproduct:

$$\beta\alpha = q^2 \alpha\beta, \quad \gamma\alpha = q^{-2}\alpha\gamma, \quad \delta\alpha = \alpha\delta$$
$$\beta\gamma = \gamma\beta + (1 - q^{-2})\alpha(\delta - \alpha)$$
$$\delta\beta = \beta\delta + (1 - q^{-2})\alpha\beta$$
$$\gamma\delta = \delta\gamma + (1 - q^{-2})\gamma\alpha$$
$$\underline{\Delta}\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \underline{\otimes} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$
$$\epsilon\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}^* = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix}$$

This is in addition to the additive coproduct above. It corresponds to the point of view of Minkowski space as Hermitian $2 \times 2$ matrices. Note that $\underline{\Delta}$ is not a $*$-algebra map in the usual sense and indeed Hermitian matrices are not a group under multiplication, but this does form a natural braided $*$-bialgebra. If we quotient by the braided determinant relation $\alpha\delta - q^2\gamma\beta = 1$, we have the unit hyperboloid in $\mathbb{R}_q^{1,3}$ which turns out to be the braided group $B_q[SU_2]$ mentioned at the end of the previous section (as obtained canonically from $\mathbb{C}_q[SU_2]$). We now have a braided antipode

$$S\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} q^2\delta + (1 - q^2)\alpha & -q^2\beta \\ -q^2\gamma & \alpha \end{pmatrix}$$

This was the first nontrivial example of a braided group (Majid 1991b) and we see that it has two $q \to 1$ limits

$$U(su_2) \longleftarrow B_q[SU_2] \to \mathbb{C}[\text{Hyperboloid} \subset \mathbb{R}^{1,3}]$$

Because most constructions in physics can be uniformly deformed by such methods (including the totally $q$-antisymmetric tensor), one finds that $q$ provides a new regulator in which infinities in quantum field theory can be in principle be encoded as poles at $q = 1$. That transmutation from the

quantum group to its braided version unifies unitary nonabelian symmetries with pseudo-Riemannian geometry is another deeper aspect of relevance to physics. In addition, $q$-constructions have their original role in quantum integrable systems, at $q$ a root of unity and for infinite-dimensional (affine) Lie algebra deformations.

## Quasi-Hopf Algebras

Although the braided category of representations of a quantum group has a trivial "associator" $\Phi_{V,W,Z}:(V\otimes W)\otimes Z\to V\otimes(W\otimes Z)$ between any three objects, a general braided category and the diagrammatic methods of "braided algebra" in the last section do not require this (one simply translates diagrams into algebra by inserting $\Phi$ as needed). A more general object that generates such categories as its representations is a "quasi-Hopf algebra." This is a generalization of Hopf algebras in which the coproduct $\Delta:H\to H\otimes H$ is not necessarily coassociative. Instead,

$$
\begin{aligned}
&(\mathrm{id}\otimes\Delta)\Delta=\phi((\Delta\otimes\mathrm{id})\Delta\ )\phi^{-1}\\
&(\mathrm{id}\otimes\epsilon\otimes\mathrm{id})\phi=1\\
&\phi_{234}(\mathrm{id}\otimes\Delta\otimes\mathrm{id})(\phi)\phi_{123}\\
&\quad=(\mathrm{id}^2\otimes\Delta)(\phi)(\Delta\otimes\mathrm{id}^2)(\phi)
\end{aligned}
$$

for some invertible element $\phi\in H\otimes H\otimes H$. The numbers denote the position in the tensor product and one says that $\phi$ is a 3-cocycle. The axioms for the antipode and quasitriangular structure $\mathcal{R}$ are also modified. The tensor product of representations is given as usual by $\Delta$, and the braiding and associator by the actions of $\mathcal{R}$ and $\phi$.

This notion, due to Drinfeld (1990), arises when one wishes to write down the quantum groups $U_q(\mathfrak{g})$ more explicitly as built on the algebras $U(\mathfrak{g})$ (recall that they are isomorphic over formal power series). Thus, for each semisimple $\mathfrak{g}$ there is a natural (quasitriangular) quasi-Hopf algebra $(U(\mathfrak{g}),\phi,\mathcal{R})$ where $U(\mathfrak{g})$ has the usual Hopf algebra structure, $\mathcal{R}$ is an exponential of the split Casimir (or inverse Killing form) in $\mathfrak{g}\otimes\mathfrak{g}$ and $\phi$ is constructed as a solution of the Knizhnik–Zamolodchikov equations coming out of conformal field theory. This is not $U_q(\mathfrak{g})$ but it has an equivalent braided category of representations. Thus, there is an element $F\in U(\mathfrak{g})^{\otimes 2}$ (extended over formal power series) such that

$$
\begin{aligned}
&\Delta_F=F(\Delta\ )F^{-1},\qquad \mathcal{R}_F=F_{21}\mathcal{R}F^{-1}\\
&\phi_F\equiv F_{12}(\Delta\otimes\mathrm{id})(F)\phi(\mathrm{id}\otimes\Delta)(F^{-1})F_{23}^{-1}=1
\end{aligned}
$$

recovers $U_q(\mathfrak{g})$ as a quasitriangular Hopf algebra built directly on the algebra $U(\mathfrak{g})$. The conjugation operations here (and a similar process regarding the antipode) are a "Drinfeld twist" of a quasi-Hopf algebra, and such twisting by any invertible $F$ such that

$$(\epsilon\otimes\mathrm{id})F=(\mathrm{id}\otimes\epsilon)F=1$$

(a cochain) does not change the representation category up to equivalence. In the present case, the twist transforms $\phi$ into $\phi_F=1$, that is, into an ordinary Hopf algebra isomorphic over formal power series to $U_q(\mathfrak{g})$. Note that in rational conformal field theory the tensor product of representations appears as a finite-dimensional commutative associative algebra (the Verlinde algebra) with integer structure constants $N^{ij}{}_k$ (this comes from the operator-product expansion of primary fields in the theory). This is because one has more precisely a truncated representation category corresponding to $q$ a root of unity, and because we are identifying equivalent representations (so $N^{ij}{}_k$ are the multiplicity in the decomposition of a tensor product of two representations). However, if one wants to know the tensor product decomposition more fully, not just its isomorphism class, this is given in a choice of bases by recoupling matrices. Computation in terms of these shows that the actual tensor product is neither commutative nor associative, but of the form above at least in the case of the WZW model.

Hopf algebra theory typically extends to the quasi-Hopf case. For example, given a quasi-Hopf algebra $H$ there is a quantum double $D(H)$ at least in the finite-dimensional case, due to the author. An example is to take $H=\mathbb{C}(G)$ and $\phi$ a 3-cocycle on $G$ in the usual sense

$$
\begin{aligned}
&\phi(y,z,w)\phi(x,yz,w)\phi(x,y,z)\\
&\quad=\phi(x,y,zw)\phi(xy,z,w)
\end{aligned}
$$

on elements of $G$ and $\phi(x,1,y)=1$. Then $(\mathbb{C}(G),\phi)$ can be viewed as a quasi-Hopf algebra. Its double $D^{\phi}(G)$ is generated by $\mathbb{C}(G)$ as a sub-quasi-Hopf algebra and by elements of $G$ with

$$
\begin{aligned}
&x\cdot y=yx\sum_s\delta_s\chi(y,x)(s),\qquad \delta_s\cdot x=x\delta_{xsx^{-1}}\\
&\Delta x=\sum_{ab=s}\frac{\phi(x,x^{-1}ax,x^{-1}bx)\phi(a,b,x)}{\phi(a,x,x^{-1}bx)}\\
&\qquad\times x\delta_a\otimes x\delta_b
\end{aligned}
$$

in terms of a basis $\{\delta_s\}$ of $\mathbb{C}(G)$, the product of $G$ on the right, and

$$\chi(x,y)(s)=\frac{\phi(x,y,y^{-1}x^{-1}sxy)\phi(s,x,y)}{\phi(x,x^{-1}sx,y)}$$

a 2-cocycle on $G$ with values in $\mathbb{C}(G)$ (the algebra is a cocycle semidirect product). There is a quasitriangular structure $\mathcal{R} = \sum \delta_x \otimes x$. This quasi-Hopf algebra first appeared in discrete topological quantum field theory related to orbifolds in the work of Dijkgraaf, Pasquier, and Roche.

There are further generalizations in the same spirit and which are linked to conformal field theories of more general type; for example, weak (quasi-) Hopf algebras in which $\Delta 1 \neq 1 \otimes 1$ but is a projector. These have been related to quantum groupoids.

Finally, we mention some applications of twisting outside of the original context. First of all, we are not limited to starting with $U(\mathfrak{g})$: starting with any Hopf algebra or quasi-Hopf algebra $H$ we can similarly twist it to another one $H_F$ with the same algebra as $H$ and $\Delta_F, \mathcal{R}_F, \phi_F$ given by conjugation as above. The representation category remains unchanged up to equivalence, so in some sense the twisted object is equivalent. Moreover, if we start with a Hopf algebra $H$ and ask $F$ to be a 2-cocycle in the sense

$$F_{12}(\Delta \otimes \mathrm{id})(F)(\mathrm{id} \otimes \Delta)(F^{-1})F_{23}^{-1} = 1$$

then $H_F$ will remain a Hopf algebra. It has conjugated antipode (see Majid 1995)

$$S_F(a) = U(Sa)U^{-1}, \qquad U = \cdot(\mathrm{id} \otimes S)(F)$$

Many Hopf algebras are twists of more standard ones, for example, the multiparameter quantum groups tend to be twists of the standard $U_q(\mathfrak{g})$. Likewise, "triangular" Hopf algebras (where $\mathcal{R}_{21}\mathcal{R} = 1$) tend to be twists of classical group or enveloping algebras.

A second application of twists is an approach to quantization. Although it can be applied to $H$ itself, this is more interesting if we think of $H$ as a background quantum group and ask to quantize objects covariant under $H$. For the sake of discussion, we start with $H$ an ordinary Hopf algebra. We twist this to $H_F$ and denote by $\mathcal{T}$ the equivalence functor from representations of $H$ to representations of $H_F$. This functor acts as the identity on all objects and all morphisms, but comes with nontrivial isomorphisms $c_{V,W} : \mathcal{T}(V) \otimes \mathcal{T}(W) \to \mathcal{T}(V \otimes W)$ for any two objects, compatible with bracketting (see Majid 1995). Given any algebraic construction covariant under $H$, we simply apply the functor $\mathcal{T}$ to all aspects of the construction and obtain an equivalent $H_F$-covariant construction. As an example, if $A$ is an $H$-covariant algebra, then applying $\mathcal{T}$ to its product we have $\mathcal{T}(\cdot) : \mathcal{T}(A \otimes A) \to \mathcal{T}(A)$. Using $c_{A,A}$ we obtain a map

$$\bullet : \mathcal{T}(A) \otimes \mathcal{T}(A) \to \mathcal{T}(A)$$
$$a \bullet b = \cdot(F^{-1} \triangleright (a \otimes b))$$

in terms of the product in $A$. Thus, we have a new algebra $A_F$ built on the same vector space as $A$ but with a modified $\bullet$ product. This is called a "covariant twist" of an algebra and should not be confused with the Drinfeld twist above. It is due to the author in the early 1990s. If $F$ is a 2-cocycle, then $A_F$ remains associative. The transmutation construction mentioned in the section "Self-dual quantum groups" or the passage from $\mathbb{R}^4_q$ to $\mathbb{R}^{1,3}_q$ are examples in quantum group theory. Other examples include the standard Moyal product on $\mathbb{R}^n$, also called noncommutative spacetime $[x_\mu, x_\nu] = \mathrm{i}\theta_{\mu\nu}$ by string theorists (*see* Bicrossproduct Hopf Algebras and Noncommutative Spacetime).

If we do not demand that $F$ is a cocycle, then the algebra $A_F$ is still associative but in the target category, which means

$$(a \bullet b) \bullet c = (\bullet(\bullet))\Phi_{A,A,A}((a \otimes b) \otimes c)$$

Such objects are called "quasialgebras." It may still be that $\Phi_{A,A,A}$ happens to be trivial ($\phi_F$ happens to act trivially) so that $A_F$ remains associative. This turns out frequently to be the case and many quantizations in physics, including $\mathbb{C}_q[G]$ but not limited to $q$-examples, can be obtained in this way. It means that although they are associative there is a hidden nonassociativity which can surface in other constructions involving $\Phi$. The physical application here is with $H = U(\mathfrak{g})$ a classical enveloping algebra, $A$ functions on a classical manifold on which $\mathfrak{g}$ acts, and a cochain $F$. In general the resulting quasialgebra will not be associative but rather a quantization of a "quasi-Poisson manifold" obeying

$$\{a, \{b, c\}\} + \text{cylic} = 2\tilde{n}(a \otimes b \otimes c)$$

Here $\tilde{n}$ is the trivector field for the action of the lowest order part of $\phi_F$ and the (quasi)Poisson bivector is the leading-order part of $F_{21}F^{-1}$. As mentioned, there are many cases where $\tilde{n}$ (and the action of the rest of $\phi_F$) happens to be trivial.

Finally, let us give a discrete example using such quantum group methods. We consider $H = \mathbb{C}(G)$ and $F \in \mathbb{C}(G \times G)$ a cochain. Twisting by this gives $H_F = (\mathbb{C}(G), \phi_F)$ a quasi-Hopf algebra where

$$\phi_F(x, y, z) = \frac{F(y, z)F((x, yz)}{F(xy, z)F(x, y)}$$

We take $A = \mathbb{C}G$ the group algebra. The action of $\mathbb{C}(G)$ on it is the diagonal one. The modified algebra $A_F$ therefore has product

$$x \bullet y = F^{-1}(x, y)xy$$

in terms of the product in $G$, and will be a quasialgebra if $F$ is not a cocycle. For example, let

$G = (\mathbb{Z}_2)^3$ which we write additively (so elements are 3-vectors with values in $\mathbb{Z}_2$) and take

$$F(\boldsymbol{x}, \boldsymbol{y}) = (-1)^{\sum_{i<j} x_i y_j + y_1 x_2 x_3 + x_1 y_2 x_3 + x_1 x_2 y_3}$$

then,

$$\phi_F(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = (-1)^{\boldsymbol{x} \cdot (\boldsymbol{y} \times \boldsymbol{z})}$$

Moreover, $A_F = \mathbb{O}$, the octonions (Albuquerque and Majid 1999). So these are a nonassociative quantization of the classical discrete space $(\mathbb{Z}_2)^3$. We see that they are in fact associative up to sign and with sign +1 when the corresponding 3-vectors are linearly independent.

## Noncommutative Geometry

In this article, we have frequently encountered the view of quantum groups and other noncommutative algebras as by definition the coordinate algebras on "noncommutative spaces." However, the "quantum groups approach" to such noncommutative geometry that emerges has a somewhat different flavor from other approaches, as we discuss now.

In fact, the problem of geometry at such a level was mentioned already by Dirac in the 1920s and led to theorems of Gelfand and Naimark in the 1940s and 1950s whereby a noncommutative $C^*$-algebra should be viewed as a noncommutative topological space, and of Serre and Swan in the 1960s whereby a finitely generated projective module should be viewed as a vector bundle. Algebraic $K$-theory led to further refinement of this picture and particularly, in the 1980s, to A Connes' formulation in terms of cyclic cohomology and "spectral triples" (*see* Noncommutative Geometry and the Standard Model; Noncommutative Tori, Yang–Mills and String Theory; Quantum Hall Effect; Hopf Algebra Structure of Renormalizable Quantum Field Theory; Path Integrals in Noncommutative Geometry). The quantum groups approach is less axiomatic, and consists of at least three disparate elements.

The first layer of the quantum groups approach is the theory of $q$-deformed groups and $q$-spaces on which they act, using braided category methods (such as braided linear spaces). The braided group additive law leads to partial derivatives and these define $q$-exterior algebras etc. This programme covered during the 1990s most of what is needed to $q$-deform physics in flat space at an algebraic level. Formulas here tend to be complex but controlled by $R$-matrices, and the correct $R$-matrix formulas can be found systematically by working with braided algebra as explained in the section "Braided groups and quantum planes." From a slightly different side, $q$-representation theory and the further theory of $q$-homogeneous spaces is intimately tied to a theory of $q$-special functions (such as the $q$-exponential function in the section "$q$-Deformation enveloping algebras") of interest in their own right (*see* $q$-Special Functions). The use of $*$-algebras in some cases completable to $C^*$-algebras is a point of contact with other approaches to noncommutative geometry but problems emerge when one considers the braiding. As a result, the natural $q$-Poincaré (plus dilation) quantum group is not even a Hopf $*$-algebra. Briefly, once one starts to braid the constructions, one may need to represent them with braided (not usual) Hilbert spaces and $q$-analysis.

The second layer of the quantum groups approach is based on "differential calculus" as a specification of an exterior algebra of differential forms or differential graded algebra (DGA). In general this is a wild problem but, as in classical geometry, the requirement of a quantum group covariance greatly narrows the possible calculi, although no longer to the point of uniqueness. The first examples of covariant calculi on the quantum group $\mathbb{C}_q[SU_2]$ were found by Woronowicz (1989). The bicovariant one of these was cast in $R$-matrix form by Jurco while the first actual classification results on the moduli of irreducible calculi were obtained by the author (the bicovariant ones are essentially in correspondence with irreducible representations $V$, with left-invariant differentials forming a braided group of the form $B(V \otimes V^*)$). Probably the most interesting feature of this theory is that for all $\mathbb{C}_q[G]$ the bicovariant $q$-calculus cannot be of classical dimensions. For example, for $\mathbb{C}_q[SU_2]$ the smallest nontrivial calculus is four dimensional. The "extra dimension" is a biinvariant 1-form $\theta$ which has the property that $[\theta, a] = \mathrm{d}a$ for all $a \in \mathbb{C}_q[SU_2]$ and which can be viewed as a spontaneously generated time (*see* Bicrossproduct Hopf Algebras and Noncommutative Spacetime). Quantum group methods also provide DGAs on finite groups, this time classified in the bicovariant case by nontrivial conjugacy classes. These therefore provide Lie structures on finite groups. One can go much further and define quantum principal bundles (with quantum groups as fiber) over general noncommutative algebras (Brzezinski and Majid 1993), associated bundles, frame bundles, and Riemannian geometry of the algebra (*see* Quantum Group Differentials, Bundles and Gauge Theory).

Again $q$-deformation provides key examples but the theory may then be applied to other situations.

For example, the permutation group $S_3$ has a natural connected calculus with dimensions $1:3:4:3:1$ (in other words the space has six points but each point has the local structure of a 4-manifold in some sense). It turns out to have a unique Levi-Civita type connection $\nabla$ for its invariant metric, with constant curvature. The use of DGAs here is in common with other approaches (e.g., Connes 1994) and indeed bundles associated to quantum group principal bundles and suitable connections can be shown to be projective modules. The approaches diverge at the level of spectral triples, however, and the examples of "Dirac operators" that emerge from quantum group methods do not usually obey the required axioms.

A third established layer of the quantum groups approach is to trade some of the noncommutativity for nonassociativity, as in the dual version of Drinfeld's construction, that is, $\mathbb{C}_q[G]$ in terms of classical $\mathbb{C}[G]$ as a (co)quasi-Hopf algebra. The general approach here is a quantization functor $\mathcal{T}$ which provides all constructions but which will typically bring out the underlying nonassociative geometry even when the noncommutative covariant algebras of interest is associative. For example, applying the functor to the classical exterior algebra $\Omega(G)$ gives a bicovariant $\Omega(\mathbb{C}_q[G])$ of classical dimensions but with nonassociative products (it is a supercoquasi-Hopf algebra). As before, one may then apply these quantum group methods to other algebras not related to $q$-deformation.

Beyond these are many recent developments, some of which are covered in other articles. Probably one of the most interesting frontiers, at the time of writing, is the exploration of links of both quantum groups and noncommutative geometry to number theory.

*See also:* Affine Quantum Groups; Axiomatic Approach to Topological Quantum Field Theory; Bicrossproduct Hopf Algebras and Noncommutative Spacetime; Braided and Modular Tensor Categories; Classical *r*-Matrices, Lie Bialgebras, and Poisson Lie Groups; Duality in Topological Quantum Field Theory; Eight Vertex and Hard Hexagon Models; Hopf Algebra Structure of Renormalizable Quantum Field Theory; The Jones Polynomial; Noncommutative Geometry and the Standard Model; Noncommutative Tori, Yang–Mills and String Theory; Path Integrals in Noncommutative Geometry; *q*-Special Functions; Quantum Group Differentials, Bundles and Gauge Theory; Quantum Hall Effect; Symmetries in Quantum Field Theory of Lower Spacetime Dimensions; Topological Quantum Field Theory: Overview; von Neumann Algebras: Subfactor Theory; Yang–Baxter Equations.

## Further Reading


Albuquerque H and Majid S (1999) Quasialgebra structure of the octonions. *Journal of Algebra* 220: 188–224.

Baxter RJ (1982) *Exactly Solvable Models in Statistical Mechanics*. New York: Academic Press.

Brzezinski T and Majid S (1993) Quantum group gauge theory on quantum spaces. *Communications in Mathematical Physics* 157: 591–638.

Connes A (1994) *Noncommutative Geometry*. San Diego: Academic Press.

Drinfeld VG (1987) Quantum groups. In: *Proceedings of the ICM*. American Mathematical Society.

Drinfeld VG (1990) Quasi-Hopf algebras. *Leningrad Mathematical Journal* 1: 1419–1457.

Faddeev LD, Reshetikhin NYu, and Takhtajan LA (1990) Quantization of Lie groups and Lie algebras. *Leningrad Mathematical Journal* 1: 193–225.

Jackson FH (1908) On $q$-functions and a certain difference operator. *Transactions of the Royal Society of Edinburgh* 46: 253–281.

Jimbo M (1985) A $q$-difference analog of $U(g)$ and the Yang–Baxter equation. *Letters in Mathematical Physics* 10: 63–69.

Larson RG and Radford DE (1988) Semisimple cosemisimple Hopf algebras. *American Journal of Mathematics* 110: 187–195.

Lusztig G (1993) *Introduction to Quantum Groups*. Boston: Birkhäser.

Majid S (1988) Hopf algebras for physics at the Planck scale. *Journal of Classical and Quantum Gravity* 5: 1587–1606.

Majid S (1990) Quasitriangular Hopf algebras and Yang–Baxter equations. *International Journal of Modern Physics A* 5: 1–91.

Majid S (1991a) Representations, duals and quantum doubles of monoidal categories. *Supplimento di Rendiconti del Circolo di Palermo, Series II* 26: 197–206.

Majid S (1991b) Examples of braided groups and braided matrices. *Journal of Mathematical Physics* 32: 3246–3253.

Majid S (1995) *Foundations of Quantum Group Theory*. Cambridge: Cambridge University Press.

Majid S (2002) *A Quantum Groups Primer,* London Mathematical Society Lecture Notes Series, vol. 291. Cambridge: Cambridge University Press.

Manin YuI (1988) Quantum groups and non-commutative geometry. Technical Report, Centre de Recherches Math, Montreal.

Sweedler ME (1969) *Hopf Algebras*. New York: Benjamin.

Woronowicz SL (1989) Differential calculus on compact matrix pseudogroups (quantum groups). *Communications in Mathematical Physics* 122: 125–170.

# *h*-Pseudodifferential Operators and Applications

**B Helffer**, Université Paris-Sud, Orsay, France



## From Classical Mechanics to Quantum Mechanics

The initial goal of semiclassical mechanics was to explore the correspondence principle, due to N Bohr in 1923, which states that one should recover the classical mechanics from the quantum mechanics as the Planck constant $h$ tends to zero. So we start with a very brief presentation of these two theories.

### Classical Mechanics

We start (with the Hamiltonian formalism) from a $C^\infty$ function $p$ on $\mathbb{R}^{2n}:(x,\xi)\mapsto p(x,\xi)$, which describes the motion of the system under consideration and is called the Hamiltonian. The variable $x$ corresponds, in the simplest case, to the position and $\xi$ to the momentum of one particle. The evolution is then described, starting at time 0 of a given point $(y,\eta)$, by the so-called Hamiltonian equations

$$\begin{aligned}\frac{\mathrm{d}x_j}{\mathrm{d}t} &= (\partial p/\partial\xi_j)(x(t),\xi(t)), \quad \text{for } j=1,\ldots,n\\ \frac{\mathrm{d}\xi_j}{\mathrm{d}t} &= -(\partial p/\partial x_j)(x(t),\xi(t)), \quad \text{for } j=1,\ldots,n\end{aligned} \qquad [1]$$

The classical trajectories are then defined as the integral curves of a vector field defined on $\mathbb{R}^{2n}$ called the Hamiltonian vector field associated with $p$ and defined by $H_p=(\partial p/\partial\xi,\ -\partial p/\partial x)$. All these definitions are more generally relevant in the framework of symplectic geometry on a symplectic manifold $M$ (but we choose, for simplicity, to explain the theory on $\mathbb{R}^{2n}$), which can be seen as the cotangent vector bundle $T^*\mathbb{R}^n$, and is the "local" model of the general situation. This space is equipped naturally with a symplectic structure defined by giving at each point a nondegenerate 2-form, which is here $\sigma:=\sum_j \mathrm{d}\xi_j\wedge\mathrm{d}x_j$. This 2-form permits us to associate canonically to a 1-form on $T^*\mathbb{R}^n_x$ a vector field on $T^*\mathbb{R}^n_x$. In this correspondence, if $p$ is a function on $T^*\mathbb{R}^n_x$, $H_p$ is associated with the differential $\mathrm{d}p$.

In this article, we consider the example of the Hamiltonian $p(x,\xi)=\xi^2+V(x)$, also called the Schrödinger Hamiltonian, as the guiding example. More specifically, the case of the harmonic oscillator, where $V$ is given by $V(x)=\sum_{j=1}^n \mu_j x_j^2$ (with $\mu_j>0$), is the most significant, which is the natural approximation of a potential near its minimum, when nondegenerate.

In the framework of the classical mechanics, the main questions could be:

- Are the trajectories bounded?
- Are there periodic trajectories?
- Is one trajectory dense in its energy surface?
- Is the energy surface compact?

The solution of these questions could be very difficult. Let us just mention the trivial fact that, if $p^{-1}(\lambda)$ is compact for some $\lambda$, then, by the conservation of energy law

$$p(x(t),y(t))=p(y,\eta) \qquad [2]$$

the whole trajectory starting of one point $(y,\eta)$ remains in the bounded set $\{p^{-1}(p(y,\eta))\}$ in $\mathbb{R}^{2n}$. This is in particular the case for the harmonic oscillator.

### Quantum Mechanics

The quantum theory was born dynamics-wise around 1920. It is structurally related to the classical mechanics in a way that we shall describe very briefly. In quantum mechanics, our basic object will be a (possibly nonbounded) self-adjoint operator defined on a dense subspace of a Hilbert space $\mathcal{H}$. In order to simplify the presentation, we shall always take $\mathcal{H}=L^2(\mathbb{R}^n)$.

This operator can be associated with $p$ by using the techniques of quantization. We choose here to present a procedure, called the Weyl quantization procedure (which was already known in 1928), which under suitable assumptions on $p$ and its derivatives, will be defined for $u\in\mathcal{S}(\mathbb{R}^n)$ by

$$\begin{aligned}p^{\mathrm{w}}(x,hD_x,h)u(x) &= (2\pi h)^{-n}\iint \exp\left(\frac{\mathrm{i}}{h}(x-y)\cdot\xi\right)\\ &\quad\times p\left(\frac{x+y}{2},\xi,h\right)u(y)\,\mathrm{d}y\,\mathrm{d}\xi\end{aligned} \qquad [3]$$

The operator $p^{\mathrm{w}}(x,hD_x,h)$ is called an $h$-pseudodifferential operator of Weyl symbol $p$. One can also write $\mathrm{Op}_h^{\mathrm{w}}(p)$ in order to emphasize that it is the operator associated to $p$ by the Weyl quantization. Here $h$ is a parameter which plays the role of the Planck constant.

Of course, one has to give a sense to these integrals and this is the object of the theory of the oscillatory integrals. If $p=1$, we observe that, by Plancherel's formula,

$$\begin{aligned}u(x) &= (2\pi h)^{-n}\iint \exp\left(\frac{\mathrm{i}}{h}(x-y)\cdot\xi\right)\\ &\quad\times u(y)\,\mathrm{d}y\,\mathrm{d}\xi\end{aligned} \qquad [4]$$

the associated operator is nothing but the identity operator. A way to rewrite any $h$-differential operator $\sum_{|\alpha|\le m} a_\alpha(x)(hD_x)^\alpha$ as an $h$-pseudodifferential operator is to apply it on both sides to [4]. In particular, we observe that if the symbol is $p(x,\xi) = \xi^2 + V(x)$, then the operator associated with $p$ by the $h$-Weyl quantization is the Schrödinger operator $-h^2\Delta + V$. Other interesting examples appear naturally in solid state physics. Let us, for example, mention the Harper operator $H$ (or almost-Mathieu; see Helffer and Sjöstrand (1989) and references therein), whose symbol is the map $(x,\xi) \mapsto \cos\xi + \cos x$, and which can also be defined, for $u \in L^2(\mathbb{R})$, by

$$(Hu)(x) = \tfrac{1}{2}(u(x+h) + u(x-h)) + \cos x\, u(x)$$

We shall later recall how to relate the properties of $p$ and those of the associated operator. More precisely, we shall describe under which conditions on $p$ the operator $p^{\mathrm{w}}(x, hD_x; h)$ is semibounded, symmetric, essentially self-adjoint, compact, with compact resolvent, trace class, Hilbert–Schmidt (see Robert (1987) for an extensive presentation). But before looking at a more general situation, let us consider the case of the Schrödinger operator $S_h = -h^2\Delta + V(x)$. If $V$ is (say, continuous) bounded from below, $S_h$, which is *a priori* defined on $\mathcal{S}(\mathbb{R}^n)$ as a differential operator, admits a unique self-adjoint extension on $L^2(\mathbb{R}^n)$. We are first interested in the nature of the spectrum. If $V(x) \to +\infty$ as $|x| \to \infty$, one can show that $S_h$, more precisely its self-adjoint realization, has compact resolvent and its spectrum consists of a sequence of eigenvalues tending to $\infty$. We are next interested in the asymptotic behavior of these eigenvalues.

In the case of the harmonic operator, corresponding to the potential

$$V(x) = \sum_{j=1}^{n} \mu_j x_j^2 \quad (\text{with } \mu_j > 0)$$

the criterion of compact resolvent is satisfied and the spectrum is described as the set of

$$\lambda_\alpha(h) = \sum_{j=1}^{n} \sqrt{\mu_j}(2\alpha_j + 1)h$$

for $\alpha \in \mathbb{N}^n$.

In this case we also have a complete description of the normalized associated eigenfunctions which are constructed recursively starting from the first eigenfunction corresponding to $\lambda_0(h) = \sum_j \sqrt{\mu_j} h$:

$$\phi_0(x;h) = \left(\prod_{j=1}^{n}\left(\frac{\sqrt{\mu_j}}{\pi h}\right)^{1/4}\right) \cdot \exp\left(-\frac{1}{2h}\sum_j \sqrt{\mu_j} x_j^2\right) \quad [5]$$

The eigenfunction $\phi_0$ is strictly positive and decays exponentially. Moreover (and here we enter in the semiclassical world), the local decay in a fixed closed set avoiding $\{0\}$ (which is measured by its $L^2$-norm) is exponentially small as $h \to 0$. In particular, this says that the eigenfunction lives asymptotically in the set $\{V(x) \le \lambda(h)\}$. This last set can also be understood as the projection by the map $(x,\xi) \mapsto x$ of the energy surface, which is classically attached to the eigenvalue $\lambda(h)$, that is, $\{(x,\xi) \in \mathbb{R}^{2n} \mid p(x,\xi) = \lambda(h)\}$. This is a typical semiclassical statement, which will be true in full generality.

### From Quantum Mechanics to Classical Mechanics: Semiclassical Mechanics

Before describing the mathematical tools involved in the exploration of the correspondence principle, let us describe a few results which are typical in the semiclassical context. They concern Weyl's asymptotics and the localization of the eigenfunctions.

**Weyl's asymptotics** We start with the case of the Schrödinger operator $S_h$, but we emphasize that the $h$-pseudodifferential techniques are not limited to this situation.

We assume that $V$ is a $C^\infty$-function on $\mathbb{R}^n$ which is semibounded and satisfies

$$\inf V < \underline{\lim}_{|x|\to\infty} V(x)$$

The Weyl theorem (which is a basic theorem in spectral theory) implies that the essential spectrum is contained in

$$\left[\underline{\lim}_{|x|\to\infty} V(x), +\infty\right]$$

It is also clear that the spectrum is contained in $[\inf V, +\infty]$. In the interval

$$I = \left[\inf V,\ \underline{\lim}_{|x|\to\infty} V(x)\right]$$

the spectrum is discrete, that is, it has only isolated eigenvalues with finite multiplicity. For any $E$ in $I$, it is consequently interesting to look at the counting function $N_h(E)$ of the eigenvalues contained in $[\inf V, E]$,

$$N_h(E) = \sharp\{\lambda_j(h); \lambda_j(h) \le E\} \quad [6]$$

The main semiclassical result is then

**Theorem 1** *With the previous assumptions, we have:*

$$\lim_{h\to 0} h^n N_h(E) = (2\pi)^{-n} \int_{V(x)\le E} (E - V(x))^{n/2}\, dx$$

The main term in the expansion of $N_h(E)$, which will be denoted by

$$W_h(E) := (2\pi h)^{-n} \int_{V(x)\le E} (E - V(x))^{n/2}\, dx$$

is called the Weyl term. It has an analog for the analysis of the counting function for Laplacians on compact manifolds (*see* Quantum Ergodicity and Mixing of Eigenfunctions and references therein), but let us emphasize that here $E$ is fixed and that one looks at the asymptotics as $h \to 0$. In the other case, $h$ is fixed and one looks at the asymptotics as $E \to +\infty$ (note that on a compact manifold and for the Laplacian, the formula $N_h(E) = N_1(E/h^2)$ permits switching between these cases).

Although this formula is rather old (first as a folk theorem), many efforts have been made by mathematicians for analyzing the remainder (see Robert (1987), Ivrii (1998) and references therein) $N_h(E) - W_h(E)$, whose behavior is again related to classical analysis. When $E$ is not a critical value of $V$, $h^{n+1}(N_h(E) - W_h(E))$ can be shown to be bounded but it appears to be $o(1)$ if the measure of the periodic points for the flow is 0 (see Ivrii (1998)).

Beyond the analysis of the counting function, one is also interested (e.g., in questions concerning the ground-state energy of an atom with a large number of particles, $N$, satisfying the Pauli exclusion principle (*see* Stability of Matter)) in other quantities like the Riesz means, which are defined, for a given $s \ge 0$, by

$$N_h^s(E) = \sum_j (E - \lambda_j)_+^s$$

The case $s = 0$ corresponds to the counting function. It is then natural to ask for the asymptotic behavior as $h \to 0$ of these functions.

We have, for example, the following result (Helffer–Robert, Ivrii–Sigal, and Ivrii; see Robert (1987) and Ivrii (1998)), which is written here in a more Hamiltonian version, when $E$ is not a critical value of $V$,

$$N_h^s(E) = (2\pi h)^{-n} \left( \int_{p_E(x,\xi)\le 0} (-p_E(x,\xi))^s\, dx \cdot d\xi \right) + \mathcal{O}(h^{\inf(1+s,2)})$$

with $p_E(x, \xi) = \xi^2 + V(x) - E$.

**Uncertainty principle and Weyl term** The Weyl term can be heuristically understood in the following way. According to the uncertainty principle, a "quantum" particle should occupy at least a volume of order $h^n$ in the phase space with the measure $dx\, d\xi$ (proportional to $(\sum_{j=1}^n d\xi_j \wedge dx_j)^n$). This guess is a consequence of the inequality

$$\frac{h}{2}\|u\|^2 \le \left( \int_{\mathbb{R}} (x - x_0)^2 |u|^2\, dx \right)^{1/2} \times \left( \int_{\mathbb{R}} \left| \left( \frac{h}{i}\frac{d}{dx} - \xi_0 \right) u \right|^2 dx \right)^{1/2}, \quad \forall u \in \mathcal{S}(\mathbb{R})$$

expressing the noncommutation of the operators $((h/i)d/dx - \xi_0)$ and (multiplication by) $(x - x_0)$. When $\|u\| = 1$ and $x_0$ (mean position) and $\xi_0$ (mean momentum) are defined by $x_0 := \int_{\mathbb{R}} x|u|^2\, dx$ and $\xi_0 := (h/i)\int_{\mathbb{R}} u'(x)\cdot \bar{u}(x)\, dx$, this inequality expresses the impossibility for a quantum particle to have a simultaneous small localization in position and momentum.

Consequently, the maximal number of "quantum" particles which can live in the region $\{p_E(x, \xi) \le 0\}$ is approximately (up to some universal multiplicative constant) the volume of this region divided by $(2\pi h)^n$.

**Lieb–Thirring inequalities and Scott's conjecture** In the case of regular potentials, we have seen that the quantity $h^n N_h^s(E)$ was asymptotically equal as $h \to 0$ to $L^{\rm cl}_{s,n}(\int_{V(x)\le E} (E - V(x))^{s+n/2}\, dx)$. For other questions occurring in atomic physics (*see* Stability of Matter), one is more interested in the existence of universal constants $M_{s,n}$ such that

$$h^n N_h^s(E) \le M_{s,n} \left( \int_{V(x)\le E} (E - V(x))^{s+n/2}\, dx \right)$$

for any $V$ and any $h$.

The best $M_{s,n}$ (which exists if $s + n/2 > 0$) is denoted by $L_{s,n}$ (for $s = 0$; this is called the Cwickel–Lieb–Rozenblium inequality). The semiclassical result gives the inequality $L_{s,n} \ge L^{\rm cl}_{s,n}$.

A still open question is the so-called Lieb–Thirring conjecture: do we have $L_{1,3} = L^{\rm cl}_{1,3}$? This is related to the question of the stability of the matter (*see* Stability of Matter). The last results in this direction have been obtained quite recently by A Laptev and T Weidl, who show, for example, the equality for $s \ge 3/2$.

The control, when $s = 1$, of a second term (for more singular potentials) for $N_h^s(E)$ was the object of the Scott conjecture, which was solved recently in many important cases by Hughes, Siedentop–Weikard, Ivrii–Sigal, and Feffermann–Secco (see Ivrii (1998), Stability of Matter, and references therein).

**Localization of the eigenfunctions** The localization property was already observed on the specific case of the harmonic oscillator. But this was a consequence

of an explicit description of the eigenfunctions. This is quite important to have a good description of the decay of the eigenfunctions (as $h \to 0$) outside the classically permitted region without having to know an explicit formula. Various approaches can be used.

The first one fits very well in the case of the Schrödinger operator (more generally to $h$-pseudodifferential operators with symbols admitting holomorphic extensions in the $\xi$ variable) and gives exponential decay. This is based on the so-called Agmon estimates (developed in the semiclassical context by Helffer–Sjöstrand and Simon). We shall not say more about this approach, which is the starting point of the analysis of the tunneling (see Helffer (1988), Dimassi and Sjöstrand (1999), and Martinez (2002)).

The second one is an elementary application of the $h$-pseudodifferential formalism which will be described later and leads, for example, to the following statement. Let $E$ in $I$ and let $(\lambda(h_j), \phi_{(h_j)}(x))$ be a sequence of spectral pairs in $I \times L^2(\mathbb{R}^n)$, where $h_j \to 0$ as $j \to +\infty, \lambda(h_j) \to E$, and $x \mapsto \phi_{(h_j)}(x)$ is an $L^2$-normalized eigenfunction associated with $\lambda(h_j)$. Let $\Omega$ be a relatively compact set in $\mathbb{R}^n$ such that

$$V^{-1}(]-\infty, E]) \cap \bar{\Omega} = \emptyset$$

Then, there exists, for all integer $N$, a constant $C_{N,\Omega}$ such that

$$\|\phi_{(h_j)}\|_{L^2(\Omega)} \leq C_{N,\Omega} \cdot h_j^N$$

A third one uses the notion of frequency set and will be discussed later (see also the book of Martinez (2002) for what can be done with the Fourier–Bros–Iagolnitzer transform as developed by J Sjöstrand).

## Brief Introduction to the *h*-Pseudodifferential Calculus

For fixed $h$, the pseudodifferential calculus has a long story starting in its modern form in the 1960s. A rather achieved version of the calculus is presented in Hörmander (1984). We will emphasize here on the semiclassical aspect of the calculus, that is, on the dependence of the calculus on the parameter $h > 0$.

### *h*-Pseudodifferential Calculus

**Basic calculus: the class $S^0$** We shall mainly discuss the most simple one called the $S^0$ calculus. Let us first say that the $S^0$ calculus is sufficient once we have suitably (micro)-localized the problem (e.g., by the functional calculus). Note that it is also sufficient for the local analysis of many problems occurring on compact manifolds.

This class of symbols $p$ is simply defined by the conditions:

$$|\partial_x^\alpha \partial_\xi^\beta p(x, \xi)| \leq C_{\alpha,\beta} \qquad [7]$$

for all $(\alpha, \beta) \in \mathbb{N}^n \times \mathbb{N}^n$. The symbols can possibly be $h$-dependent. With this symbol, one can associate an $h$-pseudodifferential operator by [3]. This operator is a continuous operator on $\mathcal{S}(\mathbb{R}^n)$ but can also be defined by duality on $\mathcal{S}'(\mathbb{R}^n)$.

The first basic analytical result is the Calderon–Vaillancourt theorem (see Hörmander (1984)) establishing the $L^2$-continuity. We also mention that if $p$ is in $L^2(\mathbb{R}^{2n})$, the associated operator is Hilbert–Schmidt. One can also give conditions on $p$ implying the trace-class property (replace the uniform control in [7] by a control in $L^1$).

The second important property is the existence of a calculus. If $a$ is in $S^0$ and $b$ is in $S^0$ then the composition $a^{\mathrm{w}}(x, hD_x) \circ b^{\mathrm{w}}(x, hD_x)$ of the two operators is a pseudodifferential operator associated with an $h$-dependent symbol $c$ in $S^0$:

$$a^{\mathrm{w}}(x, hD_x) \circ b^{\mathrm{w}}(x, hD_x) = c^{\mathrm{w}}(x, hD_x; h)$$

We see here that we immediately meet symbols admitting expansions in powers of $h$, which we shall call regular symbols, in the sense that they admit expansions of the type

$$a(x, \xi; h) \sim \sum_j a_j(x, \xi) h^j$$

$$b(x, \xi; h) \sim \sum_j b_j(x, \xi) h^j$$

In this case the Weyl symbol $c$ of the composition has a similar expansion:

$$c(x, \xi; h) \sim \left[ \exp\left( \frac{\mathrm{i}h}{2} (D_x \cdot D_\eta - D_y \cdot D_\xi) \right) \times (a(x, \xi; h) \cdot b(y, \eta; h)) \right]_{x=y;\ \xi=\eta}$$

The symbol $a_0$ is called the principal symbol. At the level of principal symbols, the rule is simply that the principal symbol of $a^{\mathrm{w}} \circ b^{\mathrm{w}}$ is the product of the principal symbols of $a^{\mathrm{w}}$ and $b^{\mathrm{w}}$: $c_0 = a_0 \cdot b_0$. Another important property is the following correspondence between commutator of two operators and Poisson brackets. The principal symbol of the commutator $(1/h)(a^{\mathrm{w}} \circ b^{\mathrm{w}} - b^{\mathrm{w}} \circ a^{\mathrm{w}})$ is $(1/\mathrm{i})\{a_0, b_0\}$, where $\{f, g\}$ is the Poisson bracket of $f$ and $g$:

$$\{f, g\}(x, \xi) = H_f g = \sum_j \left( \partial_{\xi_j} f \cdot \partial_{x_j} g - \partial_{x_j} f \cdot \partial_{\xi_j} g \right)$$

**About global classes** The class $S^0$ is far from being sufficient for analyzing the global spectral problem and we refer the reader to Hörmander (1984) or Robert (1987) for an extensive presentation of the theory and for the discussion of other quantizations. Our initial operators (think of the harmonic oscillator) do not belong to these classes of pseudodifferential operators. We are consequently obliged to construct more general classes including these examples in order to realize this localization. Once such a class is introduced, one of the main points to consider is the existence of a quasi-inverse (or parametrix) for a suitably defined elliptic operator of positive order. Following Beals–Feffermann (see also the most general Hörmander calculus in Hörmander (1984) and references therein), we introduce a scale function (possibly $h$-dependent; typically, $m(x,\xi;h) = h^{\nu} m_0(x,\xi)$) $(x,\xi) \mapsto m(x,\xi;h)$ and $C^\infty$ strictly positive weight functions $\phi$ and $\Phi$ such that $\phi \cdot \Phi \geq 1$. All these functions are strictly positive and should satisfy additional conditions on their variation and growth. The class of symbols $S^{\mathrm{reg}}(m,\phi,\Phi)$ is defined by

$$|D_x^\alpha D_\xi^\beta p(x,\xi;h)| \leq C_{\alpha,\beta}\, m(x,\xi;h)\phi(x,\xi)^{-|\alpha|}\Phi(x,\xi)^{-|\beta|}$$

These apparently complicated estimates permit actually the control of the variation of the symbol in reference balls defined by

$$\phi^{-2}(x_0,\xi_0)|x-x_0|^2 + \Phi^{-2}(x_0,\xi_0)|\xi-\xi_0|^2 \leq c$$

**Elliptic theory** As noted above, the main point is to have a large class of invertible operators, such that the inverses are also in the class. This is what we call an elliptic theory and the typical statement is:

**Theorem 2** *Let P be an h-pseudodifferential operator associated with a symbol p in* $S^{\mathrm{reg}}(m,\phi,\Phi)$. *We assume that it is elliptic in the sense that* $1/p$ *belongs to* $S^{\mathrm{reg}}(1/m,\phi,\Phi)$. *Then there exists an h-pseudodifferential operator Q with symbol in* $S^{\mathrm{reg}}(1/m,\phi,\Phi)$, *such that*

$$QP = I + R; \qquad PQ = I + S$$

*The remainders R and S are pseudodifferential operators with symbols in*

$$\bigcap_N S\left(\left(\frac{h}{\Phi\cdot\phi}\right)^N, \phi, \Phi\right)$$

These remainders are called "regularizing." Note that this notion depends strongly on the choice of the class of pseudodifferential operators! When $\phi = \Phi = 1$, we are just inverting modulo a remainder whose norm in $\mathcal{L}(L^2)$ is $\mathcal{O}(h^\infty)$ (or simply $\mathcal{O}(h)$ at the first step). With other weights like $\phi = \Phi = \sqrt{1+|x|^2+|\xi|^2}$, we invert $P$ modulo a remainder, which has, in addition, a distribution kernel in the Schwartz space $\mathcal{S}(\mathbb{R}^n \times \mathbb{R}^n)$. The invertibility modulo a compact operator (which implies the Fredholm property) is a consequence of the assumption

$$\lim_{|x|+|\xi|\to+\infty} \phi(x,\xi)\Phi(x,\xi) = +\infty$$

The proof is rather easy, once the formalism of composition and the notion of principal symbol have been understood. One can indeed start from the operator $Q_0$ of symbol $1/p$ and observe that $Q_0P = I + R_1$ holds, with $R_1$ in $\mathrm{Op}^{\mathrm{w}}(S((h/\Phi\cdot\phi),\phi,\Phi))$. The operator $(I+R_1)^{-1}Q_0 \sim (\sum_{j\geq 0}(-1)^j R_1^j)Q_0$ gives essentially the solution.

## Essential Self-Adjointness and Semiboundedness

We now sketch two applications of this calculus in spectral theory. We shall usually consider in our applications an $h$-pseudodifferential operator $P$, whose Weyl symbol $p$ is regular, that is, admitting an asymptotic expansion:

$$\text{(H0)} \quad p(x,\xi;h) \sim \sum_{j\geq 0} h^j p_j(x,\xi)$$

(We refer to Robert (1987), Hörmander (1984), and Dimassi and Sjöstrand (1999) for a more precise formulation). Moreover, we assume that

$$\text{(H1)} \quad (x,\xi) \mapsto p(x,\xi;h) \in \mathbb{R}$$

This implies, as can be immediately seen from [3], that $p^{\mathrm{w}}$ is symmetric (= formally self-adjoint):

$$\langle p^{\mathrm{w}}u, v\rangle_{L^2} = \langle u, p^{\mathrm{w}}v\rangle_{L^2}, \quad \forall u,v \in \mathcal{S}(\mathbb{R}^n)$$

The third assumption is that the principal symbol is bounded from below (and there is no restriction to assume that it is positive)

$$\text{(H2)} \quad p_0(x,\xi) \geq 0$$

This assumption implies that the operator itself is bounded from below. This result belongs to the family of the so-called Garding inequalities. More precisely, the assumption (there are other quantizations, e.g., the anti-Wick quantization, for which this result becomes trivial, the difference between the two quantizations being $\mathcal{O}(h)$) will basically give, if $m \geq 1$, the existence of a constant $C$ such that, for any $u \in \mathcal{S}(\mathbb{R}^n)$,

$$\langle Pu, u\rangle_{L^2\times L^2} \geq -C\, h||u||^2$$

Everything is proved if $m(x,\xi)=(p_0+1)$ is a scale function, if $p_j$ and their derivatives are controlled by $(p_0+1)$:

$$\text{(H3)}\quad |\partial_x^\alpha \partial_\xi^\beta p_j(x,\xi)| \le C_{\alpha,\beta,j}(p_0+1)\phi(x,\xi)^{-j-|\alpha|} \times \Phi(x,\xi)^{-j-|\beta|}$$

for all $(\alpha,\beta)\in\mathbb{N}^n\times\mathbb{N}^n$, and if there is a suitable control of the family $(N\in\mathbb{N})$ of symbols

$$\left(\frac{\phi\Phi}{(p_0+1)h}\right)^{(N+1)}\left(p-\sum_{j=0}^{N} p_j h^j\right)$$

Under these assumptions, the main result is that $P$ is, for $h$ small enough, essentially self-adjoint. This means that the operator which was initially defined on $\mathcal{S}(\mathbb{R}^n)$ by the pseudodifferential operator with symbol $p$ admits a unique self-adjoint extension.

### The Functional Calculus

It is well known by the spectral theorem for a self-adjoint operator $P$ that a functional calculus exists for Borel functions. What is important here is to find a class of functions (actually essentially $C_0^\infty$) such that $f(P)$ is a pseudodifferential operator in the same class as $P$ with simple rules of computation for the principal symbol.

We are starting from the general formula (see Dimassi and Sjöstrand (1999))

$$f(P) = -\pi^{-1}\lim_{\epsilon\to 0^+}\iint_{|\mathrm{Im}\,z|\ge\epsilon}\frac{\partial\tilde{f}}{\partial\bar{z}}(x,y)(z-P)^{-1}\,\mathrm{d}x\,\mathrm{d}y$$

which is true for any self-adjoint operator and any $f$ in $C_0^\infty(\mathbb{R})$. Here the function $(x,y)\mapsto\tilde{f}(x,y)$ (note that $z=x+\mathrm{i}y$) is a compactly supported, almost analytic extension of $f$ to C. This means that $\tilde{f}=f$ on $\mathbb{R}$ and that for any $N\in\mathbb{N}$ there exists a constant $C_N$ such that

$$\left|\frac{\partial\tilde{f}}{\partial\bar{z}}(z)\right|\le C_N\left|\mathrm{Im}\,z\right|^N$$

The main result due to Helffer–Robert (see also Dimassi and Sjöstrand (1999) and references therein) is that, for $P$ an $h$-regular pseudodifferential operator satisfying (H0)–(H3) and $f$ in $C_0^\infty(\mathbb{R})$, the operator $f(P)$ is an $h$-pseudodifferential operator, whose Weyl symbol $p_f(x,\xi;h)$ admits a formal expansion in powers of $h$:

$$p_f(x,\xi;h)\sim\sum_{j\ge 0}h^j p_{f,j}(x,\xi)$$

with

$$p_{f,0}=f(p_0)$$
$$p_{f,1}=p_1\cdot f'(p_0)$$
$$p_{f,j}=\sum_{k=1}^{2j-1}(-1)^k(k!)^{-1}d_{j,k}f^{(k)}(p_0),\quad\forall j\ge 2$$

where the $d_{j,k}$ are universal polynomial functions of the symbols $\partial_x^\alpha\partial_\xi^\beta p_\ell$, with $|\alpha|+|\beta|+\ell\le j$.

The main point in the proof is that we can construct, for $\mathrm{Im}\,z\ne 0$, a parametrix (= approximate inverse) for $(P-z)$ with a nice control as $\mathrm{Im}\,z\to 0$. The constants controlling the estimates on the symbols are exploding as $\mathrm{Im}\,z\to 0$ but the choice of the almost analytic extension of $f$ absorbs any negative power of $|\mathrm{Im}\,z|$.

As a consequence, we get that if, for some interval $I$ and some $\epsilon_0>0$,

$$\text{(H4)}\quad p_0^{-1}(I+[-\epsilon_0,\epsilon_0])\text{ is compact}$$

then the spectrum is, for $h$ small enough, discrete in $I$.

In particular, we get that, if $p_0(x,\xi)\to+\infty$ as $|x|+|\xi|\to+\infty$, then the spectrum of $P_h$ is discrete ($P_h$ has compact resolvent). Under the assumption (H4), we get more precisely the following theorem.

**Theorem 3** *Let $P$ be an $h$-regular pseudodifferential operator satisfying (H0)–(H4), with $I=[E_1,E_2]$, then, for any $g$ in $C_0^\infty([E_1,E_2])$, we have the following expansion in powers of $h$:*

$$\mathrm{tr}[g(P(h))]\sim h^{-n}\sum_{j\ge 0}h^j T_j(g),\quad\text{as } h\to 0$$

*where $g\mapsto T_j(g)$ are distributions in $\mathcal{D}'(]E_1,E_2[)$. In particular, we have*

$$T_0(g)=(2\pi)^{-n}\iint g(p_0(x,\xi))\,\mathrm{d}x\,\mathrm{d}\xi$$
$$T_1(g)=(2\pi)^{-n}\iint g'(p_0(x,\xi))p_1(x,\xi)\,\mathrm{d}x\,\mathrm{d}\xi$$

This theorem is just obtained by integration of the preceding one, because in these cases the trace of a trace-class pseudo-differential operator $\mathrm{Op}^{\mathrm{w}}(a)$ is given by the integral of the symbol $a$ over $\mathbb{R}^{2n}=\mathbb{R}_x^n\times\mathbb{R}_\xi^n$. According to [3], the distribution kernel is given by the oscillatory integral:

$$K(x,y;h)=(2\pi h)^{-n}\int_{\mathbb{R}^n}\exp\left(\frac{\mathrm{i}}{h}(x-y)\cdot\xi\right)\times a\left(\frac{x+y}{2},\xi;h\right)\mathrm{d}\xi\qquad[8]$$

and the trace of $\mathrm{Op}^{\mathrm{w}}(a)$ is the integral over $\mathbb{R}^n$ of the restriction to the diagonal of the distribution kernel:

$$K(x,x) = (2\pi h)^{-n} \int_{\mathbb{R}^n} a(x,\xi;h)\,\mathrm{d}\xi$$

Of course, one could think of using the theorem with $g$, the characteristic function of an interval, in order to get, for example, the behavior of the counting function attached to this interval. This is of course not directly possible and this will be obtained only through Tauberian theorems (Hörmander (1968), (1984), Ivrii (1998)) and at the price of additional errors.

Let us, however, remark that, if the function $g$ is not regular, then the length of the expansion depends on the regularity of $g$. So it will not be surprising that, by looking at the Riesz means, we shall get a better expansion when $s$ is large.

Anyway, one basic interest of functional calculus is to permit a localization in the energy of the operator. For a general $h$-pseudodifferential operator, it could be difficult to approximate an operator like $\exp(-\mathrm{i}tP/h)$ by suitable Fourier integral operators but approximate $\exp(-\mathrm{i}tP/h)f(P)$ for suitable compactly supported $f$ could be easier.

Another interest is that for suitable $f$ (possibly $h$-dependent) the operator $f(P)$ could have better properties than the initial operator. This idea will, for example, be applied for the theorem concerning clustering. It appears, in particular, very powerful in dimension 1, where we can in some interval of energy find a function $t \mapsto f(t;h)$ admitting an expansion in powers of $h$ such that $f(P;h)$ has the spectrum of the harmonic oscillator. This is a way to get the Bohr–Sommerfeld conditions (see Helffer–Robert (1987), together with Maslov (1972), Leray (1981), or the thesis of A Voros in 1977), which reads:

$$f(\lambda_n(h);h) \sim (2n+1)h \quad \text{modulo } \mathcal{O}(h^\infty)$$

## *h*-Fourier Integral Operators and Evolution Operators

### Classical Mechanics

Let us come back to the Hamilton equations [1]. The local existence of solutions is well known. If, in addition, we assume (H4), the energy conservation law implies global existence for these solutions, if the initial data $(y,\eta)$ belong to $p^{-1}(I)$.

We recall that $(y,\eta) \mapsto \phi^t(y,\eta) = (x(t,y,\eta),\xi(t,y,\eta))$ defines for any $t$ a canonical transformation, that is, a diffeomorphism respecting the symplectic 2-form:

$$\sigma = \sum_j \mathrm{d}\xi_j \wedge \mathrm{d}x_j$$

We shall denote by $\Lambda^t$ the graph of $\phi^t$ which is a Lagrangian submanifold (which means that at each point $m$ of the manifold the restriction of the symplectic two-form to $T_m\Lambda^t$ is 0) for the 2-form on $\mathbb{R}^{2n} \times \mathbb{R}^{2n}$: $\sum_j \mathrm{d}\eta_j \wedge \mathrm{d}y_j - \sum_j \mathrm{d}\xi_j \wedge \mathrm{d}x_j$.

When the projection $(y,\eta,x,\xi) \mapsto (\eta,x)$ gives a local system of coordinates for $\Lambda_t$ (and this will always be the case for $(\eta,x)$ in a compact set and $t$ small enough), one easily finds, using the Lagrangian character of $\Lambda^t$, a function $(\eta,x) \mapsto S_t(x,\eta)$ such that

$$\Lambda^t = \{y,\eta,x,\xi \mid y = \partial_\eta S_t,\ \xi = \partial_x S_t\}$$

This function is only defined modulo an arbitrary function of $t$. In order to get a more natural choice, we consider the Lagrangian submanifold in $\mathbb{R}^{2n} \times \mathbb{R}^{2n} \times \mathbb{R}^2$ defined as

$$\Lambda = \{y,\eta,x,\xi,t,\tau | (x,\xi) = \phi^t(y,\eta),\ \tau = -p_0(x,\xi)\} \quad [9]$$

The parametrization of $\Lambda$, by its projection $(y,\eta,x,\xi,t,\tau) \mapsto (\eta,x,t)$, will now give a natural function $(\eta,x,t) \mapsto S(t,x,\eta) = S_t(x,\eta)$ describing $\Lambda$ by

$$\Lambda = \{y,\eta,x,\xi,t,\tau | \xi = \partial_x S,\ y = \partial_\eta S,\ \tau = \partial_t S\} \quad [10]$$

We observe that we can choose

$$S(0,x,\eta) = x \cdot \eta \quad [11]$$

and that $S$ is automatically a solution of the Hamilton–Jacobi equation

$$(\partial_t S)(t,x,\eta) + p_0(x,\partial_x S(t,x,\eta)) = 0 \quad [12]$$

also called the eiconal equation.

We also observe the following property (by comparison of [9] and [10]):

$$\phi^t(\partial_\eta S(t,x,\eta),\eta) = (x,\partial_x S(t,x,\eta))$$

We have actually an explicit expression of $S(t,x,\eta)$ in term of the inverse $y(t,x,\eta)$ of the map $y \mapsto x(t,y,\eta)$:

$$\begin{aligned} S(t,x,\eta) &= y(t,x,\eta)\cdot\eta \\ &\quad + \int_0^t \Big[\sum_i \xi_i(s,y,\eta)\cdot(\partial_{\xi_i}p)(x(s,y,\eta),\xi(s,y,\eta)) \\ &\quad - p(y,\eta)\Big]\,\mathrm{d}s_{/y=y(t,x,\eta)} \end{aligned}$$

For the harmonic oscillator, easy computations give

$$p(x,\xi) = \tfrac{1}{2}(\xi^2 + x^2), \qquad H_p = (\xi,-x)$$
$$\phi^t(y,\eta) = (y\cos t + \eta\sin t,\ -y\sin t + \eta\cos t)$$

and

$$S(t,x,\eta) = -\tfrac{1}{2}(x^2+\eta^2)\cdot(\tan t) + \frac{x\cdot\eta}{\cos t}$$

### Fourier Integral Operators

We have already given in [8] the distribution kernel of an $h$-pseudodifferential operator. It appears useful to generalize this point of view by considering more generally objects defined similarly as

$$K(x,y;h) = (2\pi h)^{-r} \int_{\mathbb{R}^N} \exp\left(\frac{\mathrm{i}}{h}\phi(x,y,\theta)\right) a(x,y,\theta;h)\,\mathrm{d}\theta$$

There are a lot of examples entering in this framework. The representation of the metaplectic group in $L^2(\mathbb{R}^n)$ appears to be in this class, with the specificity that the phase is quadratic (Guillemin and Sternberg (1977)). A quite elementary case corresponds to the case when $N=0$ and $\phi(x,y)=x\cdot y$. No $\theta$ variable is present and so no integration appears with respect to $\theta$. When $a=1$, this defines essentially the Fourier transform. Under suitable conditions on $\phi$ and $a$, one can show that the associated operators are continuous on $\mathcal{S}(\mathbb{R}^n)$ (this is, of course, the case for the Fourier transform). This was done by Asada and Fujiwara, who transpose the theory developed by Hörmander (1971) in this context, and we should also mention the older (but more formal) work by Maslov (1972) (see also Leray (1981)). We actually do not need it in the semiclassical context because the case when the amplitude is with compact support is sufficient.

The basic object is first to look, thinking of the stationary-phase theorem, which gives the main contribution as $h \to 0$ in this "formal integral" (*see* Stationary Phase Approximation), at the critical set $\mathcal{C}_\phi$:

$$\mathcal{C}_\phi = \{(x,y,\theta) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^N | (\partial_\theta \phi)(x,y,\theta) = 0\}$$

In the case of a pseudodifferential operator, we find that it is included in $\{x=y\}$. Then we associate the canonical object, which is a Lagrangian submanifold called $\Lambda_\phi$ and defined as

$$\Lambda_\phi = \{(x,\xi,y,\eta) | \exists\theta \text{ s.t. } \xi = \nabla_x\phi(x,y,\theta),$$
$$\eta = -\nabla_y\phi(x,y,\theta), \nabla_\theta\phi(x,y,\theta) = 0\}$$

The assumptions on $\phi$ (which are omitted here) are given in order to get that $\Lambda_\phi$ is a regular manifold at least on the support of $a$. The associated operators are called Fourier integral operators (FIOs). L. Hörmander (1971, 1984) has developed a general and more intrinsic machinery but with a homogeneity condition on the phase which is irrelevant in the semiclassical context. This theory permits also the reduction to normal forms for Hamiltonians in continuation of what can be done in classical mechanics.

### Quantum Evolution

We just sketch how one approximates the operator $\exp(-\mathrm{i}tP/h)$ by an FIO. The formal construction is probably rather old (Maslov 1972, Fedoryuk and Maslov 1981) but the rigorous approach with estimates of the remainders was first considered by J Chazarain with rather strong assumptions. It has been later realized that we need only a local approximation of this operator and everything becomes easier.

The first approach followed by Helffer–Robert (see Robert (1987)) is to localize in energy, within the functional calculus associated to the operator $P$. If $I$ is an interval and $\chi$ is with compact support in $I$, it appears to be easier to approximate $\exp(-\mathrm{i}tP/h)\chi(P)$ when $P$ satisfies (H4) in a neighborhood of $I$.

We do not need any more assumptions at $\infty$ and the composition by $\chi(P)$ localizes the construction.

Although this construction is simple because we remain within a functional calculus which involves only functions of $P$, it is not always sufficient to localize in energy. We have then to localize through more general $h$-pseudodifferential operators and consider $\exp(-\mathrm{i}tP/h)a^{\mathrm{w}}(x,hD_x)$, where $a$ is a symbol with compact support. We shall quickly develop the first approach. The result is that one can approximate $U_\chi(t) := \chi(P)\exp(-\mathrm{i}tP/h)$ by a Fourier integral operator of the form

$$K_\chi(t,x,y;h) = (2\pi h)^{-n} \int^{\mathrm{osc}} \exp\left(-\frac{\mathrm{i}}{h}(S(t,x,\eta) - y\cdot\eta)\right)$$
$$\times\, d_\chi(t,x,\eta;h)\,\mathrm{d}\eta$$

with $d_\chi \sim \sum_j d_{\chi,j}h^j$, in order to have

$$\left\| \chi(P)\exp\left(-\frac{\mathrm{i}tP}{h}\right) - K_\chi(t) \right\|_{\mathcal{L}(L^2)} = \mathcal{O}(h^\infty)$$

Writing that $U_\chi(t)$ is a solution of $(hD_t + P)U_\chi = 0$, $(U_\chi)(0) = \chi(P)$, and expanding in powers of $h$, one gets a sequence of equations permitting to determine recursively the symbols. The first one was analyzed in [12] and reads, in the case when $P = -h^2\Delta + V$:

$$(\partial_t S)(t,x,\eta) + |\nabla_x S(t,x,\eta)|^2 + V(x) = 0$$

with the initial condition $S(0,x,\eta) = x\cdot\eta$.

This has been solved for $t$ small enough. The other equations are called transport equations. The first one is, for $a(t,x,\eta) = d_{\chi,0}(t,x,\eta)$,

$$\partial_t a + (\partial_\xi p_0)(x,\partial_x S(t,x,\eta))\cdot\partial_x a + ca = f$$

with initial condition $a(0,x,\eta) = \chi(p_0(x,\eta))$.

This type of equation is easily solved by integration along the integral curves of the vector field $\partial_t + (\partial_\xi p_0)(x,\partial_x S(t,x,\eta))\cdot\partial_x$.

## Applications

### The Frequency Set

One has already met the question of localization of the eigenfunctions. It appears important to give this localization, not only in position (in domain of $\mathbb{R}^\nu$) but directly in the phase space. This can be described by the notion of frequency set attached to a bounded family $u_h$ of functions in $L^2(\mathbb{R}^\nu)$ (or more generally of distributions in $\mathcal{S}'(\mathbb{R}^\nu)$). Here $h$ belongs to an interval $(0, h_0]$ or more generally to a subset of $\mathbb{R}^+$ having 0 as accumulation point.

**Definition 4** We shall say that $(x_0, \xi_0) \in \mathbb{R}^\nu \times \mathbb{R}^\nu$ does not belong to the frequency set of the family $u_h$ and write $(x_0, \xi_0) \notin \mathrm{FS}(u_h)$, if there exists a compactly supported function $\phi$ equal to 1 in a neighborhood of $x_0$ and a neighborhood $\mathcal{V}$ of $\xi_0$ in which the $h$-Fourier transform of $u_h$ satisfies, as $h \to 0$,

$$\int \exp\left(-\frac{\mathrm{i}x \cdot \xi}{h}\right) \phi(x) u_h(x)\, \mathrm{d}x = \mathcal{O}(h^\infty) \text{ in } \mathcal{V}$$

For example, the frequency set $\mathrm{FS}(u_h)$ of $u_h(x) = \chi(x) \exp(\mathrm{i}\,\theta(x)/h)$ with compactly supported $\chi$ is contained in $\{(x, \xi) \mid x \in \operatorname{supp} \chi, \xi = \nabla_x \theta(x)\}$, and the frequency set of the coherent state,

$$x \mapsto \Psi_{y,\eta,h}(x) = h^{-\nu/4} \cdot \exp\left(\frac{\mathrm{i}(x-y) \cdot \eta}{h}\right) \times \exp\left(-\frac{(x-y)^2}{h}\right)$$

is reduced to a point $(y, \eta)$.

In this semiclassical context, this notion seems to have been introduced by Guillemin and Sternberg (1977) and is further discussed in the book of Robert (1987) (see references therein). This is the semiclassical analog of the well-known notion of wave front set of a distribution introduced by Hörmander (1984) in the $C^\infty$-category for describing the singularities of a distribution, but note that a major difference is that the frequency set is attached to a family. If $P$ is an $h$-pseudodifferential operator with symbol in $S^0$, it is possible, as a consequence of the elliptic theory, to prove that: $\mathrm{FS}(Pu_{(h)}) \subset \mathrm{FS}(u_{(h)})$. For an FIO $F$ attached to a canonical relation $\kappa$, we get similarly: $\mathrm{FS}(Fu_{(h)}) \subset \kappa(\mathrm{FS}(u_{(h)}))$.

We also get a microlocal version of the localization result for the eigenfunctions mentioned in the first section (using again the parametrix construction).

**Theorem 5** *Let $E$ be in $I$ and let $(\lambda(h_j), \phi_{(h_j)}(x))$ be a sequence in $I \times L^2(\mathbb{R}^n)$, where $\lambda(h_j) \to E$ and $h_j \to 0$ as $j \to \infty$, $x \mapsto \phi_{(h_j)}(x)$ is an associated eigenfunction to $\lambda(h_j)$ with norm 1. Then*

$$\mathrm{FS}(\phi_{(h_j)}) \subset p_0^{-1}(E)$$

*Moreover, the frequency set of the family $\phi_{(h_j)}$ is invariant under the Hamiltonian flow $\phi^t$.*

The last statement in the above theorem is the analog of the theorem on the propagation of singularities for the solution of a partial differential equation (PDE) (see Hörmander (1984)) and is a consequence of the Egorov theorem, which will be presented in the next subsection.

Another remarkable property is that (see, e.g., the report on the lecture of T Paul in Rauch and Simon (1997), say, in the case of dimension 1, when $P$ is a harmonic oscillator, then $\exp(-\mathrm{i}t/hP)\psi_{y,\eta;h}$ is a coherent state attached to $\phi^t(y, \eta)$.

### Egorov's Theorem

Egorov's theorem plays a central role in the classical theory of PDE by permitting to reduce the study of general differential operators to the study of simpler model operators, the simplest one being $\partial/\partial x_n$ (see Hörmander (1984)). We use it here in a simple form, given in the semiclassical context by Robert (1987), and which will play an important role in the study of ergodic situations (*see* Quantum Ergodicity and Mixing of Eigenfunctions, and references therein). The theorem is the following:

**Theorem 6** *Let $P$ satisfy assumptions (H0)–(H3). For all a's in $S^0$ with compact support and all $t \in \mathbb{R}$, we have*

$$\left\| \exp\left(-\mathrm{i}\frac{t}{h}P\right) a^{\mathrm{w}}(x, hD_x) \exp\left(\mathrm{i}\frac{t}{h}P\right) - a_t^{\mathrm{w}}(x, hD_x) \right\|_{\mathcal{L}(L^2)} = \mathcal{O}(h)$$

*where*

$$a_t(x, \xi) = a(\phi^t(x, \xi))$$

*and $\phi^t$ is the flow of $H_{p_0}$, where $p_0$ is the principal symbol of $P$.*

The proof is based on the study of the operator $\exp(-\mathrm{i}(t/h)P)\, a^{\mathrm{w}}(x, hD_x) \exp(\mathrm{i}(t/h)P)$, which appears as the composition of three FIOs. But the Lagrangian manifold associated with this composition is the graph of the identity, and this is consequently a pseudodifferential operator whose "principal" symbol can be computed modulo $\mathcal{O}(h)$ as $a(\phi^t(x, \xi))$. As an immediate consequence, $FS(\exp(-\mathrm{i}tP/h)u_h) = \phi^t(FS(u_h))$.

### The Poisson Relation

We start from the harmonic oscillator

$$H(h) = \frac{1}{2}\left(-h^2\frac{d^2}{dx^2} + x^2\right)$$

Its spectrum is given by $(n+1/2)h$ $(n \geq 0)$. Its symbol is $a_0(x,\xi) = (1/2)(\xi^2 + x^2)$ and the corresponding flow, for any strictly positive level $E$, is periodic with primitive period $2\pi$. The quantity we are interested in is

$$S_h^\chi(t) := \sum_{j\in\mathbb{N}} \chi\left(\left(j+\tfrac{1}{2}\right)h\right)\exp\left(-it\left(j+\tfrac{1}{2}\right)\right)$$

Using the classical Poisson relation,

$$\sum_{k\in\mathbb{Z}} \hat{f}(k)\exp(ikx) = (2\pi)\cdot\sum_{k\in\mathbb{Z}} f(x+2k\pi)$$

one shows rather easily that the frequency set of $S_\chi$ is

$$\text{FS}\ (S_h^\chi) = \{(2k\pi,\tau)|\tau > 0,\ \tau \in \text{supp}\ \chi, k \in \mathbb{Z}\} \cup (\mathbb{R}\times\{0\}$$

This admits the following generalization, initiated in this context by Chazarain.

**Theorem 7** *Let $P$ satisfy* (H0)–(H4). *Let $\chi$ be a function with compact support in $I$ and let $t \mapsto f_\chi(t;h)$ be the family of distributions defined by*

$$f_\chi(t;h) = \text{tr}\left(\exp\left(-\frac{itP}{h}\right)\chi(P)\right)$$

*Then* $\text{FS}(f_\chi)$ *is contained in*

$$\{(t,\tau)|\tau \in \text{supp}\ (\chi) \text{ and } \exists(x,\xi) \text{ s.t. } p_0(x,\xi) = \tau, \phi^t(x,\xi) = (x,\xi)\}$$

According to the definition, we have to study

$$\int \exp\left(-\frac{it\tau}{h}\right)\theta(t) f_\chi(t;h)\mathrm{d}t$$

This takes the form

$$\int c(t,x,\eta)\exp\frac{i}{h}(-t\tau + S(t,x,\eta) - x\eta)\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}\eta$$

and can be analyzed by a nonstationary-phase theorem, in order to determine for which value of $\tau$ the quantity is $\mathcal{O}(h^\infty)$.

### Gutzwiller's Formula

The Gutzwiller formula was established formally by Gutzwiller (1971). It then appears in the context of high-energy spectral asymptotics in contributions of Colin de Verdière, Chazarain, and Duistermaat and Guillemin (see Duistermaat and Guillemin (1975), Hörmander (1984), Guillemin and Sternberg (1977); see also Semi-Classical Spectra and Closed Orbits and Quantum Ergodicity and Mixing of Eigenfunctions). In the semiclassical context, the simplest statement (cf. Chazarain, Helffer–Robert, Guillemin–Uribe, Meinrencken, Paul–Uribe, Dozias, Combescure–Ralston–Robert – see Robert (1987), Rauch and Simon (1997), Dimassi and Sjöstrand (1999), and in the recent article by Combescure *et al.* (1999) for techniques involving coherent states) can be presented in the following way. For a noncritical $E$, we introduce the energy surface $W_E = \{w \in T^*\mathbb{R}^n \,|\, p_0(w) = E\}$. Let $P(h)$ an $h$-pseudodifferential operator satisfying (H0)–(H4), with $I = \{E\}$. We also assume that

(Cl) The restriction of the flow $\phi^t_{p_0}$ to $W_E$ is clean. (A flow $\phi^t$, associated with a $C^\infty$-vector field $X$ on a manifold $W$, is called clean if the two following properties are satisfied:
- the set $\Gamma = \{(t,w) \in \mathbb{R}\times W \,|\, \phi^t(w) = w\}$ is a submanifold of $\mathbb{R}\times W$;
- in each point $\gamma = (t,w)$ of $\Gamma$, the tangent space to $\Gamma$ is given by $T_\gamma\Gamma = \{(\tau,v) \in \mathbb{R}\times T_wW \,|\, \tau X(w) + (D\phi^t)(w)\cdot v = v\}$.)

Then there exists a sequence of distributions $\gamma_k \in \mathcal{D}'(\mathbb{R})$, such that, for all $\phi \in \mathcal{S}(\mathbb{R})$ with compactly supported Fourier transform, we have the asymptotic expansion in powers of $h$:

$$\sum_{\lambda_j(h)\in[E-\epsilon_0/2,E+\epsilon_0/2]} \phi(h^{-1}(\lambda_j(h)-E)) \sim \sum_{j=0}^{\infty}\gamma_j(\hat{\phi})h^{-n+1+j} \qquad [13]$$

Moreover, the supports of the distributions are contained in the set of the periods of the periodic trajectories of the flow contained in $W_E$.

Actually, the proof gives more information on the structure of the different distributions. Let us just write the formula for $\gamma_0$:

$$\gamma_0 = (\pi)^{-n/2}\left(\frac{d}{d\lambda}\int_{p(x,\xi)\leq\lambda} dx d\xi\right)_{\lambda=E}\delta_0$$

where $\delta_0$ is the Dirac measure at 0.

### Clustering of Eigenvalues

We shall mention one typical result due to Chazarain–Helffer–Robert in this context, but inspired by previous results obtained for the Laplacian on compact manifolds (*see* Semi-Classical Spectra and Closed Orbits, Quantum Ergodicity and Mixing of Eigenfunctions and references therein, including Chazarain, Duistermaat–Guillemin, and Colin de Verdière).

Clustering means that the spectrum is concentrated around a specific sequence tending to $\infty$. This was observed in the case of the Laplacian on the sphere $S^{n-1}$ by explicit computations. Here we assume that, with $I = [E_1, E_2]$, the conditions (H0)–(H4) are satisfied and that

- (H5) $[E_1, E_2]$ does not meet the set of critical values of $p_0$.
- (H6) $\forall E \in [E_1 - \epsilon, E_2 + \epsilon]$, $W_E$ is connected.
- (H7) $\forall E \in [E_1 - \epsilon, E_2 + \epsilon]$, the Hamiltonian flow associated with $p_0$ is periodic, with period $T(E) > 0$, on $W_E$ (with $T(E)$ bounded).
- (H8) $\forall E \in [E_1 - \epsilon, E_2 + \epsilon]$, the subprincipal $p_1$ vanishes on $W_E$.

Then, under these conditions, one first observes that for a suitable $C^\infty$-function $f$ defined in a neighborhood of $[E_1, E_2]$, the period of the Hamiltonian flow associated with $f(p_0)$ can be chosen as constant and equal to $2\pi$. Extending the function $f$ suitably, one can then state the following result of Chazarain–Helffer–Robert:

**Theorem 8** *There exists $h_0$ and $C$ such that, for $0 < h \leq h_0$,*

$$\sigma(f(P(h))) \cap [E_1, E_2] \subset \bigcup_{k \in \mathbb{Z}} I_k(h)$$

*where*

$$I_k(h) = \left[ -\frac{S}{2\pi} - \frac{h}{4}\mu + kh - Ch^2, -\frac{S}{2\pi} - \frac{h}{4}\mu + kh + Ch^2 \right]$$

$$S = \int_\gamma \xi \, dx - 2\pi E$$

*for some (hence for any) periodic trajectory $\gamma$ of period $2\pi$, and $\mu$ is the Maslov index of this trajectory.*

Moreover, one can compute the multiplicity, in each of the intervals $I_k$. The property remains true (e.g., Dozias proved this (see Rauch and Simon (1997)) in the case when the assumption is made only for one energy $E$, but in intervals $[E - ah, E + ah]$, where $a$ can be large but $h$ is small enough.

**Remark 1** These results appear first in the context of high energy for Laplacians on compact manifolds. After illuminating contributions by physicists like Balian–Bloch, the main ideas (see the presentation in Semi-Classical Spectra and Closed Orbits) appear in the works of Colin de Verdière, Chazarain, Duistermaat–Guillemin (1975), and Weinstein (see also Hörmander (1984) and Quantum Ergodicity and Mixing of Eigenfunctions). The proof given in the semiclassical context is actually more general (it contains the case of the Laplacian on a Riemannian manifold) and shows that the results are true for general Hamiltonians.

**Remark 2** (the case of dimension 1). In this particular case, the flow is periodic and the above theorems gives the localization of the problem predicted by the Bohr–Sommerfeld relations and the computation of the multiplicity gives $n_k(h) = 1$ for $h$ small enough. This point of view was developed by Helffer–Robert (1987) (*see* Semi-Classical Spectra and Closed Orbits).

Similar properties have been extended to the case of integrable systems by Colin de Verdière in the high-energy context and in the semiclassical context by Charbonnel and Ivrii (see Ivrii (1998), Dimassi and Sjöstrand (1999), and references therein).

**Remark 3** Another interesting application of semi-classical analysis concerns the Schnirelman theorem treating the case when the flow is ergodic. We refer the reader to Quantum Ergodicity and Mixing of Eigenfunctions for references and to Helffer–Martinez–Robert (see Rauch and Simon (1997) for references) for the specific statement for general Hamiltonians in semiclassical analysis.

## Conclusions and Suggestions for Further Reading

In this brief survey we have tried to present some of the foundational techniques appearing in the "mathematical" semiclassical analysis. Of course, this is very limited, and semiclassical methods go far beyond the verification of the correspondence principle. One can refer to semiclassical analysis for many other problems where the same analysis (with a small parameter $h$) is relevant but where $h$ is no more the Planck constant. This could be a flux (Harper's equation) or the inverse of a flux, the inverse of a mass (Born–Oppenheimer's approximation), of an energy, or of a number of particles. We have not developed this point of view here.

The books given in the bibliography will allow the reader to discover other fields. The books by Robert (1987), Helffer (1988) and Dimassi and Sjöstrand (1999) present the basic statements of the theory. The book by Martinez (2002) is more "microlocal" in spirit. The lectures published in Rauch and Simon (1997) give a rather good idea of the state of art in the middle of the 1990s, and we also refer the reader to other articles in this encyclopedia for the presentation of the resonances (*see* Resonances), spectral problems connected with ergodicity (*see* Quantum Ergodicity and Mixing of Eigenfunctions), Kolmogorov–Arnol'd–Moser theory (*see* Normal Forms and Semi-Classical Approximation), and trace formulas (*see* Semi-Classical Spectra and Closed Orbits). The book by Ivrii (1998) gives the

most sophisticated theorems on the counting functions (including boundaries, singularities, ...) but is only written for specialists.

*See also:* Normal Forms and Semiclassical Approximation; Quantum Ergodicity and Mixing of Eigenfunctions; Resonances; Schrödinger Operators; Semiclassical Spectra and Closed Orbits; Stability of Matter; Stationary Phase Approximation.

## Further Reading

Combescure M, Ralston J, and Robert D (1999) A proof of the Gutzwiller semi-classical trace formula using coherent states decomposition. *Communications in Mathematical Physics* 202(2): 463–480.

Dimassi M and Sjöstrand J (1999) Spectral Asymptotics in the Semi-Classical Limit, *London Mathematical Society Lecture Notes Series*, vol. 268. Cambridge: Cambridge University Press.

Duistermaat JJ (1973) Fourier Integral Operators. *Courant Institut Mathematical Society*. New York: New York University.

Duistermaat JJ and Guillemin VW (1975) The spectrum of positive elliptic operators and periodic bicharacteristics. *Inventiones Mathematicae* 29: 39–79.

Fedoryuk MV and Maslov VP (1981) *Semi-Classical Approximation in Quantum Mechanics*. Dordrecht: Reidel.

Guillemin V and Sternberg S (1977) Geometric Asymptotics, *American Mathematical Surveys*, No. 14. Providence, RI: American Mathematical Society.

Gutzwiller M (1971) Periodic orbits and classical quantization conditions. *Journal of Mathematical Physics* 12: 343–358.

Helffer B (1988) *Introduction to the Semi-Classical Analysis for the Schrödinger Operator and Applications*, Springer Lecture Notes in Mathematics, No. 1336. Berlin: Springer.

Helffer B and Sjöstrand J (1984) Multiple wells in the semi-classical limit I. *Communications in Partial Differential Equations* 9(4): 337–408.

Helffer B and Sjöstrand J (1989) Analyse semi-classique pour l'équation de Harper III. *Mémoire de la SMF*, No. 39. Supplément du Bulletin de la SMF, Tome 117, Fasc.4.

Hörmander L (1968) The spectral function of an elliptic operator. *Acta Mathematica* 121: 193–218.

Hörmander L (1971) Fourier integral operators I. *Acta Mathematica* 127: 79–183.

Hörmander L (1979) The Weyl calculus of pseudodifferential operators. *Communications Pure and Applied Mathematics* 32: 359–443.

Hörmander L (1984) The Analysis of Linear Partial Differential Operators, *Grundlehren der Mathematischen Wissenschaften*. Berlin: Springer.

Ivrii V (1998) Microlocal analysis and precise spectral asymptotics, *Springer Monographs in Mathematics*. Berlin: Springer.

Leray J (1981) *Lagrangian Analysis and Quantum Mechanics. A Mathematical Structure Related to Asymptotic Expansions and the Maslov Index*. Cambridge: MIT Press. (English translation by Carolyn Schroeder.)

Martinez A (2002) An Introduction to Semi-Classical and Microlocal Analysis. *Universitext*. New York: Springer-Verlag.

Maslov VP (1972) *Théorie des perturbations et méthodes asymptotiques*. Paris: Dunod Gauthier-Villars.

Rauch J and Simon B (eds.) (1997) *Quasiclassical Methods*, The IMA Volumes in Mathematics and its Applications, vol. 95. New York: Springer-Verlag.

Robert D (1987) Autour de l'approximation semi-classique, *Progress in Mathematics*, No. 68. Boston, MA: Birkhäuser.

# Hubbard Model

**H Tasaki**, Gakushuin University, Tokyo, Japan



## Definitions

The Hubbard model is a standard theoretical model for strongly interacting electrons in a solid. It is a minimum model which takes into account both quantum many-body effects and strong nonlinear interaction between electrons. Here we review rigorous results on the Hubbard model, placing main emphasis on magnetic properties of the ground states.

Let the lattice $\Lambda$ be a finite set whose elements $x, y, \ldots \in \Lambda$ are called sites. Physically speaking, each site corresponds to an atomic site in a crystal. The Hubbard model is based on the simplest tight-binding description of electrons (Figure 1), where a single state is associated with each site.

For each $x \in \Lambda$ and $\sigma \in \{\uparrow, \downarrow\}$, we define the creation and the annihilation operators $c^\dagger_{x,\sigma}$ and $c_{x,\sigma}$, respectively, for an electron at site $x$ with spin $\sigma$. ($A^\dagger$ is the adjoint or the Hermitian conjugate of $A$.) These operators satisfy the canonical anti-commutation relations

$$\begin{aligned} \left\{c^\dagger_{x,\sigma}, c_{y,\tau}\right\} &= \delta_{x,y}\,\delta_{\sigma,\tau} \\ \left\{c^\dagger_{x,\sigma}, c^\dagger_{y,\tau}\right\} &= \{c_{x,\sigma}, c_{y,\tau}\} = 0 \end{aligned} \qquad [1]$$

for any $x, y \in \Lambda$ and $\sigma, \tau = \uparrow, \downarrow$, where $\{A, B\} = AB + BA$. The number operator is defined by

$$n_{x,\sigma} = c^\dagger_{x,\sigma}\, c_{x,\sigma} \qquad [2]$$

which has eigenvalues 0 and 1.

The Hilbert space of the model is constructed as follows. Let $\Phi_{\text{vac}}$ be a normalized vector state which satisfies $c_{x,\sigma}\Phi_{\text{vac}} = 0$ for any $x \in \Lambda$ and $\sigma = \uparrow, \downarrow$. Physically, $\Phi_{\text{vac}}$ corresponds to a state where there are no electrons in the system. For arbitrary subsets $\Lambda_\uparrow, \Lambda_\downarrow \subset \Lambda$, we define

$$\Phi_{\Lambda_\uparrow,\Lambda_\downarrow} = \left(\prod_{x\in\Lambda_\uparrow} c^\dagger_{x,\uparrow}\right)\left(\prod_{x\in\Lambda_\downarrow} c^\dagger_{x,\downarrow}\right)\Phi_{\text{vac}} \qquad [3]$$

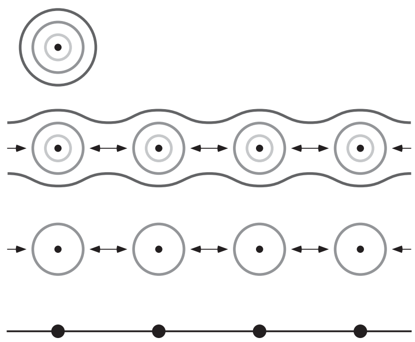



**Figure 1** A highly schematic figure which explains the philosophy of tight-binding description. (a) A single atom which has multiple electrons in different orbits. (b) When atoms come together to form a solid, electrons in the black orbits become itinerant, while those in the light gray orbits are still localized at the original atomic sites. Electrons in the gray orbits are mostly localized around the atomic sites, but tunnel to nearby gray orbits with nonnegligible probabilities. (c) We only consider electrons in the gray orbits, which are expected to play essential roles in determining various aspects of low-energy physics of the system. (d) If the gray orbit is nondegenerate, we get a lattice model in which electrons live on lattice sites and hop from one site to another. In a simplified treatment of a metal, the black and the gray orbits correspond to the 4*s* and the 3*d* bands, respectively.

in which sites in $\Lambda_\uparrow$ are occupied by up-spin electrons and sites in $\Lambda_\downarrow$ by down-spin electrons. We fix the electron number $N_e$, which is an integer satisfying $0 < N_e \le 2|\Lambda|$. (We denote by $|S|$ the number of elements in a set $S$.) The Hilbert space for the system with $N_e$ electrons is spanned by the basis states [3] with all subsets $\Lambda_\uparrow$ and $\Lambda_\downarrow$ such that $|\Lambda_\uparrow| + |\Lambda_\downarrow| = N_e$.

We define total spin operators $\hat{S}_{\text{tot}} = (\hat{S}^{(1)}_{\text{tot}}, \hat{S}^{(2)}_{\text{tot}}, \hat{S}^{(3)}_{\text{tot}})$ by

$$\hat{S}^{(\alpha)}_{\text{tot}} = \frac{1}{2} \sum_{\substack{x\in\Lambda \\ \sigma,\tau=\uparrow,\downarrow}} c^\dagger_{x,\sigma} (p^{(\alpha)})_{\sigma,\tau}\, c_{x,\tau} \qquad [4]$$

for $\alpha = 1, 2$, and 3. Here $p^{(\alpha)}$ are the Pauli matrices

$$p^{(1)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad p^{(2)} = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad p^{(3)} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad [5]$$

The operators $\hat{S}_{\text{tot}}$ are the generators of global SU(2) rotations of the spin space. As usual, we denote the eigenvalue of $(\hat{S}_{\text{tot}})^2$ as $S_{\text{tot}}(S_{\text{tot}}+1)$. The maximum possible value of $S_{\text{tot}}$ is $S_{\max} = N_e/2$ when $N_e \le |\Lambda|$, and $S_{\max} = |\Lambda| - (N_e/2)$ when $N_e \ge |\Lambda|$.

The most general Hamiltonian of the Hubbard model is

$$H = -\sum_{\substack{x,y\in\Lambda \\ \sigma=\uparrow,\downarrow}} t_{x,y}\, c^\dagger_{x,\sigma} c_{y,\sigma} + \sum_{x\in\Lambda} U_x\, n_{x,\uparrow} n_{x,\downarrow} \qquad [6]$$

Here the first term describes quantum-mechanical motion of electrons which hop around the lattice according to the amplitude $t_{x,y} = t_{y,x} \in \mathbb{R}$. Usually, $t_{x,y}$ is nonnegligible only when the two sites $x$ and $y$ are close to each other. The second term represents nonlinear interaction between electrons. There is an increase in energy by $U_x \in \mathbb{R}$ when the site $x$ is occupied by both up-spin electron and down-spin electron. We usually set $U_x > 0$ to mimic (screened) Coulomb interaction between electrons.

The Hamiltonian $H$ commutes with the total spin operator $\hat{S}^{(\alpha)}_{\text{tot}}$ for $\alpha = 1$, 2, and 3. One can thus investigate simultaneous eigenstates of $(\hat{S}_{\text{tot}})^2$ and $H$. For $S_{\text{tot}}$ in the allowed range, we denote by $E_{\min}(S_{\text{tot}})$ the lowest possible energy among the states which satisfy $(\hat{S}_{\text{tot}})^2\Phi = S_{\text{tot}}(S_{\text{tot}}+1)\Phi$.

## Wave–Particle Dualism in the Hubbard Model

It is illuminating to examine the eigenstates of the Hamiltonian [6] for the following two special cases.

First suppose that one has $U_x = 0$ for all $x \in \Lambda$, that is, the model has no interactions. For $i = 1, 2, \ldots, |\Lambda|$, let $\psi^{(i)} = (\psi^{(i)}_x)_{x\in\Lambda} \in \mathbb{C}^\Lambda$ be the single-electron eigenstate, which is the solution of the Schrödinger equation

$$-\sum_{y\in\Lambda} t_{x,y}\, \psi^{(i)}_y = \epsilon_i \psi^{(i)}_x \quad \text{for any } x \in \Lambda \qquad [7]$$

We order the energy eigenvalues as $\epsilon_i \le \epsilon_{i+1}$. By defining the corresponding creation operator by $a^\dagger_{i,\sigma} = \sum_{x\in\Lambda} \psi^{(i)}_x c^\dagger_{x,\sigma}$, we see that, for any subsets $I_\uparrow, I_\downarrow \in \{1, 2, \ldots, |\Lambda|\}$ such that $|I_\uparrow| + |I_\downarrow| = N_e$, the state

$$\Psi_{I_\uparrow, I_\downarrow} = \left(\prod_{i\in I_\uparrow} a^\dagger_{i,\uparrow}\right)\left(\prod_{i\in I_\downarrow} a^\dagger_{i,\uparrow}\right)\Phi_{\text{vac}} \qquad [8]$$

is an eigenstate of $H$ (with $U_x = 0$) with the eigenvalue $E = \sum_{i\in I_\uparrow} \epsilon_i + \sum_{i\in I_\downarrow} \epsilon_i$. The ground states are obtained by choosing $I_\uparrow, I_\downarrow$ which minimize $E$. In particular, when $N_e$ is even and the single-electron eigenenergies $\epsilon_i$ are nondegenerate, the ground state is unique and written as

$$\Phi_{\text{GS}} = \left(\prod_{i=1}^{N_e/2} a^\dagger_{i,\uparrow} a^\dagger_{i,\downarrow}\right)\Phi_{\text{vac}} \qquad [9]$$

The fact that this ground state has the minimum possible spin $S_{\text{tot}} = 0$ is known as Pauli paramagnetism.

We have seen that the Hamiltonian $H$ with $U_x = 0$ can be diagonalized by using single-electron

eigenstates $\psi^{(i)}$. When $(t_{x,y})$ has a translation invariance, each $\psi^{(i)}$ behaves as a "wave." We can say that the noninteracting models can be understood in terms of the wave picture of electrons.

Next suppose that $t_{x,y}=0$ for all $x,y\in\Lambda$, that is, the electrons do not hop. Then the Hamiltonian [6] is readily diagonalized in terms of the basis state [3], where the corresponding eigenvalue is simply $E=\sum_{x\in\Lambda_\uparrow\cap\Lambda_\downarrow}U_x$. In this case, the model is best understood in the particle picture of electrons.

We thus see that the wave–particle dualism manifests itself in the Hubbard model in an essential manner. When both the first and the second terms in the Hamiltonian [6] are present, there takes place a "competition" between wave-like nature and particle-like nature of electrons. The competition generates rich nontrivial phenomena including antiferromagnetism, ferromagnetism, metal–insulator transition, and (probably) superconductivity. To investigate these phenomena is a major motivation in the study of the Hubbard model.

## One-Dimensional Model

The Hubbard model defined on a simple one-dimensional lattice is easier to study. But it does not exhibit truly nontrivial behavior as the following classical theorem of Lieb and Mattis suggests.

**Theorem 1** *Consider a Hubbard model on a one-dimensional lattice* $\Lambda=\{1,2,\ldots,N\}$ *with open boundary conditions. We assume that* $t_{x,y}\neq 0$ *if* $|x-y|=1$, *and* $t_{x,y}=0$ *if* $|x-y|>1$. $t_{x,x}\in\mathbb{R}$ *and* $U_x\in\mathbb{R}$ *are arbitrary. Then one has* $E_{\min}(S_{\text{tot}})<E_{\min}(S_{\text{tot}}+1)$ *for any* $S_{\text{tot}}=0,1,\ldots,S_{\max}-1$ *(or* $S_{\text{tot}}=1/2,3/2,\ldots,S_{\max}-1$*).*

As a consequence, one finds that the ground states always have $S_{\text{tot}}=0$ (or $S_{\text{tot}}=1/2$) as in the noninteracting models.

The translation invariant model with $t_{x,y}=t$ if $|x-y|=1$, $t_{x,y}=0$ if $|x-y|\neq 1$, and $U_x=U$ can be solved by using the Bethe ansatz, as was first shown by Lieb and Wu. It was found that the model is insulating for all $U>0$, and there is no metal–insulator transition. (A metal–insulator transition is expected to take place in higher dimensions.) Earlier works on the Bethe ansatz were based on the assumption that the Bethe ansatz equation gives the true ground states. Recently, the existence and the uniqueness of the Bethe ansatz solution for the ground state of a finite system was proved by Goldbaum.

## Half-Filled Systems

The system in which the electron number $N_e$ is identical to the number of sites $|\Lambda|$ is said to be half-filled. Many (but not all) physical systems can be modeled as a half-filled Hubbard model.

Based on a heuristic perturbation theory, low-energy properties of half-filled models with large $U$ are expected to be similar to those of Heisenberg antiferromagnetic spin systems. There is no electrical conduction, and the spin degrees of freedom may show antiferromagnetic long-range order in the ground states.

This expectation is partly justified by the following theorem due to Lieb. A Hubbard model is said to be bipartite if the lattice $\Lambda$ can be decomposed into a disjoint union of two sublattices as $\Lambda=A\cup B$ (with $A\cap B=\emptyset$), and it holds that $t_{x,y}=0$ whenever $x,y\in A$ or $x,y\in B$. In other words, only hopping between different sublattices is allowed.

**Theorem 2** *Consider a bipartite Hubbard model. We assume* $|\Lambda|$ *is even, and the whole* $\Lambda$ *is connected through nonvanishing* $t_{x,y}$. *We also assume* $U_x=U>0$ *for any* $x\in\Lambda$. *Then the ground states of the model are nondegenerate apart from the trivial spin degeneracy, and have total spin* $S_{\text{tot}}=||A|-|B||/2$. *It also holds that* $E_{\min}(S_{\text{tot}})<E_{\min}(S_{\text{tot}}+1)$ *for any* $S_{\text{tot}}\geq ||A|-|B||/2$.

The theorem implies that, as far as the total spin is concerned, the half-filled Hubbard model behaves exactly as the Heisenberg antiferromagnet. But the existence of antiferromagnetic ordering has not been proved in any version of the Hubbard model.

To see another implication of Theorem 2, take the so-called CuO lattice in Figure 2. Here the $A$ and $B$ sublattices consist of black and white sites, respectively. One has $|A|=|\Lambda|/3$ and $|B|=2|\Lambda|/3$. Then the theorem implies that the ground state of the corresponding Hubbard model has total spin

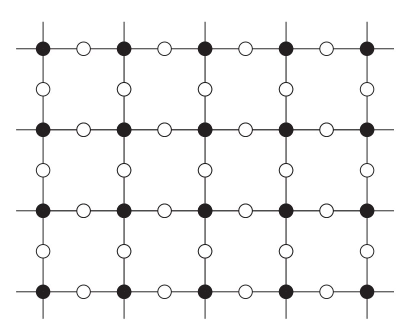


**Figure 2** An example (the so-called CuO lattice) of a bipartite lattice in which the numbers of sites in two sublattices are different. Lieb's theorem implies that the half-filled Hubbard model defined on this lattice exhibits ferrimagnetism.

$S_{tot} = ||A| - |B||/2 = |\Lambda|/6$. Since the total spin magnetic moment of the system is proportional to the number of sites $|\Lambda|$, we conclude that the model exhibits ferrimagnetism, a weaker version of ferromagnetism.

Another interesting result for the half-filled models is the following uniform density theorem by Lieb, Loss, and McCann.

**Theorem 3** *Consider a bipartite Hubbard model. $t_{x,y} \in \mathbb{R}$, $U_x \in \mathbb{R}$ are arbitrary. Suppose that the ground states are n-fold degenerate, and let $\Phi_{GS}^{(i)}$ $(i = 1, \dots, n)$ be mutually orthogonal normalized ground states. Define the correlation function by $\rho(x, y) = n^{-1} \sum_{i=1}^{n} \left\langle \Phi_{GS}^{(i)}, (c_{x,\uparrow}^{\dagger} c_{y,\uparrow} + c_{x,\downarrow}^{\dagger} c_{y,\downarrow})\, \Phi_{GS}^{(i)} \right\rangle$. ($\langle \cdot,\cdot \rangle$ is the inner product.) Then for any $x, y \in A$ or $x, y \in B$, one has $\rho(x, y) = \delta_{x,y}$.*

It is interesting that the density $\rho(x, x)$ in the ground state is always unity though the hopping matrix and interactions can be highly nonuniform.

## Ferromagnetism

Ferromagnetism is an interesting phenomenon in which the majority of the spins in the system align in the same direction. One of the original motivations to study the Hubbard model was to understand the origin of ferromagnetism in an idealized situation. Let us recall that neither the hopping term nor the interaction term in the Hamiltonian [6] favors ferromagnetism (or any other magnetic order). One must deal with the interplay between the two terms to have ferromagnetism. Here we review three rigorous examples of saturated ferromagnetism in the Hubbard model. Saturated ferromagnetism is the strongest form of ferromagnetism where the ground state has $S_{tot} = S_{max}$.

The first example is due to Nagaoka and Thouless.

**Theorem 4** *Take an arbitrary finite lattice $\Lambda$, and let $N_e = |\Lambda| - 1$. Assume that $t_{x,y} \le 0$ for any $x \ne y$, and let $U_x \to \infty$ for all $x \in \Lambda$. (Taking the limit $U_x \to \infty$ is equivalent to inhibiting x from being occupied by two electrons.) Then among the ground states of the model, there exist states with total spin $S_{tot} = S_{max} (= N_e/2)$. If the system further satisfies the connectivity condition (see below), then the ground states have $S_{tot} = S_{max} (= N_e/2)$ and are nondegenerate apart from the trivial spin degeneracy.*

The connectivity condition is a simple condition which holds in most of the lattices in two or higher dimensions, including the square lattice, the triangular lattice, or the cubic lattice. To be precise the condition requires that "by starting from any electron configuration on the lattice and by moving around the hole along nonvanishing $t_{x,y}$, one can get any other electron configuration."

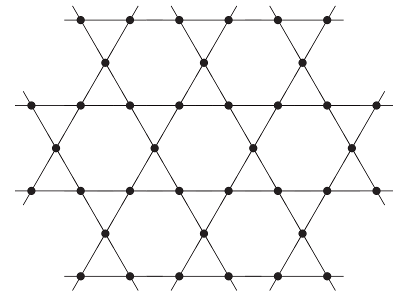


**Figure 3** The Hubbard model on the kagomé lattice is a typical example which exhibits flat-band ferromagnetism.

The requirements that $U_x \to \infty$ and $N_e = |\Lambda| - 1$ are indeed rather pathological. We still do not know if the ferromagnetism extends to more realistic situations. Heuristic studies indicate that the issue is highly delicate.

A completely different class of rigorous examples of ferromagnetism was found by Mielke. Take, for example, the kagomé lattice of Figure 3, and define a Hubbard model by setting $t_{x,y} = t < 0$ when $x$ and $y$ are neighboring, $t_{x,y} = 0$ otherwise, and $U_x = U \ge 0$ for any $x \in \Lambda$. Then the corresponding single-electron Schrödinger equation [7] has a peculiar feature that its ground states are $\{(|\Lambda|/3) + 1\}$-fold degenerate. This huge degeneracy corresponds to the fact that the lowest-energy band of the model is completely dispersionless (or flat).

**Theorem 5** *Consider the Hubbard model on the kagomé lattice with $N_e = (|\Lambda|/3) + 1$. For any $U > 0$, the ground states have $S_{tot} = S_{max} (= N_e/2)$ and are nondegenerate apart from the trivial spin degeneracy.*

There are similar examples in higher dimensions. Ferromagnetism observed in these models is called flat-band ferromagnetism.

The above examples of ferromagnetism have either singular interaction $(U_x \to \infty)$ or singular dispersion relation (highly degenerate single-electron ground states). Tasaki found a class of Hubbard models which are free from such singularities, and exhibit ferromagnetism.

For simplicity, we concentrate on the simplest model in one dimension. There are similar examples in higher dimensions. Take the one-dimensional lattice $\Lambda = \{1, 2, \dots, N\}$ with $N$ sites (where $N$ is an even integer), and impose a periodic boundary condition by identifying the site $N + 1$ with the site 1. The hopping matrix is defined by setting $t_{x,x+1} = t_{x+1,x} = -t'$ for any $x \in \Lambda$, $t_{x,x+2} = t_{x+2,x} = -t$ for even $x$, $t_{x,x+2} = t_{x+2,x} = s$ for odd $x$, and $t_{x,y} = 0$

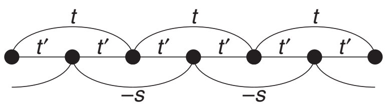



**Figure 4** An example of nonsingular Hubbard model which exhibits saturated ferromagnetism.

otherwise. Here $t > 0$ and $s > 0$ are independent parameters, but the parameter $t'$ is determined as $t' = \sqrt{2}(t + s)$.

As can be seen from Figure 4, electrons are allowed to hop to next-nearest neighbors. Thus, Theorem 1 does not apply. The single-electron ground states are not degenerate unless $s = 0$. We set $U_x = U > 0$ for any $x \in \Lambda$, and fix the electron number as $N_e = N/2$. In terms of filling factor, this corresponds to the quarter filling.

**Theorem 6** *Suppose that the two dimensionless parameters t/s and U/s are sufficiently large. Then the ground states have* $S_{tot} = S_{max}(= N/4)$ *and are nondegenerate apart from the trivial spin degeneracy.*

The theorem is valid, for example, when $t/s \geq 4.5$ if $U/s = 50$, and $t/s \geq 2.6$ if $U/s = 100$. It is crucial that the statement of the theorem is valid only when the interaction $U$ is sufficiently large. In the same model, it is also proved that low-lying excitation above the ground state has a normal dispersion relation of a spin-wave excitation.

We would like to point out that one can learn more details about the Hubbard model and further rigorous results from the review articles (Lieb 1995, Tasaki 1998a, Tasaki 1998b). One can also find references for most of the results discussed here in these review articles, especially in Lieb (1995).

As for the latest results which are not included in the above reviews, see recent publications, for example, Lieb and Wu (2003), Tasaki (2003), and Goldbaum (2005), and references therein.

## Further Reading

Goldbaum PS (2005) Existence of solutions to the Bethe Ansatz Equations for the 1D Hubbard model: finite lattice and thermodynamic limit. *Communications in Mathematical Physics* 258: 317–337 (cond-mat/0403736).

Lieb EH (1995) The Hubbard model – some rigorous results and open problems. In: Iagolnitzer D (ed.) *Proceedings of the XIth International Congress of Mathematical Physics*, Paris, 1994. pp. 392–412. International Press. cond-mat/9311033.

Lieb EH and Wu FY (2003) The one-dimensional Hubbard model: a reminiscence. *Physica A* 321: 1–27 (cond-mat/0207529).

Tasaki H (1998a) The Hubbard model – an introduction and selected rigorous results. *Journal of Physics: Condensed Matter* 10: 4353–4378 (cond-mat/9512169).

Tasaki H (1998b) From Nagaoka's ferromagnetism to flat-band ferromagnetism and beyond – an introduction to ferromagnetism in the Hubbard model. *Prog. Theor. Phys.* 99: 489–548 (cond-mat/9712219).

Tasaki H (2003) Ferromagnetism in the Hubbard model: a constructive approach. *Communications in Mathematical Physics* 242: 445–472 (cond-mat/0301071).

**Hydrodynamic Equations** *see* Interacting Particle Systems and Hydrodynamic Equations

# Hyperbolic Billiards

**Maciej P Wojtkowski**, University of Arizona, Tucson, AZ, USA and Institute of Mathematics PAN, Warsaw, Poland



## Introduction

Billiards are a class of dynamical systems with appealingly simple description. A point particle moves with constant velocity in a box of arbitrary dimension (the billiard table) and reflects elastically from the boundary (the component of velocity perpendicular to the boundary is reversed and the parallel component is preserved). Mathematically, it is a class of Hamiltonian systems with collisions defined by symplectic maps on the boundary of the phase space. The billiard dynamics defines a one-parameter group of maps $\Phi^t$ of the phase space which preserve the Lebesgue measure, and are in general only measurable due to discontinuities. The boundaries of the box are made up of pieces, concave, convex, and flat. Discontinuities occur at the orbits tangent to concave pieces of the boundary of the box. The orbits hitting two adjacent pieces ("corners") cannot be naturally

continued, which is another source of discontinuities. These singularities are not too severe so that the flow has well-defined Lyapunov exponents and Pesin structural theory is applicable (Katok and Strelcyn 1986). A billiard system is called hyperbolic if it has nonzero Lyapunov exponents on a subset of positive Lebesgue measure, and completely hyperbolic if all of its Lyapunov exponents are nonzero almost everywhere, except for one zero exponent in the direction of the flow.

Billiards in smooth strictly convex domains have no singularities, but no such examples are known to be hyperbolic.

In general, billiards exhibit mixed behavior just like other Hamiltonian systems; there are invariant tori intertwined with "chaotic sea." In hyperbolic billiards, stable behavior is excluded by the choice of the pieces in the boundary of the box, arbitrary concave pieces and special convex ones, and their particular placement. Thus, hyperbolicity is achieved by design, as in optical instruments.

It was established by Turaev and Rom-Kedar (1998) that complete hyperbolicity may be lost under generic singular perturbation of the billiard system to a smooth Hamiltonian system.

Hyperbolicity is the universal mechanism for random behavior in deterministic dynamical systems. Under suitable additional assumptions, it leads to ergodicity, mixing, $K$-property, Bernoulli property, decay of correlations, central-limit theorem, and other stochastic properties. Hyperbolic billiards provide a natural class of examples for which these properties were studied. In this article we restrict ourselves to hyperbolicity itself.

The most prominent example of a hyperbolic billiard is the gas of hard spheres. This way of looking at the system was developed in the groundbreaking papers of Sinai (see Chernov and Sinai (1987) for an exhaustive list of references). The collection of papers (Szász 2000) contains more up-to-date information. Another source on hyperbolic billiards is the book by Chernov and Markarian (2005). The books by Kozlov and Treschev (1990), and by Tabachnikov (1995) provide broad surveys of billiards from different perspectives.

## Jacobi Fields and Monotonicity

The key to understanding hyperbolicity in billiards lies in two essentially equivalent descriptions of infinitesimal families of trajectories. The basic notion is that of a Jacobi field along a billiard trajectory. Let $\gamma(t,u)$ be a family of billiard trajectories, where $t$ is time and $u$ is a parameter, $|u| < \epsilon$. A Jacobi field along $\gamma(t,0)$ is defined by $J(t) = \partial\gamma/\partial u_{|u=0}$.

Jacobi fields form a finite-dimensional vector space which can be identified with the tangent to the phase space at points along the trajectory. They contain the same information as the derivatives of the billiard flow $D\Phi^t$. In particular, the Lyapunov exponents are the exponential rates of growth of Jacobi fields.

Jacobi fields split naturally into parallel and perpendicular components to the trajectory, each of them a Jacobi field in its own right. The parallel Jacobi field carries the zero Lyapunov exponent. In the rest we discuss only the perpendicular Jacobi fields. Between collisions the Jacobi fields satisfy the differential equation $J'' = 0$, hence $J(t) = J(0) + tJ'(0)$. At a collision a Jacobi field undergoes a change by the map

$$\begin{aligned} J(t_c^+) &= \mathcal{R}J(t_c^-) \\ J'(t_c^+) &= \mathcal{R}J'(t_c^-) + \mathcal{P}^*\mathcal{K}\mathcal{P}J(t_c^+) \end{aligned} \qquad [1]$$

where $J(t_c^-)$ and $J(t_c^+)$ are Jacobi fields immediately before and after collision, $\mathcal{K}$ is the shape operator of the piece of the boundary ($\mathcal{K} = \nabla n, n$ is the inside unit normal to the boundary), and $\mathcal{P}$ is the projection along the velocity vector from the hyperplane perpendicular to the orbit to the hyperplane tangent to the boundary. Finally, $\mathcal{R}$ is the orthogonal reflection in the hyperplane tangent to the boundary.

Perpendicular Jacobi fields at a point of a trajectory can be identified with a subspace of the tangent to the phase space, the subspace perpendicular to the phase trajectory. To measure the growth/decay of Jacobi fields, we introduce a quadratic form on the tangent spaces, or equivalently on Jacobi fields, $\mathcal{Q}(J, J') = <J, J'>$ . Evaluation of $\mathcal{Q}$ on a Jacobi field is a function of time $\mathcal{Q}(t)$. Between collisions we have $\mathcal{Q}(t_2) \geq \mathcal{Q}(t_1)$ for $t_2 \geq t_1$ (monotonicity). By [1] the monotonicity at the collisions, that is, $\mathcal{Q}(t_c^+) \geq \mathcal{Q}(t_c^-)$ is equivalent to the positive semidefiniteness of the shape operator $\mathcal{K} \geq 0$, it holds for concave pieces of the boundary. If $\mathcal{K} > 0$ at a point of collision with the boundary, then for $(J, J') \neq (0,0)$, we have $\mathcal{Q}(t_2) > \mathcal{Q}(t_1)$ (strict monotonicity), assuming that the collision occurred between time $t_1$ and $t_2$.

In billiards with concave pieces of the boundary, where $\mathcal{K} \geq 0, \mathcal{K} \neq 0$, strict monotonicity may still occur after sufficiently many reflections (eventual strict monotonicity, or ESM). Such billiards are called semidispersing, and the gas of hard spheres is an example.

The role of monotonicity is revealed in the following:

**Theorem 1** (Wojtkowski 1991). *If a system is eventually strictly monotone (ESM), except on a set of orbits of zero measure, then it is completely hyperbolic.*

The theorem applies to billiard systems. It can be generalized and applied to other systems, not even Hamiltonian (see Wojtkowski (2001) for precise formulations, references and the history of this idea).

The difficulty in applying the above theorem to the gas of hard spheres lies in the gap between monotonicity and strict monotonicity. There are many orbits on which strict monotonicity is never attained (parabolic orbits). Establishing that the family of parabolic orbits has measure zero (or better yet codimension 2) is a formidable task. It was brought to conclusion in the work of Simányi (2002).

## Wave Fronts and Monotonicity

There is a geometric formulation of monotonicity (which historically preceded the one given above). Let us consider a local wave front, that is, a local hypersurface $W(0)$ perpendicular to a trajectory $\gamma(t)$ at $t=0$. Let us consider further all billiard trajectories perpendicular to $W(0)$. The points on these trajectories at time $t$ form a local hypersurface $W(t)$ perpendicular again to the trajectory (warning: at exceptional moments of time the wave front $W(t)$ may be singular). Infinitesimally wave fronts are described by the shape operator $U=\nabla n$, where $n$ is the unit normal field. $U$ is a symmetric operator on the hyperplane tangent to the wave front (and perpendicular to the trajectory $\gamma(t)$. The evolution of infinitesimal wave fronts is described by the formulas

$$\begin{aligned} U(t) &= (tI + U(0)^{-1})^{-1} \quad \text{without collisions} \\ U(t_c^+) &= \mathcal{R}U(t_c^-)\mathcal{R} + \mathcal{P}^*\mathcal{K}\mathcal{P} \quad \text{at a collision} \end{aligned} \qquad [2]$$

It follows that between collisions a wave front that is initially convex (i.e., diverging, or $U>0$) will stay convex. Moreover, any wave front after a sufficiently long run without collisions will become convex (after which the normal curvatures of the wave front will be decaying). The second part of [2] shows that after a reflection in a strictly concave boundary a convex wave front becomes strictly convex (and its normal curvatures increase). These properties are equivalent to (strict) monotonicity as formulated above. Indeed, in the language of Jacobi fields an infinitesimal wave front represents a linear subspace in the space of perpendicular Jacobi fields, that is, the tangent space. (Furthermore, it is a Lagrangian subspace with respect to the standard symplectic form.) We can follow individual Jacobi fields or whole subspaces of them. It explains the parallel of [1] and [2]. The form $\mathcal{Q}$ allows the introduction of positive and negative Jacobi fields and positive and negative Lagrangian subspaces. An infinitesimal convex wave front represents a positive Lagrangian subspace. Monotonicity is equivalent to the property that a positive Lagrangian subspace stays positive under the dynamics (it may appear that there is a loss of information in formulas [2] compared to [1], but actually they are equivalent due to the symplectic nature of the dynamics (Wojtkowski 2001).

## Design of Hyperbolic Billiards

In view of [2] it seems that a convex piece in the boundary ($K<0$) excludes monotonicity. There are two ways around this obstacle. First, we could change the quadratic form $\mathcal{Q}$ at the convex boundary. Second, we can treat convex pieces as "black boxes" and look only at incoming and outgoing trajectories. Although the second strategy seems more restrictive, all the examples constructed to date fit the black box scenario, and we will present it in more detail.

To understand this approach, let us consider a billiard table with flat pieces of the boundary and exactly one convex piece. A trajectory in such a billiard experiences visits to the convex piece separated by arbitrary long sequences of reflections in flat pieces, which do not affect the geometry of a wave front at all. Hence, whatever is the geometry of a wave front emerging from the curved piece it will become convex and very flat by the time it comes back to the curved piece of the boundary again. Hence, it follows, at least heuristically, that we must study the complete passage through the convex piece of the boundary, regarding its effect on convex, and especially flat, wave fronts.

Important difference between convex and concave pieces is that a trajectory has usually several consecutive reflections in the same convex piece; moreover, the number of such reflections is unbounded. A finite billiard trajectory is called "complete" if it contains reflections in one and the same piece of the boundary, and it is preceded and followed by reflections in other pieces.

**Definition** A complete trajectory is (strictly) $z$-monotone if for every nonzero Jacobi field the

value of the form $Q$ (increases) does not decrease between the point at the distance $z$ before the first reflection and the point at the distance $z$ after the last reflection.

A complete trajectory is parabolic if there is a nonzero Jacobi field $J$ such that $J'$ vanishes before the first and after the last reflection.

In the language of wave fronts, a complete trajectory is $z$-monotone if every diverging wave front at a distance at least $z$ from the first reflection becomes diverging after the last reflection at the distance $z$, or earlier.

It turns out that the only obstruction to monotonicity of complete trajectories is parabolicity. More precisely, if a complete trajectory is not parabolic then it is $z$-monotone for some $z > 0$.

It follows from Theorem 1 that we get a completely hyperbolic billiard if we put together curved pieces with no complete parabolic trajectories and some flat pieces, in such a way that for every two consecutive complete trajectories, being $z_1$- and $z_2$-monotone, respectively, the distance from the last reflection in the first trajectory to the first reflection in the second one is bigger than $z_1 + z_2$. Indeed, we can put together the midpoints of trajectories leaving one curved piece and hitting another one into the Poincaré section of the billiard flow and we obtain immediately ESM for the return map.

We can formulate somewhat informally two principles for the design of hyperbolic billiards.

1. *No parabolic trajectories* Convex pieces of the boundary cannot have complete parabolic trajectories.
2. *Separation* There must be enough separation (in space or in time through reflections in flat pieces) between strictly $z$-monotone trajectories according to the values of $z$.

All of the examples of hyperbolic billiards constructed up to now are designed according to these principles.

## Hyperbolic Billiards in Dimension 2

Checking the absence of parabolic trajectories is nontrivial due to the unbounded number of reflections in complete trajectories close to tangency. It was accomplished so far only in integrable, or near integrable examples, with the exception of convex scattering pieces described in the following.

Billiards in dimension 2 are understood best. First of all, there is yet another way of describing infinitesimal families of nearby trajectories. Every infinitesimal family of rays in the plane has a point of focusing (in linear approximation), possibly at infinity. This point of focusing contains the same information as the curvature of a wave front (it is the center of curvature, rather than curvature itself) and it has the advantage that it does not change between collisions. The focusing points before and after a reflection are related by the familiar mirror equation of the geometric optics:

$$-\frac{1}{f_0} + \frac{1}{f_1} = \frac{2}{d}$$

where $f_0, f_1$ are the signed distances of the points of focusing to the reflection point, $d = r\cos\theta$, $r$ being the radius of curvature of the boundary piece ($r > 0$ for a strictly convex piece), and $\theta$ the angle of incidence. The mirror equation is just the two dimensional version of [2].

It is instructive to consider an arc of a circle. A billiard in a disk is integrable due to its rotational symmetry. Let $J$ be a Jacobi field obtained by rotation of a trajectory. This family of trajectories ("the rotational family") is focused exactly in the middle between two consecutive reflections (that is where $J$ vanishes). It follows further from the mirror equation that a parallel family of orbits is focused at a distance $d/2$ after the reflection, and any family focusing somewhere between the parallel family and the rotational family will focus at a distance somewhere between $d/2$ and $d$, not only after the first reflection, but also after arbitrary long sequence of reflections.

Hence, any complete trajectory in an arc of a circle is $z$-monotone, where $2z$ is the length of a single segment of the trajectory and strictly $z'$-monotone for any $z' > z$. Two arcs of a circle separated by parallel segments form the stadium of Bunimovich (1979).

Lazutkin (1973) showed that billiards in smooth strictly convex domains are near integrable near the boundary. Donnay (1991) applied Lazutkin's coordinates to establish that for an arbitrary strictly convex arc the situation near the boundary is similar to that in a circle, that is, complete trajectories near tangency are $z$-monotone, where $z$ is of the order of the length of a single segment. In particular, no near tangent complete trajectory can be parabolic. Hence, this crucial calculation shows that if a strictly convex arc has no parabolic trajectories then any sufficiently small perturbation also has no parabolic trajectories. It follows further that any sufficiently small piece of a given strictly convex arc has no parabolic trajectories.

It turns out that in dimension 2, complete parabolic trajectories are also $z$-monotone for some

$z > 0$ (but clearly not strictly monotone) (Wojtkowski 2005). However, they are still an obstacle to complete hyperbolicity because in general nearby complete trajectories are $z$-monotone without a bound for the values of $z$, so that no separation of convex pieces is sufficient.

Integrability of the elliptic billiard allows one to establish strict monotonicity of trajectories in the semi-ellipse with endpoints on the longer axis, Wojtkowski 1986. Donnay (1991) showed that also the semi-ellipse with endpoints on the shorter axis has no parabolic trajectories provided that the eccentricity is less than $\sqrt{2}/2$. As the eccentricity goes to $\sqrt{2}/2$ the separation required to produce a hyperbolic billiard goes to infinity. Markarian *et al.* (1996) obtained explicitly the separation of the elliptic pieces needed for hyperbolicity, when the eccentricity is smaller than $\sqrt{2-\sqrt{2}}/2$.

It follows from the mirror equation that a trajectory with one reflection in a convex piece is always strictly $z$-monotone for $z > d$. Hence, if for any two consecutive reflections in convex pieces with respective values of $d$ equal to $d_1$ and $d_2$, the distance between reflections exceeds $d_1 + d_2$, then the billiard is completely hyperbolic. For one convex piece this condition, called convex scattering, turns out to be equivalent to $\mathrm{d}^2 r/\mathrm{d}s^2 < 0$, where $s$ is the arc length (Wojtkowski 1986). This leads to examples of hyperbolic billiards with one convex piece of the boundary, like the domain bounded by the cardioid.

Also, any complete trajectory in a convex scattering piece is strictly $z$-monotone for $z$ bigger than the maximum of the values of $d$ for the first and the last segment of the trajectory. This allows to find easily the explicit separation of convex scattering pieces guaranteeing hyperbolicity.

## Hyperbolic Billiards in Higher Dimensions

In higher dimensions, only two constructions of hyperbolic billiards with convex pieces in the boundary are known. The first construction by Bunimovich (1988), involves a piece of a sphere whose angular size, as seen from the center, does not exceed $\pi/2$ (Wojtkowski 1990, 2005, Bunimovich and Rehacek 1998). The second construction by Papenbrock (2000) uses two cylinders, at 90° with respect to each other to destroy integrability (Wojtkowski 2005). In both cases, the successful treatment is based on integrability of the billiard systems bounded by a sphere or a cylinder.

In both of these constructions, trajectories need to be cut into strictly monotone pieces of unbounded lengths. In the case of spherical caps, complete trajectories are $z$-monotone with unbounded value of $z$ and the geometry of the billiard table is used to separate them in time by sufficiently many reflections in flat pieces of the boundary (Wojtkowski 2005). In the case of cylinders, trajectories are cut by consecutive returns to a Poincaré section in the middle of the billiard table.

## Soft Billiards

The same ideas of monotonicity and strict monotonicity are applicable to soft billiards, where specular reflections are replaced by scatterers in which the point particle is subjected to the action of a spherically symmetric potential. As in ordinary billiards, we compare the wave fronts along trajectories before entering and after leaving scatterers. Again, in the absence of parabolic trajectories sufficient separation of the scatterers produces a completely hyperbolic system.

The conditions on the potential that guarantee the absence of parabolic trajectories were obtained by Donnay and Liverani (1991) in the two-dimensional case and by Bálint and Tóth (2006) in higher dimensions. The complete integrability of the motion of a point particle in a spherically symmetric potential is crucial in the derivation of these conditions (Wojtkowski 2005).

*See also:* Billiards in Bounded Convex Domains; Ergodic Theory; Hamiltonian Systems: Stability and Instability Theory; Hyperbolic Dynamical Systems; Polygonal Billiards; Random Matrix Theory in Physics.

## Further Reading

Bálint P and Tóth IP (2006) Hyperbolicity in multi-dimensional Hamiltonian systems with applications to soft billiards. *Nonlinearity* (to appear).

Bunimovich LA (1979) On the ergodic properties of nowhere dispersing billiards. *Communications in Mathematical Physics* 65: 295–312.

Bunimovich LA (1988) Many-dimensional nowhere dispersing billiards with chaotic behavior. *Physica D* 33: 58–64.

Bunimovich LA and Rehacek J (1998) How high-dimensional stadia look like. *Communications in Mathematical Physics* 197: 277–301.

Chernov NI and Sinai YaG (1987) Ergodic properties of some systems of 2-dimensional disks and 3-dimensional spheres. *Russian Mathematical Surveys* 42: 181–207.

Chernov NI and Markarian R (2005) *Billiards*. Providence, RI: American Mathematical Society.

Donnay V (1991) Using integrability to produce chaos: billiards with positive entropy. *Communications in Mathematical Physics* 141: 225–257.

Donnay V and Liverani C (1991) Potentials on the two-torus for which the Hamiltonian flow is ergodic. *Communications in Mathematical Physics* 135: 267–302.

Katok A and Strelcyn JM (1986) (with the collaboration of F. Ledrappier and F. Przytycki) *Invariant Manifolds, Entropy and Billiards; Smooth Maps with Singularities*, Lecture Notes in Mathematics, 1222. Springer.
Kozlov VV and Treschev DV (1990) *Billiards. A Genetic Introduction to the Dynamics of Systems with Impacts.* Providence, RI: American Mathematical Society.
Lazutkin VF (1973) On the existence of caustics for the billiard ball problem in a convex domain. *Mathematics of the USSR-Izvestiya* 7: 185–215.
Markarian R, Oliffson Kamphorst S, and Pinto de Carvalho S (1996) Chaotic properties of the elliptical stadium. *Communications in Mathematical Physics* 174: 661–679.
Papenbrock T (2000) Numerical study of a three dimensional generalized stadium billiard. *Physical Review E* 61: 4626–4628.
Simányi N (2002) The complete hyperbolicity of cylindric billiards. *Ergodic Theory and Dynamical Systems* 22: 281–302.
Szász D (ed.) (2000) *Hard Ball Systems and the Lorentz Gas.* Encyclopaedia of Mathematical Sciences, 101. Berlin: Springer.
Tabachnikov S (1995) *Billiards.* Soc. Math. Paris, France.
Turaev D and Rom-Kedar V (1998) Elliptic islands appearing in near-ergodic flows. *Nonlinearity* 11: 575–600.
Wojtkowski MP (1986) Principles for the design of billiards with nonvanishing Lyapunov exponents. *Communications in Mathematical Physics* 105: 391–414.
Wojtkowski MP (1990) Linearly stable orbits in 3-dimensional billiards. *Communications in Mathematical Physics* 129: 319–327.
Wojtkowski MP (1991) Systems of classical interacting particles with nonvanishing Lyapunov exponents. In: Arnol'd L, Crauel H, and Eckmann J-P (eds.) *Lyapunov Exponents, Proceedings, Oberwolfach 1990*, Lecture Notes in Mathematics, 1486. 243–262.
Wojtkowski MP (2001) Monotonicity, J-algebra of Potapov and Lyapunov exponents. In: *Smooth Ergodic Theory and Its Applications*, pp. 499–521. Proc. Symp. Pure Math. American Mathematical Society.
Wojtkowski MP (2005) Design of hyperbolic billiards. Preprint.

# Hyperbolic Dynamical Systems

**B Hasselblatt**, Tufts University, Medford, MA, USA



## Introduction

### Division of Smooth Dynamical Systems

Linear maps can be elliptic (complex diagonalizable with all eigenvalues on the unit circle), parabolic (all eigenvalues on the unit circle but some Jordan blocks of size at least 2), or hyperbolic (no eigenvalues on the unit circle), and for differentiable dynamical systems, that is, smooth maps or flows, one can roughly make an analogous subdivision (see Hasselblatt and Katok 2002, p. 100f). The linear maps not covered by these alternatives are those with some eigenvalues on the unit circle and others off it; the corresponding class of "partially hyperbolic" dynamical systems is usually considered in the context of hyperbolic dynamical systems with a view to studying phenomena wherein the hyperbolic behavior dominates. Thus, elliptic dynamical systems are more or less similar to isometries, with orbit separation constant or at most oscillatory but without persistent growth. KAM theory deals with elliptic systems, establishing that much of the ellipticity in an integrable Hamiltonian system persists under perturbation. Parabolic systems may have polynomial orbit separation produced by a local "shear" phenomenon; billiards in polygonal domains are an example of this. Hyperbolic dynamical systems are characterized by exponential divergence of orbits. They are of interest because of the complexity of their orbit structure with respect to both topological and statistical behavior.

Specifically, the stretching (corresponding to eigenvalues outside the unit circle in the case of linear maps) combined with the folding necessitated by compactness of the phase space produces not only highly sensitive dependence of orbit asymptotics on initial conditions, but also a close intertwining of different behaviors. On the one hand, there is a dense set of periodic points, on the other hand, an abundance of dense orbits. While there are only finitely many periodic points of a given period, their number grows exponentially as a function of the period. The entropy of these systems is positive, which indicates that the overall complexity of the orbit structure grows exponentially as a function of the length of time for which it is being tracked. In effect, the behavior of orbits is so intricate as to be quasirandom, which makes it natural to use statistical methods to describe these systems.

### History of Hyperbolic Dynamical Systems

One strand of the history of hyperbolic dynamical systems leads back to the question of the stability of the solar system and to Poincaré, in whose prize memoir on the three-body problem the possibility of "homoclinic tangles" first presented itself. For Poincaré, this was important because the resulting complexity demonstrates that this system is not integrable. We describe below how hyperbolic dynamics arises in this situation (see Figure 3).

Another strand emerged about a decade later with Hadamard's study of geodesic flows (free particle motion) on negatively curved surfaces. Hadamard noted that these exhibit the kind of sensitive dependence on initial conditions as well as the pseudorandom behavior that are central features of hyperbolic dynamics. This subject was developed much further after the advent of ergodic theory, with the Boltzmann ergodic hypothesis as an important motivation: work by numerous mathematicians, principally Hedlund and Hopf, showed that free particle motion on a negatively curved surface provides examples of ergodic mechanical systems. More than two decades later, in the 1960s, Anosov and Sinai overcame a fundamental technical hurdle and established that this is indeed the case in arbitrary dimension. This was done in the more general context of a class of dynamical systems known now as Anosov systems, which were axiomatically defined and systematically studied for the first time during this period of research in Moscow.

A greater class of dynamical systems exhibiting chaotic behavior was introduced by Smale in his seminal 1967 paper under the name of Axiom-A systems. This class includes the hyperbolic dynamics arising from homoclinic tangles, see Figure 3 (*see* Homoclinic Phenomena). Smale's motivation was his program of classifying dynamical systems under topological conjugacy, and the consequent search for structurally stable systems. Today, Axiom-A (and Anosov) systems are valued as idealized models of chaos: while the conditions defining Axiom A are too stringent to include many real-life examples, it is recognized that they have features shared in various forms by most chaotic systems. Here, we concentrate on the discrete-time context to keep notations lighter.

Partial hyperbolicity was introduced in the 1970s and has proved that a limited amount of hyperbolicity in a dynamical system can produce much of the global complexity (such as ergodicity or the presence of dense orbits) exhibited by hyperbolic systems, and can do so in a robust way. Here one imposes uniform conditions, but expansion and contraction are not assumed to occur in all directions. Stable ergodicity has been an important subject of research in the last decade.

Nonuniform hyperbolicity weakens hyperbolicity by allowing the contraction and expansion rates to be nonuniform. This was motivated by examples of systems with hyperbolicity where expansion or contraction can be arbitrarily weak or absent in places, such as the Hénon attractor, and by situations where hyperbolicity coexists with singularities, such as for (semi)dispersing billiards (*see* Hyperbolic Billiards).

With respect to both uniformly and nonuniformly hyperbolic systems, dimension theory has been a subject of much interest (computations and estimates of the fractal dimension of attractors and hyperbolic sets, which is deeply connected to dynamical properties of the system).

A different weakening of hyperbolicity, the presence of a dominated splitting, has been of interest from the a viewpoint to stability and classification of diffeomorphisms.

The study of hyperbolic dynamics has always had interactions with other sciences and other areas of mathematics. In the natural and social sciences, this is the study of chaotic motions of just about any kind. Examples of applications in related areas of mathematics are geometric rigidity (an interaction with differential geometry) and rigidity of group actions.

## Uniformly Hyperbolic Dynamical Systems

### Definitions

Let $f$ be a smooth invertible map. A compact invariant set of $f$ is said to be "hyperbolic" if at every point in this set, the tangent space splits into a direct sum of two subspaces $E^{\mathrm{u}}$ and $E^{\mathrm{s}}$ with the property that these subspaces are invariant under the differential d$f$, that is, $\mathrm{d}f(x)E^{\mathrm{u}}(x)=E^{\mathrm{u}}(f(x))$, $\mathrm{d}f(x)E^{\mathrm{s}}(x)=E^{\mathrm{u}}(f(x))$, and that d$f$ expands vectors in $E^{\mathrm{u}}$ and contracts vectors in $E^{\mathrm{s}}$, that is, there are constants $0<\lambda<1<\mu, c>0$ such that if $v\in E^{\mathrm{s}}(x)$ for some $x$, then $\|\mathrm{d}f^n v\|\leq c\lambda^n\|v\|$ for $n=1,2,\ldots$, and if $v\in E^{\mathrm{u}}(x)$ for some $x$, then $\|\mathrm{d}f^{-n}v\|\leq c\mu^{-n}\|v\|$ for $n=1,2,\ldots$.

If $E^{\mathrm{u}}=\{0\}$ in the definition above, then the invariant set is made up of attracting fixed points or periodic orbits. Similarly, if $E^{\mathrm{s}}=\{0\}$, then the orbits are repelling. If neither subspace is trivial, then the behavior is locally "saddle-like," that is to say, relative to the orbit of a point $x$, most nearby orbits diverge exponentially fast in both forward and backward time. This is why hyperbolicity is a mathematical notion of chaos.

An Anosov diffeomorphism is a smooth invertible map of a compact manifold with the property that the entire space is a hyperbolic set.

Axiom A, which is a larger class, focuses on the part of the system that is not transient. More precisely, a point $x$ in the phase space is said to be "nonwandering" if every neighborhood $U$ of $x$ contains an orbit that returns to $U$. A map is said to satisfy Axiom A if its nonwandering set is hyperbolic and contains a dense set of periodic points.

Definitions in the continuous-time case are analogous: $f$ above is replaced by the time-$t$-maps of the flow, and the tangent spaces now decompose into $E^u \oplus E^0 \oplus E^s$ where $E^0$, which is one dimensional, represents the direction of the flow lines.

A geometric way of detecting (indeed, defining) hyperbolicity is via the cone criterion: at every point there is a cone that is mapped by the differential into the interior of the corresponding cone at the image point, and a "complementary" cone family behaves similarly for the inverse.

Many continuous structures associated with a hyperbolic dynamical system are, in fact, Hölder continuous. (For a function $g$ on a metric space this is defined as the existence of $C, \alpha > 0$ such that $d(g(x), g(y)) \leq Cd(x, y)^\alpha$ whenever $x, y$ are sufficiently close to each other.) In the present article, almost every assertion of continuity could be replaced by one of Hölder continuity. This notion is natural in this context because $x_n \to y$ exponentially fast implies that $g(x_n) \to g(y)$ exponentially fast if $g$ is Hölder continuous.

## Structure and Properties

### Stable and Unstable Manifolds, Local Product Structure

Anosov and Axiom-A systems are defined by the behavior of the differential. Corresponding to the linear structures left invariant by d$f$ are nonlinear structures, namely "stable manifolds" tangent to $E^s$ and "unstable manifolds" tangent to $E^u$.

Thus, associated with an Anosov map are two families of invariant manifolds, each one of which fills up the entire phase space; they are sometimes called the stable and unstable "foliations." The leaves of these foliations are transverse at each point, that is, they intersect at positive angles, forming a kind of (topological) coordinate system. The map $f$ expands distances along the leaves of one of these foliations and contracts distances along the leaves of the other. For Axiom-A systems, one has a similar local product structure or "coordinate system" at each point in the nonwandering set, but the picture is local, and there are gaps: the stable and unstable leaves do not necessarily fill out open sets in the phase space.

There is much interest in determining the fractal dimension (box-counting or Hausdorff, say) of hyperbolic sets. So far the best dimension estimates have been made for stable slices, that is, for the intersection of a stable leaf with the hyperbolic set, and for unstable slices. Because the local coordinate systems describing the local product structure are only known to be continuous, it is not known in general whether the sum of these stable and unstable dimensions gives the dimension of the hyperbolic set (we don't even know whether all stable slices have the same fractal dimension). The problem is that an $\alpha$-Hölder-continuous map can change dimensions by a factor of $\alpha$ or $1/\alpha$. But there is evidence to suggest that something like this "dimension product structure" may often be true – this has been established for a class of solenoids.

### Transitivity and Spectral Decomposition

In addition to these local structures, Axiom-A systems have a global structure theorem known as "spectral decomposition." It says that the nonwandering set of every Axiom-A map can be written as $X_1 \cup \cdots \cup X_r$ where the $X_i$ are disjoint closed invariant sets on which $f$ is topologically transitive, that is, has a dense orbit. The $X_i$ are called "basic sets." Each $X_i$ can be decomposed further into a finite union $\bigcup X_{i,j}$, where each $X_{i,j}$ is invariant and topologically mixing under some iterate of $f$. (Topological transitivity and mixing are irreducibility conditions; transitivity means that there is no proper open invariant subset, and topological mixing says that given two open sets, from some time onward the images of one will always intersect the other.) This decomposition is reminiscent of the corresponding result for finite-state Markov chains.

### Stability

One of the reasons why hyperbolic sets are important is their "robustness": they cannot be perturbed away. More precisely, let $f$ be a map with a hyperbolic set $\Lambda$ which is locally maximal, that is, it is the largest invariant set in some neighborhood $U$. Then for every map $g$ that is $C^1$-near $f$, the largest invariant set $\Lambda'$ of $g$ in $U$ is again hyperbolic; moreover, $f$ restricted to $\Lambda$ is "topologically conjugate" to $g$ restricted to $\Lambda'$. This is mathematical shorthand for saying that not only are the two sets $\Lambda$ and $\Lambda'$ topologically indistinguishable, but the orbit structure of $f$ on $\Lambda$ is indistinguishable from that of $g$ on $\Lambda'$.

The phenomenon above brings us to the idea of "structural stability." A map $f$ is said to be structurally stable if every map $g$ $C^1$-near $f$ is topologically conjugate to $f$ (on the entire phase space). It turns out that a map is structurally stable if and only if it satisfies Axiom A and an additional condition called strong transversality.

### Chains and Shadowing

We discuss next the idea of pseudo-orbits versus real orbits. Letting $d(\cdot,\cdot)$ be the metric, a sequence of points $x_0, x_1, x_2, \ldots$ in the phase space is called an "$\varepsilon$-pseudo-orbit" or a "chain" of $f$ if $d(f(x_i), x_{i+1}) < \varepsilon$ for every $i$. Computer-generated orbits, for example, are pseudo-orbits due to round-off errors. A fact of consequence to people performing numerical experiments is that in hyperbolic systems, small errors at each step get magnified exponentially fast. For example, if the expansion rate is 3 or more, then an $\varepsilon$-error made at one step is at least tripled at each subsequent step, that is, after only $O(|\log \varepsilon|)$ iterates, the error is $O(1)$, and the pseudo-orbit bears no relation to the real one. There is, however, a theorem that says that every pseudo-orbit is "shadowed" by a real one. More precisely, given a hyperbolic set, there is a constant $C$ such that if $x_0, x_1, x_2, \ldots$ is an $\varepsilon$-pseudo-orbit, then there is a phase point $z$ such that $d(x_i, f^i(z)) < C\varepsilon$ for all $i$. Thus, paradoxical as it may first seem, this result asserts that on hyperbolic sets, each pseudo-orbit approximates a real orbit, even though it may deviate considerably from the one with the same initial condition.

The shadowing orbit corresponding to a bi-infinite pseudo-orbit is, in fact, unique. From this, one deduces easily the following Closing Lemma: For any hyperbolic set, there is a constant $C$ such that the following holds: Every finite orbit segment $x, f(x), \ldots, f^{n-1}(x)$ that nearly closes up, that is, $d(x, f^{n-1}(x)) < \varepsilon$ for some small $\varepsilon$, lies within $<C\varepsilon$ of a genuine periodic orbit of period $n$. Thus, hyperbolic sets contain many periodic points.

## Examples

### Anosov Diffeomorphisms

A large class of Anosov diffeomorphisms comes from "linear toral automorphisms," that is, maps of the $n$-dimensional torus induced by $n \times n$ matrices with integer entries, det $= \pm 1$, and no eigenvalues of modulus one. The most popular example is the map obtained from

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

sometimes called the Arnol'd cat map because of an illustration used by Arnol'd. The unstable manifolds are lines parallel to the expanding direction shown in Figure 1 and wrapped around the torus, and the stable manifolds are obtained from the orthogonal lines.

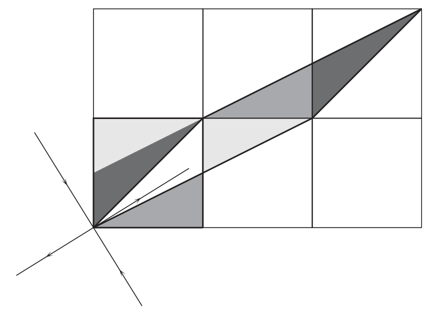


**Figure 1** A hyperbolic toral automorphism. Reproduced from Katok A and Hasselblatt B (2003) *Dynamics*: *A First Course*. Cambridge: Cambridge University Press, with permission from Cambridge University Press.

We remark that due to their structural stability, (nonlinear) perturbations of linear toral automorphisms continue to have the Anosov property. This remark applies also to all of the examples below. In fact, all known Anosov diffeomorphisms are topologically identical to a linear toral automorphism (or a slight generalization of these, infranil-manifold automorphisms).

### Geodesic Flows

Geodesic flows describe free motions of points on manifolds. Let $M$ be a manifold. Given $x \in M$ and a unit vector $v$ at $x$, there is a unique geodesic starting from $x$ in the direction $v$. The geodesic flow $\varphi^t$ is given by $\varphi^t(x, v) = (x', v')$ where $x'$ is the point $t$ units down the geodesic and $v'$ is the direction at $x'$. Geodesic flows on manifolds of strictly negative curvature are the main examples of Anosov flows. They were studied by Hadamard (ca. 1900), Hedlund and Hopf (1930s) considerably before Anosov theory was developed.

### Horseshoes

Smale's horseshoe is the prototypical example of a hyperbolic invariant set. This map, so called because it bends a rectangle $B$ into the shape of a horseshoe and puts it back on top of $B$, is shown in Figure 2. The set $\{x : f^n(x) \in B \text{ for all } n = 0, \pm 1, \pm 2, \ldots\}$ is hyperbolic. It is a two-dimensional Cantor set in $B$. The emergence of this example can be traced back directly to real-world systems.

During World War II, Cartwright and Littlewood worked on relaxation oscillations in radar circuits,

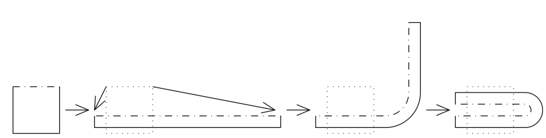


**Figure 2** The horseshoe.

consciously building on Poincaré's work. Further study of the underlying van der Pol equation by Levinson contained the first example of a structurally stable diffeomorphism with infinitely many periodic points. (Structural stability originated in 1937 but began to flourish only 20 years later.) This was brought to the attention of Smale. Inspired by Peixoto's work, who had carried out such a program in dimension 2, Smale pursued a program of studying diffeomorphisms with a view to classification (Smale 1967). Until alerted by Levinson, Smale conjectured that only Morse–Smale systems (which have only finitely many periodic points with stable and unstable sets in general position) could be structurally stable. He eventually extracted the horseshoe from Levinson's work. Smale in turn was in contact with the Russian school, where Anosov systems (then *C*- or *U*-systems) had been shown to be structurally stable, and their ergodic properties were studied by way of further development of the study of geodesic flows in negative curvature.

The appearance of horseshoes in mathematical models of real-world phenomena is quite widespread. Indeed, in a sense this is the mechanism for the production of chaotic behavior, at least in dimension 2. In disguise, one of the earliest appearances of this phenomenon occurred in the prize memoir of Poincaré, where homoclinic tangles gave a first glimpse at the serious dynamical complexity that can arise in the three-body problem in celestial mechanics. If the stable and unstable curves of a hyperbolic fixed point intersect transversely (as in Figure 3a), this engenders further such intersections and produces a complicated web of accumulations of loops or lobes of stable and unstable curves, as shown in Figure 3b. Homoclinic tangles always produce horseshoes by the Smale–Birkhoff theorem, illustrated by Figure 3c, so in trying to solve the three-body problem, Poincaré essentially discovered the possibility of nontrivial hyperbolic behavior (*see* Homoclinic Phenomena).

A related appearance of horseshoes in this context is in the work of Alekseev, who used their presence to show that capture of celestial bodies can indeed occur.

#### Solenoids

Finally we mention the solenoid, which is an example of an Axiom-A attractor (*see* Figure 4). Here the map $f$ is defined on a solid torus $M = S^1 \times D_2$, where $D_2$ is a two-dimensional disk. It is easiest to describe it in two steps: first it maps $M$ into a long thin solid torus, which is then put inside $M$

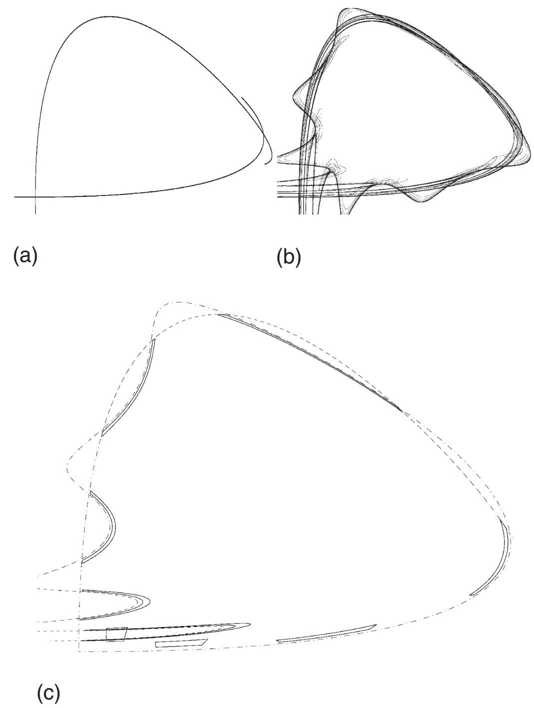



**Figure 3** Homoclinic tangles produce horseshoes. Reproduced from Katok A and Hasselblatt B (2003) *Dynamics: A First Course*. Cambridge: Cambridge University Press, with permission from Cambridge University Press.

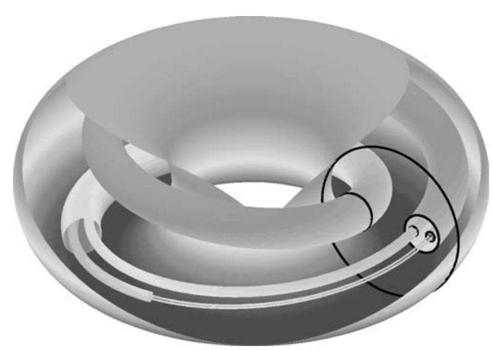


**Figure 4** The solenoid. Reproduced from Katok A and Hasselblatt B (2003) *Dynamics: A First Course*. Cambridge: Cambridge University Press, with permission from Cambridge University Press.

winding around the $S^1$ direction twice. The attractor is given by $\Lambda = \bigcap_{n \geq 0} f^n(M)$.

## Symbolic Coding of Orbits and Ergodic Theory

An important tool for studying the orbit structure of Axiom-A systems is the "Markov partition," constructed for Anosov systems by Sinai and extended to Axiom-A basic sets by Bowen. Given a partition $\{R_1, \ldots, R_k\}$ of the phase space, there is a natural way to attach to each point $x$ in the phase space a

sequence of symbols, namely $(\ldots, a_{-1}, a_0, a_1, a_2, \ldots)$, where $a_i \in \{1, 2, \ldots, k\}$ is the name of the partition element containing $f^i(x)$, that is, $f^i(x) \in R_{a_i}$ for each $i$. In general, not all sequences are realized by orbits of $f$. Markov partitions are designed so that the set of symbol sequences that correspond to real orbits has Markovian properties; it is called a shift of finite type.

The ergodic theory of Axiom-A systems has its origins in statistical mechanics. In a 1D lattice model in statistical mechanics, one has an infinite array of sites indexed by the integers; at each site, the system can be in any one of a finite number of states. Thus, the configuration space for a 1D lattice model is the set of bi-infinite sequences on a finite alphabet. Identifying this symbol space with the one coming from Markov partitions, Sinai and Ruelle were able to transport some of the basic ideas from statistical mechanics, including the notions of Gibbs states and equilibrium states, to the ergodic theory of Axiom-A systems.

The notion of equilibrium states, which is equivalent to Gibbs states for Axiom-A systems, has the following meaning in dynamical systems in general: given a potential function $\varphi$, an invariant measure is said to be an equilibrium state if it maximizes the quantity

$$h_\mu(f) - \int \varphi \mathrm{d}\mu$$

where $h_\mu(f)$ denotes the Kolmogorov–Sinai entropy of $f$ and the supremum is taken over all $f$-invariant probability measures $\mu$. In particular, when $\varphi = 0$, this measure is the measure that maximizes entropy; and when $\varphi = \log|\det(\mathrm{d}f|_{E^u})|$, it is the Sinai–Ruelle–Bowen (SRB) measure. From the physical or observational point of view, SRB measures are the most important invariant measures for dissipative dynamical systems because if $f$ is a diffeomorphism of a compact manifold $M$ and $\Lambda$ a transitive Axiom-A attractor with basin $U$, for example, $\Lambda = U = M$, then for Lebesgue-a.e. $x \in U$ and for every $\varphi \in C^0(M)$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi(f^i(x)) = \int \varphi d\mu$$

that is, Lebesgue-a.e. point is $\mu$-typical. Thus, while Axiom-A attractors will have chaotic motions, they are statistically coherent in that the asymptotic distribution of any typical orbit is given by the SRB measure.

## Periodic Points and Their Growth Properties

We discuss briefly some further results related to the abundance of periodic points in Axiom-A systems.

For an Axiom-A diffeomorphism $f$, if $P(n)$ is the number of periodic points of period $\leq n$, then $P(n) \sim e^{hn}$, where $h$ is the topological entropy of $f$. That is to say, the dynamical complexity of $f$ is reflected in its periodic behavior. An analogous result holds for Axiom-A flows. This asymptotic behavior is known to remarkably fine accuracy (Margulis 2004), and these developments used the dynamical zeta function, which sums up the periodic information of a system. In the discrete-time case, $\zeta(z) := \exp\sum_{n=1}^{\infty} P(n)z^n/n$ has been shown to be a rational function analytic on $|z| < e^{-h}$. In the continuous-time case, the zeta function is given by $\zeta(z) := \prod_\gamma (1 - \exp(-zl(\gamma)))^{-1}$, where the product is taken over all (nonstationary) periodic orbits $\gamma$ and $l(\gamma)$ is the smallest positive period of $\gamma$. This function is known to be meromorphic on a certain domain, but the location of its poles, which are intimately related to correlation decay properties of the system, remains one of the yet unresolved issues in Axiom-A theory.

## Partial Hyperbolicity and Dominated Splitting

There are various ways in which the notion of hyperbolicity described above, which we will henceforth refer to as "uniform hyperbolicity," can be extended beyond the one presented so far. This can be done with a view to weakening the conditions under which some of the salient properties of hyperbolic dynamical systems appear. The study of partially hyperbolic dynamical systems and that of dynamical systems possessing a dominated splitting is of this type. Further below, we describe a different extension motivated more by a desire to bring the results and methods of hyperbolic dynamics to bear on systems that are closer to some physical situations. This led to the study of nonuniformly hyperbolic dynamical systems.

If one views hyperbolicity as requiring that the spectrum of expansion and contraction rates is separated into two components by the unit circle, then one can consider systems where this separation is provided by a circle centered at 0 whose radius may not be 1 (partial hyperbolicity in the broad sense), or by two circles centered at 0 of which one has radius less than 1 and the other has radius greater than 1, with possibly a third component of the spectrum in the annulus between these (absolute partial hyperbolicity). Further weakenings are obtained by controlling not the whole spectrum in this absolute way, but rather ratios of expansion and contraction rates along orbits (dominated splitting

and relative partial hyperbolicity, respectively). Among the motivations for these weakenings are the desire to understand which systems are topologically transitive and robustly so (stable transitivity), and to understand which ergodic volume-preserving systems remain ergodic if perturbed within the space of volume-preserving systems (stable ergodicity).

### Pseudohyperbolicity

Let $f$ be a smooth invertible map. A compact invariant set of $f$ is said to be partially hyperbolic in the broad sense if at every point in this set, the tangent space splits into a direct sum of two subspaces $E^{\mathrm{u}}$ and $E^{\mathrm{s}}$ with the property that these subspaces are invariant under the differential d$f$, that is, $\mathrm{d}f(x)E^{\mathrm{u}}(x) = E^{\mathrm{u}}(f(x)), \mathrm{d}f(x)E^{\mathrm{s}}(x) = E^{\mathrm{u}}(f(x))$, and that there are constants $0 < \lambda < \mu, c > 0$ such that if $v \in E^{\mathrm{s}}(x)$ for some $x$ then $\|\mathrm{d}f^n v\| \le c\lambda^n \|v\|$ for $n = 1, 2, \ldots$ and if $v \in E^{\mathrm{u}}(x)$ for some $x$ then $\|\mathrm{d}f^{-n} v\| \le c\mu^{-n}\|v\|$ for $n = 1, 2, \ldots$. This is sometimes also referred to as the existence of a $(\lambda, \mu)$-splitting or pseudohyperbolicity.

### Dominated Splitting

A further weakening of this condition replaces these absolute estimates by relative ones. Let $f$ be a smooth invertible map. A compact invariant set of $f$ is said to admit a dominated splitting if at every point in this set, the tangent space splits into a direct sum of two subspaces $E^{\mathrm{u}}$ and $E^{\mathrm{s}}$ with the property that these subspaces are invariant under the differential and there are constants $\lambda \in (0, 1), c > 0$ such that if $u \in E^{\mathrm{u}}(x)$ and $v \in E^{\mathrm{s}}(x)$ for some $x$ then $\|\mathrm{d}f^n v\| / \|\mathrm{d}f^n u\| \le c\lambda^n$ for $n = 1, 2, \ldots$.

The presence of a dominated splitting has been found to yield substantial information pertinent to stability of such systems, and it plays a significant role in a program of research aiming at a classification of generic diffeomorphisms up to topological conjugacy and specifically motivated by the "Palis conjecture," which aims to describe that classification. With respect to inferring topological and ergodic (i.e., statistical) properties of the orbit structure, the stricter notion of partial hyperbolicity (in the narrow sense below) is more commonly used, but in this respect the presence of a dominated splitting is also of interest because there is evidence in support of the conjecture that stable ergodicity implies the presence of a dominated splitting.

### Partial Hyperbolicity

Let $f$ be a smooth invertible map. A compact invariant set of $f$ is said to be (absolutely) partially hyperbolic if at every point in this set, the tangent space splits into a direct sum of unstable, central, and stable directions $E^{\mathrm{u}}, E^{\mathrm{c}}$, and $E^{\mathrm{s}}$ with the property that these subspaces are invariant under the differential d$f$ and that there exist numbers $C > 0$,

$$\begin{gathered} 0 < \lambda_1 \le \mu_1 < \lambda_2 \le \mu_2 < \lambda_3 \le \mu_3 \\ \text{with } \mu_1 < 1 < \lambda_3 \end{gathered} \quad [1]$$

such that if $v \in E^{\mathrm{s}}(x), w \in E^{\mathrm{c}}(x), u \in E^{\mathrm{u}}(x)$, $n = 1, 2, \ldots$, then

$$C^{-1}{\lambda_1}^n\|v\| \le \|d_x f^n(v)\| \le C{\mu_1}^n\|v\|$$
$$C^{-1}{\lambda_2}^n\|w\| \le \|d_x f^n(w)\| \le C{\mu_2}^n\|w\|$$
$$C^{-1}{\lambda_3}^n\|u\| \le \|d_x f^n(u)\| \le C{\mu_3}^n\|u\|$$

In this case, we set $E^{\mathrm{cs}} := E^{\mathrm{c}} \oplus E^{\mathrm{s}}$ and $E^{\mathrm{cu}} := E^{\mathrm{c}} \oplus E^{\mathrm{u}}$. Following Burns–Wilkinson, we say that $f$ is "center-bunched" if $\max\{\mu_1, \lambda_3^{-1}\} < \lambda_2/\mu_2$.

As in the case of (uniformly) hyperbolic dynamical systems, the sub-bundles $E^{\mathrm{s}}$ and $E^{\mathrm{u}}$ are integrable to stable and unstable foliations $W^{\mathrm{s}}$ and $W^{\mathrm{u}}$. It is not automatic that the center-stable sub-bundle $E^{\mathrm{cs}}$ and the center-unstable sub-bundle $E^{\mathrm{cu}}$ are tangent to foliations $W^{\mathrm{cs}}$ and $W^{\mathrm{cu}}$; if this happens to be the case, the partially hyperbolic system is said to be "dynamically coherent."

Partial hyperbolicity can also be defined by a cone criterion, with suitable adaptations.

### Stable Ergodicity and Transitivity

Partial hyperbolicity was introduced as a means of providing just enough hyperbolicity to render a dynamical system ergodic or topologically transitive. These are both irreducibility conditions, and to obtain these, one rules out a Cartesian product situation by assuming something like essential accessibility: almost every two points (in the sense of volume viewed as a measure) can be connected by a curve consisting of a finite concatenation of arcs, each of which lies entirely in one stable or unstable leaf. A celebrated result in this field is in its original form (with a much stronger center-bunching assumption) due to Pugh and Shub: suppose a volume-preserving diffeomorphism is partially hyperbolic on the entire manifold. If it is dynamically coherent and center bunched and has essential accessibility, then it is ergodic (Hasselblatt and Pesin 2006).

One of the motivating aims of this theory was to obtain nonhyperbolic volume-preserving systems that are stably ergodic, that is, for which all volume-preserving $C^1$-small perturbations are also ergodic. If, in addition to the above, one assumes that essential accessibility also persists under such

perturbations and that the center bundle $E^c$ is integrable to a center foliation $W^c$ that is smooth (or "plaque-expansive"), then ergodicity is indeed stable (Hasselblatt and Pesin). There are quite a few natural examples where these assumptions hold.

While essential accessibility does not always hold, it is fairly common. The stronger property of accessibility (that any two points can be connected, not only almost every two points) is conjectured to be stable under $C^1$-perturbations and has been shown to hold for an open dense set of partially hyperbolic systems with respect to the $C^1$-topology.

Ergodicity is a measure-theoretic irreducibility notion, and topological transitivity is the topological counterpart. It can also be obtained from accessibility: a partially hyperbolic volume-preserving diffeomorphism with the accessibility property is topologically transitive (in fact, almost every orbit is dense).

There are interesting converse results as well. Any stably transitive diffeomorphism exhibits a dominated splitting. Moreover, in dimension 2 it is hyperbolic and in dimension 3 it is partially hyperbolic in the broad sense.

## Nonuniform Hyperbolicity

Applications have motivated weakening assumptions of uniform hyperbolicity to require only that "many" individual orbits exhibit hyperbolic behavior, without assuming that there are any uniform estimates on the degree of hyperbolicity.

To measure the asymptotic contraction or expansion of a vector on an exponential scale, one defines the Lyapunov exponent of a (nonzero) tangent vector $v$ at $x$ for the map $f$ to be

$$\lambda(x, v) := \lim_{n\to\infty} (1/n) \log \|Df^n(v)\| \qquad [2]$$

whenever this limit exists. Note that being positive indicates asymptotic expansion of the vector, whereas negative exponents correspond to contracting vectors. This defines a measurable but, save for exceptional circumstances, discontinuous function of $x$ and $v$. It is relatively easy to see that for a given point $x$ the function $\lambda(x, \cdot)$ can only take finitely many values, so it is natural to define nonuniform hyperbolicity as the property of having all of these finitely many values nonzero for "most" points. Given that $\lambda$ is measurable, it is natural to define "most" by using a measure that is invariant under the map $f$. Therefore, the theory of nonuniformly hyperbolic dynamical systems, much of which is due to Pesin, is based on measure theory throughout.

The fundamental fact on which this theory is based is the "Oseledets multiplicative ergodic theorem," which says that for a $C^1$-diffeomorphism of a compact Riemannian manifold the set of Lyapunov-regular points has full measure with respect to any $f$-invariant Borel probability measure.

For a Lyapunov-regular point the limit [2] exists for all $v$, so this theorem tells us that no matter which invariant measure we consider, the limit [2] makes sense for all tangent vectors at points $x$ outside a null set. (One should add that this small "bad" set can be somewhat substantial; for example, its Hausdorff dimension is usually that of the whole space.)

Accordingly, one then defines a measure to be hyperbolic if at almost every point the limit [2] is nonzero for all vectors. In this case, one says that "$f$ has nonzero Lyapunov exponents." This property can also be obtained from a cone criterion, but here the family of cones may only be invariant and eventually strictly invariant, that is, there is a cone field such that cones are mapped to cones (but not necessarily into the interior of cones), and for almost every point there is an iterate that maps a cone strictly inside the cone at the image point (i.e., into the interior). Which iterate is needed is allowed to depend on the point (*see* Hyperbolic Billiards).

It is good to keep in mind that a hyperbolic measure may be concentrated on a single point, say, in which case there is not much gained by this approach. The theory is of great interest, however, if the measure is equivalent to volume or is the "physical measure" on an attractor.

Examples of this sort are fairly common, indeed any smooth compact Riemannian manifold other than the unit circle admits a volume-preserving Bernoulli diffeomorphism with nonzero Lyapunov exponents (Dolgopyat and Pesin 2002) (and every compact smooth Riemannian manifold of dimension at least 3 carries a volume-preserving Bernoulli flow for which at almost every point the only zero Lyapunov exponent is the one in the flow direction (Hu *et al.* 2004)).

Structurally, these systems exhibit many of the features seen in uniformly hyperbolic ones (e.g., stable manifolds), but instead of being continuous these are now measurable. There are, however, (noninvariant) sets of arbitrarily large measure on which these structures are continuous. This provides a handle for pushing some of the uniform theory to this context.

There are some topological results in this area, of which one of the more remarkable ones is that any surface diffeomorphism with positive entropy contains a horseshoe. Much of the current research is

directed at the ergodic theory of these systems. A central result from the initial development of the theory is that while these systems may not be ergodic, the ergodic components are (a.e. equal to) open sets, so in particular there are at most countably many of them.

One natural question is whether nonuniformly hyperbolic systems have SRB measures, and it is answered on a case-by-case basis. There are even benign examples where this fails to be the case, but for some realistic systems, such as the Lorenz and Hénon attractors, this has been established.

Because they preserve volume, this is not an issue for billiard systems, (*see* Hyperbolic Billiards), that is, the free motion of a point mass in a cavity with elastic boundary collisions. This describes not just a toy model, but also the phase space and dynamics of a gas of convex rigid bodies. Such a gas of hard spheres in a rectangular box is semidispersing and has been studied intensely. It is now known to be hyperbolic and hoped to be ergodic. (The latter would provide a solid foundation for statistical mechanics, at least for the case of spherical molecules.) A gas of nonspherical convex rigid bodies is also a point billiard, but it is not semidispersing, which puts it beyond the range of readily available techniques for establishing ergodicity.

## Further Remarks

The historical remarks made here are significantly expanded in Hasselblatt (2002), which contains some references to yet more detailed sources as well as more detail about uniformly hyperbolic dynamical systems in a concise form. A concise but reasonably comprehensive and current account of partially hyperbolic dynamics is in Hasselblatt and Pesin, and an authoritative full presentation is in Pesin (2004). A survey of nonuniformly hyperbolic dynamics is given in Barreira and Pesin (2006), and the definitive treatment is given by Barreira *et al.*. A textbook presentation of (not only) hyperbolic dynamics is in Katok and Hasselblatt (1995) as well as Hasselblatt and Katok (2003), and much current research, including on all subjects discussed here, is surveyed in Handbook.

*See also:* Ergodic Theory; Fractal Dimensions in Dynamics; Generic Properties of Dynamical Systems; Homoclinic Phenomena; Hyperbolic Billiards; Inviscid Flows; Lyapunov Exponents and Strange Attractors; Regularization for Dynamical Zeta Functions; Singularity and Bifurcation Theory; Symmetry and Symmetry Breaking in Dynamical Systems.

## Further Reading

Barreira L, Katok A, and Pesin Y Smooth ergodic theory and nonuniformly hyperbolic dynamics (in preparation).

Barreira L and Pesin Y (2006) Smooth ergodic theory and nonuniformly hyperbolic dynamics. In: Hasselblatt B and Katok A (eds.) *Handbook of Dynamical Systems*, vol. 1B, pp. 57–263. Amsterdam: Elsevier.

Bowen R (1975) Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms. *Lecture Notes in Mathematics*, vol. 470. Berlin–New York: Springer.

Dolgopyat D and Pesin Y (2002) Every compact manifold carries a completely hyperbolic diffeomorphism. *Ergodic Theory and Dynamical Systems* 22(2): 409–435.

Hasselblatt B and Kotak A (eds.) (2002) *Handbook of Dynamical Systems*, 1B: 2006, Hasselblatt B and Kotak A (eds.); 2: 2002, Fiedler B (ed.); 3: in preparation, Broer H, Hasselblatt B, and Takens F (eds.).

Hasselblatt B (2002) Hyperbolic dynamics. In: Hasselblatt B and Katok A (eds.) *Handbook of Dynamical Systems*, vol. 1A, pp. 239–319. Amsterdam: Elsevier.

Hasselblatt B and Katok A (2003) *Dynamics: A First Course*. New York: Cambridge University Press.

Hasselblatt B and Katok A (2002) Principal structures. In: Hasselblatt B and Katok A (eds.) *Handbook of Dynamical Systems*, vol. 1A, pp. 1–203. Amsterdam: Elsevier.

Hasselblatt B and Pesin Y (2006) Partially hyperbolic dynamical systems. In: Hasselblatt B and Kotak A (eds.) *Handbook of Dynamical Systems*, vol. 1B, pp. 1–55. Amsterdam: Elsevier.

Hu H, Pesin Y, and Talitskaya A (2004) Every compact manifold carries a hyperbolic Bernoulli flow. In: Brin M, Hasselblatt B, and Pesin Y (eds.) *Modern Dynamical Systems and Applications*, pp. 347–358. Cambridge: Cambridge University Press.

Katok A and Hasselblatt B (1995) *Introduction to the Modern Theory of Dynamical Systems*. New York: Cambridge University Press.

Margulis G (2004) *On some aspects of the theory of Anosov systems. With a survey by Richard Sharp: Periodic orbits of hyperbolic flows. Translated from the Russian by Valentina Vladimirovna Szulikowska. Springer Monographs in Mathematics*. Berlin: Springer.

Pesin Y (2004) Lectures on Partial Hyperbolicity and Stable Ergodicity. *Zurich Lectures in Advanced Mathematics*. Zürich: European Mathematical Society.

Smale S (1967) Differentiable dynamical systems. *Bulletin of the American Mathematical Society* 73: 747–817.

I

# Image Processing: Mathematics

**G Aubert**, Université de Nice Sophia Antipolis, Nice, France
**P Kornprobst**, INRIA, Sophia Antipolis, France



Our society is often designated as being an "information society." It could also be defined as an "image society." This is not only because image is a powerful and widely used medium of communication, but also because it is an easy, compact, and widespread way to represent the physical world. If we think about it, it is indeed striking to realize just how much images are omnipresent in our lives through numerous applications such as medical and satellite imaging, videosurveillance, cinema, robotics, etc.

Many approaches have been developed to process these digital images, and it is difficult to say which one is more natural than the other. Image processing has a long history. Maybe the oldest methods come from 1D signal processing techniques. They rely on filter theory (linear or not), on spectral analysis, or on some basic concepts of probability and statistics. For an overview, we refer the interested reader to the book by Gonzalez and Woods (1992).

In this article, some recent mathematical concepts will be revisited and illustrated by the image restoration problem, which is presented below. We first discuss stochastic modeling which is widely based on Markov random field theory and deals directly with digital images. This is followed by a discussion of variational approaches where the general idea is to define some cost functions in a continuous setting. Next we show how the scale space theory is connected with partial differential equations (PDEs). Finally, we present the wavelet theory, which is inherited from signal processing and relies on decomposition techniques.

## Introduction

As in the real world, a digital image is composed of a wide variety of structures. **Figure 1** shows different kinds of "textures," progressive or sharp contours, and fine objects. This gives an idea of the complexity of finding an approach that allows to cope with the different structures at the same time. It also highlights the discrete nature of images which will be handled differently depending on the chosen mathematical tools. For instance, PDEs based approaches are written in a continuous setting, referring to analogous images, and once the existence and the uniqueness of the solution have been proved, we need to discretize them in order to find a numerical solution. On the contrary, stochastic approaches will directly consider discrete images in the modeling of the cost functions.

### The Image Restoration Problem

It is well known that during formation, transmission, and recording processes images deteriorate. Classically, this degradation is the result of two phenomena. The first one is deterministic and is related to the image acquisition modality, to possible defects of the imaging system (e.g., blur created by an incorrect lens adjustment or by motion). The second phenomenon is random and corresponds to the noise coming from any signal transmission. It can also come from image quantization. It is important to choose a degradation model as close as possible to reality. The random noise is usually modeled by a probabilistic distribution. In many cases, a Gaussian distribution is assumed. However, some applications require more specific ones, like the gamma distribution for radar images (speckle noise) or the Poisson distribution for tomography. Unfortunately, it is usually impossible to identify the kind of noise involved for a given real image.

A commonly used model is the following. Let $u : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ be an original image describing a real scene, and let $f$ be the observed image of the same scene (i.e., a degradation of $u$). We assume that

$$f = Au + \eta \qquad [1]$$

where $\eta$ stands for a white additive Gaussian noise and $A$ is a linear operator representing the blur (usually a convolution). Given $f$, the problem is

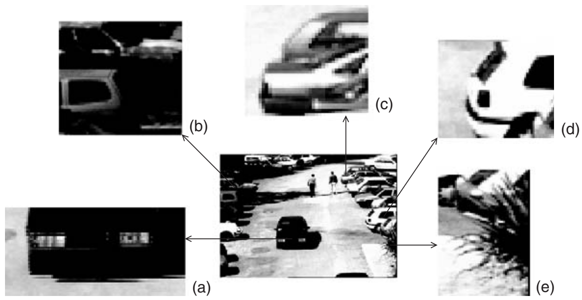



**Figure 1** Digital image example. ↷ the close-ups show examples of low resolution, low contrasts, graduated shadings, sharp transitions, and fine elements. (a) low resolution, (b) low contrasts, (c) graduated shadings, (d) sharp transitions, and (e) fine elements.

then to reconstruct $u$ knowing [1]. This problem is ill-posed, and we are able to carry out only an approximation of $u$. In this article, we will focus on the simplified model of pure denoising:

$$f = u + \eta \qquad [2]$$

## The Probabilistic Approach

### The Bayesian Framework

In this section, we show how the problem of pure denoising, that is, recovering $u$ from the equation $f = u + \eta$ knowing only some statistical information on $\eta$ can be solved by using a probabilistic approach. In this context, $f$, $u$, and $\eta$ are considered as random variables. The general idea for recovering $u$ is to maximize some prior probability. Most models involve two parts: a prior model of possible restored images $u$ and a data model expressing consistency with the observed data.

- The prior model is given by a probability space $(\Omega_u, p)$, where $\Omega_u$ is the set of all values of $u$. The model is specified by giving the probability $p(u)$ on all these values.
- The data model is a larger probability space $(\Omega_{u,f}, p)$, where $\Omega_{u,f}$ is the set of all possible values of $u$ and all possible values of the observed image $f$. This model is completed by giving the conditional probability $p(f/u)$ of any image $f$ given $u$, resulting in the joint probabilities $p(f,u) = p(f/u)p(u)$. Implicitly, we assume that the spaces $(\Omega_u)$ and $(\Omega_{u,f})$ are finite although huge.

The next step is to use a Bayesian approach introduced in image processing by Besag (1974) and Geman and Geman (1984). The probabilities $p(u)$ and $p(f/u)$ are supposed to be known and, given an observed image $f$, we seek the image $u$ which maximizes the conditional *a posteriori* probability $p(u/f)$ (MAP: Maximum A Posteriori). Thanks to the Bayes' rule, we have

$$p(u/f) = \frac{p(f/u)p(u)}{p(f)} \qquad [3]$$

Let us explain the meaning of the different terms in [3]:

- The term $p(f/u)$ expresses the probability, the likelihood, that an image $u$ is realized in $f$. It also quantifies the lack of total precision of the model and the presence of noise.
- The term $p(u)$ expresses our incomplete *a priori* information about the ideal image $u$ (it is the probability of the model, i.e., the propensity that $u$ be realized independently of the observation $f$).
- The term $p(f)$ which is the probability to observe $f$ is a constant and does not play any role when maximizing the conditional probability $p(u/f)$ with respect to $u$.

Let us remark that the problem $\max_u p(u/f)$ is equivalent to $\min_u E(u) = -\log p(f/u) - \log p(u)$. So Bayesian models lead to a minimization process.

Then the main question is how to assign these probabilities? The easiest probability to determine is $p(f/u)$. If the images $u$ and $f$ consist in a set of values $u = (u_{i,j}), i,j = 1, N$ and $f = (f_{i,j}), i,j = 1, N$, we suppose the conditional independence of $(f_{i,j}/u_{i,j})$ in any pixel:

$$p(f/u) = \prod_{i=1}^{N} p(f_{i,j}/u_{i,j})$$

and if the restoration model is of the form $f = u + \eta$ where $\eta$ is a white Gaussian noise with variance $\sigma^2$, then

$$p(f_{i,j}/u_{i,j}) = \frac{1}{\sqrt{2\pi\sigma}} \exp - \frac{(f_{i,j} - u_{i,j})^2}{2\sigma^2}$$

and

$$p(f/u) = \frac{1}{(2\pi\sigma)^{N/2}} \exp - \sum_{i,j}^{N} \frac{(f_{i,j} - u_{i,j})^2}{2\sigma^2}$$

Therefore, at this stage, the MAP reduces to minimize

$$E(u) = K_\sigma \|f - u\|^2 - \log p(u) \qquad [4]$$

where $\|.\|$ stands for the Euclidean norm on $\mathbb{R}^{N^2}$ and $K_\sigma$ is a constant. So, it remains now to assign a probability law $p(u)$. To do that, the most common way is to use the theory of Markov random fields (MRFs).

### The Theory of Markov Random Fields

In this approach, an image is described as a finite set $S$ of sites corresponding to the pixels. For each site, we associate a descriptor representing the state of the site, for example, its gray level. In order to take into account local interaction between sites, one needs to endow $S$ with a system of neighborhoods $\mathcal{V}$.

**Definition 1** For each site $s$, we define its neighborhood $\mathcal{V}(s)$ as:

$$\mathcal{V}(s) = \{t\} \text{ such that } s \notin \mathcal{V}(s) \text{ and } t \in \mathcal{V}(s) \Rightarrow s \in \mathcal{V}(t)$$

Then we associate to this neighborhood system the notion of clique: a clique is either a singleton or a set of sites which are all neighbors of each other. Depending on the neighborhood system, the family of cliques will be different and involve more and less sites. We will denote by $\mathcal{C}$ the set of all the cliques relative to a neighborhood system $\mathcal{V}$ (see Figure 2).

Before introducing the general framework of MRFs, let us define some notations. For a site $s$, $X_s$ will stand for a random variable taking its values in some set $\mathcal{E}$ (e.g., $\mathcal{E} = \{0, 1, \ldots, 255\}$) and $x_s$ will be a realization of $X_s$ and $x^s = (x_t)_{t \neq s}$ will denote an image configuration where site $s$ has been removed. Finally, we will denote by $X$ the random variable $X = (X_s, X_t, \ldots)$ with values in $\Omega = \mathcal{E}^{|S|}$.

**Definition 2** We say that $X$ is an MRF if the local conditional probability at a site $s$ is only a function of $\mathcal{V}(s)$, that is,

$$p(X_s = x_s / X^s = x^s) = p(X_s = x_s / x_t, t \in \mathcal{V}(s))$$

Therefore, the gray level at a site depends only on gray levels of neighboring pixels. Now we give the following fundamental theorem due to Hammersley–Clifford (Besag 1974) which states the equivalence between MRFs and Gibbs fields.

**Theorem 1** *Let us suppose that $S$ is finite, $\mathcal{E}$ is a discrete set and for all $x \in \Omega = \mathcal{E}^{|S|}, p(X = x) > 0$, then $X$ is an MRF relatively to a system of neighborhoods $\mathcal{V}$ if and only if there exists a family of potential functions $(V_c)_{c \in \mathcal{C}}$ such that $p(x) = (1/Z)\exp(-\sum_{c \in \mathcal{C}} V_c(x))$.*

*The function $V(x) = \sum_{c \in \mathcal{C}} V_c(x)$ is called the energy potential or the Gibbs measure and $Z$ is a normalizing constant: $Z = \exp(-\sum_{x \in \Omega} V(x))$.*

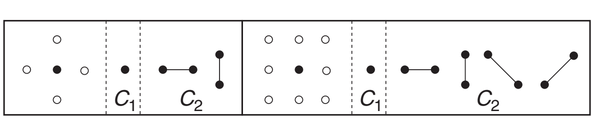



**Figure 2** Examples of neighborhood system and cliques.

If, for example, the collection of neighborhoods is the set of 4-neighbors, then the theorem says that $V(x) = \sum_{c = \{s\} \in C_1} V_c(x_s) + \sum_{c = \{(s,t)\} \in C_2} V_c(x_s, x_t)$.

### Application to the Denoising Problem

Now, given this theorem we can reformulate, thanks to [4], the restoration problem (with the change of notation $u = x$ and $u_s = x_s$): find $u$ minimizing the global energy

$$E(u) = K_\sigma \|f - u\|^2 + V(u) \qquad [5]$$

The next step is now to precise the Gibbs measure. In restoration, the potential $V(u)$ is often dedicated to impose local regularity constraints, for example, by penalizing differences between neighbors. This can be modeled using cliques of order 2 in the following manner:

$$V(u) = \beta \sum_{(s,t) \in \mathcal{C}_2} \phi(u_s - u_t)$$

where $\phi$ is a given real function. This term penalizes the difference of intensities between neighbors which may come from an edge or some noise. This discrete cost function is very similar to the gradient penalty terms in the continuous framework (see the next section). The resulting final energy is (sometimes $E(u)$ is written $E(u/f)$)

$$E(u) = K_\sigma \sum_{s \in S} (f_s - u_s)^2 + \beta \sum_{(s,t) \in \mathcal{C}_2} \phi(u_s - u_t)$$

where the constant $\beta$ is a weighting parameter which can be estimated.

The difficulty in choosing the strength of the penalty term defined by $\phi$ is to be able to penalize the noise while keeping the most salient features, that is, edges. Historically, the function $\phi$ was first chosen as $\phi(z) = z^2$ but this choice is not good since the resulting regularization is too strong introducing a blur in the image and loss of the edges. A better choice is $\phi(z) = |z|$ (Rudin *et al.* 1992) or a regularized version of this function. Of course, other choices are possible depending on the considered application and the desired degree of smoothness.

In this section, it has been shown how to model the restoration problem through MRFs and the Bayesian framework. Numerically, two main types of algorithms can be used to minimize the energy: deterministic algorithms and stochastic algorithms. The former are generally used when the global energy is strictly convex (e.g., algorithms based on

gradient descent). The latter are rather used when $E(u)$ is not convex. There are stochastic minimization algorithms mainly based on simulated annealing. Their main interest is that they always converge (almost surely) to a minimizer (this is not the case for deterministic algorithms which give only local minimizers) but they are often strongly time consuming.

We refer the reader to Li (1995) for more details about MRFs and Bayesian framework and Kirkpatrick *et al.* (1983) for more information on stochastic algorithms.

## The Variational Approach

### Minimizing a Cost Function over a Functional Space

One important issue in the previous section was the definition of $p(u)$ which gives some *a priori* on the solution. In the variational approach, this idea is also present but the way to infer it is in fact to define the more suitable functional space that describes images and their geometrical properties. The choice of a functional space sets a norm which in turn will constrain the solution to a certain smoothness.

We illustrate this idea in this section on the denoising problem [2] which can be seen as a decomposition one. This means that given the observation $f$, we look for $u$ and $\eta$ such that $f = u + \eta$, where $\eta$ incorporates all oscillations, that is, noise, and also texture. Let us define a functional to be minimized which takes into account the data $f$ and possibly some statistical informations about $\eta$:

$$\min_{(u,\eta)}\{\phi(|u|_E) \text{ such that } \psi(|\eta|_G) = \sigma \text{ with } f = u + \eta\} \qquad [6]$$

This formulation means that we look, among all decompositions $f = u + \eta$, for the one which minimizes $\phi(|u|_E)$ under the constraint $\psi(|\eta|_G) = \sigma$. Banach spaces $E$ and $G$, and functions $\phi$ and $\psi$ will be discussed in the next subsection. Since a minimization problem under constraints can be expressed with an additional term weighted by a Lagrange multiplier, the formulation [6] can be rewritten as:

$$\min_{(u,\eta)}\{\phi(|u|_E) + \lambda\psi(|\eta|_G); f = u + \eta\} \qquad [7]$$

A similar writing consists in replacing $\eta$ by $f - u$ so that [7] rewrites

$$\min_{u}\{\phi(|u|_E) + \lambda\psi(|f - u|_G)\} \qquad [8]$$

which is the classical formulation in image restoration. From a numerical point of view, the minimization is usually carried out by solving the associated Euler equations but this may be a difficult task. The main concern is the search for $E$ and $G$ and their norm (or seminorm). It is guided by the choice that an image $u$ is composed of various geometric structures (homogeneous regions, edges) while $\eta = f - u$ represents oscillations (noise and textures).

### Examples of Functional Spaces

In this section, we revisit some possible choices of functional spaces summarized in **Table 1**.

The first case (a) was inspired by the classical Tikhonov regularization. The functional space $H^1(\Omega)(\Omega \subset \mathbb{R}^2)$ is the space of functions in $L^2(\Omega)$ such that the distributional gradient $Du$ is in $L^2(\Omega)$. Unfortunately, functions in $H^1(\Omega)$ do not admit discontinuities across curves and this is a major problem with respect to image analysis since images are made of smooth patches separated by sharp variations.

Considering the problem reported in (a), Rudin *et al.* (1992) proposed to work on $\mathrm{BV}(\Omega)$, the space of bounded variations (BV) Ambrosio *et al.* (2000) defined by

$$\mathrm{BV}(\Omega) = \left\{u \in L^1(\Omega); \int_\Omega |Du| < \infty\right\}$$
$$\text{with } \int_\Omega |Du| = \sup\left\{\int_\Omega u\,\mathrm{div}\varphi\,dx;\ \varphi = (\varphi_1, \varphi_2, \ldots, \varphi_N) \in \mathcal{C}_0^1(\Omega)^N,\ |\varphi|_{L^\infty(\Omega)} \le 1\right\} \qquad [9]$$

**Table 1** Examples of functional spaces and their norm (see model [8])

| Model | $E$ and $\lvert u\rvert_E$ | $\phi(t)$ | $G$ and $\lvert u\rvert_G$ | $\psi(t)$ |
|---|---|---|---|---|
| (a) | $H^1(\Omega), \lvert u\rvert_E = \left(\int_\Omega \lvert\nabla u\rvert^2\,dx\right)^{1/2}$ | $t^2$ | $L^2(\Omega)$ with its usual norm | $t^2$ |
| (b) | $BV(\Omega), \lvert u\rvert_E = \int_\Omega \lvert Du\rvert$ | $t$ | $L^2(\Omega)$ with its usual norm | $t^2$ |
| (c) | $BV(\Omega), \lvert u\rvert_E = \int_\Omega \lvert Du\rvert$ | $t$ | $\{b \in L^2(\Omega); b = \mathrm{div}\xi, \lvert\xi\rvert_{L^\infty(\Omega)^2} \le 1, \xi\cdot N\rvert_{\partial\Omega} = 0\}$ | $t$ |

It is equivalent to define BV($\Omega$) as the space of $L^1(\Omega)$ functions whose distributional gradient $Du$ is a bounded measure and [9] is its total variation. The space BV($\Omega$) has some interesting properties:

1. lower semicontinuity of the total variation $\int_\Omega |Du|$ with respect to the $L^1(\Omega)$ topology,
2. if $u \in \mathrm{BV}(\Omega)$, we can define, for $\mathcal{H}^1$ almost everywhere $x \in S_u$, the complement of Lebesgue points (i.e., the jump set of $u$), a normal $n_u(x)$ and two approximate "right" and "left" limits $u^+(x)$ and $u^-(x)$, and
3. $Du$ can be decomposed as a sum of a regular measure, a jump measure, and a Cantor measure:

$$Du = \nabla u \, \mathrm{d}x + (u^+ - u^-) n_u \mathcal{H}^1_{/S_u} + C_u$$

where $\nabla u$ is the approximate gradient and $\mathcal{H}^1$ the one-dimensional Hausdorff measure.

This ability to describe functions with discontinuities across a hypersurface $S_u$ makes BV($\Omega$) very convenient to describe images with edges. In this context, the image restoration problem is well posed and suitable numerical tools can be proposed (Chambolle and Lions 1997).

One criticism of the model (b) in **Table 1** pointed out by Meyer (2001) is that if $f$ is a characteristic function and if $f$ is sufficiently small with respect to a suitable norm, then the model (Rudin *et al.* 1992) gives $u = 0$ and $\eta = f$ contrary to what one should expect ($u = f$ and $\eta = 0$). In fact, the main reason of this phenomenon is that the $L^2$-norm for the $\eta$ component is not the right one since very oscillating functions can have large $L^2$-norm (e.g., $f_n(x) = \cos(nx)$). To better describe such oscillating functions, Meyer (2001) introduced the space of functions which can be expressed as a divergence of $L^\infty$-fields. This work was developed in $R^N$ and this framework was adapted to bounded 2D domains by Aubert and Aujol (2005) (see (c) in **Table 1**). An example of image decomposition is shown in **Figure 3**.

In this section, we have shown how the choice of the functional spaces is closely related to the definition of a variational formulation. The functionals are written in a continuous setting and they can usually be minimized by solving the discretized Euler equations iteratively, until convergence. These PDEs and the differential operators are constrained by the energy definition but it is also possible to work directly on the equations, forgetting the formal link with the energy. Such an approach has also been much developed in the computer vision community and it is illustrated in the next section.

We refer the reader to Aubert and Kornprobst (2002) for a general review of variational approaches and PDEs as applied to image analysis.

**Figure 3** Example of image decomposition (see Aubert and Aujol (2005)).

## Scale Spaces and PDEs

Another approach to perform nonlinear filtering is to define a family of image smoothing operators $T_t$, depending on a scale parameter $t$. Given an image $f(x)$, we can define the image $u(t,x) = (T_t f)(x)$ which corresponds to the image $f$ analyzed at scale $t$. In this section, following Alvarez–Guichard–Lions–Morel (Alvarez *et al.* 1993), we show that $u(t,x)$ is the solution of a PDE provided some suitable assumptions on $T_t$.

### Basic Principles of a Scale Space

This section describes some natural assumptions to be fulfilled by scale spaces. We first assume that the output at scale $t$ can be computed from the output at a scale $t - h$ for very small $h$. This is natural, since a coarser scale view of the original picture is likely to be deduced from a finer one. $T_t$ is obtained by composition of transition filters, denoted by $T_{t+h,t}$. So the first axiom is

(A1) $T_{t+h} = T_{t+h,t} T_t \quad T_0 = \mathrm{Id}$

Another assumption is that operators act locally, that is, $(T_{t+h,t} f)(x)$ depends essentially upon the values of $f(y)$ with $y$ in a small neighborhood of $x$. Taking into account the fact that as the scale increases, no new feature should be created by the scale space, we have the local comparison principle: if an image $u$ is locally brighter than another image $v$, then this order must be conserved by the analysis. This is expressed by:

(A2) For all $u$ and $v$ such that $u(y) > v(y)$ in a neighborhood of $x$ and $y \neq x$, then for $h$ small enough, we have

$$(T_{t+h,t} u)(x) \geq (T_{t+h,t} v)(x)$$

The third assumption states that a very smooth image must evolve in a smooth way with the scale

space. Denoting the scalar product of two vectors of $\mathbb{R}^N$ by $<x, y>$, this assumption can be written as

(A3) Let $u(y) = 1/2\langle A(y-x), y-x\rangle + \langle p, y-x\rangle + c$ be a quadratic form of $\mathbb{R}^2$, $x$ fixed ($A = \nabla^2 u(x) \in S^{(2)}$ the set of $2 \times 2$ symmetric matrices, $p = \nabla u(x)$ a vector of $\mathbb{R}^2$, $c = u(x)$ a constant.). We shall say that a scale space is regular if there exists a function $F(t, x, c, p, A)$, continuous with respect to $A$, such that

$$\frac{(T_{t+h,t}u - u)(x)}{h} \to F(t, x, c, p, A) \quad \text{when } h \to 0$$

### Scale Spaces are Governed by PDEs

In the following theorem, it is stated that the former assumptions are sufficient to prove that scale spaces are in fact governed by PDEs.

**Theorem 2** *Under assumptions* A1, A2, A3, *there exists a continuous function* $F : [0, T] \times \Omega \times R \times \mathbb{R}^2 \times S^{(2)} \to \mathbb{R}$ *satisfying* $F(t,x,c,p,A) \geq F(t,x,c,p,B)$ *for all* $p \in \mathbb{R}^2$, *A and B in* $S^{(2)}$ *with* $A \geq B$ *such that*

$$\delta_t(u) = \frac{T_{t+h,t}u - u}{h} \to F(t, x, u, \nabla u, \nabla^2 u), \quad h \to 0^+ \qquad [10]$$

*uniformly for* $x \in \mathbb{R}^2$, *uniformly for* $u$.

In eqn [10], the left-hand side term can be interpreted as the partial temporal derivative with respect to $t$ so that the notion of PDEs arises. More precisely, if $f$ is continuous and uniformly bounded, then it can be established that $u(t, x) = (T_t f)(x)$ is the viscosity solution(see Definition 3) of

$$\begin{aligned} &\frac{\partial u}{\partial t} + H(t, x, u, \nabla u, \nabla^2 u) = 0 \text{ (here } H = -F) \\ &u(0, x) = f(x) \end{aligned} \qquad [11]$$

The map $H : [0, T] \times \Omega \times R \times R^2 \times S^{(2)} \to \mathbb{R}$ is called a Hamiltonian and the decreasing property of $H$ with respect to $S$ is called degenerate ellipticity.

The theory of viscosity solutions was introduced in the 1980s by Crandall and P L Lions (Crandall and Lions 1981, Crandall *et al.* 1992). When strong solutions of [11] do not exist, this theory allows to define solutions which are only continuous or even discontinuous. The definition of viscosity solutions is

**Definition 3** Let $H : \Omega \times \mathbb{R} \times \mathbb{R}^2 \times S^{(2)} \to \mathbb{R}$ be continuous and degenerate elliptic and let $u \in C^0([0, T] \times \Omega)$. Then $u$ is a viscosity solution of [11] in $[0, T] \times \Omega$ if and only if

(i) $u$ is a subsolution, that is, $\forall \phi \in C^2([0, T] \times \Omega)$, $\forall (t_0, x_0)$ a local strict maximum point of $(u - \phi)(t, x)$, we have

$$\frac{\partial \phi}{\partial t}(t_0, x_0) + H(t_0, x_0, u(t_0, x_0), \nabla \phi(t_0, x_0), \nabla^2 \phi(t_0, x_0)) \leq 0$$

(ii) $u$ is a supersolution, that is, $\forall \phi \in C^2([0, T] \times \Omega)$, $\forall (t_0, x_0)$ a local strict minimum point of $(u - \phi)(t, x)$, we have

$$\frac{\partial \phi}{\partial t}(t_0, x_0) + H(t_0, x_0, u(t_0, x_0), \nabla \phi(t_0, x_0), \nabla^2 \phi(t_0, x_0)) \geq 0$$

In this definition, it is noticeable that derivatives of $u$ are replaced by the derivatives of the test functions $\phi$. Obviously, it can be verified that this notion of weak solutions coincides with classical solution when $u$ has enough regularity.

### Diffusion Operators Coming from the Scale Space

A step further is to assume additional properties on the scale spaces and estimate the corresponding operator. Invariance properties include geometric invariance axioms, contrast invariance, or scale invariance. For example, if we assume the axioms A1–A3, gray-level shift invariance:

(I1) $T_t(0) = 0, T_t(u + c) = T_t(u) + c$ for all $u$ and all constant $c$.

and translation invariance:

(I2) $T_t(\tau_h.u) = \tau_h.(T_t u)$ for all $h$ in $\mathbb{R}^2, t \geq 0$, where $(\tau_h.u)(x) = u(x + h)$.

Then it can be established that $F$ in [10] is independent of $(x, u)$, that is, $u(t, x) = (T_t f)(x)$ is the unique viscosity solution of

$$\begin{aligned} &\frac{\partial u}{\partial t} = F(\nabla u, \nabla^2 u) \\ &u(0, x) = f(x) \end{aligned}$$

With more precise assumptions, one can even recover explicitly the operator $F$. As an example, if we look for a linear scale space which verifies some isometry assumption:

(I3) $T_t(R.u)(x) = R.(T_t u)(x)$ for all orthogonal transformation $\mathbb{R}$ on $\mathbb{R}^2$, where $(R.u)(x) = u(Rx)$.

Then it can be proved that the scale space is the unique solution of the heat equation:

$$\frac{\partial u}{\partial t} - \Delta u = 0 \qquad [12]$$
$$u(0, x) = f(x)$$

Figure 4 is an example of [12] applied to a noisy image at different scale, that is, at different time. Note that noise is quickly removed but one has to stop the evolution very early if we would like to preserve some edges. In the nonlinear cases, several operators have also been found based on curvature. For instance, under suitable axioms (Alvarez *et al.* 1993), including contrast, scale, and affine invariance, the associated scale space is

$$\frac{\partial u}{\partial t} - \operatorname{sign}(\kappa)(t\kappa)^{1/3}|\nabla u| = 0$$
$$\text{where } \kappa = \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) \qquad [13]$$
$$u(0, x) = f(x)$$

This equation is called affine morphological scale space (AMSS) and three restored images are shown in Figure 5. Some qualitative differences are shown in Figure 6.

**Figure 4** Illustration of heat equation [12].

**Figure 5** Illustration of the AMSS model [13].

**Figure 6** Some close-ups of Figures 4 and 5 showing qualitative differences after 40 iterations.

**Remark** Scale space theory has shown the formal link between some operators and PDEs. It has to be noticed that one may propose some PDEs which do not directly come from the scale space framework. Starting from [12] which performs isotropic smoothing and smears edges, many nonlinear diffusion models have been proposed to smooth images while preserving edges (see e.g., Perona and Malik (1990)). □

To know more on scale space and PDEs, we refer the reader to Weickert (1998) and Aubert and Kornprobst (2002).

## The Wavelet Approach

Before the 1980s, the Fourier transform played a major role for analyzing oscillating signals. The interest of such a transform for real application increased after the discovery of the fast Fourier transform. However, the Fourier transform has some limit. The Fourier transform extracts from the signal details of the frequency content but loses all information on the location of particular frequency. Moreover, for computing the Fourier transform $\mathcal{F}f(\lambda)$, we need to know $f(t)$ for all the real values of $t$. These difficulties can be overcome by first windowing the signal, and then by taking its Fourier transform:

$$\mathcal{F}^{\text{win}} f(\lambda, t) = \int_{\mathbb{R}} f(s) g(s - t) \mathrm{e}^{-\mathrm{i}\lambda s} \mathrm{d}s$$

where $g$ is a window function. The parameter $\lambda$ plays the role of a frequency localized around the abscissa $t$ of the temporal signal and $\mathcal{F}^{\text{win}} f(\lambda, t)$ give an information about what is happening around $s = t$, for the frequency $\lambda$. The main drawback of this method is that the window has a fixed length which is a serious disadvantage when we want to treat signals having variations of different orders of magnitude. All these issues highlighted that a mathematical theory of time–frequency representation was necessary. This was achieved with the wavelet representation. In this section, we first recall some elements of this theory (for 1D signal) and then we show how it can be applied for restoring noisy images.

### The Wavelet Decomposition

The basic idea is to construct from a function $\psi$, called mother wavelet, an orthonormal basis $\{\psi_{j,k}\}$ of $L^2(\mathbb{R})$ deduced from $\psi$ by translation and dilatation. It is required that $\psi$ be regular, oscillating (but not too much), that $\psi$ and $\mathcal{F}\psi$ are well localized and that $\psi$ has some null moments. Once this function $\psi$ is

chosen, we set $\psi_{j,k}(x) = 2^{j/2}\psi(2^j t - k), j, k \in \mathbb{Z}$. An elegant and practical way for obtaining such a basis is to construct a multiresolution analysis of $L^2(\mathbb{R})$ (Mallat 1989).

**Definition 4** A multiresolution analysis of $L^2(\mathbb{R})$ is a sequence $V_j$, $j \in \mathbb{Z}$ of subspaces of $L^2(\mathbb{R})$, with the following properties:

(i) $\bigcap_j V_j = \{0\}$,
(ii) $V_j \subset V_{j+1}$,
(iii) $\bigcup_j V_j = L^2(\mathbb{R})$,
(iv) $f(t) \in V_j$ if and only if $f(2t) \in V_{j+1}$, and
(v) There exists a regular function $\phi$ with compact support such that the family $\phi(t-k), k \in \mathbb{Z}$, is an orthonormal basis of $V_0$ for the scalar product of $L^2(\mathbb{R})$. Such a function $\phi$ is called a scaling function.

Then it is straightforward to check that the family $\phi_{j,k}(t)$ defined by $\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k)$ is an orthonormal basis of $V_j$.

A basic example of multiresolution analysis of $L^2(\mathbb{R})$ is to choose $V_0$ as the set of piecewise constant functions on $\mathbb{R}$ and take $\phi$ as the characteristic function of the interval $[0, 1)$: $\phi(t) = \chi_{[0,1)}(t)$.

Let us now look at the link between wavelet basis and multiresolution analysis. We just give main ideas, all details can be found in the work of Mallat (1989). Assume that we have a multiresolution analysis, and let us define $W_0$ as the orthogonal complement of $V_0$ in $V_1$. We build the mother wavelet $\psi$ by imposing that the family $\psi(t-k)$, $k \in \mathbb{Z}$, is an orthonormal basis of $W_0$. For example, if $\phi(t) = \chi_{[0,1)}(t)$, it can be shown that $\psi(t) = \chi_{[0,1/2)}(t) - \chi_{[1/2,1)}(t)$ (called the Haar wavelet). By change of scale, one gets that the family $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k), k \in \mathbb{Z}$, is an orthonormal basis of $W_j$, the orthogonal complement of $V_j$ in $V_{j+1}$, that is,

$$V_j \oplus W_j = V_{j+1} \qquad [14]$$

Since the $V_j$'s are a multiresolution analysis, we have $V_J = \oplus_{j=-\infty}^{J-1} W_j$ and $L^2 = \oplus_{j=-\infty}^{j=+\infty} W_j$. It is then clear that $\psi_{j,k}(t)$ is an orthonormal basis of $L^2(\mathbb{R})$, that is, for each function $f \in L^2(\mathbb{R})$, we get the following decomposition:

$$f(t) = \sum_{-\infty}^{+\infty} \sum_k f_{j,k} \psi_{j,k}(t) \quad \text{with } f_{j,k} = \langle f, \psi_{j,k} \rangle_{L^2}$$

Let us see now how in practice a multiresolution analysis can be interpreted. Let $f$ be a function in $L^2(\mathbb{R})$. We denote $A_{2^j} f$ (resp. $D_{2^j} f$) the operator which approximates $f$ (resp. gives the details of $f$) at resolution $2^j$. More precisely, $A_{2^j} f$ (resp. $D_{2^j} f$) is the projection of $f$ on $V_j$ (resp. on $W_j$):

$$A_{2^j} f(t) = \sum_{k=-\infty}^{k=+\infty} \langle f, \phi_{j,k} \rangle \phi_{j,k}(t)$$

$A_{2^j} f$ is characterized by the sequence of scalar products $A_{2^j}^d f = \{\langle f, \phi_{j,k} \rangle\}_{k \in \mathbb{Z}}$. We call $A_{2^j}^d f$ the discrete approximation of $f$ at resolution $2^j$.

In the same way, we have

$$D_{2^j} f(t) = \sum_{k=-\infty}^{k=+\infty} \langle f, \psi_{j,k} \rangle \psi_{j,k}(t)$$

$D_{2^j} f$ is characterized by the sequence of scalar products $D_{2^j}^d f = \{\langle f, \psi_{j,k} \rangle\}_{k \in \mathbb{Z}}$.

We call $D_{2^j}^d f$ the details of $f$ at resolution $2^j$. According to [14], approximation and detail are linked by the relation

$$A_{2^{j+1}} f = A_{2^j} f + D_{2^j} f$$

This means that $D_{2^j} f$ represents the details to be added to obtain from one level of approximation to the next level of approximation.

Finally, the decomposition of a signal $f$ on a wavelet basis is obtained as an accumulation of details at scale $2^j$ from 0 to $+\infty$:

$$f = \sum_{j=-\infty}^{j=+\infty} D_{2^j} f = \sum_{j=-\infty}^{j=+\infty} \sum_{k=-\infty}^{k=+\infty} \langle f, \psi_{j,k} \rangle \psi_{j,k} \qquad [15]$$

Instead of considering the sum over all dyadic levels $j$, one can sum over $j \geq J$ for a fixed $J$; in this case, we have

$$f = \sum_{k=-\infty}^{k=+\infty} \sum_{j \geq J} \langle f, \psi_{j,k} \rangle \psi_{j,k} + \sum_{k=-\infty}^{k=+\infty} \langle f, \phi_{J,k} \rangle \phi_{J,k}$$

We conclude this section by showing how we can construct a 2D wavelet basis from the 1D case. We can simply use a tensor product. Scaling function and mother wavelet are given, respectively, as follows:

$$\phi(x, y) = \phi(x)\phi(y), \qquad \Psi = (\psi^1, \psi^2, \psi^3)$$

with

$$\psi^1(x, y) = \phi(x)\psi(y)$$
$$\psi^2(x, y) = \phi(y)\psi(x)$$
$$\psi^3(x, y) = \psi(x)\psi(y)$$

As for the 1D case, $A_{2^j} f$ denotes the projection of $f$ on $V_j, D_{2^j}^1$ the horizontal details, $D_{2^j}^2$ the vertical

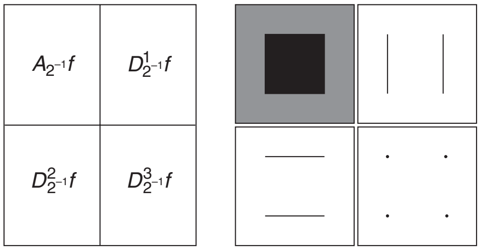



**Figure 7** Illustration on the wavelets methodology.

**Figure 8** Illustration of two regularization methods.

details, and $D^3_{2^j}$ the other details (the indice $l$ in $D^l_{2^j}$ is the same as in $\psi^l$). For a 2D image $f$, we then have the following decomposition (see Figure 7):

$$f = \sum_{\psi^l \in \Psi} \sum_{k=-\infty}^{k=+\infty} \sum_{j \geq J} \langle f, \psi_{j,k} \rangle \psi_{j,k} + \sum_{k=-\infty}^{k=+\infty} \langle f, \phi_{J,k} \rangle \phi_{J,k}$$

### Application to the Denoising Problem

We go back to the denoising problem. Our goal is to solve this problem by using a variational approach and wavelets. We recall that we have an ideal image $u$ that has been corrupted by a white Gaussian noise $\eta$ resulting in an observation $f$ with $f = u + \eta$. As it has been seen in the section "The variational approach," this question can be tackled by solving the variational problem

$$\min_u \{ \lambda \phi(|u|_E) + |f - u|_G \} \qquad [16]$$

for suitable choices of $E$, $G$, and $\phi$. Here we propose to choose $G = L^2(\Omega)$ ($\Omega$ is the domain image) and for $E$ the Besov space $B^1_1(L^1(\Omega))$ and $\phi =$ Identity. Besov spaces $B^\alpha_q(L^p(\Omega))$ are used in many domains of mathematics as harmonic analysis or approximation theory. There exist different ways for defining them. Roughly speaking, they consist of functions having $\alpha$ derivatives in $L^p(\Omega)$; the third parameter $q$ allows one to make finer distinctions in smoothness. Here we are only concerned with the Besov space $B^1_1(L^1(\Omega))$. One important property needed here is that the norm of a function in $E = B^1_1(L^1(\Omega))$ is equivalent to the $l^1$-norm of the wavelet coefficients, that is if $\{\psi_{j,k}\}$ is an orthonormal basis of $L^2(\Omega)$ and if $u_{j,k,\psi}$ are the wavelet coefficients of $u \in E$, then $|u|_E = \sum_j \sum_{k,\psi} |u_{j,k,\psi}|$.

**Remark** When one is concerned with a finite domain, then some changes must be made with respect to the construction given in [15] to obtain an orthonormal basis of $L^2(\Omega)$. To avoid further technical complications, we ignore this question. □

Let us denote, respectively, by $\{u_{j,k,\psi}\}$ and $\{f_{j,k,\psi}\}$ the wavelet coefficients of $u$ and $f$, then solving [16] amounts to finding the minimizer of the functional

$$F(u) = \lambda \sum_{j,k,\psi} |u_{j,k,\psi}| + \sum_{j,k,\psi} |u_{j,k,\psi} - f_{j,k,\psi}|^2 \qquad [17]$$

One notes immediately that minimizing problem [17] reduces to finding the minimizer $s$, given $t$, of $E(s) = |s - t|^2 + \lambda|s|$ and that the minimizer of $E(s)$ is given by $s = t - (\lambda/2)$ if $t > \lambda/2, s = 0$ if $|t| \leq \lambda/2$ and $s = t + (\lambda/2)$ if $t < -(\lambda/2)$.

Thus, we shrink the wavelet coefficients $f_{j,k,\psi}$ toward zero by an amount of $\lambda/2$ to obtain the minimizer. This is exactly the wavelet shrinkage algorithm of Donoho and Johnstone (1994). It is remarkable that the wavelet shrinkage algorithm, which has been found by using statistical tools, can also be explained via a variational approach (Chambolle *et al.* 1998). Figure 8 shows an example of the result on a noisy image.

For more details, we refer the reader to Mallat (1998).

## Conclusion

Image processing is a challenging domain of applied mathematics which has to deal with discrete and continuous representations. In this article, we have covered the core mathematical tools used in the area. The example of gray-scale image restoration allowed us to illustrate and compare the different methodologies. Naturally, as mentioned in the introduction, image processing refers to a wide variety of applications and an intensive research has been carried out on the different topics using the methodologies described here. The reader will find in the references (therein) several illustrations of challenging problems.

*See also:* Γ-Convergence and Homogenization; Convex Analysis and Duality Methods; Elliptic Differential

Equations: Linear Theory; Evolution Equations: Linear and Nonlinear; Fluid Mechanics: Numerical Methods; Fractal Dimensions in Dynamics; Free Interfaces and Free Discontinuities: Variational Problems; Geometric Measure Theory; Ginzburg–Landau Equation; Inequalities in Sobolev Spaces; Minimax Principle in the Calculus of Variations; Optimal Transportation; Partial Differential Equations: Some Examples; Stochastic Differential Equations; Variational Techniques for Ginzburg–Landau Energies; Wavelets: Applications; Wavelets: Mathematical Theory.

## Further Reading

Alvarez L, Guichard F, Lions P, and Morel J (1993) Axioms and fundamental equations of image processing. *Archive for Rational Mechanics and Analysis* 123(3): 199–257.

Ambrosio L, Fusco N, and Pallara D (2000) *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Mathematical Monographs. New York: Clarendon Press.

Aubert G and Aujol J (2005) Modeling very oscillating signals – application to image processing. *Applied Mathematics and Optimization* 51(2): 163–182.

Aubert G and Kornprobst P (2002) *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations,* Applied Mathematical Sciences, vol. 147. New York: Springer.

Besag J (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of Royal Statistical Society* 2: 192–236.

Chambolle A, DeVore R, Lee N, and Lucier B (1998) Non-linear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Transactions on Image Processing* 7(3): 319–334.

Chambolle A and Lions P (1997) Image recovery via total variation minimization and related problems. *Nümerische Mathematik* 76(2): 167–188.

Crandall M, Ishii H, and Lions P-L (1992) User's guide to viscosity solutions of second order partial differential equations. *Bulletin of the American Society* 27: 1–67.

Crandall M and Lions P (1981) Condition d'unicité pour les solutions généralisées des équations de Hamilton–Jacobi du premier ordre. *Comptes Rendus de l'Académie des Sciences* 292: 183–186.

Donoho D and Johnstone I (1994) Ideal spatial adaptation by wavelet shrinkage. *Biometrica* 81: 425–455.

Geman S and Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6(6): 721–741.

Gonzalez RC and Woods RE (1992) *Digital Image Processing*, 3rd edn. Addison-Wesley.

Kirkpatrick S, Gellat C, and Vecchi M (1983) Optimization by simulated annealing. *Science* 220: 671–680.

Li S (1995) *Markov Random Field Modeling in Computer Vision.* Tokyo: Springer.

Mallat S (1989) A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(7): 674–693.

Mallat S (1998) *A Wavelet Tour of Signal Processing*. Cambridge: Academic Press.

Meyer Y (2001) *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations,* University Lecture Series, vol. 22. Providence, RI: American Mathematical Society.

Perona P and Malik J (1990) Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(7): 629–639.

Rudin L, Osher S, and Fatemi E (1992) Nonlinear total variation based noise removal algorithms. *Physica D* 60: 259–268.

Weickert J (1998) *Anisotropic Diffusion in Image Processing.* Stuttgart: Teubner-Verlag.

# Incompressible Euler Equations: Mathematical Theory

**D Chae**, Sungkyunkwan University, Suwon, South Korea



## Introduction

In this article we present comprehensive mathematical results on the incompressible Euler equations. Our presentation is focussed on the two aspects of the equations. The first one is on the theories of classical solutions and the problem of global in time continuation/finite time blow-up of the local classical solutions. The second topic is concerned on the weak solutions, mainly for the two-dimensional (2D) Euler equations for existence and uniqueness questions.

The motion of homogeneous incompressible ideal fluid in a domain $\Omega \subset \mathbb{R}^n$ is described by the following system of Euler equations:

$$\frac{\partial v}{\partial t} + (v \cdot \nabla)v = -\nabla p \qquad [1]$$

$$\operatorname{div} v = 0 \qquad [2]$$

$$v(x,0) = v_0(x) \qquad [3]$$

where $v = (v^1, v^2, \ldots, v^n), v^j = v^j(x,t), j = 1, 2, \ldots, n$, is the velocity of the fluid flows, $p = p(x,t)$ is the scalar pressure, and $v_0(x)$ is a given initial velocity field satisfying $\operatorname{div} v_0 = 0$. Here we use the standard notion of vector calculus, denoting

$$\nabla p = \left(\frac{\partial p}{\partial x_1}, \frac{\partial p}{\partial x_2}, \dots, \frac{\partial p}{\partial x_n}\right)$$

$$(v \cdot \nabla)v^j = \sum_{k=1}^{n} v^k \frac{\partial v^j}{\partial x_k}$$

$$\operatorname{div} v = \sum_{k=1}^{n} \frac{\partial v^k}{\partial x_k}$$

Equation [1] represents the balance of momentum for each portion of fluid, while eqn [2] represents the conservation of mass of fluid during its motion, combined with the homogeneity (constant density) assumption on the fluid. Equations [1] and [2] are first obtained by Euler in 1755. Although we could consider, more generally, the inhomogeneous incompressible Euler equations, in mathematical fluid mechanics considerations the incompressible Euler equations usually mean the above system [1]–[2]. For a bounded domain with fixed boundary $\partial\Omega$, the natural boundary condition is

$$v(x,t) \cdot \nu(x) = 0 \quad \forall (x,t) \in \partial\Omega \times [0,\infty) \qquad [4]$$

where $\nu(x)$ is the unit normal vector at the boundary point $x \in \partial\Omega$. Several studies are concerned with the Cauchy problem of the system [1]–[3], where we consider the case

$$\Omega = \begin{cases} \mathbb{R}^n (\text{whole domain of } \mathbb{R}^n), \text{or} \\ \mathbb{R}^n/\mathbb{Z}^n (\text{periodic domain}) \end{cases} \qquad [5]$$

In this article for simplicity we suppose $\Omega = \mathbb{R}^n, n = 2, 3$ unless otherwise stated. We note that the Euler equation is obtained formally by setting the viscosity $= 0$, or, equivalently, Reynolds number $= \infty$ in the Navier–Stokes equations. Thus, we may view the Euler equations as the one describing approximately the extremely high Reynolds number turbulent flows. For detailed mathematical studies on the finite Reynolds number Navier–Stokes equations, see Temam (1984) and Lions (1996). For much shorter and more comprehensive review see Constantin (1995). In the study of the Euler equations the notion of vorticity, $\omega = \operatorname{curl} v$, plays a very important role. In particular, we can reformulate the system in terms of vorticity fields only as follows. We first suppose we are working in three-dimensional (3D) space, and rewrite [1] as

$$\frac{\partial v}{\partial t} - v \times \operatorname{curl} v = -\nabla\left(p + \frac{1}{2}|v|^2\right) \qquad [6]$$

Taking curl of [6], and using elementary vector identities we obtain the following vorticity formulation:

$$\frac{\partial \omega}{\partial t} + (v \cdot \nabla)\omega = \omega \cdot \nabla v \qquad [7]$$

$$\operatorname{div} v = 0, \qquad \operatorname{curl} v = \omega \qquad [8]$$

$$\omega(x,0) = \omega_0(x) \qquad [9]$$

The linear elliptic system [8] for $v$ can be solved explicitly in terms of $\omega$ to give the Biot–Savart law

$$v(x,t) = \frac{1}{4\pi}\int_{\mathbb{R}^3} \frac{(x-y) \times \omega(y,t)}{|x-y|^3}\,dy \qquad [10]$$

Substituting this $v$ into [7] formally, we obtain a integrodifferential system for $\omega$. The term in the right-hand side of [7] is called the "vortex stretching term," and is regarded as the main source of difficulties in the mathematical theory of the 3D Euler equations. In the 2D case we take the vorticity as the scalar, $\omega = \partial v^2/\partial x_1 - \partial v^1/\partial x_2$, and the evolution equation of $\omega$ becomes

$$\frac{\partial \omega}{\partial t} + (v \cdot \nabla)\omega = 0 \qquad [11]$$

combined with the 2D Biot–Savart law,

$$v(x,t) = \frac{1}{2\pi}\int_{\mathbb{R}^2} \frac{(-y_2 + x_2, y_1 - x_1)}{|x-y|^2}\,\omega(y,t)\,dy \qquad [12]$$

In many studies of the Euler equations it is convenient to introduce the notion of "particle trajectory mapping," $\Phi(\cdot,t)$ defined by

$$\frac{\partial \Phi(\alpha,t)}{\partial t} = v(\Phi(\alpha,t),t)$$
$$\Phi(\alpha,0) = \alpha, \qquad \alpha \in \Omega \qquad [13]$$

The mapping $\Phi(\cdot,t)$ transforms from the location of the initial fluid particles to the location at time $t$, and the parameter $\alpha$ is called the Lagrangian particle marker. If we denote the Jacobian of the transformation, $\det(\nabla_\alpha \Phi(\alpha,t)) = J(\alpha,t)$, then we can show easily that

$$\frac{\partial J}{\partial t} = (\operatorname{div} v)J$$

which implies the fact that the velocity field $v$ satisfies the incompressibility, $\operatorname{div} v = 0$ if and only if the mapping $\Phi(\cdot,t)$ is volume preserving. At this moment, we note that, although the Euler equations are originally derived by applying the mass conservation and the momentum balance principles, we could also derive them by applying the principle of least action to the action defined by

$$\mathcal{A}(\Phi) = \frac{1}{2}\int_{t_1}^{t_2}\int_{\Omega}\left|\frac{\partial \Phi(x,t)}{\partial t}\right|^2 dx\,dt$$

Here, $\Phi(\cdot,t): \Omega \to \Omega$ is a parametrized family of volume-preserving diffeomorphism. This variational approach to the Euler equations implies that we can

view solutions of the Euler equations as a geodesic curve in the $L^2$-metric on the infinite-dimensional manifold of volume-preserving diffeomorphisms (see for more details, e.g., Arnol'd and Khesin (1998)).

The 3D Euler equations have many conserved quantities. We list some important ones below.

1. *Energy*

$$E(t) = \frac{1}{2}\int_\Omega |v(x,t)|^2 \, dx \qquad [14]$$

2. *Helicity*

$$H(t) = \int_\Omega v(x,t) \cdot \omega(x,t) \, dx \qquad [15]$$

3. *Circulation*

$$\Gamma_{C(t)} = \oint_{C(t)} v \cdot dl \qquad [16]$$

where $C(t) = \{\Phi(\alpha,t) | \alpha \in C\}$ is the curve moving along with the fluid.

4. *Impulse*

$$I(t) = \frac{1}{2}\int_\Omega x \times \omega \, dx \qquad [17]$$

5. *Moment of impulse*

$$M(t) = \frac{1}{3}\int_\Omega x \times (x \times \omega) \, dx \qquad [18]$$

The proof of conservations of the above quantities can be carried out without difficulty by using elementary vector calculus (for details see, e.g., Chorin and Marsden (1993), Majda and Bertozzi (2002), Marchioro and Pulvirenti (1994)). The helicity above, in particular, represents the degree of knotedness of the vortex lines in the fluid, where the vortex lines are the integral curves of the vorticity fields. Arnol'd and Khesin (1998) discuss in detail aspects of helicity and other geometric aspects of the Euler equations. For the 2D Euler equations there is no analog of helicity, while the circulation conservation is replaced by the vorticity flux integral,

$$\int_{A(t)} \omega(x,t) \, dx \qquad [19]$$

where $A(t) = \{\Phi(\alpha,t) | \alpha \in A\}$ is a planar region moving along the fluid. The impulse and the moment of impulse integrals are replace by

$$\frac{1}{2}\int_\Omega (x_2, -x_1)\omega \, dx \qquad [20a]$$

and

$$-\frac{1}{3}\int_\Omega |x|^2 \omega \, dx \qquad [20b]$$

respectively.

In the 2D ideal incompressible fluids we have extra conserved quantities; namely for any $p \in [1,\infty]$ the integral

$$\int_\Omega |\omega(x,t)|^p dx \qquad [21]$$

is conserved (as a matter of fact we can extend this statement by replacing the integral by $\int_\Omega f(\omega(x,t))dx$ for any continuous function $f$). There are many known explicit solutions to the Euler equations (See e.g., Lamb (1932) and Majda and Bertozzi (2002)).

## Local Existence and the Blow-Up Problem

### The Classical Results

We first introduce some notations of function spaces. The Lebesgue space $L^p(\Omega), p \in [1,\infty]$, is the Banach space defined by the norm

$$\|f\|_{L^p} := \begin{cases} \left(\int_\Omega |f(x)|^p dx\right)^{1/p}, & p \in [1,\infty) \\ \text{ess.}\sup_{x\in\Omega} |f(x)|, & p = \infty \end{cases}$$

Let us set $\alpha := (\alpha_1, \alpha_2, \ldots, \alpha_n) \in (\mathbb{Z}_+ \cup \{0\})^n$ with $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_n$. Then, $D^\alpha := D_1^{\alpha_1} D_2^{\alpha_2} \cdots D_n^{\alpha_n}$, where $D_j = \partial/\partial x_j, j = 1, 2, \ldots, n$. For given $k \in \mathbb{Z}$ and $p \in [1,\infty)$ the Sobolev space, $W^{k,p}(\Omega)$ is the Banach space of functions consisting of functions $f \in L^p(\Omega)$ such that

$$\|f\|_{W^{k,p}} := \left(\int_\Omega |D^\alpha f(x)|^p \, dx\right)^{1/p} < \infty$$

where the derivatives are in the sense of distributions. For $p = \infty$ we replace the $L^p$-norm by the $L^\infty$ norm. In order to cooperate with the fractional derivatives of order $s \in \mathbb{R}$, we use the space $L^s_p(\Omega)$ defined by the Banach spaces norm,

$$\|f\|_{L^{s,p}} := \|(1-\Delta)^{s/2} f\|_{L^p}$$

where $(1-\Delta)^{s/2} f = \mathcal{F}^{-1}[(1+|\xi|^2)^{s/2}\mathcal{F}(f)(\xi)]$ with $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ denoting the Fourier transform and its inverse. Below we outline the key ideas of proving the local existence theorems for the Euler equations. For more details we refer the reader to Majda and Bertozzi (2002). For simplicity, we use the function space $H^m(\mathbb{R}^n) = W^{m,2}(\mathbb{R}^n), n = 2, 3$. Taking derivatives $D^\alpha$ on [1], and then taking its

$L^2$ inner product with $D^\alpha v$, and summing over the multi-indices $\alpha$ with $|\alpha| \le m$, we obtain

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|v\|_{H^m}^2 &= -\sum_{|\alpha|\le m}(D^\alpha(v\cdot\nabla)v-(v\cdot\nabla)D^\alpha v, D^\alpha v)_{L^2}\\ &\quad-\sum_{|\alpha|\le m}((v\cdot\nabla)D^\alpha v, D^\alpha v)_{L^2}\\ &\quad-\sum_{|\alpha|\le m}(D^\alpha\nabla p, D^\alpha v)_{L^2}\\ &\equiv \mathrm{I}+\mathrm{II}+\mathrm{III}\end{aligned}$$

By integration by parts, we obtain

$$\mathrm{III} = -\sum_{|\alpha|\le m}(D^\alpha p, D^\alpha \operatorname{div} v)_{L^2} = 0$$

Integrating by parts again, and using the fact that div $v = 0$, we have

$$\begin{aligned}\mathrm{II} &= -\frac{1}{2}\sum_{|\alpha|\le m}\int_{\mathbb{R}^3}(v\cdot\nabla)|D^\alpha v|^2\,\mathrm{d}x\\ &= \frac{1}{2}\sum_{|\alpha|\le m}\int_{\mathbb{R}^3}\operatorname{div} v|D^\alpha v|^2\,\mathrm{d}x = 0\end{aligned}$$

We now use the so-called commutator type of estimate,

$$\begin{aligned}&\sum_{|\alpha|\le m}\|D^\alpha(fg)-fD^\alpha g\|_{L^2}\\ &\quad\le C(\|\nabla f\|_{L^\infty}\|g\|_{H^{m-1}}+\|f\|_{H^m}\|g\|_{L^\infty})\end{aligned}$$

and obtain

$$\begin{aligned}I &\le \sum_{|\alpha|\le m}\|D^\alpha(v\cdot\nabla)v-(v\cdot\nabla)D^\alpha v\|_{L^2}\|v\|_{H^m}\\ &\le C\|\nabla v\|_{L^\infty}\|v\|_{H^m}^2\end{aligned}$$

Summarizing the above estimates, I–III, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\|v\|_{H^m}^2 \le C\|\nabla v\|_{L^\infty}\|v\|_{H^m}^2 \qquad [22]$$

Further estimate, using the Sobolev inequality, $\|\nabla v\|_{L^\infty} \le C\|v\|_{H^m}$ for $m > 5/2$, gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\|v\|_{H^m}^2 \le C\|v\|_{H^m}^3$$

Thanks to Gronwall's lemma, we have the local-in-time uniform estimate

$$\|v(t)\|_{H^m} \le \frac{\|v_0\|_{H^m}}{1-Ct\|v_0\|_{H^m}} \le 2\|v_0\|_{H^m}$$

for all $t \in [0, 1/(2C\|v_0\|_{H^m})]$. This is the key *a priori* estimate for the construction of the local solutions. The local-in-time solution of the Euler equations in the Sobolev space $H^m(\mathbb{R}^n)$ for $m > n/2+1, m \in \mathbb{Z}$, was obtained by Kato (1972). For the above-constructed local-in-time solutions, one of the most outstanding open problems in mathematical fluid mechanics is whether the solution can be continued to any future time up to infinity, or the solution will lose regularity and blow up in finite time. Even in terms of numerical experiments, the answer is not yet settled down. In the direction of solving this problem there is a celebrated results, called the Beale–Kato–Majda criterion (1984), which states

$$\limsup_{t\nearrow T_*}\|v(t)\|_{H^s} = \infty \quad \text{if and only if} \quad \int_0^{T_*}\|\omega(s)\|_{L^\infty}ds = \infty \qquad [23]$$

We outline the proof of this result below (for more details see Majda and Bertozzi (2002)). We first recall the Beale–Kato–Majda's version of the logarithmic Sobolev inequality,

$$\|\nabla v\|_{L^\infty} \le C\|\omega\|_{L^\infty}(1+\log(1+\|v\|_{H^m}))+C\|\omega\|_{L^2} \qquad [24]$$

for $m > 5/2$. Now suppose $\int_0^{T_*}\|\omega(t)\|_{L^\infty}\,\mathrm{d}t < \infty$. Taking $L^2$ inner product of [7] with $\omega$, then after integration by part we obtain

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\omega\|_{L^2}^2 &= ((\omega\cdot\nabla)v,\omega)_{L^2}\\ &\le \|\omega\|_{L^\infty}\|\nabla v\|_{L^2}\|\omega\|_{L^2}\\ &= \|\omega\|_{L^\infty}\|\omega\|_{L^2}^2\end{aligned}$$

where we used the identity $\|\nabla v\|_{L^2} = \|\omega\|_{L^2}$. Applying the Gronwall lemma, we obtain

$$\begin{aligned}\|\omega(t)\|_{L^2} &\le \|\omega_0\|_{L^2}\exp\left(\int_0^{T_*}\|\omega(s)\|_{L^\infty}\,\mathrm{d}s\right)\\ &\le C(\omega_0, T_*)\end{aligned} \qquad [25]$$

for all $t \in [0, T_*]$. Substituting [24] into [22], and combining this with [25], we have

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\|v\|_{H^m}^2\\ &\quad\le C[1+\|\omega\|_{L^\infty}[1+\log(1+\|v\|_{H^m})]\|v\|_{H^m}^2\end{aligned}$$

Applying the Gronwall's lemma, we obtain

$$\begin{aligned}\|v(t)\|_{H^m} &\le \|v_0\|_{H^m}\\ &\qquad\times\exp\left[C_1\exp\left(C_2\int_0^{T_*}\|\omega(\tau)\|_{L^\infty}\,\mathrm{d}\tau\right)\right]\\ &\le C(v_0, T_*)\end{aligned}$$

for all $t \in [0, T_*]$ and for some constants $C_1, C_2$. Thus, we proved the "necessity part" of [23], The

"sufficiency part" is an easy consequence of the Sobolev inequality,

$$\int_0^{T_*} \|\omega(s)\|_{L^\infty}\, ds \le T_* \sup_{0\le t\le T_*} \|\nabla v(t)\|_{L^\infty} \le CT_* \sup_{0\le t\le T_*} \|v(t)\|_{H^m}$$

for $m > 5/2$.

### Other Related Results

The previous local existence result in $H^m(\mathbb{R}^n), m > n/2+1$, is basically due to T Kato in 1972. He and G Ponce extended this existence result using the fractional Sobolev space, $L^s_p(\mathbb{R}^n), s > n/2+1, s \in \mathbb{R}$ in 1986. These results were further extended, using the Besov and the Triebel–Lizorkin spaces, by the present author in 2001.

For bounded domain $\Omega \subset \mathbb{R}^n$, R Temam obtained the local-existence result using the space $W^{k,p}(\Omega)$ in 1975. On the other hand, in the setting of the Hölder space, $C^{1,\alpha}(\mathbb{R}^n)$ L Lichtenstein (1925) and W Wolibner (1933) obtained local existence of solutions of the Euler equations. More recently, J-Y Chemin considered the Zygmund $C^s(\mathbb{R}^n)$, which is identical to the Hölder space $C^{[s],s-[s]}(\mathbb{R}^n)$ for noninteger $s$, where $[s]$ means the largest integer not greater than $s$, but is different from $C^{[s],0}(\mathbb{R}^n)$ for integer $s$. He proved, in 1992, local existence of solutions to the 3D Euler equations in this space in 1992. See Chemin (1998) for details of this proof.

The Beale–Kato–Majda criterion for the finite-time blow-up of the classical solutions of the 3D Euler equations has been refined recently by many authors; replacing the $L^\infty$-norm of vorticity $\omega(x,t)$ by the weaker BMO (the space of functions with bounded mean oscillations) norm (H Kozono and Y Taniuchi, 2000), and by the even weaker Besov space or Triebel–Lizorkin space norms by the present author in 2001 (see Triebel (1983) for more details on those spaces). Here we just note that these spaces are refinements of the usual Sobolev spaces. For a bounded domain case, there is a result by A Ferrari in 1993. The blow-up problem is still open even in the case of axisymmetric 3D Euler equations if there is a nonzero swirl (angular velocity). In this case, the blow-up is controlled only by the angular component of the vorticity as shown by the present author (1996). In the region off the axis, in particular, the axisymmetric 3D Euler equation has the same form as the 2D Boussinesq equations.

Some researchers also tried to approach to regularity/singularity problem of the 3D Euler equations by investigating the geometric structure of the vortex stretching term, and obtained a geometric type of blow-up criterion (P Constantin, C Fefferman, and A Majda, 1996). For more detailed review of studies in this direction see Constantin (1995).

Since the blow-up problem of the 3D Euler equation itself looks too difficult to solve, it has also been studied on the simplified model problems. In 1985, P Constantin, PD Lax, and A Majda considered the following 1D model problem of the 3D Euler equations:

$$\theta_t + (H(\theta)\theta)_x = 0, \quad \theta(x,0) = \theta_0(x)$$

where $H(\cdot)$ is the Hilbert transform defined by

$$H(\omega) = \frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{\omega(y)}{x-y}\, dy$$

They proved finite-time blow-up of this model problem by explicitly obtaining the solution. There is another, 2D model problem of the 3D Euler equations, the quasigeostrophic equations,

$$\begin{gathered} \theta_t + (u\cdot\nabla)\theta = 0 \\ u = \nabla^\perp\psi, \qquad \theta = -(-\Delta)^{1/2}\psi \\ \theta(x,0) = \theta_0(x) \end{gathered} \qquad [26]$$

where $\nabla^\perp = (-\partial_2, \partial_1)$. Contrary to the above 1D model equation, this 2D model has real physical relevance in the atmospheric science, and $\theta(x,t)$ represents the temperature of the air. The resemblance of this equation to the 3D Euler equation was first observed by P Constantin, A Majda, and E Tabak in 1994, and they derived the finite blow-up criterion of the equations. In spite of many interesting partial results, including the work by D Cordoba (1998), the blow-up problem of [26] is still open.

## The 2D Euler Equations and the Weak Solutions

### The Case of $W^{1,p}$ Weak Solutions

In 2D Euler equations, the problem of global well-posedness of the classical solutions is settled down. This is an immediate consequence of the conservation of $\|\omega(t)\|_{L^\infty}$ as stated in [21] combined with the Beale–Kato–Majda criterion [23]. On the other hand, the notion of weak solutions is not well understood. A weak solution of the Euler equations is a singular (nondifferentiable) solution of the equations. More precisely, by a weak solution of

[1]–[2] in $\Omega \times (0,T)$ we mean a vector field $v \in C([0,T); L^2_{\mathrm{loc}}(\Omega))$ satisfying the integral identity:

$$-\int_0^T \int_{\mathbb{R}^3} v(x,t) \cdot \frac{\partial \phi(x,t)}{\partial t} \, \mathrm{d}x \, \mathrm{d}t - \int_{\mathbb{R}^3} v(x,0) \cdot \phi(x,0) \, \mathrm{d}x - \int_0^T \int_{\mathbb{R}^3} v(x,t) \otimes v(x,t) : \nabla \phi(x,t) \, \mathrm{d}x \, \mathrm{d}t = 0 \quad [27a]$$

$$\int_0^T \int_{\mathbb{R}^3} v(x,t) \cdot \nabla \psi(x,t) \, \mathrm{d}x \, \mathrm{d}t = 0 \quad [27b]$$

for every vector test function $\phi = (\phi_1, \phi_2, \ldots, \phi_n) \in C_0^\infty(\Omega \times [0,T))$ satisfying div $\phi = 0$, and for every scalar test function $\psi \in C_0^\infty(\Omega \times [0,T))$. Here we used the notation $(u \otimes v)_{ij} = u_i v_j$, and $A : B = \sum_{i,j=1}^n A_{ij} B_{ij}$ for $n \times n$ matrices $A$ and $B$. We observe that [27a] and [27b] are obtained by multiplying $\phi$ and $\psi$ to [1] and [2], respectively, and integrating by parts. Thus, even the locally square-integrable vector fields, which are not differentiable in the classical sense, could be solutions of the Euler equations. For the general 3D Euler equations, we do not yet have the global existence theorems for the weak solutions. Actually, it is even suggested that we need more weaker notion of solution (the so-called "measure-valued solutions") to describe generic global solutions for the 3D Euler equations. For the 2D Euler equations, however, we have global existence theorems for $\omega_0 \in L^1(\mathbb{R}^2) \cap L^p(\mathbb{R}^2)$ for $p \in [1, \infty]$. This better situation of 2D Euler equations compared to the 3D case for the weak solutions is mainly due to the conservation law of $L^p$-norm described in [21]. Here we present briefly the existence proof of the weak solutions for 2D Euler equations in the simplest situation. We will prove the global existence of weak solutions for $\omega_0 \in L^p(\mathbb{R}^2), 1 < p < \infty$. Let $\rho_\varepsilon(x) = (1/\varepsilon^2)\rho(x/\varepsilon)$, where $\rho \in C_0^\infty(\mathbb{R}^2)$ is a standard mollifier, satisfying $\rho \geq 0$, supp $\rho \subset \{x \in \mathbb{R}^2 | |x| < 1\}$, and $\int_{\mathbb{R}^2} \rho \, \mathrm{d}x = 1$. Let $v_0$ be the velocity associated with the initial vorticity $\omega_0$, given by the Biot–Savart law [12]. Define the sequence of initial data $v_0^\varepsilon(x) = \rho_\varepsilon * v_0(x) = \int_{\mathbb{R}^2} \rho_\varepsilon(x-y) v_0(y) \, \mathrm{d}y$. For each $v_0^\varepsilon$ we have global-in-time smooth solutions $v^\varepsilon(x,t)$. Moreover, thanks to [21], we have the following estimate of the vorticity that is uniform in $\varepsilon$:

$$\|\omega(t)^\varepsilon\|_{L^p} = \|\omega_0^\varepsilon\|_{L^p} \leq \|\omega_0\|_{L^p} \quad [28]$$

where we used the property of the mollifier in the second inequality. If we take the (distributional) derivative of the Biot–Savart law [12], we find $\nabla v = K * \omega + C\omega$, where $K(x)$ is a kernel function defining a singular integral operator of the convolution type, and $C$ is a constant vector. The well-known Calderon–Zygmund inequality implies that

$$\|\nabla v\|_{L^p} \leq C_p \|\omega\|_{L^p} \quad [29]$$

Combining [28] and [29] we have

$$\sup_{0 \leq t \leq T} \|\nabla v^\varepsilon(t)\|_{L^p} \leq C(v_0), \quad \forall T > 0 \quad [30]$$

namely the sequence $\{v^\varepsilon\}$ is uniformly bounded in $L^\infty(0,T; W^{1,p}(\mathbb{R}^2))$. Next, we claim that $\{v^\varepsilon\}$ satisfies the inequality

$$\|v^\varepsilon(t_1) - v^\varepsilon(t_2)\|_{H^{-3}(\mathbb{R}^2)} \leq C\|v_0\|_2^2 |t_1 - t_2| \quad [31]$$

for all $t_1, t_2$ with $0 < t_1 \leq t_2 < T$, where $C$ is an absolute constant. Here the negative-order Sobolev space $H^{-m}(\Omega), m > 0$, is defined as the dual of $H_0^m(\Omega)$, and can be identified with the space of functions $C_0^\infty(\Omega)$ completed with metric in $H^m(\Omega)$. Indeed, let $\phi \in C_0^\infty(\mathbb{R}^2)$. Taking $L^2(\mathbb{R}^2)$ inner product of [1] with $\phi$ we have the estimates

$$\begin{aligned}
\left| \int_{\mathbb{R}^2} \frac{\partial v^\varepsilon(x,t)}{\partial t} \cdot \phi(x) \, \mathrm{d}x \right| 
&\leq \left| \int_{\mathbb{R}^2} (\phi \cdot \nabla) p^\varepsilon \, \mathrm{d}x \right| + \left| \int_{\mathbb{R}^2} \phi \cdot (v^\varepsilon \cdot \nabla) v^\varepsilon \, \mathrm{d}x \right| \\
&= \left| \int_{\mathbb{R}^2} p^\varepsilon \nabla \phi \, \mathrm{d}x \right| + \left| \int_{\mathbb{R}^2} (v^\varepsilon \cdot \nabla) \phi v^\varepsilon \, \mathrm{d}x \right| \\
&\leq \|p^\varepsilon(t)\|_{H^{-2}} \|\nabla \phi\|_{H^2} + \|v^\varepsilon(t)\|_{L^2}^2 \|\nabla \phi\|_\infty \\
&\leq C(\|p^\varepsilon(t)\|_{H^{-2}} + \|v_0^\varepsilon\|_{L^2}^2) \|\phi\|_{H^3}
\end{aligned} \quad [32]$$

where we used the Sobolev inequality $\|\nabla \phi\|_{L^\infty} \leq C\|\phi\|_{H^3}$ and the energy equality in the last step. Since [32] holds for all $\phi \in C_0^\infty(\mathbb{R}^2)$, by taking the closure of $C_0^\infty(\mathbb{R}^2)$ in $H^3(\mathbb{R}^2)$ we obtain

$$\left\| \frac{\mathrm{d}v^\varepsilon(t)}{\mathrm{d}t} \right\|_{H^{-2}} \leq C\|p^\varepsilon(t)\|_{H^{-2}} + \|v_0\|_{L^2}^2 \quad [33]$$

We now estimate $\|p^\varepsilon(t)\|_{H^{-2}}$. Taking the divergence operation on [1], we have the Poisson equation

$$\Delta p^\varepsilon = -\mathrm{div}(v^\varepsilon \cdot \nabla v^\varepsilon)$$

Let $\eta \in C_0^\infty(\mathbb{R}^2)$, then

$$\begin{aligned}
\left| \int_{\mathbb{R}^2} \Delta p^\varepsilon(x,t) \eta(x) \, \mathrm{d}x \right| &= \left| \int_{\mathbb{R}^2} \mathrm{div}(v^\varepsilon \cdot \nabla v^\varepsilon) \eta \, \mathrm{d}x \right| \\
&= \left| \int_{\mathbb{R}^2} (v^\varepsilon \cdot \nabla) v^\varepsilon \cdot \nabla \eta \, \mathrm{d}x \right| \\
&= \left| \int_{\mathbb{R}^2} (v^\varepsilon \cdot \nabla) \nabla \eta \cdot v^\varepsilon \, \mathrm{d}x \right| \\
&\leq \|v^\varepsilon(t)\|_{L^2}^2 \|\Delta^2 \eta\|_{L^\infty} \\
&\leq C\|v_0\|_{L^2}^2 \|\eta\|_{H^4}
\end{aligned} \quad [34]$$

where we used the energy equality [14] and the Sobolev inequality in the last step. Since [34] holds for all $\eta \in C_0^\infty(\mathbb{R}^2)$, taking the closure of $C_0^\infty(\mathbb{R}^2)$ in $H^4(\mathbb{R}^2)$, we obtain

$$\left| \int_{\mathbb{R}^2} \Delta p^\varepsilon(x,t)\eta(x)\,dx \right| \le C\|v_0\|_{L^2}^2 \|\eta\|_{H^4} \quad \forall \eta \in H^4(\mathbb{R}^2) \qquad [35]$$

Thus,

$$\|\Delta p^\varepsilon(t)\|_{H^{-4}} \le C\|v_0\|_{L^2}^2 \quad \forall t \in [0,T)$$

This provides us with

$$\|p^\varepsilon(t)\|_{H^{-2}} \le \|D^2 p^\varepsilon(t)\|_{H^{-4}} \le C\|\Delta p^\varepsilon(t)\|_{H^{-4}} \le C\|v_0\|_{L^2}^2$$

Combining [33] with [36], we obtain

$$\sup_{0\le t\le T} \left\| \frac{dv^\varepsilon(t)}{dt} \right\|_{H^{-2}} \le C\|v_0\|_{L^2}^2$$

Thus, from

$$v^\varepsilon(t_1) - v^\varepsilon(t_2) = \int_{t_2}^{t_1} \frac{dv^\varepsilon(t)}{dt}\,dt$$

we have

$$\|v^\varepsilon(t_1) - v^\varepsilon(t_2)\|_{H^{-2}} \le \sup_{0\le t\le T} \left\| \frac{dv^\varepsilon(t)}{dt} \right\|_{H^{-2}} |t_1 - t_2| \le C\|v_0\|_{L^2}^2 |t_1 - t_2|$$

Thus, [31] is proved as claimed. Thanks to the Aubin–Nitsche compactness lemma together with [30] and [31] we have a subsequence, denoted by the same notation, $\{v^\varepsilon\}$ and $v$ in $L^\infty(0,T;W^{1,p}(\mathbb{R}^2))$ such that

$$v^\varepsilon \to v \text{ weakly} - * \text{ in } L^\infty(0,T;W^{1,p}(\mathbb{R}^2)) \qquad [36]$$

and

$$v^\varepsilon \to v \quad \text{in } L^2_{\text{loc}}(\mathbb{R}^2 \times (0,T)) \qquad [37]$$

as $\varepsilon \to 0$. We know that as a classical solution each $v^\varepsilon$ and $v_0^\varepsilon$ satisfies

$$\int_{\mathbb{R}^2} \phi(x,0) v_0^\varepsilon(x)\,dx + \int_0^T \int_{\mathbb{R}^2} (\phi_t \cdot v^\varepsilon + \nabla\phi : v^\varepsilon \otimes v^\varepsilon)\,dx\,dt = 0 \qquad [38]$$

for all $\phi \in C_0^\infty(\mathbb{R}^2 \times [0,T))$ with $\operatorname{div}\phi = 0$ and

$$\int_0^T \int_{\mathbb{R}^2} \nabla\psi \cdot v^\varepsilon\,dx\,dt = 0 \qquad [39]$$

for all $\psi \in C_0^\infty(\mathbb{R}^2 \times [0,T))$. We can check easily that the convergence [36] and [37] is enough to pass to the limit $\varepsilon \to 0$ in [38] and [39] to obtain the corresponding equations with $v^\varepsilon$ and $v_0^\varepsilon$ replaced by $v$ and $v_0$. Thus, $v$ is a weak solution of the Euler equations with initial data $v_0$. This completes the outline of the proof of weak solutions to the 2D Euler equations.

### Notes on Further Results

The study of weak solutions of the 2D Euler equations was initiated by V Yudovich in 1963, where he proved the existence of weak solutions for initial data $\omega_0 \in L^1(\mathbb{R}^2) \cap L^\infty(\mathbb{R}^2)$. Subsequenthy, theory of weak solutions has been developed by studies of the vortex sheet problem due to DiPerna and Majda in 1987. For the existence of weak solutions to the vortex sheet initial data, namely the existence problem for initial vorticity $\omega_0 \in H^{-1}(\mathbb{R}^2) \cap \mathcal{M}(\mathbb{R}^2)$, where $\mathcal{M}(\mathbb{R}^2)$ is the space of Radon measures on $\mathbb{R}^2$, is still an outstanding open problem. The main physical motivation of this problem is to understand the dynamics of vortex sheets in the 3D turbulence. For this problem JM Delort proved existence assuming single-signedness of the initial vortex sheet in 1991. The proof is simplified by A Majda in 1993, using the conservation of moment of impulse. The result is also reproved by LC Evans and S Müller in 1994, using the weak compactness of the Hardy space. Later in 2001, MC Lopes Filho, HJ Nussenzveig Lopes and Z Xin allowed the change of sign for initial vortex sheet, but assumed special reflection symmetry to prove existence of global weak solutions. Related to this problem is the one of characterizing the precise borderline function space to which initial data belongs, and above which there is no concentration phenomenon for weakly approximating sequence of solutions; a recent analysis of this problem was done by E Tadmor in 2001.

For the uniqueness problem of the weak solutions of the 2D Euler equations, there are remarkable works by V Scheffer (1993) and A Shnirelman (1997), where they constructed explicitly an $L^2_{\text{loc}}(\mathbb{R}^2 \times \mathbb{R})$ weak solution starting from zero initial data. Also M Vishik (1999) extended the uniqueness class of the weak solutions of the 2D Euler equations, improving previous work by V Yudovich (1995). The class found by M Vishik, in particular, includes the BMO. There is another problem closely related to the weak solutions of the 2D Euler equations, called the vortex patch problem. The main question was if there is any singularity of the boundary of a patch $\Omega(t) = \{X(\alpha,t) \mid \alpha \in \Omega_0\}$, where $X(\alpha,t)$ is the particle trajectory mapping generated by a weak solution $v(x,t)$, which is evolving from the initial data $\omega_0(x) = \chi_{\Omega_0}(x)$, the characteristic function of set $\Omega_0$

with smooth boundary. The problem itself is well defined, due to the work of V Yudovich (1963), and there exists unique particle trajectories associated with such weak solutions. The problem was settled by J-Y Chemin in 1991. He proved the global-in-time preservation of the $C^{1,\alpha}$ regularity of the boundary $\partial\Omega(t)$, contrary to the previous numerical experiments. The proof of this result was later simplified by A Bertozzi and P Constantin in 1993.

Another interesting problem related to the weak solutions of the Euler equations (2D or 3D) is whether or not the energy is preserved for the weak solutions, namely if there is any "intrinsic dissipation" to the singular solutions of the ideal fluids. In 1949, L Onsager conjectured that if the weak solution of 3D Euler equations belongs to certain Hölder space, then the energy is conserved. This conjecture, in the setting of Besov space, was proved by P Constantin, W E and E S Titi in 1994. This question of possibility of dissipation of energy for weak solutions is further studied by J Duchon and R Robert in 2000. Later, in 2003 the present author considered the problem of helicity conservation for the weak solutions of the 3D Euler flows, which is related to the question of crossing/reconnections of the vortex tubes for weak solutions, and showed that for large class of weak solutions in certain Besov spaces the helicity is preserved.

*See also:* Compressible Flows: Mathematical Theory; Evolution Equations: Linear and Nonlinear; Fluid Mechanics: Numerical Methods; Interfaces and Multicomponent Fluids; Intermittency in Turbulence; Inviscid Flows; Non-Newtonian Fluids; Partial Differential Equations: Some Examples; Stability of Flows; Stochastic Hydrodynamics; Turbulence Theories; Viscous Incompressible Fluids: Mathematical Theory; Vortex Dynamics.

## Further Reading

Arnol'd VI and Khesin BA (1998) *Topological Methods in Hydrodynamics*. New York: Springer.

Beale JT, Kato T, and Majda A (1984) Remarks on the breakdown of smooth solutions for the 3-D Euler equations. *Communications in Mathematical Physics* 94: 61–66.

Chemin J-Y (1998) *Perfect Incompressible Fluids*. New York: Oxford University Press.

Chorin AJ and Marsden JE (1993) *A Mathematical Introduction to Fluid Mechanics*, 3rd edn. New York: Springer.

Constantin P (1994) Geometric statistics in turbulence. *SIAM Review* 36(1): 73–98.

Constantin P (1995) A few results and open problems regarding incompressible fluids. *Notices of the American Mathematical Society* 42(6): 658–663.

Kato T (1972) Nonstationary flows of viscous and ideal fluids in $\mathbb{R}^3$. *Journal of Functional Analysis* 9: 296–305.

Lamb H (1932) *Hydrodynamics*. Cambridge: Cambridge University Press.

Lions PL (1996) *Mathematical Topics in Fluid Mechanics, Volume 1 (Incompressible Models)*. New York: Oxford University Press.

Majda A and Bertozzi A (2002) *Vorticity and Incompressible Flow*. Cambridge: Cambridge University Press.

Marchioro C and Pulvirenti M (1994) *Mathematical Theory of Incompressible Nonviscous Fluids*. New York: Springer.

Temam R (1984) *Navier–Stokes Equations*, 3rd edn. New York: North-Holland.

Triebel H (1983) *Theory of Function Spaces*. Boston: Birkäuser Verlag.

Yudovich VI (1963) Non-stationary flow of an ideal incompressible liquid. *Computational Mathematics and Mathematical Physics* 3: 1407–1456.

# Indefinite Metric

**H Gottschalk**, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany



## Introduction

If, in a problem of quantization, state spaces with indefinite inner product are used instead of Hilbert spaces, one speaks of quantization with indefinite metric. The main domain of application is the quantization of gauge fields, like the electromagnetic vector potential $A_\mu(x)$ or Yang–Mills fields in quantum chromodynamics (QCD) and the standard model.

The conceptual problem with the indefinite metric is the occurrence of senseless negative probabilities in the formalism. Such negative probabilities, however, only arise in expectation values of fields that are not gauge invariant and hence do not correspond to observable quantities. Equivalently, the inner product of vectors generated by application of such fields to the vacuum vector with itself can be negative or null. In order to extract the observable content of an indefinite-metric quantum theory, a subsidiary condition is needed to single out the physical subspace. Restricted to this subspace, the inner product is positive semidefinite. This subsidiary condition can be seen as the implementation of a gauge, as, for example, the Lorentz gauge $\partial_\mu A^\mu(x) = 0$ in quantum electrodynamics (QED). This procedure is also known under the name Gupta–Bleuler formalism.

The use of indefinite metric in the quantization of gauge theories like QED can be avoided entirely.

This is called quantization in a physical gauge. The problem with such gauges is that they are not Lorentz invariant and that the vector potential $A^{\mu}(x)$ is not a local field. An example is the Coulomb gauge defined by $A_0(x) = 0$ and $\partial^i A_i(x) = 0$ in QED. Furthermore, Dirac spinor fields $\psi(x)$ in such gauges do not anticommute when localized in spacelike separated regions. The Dirac fields therefore are also nonlocal quantities. Although not in contrast with special relativity, as Dirac spinors and the vector potential are not gauge invariant and hence are unobservable, this leads to severe technical problems in the formulation of interacting theories. In particular, the theory of renormalization heavily uses both locality and invariance. Therefore, the Gupta–Bleuler formalism generally is the preferred quantization procedure for a gauge theory.

That a local and invariant quantization is not possible using a (positive-metric) Hilbert space has been proved by F Strocchi in a series of articles published between 1967 and 1970. If one wants to preserve locality and/or invariance of the quantized field theory, it is thus strictly necessary to give up the positivity of the state space.

A short digression into the early history of the idea might be of interest. It dates back to 1941, where the use of indefinite metric in the quantization of relativistic equations was proposed by Paul Dirac in a lecture at the London Royal Society. The negative probabilities for the bosonic vector potential were thought to be connected with the problem of negative-energy solutions of relativistic equations as a type of surrogate of the "Dirac sea" in the quantization of fermions. Furthermore, Dirac proposed that negative-energy solutions and negative probabilities would jointly lead to the cancellation of divergences in QED. The latter idea was taken up by W Heisenberg in his lectures on the theory of elementary particles held in Munich in 1961, but the generally accepted solution to the problem of ultraviolet divergences was achieved without recourse to Dirac's original motivation. In 1950 the consistent quantization of vector potential in the Lorentz gauge was formulated by S N Gupta and K Bleuler eliminating the use of negative-energy solutions. Since then the indefinite metric has become a building block of the standard theory of quantized gauge fields.

## No-Go Theorems

The strict necessity of the Gupta–Bleuler procedure for the local or covariant quantization of gauge fields has been demonstrated by F Strocchi in the form of no-go theorems for positive metric. Here we review their content for the case of the electromagnetic field. Related statements can be obtained for nonabelian gauge theories. The main problem lies in the fact that standard assumptions on the quantization of relativistic fields are in conflict with Maxwell equations that should hold as operator identities in a positive-metric theory containing no unobservable states. Let

$$F_{\nu\mu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x) \qquad [1]$$

be the quantized electromagnetic field strength tensor. Classically, the existence of $A_\mu(x)$ is guaranteed from the first set of Maxwell equations $\epsilon^{\alpha\beta\nu\mu}\partial_\beta F_{\nu\mu}(x) = 0$. Here (and henceforth) indices are raised and lowered with respect to the Minkowski metric $g_{\alpha\beta}$ and $\epsilon^{\alpha\beta\mu\nu}$ is the completely antisymmetric tensor on $\mathbb{R}^d$. Furthermore, we apply Einstein's convention on summation over repeated upper and lower indices. Standard assumptions from axiomatic quantum field theory are:

1. The field strength tensor $F_{\nu\mu}(x)$ is an operator-valued distribution acting on a (dense core of a) Hilbert space $\mathcal{H}$ with scalar product $\langle .,. \rangle$ – in the indefinite-metric case, $\langle .,. \rangle$ only needs to be an inner product.
2. $F_{\mu\nu}(x)$ transforms covariantly, that is, there is a strongly continuous unitary (with respect to $\langle .,. \rangle$) representation $U$ of the orthochronous, proper Poincaré group on $\mathcal{H}$ such that for translation $a \in \mathbb{R}^d$ combined with a restricted Lorentz transformation $\Lambda$, one has
$$\begin{aligned} &U(a,\Lambda)F_{\mu\nu}(x)U(a,\Lambda)^{-1} \\ &\quad = (\Lambda^{-1})^{\rho}{}_{\mu}(\Lambda^{-1})^{\kappa}{}_{\nu}F_{\rho\kappa}(\Lambda x + a) \end{aligned} \qquad [2]$$
3. There exists a unique (up to multiplication with $\mathbb{C}$-numbers) translation invariant vector $\Omega \in \mathcal{H}$ (the "vacuum"), that is, $U(a,1)\Omega = \Omega \forall a \in \mathbb{R}^d$.
4. The representation of the translations fulfills the spectral condition
$$\int_{\mathbb{R}^4} \langle \Phi, U(a,1)\Psi \rangle e^{ip\cdot a}\,da = 0 \qquad [3]$$
$\forall \Psi, \Phi \in \mathcal{H}$ if $p$ is not in the closed forward light cone $\bar{V}^+ = \{p \in \mathbb{R}^4 : p\cdot p \geq 0, p^0 \geq 0\}$. Here the dot is the Minkowski inner product.
So far the assumptions concerned only observable quantities. In the following, we also demand.
5. The vector potential $A_\mu(x)$ is realized as an operator-valued distribution on $\mathcal{H}$ and transforms covariantly under translations
$$U(a,1)A_\mu(x)U(a,1)^{-1} = A_\mu(x+a) \qquad [4]$$

The assumptions on the nature of the vector potential so far are rather weak. Strocchi's no-go theorems show that one cannot add further desirable properties as Lorentz covariance and/or locality without getting into conflict with the Maxwell equations:

**Theorem 1** *Suppose that the above assumptions* (1)–(3) *and* (5) *hold. If Maxwell's equations in the absence of charges,*

$$\epsilon^{\alpha\beta\nu\mu}\partial_\beta F_{\nu\mu}(x) = 0, \quad \partial^\mu F_{\mu\nu}(x) = 0 \qquad [5]$$

*are valid as δ operator identities on $\mathcal{H}$ and the gauge potential transforms covariantly*

$$U(a,\Lambda)A_\mu(x)U(a,\Lambda)^{-1} = (\Lambda^{-1})^\nu{}_\mu A_\nu(\Lambda x + a) \qquad [6]$$

*the two-point function of the electromagnetic field tensor vanishes identically*:

$$\langle \Omega, F_{\nu\mu}(x)F_{\kappa\rho}(y)\Omega\rangle = 0 \quad \forall\, x, y \in \mathbb{R}^4 \qquad [7]$$

To gain a better understanding, where the difficulties in the quantization of the Maxwell equations arise from, here is a rough sketch of the proof: Maxwell equations and covariance imply that $f_{\mu\nu\rho}(x-y) = \langle \Omega, A_\mu(x)F_{\nu\rho}(y)\Omega\rangle$ fulfills $\partial^\alpha\partial_\alpha f_{\mu\nu\rho}(x) = 0$ and hence its Fourier transform has support in the union of the forward and backward light cone. The Fourier transform thus can be split into a positive- and a negative-frequency part, and $f_{\mu\nu\rho} = f^+_{\mu\nu\rho} + f^-_{\nu\mu\rho}$ accordingly. By the general analysis of axiomatic field theory (*see* Axiomatic Quantum Field Theory), the functions $f^\pm_{\nu\mu\rho}$ are boundary values of complex analytic functions on certain tubar domains $\mathcal{T}^\pm$ transforming covariantly under a certain representation of the complex Lorentz group. By a theorem of Araki and Hepp giving a general representation of such functions and using the antisymmetry of the field tensor, the following formula can be derived:

$$f^\pm_{\mu\nu\rho}(z) = (g_{\mu\rho}\partial_\nu - g_{\mu\nu}\partial_\rho)f^\pm(z) + \epsilon_{\mu\nu\rho\alpha}\partial^\alpha h^\pm(z)$$
$$z \in \mathcal{T}^\pm \qquad [8]$$

with $f^\pm, h^\pm$ invariant under complex Lorentz transformations. Taking boundary values in $\mathcal{T}^\pm$, one obtains $f_{\mu\nu\rho} = (g_{\mu\rho}\partial_\nu - g_{\mu\nu}\partial_\rho)f + \epsilon_{\mu\nu\rho\alpha}\partial^\alpha h$, with $f = \bar{f}^+ + \bar{f}^-$ and $h = \bar{h}^+ + \bar{h}^-$, where the bar stands for the distributional boundary value. Maxwell's equations imply $\partial^\nu f_{\mu\nu\rho} = (\partial^\nu\partial_\nu g_{\mu\rho} - \partial_\mu\partial_\rho)f = 0$ and $\epsilon_\alpha{}^{\beta\nu\rho}\partial_\beta f_{\mu\nu\rho} = (\partial^\nu\partial_\nu g_{\alpha\mu} - \partial_\alpha\partial_\mu)h = 0$. The only Lorentz-invariant solutions to these equations are constant, which implies the statement of Theorem 1.

The second no-go theorem eliminates the assumption that the vector potential $A_\mu(x)$ is covariant; however, a local gauge is assumed. The result is the same as in Theorem 1:

**Theorem 2** *Suppose that the above assumptions* (1)–(5) *and Maxwell's equations hold as operator identities on $\mathcal{H}$. If, furthermore, the gauge is local, that is,*

$$[A_\mu(x), A_\nu(y)] = 0 \quad \textit{if } x - y \textit{ is spacelike} \qquad [9]$$

*the two-point function of the field strength tensor vanishes again as in Theorem 1.*

Analyzing the interplay of the covariance properties of $F_{\mu\nu}(x)$ with the locality of $A_\mu(x)$, Strocchi was able to show that the function $f_{\mu\nu\rho}(x-y)$ must have the same covariance properties as in Theorem 1, which implies the assertion of Theorem 2.

The first two no-go theorems deal with the free electromagnetic field that is not coupled to charge-carrying fields. This is, of course, already a real obstruction also for an interacting theory, since, by the LSZ formalism, one expects the asymptotic incoming and outgoing fields $A^{\mathrm{in/out}}_\mu(x), F^{\mathrm{in/out}}_{\mu\nu}(x)$ to be free. In fact, it has been proved by D Buchholz that, in the positive-metric case, such asymptotic fields can always be constructed. If one assumes a local and covariant gauge and positivity, the vanishing of the two-point function would also imply that the field $F_{\mu\nu}(x) = 0$ identically by the Reeh–Schlieder theorem.

The next no-go theorem shows that the problems connected to the quantization of the Maxwell equations are not connected only to the free electromagnetic fields. Let us assume that the second set of Maxwell equations is given by

$$\partial^\mu F_{\mu\nu}(x) = j_\nu(x) \qquad [10]$$

where $j_\nu$ is the leptonic current, that is, $j_\nu(x) = e : \psi^\dagger(x)\gamma_\nu\psi(x)$: in the case of QED, where $\psi$ is the quantized Dirac field associated with electrons and positrons. Here $:\cdot:$ stands for Wick ordering and $\gamma_\nu$ are the Dirac matrices, $\psi^\dagger = \psi^*\gamma^0$. The conservation of the current $\partial^\nu j_\nu(x) = 0$ implies that the current charge

$$Q_{\mathrm{C}} = \lim_{R\to\infty} \int_{\mathbb{R}^3}\int_{\mathbb{R}} \alpha(x^0)\chi(\boldsymbol{x}/R)j_0(x^0, \boldsymbol{x})\,\mathrm{d}x^0\,\mathrm{d}\boldsymbol{x} \qquad [11]$$

is a constant of motion, where $\alpha$ and $\chi$ are compactly supported infinitely differentiable functions with $\int_{\mathbb{R}} \alpha(x^0) = 1$ and $\chi(\boldsymbol{x}) = 1$ for $|\boldsymbol{x}| < 1$. Now, an alternative definition of charge, called gauge charge (it generates the global U(1)-gauge transformation), is given by

$$Q_{\mathrm{G}}\Omega = 0, \quad [Q_{\mathrm{C}}, A_\mu(x)] = 0 \quad \text{and}$$
$$[Q_{\mathrm{G}}, \psi(x)] = -e\psi(x) \qquad [12]$$

A third formulation of charge, the Maxwell charge $Q_{\mathrm{M}}$, can also be given by replacing $j^0(x)$ in [11] by $\partial_\nu F^{\nu 0}(x)$. Obviously, if Maxwell equations hold as operator identities, $Q_{\mathrm{C}} = Q_{\mathrm{M}}$. On observable states, all charges $Q_{\mathrm{M}}$, $Q_{\mathrm{C}}$, and $Q_{\mathrm{G}}$ ought to coincide. Strocchi's third theorem shows that this cannot be achieved within a local gauge:

**Theorem 3** *If the Maxwell equations [9] hold and the Dirac field $\psi(x)$ is local with respect to the electromagnetic field tensor $F_{\mu\nu}(x)$, that is,*

$$[F_{\mu\nu}(x), \psi(y)] = 0 \quad \text{if } x - y \text{ is spacelike} \qquad [13]$$

*then $[Q_{\mathrm{M}}, \psi(x)] = 0$, hence $Q_{\mathrm{C}} = Q_{\mathrm{M}} \neq Q_{\mathrm{C}}$.*

The proof is a simple consequence of the observation that $j_0(x) = \partial^\nu F_{\nu 0}(x) = \partial^i F_{i0}(x)$ is a three-divergence as $F_{00}(x) = 0$ by antisymmetry of $F_{\mu\nu}(x)$. Hence,

$$\begin{aligned}[Q_{\mathrm{C}}, \psi(y)] &= \lim_{R\to\infty} \int_{\mathbb{R}^4} [j_0(x), \psi(y)] \alpha(x^0) \chi(\mathbf{x}/R)\, \mathrm{d}x^0\, \mathrm{d}\mathbf{x} \\ &= -\lim_{R\to\infty} \int_{\mathbb{R}^4} [F_{i0}(x), \psi(y)] \alpha(x^0) \partial^i \chi(\mathbf{x}/R) \\ &\quad \times \mathrm{d}x^0\, \mathrm{d}\mathbf{x} = 0 \end{aligned} \qquad [14]$$

since, for $R$ sufficiently large, the support of $\alpha(x^0)\partial_i \chi(\mathbf{x}/R)$ becomes spacelike separated from $y$.

It should be noted that the proof of none of the above theorems relies on the definiteness of the inner product. The main clue of the indefinite-metric formalism, therefore, is rather to give up Maxwell equations as operator identities. In the usual positive-metric formalism, where all states in $\mathcal{H}$ are physical states, this would not be legitimate. But in indefinite metrics, many states are unobservable – in particular, those with negative "norm" $\langle \Psi, \Psi \rangle < 0$. On such states we can neglect the Maxwell equations.

## Axiomatic Framework

The formalism of axiomatic quantum field theory (*see* Axiomatic Quantum Field Theory) requires a revision in order to cover the case of gauge fields. The necessary adaptations have been elaborated by G Morchio and F Strocchi, but also earlier work of E Scheibe and J Yngvasson played a significant role in this development.

Let $\phi(x)$ be a $V'$-valued quantum field, where $V$ is a finite-dimensional $\mathbb{C}$-vector space with involution $*$. The prime stands for the (topological) dual. For the case of QED, $V$ is eight dimensional, containing four dimensions for the vector potential $A_\mu(x)$ and another four for the Dirac spinors $\psi(x), \psi^\dagger(x)$.

Such a quantum field can be reconstructed from its vacuum expectation values (Wightman functions) as follows: let $S_1 = S(\mathbb{R}^4, V)$ be the space of rapidly decreasing functions $f : \mathbb{R}^4 \to V$ endowed with the Schwarz topology. Then the Borcher's algebra $\underline{S}$ be the free, unital, involutive tensor algebra over $\mathcal{S}_1$, that is, $\underline{S} = \mathbb{C}\mathbf{1} \oplus_{n\geq 0} \mathcal{S}_1^{\otimes n}$ with the multiplication induced by the tensor product and involution $(f_1 \otimes \cdots \otimes f_n)^* = f_n^* \otimes \cdots \otimes f_1^*$. $\underline{S}$ is endowed with the direct-sum topology. One can show that any linear, normalized, continuous functional $\underline{W} : \underline{S} \to \mathbb{C}, W(\mathbf{1}) = 1$, is determined by its restrictions $W_n$ to $\mathcal{S}_1^{\otimes n}$. By the Schwarz kernel theorem, $W_n \in \mathcal{S}'(\mathbb{R}^{4n}, V^{\otimes n})$. Conversely, any such sequence of Wightman distributions $W_n$ determines a $\underline{W}$.

Given a Hermitian Wightman functional $\underline{W}$ such that $W(f^*) = \overline{W(f)}, \forall \underline{f} \in \underline{S}, \mathcal{L}_{\underline{W}} = \{\underline{f} \in \underline{S}: \underline{W}(\underline{h}^* \otimes \underline{f}) = 0\, \forall \underline{h} \in \underline{S}\}$ forms a left ideal and the inner product $\underline{W}(\underline{f}^* \otimes \underline{h})$ induces a nondegenerate inner product $\langle .,. \rangle$ on $\mathcal{H}_0 = \underline{S}/\mathcal{L}_{\underline{W}}$. Furthermore, Borchers' algebra $\underline{S}$ acts from the left on $\mathcal{H}_0$. The quantum field $\phi(x)$ defined as the restriction of this canonical representation to the space $\mathcal{S}_1 \subset \underline{S}$ according to $\phi(f) =$ "$\int_{\mathbb{R}^4} \phi^a(x) f_a(x) \mathrm{d}x$" where the index $a$ runs over a basis of $V$.

If the Wightman functional $\underline{W}$ has further properties from axiomatic QFT (*see* Axiomatic Quantum Field Theory) like invariance with respect to a given representation of the Lorentz group on $V$, translation invariance, locality, and the spectral property, the quantum field $\phi(x)$ fulfills the related requirements in analogy with the items (1)–(5) listed in the previous section for the case of the vector potential $A_\mu(x)$. The Wightman distributions $W_n$ as in the positive-metric case are related to the vacuum expectation values of the theory by

$$W_n^{a_1,\ldots,a_n}(x_1, \ldots, x_n) = \langle \Omega, \phi^{a_1}(x_1) \cdots \phi^{a_n}(x_n) \Omega \rangle \qquad [15]$$

where $\Omega$ is the equivalence class of $\mathbf{1}$ in $\mathcal{H}_0$.

The state-space $\mathcal{H}_0$ produced by the Gelfand–Naimark–Segal (GNS) construction for inner-product spaces might be too small to contain all states of physical interest. For example, in the QED case, it does not contain charged states (cf. Theorem 3). Depending on the physical problem, one might also be interested in constructing coherent or scattering states and translation-invariant states apart from the vacuum. Such states appear in problems related to symmetry breaking and confinement (the so-called $\Theta$-vacua) or in some problems of conformal QFT (*see* Boundary Conformal Field

Theory) in two dimensions. It, therefore, has become the standard point of view that one needs to make a suitable closure of $\mathcal{H}_0$ such that this closure includes the states of interest (for an alternative point of view, see the last paragraph of the following section).

Typically, larger closures are favorable, as they contain more states. One therefore focuses on maximal Hilbert closures of $\mathcal{H}_0$. A Hilbert topology $\tau$ is induced by an auxiliary scalar product $(.,.)$ on $\mathcal{H}_0$. It is admissible, if it dominates the indefinite inner product $|\langle\Phi,\Psi\rangle|^2 \leq C(\Psi,\Psi)(\Phi,\Phi)\,\forall\Psi,\Phi\in\mathcal{H}_0$ for some $C>0$. This guarantees that the inner product extends to the Hilbert space closure $\mathcal{H}$ of $\mathcal{H}_0$ with respect to $\tau$. Furthermore, there exists a self-adjoint contraction $\eta$ on $\mathcal{H}$ such that $\langle\Psi,\eta\Phi\rangle = (\Psi,\eta\Phi)\,\forall\Phi,\Psi\in\mathcal{H}$. A Hilbert topology $\tau$ is maximal if there is no admissible Hilbert topology $\tau'$ that is strictly weaker than $\mathcal{H}_0$. The classification of maximal admissible Hilbert topologies in terms of the metric operator $\eta$ is given by the following theorem:

**Theorem 4** *A Hilbert topology $\tau$ on $\mathcal{H}_0$ generated by a scalar product $(.,.)$ is maximal if and only if the metric operator $\eta$ has a continuous inverse $\eta^{-1}$ on the Hilbert space closure $\mathcal{H}$ of $\mathcal{H}_0$. In that case, one can replace $(.,)$ by the scalar product $(\Psi,\Phi)_1 = (\Psi,|\eta|\Phi)$ without changing the topology $\tau$. The new metric operator $\eta_1$ then fulfills $\eta_1^2 = 1_{\mathcal{H}}$.*

For a proof of the first statement, see the original work of Morchio and Strocchi (1980). One can easily check that $\eta_1 = \eta|\eta^{-1}|$ which implies the second assertion of the theorem. A Hilbert space $(\mathcal{H},(.,.))$ with an indefinite inner product induced by a metric operator $\eta$ with $\eta^2 = 1_{\mathcal{H}}$ is called a Krein space. For an extensive study of Krein spaces, see the monograph by Azizov and Iokhvidov (1989).

Furthermore, one can show that given a nonmaximal admissible Hilbert space topology $\tau$ induced by some $(.,.)$, one obtains a maximal admissible Hilbert topology as follows: given the metric operator $\eta$, we define a scalar product $(\Psi,\Phi)_1 = (\Psi,(1-P_0)\Phi)$ on $\mathcal{H}$ with $P_0$ the null space projector of $\eta$. Obviously, this scalar product is still admissible and it leads to a new metric operator $\eta_1$ and a new closure $\mathcal{H}_1$ of $\mathcal{H}_0$. Furthermore, it is easy to show that the scalar product $(\Psi,\Phi)_2 = (\Psi,|\eta_1|\Phi)_1$ still induces an admissible Hilbert topology which is also maximal, as $\eta_2 = \eta_1|\eta_1^{-1}|$ clearly fulfills the Krein relation $\eta_2^2 = 1_{\mathcal{H}_2}$.

The question of the existence of a Krein space closure of $\mathcal{H}_0$, therefore, reduces to the question of the existence of an admissible Hilbert topology on $\mathcal{H}_0$. The following condition on the Wightman functions $W_n$ replaces the positivity axiom in the case of indefinite-metric quantum fields:

**Theorem 5** *Given a Wightman functional $\underline{W}$, there exists an admissible Hilbert space topology $\tau$ on $\mathcal{H}_0 = \underline{\mathcal{S}}/\mathcal{L}_{\underline{W}}$ if and only if there exists a family of Hilbert seminorms $p_n$ on $\mathcal{S}_n$ such that $|W_{n+n}(f\otimes h)| \leq p_n(f)p_m(h), \forall n, m\in\mathbb{N}_0, f\in\mathcal{S}_n, h\in\mathcal{S}_m$.*

In some cases, covering also examples with nontrivial scattering in arbitrary dimension, the condition of Theorem 5 can be checked explicitly (*see* Non-trivial Models of Quantum Fields with Indefinite Metric).

It should be mentioned that different choices of the Hilbert seminorms $p_n$ lead to potentially different maximal Hilbert space closures (Hoffmann 1998, Constantinescu and Gheondea 2001). In fact, often the topology is not even Poincaré invariant and hence the states that can be approximated with local states depend on a chosen inertial frame. This fact, for the case of QED, has been interpreted in terms of physical gauges.

Many results from axiomatic field theory (*see* Axiomatic Quantum Field Theory) with positive metric also hold in the case of QFT with indefinite metric, like the PCT and the Reeh–Schlieder theorem, the irreducibility of the field algebra (for massive theories) and the Bisoniano–Wichmann theorem (*see* Algebraic Approach to Quantum Field Theory). Other classical results, like the Haag–Ruelle scattering theory and the spin and statistics theorem definitively do not hold, as has been proved by counterexamples. This is, however, far from being a disadvantage, as, for example, it permits the introduction of various gauges in the scattering theory of the vector potential $A_\mu(x)$ and fermionic scalar "ghost" fields in the BRST quantization (*see* BRST Quantization) formalism.

## Gupta–Bleuler Gauge Procedure

Here the Gupta–Bleuler gauge procedure is presented in a slightly generalized form following Steinmann's monograph. Classically, the equations of motion for the vector potential $A_\mu(x)$,

$$\partial^\nu\partial_\nu A_\mu(x) + \lambda\partial_\mu\partial^\nu A_\nu(x) = j_\mu(x) \qquad [16]$$

together with Lorentz gauge condition $B(x) = \partial_\mu A^\mu(x) = 0$ imply the Maxwell equations [10]. Here, $\lambda\in\mathbb{R}$ plays the role of a gauge parameter. As seen above, both equations, the so-called pseudo-Maxwell equations [16] and the Lorentz gauge condition $B(x) = 0$, cannot both hold as operator identities. The idea for the quantization

of the theory therefore is to give up the Lorentz gauge condition as an operator identity on the entire state space $\mathcal{H}$.

Suppose one has constructed such a theory with an indefinite inner state space $\mathcal{H}$. Already for the noninteracting theory, any invariant, spectral, local, and covariant solution requires indefinite metric, cf. the explicit formula [18] below. To complete the Gupta–Bleuler program, one needs to find a subspace of (equivalence classes of) physical states $\mathcal{H}'$ of the inner-product space $\mathcal{H}'$ such that the following conditions hold:

1. the vacuum is a physical state, that is, $\Omega \in \mathcal{H}'$;
2. observable fields like $j_\mu(x)$ and $F_{\nu\mu}(x)$ map $\mathcal{H}'$ to itself;
3. the inner product $\langle .,. \rangle$ restricted to $\mathcal{H}'$ is positive semidefinite;
4. observable fields map $\mathcal{H}''$, the set of null vectors in $\mathcal{H}'$, to itself; and
5. the Maxwell equations hold on $\mathcal{H}'$ in the sense

$$\langle \Psi, \partial^\nu F_{\nu\mu}(x)\Phi \rangle = \langle \Psi, j_\mu(x)\Phi \rangle, \quad \forall \Psi, \Phi \in \mathcal{H}' \qquad [17]$$

Then one obtains $\mathcal{H}^{\mathrm{ph}}$ as the completion of the quotient space $\mathcal{H}'/\mathcal{H}''$. The physical Hilbert space $\mathcal{H}^{\mathrm{ph}}$ contains the vacuum $\Omega$ (1), observable fields act on $\mathcal{H}^{\mathrm{ph}}$ (2) and (4), it is a Hilbert space (3) and the Maxwell equations hold on it (5).

To see that such a construction is possible, consider the noninteracting case $j_\nu(x) = 0$, that is, the limit case of vanishing electrical charge $e \to 0$, first. By taking the divergence of [16], one obtains $(1-\lambda)\partial^\nu \partial_\nu \partial^\mu A_\mu(x) = 0$. Excluding the Landau gauge ($\lambda = 1$), this implies $(\partial^\nu \partial_\nu)^2 A_\mu(x) = 0$. The most general solution for the two-point vacuum expectation values that is in agreement with [16] and the requirements of locality, translation invariance, the spectral condition, uniqueness of the vacuum, and the Lorentz covariance of $A^\mu(x)$ is then

$$\begin{aligned}\langle \Omega, A_\nu(x) A_\mu(y) \Omega \rangle &\\ = (-g_{\mu\nu} + \rho \partial_\mu \partial_\nu) D^+(x-y) &\\ + \frac{\lambda}{1-\lambda} \partial_\mu \partial_\nu E^+(x-y) & \qquad [18]\end{aligned}$$

where $D^+$ and $E^+$ are the inverse Fourier transforms of $\theta(p^0)\delta(p^2)$ and $\theta(p^0)\delta'(p^2)$ respectively, $p^2 = p \cdot p$, $\theta$ being the Heavyside function, $\delta$ the Dirac measure on $\mathbb{R}$ of mass one in zero and $\delta'$ its derivative. $\rho$ and $\lambda$ are gauge parameters, for example, the Feynman gauge corresponds to $\lambda = \rho = 0$. We have also omitted an overall factor corresponding to a field strength normalization (choice of numerical value of $\hbar$ – here $\hbar = 1$).

Using Wick's theorem and the GNS construction for inner-product spaces (cf. the preceding section), it is possible to realize a representation of the vector potential $A_\nu(x)$ as operator-valued distribution on some indefinite-metric state space $\mathcal{H}$ with Fock structure, for example, a Krein closure of the GNS space with $\Omega$ the GNS vacuum and $\mathcal{D} \subseteq \mathcal{H}$ the canonical domain of definition. In the case of Feynman gauge, the metric operator $\eta$ can be obtained by a second quantization of the operator $f_\mu \to \sum_{\nu=1}^4 g_{\mu\nu} f_\nu$ on the one-particle space $\mathcal{S}_1$.

In particular, the field $B(x)$ acts as an operator-valued distribution on $\mathcal{H}$ and, by taking the divergence of [16], it follows that $\partial^\nu \partial_\nu B(x) = 0$. Thus, $B(x) = B^+(x) + B^-(x)$ can be decomposed into a positive ("annihilation") and a negative ("creation") frequency part $B^\pm(x)$. One obtains:

**Theorem 6** *The space* $\mathcal{H}' = \{\Psi \in \mathcal{D}: B^+(x)\Psi = 0\}$ *fulfills all requirements* (1)–(5) *of the Gupta–Bleuler gauge procedure.*

Condition (1) is obvious and (2) follows from the fact that the fields $F_{\nu\mu}(x)$ and $B(x)$ commute, which can be checked on the level of two-point functions [18]. In the same spirit, one can also use [18] to check (3) and (4) by explicit calculations on the one-particle space and showing that $\mathcal{H}'$ is the Fock space over the one-particle states annihilated by $B^+(x)$. Finally, by Hermiticity of $A^\mu(x)$, $B^+(x)^* = B^-(x)$ and thus $\langle \Psi, B(x)\Phi \rangle = \langle \Psi, B^+(x)\Phi \rangle + \langle B^+(x)\Psi, \Phi \rangle = 0$. As the field $B(x)$ stands for the obstruction to Maxwell equations, this implies condition (5).

It should be noted that the physical state space $\mathcal{H}^{\mathrm{ph}}$ does not depend on the gauge parameters $\lambda, \rho$ and that it is spanned by repeated application of the field tensor $F_{\mu\nu}(x)$ to the vacuum.

By current conservation, the divergence of [16] still yields $\partial^\nu \partial_\nu B(x) = 0$ also in the interacting case where $e \neq 0$. One can then choose the same gauge condition as in Theorem 6 to define $\mathcal{H}'$. One can then try to prove that this space fulfills all the requirements of the Gupta–Bleuler procedure, for example, in the sense of perturbation theory. Using more advanced formulations as, for example, BRST quantization and Bogoliubov's local $S$-matrix formalism, this program has been completed up to a solution of the infrared problem (*see* Perturbative Renormalization Theory and BRST).

A different procedure, motivated by the necessity of coincidence of all charges $Q_{\mathrm{C}}, Q_{\mathrm{G}}$, and $Q_{\mathrm{M}}$ on the physical state space, has been elaborated by Steinmann. It deviates from the standard procedure in the sense that the physical space $\mathcal{H}'$ is not included in $\mathcal{H}$, but $\mathcal{H}^{\mathrm{ph}}$ is directly obtained from the GNS procedure after taking certain limits of Wightman functions restricted to

certain gauge-invariant algebras constructed from the Borchers algebra and a limiting procedure in a gauge parameter. The Wightman functional on this gauge-invariant algebras are positive (in the sense of perturbation theory), the limiting procedure, however, implies that the so-obtained physical states are singular (i.e., have diverging inner product) to states in $\mathcal{H}$, hence the so-defined state spaces corresponding to going to a physical gauge after solving the problem of a perturbative construction of an indefinite-metric solution, are not subspaces of $\mathcal{H}$.

*See also:* Algebraic Approach to Quantum Field Theory; Axiomatic Approach to Topological Quantum Field Theory; Axiomatic Quantum Field Theory; Boundary Conformal Field Theory; BRST Quantization; Perturbative Renormalization Theory and BRST; Quantum Fields with Indefinite Metric: Non-Trivial Models.

## Further Reading


Azizov TYa and Iokhvidov IS (1989) *Linear Operators in Spaces with an Indefinite Metric*. Chichester: Wiley-Interscience.

Bleuler K (1950) Eine neue Methode zur Behandlung der longitudinalen und skalaren Photonen. *Helvetica Physica Acta* 23: 567.

Constantinescu T and Gheondea A (2001) Representations of Hermitian kernels by means of Krein spaces II: invariant kernels. *Communications in Mathematical Physics* 216: 409–430.

Gottschalk H (2002) Complex velocity transformations and the Bisogniano–Wichmann theorem for quantum fields acting on Krein spaces. *Journal of Mathematical Physics* 43(10): 4753–4769.

Gupta SN (1950) Theory of longitudinal photons in quantum electrodynamics. *Proceedings of Physical Society A* 63: 681.

Hofmann G (1998) On GNS representations on inner product spaces: I. The structure of the representation space. *Communications in Mathematical Physics* 191: 299–323.

Morchio G and Strocchi F (1980) Infrared singularities, vacuum structure and pure phases in local quantum field theory. *Annals of the Institute Henry Poincaré* (*Mathematical Physics*). 33: 251–282.

Morchio G and Strocchi F (1983) A nonperturbative approach to the infrared problem in QED: construction of charged states. *Nuclear Physics B* 211: 471–508.

Morchio G, Pierotti D, and Strocchi F (1990) Infrared vacuum structure in two dimensional local quantum field theory models: the massless scalar field, SISSA Trieste. *Journal of Mathematical Physics* 31: 147.

Steinmann O (2000) *Perturbative Quantum Electrodynamics and Axiomatic Field Theory*. Berlin: Springer.

Strocchi F and Wightman A (1974) Proof of the charge superselection rule in local relativistic quantum field theory. *Journal of Mathematical Physics* 15(12): 2198–2224.

Yngvason J (1973) On the algebra of test functions for field operators. *Communications in Mathematical Physics* 34: 315–333.


# Index Theorems


**P B Gilkey**, University of Oregon, Eugene, OR, USA
**K Kirsten**, Baylor University, Waco, TX, USA
**R Ivanova**, University of Hawaii Hilo, Hilo, HI, USA
**J H Park**, Sungkyunkwan University, Suwon, South Korea




## Introduction

Let $g$ be a Riemannian metric on a smooth compact manifold $M$ of dimension $m$. We assume for the moment that the boundary of $M$ is empty and postpone until later a discussion of the more general setting. If $x=(x_1,\ldots,x_m)$ is a local system of coordinates on $M$, let

$$g_{ij} := g\left(\partial_i^x, \partial_j^x\right)$$

give the components of the metric tensor. Let $D$ be an operator of Laplace type on a smooth vector bundle $V$ over $M$. Adopt the Einstein convention and sum over repeated indices. Relative to a local coordinate frame for $V, D$ has the form

$$D = -\left\{g^{ij}\mathrm{Id}\partial_i^x\partial_j^x + A^k\partial_k^x + B\right\}$$

where $A^k$ and $B$ are endomorphisms (i.e., matrices) of $V$.

We assume that $V$ is equipped with a positive-definite inner product and that $D$ is self-adjoint. There is then a complete orthonormal basis $\{\phi_i\}$ for $L^2(V)$, where $\phi_i \in C^\infty(V)$ and $D\phi_i = \lambda_i\phi_i$. The collection $\{\phi_i, \lambda_i\}$ is called a discrete spectral resolution of $D$. For example, if $D = -\partial_\theta^2$ on the circle, then the discrete spectral resolution is

$$\left\{\mathrm{e}^{\sqrt{-1}n\theta}, n^2\right\}_{n\in\mathbb{Z}}$$

If we order the eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots$ and repeat each eigenvalue according to multiplicity, then there is the following estimate due to Weyl:

$$\lambda_n \sim n^{2/m} \quad \text{as} \quad n \to \infty$$

We now suppose given a pair of vector bundles $V_1$ and $V_2$ over $M$ and a $k$th-order partial differential elliptic operator

$$A : C^\infty(V_1) \to C^\infty(V_2)$$

Locally, we decompose

$$A = \sum_{|I|\le k} a_I \partial_x^I$$

where $I = (i_1, \ldots, i_m)$ is a multi-index and where

$$\partial_x^I = (\partial_1^x)^{i_1} \ldots (\partial_m^x)^{i_m}$$

The $a_I$ are linear maps from $V_1$ to $V_2$. The leading symbol of A is then defined by setting

$$\sigma_L(A)(x, \xi) := (\sqrt{-1})^k \sum_{|I|=k} a_I(x)\xi^I$$

where $\xi^I = (\xi_1)^{i_1} \ldots (\xi_m)^{i_m}$, and

$$\xi = (\xi_1, \ldots, \xi_m)$$

are local fiber coordinates on the cotangent bundle. The leading symbol is an invariantly defined map

$$\sigma_L : T^*M \to \mathrm{End}(V_1,\ V_2)$$

For example, if $V_1 = V_2$ and if $D$ is an operator of Laplace type, then the leading symbol is given by the metric tensor, that is,

$$\sigma_{\mathrm{L}}(D) = g^{ij}\xi_i\xi_j \mathrm{Id} = |\xi|^2 \mathrm{Id}$$

If $d$ is exterior differentiation, then the leading symbol is given by exterior multiplication, that is,

$$\sigma_{\mathrm{L}}(d)(\xi)\omega = \sqrt{-1}\xi \wedge \omega$$

The operator $A$ is said to be elliptic if $\sigma_{\mathrm{L}}(A)$ is an isomorphism from $V_1$ to $V_2$ for any $\xi \neq 0$. If $A$ is an elliptic partial differential operator, then

$$\begin{aligned}\mathrm{index}(A) &:= \dim \ker(A) - \dim \mathrm{coker}(A) \\ &= \dim \ker(A^*A) - \dim \ker(AA^*)\end{aligned}$$

is well defined. As the index vanishes if $m$ is odd, we assume for the most part that $m$ is even.

If $A_\varepsilon$ is a smooth one-parameter family of such operators, then index $(A_\varepsilon)$ is independent of $\varepsilon$. The index depends only on the homotopy class of the leading symbol of $A$ within the class of invertible symbols; it does not depend on the underlying metric of the manifold and it does not depend on the fiber metrics chosen for $V_1$ and $V_2$.

The Atiyah–Singer index theorem expresses the index as the integral of suitably chosen polynomials in the curvature tensor for the classical elliptic complexes and, more generally, in terms of cohomological information for general elliptic complexes. Further details appear later in the article.

The primary focus here is on the complexes which are of Dirac type, that is, complexes where $A$ is a first-order partial differential operator and where the associated second operators $D_1 := A^*A$ and $D_2 := AA^*$ are of Laplace type.

Here is a brief outline of this article. The classical elliptic complexes (de Rham, signature, spin, Dolbeault, Yang–Mills) are discussed first. Next the characteristic classes are introduced, followed by the relevant formula for the index of the classical elliptic complexes, manifolds with boundary, and the equivariant index. Index theory is an enormous topic and here only classical features are emphasized as a complete treatment is beyond the scope of a short expository note such as this one. As some guide to various applications in mathematical physics, the reader is referred to the Further Reading section.

## The Classical Elliptic Complexes

### The de Rham Complex

Let $\Lambda^p M$ be the bundle of smooth $p$ forms over $M$ and let

$$d : C^\infty(\Lambda^p M) \to C^\infty(\Lambda^{p+1} M)$$

and

$$\delta : C^\infty(\Lambda^p M) \to C^\infty(\Lambda^{p-1} M)$$

be the exterior derivative and dually the interior derivative, respectively. We set

$$\Delta := (d + \delta)^2 \quad \text{on} \quad C^\infty(\Lambda M)$$

and the decompose $\Delta = \oplus_p \Delta^p$, where $\Delta^p$ is an operator of Laplace type on $C^\infty(\Lambda^p M)$.

We have $d^2 = 0$. The de Rham cohomology groups are given by taking the quotient of the closed forms by the exact forms:

$$H^p(M; \mathbb{R}) := \frac{\ker(d : C^\infty(\Lambda^p M) \to C^\infty(\Lambda^{p+1} M))}{\mathrm{im}(d : C^\infty(\Lambda^{p-1} M) \to C^\infty(\Lambda^p M))}$$

The Hodge–de Rham theorem identifies $H^p(M; \mathbb{R})$ with the kernel of the Laplacian

$$\ker(\Delta^p) = H^p(M; \mathbb{R})$$

and with the topological cohomology groups.

If $\xi$ is a cotangent vector, let $\mathrm{e}(\xi) : \omega \to \xi \wedge \omega$ be exterior multiplication. Let $\mathrm{i}(\xi)$ be the dual operator, interior multiplication. If $\{e_i\}$ is a local

ortho-normal frame for $TM$, let $e^I = e^{i_1} \wedge \cdots \wedge e^{i_p}$, where $I = \{1 \le i_1 < \cdots < i_p \le m\}$. Then we have

$$\mathfrak{e}(e^1)e^I = \begin{cases} 0 & \text{if } i_1 = 1 \\ e^1 \wedge e^I & \text{if } i_1 > 1 \end{cases}$$
$$\mathfrak{i}(e^1)e^I = \begin{cases} e^{i_2} \wedge \cdots \wedge e^{i_p} & \text{if } i_1 = 1 \\ 0 & \text{if } i_1 > 1 \end{cases}$$

Define a Clifford module structure on $\Lambda M$ by

$$\gamma(\xi) := \mathfrak{e}(\xi) - \mathfrak{i}(\xi)$$

If $\{e_i\}$ is a local orthonormal basis for $TM$, then

$$\gamma(e^i)\gamma(e^j) + \gamma(e^j)\gamma(e^i) = -2\delta_{ij}\mathrm{Id}$$

so the usual Clifford commutation rules are satisfied. Let $\nabla$ be the Levi-Civita connection on $M$. We may then expand

$$d = \mathfrak{e}(e^i)\nabla_{e_i}, \qquad \delta = -\mathfrak{i}(e^i)\nabla_{e_i}$$
$$d + \delta = \gamma(e^i)\nabla_{e_i}$$

The de Rham complex is then defined by taking

$$\Lambda^{\mathrm{even}}M := \oplus_k \Lambda^{2k}M, \qquad \Lambda^{\mathrm{odd}}M := \oplus_k \Lambda^{2k+1}M$$
$$d + \delta : C^\infty(\Lambda^{\mathrm{even}}M) \to C^\infty(\Lambda^{\mathrm{odd}}M)$$

### The Signature Complex

The signature complex arises from a different decomposition of the exterior algebra. Let Clif $M$ be the Clifford algebra of $T^*M$; this is the universal unital algebra generated by $T^*M$ subject to the Clifford commutation relations given above:

$$\xi_1 * \xi_2 + \xi_2 * \xi_1 = -2g(\xi_1, \xi_2) \cdot \mathrm{Id}$$

We suppose $M$ is orientable and let

$$\mathrm{orn} = e_1 * \cdots * e_m \in \mathrm{Clif}\, M$$

be the orientation class. The map $\xi \to \gamma(\xi)$ extends to a unital algebra homomorphism

$$\gamma : \mathrm{Clif}\, M \to \mathrm{End}(\Lambda M)$$

$\gamma(\mathrm{orn})$ defines an endomorphism of $\Lambda M$ which is, modulo suitable sign conventions, the Hodge $\star$ operator. If $m = 2k$ is even, then

$$(d + \delta)\gamma(\mathrm{orn}) = -\gamma(\mathrm{orn})(d + \delta)$$

Set

$$\Theta := (\sqrt{-1})^k \gamma(\mathrm{orn})$$

As $\Theta^2 = \mathrm{Id}$, we can decompose

$$\Lambda M \otimes \mathbb{C} = \Lambda^+ M \oplus \Lambda^- M$$

where $\Lambda^\pm M$ are the $\pm 1$ eigenspaces of $\Theta$. The signature complex is then given by

$$(d + \delta) : C^\infty(\Lambda^+ M) \to C^\infty(\Lambda^- M)$$

### Twisted Signature Complex

Let $V$ be an auxiliary complex vector bundle over $M$ which is equipped with a unitary connection $\nabla^V$. We use the connection $\nabla^V$ on $V$ and the Levi-Civita connection on $TM$ to covariantly differentiate tensors of all types. The twisted signature complex is defined by setting

$$(d+\delta)_V := (\gamma(e^i) \otimes \mathrm{Id})\nabla_{e_i} : C^\infty(\Lambda^+ M \otimes V) \to C^\infty(\Lambda^- M \otimes V)$$

### Yang–Mills complex

This complex in dimension 4 arises from yet another decomposition of the exterior algebra. We use the discussion in the previous section to decompose

$$\Lambda^2 M = \Lambda^{2,+}M \oplus \Lambda^{2,-}M$$

into the $\pm 1$ eigenspaces of $\Theta$. Let

$$\pi : \Lambda^2 M \to \Lambda^{2,-}M$$

be orthogonal projection. The Yang–Mills complex is the 3-term sequence

$$d : C^\infty(\Lambda^0 M) \to C^\infty(\Lambda^1 M)$$

and

$$\pi d : C^\infty(\Lambda^1 M) \to C^\infty(\Lambda^{2,-}M)$$

We can wrap up this sequence to obtain an equivalent elliptic complex

$$(d + \delta) : C^\infty(\Lambda^{\mathrm{even},-}M) \to C^\infty(\Lambda^{\mathrm{odd},+}M)$$

As with the signature complex, this complex can be twisted by taking coefficients in an auxiliary vector bundle $V$. It is crucial to the study of four-dimensional geometry using Yang–Mills theory.

### Dolbeault Complex

Let $z = (z_1, \ldots, z_k)$ be a local system of holomorphic coordinates on a complex manifold $M$, where $z_i = x_i + \sqrt{-1}y_i$. We define

$$\mathrm{d}z^i := \mathrm{d}x^i + \sqrt{-1}\mathrm{d}y^i, \quad \mathrm{d}\bar{z}^i := \mathrm{d}x^i - \sqrt{-1}\mathrm{d}y^i$$
$$\partial_i^z = \tfrac{1}{2}\left(\partial_i^x - \sqrt{-1}\partial_i^y\right), \ \bar{\partial}_i^z = \tfrac{1}{2}\left(\partial_i^x + \sqrt{-1}\partial_i^y\right)$$

and decompose $d = \partial + \bar{\partial}$, where

$$\partial := \mathfrak{e}(\mathrm{d}z^i)\partial_i^z \quad \text{and} \quad \bar{\partial} := \mathfrak{e}(\mathrm{d}\bar{z}^i)\partial_i^{\bar{z}}$$

on the complexified exterior algebra. Let $\delta'$ be the adjoint of $\partial$ and $\delta''$ be the adjoint of $\bar{\partial}$. Let

$$\mathrm{d}\bar{z}^I := \mathrm{d}\bar{z}^{i_1} \wedge \cdots \wedge \mathrm{d}\bar{z}^{i_p}$$

$$\Lambda^{(0,\text{even})} := \text{Span}\{\mathrm{d}\bar{z}^I\}_{|I| \text{ is even}}$$

$$\Lambda^{(0,\text{odd})} := \text{Span}\{\mathrm{d}\bar{z}^I\}_{|I| \text{ is odd}}$$

The Dolbeault complex is then defined by

$$(\bar{\partial} + \delta'') : C^\infty(\Lambda^{(0,\text{even})}M) \to C^\infty(\Lambda^{(0,\text{odd})}M)$$

This complex can be twisted by taking coefficients in a holomorphic bundle $V$ over $M$.

### The Spin Complex

Let $M$ be orientable. Let $P_{\text{SO}}$ be the principal SO bundle of orthonormal frames for the tangent bundle. A spin structure $s$ on $M$ is a principal Spin bundle $P_{\text{Sp}}$ together with a double cover $\rho : P_{\text{Sp}} \to P_{\text{SO}}$ which respects the usual double cover $\rho : \text{Spin} \to \text{SO}$ of the structure groups. Equivalently, a spin structure is a lifting of the transition functions from SO to Spin which preserves the cocycle condition. One says that $M$ is spin if it admits a spin structure.

A manifold is orientable if and only if the first Stiefel–Whitney class of $M$ vanishes; an orientable manifold is spin if and only if the second Stiefel–Whitney class of $M$ vanishes as well; these are $\mathbb{Z}_2$-valued cohomology classes. Inequivalent spin structures are parametrized by the cohomology group $H^1(M; \mathbb{Z}_2)$ or, equivalently, by real-line bundles on $M$.

The spin representation $\mathcal{S}$ of Spin defines an associated spin bundle $\mathcal{S}M = \mathcal{S}(M, s)$. There is a natural Clifford action $c$ of $TM$ on $\mathcal{S}M$. The Levi-Civita connection lifts to define the spin connection on $\mathcal{S}$ and the Dirac operator is defined by

$$A(s) := c(\mathrm{d}x^i)\nabla_{\partial_i^x} \quad \text{on} \quad C^\infty(\mathcal{S}M)$$

Let $m = 2k$ and let $\Theta := (\sqrt{-1})^k c(\text{orn})$. Since $c(\Theta)^2 = \text{Id}$, one can decompose

$$\mathcal{S}M = \mathcal{S}^+M \oplus \mathcal{S}^-M$$

as the direct sum of the half-spin bundles to obtain the spin complex:

$$A(s) : C^\infty(\mathcal{S}^+M) \to C^\infty(\mathcal{S}^-M)$$

As with the signature complex, the spin complex can be twisted by taking coefficients in an auxiliary vector bundle $V$.

### Relating the Classic Elliptic Complexes

One has natural isomorphisms of virtual representations of the spinor group:

$$\Lambda^+ - \Lambda^- = (\mathcal{S}^+ - \mathcal{S}^-) \otimes (\mathcal{S}^+ + \mathcal{S}^-)$$

$$\Lambda^{\text{even}} - \Lambda^{\text{odd}} = (-1)^{m/2}(\mathcal{S}^+ - \mathcal{S}^-) \otimes (\mathcal{S}^+ - \mathcal{S}^-)$$

which show that the signature complex and de Rham complexes are the spin complexes with coefficients in the virtual bundles

$$\mathcal{S}^+M + \mathcal{S}^-M \quad \text{and} \quad (-1)^{m/2}(\mathcal{S}^+M - \mathcal{S}^-M)$$

respectively. If $M$ is complex and spin, then the Dolbeault complex is the spin complex with coefficients in the square root of the canonical bundle.

One can consider complex spinors to define the group $\text{Spin}^c(m)$. Any spin manifold admits a $\text{Spin}^c$ structure with trivial associated complex line bundle. Any complex manifold admits a $\text{Spin}^c$ structure with associated complex line bundle given by the canonical bundle. Thus, a complex manifold admits a $\text{Spin}^c$ structure if and only if it is possible to take a square root of the canonical line bundle; inequivalent Spin structures are parametrized by inequivalent square roots. If $M$ is orientable, then $M$ admits a $\text{Spin}^c$ structure if and only if the second Stiefel–Whitney class of $M$ lifts from $H^2(M; \mathbb{Z}_2)$ to $H^2(M; \mathbb{Z})$; in the complex setting, this lifting is performed by the first Chern class. Inequivalent $\text{Spin}^c$ structures are parametrized by $H^2(M; \mathbb{Z})$ or, equivalently, by complex line bundles over $M$.

## Characteristic Classes

### The Euler Form

Let $\nabla$ be the Levi-Civita connection on $M$. Let

$$R(x, y) := \nabla_x\nabla_y - \nabla_y\nabla_x - \nabla_{[x,y]}$$

be the curvature operator. Let $\{e_1, \ldots, e_m\}$ be a local orthonormal frame for $TM$ and let

$$R_{ijkl} := g(R(e_i, e_j)e_k, e_l)$$

give the components of the curvature relative to a local orthonormal frame. Let

$$\varepsilon^{I,J} := g(e^{i_1} \wedge \cdots \wedge e^{i_m}, e^{j_1} \wedge \cdots \wedge e^{j_m})$$

be the totally antisymmetric tensor; this is the sign of the permutation which sends $i_\nu \to j_\nu$. Let $m = 2\bar{m}$. The Euler form is given by setting

$$\mathcal{E}_m := \frac{1}{8^{\bar{m}}\pi^{\bar{m}}\bar{m}!}\varepsilon^{I,J}R_{i_1i_2j_1j_2}\ldots R_{i_{m-1}i_mj_{m-1}j_m}$$

Let $\rho_{ij} := R_{ikkj}$ and $\tau := \rho_{ii}$ be the Ricci tensor and the scalar curvature, respectively. Then,

$$\mathcal{E}_2 = \frac{1}{4\pi}\tau \quad \text{and} \quad \mathcal{E}_4 = \frac{1}{32\pi^2}\{\tau^2 - 4|\rho|^2 + |R|^2\}$$

### The Pontrjagin Forms

Since $R(x,y) = -R(y,x)$, we can regard $R$ as a 2-form-valued endomorphism of the tangent bundle. We define the Pontrjagin forms $p_i \in C^\infty(\Lambda^{4i}M)$ by expanding

$$\det\left(I + \frac{1}{2\pi}R\right) = 1 + p_1 + p_2 + \cdots$$

These differential forms are closed and the corresponding cohomology classes

$$P_i = [p_i] \in H^{4i}(M;\mathbb{R})$$

in the de Rham cohomology are independent of the particular Riemannian metric on $M$ which was chosen.

The $\hat{A}$ genus and the Hirzebruch $L$ polynomial are expressed in terms of these classes using the splitting principle. Let $A$ be a skew-symmetric matrix. One sets

$$p(A) := \det(I + A) = 1 + p_1(A) + p_2(A) + \cdots$$

As $A$ is skew symmetric, it decomposes as the direct sum of $2 \times 2$ blocks of the form

$$\begin{pmatrix} 0 & \lambda_i \\ -\lambda_i & 0 \end{pmatrix}$$

We then have

$$p(A) = \prod_\nu \{1 + \lambda_\nu^2\}$$

so

$$p_i(A) = s_i(\lambda_1^2, \lambda_2^2, \ldots)$$

where $s_i$ is the $i$th symmetric function;

$$p_1 = \sum_i \lambda_i^2, \qquad p_2 = \sum_{i<j} \lambda_i^2 \lambda_j^2$$

and so forth. Let

$$\begin{aligned} L(\vec{\lambda}) &:= \prod_\nu \frac{\lambda_\nu}{\tanh(\lambda_\nu)} \\ &= 1 + L_1(\vec{\lambda}) + L_2(\vec{\lambda}) + \cdots \\ \hat{A}(\vec{\lambda}) &:= \prod_\nu \frac{\lambda_\nu}{2\sinh\left(\frac{1}{2}\lambda_\nu\right)} \\ &= 1 + \hat{A}_1(\vec{\lambda}) + \hat{A}_2(\vec{\lambda}) + \cdots \end{aligned}$$

As $L_i$ and $\hat{A}_i$ are even symmetric functions of $\vec{\lambda}$, one can write $L_i = L_i(p_1(A), \ldots, p_k(A))$. For example,

$$\begin{aligned} L &= 1 + \tfrac{1}{3}p_1 + \tfrac{1}{45}\left(7p_2 - p_1^2\right) + \cdots \\ \hat{A} &= 1 - \tfrac{1}{24}p_1 + \tfrac{1}{5760}(7p_1^2 - 4p_2) + \cdots \end{aligned}$$

Substituting $(1/2\pi)R$ for $A$ then permits one to define the Hirzebruch polynomial $L(R)$ and the $\hat{A}$ genus $\hat{A}(R)$.

### The Chern Forms

Let $V$ be a $k$-dimensional complex vector bundle over $M$. Let $\nabla$ be a Hermitian connection on $V$ and let $\Omega$ be the associated curvature endomorphism. The Chern forms $c_i \in C^\infty(\Lambda^{2i}M)$ are defined by expanding

$$\det\left(I + \frac{\sqrt{-1}}{2\pi}\Omega\right) = 1 + c_1 + c_2 + \cdots$$

As with the Hirzebruch polynomial and the $\hat{A}$ genus, the Chern character and Todd genus are expressed in terms of the generating functions:

$$\mathrm{Td}(\vec{\lambda}) = \prod_\nu \frac{\lambda_\nu}{1 - \mathrm{e}^{-\lambda_\nu}}$$

and

$$\mathrm{ch}(\vec{\lambda}) = \sum_\nu \frac{\lambda_\nu}{\nu!}$$

One has

$$\begin{aligned} \mathrm{Td} &= 1 + \mathrm{Td}_1 + \mathrm{Td}_2 + \cdots \\ &= 1 + \tfrac{1}{2}c_1 + \tfrac{1}{12}\left(c_1^2 + c_2\right) + \cdots \\ \mathrm{Ch} &= \mathrm{ch}_0 + \mathrm{ch}_1 + \mathrm{ch}_2 + \cdots \\ &= k + c_1 + \tfrac{1}{2}\left(c_1^2 - 2c_2\right) + \cdots \end{aligned}$$

## The Index Theorem

### The Gauss–Bonnet Theorem

We return to the de Rham complex. Let

$$\chi(M) = \sum_p (-1)^p \dim H^p(M;\mathbb{R})$$

be the Euler–Poincaré characteristic; $\chi(M) = 0$ if $m$ is odd. Let $M$ have a simplicial structure with $n(k)$ cells of degree $k$; $n(0)$ is the number of vertices, $n(1)$ is the number of edges, $n(2)$ is the number of triangles, etc. Then

$$\chi(M) = \sum_k (-1)^k n(k)$$

so the Euler–Poincaré characteristic is a combinatorial invariant. By the Hodge–de Rham theorem,

$$\begin{aligned}\text{index}(d+\delta) &= \dim\ker(\Delta^{\text{even}}) - \dim\ker(\Delta^{\text{odd}})\\ &= \chi(M)\end{aligned}$$

The Chern–Gauss–Bonnet theorem expresses this invariant in terms of curvature

$$\chi(M) = \int_M \mathcal{E}_m \, \mathrm{d}x$$

where $\mathcal{E}_m$ is the Euler form given above. If one twists the de Rham complex to take coefficients in an auxiliary vector bundle $V$, then no new information results, since

$$\text{index}\{d+\delta\}_V = \chi(M)\cdot\dim(V)$$

### The Hirzebruch Signature Theorem

Let sign $(M)$ be the index of the signature complex on a manifold of dimension $4k$; the index vanishes in dimensions $m \equiv 2 \bmod 4$. Let $\star$ be the Hodge duality operator. As $\star\Delta^p\star^{-1} = \Delta^{m-p}, \star$ preserves the eigenspaces of the Laplacian. In particular, $\star$ induces an isomorphism

$$\begin{aligned}\star : H^p(M;\mathbb{R}) &= \ker(\Delta^p)\\ \to\ H^{m-p}(M;\mathbb{R}) &= \ker(\Delta^{m-p})\end{aligned}$$

which implements Poincaré duality. In dimension $2k$, $\star^2 = \text{Id}$. Decompose

$$H^{2k}(M;\mathbb{R}) = H^{2k,+}(M;\mathbb{R}) \oplus H^{2k,-}(M;\mathbb{R})$$

into the $\pm 1$ eigenspaces of $\star$; these may be identified with $\ker(\Delta^{2k,\pm})$ acting on $C^\infty(\Lambda^{2k,\pm}M)$. As the contributions to the signature away from the middle dimension cancel,

$$\text{sign}(M) = \dim H^{2k,+}(M;\mathbb{R}) - \dim H^{2k,-}(M;\mathbb{R})$$

As with the de Rham complex, there is a topological description of this invariant. If $\alpha$ and $\beta$ are closed $2k$ forms, one sets

$$\langle\alpha,\,\beta\rangle := \int_M \alpha\wedge\beta$$

One can use Stoke's theorem to see that this induces a symmetric bilinear form on the de Rham cohomology groups $H^{2k}(M;\mathbb{R})$. Poincaré duality then shows that this symmetric bilinear form is nondegenerate, so this is a form of type $(p,q)$; sign$(M)$ is the signature of this quadratic form:

$$\text{sign}(M) = q - p$$

The Hirzebruch signature formula expresses sign $(M)$ in terms of curvature; if $L$ is the Hirzebruch polynomial described above and if $m = 4k$, then

$$\text{sign}(M) = \int_M L_k$$

Let $V$ be an auxiliary coefficient bundle. Taking coefficients in $V$ then yields the formula

$$\text{sign}_V(M) = \sum_{4i+2j=m} 2^j \int_M L_i \wedge \text{ch}_j(V)$$

### The Index of the Yang–Mills Complex

Let $\text{YM}_V$ be the Yang–Mills complex with coefficients in an auxiliary vector bundle $V$, then the index can be evaluated using the formulas given above as

$$\begin{aligned}\text{index}\{\text{YM}_V\} &= \tfrac{1}{2}\{\dim(V)\chi(M) - \text{sign}(M,\,V)\}\\ &= \tfrac{1}{2}\textstyle\int_M\{\dim V\mathcal{E}_4 - \dim V L_1 - 4\,\text{ch}_2(V)\}\end{aligned}$$

### The Index of the Dolbeault Complex

If $V$ is a holomorphic bundle over a complex manifold $M$, then

$$\text{index}\{(\bar\partial + \delta'')_V\} = \sum_{2i+2j=m}\int_M \text{Td}_i(M)\wedge\text{ch}_j(V)$$

The index of the untwisted Dolbeault complex is called the arithmetic genus and denoted by ag$(M)$.

### The Index of the Spin Complex

If $M$ is a spin manifold and if $A_V$ is the Dirac operator with coefficients in an auxiliary coefficient bundle, then

$$\text{index}\{A_V\} = \sum_{4i+2j=m}\int_M \hat{A}_i(M)\wedge\text{ch}_j(M)$$

The index of the spin complex is called the $\hat{A}$ genus and is denoted by $\hat{A}(M)$. If $M$ is a Spin$^c$ manifold, the appropriate formula becomes

$$\text{index}\{A_V^c\} = \sum_{4i+2j+2k=m}\int_M \hat{A}_i(M)\wedge\text{ch}_j(M)\wedge\theta^k$$

where $\theta = \frac{1}{2}c_1(L)$, $L$ being the complex line bundle associated with the Spin$^c$ structure.

#### Properties

The classic elliptic complexes defined above are multiplicative with respect to Cartesian product. Suppose that $M_1$ and $M_2$ are Riemannian manifolds with the appropriate structures. For the signature complex, suppose $M_1$ and $M_2$ are oriented; for the Dolbeault complex, suppose $M_1$ and $M_2$ are holomorphic; for the spin complex, suppose $M_1$ and $M_2$ are spin. By taking the twisting coefficient bundle to be trivial in the interests of simplicity, one has

$$\begin{aligned}\chi(M_1\times M_2) &= \chi(M_1)\chi(M_2)\\ \mathrm{sign}(M_1\times M_2) &= \mathrm{sign}(M_1)\mathrm{sign}(M_2)\\ \mathrm{ag}(M_1\times M_2) &= \mathrm{ag}(M_1)\mathrm{ag}(M_2)\\ \hat{A}(M_1\times M_2) &= \hat{A}(M_1)\hat{A}(M_2)\end{aligned}$$

These complexes behave well under finite coverings. Let $F\to M_2\to M_1$ be a finite covering projection with $|F|$ sheets. Then

$$\begin{aligned}\chi(M_2) &= |F|\chi(M_1)\\ \mathrm{sign}(M_2) &= |F|\mathrm{sign}(M_1)\\ \mathrm{ag}(M_2) &= |F|\mathrm{ag}(M_1)\\ \hat{A}(M_2) &= |F|\hat{A}(M_1)\end{aligned}$$

The connected sum $M_1\#M_2$ is defined by punching out small disks about points $P_i$ in $M_i$ and then joining along the spherical boundaries that remain. It is necessary, of course, to smooth out the resulting corners. Note that if $M_1$ and $M_2$ are complex manifolds, then $M_1\#M_2$ is no longer a complex manifold in general. Since

$$\chi(S^m)=2,\quad \mathrm{sign}(S^m)=0,\ \text{and}\ \hat{A}(S^m)=0$$

the following additivity results follow from the integral formulas given above:

$$\begin{aligned}\chi(M_1\#M_2) &= \chi(M_1)+\chi(M_2)-2\\ \mathrm{sign}(M_1\#M_2) &= \mathrm{sign}(M_1)+\mathrm{sign}(M_2)\\ \hat{A}(M_1\#M_2) &= \hat{A}(M_1)+\hat{A}(M_2)\end{aligned}$$

#### Examples and Applications

Let $S^m$ be the standard sphere and let $\mathbb{CP}^j$ be the complex projective plane. One then has

$$\begin{aligned}&\chi(S^4)=2, &&\mathrm{sign}(S^4)=0\\ &\chi(S^2\times S^2)=4, &&\mathrm{sign}(S^2\times S^2)=0\\ &\chi(\mathbb{CP}^2)=3, &&\mathrm{sign}(\mathbb{CP}^2)=1\end{aligned}$$

In dimension 4, the Riemann–Roch formula yields

$$\mathrm{ag}(M^4)=\tfrac{1}{4}\{\chi(M)+\mathrm{sign}(M)\}$$

This would yield $\mathrm{ag}(S^4)=\frac{1}{2}$; since $\frac{1}{2}$ is not an integer, this shows that $S^4$ does not admit a complex structure; a similar argument shows that $S^n$ does not admit a complex structure for $n\neq 2,6$, and it is not known whether $S^6$ admits a holomorphic structure; it does admit an almost-complex structure.

If we set $M=\mathbb{CP}^2\#\mathbb{CP}^2$, then

$$\mathrm{ag}(M)=\tfrac{1}{4}(3+3-2+1+1)=\tfrac{3}{2}$$

and thus $\mathbb{CP}^2\#\mathbb{CP}^2$ does not admit a complex structure. These examples are typical of the use of the index theorem to prove the nonexistence of certain structures.

#### The General Index Theorem

Let $S(T^*M)$ be the sphere bundle of unit cotangent vectors and let $D(T^*M)$ be the disk bundle of cotangent vectors of length at most 1. Let

$$P : C^\infty(V_1)\to C^\infty(V_2)$$

be an elliptic pseudodifferential operator. The leading symbol $p:=\sigma_L(P)$ induces a smooth map

$$p : S(T^*M)\to \mathrm{End}(V_1,\ V_2).$$

We form $\Sigma(M)$ by gluing two copies of $D(M)$ together along their common boundary $S(M)$ and we define a bundle $\Sigma(p,V_1,V_2)$ over $\Sigma(M)$ by gluing $V_1$ to $V_2$ over $S(M)$ using the clutching function $p$. The Atiyah–Singer index theorem expresses the index of $P$ in terms of cohomological data involving the Chern class of the symbol bundle and the characteristic classes of the tangent bundle of $M$. If $\Sigma(M)$ is given a suitable orientation, then

$$\mathrm{index}(P)=\sum_{2i+4j=2m}\int_{\Sigma(M)}\mathrm{ch}_i(\Sigma(p,\ V_1,\ V_2))\wedge \mathrm{Td}_j(M)$$

It specializes to the results given above for the classical elliptic complexes. Conversely, by using $K$-theoretic methods, the index theorem in full generality can be derived from the special case of the twisted signature complex.

## Manifolds with Boundary

If the boundary of $M$ is nonempty, we must impose suitable boundary conditions.

#### Local Boundary Conditions

Choose local coordinates $x=(x^1,\ldots,x^m)$ near the boundary of $M$ so that $x^m$ is the geodesic distance to the boundary. On the boundary, we can decompose a differential form $\omega\in C^\infty(\Lambda M)$ in the form $\omega=\omega_1+\mathrm{d}x^m\wedge\omega_2$, where $\omega_1$ and $\omega_2$ are tangential

differential forms. Absolute and relative boundary conditions are defined by setting

$$\mathcal{B}_a\omega := \omega_2|_{\partial M} \quad \text{and} \quad \mathcal{B}_r\omega := \omega_1|_{\partial M}$$

Let $(d+\delta)_a$ and $(d+\delta)_r$ be the associated realizations. These operators preserve the grading of the exterior algebra $\Lambda M = \Lambda^{\text{even}}M \oplus \Lambda^{\text{odd}}M$ and define elliptic complexes

$$(d+\delta)_a : C^\infty(\Lambda^{\text{even}}M) \to C^\infty(\Lambda^{\text{odd}}M)$$
$$(d+\delta)_r : C^\infty(\Lambda^{\text{even}}M) \to C^\infty(\Lambda^{\text{odd}}M)$$

We consider a collection

$$J = \{1 \le j_1 < \cdots < j_p < m\}$$

of tangential indices and let

$$dx^J = dx^{j_1} \wedge \cdots \wedge dx^{j_p}$$

The associated absolute boundary conditions for the Laplacian are defined by

$$\begin{aligned}&\tilde{\mathcal{B}_a}(\phi_J dx^J + \psi_J dx^m \wedge dx^J)\\ &\quad = (\psi_J|_{\partial M} dx^J) \oplus (\partial_m^x \phi_J|_{\partial M}) dx^J\end{aligned}$$

If $\star$ is the Hodge operator, then one sets dually:

$$\tilde{\mathcal{B}_r}(\omega) = \tilde{\mathcal{B}_a}(\star\omega)$$

Let $\Delta_a^p$ and $\Delta_r^p$ be the associated realizations of the Laplacian with these boundary conditions. The Hodge–de Rham theorem extends to this setting to yield isomorphisms

$$\ker(\Delta_a^p) = H^p(M; \mathbb{R})$$

and

$$\ker(\Delta_r^p) = H^p(M, \partial M; \mathbb{R})$$

The Hodge $\star$ operator intertwines $\Delta_a^p$ and $\Delta_r^{m-p}$ and implements the Poincaré duality isomorphism $H^p(M;\mathbb{R}) = H^{m-p}(M, \partial M; \mathbb{R})$. This also shows that

$$\text{index}(d+\delta)_a = \sum_p (-1)^p \dim H^p(M;\mathbb{R}) = \chi(M)$$

and

$$\begin{aligned}\text{index}(d+\delta)_r &= \sum_p (-1)^p \dim H^p(M, \partial M; \mathbb{R})\\ &= \chi(M, \partial M) = \chi(M) - \chi(\partial M)\end{aligned}$$

Let $\mathcal{E}_m$ be the Euler form if $m$ is even. We set $\mathcal{E}_m = 0$ if $m$ is odd. Let $L$ be the second fundamental form. Let $A = (a_1, \ldots, a_{m-1})$ and $B = (b_1, \ldots, b_{m-1})$ be collections of distinct indices ranging from 1 to $m-1$. Set

$$\begin{aligned}\mathcal{L}_{m-1} := \sum_k &\frac{1}{\pi^k 8^k k!(m-1-2k)!\text{vol}(S^{m-1-2k})}\\ &\times \varepsilon^{A,B} R_{a_1a_2b_2b_1} \ldots R_{a_{2k-1}a_{2k}b_{2k}b_{2k-1}}\\ &\times L_{a_{2k+1}b_{2k+1}} \ldots L_{a_{m-1}b_{m-1}}\end{aligned}$$

The Chern–Gauss–Bonnet theorem generalizes to this setting to yield

$$\begin{aligned}\chi(M) &= \text{index}(d+\delta)_a\\ &= \int_M \mathcal{E}_m \text{dx} + \int_{\partial M} \mathcal{L}_{m-1} \text{dy}\end{aligned}$$

For example,

$$\begin{aligned}\chi(M^2) &= \frac{1}{4\pi}\left\{\int_{M^2} \tau \text{dx} + 2\int_{\partial M^2} L_{aa}\text{dy}\right\}\\ \chi(M^3) &= \frac{1}{8\pi}\int_{\partial M^3}\{R_{abba} + L_{aa}L_{bb} - L_{ab}L_{ab}\}\text{dy}\\ \chi(M^4) &= \frac{1}{32\pi^2}\int_{M^4}\{\tau^2 - 4|\rho|^2 + |R|^2\}\text{dx}\\ &\quad + \frac{1}{24\pi^2}\int_{\partial M^4}\{3\tau L_{aa} + 6R_{amam}L_{bb}\\ &\quad + 6R_{acbc}L_{ab} + 2L_{aa}L_{bb}L_{cc}\\ &\quad - 6L_{ab}L_{ab}L_{cc} + 4L_{ab}L_{bc}L_{ac}\}\text{dy}\end{aligned}$$

The interior integral vanishes if $m$ is odd. The boundary integral can be nonzero in any dimensions. Thus, in particular, the index of this elliptic complex can be nonzero even if $m$ is odd; $\chi(D^m) = 1$ for any $m$. The index of $(d+\delta)_r$ is computed similarly.

### Spectral Boundary Conditions

In contrast to the de Rham complex, there do not exist local boundary conditions for the signature, spin, and Dolbeault complexes. To simplify the discussion, we assume that the metric is the product near the boundary; there are appropriate compensating terms involving the second fundamental form in the more general setting. Let $A : C^\infty(V_1) \to C^\infty(V_2)$ denote either the twisted signature or the twisted spin complexes; there are some additional difficulties for the Dolbeault complex. Near the boundary, we can express

$$A = \sigma(\partial_m^x + A_{\text{T}})$$

where $A_{\text{T}}$ is a self-adjoint tangential operator of Dirac type on $V_1|_{\partial M}$ and $\sigma$ is a unitary bundle

isomorphism from $V_1|_{\partial M}$ to $V_2|_{\partial M}$. Let $\{\phi_i, \lambda_i\}$ be the discrete spectral resolution of $A_{\mathrm{T}}$. One defines

$$\eta(A_{\mathrm{T}}, s) = \sum_{\lambda_k \neq 0} \mathrm{sgn}(\lambda_k)|\lambda_k|^{-s}$$

as a measure of the spectral asymmetry of $A_{\mathrm{T}}$. This is well defined for $\mathrm{Re}(s) \gg 1$ and has a meromorphic extension to the complex plane $\mathbb{C}$. It turns out that 0 is a regular value and one defines

$$\eta(A_T) := \tfrac{1}{2}\{\eta(A_{\mathrm{T}},\, s) + \dim\, \ker(A_{\mathrm{T}})\}|_{s=0}$$

The spectral boundary conditions can now be imposed. Let $\Pi_{\geq}$ be orthogonal projection in $L^2(V_1|_{\partial M})$ on the span of the eigensections of $A_{\mathrm{T}}$ corresponding to non-negative eigenvalues and let $A_{\geq}$ be the associated realization defined by this boundary condition.

One can use the Atiyah–Patodi–Singer index theorem to generalize the relations given above to this setting. Let $f_A$ be the local integral given above that involves the Hirzebruch $L$ polynomial for the signature complex or the $\hat{A}$ genus for the spin complex. One then has

$$\mathrm{index}(A_{\geq}) = \eta(A_{\mathrm{T}}) + \int_M f_A$$

There are suitable correction formulas involving integrals of polynomials in the second fundamental form and in the curvature tensor if the structures are not product near the boundary.

## Equivariant Problems

### The Classical Lefschetz Formula

Let $M$ be a compact Riemannian manifold without boundary. Let $T$ be a smooth map from $M$ to $M$. Then pullback $T^*$ induces an action on $C^\infty(\Lambda^p M)$ which commutes with the exterior derivative $d$ and hence an action on the de Rham cohomology groups $H^p(M;\mathbb{R})$. The Lefschetz number of $T$ is then given by

$$\mathcal{L}(T) = \sum_p (-1)^p \mathrm{tr}\{T^* \text{ on } H^p(M;\mathbb{R})\}$$

To illustrate the Lefschetz number, let $M = \mathbb{T}^2$ be the two-dimensional torus. Let $e^1 := \mathrm{d}x^1$, let $e^2 := \mathrm{d}x^2$, and let $e^{12} := \mathrm{d}x^1 \wedge \mathrm{d}x^2$. Then,

$$\begin{aligned} H^0(\mathbb{T}^2;\mathbb{R}) &= 1 \cdot \mathbb{R} \\ H^1(\mathbb{T}^2;\mathbb{R}) &= e^1 \cdot \mathbb{R} + e^2 \cdot \mathbb{R} \\ H^2(\mathbb{T}^2;\mathbb{R}) &= e^{12} \cdot \mathbb{R} \end{aligned}$$

Let $T(x_1, x_2) = (n_{11}x_1 + n_{12}x_2,\ n_{21}x_1 + n_{22}x_2)$. Then,

$$\begin{aligned} T^*(1) &= 1 \\ T^*(e^1) &= n_{11}e^1 + n_{12}e^2 \\ T^*(e^2) &= n_{21}e^1 + n_{22}e^2 \\ T^*(e^{12}) &= (n_{11}n_{22} - n_{12}n_{21})e^{12} \end{aligned}$$

and, consequently, the Lefschetz number becomes

$$\begin{aligned} \mathcal{L}(T) &= \det(I - T^*) \\ &= 1 - (n_{11} + n_{22}) + (n_{11}n_{22} - n_{12}n_{21}) \end{aligned}$$

The classical Lefschetz fixed-point formula expresses $\mathcal{L}$ in terms of data for the fixed-point set $\mathcal{F}(T)$ and is an example of the equivariant index theorem. One assumes that the fixed-point set of $T$ consists of smooth submanifolds $N_1, \ldots, N_k$ and that the induced map $\mathrm{d}T_\nu$ on the normal bundles of these manifolds is nondegenerate. This means that $\det(I - \mathrm{d}T_\nu) \neq 0$, that is, that there are no infinitesimal normal directions which are left fixed. One then has

$$\mathcal{L}(T) = \sum_i \mathrm{sign}(\det(I - \mathrm{d}T_\nu))\chi(N_i)$$

### The Lefschetz Formula for the Other Classical Elliptic Complexes

Let $T$ be an orientation-preserving isometry of $M$. When dealing with the spin complex, suppose that $T$ preserves the spin structure; when dealing with the Dolbeault complex, suppose that $T$ preserves the holomorphic structure. If

$$A\,:\,C^\infty(V_1) \to C^\infty(V_2)$$

is one of the classical elliptic complexes, then by assumption $T^*$ commutes with $A$ and hence preserves the eigenspaces of the associated Laplacians. The Lefschetz number is defined by setting

$$\begin{aligned} \mathcal{L}_A(T) := &\,\mathrm{tr}(T^* \text{ on } \ker(A^*A)) \\ &- \mathrm{tr}(T^* \text{ on } \ker(AA^*)) \end{aligned}$$

Setting $T = \mathrm{Id}$, one recovers the standard index.

To simplify the discussion, we assume henceforth that $T$ is an orientation-preserving isometry of $M$ with only isolated fixed points. Let $\{\theta_1, \ldots, \theta_{m/2}\}$ be the rotation angles of $\mathrm{d}T$ at a fixed point $x$ of $T$. Set

$$\lambda_j := \cos(\theta_j) + \sqrt{-1}\sin(\theta_j)$$

We take the sum over the isolated fixed points $x$ and then the product over the rotation angles $1 \leq j \leq m/2$ to express

$$\mathcal{L}_{\text{sign}}(T) = \sum_x \prod_j \left\{ -\sqrt{-1}\cot\left(\frac{\theta_j}{2}\right) \right\}$$
$$\mathcal{L}_{\text{spin}}(T) = \sum_x \prod_j \left\{ -\frac{1}{2}\sqrt{-1}\csc\left(\frac{\theta_j}{2}\right) \right\}$$
$$\mathcal{L}_{\text{Dolb}}(T) = \sum_x \prod_j (1-\bar{\lambda}_j)^{-1}$$

In considering the spin complex, we assume $T$ preserves the spin structure. This permits us to lift $\mathrm{d}T$ from $\mathrm{SO}(m)$ to $\mathrm{Spin}(m)$ and defines liftings of the rotation angles $\theta_i$ from $[0, 2\pi]$ to $[0, 4\pi]$ in such a way that the formula given above for the spin complex is well defined. In considering the Dolbeault complex, we assume that $T$ preserves a complex structure, so the formula given above for the Dolbeault complex involving the complex eigenvalues $\lambda_j$ is well defined.

### Acknowledgements

Research of P Gilkey was partially supported by the MPI (Leipzig, Germany). Research of R Ivanova was partially supported by the UHH Seed Money Grant. Research of K Kirsten was partially supported by the Baylor University Summer Sabbatical Program and by the MPI (Leipzig, Germany). Research of J H Park was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2005-204-C00007).

*See also:* Anomalies; Clifford Algebras and their Representations; Cohomology Theories; Dirac Operator and Dirac Field; Gerbes in Quantum Field Theory; Intersection Theory; Instantons: Topological Aspects; K-Theory; Path-Integrals in Non Commutative Geometry; Quillen Determinant; Riemann Surfaces; Spinors and Spin Coefficients.

### Further Reading

Atiyah MF and Segal GB (1968) The index of elliptic operators II. *Annals of Mathematics* 87: 531–545.

Atiyah MF and Singer IM (1968) The index of elliptic operators I, III, IV, V. *Annals of Mathematics* 87: 484–530, 546–604.

Atiyah MF and Singer IM (1971) The index of elliptic operators I, III, IV, V. *Annals of Mathematics* 93: 119–138, 139–149.

Atiyah MF, Patodi VK, and Singer IM (1975) Spectral asymmetry and Riemannian geometry I. *Mathematical Proceedings of the Cambridge Philosophical Society* 77: 43–69; 78: 405–432.

Atiyah MF, Patodi VK, and Singer IM (1976) Spectral asymmetry and Riemannian geometry I. *Mathematical Proceedings of the Cambridge Philosophical Society* 79: 71–79.

Berline N, Getzler E, and Vergne M (1992) *Heat Kernels and Dirac Operators*, Grundlehren der Mathematishen Wissenschaften, vol. 298. Berlin: Springer.

Bordag M, Mohideen U, and Mostepanenko VM (2001) New developments in the Casimir effect. *Physics Reports* 353: 1–205.

Eguchi T, Gilkey PB, and Hanson AJ (1980) Gravitation, gauge theories and differential geometry. *Physics Reports* 66: 213–393.

Elizalde E, Odintsov SD, Romeo A, Bytsenko AA, and Zerbini S (1994) *Zeta Regularization Techniques with Applications*. Singapore: World Scientific.

Esposito G (1998) *Dirac Operators and Spectral Geometry*. Cambridge: Cambridge University Press.

Gilkey P (1995) *Invariance Theory, the Heat Equation, and the Atiyah–Singer Index Theorem*, 2nd edn, Studies in Advanced Mathematics. Boca Raton, FL: CRC Press.

Grubb G (1996) *The Functional Calculus of Pseudo-Differential Boundary Problems*, 2nd edn., Progress in Mathematics, vol. 65. Boston, MA: Birkhäuser Boston.

Hirzebruch F and Zagier DB (1974) *The Atiyah–Singer Index Theorem and Elementary Number Theory*. Wilmington: Publish or Perish.

Kirsten K (2001) *Spectral Functions in Mathematics and Physics*. Boca Raton, FL: Chapman and Hall/CRC Press.

Melrose R (1993) *The Atiyah–Patodi–Singer Index Theorem*, Research Notes in Mathematics, vol. 4. Wellesley, MA: A K Peters, Ltd.

Palais RS *et al.* (1965) Seminar on the Atiyah–Singer index theorem. *Annals of Mathematical Studies*, 57. Princeton: Princeton University Press.

Vassilevich DV (2003) Heat kernel expansion: user's manual. *Physics Reports* 388: 279–360.

# Inequalities in Sobolev Spaces

**M Vaugon**, Université P.-M. Curie, Paris VI, Paris, France



## Introduction

Given $1 \leq p < n$, it was shown by Sobolev that there exists a constant $K > 0$ such that, for any $u \in C_0^\infty(\mathbb{R}^n)$, the space of smooth functions with compact support in $\mathbb{R}^n$,

$$\left(\int_{\mathbb{R}^n} |u|^{p^\star}\,\mathrm{d}x\right)^{1/p^\star} \leq K\left(\int_{\mathbb{R}^n} |\nabla u|^p\,\mathrm{d}x\right)^{1/p} \qquad [1]$$

where $\nabla u$ is the gradient of $u$ and $p^\star = np/(n-p)$. It is easily seen that $p^\star$ in [1] is critical in the following sense. Let $\|\cdot\|_p$ stand for the $L^p$-norm. For $u \in C_0^\infty(\mathbb{R}^n)$, and $\lambda > 0$, let also $u_\lambda$ be the function given by $u_\lambda(x) = u(\lambda x)$. For $p$ and $q$ two real numbers,

$$\|\nabla u_\lambda\|_p = \lambda^{1-(n/p)}\|\nabla u\|_p$$
$$\|u_\lambda\|_q = \lambda^{-n/q}\|u\|_q$$

Letting $\lambda \to 0$ and $\lambda \to +\infty$, it follows that an inequality like $\|u\|_q \leq K\|\nabla u\|_p$ holds true for all $u$ (in particular for the $u_\lambda$'s) only when $q = p^\star$. To

prove [1], the approach of Sobolev was based on the straightforward representation formula

$$u(x) = -\frac{\Gamma(n/2)}{2\pi^{n/2}}\int_{\mathbb{R}^n}\sum_{k=1}^{n}\frac{x^k - y^k}{|x-y|^n}\partial_k u(y)\mathrm{d}y$$

where $\Gamma$ is the Gamma function, and on an $n$-dimensional version of a theorem of Hardy–Littlewood concerning fractional integrals that we apply to the right-hand side of the above representation formula. More direct arguments were later discovered in independent works by Gagliardo and Nirenberg. In particular, the explicit inequality

$$\left(\int_{\mathbb{R}^n}|u|^{n/(n-1)}\mathrm{d}x\right)^{(n-1)/n} \le \frac{1}{2}\prod_{k=1}^{n}\left(\int_{\mathbb{R}^n}|D_k u|\mathrm{d}x\right)^{1/n} \le \frac{1}{2}\int_{\mathbb{R}^n}|\nabla u|\mathrm{d}x \quad [2]$$

was proved to hold, where $D_k$ is the partial derivative $D_k = \partial/\partial x_k$. Inequality [2] is of the form [1] when $p = 1$, since $1^\star = n/(n-1)$. By geometric measure theory, and the coarea formula, it can be expressed as an isoperimetric type inequality.

There have been several symbols and several definitions for Sobolev spaces. Before they became generally associated with the name of Sobolev, they were sometimes referred to by other names, for instance, as "Beppo Levi spaces." We often find two definitions and two notations in the literature. For $\Omega$ a domain in $\mathbb{R}^n, p \ge 1$ real, and $u$ of class $C^m$ in $\Omega$, we let

$$\|u\|_{m,p} = \left(\sum_{0\le|\alpha|\le m}\|D^\alpha u\|_p^p\right)^{1/p} \quad [3]$$

when the right-hand side makes sense, where $\|\cdot\|_p$ is the $L^p$-norm, $\alpha = (\alpha_1, \ldots, \alpha_n)$ is a multi-index, $|\alpha| = \sum_i \alpha_i$, and $D^\alpha = D_1^{\alpha_1}\cdots D_n^{\alpha_n}$. We define

$$H^{m,p}(\Omega) = \text{the completion of } \{u \in C^m(\Omega) \text{ s.t. } \|u\|_{m,p} < +\infty\} \text{ with respect to the norm } \|\cdot\|_{m,p}$$

$$W^{m,p}(\Omega) = \{u \in L^p(\Omega) \text{ s.t. } D^\alpha u \in L^p(\Omega) \text{ for all } 0 \le |\alpha| \le m\}$$

where $D^\alpha$ is the weak (or distributional) partial derivative of $u$ with respect to the multi-index $\alpha$. Both $H^{m,p}(\Omega)$ and $W^{m,p}(\Omega)$ are Banach spaces (and even Hilbert when $p = 2$). It is easily seen that $H^{m,p}(\Omega) \subset W^{m,p}(\Omega)$, but we had to wait for the work of Meyers and Serrin to realize that $H^{m,p}(\Omega) = W^{m,p}(\Omega)$. The spaces $H^{m,p}(\Omega)$, also denoted $W^{m,p}(\Omega)$, are referred to as Sobolev spaces. The spaces $H_0^{m,p}(\Omega)$, also denoted $W_0^{m,p}(\Omega)$, are defined as the closure of $C_0^\infty(\Omega)$ in $H^{m,p}(\Omega)$, where $C_0^\infty(\Omega)$ is the space of smooth functions with compact support in $\Omega$.

Inequality [1] states that the Sobolev space $H_0^{1,p}(\mathbb{R}^n)$ is naturally embedded in the Lebesgue space $L^{p^\star}(\mathbb{R}^n)$, a particular case of what we now refer to as Sobolev embeddings.

## Sobolev Inequalities and the Sobolev Embedding Theorem in Its First Part

Let $m$ be an integer and let $p \ge 1$ be real. The Sobolev space $H^{m,p}(\mathbb{R}^n)$, also denoted by $W^{m,p}(\mathbb{R}^n)$, is defined by in one of the two equivalent ways:

$$H^{m,p}(\mathbb{R}^n) = \text{the completion of } \{u \in C^m(\mathbb{R}^n) \text{ s.t. } \|u\|_{m,p} < +\infty\} \text{ with respect to the norm } \|\cdot\|_{m,p}$$

or

$$H^{m,p}(\mathbb{R}^n) = \{u \in L^p(\mathbb{R}^n) \text{ s.t. } D^\alpha u \in L^p(\mathbb{R}^n) \text{ for all } 0 \le |\alpha| \le m\}$$

where $D^\alpha$ is the weak (or distributional) partial derivative of $u$ with respect to the multi-index $\alpha$, and $\|\cdot\|_{m,p}$ is as in [3]. The Sobolev space $(H^{m,p}(\mathbb{R}^n), \|\cdot\|_{m,p})$ is a Banach space, and even a Hilbert space when $p = 2$. The space is reflexive when $p > 1$, and we also have that $H^{m,p}(\mathbb{R}^n) = H_0^{m,p}(\mathbb{R}^n)$, where $H_0^{m,p}(\mathbb{R}^n)$ is defined as the closure of $C_0^\infty(\mathbb{R}^n)$ in $H^{m,p}(\mathbb{R}^n)$. What we usually refer to as the first part of Sobolev inequalities can be expressed as follows.

**Sobolev embeddings** (*Part I*). For $p$, $q$ two real numbers with $1 \le q < p$, and $k$, $m$ two integers with $0 \le m < k$, if $1/p = 1/q - (k-m)/n$, then $H^{k,q} \subset H^{m,p}$, and there exists $K > 0$ such that $\|u\|_{m,p} \le K\|u\|_{k,q}$ for all $u \in H^{k,q}$.

The Sobolev theorem in its first part states that the above Sobolev embeddings (resp. inequalities) hold true for the Euclidean space. A particular case of interest is when $k = 1$. In this case, we get, as in the introduction, that for any $1 \le p < n, H^{1,p}(\mathbb{R}^n) \subset L^{p^\star}(\mathbb{R}^n)$ where $p^\star = np/(n-p)$. The embedding for the Euclidean space reduces to the Sobolev inequality [1]. An important remark is that there is a hierarchy for Sobolev embeddings. In particular, that if $H^{1,1} \subset L^{n/(n-1)}, 1^\star = n/(n-1)$, then all the other embeddings $H^{k,q} \subset H^{m,p}$ hold true. Thanks to this remark, the Sobolev embedding theorem for Euclidean space easily follows from an inequality like [2]. The hierarchy for Sobolev embeddings is an easy consequence of Hölder's inequalities when $k = 1$, and of Hölder's inequalities together with Kato's inequality when $k > 1$.

There are several extensions of Sobolev inequalities in the literature. Famous extensions were discovered by Gagliardo and Nirenberg. The Nash inequality, which reads as

$$\left(\int_{\mathbb{R}^n} u^2\,\mathrm{d}x\right)^{(n+2)/n} \le K\left(\int_{\mathbb{R}^n} |u|\,\mathrm{d}x\right)^{4/n} \times \int_{\mathbb{R}^n} |\nabla u|^2\,\mathrm{d}x \qquad [4]$$

for all $u \in H^{1,2}(\mathbb{R}^n)$, is one of the Gagliardo–Nirenberg's inequalities. The Nash inequality easily follows from [1] when $p=2$ and Hölder's inequality. There are also extensions of Sobolev spaces, for instance, spaces of BV-functions or Orlicz–Sobolev spaces.

## The Sobolev Embedding Theorem in Its Second Part

For $m$ integer, let $C_B^m(\mathbb{R}^n)$ be the space of functions of class $C^m$ in $\mathbb{R}^n$ for which the norm

$$\|u\|_{C^m} = \sum_{0\le|\alpha|\le m} \sup_{x\in\mathbb{R}^n} |D^\alpha u(x)|$$

is finite. What we usually refer to as the second part of Sobolev inequalities can be expressed as follows.

**Sobolev embeddings** (*Part II*). For $q \ge 1$ a real number, and $k$, $m$ two integers with $0 \le m < k$, if $1/q - (k-m)/n < 0$, then $H^{k,q} \subset C_B^m$, and there exists $K > 0$ such that $\|u\|_{C^m} \le K\|u\|_{k,q}$ for all $u \in H^{k,q}$.

The Sobolev theorem in its second part states that the above Sobolev embeddings (resp. inequalities) hold true for the Euclidean space. Refinements were then obtained by Morrey with embeddings in Hölder spaces. Let, for instance, $C^{0,\alpha}(\mathbb{R}^n)$ be the Hölder space of continuous functions in $\mathbb{R}^n$ for which the norm

$$\|u\|_{C^{0,\alpha}} = \sup_{x\in\mathbb{R}^n} |u(x)| + \sup_{x\ne y} \frac{|u(y)-u(x)|}{|y-x|^\alpha}$$

is finite. For $k=1, m=0$, and $q \ge 1$ such that $1/q - 1/n < 0$, the embedding $H^{1,q}(\mathbb{R}^n) \subset C_B^0(\mathbb{R}^n)$ can be refined into an embedding like $H^{1,q}(\mathbb{R}^n) \subset C^{0,\alpha}(\mathbb{R}^n)$, where $\alpha \in (0,1)$ is such that $1/q - (1-\alpha)/n < 0$.

## The Case of Domains and the Kondrakov Theorem

The Sobolev embeddings in their first and second parts extend to regular domains $\Omega$. A typical condition is that $\Omega$ satisfies a cone property. When $\Omega$ is bounded, and thus of finite volume, an embedding like $H^{1,p}(\Omega) \subset L^{p^\star}(\Omega)$ implies that we also have that $H^{1,p}(\Omega) \subset L^q(\Omega)$ for all $1 \le q \le p^\star$. The Kondrakov theorem states that such embeddings are all compact, unless $q = p^\star$, in the sense that bounded sequences of functions in $H^{1,p}$ possess converging subsequences in $L^q$.

For $p \ge 1$ real, the Sobolev embedding theorem in its first part provides embeddings of $H^{1,p}$ into Lebesgue spaces when $p < n$, while the Sobolev embedding theorem in its second part provides embeddings of $H^{1,p}$ into Hölder spaces when $p > n$. For $p = n$, it is false that $H^{1,n}$ can be embedded into $L^\infty$. However, when $\Omega$ is bounded, we can prove that $\exp(u) \in L^1(\Omega)$ when $u \in H_0^{1,n}(\Omega)$, and that

$$\int_\Omega \exp(u)\,\mathrm{d}x \le K\exp(\mu\|u\|_{1,n}^n)$$

where $\mu, K > 0$ are independent of $u$. We also have that

$$\int_\Omega \exp(\mu|u|^{n/(n-1)})\,\mathrm{d}x \le K$$

for all $u \in H_0^{1,n}(\Omega)$ such that $\|\nabla u\|_n \le 1$, where $\mu$, $K > 0$ are independent of $u$. Such inequalities are often referred to as Moser–Trüdinger type inequalities.

## The Case of Riemannian Manifolds

Riemannian manifolds are natural extensions of Euclidean space. For $(M, g)$ a Riemannian manifold, $m$ integer, and $p \ge 1$ real, we define the Sobolev space $H^{m,p}(M)$ by

$$H^{m,p}(M) = \text{the completion of } \{u \in C^m(M) \text{ s.t. } \|u\|_{m,p} < +\infty\} \text{ with respect to the norm } \|\cdot\|_{m,p}$$

where $\|u\|_{m,p} = \sum_{i=0}^m \|\nabla^i u\|_p$, $\nabla^i u$ is the $i$th covariant derivative of $u$, and $\|\cdot\|_p$ is the $L^p$-norm in $(M, g)$. A notation like $\|\nabla^i u\|_p$ stands for the $L^p$-norm of the pointwise norm $|\nabla^i u|$ of $\nabla^i u$. Sobolev spaces on manifolds are Banach spaces, even Hilbert when $p=2$, and they are reflexive when $p > 1$. They do not depend on the metric when $M$ is compact.

For compact Riemannian manifolds, everything works as for bounded domains. The Sobolev embeddings in their first and second parts remain valid. The Kondrakov theorem also remains valid. However, since constant functions are in Sobolev spaces when the manifold is compact, the $L^p$-norm of $u$ in the $H^{1,p}$-norm of $u$ should be added to the right-hand side in inequalities like [1]. More precisely, if $(M,g)$ is a compact Riemannian

manifold of dimension $n$, and $1 \leq p < n$, then the inequality for the embedding $H^{1,p}(M) \subset L^{p^\star}(M)$ reads as: there exists $K > 0$ such that for any $u \in H^{1,p}(M)$,

$$\left(\int_M |u|^{p^\star} dv_g\right)^{p/p^\star} \leq K\left(\int_M |\nabla u|^p dv_g + \int_M |u|^p dv_g\right) \quad [5]$$

where $dv_g$ is the Riemannian volume element with respect to $g$. When $(M, g)$ is no longer compact, the Sobolev embedding theorem might become false. A nontrivial key observation is that a Sobolev inequality like [5] on a complete manifold $(M, g)$ implies the existence of a uniform (with respect to the center) lower bound for the volume of balls of radius 1. It follows that for any $n \geq 2$, there exist complete Riemannian $n$-manifolds $(M, g)$ for which, for any $p \in [1, n), H^{1,p}(M) \not\subset L^{p^\star}(M)$. Possible examples are warped products of the real line $\mathbb{R}$ and the $(n-1)$-sphere $S^{n-1}$. When the Ricci curvature is bounded from below, the condition that there is a uniform (with respect to the center) lower bound for the volume of balls of radius 1 is necessary and sufficient in order to get that the Sobolev embeddings are valid.

## Isoperimetric and Euclidean Type Inequalities

Let $(M, g)$ be a complete Riemannian $n$-manifold. Euclidean type inequalities are said to hold on $(M, g)$ if there exists $K > 0$ such that for any $1 \leq p < n$, and any $u \in H^{1,p}(M)$,

$$\left(\int_M |u|^{p^\star} dv_g\right)^{1/p^\star} \leq K\left(\int_M |\nabla u|^p dv_g\right)^{1/p} \quad [6]$$

where $p^\star = np/n - p$. As for the Euclidean space, if the above inequality holds for some $p_0$, then it holds, with distinct $K$, for all $p_0 \leq p < n$. In particular, if the inequality holds for $p = 1$, it holds for all $p$'s. The inequality when $p = 1$ was shown to be true by Hoffman and Spruck when the manifold is simply connected of nonpositive sectional curvature. Such manifolds are referred to as Cartan–Hadamard manifolds. The inequality when $p = 2$ is related to the nonparabolicity of the manifold, namely the existence of a minimal Green's function, and to the behavior of the minimal Green's function.

By geometric measure theory and the coarea formula, [6] when $p = 1$ is equivalent to the isoperimetric inequality

$$\mathrm{Area}_g(\partial\Omega) \geq \frac{1}{C}\mathrm{Vol}_g(\Omega)^{(n-1)/n} \quad [7]$$

where $C > 0, \Omega$ is a smooth bounded domain in $M$, $\mathrm{Area}_g(\partial\Omega)$ is the volume of $\partial\Omega$ for the metric induced by $g$, and $\mathrm{Vol}_g(\Omega)$ is the volume of $\Omega$ with respect to $g$. Moreover, the constants $C$ and $K$ (for $p = 1$) are the same in the sense that if [6] for $p = 1$ holds with $K$, then [7] holds with $C = K$, and if [7] holds with $C$, then [6] for $p = 1$ holds with $K = C$.

The sharp constant for the isoperimetric inequality [7] in Euclidean space is known. When $n = 2$ its value is $C(2) = 1/(4\pi)$ and the sharp isoperimetric inequality is the well-known inequality $L^2 \geq 4\pi A$, where $A$ is the volume of a smooth bounded domain in $\mathbb{R}^2$, and $L$ is the length of its boundary. For arbitrary $n$, the sharp constant $C(n)$ for the isoperimetric inequality is given by

$$C(n) = \frac{1}{n}\left(\frac{n}{\omega_{n-1}}\right)^{1/n} \quad [8]$$

where $\omega_{n-1}$ is the volume of the unit $(n-1)$-sphere. Moreover, still for the Euclidean space, equality holds in the sharp isoperimetric inequality if and only if $\Omega$ is a ball. A famous conjecture concerning sharp isoperimetric inequalities, often referred to as the Cartan–Hadamard conjecture, is that the sharp isoperimetric inequality holds on Cartan–Hadamard manifolds. Thanks to works by Croke, Kleiner, and Weil, the conjecture is known to be true in dimensions 2, 3, and 4. From the Bishop–Gromov comparison theorem, we also get that the only complete manifold of non-negative Ricci curvature for which the sharp isoperimetric inequality holds is the Euclidean space itself.

The sharp constants $K = K(n, p)$ for [6] when $p > 1$ have been computed in Euclidean space by Aubin, Rodemich, and Talenti. The extremal functions were also computed, where, by definition, an extremal function is a function which realizes the case of equality in the inequality. We get that

$$K(n, p) = \frac{1}{n}\left(\frac{n(p-1)}{n-p}\right)^{(p-1)/p} \times \left(\frac{\Gamma(n+1)}{\Gamma(n/p)\Gamma(n+1-n/p)\omega_{n-1}}\right)^{1/n} \quad [9]$$

where, as above, $\Gamma$ is the gamma function. Moreover, $u$ is an extremal function for the sharp inequality in Euclidean space if and only if, up to a scale factor,

$$u(x) = \left(\frac{\mu}{\mu^2 + \frac{|x-a|^{p/(p-1)}}{n(n-2)}}\right)^{(n-p)/p} \quad [10]$$

for some $\mu > 0$, and $a \in \mathbb{R}^n$. When $p = 2$, the functions $u$ in [10] are both the only extremal functions for the sharp Sobolev inequality in Euclidean space, and the only positive solutions of the equation $\Delta u = u^{2^\star - 1}$ in $\mathbb{R}^n$, where $\Delta = -\sum_i D_i^2$ is the Laplace–Beltrami operator (the usual Laplacian with a minus sign in front of it). Sharp constants are also known for several of the Gagliardo–Nirenberg inequalities in Euclidean space. The sharp constant for the Nash inequality in Euclidean space was computed by Carlen and Loss. If the sharp isoperimetric inequality holds on a complete Riemannian $n$-manifold, then the sharp inequalities [6] hold for all $1 \le p < n$.

## Sharp Inequalities on Compact Riemannian Manifolds

The study of sharp Sobolev inequalities on compact manifolds if often referred to as the $AB$ program for Sobolev inequalities. For $(M, g)$ a compact Riemannian $n$-manifold, and $1 \le p < n$, [5] can be rewritten in two different forms:

$$\left(\int_M |u|^{p^\star} \mathrm{d}v_g\right)^{1/p^\star} \le A\left(\int_M |\nabla u|^p \mathrm{d}v_g\right)^{1/p} + B\left(\int_M |u|^p \mathrm{d}v_g\right)^{1/p} \qquad [11]$$

and

$$\left(\int_M |u|^{p^\star} \mathrm{d}v_g\right)^{p/p^\star} \le A' \int_M |\nabla u|^p \mathrm{d}v_g + B' \int_M |u|^p \mathrm{d}v_g \qquad [12]$$

where $A, B, A', B'$ are positive constants independent of $u$. An easy remark is that if [12] holds with constants $A'$ and $B'$, then [11] holds with $A = (A')^{1/p}$ and $B = (B')^{1/p}$. The sharp first (resp. second) constants in [11] and [12] are defined as the lowest possible values for $A$ and $A'$ (resp. for $B$ and $B'$) in [11] and [12]. The sharp first constants are independent of the manifold and are given by $A' = A^p = K(n, p)^p$, where $K(n, p)$ is as in [9]. The sharp second constants depend on the manifold and are given by $B' = B^p = V_g^{-p/n}$, where $V_g$ is the volume of $(M, g)$. A typical question in the $AB$ program is to know whether or not we can take $A$ or $B$ to be the sharp constants in [11] and, similarly, whether or not we can take $A'$ or $B'$ to be the sharp constants in [12]. Another typical question in the $AB$ program is whether or not there are nonzero extremal functions for the saturated form of the sharp inequalities when they are valid. Concerning the $B$-part of the program, the sharp inequality [11] with $B = V_g^{-1/n}$ is true on any manifold, and constant functions are extremal functions. On the other hand, it can be proved that the stronger [12] with $B' = V_g^{-p/n}$ is always false when $p > 2$, whatever the manifold. Concerning the $A$-part of the $AB$-program, Hebey and Vaugon proved that the sharp inequality [12] with $A' = K(n, 2)^2$ is true on any manifold. In other words, for any compact Riemannian manifold $(M, g)$ of dimension $n \ge 3$, there exists $B' > 0$ such that, for any $u \in H^{1,2}(M)$,

$$\left(\int_M |u|^{2^\star} \mathrm{d}v_g\right)^{2/2^\star} \le K(n, 2)^2 \int_M |\nabla u|^2 \mathrm{d}v_g + B' \int_M |u|^2 \mathrm{d}v_g \qquad [13]$$

We then get the saturated form of [13] by taking $B' = B'(g)$ to be the lowest possible $B'$ in [13]. In general, when $p \ne 2$, we can prove that the sharp inequality [11] with $A = K(n, p)$ is true on any manifold, and that there are nonzero extremal functions for the saturated form of the sharp inequality. On the other hand, the stronger [12] with $A' = K(n, p)^p$ when $p > 2$ is false when the curvature is positive, but true when the curvature is negative. The $p = 2$ case in the $A$-part of the $AB$ program is of importance for its connection with the Yamabe problem. The $p = 1$ case in the $A$-part of the $AB$ program is of importance for its connection with the isoperimetric inequality. The $AB$ program has also been considered for Gagliardo–Nirenberg inequalities, including the Nash inequality, and Sobolev–Poincaré inequalities on compact manifolds.

## Further Reading


Adams RA (1978) *Sobolev Spaces*. San Diego: Academic Press.

Aubin T (1998) *Some Nonlinear Problems in Riemannian Geometry*. Springer Monographs in Mathematics. Berlin: Springer.

Druet O and Hebey E (2002) The AB Program in Geometric Analysis: Sharp Sobolev Inequalities and Related Problems. *Memoirs of the American Mathematical Society, 160, 761.* American Mathematical Society.

Evans LC and Gariepy RF (1992) *Measure Theory and Fine Properties of Functions*. CRC Press.

Hebey E (2000) *Nonlinear Analysis on Manifolds: Sobolev Spaces and Inequalities*. Courant Lecture Notes in Mathematics, vol. 5. American Mathematical Society.

Maz'ja VG (1985) *Sobolev Spaces*, Springer Series in Soviet Mathematics. Berlin: Springer.

# Infinite-Dimensional Hamiltonian Systems

**R Schmid**, Emory University, Atlanta, GA, USA



## Introduction

Infinite-dimensional Hamiltonian systems arise in many areas in pure and applied mathematics and in mathematical physics. These are partial differential equations (PDEs) which can be written as evolution equations (dynamical systems) in the form

$$\dot{F} = \{F, H\}$$

where $H$ is the Hamiltonian ("energy") and $\{.,.\}$ is a Poisson bracket on an infinite-dimensional phase space, called Poisson manifold. Unlike finite-dimensional Hamiltonian systems, which are ordinary differential evolution equations on finite-dimensional phase spaces, for which general existence and uniqueness theorems for solutions exist, this is not the case for PDEs. There are no general existence and uniqueness theorems for solutions of infinite-dimensional Hamiltonian systems. These have to be established case by case. This article gives only a broad mathematical framework of infinite-dimensional Hamiltonian systems. Precise definitions are presented and the concept is illustrated through physical examples.

## Hamilton's Equations on Poisson Manifolds

A Poisson manifold is a manifold $P$ (in general infinite dimensional) equipped with a bilinear operation $\{.,.\}$, called Poisson bracket, on the space $C^\infty(P)$ of smooth functions on $P$ such that:

1. $(C^\infty(P), \{.,.\})$ is a Lie algebra, that is, $\{.,.\}: C^\infty(P) \times C^\infty(P) \to C^\infty(P)$ is bilinear, skew-symmetric and satisfies the Jacobi identity $\{\{F,G\},H\} + \{\{H,F\},G\} + \{\{G,H\},F\} = 0$ for all $F, G, H \in C^\infty(P)$ and
2. $\{.,.\}$ satisfies the Leibniz rule, that is, $\{.,.\}$ is a derivation in each factor: $\{F \cdot G, H\} = F \cdot \{G, H\} + G \cdot \{F, H\}$, for all $F, G, H \in C^\infty(P)$.

The notion of Poisson manifolds was rediscovered many times under different names, starting with Lie, Dirac, Pauli, and others. The name Poisson manifold was coined by Lichnerowicz.

For any $H \in C^\infty(P)$, the Hamiltonian vector field $X_H$ is defined by

$$X_H(F) = \{F, H\}, \quad F \in C^\infty(P)$$

It follows from (2) that, indeed, $X_H$ defines a derivation on $C^\infty(P)$, hence a vector field on $P$. Hamilton's equations of motion for a function $F \in C^\infty(P)$ with Hamiltonian $H$ (energy function) are then defined by the flow (integral curves) of the vector field $X_H$, that is,

$$\dot{F} = X_H(F) = \{F, H\} \qquad [1]$$

where the overdot implies differentiation with respect to time. $F$ is then called a Hamiltonian system on $P$ with energy (Hamiltonian function) $H$.

## Examples of Poisson Manifolds and Hamilton's Equations

### Finite-Dimensional Classical Mechanics

For finite-dimensional classical mechanics, we take $P = \mathbb{R}^{2n}$ and coordinates $(q^1, \ldots, q^n, p_1, \ldots, p_n)$ with the standard Poisson bracket for any two functions $F(q^i, p_i), H(q^i, p_i)$ given by

$$\{F, H\} = \sum_{i=1}^{n} \frac{\partial F}{\partial p_i}\frac{\partial H}{\partial q^i} - \frac{\partial H}{\partial p_i}\frac{\partial F}{\partial q^i} \qquad [2]$$

Then the classical Hamilton's equations are

$$\begin{aligned} \dot{q}^i &= \{q^i, H\} = \frac{\partial H}{\partial p_i} \\ \dot{p}_i &= \{p_i, H\} = -\frac{\partial H}{\partial q^i} \end{aligned} \qquad [3]$$

$i = 1, \ldots, n$. This finite-dimensional Hamiltonian system is a system of ordinary differential equations for which there are well-known existence and uniqueness theorems, that is, it has locally unique smooth solutions, depending smoothly on the initial conditions.

**Example: harmonic oscillator** As a concrete example, consider the harmonic oscillator: here $P = \mathbb{R}^2$ and the Hamiltonian (energy) is $H(q,p) = \frac{1}{2}(q^2 + p^2)$. Then Hamilton's equations are

$$\dot{q} = p, \qquad \dot{p} = -q \qquad [4]$$

### Infinite-Dimensional Classical Field Theory

Let $V$ be a Banach space and $V^*$ its dual space with respect to a pairing $\langle .,. \rangle : V \times V^* \to \mathbb{R}$ (i.e., $\langle .,. \rangle$ is a symmetric, bilinear, and nondegenerate function). On $P = V \times V^*$, the canonical Poisson

bracket for $F, H \in C^\infty(P)$, $\varphi \in V$, and $\pi \in V^*$ is given by

$$\{F,\, H\} = \left\langle \frac{\delta F}{\delta \pi}, \frac{\delta H}{\delta \varphi} \right\rangle - \left\langle \frac{\delta H}{\delta \pi}, \frac{\delta F}{\delta \varphi} \right\rangle \qquad [5]$$

where the functional derivatives $\delta F/\delta\pi \in V, \delta F/\delta\varphi \in V^*$ are the "duals" under the pairing $\langle .\,,.\rangle$ of the partial gradients $D_1F(\pi) \in V^*, D_2F(\varphi) \in V^{**} \simeq V$. The corresponding Hamilton's equations are

$$\begin{aligned} \dot\varphi &= \{\varphi,\, H\} = \frac{\delta H}{\delta \pi} \\ \dot\pi &= \{\pi,\, H\} = -\frac{\delta H}{\delta \varphi} \end{aligned} \qquad [6]$$

As a special case in finite dimensions, if $V \simeq \mathbb{R}^n$ so $V^* \simeq \mathbb{R}^n$ and $P = V \times V^* \simeq \mathbb{R}^{2n}$, and the pairing is the standard inner product in $\mathbb{R}^n$, then the Poisson bracket [5] and Hamilton's equations [6] are identical with [2] and [3], respectively.

**Example: wave equations** As a concrete example, consider the wave equations. Let $V = C^\infty(\mathbb{R}^3)$ and $V^* = \mathrm{Den}(\mathbb{R}^3)$ (densities) and the $L^2$ pairing $\langle\varphi, \pi\rangle = \int \varphi(x)\pi(x)\,\mathrm{d}x$. Take the Hamiltonian to be

$$H(\varphi,\, \pi) = \int \Big(\tfrac{1}{2}\pi^2 + \tfrac{1}{2}|\nabla\varphi|^2 + F(\varphi)\Big)\mathrm{d}x$$

where $F$ is some function on $V$. Then Hamilton's equations [6] become

$$\dot\varphi = \pi, \quad \dot\pi = \nabla^2\varphi - F'(\varphi) \qquad [7]$$

where the prime denotes differentiation with respect to $\varphi$, which imply the wave equation

$$\frac{\partial^2\varphi}{\partial t^2} = \nabla^2\varphi - F'(\varphi) \qquad [8]$$

Different choices of $F$ give different wave equations, for example, for $F = 0$ we get the linear wave equation

$$\frac{\partial^2\varphi}{\partial t^2} = \nabla^2\varphi$$

for $F = (1/2)m\varphi$, we get the Klein–Gordon equation

$$\nabla^2\varphi - \frac{\partial^2\varphi}{\partial t^2} = m\varphi$$

So, these wave equations and the Klein–Gordon equation are infinite-dimensional Hamiltonian systems on $P = C^\infty(\mathbb{R}^3) \times \mathrm{Den}\,(\mathbb{R}^3)$.

### Cotangent Bundles

The finite-dimensional examples of Poissson brackets [2] and Hamilton's equations [3] and the infinite-dimensional examples [5] and [6] are the local versions of the general case where $P = T^*Q$ is the cotangent bundle (phase space) of a manifold $Q$ (configuration space). If $Q$ in an $n$-dimensional manifold, then $T^*Q$ is a $2n$-Poisson manifold locally isomorphic to $\mathbb{R}^{2n}$ whose Poisson bracket is locally given by [2] and Hamilton's equations are locally given by [3]. If $Q$ is an infinite-dimensional Banach manifold, then $T^*Q$ is a Poisson manifold locally isomorphic to $V \times V^*$ whose Poisson bracket is given by [5] and Hamilton's equations are locally given by [6].

### Symplectic Manifolds

All the examples above are special cases of symplectic manifolds $(P, \omega)$. This means that $P$ is equipped with a symplectic structure $\omega$ which is a closed $(\mathrm{d}\omega = 0)$, (weakly) nondegenerate 2-form on the manifold $P$. Then, for any $H \in C^\infty(P)$, the corresponding Hamiltonian vector field $X_H$ is defined by $\mathrm{d}H = \omega(X_H,\, .)$ and the canonical Poisson bracket is given by

$$\{F, H\} = \omega(X_F, X_H), \quad F, H \in C^\infty(P) \qquad [9]$$

For example, on $\mathbb{R}^{2n}$ the canonical symplectic structure $\omega$ is given by $\omega = \sum_{i=1}^n \mathrm{d}p_i \wedge \mathrm{d}q^i = \mathrm{d}\theta$, where $\theta = \sum_{i=1}^n p_i \wedge \mathrm{d}q^i$. The same formula for $\omega$ holds locally in $T^*Q$ for any finite-dimensional $Q$ (Darboux's lemma). For the infinite-dimensional example $P = V \times V^*$, the symplectic form $\omega$ is given by $\omega((\varphi_1, \pi_1), (\varphi_2, \pi_2)) = \langle\varphi_1, \pi_2\rangle - \langle\varphi_2, \pi_1\rangle$. Again, these two formulas for $\omega$ are identical if $V = \mathbb{R}^n$.

#### Remarks

(i) If $P$ is a finite-dimensional symplectic manifold, then $P$ is even dimensional.
(ii) If the Poisson bracket $\{.\,,.\}$ is nondegenerate, then $\{.\,,.\}$ comes form a symplectic form $\omega$, that is, $\{.\,,.\}$ is given by [9].

### The Lie–Poisson Bracket

Not all Poisson brackets are of the from given in the above examples [2], [5], and [9], that is, not all Poisson manifolds are symplectic manifolds. An important class of Poisson bracket is the so-called Lie–Poisson bracket. It is defined on the dual of any Lie algebra. Let $G$ be a Lie group with Lie algebra $\mathfrak{g} = T_eG \simeq$ {left-invariant vector fields on $G$} and let $[.\,,.]$ denote the Lie bracket (commutator) on $\mathfrak{g}$. Let $\mathfrak{g}^*$ be the dual of a $\mathfrak{g}$ with respect to a pairing $\langle .\,,.\rangle : \mathfrak{g}^* \times \mathfrak{g} \to \mathbb{R}$. Then, for any $F, H \in C^\infty(\mathfrak{g}^*)$ and $\mu \in \mathfrak{g}^*$, the Lie–Poisson bracket is defined by

$$\{F, H\}(\mu) = \pm \left\langle \mu, \left[\frac{\delta F}{\delta\mu}, \frac{\delta H}{\delta\mu}\right] \right\rangle \qquad [10]$$

where $\delta F/\delta\mu, \delta H/\delta\mu \in \mathfrak{g}$ are the "duals" of the gradients $DF(\mu), DH(\mu) \in \mathfrak{g}^{**} \simeq \mathfrak{g}$ under the pairing $\langle .,. \rangle$. Note that the Lie–Poisson bracket is degenerate in general, for example, for $G = \mathrm{SO}(3)$ the vector space $\mathfrak{g}^*$ is three dimensional, so the Poisson bracket [10] cannot come from a symplectic structure. This Lie–Poisson bracket can also be obtained in a different way by taking the canonical Poisson bracket on $T^*G$ (locally given by [2] and [5] and then restrict it to the fiber at the identity $T_e^*G = \mathfrak{g}^*$. In this sense, the Lie–Poisson bracket [10] is induced from the canonical Poisson bracket on $T^*G$. It is induced by the symmetry of left-multiplication, as discussed in the next section.

**Example: rigid body** A concrete example of the Lie–Poisson bracket is given by the rigid body. Here $G = \mathrm{SO}(3)$ is the configuration space of a free rigid body. Identifying the Lie algebra $(\mathfrak{so}(3), [.,.])$ with $(\mathbb{R}^3, \times)$, where $\times$ is the vector product on $\mathbb{R}^3$ and $\mathfrak{g}^* = \mathfrak{so}(3)^* \simeq \mathbb{R}^3$, the Lie–Poisson bracket translates into

$$\{F, H\}(m) = -m \cdot (\nabla F \times \nabla H) \qquad [11]$$

For any $F \in C^\infty(\mathfrak{so}(3)^*)$, we have

$$\begin{aligned}\frac{\mathrm{d}F}{\mathrm{d}t}(m) &= \nabla F \cdot \dot{m} = \{F, H\}(m) \\ &= -m \cdot (\nabla F \times \nabla H) = \nabla F \cdot (m \times \nabla H)\end{aligned}$$

hence $\dot{m} = m \times \nabla H$. With the Hamiltonian

$$H = \frac{1}{2}\left(\frac{m_1^2}{I_1^2} + \frac{m_2^2}{I_2^2} + \frac{m_3^2}{I_3^2}\right)$$

we get Hamilton's equation as

$$\dot{m}_1 = \frac{I_2 - I_3}{I_2 I_3} m_2 m_3, \qquad \dot{m}_2 = \frac{I_3 - I_1}{I_3 I_1} m_3 m_1$$
$$\dot{m}_3 = \frac{I_1 - I_2}{I_1 I_2} m_1 m_2$$

These are Euler's equations for the free rigid body.

### Reduction by Symmetries

The examples discussed so far are all canonical examples of Poisson brackets, defined either on a symplectic manifold $(P, \omega)$ or $T^*Q$, or on the dual of a Lie algebra $\mathfrak{g}^*$. Different, noncanonical Poisson brackets can arise from symmetries. Assume that a Lie group $G$ is acting in a Hamiltonian way on the Poisson manifold $(P, \{..\})$. This means that we have a smooth map $\varphi : G \times P \to P : \varphi(g, p) = g \cdot p$ such that the induced maps $\varphi_g = \varphi(g, .) : P \to P$ are canonical transformations, for each $g \in G$. In terms of Poisson manifolds, a canonical transformation is a smooth map that preserves the Poisson bracket. So, the action of $G$ on $P$ is a Hamiltonian action if $\varphi_g^*\{F, H\} = \{\varphi_g^* F, \varphi_g^* H\}$ for all $F, H \in C^\infty(P)$, $g \in G$. For any $\xi \in \mathfrak{g}$, the canonical transformations $\varphi_{\exp(t\xi)}$ generate a Hamiltonian vector field $\xi_F$ on $P$ and a momentum map $J : P \to \mathfrak{g}^*$ given by $J(x)(\xi) = F(x)$, which is $\mathrm{Ad}^*$ equivariant.

If a Hamiltonian system $X_H$ is invariant under a Lie group action, that is, $H(\varphi_g(x)) = H(x)$, then we obtain a reduced Hamiltonian system on a reduced phace space (reduced Poisson manifold). We recall the Marsden–Weinstein reduction theorem:

**Reduction Theorem** *For a Hamiltonian action of a Lie group $G$ on a Poisson manifold $(P, \{.,.\})$, there is an equivariant momentum map $J : P \to \mathfrak{g}^*$, and for every regular $\mu \in \mathfrak{g}^*$ the reduced phase space $P_\mu \equiv J^{-1}(\mu)/G_\mu$ carries an induced Poisson structure $\{.,.\}_\mu$, ($G_\mu$ the isotropy group). Any $G$-invariant Hamiltonian $H$ on $P$ defines a Hamiltonian $H_\mu$ on the reduced phase space $P_\mu$ and the integral curves of the vector field $X_H$ project onto integral curves of the induced vector field $\hat{X}_{H_\mu}$ on the reduced space $P_\mu$.*

**Example: rigid body** The rigid body discussed above can be viewed as an example of this reduction theorem. If $P = T^*G$ and $G$ is acting on $T^*G$ by the cotangent lift of the left-translation $l_g : G \to G, l_g(h) = gh$, then the momentum map $J : T^*G \to \mathfrak{g}^*$ is given by $J(\alpha_g) = T_e^* R_g(\alpha_g)$ and the reduced phase space $(T^*G)_\mu = J^{-1}(\mu)/G_\mu$ is isomorphic to the coadjoint orbit $\mathcal{O}_\mu$ through $\mu \in \mathfrak{g}^*$. Each coadjoint orbit $\mathcal{O}_\mu$ carries a natural symplectic structure $\omega_\mu$ and in this case, the reduced Lie–Poisson bracket $\{.,.\}_\mu$ on the coadjoint orbit $\mathcal{O}_\mu$ is induced by the symplectic form $\omega_\mu$ on $\mathcal{O}_\mu$ as in [9]. Furthermore, $T^*G/G \simeq \mathfrak{g}^*$, and the induced Poisson bracket $\{.,.\}_\mu$ on $\mathcal{O}_\mu$ is identical with the Lie–Poisson bracket restricted to the coadjoint orbit $\mathcal{O}_\mu \subset \mathfrak{g}^*$. For the rigid body this construction is applied to $G = \mathrm{SO}(3)$.

We now discuss some infinite-dimensional examples of reduced Hamiltonian systems.

## Infinite-Dimensional Lie Groups

A general theory of infinite-dimensional Lie groups is hardly developed. Even Bourbaki only develops a theory of infinite-dimensional manifolds, but all of the important theorems about Lie groups are stated for finite-dimensional ones.

An infinite-dimensional Lie group $\mathcal{G}$ is a group and an infinite-dimensional manifold with smooth group operations

$$m : \mathcal{G} \times \mathcal{G} \to \mathcal{G}, \quad m(g, h) = g \cdot h, \qquad C^\infty \qquad [12]$$

$$i : \mathcal{G} \to \mathcal{G}, \quad i(g) = g^{-1}, \qquad C^\infty \qquad [13]$$

Such a Lie group $\mathcal{G}$ is locally diffeomorphic to an infinite-dimensional vector space. This can be a Banach space whose topology is given by a norm $\|\cdot\|$, a Hilbert space whose topology is given by an inner product $\langle.,.\rangle$, or a Frechet space whose topology is given by a metric but not by a norm. Depending on the choice of the topology on $\mathcal{G}$, the Banach, Hilbert, or Frechet Lie groups, respectively, can be treated.

The Lie algebra $\mathfrak{g}$ of $\mathcal{G}$ is defined as $\mathfrak{g} = \{\text{left-invariant vector fields on } \mathcal{G}\} \simeq T_e\mathcal{G}$, where the isomorphism is given (as in finite dimensions) by

$$\xi \in T_e\mathcal{G} \mapsto X^\xi(g) = T_e L_g(\xi) \qquad [14]$$

and the Lie bracket on $\mathfrak{g}$ is induced by the Lie bracket of left-invariant vector fields $[\xi, \eta] = [X^\xi, X^\eta](e)$, $\xi, \eta \in \mathfrak{g}$.

These definitions in infinite dimensions are identical with the definitions in finite dimensions. The big difference although is that infinite-dimensional manifolds, hence Lie groups, are not locally compact. For Frechet Lie groups, one has the additional nontrivial difficulty of defining the differentiability of functions defined on a Frechet space. Hence, the very definition of a Frechet manifold is not canonical. This problem does not arise for Banach and Hilbert Lie groups; the differential calculus extends in a straightforward manner from $\mathbb{R}^n$ to Banach and Hilbert spaces, but not to Frechet spaces.

## Finite- versus Infinite-Dimensional Lie Groups

The lack of local compactness of infinite-dimensional Lie groups causes some deficiencies of the Lie theory in infinite dimensions. Some classical results in finite dimensions are summarized below, which are not true in general in infinite dimensions:

1. The exponential map $exp : \mathfrak{g} \to G$ is defined as follows: To each $\xi \in \mathfrak{g}$ we assign the corresponding left-invariant vector field $X^\xi$ defined by [14]. We take the flow $\varphi^\xi(t)$ of $X^\xi$ and define $exp(\xi) = \varphi^\xi(1)$. The exponential map is a local diffeomorphism from a neighborhood of zero in $\mathfrak{g}$ onto a neighborhood of the identity in $G$; hence, $exp$ defines canonical coordinates on the Lie group $G$. This is not true in infinite dimensions.
2. If $f_1, f_2 : G_1 \to G_2$ are smooth Lie group homomorphisms (i.e., $f_i(g \cdot h) = f_i(g) \cdot f_i(h), i = 1, 2$) with $T_e f_1 = T_e f_2$, then locally $f_1 = f_2$. This is not true in infinite dimensions.
3. If $H$ is a closed subgroup of $G$, then $H$ is a Lie subgroup of $G$. This is not true in infinite dimensions.
4. For any finite-dimensional Lie algebra $\mathfrak{g}$, there exists a connected Lie group $G$ whose Lie algebra is $\mathfrak{g}$, that is, such that $\mathfrak{g} \simeq T_e G$. This is not true in infinite dimensions.

Some classical finite-dimensional examples of Lie groups are the matrix groups GL($n$), SL($n$), O($n$), SO($n$), U($n$), SU($n$), Sp($n$) with smooth group operations given by matrix multiplication and matrix inversion.

## Examples of Infinite-Dimensional Lie Groups

### Abelian Gauge Group $\mathcal{G} = (C^\infty(M), +)$

Let $M$ be a finite-dimensional manifold and let $\mathcal{G} = C^\infty(M)$. With group operation being addition, that is, $m(f, g) = f + g$, $i(f) = -f$, $e = 0$. $\mathcal{G}$ is an abelian $C^\infty$ Frechet Lie group with Lie algebra $\mathfrak{g} = T_e C^\infty(M) \simeq C^\infty(M)$, with trivial bracket $[\xi, \eta] = 0$, and $exp = \text{id}$. If one completes these spaces in the $C^k$-norm, $k < \infty$ then $\mathcal{G}^k$ is a Banach Lie group, and if the $H^s$-Sobolev norm is used with $s > (1/2) \dim M$ then $\mathcal{G}^s$ is a Hilbert Lie group.

**Application of $\mathcal{G} = (C^\infty(M), +)$ to Maxwell's equations** Let $E$, $B$ be the electric and magnetic fields on $\mathbb{R}^3$; then Maxwell's equations for a charge density $\rho$ are:

$$\dot{E} = \text{curl}\, B, \qquad \dot{B} = -\text{curl}\, E \qquad [15]$$

$$\text{div}\, B = 0, \qquad \text{div}\, E = \rho \qquad [16]$$

Let $A$ be the magnetic potential such that $B = -\text{curl}\, A$. As configuration space, we take $V = \text{Vec}(\mathbb{R}^3)$, vector fields (potentials) on $\mathbb{R}^3$, so $A \in V$, and as phase space, we have $P = T^*V \simeq V \times V^* \ni (A, E)$, with the standard $L^2$ pairing $\langle A, E \rangle = \int A(x)E(x)\, dx$, and canonical Poisson bracket given by [5], which becomes

$$\{F, H\}(A, E) = \int \left( \frac{\delta F}{\delta A} \frac{\delta H}{\delta E} - \frac{\delta H}{\delta A} \frac{\delta F}{\delta E} \right) dx \qquad [17]$$

As Hamiltonian, we take the total electromagnetic energy

$$H(A,\ E) = \frac{1}{2}\int (|\text{curl } A|^2 + |E|^2)\ \mathrm{d}x$$

Then Hamilton's equations in the canonical variables $A$ and $E$ are

$$\dot{A} = \frac{\delta H}{\delta E} = E \ \Rightarrow\ \dot{B} = -\text{curl}\, E$$

and

$$\dot{E} = -\frac{\delta H}{\delta A} = -\text{curl}\,\text{curl}\, A = \text{curl}\, B$$

So the first two equations of Maxwell's equations [15] are Hamilton's equations, the third one is obtained automatically from the potential $\text{div} B = -\text{div}\,\text{curl} A = 0$ and the fourth equation, $\text{div} E = \rho$, is obtained through the following symmetry (gauge invariance): the Lie group $\mathcal{G} = (C^\infty(\mathbb{R}^3), +)$ acts on $V$ by $\varphi\cdot A = A + \nabla\varphi, \varphi\in\mathcal{G}, A\in V$. The lifted action to $V\times V^*$ becomes $\varphi\cdot(A,E) = (A+\nabla\varphi, E)$, and has the momentum map $J: V\times V^* \to \mathfrak{g}^* \simeq$ {charge densities}

$$J(A,E) = \text{div}\, E \qquad [18]$$

With $\mathfrak{g} = C^\infty(\mathbb{R}^3)$ and $\mathfrak{g}^* = \text{Den}(\mathbb{R}^3)$, we identify the elements of $\mathfrak{g}^*$ with charge densities. The Hamiltonian $H$ is $\mathcal{G}$ invariant, that is, $H(\varphi\cdot(A,E)) = H(A+\nabla\varphi, E) = H(A,E)$. Then the reduced phase space for $\rho\in\mathfrak{g}^*$ is

$$(V\times V^*)_\rho = J^{-1}(\rho)/G = \{(E,\ B)|\text{div} E = \rho,\ \text{div} B = 0\}$$

and the reduced Hamiltonian is

$$H_\rho(E,\ B) = \frac{1}{2}\int (|E|^2 + |B|^2)\ \mathrm{d}x \qquad [19]$$

The reduced Poisson bracket becomes, for any functions $F$, $H$ on $(V\times V^*)_\rho$,

$$\{F,\ H\}_\rho(E,\ B) = \int\left(\frac{\delta F}{\delta E}\cdot\text{curl}\ \frac{\delta H}{\delta B} - \frac{\delta H}{\delta E}\cdot\text{curl}\ \frac{\delta F}{\delta B}\right)\mathrm{d}x \qquad [20]$$

and a straightforward computation shows that

$$\dot{F} = \{F,\ H_\rho\}_\rho \Leftrightarrow \begin{cases}\dot{E} = \text{curl}\, B, & \dot{B} = -\text{curl}\, E\\ \text{div}\, B = 0, & \text{div}\, E = \rho\end{cases} \qquad [21]$$

So, Maxwell's equations [15], [16] form an infinite-dimensional Hamiltonian system on this reduced phase space with respect to the reduced Poisson bracket.

### Abelian Gauge Group $\mathcal{G} = (C^\infty(M, \mathbb{R} - \{0\}), \cdot)$

Let $M$ be a finite-dimensional manifold and let $\mathcal{G} = C^\infty(M, \mathbb{R} - \{0\})$, the group operation being the multiplication, that is, $m(f,g) = f\cdot g$, $i(f) = f^{-1}$, $e = 1$. For $k < \infty$, $C^k(M, \mathbb{R} - \{0\})$ is open in $C^\infty(M, \mathbb{R})$, and if $M$ is compact then $C^k(M, \mathbb{R} - \{0\})$ is a Banach Lie group. If $s > (1/2)\dim M$ then $H^s(M, \mathbb{R} - \{0\})$ is closed under multiplication, and if $M$ is compact then $H^s(M, \mathbb{R} - \{0\})$ is a Hilbert Lie group.

### Nonabelian Gauge Groups $\mathcal{G} = (C^k(M, G), \cdot)$

The abelian example can be generalized by replacing $\mathbb{R} - \{0\}$ with any finite-dimensional (nonabelian) Lie group $G$. Let $\mathcal{G} = C^k(M, G)$ with pointwise group operations $m(f,g)(x) = f(x)\cdot g(x), x\in M$ and $i(f)(x) = (f(x))^{-1}$, where "$\cdot$" and "$(\,.\,)^{-1}$" are the operations in $G$. If $k < \infty$ then $C^k(M, G)$ is a Banach Lie group. Let $g$ denote the Lie algebra of $G$, then the Lie algebra of $\mathcal{G} = C^k(M, G)$ is $\mathfrak{g} = C^k(M, g)$, with pointwise Lie bracket $[\xi,\eta](x) = [\xi(x), \eta(x)], x\in M$, the latter bracket being the Lie bracket in $g$. The exponential map $exp: g\to G$ defines the exponential map $EXP: \mathfrak{g} = C^k(M, g)\to\mathcal{G} = C^k(M, G)$, $EXP(\xi) = exp\circ\xi$, which is a local diffeomorphism. The same holds for $H^s(M, G)$ if $s > (1/2)\dim M$.

Applications of these infinite-dimensional Lie groups are in gauge theories and quantum field theory, where they appear as groups of gauge transformations.

### Loop Groups $\mathcal{G} = C^k(S^1, G)$

As a special case of the example above, we take $M = S^1$, the circle. Then $\mathcal{G} = C^k(S^1, G) = \mathcal{L}^k(G)$ is called a loop group and $\mathfrak{g} = C^k(S^1, g) = l^k(g)$ its loop algebra. They find applications in the theory of affine Lie algebras, Kac–Moody Lie algebras (central extensions), completely integrable systems, soliton equations (Toda, Korteweg–de Vries (KdV), Kadomtsev–Petviashvili (KP)), quantum field theory. Central extensions of Loop algebras are examples of infinite-dimensional Lie algebras which need not have a corresponding Lie group.

### Diffeomorphism Groups

Among the most important "classical" infinite-dimensional Lie groups are the diffeomorphism groups of manifolds. Their differential structure is not the one of a Banch Lie group as defined above. Nevertheless, they have important applications.

Let $M$ be a compact manifold (the noncompact case is technically much more complicated but similar results are true) and let $\mathcal{G} = \text{Diff}^\infty(M)$ be the group of all smooth diffeomorphisms on $M$,

group operation being the composition, that is, $m(f,g)=f\circ g$, $i(f)=f^{-1}$, $e=\mathrm{id}_M$. For $C^\infty$ diffeomorphisms, $\mathrm{Diff}^\infty(M)$ is a Frechet manifold and there are nontrivial problems with the notion of smooth maps between Frechet spaces. There is no canonical extension of the differential calculus from Banach spaces (same as for $\mathbb{R}^n$) to Frechet spaces. One possibility is to generalize the notion of differentiability. For example, if we use the so-called $C^\infty_\Gamma$ differentiability, then $\mathcal{G}=\mathrm{Diff}^\infty(M)$ becomes a $C^\infty_\Gamma$ Lie group with $C^\infty_\Gamma$ differentiable group operations. These notions of differentiability are difficult to apply to concrete examples. Another possibility is to complete $\mathrm{Diff}^\infty(M)$ in the Banach $C^k$-norm, $0\le k<\infty$, or in the Sobolev $H^s$-norm, $s>(1/2)\dim M$; $\mathrm{Diff}^k(M)$ and $\mathrm{Diff}^s(M)$ become, in this case, Banach and Hilbert manifolds, respectively. Then we consider the inverse limits of these Banach and Hilbert Lie groups, respectively:

$$\mathrm{Diff}^\infty(M)=\varprojlim \mathrm{Diff}^k(M) \qquad [22]$$

becomes the so-called inverse limit of Banach (ILB) Lie group, or with the Sobolev topologies

$$\mathrm{Diff}^\infty(M)=\varprojlim \mathrm{Diff}^s(M) \qquad [23]$$

becomes the so-called inverse limit of Hilbert (ILH) Lie group. Nevertheless, the group operations are not smooth, but have the following differentiability properties. If the diffeomorphism group is equipped with the Sobolev $H^s$-topology, then $\mathrm{Diff}^s(M)$ becomes a $C^\infty$ Hilbert manifold if $s>(1/2)\dim M$ and the group multiplication

$$m:\mathrm{Diff}^{s+k}(M)\times\mathrm{Diff}^s(M)\to\mathrm{Diff}^s(M) \qquad [24]$$

is $C^k$ differentiable; hence, for $k=0$, $m$ is only continuous on $\mathrm{Diff}^s(M)$. The inversion

$$i:\mathrm{Diff}^{s+k}(M)\to\mathrm{Diff}^s(M) \qquad [25]$$

is $C^k$ differentiable; hence, for $k=0$, $i$ is only continuous on $\mathrm{Diff}^s(M)$. The same differentiability properties of $m$ and $i$ hold in the $C^k$ topology. This situation leads to the notion of nested Lie groups.

The Lie algebra of $\mathrm{Diff}^\infty(M)$ is given by $\mathfrak{g}=T_e\mathrm{Diff}^\infty(M)\simeq\mathrm{Vec}^\infty(M)$, the space of smooth vector fields on $M$. Note that the space $\mathrm{Vec}(M)$ of all vector fields is a Lie algebra only for $C^\infty$ vector fields, but not for $C^k$ or $H^s$ vector fields if $k<\infty$, $s<\infty$, because one loses derivatives by taking brackets.

The exponential map on the diffeomorphism group is given as follows: for any vector field $X\in\mathrm{Vec}^\infty(M)$ take its flow $\varphi_t\in\mathrm{Diff}^\infty(M)$, then define $EXP:\mathrm{Vec}^\infty(M)\to\mathrm{Diff}^\infty(M):X\mapsto\varphi_1$, the flow at time $t=1$. The exponential map $EXP$ is not a local diffeomorphism; it is not even locally surjective.

Applications of $\mathrm{Diff}^\infty(M)$ occur in general relativity, where the diffeomorphism group plays the role of a symmetry group of coordinate transformations. Let $(M,g)$ be a Lorentz 4-manifold. Then the vacuum Einstein's field equations are

$$\mathrm{Ric}(g)=0$$

These are invariant under coordinate transformations, that is, under the action of $\mathrm{Diff}^\infty(M)$. Moreover, Einstein's field equations form a Hamiltonian system on the space $P=\{\text{metrics on } M\}/\mathrm{Diff}^\infty(M)$.

### Subgroups of $\mathrm{Diff}^\infty(M)$

Several subgroups of $\mathrm{Diff}^\infty(M)$ have important applications.

**Group of volume-preserving diffeomorphisms** Let $\mu$ be a volume on $M$ and $\mathcal{G}=\mathrm{Diff}^\infty_\mu(M)=\{f\in\mathrm{Diff}^\infty(M)\,|\,f^*\mu=\mu\}$ the group of volume-preserving diffeomorphisms. $\mathrm{Diff}^\infty_\mu(M)$ is a closed subgroup of $\mathrm{Diff}^\infty(M)$ with Lie algebra $\mathfrak{g}=\mathrm{Vec}^\infty_\mu(M)=\{X\in\mathrm{Vec}^\infty(M)\,|\,\mathrm{div}_\mu X=0\}$ the space of divergence free vector fields on $M$. $\mathrm{Vec}^\infty_\mu(M)$ is a Lie subalgebra of $\mathrm{Vec}^\infty(M)$.

Remark: We can neither apply the finite-dimensional theorem that if $\mathrm{Vec}^\infty_\mu(M)$ is Lie algebra then there exists a Lie group whose Lie algebra it is; nor that if $\mathrm{Diff}^\infty_\mu(M)\subset\mathrm{Diff}(M)$ is a closed subgroup then it is a Lie subgroup.

Applications of $\mathrm{Diff}^\infty_\mu(M)$ occur, for example, in fluid dynamics. Euler's equations for an incompressible fluid,

$$\frac{\partial u}{\partial t}+u\cdot\nabla u=-\nabla p,\qquad \mathrm{div}\,u=0 \qquad [26]$$

are equivalent to the equations of geodesics on $\mathrm{Diff}^\infty_\mu(M)$.

**Symplectomorphism group** Let $\omega$ be a symplectic 2-from on $M$ and $\mathcal{G}=\mathrm{Diff}^\infty_\omega(M)=\{f\in\mathrm{Diff}^\infty(M)\,|\,f^*\omega=\omega\}$ the group of canonical transformations (or symplectomorphisms). $\mathrm{Diff}^\infty_\omega(M)$ is a closed subgroup of $\mathrm{Diff}^\infty(M)$ with Lie algebra $\mathfrak{g}=\mathrm{Vec}^\infty_\omega(M)=\{X\in\mathrm{Vec}^\infty(M)\,|\,L_X\omega=0\}$ the space of locally Hamiltonian vector fields on $M$. $\mathrm{Vec}^\infty_\omega(M)$ is a Lie subalgebra of $\mathrm{Vec}^\infty(M)$.

Applications of symplectomorphism groups occur, for example, in plasma physics. Maxwell–Vlasov's

equations for a plasma density $f(x, v, t)$ generating the electric and magnetic fields $E$ and $B$ are

$$\frac{\partial f}{\partial t} + v \cdot \frac{\partial f}{\partial x} + (E + v \times B)\frac{\partial f}{\partial v} = 0$$
$$\frac{\partial B}{\partial t} = -\mathrm{curl}\, E, \qquad \frac{\partial E}{\partial t} = \mathrm{curl}\, B - J_f \qquad [27]$$
$$\mathrm{div}\, E = \rho_f, \qquad \mathrm{div}\, B = 0$$

where $J_f$ and $\rho_f$ are the current and charge densities, respectively. This coupled nonlinear system of evolution equations is an infinite-dimensional Hamiltonian system of the form $\dot{F} = \{F, H\}_{\rho_f}$ on the reduced phace space

$$\mathcal{MV} = (T^*\mathrm{Diff}^{\infty}_{\omega}(\mathbb{R}^6) \times T^*V)/C^{\infty}(\mathbb{R}^6) \qquad [28]$$

($V$ is the same space as in the example of Maxwell's equations) with respect to the following reduced Poisson bracket, which is induced via gauge symmetry from the canonical Poisson bracket on $T^*\mathrm{Diff}^{\infty}_{\omega}(\mathbb{R}^6) \times T^*V$:

$$\begin{aligned}\{F, G\}_{\rho_f}(f, E, B) &= \int f\left\{\frac{\delta F}{\delta f}, \frac{\delta G}{\delta f}\right\}\mathrm{d}x\,\mathrm{d}v \\ &+ \int\left(\frac{\delta F}{\delta E}\cdot \mathrm{curl}\frac{\delta G}{\delta B} - \frac{\delta G}{\delta E}\cdot \mathrm{curl}\frac{\delta F}{\delta B}\right)\mathrm{d}x\,\mathrm{d}v \\ &+ \int\left(\frac{\delta F}{\delta E}\cdot\frac{\partial f}{\partial v}\frac{\delta G}{\delta f} - \frac{\delta G}{\delta E}\cdot\frac{\partial f}{\partial v}\frac{\delta F}{\delta f}\right)\mathrm{d}x\,\mathrm{d}v \\ &+ \int fB\cdot\left(\frac{\partial}{\partial v}\frac{\delta F}{\delta f}\times\frac{\partial}{\partial v}\frac{\delta G}{\delta f}\right)\mathrm{d}x\,\mathrm{d}v \qquad [29]\end{aligned}$$

and with Hamiltonian

$$H(f, E, B) = \frac{1}{2}\int v^2 f(x, v, t)\mathrm{d}v + \frac{1}{2}\int(|E|^2 + |B|^2)\mathrm{d}x \qquad [30]$$

More complicated plasma models are formulated as Hamiltonian systems. For example, for the two-fluid model the phase space is constituted by coadjoint orbits of the semidirect product ($\ltimes$) of the group $\mathcal{G} = \mathrm{Diff}^{\infty}(\mathbb{R}^6) \ltimes (C^{\infty}(\mathbb{R}^6) \times C^{\infty}(\mathbb{R}^6))$. For the MHD model: $\mathcal{G} = \mathrm{Diff}^{\infty}(\mathbb{R}^6) \ltimes (C^{\infty}(\mathbb{R}^6) \times \Omega^2(\mathbb{R}^3))$.

## The KdV Equation and Fourier Integral Operators

There are many known examples of PDEs which are infinite-dimensional Hamiltonian systems, such as the Benjamin–Ono, Boussinesq, Harry Dym, KdV, and KP equations and others. In many cases, the Poisson structures and Hamiltonians are given *ad hoc* on a formal level. This is illustrated here with the KdV equation, where at least one of the three known Hamiltonian structures is well understood.

The KdV equation

$$u_t + 6uu_x + u_{xxx} = 0 \qquad [31]$$

is an infinite-dimensional Hamiltonian system with the Lie group of invertible Fourier integral operators being a symmetry group. Gardner found that with the bracket

$$\{F, G\} = \int_0^{2\pi}\frac{\delta F}{\delta u}\frac{\partial}{\partial x}\frac{\delta G}{\delta u}\mathrm{d}x \qquad [32]$$

and Hamiltonian

$$H(u) = \int_0^{2\pi}\left(u^3 + \tfrac{1}{2}u_x^3\right)\mathrm{d}x \qquad [33]$$

$u$ satisfies the KdV equation [31] if and only if

$$\dot{u} = \{u, H\}$$

An important question concerns the origin of the Poisson bracket [32] and Hamiltonian [33]. It was shown earlier that this bracket is the Lie–Poisson bracket on a coadjoint orbit of Lie group $\mathcal{G} = \mathrm{FIO}$, the group of invertible Fourier integral operators on the circle $S^1$. The latter is discussed briefly in the following.

A Fourier integral operator on a compact manifold $M$ is an operator

$$A : C^{\infty}(M) \to C^{\infty}(M) \qquad [34]$$

locally given by

$$A(u)(x) = (2\pi)^{-n}\iint \mathrm{e}^{\mathrm{i}\varphi(x,y,\xi)}a(x, \xi)u(y)\mathrm{d}y\,\mathrm{d}\xi \qquad [35]$$

where $\varphi(x, y, \xi)$ is a phase function with certain properties and the symbol $a(x, \xi)$ belongs to a certain symbol class. A pseudodifferential operator is a special kind of Fourier integral operators, locally of the form

$$P(u)(x) = (2\pi)^{-n}\iint \mathrm{e}^{\mathrm{i}(x-y)\cdot\xi}p(x, \xi)u(y)\mathrm{d}y\,\mathrm{d}\xi \qquad [36]$$

Denote by FIO and ΨDO the groups under composition (operator product) of invertible Fourier integral operators and invertible pseudodifferential operators on $M$, respectively. Then we have the following results.

Both groups ΨDO and FIO are smooth infinite-dimensional ILH Lie groups. The smoothness properties of the group operations (operator multiplication and inversion) are similar to the case of diffeomorphism groups [24] and [25]. The Lie algebra of both ILH Lie groups ΨDO and FIO is the Lie algebra of all pseudodifferential operators under the commutator bracket. Moreover, FIO is a smooth infinite-dimensional principal fiber bundle

over the diffeomorphism group of canonical transformations $\mathrm{Diff}^{\infty}_{\omega}(T^*M - \{0\})$ with structure group (gauge group) $\Psi$DO.

For the KdV equation, we take the special case where $M = S^1$. Then the Gardner bracket [32] is the Lie–Poisson bracket on the coadjoint orbit of FIO through the Schrödinger operator $P \in \Psi\mathrm{DO}$. Complete integrability of the KdV equation follows from the infinite system of conserved integrals in involution given by $H_k = \mathrm{tr}(P^k)$; in particular, the Hamiltonian [33] equals $H = H_2$.

*See also:* Bi-Hamiltonian Methods in Soliton Theory; Functional Integration in Quantum Physics; Hamiltonian Fluid Dynamics; Hamiltonian Systems: Obstructions to Integrability; Korteweg–de Vries Equation and Other Modulation Equations; Symmetries and Conservation Laws.

## Further Reading

Adams M, Ratiu TS, and Schmid R (1985) In: Kac V (ed.) *The Lie Group Structure of Diffeomorphism Groups and Invertible Fourier Integral Operators, with Applications*, MSRI Publications, vol. 4. New York: Springer.

Chernoff P and Marsden JE (1974) *Properties of Infinite Dimensional Hamiltonian Systems*, Lecture Notes in Mathematics, vol. 425. New York: Springer.

Marsden JE and Ratiu T (1994) *Introduction to Mechanics and Symmetry*. New York: Springer.

Marsden JE, Ebin GD, and Fischer A (1972) Diffeomorphism groups, hydrodynamics and relativity. In: Vanstone JR (ed.) *Proc. 13th Biennial Sem. Canadian Math. Congress*, pp. 135–279. Montreal.

Marsden JE, Weinstein A, Ratiu T, Schmid R, and Spencer RG (1983) Hamiltonian systems with symmetry, coadjoint orbits and plasma physics. *Atti Accad. Sci. Torino* 117(Suppl.): 289–340.

Olver PJ (1993) *Applications of Lie Groups to Differential Equations*. New York: Springer.

Palais R (1968) *Foundations of Global Nonlinear Analysis*. Reading, MA: Addison-Wesley.

Schmid R (1987) *Infinite Dimensional Hamiltonian Systems*. Lecture Notes, vol. 3. Naples: Bibliopolis.

Temam R (1988) *Infinite Dimensional Dynamical Systems in Mechanics and Physics*. New York: Springer.

# Instantons: Topological Aspects

**M Jardim**, IMECC–UNICAMP, Campinas, Brazil



## Introduction

Let $X$ be a closed (connected, compact without boundary) smooth manifold of dimension 4, provided with a Riemannian metric denoted by $g$. Let $\Omega^p_X$ denote space of smooth $p$-forms on $X$, that is, the sections of $\wedge^p TX$. The Hodge operator acting on $p$-forms,

$$* : \Omega^p_X \to \Omega^{4-p}_X$$

satisfies $*^2 = (-1)^p$. In particular, $*$ splits $\Omega^2_X$ into two subspaces $\Omega^{2,\pm}_X$ with eigenvalues $\pm 1$:

$$\Omega^2_X = \Omega^{2,+}_X \oplus \Omega^{2,-}_X \qquad [1]$$

Note also that this decomposition is an orthogonal one, with respect to the inner product:

$$\langle \omega_1, \omega_2 \rangle = \int_X \omega_1 \wedge *\omega_2$$

A 2-form $\omega$ is said to be self-dual if $*\omega = \omega$ and it is said to be anti-self-dual if $*\omega = -\omega$. Any 2-form $\omega$ can be written as the sum

$$\omega = \omega^+ + \omega^-$$

of its self-dual $\omega^+$ and anti-self-dual $\omega^-$ components.

Now let $E$ be a complex vector bundle over $X$ as above, provided with a connection $\nabla$, regarded as a $\mathbb{C}$-linear operator

$$\nabla : \Gamma(E) \to \Gamma(E) \otimes \Omega^1_X$$

satisfying the Leibnitz rule:

$$\nabla(f\sigma) = f\nabla\sigma + \sigma \otimes \mathrm{d}f$$

for all $f \in C^\infty(X)$ and $\sigma \in \Gamma(E)$. Its curvature $F_\nabla = \nabla \circ \nabla$ is a 2-form with values in $\mathrm{End}(E)$, that is, $F_\nabla \in \Gamma(\mathrm{End}(E)) \otimes \Omega^2_X$, satisfying the Bianchi identity $\nabla F_\nabla = 0$.

The Yang–Mills equation is

$$\nabla * F_\nabla = 0 \qquad [2]$$

It is a second-order nonlinear equation on the connection $\nabla$. It amounts to a nonabelian generalization of Maxwell equations, to which it reduces when $E$ is a line bundle; the four components of $\nabla$ are interpreted as the electric and magnetic potentials.

An instanton on $E$ is a smooth connection $\nabla$ whose curvature $F_\nabla$ is anti-self-dual as a 2-form, that is, it satisfies:

$$F^+_\nabla = 0, \quad \text{that is,} \quad * F_\nabla = -F_\nabla \qquad [3]$$

The instanton equation is still nonlinear (it is linear only if $E$ is a line bundle), but it is only first-order on the connection.

Note that if $F_\nabla$ is either self-dual or anti-self-dual as a 2-form, then the Yang–Mills equation is automatically satisfied:

$$*F_\nabla = \pm F_\nabla \Rightarrow \nabla * F_\nabla = \pm \nabla F_\nabla = 0$$

by the Bianchi identity. In other words, instantons are particular solutions of the Yang–Mills equation. Furthermore, while the Yang–Mills equation [2] makes sense over any Riemannian manifold, the instanton equation [3] is well defined only in dimension 4.

A gauge transformation is a bundle automorphism $g : E \to E$ covering the identity. The set of all gauge transformations of a given bundle $E \to X$ forms a group through composition, called the gauge group and denoted by $\mathcal{G}(E)$. The gauge group acts on the set of all smooth connections on $E$ by conjugation:

$$g \cdot \nabla = g^{-1} \nabla g$$

It is then easy to see that [3] is a gauge-invariant condition, since $F_{g\cdot\nabla} = g^{-1} F_\nabla g$. The anti-self-duality equation [3] is also conformally invariant: a conformal change in the metric does not change the decomposition [1], so it preserves self-dual and anti-self-dual 2-forms.

The topological charge $k$ of the instanton $\nabla$ is defined by the integral

$$k = -\frac{1}{8\pi^2} \int_X \operatorname{tr}(F_\nabla \wedge F_\nabla) = c_2(E) - \frac{1}{2} c_1(E)^2 \qquad [4]$$

where the second equality follows from Chern–Weil theory.

If $X$ is a smooth, noncompact, complete Riemannian manifold, an instanton on $X$ is an anti-self-dual connection for which the integral [4] converges. Note that, in this case, $k$ as above need not be an integer; however, it is always expected to be quantized, that is, always a multiple of some fixed (rational) number which depends only on the base manifold $X$.

**Summary** This note is organized as follows. After revisiting the variational approach to the anti-self-duality equation [3], we study instantons over the simplest possible Riemannian 4-manifold, $\mathbb{R}^4$ with the flat Euclidean metric. In the subsequent sections, we present 't Hooft's explicit solutions, the ADHM construction, and its dimensional reductions to $\mathbb{R}^3, \mathbb{R}^2$ and $\mathbb{R}$. We conclude by explaining the construction of the central object of study in gauge theory, the instanton moduli spaces.

## Variational Aspects of Yang–Mills Equation

Given a fixed smooth vector bundle $E \to X$, let $\mathcal{A}(E)$ be the set of all (smooth) connections on $E$. The Yang–Mills functional is defined by

$$\mathrm{YM} : \mathcal{A}(E) \to \mathbb{R}$$
$$\mathrm{YM}(\nabla) = \|F_\nabla\|^2_{L^2} = \int_M \operatorname{tr}(F_\nabla \wedge *F_\nabla) \qquad [5]$$

The Euler–Lagrange equation for this functional is exactly the Yang–Mills equation [2]. In particular, self-dual and anti-self-dual connections yield critical points of the Yang–Mills functional.

Splitting the curvature into its self-dual and anti-self-dual parts, we have

$$\mathrm{YM}(\nabla) = \|F^+_\nabla\|^2_{L^2} + \|F^-_\nabla\|^2_{L^2}$$

It is then easy to see that every anti-self-dual connection $\nabla$ is an absolute minimum for the Yang–Mills functional, and that $\mathrm{YM}(\nabla)$ coincides with the topological charge [4] of the instanton $\nabla$ times $8\pi^2$.

One can construct, for various 4-manifolds but most interestingly for $X = S^4$, solutions of the Yang–Mills equations which are neither self-dual nor anti-self-dual. Such solutions do not minimize [5]. Indeed, at least for gauge group SU(2) or SU(3), it can be shown that there are no other local minima: any critical point which is neither self-dual nor anti-self-dual is unstable and must be a "saddle point" (Bourguignon and Lawson Jr. 1981).

## Instantons on Euclidean Space

Let $X = \mathbb{R}^4$ with the flat Euclidean metric, and consider a Hermitian vector bundle $E \to \mathbb{R}^4$. Any connection $\nabla$ on $E$ is of the form $d + A$, where $d$ denotes the usual de Rham operator and $A \in \Gamma(\mathrm{End}(E)) \otimes \Omega^1_{\mathbb{R}^4}$ is a 1-form with values in the endomorphisms of $E$; this can be written as follows:

$$A = \sum_{k=1}^{4} A_k \, \mathrm{d}x^k, \quad A_k : \mathbb{R}^4 \to \mathfrak{u}(r)$$

In the Euclidean coordinates $x_1, x_2, x_3, x_4$, the anti-self-duality equation [3] is given by

$$F_{12} = F_{34}, \quad F_{13} = -F_{24}, \quad F_{14} = F_{23}$$

where

$$F_{ij} = \frac{\partial A_j}{\partial x_i} - \frac{\partial A_i}{\partial x_j} + [A_i, A_j]$$

The simplest explicit solution is the charge-1 SU(2) instanton on $\mathbb{R}^4$. The connection 1-form is given by

$$A_0 = \frac{1}{1+|x|^2} \cdot \mathrm{Im}(q\,\mathrm{d}\bar{q}) \qquad [6]$$

where $q$ is the quaternion $q = x_1 + x_2\boldsymbol{i} + x_3\boldsymbol{j} + x_4\boldsymbol{k}$, while Im denotes the imaginary part of the product quaternion; we are regarding $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ as a basis of the Lie algebra $\mathfrak{su}(2)$; from this, one can compute the curvature:

$$F_{A_0} = \left(\frac{1}{1+|x|^2}\right)^2 \cdot \mathrm{Im}(\mathrm{d}q \wedge \mathrm{d}\bar{q}) \qquad [7]$$

We see that the action density function

$$|F_{A_0}|^2 = \left(\frac{1}{1+|x|^2}\right)^2$$

has a bell-shaped profile centered at the origin and decays like $r^{-4}$.

Let $t_{\lambda,y} : \mathbb{R}^4 \to \mathbb{R}^4$ be the isometry given by the composition of a translation by $y \in \mathbb{R}^4$ with a homothety by $\lambda \in \mathbb{R}^+$. The pullback connection $t^*_{\lambda,y}A_0$ is still anti-self-dual; more explicitly,

$$A_{\lambda,y} = t^*_{\lambda,y}A_0 = \frac{\lambda^2}{\lambda^2 + |x-y|^2} \cdot \mathrm{Im}(q\,\mathrm{d}\bar{q})$$

$$F_{A_{\lambda,y}} = \left(\frac{\lambda^2}{\lambda^2 + |x-y|^2}\right)^2 \cdot \mathrm{Im}(\mathrm{d}q \wedge \mathrm{d}\bar{q})$$

Note that the action density function $|F_A|^2$ has again a bell-shaped profile centered at $y$ and decays like $r^{-4}$; the parameter $\lambda$ measures the concentration of the energy density function, and can be interpreted as the "size" of the instanton $A_{\lambda,y}$.

Instantons of topological charge $k$ can be obtained by "superimposing" $k$ basic instantons, via the so-called 't Hooft ansatz. Consider the function $\rho : \mathbb{R}^4 \to \mathbb{R}$ given by

$$\rho(x) = 1 + \sum_{j=1}^{k} \frac{\lambda_j^2}{(x-y_j)^2}$$

where $\lambda_j \in \mathbb{R}$ and $y_j \in \mathbb{R}^4$. Then the connection 1-form $A = A_\mu \mathrm{d}x_\mu$ with coefficients

$$A_\mu = i \sum_{\nu=1}^{4} \bar{\sigma}_{\mu\nu} \frac{\partial}{\partial x_\nu} \ln(\rho(x)) \qquad [8]$$

is anti-self dual; here, $\bar{\sigma}_{\mu\nu}$ are the matrices given by ($\mu, \nu = 1, 2, 3$):

$$\sigma_{\mu\nu} = \frac{1}{4\mathrm{i}}[\sigma_\mu, \sigma_\nu] \qquad \bar{\sigma}_{\mu 4} = \frac{1}{2}\sigma_\mu$$

where $\sigma_\mu$ are the Pauli matrices.

The connection [8] correspond to $k$ instantons centered at points $y_i$ with size $\lambda_i$. The basic instanton [6] is exactly (modulo gauge transformation) what one obtains from [8] for the case $k = 1$. The 't Hooft instantons form a $5k$-parameter family of anti-self-dual connections.

SU(2) instantons are also the building blocks for instantons with general structure group (Bernard *et al.* 1977). Let $G$ be a compact semisimple Lie group, with Lie algebra $\mathfrak{g}$. Let $\phi : \mathfrak{su}(2) \to \mathfrak{g}$ be any injective Lie algebra homomorphism. If $A$ is an anti-self-dual SU(2) connection 1-form, then it is easy to see that $\phi(A)$ is an anti-self-dual $G$-connection 1-form. Using [8] as an example, we have that

$$A = \mathrm{i} \sum_{\mu,\nu} \phi(\bar{\sigma}_{\mu\nu}) \frac{\partial}{\partial x_\nu} \ln(\rho(x)) \mathrm{d}x_\mu \qquad [9]$$

is a $G$-instanton on $\mathbb{R}^4$.

While this guarantees the existence of $G$-instantons on $\mathbb{R}^4$, note that the instanton [9] might be reducible (e.g., $\phi$ can simply be the obvious inclusion of $\mathfrak{su}(2)$ into $\mathfrak{su}(n)$ for any $n$) and that its charge depends on the choice of representation $\phi$. Furthermore, it is not clear whether every $G$-instanton can be obtained in this way, as the inclusion of a SU(2) instanton through some representation $\phi : \mathfrak{su}(2) \to \mathfrak{g}$.

## The ADHM Construction

All SU($r$) instantons on $\mathbb{R}^4$ can be obtained through a remarkable construction due to Atiyah, Drinfeld, Hitchin, and Manin. It starts by considering Hermitian vector spaces $V$ and $W$ of dimension $c$ and $r$, respectively, and the following data (the so-called ADHM data):

$$B_1, B_2 \in \mathrm{End}(V), \;\; i \in \mathrm{Hom}(W, V)$$
$$j \in \mathrm{Hom}(V, W)$$

Assume, moreover, that $(B_1, B_2, i, j)$ satisfy the ADHM equations:

$$[B_1, B_2] + ij = 0 \qquad [10]$$

$$[B_1, B_1^\dagger] + [B_2, B_2^\dagger] + ii^\dagger - j^\dagger j = 0 \qquad [11]$$

Now consider the following maps

$$\alpha : V \times \mathbb{R}^4 \to (V \oplus V \oplus W) \times \mathbb{R}^4$$
$$\beta : (V \oplus V \oplus W) \times \mathbb{R}^4 \to V \times \mathbb{R}^4$$

given as follows (**1** denotes the appropriate identity matrix):

$$\alpha(z_1, z_2) = \begin{pmatrix} B_1 + z_1\mathbf{1} \\ B_2 + z_2\mathbf{1} \\ j \end{pmatrix} \qquad [12]$$

$$\beta(z_1, z_2) = (-B_2 - z_2\mathbf{1} \quad B_1 + z_1\mathbf{1} \quad \mathrm{i}) \qquad [13]$$

where $z_1 = x_1 + \mathrm{i}x_2$ and $z_2 = x_3 + \mathrm{i}x_4$ are complex coordinates on $\mathbb{R}^4$. The maps [12] and [13] should be understood as a family of linear maps parametrized by points in $\mathbb{R}^4$.

A straightforward calculation shows that the ADHM equation [10] implies that $\beta\alpha = 0$ for every $(z_1, z_2) \in \mathbb{R}^4$. Therefore, the quotient $E = \ker\beta/\mathrm{im}\,\alpha = \ker\beta \cap \ker\alpha^\dagger$ forms a complex vector bundle over $\mathbb{R}^4$ or rank $r$ whenever $(B_1, B_2, i, j)$ is such that $\alpha$ is injective and $\beta$ is surjective for every $(z_1, z_2) \in \mathbb{R}^4$.

To define a connection on $E$, note that $E$ can be regarded as a sub-bundle of the trivial bundle $(V \oplus V \oplus W) \times \mathbb{R}^4$. So let $\iota : E \to (V \oplus V \oplus W) \times \mathbb{R}^4$ be the inclusion, and let $P : (V \oplus V \oplus W) \times \mathbb{R}^4 \to E$ be the orthogonal projection onto $E$. We can then define a connection $\nabla$ on $E$ through the projection formula

$$\nabla s = P\underline{d}\iota(s)$$

where $\underline{d}$ denotes the trivial connection on the trivial bundle $(V \oplus V \oplus W) \times \mathbb{R}^4$.

To see that this connection is anti-self-dual, note that projection $P$ can be written as follows:

$$P = \mathbf{1} - \mathcal{D}^\dagger\Xi^{-1}\mathcal{D}$$

where

$$\mathcal{D} : (V \oplus V \oplus W) \times \mathbb{R}^4 \to (V \oplus V) \times \mathbb{R}^4$$
$$\mathcal{D} = \begin{pmatrix} \beta \\ \alpha^\dagger \end{pmatrix}$$

and $\Xi = \mathcal{D}\mathcal{D}^\dagger$. Note that $\mathcal{D}$ is surjective, so that $\Xi$ is indeed invertible. Moreover, it also follows from [11] that $\beta\beta^\dagger = \alpha^\dagger\alpha$, so that $\Xi^{-1} = (\beta\beta^\dagger)^{-1}\mathbf{1}$.

The curvature $F_\nabla$ is given by

$$\begin{aligned} F_\nabla &= P\Big(\underline{d}(\mathbf{1} - \mathcal{D}_\mathrm{I}^\dagger\Xi^{-1}\mathcal{D})\underline{d}\Big) = P\big(\underline{d}\mathcal{D}^\dagger\Xi^{-1}(\underline{d}\mathcal{D})\big) \\ &= P\big((\underline{d}\mathcal{D}^\dagger)\Xi^{-1}(\underline{d}\mathcal{D}) + \mathcal{D}^\dagger\underline{d}(\Xi^{-1}(\underline{d}\mathcal{D})\big) \\ &= (\underline{d}\mathcal{D}^\dagger)\Xi^{-1}(\underline{d}\mathcal{D}) \end{aligned}$$

for $P(\mathcal{D}^\dagger\underline{d}(\Xi^{-1}(\underline{d}\mathcal{D}))) = 0$ on $E = \ker\mathcal{D}$. Since $\Xi^{-1}$ is diagonal, we conclude that $F_\nabla$ is proportional to $d\mathcal{D}^\dagger \wedge d\mathcal{D}$, as a 2-form.

It is then a straightforward calculation to show that each entry of $d\mathcal{D}^\dagger \wedge d\mathcal{D}$ belongs to $\Omega^{2,-}$.

The extraordinary accomplishment of Atiyah, Drinfeld, Hitchin, and Manin was to show that every instanton, up to gauge equivalence, can be obtained in this way (see, e.g., Donaldson and Kronheimer 1990). For instance, the basic SU(2) instanton [6] is associated with the following data ($c = 1, r = 2$):

$$B_1, B_2 = 0, \quad i = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad j = (0 \quad 1)$$

**Remark** The ADHM data $(B_1, B_2, i, j)$ are said to be stable if $\beta$ is surjective for every $(z_1, z_2) \in \mathbb{R}^4$, and it is said to be costable if $\alpha$ is injective for every $(z_1, z_2) \in \mathbb{R}^4$. $(B_1, B_2, i, j)$ is regular if it is both stable and costable. The quotient:

$$\{\text{regular solutions of } (10) \text{ and } (11)\}/U(V)$$

coincides with the moduli space of instantons of rank $r = \dim W$ and charge $c = \dim V$ on $\mathbb{R}^4$ (see below). It is also an example of a quiver variety (*see* Finite Dimensional Algebras and Quivers), associated to the quiver consisting of two vertices $V$ and $W$, two loop-edges on the vertex $V$ and two edges linking $V$ to $W$, one in each direction.

## Dimensional Reductions of the Anti-Self-Dual Yang–Mills Equation

As pointed out above, a connection on a Hermitian vector bundle $E \to \mathbb{R}^4$ of rank $r$ can be regarded as 1-form

$$A = \sum_{k=1}^{4} A_k(x_1, \ldots, x_4)\mathrm{d}x^k, \quad A_k : \mathbb{R}^4 \to \mathfrak{u}(r)$$

Assuming that the connection components $A_k$ are invariant under translation in one direction, say $x_4$, we can think of

$$\underline{A} = \sum_{k=1}^{3} A_k(x_1, x_2, x_3)\mathrm{d}x^k$$

as a connection on a Hermitian vector bundle over $\mathbb{R}^3$, with the fourth component $\phi = A_4$ being regarded as a bundle endomorphism $\phi : E \to E$, called a Higgs field. In this way, the anti-self-duality equation [3] reduces to the so-called Bogomolny (or monopole) equation:

$$F_{\underline{A}} = *\mathrm{d}\phi \qquad [14]$$

where $*$ is the Euclidean Hodge star in dimension 3.

Now assume that the connection components $A_k$ are invariant under translation in two directions, say $x_3$ and $x_4$. Consider

$$\underline{A} = \sum_{k=1}^{2} A_k(x_1, x_2)\mathrm{d}x^k$$

as a connection on a Hermitian vector bundle over $\mathbb{R}^2$, with the third and fourth components combined into a complex bundle endomorphism:

$$\Phi = (A_3 + \mathrm{i} \cdot A_4)(\mathrm{d}x_1 - \mathrm{i} \cdot \mathrm{d}x_2)$$

taking values on 1-forms. The anti-self-duality equation [3] is then reduced to the so-called Hitchin's equations:

$$F_{\underline{A}} = [\Phi, \Phi^*], \qquad \bar{\partial}_{\underline{A}}\Phi = 0 \qquad [15]$$

Conformal invariance of the anti-self-duality equation means that Hitchin's equations are well defined over any Riemann surface.

Finally, assume that the connection components $A_k$ are invariant under translation in three directions, say $x_2, x_3$, and $x_4$. After gauging away the first component $A_1$, the anti-self-duality equations [3] reduce to the so-called Nahm's equations:

$$\frac{\mathrm{d}T_k}{\mathrm{d}x_1} + \frac{1}{2}\sum_{j,l} \epsilon_{kjl}[T_j, T_l] = 0, \quad j, k, l = \{2, 3, 4\} \quad [16]$$

where each $T_k$ is regarded as a map $\mathbb{R} \to \mathfrak{u}(r)$.

Readers who are interested in monopoles and Nahm's equations are referred to the survey by Murray (2002) and references therein. The best source for Hitchin's equations still are Hitchin's (1987a, b) original papers. A beautiful duality, known as Nahm transform, relates the various reductions of the anti-self-duality equation to periodic instantons; see the survey article by Jardim (2004).

It is also worth mentioning the book by Mason and Woodhouse (1996), where other interesting dimensional reductions of the anti-self-duality equation are discussed, providing a deep relation between instantons and the general theory of integrable systems.

## The Instanton Moduli Space

Now fix a rank-$r$ complex vector bundle $E$ over a four-dimensional Riemannian manifold $X$. Observe that the difference between any two connections is a linear operator:

$$(\nabla - \nabla')(f\sigma) = f\nabla\sigma + \sigma \cdot \mathrm{d}f - f\nabla'\sigma - \sigma \cdot \mathrm{d}f$$
$$= f(\nabla - \nabla')\sigma$$

In other words, any two connections on $E$ differ by an endomorphism-valued 1-form. Therefore, the set of all smooth connections on $E$, denoted by $\mathcal{A}(E)$, has the structure of an affine space over $\Gamma(\mathrm{End}(E)) \otimes \Omega^1_M$.

The gauge group $\mathcal{G}(E)$ acts on $\mathcal{A}(E)$ via conjugation:

$$g \cdot \nabla := g^{-1}\nabla g$$

We can form the quotient set $\mathcal{B}(E) = \mathcal{A}(E)/\mathcal{G}(E)$, which is the set of gauge equivalence classes of connections on $E$.

The set of gauge equivalence classes of anti-self-dual connections on $E$ is a subset of $\mathcal{B}(E)$, and it is called the moduli space of instantons on $E \to X$. The subset of $\mathcal{M}_X(E)$ consisting of irreducible anti-self-dual connections is denoted $\mathcal{M}^*_X(E)$.

Since the choice of a particular vector bundle within its topological class is immaterial, these sets are usually labeled by the topological invariants (Chern or Pontrjagyn classes) of the bundle $E$. For instance, $\mathcal{M}(r, k)$ denotes the moduli space of instantons on a rank-$r$ complex vector bundle $E \to X$ with $c_1(E) = 0$ and $c_2(E) = k > 0$. It turns out that $\mathcal{M}_X(E)$ can be given the structure of a Hausdorff topological space. In general, $\mathcal{M}_X(E)$ will be singular as a differentiable manifold, but $\mathcal{M}^*_X(E)$ can always be given the structure of a smooth Riemannian manifold.

We start by explaining the notion of a $L^2_p$ vector bundle. Recall that $L^2_p(\mathbb{R}^n)$ denotes the completion of the space of smooth functions $f : \mathbb{R}^n \to \mathbb{C}$ with respect to the norm:

$$\|f\|^2_{L^2_p} = \int_X (|f|^2 + |\mathrm{d}f|^2 + \cdots + |\mathrm{d}^{(p)}f|^2)$$

In dimension $n = 4$ and for $p > 2$, by virtue of the Sobolev embedding theorem, $L^2_p$ consists of continuous functions, i.e., $L^2_p(\mathbb{R}^n) \subset C^0(\mathbb{R}^n)$. So we define the notion of a $L^2_p$ vector bundle as a topological vector bundle whose transition functions are in $L^2_p$, where $p > 2$.

Now for a fixed $L^2_p$ vector bundle $E$ over $X$, we can consider the metric space $\mathcal{A}_p(E)$ of all connections on $E$ which can be represented locally on an open subset $U \subset X$ as a $L^2_p(U)$ 1-form. In this topology, the subset of irreducible connections $\mathcal{A}^*_p(E)$ becomes an open dense subset of $\mathcal{A}_p(E)$. Since any topological vector bundle admits a compatible smooth structure, we may regard $L^2_p$ connections as those that differ from a smooth connection by a $L^2_p$ 1-form. In other words, $\mathcal{A}_p(E)$ becomes an affine space modeled over the Hilbert space of $L^2_p$ 1-forms with values in the endomorphisms of $E$. The curvature of a connection in $\mathcal{A}_p(E)$ then becomes a $L^2_{p-1}$ 2-form with values in the endomorphism bundle $\mathrm{End}(E)$.

Moreover, let $\mathcal{G}_{p+1}(E)$ be defined as the topological group of all $L^2_{p+1}$ bundle automorphisms. By virtue of the Sobolev multiplication theorem, $\mathcal{G}_{p+1}(E)$ has the structure of an infinite-dimensional

Lie group modeled on a Hilbert space; its Lie algebra is the space of $L^2_{p+1}$ sections of End($E$).

The Sobolev multiplication theorem is once again invoked to guarantee that the action $\mathcal{G}_{p+1}(E) \times \mathcal{A}_p(E) \to \mathcal{A}_p(E)$ is a smooth map of Hilbert manifolds. The quotient space $\mathcal{B}_p(E) = \mathcal{A}_p(E)/\mathcal{G}_{p+1}(E)$ inherits a topological structure; it is a metric (hence Hausdorff) topological space. Therefore, the subspace $\mathcal{M}_X(E)$ of $\mathcal{B}_p(E)$ is also a Hausdorff topological space; moreover, one can show that the topology of $\mathcal{M}_X(E)$ does not depend on $p$.

The quotient space $\mathcal{B}_p(E)$ fails to be a Hilbert manifold because in general the action of $\mathcal{G}_{p+1}(E)$ on $\mathcal{A}_p(E)$ is not free. Indeed, if $A$ is a connection on a rank-$r$ complex vector bundle $E$ over a connected base manifold $X$, which is associated with a principal $G$-bundle. Then the isotropy group of $A$ within the gauge group

$$\Gamma_A = \{g \in \mathcal{G}_{p+1}(E) | g(A) = A\}$$

is isomorphic to the centralizer of the holonomy group of $A$ within $G$.

This means that the subspace of irreducible connections $\mathcal{A}^*_p(E)$ can be equivalently defined as the open dense subset of $\mathcal{A}_p(E)$ consisting of those connections whose isotropy group is minimal, that is,

$$\mathcal{A}^*_p(E) = \{A \in \mathcal{A}_p(E) | \Gamma_A = \text{center}(G)\}$$

Now $\mathcal{G}_{p+1}(E)$ acts with constant isotropy on $\mathcal{A}^*_p(E)$; hence, the quotient $\mathcal{B}^*_p(E) = \mathcal{A}^*_p(E)/\mathcal{G}_{p+1}(E)$ acquires the structure of a smooth Hilbert manifold.

**Remark** The analysis of neighborhoods of points in $\mathcal{B}_p(E) \backslash \mathcal{B}^*_p(E)$ is very relevant for applications of the instanton moduli spaces to differential topology. The simplest situation occurs when $A$ is an SU(2) connection on a rank-2 complex vector bundle $E$ which reduces to a pair of U(1) and such $[A]$ occurs as an isolated point in $\mathcal{B}_p(E) \backslash \mathcal{B}^*_p(E)$. Then a neighborhood of $[A]$ in $\mathcal{B}_p(E)$ looks like a cone on an infinite-dimensional complex projective space.

Alternatively, the instanton moduli space $\mathcal{M}_X(E)$ can also be described by first taking the subset of all anti-self-dual connections and then taking the quotient under the action of the gauge group. More precisely, consider the map

$$\begin{aligned} \rho : \mathcal{A}_p(E) &\to L^2_p(\text{End}(E) \otimes \Omega^{2,+}_X) \\ \rho(A) &= F^+_A \end{aligned} \qquad [17]$$

Thus, $\rho^{-1}(0)$ is exactly the set of all anti-self-dual connections. It is $\mathcal{G}_{p+1}(E)$-invariant, so we can take the quotient to get

$$\mathcal{M}_X(E) = \rho^{-1}(0)/\mathcal{G}_{p+1}(E)$$

It follows that the subspace $\mathcal{M}^*_X(E) = \mathcal{B}^*_p(E) \cap \mathcal{M}_X(E)$ has the structure of a smooth Hilbert manifold. Index theory comes into play to show that $\mathcal{M}^*_X(E)$ is finite dimensional. Recall that if $D$ is an elliptic operator on a vector bundle over a compact manifold, then $D$ is Fredholm (i.e., ker $D$ and coker $D$ are finite dimensional) and its index

$$\text{ind}\, D = \dim \ker D - \dim \text{coker}\, D$$

can be computed in terms of topological invariants, as prescribed by the Atiyah–Singer index theorem. The goal here is to identify the tangent space of $\mathcal{M}^*_X(E)$ with the kernel of an elliptic operator.

It is clear that, for each $A \in \mathcal{A}_p(E)$, the tangent space $T_A\mathcal{A}_p(E)$ is just $L^2_p(\text{End}(E) \otimes \Omega^1_X)$. We define the pairing

$$\langle a, b \rangle = \int_X a \wedge *b \qquad [18]$$

and it is easy to see that this pairing defines a Riemannian metric (the so-called $L^2$-metric) on $\mathcal{A}_p(E)$.

The derivative of the map $\rho$ in [17] at the point $A$ is given by

$$\begin{aligned} d^+_A : L^2_p(\text{End}(E) \otimes \Omega^1_X) &\to L^2_{p-1}(\text{End}(E) \otimes \Omega^2_X) \\ a &\mapsto (d_A a)^+ \end{aligned}$$

so that for each $A \in \rho^{-1}(0)$ we have

$$T_A\rho^{-1}(0) = \left\{ a \in L^2_p(\text{End}(E)) \otimes \Omega^1_X \,|\, d^+_A a = 0 \right\}$$

Now for a gauge equivalence class $[A] \in \mathcal{B}^*_p(E)$, the tangent space $T_{[A]}\mathcal{B}^*_p(E)$ consists of those 1-forms which are orthogonal to the fibers of the principal $\mathcal{G}_{p+1}(E)$ bundle $\mathcal{A}^*_p(E) \to \mathcal{B}^*_p(E)$. At a point $A \in \mathcal{A}_p(E)$, the derivative of the action by some $g \in \mathcal{G}_{p+1}(E)$ is

$$-d_A \,:\, L^2_{p+1}(\text{End}(E)) \to L^2_p(\text{End}(E) \otimes \Omega^1_X)$$

Usual Hodge decomposition gives us that there is an orthogonal decomposition:

$$L^2_p(\text{End}(E) \otimes \Omega^1_X) = \text{im}\, d_A \oplus \ker d^*_A$$

which means that:

$$T_{[A]}\mathcal{B}^*_p(E) = \left\{ a \in L^2_p(\text{End}(E) \otimes \Omega^1_X) \,|\, d^*_A a = 0 \right\}$$

Thus, the pairing [18] also defines a Riemannian metric on $\mathcal{B}^*_p(E)$. Putting these together, we conclude that the space $T_{[A]}\mathcal{M}^*_X$ tangent to $\mathcal{M}^*_X(E)$ at an equivalence class $[A]$ of anti-self-dual connections can be described as follows:

$$\begin{aligned} &T_{[A]}\mathcal{M}^*_X(E) \\ &\quad = \left\{ a \in L^2_p(\text{End}(E) \otimes \Omega^1_X) \,|\, d^*_A a = d^+_A a = 0 \right\} \end{aligned} \qquad [19]$$

It turns out that the so-called deformation operator $\delta_A = d_A^* \oplus d_A$:

$$\delta_A : L_p^2(\mathrm{End}(E) \otimes \Omega_X^1) \\ \rightarrow L_{p+1}^2(\mathrm{End}(E)) \oplus L_{p-1}^2(\mathrm{End}(E) \otimes \Omega_X^2)$$

is elliptic. Moreover, if $A$ is anti-self-dual then coker $\delta_A$ is empty, so that $T_{[A]}\mathcal{M}_X^*(E) = \ker \delta_A$. The dimension of the tangent space $T_{[A]}\mathcal{M}_X^*(E)$ is then simply given by the index of the deformation operator $\delta_A$. Using the Atiyah–Singer index theorem, we have for SU($r$) bundles with $c_2(E) = k$:

$$\dim \mathcal{M}_X^*(E) = 4rk - (r^2 - 1)(1 - b_1(X) + b_+(X))$$

The dimension formula for arbitrary gauge group $G$ can be found in Atiyah *et al.* (1978).

For example, the moduli space of SU(2) instantons on $\mathbb{R}^4$ of charge $k$ is a smooth Riemannian manifold of dimension $8k - 3$. These parameters are interpreted as the $5k$ parameters describing the positions and sizes of $k$ separate instantons, plus $3(k - 1)$ parameters describing their relative SU(2) phases.

The detailed construction of the instanton moduli spaces can be found in Donaldson and Kronheimer (1990). An alternative source is Morgan's lecture notes (Friedman and Morgan 1998). It is interesting to note that $\mathcal{M}_X^*(E)$ inherits many of the geometrical properties of the original manifold $X$. Most notably, if $X$ is a Kähler manifold, then $\mathcal{M}_X^*(E)$ is also Kähler; if $X$ is a hyper-Kähler manifold, then $\mathcal{M}_X^*(E)$ is also hyper-Kähler. One expects that other geometric structures on $X$ can also be transferred to the instanton moduli spaces $\mathcal{M}_X^*(E)$.

*See also:* Characteristic Classes; Finite-Dimensional Algebras and Quivers; Gauge Theoretic Invariants of 4-Manifolds; Gauge Theory: Mathematical Applications; Integrable Systems: Overview; Index Theorems; Moduli Spaces: An Introduction; Solitons and Other Extended Field Configurations; Twistor Theory: Some Applications [in Integrable Systems, Complex Geometry and String Theory].

## Further Reading

Atiyah MF, Hitchin NJ, and Singer IM (1978) Self-duality in four-dimensional Riemannian geometry. *Proceedings of the Royal Society of London* 362: 425–461.

Bernard CW, Christ NH, Guth AH, and Weinberg EJ (1977) Pseudoparticle parameters for arbitrary gauge groups. *Physical Review D* 16: 2967–2977.

Bourguignon JP and Lawson HB Jr. (1981) Stability and isolation phenomena for Yang–Mills fields. *Communications in Mathematical Physics* 79: 189–230.

Donaldson SK and Kronheimer PB (1990) *Geometry of Four-Manifolds*. Oxford: Clarendon.

Friedman R and Morgan JW (eds.) (1998) *Gauge Theory and the Topology of Four-Manifolds*. Providence, RI: American Mathematical Society.

Hitchin N (1987a) The self-duality equations on a Riemann surface. *Proceedings of the London Mathematical Society* 55: 59–126.

Hitchin N (1987b) Stable bundles and integrable systems. *Duke Mathematical Journal* 54: 91–114.

Jardim M (2004) A survey on Nahm transform. *Journal of Geometry and Physics* 52: 313–327.

Mason LJ and Woodhouse NMJ (1996) *Integrability, Self-duality, and Twistor Theory*. New York, NY: Clarendon.

Murray M (2002) Monopoles. In: Bouwknegt P and Wu S (eds.) *Geometric Analysis and Applications to Quantum Field Theory*, Progr. Math. vol. 205, pp. 119–135. Boston, MA: Birkhauser.

# Integrability and Quantum Field Theory

**T J Hollowood**, University of Wales Swansea, Swansea, UK



## Introduction

The notion of integrability plays many different rôles in quantum field theory (QFT). In this article we interpret it in a narrow sense and describe some QFTs that are completely integrable, in the sense that there are as many integrals of motion as degrees of freedom. Necessarily this implies, since we are talking about field theories, that there is an infinite number of conserved quantities. The existence of such a tower of conserved quantities of increasing Lorentz spin implies, via the Coleman–Mandula theorem, that the theories are trivial in spacetime dimensions greater than 2. On the other hand, in $1 + 1$ dimensions there is a rich menagerie of such integrable quantum field theories (IQFTs). These theories are fascinating in their own right as nontrivial QFTs for which data like the $S$-matrix and spectrum can be determined exactly. We will describe these exact $S$-matrices for a series of seminal examples. In addition, we briefly describe the applications of these theories to statistical systems in two dimensions.

## Classical Integrable Systems and Field Theories

For a field theory to be integrable it must have an infinite number of conserved charges. Necessarily these must be spacetime symmetries which extend the Poincaré symmetry in some way. It turns out that, due to a theorem of Coleman and Mandula, such

extensions are very restrictive: they are only possible in $1+1$ dimensions (one dimension of space and one of time) apart from noninteracting theories. Below we describe some of the most important examples.

### Affine Toda Theories

These theories describe the interactions of a set of scalar fields which we write as a vector $\boldsymbol{\phi}$. The action is

$$S = \int d^2x \left( \frac{1}{2} (\partial_\mu \boldsymbol{\phi})^2 - V(\boldsymbol{\phi}) \right) \qquad [1]$$

The potential has to be very specially chosen in order that the resulting theory is integrable. The resulting theories are classified by affine Lie algebras. We shall describe only the theories related to a simply laced Lie algebra $\mathfrak{g}$ (so of ADE type). In this case, for the affine version of the theory,

$$V(\boldsymbol{\phi}) = \frac{m^2}{\beta^2} \sum_{a=0}^{r} n_a \, e^{\beta \boldsymbol{\alpha}_a \cdot \boldsymbol{\phi}} \qquad [2]$$

where $\boldsymbol{\phi}$ is an $r$-rank $\mathfrak{g}$ vector and $\boldsymbol{\alpha}_a, a=1,\ldots,r$, are a set of simple roots of $\mathfrak{g}$. The fact that we are considering the affine version of the theory means that we include the term involving the extended root (the lowest root) $\boldsymbol{\alpha}_0 = -\sum_{a=1}^{r} n_a \boldsymbol{\alpha}_a$, which defines the integers $n_a (n_0 = 1)$. If this term is absent then the potential does not have a minimum. Such nonaffine theories are interesting in their own right since they include the Liouville theory, but we shall not describe them here.

One way to expose the infinite set of conserved charges at the classical level is to write the equations of motion in Lax form. This has the form of the vanishing of the field strength, or zero-curvature condition, of an auxiliary gauge connection in $\mathfrak{g}$ with components $(A_x, A_t)$:

$$\begin{aligned} A_x &= \partial_t \boldsymbol{\phi} \cdot \boldsymbol{H} + \frac{m}{2\beta} \sum_{a=0}^{r} e^{\beta \boldsymbol{\alpha}_a \cdot \boldsymbol{\phi}/2} (e_a + f_a) \\ A_t &= \partial_x \boldsymbol{\phi} \cdot \boldsymbol{H} + \frac{m}{2\beta} \sum_{a=0}^{r} e^{\beta \boldsymbol{\alpha}_a \cdot \boldsymbol{\phi}/2} (e_a - f_a) \end{aligned} \qquad [3]$$

Here, $\{e_i, f_i\}$ are related to generators of $\mathfrak{g}$ in a Cartan–Weyl basis, via

$$\begin{aligned} e_a &= z E_{\boldsymbol{\alpha}_a}, \quad f_a = z^{-1} E_{-\boldsymbol{\alpha}_a}, \qquad a = 1, \ldots, r \\ e_0 &= z^{-h} E_{\boldsymbol{\alpha}_0}, \qquad f_0 = z^h E_{-\boldsymbol{\alpha}_0} \end{aligned} \qquad [4]$$

where $z$ is a auxiliary variable known as the spectral parameter and $h$ is the Coxeter number of $\mathfrak{g}$. Using the following commutators of $\mathfrak{g}$,

$$\begin{aligned} [E_{\boldsymbol{\alpha}_a}, E_{\boldsymbol{\alpha}_b}] &= \delta_{ab} \boldsymbol{\alpha}_a \cdot \boldsymbol{H} \\ [\boldsymbol{H}, E_{\boldsymbol{\alpha}}] &= \boldsymbol{\alpha} E_{\boldsymbol{\alpha}} \\ [E_{\boldsymbol{\alpha}_a}, E_{-\boldsymbol{\alpha}_b}] &= 0 \end{aligned} \qquad [5]$$

it is straightforward to verify that the zero-curvature condition

$$F_{xt} = \partial_x A_t - \partial_t A_x + [A_x, A_t] = 0 \qquad [6]$$

is equivalent to the equations of motion which follow from extremizing the action [1].

The fact that there exists a flat connection which depends on an auxiliary parameter $z$ is sufficient to ensure integrability. In brief, the idea is that the gauge connection can be "abelianized" by a gauge transformation:

$$\tilde{A}_\mu = U \partial_\mu U^{-1} + U A_\mu U^{-1} \quad \text{with } [\tilde{A}_t, \tilde{A}_x] = 0 \qquad [7]$$

Hence, $\partial_t \tilde{A}_x - \partial_x \tilde{A}_t = 0$. This can be done in two inequivalent ways, such that $\tilde{A}_\mu$ are polynomials in $z$ and $z^{-1}$, respectively. The corresponding coefficients are then conserved currents whose integrals give conserved charges. It can be shown that for the Toda theories these conserved charges have Lorentz spin given by an exponent $\{s_a\}$ of $\mathfrak{g}$ modulo its Coxeter number $h$:

$$\begin{array}{lll} A_n: & h = n+1, & \{1,2,3,\ldots,n\} \\ D_n: & h = 2n-2, & \{1,3,5,\ldots,2n-3,n-1\} \\ E_6: & h = 12, & \{1,4,5,7,8,11\} \\ E_7: & h = 18, & \{1,5,7,9,11,13,17\} \\ E_8: & h = 30, & \{1,7,11,13,17,19,23,29\} \end{array} \qquad [8]$$

This spectrum of conserved quantities seems to be a ubiquitous feature of IQFTs. These theories can be generalized by replacing $\mathfrak{g}$, or rather its (untwisted affinization) with any affine algebra.

### The Sinh/Sine-Gordon Theory

These theories are the simplest of the Toda theories described above, associated to the Lie algebra $A_1$. In this case there is a single field and the potential has the form

$$V(\phi) = \frac{m^2}{2\beta^2} \left( e^{\beta\phi} + e^{-\beta\phi} \right) \qquad [9]$$

We have rescaled the field by $1/\sqrt{2}$ relative to the normalization in [2]. This potential defines the "sinh-Gordon theory." However, we can also take $\beta \to i\beta$ to give the sine-Gordon theory with an action

$$S = \int d^2x \left( \frac{1}{2} (\partial_\mu \phi)^2 + \frac{m^2}{\beta^2} \cos(\beta\phi) \right) \qquad [10]$$

The sine-Gordon theory is a useful paradigm for IQFTs because it exhibits most of the features of more complicated examples. To start with, it illustrates another important property of some integrable systems; namely, the existence of solitons. In the sine-Gordon case, the minima of the potential lie at $\phi = 2n\pi/\beta$, for an integer $n$, so there is a topological

kink that separates a vacuum $n$ on the left and $n+1$ on the right, as well as an antikink. The explicit solution for the kink moving with velocity $v$ is

$$\phi(x,t) = \frac{4}{\beta}\tan^{-1}\exp(m(x\cosh\theta - t\sinh\theta - \xi)) \quad [11]$$

where $\xi$ is a constant and, since we are working in $1+1$ dimensions, we have introduced the rapidity $\theta$, in terms of which the velocity is

$$v = \tanh\theta, \qquad -\infty \le \theta \le \infty \quad [12]$$

The antikink solution is simply the negative of the above. The kinks have a mass

$$M = \frac{8m}{\beta^2} \quad [13]$$

The existence of topological solitons is not a consequence of integrability, *per se*, for example, the $\phi^4$ theory in $1+1$ dimensions also has kinks; however, in the integrable setting, the solitons have special properties that survive in the quantum theory. The first property is that multisoliton solutions can be found exactly using a variety of different techniques. They are most easily written down using the tau function, which is related to the field via

$$\phi = -\frac{1}{i\beta}\log\frac{\tau}{\tau^*} \quad [14]$$

The $N$-soliton solution can then be written compactly as

$$\tau = \sum_{\{\mu_p\}=0,1}\exp\left(\sum_{p=1}^{N}\mu_p\Phi^{(p)} + \sum_{p,q=1}^{N}\mu_p\mu_q\Upsilon^{(pq)}\right) \quad [15]$$

The sum is over the $2^N$ possibilities for which $\mu_p = 0$ or 1, for each $p$, and we have introduced

$$\Phi^{(p)} = m(x\cosh\theta_p - t\sinh\theta_p - \xi_p) \pm \frac{i\pi}{2} \quad [16]$$

The rapidity of the $p$th soliton is $\theta_p$, and the choice of sign corresponds to the kink and antikink, respectively. The "interaction coefficient" is

$$\exp\Upsilon^{(pq)} = \tanh^2\left(\tfrac{1}{2}(\theta_p - \theta_q)\right) \quad [17]$$

For example, the two-soliton solution is

$$\tau = 1 + e^{\Phi^{(1)}} + e^{\Phi^{(2)}} + e^{\Upsilon+\Phi^{(1)}+\Phi^{(2)}} \quad [18]$$

The multisoliton solutions have a natural physical interpretation as the histories of a set of solitons which scatter off each other. To make this more precise, consider the two-soliton solution [18] in more detail. Suppose that $\xi_1 < \xi_2, v_1 > v_2$. Focus on the solution in the vicinity of the first soliton, that is,

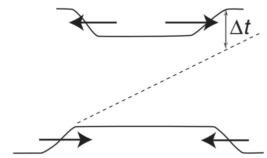



**Figure 1** Classical scattering of a kink and an antikink. The final velocities equal the initial velocities and the only effect is to introduce a velocity-dependent time delay as shown.

$x \sim v_1 t + \xi_1$. In the limit $t \to -\infty$, the solution is approximately

$$\tau \simeq 1 + e^{\Phi^{(1)}} \quad [19]$$

while, as $t \to \infty$, it is approximately

$$\tau \simeq e^{\Phi^{(2)}}\left(1 + e^{\Upsilon+\Phi^{(1)}}\right) \quad [20]$$

In both the limits, the solution represents an isolated soliton, the only difference is that the final "position offset" has been displaced: $\xi_1 \longmapsto \xi_1 - \Upsilon$. It is a consequence of integrability that the solitons interact in such a simple way. There were two solitons in the initial configuration and two in the final configuration traveling with the same velocities. The only effect is to introduce a time delay of

$$\Delta t = -\frac{\Upsilon(\theta)}{m\sinh(\theta/2)} \quad [21]$$

in the center-of-mass frame with $\theta_1 = -\theta_2 = \theta/2$, which we illustrate in Figure 1. We shall see that this kind of simple scattering is a characteristic feature of integrable field theories which extends to the quantum theory. It reflects the enormous restriction that the existence of the infinite set of integrals of motion puts on the dynamics.

## Integrability at the Quantum Level

In this section we turn to the particular implications of integrability for the field theories at the quantum level. In discussing theories in $1+1$ dimensions it is convenient, as in [12], to use the rapidity. The energy and momentum of a particle of mass $m$ are $E = m\cosh\theta$ and $p = m\sinh\theta$, respectively.

The sinh- and sine-Gordon theory, and their affine Toda generalizations, are scalar field theories with a well-behaved potential and as such they can be quantized in the conventional manner. It can be shown that integrability survives quantization and we now address its consequences. The key observation is that having an infinite set of higher-spin conserved quantities is very restrictive on the possible quantum processes. Assuming that the theory has a mass gap, the asymptotic states $|a,\theta\rangle$ are particles with rapidity

$\theta$ and additional quantum numbers needed to specify the state are indicated by the label $a$. These states are eigenstates of the conserved charges,

$$Q_s|a,\theta\rangle = q_s(a)e^{s\theta}|a,\theta\rangle \qquad [22]$$

Here, $s$ is the spin of the charge which ranges over some infinite subset of the integers. Since the charges must commute with the $S$-matrix, it follows immediately that if an incoming state of $n$ particles has a set of rapidities $\{\theta_1,\ldots,\theta_n\}$ then the outgoing state must also have $n$ particles with the same set $\{\theta_1,\ldots,\theta_n\}$: there is consequently no particle creation! For example, we have illustrated the scattering of two particles in Figure 2. The two-particle $S$-matrix element will be denoted as

$$S_{ab}^{cd}(\theta_1-\theta_2):\ |a,\theta_1;b,\theta_2\rangle \longrightarrow |c,\theta_2;d,\theta_1\rangle \qquad [23]$$

Note that masses of the incoming particles must match the outgoing ones: $m_a = m_d$ and $m_b = m_c$. We have already seen this kind of behavior with the classical scattering of solitons in the sine-Gordon theory. In spite of the fact that the scattering is purely elastic, it can be nontrivial for two reasons: if there are mass degeneracies in the theory, the quantum numbers $\{a_1,\ldots,a_n\}$ can change and, in addition, the $S$-matrix element can depend nontrivially on the momenta.

The fact that the incoming and outgoing states have the same set of momenta leads to the notion of factorizability. To see what this means, consider the case of three particles. Let us imagine that we prepare the initial state to consist of three fairly narrow wave packets in position space with momenta smeared in accordance with the uncertainty principle. The key to the following argument is the fact that the infinite set of higher-spin conserved charges (with commute with the $S$-matrix) allow one to move the positions of the three particles relative to each other in an arbitrary way. In addition, the theory has a mass gap, so interactions have a finite range. By using this freedom, we can arrange for particles 1 and 2 to interact first,

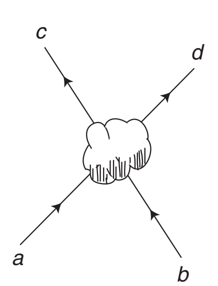



**Figure 2** The two particle $S$-matrix with particles $a$ and $b$ in the initial state and $c$ and $d$ in the final state. For consistency, $m_a = m_d$ and $m_b = m_c$.

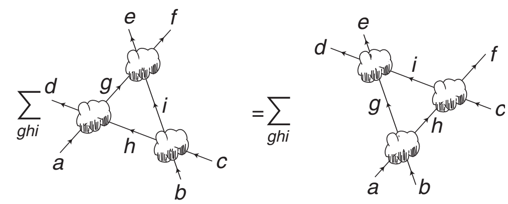



**Figure 3** The scattering of three particles can factorize in two distinct ways as illustrated, leading to a nontrivial condition: the Yang–Baxter equation.

well before they come within interaction range of the third. Subsequently, the first two particles interact with the third as on the right-hand side of Figure 3. This ability to move the wave packets around using the symmetries means that the three-particle $S$-matrix element must "factorize" into a product of three two-particle elements:

$$S_{abc}^{def}(\theta_1,\theta_2,\theta_3) = \sum_{ghi} S_{ab}^{gh}(\theta_1-\theta_2)S_{bc}^{if}(\theta_1-\theta_3)S_{gi}^{de}(\theta_2-\theta_3) \qquad [24]$$

However, we could also use the symmetries afforded by the conserved charges to shift the positions of the particles so that particle 2 and 3 interact first, as on the left-hand side of Figure 3. Since the charges commute with the $S$-matrix, the result must the same; hence, there is a nontrivial consistency condition:

$$\sum_{ghi} S_{bc}^{hi}(\theta_2-\theta_2)S_{ah}^{dg}(\theta_1-\theta_3)S_{gi}^{ef}(\theta_2-\theta_3) = \sum_{ghi} S_{ab}^{gh}(\theta_1-\theta_2)S_{bc}^{if}(\theta_1-\theta_3)S_{gi}^{de}(\theta_2-\theta_3) \qquad [25]$$

This is the celebrated Yang–Baxter equation. Notice that it is only nontrivial if there are mass degeneracies, otherwise the particles on internal lines are determined by the external particles.

The factorization of the $S$-matrix extends readily to the case of more particles in an obvious way. An $n$-body element factorizes into a two-body element for each pair of particles. One might think that considerations of the $n$-particle $S$-matrix would lead to additional constraints; however, it can readily be shown that this is not the case and that the Yang–Baxter equation acts as a basic "move" which allows one to reorder the $n$-particle $S$-matrix into an arbitrary order. Further conditions on the $S$-matrix come from the axioms of analytic $S$-matrix theory:

(i) *Unitarity*

$$\sum_{ef} S_{ab}^{ef}(\theta)S_{ef}^{cd}(-\theta) = \delta_{ac}\delta_{bd} \qquad [26]$$

(ii) *Crossing symmetry* Each particle $a$ has an antiparticle $\bar{a}$ and

$$S_{ab}^{cd}(\theta) = S_{b\bar{d}}^{\bar{a}c}(\pi i - \theta) \qquad [27]$$

(iii) *Analyticity* The $S$-matrix is a meromorphic function of $\theta$ on the physical strip, $0 \leq \mathrm{Im}\,\theta \leq \pi$. Singularities in most instances occur along the imaginary axis and the simple poles correspond to direct or cross-channel resonances. In this case, if $S_{ab}^{de}(\theta)$ has a simple pole at $\theta = \mathrm{i}u_{ab}^c$ (necessarily a nonphysical rapidity difference) in the direct channel there exists a bound state of $a$ and $b$ of mass

$$m_c^2 = m_a^2 + m_b^2 + 2m_a m_b \cos u_{ab}^c \qquad [28]$$

The situation is illustrated in Figure 4. The new particle must itself be included in the particle spectrum. The $S$-matrix elements at the pole have the form

$$S_{ab}^{de}(\theta) = \sum_c P_{ab}^c \frac{\mathrm{i}r_{ab}^c}{\theta - \mathrm{i}u_{ab}^c} P_{\bar{d}\bar{e}}^{\bar{c}} + \cdots \qquad [29]$$

where $P_{ab}^c$ can be thought of as a kind of projection operator with

$$\sum_{ab} P_{ab}^c P_{\bar{b}\bar{a}}^{\bar{d}} = \delta^{cd} \qquad [30]$$

Unitarity of the QFT requires that $r_{ab}^c$ is real and positive, although there are also examples of nonunitarity theories with exact $S$-matrices. If $ab \to c$ can occur then so can $a\bar{c} \to \bar{b}$ and $b\bar{c} \to \bar{a}$. From [28], we deduce the following identity:

$$u_{ab}^c + u_{a\bar{c}}^{\bar{b}} + u_{b\bar{c}}^{\bar{a}} = 2\pi \qquad [31]$$

The data $\{u_{ab}^c\}$ for any given scattering theory are known as the fusing angles.

(iv) *The Bootstrap equations* These give a nonlinear relation between $S$-matrix elements. The basic idea is that if particle $c$ appears as a resonance in the scattering of $a$ and $b$ then the $S$-matrix element of $c$ with another state $d$ can be deduced in terms of the scattering of $d$ with $a$ and $b$. This is illustrated in Figure 5. Using [30], we can write the resulting equation for the $S$-matrix element of $c$ and $d$ directly:

$$S_{cd}^{ef}(\theta) = \sum_{ghi} P_{\bar{a}\bar{b}}^{\bar{c}} S_{ab}^{eg}\left(\theta - \mathrm{i}\bar{u}_{a\bar{c}}^{\bar{b}}\right) S_{bd}^{hi}\left(\theta + \mathrm{i}\bar{u}_{b\bar{c}}^{\bar{a}}\right) P_{gi}^{f} \qquad [32]$$

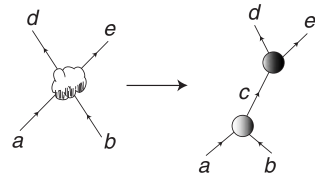



**Figure 4** Near a direct channel pole, the scattering of *a* and *b* is dominated by the bound state *c*.

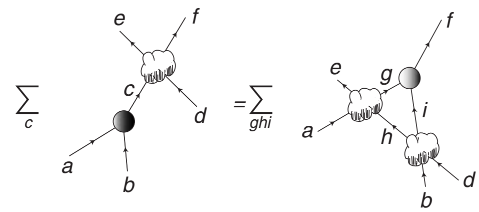



**Figure 5** The bootstrap equations result from considering the interaction of a particle *d* with the bound state *c* of *a* and *b* in two distinct ways as illustrated.

The bootstrap constraints are very powerful because they allow one to extract the $S$-matrix elements of new particles that appear as bound states. This leads to the philosophy of the “bootstrap program” where one attempts to build consistent $S$-matrices starting from the $S$-matrix for a subset of particles which act as a seed for the algorithm. The process is quite an art, but at the end one has to be satisfied that the complete analytic structure is consistent with all the axioms. The key is to be able to account for all the poles in a consistent way, either in terms of bound states, as above, or in terms of the Coleman–Thun mechanism. This allows some poles to be interpreted in ways other than the existence of a bound state. The bootstrap algorithm is very complicated in general and at the present time a complete classification of solutions is not known. However, there are a large number of known solutions which appear to be intimately related to Lie algebras and associated structures known as Yangians and quantum groups. Below we describe some of the simplest known solutions.

## Minimal *S*-Matrices

These scattering theories are in some sense the simplest. The particle spectrum is generally nondegenerate and so the Yang–Baxter equation is trivial. As is ubiquitous in the subject of IQFT, the classification of the theories is related to Lie algebras, although what seems to be important is not so much the algebra in question but rather the details of the associated root system. In this case the appropriate algebras are the simply laced algebras of ADE type. The number of particles is equal to the rank $r$ of the Lie algebra and the masses are given by the $r$ elements of one of the eigenvectors of the Cartan matrix of the algebra $\mathfrak{g}$:

$$\sum_{b=1}^{r} C_{ab} m_b = \left(2 - 2\cos\frac{\pi}{h}\right) m_a \qquad [33]$$

where $h$ is the Coxeter number of $\mathfrak{g}$. The conserved charges have spins corresponding to the exponents of $\mathfrak{g}$ modulo $h$. We briefly explain how the complete

$S$-matrix can be written down in terms of properties of the root system of $\mathfrak{g}$. Let $\boldsymbol{\Phi}$ be the set of roots of $\mathfrak{g}$, and $\boldsymbol{\alpha}_a, a=1,\ldots,r$, a set of simple roots, as in the last section. In terms of these, $C_{ab}=2\boldsymbol{\alpha}_a\cdot\boldsymbol{\alpha}_b/\boldsymbol{\alpha}_b^2$. Let $\boldsymbol{\omega}_a, a=1,\ldots,r$, be a corresponding set of fundamental weights, $\boldsymbol{\alpha}_a\cdot\boldsymbol{\omega}_b=\delta_{ab}$.

Key to defining the theories is the notation of the Weyl group of $\mathfrak{g}$, the group generated by reflections in the simple roots:

$$R_a(\boldsymbol{\alpha})=\boldsymbol{\alpha}-\frac{2\boldsymbol{\alpha}\cdot\boldsymbol{\alpha}_a}{\boldsymbol{\alpha}_a^2}\boldsymbol{\alpha}_a \qquad [34]$$

The element $w=R_1R_2\cdots R_r$ is known as a Coxeter element of the Weyl group, and it has special properties that are significant in the present context. In particular, its eigenvalues are of the form $\exp(2\pi i s_a/h)$, where $h$ is the Coxeter number of $\mathfrak{g}$ and the integers $s_a$ are the exponents of the algebra as in [8]. Note that there is always a pair with $s_1=1$ and $s_r=h-1$. Clearly, $w$ acts as a rotation in the two-dimensional space spanned by the two corresponding eigenvectors. We can define an antisymmetric function $u(\boldsymbol{\alpha},\boldsymbol{\beta})$ on roots to be $h/\pi$ times the (signed) angle between the projections of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ onto this two-dimensional eigenspace. In preparation for what follows, it is useful to also define the roots

$$\boldsymbol{\phi}_a=R_rR_{r-1}\cdots R_{a+1}(\boldsymbol{\alpha}_a) \qquad [35]$$

We can now present P Dorey's amazingly compact formula for the complete $S$-matrix. For the scattering of particle $a$ with particle $b$,

$$S_{ab}(\theta)=\prod_{\boldsymbol{\beta}\in\Gamma_b}\{1+u(\boldsymbol{\phi}_a,\boldsymbol{\beta})\}^{\boldsymbol{\omega}_a\cdot\boldsymbol{\beta}} \qquad [36]$$

In this formula $\Gamma_b$ is the set of positive roots of $\mathfrak{g}$ which lie in the orbit of $\boldsymbol{\phi}_b$ under $w$. We have also defined the building block

$$\{x\}=(x+1)(x-1)$$
$$(x)=\frac{\sinh\left(\dfrac{\theta}{2}+\dfrac{i\pi x}{h}\right)}{\sinh\left(\dfrac{\theta}{2}-\dfrac{i\pi x}{h}\right)} \qquad [37]$$

The fusing rules are also particularly elegant in the language of root systems. There is a three-point coupling between $a_i, i=1,2,3$, if there exist three roots $\boldsymbol{\alpha}^{(i)}\in\Gamma_{a_i}$ such that $\boldsymbol{\alpha}^{(1)}+\boldsymbol{\alpha}^{(2)}+\boldsymbol{\alpha}^{(3)}=0$. Furthermore, the fusing occurs in the $a_1, a_2$ channel at rapidity difference

$$iu_{a_1a_2}^{\bar{a}_3}=\frac{i\pi}{h}u(\boldsymbol{\alpha}^{(1)},\boldsymbol{\alpha}^{(2)}) \qquad [38]$$

This is Dorey's fusing rule.

For the case of $A_{n-1}$, the $S$-matrices are particularly simple. The mass spectrum is

$$m_a=m\sin\frac{\pi a}{n}, \quad a=1,\ldots,n-1 \qquad [39]$$

and Dorey's rule gives the possible fusings as $ab\to(a+b)\mathrm{mod}\ n$, which occur at the rapidity values

$$\theta=iu_{ab}=\begin{cases} i\dfrac{a+b}{n}\pi & a+b<n \\ i\left(2-\dfrac{a+b}{n}\right)\pi & a+b\geq n \end{cases} \qquad [40]$$

The charge conjugation operator maps $a\to\bar{a}=n-a$ and the explicit form for the $S$-matrix elements is

$$S_{ab}(\theta)=\{a+b-1\}\{a+b-3\}\cdots\{|a-b|+1\} \qquad [41]$$

The element $S_{ab}(\theta)$ has one direct channel pole at $\theta=iu_{ab}$ corresponding to the exchange of the particle $a+b \bmod n$, and a cross-channel pole at $\theta=iu_{a\bar{b}}$ corresponding to the exchange of particle $a-b \bmod n$.

## Affine Toda Theories

The bootstrap program has been solved for all the affine Toda theories. For the simply laced theories described earlier, the result is directly related to the minimal $S$-matrices constructed above. The only difference is that there are additional factors which depend on the coupling $\beta$ of the Toda theory but which do not introduce any additional poles onto the physical strip. These CDD factors are included by simply changing the basic building block [37]:

$$\{x\}\to\{x\}_{\mathrm{Toda}}=\frac{(x+1)(x-1)}{(x-1+B)(x+1-B)} \qquad [42]$$

where

$$B=\frac{1}{2\pi}\cdot\frac{\beta^2}{1+\beta^2/4\pi} \qquad [43]$$

The $S$-matrix structure for the Toda theories based on the nonsimply laced algebras is a good deal more complicated. Integrability is only maintained in the quantum theory if the ratios of the physical masses of the particles depend on the coupling constant $\beta$ is some very special way.

## The Sine-Gordon Theory

We have seen that the sine-Gordon theory has solitons at the classical level. At the quantum level,

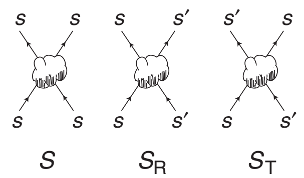



**Figure 6** Soliton scattering processes. *s* and *s*′ are the kink and antikink, respectively, or vice versa.

we expect that these kinks become *bona fide* particle states, in addition to the particle corresponding to the quantum fluctuations of the field $\phi$. Focusing on the solitons, we expect a degenerate doublet corresponding to the kink and antikink. For the scattering of two solitons, there are six allowed processes illustrated in **Figure 6**. Unitarity [26] leads to the constraints

$$\begin{aligned} &S(\theta)S(-\theta) = 1 \\ &S_{\mathrm{T}}(\theta)S_T(-\theta) + S_{\mathrm{R}}(\theta)S_{\mathrm{R}}(-\theta) = 1 \\ &S_{\mathrm{T}}(\theta)S_{\mathrm{R}}(-\theta) + S_{\mathrm{R}}(\theta)S_{\mathrm{T}}(-\theta) = 0 \end{aligned} \qquad [44]$$

while crossing symmetry [27] (using the fact that the soliton and antisoliton are antiparticles) gives

$$S(\mathrm{i}\pi-\theta) = S_T(\theta), \qquad S_{\mathrm{R}}(\mathrm{i}\pi-\theta) = S_R(\theta) \qquad [45]$$

By themselves, these constraints are rather mild; however, the complete soliton *S*-matrix must also satisfy the Yang–Baxter equation [25]. The solution to all the constraints is not unique, however, the Zamolodchikovs conjectured that the exact answer is

$$\begin{aligned} S(\theta) &= \frac{1}{\mathrm{i}\pi}\sinh\left(\frac{8\pi}{\gamma}(\mathrm{i}\pi-\theta)\right)U(\theta) \\ S_{\mathrm{T}}(\theta) &= \frac{1}{\mathrm{i}\pi}\sinh\left(\frac{8\pi}{\gamma}\theta\right)U(\theta) \\ S_{\mathrm{R}}(\theta) &= \frac{1}{\pi}\sin\left(\frac{8\pi^2}{\gamma}\right)U(\theta) \end{aligned} \qquad [46]$$

with

$$\begin{aligned} U(\theta) &= \Gamma\left(\frac{8\pi}{\gamma}\right)\Gamma\left(1+\mathrm{i}\frac{8\theta}{\gamma}\right)\Gamma\left(1-\frac{8\pi}{\gamma}-\mathrm{i}\frac{8\theta}{\gamma}\right) \\ &\quad\times\prod_{n=1}^{\infty}\frac{R_n(\theta)R(\mathrm{i}\pi-\theta)}{R_n(0)R_n(\mathrm{i}\pi)} \\ R_n(\theta) &= \frac{\Gamma\left(2n\frac{8\pi}{\gamma}+\mathrm{i}\frac{8\theta}{\gamma}\right)}{\Gamma\left((2n+1)\frac{8\pi}{\gamma}+\mathrm{i}\frac{8\theta}{\gamma}\right)} \\ &\quad\times\frac{\Gamma\left(1+2n\frac{8\pi}{\gamma}+\mathrm{i}\frac{8\theta}{\gamma}\right)}{\Gamma\left(1+(2n-1)\frac{8\pi}{\gamma}+\mathrm{i}\frac{8\theta}{\gamma}\right)} \end{aligned} \qquad [47]$$

where $\gamma = \beta^2(1-\beta^2/8\pi)^{-1}$. The reason for confidence in the conjecture is that from the soliton *S*-matrix one can complete the bootstrap program and account for all the poles in terms of particles in the theory. In particular, there is a finite set of bound states of the soliton and antisoliton, called breathers, with masses

$$m_k = 2M\sin\frac{k\gamma}{16}, \quad k = 1,2,\ldots < \frac{8\pi}{\gamma} \qquad [48]$$

Here, $M$ is the soliton mass. The bootstrap equations give the *S*-matrix for the scattering of a soliton or antisoliton with the $k$th breather,

$$S_k(\theta) = \frac{\sinh\theta+\mathrm{i}\cos\frac{k\gamma}{16}}{\sinh\theta-\mathrm{i}\cos\frac{k\gamma}{16}} \times\prod_{j=1}^{k-1}\frac{\sin^2\left(\frac{k-2j}{32}\gamma-\frac{\pi}{4}+\mathrm{i}\frac{\theta}{2}\right)}{\sin^2\left(\frac{k-2j}{32}\gamma-\frac{\pi}{4}-\mathrm{i}\frac{\theta}{2}\right)} \qquad [49]$$

while, for the scattering of breather $k$ with $l$,

$$S_{kl}(\theta) = \frac{\sinh^2\theta+\mathrm{i}\sin\left(\frac{k+l}{16}\gamma\right)\sinh\theta+\mathrm{i}\sin\left(\frac{k-l}{16}\gamma\right)}{\sinh^2\theta-\mathrm{i}\sin\left(\frac{k+l}{16}\gamma\right)\sinh\theta-\mathrm{i}\sin\left(\frac{k-l}{16}\gamma\right)} \times\prod_{j=1}^{l-1}\frac{\sin^2\left(\frac{k-l-2j}{32}\gamma+\mathrm{i}\frac{\theta}{2}\right)\cos^2\left(\frac{k+l-2j}{32}\gamma+\mathrm{i}\frac{\theta}{2}\right)}{\sin^2\left(\frac{k-l-2j}{32}\gamma-\mathrm{i}\frac{\theta}{2}\right)\cos^2\left(\frac{k+l-2j}{32}\gamma-\mathrm{i}\frac{\theta}{2}\right)} \qquad [50]$$

where we assume, without loss of generality, that $k \geq l$. The remarkable thing is that the scattering of the lowest-mass breather $m_1$ with itself,

$$S_{11}(\theta) = \frac{\sinh\theta + \mathrm{i}\sin\frac{\gamma}{8}}{\sinh\theta - \mathrm{i}\sin\frac{\gamma}{8}} \qquad [51]$$

is precisely the Toda $S$-matrix for $A_1$ with $\beta \to \mathrm{i}\beta/\sqrt{2}$ (the origin of the factor of $\sqrt{2}$ is mentioned after eqn [9]). This uniquely identifies the lowest-mass breather as being the quantum of the $\phi$ field.

The quantum structure that we have described above can be directly related to the classical scattering of solitons. In order to implement the classical limit, we can reintroduce $\hbar$ which is achieved by replacing $\beta^2$ by $\beta^2\hbar$. In this limit, the $S$-matrix elements have the form

$$S(\theta) = \exp\frac{2\mathrm{i}}{\hbar}(\delta(\theta) + \mathcal{O}(\hbar)) \qquad [52]$$

The phase $\delta(\theta)$ is related via the WKB approximation to the time delay in the classical theory of soliton scattering via

$$\delta(\theta) = \text{const.} + \int_0^\theta \mathrm{d}\theta' M\sinh(\theta/2)\Delta t(\theta) \qquad [53]$$

where $\Delta t(\theta)$ is the time delay in the center of mass (21). It is possible to verify [53] for the processes $S(\theta)$ and $S_{\mathrm{T}}(\theta)$. Note that the reflection process has no classical analogue.

## IQFT, Conformal Field Theories and Statistical Systems

We have described some IQFTs and their factorizable $S$-matrices in theories with a mass gap. We can ask the question, "what happens at very high energies compared with all the mass scales?" For a generic QFT such a limit may not exist, however, for a special class of theories the limit is a massless scale-invariant theory corresponding to a fixed point of the renormalization group. The massive theory can be thought of as a deformation of the massless theory by a particular relevant operator. At the fixed point, the Poincaré symmetry is enhanced to the full conformal group in the appropriate number of dimensions and the resulting theory is known as a conformal field theory (CFT). In $1+1$ dimensions the conformal group is infinite dimensional and so many CFTs are themselves integrable, in the sense that the complete spectrum of fields is known and their correlation functions can be constructed. Hence, an alternative way of thinking about many IQFTs is as a perturbation of a CFT by a specific relevant operator:

$$S_{\mathrm{IQFT}} = S_{\mathrm{CFT}} + g\int \mathrm{d}^2x\mathcal{O}(x) \qquad [54]$$

We will suppose that the operator has conformal dimensions $(\Delta, \bar{\Delta})$. This description of the theory can be turned around to ask the following question: which relevant deformations of a given CFT lead to IQFTs? Remarkably, since CFTs are so well understood, the question can often be answered exactly. The idea is that the conserved quantities of a CFT are all (anti-)holomorphic with respect to a holomorphic coordinate $z = x + \mathrm{i}t$. Conserved quantities include the stress tensor of spin 2 but include, in addition, an infinite tower of currents of ever increasing spin $\{T_s\}$. After perturbation, one has

$$\bar{\partial}T_s = gR^{(1)} + \cdots + g^nR^{(n)} + \cdots \qquad [55]$$

The conformal dimensions of the $R^{(n)}$ are $(s - n(1-\Delta), 1 - n(1-\Delta))$. Since the conformal dimensions of fields in a CFT are bounded below by zero, it follows that the series on the right-hand side truncates. The question of whether $T_s$ remains conserved away from the CFT boils down to the question as to whether the right-hand side has the form $\partial\Theta$, for some $\Theta$. Zamolodchikov found an ingenious counting argument which showed in certain circumstances that the right-hand side has precisely this form for some $s > 2$. This is sufficient to establish that the perturbed theory is an IQFT. In certain cases the spectrum of spins of the conserved quantities that are established by the counting argument is enough to make a connection with a known factorizable $S$-matrix.

This way of viewing IQFT as perturbations of CFTs is especially fruitful when we make the connection of the Euclidean QFT with the classical statistical mechanics of a two-dimensional system. In this connection, the Feynman path integral is reinterpreted as the sum over the configurations in the canonical ensemble with the Euclidean action interpreted as the energy. Usually, we consider statistical systems which are discrete, so typically defined on a lattice. The Euclidean QFTs are to be thought of as these statistical systems in the continuum limit where the lattice spacing is taken to zero keeping the long-range physics fixed. CFTs which have no massive degrees of freedom are identified with points of second-order phase transitions in the statistical system where correlation lengths are infinite. Perturbations of CFTs by relevant operators correspond to taking the statistical system away from criticality by changing some external parameter.

The prototypical example of such a statistical system is the Ising model. In the lattice version of

this model, there are a set spins $\{\sigma_i\}$ at each lattice site which can take the discrete values $\pm 1$. The partition function of the theory is

$$Z(H,T) = \sum_{\{\sigma_i\}} \exp\left( - T^{-1} \sum_{\langle i,j \rangle} \sigma_i \sigma_j - H \sum_i \sigma_i \right) \quad [56]$$

The Ising model is the simplest model of a ferromagnet, where $T$ is the temperature and $H$ is the external applied field. The theory has a second-order phase transition for $T = T_c$, the Curie temperature, and $H = 0$ when the competition between the energy, which favors aligning the spins, and entropy, which favors disorder, exactly balance. In the two-dimensional neighborhood of the critical point, the lattice theory admits a continuum limit which can be described as the perturbation of a CFT, describing the critical Ising model, by a pair of relevant operators with couplings $T - T_c$ and $H$. In the case of the Ising model, the CFT is simply the theory of a free massless fermion in two-dimensional Euclidean space.

It turns out that in the two-dimensional space of relevant perturbations, there are two directions which lead to IQFTs. The most obvious is changing the temperature away from $T_c$ while keeping $H = 0$. This leads to a particularly simple IQFT, that of a free massive fermion. More unexpectedly, the direction for which $H$ varies away from 0, but $T = T_c$, also leads to an IQFT. In this case, Zamolodchikov's counting argument shows that there are higher-spin conserved charges of spin including

$$s = 1, 7, 11, 13, 17, 19, \ldots \quad [57]$$

This is remarkable because, as we have described previously, there is a minimal solution of the bootstrap program that describes the scattering of eight particles which has a spectrum of conserved charges that includes these spins. It is the minimal scattering theory related to the algebra $E_8$.

The fact that the scattering theory of the off-critical Ising model in the magnetic field direction has been identified is remarkable. From the $S$-matrix one can proceed to investigate the off-critical correlation functions using a technique known as the "form factor programe." Detailed simulation of the original lattice model [56] has provided strong support for the veracity of the $E_8$ scattering theory. For instance, the two lightest masses in the scattering theory determine the ratio of the two longest correlation lengths $m_2/m_1 = 2\cos(\pi/5)$.

In general, the identification of an IQFT and the CFT at its ultraviolet limit can be more difficult to establish. One way to proceed is to use the thermodynamic Bethe ansatz. This technique involves considering the thermodynamics of a gas of the particles in a periodic box. Since the scattering is purely elastic, thermodynamic quantities can be calculated, albeit in terms of a set of coupled nonlinear integral equations. If the box is small enough, ultraviolet effects dominate and various features of the CFT can be recovered.

## Other IQFTs

There is a rich menagerie of other IQFTs that we have no space to discuss in detail. One is sigma models, whose fields take values in a Riemannian target space $\mathfrak{M}$ with an action

$$S = \int d^2x g_{ab} \partial_\mu X^a \partial^\mu X^b \quad [58]$$

where $g_{ab} dX^a\, dX^b$ is the metric of $\mathfrak{M}$. These theories are integrable at the classical level if the target space is either a group manifold of a compact simple group $G$ or a symmetric space coset $G/H$, where $H$ is a suitable subgroup of $G$. The former are known as the "principal chiral models." There are two kinds of conserved quantities, both local and nonlocal. At the quantum level, the conserved currents which imply classical integrability can be subject to quantum anomalies. An analysis of these anomalies proves that the principal chiral models are all integrable at the quantum level, while only the subset of symmetric space coset models, namely

$$\begin{array}{ll} SO(n+1)/SO(n), & SU(n)/SO(n) \\ SU(2n)/Sp(n), & SO(2n)/SO(n) \times SO(n) \\ Sp(2n)/Sp(n) \times Sp(n) & \end{array} \quad [59]$$

are quantum integrable. $S$-matrices have been proposed for all these integrable sigma models. They have a more complicated structure than most of the cases discussed here, because the particles fall into representations of the associated Lie groups and the Yang–Baxter equation, such as for the sine-Gordon solitons, is now nontrivial. Remarkably, gross features of the $S$-matrices, such as the mass spectrum fusing rules, are identical to the Toda theories or the minimal $S$-matrices.

Returning to IQFTs that are associted with deformations of CFTs, there are more general classes which are associated with the renormalization group trajectories between two nontrivial fixed points. These theories have both massless and massive degrees of freedom. Even more remarkable are the staircase models of Zamolodchikov that exhibit an infinite series of crossover behavior where the renormalization group trajectory passes close to an infinite series of fixed points in sequence.

For all of the theories described above, one might have thought more generally that integrability is a very rigid property of a theory. In general, for example, the number of external coupling constants is very limited and the mass ratios are all fixed. For

example, in Toda theories there is only an overall mass scale $m$ and the coupling $\beta$. If the form of the potential is altered in any way then integrability is lost. However, in certain circumstances, integrability appears to be a looser constraint that allows more flexibility. One class of such theories is known as the homogeneous sine-Gordon theories. These are integrable deformations of gauged WZW models associated with the coset $G/\mathrm{U}(1)^r$, where $r$ is the rank of a simple compact group $G$. In these theories there is a rich spectrum of both stable and unstable particles with masses and an $S$-matrix that depends continuously on a set of $r$ coupling constants.

*See also:* Algebraic Approach to Quantum Field Theory; Bethe Ansatz; Constructive Quantum Field Theory; Eight Vertex and Hard Hexagon Models; Functional Equations and Integrable Systems; Integrable Systems: Overview; Quantum Field Theory: A Brief Introduction; Quantum Field Theory in Curved Spacetime; Sine-Gordon Equation; Symmetries in Quantum Field Theory of Lower Spacetime Dimensions; Two-Dimensional Models; Yang–Baxter Equations.

## Further Reading


Arinshtein AE, Fateev VA, and Zamolodchikov AB (1979) Quantum $S$ matrix of the $(1+1)$-dimensional Todd chain. *Physics Letters B* 87: 389.

Braden HW, Corrigan E, Dorey PE, and Sasaki R (1990) Affine Toda field theory and exact $S$ matrices. *Nuclear Physics B* 338: 689.

Delfino G (2004) Integrable field theory and critical phenomena: the Ising model in a magnetic field. *Journal of Physics A* 37: R45 (arXiv:hep-th/0312119).

Delius GW, Grisaru MT, and Zanon D (1992) *Nuclear Physics B* 382: 365 (arXiv:hep-th/9201067).

Dorey P (1992) Root systems and purely elastic $S$ matrices. 2. *Nuclear Physics B* 374: 741 (arXiv:hep-th/9110058).

Dorey P (1998) Exact S matrices, arXiv:hep-th/9810026.

Dorey PE and Ravanini F (1993) Staircase models from affine Toda field theory. *International Journal of Modern Physics A* 8: 873 (arXiv:hep-th/9206052). (A B Zamolodchikov's original paper on the staircase models 'Resonance factorized scattering and roaming trajectories' is unpublished.).

Evans JM, Kagan D, MacKay NJ, and Young CAS (2005) Quantum, higher-spin, local charges in symmetric space sigma models. *Journal of High Energy Physics* 0501: 020 (arXiv:hep-th/0408244).

Jackiw R and Woo G (1975) Semiclassical scattering of quantized nonlinear waves. *Physical Review D* 12: 1643.

Miramontes JL and Fernandez-Pousa CR (2000) Integrable quantum field theories with unstable particles. *Physics Letters B* 472: 392 (arXiv:hep-th/9910218).

Mussardo G (1992) Off critical statistical models: factorized scattering theories and bootstrap program. *Physics Reports* 218: 215.

Olive DI and Turok N (1985) Local conserved densities and zero curvature conditions for Toda lattice field theories. *Nuclear Physics B* 257: 277.

Olshanetsky MA and Perelomov AM (1981) Classical integrable finite dimensional systems related to Lie algebras. *Physics Reports* 71: 313.

Zamolodchikov AB (1989) Integrals of motion and $S$ matrix of the (scaled) $T=T(C)$ Ising model with magnetic field. *International Journal of Modern Physics A* 4: 4235.

Zamolodchikov AB and Zamolodchikov AB (1979) Factorized $S$-matrices in two dimensions as the exact solutions of certain relativistic quantum field models. *Annals of Physics* 120: 253.

Zamolodchikov AB (1990) Thermodynamic Bethe ansatz in relativistic models. Scaling three state Potts and Lee–Yang models. *Nuclear Physics B* 342: 695.


# Integrable Discrete Systems

**O Ragnisco**, Università "Roma Tre", Rome, Italy



## Discrete Dynamical Systems

The expression "dynamical system" usually refers to a coupled system of ordinary differential equations (ODEs), namely,

$$\dot{x}_j(t) = f_j(t, x_1, \ldots, x_N), \quad j = 1, \ldots, N \qquad [1]$$

where $t$ belongs to some set of nonzero measure $I$ of the real line $\mathbb{R}$, typically an interval $[a, b]$ or a semiline or the whole line, and $x_j$ are sufficiently smooth functions from $I$ to $\mathbb{R}$ or to $\mathbb{C}$.

The system [1] is complemented by initial or boundary conditions that make it into an "initial-value" or a "boundary-value" problem. Under suitable regularity assumptions on the RHS, the existence and uniqueness of the solution of the initial-value problem is guaranteed, but in most cases the solution can be known only "approximately" either through perturbation theory or just through numerical integration. This is not the proper place to discuss finite-difference schemes for systems of ODEs: what is relevant is that such numerical schemes (think, e.g., of Euler or Runge–Kutta schemes) "discretize" the continuous independent variable $t$ by replacing it by an integer variable $n \in \mathbb{Z}$: in the simplest case, the interval $[a, b]$ is replaced by a set of $L$ equally spaced points $t_n = a + n(b-a)/L (n = 1, \ldots, L)$, the first derivative is approximated by a (forward) difference, and the system [1] is converted into a system of "difference" equations of the form

$$x_j(n+1) = x_j(n) + hF(n, x_1(n), \ldots, x_N(n)) \qquad [2]$$

where $h$ denotes the time step $(b-a)/L$.

The coupled system [2] is an example of a "discrete dynamical system," explicit (because the updated variables only depend upon the values taken

at previous discrete times), first order (only "nearest-neighbor" discrete times, $n$, $n+1$ are involved), but nonautonomous, as the RHS is allowed to depend explicitly upon the independent variable $n$, analogously to its continuum counterpart.

In the following, "autonomous" but not necessarily explicit discrete dynamical systems of a special type will be considered: in fact, we will require them to be equipped with a Hamiltonian structure, and we will define the notion of complete integrability (in the Arnol'd–Liouville sense) for such systems.

This article emphasizes on some aspects and properties of integrable discrete systems, neglecting others that could be equally important. In particular, as no nonautonomous discrete systems will be considered, discrete analogs of Painleve' equations will never be discussed in this article, and consequently the intriguing issues concerning "singularity confinement" in the discrete and "algebraic entropy" will not be touched upon (see, e.g., Grammaticos *et al.* (2004)). Similarly, neither the integrability for discrete systems in multidimensional space nor "quantum integrable mappings" will be discussed.

## Lagrangian and Hamiltonian Formulations

Following the historical path along which modern classical mechanics has been developed, first the concept of a Lagrangian map is introduced, and then Hamiltonian (in fact, symplectic) maps are defined through a proper discrete version of the Legendre transformation.

Let $x_j(n)\,(j=1,\ldots,N, n\in\mathbb{Z})$ be $N$ sequences of real numbers and let $\mathcal{L}(\boldsymbol{x},\boldsymbol{y})$ be a smooth function from $\mathbb{R}^N\times\mathbb{R}^N$ into the reals, $\boldsymbol{x}$ denoting the $N$-tuple $x_1,\ldots,x_N$. $\mathcal{L}$ is regarded as a "discrete Lagrange function": corresponding to each discrete time $n$, it is assigned a certain value $\mathcal{L}_n := \mathcal{L}(\boldsymbol{x}(n),\boldsymbol{x}(n+1))$. The corresponding discrete action functional $S[\mathcal{L}]$ is defined in a natural way:

$$S[\mathcal{L}] = \sum_{n=N_a}^{N_b} \mathcal{L}_n \qquad [3]$$

The actual "discrete trajectory" will be given by the sequence $\boldsymbol{x}(n)$ that corresponds to a "critical point" of the action [3] subject to the constraints $\delta\boldsymbol{x}(N_a)=\delta\boldsymbol{x}(N_b)=0$. Note that the values $N_a\,(N_b)$ may well possibly coincide with $-\infty\,(+\infty)$. Such "critical points" are given by the solution of the discrete Euler–Lagrange equations:

$$\left.\frac{\partial\mathcal{L}}{\partial x_j}\right|_{x_j=x_j(n),y_j=x_j(n+1)} + \left.\frac{\partial\mathcal{L}}{\partial y_j}\right|_{x_j=x_j(n-1),y_j=x_j(n)} = 0 \qquad [4]$$

It is worthwhile to remark the intrinsic nature of eqns [4], whose form turns out to be independent of the choice of a coordinate chart. In fact, by omitting the explicit dependence on $n$ and simply denoting $\boldsymbol{x}(n)=\boldsymbol{x}$, $\boldsymbol{x}(n+1)=\tilde{\boldsymbol{x}}$, $\boldsymbol{x}(n-1)=\underset{\sim}{\boldsymbol{x}}$, [4] can be cast in the form

$$\nabla_1\mathcal{L}(\boldsymbol{x},\tilde{\boldsymbol{x}}) + \nabla_2\mathcal{L}(\underset{\sim}{\boldsymbol{x}},\boldsymbol{x}) = 0 \qquad [5]$$

which makes its "implicit" nature for the updated variable $\tilde{\boldsymbol{x}}$ more transparent. Clearly, as a map from the pair $(\underset{\sim}{\boldsymbol{x}},\boldsymbol{x})$ to the pair $(\boldsymbol{x},\tilde{\boldsymbol{x}})$, it is in general a multivalued map, or a "correspondence", as it is called in the literature (Suris 2003, Veselov 1991). In order that [5] be solvable for $\tilde{\boldsymbol{x}}$, the Hessian matrix $H_{jk}=\partial^2\mathcal{L}/\partial x_j\partial y_k$ should be nondegenerate.

As will be noted shortly, the Lagrangian map [4] (or [5]) is in fact a canonical, or better a symplectic transformation on a suitably defined cotangent bundle $T^*X$ to the configuration space $X\in\mathbb{R}^N$. Namely, one defines the conjugate momentum to $\boldsymbol{x}$ as

$$\boldsymbol{p} := \nabla_2\mathcal{L}(\underset{\sim}{\boldsymbol{x}},\boldsymbol{x}) \qquad [6]$$

so that [5] can be rewritten as the following system:

$$\boldsymbol{p} = -\nabla_1\mathcal{L}(\boldsymbol{x},\tilde{\boldsymbol{x}}) \qquad [7]$$

$$\tilde{\boldsymbol{p}} = \nabla_2\mathcal{L}(\boldsymbol{x},\tilde{\boldsymbol{x}}) \qquad [8]$$

This system defines a correspondence $(\boldsymbol{x},\boldsymbol{p})\to(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{p}})$, which is indeed a "symplectic" one, as it preserves the standard symplectic form $\omega(\boldsymbol{x},\boldsymbol{p})=\sum_{j=1}^N \mathrm{d}p_j\wedge \mathrm{d}x_j$, and, of course, the associated Poisson brackets. The simplest way to recognize this property is by constructing the generating function of the corresponding canonical transformation. To this end, let us introduce

$$\mathcal{S}(\boldsymbol{x},\tilde{\boldsymbol{p}}) = -\mathcal{L} + \sum_{j=1}^N \tilde{p}_j(\tilde{x}_j - x_j) \qquad [9]$$

The discrete Euler–Lagrange equation then takes the form

$$\tilde{x}_j - x_j = \frac{\partial\mathcal{S}}{\partial\tilde{p}_j} \qquad [10]$$

$$p_j - \tilde{p}_j = \frac{\partial\mathcal{S}}{\partial x_j} \qquad [11]$$

which is canonically generated by $\mathcal{S}+\sum_j x(j)\tilde{p}(j)$. A strict analog of the Hamiltonian formulation for continuous-time Lagrangian systems does not indeed exist in the discrete-time case. One of the main consequences, well known to the specialists but worth emphasizing in the present context, is that even a symplectic map in one degree of freedom

(two-dimensional $T^*X$) is generically not integrable: the existence of an invariant function $F(\boldsymbol{x},\boldsymbol{p}) = F(\tilde{\boldsymbol{x}},\tilde{\boldsymbol{p}})$ is not entailed by the symplectic structure, so that, as discussed later, integrable maps of the standard type are indeed exceptional. On the other hand, note that invariant functions do exist whenever a Lagrangian has some additional symmetry: this is the case when a Lie group acts on the configuration space $X$ and the Lagrange function is invariant under its induced action on $X \times X$, so that a discrete version of the Noether theorem applies (Suris 2003).

## Complete Integrability

The definition of a "completely integrable" discrete-time system is now in order. Let $\Phi$ be a symplectic map on the $2N$-dimensional phase space $\mathcal{M} := (\mathbb{R}^{2N}, \mathrm{d}p \wedge \mathrm{d}q)$, equipped with $N$ smooth invariant functions $F_j$, such that

- $F_1, \ldots, F_N$ are functionally independent, that is, their gradients $\nabla F_j$ are linearly independent of $\mathcal{M}$;
- $F_1, \ldots, F_N$ are in involution:

$$\{F_j, F_k\} = 0, \quad j,k = 1,\ldots,N$$

Let $\mathcal{T}$ be a connected component of the common level set

$$\{(\boldsymbol{x},\boldsymbol{p}) \in \mathcal{T}:\ F_k(\boldsymbol{x},\boldsymbol{p}) = c_k,\ k = 1,\ldots,N\}$$

Then $\mathcal{T}$ is diffeomorphic to $\mathbb{T}^l \times \mathbb{R}^{N-l}$, for some $0 \le l \le N$; if $\mathcal{T}$ is compact, then it is diffeomorphic to an $N$-dimensional torus $\mathbb{T}^N$.

In the compact case, there exists an open ball $\Omega \in \mathbb{R}^N$ such that, in $\mathcal{T} \times \Omega$, there exist new canonical coordinates $(I_k, \phi_k), k = 1,\ldots,N;\ I_k \in \mathcal{T}, \phi_k \in \Omega$, the so-called action-angle coordinates, enjoying the following properties:

- the actions $I_k$ depend just on the $F_j$'s
- in action-angle coordinates the map is a linear shift on the $N$-dimensional torus:

$$\tilde{I}_k := \Phi(I_k) = I_k$$
$$\tilde{\phi}_k := \Phi(\phi_k) = \phi_k + \nu_k(I_1, I_2, \ldots, I_N)$$

Hence, in action-angle variables a completely integrable map is a canonical transformation from $(I,\phi)$ to $(\tilde{I}\,(=I), \tilde{\phi})$, whose generating function $W$ only depends on the action variables. It takes the form

$$\tilde{I}_k - I_k = 0 \qquad [12]$$

$$\tilde{\phi}_k - \phi_k = \frac{\partial W}{\partial I_k} := \frac{\partial}{\partial I_k} \sum_{j=1}^{N} \int_x^{\tilde{x}} \mathrm{d}x_j\ p_j(\boldsymbol{I},\boldsymbol{x}) \qquad [13]$$

## Integrable Maps of the Standard Type

As the simplest integrable models, first consider some highly nontrivial examples of "standard maps," that is, scalar discrete second-order difference equations of the following type (Suris 2003):

$$x_{n+1} - 2x_n + x_{n-1} = G(x_n; h) \qquad [14]$$

with $h$ a real parameter, which exhibit an invariant function, say

$$J(x_{n-1}, x_n) = J(x_n, x_n + 1) \qquad [15]$$

Clearly, [14] can serve as a discretization of the Newtonian equation:

$$\ddot{x} = f(x) \qquad [16]$$

if $\lim_{h\to 0} h^2 G(x;h)$ exists and is equal to $f(x)$.

All "standard maps" are Lagrangian, being stationary points of the discrete action:

$$S = \sum_{n\in\mathbb{Z}} \left( \tfrac{1}{2}\left[x_{n+1} - x_n\right]^2 + V(x_n; h) \right) \qquad [17]$$

with $G(x;h) = \partial V(x;h)/\partial x$. A point in the phase space is a pair $x_n, p_n = x_n - x_{n-1}$, and [14] is symplectic for $\mathrm{d}p \wedge \mathrm{d}x$, reading

$$x_{n+1} - x_n = p_{n+1} \qquad [18]$$

$$p_n - p_{n+1} = G(x_n; h) \qquad [19]$$

The corresponding generating function is given by $S = V(x;h) + (1/2)p_{n+1}^2$. Integrability of [19] means the existence of a function $F$ from $\mathcal{M}$ into itself such that

$$F(x_{n+1}, p_{n+1}) = F(x_n, p_n) \qquad [20]$$

where [15] and [20] are equivalent provided $J(x, x - y) = F(x,y)$.

Suris has found three families of functions $G$ that ensure integrability: a rational family, a trigonometric family, and a hyperbolic family. There is no room here to display the relevant formulas, nor to explain why, under natural analiticity assumptions both in $h$ and $x$, no other integrable family exists. However, it is worth mentioning that they turn out to be integrable discretizations of the scalar second-order differential equations [16] for the following "force" functions $f(x)$:

$$f_{\mathrm{rat}}(x) = A + Bx + Cx^2 + DX^3 \qquad [21]$$

$$f_{\mathrm{trig}}(x) = A\sin(\omega x) + B\cos(\omega x) + C\sin(\omega 2x) + D\cos(\omega 2x) \qquad [22]$$

$$f_{\mathrm{hyp}}(x) = A\exp(x) + B\exp(-x) + C\exp(2x) + D\exp(-2x) \qquad [23]$$

A curious fact is that those Newton forces that one can "discretize" in order to get integrable maps

are exactly the external forces that one can add to the internal two-body interactions of the Calogero–Moser or Calogero–Sutherland models to preserve complete integrability.

## Integrable Discrete Systems and the Lax Approach

Since, in a seminal paper, Lax (1968) introduced it for the Korteweg–de Vries (KdV) equation, the search for a "Lax representation" played a crucial role in the construction of integrable systems, both finite and infinite dimensional. In particular, the continuous time dynamical system [1] (assumed to be autonomous) is said to be equipped with a Lax representation if there exist two matrices $L$, $M$ whose entries depend upon the coordinates $x_j$, whenceforth upon the time $t$, such that the time evolution [1] can be cast in the form

$$\dot{L}(t) = [L(t), M(t)] \qquad [24]$$

Hence, the one-parameter family of matrices $L(t)$ undergoes the "isospectral" deformation:

$$L(t) = U(t)L(0)(U(t))^{-1} \qquad [25]$$

$U(t)$ being the unique solution of the linear matrix differential equation:

$$\dot{U}(t) = M(t)U(t) \qquad [26]$$

with the initial condition $U(0) = I$. Then, the existence of a Lax representation in term of, say, $k \times k$ matrices entails the existence of $k$ integrals of motion, given, for instance, by the eigenvalues of $L(t)$, or by the traces $t_l := \mathrm{tr}(L(t))^l$.

Some remarks are in order:

- In the case of a Hamiltonian system, the matrices $L$, $M$ depend, of course, on the point in the phase space.
- No guarantee exists, *a priori*, that the eigenvalues of $L$, or equivalently the traces $t_l$, be "sufficiently many" and in involution. Note, however, that in many examples the Lax matrices $L$, $M$ depend on an extra scalar parameter $\lambda$ (so that they are elements of an affine or "loop" Lie algebra), which might increase the number of integrals of motion well beyond the dimension of the matrix.

The $N$-body systems of Calogero type and Toda type are celebrated examples of integrable dynamical systems equipped with a Lax representation.

How this description can be adapted to the discrete-time case? The isospectral equation [25] suggests the proper way. One has to look for two matrices depending on the coordinates (or on the phase-space variables) $\boldsymbol{x}$ (again, they can be called $L$, $M$), such that the discrete-time evolution, modeled, for instance, by [2], can be cast in the form of a similarity transformation:

$$\tilde{L} = MLM^{-1} \qquad [27]$$

where $L = L(\boldsymbol{x})$, $\tilde{L} = L(\tilde{\boldsymbol{x}})$, and $M = M(\boldsymbol{x}, \tilde{\boldsymbol{x}})$. As usual, by denoting by $n$ the discrete time (i.e., the number of iterations), so that $\boldsymbol{x} = \boldsymbol{x}(n), \tilde{\boldsymbol{x}} = \boldsymbol{x}(n+1)$, eqn [27] implies that a discrete version of [25] holds:

$$L(n) = U(n)L(0)[U(n)]^{-1} \qquad [28]$$

where $U(n) := M(n)M(n-1)\cdots M(1)$.

As in the continuous case, the existence of a discrete Lax representation entails the existence of conserved quantities (invariants of the map or of the correspondence) but by itself it does not say anything about completeness and involutivity of such invariants. There is, however, an approach that incorporates the involutivity property in the very construction of Lax equations, both discrete and continuous, namely the "$R$-matrix approach." Indeed, from the experimental observation of a number of examples, both finite and infinite dimensional, one can assert that the matrix $M$ taking part in the "continuous" Lax representation [24] may be presented in the form (Suris 2003)

$$M = R(f(L)) \qquad [29]$$

In [29], $L$, $M$ are element of some matrix Lie algebra $\boldsymbol{g}$, $R$ is a linear map from $\boldsymbol{g}$ into itself, and $f$ is a conjugation-covariant function, namely

$$f(ALA^{-1}) = Af(L)A^{-1} \qquad [30]$$

$A$ being an arbitrary element of the group $\boldsymbol{G}$ with Lie algebra $\boldsymbol{g}$.

Polynomials in the variable $L$ with scalar coefficients are typical examples of conjugation-covariant functions. Moreover, in a matrix Lie algebra, one can identify $\boldsymbol{g}$ with its dual space $\boldsymbol{g}^*$ through the nondegenerate bilinear form provided by the trace: $(L_1, L_2) := \mathrm{tr}(L_1 L_2)$. Then, the trace $F$ of a conjugation-covariant function $f$ will be a typical example of a conjugation-invariant function, and, conversely, the gradient of a conjugation-invariant function $F$, defined as

$$\langle \nabla F, X \rangle = \frac{\mathrm{d}}{\mathrm{d}\epsilon} F(L + \epsilon X)|_{\epsilon=0} \qquad [31]$$

will be a typical example of a conjugation-covariant function. In the above setting, one can define the following Lie–Poisson bracket on $\boldsymbol{g}$:

$$\{F, G\}(L) := (L, [\nabla F, \nabla G]) \qquad [32]$$

where $F$, $G$ are arbitrary (i.e., not necessarily invariant) functions from $\boldsymbol{g}$ into $\mathbb{C}$, so that the Hamilton equation

$$\dot{L} = \{H, L\} \quad [33]$$

takes the Lax form

$$\dot{L} = [L, \nabla H] \quad [34]$$

It is immediate to check that invariant functions of $L$ are Casimir functions of [32] so that they will not generate any nontrivial flow.

Assume now that the linear mapping $R$, usually called $r$-matrix, introduced in [29], is such that it defines a new Lie bracket on $\boldsymbol{g}$, through the formula

$$[L_1, L_2]_R = \tfrac{1}{2}([L_1, R(L_2)] + [R(L_1), L_2]) \quad [35]$$

and consequently a new Lie–Poisson bracket

$$\{F, G\}_R(L) := (L, [\nabla F, \nabla G]_R) \quad [36]$$

Then the following theorem holds:

*Let $H$ be an invariant function on $\boldsymbol{g}$. Then*:

(i) *The Hamilton equations on $\boldsymbol{g}$ generated by $H$ with respect to the Poisson bracket [36] have the Lax form*

$$\dot{L} = [L, R(\nabla H)] \quad [37]$$

(ii) *The invariants of $\boldsymbol{g}$, that is, the Casimir function of the standard Lie–Poisson bracket [32], are in involution for [36] so that the corresponding flows are mutually commuting.*

A particular realization of such $R$ operator, very important for the application, arises in the so-called Adler–Kostant–Symes (AKS) construction (Adler 1979, Kostant 1979, Symes 1980), where the Lie algebra $\boldsymbol{g}$ admits a decomposition in two subalgebras, $\boldsymbol{g}_+$ and $\boldsymbol{g}_-$, so that, as linear spaces, it holds that

$$\boldsymbol{g} = \boldsymbol{g}_+ \oplus \boldsymbol{g}_- \quad [38]$$

Denoting by $\pi_\pm$ the corresponding projections, the linear mapping

$$R := \pi_+ - \pi_- \quad [39]$$

defines a new Lie bracket on $\boldsymbol{g}$, and the corresponding Lax equations take the two equivalent forms:

$$\dot{L} = [L, \pi_+(f(L))] = -[L, \pi_-(f(L))] \quad [40]$$

For the present purposes, it is of paramount importance that the AKS construction has a discrete-time version (Suris 2003).

In fact, let $\boldsymbol{G}$ be a Lie group with Lie algebra $\boldsymbol{g}$, and let $\boldsymbol{G}_+, \boldsymbol{G}_-$ be its subgroups having $\boldsymbol{g}_+, \boldsymbol{g}_-$ as Lie algebras. Then, in a certain component of the identity element $I$, any element $g$ of $\boldsymbol{G}$ is uniquely factorizable as

$$g = \Pi_+(g)\Pi_-(g), \quad \Pi_\pm(g) \in \boldsymbol{G}_\pm \quad [41]$$

Moreover, let $F: \boldsymbol{g} \to \boldsymbol{G}$ be a conjugation-covariant function. Consider now the map

$$\begin{aligned} L \to \tilde{L} &:= \Pi_+^{-1}(F(L)) \cdot L \cdot \Pi_+(F(L)) \\ &= \Pi_-(F(L)) \cdot L \cdot \Pi_-^{-1}(F(L)) \end{aligned} \quad [42]$$

and regard it as a difference equation, yielding $\tilde{L} = L(n+1)$ in terms of $L = L(n)$. Then, the following properties hold:

- For whatever function $F$, the map [42] commutes with any continuous flow [40], mapping solutions into solutions.
- It can be "explicitly integrated" with respect to the discrete time $n$, yielding

  $$L(n) = \Pi_+^{-1}(F^n(L_0)) \cdot L_0 \cdot \Pi_+(F^n(L_0)) \quad [43]$$

  or the equivalent expression in terms of the complementary projection $\Pi_-$.
- It is interpolated by the continuous flow [40] with time step $h$ if

  $$\exp(hf(L)) = F(L) \leftrightarrow f(L) = h^{-1} \log(F(L)) \quad [44]$$

  In other words, the discrete-time systems that one derives through this approach are just a sequence of pictures taken at equally spaced times of some continuous flow pertaining to the hierarchy [40]: so, by construction they are Poisson maps with an involutive family of integrals given by the conjugation-invariant functions of $L$ (typically, $\operatorname{tr} L^n$).
- As far as

  $$F(L) = I + hf(L) + o(h^2) \quad [45]$$

  the map [42] serves as an integrable exact discretization of the flow [40], sharing both its Poisson structure and its constants of the motion.

## An Integrable Discretization of the Toda Lattice

Consider a simple but an illuminating example of the above construction, showing an integrable discretization of the "open-end Toda lattice," which is described (Suris 2003) by the Newtonian equations of motion:

$$\begin{aligned} \ddot{x}_j &= \exp(x_{j+1} - x_j) - \exp(x_j - x_{j-1}) \\ j &= 1, \ldots, N \end{aligned} \quad [46]$$

and can be cast into a Hamiltonian form by setting $p_j = \dot{x}_j$; $q_j = x_j$. If, according to H Flaschka (1974), one introduces the variables

$$b_j = \dot{x}_j, \qquad a_j = \exp(x_{j+1} - x_j) \qquad [47]$$

eqn [46] takes the form

$$\dot{b}_j = a_j - a_{j-1}, \qquad \dot{a}_j = a_j(b_{j+1} - b_j) \qquad [48]$$

and enjoys the Lax representation [24] in terms of the $N \times N$ matrices:

$$L(a,b) = \sum_{k=1}^{N} a_j E_{j,j+1} + \sum_{k=1}^{N} b_j E_{j,j} + \sum_{k=1}^{N} E_{j+1,j} \qquad [49]$$

$$M(a,b) = A := \sum_{k=1}^{N} a_j E_{j,j+1} \quad \text{or}$$
$$M(a,b) = -B := \sum_{k=1}^{N} b_j E_{j,j} + \sum_{k=1}^{N} E_{j+1,j} \qquad [50]$$

In the above formula, $E_{j,k}$ is the matrix having 1 in the $jk$ position and 0 elsewhere, so that, obviously, $E_{N,N+1} = E_{N+1,N} = 0$. An inspection to [49] and [50] shows that $A$ is just the strictly upper triangular part of $L(a,b)$, while $B$ is its lower triangular part. The pair $(A,B)$ constitutes the so-called $LU$ decomposition of $L(a,b)$. One is clearly in the AKS setting, the Lie algebra $\boldsymbol{g}$ being just the algebra of $N \times N$ matrices, and the Lie subalgebras $\boldsymbol{g}_\pm$ being the strictly upper and lower triangular matrices. The tridiagonal matrix $L(a,b)$ belongs to a Poisson submanifold of $\boldsymbol{g}$, invariant under the flows [40], and a complete family of commuting integrals of motion is given, for instance, by $I_k = \mathrm{tr} L^k$.

Now, the elements of the group $\mathrm{GL}_N$, realized as the group of invertible $N \times N$ matrices, uniquely factorize into a product of an invertible lower-triangular matrix times an upper-triangular matrix with units on the diagonal, and the Lie algebras of those subgroups are just the aforementioned subalgebras $\boldsymbol{g}_\pm$. Then, one is naturally tempted to look for an integrable discretization provided by a conjugation-covariant function of the type [45], starting with the simplest possible choice, namely

$$F(L) = I + hf(L)$$

Setting

$$\begin{aligned}\tilde{L}(a,b) &:= L(\tilde{a},\tilde{b})\\ &= \Pi_+^{-1}(I+hL)\cdot L\cdot \Pi_+(I+hL)\\ &= \Pi_-(I+hL)\cdot L\cdot \Pi_-^{-1}(1+hL) \qquad [51]\end{aligned}$$

it turns out that the matrix equation [51] is equivalent to the map

$$(a,b) \to (\tilde{a},\tilde{b})$$

described by the following equations:

$$\tilde{b}_k = b_k + h\left(\frac{a_k}{\beta_k} - \frac{a_{k-1}}{\beta_{k-1}}\right)$$
$$\tilde{a}_k = a_k(\beta_{k+1} - \beta_k)$$

where $\beta_k$, which are the "field variables" entering into the $LU$ factorization [51], are explicitly and uniquely defined by the recurrent relation (amounting to a finite continued fraction):

$$\beta_k = 1 + hb_k - h^2\frac{a_{k-1}}{\beta_{k-1}}, \quad k = 1,\ldots,N \qquad [52]$$

As $a_0 = 0$, the initial condition is simply $\beta_1 = 1 + hb_1$.

It follows from the general results of the previous section that [51] is an integrable Poisson map, sharing with the continuous Toda hierarchy both the Poisson structure and the integrals of motion. Its initial-value problem can be uniquely solved in terms of the $LU$ factorization of the group element $(I + hL_0)^n$, the initial condition $L_0$ being any matrix pertaining to the tridiagonal submanifold [49]. According to [44], the interpolating Hamiltonian flow is provided by the function $f(L) = h^{-1}\log(1 + hL)$. To make contact with the discussion in the section "Lagrangian and Hamiltonian formulations," we observe that, in terms of the canonical variables $x_j$, $p_j$, the discrete Toda [51] lattice becomes the following symplectic map:

$$1 + hp_j = \exp(\tilde{x}_j - x_j) + h^2\exp(x_j - \tilde{x}_{j-1}) \qquad [53]$$

$$1 + h\tilde{p}_j = \exp(\tilde{x}_j - x_j) + h^2\exp(x_{j+1} - \tilde{x}_j) \qquad [54]$$

It can evidently be written in the discrete Newtonian form:

$$\begin{aligned}&\exp(\tilde{x}_j - x_j) - \exp(x_j - \underset{\sim}{x}_j)\\ &\quad = h^2\exp(\underset{\sim}{x}_{j+1} - x_j) - \exp(x_j - \tilde{x}_{j-1}) \qquad [55]\end{aligned}$$

whose Lagrangian function is given by

$$\mathcal{L} = \sum_{k=1}^{N}\psi(\tilde{x}_k - x_k) - h\sum_{k=1}^{N}\exp(x_{k+1} - \tilde{x}_k) \qquad [56]$$

with

$$\psi(\xi) = h^{-1}(\exp(\xi) - 1 - \xi) \qquad [57]$$

The variables $\beta_j$ acquire the following extremely simple expression in the Lagrangian coordinates $x_j, \tilde{x}_j$:

$$\beta_j = \exp(\tilde{x}_j - x_j)$$

For integrable Hamiltonian systems with long-range two-body interaction, such as Calogero–Moser type systems, and their so-called relativistic version (Ruijsenaars systems), an exact integrable discretization has also been found. However, at least

in the more natural Lax representation, the related $R$-matrix is dynamical (namely, it depends on the phase-space coordinates), and the simple factorization scheme holding for the Toda lattice system (and for the related ones) is not available.

Further knowledge on the intriguing subject of "discrete integrable systems" can be acquired by looking at the monographs and papers listed in the "Further Reading" section. In particular, the excellent book by Y B Suris, which also provides an exhaustive list of references (updated to 2003), is recommended.

*See also:* Billiards in Bounded Convex Domains; Boundary Value Problems for Integrable Equations; Calogero–Moser–Sutherland Systems of Nonrelativistic and Relativistic Type; Integrable Systems and Discrete Geometry; Integrable Systems and the Inverse Scattering Method; Integrable Systems: Overview; Painlevé Equations; Quantum Calogero–Moser Systems; Toda Lattices; Yang–Baxter Equations.

## Further Reading

Adler M (1979) On a trace functional for formal pseudo-differential operators and symplectic structures of the Korteweg–de Vries type equations. *Inventiones Mathematicae* 50: 219–248.

Grammaticos B, Kossmann-Shwarzbach Y, and Tamizhmani T (eds.) (2004) *Discrete Integrable Systems*, Lecture Notes in Physics, vol. 644. Berlin: Springer.

Kostant B (1979) The solution to the generalized Toda lattice and representation theory. *Advances in Mathematics* 34: 195–338.

Lax P (1968) Integrals of non-linear equations of evolution and solitary waves. *Communications in Pure and Applied Mathematics* 21: 467–490.

Suris YB (2003) *The Problem of Integrable Discretization: Hamiltonian Approach*, Progress in Mathematics, vol. 219. Basel: Birkhauser Verlag.

Symes W (1980) Systems of Toda type, inverse spectral problems and representation theory. *Inventiones Mathematicae* 59: 13–53.

Veselov A (1991) Integrable maps. *Russian Mathematical Surveys* 46: 1–51.

# Integrable Systems and Algebraic Geometry

**E Previato**, Boston University, Boston, MA, USA



## Historical Overview

The relevance of algebraic geometry in the theory of dynamical systems has a long history. Three models may serve as guiding threads from old to the current state of the theory. Each time algebraic geometry is used to integrate an evolution equation; this is achieved by an underlying addition rule. The very origin for this seems to be Fagnano's addition rule for the arc of a lemniscate (see Siegel (1969)). In analogy to the addition of two arcs on a circle $x^2+y^2=1$, or the duplication formula for

$$\arcsin r = \int_0^r \frac{\mathrm{d}r}{\sqrt{1-r^2}}$$

namely

$$\int_0^r \frac{\mathrm{d}r}{\sqrt{1-r^2}} = 2\int_0^u \frac{\mathrm{d}u}{\sqrt{1-u^2}}$$

if $r=2u\sqrt{1-u^2}$ (a restatement of the trigonometric identity $r=\sin(2x)=2\sin x\cos x$), Fagnano found, and proved, by substitution, a geometric rule for duplicating the arc of a lemniscate:

$$x^4+2x^2y^2+y^4=x^2-y^2$$

The length of the arc is now given by

$$s=\int_0^r \frac{\mathrm{d}r}{\sqrt{1-r^4}}$$

and later Gauss designated the limit of integration by $r=\operatorname{sinlemn}(s)$. Fagnano was able to show that

$$\int_0^r \frac{\mathrm{d}r}{\sqrt{1-r^4}} = 2\int_0^u \frac{\mathrm{d}u}{\sqrt{1-u^4}}$$

with the substitution

$$r^2=\frac{4u^2(1-u^4)}{(1+u^4)^2}$$

which is remarkable not only because it doubles the length, but also because it does so by rational functions, and in fact shows that the arc of the lemniscate can be halved by straightedge and compass. Gauss showed that the constructible fractions of an arc of a lemniscate are the same as the ones for the circle.

Thanks to subsequent work by Euler, and to the theory of abelian functions due to Abel, Jacobi, and others in the nineteenth century, we now realize that Fagnano's discovery revealed the algebraic group structure of the singular quartic curve (or of a smooth cubic, if preferred, an elliptic curve).

This is the key fact that provides the "integration by quadratures" for the simple pendulum. We follow McKean and Moll (1997) to sketch this prototype example of a system which is algebraically completely integrable (ACI), defined in the section

"Hitchin systems." Newton's law gives the equation of motion $\ddot{\theta} + \sin\theta = 0$, where $\theta$ parametrizes the position of the bob in terms of the angle the pendulum makes with the vertical axis, as it rotates about its pivot (the length has been normalized so as to match the gravitational constant). The energy is a first integral, $I = \cos\theta - 1/2\dot{\theta}^2$, and the substitution

$$x = \sqrt{\frac{2}{1-I}} \times \sin\frac{\theta}{2}$$

linearizes the motion:

$$t = \int_0^x \frac{1}{\sqrt{(1-x^2)(1-k^2x^2)}}\,\mathrm{d}x$$

with $k^2 = (1-I)/2$ between 0 and 1, precisely because of Fagnano's and Euler's addition rule.

The second striking example of addition rule, yielding solutions to a nonlinear partial differential equation (PDE), together with this first will provide the two themes of this article, and embed into an infinite-dimensional family of conservation laws that will accommodate the representation-theoretic aspect of the symmetries. In their 1895 article, Korteweg and de Vries (KdV) gave official status to the (then controversial) representation of solitary waves in shallow water:

$$u_t = 6uu_x - u_{xxx}$$

(again up to normalization) is by now the well-known KdV equation, where $u$ represents the amplitude of the wave and $x$ the direction along a canal. It so happens that by integrating twice the ordinary differential equation (ODE) obtained by the one-wave ansatz, $z = x - ct$ (where $c$ is the constant velocity), one sees that the solution $u$ and its derivative $u_z = u'$ satisfy identically an algebraic equation:

$$\begin{aligned} -cu' - 6uu' + u''' &= 0 \\ (-cu - 3u^2 + u'' + a)u' &= 0 \\ \frac{(u')^2}{2} &= u^3 + c\frac{u^2}{2} - au + b \end{aligned}$$

$$u = 2\wp + \text{const.} \quad \text{(up to a linear transformation)}$$
$$(\wp')^2 = 4\wp^3 - g_2\wp - g_3 = 4(\wp - e_1)(\wp - e_2)(\wp - e_3)$$

In disguise, then, the PDE and the Hamiltonian evolutions are the same; the motion becomes linear (and quasiperiodic) on the torus $\mathbb{C}/\Lambda$, where $\Lambda$ is the period lattice of the $\wp$ function. It took considerably greater effort to generalize this correspondence to higher genus. This article is devoted to such a correspondence as well as some of the surprising connections between complete integrability and other areas of mathematics such as: representation theory (the corresponding geometric objects are Grassmann manifolds as opposed to Jacobians); differential algebras (Weyl algebras, commutative rings of differential operators, and differential Galois theory); and reduction in symplectic geometry.

It is often helpful to highlight the relevant features in the simplest example, even if it is of special kind. The KdV equation and, as Hamiltonian counterpart, Neumann's system (see Neumann (1859)) will serve best. The abelian sum identified by Fagnano cannot be defined on points of a curve $X$ of genus $g > 0$; what one can add are points of the $g$-fold symmetric product $X^{(g)}$ up to linear equivalence, defining (up to noncanonical isomorphism) an abelian variety, the Jacobian $\mathrm{Jac}(X) = \mathbb{C}^g/\Lambda$; analytically, the Jacobian is described by abelian coordinates $z_1, \ldots, z_g$: if $\alpha_1, \ldots, \alpha_g, \beta_1, \ldots, \beta_g$ is a basis of 1-cycles on $X$ with standard intersection matrix and $\omega_1, \ldots, \omega_g$ is the dual basis of holomorphic differentials, then $z_j = \sum_{i=1}^{g} \int_{P_0}^{P_i} \omega_j$ is defined in terms of a fixed base point $P_0 \in X$ and of $(P_1, \ldots, P_g) \in X^{(g)}$ up to the period lattice $\Lambda$. It is in these coordinates that the Hamiltonian flows become linear. In canonical coordinates $q_1, \ldots, q_{g+1}, p_1, \ldots, p_{g+1}$, the harmonic oscillator

$$\begin{aligned} \dot{q}_i &= p_i \\ \dot{p}_i &= -e_i q_i \end{aligned}$$

when constrained to the unit sphere $\sum_{i=1}^{g+1} q_i^2$ has equations

$$\begin{aligned} \dot{q}_i &= p_i \\ \dot{p}_i &= -e_i q_i + q_i \sum_j (e_j q_j^2 - p_j^2) \end{aligned}$$

This system is completely integrable in the sense that there exist enough involutory invariants, $g$ generically (in the $(q_i, p_i)$ variables) independent functions on the $2g$-dimensional tangent bundle of the unit sphere with canonical symplectic structure; in fact the coefficients of the polynomial

$$f(\lambda) = \prod_{i=1}^{g+1} (\lambda - e_i)^2 \left( \left( \sum_{k=1}^{g+1} \frac{q_k^2}{\lambda - e_k} \right) \left( \sum_{k=1}^{g+1} \frac{p_k^2}{\lambda - e_k} + 1 \right) - \left( \sum_{k=1}^{g+1} \frac{q_k p_k}{\lambda - e_k} \right)^2 \right)$$

are invariant and the hyperelliptic Riemann surface $X$ whose model in the affine plane is given by $\mu^2 = f(\lambda)$ is called the spectral curve of the system. Since the polynomial $f(\lambda)$ is monic of degree $2g+1$ and has generically simple roots, $X$ has

genus $g$. A change of variables permits integration by quadratures,

$$q_i(t) = \frac{\vartheta[\eta_{2i-1}](0)\vartheta[\eta_{2i-1}](z_0 - \Delta + 2\sqrt{-1}tU)}{\vartheta[0](0)\vartheta[0](z_0 - \Delta + 2\sqrt{-1}tU)}$$

where $z_0$, $U \in \mathbb{C}^g$ are constant vectors, $\vartheta$ denotes the Riemann theta function of $X$, $\eta_k (k = 1, \ldots, 2g)$ are theta characteristics and $\Delta$ is the Riemann constant. While these are technical objects of classical Riemann function theory whose detailed definition is best found in a textbook (see, e.g., Mumford (1984)), the point here is that the motion is linearized along the line with direction $U$, on the hyperelliptic Jacobian Jac($X$), which is a $2^{g+1} : 1$ cover of the phase space.

A yet deeper fact links the integrable Hamiltonian motion and the (soliton) PDE, namely the statement that $\sum_{i=1}^{g+1}(e_i q_i^2 + p_i^2) = u(t_1, t_3)$ solves the KdV equation, where the variables are renamed as $x = t_1, t = t_3$ to denote two of the $g$ commuting Hamiltonian flows.

The Neumann system as well allows us to uncover another deep relation between dynamics and geometry, namely the moduli aspect: on the one hand, Mumford (1984) used the Neumann system to recover the equation of the spectral curve from a vanishing property of theta functions with characteristics, thereby giving the first characterization of the moduli subvariety of hyperelliptic curves in terms of thetanulls (for any genus). On the other hand, Françoise (1987) explored the relevance of the integrable system to the Picard–Fuchs equations. The fundamental link is provided by Arnol'd's theory, according to which a set of action-angle variables $(q_i, p_i)$, $i = 1, \ldots, n$, for a completely integrable Hamiltonian system can be calculated in terms of a basis $\gamma_i$ of the first homology of the fibers, which are $n$-dimensional tori, $\int_{\gamma_j} \mathrm{d}q_i = \delta_{ij}$; hence, in the case of an algebraically integrable system such as the Neumann example (or, in Françoise's paper, the Kowalevski top), in principle one can express the (coefficients of the) differential equations satisfied by the periods in terms of the commuting Hamiltonians, despite the fact that periods and Hamiltonians are transcendental functions the ones of the others. A more general family of period matrices is subject to the Gauss–Manin connection, and the question of whether its general abelian variety is Lagrangian with respect to a holomorphic symplectic structure on the family yields a cubic condition on the periods (Donagi and Markman 1996).

These are two major applications of PDEs to algebraic geometry: characterizing subvarieties of moduli spaces (of curves) and expressing the Gauss–Manin connection acting on sections of a Hodge-theoretic bundle over the moduli space in terms of the evolution equations of a completely integrable system. In the former case, the flows of the system act on the theta functions of a (fixed) curve; in the latter, the Hamiltonians are related, via the action variables, to computing the monodromy over the branch points of the base of the system. The generalization of specific (e.g., hyperelliptic) cases is very difficult to work out and remains largely open 40 years after the field of integrable equations started being actively investigated.

Before concluding this historical overview, a beautiful theory that escaped attention is worth mentioning. In the late nineteenth century, for example, Baker (1907) constructed the first genus-2 solutions of the KdV equation, although he was apparently not aware of the equation itself; in the process, he also defined what is known as the Hirota bilinear operator, a device introduced by R Hirota in the 1970s to capture an equivalent version of the KdV, or the more general Kadomtsev–Petviashvili (KP) equation,

$$(u_t - 6uu_x + u_{xxx})_x = u_{yy}$$

Just as the Lax pair allows for a linearization of the isospectral deformations, Hirota's bilinear form reveals the representation-theoretic (and algebro-geometric) nature of the equations, via the vanishing of a natural pairing on a pair of solutions, besides providing a formula for exact solutions; the definition of the bilinear operation is the following: for functions $F$ and $G$,

$$D_{t_n} F \cdot G = \left(\frac{\partial}{\partial t'_n} - \frac{\partial}{\partial t_n}\right) F(\underline{t}) G(\underline{t}') \big|_{\underline{t} = \underline{t}'}$$
$$\underline{t} = (t_1, t_2, \ldots)$$

so that Hirota's direct method gives the following solution: set $u = 2(\partial^2/\partial x^2) \log F$, then

$$\text{KdV} \Leftrightarrow \left(D_x D_t + D_x^4\right) F \cdot F = 0$$
$$\text{KP} \Leftrightarrow D_x^2 \frac{\left(D_x^4 + 3D_y^2 - 4D_x D_t\right) F \cdot F}{2F^2} = 0$$

Baker was intent on generalizing the properties of the Weierstrass $\wp$ function. He focused on genus 2 (and obtained partial results for general genus), in which case any curve is hyperelliptic,

$$f : \mu^2 = \lambda^{2g+1} + a_{2g}\lambda^{2g} + \cdots + a_0$$

and used a suitable basis of holomorphic differentials particular to the hyperelliptic case, whose

integrals give abelian coordinates $z_i$ that happen to be dual to the KdV flows,

$$\omega_1 = \frac{d\lambda}{2\mu},\ \omega_2 = \frac{\lambda d\lambda}{2\mu},\ \ldots,\ \omega_g = \frac{\lambda^{g-1}d\lambda}{2\mu}$$

to characterize the genus-2 theta function by differential equations (equivalent to the KdV hierarchy), as well as give the quartic equation for the Kummer surface in $\mathbb{P}^3$, namely the 2:1 image of the Jacobian of the curve mapped by the divisor $2\Theta$, that is, by a basis of the space of theta functions with second-order characteristics, simply as the determinant of

$$\begin{bmatrix} -a_0 & \frac{1}{2}a_1 & 2\wp_{11} & -2\wp_{12} \\ \frac{1}{2}a_1 & -(a_2+4\wp_{11}) & \frac{1}{2}a_3+2\wp_{12} & 2\wp_{22} \\ 2\wp_{11} & \frac{1}{2}a_3+2\wp_{12} & -(a_4+4\wp_{22}) & 2 \\ -2\wp_{12} & 2\wp_{22} & 2 & 0 \end{bmatrix}$$

where

$$\wp_{ij}(z) = -\frac{\partial^2}{\partial z_i \partial z_j}\log\sigma(z)$$

and the $\sigma$ function, defined in analogy to the genus-1 case, is proportional to the Riemann theta function.

To summarize this introduction, the exchange between algebraic geometry (the classification of algebraic varieties) and dynamical systems has been extremely fruitful in either direction: algebraic geometry surprisingly provides exact solutions to evolution equations that have special algebraic symmetries (and arise in nature!), and conversely those very evolutionary equations yield the structure of particularly complicated varieties, by characterizing their (rational) functions.

## Isospectral Deformations

The isospectral deformations in question have been encoded by Lax-pair equations, which take their name from Peter D Lax, who gave a version of the KdV equation in such form.

Lax pairs enter in two essentially different ways in the theory of integrable systems. The evolution equations take the form: $\partial_{t_n}L=[B,L]$, where $t_1,t_2,t_3,\ldots$ is a sequence of commuting time flows, $L$ is an operator whose coefficients depend on time, and $B$ is another operator of the same kind; since heuristically this is the infinitesimal version of the equation $L(t)=U(t)^{-1}L(0)U(t)$ (with $B=U^{-1}\partial_t U$), the spectrum of $L$ is preserved and provides conserved quantities; in fact, Moser (1980) speculated that every completely integrable system might have such a form.

In the form that immediately yields a hierarchy of PDEs, the (hierarchy of) deformations pertain to a ring of (formal) pseudodifferential operators, where the variable $x=t_1$ is singled out and $\partial$ denotes differentiation with respect to $x$:

$$L(t) \in \mathcal{D} = \left\{ \sum_{j=0}^{n} u_j(x)\partial^j, u_j \text{ analytic near } x=0 \right\}$$

$$\subset \mathcal{P} = \left\{ \sum_{-\infty}^{n} u_j(x)\partial^j \right\}$$

The multiplication rule that makes $\mathcal{P}$ into a ring (in fact, a $\mathbb{C}$-algebra) is composition:

$$\partial \circ u = u\partial + u'$$
$$\partial^{-1} \circ u = u\partial^{-1} - u'\partial^{-2} + u''\partial^{-3} - \cdots$$

We normalize $L$ by an automorphism of $\mathcal{D}$ (generated by a change of variable and conjugation by a function)

$$L = \partial^n + u_{n-2}(x)\partial^{n-2} + \cdots + u_0(x)$$

In $\mathcal{P}$ any (normalized) $L$ has a unique $n$th root, $n=\operatorname{ord} L$, of the form $\mathcal{L}=\partial+u_{-1}(x)\partial^{-1}+u_{-2}(x)\partial^{-2}+\cdots$. Finally, the deformation equations,

$$\partial_{t_n}\mathcal{L} = [(\mathcal{L}^n)_+, \mathcal{L}]$$

define the KP hierarchy, which takes its name from the first nontrivial deformation equation, known as the KP equation encountered above, if we set $x=t_1, y=t_2, t=t_3$ (notice that this reduces to KdV, up to rescaling, when the solution is independent of $y$). The algebro-geometric solutions are those with the property that only a finite number of time evolutions are independent. This turned out to be equivalent to a classical problem of elementary differential algebra, known as the Burchnall–Chaundy problem after the two co-authors who solved it in the 1920s.

The Burchnall–Chaundy problem: which $L(x)$'s have centralizer $\mathcal{C}_{\mathcal{D}}(L)$ that is larger than a polynomial ring $\mathbb{C}[L_1]$, $L_1 \in \mathcal{D}$? The key to the solution is the following fact (which clearly does not hold for operators in more than one variable, or finite-dimensional operators such as matrices): if $\operatorname{ord} L>0$ and $A,B\in\mathcal{D}$ both commute with $L$, then $[A,B]=0$; in particular, $\mathcal{C}_{\mathcal{D}}(L)$ is commutative, hence every maximal-commutative subalgebra of $\mathcal{D}$ is a centralizer. It was proved in the early 1900s by I Schur that $\mathcal{C}_{\mathcal{D}}(L)=\{\sum_{-\infty}^{N} c_j\mathcal{L}^j, c_j\in\mathbb{C}\}\cap\mathcal{D}$. It follows that centralizers are rings of affine curves: their transcendence degree over the field of coefficients is 1, and $\operatorname{Spec}\mathcal{C}(L)$ can be regarded as an affine curve $X_0$ (with natural compactification

$X$ by a smooth point at infinity). Burchnall and Chaundy proceeded to show that the rings of operators whose orders are not all multiples of a fixed integer >1, and having the same spectral curve $X$ (up to isomorphism), correspond to line bundles over $X$ (more precisely, rank-1 torsion-free sheaves); thus, the hierarchy of evolutions linearizes on Jac $X$, as indicated by the examples treated above.

In this setting, it has been very challenging to generalize the integrable flows, both to the higher-rank and to the higher-dimensional case. When all the operators in the commutative ring have order divisible by an integer $r > 1$, their common kernel defines a rank-$r$ vector bundle over the spectral curve, and although the theory in principle is similar to the case of line bundles, there are no explicit formulas for solution. On the other hand, in order that the spectrum be a variety $X$ of dimension $d > 1$ rather than a curve, it is natural to seek commutative rings of partial differential operators in $d$ variables; but again, while some constructions work in principle, explicit formulas are elusive.

The form in which Lax pairs occur for finite-dimensional Hamiltonian systems is quite different: here what is preserved is the spectrum of a finite-dimensional linear operator, a matrix. The first examples, from which the theory took off, were inspired guesses. The Neumann system described above fits in the following theory: Moser (1980) showed that the Neumann system together with other important classical examples are special cases of rank-2 perturbations (since $(2 = \dim\langle \boldsymbol{p}, \boldsymbol{q}\rangle)$) which preserve the spectrum of a matrix

$$L = A + a\boldsymbol{q} \otimes \boldsymbol{q} + b\boldsymbol{q} \otimes \boldsymbol{p} + c\boldsymbol{p} \otimes \boldsymbol{q} + d\boldsymbol{p} \otimes \boldsymbol{p}$$

where $A$ is a fixed constant matrix which can be normalized to a diagonal, $\mathrm{diag}(e_1, \ldots, e_{g+1})$,

$$\det\begin{bmatrix} a & b \\ c & d \end{bmatrix} \neq 0$$

and $\boldsymbol{u} \otimes \boldsymbol{v}$ denotes the matrix $[u_i v_j]$. The symplectic structure is the standard $\omega = \sum \mathrm{d}p_i \wedge \mathrm{d}q_i$ so that a Hamiltonian $H$ defines a flow

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i}$$

and

$$\frac{\partial G}{\partial t} = -\{H, G\} = \sum \frac{\partial H}{\partial p_i}\frac{\partial G}{\partial q_i} - \frac{\partial G}{\partial p_i}\frac{\partial H}{\partial q_i}$$

The Hamiltonian flow of

$$H = \frac{1}{2}\left( a\langle B\boldsymbol{q}, \boldsymbol{q}\rangle + (b + c)\langle B\boldsymbol{q}, \boldsymbol{p}\rangle + d\langle B\boldsymbol{p}, \boldsymbol{p}\rangle - \frac{ad - bc}{2}\sum_{i \neq j} \frac{b_i - b_j}{e_i - e_j}(q_i p_j - q_j p_i)^2 \right)$$

(where $B = \mathrm{diag}(b_1, \ldots, b_{g+1})$ is any fixed diagonal matrix) is equivalent to the Lax-pair equation $\dot{L} = [M, L]$, where $M$ is a suitable matrix:

$$M = \frac{1}{2}(b - c)[b_i \delta_{ij}] + (ad - bc)\left[\frac{b_i - b_j}{e_i - e_j}(q_i p_j - q_j p_i)\right]$$

The Weinstein–Aronszajn formula

$$\det\left( I_n - \sum_{i=1}^{r} \xi_i \otimes \eta_i \right) = \det\Big( I_r - [\langle \xi_i, \eta_j\rangle] \Big)$$

(where each of the $\xi_1, \ldots, \xi_r, \eta_1, \ldots, \eta_r$ is a $(g + 1 = n)$-vector) gives for the spectral invariants

$$\begin{aligned}\frac{l(\lambda)}{e(\lambda)} &= \frac{\det(\lambda - L)}{\det(\lambda - A)} \\ &= \det(I - ((\lambda - A)^{-1}\boldsymbol{q}) \otimes (a\boldsymbol{q} + b\boldsymbol{p}) \\ &\quad - ((\lambda - A)^{-1}\boldsymbol{p}) \otimes (c\boldsymbol{q} + d\boldsymbol{p})) \\ &= \det(I_2 - W_\lambda(\boldsymbol{q}, \boldsymbol{p}))\end{aligned}$$

with

$$W_\lambda(\boldsymbol{q}, \boldsymbol{p}) = \begin{bmatrix} \langle (\lambda - A)^{-1}\boldsymbol{q}, \boldsymbol{q}\rangle & \langle (\lambda - A)^{-1}\boldsymbol{q}, \boldsymbol{p}\rangle \\ \langle (\lambda - A)^{-1}\boldsymbol{q}, \boldsymbol{p}\rangle & \langle (\lambda - A)^{-1}\boldsymbol{p}, \boldsymbol{p}\rangle \end{bmatrix} \times \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

and $\det(I - W_\lambda(\boldsymbol{q}, \boldsymbol{p})) = 1 - \mathrm{tr}\, W_\lambda + \det W_\lambda = 1 - \phi_\lambda(\boldsymbol{q}, \boldsymbol{p})$, defining the rational function $\phi_\lambda$.

Moser also showed that the system is completely integrable and linearizes on the (generalized) Jacobian of the curve $\mu^2 = e^2(\lambda)\phi_\lambda(x, y)$. Letting $a = -1$, $b = -c = 1, d = 0$ gives the Neumann system.

The dilation $\boldsymbol{q} \mapsto \lambda\boldsymbol{q}$ gives a Lax pair with a parameter, $A \mapsto A + \lambda^2 \boldsymbol{q} \otimes \boldsymbol{q} + \lambda(\boldsymbol{q} \otimes \boldsymbol{p} - \boldsymbol{p} \otimes \boldsymbol{q})$, which makes the spectral curve look more natural. Indeed,

**Remark** (Adler and van Moerbeke 1980). The Neumann flow is equivalent to the Lax pair: $\dot{L}_1 = [M_1, L_1]$, where $L_1 = A\mu^2 + \mu(\boldsymbol{q} \otimes \boldsymbol{p} - \boldsymbol{p} \otimes \boldsymbol{q}) + \boldsymbol{q} \otimes \boldsymbol{q}$ and $M_1 = A\mu + \boldsymbol{q} \otimes \boldsymbol{p} - \boldsymbol{p} \otimes \boldsymbol{q}$. Moreover, the Hamiltonians are of Adler–Kostant–Symes (AKS) type, namely projections (with respect to an ad-invariant inner product) of gradients of orbit-invariant functions to half of the splitting of a Lie algebra.

Specifically, $\{\sum_{-\infty}^{N} A_j\mu^j | A_j \in \mathrm{gl}(n, \mathbf{C})\} = K \oplus N$, with $K = \{\sum_0^N A_j\mu^j\}$ and $N = \{\sum_{-\infty}^{-1} A_j\mu^j\}$; if the inner product is $\langle A, B\rangle = \sum_{i+j=-1} \operatorname{tr} A_iB_j$, the dual of $N$ can be identified with $K = K^{\perp}$, and the Hamiltonian for the Neumann flow can be taken to be $H = \langle (1/2)(L_1\mu^{-2})^2,\ \mu^3 I_{g+1}\rangle$ under the Lie–Poisson brackets and suitable reduction. The flows linearize on the Jacobian of the (hyperelliptic) curve $\det(L_1 - \eta) = 0$.

It is possible to recover the link between the finite and infinite integrable systems (Neumann and KdV) mentioned in the introductory overview, if we notice that squared eigenfunctions for the Lax operator $L = \mathcal{L}^2 = \partial^2 + u$ become algebraic on the spectral curve: Dubrovin *et al.* (2001) introduced the Baker function, namely the unique function $\psi(x, P)$ with the following properties:

(i) For $|x|$ sufficiently small it is meromorphic on $X \setminus \{P_\infty\}$, with pole divisor bounded by $\delta = P_1 + \cdots + P_g$, independent of $x$, such that $h^0(\delta - P_\infty) = 0$, and near $P_\infty \psi(x, P)e^{-xz} = 1 + O(z^{-1})$ is holomorphic, with $z$ chosen to be $\lambda^{1/2}$ in our case.

(ii) We let $\Omega$ be the unique meromorphic differential with zeros on $\delta$ and a double pole of the form $(-\lambda + \text{holomorphic})dz^{-1}$ at $P_\infty$. Note: (1) that Riemann–Roch show that $\Omega$ is unique. (2) We also get a characterization of the dual Baker function, defined as $\psi(x, \iota P)$ in the hyperelliptic case where $\iota$ is the involution $(\lambda, \mu) \mapsto (\lambda, -\mu)$, as meromorphic on $X \setminus \{P_\infty\}$ with poles bounded by $\delta'$ and behavior $e^{-xz}(1 + O(z^{-1}))$ near $P_\infty$, where $\delta + \delta'$ are the $2g$ zeros of $\Omega$. (3) Furthermore, $\Omega = d\lambda/\mathcal{W}(\psi,\ \phi)$, where $\mathcal{W}$ is the Wronskian (with respect to the variable $x$). Then, upon fixing a meromorphic function $h$, normalized at $P_\infty, h = \lambda^{-1/2} + \text{entire}$, with $g + 1$ fixed poles distinct from $\delta$, we have:

*If* $\rho_j = \operatorname{Res}_{e_j} h\Omega,\ q_j = \sqrt{\rho_j}\psi(x, e_j),\ p_j = \sqrt{\rho_j}\phi(x, e_j),$ *then* $\Sigma_{j=1}^{g+1} q_j^2 = 1,\ \Sigma_{j=1}^{g+1} q_jp_j = 0,\ \Sigma_{j=1}^{g+1}(e_jq_j^2 + p_j^2) = u(x)$ *and* $\{q_j, p_j\}$ *satisfy the Neumann system.*

Indeed, the constraints follow from the "residue theorem" applied to the differential $h\Omega\psi\phi$ (it has a residue of $-1$ at $P_\infty$); the differential equations $\ddot{q}_j = e_jq_j - uq_j$ follow from the assumption $L\psi = \lambda\psi$.

The function $u = -2\sum_{k=1}^{g+1}(\sum_{l\neq k} e_l)q_k^2$, evolving under suitable abelian flows, is a solution of the KdV equation; the "times" of the KdV hierarchy are linear combinations of the Neumann Hamiltonians; more precisely, of the invariant vector fields determined by the tangent directions to the image of $X$ in Jac$X$, with Abel map normalized at $P_\infty$, at some point $P$: $D_P = \sum_{k=1}^{g} \lambda(P)^{g-k} D_k$.

The other way around (Moser–Trubowitz, McKean–van Moerbeke),

> *If* $L = \partial^2 + u(x)$ *is a finite-gap operator and* $e_1, \ldots, e_{g+1}$ *are among the* $2g + 1$ *edges of the gaps, there exist constants* $\rho_1, \ldots, \rho_{g+1}$ *so that the functions* $p_j(x) = \sqrt{\rho_j}\psi(x, e_j)$ *satisfy* $\sum_1^{g+1} p_j^2(x) \equiv 1$. *Since* $L\psi_j = e_j\psi_j$, *the* $p_j(x)$ *solve the Neumann system.*

The squared eigenfunctions also provide a natural interpretation for Moser's Lax pair. If $V_\lambda$ is the kernel of $L - \lambda$, then the Baker function $\psi(x, P)$ and its dual $\phi(x, P)$ give a basis of $V_\lambda$ except at the branch points $(e_i, 0)$ where $\psi = \phi$. But then the normalized basis of $V_\lambda$ is related to $\psi$, $\phi$ by a constant matrix:

$$\begin{bmatrix} y_0 \\ y_1 \end{bmatrix} = C\begin{bmatrix} \psi \\ \phi \end{bmatrix}$$

while

$$B\begin{bmatrix} \psi \\ \phi \end{bmatrix} = \begin{bmatrix} \mu & 0 \\ 0 & -\mu \end{bmatrix}\begin{bmatrix} \psi \\ \phi \end{bmatrix}$$

if $B$ is the differential operator of the Burchnall–Chaundy ring corresponding to multiplication by $\mu$, so that

$$\begin{bmatrix} -V & U \\ W & V \end{bmatrix}^{\mathrm{T}} = M_B = C\begin{bmatrix} \mu & 0 \\ 0 & -\mu \end{bmatrix}C^{-1}$$

By evaluating at $x = 0$, we find:

$$C = \frac{1}{\mathcal{W}}\begin{bmatrix} \phi' & -\psi' \\ -\phi & \psi \end{bmatrix}\Bigg|_{x=0}$$

with $\mathcal{W} = \psi\phi' - \psi'\phi$. Finally, we calculate:

$$C\begin{bmatrix} \mu & 0 \\ 0 & -\mu \end{bmatrix}C^{-1} = \frac{\mu}{\mathcal{W}}\begin{bmatrix} \psi\phi' + \psi'\phi & 2\psi'\phi' \\ -2\psi\phi & -(\psi\phi' + \psi'\phi) \end{bmatrix}$$

so that $U(\lambda) = \psi\phi' + \psi'\phi, V(\lambda) = -2\psi\phi, W(\lambda) = 2\psi'\phi'$ are polynomials like the entries of $W_\lambda(\boldsymbol{q}, \boldsymbol{p}) \cdot e^2(\lambda)$, and the fact that $UW + V^2$ does not depend on $x$ expresses the fact that $\mathcal{W} = \text{constant}$.

An object that links the two distinct occurrences of Lax pairs is Sato's infinite-dimensional Grassmann manifold. One particular model will serve as illustration, with more general settings covered by Dickey (2003). Sato defined a one-to-one correspondence between cyclic $\mathcal{D}$-submodules $\mathcal{I}$ of $\mathcal{P}$, namely of the type $\mathcal{I} = \mathcal{D}S$ (which turns out to be equivalent to the property: $\mathcal{P} = \mathcal{I} \oplus \mathcal{P}^{(-1)}$), and subspaces of a ring of formal power series, which make up an infinite-dimensional Grassmann manifold, more

precisely elements of $\mathrm{Gr}^{\emptyset}$, the "big cell." This way, KP can be viewed as deformation of $\mathcal{D}$ modules.

There are two ways to set up the Grassmannian: (1) more direct as a limit of finite-dimensional Grassmannians; (2) more intrinsic, using the rings $\mathcal{D} \subset \mathcal{P}$.

1. Let $\dim V = m + n = N, \mathrm{Gr}(m, V) = \{m - \text{frames in } V\}/\mathrm{GL}(m) \hookrightarrow \mathbb{P}(\wedge^m V)$ via $\xi^{(0)}, \ldots, \xi^{(m-1)} \mapsto \xi^{(0)} \wedge \cdots \wedge \xi^{(m-1)}$.

If we fix a basis $e_0, \ldots, e_{N-1}$ of $V$, and write a frame in coordinates, $\xi^{(i)} = \xi_{0,i} e_0 + \cdots + \xi_{N-1,i} e_{N-1}$, then

$$\xi^{(0)} \wedge \cdots \wedge \xi^{(m-1)} = \sum_{0 \le \ell_0 < \cdots < \ell_{m-1} < N} \xi_{\ell_0 \ldots \ell_{m-1}} e_{\ell_0} \wedge \cdots \wedge e_{\ell_{m-1}}$$

with $\xi_{\ell_0 \ldots \ell_{m-1}} = \det(\xi_{\ell_i, j})_{i,j=0,\ldots,m-1}$

A point in the ambient $\mathbb{P}(\wedge^m V)$ lies in the embedded $\mathrm{Gr}(m, V) \Leftrightarrow$ its projective coordinates $\xi_{\ell_0 \ldots \ell_{m-1}} (0 \le \ell_i < N)$ satisfy the Plücker relations (PRs):

$$\sum_{i=0}^{m} (-1)^i \xi_{k_0 \ldots k_{m-2} \ell_i} \xi_{\ell_0 \ldots \hat{\ell}_i \ldots \ell_m} = 0$$

Therefore,

$$\mathrm{Gr}(m, V) = (\widetilde{\mathrm{Gr}}(m, V) \backslash \{0\})/\mathrm{GL}(1)$$

where $\widetilde{\mathrm{Gr}}(m, V) = \{(\xi_Y)_{Y \subset \Delta_{mN}}$ satisfying the PRs$\}$ is a line bundle over $\mathrm{Gr}(m, V)$, $Y$ is a Young diagram consisting of rows

$$\begin{array}{l} \boxed{\ell_{m-1} - (m-1)} \\ \vdots \\ \boxed{\ell_1 - 1} \\ \boxed{\ell_0} \end{array}$$

so it is contained in the rectangle $\Delta_{mN}$.

For the commutative diagram:

$$\begin{array}{ccc} \widetilde{\mathrm{Gr}}(m', N') & \overset{\text{project}}{\longrightarrow} & \widetilde{\mathrm{Gr}}(m, N) \\ \downarrow \text{identity} & & \downarrow \text{identity} \\ \widetilde{\mathrm{Gr}}(m', N') & \overset{\text{embed}}{\hookleftarrow} & \widetilde{\mathrm{Gr}}(m, N) \end{array}$$

the following facts can be checked. Let $m \le m', n \le n', N' = m' + n'$:

(i) if $(\xi'_Y)_{Y \subset \Delta_{m'N'}}$ satisfies PRs, so does its restriction to $Y$'s within $\Delta_{mN}$;

(ii) if $(\xi_Y)_{Y \subset \Delta_{mN}}$ satisfies PRs, so does $(\xi'_Y)_{Y \subset \Delta_{m'N'}}$ where $\xi'_Y = 0$ unless $Y \subset \Delta_{mN}$.

These facts make it possible to define: $\mathrm{Gr} = (\widetilde{\mathrm{Gr}} \backslash \{0\})/\mathrm{GL}(1)$, where $\widetilde{\mathrm{Gr}} = \{(\xi_Y)_{Y \text{ all Young diagrams}}$ satisfying all PRs$\}$

$$\begin{array}{ccc} \widetilde{\mathrm{Gr}} & \overset{\text{project}}{\longrightarrow} & \widetilde{\mathrm{Gr}}(m, N) \\ \uparrow \text{dense} & & \downarrow \text{identity} \\ \widetilde{\mathrm{Gr}}^{\text{fin}} & \overset{\text{embed}}{\hookleftarrow} & \widetilde{\mathrm{Gr}}(m, N) \end{array}$$

and

$$\begin{aligned} \widetilde{\mathrm{Gr}}^{\text{fin}} &= \{(\xi) \in \widetilde{\mathrm{Gr}} : \xi_Y = 0 \text{ for almost all } Y\} \\ &= \bigcup_{m,N} \widetilde{\mathrm{Gr}}(m, N) \end{aligned}$$

The KP time deformations are defined as follows:

$$\xi_Y(t) := \sum_{\text{all } Y'} \chi_{Y'/Y}(t) \xi_{Y'} \text{ where } \chi_{Y'/Y}(t) := \det(p_{\ell'_i - \ell_j}(t))$$

$$p_0(t) = 1, p_n(t) := \sum_{\nu_1 + 2\nu_2 + 3\nu_3 + \cdots = n} t_1^{\nu_1} t_2^{\nu_2} \ldots / (\nu_1! \nu_2! \ldots)$$

Write $\chi_{Y/\emptyset}$ as $\chi_Y$, where $\chi_Y(t) = \det(p_{\ell_i - j}(t))$ are the Schur functions.

To connect with the KP hierarchy, let

$$w_n(x, t) := (-1)^n \frac{\xi_{\Delta_{n,1}}(x + t)}{\xi_\emptyset(x + t)}$$

where $x + t = (x + t_1, t_2, \ldots)$, and $S := 1 + w_1(x, t) \partial^{-1} + \cdots$. Then $\mathcal{L} = S \partial S^{-1}$ satisfies the KP hierarchy, namely $\partial_{t_n} S = B_n S - S \partial^n$, where $B_n := (S \partial^n S^{-1})_+, \Longleftrightarrow [\partial_{t_n} - B_n, \partial_{t_k} - B_k] = 0 \Longleftrightarrow \partial_{t_n} \mathcal{L} = [(\mathcal{L}^n)_+, \mathcal{L}]$.

**Note** The Plücker coordinate $\xi_\emptyset(t) = \sum_{\text{all } Y} \chi_Y(t) \xi_Y = \tau(\xi, t)$ is a generating function for the Plücker coordinates, $\xi_Y(t) = \chi_Y(\partial_t) \xi_\emptyset(t)$, where

$$\partial_t := \left( \frac{\partial}{\partial t_1}, \frac{1}{2} \frac{\partial}{\partial t_2}, \frac{1}{3} \frac{\partial}{\partial t_3}, \ldots \right)$$

Now by reducing to $\mathrm{Gr}(m, N)$ and checking that every $\xi_Y(t)$ satisfies PRs, we have a dynamical system on $\widetilde{\mathrm{Gr}}$.

**Conclusion** (Sato). Although any $f(t) \in \mathbb{C}[[t_1, t_2, \ldots]]$ admits a formal expression of the form $\sum_Y c_Y \chi_Y(t)$, where the coefficients are

$$c_Y = \chi_Y(\partial_t) f(t)|_{t=0}$$

it represents the $\tau$ function for some $\xi \in \widetilde{\mathrm{Gr}} \Longleftrightarrow$ its coefficients satisfy the following PRs:

$$\sum_{i=0}^{m} (-1)^i \chi_{k_0 \ldots k_{m-1} \ell_i} \left( \frac{\partial_t}{2} \right) \chi_{\ell_o \ldots \hat{\ell}_i \ldots \ell_m} \left( -\frac{\partial_t}{2} \right) \tau \cdot \tau = 0$$

which is the KP hierarchy in Hirota bilinear form.

2. Let

$$V := \frac{\mathcal{P}}{\mathcal{P}x} \cong \mathcal{P}_{\text{const}} = \left\{ \sum_{-\infty<i<<\infty} a_i \partial^i, \ a_i \in \mathbb{C} \right\}$$

equipped with the induced filtration $V^{(i)}$ by order, induced by

$$\mathcal{P}^{(i)} = \left\{ \sum_{-\infty<k\leq i} a_k \partial^k, \ a_k \in \mathbb{C} \right\}$$

and define

$$\begin{aligned} \text{Gr} = \{&\text{vector subspaces } W \text{ of } V \\ &\text{s.t.} \dim(W \cap V^{(0)}) = \dim V/(W + V^{(0)}) < \infty\} \end{aligned}$$

"same size" as the reference subspace $\{\sum_{\nu\leq 0} c_\nu e_\nu : c_\nu \in \mathbb{C}\} = V^{(0)}$.

The correspondence between such a $W$ and a cyclic submodule of $\mathcal{P}$ is given as follows:

$$\mathcal{I} \mapsto W = S^{-1}V^{(0)} = \{v \in V : \mathcal{I}v \subset V^{(0)}\}$$
$$W \mapsto \mathcal{I} = \{A \in \mathcal{P} : AW \subset V^{(0)}\}$$

Generic points of particular interest in constructing KP solutions make up the "big cell":

$$\begin{aligned} \text{Gr}^{\emptyset} \underset{\text{open dense}}{\subset} \text{Gr} &\Longleftrightarrow V = W \oplus V^{(0)} \\ &\Longleftrightarrow \xi_{\emptyset} \neq 0 \text{ and a } \tau \text{ function can} \\ &\qquad \text{be defined as above} \end{aligned}$$

In standard basis of $V, e_i := \partial^{-i-1} \bmod \mathcal{P}x, \ i \in \mathbb{Z}$, the action

$$\begin{aligned} xe_i &= (i+1)e_{i+1} \\ \partial e_i &= e_{i-1} \end{aligned}.$$

gives $V$ a $\mathcal{P}$-module structure. Let $\Lambda$ be the shift operator: $\partial e_i = e_{i-1}$; then

$$\xi(t) = \mathrm{e}^{(t_1\Lambda + t_2\Lambda^2 + \cdots)} \cdot \xi$$

so, this linearizes the flows!

This survey would not be complete without an example of the formula that links the $\tau$ and the theta function; more general statements and groups of symmetries can be found in Dickey (2003). A solution of the KP hierarchy can be expressed in terms of the $\tau$ function $\tau_W$ associated with an element $W$ of $\text{Gr}(H)$, in the model $\text{Gr}(H)$, where $H = L^2(S^1), H = H_+ \oplus H_-$ with standard basis $H_+ = \langle 1, z, z^2, \ldots \rangle, H_- = \langle z^{-1}, z^{-2}, \ldots \rangle$ and $p_\pm$ the projections, $\tau_W(g) = \det(\mu_g \circ p_+ \circ \mu_{g^{-1}} \circ (p_+|_W)^{-1})$, where $g = \mathrm{e}^{\Sigma t_i z^i}$. The associated Baker function $\psi_W(g, z)$ is a function of the form

$$\psi_W(g, z) = g(z)\left(1 + \sum_{i=-\infty}^{-1} a_i z^i\right)$$

with $a_i \in \mathbb{C}[[t_1, t_2, \ldots]]$ for each $i$, such that the map $z \mapsto \psi_W(g, z)$ is an element of $g^{-1}W$. If $\phi = 1 + \sum_{i=-\infty}^{-1} a_i z^i$, then $L = \phi\partial\phi^{-1}$ is a solution of the KP hierarchy.

Moreover,

$$g^{-1}\psi_W(g, \zeta) = \frac{\tau_W\left(\left(t_\alpha - \dfrac{1}{\alpha\zeta^\alpha}\right)\right)_\alpha}{\tau_W((t_\alpha))_\alpha}$$

This is the analog of the expression for the Baker function in terms of the theta function, when $W$ corresponds to an element of the Jacobian of the spectral curve $\Gamma$ via the Krichever map

$$\psi(x, P) = \exp\left(x \int \eta - xa\right) \times \frac{\vartheta(Ux + A(P) - A(D) - \Delta)\vartheta(A(D) + \Delta)}{\vartheta(A(P) - A(D) - \Delta)\vartheta(Ux - A(D) - \Delta)}$$

where $P \in \Gamma, A(-)$ is the Abel map, $\Delta$ the Riemann constant, $U \in \mathbf{C}^g$ a suitable vector, $D$ a generic divisor of points $P_1, \ldots, P_g \in \Gamma, \eta$ a differential of the second kind, and $a$ a constant depending on the curve. For the KdV solutions, the condition on $W \in \text{Gr}^{\emptyset}$ is that $z^2 W \subset W$ and the solution is

$$u_W(x, t_2, t_3, \ldots) = 2\partial \log \tau_W(x, t_2, t_3, \ldots)$$

In the Grassmannian formulation, the Hirota bilinear operator mentioned in the introductory overview makes its third and most general appearance (we regard Baker's and Hirota's definitions as the first two – the one based on a residue formula in algebraic geometry, the other on the vanishing of a differential form):

**Definitions**

(i) In $\mathcal{P}$, it is possible to conjugate any $\mathcal{L} = \partial + u_{-1}(x)\partial^{-1} + \cdots$ into $\partial$ by a $K = 1 + v_{-1}(x)\partial^{-1} + \cdots$, determined up to elements of $\mathbf{C}[\partial] = \mathcal{C}_{\mathcal{D}}(\partial)$: $K^{-1}\mathcal{L}K = \partial$.

(ii) We define a formal Baker function for $\mathcal{L}$ as the element of the module $M$ (the free, rank-1 $\mathcal{P}$-module = space of formal expressions $f = \mathrm{e}^{xz}\tilde{f}$ where $\tilde{f} = \sum_{-\infty}^{N} f_j(x)z^j$, with generator $\mathrm{e}^{xz}$) such that $\mathcal{L}\psi = z\psi$; so that $\psi = K\mathrm{e}^{xz}$ for $K$ as in (i).

(iii) We say that the formal adjoint $A^\dagger$ of a (formal pseudo) differential operator $A = \sum_{j=-\infty}^{N} u_j(x)\partial^j$ is $A^\dagger = \sum_{j=-\infty}^{N} (-\partial)^j u_j(x)$, and that the dual Baker function $\psi^\dagger$ to $\psi = K\mathrm{e}^{\Sigma t_j z^j}$ is the Baker function of $(\mathcal{L}^\dagger)$; the operator which corresponds to $K$ in (i) is $(K^\dagger)^{-1}$, that is, $(K^\dagger)^{-1}\mathcal{L}^\dagger K^\dagger = -\partial$.

Then, the KP hierarchy is equivalent to the following formula:

$$\text{Res}_z \psi(t', z)\psi^\dagger(t, z) = 0$$

Moreover, as proved in Dickey (2003), if $\psi_1$ and $\psi_2$ are formal power series of the form $\psi_1 = Ke^{\Sigma t_i z^i}, \psi_2 = Je^{-\Sigma t_i z^i}$, for $K, J \in 1 + \mathcal{P}^{(-1)}$, satisfying the condition

$$\mathrm{Res}_z \left( \partial_{i_1}^{\alpha_1} \partial_{i_2}^{\alpha_2} \cdots \partial_{i_m}^{\alpha_m} \psi \right) \cdot \phi = 0$$

then there exists an operator $\mathcal{L}$ satisfying the Lax equations, whose wave function and adjoint wave function are $\psi_1, \psi_2$, respectively.

To conclude this overview of Lax equations, we point out that they can be viewed as zero-curvature condition for a (formal) connection (on the trivial bundle over the formal deformation space whose fiber is $\mathcal{P}$), rephrasing the fact that the time flows commute and hence define time deformations; such formulation can be found in Mulase (1984).

## Symplectic Reduction and $r$ Matrices

While the Lax-pair presentation provides natural spectral invariants, the group/representation-theoretic nature of integrability (sometimes referred to as hidden symmetries) is best seen in the context of Marsden–Weinstein reduction. We perform it in the example of a generalization of Moser's rank-2 perturbation; we extract the basic construction from Adams *et al.* (1988). A more comprehensive treatment can be found in Babelon *et al.* (2003).

**Definition** We let $M_{n,r}$ denote the space of $n \times r$ complex matrices, with $n \geq r$ and give $M = M_{n,r} \times M_{n,r}$ the symplectic structure $\omega(F, G) = \mathrm{tr}(\mathrm{d}F \wedge \mathrm{d}G^{\mathrm{T}})$ for $F, G \in M$. A rank-$r$ perturbation of a fixed $n \times n$ matrix $A$ is $L = A + FG^{\mathrm{T}}$.

**Definition** We split the formal loop algebra $\widetilde{\mathrm{gl}(r)} = \widetilde{\mathrm{gl}(r)}^+ \oplus \widetilde{\mathrm{gl}(r)}^-$ where $\widetilde{\mathrm{gl}(r)}^+$ consists of $r \times r$ matricial polynomials in $\lambda$ and $\widetilde{\mathrm{gl}(r)}^-$ of strictly negative formal power series. Under the pairing $\langle X(\lambda), Y(\lambda)\rangle = \mathrm{tr}(X(\lambda)Y(\lambda))_-$ (where the subscript $-$ means the coefficient of $\lambda^{-1}$), the dual of $\widetilde{\mathrm{gl}(r)}^+$ is identified with $\widetilde{\mathrm{gl}(r)}^-$, which therefore admits a Lie–Poisson structure.

In sketch, we consider an action on $M$ whose moment map lands in $\widetilde{\mathrm{gl}(r)}^-$; we check that the AKS flows on $\widetilde{\mathrm{gl}(r)}^-$ correspond to isospectral deformations of $L = A + FG^{\mathrm{T}}$ for flows on $M_A$; finally, we perform a Marsden–Weinstein reduction for an (equivariant) $\mathrm{GL}(r)$ action to obtain a completely integrable system on a symplectic leaf, whose flows are linear on the Jacobian of the spectral curve. We recall very briefly the general definitions.

### Moment Map

1. A smooth group action of $G$ on a symplectic manifold $(M, \omega)$ is said to be Hamiltonian if there exists a "moment map" $J : M \to g^*$ such that the Hamiltonian vector field associated with $J$ and a fixed element $\xi \in g$ is the same as the infinitesimal action associated with $\xi$. However, an infinitesimal definition is given because in the formal setup the group of a Lie algebra is often delicate to define. We recall that:
2. The Lie–Poisson structure of $g^*$ is defined by

$$\{\phi, \psi\}_{g^*}(\mu) = < \mu, [\mathrm{d}\phi(\mu), \mathrm{d}\psi(\mu)] >$$
$$\text{for } \phi, \psi \in \mathcal{C}^\infty(g^*), \quad \mu \in g^*$$

where $\mathrm{d}\phi : g^* \to g^{**}$ (which in our situation will always be identified with $g$) is defined by

$$< \mathrm{d}\phi(\mu), \nu > = \frac{\mathrm{d}}{\mathrm{d}t} \phi(\mu + t\nu)\Big|_{t=0}, \quad \mu, \nu \in g^*$$

Now we say that $J : M \to g^*$ is a moment map if

3. its linear dual $j : g \to \mathcal{C}^\infty(M)$ is a Lie-algebra homomorphism; or if
4. it is a Poisson map with respect to the Lie–Poisson structure: $\phi, \psi \in \mathcal{C}^\infty(g^*) \Rightarrow \{J^*\phi, J^*\psi\} = J^*\{\phi, \psi\}_{g^*}$. In case we do have a Hamiltonian $G$-action, then the subspace $\mathcal{C}_G^\infty(M)$ of $G$-invariant functions is a Lie subalgebra of $\mathcal{C}^\infty(M)$. If $G$ acts freely and properly on $M$, then $M/G$ is a manifold with a Poisson structure inherited from the one on $M$ through the identification $\mathcal{C}^\infty(M/G) \cong \mathcal{C}_G^\infty(M)$. The symplectic leaves of $M/G$ have the form $M_\mu = J^{-1}(\mu)/G_\mu = J^{-1}(\mathcal{O}_\mu)/G$, where $\mu \in g^*$, $G_\mu$ is the isotropy group of $\mu$ in $G$ and $\mathcal{O}_\mu$ is the $G$-orbit through $\mu$. The reduced manifold $M_\mu$ has a natural symplectic structure $\omega_\mu$ such that $i^*\omega = \pi^*\omega_\mu$, where $i : J^{-1}(\mu) \to M$ is inclusion and $\pi : J^{-1}(\mu) \to M_\mu$ is the natural projection taking points to their $G_\mu$-orbits.

This class of examples can be treated with the technique of a (classical) $r$-matrix, as follows. Given a linear map $R : g \to g$, the alternating bilinear form $[X, Y]_R = (1/2)([RX, Y] + [X, RY])$ satisfies the Jacobi identity $\Leftrightarrow$ certain quadratic conditions on $R$ are satisfied. Assuming they are, for all pairs of invariant functions $I, J$ on $g^*$, we have $\{I, J\}_R = 0$ (where $\{\,,\,\}_R$ is the attendant (Lie–Poisson) structure). Indeed, $\{I, J\}_R(\mu) = \langle [\mathrm{d}I(\mu), \mathrm{d}J(\mu)]_R, \mu\rangle = (1/2)\langle [R\mathrm{d}I(\mu), \mathrm{d}J(\mu)], \mu\rangle + (1/2)\langle [\mathrm{d}I(\mu), R\mathrm{d}J(\mu)], \mu\rangle$, but, for example, $\langle [R\mathrm{d}I(\mu), \mathrm{d}J(\mu)], \mu\rangle = \langle R\mathrm{d}I(\mu), \mathrm{ad}^*\mathrm{d}J(\mu)(\mu)\rangle = 0$.

**Remark** As is clear from the proof above, our definition of invariant need only be infinitesimal, that is, $f \in I(g^*)$ iff $< \mu, [\mathrm{d}f(\mu), X] > = 0 \ \forall \mu \in g^*$, $X \in g$. Of course, when we have a corresponding Lie group the invariants are the functions which are

invariant under the natural action, such as the symmetric functions of the eigenvalues of a matrix.

### AKS Flows

For a splitting $g = K \oplus N$, as given above, with $g^* = N^* \oplus K^*$, an example of $r$-matrix is given by $R(X) = X_+ - X_-$ (where $+, -$ denote projection to $K$, $N$): the Jacobi identity is straightforward to check. As a consequence, invariants on $g^*$ are in involution with respect to $\{,\}_R$ and these are called AKS flows, after work done independently by AKS: $\dot{X} = [\mathrm{d}f(\tilde{X})_+, X] = -[\mathrm{d}f(\tilde{X})_-, X]$, given here for the special case in which we can identify $K$ with $K^*$ and $\tilde{X}$ is the element in $K^*$ that corresponds to $X \in K$.

We now proceed to the appropriate moment maps. We generalize the constant matrix $A$ introduced above (isospectral deformations) by allowing multiple eigenvalues $\alpha_i$ of multiplicities $n_i \le r, n_1 + \cdots + n_k = n$, so that $\det(A - \lambda I) = \prod_{i=1}^k (\alpha_i - \lambda)^{n_i}$. Let $a(\lambda) = \prod_{i=1}^k (\alpha_i - \lambda)$. We split an $n \times r$ matrix $F$ into $k$ blocks $F_i$ accordingly.

**Definition/statement**

(i) $J_r^n(F, G)(x_1, \ldots, x_n) = -\sum_{j=1}^n \mathrm{tr}(F_j X_j G_j^{\mathrm{T}})$ is the moment map of the action $[(g_1, \ldots, g_n)(F, G)]_i = (F_i g_i^{-1}, G_i g_i^{\mathrm{T}})$, where $g_i \in \mathrm{GL}(r)$ so that under standard identifications $J_r^n(F, G) = -(G_1^{\mathrm{T}} F_1, \ldots, G_n^{\mathrm{T}} F_n)$ and restricting the action to the diagonal subgroup $\{(g, \ldots, g)\}$, $J_r(F, G) = -G^{\mathrm{T}} F$.

(ii) For $X(\lambda) \in \widetilde{\mathrm{gl}(r)}^+$ we define $\alpha(X(\lambda)) = (X(\alpha_1), \ldots, X(\alpha_r))$ and obtain the exact sequence

$$0 \to a(\lambda)\widetilde{\mathrm{gl}(r)}^+ \xrightarrow{\iota} \widetilde{\mathrm{gl}(r)}^+ \xrightarrow{\alpha} g_r^n \to 0$$

By dualizing, and identifying $g_r^n$ to its dual by using the trace componentwise, we get

$$\alpha^*(Y_1, \ldots, Y_n) = \sum_{i=1}^k \frac{Y_i}{\lambda - \alpha_i}$$

and finally check that $\tilde{J}_r = \alpha^* \circ J_r^n$ is a moment map. By combining (i) and (ii), we get a moment map

$$\tilde{J}_r(F, G) = \sum_{i=1}^k \frac{G_i^{\mathrm{T}} F_i}{\alpha_i - \lambda} = G^{\mathrm{T}}(A - \lambda)^{-1} F$$

which becomes injective on $\mathcal{M}/H$, where $\mathcal{M}$ is a suitable open submanifold of $M$ and $H = \mathrm{GL}(n_1) \times \cdots \times \mathrm{GL}(n_k)$ acts blockwise by $(h_i F_i, h_i^{-1\mathrm{T}} G_i)$.

(iii) We also notice that the "Moser space" $M_A = \{A + FG^{\mathrm{T}} | F, G \in \mathcal{M}\}$ of rank-$r$ perturbations can be identified with the orbit space $\mathcal{M}/G_r, G_r = \mathrm{GL}(r)$ acting as in (i).

To finish, we turn on the obvious AKS flows on $\widetilde{\mathrm{gl}(r)}^-$: the key observation is that they are isospectral for the rank-$r$ perturbation $A + FG^{\mathrm{T}}$: we see that the Poisson-commutative ring $\mathcal{F}_+$ of projected invariants defines, by composition with $\tilde{J}_r$, a Poisson-commutative ring $\mathcal{F}$ of isospectral flows on $M_{n,r} \times M_{n,r}$.

## Hitchin Systems

The Hitchin system, introduced in the late 1980s, 20 years later still encompasses the most general class of "algebraically completely integrable" systems, which we now discuss. In its most basic form, the concept of "algebraic completely integrable" (ACI) Hamiltonian system, is an extra condition on the integrability of classical mechanics, in the following sense.

A Hamiltonian system with $n$ degrees of freedom, that is, defined on a symplectic manifold $M$ of (real) dimension $2n$ is (Arnol'd–Liouville) completely integrable if it admits $n$ functions in involution whose differentials are linearly independent (possibly, generically on $M$). When $M$ is a component of the set of real points of an algebraic variety $M_{\mathbb{C}}$ and the symplectic form $\omega$ and Hamiltonian function $H$ are rational without poles on $M$, the concept of algebraic complete integrability can be introduced. For this condition to hold, we require that the vector fields corresponding to the Hamiltonians in involution still have no poles on a compactification of the fibers on $M_{\mathbb{C}}$.

**Nonexample** (Mumford 1984, §4). Consider

$$M = \mathbb{R}^2, \quad \omega = \mathrm{d}x \wedge \mathrm{d}y, \quad H = x^4 + y^4$$

Here a compactification of the fiber, the affine curve $x^4 + y^4 = c$, is the projective curve $X^4 + Y^4 = cZ^4$, which is smooth (provided $c \neq 0$) and has four points at infinity. The vector field $X_H$ defined by $H$, $X_H \rfloor \omega = -\mathrm{d}H$, is tangent to the fiber in the affine plane, but has a pole at infinity as can be checked by a change of coordinates; 4 is the lowest exponent for which this simple nonexample works!

**Note** In the algebraically completely integrable situation, the fibers are abelian varieties or extensions of such by $\mathbb{C}^{*k}$ for some power $k$. This gives rise to the issues of variations of periods over the base mentioned in the introductory overview.

The Neumann system is ACI, with integral tori given by the Jacobians of the spectral curves:

$$\Gamma : \mu^2 = g(\lambda) = \prod_1^{2g+1} (\lambda - e_i) = UW + V^2$$

where

$$L = e(\lambda)\begin{bmatrix} \sum \frac{q_i p_i}{\lambda - e_i} & \sum \frac{p_i^2}{\lambda - e_i} \\ 1 + \sum \frac{q_i^2}{\lambda - e_i} & -\sum \frac{q_i p_i}{\lambda - e_i} \end{bmatrix} = \begin{bmatrix} V & U \\ W & -V \end{bmatrix}, \qquad e(\lambda) = \prod_1^{g+1}(\lambda - e_i)$$

$U = \prod_{i=1}^{g}(\lambda - \lambda_i), \quad (\lambda_1, \ldots, \lambda_g)$ "elliptical spherical coordinates"

$$\psi = \left(1, \frac{U}{V+\mu}\right) \text{ eigenvector: } L\psi = \mu\psi$$

$$\text{divisor: } \sum_{i=1}^{g}(\lambda_i, V(\lambda_i))$$

Hitchin (1982) devised a geometrical model of the spectral curve, a compact algebraic curve contained in the surface $\mathcal{T}^*\mathbb{P}^1$, and its line bundles. He also provided subsequently (1987) a dramatic generalization.

Hitchin's construction, in the Neumann-system example, highlights the following objects:

- $L \in H^0(\mathbb{P}^1, \mathrm{End}(E) \otimes \mathcal{O}(g+1))$, $E$ rank $(r=)2$ bundle over $\mathbb{P}^1$;
- $T =$ total space of the line bundle $\mathcal{O}(g+1)$ over $\mathbb{P}^1$;
- $\mu =$ tautological section: $\mathbb{P}^1 \to T$, where $\tilde{L} - \tilde{\mu}I \in H^0(T, \mathrm{End}(\tilde{E}) \otimes \tilde{\mathcal{O}}(g+1))$ (tildes denote pullback);
- $\Gamma$: $\det(\tilde{L} - \tilde{\mu}I) = 0$. The line bundle $\psi$ (eigenvectors) is defined as the kernel of $\tilde{L} - \tilde{\mu}I$; and
- the moduli space of spectral curves is a linear system on the surface $T$. Fixing $\{e_1, \ldots, e_{g+1}\}$ in the above example gives constraints that define it as subsystem of a complete linear system, as well as providing a Poisson structure on the whole $((2g-1)+g)$-dimensional manifold (base = curves, fiber = Jacobians) which reduces to the standard $\sum \mathrm{d}p_i \wedge \mathrm{d}q_i$. Equivalent to choosing a section $s \in H^0(\mathbb{P}^1, \mathcal{O}(g-1) \otimes K_{\mathbb{P}^1}^{-1})$,

$$\begin{array}{ll} \Gamma & s \leftrightarrow (e_1, \ldots, e_{g+1}) \\ \downarrow r:1 & E \to E \otimes \mathcal{O}_{\mathbb{P}^1}(g+1) \leftrightarrow L \\ \mathbb{P}^1 & (\lambda:1) \in \mathbb{P}^1 \end{array}$$

### Generalizations

- $\mathbb{P}^1 \to$ Riemann surface $X$ of genus $g > 1$;
- $E$ stable rank-$r$ vector bundle over $X$. To give a concrete example, we will take $r = 2$ and fix $\det\ E = \mathcal{O}_X$.

### Hitchin's Abelianization Program

**Fact** (Hitchin). *Every such bundle $E$ over $X$ can be realized as the direct image of a line bundle over a spectral curve $\Gamma \overset{r:1}{\to} X$.*

We introduce the moduli space $\mathcal{M} = \mathcal{SU}_X(2, \mathcal{O}_X) = S$-equivalence classes of $E$'s, $E$ semistable rank-2 bundle over $X$, $\det E = \mathcal{O}_X$. The dimension of $\mathcal{M}$ is $3g-3$.

Hitchin (1987) proved that $\mathcal{T}^*\mathcal{M}$ is ACI (generically, there exist $3g-3$ regular functions in involution with respect to the standard symplectic structure, with invariant manifolds isomorphic to Prym $\Gamma$, where $\Gamma =$ spectral curve).

To recognize the analog of the features highlighted above, we recall that Kodaira–Spencer deformation theory gives the following description of the cotangent bundle: since a rank-$r$ vector bundle over $X$ is determined by a 1-cocycle with values in $\mathrm{GL}(r, \mathcal{O}_X)$, a first-order deformation of $E$ is given by a 1-cocycle with values in the associated bundle of Lie algebras, hence by a class in $H^1(X, \mathrm{End}(E))$, so the cotangent bundle has Serre-dual fiber $H^0(X, E \otimes E^* \otimes K)$.

**Hitchin map** $(E, \phi) \in \mathcal{T}^*\mathcal{M}$ (Higgs field, trace zero, $\phi \in H^0(X, \mathrm{End}_0(E) \otimes K)$):

$H: \phi \mapsto \det\phi$ (more generally for any $r \geq 2$, $\mathrm{tr} \wedge^i \phi \in H^0(X, K^{\otimes i})$ $i = 2, \ldots, r$;
$\mu \mapsto -\mu$ defines Prym $\Gamma, \mu^2 = \det\phi \in H^0(X, K^{\otimes 2})$ defines $\Gamma$.

### Explicit Hamiltonians for the Hitchin System

The cases in which $X$ is genus 0 and 1 were solved explicitly by Nekrasov (1996) using explicit parametrizations of the moduli spaces; this includes the case of insertions (singular curves), yielding (elliptic) Gaudin models. We report the solution for the genus-2 case (van Geemen and Previato 1996).

**Remark** The map $H$ projectivizes,

$$\bar{H}: \mathbb{P}H^0(X, \mathrm{End}_0(E) \otimes K) \to \mathbb{P}H^0(X, K^{\otimes 2})$$
$$\det(c\phi) = c^2 \det\phi$$

Coordinates on $\mathcal{T}^*\mathcal{M}$ can be given as follows:
$\Theta \subset \mathrm{Pic}^{g-1}X =$ canonical theta divisor

$$\Delta: \mathcal{M} \to |2\Theta| = \mathbb{P}^{2^g-1}$$

$$E \mapsto D_E = \{\xi \in \mathrm{Pic}^{g-1}X : h^0(E \otimes \xi) > 0\}$$

$X$ hyperelliptic $\Rightarrow \Delta$ is 2:1 except for $g = 2$ (every point of $\mathcal{M}$ is fixed under the hyperelliptic

involution), where $\mathcal{M} \cong \mathbb{P}^3$. For a vector space $V$ the Euler sequence gives

$$\mathbb{P}\mathcal{T}^*\mathbb{P}V \cong I = \{(x, h) \in \mathbb{P}V \times \mathbb{P}V^* : x \in h\}$$

In our case,

$$\mathbb{P}V \times \mathbb{P}V^* = |2\Theta| \times |2\Theta_0|$$

Define six polynomial functions $H_i$ on $\mathbb{P}^3 \times \mathbb{P}^{3*}$ by the requirement: for generic $q \in \mathbb{P}^3, (H_i = 0) \cap \mathbb{P}T_q\mathbb{P}^3 = \ell_i \cup \ell_i'$, the six pairs of bitangents to $\mathcal{K} \cap \mathbb{P}T_q^*\mathbb{P}^3$, where $\mathcal{K}$ is the Kummer surface (the remaining 16 bitangents are cut out by the tropes.)

Recall that the Grassmannian of lines in $\mathbb{P}^3, \mathrm{Gr}(2,4)$, is defined by an equation $\sum_1^6 X_i^2 = 0$ in Klein's coordinates

$$(X_1 : \ldots : X_6) \in \mathbb{P}^5$$

$$X_1 = p_{01} + p_{23}, \qquad X_2 = i(p_{01} - p_{23})$$

$$X_3 = i(p_{02} - p_{13}), \qquad X_4 = p_{02} + p_{13}$$

$$X_5 = p_{03} + p_{12}, \qquad X_6 = i(p_{03} - p_{12})$$

where $p_{ij} = Z_i W_j - W_j Z_i$ are Plücker's coordinates on the line

$$\langle (Z_0 : \ldots : Z_3)(W_0 : \ldots : W_3) \rangle \subseteq \mathbb{P}^{3*}$$

Using coordinates on the incidence variety $I$ given by the sections $\Phi_i$ of the bundle projection $\mathbb{P}\mathcal{T}^*\mathbb{P}^3 \to \mathbb{P}^3$, $\phi_i : \mathbb{P}^3 \to \mathbb{P}\mathcal{T}^*\mathbb{P}^3 = I \subset \mathbb{P}^3 \times \mathbb{P}^{3*}, q \mapsto (q, \epsilon_i(q)) = (q, X_i(q, -))$, explicitly given, for $q = (x : y : z : t)$, by

$$\epsilon_1 = (y : -x : t : -z), \quad \epsilon_2 = (y : -x : -t : z)$$

$$\epsilon_3 = (z : t : -x : -y), \quad \epsilon_4 = (z : -t : -x : y)$$

$$\epsilon_5 = (t : z : -y : -x), \quad \epsilon_6 = (t : -z : y : -x)$$

$$x_j = X_j(\langle \epsilon_i(q), p \rangle)$$

**Fact** *For a point $q \in \mathbb{P}^3, p \in \mathbb{P}T_q^*\mathbb{P}^3, p \notin \epsilon_i(q)$, the ith Klein coordinate of the line $\langle \epsilon_i(q), p \rangle$ is zero and*

$$p \in \ell_i \cup \ell_i' \Leftrightarrow H_i(p, q) = \sum_{j \neq i} \frac{x_j^2}{\lambda_i - \lambda_j} = 0$$

*with $x_j = X_j(\langle \epsilon_i(q), p \rangle)$.*

**Conclusion** In an affine patch $\mathbb{C}^3 \times \mathbb{C}^{3*} \ni (q, p) = ((x : y : z : 1), (u : v : w : -(xu + yv + zw)))$

$$H_i^a(p, q) = \sum_{j \neq i} \frac{X_j(\epsilon_i(q), p)^2}{\lambda_i - \lambda_j}$$

give six Hitchin Hamiltonians, any three of which are generically independent. The $H_i^a$ have degree $\leq 4$ in $x, y, z$ and are homogeneous of degree 2 in $u, v, w$; they Poisson-commute with respect to $\mathrm{d}x \wedge \mathrm{d}u + \mathrm{d}y \wedge \mathrm{d}v + \mathrm{d}z \wedge \mathrm{d}w$.

**Example** An example is constituted by

$$\mu^2 = (\lambda^2 - 1)(\lambda^2 - 4)(\lambda^2 - 9)$$
$$((x : y : z : 1), (u : v : w : -(xu + yv + zw)))$$
$$\in \boldsymbol{A}^3 \times \boldsymbol{A}^{3*}$$

$$\begin{aligned} H_1 = {} & uv(-70xy - 32x^3y - 18xy^3 - 10z - 32x^2z \\ & + 18y^2z) + v^2(-9 - 30y^2 - 16x^2y^2 - 9y^4 - 32xy^2 \\ & - 16z^2) + u^2(-16 - 40x^2 - 16x^4 - 9x^2y^2 + 18xyz \\ & - 9z^2) + vw(-18x + 10xy^2 + 10yz - 32x^2yz \\ & - 18y^3z - 32xz^2) + uw(32y + 10x^2y - 10xz \\ & - 32x^2z - 18xy^2z + 18yz^2) + w^2(-9x^2 - 16y^2 \\ & + 10xyz - 16x^2z^2 - 9y^2z^2) \end{aligned}$$

The concept of reduction and $r$-matrix have been generalized to Hitchin systems. Notably, Hitchin later showed that the Hamiltonians of the system appear as symbols of a heat operator that corresponds to a projectively flat connection, the quantization of the moduli space of bundles, obtained by changing the complex structure of the Riemann surface $X$.

## Other Aspects

### Special Functions

Special functions have also been traditionally significant in both algebraic geometry and integrable systems. Within the examples presented, elliptic functions gave rise to surprisingly sophisticated theories. The 1-wave solution encountered in the introduction, $u = 2\wp + \text{const.}$ in the limit when one or both periods of the Weierstrass function go to zero, becomes exponential or rational, respectively. The higher-genus analogs give rise to solitons, or rational solutions. On the other hand, the KP solutions which are doubly periodic in the $x$ variable ("elliptic solitons") were classified by Krichever (cf. Dubrovin *et al.* (2001)), as forming an ACI Hamiltonian system ("elliptic Calogero–Moser"), which, 25 years later, is still generating important work, with Hamiltonian

$$H = \sum_{i=1}^{n} p_1^2 + \frac{1}{2} \sum_{i \neq j} \wp(q_i - q_j)$$

(where $\wp$ is the Weierstrass function of a lattice $L$ with associated elliptic curve $X = \mathbb{C}/L$, $q \in X$ the origin) and $u = 2\sum_{i=1}^{n} \wp(x - x_i(t_2, t_3, \ldots))$ is a solution of the KP hierarchy for suitable time flows $t_j$ of the system $(t_1 = x)$ and KP Baker function

$$\psi(x; \alpha) = \frac{\sigma(\alpha - x)}{\sigma(\alpha) - \sigma(x)} \mathrm{e}^{(\zeta(\alpha) \cdot x)}$$

The associated spectral curves have been classified in moduli by Treibich and Verdier (cf. Treibich (2001)); Krichever produced a two-field model as

well as a universal Poisson structure for the system; Donagi and Markman (1996) realized it as a generalized Hitchin system.

More classically, elliptic potentials were the subject of much study, in particular by Lamé and Hermite in the nineteenth century and Ince in the twentieth; a sample result due to Ince makes one feel like Alice in Wonderland, who "knelt down and looked along the passage into the loveliest garden you ever saw": the Lamé operator $L = -\partial^2 + a(a+1)\wp(x - x_0)$ with real, smooth potential is finite gap (namely, almost all the periodic eigenvalues are double) iff $a \in \mathbb{Z}$ (if $a$ is positive the number of gaps is $a$). A generalization to several variables (due to Chalykh and Veselov),

$$L = -\Delta + \sum_{\alpha \in R_+} g_\alpha \wp(\langle \alpha, x \rangle)$$

where $R_+$ is the set of positive roots for a simple complex Lie algebra of rank $n$, $\langle -, - \rangle$ is some scalar product in $\mathbb{R}^n$, invariant under the action of the Weyl group, and $g_\alpha = m_\alpha(m_\alpha + 1)\langle \alpha, \alpha \rangle$ for some $m_\alpha \in \mathbb{Z}$, provides one of the few known examples of quantum completely integrable rings of differential operators in several variables. Roughly speaking, this means that the centralizer of $L$ contains $n$ operators with functionally independent symbols, where $n$ is the number of variables.

What is more, Chalykh *et al.* (2003) combine differential Galois theory and elliptic function theory to characterize (under some mild assumptions) the generalized Lamé operators that are algebraically completely integrable: the differential Galois group of the solutions is abelian.

### Duality, Fourier–Mukai Transform, and Bispectrality

Duality is a concept imported from mathematical physics; as a mathematical phenomenon, it has not reached theoretical maturity. First observed in examples, as in Fock *et al.* (2000), where different definitions of dual ACI Hamiltonian systems were given (action-angle, action–action, and quantum), it resurfaced for the Hitchin system, in more than one guise, whether it be an interchange of position and momentum variables (Gawędzki and Tran-Ngoc-Bich 1998) or a duality between the Lagrangian tori that fiber two such systems, coming from a Fourier–Mukai transform, namely a twist by the (universal) Picard line bundle:

$$\begin{array}{c} \mathcal{P} \\ \downarrow \\ \mathrm{Jac}(X) \times (H^0(X,K) = T^*\mathrm{Jac}(X)) \end{array}$$

Notably, the Picard bundle was used by Nakayashiki to give a spectacular generalization of the Burchnall–Chaundy result for a genus-2 curve $X$ (more generally, Jac($X$) is replaced by a generic abelian variety in the statement): the coordinate ring of $\mathrm{Jac}(X) - \Theta_X$ is the common spectrum of a ring of commuting $(g! \times g!)$ matrix partial differential operators in $g$ variables. The Fourier transform allowed him to extend Sato's correspondence $\partial^{-1} \leftrightarrow z$ and give $\mathcal{F}$ a unique (free, rank-$g!$) $D_{\mathrm{Jac}(X)}$-module structure, where $\mathcal{F}$ is a suitable coherent sheaf over Jac($X$) generalizing the Baker function.

In this model, the interchange of the $x$ and $z$ variables is known as bispectrality (cf. Grünbaum (2001)): a somewhat narrower question is a characterization of the differential operators $L$ in $x$ for which there exists a differential operator $B$ in $k$ and a common eigenfunction:

$$\begin{cases} L\psi(x,k) = f(k)\psi(x,k) \\ B\psi(x,k) = \theta(x)\psi(x,k) \end{cases}$$

for some functions $f, \theta$, typically polynomial. This question proved to be related with the KP hierarchy and isomonodromy deformations. When to a hierarchy there is associated an ACI Hamiltonian system (as in the Neumann case shown above), bispectrality may produce a dual system, in a sense related to the ones discussed, but somewhat mysteriously so.

## Conclusion

Many important mathematical topics and individual contributions regrettably have to go unmentioned in an article of this length. The aim was to illustrate by simplest examples the geometric nature of integrable systems and equations, in the areas of spectral curves, moduli of vector bundles over them, Grassmann manifolds, special functions, Poisson geometry, representation theory, as well as mention constructions that are not yet complete, such as spectral varieties of higher dimension, dualities sweeping vaster moduli spaces, and quantization.

*See also:* Billiards in bounded convex domains; $\bar{\partial}$-Approach to Integrable Systems; Functional Equations and Integrable Systems; Integrable Systems and Discrete Geometry; Integrable Systems and Recursion Operators on Symplectic and Jacobi Manifolds; Integrable Systems and the Inverse Scattering Method; Integrable Systems in Random Matrix Theory; Integrable Systems: Overview; Multi-Hamiltonian Systems; Recursion Operators in Classical Mechanics; Riemann–Hilbert Methods in Integrable Systems; Solitons and Kac–Moody Lie Algebras.

## Further Reading

Adams MR, Harnad J, and Previato E (1988) Isospectral Hamiltonian flows in finite and infinite dimension. *Communications in Mathematical Physics* 117: 451–500.

Adler M and van Moerbeke P (1980) Completely integrable systems, Euclidean Lie algebras and curves. Linearization of Hamiltonian systems, Jacobian varieties and representation theory. *Advances in Mathematics* 38: 267–379.

Babelon O, Bernard D, and Talon M (2003) *Introduction to Classical Integrable Systems*. Cambridge: Cambridge University Press.

Baker HF (1907) *An Introduction to the Theory of Multiply-Periodic Functions*. Cambridge: Cambridge University Press.

Chalykh O, Etingof P, and Oblomkov A (2003) Generalized Lamé operators. *Communications in Mathematical Physics* 239(1–2): 115–153.

Dickey LA (2003) *Soliton Equations and Hamiltonian Systems*, Advanced Series in Mathematical Physics, vol. 26, 2nd edn. River Edge, NJ: World Scientific.

Donagi R and Markman E (1996) Spectral covers, algebraically completely integrable Hamiltonian systems, and moduli of bundles. In: *Integrable Systems and Quantum Groups*, Lecture Notes in Mathematics, pp. 1–119. Berlin: Springer (Montecatini Terme, 1993).

Dubrovin BA, Krichever IM, and Novikov SP (2001) *Integrable Systems I, Dynamical Systems IV*, Encyclopaedia of Mathematical Science, vol. 4, pp. 177–332. Berlin: Springer.

Fock V, Gorsky A, Nekrasov N, and Rubtsov V (2000) Duality in integrable systems and gauge theories. *Journal of High Energy Physics* No. 7, pp. 40.

Françoise JP (1987) Monodromy and the Kovalevskaya top. *Astérique* 150–151; 87–108.

Gawędzki K and Tran-Ngoc-Bich P (1998) Self-duality of the $SL_2$ Hitchin integrable system at genus 2. *Communications in Mathematical Physics* 196(3): 641–670.

Grünbaum FA (2001) The bispectral problem: an overview. *In Special Functions 2000: Current Perspective and Future Directions (Tempe, AZ), 129–140*, NATO Science Series II: Mathematics, Physics, and Chemistry, vol. 30. Dordrecht: Kluwer Academic.

Hitchin N (1982) Monopoles and geodesics. *Communications in Mathematical Physics* 83(4): 579–602.

Hitchin N (1987) Stable bundles and integrable systems. *Duke Mathematical Journal* 54(1): 91–114.

McKean H and Moll V (1997) *Elliptic Curves. Function Theory, Geometry, Arithmetic*. Cambridge: Cambridge University Press.

Moser J (1980) Geometry of quadrics and spectral theory. In: *The Chern Symposium 1979*, pp. 147–188. New York–Berlin: Springer.

Mulase M (1984) Complete integrability of the Kadomtsev–Petviashvili equation. *Advances in Mathematics* 54(1): 57–66.

Mumford D (1984) *Tata Lectures on Theta II, Progr. Math.*, vol. 43. Boston: Birkhäuser.

Nekrasov N (1996) Holomorphic bundles and many-body systems. *Communications in Mathematical Physics* 180(3): 587–603.

Neumann C (1859) De problemat quodam mechanico quad ad primam integralium ultraellipticorum classem revocatur. *J. reine angew Math.* 56: 46–63.

Ol'shanetskiĭ MA, Perelomov AM, Reĭman AG, and Semenov-Tyan-Shanskiĭ MA (1987) *Integrable systems, II. Current Problems in Mathematics. Fundamental Directions*, vol. 16, pp. 86–226, 307; *Itogi Nauki i Tekhniki, Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i Tekhn. Inform.*, Moscow.

Siegel CL (1969) *Topics in Complex Function Theory*, vol. 1. New York: Wiley.

Treibich A (2001) Hyperelliptic tangential covers, and finite-gap potentials. *Uspekhi Matematicheskikh Nauk* 56(6): 89–136.

van Geemen B and Previato E (1996) On the Hitchin system. *Duke Mathematical Journal* 85(3): 659–683.


# Integrable Systems and Discrete Geometry


**A Doliwa**, University of Warmia and Mazury in Olsztyn, Olsztyn, Poland

**P M Santini**, Università di Roma "La Sapienza," Rome, Italy




## Introduction

Although the main subject of this article is the connection between integrable discrete systems and geometry, we feel obliged to begin with the differential part of the relation.

### Classical Differential Geometry and Integrable Systems

The oldest (1840) integrable nonlinear partial differential equation recorded in literature is the Lamé system

$$\frac{\partial^2 H_i}{\partial u_j \partial u_k} - \frac{1}{H_j}\frac{\partial H_j}{\partial u_k}\frac{\partial H_i}{\partial u_j} - \frac{1}{H_k}\frac{\partial H_k}{\partial u_j}\frac{\partial H_i}{\partial u_k} = 0, \quad i, j, k \quad \text{distinct} \qquad [1]$$

$$\frac{\partial}{\partial u_k}\left(\frac{1}{H_k}\frac{\partial H_j}{\partial u_k}\right) + \frac{\partial}{\partial u_j}\left(\frac{1}{H_j}\frac{\partial H_k}{\partial u_j}\right) + \frac{1}{H_i^2}\frac{\partial H_j}{\partial u_i}\frac{\partial H_k}{\partial u_i} = 0 \qquad [2]$$

describing orthogonal coordinates in the three-dimensional Euclidean space $\mathbb{E}^3$ (indices $i, j, k$ range from 1 to 3). Already in 1869, it was found by Ribaucour that the nonlinear Lamé system possesses a discrete symmetry enabling to construct, in a linear way, new solutions of the system from the old ones. He gave also a geometric interpretation of this symmetry in terms of certain spheres tangent to the coordinate surfaces of the triply orthogonal system. In 1918, Bianchi showed that the result of superposition of the Ribaucour transformations is, in a certain sense, independent of the order of their composition.

Such properties of a nonlinear equation are hallmarks of its integrability, and indeed, the Lamé system was solved using soliton techniques in 1997–98. The above example illustrates the close connection between the modern theory of integrable partial differential equations and the differential geometry of the turn of the nineteenth and twentieth centuries. A remarkable property of certain parametrized submanifolds (and then of the corresponding equations) studied that time is that they allow for transformations which exhibit the so-called "Bianchi permutability property." Such transformations called, depending on the context, the Darboux, Calapso, Christoffel, Bianchi, Bäcklund, Laplace,

Koenigs, Moutard, Combescure, Lévy, Goursat, Ribaucour, or the fundamental transformation of Jonas, can be geometrically described in terms of certain families of lines called line congruences.

In the connection between integrable systems and differential geometry, a distinguished role is played by the multidimensional conjugate nets, described by the Darboux system, which is just the first part [1] of the Lamé system with indices ranging form 1 to $N \geq 3$. On the level of integrable systems, this dominant role has the following explanation: the Darboux system, together with equations describing isoconjugate deformations of the net, forms the multicomponent Kadomtsev–Petviashvilii (KP) hierarchy, which is viewed as a master system of equations in soliton theory. In fact, in appropriate variables, the whole multicomponent KP hierarchy can be rewritten as an infinite system of the Darboux equations.

### Transition to the Discrete Domain

The recent progress in studying discrete integrable systems showed that, in many respects, they should be considered as more fundamental than their differential counterparts. Consequently, the natural problem of extending the geometric interpretation of integrable partial differential equations to the discrete domain arose, leading not only to the transition to the discrete domain of many results on the connection between the differential geometry and integrable systems, but also – and this seems to be even more important – to the description of integrability in a very elementary and purely geometric way.

At the level of integrable equations, the transition "from differential to discrete" often makes formulas more complicated and longer. On the contrary, at the geometric level, in such a transition the properties of discrete submanifolds, relevant to their integrability, become simpler and more transparent. Indeed, the mathematics necessary to understand the basic ideas of the integrable discrete geometry does not exceed the "ruler and compass constructions," and many proofs can be performed using elementary incidence geometry.

We will concentrate our attention on the multidimensional lattice made from planar quadrilaterals, which is the discrete analog of a conjugate net. Together with the discussion of its properties, which are the core of the geometric integrability, we briefly present the analytic methods of construction of these lattices and we also describe some basic multidimensional integrable reductions of them. Then we discuss integrable discrete surfaces; some of them have been found in the early period of the "case-by-case" studies. We shall however try to present them, from a unifying perspective, as reductions of the quadrilateral lattice (QL).

## Multidimensional Integrable Lattices

### The Quadrilateral Lattice

An $N$-dimensional lattice $\boldsymbol{x}:\mathbb{Z}^N \to \mathbb{R}^M$ is a lattice made from planar quadrilaterals, or a quadrilateral lattice (QL) in short, if its elementary quadrilaterals $\{\boldsymbol{x}, T_i\boldsymbol{x}, T_j\boldsymbol{x}, T_iT_j\boldsymbol{x}\}$ are planar; that is, iff the following system of discrete Laplace equations is satisfied:

$$\Delta_i\Delta_j\boldsymbol{x} = (T_iA_{ij})\Delta_i\boldsymbol{x} + (T_jA_{ji})\Delta_j\boldsymbol{x}, \quad i \neq j, \; i,j = 1,\dots,N \qquad [3]$$

where $A_{ij}:\mathbb{Z}^N \to \mathbb{R}$ are functions of the discrete variable; here $T_i$ is the translation operator in the $i$th direction, and $\Delta_i = T_i - 1$ is the corresponding difference operator. For simplicity, we work here in the affine setting neglecting projective geometric aspects of the theory.

**The geometric integrability scheme** In the case $N = 2$ the definition [3] allows one to uniquely construct, given two discrete curves intersecting in a common vertex and two functions $A_{12}, A_{21}: \mathbb{Z}^2 \to \mathbb{R}$, a quadrilateral surface. For $N > 2$ the planarity constraints [3] are instead compatible if and only if the geometric data $A_{ij}$ satisfy the nonlinear system

$$\Delta_kA_{ij} + (T_kA_{ij})A_{ik} = (T_jA_{jk})A_{ij} + (T_kA_{kj})A_{ik} \quad i,j,k \quad \text{distinct} \qquad [4]$$

This constraint has a very simple interpretation: in building the elementary cube (see Figure 1), the seven points $\boldsymbol{x}, T_i\boldsymbol{x}, T_j\boldsymbol{x}, T_k\boldsymbol{x}, T_iT_j\boldsymbol{x}, T_iT_k\boldsymbol{x}$, and $T_jT_k\boldsymbol{x}$ ($i,j,k$ are distinct) determine the eighth point $T_iT_jT_k\boldsymbol{x}$ as the unique intersection of three planes in the three-dimensional space.

The connection of this elementary geometric point of view with the classical theory of integrable systems is transparent: the planarity constraint corresponds to the set of linear spectral problems [3] and the resulting QL is characterized by the nonlinear equations [4], arising as the compatibility conditions for such spectral problems. Since the QL equations [4] are a master system in the theory of integrable equations, planarity can be viewed as the elementary geometric root of integrability. The idea

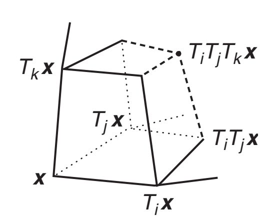



**Figure 1** The geometric integrability scheme.

that integrability be associated with the consistency of a geometric (and/or algebraic) property when increasing the dimensionality of the system is recurrent in the theory of integrable systems.

**Other forms of the Darboux system** The $i \leftrightarrow j$ symmetry of the RHS of eqns [4] implies the existence of the potentials $H_i : \mathbb{Z}^N \to \mathbb{R}$ (the Lamé coefficients) such that

$$A_{ij} = \frac{\Delta_j H_i}{H_i}, \quad i \neq j \qquad [5]$$

and then eqns [4] take the form

$$\Delta_k \Delta_j H_i - \left( T_j \frac{\Delta_k H_j}{H_j} \right) \Delta_j H_i - \left( T_k \frac{\Delta_j H_k}{H_k} \right) \Delta_k H_i = 0, \quad i, j, k \quad \text{distinct} \qquad [6]$$

which is the discrete version of the first part [1] of the Lamé system.

The Lamé coefficients allow to define the suitably normalized tangent vectors $\boldsymbol{X}_i : \mathbb{Z}^N \to \mathbb{R}^M$ by equations

$$\Delta_i \boldsymbol{x} = (T_i H_i) \boldsymbol{X}_i \qquad [7]$$

and the functions $Q_{ij} : \mathbb{Z}^N \to \mathbb{R}, i \neq j$, (the rotation coefficients) by equations

$$\Delta_i H_j = (T_i H_i) Q_{ij}, \quad i \neq j \qquad [8]$$

Then eqns [3] and [6] can be rewritten in the first-order form

$$\Delta_j \boldsymbol{X}_i = (T_j Q_{ij}) \boldsymbol{X}_j, \quad i \neq j \qquad [9]$$

$$\Delta_k Q_{ij} = (T_k Q_{ik}) Q_{kj}, \quad i, j, k \quad \text{distinct} \qquad [10]$$

The discrete Darboux system [10] implies the existence of other potentials $\rho_i$ defined by the compatible equations

$$\frac{T_j \rho_i}{\rho_i} = 1 - (T_i Q_{ji})(T_j Q_{ij}), \quad i \neq j \qquad [11]$$

The $i \leftrightarrow j$ symmetry of the RHS of eqns [11] implies the existence of yet another potential $\tau : \mathbb{Z}^N \to \mathbb{R}$,

$$\rho_i = \frac{T_i \tau}{\tau} \qquad [12]$$

which is called the $\tau$-function of the QL. In terms of the $\tau$-function, and of the functions

$$\tau_{ij} = \tau Q_{ij}, \quad i \neq j \qquad [13]$$

whose geometric interpretation will be given in a later section, the discrete Darboux equations take the following Hirota-type form:

$$(T_i T_j \tau)\tau = (T_i \tau) T_j \tau - (T_i \tau_{ji}) T_j \tau_{ij}, \quad i \neq j \qquad [14]$$

$$(T_k \tau_{ij})\tau = (T_k \tau)\tau_{ij} + (T_k \tau_{ik})\tau_{kj}, \quad i,\ j,\ k \quad \text{distinct} \qquad [15]$$

### Analytic Methods

We will show how one can construct large classes of solutions of the discrete Darboux equations and the corresponding QLs using two basic analytical methods of the soliton theory: the $\bar{\partial}$-dressing method and the algebro-geometric techniques.

**The $\bar{\partial}$-dressing method** Consider the nonlocal $\bar{\partial}$-problem

$$\bar{\partial}\chi(z) + (\hat{R}\chi)(z) = \bar{\partial}\nu(z)$$
$$\lim_{|z| \to \infty} (\chi(z) - \nu(z)) = 0 \qquad [16]$$

where $\bar{\partial} = \partial / \partial \bar{z}$, $\hat{R}$ is the integral operator

$$(\hat{R}\chi)(z) = \int_{\mathbb{C}} R(z, z')\chi(z')\, \mathrm{d}z' \wedge \mathrm{d}\bar{z}'$$

and $\nu(z)$ is a given rational function of $z$.

Let $Q_i^{\pm} \in \mathbb{C}, i = 1, \ldots, N$ be pairs of distinct points of the complex plane, which define the dependence of the kernel $R$ on the discrete variable $n \in \mathbb{Z}^N$:

$$R(z, z'; n) = \prod_{i=1}^{N} \left( \frac{z - Q_i^+}{z - Q_i^-} \right)^{n_i} \times R_0(z, z') \prod_{i=1}^{N} \left( \frac{z' - Q_i^-}{z' - Q_i^+} \right)^{n_i}$$

We consider only kernels $R_0(z, z')$ such that the nonlocal $\bar{\partial}$-problem is uniquely solvable. If $\chi(z; n)$ is the unique solution with the canonical normalization $\nu = 1$, then the function

$$\psi(z; n) = \chi(z; n) \prod_{i=1}^{N} \left( \frac{z - Q_i^-}{z - Q_i^+} \right)^{n_i}$$

satisfies the system of the Laplace equations [3] with the Lamé coefficients given by

$$H_i(n) = \lim_{z \to Q_i^+} \left( \left( \frac{z - Q_i^+}{z - Q_i^-} \right)^{n_i} \psi(z; n) \right)$$

By construction, the system of such Laplace equations is compatible, therefore the Lamé coefficients satisfy eqns [6]. To various $n$-independent measures $\mathrm{d}\mu_a$ on $\mathbb{C}$ there correspond coordinates

$$x^a(n) = \int_{\mathbb{C}} \psi(z; n) \mathrm{d}\mu_a(z)$$

of a QL $\boldsymbol{x}$, having $H_i(n)$ as the Lamé coefficients. To have real lattices, the kernel $R_0$, the points $Q_i^{\pm}$, and the measures $\mathrm{d}\mu_a$ should satisfy certain additional conditions.

One can find a similar interpretation of the normalized tangent vectors $\boldsymbol{X}_i$ and of the rotation

coefficients $Q_{ij}$. If $\chi_i(z;n)$ are the unique solutions of the nonlocal $\bar{\partial}$-problem [16] with the normalizations

$$\nu_i(z;n) = \left(\frac{Q_i^+ - Q_i^-}{z - Q_i^+}\right) \prod_{k=1,k\neq i}^{N} \left(\frac{Q_i^+ - Q_k^+}{Q_i^+ - Q_k^-}\right)^{n_k}$$

then the functions $\psi_i(z;n)$, defined by

$$\psi_i(z;n) = \prod_{k=1}^{N} \left(\frac{z - Q_k^-}{z - Q_k^+}\right)^{n_k} \chi_i(z;n)$$

satisfy the direct analog of the linear problem [9],

$$\Delta_j \psi_i(z;n) = (T_j Q_{ij}(n))\psi_j(z;n), \quad i \neq j \qquad [17]$$

where

$$Q_{ij}(n) = \lim_{z\to Q_j^+} \left( \left(\frac{z - Q_j^+}{z - Q_j^-}\right)^{n_j} \psi_i(z;n)\right)$$

Again, by construction, eqns [17] are compatible and the functions $Q_{ij}(n)$ satisfy the discrete Darboux equations [10]. The functions

$$X_i^a(n) = \int_{\mathbb{C}} \psi_i(z;n)\,\mathrm{d}\mu_a(z)$$

are coordinates of the normalized tangent vectors $\boldsymbol{X}_i$ of the QL $\boldsymbol{x}$ constructed above.

**The algebro-geometric techniques** Given a compact Riemann surface $\mathcal{R}$ of genus $g$, consider a nonspecial divisor $D = \sum_{\alpha=1}^{g} P_\alpha$. Choose $N$ pairs of points $Q_i^\pm \in \mathcal{R}$ and the normalization point $Q_\infty$. Given $n \in \mathbb{Z}^N$, there exists a unique Baker–Akhiezer function $\psi(n)$, defined as a meromorphic function on $\mathcal{R}$, with the following analytical properties: (1) as a function of $P \in \mathcal{R} \setminus \cup_{i=1}^{N} Q_i^\pm$, $\psi(n)$ may have as singularities only simple poles in the points of the divisor $D$; (2) in the points $Q_i^\pm$ function $\psi(n)$ has poles of the order $\pm n_i$; and (3) in the point $Q_\infty$ function $\psi(n)$ is normalized to 1.

When $z_i^\pm(P)$ is a local coordinate on $\mathcal{R}$ centered at $Q_i^\pm$, then condition (2) implies that the function $\psi(n)$ in a neighborhood of the point $Q_i^\pm$ is of the form

$$\psi(P;n) = \left(z_i^\pm(P)\right)^{\mp n_i} \left(\sum_{s=0}^{\infty} \xi_{s,\pm}^i(n)\left(z_i^\pm(P)\right)^s\right) \qquad [18]$$

The Baker–Akhiezer function, as a function of the discrete variable $n \in \mathbb{Z}^N$, satisfies the system of Laplace equations [3] with the Lamé coefficients $H_i(n) = \xi_{0,+}^i(n)$.

Again, by construction, the Lamé coefficients satisfy eqns [6]. To various $n$-independent measures $\mathrm{d}\mu_a$ on $\mathcal{R}$ there correspond coordinates

$$x^a(n) = \int_{\mathcal{R}} \psi(P;n)\,\mathrm{d}\mu_a(P)$$

of a QL $\boldsymbol{x}$.

We present the expression of the Baker–Akhiezer function and of the $\tau$-function of the QL in terms of the Riemann theta functions. Let us choose on $\mathcal{R}$ the canonical basis of cycles $\{a_1,\ldots,a_g,b_1,\ldots,b_g\}$ and the dual basis $\{\omega_1,\ldots,\omega_g\}$ of holomorphic differentials on $\mathcal{R}$, that is, $\oint_{a_j}\omega_k = \delta_{jk}$. Then the matrix $B$ of $b$-periods defined as $B_{jk} = \oint_{b_j}\omega_k$ is symmetric and has positively defined imaginary part. Denote by $\omega_{PQ}$ the unique differential holomorphic in $\mathcal{R}\backslash\{P,Q\}$ with poles of the first order in $P,Q$ and residues, correspondingly, 1 and –1, which is normalized by conditions $\oint_{a_j}\omega_{PQ} = 0$. The Riemann function $\theta(z;B), z \in \mathbb{C}^g$, is defined by its Fourier expansion

$$\theta(z;B) = \sum_{m\in\mathbb{Z}^g} \exp\{\pi \mathrm{i}\langle m, Bm\rangle + 2\pi\mathrm{i}\langle m, z\rangle\}$$

where $\langle\cdot,\cdot\rangle$ denotes the standard bilinear form in $\mathbb{C}^g$. Finally, the Abel map $A$ is given by $A(P) = (\int_{P_0}^{P}\omega_1,\ldots,\int_{P_0}^{P}\omega_g)$, where $P_0 \in \mathcal{R}$, and the Riemann constants vector $K$ is given by

$$K_j = \frac{1 + B_{jj}}{2} - \sum_{k\neq j}\left(\oint_{a_k}\omega_k(P)A_j(P)\omega_j\right), \quad j = 1,\ldots,g$$

The explicit form of the vacuum Baker–Akhiezer function $\psi$ can be written down with the help of the theta functions as follows:

$$\psi(n,P) = \frac{\theta\Big(A(P) + \sum_{k=1}^{N} n_k\big(A(Q_k^-) - A(Q_k^+)\big) + Z\Big)}{\theta\Big(A(Q_\infty) + \sum_{k=1}^{N} n_k\big(A(Q_k^-) - A(Q_k^+)\big) + Z\Big)} \times \frac{\theta(A(Q_\infty) + Z)}{\theta(A(P) + Z)} \exp\left(\sum_{k=1}^{N} n_k \int_{Q_\infty}^{P} \omega_{Q_k^- Q_k^+}\right)$$

where $Z = -\sum_{j=1}^{g} A(P_j) - K$.

Denote by $r_{kj}^\pm$ and $s_{kj}^\pm$ the constants in the decomposition of the abelian integrals near the point $Q_j^\pm$

$$\int_{P_0}^{P} \omega_{Q_k^- Q_k^+} \overset{P\to Q_j^\pm}{=} \mp\delta_{kj}\log z_j^\pm(P) + r_{kj}^\pm + O\Big(z_j^\pm(P)\Big)$$

$$\int_{P_0}^{P} \omega_{Q_\infty Q_k^+} \overset{P\to Q_j^\pm}{=} -\delta_{kj}\delta_{+\pm}\log z_j^\pm(P) + s_{kj}^\pm + O\Big(z_j^\pm(P)\Big)$$

Then the expression of the $\tau$-function of the QL within the subclass of algebro-geometric solutions reads

$$\tau(n) = \theta\left(\sum_{k=1}^{N} n_k\big(A\big(Q_k^-\big) - A\big(Q_k^+\big)\big) + A(Q_\infty) + Z\right) \times \prod_{k,j=1}^{N} \lambda_{kj}^{n_k n_j} \prod_{k=1}^{N} \mu_k^{n_k}$$

where

$$\lambda_{kj} = \exp\left(\frac{r_{kj}^- - r_{kj}^+}{2}\right) = \lambda_{jk}$$

$$\mu_k = \frac{1}{\lambda_{kk}} \frac{\theta(A(Q_k^+) + Z)}{\theta(A(Q_k^-) + Z)} \exp(s_{kk}^- - s_{kk}^+)$$

Finally, we remark that the geometric integrability scheme and the algebro-geometric methods work also in the finite fields setting, giving solutions of the corresponding integrable cellular automata.

### The Darboux-Type Transformations

We present the basic ideas and results of the theory of the Darboux-type transformations of the multidimensional QL.

**Line congruences and the fundamental transformation** To define the transformations we need to define first $N$-dimensional line congruences (or, simply, congruences), which are families of lines in $\mathbb{R}^M$ labeled by points of $\mathbb{Z}^N$ with the property that any two neighboring lines $\mathfrak{l}$ and $T_i\mathfrak{l}, i=1,\ldots,N$, are coplanar and therefore (eventually in the projective extension $\mathbb{P}^M$ of $\mathbb{R}^M$) intersect.

The QL $\mathcal{F}(\boldsymbol{x})$ is a fundamental transform of the QL $\boldsymbol{x}$ if the lines connecting the corresponding points of the lattices form a congruence. The superposition of a number of fundamental transformations can be compactly formulated in the vectorial fundamental transformation. The data of the vectorial fundamental transformation are: (1) the solution $\boldsymbol{Y}_i : \mathbb{Z}^N \to \mathbb{V}$, $\mathbb{V}$ being a linear space, of the linear system [9]; (2) the solution $\boldsymbol{Y}_i^* : \mathbb{Z}^N \to \mathbb{V}^*$, $\mathbb{V}^*$ being the dual of $\mathbb{V}$, of the linear system [8]. These allow to construct the linear operator-valued potential $\boldsymbol{\Omega}(\boldsymbol{Y}, \boldsymbol{Y}^*) : \mathbb{Z}^N \to L(\mathbb{V})$, defined by the following analog of eqn [7]:

$$\Delta_i \boldsymbol{\Omega}(\boldsymbol{Y}, \boldsymbol{Y}^*) = \boldsymbol{Y}_i \otimes (T_i \boldsymbol{Y}_i^*), \quad i = 1, \ldots, N \qquad [19]$$

Similarly, one defines $\boldsymbol{\Omega}(\boldsymbol{X}, \boldsymbol{Y}^*) : \mathbb{Z}^N \to L(\mathbb{V}, \mathbb{R}^M)$ and $\boldsymbol{\Omega}(\boldsymbol{Y}, H) : \mathbb{Z}^N \to \mathbb{V}$. The transforms of the lattice $\boldsymbol{x}$ and other related functions are given by

$$\mathcal{F}(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{\Omega}(\boldsymbol{X}, \boldsymbol{Y}^*)\boldsymbol{\Omega}(\boldsymbol{Y}, \boldsymbol{Y}^*)^{-1}\boldsymbol{\Omega}(\boldsymbol{Y}, H)$$

$$\mathcal{F}(H_i) = H_i - \boldsymbol{Y}_i^* \boldsymbol{\Omega}(\boldsymbol{Y}, \boldsymbol{Y}^*)^{-1}\boldsymbol{\Omega}(\boldsymbol{Y}, H), \quad i = 1, \ldots, N$$

$$\mathcal{F}(\boldsymbol{X}_i) = \boldsymbol{X}_i - \boldsymbol{\Omega}(\boldsymbol{X}, \boldsymbol{Y}^*)\boldsymbol{\Omega}(\boldsymbol{Y}, \boldsymbol{Y}^*)^{-1}\boldsymbol{Y}_i, \quad i = 1, \ldots, N$$

$$\mathcal{F}(Q_{ij}) = Q_{ij} - \boldsymbol{Y}_j^* \boldsymbol{\Omega}(\boldsymbol{Y}, \boldsymbol{Y}^*)^{-1}\boldsymbol{Y}_i, \quad i,j = 1, \ldots, N, \ i \neq j$$

$$\mathcal{F}(\rho_i) = \rho_i\left(1 + (T_i \boldsymbol{Y}_i^*)\boldsymbol{\Omega}(\boldsymbol{Y}, \boldsymbol{Y}^*)\boldsymbol{Y}_i\right), \quad i = 1, \ldots, N$$

$$\mathcal{F}(\tau) = \tau \det \boldsymbol{\Omega}(\boldsymbol{Y}, \boldsymbol{Y}^*)$$

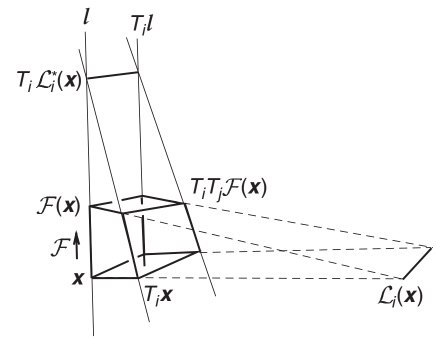



**Figure 2** The fundamental transformation as the binary transformation.

Notice that, by the coplanarity of any two neighboring lines of the congruence, also the quadrilaterals $\{\boldsymbol{x}, T_i\boldsymbol{x}, \mathcal{F}(\boldsymbol{x}), \mathcal{F}(T_i\boldsymbol{x})\}$ are planar (see Figure 2). Then the construction of the transformed lattice mimics the geometric integrability scheme. In consequence, any quadrilateral

$$\{\boldsymbol{x}, \ \mathcal{F}_1(\boldsymbol{x}), \ \mathcal{F}_2(\boldsymbol{x}), \ \mathcal{F}_1(\mathcal{F}_2(\boldsymbol{x})) = \mathcal{F}_2(\mathcal{F}_1(\boldsymbol{x}))\}$$

is planar as well. Therefore, on the discrete level, there is no difference between the lattice coordinate directions and the fundamental transformation directions. The distinction becomes visible in the limit from the QL to the conjugate net. Therefore, the vectorial description of the superposition of the fundamental transformations not only implies their permutability but also provides the explanation of the validity of the practical rule of "integrable discretization by Darboux transformations."

**The Lévy and Combescure transformations** It is easy to see that the family $\mathfrak{t}_i$ of lines passing through the points $\boldsymbol{x}$ and $T_i\boldsymbol{x}$ of a QL forms a congruence, called the $i$th tangent congruence of the lattice. When the congruence of the transformation is the $i$th tangent congruence of the lattice $\boldsymbol{x}$, then the corresponding reduction of the fundamental transformation is called the "Lévy transformation" $\mathcal{L}_i$.

It turns out that, for a generic congruence $\mathfrak{l}$, the lattice made from intersection points of the lines $\mathfrak{l}$ and $T_i^{-1}\mathfrak{l}$ is a QL, called the $i$th focal lattice of the congruence. When the fundamental transform of the lattice $\boldsymbol{x}$ is the $i$th focal lattice of the transformation congruence, then the corresponding reduction of the fundamental transformation is called the "adjoint Lévy transformation" $\mathcal{L}_i^*$.

Both Lévy transformations use only a half of the fundamental transformation data, and the corresponding reduction formulas (in the scalar case) for the lattice points read as follows:

$$\mathcal{L}_i(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{X}_i(Y_i)^{-1}\boldsymbol{\Omega}(Y, H)$$

$$\mathcal{L}_i^*(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{\Omega}(\boldsymbol{X}, Y^*)\,(Y_i^*)^{-1} H_i$$

Notice that the composition of the Lévy and the adjoint Lévy transformations gives (see Figure 2) the fundamental transformation, also called, for this reason, the binary transformation.

Another reduction of the fundamental transformation, important from a technical point of view, is the "Combescure transformation," in which the tangent lines of the transformed lattice $\mathcal{C}(\boldsymbol{x})$ are parallel to those of the lattice $\boldsymbol{x}$. The transformation formula reads

$$\mathcal{C}(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{\Omega}(\boldsymbol{X}, Y^*)$$

where only the solution $Y^*$ of the adjoint linear system [8], necessary to build the transformation congruence, is needed.

**The Laplace transformations and the geometric meaning of the Hirota equation** The Laplace transform $\mathcal{L}_{ij}(\boldsymbol{x}), i \neq j$, of the QL $\boldsymbol{x}$ is the $j$th focal lattice of its $i$th tangent congruence (see Figure 3). It is uniquely determined once the lattice $\boldsymbol{x}$ is given. The transformation formulas of the lattice points and of the $\tau$-function read as follows:

$$\mathcal{L}_{ij}(\boldsymbol{x}) = \boldsymbol{x} - \frac{1}{A_{ji}} \Delta_i \boldsymbol{x} \qquad [20]$$

$$\mathcal{L}_{ij}(\tau) = \tau_{ij} = \tau Q_{ij} \qquad [21]$$

The superpositions of Laplace transformations satisfy the following identities

$$\mathcal{L}_{ij} \circ \mathcal{L}_{ji} = \mathrm{id}$$
$$\mathcal{L}_{jk} \circ \mathcal{L}_{ij} = \mathcal{L}_{ik}$$
$$\mathcal{L}_{ki} \circ \mathcal{L}_{ij} = \mathcal{L}_{kj}$$

which allow to identify them with the Schlesinger transformations of the monodromy theory.

In the simplest case $N = 2$ one obtains the so-called Laplace sequence of two-dimensional QLs

$$\boldsymbol{x}_\ell = \mathcal{L}_{12}^\ell(\boldsymbol{x}), \qquad \tau_\ell = \mathcal{L}_{12}^\ell(\tau)$$
$$\mathcal{L}_{12}^{-1} = \mathcal{L}_{21}, \quad \ell \in \mathbb{Z}$$

Equations [14] and [21] imply that the $\tau$-functions of the Laplace sequence satisfy the celebrated Hirota equation (the fully discrete Toda system)

$$\tau_\ell T_1 T_2 \tau_\ell = (T_1 \tau_\ell)(T_2 \tau_\ell) - (T_1 \tau_{\ell-1})(T_2 \tau_{\ell+1})$$

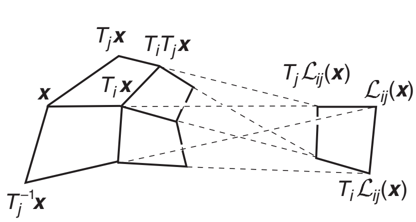



**Figure 3** The Laplace transformation $\mathcal{L}_{ij}$.

**Distinguished Integrable Reductions**

We will present here basic reductions of the multidimensional QL. The geometric criterion for their integrability is the compatibility with the geometric integrability scheme.

**The circular lattices and the Ribaucour congruences** QLs $\mathbb{Z}^N \to \mathbb{E}^M$ for which each quadrilateral is inscribed in a circle are called "circular" lattices. They are the integrable discrete analogs of submanifolds parametrized by curvature coordinates (e.g., the orthogonal coordinate systems described by the Laméequations [1]–[2]).

The integrability of circular lattices is the consequence of the fact that if the three "initial" quadrilaterals $\{\boldsymbol{x}, T_i\boldsymbol{x}, T_j\boldsymbol{x}, T_iT_j\boldsymbol{x}\}$, $\{\boldsymbol{x}, T_i\boldsymbol{x}, T_k\boldsymbol{x}, T_iT_k\boldsymbol{x}\}$, $\{\boldsymbol{x}, T_j\boldsymbol{x}, T_k\boldsymbol{x}, T_jT_k\boldsymbol{x}\}$ are circular, then also the three new quadrilaterals constructed by adding the vertex $T_iT_jT_k\boldsymbol{x}$ are circular as well (see Figure 4). In fact, all the eight vertices belong to a sphere, and, in consequence, all the vertices of any $K$-dimensional, $K = 2, \ldots, N$, elementary cell belong to a $(K-1)$-dimensional sphere.

There are various equivalent algebraic descriptions of the circular lattices:

1. the normalized tangent vectors $\boldsymbol{X}_i$ satisfy the constraint
$$\boldsymbol{X}_i \cdot T_i\boldsymbol{X}_j + \boldsymbol{X}_j \cdot T_j\boldsymbol{X}_i = 0, \quad i \neq j$$
2. the scalar function $\boldsymbol{x} \cdot \boldsymbol{x} : \mathbb{Z}^N \to \mathbb{R}$ satisfies the Laplace equations [3] of the lattice $\boldsymbol{x}$;
3. the functions $X_i^\circ = (\boldsymbol{x} + T_i\boldsymbol{x}) \cdot \boldsymbol{X}_i : \mathbb{Z}^N \to \mathbb{R}$ satisfy the same linear system [9] as the normalized tangent vectors $\boldsymbol{X}_i$; and
4. the functions $\boldsymbol{X}_i \cdot \boldsymbol{X}_i : \mathbb{Z}^N \to \mathbb{R}$ satisfy eqns [11] and thus can serve as the potentials $\rho_i$.

The Ribaucour transformation $\mathcal{R}$ is the restriction of the fundamental transformation to the class of circular lattices such that also the "side" quadrilaterals $\{\boldsymbol{x}, T_i\boldsymbol{x}, \mathcal{R}(\boldsymbol{x}), \mathcal{R}(T_i\boldsymbol{x})\}$ are circular. Again there is no geometric difference between the lattice directions and the Ribaucour transformation direction. Moreover, the quadrilaterals $\{\boldsymbol{x}, \mathcal{R}_1(\boldsymbol{x}),$

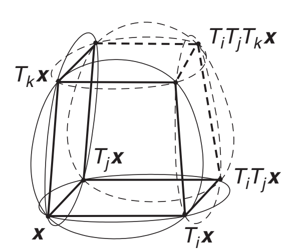



**Figure 4** The geometric integrability of circular lattices.

$\mathcal{R}_2(\boldsymbol{x}), \mathcal{R}_1(\mathcal{R}_2(\boldsymbol{x})) = \mathcal{R}_2(\mathcal{R}_1(\boldsymbol{x}))\}$ are circular as well. In consequence, the vertices of the elementary $K$-cells, $K = 2, \ldots, N$, of the circular lattice and the corresponding vertices of its Ribaucour transform are contained in a $K$-dimensional sphere. Finally, for $K = N$, one obtains a special $\mathbb{Z}^N$ family of $N$-dimensional spheres, called the Ribaucour congruence of spheres.

Algebraically, the Ribaucour transformation needs only a half of the data (necessary to build the congruence) of the fundamental transformation. The data of the vectorial Ribaucour transformation consists of the solution $\boldsymbol{Y}_i^*: \mathbb{Z}^N \to \mathbb{V}^*$, of the linear system [8]. Then, because of the circularity constraint, $\boldsymbol{Y}_i: \mathbb{Z}^N \to \mathbb{V}$ given by

$$\boldsymbol{Y}_i = (\boldsymbol{\Omega}(\boldsymbol{X}, \boldsymbol{Y}^*) + T_i\boldsymbol{\Omega}(\boldsymbol{X}, \boldsymbol{Y}^*))^{\mathrm{T}}\boldsymbol{X}_i$$

is a solution of the linear system [9], and the constraints

$$\boldsymbol{\Omega}(\boldsymbol{Y}, H) + \boldsymbol{\Omega}(X^{\circ}, \boldsymbol{Y}^*)^{\mathrm{T}} = 2\,\boldsymbol{\Omega}(\boldsymbol{X}, \boldsymbol{Y}^*)^{\mathrm{T}}\boldsymbol{x}$$
$$\boldsymbol{\Omega}(\boldsymbol{Y}, \boldsymbol{Y}^*) + \boldsymbol{\Omega}(\boldsymbol{Y}, \boldsymbol{Y}^*)^{\mathrm{T}} = 2\,\boldsymbol{\Omega}(\boldsymbol{X}, \boldsymbol{Y}^*)^{\mathrm{T}}\boldsymbol{\Omega}(\boldsymbol{X}, \boldsymbol{Y}^*)$$

are admissible.

We remark that the above constraints have a simple geometric meaning when one considers the circular lattices in $\mathbb{E}^M$ as the stereographic projections of QLs in the Möbius sphere $S^M$; that is, as a special case of QLs subjected to quadratic constraints.

**The symmetric lattice** Given a QL $\boldsymbol{x}$ with rotation coefficients $Q_{ij}$ and potentials $\rho_i$ given by [11], then the functions $\tilde{Q}_{ij}$, defined by equation

$$\rho_j T_j \tilde{Q}_{ij} = \rho_i T_i Q_{ji}, \quad i \neq j$$

and called, because of their geometric interpretation, the backward rotation coefficients, satisfy the Darboux system [10] as well. A QL is called symmetric if its forward rotation coefficients $Q_{ij}$ are also its backward rotation coefficients. Again the constraint is compatible with the geometric integrability scheme, that is, it propagates in the construction of the lattice. One can show that a QL is symmetric if and only if its rotation coefficients satisfy the following trilinear constraint:

$$(T_iQ_{ji})(T_jQ_{kj})(T_kQ_{ik}) = (T_jQ_{ij})(T_iQ_{ki})(T_kQ_{jk})$$
$$i, j, k \quad \text{distinct}$$

To obtain the corresponding reduction of the fundamental transformation we again need only half of the data. Given a solution $\boldsymbol{Y}_i^*: \mathbb{Z}^N \to \mathbb{V}^*$, of the linear system [8], then, because of the symmetric constraint, $\boldsymbol{Y}_i: \mathbb{Z}^N \to \mathbb{V}$, defined by

$$\boldsymbol{Y}_i = \rho_i (T_i \boldsymbol{Y}^*)^{\mathrm{T}}$$

is the solution of the linear system [9]; notice that, equivalently, we could start from $\boldsymbol{Y}_i$. The constraint

$$\boldsymbol{\Omega}(\boldsymbol{Y}, \boldsymbol{Y}^*) = \boldsymbol{\Omega}(\boldsymbol{Y}, \boldsymbol{Y}^*)^{\mathrm{T}}$$

is then admissible and gives a new symmetric lattice.

There are other multidimensional reductions of the QL like, for example, the $D$-invariant and Egorov lattices or discrete versions of immersions of spaces of constant negative curvature. We remark that the transformations and reductions discussed above have also a clear interpretation on the level of the analytic methods.

## Integrable Discrete Surfaces

In this section we present some distinguished examples of discrete integrable surfaces. Notice that, although the geometric integrability scheme is meaningless for $N = 2$, it can be applied indirectly, by considering the discrete surfaces, together with their transformations, as sublattices of multidimensional lattices.

We remark also that one can consider integrable evolutions of discrete curves, which give equations of the Ablowitz–Ladik hierarchy, and the corresponding integrable spin chains.

### Discrete Isothermic Nets

An isothermic lattice is a two-dimensional circular lattice $\boldsymbol{x}: \mathbb{Z}^2 \to \mathbb{E}^M$ with harmonic quadrilaterals; that is, given $\boldsymbol{x}, T_1\boldsymbol{x}$ and $T_2\boldsymbol{x}$, then the point $T_1T_2\boldsymbol{x}$ is the intersection of the circle (passing through $\boldsymbol{x}, T_1\boldsymbol{x}$ and $T_2\boldsymbol{x}$) and the line passing through $\boldsymbol{x}$ and the meeting point of the tangents to the circle at $T_1\boldsymbol{x}$ and $T_2\boldsymbol{x}$ (see Figure 5). Therefore, given two discrete curves intersecting in the common vertex $\boldsymbol{x}_0$, the unique isothermic lattice can be found using the above "ruler and compass" construction.

Algebraically the reduction looks as follows. Any oriented plane in $\mathbb{E}^M$ can be identified with the complex plane $\mathbb{C}$. Given any four complex points $z_1, z_2, z_3$, and $z_4$, their complex cross-ratio is defined by

$$q(z_1, z_2, z_3, z_4) = \frac{(z_1 - z_2)(z_3 - z_4)}{(z_2 - z_3)(z_4 - z_1)}$$

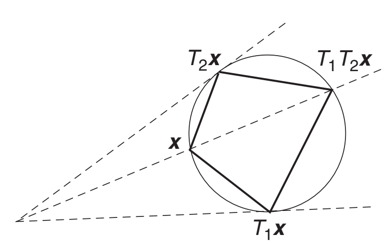



**Figure 5** Elementary quadrilaterals of the isothermic lattice.

One can show that the cross-ratio is real if and only if the four points are cocircular or collinear. In particular, a harmonic quadrilateral with vertices numbered anticlockwise has cross-ratio equal to –1. Therefore, abusing the notation (it can be formalized using Clifford algebras), the isothermic lattice is defined by the condition

$$q(\boldsymbol{x}, T_1\boldsymbol{x}, T_1T_2\boldsymbol{x}, T_2\boldsymbol{x}) = -1$$

We remark that the definition of isothermic lattices can be slightly generalized allowing for the above cross-ratio to be a ratio of two real functions of single discrete variables.

The restriction of the Ribaucour transformation to the class of isothermic lattices, named after Darboux who constructed it for isothermic surfaces, has as its data a real parameter $\lambda$ and the starting point $\mathcal{D}(\boldsymbol{x}_0)$, and can be described as follows. Given the elementary quadrilateral $\{\boldsymbol{x}, T_1\boldsymbol{x}, T_2\boldsymbol{x}, T_1T_2\boldsymbol{x}\}$ of the isothermic lattice, and given the point $\mathcal{D}(\boldsymbol{x})$, then the points $\mathcal{D}(T_1\boldsymbol{x})$ and $\mathcal{D}(T_2\boldsymbol{x})$ belong to the corresponding planes and are constructed from equations

$$q(\boldsymbol{x}, \mathcal{D}(\boldsymbol{x}), \mathcal{D}(T_1\boldsymbol{x}), T_1\boldsymbol{x}) = \lambda$$
$$q(\boldsymbol{x}, \mathcal{D}(\boldsymbol{x}), \mathcal{D}(T_2\boldsymbol{x}), T_2\boldsymbol{x}) = -\lambda$$

It turns out that the point $\mathcal{D}(T_1T_2\boldsymbol{x})$, constructed by the application of the geometric integrability scheme, is such that the quadrilateral $\{\mathcal{D}(\boldsymbol{x}), \mathcal{D}(T_1\boldsymbol{x}), \mathcal{D}(T_2\boldsymbol{x}), \mathcal{D}(T_1T_2\boldsymbol{x})\}$ is harmonic. Moreover, the construction of the Darboux transformation is compatible; that is, the new side quadrilaterals have the correct cross-ratios $\lambda$ and $-\lambda$.

There are various integrable reductions of the isothermic lattice, for example, the constant mean curvature lattice and the minimal lattice.

## Asymptotic Lattices and Their Reductions

An asymptotic lattice is a mapping $\boldsymbol{x}: \mathbb{Z}^2 \to \mathbb{R}^3$ such that any point $\boldsymbol{x}$ of the lattice is coplanar with its four nearest neighbors $T_1\boldsymbol{x}, T_2\boldsymbol{x}, T_1^{-1}\boldsymbol{x}, T_2^{-1}\boldsymbol{x}$ (see Figure 6). Such a plane is called the tangent plane of the asymptotic lattice in the point $\boldsymbol{x}$.

It can be shown that any asymptotic lattice $\boldsymbol{x}$ can be recovered from its suitably rescaled normal (to the tangent plane) field $\boldsymbol{N}: \mathbb{Z}^2 \to \mathbb{R}^3$ via the discrete analog of the Lelieuvre formulas

$$\Delta_1\boldsymbol{x} = (T_1\boldsymbol{N}) \times \boldsymbol{N}, \qquad \Delta_2\boldsymbol{x} = \boldsymbol{N} \times (T_2\boldsymbol{N}) \qquad [22]$$

By the compatibility of the Lelieuvre formulas, the normal field $\boldsymbol{N}$ satisfies the discrete Moutard equation

$$T_1T_2\boldsymbol{N} + \boldsymbol{N} = F(T_1\boldsymbol{N} + T_2\boldsymbol{N}) \qquad [23]$$

for some potential $F: \mathbb{Z}^2 \to \mathbb{R}$.

Given a scalar solution $\theta$ of the Moutard equation [23], a new solution $\mathcal{M}(\boldsymbol{N})$ of the Moutard equation, with the new potential

$$\mathcal{M}(F) = \frac{(T_1\theta)(T_2\theta)}{(T_1T_2\theta)\theta} F$$

can be found via the Moutard transformation equations

$$\mathcal{M}(T_1\boldsymbol{N}) \mp \boldsymbol{N} = \frac{\theta}{T_1\theta}(\mathcal{M}(\boldsymbol{N}) \mp T_1\boldsymbol{N}) \qquad [24]$$

$$\mathcal{M}(T_2\boldsymbol{N}) \pm \boldsymbol{N} = \frac{\theta}{T_2\theta}(\mathcal{M}(\boldsymbol{N}) \pm T_2\boldsymbol{N}) \qquad [25]$$

Now, via the Lelieuvre formulas [22], one can construct a new asymptotic lattice $\mathcal{M}(\boldsymbol{x}) = \boldsymbol{x} \pm \mathcal{M}(\boldsymbol{N}) \times \boldsymbol{N}$. The lines connecting corresponding points of the asymptotic lattices $\boldsymbol{x}$ and $\mathcal{M}(\boldsymbol{x})$ are tangent to both lattices. Such a $\mathbb{Z}^2$-family of lines in $\mathbb{R}^3$ is called Weingarten (or $W$ for short) congruence. Notice that this is not a congruence as considered earlier.

Various integrable reductions of asymptotic lattices are known in the literature: pseudospherical lattices, asymptotic Bianchi lattices and isothermally asymptotic (or Fubini–Ragazzi) lattices, and discrete (proper and improper) affine spheres.

Formally, the Moutard transformation is a reduction of the (projective version of the) fundamental transformation for the Moutard reduction of the Laplace equation. However, the geometric relation between asymptotic lattices and QLs is more subtle and the geometric scenery of this connection is the line geometry of Plücker. Straight lines in $\mathbb{R}^3 \subset \mathbb{P}^3$ are considered there as points of the so-called Plücker quadric $\mathcal{Q}_P \subset \mathbb{P}^5$. A discrete asymptotic net in $\mathbb{P}^3$, viewed as the envelope of its tangent planes, corresponds to a congruence of isotropic lines in $\mathcal{Q}_P$, whose focal lattices represent the asymptotic directions. The discrete $W$-congruences are represented by two-dimensional QLs in the Plücker quadric.

## The Koenigs Lattice

A two-dimensional QL $\boldsymbol{x}: \mathbb{Z}^2 \to \mathbb{P}^M$ is called a Koenigs lattice if, for every point $\boldsymbol{x}$ of the lattice,

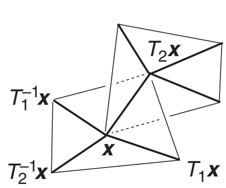



**Figure 6** Asymptotic lattices.

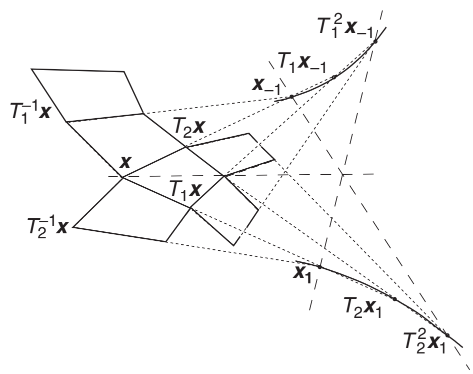



**Figure 7** The Koenigs lattice.

the six points $x_{\pm 1}, T_i x_{\pm 1}, T_i^2 x_{\pm 1}, i = 1, 2,$ of its Laplace transforms belong to a conic (see Figure 7). The nonlinear constraint in definition of the Koenigs lattice can be linearized, with the help of the Pascal "mystic hexagon" theorem, to the form that the line passing through $x$ and $T_1 T_2 x$, the line passing through $x_1$ and $T_1^2 x_{-1}$, and the line passing through $x_{-1}$ and $T_2^2 x_1$ intersect in a point.

Algebraically, the geometric Koenigs lattice condition means that the Laplace equation of the lattice in homogeneous coordinates $x : \mathbb{Z}^2 \to \mathbb{R}_*^{M+1}$ can be gauged into the form

$$T_1 T_2 x + x = T_1(Fx) + T_2(Fx) \qquad [26]$$

It turns out that, if $N$ is a solution of the Moutard equation [23], then $x = T_1 N + T_2 N$ satisfies the Koenigs lattice equation. Therefore, the algebraic theory of the discrete Koenigs lattice equation [26], its (Koenigs) transformation, and the permutability of the superpositions of such transformations is based on the corresponding theory for the Moutard equation [23].

Geometrically, the Koenigs lattices are selected from the QLs as follows. Given a two-dimensional QL $x : \mathbb{Z}^2 \to \mathbb{P}^M$ and given a congruence $\mathfrak{l}$ with lines passing through the corresponding points of the lattice. Denote by $y_i = T_i^{-1}\mathfrak{l} \cap \mathfrak{l}, i = 1, 2,$ points of the focal lattices of the congruence. For every line $\mathfrak{l}$, denote by $\imath$ the unique projective involution exchanging $y_i$ with $T_i y_i$. If, for every congruence $\mathfrak{l}$, the lattice $\mathcal{K}(x) : \mathbb{Z}^2 \to \mathbb{P}^M$, with points $\mathcal{K}(x) = \imath(x)$, is a QL, then the lattice $x$ is a Koenigs lattice. The above construction gives also the corresponding reduction of the fundamental transformation.

A distinguished reduction of the Koenigs lattice is the quadrilateral Bianchi lattice. The natural continuous limit of the corresponding equation is equivalent to the Bianchi (or hyperbolic Ernst) system describing the interaction of planar gravitational waves.

### Discrete Two-Dimensional Schrödinger Equation

In the previous sections we have discussed examples of integrable discrete geometries described by equations of hyperbolic type. Below we present some results associated with the elliptic case; it is remarkable that the QL provides a way to connect these two subjects.

Consider a solution $N : \mathbb{Z}^2 \to \mathbb{R}^3$ of the general self-adjoint five-point scheme on the star of the $\mathbb{Z}^2$ lattice

$$aT_1 N + T_1^{-1}(aN) + bT_2 N + T_2^{-1}(bN) - cN = 0 \quad [27]$$

then the lattice $x : \mathbb{Z}^2 \to \mathbb{R}^3$ obtained by the Lelieuvre type formulas

$$\begin{aligned} \Delta_1 x &= -\left(T_2^{-1} b\right) N \times T_2^{-1} N \\ \Delta_2 x &= \left(T_1^{-1} a\right) N \times T_1^{-1} N \end{aligned} \qquad [28]$$

is a QL having $N$ as normal (to the planes of elementary quadrilaterals) vector field.

The following gauge-equivalent form of eqn 27, namely

$$\frac{\Gamma}{T_1\Gamma} T_1\psi + T_1^{-1}\left(\frac{\Gamma}{T_1\Gamma}\psi\right) + \frac{\Gamma}{T_2\Gamma} T_2\psi + T_2^{-1}\left(\frac{\Gamma}{T_2\Gamma}\psi\right) - q\psi = 0 \qquad [29]$$

an integrable discretization of the Schrödinger equation

$$\frac{\partial^2 \psi}{\partial x_1^2} + \frac{\partial^2 \psi}{\partial x_2^2} - Q\psi = 0$$

is also the Lax operator associated with an integrable generalization of the Toda law to the square lattice.

The five-point scheme [27] is also a distinguished illustrative example of the sublattice theory. Indeed, it can be obtained restricting to the even sublattice $\mathbb{Z}_e^2$ the discrete Cauchy–Riemann equations

$$T_1 T_2 \phi - \phi = \mathrm{i}G(T_1\phi - T_2\phi) \qquad [30]$$

Because of the equivalence (on the discrete level!) between eqn [30] and the discrete Moutard equation [23], the five-point scheme [27] inherits integrability properties (Darboux-type transformations, superposition formulas, analytic methods of solution) from the corresponding (and simpler) integrability properties of the discrete Moutard equation.

*See also:* Bäcklund Transformations; $\bar{\partial}$-Approach to Integrable Systems; Integrable Discrete Systems; Integrable Systems and Algebraic Geometry; Integrable Systems and the Inverse Scattering Method; Integrable Systems: Overview; Nonlinear Schrödinger Equations; Sine-Gordon Equation; Stability Theory and KAM; Toda Lattices.

## Further Reading


Akhmetshin AA, Krichever IM, and Volvovski YS (1999) Discrete analogs of the Darboux–Egoroff metrics. *Proceedings of the Steklov Institute of Mathematics* 225: 16–39.

Białecki M and Doliwa A (2005) Algebro-geometric solution of the discrete KP equation over a finite field out of a hyperelliptic curve. *Communications in Mathematical Physics* 253: 157–170.

Bobenko AI (2004) Discrete differential geometry. Integrability as consistency. In: Grammaticos B, Kosmann–Schwarzbach Y, and Tamizhmani T (eds.) *Discrete Integrable Systems*, pp. 85–110. Berlin: Springer.

Bobenko AI and Seiler R (eds.) (1999) *Discrete Integrable Geometry and Physics*. Oxford: Clarendon.

Bogdanov LV and Konopelchenko BG (1995) Lattice and $q$-difference Darboux–Zakharov–Manakov systems via $\bar{\partial}$ method. *Journal of Physics A* 28: L173–L178.

Cieśliński J (1997) The spectral interpretation of $N$-spaces of constant negative curvature immersed in $\mathbb{R}^{2N-1}$. *Physics Letters A* 236: 425–430.

Doliwa A, Grinevich PG, Nieszporski M, and Santini PM (2004) Integrable lattices and their sublattices: from the discrete Moutard (discrete Cauchy–Riemann) 4-point equation to the self-adjoint 5-point scheme, nlin.SI/0410046.

Doliwa A, Mañas M, Martínez Alonso L, Medina E, and Santini PM (1999) Charged free fermions, vertex operators and transformation theory of conjugate nets. *Journal of Physics A* 32: 1197–1216.

Doliwa A, Nieszporski M, and Santini PM (2001) Asymptotic lattices and their integrable reductions. I. The Bianchi and the Fubini–Ragazzi lattices. *Journal of Physics A* 34: 10423–10439.

Doliwa A, Nieszporski M, and Santini PM (2004) Geometric discretization of the Bianchi system. *Journal of Geometry and Physics* 52: 217–240.

Doliwa A and Santini PM (2000) The symmetric, $D$-invariant and Egorov reductions of the quadrilateral latice. *Journal of Geometry and Physics* 36: 60–102.

Doliwa A, Santini PM, and Mañas M (2000) Transformations of quadrilateral lattices. *Journal of Mathematical Physics* 41: 944–990.

Klimczewski P, Nieszporski M, and Sym A (2000) Luigi Bianchi, Pasquale Calapso and solitons. *Rend. Sem. Mat. Messina, Atti del Congresso Internazionale in Onore di Pasquale Calapso*, Messina, 12–14 October 1998, pp. 223–240.

Mañas M (2001) Fundamental transformation for quadrilateral lattices: first potentials and $\tau$-functions, symmetric and pseudo-Egorov reductions. *Journal of Physics A* 34: 10413–10421.

Matsuura N and Urakawa H (2003) Discrete improper affine spheres. *Journal of Geometry and Physics* 45: 164–183.

Rogers C and Schief WK (2002) *Bäcklund and Darboux Transformations. Geometry and Modern Applications in Soliton Theory*.Cambridge: Cambridge University Press.

Schief WK (2003a) Lattice geometry of the discrete Darboux, KP, BKP and CKP equations. Menelaus' and Carnot's theorems. *Journal of Nonlinear Mathematical Physics* 10(suppl. 2): 194–208.

Schief WK (2003b) On the unification of classical and novel integrable surfaces. II. Difference geometry. *Proceedings of the Royal Society of London A* 459: 373–391.


# Integrable Systems and Recursion Operators on Symplectic and Jacobi Manifolds

**R Caseiro and J M Nunes da Costa**, Universidade de Coimbra, Coimbra, Portugal



## Introduction

Let $(M, \omega)$ be a symplectic manifold of dimension $2n$. We denote by $\sharp$ the natural isomorphism between $T^*M$ and $TM$, defined by the equation

$$i_{\sharp_\alpha}\omega = -\alpha, \quad \alpha \in T^*M \qquad [1]$$

We say that ${}^\sharp df$ is the Hamiltonian vector field defined by the Hamiltonian $f : M \to \mathbb{R}$.

Associated with the nondegenerated closed 2-form $\omega$ there is also a Poisson bracket on $C^\infty(M)$, the space of real differentiable functions on $M$, defined by

$$\{.,.\}_\omega : C^\infty(M) \times C^\infty(M) \to C^\infty(M)$$
$$(f, g) \mapsto \{f, g\}_\omega = \omega({}^\sharp df, {}^\sharp dg)$$

We say that two smooth functions $F, G : M \to \mathbb{R}$ are in involution if

$$\{F, G\}_\omega = 0 \qquad [2]$$

Suppose we have $n$ independent smooth functions in involution $H_1, \ldots, H_n$, such that the associated Hamiltonian vector fields $X_1, \ldots, X_n$ are complete on the level manifold

$$M_a = \{x \in M : H_j(x) = a_j, j = 1, \ldots, n\} \qquad [3]$$

The classical theorem of Arnol'd–Liouville states that

1. the submanifold $M_a$ is invariant with respect to each one of the Hamiltonian commuting flows generated by $H_1, \ldots, H_n$;
2. every connected component of $M_a$ is diffeomorphic to a product of a Euclidean space by a torus, $\mathbb{R}^{n-k} \times \mathbb{T}^k$;
3. there exist coordinates $f_1, \ldots, f_{n-k}, \varphi_1, \ldots, \varphi_k$ in $M_a$ such that the Hamiltonian systems in $M_a$, associated with the Hamiltonians $H_j$, have the form

$$\dot{f}_s = c^j_s \dot{\varphi}_m = \omega^j_m \quad (\omega \equiv \omega(a),\ c = \text{const.}) \qquad [4]$$

4. if $M_a$ is compact then it is diffeomorphic to $\mathbb{T}^n$ and there exists a neighborhood of $M_a$ on $M$, symplectically diffeomorphic to $\mathbb{B}^n \times \mathbb{T}^n$.

A completely integrable Hamiltonian system is a Hamiltonian vector field $X$, that admits $n$ integrals $H_1, \ldots, H_n$ satisfying the hypothesis of Arnol'd–Liouville theorem.

It may happen that a system has more than $n$ independent integrals of motion. In this case it is called superintegrable and not all the integrals are in involution. Supposing that

$$M_a = \{x \in M : H_j(x) = a_j, j = 1, \ldots, n+k\}$$

is compact and connected and that $H_1, \ldots, H_{n-k}$ commute with all the $n+k$ integrals, then $M_a$ is diffeomorphic to the torus $\mathbb{T}^{n-k}$. In particular, if the system is maximally superintegrable, that is, $k = n-1, M_a$ is diffeomorphic to $T^1 = S^1$ and all the trajectories are closed.

To prove that a system is completely integrable, we have to find a sufficient number of integrals of the system in involution. The Lax pair is an extremely powerful tool in this task, although it does not guarantee the involution of the integrals found.

A Lax pair of a vector field $X$ on a smooth manifold $M$ is a pair of operators $(L, M)$ such that

$$\dot{L} = [M, L] = ML - LM \qquad [5]$$

This equation is equivalent to

$$U^{-1}LU = L_0 \qquad [6]$$

where $U$ is the solution operator of the Cauchy problem

$$\dot{U} = MU, \quad U(0) = I \qquad [7]$$

So, the eigenvalues of $L$ are integrals of $X$. Notice that all the pairs $(L^k, M), k \in \mathbb{N}$, are Lax pairs of the system and we may conclude that the functions $\operatorname{tr} L^k, k \in \mathbb{N}$, are integrals of $X$.

The first goal of this article is to relate integrable Hamiltonian systems and recursion operators, where some of the most important properties of the latter are exhibited. Very naturally, the Poisson–Nijenhuis manifolds appear in this context and the Toda lattice is the example chosen in order to show the whole theory working in practice. Also, we see how recursion operators can help in the construction of quadratic algebras of integrals of motion and, in the last section, we present the generalization to Jacobi manifolds of the Nijenhuis structures defined for Poisson manifolds.

## Integrable Systems on Poisson–Nijenhuis Manifolds

Let $X$ be a vector field on a smooth manifold $M$. A recursion operator of $X$ is a $(1,1)$-tensor $R$ invariant of $X$:

$$\mathcal{L}_X R = 0 \qquad [8]$$

The $(1,1)$-tensors, and in particular the recursion operators, may be regarded as fiber endomorphisms of $TM$. So, given a $(1,1)$-tensor $R$, we denote by ${}^tR : T^*M \to T^*M$ the transpose of $R : TM \to TM$, that is,

$$\langle {}^tR(\alpha), X\rangle = \langle \alpha, R(X)\rangle, \quad \alpha \in T^*M,\ X \in TM \qquad [9]$$

where $\langle .,. \rangle$ denotes the canonical pairing between $T^*M$ and $TM$.

Recursion operators also generate symmetries. If $R$ is a recursion operator and $Y$ is a symmetry of $X$, that is, $[X, Y] = 0$, then $RY$ is also a symmetry of $X$. So, given a recursion operator $R$ of $X$, we may construct a sequence of symmetries of $X$, $R^kY, k \in \mathbb{N}$.

The Nijenhuis torsion of a $(1,1)$-tensor $R$ is the $(1,2)$-tensor $\mathcal{T}(R)$ defined by

$$\mathcal{T}(R)(X, Y) = [RX, RY] - R([X, RY] + [RX, Y] - R[X, Y]), \quad X, Y \in \mathfrak{X}(M) \qquad [10]$$

A Nijenhuis operator is a $(1,1)$-tensor, $R$, with vanishing Nijenhuis torsion, that is,

$$\mathcal{L}_{RX} R = R\mathcal{L}_X R \qquad [11]$$

These operators can generate sequences of closed 1-forms. If $R$ is a Nijenhuis operator and $\alpha$ is a closed 1-form such that $\mathrm{d}{}^tR(\alpha) = 0$, then $\mathrm{d}{}^tR^k(\alpha) = 0, k \in \mathbb{N}$. In the particular case of $\alpha$ being exact, that is, $\alpha = \mathrm{d}f$ and the first cohomology group being trivial, then we have a sequence of local integrals of motion $\mathrm{d}f_k = {}^tR^k(\mathrm{d}f)$.

A Nijenhuis recursion operator $R$ and a symmetry $Y$ of a vector field $X$ lead to a sequence of commuting symmetries $R^kY, k \in \mathbb{N}$,

$$[R^iY, R^jY] = 0, \quad i, j \in \mathbb{N} \qquad [12]$$

To define the integrability in terms of a $(1,1)$-tensor is of special relevance when we try to extend everything to the infinite-dimensional case.

Notice that in coordinates $(q_1, \ldots, q_n)$, the condition [8] is equivalent to

$$\dot{R} = [A, R] \qquad [13]$$

where $A$ is the $n \times n$ matrix defined by

$$A_{ij} = \left[\frac{\partial X^j}{\partial q_i}\right]$$

and $X^j = X(q_j) = \dot{q}_j$, $j = 1, \ldots, n$. So, the pair $(R, A)$ is a local Lax pair of the system and the eigenvalues of $R$ are integrals of $X$.

If a recursion operator $R$ of a vector field $X$ on a manifold $M$ has vanishing Nijenhuis torsion and $n$ doubly degenerated eigenvalues $\lambda_i$, with nowhere-vanishing differentials, $(\mathrm{d}\lambda_i)_p \neq 0$, then $X$ defines a completely integrable Hamiltonian system.

Now suppose $X$ defines a completely integrable Hamiltonian system with Hamiltonian $H$ on a symplectic manifold $(M, \omega)$. Let $(I_1, \ldots, I_n, \varphi_1, \ldots, \varphi_n)$ be the action-angle variables in a neighborhood of an invariant torus. Two cases may happen:

1. The Hamiltonian $H$ is separable in the action variable, that is,

$$H = \sum_k H_k(I_k) \qquad [14]$$

In this case, the $(1,1)$-tensor

$$R = \sum_k \lambda_k(I_k)\left(\mathrm{d}I_k \otimes \frac{\partial}{\partial I_k} + \mathrm{d}\varphi_k \otimes \frac{\partial}{\partial \varphi_k}\right) \qquad [15]$$

where $\lambda_k$ are functions with nowhere-vanishing differentials, is a recursion operator of $X$, and has vanishing Nijenhuis torsion and doubly degenerated eigenvalues.

2. The Hamiltonian has nonvanishing Hessian

$$\det\left(\frac{\partial^2 H}{\partial I_k \partial I_j}\right) \neq 0 \qquad [16]$$

In this case we may define new coordinates

$$\nu_k = \frac{\partial H}{\partial I_k}, \quad k = 1, \ldots, n \qquad [17]$$

and a new symplectic structure

$$\omega_1 = \sum_k \mathrm{d}\nu_k \wedge \mathrm{d}\varphi_k = \sum_{k,j} \frac{\partial^2 H}{\partial I_k \partial I_j} \mathrm{d}I_k \wedge \mathrm{d}\varphi_j \qquad [18]$$

The vector field $X$ is Hamiltonian with respect to $\omega_1$, with Hamiltonian

$$H = \frac{1}{2}\sum_k \nu_k^2 \qquad [19]$$

and the $(1,1)$-tensor

$$R = \sum_k \lambda_k\,(I_k)\left(\mathrm{d}\nu_k \otimes \frac{\partial}{\partial \nu_k} + \mathrm{d}\varphi_k \otimes \frac{\partial}{\partial \varphi_k}\right) \qquad [20]$$

is a recursion operator of $X$.

Nijenhuis operators also allow the construction of master symmetries from conformal ones.

A conformal symmetry of a tensor field $T$ is a vector field $Z$ such that

$$\mathcal{L}_Z T = \lambda T, \quad \text{for some constant } \lambda$$

A master symmetry of a vector field $X$ is a vector field $Y$ such that

$$[X, [X, Y]] = 0, \quad \text{but } [X, Y] \neq 0$$

Let $R$ be a recursion operator of $X_0$ and $Z_0$ be a conformal symmetry of $X_0$ and $R$ such that

$$\mathcal{L}_{Z_0} X_0 = \lambda X_0 \quad \text{and} \quad \mathcal{L}_{Z_0} R = R \qquad [21]$$

for some constants $\lambda, \mu$.

If $R$ is also a Nijenhuis operator, then defining the sequences of commuting symmetries $X_k = R^k X_0$ and of conformal symmetries $Z_k = R^k Z_0, k \in \mathbb{N}$, we have, for all $k, j \in \mathbb{N}_0$,

$$\mathcal{L}_{Z_k} R = \mu R^{k+1} \qquad [22]$$

$$[Z_k, Z_j] = \mu(j - k) Z_{j+k} \qquad [23]$$

$$[Z_k, X_j] = (\lambda + j\mu) X_{k+j} \qquad [24]$$

A bi-Hamiltonian manifold is a smooth manifold $M$ endowed with two linearly independent Poisson tensors $\Lambda_0, \Lambda_1$, compatible in the sense that their Schouten bracket vanishes, $[\Lambda_0, \Lambda_1] = 0$.

A vector field is said to be bi-Hamiltonian if it is Hamiltonian with respect to both Poisson structures. The equation that rules the flow of this vector field is said to be a bi-Hamiltonian system.

When one of the Poisson structures is obtained from the other by means of a Nijenhuis operator, we obtain a Poisson–Nijenhuis manifold. Hence, a Poisson–Nijenhuis manifold is a differentiable manifold $M$ endowed with a Poisson tensor $\Lambda$ and a $(1,1)$-tensor $R$ such that

$$R\Lambda^\sharp = \Lambda^{\sharp t} R, \quad [R\Lambda, \Lambda] = 0 \text{ and } [R\Lambda, R\Lambda] = 0$$

A classical example is the one of a bi-Hamiltonian manifold $(M, \Lambda_0, \Lambda_1)$ where $\Lambda_0$ is nondegenerated. In this case we may define the Nijenhuis operator $R = \Lambda_1^\sharp \Lambda_0^{\sharp -1}$ and the manifold $M$ is a Poisson–Nijenhuis one.

The characteristics of the Poisson–Nijenhuis manifold guarantee that all the bivectors $\Lambda_k = R^k \Lambda$ are compatible Poisson tensors and the manifold is not just bi-Hamiltonian but multi-Hamiltonian.

From what we saw, a Hamiltonian system is completely integrable if and only if it is bi-Hamiltonian

in a neighborhood of an invariant torus with the eigenvalues of the existing recursion operator providing its complete integrability. These Poisson–Nijenhuis manifolds appear quite frequently in dynamics and allow us to obtain some interesting properties easily. We finish this section with the Toda lattice. This system is a good illustration of what has been said until now.

Consider $\mathbb{R}^{2n-1}$ with coordinates $(a_1, \ldots, a_{n-1}, b_1, \ldots, b_n)$ equipped with the following compatible Poisson tensors:

$$\Lambda_0 = \frac{1}{4}\sum_{i=1}^{n-1} a_i \frac{\partial}{\partial a_i} \wedge \left(\frac{\partial}{\partial b_i} - \frac{\partial}{\partial b_{i+1}}\right) \qquad [25]$$

$$\begin{aligned}\Lambda_1 = &\sum_{i=1}^{n-1} a_i^2 \frac{\partial}{\partial b_{i+1}} \wedge \frac{\partial}{\partial b_i} - \frac{1}{4}\sum_{i=1}^{n-1} a_i \frac{\partial}{\partial a_i} \\ &\wedge \left(a_{i+1}\frac{\partial}{\partial a_{i+1}} + 2b_{i+1}\frac{\partial}{\partial b_{i+1}} - 2b_i \frac{\partial}{\partial b_i}\right)\end{aligned} \qquad [26]$$

Not only these two Poisson tensors are degenerated but also there is no Nijenhuis operator that transforms $\Lambda_0$ into $\Lambda_1$. This can be seen considering the 1-form $\sum_{i=1}^n \mathrm{d}b_i$. This 1-form belongs to the kernel of $\Lambda_0$ but not to the kernel of $\Lambda_1$. So, the bi-Hamiltonian manifold $(\mathbb{R}^{2n-1}, \Lambda_0, \Lambda_1)$ is not a Poisson–Nijenhuis one.

The Toda lattice is the bi-Hamiltonian system in $\mathbb{R}^{2n-1}$:

$$X_1 = \Lambda_0^\sharp(\mathrm{d}H_1) = \Lambda_1^\sharp(\mathrm{d}H_0) \qquad [27]$$

defined by the Hamiltonians

$$\begin{aligned} H_0 &= 2\sum_{i=1}^n b_i \\ H_1 &= 4\sum_{i=1}^{n-1} a_i^2 + 2\sum_{i=1}^n b_i^2 \end{aligned} \qquad [28]$$

that is,

$$\begin{aligned} \dot{a}_i &= a_i(b_{i+1} - b_i), \quad \text{if } 1 \le i \le n-1 \\ \dot{b}_1 &= 2a_1^2 \\ \dot{b}_i &= 2(a_i^2 - a_{i-1}^2), \quad \text{if } 2 \le i \le n-1 \\ \dot{b}_n &= -2a_{n-1}^2 \end{aligned}$$

Since we do not have a Nijenhuis operator in this setting, we are going to consider a new system in $\mathbb{R}^{2n}$ that reduces to the Toda lattice, derive a hierarchy of Hamiltonians, symmetries, Poisson tensors, conformal symmetries and the associated relations and then transport everything to $\mathbb{R}^{2n-1}$ by reduction.

Consider the Flaschka transformation

$$\begin{aligned} \pi : \mathbb{R}^{2n} &\to \mathbb{R}^{2n-1} \\ (q_1, \ldots, q_n, p_1, \ldots, p_n) &\mapsto (a_1, \ldots, a_{n-1}, b_1, \ldots, b_n) \end{aligned}$$

where

$$\begin{aligned} a_i = \frac{1}{2}\exp\left(\frac{q_i - q_{i+1}}{2}\right), &\qquad b_j = -\frac{1}{2}p_j \\ i = 1, \ldots, n-1, &\quad j = 1, \ldots, n \end{aligned} \qquad [29]$$

This application is a Poisson morphism between $(\mathbb{R}^{2n}, \widetilde{\Lambda}_0, \widetilde{\Lambda}_1)$ and $(\mathbb{R}^{2n-1}, \Lambda_0, \Lambda_1)$, where

$$\widetilde{\Lambda}_0 = \sum_{i=1}^n \frac{\partial}{\partial p_i} \wedge \frac{\partial}{\partial q_i} \qquad [30]$$

$$\begin{aligned}\widetilde{\Lambda}_1 = &\sum_{i=1}^{n-1} \mathrm{e}^{q_i - q_{i+1}} \frac{\partial}{\partial p_{i+1}} \wedge \frac{\partial}{\partial p_i} \\ &+ \sum_{i=1}^n \left(p_i \frac{\partial}{\partial q_i} \wedge \frac{\partial}{\partial p_i} + \sum_{i<j} \frac{\partial}{\partial q_j} \wedge \frac{\partial}{\partial q_i}\right)\end{aligned} \qquad [31]$$

The Poisson tensor $\widetilde{\Lambda}_0$ is nondegenerated and we may define the Nijenhuis operator $R = \widetilde{\Lambda}_1^\sharp \widetilde{\Lambda}_0^{\sharp -1}$. So, $(\mathbb{R}^{2n}, \widetilde{\Lambda}_0, \widetilde{\Lambda}_1)$ is a Poisson–Nijenhuis manifold and the bivectors of the sequence $(\widetilde{\Lambda}_k = R^k \widetilde{\Lambda}_0), k \in \mathbb{N}$, are compatible Poisson tensors.

The Toda lattice is the reduced bi-Hamiltonian system, by means of the Flaschka transformation, of the bi-Hamiltonian system

$$\widetilde{X}_1 = \widetilde{\Lambda}_0^\sharp(\mathrm{d}\widetilde{H}_1) = \widetilde{\Lambda}_1^\sharp(\mathrm{d}\widetilde{H}_0) \qquad [32]$$

where

$$\begin{aligned} \widetilde{H}_0 &= \sum_{i=1}^n p_i \\ \widetilde{H}_1 &= \sum_{i=1}^n \frac{p_i^2}{2} + \sum_{i=1}^{n-1} \mathrm{e}^{q_i - q_{i+1}} \end{aligned} \qquad [33]$$

We may define the sequence of commuting vector fields $\widetilde{X}_k = R^{n-1}\widetilde{X}_1, k \in \mathbb{N}$, and the sequence of Hamiltonians $\mathrm{d}\widetilde{H}_k = {}^t R^k(\mathrm{d}\widetilde{H}_0), k \in \mathbb{N}$, first integrals of all the vector fields $\widetilde{X}_j$ and in involution with respect to all the Poisson structures $\widetilde{\Lambda}_j$.

Moreover, considering the conformal symmetry of $\widetilde{\Lambda}_0, \widetilde{\Lambda}_1$, and $\widetilde{H}_0$ defined by

$$\widetilde{Z}_0 = \sum_{i=1}^n (n + 1 - 2i)\frac{\partial}{\partial q_i} + \sum_{i=1}^n p_i \frac{\partial}{\partial p_i} \qquad [34]$$

we have the following relations on $\mathbb{R}^{2n}$:

$$\mathcal{L}_{\widetilde{Z}_m} R = R^{m+1} \qquad [35]$$

$$[\widetilde{Z}_m, \widetilde{Z}_k] = (k-m)\widetilde{Z}_{k+m} \qquad [36]$$

$$[\widetilde{Z}_k, \widetilde{X}_{m+1}] = m\widetilde{X}_{k+m+1} \qquad [37]$$

$$\mathcal{L}_{\widetilde{Z}_k}\widetilde{\Lambda}_m = (m-k-1)\widetilde{\Lambda}_{k+m} \qquad [38]$$

$$\widetilde{Z}_k.\widetilde{H}_m = (m+n+1)\widetilde{H}_{k+m} \qquad [39]$$

Although we do not have a Nijenhuis operator on $(\mathbb{R}^{2n-1}, \Lambda_0, \Lambda_1)$, the deformation relations [35]–[39], obtained for the Poisson–Nijenhuis manifold $(\mathbb{R}^{2n}, \widetilde{\Lambda}_0, \widetilde{\Lambda}_1)$, may be reduced to the bi-Hamiltonian manifold $(\mathbb{R}^{2n-1}, \Lambda_0, \Lambda_1)$ by means of the Flaschka transformation $\pi$.

## Recursion Operators and Algebras of Integrals of Motion

A master integral of a vector field $X$ is a differentiable function $g$ such that

$$\mathcal{L}_X\mathcal{L}_X g = 0 \quad \text{and} \quad \mathcal{L}_X g \neq 0 \qquad [40]$$

So, a master integral $g$ generates an integral of motion $\mathcal{L}_X g$ of the system $X$. It is worth noticing that if $f$ and $g$ are master integrals, then not only $\mathcal{L}_X f$ and $\mathcal{L}_X g$ are integrals but also $(\mathcal{L}_X f)g - f(\mathcal{L}_X g)$ is an integral of the system. This means that several master integrals may lead to extra integrals of motion. This procedure often leads to the construction of the integrals which provide the superintegrability of the system in consideration. This is the case of, for instance, the generalized rational Calogero–Moser system or the geodesic flow on the sphere.

Recursion operators are often used to construct sequences of master symmetries of vector fields. The obvious connection between master symmetries and master integrals carries the recursion operators to this level. In many cases, the integrals of motion generated by the master integrals constructed on the basis of the existence of a recursion operator close in a quadratic algebra with respect to the Poisson structure we are considering (by quadratic algebra we mean that the brackets between the generators are polynomials of degree 2 in the generators).

Let $X$ be a vector field on a manifold $M$, $R$ a Nijenhuis operator which is also a recursion operator of $X$, and $P$ a $(1,1)$-tensor such that

$$\mathcal{L}_X P = a(R)$$

and

$$\mathcal{L}_{PX} R = b(R)$$

where $a$ and $b$ are polynomials with constant coefficients. The sequences $X_i = R^i X, Y_i = R^i(PX)$, $i \in \mathbb{N}_0, X_{-1} = Y_{-1} = 0$ satisfy

$$[X_i, X_j] = 0 \qquad [41]$$

$$[X_i, Y_j] = a(R)X_{i+j} - ib(R)X_{i+j-1} \qquad [42]$$

$$[Y_i, Y_j] = (j-i)b(R)Y_{i+j-1} \qquad [43]$$

If $(M, \Lambda)$ is a nondegenerated Poisson manifold with trivial first cohomology group, $R\Lambda$ is a bivector and $X$ and $Y$ are Hamiltonian vector fields with respect to $\Lambda$ and $R\Lambda$, that is, there exist functions $H_0, H_1, G_0$, and $G_1$ satisfying

$$X = \Lambda^\sharp(\mathrm{d}H_1) = R\Lambda^\sharp(\mathrm{d}H_0)$$
$$Y = \Lambda^\sharp(\mathrm{d}G_1) = R\Lambda^\sharp(\mathrm{d}G_0)$$

then the sequences of exact differentials

$${}^tR^i(\mathrm{d}H_1) = \mathrm{d}H_i \quad \text{and} \quad {}^tR^i(\mathrm{d}G_1) = \mathrm{d}G_i$$

may be constructed. In this case, the functions $G_j$ are master integrals of all the vector fields $X_i$ and the integrals $X_i(G_j)$ and $L^i_{k,j} = X_i(G_k)G_j - X_i(G_j)G_k$, $j, k \in \mathbb{N}_0$, close in a quadratic algebra with respect to the Poisson bracket associated with $\Lambda$.

If $M$ is not a Poisson manifold but we can find a master integral $G$ of all the vector fields $X_i$ of the sequence, then the functions $G_j = Y_j(G)$ are also master integrals of the same vector fields and the functions $X_i(G_j)$ and $L^i_{k,j} = X_i(G_k)G_j - X_i(G_j)G_k$ are integrals of $X_i$.

Now let us consider the completely integrable bi-Hamiltonian system case. In a neighborhood of an invariant torus, a completely integrable bi-Hamiltonian system may be written in the form

$$\widetilde{H}(y_1, \ldots, y_n) = y_1 + \cdots + y_n \qquad [44]$$

with

$$\Lambda_0 = \sum_{i=1}^n \frac{\partial}{\partial y_i} \wedge \frac{\partial}{\partial \phi_i}$$
$$\Lambda_1 = \sum_{i=1}^n y_i \frac{\partial}{\partial y_i} \wedge \frac{\partial}{\partial \phi_i}$$

the compatible Poisson tensors that provide the complete integrability of the bi-Hamiltonian system. In this case, we may define the recursion operator

$$R = \sum_{i=1}^n y_i \left( \frac{\partial}{\partial y_i} \otimes \mathrm{d}y_i + \frac{\partial}{\partial \phi_i} \otimes \mathrm{d}\phi_i \right)$$

for which $\Lambda_1 = R\Lambda_0$, and the bi-Hamiltonian vector field

$$X = \Lambda_0^\sharp(d\widetilde{H}) = \Lambda_1^\sharp\left[d\left(\sum_{i=1}^n \ln(y_i)\right)\right]$$

The (1, 1)-tensor

$$P = \sum_{i=1}^n \left(\phi_i \frac{\partial}{\partial \phi_i} \otimes d\phi_i + \frac{\partial}{\partial y_i} \otimes dy_i\right)$$

satisfies $\mathcal{L}_X P = \mathrm{Id}$ and $\mathcal{L}_{PX} R = 0$. So, the vector fields

$$Y_k = R^k(PX) = \sum_{i=1}^n y_i^k \phi_i \frac{\partial}{\partial \phi_i}$$

and the function $G = \sum_{i=1}^n y_i\phi_i$ help defining the functions $G_i = Y_i(G), i \in \mathbb{N}_0$.

The integrals of $X_k$

$$X_k(G_j) \quad \text{and} \quad L_{i,j}^k = X_k(G_i)G_j - G_iX_k(G_j) \qquad [45]$$

happen to close in a quadratic algebra with respect to the bracket defined by $\Lambda_0$.

## Recursion Operators on Jacobi Manifolds

In this section we extend the notion of Poisson–Nijenhuis manifold to the Jacobi setting.

Let $M$ be a smooth manifold with a bivector field $\Lambda$ and a vector field $E$. We equip the space $C^\infty(M)$ with the bracket

$$\{f, g\} = \Lambda(df, dg) + fE(g) - gE(f)$$

which is bilinear and skew-symmetric, and satisfies the Jacobi identity if and only if

$$[\Lambda, \Lambda] = -2E \wedge \Lambda \quad \text{and} \quad [E, \Lambda] = 0 \qquad [46]$$

When these conditions are satisfied, $(M, \Lambda, E)$ is called a Jacobi manifold with Jacobi bracket $\{\cdot, \cdot\}$. The pair $(C^\infty(M), \{,\})$ is a local Lie algebra in the sense of Kirillov. If the vector field $E$ identically vanishes on $M$, eqns [46] reduce to $[\Lambda, \Lambda] = 0$ and $(M, \Lambda)$ is just a Poisson manifold. But there are other examples of Jacobi manifolds that are not Poisson, for example, contact manifolds.

We denote by $(\Lambda, E)^\# : T^*M \times \mathbb{R} \to TM \times \mathbb{R}$ the vector bundle map associated with $(\Lambda, E)$, that is, for all $\alpha, \beta$ sections of $T^*M$ and $f \in C^\infty(M)$,

$$(\Lambda, E)^\#(\alpha, f) = (\Lambda^\#(\alpha) + fE, -i_E\alpha)$$

Let $\mathcal{R} : \mathfrak{X}(M) \times C^\infty(M) \to \mathfrak{X}(M) \times C^\infty(M)$ be a $C^\infty(M)$-linear map defined by

$$\mathcal{R}(X, f) = (NX + fY, i_X\gamma + gf) \qquad [47]$$

where $N$ is a tensor field of type (1, 1) on $M$, $Y \in \mathfrak{X}(M)$, $\gamma \in \Omega^1(M)$ and $g \in C^\infty(M)$. Let us denote by $\mathcal{T}(\mathcal{R})$ the Nijenhuis torsion of $\mathcal{R}$ with respect to the Lie bracket on $\mathfrak{X}(M) \times C^\infty(M)$ given by

$$[(X, f), (Z, h)] = ([X, Z], X(h) - Z(f)) \qquad [48]$$

As in the case of Poisson manifolds, if $\mathcal{R}$ has a vanishing Nijenhuis torsion, we call $\mathcal{R}$ a Nijenhuis operator.

Suppose now that $M$ is equipped with a Jacobi structure $(\Lambda_0, E_0)$ and a Nijenhuis operator $\mathcal{R}$. Then, we may define a bivector field $\Lambda_1$ and a vector field $E_1$ on $M$, by setting

$$(\Lambda_1, E_1)^\# = \mathcal{R} \circ (\Lambda_0, E_0)^\#$$

If one looks for the conditions that imply that the pair $(\Lambda_1, E_1)$ defines a new Jacobi structure on $M$ compatible with $(\Lambda_0, E_0)$, in the sense that $(\Lambda_0 + \Lambda_1, E_0 + E_1)$ is again a Jacobi structure, one finds that $\Lambda_1$ is skew-symmetric if and only if $\mathcal{R} \circ (\Lambda_0, E_0)^\# = (\Lambda_0, E_0)^\# \circ {}^t\mathcal{R}$. When $\Lambda_1$ is skew-symmetric, $(\Lambda_1, E_1)$ defines a Jacobi structure on $M$ if and only if, for all $(\alpha, f), (\beta, h) \in \Omega^1(M) \times C^\infty(M)$,

$$\mathcal{T}(\mathcal{R})\Big((\Lambda_0, E_0)^\#(\alpha, f), (\Lambda_0, E_0)^\#(\beta, h)\Big) = \mathcal{R} \circ (\Lambda_0, E_0)^\#(\mathcal{C}((\Lambda_0, E_0), \mathcal{R})((\alpha, f), (\beta, h)))$$

where $\mathcal{C}((\Lambda_0, E_0), \mathcal{R})$ is the Magri concomitant of $(\Lambda_0, E_0)$ and $\mathcal{R}$. In the case where $(\Lambda_1, E_1)$ is a Jacobi structure, it is compatible with $(\Lambda_0, E_0)$ if and only if, for all $(\alpha, f), (\beta, h) \in \Omega^1(M) \times C^\infty(M)$,

$$(\Lambda_0, E_0)^\#(\mathcal{C}((\Lambda_0, E_0), \mathcal{R})((\alpha, f), (\beta, h))) = 0$$

A Jacobi–Nijenhuis manifold $(M, (\Lambda_0, E_0), \mathcal{R})$ is a Jacobi manifold $(M, \Lambda_0, E_0)$ with a Nijenhuis operator $\mathcal{R}$ such that: (1) $\mathcal{R} \circ (\Lambda_0, E_0)^\# = (\Lambda_0, E_0)^\# \circ {}^t\mathcal{R}$ and (2) the map $(\Lambda_0, E_0)^\# \circ \mathcal{C}((\Lambda_0, E_0), \mathcal{R})$ identically vanishes. $\mathcal{R}$ is called the recursion operator of $(M, (\Lambda_0, E_0), \mathcal{R})$.

A recursion operator on a Jacobi–Nijenhuis manifold displays a hierarchy of Jacobi–Nijenhuis structures on the manifold. In fact, if $((\Lambda_0, E_0), \mathcal{R})$ is a Jacobi–Nijenhuis structure on $M$, there exists a hierarchy $((\Lambda_k, E_k), k \in \mathbb{N})$ of Jacobi structures on $M$, which are pairwise compatible. For all $k \in \mathbb{N}$, $(\Lambda_k, E_k)$ is the Jacobi structure associated with the vector bundle map $(\Lambda_k, E_k)^\#$ given by $(\Lambda_k, E_k)^\# = \mathcal{R}^k \circ (\Lambda_0, E_0)^\#$. Moreover, for all $k, l \in \mathbb{N}$, the pair $((\Lambda_k, E_k), \mathcal{R}^l)$ defines a Jacobi–Nijenhuis structure on $M$.

*See also:* Bi-Hamiltonian Methods in Soliton Theory; Classical *r*-Matrices, Lie Bialgebras, and Poisson Lie Groups; Contact Manifolds; Integrable Systems and Algebraic Geometry; Integrable Systems: Overview;

Multi-Hamiltonian Systems; Recursion Operators in Classical Mechanics.

## Further Reading

Abraham R and Marsden J (1978) *Foundations of Mechanics*, 2nd edn. Massachusetts: Benjamin-Cummings.

Kosmann-Schwarzbach Y and Magri F (1990) Poisson–Nijenhuis structures. *Annales de l'Institut Henri Poincaré, Physique Théorique* 53(1): 35–81.

Libermann P and Marle C-M (1987) Symplectic geometric and analytic mechanics. In: (ed.) *Mathematics and Its Applications*. Dordrecht, Holland: D. Reidel.

Lichnerowicz A (1978) Les variétés de Jacobi et leurs algèbres de Lie associées. *Journal de Mathématiques Pures et Appliquées Articles (9)* 57(4): 453–488.

Magri F (1997) Eight lectures on integrable systems. In: Kosmann-Schwarzbach Y *et al.* (eds.) *Integrability of Nonlinear Systems*, Proceedings of the CIMPA school, Pondicherry University, India, January 8–26, 1996, Lecture Notes in Physics, vol. 495, pp. 256–296. Berlin: Springer.

Magri F and Morosi C (1984) A geometric characterization of integrable Hamiltonian systems through the theory of Poisson–Nijenhuis manifolds. Quaderno S19, Università di Milano.

Oevel W (1987) A geometric approach to integrable systems admitting time dependent invariants. In: Ablowitz M, Fuchssteiner B, and Kruskal M (eds.) *Topics in Soliton Theory and Exactly Solvable Nonlinear Equations*, pp. 108–124. Singapore: World Scientific.

Perelomov A (1990) *Integrable Systems of Classical Mechanics and Lie Algebras*, vol. I. Basel: Birkhauser.

Vilasi G (2001) *Hamiltonian Dynamics*. Singapore: World Scientific.

# Integrable Systems and the Inverse Scattering Method

**A S Fokas**, University of Cambridge, Cambridge, UK



## Introduction

A British experimentalist, J S Russell, first observed a soliton in 1834 while riding on horseback beside a narrow barge channel. He challenged the theoreticians of the day "to predict the discovery after it happened, that is to give an *a priori* demonstration *a posterori*." This work created a controversy which, in fact, lasted almost 50 years, and which involved such distinguished scientists as Stokes and Airy. It was resolved by Korteweg and deVries in 1895, who derived the KdV equation as an approximation to water waves,

$$\frac{\partial q}{\partial t} + 6q\frac{\partial q}{\partial x} + \frac{\partial^3 q}{\partial x^3} = 0 \qquad [1]$$

This equation is a nonlinear partial differential equation (PDE) of the evolution type, where $t$ and $x$ are related to time and space respectively, and $q(x,t)$ is related to the height of the wave above the mean water level. Korteweg and de Vries were able to show that equation [1] supports a particular solution that exhibits the behavior described by Russell. This solution, which was later called 1-soliton solution, is given by

$$q_1(x - p^2 t) = \frac{p^2/2}{\cosh^2((1/2)p(x - p^2 t) + c)} \qquad [2]$$

where $p, c$ are constants. The location of this soliton at time $t$, that is, its maximum position, is given by $p^2 - 2c/p$, its velocity is given by $p^2$, and its amplitude by $p^2/2$. Thus, faster solitons are higher and narrower. It should be noted that $q_1$ is a traveling-wave solution, that is, $q_1$ depends only on the variable $X = x - p^2 t$, thus in this case the PDE [1] reduces (after integration) to the second-order ordinary differential equation (ODE)

$$-p^2 q_1(X) + 3q_1^2(X) + \frac{d^2 q_1}{dX^2}(X) = 0$$

Under the assumption that $q$ and $dq/dX$ tend to zero as $|X| \to \infty$, this ODE yields the 1-soliton solution [2].

The problem of finding a solution describing the interaction of two 1-soliton solutions is much more difficult and was not addressed by Korteweg and deVries. This question was studied by M Kruskal and N Zabusky in 1965. Studying numerically the interaction of two solutions of the form [2] (i.e., two solutions corresponding to two different $p_1$ and $p_2$), Kruskal and Zabusky discovered the defining property of solitons: after interaction, these waves regained exactly the shapes they had before. This posed a new challenge to mathematicians, namely to explain analytically the interaction properties of such coherent waves. In order to resolve this challenge one needs to develop a larger class of solutions than the 1-soliton solution. We note that eqn [1] is nonlinear and no effective method to solve such nonlinear equations existed at that time.

Gardner *et al.* (1967) not only derived an explicit solution describing the interaction of an arbitrary number of solitons, but also discovered what was to

evolve into a new method of mathematical physics. The 2-soliton solution is given by

$$q_2(x,t) = \frac{2\left(p_1^2 e^{\eta_1} + p_2^2 e^{\eta_2}\right) + 4e^{\eta_1+\eta_2}(p_1-p_2)^2 + 2A_{12}\left(p_2^2 e^{2\eta_1+\eta_2} + p_1^2 e^{\eta_1+2\eta_2}\right)}{\left(1+e^{\eta_1}+e^{\eta_2}+A_{12}e^{\eta_1+\eta_2}\right)^2} \qquad [3]$$

where

$$\eta_j = p_j x - p_j^3 t + \eta_j^0, j=1,2, \qquad A_{12} = \frac{(p_1-p_2)^2}{(p_1+p_2)^2}$$

and $p_j, \eta_j^0$ are constants. A snapshot of this solution with $p_1 = 1, p_2 = 2$ is given in **Figure 1**. After some time the taller soliton will overtake the shorter one and the only effect of the interaction will be a "phase shift," that is, a change in the position the two solitons would have reached without interaction.

Regarding the general method introduced in Gardner *et al.* (1967), we note that if eqn [1] is formulated on the infinite line, then the most interesting problem is the solution of the initial-value problem: given initial data $q(x,0) = q_0(x)$ which decay as $|x| \to \infty$, find $q(x,t)$. If $q_0$ is small and $qq_x$ can be neglected, then eqn [1] becomes linear and $q(x,t)$ can be found using the Fourier transform,

$$q(x,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx+ik^3 t}\hat{q}_0(k)\,dk \qquad [4a]$$

where

$$\hat{q}_0(k) = \int_{-\infty}^{\infty} e^{-ikx} q_0(x)\,dx \qquad [4b]$$

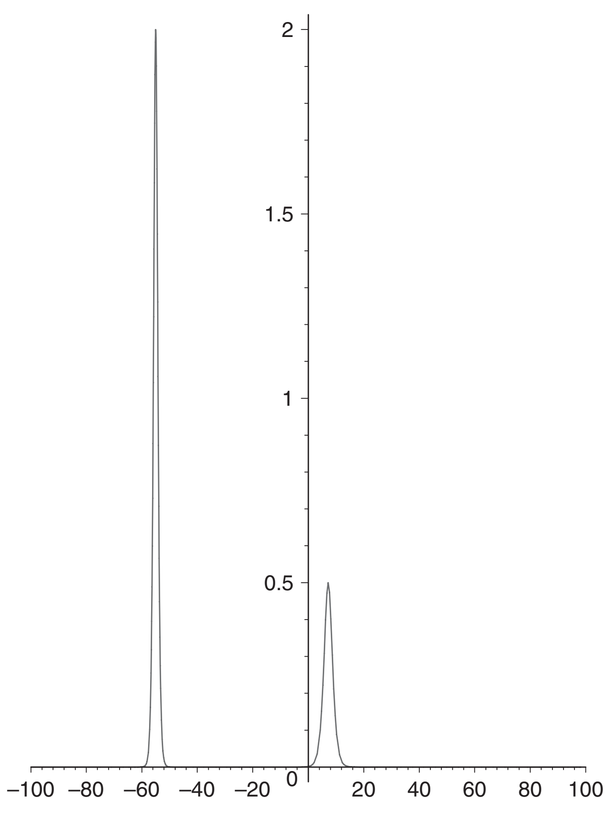



**Figure 1** A snapshot of the 2-soliton solution of the KdV equation.

The remarkable discovery of Gardner *et al.* (1967) is that for eqn [1] there exists a "nonlinear analog" of the Fourier transform capable of solving the initial-value problem even if $q_0$ is not small. Although this nonlinear Fourier transform cannot in general be written in closed form, $q(x,t)$ can be expressed through the solution of a linear integral equation, or more precisely through the solution of a linear $2 \times 2$ matrix Riemann–Hilbert (RH) problem (see the section "A nonlinear Fourier transform"). This linear integral equation is uniquely specified in terms of $q_0(x)$. For particular initial data, $q(x,t)$ can be written explicitly. For example, if $q_0(x) = q_1(x)$, where $q_1(x)$ is obtained by evaluating eqn [2] at $t=0$, then $q(x,t) = q_1(x - p^2 t)$. Similarly, if $q_0(x) = q_2(x,0)$, where $q_2(x,0)$ is obtained by evaluating eqn [3] at $t=0$, then $q(x,t) = q_2(x,t)$.

The most important question, both physically and mathematically, is the description of the long-time behavior of the solution of the initial-value problem mentioned above. If the nonlinear term of eqn [1] can be neglected, one finds a linear dispersive equation. In this case different waves travel with different wave speeds, these waves cancel each other out and the solution decays to zero as $t \to \infty$. Indeed, using the stationary-phase method to compute the large $t$ behavior of the integral appearing in eqn [4a], it can be shown that $q(x,t)$ decays like $0(1/\sqrt{t})$ as $t \to \infty$, $x/t = 0(1)$. The situation with the KdV equation is more interesting: dispersion is balanced by nonlinearity and $q(x,t)$ has a "nontrivial" asymptotic behavior as $t \to \infty$. Indeed, using a nonlinear analog of the steepest descent method discovered by Deift and Zhou (1993) to analyze the RH problem mentioned earlier, it can be shown that $q(x,t)$ asymptotes to $q_N(x,t)$, where $q_N(x,t)$ is the exact $N$-soliton solution. This underlines the physical and mathematical significance of solitons: they are the coherent structures emerging from any initial data as $t \to \infty$. This implies that if a nonlinear phenomenon is modeled by the KdV equation on the infinite line, then one can immediately predict the structure of the solution as $t \to \infty$, $x/t = 0(1)$: it will consist of $N$ ordered single solitons, where the highest soliton occurs to the right; the number $N$ and the parameters $p_j$ and $\eta_j^0$ depend on the particular initial data $q_0(x)$. It should

be noted that this result can be obtained only using the machinery of the theory of integrability, and until now cannot be obtained using standard PDE techniques.

So far we have concentrated on the KdV equation. However, there exist numerous other equations which exhibit similar behavior. Such equations are called "integrable" and the method of solving their initial-value problem is called the "inverse-scattering" or "inverse-spectral" method.

The following section presents a brief historical review of some of the important developments of soliton theory. Next, typical solitons, lumps, and dromions are given. The inverse-spectral method is discussed in the penultimate section. Finally, the extension of this method to boundary-value problems is briefly discussed.

## Important Analytical Developments in Soliton Theory

Lax (1968) introduced the so-called Lax pair formulation of the KdV. In an example, he showed that eqn [1] can be written as the compatibility condition of the following pair of linear eigenvalue equations for the eigenfunction $\psi(x,t,k)$:

$$\psi_{xx} + (q + k^2)\psi = 0 \qquad [5a]$$

$$\psi_t + (2q - 4k^2)\psi_x - (q_x + \nu)\psi = 0, \quad k \in \mathbb{C} \qquad [5b]$$

where $\nu$ is an arbitrary constant. The nonlinear Fourier transform mentioned earlier can be obtained by performing the spectral analysis of eqn [5a]. The time evolution of the associated nonlinear Fourier data, which are now called spectral data, is linear and can be determined using eqn [5b]. Following Lax's formulation, Zakharov and Shabat (1972) solved the nonlinear Schrödinger (NLS) equation

$$\mathrm{i}q_t + q_{xx} - 2\lambda|q|^2 q = 0, \quad \lambda = \pm 1 \qquad [6]$$

which has ubiquitous physical applications including nonlinear optics. Soon thereafter the sine-Gordon equation

$$q_{xx} - q_{tt} = \sin q \qquad [7]$$

and the modified KdV equation

$$q_t + 6q^2 q_x + q_{xxx} = 0 \qquad [8]$$

were solved. Since then, numerous nonlinear equations have been solved. Thus, the mathematical technique introduced by Gardner *et al.* (1967) for the solution of a particular physical equation gave rise to a new method in mathematical physics, the so-called inverse-scattering (spectral) method. Among the most important equations solved by this method are a particular two-dimensional reduction of Einstein's equation and the self-dual Yang–Mills equations.

The next important development in the analysis of integrable equations was the study of the KdV with space-periodic initial data. This occurred in the mid-1970s in the USA and in the USSR. This method involves algebraic-geometric techniques; in particular there exists a periodic analog of the $N$-soliton solution which can be expressed in terms of a certain Riemann-theta function of genus $N$.

In the mid-1970s, it was also realized that there exist integrable ODEs. For example, a stationary reduction of some of the equations introduced in connection with the space-periodic problem mentioned above led to the integration of some classical tops. Furthermore, the similarity reduction of some of the integrable PDEs led to the classical Painlevé equations. For example, letting $q = t^{-1/3}u(\xi)$, $\xi = xt^{-1/3}$ in the modified KdV equation [8], and integrating we find

$$\frac{\mathrm{d}^2}{\mathrm{d}\xi^2} + 2u^3 - \frac{1}{3}\xi u + \alpha = 0 \qquad [9]$$

where $\alpha$ is a constant. This is Painlevé II, that is, the second equation in the list of six classical ODEs introduced by Painlevé and is his school around 1900. These equations are nonlinear analogs of the linear special functions such as Airy, Bessel, etc. The connection between integrable PDEs and ODEs of the Painlevé type was established by Ablowitz and Segur (1977). Their work marked a new era in the theory of these equations. Indeed, soon thereafter Flaschka and Newell (1980) introduced an extension of the inverse-spectral method, the so-called isomonodromy method, capable of integrating these equations. The most remarkable achievement of this new development is the construction of nonlinear analogs of the classical connection formulas that exist for the linear special functions. These formulas, although rather complicated, are as explicit as the corresponding linear ones (Fokas *et al.* 2005).

It was mentioned earlier that the inverse-spectral method gives rise to a matrix RH problem. An RH problem involves the determination of a function analytic in given sectors of the complex plane, from the knowledge of the jumps of this function across the boundaries of these sectors. The algebraic-geometric method for solving the space-periodic initial-value problem can be interpreted as formulating an RH problem which can be analyzed using functions defined on a Riemann surface. Also, it was noted by Fokas and Ablowitz (1983a) and later rigorously established by Fokas and Zhou (1992) that the isomonodromy method also gives rise to a novel RH problem. This

implies the following interesting unification: Self-similar, decaying, and periodic initial-value problems for integrable evolution equations in one space variable lead to the study of the same mathematical object, namely to the RH problem.

Every integrable nonlinear evolution equation in one spatial dimension has several integrable versions in two spatial dimensions. Two such integrable physical generalizations of the Korteweg–deVries equation are the so-called Kadomtsev–Petviashvili I (KPI) and II (KPII) equations. In the context of water waves, they arise in the weakly nonlinear, weakly dispersive, weakly two-dimensional limit, and in the case of KPI when the surface tension is dominant. The NLS equation also has two physical integrable versions known as the Davey–Stewartson I (DSI), and II (DSII) equations. They can be derived from the classical water-wave problem in the shallow-water limit and govern the time evolution of the free surface envelope in the weakly nonlinear, weakly two-dimensional, nearly monochromatic limit. The KP and DS equations have several other physical applications.

A method for solving the Cauchy problem for decaying initial data for integrable evolution equations in two spatial dimensions emerged in the early 1980s. This method is sometimes referred to as the $\bar{\partial}$ ($d$-bar) method. We recall that the inverse-spectral method for solving nonlinear evolution equations on the line is based on a matrix RH problem. This problem expresses the fact that there exist solutions of the associated $x$-part of the Lax pair which are sectionally analytic. Analyticity survives in some multidimensional problems: it was shown formally by Fokas and Ablowitz (1983b) that KPI gives rise to a nonlocal RH problem. However, for other multidimensional problems, such as the KPII, the underlying eigenfunctions are nowhere analytic and the RH problem must be replaced by the $\bar{\partial}$ problem. Actually, a $\bar{\partial}$ problem had already appeared in the work of Beals and Coifman (1982) where the RH problem appearing in the analysis of one-dimensional systems was considered as a special case of a $\bar{\partial}$ problem. Soon thereafter, it was shown in Ablowitz *et al.* (1983) that KPII required the essential use of the $\bar{\partial}$ problem. The situation for the DS equations is analogous to that of the KP equations.

Multidimensional integral PDEs can support localized solutions. Actually there exist two types of localized coherent structures associated with integrable evolution equations in two spatial variables: the "lumps" and the "dromions." The spectral meaning, and therefore the genericity of these solutions was established by Fokas and Ablowitz (1983b) and Fokas and Santini (1990).

The analysis of integrable singular integro-differential equations and of integrable discrete equations, although conceptually similar to the analysis reviewed above, has certain novel features.

The fact that integrable nonlinear equations appear in a wide range of physical applications is not an accident but a consequence of the fact that these equations express a certain physical coherence which is natural, at least asymptotically, to a variety of nonlinear phenomena. Indeed, Calogero (1991) has emphasized that large classes of nonlinear evolution PDEs, characterized by a dispersive linear part and a largely arbitrary nonlinear part, after rescaling yield asymptotically equations (for the amplitude modulation) having a universal character. These "universal" equations are, therefore, likely to appear in many physical applications. Many integrable equations are precisely these "universal" models.

## Solitons, Lumps, and Dromions

Solitons, lumps, and dromions, are important not because they are exact solutions, but because they characterize the long-time behavior of integrable evolution equations in one and two space dimensions. The question of solving the initial-value problem of a given integrable PDE, and then extracting the long-time behavior of the solution is quite complicated. It involves spectral analysis, the formulation of either an RH problem or of a $\bar{\partial}$ problem, and rigorous asymptotic techniques. On the other hand, having established the importance of solitons, lumps, and dromions, it is natural to develop methods for obtaining these particular solutions directly, avoiding the difficult approaches of spectral theory. There exist several such direct methods, including the so-called Bäcklund transformations, the dressing method of Zakharov–Shabat, the direct linearizing method of Fokas–Ablowitz, and the bilinear approach of Hirota.

### Solitons

Using the bilinear approach, multisoliton solutions for a large class of integrable nonlinear PDEs in one space dimension are given in Hietarinta (2002). Here we only note that the 1-soliton solution of the NLS [6], of the sine-Gordon [7], and of the modified KdV equation [8] are given, respectively, by

$$q(x,t) = \frac{p_R \mathrm{e}^{\mathrm{i}(p_I x + (p_R^2 - p_I^2)t + \eta)}}{\cosh[p_R(x - 2p_I t) + \eta]} \quad [10]$$

$$q(px + qt) = 4 \arctan[\mathrm{e}^{px+qt+\eta}], \quad p^2 = 1 + q^2 \quad [11]$$

$$q(x - p^2 t) = \frac{\pm p}{\cosh[px - p^2 t + \eta]} \qquad [12]$$

where $p_R$, $p_I$, $\eta$, $p$, $q$ are real constants.

### Lumps

The KPI equation is

$$\partial_x[q_t + 6qq_x + q_{xxx}] = 3q_{yy} \qquad [13]$$

The 1-lump solution of this equation is given by

$$\begin{aligned} q(x,y,t) &= 2\partial_x^2 \ln\left[|L(x,y,t)|^2 + \frac{1}{4\lambda_I^2}\right], \\ L &= x - 2\lambda y + 12\lambda^2 t + a \\ \lambda &= \lambda_R + \mathrm{i}\lambda_I, \quad \lambda_I > 0 \end{aligned} \qquad [14]$$

where $\lambda$ and $a$ are complex constants.

The focusing DSII equation is

$$\mathrm{i}q_t + q_{zz} + q_{\bar{z}\bar{z}} - 2q\left(\partial_{\bar{z}}^{-1}|q|_z^2 + \partial_z^{-1}|q|_{\bar{z}}^2\right) = 0 \qquad [15]$$

where $z = x + \mathrm{i}y$, and the operator $\partial_{\bar{z}}^{-1}$ is defined by

$$\left(\partial_{\bar{z}}^{-1} f\right)(z,\bar{z}) = \frac{1}{2\mathrm{i}\pi}\int_{\mathbb{R}^2} \frac{f(\zeta,\bar{\zeta})}{\zeta - z}\,\mathrm{d}\zeta \wedge \mathrm{d}\bar{\zeta}$$

The 1-lump solution of this equation is given by

$$q(z,\bar{z},t) = \frac{\beta \mathrm{e}^{\mathrm{i}(p^2+\bar{p}^2)t + pz - \bar{p}\bar{z}}}{|z + \alpha + 2\mathrm{i}pt|^2 + |\beta|^2} \qquad [16]$$

where $\alpha, \beta, p$ are complex constants. A typical 1-lump solution is depicted in Figure 2.

### Dromions

The DSI equation is

$$\begin{aligned} \mathrm{i}q_t + \left(\partial_x^2 + \partial_y^2\right)q + qu &= 0 \\ u_{xy} &= 2\left(\partial_x^2 + \partial_y^2\right)|q|^2 \end{aligned} \qquad [17]$$

The 1-dromion solution of this equation is given by

$$\begin{aligned} q(x,y,t) &= \frac{\rho \mathrm{e}^{X - \bar{Y}}}{\alpha \mathrm{e}^{X+\bar{X}} + \beta \mathrm{e}^{-Y-\bar{Y}} + \gamma \mathrm{e}^{X+\bar{X}-Y-\bar{Y}} + \delta} \\ X &= px + \mathrm{i}p^2 t, \quad Y = qy + \mathrm{i}q^2 t \\ |\rho|^2 &= 4p_R q_R(\alpha\beta - \gamma\delta) \end{aligned} \qquad [18]$$

where $p, q$ are complex constants and $\alpha, \beta, \gamma, \delta$ are positive constants.

## A Nonlinear Fourier Transform

The solution of the initial-value problem of an integrable nonlinear evolution equation on the infinite line is based on the spectral analysis of the $x$-part of the Lax pair. Thus, for the KdV equation one must analyze eqn [5a]. This equation is the famous time-independent Schrödinger equation. We now give a physical interpretation of the relevant spectral analysis. Let KdV describe the propagation of a water wave and suppose that this wave is frozen at a given instant of time. By bombarding this water wave with quantum particles, one can reconstruct its shape from knowledge of how these particles scatter. In other words, the scattering data provide an alternative description of the wave at fixed time.

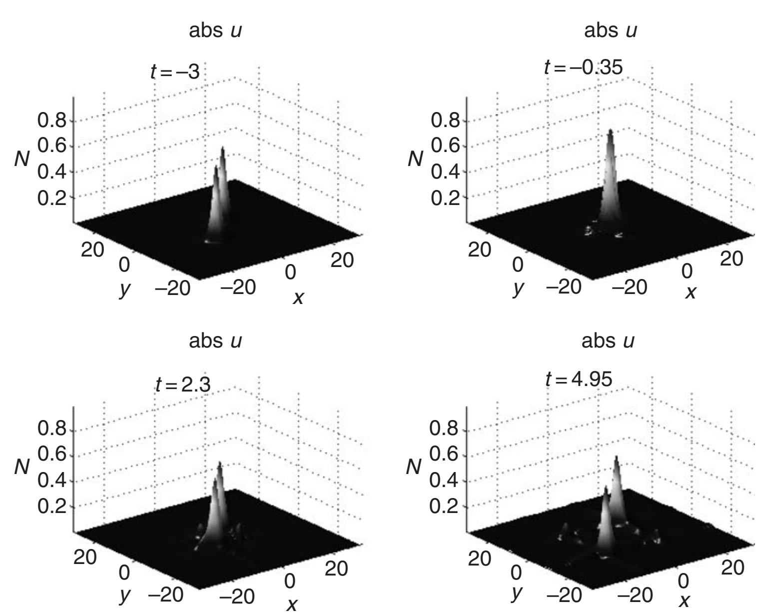



**Figure 2** A typical 1-lump solution.

The mathematical expression of this description takes the form of a linear integral equation found by Faddeev (the so-called Gel'fand–Levitan–Marchenko equation) or equivalently the form of a $2 \times 2$ matrix RH problem uniquely specified by the scattering data. This alternative description of the shape of the wave will be useful if the evolution of the scattering data is simple. This is indeed the case, namely using eqn [5b], it can be shown that the scattering data evolve linearly. Thus, this highly nontrivial change of variables from the physical to scattering space provides a linearization of the KdV equation.

In what follows we will describe some of the relevant mathematical formulas. We first "assume" that there exists a real solution $q(x, t)$ of the initial-value problem which has sufficient smoothness and which decays for all $t$ as $|x| \to \infty$. We then discuss how this assumption can be eliminated.

As it was mentioned earlier most of the analysis of the inverse-scattering transform is carried out on the $x$-part of the Lax pair, that is, on eqn [5a]. Hence, we first concentrate on eqn [5a] and for convenience of notation we suppress the time dependence.

### The Direct Problem

As $|x| \to \infty, q \to 0$, thus there exist solutions of eqn [5a] which tend to $\exp[\pm ikx]$ as $|x| \to \infty$. Let $\psi(k, x)$ and $\hat{\psi}(k, x)$ denote solutions of eqn [5a] with the following asymptotic property:

$$\psi \to e^{ikx}, \quad \hat{\psi} \to e^{-ikx}, \quad \text{as } x \to \infty, \quad k \in \mathbb{R} \qquad [19]$$

Under the transformation $k \to -k$, eqn [5a] remains invariant and the boundary condition for $\psi$ is mapped to the boundary condition for $\hat{\psi}$. Hence

$$\hat{\psi}(k, x) = \psi(-k, x) \qquad [20]$$

We denote by $\phi(k, x)$ the solution of eqn [5a] which tends to $\exp[-ikx]$ as $x \to -\infty$,

$$\phi \to e^{-ikx}, \quad \text{as } x \to -\infty, \quad k \in \mathbb{R} \qquad [21]$$

It is more convenient to work with eigenfunctions (i.e., solutions of [5a]) normalized to unity as $x \to \infty$, thus we introduce $M(k, x)$ and $N(k, x)$ as follows:

$$M = \phi e^{ikx}, \qquad N = \psi e^{-ikx} \qquad [22]$$

The functions $M$ and $N$ can be expressed in terms of $q$ through the solution of linear Volterra integral equations. Indeed, $M$ satisfies

$$\begin{gathered} M_{xx} - 2ikM_x = -qM, \quad k \in \mathbb{R} \\ M \to 1, \quad x \to -\infty \end{gathered} \qquad [23]$$

The homogeneous version of [23] has solutions 1 and $e^{2ikx}$. Thus,

$$M = c_1 + c_2 e^{2ikx} + M_p \qquad [24]$$

where $c_1, c_2$ are constants and $M_p$ is given by

$$M_p = u_1(x) + u_2(x) e^{2ikx} \qquad [25]$$

The functions $u_1, u_2$ satisfy

$$u_1' + e^{2ikx} u_2' = 0, \qquad 2ike^{2ikx} u_2' = -qM$$

Thus,

$$\begin{aligned} u_1(x) &= \frac{1}{2ik} \int_{-\infty}^{x} d\xi q(\xi) M(k, \xi), \\ u_2(x) &= -\frac{1}{2ik} \int_{-\infty}^{x} d\xi e^{-2ik\xi} q(\xi) M(k, \xi) \end{aligned} \qquad [26]$$

Substituting [25] and [26] into [24] and using the boundary condition [23], we find

$$\begin{aligned} M(k, x) & \\ = 1 + \frac{i}{2k} & \int_{-\infty}^{x} d\xi (-1 + e^{2ik(x-\xi)}) q(\xi) M(k, \xi) \end{aligned} \qquad [27]$$

Similarly, one may establish that $N$ satisfied

$$\begin{aligned} N(k, x) & \\ = 1 + \frac{i}{2k} & \int_{x}^{\infty} d\xi (-1 + e^{-2ik(x-\xi)}) q(\xi) N(k, \xi) \end{aligned} \qquad [28]$$

The kernel of eqn [27], as a function of $k$, is bounded and analytic for $\operatorname{Im} k > 0$. Thus, if $q \in L_1, M(k, x)$ as a function of $k$ is holomorphic for $\operatorname{Im} k > 0$. Similarly, $N(k, x)$ as a function of $k$ is holomorphic for $\operatorname{Im} k > 0$.

Thus, we have found particular solutions of eqn [5a] which are holomorphic for $\operatorname{Im} k > 0$. Furthermore, these solutions are simply related for $k$ real. Indeed, the linear independence of solutions of the second-order ODE [5a] implies

$$\phi(k, x) = a(k)\hat{\psi}(k, x) + b(k)\psi(k, x), \quad k \in \mathbb{R}$$

Using [20] and replacing $\phi$ and $\psi$ in terms of $M$ and $N$, we find

$$\begin{gathered} \frac{M(k, x)}{a(k)} = N(-k, x) + \rho(k) e^{2ikx} N(k, x) \\ \rho(k) = \frac{b(k)}{a(k)}, \quad k \in \mathbb{R} \end{gathered} \qquad [29]$$

The functions $a(k)$ and $b(k)$ are given by

$$a(k) = 1 - \frac{\mathrm{i}}{2k}\int_{-\infty}^{\infty} \mathrm{d}\xi q(\xi) M(k,\xi), \quad k \in \mathbb{R}$$
$$b(k) = \frac{\mathrm{i}}{2k}\int_{-\infty}^{\infty} \mathrm{d}\xi q(\xi) M(k,\xi)\mathrm{e}^{-2\mathrm{i}k\xi}, \quad k \in \mathbb{R} \qquad [30]$$

Indeed as $x \to \infty, N \to 1$, thus, eqn [29] implies

$$M \to a(k) + b(k)\mathrm{e}^{2\mathrm{i}kx} \text{ as } x \to \infty \qquad [31]$$

On the other hand, eqn [27] implies that

$$M \to 1 + \frac{\mathrm{i}}{2k}\int_{-\infty}^{\infty} \mathrm{d}\xi(-1 + \mathrm{e}^{2\mathrm{i}k(x-\xi)}q(\xi)M(k,\xi)) \quad x \to \infty \qquad [32]$$

Comparing eqns [31] and [32], we find eqns [30].

The expression for $a(k)$ implies that this function is also holomorphic for $\mathrm{Im}\, k > 0$.

In summary, in the "direct problem," we have found particular solutions of eqn [5a] which are sectionally holomorphic:

$$\begin{pmatrix} M(k,x) \\ N(k,x) \end{pmatrix} \text{ and } \begin{pmatrix} M(-k,x) \\ N(-k,x) \end{pmatrix}$$

are holomorphic for $\mathrm{Im}\, k > 0$ and $\mathrm{Im}\, k < 0$, respectively. These solutions, which are characterized in terms of $q$ by eqns [27] and [28], are simply related by eqn [29].

### The Inverse Problem

Equation [28] expresses $N$ in terms of $q$. Is it possible to find an alternative expression for $N$ in terms of some appropriate "spectral data"? The answer is positive and is a direct consequence of the fact that eqn [29] defines the "jump condition" of an RH problem. Indeed, it can be shown that $a(k)$ may have simple zeros $k_1, \ldots, k_n$ in the positive imaginary axis of the $k$-complex plane. Hence, in general, $M/a$ can be expressed in the form

$$\frac{M(k,x)}{a(k)} = \mathcal{M}(k,x) + \sum_{j=1}^{n} \frac{A_j(x)}{k - \mathrm{i}p_j}, \quad p_j > 0$$

where $\mathcal{M}(k,x)$ as a function of $k$ is holomorphic for $\mathrm{Im}\, k > 0$. It can also be shown that $A_j(x) = C_j \exp[-2p_j, x]N(k_j, x)$. Hence eqn [29] becomes

$$\mathcal{M}(k,x) - N(-k,x) = \sum_{j=1}^{n} \frac{C_j \mathrm{e}^{-2p_j x} N(\mathrm{i}p_j, x)}{k - \mathrm{i}p_j} + \rho(k)\mathrm{e}^{2\mathrm{i}kx}N(k,x), \quad k \in \mathbb{R}$$

Taking the $(-)$ projection of this equation, and using the fact that both $\mathcal{M}$ and $N$ tend to 1 as $k \to \infty$, we find

$$N(k,x) - \frac{1}{2\mathrm{i}\pi}\int_{-\infty}^{\infty} \frac{\mathrm{d}l \rho(l)\mathrm{e}^{2\mathrm{i}lx}N(l,x)}{l + k + \mathrm{i}0} = 1 - \sum_{j=1}^{n} \frac{C_j \mathrm{e}^{-2p_j x}}{k + \mathrm{i}p_j} N(\mathrm{i}p_j, x) \qquad [33]$$

In summary, this equation expressed $N(k,x)$ in terms of the scattering data $(\rho(k), \{C_j, p_j\}_1^n)$.

Since both eqns [28] and [33] are associated with the same $q$, these equations can be used to obtain the following expression for $q$:

$$q = -2\frac{\partial}{\partial x}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}l \rho(l)\mathrm{e}^{2\mathrm{i}lx}N(l,x) - \mathrm{i}\sum_{j=1}^{n} C_j e^{-2p_j x} N(\mathrm{i}p_j, x)\right] \qquad [34]$$

Indeed, eqn [28] implies

$$\lim_{k\to\infty} N(k,x) = 1 - \frac{\mathrm{i}}{2k}\int_{x}^{\infty} \mathrm{d}\xi q(\xi)$$

Comparing this expression with the large-$k$ behavior of eqn [33], we find [34].

### Time Dependence of the Scattering Data

We now use eqn [5b] to compute the time dependence of the scattering data by evaluating eqn [5b] as $x \to -\infty$ we find $\nu = 4\mathrm{i}k^3$. Then, evaluating it as $x \to \infty$ and using

$$\phi \sim a\mathrm{e}^{-\mathrm{i}kx} + b\mathrm{e}^{\mathrm{i}kx}, \quad x \to +\infty$$

we find

$$a_t = 0, \qquad b_t = 8\mathrm{i}k^3 b$$

Hence,

$$a(t,k) = a(0,k), \qquad \rho(t,k) = \rho(0,k)\mathrm{e}^{8\mathrm{i}k^3 t} \qquad [35]$$

Thus,

$$p_j(t) = p_j(0), \qquad C_j(t) = C_j(0)\mathrm{e}^{8p_j^3 t} \qquad [36]$$

The above formal results motivate the following definitions (for simplicity, we assume that $a(k)$ has no zeros). Given a decaying real function

$q_0(x)$, $x \in \mathbb{R}$, define $M_0(k,x)$ as the solution of the linear Volterra integral equation

$$M_0(k,x) = 1 + \frac{\mathrm{i}}{2k}\int_{-\infty}^{x} \mathrm{d}\xi\left(-1 + \mathrm{e}^{2\mathrm{i}k(x-\xi)}q(\xi)M_0(k,\xi)\right)$$
$$\operatorname{Im} k \geq 0$$

Given $M_0(k,x)$, define $a_0(k)$ and $b_0(k)$ by

$$M_0(k,x) \to a_0(k) + b_0(k)\mathrm{e}^{2\mathrm{i}kx}, \quad x \to \infty, \quad k \in \mathbb{R}$$

Given $a_0$ and $b_0$, define $N(k,x,t)$ by the solution of the linear integral equation

$$N(k,x,t) - \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}l \frac{b_0(l)}{a_0(l)} \mathrm{e}^{8\mathrm{i}l^3 t + 2\mathrm{i}lx} \frac{N(l,x,t)}{l + k + \mathrm{i}0} = 1$$

A theorem of Gohberg and Krein implies that this equation has a unique global solution. Given $a_0, b_0, N$, define $q(x,t)$ by

$$q(x,t) = -\frac{1}{\pi}\frac{\partial}{\partial x}\int_{-\infty}^{\infty} \mathrm{d}k \frac{b_0(k)}{a_0(k)} \mathrm{e}^{8\mathrm{i}k^3 t + 2\mathrm{i}kx} N(k,x,t)$$

Then it can be shown that $q(x,t)$ satisfies the KdV equation and $q(x,0) = q_0(x)$.

## A Unification

After the emergence of a method for solving the initial-value problem for nonlinear integrable evolution equations in one and two space variables, the most outstanding open problem in the analysis of these equations became the solution of initial boundary-value problems. A general approach for solving such problems for evolution equations in one space dimension was provided by Fokas (1997). This approach has already been used for the study of nonlinear integrable evolution PDEs on the half-line (Fokas 2002, 2005), on the interval, and in a time-dependent domain. An important advantage of this new method is that it yields the formulation of a matrix RH problem (or a $\bar{\partial}$ problem in the case of a convex time-dependent domain), which although has more complicated jump matrices than the analogous problem on the infinite line, it still has an explicit exponential $(x,t)$ dependence. This fact allows one to describe effectively the asymptotic properties of the solution, using the powerful Deift–Zhou method (Deift and Zhou 1993). For example, the long-time asymptotics of boundary-value problems on the half line are discussed in Fokas and Its (1996).

It is remarkable that the above results have motivated the discovery of a new method for solving boundary-value problems, not only for linear evolution PDEs, but also for linear elliptic PDEs in two dimensions. This includes the Laplace, the biharmonic and the Helmholtz equations in a convex polygon (Dassios and Fokas 2005). In a most recent development, this method has also been applied to certain classes of linear PDEs with variable coefficients. This highly unexpected development unifies and extends several classical branches of mathematics. In particular, it unifies the classical transform methods for simple linear PDEs as well as the method of images, the treatment of linear PDEs via certain ingenious techniques such as the Wiener–Hopf technique, the formulation of Ehrenpreis type integral representations, and the solution of integrable nonlinear PDEs via the inverse-scattering transform. Furthermore, it extends these results to arbitrary domains and to certain classes of PDEs with variable coefficients.

Regarding linear equations we note the following:

Almost as soon as linear two-dimensional PDEs made their appearance, d'Alembert and Euler discovered a general approach for constructing large classes of their solutions. This approach involved separating variables and superimposing solutions of the resulting ODEs. The method of separation of variables naturally led to the solution of PDEs by a transform pair. The prototypical such pair is the direct and the inverse Fourier transforms; variations of this fundamental transform include the Laplace, Mellin, sine, cosine transforms, and their discrete analogs.

The proper transform for a given boundary-value problem is specified by the PDE, by the domain, and by the given boundary conditions. For some simple boundary-value problems, there exists an algorithmic procedure for deriving the associated transform. This procedure involves constructing the Green's function of a single eigenvalue equation, and integrating this Green's function in the $k$-complex plane, where $k$ denotes the eigenvalue.

The transform method has been enormously successful for solving a great variety of initial- and boundary-value problems. However, for sufficiently complicated problems the classical transform method fails. For example, there does not exist a proper analog of the sine transform for solving a third-order evolution equation on the half-line. Similarly, there do not exist proper transforms for solving boundary-value problems for elliptic equations even of second order and in simple domains. The failure of the transform method led to the development of several ingenious but *ad hoc* techniques, which include: conformal mappings for the Laplace and the biharmonic equations; the Jones method and the formulation of the Wiener–Hopf factorization problem; the use of some integral representation, such as that of Sommerfeld; the

formulation of a difference equation, such as the Malyuzhinet's equation. The use of these techniques has led to the solution of several classical problems in acoustics, diffraction, electromagnetism, fluid mechanics, etc. The Wiener–Hopf technique played a central role in the solution of many of these problems.

A crucial role in the new method is played by the global equation satisfied by the boundary values of $q$ and of its derivatives. For evolution equations and for elliptic equations with simple boundary conditions, this involves the solution of a system of algebraic equations, while for elliptic equations with arbitrary boundary conditions, it involves the solution of an RH problem. For simple polygons, this RH problem is formulated on the infinite line, thus it is equivalent to a Wiener–Hopf problem. This explains the central role played by the Wiener–Hopf technique in many earlier works.

For linear PDEs, the explicit $x_1, x_2$ dependence of $q(x_1, x_2)$ is consistent with the Ehrenpreis formulation of the solution. Thus, this method provides the concrete implementation as well as the generalization to concave domains of this fundamental principle. For nonlinear equations, it provides the extension of the Ehrenpreis principle to integrable nonlinear PDEs.

*See also:* Boundary value Problems for Integrable Equations; $\bar{\partial}$-Approach to Integrable Systems; Integrable Systems and Algebraic Geometry; Integrable Discrete Systems; Integrable Systems and Discrete Geometry; Integrable Systems in Random Matrix Theory; Integrable Systems: Overview; Korteweg–de Vries Equation and Other Modulation Equations; Partial Differential Equations: Some Examples; Riemann–Hilbert Methods in Integrable Systems; Sine-Gordon Equation; Toda lattices; Twistor Theory: Some Applications [in Integrable Systems, Complex Geometry and String Theory].

## Further Reading


Ablowitz MJ and Segur H (1977) Exact linearization of a Painlevé transcendent. *Physical Review Letters* 38: 1103–1106.

Ablowitz MJ, Yaakov DB, and Fokas AS (1983) On the inverse scattering transform for the Kadomtsev–Petviashvili equation. *Studies in Applied Mathematics* 69: 135–142.

Beals R and Coifman RR (1982) Scattering, transformations spectrales, et equations d'evolution nonlineaire. I. In: *Seminaire Goulaouic–Meyer–Schwartz, Exposè 21*. École Polytechnique, Palaiseau.

Calogero F (1991) In: Zakharov VE (ed.) *What Is Integrability?*, Springer.

Dassios G and Fokas AS (2005) The basic elliptic equations in an equilateral triangle. *Proceedings of the Royal Society of London, Series A* 461: 2721–2748.

Deift PA and Zhou X (1993) A steepest descent method for oscillatory Riemann–Hilbert problems. *Annals of Mathematics* 137: 245–338.

Flaschka H and Newell AC (1980) Monodromy and spectrum-preserving deformation I. *Communications in Mathematical Physics* 76: 65–116.

Fokas AS (1997) A unified transform method for solving linear and certain nonlinear PDE's. *Proceedings of the Royal Society of London, Series A* 453: 1411–1443.

Fokas AS (2002) Integrable nonlinear evolution equations on the half-line. *Communications in Mathematical Physics* 230: 1–39.

Fokas AS (2005) A generalised Dirichlet to Neumann map for certain nonlinear evolution PDEs. *Communications on Pure and Applied Mathematics* 58: 639–670.

Fokas AS and Ablowitz MJ (1983a) On the initial value problem of the second Painlevé transcendent. *Communications in Mathematical Physics* 91: 381–403.

Fokas AS and Ablowitz MJ (1983b) On the inverse scattering of the time dependent Schrödinger equation and the associated KPI equation. *Studies in Applied Mathematics* 69: 211–228.

Fokas AS and Its AR (1996) The linearization of the initial-boundary value problem of the nonlinear Schrödinger equation. *SIAM Journal on Mathematical Analysis* 27: 738–764.

Fokas AS and Santini PM (1990) Dromions and a boundary value problem for the Davey–Stewartson I equation. *Physica D* 44: 99–130.

Fokas AS and Zhou X (1992) On the solvability of Painlevé II and IV. *Communications in Mathematical Physics* 144: 601–622.

Fokas AS, Its AR, Kapaev AA, and Novokshenov VY (2006) *Painlevé Transcendents: The Riemann–Hilbert Approach*. Providence, RI: American Mathematical Society.

Gardner CS, Greene JM, Kruskal MD, and Miura RM (1967) Method for solving the Korteweg–de Vries equation. *Physical Review Letters* 19: 1095–1097.

Hietarinta J (2002) Scattering of solitons and dromions. In: Sabatier P and Pike E (eds.) *Scattering*. San Diego: Academic Press.

Lax PD (1968) Integrals of nonlinear equations of evolution and solitary waves. *Communications on Pure and Applied Mathematics* 21: 467–490.

Zakharov V and Shabat A (1972) Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media. *Soviet Physics – JEPT* 34: 62–69.

# Integrable Systems in Random Matrix Theory

**C A Tracy**, University of California at Davis, Davis, CA, USA
**H Widom**, University of California at Santa Cruz, Santa Cruz, CA, USA



An earlier version of this article was originally published as "Distribution functions for largest eigenvalues and their applications". *Proceedings of the International Congress of Mathematics, Volume 1* (2002), pp. 587–596. Beijing, China: Higher Education Press, with permission.

## Random Matrix Models

A random matrix model is a probability space $(\Omega, \mathcal{P}, \mathcal{F})$ where the sample space $\Omega$ is a set of matrices. There are three classic finite $N$ random matrix models (see, e.g., Mehta (1991)):

1. Gaussian orthogonal ensemble ($\beta = 1$):
   (a) $\Omega = N \times N$ real symmetric matrices;
   (b) $\mathcal{P} =$ "unique" measure that is invariant under orthogonal transformations and the matrix elements are i.i.d. random variables; explicitly, the density is

$$c_N \exp\left(-\mathrm{tr}(A^2)\right) \mathrm{d}A \qquad [1]$$

   where $c_N$ is a normalization constant and $\mathrm{d}A = \prod_i \mathrm{d}A_{ii} \prod_{i<j} \mathrm{d}A_{ij}$, the product Lebesgue measure on the independent matrix elements.
2. Gaussian unitary ensemble ($\beta = 2$):
   (a) $\Omega = N \times N$ Hermitian matrices;
   (b) $\mathcal{P} =$ "unique" measure that is invariant under unitary transformations and the (independent) real and imaginary matrix elements are i.i.d. random variables; and
3. Gaussian symplectic ensemble ($\beta = 4$) (see Mehta (1991) for a definition).

Generally speaking, the interest lies in the $N \to \infty$ limit of these models. Here we concentrate on one aspect of this limit. In all three models the eigenvalues, which are random variables, are real and with probability 1 they are distinct. If $\lambda_{\max}(A)$ denotes the largest eigenvalue of the random matrix $A$, then for each of the three Gaussian ensembles we introduce the corresponding distribution function

$$F_{N,\beta}(t) := \mathrm{P}_\beta(\lambda_{\max} < t), \quad \beta = 1, 2, 4$$

The basic limit laws (see Tracy and Widom (1996) and references therein) state that

$$F_\beta(s) := \lim_{N\to\infty} F_{N,\beta}\left(2\sigma\sqrt{N} + \frac{\sigma s}{N^{1/6}}\right), \quad \beta = 1, 2, 4 \qquad [2]$$

exist and are given explicitly by

$$\begin{aligned} F_2(s) &= \det\left(I - K_{\mathrm{Airy}}\right) \\ &= \exp\left(-\int_s^\infty (x - s) q^2(x)\, \mathrm{d}x\right) \end{aligned}$$

where

$$K_{\mathrm{Airy}} \doteq \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\mathrm{Ai}(y)}{x - y}$$
$$\text{acting on } L^2(s, \infty) (\text{Airy kernel})$$

and $q$ is the unique solution to the Painlevé II equation

$$q'' = sq + 2q^3$$

satisfying the condition

$$q(s) \sim \mathrm{Ai}(s) \quad \text{as } s \to \infty$$

$\sigma$ in eqn [2] is the standard deviation of the Gaussian distribution on the off-diagonal matrix elements. For the normalization we have chosen $\sigma = 1/\sqrt{2}$; however, for subsequent comparisons, the normalization $\sigma = \sqrt{N}$ is perhaps more natural.

The orthogonal and symplectic distribution functions are

$$F_1(s) = \exp\left(-\frac{1}{2}\int_s^\infty q(x)\, \mathrm{d}x\right)(F_2(s))^{1/2}$$
$$F_4(s/\sqrt{2}) = \cosh\left(\frac{1}{2}\int_s^\infty q(x)\, \mathrm{d}x\right)(F_2(s))^{1/2}$$

Graphs of the densities $\mathrm{d}F_\beta/\mathrm{d}s$ are in the adjacent figure and some statistics of $F_\beta$ can be found in **Figure 1**.

The Airy kernel is an example of an integrable integral operator and a general theory is developed in Tracy and Widom (1994). A vertex operator approach to these distributions (and many other closely related distribution functions in random matrix theory) was initiated by Adler, Shiota, and van Moerbeke (see the review article var Moerbeke (2001) for further developments of this latter approach).

Historically, the discovery of the connection between Painlevé functions ($P_{\mathrm{III}}$ in this case) and Toeplitz/Fredholm determinants appears in work of Wu *et al.* (1976) on the spin–spin correlation functions of the two-dimensional Ising model. Painlevé functions first appear in random matrix theory in

| $\beta$ | $\mu_\beta$ | $\sigma_\beta$ | $S_\beta$ | $K_\beta$ |
|---|---|---|---|---|
| 1 | –1.20653 | 1.2680 | 0.293 | 0.165 |
| 2 | –1.77109 | 0.9018 | 0.224 | 0.093 |
| 4 | –2.30688 | 0.7195 | 0.166 | 0.050 |

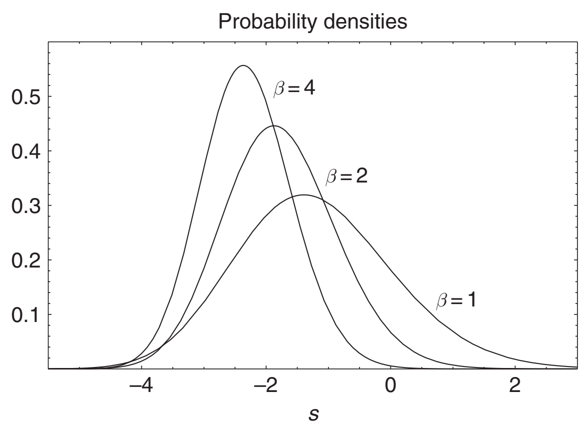



**Figure 1** The mean ($\mu_\beta$), standard deviation ($\sigma_\beta$), skewness ($S_\beta$), and kurtosis ($K_\beta$) of $F_\beta$.

Jimbo *et al.* (1980) where they prove that the Fredholm determinant of the sine kernel is expressible in terms of $P_V$. Gaudin (using Mehta's then newly invented method of orthogonal polynomials (Porter 1965)) was the first to discover the connection between random matrix theory and Fredholm determinants.

### Universality Theorems

A natural question is to ask whether the above limit laws depend upon the underlying Gaussian assumption on the probability measure. To investigate this for unitarily invariant measures ($\beta = 2$), one replaces in [1]

$$\exp\left(-\mathrm{tr}(A^2)\right) \rightarrow \exp(-\mathrm{tr}(V(A)))$$

Bleher and Its (1999) choose

$$V(A) = gA^4 - A^2, \quad g > 0$$

and subsequently a large class of potentials $V$ was analyzed by Deift *et al.* (1999). These analyses require proving new Plancherel–Rotach type formulas for nonclassical orthogonal polynomials. The proofs use Riemann–Hilbert methods. It was shown that the generic behavior is GUE; hence, the limit law for the largest eigenvalue is $F_2$. However, by finely tuning the potential new universality classes will emerge at the edge of the spectrum. For $\beta = 1, 4$ a universality theorem was proved by Stojanovic (2000) for the quartic potential.

In the case of noninvariant measures, Soshnikov (1999) proved that for real symmetric Wigner matrices (complex Hermitian Wigner matrices), the limiting distribution of the largest eigenvalue is $F_1$ (respectively, $F_2$). (A symmetric Wigner matrix is a random matrix whose entries on and above the main diagonal are independent and identically distributed random variables with distribution function $F$. Soshnikov assumes that $F$ is even and all moments are finite.) The significance of this result is that non-Gaussian Wigner measures lie outside the "integrable class" (e.g., there are no Fredholm determinant representations for the distribution functions) yet the limit laws are the same as in the integrable cases.

## Appearance of $F_\beta$ in Limit Theorems

In this section we briefly survey the appearances of the limit laws $F_\beta$ in widely differing areas.

### Combinatorics

A major breakthrough occurred with the work of Baik, Deift, and Johansson (see Baik *et al.* (2000) and references therein) when they proved that the limiting distribution of the length of the longest increasing subsequence in a random permutation is $F_2$. Precisely, if $\ell_N(\sigma)$ is the length of the longest increasing subsequence in the permutation $\sigma \in S_N$, then

$$P\left(\frac{\ell_N - 2\sqrt{N}}{N^{1/6}} < s\right) \rightarrow F_2(s)$$

as $N \rightarrow \infty$. Here the probability measure on the permutation group $S_N$ is the uniform measure. Further discussion of this result can be found in Johansson (2000b).

Baik and Rains (2001) showed by restricting the set of permutations (and these restrictions have natural symmetry interpretations) that $F_1$ and $F_4$ also appear. Even the distributions $F_1^2$ and $F_2^2$ (Tracy and Widom 1999) arise. By the Robinson–Schensted–Knuth correspondence, the Baik–Deift–Johansson result is equivalent to the limiting distribution on the number of boxes in the first row of random standard Young tableaux. (The measure is the push-forward of the uniform measure on $S_N$.) These same authors conjectured that the limiting distributions of the number of boxes in the second, third, etc., rows were the same as the limiting distributions of the next-largest, next-next-largest, etc., eigenvalues in GUE. Since these eigenvalue distributions were also found in Tracy and Widom (1996), they were able to compare the then unpublished numerical work of Odlyzko and Rains (2000) with the predicted results of random matrix theory. Subsequently, Baik *et al.* (2000) proved the conjecture for the second row. The full conjecture was proved by Okounkov (2000) using topological methods and by,

among others, Johansson (2001) using analytical methods. For an interpretation of the Baik–Deift–Johansson result in terms of the card game patience sorting, see the very readable review paper by Aldous and Diaconis (1999).

### Growth Processes

Growth processes have an extensive history both in the probability literature and the physics literature (see, e.g., Meakin (1998) and references therein), but it was only recently that Johansson (2002b) proved that the fluctuations about the limiting shape in a certain growth model ("corner growth model") are $F_2$. Johansson further pointed out that certain symmetry constraints (inspired from the Baik and Rains (2001) work) lead to $F_1$ fluctuations (*see* Growth Processes in Random Matrix Theory).

Subsequently, Baik and Rains (2000) and Gravner *et al.* (2002) have shown the same distribution functions appearing in closely related lattice growth models. Prähofer and Spohn (2000) reinterpreted the work of Baik *et al.* in terms of the physicists' polynuclear growth (PNG) model thereby clarifying the role of the symmetry parameter $\beta$. For example, $\beta=2$ describes growth from a single droplet, whereas $\beta=1$ describes growth from a flat substrate. They also related the distribution functions $F_\beta$ to fluctuations of the height function in the KPZ equation (Kardar *et al.* 1986, Meakin 1998). (The connection with the KPZ equation is heuristic.) Thus, one expects on physical grounds that the fluctuations of any growth process falling into the $1+1$KPZ universality class will be described by the distribution functions $F_\beta$ or one of the generalizations by Baik and Rains (2000). Such a physical conjecture can be tested experimentally. Earlier Myllys *et al.* established experimentally that a slow, flameless burning process in a random medium (paper!) is in the $1+1$KPZ universality class. This sequence of events is a rare instance in which new results in mathematics inspire new experiments in physics.

In the context of the PNG model, Prähofer and Spohn have given a process interpretation, the Airy process, of $F_2$.

There is an extension of the growth model in Gravner *et al.* (2002) to growth in a random environment. In Gravner *et al.* (2002) the following model of interface growth in two dimensions is considered by introducing a height function on the sites of a one-dimensional integer lattice with the following update rule: the height above the site $x$ increases to the height above $x-1$, if the latter height is larger; otherwise, the height above $x$ increases by 1 with probability $p_x$. It is assumed that the $p_x$ are chosen independently at random with a common distribution function $F$, and that the initial state is such that the origin is far above the other sites. In the pure regime, Gravner–Tracy–Widom identify an asymptotic shape and prove that the fluctuations about that shape, normalized by the square root of the time, are asymptotically normal. This contrasts with the quenched version: conditioned on the environment and normalized by the cube root of time, the fluctuations almost surely approach the distribution function $F_2$. We mention that these same authors find, under some conditions on $F$ at the right edge, a composite regime where now the interface fluctuations are governed by the extremal statistics of $p_x$ in the annealed case while the fluctuations are asymptotically normal in the quenched case.

### Random Tilings

The Aztec diamond of order $n$ is a tiling by dominoes of the lattice squares $[m,m+1]\times[\ell,\ell+1], m,n\in\mathbb{Z}$, that lie inside the region $\{(x,y):|x|+|y|\le n+1\}$. A domino is a closed $1\times 2$ or $2\times 1$ rectangle in $\mathbb{R}^2$ with corners in $\mathbb{Z}^2$. A typical tiling is shown in **Figure 2**. One observes that near the center the tiling appears random, called the temperate zone, whereas near the edges the tiling is frozen, called the polar zones. As $n\to\infty$ the boundary between the temperate zone and the polar zones (appropriately scaled) converges to a circle ("arctic circle theorem"). Johansson (2002a) proved that the fluctuations about this limiting circle are $F_2$.

### Statistics

Johnstone (2001) considers the largest principal component of the covariance matrix $X^tX$ where $X$ is an $n\times p$ data matrix all of whose entries are independent standard Gaussian variables and proves that for appropriate centering and scaling, the limiting distribution equals $F_1$ in the limit $n,p\to\infty$ with $n/p\to\gamma\in\mathbb{R}^+$. Soshnikov has removed the Gaussian assumption but requires that $n-p=\mathrm{O}(p^{1/3})$. Thus, we can anticipate applications of the distributions $F_\beta$ (and particularly $F_1$) to the statistical analysis of large data sets.

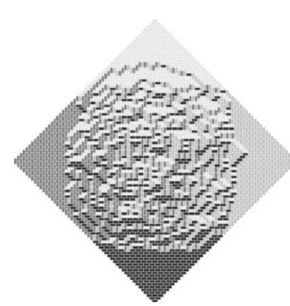


**Figure 2** Random tilings.

### Queuing Theory

Glynn and Whitt (1991) consider a series of $n$ single-server queues each with unlimited waiting space with a first-in and first-out service. Service times are i.i.d. with mean one and variance $\sigma^2$ with distribution $V$. The quantity of interest is $D(k,n)$, the departure time of customer $k$ (the last customer to be served) from the last queue $n$. For a fixed number of customers, $k$, they prove that

$$\frac{D(k,n)-n}{\sigma\sqrt{n}}$$

converges in distribution to a certain functional $\hat{D}_k$ of $k$-dimensional Brownian motion. They show that $\hat{D}_k$ is independent of the service time distribution $V$. It was shown in, for example, Gravner *et al.* (2002) that $\hat{D}_k$ is equal in distribution to the largest eigenvalue of a $k\times k$ GUE random matrix. This fascinating connection has been greatly clarified in recent work of O'Connell and Yor (2002).

From Johansson (2002), it follows for $V$ Poisson that

$$P\left(\frac{D(\lfloor xn\rfloor,n)-c_1 n}{c_2 n^{1/3}}<s\right)\to F_2(s)$$

as $n\to\infty$ for some explicitly known constants $c_1$ and $c_2$ (depending upon $x$).

### Superconductors

Vavilov *et al.* (2001) have conjectured (based upon certain physical assumptions supported by numerical work) that the fluctuation of the excitation gap in a metal grain or quantum dot induced by the proximity to a superconductor is described by $F_1$ for zero magnetic field and by $F_2$ for nonzero magnetic field. They conclude their paper with the remark:

> The universality of our prediction should offer ample opportunities for experimental observation.

## Acknowledgments

This work was supported by the National Science Foundation through grants DMS-9802122 and DMS-9732687.

*See also:* Determinantal Random Fields; Growth Processes in Random Matrix Theory; Integrable Systems and Algebraic Geometry; Integrable Systems and the Inverse Scattering Method; Integrable Systems: Overview; Quantum Calogero–Moser Systems; Random Partitions; Random Matrix Theory in Physics; Symmetry Classes in Random Matrix Theory; Toeplitz Determinants and Statistical Mechanics.

## Further Reading


Aldous D and Diaconis P (1999) Longest increasing subsequences: from patience sorting to the Baik–Deift–Johansson theory. *Bulletin of American Mathematical Society* 36: 413–432.

Baik J, Deift P, and Johansson K (2000) On the distribution of the length of the second row of a Young diagram under Plancherel measure. *Geometry of Functional Analysis* 10: 702–731.

Baik J and Rains EM (2000) Limiting distributions for a polynuclear growth model. *Journal of Statistical Physics* 100: 523–541.

Baik J and Rains EM (2001) Symmetrized random permutations. In: Bleher P and Its A (eds.) *Random Matrix Models and Their Applications*, Math. Sci. Res. Inst. Publications, vol. 40, pp. 1–19. Cambridge: Cambridge University Press.

Bleher P and Its A (1999) Semiclassical asymptotics of orthogonal polynomials, Riemann–Hilbert problem, and universality in the matrix model. *Annals of Mathematics* 150: 185–266.

Deift P, Kriecherbauer T, McLauglin K.T-R, Venakides S, and Zhou X (1999) Uniform asymptotics for polynomials orthogonal with respect to varying exponential weight and applications to universality questions in random matrix theory. *Communications on Pure and Applied Mathematics* 52: 1335–1425.

Glynn PW and Whitt W (1991) Departure times from many queues. *Annals of Applied Probability* 1: 546–572.

Gravner J, Tracy CA, and Widom H (2002) A growth model in a random environment. *Annals of Probability* 30: 1340–1368.

Jimbo M, Miwa T, Môri Y, and Sato M (1980) Density matrix of an impenetrable Bose gas and the fifth Painlevé transcendent. *Physica D* 1: 80–158.

Johansson K (2000) Shape fluctuations and random matrices. *Communications in Mathematical Physics* 209: 437–476.

Johansson K (2001) Discrete orthogonal polynomial ensembles and the Plancherel measure. *Annals of Mathematics* 153: 259–296.

Johansson K (2002a) Non-intersecting paths, random tilings and random matrices. *Probability Theory and Related Fields* 123: 225–280.

Johansson K (2002b) Toeplitz determinants, random growth and determinantal processes. In: *Proceedings of the ICM*, Beijing, ICM vol. 3, pp. 53–62.

Johnstone I (2001) On the distribution of the largest principal component. *Annals of Statistics* 29: 295–327.

Kardar M, Parisi G, and Zhang Y-C (1986) Dynamical scaling of growing interfaces. *Physical Review Letters* 56: 889–892.

Meakin P (1998) *Fractals, Scaling and Growth Far from Equilibrium.* Cambridge: Cambridge University Press.

Mehta ML (1991) *Random Matrices*, 2nd edn. Academic Press.

Myllys M, Maunuksela J, Alava M, Ala-Nissila T, Merikoski J, and Timonen J Kinetic roughening in slow combustion of paper. *Physical Review E* 64: 036101-1–036101-12.

Odlyzko AM and Rains EM On the longest increasing subsequences in random permutations. In: Grinberg EL, Berhanu S, Knopp M, Mendoza G, and Quinto ET (eds.) *Analysis, Geometry, Number Theory: The Mathematics of Leon Ehrenpreis*, American Mathematical Society. pp. 439–451.

O'Connell N and Yor M (2002) A representation for non-colliding random walks. *Electronic Communications in Probability* 7: 1–12.

Okounkov A (2000) Random matrices and random permutations. *International Mathematics Research Notices* 20: 1043–1095.

Porter CE (1965) *Statistical Theories of Spectra: Fluctuations.* Academic Press.

Prähofer M and Spohn H (2000) Universal distributions for growth processes in $1+1$ dimensions and random matrices. *Physical Review Letters* 84: 4882–4885.

Soshnikov A (1999) Universality at the edge of the spectrum in Wigner random matrices. *Communications in Mathematical Physics* 207: 697–733.

Stojanovic A (2000) Universality in orthogonal and symplectic invariant matrix models with quartic potential. *Mathematical Physics Analysis and Geometry* 3: 339–373.

Tracy CA and Widom H (1994) Fredholm determinants, differential equations and matrix models. *Communications in Mathematical Physics* 163: 151–174.

Tracy CA and Widom H (1996) On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics* 177: 727–754.

Tracy CA and Widom H (1999) Random unitary matrices, permutations and Painlevé. *Communications in Mathematical Physics* 207: 665–685.

van Moerbeke P (2001) Integrable lattices: random matrices and random permutations. In: Blcher P and Its A (eds.) *Random Matrix Models and Their Applications*, Math. Sci. Res. Inst. Publications vol. 40, pp. 321–406. Cambridge: Cambridge Univ. Press.

Vavilov MG, Brouwer PW, Ambegaokar V, and Beenakker CWJ (2001) Universal gap fluctuations in the superconductor proximity effect. *Physical Review Letters* 86: 874–877.

Wu TT, McCoy BM, Tracy CA, and Barouch E (1976) Spin–spin correlation functions of the two-dimensional Ising model: exact theory in the scaling region. *Physical Review B* 13: 316–374.

# Integrable Systems: Overview

**Francesco Calogero**, University of Rome, Rome, Italy and Institute Nazionale di Fisica Nucleare, Rome, Italy



## Introduction

This section introduces some elementary notions and sets the (mathematically low brow) tone of this presentation.

A *dynamical system* is characterized by an evolution equation the general structure of which reads

$$Q_t = F \qquad [1]$$

Here $Q \equiv Q(x,t)$ is the dependent variable, and it might be a scalar, a vector, a matrix, you name it. The focus of interest is on its evolution as function of the (real, scalar) "time" variable $t$. The *a priori* unknown quantity $Q$ might moreover depend on another independent "space" variable (scalar or vector) $x$, $Q \equiv Q(x,t)$. The appended variable $t$ in the left-hand side of the above equation denotes partial differentiation, and this notation will be used throughout, although when $t$ is the only independent variable differentiation with respect to it might be instead denoted by a superimposed dot:

$$Q_t \equiv \frac{\partial Q(x,t)}{\partial t},\ Q_x \equiv \frac{\partial Q(x,t)}{\partial x},\ \dot{Q} \equiv \frac{dQ(t)}{dt}$$

The quantity in the right-hand side of the evolution equation (1), which has of course the same (scalar, vector, matrix) character as $Q$, is an *assigned* function of $t$, $x$ and $Q$, $F \equiv (x,t,Q)$ (more generally, its dependence on $Q$ might be functional, see below). A typical example of the dynamical systems we shall consider is the $N$-body problem characterized by the Newtonian equations of motion

$$\ddot{q}_n = -\omega^2 q_n + 2g^2 \sum_{m=1,\, m\neq n}^{N} (q_n - q_m)^{-3},\ n = 1, 2, \ldots N \qquad [2]$$

where the dependent variable is the $N$-vector $\vec{q} \equiv (q_1, \ldots q_N)$, the components of which are the "particle coordinates" $q_n \equiv q_n(t)$. Note however that these equations of motion are of *second-order* in time (contrary to (1)); but they can of course be reformulated as *first-order* ODEs indeed their Hamiltonian version, derived in the standard manner from the Hamiltonian

$$H = \frac{1}{2}\sum_{n=1}^{N}(p_n^2 + \omega^2 q_n^2) + \frac{g^2}{2}\sum_{m,n=1;m\neq n}^{N}(q_n - q_m)^{-2} \qquad [3a]$$

reads

$$\dot{q}_n = p_n \qquad [3b]$$

$$\dot{p}_n = -\omega^2 q_n + 2g^2 \sum_{m=1,\, m\neq n}^{N} (q_n - q_m)^{-3},\ n = 1, 2, \ldots N \qquad [3c]$$

Other typical examples are the ("Korteweg-de Vries", "Burgers", "Nonlinear Schrödinger", "sine Gordon") PDEs satisfied by the scalar dependent variable $q \equiv q(x,t)$,

$$q_t = -q_{xxx} + 2q_x q = \left(-q_{xx} + q^2\right)_x \qquad [4]$$

$$q_t = -q_{xx} + 2q_x q = \left(-q_x + q^2\right)_x \qquad [5]$$

$$q_t = i\left[q_{xx} + s|q|^2 q\right],\ s = \pm \qquad [6]$$

$$q_t - q_x = s,\, s_t + s_x = \sin q \qquad [7]$$

as well as the integrodifferential ("Benjamin–Ono") equation

$$q_t = P\int_{-\infty}^{\infty} dy \frac{q_{yy}(y)}{x-y} + q_x q \qquad [8]$$

and the ("Kadomtsev–Petviashvili") PDE satisfied by the scalar dependent variable $q \equiv q(x, y, t)$,

$$q_{tx} = (-q_{xxx} + q_x q)_x + s q_{yy},\ s = \pm \qquad [9]$$

This last equation should of course be reformulated as an integrodifferential equation to fit with (1).

These are all examples of *integrable systems* (see below). In this presentation we restrict attention to dynamical systems of these general types, without considering evolutions in which the space variable, and/or the time variable, and/or the dependent variable, only take discrete values, forsaking thereby the discussion of discrete evolution equations, cellular automata and functional equations, see other entries of this Encyclopedia. We shall consider mainly the "initial-value problem" in which the solution is assigned at the initial time, say at $t = 0$,

$$Q(x, 0) = Q_0(x)$$

and the subsequent evolution of the dependent variable, namely the values taken by $Q(x, t)$ for $t > 0$, is the focus of attention. Note however that, except when there is no dependence at all on the space variable $x$ (see for instance (2)), the functional class to which $Q(x, t)$ belongs as regards its $x$-dependence should be specified (and the assigned initial-value $Q_0(x)$ should of course belong to this functional class). A typical class of functions are those vanishing (adequately fast) at (spatial) infinity; another typical class are those characterized by periodicity properties as functions of $x$; and still another class are those restricted to a finite spatial domain (for instance, the positive $x$-axis, $x > 0$, or a finite interval, $a \le x \le b$), in which cases the initial-value problem must be supplemented by assigning boundary conditions. These latter class of problems, called initial/boundary-value problems, are generally more difficult; even the identification of which boundary conditions are adequate to identify uniquely the solution may be a nontrivial task. In the following we will always focus on the simpler class of problems characterized by solutions defined in the entire space region and vanishing (sufficiently fast) asymptotically (far away).

Thus, in the spirit of the initial-value problem, a dynamical system is generally characterized by assigning its evolution equation, the functional class to which its solutions are required to belong, and possibly in addition some (additional) restriction on the set of initial data.

Let us finally mention that, aside from considering the initial-value problem, the study of dynamical systems may focus on the identification of special (classes of) solutions, for instance those obtained by using symmetry properties of the evolution equation under consideration (yielding, say, "similarity solutions"), and, in the integrable case, "solitonic" and "multisolitonic" solutions (see below).

## Integrable dynamical systems

The solution of a dynamical system, however simple the equation that defines its time evolution, see (1), may be extremely complicated, indeed its time-dependence might feature one or more of the characteristics of *deterministic chaos*, such as a *sensitive dependence* on the initial data. But there are "exceptional" dynamical systems, the behavior of which is instead, in some sense, *simple*. Such systems are termed – in the least technical sense of the word – "integrable".

This characterization can be made precise for Hamiltonian systems with a finite number $N$ of degrees of freedom, the equations of motion of which read

$$\dot{q}_n = \frac{\partial H(\vec{p}, \vec{q})}{\partial p_n}, \dot{p}_n = -\frac{\partial H(\vec{p}, \vec{q})}{\partial q_n}, n = 1, \ldots N$$

Such a system is *integrable* if there exist, in addition to the Hamiltonian $H(\vec{p}, \vec{q}) \equiv H^{(1)}(\vec{p}, \vec{q})$ itself, $N - 1$ other (nontrivial and functionally independent) constants of motion $H^{(m)}(\vec{p}, \vec{q})$ *in involution*, namely such that their Poisson brackets vanish:

$$\left\{H^{(n)}, H^{(m)}\right\} \equiv \sum_{\ell=1}^{N}\left[\frac{\partial H^{(n)}(\vec{p}, \vec{q})}{\partial q_\ell}\frac{\partial H^{(m)}(\vec{p}, \vec{q})}{\partial p_\ell} - \frac{\partial H^{(m)}(\vec{p}, \vec{q})}{\partial q_\ell}\frac{\partial H^{(n)}(\vec{p}, \vec{q})}{\partial p_\ell}\right] = 0, \quad n, m = 1, \ldots, N$$

Let us however emphasize the crucial role of the words "there exist", as used just above. For definiteness let us require that the constants of motion $H^{(n)}(\vec{p}, \vec{q})$ be *analytic functions* of their $2N$ arguments, and *not excessively multivalued*: they might feature some branch points, but not so many to vanify their effectiveness in constraining the time evolution of the dynamical variables $q_n(t), p_n(t)$ sufficiently to avoid their behavior from being too complicated. On the other hand it is of course not necessary that these functions $H^{(n)}(\vec{p}, \vec{q})$ be *explicitly* known.

When these conditions hold it is in principle possible ("Liouville theorem") to identify a

*canonical transformation* from the canonical coordinates and momenta $q_n$ and $p_n$ to *action-angle variables* $\theta_n$ and $I_n$ such that

$$I_n = H^{(n)}(\vec{p}, \vec{q}) \qquad [10]$$

Then these action variables evolve trivially,

$$I_n(t) = I_n(0), \theta_n(t) = \theta_n(0) + I_n(0)t, n = 1, \ldots N$$

Note that, once these new canonical variables are identified, the solution of the initial-value problem for the original Hamiltonian problem is provided directly by the expressions of the action-angle variables $\theta_n$ and $I_n$ in terms of the original variables $q_n$ and $p_n$, as well as the expressions of the latter in terms of the former. The second step of this procedure requires inverting the expressions (10), and the corresponding expressions of the angle variables $\theta_n$ in terms of the original variables $q_n$ and $p_n$; a necessary condition in order that this step allow to identify uniquely, at least in principle, the original canonical variables $q_n$ and $p_n$ in terms of the action-angle variables $I_n$ and $\theta_n$ – hence imply a *simple* time-evolution of these original variables – is the requirement, as mentioned above, that the expressions of the constants of the motion $H^{(n)}(\vec{p}, \vec{q})$ in terms of their arguments $q_n$ and $p_n$ not be excessively multivalued.

The statements outlined above can be rigorously formulated for finite-dimensional Hamiltonian systems, and they can be heuristically extended to all analogous dynamical systems with a finite number of degrees of freedom, even if they are not Hamiltonian.

A system with $N$ degrees of freedom might possess more than $N$ constants of motion. Such a system that possesses $2N - 1$ (nontrivial and functionally independent) constants of motion (the maximal number, to avoid the evolution being frozen) is called *superintegrable*, and its evolution is in some sense analogous to that of a system with a single degree of freedom, in particular *all* its *confined* and *nonsingular* motions are then *completely periodic*,

$$q_n(t + T) = q_n(t), p_n(t + T) = p_n(t), n = 1, \ldots, N$$

The period $T$ depends generally on the initial data. If it does not, at least for an open set of such data having full dimensionality in phase space, the system is called *isochronous*: *all* its motions in that phase space region are then *completely periodic with the same period*.

A dynamical system might be integrable in a region of its "natural" phase space, and nonintegrable in another region. Sometimes such systems are referred to as *partially integrable*. There even are systems which are *isochronous* (hence *superintegrable*) in a region of their phase space, and behave instead *chaotically* in another region. These regions are generally separated by boundaries where the evolution of the system runs into singularities, and the constants of motion associated with the integrable behavior become *excessively multivalued* in the regions where the behavior is *chaotic*. (*see* Isochronous Systems).

Dynamical systems featuring an additional space variable $x$ (see Section 1) can be interpreted as *infinite-dimensional* dynamical systems (by considering the variable $x$ as a continuous label for the dependent variable $Q$). Accordingly, a necessary condition in order that such systems be considered *integrable* is the requirement that they possess an *infinite* number of constants of the motion. But – even for such systems that allow a Hamiltonian formulation – this condition cannot be considered sufficient (due to the inherent ambiguities in the counting of infinities), and in fact a completely cogent, universally accepted definition of *integrability* for infinite-dimensional dynamical systems is still lacking (various definitions can of course be given in special contexts). It is nevertheless rather well understood by practitioners what is meant by such a term at least for *integrable* equations such as those indicated at the end of the previous section, which generally give rise to the *solitonic* phenomenology – as explained below.

The study of integrable systems has an illustrious history, to which many eminent mathematicians and mathematical physicists contributed after the Newtonian revolution: Euler, Jacobi, Poincaré, Painlevé, Kowalewskaya, Kolmogorov, Moser ... Below we report – most tersely – on the bloom that this topic has witnessed over the last 3–4 decades, without being generally able, due to space constraints, to attribute the appropriate credit to the many colleagues, most of them still living, who contributed to this endeavor. For more detailed treatments of the topics outlined below, of related developments not mentioned here, and of such credits, the interested reader is referred to the bibliography given below, including the additional references traceable from there.

### Integrable many-body problems

An important class of integrable dynamical systems is provided by $N$-body problems characterized by Hamiltonians such as

$$H(\vec{p}, \vec{q}) = \frac{1}{2}\sum_{n=1}^{N} p_n^2 + V(\vec{q}) \qquad [11]$$

with a potential energy $V(\vec{q})$ that includes "external" and "two-body" forces,

$$V(\vec{q}) = \sum_{n=1}^{N} V^{(1)}(q_n) + \frac{1}{2}\sum_{m,n=1;m\neq n}^{N} V^{(2)}(q_n - q_m),$$
$$V^{(2)}(-q) = V^{(2)}(q) \qquad [12]$$

The corresponding Hamiltonian and Newtonian equations of motion read

$$\dot{q}_n = p_n, \dot{p}_n = -\frac{\partial V^{(1)}(q_n)}{\partial q_n} - \sum_{m=1, m\neq n}^{N} \frac{\partial V^{(2)}(q_n - q_m)}{\partial q_n},$$

$$\ddot{q}_n = -\frac{\partial V^{(1)}(q_n)}{\partial q_n} - \sum_{m=1, m\neq n}^{N} \frac{\partial V^{(2)}(q_n - q_m)}{\partial = q_n} \quad [13]$$

**The Lax pair and the constants of motion** Suppose that two $N \otimes N$ matrices $L \equiv L(\vec{p}, \vec{q})$ and $M \equiv M(\vec{p}, \vec{q})$ could be found such that the matrix "Lax equation"

$$\dot{L} = [L, M] \quad [14]$$

be equivalent to the Hamiltonian equations of motion (13). Here and throughout the notation $[A, B]$ denotes the commutator:

$$[A, B] \equiv A\,B - B\,A$$

Because this matrix equation clearly entails that the $N$ traces

$$T_n = \text{trace}\,[L^n], n = 1, \ldots, N$$

are constants of the motion,

$$\dot{T}_n = 0, n = 1, \ldots, N$$

the possibility to write the Hamiltonian equations (13) in the Lax form (14) yields as a bonus $N$ constants of the motion, namely it entails that the Hamiltonian system under consideration is *integrable*. (One must moreover show that these constants of motion are *in involution*; this is usually the case).

Hence a route to identify *integrable* $N$-body problems is via the search of Lax pairs $L$, $M$ of matrices such that (14) correspond to (13), with an appropriate assignment of the potential energy (12). For $N > 2$ this is a nontrivial task, because (13) is a system of $2N$ ODEs in $2N$ unknowns, while the matrix Lax equation (14) amounts to a system of $N^2$ ODEs.

**Functional equations and the identification of integrable many-body problems** A convenient *ansatz* to identify a *Lax pair* suitable for the purpose outlined above reads as follows:

$$L_{nm} = p_n \text{ for } n = m, L_{nm} = \alpha(q_n - q_m) \text{ for } m \neq n,$$

$$M_{nm} = \sum_{\ell=1, \ell\neq n}^{N} \beta(q_n - q_\ell) \text{ for } n = m,$$

$$M_{nm} = \gamma(q_n - q_m), \text{ for } m \neq n$$

where $\alpha(q), \beta(q)$ and $\gamma(q)$ are 3 functions to be determined. It is then easily seen that these functions may be assigned so that the corresponding Lax equation (14) be equivalent to the Hamiltonian equations (13) with

$$V^{(1)}(q) = 0 \quad [15a]$$

$$V^{(2)}(q) = \alpha(q)\alpha(-q) \quad [15b]$$

provided the function $\alpha(x)$ satisfies the *functional equation*

$$\frac{\alpha(x)\alpha'(y) - \alpha(y)\alpha'(x)}{\alpha(x+y)} = \beta(x) - \beta(y), \beta(x) = \beta(-x)$$

The general solution of this functional equation yields via [15b] the two-body potential

$$V^{(2)}(q) = g^2\, a^2 \wp(a\,q|\omega, \omega') \quad [16]$$

where $g$ and $a$ are two arbitrary constants and $\wp(x|\omega, \omega')$ is the Weierstrass elliptic function (with semiperiods $\omega$ and $\omega'$, as well arbitrary). One concludes therefore that the $N$-body problem characterized by the Hamiltonian (11) with (12), (15a) and (16) is *integrable*.

This Hamiltonian system has played, since the mid-seventies, a seminal role in the developments of finite-dimensional integrable systems that occurred over the last few decades. However, since the Weierstrass function is doubly-periodic, from a "physical" point of view this $N$–body problem is rather unrealistic, or perhaps rather suited for the study of crystalline configurations, including their statistical mechanics. But there are two special cases, obtained by assigning an infinite value to one or both of the semiperiods of the Weierstrass function in (16), that qualify $V^{(2)}(q)$ as a physical two-body potential:

$$V^{(2)}(q) = \frac{g^2\, a^2}{\sinh^2(a\,q)} \quad [17a]$$

$$V^{(2)}(q) = \frac{g^2}{q^2} \quad [17b]$$

(Of course the second of these two-body potentials, (17b), is merely the special case of the first, (17b), corresponding to $a = 0$). These Hamiltonian models are then naturally interpretable as one-dimensional many-body problems with repulsive two-body forces singular at zero separation and vanishing at large distances. Actually the fact that these systems are *integrable* is far from remarkable, since it is generally true that any many-body problem characterized by repulsive forces vanishing at large distances (hence causing *unconfined* motions) is *integrable*: indeed in such models the particles eventually separate and move freely, so that their trajectories cannot display the extreme complication

characterizing a *chaotic* (i.e., *nonintegrable*) behavior. But these models are in fact *superintegrable* and they (as well as various *integrable* extensions of them) feature many (physically and mathematically) interesting properties. For instance the asymptotic behavior of their trajectories,

$$q_n(t) = p_n^{(\pm)}t + q_n^{(\pm)} + o(1), p_n(t) = p_n^{(\pm)} + o(1) \quad \text{as } t \to \pm\infty, n = 1, \ldots, N \tag{18}$$

is characterized by the simple rules

$$p_n^{(+)} = p_{N+1-n}^{(-)}, n = 1, \ldots, N,$$
$$q_n^{(+)} = q_n^{(-)} + \sum_{m=1, m\neq n}^{N} \Delta\left(p_m^{(-)} - p_n^{(-)}; g, a\right) \quad n = 1, \ldots, N \tag{19}$$

with

$$\Delta(p; g, a) = \text{sign}(p)\frac{\log\left[1 + (ga/p)^2\right]}{2a}$$

The formula (19) indicates that the shift $q_n^{(+)} - q_n^{(-)}$ among the asymptotic positions of the particles (see (18)) is merely a sum of *two-body* shifts $\Delta$ (which incidentally vanish altogether if $a = 0$, namely in the (17b) case), and it only depends on the velocities $p_n^{(-)}$ of the particles in the remote past (*not* on the corresponding asymptotic positions $q_n^{(-)}$, in spite of their relevance in determining the order in which the different particles approach each other through the motion).

A generalization of the above model in the (17b) case – nontrivial inasmuch as it yields *confined* motions – is characterized by the additional presence in the potential (12) of the one-body potential

$$V^{(1)}(q) = \tfrac{1}{2}\,\omega^2 q^2 \tag{20}$$

yielding the Hamiltonian (3a). This model is *integrable*, indeed *superintegrable*, indeed *isochronous*, all its (real) solutions being *completely periodic* with period

$$T = \frac{2\pi}{\omega} \tag{21}$$

A neat way to understand this result is by noting that, if $\tilde{q}(t)$ is a (possibly *complex*) solution of the model discussed above (in this subsection, with the two-body potential (17b) and no one-body potential, see (15a)), then

$$q_n(t) = \exp(-i\omega t)\tilde{q}_n(\tau), \tau = \frac{\exp(2\,i\omega t) - 1}{2\,i\omega}$$

provides a (possibly *real*) solution of the Newtonian equations of motion (2), namely of the same model but with the additional one-body potential (20). Remarkably this model was solved firstly in the quantal case (at the beginning of the seventies), and only a few years later in the classical case considered here (by J. Moser, who, for the $\omega = 0$ case, introduced the special version of the Lax matrix appropriate for this case).

Another class of many-body problems, introduced in the mid-sixties by M. Toda, played a seminal role in the study of *integrable* dynamical systems, indeed the first application (independently by H. Flaschka and S. Manakov) of the Lax approach to *integrable* many-body problems occurred in that context. This model is often referred to as the Toda lattice, because its (two-body) interaction (of exponential type) is only assumed to act among "nearest neighbors".

A particularly interesting, and just as integrable, generalization of this class of Hamiltonian many-body problems features an extra parameter, say $c$, which might be considered to play the role of "speed of light". These models reduce to those considered above for $c = \infty$, and for finite $c$ they are invariant under the Poincaré group of coordinate transformations (while of course the many-body problems described above are invariant under the Galilei group). They are sometimes termed RS models, to recognize those who first introduced them (S. Ruijsenaars and H. Schneider) as well as the possibility to interpret them in some sense as "relativistic" generalizations of the "nonrelativistic" models described above.

**Reduction of the solution to algebraic operations** The solution of the models described above can actually be reduced to purely algebraic operations. For instance for the model characterized by the Newtonian equations of motion (2) such a solution of the initial-value problem is provided by the following prescription: the particle coordinates $q_n(t)$ coincide with the $N$ eigenvalues of the $N \otimes N$ matrix:

$$\tilde{Q}_{nm}(t) = q_n(0)\cos(\omega t) + \dot{q}_n(0)\frac{\sin(\omega t)}{\omega} \text{ for } n = m,$$
$$\tilde{Q}_{nm}(t) = \frac{ig\ \sin(\omega t)}{\omega[q_n(0) - q_m(0)]} \text{ for } n \neq m$$

**Many-body problems related to the motion of the zeros of linear PDEs** Another convenient approach to manufacture and investigate *integrable* many-body problems is by identifying the motion of the particles with that of the zeros of (polynomial)

solutions of linear (hence solvable) evolution PDEs. Assume for instance that the monic polynomial

$$\psi(z,t) = x^N + \sum_{m=1}^{N} c_m(t) x^{N-m} = \prod_{n=1}^{N} [z - z_n(t)] \qquad [22]$$

satisfies the (compatible) linear PDE

$$\begin{aligned}&\left[A_0 + A_1 z + A_2 z^2 + A_3 z^3\right]\psi_{zz}\\ &\quad + \left[B_0 + B_1 z - 2(N-1)A_3 z^2\right]\psi_z\\ &\quad + C\psi_{tt} + [E - (N-1)D_2 z]\psi_t\\ &\quad + \left[D_0 + D_1 z + D_2 z^2\right]\psi_{zt}\\ &\quad - [N(N-1)(A_2 - A_3 z) + NB_1]\psi = 0 \qquad [23]\end{aligned}$$

where the letters $A_0, A_1, A_2, A_3, B_0, B_1, C, D_0, D_1, D_2, E$ denote 11 arbitrary constants. Then the zeros $z_n(t)$ evolve according to the system of ODEs

$$\begin{aligned}C\ddot{z}_n + E\dot{z}_n &= B_0 + B_1 z_n - 2(N-1)A_3 z_n^2\\ &\quad + \sum_{m=1,\, m\neq n}^{N} (z_n - z_m)^{-1}\\ &\quad \times [2C\dot{z}_n\dot{z}_m - (\dot{z}_n + \dot{z}_m)(D_0 + D_1 z_n)\\ &\quad - D_2 z_n(\dot{z}_n z_m + \dot{z}_m z_n)\\ &\quad + 2\left(A_0 + A_1 z_n + A_2 z_n^2 + A_3 z_n^3\right)] \qquad [24]\end{aligned}$$

interpretable as the Newtonian equations of motion of an $N$-body problem with one- and two-body (velocity-dependent) forces. This problem is *integrable*, indeed its solution can be reduced to the algebraic problem of finding the zeros of the polynomial $\psi(z,t)$, see (22), whose time evolution can be ascertained by solving the linear PDE (23), itself a purely algebraic problem as it amounts to solving the system of (constant coefficients, linear) ODEs implied via (22) by this PDE (23) for the $N$ coefficients $c_m(t)$.

This class of many-body problems is rather rich, thanks to the arbitrariness of the 11 constants it features. Several subcases, characterized by special choices of these constants, are suitable to display a gamut of different phenomenological behaviors: confined and nonconfined motions, periodic and nonperiodic evolutions, limit cycles, Hamiltonian cases, . . . .

**Solvable many-body problems in the plane** The many-body problems considered above were all essentially *one-dimensional*. But via a simple trick it is possible to obtain from some of them *many-body problems in the plane* (which should of course be *rotation-invariant* to be certified as such). Consider for instance the special case of the above model, (24), with $C=1$ and with $A_0 = A_1 = A_3 = B_0 = D_0 = D_2 = 0$ so that its equations of motion,

$$\begin{aligned}\ddot{z}_n + E\dot{z}_n &= B_1 z_n + \sum_{m=1,\, m\neq n}^{N} (z_n - z_m)^{-1}\\ &\quad \times \left[2\dot{z}_n\dot{z}_m - D_1(\dot{z}_n + \dot{z}_m)z_n + 2A_2 z_n^2\right] \qquad [25]\end{aligned}$$

are invariant under rescaling of the dependent variables ($z_n \Longrightarrow c z_n$). Let us then assume to work in the *complex* rather than the *real*, and let us set

$$E = \gamma + i\omega,\ A_2 = \alpha + i\tilde{\alpha},\ B_1 = \beta + i\tilde{\beta},\\ D_1 = \delta + i\tilde{\delta}$$

where the Greek letter indicate now *real* constants, and let us moreover relate the $N$ complex coordinates $z_n$ to $N$ two-vectors $\vec{r}_n$ in the horizontal plane via the self-evident positions

$$z_n = x_n + iy_n, \vec{r}_n = (x_n, y_n, 0); \hat{k} = (0,0,1) \qquad [26]$$

It is then easily seen that the *integrable* equations of motion (25) become the following *rotation-invariant* Newtonian equations of motion identifying a (no less *integrable*) $N$-body problem *in the plane*:

$$\begin{aligned}&\ddot{\vec{r}}_n + \left(\gamma + \omega\hat{k}\wedge\right)\dot{\vec{r}}_n\\ &\quad = \left(\beta + \tilde{\beta}\hat{k}\wedge\right)\vec{r}_n + \sum_{m=1,\, m\neq n}^{N} r_{nm}^{-2}\\ &\quad \Big[2\left\{\dot{\vec{r}}_n\left(\dot{\vec{r}}_m\cdot\vec{r}_{nm}\right) + \dot{\vec{r}}_m\left(\dot{\vec{r}}_n\cdot\vec{r}_{nm}\right) - \vec{r}_{nm}\left(\dot{\vec{r}}_n\cdot\dot{\vec{r}}_m\right)\right\}\\ &\quad - \left(\delta + \tilde{\delta}\hat{k}\wedge\right)\Big\{\left(\dot{\vec{r}}_n + \dot{\vec{r}}_m\right)\left[r_n^2 - (\vec{r}_n\cdot\vec{r}_m)\right]\\ &\quad - \vec{r}_n\left[\vec{r}_m\cdot\left(\dot{\vec{r}}_n + \dot{\vec{r}}_m\right)\right] + \vec{r}_m\left[\vec{r}_n\cdot\left(\dot{\vec{r}}_n + \dot{\vec{r}}_m\right)\right]\Big\}\\ &\quad + 2\left(\alpha + \tilde{\alpha}\hat{k}\wedge\right)\left\{\vec{r}_n\left[r_n^2 - 2(\vec{r}_n\cdot\vec{r}_m)\right] + \vec{r}_m r_n^2\right\}\Big] \qquad [27]\end{aligned}$$

Here and below we use the short-hand notation $\vec{r}_{nm} = \vec{r}_n - \vec{r}_m$ entailing $r_{nm}^2 = r_n^2 + r_m^2 - 2\vec{r}_n\cdot\vec{r}_m$, the symbol $\wedge$ denotes the three-dimensional vector product so that $\hat{k}\wedge\vec{r}_n = (-y_n, x_n, 0)$ (see (26)), and the rest of the notation is self-evident. Note that these *rotation-invariant* Newtonian equations of motion are also *translation-invariant* if $\beta = \tilde{\beta} = \delta = \tilde{\delta} = \alpha = \tilde{\alpha} = 0$.

**The "goldfish" model** The attribute of "goldfish" has been attributed to the special case of the above model with *all* "coupling constants" vanishing, thanks to the neatness of its equations of motion, which in their *complex* version read

$$\ddot{z}_n = 2\sum_{m=1,\, m\neq n}^{N} \frac{\dot{z}_n\dot{z}_m}{z_n - z_m}, n = 1, \ldots, N$$

and in their *real* ("physical") version as Newtonian equations of motion of an $N$-body problem in the horizontal plane read

$$\ddot{\vec{r}}_n = 2 \sum_{m=1, m\neq n}^{N} \frac{\dot{\vec{r}}_n\left(\dot{\vec{r}}_m \cdot \vec{r}_{nm}\right) + \dot{\vec{r}}_m\left(\dot{\vec{r}}_n \cdot \vec{r}_{nm}\right) - \vec{r}_{nm}\left(\dot{\vec{r}}_n \cdot \dot{\vec{r}}_m\right)}{r_{nm}^2}$$
$$n = 1, \ldots, N$$

(This name has also been attributed to some extensions of this model, see the entry *Isochronous Systems* in this Encyclopedia). This model is invariant under time rescaling ($t \Rightarrow ct$), in its physical version it is *translation-* and *rotation-invariant*, it only features *two-body* forces and in spite of their *velocity-dependence* it is *Hamiltonian* (it is in fact a simple instance of the RS models mentioned above). The solution of its initial-value problem (in its *complex* version) is given by a remarkably neat rule: the $N$ coordinates $z_n(t)$ are the $N$ roots of the following algebraic equations in $z$:

$$\sum_{n=1}^{N} \frac{\dot{z}_n(0)}{z - z_n(0)} = \frac{1}{t} \quad [28]$$

The phenomenology of its *generic* solution is also remarkable, corresponding to the "game of musical chairs": in the remote past all particles but one are *almost* at rest in $N-1$ positions ("sitting in $N-1$ chairs") and one particle comes in from infinity, moving initially as a free particle; as it approaches, all the particles begin to move around ("dancing"); in the remote future one particle goes away (moving eventually with the same speed as the incoming particle), and all the others settle down in the same $N-1$ positions ("of the $N-1$ chairs"), but with the possibility that the outgoing particle be different from the incoming one, and that the other particles have reshuffled their "seating".

Another remarkable version (also *translation-* and *rotation-invariant*, as well as *Hamiltonian*) of the $N$-body model in the plane (27) obtains if all the "coupling constants" vanish except $\omega$. Then *all* its *nonsingular* solutions – which are given by the same prescription indicated just above, except for the replacement of $\frac{1}{t}$ with $\frac{i\omega}{\exp(i\omega t)-1}$ in the right-hand side of (28) – are *completely periodic* with periods which are an integer multiple – no larger than a number depending on $N$, generally (much) smaller than $N!$ – of $T$ (see (21)), the domains of phase space that give rise to solutions with different periodicity being separated from each other by boundaries characterized by lower-dimensional sets of initial data yielding trajectories that run into *singularities* corresponding to *particle collisions* (note that when two or more particles collide their individuality gets lost, and their velocities diverge).

**Integrable many-body problems in spaces with arbitrary dimensions** *Integrable*, or even *solvable*, many-body problems in spaces with more than two dimensions – with *rotation-invariant* equations of motion of Newtonian type – can be manufactured by starting from an appropriate *integrable*, or *solvable*, second-order *matrix* evolution equation, and by then parametrizing the evolving matrix in terms of multidimensional vectors so as to transform the matrix evolution equation into a *covariant* – hence *rotation-invariant* – system of evolution equations for these vectors, interpretable as Newtonian equations of motion of a many-body problem in multidimensional space.

For instance the matrix equation

$$\dot{M} = AM + MA + M^3$$

is *integrable*. Here $M \equiv M(t)$ is a square matrix of arbitrary order and $A$ is an *arbitrary* constant matrix. By parametrizing appropriately these two matrices one concludes that either one of the following two Newtonian systems of ODEs is *integrable*:

$$\ddot{\vec{r}}_{nm} = \sum_{\nu=1}^{N} \alpha_{n\nu}\vec{r}_{\nu m} + \sum_{\mu=1}^{M}\sum_{\nu=1}^{N} \vec{r}_{n\mu}\left(\vec{r}_{\nu\mu}\cdot\vec{r}_{\nu m}\right)$$
$$n = 1, \ldots, N, m = 1, \ldots, M,$$
$$\ddot{\vec{r}}_{nm} = \sum_{\nu=1}^{N} \alpha_{n\nu}\vec{r}_{\nu m} + \sum_{\mu=1}^{M}\sum_{\nu=1}^{N} \vec{r}_{\nu\mu}\left(\vec{r}_{\nu\mu}\cdot\vec{r}_{nm}\right)$$
$$n = 1, \ldots, N, m = 1, \ldots, M.$$

Here $N$ and $M$ are *arbitrary* positive integers, the $NM$ constants $\alpha_{nm}$ are also *arbitrary*, the $NM$ "particle coordinates" $\vec{r}_{nm} \equiv \vec{r}_{nm}(t)$ are $S$-vectors, with $S$ an *arbitrary* positive integer, and the dots sandwiched among these $S$-vectors denote the standard scalar product in $S$-dimensional space.

Let us emphasize the physical relevance of this class of many-body problems, characterized by *linear* and *cubic* forces. This is reinforced by the fact that these models are *Hamiltonian*.

**Nonlinear harmonic oscillators** Two classes of *integrable* systems obtain from the classes written above by first setting to zero all the constants $\alpha_{nm}$ and by then performing the change of variables

$$\vec{w}_{nm}(t) = \exp(i\omega t)\vec{r}_{nm}(\tau), \tau = \frac{\exp(i\omega t) - 1}{i\omega} \quad [29]$$

with $\omega > 0$. The corresponding Newtonian equations of motion read

$$\ddot{\vec{w}}_{nm} - 3i\omega\dot{\vec{w}}_{nm} - 2\vec{w}_{nm} = \sum_{\mu=1}^{M}\sum_{\nu=1}^{N}\vec{w}_{n\mu}(\vec{w}_{\nu\mu}\cdot\vec{w}_{\nu m})$$

$$n = 1, \ldots, N, m = 1, \ldots, M,$$

$$\ddot{\vec{w}}_{nm} - 3i\omega\dot{\vec{w}}_{nm} - 2\vec{w}_{nm} = \sum_{\mu=1}^{M}\sum_{\nu=1}^{N}\vec{w}_{\nu\mu}(\vec{w}_{\nu\mu}\cdot\vec{w}_{nm})$$

$$n = 1, \ldots, N, m = 1, \ldots, M$$

These equations of motion cause the $N\,M$ evolving $S$-vectors $\vec{w}_{nm} \equiv \vec{w}_{nm}(t)$ to be *complex* (see the second term in their left-hand sides), but a *real* system (with double the number of dependent variables) can be easily obtained by setting

$$\vec{w}_{nm} = \vec{u}_{nm} + i\vec{v}_{nm}$$

Remarkably (but clearly suggested by (29)), *all* the *nonsingular* solutions of each of these two many-body problems are *completely periodic*, with a period which is an integer multiple of the period $T$, see (21). This justifies the title given to this subsection. It also shows that these are *isochronous systems* (*see* Isochronous Systems).

## Integrable nonlinear PDEs

As indicated in Section 1 another class of integrable systems are nonlinear evolution PDEs. In this section we outline (some of) their properties, focussing mainly on the Korteweg-de Vries PDE (4), the solution of which by C. S. Gardner, J. M. Greene, M. D. Kruskal and R. M. Miura in the mid-sixties was the opening shot of a major scientific development which is still blooming. Other important early steps of this development were, in the late sixties, the introduction by P. D. Lax of what is now called the Lax pair technique, and at the beginning of the seventies the solution by V. E. Zakharov and A. B. Shabat of the Nonlinear Schrödinger equation (6) – an evolution PDE of great applicative importance. Subsequently many researchers developed various techniques to identify, classify and investigate *integrable* nonlinear PDEs, a continuing activity for an overall appraisal of which the interested reader is referred to the bibliography reported below.

Here we outline one of the approaches to obtaining these results; other approaches are tersely mentioned below.

### Identification and investigation of integrable PDEs via the inverse spectral transform technique

The class of *linear dispersive* evolution PDEs reads

$$u_t(x,t) = -i\omega\left(-i\frac{\partial}{\partial x}\right)u(x,t), -\infty < x < \infty \quad [30]$$

where the "dispersion function" $\omega(z)$ is, say, a (*real*) polynomial (which must be *odd* to guarantee that this PDE be *real*). The solution of this PDE is achieved via the introduction of the Fourier transform $\hat{u}(k,t)$,

$$u(x,t) = (2\pi)^{-1}\int_{-\infty}^{\infty} dk \exp(i\,kx)\,\hat{u}(k,t) \quad [31a]$$

$$\hat{u}(k,t) = \int_{-\infty}^{\infty} dx \exp(-i\,kx)u(x,t) \quad [31b]$$

whose evolution corresponding to (30) is then given by the simple linear ODE

$$\hat{u}_t(k,t) = -i\omega(k)\hat{u}(k,t), -\infty < k < \infty \quad [32a]$$

which can be immediately integrated:

$$\hat{u}(k,t) = \hat{u}(k,0)\exp[-i\omega(k)t] \quad [32b]$$

Thus the solution of the initial-value problem of (30) is achieved via three steps: (i) at the initial time one obtains the initial value of the Fourier transform, $\hat{u}(k,0)$, from the initial datum $u(x,0)$ (via (31b)); (ii) one then obtains $\hat{u}(k,t)$ (via (32b)); (iii) one finally obtains $u(x,t)$ (via (31a)). From these formulas the main features of the resulting phenomenology are easily evinced (even when the above integrals cannot be explicitly performed).

A class of *integrable nonlinear* evolution PDEs reads

$$u_t(x,t) = \alpha(\mathbf{R})u_x(x,t) \quad [33]$$

where the assigned function $\alpha(z)$ is again, say, a (*real*) polynomial, while $\mathbf{R}$ is now the integrodifferential "recursion operator" defined by the following formula that specifies its action on a generic function $f(x,t)$ (vanishing asymptotically so as to allow all integrations to converge):

$$\mathbf{R}\,f(x,t) = f_{xx}(x,t) - 4u(x,t)f(x,t) + 2u_x(x,t)\int_x^{\infty} dy\,f(y,t) \quad [34]$$

Note that the presence of the time variable $t$ plays no relevant role (it is merely parametric). A remarkable property of this operator – which depends on $u(x,t)$ – is that any power of it acting

on $u_x(x,t)$ yields a nonlinear combination of $u(x,t)$ and its $x$-derivatives – without any left-over integration, in fact yielding a result which is itself an exact $x$-derivative, ready for exact integration in case of a further application of **R**, see the last term in the right-hand side of (34). For instance

$$\mathbf{R}u_x = u_{xxx} - 6u_x u = \left(u_{xx} - 3u^2\right)_x,$$
$$\mathbf{R}^2 u_x = u_{xxxxx} - 10u_{xxx}u - 20u_{xx}u_x + 30u_x u^2 = \left(u_{xxxx} - 10u_{xx}u - 5u_x^2 + 10u^3\right)_x$$

and so on. Hence the simplest nonlinear evolution equation contained in the class (33) is the Korteweg-de Vries (KdV) equation

$$u_t + u_{xxx} = 6u_x u \qquad [35]$$

(corresponding to $\alpha(z) = -z$; and note the identity with (4), via the trivial rescaling $q(x,t) = 3\,u(x,t)$). Note that, if one neglects all nonlinear contributions, the class (33) reduces to (30) with

$$\omega(z) = -z\alpha\left(-z^2\right)$$

The solution of this class of *nonlinear* PDEs, (33), is given by a somewhat analogous procedure to that described above for the class of *linear* dispersive PDEs (30).

Firstly, one introduces the *spectral transform*, a nonlinear generalization of the Fourier transform which indeed reduces to it if nonlinear effects are altogether neglected. That relevant for the class of PDEs (33) is based on the spectral problem associated with the linear Schrödinger operator

$$\mathbf{L} = -\left(\frac{\partial}{\partial x}\right)^2 + u(x,t), -\infty < x < \infty \qquad [36]$$

Via it, the spectral transform

$$S[u(x,t)] = \{R(k,t), -\infty < k < \infty; p_n, \rho_n(t), n = 1, \ldots, N\} \qquad [37]$$

is introduced. Here the function $R(k,t)$ is the "reflection coefficient" associated to the eigenvalue $k^2$ of the *continuous* spectrum of **L**, while the nonnegative number $N$ gives the number of *discrete* eigenvalues of **L**, and the positive quantities $p_n$ and $\rho_n(t)$ are associated to these discrete eigenvalues, specifically $-p_n^2$ are the "binding energies", and $\rho_n(t)$ the "normalization coefficients", associated to the "bound states" possessed by the "potential" $u(x,t)$. (All this terminology comes from the interpretation of the above spectral problem in quantum-mechanical terms). And it can be shown not only that there is a one-to-one correspondence among a function $u(x,t)$ and its spectral transform $S[u(x,\,t)]$, but moreover that both the *direct spectral problem* to compute $S[u(x,t)]$ from $u(x,t)$ (arbitrarily assigned within an appropriate class), and the *inverse spectral problem* to compute $u(x,t)$ from $S[u(x,\,t)]$ (arbitrarily assigned within an appropriate class), only entail solving *linear* equations (an ODE in the former case, a Fredholm integral equation in the latter case).

Note that, in the above definition of the spectral transform, the time variable $t$ plays merely a parametric role. But the usefulness of this spectral transform to solve the PDE (33) resides in the fact that, if $u(x,\,t)$ evolves in time according to this PDE, the corresponding evolution of the spectral transform is quite simple: the number $N$ and the positive numbers $p_n$ are time-independent (as already implied by our notation), while the time evolution of the reflection coefficient $R(k,t)$ and of the normalization coefficients $\rho_n(t)$ is given by the simple *linear* ODEs

$$R_t(k,t) = 2ik\alpha\left(-4k^2\right)R(k,t), -\infty < k < \infty \qquad [38a]$$

$$\dot{\rho}_n(t) = -2p_n\alpha(4p_n^2)\rho_n(t), n = 1, \ldots, N \qquad [38b]$$

which can be readily integrated:

$$R(k,t) = R(k,0)\exp\left[2ik\alpha\left(-4k^2\right)t\right] \qquad [39a]$$

$$\rho_n(t) = \rho_n(0)\exp\left[-2p_n\alpha(4p_n^2)t\right] \qquad [39b]$$

Hence the solution of the initial-value problem for the class of nonlinear PDEs (33) can now be achieved via the following three steps: (i) at the initial time, via the solution of the *direct spectral problem*, the spectral transform $S[u(x,0)]$ (see (37)) is obtained (from $u(x,0)$, arbitrarily assigned within an appropriate class); (ii) the spectral transform at time $t$ is then obtained via (39); (iii) by solving the *inverse spectral problem*, $u(x,t)$ is obtained from $S[u(x,t)]$ (see (37)).

The analogy of this procedure to that outlined above for the class of linear dispersive PDEs (30) is clear, and the fact that in this manner the solution of the initial-value problem for the nonlinear PDEs (33) can be achieved via a sequence of steps involving only the solution of *linear* problems is an indication of the *integrable* character of this class of *nonlinear* evolution PDEs. And it allows to gain thereby a lot of insight on the behavior of these solutions, and also to construct classes of explicit solutions of these equations, as we now indicate.

### Solitons

The *integrable* nonlinear PDE (33) possesses the *single-soliton* solution

$$u(x,t) = \frac{-2p^2}{\cosh^2\{p[x-\xi(t)]\}} \qquad [40a]$$

$$\xi(t) = (2p)^{-1}\log\left[\frac{\rho(t)}{2p}\right] = \xi(0) + vt,$$
$$v = -\alpha(4p^2) \qquad [40b]$$

to which corresponds the simple *spectral transform*

$$S[u(x,t)] = \{R(k,t) = 0; p_1 = p,$$
$$\rho_1(t) = \rho(t) = \rho(0)\exp[-2p\alpha(4p^2)t]; N = 1\} \qquad [41]$$

This solution, (40), describes a localized wave of constant shape moving with the constant speed $v$: the "soliton". It is characterized by two (*real*) parameters, $\xi(0)$ and $p$. The first identifies the initial location of the *soliton*; its arbitrariness corresponds to the translation invariant character of (33). The second, $p$, the spectral significance of which is clear from (41), determines the shape of the *soliton* (both its "height" $2p^2$ and its "width" $\frac{1}{p}$) as well as its speed $v$ (see (40b)); note that the shape is identical for all the nonlinear evolution PDEs of the class (33), while the speed depends on the function $\alpha(z)$, see (40b), namely it depends on which specific equation of the class (33) one is considering. For instance for the KdV equation (35), corresponding to $\alpha(z) = -z$, the speed of the soliton is

$$v = 4p^2 \qquad [42]$$

thus *all* solitons of the KdV equation move from left to right, and taller and thinner solitons move faster than less tall and more fat ones.

More generally, every PDE of the class (33) possesses the $N$-soliton solution

$$u(x,t) = -2\left(\frac{\partial}{\partial x}\right)^2 \log\det[\mathbf{I} + \mathbf{C}(x,t)] \qquad [43a]$$

Here $\mathbf{I}$ is the $N \otimes N$ unit matrix and $\mathbf{C}(t)$ is the $N \otimes N$ matrix

$$C_{mn}(x,t) = [\rho_m(t)\rho_n(t)]^{1/2}\frac{\exp[-(p_m+p_n)x]}{p_m+p_n} \qquad [43b]$$

where the time-evolution of the $\rho_n(t)$'s is given by (39b). Indeed the spectral transform of this solution is given by (37) with $R(k,t) = 0$ and $\rho_n(t)$ given by (39b). To discuss the multisolitonic phenomenology, let us focus on the KdV equation, so that the speed of each soliton is given by the simple formula (42) and let us order the $N$ positive numbers $p_n$ in increasing order,

$$p_1 < p_2 < \cdots < p_N$$

so that the corresponding soliton velocities, $v_n = 4p_n^2$, are as well ordered in increasing order:

$$v_1 < v_2 < \cdots < v_N$$

The $N$-soliton solution (43) is not so transparent, especially if $N$ is large, but it becomes quite simple in the remote past and future:

$$u(x,t) \approx \sum_{n=1}^{N} \frac{-2p_n^2}{\cosh^2\{p_n[x-\xi_n(t)]\}},$$
$$\xi_n(t) = \xi_n^{(\pm)} + v_n t,\ t \to \pm\infty$$

with the $2N$ (*real*) constants $\xi_n^{(\pm)}$ related to one another (see below). It is thus seen that, both in the remote past and future, the $N$-soliton solution (43) splits into the sum of $N$ separated solitons. In the remote past the solitons are ranged, from left to right, in order of decreasing amplitude, and they move to the right with speeds ordered in decreasing magnitude; then the taller and faster solitons gradually catch up and eventually "overtake" the fatter and slower ones (the quotation marks underscore the fact that whenever two, or possibly more, solitons get together, their individuality is in fact lost: for a while the solution might have just one peak, or instead the "overtaking" of two solitons may rather appear as an "exchange of identity", with the taller soliton becoming fatter and the fatter becoming taller as they get close together until they separate again because the one in front, having become taller, speeds up while the one behind, having become fatter, slows down). The final outcome is of course that the order of the solitons gets altogether reversed, with the taller and faster heading the escape to the right. The most remarkable aspect of this phenomenology is that precisely the same solitons that existed in the remote past are found in the remote future, the only effect of their "interaction" having been to shift the position of the *n-th* soliton, relative to what it would have been if it had been moving in isolation, by the amount

$$\Delta_n = \xi_n^{(+)} - \xi_n^{(-)}$$

These $N$ shifts are moreover determined (while either the $N$ quantities $\xi_n^{(-)}$ or the $N$ quantities $\xi_n^{(+)}$ can be arbitrarily assigned), being given by the simple rule

$$\Delta_n = \sum_{m=1}^{n-1} \Delta(p_n, p_m) - \sum_{m=n+1}^{N} \Delta(p_n, p_m) \qquad [44a]$$

$$\Delta(p_n, p_m) = \frac{1}{p_n}\log\left(\frac{p_n + p_m}{|p_n - p_m|}\right) \qquad [44b]$$

Of course in (44a) a sum vanishes if its lower limit exceeds its upper limit.

This formula (44), has a simple phenomenological significance. From the two-soliton case ($N = 2$) it is seen that in a *two-body* encounter the taller and faster soliton gets *advanced* by the amount $\Delta(p_2, p_1)$, while the slower and fatter one gets *delayed* by the amount $\Delta(p_1, p_2)$. Hence the overall shift (44) experienced by the $n$-th soliton in the $N$-soliton case is the sum of the $n - 1$ positive shifts derived from its "overtaking" $n - 1$ slower solitons and the $N - n$ negative shifts derived from its being "overtaken" by $N - n$ faster solitons. This outcome is obvious when each two-soliton encounter occurs *separately*, but is quite nontrivial in the general case when, at some intermediate time, several solitons might all encounter simultaneously.

This soliton phenomenology strongly suggest ascribing to each soliton an individuality, even though in configuration space it only shows up as a separate entity in the remote past and future. The separated identity of each soliton is instead quite clear in the spectral transform context, since each of them corresponds to a (time-independent) discrete eigenvalue of the spectral problem. Indeed in the spectral context this identity is clear also for the generic solution of the class of integrable nonlinear PDEs (33) which, in contrast to the purely solitonic solution (43), is *not* characterized by a vanishing reflection coefficient $R(k, t)$. And indeed, even in configuration space, the soliton phenomenology described above is still featured by a *generic* solution (each of which is characterized, via its spectral transform (37), by the number $N$ of its solitons), up to the additional presence of a "background" component of this solution (corresponding to the nonvanishing reflection coefficient $R(k, t)$), which however behaves in a manner analogous to the solution of the linear, dispersive part of the PDE under consideration, becoming eventually locally small due to its dispersive character.

**Kinks, breathers, boomerons and trappons, dromions** The solitonic phenomenology described above for the class of integrable PDEs (33), and in particular for the KdV equation (35), is more or less common to all *integrable* nonlinear evolution PDEs – of which many other classes exist besides (33). But there also are some significant differences, some of which we now review tersely.

For certain *integrable* PDEs the typical shape of the soliton is not localized, but it rather has the form of a "kink". Some *integrable* PDEs also feature additional kinds of localized "solitons" which, in isolation, move overall with constant speed as ordinary solitons, but feature in addition a time-dependent amplitude modulation and are therefore called "breathers". For *integrable* matrix nonlinear evolution PDEs – or, equivalently, for *integrable* systems of coupled PDEs – the new phenomenology may emerge of solitons that, even in isolation, move with a variable speed, the change of which over time is correlated with the variable interplay of the amplitudes of the different components of the solution: typically such solitons come in from one side in the remote past and boomerang back to that side in the remote future ("boomerons"), or they may be trapped to oscillate around some fixed position ("trappons"); and there are *integrable* evolution equations in which both these types of solitons are simultaneously present in a generic solution. All these phenomenologies refer to the simpler class of *integrable* evolution PDEs in $1 + 1$ (one space and one time) variables, with asymptotically vanishing boundary conditions (at large space distances; or perhaps asymptotically constant, as in the case of kinks). There also exist *integrable* evolution PDEs in $2 + 1$ dimensions (such as the KP equation (9)) the generic solution of which may feature localized soliton-like components, although in this case appropriate boundary conditions play a crucial role (for this reason such solitons have been called "dromions", hinting at their being to some extent driven by the boundary conditions, as objects moving in a stadium).

While there are quite many (classes of) integrable PDEs in $1 + 1$ dimensions, there are only a few in $2 + 1$ dimensions, and there is a widespread belief that no integrable PDEs exist in $D + 1$ dimensions with $D > 2$. But already in the early days of soliton theory it was pointed out that there do exist quite many (classes of) integrable PDEs in $1 + D$ dimensions (namely, one space and $D$ time variables) and that it is quite possible via a different formulation of the initial-value problem to interpret such equations as (no less *integrable*) PDEs in $D + 1$ dimensions ($D$ space and one time variables); and *integrable* PDEs in $D + 1$ dimensions have also been identified and investigated in the context of (the simpler class of) *C-integrable* PDEs (see below).

### Other properties of integrable PDEs

For the *linear* evolution equations (30) the main message implied by their solvability via the Fourier transform is, that the time-evolution is much simpler in Fourier space (see (32)) than in configuration

space. This has a profound impact on the understanding of all phenomena describable by such equations, to the extent of determining the kind of experimental tools better suited to understand the underlining physics (for instance, the use of monochromatic beams of light, the use of high-energy particle accelerators, and so on). The same kind of message is as well relevant for the class of *integrable* nonlinear PDEs solvable via the spectral transform technique – even more so inasmuch as the time-evolution is in this case so much simpler in the spectral space (being actually *linear* there, see (38) and (39)) than in configuration space (where the evolution is *nonlinear*, see (33)). It is indeed the basis for the possession by the class of *integrable* nonlinear PDEs (33) of several other remarkable properties as outlined tersely in the following subsections.

**Bäcklund transformations** A Bäcklund transformation is a formula relating two functions, say $u^{(0)}(x,t)$ and $u^{(1)}(x,t)$, so that, if one of them satisfies a (generally nonlinear) PDE, the other one satisfies the same PDE. In the context of the class (33) of *integrable* PDEs, such a (class of) Bäcklund transformations is provided by the formula

$$g(\mathbf{\Lambda})\left[u^{(0)}(x,t)-u^{(1)}(x,t)\right]+h(\mathbf{\Lambda})\mathbf{\Gamma}\ 1=0 \qquad [45]$$

where $g(z)$ and $h(z)$ are two (*a priori* arbitrary) entire functions (say, two polynomials), while $\mathbf{\Lambda}$ and $\mathbf{\Gamma}$ are two integrodifferential operators the effect of which on a function $f(x,t)$ (such that all relevant integrations are convergent) reads

$$\begin{aligned}\mathbf{\Gamma}f(x,t)=&\left[u_x^{(0)}(x,t)+u_x^{(1)}(x,t)\right]f(x,t)\\&+\left[u^{(0)}(x,t)-u^{(1)}(x,t)\right]\\&\times\int_x^\infty dy\left[u^{(0)}(y,t)-u^{(1)}(y,t)\right]f(y,t)\end{aligned} \qquad [46a]$$

$$\begin{aligned}\mathbf{\Lambda}f(x,t)=&f_{xx}(x,t)-2\left[u^{(0)}(x,t)+u^{(1)}(x,t)\right]f(x,t)\\&+\mathbf{\Gamma}\int_x^\infty dyf(y,t)\end{aligned} \qquad [46b]$$

Note that here the variable $t$ plays no relevant role (its presence is merely parametric), and that $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$ depend (in a symmetrical way) on $u^{(0)}(x,t)$ and $u^{(1)}(x,t)$, whose presence causes the Bäcklund transformation (45) to be *nonlinear* in these functions. Also important is the observation that, for $u^{(0)}(x,t)=u^{(1)}(x,t)=u(x,t)$, the operator $\mathbf{\Lambda}$ becomes the recursion operator $\mathbf{R}$, see (34).

The reason why the formulas (45) constitute a class of Bäcklund transformations is because – as a property of the spectral transform based on the linear Schrödinger operator $\mathbf{L}$, see (36) – if two "potentials" $u^{(0)}(x,t)$ and $u^{(1)}(x,t)$ are related by (45), the corresponding "reflection coefficients" $R^{(0)}(k,t)$ and $R^{(1)}(k,t)$ are related *algebraically*, as follows:

$$\begin{aligned}&g(-4k^2)\left[R^{(0)}(k,t)-R^{(1)}(k,t)\right]\\&\quad+2ikh(-4k^2)\left[R^{(0)}(k,t)+R^{(1)}(k,t)\right]=0\end{aligned} \qquad [47a]$$

entailing

$$R^{(1)}(k,t)=R^{(0)}(k,t)\frac{g(-4k^2)+2ikh(-4k^2)}{g(-4k^2)-2ikh(-4k^2)} \qquad [47b]$$

Clearly this formula entails that, if $R^{(0)}(k,t)$ satisfies (38a), so does $R^{(1)}(k,t)$. Hence, as the fact that $R^{(0)}(k,t)$ satisfies (38a) is a consequence of the fact that $u^{(0)}(x,t)$ satisfies (33), likewise the fact that $R^{(1)}(k,t)$ satisfies (38a) provides the basis for concluding that $u^{(1)}(x,t)$ also satisfies (33).

The simpler version of the Bäcklund transformation (45) obtains by setting $g(z)=-2ph(z)$ with $p$ an arbitrary constant, hence it reads

$$\begin{aligned}&w_x^{(0)}(x,t)+w_x^{(1)}(x,t)\\&\quad=2p\left[w^{(0)}(x,t)-w^{(1)}(x,t)\right]\\&\qquad-\frac{1}{2}\left[w^{(0)}(x,t)-w^{(1)}(x,t)\right]^2\end{aligned} \qquad [48]$$

Here and below we use for convenience the functions $w^{(j)}(x,t)$ related to $u^{(j)}(x,t)$ as follows:

$$\begin{aligned}&w^{(j)}(x,t)=\int_x^\infty dy\,u^{(j)}(y,t),\\&w_x^{(j)}(x,t)=-u^{(j)}(x,t)\end{aligned} \qquad [49]$$

A convenient application of Bäcklund transformations is to yield new solutions of (33) from known solutions; for instance from the trivial solution $u^{(0)}(x,t)=w^{(0)}(x,t)=0$ the single-soliton solution (40) can be readily obtained via (48) and (49) (of course an appropriate time-dependence must be attributed to the $x$-independent "integration constant" that obtains from the integration of (48), which is an ODE in the independent variable $x$).

Another important property of Bäcklund transformations is their *commutativity*. Consider two sets of two polynomials, $g^{(m)}(z)$ and $h^{(m)}(z), m=1,2$, and the two Bäcklund transformations (45) they generate, say BT1 and BT2. Take as starting point some function $u^{(0)}(x)$ and associate to it two functions,

$u^{(1)}(x)$ respectively $u^{(2)}(x)$, obtained from $u^{(0)}(x)$ via these two Bäcklund transformations, BT1 respectively BT2. Then obtain a new function, say $u^{(12)}(x)$, from $u^{(1)}(x)$ via BT2; and likewise obtain $u^{(21)}(x)$ from $u^{(2)}(x)$ via BT1. The property of commutativity entails that, provided an appropriate choice is made of integration constants (see (45)),

$$u^{(12)}(x) = u^{(21)}(x) \qquad [50]$$

This property is highly nontrivial when viewed, as we just did, in configuration space; it is instead rather obvious in the spectral space, indeed the corresponding property for the "reflection coefficients" reads (in self-evident notation, see (47b))

$$R^{(12)}(k) = R^{(21)}(k) = R^{(0)}(k)\, B^{(1)}(k)\, B^{(2)}(k) \qquad [51a]$$

$$B^{(m)}(k) = \frac{g^{(m)}(-4\,k^2) + 2ikh^{(m)}(-4\,k^2)}{g^{(m)}(-4\,k^2) - 2ikh^{(m)}(-4\,k^2)}, \quad m = 1, 2 \qquad [51b]$$

hence it corresponds simply to the commutativity of the ordinary product.

**Nonlinear superposition principle** Another remarkable property of the class of evolution equations (33) is a straightforward consequence of the commutativity property, (50), of Bäcklund transformations. It reads (hereafter with a slight abuse of language we refer to "solutions" $w^{(j)}$ even though the actual solutions are the functions $u^{(j)}$ related to the $w^{(j)}$ by (49))

$$w^{(12)} = w^{(21)} = w^{(0)} - \frac{2(p_1+p_2)\left(w^{(1)} - w^{(2)}\right)}{2(p_1-p_2) + w^{(1)} - w^{(2)}} \qquad [52]$$

where $w^{(0)} \equiv w^{(0)}(x,t)$ is an arbitrary solution of (33), $w^{(1)} \equiv w^{(1)}(x,t)$ respectively $w^{(2)} \equiv w^{(2)}(x,t)$ are likewise the solutions of the same PDE related to $w^{(0)}$ by the Bäcklund transformation (48) with $p = p_1$ respectively $p = p_2$, and $w^{(12)}(x,t) = w^{(21)}(x,t)$ is another solution of the same PDE. Note that this formula, for which the title of this subsection seems appropriate, provides a completely explicit, rational expression of a new solution of (33) in terms of three other solutions of the same equation: an arbitrary solution $w^{(0)}$, and the two solutions $w^{(1)}$ and $w^{(2)}$ related to it by a simple Bäcklund transformation, see (48).

**Soliton ladder** A simple application of the preceding formula is to start from the trivial solution

$$w^{(0)} = 0 \qquad [53]$$

so that (see (48))

$$w^{(j)}(x,t) = -2p_j\left[1 - \tan\left\{p_j\left[x - x_0^{(j)} + \alpha(4p^2)t\right]\right\}\right], \quad j = 1, 2 \qquad [54a]$$

where, in order that this function be real, either

$$\mathrm{Im}\left[x_0^{(j)}\right] = 0 \qquad [54b]$$

or

$$\mathrm{Im}\left[x_0^{(j)}\right] = \frac{\pi}{2p_j} \qquad [54c]$$

Via (49), the expression (54a) with (54b) yields, for each value of $j$, a version of the *single-soliton* solution (40). Insertion of (53) and (54a) in (52) yields, via (49), the *two-soliton* solution of (33), provided $0 < p_1 < p_2$ and $x_0^{(1)}$ satisfies (54b) while $x_0^{(2)}$ satisfies (54c) (otherwise the solution produced by (52) is complex or singular).

Having thus obtained the *two-soliton* solution, one can apply the nonlinear superposition formula (52) to get the three-soliton solutions, by inserting in place of $w^{(0)}$ the *single-soliton* expression (54a) (with parameter, say, $p_1$) and in place of $w^{(1)}$ and $w^{(2)}$ the *two-soliton* expression (with parameters $p_1$ and $p_2$ respectively $p_1$ and $p_3$); and the process can be continued, as suggested by the title of this subsection. In this manner the multisolitonic solution can be constructed by a sequence of purely algebraic operations: and simple rules can be given, detailing the restrictions on the soliton parameter $p_n$ and the reality properties of the constants $x_0^{(n)}$ ((54b) or (54c)) to insure that the solution so arrived at be real and nonsingular, and thus coincide with (43).

**Conservation laws** As mentioned above, *integrable* evolution PDEs are interpretable as *infinite-dimensional* dynamical systems. It is therefore natural that they possess an *infinite* number of conserved quantities. For instance every PDE of the class [33] possesses the following infinite sequence of conserved quantities:

$$C_n = \frac{(-1)^n}{2n+1}\int_{-\infty}^{\infty} dx\, \mathbf{R}^n[xu_x(x,t) + 2u(x,t)], \quad n = 0, 1, 2, \ldots, \qquad [55a]$$

where $\mathbf{R}$ is the recursion operator (34). An alternative definition for this sequence is

$$C_n = \frac{(-1)^n}{2n+1}\int_{-\infty}^{\infty} dx \tilde{\mathbf{R}}^n u(x,t), \quad n = 0, 1, 2, \ldots, \qquad [55b]$$

where the integrodifferential operator $\tilde{\mathbf{R}}$ is in some sense the adjoint of $\mathbf{R}$, being defined by the formula

$$\tilde{\mathbf{R}}f(x,t) = f_{xx}(x,t) - 4u(x,t)f(x,t) + 2\int_x^\infty dy\, u(y,t)f_y(y,t) \qquad [55c]$$

that specifies its action on a generic function $f(x,t)$ (such that the integration converge). The first 3 of these conserved quantities read as follows:

$$C_0 = \int_{-\infty}^{\infty} dx\, u(x,t),$$

$$C_1 = \int_{-\infty}^{\infty} dx\, u^2(x,t),$$

$$C_2 = \int_{-\infty}^{\infty} dx\, \left[2\,u^3(x,t) + u_x^2(x,t)\right]$$

These constants of the motion (55) are functionally independent and, in the context of a Hamiltonian formulation characterized by the Poisson bracket

$$\{A,B\} = \int_{-\infty}^{\infty} dx\, \frac{\delta A}{\delta u(x)} \frac{\partial}{\partial x} \frac{\delta B}{\delta u(x)}$$

(where $A$ and $B$ are functionals of $u(x)$ and $\delta/\delta u(x)$ denotes the functional derivative), they are *in involution*,

$$\{C_n, C_m\} = 0$$

Note that, in this context, the KdV PDE (35) coincides with the Hamiltonian equation

$$u_t(x,t) = \{u(x,t), H\} = \left(\frac{\partial}{\partial x}\right) \frac{\delta H}{\delta u(x,t)}$$

with

$$H = \frac{1}{2}C_2 = \frac{1}{2}\int_{-\infty}^{\infty} dx\, \left[2\,u^3(x,t) + u_x^2(x,t)\right]$$

Several alternative sequences of constants of motion also exist. For instance another infinite sequence is provided by the two equivalent formulas

$$c_n = (-1)^n \int_{-\infty}^{\infty} dx\, \hat{\mathbf{R}}^{2n} \cdot 1 \qquad [56a]$$

$$c_n = (-1)^n \int_{-\infty}^{\infty} dx\, \mathbf{\Lambda}_0^n\, u(x,t) \qquad [56b]$$

with the integrodifferential operators $\hat{\mathbf{R}}$ and $\mathbf{\Lambda}_0$ defined by the formulas

$$\hat{\mathbf{R}}f(x,t) = f_x(x,t) - \int_{-\infty}^{x} dy\, u(y,t)\, f(y,t),$$

$$\mathbf{\Lambda}_0 f(x,t) = f_{xx}(x,t) - 2\,u(x,t)\, f(x,t) + u_x(x,t)\int_x^\infty dy\, f(y,t) + u(x,t)\int_x^\infty dy\, u(y,t)\int_y^\infty dz\, f(z,t)$$

Note that the integrodifferential operator $\mathbf{\Lambda}_0$ is just $\mathbf{\Lambda}$, see (46), with $u^{(0)}(x,t)=0$ and $u^{(1)}(x,t)=u(x,t)$.

The constants $c_n$ are also all independent of each other, but there is a relationship between the constants of the two sequences, (55) and (56),

$$\sum_{n=0} c_n\, z^{2n+1} = \sin\left[\sum_{n=0} C_n\, z^{2n+1}\right]$$

which is to be understood by expanding the right-hand side in powers of $z$ and then equating the coefficients of equal powers of $z$:

$$c_0 = C_0,$$
$$c_1 = C_1 - \tfrac{1}{6}C_0^3,$$
$$c_2 = C2 - \tfrac{1}{2}C_0^2 C_1 + \tfrac{1}{120}C_0^5$$

and so on.

Of course all these conservation laws are applicable to the class of solutions of (33) defined for all (*real*) values of $x$ and vanishing asymptotically (as $x \to \pm\infty$). But they can also be reformulated as *local* "continuity equations". And – rather remarkably – all these results hold as well for the explicitly time-dependent class of PDEs that obtains if one allows the polynomial $\alpha(z)$ in the right-hand side of (33) to feature an *arbitrary* time-dependence, say

$$\alpha(z,t) = \sum_{m=0}^{M} \alpha_m(t) z^m \qquad [57]$$

Finally let us note that there is an additional conserved quantity for this (generalized) class of PDEs,

$$C = \int_{-\infty}^{\infty} dx\left[xu(x,t) + \int_0^t dt'\alpha(\tilde{\mathbf{R}},t')u(x,t)\right]$$

with $\tilde{\mathbf{R}}$ defined by (55c). This implies that, for the generic solution of this (generalized) class of PDEs the *center of mass*

$$X(t) = \frac{\int_{-\infty}^{\infty} dx\, x\, u(x,t)}{\int_{-\infty}^{\infty} dx\, u(x,t)}$$

moves according to the formula

$$X(t) = X_0 + \sum_{m=0}^{M} (-1)^{m+1} (2\,m+1) \left(\frac{C_m}{C_0}\right) \times \int_0^t dt' \alpha_m(t'), X_0 = \frac{C}{C_0}$$

Hence for all the *autonomous* evolution PDEs of the class (33) (with $\alpha(z,t) = \alpha(z), \alpha_m(t) = \alpha_m$, see (57)) the center of mass of the *generic* solution moves uniformly,

$$X(t) = X_0 + Vt$$

with the (constant) speed

$$V = \sum_{m=0}^{M} (-1)^{m+1} (2\,m+1) \left(\frac{C_m}{C_0}\right) \alpha_m$$

#### Other techniques to identify, classify and investigate integrable PDEs

The spectral transform approach on which we focussed above is just one of the various techniques used to identify and investigate *integrable* nonlinear evolution PDEs. (Incidentally; because the less standard aspect of this approach is the *inverse* transformation to reconstruct, in the framework of the spectral problem, the "potential" $u(x)$ from its spectral transform, this approach is often called the Inverse Spectral, or Scattering, Transform method – abbreviated as IST). In this subsection we tersely mention some other approaches, referring to the literature indicated below for more adequate treatments.

An approach starts from a trivially integrable PDE – say, linear and autonomous, see for instance (30) – and performs a nonlinear change of dependent, and possibly as well of independent, variables. The PDE thus obtained is generally *integrable*, indeed the term *C-integrable* is used to denote such equations (to distinguish them from the *S-integrable* equations solvable via IST: the letter C refers to the Change of variables, the letter S to the Spectral, or Scattering, transform). A simple instance of *C-integrable* equations is the Burgers equation (5), which is linearized via the change of dependent variable

$$\tilde{q}(x,t) = q(x,t) \exp\left[-\int_{-\infty}^{x} dy q(y,t)\right]$$

$$q(x,t) = \frac{\tilde{q}(x,t)}{1 - \int_{-\infty}^{x} dy\, \tilde{q}(y,t)}$$

entailing the linear PDE

$$\tilde{q}_t + \tilde{q}_{xx} = 0$$

A second example is the "Liouville equation"

$$u_{xt} = \exp(u) \qquad [58a]$$

or equivalently, in "light-cone coordinates" ($\xi = x + t$, $\tau = -x + t$)

$$u_{\tau\tau} - u_{\xi\xi} = \exp(u) \qquad [58b]$$

the general solution of which reads

$$u(x,t) = f(x) - g(t) - 2\log\left\{a \int_{x_0}^{x} dx' \exp[f(x')] + (2a)^{-1} \int_{t_0}^{t} dt' \exp[-g(t')]\right\}$$

with $f(x)$ and $g(t)$ arbitrary functions and $x_0, t_0$, $a$ arbitrary constants. And a third example is the Eckhaus equation

$$q_t = i\left\{q_{xx} + \left[2\left(|q|^2\right)_x + |q|^4\right] q\right\} \qquad [59]$$

which is linearized by the transformation

$$\hat{q}(x,t) = q(x,t) \exp\left[\int_{-\infty}^{x} dy |q(y,t)|^2\right]$$

$$q(x,t) = \frac{\hat{q}(x,t)}{\sqrt{1 + 2\int_{-\infty}^{x} dy |\hat{q}(y,t)|^2}}$$

entailing the linear PDE

$$\hat{q}_t = i\hat{q}_{xx}$$

Thanks to the simplicity of the technique to solve them, *C-integrable* PDEs provide a convenient tool to investigate the phenomenology associated with *nonlinear* PDEs. For instance the Burgers equation (5), which possesses kink-like solitons, is a simple nonlinear generalization of the heat equation; and the "relativistic invariance" of the Liouville equation, see (58b), makes it a convenient "toy model" in the context of relativistic field theory. The Eckhaus equation, (59), provides an interesting theoretical tool because of its similarity with the phenomenologically important NLS equation (6), as well as the fact that, thanks to its *C-integrability*, the structure of its solutions – which feature a remarkable solitonic zoology, including the possibility of "anelastic" solitonic reactions – can be studied in considerable detail, entailing an understanding of why such anelastic reactions are unlikely to be featured by solutions obtained in the context of the initial-value problem.

*C-integrable* PDEs are generally as well *S-integrable*, being generally associable with a spectral problem that can be explicitly solved; the converse, instead, is not generally true. Hence *C-integrability* represents a higher level of integrability than *S-integrability*; a ranking that is quite useful in spite of its lack of strict cogency caused by the possibility to consider also the transformation from a function to its spectral transform as a change of (dependent) variable.

The Lax approach, described in some detail above in the context of finite-dimensional integrable dynamical systems, was in fact originally invented in the context of integrable PDEs. For instance the KdV equation (35) corresponds to the (operator) Lax equation (to be compared with the matrix Lax equation (14))

$$\mathbf{L}_t = [\mathbf{L}, \mathbf{M}]$$

where now the Schrödinger operator $\mathbf{L}$ is defined by (36) (so that $\mathbf{L}_t = u_t(x,t)$) and the operator $\mathbf{M}$ is defined as follows:

$$\mathbf{M} = -4\left(\frac{\partial}{\partial x}\right)^3 + 6u(x,t)\frac{\partial}{\partial x} + 3u_x(x,t)$$

Closely connected with this approach is the AKNS method (due to M. J. Ablowitz, D. J. Kaup, A. C. Newell and H. Segur), based on the observation that the KdV equation (35) coincides with the integrability condition

$$\psi_{xxt} = \psi_{txx} \qquad [60]$$

for the following pair of linear PDEs (the first of which is just the eigenvalue equation for the Schrödinger operator $\mathbf{L}$, see (36)) satisfied by the function $\psi(x,k,t)$ :

$$\psi_{xx} = \left[u(x,t) - k^2\right]\psi \qquad [61a]$$

$$\psi_t = \left[-u_x(x,t) + 4ik^3\right]\psi + 2\left[u(x,t) + 2k^2\right]\psi_x \qquad [61b]$$

and, more generally, that every equation of the class (33) coincides with the integrability condition (60) for the eigenvalue equation (61a) and the equation

$$\psi_t = a(x,k,t)\,\psi + b(x,k,t)\,\psi_x \qquad [61c]$$

with an appropriate choice of the two functions $a(x,k,t)$ and $b(x,k,t)$. Indeed this *ansatz*, (61c), with $a(x,k,t)$ and $b(x,k,t)$ low-order polynomials in $k$, provides a quite straightforward technique to identify the simpler equations of the class (33); ditto for the extension of this approach based on more general eigenvalue problems than (61a).

Another powerful approach suitable to identify and investigate *integrable* PDEs is the so-called "dressing method" (introduced by V. E. Zakharov and A. B. Shabat and pursued by many others), in which one starts again (as in the approach leading to *C-integrable* equations) from an easily solvable evolution equation and then performs transformations (less elementary than just a change of variables) that modify ("dress") the original equation, obtaining thereby new (nontrivial and interesting) evolution equations, the *integrability* of which hinges on the control one has on the (dressing) transformation relating (both ways) the solutions of the new equations with those of the original equation. Of course many specific techniques are accommodated within this (admittedly vague) description; we must confine our remarks here to noting the crucial role that the Riemann-Hilbert problem generally plays in this context (indeed the Riemann-Hilbert problem also lies at the core of the solvability of the inverse spectral problem, although techniques not explicitly relying on it are also available).

*Algorithmic* approaches, particularly suitable to manufacture multisolitonic solutions and to identify nonlinear PDEs that are *integrable* inasmuch as they feature such solutions, were developed already at the beginning of the 70's. The pioneer of this approach was R. Hirota; less than a decade later a more sophisticated and general development – the so-called "tau-function" method – was invented by M. Sato and his pupils/collaborators.

Finally let us mention that many remarkable connections exist among *integrable* PDEs and *integrable* finite-dimensional dynamical systems such as those discussed above; for instance the time-evolution (taking generally place in the complex plane) of the poles of *rational* solutions of certain *integrable* PDEs obey the equations of motion of *integrable* dynamical systems interpretable as many-body problems.

### Why are certain nonlinear PDEs both integrable and widely applicable?

Several integrable PDEs play a key role in various applicative contexts, justifying the question figuring as title of this subsection. A metamathematical but enlightening, and heuristically quite useful, reply to this question reads as follows.

Consider as starting point a *large* class of nonlinear PDEs, and associate to it via some kind of asymptotic limit procedure a *single* nonlinear

PDE – to which it is then justified to attribute a certain *universal* character. If this procedure corresponds to a *physically* (or, more generally, *applicatively*) significant limit, it stands to reason that this universal PDE play a role in *several* applicative contexts (because the original class of PDEs, being *large*, certainly contains *several* equations of applicative relevance). And if the limit procedure is in some sense *asymptotically exact*, and it therefore *preserves* the property of *integrability*, it is also likely that this universal PDE be *integrable*, because for this it is sufficient that the original, *large* class of PDEs contain just one *integrable* PDE.

For instance most phenomena characterized by a dominant dispersive plane wave in a weakly nonlinear context can be shown, via an asymptotically exact multiscale expansion, to be modeled by the Nonlinear Schroedinger equation (6), the solution of which provides then the evolution, in appropriately rescaled "slow" and "coarse-grained" time and space variables, of the amplitude modulation of the dominant dispersive wave. This explains why this nonlinear PDE plays a key role in so many, disparate applicative contexts, and it also implies, in the light of the above argument, its integrability.

The reasoning outlined above is quite robust, and it allows to infer that, if instead the universal limit equation is *not* integrable, then the large class of PDEs from which it originates cannot contain *any* integrable equation, providing thereby the point of departure to obtain (quite useful) *necessary conditions for integrability*. Indeed these conditions are adequate to distinguish among different levels of integrability, for instance among *C-integrability* and *S-integrability*; with the Eckhaus equation (59) playing in this context a somewhat analogous role for *C-integrable* PDEs to that played by the Nonlinear Schrödinger equation (6) for *S-integrable* PDEs.

## Outlook

Many more important developments than could be covered in this overview have occurred in the last few decades; for these we refer to the books listed below (and there are many more), and to the literature cited there.

Let us end this entry by emphasizing that both the study of *integrable systems*, and its application to phenomenologically interesting situation – including technological innovations, for instance in nonlinear optics and telecommunications – are still in the forefront of current research; although perhaps the "heroic era" of this field of study is over.

*See also:* Abelian Higgs Vortices; Bäcklund Transformations; Bethe Ansatz; Bifurcations of Periodic Orbits; Bi-Hamiltonian Methods in Soliton Theory; Billiards in Bounded Convex Domains; Boundary-Value Problems for Integrable Equations; Breaking Water Waves; Calogero–Moser–Sutherland Systems of Nonrelativistic and Relativistic Type; Cauchy Problem for Burgers-type Equations; Cellular Automata; Classical *r*-Matrices, Lie Bialgebras, and Poisson Lie Groups; $\bar{\partial}$-Approach to Integrable Systems; Einstein Equations: Exact Solutions; Functional Equations and Integrable Systems; Ginzburg–Landau Equation; Hamiltonian Systems: Obstructions to Integrability; Holonomic Quantum Fields; Instantons: Topological Aspects; Integrability and Quantum Field theory; Integrable Discrete Systems; Integrable Systems and Algebraic Geometry; Integrable Systems and Discrete Geometry; Integrable Systems and the Inverse Scattering Method; Integrable Systems in Random Matrix Theory; Inverse Problem in Classical Mechanics; Isochronous Systems; Isomonodromic Deformations; Integrable Systems and Recursion Operators on Symplectic and Jacobi Manifolds; Korteweg–de Vries Equation and Other Modulation Equations; Multi-Hamiltonian Systems; Nonlinear Schrödinger Equations; Ordinary Special Functions; Painlevé Equations; Peakons; *q*-Special Functions; Quantum Calogero–Moser Systems; Quantum *n*-Body Problem; Random Matrix Theory in Physics; Recursion Operators in Classical Mechanics; Riemann–Hilbert Methods in Integrable Systems; Riemann–Hilbert Problem; Separation of Variables for Differential Equations; Sine-Gordon Equation; Solitons and Kac–Moody Lie Algebras; Solitons and Other Extended Field Configurations; Twistors; Toda Lattices; Vortex Dynamics; WDVV Equations and Frobenius Manifolds; Yang–Baxter Equations.

## Further Reading


Ablowitz MJ and Clarkson PA (1991) *Solitons, Nonlinear Evolution Equations and Inverse Scattering*. Cambridge: Cambridge University Press.

Ablowitz MJ and Segur H (1981) *Solitons and the Inverse Scattering Transform*. Philadelphia: SIAM.

Babelon O, Bernard D, and Talon M (2003) *Introduction to Classical Integrable Systems*. Cambridge: Cambridge University Press.

Bullough RK and Caudrey PJ (eds.) (1980) *Solitons*. Heisenberg: Springer.

Calogero F (ed.) (1978) *Nonlinear Evolution Equations Solvable by the Spectral Transform*. London: Pitman.

Calogero F (2001) *Classical Many-Body Problems Amenable to Exact Treatments*. Heidelberg: Springer.

Calogero F and Degasperis A (1982) *Spectral Transform and Solitons. I*. Amsterdam: North Holland.

Dodd RK, Eilbeck JC, Gibbon JD, and Morris HC (1982) *Solitons and Non-linear Wave Equations*. New York: Academic Press.

Faddeev LD and Takhtajan LA (1987) *Hamiltonian Methods in the Theory of Solitons*. Heidelberg: Springer.

Hoppe J (1992) *Lectures on Integrable Systems*. Heidelberg: Springer.
Konopelchenko GB (1987) *Nonlinear Integrable Equations*. Heidelberg: Springer.
Novikov SP, Manakov SV, Pitaevskii LP, and Zakharov VE (1984) *Theory of Solitons: the Inverse Scattering Method*. New York: Plenum Press.
Moser J (ed.) (1975) *Dynamical Systems, Theory and Applications*. Heidelberg: Springer.
Moser J (1981) *Integrable Hamiltonian Systems and Spectral Theory*. Pisa: Scuola Normale Superiore.
Perelomov AM (1990) *Integrable Systems of Classical Mechanics and Lie Algebras*. Basel: Birkhauser.
Toda M (1981) *Theory of Nonlinear Lattices*. Heidelberg: Springer.
van Diejen JF and Vinet L (eds.) (2000) *Calogero-Moser-Sutherland Models*. Heidelberg: Springer.
Zakharov VE (ed.) (1991), *What is Integrability?*. Heidelberg: Springer.

# Interacting Particle Systems and Hydrodynamic Equations

**C Landim**, IMPA, Rio de Janeiro, Brazil
and UMR 6085, Université de Rouen, France



## Introduction

We present the theory of hydrodynamic behavior of interacting particle systems in the context of exclusion processes, in which no more than one particle per site is allowed.

Denote by $\mathbb{T}_N = \mathbb{Z}/N\mathbb{Z}$ the discrete torus with $N$ points and let $\mathbb{T}_N^d = (\mathbb{T}_N)^d$. The state space $\mathcal{E}_N = \{0,1\}^{\mathbb{T}_N^d}$ consists of all configurations obtained by distributing particles on the discrete torus $\mathbb{T}_N^d$ respecting the exclusion rule which prevents more than one particle per site. The configurations are denoted by the Greek letter $\eta$ so that $\eta(x)$ is equal to 0 or 1 if site $x \in \mathbb{T}_N^d$ is vacant or occupied for the configuration $\eta$.

Denote by $\{\tau_x : x \in \mathbb{Z}^d\}$ the group of translations in $\mathcal{E}_N$: $(\tau_x\eta)(z) = \eta(x+z)$ for each $x$, $z$ in $\mathbb{Z}^d$. Here and below summations are performed modulo $N$. A function $f : \{0,1\}^{\mathbb{Z}^d} \to \mathbb{R}$ with finite support is called a cylinder function.

Fix a family of non-negative cylinder functions $c_j$, $1 \le j \le d$. Let $c_{x,x+e_j}(\eta) = c_j(\tau_x\eta)$ and consider the Markov process $\{\eta_t : t \ge 0\}$ on $\mathcal{E}_N$ with generator $L_N$ given by

$$(L_N f)(\eta) = \sum_{j=1}^{d} \sum_{x \in \mathbb{T}_N^d} c_{x,x+e_j}(\eta)[f(\sigma^{x,x+e_j}\eta) - f(\eta)] \qquad [1]$$

Here, $\{e_1, \ldots, e_d\}$ stands for the canonical basis of $\mathbb{R}^d$ and $\sigma^{x,y}\eta$ for the configuration obtained from $\eta$ by exchanging the occupation variables $\eta(x)$ and $\eta(y)$:

$$(\sigma^{x,y}\eta)(z) = \begin{cases} \eta(z) & \text{if } z \ne x, y \\ \eta(y) & \text{if } z = x \\ \eta(x) & \text{if } z = y \end{cases} \qquad [2]$$

In this dynamics at each bond $\{x, x+e_j\}$ the occupation variables $\eta(x), \eta(x+e_j)$ are exchanged at rate $c_{x,x+e_j}(\eta)$. This happens simultaneously and independently at each bond.

Notice that the total number of particles is conserved by the dynamics since only exchanges are allowed. Denote by $\Sigma_{N,K}(0 \le K \le |\mathbb{T}_N^d|)$ the hyperplane of all configurations $\eta$ of $\mathcal{E}_N$ with $K$ particles. Assume that the rates $c_j$ are nondegenerate for $\eta_t$ to be an irreducible Markov process on each $\Sigma_{N,K}$.

For $0 \le \alpha \le 1$, denote by $\nu_\alpha^N$ the Bernoulli product measure of parameter $\alpha$ on $\mathcal{E}_N$. Under $\nu_\alpha^N$, the variables $\{\eta(x), x \in \mathbb{T}_N^d\}$ are independent, with marginals given by

$$\nu_\alpha^N\{\eta(x) = 1\} = \alpha = 1 - \nu_\alpha^N\{\eta(x) = 0\}$$

Assume that the measures $\nu_\alpha^N, 0 \le \alpha \le 1$ are stationary for the Markov process $\eta_t$. An elementary computation shows that this is the case if each function $c_j$ does not depend on $\eta(0)$, $\eta(e_j)$, in which case the process is in fact reversible with respect to $\nu_\alpha^N$.

Let $\mathcal{M}_+(\mathbb{T}^d)$ be the space of finite positive measures on the torus $\mathbb{T}^d$ endowed with the weak topology. For each configuration $\eta$, let $\pi^N = \pi^N(\eta, \mathrm{d}u)$ be the positive measure on $\mathbb{T}^d$ obtained by assigning mass $N^{-d}$ to each particle:

$$\pi^N := N^{-d} \sum_{x \in \mathbb{T}_N^d} \eta(x)\delta_{x/N}(\mathrm{d}u) \qquad [3]$$

where $\delta_u$ stands for the Dirac measure on $u$. The measure $\pi^N$ is called the empirical measure associated to the configuration $\eta$. The integral of a continuous function $G : \mathbb{T}^d \to \mathbb{R}$ with respect to $\pi^N$ is denoted by

$$\langle \pi^N, G \rangle = N^{-d} \sum_{x \in \mathbb{T}_N^d} G(x/N)\eta(x)$$

Fix a density profile $\rho_0 : \mathbb{T}^d \to [0,1]$. A sequence of probability measures $\mu^N$ on $\mathcal{E}_N$ is said to be associated to $\rho_0$ if $\pi^N$ converges in probability to $\rho_0(u)\mathrm{d}u$ under $\mu^N$:

$$\lim_{N\to\infty} \mu^N \left\{ \left| \langle \pi^N, G \rangle - \int_{\mathbb{T}^d} G(u)\rho_0(u)\, \mathrm{d}u \right| > \delta \right\} = 0$$

for all continuous functions $G: \mathbb{T}^d \to \mathbb{R}$ and all $\delta > 0$. For a continuous profile $\rho_0$ consider, for instance, the product measure $\nu^N_{\rho_0(\cdot)}$ on $\mathcal{E}_N$ whose marginals are given by

$$\nu^N_{\rho_0(\cdot)}\{\eta(x) = 1\} = \rho_0(x/N)$$

It is easy to check that the sequence of probability measures $\nu^N_{\rho_0(\cdot)}$ is associated to $\rho_0$.

Denote by $W_{x,x+e_j}$ the instantaneous current of particles from $x$ to $x+e_j$. This is the rate at which a particle jumps from $x$ to $x+e_j$ minus the rate at which a particle jumps from $x+e_j$ to $x$:

$$W_{x,x+e_j} = \{\eta(x) - \eta(x+e_j)\}c_{x,x+e_j}(\eta)$$

Suppose that the mean value of the current vanishes under all stationary states $\nu^N_\alpha$. This denotes that the average displacement of each particle vanishes in the mean. In particular, in view of the central limit theorem, to observe an evolution of the density in the macroscopic scale, a diffusive rescaling of time is needed. On the other hand, if there is a net flux of particles, the evolution has to be examined in the Euler scale $tN$.

Denote by $\theta(N)$ the time rescaling: $N^2$ if the mean displacement of particles vanishes and $N$ otherwise. For each probability measure $\mu^N$ on $\mathcal{E}_N$, let $\mathbb{P}_{\mu^N}$ be the probability measure on the path space $D(\mathbb{R}_+, \mathcal{E}_N)$ induced by $\mu^N$ and the Markov process $\eta_t$ speeded up by $\theta(N)$. Expectation with respect to $\mathbb{P}_{\mu^N}$ is denoted by $\mathbb{E}_{\mu^N}$.

Denote by $\pi^N_t(\mathrm{d}u) = \pi^N(\eta_{t\theta(N)}, \mathrm{d}u)$ the empirical measure at time $t$. Fix a density profile $\rho_0: \mathbb{T}^d \to [0,1]$ and a sequence of probability measures $\mu^N$ on $\mathcal{E}_N$ associated to $\rho_0$. The goal of the theory of hydrodynamic limit of interacting particle systems is to show that for each $t > 0, \pi^N_t$ converges, as $N \uparrow \infty$, to a deterministic path $\pi(t, \mathrm{d}u) = \rho(t,u)\mathrm{d}u$ whose density $\rho$ is the solution of some partial differential equation, called the hydrodynamic equation.

The main tools available are entropy production and Dirichlet forms. Denote by $H_N(\mu^N|\nu^N)$ the entropy of a probability measure $\mu^N$ on $\mathcal{E}_N$ with respect to a reference probability measure $\nu^N$:

$$H_N(\mu^N|\nu^N) = \sup_f \left\{ \int_{\mathcal{E}_N} f\,\mathrm{d}\mu^N - \log \int_{\mathcal{E}_N} \mathrm{e}^f\,\mathrm{d}\nu^N \right\}$$

where the supremum is carried over all functions $f: \mathcal{E}_N \to \mathbb{R}$.

It follows from the general theory of Markov processes that the entropy of the state of the process with respect to an invariant state decreases in time. The rate at which the entropy production decreases can be estimated by the Dirichlet form: let $S^N_t$ be the semigroup associated to the generator $L_N$ defined in [1] speeded up by $\theta(N)$. An elementary computation gives that

$$\begin{aligned} &H_N(\mu^N S^N_t|\nu^N_\alpha) + 2\theta(N)\int_0^t \mathrm{d}s\, I^N_\alpha(\mu^N S^N_s) \\ &\quad\le H_N(\mu^N|\nu^N_\alpha) \end{aligned}$$

Here, $I^N_\alpha(\mu^N)$ is the convex and lower semicontinuous functional given by

$$I^N_\alpha(\mu^N) = -\langle f^{1/2}, L_N f^{1/2}\rangle_{\nu^N_\alpha}$$

where $f$ stands for the Radon–Nikodym derivative $\mathrm{d}\mu^N/\mathrm{d}\nu^N_\alpha$ and $\langle\cdot,\cdot\rangle_{\nu^N_\alpha}$ for the scalar product in $L^2(\nu^N_\alpha)$.

Therefore, if the initial state $\mu^N$ has entropy with respect to a reference measure $\nu^N_\alpha$ bounded by $C_0 N^d$, by convexity of $I^N_\alpha$,

$$\begin{aligned} &N^{-d} H_N(\mu^N S^N_t|\nu^N_\alpha) \\ &\quad + 2t\theta(N)N^{-d} I^N_\alpha\left(t^{-1}\int_0^t \mathrm{d}s \mu^N S^N_s\right) \le C_0 \end{aligned} \quad [4]$$

for all $t \ge 0$. This elementary estimate plays a fundamental role in the following sections.

## The Entropy Method

Consider an exclusion process with generator given by [1]. Fix $T > 0$, a density profile $\rho_0: \mathbb{T}^d \to [0,1]$ and a sequence of probability measures $\mu^N$ associated to $\rho_0$. Let $\mathbb{Q}_{\mu^N}$ be the measure on the path space $D([0,T], \mathcal{M}_+(\mathbb{T}^d))$ induced by the process $\pi^N_t$ and the initial state $\mu^N$.

To prove that $\pi^N_t$ converges to $\rho(t,u)\mathrm{d}u$ in probability, we first show that the sequence $\mathbb{Q}_{\mu^N}$ converges to the probability measure $\mathbb{Q}^*$ concentrated on the deterministic trajectory $\rho(t,u)\mathrm{d}u$, whose density is the solution of some partial differential equation with initial condition $\rho_0$. It follows from this result and general arguments that $\pi^N_t$ converges to $\rho(t,u)\mathrm{d}u$ for each $0 \le t \le T$.

To prove that $\mathbb{Q}_{\mu^N}$ converges to $\mathbb{Q}^*$, assume that we are able to prove tightness of the sequence $\mathbb{Q}_{\mu^N}$. Since there is at most one particle per site, all limit points $\mathbb{Q}^*$ of the sequence $\mathbb{Q}_{\mu^N}$ are concentrated on trajectories $\pi(t,\mathrm{d}u) = \rho(t,u)\mathrm{d}u$, which are absolutely continuous with respect to Lebesgue.

To characterize the limit points $\mathbb{Q}^*$, fix a smooth function $G: \mathbb{T}^d \to \mathbb{R}$ and consider the martingale

$$\begin{aligned} M^{G,N}_t &= \langle \pi^N_t, G\rangle - \langle \pi^N_0, G\rangle \\ &\quad - \int_0^t \theta(N) L_N \langle \pi^N_s, G\rangle\,\mathrm{d}s \end{aligned} \quad [5]$$

An elementary computation of its quadratic variation shows that $M_t^{G,N}$ vanishes in $L^2(\mathbb{P}_{\mu^N})$ as $N \uparrow \infty$.

Denote by $\mathcal{C}_0$ the space of cylinder functions which have zero mean with respect to all invariant states $\nu_\alpha^N$. Assume that the currents $W_{0,e_j}, 1 \le j \le d$, belong to $\mathcal{C}_0$ so that a diffusive scaling $\theta(N) = N^2$ is in force. Notice that

$$L_N \eta(x) = \sum_{j=1}^{d} W_{x-e_j,x} - W_{x,x+e_j}$$

In particular, after a summation by parts, the integral term on the right-hand side of [5] can be written as

$$\int_0^t N^{1-d} \sum_{j=1}^{d} \sum_{x \in \mathbb{T}_N^d} (\nabla_{u_j}^N H)(x/N) W_{x,x+e_j}(s)\, \mathrm{d}s \qquad [6]$$

where $(\nabla_{u_j}^N H)(x/N) = N\{H(x+e_j/N) - H(x/N)\}$. Notice that this sum is in principle of order $N$.

To illustrate the entropy method, consider the symmetric simple exclusion process obtained by taking $c_j = 1/2$ in [1] and observe that the current $W_{0,e_j} = (1/2)\{\eta(0) - \eta(e_j)\}$. A second summation by parts permits to rewrite the martingale [5] as

$$\langle \pi_t^N, G \rangle - \langle \pi_0^N, G \rangle - \frac{1}{2} \int_0^t \langle \pi_s^N, \Delta_N G \rangle\, \mathrm{d}s$$

where $\Delta_N$ is the discrete Laplacian.

Since the martingale $M_t^{G,N}$ vanishes in $L^2(\mathbb{P}_{\mu^N})$, as $N \uparrow \infty$, all limit points $\mathbb{Q}^*$ are concentrated on weak solutions of the linear heat equation. It remains to recall that there is a unique weak solution of the Cauchy problem for the heat equation to conclude that the sequence $\mathbb{Q}_{\mu^N}$ converges to $\mathbb{Q}^*$, the measure concentrated on the deterministic path $\pi_t(\mathrm{d}u) = \rho(t,u)\mathrm{d}u$ whose density $\rho$ is the solution of the heat equation with initial condition $\rho_0$.

The symmetric simple exclusion process has the very special property that the martingale $M_t^{G,N}$ can be written as a function of the empirical measure. This is not the case for all the other models, for which a further argument is needed to close eqn [5] in terms of the empirical measure.

To present the additional arguments needed, assume that $c_j(\eta) = 1 + [\eta(-e_j) + \eta(2e_j)]$. In this case, the current $W_{0,e_j}$ is equal to

$$\{\eta(0) - \eta(e_j)\} + \{\eta(0)\eta(-e_j) - \eta(e_j)\eta(2e_j)\} + \{\eta(0)\eta(2e_j) - \eta(-e_j)\eta(e_j)\}$$

A second summation by parts in [6] permits to rewrite it as

$$\int_0^t N^{-d} \sum_{j=1}^{d} \sum_{x \in \mathbb{T}_N^d} (\partial_{u_j}^2 H)(x/N) \tau_x h(\eta_{sN^2})\, \mathrm{d}s + o_N(1) \qquad [7]$$

where $h(\eta) = \eta(0) + 2\eta(0)\eta(-e_j) - \eta(0)\eta(2e_j)$. The remainder $o_N(1)$ appears because we replaced discrete space derivatives by continuous ones.

In contrast with the symmetric simple exclusion process, the martingale $M_t^{G,N}$ defined in [5] is not a function of the empirical measure and an argument is needed to close the equation.

For each positive integer $\ell$ and $d$-dimensional integer $x$, denote by $\eta^\ell(x)$ the empirical density of particles in a box of length $2\ell + 1$ centered at $x$:

$$\eta^\ell(x) = \frac{1}{(2\ell+1)^d} \sum_{|y-x| \le \ell} \eta(y)$$

For a cylinder function $h : \mathcal{E}_N \to \mathbb{R}$, let $\tilde{h}(\alpha)$ be the expected value of $h$ with respect to the invariant state $\nu_\alpha^N : \tilde{h}(\alpha) = E_{\nu_\alpha^N}[h(\eta)]$. For $\ell \ge 1$ and a cylinder function $h$, let

$$V_\ell(\eta) = \left| \frac{1}{(2\ell+1)^d} \sum_{|y| \le \ell} (\tau_y h)(\eta) - \tilde{h}(\eta^\ell(0)) \right|$$

**Theorem 1** *Consider a sequence of probability measures $m^N$ on $\mathcal{E}_N$ such that $I_\alpha^N(m^N) \le C_0 N^{d-2}$ for some $0 < \alpha < 1$ and some finite constant $C_0$. Then,*

$$\limsup_{\varepsilon \to 0} \limsup_{N \to \infty} \mathbb{E}_{\mu^N} \left[ N^{-d} \sum_{x \in \mathbb{T}_N^d} \tau_x V_{\varepsilon N}(\eta) \right] = 0$$

This statement, due to Guo *et al.* (1988), permits the replacement of a local function $h$ by a function of the density of particles over a macroscopic cube. It is the main step in the proof of the hydrodynamic behavior of gradient systems, defined below, and its proof can be found in Kipnis and Landim (1999, chapter 5).

Assume that the sequence $\mu^N$ has entropy with respect to a reference invariant state $\nu_\alpha^N$ bounded by $C_0 N^d$ for some finite constant $C_0$. It follows from [4] that the sequence of measures $T^{-1} \int_0^T \mathrm{d}s \mu^N S_s^N$ satisfies the assumptions of Theorem 1. Therefore, due to the presence of the time integral, we may replace the cylinder function $h$ in [7] by $\tilde{h}(\eta^{\varepsilon N}(x))$. Since $\eta^{\varepsilon N}(0)$ can be written as $\langle \pi^N, \iota_\varepsilon \rangle$, where $\iota_\varepsilon = (2\varepsilon)^{-d} 1\{[-\varepsilon, \varepsilon]^d\}$, we now have expressed the martingale [5] in terms of the empirical measure.

Repeating the arguments presented for the symmetric simple exclusion process, we may conclude that all limit points $\mathbb{Q}^*$ of the sequence $\mathbb{Q}_{\mu^N}$ are concentrated on paths $\pi_t(\mathrm{d}u) = \rho(t,u)\mathrm{d}u$, whose density $\rho$ is a weak solution of the parabolic equation

$$\begin{cases} \partial_t \rho = \Delta(\rho + \rho^2) \\ \rho(0, \cdot) = \rho_0(\cdot) \end{cases}$$

because $\tilde{h}(\alpha) = \alpha + \alpha^2$ for $h(\eta) = \eta(0) + 2\eta(0)\eta(-e_j) - \eta(0)\eta(2e_j)$. It remains to show the uniqueness of weak solutions of this differential equation to conclude.

The second integration by parts in [6] was possible because the currents could be written as the difference of local functions and their translations, a very special property not shared by most interacting particle systems. Processes with this attribute are called gradient systems.

## Nongradient Models

Consider an exclusion process with rates $c_j(\eta) = 1 + \eta(-e_j)$, in which case the current is given by

$$W_{0,e_j} = \{\eta(0) - \eta(e_j)\} + \{\eta(0) - \eta(e_j)\}\eta(-e_j)$$

a cylinder function in $\mathcal{C}_0$.

Fix $T > 0$, a density profile $\rho_0 : \mathbb{T}^d \to [0,1]$ and a sequence of probability measures $\mu^N$ associated to $\rho_0$ and having entropy with respect to a reference invariant state $\nu_\alpha^N$ bounded by $C_0 N^d$ for some finite constant $C_0$. Recall the definition of the sequence of measures $\mathbb{Q}_{\mu^N}$, assumed to be tight.

To characterize the limit points of $\mathbb{Q}_{\mu^N}$, fix a smooth function $G : \mathbb{T}^d \to \mathbb{R}$ and examine the martingale $M_t^{G,N}$ introduced in [5]. After an integration by parts, the integral term of the martingale becomes [6]. While a second integration by parts is possible for the first part of the current $\eta(0) - \eta(e_j)$, the second piece remains

$$\int_0^t N^{1-d} \sum_{j=1}^d \sum_{x \in \mathbb{T}_N^d} (\nabla_{u_j}^N H)(x/N) \tau_x w_j(\eta_{sN^2})\, \mathrm{d}s \qquad [8]$$

where $w_j = \{\eta(0) - \eta(e_j)\}\eta(-e_j)$. Notice the extra factor $N$ multiplying the sum and that $w_j$ belongs to $\mathcal{C}_0$. The next result and Theorem 4 are due to Varadhan (1994).

**Theorem 2** *Consider a sequence of probability measures $m^N$ on $\mathcal{E}_N$ such that $H_N(m^N | \nu_\alpha^N) \le C_0 N^d$ for some $0 < \alpha < 1$ and some finite constant $C_0$. Fix a smooth function $G : \mathbb{T}^d \to \mathbb{R}$ and a cylinder function $\Psi$ in $\mathcal{C}_0$. There exists a seminorm $\|\cdot\|_\alpha$ such that*

$$\limsup_{N\to\infty} \left\{ \mathbb{E}_{m^N}\left[ \left| \int_0^T \mathrm{d}s N^{1-d} \sum_{x\in\mathbb{T}_N^d} G(x/N)\tau_x \Psi(\eta_{sN^2}) \right| \right] \right\}^2 \le C_0 T \|G\|_2^2 \sup_{0\le\alpha\le 1} \|\Psi\|_\alpha^2 \qquad [9]$$

The explicit form of the seminorm $\|\cdot\|_\alpha$ can be found in Kipnis and Landim (1999, chapter 7). The proof of Theorem 2 requires a sharp estimate on the spectral gap of the generator $L_N$. Denote by $\Lambda_\ell$ the cube $\{-\ell, \ldots, \ell\}^d$ and by $L_{\Lambda_\ell}$ the restriction of the generator $L_N$ to the cube $\Lambda_\ell$, obtained by suppressing all jumps from $\Lambda_\ell$ (resp. $\Lambda_\ell^c$) to $\Lambda_\ell^c$ (resp. $\Lambda_\ell$). For $0 \le K \le |\Lambda_\ell|$, let $\nu_{\Lambda_\ell, K}$ be the uniform measure on the configurations of $\{0,1\}^{\Lambda_\ell}$ with $K$ particles. The following estimate is needed in the proof of Theorem 2:

**Theorem 3** *There exists a finite constant $C_0$ such that*

$$\langle f, f\rangle_{\nu_{\Lambda_\ell,K}} \le C_0 \ell^2 \langle f, (-L_{\Lambda_\ell}) f\rangle_{\nu_{\Lambda_\ell,K}}$$

*for all $\ell \ge 1, 0 \le K \le |\Lambda_\ell|$ and zero-mean function $f$ in $L^2(\nu_{\Lambda_\ell,K})$.*

This result is due to Quastel (1992) for symmetric simple exclusion processes. Yau developed a general method to prove sharp estimates for the spectral gap of the generator for conservative dynamics (see Lu and Yau (1993) and Yau (1997)).

Since the parallelogram identity is easy to check, by polarization we can define a semi-inner product $\ll \cdot, \cdot \gg_\alpha$ from the seminorm $\|\cdot\|_\alpha$. Denote by $\mathcal{H}_\alpha$ the Hilbert space induced by $\mathcal{C}_0$ and the semi-inner product $\ll \cdot, \cdot \gg_\alpha$ .

Denote by $L$ the generator [1] extended to $\mathbb{Z}^d$. Notice that $Lf$ belongs to $\mathcal{C}_0$ for any cylinder function $f$, and that the gradients $\eta(e_j) - \eta(0)$, and the currents $w_j, 1 \le j \le d$, also belong to $\mathcal{C}_0$. The next result states that all functions in $\mathcal{H}_\alpha$ can be written as a linear combination of gradients and cylinder functions in the image of the generator.

**Theorem 4** *Denote by $L\mathcal{C}_0$ the space $\{Lg : g \in \mathcal{C}_0\}$. For each $0 \le \alpha \le 1$,*

$$\mathcal{H}_\alpha = \overline{L\mathcal{C}_0} \oplus \{\eta(e_j) - \eta(0) : 1 \le j \le d\}$$

In particular, there exists a matrix $\{D_{i,j}(\alpha) : 1 \le i, j \le d\}$ and a sequence of functions $\{f_{i,k}(\alpha, \cdot) \in \mathcal{C}_0 : k \ge 1\}, 1 \le i \le d$, for which

$$w_i + \sum_{j=1}^d D_{i,j}(\alpha)\{\eta(e_j) - \eta(0)\} - Lf_{i,k}(\alpha, \cdot)$$

vanishes in $\mathcal{H}_\alpha$ as $k \uparrow \infty$. For reversible systems (and more generally for generators satisfying a sector condition), it can be shown that the sequence of local functions $f_{i,k}(\alpha, \eta)$ can be taken independent of $\alpha : f_{i,k}(\alpha, \eta) = f_{i,k}(\eta)$. Moreover, with a little extra effort, one obtains a bound uniform in $\alpha$:

$$\inf_{f\in\mathcal{C}_0} \sup_{0\le\alpha\le 1} \left\| w_i + \sum_{j=1}^d D_{i,j}(\alpha)\{\eta(e_j) - \eta(0)\} - Lf \right\|_\alpha = 0 \qquad [10]$$

This estimate together with some algebraic relations in $\mathcal{H}_\alpha$ give a variational formula for the matrix $D_{i,j}$:

for every vector $v$ in $\mathbb{R}^d$,

$$v \cdot D(\alpha)v = \frac{1}{\alpha(1-\alpha)} \inf_{f \in \mathcal{C}_0} \left\| \sum_{i=1}^{d} v_i w_i - Lf \right\|_{\alpha}^{2} \qquad [11]$$

It can also be shown that the matrix $D$ is continuous and strictly elliptic.

We may now complete the proof of the hydrodynamic behavior. Recall that the main difficulty was to express formula [8] in terms of the empirical measure. Fix $1 \le i \le d$ and consider a sequence of cylinder functions $\{f_{i,k} : k \ge 1\}$ satisfying [10] asymptotically as $k \uparrow \infty$. Adding and subtracting the expression $\sum_{1 \le k \le d} D_{j,k}(\eta^{\varepsilon N}(0))\{\eta^{\varepsilon N}(e_j) - \eta^{\varepsilon N}(0)\} - Lf_{j,k}$, [8] becomes the sum of three terms.

The first one is just the expression which appears inside the expectation in [9] with $G = (\nabla^N_{u_j} H)$ and $\Psi$ given by

$$w_j + \sum_{k=1}^{d} D_{j,k}(\eta^{\varepsilon N}(0))\{\eta^{\varepsilon N}(e_j) - \eta^{\varepsilon N}(0)\} - Lf_{j,k}$$

Since the sequence of measure $\mu^N$ satisfies the assumptions of Theorem 2, a modification of the proof of this theorem, to take into account the dependence of $\Psi$ on $N$ and $\varepsilon$, shows that the limit of the expectation of the absolute value of the first term in the decomposition, as $N \uparrow \infty$ and then $\varepsilon \downarrow 0$, is bounded by

$$C_0\, T \|\partial_{u_j} H\|_2^2 \sup_{0 \le \alpha \le 1} \|\Psi_{j,\alpha}\|_{\alpha}^2$$

where

$$\Psi_{j,\alpha} = w_j + \sum_{k=1}^{d} D_{j,k}(\alpha)\{\eta(e_j) - \eta(0)\} - Lf_{j,k}$$

By [10], the penultimate expression vanishes as $k \uparrow \infty$.

The second term in the decomposition is

$$\int_0^t \mathrm{d}s N^{1-d} \sum_{j,k=1}^{d} \sum_{x \in \mathbb{T}_N^d} (\nabla^N_{u_j} H)(x/N) \tau_x L f_{j,k}(\eta_{sN^2})$$

The presence of the generator $L$ and the diffusive rescaling of time permit to show that the expectation of the absolute value of this expression is of order $N^{-1}$ for each fixed $k$.

Finally, the third term is equal to

$$-\int_0^t \mathrm{d}s N^{1-d} \sum_{j,k=1}^{d} \sum_{x \in \mathbb{T}_N^d} (\nabla^N_{u_j} H)(x/N) D_{j,k} \times \left(\eta^{\varepsilon N}_{sN^2}(x)\right)\left\{\eta^{\varepsilon N}_{sN^2}(x+e_k) - \eta^{\varepsilon N}_{sN^2}(x)\right\}$$

A second integration by parts is now possible and one obtains that the previous expression is equal to

$$\int_0^t \mathrm{d}s N^{-d} \sum_{j,k=1}^{d} \sum_{x \in \mathbb{T}_N^d} \left(\partial^2_{u_j,u_k} H\right)(x/N)\, d_{j,k}\left(\eta^{\varepsilon N}_{sN^2}(x)\right) + o_N(1)$$

where $d'_{j,k} = D_{j,k}$. We have already seen in the derivation of the hydrodynamic equation for gradient systems that this sum can be expressed as a function of the empirical measure. Since all limit points are concentrated on paths $\pi_t(\mathrm{d}u)$ which are absolutely continuous, this integral converges to

$$\sum_{j,k=1}^{d} \int_0^t \mathrm{d}s \int_{\mathbb{T}^d} \mathrm{d}u \left(\partial^2_{u_j,u_k} H\right)(u)\, \mathrm{d}_{j,k}(\rho(s,u))$$

Since the martingale [5] vanishes, all limit points are concentrated on trajectories $\pi_t(\mathrm{d}u) = \rho(t,u)\mathrm{d}u$ which are weak solutions of

$$\partial_t \rho = \sum_{j,k=1}^{d} \partial_{u_j}\left\{\left[\delta_{j,k} + D_{j,k}(\rho)\right]\partial_{u_k}\rho\right\}$$

where $D$ is the strictly elliptic and continuous matrix given by the variational formula [11]. Here, the identity matrix $\delta_{j,k}$ comes from the first piece of the current which permitted a second integration by parts. A uniqueness result of weak solutions of the Cauchy problem with initial condition $\rho_0$ concludes the proof of the hydrodynamic behavior of this nongradient system.

## Hyperbolic Equations

Consider the asymmetric simple exclusion process obtained by setting $c_j(\eta) = \eta(0)[1-\eta(e_j)]$ in formula [1]. Notice that the current $W_{0,e_j} = \eta(0)[1-\eta(e_j)]$ has mean $\alpha(1-\alpha)$ with respect to the invariant state $\nu^N_\alpha$, suggesting the Euler rescaling of time $\theta(N) = N$.

Let $\le$ be the partial order on $\mathcal{E}_N$ defined by $\eta \le \xi$ if $\eta(x) \le \xi(x)$ for every $x$ in $\mathbb{T}_N^d$. The asymmetric exclusion process is attractive: there exists a stochastic evolution on $\mathcal{E}_N \times \mathcal{E}_N$ with the following two properties: (1) it preserves the order, in the sense that $\eta_t \le \xi_t$ for all $t \ge 0$ if $\eta_0 \le \xi_0$ and (2) each coordinate evolves according to the original asymmetric exclusion dynamics. This coupling, which may be constructed by letting particles jump together as much as possible, is the main tool in the derivation of the hydrodynamic equation of asymmetric processes.

Fix a smooth function $G : \mathbb{T}^d \to \mathbb{R}$ and recall definition [5] of the martingale $M_t^{G,N}$. An elementary computation shows that the quadratic variation of this martingale vanishes as $N \uparrow \infty$. On the other

hand, after an integration by parts, the integral term of the martingale becomes

$$\int_0^t N^{-d} \sum_{j=1}^{d} \sum_{x\in\mathbb{T}_N^d} (\nabla_{u_j}^N H)(x/N)\eta_{sN}(x) \times [1-\eta_{sN}(x+e_j)]\,ds$$

Assume that the state of the process at any macroscopic time $s$ is close to a product measure associated to some profile $\rho(s,\cdot)$. Since the martingale vanishes asymptotically, taking expectations in [5], we obtain that the density profile should be a weak solution of the quasilinear hyperbolic equation

$$\partial_t \rho + \sum_{j=1}^{d} \partial_{u_j} F(\rho) = 0 \qquad [12]$$

where $F(a) = a(1-a)$.

It is well known that solutions of this equation may develop shocks even if the initial profile $\rho_0(\cdot)$ is smooth and that there is no uniqueness of weak solutions. Several criteria have been introduced to select the relevant solution among the weak solutions. Kružkov (1970), for instance, in the case where density profile $\rho_0 : \mathbb{T}^d \to \mathbb{R}$ is bounded, proved that there exists a unique measurable function $\rho$ which satisfies the entropy condition

$$\partial_t|\rho - c| + \sum_{i=1}^{d} \partial_{u_i}|F(\rho) - F(c)| \le 0 \qquad [13]$$

in the sense of distributions on $(0,\infty)\times\mathbb{T}^d$, for every $c\in\mathbb{R}$, and which converges to the initial condition in $L^1(\mathbb{T}^d)$ as $t\downarrow 0$: $\lim_{t\to 0}\|\rho_t - \rho_0\|_1 = 0$.

Fix $T>0$ and a density profile $\rho_0 : \mathbb{T}^d \to [0,1]$. To couple the original process with another one starting from a different initial sate, we need to impose the initial distribution to be of product form. Consider, therefore, a sequence of "product" probability measures $\mu^N$ associated to $\rho_0$ and recall the definition of the sequence of measures $\mathbb{Q}_{\mu^N}$ given in the section "The entropy method," assumed to be tight.

We have to prove that all limit points are concentrated on entropy solutions of [12]. Coupling the original process $\eta_t$ with another one, denoted by $\xi_t$, starting from the Bernoulli product measure with density $\alpha$, and examining the time evolution of $\sum_{x\in\mathbb{T}_N^d}|\eta_{tN}(x)-\xi_{tN}(x)|$, we derive an entropy inequality at the microscopic level: let $\bar\mu^N$ be a sequence of probability measures on the product space $\mathcal{E}_N\times\mathcal{E}_N$ whose first coordinate is $\mu^N$. Denote by $\mathbb{P}_{\bar\mu^N}^N$ the measure on the path space $D([0,T],\mathcal{E}_N\times\mathcal{E}_N)$ induced by $\bar\mu^N$ and the coupling informally described at the beginning of this section. Rezakhanlou (1991) proved the following theorem:

**Theorem 5** *For every smooth positive function $H$ with compact support in $(0,\infty)\times\mathbb{T}^d$ and every $\varepsilon>0$,*

$$\lim_{\ell\to\infty}\lim_{N\to\infty}\mathbb{P}_{\bar\mu^N}^N\Bigg[\int_0^\infty dt\, N^{-d}\sum_{x\in\mathbb{T}_N^d}\Big\{\partial_t H(t,x/N)\big|\eta_t^\ell(x)-\xi_t^\ell(x)\big| + \sum_{i=1}^{d}(\partial_{u_i}H)(t,x/N)\big|F(\eta_t^\ell(x))-F(\xi_t^\ell(x))\big|\Big\}\ge -\varepsilon\Bigg] = 1$$

If we now assume that the second coordinate $\xi_t$ is initially distributed according to the stationary state $\nu_\alpha^N$, it is not difficult to replace $\xi_t^\ell$ in the above formula by $\alpha$, obtaining a microscopic version of the entropy inequality.

In the one-dimensional nearest-neighbor case, by coupling arguments, we may replace the average $\eta^\ell(0)$ over a large microscopic box by an average $\eta^{\varepsilon N}(0)$ over a small macroscopic box, deriving the entropy inequality [13]. To conclude the proof it remains to show, by means of coupling argument again, that the density profile at time $t$ converges in $L^1(\mathbb{T}^d)$ to the initial condition as $t\downarrow 0$.

In higher dimensions or in the one-dimensional non-nearest-neighbor case, it has not been proved that replacement of $\eta^\ell(0)$ by $\eta^{\varepsilon N}(0)$ is allowed. One is thus forced to consider measure-valued solutions of eqn [12]. Details can be found in Kipnis and Landim (1999, chapter 8).

## Relative Entropy Method

The relative entropy method, due to Yau (1991), is based on the analysis of the time evolution of the entropy of the state of the process with respect to the product measure associated to the solution of the hydrodynamic equation.

While the entropy method requires uniqueness of weak solutions and proves the existence of weak solutions, the relative entropy method requires the existence of a smooth solution and proves the uniqueness of such smooth solutions.

Consider the exclusion process with rates $c_j(\eta) = 1 + [\eta(-e_j) + \eta(2e_j)]$. We have seen that the hydrodynamic equation of this model is given by the nonlinear parabolic equation

$$\partial_t \rho = \Delta\{\rho + \rho^2\} \qquad [14]$$

Fix a profile $\rho_0 : \mathbb{T}^d \to [0,1]$ bounded away from 0 and 1: $0<\delta\le\rho_0(u)\le 1-\delta$. Let $\rho(t,u)$ be the solution of the hydrodynamic equation [14] with initial condition $\rho_0$ and denote by $\nu_{\rho(t,\cdot)}^N$ the product

measure with slowly varying parameter associated to the profile $\rho(t,\cdot)$:

$$\nu^N_{\rho(t,\cdot)}\{\eta;\eta(x)=1\}=\rho(t,x/N),\quad \text{for } x\in\mathbb{T}^d_N$$

**Theorem 6** *Let* $\{\mu^N : N\geq 1\}$ *be a sequence of probability measures on* $\mathcal{E}_N$ *whose entropy with respect to* $\nu^N_{\rho_0(\cdot)}$ *is of order* $o(N^d)$:

$$H_N\left(\mu^N|\nu^N_{\rho_0(\cdot)}\right)=o(N^d)$$

*Then, the relative entropy of the state of the process at the macroscopic time t with respect to* $\nu^N_{\rho(t,\cdot)}$ *is also of order* $o(N^d)$:

$$H_N\left(\mu^N S^N_t|\nu^N_{\rho(t,\cdot)}\right)=o(N^d)\quad \textit{for every } t\geq 0$$

It is not difficult to deduce from this result a strong version of the hydrodynamic limit behavior of the interacting particle system:

**Corollary 1** *Under the assumptions of the theorem, for every cylinder function* $\Psi$ *and every continuous function* $H:\mathbb{T}^d\to\mathbb{R}$,

$$\lim_{N\to\infty} E_{\mu^N S^N_t}\left[\left|N^{-d}\sum_{x\in\mathbb{T}^d_N} H(x/N)\tau_x\Psi(\eta)\right.\right.$$
$$\left.\left.-\int_{\mathbb{T}^d} H(u)\tilde{\Psi}(\rho(t,u))\,\mathrm{d}u\right|\right]=0$$

The relative entropy method can be extended to nongradient systems and to asymmetric processes, whose macroscopic evolution is described by quasilinear hyperbolic equations, up to the first shock.

The hydrodynamic behavior of an interacting particle system corresponds to a law of large numbers for the empirical measure. The central limit theorem is well understood in equilibrium, but remains to this date an important open question in nonequilibrium. The large deviations for diffusive systems have also been investigated, as well as the hydrodynamic behavior of systems in contact with reservoirs. The Navier–Stokes equations have been derived as a correction of the hydrodynamic equation of asymmetric particle systems. We refer to Kipnis and Landim (1999) for further details.

*See also:* Boltzmann Equation (Classical and Quantum); Bose–Einstein Condensates; Breaking Water Waves; Fourier Law; Interacting Stochastic Particle Systems; Macroscopic Fluctuations and Thermodynamic Functionals; Multi-Scale Approaches.

## Further Reading


De Masi A and Presutti E (1991) *Mathematical Methods for Hydrodynamic Limits*, Lecture Notes in Mathematics, vol. 1501. New York: Springer.

Fritz J (2001) *An Introduction to the Theory of Hydrodynamic Limits,* Lectures in Mathematical Sciences, vol. 18. Tokyo: The University of Tokyo, ISSN 09198140.

Guo MZ, Papanicolaou GC, and Varadhan SRS (1988) Nonlinear diffusion limit for a system with nearest neighbor interactions. *Communications in Mathematical Physics* 118: 31–59.

Jensen L and Yau HT (1999) *Hydrodynamical Scaling Limits of Simple Exclusion Models*, IAS/Park City Mathematical Series, vol. 6, pp. 167–225. Providence, RI: American Mathematical Society.

Kipnis C and Landim C (1999) Scaling Limits of Interacting Particle Systems. *Grundlheren der mathematischen Wissenschaften*, vol. 320. New York: Springer.

Kružkov SN (1970) First order quasilinear equations in several independent variables. *Matematicheskii Sbornik* 10: 217–243.

Landim C (2004) *Hydrodynamic Limits of Interacting Particle Systems*, ICTP Lecture Notes, vol. 17, pp. 57–100. Trieste: Abdus Salam International Centre for Theoretical Physics.

Lu SL and Yau HT (1993) Spectral gap and logarithmic Sobolev inequality for Kawasaki and Glauber dynamics. *Communications in Mathematical Physics* 156: 399–433.

Quastel J (1992) Diffusion of color in the simple exclusion process. *Communications on Pure and Applied Mathematics* XLV: 623–679.

Rezakhanlou F (1991) Hydrodynamic limit for attractive particle systems on $Z^d$. *Communications in Mathematical Physics* 140: 417–448.

Spohn H (1991) *Large Scale Dynamics of Interacting Particles.* Berlin: Springer.

Varadhan SRS (1994) Nonlinear diffusion limit for a system with nearest neighbor interactions II. In: Elworthy KD and Ikeda N (eds.) *Asymptotic Problems in Probability Theory: Stochastic Models and Diffusion on Fractals*, Pitman Research Notes in Mathematics, vol. 283, pp. 75–128. New York: Wiley.

Varadhan SRS (2000) Lectures on hydrodynamic scaling. *Fields Institute Communications* 27: 3–40.

Yau HT (1991) Relative entropy and hydrodynamics of Ginzburg–Landau models. *Letters in Mathematical Physics* 22: 63–80.

Yau HT (1997) Logarithmic Sobolev inequality for generalized simple exclusion processes. *Probability Theory and Related Fields* 109: 507–538.

# Interacting Stochastic Particle Systems

**H Spohn**, Technische Universität München, Garching, Germany



## Introduction

According to the basic principles of mechanics, the motion of atoms and molecules is governed, in the semiclassical approximation, by the deterministic Hamiltonian equations of motion. While all evidence points in this direction, for many problems this Hamiltonian approach is so complicated that it hardly yields any useful results. A simple example are many ($10^9$) polystyrene balls (size $1\,\mu m$) immersed in water. The Hamiltonian description would have to deal with the degrees of freedom of all the fluid molecules and all the polystyrene balls. Clearly, a more useful approach is to collect the incessant bombardment of a polystyrene ball by water molecules into a stochastic force acting on the ball with postulated statistical properties. For example, following Einstein, one could regard successive collisions as independent and occurring after an exponentially distributed waiting time. In addition to such stochastic forces, the polystyrene balls are charged and interact with each other through the screened Coulomb force.

On the one-particle level, stochastic models have a long tradition within statistical physics. Considerable part of the classical theory of Markov processes is the mathematical response to such type of description. The aspect of interaction is more recent. Its origin can be traced back to the Metropolis algorithm in early computer simulations ($\cong$1953). It was recognized that the Hamiltonian dynamics is a rather slow tool to statistically sample the Gibbs equilibrium distribution $Z^{-1} \exp[-H/k_B T]$. A more efficient route is to devise a stochastic algorithm which has as its unique stationary measure the Gibbs distribution. Such schemes are now known as Markov Chain Monte Carlo and of extremely wide use, not only in statistical physics but also in quantum chromodynamics (QCD) and other quantum field theories. The time appearing in the stochastic algorithm has no physical significance; it merely counts how often a certain operation is performed.

The second clearly identifiable push toward the use of interacting stochastic particle systems came from the study of critical dynamics. Close to a point of second-order phase transition, the equilibrium properties are very effectively handled by means of statistical field theories. Thus, it was natural to search for an extension into the time domain, which then led to time-dependent Ginzburg–Landau theories, where now time refers to physical time. These are interacting stochastic models, where one keeps only a few basic fields, together with their behavior under time reversal, their vector character, and whether they are dynamically conserved or not.

In probability theory, interacting stochastic particle systems date back to the seminal papers by M Kac in 1956 and independently by R L Dobrushin and by F Spitzer in 1970. Spitzer was motivated by spin-flip and spin-exchange dynamics, while Dobrushin had the vision of many locally interacting components. In the early days, one of the prime goals was the construction of the stochastic process in infinite volume, an enterprise which had important mathematical spin-off, for example, the theory of Dirichlet forms on function spaces. Physical models offer a rich menu to the probabilist, but there is also considerable input from other areas. To give just one example: in queueing theory one considers queues in series, that is, a customer served at one counter immediately moves on to the next one. If one regards as field the number of customers at each counter, one has an interacting stochastic particle system, the interaction being mediated through the servers.

This article is split into two sections. In the first one, we list and explain a few prototypical interacting stochastic particle systems. Of course, the list is hardly exhaustive and we restrict ourselves from the outset to models from statistical physics. In the second part, we summarize prominent lines of recent research. Again the wealth of material is overwhelming and we draw the line according to the rules of mathematical physics.

## Model Systems

Our list is determined by the intrinsic mathematical properties of the stochastic particle system. Alternatively, a classification is possible according to the physical system, which would, however, be less transparent for our purposes. We restrict ourselves to models with only position-like degrees of freedom, but if needed velocity-like fields may be included. The most basic distinction is the behavior under time reversal. A model is called (statistically) "time reversible" if a particular history and its time-reversed image have the same probability. Technically, one imposes this through the condition of detailed balance. Nonreversible systems are much less explored, but currently a very active area of research.

**Reversible Models**

1. *Spin-flip, Glauber dynamics*. One considers spins attached to the sites of a regular lattice, which for symplicity we take as the hypercubic lattice $\mathbb{Z}^d$. The spin at site $x \in \mathbb{Z}^d$ is denoted by $\sigma_x = \pm 1$ and the whole spin configuration is denoted by $\sigma$. Thus, the state space of the Markov processs is $\{-1, 1\}^{\mathbb{Z}^d} = \Omega$. Spin configurations evolve in time through random spin flips, that is, through a change from $\sigma_x$ to $-\sigma_x$ according to configuration-dependent rates $c_x(\sigma)$. $c_x(\sigma)$ is local, in the sense that it depends only on the spins close to $x$, and is translation invariant, that is, if $\tau_y$ is the shift by $y$, then $c_{x+y}(\tau_y \sigma) = c_x(\sigma)$. If the current spin configuration is $\sigma(t)$, then after a short time $\mathrm{d}t$

$$\sigma_x(t+\mathrm{d}t) = \begin{cases} \sigma_x(t) & \text{with probability } 1 - c_x(\sigma(t))\,\mathrm{d}t \\ -\sigma_x(t) & \text{with probability } c_x(\sigma(t))\,\mathrm{d}t \end{cases}$$

The update is performed independently at each lattice site. Technically, it is more concise to specify the generator, $L$, of the Markov process. It acts on local functions $f : \Omega \to \mathbb{R}$ and is given by

$$Lf(\sigma) = \sum_{x \in \mathbb{Z}^d} c_x(\sigma)(f(\sigma^x) - f(\sigma)) \qquad [1]$$

where $\sigma^x$ denotes the configuration $\sigma$ with the spin at site $x$ reversed. The transition probability from the configuration $\sigma$ to the configuration $\sigma'$ in time $t \geq 0$ is given by the matrix element $(\mathrm{e}^{Lt})_{\sigma,\sigma'}$ of the Markov semigroup $\mathrm{e}^{Lt}$.

To impose time reversibility, one needs an energy function $H(\sigma)$ constructed according to the rules of equilibrium statistical mechanics. The condition of detailed balance then reads

$$c_x(\sigma) = c_x(\sigma^x)\mathrm{e}^{-\beta(H(\sigma^x) - H(\sigma))} \qquad [2]$$

with $\beta = 1/k_{\mathrm{B}}T$ the inverse temperature. Note that on the right only energy differences appear, which are always well defined. In finite volume the unique invariant measure is the Gibbs measure $Z^{-1}\mathrm{e}^{-\beta H}$.

2. *Spin-exchange, Kawasaki dynamics, stochastic lattice gases*. We model particles hopping on the lattice $\mathbb{Z}^d$ and switch to the occupation variables $\eta_x$, where $\eta_x = 0$ stands for site $x$ empty and $\eta_x = 1$ stands for site $x$ occupied. The state space is $\Omega = \{0, 1\}^{\mathbb{Z}^d}$. Since the number of particles is conserved, the basic dynamical process is a random jump of a particle from $x$ to a nearby site $y$, provided $\eta_y = 0$. Therefore, we specify the exchange rates $c_{xy}(\eta)$ between $x$ and $y$. They are local, translation invariant and symmetric, that is, $c_{xy}(\eta) = c_{yx}(\eta)$. The generator now reads

$$Lf(\eta) = \frac{1}{2}\sum_{x,y \in \mathbb{Z}^d} c_{xy}(\eta)(f(\eta^{xy}) - f(\eta)) \qquad [3]$$

where $\eta^{xy}$ is the configuration $\eta$ with the occupancies at sites $x$ and $y$ exchanged.

The condition of detailed balance refers to the exchange and reads

$$c_{xy}(\eta) = c_{xy}(\eta^{xy})\mathrm{e}^{-\beta(H(\eta^{xy}) - H(\eta))} \qquad [4]$$

In [4] we can freely add to $H$ the chemical potential $-\mu \sum_x \eta_x$. Thus for stochastic lattice gases there is a one-parameter family of invariant measures, labeled by the chemical potential $\mu$.

3. *Interacting Brownian motions*. These motions model, for example, suspensions as mentioned in the "Introduction". One considers a box $\Lambda \subset \mathbb{R}^d$ containing $N$ Brownian particles. The $j$th Brownian particle has position $x_j \in \Lambda$. Thus, the state space of the Markov process is $\Lambda^N$. We assume that the Brownian particles interact through a (sufficiently local) even pair potential $U$. Then the total potential energy is

$$H(\underline{x}) = \frac{1}{2}\sum_{i,j=1}^{N} U(x_i - x_j), \quad \underline{x} = (x_1, \ldots, x_N) \qquad [5]$$

The dynamics of the Brownian particles is given through the stochastic differential equations

$$\begin{aligned} \mathrm{d}x_j(t) = &- \sum_{i=1, i\neq j}^{N} \nabla U(x_j(t) - x_i(t))\ \mathrm{d}t \\ &+ \sqrt{2D_0}\mathrm{d}W_j(t), \quad j = 1, \ldots, N \end{aligned} \qquad [6]$$

$W_j(t), j = 1, \ldots, N$, are a collection of independent Brownian motions and $D_0$ is the diffusion coefficient of a single Brownian particle. Equation [6] has to be supplemented with suitable boundary conditions at the surface $\partial\Lambda$. Since the forces in [6] are the gradient of a potential, time reversibility is automatically satisfied with the invariant measure being $Z_N^{-1} \exp(-H(\underline{x})/D_0)\,\mathrm{d}x_1 \cdots \mathrm{d}x_N$.

4. *Ginzburg–Landau models*. Ginzburg–Landau models should be viewed as discretized versions of stochastic partial differential equations. At every lattice site $x \in \mathbb{Z}^d$, there is a real-valued field $\phi_x \in \mathbb{R}$, a field configuration being denoted by $\phi$. Formally, the state space is $\mathbb{R}^{\mathbb{Z}^d}$. Since the single-site space is noncompact, some growth condition at

infinity must be imposed. Next we give ourselves an energy, $H(\phi)$, one standard example being

$$H(\phi) = \sum_{x,y\in\mathbb{Z}^d,|x-y|=1} (\phi_x - \phi_y)^2 + \sum_{x\in\mathbb{Z}^d} V(\phi_x) \qquad [7]$$

The on-site potential increases sufficiently rapidly, so as to make large field values unlikely. The $\phi$-field evolves according to the set of stochastic differential equations

$$\mathrm{d}\phi_x(t) = -\frac{\partial H}{\partial \phi_x}(\phi(t))\mathrm{d}t + \sqrt{2/\beta}\,\mathrm{d}W_x(t), \quad x \in \mathbb{Z}^d \qquad [8]$$

where $\{W_x(t), x \in \mathbb{Z}^d\}$ is a collection of independent Brownian motions. If $V(\phi_x) = \phi_x^2$, then $\phi(t)$ is a Gaussian field theory. To have an Ising-type phase transition, one would have to choose $V(\phi_x) = \lambda\phi_x^2 + \phi_x^4$.

It is rather simple to modify [8] as to incorporate a conservation law. To each directed bond $(x, y)$, $|x - y| = 1$, one associates the current $j_{xy} = -j_{yx}$. If $e$ is a unit vector, $|e| = 1$, then

$$\mathrm{d}\phi_x(t) + \sum_{e,|e|=1} j_{xx+e}(t)\mathrm{d}t = 0, \quad x \in \mathbb{Z}^d \qquad [9]$$

The current has both a deterministic part, given through the gradient of a chemical potential, and a random part:

$$j_{xy}(t)\mathrm{d}t = -\left(\frac{\partial H}{\partial \phi_x} - \frac{\partial H}{\partial \phi_y}\right)(\phi(t))\mathrm{d}t + \mathrm{d}W_{xy}(t), \quad |x - y| = 1 \qquad [10]$$

where $W_{xy}(t) = -W_{yx}(t)$ is a collection of independent Brownian motions labeled by nearest-neighbor bonds. The conserved quantity is $\sum_x \phi_x$. Again, the dynamics has a one-parameter family of stationary measures labeled by the "magnetic field". Since in [8] and [10] the drift is the gradient of a potential, Ginzburg–Landau models are reversible.

5. *Interface dynamics*. The scalar field $\phi$ describes the location of an interface. The energy of an interface does not depend on its absolute displacements. Thus, interface models are special Ginzburg–Landau models, which have an energy $H(\phi)$ invariant under the global shift $\phi_x \to \phi_x + a$ for all $x \in \mathbb{Z}^d$. An example is

$$H(\phi) = \sum_{x,y\in\mathbb{Z}^d,|x-y|=1} V(\phi_x - \phi_y) \qquad [11]$$

with even $V$. Note that in order to have a normalizable equilibrium measure, the interface must be pinned somewhere.

6. *Several components*. For lattice gases, there may be several components. In a Ginzburg–Landau theory instead of a scalar, Ising-like field, one could consider a vector-valued, Heisenberg-like, field and require the energy to be invariant under global rotations of the field variables. The construction is as before and we do not have to repeat it.

7. *Constrained*, *glassy dynamics*. The constraint is enforced by setting some of the rates equal to zero. For example, in the case of standard Glauber dynamics, one could allow for a spin-flip only if at least two neighboring spins have the opposite sign. The Gibbs measure is still invariant, but the approach to equilibrium will be slowed down due to the constraint. It may even happen that the configuration space splits into several invariant subsets.

After this long and still incomplete list, let us turn to the nonreversible models.

### Nonreversible Models

Mathematically, one merely has to drop the condition of detailed balance. To have a more concrete example, let $L_i$ be the generator for the Glauber dynamics satisfying detailed balance with inverse temperature $\beta_i, i = 1, 2$. Then $L = L_1 + L_2$ generates a nonreversible dynamics provided $\beta_1 \neq \beta_2$. Physically, it corresponds to coupling the spins to two bulk thermal reservoirs of different temperatures. Our example leads to a general point which should be noted: While reversible models have a wide range of physical applicability, for nonreversible models nonequilibrium conditions have to be maintained over sufficiently long time spans, which poses considerable difficulties experimentally. Thus on a theoretical level, the efforts go into exploring properties of, say, semirealistic models.

Very roughly there are two broad classes of nonreversible models.

**Boundary-driven models** We consider a finite volume $\Lambda$. Inside $\Lambda$ the dynamics is reversible as explained before. At the boundary $\partial\Lambda$ the system is coupled to particle, resp. energy, reservoirs. In case the boundary chemical potential, resp. temperature, is not uniform, the dynamics is nonreversible. To be more concrete let us reconsider the lattice gas discussed in item (2) (see the discussion following eqn [2]). Inside $\Lambda$ the generator $L_\Lambda$ is given by [3] and satisfies detailed balance [4]. The boundary generator is

$$L_{\partial\Lambda} f(\eta) = \sum_{x\in\partial\Lambda} c_x(\eta)(f(\eta^x) - f(\eta)) \qquad [12]$$

where the notation is as in [1] with $\{-1, 1\}$ substituted by $\{0, 1\}$. $c_x(\eta)$ satisfies [2] with the same $\beta$ as in the bulk, but a chemical potential $\mu_x$ depending on $x \in \partial\Lambda$. $\mu_x$ controls the injection/

absorption of particles at $x$. The generator for the nonreversible dynamics is then

$$L = L_\Lambda + L_{\partial\Lambda} \qquad [13]$$

**Bulk-driven models** A prototype is the two-temperature model mentioned above. More widely studied is a nonconservative force acting globally. Here the standard example are particles moving in $\Lambda$ with periodic boundary conditions and subject to an additional uniform force field of strength $F$, which clearly cannot be written as the gradient of a potential. In the case of Brownian particles, by changing to a comoving frame of reference, one would be back to the reversible case $F=0$. For lattice gases the lattice provides a fixed frame and the driven model has properties very different from the undriven one. This leads us to:

8. *Driven lattice gases*. The generator $L$ is still given by [3]. Formally, we insert in [4] instead of $H$ the Hamiltonian $H(\eta) - \sum_x (F \cdot x)\eta_x$. The exchange rates then satisfy the condition of "local" detailed balance as

$$c_{xy}(\eta) = c_{xy}(\eta^{xy})\, e^{-\beta(H(\eta^{xy})-H(\eta))} \times e^{-\beta(F\cdot(x-y))(\eta_x-\eta_y)} \qquad [14]$$

This means, particles preferentially jump in the direction of $F$. On the infinite lattice the dynamics admits two classes of stationary measures. First, there is the Gibbs measure with particles piling up along $F$ and formally given by

$$Z^{-1}\, e^{-\beta(H(\eta)-\sum_x (F\cdot x)\eta_x)} \qquad [15]$$

With respect to this measure the dynamics is reversible. Second, there are translation invariant measures with nonzero steady-state current. This cannot happen for reversible models. A very widely studied particular case is the asymmetric simple exclusion process for which $d=1, H(\eta)=0$, and jumps are only to nearest-neighbor sites.

## Items of Interest

As there are thousands of research papers in mathematical physics alone, it is literally impossible to provide any sort of summary. On the other hand, the type of questions investigated are generic. Thus, we just explain what one would like to understand without paying much attention to the fractal boundary between "proven" and "unproven." For the construction of the stochastic processes listed above, there is a well-developed probabilistic theory available. Thus, the main focus is on "qualitative properties" of the stochastic particle system. As in the previous section, we distinguish between reversible and nonreversible models.

### Reversible Models

1. *Equilibrium state*. The most basic question concerns the classification of invariant measures in infinite volume. By construction, they are the Gibbs measures for the Hamiltonian appearing in the condition of detailed balance. In principle there could be more, which so far has been excluded only in dimension 1 or 2. Properties of the invariant measure belong to the domain of equilibrium statistical mechanics.

Thus we can turn directly to:

2. *Spectral analysis of the generator L*. We fix some extreme Gibbs measure stationary for $L$. By detailed balance, $e^{Lt}$ is a symmetric Markov semigroup in $L^2(\Omega, \mu)$. Hence, $L$ is self-adjoint and $L \leq 0$. Furthermore, it has a nondegenerate eigenvalue 0. The rate of approach to equilibrium is determined by the spectral gap of $L$. Related are log-Sobolev inequalities which serve as a stronger notion. For models with a conservation law, there is no spectral gap. Thus, the more appropriate question is to study how fast the gap vanishes as the volume $\Lambda$ increases. In the case of independent components, the spectral subspaces for $L$ are organized as single excitation, double excitation etc. Such a structure persists as the interaction is turned on which, on a mathematical level, is similar to the particle spectrum of a quantum field theory.

Physically more directly relevant are:

3. *Spacetime correlations*. To be concrete, let us consider a Ginzburg–Landau field theory $\phi_x(t)$ starting with a translation invariant Gibbs measure $\mu$. Then $\phi_x(t)$ is a spacetime stationary process. The two-point correlation function is the covariance

$$\langle \phi_x(t)\phi_0(0)\rangle - \langle \phi_0(0)\rangle^2 \qquad [16]$$

Its Fourier transform is directly linked to energy–momentum resolved scattering intensity from a probe which is modeled by the respective Ginzburg–Landau theory. For $t=0$, the expression [16] is the static correlation, again belonging to the domain of equilibrium statistical mechanics. The time decay depends on whether the field is dynamically conserved or not.

Correlation functions do not always capture the physics of the system well. This is certainly true for:

4. *Dynamics at low temperatures*. Let us consider the Glauber dynamics for the ferromagnetic Ising model in the finite but large volume $\Lambda$. Then there is a very high free energy barrier between configurations typical for the + phase and those typical for the − phase. If one starts the spin system in the + phase, one

may study through which configurations the system moves to the $-$ phase and how much time such a process will take. If the two phases are symmetric with the external magnetic field $h = 0$, the spin system tunnels, while for $h < 0$ and small the $+$ phase is metastable. Another widely studied situation, also experimentally, is the quenching from high to low temperatures. In our context this means that the initial measure is Bernoulli, while the Glauber dynamics runs at low temperatures. Then spin clusters coarsen as time proceeds developing well-defined interfaces which are governed through motion by mean curvature.

Close to a point of second-order phase transition, one has to deal with:

5. *Critical dynamics*. The usual Glauber dynamics becomes very slow at the critical point and reliable equilibrium is hard to achieve. It is thus a challenge to design faster algorithms. One proposal is the Swendsen–Wang algorithm which is based on the Fortuin–Kasteleyn representation and flips a whole cluster of spins simultaneously.

So far we concentrated on statistical properties. Researchers have been fascinated by the observation that for stochastic particle systems, the transition to a deterministic macroscopic evolution can be handled with full rigor. Such a program has been baptized:

6. *Hydrodynamic limit*, which is meaningful only for particle systems with one or several conservation laws. Let us discuss then a reversible lattice gas with Hamiltonian $H$. We start the dynamics with a state of local equlibrium which is Gibbs with a slowly varying chemical potential, that is,

$$Z^{-1} \exp\left[-\beta\left(H(\eta) - \sum_x \mu(\varepsilon x)\eta_x\right)\right], \quad \varepsilon \ll 1 \quad [17]$$

Such a measure is almost time invariant. For small $\varepsilon$, at least approximately, such a structure should persist in the course of time at the expense of properly regulating the chemical potential. For our example, the correct timescale is $\varepsilon^{-2}t$ in microscopic units, and the evolution equation for the density, related thermodynamically to the chemical potential, is a nonlinear diffusion equation of the form

$$\frac{\partial}{\partial t}\rho_t = \nabla \cdot D(\rho_t)\nabla\rho_t \quad [18]$$

We turn to the nonreversible models.

### Nonreversible Models

While for reversible models the study of the stationary Gibbs measure is its own field of inquiry, here the first entry must be:

7. *Nonequilibrium steady state*. This steady state is determined through the dynamics, since the stationary measure $\mu$ has to satisfy $\mu(Lf) = 0$ for a sufficiently large class of functions $f$. As in equilibrium, phase transitions may occur. In the nonconservative case it would mean that the infinitely extended system has several extreme stationary measures. In the conservative case, say with the density as locally conserved field, it would mean that there is an interval of densities for which there is no extreme stationary measure. Given the nonequilibrium steady state, one may wonder about its typical fluctuations and large deviations. In contrast to thermal equilibrium, weak long-range correlations are the rule.

8. *Spacetime correlations in the steady state*. Through the bulk drive the power-law decay of time correlations may change. For example for the symmetric and asymmetric exclusion process, the steady states are Bernoulli with density $\rho$, denoted by $\langle\cdot\rangle_\rho$. For the on-site density–density correlation, one finds, for large $t$,

$$\langle\eta_0(t)\eta_0(0)\rangle_{1/2} - \frac{1}{4} \cong \begin{cases} t^{-1/2} & \text{for } F = 0 \\ t^{-2/3} & \text{for } F \neq 0 \end{cases} \quad [19]$$

9. *Hydrodynamic limit*. The concept of slowly varying conserved fields remains valid; only local equilibrium must be replaced by local stationarity. Generically, there are nonzero currents in the steady state. Therefore, the macroscopic fields change on the timescale $\varepsilon^{-1}t$ (cf. item (5)) and are governed by a hyperbolic conservation law of the form

$$\frac{\partial}{\partial t}\rho_t + \operatorname{div} j(\rho_t) = 0 \quad [20]$$

in the case of a single conservation law. Here, $j(\rho)$ is the average steady state in the stationary measure at density $\rho$. Several conservation laws have an intriguing rich variety of solutions. Even on the level of continuum partial differential equations, such systems of hyperbolic conservation laws still pose unresolved basic problems.

*See also:* Ginzburg–Landau Equation; Glassy Disordered Systems: Dynamical Evolution; Interacting Particle Systems and Hydrodynamic Equations; Macroscopic Fluctuations and Thermodynamic Functionals; Stochastic Differential Equations.

## Further Reading

Binder K and Heermann D (2002) *Monte Carlo Simulations in Statistical Physics*. Berlin: Springer.

Kac M (1959) *Probability and Related Topics in the Physical Sciences*. London: Interscience.

Kipnis C and Landim C (1999) Scaling Limits of Interacting Particle Systems. *Grundlehren*, vol. 320. Berlin: Springer.

Liggett TM (1985) *Interacting Particle Systems*. Berlin: Springer.
Liggett TM (1999) Stochastic Interacting Systems: Contact, Voter and Exclusion Processes. *Grundlehren*, vol. 324. Berlin: Springer.
Marro J and Dickmann R (1999) *Nonequilibrium Phase Transitions in Lattice Models*. Cambridge: Cambridge University Press.
Martinelli F (1999) Lecture on Glauber Dynamics for Discrete Spin Models. *Lecture Notes in Mathematics*, vol. 1717. Berlin: Springer.
Schmittmann B and Zia RKP (1995) In: Domb C and Lebowitz JL (eds.) Statistical Mechanics of Driven Diffusive Systems, *Phase Transitions and Critical Phenomena*, vol. 17, London: Academic Press.
Spitzer F (1970) Interaction of Markov processes. *Advances in Mathematics* 5: 246–290.
Spohn H (1991) Large Scale Dynamics of Interacting Particles. *Texts and Monographs in Physics*. Heidelberg: Springer.

# Interfaces and Multicomponent Fluids

**J Kim and J Lowengrub**, University of California at Irvine, Irvine, USA



## Introduction

Many important industrial problems involve flows with multiple constitutive components. Examples include extractors, separators, reactors, sprays, polymer blends, and microfluidic applications such as DNA analysis, and protein crystallization. Due to inherent nonlinearities, topological changes, and the complexity of dealing with unknown, active, and moving surfaces, multiphase flows are challenging. Much effort has been put into studying such flows through analysis, asymptotics, and numerical simulation. Here, we focus on review on studies of multicomponent fluids using continuum numerical methods.

There are many ways to characterize moving interfaces. The two main approaches to simulating multiphase and multicomponent flows are interface tracking and interface capturing. In interface-tracking methods (examples include boundary-integral, volume-of-fluid, front-tracking, immersed-boundary, and immersed-interface methods), Lagrangian (or semi-Lagrangian) particles are used to track the interfaces. In (BIMs), the flow equations are mapped from the immiscible fluid domains to the sharp interfaces separating them thus reducing the dimensionality of the problem (the computational mesh discretizes only the interface). In interface-capturing methods such as level-set and phase-field methods, the interface is implicitly captured by a contour of a particular scalar function.

The equations governing the motion of an unsteady, viscous, incompressible, immiscible two-fluid system are the Navier–Stokes equations (the subscript $i$ denotes the $i$th flow component):

$$\rho_i\left(\frac{\partial \boldsymbol{u}_i}{\partial t} + \boldsymbol{u}_i \cdot \nabla \boldsymbol{u}_i\right) = \nabla \cdot \sigma_i + \rho_i \boldsymbol{g}, \quad i = 1, 2 \qquad [1]$$

$$\sigma_i = -p_i \boldsymbol{I} + 2\eta_i \boldsymbol{D}_i \qquad [2]$$

where $\rho_i$ is the density, $\boldsymbol{u}_i$ is the fluid velocity, $p_i$ is the pressure, $\eta_i$ is the viscosity, and $\boldsymbol{g}$ is the gravitational acceleration vector. In eqn [2], $\sigma_i$ is the stress tensor, $\boldsymbol{I}$ is the identity matrix, and $\boldsymbol{D}_i$ is the rate of deformation tensor and defined as $\boldsymbol{D}_i = (1/2)(\nabla \boldsymbol{u}_i + \nabla \boldsymbol{u}_i^{\mathrm{T}})$. The velocity field is subject to the incompressibility constraint,

$$\nabla \cdot \boldsymbol{u}_i = 0 \qquad [3]$$

We let $\Gamma$ denote the fluid interface. The effect of surface tension is to balance the jump of the normal stress along the fluid interface. This gives rise to a Laplace–Young condition for the discontinuity of the normal stress across $\Gamma$:

$$[\sigma \boldsymbol{n}]_\Gamma = \tau \kappa \boldsymbol{n} \qquad [4]$$

where $[\sigma]_\Gamma$ denotes the jump $\sigma_2 - \sigma_1$ across $\Gamma, \kappa$ is the curvature of $\Gamma$ (positive for a spherical interface), $\tau$ is the surface tension coefficient which is assumed to be constant, and $\boldsymbol{n}$ is the unit normal vector along $\Gamma$ directed toward fluid 2. The fluid velocity is continuous across $\Gamma$.

In order to circumvent the problems associated with implementing the Laplace–Young calculation at the exact interface boundary, Brackbill and collaborators developed a method referred to as the continuum surface force (CSF) method. See the review by Scardovelli and Zaleski (1999). In this method, the surface tension jump condition is converted into an equivalent singular volume force that is added to the Navier–Stokes equations. Typically, the singular force is smoothed and acts only in a finite transition region across the interface. The system of equations [1]–[2] and the boundary condition, eqn [4] can be combined into the following distribution formulation that holds in both phases:

$$\rho(\boldsymbol{u}_t + \boldsymbol{u} \cdot \nabla \boldsymbol{u}) = -\nabla p + \nabla \cdot (2\eta \boldsymbol{D}) + \rho \boldsymbol{g} + \boldsymbol{F}_{\text{sing}}, \quad \nabla \cdot \boldsymbol{u} = 0 \qquad [5]$$

where the subscript $i$ is dropped (i.e., it is understood that $\boldsymbol{u} = \boldsymbol{u}_i$ in fluid $i$, etc.,) and $\boldsymbol{F}_{\text{sing}}$ is singular

surface tension force that is given by $F_{sing} = -\tau\kappa\delta_\Gamma n$, where $\delta_\Gamma$ is the surface delta-function.

## Numerical Methods for Multicomponent Fluid Flows

### Interface-Tracking Methods

**Boundary-integral methods (BIMs)** BIMs can be highly accurate for modeling free surface flows with relatively regular interface topologies. The BIM was apparently first used by Rosenhead in 1932 to study vortex sheet roll-up. In this approach, the interface is explicitly tracked, but the flow solution in the entire domain is deduced solely from information possessed by discrete points along the interface.

BIMs have been used for both inviscid and Stokes flows. For a review of Stokes flow computations, see Pozrikidis (2001), and for a review of computations of inviscid flows, see Hou *et al.* (2001). For flows with both inertia and viscosity, volume integrals must be incorporated into the formulation.

When inertial forces are negligible (left-hand side term of eqn [1] is dropped), the velocity $u(x_0)$ at a given point $x_0$ on the interface can be obtained by means of the boundary-integral formulation,

$$(\lambda+1)u(x_0) = 2u_\infty(x_0) - \frac{1}{4\pi}\int_\Gamma f(x)G(x_0,x) \times n(x)\,ds(x) \qquad [6]$$

$$-\frac{\lambda-1}{4\pi}\int_\Gamma u(x)\cdot T(x_0,x)\cdot n(x)\,ds(x) \qquad [7]$$

where $\lambda$ is the viscosity ratio, $u_\infty$ is an imposed velocity prevailing in the absence of the interfaces, and $f(x)$ is the capillary force function $f=\tau\kappa$. The tensors $G$ and $T$ are the Stokeslet and stresslet, respectively:

$$G(x_0,x) = \frac{I}{r} + \frac{\hat{x}\hat{x}}{r^3}, \qquad T(x_0,x) = -\frac{6\hat{x}\hat{x}\hat{x}}{r^5} \qquad [8]$$

$$\text{where } \hat{x} = x - x_0, \quad r = |\hat{x}| \qquad [9]$$

The boundary conditions at the interface, that is, the stress balance equation [4] and continuity of the velocity across the interface, are automatically satisfied by the boundary-integral formulation.

The normal velocity of the interface $\Gamma(x,t)$ is given by

$$\frac{dx}{dt}\cdot n(x) = u(x,t)\cdot n(x) \qquad [10]$$

The shape of the interface does not depend on the tangential velocity and there are many possible choices that can be taken, see Hou *et al.* (2001).

The principal advantages gained by using BIMs are the reduction of the flow problem by one dimension since the formulation involves quantities defined on the interface only and the potential for highly accurate solutions if the flow has topologically regular interfaces. In addition, highly efficient adaptive surface mesh refinement algorithms have recently been developed to improve the performance and accuracy of the methods (Cristini *et al.* 2001). The main disadvantages are the development of accurate quadratures of integrals with singular kernels (particularly in 3D) and the need for local surgery of the interface in the event of topological changes.

BIMs have been successfully used for simulations of complex multiphase flows: drop deformation and breakup; jets; capillary waves; mixing; drop-to-drop interaction; suspension of liquid drops in viscous flow (e.g., see Cristini *et al.* (2001), Hou *et al.* (2001), and Pozrikidis (2001) and the references therein).

**Volume-of-fluid (VOF) method** In the VOF method (see Scardovelli and Zaleski (1999) for a recent review), the location of the interface is determined by the volume fraction $c_{ij}$ of fluid 1 in the computational cell, $\Omega_{ij}$. In cells containing the interface $0 < c_{ij} < 1$, $c_{ij} = 1$ in cells containing fluid 1, and $c_{ij} = 0$ in cells containing fluid 2 as shown in Figure 1b.

A VOF algorithm is divided into two parts: a reconstruction step and a propagation step. A typical interface reconstruction is shown in Figure 1c. In the piecewise linear interface construction (PLIC) method, the true interface, as shown in Figure 1a, is approximated by a straight line perpendicular to an interface normal vector $n_{ij}$ in each cell $\Omega_{ij}$. The normal vector $n_{ij}$ is determined from the volume fraction gradient using data from neighboring cells. With given a volume fraction $c_{ij}$

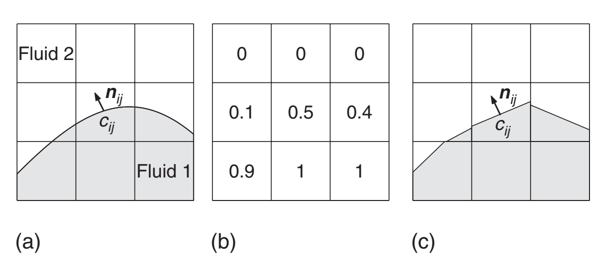



**Figure 1** VOF representation of an interface: (a) actual interface, (b) volume fraction, and (c) an approximation to the interface is produced using an interface reconstruction method such as piecewise linear approximation as shown.

and a normal vector $\boldsymbol{n}_{ij}$, the interface is given by the straight line with normal $\boldsymbol{n}_{ij}$ such that area beneath the line in cell $\Omega_{ij}$ is equal to $c_{ij}$. More recently, parabolic reconstructions of the interface have been used to gain higher-order accuracy for the surface tension force (e.g., the "parabolic reconstruction of surface tension" or PROST algorithm).

Once the interface has been reconstructed, its motion by the underlying flow field must be modeled by a suitable advection algorithm. The key here is that the explicit interface reconstruction enables fluxes to be developed that exactly conserve mass and do not diffuse the interface.

Capillary effects may be represented by the continuous surface stress (Scardovelli and Zaleski 1999),

$$\boldsymbol{T} = -\tau(\boldsymbol{I} - \boldsymbol{n} \otimes \boldsymbol{n})|\nabla\tilde{c}|, \quad \boldsymbol{F}_{\text{sing}} = -\nabla \cdot \boldsymbol{T} \qquad [11]$$

where $\tilde{c}$ is a smoothed version of the volume fraction. For the flows in which the capillary force is the dominant physical mechanism, the PROST algorithm discussed above can be used to significantly reduce spurious currents due to inaccurate representation of surface tension terms and associated pressure jump in normal stress.

The distribution form of the fluid equations [5] is typically solved using a variant of the projection method for incompressible single phase flows.

VOF methods are popular and have been used in commercial multiphase flow codes, in models of inkjet printers, flows with surfactants and in many other applications (e.g., see Scardovelli and Zaleski (1999) and James and Lowengrub (2004) and the references therein). The principal advantage of VOF methods is their inherent volume-conserving property. Nevertheless, spurious bubbles and drops may be created. The reconstruction of the interface from the volume fractions and the computation of geometric quantities such as curvature are typically less accurate than other methods discussed here since the curvature and normal vectors are obtained by differentiating a nearly discontinuous function (volume fraction).

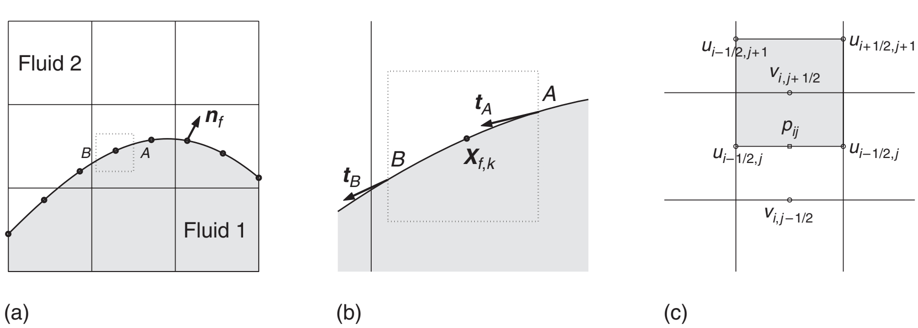
**Front-tracking methods** The basic idea behind the original front-tracking method is the use of two grids as illustrated in Figure 2. One is a standard, Eulerian finite difference mesh that is used to solve the fluid equations. The other is a discretized interface mesh that is used to explicitly track the interface and compute surface tension force which is then transferred to the finite difference mesh via a discrete delta-function. Front tracking was first proposed by Richtmyer and Morton and further developed by Glimm and co-workers.

A similar approach was taken by Unverdi and Tryggvason (see Tryggvason *et al.* (2001) and Peskin (2002) for recent reviews), who combined a moving grid description of the interface with flow computations on a fixed grid. In this immersed-boundary approach, all the fluid phases are treated together by solving a single set of governing equations. This method has its roots in the original marker-and-cell (MAC) method, where marker particles are used to identify each fluid and the immersed-boundary method of Peskin and McQueen, that was designed to track moving elastic boundaries in homogeneous fluids.

The interface is represented discretely by Lagrangian markers that are connected to form a front which lies within and moves through a stationary Eulerian mesh.

In Tryggvason's original implementation, the basic structural unit is a line segment. Since the interface moves and deforms during the computation, interface elements must occasionally be added or deleted to maintain regularity and stability. In the event of merging/breakup, elements must be relinked to effect a change in topology.

The interface is represented using an ordered list of marker particles $\boldsymbol{x}_k = ((x_1)_k, (x_2)_k), 1 \le k \le N$.




(a) (b) (c)

**Figure 2** (a) The basic idea in the front-tracking method is to use two grids – a stationary finite difference mesh and a moving Lagrangian mesh, which is used to track the interface. (b). Blow-up of the subgrid control volume in (a). (c) Control volume for the Eulerian mesh, $\Omega_{i,j+(1/2)}$.

The first step in this algorithm is the advection of the marker particles. A simple bilinear interpolation is used to find the velocity inside each grid cell (indicated in Figure 2c). The marker particles are then advected in a Lagrangian manner. Once the points have been advected, a list of connected polynomials $(p_i^x(s), p_i^y(s))$ is constructed using the marker particles. This gives a parametric representation of the interface, with $s$ typically an approximation of the arclength. Both lists are ordered and thus identify the topology of the interface. In later works, higher-order polynomials have been used (e.g., cubic splines) and semi-Lagrangian evolutions have been implemented where other tangential velocities have been used.

As the interface evolves, the markers drift along the interface following tangential velocities and more markers may be needed if the interface is stretched by the flow. Typically, the markers are redistributed along the interface to maintain an accurate interface representation.

Next, we compute the surface tension force,

$$\boldsymbol{F}_{\text{sing}}(\boldsymbol{x}, t) = \int_{\Gamma(t)} \tau \kappa_f \delta(\boldsymbol{x} - \boldsymbol{x}_f(s)) \boldsymbol{n}_f \mathrm{d}s \qquad [12]$$

where the subscript $f$ means values evaluated at the interface $\Gamma(t)$ and $s$ is arclength. The discrete numerical implementation of this distribution onto the fixed grid is in the form of a sum over interface elements, $\boldsymbol{x}_{f,k}$:

$$\boldsymbol{F}_{ij}(\boldsymbol{x}) = \sum_k \boldsymbol{f}_k \delta(\boldsymbol{x} - \boldsymbol{x}_{f,k}) \Delta s_k \qquad [13]$$

where $\Delta s_k$ is the average of the straight line distances from the point $\boldsymbol{x}_{f,k}$ to the two neighboring points $\boldsymbol{x}_{f,k+1}$ and $\boldsymbol{x}_{f,k-1}$ as indicated by the subgrid control volume shown in Figures 2a and 2b. The delta-function is typically taken to be Peskin's discrete Dirac delta-function:

$$\delta(\boldsymbol{x} - \boldsymbol{x}_{f,k}) = \begin{cases} \prod_{i=1}^{2} \dfrac{1}{4h}\left(1 + \cos\dfrac{\pi[\boldsymbol{x}_i - (\boldsymbol{x}_{f,i})_k]}{2h}\right) & \text{if } |\boldsymbol{x} - \boldsymbol{x}_{f,k}| \le 2h \\ 0 & \text{otherwise} \end{cases} \qquad [14]$$

Other higher-order alternative forms of the regularized delta-function using the product formula have recently been proposed.

Using the Frenet relation, the surface tension force on a short segment of the front is given by

$$\boldsymbol{f}_k = \int_A^B \tau \kappa_f \boldsymbol{n}_f \mathrm{d}s = \int_A^B \tau \frac{\partial \boldsymbol{t}_f}{\partial s} \mathrm{d}s = \tau(\boldsymbol{t}_B - \boldsymbol{t}_A) \qquad [15]$$

where $A$ and $B$ are the segment endpoints that lie on the boundary of the subgrid control volume (Figures 2a and 2b), and $\boldsymbol{t}_f$ is a tangent vector computed by fitting a polynomial to the endpoints of each element.

In the case of flows with varying density and/or viscosity between the fluid components, there is a need to calculate the phase indicator function $I(\boldsymbol{x}, t)$ (defined by interface geometry and position), which has the value 0 in fluid 1 and 1 in fluid 2. The indication function can be determined via the solution of the equation

$$\Delta I(\boldsymbol{x}, t) = \nabla \cdot \int_{\Gamma(t)} \boldsymbol{n}_f \delta(\boldsymbol{x} - \boldsymbol{x}_f(s, t)) \mathrm{d}s \qquad [16]$$

This equation is discretized on the Eulerian mesh and a discrete delta-function (e.g., eqn [14]) is used. The fluid properties such as density and viscosity are determined via the indicator function, that is, $\rho(\boldsymbol{x}, t) = \rho_1 + (\rho_2 - \rho_1) I(\boldsymbol{x}, t)$, etc.

As in the volume of fluid algorithm, the distribution form of the Navier–Stokes equations [5] are typically solved using a version of Chorin's projection method.

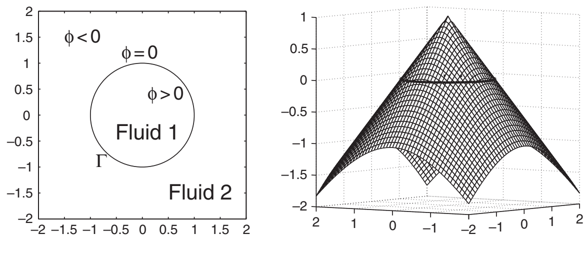
An alternative flow solver that can be used to integrate the flow equations in the presence of an interface is the immersed-interface method (IIM). The IIM was developed by Leveque and Li (see the review Li 2003), and can be used together with front-tracking as well as level-set methods.

The IIM directly incorporates jump conditions for the normal stress into the finite difference stencil. The key idea of this method is to use the jump conditions in Taylor series expansions of pressure and velocity near interfaces to derive difference equations that achieve pointwise second-order accuracy.

The principal advantage of front-tracking algorithms is their inherent accuracy, due in part to the ability to use a large number of grid points on the interface. Front-tracking methods can be complicated to implement, particularly in 3D, but give the precise location and geometry of the interface. In addition, explicit front tracking permits more than one interface to be present in a single computational cell without coalescence, which can be important in dense bubbly flows, emulsions, etc. One of major handicaps of front-tracking methods is the difficulty in modeling topological changes of the interface such as breakup and coalescence without *ad hoc* cut-and-connect and reconnecting parameterized interface (particularly, difficulties in 3D).

### Interface-Capturing Methods

**Level-set method** Level-set methods, introduced by Osher and Sethian (see the recent review papers (Osher and Fedkiw 2001, Sethian and Smereka




(a) (b)

**Figure 3** (a) Zero contour of $\phi$ representing the interface $\Gamma$. (b) Surface of $\phi$ with zero contour.

2003) and the recent texts (Osher and Fedkiw 2002, Sethian 1999)), are popular computational techniques for tracking moving interfaces. These methods rely on an implicit representation of the interface as the zero set of an auxiliary function (level-set function). The application of these methods to incompressible, multiphase flows started with the work of Osher, Merriman, Sussman, Smereka, Hou, and their collaborators.

In the level-set method, the level-set function $\phi(\boldsymbol{x}, t)$ is defined as follows (see Figure 3):

$$\phi(\boldsymbol{x}, t) \begin{cases} > 0 & \text{if } \boldsymbol{x} \in \text{fluid 1} \\ = 0 & \text{if } \boldsymbol{x} \in \Gamma \text{ (the interface between fluids)} \\ < 0 & \text{if } \boldsymbol{x} \in \text{fluid 2} \end{cases}$$

and the evolution of $\phi$ is given by

$$\phi_t + \boldsymbol{u} \cdot \nabla\phi = 0 \qquad [17]$$

which means that the interface moves with fluid.

To keep the interface geometry well resolved, the level-set function $\phi$ should be a distance function near the interface. However, under the evolution [17] it will not necessarily remain as such. We note that special velocity extensions $\boldsymbol{v}$ off the interface (i.e., $\boldsymbol{v} = \boldsymbol{u}$ at the interface, $\boldsymbol{v} \neq \boldsymbol{u}$ away from interface) have been recently developed to better maintain $\phi$ as a distance function (e.g., Sethian and Smereka (2003) and Macklin and Lowengrub (2005)). Typically, a reinitialization step (solving a Hamilton–Jacobi type equation, eqn [18]) below, is performed to keep $\phi$ as a distance function near the interface while keeping original zero-level set unchanged. More specifically, given a level-set function, $\phi$, at time t, the contours are redistributed according to the steady-state solution of the equation

$$\frac{\partial d}{\partial \tau} = S_\epsilon(\phi)(1 - |\nabla d|), \quad d(\boldsymbol{x}, 0) = \phi(\boldsymbol{x}) \qquad [18]$$

where $S_\epsilon$ is the smoothed sign function defined as

$$S_\epsilon(\phi) = \frac{\phi}{\sqrt{\phi^2 + \epsilon^2}} \qquad [19]$$

where $\epsilon$ is usually is one or two grid lengths. After solving eqn [18] to steady state $\phi(\boldsymbol{x}, t)$ is then replaced by $d(\boldsymbol{x}, \tau_{\text{steady}})$. Note that $d(\boldsymbol{x}, \tau_{\text{steady}})$ is typically a good approximation of the signed distance function.

The density and viscosity are defined as

$$\rho(\phi) = \rho_2 + (\rho_1 - \rho_2)H_\epsilon(\phi)$$

and

$$\eta(\phi) = \eta_2 + (\eta_1 - \eta_2)H_\epsilon(\phi) \qquad [20]$$

where $H_\epsilon(\phi)$ is the smoothed Heaviside function given by

$$H_\epsilon(\phi) = \begin{cases} 0 & \text{if } \phi < -\epsilon \\ \frac{1}{2}\left[1 + \frac{\phi}{\epsilon} + \frac{1}{\pi}\sin(\pi\phi/\epsilon)\right] & \text{if } |\phi| \le \epsilon \\ 1 & \text{if } \phi > \epsilon \end{cases}$$

The mollified delta-function is $\delta_\epsilon(\phi) = \mathrm{d}H_\epsilon/\mathrm{d}\phi$. The surface tension force is given as

$$\boldsymbol{F}_{\text{sing}} = -\tau\nabla \cdot \left(\frac{\nabla\phi}{|\nabla\phi|}\right)\delta_\epsilon(\phi)\frac{\nabla\phi}{|\nabla\phi|} \qquad [21]$$

The fluid equations [5] are solved using projection methods, the IIM or the ghost-fluid (GF) method (e.g., Osher and Fedkiw (2001, 2002) and Fedkiw *et al.* (2003)). The GF method is similar to the IIM in that jump discontinuities are incorporated in the finite difference stencil. In the GF algorithm, subcell resolution is used to mark the interface position and the values of discontinuous quantities are artificially extended to grid points neighboring the interface via extrapolation. A fully second order accurate GF method for moving interfaces has recently been developed (Macklin and Lowengrub 2005).

Applications of the level-set method include multiphase flows, viscoelastic fluid flows and fluid–structure interactions (e.g., see the reviews Osher and Fedkiw (2001, 2002), Sethian (1999), and Sethian and Smereka (2003)).

Advantages of the level-set algorithm include the simplicity with which it can be implemented, the ability to capture merging and breakup of interfaces automatically, and the ease with which the interface geometry can be described using the level-set function. A disadvantage of the level-set method is that mass is not conserved.

Accurate numerical simulations of multiphase flow and topology transitions require the computational mesh to resolve both the macroscales (e.g., droplet size, flow geometry) and the microscales to accurately capture local interface geometries near contact region, van der Waals forces, surfactant distribution, and Marangoni stresses. Adaptive mesh

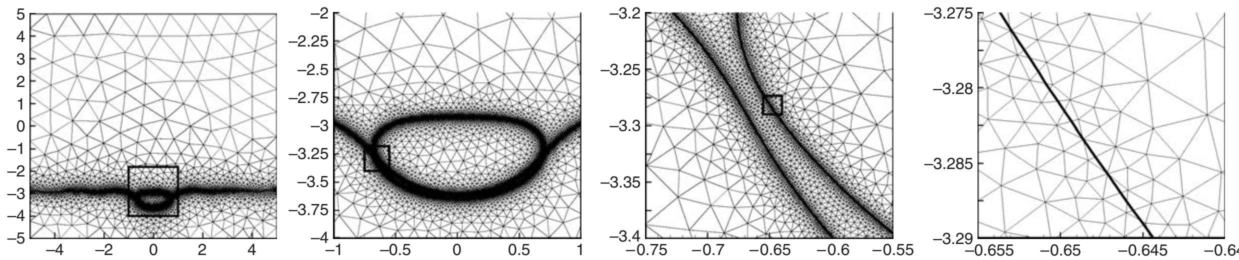


**Figure 4** Each of the first three figures has a boxed region that is magnified in the next figure. The rates of magnification are 5, 10, 40/3, respectively. The meshes in the figure are used to simulate the drop-impacting interface problem. Source: Zheng X, Anderson A, Lowengrub JS, and Cristini V (unpublished).

algorithms have recently been used greatly to increase accuracy and computational efficiency in level-set methods. Typically, the methods involve Cartesian adaptive mesh refinement. Problems tackled using this approach include droplet formation in inkjet printers and wake development behind a ship. Another approach, recently developed, is to use adaptive unstructured mesh refinement (Zheng *et al.* 2005), as shown in **Figure 4**, in which the impact of a drop onto a fluid interface is captured.

### Hybrid Methods

More recently, a number of hybrid methods, which combine good features of each algorithm, have been developed. These include coupled level-set volume-of-fluid (CLSVOF) algorithms, particle level-set methods, marker-VOF methods and level-contour front-tracking methods.

Level-set and VOF methods have recently been combined. The volume fraction is used to maintain volume conservation, while the level-set function is used to describe the interface geometry. After every time step, the volume-fraction function and level-set function are made compatible. The coupling between the level-set function $\phi$ and the VOF function $c$ occurs through the normal of the reconstructed interface and through the fact that the level-set function is reset to the exact signed normal distance to the reconstructed interface (where the area below the reconstructed interface is given by the volume-fraction function).

In the particle level-set method, Lagrangian disconnected marker particles are randomly positioned near the interface and are passively advected by the flow in order to rebuild the level-set function in under-resolved zones, such as high-curvature regions and near filaments. In these regions, the standard nonadaptive level-set method regularizes excessively the interface structure and mass is lost. The use of marker particles significantly ameliorates these difficulties.

Recently, a hybrid method has been developed, which uses both marker particles, to reconstruct and move the interface, and the volume-fraction function to conserve volume. In this approach, a smooth motion of the interface, typical of marker methods is obtained together with volume conservation, as in standard VOF methods. This work improves both the accuracy of interface tracking, when compared to standard VOF methods, and the conservation of mass, with respect to the original marker method.

Finally, a hybrid method that combines a level contour reconstruction technique with front-tracking methods has recently been developed to automatically model the merging and breakup of interfaces in three-dimensional flows.

### Phase-Field Method

Phase-field, or diffuse-interface, models are an increasingly popular choice for modeling the motion of multiphase fluids (see Anderson *et al.* (1998) for a recent review). In the phase-field model, sharp fluid interfaces are replaced by thin but nonzero thickness transition regions where the interfacial forces are smoothly distributed. The basic idea is to introduce a conserved order parameter (e.g., mass concentration) that varies continuously over thin interfacial layers and is mostly uniform in the bulk phases (see **Figure 5**).

For density-matched binary liquids (let $\rho = 1$ for simplicity), the coupling of the convective Cahn–Hilliard equation for the mass concentration with a modified momentum equation that includes a phase-field-dependent surface force is known as Model H (Hohenberg and Halperin 1977). In the case of fluids with different densities a phase-field model has been proposed by Lowengrub and Truskinovsky. Complex flow morphologies and topological transitions such as coalescence and interface breakup can be captured naturally and in a mass-conservative and energy-dissipative fashion since there is an associated free energy functional.

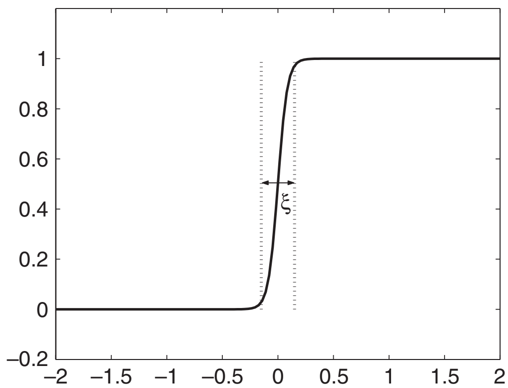



**Figure 5** A concentration prome across an interface with interface thickness, $\xi$.

The phase field is governed by the following advective Cahn–Hilliard equation:

$$\frac{\partial c}{\partial t} + \boldsymbol{u} \cdot \nabla c = \nabla \cdot (M(c)\nabla \mu) \quad [22]$$

$$\mu = F'(c) - \epsilon^2 \Delta c \quad [23]$$

where $M(c) = c(1-c)$ is the mobility, $F(c) = (1/4)c^2(1-c)^2$ is a Helmholtz free energy that describe the coexistence of immiscible phases, and $\epsilon$ is a measure of interface thickness and $\epsilon \sim \xi$ (see Figure 5). It can be shown that in the sharp interface limit $\epsilon \to 0$, the classical Navier–Stokes system equations and jump conditions are recovered.

The singular surface tension force is $\boldsymbol{F}_{\text{sing}} = -6\sqrt{2}\tau\epsilon\nabla \cdot (\nabla c \otimes \nabla c)$, where $\tau$ is the surface tension coefficient. An alternative surface tension force formulation based on the CSF is $\boldsymbol{F}_{\text{sing}} = -6\sqrt{2}\tau\epsilon\nabla \cdot (\nabla c/|\nabla c|)|\nabla c|\nabla c$.

Recently, very efficient nonlinear multigrid methods have been developed to solve implicit discretizations of the Cahn–Hilliard equation (e.g., Kim *et al.* (2004)). These schemes have been combined with projection methods to solve the Navier–Stokes equations to perform simulations of multiphase flows.

An example of simulation of liquid thread breakup using a phase-field method is shown in Figure 6. A long cylindrical thread of a viscous fluid 1 is in an infinite mass of another viscous fluid 2. If the thread becomes varicose with wavelength $\lambda$, the equilibrium of the column is unstable, provided $\lambda$ exceeds the circumference of the cylinder. This is the Rayleigh capillary instability that results in surface-tension-driven breakup of the thread.

An advantage of the phase-field approach is that it is straightforward to include more complex physical effects. For example, the binary model can be straightforwardly extended to describe three-component flows as follows.

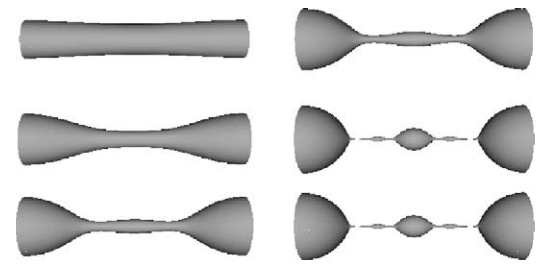


**Figure 6** Time evolution leading to multiple pinch-offs. The evolution is from top to bottom and left to right. The domain is axisymmetric, the initial velocities are zero everywhere, and the concentration field is given by $c(r,z) = 0.5(1 - \tanh((r - 0.5 - 0.05\cos(z))/(2\sqrt{2}\epsilon)))$ on $\Omega = (0,\pi) \times (0, 2\pi)$. Densities are matched and viscosity ratio is 0.5.

Consider a ternary mixture and denote the composition of components 1, 2, and 3, expressed as mass fractions, by $c_1, c_2$, and $c_3$, respectively. Therefore,

$$\sum_{i=1}^{3} c_i = 1, \quad 0 \le c_i \le 1 \quad [24]$$

The composition of a ternary mixture (A, B, and C) can be mapped onto an equilateral triangle (the Gibbs triangle (Porter and Easterling 1993)) whose corners represent 100% concentration of A, B, or C as shown in Figure 7a. Mixtures with components lying on lines parallel to $\overline{BC}$ contain the same percentage of A, those with lines parallel to $\overline{AC}$ have the same percentage of B concentration, and analogously for the C concentration. In Figure 7a, the mixture at the position marked '∘' contains 60% A, 10% B, and 30% C. Because the concentrations sum to unity, only two of them need to be determined, say $c_1, c_2$.

The evolution of $c_1$ and $c_2$ is governed by the following advective ternary Cahn–Hilliard equation:

$$\frac{\partial c_1}{\partial t} + \boldsymbol{u} \cdot \nabla c_1 = \nabla \cdot (M(c_1, c_2)\nabla \mu_1) \quad [25]$$

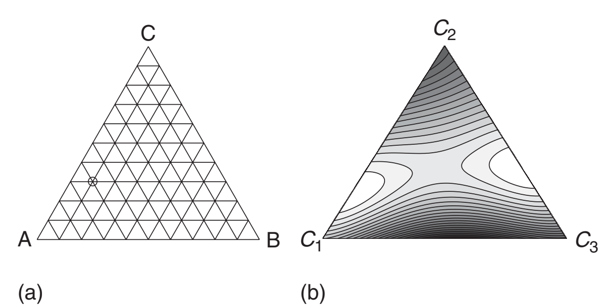



**Figure 7** (a) Gibbs triangle. (b) Contour plot of the free energy $F(c_1, c_2)$ on the Gibbs triangle.

$$\frac{\partial c_2}{\partial t} + \boldsymbol{u} \cdot \nabla c_2 = \nabla \cdot (M(c_1, c_2) \nabla \mu_2) \qquad [26]$$

$$\mu_1 = \frac{\partial F(c_1, c_2)}{\partial c_1} - \epsilon^2 \Delta c_1 - 0.5\epsilon^2 \Delta c_2 \qquad [27]$$

$$\mu_2 = \frac{\partial F(c_1, c_2)}{\partial c_2} - 0.5\epsilon^2 \Delta c_1 - \epsilon^2 \Delta c_2 \qquad [28]$$

where $M(c_1, c_2) = \sum_{i<j}^{3} c_i c_j$ is the mobility and $F(c_1, c_2)$ is the Helmholtz free energy that can be used to model the miscibility of the components. An example of a free energy (used in the simulation shown in Figure 8 below) for which fluids 1 and 3 are immiscible and fluid 2 is preferentially miscible with fluid 3 is:

$$F(c_1, c_2) = 2c_1^2(1 - c_1 - c_2)^2 + (c_1 + 0.2)(c_2 - 0.2)^2 + (1.2 - c_1 - c_2)(c_2 - 0.4)^2$$

The contours of $F$ on the Gibbs triangle are shown in Figure 7b.

The singular surface tension force is $\boldsymbol{F}_{\text{sing}} = -6\sqrt{2}\epsilon \sum_{i=1}^{3} \tau_i \nabla \cdot (\nabla c_i \otimes \nabla c_i)$, where the physical surface tension coefficients $\tau_{ij}$ between two fluids $i$ and $j$ are decomposed into the phase-specific surface tensions $\tau_i$ such that $\tau_{ij} = \tau_i + \tau_j$.

As a demonstration of the evolution possible in partially miscible liquid systems, we present an example in which there is a gravity-driven (Rayleigh–Taylor) instability that enhances the transfer of a preferentially miscible contaminant from one immiscible fluid to another in 2D. In this system, the ternary Cahn–Hilliard system is solved using nonlinear multigrid methods and a projection method (Kim and Lowengrub (in press)) is used to solve the flow equations [5].

In Figure 8 (first column), the top half of the domain initially consists of a mixture of fluids 1 and 2, and the bottom half consists of fluid 3, which is immiscible with fluid 1. The contours of $c_1, c_2$, and $c_3$ are visualized in gray-scale where darker regions denote larger values of $c_1, c_2$, and $c_3$, respectively. In the top row, the contours of fluid 1 are shown, the middle and bottom rows correspond to fluids 2 and 3, respectively.

Fluid 2 is preferentially miscible with fluid 3. Fluid 1 is assumed to be the lightest and fluid 2 the heaviest. The density of the 1/2 mixture is heavier than that of fluid 3, so the density gradient induces the Rayleigh–Taylor instability.

The evolution of the three phases is shown in Figure 8. As the simulation begins, the 1/2 mixture falls and fluid 2 diffuses into fluid 3. A characteristic Rayleigh–Taylor (inverted) mushroom forms, the

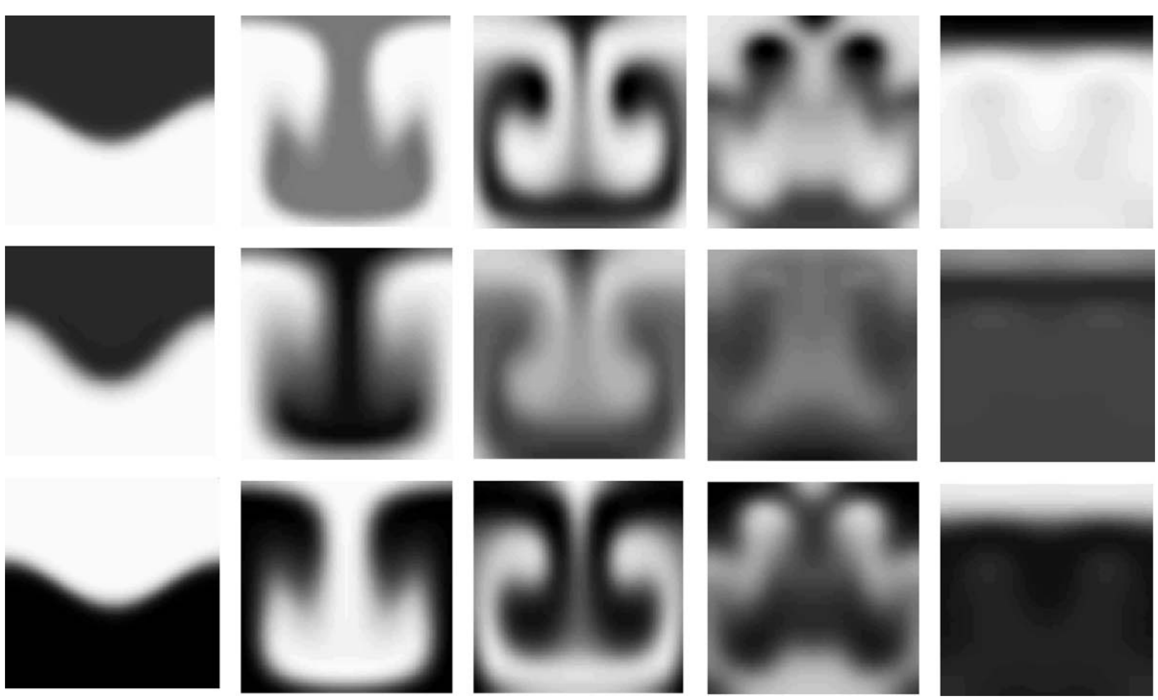


**Figure 8** Evolution of concentration of fluid 1 (top row), 2 (middle row), and 3 (bottom row). The contours of $c_1, c_2$, and $c_3$ are visualized in gray-scale where darker regions denote larger values of $c_1, c_2$, and $c_3$, respectively.

surface area of the 1/3 interface increases, and vorticity is generated and shed into the bulk. As fluid 2 is diffused from fluid 1, the pure fluid 1 rises to the top as shown in **Figure 8**. Imagining that fluid 2 is a contaminant in fluid 1, this configuration provides an efficient means of cleansing fluid 1 since the buoyancy-driven flow enhances the diffusional transfer of fluid 2 from fluid 1 to fluid 3.

The advantages of the phase-field method are: (1) topology changes are automatically described; (2) the composition field $c$ has a physical meaning not only near interface but also in the bulk phases; (3) complex physics can easily be incorporated into the framework, the methods can be straightforwardly extended to multicomponent systems, and miscible, immiscible, partially miscible, and lamellar phases can be modeled.

Associated with diffuse interfaces is a small scale $\epsilon$, proportional to the width of the interface. In real physical systems describing immiscible fluids, $\epsilon$ can be vanishingly small. However, for numerical accuracy $\epsilon$ must be at least a few grid lengths in size. This can make computations expensive. One way of ameliorating this problem is to adaptively refine the grid only near the transition layer. Such methods are under development by various research groups.

Phase-field methods have been used to model viscoelastic flow, thermocapillary flow, spinodal decomposition, the mixing and interfacial stretching, in a shear flow, droplet breakup process, wave-breaking and sloshing, the fluid motion near a moving contact line, and the nucleation and annihilation of an equilibrium droplet (see the references in the review paper Anderson *et al.* (1998)).

## Conclusions and Future Directions

In this paper we have reviewed the basic ideas of interface-tracking and interface-capturing methods that are critical in simulating the motion of interfaces in multicomponent fluid flows. The differences between these various formulations lie in the representation and the reconstruction of interfaces. The advantages and disadvantages of the algorithms have been discussed. While there has been much progress on the development of robust multifluid solvers, there is much more work to be done. Promising future directions for research include the incorporation of adaptive mesh refinement into the algorithms and the development of efficient hybrid schemes that combine the best features of individual methods.

*See also:* Breaking Water Waves; Capillary Surfaces; Fluid Mechanics: Numerical Methods; Incompressible Euler Equations: Mathematical Theory; Inviscid Flows; Non-Newtonian Fluids; Partial Differential Equations: Some Examples; Viscous Incompressible Fluids: Mathematical Theory; Vortex Dynamics.

## Further Reading

Anderson DM, McFadden GB, and Wheeler AA (1998) Diffuse-interface methods in fluid mechanics. *Ann. Rev. Fluid Mech.* 30: 139–165.

Cristini V, Blawzdziewicz J, and Loewenberg M (2001) An adaptive mesh algorithm for evolving surfaces: simulations of drop breakup and coalescence. *Journal of Computational Physics* 168: 445–463.

Fedkiw RP, Sapirop G, and Shu C-W (2003) Shock capturing, level sets and PDE based methods in computer vision and image processing: a review of Osher's contributions. *Journal of Computational Physics* 185: 309–341.

Hohenberg PC and Halperin BI (1977) Theory of dynamic critical phenomena. *Reviews of Modern Physics* 49: 435–479.

Hou TY, Lowengrub JS, and Shelley MJ (2001) Boundary integral methods for multicomponent fluids and multiphase materials. *Journal of Computational Physics* 169: 302–362.

James AJ and Lowengrub J (2004) A surfactant-conserving volume-of-fluid method for interfacial flows with insoluble surfactant. *Journal of Computational Physics* 201: 685–722.

Kim JS, Kang KK, and Lowengrub JS (2004) Conservative multigrid methods for Cahn–Hilliard fluids. *Journal of Computational Physics* 193: 511–543.

Kim JS and Lowengrub JS Phase field modeling and simulation of three-phase flows, Int. Free Bound (in press).

Li Z (2003) An overview of the immersed interface method and its applications. *Taiwanese J. Math.* 7: 1–49.

Macklin P and Lowengrub JS (2005) Evolving interfaces via gradients of geometry-dependent interior Poisson problems: application to tumor growth. *Journal of Computational Physics* 203: 191–220.

Osher S and Fedkiw RP (2001) Level set methods: an overview and some recent results. *Journal of Computational Physics* 169: 463–502.

Osher S and Fedkiw RP (2002) *Level Set Methods and Dynamic Implicit Surfaces*. Springer.

Peskin CS (2002) The immersed boundary method. *Acta Numerica* 1–39.

Porter DA and Easterling KE (1993) *Phase Transformations in Metals and Alloys*. Van Nostrand Reinhold.

Pozrikidis C (2001) Interfacial dynamics for Stokes flow. *Journal of Computational Physics* 169: 250–301.

Scardovelli R and Zaleski S (1999) Direct numerical simulation of free-surface and interfacial flow. *Annu. Rev. Fluid Mech.* 31: 567–603.

Sethian JA (1999) *Level-Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision and Materials Science*. Cambridge, MA: Cambridge University Press.

Sethian JA and Smereka P (2003) Level set methods for fluid interfaces. *Annu. Rev. Fluid Mech.* 35: 341–372.

Tryggvason G, Bunner B, Esmaeeli A, Juric D, Al-Rawahi N *et al.* (2001) A front-tracking method for the computations of multiphase flow. *Journal of Computational Physics* 169: 708–759.

Zheng X, Lowengrub J, Anderson A, and Cristini V (2005) Adaptive unstructured volume remeshing II. Application to two- and three-dimensional level-set simulations of multiphase flow. *Journal of Computational Physics* 208: 626–650.

# Intermittency in Turbulence

**J Jiménez**, Universidad Politecnica de Madrid, Madrid, Spain



## Introduction

Intermittency has several meanings in turbulence. The oldest one, now most often labeled "external" or "large-scale" intermittency, refers to the coexistence of turbulent and laminar regions in inhomogeneous turbulent flows, such as in boundary layers or in free shear layers. In those cases, the interface between laminar irrotational flow and turbulent vortical fluid is typically sharp and corrugated. An observer sitting near the edge of the layer is immersed in turbulent fluid only part of the time.

The intermittency coefficient $\gamma$ measures the fraction of turbulent fluid over the sampling universe over which the statistics are taken. For example, in a boundary layer such as that in Figure 1, the intermittency coefficient as a function of wall distance measures the fraction of turbulent fluid at a given distance from the wall. External intermittency is important in any attempt to model realistic turbulent flows, which are almost always inhomogeneous. Consider, for example, the classical homogeneous relation in eqn [1] between the mean kinetic energy $K$ of the turbulent fluctuations and the energy dissipation rate $\varepsilon$:

$$\varepsilon = C\frac{K^{3/2}}{L} \qquad [1]$$

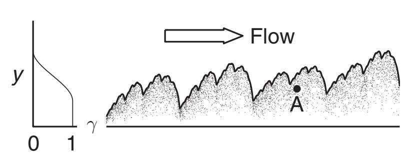



**Figure 1** Sketch of a turbulent boundary layer, and of the associated intermittency factor. An observer such as A, at a distance $y$ from the wall, only sees turbulent flow for a fraction $\gamma$ of the time.

where $L$ is the length scale of the largest eddies, and $C \approx 0.1$ is an experimentally determined constant. Such relations are often implicit in turbulent models, and they have to be modified to account for intermittency. Equation [1] only holds within the turbulent regions where the energy and the dissipation rates are $K_T$ and $\varepsilon_T$, while the overall mean values used in the modeling conservation equations are $K = \gamma K_T$ and $\varepsilon = \gamma\varepsilon_T$. The true overall relation should therefore be

$$\varepsilon = C\gamma^{-1/2}\frac{K^{3/2}}{L} \qquad [2]$$

which may differ substantially from eqn [1], especially near the edge of the layer. Experimental values and rough theoretical estimates for the distribution of the intermittency coefficient are available for most practical turbulent flows.

## Internal Intermittency

While the external intermittency just described is probably the most important one from the point of view of applications, it is not the most interesting from the theoretical point of view. Turbulence is a multiscale phenomenon which is inhomogeneous at all length scales, from the largest ones to the inner viscous cutoff (*see* Turbulence Theories). Moreover, this inhomogeneity goes beyond what could be expected just from the statistics of a random process. Consider, for example, the velocity difference $\Delta u$ between two points separated by a distance $r$. The original Kolmogorov formulation of the energy cascade assumes that the probability density function (PDF), $p(\Delta u)$, is a universal function in the inertial range of scales, whose only parameter is a velocity scale depending on $r$. It then follows from Kolmogorov's analysis that

$$p(\Delta u) = F\left[\Delta u/(\bar{\varepsilon}r)^{1/3}\right] \qquad [3]$$

where $\bar{\varepsilon}$ is the average energy transfer rate across scales per unit mass, and the average $\overline{(\ )}$ is taken either over the whole flow or over a suitably designed ensemble of experiments. In an equilibrium system,

global energy conservation implies that $\bar{\varepsilon}$ is equal to the average viscous dissipation per unit mass:

$$\bar{\varepsilon} = \nu\overline{|\nabla \boldsymbol{u}|^2} \qquad [4]$$

In eqn [4], the kinematic viscosity of the fluid is $\nu$, and $|\nabla \boldsymbol{u}|$ is the $L_2$-norm of the velocity gradient tensor. Equation [3] is valid as long as the separation $r$ is much larger than the Kolmogorov viscous cutoff $\eta = (\nu^3/\bar{\varepsilon})^{1/4}$, and much smaller than the integral scale of the largest eddies $L_\varepsilon = u'^3/\bar{\varepsilon}$, where $u'$ is the root-mean-square value of the fluctuations of one velocity component. The extent of this inertial range is a function of the Reynolds number $Re_L = u'L_\varepsilon/\nu$:

$$L_\varepsilon/\eta = Re_L^{3/4} \qquad [5]$$

The strict similarity hypothesis in eqn [3] is not well satisfied by experiments. While the velocity distribution at a given point is approximately Gaussian, Figure 2a shows that the velocity increments become increasingly non-Gaussian as the spatial separation is made much smaller than $L_\varepsilon$. It was also soon noted that the dependence of eqn [3] on a single parameter such as $\bar{\varepsilon}$ was theoretically suspect, since it is difficult to see how the PDFs of a whole set of local properties, such as the $\Delta u$ for different intervals, could depend only on a single global property. Kolmogorov himself sought to bypass that difficulty by substituting eqn [3] by a “refined similarity” hypothesis,

$$p(\Delta u) = F\left[\Delta u/(\varepsilon_r r)^{1/3}\right] \qquad [6]$$

where $\varepsilon_r$ is no longer a global average, but the mean value of the dissipation over a ball of radius of order $r$ centered at the midpoint of the interval. This refined similarity is better satisfied by experiments (see Figure 2b), although, from the practical point of view, it just transfers the problem of characterizing $\Delta u$ to that of characterizing the statistics of $\varepsilon_r$.

It has become customary to measure the behavior of $p(\Delta u)$ in terms of its structure functions,

$$S(n) = \int_{-\infty}^{\infty} \Delta u^n p(\Delta u)\,\mathrm{d}\Delta u \qquad [7]$$

which can be normalized as generalized flatness factors,

$$\sigma(n) = S(n)/S(2)^{n/2} \qquad [8]$$

It follows from the strict similarity hypothesis [3] that

$$S(n) \sim r^{n/3} \qquad [9]$$

and that all the $\sigma(n)$ should be independent of the separation.

For example, the fourth-order flatness of a Gaussian distribution is $\sigma(4) = 3$. Figure 3 shows that this is not true. The flatness increases as the separation decreases, and it only levels off at lengths of the order of the Kolmogorov viscous scale. For separations in that viscous range the flow is smooth, $\Delta u \approx (\partial_x u)r$, and

$$\sigma(n) \approx \overline{(\partial_x u)^n}/\overline{(\partial_x u)^2}^{\,n/2} \qquad [10]$$

It follows from eqn [10] and from Figure 3 that the velocity gradients become increasingly non-Gaussian as $L_\varepsilon$ and $\eta$ separate at high Reynolds numbers. The velocity differences across intervals which are large with respect to $\eta$ also become very non-Gaussian when $r \ll L_\varepsilon$.

Because the velocity difference between two points which are not too close to each other can be expressed as the sum of velocity differences over subintervals, a loose application of the central limit

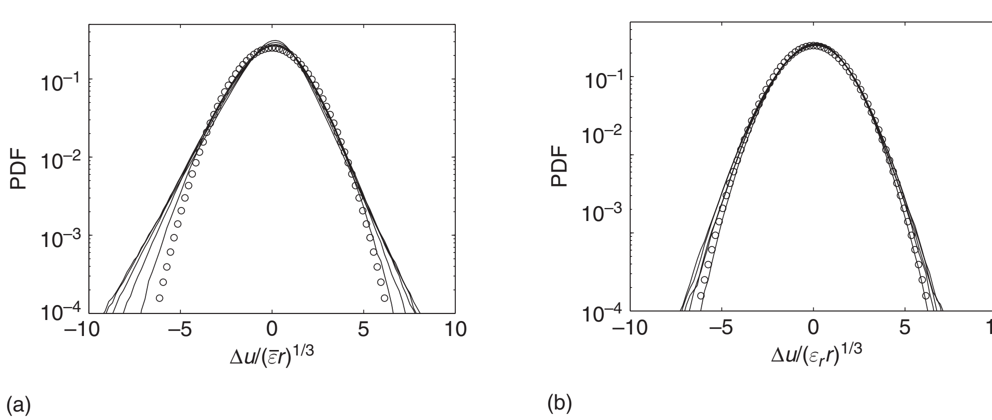



**Figure 2** PDFs of the differences of the velocity component in the direction of the separation (for separations in the inertial range of scales). $r/L_\varepsilon = 0.02-0.36$, increasing by factors of 2; equivalent to $r/\eta = 180-3000$. Nominally isotropic turbulence at Reynolds number $\mathrm{Re}_L = 10^5$.. (a) $\Delta u$ is normalized with the global energy dissipation rate $\bar{\varepsilon}$; distributions are wider as the separation decreases. (b) $\Delta u$ is scaled with the locally averaged dissipation over the separation interval. Data courtesy of H Willaime and P Tabeling.

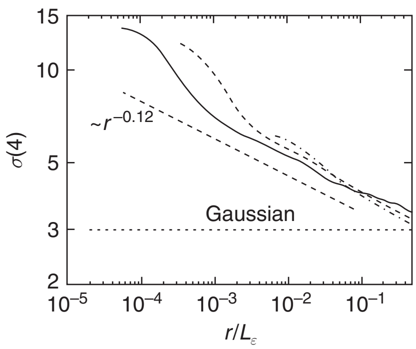



**Figure 3** Fourth-order flatness of the differences of the velocity component in the direction of the separation, for separations in the inertial range of scales, $r/L_\varepsilon = 0.5$ to $r/\eta = 2$.. The Reynolds numbers of the different flows range from $Re_L = 1800$ to $10^6$.. Data in part courtesy of H Willaime, P Tabeling, and R A Antonia.

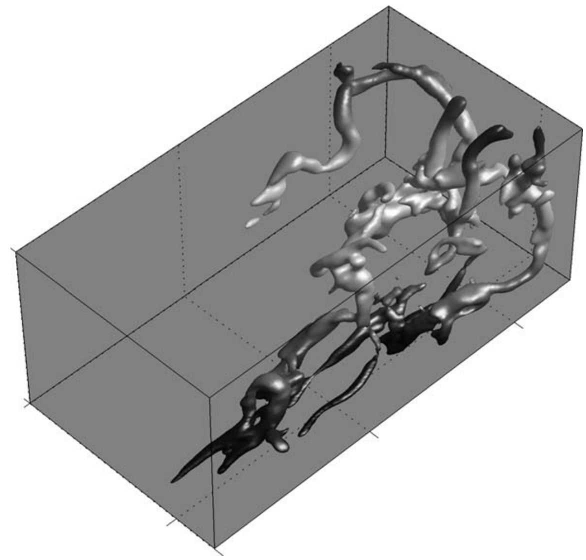


**Figure 4** Intense vortex tangle in the logarithmic layer of a turbulent channel. The vortex diameters are of the order of $10\eta$, and the size of the bounding box is of the order of the channel width. Reproduced with permission of J C del Álamo.

theorem would suggest that its PDF should be roughly Gaussian. The key conditions for that to happen are that the summands should be mutually independent, that their magnitudes should be comparable, and that each of them has a probability distribution with a finite variance. The first of those three conditions is probably a good approximation if the separation is much longer than the viscous cutoff, but the second one depends on the structure of the flow. The experimental non-Gaussian behavior suggests the existence of occasional very strong velocity jumps. In the viscous range of scales, those structures have been identified both experimentally and numerically as very strong linear vortices, in whose neighborhoods the strongest gradients are generated. An example of a tangle of such structures is shown in Figure 4.

In another example, the vorticity in decaying two-dimensional turbulence concentrates very quickly into relatively few strong compact vortices, which are stable except when they interact with each other. The velocity field is dominated by them, and the flatness of the velocity increments reaches values of the order of $\sigma(4) \approx 50$–$100$, even at moderate Reynolds numbers. That case is interesting because something can be said about the probability distribution of the velocity gradients. We have noted that the PDF of a sum of mutually comparable independent random variables with finite variances tends to Gaussian when the number of summands is large. This well-known theorem is a particular case of a more general result about sums of random variables whose incomplete second moments diverge as

$$\mu_2(s) = \int_{-s}^{s} x^2 p(x) \mathrm{d}x \sim s^{2-\alpha} \quad \text{when } s \to \infty \qquad [11]$$

When $0 < \alpha \leq 2$, the sums of such variables tend to a family of "stable" distributions parametrized by $\alpha$. The Gaussian case is the limit of that family when $\alpha = 2$. In the case of two-dimensional vortices with very small cores, the velocity gradients at a distance $R$ from the center of the vortex behave as $1/R^2$. If we take $s$ in eqn [11] to be one of those velocity derivatives, its probability distribution is proportional to the area covered by gradients with a given magnitude, and

$$\mu_2(s) \sim \int_0^{s^{1/2}} R^{-4} 2\pi R \, \mathrm{d}R \sim s^{-1} \qquad [12]$$

The velocity derivatives at any point, which are the sums of the velocity derivatives induced by all the randomly distributed neighboring vortices, should therefore be distributed according to the stable distribution with $\alpha = 1$, which is Cauchy's

$$p(s) = \frac{c}{\pi(c^2 + s^2)} \qquad [13]$$

This distribution has no moments for $n > 1$. Its tails decay as $s^{-2}$, and the distribution of the gradients essentially reflects the properties of the closest vortex. In real two-dimensional turbulent flows, the distribution [13] is followed fairly well, but its extreme tails only reach to the maximum values of the velocity gradient found within the viscous vortex cores, which are not exactly point vortices.

Other similar general results can be derived that link the behavior of the structure functions with the properties of the stable distributions corresponding to the type of flow singularities expected in the limit of infinite Reynolds number.

The common feature of the two cases just described is the presence of strong structures that live for long times because viscosity stabilizes them. They are therefore more common than what could be expected on purely statistical grounds. They are responsible for the tails of the probability distributions of the velocity derivatives, but they are not the only intermittent features of turbulent flows. The increase of the flatness in Figure 3 below $r \approx 50\eta$ is clearly connected with the presence of the coherent vortices, but even for larger separations there is a smooth evolution of $\sigma(4)$ that suggests that the formation of intense structures is a gradual process that takes place across the inertial range. Much less is known about those hypothetical inertial structures than about the viscous ones.

We can now recast the problem of intermittency in Navier–Stokes turbulence into geometric terms. The defining empirical observation for that system is that the energy dissipation given by eqn [4] does not vanish even in the infinite Reynolds number limit in which $\nu \to 0$. This means that the flow has to become singular as $\overline{|\nabla \boldsymbol{u}|}L_\epsilon/u' \sim Re_L^{1/2}$. The strict similarity approximation assumes that those singularities are uniformly distributed across the flow, but the experimental evidence just discussed shows that this is not true. The singularities are distributed inhomogeneously, and the inhomogeneity develops across the inertial cascade. The problem of intermittency is to characterize the geometry of the support of the flow singularities in the limit of infinite Reynolds number.

In the absence of detailed physical mechanisms for the dynamics of the inertial range, most intermittency models are based on plausible processes compatible with the invariances of the inviscid Euler equations. The precise power law given in eqn [9] for the structure functions depends on the strict similarity hypothesis [3], but the fact that it is a power law only depends on the scaling invariances of the equations of motion. The energies and sizes of the eddies in the inertial range are too small for the integral scales of the flow to be relevant, and too large for the viscosity to be important. They therefore have no intrinsic velocity or length scales. Under those conditions, any function of the velocity which depends on a length has to be a power. Consider a quantity with dimensions of velocity, such as $u(r) = S(n)^{1/n}$, which is a function of a distance such as $r$. On dimensional grounds we should be able to write it as

$$u(r) = UF(\rho) \qquad [14]$$

where $\rho = r/L$, and $L$ and $U(L)$ are arbitrary length and velocity scales. The value of $u(r)$ should not depend on the choice of units, and we can differentiate eqn [14] with respect to $L$ to give

$$\partial_L u = (\mathrm{d}U/\mathrm{d}L)F(\rho) - U\rho L^{-1}(\mathrm{d}F/\mathrm{d}\rho) = 0 \qquad [15]$$

which can only be satisfied if

$$\frac{\mathrm{d}F}{\mathrm{d}\rho} = \zeta F \Rightarrow F \sim \rho^{\zeta} \qquad [16]$$

and $\zeta = L(\mathrm{d}U/\mathrm{d}L)/U$ is constant. This suggests generalizing eqn [9] to

$$S(n) \sim r^{\zeta(n)} \qquad [17]$$

where the exponents are empirically adjusted. Only $\zeta(3) = 1$ can be derived directly from the Navier–Stokes equations. Equation [17] implies that $\sigma(n)$ satisfies a power law with exponent $\zeta(n) - n\zeta(2)/2$. In Figure 3, for example, the flatness follows a reasonably good power law outside the viscous range, consistent with $\zeta(4) - 2\zeta(2) \approx -0.12$. The anomalous behavior near the viscous limit, and similar limitations at the largest scales, mean that only very high Reynolds number flows can be used to measure the scaling exponents, and that the range over which they are measured is never very large. Moreover, the integrand of the higher-order structure functions peaks at the extreme tails of the probability distributions of the velocity differences, which implies that very long experimental samples have to be used to accumulate enough statistics to measure the high-order exponents. For these and for other reasons, the scaling exponents above $n \gtrsim 8-10$ are poorly known. This is unfortunate because we will see later that some of the most interesting intermittency properties of the velocity field, such as the nature of the flow singularities in the infinite Reynolds number limit, depend on the behavior of the $\zeta(n)$ for large $n$.

Experimental values for the scaling exponents are given in Table 1. They are generally smaller than the ones predicted by the strict similarity approximation, implying that the moments of the velocity differences decrease with the separation more slowly than they would if they were self-similar, and suggesting that new stronger structures become important as the scale decreases.

Note that we have included in the table values for odd-order powers. Up to now we have not specified

**Table 1** Longitudinal scaling exponents

| *Order* | *Experimental* | *Strict similarity* |
|---|---|---|
| 2 | 0..70 ± ..01 | 0.667 |
| 3 | 1.00 | 1 |
| 4 | 1..30 ± ..03 | 1.333 |
| 5 | 1..56 ± ..04 | 1.667 |
| 6 | 1..79 ± ..03 | 2.000 |
| 7 | 1..99 ± ..10 | 2.333 |
| 8 | 2..22 ± ..05 | 2.667 |

The values on the second column are averages from different experiments, and the standard deviations reflect scatter among experiments. The third column is the value from the strict similarity equation [9].

which velocity component is being analyzed, but most experiments refer to the one in the direction of the separation. That is the easiest case to measure, specially if time is used as a surrogate for distance, and those PDFs are not symmetric even in isotropic turbulence. Negative increments are more common than positive ones because of the extra energy required to stretch a vortex, and the effect is clearly visible in the distributions in Figure 2. Those longitudinal odd-order structure functions do not vanish, and their scaling exponents are the ones given in the table. The transverse structure functions are those in which the velocity component is normal to the separation, and their odd-order moments vanish by symmetry in isotropic turbulence. There has been a lot of discussion about whether the longitudinal scaling exponents of even orders differ from the transverse ones. Early results suggested that the latter are lower than the former, undermining the case for intermittency theories based on similarity arguments, and suggesting that a more mechanistic approach was needed. The present consensus seems to be that both sets of exponents are equal, but that there are residual effects of low Reynolds numbers and of flow anisotropy that are difficult to avoid experimentally. The question is still open.

## Multiplicative Models

The most successful phenomenological models for the geometry of intermittency are based on the concept of a multiplicative cascade. Consider some flow property $v$, such as the locally averaged energy transfer rate by eddies of size $r_k$, which cascades into smaller eddies of size $r_{k+1}$ which is some fraction of $r_k$. Denote by $p_k(v_k)$ the probability distribution of the value of $v$ at the step $k$ of the cascade.

Assume that the cascade is Markovian in the sense that the probability distribution of $v_k$ depends only on its value in the previous step,

$$p_{k+1}(v_{k+1}) = \int p_{\mathrm{T}}(v_{k+1}|v_k;\, k)p_k(v_k)\,\mathrm{d}v_k \qquad [18]$$

This is in contrast to some more complicated functional dependence, such as on the values of $v_k$ in some extended spatial neighborhood, or on several previous cascade stages. This assumption intuitively implies that $v_{k+1}$ evolves faster, or on a smaller scale, than $v_k$, and that it is in some kind of equilibrium with its precursor. If the cascade is deterministic in that sense, $v_k$ can be represented as a product

$$v_k/v_0 = q_k q_{k-1} \ldots q_1 \qquad [19]$$

in which the factors $q_k = v_k/v_{k-1}$ are statistically independent of each other.

If the underlying process is invariant to scaling transformations, the transition probability density function has to have the form

$$p_{\mathrm{T}}(v_{k+1}|v_k) = v_k^{-1} w(q_{k+1};\, k) \qquad [20]$$

The multiplicative model works most naturally for positive variables, and we will assume that to be the case in the following, but most results can be generalized to arbitrary distributions. We will also assume for simplicity that all the cascade steps are equivalent, so that the distribution $w(q)$ of the multiplicative factors is independent of $k$, and depends only on our choice for $r_{k+1}/r_k$.

Local deterministic self-similar cascades lead naturally to intermittent distributions, in the sense that the high-order flatness factors for $v_k$ become arbitrarily large as $k$ increases. It follows from eqns [18]–[20] that the $n$th order moment for $p_k$ can be written as

$$S_k(n) = \int \xi^n p_k(\xi)\,\mathrm{d}\xi = S_0(n)S_w(n)^k \qquad [21]$$

where $S_w(n)$ is the $n$th order moment of the multiplicative factor $q$, and $n$ is any real number for which the integral exists. If we define flatness factors as in eqn [7], we can rewrite eqn [21] as

$$\sigma_k(n) = \sigma_0(n)\sigma_w(n)^k \qquad [22]$$

It follows from Chebichev's inequality that

$$S(n) \geq S(n-2)S(2) \geq S(n-4)S(2)^2 \ldots \qquad [23]$$

from where

$$1 \le \sigma(4) \le \sigma(6) \ldots \quad [24]$$

which is true for any distribution of positive numbers. Equality only holds for trivial distributions concentrated on a single value. The product in eqn [22] therefore increases without bound with the number of cascade steps, and the flatness factors diverge.

It is tempting to substitute $k$ in [21] by a continuous variable, in which case the PDFs form a continuous semigroup generated by infinitesimal scaling steps. This leads to beautiful theoretical developments, but it is not necessarily a good idea from the physical point of view. For example, while it might be reasonable to assume that the properties of an eddy of size $r$ depend only on those of the eddy of size $2r$ from which it derives, the same argument is weaker when applied to eddies of almost equal sizes. We will restrict ourselves here to the discrete case.

### Limiting Distributions

The multiplicative process just described can be summarized as a family of distributions $p_k(v_k)$ such that the probability density for the product of two variables is

$$p(v_{k_1}v_{k_2}) = p_{k_1+k_2}(v_{k_1+k_2}) \quad [25]$$

and it is natural to ask whether there is a limiting distribution for large $k$. We know that, in the case of sums, rather than products, such distributions tend to be Gaussian under fairly general conditions, and the first attempt to analyze [25] was to reduce it to a sum by defining

$$z = k^{-1}\log(v_k/v_0) \quad [26]$$

The argument was that $z$ would tend to a Gaussian distribution, and that the limiting distribution for $v_k$ would be lognormal. This was soon shown to be incorrect. The central part of the distribution approaches lognormality, but the tails do not, because the central limit theorem says nothing about their behavior. The family of lognormal distributions is a fixed point of eqn [25], but it is unstable, and it is only attained if the individual generating distributions are themselves lognormal.

The lognormal distribution has moments

$$S_w(n) = \exp(an + bn^2) \quad [27]$$

which are conserved under [21], so that the product of lognormally distributed variables stays lognormal. The moments in eqn [27] are generated by the recursive relation

$$Q_w(n) = \frac{S_w(n+3)S_w^3(n+1)}{S_w(n)S_w^3(n+2)} = 1 \quad [28]$$

with suitable conditions for $n < 2$. Under [21], $Q_k(n) = Q_w^k(n)$, and it is clear that only when all the $Q_w(n)$ are exactly equal to 1 do they continue to be so under multiplication. Otherwise, any $Q_w$ initially larger than 1 tends to infinity after enough cascade steps, while any one initially smaller than 1 tends to 0. Only an exactly lognormal distribution of the generating factors results in a lognormal limiting distribution, and even small errors lead to very different patterns of moments. This contrasts with the situation for sums of random variables, in which the Gaussian distribution is not only a fixed point, but also has a very large basin of attraction.

### Multifractals

The problem with using the transformation [26] to find the limiting distribution of a multiplicative process is not so much the technique of analyzing the statistics of products in terms of those of sums, but the inappropriate use of the central limit theorem. It can be bypassed by using instead the theory of large deviations of sums of random variables. The key result is obtained by expanding the characteristic function of $p_k$ when $k \gg 1$, and states that

$$p_k(v_k) \approx \left(\frac{-\phi_0''}{2\pi k}\right)^{1/2} \mathrm{e}^{k[\phi(z)-z]} \quad [29]$$

where $z$ is defined as in [26] and $\phi$, which plays the role of an entropy, is a smooth function of $z$. Primes stand for derivatives with respect to $z$. Let us define $z_n$ as the point where

$$\phi_n' \equiv \phi'(z_n) = -n \quad [30]$$

which corresponds to the location of the maximum of $\phi + nz$. The entropy $\phi$ can be computed from the moments of the transition probability density. Using Laplace's method to expand the $n$th moment of $p_k$, we obtain

$$S_k(n) = \int_{-\infty}^{\infty} k\mathrm{e}^{k(n+1)z} p_k(v_k)\,\mathrm{d}z \approx \left(\frac{\phi_0''}{\phi_n''}\right)^{1/2} \mathrm{e}^{k(\phi_n + nz_n)} \quad [31]$$

from where, using [21],

$$\lambda_n \equiv \log S_w(n) = \phi(z_n) + nz_n \quad [32]$$

The essence of Laplace's approximation is that, for $k \gg 1$, most of the contribution to the integral in eqn [31] comes from the neighborhood of $z_n$, so that

it makes sense to consider each such neighborhood as a separate "component" of the cascade.

The geometric interpretation of this classification into components as a multifractal was developed in the context of three-dimensional homogeneous turbulence. We have up to now assumed very little about the nature of each cascade step, but it is natural in turbulence to interpret it as the process in which eddies decay to a smaller geometric scale. The argument works for any variable for which scale similarity can be invoked, but we have seen that most experiments are done for the magnitude of the velocity increments across a distance $r$. If we assume for simplicity that $r_k/r_{k+1} = \mathrm{e}$, so that $r_k/r_0 = \exp(-k)$, eqns [26] and [29] can be written as

$$v_k/v_0 = (r_k/r_0)^{-z_n}, \qquad p_k(z_n) \sim (r_k/r_0)^{-\phi_n} \qquad [33]$$

The multifractal interpretation is that the "component" indexed by $n$, whose velocity increments are "singular" in terms of $r$ with exponent $z_n$, lies on a fractal whose volume is proportional to its probability, and which therefore has a dimension $D(z_n) = 3 + \phi_n$.

Note that eqn [32] implies that the scaling exponents in eqn [17] can now be expressed as

$$\zeta(n) = -\log S_w(n) = -\lambda_n \qquad [34]$$

There was an enumeration there of several things which are equivalent: the exponents, the spectra, the distribution, and the limiting distribution $p_\infty(v)$ – univocally determine each other. Note however that different quantities have different scaling exponents. For example, it follows from eqn [6] that, if the scaling exponents for the local dissipation are $\zeta_\varepsilon(n)$, the exponents for $\Delta u$ would be $\zeta_{\Delta u}(n) = n/3 + \zeta_\varepsilon(n/3)$.

Some properties can be easily derived from the previous discussion. If we assume, for example, that the multiplicative factor $q$ is bounded above by $q_\mathrm{b}$, which is reasonable for many physical systems, eqn [26] implies that $z_n \leq \log q_\mathrm{b}$. In fact, if the transition probability behaves near $q_\mathrm{b}$ as $w(q) \sim (q_\mathrm{b} - q)^\beta$, the scaling exponents tend to

$$\lambda_n = n \log q_\mathrm{b} - (\beta + 1) \log n + \mathrm{O}(1) \qquad [35]$$

for $n \gg 1$. In the case in which $w(q)$ has a concentrated component at $q = q_\mathrm{b}$, the $\log n$ is missing in eqn [35]. In all cases, the singularity exponent of the set associated with $n \to \infty$ is $z_\infty = \log q_\mathrm{b}$, because the very high moments are dominated by the largest possible multiplier. In the case of a concentrated distribution the dimension of this set approaches a finite limit, but otherwise

$$D(n) \approx -(\beta + 1) \log n \qquad [36]$$

which becomes infinitely negative. This should not be considered a flaw. The set of events which only happen at isolated points and at isolated instants has dimension $D = -1$ in three-dimensional space, and those which only happen at isolated instants, and only under certain circumstances, have still lower negative dimensions. Sets with very negative dimensions are however extremely sparse, and are difficult to characterize experimentally.

The multifractal spectrum of the velocity differences in three-dimensional Navier–Stokes turbulence has been measured for several flows in terms of the scaling exponents, and appears to be universal. The probability distribution $w(q)$ of the multipliers has also been measured directly, and agrees well with the values implied by the exponents. It is also approximately independent of $r$, although not completely, perhaps due to the same experimental problems of anisotropy and limited Reynolds number which plague the measurement of the scaling exponents. There has been extensive theoretical work on the consequences of imposing various physical constraints on the multipliers, specially the conservation requirement that the average value of the dissipation has to be conserved across each cascade step. Several simple models have been proposed for the transition distribution which approximate the experimental exponents well, but the relation lacks specificity. Models that are very different give very similar results, and it is impossible to choose among them using the available data.

Multiplicative cascades and the resulting intermittency are not limited to Navier–Stokes turbulence. The equations of motion have only entered the discussion in this section through the assumption of scaling invariance. Multifractal models have in fact been proposed for many chaotic systems, from social sciences to economics, although the geometric interpretation is hard to justify in most of them. It is also important to realize that the fact that a given process can in principle be described as a cascade does not necessarily mean that such a description is a good one. Neither does a cascade imply a multiplicative process. For each particular case, we need to provide a dynamical mechanism that implements both the cascade and the transition multipliers. In three-dimensional Navier–Stokes turbulence, the basic transport of energy to smaller scales and to higher gradients is vortex stretching. The differential strengthening and weakening of the vorticity under axial stretching and compression also provide a natural way of introducing the self-similar transition probabilities of the local dissipation.

Examples of nonintermittent cascades abound. We have already mentioned that the vorticity in

decaying two-dimensional turbulence gets concentrated into stable vortex cores which eventually block the decay. The resulting enstrophy distribution is highly intermittent, but it is not well described by a multifractal. Conversely, forced two-dimensional turbulence is dominated by an inverse energy cascade to larger scales, which is not intermittent.

In addition, the intermittency of some systems is not a small-scale effect. Turbulent mixing of a passive scalar, which is the key process in turbulent heat transfer and in the atmospheric dispersion of pollutants, is an extremely intermittent phenomenon. The gradients of the scalar tend to be very localized, but they concentrate in sheets, narrow in thickness but otherwise extended. Some progress has recently been made on a simplified model due to Kraichnan for this problem, which is the linear stirring of a passive scalar by a random noise with delta correlation. Its statistics have been computed analytically, but the constraints of linearity and of uncorrelated forcing are strong, and the same methods do not appear to be extensible to mixing by real turbulence (*see* Lagrangian Dispersion (Passive Scalar)). Another problem in which intermittency is confined to large-scale surfaces is the motion of a three-dimensional pressureless gas, which has been used as a model for hypersonic turbulence and for the large-scale evolution of dark matter in the early universe.

In summary, intermittency is a fascinating property of many random systems, including three-dimensional Navier–Stokes turbulence, which interferes, sometimes strongly, with their description by simple cascade models. Significant advances have been made in its quantitative kinematic analysis. In some cases we also have a qualitative understanding of its roots. But in very few cases do we understand it well enough to make quantitative predictions.

*See also:* Ergodic Theory; Incompressible Euler Equations: Mathematical Theory; Lagrangian Dispersion (Passive Scalar); Turbulence Theories; Vortex Dynamics; Wavelets: Applications.

## Further Reading

Feller W (1971) *An Introduction to Probability Theory and Its Applications*, 2nd edn., vol. 2, pp. 169–172 and 574–581. New York: Wiley.

Frisch U (1995) *Turbulence. The Legacy of A.N. Kolmogorov*, pp. 159–192. Cambridge: Cambridge University Press.

Jiménez J (1998) Small scale intermittency in turbulence. *European Journal of Mechanics B: Fluids* 17: 405–419.

Jiménez J (2001) Intermittency and cascades. *Journal of Fluid Mechanics* 409: 99–120.

Lanford OE (1973) Entropy and equilibrium states in classical mechanics. In: Lenard A (ed.) *Statistical Mechanics and Mathematical Problems*, Lecture Notes in Physics, vol. 20, pp. 1–113. Berlin: Springer.

Nelkin M (1994) Universality and scaling in fully developed turbulence. *Advances in Physics* 43: 143–181.

Paladin G and Vulpiani A (1987) Anomalous scaling laws in multifractal objects. *Physics Reports* 156: 147–225.

Pope SB (2000) *Turbulent Flows*, pp. 167–173 and 254–263. Cambridge: Cambridge University Press.

Schroeder M (1991) *Fractals, Chaos, Power Laws*, Sect. 9. New York: W.H. Freeman.

Sreenivasan KR and Stolovitzky G (1995) Turbulent cascades. *Journal of Statistical Physics* 78: 311–333.

Vassilicos JC (ed.) (2001) *Intermittency in Turbulent Flows*. Cambridge: Cambridge University Press.

# Intersection Theory

**A Kresch**, University of Warwick, Coventry, UK



## Introduction

Intersection theory is the theory that governs the rigorous definition of intersections of cycles. This can take place in a variety of mathematical contexts, for instance, the intersections of two cycles on an oriented manifold in algebraic topology, of two currents on a differentiable manifold in differential geometry, or of two subvarieties on a nonsingular algebraic variety in algebraic geometry.

In algebraic geometry the theory is especially well developed (Fulton 1998). A cycle on an algebraic variety (or scheme) is a formal linear combination of irreducible closed subvarieties. These are subject to an equivalence relation called rational equivalence. For every rational function on every subvariety, its zero set is deemed rationally equivalent to its poles (with appropriate multiplicities).

As an example, in the complex projective plane $\mathbb{CP}^2$, any two lines are rationally equivalent since the ratio of two linear forms will vanish on one line and have a pole along the other. Similarly, a curve of degree $d$ is rationally equivalent to $d$ lines. Any two points in $\mathbb{CP}^2$ can be joined by a line (a copy of

$\mathbb{CP}^1$), and a rational function on $\mathbb{CP}^1$ can be chosen to vanish at one point and have a pole at the other. The groups of cycles modulo rational equivalence, known as Chow groups, are

$CH_2(\mathbb{CP}^2) \cong \mathbb{Z}$, generated by the fundamental class $[\mathbb{CP}^2]$

$CH_1(\mathbb{CP}^2) \cong \mathbb{Z}$, generated by the class of a line

$CH_0(\mathbb{CP}^2) \cong \mathbb{Z}$, generated by the class of a point

Two distinct lines $\ell_1$ and $\ell_2$ meeting at a point $p$ have this point as their intersection-theoretic product:

$$[\ell_1] \cdot [\ell_2] = [p] \qquad [1]$$

Intersection theory must also provide a self-intersection $[\ell_1] \cdot [\ell_1]$. Because $\ell_1$ and $\ell_2$ are rationally equivalent, this must also be the class of a point, but symmetry precludes the choice of a distinguished point on $\ell_1$. Instead, $[\ell_1] \cdot [\ell_1]$ is declared to be the rational equivalence class of a point on $\ell_1$, an element of $CH_0(\ell_1)$ rather than a specific cycle. This example illustrates that intersections cannot generally be defined on the level of cycles.

## Algebraic Intersection Products

### Refined Intersections

For a general nonsingular variety $X$, say of dimension $m$, if $U$ and $V$ are subvarieties of $X$ of respective dimensions $c$ and $d$, then there is a refined intersection product

$$[U] \cdot [V] \in CH_{c+d-m}(U \cap V) \qquad [2]$$

The traditional definition of the intersection product is based on two ideas. First, given two cycles that intersect properly, which by definition means that no component of their intersection has codimension less than the sum of the codimensions of the given cycles, the intersection product should be a formal sum of these components, each with a multiplicity that correctly reflects the geometry of the intersection. Second, given two arbitrary cycles, it should be possible to replace one of them by a rationally equivalent cycle which intersects the other properly.

While these ideas are simple, it took several decades for them to be carried out successfully. The case of curves on a surface meeting at a point was understood in the nineteenth century. Generalizing the classically understood canonical divisor class on a variety, work in the 1930s by Severi, Todd, and others showed that there are groups of equivalence classes of cycles in which canonical invariants of higher degrees can be defined (in modern language, higher Chern classes of the tangent bundle). Weil's foundations for algebraic geometry of the 1940s included a study of intersections of cycles. It was not until the 1950s that the notion of Chow groups was formalized and intersection theory was properly developed in this context. Chevalley, Chow, Samuel, Severi, and others contributed essential components of the theory. In an interesting parallel development, an intersection theory based on intersection multiplicities in algebraic topology was put forth by Alexander and Lefschetz in the 1920s, a decade before the introduction of the cup product in cohomology.

### Deformation to the Normal Cone

In the 1970s, Fulton and MacPherson established a construction of the intersection product in algebraic intersection theory that does not require moving cycles into general position. To accomplish this, they used an elegant geometric construction known as deformation to the normal cone.

Let $i : X \to Y$ be an embedding of codimension $d$ of nonsingular varieties. Let $V$ be a subvariety of $Y$ of dimension $k$ whose intersection with $X$ is of interest. We may view $X$ as the zero set of a section $s$ of some algebraic vector bundle $E$ on $Y$. By

$$(y, \lambda) \mapsto (\lambda^{-1} s(y), \lambda)$$

we have a map of the product of $Y$ with the punctured affine line, $Y \times (\mathbb{A}^1 \setminus \{0\})$, into $E \times \mathbb{A}^1$. We denote the closure of the image by $M^\circ_X Y$. An alternative, more intrinsic description is in terms of the blowup construction of algebraic geometry:

$$M^\circ_X Y = Bl_{X \times \{0\}}(Y \times \mathbb{A}^1)$$

Geometrically, $M^\circ_X Y$ has a copy of $Y$ over each $\lambda \neq 0$ and a copy of the normal bundle $N_X Y$ over $\lambda = 0$. This is the key construction that Fulton and MacPherson make use of. The same construction applied to $V$, that is, the closure of $V \times (\mathbb{A}^1 \setminus \{0\})$ in $M^\circ_X Y$, has over 0 a sort of singular normal bundle known as the normal cone

$$C_{X \cap V} V \subset N_X Y|_{X \cap V}$$

One of the properties of Chow groups is that they are unchanged upon pullback to the total space of a vector bundle (apart from the obvious dimension shift). The refined intersection of $V$ with $X$, denoted $i^![V]$, is defined to be the unique element of $CH_{k-d}(X \cap V)$ whose pullback to $N_X Y$ is equal to $[C_{X \cap V} V]$.

This single construction encompasses and interpolates between two extreme cases of intersections:

$$i^![V] = [X \cap V] \quad \text{when } X \text{ and } V \text{ meet transversely} \qquad [3]$$

$$i^![V] = c_d(N_X Y) \cap [V] \quad \text{when } V \subset X \qquad [4]$$

Equation [3] makes reference to transverse intersection, a notion that is stronger than proper intersection. In situations when it applies, for example, in eqn [1], it signifies that intersection operations behave as one might expect. Equation [4] includes the self-intersection formula which says that $[X] \cdot [X]$ is equal to the top Chern class of $N_X Y$.

With this construction, which is well documented in Fulton (1998), the general refined intersection in eqn [2] is obtained by reduction to the diagonal. Let $\Delta_X$ denote the diagonal inclusion $X \to X \times X$ of the nonsingular variety $X$. For subvarieties $U$ and $V$ of $X$, we define

$$[U] \cdot [V] = \Delta_X^![U \times V] \qquad [5]$$

Equation [5] makes the Chow groups of $X$ into a ring, the Chow ring $CH^*(X)$, which is graded by codimension by setting

$$CH^k(X) = CH_{m-k}(X)$$

## Links with Topology

### Cycle Map to Homology

For algebraic varieties over the complex numbers, there is a cycle map which links the Chow groups with a topological homology group. If $X$ is an algebraic variety over $\mathbb{C}$, then let $H_*(X)$ denote the Borel–Moore homology of $X$, that is, the homology of locally finite singular chains on $X$ (viewed as a topological space with the classical topology). If $X$ is embedded as a closed subset of an oriented differentiable manifold $M$, then there are identifications

$$H_i(X) \cong H^{n-i}(M, M \setminus X) \qquad [6]$$

where $n$ is the dimension of $M$. There is a cycle class map

$$CH_k(X) \to H_{2k}(X)$$

which sends the class of each irreducible subvariety $Z$ of dimension $k$ in $X$ to its fundamental class $[Z] \in H_{2k}(X)$.

Let $M$ be an oriented differentiable manifold of dimension $n$ and let $X$ and $Y$ be closed subsets of $M$. Then the cup product $H^i(M, M\setminus X) \otimes H^j(M, M\setminus Y) \to H^{i+j}(M, M\setminus(X \cap Y))$ induces, via eqn [6], an intersection product

$$H_i(X) \otimes H_j(Y) \to H_{i+j-n}(X \cap Y)$$

which is the topological analog of the refined intersection product of eqn [5]. The products are compatible via the cycle class map. The topology of complex algebraic varieties and the compatibilities between algebraic and topological intersections are discussed in Fulton (1998). An interesting application of this interplay of intersection theories is the convolution product in Borel–Moore homology, which is important in geometric representation theory (see Chriss and Ginzburg (1997)).

### Riemann–Roch Theorems

The classical Riemann–Roch theorem relates the dimensions of linear systems on an algebraic curve (algebraic quantities) with their degrees and the curve's genus (topological quantities). The Hirzebruch–Riemann–Roch theorem states that on a nonsingular projective variety $X$, if $E$ is an algebraic vector bundle on $X$ and $\chi(E)$ denotes its Euler characteristic (the alternating sum of the ranks of the sheaf-theoretic cohomology groups), then

$$\chi(E) = \int_X \mathrm{ch}(E) \cdot \mathrm{td}(T_X) \qquad [7]$$

where $\int_X$ denotes the degree of the zero-dimensional component of the quantity that follows, and the Chern character $\mathrm{ch}(E)$ and Todd class $\mathrm{td}(T_X)$ are certain standard universal polynomials of Chern classes.

Grothendieck had the inspired idea that eqn [7] could be generalized to a covariance property for the Chern character times the Todd class. If $X$ and $Y$ are nonsingular varieties and $f : X \to Y$ is a projective morphism (or, more generally, a proper morphism), then there is a well-defined push-forward $f_*$ on Chow groups. There is also a kind of push-forward for vector bundles. The Grothendieck group of vector bundles on $X$, denoted $K^0(X)$, is the group of formal linear combinations of vector bundles, modulo the relations $[E] = [E'] + [E'']$ whenever $E'$ is a sub-bundle of $E$ with quotient bundle $E''$. Every coherent sheaf $\mathcal{F}$ has a well-defined class in $K^0(X)$, namely, the alternating sum of $[E_i]$ where $E_\bullet$ is any finite resolution of $\mathcal{F}$ by vector bundles (locally free sheaves). The push-forward $f_*[E]$ is defined as the alternating sum of the classes in $K^0(Y)$ of the higher direct images $R^i f_* E$. The Grothendieck–Riemann–Roch theorem states that

$$\mathrm{ch}(f_*[E]) \cdot \mathrm{td}(T_Y) = f_*(\mathrm{ch}(E) \cdot \mathrm{td}(T_X)) \qquad [8]$$

in $CH_*(Y) \otimes \mathbb{Q}$. Notice that eqn [7] represents the case that $Y$ is a point.

There is an even more general formulation valid for singular varieties. It is necessary to work with a homology version of the Grothendieck group, namely, the Grothendieck group $K_0(X)$ of coherent sheaves on $X$. The Baum–Fulton–MacPherson version of the Grothendieck–Riemann–Roch theorem prescribes transformations

$$\tau_X : K_0(X) \to CH_*(X) \otimes \mathbb{Q} \qquad [9]$$

which are covariant for proper morphisms. When $X$ is nonsingular, $\tau_X$ is given by the "Chern character" times the "Todd class", and covariance becomes eqn [8].

In the case of varieties over the complex numbers, there is also a transformation from the algebraic Grothendieck group $K_0(X)$ to a topological analog, satisfying various compatibilities. The composition with the homology Chern character gives Riemann–Roch transformations $K_0(X) \to H_*(X; \mathbb{Q})$ satisfying properties akin to those of eqn [9].

### The Analytic Setting

The Atiyah–Singer index theorem stands as an important generalization of the Hirzebruch–Riemann–Roch theorem. The index of an elliptic differential operator on a differentiable manifold plays the role of the Euler characteristic, and is equated with a topological quantity. One of the consequences of the index theorem is the validity of eqn [7] for general compact complex manifolds.

More in the domain of pure analysis is the question of intersecting two currents on a differentiable manifold. Currents arise naturally out of Chern–Weil theory. To each current is associated a wave front, a subset of the cotangent bundle that reflects the geometry of the singular set of the current. A current can be pulled back to an embedded submanifold whenever the embedding is transverse to the wave front. By reduction to the diagonal, this gives an intersection of two currents with transverse wave fronts which reduces to the usual wedge product in the case of smooth differential forms (see Hörmander (1990)).

## Applications of Intersection Theory

### Enumerative Geometry

Intersection theory has proved to be a useful tool in diverse areas such as enumerative geometry, singularity theory, and moduli problems. Enumerative problems have intrigued generations of geometers. Chasles, Maillard, Schubert, and Zeuthen are among the geometers of the second half of the nineteenth century who solved an impressive array of problems, including, as a notable example, Steiner's five conics problem to determine the number of plane conics tangent to five given conics in general position.

In modern terms, the successful solution to an enumerative problem involves setting up a space which parametrizes the geometric objects being counted, suitably compactified, and carrying out an intersection-theoretic computation on this space. Steiner's problem illustrates how "excess intersection" can occur and cause difficulty. Inside the $\mathbb{CP}^5$ of plane conics, including degenerate conics, those tangent to a given conic constitute a sextic hypersurface. So $6^5 = 7776$ would appear plausible; this was, in fact, the originally proposed solution. However, the most degenerate conics, the double lines, all appear as limits of families of conics tangent to any given conic. The refined intersection of five of these sextics has a cycle class of degree 4512 supported on the Veronese surface of double lines. This leaves 3264, the correct answer given by Chasles in 1864. The issue of providing rigorous foundations for these kinds of calculations was recognized by Hilbert, who set it as the 15th of his 23 major mathematical problems outlined in 1900. A good survey of early and modern efforts in enumerative geometry can be found in Kleiman and Thorup (1987).

### Singularity Theory and Degeneracy Loci

In any situation where a geometric object is described by parameters, there will be values of the parameter at which the geometry changes qualitatively. The significance of this is evident in the space of conics above. Singularity theory is concerned with the loci in parameter spaces on which these transitions can occur. Let $\pi : Y \to P$ be a map of differential manifolds, or of nonsingular algebraic varieties, which is generally (but not everywhere) submersive, so that there are singular fibers. Let $d$ denote the dimension of $P$, which can be considered as a parameter space, and let $c$ be the dimension of $Y$. Consider the loci

$$\bar{S}_k(\pi) = \{y \in Y \mid \operatorname{rk}(T_{y,Y} \to T_{\pi(y),P}) \le d - k\}$$

of singularity theory. Thom made an influential study of these in the 1950s, and Porteous in 1971 gave the following formula, now called the Thom–Porteous formula:

$$[\bar{S}_k(\pi)] = s_{((k+c-d)^k)}(\pi^* T_P - T_Y) \qquad [10]$$

The symbol on the right is shorthand for $s_{(k+c-d,\dots,k+c-d)}$, the case $a_1 = \cdots = a_k = k + c - d$ of the Schur determinant $s_{(a_1,\dots,a_k)} = \det(s_{a_i+j-i})_{1 \le i,j \le k}$, and for vector bundles $E$ and $F$ the $s_i(F - E)$ are

defined by the formula $s(F-E)=\sum_i(-1)^i c_i(E)/\sum_i(-1)^i c_i(F)$. In algebraic intersection theory, eqn [10] has the precise meaning that when $\bar{S}_k(\pi)$ has the expected codimension $k(k+c-d)$ in $Y$ (or is empty), its cycle class is equal to the given polynomial in Chern classes. The Thom–Porteous formula applies to the degeneracy loci of arbitrary maps of vector bundles $E \to F$. Degeneracy loci constitute an active area of research in intersection theory, and there are generalizations, for example, to cases where there are more bundles or bundle maps with symmetry (see Fulton and Pragacz (1998)).

### Moduli Spaces

The parameter spaces that have appeared often admit interpretations as moduli spaces. Moduli problems start with geometric objects to be classified, and ask for families of these objects over an arbitrary base space to be represented as faithfully as possible by maps from the base space to some space called a moduli space. For enumerative applications it is most useful for the moduli space to be compact. One of the principal examples is the moduli of algebraic curves of given genus $g$: for $g \geq 2$, the moduli space of smooth curves $M_g$ has a compactification $\bar{M}_g$ by stable curves, as defined and studied by Deligne and Mumford. While the $M_g$ are singular, the singularities are mild enough to permit the definition of an intersection theory for $M_g$ and $\bar{M}_g$, as was done by Mumford in the 1980s. More generally, if $X$ is a complex projective variety, Kontsevich's spaces of stable maps $\bar{M}_{g,n}(X,\beta)$ compactify the moduli of genus $g$ curves with $n$ marked points together with algebraic maps to $X$ having image in homology class $\beta \in H_2(X)$. These spaces, and some high-powered intersection theory that takes place on them, are vitally important in Gromov–Witten theory. $K$-theory also provides an alternative approach to intersection products in algebraic geometry.

## Extensions and Related Theories

### Motives and Higher Chow Groups

Intersection theory has evolved into a mature theory with numerous extensions and offshoots. Many of these are a result of endeavors to forge links with other branches of mathematics. One of the extensions, higher Chow groups, has its roots in a basic property of intersection theory, the excision property, which states that if $X$ is a variety and $U \subset X$ an open subvariety, with $Z = X \setminus U$, then the inclusion and restriction maps fit into a right exact sequence

$$CH_*Z \to CH_*X \to CH_*U \to 0$$

This is reminiscent of the long exact homology sequence of a pair in algebraic topology. Indeed, there is a corresponding long exact sequence of Borel–Moore homology groups, but the elementary algebraic theory lacks such a long exact sequence. Bloch introduced higher Chow groups in the 1980s to fill this gap. The theory, which is quite complicated, provides groups $CH_*(X,j)$, with $CH_*(X,0)=CH_*X$, such that there is a long exact sequence

$$\cdots \to CH_*(U,j+1) \to CH_*(Z,j) \to CH_*(X,j) \to CH_*(U,j) \to \cdots$$

These groups are closely connected to algebraic $K$-theory and also to a related theory called motivic cohomology.

Motives, a sort of universal cohomology theory envisaged by Grothendieck, conjecturally form a category which can be extended to a bigger category of mixed motives that reflects mixed structures in cohomology, such as mixed Hodge structures. Recently, Voevodsky *et al.* (2000) have introduced motivic cohomology groups which form an integral part of a homotopy theory for algebraic varieties. Voevodsky's work, including a proof of the Milnor conjecture of $K$-theory, earned him a Fields Medal in 2002.

### Arithmetic Intersection Theory

There is an arithmetic version of intersection theory which applies to an arithmetic scheme $X$, which is, informally, a scheme defined over every prime field (all finite fields $\mathbb{F}_p$ and also $\mathbb{Q}$) in a consistent way. This means that $X$ can be base-extended to any field. In situations where the complex variety $X(\mathbb{C})$ is nonsingular, there is an arithmetic Chow ring $\widehat{CH}^*(X)$, introduced by Gillet and Soulé in 1990. Elements of $\widehat{CH}^*(X)$ are equivalence classes of pairs $(Z,g)$ where $Z$ is an algebraic cycle on $X$ and $g$ is known as a Green current for $Z$, a current on $X(\mathbb{C})$ satisfying the relation

$$\frac{\mathrm{i}}{2\pi}\partial\bar{\partial}g + \delta_{Z(\mathbb{C})} = \omega \qquad [11]$$

for some smooth differential form $\omega$ satisfying some conditions. Here, $\delta_Z(\mathbb{C})$ denotes the current of integration along $Z(\mathbb{C})$. The point to notice is that eqn [11] relates analysis (the Green current) and algebra (the cycle) on $X$ on one side with topology on the other, as $\omega$ will be a closed form whose class in de Rham cohomology is Poincaré dual to $[Z(\mathbb{C})]$.

Arithmetic intersection theory is used to define arithmetic height functions. Height functions have important applications to Diophantine problems, and were an essential component of the proof by Faltings of the Mordell conjecture, which earned him a Fields Medal in 1986. Arithmetic intersection theory grew

out of an earlier theory of Arakelov, in which $X(\mathbb{C})$ is endowed with a Kähler metric, and the form $\omega$ in eqn [11] is required to be harmonic. The Arakelov Chow group is only a ring when harmonic forms are closed under wedge product, which is not the case generally but which is true in some interesting cases, for example, for Grassmannian varieties. Arakelov treated the case of arithmetic surfaces, that is, the case when $X(\mathbb{C})$ is an algebraic curve ("surface" refers to a second dimension in the arithmetic direction), and introduced a pairing of arithmetic divisors, in analogy with the usual pairing of divisors on an algebraic surface. Arakelov's work, its subsequent generalizations, and more recent developments are covered in Faltings (1992).

### Equivariant Theories and Stacks

Moduli problems such as those mentioned previously are often best represented not by traditional varieties, but by a more sophisticated sort of object called a stack. Taking inspiration from Mumford's intersection theory on $M_g$, intersection theory on algebraic stacks has grown into a mature theory in its own right. Examples of stacks include orbifolds, for which there is the Chen–Ruan (orbifold) cohomology theory as well as an algebraic analog due to Abramovich, Graber, and Vistoli (see Abramovich, *et al.* (2002)). Another class of examples are quotient stacks of a variety by the action of an algebraic group. In these cases the Chow groups of the stack are equivariant Chow groups, part of a rich theory modeled on equivariant cohomology in algebraic topology. Behrend (2002) provides a nice survey of stacks, equivariant intersection theory, and their uses in Gromov–Witten theory. The Bott residue formula is an important tool in equivariant intersection theory which is particularly well suited to making concrete calculations, for example, in enumerative geometry. A description with nice examples can be found in Ellingsrud and Strømme (1996).

*See also:* Cohomology Theories; Hamiltonian Group Actions; Index Theorems; *K*-Theory; Moduli Spaces: An Introduction.

## Further Reading


Abramovich D, Graber T, and Vistoli A (2002) Algebraic orbifold quantum products. In: Adem A, Morava J, and Ruan Y (eds.) *Orbifolds in Mathematics and Physics (Madison 2001)*, Contemporary Mathematics vol. 310, pp. 1–24. Providence: American Mathematical Society.

Behrend K (2002) Localization and Gromov–Witten invariants. In: de Bartolomeis P, Dubrovin B, and Reina C (eds.) *Quantum Cohomology (Cetraro 1997)*, Lecture Notes in Mathematics vol. 1776, pp. 3–38. Berlin: Springer.

Chriss N and Ginzburg V (1997) *Representation Theory and Complex Geometry*. Boston: Birkhäuser.

Ellingsrud G and Strømme SA (1996) Bott's formula and enumerative geometry. *Journal of the American Mathematical Society* 9: 175–193.

Faltings G (1992) *Lectures on the Arithmetic Riemann–Roch Theorem*. Princeton: Princeton University Press.

Fulton W (1998) *Intersection Theory*, 2nd edn. Berlin: Springer.

Fulton W and Pragacz P (1998) *Schubert Varieties and Degeneracy Loci*. Berlin: Springer.

Hörmander L (1990) *The Analysis of Linear Partial Differential Operators*, 2nd edn, vol. 1. Berlin: Springer.

Kleiman SL and Thorup A (1987) Intersection theory and enumerative geometry: a decade in review. In: Bloch S. *et al.* (ed.) *Algebraic Geometry (Brunswick, Maine, 1985)*, Proceedings of Symposia in Pure Mathematics, part 2 vol. 46, pp. 321–370. Providence: American Mathematical Society.

Voevodsky V, Suslin A, and Friedlander EM (2000) *Cycles, Transfers, and Motivic Homology Theories*. Princeton: Princeton University Press.


# Inverse Problem in Classical Mechanics

**R G Novikov**, Université de Nantes, Nantes, France



## Formulation of the Problem

Consider the Newton equation

$$\ddot{x} = F(x), \quad F(x) = -\nabla v(x), \qquad x \in \mathbb{R}^d \qquad [1]$$

where

$$v \in C^2(\mathbb{R}^d, \mathbb{R})$$
$$|\partial_x^j v(x)| \le c_{|j|}(1 + |x|)^{-\alpha-|j|} \qquad [2]$$
$$\text{for } x \in \mathbb{R}^d, |j| \le 2, \text{and some } \alpha > 1, c_{|j|} \ge 0$$

(where $j$ is the multi-index $j \in (\mathbb{N} \cup \{0\})^d, |j| = \sum_{n=1}^d j_n$). In classical mechanics, eqn [1] describes the dynamics of a particle with the mass $m = 1$ in the force field $F$ with the potential $v$. For eqn [1] the energy $E = (1/2)(\dot{x}(t))^2 + v(x(t))$ is an integral of motion.

Under the assumptions [2], it follows that (Reed and Simon 1979): for any $(p_-, x_-) \in \mathbb{R}^{2d}, p_- \ne 0$, eqn [1] has a unique solution $x \in C^2(\mathbb{R}, \mathbb{R}^2)$ such that

$$x(t) = p_- t + x_- + y_-(t)$$
$$y_-(t) \to 0, \quad \dot{y}_-(t) \to 0, \qquad \text{as } t \to -\infty \qquad [3]$$

in addition, for almost any $(p_-, x_-)$

$$x(t) = a(p_-, x_-)t + b(p_-, x_-) + y_+(t)$$
$$a(p_-, x_-) \neq 0,\ y_+(t) \to 0,\ \dot{y}_+(t) \to 0 \quad [4]$$
$$\text{as } t \to +\infty$$

furthermore, the set $\mathcal{D}$ of all $(p_-, x_-) \in \mathbb{R}^{2d}, p_- \neq 0$, for which [4] holds for fixed $v$, is an open subset of $\mathbb{R}^{2d}$ and $\mathcal{M}es(\mathbb{R}^{2d}\backslash\mathcal{D}) = 0$.

We say that $a, b$ arising in [4] (and defined on $\mathcal{D}$) are the scattering data for eqn [1]. In addition, the scattering data $a, b$ at fixed energy $E > 0$ means $a, b$ on $\{(p_-, x_-) \in \mathcal{D} \,|\, p_-^2/2 = E\}$. Roughly speaking, for a particle moving according to [1], the functions $a, b$ relate the free motion at time $t \to -\infty$ with the free motion at time $t \to +\infty$.

Note that

$$a(p_-, x_- + t_0 p_-) = a(p_-, x_-)$$
$$b(p_-, x_- + t_0 p_-) = b(p_-, x_-) + t_0 a(p_-, x_-) \quad [5]$$
$$(p_-, x_-) \in \mathcal{D}, \quad t_0 \in \mathbb{R}$$

Formula [5] imply that $a, b$ on $\mathcal{D}$ are uniquely determined by $a, b$ on $\{(p_-, x_-) \in \mathcal{D} \,|\, p_- x_- = 0\}$, where $p_- x_-$ is the scalar product of $p_-$ and $x_-$.

If $v(x) \equiv 0$, then $a(p_-, x_-) = p_-, b(p_-, x_-) = x_-, (p_-, x_-) \in \mathbb{R}^d, p_- \neq 0$. Therefore, it is convenient to use for $a$, $b$ the following representation:

$$a(p_-, x_-) = p_- + a_{\text{sc}}(p_-, x_-)$$
$$b(p_-, x_-) = x_- + b_{\text{sc}}(p_-, x_-), \quad (p_-, x_-) \in \mathcal{D} \quad [6]$$

where the subscript sc is an abbreviation of the word "scattering."

The direct scattering problem for eqn [1], under the assumptions [2], consists in the following: given $v$, find $a, b$.

The inverse-scattering problem for eqn [1], under the assumptions [2], consists in the following: given $a, b$ (or some partial information about $a, b$), find $v$.

In the present article, we discuss, mainly, the aforementioned inverse-scattering problem.

## Abel's Result of 1826

Consider the Newton equation [1] in dimension $d = 1$ for $x \in ]-\infty, x_1], x_1 > 0$, where

$$v \in C^2(]-\infty, x_1], \mathbb{R})$$
$$v(x) = 0 \quad \text{for } x < 0 \quad [7]$$
$$\frac{dv(x)}{dx} > 0 \quad \text{for } 0 < x < x_1$$

Under the assumptions [7], for any $p_- > 0$, where $E = p_-^2/2 < v(x_1)$, eqn [1] has a unique solution $x \in C^2(\mathbb{R}, ]-\infty, x_1])$ such that

$$x(t) = p_- t \quad \text{for } t \leq 0 \quad [8]$$

in addition,

$$x(t) = -p_- t + b(p_-) \quad \text{as } t \to +\infty \quad [9]$$

Let

$$T(E) = \frac{b(\sqrt{2E})}{\sqrt{2E}}, \quad 0 < E < v(x_1),\ \sqrt{2E} > 0 \quad [10]$$

($T(E)$ is the time during which a particle starting at $x = 0$ with the impulse $p_- = \sqrt{2E}$ returns to $x = 0$).

Let $x(v)$, $v \in [0, v(x_1)]$, be the inverse function to $v(x)$, $x \in [0, x_1]$. Then (under the assumptions [7]),

$$T(E) = \sqrt{2}\int_0^E (E - v)^{-1/2}\frac{dx(v)}{dv}dv$$
$$0 < E < v(x_1) \quad [11]$$

$$x(v) = \frac{1}{\sqrt{2}\pi}\int_0^v (v - E)^{-1/2}T(E)dE$$
$$0 < v < v(x_1) \quad [12]$$

Actually, the formulas [11], [12] relating the travel time $T$ and the potential $v$ are the results from Abel (1826) (see also Keller (1976) for a discussion of this result). Formula [11] is a result on direct scattering, whereas [12] is a result on inverse scattering. In addition, if $T(E), 0 < E < v(x_1)$, is given, then [11] is the Abel integral equation for $x(v), 0 < v < v(x_1)$, and [12] solves this equation.

Concerning further results on inverse scattering for the one-dimensional Newton equation, see Keller (1976) and Astaburuaga *et al.* (1991). Note that for the one-dimensional case the scattering data $a$, $b$ do not in general determine $v$ uniquely.

The Abel integral equation and the Abel formula solving this equation were used also, in particular, by Firsov (1953) and Keller *et al.* (1956), where inverse scattering was considered for the three-dimensional Newton equation at fixed energy for the case of spherically symmetric monotonous decreasing potential in $|x|$.

Note also that the Abel method for solving the integral equation [11] was used by Radon (1917) for finding the inversion formula for the Radon transformation. In the next section, we reduce the inverse-scattering problem for the Newton equation [1] in dimension $d \geq 2$, under the assumptions [2], to the inversion problem for the X-ray transformation (i.e., the Radon transformation along straight lines).

## Inverse Scattering for the Multidimensional Newton Equation

Consider

$$T\mathbb{S}^{d-1} = \{(\theta, x) \in \mathbb{S}^{d-1} \times \mathbb{R}^d \mid \theta x = 0\} \qquad [13]$$

Consider the X-ray transformation $P$ defined by the formula

$$Pf(\theta, x) = \int_{\mathbb{R}} f(t\theta + x)\mathrm{d}t, \quad (\theta, x) \in T\mathbb{S}^{d-1} \qquad [14]$$

where

$$f \in C(\mathbb{R}^d, \mathbb{R}^m)$$
$$f(x) = O(|x|^{-\beta}) \quad \text{as } |x| \to \infty \text{ for some } \beta > 1 \qquad [15]$$

Consider the functions $a_{sc}, b_{sc}$ of [6]

**Theorem 1** (Novikov 1999). *For the Newton equation* [1], *under the assumptions* [2], *the following formulas hold*:

$$PF(\theta, x) = \lim_{s\to+\infty} s a_{sc}(s\theta, x), \quad (\theta, x) \in T\mathbb{S}^{d-1} \qquad [16]$$

$$Pv(\theta, x) = \lim_{s\to+\infty} s^2 \theta b_{sc}(s\theta, x), \quad (\theta, x) \in T\mathbb{S}^{d-1} \qquad [17]$$

*in addition*,

$$|PF(\theta, x) - s a_{sc}(s\theta, x)| \le \frac{d^3 c^2 2^{2\alpha+4}}{\alpha(\alpha-1)(1+|x|/\sqrt{2})^{2\alpha-1}} \frac{s^3}{(s/\sqrt{2}-1)^4} \qquad [18]$$

$$|Pv(\theta, x) - s^2\theta b_{sc}(s\theta, x)| \le \frac{d^3 c^2 2^{2\alpha+4}}{\alpha(\alpha-1)^2(1+|x|/\sqrt{2})^{2\alpha-2}} \frac{s^4}{(s/\sqrt{2}-1)^5} \qquad [19]$$

*for* $(\theta, x) \in T\mathbb{S}^{d-1}, s \ge z(d, c, \alpha, |x|)$, *where* $\theta b_{sc}$ *is the scalar product of* $\theta$ *and* $b_{sc}$, $z$ *is the root of the equation*

$$\frac{d^2 c 2^{\alpha+2}}{(\alpha-1)(1+|x|/\sqrt{2})^{\alpha-1}} \frac{z^2}{(z/\sqrt{2}-1)^3} = 1$$
$$z \in ]\sqrt{2}, +\infty[ \qquad [20]$$

$c = \max(c_1, c_2)$ (*and* $\alpha, c_1, c_2$ *are the constants of* [2]).

Theorem 1 gives a method for finding $PF$ and $Pv$ from $a_{sc}$ and $b_{sc}$ at high energies. It has been proved in Novikov (1999) by means of analysis of the following nonlinear integral equation for the function $y_-$ of [3]:

$$y_-(t) = A_{p_-, x_-}(y_-)(t)$$

where

$$A_{p_-, x_-}(u)(t) = \int_{-\infty}^{t} \int_{-\infty}^{\tau} F(p_- s + x_- + u(s))\mathrm{d}s\, \mathrm{d}\tau$$
$$p_- \neq 0$$

In dimension $d \ge 2$, Theorem 1 and methods for the reconstruction of $f$ from $Pf$ (Gelfand *et al.* 1980, Natterer 1986, Novikov 1999) give a method for the reconstruction of $F$ and $v$ from the scattering data $a$, $b$ at high energies. Note that for $d = 1$ Theorem 1 is valid but $f$ cannot be uniquely reconstructed from $Pf$.

Theorem 1 is an analog of the Born formula for the Schrödinger equation at high energies (see, e.g., Faddeev (1956), Enss and Weber (1995), and Novikov (1998) as regards this Born formula and its variations). On the other hand, Theorem 1 was preceded by a result of Gerver and Nadirashvili (1983) on the high-energy asymptotics for the travel time between boundary points for the Newton equation in a bounded strictly convex domain with smooth boundary. There is a considerable similarity between this result and Theorem 1.

We continue our review on inverse scattering for the multidimensional Newton equation, and make the following well-known observation.

**Observation 1** *Suppose that* $v(x) > E > 0$ *for* $x \in \mathcal{U}$, *where* $\mathcal{U}$ *is a compact subset of* $\mathbb{R}^d$. *Then the scattering data* $a, b$ *for energies smaller than or equal to* $E$ *contain no information about* $v(x)$ *for* $x \in \mathcal{U}$.

In addition to Theorem 1 and Observation 1, one has the following conjecture.

**Conjecture 1** (Novikov 1999). *Suppose that* $v$ *satisfies* [2], $d \ge 2$, *and the energy* $E$ *is sufficiently large*, $E > E(v)$. *Then the scattering data* $a, b$ *at fixed energy* $E$ *uniquely determine* $v$.

Gerver and Nadirashvili (1983) proved a result similar to Conjecture 1 for the case of the Newton equation in a bounded strictly convex domain $G$ with smooth boundary. Their proof of this result contains no reconstruction method but does contain a stability estimate. It is based on the Maupertuis principle and the results of Muhometov and Romanov (1978), Beylkin (1979), and Bernstein and Gerver (1980). For the case $v \in C^2(\mathbb{R}^d, \mathbb{R})$, $\operatorname{supp} v \subset G$ (where $G$ has the properties mentioned above), in Novikov (1999) a connection between the boundary-value data of Gerver and Nadirashvili (1983) and the scattering data $a, b$ is given and it is shown that for $d \ge 2$ the scattering data $a, b$ and the domain $G$ uniquely determine $v$ at fixed sufficiently large energy $E > E(v, G)$.

For more information concerning results mentioned above, see Novikov (1999) and Gerver and Nadirashvili (1983). One can see from the review of this section that very few results on inverse scattering for the multidimensional Newton equation are given in the literature, at present. It should

be remarked that the inverse-scattering theory in multidimensions is much more developed for the Schrödinger equation than for the Newton equation.

## Inverse Scattering for the Schrödinger Equation in Multidimensions

The inverse-scattering theory for the multidimensional Schrödinger equation has been developed by many authors (see, e.g., surveys given in Grinevich (2000) and Novikov (2001)).

Quantum-mechanical analogs of Theorem 1 appear, for example, in Faddeev (1956), Enss and Weder (1995), Novikov (1998) (see also references therein). Similarly, the quantum-mechanical analogs of Conjecture 1 have been proved, for example, in Novikov (1992, 1994) and Grinevich and Novikov (1995) (see also references therein). On the other hand, as a rule, classical-mechanical analogs of results of the works on inverse Schrödinger scattering in multidimensions are unknown. This leads to many open problems. For the one-dimensional case some results on finding classical limits of results on inverse Schrödinger scattering are given in Lax and Levermore (1983) and Bogdanov (1985). Note that inverse scattering for the two-dimensional Schrödinger equation at fixed energy (see Novikov (1992), Grinevich and Novikov (1995), and Grinevich (2000) and references therein) has considerable similarity with inverse scattering for the one-dimensional Schrödinger equation. Therefore, an interesting open problem consists in extending the aforementioned study of Lax and Levermore (1983) and Bogdanov (1985) to the case of inverse scattering for the two-dimensional Schrödinger equation at fixed energy. Perhaps, in this way one can find proper two-dimensional analogs of the Abel formulas [11] and [12].

## Further Reading


Abel NH (1826) Auflösung einer mechanis chen Aufgabe. *J. Reine Angew. Math.* 1: 153–157 (German) (French translation: Résolution d'un problème de mécanique. In: Sylow L and Lie S (eds.) *Oeuvres complètes de Niels Henrik Abel*, vol. 1, pp. 97–101. Grondahl: Christiania (Oslo), (1881)).

Astaburuaga MA, Fernandez C, and Cortés VH (1991) The direct and inverse problem in Newtonian scattering. *Proceedings of the Royal Society of Edinburgh Section A* 118: 119–131.

Bernstein IN and Gerver ML (1980) A condition of distinguishability of metrics by hodographs. *Computational Seismology* 13: 50–73 (Russian).

Beylkin G (1979) Stability and uniqueness of the solution of the inverse kinematic problem of seismology in higher dimensions. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.* (LOMI) 84: 3–6 (Russian) (English translation *Journal of Soviet Mathematics* 21: 251–254 (1983)).

Bogdanov IV (1985) Classical limit of the quantum inverse scattering problem. *Teor. Mat. Fiz.* 65(1): 35–43 (Russian) (English translation. *Theoretical and Mathematical Physics* 65: 992–998 (1985)).

Enss V and Weder R (1995) Inverse potential scattering: A geometrical approach. In: Feldman J, Froese R, and Rosen L (eds.) *Mathematical Quantum Theory II: Schrödinger Operators*, CRM Proc. Lecture Notes, vol. 8, pp. 151–162. Providence, RI: American Mathematical Society.

Faddeev LD (1956) Uniqueness of solution of the inverse scattering problem. *Vestnik Leningrad Univ.* 11(7): 126–130 (Russian).

Firsov OB (1953) Determination of the force acting between atoms via differential effective elastic cross section. *Zh. Eksper. Teoret. Fiz.* 24: 279–283 (Russian).

Gelfand IM, Gindikin SG, and Graev MI (1980) Integral geometry in affine and projective spaces. Itogi Nauki i Tekhniki. *Sov. Prob. Mat.* 16: 53–226 (Russian) (English translation. *Journal of Soviet Mathematics* 18: 39–167 (1980)).

Gerver ML and Nadirashvili NS (1983) Inverse problem of mechanics at high energies. *Computational Seismology* 15: 118–125 (Russian).

Grinevich PG (2000) Scattering transform for the two-dimensional Schrödinger operator with decreasing at infinity potential at fixed non-zero energy. *Usp. Math. Nauk* 55(6): 3–70 (Russian) (English translation. *Russian Mathematical Surveys* 55: 1015–1083).

Grinevich PG and Novikov RG (1995) Transparent potentials at fixed energy in dimension two. Fixed-energy dispersion relations for the fast decaying potentials. *Communications in Mathematical Physics* 174: 409–446.

Keller JB (1976) Inverse problems. *The American Mathematical Monthly* 83: 107–118.

Keller JB, Kay I, and Shmoys J (1956) Determination of the potential from scattering data. *Physical Review* 102: 557–559.

Lax PD and Levermore CD (1983) The small dispersion limit of the Korteweg–de Vries equation. I–III. *Communications in Pure and Applied Mathematics* 36: 253–290, 571–593, 809–830.

Muhometov RG and Romanov VG (1978) On the problem of finding an isotropic Riemannian metric in an *n*-dimensional space. *Dokl. Akad. Nauk SSSR* 243(1): 41–44 (Russian) (English translation *Soviet Mathematics Doklady*. 19: 1330–1333).

Natterer F (1986) *The Mathematics of Computerized Tomography*. Stuttgart: Teubner.

Novikov RG (1992) The inverse scattering problem on a fixed energy level for the two-dimensional Schrödinger operator. *Journal of Functional Analysis* 103: 409–463.

Novikov RG (1994) The inverse scattering problem at fixed energy for the three-dimensional Schrödinger equation with an exponentially decreasing potential. *Communications in Mathematical Physics* 161: 569–595.

Novikov RG (1998) On inverse scattering for the *N*-body Schrödinger equation. *Journal of Functional Analysis* 159: 492–536.

Novikov RG (1999) Small angle scattering and X-ray transform in classical mechanics. *Arkiv för Matematik* 37: 141–169.

Novikov RG (2001) Scattering for the Schrödinger equation in multidimensions. Non-linear $\bar{\partial}$-equation, characterization of scattering data and related results. In: Pike ER and Sabatier P (eds.) *Scattering*, ch. 6.2.4. New York: Academic Press.

Radon J (1917) Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Ber. Verh. Sächs. Akad. Wiss. Leipzig, Math.-Nat. K1* 69: 262–267.

Reed M and Simon B (1979) *Methods of Modern Mathematical Physics. III. Scattering Theory*. New York: Academic Press.

**Inverse Problems in Wave Propagation** *see* Boundary Control Method and Inverse Problems of Wave Propagation

# Inviscid Flows

**R Robert**, Université Joseph Fourier, Saint Martin D'Hères, France



## Introduction

The equations governing the motion of an ideal (inviscid) fluid were derived by Euler in 1755. They were, together with the equation of vibrating strings, the first partial differential equations introduced in the field of mathematical physics. While several partial differential equations, coming from the modeling of physical phenomena, have had a satisfactory mathematical solution, it is piquant to note that the old Euler equations remain essentially unsolved. Together with the Navier–Stokes equations of viscous fluids, the Euler equations play a central role in the modern analysis of partial differential equations.

The mathematical difficulties encountered in the study of Euler equations seem to be deeply linked with the understanding of turbulence, which remains one of the great open problems in the field of macroscopic physics.

The relevance of Euler equations as a model of fluid flow is rather subtle, and the discussion is far from closed. On the one hand, Euler equations have disturbing aspects, which, in their most visible form, yield paradoxes. On the other hand, the systematic recourse to some viscosity seems to put a serious obstacle to a proper understanding of turbulence. In this article we will try to give some insight into this issue.

To be rigorous, every fluid has some compressibility, that is to say the density varies with the pressure. Compressibility gives rise to pressure waves, which propagate in the fluid with some finite speed. When the velocity of the fluid particles is slow relative to the speed of the pressure waves, it is legitimate to make the approximation that the flow is incompressible; it is the case for meteorological flows, for example. Then, there are no more pressure waves; nevertheless the motion can be very unstable and intricate (turbulent). Although very often in physical flows these two features coexist, following the tradition, we clearly separate the compressible and incompressible cases.

## The Equations of the Perfect Fluid

Until now a rigorous derivation of the fluid equations from a system of interacting particles governed by Newton's laws is not known. Thus, the mathematical models of fluid motion result from heuristic considerations.

Let us specify some notations.

The fluid motion is supposed to take place in some domain (not necessarily bounded) $\Omega$ of the physical space $\Re^3$.

We shall use the so-called Eulerian description of the fluid motion: $\rho(t,\boldsymbol{x})$ denotes the local density of the fluid at time $t$ and position $\boldsymbol{x}$, and $\boldsymbol{u}(t,\boldsymbol{x})$ the velocity of the fluid particle located at $\boldsymbol{x}$ at time $t$.

The first equation (conservation equation) expresses the conservation of mass:

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho \boldsymbol{u}) = 0 \qquad [1]$$

The second equation (momentum equation) expresses Newton's law (in the absence of internal friction):

$$\rho\left(\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u}\right) = -\nabla p \qquad [2]$$

where the scalar function p($t$, $\boldsymbol{x}$) is the pressure inside the fluid, and

$$(\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = \sum_i u_i \partial_i \boldsymbol{u}$$

With [1] and [2], we have five scalar unknown functions $(\rho, u_i, p)$ and only four equations. To get a

closed set of equations, we need to add a supplementary relationship:

$$\operatorname{div}(\boldsymbol{u}) = 0, \quad \text{for the incompressible flows} \qquad [3]$$

In the case of compressible flows, eqns [1] and [2] must be completed by a thermodynamical description of the fluid, which yields a relationship between $\rho, p$, the internal energy, the specific entropy, etc. We will only consider here the simple case of an isentropic gas which is modeled by the relationship

$$p = p(\rho) \qquad [4]$$

with $p(\rho) = c\,p^{\gamma}$ for a perfect gas ($c > 0, \gamma > 1$).

#### Condition at the Boundary $\partial\Omega$ of the Domain

In the case of a perfect fluid, we simply have to write that the velocities of the fluid particles at the boundary are tangent to the boundary, that is,

$$\boldsymbol{u} \cdot \boldsymbol{n} = 0 \quad \text{on } \partial\Omega \qquad [5]$$

where $\boldsymbol{n}$ denotes the unit normal vector to the boundary (pointing outward).

### The Incompressible Perfect Fluid: Main Properties of Smooth Flows

We shall suppose $\rho = 1$. Equations [1]–[3] and [5] then yield the classical Euler system:

$$\begin{aligned} &\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = -\nabla p \quad \text{on } \Omega \\ &\operatorname{div} \boldsymbol{u} = 0, \quad \boldsymbol{u} \cdot \boldsymbol{n} = 0 \text{ on } \partial\Omega \end{aligned} \qquad [6]$$

#### The Constants of the Motion

Let us examine the constants of the motion of the dynamical system defined by [6], that is, the functionals which are conserved by the motion of the fluid.

First we have the classical constants of motion associated with the natural symmetries by Noether's theorem.

The time translational invariance of the system implies that the kinetic energy is conserved:

$$E_c = \frac{1}{2}\int_{\Omega} \boldsymbol{u}^2 \,\mathrm{d}\boldsymbol{x}$$

In the case $\Omega = \Re^3$, the homogeneity of space implies the conservation of the impulsion:

$$\int_{\Omega} \boldsymbol{u} \,\mathrm{d}\boldsymbol{x}$$

The space isotropy, on the other hand, yields the conservation of the angular momentum:

$$\int_{\Omega} \boldsymbol{x} \wedge \boldsymbol{u} \,\mathrm{d}\boldsymbol{x}$$

There is a more hidden constant of the motion, called helicity, which was discovered in 1961 by J J Moreau (1961) (see, e.g., Serre (1979)).

Let us define the vorticity of the flow:

$$\omega = \operatorname{curl} \boldsymbol{u}$$

then the helicity is

$$\int_{\Omega} \omega \cdot \boldsymbol{u} \,\mathrm{d}\boldsymbol{x}$$

Of course, here, we suppose $\boldsymbol{u}$ to be vanishing at infinity in such a manner that the above integrals make sense.

One may wonder about the existence of other constants of the motion of the form (first-order functionals):

$$\int F(\boldsymbol{x}, \boldsymbol{u}(\boldsymbol{x}), \nabla \boldsymbol{u}(\boldsymbol{x}))\mathrm{d}\boldsymbol{x}$$

The answer, due to Serre (1979), is that any functional of the above form which is conserved by the flow is a linear function of the energy, the impulsion, the angular momentum, the helicity plus a trivial term (i.e., taking the same value for any field $\boldsymbol{u}$ such that $\operatorname{div} \boldsymbol{u} = 0$).

#### Beltrami Equation and Kelvin's Theorem

Another important issue is to know how the vorticity field evolves in a regular flow. If we apply the operator curl to the equation [6] in order to eliminate the pressure term, we get:

$$\frac{\partial \omega}{\partial t} + (\boldsymbol{u} \cdot \nabla)\omega - (\omega \cdot \nabla)\boldsymbol{u} = 0 \qquad [7]$$

which is the Beltrami equation.

To exploit the Beltrami equation, we need the Lagrangian flow $\varphi(t, \boldsymbol{x})$, associated with the field $\boldsymbol{u}$, which is defined by the differential equation:

$$\frac{\partial \varphi}{\partial t}(t, \boldsymbol{x}) = \boldsymbol{u}(t, \varphi(t, \boldsymbol{x})), \quad \varphi(0, \boldsymbol{x}) = \boldsymbol{x}$$

Then we can state the following proposition.

**Proposition** *During the smooth motion of an incompressible perfect fluid, we have:*

$$\omega(t, \varphi(t, \boldsymbol{x})) = D\varphi(t, \boldsymbol{x})[\omega(0, \boldsymbol{x})], \quad \text{for all } t, \boldsymbol{x}$$

*where $D\varphi(t, \boldsymbol{x})$ denotes the derivative at the point $\boldsymbol{x}$ ($t$ fixed) of the mapping $\boldsymbol{x} \to \varphi(t, \boldsymbol{x})$.*

The first consequence of this result is to point out the class of irrotational flows, for which

$\omega(t, \boldsymbol{x}) = 0$. Indeed, if the vorticity vanishes initially, it follows from the proposition that it will vanish for ever.

Another consequence is the behavior of vortex lines. By definition, a vortex line is any integral curve of the vorticity field. More precisely, a vorticity line at time $t$, $C(s)$ is defined by the differential equation

$$\frac{\mathrm{d}C}{\mathrm{d}s}(s) = \omega(t, C(s))$$

Now we can check that vortex lines are merely transported by the flow: if $C(s)$ is a vortex line at time $t = 0, \varphi(t, C(s))$ is a vortex line at time $t$.

We end this section with the famous Kelvin's circulation theorem (1869) (see, e.g., Marchioro and Pulvirenti (1994)).

**Theorem** *Let L be a closed (oriented) contour drawn inside the fluid. We suppose that L is transported by the flow; $\varphi_t(L)$ denotes the contour at time t. Then the circulation of the velocity field $\boldsymbol{u}(t, \boldsymbol{x})$ along $\varphi_t(L)$ is independent of t.*

### Stationary Solutions: D'Alembert's Paradox

Let us focus now on the flow around a bounded body $\overline{\Omega}$, whose complement $\overline{\Omega}^c$ will be supposed to be simply connected.

A stationary solution $u(\boldsymbol{x}), p(\boldsymbol{x})$ satisfies:

$$(\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = -\nabla p$$
$$\operatorname{div} \boldsymbol{u} = 0, \quad \boldsymbol{u} \cdot \boldsymbol{n} = 0 \quad \text{on } \partial\Omega$$

But since $(\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = \nabla(\frac{1}{2}\boldsymbol{u}^2) + (\operatorname{curl} \boldsymbol{u}) \wedge \boldsymbol{u}$, any stationary field $\boldsymbol{u}(\boldsymbol{x})$ satisfying curl $\boldsymbol{u} = 0$, div $\boldsymbol{u} = 0$, $\boldsymbol{u} \cdot \boldsymbol{n} = 0$ on $\partial\Omega$, defines a stationary solution with associated pressure $p = -\frac{1}{2}\boldsymbol{u}^2$.

We also need to specify a condition at infinity for the field $\boldsymbol{u}$. We impose that the velocity is equal (at infinity) to some constant value $U$. Since $\overline{\Omega}^c$ is simply connected, the condition curl $\boldsymbol{u} = \boldsymbol{0}$ implies that the flow is potential, that is, there is a scalar function $F(\boldsymbol{x})$ such that $\boldsymbol{u} = U + \nabla F$.

Thus, the determination of an irrotational flow around an obstacle amounts to solving the following exterior Neuman problem.

Find $F$ satisfying:

$$\Delta F = 0 \quad \text{in } \overline{\Omega}^c$$
$$\frac{\partial F}{\partial n} = -\mathrm{U} \cdot \boldsymbol{n} \quad \text{on } \partial\Omega$$
$$\nabla F = 0 \quad \text{at infinity}$$

This problem is well known and has a unique solution, which satisfies, at infinity:

$$F(\boldsymbol{x}) = \mathrm{O}(1/|\boldsymbol{x}|^2) \qquad \nabla F(\boldsymbol{x}) = \mathrm{O}(1/|\boldsymbol{x}|^3)$$

Then a classical calculation (integration by parts) gives the resulting force exerted by the flow on the body:

$$R = -\int_{\partial\Omega} \mathrm{p}\boldsymbol{n} \,\mathrm{d}\sigma = \int_{\partial\Omega} \frac{1}{2}\boldsymbol{u}^2\boldsymbol{n} \,\mathrm{d}\sigma = 0$$

This property of inviscid potential flows was first noticed by Jean Le Rond d'Alembert (1717–1783). Furthermore, d'Alembert performed a series of experiments to measure the drag on a sphere in a flowing fluid and he expected that the force would go to zero as the viscosity of the fluid approached zero. But this was not the case: the drag seemed to converge toward a nonzero value. Hence, this property was called d'Alembert's paradox.

Of course, d'Alembert's paradox tells us that something is going wrong: this model of flow around a body is not physically relevant. But it is not obvious to identify precisely what is going wrong.

Physics tells us that in a flow around a flying airplane, the viscous term (as measured by an dimensionless number called Reynolds number) is very small. The main effect of the viscosity is then to alter the limit condition at the boundary of the body. The relevant boundary condition is no longer $\boldsymbol{u} \cdot \boldsymbol{n} = 0$, but the purely viscous condition $\boldsymbol{u} = 0$, or more realistically a condition of friction type (turbulent boundary condition).

A common approach is to disqualify the perfect-fluid model in arguing that this modification of the boundary condition has important consequences on the flow near the body (giving rise to a turbulent boundary layer, for example).

It seems to us that such a disqualification of the perfect-fluid model discards prematurely interesting issues. Indeed, we must notice first that the stationary solution on which d'Alembert's reasoning is based is highly unstable and not acceptable physically. Thus, a realistic solution would necessarily be either nonstationary or with some vorticity. On this basis, we can imagine other scenarios to explain the existence of a resulting force exerted on the body. For example, we may imagine a stationary solution with a discontinuous velocity field (i.e., with a vortex sheet). The process conducive to such a stationary solution is called Prandtl's scenario (Batchelor 1967). The mathematical proof that Prandtl's scenario does exist is a difficult (open) issue, which seems closely related to the (probable) nonuniqueness of weak solutions of the Cauchy problem.

## The Cauchy Problem for the Incompressible Perfect Fluid

### The Case $\Omega \subset \Re^3$

In the Cauchy problem, given an initial velocity field $\boldsymbol{u}_0(\boldsymbol{x})$, we want to determine the corresponding solution $\boldsymbol{u}(t,\boldsymbol{x})$ of [6] at each time $t$.

The first significant result on the Cauchy problem for three-dimensional Euler equations was given by Kato (1975).

**Theorem** *For $\boldsymbol{u}_0$ in the Sobolev space $H^s(\Re^3)$, for $s > 5/2$, there is $T > 0$ and a unique classical solution (of the Cauchy problem) $\boldsymbol{u}(t,\boldsymbol{x})$ on $[0,T] \times \Re^3$. $\boldsymbol{u}$ depends continuously on t in the space $H^s$.*

By a classical solution we mean that the field $\boldsymbol{u}(t,\boldsymbol{x})$ is derivable in terms of the variables $t$, $\boldsymbol{x}$ and satisfies the equations in the usual sense.

Here $H^S(\Re^3)$ denotes the Sobolev space of the fields $\boldsymbol{u}$, which are square integrable and with spatial derivatives of order $s$ (in the case where $s$ is an integer) also square integrable.

**Remark** These results have been generalized to some extent during the last few decades, but the following issues are still open:

1. Do singularities occur at a finite time for such regular solutions?
2. For a less regular initial datum, do weak solutions exist (in the sense of distributions)?

### The Case $\Omega \subset \Re^2$

This case is better understood, the first mathematical results trace back to Lichtenstein (1925) and Wolibner (1933); they take a plain form with the famous theorem of Youdovitch 1963 (see, e.g., Chemin (1995)).

In two dimensions, the vorticity $\omega = \operatorname{curl} \boldsymbol{u}$ identifies with a scalar function, and the Beltrami equation becomes

$$\frac{\partial \omega}{\partial t} + \operatorname{div}(\omega \boldsymbol{u}) = 0 \quad [8a]$$

$$\operatorname{curl} \boldsymbol{u} = \omega \quad [8b]$$

$$\operatorname{div} \boldsymbol{u} = 0, \quad \boldsymbol{u} \cdot \boldsymbol{n} = 0 \quad \text{on } \partial\Omega \quad [8c]$$

This formulation, which appears as a transport equation [8a] for $\omega$, coupled with the elliptic system [8b]–[8c], which determines $\boldsymbol{u}$ from $\omega$, is particularly convenient.

The constants of motion associated with the usual symmetries, of course, persist; notice, however, that the helicity degenerates since, in two dimensions, $\omega \cdot \boldsymbol{u} = 0$. But now from [8a] we see that $\omega$ is merely convected by the incompressible velocity field $\boldsymbol{u}$. We deduce that, for any continuous function $f$, the functional

$$\int_\Omega f(\omega(t,\boldsymbol{x}))\mathrm{d}\boldsymbol{x}$$

is a constant of motion.

Thus, a specific feature of the two-dimensional case is to introduce an infinite set of constants of motion. By a skilful exploitation of this fact, Youdovitch succeeded in proving the following result.

**Theorem** *For a given $\omega_0$ in the space $L^\infty(\Omega)$, there is a unique weak solution $\omega(t,\boldsymbol{x})$ of [8], such that $\omega(t,\boldsymbol{x})$ is in $L^\infty(\Omega)$ for all $t$, and $\omega$ depends continuously on t in the space $L^p, 1 \le p < \infty$.*

$L^p$ denotes, in a standard way, the Lebesgue space of the functions $f$ such that $|f|^p$ is integrable over $\Omega$ and $L^\infty(\Omega)$, the space of measurable bounded functions on $\Omega$.

Thus, if we limit ourselves to initial data with bounded scalar vorticity, the Cauchy problem for the two-dimensional incompressible perfect fluid is satisfactorily solved. The situation is much more intricate if we consider a less regular initial datum (e.g., if $\omega_0$ is a measure supported by a curve (vortex sheet)).

#### Arnol'd's Work on Two-Dimensional Inviscid Flows

Youdovitch's theorem implies that the incompressible Euler equations, with $\omega_0$ in $L^\infty(\Omega)$, is a satisfactory model of two-dimensional flows – an important issue to study further the properties of this model.

A famous result due to Arnol'd (see Arnol'd and Khesin (1998) and Marchioro and Pulvirenti (1994)) deals with the nonlinear stability of the stationary solutions.

Let us determine the smooth stationary solutions of the two-dimensional Euler equations in a bounded domain $\Omega$ of the plane. We have to solve:

$$(\boldsymbol{u} \cdot \nabla)\omega = \boldsymbol{0} \quad [9a]$$

$$\operatorname{curl} \boldsymbol{u} = \omega \quad [9b]$$

$$\operatorname{div} \boldsymbol{u} = 0, \quad \boldsymbol{u} \cdot \boldsymbol{n} = 0 \quad \text{on } \partial\Omega \quad [9c]$$

Since we have $\operatorname{div} \boldsymbol{u} = 0$, we may introduce the stream function of $\boldsymbol{u}$, $\psi$, which is given by the Dirichlet's problem:

$$-\Delta\psi = \omega, \quad \psi = 0 \text{ on } \partial\Omega$$

so that $\boldsymbol{u} = \operatorname{curl} \psi$.

The system [9] becomes:

$$\nabla\psi \wedge \nabla\omega = 0, \quad -\Delta\psi = \omega, \quad \psi = 0 \text{ on } \partial\Omega$$

Let us focus on solutions which are characterized by a relationship $\omega = f(\psi)$, where $f$ is a smooth function. Such solutions are given by the resolution of the following nonlinear elliptic problem:

$$-\Delta\psi = f(\psi), \quad \psi = 0 \text{ on } \partial\Omega \qquad [10]$$

This problem has always at least a solution, for example, if $f$ is a bounded function of $\psi$.

Let $\psi^*$ be a solution of [10], and $\omega^* = f(\psi^*)$ the corresponding vorticity function. We shall say that the stationary solution $\omega^*$ is stable in the $L^2$-norm if:

For all $\varepsilon > 0$, there is a $\delta > 0$, such that for all initial datum $\omega_0$ in $L^\infty(\Omega)$ satisfying

$$\int_\Omega (\omega^* - \omega_0)^2 \mathrm{d}\boldsymbol{x} \le \delta, \text{ we have :}$$
$$\int_\Omega (\omega^* - \omega(t))^2 \,\mathrm{d}\boldsymbol{x} \le \varepsilon, \text{for all } t$$

where $\omega(t)$ denotes the solution of the Cauchy problem associated with the initial datum $\omega_0$ by Youdovitch's theorem.

Now we can state the following result.

**Theorem** (Arnol'd) *Let $\omega$ be a stationary solution given by [10]. We assume that one of the following assumptions holds*:

(C1) *There are positive constants $c_1, c_2$, such that*

$$c_1 \le f' \le c_2$$

(C2) *There are positive constants $c_1, c_2$, with $c_2 < \lambda_1$ (first eigenvalue of the Dirichlet problem on the domain $\Omega$) such that*:

$$c_1 \le -f' \le c_2$$

*Then $\omega$ is stable in the $L_2$-norm.*

**Remarks**

(i) This result was the first nonlinear stability result for stationary flows.
(ii) The proof makes use of the conservation of the functionals of the vorticity field.

Another significant contribution of Arnol'd to hydrodynamics was to reveal the geometrical aspect of the instability of the perfect-fluid motion. We give a brief insight into this issue.

Let us come back to the Lagrangian description of motion. We want to determine the function $\varphi(t, \boldsymbol{x})$. Each mapping $\varphi_t(\boldsymbol{x}) = \varphi(t, \boldsymbol{x})$ is, for $t$ fixed, a diffeomorphism of $\bar{\Omega}$ preserving the Lebesgue measure and the orientation (equivalently stated, it is an element of $\mathrm{SDiff}(\bar{\Omega})$).

In other words, a fluid motion is a curve $t \to \varphi_t$ drawn on the "manifold" $M = \mathrm{SDiff}(\bar{\Omega})$ (the configuration space of the system).

At time $t$, the relationship

$$\frac{\partial\varphi}{\partial t}(t, \boldsymbol{x}) = \boldsymbol{u}(t, \varphi(t, \boldsymbol{x}))$$

states that the velocity field $\boldsymbol{u}(t, \varphi_t(\boldsymbol{x}))$ belongs to the space tangent to $M$ at $\varphi_t$. The tangent space at $\varphi$ to $M$ is the space of vector fields $\boldsymbol{v}(\varphi(\boldsymbol{x}))$, where $\boldsymbol{v}(\boldsymbol{x})$ is an incompressible vector field on $\bar{\Omega}$ satisfying $\boldsymbol{v}\cdot\boldsymbol{n} = 0$ on $\partial\Omega$. This space is naturally endowed with a norm given by the kinetic energy

$$\frac{1}{2}\int_\Omega \boldsymbol{v}(\boldsymbol{x})^2 \mathrm{d}\boldsymbol{x}$$

and thus $M$ is endowed with a Riemannian structure.

It is easy to check that the perfect-fluid motions correspond to the curves $\varphi_t$ drawn on $M$ which are the critical points of the action integral:

$$\frac{1}{2}\int_{t_1}^{t_2} \mathrm{d}t \int_\Omega \left|\frac{\partial\varphi}{\partial t}(t, \boldsymbol{x})\right|^2 \mathrm{d}\boldsymbol{x}, \quad \text{for all } t_1 < t_2$$
$$\text{(with the constraints } \varphi(t_1, .) = \varphi_1, \varphi(t_2, .) = \varphi_2)$$

That is to say, the perfect-fluid motions are the geodesics of the Riemannian manifold $M$.

The main interest of this geometric framework is to bring back, at least formally, the perfect-fluid motions to well-known objects. Indeed, we know that the Riemannian curvature of a manifold has a profound impact on the behavior of geodesics on it. If the Riemannian curvature is positive, then nearby geodesics oscillate about one another, and if the curvature is negative, geodesics rapidly diverge from one another. More precisely, the stability of geodesics is expressed in terms of the curvature by means of Jacobi's equation [1]. If $\varphi_t$ is a geodesic curve starting from $\varphi_0$, with velocity field $\boldsymbol{v}(t)$ (whose norm is supposed equal to 1), if the sectional curvature of the manifold in all the 2-planes containing $\boldsymbol{v}(t)$ is less than $-c(<0)$, a perturbation of the initial datum will increase at least as $\exp(ct)$:

$$d(\varphi_t, \tilde{\varphi}_t) \ge d(\varphi_0, \tilde{\varphi}_0)\exp(ct)$$

where $\tilde{\varphi}_0$ denotes the perturbed initial datum and $d$ the geodesic distance on the manifold. Moreover, if the curvature at every point and for all the sections is less than $-c$, and if $M$ is compact, then the geodesic flow, that is, the one-parameter group of transformations $(\varphi_0, \boldsymbol{v}(0)) \to (\varphi_t, \boldsymbol{v}(t))$, is mixing (in the usual meaning of ergodic theory). Arnol'd succeeded in calculating the sectional curvature for flows on the two-dimensional torus; he showed that the

curvature is negative for "most" of the sections. This gives an enlightening geometrical picture of the instability of Lagrangian flows.

It was tempting to connect the above considerations on the instability of two-dimensional flows with the problem of weather forecast. In 1963 EN Lorenz stated that a two-week forecast would be a theoretical bound for predicting the atmospheric motion. Lorenz's assertion was based on numerical simulations. He took as model for the large-scale atmospheric motion the two-dimensional Euler equations on the torus, which he truncated to a small number of Fourier modes (about 20). This model is highly unstable and displays exponential sensitivity with respect to the initial datum. However, the parallel between the behavior of this system and the instability of the Lagrangian flow is misleading. On the one hand, if we again do the Lorenz computations on Euler equations, taking into account a large number of Fourier modes, we note a striking phenomenon: the flow has a tendency to self-organize into large vortices, called coherent structures, and simultaneously the exponential sensitivity, as measured in terms of the energy norm of the velocity field, disappears. On the other hand, the problem of predicting the Lagrangian flow is very different, the Lagrangian flow can be exponentially unstable, while the corresponding velocity field quietly converges, in the energy norm, towards some equilibrium. We must keep in mind that the meteorologist aims to predict the values of the velocity field at some future time and not the trajectories of the fluid particles. In fact, it appears that Lorenz has ignored phenomena of a statistical nature which occur when a large number of degrees of freedom are considered; thus, his theoretical bound for the prediction of the atmospheric motion has no definite basis. More detailed reflections on this issue can be found in Robert and Rosier (2001).

## The Cauchy Problem for the Euler Equations for Compressible Inviscid Fluids

As remarked in the introduction, compressible flows yield pressure waves. The equations of motion being nonlinear, these waves interact in an intricate manner giving rise to shocks. This is the main feature of compressible fluid flows. Compressible flows are situated in the more general domain of nonlinear hyperbolic systems, which were intensively studied during the last decades. We only give here an example of the kind of result which can be obtained.

The following theorem, which states that for a set of regular initial data, shocks do not occur till some finite time, is a consequence of a more general result on hyperbolic systems due to Majda (1984).

We consider $\Omega = \Re^3$ and the system [1], [2], [4].

**Theorem** *Assume* $p_0, \boldsymbol{u}_0 \in H^S \cap L^\infty(\Re^3)$, *with* $s > 5/2$ *and* $p_0(\boldsymbol{x}) > 0$. *Then there is a finite time* $T > 0$, *depending on the* $H^s$ *and* $L^\infty$ *norms of the initial data, such that the Cauchy problem for* [1], [2], [4] *has a unique bounded smooth solution* $p$, $\boldsymbol{u} \in C^1([0, T] \times \Re^3)$, *with* $p(t, \boldsymbol{x}) > 0$ *for all* $t, \boldsymbol{x}$.

## Inviscid Flows and Turbulence

Loosely speaking, turbulence is the intricate motion of a slightly viscous flow. Going back to the first half of the last century, there are two main approaches to turbulence. The first is due to Leray. The dissipation of energy is a characteristic feature of three-dimensional turbulence, and Leray thought that, even if very small, the viscosity of the fluid plays an important role, so that to understand turbulence the first step is to study the Navier–Stokes equations. A radically different approach is due to Onsager. Onsager (1949) started with the fundamental remark that the 4/5 law of turbulence, which relates the dissipation of energy to the increments of the velocity field, does not involve viscosity. Furthermore, he observed that the proof of the conservation of energy for the solutions of Euler equations uses an integration by parts which supposes some regularity of the velocity field. He then imagined that an inviscid dissipation mechanism, due to a lack of regularity of the solutions, was at work in Euler equations. In modern terminology, he suggested to model turbulent flows by nonregular (weak) solutions satisfying the Euler equations in the sense of distributions. He also conjectured that if a solution satisfies a Hölder regularity condition of order $>1/3$, then the energy would be conserved.

Onsager's views were revolutionary and forgotten for a long time. Recent works, such as the proof of Onsager's conjecture, the construction of weak solutions with energy dissipation, and the discovery of the explicit local form of the energy dissipation for weak solutions, show a renewed interest in these views (see, e.g., Constantin and Titi (1994), Eyink (1994), Robert (2003), and Shnirelman (2003)).

*See also:* Compressible flows: Mathematical Theory; Dissipative Dynamical Systems of Infinite Dimension; Hyperbolic Dynamical Systems; Incompressible Euler Equations: Mathematical Theory; Non-Newtonian Fluids; Partial Differential Equations: Some Examples; Chaos and Attractors; Turbulence Theories.

## Further Reading


Arnold VI and Khesin BA (1998) *Topological Methods in Hydrodynamics*. Berlin: Springer.

Batchelor GK (1967) *An Introduction to Fluid Dynamics*. Cambridge: Cambridge University Press.

Chemin JY (1995) Fluides parfaits incompressibles. *Asterisque* no 230, SMF.

Chen GQ and Wang D (2002) The Cauchy problem for the Euler equations for compressible fluids. In: Friedlander S and Serre D (eds.) *Handbook of Mathematical Fluid Dynamics*, vol. 1. Amsterdam: Elsevier.

Constantin P, Weinan E, and Titi ES (1994) Onsager's conjecture on the energy conservation for solutions of Euler's equation. *Communications in Mathematical Physics* 165: 207–209.

Eyink G (1994) Energy dissipation vithout viscosity in ideal hydrodynamics. *Physica D* 78(3–4): 222–240.

Frisch U (1995) *Turbulence*. Cambridge: Cambridge University Press.

Kato T (1975) *Quasi-Linear Equations of Evolution with Applications to Partial Differential Equations*, Lecture Notes in Math. vol. 448, pp. 25–70. New York: Springer.

Majda A (1984) *Compressible Fluid Flow and Systems of Conservation Laws in Several Space Variables*. New York: Springer.

Marchioro C and Pulvirenti M (1994) *Mathematical Theory of Incompressible Non-Viscous Fluids*. Berlin: Springer.

Onsager L (1949) Statistical hydrodynamics. *Nuovo Cimento* (suppl. 6): 279–291.

Robert R (2003) Statistical hydrodynamics. In: Friedlander S and Serre D (eds.) *Handbook of Mathematical Fluid Dynamics*, vol. 2. Amsterdam: Elsevier.

Robert R and Rosier C (2001) Long-range predictibility of atmospheric flows. *Nonlinear Processes in Geophysics* 8: 55–67.

Serre D (1979) Les invariants du premier ordre de l'equation d'Euler en dimension 3. *Comptes Rendus de l'Academie des Sciences Paris, Serie A* 289: 267–270.

Shnirelman A (2003) Weak solutions of incompressible Euler equations. In: Friedlander S and Serre D (eds.) *Handbook of Mathematical Fluid Dynamics*, vol. 2. Amsterdam: Elsevier.


**Ising Model** *see* Two-Dimensional Ising Model

# Isochronous Systems

**Francesco Calogero**, University of Rome, Rome, Italy and Institute, Nazionale de Fisica Nucleare, Rome, Italy



## Introduction

This paper reviews recent developments, following closely (sometimes *verbatim*) the review paper Calogero F (2004c) (see the Bibliography below); for more traditional investigations of *isochronous* systems see other entries of this Encyclopedia (and for the mathematical investigation of isochronous centers in the plane, related to the 16th Hilbert problem, see for instance the survey paper referred to at the end of this entry).

The *isochronous* systems treated herein are characterized by the property to possess an *open* domain having full dimensionality in their phase space such that *all* the motions evolving from a set of initial data in it are *completely periodic* with the same fixed period. The natural measure of this *open* domain might, or it might not, be *infinite* when the measure of the entire phase space is itself *infinite*: for instance, if the entire phase space is the two-dimensional Euclidian plane, such a domain might be the exterior, or the interior, of a circle of finite radius.

It is justified to call such systems *superintegrable*, or perhaps *partially superintegrable* inasmuch as the property of *isochronicity* of all their motions holds only in a subregion of the entire phase space. This terminology is justified by the observation that, roughly speaking, all confined motions of a *superintegrable* system – in which all but one of the degrees of freedom are constrained by the existence of the *maximal* possible number of constants of motion – are *completely periodic*, although not necessarily all with a fixed period – entailing that *isochronicity* entails *superintegrability*, while the converse is not the case (see the entry *Integrable systems* in this Encyclopedia).

A simple trick – amounting essentially to a change of independent, and possible as well of dependent, variables, allows to deform a largely arbitrary dynamical system so that the deformed system obtained from it be *isochronous*. This "trick", which is now explained, entails therefore that *isochronous systems are not rare*. Below we provide several examples; others can be found in the further reading suggested at the end of this entry, and/or can be manufactured *ad libitum* using the trick.

## The Trick

We now show that, given a largely arbitrary dynamical system, it is possible to introduce a deformed version of it featuring a *real* constant $\omega$, that has the following properties: for $\omega = 0$, it coincides with the original, undeformed system; for $\omega > 0$, it possesses an *open* region having full dimensionality in its phase space such that *all* solutions evolving from an initial datum in it are *completely periodic* with a period $\tilde{T}$ which is a finite integer multiple, or perhaps a simple fraction, of the basic period

$$T = \frac{2\pi}{\omega} \qquad [1]$$

Let us indeed, consider a quite general dynamical system which we write as follows:

$$\zeta' = F(\zeta; \tau) \qquad [2]$$

Here $\zeta \equiv \zeta(\tau)$ is the dependent variable, which might be a scalar, a vector, a tensor, a matrix, you name it. The independent variable is $\tau$, and the main limitation on the dynamical system [2] is that it be permissible to treat this variable as *complex*; this requires that the derivative with respect to this *complex* variable $\tau$ that appears in the left-hand side of the evolution equation [2] make sense, namely that this dynamical system be *analytic*, entailing that the dependent variable $\zeta$ be an *analytic* function of the complex variable $\tau$ (but this does not require $\zeta(\tau)$ to be a holomorphic nor a meromorphic function of $\tau; \zeta(\tau)$ might feature all sorts of singularities, including branch points, in the complex $\tau$-plane, indeed this will generally happen since we generally assume the evolution equation (??) to be *nonlinear*). The quantity $F$ in the right-hand side of [2] – which has of course the same scalar, vector, matrix... character as $\zeta$ – might depend (arbitrarily but analytically) on $\zeta$ as well as on $\tau$. (Let us also emphasize that this approach is as well applicable to more general dynamical systems that also feature other, "spacelike", independent variables, for instance are a system of PDEs rather than ODEs; the interested reader is referred to the literature cited below).

In spite of the generality of this dynamical system, [2], there generally holds a result ("Theorem of existence, uniqueness and analyticity") that characterizes the solution $\zeta(\tau)$ of its initial-value problem determined by the assignment

$$\zeta(0) = \zeta_0$$

Here, for notational simplicity, we assign the initial datum $\zeta_0$ at $\tau = 0$; and we assume of course that the right-hand side of [2] is not singular for $\tau = 0$ and $\zeta = \zeta_0$. The relevant result guarantees, not only for the initial datum $\zeta_0$, but for a (sufficiently small but open) set of initial data in its neighborhood, the existence of a circular disk in the complex $\tau$-plane, centered at $\tau = 0$ (where the initial data are assigned) and having a *nonvanishing* radius $\rho$, such that the solutions $\zeta(\tau)$ corresponding to these initial data are *holomorphic* in it, namely for $|\tau| < \rho$ (and note that if $\zeta(\tau)$ is a multicomponent object, the property to be *holomorphic* is featured by each and everyone of its components).

Let us now introduce the following changes of dependent and independent variables:

$$z(t) = \exp(i\lambda\omega t)\zeta(\tau) \qquad [3a]$$

$$\tau \equiv \tau(t) = \frac{\exp(i\omega t) - 1}{i\omega} \qquad [3b]$$

This transformation is called "the trick". The essential part of it is the change of *independent* variable [3b]: and let us re-emphasize that, here and hereafter, the new *independent* variable $t$ is considered as the *real*, "physical time" variable. Note that [3b] entails

$$\tau(0) = 0, \ \dot{\tau}(0) = 1$$

and, most importantly, that $\tau(t)$ is a periodic function of $t$ with period $T$, see [1]. More specifically, as the time $t$ increases from zero onwards, the complex variable $\tau$ travels counterclockwise round and round on the circle $C$ the diameter of which, of length $2/\omega$, lies on the imaginary axis in the complex $\tau$-plane, with one extreme at the origin, $\tau = 0$, and the other at the point $\tau = 2i/\omega$, making a full circle in the time interval $T$. As for the prefactor $\exp(i\lambda\omega t)$ that multiplies $\zeta(\tau)$ in the right-hand side of [3a], its purpose is to allow, via an appropriate choice of the parameter $\lambda$, the deformed system, see below, to have a neater look; however this choice is hereafter restricted by the condition that $\lambda$ be *real* and *rational*, say

$$\lambda = \frac{p}{q}$$

with $p$ and $q$ two coprime integers and $q > 0$. This restriction is essential to guarantee, via [3], that if $\zeta(\tau)$ is *holomorphic* in $\tau$ in the (closed) disk encircled by the circle $C$, then $z(t)$ is *completely periodic* (namely, each and everyone of its components is *periodic*) with the period

$$\tilde{T} = qT \qquad [4]$$

The deformed dynamical system is the one that obtains from [2] via the trick [3]. It clearly reads as follows:

$$\dot{z} = i\lambda\omega z + \exp[i(\lambda+1)\omega t] \times F\left(\exp(-i\lambda\omega t)z; \frac{\exp(i\omega t)-1}{i\omega}\right) \quad [5]$$

And it is plain, on the basis of the arguments we just gave, that this system is *isochronous*, a sufficient condition for the *complete periodicity* with period $\tilde{T}$, see [4], of its solutions being provided by the inequality

$$\frac{2}{\omega} < \rho$$

which can clearly be satisfied by initial data situated inside an *open* domain of such data, at least provided $\omega$ is sufficiently large (actually, in all the examples reported below no restriction on the value of $\omega$ is required, namely such an open domain exists for any arbitrary value of $\omega > 0$).

## Examples

In this subsection we report tersely several examples of *isochronous* dynamical systems; in each case we also provide the reference where more information can be found. Except when explicitly otherwise mentioned, these dynamical systems are to be considered in the *complex* context.

The first example we report is a Hamiltonian $N$-body problem which is a generalization of a well-known *integrable* (indeed, *superintegrable*) system (*see* Integrable Systems: Overview). It is characterized by the (normal) Hamiltonian

$$H(\underline{z},\underline{p}) = \frac{1}{2}\sum_{n=1}^{N}\left(p_n^2+\omega^2 z_n^2\right) + \frac{1}{4}\sum_{m,n=1;m\neq n}^{N}\sum_{k=1}^{K}\frac{f_{nm}^{(k)}}{k(z_n-z_m)^{2k}} \quad [6a]$$

and correspondingly by the Newtonian equations of motion

$$\ddot{z}_n + \omega^2 z_n = \sum_{m=1,m\neq n}^{N}\sum_{k=1}^{K}\frac{f_{nm}^{(k)}}{(z_n-z_m)^{1+2k}} \quad [6b]$$

Here the $\frac{1}{2}N(N-1)K$ "coupling constants" $f_{nm}^{(k)}$ are arbitrary, except for the symmetry restriction $f_{nm}^{(k)} = f_{mn}^{(k)}$ (see [6a]).

The next example we report is a *real* $N$-body problem in the horizontal plane, characterized by the Newtonian equations of motions

$$\ddot{\vec{r}}_n = \omega\hat{k}\wedge\dot{\vec{r}}_n + 2\sum_{m=1,m\neq n}^{N}\left(\alpha_{nm}+\beta_{nm}\hat{k}\wedge\right) \times \frac{\left[\dot{\vec{r}}_n\left(\dot{\vec{r}}_m\cdot\vec{r}_{nm}\right)+\dot{\vec{r}}_m\left(\dot{\vec{r}}_n\cdot\vec{r}_{nm}\right)-\vec{r}_{nm}\left(\dot{\vec{r}}_n\cdot\dot{\vec{r}}_m\right)\right]}{r_{nm}^2} \quad [7]$$

Here $\vec{r}_n \equiv (x_n, y_n, 0)$ is a *real* two-vector in the horizontal plane, $\hat{k} \equiv (0,0,1)$ is the unit vector orthogonal to the horizontal plane, the symbol $\wedge$ denotes the (three-dimensional) vector product so that $\hat{k}\wedge\vec{r}_n = (-y_n, x_n, 0)$, and we use the short-hand notation $\vec{r}_{nm} = \vec{r}_n - \vec{r}_m$ entailing $r_{nm}^2 = r_n^2 + r_m^2 - 2\vec{r}_n\cdot\vec{r}_m$. Note that these equations are *translation-* and *rotation-invariant*; and they are *Hamiltonian*, although the corresponding Hamiltonian function is *not* of normal type (kinetic plus potential energy).

The $N(N-1)$ "coupling constants" $\alpha_{nm}$ and $\beta_{nm}$ are of course real, but they are otherwise *arbitrary* except for the symmetry restrictions $\alpha_{nm} = \alpha_{mn}$, $\beta_{nm} = \beta_{mn}$ which are required in order that this system be *Hamiltonian*. If *all* these coupling constants *vanish*, this dynamical system has a clear physical interpretation: it describes the motion of $N$ equal, electrically charged, point particles, moving in the horizontal plane under the effect of a magnetic field orthogonal to that plane (in the approximation in which the electrostatic interparticle interaction is neglected). In that case each particle moves on a circle, the center and radius of which depend on the initial data, while the time taken to go round it is, in all cases, $T$, see [1]. If the $\frac{1}{2}N(N-1)$ coupling constants $\beta_{nm}$ vanish, $\beta_{nm} = 0$, and the $\frac{1}{2}N(N-1)$ coupling constants $\alpha_{nm}$ all equal unity, $\alpha_{nm} = 1$, the system is a well-known *integrable* (indeed *solvable*) system; and this is as well the case if the $\frac{1}{2}N(N-1)$ coupling constants $\beta_{nm}$ vanish, $\beta_{nm} = 0$, and the $\frac{1}{2}N(N-1)$ coupling constants $\alpha_{nm}$ equal minus one half, and only act among "nearest neighbors", $\alpha_{nm} = -\frac{1}{2}(\delta_{m,n+1}+\delta_{m,n-1})$ (see the entry *Integrable systems* in this Encyclopedia).

Because of its many interesting features as well as the neatness of its equations of motion (especially in their complex version, see below) the honorary title of "goldfish" has been attributed to this model, characterized by the Newtonian equations of motion in the plane [7]. A more detailed discussion of it – in particular of its behavior for initial data *outside* of the region yielding *isochronous* motions – is made in the next section.

Several interesting classes of *isochronous* dynamical systems are reported in Calogero F. (2004b).

We only mention here a remarkably general example, characterized by the Newtonian equations of motion

$$\ddot{\underline{z}} + i\omega\dot{\underline{z}} = \sum_{k=1}^{K} \underline{f}^{(-k)}(\underline{z}, \dot{\underline{z}} + i\omega\underline{z})$$

where $\underline{z} \equiv (z_1, \ldots, z_N)$ is the $N$-vector whose *complex* components $z_n \equiv z_n(t)$ are the dependent variables, while the "forces" $\underline{f}^{(-k)}(\underline{z}, \tilde{\underline{z}})$ are required to be *analytic* in all their arguments and to satisfy the scaling properties

$$\underline{f}^{(-k)}(\alpha\underline{z}, \tilde{\underline{z}}) = \alpha^{-k}\underline{f}^{(-k)}(\underline{z}, \tilde{\underline{z}})$$

which however entail no restriction on the velocity-dependence of these forces, namely on the dependence of $\underline{f}^{(-k)}(\underline{z}, \tilde{\underline{z}})$ on the (components of the) second, $\tilde{\underline{z}}$, of its two $N$-vector arguments.

The next example we report is characterized by the Newtonian equations of motion

$$\ddot{\vec{r}}_n + i\omega\dot{\vec{r}}_n + 2\omega^2\vec{r}_n = \sum_{m=1, m\neq n}^{N} \frac{M_m \vec{r}_{mn}}{r_{mn}^3}$$

where we assume the $N$ dependent variables $\vec{r}_n \equiv \vec{r}_n(t)$ to be three-vectors (although the property of *isochronicity* of this deformed system would hold no less if these were $S$-vectors, with $S$ an arbitrary positive integer) and we use the short-hand notation $\vec{r}_{mn} = \vec{r}_m - \vec{r}_n$. This system is (perhaps) remarkable inasmuch as it represents a (*complex*) deformation of the classical $N$-body gravitational problem, to which it clearly reduces for $\omega = 0$.

The next example we report is characterized by the following (first-order) equations of motion of oscillator type:

$$\dot{x}_n - ip_n\omega x_n = f_n(\underline{x}, \underline{y}), \;\; n = 1, \ldots, N$$
$$\dot{y}_m + iq_m\omega y_m = g_m(\underline{x}, \underline{y}), \;\; m = 1, \ldots, M \qquad [8]$$

Here the $N$-vector $\underline{x}$, respectively the $M$-vector $\underline{y}$, have as components the $N + M$ *complex* dependent variables $x_n \equiv x_n(t), y_m \equiv y_m(t)$; the $N + M$ parameters $p_n, q_m$ are all *nonnegative integers* (or they could be *nonnegative rational* numbers); and the $N + M$ *complex* functions $f_n, g_m$ are restricted by the following conditions (which are sufficient to guarantee the *isochronicity* of this dynamical system):

(1) $f_n(\underline{x}, \underline{y})$ and $g_m(\underline{x}, \underline{y})$ are *holomorphic* at $\underline{x} = 0, \underline{y} = 0$;
(2) $\lim_{\varepsilon\to 0}[\varepsilon^{-1}\underline{f}(\varepsilon\underline{x}, \varepsilon\underline{y})] = \underline{0}$, $\lim_{\varepsilon\to 0}[\varepsilon^{-1}\underline{g}(\varepsilon\underline{x}, \varepsilon\underline{y})] = \underline{0}$;
(3) $\underline{f}(\underline{x}, \underline{y})$ and $\underline{g}(\underline{x}, \underline{y})$ are *polynomial* in the $y_m$;
(4a) $\lim_{\varepsilon\to 0}[\varepsilon^{-1-p_n} f_n(\varepsilon^{\underline{p}}\underline{x}, \varepsilon^{-\underline{q}}\underline{y})] = \text{nondivergent}, n = 1, \ldots, N$;
(4b) $\lim_{\varepsilon\to 0}[\varepsilon^{-1+q_m} g_m(\varepsilon^{\underline{p}}\underline{x}, \varepsilon^{-\underline{q}}\underline{y})] = \text{nondivergent}, m = 1, \ldots, M$.

In the conditions (4a) and (4b) the notation $\varepsilon^{\underline{p}}\underline{x}$ indicates of course the $N$-vector of components $\varepsilon^{p_n}x_n$, and likewise $\varepsilon^{-\underline{q}}\underline{y}$ is the $M$-vector of components $\varepsilon^{-q_m}y_m$.

Note that this dynamical system, [8], includes the *Hamiltonian* case characterized by the restrictions

$$N = M, p_n = q_n, f_n(\underline{x}, \underline{y}) = \frac{\partial V(\underline{x}, \underline{y})}{\partial y_n}, g_n(\underline{x}, \underline{y}) = -\frac{\partial V(\underline{x}, \underline{y})}{\partial x_n}$$

which imply that the equations of motion [8] are just the Hamiltonian equations entailed by the Hamiltonian function

$$H(\underline{x}, \underline{y}) = i\omega \sum_{n=1}^{N} p_n x_n y_n + V(\underline{x}, \underline{y})$$

*isochronicity* being now guaranteed by the following conditions on the function $V(\underline{x}, \underline{y})$:

(1) $V(\underline{x}, \underline{y})$ is *holomorphic* at $\underline{x} = 0, \underline{y} = 0$;
(2) $\lim_{\varepsilon\to 0}[\varepsilon^{-2} V(\varepsilon\underline{x}, \varepsilon\underline{y})] = \underline{0}$;
(3) $V(\underline{x}, \underline{y})$ is *polynomial* in the $y_n$;
(4) $\lim_{\varepsilon\to 0}[\varepsilon^{-1} V(\varepsilon^{\underline{p}}\underline{x}, \varepsilon^{-\underline{p}}\underline{y})] = \text{nondivergent}$.

The last two examples we report can be characterized as assemblies of *non-linear harmonic oscillators*, inasmuch as these two dynamical systems (which are actually special cases of more general systems) have the remarkable property that their *generic* solutions (namely, *all* their solutions, except for a lower-dimensional set of *singular* solutions in which one or more of the "moving particles" shoot off to infinity at a finite time) are *completely periodic* with the fixed period $T$, see [1]. Their Newtonian equations of motion read

$$\ddot{\vec{z}}_{nm} - 3i\omega\dot{\vec{z}}_{nm} - 2\omega^2\vec{z}_{nm} = c\sum_{\nu=1}^{N}\sum_{\mu=1}^{M} \vec{z}_{n\mu}(\vec{z}_{\nu\mu}\cdot\vec{z}_{\nu m})$$

$$\ddot{\vec{z}}_{nm} - 3i\omega\dot{\vec{z}}_{nm} - 2\omega^2\vec{z}_{nm} = c\sum_{\nu=1}^{N}\sum_{\mu=1}^{M} \vec{z}_{\nu\mu}(\vec{z}_{\nu\mu}\cdot\vec{z}_{nm})$$

These are two (different) systems of $NM$ Newtonian equations of motion satisfied by the $NM$ *complex* $S$-vectors $\vec{z}_{nm}$ (with $S$ an *arbitrary* positive integer); hence here the index $n$ runs from 1 to $N$, and the index $m$ runs from 1 to $M$, with $N$ and $M$ two *arbitrary* positive integers, while $c$ is of course an *arbitrary* complex constant (which might actually be rescaled away). The dot sandwiched between two

$S$-vectors denotes the standard (Euclidian) scalar product, entailing the rotation-invariant character, in $S$-dimensional space, of these equations of motion. Since these systems only feature *linear* and *cubic* forces, these models are remarkably close to physics; and they become even more applicable if they are written in their *real* versions, that obtain in an obvious manner by setting

$$\vec{z}_{nm} = \vec{x}_{nm} + i\vec{y}_{nm}, c = a + ib$$

In contrast to what we did for the previous examples, let us outline here the derivation of these results. Actually the two systems of Newtonian equations written above are merely special subcases, corresponding to appropriate parametrizations of a square matrix $M$ (of appropriate rank) in terms of $S$-vectors, of the following nonlinear *matrix* evolution equation:

$$\ddot{M} - 3i\omega\dot{M} - 2\omega^2 M = cM^3 \qquad [9]$$

Hence the findings reported above are merely special cases of the more general result according to which the *generic* solution of this nonlinear *matrix* evolution equation – with $M \equiv M(t)$ a square matrix of *arbitrary* rank – is *periodic* with period $T$, see [1]:

$$M(t + T) = M(t)$$

And this result is an immediate consequence, via the following *matrix* version of the trick

$$M(t) = \exp(i\omega t)\Psi(\tau), \tau = \frac{\exp(i\omega t) - 1}{i\omega} \qquad [10]$$

of a previous result due to V. I. Inozemtsev, according to which the matrix evolution equation

$$\Psi'' = c\Psi^3$$

which clearly corresponds to [9] via [10], is *integrable* and *all* its solutions $\Psi(\tau)$ are *meromorphic* functions of the independent variable $\tau$.

## The Transition to Deterministic Chaos

In this section we illustrate, using the *real* $N$-body problem in the plane characterized by the Newtonian equations of motion [7], the behavior of an *isochronous* system of the kind described above when the initial data fall *outside* of the region yielding *isochronous* motions.

To do this it is convenient to use the *complex* version of the equations of motion [7], that obtain from [7] by setting

$$z_n = x_n + iy_n, \vec{r}_n = (x_n, y_n, 0),$$
$$\hat{k} = (0, 0, 1), a_{nm} = \alpha_{nm} + i\beta_{nm} \qquad [11]$$

and read as follows:

$$\ddot{z}_n = i\omega\dot{z}_n + 2\sum_{m=1, m\neq n}^{N} \frac{a_{nm}\dot{z}_n\dot{z}_m}{z_n - z_m} \qquad [12]$$

The main tool of our analysis is the (particularly simple) version of the trick appropriate to this model,

$$z_n(t) = \zeta_n(\tau), \tau = \frac{\exp(i\omega t) - 1}{i\omega} \qquad [13a]$$

entailing

$$z_n(0 = \zeta_n(0), \dot{z}_n(0) = \zeta'_n(0) \qquad [13b]$$

that relates our equations of motion [12] to the equations of motion

$$\zeta''_n = 2\sum_{m=1, m\neq n}^{N} \frac{a_{nm}\zeta'_n\zeta'_m}{\zeta_n - \zeta_m} \qquad [14]$$

These equations of motion, together with the initial data $\zeta_n(0), \zeta'_n(0)$ (see [13b]) define the solutions $\zeta_n \equiv \zeta_n(\tau)$ in the *complex* $\tau$-plane. The "physical" evolution of the points $z_n \equiv z_n(t)$ as functions of the *real* time variable $t$ is then given by the evolution of the corresponding coordinates $\zeta_n(\tau)$, see [13a], as the *complex* variable $\tau$ travels round and round on the circle $C$ in the *complex* $\tau$-plane, the diameter of which of length $2/\omega$, has one extreme at the origin $\tau = 0$ and the other on the positive imaginary axis at $\tau = 2i/\omega$. It is therefore clear that the behavior of $z_n(t)$ as a function of the *real*, "physical time" variable $t$ depends on the analytic structure of $\zeta_n(\tau)$ as function of the *complex* variable $\tau$, in particular of the singularities, if any, of this function $\zeta_n(\tau)$ that fall in the disk $D$ encircled by the circle $C$ in the *complex* $\tau$-plane.

Let us tersely review the relevant analysis. We recall first of all that (it can be proven that) there exists in phase space an *open* region of initial data $z_n(0), \dot{z}_n(0)$, characterized by *large* values of the moduli $|z_n(0) - z_m(0)|$ of the initial interparticle distances and by *small* values of the moduli of the initial particle velocities $|\dot{z}_n(0)|$ (see [14] and [13b]), that guarantees (all components $\zeta_n(\tau)$ of) the corresponding solution $\underline{\zeta}(\tau)$ of [14] to be *holomorphic* in (a disk of radius $\rho$ centered at the origin $\tau = 0$ of the *complex* $\tau$-plane that includes) the circle $C$, hence the corresponding solution $\underline{z}(t)$ to be *completely periodic* with period $T$, see [13a] and [1]. This result guarantees the *isochronous* character of this model, [12], for any arbitrarily given assignment of the coupling constants $a_{nm}$.

Next, let us restrict, for simplicity, our consideration to models [12] in which the coupling constants $a_{nm}$ are *real* and *nonnegative*,

$$a_{nm} \geq 0 \qquad [15]$$

Then the singularities of the generic solution $\underline{\zeta}(\tau)$ of [14] – which occur at values $\tau_b$ of $\tau$ where *two* coordinates $\zeta_n(\tau)$ coincide, say $\zeta_\mu(\tau_b) = \zeta_\nu(\tau_b) = b$ (see the right-hand side of [14]) – are *branch points* characterized by the exponent, say,

$$\gamma = \gamma_{\mu\nu} = \frac{1}{1 + a_{\mu\nu}} \qquad [16]$$

so that in their neighborhood, namely for $\tau \approx \tau_b$,

$$\zeta_s(\tau) = b \pm c(\tau - \tau_b)^\gamma + v(\tau - \tau_b) + \sum_{k=1}^{\infty} \sum_{\ell, m=0 \, \ell+m \geq 1}^{\infty} \varphi_{k\ell m}^{(s)} (\tau - \tau_b)^{k + \ell\gamma + m(1-\gamma)}, \quad s = \mu, \nu \qquad [17a]$$

$$\zeta_n(\tau) = b_n + v_n(\tau - \tau_b) + \sum_{k=1}^{\infty} \sum_{\ell=\delta_{k1}}^{\infty} \sum_{m=0}^{\infty} \varphi_{k\ell m}^{(n)} (\tau - \tau_b)^{k + \ell\gamma + m(1-\gamma)}, \quad n \neq \mu, \nu \qquad [17b]$$

The $\pm$ sign in front of $c$ in the right- hand side of the first, [17a], of these formulas indicates that one sign must be chosen for $s = \mu$, the opposite for $s = \nu$. Note that here the $4 + 2(N-2) = 2N$ constants $\tau_b, b, c, v, b_n, v_n$ are *a priori* arbitrary – except for the obvious restrictions $b_n \neq b, b_n \neq b_m$ – while the coefficients $\varphi_{k\ell m}^{(s)}, \varphi_{k\ell m}^{(n)}$ can be computed from these constants, recursively, by inserting this *ansatz*, [17], in the equations of motion [14]. The fact that the number, $2N$, of *a priori* undetermined coupling constants equals the number of arbitrary initial data for this system of ODEs, [14], indicates that this kind of branch points, characterized by the exponents $\gamma_{nm}$, see [16], is the typical singularity featured by the generic solution $\underline{\zeta}(\tau)$ of [14]. (Branch points with different exponents may appear, but only in *nongeneric* solutions $\underline{\zeta}(\tau)$ which, at some value $\tau_b$ of $\tau$, feature the coincidence of more than two components, say $\zeta_\mu(\tau_b) = \zeta_\nu(\tau_b) = \zeta_\lambda(\tau_b)$).

We conclude therefore that the *generic* solution $\underline{\zeta}(\tau)$ of [14] features a, generally infinite, number of branch points, that generally affect each of its components $\zeta_n(\tau)$, and which are characterized, for the class of models to which we are restricting attention here, see [15]) by (*real*) exponents $\gamma_{nm}$, see [16], which are then clearly characterized by the inequalities

$$0 < \gamma_{nm} \leq 1$$

What does this tell us about the *generic* solution $\underline{z}(t)$ of the equations of motions of primary interest to us, [12], in particular about its evolution as function of the *real* "time" variable $t$?

To the solution $\underline{\zeta}(\tau)$ is associated a Riemann surface the structure of which is determined by the character and distribution of the branch points of $\underline{\zeta}(\tau)$ in the complex $\tau$-plane (each of which is generally featured by each component $\zeta_n(\tau)$ of $\underline{\zeta}(\tau)$, although generally not in the same way: see [17]), and we know from [13a] that the values taken by $\underline{z}(t)$ as $t$ evolves from $t = 0$ towards $t = \infty$ coincide with the values taken by $\underline{\zeta}(\tau)$ as the independent variable $\tau$ travels, on that Riemann surface associated with $\underline{\zeta}(\tau)$, counterclockwise round and round on the circle $C$ defined above (the diameter of which lies on the imaginary axis in the complex $\tau$-plane, with one end at $\tau = 0$ and the other at $\tau = 2i/\omega$), employing a period $T$, see [1], to make each full round. Hence the behavior of the solution $\underline{z}(t)$ of [12] depends on the structure of the Riemann surface associated with the corresponding solution $\underline{\zeta}(\tau)$ of [14], and specifically on the number of different sheets of that surface that are visited as one travels on it before returning, if ever, to the main sheet from which the travel started at $t = \tau = 0$.

If no other sheet is visited besides the main one, the corresponding solution $\underline{z}(t)$ is of course periodic with period $T$, see [1] and [13a],

$$\underline{z}(t + T) = \underline{z}(t) \qquad [18]$$

This happens provided no branch point is featured by $\underline{\zeta}(\tau)$ on its main sheet *inside* the circle $C$; and, as already indicated above, it has been proven (even in the more general case with arbitrary coupling constants $a_{nm}$) that there is an *open* region having full dimensionality in the phase space of initial data, see [13b], that yields such an outcome, implying the *isochronicity* of the model characterized by the Newtonian equations of motion [12]. This region $R$ of initial data has a boundary – a lower-dimensional domain in the phase space of initial data – out of which emerge motions leading, at a time $t_b$ smaller than $T$, to a "particle collision", say $z_\nu(t_b) = z_\mu(t_b)$.

The character of the solution $\underline{z}(t)$ yielded by initial data *outside* of the region $R$ depends on the structure of the Riemann surface associated with the corresponding solution $\underline{\zeta}(\tau)$. This is mainly determined by the values of the branch point exponents $\gamma_{nm}$, which are themselves determined by the values of the coupling constants $a_{nm}$, see [17] and [16]. Let us focus on the (more interesting) case in which these constants $a_{nm}$ are *rational* numbers, entailing that the coefficients $\gamma_{nm}$ determining the

character of the branch points are as well *rational*, see [16], so that each of the cuts associated with them opens the way, in the Riemann surface, to a *finite* number of sheets. There are then two possibilities, each generally characterized by *open* regions of initial data having full dimensionality in phase space, the boundaries of which always are (lower-dimensional) domains out of which emerge motions leading, in a time $t_b$ smaller than $T$, to a "particle collision".

One possibility is that the number $B$ of sheets visited before returning to the main sheet be *finite*, $B < \infty$; the corresponding solutions $\underline{z}(t)$ are then *completely periodic* with period $\tilde{T} = (B+1)T$, $\underline{z}(t+\tilde{T}) = \underline{z}(t)$.

Another possibility is that the number of new sheets visited be *unlimited*, namely that the structure of the Riemann surface be such that, by traveling round and round on it along the circle $C$ one *never* returns back to the main sheet. This can happen, even if the exponents $\gamma_{nm}$ are *all rational* so that via the cuts associated to each of them access is gained to only a *finite* number of new sheets, because of the possibility that an *infinity* of branch points be located inside the circle $C$ on the *infinite* sheets associated to these branch points, via a never ending mechanism of branch points nesting. Whenever this happens the corresponding solution $\underline{z}(t)$ is *aperiodic*; and it is moreover likely that it then be *chaotic*, in the sense of displaying a *sensitive dependence* on its initial data. Indeed this will happen whenever some ones out of this infinity of branch points fall arbitrarily close to the contour $C$, because then a *minute* change in the initial data, to which there will correspond a *minute* change in the pattern of these branch points of $\underline{\zeta}(\tau)$ in the complex $\tau$-plane, will cause some relevant branch point to cross over from outside the circle $C$ to inside it, or viceversa, and this will eventually affect quite significantly the time evolution of $\underline{z}(t)$, by causing a change in the sequence of sheets that get visited by traveling along the circle $C$ on the Riemann surface associated to the corresponding $\underline{\zeta}(\tau)$.

This phenomenology has a clear "physical interpretation", which can be qualitatively understood as follows. The $N$-body problem characterized by the Newtonian equations of motion [12] generally yields confined motions, the trajectory of each particle tending to wind round and round – it would indeed reduce to a circle were it not for the interaction with the other particles. A possibility, as we know, is that this $N$-body motion be *completely periodic*, with the same period $T$ that characterizes the circular motion of each particle when the two-body interparticle interaction is altogether missing ($a_{nm} = 0$). Another possibility, in the case discussed above with *rational* coupling constants, is that there exist other motions which are as well *completely periodic*, but with periods which are *integer multiples* of $T$. A third possibility, which cannot *a priori* be excluded, is that there also exist motions which are *aperiodic* but in some way overall *ordered*, perhaps featuring trajectories that eventually wind up around *limit cycles*. And still another possibility is that the motions described by the solution $\underline{z}(t)$ be *aperiodic* and *disordered*. In this case the physical mechanism causing a sensitive dependence on the initial data can be understood as follows. Such disordered motions necessarily feature *near misses*, in which, typically, two particles pass *quite close* to each other (while the probability that an actual collision occur among *point particles* moving *in a plane* is of course *a priori* nil). Such a *near miss* in the motion described by $\underline{z}(t)$ corresponds – see the discussion above – to a branch point of the corresponding solution $\underline{\zeta}(\tau)$ occurring *quite close* to the circle $C$ in the complex $\tau$-plane (which is the one-dimensional region of the two-dimensional complex $\tau$-plane in which the values of $\underline{\zeta}(\tau)$ correspond to the values $\underline{z}(t)$ describing the motion of physical particles moving as functions of the time $t$); and in the generic case of a *two-body* near miss, there is a correspondence between the fact that such a branch point occur *just inside*, or *just outside*, the circle $C$, and the way the particles pass, on one or the other side, by each other. Likewise, the tiny change in the initial data that causes, in the context of the solutions $\underline{\zeta}(\tau)$ – see the discussion above – a branch point of $\underline{\zeta}(\tau)$ to pass from inside to outside the circle $C$, or viceversa, corresponds, in the context of the "physical" solutions $\underline{z}(t)$, to a change occurring in the corresponding *near miss*, from the case in which the two particles involved in it slide by each other on one side to the case in which they instead slide by each other on the other side – entailing a significant change in the subsequent motion (indeed, the closer a near miss, the more it affects the motion, due to the singularity of the two-body interaction at zero separation, see [12]).

The phenomenology outlined here does indeed occur in this goldfish model. It also occurs – rather similarly if more simply, since in this case only square-root branch points occur, irrespective of the values of the coupling constants – in the model [6] with $K = 1$. Indeed, it is clear that this phenomenology provides a paradigm of rather general applicability for the transition from *isochronicity* to *deterministic chaos*, indeed perhaps for the generic *onset* of *deterministic chaos*.

*See also:* Bifurcations of Periodic Orbits; Calogero–Moser–Sutherland Systems of Nonrelativistic and Relativistic Type; Integrable Systems: Overview; Quantum Calogero–Moser Systems; Synchronization of Chaos.

## Further Reading


Calogero F (2001) *Classical Many-Body Problems Amenable to Exact Treatments*. Heidelberg: Springer.

Calogero F (2004a) Partially superintegrable (indeed isochronous) systems are not rare. In: Shabat AB, Gonzalez-Lopez A, Mañas M, Martinez Alonso L, and Rodriguez MA (eds.) *New Trends in Integrability and Partial Solvability*, NATO Science Series, II. Mathematics, Physics and Chemistry, Proceedings of the NATO Advanced Research Workshop held in Cadiz, Spain, 2–16 June 2002, vol. 132, pp. 49–77. Kluwer.

Calogero F (2004b) Isochronous dynamical systems. *Applicable Anal*. (in press).

Calogero F (2004c) Isochronous systems. *Proceedings of the VI International Conference on Geometry, Integrability and Quantization*, Varna, June 2004. Edited by Mladenov IM and Hirshfeld AC, Sofia, Bulgaria, 2005, pp. 11–61 (ISBN 954-84952-9-5).

Calogero F and Françoise J-P (2002) Periodic motions galore: how to modify nonlinear evolution equations so that they feature a lot of periodic solutions. *Nonlinear Math. Phys* 9: 99–125.

Calogero F and Françoise J-P (2003) Nonlinear evolution ODEs featuring many periodic solutions. *Theor. Mat. Fis* 137: 1663–1675.

Calogero F and Françoise J-P (2002) Isochronous motions galore: nonlinearly coupled oscillators with lots of isochronous solutions. In: *Proceedings of the Workshop on Superintegrability in Classical and Quantum Systems,* Centre de Recherches Mathématiques (CRM), Université de Montréal, September 2002, CRM Proceedings and Lecture Notes, American Mathematical Society, 2004, vol. 37, pp. 15–27.

Chavarriga J and Sabatini M (1999) A survey of isochronous systems. *Qualitative Theory Dyn. Syst* 1: 1–70.

Mariani M and Calogero F (2005) Isochronous PDEs. *Yadernaya Fizika* (*Russian Journal of Nuclear Physics*) 68: 958–968.


# Isomonodromic Deformations

**V P Kostov**, Université de Nice – Sophia Antipolis, Nice, France



## Introduction

In this article we consider families of linear differential equations whose monodromy data do not depend on the parameters. Such families are called isomonodromic deformations of any of the equations of the family (for the definitions of a regular and Fuchsian linear system and of their monodromy groups, see Riemann–Hilbert Problem).

## Schlesinger's Equation

The best-studied example of an isomonodromic deformation is the Fuchsian system on Riemann's sphere $\mathbf{C}P^1 = \mathbf{C} \cup \infty$ considered by L Schlesinger:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \left(\sum_{j=1}^{p+1} \frac{A_j}{t - a_j}\right) X \qquad [1]$$

Here the poles $a_j \in \mathbf{C}$ are free parameters and the matrices-residua $A_j$ depend analytically on $a := (a_1, \ldots, a_{p+1})$; therefore, system [1] is in fact a family of linear systems which is an analytic deformation of the system obtained for $a_j = a_j^0$.

One can think of system [1] as defined by the Pfaffian system

$$\mathrm{d}X = \omega_{\mathrm{s}} X, \quad \omega_{\mathrm{s}} = \sum_{j=1}^{p+1} \frac{A_j}{t - a_j} \mathrm{d}(t - a_j) \qquad [2]$$

Suppose first that the poles $a_j$ vary within small nonintersecting disks of the points $a_j^0$, so small that the standard system of generators of the monodromy group could be defined by one and the same contours for all values of the parameters $a_j$ (see **Figure 1** from Riemann–Hilbert Problem). Suppose also that one chooses $\infty$ as base point and that one has

$$X|_{t=\infty} = I \qquad [3]$$

(where $I$ is the identity matrix) for all values of the parameters $a_j$. Finally, suppose that all matrices $A_j$ are nonresonant, that is, without two eigenvalues differing by a nonzero integer. Then the following conditions are necessary and sufficient for system [1] to be isomonodromic:

$$\mathrm{d}A_i(a) = -\sum_{j=1, j\neq i}^{p+1} \frac{[A_i(a), A_j(a)]}{a_i - a_j} \mathrm{d}(a_i - a_j)$$
$$i = 1, \ldots, p+1 \qquad [4]$$

This system (called Schlesinger's equations) results from the Frobenius integrability condition $\mathrm{d}\omega_{\mathrm{s}} = \omega_{\mathrm{s}} \wedge \omega_{\mathrm{s}}$ of system [2].

**Remarks 1**

(i) To find the matrices-residua $A_j$ as functions of $a$ and given their values $A_j|_{a=a^0}$ is a Cauchy problem. It is solvable for $a$ close to $a^0$ and the matrices $A_j$ are analytic in $a$.

(ii) The differential of $A_i$ being a commutator $[A_i, .]$, the matrix $A_i$ remains within its conjugacy class throughout the deformation.

(iii) Schlesinger's equations are the necessary and sufficient conditions for isomonodromy also in the case when system [1] has a logarithmic pole at $\infty$ whose matrix-residuum does not change throughout the deformation. In this case the solution to system [1] in its Levelt's decomposition at $\infty$ (*see* Riemann–Hilbert Problem) equals $U_\infty(1/t)t^{-D_\infty}t^{-E_\infty}G$, where $D_\infty$ is a diagonal matrix with integer entries, $E_\infty$ is an upper-triangular constant matrix, and $U_\infty$ is holomorphic at $\infty$ and such that $U_\infty(0)=I$.

**Definition 2** The deformation satisfying condition [4] with initial condition [3] for the solution to system [1] is called the normalized Schlesinger deformation.

**Remark 3** When the matrices-residua $A_j$ are nonresonant, then every isomonodromic deformation of system [1] with $a_j=a_j^0$ is either the normalized Schlesinger deformation or is a nonnormalized Schlesinger deformation, that is, obtained from the normalized one by a change of variables $X\mapsto C(a)X, C(a)\in GL(n,\mathbf{C})$. In this way, one has $X|_{t=\infty}=C(a)$ instead of [3] and the deformation is described by a Pfaffian system with a form of the kind $\omega_\mathrm{n}=\omega_\mathrm{s}+\sum_{j=1}^{p+1}\gamma_j(a)\mathrm{d}a_j$.

**Example 4** The following one-parameter Fuchsian family is an isomonodromic Schlesinger deformation:

$$\frac{\mathrm{d}X}{\mathrm{d}t}=\left(\sum_{j=1}^{p+1}\frac{A_j}{t-ba_j^0}\right)X$$

Here the matrices $A_j$ are constant and the parameter $b$ takes nonzero values. Indeed, one either checks directly that there holds condition [4] or one makes the change of time (which does not change monodromy) $t\mapsto bt$ after which the parameter $b$ disappears.

A A Bolibrukh has shown that in the resonant case every isomonodromic deformation of a Fuchsian system is described by an integrable Pfaffian system with 1-form $\omega=\omega_\mathrm{n}+\omega_\mathrm{m}$, where the meromorphic 1-form $\omega_\mathrm{m}$ vanishes at $\infty$ and has poles of orders $\le r_j$ along the hyperplanes $\{x-a_j=0\}$; here $r_j$ is the largest nonzero integer difference between two eigenvalues of the matrix $A_j$.

Consider now Schlesinger's equation in the global situation, that is, when the poles $a_j$ belong to the universal covering $Z$ of the space $\mathbf{C}^n\backslash\Delta$, where $\Delta$ is the "diagonal," that is, the union of all sets $\{a_i=a_j\}$, $i\neq j$. Suppose that the matrices $A_j$ are nonresonant. There are values of $a$ (their set is denoted by $\Theta$) for which some entries of some of the matrices-residua $A_j$ tend to $\infty$. Typically, at such points the matrices $A_j$ have poles of second order; this is a result due to Bolibrukh. Indeed, set $A_j=Q_j^{-1}J_jQ_j$, where $J_j$ is the Jordan normal form of $A_j$; hence, this is a constant matrix; we assume that $Q_j\in SL(n,\mathbf{C})$. Typically, at points of $\Theta$ the matrices $Q_j$ and $Q_j^{-1}$ have simple poles, which makes a pole of second order for $A_j$.

B Malgrange and, independently, T Miwa have proved that system [4] is completely integrable and that it has the Painlevé property: "The only movable singularities of its solutions are poles." (The fixed singularities of the solutions are, by definition, along the points of $Z$ which are over $\Delta$. The positions of the movable singularities depend on the initial condition, that is, on the values of the matrices $A_j$ for $a=a^0$.) In other words, the solutions to Schlesinger's equation are matrices meromorphic on $Z$.

**Theorem 5** *The set $\Theta$ of movable singular points of the Schlesinger equation is the set of zeros of a function $\tau$ (the Miwa $\tau$-function) holomorphic on $Z$ and such that*

$$\frac{1}{2}\sum_{i,j,i\neq j}\frac{\mathrm{tr}(A_i(a)A_j(a))\mathrm{d}(a_i-a_j)}{a_i-a_j}=\mathrm{d}\log(\tau(a))$$

Some improvements of this result are due to Malgrange and Bolibrukh.

## Isomonodromy and Confluence

The idea to consider a linear system of ordinary differential equations with a pole of order higher than 1 as embedded into a family of Fuchsian systems with confluence of the poles has been proposed by V I Arnol'd in 1984 and independently by J-P Ramis in 1988. The idea has been used by A Duval, B Khesin, A A Glutsyuk, and other authors. In particular, it is interesting to relate the Stokes multipliers (defined in the next section) of the system obtained as a result of a confluence to the monodromy groups of the

Fuchsian systems obtained for values of the parameters before the confluence occurs.

**Example 6** Consider the one-parameter family of linear systems:

$$(t^2 - \lambda)\mathrm{d}X/\mathrm{d}t = (A(\lambda)t + B(\lambda))X \qquad [5]$$

Here the matrices $A$, $B$, and $X$ are $n \times n$. Suppose that $t \in \mathbf{C}$ (i.e., we do not consider singularity at $\infty$), $\lambda \in (\mathbf{C}, 0)$. Then for $\lambda \neq 0$ the system is Fuchsian – it has two logarithmic poles at $\pm\lambda^{1/2}$ whose confluence for $\lambda = 0$ gives as a result a pole at 0 which might be of order 2 if $B(0) \neq 0$ or 1 if $B(0) = 0$.

In this section we consider only the situation when the family producing the confluence is isomonodromic for values of the parameters before the confluence.

**Example 7** This is the case of family [5] with $B \sim 0$ and $A$ being a constant nonresonant $n \times n$ matrix. Indeed, the change of time $t \mapsto \lambda^{1/2}t(*)$ transforms the family into the family $(t^2 - 1)\mathrm{d}X/\mathrm{d}t = tAX$ (independent of $\lambda$) which is a Fuchsian system (at $\infty$ as well).

Suppose now that $t \in \mathbf{C}P^1$ (i.e., we consider the singularity at $\infty$ as well). Hence, the monodromy operator $M_\infty$ around $\infty$ is independent of $\lambda$ up to conjugacy (it is conjugate to $\exp(-2\pi \mathrm{i} A)$). On the other hand, consider the monodromy operator $M'$ defined by a contour circumventing counterclockwise both poles at $\pm\lambda^{1/2}$ (one can choose as such a contour a circumference centered at the origin and of sufficiently large radius). It equals $M_\infty^{-1}$, and it is well defined for $\lambda = 0$ as well. (This is not the case of the monodromy operators defined by contours circumventing only one of the poles at $\pm\lambda^{1/2}$.) Hence, up to conjugacy $M'$ is independent of $\lambda$. As $M'$ is in a sense the only monodromy operator that can be defined by a contour depending continuously on $\lambda$ for all $\lambda \in (\mathbf{C}, 0)$ and not passing through a pole of the system, one can say that the family is strongly isomonodromic.

**Example 8** Consider now family [5] with $n = 2$,

$$A = \begin{pmatrix} d & 0 \\ 0 & d \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & \lambda \\ 0 & 0 \end{pmatrix}$$

where $d \in \mathbf{C}$. For $\lambda \neq 0$ the family is isomonodromic – the change of time $(*)$ followed by the change of variables

$$X \mapsto \begin{pmatrix} \lambda^{1/2} & 0 \\ 0 & 1 \end{pmatrix} X(**)$$

brings the family to the form

$$(t^2 - 1)\frac{\mathrm{d}X}{\mathrm{d}t} = \left(\begin{pmatrix} d & 0 \\ 0 & d \end{pmatrix} t + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right) X$$

which is independent of $\lambda$, hence, isomonodromic. However, the change of variables $(**)$ is not defined for $\lambda = 0$. The monodromy operator $M'$ (defined as above) is scalar for $\lambda = 0$ and conjugate to a Jordan block of size 2 for $\lambda \neq 0$. Hence, the family is not strongly isomonodromic.

The following example is closely connected to singularity theory. It has been suggested by F Pham.

**Example 9** Consider the Abelian integrals

$$I_1 = \int \mathrm{d}x/(x^3 + sx + t) \quad \text{and}$$
$$I_2 = \int x\,\mathrm{d}x/(x^3 + sx + t)$$

taken over a closed contour $\gamma$ belonging to a nonsingular fiber of the function $f(x) = x^3 + sx + t$. Suppose that $x^3 + sx + t \neq 0$ on $\gamma$. Obviously, $I_1$ and $I_2$ depend only on $[\gamma]$, the class of homotopy equivalence of $\gamma$. Set

$$x^3 + sx + t = (x - x_1)(x - x_2)(x - x_3),$$
$$x_j = x_j(s, t)$$

Then one has

$$I_k = 2\pi \mathrm{i} \sum_{j=1}^{3} \delta_{k,j} x_j^{k-1} / \left(3x_j^2 + s\right), \quad k = 1, 2$$

where the integers $\delta_{k,j}$ depend only on $[\gamma]$ (the contour $\gamma$ is homotopy equivalent to a linear combination with integer coefficients of small loops around the roots of $f$; the integral along such a loop is computed using residua). Note that

$$\dot{x}_j := \mathrm{d}x_j/\mathrm{d}t = -1/\left(3x_j^2 + s\right)$$

An easy computation shows that the integrals $I_1, I_2$ satisfy the following Picard–Fuchs system of differential equations:

$$-t\dot{I}_1 - 2s\dot{I}_2/3 = 2I_1/3$$
$$2s^2\dot{I}_1/9 - t\dot{I}_2 = I_2/3$$

The system admits also a presentation of the form

$$\left(t^2 + \frac{4s^3}{27}\right)\begin{pmatrix} \dot{I}_1 \\ \dot{I}_2 \end{pmatrix} = \begin{pmatrix} -2t/3 & 2s/9 \\ -4s^2/27 & -t/3 \end{pmatrix}\begin{pmatrix} I_1 \\ I_2 \end{pmatrix}$$

Here the unknown variables form a vector column of length 2; to obtain a $2 \times 2$ matrix, one has to choose another contour $\gamma'$ (linearly independent

with $\gamma$ as a linear combination of the loops around the roots $x_i$) which gives the second column of the matrix. The system is strongly isomonodromic – its matrix-residuum at $\infty$ equals diag(2/3, 1/3); hence, the monodromy operator $M'$ up to conjugacy equals $\mathrm{diag}(\exp(-4\pi i/3),\ \exp(-2\pi i/3))$.

A A Bolibrukh has considered the possibility of confluence of poles in Schlesinger's equation (i.e., the possibility to have equalities of the form $a_i = a_j$ in system [1]). He has considered the so-called normalized isomonodromic confluences, that is, isomonodromic confluences defined by Pfaffian systems with coefficient forms $\omega = \omega_s + \omega_m$ alone (see the previous section). He has shown that a normalized isomonodromic confluence of singular points of Fuchsian systems of linear differential equations on Riemann's sphere can only lead to a system with regular singular points. This is a partial answer to a problem stated by V I Arnol'd: how to express a system with regular singular points as a limit of Fuchsian systems?

## Other Results

In the case of a linear system with irregular singular point, isomonodromy means that the formal monodromy and the Stokes multipliers do not change throughout the deformation. The formal monodromy can be computed from the formal normal form (the latter can be found algorithmically; this is due to H Turrittin). Consider, for simplicity, the nonresonant case, that is, the case when the leading matrix in the Laurent series of the system at the singular point has distinct eigenvalues (this definition differs from the one in the case of a Fuchsian singular point). The Stokes multipliers are linear operators acting on the solution space. They are defined as follows: there exist sectors of maximal opening centered at the singular point on each of which the solution is uniquely defined by its asymptotic development. Two solutions $X_1$, $X_2$ having one and the same asymptotic development in two overlapping sectors are related by $X_1 = X_2 C$, where $C$ is a Stokes multiplier. The monodromy operator is expressed as a product of the operator of formal monodromy and the Stokes multipliers. Isomonodromic deformations of systems with irregular singular points have been constructed by B Malgrange. Isomonodromic deformations have been used by Y Sibuya and C H Lin and by Y Sibuya and T J Tabara to investigate Stokes multipliers.

At the beginning of the twentieth century, P Painlevé and B Gambier have classified the differential equations of second order,

$$u_{xx} = R(x, u, u_x) \qquad [6]$$

(where $R$ is analytic in $x$ and rational in $u$ and $u_x$) whose solutions do not have branch-type movable singularities. From the 50 equations (up to local transformation) discovered by them only six are not reduced to linear ones. These are the so-called Painlevé equations. They appear often as isomonodromy conditions for families of linear differential equations and this has given the idea to develop the isomonodromic deformation method. It consists in associating with eqn [6] a linear system

$$\mathrm{d}\Psi/\mathrm{d}\lambda = A(\lambda, x, u, u_x)\Psi \qquad [7]$$

with matrix-valued coefficients rational in $\lambda$. The deformation of the coefficients in $x$ is described by eqn [6] in such a way that the monodromy data of system [7] remain the same. Thus, the monodromy data of system [7] are first integrals of eqn [6].

**Example 10** The Painlevé II equation

$$u_{xx} - xu - 2u^3 = \nu$$

is associated with the system

$$\frac{\mathrm{d}\Psi}{\mathrm{d}\lambda} = \begin{pmatrix} -4i\lambda^2 - ix - 2iu^2 & 4i\lambda u - 2u_x - \dfrac{i\nu}{\lambda} \\ -4i\lambda u - 2u_x + \dfrac{i\nu}{\lambda} & 4i\lambda^2 + ix + 2iu^2 \end{pmatrix}\Psi$$

The idea to present the Painlevé equations as isomonodromy conditions originate from the works of Fuchs (1907) and Garnier (1912). It has been used, for example, in the papers of Flaschka and Newell (1980), Jimbo and Miwa (1981), and Its and Novokshenov (1986).

*See also:* Holonomic Quantum Fields; Integrable Systems: Overview; Painlevé Equations; Riemann–Hilbert Problem; WDVV Equations and Frobenius Manifolds.

## Further Reading

Arnol'd VI and Ilyashenko YuS (1988) Ordinary differential equations. In: *Dynamical Systems I*, Encyclopedia of Mathematical Sciences, t. 1. Berlin: Springer.

Bolibrukh AA (1997) On isomonodromic deformations of Fuchsian systems. *Journal of Dynamical and Control Systems* 3(4): 589–604.

Bolibrukh AA (1998) On isomonodromic confluences of Fuchsian singularities. *Proceedings of the Steklov Institute of Mathematics* 221: 117–132 (translation from *Trudy Matematicheskogo Instituta Imeni Steklova* 221: 127–142 (1998)).

Bolibrukh AA (2000) On orders of movable poles in Schlesinger's equation. *Journal of Dynamical and Control Systems* 6(1): 57–73.

Bolibrukh AA (2001) Regular singular points as isomonodromic confluences of Fuchsian singular points. *Russian Mathematical*

*Surveys* 56(4): 745–746 (translation from *Uspekhi Matematicheskikh Nauk* 56(4): 135–136 (2001)).

Flaschka H and Newell AC (1980) Monodromy and spectrum preserving deformations I. *Communications in Mathematical Physics* 76: 67–116.

Fokas AS and Ablowitz MJ (1982) On a unified approach to transformations and elementary solutions of Painlevé equations. *Journal of Mathematical Physics* 23(11): 2033–2042.

Fuchs R (1907) *Mathematical Annals* 63: 301–321.

Garnier R (1912) *Annales Scientifiques de l'Ecole Normale Supérieure* 29: 1–126.

Its AR and Novokshenov VYu (1986) *The Isomonodromic Deformation Method in the Theory of Painlevé Equations*, Lecture Notes in Mathematics, vol. 1191, p. 313. Berlin: Springer.

Jimbo M and Miwa T (1981) Monodromy preserving deformations of linear ordinary differential equations with rational coefficients, II. *Physica D* 2: 407–448.

Lin C-H and Sibuya Y (1990) Some applications of isomonodromic deformations to the study of Stokes multipliers. *Journal of the Faculty of Sciences, University of Tokyo, Section IA* 36(3): 649–663.

Malgrange B (1983) Sur les déformations isomonodromiques. I. *Singularités régulières*, pp. 401–426. II. *Singularités*irrégulières. Mathematics *and Physics* (Paris, 1979/1982), Progress in Mathematics, vol. 37. Boston, MA: Birkhäuser.

Schlesinger L (1912) Über eine Klasse von Differentialsystemen beliebiger Ordnung mit festen kritischen Punkten. *J. Reine Angew. Math* 141: 96–145.

Sibuya Y (1990) *Linear Differential Equations in the Complex Domain: Problems of Analytic Continuation*. Providence, RI: American Mathematical Society.

Ueno K (1980) Monodromy preserving deformations of linear differential equations with irregular singular points. *Proceedings of the Japanese Academy Series A Mathematical Sciences* 56(3): 97–102.

# J

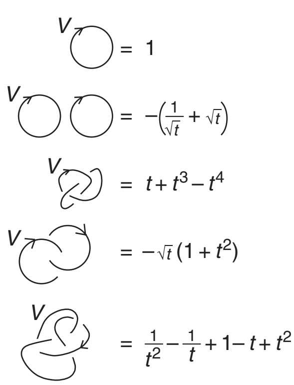
## The Jones Polynomial

**V F R Jones**, University of California at Berkeley, Berkeley, CA, USA



### Introduction

A “link” is a finite family of disjoint, smooth, oriented or unoriented, closed curves in $\mathbb{R}^3$ or equivalently $S^3$. A “knot” is a link with one component. The “Jones polynomial” $V_L(t)$ is a Laurent polynomial in the variable $\sqrt{t}$ which is defined for every oriented link $L$ but depends on that link only up to orientation-preserving diffeomorphism, or equivalently isotopy, of $\mathbb{R}^3$. Links can be represented by diagrams in the plane and the Jones polynomials of the simplest links are given below.

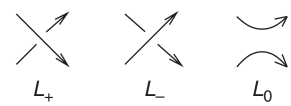



The Jones polynomial of a knot (and generally a link with an odd number of components) is a Laurent polynomial in $t$.

The most elementary ways to calculate $V_L(t)$ use the “linear skein theory” ideas of Conway (1970). Indeed, it is not hard to see by induction that $V_L(t)$ is defined by its invariance under isotopy, the normalization $V_{\bigcirc}(t)=1$ and the skein formula

$$\frac{1}{t}V_{L_+} - tV_{L_-} = \left(\sqrt{t} - \frac{1}{\sqrt{t}}\right)V_{L_0}$$

which holds for any three oriented links having diagrams which are identical except near one crossing where they differ as below.




As such the Jones polynomial resembles the Alexander (1928) polynomial $\Delta_L(t)$ which can be calculated in exactly the same manner as $V_L(t)$ except that the skein relation becomes

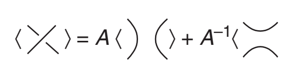
$$\Delta_{L_+} - \Delta_{L_-} = \left(\sqrt{t} - \frac{1}{\sqrt{t}}\right)\Delta_{L_0}$$

A two-variable generalization $P_L$ of both $\Delta_L$ and $V_L$, sometimes called the HOMFLYPT polynomial, was found in Freyd *et al.* (1985) and Przytycki and Traczyk (1988). It satisfies the most general skein relation

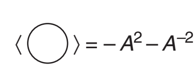
$$xP_{L_+} + yP_{L_-} + zP_{L_0} = 0$$

for homogeneous variables $x, y$, and $z$.

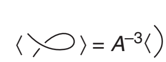
The other skein-like definition of $V_L$ was found in Kauffman (1987). Begin with unoriented link diagrams up to planar istotopy. The Kauffman bracket $\langle L\rangle$ of such a diagram is calculated using




where the $\langle\cdot\rangle$ notation means that the relation may be applied to that part of the link diagrams inside the bracket, the rest of the diagrams being identical. If $\langle L\rangle$ were to be an invariant of three-dimensional isotopy it is easy to see that




which further implies




Thus, $\langle L\rangle$ cannot be a three-dimensional isotopy invariant as such. However, if $L$ is given an

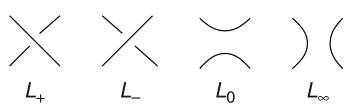
orientation (then called $\vec{L}$), a simple renormalization solves the problem and it is true that

$$(*) \qquad V_L(A^4) = A^{-3\,\mathrm{writhe}\,(\vec{L})}\langle L\rangle$$

where writhe ($\vec{L}$) is the sum over the crossings of $L$ of +1 for a positive crossing ( ) and –1 for a negative crossing ( ).

The formula ($*$) is readily proved by induction but a more structural proof will be discussed later on, connected with physics. If the crossings in a link alternate between over and under as one follows the string around, the highest and lowest degree terms in the Kauffman bracket can readily be located. This led to the proof of some old conjectures about alternating knots in Murasugi (1987), Kauffman (1987), and Thistlethwaite (1987).

The Kauffman two-variable polynomial $F_L(a,x)$ is defined in Kauffman (1990) by considering the linear skein relation involving all four possibilities at a crossing:

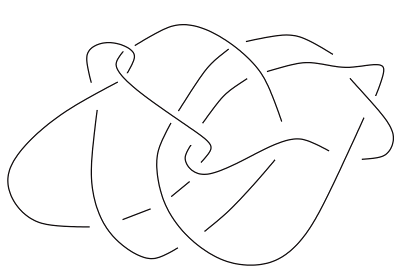



This polynomial contains $V_L(T)$ as a specialization but not the Alexander polynomial.

The above polynomials are quite powerful at distinguishing links one from another, including links from their mirror images, which corresponds for the Jones polynomial to replacing $t$ by $t^{-1}$. More power can be added to the polynomials if simple geometric operations are allowed. "Cabling" entails replacing a single strand with several parallel copies and the polynomials of cables of a link are also isotopy invariants if attention is paid to the writhe of a diagram.

The following problem, however, is open at the time of writing this article: "Does there exist a knot in $\mathbb{R}^3$, different from the unknot ○, whose Jones polynomial is equal to 1?"

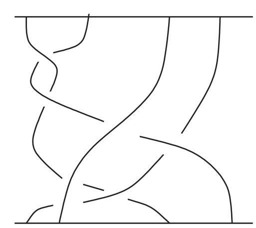
For links with more than one component, it is known (Thistlethwaite 2001, Eliahou *et al.* 2003) that the answer to the corresponding question is yes, the simplest example being:



One of the reasons that the question above has not been answered is presumably that, unlike with the Alexander polynomial, we have little intuitive understanding of the meaning of the "$t$" in $V_L(t)$. Perhaps, the most promising theory in this context is in Khovanov (2000) where a complex is constructed whose Euler characteristic, in an appropriately graded sense, is the Jones polynomial. The homology of the complex is a finer invariant of links known as "Khovanov homology."

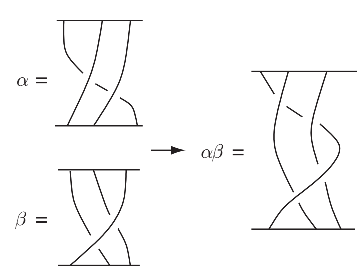
## Braids

A braid (see Birman (1974)) on $n$ strings is a collection of curves in $\mathbb{R}^3$ joining $n$ points in a horizontal plane to the $n$ points directly below them on another horizontal plane. If the endpoints of the braid are on a straight line, the braid can be drawn as in the example below (where $n = 4$).

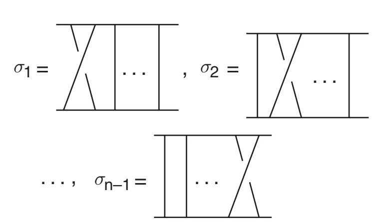


The crucial property of a braid is that the tangent vector to the curves can never be horizontal. Braids are considered up to isotopies which are supported between the top and bottom planes.

Braids on $n$ strings form a group, called $B_n$, under concatenation (plus some isotopy) as below:




Let $\sigma_1, \sigma_2, \ldots, \sigma_{n-1}$ be the braids below:



Artin's presentation (Birman 1974) of the braid group is on the generators $\sigma_1, \sigma_2, \ldots, \sigma_{n-1}$ with the relations

$$\sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1} \quad \text{for } 1 \le i \le n-2$$
$$\sigma_i\sigma_j = \sigma_j\sigma_i \quad \text{if } |i-j| \ge 2$$

Thus, to find linear representations of $B_n$, it suffices to find matrices $\rho_1, \rho_2, \ldots, \rho_{n-1}$ satisfying the above relations (with $\sigma$ replaced by $\rho$). One such representation (of dimension $n$) called the (nonreduced) Burau representation is given by the row-stochastic matrices

$$\rho_1 = \begin{pmatrix} 1-t & t & 0 & 0 & \ldots \\ 1 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 1 & 0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}$$

$$\rho_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & \ldots \\ 0 & 1-t & t & 0 & \ldots \\ 0 & 1 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}, \quad \ldots$$

$$\ldots, \quad \rho_{n-1} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 1-t & t \\ 0 & \ldots & 1 & 0 \end{pmatrix}$$

This representation is known not to be faithful for $n \ge 5$ but faithful for $n \le 3$. The case $n = 4$ remains open. (See Moody (1991), Long and Paton (1993), and Bigelow (1999)).

Braids can be viewed in several ways, which lead to several generalizations. For instance, identifying the vertical axis for a braid with time and taking the intersection of horizontal planes with the braids shows that elements of $B_n$ can be thought of as motions of $n$ distinct points in the plane. Thus, it is natural that

$$B_n \cong \pi_1(\{\mathbb{C}^n \backslash \Delta\}/S^n)$$

when $\Delta$ is the set $\{(z_1, \ldots, z_n) | z_i = z_j \text{ for some } i \ne j\}$ and the symmetric group $S_n$ acts freely on $\mathbb{C}^n \backslash \Delta$ by permuting coordinates. But $\Delta$ is the zero-set of the frequently encountered function

$$\prod_{i<j}(z_i - z_j)$$

so the braid group may naturally be generalized as the fundamental group of $\mathbb{C}^n$ minus the singular set of some algebraic function (Birman 1974). Or, motions of points can be extended to motions of the whole plane and a braid defines a diffeomorphism of the plane minus $n$ points. Thus, the braid group may be generalized as the "mapping class group" of a surface with marked points (Birman 1974).

## The Temperley–Lieb Algebra

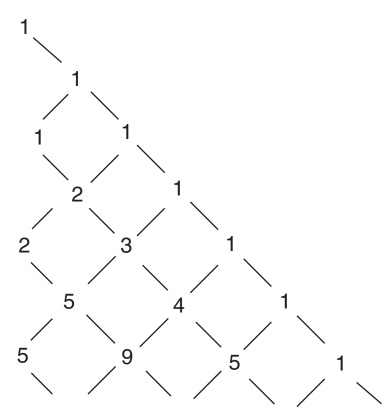
If $\tau \in \mathbb{C}$ one may define the algebra $TL(n, \tau)$ with identity 1 and generators $e_1, e_2, \ldots, e_{n-1}$ subject to the following relations:

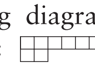
$$e_i^2 = e_i$$
$$e_i e_{i\pm1} e_i = \tau e_i$$
$$e_i e_j = e_j e_i \quad \text{if } |i-j| \ge 2$$

Counting reduced words on the $e_i$'s shows that

$$\dim\{TL(n, \tau)\} \le \frac{1}{n+1}\binom{2n}{n}$$

and in Jones (1983) it is shown that these numbers, the Catalan numbers, are indeed the dimensions of the Temperley–Lieb algebras. In the obvious way, $TL(n, \tau) \subseteq TL(n+1, \tau)$. If $\tau^{-1}$ is not in the set $\{4\cos^2 q\pi; q \in \mathbb{Q}\}, TL(n, \tau)$ is semisimple and its structure is given by the following Bratteli diagram:




where the integers on each row are the dimensions of the irreducible representations of $TL(n, \tau)$ and the diagonal lines give the restriction of representations of $TL(n, \tau)$ to $TL(n-1, \tau)$. These representations are naturally indexed by Young diagrams with $n$ boxes and at most two rows:  with the diagonal lines in the Bratteli diagram corresponding to removal/addition of a box. The dimension of the representation corresponding to the diagram whose second row has $r$ boxes ($r \le n$) is

$$\binom{n}{r} - \binom{n}{r-1}$$

One may attempt to make $TL(n,\tau)$ into a C*-algebra and look for Hilbert space representations (with $e_i \neq 0$), by imposing $e_i^* = e_i$. From (Wenzl 1987), this is only possible (for all $n$) when

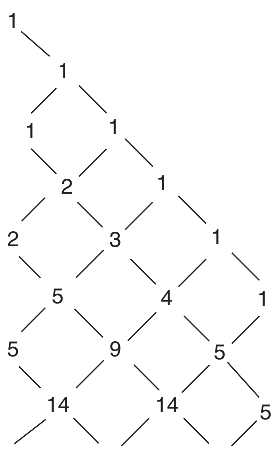
1. $\tau \in \mathbb{R}$, $0 < \tau \leq 1/4$, or
2. $\tau^{-1} \in \{4\cos^2 \pi/m,\ m = 3,4,5,\ldots\}$.

The proof uses the fact that $f_n$, inductively defined by

$$f_{n+1} = f_n - \frac{[2]_q[n+1]_q}{[n+2]_q} f_n e_{n+1} f_n$$

must be an orthogonal projection with $e_i f_n = f_n e_i = 0$ for $i \leq n$. These $f_n$ are sometimes called Jones–Wenzl idempotents. (Here $\tau^{-1} = 2 + q^2 + q^{-2}$ and for this and later formulas we define the quantum integer $[n]_q = (q^n - q^{-n})/(q - q^{-1})$).

When $\tau^{-1} = 4\cos^2(\pi/m)$, the Hilbert space representations decompose according to Bratteli diagrams obtained by truncating – eliminating the 1 on the $m$th row, and all representations below and to the right of it, so that for $m = 7$ we would obtain




In terms of Young diagrams, this corresponds to only taking those diagrams whose row lengths differ by at most $m-2$. The existence of these Hilbert space representations is from Jones (1983).

The Temperley–Lieb algebras arose in Jones (1983) as orthogonal projections onto subfactors of $II_1$ factors. As such the Hilbert space structure was manifest. The trace on a $II_1$ factor also yielded a trace on the $TL(n,\tau)$.

To be precise, there is for each $m$ a unique linear map $\mathrm{tr} : TL(n,\tau) \to \mathbb{C}$ with:

1. $\mathrm{tr}(1) = 1$
2. $\mathrm{tr}(ab) = \mathrm{tr}(ba)$
3. $\mathrm{tr}(xe_{n+1}) = \tau\,\mathrm{tr}(x)$ for $x \in TL(n+1,\tau)$.

This trace may be calculated either from (1), (2), and (3), or using the representations, as a weighted sum of ordinary matrix traces. The weight for the representation of $TL(n,\tau)$, the second row of whose Young diagram has $r$ boxes, is

$$\frac{[n-r+1]_q}{([2]_q)^n}$$

Thus, if $x \in TL(n,\tau)$ and $\pi_r$ is the $\binom{n}{r} - \binom{n}{r-1}$ dimensional irreducible representation, then

$$\mathrm{tr}(x) = \frac{1}{(q+q^{-1})^n} \sum_{r=0}^{[n/2]} [n-r+1]_q \,\mathrm{trace}\,(\pi_r(x))$$

One also has

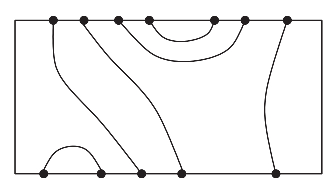
$$\mathrm{tr}(f_n) = \frac{[n+2]_q}{([2]_q)^{n+1}}$$

so that the disappearance of the "1" from the Bratteli diagram is mirrored by the vanishing of the trace of the corresponding projection.

Positivity of tr, $\mathrm{tr}(a^*a) \geq 0$, is responsible for all the Hilbert space structures. To explicitly construct the Hilbert space representations, one may use the GNS construction: take the quotient of the $*$-algebra by the kernel of the form $\langle a,b \rangle = \mathrm{tr}(b^*a)$ which makes this quotient a Hilbert space on which $TL(n,\tau)$ will act with the $e_i$'s as orthogonal projections. Explicit bases can be obtained easily if desired, using paths on the Bratteli diagram, or Young tableaux.

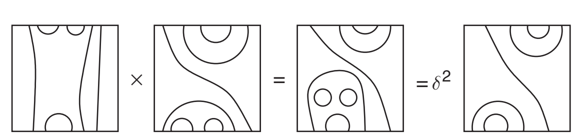
A useful diagrammatic presentation of $TL(n,\tau)$ was discovered in Kauffman (1987). A (Kauffman) TL diagram (for non-negative integers $m$ and $n$) is a rectangle with $n$ marked points on the top and $m$ on the bottom with nonintersecting smooth curves inside the rectangle connecting the boundary points as illustrated below.

A (5, 7)-diagram

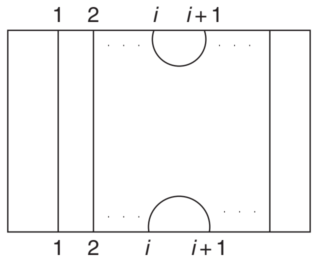


Two Kauffman TL diagrams are considered the same if they connect the same pairs of boundary points.

The vector space $TL(m,n,\delta)$ with basis the set of $(m,n)$ diagrams, and $\delta \in \mathbb{C}$, becomes a category with this concatenation together with the rule that closed curves may be removed, each one counting a (multiplicative) factor of $\delta$. We illustrate their product in $TL(m,n,\delta)$ below:

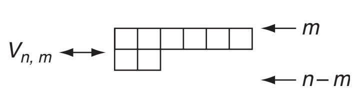


Of special interest is the algebra $TL(n, n, \delta)$. If we define $E_i$ to be the diagram below:




then $E_i^2 = \delta E_i$, $E_i E_{i\pm 1} E_i = E_i$, and $E_i E_j = E_j E_i$ for $|i - j| \geq 2$. Thus, provided $\delta \neq 0$, we have an isomorphism between $TL(n, \delta^{-2})$ and $TL(n, n, \delta)$ by mapping $e_i$ to $(1/\delta)E_i$.

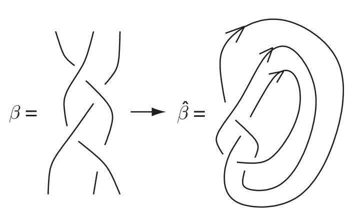
One of the nicest features of the Kauffman diagrams is that they yield simple explicit bases for the irreducible representations. To see this, call a curve in a diagram a "through-string" if it connects the top of the rectangle to the bottom. Then all $(m, n)$ diagrams are filtered by the number of through-strings and if we let $TL(m, n, k, \delta)$ be the span of $(m, n)$ diagrams with at most $k$ through-strings, we have $TL(k, n, \delta)TL(n, m, k, \delta) \subseteq TL(k, m, k, \delta)$. Thus, $V_{n,m} = TL(n, m, m, \delta)/TL(n, m, m-1, \delta)$ is a $TL(n, \delta^{-2})$-module, a basis of which is given by $(m, n)$-diagrams with $m$ through-strings $(m \leq n)$. The number of such diagrams is $\binom{n}{m} - \binom{n}{m-1}$ and it follows from Jones (1983) that all these representations are irreducible for "generic" $\delta$ (i.e., $\delta \notin \{2 \cos \mathbb{Q}\pi\}$) and that they may be identified with those indexed by Young diagrams as below:




The invariant inner product on $V_{n,m}$ is defined by $\langle v, w \rangle = w^* v$ for the natural identification of $V_{m,m}$ with $\mathbb{C}$ ($^*$ is the obvious involution from $(m, n)$ diagrams to $(n, m)$ diagrams.).

## The Original Definition of $V_L(t)$

Given a braid $\beta \in B_n$ one may form an oriented link $\hat{\beta}$ called the closure of $\beta$ by tying the top of the braid to the bottom as illustrated below:




All oriented links occur in this way (Birman 1974) but if $\alpha \in B_n$, $\alpha\beta\alpha^{-1}$ and $\beta\sigma_n^{\pm 1}$ (in $B_{n+1}$) have the same closure.

**Theorem 1** (Markov) (*Birman 1974*). *Let $\sim$ be the equivalence relation on $\coprod_{n=1}^{\infty} B_n$ (all braids on any number of strings) generated by the two "moves" $\beta \sim \beta\sigma_n^{\pm 1}$ and $\beta \sim \alpha\beta\alpha^{-1}$. Then $\beta_1 \sim \beta_2$ if and only if the links $\hat{\beta}_1$ and $\hat{\beta}_2$ are the same.*

It is easily checked that, if $1, e_1, e_2, e_3, \ldots$ satisfy the TL relations of the section "The Temperley–Lieb algebra," then sending $\sigma_i$ to $(t + 1)e_i - 1$ (with $\tau^{-1} = 2 + t + t^{-1}$) defines a representation $\rho_n$ of $B_n$ inside $TL(n, \tau)$ for each $n$. The representation is unitary for the $C^*$-algebra structure when $\tau^{-1} = 4 \cos^2 \pi/n$, $n = 3, 4, 5, \ldots$ (and $t = e^{\pm 2\pi i/n}$). It is an open question whether $\rho_n$ is faithful for all $n$. It contains the Burau representation as a direct summand.

Combining the properties of the trace tr defined on TL with Markov's theorem, one obtains immediately that, for $\alpha \in B_n$, the following function of $t$ depends only on $\hat{\alpha}$:

$$\left(-\sqrt{t} - \frac{1}{\sqrt{t}}\right)^{n-1} \sqrt{t}^{\,-e} \operatorname{tr}(\rho_n(\alpha))$$

(here $e \in \mathbb{Z}$ is the "exponent sum" of $\alpha$ as a word on $\sigma_1, \sigma_2, \ldots, \sigma_{n-1}$).

A simple check using the (oriented) skein-theoretic definition of the Jones polynomial shows that this function of $t$ is precisely $V_{\hat{\alpha}}(t)$. This is how $V_L(t)$ was first discovered in Jones (1985).

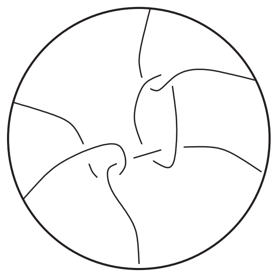
Although less elementary, this approach to $V_L(t)$ does have some advantages. Let us mention a few.

1. One may use representation theory to do calculations. For instance, using the weighted sum of ordinary traces to calculate tr as in the section "The Temperley–Lieb algebra," one obtains readily the Jones polynomial of a torus knot (i.e., $\hat{\alpha}$ where $\alpha = (\sigma_1\sigma_2 \cdots \sigma_{p-1})^q \in B_p$ if $p$ and $q$ are relatively prime). It is

$$\frac{t^{(p-1)(q-1)/2}}{1 - t^2}(1 - t^{p+1} - t^{q+1} - t^{p+q})$$

2. If one restricts attention to links realizable as $\hat{\alpha}$ for $\alpha \in B_n$ for fixed $n$, the computation of $V_{\hat{\alpha}}(t)$ can be performed in polynomial time as a function of the number of crossings in $\hat{\alpha}$. Thus, one has computational access to rather complicated families of links.
3. Unitarity of the representation when $t = e^{\pm\frac{2\pi i}{n}}$ can be used to bound the size of $|V_L(t)|$. For instance, if $\alpha \in B_k$ and $V_{\hat{\alpha}}(t) = (-\sqrt{t} - (1/\sqrt{t}))^{k-1}$, then $\alpha$ is in the kernel of $\rho_n$, and $|V_{\hat{\beta}}(e^{\pm 2\pi i/n})| \leq (2 \cos \pi/n)^{k-1}$ for any other $\beta \in B_k$.

The representation of the braid group inside the TL algebra should be thought of as an extension of the Jones polynomial to "special knots with boundary." The coefficients of the words in the $e_i$'s (or equivalently the Kauffman TL diagrams) are all invariants of the braid. We can further remove the braid restriction and consider arbitrary knots and links with boundary, known as "tangles" (Conway 1970).




Tangles may be oriented or not and their invariants may be evaluated either by reduction to a system of elementary tangles using skein relations or by organizing the tangle and representing it in an algebra. See Turaev (1994).

A similar algebraic approach is available for the HOMFLYPT and Kauffman two-variable polynomials. The algebra playing the role of the TL algebra is the Hecke algebra for HOMFLYPT (Freyd *et al.* 1985, Jones 1987) and the BMW algebra (Birman and Wenzl 1989, Murakami 1990) for the Kauffman polynomial. The BMW algebra was discovered after the Kauffman polynomial in order to provide an analog of the TL and Hecke algebras. For detailed analysis of the Hilbert space and other structures for both Hecke and BMW algebras, see Wenzl (1988) and Wenzl (1990).

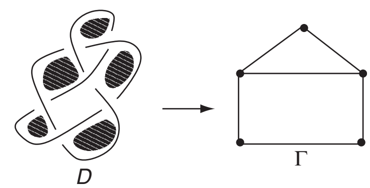
## Connections with Statistical Mechanics

One might say that turning a knot into a braid organizes the knot by "putting it on a lattice," thereby creating a physical model with the crossings of the knot as interactions. Taking the trace of the braid is evaluating the partition function with periodic (vertical) boundary conditions.

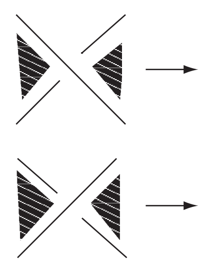
This is more than wishful thinking. The Temperley–Lieb algebra arose from transfer matrices in both the Potts and ice-type models in two dimensions (Temperley and Lieb 1971) and each "$e_i$" implements the addition of one more interaction to the system. (The same $e_i$'s as in the ice-type models were rediscovered in the subfactor context in Pimsner and Popa (1986)). Thus, the Jones polynomial of a closed braid is the partition function for a statistical mechanical model on the braid. In Jones (1983), it is observed that knowledge of the Jones polynomial for a family of links called French sinnets would constitute a solution of the Potts model in two dimensions.

In Temperley and Lieb (1971), the TL relations are used to establish the mathematical equivalence of the Potts and ice-type (six-vertex) models. In Baxter (1982, chapter 12), this equivalence is shown for Potts models on an arbitrary planar graph. In view of this, it is not surprising that statistical mechanical models can be defined directly on link diagrams to give explicit formulas for $V_L(t)$ (and other invariants) as partition functions. This works most easily for the $Q$-state Potts model.

Given an unoriented link diagram $D$, shade the regions of the plane black and white and form the planar graph $\Gamma$ whose vertices are the black regions and whose edges are the crossings as below:




Assign + and − to each edge according to the following scheme:




Fix $Q \in \mathbb{N}$ and two symmetric matrices $w_\pm(a,b)$ for $1 \le a, b \le Q$. The partition function of the diagram is then

$$Z_D = \sum_{\text{states}} \prod_{\text{edges of } \Gamma} w_\pm(\sigma, \sigma')$$

where a "state" is a function from the vertices of $\Gamma$ to $\{1, 2, \ldots, Q\}$ and, given an edge of $\Gamma$ and a state, $\sigma$ and $\sigma'$ denote the values of the state at the ends of that edge ($w_+$ and $w_-$ are used according to the sign of the edge).

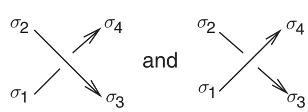
The "Potts model" is defined by the property that the "Boltzmann weights" $w_\pm(\sigma, \sigma')$ depend only on whether $\sigma = \sigma'$ or not. It is a miracle that the choice (with $Q = 2 + t + t^{-1}$)

$$w_\pm(\sigma, \sigma') = \begin{cases} t^{\pm 1} & \text{if } \sigma = \sigma' \\ -1 & \text{otherwise} \end{cases}$$

gives the Jones polynomial of the link defined by $D$ as its partition function (up to a simple normalization). See Jones (1989) for details.

It is natural to look for other choices of $w_\pm$ which give knot invariants. The Fateev–Zamolodchikov (1982) model gives a classical knot invariant but besides that (and some variants on the Jones polynomial) there is only one other known choice of any interest, discovered in Jaeger (1992). In this case, $Q=100$ and the Boltzmann weights are symmetric under the action of the Higman–Sims group on the Higman–Sims graph with 100 vertices. The knot invariant is a special value of the Kauffman two-variable polynomial.

The other side of Temperley–Lieb equivalence is the "ice-type" model which is a "vertex model." That is to say the "spins" reside on the edges of a graph and the interactions occur at the vertices. To use vertex models in knot theory, the knot projection $D$ itself is the (4-valent) graph. The ice-type model has two spin states per edge so that a state of the system is a function from the edges of the graph to the set $\{\pm\}$; the Boltzmann weights are given by two $4\times 4$ matrices $w_\pm(\sigma_1,\sigma_2,\sigma_3,\sigma_4)$ where the $\sigma$'s are $\pm 1$ and $w_+$ and $w_-$ are the contributions of




to the partition function, respectively. Furthermore, we may think of a state as a locally constant function $\sigma$ on $D$ so for any $f:\{\pm 1\}\to\mathbb{R}$ we may form the term $\int_D f(\sigma)\mathrm{d}\theta$ corresponding to interaction with an external field ($\mathrm{d}\theta$ is the curvature or change of angle form on $D$). Then the partition function is

$$Z_D=\sum_{\text{states}}\left(\prod_{\text{crossings of } D} w_\pm(\sigma_1,\sigma_2,\sigma_3,\sigma_4)\right)\mathrm{e}^{\int_D f(\sigma)\mathrm{d}\theta}$$

A (nonphysical) specialization of the six-vertex model yields values of $f$ and $w_\pm$ for which $Z_D$ is a link invariant equal to $V_L(t)$. See Jones (1989).

As with the Potts model, one may try to generalize to more general $w_\pm$ and $f$. This is much more successful for these "vertex" models than it was for models like the Potts model. The theory of quantum groups (Jimbo 1986, Drinfeld 1987, Rosso 1988) allows one to obtain link invariants (as partition functions for vertex models) for each simple finite-dimensional Lie algebra $\mathfrak{A}$ and each assignment of an irreducible representation of $\mathfrak{A}$ to the components of the link. The images of the braid generators $\sigma_i$ in the corresponding braid group representations are called "$R$-matrices." It is the Yang–Baxter equation that gives isotopy invariance of the partition function. In this way, one obtains (by an infinite family of one-variable specializations) the HOMFLYPT polynomial ($\mathrm{sl}_n$) and the Kauffman polynomial (orthogonal and symplectic algebras) and more polynomials. The geometric operation of cabling corresponds to the tensor product of representations.

## Connections with Quantum Field Theory

### Conformal Field Theory

If $\varphi$ is a (multicomponent) field in one chiral half of a two-dimensional conformal field theory (CFT), the correlation functions

$$\langle\varphi(z_1)\varphi(z_2)\cdots\varphi(z_n)\rangle$$

(where $z_i\in\mathbb{C}$) are expected to be singular if $z_i=z_j$ for some $i\neq j$, holomorphic otherwise and satisfy a linear differential equation. Thus, analytic continuation should determine a unitary monodromy representation of $\pi_1(\mathbb{C}^n\backslash\{(z_1,z_2,\ldots,z_n)|z_i=z_j$ for some $i\neq j\})$ on the vector space of solutions to the differential equation near a point. In Tsuchiya and Kanie (1988), these representations were calculated for the SU(2) WZW (Wess–Zumino–Witten) model, where the differential equation is known as the Khniznik–Zamolodchikov equation. The corresponding braid group representations were shown to be those obtained in the section "The original definition of $V_L(t)$" and cablings thereof.

### Topological Quantum Field Theory

In Witten (1989), the following formula appears:

$$\begin{aligned}V_L(\mathrm{e}^{2\pi\mathrm{i}/(k+2)})&\\&=\int_A\exp\left\{\frac{i}{\hbar}\int_{S^3}\mathrm{tr}(A\wedge\mathrm{d}A+2/3\,A\wedge A\wedge A)\right\}\\&\quad\times\prod_j\mathrm{tr}\left(\mathrm{Pexp}\oint_j A\right)[\mathfrak{D}A]\end{aligned}$$

where $A$ ranges over all functions from $S^3$ to the Lie algebra su(2), modulo the action of the gauge group SU(2). Also $\hbar=\pi/k$ and $j$ runs over the components of the link $L$, to each of which is assigned an irreducible representation of SU(2). Parallel transport around a component $j$ using $A$ yields the linear map $\mathrm{Pexp}\oint_j A$ whose trace is constant modulo gauge transformations. And $[\mathfrak{D}A]$ is a fictitious diffeomorphism invariant measure on all $A$'s modulo gauge transformation.

There are at least two ways to interpret this formula.

1. As a solvable topological quantum field theory (TQFT) in $2+1$ dimensions, according to Witten (1988) and Atiyah (1988, 1989). One is then obliged to expand the context and conclude that $V_L(e^{2\pi i/n})$ is defined for (possibly empty) links in an arbitrary 3-manifold. The TQFT axioms then provide an explicit formula for the invariant if the 3-manifold is obtained from surgery on a link. In particular, the invariant of a 3-manifold without a link is a statistical mechanics type sum over assignments of irreducible representations of SU(2) to the components of the surgery link. The key condition making this sum finite is that only representations up to a certain dimension (determined by $n$) are allowed. This is the vanishing of the Jones–Wenzl idempotent of the section "The Temperley–Lieb algebra." This explicit formula was rigourously shown to be a manifold invariant in Reshetikhin and Turaev (1991). For a more simple treatment, see Lickorish (1997) and for the whole TQFT treatment, see Blanchet *et al.* (1995).
2. As a perturbative QFT. The stationary-phase Feynman diagram technique may be applied to obtain the coefficients of the expansion of Witten's formula in powers of $\hbar$ or equivalently $1/n$. These coefficients are known to be "finite type" or Vassiliev invariants and have expressions as integrals over configurations of points on the link, see Vassiliev (1990) and Bar-Natan (1995).

### Algebraic Quantum Field Theory

In the Haag–Kastler operator algebraic framework of quantum field theory (Haag 1996), statistics of quantum systems were interpreted in Doplicher *et al.* (1971, 1974) (DHR) in terms of certain representations of the symmetric group corresponding to permuting regions of spacetime. To obtain the symmetric group, the dimension of spacetime needs to be sufficiently large. It was proposed in Fredenhagen *et al.* (1989) that the DHR theory should also work in low dimensions with the braid group replacing the symmetric group, and that unitary braid group representations defined above should be the ones occurring in quantum field theory. The "statistical dimension" of the DHR theory turns up as the square root of the index of a subfactor (this connection was clearly established in Longo (1989, 1980)). The mathematical issue of the existence of quantum fields with braid statistics was established in Wassermann (1998) using the language of loop group representations. Actual physical systems with nonabelian braid statistics have not yet been found but have been proposed in Freedman (2003) as a mechanism for quantum computing.

## Acknowledgments

The author would like to thank Tsou Sheung Tsun for her help in preparing this article.

The work of this author was supported in part by NSF Grant DMS 0401734, the NZIMA, and the Swiss National Science Foundation.

*See also:* Braided and Modular Tensor Categories; $C^*$-Algebras and their Classification; Hopf Algebras and *q*-Deformation Quantum Groups; Knot Homologies; Knot Invariants and Quantum Gravity; Knot Theory and Physics; Large-*N* and Topological Strings; Mathematical Knot Theory; Schwarz-Type Topological Quantum Field Theory; String Field Theory; Topological Knot Theory and Macroscopic Physics; Topological Quantum Field Theory: Overview; von Neumann Algebras: Introduction, Modular Theory, and Classification Theory; von Neumann Algebras: Subfactor Theory; Yang–Baxter Equations.

## Further Reading

Alexander JW (1928) Topological invariants of knots and links. *Transactions of the American Mathematical Society* 30(2): 275–306.

Atiyah M (1988) Topological quantum field theories. *Institut des Hautes Etudes Scientifiques Publications Mathématiques* 68: 175–186.

Bar-Natan D (1995) On the Vassiliev knot invariants. *Topology* 34(2): 423–472.

Baxter RJ (1982) *Exactly Solved Models in Statistical Mechanics*. London: Academic Press.

Bigelow S (1999) The Burau representation is not faithful for $n=5$. *Geometric Topology* 3: 397–404.

Birman JS (1974) Braids, Links, and Mapping Class Groups. *Annals of Mathematics Studies, No. 82*. Princeton, NJ: Tokyo: Princeton University Press; University of Tokyo Press.

Birman JS and Wenzel H (1989) Braids, link polynomials and a new algebra. *Transaction of the American Mathematical Society* 313(1): 249–273.

Blanchet C, Habegger N, Masbaum G, and Vogel P (1995) Topological quantum field theories derived from the Kauffman bracket. *Topology* 34(4): 883–927.

Conway JH (1970) An enumeration of knots and links, and some of their algebraic properties. *Computational Problems in Abstract Algebra*, Proc. Conf., Oxford, 1967 and pp. 329–358.

Doplicher S, Haag R, and Roberts JE (1971) Local observables and particle statistics, I. *Communications in Mathematical Physics* 23: 199–230.

Doplicher S, Haag R, and Roberts JE (1974) Local observables and particle statistics, II. *Communication in Mathematical Physics* 35: 49–85.

Drinfeld VG (1987) Quantum groups. In: *Proceedings of the International Congress of Mathematicians*, Berkeley, CA, 1986, vol. 1 and 2, pp. 798–820. Providence, RI: American Mathematical Society.

Eliahou S, Kauffman LH, and Thistlethwaite MB (2003) Infinite families of links with trivial Jones polynomial. *Topology* 42(1): 155–169.

Fateev VA and Zamolodchikov AB (1982) Self-dual solutions of the star-triangle relations in $Z_N$-models. *Physics Letters A* 92(1): 37–39.

Fredenhagen K, Rehren KH, and Schroer B (1989) Superselection sectors with braid group statistics and exchange algebras. *Communications in Mathematical Physics* 125: 201–226.

Freedman MH (2003) A magnetic model with a possible Chern–Simons phase. With an appendix by F. Goodman and H. Wenzl. *Communications in Mathematical Physics* 234(1): 129–183.

Freyd P, Yetter D, Hoste J, Lickorish WBR, Millett K, and Ocneanu A (1985) A new polynomial invariant of knots and links. *Bulletin of the American Mathematical Society (NS)* 12(2): 239–246.

Haag R (1996) *Local Quantum Physics*. Berlin: Springer.

Jaeger F (1992) Strongly regular graphs and spin models for the Kauffman polynomial. *Geometriae Dedicata* 44(1): 23–52.

Jimbo M (1986) A $q$-analogue of $U(gl(N+1))$, Hecke algebra, and the Yang–Baxter equation. *Letters in Mathematical Physics* 11(3): 247–252.

Jones VFR (1983) Index for subfactors. *Inventiones Mathematicae* 72: 1–25.

Jones VFR (1985) A polynomial invariant for knots via von Neumann algebras. *Bulletin of the American Mathematical Society* 12: 103–112.

Jones VFR (1987) Hecke algebra representations of braid groups and link polynomials. *Annals of Mathematics* 126(2): 335–388.

Jones V (1989) On knot invariants related to some statistical mechanical models. *Pacific Journal of Mathematics* 137: 311–388.

Kauffman LH (1987) State models and the Jones polynomial. *Topology* 26(3): 395–407.

Kauffman LH (1990) An invariant of regular isotopy. *Transactions of the American Mathematical Society* 318(2): 417–471.

Khovanov M (2000) A categorification of the Jones polynomial. *Duke Mathematical Journal* 101(3): 359–426.

Lickorish WBR (1997) An Introduction to Knot Theory, *Graduate Texts in Mathematics*, vol. 175. New York: Springer.

Long DD and Paton M (1993) The Burau representation is not faithful for $n \geq 6$. *Topology* 32(2): 439–447.

Longo R (1989) Index of subfactors and statistics of quantum fields I. *Communications in Mathematical Physics* 126: 217–247.

Longo R (1990) Index of sub factor and statistics of quantum fields II. *Communications in Mathematical Physics* 130: 285–309.

Moody JA (1991) The Burau representation of the braid group $B_n$ is unfaithful for large $n$. *Bulletin of the American Mathematical Society (NS)* 25(2): 379–384.

Murakami J (1990) The representations of the $q$-analogue of Brauer's centralizer algebras and the Kauffman polynomial of links. *Publ. Res. Inst. Math. Sci.* 26(6): 935–945.

Murasugi K (1987) Jones polynomials and classical conjectures in knot theory. *Topology* 26(2): 187–194.

Pimsner M and Popa S (1986) Entropy and index for subfactors. *Annales Scientifiques de l'Ecole Normale Supérieure* 19: 57–106.

Przytycki JH and Traczyk P (1988) Invariants of links of Conway type. *Kobe Journal of Mathematics* 4(2): 115–139.

Rosso M (1988) Groupes quantiques et modèles vertex de V. Jones en théorie des nœuds. (French) [Quantum groups and V. Jones's vertex models for knots]. *Comptes Rendus de l'Academie des Sciences Paris Séries I Mathematiques* 307(6): 207–210.

Reshetikhin N and Turaev VG (1991) Invariants of 3-manifolds via link polynomials and quantum groups. *Inventiones Mathematicae* 103(3): 547–597.

Temperley HNV and Lieb EH (1971) Relations between the 'percolation' and 'colouring' problem and other graph-theoretical problems associated with regular planar lattices: some exact results for the 'percolation' problem. *Proceedings of the Royal Society of London Series A* 322(1549): 251–280.

Thistlethwaite MB (1987) A spanning tree expansion of the Jones polynomial. *Topology* 26(3): 297–309.

Thistlethwaite M (2001) Links with trivial Jones polynomial. *Journal of Knot Theory and Its Ramifications* 10(4): 641–643.

Tsuchiya A and Kanie Y (1988) Vertex operators in conformal field theory on $P^1$ and monodromy representations of braid group. In: *Conformal Field Theory and Solvable Lattice Models*, (Kyoto, 1986), Adv. Stud. Pure Math., vol. 16, pp. 297–372. Boston, MA: Academic Press.

Turaev VG (1994) Quantum Invariants of Knots and 3-Manifolds. *de Gruyter Studies in Mathematics*, vol. 18. Berlin: Walter de Gruyter.

Vassiliev VA (1990) Cohomology of knot spaces. In: *Theory of Singularities and Its Applications*, Advances in Soviet Mathematics, vol. 1, pp. 23–69. Providence, RI: American Mathematical Society.

Wassermann A (1998) Operator algebras and conformal field theory. III. Fusion of positive energy representations of LSU($N$) using bounded operators. *Inventiones Mathematical* 133(3): 467–538.

Wenzl H (1987) On sequences of projections. *C. R. Math. Rep. Acad. Sci. Canada* 9(1): 5–9.

Wenzl H (1988) Hecke algebras of type $A_n$ and subfactors. *Inventiones Mathematical* 92(2): 349–383.

Wenzl H (1990) Quantum groups and subfactors of type $B$, $C$, and $D$. *Communications in Mathematical Physics* 133(2): 383–432.

Witten E (1988) Topological quantum field theory. *Communications in Mathematical Physics* 117(3): 353–386.

Witten E (1989) Quantum field theory and the Jones polynomial. *Communication in Mathematical Physics* 121(3): 351–399.

# K

**Kac–Moody Lie Algebras** *see* Solitons and Kac–Moody Lie Algebras

## KAM Theory and Celestial Mechanics

**L Chierchia**, Università degli Studi "Roma Tre," Rome, Italy



### Introduction

Kolmogorov–Arnol'd–Moser (KAM) theory deals with the construction of quasiperiodic trajectories in nearly integrable Hamiltonian systems and it was motivated by classical problems in celestial mechanics such as the $n$-body problem. Notwithstanding the formidable bulk of results, ideas and techniques produced by the founders of the modern theory of dynamical systems, most notably by H Poincaré and G D Birkhoff, the fundamental question about the persistence under small perturbations of invariant tori of an integrable Hamiltonian system remained completely open until 1954. In that year, A N Kolmogorov stated what is now usually referred to as the KAM theorem (in the real-analytic setting) and gave a precise outline of its proof, presenting a strikingly new and powerful method to overcome the so-called small-divisor problem (resonances in Hamiltonian dynamics produce, in the perturbation series, divisors which may become arbitrarily small, making convergence argument extremely delicate). Subsequently, KAM theory has been extended and applied to a large variety of different problems, including infinite-dimensional dynamical systems and partial differential equations with Hamiltonian structure. However, establishing the existence of quasiperiodic motions in the $n$-body problem turned out to be a longer story, which only very recently has reached a satisfactory level; the point being that the $n$-body problems present strong degeneracies, which violate the main hypotheses of the KAM theorem.

This article gives an account of the ideas and results concerning the construction of quasiperiodic solutions in the planetary $n$-body problem. The synopsis of the article is the following.

The next section gives the analytical description of the planetary $(1+n)$-body problem.

In the subsection "Kolmogorov's theorem and the RPC3BP (1954)," original version of the KAM theorem is recalled, giving an outline of its proof and showing its implications for the simplest many-body case, namely, the restricted, planar, and circular three-body problem.

In the section "Arnol'd's theorem," the existence of a positive measure set of initial data in phase space giving rise to quasiperiodic motions near coplanar and nearly circular unperturbed Keplerian trajectories is presented. The rest of the section is devoted to the proof of Arnol'd's theorem following the historical developments: Arnol'd's proof (1963a) for the planar three-body case is presented, the extension to the spatial three-body case due to Laskar and Robutel (1995) is discussed, and Herman's proof – in the form given by Féjoz in 2004 – of the general spatial $(1+n)$-case is presented.

In the section "Lower dimensional tori," a brief discussion of the construction of lower-dimensional elliptic tori bifurcating from the Keplerian unperturbed motions is given (these results have been established in the early 2000s).

Finally, the problem of taking into account real astronomical parameter values is considered and a recent result on an application of (computer-assisted) KAM techniques to the solar subsystem formed by Sun, Jupiter, and the asteroid Victoria is briefly mentioned.

### The Planetary $(1+n)$-Body Problem

The evolution of $(1+n)$-body systems (assimilated to point masses) interacting only through gravitational attraction is governed by Newton's equations.

If $u^{(i)} \in \mathbb{R}^3$ denotes the position of the $i$th body in a given reference frame and if $m_i$ denotes its mass, then Newton's equations read

$$\frac{d^2 u^{(i)}}{dt^2} = -\sum_{\substack{0 \le j \le n \\ j \ne i}} m_j \frac{u^{(i)} - u^{(j)}}{|u^{(i)} - u^{(j)}|^3}, \quad i = 0, 1, \ldots, n \quad [1]$$

Here the gravitational constant is taken to be equal to 1 (which amounts to rescale the time $t$). Equations [1] are equivalent to the standard Hamilton's equations corresponding to the Hamiltonian function

$$\mathcal{H}_{\text{New}} := \sum_{i=0}^{n} \frac{|U^{(i)}|^2}{2m_i} - \sum_{0 \le i < j \le n} \frac{m_i m_j}{|u^{(i)} - u^{(j)}|} \quad [2]$$

where $(U^{(i)}, u^{(i)})$ are standard symplectic variables and the phase space is the "collisionless domain" $\widehat{\mathcal{M}} := \{U^{(i)}, u^{(i)} \in \mathbb{R}^3 : u^{(i)} \ne u^{(j)}, 0 \le i \ne j \le n\}$; the symplectic form is the standard one: $\sum_i dU^{(i)} \wedge du^{(i)} := \sum_{i,k} dU_k^{(i)} \wedge du_k^{(i)}$; $|\cdot|$ denotes the standard Euclidean norm. Introducing the symplectic coordinate change $(U, u) = \phi_{\text{hel}}(R, r)$,

$$\phi_{\text{hel}} : \begin{cases} u^{(0)} = r^{(0)}, \quad u^{(i)} = r^{(0)} + r^{(i)} \quad (i = 1, \ldots, n) \\ U^{(0)} = R^{(0)} - \sum_{i=1}^{n} R^{(i)}, \quad U^{(i)} = R^{(i)} \\ (i = 1, \ldots, n) \end{cases} \quad [3]$$

one sees that the Hamiltonian $\mathcal{H}_{\text{hel}} := \mathcal{H}_{\text{New}} \circ \phi_{\text{hel}}$ does not depend upon $r^{(0)}$ (recall that a local diffeomorphism is called symplectic if it preserves the symplectic form). This means that $R^{(0)}$ ($\equiv$ total linear momentum) is a global integral of motion. Without loss of generality, one can restrict attention to the invariant manifold $\mathcal{M}_0 := \{R^{(0)} = 0\}$ (invariance of eqn [1] by changes of inertial reference frames).

In the "planetary" case, one assumes that one of the bodies, say $i = 0$ (the Sun), has mass much larger than that of the other bodies (this accounts for the index "hel," which stands for "heliocentric"). To make the perturbative character of the problem transparent, one may introduce the following rescalings. Let

$$m_i = \varepsilon \bar{m}_i, \quad X^{(i)} = \frac{R^{(i)}}{\varepsilon m_0^{5/3}}, \quad x^{(i)} = \frac{r^{(i)}}{m_0^{2/3}} \quad (i = 1, \ldots, n) \quad [4]$$

and rescale time by a factor $\varepsilon m_0^{7/3}$ (which amounts to dividing the new Hamiltonian by such a factor); then, the flow of the Hamiltonian $\mathcal{H}_{\text{hel}}$ on $\mathcal{M}_0$ is equivalent to the flow of the Hamiltonian

$$\mathcal{H}_{\text{plt}} := \sum_{i=1}^{n} \left( \frac{|X^{(i)}|^2}{2\mu_i} - \frac{\mu_i M_i}{|x^{(i)}|} \right) + \varepsilon \sum_{1 \le i < j \le n} \left( X^{(i)} \cdot X^{(j)} - \frac{\bar{m}_i \bar{m}_j / m_0^2}{|x^{(i)} - x^{(j)}|} \right) \quad [5]$$

on the phase space $\mathcal{M} := \{X^{(i)}, x^{(i)} \in \mathbb{R}^3 : 1 \le i \le n$ and $0 \ne x^{(i)} \ne x^{(j)}\}$ with respect to the standard symplectic form $\sum_{i=1}^{n} dX^{(i)} \wedge dx^{(i)}$; the mass parameters are defined as

$$M_i := 1 + \varepsilon \frac{\bar{m}_i}{m_0}, \qquad \mu_i := \frac{\bar{m}_i}{m_0 + \varepsilon \bar{m}_i} = \frac{\bar{m}_i}{m_0} \frac{1}{M_i} \quad [6]$$

The following observations can be made:

1. The Hamiltonian

$$\mathcal{H}_{\text{plt}}^{(0)} := \sum_{i=1}^{n} \left( \frac{|X^{(i)}|^2}{2\mu_i} - \frac{\mu_i M_i}{|x^{(i)}|} \right)$$

   is integrable and represents the sum of $n$ two-body systems formed by the Sun and the $i$th planet (disregarding the interaction with the other planets).
2. The transformation $\phi_{\text{hel}}$ in eqn [3] preserves the total angular momentum $\widehat{C} := \sum_{i=0}^{n} U^{(i)} \times u^{(i)}$, which is a vector-valued integral for $\mathcal{H}_{\text{New}}$. Thus, the three components, $C_k$, of $C := \sum_{i=1}^{n} X^{(i)} \times x^{(i)}$ (which is proportional to $\widehat{C}$ and is termed the "total angular momentum"), are integrals for $\mathcal{H}_{\text{plt}}$. The integrals $C_k$ do not commute: if $\{\cdot,\cdot\}$ denotes the standard Poisson bracket, then $\{C_1, C_2\} = C_3$ (and, cyclically, $\{C_2, C_3\} = C_1, \{C_3, C_1\} = C_2$). Nevertheless, one can form two (independent) commuting integrals, for example, $|C|^2$ and $C_3$. This shows that the (spatial) $(1+n)$-body problem has $(3n-2)$ degrees of freedom.
3. An important special case is the planar $(1+n)$-body problem. In such a case, one assumes that all the "single" angular momenta $C^{(i)} := X^{(i)} \times x^{(i)}$ are parallel. In this case, the motion takes place on a fixed plane orthogonal to $C$ and (up to a rotation of the reference frame) one can take, as symplectic variables, $X^{(i)}, x^{(i)} \in \mathbb{R}^2$. The Hamiltonian $\mathcal{H}_{\text{pln}}$ governing the dynamics of the planar $(1+n)$-body problem is, then, given on the right-hand side of eqn [5] with $X^{(i)}$, $x^{(i)} \in \mathbb{R}^2$. Notice that the planar $(1+n)$-body problem has $2n$ degrees of freedom.
4. For a deeper understanding of the perturbation theory of the planetary many-body problem, it is necessary to find "good" sets of symplectic coordinates, which the founders of celestial

mechanics (most notably, Jacobi, Delaunay, and Poincaré) have done. In particular, Delaunay introduced an analytic set of symplectic "action-angle" variables. Recall the Delaunay variables for the two-body "reduced Hamiltonian"

$$\mathcal{H}_{\text{Kep}} = \frac{|X|^2}{2\mu} - \frac{\mu M}{|x|}$$

Let $\{k_1, k_2, k_3\}$ be a standard orthonormal basis in the $x$-configuration space; let the angular momentum $C = X \times x$ be nonparallel to $k_3$ and let the energy $E = \mathcal{H}_{\text{Kep}} < 0$. In such a case, $x(t)$ describes an ellipse lying in the plane orthogonal to $C$, with focus in the origin and fixed symmetry axes. Let $a$ be the semimajor axis of the ellipse spanned by $x$; $\imath$ (the inclination) be the angle between $k_3$ and $C$; $G = |C|$; $\Theta = G \cos \imath = C \cdot k_3$; $L = m\sqrt{Ma}$; $\ell$ be the mean anomaly of $x$ ($:= 2\pi$ times the normalized area spanned by $x$ measured from the perihelion $P$, which is the point of the ellipse closest to the origin); $\theta$ be the angle between $k_1$ and $N := k_3 \times C$ (:= oriented "node"); and $g$ be the argument of the perihelion (:= the angle between $N$ and $(O, P)$). Then (letting $\mathbb{T} := \mathbb{R}/(2\pi\mathbb{Z})$)

$$\begin{aligned} &(L, G, \Theta) \in \{L > 0\} \times \{G > \Theta > 0\} \\ &(\ell, g, \theta) \in \mathbb{T}^3 \end{aligned} \qquad [7]$$

are conjugate symplectic coordinates and if $\phi_{\text{Del}}$ is the corresponding symplectic map, then $\mathcal{H}_{\text{Kep}} \circ \phi_{\text{Del}} = -(\mu^3 M^2)/(2L^2)$.

Note that the Delaunay variables become singular when C is vertical (the node is no more defined) and in the circular limit (the perihelion is not unique). In these cases different variables have to be used.

5. Let $(X^{(i)}, x^{(i)}) = \phi_{\text{Del}}((L_i, G_i, \Theta_i), (\ell_i, g_i, \theta_i))$. Then $\mathcal{H}_{\text{plt}}$ expressed in the Delaunay variables $\{(L_i, G_i, \Theta_i), (\ell_i, g_i, \theta_i): 1 \le i \le n\}$ becomes

$$\mathcal{H}_{\text{Del}} = \mathcal{H}^{(0)}_{\text{Del}} + \varepsilon \mathcal{H}^{(1)}_{\text{Del}}, \quad \mathcal{H}^{(0)}_{\text{Del}} := -\sum_{i=1}^{n} \frac{\mu_i^3 M_i^2}{2L_i^2} \qquad [8]$$

Note that the number of action variables on which the integrable Hamiltonian $\mathcal{H}^{(0)}_{\text{Del}}$ depends is strictly less than the number of degrees of freedom. This "proper degeneracy," as we shall see in next sections, brings in an essential difficulty one has to face in the perturbative approach to the many-body problem. In fact, this feature of the many-body problem is common to several other problems of celestial mechanics.

## Maximal KAM Tori

### Kolmogorov's Theorem and the RPC3BP (1954)

Kolmogorov's invariant tori theorem deals with the persistence, in nearly integrable Hamiltonian systems, of Lagrangian (maximal) tori, which, in general, foliate the integrable limit. Kolmogorov (1954) stated his theorem and gave a precise outline of the proof. Let us briefly recall this milestone of the modern theory of dynamical systems.

Let $\mathcal{M} := B^d \times \mathbb{T}^d$ ($B^d$ being a $d$-dimensional ball in $\mathbb{R}^d$ centered at the origin) be endowed with the standard symplectic form $\mathrm{d}y \wedge \mathrm{d}x := \sum \mathrm{d}y_i \wedge \mathrm{d}x_i$ ($y \in B^d$, $x \in \mathbb{T}^d$). A Hamiltonian function $N$ on $\mathcal{M}$ having a Lagrangian invariant $d$-torus of energy $E$ on which the $N$-flow is conjugated to the linear dense translation $x \to x + \omega t$, $\omega \in \mathbb{R}^d \backslash \mathbb{Q}^d$ can be put in the form

$$\begin{aligned} &N := E + \omega \cdot y + Q(y, x) \\ &\quad \partial_y^\alpha Q(0, x) = 0, \quad \forall \alpha \in \mathbb{N}^d, \quad |\alpha| \le 1 \end{aligned} \qquad [9]$$

(as usual, $|\alpha| = \alpha_1 + \cdots + \alpha_d$, $\omega \cdot y := \sum_{i=1}^d \omega_i y_i$, and $\partial_y^\alpha = \partial_{y_1}^{\alpha_1} \cdots \partial_{y_d}^{\alpha_d}$); in such a case, the Hamiltonian $N$ is said to be in Kolmogorov normal form. The vector $\omega$ is called the "frequency vector" of the invariant torus $\{y = 0\} \times \mathbb{T}^d$. The Hamiltonian $N$ is said to be nondegenerate if

$$\det \langle \partial_y^2 Q(0, \cdot) \rangle \neq 0 \qquad [10]$$

where the brackets denote average over $\mathbb{T}^d$ and $\partial_y^2$ the Hessian with respect to the $y$-variables.

We recall that a vector $\omega \in \mathbb{R}^d$ is said to be "Diophantine" if there exist $\kappa > 0$ and $\tau \ge d - 1$ such that

$$|\omega \cdot k| \ge \frac{\kappa}{|k|^\tau}, \quad \forall k \in \mathbb{Z}^d \backslash \{0\} \qquad [11]$$

The set $\mathcal{D}^d$ of all Diophantine vectors in $\mathbb{R}^d$ is a set of full Lebesgue measure. We also recall that Hamiltonian trajectory is called quasiperiodic with (rationally independent) frequency $\omega \in \mathbb{R}^d$ if it is conjugate to the linear translation $\theta \in \mathbb{T}^d \to \theta + \omega t \in \mathbb{T}^d$.

**Theorem** (Kolmogorov 1954) *Consider a one-parameter family of real-analytic Hamiltonian functions $H_\varepsilon := N + \varepsilon P$ where $N$ is in Kolmogorov normal form (as in eqn [9]) and $\varepsilon \in \mathbb{R}$. Assume that $\omega$ is Diophantine and that $N$ is nondegenerate. Then, there exists $\varepsilon_0 > 0$ and for any $|\varepsilon| \le \varepsilon_0$, a real-analytic symplectic transformation $\phi_\varepsilon : \mathcal{M} \to \mathcal{M}$ putting $H_\varepsilon$ in Kolmogorov normal form, $H_\varepsilon \circ \phi_\varepsilon = N_\varepsilon$, with $N_\varepsilon := E_\varepsilon + \omega \cdot y' + Q_\varepsilon(y', x')$. Furthermore, $|E_\varepsilon - E|$, $\|\phi_\varepsilon - \mathrm{id}\|_{C^2}$, and $\|Q_\varepsilon - Q\|_{C^2}$ are small with $\varepsilon$.*

In other words, the Lagrangian unperturbed torus $\mathcal{T}_0 := \{y = 0\} \times \mathbb{T}^d$ persists under small perturbation and is smoothly deformed into the $H_\varepsilon$-invariant torus $\mathcal{T}_\varepsilon := \phi_\varepsilon(\{y' = 0\} \times \mathbb{T}^d)$; the dynamics on such torus, for all $|\varepsilon| \le \varepsilon_0$, consists of dense quasiperiodic trajectories. Note that the $H_\varepsilon$-flow on $\mathcal{T}_\varepsilon$ is analytically conjugated by $\phi_\varepsilon$ to the translation $x' \to x' + \omega t$ with the same frequency vector of $N$, while the energy of $\mathcal{T}_\varepsilon$, namely $E_\varepsilon$, is in general different from the energy $E$ of $\mathcal{T}_0$.

Kolmogorov's proof is based on an iterative (Newton) scheme. The map $\phi_\varepsilon$ is obtained as $\lim_{k\to\infty} \phi^{(1)} \circ \cdots \circ \phi^{(k)}$, where the $\phi^{(j)}$'s are ($\varepsilon$-dependent) symplectic transformations of $\mathcal{M}$ successively closer to the identity. It is enough to describe the construction of $\phi^{(1)}$; $\phi^{(2)}$ is then obtained by replacing $H_\varepsilon$ with $H_\varepsilon \circ \phi^{(1)}$, and so on. The map $\phi^{(1)}$ is $\varepsilon$-close to the identity and it is generated by $g(y', x) := y' \cdot x + \varepsilon(b \cdot x + s(x) + y' \cdot a(x))$, where $s$ and $a$ are (resp. scalar- and vector-valued) real-analytic functions on $\mathbb{T}^d$ with zero average and $b \in \mathbb{R}^d$; this means that the symplectic map $\phi^{(1)} : (y', x') \to (y, x)$ is implicitly given by the relations $y = \partial_x g$ and $x' = \partial_{y'} g$. It is easy to see that there exists a unique $g$ of the above form such that for a suitable $\varepsilon_0 > 0$,

$$H_\varepsilon \circ \phi^{(1)} = E_1 + \omega \cdot y' + Q_1(y', x') + \varepsilon^2 P_1 \quad \forall\ |\varepsilon| \le \varepsilon_0 \qquad [12]$$

with $\partial_y^\alpha Q_1(0, x') = 0$, for any $\alpha \in \mathbb{N}^d$ and $|\alpha| \le 1$; here, $E_1, Q_1$, and $P_1$ depend on $\varepsilon$ and, for a suitable $c_1 > 0$ and for $|\varepsilon| \le \varepsilon_0$, $|E - E_1| \le c_1|\varepsilon|$, $\|Q - Q_1\|_{C^2} \le c_1|\varepsilon|$, and $\|P_1\|_{C^2} \le c_1$.

Notice that the symplectic transformation $\phi^{(1)}$ is actually the composition of two "elementary" transfomations: $\phi^{(1)} = \phi_1^{(1)} \circ \phi_2^{(1)}$ where $\phi_2^{(1)}:\ (y', x') \to (\eta, \xi)$ is the symplectic lift of the $\mathbb{T}^d$-diffeomorphism given by $x = \xi + \varepsilon a(\xi)$ (i.e., $\phi_2^{(1)}$ is the symplectic map generated by $y' \cdot \xi + \varepsilon y' \cdot a(\xi)$), while $\phi_1^{(1)} : (\eta, \xi) \to (y, x)$ is the angle-dependent action translation generated by $\eta \cdot x + \varepsilon(b \cdot x + s(x))$; $\phi_2^{(1)}$ acts in the "angle direction" and straightens out the flow up to order $O(\varepsilon^2)$, while $\phi_1^{(1)}$ acts in the "action direction" and is needed to keep the frequency of the torus fixed.

Since $H_\varepsilon \circ \phi_1 =: N_1 + \varepsilon^2 P_1$ is again a perturbation of a nondegenerate Kolmogorov normal form (with same frequency vector $\omega$), one can repeat the construction by obtaining a new Hamiltonian of the form $N_2 + \varepsilon^4 P_2$. Iterating, after $k$ steps, one gets a Hamiltonian $N_k + \varepsilon^{2^k} P_k$. Carrying out the (straightforward but lengthy) estimates, one can check that $\|P_k\|_{C^2} \le c_k \le c^{2^k}$, for a suitable constant $c > 1$ independent of $k$ (the fast growth of the constant $c_k$ is due to the presence of the small divisors appearing in the explicit construction of the symplectic transformations $\phi^{(j)}$). Thus, it is clear that taking $\varepsilon_0$ small enough the iterative procedure converges (superexponentially fast) yielding the thesis of the above theorem.

6. While the statement of the invariant tori theorem and the outline of the proof are very clearly explained in Kolmogorov (1954), Kolmogorov did not fill out the details nor gave any estimates. Some years later, Arnol'd (1963a) published a detailed proof, which, however, did not follow Kolmogorov's idea. In the same year, J K Moser published his invariant curve theorem (for area-preserving twist diffeomorphisms of the annulus) in smooth setting. The bulk of techniques and theorems stemmed out from these works is normally referred to as KAM theory; for reviews, see Arnol'd (1988) or Bost (1984–85). A very complete version of the "KAM theorem" both in the real-analytic and in the smooth case (with optimal smoothness assumptions) is given in Salamon (2004); the proof of the real-analytic part is based on Kolmogorov's scheme. The KAM theory of M Herman, used in his approach to the planetary problem, is based on the abstract functional theoretical approach of R Hamilton (which, in turn, is a development of Nash–Moser implicit function theorem; see Bost (1984–85) for references); it is interesting, however, to note that the heart of Herman's KAM method is based on the above-mentioned Kolmogorov's transformation $\phi^{(1)}$ (compare Féjoz (2002)).
7. In the nearly integrable case, one considers a one-parameter family of Hamiltonians $H_0(I) + \varepsilon H_1(I, x)$ with $(I, x) \in \mathcal{M} := U \times \mathbb{T}^d$ standard symplectic action-angle variables, $U$ being an open subset of $\mathbb{R}^d$. When $\varepsilon = 0$, the phase space $\mathcal{M}$ is foliated by $H_0$-invariant tori $\{I_0\} \times \mathbb{T}^d$, on which the flow is given by $x \to x + \partial_y H_0(I_0)t$. If $I_0$ is such that $\omega := \partial_y H_0(I_0)$ is Diophantine and if $\det \partial_y^2 H_0(I_0) \ne 0$, then from Kolmogorov's theorem it follows that the torus $\{I_0\} \times \mathbb{T}^d$ persists under perturbation. In fact, introduce the symplectic variables $(y,\ x)$ with $y = I - I_0$ and let $N(y) := H_0(I_0 + y)$, which by Taylor's formula can be written as $H_0(I_0) + \omega \cdot y + Q(y)$ with $Q(y)$ quadratic in $y$ and $\partial_y^2 Q(0) = \partial_y^2 H_0(I_0)$ invertible. One can then apply Kolmogorov's theorem with $P_1(y, x) := H_1(I_0 + y, x)$.

Notice that Kolmogorov's nondegeneracy condition $\det \partial_y^2 H_0(I_0) \ne 0$ simply means that the frequency map

$$I \in B^d \subset U \to \omega(I) := \partial_y H_0(I) \qquad [13]$$

is a local diffeomorphism ($B^d$ being a ball around $I_0$).

8. The symplectic structure implies that if $n$ denotes the number of degrees of freedom (i.e., half of the dimension of the phase space) and $d$ is the number of independent frequencies of a quasiperiodic motion, then $d \leq n$; if $d = n$, the quasiperiodic motion is called maximal. Kolmogorov's theorem gives sufficient conditions in order to get maximal quasiperiodic solutions. In fact, Kolmogorov's nondegeneracy condition is an open condition and the set of Diophantine vectors is a set of full Lebesgue measure. Thus, in general, Kolmogorov's theorem yields a positive invariant measure set spanned by maximal quasiperiodic trajectories.

As mentioned above, the planetary many-body models are properly degenerate and violate Kolmogorov's nondegeneracy conditions and, hence, Kolmogorov's theorem – clearly motivated by celestial mechanics – cannot be applied.

There is, however, an important case to which a slight variation of Kolmogorov's theorem can be applied (Kolmogorov did not mention this in 1954). The case referred to here is the simplest nontrivial three-body problem, namely, the restricted, planar, and circular three-body problem (RPC3BP for short). This model, largely investigated by Poincaré, deals with an asteroid of "zero mass" moving on the plane containing the trajectory of two unperturbed major bodies (say, Sun and Jupiter) revolving on a Keplerian circle. The mathematical model for the restricted three-body problem is obtained by taking $n = 2$ and setting $m_2 = 0$ in eqn [1]: the equations for the two major bodies ($i = 0, 1$) decouple from the equation for the asteroid ($i = 2$) and form an integrable two-body system; the problem then consists in studying the evolution of the asteroid $u^{(2)}(t)$ in the given gravitational field of the primaries. In the circular and planar cases, the motion of the two primaries is assumed to be circular and the motion of the asteroid is assumed to take place on the plane containing the motion of the two primaries; in fact (to avoid collisions), one considers either inner or outer (with respect to the circle described by the relative motion of the primaries) asteroid motions. To describe the Hamiltonian $\mathcal{H}_{\text{rcp}}$ governing the motion of the RCP3BP problem, introduce planar Delaunay variables $((L, G), (\ell, \hat{g}))$ for the asteroid (better, for the reduced heliocentric Sun–asteroid system). Such variables, which are closely related to the above (spatial) Delaunay variables, have the following physical interpretation: $G$ is proportional to the absolute value of the angular momentum of the asteroid, $L$ is proportional to the square root of the semimajor axis of the instantaneous Sun–asteroid ellipse, $\ell$ is the mean anomaly of the asteroid, while $\hat{g}$ the argument of the perihelion. Then, in suitably normalized units, the Hamiltonian governing the RPC3BP is given by

$$\mathcal{H}_{\text{rcp}}(L, G, \ell, g; \varepsilon) := -\frac{1}{2L^2} - G + \varepsilon \mathcal{H}_1(L, G, \ell, g; \varepsilon) \quad [14]$$

where $g := \hat{g} - \tau, \tau \in \mathbb{T}$ being the longitude of Jupiter; the variables $((L, G), (\ell, g))$ are symplectic coordinates (with respect to the standard symplectic form); the normalizations have been chosen so that the relative motion of the primary bodies is $2\pi$ periodic and their distance is 1; the parameter $\varepsilon$ is (essentially) the ratio between the masses of the primaries; the perturbation $\mathcal{H}_1$ is the function $x^{(2)} \cdot x^{(1)} - 1/|x^{(2)} - x^{(1)}|$ expressed in the above variables, $x^{(2)}$ being the heliocentric coordinate of the asteroid and $x^{(1)}$ that of the planet (Jupiter): such a function is real-analytic on $\{0 < G < L\} \times \mathbb{T}^2$ and for small $\varepsilon$ (for complete details, see, e.g., Celletti and Chierchia (2003)).

The integrable limit

$$\mathcal{H}_{\text{rcp}}^{(0)} := \mathcal{H}_{\text{rcp}}|_{\varepsilon = 0} = -1/(2L^2) - G$$

has vanishing Hessian and, hence, violates Kolmogorov's nondegeneracy condition (as described in item (7) above). However, there is another nondegeneracy condition which leads to a simple variation of Kolmogorov's theorem, as explained briefly below.

Kolmogorov's nondegeneracy condition $\det_y^2 H_0(I_0) \neq 0$ allows one to fix $d$-parameters, namely, the $d$-components of the (Diophantine) frequency vector $\omega = \partial_y H_0(I_0)$. Instead of fixing such parameters, one may fix the energy $E = H_0(I_0)$ together with the direction $\{s\omega : s \in \mathbb{R}\}$ of the frequency vector: for example, in a neighborhood where $\omega_d \neq 0$, one can fix $E$ and $\omega_i/\omega_d$ for $1 \leq i \leq d - 1$. Notice also that if $\omega$ is Diophantine, then so is $s\omega$ for any $s \neq 0$ (with same $\tau$ and rescaled $\kappa$). Now, it is easy to check that the map $I \in H_0^{-1}(E) \to (\omega_1/\omega_d, \ldots, \omega_{d-1}/\omega_d)$ is (at fixed energy $E$) a local diffeomorphism if and only if the $(d + 1) \times (d + 1)$ matrix

$$\begin{pmatrix} \partial_y^2 H_0 & \partial_y H_0 \\ \partial_y H_0 & 0 \end{pmatrix}$$

evaluated at $I_0$ is invertible (here the vector $\partial_y H_0$ in the upper right corner has to be interpreted as a column while the vector $\partial_y H_0$ in the lower left corner has to be interpreted as a row). Such

"iso-energetic nondegeneracy" condition, rephrased in terms of Kolmogorov's normal forms, becomes

$$\det\begin{pmatrix} \langle \partial_y^2 Q(0,\cdot)\rangle & \omega \\ \omega & 0 \end{pmatrix} \neq 0 \qquad [15]$$

Kolmogorov's theorem can be easily adapted to the fixed energy case. Assuming that $\omega$ is Diophantine and that $N$ is isoenergetically nondegenerate, the same conclusion as in Kolmogorov's theorem holds with $N_\varepsilon := E + \omega_\varepsilon \cdot y' + Q_\varepsilon(y', x')$, where $\omega_\varepsilon = \alpha_\varepsilon \omega$ and $|\alpha_\varepsilon - 1|$ is small with $\varepsilon$.

In the RCP3BP case, the isoenergetic nondegeneracy is met, since

$$\det\begin{pmatrix} \partial^2_{(L,G)}\mathcal{H}^{(0)}_{\rm rcp} & \partial_{(L,G)}\mathcal{H}^{(0)}_{\rm rcp} \\ \partial_{(L,G)}\mathcal{H}^{(0)}_{\rm rcp} & 0 \end{pmatrix} = \frac{3}{L^4}$$

Therefore, one can conclude that on each negative energy level, the RCP3BP admits a positive measure set of phase points, whose time evolution lies on two-dimensional invariant tori (on which the flow is analytically conjugate to linear translation by a Diophantine vector), provided the mass ratio of the primary bodies is small enough; such persistent tori are a slight deformation of the unperturbed "Keplerian" tori corresponding to the asteroid and the Sun revolving on a Keplerian ellipse on the plane where the Sun and the major planet describe a circular orbit.

In fact, one can say more. The phase space for the RCP3BP is four dimensional, the energy levels are three dimensional, and Kolmogorov's invariant tori are two dimensional. Thus, a Kolmogorov torus separates the energy level, on which it lies, into two invariant components, and two Kolmogorov's tori form the boundary of a compact invariant region so that any motion starting in such region will never leave it. Thus, the RCP3BP is "totally stable": in a neighborhood of any phase point of negative energy, if the mass ratio of the primary bodies is small enough, the asteroid stays forever on a nearly Keplerian ellipse with nearly fixed orbital elements $L$ and $G$.

## Arnol'd's Theorem

Consider again the planetary $(1+n)$-body problem governed by the Hamiltonian $\mathcal{H}_{\rm plt}$ in eqn [5]. In the integrable approximation, governed by the Hamiltonian $\mathcal{H}^{(0)}_{\rm plt}$, the $n$ planets describe Keplerian ellipses focused on the Sun. Arnol'd (1963b) has stated the following theorem.

**Theorem** (Arnol'd 1963b) *Let $\varepsilon > 0$ be small enough. Then, there exists a bounded, $\mathcal{H}_{\rm plt}$-invariant set $\mathcal{F}(\varepsilon) \subset \mathcal{M}$ of positive Lebesgue measure corresponding to planetary motions with bounded relative distances; $\mathcal{F}(0)$ corresponds to Keplerian ellipses with small eccentricities and small relative inclinations.*

This theorem represents a major achievement in celestial mechanics solving more than tri-ćentennial mathematical problem. Arnol'd (1963b) gave a complete proof of this result only in the planar three-body case and gave some indications of how to extend his approach to the general situation. However, to give a full proof of Arnol'd's theorem in the general case turned out to be more than a technical problem and new ideas were needed: the complete proof (due, essentially, to M Herman) has been given only in 2004.

In the following subsections, we briefly review the history and the ideas related to the proof of Arnol'd's theorem. As for credits: the proof of Arnol'd's theorem in the planar 3BP case is due to Arnol'd himself (Arnol'd 1963b); the spatial 3BP case is due to Laskar and Robutel (1995) and Robutel (1995); the general case is due to Herman (1998) and Féjoz (2004). The exposition we have given does not always follow the original references.

**The planar three-body problem** Recall the Hamiltonian $\mathcal{H}_{\rm pln}$ of the planar $(1+n)$-body problem given in item (3) of the section "The planetary $(1+n)$-body problem." A convenient set of symplectic variables for nearly circular motions are the "planar Poincaré variables." To describe such variables, consider a single, planar two-body system with Hamiltonian

$$\frac{|X|^2}{2\mu} - \frac{\mu M}{|x|}, \quad X \in \mathbb{R}^2, \quad 0 \neq x \in \mathbb{R}^2$$
$$\text{(with respect to } dX \wedge dx) \qquad [16]$$

and introduce – as done before formula [14] for $\mathcal{H}^{(0)}_{\rm rcp}$ – planar Delaunay variables $((L,G),(\ell,g))$ (here, $g = \hat{g} =$ argument of the perihelion). To remove the singularity of the Delaunay variables near zero eccentricities, Poincaré introduced variables $((\Lambda,\eta),(\lambda,\xi))$ defined by the following formulas:

$$\begin{aligned} &\Lambda = L, \quad H = L - G \\ &\lambda = \ell + g, \quad h = -g \\ &\sqrt{2H}\cos h = \eta \\ &\sqrt{2H}\sin h = \xi \end{aligned} \qquad [17]$$

As Poincaré showed, such variables are symplectic and analytic in a neighborhood of $(0,\infty) \times \mathbb{T} \times \{0,0\}$; notice that the symplectic map $((\Lambda,\eta),(\lambda,\xi)) \to (X,x)$ depends on the parameters $\mu, M$, and $\varepsilon$. In Poincaré variables, the two-body Hamiltonian in eqn [16]

becomes $-\kappa/(2\Lambda^2)$, with $\kappa := (\mu/m_0)^3/M$. Now, re-insert the index $i$, let $\phi_i : ((\Lambda_i, \eta_i), (\lambda_i, \xi_i)) \to (X^{(i)}, x^{(i)})$ and $\phi(\Lambda, \eta, \lambda, \xi) = (\phi_1(\Lambda_1, \eta_1, \lambda_1, \xi_1), \ldots, \phi_n(\Lambda_n, \eta_n, \lambda_n, \xi_n))$. Then, the Hamiltonian for the planar $(1+n)$-body problem takes the form

$$\begin{aligned} \mathcal{H}_{\text{pln}} \circ \phi &= \mathcal{H}_0(\Lambda) + \varepsilon \mathcal{H}_1(\Lambda, \lambda, \eta, \xi) \\ \mathcal{H}_0 &:= -\frac{1}{2}\sum_{i=1}^{n} \frac{\kappa_i}{\Lambda_i^2}, \quad \kappa_i := \left(\frac{\mu_i}{m_0}\right)^3 \frac{1}{M_i} \\ \mathcal{H}_1 &:= \mathcal{H}_1^{\text{compl}} + \mathcal{H}_1^{\text{princ}} \end{aligned} \qquad [18]$$

where the so-called "complementary part" $\mathcal{H}_1^{\text{compl}}$ and the "principal part" $\mathcal{H}_1^{\text{princ}}$ of the perturbation are, respectively, the functions

$$\sum_{1 \le i < j \le n} X^{(i)} \cdot X^{(j)} \quad \text{and} \quad \sum_{1 \le i < j \le n} \frac{\mu_i \mu_j}{m_0^2} \frac{1}{|x^{(i)} - x^{(j)}|} \qquad [19]$$

expressed in Poincaré variables.

The scheme of proof of Arnol'd's theorem in the planar, three-body case (one star, $n = 2$ planets) is as follows. The Hamiltonian is given by eqn [13] with $n = 2$; the phase space is eight dimensional (four degrees of freedom). This system, as mentioned several times, is properly degenerate and Kolmogorov's theorem cannot be applied directly; furthermore, a full (four-dimensional) set of action variables needs to be identified.

A first observation is that, in the planetary model, there are "fast variables" (the $\lambda_i$'s describing the revolutions of the planets) and "secular variables" (the $\eta_i$'s and $\xi_i$'s describing the variations of position and shape of the instantaneous Keplerian ellipses). By averaging theory (see, e.g., Arnol'd (1998)), one can "neglect," in nonresonant regions, the fast-angle dependence up to high order in $\varepsilon$ obtaining an effective Hamiltonian, which, up to $O(\varepsilon^2)$, is given by the "secular" Hamiltonian

$$\begin{aligned} \mathcal{H}_{\text{sec}} &:= \mathcal{H}_0(\Lambda) + \varepsilon \bar{\mathcal{H}}_1(\Lambda, \eta, \xi) \\ \bar{\mathcal{H}}_1(\Lambda, \eta, \xi) &:= \int \mathcal{H}_1 \frac{d\lambda}{(2\pi)^2} \end{aligned} \qquad [20]$$

"Nonresonant region" means, here, an open $\Lambda$-set where $\partial_\Lambda \mathcal{H}_0 \cdot k \neq 0$ for $k \in \mathbb{Z}^2, |k_1| + |k_2| \le K$ and for a suitable $K \ge 1$.

In order to analyze the secular Hamiltonian, we shall beriefly consider $\bar{\mathcal{H}}_1$ as a function of the symplectic variables $\eta$ and $\xi$, regarding the "slow actions" $\Lambda_i$ as parameters.

For symmetry reasons, $\bar{\mathcal{H}}_1$ is even in $(\eta, \xi)$ and the point $(\eta, \xi) = (0, 0)$ is an elliptic equilibrium for $\bar{\mathcal{H}}_1$: the eigenvalues of the matrix $S\partial^2_{(\eta,\xi)}\bar{\mathcal{H}}_1(\Lambda, 0, 0)$, $S$ being the standard symplectic matrix, are purely imaginary numbers $\{\pm i\Omega_1, \pm i\Omega_2\}$. The real numbers $\{\Omega_i\}$ are symplectic invariants of the secular Hamiltonian and are usually called first (or linear) Birkhoff invariants. In a neighborhood of an elliptic equilibrium, one can use Birkhoff's normal form theory (see, e.g., Siegel (1971)): if the linear invariants $(\Omega_1, \Omega_2)$ are nonresonant up to order $r$ (i.e., if $\Omega \cdot k := \Omega_1 k_1 + \Omega_2 k_2 \neq 0$ for any $k \in \mathbb{Z}^2$ such that $|k_1| + |k_2| \le r$), then one can find a symplectic transformation $\phi_{\text{Bir}}$ so that

$$\bar{\mathcal{H}}_1 \circ \phi_{\text{Bir}} = F(J_1, J_2; \Lambda) + o_r, \quad J_j = \frac{\eta_j^2 + \xi_j^2}{2} \qquad [21]$$

where $F$ is a polynomial of degree $[r/2]$ of the form $\Omega_1 J_1 + \Omega_2 J_2 + (1/2)\mathcal{M}J \cdot J + \cdots$, $\mathcal{M} = \mathcal{M}(\Lambda)$ being a $(2 \times 2)$ matrix (and $o_r/|J|^{r/2} \to 0$ as $|J| \to 0$). Arnol'd, using computations performed by Le Verrier, checked the nonresonance condition up to order $r = 6$ in the asymptotic regime $a_1/a_2 \to 0$ (where $a_i$ denote the semimajor axes of approximate Keplerian ellipses of the two planets); these computations represent one of the most delicate parts of the paper.

Thus, combining averaging theory and Birkhoff normal form theory, one can construct a symplectic change of variables defined on an open subset of the phase space (avoiding some linear resonances) $(\Lambda, \lambda, \eta, \xi) \to (\Lambda', \lambda', J, \varphi)$, where $\eta_j + i\xi_j = \sqrt{2J_j}\exp(i\varphi_j)$, casting the three-body Hamiltonian into the form

$$\begin{aligned} &\mathcal{H}_0(\Lambda') + \varepsilon\big(\Omega(\Lambda') \cdot J + \tfrac{1}{2}\mathcal{M}(\Lambda')J \cdot J\big) \\ &\quad + \varepsilon^2 \mathcal{F}_1(\Lambda', J) + \varepsilon^p \mathcal{F}_2(\Lambda', \lambda', J, \varphi) \\ &\quad := \widetilde{\mathcal{H}}_0(\Lambda', J; \varepsilon) + \varepsilon^p \mathcal{F}_2(\Lambda', \lambda', J, \varphi) \end{aligned} \qquad [22]$$

for a suitable prefixed order $p \ge 3$; notice that the nonresonance condition needed to apply averaging theory is not particularly hard to check since it involves the unperturbed and completely explicit Kepler Hamiltonian $\mathcal{H}_0$. The idea is now to consider $\varepsilon^p \mathcal{F}_2$ as a perturbation of the completely integrable Hamiltonian $\widetilde{\mathcal{H}}_0$ and to apply Kolmogorov's theorem. Finally, one can check the Kolmogorov's nondegeneancy condition, which since

$$\det \partial^2_{(\Lambda', J)} \widetilde{\mathcal{H}}_0(\Lambda', J'; \varepsilon) = \varepsilon^2\big((\det \mathcal{H}_0'') \det \mathcal{M} + O(\varepsilon)\big)$$

amounts to check the invertibility of the matrix $\mathcal{M}$. Such a condition is also checked in Arnol'd (1963b) with the aid of Le Verrier's tables and in the asymptotic regime $a_1/a_2 \to 0$.

**The spatial three-body problem** In order to extend the previous argument to the spatial case, Arnol'd suggested connecting the planar and spatial case through a limiting procedure. Such strategy presents

analytical problems (the symplectic variables for the spatial case become singular in the planar limit), which have not been overcome. However, the particular structure of the three-body case allows one to derive a four-degree-of-freedom Hamiltonian, to which the proof of the planar case can be easily adapted. The procedure described below is based on the classical Jacobi's reduction of the nodes.

First, we inroduce a convenient set of symplectic variables. Let, for $i=1,2,((L_i,G_i,\Theta_i),(\ell_i,g_i,\theta_i))$ denote the Delaunay variables introduced in items (5) and (6) above: these are the Delaunay variables associated to the two-body system, Sun–*i*th planet. Then, as Poincaré showed, the variables $((\Lambda_i^*,\lambda_i^*),(\eta_i^*,\xi_i^*),(\Theta_i,\theta_i))$, where

$$\begin{aligned}\Lambda_i^* &= L_i\\ \lambda_i^* &= \ell_i + g_i\\ \eta_i^* &= \sqrt{2(L_i-G_i)}\cos g_i\\ \xi_i^* &= -\sqrt{2(L_i-G_i)}\sin g_i\end{aligned} \qquad [23]$$

are symplectic and analytic near circular, non-coplanar motions; for a detailed discussion of these and other sets of interesting classical variables, see, for example, Biasco *et al.* (2003) and references therein; the asterisk is introduced to avoid confusion with a closely related but different set of Poincaré variables (see below). Let us denote by

$$\mathcal{H}_{3\mathrm{bp}} := \mathcal{H}^{(0)}(\Lambda^*) + \varepsilon\mathcal{H}^{(1)}(\Lambda^*,\lambda^*,\eta^*,\xi^*,\Theta,\theta)$$

the Hamiltonian equation [8] (with $n=2$) expressed in terms of the symplectic variables $((\Lambda^*,\lambda^*),(\eta^*,\xi^*),(\Theta,\theta)),\Lambda^*=(\Lambda_1^*,\Lambda_2^*)$, etc. Recalling the physical meaning of the Delaunay variables, one realizes that $\Theta_1+\Theta_2$ is the vertical component, $C_3=C\cdot k_3$, of the total argument $C=C^{(1)}+C^{(2)}$, where $C^{(i)}$ denotes the angular momentum of the *i*th planet with respect to the origin of an inertial heliocentric frame $\{k_1,k_2,k_3\}$. This suggests that the symplectic variables can be introduced:

$$(\Lambda^*,\lambda^*,\eta^*,\xi^*,\Psi,\psi) = \phi(\Lambda^*,\lambda^*,\eta^*,\xi^*,\Theta,\theta)$$

with $(\Psi_1,\Psi_2,\psi_1,\psi_2):=(\Theta_1,\Theta_1+\Theta_2,\theta_1-\theta_2,\theta_2)$.

Let

$$\mathcal{H}_{3\mathrm{bp}}^* := \mathcal{H}_{3\mathrm{bp}}\circ\phi^{-1}$$

denote the Hamiltonian of the spatial three-body problem in these symplectic variables. Since the Poisson bracket of $\Psi_2=\Theta_1+\Theta_2$ and $\mathcal{H}_{3\mathrm{bp}}^*$ vanishes ($C_3$ being an integral for the $\mathcal{H}_{3\mathrm{bp}}$-flow), the conjugate angle $\psi_2$ is cyclic for $\mathcal{H}_{3\mathrm{bp}}^*$, that is,

$$\mathcal{H}_{3\mathrm{bp}}^* = \mathcal{H}_{3\mathrm{bp}}^*(\Lambda^*,\lambda^*,\eta^*,\xi^*,\Psi_1,\Psi_2,\psi_1)$$

Now (because the total angular momentum $C$ is preserved), one may restrict attention to the ten-dimensional invariant (and symplectic) submanifold $\mathcal{M}_{\mathrm{ver}}$ defined by fixing the total angular momentum to be vertical. Such submanifold is easily described in terms of Delaunay variables; in fact, $C\cdot k_1=0=C\cdot k_2$ is equivalent to

$$\theta_1-\theta_2=\pi \quad\text{and}\quad G_1^2-\Theta_1^2=G_2^2-\Theta_2^2 \qquad [24]$$

Thus, $\mathcal{M}_{\mathrm{ver}}^* := \phi(\mathcal{M}_{\mathrm{ver}})$ is given by

$$\mathcal{M}_{\mathrm{ver}}^* = \left\{\psi_1=\pi,\ \Psi_1=\widehat{\Psi}_1(\Lambda^*,\eta^*,\xi^*;\Psi_2)\right\}$$

with

$$\widehat{\Psi}_1 := \frac{\Psi_2}{2} + \frac{(\Lambda_1^*-H_1^*)^2-(\Lambda_2^*-H_2^*)^2}{2\Psi_2}$$
$$H_i^* := \frac{\eta_i^{*2}+\xi_i^{*2}}{2}$$

Since $\mathcal{M}_{\mathrm{ver}}^*$ is invariant for the flow $\phi_*^t$ of $\mathcal{H}_{3\mathrm{bp}}^*,\psi_1(t)\equiv\pi$ and $\dot\psi_1\equiv 0$ for motions starting on $\mathcal{M}_{\mathrm{ver}}^*$, which implies that $(\partial_{\Psi_1}\mathcal{H}_{3\mathrm{bp}}^*)|_{\mathcal{M}_{\mathrm{ver}}^*}=0$. This fact allows one to introduce, for fixed values of the vertical angular momentum $\Psi_2=c\neq 0$, the following reduced Hamiltonian

$$\begin{aligned}&\mathcal{H}_{\mathrm{red}}^c(\Lambda^*,\lambda^*,\eta^*,\xi^*)\\ &\quad := \mathcal{H}_{3\mathrm{bp}}^*(\Lambda^*,\lambda^*,\eta^*,\xi^*,\ \widehat{\Psi}_1(\Lambda^*,\eta^*,\xi^*;c),c,\pi)\end{aligned}$$

on the eight-dimensional phase space $\mathcal{M}_{\mathrm{red}} := \{\Lambda_i^*>0,\ \lambda\in\mathbb{T}^2,(\eta^*,\xi^*)\in B^4\}$ endowed with the standard symplectic form $d\Lambda^*\wedge d\lambda^*+d\eta^*\wedge d\xi^*$ ($B^4$ being a ball around the origin in $\mathbb{R}^4$). In fact, the (standard) Hamilton's equations for $\mathcal{H}_{\mathrm{red}}^c$ are immediately recognized to be a subsystem of the full (standard) Hamilton's equations for $\mathcal{H}_{3\mathrm{bp}}$ when the initial data are restricted on $\mathcal{M}_{\mathrm{ver}}^*$ and the constant value of $\Psi_2$ is chosen to be $c$. More precisely, if the Hamiltonian flow of $\mathcal{H}_{\mathrm{red}}^c$ on $\mathcal{M}_{\mathrm{red}}$ is denoted by $\phi_c^t$, then

$$\begin{aligned}&\phi_*^t\left(z^*,\widehat{\Psi}_1(\Lambda^*,\eta^*,\xi^*;c),c,\pi,\psi_2\right)\\ &\quad = \left(\phi_c^t(z^*),\widehat{\Psi}_1(t),c,\pi,\psi_2(t)\right)\end{aligned} \qquad [25]$$

where we have used the shorthand notations: $z^*=(\Lambda^*,\lambda^*,\eta^*,\xi^*)\in\mathcal{M}_{\mathrm{red}}$; $\widehat{\Psi}_1(t)=\widehat{\Psi}_1\circ\phi_c^t(z^*)$; $\psi_2(t)=\psi_2+\int_0^t\partial_{\Psi_2}\mathcal{H}_{3\mathrm{bp}}^*(\phi_c^s(z^*),\widehat{\Psi}_1(s),c,\pi)ds$. At this point, the scheme used for the planar case may be easily adapted to the present situation. The nondegeneracy conditions have been checked in Robutel (1995) where indications, based on a computer program, have been given for the validity of the theorem in a wider set of initial data.

Notice that the dimension of the reduced phase space of the spatial case is 8, which is also the dimension of the phase space of the planar case.

Therefore, also the Lagrangian tori obtained with this procedure have the same dimension of the tori obtained in the planar case (i.e., four).

**The general case** Consider the general case following the strategy of M Herman as presented by Féjoz (2004), to which the reader is referred for complete proofs and further references.

The symplectic variables used in Féjoz (2004), to cope with the spatial planetary $(1+n)$-body problem (Sun and $n$ planets), are closely related to the variables defined in eqn [23]. For $1 \le i \le n$, let $((L_i, G_i, \Theta_i), (\ell_i, g_i, \theta_i))$ denote the Delaunay variables associated with the two-body system, Sun–$i$th planet. Then (as shown by Poincaré) the variables $((\Lambda_i, \lambda_i), (\eta_i, \xi_i), (p_i, q_i))$, where $\Lambda_i = L_i$, $\lambda_i = \ell_i + g_i + \theta_i$, and

$$\begin{aligned} \eta_i &= \sqrt{2(L_i - G_i)}\cos(g_i + \theta_i) \\ \xi_i &= -\sqrt{2(L_i - G_i)}\sin(g_i + \theta_i) \\ p_i &= \sqrt{2(G_i - \Theta_i)}\cos\theta_i \\ q_i &= -\sqrt{2(G_i - \Theta_i)}\sin\theta_i \end{aligned} \qquad [26]$$

are symplectic and analytic near circular, noncoplanar motions (see, e.g., Biasco *et al.* (2003)). Let

$$\mathcal{H}_{\text{nbp}} := \mathcal{H}^{(0)}(\Lambda) + \varepsilon \mathcal{H}^{(1)}(\Lambda, \lambda, \eta, \xi, p, q) \qquad [27]$$

denote the Hamiltonian (eqn [8]) expressed in terms of the Poincaré symplectic variables $((\Lambda, \lambda), (\eta, \xi), (p, q))$, $\Lambda = (\Lambda_1, \ldots, \Lambda_n)$, etc.

As the number of the planets increases, the degeneracies become stronger and stronger. Furthermore, a clean reduction, such as the reduction of the nodes, is no more available if $n > 2$. To overcome these problems Herman proposed a new approach, which is described below.

Instead of Kolmogorov's nondegeneracy assumption – which says that the frequency map [13] $I \to \omega(I)$ is a local diffeomorphism – one may consider weaker nondegeneracy conditions. In particular, in Féjoz (2004), one considers nonplanar frequency maps. A smooth curve $u \in A \to \omega(u) \in \mathbb{R}^d$, where $A$ is an open nonempty interval, is called "nonplanar" at $u_0 \in A$ if all the $u$-derivatives up to order $(d-1)$ at $u_0, \omega(u_0), \omega'(u_0), \ldots, \omega^{(d-1)}(u_0)$ are linearly independent in $\mathbb{R}^d$; a smooth map $u \in A \subset \mathbb{R}^p \to \omega(u) \in \mathbb{R}^d, p \le d$, is called nonplanar at $u_0 \in A$ if there exists a smooth curve $\varphi : \hat{A} \to A$ such that $\omega \circ \varphi$ is nonplanar at $t_0 \in \hat{A}$ with $\varphi(t_0) = u_0$. A S Pyartli has proved (see, e.g., Féjoz (2004)) that if the map $u \in A \subset \mathbb{R}^p \to \omega(u) \in \mathbb{R}^d$ is nonplanar at $u_0$, then there exists a neighborhood $B \subset A$ of $u_0$ and a subset $C \subset B$ of full Lebesgue measure (i.e., $\text{meas}(C) = \text{meas}(B)$) such that $\omega(u)$ is Diophantine for any $u \in C$. The nonplanarity condition is weaker than Kolmogorov's nondegeneracy conditions; for example, the map

$$\begin{aligned} \omega(I) &:= \partial_I \left( \frac{I_1^4}{4} + I_1^2 I_2 + I_1 I_3 + I_4 \right) \\ &= \left( I_1^3 + 2 I_1 I_2 + I_3, I_1^2, I_1, 1 \right) \end{aligned}$$

violates both Kolmogorov's nondegeneracy and the isoenergetic nondegeneracy conditions but is nonplanar at any point of the form $(I_1, 0, 0, 0)$, since $\omega(I_1, 0, 0, 0) = (I_1^3, I_1^2, I_1, 1)$ is a nonplanar curve (at any point).

As in the three-body case, the frequency map is that associated with the averaged secular Hamiltonian

$$\begin{aligned} \mathcal{H}_{\text{sec}} &:= \mathcal{H}^{(0)}(\Lambda) + \varepsilon \bar{\mathcal{H}}^{(1)} \\ \bar{\mathcal{H}}^{(1)}(\Lambda, \eta, \xi, p, q) &:= \int \mathcal{H}^{(1)} \frac{d\lambda}{(2\pi)^n} \end{aligned} \qquad [28]$$

which has an elliptic equilibrium at $\eta = \xi = p = q = 0$ (as above, $\Lambda$ is regarded as a parameter). It is a remarkably well-known fact that the quadratic part of $\bar{\mathcal{H}}^{(1)}$ does not contain "mixed terms," namely,

$$\begin{aligned} \bar{\mathcal{H}}^{(1)} = \bar{\mathcal{H}}_0^{(1)} + \varepsilon \big( & \mathcal{Q}_{\text{pln}}\, \eta \cdot \eta + \mathcal{Q}_{\text{pln}}\, \xi \cdot \xi + \mathcal{Q}_{\text{spt}}\, p \cdot p \\ & + \mathcal{Q}_{\text{spt}}\, q \cdot q + O_4 \big) \end{aligned} \qquad [29]$$

where the function $\bar{\mathcal{H}}_0^{(1)}$ and the symmetric matrices $\mathcal{Q}_{\text{pln}}$ and $\mathcal{Q}_{\text{spt}}$ depend upon $\Lambda$ while $O_4$ denotes terms of order 4 in $(\eta, \xi, p, q)$. The eigenvalues of the matrices $\mathcal{Q}_{\text{pln}}$ and $\mathcal{Q}_{\text{spt}}$ are the first Birkhoff invariants of $\bar{\mathcal{H}}^{(1)}$ (with respect to the symplectic variables $(\eta, \xi, p, q)$). Let $\sigma_1, \ldots, \sigma_n$ and $\varsigma_1, \ldots, \varsigma_n$ denote, respectively, the eigenvalues of $\mathcal{Q}_{\text{pln}}$ and $\mathcal{Q}_{\text{spt}}$; then the frequency map for the $(1+n)$-body problem will be defined as (recall eqn [18])

$$\Lambda \to (\hat{\omega}, \varepsilon \Omega) \qquad [30]$$

with

$$\begin{aligned} \hat{\omega} &:= \left( \frac{\kappa_1}{\Lambda_1^3}, \ldots, \frac{\kappa_n}{\Lambda_n^3} \right) \\ \Omega &:= (\sigma, \varsigma) := ((\sigma_1, \ldots, \sigma_n), (\varsigma_1, \ldots, \varsigma_n)) \end{aligned} \qquad [31]$$

Herman pointed out, however, that the frequencies $\sigma$ and $\varsigma$ satisfy two independent linear relations, namely (up to renumbering the indices),

$$\varsigma_n = 0, \qquad \sum_{i=1}^{n} (\sigma_i + \varsigma_i) = 0 \qquad [32]$$

which clearly prevents the frequency map to be nonplanar; the second relation in eqn [32] is usually

called "Herman resonance" (while the first relation is a well-known consequence of rotation invariance).

The degeneracy due to rotation invariance may be easily taken care of by considering (as in the three-body case) the $(6n-2)$-dimensional invariant symplectic manifold $\mathcal{M}_{\text{ver}}$, defined by taking the total angular momentum C to be vertical, that is, $C \cdot k_1 = 0 = C \cdot k_2$. But, when $n > 2$, Jacobi's reduction of the nodes is no more available and to get rid of the second degeneracy (Herman's resonance), the authors bring in a nice trick, originally due – once more! – to Poincaré. In place of considering $\mathcal{H}_{\text{nbp}}$ restricted on $\mathcal{M}_{\text{ver}}$, Féjoz considers the modified Hamiltonian

$$\mathcal{H}^{\delta}_{\text{nbp}} := \mathcal{H}_{\text{nbp}} + \delta C_3^2, \quad C_3 := C \cdot k_3 = |C| \qquad [33]$$

where $\delta \in \mathbb{R}$ is an extra artificial parameter. By an analyticity argument, it is then possible to prove that the (rescaled) frequency map

$$(\Lambda, \delta) \rightarrow (\hat{\omega}, \sigma_1, \ldots, \sigma_n, \varsigma_1, \ldots, \varsigma_{n-1}) \in \mathbb{R}^{3n-1}$$

is nonplanar on an open dense set of full measure and this is enough to find a positive measure set of Lagrangian maximal $(3n-1)$-dimensional invariant tori for $\mathcal{H}^{\delta}_{\text{nbp}}$; but, since $\mathcal{H}^{\delta}_{\text{nbp}}$ and $\mathcal{H}_{\text{nbp}}$ commute, a classical Lagrangian intersection argument allows one to conclude that such tori are invariant also for $\mathcal{H}_{\text{nbp}}$ yielding the complete proof of Arnol'd's theorem in the general case. Notice that this argument yields $(3n-1)$-dimensional tori, which in the three-body case means five dimensional. Instead, the tori found in the section "The spatial three-body problem" are four dimensional. The point is that in the reduced phase space, the motion of the nodeline – denoted as $\psi_2(t)$ in eqn [25] – does not appear.

We conclude this discussion by mentioning that the KAM theory used in Féjoz (2004) is a modern and elegant function-theoretic reformulation of the classical theory and is based on a $C^{\infty}$ local inversion theorem (F Sergeraert and R Hamilton) on "tame" Frechet spaces (which, in turn, is related to the Nash–Moser implicit function theorem; see Bost (1984–85)).

## Lower Dimensional Tori

The maximal tori for the many-body problems described above are found near the elliptic equilibria given by the decoupled Keplerian motions. It is natural to ask what happens of such elliptic equilibria when the interaction among planets is taken into account. Even though no complete answer has yet been given to such a question, it appears that, in general, the Keplerian elliptic equilibria "bifurcate" into elliptic $n$-dimensional tori. This section presents a short and nontechnical account of the existing results on the matter (the general theory of lower-dimensional tori is, mainly, due to J K Moser and S M Graff for the hyperbolic case and V K Melnikov, H Eliasson, and S B Kuksin for the technically more difficult elliptic case; for references, see, e.g., Chierchia *et al.* (2004)).

The normal form of a Hamiltonian admitting an $n$-dimensional elliptic invariant torus $\mathcal{T}$ of energy $E$, proper frequencies $\hat{\omega} \in \mathbb{R}^n$, and "normal frequencies" $\Omega \in \mathbb{R}^p$ in a $2d$-dimensional phase space with $d = n + p$ is given by

$$N := E + \hat{\omega} \cdot y + \sum_{j=1}^{p} \Omega_j \frac{\eta_j^2 + \xi_j^2}{2} \qquad [34]$$

Here the symplectic form is given by $dy \wedge dx + d\eta \wedge d\xi, y \in \mathbb{R}^n, x \in \mathbb{T}^n, (\eta, \xi) \in \mathbb{R}^{2p}; \mathcal{T}$ is then given by $\mathcal{T} := \{y = 0\} \times \{\eta = \xi = 0\}$. Under suitable assumptions, a set of such tori persists under the effect of a small enough perturbation $P(y, x, \eta, \xi)$. Clearly, the union of the persistent tori (if $n < d$) forms a set of zero measure in phase space; however, in general, $n$-parameter families persist.

In the many-body case considered in this article, the proper frequencies are the Keplerian frequencies given by the map $\Lambda \rightarrow \hat{\omega}(\Lambda)$ (eqn [31]), which is a local diffeomorphism of $\mathbb{R}^n$. The normal frequencies $\Omega$, instead, are proportional to $\varepsilon$ and are the first Birkhoff invariants around the elliptic equilibria as discussed above. Under these circumstances, the main nondegeneracy hypothesis needed to establish the persistence of the Keplerian $n$-dimensional elliptic tori boils down to the so-called Melnilkov condition:

$$\Omega_j \neq 0 \neq \Omega_i - \Omega_j, \quad \forall j \neq i \qquad [35]$$

Such condition has been checked for the planar three-body case in Féjoz (2002), for the spatial three-body case in Biasco *et al.* (2003) and for the planar $n$-body case in Biasco *et al.* (2004). The general spatial case is still open: in fact, while it is possible to establish lower-dimensional elliptic tori for the modified Hamiltonian $\mathcal{H}^{\delta}_{\text{nbp}}$ in [33], it is not clear how to conclude the existence of elliptic tori for the actual Hamiltonian $\mathcal{H}_{\text{nbp}}$ since the argument used above works only for Lagrangian (maximal) tori; on the other hand, the direct asymptotics techniques used in Biasco *et al.* (2003) do not extend easily to the general spatial case.

Clearly, the lower-dimensional tori described in this section are not the only ones that arise in $n$-body dynamics. For more lower-dimensional tori in the planar three-body case, see Féjoz (2002).

## Physical Applications

The above results show that, in principle, there may exist "stable planetary systems" exhibiting quasiperiodic motions around coplanar, circular Keplerian trajectories – in the Newtonian many-body approximation – provided the masses of the planets are much smaller than the mass of the central star.

A quite different question is: in the Newtonian many-body approximation, is the solar system or, more in generally, a solar subsystem stable?

Clearly, even a precise mathematical reformulation of such a question might be difficult. However, it might be desirable to develop a mathematical theory for important physical models, taking into account observed parameter values.

As a very preliminary step in this direction, consider one of the results of Celletti and Chierchia (see Celletti and Chierchia (2003), and references therein).

In Celletti and Chierchia (2003), the (isolated) subsystem formed by the Sun, Jupiter, and asteroid Victoria (one of the main objects in the Asteroidal belt) is considered. Such a system is modeled by an order-10 Fourier truncation of the RPC3BP, whose Hamiltonian has been described in the section "Kolmogorov's theorem and the RPC3BP (1954)." The Sun–Jupiter motion is therefore approximated by a circular one, the asteroid Victoria is considered massless, and the motions of the three bodies are assumed to be coplanar; the remaining orbital parameters (Jupiter/Sun mass ratio, which is approximately 1/1000; eccentricity and semimajor axis of the osculating Sun–Victoria ellipse; and "energy" of the system) are taken to be the actually observed values. For such a system, it is proved that there exists an invariant region, on the observed fixed energy level, bounded by two maximal two-dimensional Kolmogorov tori, trapping the observed orbital parameters of the osculating Sun–Victoria ellipse.

As mentioned above, the proof of this result is computer assisted: a long series of algebraic computations and estimates is performed on computers, keeping a rigorous track of the numerical errors introduced by the machines.

## Acknowledgments

The author is indebted to J Féjoz for explaining his work on Herman's proof of Arnol'd's theorem prior to publication. The author is also grateful for the collaborations with his colleagues and friends L Biasco, E Valdinoci, and especially A Celletti.

*See also:* Averaging Methods; Diagrammatic Techniques in Perturbation Theory; Gravitational *N*-Body Problem (Classical); Hamiltonian Systems: Stability and Instability Theory; Hamilton–Jacobi Equations and Dynamical Systems: Variational Aspects; Korteweg–de Vries Equation and Other Modulation Equations; Stability Problems in Celestial Mechanics; Stability Theory and KAM.

## Further Reading

Arnol'd VI (1963a) Proof of a Theorem by A. N. Kolmogorov on the invariance of quasi-periodic motions under small perturbations of the Hamiltonian. *Russian Mathematical Survey* 18: 13–40.

Arnol'd VI (1963b) Small denominators and problems of stability of motion in classical and celestial mechanics. *Uspehi MatematiĬceskih Nauk* 18(6(114)): 91–192.

Arnol'd VI (ed.) (1988) *Dynamical Systems III*, Encyclopedia of Mathematical Sciences, vol. 3. Berlin: Springer.

Biasco L, Chierchia L, and Valdinoci E (2003) Elliptic two-dimensional invariant tori for the planetary three-body problem. *Archives for Rational and Mechanical Analysis* 170: 91–135.

Biasco L, Chierchia L, and Valdinoci E (2004) N-dimensional invariant tori for the planar (N + 1)-body problem. *SIAM Journal on Mathematical Analysis,* to appear, pp. 27 (http://www.mat.uniroma3.it/users/chierchia/WWW/english_version.html).

Bost JB (1984/85) Tores invariants des systèmes dynamiques hamiltoniens. *Séminaire Bourbaki expose* 639: 113–157.

Celletti A and Chierchia L (2003) KAM stability and celestial mechanics. *Memoirs of the AMS,* to appear, pp. 116. (http://www.mat.uniroma3.it/users/chierchia/WWW/english_version.html).

Chierchia L and Qian D (2004) Moser's theorem for lower dimensional tori. *Journal of Differential Equations* 206: 55–93.

Féjoz J (2002) Quasiperiodic motions in the planar three-body problem. *Journal of Differential Equations* 183(2): 303–341.

Féjoz J (2004) Démonstration du "théorème d'Arnol'd" sur la stabilité du système planétaire (d'après Michael Herman). *Ergodic Theory & Dynamical Systems* 24: 1–62.

Herman MR (1998) Démonstration d'un théorème de V.I. Arnol'd. *Séminaire de Systèmes Dynamiques and manuscripts*.

Kolmogorov AN (1954) On the conservation of conditionally periodic motions under small perturbation of the Hamiltonian. *Doklady Akademii Nauk SSR* 98: 527–530.

Laskar J and Robutel P (1995) Stability of the planetary three-body problem. I: Expansion of the planetary Hamiltonian. *Celestial Mechanics & Dynamical Astronomy* 62(3): 193–217.

Robutel P (1995) Stability of the planetary three-body problem. II: KAM theory and existence of quasi-periodic motions. *Celestial Mechanics & Dynamical Astronomy* 62(3): 219–261.

Salamon D (2004) The Kolmogorov–Arnol'd–Moser theorem. *Mathematical Physics Electronic Journal* 10: 1–37.

Siegel CL and Moser JK (1971) *Lectures on Celestial Mechanics.* Berlin: Springer.

# Kinetic Equations

**C Bardos**, Université de Paris 7, Paris, France



## Introduction

In most physical cases, the evolution of a system of $N$ indistinguishable interacting particles $X_N = (x_1, x_2, \ldots, x_N)$ with velocities $V_N = (v_1, v_2, \ldots, v_N)$ is described by a Hamiltonian system

$$\begin{aligned} \frac{\mathrm{d}X_N}{\mathrm{d}t} &= \frac{\partial H(X_N, V_N)}{\partial V_N} \\ \frac{\mathrm{d}V_N}{\mathrm{d}t} &= -\frac{\partial H(X_N, V_N)}{\partial X_N} \end{aligned} \qquad [1]$$

in the phase space $\boldsymbol{R}_X^{dN} \times \boldsymbol{R}_V^{dN}$. When $N$ becomes large, it is natural to consider replacing the above discrete phase space by a continuous phase space of dimension $1 \le d \le 3$, $\boldsymbol{R}_x^d \times \boldsymbol{R}_v^d$ and to introduce a measure $f(x, v, t)$ that describes the density of particles which, at the point $x \in \boldsymbol{R}^d$ and at time $t$, have velocity $v$. This measure may also be interpreted as a generalization of the empirical measure

$$\mu_N(t) = \frac{1}{N} \sum_{1 \le i \le N} \delta_{x_i(t), v_i(t)}$$

defined in the phase space $\boldsymbol{R}_x^d \times \boldsymbol{R}_v^d$ by the above system of $N$ particles. In this way, one constructs a link between the microscopic and the macroscopic descriptions. The macroscopic physical quantities are, for instance, the first moments of this density:

$$\begin{aligned} \rho(x,t) &= \int_{\boldsymbol{R}_v^d} f(x,v,t)\,\mathrm{d}v \ \text{(density)} \\ \rho(x,t)u(x,t) &= \int_{\boldsymbol{R}_v^d} v f(x,v,t)\,\mathrm{d}v \ \text{(momentum)} \\ \rho(x,t)E(x,t) &= \int_{\boldsymbol{R}_v^d} \frac{|v|^2}{2} f(x,v,t)\,\mathrm{d}v \ \text{(energy)} \end{aligned}$$

Kinetic theory studies the intermediate stage shown in Figure 1.

Its first successes were related to classical thermodynamics and in particular to the molecular hypothesis. The contributions of Maxwell (1860, 1872) and of Boltzmann (1867) led to the "Boltzmann" equation, described in the companion article of Mario Pulvirenti (*see* Boltzmann Equation (Classical and Quantum)). In 1905, Lorentz used the same point of view to describe the motion of electrons in a metal. However, the different physical context leads to some basic differences between the Boltzmann equation and the Lorentz equation. The Boltzmann equation is derived under the assumption that the driving forces result from collisions between pairs of molecules. Therefore, the problem is nonlinear with a quadratic nonlinearity. In the Lorentz model the driving force is the interaction of the electrons with the atoms of the metal, which remain fixed. Collisions between electrons are ignored, so that the Lorentz equation is linear.

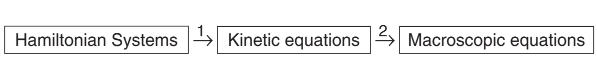



**Figure 1** Illustration of the role of kinetic equations in linking microscopic and macroscopic properties.

The most general form of a kinetic equation is as follows:

$$\begin{aligned} &\partial_t f(x,v,t) + \nabla_v H_f \cdot \nabla_x f(x,v,t) \\ &\quad - \nabla_x H_f \cdot \nabla_v f(x,v,t) = C(f) \end{aligned} \qquad [2]$$

The term $C(f)$ represents the effect of interactions either between particles or with the background. Without this term, the eqn [2] is reduced to the classical Liouville equation

$$\begin{aligned} &\partial_t f(x,v,t) + \nabla_v H_f \cdot \nabla_x f(x,v,t) \\ &\quad - \nabla_x H_f \cdot \nabla_v f(x,v,t) = 0 \end{aligned} \qquad [3]$$

which says that the function $f$ is transported by the flow of the Hamiltonian $H_f(x, v)$. This Hamiltonian depends on the model and may involve the unknown function $f$ itself. In the simplest case $H(x, v) = |v|^2/2$, eqn [3] and its solutions are given by

$$\begin{aligned} \partial_t f(x,v,t) + v \cdot \nabla_v f(x,v,t) &= 0 \\ f(x,v,t) &= f(x - vt, v, 0) \end{aligned} \qquad [4]$$

Nowadays kinetic equations appear in a variety of sciences and applications, such as astrophysics, aerospace engineering, nuclear engineering, particle–fluid interactions, semiconductor technology, social sciences, and biology, for example in chemotaxis and immunology.

They are used first to model phenomena and then to obtain a qualitative and quantitative description of situations involving sufficiently many particles so as to prohibit any computation at the level of particles, and yet the medium is still too rarefied to allow the use of macroscopic equations. As detailed in the next section, a macroscopic description requires that the function $f(x, v, t)$ be close to local thermodynamical equilibrium. For classical and quantum Boltzmann equations (*see* Boltzmann

Equation (Classical and Quantum)) these equilibria are either Maxwellian, Bose–Einstein, or Fermi–Dirac distributions.

Several effects, especially the influence of the boundary, may prevent the system from reaching local thermodynamical equilibrium and, therefore, even in macroscopic descriptions, kinetic equations may still be used to take into account the effect of the boundary. In this case, the term "Knudsen boundary layer" is currently used.

Finally, one should keep in mind that there exist some macroscopic phenomena which cannot be deduced from the corresponding microscopic physics by the mediation of a kinetic equation. Once again, returning to the companion article (*see* Boltzmann Equation (Classical and Quantum)) one observes that, since the only equilibria are Maxwellian, the macroscopic equations are those describing perfect gases. A real gas with a nontrivial van der Waals law is "too dense" to be explained by this theory. The alternative seems to go directly from the microscopic direction to the macroscopic description. This is a subject which is still under investigation and for which the reader may consult Olla *et al.* (1993).

## Kinetic Equations Entropy and Irreversibility

At the level of particles, the basic laws of physics are reversible. Yet these same laws are not reversible when seen at the level of a macroscopic description. This lack of reversibility is measured by the decay of entropy (mathematicians prefer convex functions; therefore, the mathematical entropy considered in this contribution is the negative of the physical entropy, and with irreversibility it decays). The kinetic equations lie in between, as shown in **Figure 1**; the decay of entropy should appear along one of the two arrows of this diagram.

Since the appearance of irreversibility is related to loss of information and averaging, it should be driven by a "mixing" process.

In general two mechanisms are responsible for such effects:

1. an ergodic or a relaxation mechanism by which a process averages itself; and
2. the introduction of some external random parameter. Observable quantities are then defined as averages over that parameter.

It seems important to compare these two "processes." This will be illustrated below with the most classical examples of the theory.

### The Diffusion Limit for the Neutron Transport Equation

Equations very similar to the one introduced by Lorentz are used to describe the interaction of neutrons with atoms in a nuclear reactor: this is the reason why these types of equations are often called neutron transport equations. An important issue is the derivation of a macroscopic diffusion equation. Assuming that neutrons are not subject to acceleration effects, considering the problem with constant modulus of velocity ($|v|=1$), introducing a "small" parameter $\epsilon$ which here corresponds to the absorption of the medium, one can study the following simplified model:

$$\epsilon\partial_t f_\epsilon + v\cdot\nabla_x f_\epsilon + \frac{\sigma(x)}{\epsilon}\left(f_\epsilon - \int_{|v'|=1} k(v,v')f_\epsilon(v')\mathrm{d}v'\right) = 0 \qquad [5]$$

In [5] one assumes, for the kernel $k(v,v')$, the following properties:

$$\forall v,\, v',\quad k(v,v') = k(v',v),\ 0 < k(v,v') \qquad \int_{|v'|=1} k(v,v')\mathrm{d}v = 1 \qquad [6]$$

and denotes by $K$ the operator

$$f \mapsto Kf = \int_{|v'|=1} k(v,v')f(v')\mathrm{d}v'$$

In the simplest case (say without boundary) eqn [5] is well-posed both for positive and negative time but hypothesis [6] has the following important consequences:

1. For positive time, it defines, for each $\epsilon > 0$, a contraction semigroup in any $L^p$ space and, therefore, the sequence of solutions or a subsequence thereof converges, say weakly, to a limit $f(x,v,t)$.
2. One also observes that $v \mapsto 1$ is (up to a multiplicative constant) the only solution of the equation

$$f - Kf = f(v) - \int_{|v'|=1} k(v,v')f(v')\mathrm{d}v = 0 \qquad [7]$$

   Therefore, the $\epsilon^{-1}$ in front of the collision term forces the limit $f(x,v,t)$ to be independent of $v$. In this simple problem, this is the thermodynamical equilibrium.

Dividing by $\epsilon$ and integrating over $|v|=1$ gives the relation

$$\partial_t \int_{|v|=1} f_\epsilon(x,v,t)\mathrm{d}v + \nabla_x \frac{1}{\epsilon}\int_{|v|=1} v f_\epsilon(x,v,t)\mathrm{d}v = 0 \qquad [8]$$

Now using the Fredholm alternative implies the existence and uniqueness of a function $v \mapsto \beta(v)$ such that

$$\beta(v) - \int_{|v'|=1} k(v,v')\beta(v')\mathrm{d}v' = v, \int_{|v'|=1} \beta(v')\mathrm{d}v' = 0 \qquad [9]$$

Multiply eqn [5] by $\beta(v)$ and integrate over $|v| = 1$ to obtain

$$\lim_{\epsilon\to 0} \frac{\sigma(x)}{\epsilon} \int_{|v|=1} ((I-K)\beta)(v) f_\epsilon(x,v,t)\mathrm{d}v = \lim_{\epsilon\to 0} \frac{\sigma(x)}{\epsilon} \int_{|v|=1} \beta(v)(I-K) f_\epsilon(x,v,t)\mathrm{d}v = -\lim_{\epsilon\to 0} \nabla_x \int_{|v|=1} \beta(v)\otimes v f_\epsilon(x,v,t)\mathrm{d}v \qquad [10]$$

Since the operator $(I-K)$ is self-adjoint non-negative, with 0 as the leading eigenvalue, the matrix

$$D = \int_{|v|=1} \beta(v)\otimes v\mathrm{d}v = \int_{|v|=1} \beta(v)\otimes (I-K)\beta(v)\mathrm{d}v$$

is positive definite, and one finally obtains the diffusion equation

$$\partial_t f - \nabla_x \left( \frac{D}{\sigma(x)} \nabla_x f \right) = 0 \qquad [11]$$

The above derivation is an example of what is called the "moments method." It is implicit even in the papers of Maxwell. It has been systematically used in several domains:

- To understand the relation between the Boltzmann equation and the Euler and Navier–Stokes equations (Golse 2005);
- To compute the critical size of a nuclear assembly. One shows that this size is well approximated by the size of the domain for which the Laplacian, with appropriate boundary conditions, has leading eigenvalue 0. It is for the spectral analysis of this problem that the averaging lemma (see the section "Some specific mathematical tools") was derived.
- To analyze the macroscopic limit for the solution of the radiative transfer equations, which describe the propagation of the intensity of photons in a large class of phenomena ranging from stellar atmospheres to the cooling of glass, including optical tomography in biomedical imaging. In a simplified form, the so-called "grey model," these equations can be reduced to

$$\epsilon\partial_t I_\epsilon(x,v,t) + v\cdot\nabla_x I_\epsilon(x,v,t) + \frac{1}{\epsilon}\sigma\Big(\frac{1}{4\pi}\int_{|v'|=1} I_\epsilon(x,v',t))\mathrm{d}v'\Big)\Big(I_\epsilon(x,v,t) - \frac{1}{4\pi}\int_{|v'|=1} I_\epsilon(x,v',t))\,\mathrm{d}v'\Big) = 0 \qquad [12]$$

In contrast to the previous example, the problem is, in many cases, nonlinear. The opacity $\sigma$ is a positive function that depends on the intensity $I_\epsilon$ through

$$\tilde{I}(x,t) = \frac{1}{4\pi}\int_{|v'|=1} I_\epsilon(x,v',t)\,\mathrm{d}v'$$

and which goes to $\infty$ with $\tilde{I}_\epsilon$ going to zero. The moments method can be applied with the averaging lemma, and one shows that the limit of $I_\epsilon$ is a function that is independent of $v$ and satisfies the following degenerate parabolic equation:

$$\partial_t I - \nabla_x \left( \frac{1}{3\sigma(I)} \nabla_x I \right) = 0 \qquad [13]$$

This equation is similar to the one obtained in the description of porous media and contains the following information: for initial data $I(x,0)$ with compact support, in contrast to the behavior of solutions of the standard diffusion equation, the solution $I(x,t)$ remains compactly supported in $x$. The boundary of this support is the thermal front and for a finite time, up to saturation (by water in porous media, by reacted deuterium in laser-confined fusion), this front remains fixed.

What made the analysis of the above macroscopic limit simple was the existence of an $\epsilon > 0$ dependent process which, for vanishing $\epsilon$, forces the solution to converge to a "thermodynamical" equilibrium. The irreversibility was already present in the first arrow of **Figure 1**. This is what made the analysis of the second arrow simple. The subtleties of the appearance of the irreversibility in the first arrow may be well explained by the next examples.

### The Linear Billiard Model

In the absence of an external electric field, the model proposed by Lorentz could be viewed as a limit of a system of particles evolving freely between spherical obstacles and reflecting on these obstacles according to the law of geometric optics. Along these lines, two types of results have been proved in two space variables.

In 1973, Gallavotti considered the case where the obstacles are randomly spaced under a Poisson configuration and proved the following theorem:

**Theorem 1** *Consider obstacles(balls) of radius $\epsilon$ and center $c_i$. Assume that the probability of finding exactly N such obstacles in a bounded measurable set $\Lambda \subset \boldsymbol{R}^2$ is given by the "Poisson law"*

$$P(\mathrm{d}\boldsymbol{c}_N) = \mathrm{e}^{-\mu_\epsilon|\Lambda|}\frac{\mu_\epsilon^N}{N!}\,\mathrm{d}c_1\,\mathrm{d}c_2\cdots\mathrm{d}c_N \qquad [14]$$

*with*

$$\boldsymbol{c}_N = c_1, c_2, \ldots, c_N \quad \text{and} \quad \mu_\epsilon = \frac{\mu}{\epsilon} \qquad [15]$$

*Denote by $\boldsymbol{E}^\epsilon$ the expectation with respect to the above Poisson distribution. For given $\epsilon$ and $\boldsymbol{c}_N$ introduce*

$$\mathrm{O}_{\boldsymbol{c}_N,\epsilon} = \boldsymbol{R}^2 \setminus \cup_{1\le i\le N}\{|x-c_i|\le\epsilon\} \qquad [16]$$

*and $f_{\boldsymbol{c}_N,\epsilon}$, the solution of the problem*

$$\partial_t f_{\boldsymbol{c}_N,\epsilon}(x,v,t) + v\cdot\nabla_x f_{\boldsymbol{c}_N,\epsilon}(x,v,t) = 0 \quad \text{in } \mathrm{O}_{\boldsymbol{c}_N,\epsilon}\times S^1 \qquad [17]$$

*with specular reflection on the boundary and v-independent initial data*:

$$f_{\boldsymbol{c}_N,\epsilon}(x,v,0) = \phi(x) \quad \text{in } \mathrm{O}_{\boldsymbol{c}_N,\epsilon}\times S^1 \qquad [18]$$

*Then*

$$h_\epsilon(x,t,\Omega) = \boldsymbol{E}^\epsilon[f_{\boldsymbol{c}_N,\epsilon}] \qquad [19]$$

*converges weakly for $t \ge 0$ to the solution of the transport equation*

$$\partial_t f(x,v,t) + v\cdot\nabla f(x,v,t) + \mu\left[2f(x,v) - \frac{1}{4}\int_S f(x,v')|v-v'|\mathrm{d}v'\right] = 0 \qquad [20]$$

$$f(x,v,0) = \phi(x) \quad \text{in } \boldsymbol{R}^2\times S^1 \qquad [21]$$

The situation is completely different when the obstacles are periodically spaced, a situation which seems closer to Lorentz's original idea. Golse (2003) (and previous contributions quoted in this article) obtained the following result:

**Theorem 2** *Assume that the obstacles are periodically spaced and conveniently scaled, defining the domain*

$$\mathrm{O}_\epsilon = \boldsymbol{R}^2 \setminus \underset{j\in\boldsymbol{Z}^2}{\cup}\{x, |x-\epsilon j|\le\epsilon^2\} \qquad [22]$$

*Then there exists a family of continuous uniformly bounded initial data such that no subsequence extracted from the family of solutions of*

$$\partial_t f_\epsilon + v\cdot\nabla_x f_\epsilon = 0 \quad \text{in } \mathrm{O}_\epsilon\times S^1 \qquad [23]$$

*with specular reflections on the boundary, converges to solutions of equations of the type* [20].

This pathology is related to the existence of particles that can travel freely for a very long time before meeting the obstacles, and the proof with some arithmetic (Diophantine approximations and continued fractions) relies on the analysis of such trajectories.

A comparison between the Theorems 1 and 2 shows that the ergodic property of the free flow on the periodic lattice is not strong enough to lead to a collisional kinetic equation unless some complementary randomness is introduced.

The examples of this section should be compared with the rigorous derivation of the Boltzmann equation by Lanford (*see* Boltzmann Equation (Classical and Quantum)). The reader should observe that this derivation corresponds to the same type of scaling (finite mean free path). However, no extra randomness is needed in this case. The proof uses the fact that configurations leading only to a finite number of binary collisions are of full measure. This corresponds to an ergodicity property which is enforced by the fact that the problem is genuinely nonlinear.

## Mean-Field Scaling and Vlasov Equations

The neutron transport equation is devoted to the interaction with obstacles and the Boltzmann equation to binary collisions. A simpler situation from the mathematical point of view corresponds to the case where each particle is under the action of the average of all other particles. Then the name "mean field limit" is used. The simplest example is the derivation of a Vlasov-type equation from a system of $N$ classical particles interacting with a $C^2$ potential $V(|x|)$. The following Hamiltonian is used:

$$H(x_1,\ldots,x_N,v_1,\ldots,v_N) = \sum_{1\le k\le N}\frac{|v_k|^2}{2} + \frac{1}{2N}\sum_{1\le l\ne k\le N}V(|x_k-x_l|) \qquad [24]$$

and the name mean-field scaling is related to the factor $N^{-1}$ before the potential. Assuming that the particles are undistinguishable, one introduces the joint probability density $F_N \equiv F_N(x_1,\ldots,x_N,$

$v_1, \ldots, v_N$) in the $N$-particle phase space, which satisfies the Liouville equation

$$\partial_t F_N + \{H_N, F_N\} := \partial_t F_N + \sum_{1\le k\le N} \Big( v_k \nabla_{x_k} F_N - \frac{1}{2N} \sum_{1\le l\ne k\le N} \nabla_{x_k}(V(|x_k - x_l|)) \times \nabla_{v_k} F_N \Big) = 0 \qquad [25]$$

From [25], with the notations

$$X_n = (x_1, \ldots, x_n), \quad V_n = (v_1, \ldots, v_n)$$
$$X_N^n = (x_{n+1}, \ldots, x_N), \quad V_N^n = (x_{n+1}, \ldots, x_N)$$

one deduces an infinite hierarchy of equations for the marginals

$$F_N^n(X_n, V_n, t) = \int f_N(X_N, V_N, t) \mathrm{d}X_N^n \, \mathrm{d}V_N^n$$
$$\text{for } 1 \le n \le N, \; F_N^n \equiv 0 \text{ for } N < n:$$

$$\partial_t F^n(X_n, V_n, t) + \sum_{1\le i\le n} v_n \nabla_{x_i} F_N^n(X_n, V_n, t) - \frac{1}{N} \sum_{1\le i<j\le n} \nabla_{v_i}(\nabla_{x_i} V(|x_i - x_j|) F_N^n(X_n, V_n, t)) - \frac{N-n}{N} \Big( \sum_{1\le i\le n} \nabla_{v_i} \iint \nabla_{x_i} V(|x_i - x^*|) \times F_N^{n+1}(X_n, V_n, x^*, v^*, t) \mathrm{d}x^* \mathrm{d}v^* \Big) = 0 \qquad [26]$$

Letting $N$ go to infinity, one obtains "formally," for the distribution functions,

$$F^n = \lim_{N\to\infty} F_N^n$$

the Vlasov hierarchy:

$$\partial_t F^n(X_n, V_n, t) + V_n \cdot \nabla_{X_n} F^n(X_n, V_n, t) - \sum_{1\le i\le n} \nabla_{v_i} \Big( \iint \nabla_{x_i} V(|x_i - x^*|) \times F_N^{n+1}(X_n, V_n, x^*, v^*, t) \mathrm{d}x^* \mathrm{d}v^* \Big) = 0 \qquad [27]$$

Observe that for any density $F(x, v, t)$ that satisfies

$$\iint F(x, v, t) \mathrm{d}x \, \mathrm{d}v = 1, \quad F(x, v, t) \ge 0 \qquad [28]$$

and is a solution of the $V$ potential Vlasov equation:

$$\partial_t F(x, v, t) + v \cdot \nabla_x F(x, v, t) - \Big( \iint \nabla_x V(|x - x^*|) F(x^*, v^*) \mathrm{d}x^* \mathrm{d}v^* \Big) \times \nabla_v F(x, v, t) = 0 \qquad [29]$$

the factorization formula

$$F^n(X_n, V_n, t) = \prod_{1\le i\le n} F(x_i, v_i, t) \qquad [30]$$

defines a solution of the above Vlasov hierarchy.

A uniqueness argument implies that any solution of the Vlasov hierarchy which is factorized at time zero will remain factorized at any subsequent time. Such a property, also observed for the hierarchy leading to the Boltzmann equation, is called the propagation of chaos. To make the proof rigorous, one has to analyze the limiting process in the hierarchy and prove the uniqueness of the solution of the infinite hierarchy. For a smooth potential, this has been done by Braun and Hepp in 1977 and by Spohn in 1981. An interesting approach consists, following Dobrushin, in introducing the Wasserstein distance; see Golse (2003) for a detailed exposition.

In the case of the Vlasov–Poisson equation [29] with $V(|x|) = 1/4\pi|x|$ the potential turns out to be too singular for the above derivation. In particular, the corresponding solution of the $N$-particle problem is not uniformly defined. However, for the corresponding equation (and for variants thereof, including the effect of the magnetic field, the Vlasov–Maxwell system) a series of mathematical results concerning existence and stability of solutions have been obtained. An excellent recent exposition of these results can be found in the book of Glassey (1996).

Equation [29] as well as the original system turns out to be fully reversible. Neither irreversibility nor averaging has appeared in the limit process which corresponds to the first arrow of **Figure 1**; this is due to the "weak coupling." Therefore, irreversibility should now appear on the second arrow. Integrating eqn [29] with respect to $v$ gives the relation (often called Fick's law):

$$\partial_t \rho(x, t) + \nabla_x \int v F(x, v, t) \mathrm{d}v = 0 \qquad [31]$$

But now expressing the current $j = \int v F(x, v, t) \mathrm{d}v$ in terms of macroscopic variables turns out to be a difficult issue in the absence of a "relaxation" effect. Up to now there has been no derivation of such macroscopic equations from first principles.

The same type of problems exist for the two-dimensional Euler equation, which is in some sense very similar to the Vlasov equation. It has been observed that these equations develop for "turbulent initial data" a kind of "mixing process" leading to coherent structures that would play the role of thermodynamical equilibrium (in the absence of relaxation). The Jupiter red spot is the most

well-known example of such a structure. These coherent structures are obtained by maximizing an entropy which does not come directly from the dynamics but which is inspired by similar problems in statistical mechanics. Finally, one has to take into account in this construction the existence of an infinite set of conserved quantities: for each regular function $G$, vanishing at infinity, one has

$$\frac{d}{dt}\iint G(F(x,v,t))\mathrm{d}x\,\mathrm{d}v = 0$$

This approach was already started by Onsager in 1945 and pursued by many scientists. A recent reference is the article of Chavanis and Sommeria (1998).

## Derivation of Kinetic Equations from the Schrödinger Equation

Oscillatory solutions of the Schrödinger equation, with wavelength of the order of the Planck constant, tend to behave like particles. This is described in detail by different tools of high-frequency approximation. In particular, the limit of the Wigner transform of the density $\psi(x,t)\otimes\overline{\psi}(y,t)$:

$$W(x,\xi,t) = \frac{1}{(2\pi)^{3d}}\int_{\boldsymbol{R}^{3d}} \mathrm{e}^{-\mathrm{i}\xi y}\psi\left(x+\frac{\hbar y}{2},t\right) \otimes\overline{\psi}\left(x-\frac{\hbar y}{2},t\right)\mathrm{d}y \qquad [32]$$

is a solution of a Liouville equation. Therefore, one should expect that in the presence of "many" obstacles ("many potentials") the limit should be given by a kinetic equation. As shown by the previous section the introduction of randomness seems compulsory in reaching this goal.

Consider a big cube $\Lambda=\Lambda_L$ of size $L$ in $\boldsymbol{R}^3$. Let $\omega=(x_\alpha), \alpha=1,2,\ldots,N$ denote the configuration of random obstacles distributed uniformly in $\Lambda$. The density of obstacles is $\rho=N/L^3$ and the expectation with respect to this uniform measure is denoted by

$$E := \prod_{1\le\alpha\le N}\left(L^{-d}\int \mathrm{d}x_\alpha\right)$$

With $V(|x|)$ a smooth, short-range potential, the random potential created by the obstacles is

$$V_\omega(x) = \sum_{1\le\alpha\le N} V(|x-x_\alpha|)$$

then one of the typical results (low-density limit, which corresponds to the quantum version of Gallavotti classical result) obtained, reads as follows:

**Theorem 3** (Erdös and Yau 1988) *Assume that the density of obstacles is $\rho=\rho_0\epsilon$ with a fixed $\rho_0$. Denote by $\psi_\omega^\epsilon(t)$ the solution of the Schrödinger equation*

$$\mathrm{i}\partial_t\psi_\omega^\epsilon = -\tfrac{1}{2}\Delta_x\psi_\omega^\epsilon + V_\omega\psi_\omega^\epsilon \qquad [33]$$

*with initial condition localized and oscillating at the scale $\epsilon$, that is, with $h$ and $S$ smooth*

$$\psi_\omega^\epsilon(0) = \epsilon^{3/2}h(\epsilon x)\exp\left(\mathrm{i}\frac{S(x)}{\epsilon}\right) \qquad [34]$$

*Consider the density matrix $\rho_\omega^\epsilon(t,x,y)=\psi_\omega^\epsilon(t,x)\otimes\overline{\psi}_\omega^\epsilon(t,y)$ and its Wigner transform*

$$W_\omega^\epsilon(x,\xi,t) = \frac{1}{(2\pi)^3}\int_{\boldsymbol{R}^3}\mathrm{e}^{-\mathrm{i}\xi y}\rho_\omega^\epsilon\left(t,x+\frac{\epsilon y}{2},x-\frac{\epsilon y}{2}\right)\mathrm{d}y \qquad [35]$$

*Then for any $t>0, EW_\omega^\epsilon(t)$ converges weakly with $\epsilon$ going to zero to a solution $F(t)$ of the kinetic equation*

$$\partial_t F(t,x,\xi)+\xi\cdot\nabla_x F(t,x,\xi) = \int |T(\xi,\xi')|^2\delta(|\xi|^2-|\xi'|^2)(F(t,x,\xi') - F(t,x,\xi))\mathrm{d}\xi' \qquad [36]$$

*where $T$ is the amplitude of the scattering operator associated to the Schrödinger equation with the short range potential $V$.*

The proof uses several ingredients including scattering theory with expansion in term of Dyson series; see Erdös and Yau (1998).

## Semiconductor Modeling

In modern computers, the electronic devices are so small that the electric current may have no space/time to reach a thermodynamical equilibrium. Therefore, this turns out to be a field where the kinetic equations are the most naturally used. Details of what can be deduced from a mathematical analysis can be found in Poupaud (1994). The equations involve the distribution of electrons $f_\mathrm{e}(x,k,t)$ and holes $f_\mathrm{h}(x,k,t)$ and have the following form:

$$\epsilon\partial_t f_\mathrm{e}(t,x,k)+v_\mathrm{e}(k)\nabla_x f_\mathrm{e}(t,x,k) + \frac{q}{\hbar}\nabla_x U(t,x)\cdot\nabla_k f_\mathrm{e}(t,x,k) = \frac{1}{\epsilon}(Q_\mathrm{e}(f_\mathrm{e})(t,x,k)+R_e(f_\mathrm{e},f_\mathrm{h})(t,x,k)) \qquad [37]$$

$$\epsilon\partial_t f_\mathrm{h}(t,x,k)+v_\mathrm{h}(k)\nabla_x f_\mathrm{h}(t,x,k) - \frac{q}{\hbar}\nabla_x U(t,x)\cdot\nabla_k f_\mathrm{h}(t,x,k) = \frac{1}{\epsilon}Q_\mathrm{h}(f_\mathrm{h})(t,x,k)+R_\mathrm{h}(f_\mathrm{h},f_\mathrm{e})(t,x,k) \qquad [38]$$

The variable $k$ ranges over a torus $B$ of $\boldsymbol{R}^3$ which, in physics books, carries the name of Brillouin zone. The velocities of propagation of electrons and holes are determined in terms of the energy band by the formula

$$v_{\text{e,h}} = \frac{1}{\hbar}\nabla_k E_{\text{e,h}}(k) \quad [39]$$

The potential $U$ is determined in terms of the doping profile $C(x)$, the conductivity $\epsilon_r$, and the density of electrons and holes according to the formula

$$-\Delta_x U(t,x) = \frac{q}{\epsilon_r}\left(C(x) - \frac{1}{|B|}\int_B f_{\text{e}}(t,x,k)\mathrm{d}k + \frac{1}{|B|}\int_B f_{\text{h}}(t,x,k)\mathrm{d}k\right) \quad [40]$$

Finally $Q_{\text{e,h}}$ and $R_{\text{e,h}}$ are binary integral operators in the variable $k \in B$ which model collisions and generation–recombination processes. Concerning the "mathematical approach" the situation is as follows.

The relations [39] can be deduced from the high-frequency analysis of the solution of the Schrödinger equation

$$\mathrm{i}\hbar\partial_t\psi = -\frac{\hbar^2}{2}\Delta\psi + V\left(\frac{x}{\hbar}\right)\psi \quad [41]$$

with $V$ a periodic potential constructed on the dual lattice of $B$. The method uses the Bloch decomposition of the solution and the Wigner series (Poupaud 1994). No mathematical derivation of the collisions operator is currently available. The situation should be compared to what is said in the section "Derivation of kinetic equations from the Schrödinger equation," but in a much more complicated setting.

On the other hand, the collision operators $Q_{\text{e,h}}$ and $R_{\text{e,h}}$, as given by phenomenological arguments, have enough good relaxation properties to allow a rigorous limit of the system [37]–[38] for $\epsilon$ going to zero (Poupaud 1994). This leads to the justification of the so-called drift–diffusion models and to the possibility of constructing correctors (with respect to $\epsilon$) and to treating the effect of heterojunctions by boundary layer analysis.

## Some Specific Mathematical Tools

Few proofs were given in the above exposition and details would not be suitable for a review article. However, the mathematical approach to kinetic equations has generated some new tools, and it may be useful to give the most prominent ones.

### The Averaging Lemma

Compactness results appear in spectral theory and in the construction of solutions of nonlinear equations (whenever strong convergence is needed for the limit). Being hyperbolic, the transport operator $v\cdot\nabla_x$ propagates singularities along characteristics. Therefore, at first sight it seems hopeless that one might obtain any regularizing effect from the free streaming part of a kinetic model. The key to obtaining regularizing effects from the transport operator $v\cdot\nabla_x$ is to seek those effects not on the number density itself, but on velocity averages thereof; in other words, on the macroscopic densities.

Here is the prototype of all velocity averaging results.

**Theorem 4** *Let $F_\epsilon$ be a bounded family in $L^2(\boldsymbol{R}^d\times\boldsymbol{R}^d)$. Assume that the family $v\cdot\nabla_x F_\epsilon$ is also bounded in $L^2(\boldsymbol{R}^d\times\boldsymbol{R}^d)$. Then, for each $\phi\in L^2(\boldsymbol{R}^d)$, the family of moments $\rho_\epsilon(x)$ defined by*

$$\rho_\epsilon(x) = \int_{\boldsymbol{R}^d} F_\epsilon(x,v)\phi(v)\mathrm{d}v$$

*is relatively compact in $L^2(\boldsymbol{R}^d)$.*

For the proof one starts with the expression $G_\epsilon = F_\epsilon + v\cdot\nabla_x F_\epsilon$ takes the Fourier transform with respect to $x$ of this relation and writes for $\hat{\rho}_\epsilon(\xi)$ the expression

$$\hat{\rho}_\epsilon = \int_{\boldsymbol{R}^d}\frac{\hat{G}_\epsilon(\xi,v)\phi(v)\mathrm{d}v}{1+\mathrm{i}v.\xi} \quad [42]$$

Then use the Cauchy–Schwarz inequality to obtain

$$|\hat{\rho}_\epsilon|^2 \le \left(\int_{\boldsymbol{R}^d}\frac{|\phi(v)|^2\mathrm{d}v}{1+|v\cdot\xi|^2}\right)\int_{\boldsymbol{R}^d}|G_\epsilon(\xi,v)|^2\mathrm{d}v \quad [43]$$

and complete the proof by standard arguments.

The averaging lemma was first observed by Agoshkov (1984) for abstract results concerning the regularity of solutions of kinetic equations in domains with boundary. Independently, it was rediscovered in the improved form given above by Golse, Perthame, and Sentis (1985) and used for the spectral theory in the diffusion approximation. The extension to $L^p, p>1$, spaces and to $L^1$ (with use of entropy estimate) were instrumental in proving the validity of the Rosseland approximation for the radiative transfer equations and for the proof of existence by Lions and Di Perna of renormalized solutions of the Boltzmann equation. A more refined result needs to be used to establish the incompressible limit of the solutions of the Boltzmann equations; see Golse (2005) for details and a complete list of references.

### The Dispersive Property

Consider for the solutions in $\mathbf{R}_x^d \times \mathbf{R}_v^d$ of the elementary kinetic equations

$$\partial_t f + v \cdot \nabla_x f = 0, \quad f(x, v, 0) = f^0(x, v) \qquad [44]$$

the local density

$$\rho(x,t) = \int_{\mathbf{R}_v^d} f(x,v,t)\mathrm{d}v \qquad [45]$$

From the relation

$$\begin{aligned}|\rho(x,t)| &= \int_{\mathbf{R}_v^d} f(x,v,t)\mathrm{d}v \\ &= \int_{\mathbf{R}_v} f^0(x - vt, v, t)\mathrm{d}v \\ &\leq \int_{\mathbf{R}^d} \sup_{w \in \mathbf{R}^d} |f^0(x - vt, w)|\mathrm{d}v \end{aligned} \qquad [46]$$

deduce with an elementary change of variable the following estimate, which carries the name of dispersion lemma,

$$|\rho(x,t)| \leq \frac{1}{|t|^d} \|f^0\|_{L^1(\mathbf{R}_x^d; L^\infty(\mathbf{R}_v^d))} \qquad [47]$$

From interpolation and duality arguments follows:

**Proposition 1** *The macroscopic density* $\rho$ *defined by* [45] *satisfies the inequality*

$$\|\rho\|_{L^q(\mathbf{R}_t; L^p(\mathbf{R}_x^d))} \leq C(d)\|f^0\|_{L^a(\mathbf{R}^{2d})} \qquad [48]$$

*for any choice of real numbers* $a$, $p$, *and* $q$ *such that*

$$\begin{aligned} &1 \leq p < \frac{d}{d-1}, \quad \frac{2}{q} = \frac{d}{1 - \frac{1}{p}} \\ &1 \leq a = \frac{2p}{p+1} < \frac{2d}{2d-1} \end{aligned} \qquad [49]$$

The values $a = 1$, $p = 1$, and $q = \infty$ are obvious. The other limiting values are the interesting ones. They are given by $p = d/(d-1)$, that is, $p = d'$ then $q = 2$ and $a = 2d/(2d-1)$.

These inequalities carry the name of Strichartz inequalities because they are very similar to classical inequalities obtained by Strichartz for the solution of the free Schrödinger equation. This should not be surprising since the Wigner transform of the densities

$$\begin{aligned} f(x,v,t) = \frac{1}{(2\pi)^d} \int & \mathrm{e}^{-\mathrm{i}yv} \psi(x + \tfrac{1}{2}y, t) \\ & \otimes \overline{\psi}(x - \tfrac{1}{2}y, t)\mathrm{d}y \end{aligned} \qquad [50]$$

then turns out to be a solution of the transport equation

$$\partial_t f + v \cdot \nabla_x f = 0 \qquad [51]$$

However, the estimates for kinetic equations are not easily translated into estimations for the Schrödinger equation because the properties of the initial data in terms of norms cannot be simply estimated in terms of the inverse Wigner transform. Spaces with Fourier transform in $L^p, p \neq 2$, are not easy to characterize and not natural for the Schrödinger equation. The above estimates have been very useful in analyzing the large-time behavior of solutions and also in proving the regularity of the three-dimensional Vlasov equation.

### The Entropy and Entropy Dissipation

For solutions of the Boltzmann equation the Boltzmann $H$ function

$$H(f) = \int_{\mathbf{R}^3 \times \mathbf{R}^3} f(x,v) \log f(x,v)\mathrm{d}x\,\mathrm{d}v$$

decreases in time and the same is true for the relative entropy to an absolute Maxwellian $M(v) = (2\pi)^{-3/2}\mathrm{e}^{-|v|^2/2}$:

$$H(F|M) = \int_{\mathbf{R}^3 \times \mathbf{R}^3} \left(f \ln\left(\frac{f}{M}\right) - f + M\right)\mathrm{d}x\,\mathrm{d}v$$

This leads to the systematic introduction in the theory of the notion of relative entropy. It turned out to be instrumental in proving relaxation toward equilibrium of solutions of kinetic (or similar) equations and for the analysis of hydrodynamical limits.

A striking example considered by Desvillettes and Villani is the linearized Fokker–Planck equation in any space dimension:

$$\begin{aligned} &\partial_t F + v \cdot \nabla_x F - \nabla_x V(x) \cdot \nabla_v F \\ &\quad = \nabla_v(\nabla_v F + Fv) \end{aligned} \qquad [52]$$

When $x \mapsto V(x)$ is a smooth potential strictly convex at infinity, this system has a unique steady state given by the relation

$$F_\infty(x,v) = \mathrm{e}^{-V(x)}M(v) = \mathrm{e}^{-V(x)}\frac{\mathrm{e}^{-|v|^2/2}}{(2\pi)^{d/2}} \qquad [53]$$

For any solution of [52] one has

$$\partial_t H(F|M) + \int_{\mathbf{R}^d \times \mathbf{R}^d} F|\nabla_v \log\frac{F}{M}|^2\,\mathrm{d}x\,\mathrm{d}v = 0 \qquad [54]$$

which says that the entropy dissipation is the relative Fisher information (with respect to $v$) of $F$. Now, to study the relaxation to equilibrium, one uses the logarithmic Sobolev inequality:

$$H(F|M) \leq \frac{1}{2}\int_{\mathbf{R}^d \times \mathbf{R}^d} F|\nabla_v \log\frac{F}{M}|^2\,\mathrm{d}x\,\mathrm{d}v \qquad [55]$$

Details, references, and extensions can be found in Arnold *et al.* (2004).

## Conclusions

Kinetic equations have been studied since the end of the nineteenth century, both from the physical and mathematical points of view, but it seems that since the middle of the last century the interest in this approach has considerably increased.

The fact that these equations are well adapted to the description of media which have not "thermalized" (because they are too rarefied or because the domain where they evolve is too small) has been a basic reason for their use in many applied fields; to the ones already quoted one may add the analysis of the air between the reading head and a compact disk, the computations of the characteristics of an ionic motor, and many others.

As a consequence, mathematical progress has been very important. Without going into the details, this contribution is focused on this, and in particular on what can be obtained by the deterministic approach and where the introduction of randomness seems compulsory.

The kinetic formulation turned out to be well adapted to large-scale computers, in particular with Monte Carlo simulations. One should observe that the point of view of modern functional analysis contributes stability estimates to the understanding and improvement of numerical methods. For an introduction to such numerical methods, the reader should first concentrate on the Boltzmann equation itself, which has been one of the basic motivations; consult the book of Sone (2002) the references therein and in particular the book of Bird (1994).

*See also:* Boltzmann Equation (Classical and Quantum); Breaking Water Waves; Einstein's Equations with Matter; Fourier Law; Interacting Stochastic Particle Systems; Nonequilibrium Statistical Mechanics: Dynamical Systems Approach; Partial Differential Equations: Some Examples; Quantum Dynamical Semigroups.

## Further Reading

Arnold A, Carrillo J, Dolbeault J, *et al.* (2004) Entropies and equilibria of many-particle systems: an essay on recent research. *Monatschefte für Mathematik* 142(1): 35–43.

Bird GA (1994) *Molecular Gas Dynamics and the Direct Simulation of Gas Flows.* Oxford: Oxford University Press.

Chavanis PH and Sommeria J (1998) *Monthly Notices of the Royal Astronomical Society* 296: 569.

Erdös L and Yau HT (1998) Linear Boltzmann equation as a scaling weak limit of quantum Lorentz gas. *Advances in Differential Equations and Mathematical Physics, Contemporary Mathematics* 217: 137–155.

Glassey RT (1996) *The Cauchy Problem in Kinetic Theory.* Philadelphia, PA: SIAM.

Golse F (2003a) On the statistics of free-path lengths for the periodic Lorentz gas. In: Zambrini J (ed.) *Proceedings of the XIVth International Congress of Mathematical Physics, Lisbon 2003.* Singapore: World Scientific.

Golse F (2003b) The mean-field limit for the dynamics of large particle systems. *Journées Equations aux dérivées partielles Forges-les-Eaux*, 2 6 juin 2003 GDR 2434 (CNRS).

Golse F (2005) The Boltzmann equation and its hydrodynamic limits. In: Dafermos C and Feireisl E (eds.) *Handbook of Differential Equations, vol. 2: Evolutionary Equations.* Boston: Elsevier.

Olla S, Varadhan R, and Yau HT (1993) Hydrodynamical limit for Hamilton system with weak noise. *Communications in Mathematical Physics* 155: 523–560.

Poupaud F (1994) Mathematical theory of kinetic equations for transport modelling in semiconductors. In: Perthame B (ed.) *Advances in Kinetic Theory and Computing Series in Applied Mathematics*, vol. 22, pp. 141–168. Singapore: World Scientific.

Sone Y (2002) *Kinetic Theory and Fluid Dynamics.* Boston: Birkhäuser.

# Knot Homologies

**J Rasmussen**, Princeton University, Princeton, NJ, USA



## Introduction

A knot homology is a theory which assigns to a knot $K$ (or link $L$) in $S^3$ a graded homology group whose graded Euler characteristic is a knot polynomial associated to $K$. In all known examples, the knot polynomials in question are specializations of the HOMFLY polynomial $P_K(a, q)$, which we take to be determined by the skein relation

$$aP(\times) - a^{-1}P(\times) = (q - q^{-1})P(\circlearrowleft\circlearrowright) \qquad [1]$$

and normalized so that $P$ of the unknot is equal to 1. Let $P_N(K)$ be the specialization of $P_K$ given by

$$P_N(K) = P_K(q^N, q) \qquad [2]$$

Then for each $N \geq 0$, there is a bigraded knot homology $H_N^{i,j}(K)$, which satisfies

$$P_N(K) = \sum_{i,j} (-1)^i q^j \dim H_N^{i,j}(K) \qquad [3]$$

We refer to the first grading $i$ as the homological grading, and the second grading $j$ as the polynomial or $q$-grading.

The idea of a knot homology was introduced by Khovanov (2000) in a seminal paper, in which he defined the homology theory corresponding to the

Jones polynomial ($N=2$). In subsequent work, he defined such a theory for $N=3$, and then, in collaboration with Rozansky, for any $N>0$. Recently, the two authors have introduced a triply graded homology theory $\mathcal{H}^{i,j,k}(K)$ whose graded Euler characteristic gives the entire HOMFLY polynomial:

$$P_K(a,q)=\sum_{i,j,k}(-1)^i q^j a^k \dim \mathcal{H}^{i,j,k}(K) \qquad [4]$$

All of these theories are combinatorial in nature.

In contrast, the knot homology for $N=0$ arises from a very different source – the Heegaard Floer homology of Ozsváth and Szabó. This theory traces its roots back to invariants of 3- and 4-manifolds defined using Seiberg–Witten and Donaldson theory. The definition of $H_0(K)$ is not combinatorial, but because of its connections with these invariants, the theory is known to carry a good deal of geometric information about the knot $K$. The interplay between the two apparently different sorts of knot homologies ($N>0$ and $N=0$) has enhanced our understanding of both sides.

This article will mostly focus on the cases $N=0$ and $N=2$, which are the oldest and best-studied examples of knot homologies and are related to the two best-known specializations of the HOMFLY polynomial – the Alexander and Jones polynomials. We have chosen to use a uniform notation to emphasize the similarities between theories, but the reader should be aware that other notation is more common in the literature. $H_0$ is often referred to as the knot Floer homology (written *HFK*), and is usually normalized with a polynomial grading of $j'=j/2$, corresponding to the substitution $t=q^2$, which gives the standard normalization of the Alexander polynomial. $H_2$ is generally called the reduced Khovanov homology, and often denoted by $\mathrm{Kh}_{\mathrm{r}}$ or $\mathrm{Kh}_{\mathrm{red}}$.

## Construction

Seen from a distance, all knot homologies are defined in much the same way. Given a knot $K$, we must first choose some additional data $D$ which give a concrete geometric presentation of the knot. Using this data, we write down a bigraded chain complex $(C_N^{i,j}(D), d_N)$. This complex depends on our initial choice of $D$, but when we take homology, we are left with groups $H_N^{i,j}(K)$ which are invariants of the knot $K$ (cf. the simplicial homology of a topological space $X$, where the chain groups depend on the choice of some initial geometric data – a triangulation of $X$ – but the homology groups are invariants of $X$).

In all cases, the generators of $C_N(D)$ correspond naturally to terms which appear in a classical model for computing $P_N(K)$. In other words, we can write

$$P_N(K)=\sum_{\sigma\in S}(-1)^{i(\sigma)}q^{j(\sigma)} \qquad [5]$$

where the sum runs over a set of states S determined by $D$, and the functions $i$ and $j$ are also determined by $D$. $C_N^{i,j}(D)$ is the free abelian group generated by $\{\sigma\in S|i(\sigma)=i, j(\sigma)=j\}$ and the differential $d_N$ is chosen to preserve the $j$-grading: $j(d_N x)=j(x)$. It follows that $C_N(D)$ decomposes into an infinite direct sum of complexes, one for each value of $j$, and [3] is a consequence of [5].

Beyond these global similarities, the definition of $C_N(D)$ varies with the value of $N$. In the second half of the article, we give explicit details of the constructions for $N=0$ and $N=2$.

## Filtered Complexes and Deformations

An important characteristic shared by all the $C_N$'s is the existence of deformations with homology $\mathbb{Z}$. Recall that $(C_N(D), d_N)$ is a graded chain complex: $j(d_N x)=j(x)$. By a deformation of such a complex, we mean a new chain complex $(C_N(D), d_N+d'_N)$ in which the underlying group remains the same, but the differential has been perturbed by the addition of a new term $d'_N$ which strictly raises the $j$-grading: $j(d'_N(x))>j(x)$.

Any deformation of a graded complex is naturally a filtered complex, and as such, gives rise to a spectral sequence. The $E_0$ term of this spectral sequence is the original unperturbed complex $(C_N(D), d_N)$, so the underlying group of the $E_1$ term is just $H_N(K)$. Thus, it is independent of the choice of initial data $D$. In fact, it can often be shown that all terms in the spectral sequence beyond the first one are invariants of $K$. This is known to be the case for $N=0$ and $N=2$, and is most likely true for all other $N$ as well (cf. the Leray–Serre sequence associated to a fibration, where the first two terms depend on a choice of geometric data but the $E_2$ and higher terms are all invariants of the fibration).

For each value of $N$, $C_N(D)$ admits a natural deformation whose homology is $\mathbb{Z}$ in homological grading 0, and zero in every other grading. When $N=0,2$, the filtration grading of this generator is known to be an invariant of $K$. (This is probably the case for $N>2$ as well.) Equivalently, this is the $j$-grading of the surviving copy of $\mathbb{Z}$ in the spectral sequence. When $N=0$, this invariant is conventionally normalized to be half the $j$-grading of the

generator, and is called $\tau(K)$. When $N=2$, it is called $s(K)$.

## Geometric Properties

Some elementary properties of the $H_N$'s generalize those of the HOMFLY polynomial. If $K_1\#K_2$ denotes the connected sum of $K_1$ and $K_2$, then over $\mathbb{Q}$

$$H_N(K_1\#K_2) \cong H_N(K_1) \otimes H_N(K_2) \qquad [6]$$

and if $\bar{K}$ is the mirror image of $K$,

$$H_N^{i,j}(\bar{K}) \cong H_N^{-i,-j}(K) \qquad [7]$$

Moreover, $H_0$ satisfies an additional symmetry

$$H_0^{i,j}(K) \cong H_0^{i-j,-j}(K) \qquad [8]$$

generalizing the symmetry of the Alexander polynomial: $P_0(q)=P_0(q^{-1})$. (With integer coefficients, these equalities all hold at the chain level. The correct statements about the homology can be obtained from the Kunneth formula and universal coefficient theorem.)

$H_N(K)$ also contains deeper information related to the genus of surfaces bounding $K$. If $K$ is a knot in $S^3$, recall that $g(K)$ – the Seifert genus of $K$ – is the minimal genus of an orientable surface smoothly embedded in $S^3$ and bounding $K$. If we view $S^3$ as the boundary of the 4-ball $B^4$, we can define a second quantity $g_*(K)$ – the slice genus – by relaxing the requirement that the surface be embedded in $S^3$ and instead requiring it to be embedded in $B^4$.

Both $s(K)$ and $\tau(K)$ give lower bounds on the slice genus of $K$:

$$|\tau(K)| \leq g_*(K) \qquad [9]$$

$$|s(K)| \leq 2g_*(K) \qquad [10]$$

These bounds are far from independent. In fact, in all known examples, $s(K)=2\tau(K)$. It is an open problem to determine whether this is true for all knots.

From [6], it follows that $s$ and $\tau$ are additive under connected sum. Thus, both invariants define homomorphisms from the concordance group of knots in $S^3$ to $\mathbb{Z}$. The inequalities in eqns [9] and [10] are not always sharp, but there is one case where equality is known to hold. This is when $K$ is represented by a diagram with all positive crossings (or, more generally, $K$ is quasipositive.) In this case, the slice genus is also equal to the Seifert genus, and all three are easily computed using Seifert's algorithm.

The proof of [10] depends on the fact that for $N>0, H_N$ is functorial in the following sense. If $S \subset S^3 \times [0,1]$ is a smoothly embedded, orientable cobordism between links $L_1$ and $L_2$, then for each $N>0$, there is an induced map $\phi_N^S : H_N(L_1) \to H_N(L_2)$. $\phi_N^S$ is a graded map: it preserves the homological grading, and lowers the $j$-grading by $(N-1)\chi(S)$. Under deformation, it becomes a filtered map which induces a rational isomorphism on the deformed homologies.

### $H_0$ and Heegaard Floer Homology

The proof of [9] depends on the close connection between the knot Floer homology and the Heegaard Floer homology. Roughly speaking, the Heegaard Floer groups of 3-manifolds obtained by surgery on $K$ are determined by the groups $H_0^{i,j}(K)$ together with additional differentials obtained by relaxing the requirement that $n_z(\phi)=n_w(\phi)=0$. The relation with the slice genus again arises by studying maps induced by cobordisms, but in this case, the relevant cobordism is the surgery cobordism between $S^3$ and the 0-surgery on $K$.

This connection also leads to another important property of $H_0$: it detects the Seifert genus. If we let $M(K)$ be the largest value of $j$ for which the group $H_0^{*,j}(K)$ is nontrivial, then

$$M(K) = 2g(K) \qquad [11]$$

This fact generalizes a well-known inequality involving the degree of the Alexander polynomial: if $m(K)$ is the largest power of $q$ appearing in $P_0(K)$, then $m(K) \leq 2g(K)$.

## Computations

The difficulty of computing $H_N(K)$ varies with the value of $N$. When $N=1$, the theory is essentially trivial: $H_1^{0,0}(K) \cong \mathbb{Z}$ for any knot $K$, and all other groups vanish. Of the remaining knot homologies, $H_2(K)$ is the easiest to compute. The theory for alternating knots was worked out by E S Lee, and extensive calculations have also been made for nonalternating knots using computer programs written by Bar-Natan and Shumakovitch.

Computing $H_0$ is more difficult, on account of the noncombinatorial nature of $d_0$. Three families of knots for which $H_0$ is well understood are alternating knots, (1,1) knots (described in the next section), and knots which admit lens space surgeries. Beyond this, there is an array of techniques which may or may not work in any given case. The best of these is probably a setup introduced by Ozsváth and Szabó, in which the generators of $C_0(D)$ correspond to states in the Kauffman state model of the Alexander polynomial. Combining this method with the known

results for alternating knots and (1,1) knots gives a fairly good understanding of $H_0(K)$ for knots with 10 or fewer crossings; for larger knots, relatively little is known.

Few computations of $H_N$ for $N > 2$ have been made, although the definition in this case is purely combinatorial.

### Thin and Thick Knots

For simple knots, both $H_0$ and $H_2$ are thin. This means that there exists a constant $c_N(K)(N = 0, 2)$ such that $H_N^{i,j}(K)$ is trivial unless $j - 2i = c_N(K)$. In such cases, we necessarily have $c_0(K) = 2\tau(K)$ (resp. $c_2(K) = s(K)$), and $H_N(K)$ is completely determined by $c_N(K)$ and $P_N(K)$. The relationship is best expressed in terms of the Poincaré polynomial of $H_N(K)$:

$$\mathcal{P}_N(K) = \sum_{i,j} t^i q^j \dim H_N^{i,j}(K) = (-t)^{-c_N(K)/2} P_N(K)(q(-t)^{1/2}) \qquad [12]$$

If $K$ is an alternating knot, both $H_0(K)$ and $H_2(K)$ are thin, and $c_0(K) = c_2(K) = \sigma(K)$. (Note that in this case the bound on $g_*(K)$ coming from $\tau$ and $s$ coincides with the classical bound coming from the signature.) Many nonalternating knots are thin as well; in all examples in which both groups have been computed, either both $H_0(K)$ and $H_2(K)$ are thin, or neither is. In addition, all such knots appear to have $c_0(K) = c_2(K) = \sigma(K)$.

Those knots whose homologies are not thin are called *thick*. There are a dozen such knots with ten or fewer crossings: using the standard numbering in the knot tables (see, e.g., Rolfsen (1976)) these are $8_{19}, 9_{42}, 10_{124}, 10_{128}, 10_{132}, 10_{136}, 10_{139}, 10_{145}, 10_{152}, 10_{153}, 10_{154}$, and $10_{161}$. It is a curious and as yet unexplained coincidence that, for all of these knots, the ranks of $H_0(K)$ and $H_2(K)$ are equal.

There is an analogous notion of thinness when $N > 2$, but there exist alternating knots for which $H_N$ cannot be thin for $N \gg 0$ (this can be seen from the HOMFLY polynomials).

## Construction of $H_0$

We now turn to a more detailed description of the definition of $H_0(K)$. The geometric data $D$ used to define $C_0$ is a Heegaard diagram for the complement of $K$. One convenient way to specify such a diagram is by a doubly pointed Heegaard diagram of $S^3$. The data for such a diagram consist of a surface $\Sigma$ of genus $g$, two $g$-tuples of attaching circles $\{\alpha_1, \ldots, \alpha_g\}$ and $\{\beta_1, \ldots, \beta_g\}$ on $\Sigma$, and two points $z, w \in \Sigma$ which are disjoint from all the $\alpha$'s and $\beta$'s. Each set of attaching circles is composed of $g$ disjoint simple closed curves, arranged so that when $\Sigma_g$ is cut along them the result is a sphere with $2g$ holes. Any such set of attaching circles determines a unique genus-$g$ handlebody $H$ with boundary $\Sigma$ and the property that each attaching circle bounds a disk in $H$.

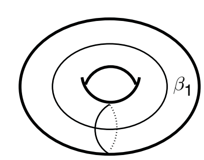



**Figure 1** Heegaard splitting of $S^3$ corresponding to the standard decomposition of $S^3$ into two solid tori.

The choice of $\alpha$ and $\beta$ curves determines the underlying 3-manifold in which the knot is embedded. Starting with $\Sigma \times [0,1]$, we fill in one component of the boundary with the handlebody determined by the $\alpha$-curves, and the other component with the handlebody determined by the $\beta$-curves to obtain a closed 3-manifold. By hypothesis, this manifold is required to be $S^3$. A simple Heegaard diagram of $S^3$ with $g = 1$ is shown in Figure 1.

To go from a doubly pointed Heegaard diagram to a diagram of the knot complement, we remove neighborhoods of $z$ and $w$ and replace them with a tube to get a surface $\Sigma'$ of genus $g + 1$. We also add an additional $\alpha$-handle $\alpha_{g+1}$, which runs from $z$ to $w$ in $\Sigma$ in such a way that it does not intersect the other $\alpha$'s, and then comes back over the tube. This process is illustrated in Figure 2.

A Heegaard diagram of $S^3 - K$ determines a presentation of $\pi_1(S^3 - K)$ with one generator $x_i$ for each $\alpha$-circle and one relator $w_j$ for each $\beta$-circle. To find the relator $w_j$, one travels along $\beta_j$, recording each intersection with some $\alpha_i$ by appending $x_i^{\pm 1}$ to the relator. The sign is determined by the sign of the intersection. As an example, consider the two doubly pointed diagrams of Figure 3, both of which correspond to the same Heegaard diagram of $S^3$. (It is isotopic to the one shown in Figure 1.) The fundamental groups of the associated knot complements can be read off from the corresponding genus-2 Heegaard splittings. Starting from the point where

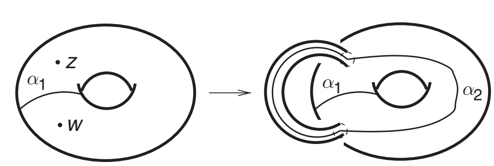
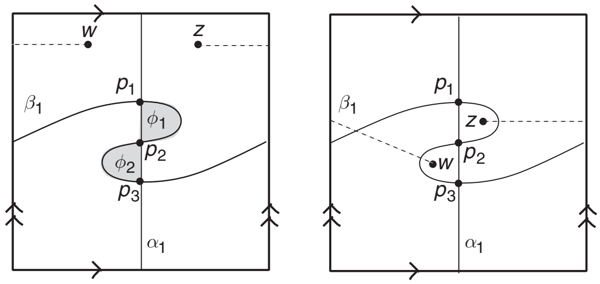



**Figure 2** Going from a doubly pointed diagram to a Heegaard diagram of the knot complement.




(a) (b)

**Figure 3** Doubly pointed Heegaard diagrams for the unknot and the trefoil. Opposite sides of the square are identified to form a torus. The dotted line represents $\alpha_2$.

$\beta_1$ intersects the left-hand side of the square and moving to the right, we get

$$\pi_1(S^3 - K_1) = \langle x_1, x_2 | x_1 x_1^{-1} x_1 = 1 \rangle$$
$$\pi_1(S^3 - K_2) = \langle x_1, x_2 | x_2 x_1 x_2^{-1} x_1^{-1} x_2^{-1} x_1 = 1 \rangle$$

The first group is isomorphic to $\mathbb{Z}$, and the knot in Figure 3a is the unknot. The second is isomorphic to $\pi_1$ of the complement of the trefoil knot, and in fact the knot in Figure 3b is the left-handed trefoil.

The definition of $C_0(D)$ is based on a classical method for computing the Alexander polynomial known as the Fox calculus, which takes as its input a presentation of $\pi_1(S^3 - K)$. According to Fox calculus,

$$P_0(K) = \pm q^n \det(d_{x_i} w_j)_{1 \le i,j \le g} \qquad [13]$$

Here $d_{x_i} w_j$ is an element of the group ring

$$\mathbb{Z}[H_1(S^3 - K)] \cong \mathbb{Z}[q^{\pm 2}]$$

It is determined by the following rules:

$$d_{x_i} x_j = \delta_{ij} \qquad [14]$$

$$d_{x_i} ab = d_{x_i} a + |a| d_{x_i} b \qquad [15]$$

$$d_{x_i} x_i^{-1} = -|x_i^{-1}| \qquad [16]$$

where

$$|\cdot| : \pi_1(S^3 - K) \to H_1(S^3 - K) \cong \mathbb{Z} = \langle q^2 \rangle \qquad [17]$$

is the abelianization map. The factor of $\pm q^n$ is chosen so that $P_0(K)(1) = 1$ and $P_0(K)(q) = P_0(K)\ (q^{-1})$.

As an example, consider the two presentations above. In the first presentation, $|\cdot|$ sends $x_1$ to 1 and $x_2$ to $q^2$, so

$$\begin{aligned} d_{x_1}(x_1 x_1^{-1} x_1) &= 1 - |x_1 x_1^{-1}| + |x_1 x_1^{-1}| \\ &= 1 - 1 + 1 \\ &= 1 \end{aligned} \qquad [18]$$

which is the Alexander polynomial of the unknot. If we abelianize the relator in the second presentation, we see that $|x_1| = |x_2| = q^2$, so

$$d_{x_1}(x_2 x_1 x_2^{-1} x_1^{-1} x_2^{-1} x_1)$$
$$= |x_2| - |x_2 x_1 x_2^{-1} x_1^{-1}| + |x_2 x_1 x_2^{-1} x_1^{-1} x_2^{-1}| \qquad [19]$$
$$= q^2 - 1 + q^{-2} \qquad [20]$$

which is the Alexander polynomial of the trefoil.

When $g = 1$, the complex $C_0(D)$ is generated by the points of $\alpha_1 \cap \beta_1$. These intersection points may be naturally identified with the appearances of the generator $x_1$ in $w_1$, and thus with the monomials appearing in $d_{x_1} w_1$. For example, the three monomials which appear on the right-hand sides of eqns [18] and [19] correspond, respectively, to the points labeled $p_1, p_2$, and $p_3$ in Figure 3. The $j$-grading of each generator is given by the exponent of $q$ which the corresponding monomial contributes to the Alexander polynomial. Thus, all three generators in Figure 3a have $j$-grading 0, while in Figure 3b, the generators $p_1, p_2$, and $p_3$ have $j$-gradings 2, 0, and –2 respectively.

For general $g$, the monomials appearing in the determinant of eqn [13] correspond to intersection points of the two totally real tori $\alpha = \alpha_1 \times \cdots \times \alpha_g$ and $\beta = \beta_1 \times \cdots \times \beta_g$ inside the symmetric product $\mathrm{Sym}^g \Sigma$. The knot Floer homology is the Lagrangian Floer homology of $\alpha$ and $\beta$ inside the symplectic manifold $\mathrm{Sym}^g(\Sigma - z - w)$. The generators of $C_0(D)$ are the points of $\alpha \cap \beta$; the differential is defined by counting holomorphic disks with boundary on $\alpha$ and $\beta$. To be precise, for $x \in \alpha \cap \beta$,

$$d_0 x = \sum_{\substack{\phi \in \pi_2(x,y), \mu(\phi)=1 \\ n_z(\phi) = n_w(\phi) = 0}} \#\mathcal{M}(\phi) y \qquad [21]$$

Here $\pi_2(x, y)$ denotes the set of homotopy classes of maps of the strip $D = \{a + ib \mid b \in [0, 1]\}$ into $\mathrm{Sym}^g \Sigma$ which take the right-hand boundary to $\alpha$ and the left-hand boundary to $\beta$, and which limit to $x$ as $b \to -\infty$ and to $y$ as $b \to \infty$. $\mu(\phi)$ denotes the formal dimension of the space of pseudoholomorphic disks in this homotopy class. There is a natural action by translation on the space of such maps, so when $\mu(\phi) = 1$ we can divide out by this action and obtain an oriented zero-dimensional moduli space $\mathcal{M}(\phi)$. Finally, by $n_z(\phi)$ and $n_w(\phi)$ we denote the intersection number of such a strip with the divisors determined by $z$ and $w$ inside of $\mathrm{Sym}^g \Sigma$. The requirement that they vanish forces the strip to lie

in $\mathrm{Sym}^g(\Sigma - z - w)$. It can be shown that, for $\phi \in \pi_2(x, y)$,

$$j(x) - j(y) = n_z(\phi) - n_w(\phi) \qquad [22]$$

so $j(d_0 x) = j(x)$.

When $g = 1$, computing the differential amounts to counting maps of the strip into the Heegaard torus. This can be done algorithmically using the Riemann mapping theorem, so computation of $H_0$ is purely combinatorial. Knots of this form are called (1,1) knots. They are one of our few windows into the behavior of $H_0$ for large knots.

As an example, consider the diagram of Figure 3a. The two shaded regions represent the domains of classes $\phi_1 \in \pi_2(p_1, p_2)$ and $\phi_3 \in \pi_2(p_3, p_2)$. The Riemann mapping theorem implies that up to reparametrization, there is a unique holomorphic map of the strip into each region, so $\#\mathcal{M}(\phi_1) = \pm 1 = \#\mathcal{M}(\phi_2)$. The differential in $C_0(D_1)$ is given by

$$d_0(p_1) = \pm p_2 = d_0(p_3)$$
$$d_0(p_2) = 0$$

and $H_0(U) \cong \mathbb{Z}$. This reflects the fact that we could have chosen the more efficient diagram of $S^3 - U$ shown in Figure 1, simply by moving $\beta_1$ to remove two of the intersection points.

For comparison, consider the diagram for the trefoil shown in Figure 3b. All three generators of $C_0(D_2)$ have different $j$-gradings, so we must have $d_0 \equiv 0$. Thus, $H_0(T) \cong \mathbb{Z}^3$. The two disks $\phi_1$ and $\phi_2$ are still present, but now $n_z(\phi_1) = n_w(\phi_2) = 1$, so neither disk contributes to the differential. This is reflected in the fact that $\beta_1$ cannot be moved to reduce the number of intersection points without passing through either $z$ or $w$.

#### Deformations

In this case, finding an appropriate deformation of $C_0(D)$ is simple: we just drop the condition that $n_z(\phi) = 0$ in the definition of the differential. If a homotopy class $\phi \in \pi_2(x, y)$ contributes nontrivially to the sum, it must have a holomorphic representative, which necessarily intersects the divisor in $\mathrm{Sym}^g\Sigma$ defined by $z$ non-negatively. Thus, $n_z(\phi) \geq 0$. From [22], it follows that $j(x) - j(y) = n_z(\phi) \geq 0$, so this new differential has the form $d_0 + d_0'$, where $d_0'$ strictly lowers the $j$-grading.

The fact that the homology of $C_0(D)$ with respect to the perturbed differential is $\mathbb{Z}$ goes back to the knot Floer homology's roots in Heegaard Floer homology. By dropping the condition that $n_z(\phi) = 0$, we have effectively forgotten about the basepoint $z$, and thus about the knot. The new complex simply computes the Heegaard Floer group $\widehat{HF}(S^3)$, which is isomorphic to $\mathbb{Z}$. When $g = 1$, this can be seen directly: if we remove the basepoint $z$, any genus-1 Heegaard diagram of $S^3$ can be isotoped into the standard diagram of Figure 1.

### Construction of $H_2$

In this case, the geometric data $D$ needed to define the chain complex $C_2(D)$ is a planar diagram of the knot, and the classical model on which the construction of $C_2(D)$ is based is the Kauffman state model for the Jones polynomial. There is a related homology theory $\tilde{H}_2(D)$, known as the unreduced Khovanov homology, whose graded Euler characteristic is $(q + q^{-1})P_2(K)$. This is the original categorification of the Jones polynomial defined in Khovanov (2000).

To construct $\tilde{C}_2(D)$, we consider complete resolutions of the planar diagram $D$. As shown in Figure 4, there are two different ways to resolve each crossing of $D$. If $D$ has $n$ crossings, there will be $2^n$ ways to resolve all $n$, one for each vertex of the cube $[0, 1]^n$. To a vertex $v$, we associate the crossingless planar diagram $D_v$ obtained from the corresponding resolution of $D$. Thus, each vertex of the cube is decorated by a 1-manifold $D_v$.

If $e$ is an edge joining vertices $v_0$ and $v_1$ (where $v_0$ has one more 0 coordinate than $v_1$), we write $e: v_0 \to v_1$, and decorate $e$ with a two-dimensional cobordism $S_e$ from $D_{v_0}$ to $D_{v_1}$. $S_e$ is a product cobordism outside a neighborhood of a single crossing, where it is the one-handle cobordism between the 0-resolution and the 1-resolution. The resulting cobordism is necessarily composed of a union of product cobordisms (cylinders) together with a single nontrivial cobordism (a pair of pants). Thus, starting from $D$, we have constructed an $n$-dimensional cube whose vertices are decorated by 1-manifolds and whose edges are decorated by cobordisms between them. This is the cube of resolutions of $D$.

The next step in the construction of $\tilde{C}_2(D)$ is to apply a graded $(1 + 1)$-dimensional TQFT $\mathcal{A}$ to the cube of resolutions. $\mathcal{A}$ is a functor which associates to each 1-manifold $X$ a group $\mathcal{A}(X)$, and to each two-dimensional cobordism $W: X_1 \to X_2$ a homomorphism $\mathcal{A}(W): \mathcal{A}(X_1) \to \mathcal{A}(X_2)$. If we apply $\mathcal{A}$ to all the manifolds and cobordisms of the cube of

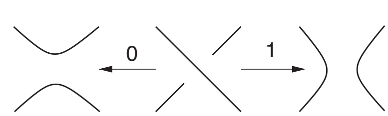



**Figure 4** 0- and 1-resolutions of a crossing.

**Table 1** Summary of cube of resolutions

| Vertex $v$ | $\to$ | 1-*manifold* $D_v$ | $\to$ | Group $\mathcal{A}(D_v)$ |
|---|---|---|---|---|
| Edge $e: v_1 \to v_2$ | $\to$ | Cobordism $S_e: D_{v_1} \to D_{v_2}$ | $\to$ | Homomorphism $\mathcal{A}(S_e): \mathcal{A}(D_{v_1}) \to \mathcal{A}(D_{v_2})$ |

resolutions, we obtain a new cube, decorated with groups and cobordisms between them. This process is summarized in Table 1.

We can now describe the chain complex $\tilde{C}_2(D)$. As a group,

$$\tilde{C}_2(D) = \bigoplus_v \mathcal{A}(D_v) \quad [23]$$

where the sum runs over all vertices of the cube of resolutions. For $x \in \mathcal{A}(D_v)$, the differential is given by

$$d_2 x = \sum_{e:v \to v'} (-1)^{s(e)} \mathcal{A}(S_e)(x) \quad [24]$$

The signs in this sum are determined by assigning a sign $(-1)^{s(e)}$ to each edge $e$ in such a way that every two-dimensional face of the cube has an odd number of $-$ signs on its edges. (This ensures that $d^2 = 0$.) There are many ways to do this, but they all result in isomorphic complexes.

The homological grading $i$ on $\tilde{C}_2(D)$ is easily determined. For $x \in \mathcal{A}(D_v)$, we set $i(x) = i(v) - c(D)$, where $i(v)$ is the sum of all the coordinates of $v$, and $c(D)$ is a constant. Clearly, $i(d_2 x) = i(x) + 1$. In order to have invariance, it turns out that $c(D)$ must be chosen to be equal to the number of negative crossings in $D$.

It remains to specify the TQFT $\mathcal{A}$. At the level of groups, $\mathcal{A}(S^1)$ is a free abelian group of rank 2:

$$\mathcal{A}(S^1) = A = \langle 1, X \rangle \quad [25]$$

General principles then imply that

$$\mathcal{A}\Big(\coprod^n S^1\Big) = A^{\otimes n} \quad [26]$$

To specify the maps induced by cobordisms, it is enough to describe the maps associated to the two pairs of pants shown in Figure 5. They are given by

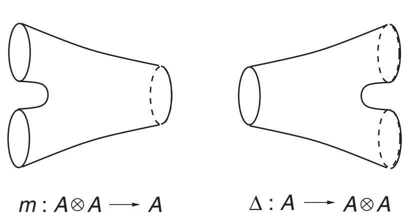



**Figure 5** Maps induced by pairs of pants.

$$m(1 \otimes 1) = 1$$
$$\Delta(1) = 1 \otimes X + X \otimes 1 \quad [27]$$

$$m(1 \otimes X) = m(X \otimes 1) = X$$
$$\Delta(X) = X \otimes X \quad [28]$$

$$m(X \otimes X) = 0 \quad [29]$$

Note that the multiplication $m$ makes $A$ into a commutative ring isomorphic to $\mathbb{Z}[X]/(X^2)$.

$\mathcal{A}$ is a graded TQFT. In other words, there is a grading $q$ on $A$ and its tensor products, determined by

$$q(1) = 1$$
$$q(a \otimes b) = q(a) + q(b) \quad [30]$$

$$q(X) = -1 \quad [31]$$

From eqns [27]–[29], it is easy to see that

$$q(m(a \otimes b)) = q(a \otimes b) - 1$$
$$q(\Delta(a)) = q(a) - 1 \quad [32]$$

If we define $j(x) = k(D) + q(x) + i(x)$, it follows that $j(d_2 x) = j(x)$. Taking the graded Euler characteristic gives

$$\chi(\tilde{C}_2(D)) = q^{k(D)} \sum_v (-q)^{i(v)} (q + q^{-1})^{n_v} \quad [33]$$

where $n_v$ is the number of components of $D_v$. If we define $k(D)$ to be the writhe of $D$, this is precisely Kauffman's formula for the unnormalized Jones polynomial.

Figure 6 illustrates $\tilde{C}_2(D)$ for a simple two-crossing link. The figure shows the original link (in the center), the cube of resolutions, and basis vectors for $\tilde{C}_2(D)$, together with their $j$-gradings. We leave it to the reader to check that the homology $\tilde{H}_2(L)$ is four dimensional, supported in $j$-gradings 1 and 3 at the vertex labeled 00, and in gradings 5 and 7 at the vertex labeled 11.

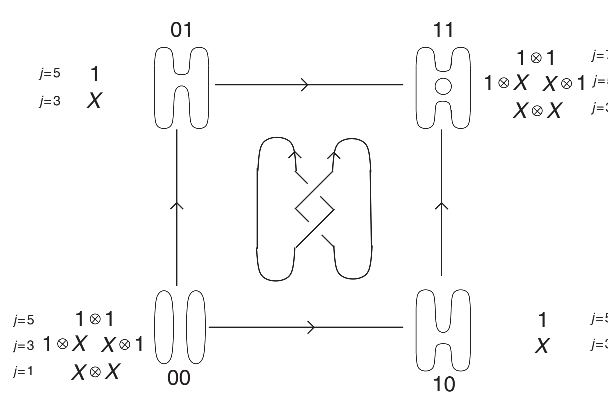



**Figure 6** The cube of resolutions for the Hopf link.

To get the reduced chain complex $C_2(D)$, we must divide the graded Euler characteristic by a factor of $(q+q^{-1})$. This is accomplished by choosing a marked point on $K$ and requiring that for each resolution $D_v$, the vector associated to the circle containing the marked point lie in the subspace of $A$ spanned by $X$. If $D$ is a diagram of a knot, the resulting homology $H_2(K)$ is independent of the choice of marked point. For links, $H_2(L)$ depends on the component of the link on which the marked point lies.

### Deformations

Deformations in the $N=2$ theory are constructed using a technique introduced by E S Lee. The idea is to replace the graded TQFT $\mathcal{A}$ with a filtered TQFT $\mathcal{A}'$. As a group, we still have $\mathcal{A}(S^1)=A$, but the multiplication and comultiplication maps are perturbations of those for $\mathcal{A}$:

$$m'(1\otimes 1)=1 \qquad \Delta'(1)=1\otimes X+X\otimes 1-r1\otimes 1 \quad [34]$$

$$m'(1\otimes X)=m'(X\otimes 1)=X \qquad \Delta'(X)=X\otimes X+s1\otimes 1 \quad [35]$$

$$m'(X\otimes X)=rX+s \quad [36]$$

The new terms involving $r$ and $s$ have $q$ gradings strictly greater than the terms which are shared with eqns [27]–[29]. Thus, the differential defined by replacing $m$ and $\Delta$ by $m'$ and $\Delta'$ will be a perturbation of the original differential on $\tilde{C}_2(D)$.

The simplicity of the homology with respect to the new differential depends on the fact that when the polynomial $X^2-rX-s$ has simple roots, the TQFT $\mathcal{A}'$ decomposes as a direct sum of two one-dimensional TQFTs. This implies that for a knot, the deformed homology $\tilde{H}_2'(K)$ decomposes as a direct sum of two copies of $H_1(K)$. This group is always isomorphic to $\mathbb{Z}$, so $\tilde{H}_2'(K)\cong\mathbb{Z}\oplus\mathbb{Z}$. If $s=0$, the same strategy can be used to define deformations of the reduced chain complex $C_2(D)$. In this case, we find that the deformed homology is isomorphic to a single copy of $\mathbb{Z}$.

*See also:* Floer Homology; Gauge Theory: Mathematical Applications; The Jones Polynomial; Knot Theory and Physics; Topological Quantum Field Theory: Overview.

### Further Reading


Bar-Natan D (2002) On Khovanov's categorification of the Jones polynomial. *Algebraic and Geometric Topology* 2: 337–370.
Crowell R and Fox R (1963) *Introduction to Knot Theory*. Boston: Ginn and Co.
Kauffman L (1987) State models and the Jones polynomial. *Topology* 26: 395–407.
Khovanov M (2000) A categorification of the Jones polynomial. *Duke Mathematical Journal* 101: 359–426.
Ozsváth P and Szabó Z (2004) Heegaard diagrams and holomorphic disks. In: Donaldson S, Eliashberg Y, and Gromov M (eds.) *Different Faces of Geometry*, pp. 301–348. New York: Kluwer/Plenum.
Ozsváth P and Szabó Z (2004) Holomorphic disks and topological invariants for closed three-manifolds. *Annals of Mathematics* 159: 1027–1158.
Ozsváth P and Szabó Z (2004) Holomorphic disks and knot invariants. *Advances in Mathematics* 186: 58–116.
Rolfsen D (1976) *Knots and Links*, Mathematics Lecture Series, No. 7. Berkeley: Publish or Perish.
Viro O (2004) Khovanov homology, its definitions and ramifications. *Fundamentals of Mathematics* 184: 317–342.


## Knot Invariants and Quantum Gravity


**R Gambini**, Universidad de la República, Montevideo, Uruguay
**J Pullin**, Louisiana State University, Baton Rouge, LA, USA




### Introduction

As in all other physical theories, one expects that gravitational phenomena will ultimately be ruled by quantum mechanics. This requires to consider the quantization of the best available theory of gravity, namely Einstein's general relativity. This problem has been considered since the 1930s (*see* Loop Quantum Gravity). The application of the rules of quantum mechanics to general relativity is immediately problematic. Unlike other physical interactions, general relativity describes gravitational phenomena through a distortion of spacetime rather than through a field living in spacetime. Therefore, its quantization is bound to be very different from that of other physical theories. In particular, the well-established framework of perturbative quantum field theory, used with remarkable success in describing electroweak and strong interactions (in the latter case at least in certain regimes), runs into trouble when applied to general relativity. At present, it is not clear if this is a fundamental problem or if there might exist an implementation of perturbative quantum field theory that works well in the gravitational case. On the

other hand, there exist examples of field theories where perturbative methods fail but that nevertheless can be quantized. This suggests that the consideration of nonperturbative techniques in the quantization of the gravitational field could be a promising avenue.

In particular, canonical quantization methods appear attractive for attempting a nonperturbative quantization of gravity. Canonical methods force the introduction, in a clear way, of a Hilbert space of states and definition of the quantum operators of interest. The application of canonical methods to classical general relativity was pioneered by Dirac and Bergmann in the late 1950s. During the 1960s, the resulting canonical theories were considered in a quantum setting by DeWitt. At the time it appeared that making progress in the canonical quantization of general relativity was going to be quite a challenge. In particular, the canonical theory has constraints, which have to be implemented as operator identities quantum mechanically. The wave functions were functionals of the spatial metric of spacetime. One of the operator identities to be satisfied implies that the wave functions only depend on properties of the spatial metric that are invariant under spatial diffeomorphisms. This is a direct consequence of general relativity being a theory that is independent of coordinate choice since a diffeomorphism changes the assignment of coordinates to points in the manifold. Finding such wave functions already presented a challenge, since there is no well-grounded mathematical theory of functionals of diffeomorphism-invariant classes of metrics. Moreover, the other operator identity to be imposed, known as the Hamiltonian constraint or Wheeler–DeWitt equation, was a nonpolynomial complicated operator equation that does not admit a simple geometrical interpretation and needs to be regularized. Since one does not have a background metric to rely upon, traditional regularization techniques of quantum field theory are not suitable to deal with the Hamiltonian constraint.

These difficulties severely hampered development of canonical methods for the quantization of general relativity for approximately two decades. The situation started to change when Ashtekar noticed that one could choose a different set of variables to describe general relativity canonically. Instead of using as variable the spatial metric $q_{ab}$, Ashtekar chooses to use a set of (densitized) frame fields $\tilde{E}_i^a$. The relationship between the metric and the densitized frames is $\det(q^{ab})q^{ab} = \tilde{E}_i^a \tilde{E}_i^b$ and we are assuming the Einstein summation convention, that is, the index $i$ is summed from 1 to 3 (such an index labels which vector in the triad one is referring to). The resulting theory has an additional symmetry with respect to usual general relativity, in the sense that it is invariant under the choice of frame. This symmetry operates on the index $i$ as if it were an SO(3) symmetry. As canonical momenta the usual choice is to pick the extrinsic curvature of the 3-geometry. Ashtekar chooses a variable related to it that behaves under frame transformations as an SO(3) connection, $A_a^i$. The resulting theory is therefore cast in terms of a canonical pair $(\tilde{E}_i^a, A_a^i)$, with $i$ an SO(3) index. One can therefore consider the canonical pair as that of a Yang–Mills theory associated with the SO(3) group. In fact, associated with the extra symmetry under triad rotations the theory has a new set of constraints that take the form of a Gauss law, $D_a\tilde{E}_i^a = 0$ with $D_a$ the covariant derivative formed with the connection $A_a^i$. This allows us to view the phase space of a Yang–Mills theory as the kinematical arena on which to discuss quantum gravity. The theory is of course different from the Yang–Mills theory. In particular, it still has constraints that imply that it is invariant under spacetime diffeomorphisms. In the canonical picture, these constraints appear asymmetrically as one constraint is associated with time evolution ("Hamiltonian constraint") and a set of three constraints is associated with spatial diffeomorphisms ("diffeomorphism constraint").

If one quantizes the theory starting from the Ashtekar formulation, given the resemblance with Yang–Mills theory, the natural choice for a representation of the quantum wave functions is to consider wave functions of the connection $\Psi[A]$ that are invariant under SO(3) transformations. Such a representation is known as "connection representation." There is significant experience in Yang–Mills theory in constructing such wave functions. In particular, it is known that if one considers the parallel transport operator defined by a connection around a closed curve (holonomy) and one takes its trace ("Wilson loop"), the resulting object is invariant under SO(3) transformations. What is more important, the set of traces of holonomies along all possible closed loops is an overcomplete basis for all gauge-invariant functions. More recently, it has been shown that one can construct a less redundant complete basis using techniques from spin networks. We will discuss later on how to do this.

Since any gauge-invariant functional can be expanded in the basis of Wilson loops, one can choose to represent it through the coefficients of such an expansion. These coefficients are functions of the curve upon which the corresponding element of the basis of Wilson loops is based. The representation of wave functions in terms of such coefficients is called "loop representation." Wave

functions in the loop representation are functions of a closed curve (more precisely of families of closed curves, or spin networks, as we will discuss below).

We still have to deal with the diffeomorphism and Hamiltonian constraints. The diffeomorphism constraint when written in the loop representation implies that the wave functions are not functions of loops but rather of topologically invariant properties of the loops under general diffeomorphisms of the spatial manifold containing the loops. Such functions are technically known in the mathematical literature as "knot invariants." This is the first point of connection between knot invariants and quantum gravity; they constitute the kinematical arena of the theory. One still has to deal with the Hamiltonian constraint, which has to be imposed as an operator equation. We shall see that knot theory also seems to have a lot to say about solutions of the Hamiltonian constraint. This is quite remarkable, since the Hamiltonian constraint embodies in detail the specific dynamics of Einstein's theory of gravitation, and to our knowledge this is an input that has never gone into the ideas of knot theory.

In terms of the Ashtekar variables, the Hamiltonian constraint takes the form

$$H = \boldsymbol{E}^a \cdot \boldsymbol{E}^b \times (\boldsymbol{B}^c + \Lambda \boldsymbol{E}^c)\epsilon_{abc} \qquad [1]$$

where we have used a conventional vector notation for the frame indices and kept explicit the spatial indices. $\epsilon_{abc}$ is the Levi-Civita totally antisymmetric tensor. We have included a possible cosmological constant $\Lambda$. The Ashtekar formulation can be constructed in different ways. In the original formulation, the connection $A_a^i$ was a complex variable and the Hamiltonian took the form we listed above. However, the resulting theory was only equivalent to real general relativity if the variables satisfied certain reality conditions. One can choose to use a real connection instead, but then the Hamiltonian constraint has additional terms. At the moment, we will concentrate on the constraint as listed above. The constraint has to be implemented as a quantum operator acting on wave functions. Since it involves the product of operators, it needs to be regularized. Most regularization methods are problematic in this context, since they use a metric, and here the metric is a quantum operator, not an external fixed quantity. If we ignore these difficulties, one observes that, if one were to choose a quantum state, for instance in the connection representation, for which,

$$\Lambda \hat{E}_i^a \Psi[A] = -\hat{B}_i^a \Psi[A] \qquad [2]$$

the state would be annihilated by the Hamiltonian constraint, and this would be true no matter what regularization was chosen. Classically, the condition $E_i^a \sim B_i^a$ is satisfied for the de Sitter geometry, so one could envision the state as a quantum state associated with such geometry. The exact solution of the above equation is given by a state that is the exponential of the integral on the spatial slice of the Chern–Simons form built from the connection

$$\Psi_{\text{CS}}[A] = \exp\left(k \int \mathrm{d}^3x \,\mathrm{tr}(A \wedge dA + \frac{2}{3} A \wedge A \wedge A)\right) \qquad [3]$$

and the constant $k$ needs to be chosen as $k = 6/\Lambda$ for the state to be a solution.

One can ask, "what is the expression of this state in the loop representation?" To answer this, one needs to compute the coefficients of its expansion in the basis of Wilson loops $W_\gamma[A]$, where as we stated earlier, $\gamma$ should be a collection of (intersecting) loops (later we will discuss the generalization to spin networks). The expression for the coefficients will be a function only of the loops $\gamma$ and is given by

$$\Psi_{\text{CS}}[\gamma] = \int DA\, W_\gamma[A] \Psi_{\text{CS}}[A] \qquad [4]$$

This expression is invariant under diffeomorphisms of the manifold or, equivalently, under smooth deformations of the curve $\gamma$. That is, it is what in the mathematical literature is called "knot invariant." In fact, this integral has been studied by Witten in the context of Chern–Simons theory and has been shown to be related to the Kauffman bracket knot polynomial, which in turn is related to the celebrated Jones polynomial. Therefore, the implication of these results is that the Kauffman bracket knot polynomial appears to be the representation in the loop representation of a state of quantum gravity that solves the quantum Einstein equations (with a cosmological constant). The reader may be intrigued by the word "polynomial" in this context. It should be noted that the Chern–Simons state $\Psi_{\text{CS}}[A]$ depended on a parameter $k$, which had to take a certain value for it to solve the quantum Einstein equations. The resulting knot invariant is a polynomial in $\exp(k)$. If one expands out the result, an infinite power series in $k$ results. There will be infinite coefficients in the series, but they are just combination of the finite number of coefficients of the polynomial. Knot polynomials are a powerful tool for analyzing and distinguishing knots. The coefficients of the polynomials are all knot invariants. Typically, for "simple" knots, the first few coefficients of the knot polynomial are nonzero. As one considers more complicated knottings, higher

coefficients become nonvanishing. The ultimate goal of knot theory is to be able to consider two arbitrary knots and to unambiguously determine if the two knots are related by a smooth transformation. The knot polynomials appear as promising tools for achieving this task that has remained elusive up to now.

Returning to quantum gravity, to have a well-known knot polynomial as a solution of the quantum Einstein equations is a remarkable fact. The first connection we outlined between knot theory and quantum gravity was less unexpected: if one describes a theory that is diffeomorphism invariant in terms of loops, the appearance of knots is inevitable. But we are now finding that knot invariants from the mathematical literature, which were constructed without any knowledge of the details of the dynamics of the Einstein equations, seem to manage to solve such equations. This is either a big coincidence or a pointer to some unexplained deep connection yet to be understood. Notice, for instance, that other theories of gravity would not have the Kauffman bracket as a quantum state.

There is a certain technicality about the Kauffman bracket that makes it difficult to argue with precision that it is a state of quantum gravity. To understand this technicality better, it is perhaps best to concentrate on the form of the quantum state written above if the connection is an abelian connection. In that case, the integral in question,

$$\Psi_{\text{CS abelian}}[\gamma] = \int DA \oint_\gamma \mathrm{d}y^a \exp(\mathrm{i}A_a) \times \exp\left(\int \mathrm{d}^3x \epsilon^{abc} A_a \partial_b A_c\right) \quad [5]$$

by turning it into a Gaussian integral. The result is

$$\Psi_{\text{CS abelian}}[\gamma] = \oint_\gamma \mathrm{d}x^a \oint_\gamma \mathrm{d}y^b \epsilon_{abc} \frac{(x-y)^c}{|x-y|} \quad [6]$$

This integral has problems, since the integrand is ill-defined when $x = y$. Notice that the integral would be well defined if the two contour integrals were evaluated on different, nonintersecting curves. The result would be the well-known formula for the Gauss linking number of the curves, yielding zero if they are not linked and and integer multiple of $4\pi$ if they were. So the integral we were trying to compute was actually the Gauss linking number of the curve with itself. Such a quantity is not well defined for ordinary curves. To deal with this problem, mathematicians introduced the concept of framed knots. A framed knot is a curve with a prescription to determine a second curve from it. One way to see it is to construct another curve that is "infinitesimally close" in space to the original one. It is clear that there is no canonical way to compute such a second curve. Then, when one considers quantities like the self-linking number, one makes them well defined by evaluating the two integrals on the two curves, the original one and the one yielded by the prescription. In reality, the notion of framing is a bit more elaborate than what we hint at here, since one could consider invariants constructed with more than two integrals and could still be ill-defined if one only considers two curves. The notion has to be extended as well to handle intersections in the curves. We will ignore these subtleties in this discussion.

The Kauffman bracket knot invariant is an invariant of framed knots, just like the self-linking number. It is not well defined for a single curve. It requires a framing of the knot. In quantum gravity, there is no compelling reason to consider framed curves. It is true that framed curves arise naturally in $q$-deformed field theories and perhaps a $q$-deformed version of quantum gravity is what needs to be considered to accommodate the Chern–Simons state, but at the moment there are no proposals along these lines that have widespread consensus.

So, it appears the Kauffman bracket does not have a natural role to play as a state of quantum gravity. However, it is known that the frame dependence of the Kauffman bracket knot polynomial can be captured in an overall factor that depends on the self-linking number. If one strips the polynomial of this factor, one gets the Jones polynomial, which is a knot invariant of single curves. Could it be that this polynomial has a chance of being a solution of the quantum Einstein equations?

To determine this, the analogy with Chern–Simons theory is no longer useful, since there is no straightforward way to transform the relation between the Kauffman and Jones polynomials into relations between states in the connection representation. To analyze if the Jones polynomial could be a solution of the quantum Einstein equations, one needs to write the quantum Einstein equations directly in terms of loops.

There have been several attempts to rewrite the quantum Einstein equations directly in the loop representation. In one of these attempts, the curvature that appears in the Hamiltonian constraint was represented by the "loop derivative." This is a differential operator that can be introduced in the space of loops by considering that two loops that differ by a small element of area are "close." One can build an attractive differential calculus in loop space that actually encodes many of the kinematical properties that are useful to formulate Yang–Mills theory.

The Hamiltonian constraint in terms of the loop derivative is an operator that has an explicit form. The coefficients of the Jones polynomial can also be given an explicit form by computing perturbatively the integral in the Chern–Simons theory. The results are generalizations of the types of integrals that arise in the self-linking number, but involving a larger number of integrals. One can therefore envisage carrying out an explicit computation in which one checks if the coefficients of the Jones polynomial are annihilated or not by the Hamiltonian constraint of quantum gravity in the loop representation. Such a calculation has been carried out for the first few coefficients. It turns out that the second coefficient (the first coefficient is normalized to unity, so it trivially satisfies the constraint) is indeed annihilated by the Hamiltonian constraint of vacuum quantum gravity (with zero cosmological constant). It has been shown that the third coefficient is not, and there are good arguments to indicate that other coefficients will not be states of quantum gravity.

So, a remarkable result has been found in that one of the coefficients of the Jones polynomial (related to the Arf and Casson invariants) is annihilated by a version of the quantum Hamiltonian constraint of general relativity. The result is quite nontrivial; it requires a fair amount of calculation to actually show that the coefficient is annihilated. The meaning of this quantum state and the deep reason why it is annihilated remain at present a mystery.

The quantum Hamiltonian constraint based on the loop derivative makes certain assumptions about the space of functions one is using to quantize the theory. In quantum field theory, not all classical operators have a well-defined quantum counterpart. The choice being made is to assume that the curvature $F_{ab}$ is a well-defined quantum operator defined by the loop derivative. Differentiability of knot polynomials is not a new idea. It is the core idea of the Vassiliev knot invariants, which are defined by a set of identities, one of them acting as a "derivative in knot space." It can be shown that the loop derivative is a concrete implementation of the Vassiliev derivative and, therefore, Vassiliev invariants are the "arena" in which this version of quantum gravity takes place.

The Hamiltonian based on the loop derivative has problems, in the sense that it is obtained by a regularization procedure that requires extra external geometric structures. This is common practice in Yang–Mills theory, where one has at hand a fixed external background metric. However, in gravity the geometry is a dynamical object and, if one constructs expressions that resort to some fixed external geometry, one gets inconsistencies. In particular, it is expected that the Hamiltonian based on the loop derivative will not reproduce the correct Poisson algebra of canonical general relativity. This sort of problem plagued early attempts to construct a quantum version of the Hamiltonian constraint in the early 1990s.

A point that we mentioned earlier but did not elaborate upon, is that the Wilson loops constitute an overcomplete basis of states. Therefore, if one takes a quantum state and expands it on such a basis, one gets that the coefficients of the expansion satisfy certain identities, called the Mandelstam identities. These are nonlinear identities that states in the loop representation have to satisfy. These identities are very inconvenient at the time of constructing quantum states. The identities stem from the fact that if one chooses a matrix representation of the group of interest, the fact that one is in a given representation is indicated by certain identities the matrices satisfy. To break free from these constraints, one possibility is to consider multiple representations when constructing Wilson loops. To do this, one considers piecewise-continuous graphs with intersections (the nonintersecting case is a trivial subcase). Along the lines connecting the intersections one considers holonomies in a given representation for a given line. In the case of the group SU(2), which is the one of interest in quantum gravity, such representations are labeled by a (half-) integer. One then considers invariant tensors in the group to "tie the holonomies together" at intersections. The resulting object is a gauge-invariant object for a given connection based on a "spin network." The latter is an embedded piecewise-continuous graph with an assignment of integers to each of its lines and an assignment of "intertwiners" at each intersection (if the intersections are trivalent or lower, one can choose canonical intertwiners and forget about them).

One can then consider the "spin network representation" in which one expands gauge-invariant states in terms of the basis of Wilson nets. Knot polynomials for these types of graphs have been considered in the mathematical literature ("polynomials of colored graphs"). The construction with the Chern–Simons state can be repeated, and there exist suitable generalizations of the Kauffman bracket and Jones polynomials. The Hamiltonian based on the loop derivative can also be introduced in this context; again, its action is well defined on suitable generalizations of Vassiliev invariants for these kinds of graphs. This opens the possibility of encoding the quantum dynamics of general relativity as a combinatorial action in the space of Vassiliev invariants.

An alternative Hamiltonian based on assuming that the holonomies and the volume operators are well defined quantum mechanically (but not the curvature) has been introduced that has the advantage of not

requiring external structures for its regularization. In fact, it can be explicitly checked that it satisfies the correct Poisson algebra without anomalies at the quantum level. The exploration of the action of this Hamiltonian constraint on knot polynomials has not been carried out as systematically as for the one based on the loop derivative, but it has been explicitly shown that the first coefficient in the expansion of the Jones polynomial is annihilated by this Hamiltonian constraint. The first coefficient, written in terms of loops, was simply the numeral 1 and was automatically annihilated. In terms of spin network states, the first coefficient is the "chromatic evaluation" of the network (the result of computing the Wilson loop on a connection that is pure gauge). It is somewhat nontrivial to show that this quantity is actually annihilated by the Hamiltonian constraint in question.

At the moment, the issue of what the correct Hamiltonian constraint is that describes a realistic and physically correct theory of quantum gravity is still open to debate. There are certain concerns that the action of the operators considered up to now is too simple to encompass the true dynamics of general relativity. Constructing a semiclassical theory that could confirm or deny the viability of the proposals is a complicated task, since one has to make contact with physics that is not diffeomorphism invariant in the context of a theory that is. Moreover, in canonical quantum gravity, there exists the "problem of time." Since the Hamiltonian vanishes, the dynamics implied by it is trivial, and one has to disentangle the true dynamics by relational constructions among the variables of the theory. One then needs to compare the resulting predictions with classical general relativity.

Whether the current proposals are viable and whether knot theory will play a role at a "kinematical level" or it will actually play a key role in the detailed dynamics of quantum general relativity is yet to be seen. It is reassuring that in partial constructions, celebrated knot polynomials have appeared to have some knowledge of the dynamics of the Einstein equations.

Quantum gravity being an unfinished symphony, we cannot entirely conclude how great an impact knot theory will have on it in the end. One can only note that beautiful mathematical results seem to tie in naturally with the partial constructions that have been carried out thus far.

*See also:* BF Theories; Braided and Modular Tensor Categories; Finite-Type Invariants; Finite-type invariants of 3-Manifolds; The Jones Polynomial; Knot Theory and Physics; Loop Quantum Gravity; Mathematical Knot Theory; Quantum Dynamics in Loop Quantum Gravity; Quantum Geometry and its Applications; Yang–Baxter Equations.

## Further Reading

Ashtekar A (2005) Gravity and the quantum. *New Journal of Physics* 7: 200.

Gambini R and Pullin J (1996) *Loops, Knots, Gauge Theories and Quantum Gravity*. Cambridge: Cambridge University Press.

Rovelli C (2001) Notes for a brief history of quantum gravity. In: Ruffini R (ed.) *Rome 2000, Recent Developments in Theoretical and Experimental General Relativity, Gravitation, and Relativistic Field Theories*, p. 742. Singapore: World Scientific.

Rovelli C (2004) *Quantum Gravity*. Cambridge: Cambridge University Press.

Smolin L (2001) *Three Roads to Quantum Gravity*. New York: Basic Books.

Thiemann T, *Introduction to Modern Canonical General Relativity*. Cambridge University Press (arXiv.org:gr-qc/0110034) (to appear).

# Knot Theory and Physics

**L H Kauffman**, University of Illinois at Chicago, Chicago, IL, USA



## Introduction

This article is an introduction to some of the relationships between knot theory and theoretical physics. Knots themselves are macroscopic physical phenomena in three-dimensional space, occurring in rope, vines, telephone cords, polymer chains, DNA, certain species of eel, and many other places in the natural and man-made world. The study of topological invariants of knots leads to relationships with statistical mechanics and quantum physics. This is a remarkable and deep situation where the study of a certain (topological) aspects of the macroscopic world is entwined with theories developed for the subtleties of the microscopic world. The present article is an introduction to the mathematical side of these connections, with some hints and references to the related physics.

We begin with a short introduction to knots, links, braids, and the bracket polynomial invariant of knots and links. The article then discusses Vassiliev invariants of knots and links, and how these invariants are naturally related to Lie algebras and to Witten's gauge-theoretic approach. This part

of the article is an introduction to how Vassiliev invariants in knot theory arise naturally in the context of Witten's functional integral.

The article is divided into several sections beyond the introduction. Section two is a quick introduction to the topology of knots and links. The third one discusses Vassiliev invariants and invariants of rigid vertex graphs. The fourth section introduces the basic formalism and shows how Witten's functional integral is related directly to Vassiliev invariants. The fifth section discusses the loop transform and loop quantum gravity in this context. The final section is an introduction to topological quantum field theory, and to the use of these techniques in producing unitary representations of the braid group, a topic of intense interest in quantum information theory.

## Knots, Braids, and Bracket Polynomial

The purpose of this section is to give a quick introduction to the diagrammatic theory of knots, links, and braids. A knot is an embedding of a circle in three-dimensional space, taken up to ambient isotopy. That is, two knots are regarded as equivalent if one embedding can be obtained from the other through a continuous family of embeddings of circles in 3-space. A link is an embedding of a disjoint collection of circles, taken up to ambient isotopy. Figure 1 illustrates a diagram for a knot. The diagram is regarded both as a schematic picture of the knot, and as a plane graph with extra structure at the nodes (indicating how the curve of the knot passes over or under itself by standard pictorial conventions).

Ambient isotopy is mathematically the same as the equivalence relation generated on diagrams by the Reidemeister moves. These moves are illustrated in Figure 2. Each move is performed on a local part of the diagram that is topologically identical to the part of the diagram illustrated in this figure (these figures are representative examples of the types of Reidemeister moves) without changing the rest of the diagram. The Reidemeister moves are useful in

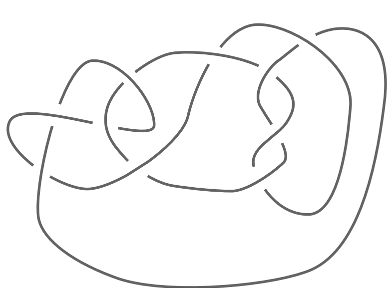


Figure 1 A knot diagram.

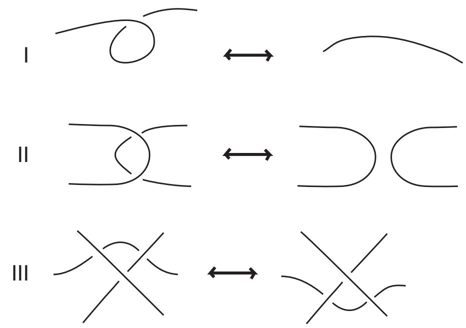



Figure 2 The Reidemeister moves.

doing combinatorial topology with knots and links, notably in working out the behavior of knot invariants. A knot invariant is a function defined from knots and links to some other mathematical object (such as groups or polynomials or numbers) such that equivalent diagrams are mapped to equivalent objects (isomorphic groups, identical polynomials, identical numbers).

Another significant structure related to knots and links is the Artin braid group. A braid is an embedding of a collection of strands that have their ends in two rows of points that are set one above the other with respect to a choice of vertical. The strands are not individually knotted and they are disjoint from one another. See Figures 3–5 for illustrations of braids and moves on braids. Braids can be multiplied by attaching the bottom row of one braid to the top row of the other braid. Taken up to ambient isotopy, fixing the endpoints, the braids form a group under this notion of multiplication. In Figure 3 we illustrate the form of the basic generators of the braid group, and the form of the relations among these generators. Figure 4 illustrates how to close a braid by attaching the top strands to the bottom strands by a collection of parallel arcs. A key theorem of Alexander states that every knot or link can be represented as a closed braid. Thus, the theory of braids is critical to the

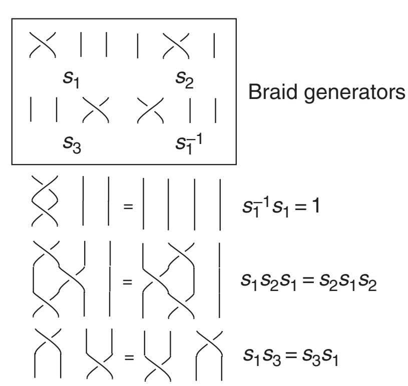



Figure 3 Braid generators.

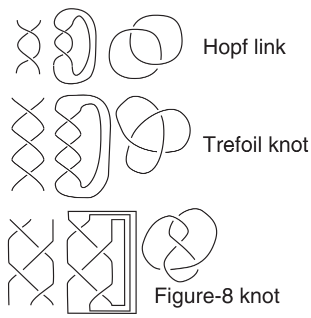



**Figure 4** Closing braids to form knots and links.

theory of knots and links. Figure 5 illustrates the famous Borrowmean rings (a link of three unknotted loops such that any two of the loops are unlinked) as the closure of a braid.

We now discuss a significant example of an invariant of knots and links, the bracket polynomial. The bracket polynomial can be normalized to produce an invariant of all the Reidemeister moves. This normalized invariant is known as the Jones (1985) polynomial. The Jones polynomial was originally discovered by a different method than the one given here.

The bracket polynomial, $\langle K\rangle = \langle K\rangle(A)$, assigns to each unoriented link diagram $K$ a Laurent polynomial in the variable $A$, such that

1. If $K$ and $K'$ are regularly isotopic diagrams, then $\langle K\rangle = \langle K'\rangle$.
2. If $K \amalg O$ denotes the disjoint union of $K$ with an extra unknotted and unlinked component $O$ (also called "loop" or "simple closed curve" or "Jordan curve"), then

   $$\langle K \amalg O\rangle = \delta\langle K\rangle$$

   where

   $$\delta = -A^2 - A^{-2}$$

3. $\langle K\rangle$ satisfies the following formulas:

   $$\langle \chi\rangle = A\langle \asymp\rangle + A^{-1}\langle )(\rangle$$
   $$\langle \overline{\chi}\rangle = A^{-1}\langle \asymp\rangle + A\langle )(\rangle$$

where the small diagrams represent parts of larger diagrams that are identical except at the site indicated in the bracket. We take the convention that the letter chi, $\chi$, denotes a crossing where the curved line is crossing over the straight segment. The barred letter denotes the switch of this crossing, where the curved line is undercrossing the straight segment.

In computing the bracket, one finds the following behavior under Reidemeister move I:

$$\langle \gamma\rangle = -A^3\langle \smile\rangle$$

and

$$\langle \overline{\gamma}\rangle = -A^{-3}\langle \smile\rangle$$

where $\gamma$ denotes a curl of positive type as indicated in Figure 6, and $\overline{\gamma}$ indicates a curl of negative type, as also seen in this figure. The type of a curl is the sign of the crossing when we orient it locally. Our convention of signs is also given in Figure 6. Note that the type of a curl does not depend on the orientation we choose. The small arcs on the right-hand side of these formulas indicate the removal of the curl from the corresponding diagram.

The bracket is invariant under regular isotopy and can be normalized to an invariant of ambient isotopy by the definition

$$f_K(A) = (-A^3)^{-w(K)}\langle K\rangle(A)$$

where we chose an orientation for $K$, and where $w(K)$ is the sum of the crossing signs of the oriented link $K$. $w(K)$ is called the writhe of $K$. The convention for crossing signs is shown in Figure 6.

### The State Summation

In order to obtain a closed formula for the bracket, we now describe it as a state summation. Let $K$ be any unoriented link diagram. Define a state, $S$, of $K$ to be a choice of smoothing for each crossing of $K$. There are two choices for smoothing a given crossing, and thus there are $2^N$ states of a diagram with $N$ crossings. In a state we label each smoothing with $A$ or $A^{-1}$ as in the expansion formula for the bracket. The label is called a vertex weight of the

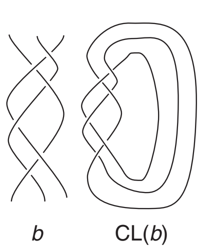



**Figure 5** Borromean rings as a braid closure.

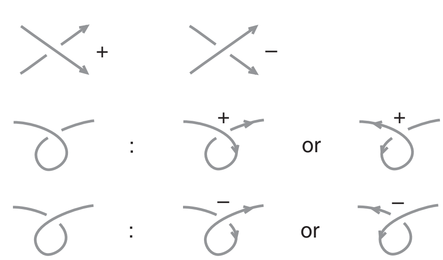



**Figure 6** Crossing signs and curls.

state. There are two evaluations related to a state. The first one is the product of the vertex weights, denoted $\langle K|S\rangle$. The second evaluation is the number of loops in the state $S$, denoted $\|S\|$.

Define the state summation, $\langle K\rangle$, by the formula

$$\langle K\rangle = \sum_S \langle K|S\rangle \delta^{\|S\|-1}$$

It follows from this definition that $\langle K\rangle$ satisfies the equations

$$\langle \chi \rangle = A\langle \asymp \rangle + A^{-1}\langle )( \rangle$$
$$\langle K \amalg O\rangle = \delta\langle K\rangle$$
$$\langle O\rangle = 1$$

The first equation expresses the fact that the entire set of states of a given diagram is the union, with respect to a given crossing, of those states with an $A$-type smoothing and those with an $A^{-1}$-type smoothing at that crossing. The second and the third equation are clear from the formula defining the state summation. Hence, this state summation produces the bracket polynomial as we have described it at the beginning of the section.

**Remark** By a change of variables one obtains the original Jones polynomial, $V_K(t)$, for oriented knots and links from the normalized bracket:

$$V_K(t) = f_K(t^{-1/4})$$

**Remark** The bracket polynomial provides a connection between knot theory and physics, in that the state summation expression for it exhibits it as a generalized partition function defined on the knot diagram. Partition functions are ubiquitous in statistical mechanics, where they express the summation over all states of the physical system of probability weighting functions for the individual states. Such physical partition functions contain large amounts of information about the corresponding physical system. Some of this information is directly present in the properties of the function, such as the location of critical points and phase transition. Some of the information can be obtained by differentiating the partition function, or performing other mathematical operations on it.

In fact, by defining a generalization of the bracket polynomial, defined on knot diagrams but not invariant under the Reidemeister moves, we can capture significant partition functions that are physically meaningful. There is no room in this survey to detail how this generalization can be used to express the Potts model for planar graphical configurations, and how it expresses the relationship between the Potts model and the Temperley–Lieb algebra in diagrammatic form. There is much more in this connection with statistical mechanics in that the local weights in a partition function are often expressed in terms of solutions to a matrix equation called the Yang–Baxter equation, that turns out to fit perfectly invariance under the third Reidemeister move. As a result, there are many ways to define partition functions of knot diagrams that give rise to invariants of knots and links. The subject is intertwined with the algebraic structure of Hopf algebras and quantum groups, useful for producing systematic solutions to the Yang–Baxter equation. In fact, Hopf algebras are deeply connected with the problem of constructing invariants of three-dimensional manifolds in relation to invariants of knots. We have chosen, in this survey article, not to discuss the details of these approaches, but rather to proceed to Vassiliev invariants and the relationships with Witten's functional integral. The reader is referred to Kauffman (1987, 1994, 2002), Jones (1985), and Reshetikhin and Turaev (1991) for more information about relationships of knot theory with statistical mechanics, Hopf algebras, and quantum groups. For topology, the key point is that Lie algebras can be used to construct invariants of knots and links. This is shown nowhere more clearly than in the theory of Vassiliev invariants that we take up in the next section.

## Vassiliev Invariants and Invariants of Rigid Vertex Graphs

In this section we study the combinatorial topology of Vassiliev invariants. As we shall see, by the end of this section, Vassiliev invariants are directly connected with Lie algebras, and representations of Lie algebras can be used to construct them. This aspect of link invariants is one of the most fundamental for connections with physics. Just as symmetry considerations in physics lead to a fundamental relationship with Lie algebras, topological invariance leads to a fundamental relationship of the theory of knots and links with Lie algebras.

If $V(K)$ is a (Laurent polynomial valued or, more generally, commutative ring valued) invariant of knots, then it can be naturally extended to an invariant of rigid vertex graphs by defining the invariant of graphs in terms of the knot invariant via an "unfolding of the vertex." That is, we can regard the vertex as a "black box" and replace it by any tangle of our choice. Rigid vertex motions of the graph preserve the contents of the black box, and hence implicate ambient isotopies of the link obtained by replacing the black box by its contents.

Invariants of knots and links that are evaluated on these replacements are then automatically rigid vertex invariants of the corresponding graphs. If we set up a collection of multiple replacements at the vertices with standard conventions for the insertions of the tangles, then a summation over all possible replacements can lead to a graph invariant with new coefficients corresponding to the different replacements. In this way, each invariant of knots and links implicates a large collection of graph invariants.

The simplest tangle replacements for a 4-valent vertex are the two crossings, positive and negative, and the oriented smoothing. Let $V(K)$ be any invariant of knots and links. Extend $V$ to the category of rigid vertex embeddings of 4-valent graphs by the formula

$$V(K_*) = aV(K_+) + bV(K_-) + cV(K_0)$$

where $K_+$ denotes a knot diagram $K$ with a specific choice of positive crossing, $K_-$ denotes a diagram identical to the first with the positive crossing replaced by a negative crossing and $K_*$ denotes a diagram identical to the first with the positive crossing replaced by a graphical node.

There is a rich class of graph invariants that can be studied in this manner. The Vassiliev invariants (Bar-Natan 1995) constitute the important special case of these graph invariants where $a = +1, b = -1$ and $c = 0$. Thus, $V(G)$ is a Vassiliev invariant if

$$V(K_*) = V(K_+) - V(K_-)$$

Call this formula the exchange identity for the Vassiliev invariant $V$. See **Figure 7**.

$V$ is said to be of finite type $k$ if $V(G) = 0$ whenever $|G| > k$, where $|G|$ denotes the number of (4-valent) nodes in the graph $G$. The notion of finite type is of extraordinary significance in studying these invariants. One reason for this is the following basic lemma.

**Lemma** *If a graph G has exactly k nodes, then the value of a Vassiliev invariant $v_k$ of type k on G, $v_k(G)$, is independent of the embedding of G.*

**Proof** Omitted. □

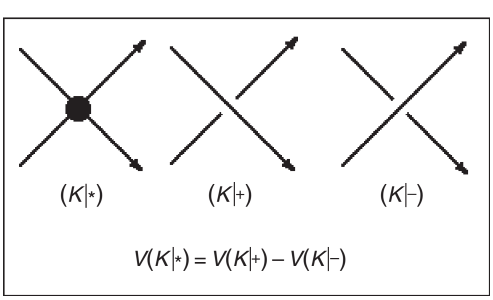



**Figure 7** Exchange identity for Vassiliev invariants.

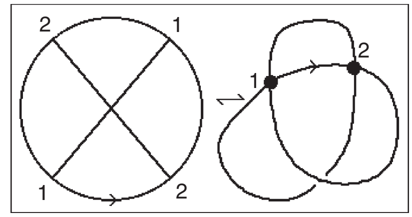



**Figure 8** Chord diagrams.

The upshot of this lemma is that Vassiliev invariants of type $k$ are intimately involved with certain abstract evaluations of graphs with $k$ nodes. In fact, there are restrictions (the four-term relations) on these evaluations demanded by the topology and it follows from results of Kontsevich (see Bar-Natan (1995) that such abstract evaluations actually determine the invariants. The knot invariants derived from classical Lie algebras are all built from Vassiliev invariants of finite type. All of this is directly related to Witten's functional integral (Witten 1989).

In the next few figures we illustrate some of these main points. In **Figure 8** we show how one associates a so-called chord diagram to represent the abstract graph associated with an embedded graph. The chord diagram is a circle with arcs connecting those points on the circle that are welded to form the corresponding graph. In **Figure 9** we illustrate how the four-term relation is a consequence of topological invariance.

In Figure 10 we show how the four-term relation is a consequence of the abstract pattern of the commutator

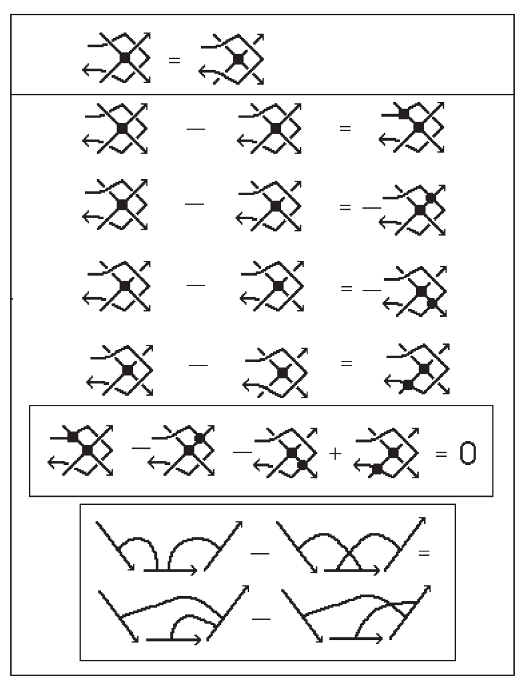


**Figure 9** The four-term relation from topology.

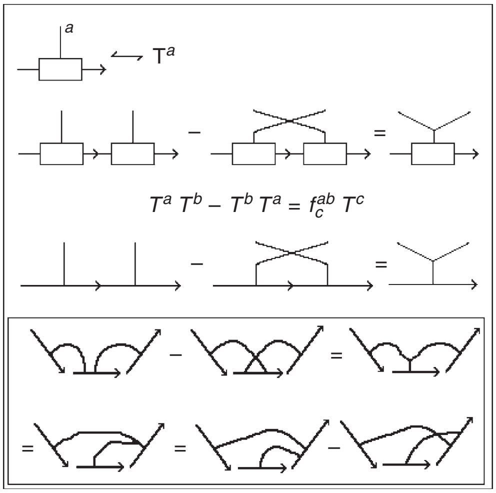



**Figure 10** The four-term relation from categorical Lie algebra.

identity for a matrix Lie algebra. That is, we show how a diagrammatic version of the formula

$$T^a T^b - T^b T^a = f^{ab}_c T^c$$

fits directly with the four-term relation. The formula we have quoted here states that the commutator of the matrices $T^a$ and $T^b$ is equal to a sum of the matrices $T^c$ with coefficients (the structure coefficients of the Lie algebra) $f^{ab}_c$. Such a relation is the most concrete way to define a matrix Lie algebra. There are other levels of abstraction that can be employed here. The same diagrammatic can be interpreted directly in terms of the Jacobi identity that defines a Lie algebra. We shall content ourselves with this matrix point of view here, and add that it is assumed here that the structure coefficients are invariant under cyclic permutation, an assumption that is not needed in the general case. The four-term relation is directly related to a categorical generalization of Lie algebras.

Figure 11 illustrates how the weights are assigned to the chord diagrams in the Lie algebra case – by inserting Lie algebra matrices into the circle and taking a trace of a sum of matrix products. The relationship between Vassiliev invariants and Lie

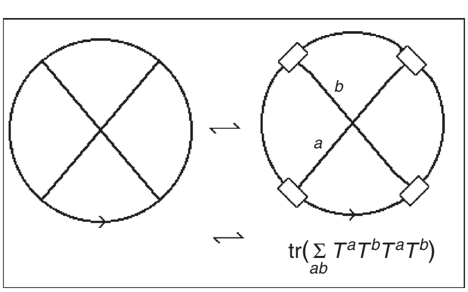



**Figure 11** Calculating Lie algebra weights.

algebras has been known since Bar-Natan's thesis (see also Kauffman (1995). In Bar-Natan (1995) the reader will find a good account of Kontsevich's theorem, showing how Lie algebra weight systems, and in fact any weight system satisfying the four-term relation, can be used to construct knot invariants. Conceptually, the ideas behind the Kontsevich theorem are directly related to Witten's approach to knot invariants via quantum field theory. We give an exposition of this approach in the next section of this article.

**Example** Let $P_K(t) = f_K(\mathrm{e}^t)$ $(A = \mathrm{e}^t)$ where $f_K(A)$ is the normalized bracket polynomial invariant discussed in the last section. Then $P_K(t)$ is expressed as a power series in $t$ with coefficients $v_n(K)$, $n = 0, 1, 2, \ldots$, that are invariants of the knot or link $K$. It is not hard to show that these coefficient invariants (extended to graphs so that the Vassiliev exchange identity is satisfied) are Vassiliev invariants of finite type. In fact, most of the so-called polynomial invariants of knots and links (relatives of the bracket and Jones polynomials) give rise to Vassiliev invariants in just this way. Thus, Vassiliev invariants of finite type are ubiquitous in this area of knot theory. One can think of Vassiliev invariants as building blocks for the other invariants, or that these invariants are sources of Vassiliev invariants.

## Vassiliev Invariants and Witten's Functional Integral

Edward Witten (1989) proposed a formulation of a class of 3-manifold invariants as generalized Feynman integrals taking the form $Z(M)$, where

$$Z(M) = \int DA\mathrm{e}^{(\mathrm{i}k/4\pi)S(M,A)}$$

Here $M$ denotes a 3-manifold without boundary and $A$ is a gauge field (also called a gauge potential or gauge connection) defined on $M$. The gauge field is a 1-form on a trivial $G$-bundle over $M$ with values in a representation of the Lie algebra of $G$. The group $G$ corresponding to this Lie algebra is said to be the gauge group. In this integral, the action $S(M, A)$ is taken to be the integral over $M$ of the trace of the Chern–Simons 3-form $A \wedge \mathrm{d}A + (2/3)A \wedge A \wedge A$. (The product is the wedge product of differential forms.) $Z(M)$ integrates over all gauge fields modulo gauge equivalence.

The formalism and internal logic of Witten's integral supports the existence of a large class of topological invariants of 3-manifolds and associated invariants of knots and links in these manifolds.

The invariants associated with this integral have been given rigorous combinatorial descriptions but questions and conjectures arising from the integral formulation are still outstanding. Specific conjectures about this integral take the form of just how it implicates invariants of links and 3-manifolds, and how these invariants behave in certain limits of the coupling constant $k$ in the integral. Many conjectures of this sort can be verified through the combinatorial models. On the other hand, the really outstanding conjecture about the integral is that it exists! At the present time there is no measure theory or generalization of measure theory that supports it in full generality. Here is a formal structure of great beauty. It is also a structure whose consequences can be verified by a remarkable variety of alternative means.

The formalism of the Witten integral implicates invariants of knots and links corresponding to each classical Lie algebra. In order to see this, we need to introduce the Wilson loop. The Wilson loop is an exponentiated version of integrating the gauge field along a loop $K$ in three space that we take to be an embedding (knot) or a curve with transversal self-intersections. For this discussion, the Wilson loop will be denoted by the notation

$$W_K(A)$$

to denote the dependence on the loop $K$ and the field $A$. It is usually indicated by the symbolism $\mathrm{tr}(Pe^{\oint_K A})$. Thus,

$$W_K(A) = \mathrm{tr}\left(Pe^{\oint_K A}\right)$$

Here the $P$ denotes path ordered integration – we are integrating and exponentiating matrix valued functions, and so must keep track of the order of the operations. The symbol tr denotes the trace of the resulting matrix. This Wilson loop integration exists by normal means and does not require functional integration.

With the help of the Wilson loop functional on knots and links, Witten writes down a functional integral for link invariants in a 3-manifold $M$:

$$Z(M,K) = \int DA e^{(ik/4\pi)S(M,A)} \mathrm{tr}\left(Pe^{\oint_K A}\right) = \int DA e^{(ik/4\pi)S} W_K(A)$$

Here $S(M,A)$ is the Chern–Simons Lagrangian, as in the previous discussion. We abbreviate $S(M,A)$ as $S$ and write $W_K(A)$ for the Wilson loop. Unless otherwise mentioned, the manifold $M$ will be the three-dimensional sphere $S^3$.

An analysis of the formalism of this functional integral reveals quite a bit about its role in knot theory. One can determine how the Witten integral behaves under a small deformation of the loop $K$.

**Theorem**

(i) *Let $Z(K) = Z(S^3, K)$ and let $\delta Z(K)$ denote the change of $Z(K)$ under an infinitesimal change in the loop $K$. Then*

$$\delta Z(K) = (4\pi i/k) \int dA e^{(ik/4\pi)S} [\mathrm{Vol}] T_a T_a W_K(A)$$

*where* $\mathrm{Vol} = \epsilon_{rst} dx^r dx^s dx^t$.

*The sum is taken over repeated indices, and the insertion is taken of the matrices $T_a T_a$ at the chosen point $x$ on the loop $K$ that is regarded as the center of the deformation. The volume element $\mathrm{Vol} = \epsilon_{rst} dx_r dx_s dx_t$ is taken with regard to the infinitesimal directions of the loop deformation from this point on the original loop.*

(ii) *The same formula applies, with a different interpretation, to the case where $x$ is a double point of transversal self-intersection of a loop $K$, and the deformation consists in shifting one of the crossing segments perpendicularly to the plane of intersection so that the self-intersection point disappears. In this case, one $T_a$ is inserted into each of the transversal crossing segments so that $T_a T_a W_K(A)$ denotes a Wilson loop with a self-intersection at $x$ and insertions of $T_a$ at $x + \epsilon_1$ and $x + \epsilon_2$, where $\epsilon_1$ and $\epsilon_2$ denote small displacements along the two arcs of $K$ that intersect at $x$. In this case, the volume form is nonzero, with two directions coming from the plane of movement of one arc, and the perpendicular direction is the direction of the other arc.*

**Remark** One shows that the result of a topological variation has an analytic expression that is zero if the topological variation does not create a local volume. Thus, we have shown that the integral of $e^{(ik/4\pi)S(A)} W_K(A)$ is topologically invariant as long as the curve $K$ is moved by the local equivalent of regular isotopy.

In the case of switching a crossing, the key point is to write the crossing switch as a composition of first moving a segment to obtain a transversal intersection of the diagram with itself, and then to continue the motion to complete the switch. Up to the choice of our conventions for constants, the switching formula is, as shown below (see Figure 12).

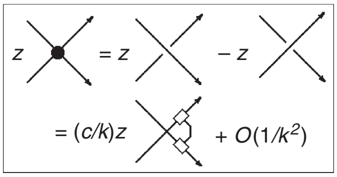



**Figure 12** The difference formula.

$$\begin{aligned}Z(K_+) - Z(K_-) \\ &= (4\pi \mathrm{i}/k) \int DA \mathrm{e}^{(\mathrm{i}k/4\pi)S} T_a T_a \langle K_{**}|A\rangle \\ &= (4\pi \mathrm{i}/k) Z(T^a T^a K_{**})\end{aligned}$$

where $K_{**}$ denotes the result of replacing the crossing by a self-touching crossing. We distinguish this from adding a graphical node at this crossing by using the double-star notation.

A key point is to notice that the Lie algebra insertion for this difference is exactly what is done (in chord diagrams) to make the weight systems for Vassiliev invariants (without the framing compensation). Thus, the formalism of the Witten functional integral takes one directly to these weight systems in the case of the classical Lie algebras. In this way, the functional integral is central to the structure of the Vassiliev invariants.

## The Loop Transform and Quantum Gravity

Suppose that $\psi(A)$ is a (complex-valued) function defined on gauge fields. Then we define formally the loop transform $\widehat{\psi}(K)$, a function on embedded loops in three-dimensional space, by the formula

$$\widehat{\psi}(K) = \int DA \psi(A) W_K(A)$$

If $\Delta$ is a differential operator defined on $\psi(A)$, then we can use this integral transform to shift the effect of $\Delta$ to an operator on loops via integration by parts:

$$\begin{aligned}\widehat{\Delta\psi}(K) &= \int DA \Delta\psi(A) W_K(A) \\ &= -\int DA \psi(A) \Delta W_K(A)\end{aligned}$$

When $\Delta$ is applied to the Wilson loop, the result can be an understandable geometric or topological operation. One can illustrate this situation with operators $G$ and $H$:

$$\begin{aligned}G &= -F^a_{ij} \mathrm{d}x^i \delta/\delta A^a_j(x) \\ H &= -\epsilon_{ars} F^a_{ij} \delta/\delta A^s_i \delta/\delta A^r_j\end{aligned}$$

with summation over the repeated indices. Each of these operators has the property that its action on the Wilson loop has a geometric or topological interpretation. One has

$$\widehat{G\psi}(K) = \delta\widehat{\psi}(K)$$

where this variation refers to the effect of varying $K$. As we saw in the previous section, this means that if $\widehat{\psi}(K)$ is a topological invariant of knots and links, then $\widehat{G\psi}(K) = 0$ for all embedded loops $K$. This condition is a transform analog of the equation $G\psi(A) = 0$. This equation is the differential analog of an invariant of knots and links. It may happen that $\delta\widehat{\psi}(K)$ is not strictly zero, as in the case of our framed knot invariants. For example, with

$$\psi(A) = \exp\left((\mathrm{i}k/4\pi) \int tr(A \wedge dA + (2/3) A \wedge A \wedge A)\right)$$

we conclude that $\widehat{G\psi}(K)$ is zero for flat deformations (in the sense of the previous section) of the loop $K$, but can be nonzero in the presence of a twist or curl. In this sense, the loop transform provides a subtle variation on the strict condition $G\psi(A) = 0$.

In Ashtekar *et al.* (1992) and other publications by Ashtekar, Rovelli, Smolin, and their colleagues, the loop transform is used to study a reformulation and quantization of Einstein gravity. The differential-geometric gravity theory of Einstein is reformulated in terms of a background gauge connection and in the quantization, the Hilbert space consists in functions $\psi(A)$ that are required to satisfy the constraints $G\psi = 0$ and $H\psi = 0$. Thus, we see that $\widehat{G}(K)$ can be partially zero in the sense of producing a framed knot invariant, and that $\widehat{H}(K)$ is zero for non-self-intersecting loops. This means that the loop transforms of $G$ and $H$ can be used to investigate a subtle variation of the original scheme for the quantization of gravity. This program is being actively pursued by a number of researchers. The Vassiliev invariants arising from a topologically invariant loop transform are of significance to this theory.

## Braiding, Topological Quantum Field Theory, and Quantum Computing

The purpose of this section is to discuss in a very general way how braiding is related to topological quantum field theory and to the enterprise (Freedman *et al.* 2002) of using this sort of theory as a model for anyonic quantum computation. The ideas in the subject of topological quantum field theory are well expressed by Michael Atiyah (1990)

and Edward Witten (1989). The simplest case of this idea is C N Yang's original interpretation of the Yang–Baxter equation. Yang articulated a quantum field theory in one dimension of space and one dimension of time, in which the $R$-matrix giving the scattering amplitudes for an interaction of two particles whose (let us say) spins corresponded to the matrix indices so that $R_{ab}^{cd}$ is the amplitude for particles of spin $a$ and spin $b$ to interact and produce particles of spin $c$ and $d$. Since these interactions are between particles in a line, one takes the convention that the particle with spin $a$ is to the left of the particle with spin $b$, and the particle with spin $c$ is to the left of the particle with spin $d$. If one follows the concatenation of such interactions, then there is an underlying permutation that is obtained by following strands from the bottom to the top of the diagram (thinking of time as moving up the page). Yang designed the Yang–Baxter equation for $R$ so that the amplitudes for a composite process depend only on the underlying permutation corresponding to the process and not on the individual sequences of interactions.

In taking over the Yang–Baxter equation for topological purposes, we can use the same interpretation, but think of the diagrams with their under- and over-crossings as modeling events in a spacetime with two dimensions of space and one dimension of time. The extra spatial dimension is taken in displacing the woven strands perpendicular to the page, and allows the use of braiding operators $R$ and $R^{-1}$ as scattering matrices. Taking this picture to heart, one can add other particle properties to the idealized theory. In particular, one can add fusion and creation vertices where, in fusion, two particles interact to become a single particle and, in creation, one particle changes (decays) into two particles. Matrix elements corresponding to trivalent vertices can represent these interactions (see Figure 13).

Once one introduces trivalent vertices for fusion and creation, there is the question how these interactions will behave in respect to the braiding operators. There will be a matrix expression for the compositions of braiding and fusion or creation as indicated in Figure 14. Here we will restrict ourselves to showing the diagrammatics with the intent of giving the reader a flavor of these

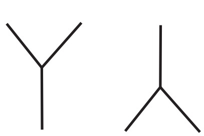


**Figure 13** Creation and fusion.

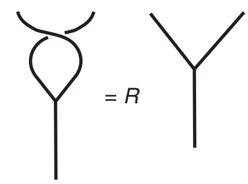



**Figure 14** Braiding.

structures. It is natural to assume that braiding intertwines with creation as shown in Figure 15 (similarly with fusion). This intertwining identity is clearly the sort of thing that a topologist will love, since it indicates that the diagrams can be interpreted as embeddings of graphs in three-dimensional space. Figure 16 illustrates the Yang–Baxter equation. The intertwining identity is an assumption like the Yang–Baxter equation itself, which simplifies the mathematical structure of the model.

It is to be expected that there will be an operator that expresses the recoupling of vertex interactions as shown in Figure 17 and labeled by $Q$. The actual formalism of such an operator will parallel the mathematics of recoupling for angular momentum (see, e.g., Kauffman (1994)). If one just considers the abstract structure of recoupling then one sees that for trees with four branches (each with a single root) there is a cycle of length 5, as shown in

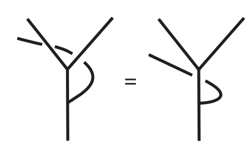



**Figure 15** Intertwining.

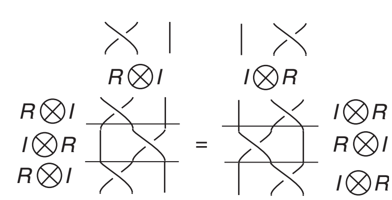



**Figure 16** Yang–Baxter equation.

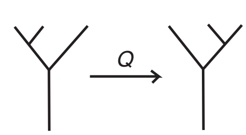



**Figure 17** Recoupling.

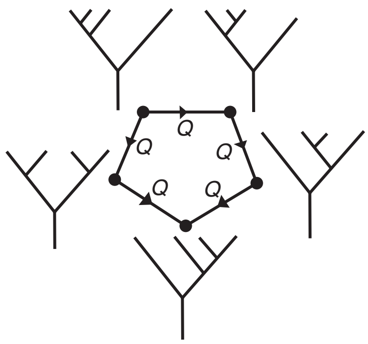



Figure 18 Pentagon identity.

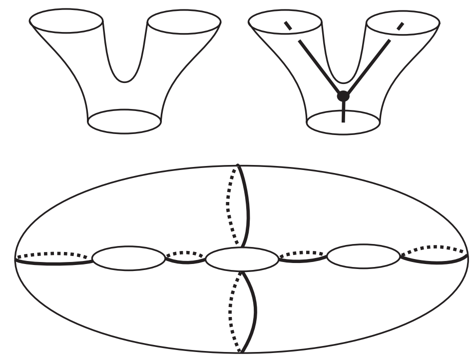


Figure 20 Decomposition of a surface into trinions.

Figure 17. One can start with any pattern of three vertex interactions and go through a sequence of five recouplings that bring one back to the same tree from which one started. It is a natural simplifying axiom to assume that this composition is the identity mapping. This axiom is called the pentagon identity (Figure 18).

Finally, there is a hexagonal cycle of interactions between braiding, recoupling and the intertwining identity as shown in Figure 19. One says that the interactions satisfy the hexagon identity if this composition is the identity.

A three-dimensional topological quantum field theory is an algebra of interactions that satisfies the Yang–Baxter equation, the intertwining identity, the pentagon identity and the hexagon identity. There is no room in this summary to detail the way that these properties fit into the topology of knots and three-dimensional manifolds, but a sketch is in order. For the case of topological quantum field theory related to the group SU(2) there is a construction based entirely on the combinatorial topology of the bracket polynomial (see the section "Knots, braids, and bracket polynomial"). For more information on this approach, the reader is referred to Kauffman (1994, 2002).

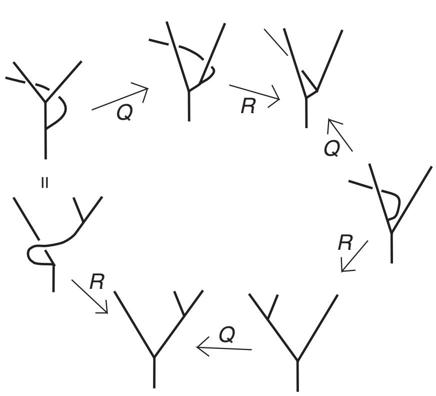



Figure 19 Hexagon identity.

It turns out that the algebraic properties of a topological quantum field theory give it enough power to rigourously model three manifold invariants described by the Witten integral. This is done by regarding the 3-manifold as a union of two handlebodies with boundary an orientable surface $S_g$ of genus $g$. The surface is divided up into trinions as illustrated in Figure 20. A trinion is a surface with boundary that is topologically equivalent to a sphere with three punctures. In Figure 20 we illustrate two trinions, the second shown as a neighborhood of a trivalent vertex, and a surface of genus 3 that is decomposed into three trinions. It turns out that there is a way to associate a vector space $V(S_g)$ to a surface with a trinion decomposition, defined in terms of the associated topological quantum field theory, such that the isomorphism class of the vector space $V(S_g)$ does not depend upon the choice of decomposition. This independence is guaranteed by the braiding, hexagon, and pentagon identities in such a way that one can associate a well-defined vector $|M_\epsilon\rangle$ in $V(S_g)$ whenever $M$ is a 3-manifold whose boundary is $S_g$. Furthermore, if a closed 3-manifold $M^3$ is decomposed along a surface $S_g$ into the union of $M_-$ and $M_+$, where these parts are otherwise disjoint 3-manifolds with boundary $S_g$, then the inner product $I(M) = \langle M_-|M_+\rangle$ is, up to normalization, an invariant of the 3-manifold $M_3$. With the definition of topological quantum field theory given above, knots and links can be incorporated as well, so that one obtains a source of invariants $I(M^3, K)$ of knots and links in orientable 3-manifolds.

The invariant $I(M^3, K)$ can be formally compared with the Witten integral

$$Z(M^3, K) = \int DA e^{(ik/4\pi)S(M,A)} W_K(A)$$

It can be shown that up to limits of the heuristics, $Z(M, K)$ and $I(M^3, K)$ are essentially equivalent for appropriate choice of gauge groups.

This point of view leads to more abstract formulations of topological quantum field theories as ways to associate vector spaces and linear transformations to manifolds and cobordisms of manifolds. (A cobordism of surfaces is a 3-manifold whose boundary consists of these surfaces.)

As the reader can see, a three-dimensional TQFT is, at base, a highly simplified theory of point-particle interactions in $(2+1)$-dimensional spacetime. It can be used to articulate invariants of knots and links and invariants of 3-manifolds. The reader interested in the SU(2) case of this structure and its implications for invariants of knots and 3-manifolds can consult Kauffman (1994, 2002) and Crane (1991). One expects that physical situations involving $2+1$ spacetime will be approximated by such an idealized theory. It is thought, for example, that aspects of the quantum Hall effect will be related to topological quantum field theory (Wilczek 1990). One can imagine a physics where the geometrical space is two dimensional and the braiding of particles corresponds to their interactions through circulating around one another in the plane. Anyons are particles that do not just change their wave functions by a sign under interchange, but rather by a complex phase or even a linear combination of states. It is hoped that TQFT models will describe applicable physics. One can think about the possible applications of anyons to quantum computing. The TQFTs then provide a class of anyonic models where the braiding is essential to the physics and to the quantum computation.

The key point in the application and relationship of TQFT and quantum information theory is, in our opinion, contained in the structure illustrated in Figure 21. There we show a more complex braiding operator, based on the composition of recoupling with the elementary braiding at a vertex. (This structure is implicit in the hexagon identity of Figure 19.) The new braiding operator is a source of unitary representations of braid group in situations (which exist mathematically) where the recoupling transformations are themselves unitary. This kind of pattern is utilized in the work of Freedman *et al.* (2002) and in the case of classical angular momentum formalism has been dubbed a "spin-network quantum simulator" by Rasetti and collaborators (see, e.g., Marzuoli and Rasetti (2002). Kauffman and Lomonaco (2006) show how certain natural deformations (Kauffman 1994) of Penrose (1969) spin networks can be used to produce such the Freedman–Kitaev model for anyonic topological quantum computation. It is legitimate to speculate that networks of this kind are present in physical reality.

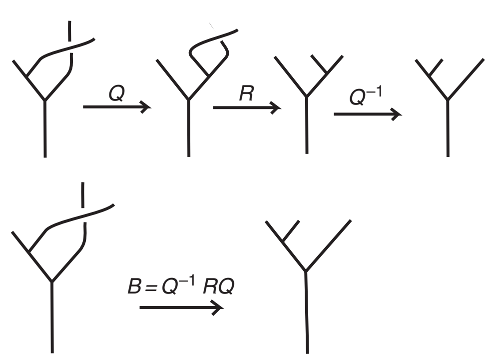



**Figure 21** A more complex braiding operator.

Quantum computing can be regarded as a study of the structure of the preparation, evolution, and measurement of quantum systems. In the quantum computation model, an evolution is a composition of unitary transformations (usually finite-dimensional over the complex numbers). The unitary transformations are applied to an initial state vector that has been prepared prior to this process. Measurements are projections to elements of an orthonormal basis of the space upon which the evolution is applied. The result of measuring a state $|\psi\rangle$, written in the given basis, is probabilistic. The probability of obtaining a given basis element from the measurement is equal to the absolute square of the coefficient of that basis element in the state being measured.

It is remarkable that the above lines constitute an essential summary of quantum theory. All applications of quantum theory involve filling in details of unitary evolutions and specifics of preparations and measurements. Such unitary evolutions can be seen as approximated arbitrarily closely by representations of the Artin braid group. The key to the anyonic models of quantum computation via topological quantum field theory, or via deformed spin networks, is that all unitary evolutions can be approximated by a single coherent method for producing representations of the braid group. This beautiful mathematical fact points to a deep role for topology in the structure of quantum physics.

The future of knots, links, and braids in relation to physics will be very exciting. There is no question that unitary representations of the braid group and quantum invariants of knots and links play a fundamental role in the mathematical structure of quantum mechanics, and we hope that time will show us the full meaning of this relationship.

## Acknowledgments

It gives the author pleasure to thank the National Science Foundation for support of this research

under NSF Grant DMS-0245588. Much of this effort was sponsored by the Defense Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory, Air Force Materiel Command, USAF, under agreement F30602-01-2-05022. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Defense Advanced Research Projects Agency, the Air Force Research Laboratory, or the US Government.

*See also:* Chern–Simons Models: Rigorous Results; Functional Integration in Quantum Physics; Knot Homologies; Knot Invariants and Quantum Gravity; Large-*N* and Topological Strings; Loop Quantum Gravity; Mathematical Knot Theory; Schwarz-Type Topological Quantum Field Theory; The Jones Polynomial; Topological Knot Theory and Macroscopic Physics; Topological Quantum Field Theory: Overview; Two-Dimensional Conformal Field Theory and Vertex Operator Algebras; von Neumann Algebras: Introduction, Modular Theory, and Classification Theory; Yang–Baxter Equations.

## Further Reading

Alexander JW (1923) Topological invariants of knots and links. *Transactions of the American Mathematical Society* 20: 275–306.
Atiyah MF (1990) *The Geometry and Physics of Knots*. Cambridge: Cambridge University Press.
Ashtekar A, Rovelli C, and Smolin L (1992) Weaving a classical geometry with quantum threads. *Physical Review Letters* 69: 237.
Bar-Natan D (1995) On the Vassiliev knot invariants. *Topology* 34: 423–472.
Crane L (1991) 2-d physics and 3-d topology. *Communications Mathematical Physics* 135(3): 615–640.
Freedman MH, Kitaev A, and Wang Z (2002) Simulation of topological field theories by quantum computers. *Communications in Mathematical Physics* 227: 587–603 (quant-ph/0001071).
Jones VFR (1985) A polynomial invariant for links via von Neumann algebras. *Bulletin of the American Mathematical Society* 129: 103–112.
Kauffman LH (1987) State models and the Jones polynomial. *Topology* 26: 395–407.
Kauffman LH (1994) *Temperley–Lieb Recoupling Theory and Invariants of Three-Manifolds*, Annals Studies, vol. 114. Princeton: Princeton University Press.
Kauffman LH (1991) *Knots and Physics* (2nd edn., 1993, 3rd edn., 2002). Singapore: World Scientific Publishers.
Kauffman LH (1995) Functional integration and the theory of knots. *Journal of Mathematical Physics* 36(5): 2402–2429.
Kauffman LH (2002) Quantum computation and the Jones polynomial. In: Lomonaco S, Jr. (ed.) *Quantum Computation and Information*, AMS CONM/305. Providence, RI: American Mathematical Society pp. 101–137.
Kauffman LH and Lomonaco SJ, Jr. (2002) *Quantum entanglement and topological entanglement*. *New Journal of Physics* 4: 73.1–73.18 (http://www.njp.org).
Kauffman LH and Lomonaco SJ, Jr. (2006) Spin networks and anyonic topological quantum computing (in preparation).
Marzuoli A and Rasetti M (2002) Spin network quantum simulator. *Physics Letters A* 306: 79–87.
Penrose R (1969) Angular momentum: an approach to combinatorial spacetime. In: Bastin T (ed.) *Quantum Theory and Beyond*. Cambridge: Cambridge University Press.
Reshetikhin NY and Turaev V (1991) Invariants of three manifolds via link polynomials and quantum groups. *Inventiones Mathematicae* 103: 547–597.
Wilczek F (1990) *Fractional Statistics and Anyon Superconductivity*. Berlin–Heidelberg–New York: Springer-Verlag.
Witten E (1989) Quantum field Theory and the Jones Polynomial. *Communications in Mathematical Physics* 121: 351–399.

# Kontsevich Integral

**S Chmutov and S Duzhin**, Petersburg Department of Steklov Institute of Mathematics, St. Petersburg, Russia



## Introduction

The Kontsevich integral was invented by Kontsevich (1993) as a tool to prove the fundamental theorem of the theory of finite-type (Vassiliev) invariants (see Bar-Natan (1995a)). It provides an invariant exactly as strong as the totality of all Vassiliev knot invariants.

The Kontsevich integral is defined for oriented tangles (either framed or unframed) in $\mathbb{R}^3$; therefore, it is also defined in the particular cases of knots, links, and braids (see Figure 1).

As a starter, we give two examples where simple versions of the Kontsevich integral have a straightforward geometrical meaning. In these examples, as well as in the general construction of the Kontsevich integral, we represent 3-space $\mathbb{R}^3$ as the product of a real line $\mathbb{R}$ with coordinate $t$ and a complex plane $\mathbb{C}$ with complex coordinate $z$.

**Example 1** The number of twists in a braid with two strings $z_1(t)$ and $z_2(t)$ placed in the slice $0 \leq t \leq 1$ (see Figure 2) is equal to

$$\frac{1}{2\pi i}\int_0^1 \frac{dz_1 - dz_2}{z_1 - z_2}$$

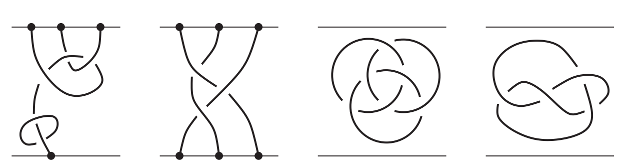


**Figure 1** A tangle, a braid, a link, and a knot.

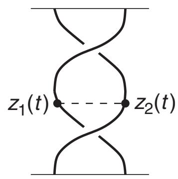



**Figure 2** Counting the number of twists.

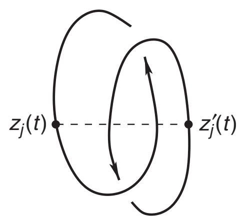



**Figure 3** Counting the linking number.

**Example 2** The linking number of two spatial curves $K$ and $K'$ (see Figure 3) can be computed as

$$\mathrm{lk}(K, K') = \frac{1}{2\pi \mathrm{i}} \int_{m<t<M} \sum_j \varepsilon_j \frac{\mathrm{d}(z_j(t) - z'_j(t))}{z_j(t) - z'_j(t)}$$

where $m$ and $M$ are the minimum and the maximum values of $t$ on the link $K \cup K'$, $j$ is the index that enumerates all possible choices of a pair of strands of the link as functions $z_j(t)$, $z'_j(t)$ corresponding to $K$ and $K'$, respectively, and $\varepsilon_j = \pm 1$ according to the parity of the number of chosen strands that are oriented downwards.

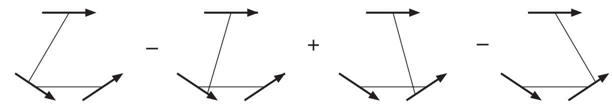
The Kontsevich integral can be regarded as a far-going generalization of these formulas. It aims at encoding all information about how the horizontal chords on the knot (or tangle) rotate when moved in the vertical direction. From a more general viewpoint, the Kontsevich integral represents the monodromy of the Knizhnik–Zamolodchikov connection in the complement to the union of diagonals in $\mathbb{C}^n$ (see Bar-Natan (1995a) and Ohtsuki (2002)).

## Chord Diagrams and Weight Systems

### Algebras $\mathcal{A}(p)$

The Kontsevich integral of a tangle $T$ takes values in the space of chord diagrams supported on $T$.

Let $X$ be an oriented one-dimensional manifold, that is, a collection of $p$ numbered oriented lines and $q$ numbered oriented circles. A chord diagram of order $n$ supported on $X$ is a collection of $n$ pairs of unordered points in $X$, considered up to an orientation- and component-preserving diffeomorphism. In the vector space formally generated by all chord diagrams of order $n$, we distinguish the subspace spanned by all four-term relations

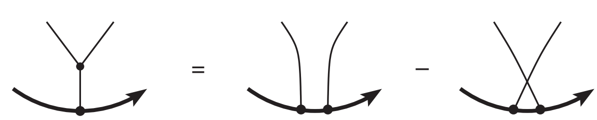



where thin lines designate chords, while thick lines are pieces of the manifold $X$. Apart from the fragments shown, all the four diagrams are identical. The quotient space over all such combinations is denoted by $\mathcal{A}_n(X) = \mathcal{A}_n(p, q)$. Let $\mathcal{A}(p, q) = \oplus_{n=0}^{\infty} \mathcal{A}_n(p, q)$ and let $\hat{\mathcal{A}}(p, q)$ be the graded completion of $\mathcal{A}(p, q)$ (i.e., the space of formal infinite series $\sum_{i=0}^{\infty} a_i$ with $a_i \in \mathcal{A}_i(p, q)$). If, moreover, we divide $\mathcal{A}(p, q)$ by all "framing independence" relations (any diagram with an isolated chord, i.e., a chord joining two adjacent points of the same connected component of $X$, is set to 0), then the resulting space is denoted by $\mathcal{A}'(p, q)$, and its graded completion by $\hat{\mathcal{A}}'(p, q)$.

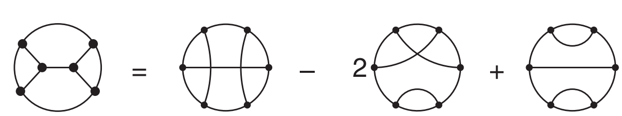
The spaces $\mathcal{A}(p, 0) = \mathcal{A}(p)$ have the structure of an algebra (the product of chord diagrams is defined by concatenation of underlying manifolds in agreement with the orientation). Closing a line component into a circle, we get a linear map $\mathcal{A}(p, q) \to \mathcal{A}(p-1, q+1)$ which is an isomorphism when $p = 1$. In particular, $\mathcal{A}(S^1) \cong \mathcal{A}(\mathbb{R}^1)$ has the structure of an algebra; this algebra is denoted simply by $\mathcal{A}$; the Kontsevich integral of knots takes its values in its graded completion $\hat{\mathcal{A}}$. Another algebra of special importance is $\hat{\mathcal{A}}(3) = \hat{\mathcal{A}}(3, 0)$, because it is where the Drinfeld associators live.

### Hopf Algebra Structure

The algebra $\mathcal{A}(p)$ has a natural structure of a Hopf algebra with the coproduct $\delta$ defined by all ways to split the set of chords into two disjoint parts. To give a convenient description of its primitive space, one can use generalized chord diagrams. We now allow trivalent vertices not belonging to the supporting manifold and use STU relations (Bar-Natan 1995a)




to express the generalized diagrams as linear combinations of conventional chord diagrams, for example,



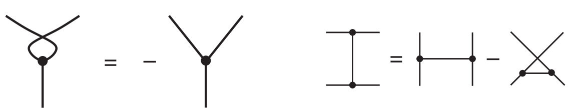

Then the primitive space coincides with the subspace of $\mathcal{A}(p)$ spanned by all connected generalized chord diagrams ("connected" means that they remain connected when the supporting manifold $X$ is disregarded).

### Weight Systems

A "weight system" of degree $n$ is a linear function on the space $\mathcal{A}_n$. Every Vassiliev invariant $v$ of degree $n$ defines a weight system symb($v$) of the same degree called its "symbol."

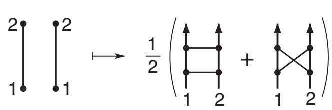
### Algebras $\mathcal{B}(p)$

Apart from the spaces of chord diagrams modulo four-term relations, there are closely related spaces of Jacobi diagrams. A Jacobi diagram is defined as a unitrivalent graph, possibly disconnected, having at least one vertex of valency 1 in each connected component and supplied with two additional structures: a cyclic order of edges in each trivalent vertex and a labeling of univalent vertices taking values in the set $\{1, 2, \ldots, p\}$. The space $\mathcal{B}(p)$ is defined as the quotient of the vector space formally generated by all $p$-colored Jacobi diagrams modulo the two types of relations:
Antisymmetry: IHX:




The disjoint union of Jacobi diagrams makes the space $\mathcal{B}(p)$ into an algebra.

The symmetrization map $\chi_p: \mathcal{B}(p) \to \mathcal{A}(p)$, defined as the average over all ways to attach the legs of color $i$ to $i$th connected component of the underlying manifold




is an isomorphism of vector spaces (the formal PBW isomorphism (Bar-Natan 1995a, Le and Murakami 1995) which is not compatible with the multiplication. The relation between $A(p)$ and $\mathcal{B}(p)$ very much resembles the relation between the universal enveloping algebra and the symmetric algebra of a Lie algebra. The algebra $\mathcal{B} = \mathcal{B}(1)$ is used to write out the explicit formula for the Kontsevich integral of the unknot (see Bar-Natan *et al.* (2003) and below).

## The Construction

### Kontsevich's Formula

We will explain the construction of the Kontsevich integral in the classical case of (closed) oriented knots; for an arbitrary tangle $T$, the formula is the same; only the result is interpreted as an element of $\hat{\mathcal{A}}(T)$. As above, represent three-dimensional space $\mathbb{R}^3$ as a direct product of a complex line $\mathbb{C}$ with coordinate $z$ and a real line $\mathbb{R}$ with coordinate $t$. The integral is defined for Morse knots, that is, knots $K$ embedded in $\mathbb{R}^3 = \mathbb{C}_z \times \mathbb{R}_t$ in such a way that the coordinate $t$ restricted to $K$ has only nondegenerate (quadratic) critical points. (In fact, this condition can be weakened, but the class of Morse knots is broad enough and convenient to work with.)

The Kontsevich integral $Z(K)$ of the knot $K$ is the following element of the completed algebra $\hat{\mathcal{A}}'$:

$$Z(K) = \sum_{m=0}^{\infty} \frac{1}{(2\pi \mathrm{i})^m} \times \int_{\substack{t_{\min} < t_m < \cdots < t_1 < t_{\max} \\ t_j \text{ are noncritical}}} \sum_{P=\{(z_j, z'_j)\}} (-1)^{\downarrow_P} \times D_P \bigwedge_{j=1}^{m} \frac{\mathrm{d}z_j - \mathrm{d}z'_j}{z_j - z'_j}$$

### Explanation of the Constituents

The real numbers $t_{\min}$ and $t_{\max}$ are the minimum and the maximum of the function $t$ on $K$.

The integration domain is the $m$-dimensional simplex $t_{\min} < t_m < \cdots < t_1 < t_{\max}$ divided by the critical values into a certain number of "connected components." For example, Figure 4 shows an embedding of the unknot where, for $m = 2$, the integration domain has six connected components.

The number of summands in the integrand is constant in each connected component of the integration domain, but can be different for different components. In each plane $\{t = t_j\} \subset \mathbb{R}^3$ choose an unordered pair of distinct points $(z_j, t_j)$ and $(z'_j, t_j)$ on $K$, so that $z_j(t_j)$ and $z'_j(t_j)$ are continuous branches of the knot. We denote by $P = \{(z_j, z'_j)\}$ the collection of such pairs for $j = 1, \ldots, m$. The integrand is the sum over all choices of the pairing $P$. In the example above for the component $\{t_{\min} < t_1 < c_1,\ c_2 < t_2 < t_{\max}\}$, we have only one possible pair of points on the levels $\{t = t_1\}$ and $\{t = t_2\}$. Therefore, the sum over $P$ for this component consists of only one summand. Unlike this, in the component $\{t_{\min} < t_1 < c_1,\ c_1 < t_2 < c_2\}$, we still have only one

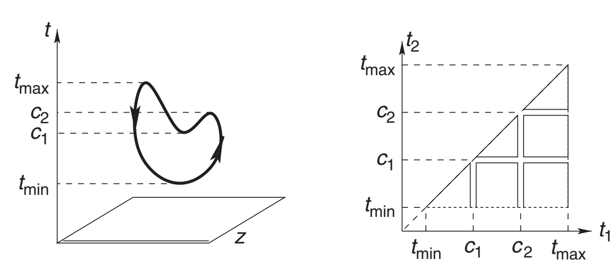



**Figure 4** Connected components.

possibility for the level $\{t = t_1\}$, but the plane $\{t = t_2\}$ intersects our knot $K$ in four points. So we have $\binom{4}{2} = 6$ possible pairs $(z_2, z_2')$, and the total number of summands is six (see Figure 5).

For a pairing $P$, the symbol "$\downarrow_P$" denotes the number of points $(z_j, t_j)$ or $(z_j', t_j)$ in $P$, where the coordinate $t$ decreases along the orientation of $K$.

Fix a pairing $P$. Consider the knot $K$ as an oriented circle and connect the points $(z_j, t_j)$ and $(z_j', t_j)$ by a chord. Up to a diffeomorphism, this chord does not depend on the value of $t_j$ within a connected component. We obtain a chord diagram with $m$ chords. The corresponding element of the algebra $\mathcal{A}'$ is denoted by $D_P$. Figure 5, for each connected component in our example, shows one of the possible pairings, the corresponding chord diagram with the sign $(-1)^{\downarrow}_P$ and the number of summands of the integrand (some of which are equal to zero in $\mathcal{A}'$ due to the framing independence relation).

Over each connected component, $z_j$ and $z_j'$ are smooth functions of $t_j$.

By

$$\bigwedge_{j=1}^{m} \frac{dz_j - dz_j'}{z_j - z_j'}$$

we mean the pullback of this form to the integration domain of variables $t_1, \ldots, t_m$. The integration domain is considered with the orientation of the space $\mathbb{R}^m$ defined by the natural order of the coordinates $t_1, \ldots, t_m$.

By convention, the term in the Kontsevich integral corresponding to $m = 0$ is the (only) chord diagram of order 0 with coefficient 1. It represents the unit of the algebra $\mathcal{A}'$.

#### Framed Version of the Kontsevich Integral

Let $K$ be a framed oriented Morse knot with writhe number $w(K)$. Denote the corresponding knot without framing by $\bar{K}$. The framed version of the Kontsevich integral can be defined by the formula

$$Z^{\mathrm{fr}}(K) = \mathrm{e}^{(w(K)/2)\Theta} \cdot Z(\bar{K}) \in \hat{\mathcal{A}}$$

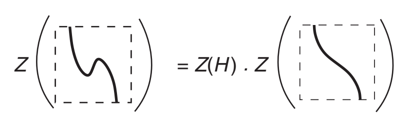
where $\Theta$ is the chord diagram with one chord and the integral $Z(\bar{K}) \in \hat{\mathcal{A}}'$ is understood as an element of the completed algebra $\hat{\mathcal{A}}$ (without one-term relations) by virtue of a natural inclusion $\mathcal{A}' \to \mathcal{A}$ defined as identity on the primitive subspace of $\mathcal{A}'$ (see Goryunov (1999) and Le and Murakami (1996)).

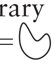
## Basic Properties

#### Constructing the Universal Vassiliev Invariant

The Kontsevich integral $Z(K)$

1. converges for any Morse knot $K$,
2. is invariant under deformations of the knot in the class of Morse knots, and
3. behaves in a predictable way under the deformation that adds a pair of new critical points to a Morse knot:




Here the first and the third pictures depict two embeddings of an arbitrary knot, differing only in the shown fragment, $H =$  is the "hump" (unknot embedded in $\mathbb{R}^3$ in the specified way), and the product is the product in the completed algebra $\hat{\mathcal{A}}'$

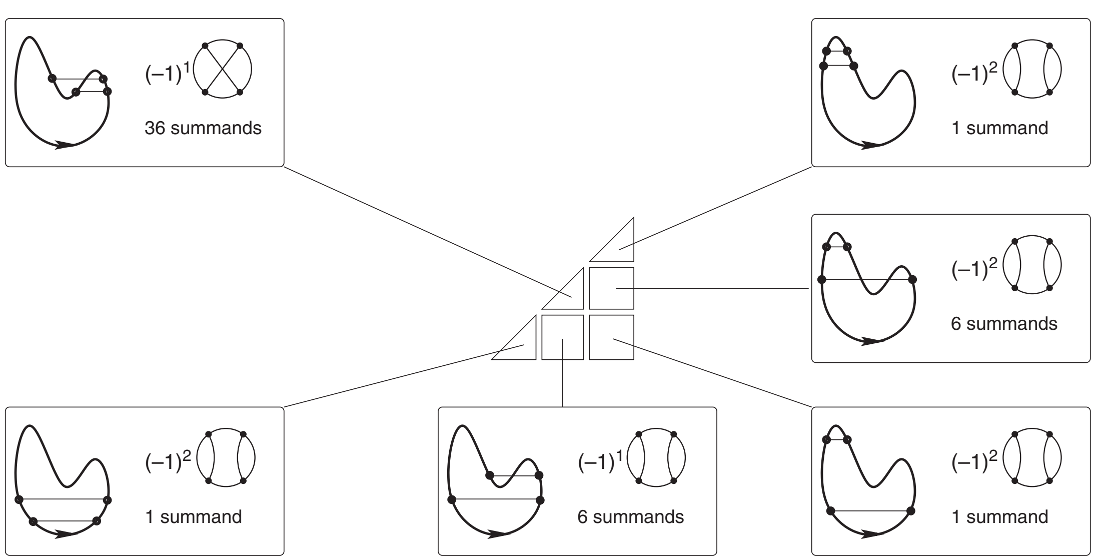



**Figure 5** Pairings and chord diagrams.

of chord diagrams. The last equality allows one to define a genuine knot invariant by the formula

$$I(K) = Z(K)/Z(H)^{c/2}$$

where $c$ denotes the number of critical points of $K$ and the ratio means the division in the algebra $\hat{\mathcal{A}}'$ according to the rule $(1+a)^{-1} = 1 - a + a^2 - a^3 + \cdots$ .

The expression $I(K)$ is sometimes referred to as the "final" Kontsevich integral as opposed to the "preliminary" Kontsevich integral $Z(K)$. It represents a universal Vassiliev invariant in the following sense: *Let $w$ be a weight system, that is, a linear functional on the algebra $\hat{\mathcal{A}}'$. Then the composition $w(I(K))$ is a numerical Vassiliev invariant, and any Vassiliev invariant can be obtained in this way.*

The final Kontsevich integral for framed knots is defined in the same way, using the hump $H$ with zero writhe number.

#### Is Universal Vassiliev Invariant Universal?

At present, it is not known whether the Kontsevich integral separates knots, or even if it can tell the orientation of a knot. However, the corresponding problem is solved, in the affirmative, in the case of braids and string links (theorem of Kohno–Bar-Natan (Bar-Natan 1995b, Kohno 1987).

#### Omitting Long Chords

We will state a technical lemma which is highly important in the study of the Kontsevich integral. It is used in the proof of the multiplicativity, in the combinatorial construction, etc.

Suppose we have a Morse knot $K$ with a distinguished tangle $T$ (**Figure 6**). Let $m$ and $M$ be the maximal and minimal values of $t$ on the tangle $T$. In the horizontal planes between the levels $m$ and $M$, we can distinguish two kinds of chords: "short" chords that lie either inside $T$ or inside $K \setminus T$, and "long" chords that connect a point in $T$ with a point in $K \setminus T$. Denote by $Z_T(K)$ the expression defined by the same formula as the Kontsevich integral $Z(K)$ where only short chords are taken into consideration. More exactly, if $C$ is a connected component of the integration domain whose projection on the coordinate axis $t_j$ is entirely contained in the segment $[m, M]$, then in the sum over the pairings $P$ we include only those pairings that include short chords.

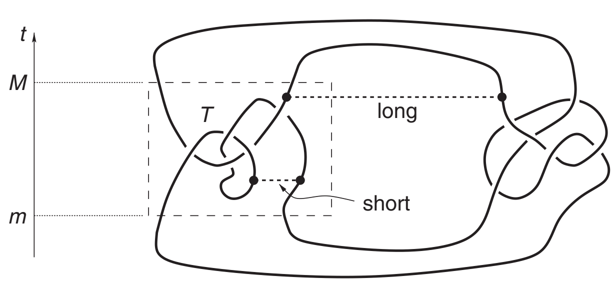



**Figure 6** Short and long chords.

**Lemma** "Long" chords can be omitted when computing the Kontsevich integral: $Z_T(K) = Z(K)$.

#### Kontsevich's Integral and Operations on Knots

The Kontsevich integral behaves in a nice way with respect to the natural operations on knots, such as mirror reflection, changing the orientation of the knot, mutation of knots (see Chmutov and Duzhin (2001)), cabling (see Willerton (2002)). We give some details regarding the first two items.

**Fact 1** Let $R$ be the operation that sends a knot to its mirror image. Define the corresponding operation $R$ on chord diagrams as multiplication by $(-1)^n$, where $n$ is the order of the diagram. Then the Kontsevich integral commutes with the operation $R$: $Z(R(K)) = R(Z(K))$, where by $R(Z(K))$ we mean simultaneous application of $R$ to all the chord diagrams participating in $Z(K)$.

**Corollary** *The Kontsevich integral $Z(K)$ and the universal Vassiliev invariant $I(K)$ of an amphicheiral knot $K$ consist only of even order terms. (A knot $K$ is called "amphicheiral," if it is equivalent to its mirror image: $K = R(K)$.)*

**Fact 2** Let $S$ be the operation on knots which inverts their orientation. The same letter will also denote the analogous operation on chord diagrams (inverting the orientation of the outer circle or, which is the same thing, axial symmetry of the diagram). Then the Kontsevich integral commutes with the operation $S$ of inverting the orientation: $Z(S(K)) = S(Z(K))$.

**Corollary** *The following two assertions are equivalent:*

(i) *Vassiliev invariants do not distinguish the orientation of knots and*
(ii) *all chord diagrams are symmetric: $D = S(D)$ modulo four-term relations.*

The calculations of Kneissler (1997) show that up to order 12 all chord diagrams are symmetric. For bigger orders, the problem is still open.

#### Multiplicative Properties

The Kontsevich integral for tangles is multiplicative:

$$Z(T_1) \cdot Z(T_2) = Z(T_1 \cdot T_2)$$

whenever the product $T_1 \cdot T_2$, defined by vertical concatenation of tangles, exists. Here, the product

on the left-hand side is understood as the image of the element $Z(T_1) \otimes Z(T_2)$ under the natural map $\mathcal{A}(T_1) \otimes \mathcal{A}(T_2) \to \mathcal{A}(T_1 \cdot T_2)$.

This simple fact has two important corollaries:

1. For any knot $K$, the Kontsevich integral $Z(K)$ is a group-like element of the Hopf algebra $\hat{\mathcal{A}}'$, that is,

$$\delta(Z(K)) = Z(K) \otimes Z(K)$$

where $\delta$ is the comultiplication in $\mathcal{A}$ defined above.

2. The final Kontsevich integral, taken in a different normalization

$$I'(K) = Z(H)I(K) = \frac{Z(K)}{Z(H)^{c/2-1}}$$

is multiplicative with respect to the connected sum of knots:

$$I'(K_1 \# K_2) = I'(K_1)I'(K_2)$$

### Arithmetical Properties

For any knot $K$ the coefficients in the expansion of $Z(K)$ over an arbitrary basis consisting of chord diagrams are rational (see Kontsevich (1993), Le and Murakami (1996), and below).

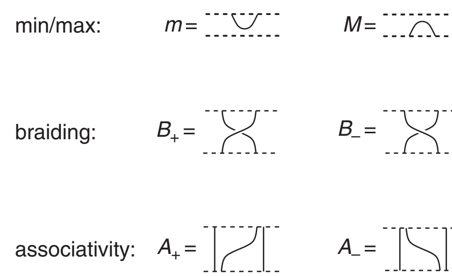
## Combinatorial Construction of the Kontsevich Integral

### Sliced Presentation of Knots

The idea is to cut the knot into a number of standard simple tangles, compute the Kontsevich integral for each of them, and then recover the integral of the whole knot from these simple pieces.

More exactly, we represent the knot by a family of plane diagrams continuously depending on a parameter $\varepsilon \in (0, \varepsilon_0)$ and cut by horizontal planes into a number of slices with the following properties.

1. At every boundary level of a slice (dashed lines in the pictures below), the distances between various strings are asymptotically proportional to different whole powers of the parameter $\varepsilon$.
2. Every slice contains exactly one special event and several strictly vertical strings which are farther away (at lower powers of $\varepsilon$) from any string participating in the event than its width.
3. There are three types of special events:




where, in the two last cases, the strings may be replaced by bunches of parallel strings which are closer to each other than the width of this event.

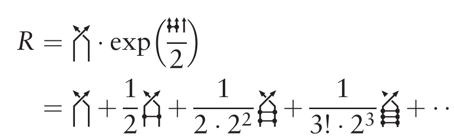
### Recipe of Computation of the Kontsevich Integral

Given such a sliced representation of a knot, the combinatorial algorithm to compute its Kontsevich integral consists in the following:

1. Replace each special event by a series of chord diagrams supported on the corresponding tangle according to the rule

$$m, M \mapsto 1$$
$$B_+ \mapsto R, \quad B_+ \mapsto R^{-1}$$
$$A_+ \mapsto \Phi, \quad A_- \mapsto \Phi^{-1}$$

where



$$\Phi = 1 - \frac{\zeta(2)}{(2\pi i)^2}[a, b] - \frac{\zeta(3)}{(2\pi i)^3}([a,[a,b]] + [b,[a,b]]) + \cdots$$

($\Phi \in \hat{\mathcal{A}}(3)$ is the Knizhnik–Zamolodchikov Drinfeld associator defined below; it is an infinite series in two variables $a =$ [diagram], $b =$ [diagram]).

2. Compute the product of all these series from top to bottom taking into account the connection of the strands of different tangles, thus obtaining an element of the algebra $\hat{\mathcal{A}}'$.

To accomplish the algorithm, we need two auxiliary operations on chord diagrams:

1. $S_i : \mathcal{A}(p) \to \mathcal{A}(p)$ defined as multiplication by $(-1)^k$ on a chord diagram containing $k$ endpoints of chords on the string number $i$. This is the correction term in the computation of $R$ and $\Phi$ in the case when the tangle contains

some strings oriented downwards (the upwards orientation is considered as positive).

2. $\Delta_i : \mathcal{A}(p) \to \mathcal{A}(p+1)$ acts on a chord diagram $D$ by doubling the $i$th string of $D$ and taking the sum over all possible lifts of the endpoints of chords of $D$ from the $i$th string to one of the two new strings. The strings are counted by their bottom points from left to right. This operation can be used to express the combinatorial Kontsevich integral of a generalized associativity tangle (with strings replaced by bunches of strings) in terms of the combinatorial Kontsevich integral of a simple associativity tangle.

### Example

Using the combinatorial algorithm, we compute the Kontsevich integral of the trefoil knot $3_1$ to the terms of degree 2. A sliced presentation for this knot shown in Figure 7 implies that $Z(3_1) = S_3(\Phi) R^{-3} S_3(\Phi^{-1})$ (here the product from left to right corresponds to the multiplication of tangles from top to bottom). Up to degree 2, we have

$$\Phi = 1 + \tfrac{1}{24}[a,b] + \cdots$$
$$R = X\left(1 + \tfrac{1}{2}a + \tfrac{1}{8}a^2 + \cdots\right)$$

where $X$ means that the two strands in each term of the series must be crossed over at the top. The operation $S_3$ changes the orientation of the third strand, which means that $S_3(a) = a$ and $S_3(b) = -b$. Therefore,

$$S_3(\Phi) = 1 - \tfrac{1}{24}[a,b] + \cdots$$
$$S_3(\Phi^{-1}) = 1 + \tfrac{1}{24}[a,b] + \cdots$$
$$R^{-3} = X\left(1 - \tfrac{3}{2}a + \tfrac{9}{8}a^2 + \cdots\right)$$

and

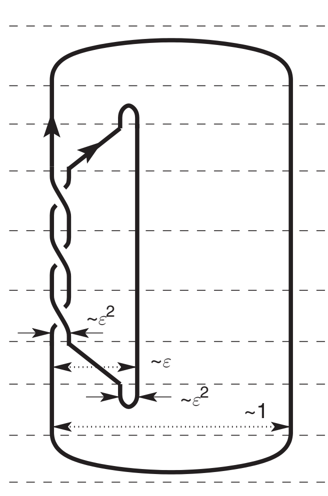



**Figure 7** A sliced presentation of the trefoil.

$$\begin{aligned} Z(3_1) &= \left(1 - \tfrac{1}{24}[a,b] + \cdots\right) \\ &\quad \times X\left(1 - \tfrac{3}{2}a + \tfrac{9}{8}a^2 + \cdots\right)\left(1 + \tfrac{1}{24}[a,b] + \cdots\right) \\ &= 1 - \tfrac{3}{2}Xa - \tfrac{1}{24}abX + \tfrac{1}{24}baX \\ &\quad + \tfrac{1}{24}Xab - \tfrac{1}{24}Xba + \tfrac{9}{8}Xa^2 + \cdots \end{aligned}$$

Closing these diagrams into the circle, we see that in the algebra $\mathcal{A}$ we have $Xa = 0$ (by the framing independence relation), then $baX = Xab = 0$ (by the same relation, because these diagrams consist of two parallel chords) and $abX = Xba = Xa^2 = \otimes$. The result is $Z(3_1) = 1 + (25/24)\otimes + \cdots$. The final Kontsevich integral of the trefoil (in the multiplicative normalization) is thus equal to

$$\begin{aligned} I'(3_1) &= Z(3_1)/Z(H) \\ &= \left(1 + \frac{25}{24}\otimes + \cdots\right) \Big/ \left(1 + \frac{1}{24}\otimes + \cdots\right) \\ &= 1 + \otimes + \cdots \end{aligned}$$

### Drinfeld Associator and Rationality

The Drinfeld associator used as a building block in the combinatorial construction of the Kontsevich integral can be defined as the limit

$$\Phi_{KZ} = \lim_{\varepsilon \to 0} \varepsilon^{-b} Z(AT_\varepsilon)\varepsilon^a$$

where $a =$ ǂǂ|, $b =$ |ǂǂ, and $AT_\varepsilon$ is the positive associativity tangle (special event $A_+$ shown above) with the distance between the vertical strands constant 1 and the distance between the close endpoints equal to $\varepsilon$. An explicit formula for $\Phi_{KZ}$ was found by Le and Murakami (1996); it is written as a nested summation over four variable multi-indices and therefore does not provide an immediate insight into the structure of the whole series; we confine ourselves by quoting the beginning of the series (note that $\Phi_{KZ}$ is a group-like element in the free associative algebra with two generators; hence, its logarithm belongs to the corresponding free Lie algebra):

$$\begin{aligned} \log(\Phi_{KZ}) = &- \zeta(2)[x,y] - \zeta(3)([x,[x,y]] + [y,[x,y]]) \\ &- \frac{\zeta(2)^2}{10}(4[x,[x,[x,y]]] + [y,[x,[x,y]]] \\ &+ 4[y,[y,[x,y]]]) \\ &- \zeta(5)([x,[x,[x,[x,y]]]] + [y,[y,[y,[x,y]]]]) \\ &+ (\zeta(2)\zeta(3) - 2\zeta(5))([y,[x,[x,[x,y]]]] \\ &+ [y,[y,[x,[x,y]]]]) \\ &+ \left(\tfrac{1}{2}\zeta(2)\zeta(3) - \tfrac{1}{2}\zeta(5)\right)[[x,y],[x,[x,y]]] \\ &+ \left(\tfrac{1}{2}\zeta(2)\zeta(3) - \tfrac{3}{2}\zeta(5)\right)[[x,y],[y,[x,y]]] \\ &+ \cdots \end{aligned}$$

where $x = (1/2\pi\mathrm{i})a$ and $y = (1/2\pi\mathrm{i})b$. In general, $\Phi_{\mathrm{KZ}}$ is an infinite series whose coefficients are "multiple zeta values" (Le and Murakami 1996, Zagier 1994)

$$\zeta(a_1, \ldots, a_n) = \sum_{0<k_1<k_2<\cdots<k_n} k_1^{-a_1} \ldots k_n^{-a_n}$$

There are other equivalent definitions of $\Phi_{\mathrm{KZ}}$, in particular one in terms of the asymptotical behavior of solutions of the simplest Knizhnik–Zamolodchikov equation

$$\frac{\mathrm{d}G}{\mathrm{d}z} = \left(\frac{a}{z} + \frac{b}{z-1}\right) G$$

where $G$ is a function of a complex variable taking values in the algebra of series in two noncommuting variables $a$ and $b$ (see Drinfeld (1991)).

It turns out (theorem of Le and Murakami (1996)) that the combinatorial Kontsevich integral does not change if $\Phi_{\mathrm{KZ}}$ is replaced by another series in $\hat{\mathcal{A}}(3)$ provided it satisfies certain axioms (among which the pentagon and hexagon relations are the most important, see Drinfeld (1991) and Le and Murakami (1996)).

Drinfeld (1991) proved the existence of an associator $\Phi_{\mathbb{Q}}$ with rational coefficients. Using it instead of $\Phi_{\mathrm{KZ}}$ in the combinatorial construction, we obtain the following:

**Theorem** (Le and Murakami 1996). *The coefficients of the Kontsevich integral of any knot (tangle) are rational when $Z(K)$ is expanded over an arbitrary basis consisting of chord diagrams.*

## Explicit Formulas for the Kontsevich Integral

### The Wheels Formula

Let O be the unknot; the expression $I(O) = Z(H)^{-1}$ is referred to as the "Kontsevich integral of the unknot." A closed form formula for $I(O)$ was proved in Bar-Natan *et al.* (2003):

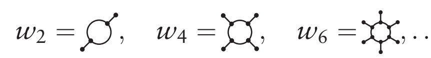
**Theorem**

$$I(O) = \exp \sum_{n=1}^{\infty} b_{2n} w_{2n}$$
$$= 1 + \left(\sum_{n=1}^{\infty} b_{2n} w_{2n}\right) + \frac{1}{2}\left(\sum_{n=1}^{\infty} b_{2n} w_{2n}\right)^2 + \cdots$$

Here $b_{2n}$ are modified Bernoulli numbers, that is, the coefficients of the Taylor series

$$\sum_{n=1}^{\infty} b_{2n} x^{2n} = \frac{1}{2} \ln \frac{\mathrm{e}^{x/2} - \mathrm{e}^{-x/2}}{x}$$

$(b_2 = 1/48, b_4 = -1/5760, b_6 = 1/362\,880, \ldots)$, and $w_{2n}$ are the "wheels," that is, Jacobi diagrams of the form



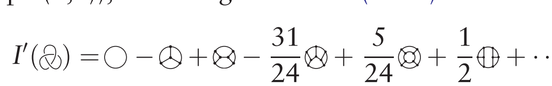
The sums and products are understood as operations in the algebra of Jacobi diagrams $\mathcal{B}$, and the result is then carried over to the algebra of chord diagrams $\mathcal{A}$ along the isomorphism $\chi$.

### Generalizations

There are several generalizations of the wheels formula.

1. Rozansky's rationality conjecture (Rozansky 2003) proved by Kricker (2000) affirms that the Kontsevich integral of any (framed) knot can be written in a form resembling the wheels formula. Let us call the "skeleton" of a Jacobi diagram the regular 3-valent graph obtained by "shaving off" all univalent vertices. Then the wheels formula says that all diagrams in the expansion of $I(O)$ have one and the same skeleton (circle), and the generating function for the coefficients of diagrams with $n$ legs is a certain analytic function, more or less rational in $\mathrm{e}^x$. In the same way, the theorem of Rozansky and Kricker states that the terms in $I(K) \in \hat{\mathcal{B}}$, when arranged by their skeleta, have the generating functions of the form $p(\mathrm{e}^x)/A_K(\mathrm{e}^x)$, where $A_K$ is the Alexander polynomial of $K$ and $p$ is some polynomial function. Although this theorem does not give an explicit formula for $I(K)$, it provides a lot of information about the structure of this series.
2. Marché gives a closed form formula for the Kontsevich integral of torus knots $T(p, q)$.

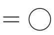
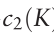
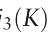
The formula of Marché, although explicit, is rather intricate, and here, by way of example, we only write out the first several terms of the final Kontsevich integral $I'$ for the trefoil (torus knot of type (2,3)), following Willerton (2002):



### First Terms of the Kontsevich Integral

A Vassiliev invariant $v$ of degree $n$ is called "canonical" if it can be recovered from the Kontsevich integral by applying a homogeneous weight system, that is, if $v = \mathrm{symb}(v) \circ I$. Canonical invariants define a grading in the filtered space of Vassiliev invariants which is consistent with the filtration. If the Kontsevich integral is expanded

over a fixed basis in the space of chord diagrams $\hat{\mathcal{A}}'$, then the coefficient of every diagram is a canonical invariant. According to Stanford (2001) and Willerton (2002), the expansion of the final Kontsevich integral up to degree 4 can be written as follows:

$I'(K) =$  $- c_2(K)$  $- \frac{1}{6} j_3(K)$ 

$+ \frac{1}{48}\left(4j_4(K) + 36c_4(K) - 36c_2^2(K) + 3c_2(K)\right)$

$+ \frac{1}{24}\left(-12c_4(K) + 6c_2^2(K) - c_2(K)\right)$

$+ \frac{1}{2}c_2^2(K)$ $+ \cdots$

where $c_n$ are coefficients of the Conway polynomial $\nabla_K(t) = \sum c_n(K)t^n$ and $j_n$ are modified coefficients of the Jones polynomial $J_K(e^t) = \sum j_n(K)t^n$. Therefore, up to degree 4, the basic canonical Vassiliev invariants of unframed knots are $c_2,\ j_3,\ j_4,\ c_4 + (1/12)c_2$, and $c_2^2$.


## Acknowledgments

S Duzhin was partially supported by presidential grant NSh-1972.2003.1.


*See also:* Finite-Type Invariants; Mathematical Knot Theory.

## Further Reading


Bar-Natan D (1995a) On the Vassiliev knot invariants. *Topology* 34: 423–472.

Bar-Natan D (1995b) Vassiliev homotopy string link invariants. *Journal of Knot Theory and Its Ramifications* 4-1: 13–32.

Bar-Natan D, Le TQT, and Thurston DP (2003) Two applications of elementary knot theory to Lie algebras and Vassiliev invariants. *Geometry and Topology* 7(1): 1–31 (arXiv:math. QA/0204311).

Cartier P (1993) Construction combinatoire des invariants de Vassiliev–Kontsevich des nœuds. *Comptes Rendus de l'Académie des Sciences, Série Mathématique Paris* 316(I): 1205–1210.

Chmutov S and Duzhin S (2001) The Kontsevich integral. *Acta Applicandae Mathematicae* 66: 155–190.

Drinfeld VG (1991) On quasi-triangular quasi-Hopf algebras and a group closely connected with Gal($\overline{\mathbb{Q}}/\mathbb{Q}$). *Leningrad Mathematical Journal* 2: 829–860.

Goryunov V (1999) Vassiliev invariants of knots in $R^3$ and in a solid torus. *American Mathematical Society Translations* 190(2): 37–59.

Kneissler J (1997) The number of primitive Vassiliev invariants up to degree twelve, arXiv:math.QA/9706022, June 1997.

Kohno T (1987) Monodromy representations of braid groups and Yang–Baxter equations. *Annales de l'Institut Fourier* 37: 139–160.

Kontsevich M (1993) Vassiliev's knot invariants. *Advances in Soviet Mathematics* 16(2): 137–150.

Kricker A (2000) The lines of the Kontsevich integral and Rozansky's rationality conjecture. Preprint arXiv:math.GT/0005284.

Le TQT and Murakami J (1995) Kontsevich's integral for the Homfly polynomial and relations between values of multiple zeta functions. *Topology and Its Applications* 62: 193–206.

Le TQT and Murakami J (1996) The universal Vassiliev–Kontsevich invariant for framed oriented links. *Compositio Mathematica* 102: 41–64.

Lescop C (1999) Introduction to the Kontsevich integral of framed tangles. Preprint in CNRS Institut Fourier (PostScript file available online at http://www-fourier.ujf-grenoble.fr/~lescop/publi.html).

Marché J (2004) A computation of Kontsevich integral of torus knots, arXiv:math.GT/0404264.

Ohtsuki T (2002) *Quantum Invariants. A Study of Knots, 3-Manifolds and their Sets*. Series on Knots and Everything, vol. 29. Singapore: World Scientific.

Rozansky L (2003) A rationality conjecture about Kontsevich integral of knots and its implications to the structure of the colored Jones polynomial. *Topology and Its Applications* 127: 47–76. Preprint arXiv:math.GT/0106097.

Stanford T Some computational results on mod 2 finite-type invariants of knots and string links. In: *Invariants of Knots and 3-Manifolds* (Kyoto 2001), pp. 363–376.

Willerton S (2002) An almost integral universal Vassiliev invariant of knots. In: *Algebraic and Geometric Topology*, vol. 2, pp. 649–664. Preprint arXiv:math.GT/0105190.

Zagier D (1994) Values of Zeta Functions and their Applications. *First European Congress of Mathematics Progress in Mathematics* vol. 120, pp. 497–512. Basel: Birkhauser.


# Korteweg–de Vries Equation and Other Modulation Equations


**G Schneider**, Universität Karlsruhe, Karlsruhe, Germany

**E Wayne**, Boston University, Boston, MA, USA




Modulation equations are simplified equations used to model complicated physical systems. Typically they are derived from the fundamental partial differential equations that describe the system via asymptotic analysis. Furthermore, the modulation equations are in a sense "universal" in that many different physical systems are described by the same modulation equation. This comes about because the form of the modulation equation depends on only a very few, qualitative features of the original partial differential equation. Thus, they serve a sort of "normal form" for these partial differential equations and as such justify greater study than their apparently special character might otherwise merit.

The Korteweg–de Vries (KdV) equation

$$\partial_t u = \partial_x^3 u + 6u\partial_x u, \quad u = u(x,t),\ x \in \mathbf{R},\ t \geq 0 \quad [1]$$

was one of the earliest modulation equations to be intensively studied. It was derived in an attempt to understand the propagation of solitary waves on the surface of water in a channel of finite depth. The KdV equation was first derived by Boussinesq but then independently rederived and studied in detail by Korteweg and de Vries. (For an interesting discussion of the early history of the KdV equation see Pego and Weinstein (1997).)

## Derivation of the KdV Equation

As mentioned above, the KdV equation is a sort of normal form describing the propagation of small-amplitude, long-wavelength disturbances in a variety of different physical systems. In this section we describe in detail how it arises as an approximation to the Fermi–Pasta–Ulam (FPU) model of coupled, nonlinear oscillators. Although the KdV equation is most commonly encountered as an approximation to water waves, its study as an approximation to the FPU model was extremely important historically because it was in this context that its complete integrability was discovered by Miura (1968) and Gardner *et al.* (1974).

Consider an infinite set of particles of mass $m=1$ at positions $q_j(t), j\in\mathbb{Z}$, interacting with their nearest neighbors via a potential $\mathcal{V}(q)$. Newton's equations for the motion of such particles are:

$$\frac{\mathrm{d}^2 q_j}{\mathrm{d}t^2} = \mathcal{V}'(q_{j+1}(t) - q_j(t)) - \mathcal{V}'(q_j(t) - q_{j-1}(t)), \quad j\in\mathbb{Z} \qquad [2]$$

If we rewrite these equations in terms of the difference variables $r(j,t) = q_{j+1}(t) - q_j(t)$, then [2] becomes

$$\frac{\mathrm{d}^2 r}{\mathrm{d}t^2}(j,t) = \mathcal{V}'(r(j+1,t)) + \mathcal{V}'(r(j-1,t)) - 2\mathcal{V}'(r(j,t)), \quad j\in\mathbb{Z} \qquad [3]$$

We are interested in small-amplitude, long-wavelength, solutions of [3]. One way of studying such motions is to change the lattice spacing in [3] from 1 to $h$ and then let $h$ tend to zero. A nice derivation of the KdV equation from that point of view is contained in Ablowitz and Segur (1981). Here, following Schneider and Wayne (1999), we will keep the lattice spacing fixed at 1 and rescale the spatial variable in the KdV equation. This is closer to the approximation method used in the water wave problem.

Since we want to focus on small-amplitude, long-wavelength solutions of [3], we begin by making the hypothesis that there exists some real-valued function $R(x,t)$ such that the solution of [3] can be written as

$$r(j,t) = \varepsilon^2 R(\varepsilon j, t) \qquad [4]$$

The prefactor $\varepsilon^2$ insures that the solution is of small amplitude while rescaling $j \to \varepsilon j$ means that phenomena that occur on length scales of $\mathcal{O}(1)$ in the equation for $R$ will occur on length scales of $\mathcal{O}(1/\varepsilon)$ in the original equation – that is, they will be long-wavelength solutions. The differing powers of $\varepsilon$ chosen for rescaling the amplitude and the spatial scale are chosen so that the dispersive and nonlinear effects will balance each other. Inserting [4] into [3] and expanding to lowest order in $\varepsilon$ we find that the nonvanishing terms of lowest order in $\varepsilon$ are

$$\frac{\partial^2 R}{\partial t^2} = \varepsilon^2 \mathcal{V}''(0)\frac{\partial^2 R}{\partial x^2} \qquad [5]$$

This is just the wave equation and thus to leading order we expect solutions of [3] to split into a left- and right-moving waves, each moving with speed $c_\varepsilon = \varepsilon\sqrt{\mathcal{V}''(0)}$. (We assume that $c^2 \equiv \mathcal{V}''(0) > 0$.) Thus, we make a refinement of the hypothesized form of the solution and replace [4] by

$$r(j,t) = \varepsilon^2 U(\varepsilon(j+ct), \varepsilon^3 t) + \varepsilon^2 V(\varepsilon(j-ct), \varepsilon^3 t) + \varepsilon^4 \varphi(\varepsilon j, \varepsilon t) \qquad [6]$$

The presence of the term $\varepsilon^4\varphi$ may be somewhat surprising. We will discuss the reason for its appearance in more detail below, but for the moment we mention merely that its presence does not affect the fact that to leading order the solution is approximated by the left- and right-moving waves represented by the $\varepsilon^2 U$ and $\varepsilon^2 V$ terms, respectively. We also note that the additional time dependence $\varepsilon^3 t$ in $U$ and $V$ is chosen, as is typical in the multiscale method to incorporate the higher-order terms omitted in [5] into the evolution.

Substituting [6] into [3] and expanding the resulting equation in $\varepsilon$ we find that the lowest order in $\varepsilon$ that occurs is $\mathcal{O}(\varepsilon^4)$ and these terms all cancel exactly because of the form of our hypothesized solution. The terms of $\mathcal{O}(\varepsilon^6)$ are:

$$\{2c\partial_X\partial_T U - 2c\partial_X\partial_T V + \partial_\tau^2\varphi\} = c^2\left(\tfrac{1}{12}\partial_X^4 U + \tfrac{1}{12}\partial_X^4 V + \partial_\xi^2\varphi\right) + \tfrac{1}{2}\mathcal{V}'''(0)\{\partial_X^2(U^2 + V^2 + 2UV)\} \qquad [7]$$

Here, $X, T, \xi$, and $\tau$ represent the rescaled independent variables, that is, $U = U(X,T), V = V(X,T)$, and $\varphi = \varphi(\xi,\tau)$.

Note that if it were not for the presence of the term $2UV$ on the right-hand side of this last equation the equations for $U$ and $V$ would completely decouple, that is, there would be no interaction between the left- and right-moving parts of the solution to this order. At this point, we can take advantage of the (heretofore) arbitrary function $\varphi$. If we assume that $U$ and $V$ are given, we can choose $\varphi$ to satisfy the inhomogeneous wave equation:

$$\partial_\tau^2 \varphi = c^2 \partial_\xi^2 \varphi + \mathcal{V}'''(0)\partial_X^2(UV) \qquad [8]$$

Then, provided $\varphi$ remains of $\mathcal{O}(1)$ over the timescales of interest (which one can verify *a posteriori*), we see that all terms of $\mathcal{O}(\varepsilon^6)$ in the expansion of [3] will vanish provided

$$\begin{aligned} 2\partial_T U &= \frac{c}{12}\partial_X^3 U + \frac{1}{2c}\mathcal{V}'''(0)\partial_X(U^2) \\ -2\partial_T V &= \frac{c}{12}\partial_X^3 V + \frac{1}{2c}\mathcal{V}'''(0)\partial_X(V^2) \end{aligned} \qquad [9]$$

This means that the left- and right-moving parts of the solution satisfy a pair of uncoupled KdV equations.

**Remark 1** To rewrite [9] in the standard form [1] one can make a simple rescaling – for instance, choose $X = \alpha x, T = t$ and $u(x,t) = \beta U(\alpha x, t)$, with $\alpha = (c/24)^{(1/3)}$ and $\beta = \mathcal{V}'''(0)/(12c\alpha)$.

We can now comment on the reasons we chose the particular scalings of the amplitude and of the independent variables used in [6]. The terms $\partial_X^2 U^2$ and $\partial_X^2 V^2$ are the lowest-order contributions from the nonlinear part of [3], while the terms $\partial_X^4 U$ and $\partial_X^4 V$ represent the lowest-order contributions from the linear part of the equation, except for the "trivial" translation that comes for [5]. In particular, in the absence of nonlinear effects the terms $\partial_X^4 U$ and $\partial_X^4 V$ (or equivalently, the terms $\partial_X^3 U$ and $\partial_X^3 V$ in [9]) would cause traveling waves to "disperse" and thus, the KdV equation represents a balance between nonlinear and dispersive effects. It is this balance between dispersion and nonlinearity which permits traveling-wave solutions to propagate without change of form (see the section "Integrability of the KdV equation").

More generally, we expect the KdV equation to arise as a modulation equation whenever a small-amplitude, long-wavelength linear wave is simultaneously perturbed by dispersive and nonlinear effects of the same order of magnitude. This is, of course, oversimplified. For instance, the original equation may have no quadratic terms in the nonlinearity, for instance, which means that the term $\partial_X U^2$ in the modulation equation will be replaced by a term like $\partial_X U^p$, for $p > 2$ – this leads to the modified KdV equation as the appropriate modulation equation. Or, for certain parameter values in the original equation the coefficient in front of the leading-order dispersive term may vanish, in which case a fifth-order modulation equation known as the Kawahara equation is more appropriate. However, both of these cases are in some sense nongeneric and the relatively weak hypotheses needed to obtain the KdV equation as the appropriate modulation equation indicate why it is encountered in so many diverse circumstances. We note, however, that the multiscale method used above to derive the KdV equation does not give a unique choice for the appropriate modulation equation at any given order of approximation and we discuss in a later section some other equations that could be used as models in the situation above.

## Validity of the KdV Approximation

While the above derivation of the KdV equation is simple and intuitive one may wonder how accurate an approximation it actually provides to the true solutions of [3] (or to the evolution of water waves, probably the most important physical situation in which the KdV approximation is used). In particular, note that in the notation of [9] the phenomena intrinsic to the KdV equation occur on timescales $T = \mathcal{O}(1)$. However, this corresponds to a very long timescale $t = \mathcal{O}(1/\varepsilon^3)$ in the original FPU model and it could easily be the case that although the error made in derivation of the KdV approximation at any given time is quite small, over these very long timescales the errors could accumulate in such a way as to destroy the accuracy of the approximation.

The KdV and other modulation equations have been used since the nineteenth century but only relatively recently have rigorous estimates of the accuracy of this approximation been proved. In fact, the first estimates demonstrating that the KdV equation actually provided an accurate approximation to the true motion of water waves over the timescales expected from the heuristic derivation were not proved until Craig (1985). More recently, powerful general methods have been developed to justify not just the KdV equation but other modulation equations like the nonlinear Schrödinger equation and Ginzburg–Landau equation as well.

For instance, the following method, introduced in Kirrmann *et al.* (1992), has been used to justify the use of modulation equations in the water-wave problem, the evolution of Taylor–Couette patterns in viscous fluids, and a number of other

circumstances. We will explain it in the context of a general, abstract evolution equation to indicate its generality. Suppose that one wishes to approximate the small-amplitude solutions of a general evolution equation (or system of such equations) of the form

$$\partial_t u = Lu + \mathcal{N}(u) \qquad [10]$$

where $L$ is a linear operator and $\mathcal{N}$ represents the nonlinear terms. Suppose that via some formal analysis like that in the previous section we have derived a function $\varepsilon^2\psi$ that is believed to be a good approximation to a true solution of [10]. In that example, for instance, $\varepsilon^2\psi$ would be the sum of the solutions of the two KdV equations in [9], and in general it will be given by the solution of the modulation equation that is expected to approximate [10]. We must show that the difference between $\varepsilon^2\psi$ and a true solution of [10] remains small over the timescales of interest. We write this difference as $u - \varepsilon^2\psi = \varepsilon^\beta R$ so that if $\beta > 2$, and if $R = \mathcal{O}(1), \varepsilon^2\psi$ does provide the leading-order approximation to the true solution. We can make $R|_{t=0}$ small by choosing the initial conditions of our modulation equation appropriately and thus we need to follow how $R$ evolves in time. If we use the equation satisfied by $u$ we see that $R$ evolves as

$$\begin{aligned}\partial_t R = LR + \varepsilon^{-\beta}\big[&\mathcal{N}(\varepsilon^2\psi + \varepsilon^\beta R) \\ &-\mathcal{N}(\varepsilon^2\psi)\big] + \varepsilon^{-\beta}\mathrm{Res}(\varepsilon^2\psi)\end{aligned} \qquad [11]$$

where $\mathrm{Res}(\varepsilon^2\psi) = L(\varepsilon^2\psi) + \mathcal{N}(\varepsilon^2\psi) - \partial_t(\varepsilon^2\psi)$, the "residual" of our approximation is simply the amount by which the approximation fails to satisfy the original equation at any given time. In the example in the previous section the residual would include the terms $\mathcal{O}(\varepsilon^8)$ that we ignored in our expansion.

One must now, in any given example consider three points:

1. The linear evolution of $R$:

$$\partial_t R = LR + D\mathcal{N}(\varepsilon^2\psi)R \qquad [12]$$

   Controlling the solutions of this linear, but nonconstant coefficient partial differential equation is often the most difficult step in proving that solutions of the modulation equation give accurate approximations to the true solution. One can frequently find norms that are preserved by solutions of the leading-order equation $\partial_t R = LR$. However, the term $D\mathcal{N}(\varepsilon^2\psi) = \mathcal{O}(\varepsilon^2)$ if $\mathcal{N}$ is a quadratic nonlinearity. Over the very long timescales (i.e., $\mathcal{O}(\varepsilon^{-3})$) of interest in these approximation problems this $\mathcal{O}(\varepsilon^2)$ term can cause uncontrolled growth of $R$, leading to a breakdown in the approximation. In order to control [12] one must typically make use of some special features of the problem under consideration. For instance, it is sometimes possible to make a coordinate transformation which eliminates the terms of $\mathcal{O}(\varepsilon^2)$ on the right-hand side of [12], after which relatively standard methods suffice to control the solutions of [12].
2. The nonlinear terms in [11]: these terms are of the form $\varepsilon^{-\beta}[\mathcal{N}(\varepsilon^2\psi + \varepsilon^\beta R) - \mathcal{N}(\varepsilon^2\psi)] - D\mathcal{N}(\varepsilon^2\psi)R$. From Taylor's theorem we see that, if the nonlinear term is reasonably smooth, these terms are of $\mathcal{O}(\varepsilon^\beta)$. If $\beta > 3$, these terms are small and can be controlled over the timescales of interest by a straightforward application of Gronwall's inequality or standard "energy estimates."
3. Finally, one must consider the influence of the inhomogeneous terms $\varepsilon^{-\beta}\,\mathrm{Res}(\varepsilon^2\psi)$. Note that if this term is small enough, say $\mathcal{O}(\varepsilon^\beta)$, with $\beta \geq 3$ this term can also be controlled over the relevant timescales by an application of the Gronwall inequality. In order to make this term small, we need to be sure that our approximation $\varepsilon^2\psi$ fails to solve the true equation at any given time by a small amount. In doing so, we can exploit the fact that we can add to our leading-order approximation terms of higher order without affecting the fact that to leading order the true solution is still approximated by the solution of the modulation equation. This is the role of the term $\varepsilon^4\varphi$ in the approximation [6] in the previous section. The leading-order approximation is given by the functions $U$ and $V$ which solve the KdV equations but by adding the additional term $\varepsilon^4\varphi$ to the approximation we cancel the remaining terms of $\mathcal{O}(\varepsilon^6)$ in [7], thereby reducing the size of the residue in that example to $\mathcal{O}(\varepsilon^8)$. This method works in other examples as well so that the inhomogeneous term in [11] can usually be treated by this means. However, in each case, we must prove that the additional terms one adds to the approximation remain bounded over the timescales of interest and demonstrating this fact may not be as easy as it was in the case of the FPU model where the additional term satisfied a simple wave equation.

Using this approach one can show that the approximation derived heuristically in the previous section does accurately model the behavior of solutions of the FPU model over the expected timescales. More precisely, if $r(j, t)$ is the solution of [3] and if $U$ and $V$ are the solutions of the modulation equations [9] (with appropriately chosen, small-amplitude, long-wavelength initial

conditions), one can prove (see Schneider and Wayne (1999)) that for any $T_0 > 0$ there is an $\varepsilon_0 > 0$ and $C > 0$ such that for all $0 < \varepsilon < \varepsilon_0$,

$$\sup_{t\in[0,T_0/\varepsilon^3]} \|r(\cdot,t) - (\varepsilon^2 U(\varepsilon(\cdot + ct), \varepsilon^3 t) + (\varepsilon^2 V(\varepsilon(\cdot - ct), \varepsilon^3 t))\|_{\ell^\infty} \le C\varepsilon^{7/2}$$

One can also use this method to show that the solution of the water-wave problem with general small-amplitude, long-wavelength, initial data can be approximated by the sum of the solutions of a pair of uncoupled KdV equations (Schneider and Wayne 2000), one representing the left-moving part of the solution and one representing the right-moving part of the solution, though in this context the technical difficulties associated with the existence theory for the water-wave problem mean the details are quite a lot more complicated.

## Integrability of the KdV Equation

One reason that normal forms for systems of ordinary differential equations are so useful is that they are frequently integrable – that is, they possess sufficiently many integrals, or constants of motion, that essentially explicit formulas for their solutions can be obtained. Remarkably, the same is true for the KdV equation and for many other modulation equations. An argument for why this is so has been put forth by Calogero and Eckhaus based on the universality of these equations – see Calogero and Eckhaus (1987) and references therein, as well as the article Integrable Systems: Overview for more on this point.

Recall that Boussinesq and Korteweg and de Vries introduced the KdV equation to study solitary traveling waves on a fluid surface. For [1], one has an explicit family of such solutions given by:

$$u(x,t) = 2A^2 \operatorname{sech}^2(A[x + 4A^2 t]), \quad A \ge 0$$

Note that from this formula one sees that waves of large amplitude are narrower and travel faster than waves of small amplitude.

In a famous numerical study, Zabusky and Kruskal made a remarkable discovery. They considered solutions of the KdV equation in which a solitary wave of large amplitude overtook one of smaller amplitude. They found that after a highly nonlinear interaction the two solitary waves re-emerged with their original amplitudes and speeds and the only reminder of their interaction was a phase shift in their relative positions. Their discovery began a search for a mathematical explanation of this remarkable "nonlinear superposition principle" which culminated with the solution of the KdV equation via the method of inverse scattering and the identification of the KdV equation as an infinite-dimensional, completely integrable Hamiltonian system.

We begin by describing how a transformation discovered by Miura (1968) and then generalized by Gardner *et al.* (1974) leads very easily to the conclusion that there are infinitely many conserved quantities for the KdV equation. The basic idea is that given a transformation which maps solutions of one equation to solutions of a second, the existence of simple or "obvious" conserved quantities for the first equation may lead, via the transformation, to more complicated conserved quantities for the second.

Given $u = u(x,t)$, define $w(x,t)$ implicitly via the formula

$$u(x,t) = w(x,t) + \mathrm{i}\varepsilon\partial_x w(x,y) + \varepsilon^2 (w(x,t))^2 \quad [13]$$

Note that if $w$ is smooth enough and $\varepsilon$ is small, we can invert this relation recursively to obtain $w$ in terms of $u$ via the formula

$$\begin{aligned} w = u &- \mathrm{i}\varepsilon\partial_x u - \varepsilon^2(u^2 + \partial_x^2 u) \\ &+ \mathrm{i}\varepsilon^3(\partial_x^3 u + 4u\partial_x^2 u) + \varepsilon^4(2u^3 + 5(\partial_x u)^2 \\ &+ 6u\partial_x^2 u + \partial_x^4 u) + \mathcal{O}(\varepsilon^5) \end{aligned} \quad [14]$$

Now compute

$$\begin{aligned} &\partial_t u - \partial_x^3 u - 6u\partial_x u \\ &= \{\partial_t w - 6w\partial_x w - 6\varepsilon^2 w^2 \partial_x w - \partial_x^3 w\} \\ &\quad + 2\varepsilon^2 w\{\partial_t w - 6w\partial_x w - 6\varepsilon^2 w^2 \partial_x w - \partial_x^3 w\} \\ &\quad + \mathrm{i}\varepsilon\partial_x\{\partial_t w - 6w\partial_x w - 6\varepsilon^2 w^2 \partial_x w - \partial_x^3 w\} \end{aligned} \quad [15]$$

From this we see immediately that if $w$ satisfies the modified KdV equation

$$\partial_t w = 6(w\partial_x w + \varepsilon^2 w^2 \partial_x w) + \partial_x^3 w \quad [16]$$

then $u$, defined by [13] satisfies the KdV equation. However, one also sees immediately that the integral of $w$ is a conserved quantity of [16] for all values of $\varepsilon$, that is, if we define $\mathcal{I}_\varepsilon(t) = \int w(x,t)\,\mathrm{d}x$, then $\mathcal{I}_\varepsilon$ is a constant for all values of $\varepsilon$. (We will assume here that $w$ is defined on the real line, and that $w$ and its derivatives go to zero as $|x|$ tends to infinity. Similar results hold for $x$ running over a finite interval with periodic boundary conditions.) But this in turn immediately implies that if we use [14] to expand $\mathcal{I}_\varepsilon$ in powers of $\varepsilon$ the coefficients in this expansion must also be constants in time. Since these coefficients will be expressed as integrals of $u$ and its derivatives, they will give us (infinitely many)

conserved quantities for the KdV equation! Looking at the first few of these we find:

1. $K_0 = \int u(x,t)\,dx$. The conservation of this quantity follows immediately from the form of the KdV equation.
2. $K_1 = \int \partial_x u(x,t)\,dx = 0$, if we assume that $u$ and its derivatives tend to zero as $|x|$ tends to infinity. Thus, we gain no new information from this quantity and in fact, all the integrals coming from the odd powers of $\varepsilon$ turn out to be "trivial" so we ignore them and focus just on the even powers of $\varepsilon$.
3. $K_2 = \int (u^2 + \partial_x^2 u)\,dx = \int u^2\,dx$. That this is a conserved quantity is again easy to see directly from the KdV equation, just by multiplying the equation by $u$ and integrating with respect to $x$.
4. $K_4 = \int (3u^2 + 5(\partial_x u)^2 + 6u\partial_x^2 u + \partial_x^4 u)\,dx = \int (3u^2 - (\partial_x u)^2)\,dx$. The origin of this integral is not so obvious and we comment further on its meaning below.

Clearly by continuing this procedure we can generate an infinite number of conserved quantities for the KdV equation. Indeed, if one chose another conserved quantity for the modified KdV equation, [16], say $\int w^2(x,t)\,dx$ one could generate another sequence of conserved quantities via this same procedure. However, Kruskal, Miura, Gardner, and Zabusky proved that in fact, all of the conserved quantities that can be written as polynomials in $u$ and its derivatives are already obtained by the procedure above.

The constant of the motion $K_4$ found above is of particular interest because one can write the KdV equation as

$$u_t = \partial_x\left(\frac{\delta K_4}{\delta u}\right) \qquad [17]$$

where $\delta/\delta u$ denotes the variational derivative of $K_4$ with respect to $u(x)$. One can interpret this equation as a Hamiltonian system where $\partial_x$ defines the (nonstandard) symplectic structure and remarkably, Zhakarov and Faddeev (1971) proved that the KdV equation is actually a completely integrable Hamiltonian system. In particular, there exists a canonical transformation such that with respect to the new coordinates the Hamiltonian is a function only of the action variables (and hence in particular, the action variables remain constant in time). The transform which brings the Hamiltonian into its action-angle form is known as the inverse spectral transform and its details would take us beyond the limits of this article. However, very briefly, by observing that the Miura transformation [13] defines a Ricatti differential equation, and using the transformation that converts the Ricatti equation to a linear ordinary differential equation one can relate the solution of the KdV equation to an eigenvalue problem for a linear Schrödinger operator. The potential term in the Schrödinger operator is given by the solution $u(x,t)$ of the KdV equation. Remarkably, it turns out that the eigenvalues of this Schrödinger operator are constants of the motion if $u$ is a solution of the KdV equation and are very closely related to the action variables for the Hamiltonian system. For more details on the inverse-scattering method and its use in solving the KdV equation we refer the reader to the monographs of Ablowitz and Segur (1981), Newell (1985), or the recent book by Kappeler and Pöschel (2003) which develops the theory for the KdV equation on a finite interval with periodic boundary conditions in a particularly elegant fashion.

## Other Mathematical Aspects of the KdV Equation

In addition to the inverse-scattering transform approach, more traditional approaches to the existence and uniqueness of solutions have also been studied, starting with Temam's proof of the well-posedness of solutions of the KdV equation with periodic boundary conditions in the Sobolev space $H^2$. Noting that the Hamiltonian for the KdV equation described in the preceding section is closely related to the $H^1$ norm, this might seem a natural space in which to study well-posedness, but surprisingly Kenig, Ponce, and Vega, and Bourgain showed that the equation is also well posed in Sobolev spaces $H^s$, with $s < 1$ and more recent work has extended the global well-posedness results to Sobolev spaces of small negative order. Aside from their intrinsic interest, these results have other physical implications. If one wishes to study statistical aspects of the behavior of ensembles of solutions of these equations, statistical mechanics suggests that the natural invariant measure for these equations is given by the Gibbs' measure. However, the Gibbs' measure is typically supported on functions less regular than $H^1$, so that in order to define and study this measure one needs to know that solutions of the equation are well behaved in such spaces.

Another natural mathematical question arises from the fact that the KdV equation is only an approximation to the original physical equation. Viewed from another perspective, the original system can be seen as a perturbation of the KdV equation. It then becomes natural to ask whether the special features of the KdV equation are preserved under perturbation. Viewing the KdV equation as a completely integrable Hamiltonian system this is

very analogous to the questions studied by the Kolmogorov–Arnol'd–Moser (KAM) theory and has led to a development of KAM-like results for a number of different partial differential equations like the KdV equation. The results are somewhat technical in nature but roughly speaking they say that if one considers the KdV equation with periodic boundary conditions, temporally periodic or quasi-periodic solutions will persist under small perturbations. The situation is more complicated and less well understood for the equation on the whole line due to the presence of a continuum of scattering states. For a very thorough review of the problem with periodic boundary conditions see Kappeler and Pöschel (2003).

## Other Modulation Equations

As we stressed in its derivation, the KdV equation is an appropriate modulation equation for small-amplitude, long-wavelength solutions in dispersive nonlinear partial differential equations. However, as mentioned in the section "Derivation of the KdV equation" the method of multiple scales does not give a unique modulation equation even in this specific physical regime. Already in his original studies Boussinesq derived at least three different model equations for small-amplitude, long-wavelength water waves and a variety of such models continue to be studied today. For instance, an easy variation in the derivation of the KdV equation leads to the regularized long wave, or Benjamin–Bona–Mahoney equation in which the $\partial_x^3 u$ term in the KdV equation is replaced by the term $\partial_x^2 \partial_t u$. The validity of these alternatives to the KdV equation can also be studied with the aid of the methods described in the section "Validity of the Kdv approximation."

There have been many discussions of which of these modulation equations is the "correct" one. while they may all yield equivalent approximations to the original physical problem the KdV equation has at least two advantages: it is independent of the expansion parameter $\varepsilon$, and it is completely integrable. None of the other equations that have been proposed as approximations to these small-amplitude, long-wavelength phenomena share both of these properties.

If we think in terms of the Fourier transforms of the long-wavelength functions studied above they are solutions whose Fourier transform is concentrated near zero. One can also ask about modulation equations for solutions whose Fourier transform is concentrated about nonzero wave numbers. Such solutions represent a wave train with some fixed underlying wavelength, $\lambda_c$, modulated on a much longer length scale, $\lambda_c/\varepsilon$.

If we make the ansatz that the solution has the form

$$u(x,t) \approx \varepsilon A(\varepsilon(x - c_g t), \varepsilon^2 t) e^{i2\pi(x - c_p t)/\lambda_c} + \text{complex conjugate} \qquad [18]$$

and insert this hypothesized form of the solution into the original equation, then under mild assumptions on the form and properties of the original equation, similar to those under which we derived the KdV equation in an earlier section we find that to the lowest, nontrivial order in $\varepsilon$, the amplitude $A$ evolves according to the nonlinear Schrödinger equation

$$-i\partial_T A = c_1 \partial_X^2 A + c_2 A|A|^2 \qquad [19]$$

If $c_1$ and $c_2$ are both real, the nonlinear Schrödinger equation can also be solved via the inverse-scattering method and it represents another completely integrable modulation equation.

In this article, we have discussed modulation equations only for Hamiltonian, or conservative systems. However, similar equations have also played an important role in the study of dissipative equations like the Navier–Stokes equation. The most common modulation context in that setting is the Ginzburg–Landau equation, which can be derived as a modulation equation for Taylor–Couette rolls or for the convection rolls in the Rayleigh–Bénard problem. Like the nonlinear Schrödinger equation, the Ginzburg–Landau equation describes how slow variations of the amplitude of an underlying periodic pattern evolve and as such it arises in a host of other situations in addition to the fluid dynamics examples mentioned above. For an extensive review of the applications of the Ginzburg–Landau equation, as well as its mathematical properties and some special solutions, see the recent article of Mielke (2002).

*See also:* Bi-Hamiltonian Methods in Soliton Theory; Central Manifolds, Normal Forms; Hamiltonian Fluid Dynamics; Infinite-Dimensional Hamiltonian Systems; Integrable Systems and the Inverse Scattering Method; Integrable Systems: Overview; KAM Theory and Celestial Mechanics; Multiscale Approaches; Partial Differential Equations: Some Examples; WDVV Equations and Frobenius manifolds.

## Further Reading


Ablowitz MJ and Segur H (1981) Solitons and the Inverse Scattering Transform. *SIAM Studies in Applied Mathematics* vol. 4. Philadelphia: Society for Industrial and Applied Mathematics (SIAM).

Calogero F and Eckhaus W (1987) Nonlinear evolutions equations, rescalings, model PDE's and their integrability, i. *Inverse Problems* 3: 229–262.

Craig W (1985) An existence theory for water waves and the Boussinesq and Korteweg–de Vries scaling limits. *Communications in Partial Differential Equations* 10(8): 787–1003.
Gardner CS, Greene JM, Kruskal MD, and Miura RM (1974) Korteweg–deVries equation and generalization, VI. Methods for exact solution. *Communications on Pure and Applied Mathematics* 27: 97–133.
Kappeler T and Pöschel J (2003) KdV and KAM, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge, *A Series of Modern Surveys in Mathematics*. vol. 45 (*Results in Mathematics and Related Areas. 3rd Series, A Series of Modern Surveys in Mathematics*). Berlin: Springer.
Kirrmann P, Schneider G, and Mielke A (1992) The validity of modulation equations for extended systems with cubic nonlinearities. *Proceedings of the Royal Society of Edinburgh Sect. A* 122(1–2): 85–91.
Mielke A (2002) The Ginzburg–Landau equation in its role as a modulation equation. In: *Handbook of Dynamical Systems*, vol. 2. pp. 759–834. Amsterdam: North-Holland.
Miura RM (1968) Korteweg–de Vries equation and generalizations. I. A remarkable explicit nonlinear transformation. *Journal of Mathematical Physics* 9: 1202–1204.
Newell AC (1985) Solitons in Mathematics and Physics. *CBMS-NSF Regional Conference Series in Applied Mathematics*, vol. 48. Philadelphia: Society for Industrial and Applied Mathematics (SIAM).
Pego RL and Weinstein MI (1997) Convective linear stability of solitary waves for Boussinesq equations. *Studies in Applied Mathematics* 99(4): 311–375.
Schneider G and Wayne CE (1999) Counter-propagating waves on fluid surfaces and the continuum limit of the Fermi–Pasta–Ulam model. In: Fiedler B, Gröger K, and Sprekels J (eds.) *International Conference on Differential Equations*, (Berlin, 1999) vols. 1, 2, pp. 390–404. River Edge, NJ: World Scientific.
Schneider G and Wayne CE (2000) The long-wave limit for the water wave problem. I. The case of zero surface tension. *Communications on Pure and Applied Mathematics* 53(12): 1475–1535.
Zakharov VE and Faddeev LD (1971) The Korteweg–de Vries equation is a fully integrable Hamiltonian system. *Functional Analysis and its Applications* 5: 280–287.

# *K*-Theory

**V Mathai**, University of Adelaide, Adelaide, SA, Australia



*K*-theory was invented in the category of algebraic vector bundles over algebraic varieties by A Grothendieck, who was directly motivated by the Hirzebruch–Riemann–Roch theorem which he subsequently greatly generalized. He also defined *K*-homology in terms of coherent sheaves and established the basic properties of *K*-theory and *K*-homology including Poincaré duality for nonsingular varieties. The origin for the choice of the letter *K* in *K*-theory was apparently the German word “Klasse.”

Using the formalism of Grothendieck, M F Atiyah and F Hirzebruch (cf. Karoubi 1978), developed topological *K*-theory in the category of topological (complex) vector bundles over topological spaces. It is this theory that will be the first principal focus of this article. A topological (complex) vector bundle over a compact topological space $X$ is a topological space $E$ together with a continuous map $p: E \to X$ that is onto, such that $p^{-1}(x)$ is a vector space that is isomorphic to $\mathbb{C}^n$ for all $x \in X$, and there is an open cover $\{U\}$ of $X$ together with homeomorphisms $h_U: p^{-1}(U) \to U \times \mathbb{C}^n$ called “local trivializations” with the property that $h_V \circ h_U^{-1}: U \cap V \times \mathbb{C}^n \to U \cap V \times \mathbb{C}^n$ is of the form $(\mathrm{Id}, g_{UV})$, where $g_{UV}: U \cap V \to \mathrm{GL}(n, \mathbb{C})$ are continuous maps satisfying the following cocycle condition on triple overlaps, $g_{UV} g_{VW} g_{WU} = 1$. $X \times \mathbb{C}^n$ is called the trivial vector bundle. Two vector bundles $p: E \to X$ and $q: F \to X$ over $X$ are said to be isomorphic if there is a homeomorphism $\phi: E \to F$ with the property that $p = q \circ \phi$, and which is a linear isomorphism when restricted to each fiber. The direct sum and tensor product of vector spaces carries over to vector bundles. There are canonical isomorphisms $E \oplus F \cong F \oplus E$ and $E \otimes F \cong F \otimes E$, making the set $\mathrm{Vect}(X)$ of isomorphism classes of complex vector bundles over $X$ into a commutative semiring. $\mathrm{Vect}(X)$ can be made into the commutative ring $K^0(X)$ as follows. $K^0(X)$ is generated by pairs $([E], [F])$, together with the relation $([E], [F]) = ([E'], [F'])$ if $E \oplus F' \oplus G \cong E' \oplus F \oplus G$ for some $[G] \in \mathrm{Vect}(X)$. Also $K^1(X)$ is defined to be the group of homotopy classes of continuous maps from $X$ to the infinite unitary group. Around the same time, R Bott proved his celebrated periodicity theorem, which says that the odd homotopy group of the (infinite) unitary group is the integers, whereas the even homotopy groups are all trivial. Incorporating Bott's periodicity theorem for the unitary group into *K*-theory, Atiyah and Hirzebruch proved that topological *K*-theory $K^\bullet(X) = K^0(X) \oplus K^1(X)$ is a periodic generalized cohomology theory, and in what follows, the notation $K^n(X)$ means $n$ modulo 2. If $M$ is not compact, then we can compactify $M$ by adding to it a point $+$ “at infinity,” and denote it by $M^+$. Let $\iota: + \to M^+$ be the inclusion, inducing the pullback

map $\iota^!: K^\bullet(M^+) \to K^\bullet(+) \cong \mathbb{Z}$. Then $K^\bullet(M)$ is defined to be $\ker(\iota^!)$, also called the reduced $K$-theory. If $X_1$ is a closed subset of $X$, the $K$-theory of the pair $(X, X_1)$ is defined as the reduced $K$-theory of the quotient space $X/X_1$. A fundamental computation of Bott is the computation of the $K$-theory of Euclidean space, $K^n(\mathbb{R}^n) \cong \mathbb{Z}$ with canonical generator called the Bott class $b \in K^n(\mathbb{R}^n)$, and $K^{n-1}(\mathbb{R}^n) = \{0\}$.

Some of the basic properties of $K$-theory are listed as follows. Details can be found in Karoubi (1978).

1. *Pullback* If $f: N \to M$ is a continuous map, then given a vector bundle $\pi: E \to M$ over $M$, the pullback vector bundle is defined as $f^*(E) = \{(x, v) \in N \times E : f(x) = \pi(v)\}$ over $N$. This induces a pullback homomorphism, $f^!: K^\bullet(M) \to K^\bullet(N)$.
2. *Push-forward* Let $f: N \to M$ be a smooth proper map between compact manifolds which is $K$-oriented, that is, $TN \oplus f^*TM$ is a spin$^\mathbb{C}$ vector bundle over $N$. Then there is a pushforward homomorphism, also called a Gysin map, $f_!: K^\bullet(N) \to K^{\bullet+d}(M)$. where $d = \dim M - \dim N$, whose construction will be explained in the next section.
3. *Homotopy* If $f: N \to M$ and $g: N \to M$ are homotopic maps, then the pullback maps $f^! = g^!$ are equal. If in addition, $f$ and $g$ are $K$-oriented, proper maps which are homotopic via proper maps, then the Gysin maps $f_! = g_!$ are equal.
4. *Excision* Let $M_1$ be a closed subset of $M$ and $U$ be an open subset of $M$ such that $U$ is contained in the interior of $M_1$. Then the inclusion of pairs $(M \setminus U, M_1 \setminus U) \hookrightarrow (M, M_1)$ induces an isomorphism in $K$-theory, $K^\bullet(M, M_1) \cong K^\bullet(M \setminus U, M_1 \setminus U)$.
5. *Exactness* Let $M_1$ be a closed subset of $M$. Then there is a six-term exact sequence in $K$-theory,

$$\begin{array}{ccccc} K^0(M, M_1) & \longrightarrow & K^0(M) & \longrightarrow & K^0(M_1) \\ \uparrow & & & & \downarrow\Delta \\ K^1(M_1) & \longleftarrow & K^1(M) & \longleftarrow & K^1(M, M_1) \end{array}$$

6. *Cup product* There is a canonical map given by external tensor product, $K^i(M) \otimes K^j(N) \to K^{i+j}(M \times N)$. When $N = M$, one can compose this with the homomorphism induced by the diagonal map $M \to M \times M$ given by $x \to (x, x)$, to get a cup product, $K^p(M) \otimes K^q(M) \to K^{p+q}(M)$.
7. *Bott periodicity* This is arguably the most important property of $K$-theory. It says that the zero-section embedding $\iota^M: M \hookrightarrow M \times \mathbb{R}^n$ induces a Gysin isomorphism, $\iota^M{}_!: K^\bullet(M) \xrightarrow{\cong} K^{\bullet+n}(M \times \mathbb{R}^n)$, which is given as follows. Let $\pi_M: M \times \mathbb{R}^n \to M$ and $\pi_{\mathbb{R}^n}: M \times \mathbb{R}^n \to \mathbb{R}^n$ denote the projections onto the factors, and $b = \iota_! 1 \in K^n(\mathbb{R}^n)$ the Bott element, where $\iota: \{0\} \hookrightarrow \mathbb{R}^n$ is the inclusion of the origin. Then the Bott periodicity isomorphism is given by $\iota^M{}_!(x) = \pi^!_M(x) \cup \pi^!_{\mathbb{R}^n}(b) \in K^{\bullet+n}(M \times \mathbb{R}^n)$ for all $x \in K^\bullet(M)$.

Using the fact that any vector bundle over a contractible space is trivial, together with Bott's periodicity theorem, one deduces the calculation of the $K$-theory of spheres. The calculation for the odd-dimensional spheres given, $K^0(S^{2n-1}) \cong \mathbb{Z} \cong K^1(S^{2n-1})$, and for the even-dimensional spheres $K^0(S^{2n-1}) \cong \mathbb{Z}^2$ and $K^1(S^{2n}) \cong \{0\}$, for all $n \geq 1$.

There is a natural homomorphism of rings called the Chern character, $\mathrm{Ch}: K^\bullet(X) \to H^\bullet(X, \mathbb{Q})$ which is characterized by the following axioms:

1. *Naturality* If $f: N \to M$ is a smooth map, and if $E$ is a vector bundle over $M$, then $\mathrm{Ch}(f^!(E)) = f^*(\mathrm{Ch}(E))$.
2. *Additivity* $\mathrm{Ch}(E \oplus F) = \mathrm{Ch}(E) + \mathrm{Ch}(F)$.
3. *Normalization* If $L$ is the canonical line bundle over $\mathbb{C}P^n$ which restricts to the Hopf line bundle over $\mathbb{C}P^1$, then $\mathrm{Ch}(L) = \exp(x)$, where $x$ is the generator of $H^2(\mathbb{C}P^n, \mathbb{Z}) \cong \mathbb{Z}$.

Atiyah and Hirzebruch, cf. Karoubi (1978), also proved that the Chern character induces an isomorphism of the rings $K^\bullet(X) \otimes \mathbb{Q}$ and $H^\bullet(X, \mathbb{Q})$. The Chern–Weil representative of the Chern character is $\mathrm{tr}(\exp((i/2\pi)\Omega_E))$, where $\Omega_E$ is the curvature of a Hermitian connection on $E$.

There are many variants of $K$-theory, such as $KO$-theory, where the unitary group is replaced by the orthogonal group, which is periodic of order eight, and $G$-equivariant $K$-theory, where $G$ is a compact Lie group. $K$-theory and its variants have many interesting applications such as determining the maximum number of linearly independent vector fields on spheres, which is due to Adams, cf. Karoubi (1978). We will content ourselves with the description of two important applications.

## Grothendieck–Riemann–Roch Theorem for Smooth Manifolds

Recall that an oriented real vector bundle $E$ over $M$ is said to be a spin$^\mathbb{C}$ vector bundle if the bundle of oriented frames on $E$, $\mathrm{SO}(E)$ has a circle bundle $\mathrm{Spin}^\mathbb{C}(E)$ such that the restriction to each fiber yields the central extension $0 \to \mathrm{U}(1) \to \mathrm{Spin}^\mathbb{C}(n) \to \mathrm{SO}(n) \to 0$ that defines the group $\mathrm{Spin}^\mathbb{C}(n)$, where $n$ is the rank of $E$. It turns out that the obstruction to the existence of a spin$^\mathbb{C}$ structure on $E$ is the third integral Stieffel–Whitney class of $E$, $W_3(E) \in H^3(M, \mathbb{Z})$.

A generalization of Bott periodicity is the Thom isomorphism in $K$-theory. It says that if $\pi: E \to M$ is a rank-$n$ spin$^{\mathbb{C}}$ vector bundle over $M$, then the zero-section embedding $\iota^M: M \hookrightarrow E$ induces a Gysin isomorphism, $\iota^M{}_!: K^\bullet(M) \cong K^{\bullet+n}(E)$, which is given as follows. There is a canonical element $\iota_!^M 1 \in K^n(E)$ called the Thom class in $K$-theory, which is characterized by the property that $\iota_! 1$ restricts to give the Bott class on each fiber. Then the Thom isomorphism in $K$-theory is given by $\iota^M{}_!(x) = \pi^!(x) \cup \iota_!^M 1 \in K^{\bullet+n}(E)$ for all $x \in K^\bullet(M)$. For canonical representatives of the Thom class, cf. Mathai–Quillen Formalism, or Mathai and Quillen (1986).

Recall the definition of the Gysin map for smooth embeddings. Let $X$ be a smooth, compact manifold, and $Y$ a smooth manifold. Let $h: X \to Y$ be a smooth embedding that is $K$-oriented. Since $TX \oplus TX$ has a canonical almost-complex structure, it follows that the normal bundle $N_Y X = h^*(TY)/TX$ is a spin$\mathbb{C}$ vector bundle. If $\iota^X: X \hookrightarrow N_Y X$ is the zero-section embedding, then we have the Thom isomorphism $\iota_!^X: K^\bullet(X) \xrightarrow{\cong} K^{\bullet+n}(N_Y X)$, where $n = \dim(Y) - \dim(X)$ is the codimension of the embedding. Upon choosing a Riemannian metric on $Y$, there is a diffeomorphism $\Phi$ from a tubular neighborhood $U$ of $h(X)$ onto a neighborhood of the zero section in the normal bundle $\iota(X)$. That is, $\Phi^!: K^\bullet(N_Y X) \xrightarrow{\cong} K^\bullet(U)$. For any open subset $j: U \hookrightarrow Y$, the extension by zero defines a homomorphism $j: K^\bullet(U) \to K^\bullet(Y)$. Then the Gysin map of the embedding $h$ is defined as $h_! = j \circ \Phi^! \circ \iota_!^X: K^\bullet(X) \to K^{\bullet+n}(Y)$, which turns out to be independent of the choices made.

Next recall the definition of the Gysin map for smooth submersions. Let $\pi: Y \to Z$ be a smooth submersion of smooth manifolds, which is $K$-oriented and a proper map. Since every smooth compact manifold can be smoothly embedded in $\mathbb{R}^{2q}$ for $q$ sufficiently large, a parametrized version yields an embedding $\kappa: Y \hookrightarrow Z \times \mathbb{R}^{2q}$ that is spin$\mathbb{C}$. Therefore the Gysin map is a homomorphism $\kappa_!: K^\bullet(Y) \to K^{\bullet+a}(Z \times \mathbb{R}^{2q})$, where $a = \dim(Z) + 2q - \dim(Y)$. Let $\iota^Z: Z \hookrightarrow Z \times \mathbb{R}^{2q}$ denote the zero-section embedding. Then we have the Thom isomorphism $\iota_!^Z: K^\bullet(Z) \xrightarrow{\cong} K^{\bullet+2q}(Z \times \mathbb{R}^{2q})$. Then the Gysin map of the submersion $\pi$ is defined as $\pi_! = \kappa_! \circ (\iota_!^Z)^{-1}: K^\bullet(Y) \to K^{+b}(Z)$, where $b = \dim(Y) - \dim(Z)$, and turns out to be independent of the choices made.

Let $f: N \to M$ be a smooth proper map that is $K$-oriented. Then $f$ can be canonically factored, first into the smooth embedding $\mathrm{gr}(f): N \hookrightarrow N \times M$, which is the graph of the function, that is, $\mathrm{gr}(f)(x) = (x, f(x))$, and which is $K$-oriented. The Gysin map is $\mathrm{gr}(f)_!: K^\bullet(N) \to K^{\bullet+\dim(M)}(N \times M)$. Second, the projection $p_M: N \times M \to M$ is a $K$-oriented proper submersion, when restricted to the image of $\mathrm{gr}(f)$. The Gysin map is $p_{M!}: K^\bullet(M \times N) \to K^{\bullet+b}(M)$, where $b = \dim(N)$. The Gysin map of $f$ is defined as $f_! = p_{M!} \circ \mathrm{gr}(f)_!: K^\bullet(N) \to K^{\bullet+d}(M)$, where $d = \dim(M) + \dim(N)$.

Given such a smooth proper map $f: N \to M$ that is $K$-oriented. Then there are Gysin maps in cohomology, $f_*: H^\bullet(N, \mathbb{Q}) \to H^{\bullet+d}(M, \mathbb{Q})$ (where we consider the $\mathbb{Z}_2$-grading given by even and odd degree), and in $K$-theory, $f_!: K^\bullet(N) \to K^{\bullet+d}(M)$ which increases the degree by $d = \dim(M) + \dim(N)$. The Grothendieck–Riemann–Roch theorem due to Atiyah and Hirzebruch, cf. Karoubi 1978, in the smooth category can be phrased as the commutativity of the diagram,

$$\begin{array}{ccc} K^\bullet(N) & \xrightarrow{f_!} & K^{\bullet+d}(M) \\ {\scriptstyle \mathrm{Todd}(TN)\cup \mathrm{Ch}}\downarrow & & \downarrow{\scriptstyle \mathrm{Todd}(TM)\cup \mathrm{Ch}} \\ H^\bullet(N, \mathbb{Q}) & \xrightarrow{f_*} & H^{\bullet+d}(M, \mathbb{Q}) \end{array}$$

That is,

$$\mathrm{Ch}(f_!(\xi)) \cup \mathrm{Todd}(TM) = f_*(\mathrm{Ch}(\xi) \cup \mathrm{Todd}(TN))$$

for all $\xi \in K^\bullet(N)$, where $\mathrm{Todd}(E)$ is the Todd genus characteristic class of a Hermitian vector bundle $E$ over $M$. The Chern–Weil representative of the Todd genus is

$$\sqrt{\det\left(\frac{(\mathrm{i}/2\pi)\Omega_E}{\tanh((\mathrm{i}/2\pi)\Omega_E)}\right)}$$

where $\Omega_E$ is the curvature of a Hermitian connection on $E$. There are many useful variants of this beautiful formula.

## The Atiyah–Singer Index Theorem

The 2004 Abel Prize citation mentions the Atiyah–Singer (1971) index theorem as being one of the greatest achievements of twentieth-century mathematics. It has stimulated considerable interaction between mathematicians and mathematical physicists. We content ourselves here with a rudimentary description of the results.

Let $\mathcal{F}$ be the space of all Fredholm operators on an infinite-dimensional complex Hilbert space $\mathcal{H}$. Recall that an operator $A$ is said to be Fredholm if both the kernel and cokernel of $A$ are finite dimensional. The index of such a Fredholm operator is $\mathrm{index}(A) = \dim(\ker(A)) - \dim(\mathrm{coker}(A)) \in \mathbb{Z}$. The index map is continuous, so it induces a map on the connected components of $\mathcal{F}$, which turns out to be an isomorphism.

$K$-theory is naturally related to the space of all Fredholm operators $\mathcal{F}$ endowed with the norm topology. Any continuous map $A: X \to \mathcal{F}$ from a compact space to $\mathcal{F}$ has an index in $K^0(X)$, which is given by $\text{index}(A) = \ker(A) - \text{coker}(A)$ in the special case when $\dim(\ker(A))(x)$ is constant in $x \in X$. In general, one uses the fact that the index is stable under compact perturbation, and shows that one can always achieve the special case after a compact perturbation. It is again the case that the index map is continuous, and so induces a map, index: $[X, \mathcal{F}] \to K^0(X)$, which turns out to be an isomorphism, thanks to a fundamental theorem of Kuiper which proves that the group of all invertible operators on an infinite-dimensional complex Hilbert space is contractible in the norm topology.

Now let $\pi: N \to Z$ be a fiber bundle with typical fiber a smooth compact manifold $M$, where $N$ and $Z$ are also smooth compact manifolds. Consider a smooth family of elliptic operators $D = \{D_z\}_{z \in Z}$ along the fibers of $\pi$, parametrized by $Z$, where $D_z : C^\infty(\pi^{-1}(z), E|_{\pi^{-1}(z)}) \to C^\infty(\pi^{-1}(z), F|_{\pi^{-1}(z)})$ and $E$, $F$ are vector bundles over $N$. Such a family of elliptic operators has a symbol

$$\sigma(D) : \pi^*(E) \to \pi^*(F)$$

where $\pi: T^*(N/Z) \to N$ is the projection and $T^*(N/Z)$ is the vertical cotangent bundle. Ellipticity for the family is the condition that $\sigma(D)$ is an isomorphism outside the zero section, so that the triple $(\pi^*(E), \pi^*(F), \sigma(D))$ determines an element in $K^0(T^*(N/Z))$ denoted by $\sigma(D)$.

The analytic index of the family $D$ is $\text{index}(D) \in K^0(Z)$, and it turns out that it only depends on the class of the symbol $\sigma(D) \in K^0(T^*(N/Z))$, so the analytic index can be viewed as a homomorphism,

$$\text{index} : K^0(T^*(N/Z)) \to K^0(Z)$$

Consider an embedding $\iota: N \hookrightarrow Z \times \mathbb{R}^n$ that is compatible with the projection $\pi: N \to Z$. The fiberwise differential is an embedding $d\iota: T(N/Z) \to Z \times \mathbb{R}^{2n}$, which induces a Gysin map

$$d\iota_! : K^0(T(N/Z)) \to K^0(Z \times \mathbb{R}^{2n})$$

upon identifying $T^*(N/Z)$ with $T(N/Z)$. Let $j: Z \to Z \times \mathbb{R}^{2n}$ be the inclusion $j(z) = (z, 0)$. It induces the Bott isomorphism $j_! : K^0(Z) \cong K^0(Z \times \mathbb{R}^{2n})$. The topological index of the family $D$ is, by definition,

$$\text{index}_t = j_!^{-1} \circ d\iota_! : K^0(T^*(N/Z)) \to K^0(Z)$$

The Atiyah–Singer (1971) index theorem for families of elliptic operators $D$ asserts the equality of the analytic index and the topological index,

$$\text{index}(D) = \text{index}_t(\sigma(D)) \in K^0(Z)$$

Combined with the Grothendieck–Riemann–Roch theorem, one has the following exquisite formula in $H^\bullet(Z, \mathbb{Q})$:

$$\text{Ch}(\text{index}(D)) = \rho_* \pi_* \{\text{Todd}(T^*_{\mathbb{C}}(N/Z)) \cup \text{Ch}(\sigma(D))\}$$

where $\rho: T^*_{\mathbb{C}}(N/Z) \to N$ is the projection.

The map sending a complex vector bundle $E$ over $Z$ to its determinant line bundle $\det(E) = \Lambda^{\max} E$ induces a homomorphism, $\det: K^0(Z) \to \pi_0(\text{Pic}(Z))$, where $\pi_0(\text{Pic}(Z))$ denotes the isomorphism classes of complex line bundles over $Z$. Then

$$c_1(\det(\text{index}(D))) = \left\{\rho_* \pi_* \{\text{Todd}(T^*_{\mathbb{C}}(N/Z)) \cup \text{Ch}(\sigma(D))\}\right\}^{[2]}$$

where $^{[2]}$ denotes the degree-2 component, and the left-hand side denotes the first Chern class of the determinant line bundle of the index class. This formula is often used in the study of anomalies in physics.

## *K*-Theory of $C^*$-Algebras

The Gelfand–Naimark theorem asserts that unital abelian $C^*$-algebras $A$ can be identified with the space of continuous functions $C(X)$, where $X$ is the compact Hausdorff space known as the spectrum of $A$, consisting of characters of $A$. Conversely, given a compact Hausdorff space $X$, the characters of $C(X)$ consist of the evaluation maps at points of $X$.

Let $E$ be a vector bundle over $X$. Then there is a vector bundle $F$ over $X$ such that $E \oplus F \cong X \times \mathbb{C}^n$. Setting $A = C(X)$, $\mathcal{M} = C(X, E)$, $\mathcal{N} = C(X, F)$, we see that $\mathcal{M} \oplus \mathcal{N} \cong A^n$, showing that each vector bundle $E$ over $X$ determines a canonical finite projective module $\mathcal{M}$ over $A$. The converse is also true and is a result of Serre and Swan, cf. Blackadar (1986), which asserts that every finite projective module $\mathcal{M}$ over $A$ is the space of all continuous sections of a vector bundle over $X$. So we have an equivalence of the category of vector bundles over $X$ and the category of finite projective modules over $A$.

This motivates the following generalization of topological $K$-theory for a general unital $C^*$-algebra $A$. Let Proj$(A)$ denote the isomorphism classes of finite projective modules over $A$. It is a commutative semigroup under the operation of direct sum, which can be made into the commutative group $K_0(A)$ as follows: $K_0(A)$ is generated by pairs $([\mathcal{M}], [\mathcal{N}])$, together with the relation $([\mathcal{M}], [\mathcal{N}]) = ([\mathcal{M}'], [\mathcal{N}'])$ if $\mathcal{M} \oplus \mathcal{N}' \oplus \mathcal{G} \cong \mathcal{M}' \oplus \mathcal{N} \oplus \mathcal{G}$ for some $[\mathcal{G}] \in$ Proj

$(A)$. Also $K_1(A) = \pi_0(\mathrm{GL}(\infty, A))$ where $\mathrm{GL}(\infty, A)$ denotes the direct limit of $\mathrm{GL}(n, A)$ where $(\mathrm{GL}(n, A)$ embeds in $\mathrm{GL}(n+1, A)$ as $1 \oplus \mathrm{GL}(n, A)$. Then, defining $K_j(A) = \pi_{j-1}(\mathrm{GL}(\infty, A))$ for $j \geq 1$, together with generalized Bott periodicity which asserts that there is a canonical isomorphism $\pi_{j-1}(\mathrm{GL}(\infty, A)) \cong \pi_{j+1}(\mathrm{GL}\ (\infty, A))$, we see that $K_\bullet(A) = K_0(A) \oplus K_1(A)$ is a generalized periodic cohomology theory. If $A$ is a $C^*$-algebra without unit, then consider $A^+ = A \oplus \mathbb{C}$, with product given by $(a, \lambda)(b, \mu) = (ab + a\mu + b\lambda, \lambda\mu)$ with unit $(0, 1)$. The projection $p: A^+ \to \mathbb{C}$ defined as $p(a, \lambda) = \lambda$ induces a map $p_!: K_\bullet(A^+) \to K_\bullet(\mathbb{C})$. In the nonunital case, $K_\bullet(A)$ is defined as $\ker(p_!)$. Observe that $K_1(A) = K_1(A^+)$, but this is often not the case with $K_0$. It is easy to see that when $A$ has a unit, then the two definitions of $K_0$ agree. An important caveat in the case of noncommutative $C^*$-algebras is that the $K$-theory is often not a ring as there is no analog of the tensor product operation.

Some of the basic properties of $K$-theory are listed as follows. Details can be found in Blackadar (1986).

1. *Cup product* A continuous bilinear map of $C^*$-algebras, $A \times B \to C$, induces a cup product, $K_i(A) \otimes K_j(B) \to K_{i+j}(C)$.
   In particular, the continuous product $A \times A \to A$ induces a cup product homomorphism, $K_i(A) \otimes K_j(A) \to K_{i+j}(A)$.
2. *Induced homomorphism* If $f: A \to B$ is a homomorphism of $C^*$-algebras, then there is an induced homomorphism, $f_!: K_\bullet(A) \to K_\bullet(B)$.
3. *Homotopy* If $f: A \to B$ and $g: A \to B$ are homomorphisms of $C^*$-algebras that are homotopic, the induced homomorphisms on $K$-theory $f_* = g_*$ are equal.
4. *Excision* If $I$ is a closed two-sided ideal in $A$, then there is a six-term exact sequence in $K$-theory,

$$\begin{array}{ccccc} K_0(I) & \longrightarrow & K_0(A) & \longrightarrow & K_0(A/I) \\ \uparrow & & & & \downarrow\Delta \\ K_1(A/I) & \longleftarrow & K_1(A) & \longleftarrow & K_1(I) \end{array}$$

5. *Morita invariance* The inclusion homomorphism of $A$ into the top left of the diagonal in $M_n(A)$ induces an isomorphism in $K$-theory, $K_\bullet(A) \cong K_\bullet(M_n(A))$.
6. *Continuity* Let $A = \lim_{n\to\infty} A_n$ be a $C^*$-direct limit. Then, $K_\bullet(A) = \lim_{n\to\infty} K_\bullet(A_n)$.
7. *Stability* Let $\mathcal{K}$ be a $C^*$-algebra of all compact operators on an infinite-dimensional complex Hilbert space. Then since $\mathcal{K} = \lim_{n\to\infty} M_n(\mathbb{C})$ is a $C^*$-direct limit, we see that $K_\bullet(A \otimes \mathcal{K}) = \lim_{n\to\infty} K_\bullet(A \otimes M_n(\mathbb{C})) = K_\bullet(A)$.
8. *Bott periodicity* The continuous product $A \times \mathbb{C} \to A$ induces the cup product $K_i(A) \otimes K_j(\mathbb{C}) \to K_{i+j}(A)$. The computation by Bott asserts that there is a canonical element $b \in K_2(\mathbb{C})$ that gives an isomorphism $K_2(\mathbb{C}) \cong \mathbb{Z}$, and Bott periodicity asserts that the cup product with $b$ gives rise to an isomorphism $K_i(A) \cong K_{i+2j}(A)$.

We mention in passing that Connes has defined a Chern character homomorphism, $\mathrm{Ch}: K_\bullet(A) \to HE_\bullet(A)$, mapping into the entire cyclic homology of $A$, having similar properties as the ordinary Chern character. Due to space constraints, it will not be defined here.

## A $C^*$-Algebra Generalization of the Atiyah–Singer Index Theorem and the Baum–Connes Conjecture

We content ourselves here with a rudimentary account of the $C^*$-algebra generalization of the Atiyah–Singer index theorem and the Baum–Connes conjecture, and its relevance to the quantum Hall effect and strict deformation quantization. Let $A$ be a $C^*$-algebra.

Let $\mathcal{H}_A = A \otimes \mathcal{H}$, which is the analog of a Hilbert space. Let $\mathcal{F}_A$ be the space of all $A$-Fredholm operators on $\mathcal{H}_A$. Recall that an operator $T$ is said to be $A$-Fredholm if both the kernel and cokernel of $T + K$ are closed and finitely generated projective modules, where $K$ is an $A$-compact operator. The space of $A$-compact operators is by definition the closure of the $A$-finite rank operators. The index of $T$ is

$$\mathrm{index}(T) = [\ker(T+K)] - [\mathrm{coker}(T+K)] \in K_0(A)$$

The index map turns out to be well defined and independent of the choice of $A$-compact perturbation $K$. It is continuous, so it induces a map on the connected components of $\mathcal{F}_A$, which turns out to be an isomorphism, by a theorem of Mingo (cf. Rosenberg (1983, 1989)).

Now let $M$ be a smooth compact manifold. An $A$-vector bundle over $M$ is a locally trivial Banach vector bundle $E$ over $M$ whose fibers have the structure of finitely generated left $A$-modules, with morphisms respecting the $A$-module structure. The isomorphism classes of $A$-vector bundles over $M$ form a commutative semigroup under direct sums, and the associated commutative group is easily identified with $K_0(C(M) \otimes A)$. Let $D: C^\infty(M, E) \to C^\infty(M, F)$ be an elliptic $A$-operator acting between smooth sections of $A$-vector bundles $E$, $F$ over $M$. It

turns out that by elliptic regularity, such an operator is $A$-Fredholm, and has an analytic index,

$$\text{index}(D) \in K_0(A)$$

Associated to each such operator is a symbol

$$\sigma(D) : \pi^*(E) \to \pi^*(F)$$

where $\pi : T^*M \to M$ is the projection. Ellipticity is the condition that $\sigma(D)$ is an isomorphism outside the zero section, so that the triple $(\pi^*(E), \pi^*(F), \sigma(D))$ determines an element in $K_0(C_0(T^*M) \otimes A)$ denoted by $\sigma(D)$. It turns out that the analytic index of $D$ depends only on the class $\sigma(D) \in K_0(C_0(T^*M) \otimes A)$. Therefore, the analytic index can be viewed as a homomorphism,

$$\text{index} : K_0(C_0(T^*M) \otimes A) \to K_0(A)$$

Consider an embedding $\iota : M \hookrightarrow \mathbb{R}^n$, which induces an embedding $d\iota : TM \to \mathbb{R}^{2n}$. The associated Gysin map is $d\iota_! : K_0(C_0(T^*M) \otimes A) \to K_0(C_0(\mathbb{R}^{2n}) \otimes A)$. Let $j : \{0\} \to \mathbb{R}^{2n}$ denote inclusion of the origin in $\mathbb{R}^{2n}$. It induces a Gysin map $j_! : K_0(A) \to K_0(C_0(\mathbb{R}^{2n}) \otimes A)$ which is the Bott periodicity isomorphism. Then the topological index is the homomorphism

$$\text{index}_t = j_!^{-1} \circ d\iota_! : K_0(C_0(T^*M) \otimes A) \to K_0(A)$$

The $C^*$-generalization of the Atiyah–Singer index theorem due to Mishchenko–Formenko, cf. Kasparov (1988), asserts the equality of the analytic index and the topological index,

$$\text{index}(D) = \text{index}_t(\sigma(D)) \in K_0(A)$$

Now let $M$ be a compact even-dimensional spin$^{\mathbb{C}}$ manifold. Then there is a spin$^{\mathbb{C}}$ Dirac operator $D : C^\infty(M, S^+) \to C^\infty(M, S^-)$, where $S^\pm$ is the bundle of half-spinors on $T^*M \otimes L$, where $L$ is a line bundle over $M$ with the property that the first Chern class of $L$ modulo 2, $c_1(L) \bmod 2$ is equal to the second Stieffel–Whitney class of $M$, $w_2(M)$. Let $\Gamma$ be a torsion-free discrete group, and $B\Gamma$ be its classifying space. It is a paracompact space with the property that it is the quotient of $\Gamma$ acting freely on a contractible space $E\Gamma$. Let $C_r^*(\Gamma)$ denote the reduced group $C^*$-algebra, and consider the canonical flat $C_r^*(\Gamma)$ bundle $\mathcal{V}$ over $B\Gamma$ defined as follows:

$$\mathcal{V} = \{E\Gamma \times C_r^*(\Gamma)\}/\Gamma$$

where $\Gamma$ acts on the left on $C_r^*(\Gamma)$ and on the right on $E\Gamma$. Let $f : M \to B\Gamma$ be a continuous map. Then $f^*\mathcal{V}$ is a flat $C_r^*(\Gamma)$-bundle over $M$. Upon choosing a flat connection on $f^*\mathcal{V}$, we can couple the spin$^{\mathbb{C}}$ Dirac operator $D_{\mathcal{V}}$ to act on sections of $S^\pm \otimes f^*\mathcal{V}$. The ellipticity of $D_{\mathcal{V}}$ ensures that it is a $C_r^*(\Gamma)$-Fredholm operator, so it has an analytic index, $\text{index}(D_{\mathcal{V}}) \in K_0(C_r^*(\Gamma))$ by the earlier discussion, which is also equal to the topological index $\text{index}_t(\sigma(D_{\mathcal{V}})) \in K_0(C_r^*(\Gamma))$.

By Baum, Connes, and Douglas, the $K$-homology of $B\Gamma$, $K_0(B\Gamma)$, is generated by the triples $(M, E, f)$ as described above, modulo relations that we will not present here because of space constraints. The assembly map

$$\mu : K_0(B\Gamma) \to K_0(C_r^*(\Gamma))$$

is a homomorphism given by $\mu([(M, E, f)]) = \text{index}(D_{\mathcal{V}})$. The Baum–Connes conjecture asserts that $\mu$ is an isomorphism. There are variants of this conjecture when $\Gamma$ has torsion. The Baum–Connes conjecture has been verified when $\Gamma$ is an amenable group or, for instance, a word hyperbolic group. There are also variants of this conjecture for certain foliations and groupoids, and is an extremely active area of research. The injectivity of the assembly map is related to the Novikov conjecture on the homotopy invariance of the higher signatures (Kasparov 1988), and the obstructions to the existence of Riemannian metrics of positive scalar curvature on compact spin manifolds (Rosenberg 1983, 1989). A variant of the Baum–Connes conjecture, where the reduced group $C^*$-algebra is replaced by the twisted reduced group $C^*$-algebra, is used in the analysis of the noncommutative geometry approach to the integer and fractional quantum Hall effect, and also the gaps in the spectrum of magnetic Schrödinger operators (Bellissard *et al.* 1994, Marcolli and Mathai 2001).

## Twisted *K*-theory and the Chern Character

We begin by reviewing some results due to Dixmier and Douady (1963). Let $M$ be a smooth manifold, let $\mathcal{H}$ denote an infinite-dimensional, separable, Hilbert space and let $\mathcal{K}$ be the $C^*$-algebra of compact operators on $\mathcal{H}$. Let $\text{U}(\mathcal{H})$ denote the group of unitary operators on $\mathcal{H}$ endowed with the strong operator topology and let $\text{PU}(\mathcal{H}) = \text{U}(\mathcal{H})/\text{U}(1)$ be the projective unitary group with the quotient space topology, where $\text{U}(1)$ consists of scalar multiples of the identity operator on $\mathcal{H}$ of norm equal to 1. Since $\text{U}(\mathcal{H})$ is contractible in the operator norm topology, it follows that $\text{PU}(\mathcal{H}) = B\text{U}(1)$ is an Eilenberg–MacLane space $K(\mathbb{Z}, 2)$. Therefore, $B\text{PU}(\mathcal{H})$ is an Eilenberg–MacLane space $K(\mathbb{Z}, 3)$. That is, principal $\text{PU}(\mathcal{H})$ bundles $P$ over $X$ are classified up to isomorphism by

the Dixmier–Douady class $DD(P)$ in $H^3(X,\mathbb{Z})$ and conversely.

For $g \in \mathrm{U}(\mathcal{H})$, let $\mathrm{Ad}(g)$ denote the automorphism $T \to gTg^{-1}$ of $\mathcal{K}$. As is well known, Ad is a continuous homomorphism of $\mathrm{U}(\mathcal{H})$, given the strong operator topology, onto $\mathrm{Aut}(\mathcal{K})$ with kernel the circle of scalar multiples of the identity where $\mathrm{Aut}(\mathcal{K})$ is given the point-norm topology. Under this homomorphism we may identify $\mathrm{PU}(\mathcal{H})$ with $\mathrm{Aut}(\mathcal{K})$. Define an Azumaya bundle to be a locally trivial bundle $\mathcal{E}$ over $X$ with fiber $\mathcal{K}$ and structure group $\mathrm{Aut}(\mathcal{K})$. They are of the form $\mathcal{K}_P = \{P \times \mathcal{K}\}/\mathrm{PU}(\mathcal{H})$ and isomorphism classes of Azumaya bundles are also parametrized by their Dixmier–Douady class $\mathrm{DD}(P)$ in $H^3(X,\mathbb{Z})$ and conversely.

Since $\mathcal{K}\otimes\mathcal{K}\cong\mathcal{K}$, the isomorphism classes of locally trivial bundles over $X$ with fiber $\mathcal{K}$ and structure group $\mathrm{Aut}(\mathcal{K})$ form a group under the tensor product, where the inverse of such a bundle is the conjugate bundle. This group is known as the infinite Brauer group and is denoted by $\mathrm{Br}^\infty(X)$. So, a restatement of the Dixmier–Douady theorem is that $\mathrm{Br}^\infty(X)\cong H^3(X,\mathbb{Z})\,.\,H^3(X,\mathbb{Z})$ can also be described in terms of bundle gerbes (Murray 1996).

The twisted $K$-theory, $K^\bullet(X,P)$, is defined as the $K$-theory of the $C^*$-algebra of continuous sections of the Azumaya bundle $\mathcal{K}_P, K_\bullet(C(X,\mathcal{K}_P))$. It was studied in the torsion case by Donovan and Karoubi, where one can replace the compact operators $\mathcal{K}$ by finite-dimensional matrices, and was studied in the general case by Rosenberg (1983, 1989). Let $\mathcal{F}$ be the space of all Fredholm operators endowed with the norm topology. Then, one can form the bundle of Fredholm operators $\mathcal{F}_P = \{P\times\mathcal{F}\}/\mathrm{PU}(\mathcal{H})$, where $\mathrm{PU}(\mathcal{H})$ acts on $\mathcal{F}$ via the adjoint action. Consider the fibration $\mathcal{K}_P \to \mathcal{F}_P \to \mathrm{GL}(\mathcal{C}_P)$, where $\mathcal{C}_P = \{P\times\mathcal{C}\}/\mathrm{PU}(\mathcal{H})$ and $\mathcal{C} = \mathcal{B}(\mathcal{H})/\mathcal{K}$ is the Calkin algebra. Since $\pi_0(C(X,\mathcal{K}_P)) = \{0\}$, we see that $\pi_0(C(X,\mathcal{F}_P)) = \pi_0(C(X,\mathrm{GL}(\mathcal{C}_P)))$. Consider the short exact sequence of $C^*$-algebras,

$$0\to C(X,\mathcal{K}_P)\to C(X,\mathcal{B}_P)\to C(X,\mathcal{C}_P)\to 0$$

where $\mathcal{B}_P = \{P\times\mathcal{B}(\mathcal{H})\}/\mathrm{PU}(\mathcal{H})$ and where $\mathrm{PU}(\mathcal{H})$ acts on $\mathcal{B}(\mathcal{H})$ via the adjoint action. It gives rise to a six-term exact sequence

$$\begin{array}{ccccc} K_0(C(X,\mathcal{K}_P)) & \longrightarrow & K_0(C(X,\mathcal{B}_P)) & \longrightarrow & K_0(C(X,\mathcal{C}_P)) \\ {\scriptstyle\text{index}}\uparrow & & & & \downarrow \\ K_1(C(X,\mathcal{C}_P)) & \longleftarrow & K_1(C(X,\mathcal{B}_P)) & \longleftarrow & K_1(C(X,\mathcal{K}_P)) \end{array}$$

By definition, $K_1(C(X,\mathcal{C}_P))\cong\pi_0(C(X,\mathrm{GL}(\infty,\mathcal{C}_P)))$ and a standard argument shows that this is also equal to $\pi_0(C(X,\mathrm{GL}(\mathcal{C}_P)))$. By Kuiper's theorem, it is not difficult to see that $K_\bullet(C(X,\mathcal{B}_P))=\{0\}$. Therefore,

$$\mathrm{index}\colon \pi_0(C(X,\mathcal{F}_P))\to K^0(X,P)$$

is an isomorphism. Let $X_1$ be a closed subset of $X$, and $I_{X_1}$ be the closed ideal of sections of $\mathcal{K}_P$ that vanish on $X_1$. Then $K^\bullet(X,X_1,P)$ is by definition $K_\bullet(I_{X_1})$. A geometric description of twisted $K$-theory in terms of modules for bundle gerbes is described in Bouwknegt *et al.* (2002).

Some of the basic properties of twisted $K$-theory are listed as follows. Many of these properties follow from the corresponding properties for the $K$-theory of $C^*$-algebras. See Atiyah and Segal and Bouwknegt *et al.* (2002).

1. *Normalization* If $P$ is trivial, then $K^\bullet(M,P) = K^\bullet(M)$.
2. *Module property* $K^\bullet(M,P)$ is a module over $K^0(M)$.
3. *Pullback* If $f: N\to M$ is a continuous map, and $P$ a principal $\mathrm{PU}(\mathcal{H})$ bundle over $M$, then there is a pullback homomorphism $f: K^\bullet(M,P)\to K^\bullet(N,f(P))$.
4. *Push-forward* Let $f: N\to M$ be a smooth proper map between compact manifolds which is $K$-oriented, that is, $TN\oplus f^*TM$ is a $\mathrm{spin}^{\mathbb{C}}$ vector bundle over $N$. Let $P$ be a principal $\mathrm{PU}(\mathcal{H})$ bundle over $M$. Then there is a pushforward homomorphism, also called a Gysin map, $f: K^\bullet(N,f^!(P))\to K^{\bullet+d}(M,P)$, where $d = \dim M - \dim N$.
5. *Homotopy* If $f: N\to M$ and $g: N\to M$ are homotopic maps, then the pullback maps $f^! = g^!$ are equal. If in addition, $f$ and $g$ are $K$-oriented, then the pushforward maps $f_! = g_!$ are equal.
6. *Excision* Let $M_1$ be a closed subset of $M$ and $U$ be an open subset of $M$ such that $U$ is contained in the interior of $M_1$. Then the inclusion of pairs $(M\setminus U, M_1\setminus U)\hookrightarrow(M,M_1)$ induces an isomorphism in $K$-theory, $K^\bullet(M,M_1,P)\cong K^\bullet(M\setminus U, M_1\setminus U, P|_{M\setminus U})$.
7. *Exactness* Let $M_1$ be a closed subset of $M$ and $\iota: M_1\to M$ be the inclusion. Let $P$ be a principal $\mathrm{PU}(\mathcal{H})$ bundle over $M$. Then the short exact sequence

$$0\to I_{M_1}\to C(M,\mathcal{K}_P)\to C\left(M_1,\mathcal{K}_{P|_{M_1}}\right)\to 0$$

gives rise to the six-term exact sequence in $K$-theory,

$$\begin{array}{ccccc} K^0(M,M_1,P) & \longrightarrow & K^0(M,P) & \longrightarrow & K^0(M_1,\iota^!(P)) \\ \uparrow & & & & \downarrow\Delta \\ K^1(M_1,\iota^!(P)) & \longleftarrow & K^1(M,P) & \longleftarrow & K^1(M,M_1,P) \end{array}$$

8. *Cup product* Let $P$ be a principal $\mathrm{PU}(\mathcal{H})$ bundle over $M$ and $Q$ be a principal $\mathrm{PU}(\mathcal{H})$ bundle over $N$. An identification $\mathcal{H} \otimes \mathcal{H} \cong \mathcal{H}$ gives rise to a principal $\mathrm{PU}(\mathcal{H})$ bundle $P \otimes Q$ over $M \times N$ whose Dixmier–Douady invariant is $DD(P \otimes Q) = p_1^*(DD(P)) + p_2^*(DD(Q))$, where $p_j$ denote projections onto the $j$th factor, $j = 1, 2$. Then there is a canonical map given by external tensor product,
$$K^i(M, P) \otimes K^j(N, Q) \to K^{i+j}(M \times N, P \otimes Q)$$
called the cup product.
9. *Bott periodicity* Let $P$ be a principal $\mathrm{PU}(\mathcal{H})$ bundle over $M$. Bott periodicity says that there is a canonical isomorphism
$$K^\bullet(M, P) \cong K^{\bullet+n}(M \times \mathbb{R}^n, \pi(P))$$
where $\pi : M \times \mathbb{R}^n \to M$ is the projection onto the first factor. Let $b \in K^n(\mathbb{R}^n)$ be the Bott element. Then the isomorphism above is given by $\pi^!(x) \cup b \in K^{\bullet+n}(M \times \mathbb{R}^n, \pi^!(P))$ for all $x \in K^\bullet(M, P)$.

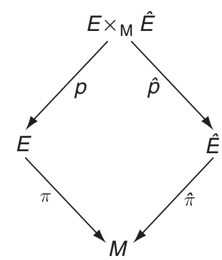
There is a natural homomorphism of rings called the twisted Chern character, which depends both on a choice of $P$ and a de Rham representative $H$ of $DD(P)$,
$$\mathrm{Ch}_P : K^\bullet(M, P) \to H^\bullet(M, H)$$
Here $H^\bullet(M, H)$ denotes the twisted cohomology, which is by definition the cohomology of the complex $(\Omega^\bullet(M), d - H\wedge)$. The twisted Chern character is characterized by the following axioms:

1. *Naturality* If $f : N \to M$ is a smooth map, and if $x \in K^\bullet(M, P)$, then $\mathrm{Ch}_{f(P)}(f^!(x)) = f^*(\mathrm{Ch}_P(x))$.
2. *Additivity* If $x, y \in K^\bullet(M, P)$, then $\mathrm{Ch}_P(x \oplus y) = \mathrm{Ch}_P(x) + \mathrm{Ch}_P(y)$.
3. $\mathrm{Ch}_P$ respects the $K^0(M)$-module structure of $K^0(M, P)$.
4. *Normalization* If $P$ is trivial, then $\mathrm{Ch}_P$ reduces to the ordinary Chern character Ch.

It turns out that the twisted Chern character induces an isomorphism of the rings $K^\bullet(M, P) \otimes \mathbb{Q}$ and $H^\bullet(M, H)$. The Chern–Weil representative of the twisted Chern character is derived in Bouwknegt *et al.* (2002).

## Twisted *K*-Theory and Duality in Type II String Theories

Let $E$ be an oriented $S^1$-bundle over $M$,
$$\begin{array}{ccc} S^1 & \longrightarrow & E \\ & & \pi\downarrow \\ & & M \end{array}$$
characterized by its first Chern class $c_1(E) \in H^2(M, \mathbb{Z})$, in the presence of (possibly nontrivial) $H$-flux $H \in H^3(E, \mathbb{Z})$. We will argue that the $T$-dual of $E$ is again an oriented $S^1$-bundle over $M$, denoted by $\hat{E}$,
$$\begin{array}{ccc} \hat{S}^1 & \longrightarrow & \hat{E} \\ & & \hat{\pi}\downarrow \\ & & M \end{array}$$
supporting $H$-flux $\hat{H} \in H^3(\hat{E}, \mathbb{Z})$, such that
$$c_1(\hat{E}) = \pi_* H, \qquad c_1(E) = \hat{\pi}_* \hat{H}$$
where $\pi_* : H^k(E, \mathbb{Z}) \to H^{k-1}(M, \mathbb{Z})$ and, similarly, $\boldsymbol{\pi}_*$ denote the pushforward maps. Then we can form the following commutative diagram:




The correspondence space $E \times_M \hat{E}$ is a circle bundle over $E$ with first Chern class $\pi^*(c_1(\hat{E}))$, and it is also a circle bundle over $\hat{E}$ with first Chern class $\boldsymbol{\pi}^*(c_1(E))$, by the commutativity of the diagram above. If $\hat{E} = E$ or if $\hat{E} = M \times S^1$, then the correspondence space $E \times_M \hat{E}$ is diffeomorphic to $E \times S^1$.

$T$-duality gives an isomorphism of the twisted $K$-theories of $E$ and $\hat{E}$ as well as an isomorphism between the twisted cohomologies of $E$ and $\hat{E}$, and can be expressed in the following commutative diagram:
$$\begin{array}{ccc} K^\bullet(E, P) & \xrightarrow{T} & K^{\bullet+1}(\hat{E}, \hat{P}) \\ \mathrm{Ch}_P\downarrow & & \downarrow\mathrm{Ch}_{\hat{P}} \\ H^\bullet(E, H) & \xrightarrow{T_*} & H^{\bullet+1}(\hat{E}, \hat{H}) \end{array}$$
where the horizontal arrows are isomorphisms. Here $P$ is a principal $\mathrm{PU}(\mathcal{H})$ bundle over $E$ such that $DD(P) = H$ and $\hat{P}$ is a principal $\mathrm{PU}(\mathcal{H})$ bundle over $\hat{E}$ such that $DD(\hat{P}) = H$. We refer to Bouwknegt *et al.* (2004) for details. The $T$-duality isomorphism above gives compelling evidence that a type II$A$ string theory $A$ on a circle bundle of radius $R$ in the presence of a background $H$-flux, and a type II$B$ string theory $B$ on a "T-dual" circle bundle of radius

$1/R$ in the presence of a "$T$-dual" background H-flux, are equivalent in the sense that the string states of string theory A are in canonical one-to-one correspondence with the string states of string theory B.

We briefly mention two other applications of twisted $K$-theory. Consider the adjoint action of a compact connected simple Lie group $G$ on itself, and the corresponding twisted $G$-equivariant $K$-theory, twisted by a multiple of the generator of $H^3(G, \mathbb{Z})$. The relevance of the equivariant case to conformal field theory was highlighted by the result of Freed, Hopkins and Teleman (see Freed (2002)) that it is graded isomorphic to the Verlinde algebra of $G$, with a shift given by the dual Coxeter number. Here the Verlinde algebra consists of equivalence classes of positive-energy representations of the loop group of $G$ which was originally shown to be a ring in a rather nontrivial way. On the other hand, the ring structure of the twisted $G$-equivariant $K$-theory of $G$ is just induced by the product on $G$, which makes this result all the more remarkable.

Fractional analytic index theory, developed in Mathai *et al.* is a generalization of Atiyah–Singer index theory, assigning a fractional-valued analytic index to each projective elliptic operator on a compact manifold, where the fraction need not be an integer. These projective elliptic operators act on projective vector bundles, where the usual compatibility condition on triple overlaps to give a global vector bundle, may fail by a scalar factor. These are the geometric objects in twisted $K$-theory, when the twist is torsion. In Mathai *et al.*, a fractional index theorem is proved, computing the fractional-valued analytic index of projective elliptic operators essentially in terms of topological data. The Dirac operator in the absence of a spin structure is also defined there for the first time resolving a long standing mystery, and its index is computed.

Some topics not covered in this brief account of $K$-theory include: $KK$-theory, cf. Blackadar (1986) and Kasparov (1988), which is natural setting for the Atiyah–Singer index theorem and its generalizations, as well as higher algebraic $K$-theory.

*See also:* $C^*$-Algebras and Their Classification; Characteristic Classes; Cohomology Theories; Equivariant Cohomology and the Cartan Model; Gerbes in Quantum Field Theory; Index Theorems; Intersection Theory; Mathai–Quillen Formalism; Spectral Sequences.

## Further Reading

Atiyah MF and Segal G Twisted *K*-theory, math.KT/0407054.

Atiyah MF and Singer IM (1971) The index of elliptic operators, IV. *Annals of Mathematics* 93: 119–138.

Bellissard J, van Elst A, and Schulz-Baldes H (1994) The noncommutative geometry of the quantum Hall effect. *Journal of Mathematical Physics* 35: 5373–5451.

Blackadar B (1986) *K*-Theory for Operator Algebras. In: *Mathematical Sciences Research Institute Publications*, vol. 5, viii+338 pp. New York: Springer.

Bouwknegt P, Carey A, Mathai V, Murray MK, and Stevenson D (2002) Twisted *K*-theory and the *K*-theory of bundle gerbes. *Communications in Mathematical Physics* 228(1): 17–49.

Bouwknegt P, Evslin J, and Mathai V (2004) T-duality: topology change from H-flux. *Communications in Mathematical Physics* 249: 383 (hep-th/0306062).

Bouwknegt P, Evslin J, and Mathai V (2004b) On the topology and flux of T-dual manifolds. *Physical Review Letters* 92: 181601.

Dixmier J and Douady A (1963) Champs continues d'espaces hilbertiens at de C*-algèbres. *Bull. Soc. Math. France* 91: 227–284.

Freed DS (2002) Twisted *K*-theory and loop groups. *Proceedings of the International Congress of Mathematicians* (Beijing, 2002) vol. III pp. 419–430.

Karoubi M (1978) *K-theory. An Introduction* Grundlehren der Mathematischen Wissenschaften, Band 226, xviii+308 pp. Berlin: Springer.

Kasparov GG (1988) Equivariant *KK*-theory and the Novikov conjecture. *Invent. Math.* 91(1): 147–201.

Marcolli M and Mathai V (2001) Twisted index theory on good orbifolds, II: fractional quantum numbers. *Communications in Mathematical Physics* 217(1): 55–87.

Mathai V, Melrose RB, and Singer IM, The fractional analytic index, math.DG/0402329.

Mathai V and Quillen DG (1986) Superconnections, Thom classes and equivariant differential forms. *Topology* 26: 85–110.

Murray MK (1996) Bundle gerbes. *J. London Math. Soc.* 54: 403–416.

Rosenberg J (1983) C*-algebras, positive scalar curvature, and the Novikov conjecture. *Inst. Hautes tudes Sci. Publ. Math* 58: 197–212.

Rosenberg J (1989) Continuous-trace algebras from the bundle theoretic point of view. *J. Austral. Math. Soc. Ser. A* 47(3): 368–381.

# L

## Lagrangian Dispersion (Passive Scalar)

**G Falkovich**, Weizmann Institute of Science, Rehovot, Israel



### Introduction

To describe transport by a random flow, one needs to apply the statistical methods to the motion of fluid particles, that is, to the Lagrangian dynamics. We first present the propagators describing evolving probability distributions of different configurations of fluid particles. We then use those propagators to describe decay and steady states of a passive scalar field transported by random flows.

Consider an evolution of a passive scalar tracer $\theta(\boldsymbol{r},t)$ in a random flow. The mean value of the scalar tracer at a given point is an average over values brought by different trajectories:

$$\langle \theta(\boldsymbol{r},s)\rangle = \int \mathcal{P}(\boldsymbol{r},s;\boldsymbol{R},0)\,\theta(\boldsymbol{R},0)\,\mathrm{d}\boldsymbol{R} \qquad [1]$$

Here, $\mathcal{P}(\boldsymbol{r},s;\boldsymbol{R},t)$ is the probability density function (PDF) to find the particle at time $t$ at position $\boldsymbol{R}$ given its position $\boldsymbol{r}$ at time $s$. That PDF is called the propagator or the Green function. Multipoint correlation functions of the tracer

$$C_N(\underline{\boldsymbol{r}},s) \equiv \langle \theta(\boldsymbol{r}_1,s)\ldots\theta(\boldsymbol{r}_N,s)\rangle = \int \mathcal{P}_N(\underline{\boldsymbol{r}},s;\underline{\boldsymbol{R}},0)\,\theta(\boldsymbol{R}_1,0)\ldots\theta(\boldsymbol{R}_N,0)\,\mathrm{d}\underline{\boldsymbol{R}} \qquad [2]$$

are expressed via the multiparticle Green functions $\mathcal{P}_N$ which are the joint PDFs of the equal-time positions $\underline{\boldsymbol{R}}=(\boldsymbol{R}_1,\ldots,\boldsymbol{R}_N)$ of $N$ fluid trajectories.

The trajectory of the fluid particle that passes at time $s$ through the point $\boldsymbol{r}$ is described by the vector $\boldsymbol{R}(t;\boldsymbol{r},s)$ which satisfies $\boldsymbol{R}(t;\boldsymbol{r},t)=\boldsymbol{r}$ and the stochastic equation

$$\dot{\boldsymbol{R}} = \boldsymbol{v}(\boldsymbol{R},t) + \boldsymbol{u}(t) \qquad [3]$$

Here, $\boldsymbol{u}(t)$ describes the molecular Brownian motion with zero average and covariance $\langle u^i(t)u^j(t')\rangle = 2\kappa\delta^{ij}\delta(t-t')$. We also consider macroscopic velocity $\boldsymbol{v}$ as random with various statistical properties in space and time. There is a clear scale separation between macroscopic velocity $\boldsymbol{v}$ and molecular diffusion $\boldsymbol{u}$ that allows one to treat them separately.

Using [3], one can write the Green's function as an integral over paths that satisfy $\boldsymbol{q}(s)=\boldsymbol{r}$ and $\boldsymbol{q}(t)=\boldsymbol{R}$:

$$\mathcal{P}(\boldsymbol{r},s;\boldsymbol{R},t) = \left\langle \int \mathcal{D}\boldsymbol{p}\mathcal{D}\boldsymbol{q}\exp\left(-\int_s^t \imath\boldsymbol{p}(\tau)\cdot[\dot{\boldsymbol{q}}(\tau) - \boldsymbol{v}(\boldsymbol{q}(\tau),\tau) - \boldsymbol{u}(\tau)]\,\mathrm{d}\tau\right)\right\rangle_{\boldsymbol{v},\boldsymbol{u}} \qquad [4]$$

$$= \left\langle \int \mathcal{D}\boldsymbol{p}\mathcal{D}\boldsymbol{q}\,\exp\left(-\int_s^t [\imath\boldsymbol{p}(\tau)\cdot(\dot{\boldsymbol{q}}(\tau) - \boldsymbol{v}(\boldsymbol{q}(\tau),\tau)) + \kappa\boldsymbol{p}^2(\tau)]\,\mathrm{d}\tau\right)\right\rangle_{\boldsymbol{v}} \qquad [5]$$

$$= \left\langle \int \mathcal{D}\boldsymbol{q}\exp\left(-\frac{1}{4\kappa}\int_s^t [\dot{\boldsymbol{q}}(\tau) - \boldsymbol{v}(\boldsymbol{q}(\tau),\tau)]^2\,\mathrm{d}\tau\right)\right\rangle_{\boldsymbol{v}} = \langle P(\boldsymbol{r},s;\boldsymbol{R},t|\boldsymbol{v})\rangle_{\boldsymbol{v}} \qquad [6]$$

The integration over the auxiliary field $\boldsymbol{p}$ in [4] enforces the delta function of [3]. One passes from [4] to [5] by averaging over the Gaussian Brownian noise, and from [5] to [6] by calculating Gaussian integral over $\boldsymbol{p}$.

Generally, exact calculations are only possible for Gaussian random processes short-correlated in time-like in [5]. The simplest case is the Brownian motion when the advection is absent. One then obtains from [6] the Gaussian PDF of the displacement:

$$\mathcal{P}(\boldsymbol{R},t) = (4\pi\kappa t)^{-d/2}\mathrm{e}^{-R^2/(4\kappa t)} \qquad [7]$$

which satisfies the heat equation $(\partial_t - \kappa\nabla^2)\,\mathcal{P}(\boldsymbol{r},t)=0$. The short-correlated case is far from being an exotic exception but rather presents a long-time limit of an integral of any finite-correlated random function. Indeed, such an integral can be presented as a sum of many independent equally distributed random numbers,

the statistics of such sums is a subject of the central limit theorem. One can move beyond the central limit theorem considering the correlation time finite (yet small comparing to the time of evolution). Such generalization is the subject of the large deviation theory. Consider some quantity $X$ which is an integral of some random function over time $t$ much larger than the correlation time $\tau$. At $t \gg \tau$, $X$ behaves as a sum of many independent identically distributed random numbers $y_i$: $X = \sum_1^N y_i$ with $N \propto t/\tau$. The generating function $\langle e^{zX} \rangle$ of the moments of $X$ is the product, $\langle e^{zX} \rangle = e^{NS(z)}$, where we have denoted $\langle e^{zy} \rangle \equiv e^{S(z)}$ (assuming that the generating function $\langle e^{zy} \rangle$ exists for all complex $z$). The PDF $\mathcal{P}(X)$ is given by the inverse Laplace transform $(2\pi i)^{-1} \int e^{-zX+NS(z)} dz$ with the integral over any axis parallel to the imaginary one. For $X \propto N$, the integral is dominated by the saddle point $z_0$ such that $S'(z_0) = X/N$ and

$$\mathcal{P}(X) \propto e^{-NH(X/N - \langle y \rangle)} \qquad [8]$$

Here $H = -S(z_0) + z_0 S'(z_0)$ is the function of the variable $X/N - \langle y \rangle$; it is called entropy function as it appears also in the thermodynamic limit in statistical physics. A few important properties of $H$ (also called rate or Cramér function) may be established independently of the distribution $\mathcal{P}(y)$. It is a convex function which takes its minimum at zero, that is, for $X$ equal to the mean value $\langle X \rangle = NS'(0)$. The minimal value of $H$ vanishes since $S(0) = 0$. The entropy is quadratic around its minimum with $H''(0) = \Delta^{-1}$, where $\Delta = S''(0)$ is the variance of $y$. We thus see that the mean value $\langle X \rangle = N\langle y \rangle$ grows linearly with $N$. The fluctuations $X - \langle X \rangle$ on the scale $\mathcal{O}(N^{1/2})$ are governed by the central limit theorem that states that $(X - \langle X \rangle)/N^{1/2}$ becomes for large $N$ a Gaussian random variable with variance $\langle y^2 \rangle - \langle y \rangle^2 \equiv \Delta$ as in [7]. Finally, its fluctuations on the larger scale $\mathcal{O}(N)$ are governed by the large deviation form [8]. The possible non-Gaussianity of the $y$'s leads to a nonquadratic behavior of $H$ for (large) deviations from the mean, starting from $X - \langle X \rangle / N \simeq \Delta / S'''(0)$. Note that if $y$ is Gaussian, then $X$ is Gaussian too for any $t$, but the universal formula [8] with $H = (X - N\langle y \rangle)^2 / 2N\Delta$ is valid only for $t \gg \tau$.

## Single-Particle Diffusion

For the pure advection without noise, the displacement of the single Lagrangian trajectory is $\boldsymbol{R}(t) - \boldsymbol{R}(0) = \int_0^t \boldsymbol{V}(s)\, ds$, with $\boldsymbol{V}(t) = \boldsymbol{v}(\boldsymbol{R}(t), t)$ being the Lagrangian velocity. One can show that $\boldsymbol{V}(t)$ is statistically stationary in the frame of reference with no mean flow and under statistical homogeneity and stationarity of the incompressible Eulerian velocities. For $\kappa = 0$, the mean square displacement satisfies the equation

$$\frac{d}{dt} \langle [\boldsymbol{R}(t) - \boldsymbol{R}(0)]^2 \rangle = 2 \int_0^t \langle \boldsymbol{V}(0) \cdot \boldsymbol{V}(s) \rangle\, ds \qquad [9]$$

The behavior of the displacement is crucially dependent on the Lagrangian correlation time $\tau$ of $\boldsymbol{V}(t)$ defined by

$$\int_0^\infty \langle \boldsymbol{V}(0) \cdot \boldsymbol{V}(s) \rangle\, ds = \langle \boldsymbol{v}^2 \rangle \tau \qquad [10]$$

No general relation between the Eulerian and the Lagrangian correlation times has been established, except for the case of short-correlated velocities. For times $t \ll \tau$, the two-point function in [9] is approximately equal to $\langle \boldsymbol{V}(0)^2 \rangle = \langle \boldsymbol{v}^2 \rangle$. The fluid particle transport is then ballistic with $\langle [\boldsymbol{R}(t) - \boldsymbol{R}(0)]^2 \rangle \simeq \langle \boldsymbol{v}^2 \rangle t^2$ and the PDF $\mathcal{P}(\boldsymbol{R}, t)$ is determined by the whole single-time velocity PDF. When the correlation time of $\boldsymbol{V}(t)$ is finite (a generic situation in a turbulent flow where $\tau$ is of order of a large-scale turnover time), an effective diffusive regime is expected to arise for $t \gg \tau$ with $\langle (\boldsymbol{R}(t) - \boldsymbol{R}(0))^2 \rangle \simeq 2\langle \boldsymbol{v}^2 \rangle \tau t$. Indeed, the particle displacements over time segments much larger than $\tau$ are almost independent. At long times, the displacement $\delta \boldsymbol{R}(t)$ behaves then as a sum of many independent variables and falls into the class of stationary processes treated in the previous section. In other words, $\delta \boldsymbol{R}(t)$ for $t \gg \tau$ becomes a Brownian motion in $d$ dimensions, normally distributed with $\langle \delta R^i(t) \delta R^j(t) \rangle \simeq D_e^{ij} t$, where the so-called eddy diffusivity tensor is as follows:

$$D_e^{ij} = \frac{1}{2} \int_0^\infty \langle V_i(0) V_j(s) + V_j(0) V_i(s) \rangle\, ds \qquad [11]$$

The symmetric second-order tensor $D_e^{ij}$ is the only characteristics of the velocity which matters in this limit of $t \gg \tau$. The trace of the tensor is equal to $\langle \boldsymbol{v}^2 \rangle \tau$, that is, equal to the large-time value of the integral in [9], while its tensorial properties reflect the rotational symmetries of the advecting velocity field. If the latter is isotropic, the tensor reduces to a diagonal form characterized by a single scalar value $D_e$. The main problem of turbulent diffusion is to obtain the effective diffusivity tensor given the velocity field $\boldsymbol{v}$ and the value of the molecular diffusivity $\kappa$.

## Two-Particle Dispersion in Smooth Flows

Even when velocity $\boldsymbol{v}(\boldsymbol{R}, t)$ is a smooth function of the coordinates, Lagrangian dynamics can be quite

complicated. Indeed, $d$ ordinary differential equations $\dot{\boldsymbol{R}} = \boldsymbol{v}(\boldsymbol{R}, t)$ generally produce chaotic dynamics (for $d \geq 3$ already for steady flows and for $d = 2$ for time-dependent flows). The tools for the description of what is called chaotic advection are similar to those of the theory of dynamical chaos. The description consistently exploits two simple ideas: to single out the variables that can be represented by the sum of a large number of independent random quantities and to separate variables that fluctuate on different timescales.

The distance, $\boldsymbol{R}_{12} = \boldsymbol{R}_1 - \boldsymbol{R}_2$, between two fluid particles with trajectories $\boldsymbol{R}_i(t) = \boldsymbol{R}(t; \boldsymbol{r}_i)$ passing at $t = 0$ through points $\boldsymbol{r}_i$ satisfies the equation

$$\dot{\boldsymbol{R}}_{12} = \boldsymbol{v}(\boldsymbol{R}_1, t) - \boldsymbol{v}(\boldsymbol{R}_2, t) \qquad [12]$$

If the velocity field can be considered smooth on the scale $R_{12}$, then one expands $\boldsymbol{v}(\boldsymbol{R}_1, t) - \boldsymbol{v}(\boldsymbol{R}_2, t) = \sigma(\boldsymbol{t}, \boldsymbol{R}_1)\boldsymbol{R}_{12}$, introducing the strain matrix $\sigma$ which can be treated as independent of $\boldsymbol{R}_{12}$. The distance thus satisfies locally a linear system of ordinary differential equations (we omit subscripts replacing $\boldsymbol{R}_{12}$ by $\boldsymbol{R}$)

$$\dot{\boldsymbol{R}}(t) = \sigma(t)\boldsymbol{R}(t) \qquad [13]$$

This equation, with the strain treated as given and $R(0) = r$, may be explicitly solved for arbitrary $\sigma(t)$ only in the 1D case

$$\ln[R(t)/r] = \ln W(t) = \int_0^t \sigma(s)\, \mathrm{d}s \equiv X \qquad [14]$$

When $t$ is much larger than the correlation time $\tau$ of the strain, the variable $X$ is a sum of $N$ independent equally distributed random numbers with $N = t/\tau$ and one can apply [8]. In the multidimensional case, to use the large deviation theory, one introduces the evolution matrix $W$ such that $\boldsymbol{R}(t) = W(t)\boldsymbol{R}(0)$. The modulus $R$ is expressed via the positive symmetric matrix $W^{\mathrm{T}}W$. In almost every realization of the strain, the matrix $t^{-1} \ln W^{\mathrm{T}}W$ stabilizes at $t \to \infty$, that is, its eigenvectors tend to $d$-fixed orthonormal eigenvectors $\boldsymbol{f}_i$. To understand that intuitively, consider some fluid volume, say a sphere, which evolves into an elongated ellipsoid at later times. As time increases, the ellipsoid is more and more elongated and it is less and less likely that the hierarchy of the ellipsoid axes will change. The limiting eigenvalues

$$\lambda_i = \lim_{t \to \infty} t^{-1} \ln |W\boldsymbol{f}_i| \qquad [15]$$

are called Lyapunov exponents. The major property of the Lyapunov exponents is that they are realization independent if the flow is ergodic (i.e., spatial and temporal averages coincide). The relation [15] states that two fluid particles separated initially by $\boldsymbol{r}$ pointing into the direction $\boldsymbol{f}_i$ will separate (or converge) asymptotically as $\exp(\lambda_i t)$. The incompressibility constraints $\det(W) = 1$ and $\sum \lambda_i = 0$ imply that a positive Lyapunov exponent will exist whenever at least one of the exponents is nonzero. Consider indeed

$$E(n) = \lim_{t \to \infty} t^{-1} \ln\langle [R(t)/r]^n \rangle \qquad [16]$$

whose derivative at the origin gives the largest Lyapunov exponent $\lambda_1$. The function $E(n)$ obviously vanishes at the origin. Furthermore, $E(-d) = 0$, that is, incompressibility and isotropy make that $\langle R^{-d} \rangle$ is time independent as $t \to \infty$. Apart from $n = 0$, $-d$, the convex function $E(n)$ cannot have other zeroes if it does not vanish identically. It follows that $\mathrm{d}E/\mathrm{d}n$ at $n = 0$, and thus $\lambda_1$, is positive. A simple way to appreciate intuitively the existence of a positive Lyapunov exponent is to consider the saddle-point 2D flow $v_x = \lambda x, v_y = -\lambda y$ with the axes randomly rotating after time interval $T$. A vector initially at the angle $\phi$ with the $x$-axis will be stretched after time $T$ if $\cos\phi \geq [1 + \exp(2\lambda T)]^{-1/2}$, that is, the measure of the stretching directions is larger than $1/2$.

A major consequence of the existence of a positive Lyapunov exponent for any random incompressible flow is the exponential growth of the interparticle distance $R(t)$. In a smooth flow, it is also possible to analyze the statistics of the set of vectors $\boldsymbol{R}(t)$ and to establish a multidimensional analog of [8]. The idea is to reduce the $d$-dimensional problem to a set of $d$ scalar problems for slowly fluctuating stretching variables excluding the fast fluctuating angular degrees of freedom. Consider the matrix $I(t) = W(t)W^{\mathrm{T}}(t)$, representing the tensor of inertia of a fluid element such as the above-mentioned ellipsoid. The matrix is obtained by averaging $R^i(t)R^j(t)d/\ell^2$ over the initial vectors of length $\ell$ and $I(0) = 1$. Introducing the variables that describe stretching as the lengths of the ellipsoid axis $e^{2\rho_1}, \ldots, e^{2\rho_d}$, one can deduce similarly to [8] the asymptotic PDF:

$$\begin{aligned} &\mathcal{P}(\rho_1, \ldots, \rho_d; t) \\ &\quad \propto \exp[-t\, H(\rho_1/t - \lambda_1, \ldots, \rho_{d-1}/t - \lambda_{d-1})] \\ &\qquad \times \theta(\rho_1 - \rho_2) \ldots \theta(\rho_{d-1} - \rho_d) \\ &\qquad \times \delta(\rho_1 + \cdots + \rho_d) \end{aligned} \qquad [17]$$

The entropy function $H$ depends on the statistics of $\sigma$. In the $\delta$-correlated case, $H$ is everywhere quadratic:

$$H(\boldsymbol{x}) \propto d^{-1} \sum_{i=1}^{d} x_i^2, \qquad \lambda_i \propto d(d - 2i + 1) \qquad [18]$$

## Two-Particle Dispersion in Nonsmooth Flows

To consider dispersion in the inertial interval of turbulence, one should assume $\delta \boldsymbol{v}(\boldsymbol{r}, t)| \propto r^{\alpha}$, where generally $\alpha < 1$. Rewriting then eqn [12] for the distance between two particles as $\dot{\boldsymbol{R}} = \delta \boldsymbol{v}(\boldsymbol{R}, t)$, we infer that $\mathrm{d}R^2/\mathrm{d}t = 2\boldsymbol{R} \cdot \delta \boldsymbol{v}(\boldsymbol{R}, t) \propto R^{1+\alpha}$. It suggests

$$R(t)^{1-\alpha} - R(0)^{1-\alpha} \propto t \qquad [19]$$

For large $t$, $R(t) \propto t^{1/(1-\alpha)}$, with the dependence of the initial separation quickly forgotten. Of course, for the random process $\boldsymbol{R}(t)$, relation [19] is of the mean-field type and should pertain (if true) to the large-time behavior of the averages ($\langle R(t)^p \rangle \propto t^{p/(1-\alpha)}$, for $p > 0$) implying their super-diffusive growth, faster than the diffusive one $\propto t^{p/2}$. The power-law scaling may be amplified to the scaling behavior of the PDF of the interparticle distance, $\mathcal{P}(R, t) = \lambda \mathcal{P}(\lambda R, \lambda^{1-\alpha} t)$. The power-law growth of the second moment, $\langle R(t)^2 \rangle \propto t^3$, is the celebrated Richardson dispersion relation, which was the first quantitative phenomenological prediction in developed turbulence. It seems to be confirmed by experimental data and the numerical simulations. It is important to remark that, even assuming the validity of the Richardson relation, it is impossible to establish general large-time properties of the PDF $\mathcal{P}(R; t)$ such as those for the single-particle PDF of the distance between two particles. This is because the correlation time of the Lagrangian velocity difference, $R/\delta v(R) \propto \langle R^2 \rangle^{1/3} \propto t$, is comparable with the total time of the process.

It is instructive to contrast the exponential growth [16] of the distance between the trajectories with the power-law growth [19]. In a smooth flow, the closer two trajectories are initially, the more time is needed to effectively separate them. In a nonsmooth turbulent flow, the trajectories separate in a finite time independent of their initial distance $R(0)$, provided that the latter is also in the inertial range. This explosive separation of trajectories results in a breakdown of the deterministic Lagrangian flow since the trajectories cannot be labeled by the initial conditions. That agrees with the fundamental theorem stating that the ordinary differential equation $\dot{\boldsymbol{R}} = \boldsymbol{v}(\boldsymbol{R}, t)$ does not have unique solution if $\boldsymbol{v}(\boldsymbol{r}, t)$ is non-Lipschitz. As shown by the example of the equation $\dot{x} = |x|^{\alpha}$ with two solutions $x = [(1-\alpha)t]^{1/(1-\alpha)}$ and $x = 0$ both starting at zero, one should expect multiple Lagrangian trajectories starting or ending at the same point for velocity fields with $\alpha < 1$. Even though the deterministic Lagrangian description breaks down, the statistical description is still possible and one can make sense of propagators like $P(\boldsymbol{r}, s; \boldsymbol{R}, t|\boldsymbol{v})$. They are expected to be weak solutions of the equation $[\partial_t - \nabla \cdot \boldsymbol{v}(\boldsymbol{R}, t)] P(\boldsymbol{r}, s; \boldsymbol{R}, t|\boldsymbol{v}) = 0$ in the nonsmooth case. According to this assumption, the Lagrangian trajectories behave stochastically already in a given velocity field and for negligible molecular diffusivity – and not only due to a random noise or to random fluctuations of the velocities.

The general conjecture about the existence and diffuse nature of propagators is known to be true for the Gaussian ensemble of velocities decorrelated in time (Kraichnan 1968):

$$\langle v_i(\boldsymbol{r}, t) v_j(\boldsymbol{r}', t') \rangle = 2\delta(t - t') D_{ij}(\boldsymbol{r} - \boldsymbol{r}') \qquad [20]$$

Here the Lagrangian velocity $\boldsymbol{v}(\boldsymbol{R}, t)$ has the same white noise temporal statistics as the Eulerian velocity $\boldsymbol{v}(\boldsymbol{r}, t)$ for fixed $\boldsymbol{r}$ and the displacement along a Lagrangian trajectory $\boldsymbol{R}(t) - \boldsymbol{R}(0)$ is a Brownian motion for all times. To model nonsmooth velocity field of turbulence, we choose $D^{ij}(\boldsymbol{r}) = D_0 \delta^{ij} - (1/2) d^{ij}(\boldsymbol{r})$ and

$$d^{ij}(\boldsymbol{r}) = D_1 [(d - 1 + \xi) \delta^{ij} r^{\xi} - \xi r^i r^j r^{\xi - 2}] \qquad [21]$$

Here $D_0$ gives the eddy diffusivity of a single fluid particle (discussed earlier), whereas $d_{ij}(\boldsymbol{r})$ describes the statistics of the velocity differences. For $0 < \xi < 2$, the Kraichnan ensemble is supported on the velocities that are Hölder continuous in space with a fixed exponent $\alpha$ arbitrarily close to $\xi/2$. It mimics this way the main property of turbulent velocities. The rough (distributional) behavior of Kraichnan velocities in time, although not very physical, is not expected to modify essentially the qualitative properties of propagators (it is the spatial regularity, not the temporal one, of a vector field that is crucial for the uniqueness of its trajectories).

In exactly the same way as one derives [6] and [7] from [4], one gets $\mathcal{P}(\boldsymbol{R}, t) = |\hat{\beta}|^{1/2} (4\pi t)^{-d/2} \mathrm{e}^{-\beta_{ij} R_i R_j / 4t}$, where $(\hat{\beta}^{-1})_{ij} = D_{ij}(0) + \kappa \delta_{ij}$. In much the same way one can examine the two-particle PDF. The PDF $\mathcal{P}_2(r, s; R, t)$ of the distance $R$ between two particles satisfies the equation

$$(\partial_t - M_2) \mathcal{P}_2(r, s; R, t) = \delta(t - s) \delta(r - R) \qquad [22]$$

where $M_2 = -D_1 (d-1) r^{1-d} \partial_r r^{d-1+\xi} \partial_r$ and [22] can be readily solved:

$$\lim_{r \to 0} \mathcal{P}_2(r, s; R, t) \propto \frac{R^{d-1}}{|t - s|^{d/(2-\xi)}} \times \exp\left[-\mathrm{const.} \frac{R^{2-\xi}}{|t - s|}\right] \qquad [23]$$

That confirms the diffusive character of the limiting process describing the Lagrangian trajectories in fixed non-Lipschitz velocities: the endpoints of the process stay at finite distance when the initial points converge. The PDF [23] changes from Gaussian to log–normal when $\xi$ changes from 0 to 2. The Richardson dispersion $\langle R^2(t)\rangle \propto t^3$ is reproduced for $\xi = 4/3$.

## Multiparticle Propagators

In studying multiparticle statistics, an important question is what memory of the initial configuration remains when final distances far exceed initial ones. To answer this question, one must analyze the conservation laws of turbulent diffusion. Many-particle evolution in nonsmooth velocities exhibits nontrivial statistical integrals of motion (martingales) that are proportional to the positive powers of the distances. The integrals involve geometry in such a way that the distance growth is balanced by the decrease of the shape fluctuations. The existence of multiparticle conservation laws indicates the presence of a long-time memory and is a reflection of the coupling among the particles due to the simple fact that they are all in the same velocity field. The conserved quantities may be easily built for the limiting cases. Already for a smooth velocity, the $d$-volume $\epsilon_{i_1 i_2 \ldots i_d} R_{12}^{i_1} \ldots R_{1d}^{i_d}$ is indeed preserved for $(d+1)$ Lagrangian trajectories. In the opposite case of a very irregular velocity, the fluid particles undergo a Brownian motion. The distances between the Brownian particles grow according to $\langle R_{nm}^2(t)\rangle = R_{nm}^2(0) + Dt$. The statistical integrals of motion are $\langle R_{nm}^2 - R_{pr}^2\rangle$, $\langle 2(d+2)R_{nm}^2 R_{pr}^2 - d(R_{nm}^4 + R_{pr}^4)\rangle$, and an infinity of similarly built harmonic polynomials (zero modes of Laplacian).

The statistics of the relative motion of $N$ particles is described by the joint PDF averaged over rigid translations: $\mathcal{P}_N^{\mathrm{rel}}(\underline{\boldsymbol{r}}, s; \underline{\boldsymbol{R}}, t) = \int \mathcal{P}_N(s, \underline{\boldsymbol{r}}; \underline{\boldsymbol{R}} + \underline{\rho}, t)\, \mathrm{d}\rho$. For smooth velocities,

$$\mathcal{P}_N^{\mathrm{rel}}(\underline{\boldsymbol{r}}, 0; \underline{\boldsymbol{R}}, t) = \int \left\langle \prod_{n=1}^{N} \delta(\boldsymbol{R}_n + \rho - W(t)\boldsymbol{r}_n) \right\rangle \mathrm{d}\rho \qquad [24]$$

Such PDF depends only on the statistics of the evolution matrix $W(t)$ discussed earlier. Under the evolution governed by $W(t)$, all distances between points grow exponentially for large times while their ratios $R_{nm}/R_{kl}$ tend to a constant. For whatever initial positions, asymptotically in time, the points tend to be situated on the line. Note that the existence of deterministic trajectories leads to the collapse property $\lim_{\boldsymbol{r}_N \to \boldsymbol{r}_{N-1}} \mathcal{P}_N^{\mathrm{rel}}(\underline{\boldsymbol{r}}; \underline{\boldsymbol{R}}; t) = \mathcal{P}_{N-1}^{\mathrm{rel}}(\underline{\boldsymbol{r}}'; \underline{\boldsymbol{R}}'; t)\, \delta(\boldsymbol{R}_{N-1} - \boldsymbol{R}_N)$, where $\underline{\boldsymbol{R}}' = (\boldsymbol{R}_1, \ldots, \boldsymbol{R}_{N-1})$.

The long-time asymptotics of the propagators in the nonsmooth case can be found explicitly for the Kraichnan ensemble of velocities:

$$(\partial_t + M_N)\mathcal{P}_N^{\mathrm{rel}}(\underline{\boldsymbol{r}}, s; \underline{\boldsymbol{R}}, t) = \delta(t-s)\delta(\underline{\boldsymbol{R}} - \underline{\boldsymbol{r}}) \qquad [25]$$

$$M_N = \sum_{n<m} d^{ij}(\boldsymbol{r}_{nm}) \nabla_{r_n^i} \nabla_{r_m^j} \qquad [26]$$

When initial points get close or final points far apart and time gets large, the multiparticle PDF is factorized:

$$\lim_{\lambda \to 0} \mathcal{P}_N^{\mathrm{rel}}(\lambda \underline{\boldsymbol{r}}, 0; \underline{\boldsymbol{R}}, t) = \sum_{\beta} \lambda^{\zeta_\beta} f_\beta(\underline{\boldsymbol{r}}) g_\beta(\underline{\boldsymbol{R}}, t) \qquad [27]$$

where $f_\beta$ must be taken as zero modes of $M_N^\dagger$ and its powers while $\partial_t g_\beta = -M_N g_\beta$. The remarkable feature of the zero modes of $M_N^\dagger$ is that they are conserved in mean by the Lagrangian evolution:

$$\partial_t \langle f(\underline{\boldsymbol{R}}(t))\rangle = \int f(\underline{\boldsymbol{R}}) M_N \mathcal{P}_N^{\mathrm{rel}}(\underline{\boldsymbol{r}}, 0; \underline{\boldsymbol{R}}, t)\, \mathrm{d}\underline{\boldsymbol{R}}'$$
$$= \int \mathcal{P}_N^{\mathrm{rel}}(\underline{\boldsymbol{r}}, 0; \underline{\boldsymbol{R}}, t) M_N^\dagger f(\underline{\boldsymbol{R}})\, \mathrm{d}\underline{\boldsymbol{R}}' = 0$$

The scaling exponents of the zero modes depend, in a nontrivial way, on the number of particles $N$. For $\xi \ll 1$ and $d \gg 1$, one finds

$$\zeta_N = \frac{N}{2}(2-\xi) - \frac{N(N-2)}{2(d+2)}\xi \qquad [28]$$

## Passive Scalar

For practical applications, for example, in the diffusion of pollution, the most relevant quantity is the average $\langle \theta(\boldsymbol{r}, t)\rangle$ which can be expressed via the single-particle propagator. As discussed earlier, for times longer than the Lagrangian correlation time, the particle diffuses and $\langle \theta \rangle$ obeys the effective heat equation

$$\partial_t \langle \theta(\boldsymbol{r}, t)\rangle = \left(D_{\mathrm{e}}^{ij} + \kappa \delta_{ij}\right) \nabla_i \nabla_j \langle \theta(\boldsymbol{r}, t)\rangle \qquad [29]$$

with the eddy diffusivity $D_{\mathrm{e}}^{ij}$ given by [11]. The simplest decay problem is that of a uniform scalar spot of size $L$ released in the fluid. Its averaged spatial distribution at later times is given by the solution of [11] with the appropriate initial condition. On the other hand, the decay of the scalar in the spot is governed by the multipoint Lagrangian propagators. Taking the point of measurement inside the spot, consider the single-point moment $\langle \theta^N \rangle(t)$ described by [2]. If there is no molecular diffusion and the trajectories are unique (spatially smooth velocity), particles that end at the same

point remained together throughout the evolution and all the moments are preserved. On the contrary, when velocity is nonsmooth and the propagator is diffusive, we expect the decay even at the limit $\kappa \to 0$. This is an example of the so-called dissipative anomaly: the symmetry $t \to -t$ remains broken even when the symmetry-breaking factor $\kappa$ goes to zero. Consider a spherical spot of $\theta$ released in a spatially smooth incompressible 3D flow with $\lambda_1 > \lambda_2 > 0 > \lambda_3$. During the time less than $t_d = |\lambda_3|^{-1} \ln(L/r_d)$, diffusion is unimportant and $\theta$ inside the spot does not change. At larger time, the dimensions of the spot with negative Lyapunov exponents are frozen at $r_d$, while the rest keep growing exponentially, resulting in an exponential growth of the total volume $\exp(\rho_1 + \rho_2)$. That leads to an exponential decay of scalar moments averaged over velocity statistics: $\langle [\theta(t)]^N \rangle \propto \exp(-\gamma_N t)$. The decay rates $\gamma_N$ can be expressed via the PDF [18] of stretching variables $\rho_i$. Since $\theta$ decays as the inverse volume,

$$\langle [\theta(t)]^N \rangle \propto \int d\rho_1 d\rho_2 \exp[-tH(\rho_1/t - \lambda_1, \rho_2/t - \lambda_2) - N(\rho_1 + \rho_2)] \qquad [30]$$

At large $t$, the integral is determined by the saddle point. At small $N$, the saddle point lies within the parabolic domain of $H$ so $\gamma_N$ increases with $N$ quadratically. At large $N$, the main contribution is due to the realization with smallest possible spot of size $L$ so $\gamma_N$ saturates.

For the decay in incompressible nonsmooth flow, using the Kraichnan model one gets

$$\langle \theta^{2n}(t) \rangle = \int \mathcal{P}_{2n}(\underline{0}; \underline{R}; -1)\, C_{2n}\left(t^{1/(2-\xi)}\underline{R}, 0\right) d\underline{R} \qquad [31]$$

When $J_0 = \int C_2(\boldsymbol{r}, t)\, d\boldsymbol{r} \neq 0$, the function $t^{d/(2-\xi)}$ $C_2(t^{1/(2-\xi)}\boldsymbol{r}, 0)$ tends to $J_0\, \delta(\boldsymbol{r})$ in the long-time limit and [31] is reduced to

$$\langle \theta^{2n}(t) \rangle \approx (2n-1)!!\, J_0^n t^{nd/(\xi-2)} \times \int \mathcal{P}_{2n}(\underline{0}; \boldsymbol{R}_1, \boldsymbol{R}_1, \ldots, \boldsymbol{R}_n, \boldsymbol{R}_n; -1)\, d\underline{R} \qquad [32]$$

The decay is self-similar: $\mathcal{P}(t, \theta) = t^{d/2(2-\xi)}$ $Q(t^{d/2(2-\xi)}\theta)$. That means that the PDF of $\theta/\sqrt{\bar{\epsilon}}$ is asymptotically time independent, with $\bar{\epsilon}(t) = \kappa \langle (\nabla\theta)^2 \rangle$ being time-dependent (decreasing) dissipation rate. This should be contrasted with the lack of self-similarity for the smooth case.

One can also consider steady state of $\theta$ pumped by a source $\phi(\boldsymbol{r}, t)$:

$$\partial_t \theta + (\boldsymbol{v} \cdot \nabla)\theta + \kappa \Delta \theta = \phi \qquad [33]$$

Assuming that pumping is white Gaussian with a zero mean and variance $\overline{\phi(\boldsymbol{r}_1, t_1)\phi(\boldsymbol{r}_2, t_2)} = \chi(\boldsymbol{r}_{12})\delta(t_2 - t_1), \boldsymbol{r}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_j$, one can express the correlation functions via the multiparticle propagators. For example, assuming zero conditions at the distant past and space homogeneity, one gets

$$C_2(\boldsymbol{r}, t) = \int_{-\infty}^{t} dt' \int P(\boldsymbol{R}, \boldsymbol{r}, t')\chi(\boldsymbol{R})\, d\boldsymbol{R} \qquad [34]$$

The function $\chi(\boldsymbol{R})$ is nonzero within the correlation scale $L$ of the pumping which restricts integration to $R(t) < L$. For smooth velocity, this gives $F_2(r) = |\lambda_3|^{-1}\chi(0)\ln(L/r)$ at $r < L$. For nonsmooth velocity, the statistics of scalar fluctuations at small scales is described by the set of structure functions $S_N(r) \equiv \langle [\theta(r) - \theta(0)]^N \rangle \propto r^{\zeta_N}$ with the scaling exponents determined by the zero modes (see Falkovich *et al.* (2001)). Therefore, existence of Lagrangian statistical invariants explains the anomalous scaling of passive scalar (here, anomaly means that scale invariance broken by pumping is not restored even when the pumping scale goes to infinity).

*See also*: Anomalies; Intermittency in Turbulence; Large Deviations in Equilibrium Statistical Mechanics; Lyapunov Exponents and Strange Attractors; Random Walks in Random Environments; Stochastic Differential Equations; Turbulence Theories.

## Further Reading

Ellis R (1995) *Entropy, Large Deviations, and Statistical Mechanics*. Berlin: Springer.

Falkovich G, Gawedzki K, and Vergassola M (2001) Particles and fields in fluid turbulence. *Reviews of Modern Physics* 73: 913–975.

Kraichnan RH (1968) Small-scale structure of a scalar field convected by turbulence. *Physics of Fluids* 11: 945–963.

Majda A and Kramer PR (1999) Simplified models for turbulent diffusion: theory, numerical modelling, and physical phenomena. *Physics Reports* 314: 237–574.

Monin A and Yaglom A (1979) *Statistical Fluid Mechanics*. Boston: MIT Press.

Pope SB (1994) Lagrangian PDF methods for turbulent flows. *Annual Review of Fluid Mechanics* 26: 23–63.

Zinn-Justin J (1989) *Quantum Field Theory and Critical Phenomena*. Oxford: Science Publishing.

# Large Deviations in Equilibrium Statistical Mechanics

**S Shlosman**, Université de Marseille, Marseille, France



## Introduction

Large deviation theory (LDT) deals with the study of probabilities of extremely rare events. As an example, consider the case of independent identically distributed random variables $\sigma_1, \ldots, \sigma_N$ with the mean value $\mathbb{E}(\sigma_i) = m$. Then the typical deviations of the sum $M_N = \sigma_1 + \cdots + \sigma_N$ from its mean value $Nm$ are of the order of $\sqrt{N}$, while in LDT we study the probabilities of the deviations which are linear in $N$. In "good" cases we know that for $b > 0$

$$\Pr\{M_N - Nm \geq bN\} \sim \exp\{-I(b)N\} \quad \text{as } N \to \infty \qquad [1]$$

where $I(\cdot) > 0$ is the "rate" function.

Questions of LDT are very natural in statistical mechanics, and they have deep physical meaning, notwithstanding the fact that the corresponding events are rare. One reason is that (some) rare events in the grand canonical ensemble become typical events in the canonical ensemble.

An interesting feature of LDT in statistical mechanics is that the behavior [1] of LD is not universal, and sometimes is replaced by a nonclassical one:

$$\Pr\{M_N - Nm \geq bN\} \sim \exp\{-\tilde{I}(b)N^{\nu}\} \qquad [2]$$

with $\nu < 1$. That usually happens in the "phase transition" regime, and then the quantity $\tilde{I}(b)$, as well as the exponent $\nu$, have very much to do with the geometry of a droplet of one phase formed inside the other.

Below, we will illustrate all these features on the example of the Ising model.

## The Ising Model in the Finite Box

Our random variables $\sigma_x$ will take values $\pm 1$, with $x \in \mathbb{Z}^d$. They are called spins. For every finite box $\Lambda \subset \mathbb{Z}^d$, we will define Gibbs states in $\Lambda$. To do this we need the Hamiltonians

$$H_{\Lambda,\xi}(\sigma) = -\sum_{\substack{x,y \text{ n.n.} \\ x,y \in \Lambda}} \sigma_x \sigma_y - \sum_{\substack{x,y \text{ n.n.} \\ x \in \Lambda, y \notin \Lambda}} \sigma_x \xi_y$$

Here, $\xi$ is some spin configuration on $\mathbb{Z}^d$, which is called "boundary condition," while $\sigma \in \Omega_\Lambda$ is any spin configuration in $\Lambda$.

The "grand canonical Gibbs measure" $\mu_{\Lambda,\xi,T}$ in $\Lambda$ with boundary condition $\xi$ at inverse temperature $\beta = T^{-1}$ is given by

$$\mu_{\Lambda,\xi,T}(\sigma) = Z^{-1}_{\Lambda,\xi,T} \exp(-\beta H_{\Lambda,\xi}(\sigma)) \qquad [3]$$

where

$$Z_{\Lambda,\xi,T} = \sum_{\sigma \in \Omega_\Lambda} \exp(-\beta H_{\Lambda,\xi}(\sigma))$$

is called "partition function"; it makes the measure [3] to be a probability distribution.

The boundary condition $\xi \equiv +1(-1)$ will be denoted by $+(-)$. For every value of $T$, the Gibbs measures $\mu_{\Lambda(l),\pm,T}$ with $(\pm)$-boundary condition in the cubic box $\Lambda(l)$ of size $l$ converge, as $l \to \infty$, to the probability measures that we will denote by $\mu_{\pm,T}$. If the two happen to be different, then $\mu_{+,T}$ is called the (+)-phase, and $\mu_{-,T}$ the (–)-phase. That happens to be the case iff the temperature $T$ is lower than the critical temperature $T_c = T_c(d)$. The critical temperature depends on dimension; $T_c(1) = 0$, while $T_c(d) > 0$ for $d \geq 2$. The expectation

$$\mathbb{E}_{\mu_{+,T}}(\sigma_0) \equiv m(\beta)$$

is called spontaneous magnetization; $m(\beta) > 0$ iff $\beta > T_c^{-1}$.

## LD Properties of the Gibbs States $\mu_{\Lambda(l),-,T}$

In what follows, we will discuss the LD properties of the sum $M_\Lambda = \sigma_1 + \cdots + \sigma_{|\Lambda|}$, where the spins $\sigma_x$, $x \in \Lambda$, are distributed according to the Gibbs state $\mu_{\Lambda,-,T}$. Note that $\mathbb{E}_{\mu_{-,T}}(\sigma_0) = -m(\beta)$.

### Classical Case

If we look on the LDs of the sum $M_\Lambda$ when the temperature $T$ is high enough (in which case the limiting states $\mu_{+,T}$ and $\mu_{-,T}$ coincide), or else if the temperature is low, and the deviations are negative – that is, we consider the events $M_\Lambda + |\Lambda| m(T^{-1}) \leq b|\Lambda|$ with $b < 0$ – then their probabilities behave classically:

There exists a (high) temperature $T_0$ such that if $T > T_0$, then

$$\lim_{\Lambda \to \mathbb{Z}^d} \frac{1}{|\Lambda|} \Pr\{M_\Lambda + |\Lambda| m(T^{-1}) \leq b|\Lambda|\} = -I_T(b) \quad \text{for} \quad b \leq 0 \qquad [4]$$

$$\lim_{\Lambda \to \mathbb{Z}^d} \frac{1}{|\Lambda|} \Pr\{M_\Lambda + |\Lambda| m(T^{-1}) \geq b|\Lambda|\} = -I_T(b) \quad \text{for} \quad b \geq 0 \qquad [5]$$

where the function $I_T(b) \geq 0$ is strictly concave on the segment $(m(T^{-1}) - 1, m(T^{-1}) + 1)$. It vanishes at only one point $b = 0$.

There exists a (low) temperature $T_1$ such that if $T < T_1$, then the relation [4] holds with the function $I_T(b) > 0$ strictly concave on the segment $(m(T^{-1}) - 1, 0)$. The limit [5] also does exist, but it can vanish once we are in the phase transition region. In order to see some nontrivial behavior, we have to change the normalization $1/|\Lambda|$ in [5].

### Nonclassical Case

The proper normalization happens to be the surface term, $1/|\Lambda|^{(d-1)/d}$:

There exists a temperature $T_1$ such that if $T < T_1$, then

$$\lim_{\Lambda\to\mathbb{Z}^d} \frac{1}{|\Lambda|^{(d-1)/d}} \Pr\{M_\Lambda + |\Lambda| m(T^{-1}) \geq b|\Lambda|\} = -\mathcal{W}_T(b) \quad \text{for} \quad b > 0 \qquad [6]$$

The function $\mathcal{W}_T(b)$ obeys $\mathcal{W}_T(b) = b^{(d-1)/d} w_T$, with $w_T > 0$, provided the value $b > 0$ is not too large: $b \leq b(d)$, where $b(d)$ is some constant, depending on the dimension and temperature; one can show that $b(d) \geq 1/2^d$. For larger $b$'s the dependence is more complex.

The key object here is the constant $w_T$. To obtain it, one has to solve the following variational problem. Let $\tau_T(\theta)$, $\theta \in S^{d-1}$ be the surface tension between the (+)-phase and the (–)-phase of the Ising model at the temperature $T$. Then, for every closed compact (hyper)surface $M^{d-1} \subset \mathbb{R}^d$, we define its surface energy as

$$\mathcal{W}_T(M) = \int_M \tau_T(\theta_s)\,\mathrm{d}s$$

where $\theta_s$ is the normal vector to $M$ at $s \in M$. The functional $\mathcal{W}_T(M)$ has the meaning of the energy of the $M$-shaped droplet of the (+)-phase floating in the (–)-phase. It is called the "Wulff functional." Let $\mathfrak{W}_T$ be the surface which minimizes $\mathcal{W}_T(\cdot)$ over all the surfaces enclosing the unit volume. Such a minimizer does exist and is unique up to translation. It is called the "Wulff shape." The value $w_T$ is just the surface energy of the Wulff shape:

$$w_T = \mathcal{W}_T(\mathfrak{W}_T)$$

The value $b(d)$ is defined as the maximal value of $b$'s, for which the dilatation $b^{1/d}\mathfrak{W}_T$ can fit into the unit cube. For higher values of $b$, the shape of the (+)-phase droplet in the cube with (–)-boundary condition is deformed by its walls, so its surface energy is given by a more complicated variational problem.

## Moderate Deviations and the Droplet Condensation

The reason behind the different order of the probabilities of the events $M_\Lambda + |\Lambda| m(T^{-1}) \leq b|\Lambda|$, $b < 0$, and $M_\Lambda + |\Lambda| m(T^{-1}) \geq b|\Lambda|$, $b > 0$, at low temperatures is the following. A typical configuration contributing to the first event contains many small droplets of (–)-spins, of size $\leq \ln|\Lambda|$, floating in the sea of (+)-spins. On the contrary, in the case of the second event a typical configuration contains, in addition to small droplets, one large droplet of the size of $\Lambda$. It has a random shape, but in the limit $\Lambda \to \mathbb{Z}^d$ that shape converges to a nonrandom one, which happens to be the Wulff shape $\mathfrak{W}_T$. (The precise meaning of that statement depends on dimension; in case $d = 2$ the convergence holds in the Hausdorff metrics, while in higher dimensions it is known only in $L^1$ sense.) That statement makes the following question natural: consider the event

$$M_\Lambda - \mathbb{E}(M_\Lambda) \geq |\Lambda|^\alpha, \quad 0 < \alpha < 1$$

For which $\alpha$ should we expect, in addition to microscopic (+)-droplets of size $\leq \ln|\Lambda|$, the formation of a large droplet, of volume $\sim |\Lambda|^\alpha$, in a corresponding typical configuration? In other words, how many extra (+)-spins should we pump into our systems in order for the microscopic droplets to condense into a macroscopic one? (In the formulation of this question, we have to use the expectation $\mathbb{E}(M_\Lambda)$ instead of the asymptotically equivalent quantity $-|\Lambda| m(T^{-1})$. The difference, $\mathbb{E}(M_\Lambda) + |\Lambda| m(T^{-1}) \sim O(|\partial\Lambda|)$, being irrelevant in the LD case, becomes significant here.)

The answer is the following:

- if $\alpha < d/(d+1)$, then a typical configuration contains only microscopic droplets;
- if $\alpha > d/(d+1)$, then any typical configuration contains, in addition to microscopic droplets, one large droplet of volume $\sim |\Lambda|^\alpha$.

Therefore, the condensation happens at the value $\alpha = d/(d+1)$. This picture has its counterpart in the behavior of the probabilities of "moderate deviations" (MD), that is, events when $M_\Lambda + |\Lambda| m(T^{-1}) \geq |\Lambda|^\alpha$:

- if $\alpha < d/(d+1)$, then the deviation is due to independent fluctuations of sizes of many small droplets, and the usual Gaussian behavior holds:

$$\Pr\{M_\Lambda - \mathbb{E}(M_\Lambda) \geq |\Lambda|^\alpha\} \sim \exp\left\{-\frac{(|\Lambda|^\alpha)^2}{2\mathrm{Var}(M_\Lambda)}\right\} = \exp\left\{-c|\Lambda|^{2\alpha-1}\right\}$$

- if $\alpha > d/(d+1)$, then the deviation is due to the formation of a large droplet, and so

$$\Pr\{M_\Lambda - \mathbb{E}(M_\Lambda) \geq |\Lambda|^\alpha\} \sim \exp\left\{-c'|\Lambda|^{\alpha((d-1)/d)}\right\}$$

Note that the two estimates match at $\alpha = d/(d+1)$.

## Other Questions

There are many related questions; some are partially solved, others are widely open, if considered on a rigorous mathematical level.

One can ask about the asymptotic behavior of probabilities of the events like

$$M_\Lambda - \mathbb{E}(M_\Lambda) = b_\Lambda$$

where the values $b_\Lambda$ lie in the LD or MD region. The difference between such questions and those treated above is of the same nature as the difference between the integral and the local limit theorems. Partial answers to them are given in Dobrushin and Shlosman (1994).

Many results about the Wulff shape and its relation to the Ising model are known, starting by Dobrushin *et al.* (1992). Some are still challenging. One such question concerns the so-called roughening phase transition. It is known rigorously that the Wulff shape $\mathfrak{W}_T$ in the $d \geq 3$ Ising model has flat facets at low temperatures $T$. It is believed that such a feature holds true only for $T < T_R$, where the roughening temperature $T_R$ is strictly less than the critical temperature $T_c(d)$ for $d = 3$. At the temperatures $T \in (T_R, T_c(3))$, the Wulff shape $\mathfrak{W}_T$ does not have facets. This conjecture seems to be very difficult.

The question about the typical behavior of the MD of the Ising model at the threshold value $M_\Lambda - \mathbb{E}(M_\Lambda) \sim |\Lambda|^{d/(d+1)}$ was recently answered in Biscup *et al.* (2003).

## Further Reading


Biscup M, Chayes L, and Kotecky R (2003) Critical region for droplet formation in the two-dimensional Ising model. *Communications in Mathematical Physics* 242: 137–183.

Dobrushin RL and Shlosman SB (1994) Large and moderate deviations in the Ising model. In: Dobrushin RL (ed.) *Probability Contributions to Statistical Mechanics*, Advances in Soviet Mathematics, vol. 18, pp. 91–220. Providence, RI: American Mathematical Society.

Dobrushin RL, Kotecky R, and Shlosman SB (1992) *Wulff Construction: A Global Shape from Local Interaction.* AMS Translations Series. Providence, RI: American Mathematical Society.


# Large-*N* and Topological Strings

**R Gopakumar**, Harish-Chandra Research Institute, Allahabad, India



## Introduction

Topological strings have been well studied since they were introduced in the early 1990s. Essentially, they are simplified string theories that capture the information about a sector of the full (or "physical") string theory. Thus, while sharing many of the structural features of usual string theory, they hold out the possibility of being amenable to explicit calculations. This is especially true with regard to stringy quantum corrections (the higher genus contributions from the point of view of the string world sheet), which are normally rather intractable in the full physical string theory. This has allowed them to play a useful role in enhancing the understanding of string theory and many of its mysterious quantum properties, such as the various dualities.

In particular, in the last several years, topological strings have served as an important laboratory for testing and understanding the connection between the large-$N$ expansion of gauge theories and closed-string theories. In this article we will sketch how this connection is illustrated in a duality between large-$N$ Chern–Simons gauge theory and closed topological string theories. We will survey the origin and current status of these developments and indicated some of its remarkable mathematical ramifications.

## Background

In order to appreciate the conjecture relating the Chern–Simons theory and topological string theories, we need to go back to the seminal work of 't Hooft, who pointed to the connection between the large-$N$ expansion of gauge field theories and string theories.

The starting point is a gauge field theory (with, say, gauge group $U(N)$), where we take the limit of the rank $N$ of the gauge group to infinity (see Brezin

and Wadia (1993) for a collection of papers on the topic). The idea is then to make an expansion in inverse powers of $N$ for various observables such as the free energy and correlation functions. For definiteness, let us take a gauge theory containing only gauge fields $A$ in the adjoint representation of U($N$). The quantum theory is (schematically) defined by the path integral

$$Z = \int [\mathcal{D}A] e^{iS(A)} \qquad [1]$$

For now, the action $S(A)$ for the gauge fields is left unspecified. It could be either the usual Yang–Mills functional or of the Chern–Simons form which we describe below. $S(A)$ is normalized in such a way that the gauge coupling constant, denoted by $\kappa$, only appears via an overall multiplicative factor of $1/\kappa$.

Then the expression, for instance, for the free energy $F = \ln Z$ has an expansion in a power series in $\kappa$, whose individual terms are given by the usual Feynman diagrammatic rules. Namely, we have is a sum over connected vacuum diagrams (those without any external legs) formed from the vertices determined by the action $S(A)$. Even without going into the details of the action, we can write down the dependence on $N$ and $\kappa$ coming from a diagram with $h$ faces, $V$ vertices, and $E$ edges. Every edge is associated with a propagator (arising from the inverse of the quadratic term in $S(A)$) and thus comes with a weight of $\kappa$. Every vertex, coming from the cubic and higher-order terms in $S(A)$, comes with a factor of $\kappa^{-1}$. There is a factor of $N$ coming from summing over the color indices that circulate in every loop (face). We thus get a weight of $N^h \kappa^{E-V}$ and so the total contribution to the free energy can be organized as

$$\begin{aligned} F &= \sum_{g=0,h=1}^{\infty} C_{g,h} N^h \kappa^{2g-2+h} \\ &= \sum_{g=0,h=1}^{\infty} C_{g,h} N^{2-2g} \lambda^{2g-2+h} \end{aligned} \qquad [2]$$

Here we have defined $\lambda \equiv \kappa N$, the 't Hooft coupling, as the combination that will be kept fixed when taking the limit of large $N$. We have also used the fact that $V - E + h = 2 - 2g$, where $g$ is the number of handles of the closed two-dimensional surface one can associate with the Feynman diagram. (It is best to visualize the Feynman diagram as a "fatgraph" which forms the skeleton of a closed Riemann surface.) The coefficients $C_{g,h}$ represent the sum of the contributions from all genus $g$ diagrams with $h$ boundaries and depend on the details of the theory.

We note that the reorganization of the contributions to the free energy is reminiscent of the genus expansion in a string theory. In fact, eqn [2] as it stands looks like an open-string expansion on world sheets with $g$ handles and $h$ boundaries. Indeed, in many cases the gauge theory arises as a limit of an open-string theory. (Recall that a massless nonabelian gauge boson is one of the low-lying excitations of an open-string theory.) So the double expansion in terms of $g$ and $h$ is not too surprising.

However, the interesting conjecture of 't Hooft is in the relation to closed-string theory. Note that the expansion in inverse powers of $N$ depends only on the number of handles $g$. In fact, $1/N$ seems to play the role of closed-string coupling in that it suppresses higher genus diagrams. The total contribution to a given genus $g$ comes from summing over all the holes $h$ in eqn [2], for example,

$$F = \sum_{g=0}^{\infty} N^{2-2g} F_g(\lambda) \qquad [3]$$

The conjecture is to identify this with a closed-string expansion in which $F_g(\lambda)$ is a closed-string amplitude on a genus $g$ Riemann surface. (In carrying out the sum over the holes, we have assumed the existence of a radius of convergence. This is plausible since the number of planar diagrams ($g = 0$), for instance, grows only exponentially with the number of holes.) The question, since 't Hooft, has been: what is this closed-string theory? In other words, what is the background on which the closed string propagates?

A breakthrough came from Maldacena's identification of the background for the particular case of U($N$) $\mathcal{N} = 4$ supersymmetric Yang–Mills theory. His conjecture was that this theory is dual to type IIB closed-string theory on $AdS_5 \times S^5$ with a curvature scale set by $\lambda$ and with closed-string coupling $\propto \lambda/N$. This proposal passed a number of nontrivial checks and is widely held to be true. It also stimulated the search for closed-string duals to other large-$N$ gauge theories.

In what follows, we explain how the conjecture of 't Hooft has a nice realization in the case of three-dimensional U($N$) Chern–Simons gauge theory on $S^3$. The dual closed-string theory, obtained by summing over the holes, turns out to be the A-model topological string on the (six-dimensional) resolved conifold background. The parameter $\lambda$ maps into a Kahler parameter in the closed-string

geometry and once again the closed-string coupling is $\propto \lambda/N$.

## The Large-*N* Expansion of Chern–Simons Theory

Nonabelian Chern–Simons theory is based on the following action functional for the U($N$) gauge connection $A$:

$$S_{\mathrm{CS}}(A) = \frac{k}{4\pi}\int_M \mathrm{tr}\left(A \wedge \mathrm{d}A + \tfrac{2}{3}A \wedge A \wedge A\right) \quad [4]$$

Here $M$ is a three-dimensional manifold. $k$ is called the level and is integer quantized for the path-integral equation [1] to be single valued. Note that, classically, $\kappa$ as defined earlier is proportional to $1/k$. One of the nice properties of $S_{\mathrm{CS}}(A)$ is that it is independent of the metric on $M$, unlike the Yang–Mills functional. Thus, it is a prototype of a topological field theory. In fact, the observables in this theory capture topological information about the 3-manifold $M$.

Witten succeeded in quantizing the Chern–Simons theory by relating its Hilbert space to the space of conformal blocks in the two-dimensional U($N$) WZW theory. (for more details on the quantization, *see* Chern-Simons Models: Rigorous Results). Here, merely the answers for various observables in the theory will be quoted. In particular, the free energy for the theory on $S^3$ can be written in a completely explicit form:

$$\begin{aligned} Z(S^3, N, k) &= \exp F(S^3, N, k) \\ &= \frac{1}{(N+k)^{N/2}} \prod_{j=1}^{N-1}\left(2\sin\frac{j\pi}{N+k}\right)^{N-j} \end{aligned} \quad [5]$$

One of the features one observes in the quantization is the shift ("finite renormalization") of the effective level from $k$ to $k+N$. This can also be seen in perturbation theory. Consequently, while taking the large-$N$ limit, the natural quantity to be held fixed as the 't Hooft coupling is $\lambda = 2\pi N/(k+N)$.

We can then carry out the 't Hooft expansion in powers of $\lambda$ and $1/N$, of expressions, for example, for the free energy in eqn [5]:

$$\begin{aligned} F = {} & \frac{N^2}{2}\left(\log\lambda - \frac{3}{2}\right) - \frac{1}{12}\log N + \zeta'(-1) \\ & + \sum_{g=2}^{\infty}\frac{1}{N^{2g-2}}\frac{B_{2g}}{2g(2g-2)} \\ & + \sum_{g=0}^{\infty}\sum_{h=2}^{\infty} F_{g,h}\lambda^{2g-2+h} \end{aligned} \quad [6]$$

The coefficents $F_{g,h}$ are nonzero only for even $h$ and are given by

$$\begin{aligned} F_{0,h} &= -\frac{2\zeta(h-2)}{(2\pi)^{h-2}(h-2)h(h-1)} \\ F_{1,h} &= \frac{\zeta(h)}{6(2\pi)^h h} \\ F_{g,h} &= \frac{2\zeta(2g-2+h)}{(2\pi)^{2g-2+h}}\binom{2g-3+h}{h} \\ &\quad \times \frac{B_{2g}}{2g(2g-2)} \end{aligned} \quad [7]$$

where the last line is for $g > 1$. $B_{2g}$ are the Bernoulli numbers. The first few terms in eqn [6] are nonperturbative contributions which do not have a Feynman-diagram interpretation. The power series in $\lambda$ is, on the other hand, of the same form as eqn [2]. In fact, there is an open-string interpretation for these terms which will be considered later.

Given the explicit form of the answer, we can carry out the summation over the holes $h$. Using some resummation techniques, we find

$$F = \sum_{g=0}^{\infty}\left(-\mathrm{i}\frac{t}{N}\right)^{2g-2} F_g(t) \quad [8]$$

with $t \equiv \mathrm{i}\lambda$ and

$$\begin{aligned} F_g(t) = {} & \frac{(-1)^g |B_{2g}B_{2g-2}|}{2g(2g-2)(2g-2)!} \\ & + \frac{|B_{2g}|}{2g(2g-2)!}\sum_{n=1}^{\infty} n^{2g-3}\mathrm{e}^{-nt} \end{aligned} \quad [9]$$

(This expression is for $g > 1$. There are very similar expressions for genus 0 and 1 as well.) With the identification of the string coupling $g_s = -\mathrm{i}t/N$, the $F_g(t)$ actually turn out to be the genus $g$ amplitudes of a closed topological string, in line with the general expectation of the previous section. This is explained in the following.

## Topological Strings

Physical strings are defined in terms of a two-dimensional sigma model (the theory on the world sheet) made reparametrization invariant by coupling to two-dimensional gravity. Topological strings are simpler versions of this, where the world-sheet theory is a two-dimensional topological sigma model. The latter is defined in terms of a sigma model (usually with $N = 2$ superconformal symmetry) with an additional twist which drastically cuts down the physical states to a subset of the low-lying modes. There are actually two inequivalent twists

denoted by $A$ and $B$, respectively, but we will restrict to the $A$ twist in this article. One of the simplifications of the $A$ twisted sigma model is that the path integral localizes to contributions from only holomorphic maps from the world sheet to the target space (which will be taken to be a Calabi–Yau 3-fold). Also, all the observables in the theory depend only on the Kahler parameters of the target space and not the complex structure parameters (*see* Topological Sigma Models as well as the book by Hori *et al.* (2003) for more details).

The topological string theory is defined by an appropriate integration of the observables of the topological sigma model over the moduli space of the world-sheet Riemann surface. For instance, the free energy of the string theory at genus $g$ is given by

$$F_g^{\text{top}}(t) = \int_{\mathcal{M}_g} < \prod_{i=1}^{6g-6} (b, \mu_i) >_X \qquad [10]$$

Here $b$ is one of the reparametrization ghost fields on the world sheet and $\mu_i$ are Beltrami differentials. The averaging is with respect to the world-sheet sigma model for the Calabi–Yau target $X$, as the subscript indicates. We have also shown the dependence of $F_g$ on the Kahler parameters of $X$, collectively denoted by $t$. The localization to the holomorphic maps in the path integral implies that $F_g^{\text{top}}(t)$ takes the generic form

$$F_g^{\text{top}}(t) = \sum_\beta N_{g,\beta} q^\beta \qquad q^\beta \equiv \prod_i q_i^{n_i} \qquad [11]$$

Here $q_i = \mathrm{e}^{-t_i}$ and $n_i$ are the integer coefficents labeling the element $\beta \in H^2(X)$. This is in the same basis of two cycles of $H^2(X)$ in terms of which the complex Kahler parameters $t_i$ are expressed. (Recall that in string theory the Kahler parameters are complexified because of the presence of an additional 2-form field.) The $N_{g,\beta}$ are the Gromov–Witten invariants for $X$ and are in general rational numbers. For nonzero $\beta$, the corresponding terms are often called world-sheet instanton contributions since they correspond to topologically nontrivial maps from the world sheet to 2-cycles in the target space. The all-genus free energy of the topological string is also defined to be

$$F^{\text{top}}(t, g_s) = \sum_{g=0}^{\infty} g_s^{2g-2} F_g^{\text{top}}(t) \qquad [12]$$

with $g_s$ being the string coupling.

Since topological strings are related to physical strings by a twist on the world sheet, it is natural that topological string computations are related to computations in the physical string theory. In fact, as shown by Antoniadis, Gava, Narain, and Taylor as well as Bershadsky, Cecotti, Ooguri, and Vafa, observables such as $F_g^{\text{top}}(t)$ are related to special superpotential terms in the type II string compactification on the Calabi–Yau $X$. Using duality to M-theory, these answers were reinterpreted by Gopakumar and Vafa in terms of contributions coming from BPS states of wrapped D-branes. This gives a completely different perspective on topological strings. For instance, the all-genus free energy can naturally be reorganized as

$$F^{\text{top}}(t, g_s) = \sum_{g=0}^{\infty} \sum_\beta \sum_{d=1}^{\infty} n_\beta^g \frac{1}{d} \left(2 \sin \frac{d g_s}{2}\right)^{2g-2} q^{d\beta} \qquad [13]$$

where the $n_\beta^g$ are integer invariants (Gopakumar–Vafa) since they count the number of BPS states. This will prove to be useful in extracting all-genus answers for topological string amplitudes, which is normally quite difficult using the perturbative definition given earlier.

## The Large-*N* Dual to Chern–Simons Theory

We are now in a position to state the duality (Gopakumar and Vafa 1999) between large-$N$ Chern–Simons theory and topological strings in a precise way. The conjecture is that the closed topological string theory on the $S^2$ resolved conifold geometry is exactly dual to the U($N$) Chern–Simons theory on $S^3$. The resolved conifold geometry is a noncompact Calabi–Yau 3-fold described by the equation

$$xy - zw = 0 \qquad [14]$$

where the singularity is resolved by a 2-sphere $x = \rho z$, $w = \rho y$. The resulting space can thus be characterized as an $O(-1) + O(-1)$ bundle over $P^1$. It has a single Kahler parameter $t$ for the nontrivial 2-cycle of the $S^2$. In addition, the string theory is characterized by the string coupling $g_s$. These parameters map on the gauge theory side to the 't Hooft parameter $\lambda$ and $N$ via the dictionary

$$t = \mathrm{i}\lambda, \qquad g_s = \frac{\lambda}{N} \qquad [15]$$

This conjecture can be checked by comparing various exact calculations in the Chern–Simons theory with corresponding calculations in the topological string on this conifold background. The use of the duality to M-theory enables us to make exact computations on this side as well. One of the

nontrivial checks of this duality comes from a comparison of the free energies. In eqns [8] and [9], we already have carried out the sum over the holes in the Chern–Simons theory and organized it as a closed-string genus expansion. Note that these expressions are already of the form [11] expected of a closed topological string. One simply has to check that it is indeed that on the $S^2$ resolved conifold.

In the language of the integer invariants $n^g_\beta$, the $S^2$ resolved conifold is particularly simple. The only nonzero invariant is $n^0_1 = 1$. Physically, this corresponds to a single brane wrapped on the genus-zero $S^2$. Putting this into eqn [13], and making the expansion in powers of $g_s$, we find exactly eqn [9] for the genus-$g$ contribution to the free energy. This is quite a remarkable agreement and represents a triumph for the ideas of large-$N$ duality.

## Geometric Transitions and Large-*N* Duality

To understand the reason for this duality a bit better, we utilize an old observation of Witten that Chern–Simons theory is an open topological string theory. As mentioned earlier, the expansion [2] (or [6]) is suggestive of an open-string expansion in terms of handles and holes. Witten observed that open topological strings on the noncompact 3-fold $T^*M$ (with Dirichlet boundary conditions on $M$ for the end points of the string) is Chern–Simons theory on $M$. In fact, in the modern language of D-branes, we would say that U($N$) Chern–Simons theory is the world-volume theory of $N$ D-branes wrapped on $M$, for the topological A-model on $T^*M$.

In particular, Chern–Simons theory on $S^3$ is the theory of branes wrapped on $S^3$ inside $T^*S^3$. The latter is the conifold geometry but now deformed by a nonzero size $S^3$. It is described by the equation

$$xy - zw = \mu \qquad [16]$$

where $\mu$ is the deformation which parametrizes the size of the $S^3$.

The above large-$N$ duality can be considered as an open–closed string duality. Namely, that the theory of open A-model topological strings on the $S^3$ resolved conifold (with $N$ D-branes) is dual to closed A-model topological strings on the $S^2$ resolved conifold. Cast in this way, we see that the duality involves a transition in the background geometry in going from the open-string to the closed-string description. The sum over the holes changes the background. The $S^3$, as it were, shrinks to zero size and a transverse $S^2$ opens up. This geometric transition makes the connection between the Chern–Simons theory and the closed topological string somewhat less mysterious. Maldacena's conjecture for super Yang–Mills involves a similar passage from D-branes in flat space to a closed-string theory on anti-de Sitter space. In fact, it appears as if the best way to understand 't Hooft's idea in generality is to think of it as an open–closed string duality.

## Further Checks and Consequences

The free energy is not the only gauge-invariant observable in Chern–Simons theory. One important class of observables, which played an important role in the connection with knot invariants, are the Wilson loop expectation values. Given a knot $\mathcal{K}$ in $S^3$, we can define, in terms of an arbitrary representation $R$ of U($N$), the trace of the holonomy around the knot averaged with respect to the Chern–Simons path-integral measure:

$$W_R(\mathcal{K}) = < \mathrm{tr}_R\left(P \exp \mathrm{i} \oint_{\mathcal{K}} A\right) > \qquad [17]$$

$P$ denotes path ordering. Similarly, we can also define the expectation values of links: products of traces of holonomies around various interlinked paths. The nonperturbative solution of Chern–Simons theory gives exact answers for the expectation values of these Wilson loops. The discussion below is, however, confined to knots.

Since the trace of holonomies is being considered in different representations, it makes sense to study the generating functional

$$\begin{aligned} Z(U, V) &= \sum_R \mathrm{tr}_R(U)\mathrm{tr}_R(V) \\ &= \exp\left[\sum_{n=1}^{\infty} \frac{1}{n} \mathrm{tr}\ U^n \mathrm{tr} V^n\right] \end{aligned} \qquad [18]$$

The source $V$ here is a U($M$) matrix, unrelated to the U($N$) holonomy $U$ around $\mathcal{K}$. The second equality in [18] follows from use of the Frobenius formula. It was shown by Ooguri and Vafa that this generating functional is the natural object from the point of view of the open–closed string duality.

We have already mentioned that the U($N$) Chern–Simons theory can be thought of as the theory of $N$ topological D-branes wrapped on the Lagrangian $S^3$ cycle inside $T^*S^3$. For a knot $\mathcal{K}$ in the $S^3$, we consider another Lagrangian 3-cycle $\hat{C}_{\mathcal{K}}$ in $T^*S^3$ which intersects the $S^3$ exactly in $\mathcal{K}$. A canonical construction for $\hat{C}_{\mathcal{K}}$ is

$$\hat{C}_{\mathcal{K}} = \left\{(q(s), p) \in T^*S^3 | \sum_i p_i \dot{q}_i = 0\right\} \qquad [19]$$

where the knot $\mathcal{K}$ is parametrized by the closed curve $q(s)$. By construction, $\hat{C}_{\mathcal{K}}$ intersects the $S^3$ in $\mathcal{K}$. Now consider $M$ D-branes wrapped on $\hat{C}_{\mathcal{K}}$. One now has to consider the fields coming from the strings stretching between the two sets of branes. One can show that integrating out these fields (which are in the bifundamental of the product group $\mathrm{U}(N) \times \mathrm{U}(M)$) modifies the original Chern–Simons action to

$$S_{\mathrm{eff}}(A) = S_{\mathrm{CS}}(A) + \sum_{n=1}^{\infty} \frac{1}{n} \mathrm{tr} U^n \mathrm{tr} V^n \qquad [20]$$

Here $V$ is the holonomy around $\mathcal{K}$ of the $\mathrm{U}(M)$ gauge field $\tilde{A}$. Thus, this configuration of $M$ probe branes gives rise exactly to the generating function eqn [18] for Wilson loops of $\mathcal{K}$.

The geometric transition which relates the Chern–Simons theory to the closed-string theory now suggests what one needs to do to compute this generating function on the closed-string side. We have to follow the configuration of the $M$ probe branes on $\hat{C}_{\mathcal{K}}$ through the conifold transition in which the $S^3$ shrinks and one blows up the $S^2$. It is not easy in general to figure out the Lagrangian cycle $C_{\mathcal{K}}$ which results from following $\hat{C}_{\mathcal{K}}$ through the transition. It has only been done in a class of knots including the simple unknot. But assuming we know $C_{\mathcal{K}}$, the generating function for Wilson loops is given by the free energy on the $S^2$ resolved conifold in the presence of $M$ probe branes on $C_{\mathcal{K}}$. This requires one to know more than the closed-string partition function computed earlier. We now also need to compute amplitudes for world sheets with boundary on $C_{\mathcal{K}}$. These are called open-string Gromov–Witten invariants and the study of this subject is in its infancy. For simple knots such as the unknot, for which $C_{\mathcal{K}}$ is known, these can be computed. One finds again a remarkable agreement with the nonperturbative answers of Chern–Simons theory. Thus, the computation of knot invariants gets related to open-string Gromov–Witten invariants. There have been a number of other tests involving more general knots and links. One also has to be careful of subtleties such as in the choice of framing. The reader is referred to the articles by Marino (2002, 2004) for these topics.

## Conclusions

The large-$N$ duality of 't Hooft is realized in Chern–Simons theory in a very explicit way. Thanks to the analytic control we have over both Chern–Simons theory as well as closed topological strings, the conjecture passes very nontrivial checks that extend to all-genus case. This is more than we can do in the AdS/CFT conjecture where most computations are at tree level in the supergravity limit. In contrast, here we see the essential stringiness of the closed-string dual to Chern–Simons theory.

Also, by viewing it as an open–closed string duality, many aspects of the correspondence were clarified. It, therefore, provides a useful toy model for a general understanding of open–closed string duality. Indeed, a proof of this duality using world sheet techniques has been proposed by Ooguri and Vafa. One would like to carry over some of the intuition that operates in this duality to the case of other physically interesting gauge theories.

From the mathematical point of view, as already indicated, this duality leads to previously unsuspected relations between Gromov–Witten invariants and invariants of 3-manifolds, including those of knots. In fact, by considering more general geometric transitions and using this duality locally, one can learn about all-genus topological string amplitudes for a wide class of noncompact toric geometries. This line of development culminated in the formulation of the topological vertex by Aganagic, Klemm, Marino, and Vafa, which captures the essence of the topological closed-string amplitudes for noncompact toric geometries. As in the case of the general correspondence between the gauge theory and gravity, this duality sheds new light on both sides of the equation. We learn to see new integrality properties in knot and 3-manifold invariants which have an interpretation in terms of enumerative problems in 3-folds. The surprises that such a deep connection presages have not yet been exhausted.

*See also:* AdS/CFT Correspondence; Chern–Simons Models: Rigorous Results; Duality in Topological Quantum Field Theory; Free Probability Theory; The Jones Polynomial; Knot Theory and Physics; Large-*N* Dualities; Quantum 3-Manifold Invariants; Schwarz-Type Topological Quantum Field Theory; String Field Theory; Topological Gravity, Two-Dimensional; Topological Quantum Field Theory: Overview.

## Further Reading

Brezin E and Wadia SR (eds.) (1993) *The Large N Expansion in Quantum Field Theory and Statistical Physics*. Singapore: World Scientific.

Gopakumar R and Vafa C (1999) On the gauge theory/geometry correspondence. *Advances in Theoretical and Mathematical Physics* 3: 1415 (arXiv:hep-th/9811131).

Hori K, Katz S, Klemm A, Pandharipande R, and Thomas R (2003) *Mirror Symmetry*. New York: AMS Publishing.

Marino M (2002) Enumerative geometry and knot invariants (arXiv:hep-th/0210145).

Marino M (2004) Chern–Simons theory and topological strings (arXiv:hep-th/0406005).

# Large-*N* Dualities

**A Grassi**, University of Pennsylvania, Philadelphia, PA, USA



## Introduction

Gopakumar and Vafa (1999) conjectured that U($N$) Chern–Simons gauge theory on $S^3$ is dual, for large values of $N$, to a closed topological string theory on a suitable Calabi–Yau 3-fold $X$. They suggested that this duality is realized by a geometric "transition," a topological surgery which can be realized by birational contractions followed by the complex deformations of Calabi–Yau varieties. Here we will give some general comments on the history of this conjecture and then present some of its mathematical implications; we will focus on the geometric transition and the novel mathematics that it has generated.

A duality relating gauge theories and string theories (with gravity) was first conjectured by 't Hooft (1974). In 1998 Maldacena conjectured a duality between Yang–Mills gauge theory with $N=4$ SUSY on a four-dimensional manifold $M$ and IIB type closed string on the anti-de Sitter space $\mathrm{AdS}^5 \times S^5$. Chern–Simons string theory is a three-dimensional theory and purely topological, hence it is in principle simpler than four-dimensional Yang–Mills theory, which also involves a metric.

In this survey, we discuss the IIA open/closed dualities: we will mostly be concerned with the partition function, that is we will be working in the context of "topological strings." The duality has been extended to a duality of strings, adding fluxes on the closed sector and branes on the open sector. There is much mathematical evidence supporting the conjecture.

## Overview

The conjecture says that U($N$) Chern–Simons gauge theory on $S^3$ is dual, for large values of $N$, to type IIA closed topological string theory on a suitable Calabi–Yau manifold $X$. A starting point for the geometry, and its mathematical implications, is that $S^3$ can be thought of as a vanishing cycle in a local Calabi–Yau manifold $Y = T^*S^3$, which deforms to a singular Calabi–Yau $Y_0$; $X$ is a Calabi–Yau birational resolution of $Y_0$. $X$ are $Y$ are related by a geometric transition. In fact, Witten showed that quantum Chern–Simons theory on $S^3$ can be thought of as open IIA (with U($N$) branes) on $Y = T^*S^3$; thus, a more general conjecture says, loosely speaking, that open IIA theory on a Calabi–Yau manifold $Y$ is dual, for large $N$, to closed IIA on a Calabi–Yau $X$ which is related to $Y$ via a geometric transition. A consequence of a physics "duality" is a matching of the free energies of the dual theories. In this particular case, if the conjecture is true, the Chern–Simons free energy $Z(S^3, \mathrm{U}(N))$ should determine, and be determined by, the closed prepotential $\mathcal{F}_{\mathrm{cl}}(X,t)$. Note that $Z(S^3, \mathrm{U}(N))$ is purely topological, and that $\mathcal{F}_{\mathrm{cl}}(X,t)$ includes all genera, as we will discuss later. A mathematical application is computing Gromov–Witten invariants for higher genus via large-$N$ dualities (Mariño 2004). Another consequence involves the matching of the observable in $S^3$ and $X$.

This conjecture is now supported by a vast amount of evidence. Vafa, Gopakumar and Ooguri noted, via a string-theory analysis, that topological and knot invariants of $S^3$ (computed through U($N$) Chern–Simons theory on $S^3$) determine and are determined by, for large $N$, the Gromov–Witten invariants of $X$ in a neighborhood of the exceptional locus of the birational contraction $X \to Y_0$.

The extension to the full string theory would say that open string of type IIA compactified on a Calabi–Yau manifold $Y$ with branes is conjectured to be dual to closed string of type IIA compactified on a Calabi–Yau manifold $X$ with fluxes, if $X$ and $Y$ are related by a geometric transition.

A mathematical consequence of this statement is that the closed Gromov–Witten invariants of $X$ agree, with a suitable identification of the parameters, with combinations of open Gromov–Witten invariants and knot invariants of $Y$. This has been shown to hold for some classes of examples.

This circle of ideas has stimulated much work in physics and mathematics on the nature of the mathematical correspondence behind this duality, as well as the property of the enumerative and topological invariants involved. The "mirrors" of the above transitions have been studied in a series of papers, starting with the work of Dijkgraaf and Vafa (2002).

The mathematics behind the open/closed dualities is still not understood: it is reasonable to speculate that the natural setup is a framework of symplectic field theory.

We shall start by discussing the principal topics of this large-$N$ duality: Chern–Simons quantum field theory, IIA closed prepotential (and Gromov–Witten invariants), and Chern–Simons as open string (and IIA open prepotential). Next we shall study the geometric transitions and conclude with some mathematical predictions of the duality.

We shall not discuss some other interesting implications of this duality. For example, we shall not discuss its mirror IIB duality: it is known that the part of the closed prepotential in IIA corresponding to rational curves can be expressed as its IIB mirror dual with periods over certain suitable cycles; the IIA open contribution corresponding to open discs is expressed in terms of integrals over chains and the Abel–Jacobi map. We only remark that this large-$N$ duality has also been interpreted as a duality between seven-dimensional manifolds with $G_2$ holonomy.

## History

The chronology of various important contributions in the field of large-$N$ duality is as follows:

- 1976: 't Hooft's conjecture
- 1988: Clemens introduces transitions
- 1988: Witten introduces quantum Chern–Simons theory on 3-manifolds
- 1992: Witten discusses Chern–Simons theory as open string
- 1998: Gopakumar–Vafa–Ooguri
- 2001: Verification for unknot, Katz–Liu, Li, and Song
- 2001: Lift to manifolds with $G_2$ holonomy
- 2002: The conjecture verified for many examples of conifold transitions, including compact case; the topological vertex is introduced
- 2003: Relations with Donaldson–Thomas invariants

## Background

The varieties of interest in the physical theory must satisfy certain "supersymmetry" conditions; in particular, a complex algebraic manifold is required to be Calabi–Yau, a real seven-dimensional Riemannian manifold is required to have $G_2$ holonomy group. Also of particular interest are the Lagrangian real submanifolds of the Calabi–Yau 3-folds. By a Calabi–Yau manifold $X$ we mean a manifold with $c_1(X)=0, h^0(\Omega^k)=0$, where $\Omega^k$ is the sheaf of holomorphic $k$-forms, and $0<k<\dim(X)$. If $\dim X \geq 2$, we also assume that $X$ is simply connected, but not necessarily compact. For example, if $\dim(X)=1$, $X$ is a torus, if $\dim(X)=2$, $X$ is a $K3$ surface, if $\dim(X)\geq 3$, $X$ is simply called a Calabi–Yau manifold. A compact Kähler manifold $(M,g,J)$ of complex dimension $m\geq 3$ is a Calabi–Yau variety if and only if its holonomy is $SU(m)$. A subvariety $L$ of a symplectic manifold $(X,\omega)$ is Lagrangian if $\omega_{|L}=0$ and $\dim L=(1/2)\dim X$. Sometimes we consider noncompact manifolds, thought of as neighborhoods of a compact projective Calabi–Yau manifold. Typically, our symplectic manifold is a Calabi–Yau 3-fold $(X,\omega)$ together with its Kähler form $\omega$. If there exists an antiholomorphic involution, then the fixed locus is a Lagrangian submanifold.

## The Dualities

We will take the point of view that dualities in physics imply relations between geometric invariants, without dwelling on the physics of the dualities themselves. A consequence of a physics "duality" is the matching of the prepotential of two dual string theories.

### A Few Comments on Chern–Simons Theory: Free Energy (Partition Function)

Let $L$ be a closed oriented manifold together with a principal $G$-bundle. The classical Chern–Simons action is defined as $S(L,A)=\int_L \alpha(A)$, where $\alpha$ is a 3-form on $L$ which depends on a connection $A$ and a suitable bilinear invariant form on the Lie algebra $\mathfrak{g}$. It is well defined under gauge transformations modulo the integers; $e^{2\pi i S(L,A)}$ is well defined. In the large-$N$ dualities considered here, the groups of interests are $SU(N)$ and $U(N)$. The first check of the duality was found with $G=SU(N)$ and $M=S^3$; later it was discovered that the correct group for the matching of the observables must be $U(N)$, while both can be used for the free energies. We shall consider $G=SU(N)$ and $M=S^3$. Without loss of generality, the bundle can be taken to be the product $U(N)\times S^3$; any bilinear invariant form on the Lie algebra $\mathfrak{su}(N)$ is necessarily an integer multiple $k$ of the Cartan–Killing form on the Lie algebra. Then $S=S(k,A)$ and

$$S(k,A)=-\frac{k}{8\pi^2}\int_{S^3}\mathrm{tr}\left(A\wedge dA+\tfrac{2}{3}A\wedge A\wedge A\right)$$

where $k$ is the "level" of the theory. Witten defines the quantum Chern–Simons theory by taking the integral of the Chern–Simons action over all possible connections $\mathcal{A}$ modulo gauge equivalence $\mathcal{G}$:

$$\begin{aligned} &Z(S^3,SU(N))\\ &\quad=\int_{\mathcal{A}/\mathcal{G}}(D\mathcal{A})e^{2\pi i S(A)}\\ &\quad=\int_{\mathcal{A}/\mathcal{G}}(D\mathcal{A})\exp\left(-\frac{ki}{4\pi}\int_{S^3}\mathrm{tr}\left(A\wedge dA+\tfrac{2}{3}A\wedge A\wedge A\right)\right)\end{aligned}$$

Witten shows how to calculate the free energy $Z(S^3,SU(N))$ through topological surgery, assuming $Z(S^2\times S^1)=1$. Witten also defines the partition

function of knots and links in $L$ (the "expectation values"), which are knot and link invariants. The expectation values are computed by evaluating the trace of the holonomy transformation of a U($N$) connection around the knot, and then taking a suitable average of the U($N$) connections. These invariants depend on a choice of the framing of the knot (or link).

The explicit computations involve physics, representation theory, and topology. If $L = S^3$, then:

$$Z(S^3, \mathrm{SU}(N)) = (k+N)^{-N/2}\sqrt{\frac{k+N}{N}} \times \prod_{j=1}^{N-1}\left\{2\sin\left(\frac{j\pi}{k+N}\right)\right\}^{N-j}$$

Reshetikin and Turaev, among others, described mathematically the Chern–Simons free energy and the expectation values.

### A Few Comments on Closed-String Theory: Free Energy (Prepotential)

In IIA closed-string theory on $X$, a Calabi–Yau manifold, one considers holomorphic stable maps of closed Riemann surfaces of genus $g, \phi : \Sigma_g \to X$, with $\phi_*(\Sigma_g) = [\beta] \in H_2(X, \mathbb{Z})$, for all genera $g$ and homology classes $\beta \in H_2(X, \mathbb{Z})$.

Then one forms the closed prepotential $\mathcal{F}_{\mathrm{cl}}(X, t)$, which encodes the enumerative invariants of such maps to $X$, and which depends on the Kähler parameters $t$ of $X$. Sometimes the prepotential is also called "free energy" in the physics literature or Gromov–Witten prepotential, as it contains the Gromov–Witten invariants of $X$. Setting $\mathcal{F}_g(q) = \sum_{\beta \in H_2(X,\mathbb{Z})} C_{g,\beta} q^\beta$, the closed prepotential is defined as

$$\mathcal{F}_{\mathrm{cl}}(X, q) = \sum_{g>0}^{\infty} g_s^{2g-2} \mathcal{F}_g(q)$$

Here $q$ is a formal variable such that $q^{\beta_1+\beta_2} = q_1^\beta \cdot q_2^\beta$ (for $\beta_1, \beta_2 \in H_2(X, \mathbb{Z})$) and $g_s$ is the string coupling constant. $C_{g,\beta}$ are the genus $g$ Gromov–Witten invariants of $X$, corresponding to the class $\beta$ and they have been defined as

$$C_{g,\beta} = \int_{[\overline{M}_{g,0}(X,\beta)]^{\mathrm{virt}}} 1$$

It is difficult to explicitly compute the invariants $C_{g,\beta}$; in particular, there is no known general method for calculating these invariants. They are computed mostly via "localization" methods, in the presence of a suitable torus action. In the case of $g = 0$ the invariants are often computed via IIA–IIB duality, calculating certain periods in the mirror manifold $W$.

**Example** (Faber–Pandharipande). Let $X \cong \mathcal{O}_{\mathbb{P}^1}(-1) \oplus \mathcal{O}_{\mathbb{P}^1}(-1)$; $X$ is a neighborhood of a rigid rational curve, which can be thought of as a local Calabi–Yau manifold; then all the effective curves $\beta \in H_2(X, \mathbb{Z})$ must be of the form $\beta = d[\mathbb{P}^1]$, $\forall d \in \mathbb{N}$. Faber and Pandharipande showed that

$$\mathcal{F}_{\mathrm{cl}}(X, q) = \sum_{d=1}^{\infty} \frac{q^d}{2\sin(dg_s/2)^2} \qquad [1]$$

This formula was proved with localization methods after it was conjectured by Gopakumar and Vafa using large-$N$ dualities. In fact, a consequence of a duality between two theories is the matching of the free energies of two dual string theories. In this particular case, the conjectures imply that Chern–Simons free energy determines, and is determined by, the all-genus closed prepotential of a suitable Calabi–Yau manifold $X$:

$$Z(S^3, \mathrm{U}(N)) \rightleftarrows \mathcal{F}_{\mathrm{cl}}(X, t)$$

Note that the left-hand side is purely topological, as we saw in the previous section, while the right-hand side is holomorphic.

The *trait d'union* between the two prepotentials is given by the interpretation of Chern–Simons theory on $S^3$ as open-string theory on $T^*S^3$ and the geometric transition.

### A Few Comments on Open-String Theory with Branes: Open Prepotential

Let $Y$ be a Calabi–Yau manifold together with $\{\cup L_i\}$, Lagrangian submanifolds; to each submanifold $L_i$ is assigned a gauge group $G_i : L_i$ is wrapped with $G_i$-branes. Here we shall focus on the case $G_i = \mathrm{U}(N_i)$ and we will write $(Y; L_i, \mathrm{U}(N_i))$.

Witten shows that the open prepotential $\mathcal{F}_{\mathrm{op}}(Y, \lambda, t_{\mathrm{op}}, g_s)$ depends on 't Hooft's coupling constants $\lambda_i$ associated to Chern–Simons theory on the Lagrangian submanifolds $(L_i, \mathrm{U}(N_i))$, together with the open Kähler parameters $t_{\mathrm{op}} \in H_2(X; \cup L_i, \mathbb{Z})$, and the string coupling constant $g_s$. To describe the open prepotential, Witten argues, we consider all maps of Riemann surfaces with boundary to $Y$, with the condition that the boundaries are mapped to the Lagrangian submanifolds $L_i$; one should also include all the "highly degenerate holomorphic maps," in particular those which contract $\Sigma_{g,h}$ to a "ribbon graph" on the Lagrangian $\cup L_i$. The contribution of these highly degenerate maps is captured by the quantum Chern–Simons theory of the Lagrangians $\{L_i, \mathrm{U}(N_i)\}$.

**Application 1** (Chern–Simons free energy as open prepotential). Let us consider open IIA on $Y = T^*S^3$ with U($N$)-branes wrapped on $L = S^3$: $L$ is a Lagrangian submanifold with the standard symplectic structure; note that in $T^*L$ there are no nontrivial homology curves. Then, according to Witten, the corresponding open prepotential $\mathcal{F}_{\text{op}}(Y, \cup L_i)$ must only depend on the "highly degenerate" maps and must consist of the Chern–Simons term $F_{\text{CS}}$ on $L = S^3$. In particular,

$$F_{\text{CS}} = \log Z(S^3) = \mathcal{F}_{\text{op}}(Y, \lambda, g_s)$$

where $\lambda = 2N\pi/(k+N)$ is the 't Hooft coupling constant. Periwal (1993) showed that, for large $N$, $\log Z(S^3)$ could be expanded as a closed-string expansion:

$$F_{\text{CS}}(\lambda) = \sum_{g \geq 0} \mathcal{F}_g(\lambda) g_s^{2\ 2g}$$

where $g_s =: 2\pi/(k+N)$ is the Chern–Simons coupling constant. In 1998 Gopakumar and Vafa, using physics arguments, deduced that the expansion would have the closed form [1], which was later proved by Faber and Pandharipande.

The explicit description of the open prepotential in the presence of homology classes is not known; one would need to combine the enumerative invariants of open maps together with the quantum Chern–Simons factor. We shall discuss an approach at the end of this note, but consider first the geometric transition.

## The Transition

The conjecture says that $U(N)$ Chern–Simons gauge theory on $S^3$ is dual, for large values of $N$, to IIA closed topological string theory on a suitable Calabi–Yau manifold $X$. A starting point to find such $X$ is that $S^3$ is a Lagrangian 3-cycle in the manifold $Y = T^*S^3$; performing a topological surgery by replacing $S^3$ with $S^2$ one obtains a (local) Calabi–Yau manifold $X$, on which the dual IIA theory is compactified. The key observation is that $Y$ can be identified with the algebraic variety of equation $\{xy - zw = t\} \subset \mathbb{C}^4$ and that this is a complex smoothing (in fact the Milnor fiber) of $Y_0$ with equation $\{xy - zw = 0\} \subset \mathbb{C}^4$. On the other hand, $X$ is a small resolution of this singularity, where $\mathbb{P}^1$ is the exceptional locus of the birational contraction. The origin is an "ordinary double point" singularity and the nontrivial sphere $S^3 \subset Y$ is the vanishing cycle of the degeneration. The manifolds involved are noncompact: the exceptional curve $[\mathbb{P}^1] = t$ is the only nontrivial homology class in $X$, and the enumerative invariants in $X$ can be thought as the contribution of the exceptional curve in a neighborhood of a Calabi–Yau manifold. We shall present the steps leading to this construction and the evidence for the conjecture.

### The Local Construction of X

Let $Y_\mu = \{(w_1, \ldots, w_4) \in \mathbb{C}^4$ such that $\sum_{j=1}^4 w_j^2 = \mu\}$.

**Proposition 1** *Let $\mu$ be a nonzero real positive parameter; then:*

- *$L = S^3 \subset T^*S^3$ is a Lagrangian submanifold of $T^*S^3$ with its standard symplectic structure;*
- *$T^*S^3 \cong Y_\mu$ and $L \cong L_\mu \stackrel{\text{def}}{=} \{\text{Re}(\sum_{j=1}^4 w_j^2 = \mu)\}$.*

In fact, we can embed $T^*S^3$ in $\mathbb{R}^8$ as

$$\sum_{j=1}^{4} q_j^2 = 1, \qquad \sum_{j=1}^{4} q_j p_j = 0$$

where $S^3 = \{p_i = 0\}$; consider then the morphism $\mathbb{C}^4 \to \mathbb{R}^8$ defined by setting

$$q_j = \frac{\text{Re}(w_j)}{\sqrt{\mu + \sum_i v_i^2}}, \qquad p_j = \text{Im}(w_j)$$

which induces the diffeomorphism $Y_\mu \cong T^*S^3$ of the statement.

**Remark 1** Let $Y_0 = \{\sum_{j=1}^4 w_j^2 = 0\} \subset \mathbb{C}^4$; then:

- $Y_0$ is singular at the origin,
- $Y_\mu$ is a complex deformation of $Y_0$, and
- $L_\mu$ is called a "vanishing cycle."

With a change of coordinates we can write the equation of $Y_\mu$ as $\{xy - zw = 0\}$; the singularity is still at the origin. This singularity is an ordinary double point, which is often referred in physics literature as "the conifold singularity." Let $X \subset \mathbb{C}^4 \times \mathbb{P}^1$ be defined:

$$\lambda z + \nu y = 0, \qquad \lambda x + \nu w = 0$$

$[\lambda, \nu] \in \mathbb{P}^1$.

**Remark 2** $X$ is smooth and the morphism

$$\phi : X \longrightarrow Y_0, \quad ((x, y, z, w), [\lambda, \mu]) \mapsto (x, y, z, w)$$

is an isomorphism $\phi_{|X \setminus \mathbb{P}^1} : (X \setminus \mathbb{P}^1) \simeq (Y_0 \setminus \{0\})$ and $\mathbb{P}^1 \mapsto (0, 0, 0, 0, ) \subset \mathbb{C}^4$. $\phi$ is a small (nondivisorial) birational resolution of the singularity at the origin. $Y_\mu$ is a deformation (smoothing) of $Y_0$. Note that topologically $S^3 \cong L_\mu \subset Y_\mu$ has been replaced by $\mathbb{P}^1 \cong S^2 \subset X$. The algebraic properties of the topological surgery between $Y_\mu$ and $X$ were first studied by Clemens in 1988.

### Transitions in Geometry

A transition between $X$ and $Y$ is a birational contraction from a smooth Calabi–Yau $X$ to a singular variety $Y_0$ followed by a complex deformation to another smooth Calabi–Yau manifold $Y$:

$$\begin{array}{ccc} & & X \\ & & \downarrow \\ Y & \leftsquigarrow & Y_0 \end{array}$$

The vanishing cycles of the complex deformation $\cup L_i$ are always Lagrangian submanifolds of $Y$. The transition makes sense if $\dim(X) = \dim(Y) \geq 2$ and it is nontrivial if $\dim(X) = \dim(Y) \geq 3$, when the topology of $X$ is different from the topology of $Y$. The possible transitions among Calabi–Yau 3-folds have been classified.

**Conjecture 1** Let $X$ and $Y$ be Calabi–Yau manifolds related by a geometric transition: then IIA open theory with $U(U)$ branes compactified on $(Y, \cup L_i)$ is dual to IIA closed theory compactified on $X$ (with fluxes).

As a consequence:

**Conjecture 2** Let $X$ and $Y$ be Calabi–Yau manifolds related by a geometric transition: then $\mathcal{F}_{\text{op}}(Y, \lambda, g_s, t_{\text{op}}) = \mathcal{F}_{\text{cl}}(X, q, g_s)$ for a suitable identification of the parameters.

The results stated in the previous section can be summarized in the the following statement, which is the proof of the above conjecture for the special case of a local conifold transition:

**Theorem 1** *Let* $X \cong \mathcal{O}_{\mathbb{P}1}(-1) \oplus \mathcal{O}_{\mathbb{P}1}(-1)$ *and* $Y = T^*S^3$ *with* $U(N)$ *branes wrapped on* $L = S^3$. *Then* $X$ *and* $Y$ *are related by a conifold transition and* $\log F_{\text{CS}}(S^3) = \mathcal{F}_{\text{op}}(Y, \lambda) = \mathcal{F}_{\text{cl}}(X, q)$, *with the identification*

$$\lambda = \frac{2N\pi}{k+N} = q, \qquad g_s = \frac{2\pi}{k+N}$$

This matching of the free energies is supporting evidence for the large-$N$ conjecture. At this moment, we still do not know if Conjectures 1 and 2 hold for more general transitions.

## A Few Comments on Knots and Links

Later, Ooguri and Vafa extended the conjecture to the observables, that is, by adding knots and links in $S^3$; the guiding principle is that a knot (or link) $\mathcal{C} \subset S^3$ should determine a noncompact Lagrangian submanifold $\mathcal{L}_{\mathcal{C}} \subset X$; it is conjectured that the knot (and link) invariants, expressed as expectation values, should determine and be determined by the enumerative invariants of morphisms of bounded Riemann surfaces, with boundaries mapped onto $\mathcal{L}_{\mathcal{C}}$. We refer to these invariants as open Gromov–Witten invariants. While both statements have been verified with mathematical techniques only when $\mathcal{C}$ is the unknot, there is much supporting evidence for the conjecture in general. We will not describe these aspects here but only make a few remarks.

The expectation values of a knot $\mathcal{C}$ are computed by taking first the trace of a holonomy matrix of a $U(N)$ connection $A$ along $\mathcal{C}$ and then integrating over all connections (modulo gauge equivalence). As for the case of the Chern–Simons free energy, the definition of expectation values has been worked out both in the realm of physics and of mathematics. The expectation values are knot and link invariants, and depend on a choice of the framing of the knot (or link). The open Gromov–Witten invariants have not yet been constructed, as we shall discuss in the following section; however, starting with the work of Katz and Liu, Li and Song open invariants have been successfully calculated in the presence of a torus action. The resulting invariants do depend on the choice of the torus action, which has been shown to match the choice of the framing of the knot (or link).

## More on the Open Prepotential

The open Gromov–Witten invariants, in analogy with the closed case, should "count" in an appropriate sense open morphisms; at this point, it is not known how to define this quantity. To proceed in analogy with the closed case, one would need to define the appropriate moduli space of open maps and its virtual fundamental class. On the other hand, open invariants have been successfully calculated in the presence of a torus action, assuming the existence of the moduli and virtual fundamental class and that the Atiyah–Bott localization theorems can be applied. We shall follow this approach in sketching how the IIA prepotential has been computed in many examples.

### Open Invariants

Let $[\beta] \in H_2(Y; \cup L_i, \mathbb{Z})$ be the relative homology class of Riemann surfaces in $Y$ with boundary on the union of the Lagrangian 3-cycles $\cup_i L_i$ and a class $[\gamma] \in H_1(\cup L_i)$.

If $\Sigma_{g,h}$ is a Riemann surface of genus $g$ and $h$ boundary components, let $\phi: \Sigma_{g,h} \to Y$ be a morphism with

$$\phi_*(\Sigma_{g,h}) = [\beta] \in H_2(Y; \cup L_i, \mathbb{Z})$$

The open generating function is

$$F_o(Y, \cup L_i, t_{\text{op}}, g_s) = \sum_{g,h \geq 0}^{\infty} g_s^{2g-2+h} F_{g,h}(t_{\text{op}})$$

with

$$F_{g,h}(t_{\text{op}}) = \sum_{\beta,\gamma} C_{g,h,\beta,\gamma} q^{\beta} y^{\gamma}$$

Here $q$ and $y$ are formal variables such that $q^{\beta_1+\beta_2} = q^{\beta_1} \cdot q^{\beta_2}$ and $y^{h_1+h_2} = y^{h_1} \cdot y^{h_2}$, for $\beta_1, \beta_2 \in H_2(Y; \cup L_i, \mathbb{Z}), \gamma_1, \gamma_2 \in H_1(\cup L_i, \mathbb{Z})$; $t_{\text{op}}$ is the open Kähler parameter, $g_s$ is the string coupling constant and $C_{g,h,\beta,\gamma}$ should "count" in an appropriate sense the maps $\phi$.

**Example** (Ooguri–Vafa; Katz–Liu; Li–Song). If $Y = \mathcal{O}_{\mathbb{P}_1}(-1) \oplus \mathcal{O}_{\mathbb{P}_1}(-1)$, then $t$ is the class of the $\mathbb{P}^1 \cong S^2$, $t/2$ represents the class of the lower hemisphere in $S^2$. The Lagrangian $L$ is the Lagrangian $\mathcal{L}$ in the previous sections, which corresponds to the unknot in $S^3 \subset Y$; it is the fixed locus of an antiholomorphic involution on $X$ and it intersects $S^2$ in an equator. Then, for a suitable choice of the torus action:

$$F_o(Y, \cup L_i, t_{\text{op}}, g_s) = \sum_d \frac{y^d}{2d \sin(d\lambda/2)} \, \mathrm{e}^{-dt/2}$$

There is a complete form for more general torus actions. The above formula was first computed by Ooguri and Vafa, using string-theory arguments, and then computed by the mathematicians, Katz and Liu, and Li and Song.

### More on the Open IIA Prepotential

If there is only one rigid open curve in $Y$, say a disk $C$, with boundary on $L \subset Y$, then, as Witten showed, the open prepotential is a combination of the open enumerative invariants as described above with $\beta = d[C]$ and $\gamma = \partial C$ and the expectation values of the unknot $\partial C$. The variable $Y$ is changed in the trace of the holonomy of a connection.

In the presence of a torus action, one can treat the fixed locus as if it were rigid and proceed accordingly.

With these techniques, Conjecture 2 has been verified for many cases of conifold transitions, with $t_{\text{op}}$ nontrivial, for a suitable identifications of the parameters, including when both $X$ and $Y$ are compact manifolds (Diaconescu–Florea 2003).

*See also:* AdS/CFT Correspondence; Chern–Simons Models: Rigorous Results; Large-*N* and Topological Strings; Mirror Symmetry: A Geometric Survey; String Field Theory.

## Further Reading

Clemens CH (1983) Double solids. *Advances in Mathematics* 47: 107–230.

Diaconescu D-E, Florea B (2003) Large N duality of compact Calabi–Yau three-folds, hep-th/0302076.

Dijkgraaf R and Vafa C (2002) Matrix models, topological strings and supersymmetric gauge theories. *Nuclear Physics B* 644(3): 3–20 (hep-th/0206255).

Faber C and Pandharipande R (2000) Hodge integrals and Gromov–Witten theory. *Inventiones Mathematicae* 139: 173–199 (math.AG/9810173).

Freed DS (1995) Classical Chern–Simons theory, Part 1. *Advances in Mathematics* 113: 237–303.

Gopakumar R and Vafa C (1999) On the gauge theory/geometry correspondence. *Advances in Theoretical and Mathematical Physics* 3: 1415–1443.

Katz S and Liu CCM (2002) Enumerative geometry of stable maps with Lagrangian boundary conditions and multiple covers of the disk. *Advances in Theoretical and Mathematical Physics* 5: 1–49.

Li J and Song YS (2001) Open string instantons and relative stable morphisms, hep-th/0103100.

Mariño M (2004) Chern–Simons theory and topological strings, hep-th/0406005.

Ooguri HH and Vafa C (2000) Knot invariants and topological strings. *Nuclear Physics B* 577: 419–438.

Periwal V (1993) Topological closed-string interpretation of Chern–Simons theory. *Physical Review Letters* 71: 1295–1298.

't Hooft G (1974) A planar diagram theory for strong interactions. *Nuclear Physics B* 72: 461–473.

Witten E (1989) Quantum field theory and the Jones polynomial. *Communications in Mathematical Physics* 121: 351–399.

Witten E (1995) Chern–Simons gauge theory as a string theory. In: *The Floer Memorial Volume*, Progress in Mathematics, vol. 133, pp. 637–678. Basel: Birkhäuser (hep-th/9207094).

# Lattice Gauge Theory

**A Di Giacomo**, Università di Pisa, Pisa, Italy



## Introduction

As a prototype of lattice gauge theory, quantum chromodynamics (QCD) will be considered in this article. All statements about QCD can easily be extended to other theories, with different gauge group and different content of particles.

QCD is a gauge theory with gauge group SU(3) (color group), coupled to spin-1/2 particles (quarks) belonging to the fundamental representation of the color group. There exist in Nature six different species (flavors) of quarks, with masses ranging from $m_{\text{up}} \sim 5\,\text{MeV}$ to $m_{\text{top}} \sim 180\,\text{GeV}$: the values of these masses are determined by other interactions and can be treated as input parameters of the theory as well as the number of quark flavors. In standard notation, the Lagrangian reads

$$L = -\frac{1}{2}\text{tr}(G_{\mu\nu}G_{\mu\nu}) + \sum_f \bar{\psi}_f(\text{i}\,\not{D} \;-\; m_f)\psi_f \qquad [1]$$

The sum runs over the six quark flavors $f$. $G_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\nu + \text{i}g[A_\mu, A_\nu]$ is the field strength tensor, $A_\mu = \sum T^a A^a_\mu$ the (gluon) gauge field, $T^a(a = 1, \ldots, 8)$ are the eight generators of the gauge group in the fundamental representation, normalized as $\text{tr}(T^a T^b) = (1/2)\delta^{ab}$. $\psi_f$ is a color triplet of fields. Under a gauge transformation $U(x)$,

$$\psi_f(x) \to \psi'_f(x) = U(x)\psi_f(x) \qquad [2]$$

$$\begin{aligned} A_\mu(x) &\to A'_\mu(x) \\ &= U(x)A_\mu U^\dagger(x) + \text{i}U(x)\partial_\mu U^\dagger(x) \end{aligned} \qquad [3]$$

$D_\mu$ is the covariant derivative of $\psi$

$$D_\mu\psi_f = (\partial_\mu - \text{i}gA_\mu)\psi_f \qquad [4]$$

and transforms like $\psi_f$ by construction.

$L$ is invariant under the gauge transformation equations [2] and [3]. As a consequence of gauge invariance, the theory has one single coupling constant $g$.

To make connection with the observations, one has to solve the theory, that is, one has to construct a Hilbert space on which the fields act as operators obeying the equations of motion and the canonical commutation relations. In textbook field theory, this is done by splitting the Lagrangian $L$ into two parts:

$$L = L_0 + L_{\text{I}} \qquad [5]$$

with $L_0$ the part of $L$ which is bilinear in the fields and $L_{\text{I}}$ the rest. $L_0$ can be solved exactly since it describes free particles and the corresponding equations of motion are linear. The resulting Hilbert space is the Fock space of free particles. $L_{\text{I}}$ is treated as a perturbation producing scattering between the fundamental particles. This approach works well in quantum electrodynamics, where the observed particles (electrons and photons) coincide with the excitations of the fundamental fields of the Lagrangian.

In QCD, the fundamental excitations (the quarks and the gluons) are observed as particles neither in Nature nor as a product of high-energy collisions between elementary particles. This feature is known as confinement of color. The conjecture is that excitations with nontrivial color are forbidden to propagate as free particles. However, if hadrons are probed at short distances by photons or by leptons, everything works as if they were composite states of quarks. The accepted explanation relies on asymptotic freedom: the effective coupling constant becomes small at short distances (high momentum transfers) and the constituents behave as free particles.

At large distances, the fundamental excitations are not observed, the interaction is strong and the perturbative picture describing scattering between quarks and gluons is not adequate for the real world.

An alternative quantization procedure is needed which does not rely on perturbation theory. A formally exact quantization procedure is the Feynman path integral. The solution of the theory is given in terms of a functional integral $Z[J]$, which generates the correlators of the fields in the ground state (vacuum). Indicating symbolically the Lagrangian coordinates, namely the fields, by a single symbol $\Phi$, one has

$$Z[J] = \int \prod_x \text{d}\Phi(x)\exp\left[-S[\Phi] - \int J(x)\Phi(x)\text{d}x\right] \qquad [6]$$

The connected Euclidean vacuum correlators are given in terms of functional derivatives of $Z[J]$

$$\begin{aligned} &< 0|T(\Phi(x_1)\Phi(x_2)\cdots\Phi(x_n))|0 >_{\text{conn}} \\ &= \frac{1}{Z[0]}\frac{\delta^n Z[J]}{\delta J(x_1)\delta J(x_2)\cdots\delta J(x_n)}\bigg|_{J(x)=0} \end{aligned} \qquad [7]$$

"Euclidean" means that they are analytic continuations to imaginary times. Going to Euclidean system is necessary to isolate the vacuum state. The amplitudes can be analytically continued back to Minkowski space. The Hilbert space and all the physical observables can be constructed in terms of the correlators, a property known as reconstruction theorem. Formally (i.e., assuming that everything makes sense only if the functional integral exists),

$$\begin{aligned} &< 0|T(\Phi(x_1)\Phi(x_2)\cdots\Phi(x_n))|0 >_{\text{conn}} \\ &= \frac{1}{Z}\int\prod_x \mathrm{d}\Phi(x) \\ &\times \exp(-S[\Phi])\Phi(x_1)\Phi(x_2)\cdots\Phi(x_n) \end{aligned} \qquad [8]$$

The continuation to imaginary time changes sign to the kinetic energy, and $Z$ formally becomes the partition function of a four-dimensional statistical model with Hamiltonian $S_{\mathrm{E}}[\Phi]$, a general fact in Feynman integrals.

By definition of functional integral, $Z$ is defined by discretizing a finite volume $V$ of spacetime to a finite set of points and then sending their number to infinity, making a set dense in $V$. If the limit exists, a $Z_V$ is obtained. The volume $V$ is then sent to infinity, to cover the whole spacetime (thermodynamical limit) and $Z_V$ eventually converges to $Z$. A rigorous proof of the existence of these limits does not exist for QCD, but there are qualitative arguments that this is the case, which will be presented below.

In the lattice formulation of field theory, a regular lattice, usually cubic, is taken as a discretization of spacetime.

From the very definition of Feynman integral, it follows that the formulation of field theory on the lattice is nothing but an approximation to the limit which defines $Z$. It will provide a good approximation if the lattice spacing is small enough with respect to the physical lengths involved and if the lattice is large compared to them.

Perturbation theory amounts to split the action into a bilinear term $S_0$ and an interaction term $S_{\mathrm{I}}$ containing the higher powers of the fields. The $Z$ integral is then computed by expanding the weight in a power series of $S_{\mathrm{I}}$:

$$\begin{aligned} &\int\prod_x \mathrm{d}\Phi(x)\exp(-S_0 - S_{\mathrm{I}}) \\ &= \int\prod_x \mathrm{d}\Phi(x)\exp(-S_0)\sum_n \frac{(-S_{\mathrm{I}})^n}{n!} \end{aligned} \qquad [9]$$

The Feynman integral thus becomes Gaussian, can be computed, and gives the usual perturbative expansion. The two limits (integral and series expansion) do not commute in general. For QCD, there are indeed arguments that the renormalized perturbative expansion does not converge and is plagued by singularities known as renormalons.

## Wilson's Formulation

For field theories of scalar particles, the lattice discretization is performed by assigning a value of the field to each site of the lattice. The Wilson formulation for gauge theories is not made in terms of the fields $A_\mu$, which are defined in the Lie algebra of the gauge group, but in terms of parallel transports, which are elements of the group itself. The building blocks are parallel transports along links parallel to spacetime axes connecting neighboring sites

$$\begin{aligned} &U_\mu(x) \\ &\equiv P\exp\left(\mathrm{i}g\int_x^{x+\hat\mu} A_\rho \mathrm{d}x^\rho\right) \approx \exp(\mathrm{i}gaA_\mu(x)) \end{aligned} \qquad [10]$$

where $\hat\mu$ indicates the vector of length $a$ in the $\mu$ direction and $P$ the ordered product. The last approximate equality is valid in the limit of small lattice spacing $a$. $g$ is the coupling constant.

Under a gauge transformation $V(x)$,

$$U_\mu(x) \to V(x)U_\mu(x)V^\dagger(x+\hat\mu) \qquad [11]$$

It follows from eqn [11] that the parallel transport along a closed path is gauge invariant. The density of action can be written in terms of the parallel transport along the elementary square of links in the hyperplanes $\mu\nu$ $\Pi_{\mu\nu}$, known as plaquette:

$$\prod_{\mu\nu} = \mathrm{tr}[U_\mu(x)U_\nu(x+\hat\mu)U_\mu^\dagger(x+\hat\nu)U_\nu^\dagger(x)] \qquad [12]$$

By expanding in powers of $a$, one easily finds

$$\prod_{\mu\nu} = N_{\mathrm{c}} - \frac{1}{2}a^4\mathrm{tr}[G_{\mu\nu}G_{\mu\nu}] + O(a^6) \qquad [13]$$

with $N_{\mathrm{c}}$ the number of colors, 3 for QCD. The lattice action can be defined as

$$S = \sum_{x\mu\nu}\beta\left(1 - \frac{1}{N_{\mathrm{c}}}\Pi_{\mu\nu}\right) \qquad [14]$$

with $\beta = 2N_{\mathrm{c}}/g^2$, and tends to the continuum action as $a \to 0$, $O(a^2)$. An infinite number of higher-order terms in $a$ exist, which come from the expansion of the links, but they are expected to be irrelevant in the continuum limit $a \to 0$.

The measure of the Feynman integral is assumed to be the Haar measure of the gauge group for each link, which again can be shown to tend to the continuum measure in the continuum limit.

Everything is gauge invariant, contrary to the perturbative formulation, where a gauge fixing is required to define the vector meson propagator.

By Weierstrass theorem, the integral is finite for any finite number of links, the gauge group being compact.

Any other choice of the lattice action differing from the Wilson action of eqn [14] by terms of higher order in $a$ will have the same continuum limit: there is significant freedom in the choice of the action.

In the language of statistical mechanics, the Euclidean lattice formulation is a spin model. Different choices of the action correspond to different spin models. In the vicinity of a second-order phase transition, however, the correlation length becomes large with respect to the lattice spacing and all the irrelevant terms become negligible. All the spin models at the critical point belong to the same universality class and define the same field theory.

This is what happens for QCD because of asymptotic freedom. By renormalization group arguments, the lattice spacing behaves as

$$a(\beta) \approx \frac{1}{\Lambda}\exp(-b_0\beta) \qquad [15]$$

at sufficiently large $\beta$, where $-b_0$ is the coefficient of lowest-order term of the $\beta$-function, $b_0$ is positive and $\Lambda$ is a physical scale. As $\beta \to \infty$, $a$ tends exponentially to zero in physical units and the coarse structure of the lattice becomes unimportant, indicating that the short-distance limit in the definition of the Feynman integral exists. The theory also develops a mass scale $\Lambda$ which insures the existence of a finite correlation length and hence of the thermodynamical limit. In practice, when $\beta$ is increased, the lattice space becomes exponentially small in physical units. As a consequence, however, the physical scale becomes exponentially large in lattice units, and an exponentially large lattice is needed to insure the large-distance convergence. This makes life difficult if the Feynman integral has to be computed numerically.

## Quarks

Fermion fields are defined on lattice sites. The naive lattice transcription of the fermion term in eqn [1] consists in replacing the covariant derivatives by finite differences with parallel transports to make the result gauge covariant. In principle, $D_\mu\psi(x) = U^\dagger(x)\psi(x+\hat{\mu}) - \psi(x)$ is a correct definition. In practice, a more symmetric difference is used which is correct $O(a^2)$, namely

$$D^{\mathrm{L}}_\mu\psi(x) = \frac{1}{2}[U(x)\psi(x+\hat{\mu}) - U^\dagger(x-\hat{\mu})\psi(x-\hat{\mu})] \qquad [16]$$

The fermionic Lagrangian then reads

$$\sum_x \bar{\psi}(x)[\mathrm{i}\not{D}^{\mathrm{L}} - m]\psi(x) \equiv \sum_{x,x'\alpha\beta} \bar{\psi}_\alpha(x) M^{-1}_{\alpha\beta(x,x')}\psi_\beta(x') \qquad [17]$$

It is convenient to indicate this expression in the form $S_f = \bar{\psi}M^{-1}\psi$, where $\psi$ is a large column whose elements are labeled by the site $x$ and by the component $\alpha$. The functional integral over $\psi$ can explicitly be done by using the standard rules of integration on Grassman variables, since the action is bilinear,

$$Z = \int \prod \mathrm{d}U_\mu(x)\mathrm{d}\psi(x)\mathrm{d}\bar{\psi}(x) \times \exp(-S_{\mathrm{E}}[U] - \bar{\psi}M\psi) \qquad [18]$$

The result is

$$Z = \int \prod \mathrm{d}U_\mu(x)\exp(-S_{\mathrm{E}}[U])\det M \qquad [19]$$

The effect of fermions is to multiply the weight by a functional determinant which depends on the gauge field configuration.

A problem exists, however, in this procedure already at the level of free fermions, that is, putting $U = 1$ in the action and in the determinant of eqn [18]. The equation of motion reads, in Fourier transform,

$$\sum_\mu \left[\gamma_\mu \sin\left(2\pi\frac{k_\mu}{L}\right) - m\right]\tilde{\psi}(k) = 0 \qquad [20]$$

With respect to the continuum, the momentum $p_\mu = 2\pi k_\mu/L$ has been replaced by its sinus. At small values of $p_\mu$, eqn [20] coincides with the Dirac equation. However, an alternative solution exists at $p_\mu \approx \pi$, for each $\mu$ independently. The new equation differs from the other by a change of sign of $\gamma_\mu$. Changing sign of one of the gammas means changing sign to $\gamma^5 \equiv \gamma^1\gamma^2\gamma^3\gamma^4$; which is the chirality of the fermion. Instead of one fermion, we then have $2^4 = 16$ fermion species, organized in pairs with opposite chiralities. It is impossible to have a single fermion with a given chirality. A number of recipes have been proposed to circumvent this artifact of the lattice regulation, for example, introduce by hand a term in the action which removes the spurious particles in the limit of zero lattice spacing (Wilson's fermions); double the lattice spacing by constructing two sublattices on even and odd sites, respectively, which propagate fermions of opposite chirality (staggered fermions),

so that the argument of the sinus in the derivative is doubled. More recently, an idea which goes back to Ginsparg and Wilson has been implemented, which consists in replacing a strictly local equation of motion like eqn [20] by an equation with the same continuum limit which is nonlocal, but with a nonlocality falling off exponentially at large distances, a recipe which makes propagation of chiral fermions possible. This is an important improvement, even if very demanding in computer power.

## Numerical Simulations

Solving analytically the lattice version of QCD would allow one to follow constructively all the steps which bring to the definition of $Z$, that is, the ultraviolet and the infrared limit, as explained earlier. Presently that is out of reach. Also an attempt by Wilson to solve the lattice renormalization group equations by techniques of decimation is not conclusive.

The problem can be attacked numerically. One way would be to compute the integral numerically. That is, however, prohibitive: it would be like solving exactly the equations of motion for the molecules of a gas. The lattice theory is in fact a four-dimensional statistical mechanics with the Boltzmann factor $\beta = 2N_c/g^2$ and Hamiltonian equal to the Euclidean action. As in statistical mechanics the way out is to create a significant sample of configurations with weight $\exp(-\beta S_E)$ and to determine the field correlators which describe physics by an average on this ensemble. This is done by Monte Carlo techniques.

The basic principle is to start from an arbitrary field configuration and make a sequence of random changes, normally on a single link at a time, with uniform probability in the group measure so as to converge toward the equilibrium distribution $\exp(-\beta S_E)$. For that purpose, the probability $P_{C'C}$ to change from a configuration $C$ to another $C'$ is constrained to obey the detailed balance relation

$$P_{C'C}\exp(-\beta S[C]) = P_{CC'}\exp(-\beta S[C']) \qquad [21]$$

A common algorithm is known as metropolis. The way to implement the condition (eqn [21]) is to accept the new trial configuration $C'$ if $S[C'] \leq S[C]$, and to accept it with probability $\exp(-\beta[S(C') - S(C)])$ if $S[C'] \geq S[C]$. An alternative method is known as "heat-bath". If the probability of the configuration for one link at a fixed value of the other variables is explicitly known, the change can be accepted with that probability.

In the presence of dynamical quarks, the integral eqn [18] is converted into an integral on bosonic variables by inverting the matrix $M$:

$$Z = \int \prod \mathrm{d}U_\mu(x)\,\mathrm{d}\phi(x)\,\mathrm{d}\phi(x)^\dagger \times \exp(-S_E[U] - \phi^\dagger[M^\dagger M]^{-1}\phi) \qquad [22]$$

The property has been used such that $\int \prod \mathrm{d}\phi(x)\,\mathrm{d}\phi^\dagger(x)\,\exp(-\phi^\dagger[M^\dagger M]^{-1}\phi) = |\det M|$. A metropolis updating is then performed on the combined $U_\mu$ and $\phi$ variables. To have a choice of the trial uniform in the measure, an algorithm is commonly used which is based on ergodicity, known as hybrid molecular dynamics. A fictitious conjugate momentum is associated with all variables, and a fictitious Hamiltonian is defined by adding to the action, considered as a potential energy, the sum of the squares of the conjugate momenta. A classical evolution is then performed in time by small steps which should displace the state in phase space ergodically: the evolution is called a trajectory. After a number of steps, a metropolis test is made as explained above.

Typically, the computer time needed to produce a significant configuration is proportional to the volume $V$ of the lattice for pure gauge systems, to $V^{5/4}$ in the hybrid algorithm for full QCD.

As explained before, in order to have a good approximation to the Feynman integral the lattice spacing has to be small compared to the physical scales, for example, with respect to the Compton wavelength of the heaviest quark. On the other hand, to control volume effects it has to be large compared to the biggest physical length, for example, with respect to the Compton wavelength of the lightest quark. Since there is a factor $m_{\text{top}}/m_{\text{up}} \approx 3 \times 10^3$ between these two lengths, the lattice size needed would be prohibitive from numerical point of view. In practice, lattices of size $L^4$ are affordable with $L \leq 64 - 128$. For this reason, only the light quarks $u$, $d$, $s$ are kept, which have mass smaller than the typical scale of the theory, which can be identified as the square root of the string tension. In the limit in which light quark masses are small compared to QCD scale, the Lagrangian is invariant under any unitary mixing of them. A global SU(3) invariance exists, which is known as flavor symmetry, and is broken by the difference of quark masses. Heavier quarks can be described by an effective theory, since they have negligible dynamical effects at low energies.

## A Selection of Physics Results

### String Tension

A big excitement followed the first numerical calculations by M Creutz at the beginning of the 1980s in which the static potential $V(r)$ between a quark and an antiquark was computed in pure-gauge theory on the lattice. One way to measure it is to measure the correlator of two Polyakov lines at a distance $r$ on a significant ensemble of field configurations. The Polyakov line is the parallel transport in the fundamental representation along the time axis across the lattice: with periodic boundary conditions it is a closed loop, and hence it is gauge invariant. It can be proved that the log of this correlator is equal to $-V(r)aL_t$ with $L_t$ the extension of the lattice in the time direction. It was found that

$$V(r) = \sigma r \qquad [23]$$

The parameter $\sigma$ is known as string tension. A potential of the form eqn [23] means confinement: an infinite amount of energy is required to pull apart the particles at infinite distance. The parameter $\sigma$ can be determined phenomenologically from the mass spectrum of the mesons and $\sigma 2\pi \approx 1\,\text{GeV}$. What is measured on the lattice is

$$\sigma a(\beta)^2 n^2 \qquad [24]$$

where $n$ is the distance of the two Polyakov lines in lattice spacings and $a(\beta)$ the lattice spacing in physical units. In fact, the computer only produces pure numbers. If the lattice QCD belongs to the same universality class of QCD at the critical point, that is, if the lattice really defines QCD, the dependence of $a(\beta)$ on $\beta$ is dictated by the $\beta$-function of the renormalization group. At sufficiently large $\beta = 6/g^2$,

$$a(\beta) \approx \frac{1}{\Lambda_{\text{latt}}} \exp(-b_0 \beta) \qquad [25]$$

with $b_0 = (11/3)N_c/16\pi^2$. $\Lambda_{\text{latt}}$ is the energy scale of the theory. The measurement of the potential gives indeed a dependence of the lattice spacing on $\beta$ consistent with eqn [25] and allows one to determine $\sigma/\Lambda^2_{\text{latt}}$. The absolute value of the lattice spacing can be determined by comparison with the physical value of the string tension. The theory is able to produce a physical scale. The correlation length is finite and as a consequence the infrared limit of the Feynman integral exists.

### Mass Spectrum

Any operator with the quantum numbers of a particle can be used as interpolating field for it. The correlator of the operator at large distances behaves like a sum of exponentials $\exp(-mr)$ with $m$ the masses of the particles with the same quantum numbers. At large distances the lightest particle dominates, especially if the operator has a good overlap, that is, if its matrix element between vacuum and the state of the particle is the biggest. From the correlators $mr$ can be determined. On the lattice $r = na(\beta)$ so that, by eqn [25] what is really determined is the ratio $m/\Lambda_{\text{latt}}$. If $\Lambda_{\text{latt}}$ has been determined, for example, from the string tension, the mass of the particle results in physical units. Alternatively, the ratios of any two masses can be determined and the scale fixed by the value of one of them. A good agreement is obtained already in pure gauge (quenched approximation) indicating that the quark loops are relevant at the level of 10% typically. This fact supports the idea that the large $N_c$-limit is a good approximation to reality, quark loops being nonleading in that limit. The light particle masses are more difficult to compute, being sensitive to the masses of light quarks which cannot be taken at realistic values due to computational difficulties: large lattices are required and big fluctuations are present near the chiral point. The spectrum of particles made of heavy quarks can be computed using effective theories, and nicely fits experiment. A byproduct is a precise determination of the gauge coupling constant, competitive with phenomenological determinations from short distance perturbative QCD.

### Weak Interaction Matrix Elements

There exist matrix elements of currents (or products thereof) entering in weak amplitudes which involve large distances and are not computable in perturbation theory. Lattice can be used to evaluate them. Renormalization problems can appear in this approach when the cutoff is removed, which, however, are not difficulties of principle but only of technical nature. This activity is of fundamental importance to have precise predictions in order to understand the limits of the standard model.

### Finite-Temperature QCD and the Deconfinement Transition

The static thermodynamics of a system of fields is described by the partition function

$$Z_T = \text{tr}[\exp(-H/T)] \qquad [26]$$

It is easy to show that $Z_T$ is equal to the Euclidean Feynman integral on the imaginary time interval $(0, 1/T)$ with boundary conditions in time periodic for bosons and antiperiodic for fermions. Indeed, the

Boltzmann factor is formally an imaginary time evolution by 1/*T*. A lattice of extension $L_t L_S^3$ with $L_s \gg L_t$ provides the partition function at a temperature $T = 1/aL_t$, if $a$ is the lattice spacing in physical units.

Finite-temperature simulations are important to investigate the transition from the phase in which color is confined to a phase in which quarks and gluons can propagate as free particles. This phase is called deconfined phase or quark gluon plasma.

Big experiments at Brookhaven and at CERN are looking for this phase transition in high-energy collisions between heavy nuclei, but no definite evidence has yet been produced for it. Lattice simulations instead definitely prove that such a transition exists. For pure SU(3) gauge theory (quenched) at $T \approx 270\,\text{MeV}$, a first-order phase transition is observed, at which the string tension vanishes. In a more realistic theory with dynamical quarks, a transition is also observed at $T \approx 160\,\text{MeV}$, where chiral symmetry, which is spontaneously broken at zero temperature, is restored. This transition is also associated to deconfinement even if, in the presence of light quarks, the string tension does not exist. Indeed, when pulling apart a quark and an antiquark, an instability for production of quark–antiquark pairs sets in when the potential energy becomes large enough, which physically manifests itself as a production of light mesons. An alternative order parameter is needed. The possibility of defining alternative order parameters is discussed in next section.

The equation of state can also be studied relating internal energy to pressure, which is useful to understand heavy ion collisions.

From the features of the deconfinement transition, information can be extracted on the mechanisms by which QCD confines color.

A connected issue is the behavior of QCD at nonzero baryon density or chemical potential. The corresponding thermodynamics is described by a grand canonical ensemble

$$Z_\mu = \text{tr}[\exp[-(H + \mu N)/T]] \qquad [27]$$

where $N = \int d^3x \psi^\dagger \psi$ is the baryon number operator and $\mu$ the chemical potential. In the process of converting the partition function $Z_\mu$ into a Feynman integral, the term $H$ at the exponent of eqn [27] generates the Euclidean action, which is real. The term proportional to $N$ becomes imaginary. The integral is well defined, but the analogy with a four-dimensional statistical mechanics is broken, the effective Hamiltonian being non-Hermitian and no sampling can be made. Approximate methods have been developed, but the problem is open. Exploring numerically the region of phase space with $\mu \neq 0$ would be interesting, since a rich structure is expected, which could be relevant to dense systems such as neutron stars.

### Mechanisms of Color Confinement

Understanding how QCD manages to confine color is one of the most fascinating problems in field theory.

To prove confinement, one should, in principle, prove that, at zero temperature, no gauge-invariant quasilocal operator exists, carrying nontrivial color and obeying cluster property at large distances. This proof is not known. There exists evidence form lattice simulations that a string tension exists, as discussed before. In any case, a guess can be made of the physical mechanism of confinement. If confinement is an absolute property reflecting a symmetry property of the vacuum, an order parameter should exist which discriminates between confined and deconfined phase, and the transition between the two phases has to be a true transition. Observing a crossover in some part of the boundary between the two phases would disprove this view. A lattice determination of the order of the deconfining transition is therefore of fundamental importance.

A possible mechanism of confinement proposed by G 't Hooft is dual superconductivity of the vacuum: dual means interchange of electric with magnetic with respect to ordinary superconductors. In the same way as the magnetic field is constrained into Abrikosov flux tubes in an ordinary superconductor, the chromoelectric field acting between a quark and an antiquark would be constrained into flux tubes by a dual Meissner effect producing an energy proportional to the distance, or a string tension.

This mechanism can be investigated by lattice simulations, by checking if any magnetically charged operator exists whose vacuum expectation value is nonzero in the confined phase signaling condensation of magnetic charges and zero in the deconfined phase. Progress has been made in this direction which, however, is not yet conclusive. Chromoelectric flux tubes between $q$–$\bar{q}$ pairs are observed in lattice field configurations.

### Topology

Euclidean QCD admits classical solutions with finite action and with a nontrivial topology which makes them stable. These solutions, known as instantons or multi-instantons, realize a mapping of the three-dimensional sphere at infinity on the gauge group, and the topological charge is the winding number of this mapping. The Jacobian of this mapping is the Chern

current $K_\mu$ and its divergence $\partial_\mu K_\mu(x) \equiv Q(x)$ is the density of topological charge. $Q = \int d^4x\, Q(x)$ is the topological charge which has integer values. Explicitly,

$$Q(x) = \frac{g^2}{16\pi^2} \mathrm{tr}[G_{\mu\nu} G^*_{\mu\nu}] \qquad [28]$$

with $G^*_{\mu\nu} = (1/2)\epsilon_{\mu\nu\rho\sigma} G_{\rho\sigma}$ the dual field strength tensor.

$Q(x)$ plays an important role in hadron physics, being related to the anomaly of the flavor singlet axial current $J^5_\mu = \sum_f \bar{\psi}\gamma^5\gamma_\mu\psi_f$. $J^5_\mu$ is conserved at the classical level in the chiral limit $m_f = 0$, but this symmetry does not survive quantization. In fact,

$$\partial_\mu J^5_\mu = 2N_f Q(x) \qquad [29]$$

A consequence of eqn [29] is the high mass $m_{\eta'} \approx 1\,\mathrm{GeV}$ of the flavor singlet partner $\eta'$ of the pseudoscalar flavor octet. An $N_c \to \infty$ argument by Witten and Veneziano relates $m_{\eta'}$ to the response of the quenched (no quark) vacuum to topological excitation, the topological susceptibility $\chi \equiv \int d^4x < 0|TQ(x)Q(0)|0 >$. The relation is

$$\frac{2N_f}{f_\pi^2}\chi = [m^2_{\eta'} + m^2_\eta - 2m^2_K][1 + O(1/N_c)] \qquad [30]$$

This approximate relation has been checked on the lattice. $\chi$ has been determined by different methods which agree in confirming it. This is an important verification of QCD.

Instantons are stable solutions in the continuum, approximately stable in the lattice discretized version. A cooling procedure which locally freezes short-distance quantum fluctuations would leave the instantons untouched if they were stable. On the lattice the instanton is stable anyhow if the distance in correlation reached by the local cooling procedure is small compared to the size of the instanton: cooling is indeed a diffusion process and the distance involved grows as the square root of the number of cooling iterations. Instanton configurations can nicely be exposed by cooling.

*See also:* Anomalies; Quantum Chromodynamics; Renormalization: General Theory; Spin Foams; Symmetry Breaking in Field Theory.

## Further Reading


Creutz M (1983) *Quarks, Gluons and Lattices*. Cambridge: Cambridge University Press.

Feynman RP (1948) Space–time approach to nonrelativistic quantum mechanics. *Reviews of Modern Physics* 20: 367.

Feynman RP and Hibbs AR (1965) *Quantum Mechanics and Path Integral*. New York: McGraw-Hill.

Ginsparg PH and Wilson KG (1982) A remnant of chiral symmetry on the lattice. *Physical Review D* 25: 2649.

Gottlieb S, Liu W, Toussaint D, Renken RL, and Sugar RL (1987) Hybrid-molecular-dynamics algorithms for the numerical simulation of QCD. *Physical Review D* 35: 2531.

Kogut J and Susskind L (1975) Hamiltonian formulation of Wilson's lattice gauge theories. *Physical Review D* 11: 395.

't Hooft G (1981) Topology of the gauge condition and new confinement phases in non-abelian gauge theories. *Nuclear Physics B* 190: 455.

Veneziano G (1979) U(1) without instantons. *Nuclear Physics B* 159: 213.

Wilson KG (1974) Confinement of quarks. *Physical Review D* 10: 2445–2459.

Wilson KG (1983) The renormalization group and critical phenomena (Nobel Lecture). *Reviews of Modern Physics* 55: 583.

Witten E (1979) Current algebra theorems for the U(1) goldstone boson. *Nuclear Physics B* 156: 269.


# Leray–Schauder Theory and Mapping Degree


**J Mawhin**, Université Catholique de Louvain, Louvain-la-Neuve, Belgium




## Introduction

The Leray–Schauder theory gives a powerful and versatile continuation method for proving the existence, multiplicity, and bifurcation of solutions of nonlinear operator, differential and integral equations. Let $X$ and $Y$ be topological spaces, $A \subset X$, $f: X \to Y$, a continuous mapping, and $y \in Y$. The fundamental idea of a continuation method to solve the equation $f(x) = y$ in $A$ consists in embedding it into a one-parameter family of equations

$$F(x, \lambda) = z(\lambda) \qquad [1]$$

where the continuous functions $F: X \times [0,1] \to Y$, $z: [0,1] \to Y$ are chosen in such a way that $F(\cdot, 1) = f$, $z(1) = y$ and

1. equation $F(x, 0) = z(0)$ has a nonempty set of solutions in $A$;
2. one of those solutions at least can be continued into a solution in $A$ of [1] for each $\lambda \in [0,1]$.

Simple examples show that Assertion 2 can be violated when all solutions of [1] leave $A$ after some

$\lambda^* \in ]0,1[$. A way to avoid such a situation consists in "closing the boundary," through the "boundary condition":

$$F(x,\lambda) \neq z(\lambda) \quad \text{for each} \quad (x,\lambda) \in \partial A \times [0,1]$$

When this condition is satisfied, Assertion 2 can still fail when two existing solutions for $\lambda$ small disappear after coalescing at some $\lambda_0 < 1$. Losing all solutions through this process can be eliminated by reinforcing Assumption 1 into

2′. Equation $F(x,0) = z(0)$ has a "robust" nonempty set of solutions in $A$.

This statement can be made precise through the concept of topological degree of a mapping, an "algebraic" count of the number of its zeros. In a finite-dimensional setting, this concept was introduced by Kronecker for smooth mappings and by Brouwer for continuous mappings. Its extension by Leray and Schauder to some classes of mappings in Banach spaces made much wider applications to nonlinear differential and integral equations possible.

## Topological Degree of a Mapping

If $U \subset \mathbb{R}^n$ is a bounded open set, $z \in \mathbb{R}^n$ and $F : \bar{U} \to \mathbb{R}^n$ is a $C^1$ mapping such that $z \notin F(\partial U)$ and $\det F'(x) \neq 0$ on $F^{-1}(z)$, the Brouwer degree $\deg_B[F,U,z]$ is defined (analytically) by

$$\begin{aligned}\deg_B[F,U,z] &:= \sum_{x \in F^{-1}(z)} \operatorname{sign} \det F'(x) \\ &= \sum_{x \in F^{-1}(z)} (-1)^{\sigma(x)}\end{aligned}$$

where $\sigma(x)$ is the sum of the multiplicities of the negative eigenvalues of $F'(x)$. The case of a continuous $F$ such that $z \notin F(\partial U)$ is treated by approximating $F$ through mappings of the above type, and showing that the corresponding degrees stabilize to an unique value, defining $\deg_B[F,U,z]$ in the general case. This number remains the same under sufficiently small perturbations of $F$ and/or $z$, which expresses the "robustness" mentioned above. When $n = 2$ and $U$ is bounded by a closed Jordan curve, then $\deg_B[F,U,0]$ is nothing but the winding number of $F/\|F\|$ along $\partial U$.

Leray and Schauder have extended Brouwer degree to the important class of compact perturbations of identity in a normed space. A compact mapping $f : A \to B$ between metric spaces is a continuous mapping on $A$ such that $f(A)$ is relatively compact. If $f : A \to B$ is continuous and compact on each bounded $B \subset A$, $f$ is called "completely continuous" on $A$.

If $X$ is a real normed space, $U \subset X$ an open bounded set, $f : \bar{U} \to X$ compact, and $z \notin (I - f)(\partial U)$, the Leray–Schauder degree $\deg_{LS}[I - f, U, z]$ of $I - f$ in $U$ over $z$ is constructed from Brouwer degree by approximating the compact mapping $f$ over $\bar{U}$ by mappings $f_\epsilon$ with range in a finite-dimensional subspace $X_\epsilon$ of $X$ containing $z$. One shows that the values of the Brouwer degrees $\deg_B[(I - f_\epsilon)|_{X_\epsilon}, U \cap X_\epsilon, z]$ stabilize for sufficiently small positive $\epsilon$ to a common value which defines $\deg_{LS}[I - f, U, z]$.

Again, this topological degree is an algebraic count of the number of elements of $(I - f)^{-1}(z)$, equal to 0 when $z \notin (I - f)(U)$. When $f$ is of class $C^1$, and $I - f'(x)$ invertible at each fixed point $x \in (I - f)^{-1}(z)$, $(I - f)^{-1}(z)$ is finite and the Leray–Schauder formula holds:

$$\deg_{LS}[I - f, U, z] = \sum_{x \in (I-f)^{-1}(z)} (-1)^{\sigma(x)} \qquad [2]$$

where $\sigma(x)$ is the sum of the algebraic multiplicities of the eigenvalues of $f'(x)$ contained in $[1, +\infty]$.

Let $I = [0,1]$. For $A \subset X \times I$, and $\lambda \in I$, we write $A_\lambda = \{x \in X : (x,\lambda) \in A\}$. The Leray–Schauder degree inherits the basic properties of Brouwer degree:

1. *Additivity*. If $U = U_1 \cup U_2$, where $U_1$ and $U_2$ are open and disjoint, and if $z \notin (I - f)(\partial U_1) \cup (I - f)(\partial U_2)$, then
$$\begin{aligned}\deg_{LS}[I - f, U, z] = {} & \deg_{LS}[I - f, U_1, z] \\ & + \deg_{LS}[I - f, U_2, z]\end{aligned}$$
2. *Existence*. If $\deg_{LS}[I - f, U, z] \neq 0$, then $z \in (I - f)(U)$.
3. *Homotopy invariance*. Let $\Omega \subset X \times I$ be a bounded open set, and let $F : \bar{\Omega} \to X$ be compact. If $x - F(x,\lambda) \neq z$ for each $(x,\lambda) \in \partial\Omega$, then $\deg_{LS}[I - F(\cdot,\lambda), \Omega_\lambda, z]$ is independent of $\lambda$.

In particular, if $a$ is an isolated fixed point of $f$, and $B(a,r)$ denotes the open ball of center $a$ and radius $r$, $\deg_{LS}[I - f, B(a,r), 0]$ is defined and independent of $r$ for sufficiently small $r > 0$. Its value is called the "Leray–Schauder index" of $I - f$ at $a$, and denoted by $\operatorname{ind}_{LS}[I - f, a]$.

## Fixed-Point Theorems for Compact Perturbations of Identity in a Normed Space

An important application of Leray–Schauder degree is the obtention of general fixed point theorems for compact mappings in normed spaces based on

continuation along a parameter. If $F: A \subset X \times I \to X$, we denote by $\Sigma^A$ the (possibly empty) solution set defined by

$$\Sigma^A = \{(x, \lambda) \in A : x = F(x, \lambda)\}$$

Let $\Omega \subset X \times I$ be a bounded open set and $F: \bar{\Omega} \to X$ be a compact mapping. The general Leray–Schauder fixed-point theorem goes as follows:

**Theorem** *If the following conditions hold*:

(i) $\Sigma^{\bar{\Omega}} \cap \partial\Omega = \emptyset$ (*a priori estimate*)
(ii) $\deg_{LS}[I - F(\cdot, 0), \Omega_0, 0] \neq 0$ (*degree condition*), *then* $\Sigma^{\bar{\Omega}}$ *contains a continuum* $\mathcal{C}$ *along which* $\lambda$ *takes all values in* $I$. In other words, $\Sigma^{\bar{\Omega}}$ contains a compact connected subset $\mathcal{C}$ connecting $\Sigma_0^{\bar{\Omega}}$ to $\Omega_1$. If one refines Assumption (ii) into
(iii) $\Sigma_0^{\bar{\Omega}}$ *is a finite nonempty set* $\{a_1, \ldots, a_\mu\}$ and $\mathrm{ind}_{LS}[I - F(\cdot, 0), a_1] \neq 0$, the conclusion takes the form of an "alternative": *if assumptions* (i) *and* (iii) *hold*, *then* $(a_1, 0)$ *belongs either to a continuum in* $\Sigma^{\bar{\Omega}}$ *containing one of the points* $(a_2, 0), \ldots, (a_\mu, 0)$, *or to a continuum in* $\Sigma^{\bar{\Omega}}$ *along which* $\lambda$ *takes all the values in* $I$.

Condition (iii) automatically holds in the following important special case: If $\Sigma^{\bar{\Omega}} \cap \partial\Omega = \emptyset$, $F(\cdot, 0) = 0$, and $0 \in \Omega_0$, then $\Sigma^{\bar{\Omega}}$ contains a continuum $\mathcal{C} \ni (0, 0)$ along which $\lambda$ takes all values in $I$. When dealing with the fixed-point problem $x = f(x)$ with $f: \bar{U} \subset X \to X$ compact, $U$ open and bounded, a natural choice is $F(x, \lambda) = \lambda f(x), \Omega = U \times I$, giving the statement: If $0 \in U$ and if $x \neq \lambda f(x)$ for each $(x, \lambda) \in \partial U \times I$, then $\{(x, \lambda) \in \bar{U} \times I : x = \lambda f(x)\}$ contains a continuum $\mathcal{C} \ni (0, 0)$ along which $\lambda$ takes all values in $I$.

Condition (i) requires the *a priori* knowledge of the localization of the solution set $\Sigma^{\bar{\Omega}}$ and is in general very difficult to check. An important special case occurs when $\Sigma^X$ is *a priori* bounded: if $F$ is completely continuous on $X \times I$, $F(\cdot, 0) = 0$, and $\Sigma^X \subset B(r) \times I$ for some $r > 0$, then $\Sigma^X$ contains a continuum $\mathcal{C} \ni (0, 0)$ along which $\lambda$ takes all values in $I$. Its special case with $F(\lambda, x) = \lambda f(x)$ can be stated as Schaefer's alternative: Let $f: X \to X$ be completely continuous. Then either there exists, for each $\lambda \in [0, 1]$, at least one $x \in X$ such that $x = \lambda f(x)$, or the fixed point set $\{x \in X: x = \lambda f(x), 0 < \lambda < 1\}$ is unbounded in $X$. Schaefer's alternative is equivalent to the following Schauder fixed-point theorem:

**Theorem** *Any compact mapping* $f: \overline{B(r)} \to \overline{B(r)}$ *has a fixed point*.

A simple consequence of Schauder's theorem is that, for any continuous and bounded $g: \mathbb{R} \to \mathbb{R}$, any open bounded $D \subset \mathbb{R}^n$, any $\lambda$ different from an eigenvalue of $-\Delta$ on $D$ with Dirichlet boundary conditions, the nonlinear Dirichlet problem

$$\begin{aligned} \Delta u + \lambda u + g(u) &= h(x) \quad \text{in } D \\ u &= 0 \quad \text{on } \partial D \end{aligned}$$

has a weak solution for each $h \in L^2(D)$.

An interesting consequence of Leray–Schauder theorem with $\Sigma^X$ *a priori* bounded is that, for any bounded domain $D \subset \mathbb{R}^n$ with $\partial D$ of class $C^2$, the Dirichlet problem for the equation of surfaces with constant mean curvature $\Lambda$

$$\begin{aligned} (1 + \|\nabla u\|^2)\Delta u - \sum_{i,j=1}^{n} \partial_i u \, \partial_j u \, \partial_{ij}^2 u \\ = n\Lambda(1 + \|\nabla u\|^2)^{3/2} \end{aligned}$$

has a unique solution for arbitrary smooth boundary data if and only if the mean curvature of the boundary $\partial D$ is everywhere greater than $[n/(n-1)]|\Lambda|$.

The use of auxiliary continuous functionals gives a fixed-point theorem in the absence of *a priori* bounds:

**Theorem** (Capietto–Mawhin–Zanolin). *Let* $\Omega \subset X \times I$ *be an open set and* $F: \bar{\Omega} \to X$ *be completely continuous. If* $\Sigma_0^{\bar{\Omega}}$ *is bounded*, $\deg_{LS}[I - F(\cdot, 0), U_0, 0] \neq 0$ *for some open bounded neighborhood* $U_0$ *of* $\Sigma_0^{\bar{\Omega}}$, *and if there exists a continuous mapping* $\varphi: X \times I \to \mathbb{R}_+$, *proper on* $\Sigma^{\bar{\Omega}}$, *and* $c_- < \min_{\Sigma_0^{\bar{\Omega}}} \varphi(\cdot, 0) \leq \max_{\Sigma_0^{\bar{\Omega}}} \varphi(\cdot, 0) < c_+$ *such that* $\Sigma^{\Omega} \notin \{c_-, c_+\}$ *and* $\Sigma^{\partial\Omega} \notin [c_-, c_+]$, *then* $\Sigma^{\bar{\Omega}}$ *contains a continuum* $\mathcal{C}$ *along which* $\lambda$ *takes all values in* $I$.

This result implies, for example, that for $g: \mathbb{R} \to \mathbb{R}$ continuous, odd and superlinear ($\lim_{|u| \to \infty} g(u)/u = +\infty$), and $p: [0, 1] \times \mathbb{R}^2$ with at most linear growth in $u$ and $u'$ at infinity, the two-point boundary-value problem

$$u'' + g(u) = p(t, u, u'), \quad u(0) = u(1) = 0$$

has, for all sufficiently large $j$, at least one solution $u_j$ having exactly $j + 1$ zeros on $[0, 1]$, and $\|u_j\|_{C^1} \to \infty$ if $j \to \infty$.

## Extensions of Leray–Schauder degree

Fixed-point theorems for operators between suitable nonlinear spaces can also be proved using topological continuation arguments. For example, if $C \subset X$ is a nonempty convex set, one has the following extension of a result of the previous section to mappings in $C$: if $U \subset C$ is open and bounded, $F: \mathrm{cl}_C U \times I \to C$ compact and such that $x \neq F(x, \lambda)$ for each $(x, \lambda) \in \partial_C U \times I$, $F(\cdot, 0) = x_0 \in U$, then

$F(\cdot,\lambda)$ has a fixed point in $U$ for each $\lambda \in I$. The special case where $C$ is a wedge is useful in finding positive solutions of nonlinear differential or integral equations. For nonlinear spaces, the degree has to be replaced by the fixed-point index, which generalizes both the "Hopf–Lefschetz number" and Leray–Schauder degree.

The Leray–Schauder degree also has been extended to other classes of operators. Compact operators can be replaced by $k$-set-contractive or condensing mappings $f$, with respect to various measures of noncompactness, and fixed-point problems can be replaced by problems of the form $x \in F(x)$ for multivalued mappings $F$. Equivariant degree theories have been developed when $U$ is invariant and $f$ equivariant with respect to the action of some compact Lie group $G$ on $X$. The special case of $G = S^1$ is of special importance in the study of periodic solutions of autonomous differential systems. Degree theories have also been constructed for various classes of mappings between two different Banach spaces or manifolds, which include monotone-like and nonlinear Fredholm operators. We just describe a simple but useful situation in this direction.

Many differential equations, when expressed as equations in an abstract space, do not have the fixed-point form but can be written as $Lx = Nx$ with $L: D(L) \subset X \to Z$ linear, $N: \bar{U} \to Z$, $X$ and $Z$ real normed spaces. If $L$ is invertible, the equation is trivially equivalent to the fixed-point problem $x = L^{-1}Nx$, to which Leray–Schauder theory can be applied when $L^{-1}N$ is compact. The situation is more delicate when $L$ has no inverse. If $L$ is a linear Fredholm mapping of index zero (its range $R(L)$ is closed and has a finite codimension equal to the dimension of its null space $N(L)$), the set $\mathcal{F}(L)$ of linear continuous mappings of finite rank $A: X \to Z$ such that $L + A: D(L) \to Z$ is a bijection is nonempty and the compactness of $(L+A)^{-1}G$ does not depend upon the choice of $A \in \mathcal{F}(L)$. $G$ is then called "$L$-compact" on $E$, and "$L$-completely continuous" on $E$ when compact on each bounded set of $E$.

The following continuation theorem for perturbed Fredholm mapping of index zero holds.

**Theorem** *Let $\Omega \subset X \times I$ be open and bounded, $L: D(L) \subset X \to Z$ linear Fredholm of index zero, $N: \bar{\Omega} \to Z$ $L$-compact, and let $\Sigma = \{(x,\lambda) \in (D(L) \times I) \cap \bar{\Omega} : Lx = N(x,\lambda)\}$. If*

(i) $\Sigma \cap \partial\Omega \neq \emptyset$ (*a priori estimate*),
(ii) $N(\bar{\Omega}_0 \times \{0\}) \subset Y$, *with* $Y \oplus R(L) = Z$ (*transversality condition*), *and*
(iii) $\deg_B[N(\cdot\,,0)|_{\ker L},\ \Omega_0 \cap \ker L,\ 0] \neq 0$ (*degree condition*)

*then $\Sigma$ contains a continuum $\mathcal{C}$ along which $\lambda$ takes all values in $I$.*

When dealing with equation $Lx = f(x)$ with $f$ $L$-completely continuous, an interesting special case of the above result follows from the choice $N(x,\lambda) = \lambda f(x) + (1-\lambda)Qf(x)$, with $Q: Z \to Z$ a projector such that $N(Q) = R(L)$. In this case, the homotopy is equivalent to

$$Lx = \lambda f(x) \quad (\lambda \in ]0,1])$$
$$Qf(x) = 0;\ x \in N(L) \quad (\lambda = 0)$$

An application (among many) of this result, for $g: \mathbb{R} \to \mathbb{R}$ continuous such that $-\infty < \limsup_{u \to -\infty} g(u) < \liminf_{u \to +\infty} g(u) < +\infty, D \subset \mathbb{R}^n$ open, bounded, $\lambda_k$ an eigenvalue of the Dirichlet problem for $-\Delta$ on $D$, is the weak solvability of the nonlinear problem

$$\Delta u + \lambda_k u + g(u) = h(x) \quad \text{in } D$$
$$u = 0 \quad \text{on } \partial D$$

for each $h \in L^2(D)$ such that

$$\int_D h(x)\varphi(x)\,\mathrm{d}x < \left[\lim \sup_{u \to -\infty} g(u)\right] \times \int_D \varphi^+(x)\,\mathrm{d}x - \left[\lim \inf_{u \to +\infty} g(u)\right] \int_D \varphi^-(x)\,\mathrm{d}x$$

for all eigenfunctions $\varphi$ associated to $\lambda_k$. The addition of the nonlinearity $g$ "widens" the range $\{h \in L^2(D) : \int_D h\varphi = 0\}$ of the corresponding linear problem.

## Bifurcation Theory

Leray–Schauder degree is a powerful tool in bifurcation theory, where, given a family $\mathcal{F}$ of solutions, one tries to detect and analyze other ones branching or bifurcating from $\mathcal{F}$. Consider the equation

$$x = \lambda Lx + R(x,\lambda) \qquad [3]$$

in a real normed space $X$, where $L: X \to X$, linear, and $R: X \times \mathbb{R} \to X$ are completely continuous, and $R(0,\lambda) = 0$ for each $\lambda \in \mathbb{R}$. Thus, $\{(0,\lambda) : \lambda \in \mathbb{R}\}$ is the trivial solution set of [3]. A bifurcation point $(\lambda^*, 0)$ for [3] is the limit of a sequence $(\lambda_k, x_k)$ of solutions of [3] in $\mathbb{R}\backslash\{0\}$.

If

$$\lim_{x \to 0} \frac{\|R(x,\lambda)\|}{\|x\|} = 0 \quad \text{uniformly on bounded } \lambda\text{-sets} \qquad [4]$$

it is easy to prove that if $(\lambda^*, 0)$ is a bifurcation point for [3], then $\lambda^*$ is a characteristic value (reciprocal of an eigenvalue) of $L$. Leray–Schauder theory gives a partial

converse to this result known as Krasnosel'skii's bifurcation theorem:

**Theorem** *For each real characteristic value* $\lambda^*$ *of* $L$ *with odd algebraic multiplicity*, $(\lambda^*, 0)$ *is a bifurcation point of* [3]. Of fundamental importance in the proof is the special case of [2] with $f = L$ and $N(I - L) = \{0\}$.

Another fruitful concept is Krasnosel'skii's bifurcation from infinity. We say $(\lambda^*, \infty)$ is a bifurcation point for [3] if there exists a sequence $(\lambda_n, x_n)$ of solutions of [3] such that $\lambda_n \to \lambda^*$ and $\|x_n\| \to \infty$. The corresponding bifurcation result goes as follows (Krasnosel'skii): if

$$\lim_{\|x\|\to\infty} \frac{\|R(x,\lambda)\|}{\|x\|} = 0 \quad \text{uniformly on bounded } \lambda\text{-sets} \qquad [5]$$

then, for each real characteristic value $\lambda^*$ of $L$ with odd algebraic multiplicity, $(\lambda^*, \infty)$ is a bifurcation point of [3].

Global versions of Krasnosel'skii's theorems can be given, whose statements are reminiscent of Leray–Schauder's alternative theorem. Let $\mathcal{S}$ denote the closure in $\mathbb{R} \times X$ of the set of $(\lambda, x) \in \mathbb{R} \times (X \setminus \{0\})$ satisfying [3]. For bifurcation from zero, one has Rabinowitz global bifurcation theorem:

**Theorem** *If* [4] *holds and* $\lambda^*$ *is a real characteristic value of* $L$ *with odd algebraic multiplicity, then* $\mathcal{S}$ *contains a component* $\mathcal{C}$ *which either is unbounded, or contains* $(\lambda^{**}, 0)$, *where* $\lambda^{**} \neq \lambda^*$ *is a characteristic value of* $L$.

As an application, one can show that the nonlinear Sturm–Liouville problem

$$-(p(x)u')' + q(x)u = \lambda a(x)u + h(x,u,u',\lambda) \quad (x \in ]0,1[)$$
$$a_0 u(0) + b_0 u'(0) = a_1 u(1) + b_1 u'(1) = 0$$

with $p \in C^1$ positive, $q$, $a$, $h$ continuous, $a$ positive, $(a_0^2 + b_0^2)(a_1^2 + b_1^2) \neq 0$ and $h(x,u,v) = o(|u| + |v|)$ if $|u| + |v| \to 0$ uniformly on compact $\lambda$-intervals, has, for each $k \in \mathbb{N}$, an unbounded component of solution $\mathcal{C}_k$ in $\mathbb{R} \times C^1([0,1])$ emanating from $(\lambda_k, 0)$, with $\lambda_k$ an eigenvalue of the problem with $h \equiv 0$ (Rabinowitz).

One has also global bifurcation from infinity: if [5] holds and if $\lambda^*$ is a real characteristic value of $L$ with odd algebraic multiplicity, then [3] has an unbounded component of solutions $\mathcal{D}$ which contains $(\lambda^*, \infty)$.

*See also:* Bifurcation Theory; Bifurcations in Fluid Dynamics; Bifurcations of Periodic Orbits; Minimal Submanifolds; Minimax Principle in the Calculus of Variations; Partial Differential Equations: Some Examples; Riemann–Hilbert Problem; Topological Defects and Their Homotopy Classification; Viscous Incompressible Fluids: Mathematical Theory.

## Further Reading

Cronin J (1964) *Fixed Points and Topological Degree in Nonlinear Analysis*. Providence: American Mathematical Society.

Deimling K (1985) *Nonlinear Functional Analysis*. Berlin: Springer.

Fitzpatrick P, Martelli M, Mawhin J, and Nussbaum R (1993) *Topological Methods for Ordinary Differential Equations*. Berlin: Springer.

Fonseca I and Gangbo W (1995) *Degree Theory in Analysis and Applications*. Oxford: Oxford Science Publisher.

Geba K and Rabinowitz P (1985) *Topological Methods in Bifurcation Theory*. Montréal: Presses Univ. Montréal.

Granas A and Dugundji J (2003) *Fixed Point Theory*. New York: Springer.

Ize J and Vignoli A (2003) *Equivariant Degree Theory*. Berlin: de Gruyter.

Krasnosel'skii MA (1963) *Topological Methods in the Theory of Nonlinear Integral Equations*. Oxford: Pergamon.

Krasnosel'skii MA and Zabreiko PP (1984) *Geometrical Methods of Nonlinear Analysis*. Berlin: Springer.

Krawcewicz W and Wu J (1997) *Theory of Degrees with Applications to Bifurcations and Differential Equations*. New York: Wiley.

Leray J and Schauder J (1934) Topologie et équations fonctionnelles. *Annales Scientifiques de l'Ecole Normale Supérieure* 51(3): 45–78.

Lloyd NG (1978) *Degree Theory*. Cambridge: Cambridge University Press.

Matzeu M and Vignoli A (eds.) (1995–97) *Topological Nonlinear Analysis. Degree, Singularity and Variations* I, II. Basel: Birkhäuser.

Mawhin J (1979) *Topological Degree Methods in Nonlinear Boundary Value Problems*. Providence: American Mathematical Society.

Petryshyn WW (1995) *Generalized Topological Degree and Semilinear Equations*. Cambridge: Cambridge University Press.

Rothe E (1986) *Introduction to Various Aspects of Degree Theory in Banach Spaces*. Providence: American Mathematical Society.

Schwartz JT (1969) *Nonlinear Functional Analysis*. New York: Gordon and Breach.

Zeidler E (1986–88) *Nonlinear Functional Analysis and Its Applications*, vols. I–IV. New York: Springer.

**Lie Bialgebras** *see* Classical *r*-Matrices, Lie Bialgebras, and Poisson Lie Groups

# Lie Groups: General Theory

**R Gilmore**, Drexel University, Philadelphia, PA, USA



## Introduction

Local continuous transformations were introduced by Lie as a tool for solving ordinary differential equations. In this program, he followed the spirit of Galois, who used finite groups to develop algorithms for solving algebraic equations (the general quadratic, cubic, and quartic), or else to prove that some equations (the generic quintic) could not be solved by quadrature.

Lie's work led eventually to the definition and study of Lie groups. Lie groups are beautiful in their own right – so beautiful that they have been studied independently of their origin as a tool for solving differential equations and studying the special functions determined by certain classes of these equations.

## Lie Groups

Lie groups exist at the interface of the two great divisions of mathematics: algebra and topology. Their algebraic properties derive from the group axioms. Their geometric properties arise from the parametrization of the group elements by points in a differentiable manifold. The rigidity of these structures arises from the continuity requirements imposed on the group composition and inversion maps.

The algebraic axioms are standard.

**Definition** A group $G$ consists of a set $g_i, g_j, g_k, \ldots \in G$ together with a combinatorial operation $\circ$ that satisfy the four axioms:

(i) *Closure*. If $g_i \in G, g_j \in G$, then $g_i \circ g_j \in G$.
(ii) *Associativity*. If $g_i, g_j, g_k \in G$, then $(g_i \circ g_j) \circ g_k = g_i \circ (g_j \circ g_k)$.
(iii) *Identity*. There is a unique operation $e \in G$ that satisfies $e \circ g_i = g_i = g_i \circ e$.
(iv) *Inverse*. Every group operation $g_i \in G$ has an inverse, denoted $g_i^{-1}$, that satisfies $g_i \circ g_i^{-1} = e = g_i^{-1} \circ g_i$.

Lie groups have more structure than groups. In particular, each $g_i \in G$ is a point in an $n$-dimensional manifold $M^n$. That is, the subscript $i$ actually identifies a point $x \in M^n$, so that we can write $g_i = g(x)$ or most simply $g_i = x$. The group multiplication can be expressed in the form $g_i \circ g_j = g_k \to g(x) \circ g(y) = g(z)$, where $x \in M^n$, $y \in M^n, z = \phi(x, y) \in M^n$. The group inversion map can be expressed in the form $g(x) \to g(x)^{-1} = g(y)$, $y = \psi(x) \in M^n$. The topological axioms for Lie groups can be taken as:

(v) *Continuity of composition*. The mapping $z = \phi(x, y)$ defined by the group composition law is differentiable.
(vi) *Continuity of inversion*. The mapping $y = \psi(x)$ defined by the group inversion law is differentiable.

The dimension of the Lie group is the dimension of the manifold that parametrizes the operations in the group.

The most familiar examples of Lie groups consist of $n \times n$ nonsingular matrices over the fields $R$, $C$, $Q$ of real numbers, complex numbers, and quaternions. For example, the set of $2 \times 2$ real unimodular matrices

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad ad - bc = 1$$

is a three-dimensional submanifold embedded in $R^{2^2} = R^4$.

## Matrix Lie Groups

Not every Lie group is a matrix group. Yet, it is a surprising and useful result that almost every Lie group encountered in physics is a matrix Lie group. These are all subgroups of the general linear groups $\mathrm{GL}(n; F)$ of $n \times n$ nonsingular matrices over the field $F$ ($R$, $C$, $Q$). These groups have real dimension $n^2 \times (1, 2, 4)$, respectively. The special linear subgroups $\mathrm{SL}(n; F)$ are defined as the subgroups of $n \times n$ matrices with determinant $+1$: $M \in \mathrm{SL}(n; F)$ if $\det M = +1$. This definition is problematic for quaternions, as they do not commute. To avoid this problem, it is useful to map quaternions into $2 \times 2$ complex matrices in the same way complex numbers can be mapped into $2 \times 2$ real matrices:

$$a + \mathrm{i}b \to \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$$

$$q_0 + \mathcal{I}q_1 + \mathcal{J}q_2 + \mathcal{K}q_3 \to \begin{bmatrix} q_0 + \mathrm{i}q_3 & \mathrm{i}q_1 + q_2 \\ \mathrm{i}q_1 - q_2 & q_0 - \mathrm{i}q_3 \end{bmatrix}$$

Here $(1, \mathrm{i})$ are basis vectors for $C^1$ considered as a real two-dimensional linear vector space,

$(1, \mathcal{I}, \mathcal{J}, \mathcal{K})$ are basis vectors for $Q^1$ considered as a real four-dimensional linear vector space, and $(a, b)$ and $(q_0, q_1, q_2, q_3)$ are all real. The squares of the imaginary quantities i and $\mathcal{I}, \mathcal{J}, \mathcal{K}$ are all –1: $i^2 = -1$; $\mathcal{I}^2 = \mathcal{J}^2 = \mathcal{K}^2 = -1$ and the imaginary quaternion basis elements anticommute: $\{\mathcal{I}, \mathcal{J}\} = \{\mathcal{J}, \mathcal{K}\} = \{\mathcal{K}, \mathcal{I}\} = 0$. The unimodular subgroup $SL(n; Q)$ of $GL(n; Q)$ is obtained by replacing each quaternion matrix element by a $2 \times 2$ complex matrix, setting the determinant of the resulting $2n \times 2n$ matrix group to +1, and then mapping each of the $n^2$ complex $2 \times 2$ matrices back to quaternions.

Many other important groups are defined by imposing linear or quadratic constraints on the $n^2$ matrix elements of $GL(n; F)$ or $SL(n; F)$. The compact metric-preserving groups $U(n; F)$ leave invariant lengths (preserve a positive-definite metric $g = I_n$) in linear vector spaces. The matrices $M \in U(n; F)$ satisfy $M^\dagger I_n M = I_n$. These conditions define the orthogonal groups $O(n) = U(n; R)$ and the unitary groups $U(n) = U(n; C)$. Their noncompact counterparts $O(p, q)$ and $U(p, q)$ leave invariant nonsingular indefinite metrics

$$g = I_{p,q} = \begin{bmatrix} I_p & 0 \\ 0 & -I_q \end{bmatrix}$$

in real and complex $n = (p + q)$-dimensional linear vector spaces: $M^\dagger I_{p,q} M = I_{p,q}$.

Intersections of matrix Lie groups are also Lie groups. The special metric-preserving groups are intersections of the special linear groups $SL(n; F) \subset GL(n; F)$ (with $F = Q$, $SL(n; Q)$ is defined as described above) and the metric-preserving subgroups $U(n; F) \subset GL(n; F)$:

$$\begin{aligned}
&SL(n; R) \cap U(n; R) = SO(n), && n(n-1)/2 \\
&SL(n; C) \cap U(n; C) = SU(n), && n^2 - 1 \\
&SL(n; Q) \cap U(n; Q) = Sp(n) = USp(2n), && n(2n+1)
\end{aligned}$$

The real dimensions of these groups are given in the right-hand column. Under the replacement of quaternions by $2 \times 2$ complex matrices, the group of $n \times n$ metric-preserving and unimodular matrices $Sp(n)$ over $Q$ is identified as $USp(2n)$, an isomorphic group of $2n \times 2n$ matrices over $C$.

Noncompact forms $SO(p, q)$, $SU(p, q)$, and $Sp(p, q) = USp(2p, 2q)$ are defined similarly.

The Lie group SU(2) rotates spin states to spin states in a complex two-dimensional linear vector space. It leaves lengths, inner products, and probabilities invariant. If an interaction is spin independent, only an invariant ("Casimir invariant") constructed from the spin operators can appear in the Hamiltonian. The same group can act in isospin space, rotating proton to neutron states. The Lie group SU(3) similarly rotates quark states or color states into quark states or color states, respectively. The Lie group SU(4) rotates spin–isospin states into themselves. The conformal group SO(4, 2) leaves angles but not lengths in spacetime invariant. It is the largest group that leaves the source-free Maxwell equations invariant. It is also the largest group that transforms all the (bound, scattering, and parabolic) hydrogen atom states into themselves.

Lie groups such as the Poincaré group (inhomogeneous Lorentz group) and the Galilei group have the matrix structures

$$\left[\begin{array}{cccc|c} & & & & t_1 \\ & O(3,1) & & & t_2 \\ & & & & t_3 \\ & & & & t_4 \\ \hline 0 & 0 & 0 & 0 & 1 \end{array}\right] \begin{bmatrix} x \\ y \\ z \\ ct \\ 1 \end{bmatrix}$$

$$\left[\begin{array}{ccc|cc} & & & v_1 & t_1 \\ & O(3) & & v_2 & t_2 \\ & & & v_3 & t_3 \\ \hline 0 & 0 & 0 & 1 & t_4 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right] \begin{bmatrix} x \\ y \\ z \\ t \\ 1 \end{bmatrix}$$

respectively. In these transformations $\boldsymbol{t} = (t_1, t_2, t_3)$ describes translations in the space ($x$-, $y$-, and $z$-) directions, $\boldsymbol{v} = (v_1, v_2, v_3)$ describes boosts, and $t_4$ resets clocks. The matrices in these defining matrix representations are reducible.

The Heisenberg covering group $H_4$ is a four-dimensional Lie group with a simple $3 \times 3$ matrix structure:

$$\text{Heisenberg covering group} = H_4 = \begin{bmatrix} 1 & l & d \\ 0 & n & r \\ 0 & 0 & 1 \end{bmatrix}, \quad n \neq 0$$

This matrix representation of $H_4$ is faithful but nonunitary.

## "Linearization" of a Lie Group

At the topological level, a Lie group is homogeneous. That is, every point in a manifold that parametrizes a Lie group looks like every other point. At the algebraic level, this is not true – the identity group operation $e$ is singled out as an exceptional group element. At the analytic level, the group composition law $z = \phi(x, y)$ is nonlinear, and can therefore be arbitrarily complicated.

The study of Lie groups is enormously simplified by exploiting these three observations. Specifically, it is useful to "linearize" the group multiplication law in the neighborhood of the identity. The linearization leads to a local Lie group. This is a linear vector space on which there is an additional structure. Once the local Lie group properties are known in the neighborhood of the identity, they are known everywhere else in the group, since the group is homogeneous.

A Lie group is linearized in the neighborhood of the identity by expressing an operator near the identity in the form $g(\epsilon) = I + \epsilon X$, where the local Lie group operator $\epsilon X = \delta x^i X_i$, the $X_i$ are $n$ linearly independent vector fields on the manifold $M^n$, and the small coordinates $\delta x^i$ measure the distance (in some rough sense) of $g(\epsilon)$ from the point that parametrizes the identity group operation $e = g(0)$. For another group operation $g(\delta Y) = I + \delta Y$ in the neighborhood of the identity, the following holds.

1. The product $g(\epsilon X)g(\delta Y) = (I + \epsilon X)(I + \delta Y) = I + (\epsilon X + \delta Y) + (\text{h.o.t})$ is in the local Lie group.
2. The commutator $g_i \circ g_j \circ g_i^{-1} \circ g_j^{-1}$ in the group leads to

$$\begin{aligned} &g(\epsilon X)g(\delta Y)g(\epsilon X)^{-1}g(\delta Y)^{-1} \\ &\quad = I + \tfrac{1}{2}\epsilon\delta(XY - YX) + \text{h.o.t} \\ &\quad = I + \tfrac{1}{2}\epsilon\delta[X, Y] + \text{h.o.t} \end{aligned}$$

   in the local Lie group.

The first condition shows that the local Lie group is a linear vector space. The $n$ vector fields $X_i$ can be chosen as a set of basis vectors in this space.

The second condition shows that the commutator of two vectors in this linear vector space is also in this linear vector space. The commutator endows this linear vector space with an additional combinatorial operation ("vector multiplication") and provides it with the structure of an algebra, called a Lie algebra.

**Definition** A Lie algebra $\mathfrak{la}$ consists of a set of operators $X, Y, Z, \ldots$, together with the operations of vector addition, scalar multiplication, and commutation $[X,Y]$ that satisfy the following three axioms:

(i) *Closure* (*linear vector space*). If $X, Y \in \mathfrak{la}$, $\alpha X + \beta Y \in \mathfrak{la}$ and $[X, Y] \in \mathfrak{la}$.
(ii) *Antisymmetry*. $[X, Y] = -[Y, X]$.
(iii) *Jacobi identity*. $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$.

The structure of a Lie algebra, or local Lie group, is summarized by the structure constants, defined in terms of the basis vectors $X_i$, by

$$[X_i, X_j] = c_{ij}{}^k X_k \quad \text{summation convention}$$

The structure constants $c_{ij}{}^k$ are components of a third-order tensor, covariant and antisymmetric in two indices ($c_{ij}{}^k = -c_{ji}{}^k$) and contravariant in the third. These components obey the Jacobi identity, which places a quadratic constraint on them:

$$c_{ij}{}^s c_{sk}{}^t + c_{jk}{}^s c_{si}{}^t + c_{ki}{}^s c_{sj}{}^t = 0$$

Linearization of a Lie group generates a Lie algebra. A Lie group can be recovered by the inverse process. This is the exponential operation. A group operation a finite distance from the origin (the point identified with the identity group operation) of the manifold that parametrizes the Lie group can be obtained from the limiting procedure ($\epsilon = 1/K \to 0$):

$$g(X) = \lim_{K\to\infty} \prod \left(I + \frac{1}{K}X\right)^K = e^X = \text{EXP}(X)$$

The exponential operation is well defined for real numbers, complex numbers, quaternions, $n \times n$ matrices over these fields, and vector fields.

A 1:1 correspondence between Lie groups and Lie algebras does not exist. Isomorphic Lie groups have isomorphic Lie algebras. But nonisomorphic Lie groups may also possess isomorphic Lie algebras. The best known examples of nonisomorphic Lie groups and their isomorphic Lie algebras are

$$\begin{aligned} &SO(3) \neq SU(2), && \mathfrak{so}(3) = \mathfrak{su}(2) \\ &SO(4) \neq SU(2) \times SU(2), && \mathfrak{so}(4) = \mathfrak{su}(2) + \mathfrak{su}(2) \\ &SO(5) \neq Sp(2) = USp(4), && \mathfrak{so}(5) = \mathfrak{sp}(2) = \mathfrak{usp}(4) \end{aligned}$$

There is a 1:1 correspondence between Lie algebras and "locally" isomorphic Lie groups. This has been extended to global Lie groups by a beautiful theorem due to E Cartan.

**Theorem** (Cartan) *There is a 1:1 correspondence between Lie algebras and simply connected Lie groups. Every Lie group with the same Lie algebra is either the simply connected* ("*universal covering*") *group or is the quotient of this universal covering group by one of its discrete invariant subgroups.*

This relation is summarized in Figure 1.

As a concrete example, the Lie algebra of SO(3), which is the group of real $3 \times 3$ matrices satisfying $M^\dagger I_3 M = I_3$ and $\det(M) = +1$, is spanned by the three "angular momentum vector

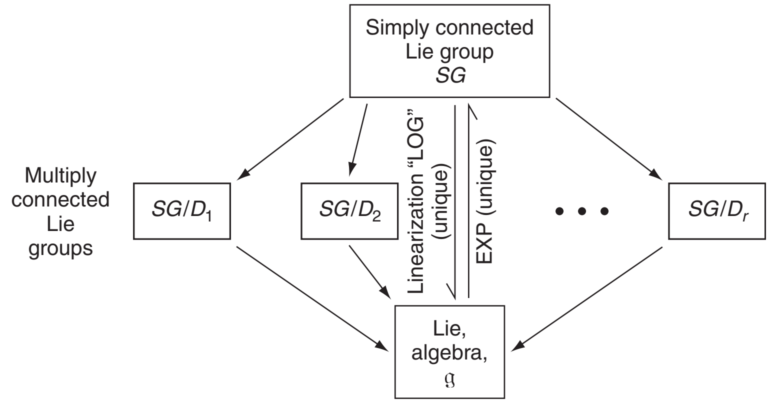



**Figure 1** Cartan's theorem states that there is a 1:1 correspondence between Lie algebras and simply connected Lie groups. All other Lie groups with this Lie algebra are quotients of the covering group by one of its discrete invariant subgroups $D_j \subseteq D_{\text{Max}}$. There is a relation between the discrete invariant subgroup $D_j$ and the homotopy group of $SG/D_j$. Reproduced with permission from Gilmore R (1974) *Lie Groups, Lie Algebras, and Some of Their Applications*. New York: Wiley.

fields" $L_i(x) = \epsilon_{ijk} x^j \partial_k$ or the three angular momentum matrices

$$L_1 = L_{23} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & +1 \\ 0 & -1 & 0 \end{bmatrix}$$

$$L_2 = L_{31} = -L_{13} = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ +1 & 0 & 0 \end{bmatrix}$$

$$L_3 = L_{12} = \begin{bmatrix} 0 & +1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The Lie group SU(2) is the group of complex $2 \times 2$ matrices satisfying $M^\dagger I_2 M = I_2$ and $\det(M) = +1$. Its Lie algebra is spanned by the three spin matrices $S_j = (\mathrm{i}/2)\sigma_j$, which are multiples of the Pauli spin matrices $\sigma_j$:

$$S_1 = \frac{\mathrm{i}}{2}\begin{bmatrix} 0 & +1 \\ +1 & 0 \end{bmatrix}, \quad S_2 = \frac{\mathrm{i}}{2}\begin{bmatrix} 0 & -i \\ +i & 0 \end{bmatrix}$$

$$S_3 = \frac{\mathrm{i}}{2}\begin{bmatrix} +1 & 0 \\ 0 & -1 \end{bmatrix}$$

The two Lie algebras are isomorphic as they share isomorphic commutation relations $[J_1, J_2] = -J_3$ (and cyclic), $J_j = L_j$ or $J_j = S_j$. The group SU(2) is simply connected. Its maximal discrete invariant subgroup $D$ consists of all multiples of the identity, $\alpha I_2$, so that $\alpha = \pm 1$. According to Cartan's theorem, $\mathrm{SO}(3) = \mathrm{SU}(2)/D_2, D_2 = \{I_2, -I_2\}$. The group SO(3) is doubly connected, with a two-element homotopy group.

## Matrix Lie Algebras

A deep theorem of Ado guarantees that every Lie algebra is equivalent to a matrix Lie algebra, even though the same is not true of Lie groups.

Sets of $n \times n$ matrices that close under vector addition, scalar multiplication, and commutation ($M_1 \in \mathfrak{la}, M_2 \in \mathfrak{la} \Rightarrow [M_1, M_2] = M_1 M_2 - M_2 M_1 \in \mathfrak{la}$) form matrix Lie algebras. The antisymmetry properties and Jacobi identity are guaranteed by matrix multiplication.

Lie algebras for the general linear groups $\mathrm{GL}(n; F)$ consist of $n \times n$ matrices over $F$. Lie algebras for the special linear groups $\mathrm{SL}(n; F)$ consist of traceless $n \times n$ matrices. The Lie algebras of the unitary groups consist of anti-Hermitian matrices. The Lie algebras of $\mathrm{U}(p, q; F)$ consist of matrices that obey

$$M^\dagger I_{p,q} + I_{p,q} M = 0, \quad M \in \mathfrak{u}(p, q; F)$$

The matrix Lie algebras of other matrix Lie groups are obtained by constructing the most general Lie group operation in the neighborhood of the identity by linearization. For example, the Lie algebra of the Heisenberg covering group $H_4$ is

$$\begin{bmatrix} 1 & l & d \\ 0 & n & r \\ 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & \delta l & \delta d \\ 0 & 1 + \delta n & \delta r \\ 0 & 0 & 1 \end{bmatrix}$$

$$\rightarrow I_3 + \delta n\, N + \delta r\, R + \delta l\, L + \delta d\, D$$

$$N \simeq a^\dagger a \qquad\qquad R \simeq a^\dagger$$

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$L \simeq a \qquad D \simeq I = [a, a^\dagger]$$
$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The four $3 \times 3$ matrices $N$, $R$, $L$, $D$ that span the Lie algebra $\mathfrak{h}_4$ of $H_4$ satisfy commutation relations isomorphic with the commutation relations satisfied by the photon operators ($a^\dagger a$, $a^\dagger$, $a$, $I = [a, a^\dagger]$). The $3 \times 3$ matrix representations of the group $H_4$ and the algebra $\mathfrak{h}_4$ are faithful. The representation of $H_4$ is nonunitary and that of $\mathfrak{h}_4$ is non-Hermitian.

There is a simple way to relate a large class of operator Lie algebras to matrix Lie algebras. If $A, B, C, \ldots$ belong to a Lie algebra of $n \times n$ matrices with $[A, B] = C$, the matrix-to-operator mapping

$$A \to \mathcal{A} = x^i A_i{}^j \partial_j$$

preserves commutation relations, for

$$\begin{aligned}[\mathcal{A}, \mathcal{B}] &= \left[x^i A_i{}^j \partial_j, x^r B_r{}^s \partial_s\right] \\ &= x^i A_i{}^j \left[\partial_j, x^r\right] B_r{}^s \partial_s - x^r B_r{}^s \left[\partial_s, x^i\right] A_i{}^j \partial_j \\ &= x^i A_i{}^j B_j{}^s \partial_s - x^r B_r{}^i A_i{}^j \partial_j = x^i [A, B]_i{}^j \partial_j = \mathcal{C}\end{aligned}$$

This relation depends on the bilinear products $x^i \partial_j$ satisfying commutation relations

$$\left[x^i \partial_j, x^r \partial_s\right] = x^i \partial_s \delta_j{}^r - x^r \partial_j \delta_s{}^i$$

These commutation relations are satisfied by products of creation and annihilation operators $a_i^\dagger a_j$ for either bosons ($b_i^\dagger b_j$) or fermions ($f_i^\dagger f_j$). These matrix-to-operator mappings can be extended to include bilinear products such as $x^i x^j$, $x^i \partial_j$, $\partial_i \partial_j$ and their boson and fermion counterparts $a_i a_j$, $a_i^\dagger a_j$, $a_i^\dagger a_j^\dagger$. For example, the vector fields associated with the operator $J_1$ for SO(3) and SU(2) are $x^i (L_1)_i{}^j \partial_j = x^2 \partial_3 - x^3 \partial_2$ and $u^i (S_1)_i{}^j \partial_j = (\mathrm{i}/2)(u^1 \partial_2 + u^2 \partial_1)$.

Boson and fermion bilinear products $a_i^\dagger a_j (1 \leq i, j \leq n)$ are isomorphic to $\mathfrak{u}(n)$. Boson bilinear products $b_i b_j$, $b_i^\dagger b_j$, $b_i^\dagger b_j^\dagger$ are isomorphic to $\mathfrak{usp}(2n)$ while fermion bilinear products $f_i f_j, f_i^\dagger f_j, f_i^\dagger f_j^\dagger$ are isomorphic to $\mathfrak{so}(2n)$.

## Structure of Lie Algebras

The study of Lie algebras is greatly facilitated by studying their structure. The structure is determined by the commutation properties of the Lie algebra.

### Invariant Subalgebra

If a Lie algebra has an invariant subalgebra, then the commutator of anything in the algebra with anything in the subalgebra is in the subalgebra. Suppose $\mathfrak{a}$ is a linear vector subspace of $\mathfrak{g}$. If $[\mathfrak{g}, \mathfrak{a}] \subseteq \mathfrak{a}$, then $\mathfrak{a}$ is an invariant subspace of $\mathfrak{g}$. In particular, $[\mathfrak{a}, \mathfrak{a}] \subseteq \mathfrak{a}$ and $\mathfrak{a}$ is therefore also a subalgebra of $\mathfrak{g}$: it is an invariant subalgebra in $\mathfrak{g}$.

**Example** The Lie algebra $\mathfrak{iso}(3)$ consists of the three rotation operators $L_{ij} = x^i \partial_j - x^j \partial_i$ and the three displacement operators $P_k = \partial_k$. The subset of displacement operators is an invariant subspace in $\mathfrak{iso}(3)$, since it is mapped into itself by all commutators. It is also a subalgebra in $\mathfrak{iso}(3)$. This particular invariant subalgebra is commutative.

### Solvable Algebra

If $\mathfrak{g}$ is a Lie algebra, the linear vector space obtained by taking all possible commutators of the operators in $\mathfrak{g}$ is called the "derived" algebra: $[\mathfrak{g}, \mathfrak{g}] = \mathfrak{g}^{(1)} \subseteq \mathfrak{g}$. If $\mathfrak{g}^{(1)} = \mathfrak{g}$, there is no point in continuing this process. If $\mathfrak{g}^{(1)} \subset \mathfrak{g}$, it is useful to define $\mathfrak{g} = \mathfrak{g}^{(0)}$ and to continue this process by defining $\mathfrak{g}^{(2)}$ as the derived algebra of $\mathfrak{g}^{(1)}$: $\mathfrak{g}^{(2)} = [\mathfrak{g}^{(1)}, \mathfrak{g}^{(1)}]$. We can continue in this way, defining $\mathfrak{g}^{(n+1)}$ as the algebra derived from $\mathfrak{g}^{(n)}$. Ultimately (for finite-dimensional Lie algebras), either $\mathfrak{g}^{(n+1)} = 0$ or $\mathfrak{g}^{(n+1)} = \mathfrak{g}^{(n)}$ for some $n$. If the former case occurs,

$$\mathfrak{g} = \mathfrak{g}^{(0)} \supset \mathfrak{g}^{(1)} \supset \mathfrak{g}^{(2)} \supset \cdots \supset \mathfrak{g}^{(n)} \supset \mathfrak{g}^{(n+1)} = 0$$

the Lie algebra $\mathfrak{g}^{(0)}$ is called solvable. Each algebra $\mathfrak{g}^{(i)}$ is an invariant subalgebra of $\mathfrak{g}^{(j)}$, $i > j$.

**Example** The Lie algebra spanned by the boson number, creation, annihilation, and identity operators is solvable. The series of derived algebras has dimensions 4, 3, 1, 0.

| $\mathfrak{g}^{(0)}$ | $\mathfrak{g}^{(1)}$ | $\mathfrak{g}^{(2)}$ | $\mathfrak{g}^{(3)}$ |
|---|---|---|---|
| $a^\dagger a$ | – | – | – |
| $a^\dagger$ | $a^\dagger$ | – | – |
| $a$ | $a$ | – | – |
| $I$ | $I$ | $I$ | – |

### Semidirect Sum Algebra

When a Lie algebra $\mathfrak{g}$ has an invariant subalgebra $\mathfrak{a}$, the linear vector space of the Lie algebra $\mathfrak{g}$ can be written as the direct sum of the linear vector subspace of the subalgebra $\mathfrak{a}$ plus a complementary subspace $\boldsymbol{b}$. The subspace $\boldsymbol{b}$ is generally not by itself a Lie algebra. The Lie algebra $\mathfrak{g}$ is written as a semidirect sum of the two subspaces. The semidirect

sum structure satisfies the commutation relations shown:

$$\mathfrak{g} = \boldsymbol{b} \wedge \mathfrak{a} \qquad \begin{array}{l} [\boldsymbol{b}, \boldsymbol{b}] \subseteq \boldsymbol{b} \wedge \mathfrak{a} \\ [\boldsymbol{b}, \mathfrak{a}] \subseteq \mathfrak{a} \\ [\mathfrak{a}, \mathfrak{a}] \subseteq \mathfrak{a} \end{array}$$

The subspace $\boldsymbol{b}$ can be given the structure of an algebra modulo the component of the commutator in $\mathfrak{a}$: $\mathfrak{b} = \mathfrak{g} \bmod \mathfrak{a}$.

**Example** The three-dimensional Lie algebra spanned by the photon operators $a^{\dagger}$, $a$, $I$ has a semidirect sum decomposition where $\boldsymbol{b}$ is spanned by $a^{\dagger}$, $a$ and $\mathfrak{a}$ is spanned by $I$. The subspace $\boldsymbol{b}$ is not closed under commutation, and $\mathfrak{a}$ is commutative. The Lie algebra $\mathfrak{iso}(3)$ also has the structure of a semidirect sum, with $\boldsymbol{b} = \mathfrak{b} = \mathfrak{so}(3)$ and the invariant subalgebra $\mathfrak{a}$ is spanned by the three displacement operators $P_k$.

#### Nonsemisimple Algebra

A Lie algebra is nonsemisimple if it has a solvable invariant subalgebra.

**Example** The Lie algebra spanned by bilinear products of photon creation and annihilation operators $a_i^{\dagger} a_j$, creation operators $a_i^{\dagger}$, annihilation operators $a_j$, and the identity operator $I (1 \le i, j \le n)$ is nonsemisimple. The solvable invariant subalgebra is spanned by the $2n + 2$ operators consisting of the single photon operators $a_i^{\dagger}, a_j$, the identity operator $I$, and the total number operator $\hat{n} = \sum_{i=1}^{n} a_i^{\dagger} a_i$.

#### Semisimple Algebra

A Lie algebra is semisimple if it has no solvable invariant subalgebras.

**Example** The Lie algebra $\mathfrak{so}(4)$ is semisimple. This Lie algebra has two invariant subalgebras, both isomorphic to $\mathfrak{so}(3)$. The direct sum decomposition

$$\mathfrak{so}(4) = \mathfrak{so}(3) + \mathfrak{so}(3)$$

is well known to physical chemists and is responsible for the dualities that exist between rotating and laboratory frame descriptions of molecular systems.

#### Simple Algebra

A Lie algebra is simple if it has no invariant subalgebras at all. The prettiest page in the theory of Lie groups is the classification theory of the simple Lie algebras. We turn to this subject now.

## Lie Algebra Tools

Two powerful tools have been developed for studying the structure of a Lie algebra. These are the regular representation and the Cartan–Killing form.

#### Regular Representation

This representation assigns the structure constants to a set of $n$ $n \times n$ matrices according to

$$X_{\alpha} \to R(X_{\alpha})_{\mu}{}^{\nu} = c_{\alpha\mu}{}^{\nu}, \qquad [X_{\alpha}, X_{\mu}] = c_{\alpha\mu}{}^{\nu} X_{\nu}$$

The matrices of the regular representation contain exactly as much information as the components of the structure tensor. They can be studied by standard linear algebra methods. For example, a secular equation can be used to put the commutation relations into canonical form.

The structure of the matrices of the regular representation determines the structure of the Lie algebra. The identification is carried out according to the usual rules of representation theory, as shown in Figure 2. If a basis $X_{\alpha}$ can be found in which all the matrices of the regular representation are simultaneously reducible, the algebra possesses an invariant subalgebra. If the representation is not fully reducible, the invariant subalgebra is solvable. If the regular representation is fully reducible, the algebra consists of the direct sum of two (or more) smaller, mutually commuting subalgebras. If the regular representation is irreducible, the algebra is simple.

If a Lie algebra is solvable ($\mathfrak{solv}$), all matrices in the regular representation can be transformed to upper triangular matrices. If the Lie algebra is nilpotent ($\mathfrak{nil} \subset \mathfrak{solv}$), the diagonal matrix elements in the upper triangular matrices are zero. The converses are also true.

#### Cartan–Killing Form

The Cartan–Killing form is a second-order symmetric tensor that is constructed from the third-order antisymmetric tensor $c_{\alpha\mu}{}^{\nu}$ by cross-contraction

$$g_{\alpha\beta} = c_{\alpha\mu}{}^{\nu} c_{\beta\nu}{}^{\mu} = g_{\beta\alpha} = \operatorname{tr} R(X_{\alpha}) R(X_{\beta}) = (X_{\alpha}, X_{\beta}) = (X_{\beta}, X_{\alpha})$$

The metric $g_{\alpha\beta}$ can be used to place an inner product $(X_{\alpha}, X_{\beta})$ on this linear vector space. This inner product is not necessarily positive definite.

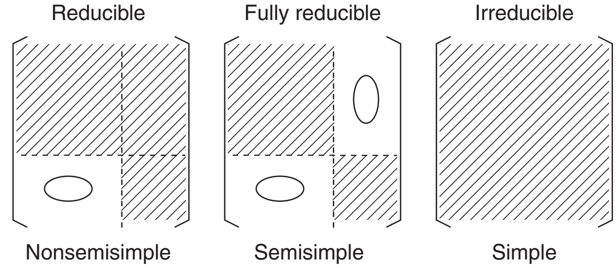



**Figure 2** When the regular matrix representation of a Lie algebra is reducible, fully reducible, or irreducible, the Lie algebra is nonsemisimple, semisimple, or simple.

The matrix $g_{\alpha\beta}$ can also be treated by standard linear algebra methods. Since it is real and symmetric, it can be diagonalized. If there are $n_-$ negative eigenvalues, $n_+$ positive eigenvalues, and $n_0$ vanishing eigenvalues ($n = n_- + n_+ + n_0$), the Lie algebra has a corresponding linear vector space decomposition of the form

$$\mathfrak{g} = \mathfrak{g}_- + \mathfrak{g}_+ + \mathfrak{g}_0$$

The inner product is positive definite on the subspace $\mathfrak{g}_+$ and negative definite on $\mathfrak{g}_-$. We call $\mathfrak{g}_0$ the singular subspace. The subspace $\mathfrak{g}_0$ is closed under commutation and in fact is a nilpotent invariant subalgebra of $\mathfrak{g}$.

## Decomposition of Lie Algebras

The most general Lie algebra $\mathfrak{g}$ is the semidirect sum of a semisimple Lie algebra $\mathfrak{SS}$ and a solvable invariant subalgebra $\mathfrak{Solv}$:

$$\mathfrak{g} = \mathfrak{SS} \wedge \mathfrak{Solv} \qquad \begin{aligned} [\mathfrak{SS}, \mathfrak{SS}] &= \mathfrak{SS} \\ [\mathfrak{SS}, \mathfrak{Solv}] &\subseteq \mathfrak{Solv} \\ [\mathfrak{Solv}, \mathfrak{Solv}] &\subset \mathfrak{Solv} \end{aligned}$$

The decomposition of $\mathfrak{g}$ into its component parts is accomplished by a simple two-step algorithm.

1. Compute the Cartan–Killing metric for $\mathfrak{g}$ and determine the singular subspace. If there is none, stop. If the dimension of $\mathfrak{g}_0 > 0, \mathfrak{nil} = \mathfrak{g}_0$ is the maximal nilpotent invariant subalgebra of $\mathfrak{g}$.
2. Compute the structure constants of the Lie algebra $\mathfrak{g}' = \mathfrak{g} - \mathfrak{nil} = \mathfrak{g} \bmod \mathfrak{nil} = \mathfrak{g}/\mathfrak{nil}$, the Cartan–Killing metric tensor on $\mathfrak{g}'$, and the decomposition $\mathfrak{g}' = \mathfrak{g}'_- + \mathfrak{g}'_+ + \mathfrak{g}'_0$. Then $\mathfrak{a} = \mathfrak{g}'_0$ is abelian and invariant in $\mathfrak{g}'$. In fact, $\mathfrak{a}$ is the largest abelian invariant subalgebra in $\mathfrak{g}'$.

The algorithm stops here, for the algebra $\mathfrak{g}'' = \mathfrak{g}' \bmod \mathfrak{a} = \mathfrak{g}'/\mathfrak{a} = \mathfrak{g}'_- + \mathfrak{g}'_+$ has no singular subspace under its Cartan–Killing metric.

Under this algorithm, the decomposition of $\mathfrak{g}$ into its semisimple part and its maximal solvable invariant subalgebra is

$$\mathfrak{g} = \left(\mathfrak{g}'_- + \mathfrak{g}'_+\right) \wedge \left(\mathfrak{g}'_0 \wedge \mathfrak{g}_0\right)$$

The maximum solvable invariant subalgebra $\mathfrak{Solv}$ in $\mathfrak{g}$ is the semidirect sum of $\mathfrak{a}$ and $\mathfrak{nil}$: $\mathfrak{Solv} = \mathfrak{g}'_0 \wedge \mathfrak{g}_0 = \mathfrak{a} \wedge \mathfrak{nil}$. In addition, $\mathfrak{SS} = \mathfrak{g} \bmod \mathfrak{Solv} = \mathfrak{g}/\mathfrak{Solv} = \mathfrak{g}'_- + \mathfrak{g}'_+$. The subspace $\mathfrak{g}'_-$ is closed under commutation and exponentiates into a compact subgroup of $G'$. The subspace $\mathfrak{g}'_+$ exponentiates to a noncompact coset in $G'$ that is simply connected.

Every element in a semisimple Lie algebra can be expressed as the commutator of two elements in the Lie algebra. In this sense, a semisimple algebra reproduces itself under commutation.

To illustrate this algorithm, we tear apart the eight-dimensional Lie algebra spanned by the photon operators $a_i^\dagger a_j, 1 \le i, j \le 2$ and $a_3^\dagger a_3$, $a_3^\dagger$, $a_3$, $I$, where the photon operators obey $[a_i, a_j^\dagger] = \delta_{ij} I$. The regular representative of the general linear combination $X = \sum_{ij} m_{ij} a_i^\dagger a_j + n a_3^\dagger a_3 + r a_3^\dagger + l a_3 + \delta I$ is

$$R(X) = \begin{bmatrix} 0 & & -m_{12} & m_{21} & & & & \\ & 0 & m_{12} & -m_{21} & & & & \\ -m_{21} & m_{21} & +m_{11} - m_{22} & 0 & & & & \\ m_{12} & -m_{12} & 0 & -m_{11} + m_{22} & & & & \\ & & & & 0 & & & \\ & & & & & n & & l \\ & & & & & & -n & -r \\ & & & & & & & 0 \end{bmatrix} \begin{matrix} a_1^\dagger a_1 \\ a_2^\dagger a_2 \\ a_1^\dagger a_2 \\ a_2^\dagger a_1 \\ a_3^\dagger a_3 \\ a_3^\dagger \\ a_3 \\ I \end{matrix}$$

The Cartan–Killing inner product is the trace of the square of this matrix:

$$\begin{aligned} (X, X) = \operatorname{tr} R(X)^2 &= 2(m_{11} - m_{22})^2 \\ &+ 8 m_{12} m_{21} + 2n^2 \end{aligned}$$

The subspace $\mathfrak{g}_0$ is spanned by $a_1^\dagger a_1 + a_2^\dagger a_2$, $a_3^\dagger$, $a_3, I$, leaving the four operators $a_1^\dagger a_1 - a_2^\dagger a_2, a_1^\dagger a_2$, $a_2^\dagger a_1$, $a_3^\dagger a_3$ to span $\mathfrak{g}'$. A simple calculation shows that $\mathfrak{g}'_0$ is spanned by $a_3^\dagger a_3$. As a result:

| *Subspace* | *Spanned by* |
|---|---|
| $\mathfrak{g}'_+$ | $a_1^\dagger a_1 - a_2^\dagger a_2$, $\frac{1}{\sqrt{2}}(a_1^\dagger a_2 + a_2^\dagger a_1)$ |
| $\mathfrak{g}'_-$ | $\frac{1}{\sqrt{2}}(a_1^\dagger a_2 - a_2^\dagger a_1)$ |
| $\mathfrak{g}'_0$ | $a_3^\dagger a_3$ |
| $\mathfrak{g}_0$ | $a_1^\dagger a_1 + a_2^\dagger a_2, a_3^\dagger, a_3, I$ |

The Lie algebra is the direct sum $\mathfrak{g} = \mathfrak{sl}(2; R) + \mathfrak{u}(1) + \mathfrak{h}_4$.

## Structure of Semisimple Lie Algebras

The Cartan–Killing metric $g_{\alpha\beta}$ is nonsingular on a semisimple Lie algebra. The metric and its inverse $g^{\alpha\beta}$, can be used to raise and lower indices. In particular, the tensor whose components are $c_{\alpha\beta\gamma} = c^{\mu}_{\alpha\beta} g_{\mu\gamma}$ is third-order antisymmetric: $c_{\alpha\beta\gamma} = c_{\beta\gamma\alpha} = c_{\gamma\alpha\beta} = -c_{\beta\alpha\gamma} \ldots$. Classification of semisimple Lie algebras is equivalent to classifying such tensors.

Another useful way to describe semisimple Lie algebras is to search for a canonical structure for the commutation relations. A useful canonical form is an eigenvalue form

$$[X, Y] = \lambda Y$$

In a basis $X_i$, with $X = x^i X_i$ and $Y = y^j X_j$, this equation reduces to a standard eigenvalue equation for the regular representation

$$\sum_j \sum_k y^j \left( R(x^i X_i)_j{}^k - \lambda \delta_j{}^k \right) X_k = 0$$

Thus, the search for a standard form for the commutation relations reduces to a study of the secular equation

$$\det(R(X) - \lambda I) = \sum_{j=0}^{n} (-\lambda)^{n-j} \phi_j(X) = 0 \qquad [1]$$

The coefficients $\phi_j(X)$ are homogeneous polynomials of degree $j$ in the coefficients $x^i$ of $X = x^i X_i$.

In order to extract maximum information from this secular equation, a generic vector $X \in \mathfrak{g}$ is chosen. Such a choice minimizes all degeneracies. With a generic choice of $X \in \mathfrak{g}$, it is useful to define the rank, $l$, of the Lie algebra $\mathfrak{g}$ as:

1. the number of functionally independent coefficients $\phi_j(X)$ in the secular equation;
2. the number of independent roots, $\alpha_1, \alpha_2, \ldots, \alpha_l$ of the secular equation;
3. the dimension of the subspace $H \subset \mathfrak{g}$ that commutes with $X$; and
4. the number of independent (Casimir) operators that commute with all $X_i : \mathcal{C}_j(X) = \phi_j(x^i \to X_i)$: $[\mathcal{C}_j(X), X_i] = 0$.

For example, for $\mathfrak{so}(3)$ or $\mathfrak{su}(2)$, the secular equation for $X = x^i X_i$ is

$$\det\left[\begin{bmatrix} 0 & x_3 & -x_2 \\ -x_3 & 0 & x_1 \\ x_2 & -x_1 & 0 \end{bmatrix} - \lambda I_3\right]$$
$$= (-\lambda)^3 + (-\lambda)\phi_2(x) = 0$$

where $\phi_2(x) = x_1^2 + x_2^2 + x_3^2$. The rank is $l = 1$. There is one independent coefficient $\phi_2(x)$ and one independent root of this equation, $\alpha_1 = \sqrt{-\delta_{ij} x^i x^j} = \mathrm{i}\sqrt{x \cdot x}$. The only linear operators that commute with $X$ are scalar multiples of $X$. There is one independent homogeneous operator that commutes with all generators $X_i$, obtained by the substitutions $x^i \to L_i$ (for $\mathfrak{so}(3)$) or $x^i \to S_i$ (for $\mathfrak{su}(2)$):

$$\mathcal{C}^2(L) = \phi_2(x_i \to L_i) = L_1^2 + L_2^2 + L_3^2$$

The secular equation [1] is over the field of real numbers. This is not an algebraically closed field. There is no guarantee that the number of independent functions $\phi_j(x)$ in the secular equation is equal to the number of (real) roots of this equation until we extend the field from $R$ to $C$, which is algebraically closed. As a result, the classification of semisimple Lie algebras is done over complex numbers. After the complex extensions of the simple Lie algebras have been classified, their different inequivalent real forms can be determined.

## Root Spaces

When the secular equation for the regular representation of a generic element in a Lie algebra is solved, the commutation relations can be put into a simple and elegant canonical form. This canonical form depends on the rank, $l$, of the Lie algebra, not the dimension, $n$, of the Lie algebra. This provides a very useful simplification, as $n \sim l^2$.

For this canonical form, the independent roots $\alpha_1(x), \alpha_2(x), \ldots, \alpha_l(x)$ are gathered into a single vector $\boldsymbol{\alpha}$ with $l$ components. The vectors $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_l)$ are called root vectors. The root vectors exist in an $l$-dimensional space on which a positive-definite inner product can be defined. The root vectors for a rank-$l$ semisimple Lie algebra $\mathfrak{g}$ span this Euclidean space. The basis vectors of $\mathfrak{g}$ can be identified with the roots in the root space.

The roots in a root space have the following properties:

1. A positive-definite metric can be placed on the root space.
2. The vector **0** is a root.
3. The root **0** is $l$-fold degenerate.
4. If $\boldsymbol{\alpha}$ is a root and $c\boldsymbol{\alpha}$ is a root, $c = \pm 1, 0$.
5. If $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are roots,

$$\boldsymbol{\beta}' = \boldsymbol{\beta} - \frac{2\boldsymbol{\alpha} \cdot \boldsymbol{\beta}}{\boldsymbol{\alpha} \cdot \boldsymbol{\alpha}} \boldsymbol{\alpha}$$

   is also a root and $2\boldsymbol{\alpha} \cdot \boldsymbol{\beta} / \boldsymbol{\alpha} \cdot \boldsymbol{\alpha}$ is an integer, $n_1$. In fact, $\boldsymbol{\beta}'$ is the root obtained by reflecting $\boldsymbol{\beta}$ in the hyperplane orthogonal to $\boldsymbol{\alpha}$.
6. The set of reflections generated by nonzero roots itself forms a group, the Weyl group of the Lie algebra.

7. The angle between roots $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is determined by

$$\cos^2(\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{\boldsymbol{\alpha}\cdot\boldsymbol{\beta}}{\boldsymbol{\alpha}\cdot\boldsymbol{\alpha}}\frac{\boldsymbol{\alpha}\cdot\boldsymbol{\beta}}{\boldsymbol{\beta}\cdot\boldsymbol{\beta}} = \frac{n_1}{2}\frac{n_2}{2} = 0, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, 1$$

   The integers $n_1, n_2$ for noncolinear roots are constrained by $|n_1 n_2| < 4$.

8. The relative lengths of the roots are determined by the angles between them:

| $\cos^2(\theta(\boldsymbol{\alpha},\boldsymbol{\beta}))$ | $\theta(\boldsymbol{\alpha},\boldsymbol{\beta})$ | $\boldsymbol{\alpha}\cdot\boldsymbol{\alpha}/\boldsymbol{\beta}\cdot\boldsymbol{\beta}$ |
|---|---|---|
| 3/4 | $30°, 150°$ | $3^{\pm1}$ |
| 2/4 | $45°, 135°$ | $2^{\pm1}$ |
| 1/4 | $60°, 120°$ | 1 |

9. When the roots are normalized so that

$$\sum_{\alpha\neq0}\alpha_i\alpha_j = \delta_{ij} \quad \text{or} \quad \sum_{\alpha\neq0}\boldsymbol{\alpha}\cdot\boldsymbol{\alpha} = l$$

   the commutation relations can be placed in the canonical form presented in the next section.

It is possible to build up all possible root space diagrams using an "Aufbau" construction. We start with a rank-1 root space. This consists of three roots in $R^1$: $\boldsymbol{\alpha}, 0, -\boldsymbol{\alpha}$.

To construct rank-2 root spaces, a new noncolinear root $\boldsymbol{\beta}$ is adjoined to the two nonzero roots. The new root and the old roots span $R^2$. The new root can only have a limited set of angles with the roots already present. The set of roots $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ is completed by reflection in hyperplanes orthogonal to all roots present. If any pair of roots violates the angle conditions, the result is not a root space. In this way, the rank-2 root spaces $G_2(30°), B_2 = C_2(45°), A_2(60°)$, and $D_2 = A_1 + A_1(90°)$ are constructed from $A_1$. Proceeding in this way, it is possible to construct rank-3 root spaces $(B_3, C_3, A_3 = D_3)$ from the rank-2 root spaces, the rank-4 root spaces from the rank-3 root spaces, and so forth. Ultimately, there are four unending chains $A_n, B_n, C_n, D_n$ and five exceptional root spaces $G_2, F_4, E_6, E_7, E_8$. The rank-2 root spaces are shown in Figure 3 and the rank-3 root spaces are shown in

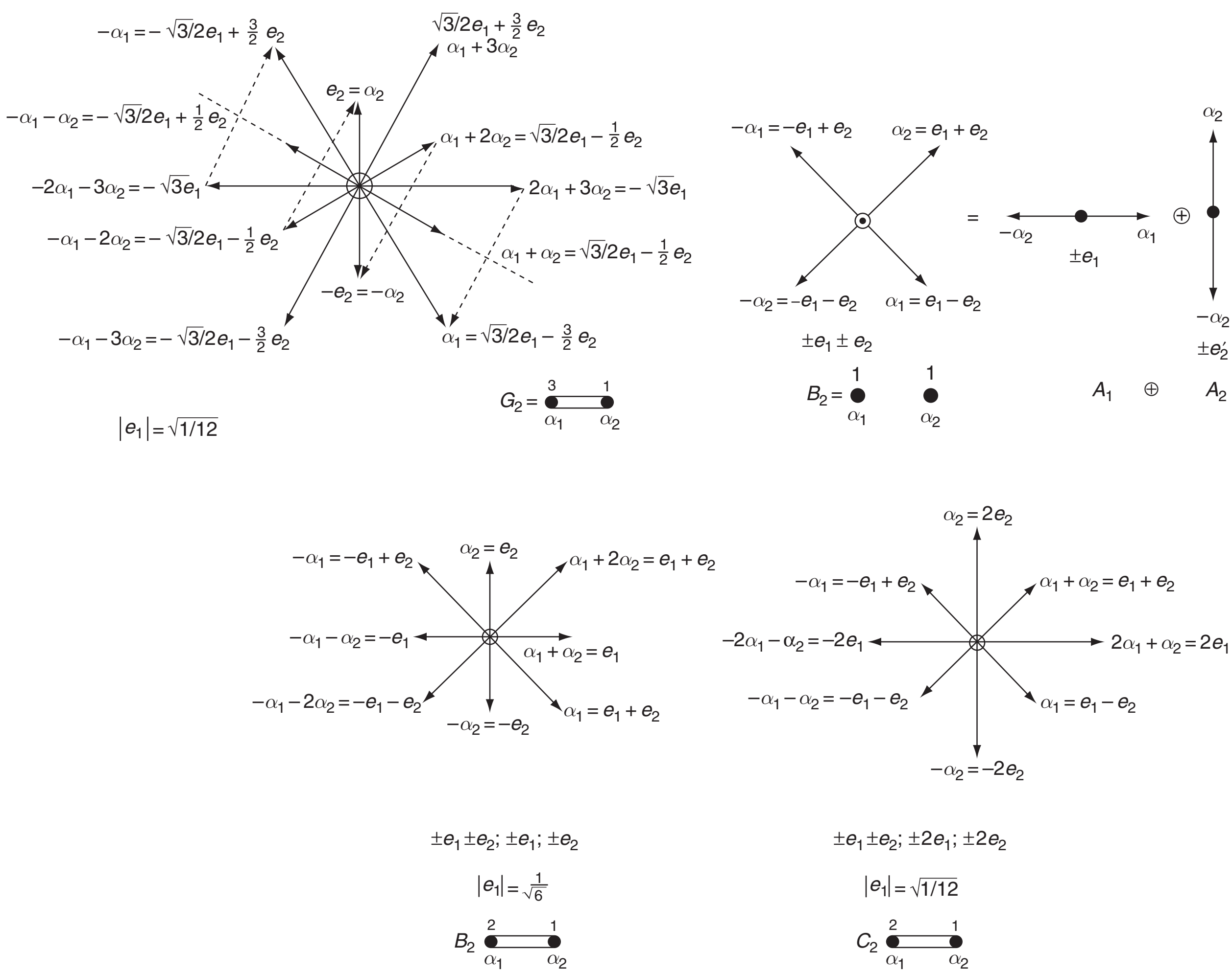



**Figure 3** Rank-2 root spaces: $G_2$ 30°, $B_2 = C_2$ 45°, $A_2$ 60°, $D_2 = A_1 + A_1$ 90°.

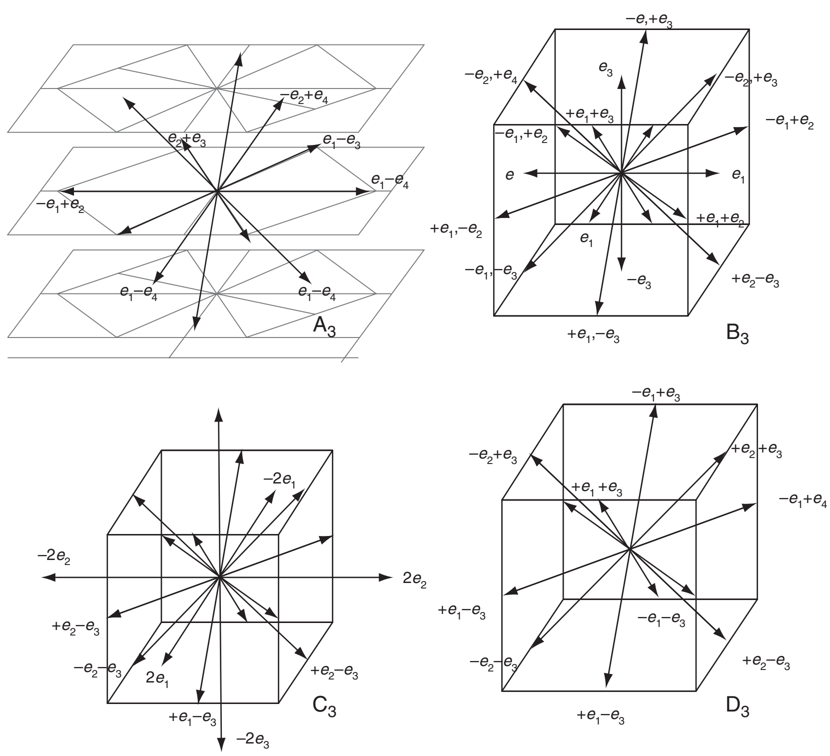



**Figure 4** Rank-3 root spaces: $A_3, B_3, C_3, D_3 = A_3$.

Figure 4. The normalization factors (cf. point (9) above) are shown for the rank-2 root spaces in Figure 3.

## Canonical Commutation Relations

The canonical commutation relations are expressed in terms of root vectors. The $l$ operators in $\mathfrak{g}$ with the $l$-fold degenerate root vector $\mathbf{0}$ are $H_1, H_2, \ldots, H_l$. These $l$ operators mutually commute. In a matrix Lie algebra, they can be taken as simultaneously commuting diagonal matrices. Associated with each nonzero root $\boldsymbol{\alpha} \neq 0$, there is exactly one basis vector, $E_{\boldsymbol{\alpha}}$, in $\mathfrak{g}$. The canonical commutation relations are expressed in terms of the roots as follows:

$$
\begin{aligned}
[H_i, H_j] &= 0 \qquad\qquad 1 \le i, j \le l \\
[H_i, E_{\boldsymbol{\alpha}}] &= \alpha_i E_{\boldsymbol{\alpha}} \\
[E_{\boldsymbol{\alpha}}, E_{-\boldsymbol{\alpha}}] &= \boldsymbol{\alpha} \cdot \boldsymbol{H} \\
[E_{\boldsymbol{\alpha}}, E_{\boldsymbol{\beta}}] &= \begin{cases} N_{\boldsymbol{\alpha\beta}} E_{\boldsymbol{\alpha}+\boldsymbol{\beta}} & \boldsymbol{\alpha} + \boldsymbol{\beta} \text{ a root} \\ 0 & \boldsymbol{\alpha} + \boldsymbol{\beta} \text{ not a root} \end{cases}
\end{aligned}
$$

The structure constants $N_{\boldsymbol{\alpha\beta}}$ are determined from a recursion relation derived from a chain of roots $\boldsymbol{\beta} - m\,\boldsymbol{\alpha}, \boldsymbol{\beta} - (m-1)\boldsymbol{\alpha}, \ldots, \boldsymbol{\beta} + (n-1)\boldsymbol{\alpha}, \boldsymbol{\beta} + n\boldsymbol{\alpha}$, where $\boldsymbol{\beta} - (m+1)\boldsymbol{\alpha}$ and $\boldsymbol{\beta} + (n+1)\ \boldsymbol{\alpha}$ are not roots (cf. Figure 5). The structure constants are

$$N^2_{\boldsymbol{\alpha},\boldsymbol{\beta}} = \tfrac{1}{2} n(1+m)(\boldsymbol{\alpha} \cdot \boldsymbol{\alpha})$$

The operators $\boldsymbol{H}$ and $E_{\boldsymbol{\alpha}}$ are often called diagonal and shift operators, respectively. They are generalizations of the shift operators $J_3$ and $J_\pm$ of angular momentum theory. The general idea is as follows. Since the operators $H_i$ mutually commute, the matrices $\Gamma(H_i)$ representing these operators can be chosen as diagonal in any matrix representation.

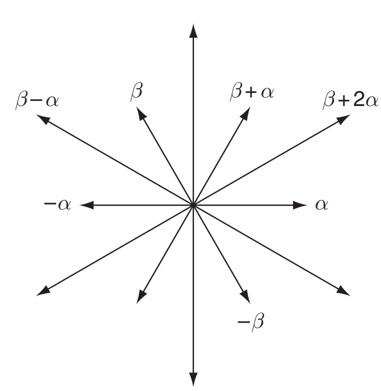



**Figure 5** An $\alpha$ chain containing $\beta$.

The action of any of these operators on a basis vector in this representation is $H_i|\boldsymbol{m}\rangle = m_i|\boldsymbol{m}\rangle$. The operator $E_\alpha$ shifts the eigenvalue of $\boldsymbol{H}$ according to

$$\boldsymbol{H}(E_\alpha|\boldsymbol{m}\rangle) = ([\boldsymbol{H}, E_\alpha] + E_\alpha \boldsymbol{H})|\boldsymbol{m}\rangle = (\boldsymbol{\alpha} + \boldsymbol{m})(E_\alpha|\boldsymbol{m}\rangle)$$

In this sense the operators $E_\alpha$ act on basis vectors $|\boldsymbol{m}\rangle$ in such a way that the eigenvalue $\boldsymbol{m}$ is shifted by $\boldsymbol{\alpha}$ to $\boldsymbol{m} + \boldsymbol{\alpha}$.

For the simple classical Lie algebras, the roots can be expressed in terms of an orthogonal Euclidean basis set as shown in Table 1 and Figures 3 and 4 for the rank-2 and rank-3 root spaces. The roots for the five remaining inequivalent simple Lie algebras ("exceptional" algebras) are shown in Table 2.

The diagonal and shift operators for several of the classical Lie algebras can be related to bilinear products of boson or fermion creation and annihilation operators. For $\mathfrak{u}(n)$, the bilinear products $a_i^\dagger a_j$ are related to $E_\alpha$ with $\boldsymbol{\alpha} = \boldsymbol{e}_i - \boldsymbol{e}_j$, $1 \le i \ne j \le n$, and $H_i = a_i^\dagger a_i$. This holds for either boson or fermion operators. For $\mathfrak{sp}(2n; R)$, we have the identifications with bilinear products of boson operators as follows: $+\boldsymbol{e}_i + \boldsymbol{e}_j \leftrightarrow b_i^\dagger b_j^\dagger$, $+\boldsymbol{e}_i - \boldsymbol{e}_j \leftrightarrow b_i^\dagger b_j$, $-\boldsymbol{e}_i - \boldsymbol{e}_j \leftrightarrow b_i b_j$, and $H_i = b_i^\dagger b_i$. In particular, $+2\boldsymbol{e}_i \leftrightarrow b_i^{\dagger 2}$ and $-2\boldsymbol{e}_i \leftrightarrow b_i^2$. For $\mathfrak{so}(2n)$, we have the identifications with bilinear products of fermion operators as follows: $+\boldsymbol{e}_i + \boldsymbol{e}_j \leftrightarrow f_i^\dagger f_j^\dagger$, $+\boldsymbol{e}_i - \boldsymbol{e}_j \leftrightarrow f_i^\dagger f_j$, $-\boldsymbol{e}_i - \boldsymbol{e}_j \leftrightarrow f_i f_j$, and $H_i = f_i^\dagger f_i$. In particular, $f_i^\dagger f_i^\dagger = f_i^2 = 0$. These identifications make it a relatively simple matter to construct unitary matrix representations of the compact Lie groups SU($n$) that are symmetric or antisymmetric, of USp($2n$) that are symmetric, and of SO($2n$) that are antisymmetric (bosons $\leftrightarrow$ symmetric, fermions $\leftrightarrow$ antisymmetric).

## Dynkin Diagrams

Every root in a rank-$l$ root space can be represented as a linear combination of $l$ "basis roots." These basis roots can be chosen in such a way that all coefficients are integers. In fact, the basis roots can be chosen so that all linear combinations that are roots involve only positive integers (and zero) or only negative integers and zero. This comes about because every shift operator $E_\delta$ can be written as a multiple commutator

$$E_\delta \sim [E_\alpha, [E_\beta, E_\gamma]], \quad \boldsymbol{\delta} = \boldsymbol{\alpha} + \boldsymbol{\beta} + \boldsymbol{\gamma}$$

One simple way to construct such a basis set of fundamental roots is to construct an $(l-1)$-dimensional plane through the origin of the root space that contains no nonzero roots, and choose as $l$ fundamental roots the $l$ roots on one side of this hyperplane that are closest to it. For the classical simple Lie algebras, the fundamental roots are:

| *Root Space* | $\alpha_1$ | $\alpha_2$ | $\alpha_{l-1}$ | $\alpha_l$ |
|---|---|---|---|---|
| $A_{l-1}$ | $\boldsymbol{e}_1 - \boldsymbol{e}_2$ | $\boldsymbol{e}_2 - \boldsymbol{e}_3$ | $\boldsymbol{e}_{l-1} - \boldsymbol{e}_l$ | |
| $D_l$ | $\boldsymbol{e}_1 - \boldsymbol{e}_2$ | $\boldsymbol{e}_2 - \boldsymbol{e}_3$ | $\boldsymbol{e}_{l-1} - \boldsymbol{e}_l$ | $\boldsymbol{e}_{l-1} + \boldsymbol{e}_l$ |
| $B_l$ | $\boldsymbol{e}_1 - \boldsymbol{e}_2$ | $\boldsymbol{e}_2 - \boldsymbol{e}_3$ | $\boldsymbol{e}_{l-1} - \boldsymbol{e}_l$ | $+1\boldsymbol{e}_l$ |
| $D_l$ | $\boldsymbol{e}_1 - \boldsymbol{e}_2$ | $\boldsymbol{e}_2 - \boldsymbol{e}_3$ | $\boldsymbol{e}_{l-1} - \boldsymbol{e}_l$ | $+2\boldsymbol{e}_l$ |

**Table 1** Roots for the simple classical Lie groups and algebras

| *Group* | *Algebra* | *Root space* | *Rank* | *Roots* | *Conditions* |
|---|---|---|---|---|---|
| SU($l$) | $\mathfrak{su}(l)$ | $A_{l-1}$ | $l-1$ | $+\boldsymbol{e}_i - \boldsymbol{e}_j$ | $1 \le i \ne j \le l$ |
| SO($2l$) | $\mathfrak{so}(2l)$ | $D_l$ | $l$ | $\pm\boldsymbol{e}_i \pm \boldsymbol{e}_j$ | $1 \le i < j \le l$ |
| SO($2l+1$) | $\mathfrak{so}(2l+1)$ | $B_l$ | $l$ | $\pm\boldsymbol{e}_i \pm \boldsymbol{e}_j$, $\pm\boldsymbol{e}_k$ | $1 \le i < j, k \le l$ |
| Sp($l$) = USp($2l$) | $\mathfrak{sp}(l) = \mathfrak{usp}(2l)$ | $C_l$ | $l$ | $\pm\boldsymbol{e}_i \pm \boldsymbol{e}_j$, $\pm 2\boldsymbol{e}_k$ | $1 \le i < j, k \le l$ |

**Table 2** Roots for the simple exceptional Lie algebras

| *Root space* | *Rank* | *Dimension* | *Roots* | *Conditions* |
|---|---|---|---|---|
| $G_2$ | 2 | 14 | $+\boldsymbol{e}_i - \boldsymbol{e}_j$<br>$\pm[(\boldsymbol{e}_i + \boldsymbol{e}_j) - 2\boldsymbol{e}_k]$ | $1 \le i \ne j \ne k \le 3$ |
| $F_4$ | 4 | 52 | $\pm\boldsymbol{e}_i \pm \boldsymbol{e}_j$, $\pm 2\boldsymbol{e}_i$<br>$\pm\boldsymbol{e}_1 \pm \boldsymbol{e}_2 \pm \boldsymbol{e}_3 \pm \boldsymbol{e}_4$ | $1 \le i < j \le 4$ |
| $E_6$ | 6 | 78 | $\pm\boldsymbol{e}_i \pm \boldsymbol{e}_j$<br>$\frac{1}{2}(\pm\boldsymbol{e}_1 \pm \boldsymbol{e}_2 \pm \boldsymbol{e}_3 \pm \boldsymbol{e}_4 \pm \boldsymbol{e}_5) \pm \frac{\sqrt{3}}{4}\boldsymbol{e}_6$ | $1 \le i < j \le 5$<br>[a] |
| $E_7$ | 7 | 133 | $\pm\boldsymbol{e}_i \pm \boldsymbol{e}_j$<br>$\frac{1}{2}(\pm\boldsymbol{e}_1 \pm \boldsymbol{e}_2 \pm \boldsymbol{e}_3 \pm \boldsymbol{e}_4 \pm \boldsymbol{e}_5 \pm \boldsymbol{e}_6) \pm \frac{\sqrt{2}}{4}\boldsymbol{e}_7$ | $1 \le i < j \le 6$<br>[b] |
| $E_8$ | 8 | 248 | $\pm\boldsymbol{e}_i \pm \boldsymbol{e}_j$<br>$\frac{1}{2}(\pm\boldsymbol{e}_1 \pm \boldsymbol{e}_2 \pm \boldsymbol{e}_3 \pm \boldsymbol{e}_4 \pm \boldsymbol{e}_5 \pm \boldsymbol{e}_6 \pm \boldsymbol{e}_7 \pm \boldsymbol{e}_8)$ | $1 \le i < j \le 8$<br>[a] |

[a]Even number of + signs.
[b]Even number of + signs within bracket.

All roots in the rank-2 root spaces have been expressed in terms of both two orthogonal vectors and two fundamental roots in Figure 3.

If $\alpha_i$ and $\alpha_j$ are fundamental roots, their inner product is zero or negative

$$\cos(\alpha_i, \alpha_j) = 0, -\sqrt{\frac{1}{4}}, -\sqrt{\frac{2}{4}}, -\sqrt{\frac{3}{4}}$$

This information has been used to classify the root spaces of the inequivalent simple Lie algebras (over $C$). The procedure is as follows. Each of the $l$ fundamental roots in a rank-$l$ root space is represented by a dot in a plane. Dots representing roots $\alpha_i$ and $\alpha_j$ are connected by $n_{ij}$ lines, where $\cos(\alpha_i, \alpha_j) = -\sqrt{n_{ij}/4}$. Orthogonal roots are not connected by any lines. Such diagrams are called Dynkin diagrams. Disconnected Dynkin diagrams describe semisimple Lie algebras. Connected Dynkin diagrams classify simple Lie algebras.

The properties of Dynkin diagrams arise from two simple observations:

**O1:** The root space is positive definite.

**O2:** If $u$ is a unit vector and $v_i$ are an orthonormal set of vectors,

$$\sum (u \cdot v_i)^2 \leq 1$$

These two observations lead to three important properties of Dynkin diagrams.

**D1:** There are no loops. If $\alpha_i$ $(i = 1, 2, \ldots, k)$ are in a loop, then there are at least as many lines as vertices. With $u_i = \alpha_i/|\alpha_i|$,

$$\left( \sum_{i=1}^{k} u_i, \sum_{j=1}^{k} u_j \right) = k + 2 \sum_{i<j}^{k} u_i \cdot u_j > 0$$

Since $2u_i \cdot u_j \leq -1$ if $u_i \cdot u_j \neq 0$, there cannot be as many lines as vertices.

**D2:** The number of lines connected to any node is $<4$. If $\alpha_i$ are connected to $v$, then with $u_i = \alpha_i/|\alpha_i|$,

$$\sum (v \cdot u_i)^2 = \sum n_i/4 < 1$$

since $v$ is linearly independent of the $\alpha_i$.

**D3:** A simple chain connecting any two nodes can be shrunk. If the original diagram is allowed, the shrunk diagram is also allowed, and conversely. Since the shrunk diagram in Figure 6 violates *D2*, the original is not an allowed Dynkin diagram.

According to these results, the maximum number of lines that can be attached to a vertex is three. If a vertex is attached to three lines, it can be connected to three (one line each) other vertices, two (two plus one) other vertices, or only one other vertex (all three lines). This last case describes Dynkin diagram $G_2$ (cf. Figures 3 and 5).

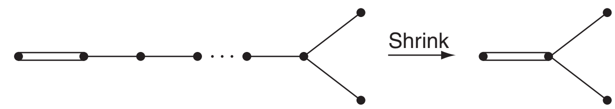



**Figure 6** A chain with single links can be removed from a diagram. If the original is an allowed Dynkin diagram, the shrunk diagram is also allowed, and conversely.

The only remaining possibilities are shown in Figure 7.

For diagrams of type ($B$, $C$, $F$) we define vectors

$$u = \sum_{i=1}^{p} i u_i \qquad v = \sum_{j=1}^{q} j v_j$$

where as usual $u_i, v_j$ are unit vectors $\alpha_k/|\alpha_k|$. The Schwartz inequality applied to $u$ and $v$ leads to the inequality

$$\left(1 + \frac{1}{p}\right)\left(1 + \frac{1}{q}\right) > 2$$

The solutions with $p \geq q$ are

| $p$ | $q$ | Root space | Constraint |
|---|---|---|---|
| arbitrary | 1 | $B_l, C_l$ | $i = p + 1$ |
| 2 | 2 | $F_4$ | |

For diagrams of type ($D$, $E$), we define vectors

$$u = \sum_{i=1}^{p-1} i u_i, \qquad v = \sum_{j=1}^{q-1} j v_j, \qquad w = \sum_{k=1}^{r-1} k w_k$$

where as usual $u_i, v_j, w_k$ are unit vectors $\alpha_m/|\alpha_m|$. With similar arguments, we obtain the inequality

$$\frac{1}{p} + \frac{1}{q} + \frac{1}{r} > 2$$

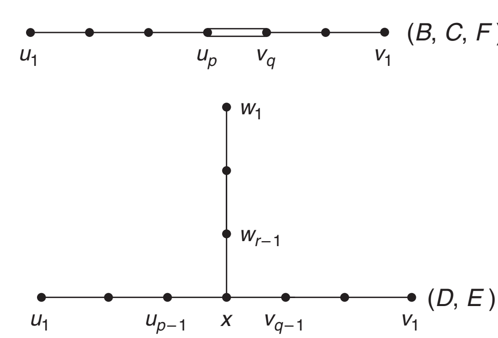



**Figure 7** The only remaining candidate Dynkin diagrams have either two vertices ($B$, $C$, $F$) or one vertex ($D$, $E$) connected to three lines.

The solutions with $p \geq q \geq r$ are

| $p$ | $q$ | $r$ | *Root space* | *Regular Euclidean solid* |
|---|---|---|---|---|
| arbitrary | 2 | 2 | $D_p + 2$ | |
| 3 | 3 | 2 | $E_6$ | Tetrahedron |
| 4 | 3 | 2 | $E_7$ | Cube–octahedron |
| 5 | 3 | 2 | $E_8$ | Icosahedron–dodecahedron |

All allowed Dynkin diagrams are shown in Figure 8. In these diagrams roots making an angle of 120° with each other (joined by single lines) have equal length. Roots joined by double lines or triple lines have different lengths. The arrows on double lines indicated the shorter and longer roots. Arrows point to longer roots. The root space $G_2$ and $F_4$ are self-dual, so it does not matter which way the arrow points.

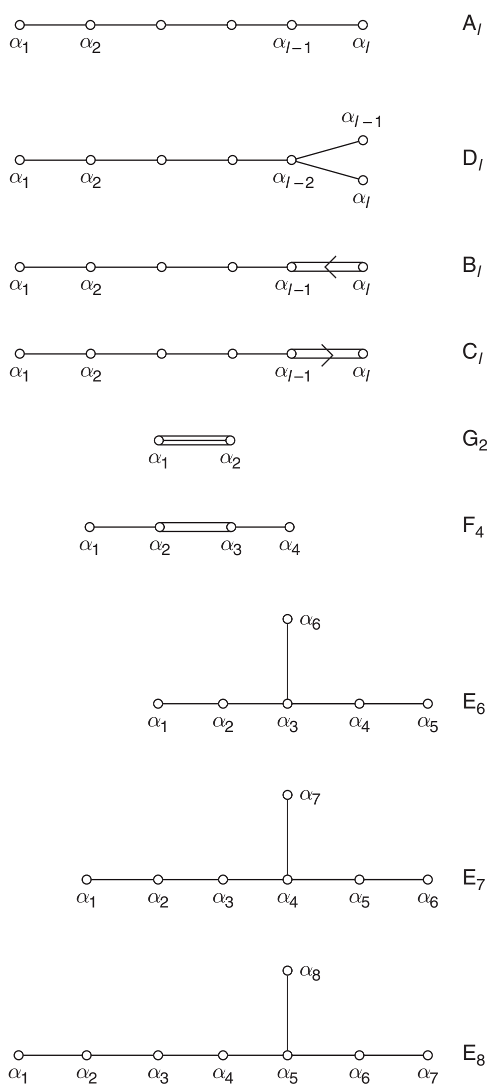



**Figure 8** Four infinite series ($A_l, D_l, B_l, C_l$) of Dynkin diagrams exist and correspond to the classical simple Lie groups (SU $(l+1)$, SO$(2l)$, SO$(2l+1)$, USp$(2l)$). The five exceptional Dynkin diagrams include a short finite series ($E_l, l = 6, 7, 8$), $F_4$, and $G_2$.

Coxeter–Dynkin diagrams also appear in classical geometry and catastrophe theory.

## Real Forms

The metric tensor $g_{\mu\nu}$ for a simple Lie algebra (over $C$) in the canonical basis $\boldsymbol{H}, E_{\boldsymbol{\alpha}}$ is

$$
g \rightarrow \left[\begin{array}{cccc|cc|cc|c}
1 &&&&&&&&\\
&1&&&&&&&\\
&&\ddots&&&&&&\\
&&&1&&&&&\\ \hline
&&&&0&1&&&\\
&&&&1&0&&&\\ \cline{5-8}
&&&&&&0&1&\\
&&&&&&1&0&\\ \cline{5-8}
&&&&&&&&\ddots
\end{array}\right]
\begin{array}{c} H_1\\ H_2\\ \vdots\\ H_l\\ E_{+\alpha}\\ E_{-\alpha}\\ E_{+\beta}\\ E_{-\beta}\\ \vdots \end{array} \qquad [2]
$$

In this basis, the Lie algebra decomposes into positive- and negative-definite subspaces according to

$$\mathfrak{g} = \mathfrak{g}_+ + \mathfrak{g}_-$$

$$
\begin{array}{lll}
\mathfrak{g}_+ \text{ spanned by} & H_i, & (E_{+\alpha} + E_{-\alpha})/\sqrt{2} \\
\mathfrak{g}_- \text{ spanned by} & & (E_{+\alpha} - E_{-\alpha})/\sqrt{2}
\end{array}
$$

The choice of basis suggested above diagonalizes the Cartan–Killing form in eqn [2]: $g \rightarrow I_{p,q}$, with $p = l + (1/2)(n - l)$ positive values +1 on the diagonal and $q = (1/2)(n - l)$ values –1 on the diagonal. The trace of this matrix is the trace of $g$: $+l$.

An arbitrary element in this (complex) Lie algebra is a linear superposition of the form

$$X = \sum_i h^i H_i + \sum_{\alpha \neq 0} e^{\alpha} E_{\alpha} \qquad [3]$$

where all $n$ coefficients $h^i, e^{\alpha}$ are complex. If all these coefficients are taken real, the resulting Lie algebra closes under commutation and describes a noncompact Lie group. The subalgebra describing the maximal compact subgroup is spanned by the

linear combinations $(E_{+\alpha} - E_{-\alpha})/\sqrt{2}$. The remaining operators exponentiate to a noncompact coset

$$\mathrm{EXP}\left\{h^i H_i + e_+^{\alpha}(E_{+\alpha} + E_{-\alpha})/\sqrt{2}\right\}$$

which is topologically equivalent to $R^K, K = l + (1/2)(n-l) = (1/2)(n+l)$. Of all the real forms of the complex Lie algebra described by this set of canonical commutation relations (or root space, or Dynkin diagram), this is the least compact real form.

The compact real form is obtained from [3] by taking linear combinations

$$X = \sum_i \mathrm{i}h^i H_i + \sum_{\alpha\neq 0} \mathrm{i}e_+^{\alpha}(E_{+\alpha} + E_{-\alpha})/\sqrt{2} + \sum_{\alpha\neq 0} e_-^{\alpha}(E_{+\alpha} - E_{-\alpha})/\sqrt{2}$$

where $h^i, e_+^{\alpha}, e_-^{\alpha}$ are real. The compact real forms of the simple Lie algebras are:

| *Root space* | *Group* |
|---|---|
| $A_l - 1$ | SU($l$) |
| $D_l$ | SO($2l$) |
| $B_l$ | SO($2l+1$) |
| $C_l$ | USp($2l$) = Sp($l$) |

If the imaginary factor i is absorbed into the Cartan–Killing metric, this metric is diagonal, all matrix elements are –1, the trace of this form is $-n$, and the linear combinations for $X$ are real.

Every complex simple Lie algebra (i.e., simple Lie algebra over $C$) has a spectrum of inequivalent real forms. These can all be obtained from the compact real form by an analog of Minkowski's "rotation trick," derived by Cartan. Cartan introduced a metric-preserving linear mapping ("involutive automorphism") $T : \mathfrak{g} \to \mathfrak{g}$ with the property $T^2 = I$ and $(TX, TY) = (X, Y)$, with $X, Y \in \mathfrak{g}$. The operator $T$ has eigenvalues $\pm 1$ and induces a decomposition ("Cartan decomposition") in $\mathfrak{g}$ as follows:

$$\begin{array}{lccccc} & T(\mathfrak{g}) & = & T(\mathfrak{k}) & + & T(\mathfrak{p}) \\ \mathfrak{g} = \mathfrak{k} + \mathfrak{p} & & & \downarrow & & \downarrow \\ & & & \mathfrak{k} & - & \mathfrak{p} \end{array}$$

As a result, the subspaces $\mathfrak{k}$ and $\mathfrak{p}$ are orthogonal. The subspaces obey the following commutation and inner-product properties:

$$\begin{array}{ll} [\mathfrak{k}, \mathfrak{k}] \subseteq \mathfrak{k}, & (\mathfrak{k}, \mathfrak{k}) < 0 \\ [\mathfrak{k}, \mathfrak{p}] \subseteq \mathfrak{p}, & (\mathfrak{k}, \mathfrak{p}) = 0 \\ [\mathfrak{p}, \mathfrak{p}] \subseteq \mathfrak{k}, & (\mathfrak{p}, \mathfrak{p}) < 0 \end{array}$$

Under the analytic continuation $\mathfrak{p} \to \mathrm{i}\mathfrak{p}$, the compact Lie algebra $\mathfrak{g}$ is rotated to a noncompact Lie algebra $\mathfrak{g}'$ whose commutation relations and inner-product properties are

$$\mathfrak{g} = \mathfrak{k} + \mathfrak{p} \quad \to \quad \mathfrak{g}' = \mathfrak{k} + \mathfrak{p}'$$

$$\begin{array}{ll} [\mathfrak{k}, \mathfrak{k}] \subseteq \mathfrak{k}, & (\mathfrak{k}, \mathfrak{k}) < 0 \\ [\mathfrak{k}, \mathfrak{p}'] \subseteq \mathfrak{p}', & (\mathfrak{k}, \mathfrak{p}') = 0 \\ [\mathfrak{p}', \mathfrak{p}'] \subseteq \mathfrak{k}, & (\mathfrak{p}', \mathfrak{p}') > 0 \end{array}$$

The maximal compact subalgebra of $\mathfrak{g}'$ is $\mathfrak{k}$. The subspace $\mathfrak{p}'$ exponentiates to a simply connected submanifold on which the Cartan–Killing metric is positive definite. This manifold is topologically equivalent to $R^K$, $K = \dim \mathfrak{p}$. It is not geometrically equivalent to $R^K$ once an invariant metric is placed on it.

Three linear mappings that satisfy $T^2 = I$ suffice to generate all real forms of all the simple classical Lie algebras.

## Block Matrix Decomposition

The compact Lie algebra $\mathfrak{u}(n; F)$ has a block submatrix decomposition ($n = p + q$):

$$\mathfrak{u}(n; F) = \begin{bmatrix} A_p & 0 \\ 0 & A_q \end{bmatrix} + \begin{bmatrix} 0 & +B \\ -B^{\dagger} & 0 \end{bmatrix}$$

where $A_p^{\dagger} = -A_p, A_q^{\dagger} = -A_q$ and $B$ is an arbitrary $p \times q$ matrix over $F$. Under the map

$$T(\mathfrak{g}) = I_{p,q}\mathfrak{g}I_{p,q}, \ I_{p,q} = \begin{bmatrix} I_p & 0 \\ 0 & -I_q \end{bmatrix}$$

the diagonal subspace

$$\begin{bmatrix} A_p & 0 \\ 0 & A_q \end{bmatrix}$$

has eigenvalue +1 and the off-diagonal subspace

$$\begin{bmatrix} 0 & +B \\ -B^{\dagger} & 0 \end{bmatrix}$$

has eigenvalue –1. Under the Cartan rotation

$$\mathfrak{u}(n; F) \to \mathfrak{u}(p, q; F) = \begin{bmatrix} A_p & 0 \\ 0 & A_q \end{bmatrix} + \begin{bmatrix} 0 & +B \\ +B^{\dagger} & 0 \end{bmatrix}$$

The real forms of the classical Lie groups obtained in this way are

$$D_n, B_n$$

$$\begin{matrix} \mathrm{SO}(2n) & & \\ & \rightarrow & \mathrm{SO}(p,q) \\ \mathrm{SO}(2n+1) & & \end{matrix}$$

$$A_{n-1}$$
$$\mathrm{SU}(n) \rightarrow \mathrm{SU}(p,q)$$

$$C_n$$
$$\mathrm{Sp}(n) \rightarrow \mathrm{Sp}(p,q)$$
$$\mathrm{USp}(2n) \rightarrow \mathrm{USp}(2p,2q)$$

### Subfield Restriction

The Lie algebra $\mathfrak{su}(n)$ of complex traceless anti-Hermitian matrices has a subalgebra $\mathfrak{so}(n)$ of real antisymmetric matrices. The algebra $\mathfrak{su}(n)$ can be expressed in terms of real $n \times n$ antisymmetric matrices $A_n$ and traceless symmetric matrices $S_n$:

$$\mathfrak{su}(n) = \mathfrak{so}(n) + [\mathfrak{su}(n) - \mathfrak{so}(n)] = A_n + \mathrm{i}S_n$$

The Cartan rotation is

$$\begin{aligned} \mathfrak{su}(n) \rightarrow \mathfrak{sl}(n;R) &= \mathfrak{so}(n) + \mathrm{i}[\mathfrak{su}(n) - \mathfrak{so}(n)] \\ &= A_n + S_n \end{aligned}$$

The classical Lie group generated by this transformation is $\mathrm{SL}(n;R)$.

A similar rotation can be carried out on unitary matrices over the quaternion field, $\mathfrak{u}(n;Q) = \mathfrak{sp}(n)$. This algebra contains the subalgebra $\mathfrak{u}(n)$ in which quaternions $q = q_0 + \mathcal{I}q_1 + \mathcal{J}q_2 + \mathcal{K}q_3$ are restricted to complex numbers $q = q_0 + \mathrm{i}q_1$. There is a natural decomposition

$$\mathfrak{sp}(n) = \mathfrak{u}(n) + [\mathfrak{sp}(n) - \mathfrak{u}(n)]$$

It is useful at this point to replace each quaternion matrix element by a $2 \times 2$ complex matrix: $\mathfrak{sp}(n) \rightarrow \mathfrak{usp}(2n)$. This is a unitary representation of the symplectic algebra. Replacing the complex matrix elements in $\mathfrak{u}(n)$ by $2 \times 2$ real matrices simultaneously generates a real matrix representation of $\mathfrak{u}(n)$ named $\mathfrak{ou}(2n)$. This is an orthogonal representation of the unitary algebra. The decomposition above is

$$\begin{aligned} \mathfrak{sp}(n) &\rightarrow \mathfrak{u}(n) + [\mathfrak{sp}(n) - \mathfrak{u}(n)] \\ &\rightarrow \mathfrak{ou}(2n) + [\mathfrak{usp}(2n) - \mathfrak{ou}(2n)] = A_{2n} + \mathrm{i}S_{2n} \end{aligned}$$

where as before $A_{2n}$ and $S_{2n}$ are $2n \times 2n$ antisymmetric and symmetric matrices. The Cartan rotation maps this to $\mathfrak{sp}(2n;R)$,

$$\mathfrak{usp}(2n) \rightarrow \mathfrak{sp}(2n;R) = A_{2n} + S_{2n}$$

The classical Lie group generated in this way is $\mathrm{Sp}(2n;R)$. Matrices in this group satisfy the quadratic constraint $M^t G M = G, G^t = -G, \det(G) \neq 0$. The real symplectic groups leave invariant Hamilton's equations of motion: $\mathrm{d}p_i/\mathrm{d}t = -\partial H/\partial q_i$, $\mathrm{d}q_i/\mathrm{d}t = +\partial H/\partial p_i$.

### Field Embeddings

The image of $\mathfrak{u}(n) \rightarrow \mathfrak{ou}(2n)$ consists of a set of $2n \times 2n$ antisymmetric matrices of dimension $n^2$. These matrices form a subset of $\mathfrak{so}(2n)$, which consists of $2n \times 2n$ antisymmetric matrices of dimension $2n(2n-1)/2$. As a result, $\mathfrak{ou}(2n)$ is a subalgebra in $\mathfrak{so}(2n)$. Thus, $\mathfrak{ou}(2n) \sim \mathfrak{k}$ and $\mathfrak{so}(2n) \sim \mathfrak{g}$ and we have a Cartan decomposition

$$\begin{matrix} \mathfrak{so}(2n) = & \mathfrak{ou}(2n) + & [\mathfrak{so}(2n) - \mathfrak{ou}(2n)] & \\ & \downarrow & \downarrow & \\ & \mathfrak{ou}(2n) + & \mathrm{i}[\mathfrak{so}(2n) - \mathfrak{ou}(2n)] & = \mathfrak{so}^*(2n) \end{matrix}$$

In the same way, the image of $\mathfrak{sp}(2n) \rightarrow \mathfrak{usp}(2n)$ consists of an $n(2n+1)$-dimensional set of $2n \times 2n$ anti-Hermitian matrices. This is a subset of $\mathfrak{su}(2n)$, which has dimension $(2n)^2 - 1$. It is also a subalgebra of $\mathfrak{su}(2n)$. Thus, $\mathfrak{usp}(2n) \sim \mathfrak{k}$ and $\mathfrak{su}(2n) \sim \mathfrak{g}$, so we have a Cartan decomposition

$$\begin{matrix} \mathfrak{su}(2n) = & \mathfrak{usp}(2n) + & [\mathfrak{su}(2n) - \mathfrak{usp}(2n)] & \\ & \downarrow & \downarrow & \\ & \mathfrak{usp}(2n) + & \mathrm{i}[\mathfrak{su}(2n) - \mathfrak{usp}(2n)] & = \mathfrak{su}^*(2n) \end{matrix}$$

These real forms are summarized in Table 3.

**Table 3** Real forms of the simple classical Lie algebras

| *Mapping* | *Real form* | *Maximal compact subalgebra* | *Root space* | *Condition* |
|---|---|---|---|---|
| Block submatrix | $\mathfrak{so}(p,q)$ | $\mathfrak{so}(p) + \mathfrak{so}(q)$ | $D_n$ | $p+q=2n$ |
| | $\mathfrak{so}(p,q)$ | $\mathfrak{so}(p) + \mathfrak{so}(q)$ | $B_n$ | $p+q=2n+1$ |
| | $\mathfrak{su}(p,q)$ | $\mathfrak{u}(1) + \mathfrak{su}(p) + \mathfrak{su}(q)$ | $A_{n-1}$ | $p+q=n$ |
| | $\mathfrak{sp}(p,q) = \mathfrak{usp}(2p,2q)$ | $\mathfrak{usp}(2p) + \mathfrak{usp}(2p)$ | $C_n$ | $p+q=n$ |
| Subfield restriction | $\mathfrak{sl}(n;R)$ | $\mathfrak{so}(n)$ | $A_{n-1}$ | |
| | $\mathfrak{sp}(2n;R)$ | $\mathfrak{u}(n)$ | $C_n$ | |
| Field embedding | $\mathfrak{so}^*(2n)$ | $\mathfrak{u}(n)$ | $D_n$ | |
| | $\mathfrak{su}^*(2n)$ | $\mathfrak{sp}(n) = \mathfrak{usp}(2n)$ | $A_{2n-1}$ | |

**Table 4** Equivalence among real forms of the simple classical Lie algebras

| $A_1$ | $= B_1$ | $= C_1$ | $\chi$ |
|---|---|---|---|
| $\mathfrak{su}(2)$ | $= \mathfrak{so}(3)$ | $= \mathfrak{sp}(1) = \mathfrak{usp}(2)$ | −3 |
| $\mathfrak{su}(1,1) = \mathfrak{sl}(2;R)$ | $= \mathfrak{so}(2,1)$ | $= \mathfrak{sp}(2;R)$ | +1 |
| $D_2$ | $= A_1$ | $+ A_1$ | $\chi$ |
| $\mathfrak{so}(4)$ | $= \mathfrak{so}(3)$ | $+ \mathfrak{so}(3)$ | −6 |
| $\mathfrak{so}^*(4)$ | $= \mathfrak{so}(3)$ | $+ \mathfrak{so}(2,1)$ | −2 |
| $\mathfrak{so}(3,1)$ | $= \mathfrak{sl}(2;C)$ | | 0 |
| $\mathfrak{so}(2,2)$ | $= \mathfrak{so}(2,1)$ | $+ \mathfrak{so}(2,1)$ | +2 |
| $B_2$ | $= C_2$ | | $\chi$ |
| $\mathfrak{so}(5)$ | $= \mathfrak{sp}(2) = \mathfrak{usp}(4)$ | | −10 |
| $\mathfrak{so}(4,1)$ | $= \mathfrak{sp}(1,1) = \mathfrak{usp}(2,2)$ | | −2 |
| $\mathfrak{so}(3,2)$ | $= \mathfrak{sp}(4;R)$ | | +2 |
| $D_3$ | $= A_3$ | | $\chi$ |
| $\mathfrak{so}(6)$ | $= \mathfrak{su}(4)$ | | −15 |
| $\mathfrak{so}(5,1)$ | $= \mathfrak{su}^*(4)$ | | −5 |
| $\mathfrak{so}^*(6)$ | $= \mathfrak{su}(3,1)$ | | −3 |
| $\mathfrak{so}(4,2)$ | $= \mathfrak{su}(2,2)$ | | +1 |
| $\mathfrak{so}(3,3)$ | $= \mathfrak{sl}(4;R)$ | | +3 |

**Table 5** Real forms of the exceptional Lie algebras

| | | *Maximal compact subgroup* | |
|---|---|---|---|
| *Root space* | *Class*$_{Rank(Character)}$ | *Root space* | *Dimension* |
| $G_2$ | $G_{2(-14)}$ | $G_2$ | 14 |
| | $G_{2(+2)}$ | $A_1 + A_1$ | 6 |
| $F_4$ | $F_{4(-52)}$ | $F_4$ | 52 |
| | $F_{4(-20)}$ | $B_4$ | 36 |
| | $F_{4(+4)}$ | $C_3 + A_1$ | 24 |
| $E_6$ | $E_{6(-78)}$ | $E_6$ | 78 |
| | $E_{6(-26)}$ | $F_4$ | 52 |
| | $E_{6(-14)}$ | $D_5 + D_1$ | 46 |
| | $E_{6(+2)}$ | $A_5 + A_1$ | 38 |
| | $E_{6(+6)}$ | $C_4$ | 36 |
| $E_7$ | $E_{7(-133)}$ | $E_7$ | 133 |
| | $E_{7(-25)}$ | $E_6 + D_1$ | 79 |
| | $E_{7(-5)}$ | $D_6 + A_1$ | 69 |
| | $E_{7(+7)}$ | $A_7$ | 63 |
| $E_8$ | $E_{8(-248)}$ | $E_8$ | 248 |
| | $E_{8(-24)}$ | $E_7 + A_1$ | 136 |
| | $E_{8(+8)}$ | $D_8$ | 120 |

The root spaces $A_1[\mathrm{SU}(2)], B_1[\mathrm{SO}(3)]$, and $C_1[\mathrm{U}(1;Q) \simeq \mathrm{USp}(2;C)]$ are equivalent. As a result, the different real forms of their complex extensions are related to each other. Similar remarks hold for the real forms of $B_2 = C_2$, $D_2 = A_1 + A_1$, and $D_3 = A_3$. The relations among these real forms are summarized in Table 4. This table is useful in inferring "spinor representations" among classical groups. Thus, SO(3) has spinor representations based on SU(2) and Sp(1); SO(4) has spinor representations based on $\mathrm{SU}(2) \times \mathrm{SU}(2)$; SO(5) has spinor representations based on USp(4); and SO(6) has spinor representations based on SU(4).

For completeness, the real forms for the exceptional Lie algebras are collected in Table 5.

Real forms of the complex extension of a simple Lie algebra are almost uniquely distinguished by an index. This is the trace of the Cartan–Killing form [2], once the appropriate factors of i have been absorbed into it. If $n_c$ is the dimension of the maximal compact subgroup, $\chi = \mathrm{tr}(g) = +1(n - n_c) -1(n_c) = n - 2n_c$. The index ranges from $-n$ for the compact real form (for which $n_c = n$) to $+l$ for the least compact real form.

## Riemannian Symmetric Spaces

Exponentiation lifts Lie algebras to Lie groups and subspaces in Lie algebras into submanifolds in Lie groups. In particular, exponentiation of a Cartan decomposition

$$\begin{array}{ccccc} \mathfrak{g} & = & \mathfrak{k} & + & \mathfrak{p} \\ \downarrow & & \downarrow & & \downarrow \\ G & = & K & \times & (P = G/K) \end{array}$$

lifts the subspace $\mathfrak{p}$ to the quotient $(P = G/K)$.

A metric may be defined on the Lie group $G$ as follows. Define the distance between the identity and some nearby point $g(\epsilon) = \mathrm{EXP}(\epsilon X) = \mathrm{EXP}(\delta x^i X_i)$ by

$$\mathrm{d}s^2(0) = G_{rs}\delta x^r \delta x^s$$

Move $I$ and $g(\epsilon)$ to the neighborhood of any point $g(x) \in G$ by left multiplication: $g(x)I \to g(x)$, $g(x)g(\delta x^i X_i) \to g((x + \mathrm{d}x)^i X_i)$. The infinitesimals $\mathrm{d}x^i(x)$ at $x$ (defined by $g(x)$) and $\delta x^i = \mathrm{d}x^i(0)$ at $I$ are linearly related,

$$\delta x^i = M^i{}_j(x)\, \mathrm{d}x^j(x)$$

By requiring that the distance $\mathrm{d}s$ between $I$ and $g(\delta x^i X_i)$ at the identity be the same as the distance between $g(x^i X_i)I$ and $g(x^i X_i)g(\delta x^i X_i) = g((x + \mathrm{d}x)^i X_i)$ at $g(x^i X_i)$ leads to the condition

$$\begin{aligned} \mathrm{d}s^2 &= G_{rs}(0)\delta x^r \delta x^s \\ &= G_{rs}(0) M^r{}_i(x) M^s{}_j(x)\, \mathrm{d}x^i(x)\, \mathrm{d}x^j(x) \\ &= G_{ij}(x)\, \mathrm{d}x^i(x)\, \mathrm{d}x^j(x) \end{aligned}$$

An invariant metric $G(x)$ over the Lie group $G$ is defined by

$$\begin{aligned} G_{ij}(x) &= G_{rs}(0) M^r{}_i(x) M^s{}_j(x) \\ G(x) &= M^t(x) G(0) M(x) \end{aligned}$$

It is useful to identify $G(0)$ with the Cartan–Killing inner product on $\mathfrak{g}$. Since $M(x)$ is nonsingular, the signature of $G(x)$ is invariant over the group.

The invariant metric on $G$ can be restricted to subspaces $K \subset G$ and $P = G/K \subset G$. The signature on these subspaces is the same as the signature on the subspaces $\mathfrak{k}$ and $\mathfrak{p}$ in $\mathfrak{g}$. Thus, if $G$ is compact, the invariant metric is negative definite on $K$ and on $P = G/K$ and positive definite on the analytically continued space $P' = G'/K$. In short, it is definite (negative, positive) on $P, P'$. These spaces are Riemannian spaces and they are globally symmetric. They have been investigated by studying the properties of the secular equation of the Lie algebra $\mathfrak{g}$, restricted to the subspace $\mathfrak{p}$:

$$\det\left[R(p^i P_i) - \lambda I\right] = \sum_j (-\lambda)^{n-j} \hat{\phi}_j(p) = 0 \qquad [4]$$

where the $P_i$ are basis vectors that span $\mathfrak{p}$. The coefficients $\hat{\phi}_j(p)$ in the secular equation [4] for Riemannian symmetric spaces are related to the coefficients $\phi_j(x)$ in the secular equation [1] for Lie algebras. A rank for the Riemannian symmetric space $P = \mathrm{EXP}(\mathfrak{p})$ can be defined from the secular equation following exactly the prescription followed for the Lie algebra $\mathfrak{g}$. The rank of the Riemannian symmetric space $P = \mathrm{EXP}(\mathfrak{p})$ is

1. the number of functionally independent coefficients $\hat{\phi}_j(p)$ in the secular equation;
2. the number of independent roots of the secular equation;
3. the dimension of the maximal Euclidean subspace in $P$; and
4. the number of independent (Laplace–Beltrami) operators that commute with all displacement operators $P_i$: $\Delta_j(P) = \hat{\phi}_j(p^i \to P_i)$.

Rank-1 Riemannian symmetric spaces are isotropic as well as homogeneous.

Tables 3 and 5 contain all the information required to enumerate all the classical and exceptional Riemannian symmetric spaces. All the classical Riemannian symmetric spaces are tabulated in Table 6. The exceptional Riemannian symmetric spaces can be constructed from the information in Table 5 following the procedure used to construct Table 6 from Table 3.

As particular examples of Riemannian symmetric spaces we consider the compact spaces $\mathrm{SO}(p+q)/[\mathrm{SO}(p) \times \mathrm{SO}(q)]$ and their noncompact counterparts $\mathrm{SO}(p,q)/[\mathrm{SO}(p) \times \mathrm{SO}(q)]$. These spaces have rank $\min(p,q)$, dimension $pq$, and can be represented explicitly in matrix form as

$$\left[\begin{array}{c|c} 0 & X \\ \hline \sigma X^t & 0 \end{array}\right] \to \mathrm{EXP}\left[\begin{array}{c|c} 0 & X \\ \hline \sigma X^t & 0 \end{array}\right] = \left[\begin{array}{c|c} D_p & Y \\ \hline \sigma Y^t & D_q \end{array}\right]$$

Here $X$ is a $p \times q$ matrix and $\sigma = +1$ for the noncompact case and –1 for the compact case. The block diagonal matrices $D_p$ and $D_q$ are defined from the metric-preserving conditions ($M^t I_{p+q} M = I_{p+q}$, $M^t I_{p,q} M = I_{p,q}$)

$$D_p^2 = I_p + \sigma Y Y^t, \qquad D_q^2 = I_q + \sigma Y^t Y$$

The $pq$ coordinates in the Riemannian symmetric spaces can be taken as the $pq$ elements of the submatrix $Y$.

These Riemannian symmetric spaces can be treated as algebraic submanifolds in $R^K, K = pq + (1/2)q(q+1)$. The $K$ coordinates on $R^K$ can be identified with the $pq$ matrix elements of $Y$ and the $(1/2)q(q+1)$ matrix elements of the real symmetric matrix $D_q$. These coordinates obey the $(1/2)q(q+1)$ algebraic constraints defined by

$$D_q^2 - \sigma Y^t Y = I_q$$

For SO(3)/SO(2) and SO(2,1)/SO(2), this condition is determined from the matrix

$$\left[\begin{array}{c|c} \left[I_2 + \begin{pmatrix} \sigma x & \sigma y \end{pmatrix}\begin{bmatrix} x \\ y \end{bmatrix}\right]^{1/2} & \begin{matrix} x \\ y \end{matrix} \\ \hline \sigma x \quad \sigma y & z \end{array}\right] \quad \text{to be}$$

$$z^2 - \sigma(x^2 + y^2) = 1$$

**Table 6** All classical Riemannian symmetric spaces

| *Root space* | *Quotient* | *Dimension* | *Rank* | $\chi$ |
|---|---|---|---|---|
| $A_{p+q-1}$ | $\mathrm{SU}(p,q)/\mathrm{S}[\mathrm{U}(p) \otimes \mathrm{U}(q)]$ | $2pq$ | $\min(p,q)$ | $1-(p-q)^2$ |
| $A_{n-1}$ | $\mathrm{SL}(n;R)/\mathrm{SO}(n)$ | $\frac{1}{2}(n+2)(n-1)$ | $n-1$ | $n-1$ |
| $A_{2n-1}$ | $\mathrm{SU}^*(2n)/\mathrm{USp}(2n)$ | $(2n+1)(n-1)$ | $n-1$ | $-2n-1$ |
| $B_{p+q}$ | $\mathrm{SO}(p,q)/\mathrm{SO}(p) \otimes \mathrm{SO}(q)$ | $pq$ | $\min(p,q)$ | $pq - \frac{1}{2}p(p-1) - \frac{1}{2}q(q-1)$ |
| $D_{p+q}$ | $\mathrm{SO}(p,q)/\mathrm{SO}(p) \otimes \mathrm{SO}(q)$ | $pq$ | $\min(p,q)$ | $pq - \frac{1}{2}p(p-1) - \frac{1}{2}q(q-1)$ |
| $D_n$ | $\mathrm{SO}^*(2n)/\mathrm{U}(n)$ | $n(n-1)$ | $n/2$ | $-n$ |
| $C_{p+q}$ | $\mathrm{USp}(2p,2q)/\mathrm{USp}(2p) \otimes \mathrm{USp}(2q)$ | $4pq$ | $\min(p,q)$ | $-2(p-q)^2 - (p+q)$ |
| $C_n$ | $\mathrm{Sp}(2n;R)/\mathrm{U}(n)$ | $n(n+1)$ | $n$ | $+n$ |

For $\sigma=-1$, the space is the sphere $S^2$ defined by $z^2+(x^2+y^2)=1$. For $\sigma=+1$, the space is the two-sheeted hyperboloid $H_2^2$ defined by $z^2-(x^2+y^2)=1$. More specifically, it is the upper sheet containing $(0,0,1)$ of the two-sheeted hyperboloid. The second sheet occurs in the coset O(2,1)/SO(2). The symmetric spaces $\mathrm{SO}(n+1)/\mathrm{SO}(n)$ and $\mathrm{SO}(n,1)/\mathrm{SO}(n)$ are the sphere $S^n$ and the upper sheet of the two-sheeted hyperboloid $H_{2+}^n$. Both have dimension $n$ and rank 1. The spaces are simply connected, homogeneous, and isotropic.

For $\mathrm{SO}(4,2)/\mathrm{SO}(4)\times\mathrm{SO}(2)$, the eight-dimensional algebraic manifold is defined by the three constraints in $R^{11}$:

$$\begin{bmatrix} y_9 & y_{10} \\ y_{10} & y_{11} \end{bmatrix}^2 - \sigma \begin{bmatrix} y_1 & y_2 & y_3 & y_4 \\ y_5 & y_6 & y_7 & y_8 \end{bmatrix} \begin{bmatrix} y_1 & y_5 \\ y_2 & y_6 \\ y_3 & y_7 \\ y_4 & y_8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The compact analytically continued space $\mathrm{SO}(6)/\mathrm{SO}(4)\times\mathrm{SO}(2)$ is obtained by setting $\sigma=-1$. These spaces have dimension 8 and rank 2. They are homogeneous but not isotropic. For each, there are "two inequivalent directions." There are two independent Laplace–Beltrami operators on these spaces, one quadratic and one quartic.

The complete list of globally symmetric pseudo-Riemannian symmetric spaces can be constructed almost as easily. Two linear operators, $T_1$ and $T_2$, are introduced that obey $T_1^2=I, T_2^2=I, T_1T_2=T_2T_1\neq I$. The two are used to split $\mathfrak{g}$ into subspaces

$$T_1\,\mathfrak{g}_{\sigma\tau}=\sigma\,\mathfrak{g}_{\sigma\tau}, \qquad T_2\,\mathfrak{g}_{\sigma\tau}=\tau\,\mathfrak{g}_{\sigma\tau}$$

where $\sigma=\pm1, \tau=\pm1$. The decomposition and double rotation

$$\begin{aligned} \mathfrak{g} &= \mathfrak{g}_{++}+\mathfrak{g}_{+-}+\mathfrak{g}_{-+}+\mathfrak{g}_{--} \\ &\downarrow T_1 \\ \mathfrak{g}' &= \mathfrak{g}_{++}+\mathfrak{g}_{+-}+\mathrm{i}(\mathfrak{g}_{-+}+\mathfrak{g}_{--}) \\ &\downarrow T_2 \\ \mathfrak{g}'' &= \mathfrak{g}_{++}+i\mathfrak{g}_{+-}+\mathrm{i}(\mathfrak{g}_{-+}+\mathrm{i}\mathfrak{g}_{--}) \end{aligned}$$

generates a noncompact subgroup $K''$ as well as a pseudo-Riemannian symmetric space $P''$:

$$K''=\mathrm{EXP}(\mathfrak{g}_{++}+\mathrm{i}\mathfrak{g}_{+-}), \qquad P''=\mathrm{EXP}(\mathrm{i}\mathfrak{g}_{-+}+\mathfrak{g}_{--})$$

These have also been classified.

The simplest example of a pseudo-Riemannian symmetric space is SO(2,1)/SO(1,1):

$$\mathfrak{so}(2,1)\rightarrow \left[\begin{array}{cc|c} 0 & \theta_3 & \theta_2 \\ -\theta_3 & 0 & \theta_1 \\ \hline \theta_2 & \theta_1 & 0 \end{array}\right] \rightarrow \left[\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & 0 & \theta_1 \\ \hline 0 & \theta_1 & 0 \end{array}\right]$$

$$+\left[\begin{array}{cc|c} 0 & \theta_3 & \theta_2 \\ -\theta_3 & 0 & 0 \\ \hline \theta_2 & 0 & 0 \end{array}\right] \rightarrow M=\left[\begin{array}{cc|c} z & x & y \\ -x & * & * \\ \hline y & * & * \end{array}\right]$$

The metric-preserving condition $M^tI_{2,1}M=I_{2,1}$ leads to the constraint equation $z^2+x^2-y^2=1$. This space is the single-sheeted hyperboloid $H_1^2$. It is two dimensional and has rank 1, but it is not isotropic. Intersections with the plane $x=0$ are hyperbolas and with the planes $y=$ const. are circles. This space is not simply connected.

## Summary

Lie groups are among the most powerful mathematical tools available to physicists. They play a major role in physics because they occur as transformation groups from coordinate system to coordinate system in real space (rotation group SO(3), Lorentz group O(3,1), Galilei group, Poincaré group $I$SO(3,1)) or in spaces describing internal degrees of freedom (SU(2) for spin or isospin, SU(3) for quarks and color, SU(4) for spin–isospin, etc.).

It is remarkable that a beautiful classification theory for simple (the building blocks) Lie groups exists, because of the rather amorphous nature of the definition of a Lie group. In a search for structure, the first step in the analysis of Lie groups is linearization of the group multiplication law in the neighborhood of the identity to a linear vector space on which there is a Lie algebra structure. This in itself is sufficient to create a strong connection to quantum mechanics. Although there is not a 1:1 correspondence between Lie groups and their Lie algebras, there is a very beautiful connection between them. This relates algebra (discrete invariant subgroups) and topology (homotopy groups) in an elegant way.

The structure of Lie algebras is described using tools from linear algebra: secular equations and inner products. Together, these tools are used to reduce Lie algebras to their basic units: nilpotent and solvable invariant subalgebras, and semisimple and simple Lie algebras. The commutation relations for simple Lie algebras can be put into a canonical form using another miracle of this theory: a positive-definite root space that summarizes the properties of the secular equation and the Cartan–Killing inner

product. As the secular equation can only be solved exactly over an algebraically closed field, the classification of simple Lie algebras covers complex Lie algebras. Each complex extension has several real forms, which are easily classified.

Even more remarkable is the connection between simple Lie groups and Riemannian spaces that "look the same everywhere." All Riemannian symmetric spaces are quotients of a simple Lie group by a subgroup that is maximal in some precise sense (Cartan decomposition sense). Cartan was able to classify all Riemannian symmetric spaces as a consequence of his classification of all the real forms of all the simple Lie groups. The algebraic tools used to classify Lie algebras (secular equations, Dynkin diagrams) were used again to classify these spaces (Dynkin diagrams → Araki–Satake diagrams). These spaces are classified by a root space, group–subgroup pair, dimension, rank, and character. Construction of invariant operators (Casimir invariants, Laplace–Beltrami operators) is algorithmic.

Nonsemisimple Lie groups/algebras can be constructed from simple Lie algebras by carefully introducing singular change of basis transformations. This leads to "group contraction," not discussed above. In this way, the Poincaré group can be constructed systematically from the groups SO(3,2) or SO(4,1): $SO(3,2) \rightarrow ISO(3,1), SO(4,1) \rightarrow ISO(3,1)$ in the limit of "large $R$." Here, $R$ is the "radius" of some universe of hyperbolic nature, with signature (3,2) or (4,1). The Galilei group can be constructed by contraction from the Poincaré group in the limit $c = 3 \times 10^{10}\,\mathrm{cm\,s^{-1}} \rightarrow \infty$.

We have not discussed here the theory of the representations of Lie groups. A beautiful theorem by Wigner and Stone guarantees that the tensor representations of a compact group are complete. Gel'fand has given expressions for the complete set of tensor representations of the classical compact Lie groups. They are expressed by "dressing" the appropriate Dynkin diagrams or else in terms of irreducible representations of the symmetric group $S_n$. Gel'fand has also given explicit, analytic, closed-form expressions for the matrix elements of any of the shift operators in any of these representations. For the noncompact real forms, most of the unitary irreducible representations can be obtained from these expressions for matrix elements ("master analytic representation") by appropriate analytic continuation.

Since Lie groups exist at the interface of algebra and topology, it is to be expected that there is a very close relation with the theory of special functions. In fact, the theory of special functions forms an important chapter in the theory of Lie groups. On the topological side, the shift operators $E_{\boldsymbol{\alpha}}$ (think $J_{\pm}$) have coordinate representations $\langle x'|E_{\boldsymbol{\alpha}}|x\rangle$ involving first-order differential operators. On the algebraic side, the matrix elements $\langle \boldsymbol{n}'|E_{\boldsymbol{\alpha}}|\boldsymbol{n}\rangle$ are square roots of products of integers (divided by products of integers). These topological and algebraic expressions are related to each other in a myriad of ways. All of the standard properties of special functions (Rodriguez formulas, recursion relations in coordinates and indices, differential equations, generating functions, etc.) occur in a systematic way in a Lie-theoretic formulation of this subject.

Finally, no review or even book could do justice to the applications that Lie group theory finds in physics.

The rich interplay that exists between freedom and rigidity of structure found in Lie group theory can be found in only the purest works of art – for example, the fugues of Bach.

*See also:* Classical Groups and Homogeneous Spaces; Compact Groups and their Representations; Cosmology: Mathematical Aspects; Equivariant Cohomology and the Cartan Model; Finite-Type Invariants of 3-Manifolds; Functional Equations and Integrable Systems; Lie Superalgebras and Their Representations; Lie, Symplectic, and Poisson Groupoids and Their Lie Algebroids; Measure on Loop Spaces; Quasiperiodic Systems; Symmetry and Symplectic Reduction; Symmetry Classes in Random Matrix Theory; Toda Lattices.

## Further Reading

Barut AO and Raczka (1986) *Theory of Group Representations and Applications*. Singapore: World Scientific.

Gilmore R (1974) *Lie Groups, Lie Algebras, and Some of Their Applications*. New York: Wiley (republished (2005); New York: Dover).

Helgason S (1962) *Differential Geometry and Symmetric Spaces*. New York: Academic Press.

Helgason S (1978) *Differential Geometry, Lie Groups, and Symmetric Spaces*. New York: Academic Press.

Talman JD (1968) *Special Functions, a Group Theoretical Approach, Based on Lectures by Eugene P. Wigner*. New York: Benjamin.

# Lie Superalgebras and Their Representations

**L Frappat**, Université de Savoie, Chambery-Annecy, France



## Basic Definitions

Let $\mathcal{A}$ be an algebra over a field $\mathbb{K}$ of characteristic zero (usually $\mathbb{K}=\mathbb{R}$ or $\mathbb{C}$) with internal laws $+$ and $*$. One sets $\mathbb{Z}_2=\mathbb{Z}/2\mathbb{Z}=\{\bar{0},\bar{1}\}$. $\mathcal{A}$ is called a superalgebra or $\mathbb{Z}_2$-graded algebra if it can be written into a direct sum of two spaces $\mathcal{A}=\mathcal{A}_{\bar{0}}\oplus\mathcal{A}_{\bar{1}}$, such that

$$\mathcal{A}_{\bar{0}}*\mathcal{A}_{\bar{0}}\subset\mathcal{A}_{\bar{0}},\quad \mathcal{A}_{\bar{0}}*\mathcal{A}_{\bar{1}}\subset\mathcal{A}_{\bar{1}},\quad \mathcal{A}_{\bar{1}}*\mathcal{A}_{\bar{1}}\subset\mathcal{A}_{\bar{0}}$$

Elements of $\mathcal{A}_{\bar{0}}$ are called even or of degree 0 while elements of $\mathcal{A}_{\bar{1}}$ are called odd or of degree 1. A superalgebra $\mathcal{A}$ is called associative if $(X*Y)*Z=X*(Y*Z)$ for all $X,Y,Z\in\mathcal{A}$. It is called commutative if $X*Y=(-1)^{\deg X.\deg Y}Y*X$ for all $X,Y\in\mathcal{A}$, where $\deg X$ is the degree of the element $X$.

A homomorphism $\Phi$ from a superalgebra $\mathcal{A}$ into a superalgebra $\mathcal{A}'$ is a linear application from $\mathcal{A}$ into $\mathcal{A}'$ which respects the $\mathbb{Z}_2$-gradation, that is, $\Phi(\mathcal{A}_{\bar{0}})\subset\mathcal{A}'_{\bar{0}}$ and $\Phi(\mathcal{A}_{\bar{1}})\subset\mathcal{A}'_{\bar{1}}$.

A Lie superalgebra $\mathcal{G}$ over a field $\mathbb{K}$ of characteristic zero (usually $\mathbb{K}=\mathbb{R}$ or $\mathbb{C}$) is a superalgebra in which the product, denoted $[\![\,,]\!]$, satisfies the following properties:

**$\mathbb{Z}_2$-gradation**

$$[\![\mathcal{G}_i,\mathcal{G}_j]\!]\subset\mathcal{G}_{i+j}\quad(i,j\in\mathbb{Z}_2)$$

**Graded-antisymmetry**

$$[\![X_i,X_j]\!]=-(-1)^{\deg X_i.\deg X_j}[\![X_j,X_i]\!]$$

**Generalized Jacobi identity**

$$\begin{aligned}&(-1)^{\deg X_i.\deg X_k}[\![X_i,[\![X_j,X_k]\!]]\!]\\&\quad+(-1)^{\deg X_j.\deg X_i}[\![X_j,[\![X_k,X_i]\!]]\!]\\&\quad+(-1)^{\deg X_k.\deg X_j}[\![X_k,[\![X_i,X_j]\!]]\!]=0\end{aligned}$$

Note that $\mathcal{G}_{\bar{0}}$ is a Lie algebra, called the even or bosonic part of $\mathcal{G}$, while $\mathcal{G}_{\bar{1}}$, called the odd or fermionic part of $\mathcal{G}$, is not an algebra.

An associative superalgebra $\mathcal{G}=\mathcal{G}_{\bar{0}}\oplus\mathcal{G}_{\bar{1}}$ over the field $\mathbb{K}$ acquires the structure of a Lie superalgebra by taking for the product $[\![\,,]\!]$ of two elements $X,Y\in\mathcal{G}$ the Lie superbracket (also called supercommutator or graded commutator)

$$[\![X,Y]\!]=X*Y-(-1)^{\deg X.\deg Y}Y*X$$

The notation $[\![\,,]\!]$ for the supercommutator is used to avoid confusion with the usual commutator $[X,Y]=X*Y-Y*X$.

A Lie superalgebra $\mathcal{G}$ is $\mathbb{Z}$-graded if it can be written as a direct sum of finite-dimensional $\mathbb{Z}_2$-graded subspaces $\mathcal{G}_i$ such that

$$\mathcal{G}=\bigoplus_{i\in\mathbb{Z}}\mathcal{G}_i,\quad\text{where }[\![\mathcal{G}_i,\mathcal{G}_j]\!]\subset\mathcal{G}_{i+j}$$

The $\mathbb{Z}$-gradation is said to be consistent with the $\mathbb{Z}_2$-gradation if

$$\mathcal{G}_{\bar{0}}=\sum_{i\in\mathbb{Z}}\mathcal{G}_{2i}\quad\text{and}\quad\mathcal{G}_{\bar{1}}=\sum_{i\in\mathbb{Z}}\mathcal{G}_{2i+1}$$

It follows that $\mathcal{G}_0$ is a Lie subalgebra and that each $\mathcal{G}_i(i\neq0)$ is a $\mathcal{G}_0$-module.

A subalgebra $\mathcal{K}=\mathcal{K}_{\bar{0}}\oplus\mathcal{K}_{\bar{1}}$ of a Lie superalgebra $\mathcal{G}$ is a subset of elements of $\mathcal{G}$ which forms a vector subspace of $\mathcal{G}$ that is closed with respect to the Lie product of $\mathcal{G}$ such that $\mathcal{K}_{\bar{0}}\subset\mathcal{G}_{\bar{0}}$ and $\mathcal{K}_{\bar{1}}\subset\mathcal{G}_{\bar{1}}$. A subalgebra $\mathcal{K}$ of $\mathcal{G}$ is called a proper subalgebra of $\mathcal{G}$ if $\mathcal{K}\neq\mathcal{G}$. An ideal $\mathcal{I}$ of $\mathcal{G}$ is a subalgebra of $\mathcal{G}$ such that $[\![\mathcal{G},\mathcal{I}]\!]\subset\mathcal{I}$, that is, $X\in\mathcal{G},Y\in\mathcal{I}\Rightarrow[\![X,Y]\!]\in\mathcal{I}$. An ideal $\mathcal{I}$ of $\mathcal{G}$ is called a proper ideal of $\mathcal{G}$ if $\mathcal{I}\neq\mathcal{G}$. If $\mathcal{I}$ and $\mathcal{I}'$ are two ideals of $\mathcal{G}$, $[\![\mathcal{I},\mathcal{I}']\!]$ is an ideal of $\mathcal{G}$.

The definitions of the centralizer, the center, and the normalizer of a Lie superalgebra follow those of a Lie algebra. Let $\mathcal{S}$ be a subset of elements in the Lie superalgebra $\mathcal{G}$. The centralizer $\mathcal{C}_{\mathcal{G}}(\mathcal{S})$ is the subset of $\mathcal{G}$ given by

$$\mathcal{C}_{\mathcal{G}}(\mathcal{S})=\{X\in\mathcal{G}\,|\,[\![X,Y]\!]=0,\ \forall Y\in\mathcal{S}\}$$

The center $\mathcal{Z}(\mathcal{G})$ of $\mathcal{G}$ is the set of elements of $\mathcal{G}$ which commute with any element of $\mathcal{G}$ (in other words, it is the centralizer of $\mathcal{G}$ in $\mathcal{G}$):

$$\mathcal{Z}(\mathcal{G})=\{X\in\mathcal{G}\,|\,[\![X,Y]\!]=0,\ \forall Y\in\mathcal{G}\}$$

The normalizer $\mathcal{N}_{\mathcal{G}}(\mathcal{S})$ is the subset of $\mathcal{G}$ given by

$$\mathcal{N}_{\mathcal{G}}(\mathcal{S})=\{X\in\mathcal{G}\,|\,[\![X,Y]\!]\in\mathcal{S},\ \forall Y\in\mathcal{S}\}$$

The Lie superalgebra $\mathcal{G}$ is said to be nilpotent if considering the series $[\![\mathcal{G},\mathcal{G}^{[i-1]}]\!]=\mathcal{G}^{[i]}$ with $\mathcal{G}^{[0]}=\mathcal{G}$, then there exists an integer $n$ such that $\mathcal{G}^{[n]}=\{0\}$.

The Lie superalgebra $\mathcal{G}$ is said to be solvable if considering the series $[\![\mathcal{G}^{(i-1)},\mathcal{G}^{(i-1)}]\!]=\mathcal{G}^{(i)}$ with $\mathcal{G}^{(0)}=\mathcal{G}$, then there exists an integer $n$ such that $\mathcal{G}^{(n)}=\{0\}$. A Lie superalgebra $\mathcal{G}$ is solvable if and only if $\mathcal{G}_{\bar{0}}$ is solvable.

Let $\mathcal{G}$ be a noncommutative Lie superalgebra. The Lie superalgebra $\mathcal{G}$ is called simple if it does not contain any nontrivial ideal. The Lie superalgebra $\mathcal{G}$ is called semisimple if it does not

contain any nontrivial solvable ideal. Let us recall that if $\mathcal{A}$ is a semisimple Lie algebra, it can be written as the direct sum of simple Lie algebras $\mathcal{A}_i : \mathcal{A} = \oplus_i \mathcal{A}_i$. This is not the case for superalgebras.

Let $\mathcal{G} = \mathcal{G}_{\bar{0}} \oplus \mathcal{G}_{\bar{1}}$ be a Lie superalgebra and $\mathcal{V} = \mathcal{V}_{\bar{0}} \oplus \mathcal{V}_{\bar{1}}$ be a $\mathbb{Z}_2$-graded vector space. Consider the algebra $\mathrm{End}\,\mathcal{V}$ of endomorphisms of $\mathcal{V}$, which naturally acquires a superalgebra structure by $\mathrm{End}\,\mathcal{V} = \mathrm{End}_{\bar{0}}\,\mathcal{V} \oplus \mathrm{End}_{\bar{1}}\,\mathcal{V}$, where $\mathrm{End}_i\,\mathcal{V} = \{\phi \in \mathrm{End}\,\mathcal{V} | \phi(\mathcal{V}_j) \subset \mathcal{V}_{i+j}\}$. A linear representation $\pi$ of $\mathcal{G}$ is a homomorphism of $\mathcal{G}$ into $\mathrm{End}\,\mathcal{V}$, that is,

$$\pi(\alpha X + \beta Y) = \alpha\pi(X) + \beta\pi(Y)$$
$$\pi([\![X, Y]\!]) = [\![\pi(X), \pi(Y)]\!]$$
$$\pi(\mathcal{G}_{\bar{0}}) \subset \mathrm{End}_{\bar{0}}\mathcal{V} \quad \text{and} \quad \pi(\mathcal{G}_{\bar{1}}) \subset \mathrm{End}_{\bar{1}}\mathcal{V}$$

for all $X, Y \in \mathcal{G}$ and $\alpha, \beta \in \mathbb{C}$. The vector space $\mathcal{V}$ is the representation space. The vector space $\mathcal{V}$ has the structure of a $\mathcal{G}$-module by $X(\boldsymbol{v}) = \pi(X)\boldsymbol{v}$ for $X \in \mathcal{G}$ and $\boldsymbol{v} \in \mathcal{V}$. The dimension (resp. superdimension) of the representation $\pi$ is the dimension (resp. graded dimension) of the vector space $\mathcal{V}$: $\dim \pi = \dim \mathcal{V}_{\bar{0}} + \dim \mathcal{V}_{\bar{1}}$ and $\mathrm{sdim}\pi = \dim \mathcal{V}_{\bar{0}} - \dim \mathcal{V}_{\bar{1}}$. In particular, the representation $\mathrm{ad} : \mathcal{G} \to \mathrm{End}\,\mathcal{G}$ ($\mathcal{G}$ being considered as a $\mathbb{Z}_2$-graded vector space) such that $\mathrm{ad}(X)Y = [\![X, Y]\!]$ is called the adjoint representation of $\mathcal{G}$.

In the basis $(e_1, \ldots, e_m, e_{m+1}, \ldots, e_{m+n})$ of $\mathcal{V} = \mathcal{V}_{\bar{0}} \oplus \mathcal{V}_{\bar{1}}$ (called homogeneous basis), where $\dim \mathcal{V}_{\bar{0}} = m$ and $\dim \mathcal{V}_{\bar{1}} = n$, an element of $\mathcal{G}$ is represented by the matrix

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

where $A$, $B$, $C$, and $D$ are $m \times m, m \times n, n \times m$, and $n \times n$ matrices, respectively. Even elements correspond to block diagonal matrices (i.e., $B = C = 0$), odd elements to block antidiagonal matrices (i.e., $A = D = 0$). One defines the supertrace function denoted by str:

$$\mathrm{str}(M) = \mathrm{tr}(A) - \mathrm{tr}(D)$$

To a given representation $\pi$ of $\mathcal{G}$, one can associate a bilinear form $B_\pi$ on $\mathcal{G}$ as

$$B_\pi(X, Y) = \mathrm{str}(\pi(X)\pi(Y)), \quad \forall X, Y \in \mathcal{G}$$

$\pi(X)$ are the matrices of the generators $X$ in the representation $\pi$ and str denotes the supertrace. A bilinear form $B$ on $\mathcal{G}$ is called

1. consistent if $B(X, Y) = 0$ for all $X \in \mathcal{G}_{\bar{0}}$ and all $Y \in \mathcal{G}_{\bar{1}}$,
2. supersymmetric if, for all $X, Y \in \mathcal{G}$,
$$B(X, Y) = (-1)^{\deg X . \deg Y} B(Y, X)$$
3. invariant if, for all $X, Y, Z \in \mathcal{G}$,
$$B([\![X, Y]\!], Z) = B(X, [\![Y, Z]\!])$$

The bilinear form associated to the adjoint representation of $\mathcal{G}$ is called the Killing form on $\mathcal{G} : K(X, Y) = \mathrm{str}(\mathrm{ad}(X)\mathrm{ad}(Y))$. It is consistent, supersymmetric, and invariant.

## Classification of Simple Lie Superalgebras

The simple Lie superalgebras have been classified by V G Kac. One distinguishes two general families: the classical Lie superalgebras and the Cartan type superalgebras.

### Classical Lie Superalgebras

A simple Lie superalgebra $\mathcal{G} = \mathcal{G}_{\bar{0}} \oplus \mathcal{G}_{\bar{1}}$ is called classical if the representation of the even subalgebra $\mathcal{G}_{\bar{0}}$ on the odd part $\mathcal{G}_{\bar{1}}$ is completely reducible. The superalgebra is said to be of type I if the representation of $\mathcal{G}_{\bar{0}}$ on $\mathcal{G}_{\bar{1}}$ is the direct sum of two irreducible representations of $\mathcal{G}_{\bar{0}}$. In that case, one has $\mathcal{G}_{\bar{1}} = \mathcal{G}_{-1} \oplus \mathcal{G}_1$ with

$$[\![\mathcal{G}_{-1}, \mathcal{G}_1]\!] = \mathcal{G}_{\bar{0}} \quad \text{and} \quad [\![\mathcal{G}_{\pm 1}, \mathcal{G}_{\pm 1}]\!] = 0$$

The superalgebra is said to be of type II if the representation of $\mathcal{G}_{\bar{0}}$ on $\mathcal{G}_{\bar{1}}$ is irreducible.

A classical Lie superalgebra $\mathcal{G}$ is called basic if there exists a nondegenerate invariant bilinear form on $\mathcal{G}$. The basic Lie superalgebras split into four infinite families: $A(m, n)$ or $\mathrm{sl}(m+1|n+1)$ for $m \neq n$ and $A(n, n)$ or $\mathrm{sl}(n+1|n+1)/\mathcal{Z} = \mathrm{psl}(n+1|n+1)$, where $\mathcal{Z}$ is a one-dimensional center for $m = n$ (unitary series), $B(m, n)$ or $\mathrm{osp}(2m+1|2n)$, $C(n)$ or $\mathrm{osp}(2|2n)$, $D(m, n)$ or $\mathrm{osp}(2m\,|\,2n)$ (orthosymplectic series); and three exceptional superalgebras $F(4)$, $G(3)$, and $D(2, 1; \alpha)$, the last one being actually a one-parameter family of superalgebras. The classical Lie superalgebras which are not basic are called strange, and correspond to two infinite families denoted by $P(n)$ and $Q(n)$.

A basic Lie superalgebra $\mathcal{G} = \mathcal{G}_{\bar{0}} \oplus \mathcal{G}_{\bar{1}}$ admits a consistent $\mathbb{Z}$-gradation $\mathcal{G} = \oplus_{i \in \mathbb{Z}} \mathcal{G}_i$ (called distinguished), such that (see Tables 1 and 2)

- for superalgebras of type I, $\mathcal{G}_i = 0$ for $|i| > 1$ and $\mathcal{G}_{\bar{0}} = \mathcal{G}_0, \mathcal{G}_{\bar{1}} = \mathcal{G}_{-1} \oplus \mathcal{G}_1$ and
- for superalgebras of type II, $\mathcal{G}_i = 0$ for $|i| > 2$ and $\mathcal{G}_{\bar{0}} = \mathcal{G}_{-2} \oplus \mathcal{G}_0 \oplus \mathcal{G}_2, \mathcal{G}_{\bar{1}} = \mathcal{G}_{-1} \oplus \mathcal{G}_1$.

### Cartan Type Superalgebras

The Cartan type Lie superalgebras are the simple Lie superalgebras in which the representation of the even subalgebra on the odd part is not completely

**Table 1** $\mathbb{Z}_2$-gradation of the classical Lie superalgebras

| *Superalgebra* $\mathcal{G}$ | $\mathcal{G}_{\bar{0}}$ | $\mathcal{G}_{\bar{1}}$ |
|---|---|---|
| $A(m-1,n-1)$ | $A_{m-1}\oplus A_{n-1}\oplus U(1)$ | $(m,\bar{n})\oplus(\bar{m},n)$ |
| $A(n-1,n-1)$ | $A_{n-1}\oplus A_{n-1}$ | $(n,\bar{n})\oplus(\bar{n},n)$ |
| $C(n+1)$ | $C_n\oplus U(1)$ | $(2n)\oplus(2n)$ |
| $B(m,n)$ | $B_m\oplus C_n$ | $(2m+1,2n)$ |
| $D(m,n)$ | $D_m\oplus C_n$ | $(2m,2n)$ |
| $F(4)$ | $A_1\oplus B_3$ | $(2,8)$ |
| $G(3)$ | $A_1\oplus G_2$ | $(2,7)$ |
| $D(2,1;\alpha)$ | $A_1\oplus A_1\oplus A_1$ | $(2,2,2)$ |
| $P(n)$ | $A_n$ | $[2]\oplus[1^{n-1}]$ |
| $Q(n)$ | $A_n$ | $\mathrm{ad}(A_n)$ |

reducible. They are classified into four infinite families called $W(n)$ with $n\geq 2, S(n)$ with $n\geq 3$, $\widetilde{S}(n)$, and $H(n)$ with $n\geq 4$. $S(n)$ and $\widetilde{S}(n)$ are called special Cartan type Lie superalgebras and $H(n)$ Hamiltonian Cartan type Lie superalgebras.

## Classical Lie Superalgebras

The classical Lie superalgebras are described as matrix superalgebras as follows. Let $\mathcal{V}=\mathcal{V}_{\bar{0}}\oplus\mathcal{V}_{\bar{1}}$ be a $\mathbb{Z}_2$-graded vector space, with $\dim\mathcal{V}_{\bar{0}}=m$, $\dim\mathcal{V}_{\bar{1}}=n$. The Lie superalgebra $\mathrm{gl}(m|n)$ is defined as the superalgebra $\mathrm{End}\,\mathcal{V}=\mathrm{End}_{\bar{0}}\,\mathcal{V}\oplus\mathrm{End}_{\bar{1}}\,\mathcal{V}$ supplied with the Lie superbracket.

The unitary superalgebra $A(m-1,n-1)=\mathrm{sl}(m\,|\,n)$ is defined as the superalgebra of matrices $M\in\mathrm{gl}(m|n)$ satisfying the supertrace condition $\mathrm{str}(M)=0$. In the case $m=n, \mathrm{sl}(n|n)$ contains a one-dimensional ideal $\mathcal{I}$ generated by $\mathbb{I}_{2n}$ and one sets $A(n-1,n-1)=\mathrm{sl}(n\,|\,n)/\mathcal{I}\equiv\mathrm{psl}(n\,|\,n)$.

The orthosymplectic superalgebra $\mathrm{osp}(m\,|\,2n)$ is defined as the superalgebra of matrices $M\in\mathrm{gl}(m\,|\,n)$ satisfying the conditions

$$A^{\mathrm{t}}=-A,\quad D^{\mathrm{t}}G=-GD,\quad B=C^{\mathrm{t}}G$$

where t denotes the usual transposition and the matrix $G$ is given by

$$G=\begin{pmatrix}0 & \mathbb{I}_n\\ -\mathbb{I}_n & 0\end{pmatrix}$$

The strange superalgebra $P(n)$ is defined as the superalgebra of matrices $M\in\mathrm{gl}(n\,|\,n)$ satisfying the conditions

$$A^{\mathrm{t}}=-D,\quad B^{\mathrm{t}}=B,\quad C^{\mathrm{t}}=-C,\quad \mathrm{tr}(A)=0$$

The strange superalgebra $\widetilde{Q}(n)$ is defined as the superalgebra of matrices $M\in\mathrm{gl}(n\,|\,n)$ satisfying the conditions

$$A=D,\quad B=C,\quad \mathrm{tr}(B)=0$$

The superalgebra $\widetilde{Q}(n)$ has a one-dimensional center $\mathcal{Z}$. The simple superalgebra $Q(n)$ is given by $Q(n)=\widetilde{Q}(n)/\mathcal{Z}$.

## Structure of the Classical Lie Superalgebras

Let $\mathcal{G}=\mathcal{G}_{\bar{0}}\oplus\mathcal{G}_{\bar{1}}$ be a classical Lie superalgebra. A Cartan subalgebra $\mathcal{H}$ of $\mathcal{G}$ is defined as a Cartan subalgebra of $\mathcal{G}_{\bar{0}}$, that is, the maximal nilpotent subalgebra of $\mathcal{G}_{\bar{0}}$ coinciding with its own normalizer: $\mathcal{H}=\{X\in\mathcal{G}_{\bar{0}}\,|\,[\![X,\mathcal{H}]\!]\subseteq\mathcal{H}\}$. It follows that the Cartan subalgebras of a Lie superalgebra are conjugate since the Cartan subalgebras of a Lie algebra are conjugate and any inner automorphism of the even part $\mathcal{G}_{\bar{0}}$ can be extended to an inner automorphism of $\mathcal{G}$; hence, they all have the same dimension. By definition, the dimension of a Cartan subalgebra $\mathcal{H}$ is the rank of $\mathcal{G}:\mathrm{rank}\,\mathcal{G}=\dim\mathcal{H}$.

A classical Lie superalgebra $\mathcal{G}$ with Cartan subalgebra $\mathcal{H}$ can be decomposed as $\mathcal{G}=\bigoplus_{\alpha\in\mathcal{H}^*}\mathcal{G}_\alpha$ ($\mathcal{H}^*$ is the dual of $\mathcal{H}$), where

$$\mathcal{G}_\alpha=\{x\in\mathcal{G}\,|\,[\![h,x]\!]=\alpha(h)x, h\in\mathcal{H}\}$$

The set $\Delta\subset\mathcal{H}^*$

$$\Delta=\{\alpha\in\mathcal{H}^*|\mathcal{G}_\alpha\neq 0\}$$

is by definition the root system of $\mathcal{G}$. A root $\alpha$ is called even (resp. odd) if $\mathcal{G}_\alpha\cap\mathcal{G}_{\bar{0}}\neq\emptyset$ (resp.

**Table 2** $\mathbb{Z}$-gradation of the classical basic Lie superalgebras

| *Superalgebra* $\mathcal{G}$ | $\mathcal{G}_0$ | $\mathcal{G}_1\oplus\mathcal{G}_{-1}$ | $\mathcal{G}_2\oplus\mathcal{G}_{-2}$ |
|---|---|---|---|
| $A(m-1,n-1)$ | $A_{m-1}\oplus A_{n-1}\oplus U(1)$ | $(m,\bar{n})\oplus(\bar{m},n)$ | |
| $A(n-1,n-1)$ | $A_{n-1}\oplus A_{n-1}$ | $(n,\bar{n})\oplus(\bar{n},n)$ | |
| $C(n+1)$ | $C_n\oplus U(1)$ | $(2n)_+\oplus(2n)_-$ | |
| $B(m,n)$ | $B_m\oplus A_{n-1}\oplus U(1)$ | $(2m+1,n)\oplus(2m+1,\bar{n})$ | $[2]\oplus[2^{n-1}]$ |
| $D(m,n)$ | $D_m\oplus A_{n-1}\oplus U(1)$ | $(2m,n)\oplus(2m,\bar{n})$ | $[2]\oplus[2^{n-1}]$ |
| $F(4)$ | $B_3\oplus U(1)$ | $8_+\oplus 8_-$ | $1_+\oplus 1_-$ |
| $G(3)$ | $G_2\oplus U(1)$ | $7_+\oplus 7_-$ | $1_+\oplus 1_-$ |
| $D(2,1;\alpha)$ | $A_1\oplus A_1\oplus U(1)$ | $(2,2)_+\oplus(2,2)_-$ | $1_+\oplus 1_-$ |

$\mathcal{G}_\alpha \cap \mathcal{G}_{\bar{1}} \neq \emptyset$). The set of even roots $\Delta_{\bar{0}}$ is the root system of the even part $\mathcal{G}_{\bar{0}}$ of $\mathcal{G}$. The set of odd root $\Delta_{\bar{1}}$ is the weight system of the representation of $\mathcal{G}_{\bar{0}}$ in $\mathcal{G}_{\bar{1}}$. One has $\Delta = \Delta_{\bar{0}} \cup \Delta_{\bar{1}}$. A root can be both even and odd (however this only occurs in the case of the superalgebra $Q(n)$). The vector space spanned by all the possible roots is called the root space. It is the dual $\mathcal{H}^*$ of the Cartan subalgebra $\mathcal{H}$ as vector space.

Except for $A(1,1), P(n)$, and $Q(n)$, using a nondegenerate invariant bilinear form $B$ on the superalgebra $\mathcal{G}$, one can define a bilinear form $(\cdot,\cdot)$ on the root space $\mathcal{H}^*$ by $(\alpha_i,\alpha_j) = B(H_i,H_j)$, where the $H_i$ form a basis of $\mathcal{H}$. The following properties hold:

1. $\mathcal{G}_{(\alpha=0)} = \mathcal{H}$ except for $Q(n)$.
2. dim $\mathcal{G}_\alpha = 1$ when $\alpha \neq 0$ except for $A(1,1), P(2)$, $P(3)$, and $Q(n)$.
3. Except for $A(1,1), P(n), Q(n)$, one has
   (a) $[\![\mathcal{G}_\alpha, \mathcal{G}_\beta]\!] \neq 0$ if and only if $\alpha, \beta, \alpha+\beta \in \Delta$,
   (b) $(\mathcal{G}_\alpha, \mathcal{G}_\beta) = 0$ for $\alpha + \beta \neq 0$,
   (c) if $\alpha \in \Delta$ (resp. $\Delta_{\bar{0}}, \Delta_{\bar{1}}$), then $-\alpha \in \Delta$ (resp. $\Delta_{\bar{0}}, \Delta_{\bar{1}}$), and
   (d) $\alpha \in \Delta \Rightarrow 2\alpha \in \Delta$ if and only if $\alpha \in \Delta_{\bar{1}}$ and $(\alpha,\alpha) \neq 0$.

In the rest of this section, we restrict to the case of a basic Lie superalgebra $\mathcal{G}$ of rank $r$, with Cartan subalgebra $\mathcal{H}$ and root system $\Delta = \Delta_{\bar{0}} \cup \Delta_{\bar{1}}$. Then $\mathcal{G}$ admits a Borel decomposition $\mathcal{G} = \mathcal{N}^+ \oplus \mathcal{H} \oplus \mathcal{N}^-$, where $\mathcal{N}^\pm$ are subalgebras such that $[\mathcal{H}, \mathcal{N}^\pm] \subset \mathcal{N}^\pm$ with dim $\mathcal{N}^+ =$ dim $\mathcal{N}^-$. If $\mathcal{G} = \mathcal{H} \bigoplus_\alpha \mathcal{G}_\alpha$ is the root decomposition of $\mathcal{G}$, a root $\alpha$ is called positive if $\mathcal{G}_\alpha \cap \mathcal{N}^+ \neq \emptyset$ and negative if $\mathcal{G}_\alpha \cap \mathcal{N}^- \neq \emptyset$. A root is called simple if it cannot be decomposed into a sum of positive roots. The set of all simple roots is called a simple root system of $\mathcal{G}$ and is denoted here by $\Delta^0$. The set $\mathcal{B} = \mathcal{H} \oplus \mathcal{N}^+$ is called a Borel subalgebra of $\mathcal{G}$. Such a Borel subalgebra is solvable but not maximal solvable. Indeed, adding to $\mathcal{B}$ a negative simple isotropic root generator (i.e., a generator associated to an odd root of zero length), the obtained subalgebra is still solvable since the superalgebra sl(1|1) is solvable. However, $\mathcal{B}$ contains a maximal solvable subalgebra $\mathcal{B}_{\bar{0}}$ of the even part $\mathcal{G}_{\bar{0}}$.

In general, for a basic Lie superalgebra $\mathcal{G}$, there are many inequivalent classes of conjugacy of Borel subalgebras (while for the simple Lie algebras, all Borel subalgebras are conjugate).

To each class of conjugacy of Borel subalgebras of $\mathcal{G}$ is associated a simple root system $\Delta^0$. Hence, contrary to the Lie algebra case, to a given basic Lie superalgebra $\mathcal{G}$ will be associated in general many inequivalent simple root systems, up to a transformation of the Weyl group $W(\mathcal{G})$ of $\mathcal{G}$ (the Weyl group of a basic Lie superalgebra being generated by the Weyl reflections with respect to the even roots; under a transformation of $W(\mathcal{G})$, a simple root system will be transformed into an equivalent one with the same Dynkin diagram). The generalization of the Weyl group for a basic Lie superalgebra $\mathcal{G}$ gives a method for constructing all the simple root systems of $\mathcal{G}$ and hence all the inequivalent Dynkin diagrams of $\mathcal{G}$. For $\alpha \in \Delta_{\bar{1}}$, one defines

$$
\begin{aligned}
w_\alpha(\beta) &= \beta - 2\frac{(\alpha,\beta)}{(\alpha,\alpha)}\alpha \quad \text{if } (\alpha,\alpha) \neq 0 \\
w_\alpha(\beta) &= \beta + \alpha \quad \text{if } (\alpha,\alpha) = 0,\ (\alpha,\beta) \neq 0 \\
w_\alpha(\beta) &= \beta \quad \text{if } (\alpha,\alpha) = (\alpha,\beta) = 0 \\
w_\alpha(\alpha) &= -\alpha
\end{aligned}
$$

Note that the transformation associated to an odd root $\alpha$ of zero length cannot be lifted to an automorphism of the superalgebra since $w_\alpha$ transforms even roots into odd ones, and vice versa, and the $\mathbb{Z}_2$-gradation would not be respected. A simple root system $\Delta^0$ being given, from any root $\alpha \in \Delta^0$ such that $(\alpha,\alpha) = 0$, one constructs the simple root system $w_\alpha(\Delta^0)$, where $w_\alpha$ is the generalized Weyl reflection with respect to $\alpha$ and one repeats the procedure on the obtained system until no new basis arises.

In the set of all inequivalent simple root systems of a basic Lie superalgebra, there is one simple root system that plays a particular role, the distinguished simple root system, for which the number of odd roots is equal to one, constructed as follows. Consider the distinguished $\mathbb{Z}$-gradation of $\mathcal{G}, \mathcal{G} = \oplus_{i\in\mathbb{Z}} \mathcal{G}_i$. The even simple roots are given by the simple root system of the Lie subalgebra $\mathcal{G}_0$ and the odd simple root is the lowest weight of the representation $\mathcal{G}_1$ of $\mathcal{G}_0$. See Table 3 for the root systems and Table 4 for the distinguished simple root systems of the basic Lie superalgebras.

Let $\Delta^0 = (\alpha_1, \ldots, \alpha_r)$ be a simple root system of $\mathcal{G}$, such that $(\alpha_i,\alpha_j) \in \mathbb{Z}$ and $|\min(\alpha_i,\alpha_j)| = 1$ if $(\alpha_i,\alpha_j) \neq 0$. Then one defines the symmetric Cartan matrix $a$ with integer entries as $a_{ij} = (\alpha_i,\alpha_j)$. One associates to $\Delta^0$ a Dynkin diagram according to the following rules:

1. One associates to each simple even root a white dot, to each simple odd root of nonzero length ($a_{ii} \neq 0$) a black dot, and to each simple odd root of zero length ($a_{ii} = 0$) a gray dot.

**Table 3** Root systems $\Delta_{\bar{0}}, \Delta_{\bar{1}}$ of the basic Lie superalgebras

| *Superalgebra* $\mathcal{G}$ | $\Delta_{\bar{0}}$ | $\Delta_{\bar{1}}$ |
|---|---|---|
| $A(m-1,n-1)$ | $\varepsilon_i - \varepsilon_j, \delta_k - \delta_l$ | $\pm(\varepsilon_i - \delta_k)$ |
| $B(m,n)$ | $\pm\varepsilon_i \pm \varepsilon_j,\ \pm\varepsilon_i,\ \pm\delta_k \pm \delta_l,\ \pm 2\delta_k$ | $\pm\varepsilon_i \pm \delta_k,\ \pm\delta_k$ |
| $B(0,\ n)$ | $\pm\delta_k \pm \delta_l,\ \pm 2\delta_k$ | $\pm\delta_k$ |
| $C(n+1)$ | $\pm\delta_k \pm \delta_l,\ \pm 2\delta_k$ | $\pm\varepsilon \pm \delta_k$ |
| $D(m,n)$ | $\pm\varepsilon_i \pm \varepsilon_j,\ \pm\delta_k \pm \delta_l,\ \pm 2\delta_k$ | $\pm\varepsilon_i \pm \delta_k$ |
| $F(4)$ | $\pm\delta,\ \pm\varepsilon_i \pm \varepsilon_j,\ \pm\varepsilon_i$ | $\frac{1}{2}(\pm\varepsilon_1 \pm \varepsilon_2 \pm \varepsilon_3 \pm \delta)$ |
| $G(3)$ | $\pm 2\delta,\ \pm\varepsilon_i, \varepsilon_i - \varepsilon_j$ | $\pm\delta,\ \pm\varepsilon_i \pm \delta$ |
| $D(2,1;\alpha)$ | $\pm 2\varepsilon_i$ | $\pm\varepsilon_1 \pm \varepsilon_2 \pm \varepsilon_3$ |

$1 \le i,j \le m, 1 \le k,l \le n$ for $A(m-1,n-1), B(m,n), C(n+1), D(m,n)$. $1 \le i,j \le 3$ for $F(4), G(3), D(2,1;\alpha)$, with $\varepsilon_1 + \varepsilon_2 + \varepsilon_3 = 0$ in the case of $G(3)$. For $A(n-1,n-1)$, one has to add the condition $\varepsilon_1 + \cdots + \varepsilon_n = \delta_1 + \cdots + \delta_n$.

**Table 4** Distinguished simple root systems of the basic Lie superalgebras

| *Superalgebra* $\mathcal{G}$ | *Distinguished simple root system* $\Delta^0$ |
|---|---|
| $A(m-1,n-1)$ | $\delta_1 - \delta_2, \ldots, \delta_{n-1} - \delta_n, \delta_n - \varepsilon_1, \varepsilon_1 - \varepsilon_2, \ldots, \varepsilon_{m-1} - \varepsilon_m$ |
| $B(m,n)$ | $\delta_1 - \delta_2, \ldots, \delta_{n-1} - \delta_n, \delta_n - \varepsilon_1, \varepsilon_1 - \varepsilon_2, \ldots, \varepsilon_{m-1} - \varepsilon_m, \varepsilon_m$ |
| $B(0,n)$ | $\delta_1 - \delta_2, \ldots, \delta_{n-1} - \delta_n, \delta_n$ |
| $C(n)$ | $\varepsilon - \delta_1, \delta_1 - \delta_2, \ldots, \delta_{n-1} - \delta_n, 2\delta_n$ |
| $D(m,n)$ | $\delta_1 - \delta_2, \ldots, \delta_{n-1} - \delta_n, \delta_n - \varepsilon_1, \varepsilon_1 - \varepsilon_2, \ldots, \varepsilon_{m-1} - \varepsilon_m, \varepsilon_{m-1} + \varepsilon_m$ |
| $F(4)$ | $\frac{1}{2}(\delta - \varepsilon_1 - \varepsilon_2 - \varepsilon_3), \varepsilon_3, \varepsilon_2 - \varepsilon_3, \varepsilon_1 - \varepsilon_2$ |
| $G(3)$ | $\delta + \varepsilon_3, \varepsilon_1, \varepsilon_2 - \varepsilon_1$ |
| $D(2,1;\alpha)$ | $\varepsilon_1 - \varepsilon_2 - \varepsilon_3, 2\varepsilon_2, 2\varepsilon_3$ |

2. The $i$th and $j$th dots are joined by $\eta_{ij}$ lines where

$$\eta_{ij} = \frac{2|a_{ij}|}{\min(|a_{ii}|, |a_{jj}|)} \qquad \text{if } a_{ii}.a_{jj} \neq 0$$

$$\eta_{ij} = \frac{2|a_{ij}|}{\min(|a_{ii}|, 2)} \qquad \text{if } a_{ii} \neq 0 \text{ and } a_{jj} = 0$$

$$\eta_{ij} = |a_{ij}| \qquad \text{if } a_{ii} = a_{jj} = 0$$

3. We add an arrow on the lines connecting the $i$th and $j$th dots when $\eta_{ij} > 1$, pointing from $i$ to $j$ if $a_{ii}.a_{jj} \neq 0$ and $|a_{ii}| > |a_{jj}|$ or if $a_{ii} = 0, a_{jj} \neq 0$, $|a_{jj}| < 2$, and pointing from $j$ to $i$ if $a_{ii} = 0$, $a_{jj} \neq 0, |a_{jj}| > 2$.
4. For $D(2,1;\alpha), \eta_{ij} = 1$ if $a_{ij} \neq 0$ and $\eta_{ij} = 0$ if $a_{ij} = 0$. No arrow is put on the Dynkin diagram.

The distinguished Dynkin diagrams of the basic Lie superalgebras are listed in Table 5.

## Representation Theory of Basic Lie Superalgebras

We restrict in the following to the basic Lie superalgebras. We assume that $\mathcal{G} \neq \mathrm{psl}(n,n)$ but the following results still hold for $\mathrm{sl}(n\,|\,n)$. Let $\mathcal{G} = \mathcal{N}^+ \oplus \mathcal{H} \oplus \mathcal{N}^-$ be a Borel decomposition of $\mathcal{G}$ where $\mathcal{N}^+$ (resp. $\mathcal{N}^-$) is spanned by the positive (resp. negative) root generators of $\mathcal{G}, \mathcal{H}$ is a Cartan subalgebra, and $\mathcal{H}^*$ is the dual of $\mathcal{H}$. A representation $\pi : \mathcal{G} \to \mathrm{End}\,\mathcal{V}$ with representation space $\mathcal{V}$ is called a highest-weight representation with highest weight $\Lambda \in \mathcal{H}^*$ if there exists a nonzero vector $\boldsymbol{v}_\Lambda \in \mathcal{V}$ such that

$$\mathcal{N}^+ \boldsymbol{v}_\Lambda = 0$$

$$h(\boldsymbol{v}_\Lambda) = \Lambda(h)\boldsymbol{v}_\Lambda (h \in \mathcal{H})$$

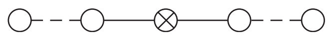
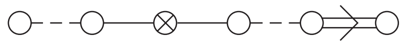
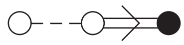
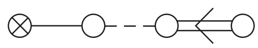
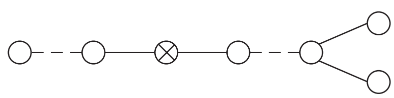
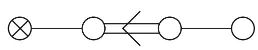
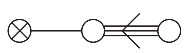
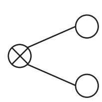
The $\mathcal{G}$-module $\mathcal{V}$ is called a highest-weight module, denoted by $\mathcal{V}(\Lambda)$, and the vector $\boldsymbol{v}_\Lambda \in \mathcal{V}$ a highest-weight vector. From now on, $\mathcal{H}$ is the distinguished Cartan subalgebra of $\mathcal{G}$ with basis of generators $(H_1, \ldots, H_r)$ where $r = \mathrm{rank}\,\mathcal{G}$ and $H_s$ denotes the Cartan generator associated to the odd simple root. The Kac–Dynkin labels are defined by

$$a_i = 2\frac{(\Lambda, \alpha_i)}{(\alpha_i, \alpha_i)} \quad \text{for} \quad i \neq s \quad \text{and} \quad a_s = (\Lambda, \alpha_s)$$

A weight $\Lambda \in \mathcal{H}^*$ is called a dominant weight if $a_i \geq 0$ for all $i \neq s$, integral if $a_i \in \mathbb{Z}$ for all $i \neq s$, and integral dominant if $a_i \in \mathbb{Z}_{\geq 0}$ for all $i \neq s$. A necessary condition for the highest-weight representation of $\mathcal{G}$ with highest weight $\Lambda$ to be finite dimensional is that $\Lambda$ be an integral dominant weight.

One then defines the Kac module. Consider $\mathcal{G} = \oplus_{i \in \mathbb{Z}} \mathcal{G}_i$ the distinguished $\mathbb{Z}$-gradation of $\mathcal{G}$ and let $\mathcal{K} = \mathcal{G}_0 \oplus \mathcal{N}^+$, where $\mathcal{N}^+ = \oplus_{i>0} \mathcal{G}_i$, be a subalgebra of $\mathcal{G}$. Denote by $\mathcal{U}(\mathcal{G})$ and $\mathcal{U}(\mathcal{K})$ the corresponding universal enveloping superalgebras. Let $\Lambda \in \mathcal{H}^*$ be an integral dominant weight and $\mathcal{V}_0(\Lambda)$ be the $\mathcal{G}_0$-module with highest weight $\Lambda$, which is extended to a $\mathcal{K}$-module by setting

**Table 5** Distinguished Dynkin diagrams of the basic Lie superalgebras

| *Superalgebra $\mathcal{G}$* | *Distinguished Dynkin diagram* |
|---|---|
| $A(m-1, n-1)$ |  |
| $B(m, n)$ |  |
| $B(0, n)$ |  |
| $C(n+1)$ |  |
| $D(m, n)$ |  |
| $F(4)$ |  |
| $G(3)$ |  |
| $D(2, 1; \alpha)$ |  |

$\mathcal{N}^+\mathcal{V}_0(\Lambda) = 0$. From this $\mathcal{K}$-module, it is possible to construct a $\mathcal{G}$-module in the following way. One considers the factor space $\mathcal{U}(\mathcal{G}) \otimes_{\mathcal{U}(\mathcal{K})} \mathcal{V}_0(\Lambda)$ consisting of elements of $\mathcal{U}(\mathcal{G}) \otimes \mathcal{V}_0(\Lambda)$ such that the elements $h \otimes \boldsymbol{v}$ and $1 \otimes h(\boldsymbol{v})$ have been identified for $h \in \mathcal{K}$ and $\boldsymbol{v} \in \mathcal{V}_0(\Lambda)$. This space acquires the structure of a $\mathcal{G}$-module by setting $g(\boldsymbol{u} \otimes \boldsymbol{v}) = g\boldsymbol{u} \otimes \boldsymbol{v}$ for $\boldsymbol{u} \in \mathcal{U}(\mathcal{G}), g \in \mathcal{G}$, and $\boldsymbol{v} \in \mathcal{V}_0(\Lambda)$. This $\mathcal{G}$-module is called the induced module from the $\mathcal{K}$-module $\mathcal{V}_0(\Lambda)$ and denoted by $\mathrm{Ind}_{\mathcal{K}}^{\mathcal{G}}\mathcal{V}_0(\Lambda)$. For example, in the case of type I basic Lie superalgebras, if $\{f_1, \ldots, f_d\}$ denotes a basis of odd generators of $\mathcal{G}/\mathcal{K}$, then

$$\mathrm{Ind}_{\mathcal{K}}^{\mathcal{G}}\mathcal{V}_0(\Lambda) = \bigoplus_{1 \le i_1 < \cdots < i_k \le d} f_{i_1} \ldots f_{i_k}\mathcal{V}_0(\Lambda)$$

The Kac module $\overline{\mathcal{V}}(\Lambda)$ is defined as follows:

1. For a superalgebra $\mathcal{G}$ of type I (the odd part is the direct sum of two irreducible representations of the even part), the Kac module is the induced module
$$\overline{\mathcal{V}}(\Lambda) = \mathrm{Ind}_{\mathcal{K}}^{\mathcal{G}}\ \mathcal{V}_0(\Lambda)$$
2. For a superalgebra $\mathcal{G}$ of type II (the odd part is an irreducible representation of the even part), the induced module $\mathrm{Ind}_{\mathcal{K}}^{\mathcal{G}}\ \mathcal{V}_0(\Lambda)$ contains a submodule $\mathcal{M}(\Lambda) = \mathcal{U}(\mathcal{G})\mathcal{G}_{-\psi}^{b+1}\mathcal{V}_0(\Lambda)$, where $\psi$ is the longest simple root of $\mathcal{G}_{\bar{0}}$ which is hidden behind the odd simple root – that is, the longest simple root of $\mathrm{sp}(2n)$ in the case of $\mathrm{osp}(m\,|\,2n)$ and the simple root of $\mathrm{sl}(2)$ in the case of $F(4), G(3)$, and $D(2, 1; \alpha)$ – and $b = 2(\Lambda, \psi)/(\psi, \psi)$ is the component of $\Lambda$ with respect to $\psi$. The Kac module is defined as the quotient of the induced module $\mathrm{Ind}_{\mathcal{K}}^{\mathcal{G}}\ \mathcal{V}_0(\Lambda)$ by the submodule $\mathcal{M}(\Lambda)$:
$$\overline{\mathcal{V}}(\Lambda) = \mathrm{Ind}_{\mathcal{K}}^{\mathcal{G}}\ \mathcal{V}_0(\Lambda)/\mathcal{U}(\mathcal{G})\mathcal{G}_{-\psi}^{b+1}\mathcal{V}_0(\Lambda)$$

In the case where the Kac module is not simple, it contains a maximal submodule $\mathcal{I}(\Lambda)$ and the quotient module $\mathcal{V}(\Lambda) = \overline{\mathcal{V}}(\Lambda)/\mathcal{I}(\Lambda)$ is a simple module.

The fundamental result concerning the representations of basic Lie superalgebras is the following:

1. Any finite dimensional irreducible representation of $\mathcal{G}$ is of the form $\mathcal{V}(\Lambda) = \overline{\mathcal{V}}(\Lambda)/\mathcal{I}(\Lambda)$, where $\Lambda$ is an integral dominant weight.
2. Any finite-dimensional simple $\mathcal{G}$-module is uniquely characterized by its integral dominant weight $\Lambda$: two $\mathcal{G}$-modules $\mathcal{V}(\Lambda)$ and $\mathcal{V}(\Lambda')$ are isomorphic if and only if $\Lambda = \Lambda'$.
3. The finite-dimensional simple $\mathcal{G}$-module $\mathcal{V}(\Lambda) = \overline{\mathcal{V}}(\Lambda)/\mathcal{I}(\Lambda)$ has the weight decomposition
$$\mathcal{V}(\Lambda) = \bigoplus_{\lambda \le \Lambda} \mathcal{V}_\lambda$$
with
$$\mathcal{V}_\lambda = \{\boldsymbol{v} \in \mathcal{V} | h(\boldsymbol{v}) = \lambda(h)\boldsymbol{v}, h \in \mathcal{H}\}$$

The presence of odd roots will have another important consequence in the representation theory of superalgebras. Indeed, one might find that in certain representations, weight vectors, different from the highest one specifying the representation, are annihilated by all the generators corresponding to positive roots. Such vector have, of course, to be decoupled from the representation. Representations of this kind are called atypical, while the other irreducible representations not suffering this pathology are called typical. For a basic Lie superalgebra $\mathcal{G}$ with root system $\Delta$, one defines $\overline{\Delta_{\bar{0}}} = \{\alpha \in \Delta_{\bar{0}} | \alpha/2 \notin \Delta_{\bar{1}}\}$ and $\overline{\Delta_{\bar{1}}} = \{\alpha \in \Delta_{\bar{1}} | 2\alpha \notin \Delta_{\bar{0}}\}$. Let $\rho_0$ be the half-sum of the roots of $\Delta_{\bar{0}}^+$, $\rho_1$ the half-sum of the roots of $\Delta_{\bar{1}}^+$, and $\rho = \rho_0 - \rho_1$. The representation $\pi$ with highest weight $\Lambda$ is called typical if

$$(\Lambda + \rho, \alpha) \neq 0 \quad \text{for all } \alpha \in \overline{\Delta_{\bar{1}}^+}$$

The highest weight $\Lambda$ is then called typical. If there exists some $\alpha \in \overline{\Delta_{\bar{1}}^+}$ such that $(\Lambda + \rho, \alpha) = 0$,

the representation $\pi$ and the highest weight $\Lambda$ are called atypical. The number of distinct elements $\alpha \in \overline{\Delta}_{\bar{1}}^{+}$ for which $\Lambda$ is atypical is the degree of atypicality of the representation $\pi$. If there exists one and only one $\alpha \in \overline{\Delta}_{\bar{1}}^{+}$ such that $(\Lambda + \rho, \alpha) = 0$, the representation $\pi$ and the highest weight $\Lambda$ are called singly atypical.

The Kac module $\overline{\mathcal{V}}(\Lambda)$ is a simple $\mathcal{G}$-module if and only if the highest weight $\Lambda$ is typical. All the finite-dimensional representations of $B(0,n)$ are typical. All the finite-dimensional representations of $C(n+1)$ are either typical or singly atypical.

The dimension of a typical finite-dimensional representation $\mathcal{V}$ of $\mathcal{G}$ is given by

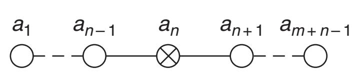
$$\dim \mathcal{V}(\Lambda) = 2^{\dim \Delta_{\bar{1}}^{+}} \prod_{\alpha \in \Delta_{\bar{0}}^{+}} \frac{(\Lambda + \rho, \alpha)}{(\rho_0, \alpha)}$$

where $\dim \mathcal{V}_{\bar{0}}(\Lambda) = \dim \mathcal{V}_{\bar{1}}(\Lambda)$ if $\mathcal{G} \neq B(0,n)$, and if $\mathcal{G} = B(0,n)$,

$$\dim \mathcal{V}_{\bar{0}}(\Lambda) - \dim \mathcal{V}_{\bar{1}}(\Lambda) = \prod_{\alpha \in \overline{\Delta}_{\bar{0}}^{+}} \frac{(\Lambda + \rho, \alpha)}{(\overline{\rho}_0, \alpha)}$$

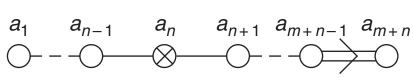
The atypicality conditions are the following:

- For $A(m,n)$ with $\Lambda = (a_1, \ldots, a_{m+n-1})$




$$\sum_{k=i}^{n-1} a_k - \sum_{k=n+1}^{j} a_k + a_n = i + j - 2n$$

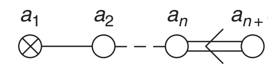
  where $1 \leq i \leq n \leq j \leq m+n-1$.
- $B(m,n)$ with $\Lambda = (a_1, \ldots, a_{m+n})(m \neq 0)$




$$\sum_{q=i}^{n} a_q - \sum_{q=n+1}^{j} a_q = i + j - 2n$$

$$\sum_{q=i}^{n} a_q - \sum_{q=n+1}^{j} a_q - 2 \sum_{q=j+1}^{m+n-1} a_q - a_{m+n} = 2m + i - j - 1 = 0$$

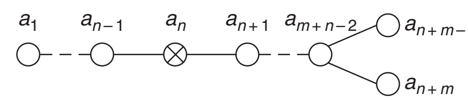
  where $1 \leq i \leq n \leq j \leq m+n-1$.
- $C(n+1)$ with $\Lambda = (a_1, \ldots, a_{n+1})$




$$a_1 - \sum_{q=2}^{i} a_q - i + 1 = 0$$

$$a_1 - \sum_{q=2}^{i} a_q - 2 \sum_{q=i+1}^{n+1} a_q - 2n + i - 1 = 0$$

  where $1 \leq i \leq n$.
- $D(m \mid n)$ with $\Lambda = (a_1, \ldots, a_{m+n})$




$$\sum_{q=i}^{n} a_q - \sum_{q=n+1}^{j} a_q = i + j - 2n$$

  where $1 \leq i \leq n \leq j \leq m+n-1$

$$\sum_{q=i}^{n} a_q - \sum_{q=n+1}^{m+n-2} a_q - a_{m+n} = m - n + i - 1$$

  where $1 \leq i \leq n$

$$\sum_{q=i}^{n} a_q - \sum_{q=n+1}^{j} a_q - 2 \sum_{q=j+1}^{m+n-2} a_q = a_{m+n-1} + a_{m+n} + 2m + i - j - 2$$

  where $1 \leq i \leq n \leq j \leq m+n-2$

*See also:* Lie Groups: General Theory; Lie, Symplectic, and Poisson Groupoids and Their Lie Algebroids.

## Further Reading


Frappat L, Sciarrino A, and Sorba P (2000) *Dictionary on Lie Algebras and Superalgebras*. London: Academic Press.

Kac VG (1977a) Lie superalgebras. *Advances in Mathematics* 26: 8.

Kac VG (1977b) A sketch of Lie superalgebra theory. *Communications in Mathematical Physics* 53: 31.

Kac VG (1978) *Representations of Classical Lie Superalgebras*, Lecture Notes in Mathematics, vol. 676. Berlin: Springer.

# Lie, Symplectic, and Poisson Groupoids and Their Lie Algebroids

**C-M Marle**, Université P.-M. Curie, Paris VI, Paris, France



## Introduction

Groupoids are mathematical structures able to describe symmetry properties more general than those described by groups. They were introduced (and named) by H Brandt in 1926. Around 1950, Charles Ehresmann used groupoids with additional structures (topological and differentiable) as essential tools in topology and differential geometry. In recent years, Mickael Karasev, Alan Weinstein, and Stanisław Zakrzewski independently discovered that symplectic groupoids can be used for the construction of noncommutative deformations of the algebra of smooth functions on a manifold, with potential applications to quantization. Poisson groupoids were introduced by Alan Weinstein as generalizations of both Poisson Lie groups and symplectic groupoids.

We present here the main definitions and first properties relative to groupoids, Lie groupoids, Lie algebroids, symplectic and Poisson groupoids and their Lie algebroids.

## Groupoids

### What is a Groupoid?

Before stating the formal definition of a groupoid, let us explain, in an informal way, why it is a very natural concept. The easiest way to understand that concept is to think of two sets, $\Gamma$ and $\Gamma_0$. The first one, $\Gamma$, is called the "set of arrows" or "total space" of the groupoid, and the other one, $\Gamma_0$, the "set of objects" or "set of units" of the groupoid. One may consider an element $x \in \Gamma$ as an arrow going from an object (a point in $\Gamma_0$) to another object (another point in $\Gamma_0$). The word "arrow" is used here in a very general sense: it means a way for going from a point in $\Gamma_0$ to another in $\Gamma_0$. One should not consider an arrow as a line drawn in the set $\Gamma_0$ joining the starting point of the arrow to its endpoint: this happens only for some special groupoids. Rather, one should think of an arrow as living outside $\Gamma_0$, with only its starting point and its endpoint in $\Gamma_0$, as shown in Figure 1.

The following ingredients enter the definition of a groupoid.

1. Two maps $\alpha : \Gamma \to \Gamma_0$ and $\beta : \Gamma \to \Gamma_0$, called the "target map" and the "source map" of the groupoid. If $x \in \Gamma$ is an arrow, $\alpha(x) \in \Gamma_0$ is its endpoint and $\beta(x) \in \Gamma_0$ its starting point.
2. A "composition law" on the set of arrows; we can compose an arrow $y$ with another arrow $x$, and get an arrow $m(x, y)$, by following first the arrow $y$, then the arrow $x$. Of course, $m(x, y)$ is defined if and only if the target of $y$ is equal to the source of $x$. The source of $m(x, y)$ is equal to the source of $y$, and its target is equal to the target of $x$, as illustrated in Figure 1. It is only by convention that we write $m(x, y)$ rather than $m(y, x)$: the arrow which is followed first is on the right, by analogy with the usual notation $f \circ g$ for the composition of two maps $g$ and $f$. When there is no risk of confusion, we write $x \circ y$, or $x\,.\,y$, or even simply $xy$ for $m(x, y)$. The composition of arrows is associative.
3. An "embedding" $\varepsilon$ of the set $\Gamma_0$ into the set $\Gamma$, which associates a unit arrow $\varepsilon(u)$ with each $u \in \Gamma_0$. That unit arrow is such that both its source and its target are $u$, and it plays the role of a unit when composed with another arrow, either on the right or on the left: for any arrow $x$, $m(\varepsilon(\alpha(x)), x) = x$, and $m(x, \varepsilon(\beta(x))) = x$.
4. Finally, an "inverse map" $\iota$ from the set of arrows onto itself. If $x \in \Gamma$ is an arrow, one may think of $\iota(x)$ as the arrow $x$ followed in the reverse sense. We often write $x^{-1}$ for $\iota(x)$.

Now we are ready to state the formal definition of a groupoid.

**Definition 1** A groupoid is a pair of sets $(\Gamma, \Gamma_0)$ equipped with the structure defined by the following data:

(i) an injective map $\varepsilon : \Gamma_0 \to \Gamma$, called the unit section of the groupoid;

(ii) two maps $\alpha : \Gamma \to \Gamma_0$ and $\beta : \Gamma \to \Gamma_0$, called, respectively, the target map and the source map; they satisfy

$$\alpha \circ \varepsilon = \beta \circ \varepsilon = \mathrm{id}_{\Gamma_0} \qquad [1]$$

(iii) a composition law $m : \Gamma_2 \to \Gamma$, called the product, defined on the subset $\Gamma_2$ of $\Gamma \times \Gamma$, called the set of composable elements,

$$\Gamma_2 = \{(x, y) \in \Gamma \times \Gamma;\ \beta(x) = \alpha(y)\} \qquad [2]$$

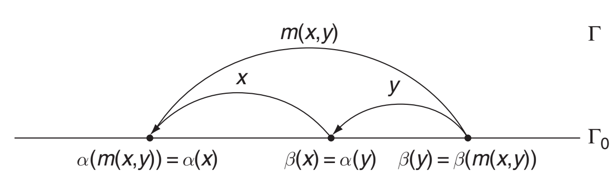



**Figure 1** Two arrows $x$ and $y \in \Gamma$, with the target of $y$, $\alpha(y) \in \Gamma_0$, equal to the source of $x$, $\beta(x) \in \Gamma_0$, and the composed arrow $m(x, y)$.

which is associative, in the sense that whenever one side of the equality

$$m(x, m(y, z)) = m(m(x, y), z) \qquad [3]$$

is defined, the other side is defined too, and the equality holds; moreover, the composition law $m$ is such that for each $x \in \Gamma$,

$$m(\varepsilon(\alpha(x)), x) = m(x, \varepsilon(\beta(x))) = x \qquad [4]$$

(iv) a map $\iota : \Gamma \to \Gamma$, called the inverse, such that, for every $x \in \Gamma$, $(x, \iota(x)) \in \Gamma_2$ and $(\iota(x), x) \in \Gamma_2$, and

$$m(x, \iota(x)) = \varepsilon(\alpha(x)), \quad m(\iota(x), x) = \varepsilon(\beta(x)) \qquad [5]$$

The sets $\Gamma$ and $\Gamma_0$ are called, respectively, the total space and the set of units of the groupoid, which is itself denoted by $\Gamma \overset{\alpha}{\underset{\beta}{\rightrightarrows}} \Gamma_0$.

### Identification and Notations

In what follows, by means of the injective map $\varepsilon$, we will identify the set of units $\Gamma_0$ with the subset $\varepsilon(\Gamma_0)$ of $\Gamma$. Therefore, $\varepsilon$ will be the canonical injection in $\Gamma$ of its subset $\Gamma_0$.

For $x$ and $y \in \Gamma$, we will sometimes write $x\,.\,y$, or even simply $xy$ for $m(x, y)$, and $x^{-1}$ for $\iota(x)$. In addition, we will write "the groupoid $\Gamma$" for "the groupoid $\Gamma \overset{\alpha}{\underset{\beta}{\rightrightarrows}} \Gamma_0$."

### Properties and Comments

The above definitions have the following consequences.

**Involutivity of the inverse map** The inverse map $\iota$ is involutive:

$$\iota \circ \iota = \mathrm{id}_\Gamma \qquad [6]$$

We have indeed, for any $x \in \Gamma$,

$$\begin{aligned}\iota \circ \iota(x) &= m(\iota \circ \iota(x), \beta(\iota \circ \iota(x)))\\ &= m(\iota \circ \iota(x), \beta(x)) = m(\iota \circ \iota(x), m(\iota(x), x))\\ &= m(m(\iota \circ \iota(x), \iota(x)), x) = m(\alpha(x), x) = x\end{aligned}$$

**Unicity of the inverse** Let $x$ and $y \in \Gamma$ be such that

$$m(x, y) = \alpha(x) \quad \text{and} \quad m(y, x) = \beta(x)$$

Then we have

$$\begin{aligned}y &= m(y, \beta(y)) = m(y, \alpha(x))\\ &= m(y, m(x, \iota(x))) = m(m(y, x), \iota(x))\\ &= m(\beta(x), \iota(x)) = m(\alpha(\iota(x)), \iota(x)) = \iota(x)\end{aligned}$$

Therefore for any $x \in \Gamma$, the unique $y \in \Gamma$ such that $m(y, x) = \beta(x)$ and $m(x, y) = \alpha(x)$ is $\iota(x)$.

**The fibers of $\alpha$ and $\beta$ and the isotropy groups** The target map $\alpha$ (resp. the source map $\beta$) of a groupoid $\Gamma \overset{\alpha}{\underset{\beta}{\rightrightarrows}} \Gamma_0$ determines an equivalence relation on $\Gamma$: two elements $x$ and $y \in \Gamma$ are said to be $\alpha$-equivalent (resp. $\beta$-equivalent) if $\alpha(x) = \alpha(y)$ (resp. if $\beta(x) = \beta(y)$). The corresponding equivalence classes are called the $\alpha$-fibers (resp. the $\beta$-fibers) of the groupoid. They are of the form $\alpha^{-1}(u)$ (resp. $\beta^{-1}(u)$), with $u \in \Gamma_0$.

For each unit $u \in \Gamma_0$, the subset

$$\begin{aligned}\Gamma_u &= \alpha^{-1}(u) \cap \beta^{-1}(u)\\ &= \{x \in \Gamma;\ \alpha(x) = \beta(x) = u\}\end{aligned} \qquad [7]$$

is called the "isotropy group" of $u$. It is indeed a group, with the restrictions of $m$ and $\iota$ as composition law and inverse map.

**A way to visualize groupoids** We have seen (Figure 1) a way in which groupoids may be visualized, by using arrows for elements in $\Gamma$ and points for elements in $\Gamma_0$. There is another very useful way to visualize groupoids, shown in Figure 2.

The total space $\Gamma$ of the groupoid is represented as a plane, and the set $\Gamma_0$ of units as a straight line in that plane. The $\alpha$-fibers (resp. the $\beta$-fibers) are represented as parallel straight lines, transverse to $\Gamma_0$.

### Examples of Groupoids

**The groupoid of pairs** Let $E$ be a set. The "groupoid of pairs" of elements in $E$ has, as its total space, the product space $E \times E$. The diagonal $\Delta_E = \{(x, x); x \in E\}$ is its set of units, and the target and source maps are

$$\alpha : (x, y) \mapsto (x, x), \qquad \beta : (x, y) \mapsto (y, y)$$

Its composition law $m$ and inverse map $\iota$ are

$$\begin{aligned}m((x, y), (y, z)) &= (x, z)\\ \iota((x, y)) &= (x, y)^{-1} = (y, x)\end{aligned}$$

**Groups** A group $G$ is a groupoid with set of units $\{e\}$, with only one element $e$, the unit element of the

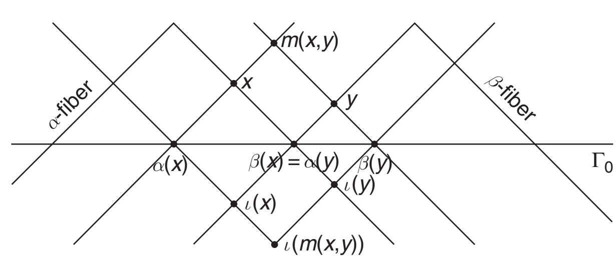



**Figure 2** A way to visualize groupoids.

group. The target and source maps are both equal to the constant map $x \mapsto e$.

**Definition 2** A topological groupoid is a groupoid $\Gamma \overset{\alpha}{\underset{\beta}{\rightrightarrows}} \Gamma_0$ for which $\Gamma$ is a (maybe non-Hausdorff) topological space, $\Gamma_0$ a Hausdorff topological subspace of $\Gamma$, $\alpha$ and $\beta$ surjective continuous maps, $m : \Gamma_2 \to \Gamma$ a continuous map, and $\iota : \Gamma \to \Gamma$ a homeomorphism.

A Lie groupoid is a groupoid $\Gamma \overset{\alpha}{\underset{\beta}{\rightrightarrows}} \Gamma_0$ for which $\Gamma$ is a smooth (maybe non-Hausdorff) manifold, $\Gamma_0$ a smooth Hausdorff submanifold of $\Gamma$, $\alpha$ and $\beta$ smooth surjective submersions (which implies that $\Gamma_2$ is a smooth submanifold of $\Gamma \times \Gamma$), $m : \Gamma_2 \to \Gamma$ a smooth map, and $\iota : \Gamma \to \Gamma$ a smooth diffeomorphism.

### Properties of Lie Groupoids

**Dimensions** Let $\Gamma \overset{\alpha}{\underset{\beta}{\rightrightarrows}} \Gamma_0$ be a Lie groupoid. Since $\alpha$ and $\beta$ are submersions, for any $x \in \Gamma$, the $\alpha$-fiber $\alpha^{-1}(\alpha(x))$ and the $\beta$-fiber $\beta^{-1}(\beta(x))$ are submanifolds of $\Gamma$, both of dimension $\dim \Gamma - \dim \Gamma_0$. The inverse map $\iota$, restricted to the $\alpha$-fiber through $x$ (resp. the $\beta$-fiber through $x$), is a diffeomorphism of that fiber onto the $\beta$-fiber through $\iota(x)$ (resp. the $\alpha$-fiber through $\iota(x)$). The dimension of the submanifold $\Gamma_2$ of composable pairs in $\Gamma \times \Gamma$ is $2 \dim \Gamma - \dim \Gamma_0$.

**The tangent bundle of a Lie groupoid** Let $\Gamma \overset{\alpha}{\underset{\beta}{\rightrightarrows}} \Gamma_0$ be a Lie groupoid. Its tangent bundle $T\Gamma$ is a Lie groupoid, with $T\Gamma_0$ as set of units, $T\alpha : T\Gamma \to T\Gamma_0$ and $T\beta : T\Gamma \to T\Gamma_0$ as target and source maps. Let us denote by $\Gamma_2$ the set of composable pairs in $\Gamma \times \Gamma$, by $m : \Gamma_2 \to \Gamma$ the composition law, and by $\iota : \Gamma \to \Gamma$ the inverse. Then the set of composable pairs in $T\Gamma \times T\Gamma$ is simply $T\Gamma_2$, the composition law on $T\Gamma$ is $Tm : T\Gamma_2 \to T\Gamma$, and the inverse is $T\iota : T\Gamma \to T\Gamma$.

When the groupoid $\Gamma$ is a Lie group $G$, the Lie groupoid $TG$ is a Lie group too. We will see that the cotangent bundle of a Lie groupoid is a Lie groupoid, and more precisely a symplectic groupoid.

**Isotropy groups** For each unit $u \in \Gamma_0$ of a Lie groupoid, the isotropy group $\Gamma_u$ (defined earlier) is a Lie group.

### Examples of Topological and Lie Groupoids

**Topological groups and Lie groups** A topological group (resp. a Lie group) is a topological groupoid (resp. a Lie groupoid) whose set of units has only one element $e$.

**Vector bundles** A smooth vector bundle $\pi : E \to M$ on a smooth manifold $M$ is a Lie groupoid, with the base $M$ as set of units (identified with the image of the zero section); the source and target maps both coincide with the projection $\pi$; the product and the inverse maps are the addition $(x, y) \mapsto x + y$ and the opposite map $x \mapsto -x$ in the fibers.

**The fundamental groupoid of a topological space** Let $M$ be a topological space. A "path" in $M$ is a continuous map $\gamma : [0, 1] \to M$. We denote by $[\gamma]$ the homotopy class of a path $\gamma$ and by $\Pi(M)$ the set of homotopy classes of paths in $M$ (with fixed end-points). For $[\gamma] \in \Pi(M)$, we set $\alpha([\gamma]) = \gamma(1)$, $\beta([\gamma]) = \gamma(0)$, where $\gamma$ is any representative of the class $[\gamma]$. The concatenation of paths determines a well-defined composition law on $\Pi(M)$, for which $\Pi(M) \overset{\alpha}{\underset{\beta}{\rightrightarrows}} M$ is a topological groupoid, called the "fundamental groupoid" of $M$. The inverse map is $[\gamma] \mapsto [\gamma^{-1}]$, where $\gamma$ is any representative of $[\gamma]$ and $\gamma^{-1}$ is the path $t \mapsto \gamma(1 - t)$. The set of units is $M$, if we identify a point in $M$ with the homotopy class of the constant path equal to that point.

When $M$ is a smooth manifold, the same construction can be made with piecewise smooth paths, and the fundamental groupoid $\Pi(M) \overset{\alpha}{\underset{\beta}{\rightrightarrows}} M$ is a Lie groupoid.

## Symplectic and Poisson Groupoids

### Symplectic and Poisson Geometry

Let us recall some definitions and results in symplectic and Poisson geometry, used in the next sections.

**Symplectic manifolds** A "symplectic form" on a smooth manifold $M$ is a differential 2-form $\omega$, which is closed, that is, which satisfies

$$d\omega = 0 \qquad [8]$$

and nondegenerate, that is, such that for each point $x \in M$ and each nonzero vector $v \in T_x M$, there exists a vector $w \in T_x M$ such that $\omega(v, w) \neq 0$. Equipped with the symplectic form $\omega$, a smooth manifold $M$ is called a "symplectic manifold" and denoted by $(M, \omega)$.

The dimension of a symplectic manifold is always even.

**The Liouville form on a cotangent bundle** Let $N$ be a smooth manifold, and $T^*N$ be its cotangent bundle. The Liouville form on $T^*N$ is the 1-form $\theta$ such that, for any $\eta \in T^*N$ and $v \in T_\eta(T^*N)$,

$$\theta(v) = \langle \eta, T\pi_N(v) \rangle \qquad [9]$$

where $\pi_N : T^*N \to N$ is the canonical projection.

The 2-form $\omega = d\theta$ is symplectic, and is called the "canonical symplectic form" on the cotangent bundle $T^*N$.

**Poisson manifolds** A Poisson manifold is a smooth manifold $P$ equipped with a bivector field (i.e., a smooth section of $\wedge^2 TP$) $\Pi$ which satisfies

$$[\Pi, \Pi] = 0 \qquad [10]$$

the bracket on the left-hand side being the Schouten bracket. The bivector field $\Pi$ will be called the Poisson structure on $P$. It allows us to define a composition law on the space $C^\infty(P, \mathbb{R})$ of smooth functions on $P$, called the Poisson bracket and denoted by $(f, g) \mapsto \{f, g\}$, by setting, for all $f$ and $g \in C^\infty(P, \mathbb{R})$ and $x \in P$,

$$\{f, g\}(x) = \Pi(\mathrm{d}f(x), \mathrm{d}g(x)) \qquad [11]$$

That composition law is skew-symmetric and satisfies the Jacobi identity, therefore turns $C^\infty(P, \mathbb{R})$ into a Lie algebra.

**Hamiltonian vector fields** Let $(P, \Pi)$ be a Poisson manifold. We denote by $\Pi^\sharp : T^*P \to TP$ the vector bundle map defined by

$$\langle \eta, \Pi^\sharp(\zeta) \rangle = \Pi(\zeta, \eta) \qquad [12]$$

where $\zeta$ and $\eta$ are two elements in the same fiber of $T^*P$. Let $f : P \to \mathbb{R}$ be a smooth function on $P$. The vector field $X_f = \Pi^\sharp(\mathrm{d}f)$ is called the Hamiltonian vector field associated to $f$. If $g : P \to \mathbb{R}$ is another smooth function on $P$, the Poisson bracket $\{f, g\}$ can be written as

$$\{f, g\} = \langle \mathrm{d}g, \Pi^\sharp(\mathrm{d}f) \rangle = -\langle \mathrm{d}f, \Pi^\sharp(\mathrm{d}g) \rangle \qquad [13]$$

**The canonical Poisson structure on a symplectic manifold** Every symplectic manifold $(M, \omega)$ has a Poisson structure, associated to its symplectic structure, for which the vector bundle map $\Pi^\sharp : T^*M \to M$ is the inverse of the vector bundle isomorphism $v \mapsto -\mathrm{i}(v)\omega$. We will always consider that a symplectic manifold is equipped with that Poisson structure, unless otherwise specified.

**The KKS Poisson structure** Let $\mathcal{G}$ be a finite-dimensional Lie algebra. Its dual space $\mathcal{G}^*$ has a natural Poisson structure, for which the bracket of two smooth functions $f$ and $g$ is

$$\{f, g\}(\xi) = \langle \xi, [\mathrm{d}f(\xi), \mathrm{d}g(\xi)] \rangle \qquad [14]$$

with $\xi \in \mathcal{G}^*$, the differentials $\mathrm{d}f(\xi)$ and $\mathrm{d}g(\xi)$ being considered as elements in $\mathcal{G}$, identified with its bidual $\mathcal{G}^{**}$. It is called the Kirillov, Kostant, and Souriau (KKS) Poisson structure on $\mathcal{G}^*$.

**Poisson maps** Let $(P_1, \Pi_1)$ and $(P_2, \Pi_2)$ be two Poisson manifolds. A smooth map $\varphi : P_1 \to P_2$ is called a Poisson map if, for every pair $(f, g)$ of smooth functions on $P_2$,

$$\{\varphi^* f, \varphi^* g\}_1 = \varphi^*\{f, g\}_2 \qquad [15]$$

**Product Poisson structures** The product $P_1 \times P_2$ of two Poisson manifolds $(P_1, \Pi_1)$ and $(P_2, \Pi_2)$ has a natural Poisson structure: it is the unique Poisson structure for which the bracket of functions of the form $(x_1, x_2) \mapsto f_1(x_1) f_2(x_2)$ and $(x_1, x_2) \mapsto g_1(x_1) g_2(x_2)$ (where $f_1$ and $g_1 \in C^\infty(P_1, \mathbb{R}), f_2$ and $g_2 \in C^\infty(P_2, \mathbb{R})$) is

$$(x_1, x_2) \mapsto \{f_1, g_1\}_1(x_1)\{f_2, g_2\}_2(x_2)$$

The same property holds for the product of any finite number of Poisson manifolds.

**Symplectic orthogonality** Let $(V, \omega)$ be a symplectic vector space, that means a real, finite-dimensional vector space $V$ with a skew-symmetric nondegenerate bilinear form $\omega$. Let $W$ be a vector subspace of $V$. The "symplectic orthogonal" of $W$ is

$$\text{orth}\, W = \{v \in V; \omega(v, w) = 0 \text{ for all } w \in W\} \qquad [16]$$

It is a vector subspace of $V$, which satisfies

$$\dim W + \dim(\text{orth}\, W) = \dim V, \quad \text{orth}(\text{orth}\, W) = W$$

The vector subspace $W$ is said to be isotropic if $W \subset \text{orth}\, W$, coisotropic if $\text{orth}\, W \subset W$, and Lagrangian if $W = \text{orth}\, W$. In any symplectic vector space, there are many Lagrangian subspaces; therefore, the dimension of a symplectic vector space is always even; if $\dim V = 2n$, the dimension of an isotropic (resp. coisotropic, resp. Lagrangian) vector subspace is $\leq n$ (resp. $\geq n$, resp. $= n$).

**Coisotropic and Lagrangian submanifolds** A submanifold $N$ of a Poisson manifold $(P, \Pi)$ is said to be coisotropic if the bracket of two smooth functions, defined on an open subset of $P$ and which vanish on $N$, vanishes on $N$ too. A submanifold $N$ of a symplectic manifold $(M, \omega)$ is coisotropic if and only if for each point $x \in N$, the vector subspace $T_xN$ of the symplectic vector space $(T_xM, \omega(x))$ is coisotropic. Therefore, the dimension of a coisotropic submanifold in a $2n$-dimensional symplectic manifold is $\geq n$; when it is equal to $n$, the submanifold $N$ is said to be Lagrangian.

**Poisson quotients** Let $\varphi : M \to P$ be a surjective submersion of a symplectic manifold $(M, \omega)$ onto a

manifold $P$. The manifold $P$ has a Poisson structure $\Pi$ for which $\varphi$ is a Poisson map if and only if orth(ker $T\varphi$) is integrable. When that condition is satisfied, that Poisson structure on $P$ is unique.

**Poisson Lie groups** A Poisson Lie group is a Lie group $G$ with a Poisson structure $\Pi$, such that the product $(x,y)\mapsto xy$ is a Poisson map from $G\times G$, endowed with the product Poisson structure, into $(G,\Pi)$. The Poisson structure of a Poisson Lie group $(G,\Pi)$ always vanishes at the unit element $e$ of $G$. Therefore, the Poisson structure of a Poisson Lie group never comes from a symplectic structure on that group.

**Definition 3** A symplectic groupoid (resp. a Poisson groupoid) is a Lie groupoid $\Gamma \overset{\alpha}{\underset{\beta}{\rightrightarrows}} \Gamma_0$ with a symplectic form $\omega$ on $\Gamma$ (resp. with a Poisson structure $\Pi$ on $\Gamma$) such that the graph of the composition law $m$

$$\{(x,y,z)\in\Gamma\times\Gamma\times\Gamma;(x,y)\in\Gamma_2 \textit{ and } z=m(x,y)\}$$

is a Lagrangian submanifold (resp. a coisotropic submanifold) of $\Gamma\times\Gamma\times\bar{\Gamma}$ with the product symplectic form (resp. the product Poisson structure), the first two factors $\Gamma$ being endowed with the symplectic form $\omega$ (resp. with the Poisson structure $\Pi$), and the third factor $\bar{\Gamma}$ being $\Gamma$ with the symplectic form $-\omega$ (resp. with the Poisson structure $-\Pi$).

The next theorem states important properties of symplectic and Poisson groupoids.

**Theorem 4** *Let $\Gamma \overset{\alpha}{\underset{\beta}{\rightrightarrows}} \Gamma_0$ be a symplectic groupoid with symplectic 2-form $\omega$ (resp. a Poisson groupoid with Poisson structure $\Pi$). We have the following properties.*

(i) *For a symplectic groupoid, given any point $c\in\Gamma$, each one of the two vector subspaces of the symplectic vector space $(T_c\Gamma,\omega(c))$,*

$$T_c(\beta^{-1}(\beta(c))) \quad \textit{and} \quad T_c(\alpha^{-1}(\alpha(c)))$$

*is the symplectic orthogonal of the other one. For a symplectic or Poisson groupoid, if $f$ is a smooth function whose restriction to each $\alpha$-fiber is constant, and $g$ a smooth function whose restriction to each $\beta$-fiber is constant, then the Poisson bracket $\{f,g\}$ vanishes identically.*

(ii) *The submanifold of units $\Gamma_0$ is a Lagrangian submanifold of the symplectic manifold $(\Gamma,\omega)$ (resp. a coisotropic submanifold of the Poisson manifold $(\Gamma,\Pi)$).*

(iii) *The inverse map $\iota:\Gamma\to\Gamma$ is an antisymplectomorphism of $(\Gamma,\omega)$, that is, it satisfies $\iota^*\omega=-\omega$ (resp. an anti-Poisson diffeomorphism of $(\Gamma,\Pi)$, i.e., it satisfies $\iota_*\Pi=-\Pi$).*

**Corollary 5** *Let $\Gamma \overset{\alpha}{\underset{\beta}{\rightrightarrows}} \Gamma_0$ be a symplectic groupoid with symplectic 2-form $\omega$ (resp. a Poisson groupoid with Poisson structure $\Pi$). There exists on $\Gamma_0$ a unique Poisson structure $\Pi_0$ for which $\alpha:\Gamma\to\Gamma_0$ is a Poisson map, and $\beta:\Gamma\to\Gamma_0$ an anti-Poisson map (i.e., $\beta$ is a Poisson map when $\Gamma_0$ is equipped with the Poisson structure $-\Pi_0$).*

### Examples of Symplectic and Poisson Groupoids

**The cotangent bundle of a Lie groupoid** Let $\Gamma \overset{\alpha}{\underset{\beta}{\rightrightarrows}} \Gamma_0$ be a Lie groupoid.

We have seen above that its tangent bundle $T\Gamma$ has a Lie groupoid structure, determined by that of $\Gamma$. Similarly (but much less obviously), the cotangent bundle $T^*\Gamma$ has a Lie groupoid structure determined by that of $\Gamma$. The set of units is the conormal bundle to the submanifold $\Gamma_0$ of $\Gamma$, denoted by $\mathcal{N}^*\Gamma_0$. We recall that $\mathcal{N}^*\Gamma_0$ is the vector sub-bundle of $T^*_{\Gamma_0}\Gamma$ (the restriction to $\Gamma_0$ of the cotangent bundle $T^*\Gamma$), whose fiber $\mathcal{N}^*_p\Gamma_0$ at a point $p\in\Gamma_0$ is

$$\mathcal{N}^*_p\Gamma_0=\left\{\eta\in T^*_p\Gamma;\langle\eta,v\rangle=0 \text{ for all } v\in T_p\Gamma_0\right\}$$

To define the target and source maps of the Lie algebroid $T^*\Gamma$, we introduce the notion of "bisection" through a point $x\in\Gamma$. A bisection through $x$ is a submanifold $A$ of $\Gamma$, with $x\in A$, transverse both to the $\alpha$-fibers and to the $\beta$-fibers, such that the maps $\alpha$ and $\beta$, when restricted to $A$, are diffeomorphisms of $A$ onto open subsets $\alpha(A)$ and $\beta(A)$ of $\Gamma_0$, respectively. For any point $x\in M$, there exist bisections through $x$. A bisection $A$ allows us to define two smooth diffeomorphisms between open subsets of $\Gamma$, denoted by $L_A$ and $R_A$ and called the left and right translations by $A$, respectively. They are defined by

$$L_A : \alpha^{-1}(\beta(A))\to\alpha^{-1}(\alpha(A))$$
$$L_A(y)=m\left(\beta|_A^{-1}\circ\alpha(y),y\right)$$

and

$$R_A : \beta^{-1}(\alpha(A))\to\beta^{-1}(\beta(A))$$
$$R_A(y)=m\left(y,\alpha|_A^{-1}\circ\beta(y)\right)$$

The definitions of the target and source maps for $T^*\Gamma$ rest on the following properties. Let $x$ be a point in $\Gamma$ and $A$ be a bisection through $x$. The two vector subspaces, $T_{\alpha(x)}\Gamma_0$ and $\ker T_{\alpha(x)}\beta$, are complementary in $T_{\alpha(x)}\Gamma$. For any $v\in T_{\alpha(x)}\Gamma, v-T\beta(v)$ is in $\ker T_{\alpha(x)}\beta$. Moreover, $R_A$ maps the fiber

$\beta^{-1}(\alpha(x))$ into the fiber $\beta^{-1}(\beta(x))$, and its restriction to that fiber does not depend on the choice of $A$; it depends only on $x$. Therefore, $TR_A(v - T\beta(v))$ is in $\ker T_x\beta$ and does not depend on the choice of $A$. We can define the map $\widehat{\alpha}$ by setting, for any $\xi \in T_x^*\Gamma$ and any $v \in T_{\alpha(x)}\Gamma$,

$$\langle \widehat{\alpha}(\xi), v \rangle = \langle \xi, TR_A(v - T\beta(v)) \rangle$$

Similarly, we define $\widehat{\beta}$ by setting, for any $\xi \in T_x^*\Gamma$ and any $w \in T_{\beta(x)}\Gamma$,

$$\left\langle \widehat{\beta}(\xi), w \right\rangle = \langle \xi, TL_A(w - T\alpha(w)) \rangle$$

We see that $\widehat{\alpha}$ and $\widehat{\beta}$ are unambiguously defined, smooth, and take their values in the submanifold $\mathcal{N}^*\Gamma_0$ of $T^*\Gamma$. They satisfy

$$\pi_\Gamma \circ \widehat{\alpha} = \alpha \circ \pi_\Gamma, \qquad \pi_\Gamma \circ \widehat{\beta} = \beta \circ \pi_\Gamma$$

where $\pi_\Gamma : T^*\Gamma \to \Gamma$ is the cotangent bundle projection.

Let us now define the composition law $\widehat{m}$ on $T^*\Gamma$. Let $\xi \in T_x^*\Gamma$ and $\eta \in T_y^*\Gamma$ be such that $\widehat{\beta}(\xi) = \widehat{\alpha}(\eta)$. This implies $\beta(x) = \alpha(y)$. Let $A$ be a bisection through $x$ and $B$ a bisection through $y$. There exist a unique $\xi_{h\alpha} \in T^*_{\alpha(x)}\Gamma_0$ and a unique $\eta_{h\beta} \in T^*_{\beta(y)}\Gamma_0$ such that

$$\xi = \left(L_A^{-1}\right)^* \left(\widehat{\beta}(\xi)\right) + \alpha_x^* \xi_{h\alpha}$$
$$\eta = (R_B^{-1})^* (\widehat{\alpha}(\xi)) + \beta_y^* \eta_{h\beta}$$

Then $\widehat{m}(\xi, \eta)$ is given by

$$\widehat{m}(\xi, \eta) = \alpha_{xy}^* \xi_{h\alpha} + \beta_{xy}^* \eta_{h\beta} + (R_B^{-1})^* \left(L_A^{-1}\right)^* \left(\widehat{\beta}(x)\right)$$

We observe that in the last term of the above expression we can replace $\widehat{\beta}(\xi)$ by $\widehat{\alpha}(\eta)$, since these two expressions are equal, and that $(R_B^{-1})^*(L_A^{-1})^* = (L_A^{-1})^*(R_B^{-1})^*$, since $R_B$ and $L_A$ commute.

Finally, the inverse $\widehat{\iota}$ in $T^*\Gamma$ is $\iota^*$.

With its canonical symplectic form, $T^*\Gamma \overset{\widehat{\alpha}}{\underset{\widehat{\beta}}{\rightrightarrows}} \mathcal{N}^*\Gamma_0$ is a symplectic groupoid. When the Lie groupoid $\Gamma$ is a Lie group $G$, the Lie groupoid $T^*G$ is not a Lie group, contrary to what happens for $TG$. This shows that the introduction of Lie groupoids is not at all artificial: when dealing with Lie groups, Lie groupoids are already with us! The set of units of the Lie groupoid $T^*G$ can be identified with $\mathcal{G}^*$ (the dual of the Lie algebra $\mathcal{G}$ of $G$), identified itself with $T_e^*G$ (the cotangent space to $G$ at the unit element $e$). The target map $\widehat{\alpha} : T^*G \to T_e^*G$ (resp. the source map $\widehat{\beta} : T^*G \to T_e^*G$) associates to each $g \in G$ and $\xi \in T_g^*G$, the value at the unit element $e$ of the right-invariant 1-form (resp. the left-invariant 1-form) whose value at $x$ is $\xi$.

**Poisson Lie groups as Poisson groupoids** Poisson groupoids were introduced by Alan Weinstein as a generalization of both symplectic groupoids and Poisson Lie groups. Indeed, a Poisson Lie group is a Poisson groupoid with a set of units reduced to a single element.

## Lie Algebroids

The notion of a Lie algebroid, due to Jean Pradines, is related to that of a Lie groupoid in the same way as the notion of a Lie algebra is related to that of a Lie group.

**Definition 6** A Lie algebroid over a smooth manifold $M$ is a smooth vector bundle $\pi : A \to M$ with base $M$, equipped with

(i) a composition law $(s_1, s_2) \mapsto \{s_1, s_2\}$ on the space $\Gamma^\infty(\pi)$ of smooth sections of $\pi$, called the bracket, for which that space is a Lie algebra; and
(ii) a vector bundle map $\rho : A \to TM$, over the identity map of $M$, called the anchor map, such that, for all $s_1$ and $s_2 \in \Gamma^\infty(\pi)$ and all $f \in C^\infty(M, \mathbb{R})$,

$$\{s_1, fs_2\} = f\{s_1, s_2\} + ((\rho \circ s_1) \cdot f)s_2 \qquad [17]$$

### Examples

**Lie algebras** A finite-dimensional Lie algebra is a Lie algebroid (with a base reduced to a point and the zero map as anchor map).

**Tangent bundles and their integrable sub-bundles** A tangent bundle $\tau_M : TM \to M$ to a smooth manifold $M$ is a Lie algebroid, with the usual bracket of vector fields on $M$ as composition law, and the identity map as anchor map. More generally, any integrable vector sub-bundle $F$ of a tangent bundle $\tau_M : TM \to M$ is a Lie algebroid, still with the bracket of vector fields on $M$ with values in $F$ as composition law and the canonical injection of $F$ into $TM$ as anchor map.

**The cotangent bundle of a Poisson manifold** Let $(P, \Pi)$ be a Poisson manifold. Its cotangent bundle $\pi_P : T^*P \to P$ has a Lie algebroid structure, with $\Pi^\sharp : T^*P \to TP$ as anchor map. The composition law is the bracket of 1-forms. It will be denoted by $(\eta, \zeta) \mapsto [\eta, \zeta]$ (in order to avoid any confusion with the Poisson bracket of functions). It is given by the formula, in which $\eta$ and $\zeta$ are 1-forms and $X$ a vector field on $P$:

$$\langle [\eta, \zeta], X \rangle = \Pi(\eta, \mathrm{d}\langle \zeta, X \rangle) + \Pi(\mathrm{d}\langle \eta, X \rangle, \zeta) + (\mathcal{L}(X)\Pi)(\eta, \zeta) \qquad [18]$$

We have denoted by $\mathcal{L}(X)\Pi$ the Lie derivative of the Poisson structure $\Pi$ with respect to the vector

field $X$. Another equivalent formula for that composition law is

$$[\zeta,\eta]=\mathcal{L}(\Pi^\sharp\zeta)\eta-\mathcal{L}(\Pi^\sharp\eta)\zeta-\mathrm{d}(\Pi(\zeta,\eta)) \qquad [19]$$

The bracket of 1-forms is related to the Poisson bracket of functions by

$$[\mathrm{d}f,\mathrm{d}g]=\mathrm{d}\{f,g\} \quad \text{for all } f \text{ and } g\in C^\infty(P,\mathbb{R}) \qquad [20]$$

### Properties of Lie Algebroids

Let $\pi : A$ be a Lie algebroid with anchor map $\rho : A\to TM$.

**A Lie algebras homomorphism** For any pair $(s_1,s_2)$ of smooth sections of $\pi$,

$$\rho\circ\{s_1,s_2\}=[\rho\circ s_1,\rho\circ s_2]$$

which means that the map $s\mapsto\rho\circ s$ is a Lie algebra homomorphism from the Lie algebra of smooth sections of $\pi$ into the Lie algebra of smooth vector fields on $M$.

**The generalized Schouten bracket** The composition law $(s_1,s_2)\mapsto\{s_1,s_2\}$ on the space of sections of $\pi$ extends into a composition law on the space of sections of exterior powers of $(A,\pi,M)$, which is called the "generalized Schouten bracket." Its properties are the same as those of the usual Schouten bracket. When the Lie algebroid is a tangent bundle $\tau_M : TM\to M$, that composition law reduces to the usual Schouten bracket. When the Lie algebroid is the cotangent bundle $\pi_P : T^*P\to P$ to a Poisson manifold $(P,\Pi)$, the generalized Schouten bracket is the bracket of forms of all degrees on the Poisson manifold $P$, introduced by J-L Koszul, which extends the bracket of 1-forms used earlier.

**The dual bundle of a Lie algebroid** Let $\varpi : A^*\to M$ be the dual bundle of the Lie algebroid $\pi : A\to M$. There exists on the space of sections of its exterior powers a graded endomorphism $\mathrm{d}_\rho$, of degree 1 (that means that if $\eta$ is a section of $\wedge^k A^*, d_\rho(\eta)$ is a section of $\wedge^{k+1}A^*$). That endomorphism satisfies

$$d_\rho\circ d_\rho=0$$

and its properties are essentially the same as those of the exterior derivative of differential forms. When the Lie algebroid is a tangent bundle $\tau_M : TM\to M, d_\rho$ is the usual exterior derivative of differential forms.

On the spaces of sections of the exterior powers of a Lie algebroid and of its dual bundle we can develop a differential calculus very similar to the usual differential calculus of vector and multivector fields and differential forms on a manifold. Operators such as the interior product, the exterior derivative, and the Lie derivative can still be defined and have properties similar to those of the corresponding operators for vector and multivector fields and differential forms on a manifold.

The total space $A^*$ of the dual bundle of a Lie algebroid $\pi : A\to M$ has a natural Poisson structure: a smooth section $s$ of $\pi$ can be considered as a smooth real-valued function on $A^*$ whose restriction to each fiber $\varpi^{-1}(x)(x\in M)$ is linear; this property allows us to extend the bracket of sections of $\pi$ (defined by the Lie algebroid structure) to obtain a Poisson bracket of functions on $A^*$. When the Lie algebroid $A$ is a finite-dimensional Lie algebra $\mathcal{G}$, the Poisson structure on its dual space $\mathcal{G}^*$ is the KKS Poisson structure discussed earlier.

### The Lie Algebroid of a Lie Groupoid

Let $\Gamma\overset{\alpha}{\underset{\beta}{\rightrightarrows}}\Gamma_0$ be a Lie groupoid. Let $A(\Gamma)$ be the intersection of ker $T\alpha$ and $T_{\Gamma_0}\Gamma$ (the tangent bundle $T\Gamma$ restricted to the submanifold $\Gamma_0$). We see that $A(\Gamma)$ is the total space of a vector bundle $\pi : A(\Gamma)\to\Gamma_0$, with base $\Gamma_0$, the canonical projection $\pi$ being the map which associates a point $u\in\Gamma_0$ to every vector in ker $T_u\alpha$. In this section, we define a composition law on the set of smooth sections of that bundle, and a vector bundle map $\rho : A(\Gamma)\to T\Gamma_0$, for which $\pi : A(\Gamma)\to\Gamma_0$ is a Lie algebroid, called the Lie algebroid of the Lie groupoid $\Gamma\overset{\alpha}{\underset{\beta}{\rightrightarrows}}\Gamma_0$.

We observe first that for any point $u\in\Gamma_0$ and any point $x\in\beta^{-1}(u)$, the map $L_x : y\mapsto L_x y=m(x,y)$ is defined on the $\alpha$-fiber $\alpha^{-1}(u)$, and maps that fiber into the $\alpha$-fiber $\alpha^{-1}(\alpha(x))$. Therefore, $T_uL_x$ maps the vector space $A_u=\ker T_u\alpha$ onto the vector space ker $T_x\alpha$, tangent at $x$ to the $\alpha$-fiber $\alpha^{-1}(\alpha(x))$. Any vector $w\in A_u$ can therefore be extended into the vector field along $\beta^{-1}(u), x\mapsto\widehat{w}(x)=T_uL_x(w)$. More generally, let $w : U\to A(\Gamma)$ be a smooth section of the vector bundle $\pi : A(\Gamma)\to\Gamma_0$, defined on an open subset $U$ of $\Gamma_0$. By using the above-described construction for every point $u\in U$, we can extend the section $w$ into a smooth vector field $\widehat{w}$, defined on the open subset $\beta^{-1}(U)$ of $\Gamma$, by setting, for all $u\in U$ and $x\in\beta^{-1}(u)$:

$$\widehat{w}(x)=T_uL_x(w(u))$$

We have defined an injective map $w\mapsto\widehat{w}$ from the space of smooth local sections of $\pi : A(\Gamma)\to\Gamma_0$, into a subspace of the space of smooth vector fields defined on open subsets of $\Gamma$. The image of that map is the space of smooth vector fields $\widehat{w}$, defined on open subsets $\widehat{U}$ of $\Gamma$ of the form $\widehat{U}=\beta^{-1}(U)$, where

$U$ is an open subset of $\Gamma_0$, which satisfy the two properties:

1. $T\alpha \circ \widehat{w} = 0$,
2. for every $x$ and $y \in \widehat{U}$ such that $\beta(x) = \alpha(y)$, $T_y L_x(\widehat{w}(y)) = \widehat{w}(xy)$.

These vector fields are called "left-invariant vector fields" on $\Gamma$.

The space of left-invariant vector fields on $\Gamma$ is closed under the bracket operation. We can therefore define a composition law $(w_1, w_2) \mapsto \{w_1, w_2\}$ on the space of smooth sections of the bundle $\pi : A(\Gamma) \to \Gamma_0$ by defining $\{w_1, w_2\}$ as the unique section such that

$$\widehat{\{w_1, w_2\}} = [\widehat{w}_1, \widehat{w}_2]$$

Finally, we define the anchor map $\rho$ as the map $T\beta$ restricted to $A(\Gamma)$. With that composition law and that anchor map, the vector bundle $\pi : A(\Gamma) \to \Gamma_0$ is a Lie algebroid, called the Lie algebroid of the Lie groupoid $\Gamma \overset{\alpha}{\underset{\beta}{\rightrightarrows}} \Gamma_0$.

We could exchange the roles of $\alpha$ and $\beta$ and use right-invariant vector fields instead of left-invariant vector fields. The Lie algebroid obtained remains the same, up to an isomorphism.

When the Lie groupoid $\Gamma \overset{\alpha}{\underset{\beta}{\rightrightarrows}}$ is a Lie group, its Lie algebroid is simply its Lie algebra.

### The Lie Algebroid of a Symplectic Groupoid

Let $\Gamma \overset{\alpha}{\underset{\beta}{\rightrightarrows}} \Gamma_0$ be a symplectic groupoid, with symplectic form $\omega$. As we have seen above, its Lie algebroid $\pi : A \to \Gamma_0$ is the vector bundle whose fiber, over each point $u \in \Gamma_0$, is $\ker T_u\alpha$. We define a linear map $\omega_u^\flat : \ker T_u\alpha \to T_u^*\Gamma_0$ by setting, for each $w \in \ker T_u\alpha$ and $v \in T_u\Gamma_0$,

$$\left\langle \omega_u^\flat(w), v \right\rangle = \omega_u(v, w)$$

Since $T_u\Gamma_0$ is Lagrangian and $\ker T_u\alpha$ complementary to $T_u\Gamma_0$ in the symplectic vector space $(T_u\Gamma, \omega(u))$, the map $\omega_u^\flat$ is an isomorphism from $\ker T_u\alpha$ onto $T_u^*\Gamma_0$. By using that isomorphism for each $u \in \Gamma_0$, we obtain a vector bundle isomorphism of the Lie algebroid $\pi : A \to \Gamma_0$ onto the cotangent bundle $\pi_{\Gamma_0} : T^*\Gamma_0 \to \Gamma_0$.

As seen in Corollary 5, the submanifold of units $\Gamma_0$ has a unique Poisson structure $\Pi$ for which $\alpha : \Gamma \to \Gamma_0$ is a Poisson map. Therefore, the cotangent bundle $\pi_{\Gamma_0} : T^*\Gamma_0 \to \Gamma_0$ to the Poisson manifold $(\Gamma_0, \Pi)$ has a Lie algebroid structure, with the bracket of 1-forms as composition law. That structure is the same as the structure obtained as a direct image of the Lie algebroid structure of $\pi : A(\Gamma) \to \Gamma_0$, by the above-defined vector bundle isomorphism of $\pi : A \to \Gamma_0$ onto the cotangent bundle $\pi_{\Gamma_0} : T^*\Gamma_0 \to \Gamma_0$. The Lie algebroid of the symplectic groupoid $\Gamma \overset{\alpha}{\underset{\beta}{\rightrightarrows}} \Gamma_0$ can therefore be identified with the Lie algebroid $\pi_{\Gamma_0} : T^*\Gamma_0 \to \Gamma_0$, with its Lie algebroid structure of cotangent bundle to the Poisson manifold $(\Gamma_0, \Pi)$.

### The Lie Algebroid of a Poisson Groupoid

The Lie algebroid $\pi : A(\Gamma) \to \Gamma_0$ of a Poisson groupoid has an additional structure: its dual bundle $\varpi : A(\Gamma)^* \to \Gamma_0$ also has a Lie algebroid structure, compatible in a certain sense (indicated below) with that of $\pi : A(\Gamma) \to \Gamma_0$.

The compatibility condition between the two Lie algebroid structures on the two vector bundles in duality $\pi : A \to M$ and $\varpi : A^* \to M$ can be written as follows:

$$d_*[X, Y] = \mathcal{L}(X)d_*Y - \mathcal{L}(Y)d_*X \qquad [21]$$

where $X$ and $Y$ are two sections of $\pi$, or, using the generalized Schouten bracket of sections of exterior powers of the Lie algebroid $\pi : A \to M$,

$$d_*[X, Y] = [d_*X, Y] + [X, d_*Y] \qquad [22]$$

In these formulas $d_*$ is the generalized exterior derivative, which acts on the space of sections of exterior powers of the bundle $\pi : A \to M$, considered as the dual bundle of the Lie algebroid $\varpi : A^* \to M$.

These conditions are equivalent to the similar conditions obtained by exchange of the roles of $A$ and $A^*$.

When the Poisson groupoid $\Gamma \overset{\alpha}{\underset{\beta}{\rightrightarrows}} \Gamma_0$ is a symplectic groupoid, we have seen that its Lie algebroid is the cotangent bundle $\pi_{\Gamma_0} : T^*\Gamma_0 \to \Gamma_0$ to the Poisson manifold $\Gamma_0$ (equipped with the Poisson structure for which $\alpha$ is a Poisson map). The dual bundle is the tangent bundle $\tau_{\Gamma_0} : T\Gamma_0 \to \Gamma_0$, with its natural Lie algebroid structure defined earlier.

When the Poisson groupoid is a Poisson Lie group $(G, \Pi)$, its Lie algebroid is its Lie algebra $\mathcal{G}$. Its dual space $\mathcal{G}$ has a Lie algebra structure, compatible with that of $\mathcal{G}$ in the above-defined sense, and the pair $(\mathcal{G}, \mathcal{G}^*)$ is called a Lie bialgebra.

Conversely, if the Lie algebroid of a Lie groupoid is a Lie bialgebroid (i.e., if there exists on the dual vector bundle of that Lie algebroid a compatible structure of Lie algebroid, in the above-defined sense), that Lie groupoid has a Poisson structure for which it is a Poisson groupoid.

### Integration of Lie Algebroids

According to Lie's third theorem, for any given finite-dimensional Lie algebra, there exists a Lie group whose Lie algebra is isomorphic to that Lie algebra. The same property is not true for Lie algebroids and Lie groupoids. The problem of

finding necessary and sufficient conditions under which a given Lie algebroid is isomorphic to the Lie algebroid of a Lie groupoid remained open for more than 30 years, although partial results were obtained. A complete solution of that problem was recently obtained by M Crainic and R L Fernandes. Let us briefly sketch their results.

Let $\pi : A \to M$ be a Lie algebroid and $\rho : A \to TM$ its anchor map. A smooth path $a : I = [0,1] \to A$ is said to be admissible if, for all $t \in I$, $\rho \circ a(t) = (\mathrm{d}/\mathrm{d}t)(\pi \circ a)(t)$. When the Lie algebroid $A$ is the Lie algebroid of a Lie groupoid $\Gamma$, it can be shown that each admissible path in $A$ is, in a natural way, associated to a smooth path in $\Gamma$ starting from a unit and contained in an $\alpha$-fiber. When we do not know whether $A$ is the Lie algebroid of a Lie groupoid or not, the space of admissible paths in $A$ still can be used to define a topological groupoid $\mathcal{G}(A)$ with connected and simply connected $\alpha$-fibers, called the Weinstein groupoid of $A$. When $\mathcal{G}(A)$ is a Lie groupoid, its Lie algebroid is isomorphic to $A$, and when $A$ is the Lie algebroid of a Lie groupoid $\Gamma$, $\mathcal{G}(A)$ is a Lie groupoid and is the unique (up to an isomorphism) Lie groupoid with connected and simply connected $\alpha$-fibers with $A$ as Lie algebroid; moreover, $\mathcal{G}(A)$ is a covering groupoid of an open sub-groupoid of $\Gamma$. Crainic and Fernandes have obtained computable necessary and sufficient conditions under which the topological groupoid $\mathcal{G}(A)$ is a Lie groupoid, that is, necessary and sufficient conditions under which $A$ is the Lie algebroid of a Lie groupoid.

*See also:* Classical *r*-Matrices, Lie Bialgebras, and Poisson Lie Groups; Lie Superalgebras and Their Representations; Lie Groups: General Theory; Nonequilibrium Statistical Mechanics (Stationary): Overview; Poisson Reduction.

## Further Reading


Cannas da Silva A and Weinstein A (1999) *Geometric Models for Noncommutative Algebras*, Berkeley Mathematics Lecture Notes 10. Providence: American Mathematical Society.

Crainic M and Fernandes RL (2003) Integrability of Lie brackets. *Annals of Mathematics* 157: 575–620.

Dazord P and Weinstein A (eds.) (1991) *Symplectic Geometry, Groupoids and Integrable Systems*, Mathematical Sciences Research Institute Publications. New York: Springer.

Karasev M (1987) Analogues of the objects of Lie group theory for nonlinear Poisson brackets. *Mathematics of the USSR. Izvestiya* 28: 497–527.

Libermann P and Marle Ch-M (1987) *Symplectic Geometry and Analytical Mechanics*. Dordrecht: Kluwer.

Mackenzie KCH (1987) *Lie Groupoids and Lie Algebroids in Differential Geometry*, London Mathematical Society Lecture Notes Series 124. Cambridge: Cambridge University Press.

Marsden JE and Ratiu TS (eds.) (2005) *The Breadth of Symplectic and Poisson Geometry*, Festschrift in Honor of Alan Weinstein. Boston: Birkhäuser.

Ortega J-P and Ratiu TS (2004) *Momentum Maps and Hamiltonian Reduction*. Boston: Birkhäuser.

Vaisman I (1994) *Lectures on the Geometry of Poisson Manifolds*. Basel: Birkhäuser.

Weinstein A (1996) *Groupoids: unifying internal and external symmetry, a tour through some examples*, Notices of the American Mathematical Society. vol. 43, pp. 744–752. Rhode Island: American Mathematical Society.

Xu P (1995) On Poisson groupoids. *International Journal of Mathematics* 6(1): 101–124.

Zakrzewski S (1990) Quantum and classical pseudogroups, I and II. *Communications in Mathematical Physics* 134: 347–370 and 371–395.


# Liquid Crystals

**O D Lavrentovich**, Kent State University, Kent, OH, USA



Liquid crystals represent an important state of matter, intermediate between regular solids with long-range positional order of atoms or molecules (often accompanied by the orientational order, as in the case of molecular crystals) and isotropic fluids with neither positional nor orientational long-range order. The basic feature of liquid crystals is orientational order of building units, which might be individual molecules or their aggregates, and complete or partial absence of the long-range positional order. Molecular interactions responsible for orientation order in liquid crystals are relatively weak (most liquid crystals melt into the isotropic phase at around 100–150 °C). As a result, the structural organization of liquid crystals, most importantly, the direction of molecular orientation, is very sensitive to the external factors, such as electromagnetic field and boundary conditions. This sensitivity opened the doors for applications of liquid crystals, including in information displays and flat-panel TVs.

Liquid crystals, discovered more than 100 years ago, represent nowadays one of the best studied classes of soft matter, along with colloids, polymer solutions and melts, gels and foams. There is an extensive literature on physical phenomena in liquid crystals, their chemical structure and material parameters, display applications, etc.

## Thermotropic and Lyotropic Systems

Depending on the way the liquid crystalline state (also known as "mesophase") is produced, one distinguishes thermotropic and lyotropic liquid crystals. Thermotropic liquid crystalline state can exist in a certain temperature range for the materials made of strongly anisometric molecules, either elongated (calamitic molecules) or disk-like (discotic molecules). Upon heating, many substances of this type yield the following phase sequence: solid crystal–liquid crystal–isotropic fluid.

Lyotropic liquid crystals form only in the presence of a solvent, such as water or oil. Most commonly, lyotropic mesophases are formed by solutions of anisometric amphiphilic molecules (such as soaps, phospholipids, and surfactants). Amphiphilic molecules have two distinct parts: a (polar) hydrophilic head and a (nonpolar) hydrophobic tail (generally, an aliphatic chain). This feature gives rise to a special "self-organization" of amphiphilic molecules in solvents. Mesomorphic states also might be formed in the solutions of certain polymers; polymers might also form thermotropic (solvent-free) liquid crystals.

There are four basic types of liquid crystalline phases, classified according to the dimensionality of the translational correlations of building units: nematic (no translational correlations), smectic (1D correlations), columnar (2D correlations), and various 3D-correlated structures, such as cubic phases and blue phases.

"Uniaxial nematic," noted UN, is an optically uniaxial fluid phase. The unit vector along the optic axis is called the director $\boldsymbol{n}$, $\boldsymbol{n}^2 = 1$; it indicates the average orientation of the molecular axes (see Figure 1). Even when the molecules are polar, head-to-head overlapping and flip-flops establish centrosymmetric arrangement in the nematic bulk. Thus, $\boldsymbol{n}$ and $-\boldsymbol{n}$ are equivalent notations. It is important to realize that $\boldsymbol{n}$ specifies only the direction of orientation but not the degree of orientational order. In biaxial nematics (BN), the symmetry point group is one of a prism. A BN phase is characterized by three directors, $\boldsymbol{n}$, $\boldsymbol{l}$, and $\boldsymbol{m} = \boldsymbol{n} \times \boldsymbol{l}$, such that $\boldsymbol{n} \equiv -\boldsymbol{n}, \boldsymbol{l} \equiv -\boldsymbol{l}$, and $\boldsymbol{m} \equiv -\boldsymbol{m}$.

When the building unit (molecule or aggregate) is chiral, that is, not equal to its mirror image, UN might show a helicoidal structure. It is then called a cholesteric phase denoted Ch or N*. Note that UN, BN, and N* phases are liquid phases (no long-range correlations in molecular positions).

"Smectics" are layered phases with a quasi-long-range 1D translational order of centers of molecules in a direction normal to the layers (see Figure 2). This positional order is not exactly the long-range order as in regular 3D crystals: as shown by Landau and Peierls, the fluctuative displacements of layers in 1D lattice diverge logarithmically with the size of the sample. However, for regular materials with smectic period of the order of 1 nm, the effect is noticeable only on scales of 1 mm and larger. In smectic A (SmA), the molecules within the layers show fluid-like arrangement, with no long-range in-plane positional order; it is a uniaxial medium with the optic axis $\boldsymbol{n}$ perpendicular to the layers (see Figure 2). Some materials, such as octylcyanobiphenyl (see Figure 1b), show both UN and SmA phase (at somewhat lower temperatures). In the lyotropic version of SmA, the so-called lamellar $L_\alpha$ phase, the amphiphilic molecules arrange into bilayers. If the solvent is water, the exterior surfaces of the bilayer are formed by polar heads; the hydrophobic tails are

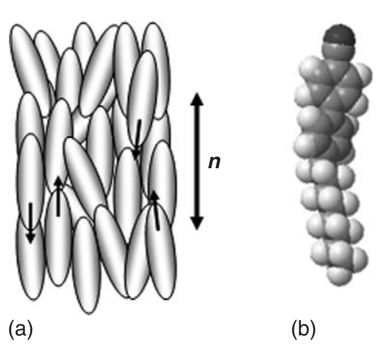



**Figure 1** (a) Nematic (uniaxial) type of ordering in thermotropic liquid crystals; the molecular long axes are on average aligned along the director $\boldsymbol{n}$; (b) a molecule of octylcyanobiphenyl, a typical thermotropic liquid crystalline material capable of both nematic and SmA types of ordering.

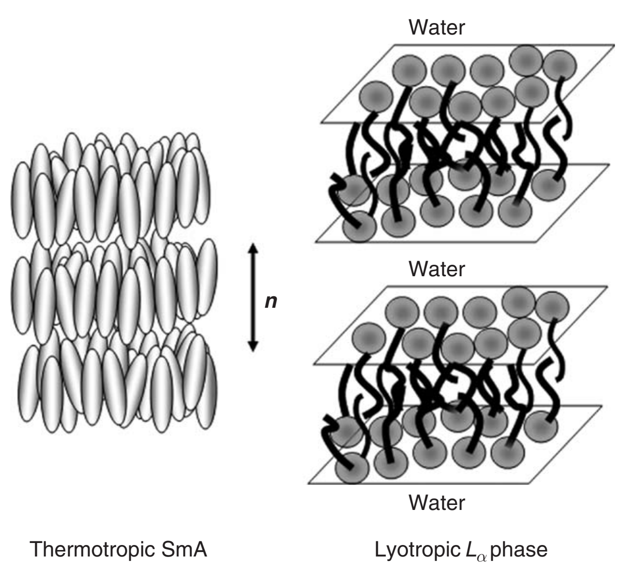



**Figure 2** SmA type of ordering in the thermotropic SmA liquid crystal (left) and the lyotropic analog, $L_\alpha$ phase (right) formed by equidistant arrangement of amphiphilic bilayers in water.

hidden in the middle of the bilayer (note that membranes of many biological cells are organized in the similar way). The periodic structure of alternating surfactant and water layers gives rise to the $L_\alpha$ phase (see Figure 2). Interestingly, the structure might retain its smectic ordering even when strongly diluted, being stabilized by thermal fluctuations of bilayers.

Other types of smectics show in-plane order, caused, for example, by a collective tilt of the rod-like molecules with respect to the normals to the layers (the so-called SmC). In chiral materials, the tilt of the molecules might lead to the helicoidal structure; we do not consider them here, although the chiral SmC phase is of considerable interest for applications in fast-switching optical devices.

"Columnar phases" are most frequently formed by hexagonal packing of cylindrical aggregates, as in the case of thermotropic materials formed by disc-like molecules. The positional order is 2D only, as the intermolecular distances along the axes of the aggregates are not regular.

"3D-correlated structures" demonstrate a periodic structure along all three coordinates, but they are still different from the 3D crystals, as the periodicity is caused by the repetition of molecular orientations rather than by regular repetition of the molecular centers of mass. For example, in cubic lyotropic phases, the 3D network is formed by periodically curved layers of amphiphilic molecules; the molecules are free to move within the layers.

## Order Parameter

The concept of an order parameter (OP) has emerged in its modern form in the Landau model of phase transitions and has been later expanded to describe other features such as topologically stable defects in the ordered media. The OP of the liquid crystal can be related to the anisotropy of macroscopic properties such as diamagnetic or dielectric susceptibility. Measuring these anisotropies allows one to determine the degree of orientational order. The magnetic measurements are especially convenient compared with their electric counterparts, as in this case the local field acting on the molecules differs very little from the external field. In UN, the components of the (symmetric) magnetic susceptibility tensor $\overline{\overline{\chi}}$ read in the frame in which the $z$-axis is parallel to the director $\boldsymbol{n}$, as

$$\overline{\overline{\chi}} = \begin{pmatrix} \chi_\perp & 0 & 0 \\ 0 & \chi_\perp & 0 \\ 0 & 0 & \chi_\parallel \end{pmatrix} \qquad [1]$$

The quantity $\chi_a = \chi_\parallel - \chi_\perp$ is called the anisotropy of the magnetic susceptibility. In most thermotropic UNs, $\chi_\parallel < 0$ and $\chi_\perp < 0$ (diamagnetism), and $\chi_a > 0$, so that $\boldsymbol{n}$ orients along the applied magnetic field. In the isotropic phase, $\chi_a = 0$; in UN, $\chi_a$ is determined by (1) molecular susceptibilities of individual molecules and (2) degree of molecular order. For the latter, one can chose the temperature-dependent quantity $s(T) = (1/2)\langle 3\cos^2\theta - 1\rangle$, where $\theta$ is the angle between the axis of an individual molecule and the director $\boldsymbol{n}$ and $\langle \ldots \rangle$ means an average over molecular orientations. The OP is thus the traceless symmetric tensor $\overline{\overline{Q}}$ with the components that vanish in the isotropic phase, and are proportional to $\chi_a$ in the UN phase:

$$\overline{\overline{Q}} = Q\begin{pmatrix} -\chi_a/3 & 0 & 0 \\ 0 & -\chi_a/3 & 0 \\ 0 & 0 & 2\chi_a/3 \end{pmatrix} \qquad [2]$$

One can choose the constant $Q$ in such a way that in an arbitrary coordinate system, where $\chi_{ij} = \chi_\perp \delta_{ij} + \chi_a n_i n_j$,

$$Q_{ij} = s(T)\left(n_i n_j - \tfrac{1}{3}\delta_{ij}\right) \qquad [3]$$

The tensor OP allows one to describe the biaxial nematic phase as well:

$$Q_{ij} = s(T)\left(n_i n_j - \tfrac{1}{3}\delta_{ij}\right) + b(T)\left(l_i l_j - m_i m_j\right) \qquad [4]$$

where $\boldsymbol{n}$, $\boldsymbol{l}$, and $\boldsymbol{m}$ are three orthogonal directors and $b$ is the "biaxiality parameter"; $b = 0$ in UN.

## Elasticity of the Nematic Phase

In real samples of liquid crystals, the average molecular orientation changes from point to point because of the external fields, boundary conditions, presence of foreign particles, etc. The OP becomes spatially nonuniform, $Q_{ij}(\boldsymbol{r})$. In most problems of practical interest, the typical scale of distortions is much larger than the molecular scale; the deformations are weak in the sense that the scalar part of the OP, $s(T)$, remains constant despite the spatial gradients of the director field $\boldsymbol{n}(\boldsymbol{r})$.

The free-energy density associated with the (small) deformations of the UN, classified as splay, twist, and bend of the director (see Figure 3) writes in terms of the director gradients $n_{i,j} = (\partial \boldsymbol{n_i}/\partial \boldsymbol{x_j})$ as

$$f_{FO} = \tfrac{1}{2}K_1(\operatorname{div}\boldsymbol{n})^2 + \tfrac{1}{2}K_2(\boldsymbol{n}\cdot\operatorname{curl}\boldsymbol{n})^2 + \tfrac{1}{2}K_3(\boldsymbol{n}\times\operatorname{curl}\boldsymbol{n})^2 \qquad [5]$$

and is known as the Frank–Oseen energy density with Frank elastic constants of splay ($K_1$), twist ($K_2$), and bend ($K_3$); all three are necessarily positive definite; the

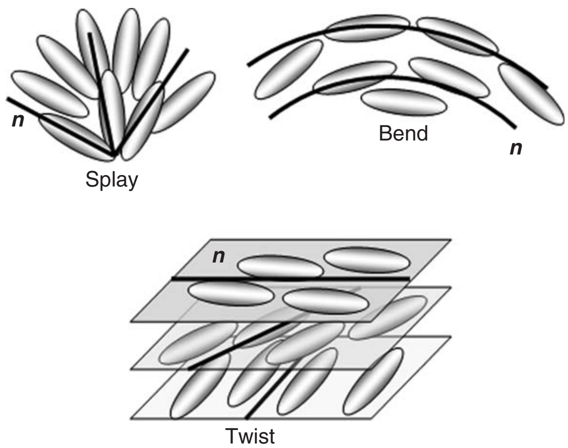



**Figure 3** Basic types of director distortions in the bulk of the uniaxial nematic.

dimensionality is that of a force. The elastic constants can be estimated as the typical energy of molecular interactions responsible for the orientational order divided by the characteristic length (a molecular size): $K \sim U/l \sim k_B T/l \sim 4 \times 10^{-21}\text{J}/10^{-9} \sim 4\,\text{pN}$, which yields a good estimate for many thermotropic UNs, as the experimental values are between 1 and 10 pN. The energy density [5] is often supplemented with the so-called divergence terms:

$$f_{13} + f_{24} = K_{13}\ \text{div}(\boldsymbol{n}\,\text{div}\,\boldsymbol{n}) - K_{24}\ \text{div}(\boldsymbol{n}\,\text{div}\,\boldsymbol{n} + \boldsymbol{n} \times \text{curl}\,\boldsymbol{n}) \qquad [6]$$

The $K_{24}$ term can be re-expressed as a quadratic form of the first derivatives whereas the $K_{13}$ term is proportional to the second derivatives $n_{i,jk}$ and thus might in principle be comparable to $f_{FO} \sim n_{i,j}n_{k,l}$. The volume integrals of these terms can be re-expressed as the surface integrals by virtue of the Gauss theorem (but only when the elastic moduli $K_{13}$ and $K_{24}$ are constant which might not be the case at certain interfaces and at the core of defects). Therefore, when one seeks for equilibrium director configurations by minimizing the total free-energy functional $\int (f_{FO} + f_{13} + f_{24})\mathrm{d}V$, the $K_{13}$ and $K_{24}$ terms do not enter the Euler–Lagrange variational derivative for the bulk. However, they can contribute to the energy and influence the equilibrium director through boundary conditions at the surface. Usually, $K_{24}$ term is retained when the system experiences a topological change of the director field. The $K_{13}$ term is often neglected; very little is known about $K_{13}$ value.

In the presence of external field, the free-energy density acquires additional terms. For example, for the magnetic field $\boldsymbol{B}$, the energy density [5], [6] should be supplemented by the term $-(1/2)\mu_0^{-1}\chi_a(\boldsymbol{B}\cdot\boldsymbol{n})^2$, where $\mu_0 = 4\pi \times 10^{-7}\,\text{Hm}^{-1}$ is the magnetic permeability of free space (magnetic constant).

The possibility to orient the director by an applied electric or magnetic field leads to numerous practical applications. Any actual liquid crystal cell is confined; say, by a pair of parallel glass plates. The molecular interactions between the liquid crystal and the boundary substrates are anisotropic. This anisotropy establishes one (sometimes more) preferred orientation of $\boldsymbol{n}$ at the boundary, the so-called "easy axis." The phenomenon is called the "surface anchoring." Orienting action of the substrates usually keeps the director uniform if the external field is absent. However, the external field can overcome both the "anchoring" at the surfaces and the elasticity of the nematic bulk and reorient the director. This is the "Frederiks effect," first discovered for the magnetic case. When the field is removed, the surface anchoring restores the original director structure. Thus, one can use the external field and surface anchoring to switch the liquid crystal orientation back and forth. The dielectric version of the effect is used in electrooptic devices, including displays. The liquid crystal is usually sandwiched between two transparent electroconductive plates (e.g., glass covered with indium tin oxide) coated with a suitable alignment layer. The voltage across the cell controls the director configuration and thus the optical properties of the cell.

## Elasticity of the Smectic A Phase

For the SmA phase, the elastic free-energy density should be modified to take into account (1) restrictions that the layered structure imposes onto the director twist and bend, and (2) elastic cost of changes in the thickness of the layers:

$$f = \tfrac{1}{2}K_1(\text{div}\ \boldsymbol{n})^2 + \tfrac{1}{2}B\gamma^2 \qquad [7]$$

where $B$ is the Young modulus (layers compressibility modulus) and $\gamma = (d - d_0)/d_0$, the relative difference between the equilibrium period $d_0$ and the actual layer thickness measured along the director $\boldsymbol{n}$. The ratio of $K_1$ to $B$ defines an important length scale

$$\lambda = \sqrt{K_1/B} \qquad [8]$$

called "the penetration length"; $\lambda$ is of the order of the layer separation but diverges when the system approaches the SmA–nematic transition. The splay constant $K_1$ in the SmA phase is of the same order as in a nematic phase stable at higher temperatures. With $\lambda \approx d_0 \approx (1 \div 3)\,\text{nm}$, one finds

$B \sim 10^6 \div 10^7\,\text{N/m}^2$, a value that is $10^3$ to $10^4$ times smaller than the compressibility modulus in a solid.

The SmA elastic free-energy density is often written in terms of the mean curvature $H = (1/2)(\sigma_1 + \sigma_2)$ and the Gaussian curvature $G = \sigma_1\sigma_2$ of the layers:

$$f = \tfrac{1}{2}K_1(\sigma_1 + \sigma_2)^2 + \overline{K}\sigma_1\sigma_2 + \tfrac{1}{2}B\gamma^2 \qquad [9]$$

As compared with eqn [7], it is supplemented by the divergence saddle-splay term $\overline{K}$; $-2K_1 < \overline{K} \leq 0$ (for the system of flat layers to be energetically stable); $\sigma_1 = 1/R_1$ and $\sigma_2 = 1/R_2$ are the local values of the principal curvatures of the smectic layers.

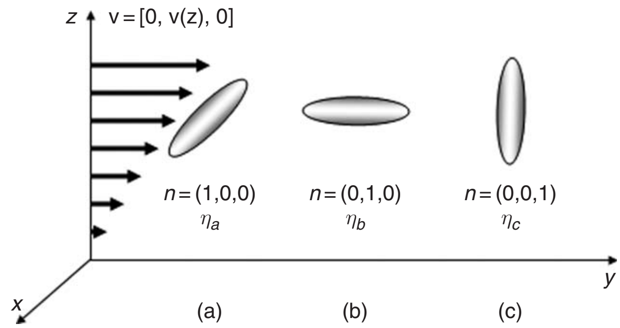



**Figure 4** Miezowicz geometries for effective viscosities of the uniaxial nematic.

## Dynamics

Liquid crystals are fluids; they can flow preserving the orientational order. Flow imposes an orientational torque on the liquid crystals. Most often, the director tends to realign along the direction of flow. There is also an inverse effect: director distortions can cause the flow. This "backflow" effect is of importance in liquid crystal displays. In the approximation of a constant scalar OP, the hydrodynamics of liquid crystals is described in terms of seven unknown variables: (1) mass density $\rho(\boldsymbol{r}, t)$, (2) three components of the velocity field $\boldsymbol{v}(\boldsymbol{r}, t)$, (3) energy density, and (4) two components of the director field $\boldsymbol{n}(\boldsymbol{r}, t)$. These variables are found from seven equations

1. conservation of mass,
2. three equations for the conserved components of the linear momentum,
3. entropy balance equation, and
4. two director dynamics equations.

In contrast to an isotropic fluid, the stress tensor depends not only on the gradients of the velocity, but also on the director components. UN phase should be characterized by five different viscosity constants. The number of viscosities reduces to three, when the director distortions are small. These three can be chosen as the effective viscosities for three idealized geometries of flow, also known as Miezowicz geometries, in which one assumes that the director is fixed (e.g., by a strong magnetic field) (see Figure 4):

When $\boldsymbol{n} = (1, 0, 0)$ is perpendicular to both the flow direction and the velocity gradient, the UN behaves as an isotropic fluid with a viscosity $\eta_a$; however, director fluctuations coupled with the certain values of the viscosity coefficients might destabilize the initial director orientation (see Figure 4a). When $\boldsymbol{n}$ is parallel to the flow (Figure 4b) or parallel to the velocity gradient (Figure 4c), the corresponding viscosities $\eta_b$ and $\eta_c$ are generally different from $\eta_a$ and from each other; $\eta_b < \eta_a < \eta_c$ for a typical thermotropic UN material composed of the rod-like elongated molecules. The result $\eta_b < \eta_c$ can be explained by assuming that the friction correlates with the cross section of the molecules seen by the flow.

## Topological Defects

### Experimental Observations

When a thick UN sample (say, 100 μm thick) with no special aligning layers is viewed under the microscope, one usually observes a number of mobile flexible lines, the so-called disclinations. The disclinations are seen as thin and thick threads (see Figure 5). Thin threads strongly scatter light and show up as sharp lines. These are truly topologically stable defect lines, along which the nematic symmetry of rotation is broken. The disclinations are topologically stable in the sense that no continuous deformation can transform them into a uniform state, $\boldsymbol{n}(\boldsymbol{r}) = \text{const}$. Thin disclinations are singular in the sense that the director is not defined along the core of the defect line. Thick threads are line defects only in appearance; they are not singular disclinations. The director is smoothly curved and well defined everywhere, except, perhaps, at a number of point defects, the so-called hedgehogs (see Figure 5).

In thin UN samples (1–50 μm) with the director tangential to the bounding plates, the disclinations are often perpendicular to the plates. Under a microscope with two crossed polarizers, one can see the ends of the disclinations as centers with emanating pairs of dark brushes (see Figure 6) giving rise to the so-called "Schlieren texture." The dark brushes display the areas where $\boldsymbol{n}$ is either in

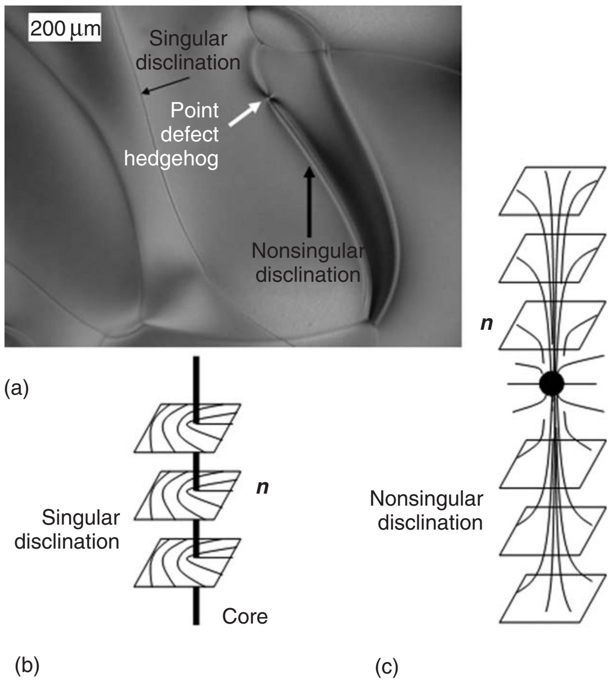



**Figure 5** (a) Thin singular disclinations and thick nonsingular threads in the nematic (n-pentylcyanobiphenyle (5CB)) bulk. Crossed polarizers; (b, c) typical director configurations associated with thin and thick lines; thick lines are often associated with point defects in the nematic bulk – hedgehogs.

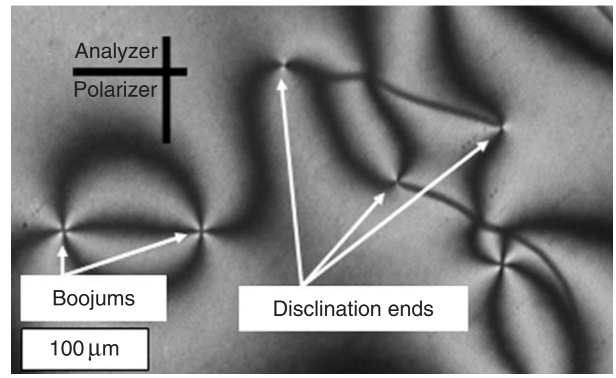



**Figure 6** Schlieren texture of a thin (13 μm) slab of 5CB. Centers with two and four brushes are the ends of singular disclinations and point defects – boojums, respectively. Tangential director orientation. Crossed polarizers.

the plane of polarization of light or in the perpendicular plane. The director rotates by an angle $\pm\pi$ when one goes around the end of the disclination at the surface. Centers with four emanating brushes are also observed; they correspond to point defects located at the surface, the so-called boojums, (see Figure 6). The director undergoes a $\pm 2\pi$ rotation around these four-brush centers. The principal difference between the centers with two brushes (ends of singular lines) and centers with four brushes (surface point defects) can be seen after a gentle shift of one of the bounding plates with respect to the other. Upon shear-induced separation in the plane of observation, the centers with two brushes are clearly seen as connected by a singular trace – disclination, while the centers with four brushes separate without a visible singularity between them.

The intensity of linearly polarized light coming through a uniform UN slab depends on the angle $\beta$ between the polarization direction and the projection of the director $\boldsymbol{n}$ onto the slab's plane:

$$I = I_0 \sin^2 2\beta \sin^2\left[\frac{\pi h}{\lambda}\left(n_{e,\text{eff}} - n_o\right)\right] \qquad [10]$$

where $I_0$ is the intensity of incident light, $\lambda$ is the wavelength of the light, $n_{e,\text{eff}}$ is the effective refractive index that depends on the ordinary index $n_o$, extraordinary index $n_e$, and the director orientation. Equation [10] allows one to relate the number $|k|$ of director rotations by $\pm 2\pi$ around the defect core, to the number $B$ of brushes:

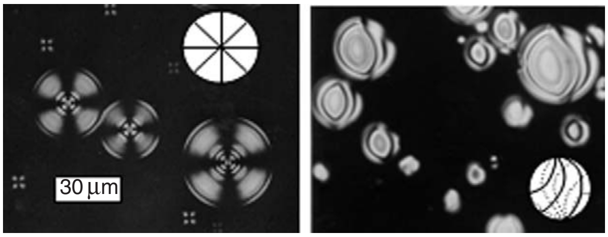
$$|k| = B/4 \qquad [11]$$

Taken with a sign that specifies the direction of rotation, $k$ is called the "strength of disclination," and is related to a more general concept of a topological charge (but does not coincide with it). Note that $I = 0$ when $\boldsymbol{n}$ is perpendicular to the plates (so-called homeotropic state), as $n_{e,\text{eff}} = n_o$. The homeotropic state is used as one of the ground states in modern flat-panel TV sets. By applying the electric field, one tilts the director so that $n_{e,\text{eff}} \neq n_o$ and the cell (or the corresponding pixel in the liquid crystal panel) becomes transparent.

### Nematic Droplets

When left intact, textures with defects in flat samples relax into a more or less uniform state. Disclinations with positive and negative $k$ find each other and annihilate. There are, however, situations when the equilibrium state requires topological defects. Nematic droplets suspended in an isotropic matrix such as glycerin, water, polymer, etc., (see Figure 7) and inverted systems, such as water droplets in a nematic matrix are the most evident examples.

Consider a spherical nematic droplet of a radius $R$ and the balance of the surface anchoring energy $\sim W_a R^2$ ($W_a$ is the surface anchoring coefficient), and the elastic energy $\sim KR; K$ is some averaged Frank constant. Small droplets with $R << K/W_a$ avoid spatial variations of $\boldsymbol{n}$ at the expense of violated boundary conditions. In contrast, large droplets, $R >> K/W_a$, satisfy boundary conditions by aligning $\boldsymbol{n}$ along the




(a) (b)

**Figure 7** Polarizing-microscope texture of spherical nematic droplets suspended in glycerin. (a) The director configuration is radial and normal to the spherical surface; the inset shows the point-defect hedgehog in the center of the droplet. (b) Tangential director orientation at the interface results in the bipolar structure with two defects-boojums at the poles. The director is twisted because of the smallness of the twist elastic constant as compared to the splay and bend constants.

preferred direction(s) at the surface. Since the surface is a sphere, the result is the distorted director in the bulk, for example, a radial hedgehog when the surface orientation is normal (see Figure 7). The characteristic radius $R$ is macroscopic (microns), as $K \sim 10\,\mathrm{pN}$ and $W_a \sim 10^{-5}$–$10^{-6}\,\mathrm{J\,m^{-2}}$. Point defects in large nematic droplets must satisfy restrictions on their topological characteristics that have their roots in the Poincaré and Gauss theorems of differential geometry.

## Topological Classification of Defects in UN

The language of topology, or, more precisely, of homotopy theory, allows one to associate the character of ordering of a medium and the types of defects arising in it, to find the laws of decay, merger and crossing of defects, to trace out their behavior during phase transitions, etc. The key point is occupied by the concept "of topological invariant," also called a "topological charge," which is inherent in every defect. The stability of the defect is guaranteed by the conservation of its charge. Homotopy classification of defects includes three steps.

First, one defines the OP of the system. In a nonuniform state, the OP is a function of coordinates.

Second, one determines the OP (or degeneracy) space $\boldsymbol{R}$, that is, the manifold of all possible values of the OP that do not alter the thermodynamical potentials of the system. In the UN, $\boldsymbol{R}$ is a unit sphere denoted $S^2/Z_2$ (also called the projective plane $RP_2$) with pairs of diametrically opposite points being identical. Every point of $S^2/Z_2$ represents a particular orientation of $\boldsymbol{n}$. Since $\boldsymbol{n} \equiv -\boldsymbol{n}$, any two diametrically opposite points at $S^2/Z_2$ describe the same state.

The function $\boldsymbol{n}(\boldsymbol{r})$ maps the points of the nematic volume into $S^2/Z_2$. The mappings of interest are those of $i$-dimensional "spheres" enclosing defects. A line defect is enclosed by a linear contour, $i = 1$; a point defect is enclosed by a sphere, $i = 2$, etc.

Third, one defines the homotopy groups $\pi_i(\boldsymbol{R})$. The elements of these groups are mappings of $i$-dimensional spheres enclosing the defect in real space into the OP space. To classify the defects of dimensionality $t'$ in a $t$-dimensional medium, one has to know the homotopy group $\pi_i(\boldsymbol{R})$ with $i = t - t' - 1$.

Each element of $\pi_i(\boldsymbol{R})$ corresponds to a class of topologically stable defects; all these defects are equivalent to one another under continuous deformations. The elements of homotopy groups are topological charges of the defects. For UN, the homotopy group $\pi_1(S^2/Z_2) = Z_2 = \{0, 1/2\}$ is composed of two elements; there is thus only one class of topologically stable defects (that appear as thin singular lines under the microscope, see Figure 5) with the addition rules $1/2 + 1/2 = 0$ and $1/2 + 0 = 1/2$ describing interaction of disclinations. The topological point defects in the bulk (hedgehogs) are described by the second homotopy group, $\pi_2(S^2/Z_2) = Z = \{0, 1, 2, \ldots\}$, and can be labeled by integer topological charges. The simplest point defect is a "radial" hedgehog, seen in the center of the radial droplet (see Figure 7a). Boojums are special point defects that, in contrast to hedgehogs, can exist only at the boundary of the medium (see Figure 7b).

The relative stability of stable disclinations depends on the Frank elastic constants of splay ($K_{11}$), twist ($K_{22}$), bend ($K_{33}$) and saddle-splay ($K_{24}$) in the Frank–Oseen elastic free-energy density functional; the role of the elastic constant $K_{13}$ in the structure of defects is not clarified yet.

Consider the simplest case of "planar" disclinations with $\boldsymbol{n}$ perpendicular to the line. In this case, the $K_{24}$-term in the line's energy is zero. Assuming $K_{11} = K_{22} = K_{33} = K$, by minimizing the bulk integral of [5], one finds the equilibrium director configuration around the line of strength $k$

$$\boldsymbol{n} = \{\cos[k\varphi + c], \sin[k\varphi + c], 0\} \qquad [12]$$

where $\varphi = \arctan(y/x)$, $x$ and $y$ are Cartesian coordinates normal to the line, $c$ is a constant. The energy per unit length of a straight planar disclination is

$$F_{1l} = \pi K k^2 \ln\frac{L}{r_c} + F_c \qquad [13]$$

where $L$ is the characteristic size of the system, $r_c$ and $F_c$ are, respectively, the radius and the energy of the disclination core, a region in which the distortions are too strong to be described by a phenomenological theory.

The restriction of planar director distortions does not allow the model to grasp the crucial difference between half-integer and integer $k$'s. The lines of integer $k$, as already discussed, are fundamentally unstable, as the director can be reoriented along the axis. This "escape in the third dimension," is usually energetically favorable, since the singular core is eliminated. When opposite directions of the "escape" meet, a point defect hedgehog is formed, as illustrated in Figure 5c.

Unlike point defects such as vacancies in solids, topological point defects in nematics cause disturbances over the whole volume. The curvature energy of the point defect is proportional to the size $R$ of the system. For example, for the radial hedgehog with $\boldsymbol{n} = (x, y, z)/\sqrt{x^2 + y^2 + z^2}$, and the hyperbolic hedgehog with $\boldsymbol{n} = (-x, -y, z)/\sqrt{x^2 + y^2 + z^2}$, one finds, respectively,

$$F_{\text{rh}} = 8\pi R(K_{11} - K_{24}) + F_{\text{cr}} \quad \text{and}$$
$$F_{\text{hh}} = 8\pi R\left(\frac{K_{11}}{5} + \frac{2K_{33}}{15} + \frac{K_{24}}{3}\right) + F_{\text{ch}} \qquad [14]$$

## Defects in Smectics

Layered structure of smectics leads to linear defects of positional order, dislocations, in addition to disclinations. There is also a special class of distortions known as focal conic domains (FCDs) that are associated with large-scale curvatures of layers. Imagine that because of the boundary conditions, flow, or the external fields, the smectic layers are curved over the scale much larger than the thickness of the layers. It is easy to see from eqn [9] that the curved layers will prefer to maintain their equidistance, as the curvature energy is much smaller than the layers dilation energy at the large scales of deformations. Generally, the family of equidistant curved surfaces is associated with the focal surfaces at which the principal curvatures diverge. These focal surfaces are thus energetically very costly. A radical way to reduce the elastic energy would be to decrease the dimensionality of the focal surfaces, say, by transforming them into lines and points. The latter case corresponds simply to a system of concentric spherical layers. The former is more complicated and corresponds to FCDs in

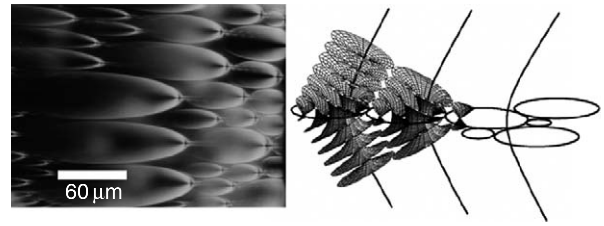



**Figure 8** SmA phase with FCDs based on the confocal pairs of ellipses and hyperbolas; the scheme on the right shows the arrangement of the elliptic bases and smectic layers wrapped around the confocal pairs of defects. Reproduced from Lavrentovich OD (2003) In: Arodz *et al.* (eds.) *Patterns of Symmetry Breaking*. Dordrecht: Kluwer Academic Publishers, with kind permission of Springer Science and Business Media.

which the focal surfaces are represented by pairs of confocal lines: ellipse and hyperbola (limiting case: circle and straight line), and the pair of confocal parabolae. Experiments confirm that the FCDs are the most frequent type of structural deformations in smectic materials see Figure 8.

## Conclusion

To summarize, over the last few decades, liquid crystals transformed from a mysterious and curious form of condensed matter into a key technological material, thanks to the progress in the understanding of their elastic, optical, and viscous properties. However, the intrinsic complexity of these materials still leaves plenty of room for further studies, not only of an applied nature, but also fundamental. In the field of thermotropic liquid crystals, researchers continue to discover new types of structural organization, such as the phases formed by "banana-shaped" molecules that are dramatically different from the phases formed by "regular" rod-like and disk-like molecules. There is a continuous work to sharpen our understanding of even the "old" problems, such as mechanisms of surface alignment, nature and quantitative values of the elastic constants $K_{13}, K_{24}$, and $\bar{K}$. Even in the case of the electric Frederiks effect that is at the heart of modern applications, the search continues as the corresponding process of director reorientation is generally very complex. In addition to the dielectric torque, it is controlled by various factors, for example, a nonlocal character of the electric field in the anisotropic medium, finite electric conductivity, flexoelectric effect (i.e., electric polarization brought about by the director deformations), surface electric polarization at the bounding plates, dependence of the dielectric and other material properties on the frequency of the applied field which might be comparable with the

characteristic frequency of dielectric relaxation, coupling of the director reorientation and the material's flows, appearance of topological defects, etc. Many research efforts nowadays are focused on composite systems, such as liquid crystal colloids and polymer–liquid crystal composites. Over the next decade or so, one would expect that the emphasis in fundamental studies will gradually shift from the thermotropic liquid crystals to their lyotropic counterparts, as the lyotropic type of orientational order is featured by many systems of biological significance, such as solutions of DNA, f-actin, etc.

*See also:* Non-Newtonian Fluids; Topological Defects and Their Homotopy Classification.

## Further Reading


Barbero G and Evangelista LR (2001) In: Ong HC (ed.) *An Elementary Course on the Continuum Theory for Nematic Liquid Crystals, Series on Liquid Crystals*. Singapore: World Scientific.

Blinov LM and Chigrinov VG (1996) *Electrooptic Effects in Liquid Crystal Materials*. New York: Springer.

Chaikin PM and Lubensky TC (1995) *Principles of Condensed Matter Physics*. Cambridge University Press.

Chandrasekhar S (1992) *Liquid Crystals*, 460 pp. Cambridge: Cambridge University Press.

Frank FC (1958) On the theory of liquid crystals. *Transactions of the Faraday Society* 25: 19–28.

de Gennes PG and Prost J (1993) *The Physics of Liquid Crystals*. Oxford: Oxford Science Publications.

Hartshorne NH and Stuart A (1970) *Crystals and the Polarizing Microscope*. New York: American Elsevier.

Kitzerow HS and Bahr C (eds.) (2001) *Chirality in Liquid Crystals*, 502 pp. New York: Springer.

Kleman M and Lavrentovich OD (2003) *Soft Matter Physics: An Introduction*. New York, NY: Springer.

Larsen RG (1999) *The Structure and Rheology of Complex Fluids*, 664 pp. New York: Oxford University Press.

Lavrentovich OD, Pasini P, Zannoni C, and Zumer S (eds.) (2001) *Defects in Liquid Crystals: Computer Simulations, Theory and Experiments*, 344 pp. Dordrecht: Kluwer Academic Publishers.

Sonin AA (1995) *The Surface Physics of Liquid Crystals*, 180 pp. Australia: Gordon and Breach.

Wu ST and Yang DK (2001) *Reflective Liquid Crystal Displays*, 336 pp. Chichester: Wiley.


# Ljusternik–Schnirelman Theory

**J Mawhin**, Université de Louvain, Louvain-la-Neuve, Belgium



## Introduction

Using Lagrange multipliers, the smallest and the largest eigenvalue of a symmetric quadratic form

$$Q(u) = \sum_{j,k=1}^{n} a_{jk} u_j u_k \quad (a_{jk} = a_{kj})$$

can be obtained by minimizing and maximizing $Q$ on the unit sphere $S^{n-1} = \{u \in \mathbb{R}^n : \|u\| = 1\}$. If the corresponding extremum is reached at $u^*$, then $u^*$ is an associated eigenvector.

In the setting of integral or partial differential equations, a "recursive variational method" has been proposed to determine all the eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ and corresponding eigenvectors $u^1, u^2, \ldots, u^n$ of $Q$ or, in modern terms, of the associated symmetric matrix $A = (a_{ij})$:

$$\lambda_1 = \min_{\|u\|=1} Q(u) \quad (= Q(u^1))$$

$$\lambda_j = \min_{\|u\|=1, u \cdot u^1 = 0, \ldots, u \cdot u^{j-1} = 0} Q(u) \quad (= Q(u^j)) \quad (j = 2, \ldots, n)$$

Further considerations have led to a nonrecursive minimum–maximum principle:

$$\lambda_j = \min_{\{X^j \subset \mathbb{R}^n \,:\, \dim X^j = j\}} \; \max_{\{u \in X^j \,:\, \|u\| = 1\}} Q(u) \quad (1 \leq j \leq n)$$

and to a dual maximum–minimum principle (Weyl):

$$\lambda_j = \max_{\{p_1, \ldots, p_{j-1} \in \mathbb{R}^n\}} \; \min_{\{\|u\|=1, u \cdot p_i = 0, 1 \leq i \leq j-1\}} Q(u) \quad (1 \leq j \leq n)$$

These principles have been widely used in various existence and approximation questions of mathematical physics, and extensions have been made to the abstract setting of symmetric bilinear forms in Hilbert spaces.

Around 1930, Ljusternik and Schnirelman have extended this theory beyond the frame of quadratic forms, replacing $Q$ by a differentiable real-valued function $f$ and the unit sphere by a finite-dimensional compact differentiable manifold $M$. Their aim was the obtention of the "critical points" of $f$ on $M$, that is, the points $u \in M$ where the differential $f'(u)$ of $f$ at $u$ (as a linear functional on the tangent space $T_u M$ to $M$) is equal to zero, and of the corresponding critical values, that is, the values of $f$ at critical points. When $M$ is a sphere, the

critical points are nontrivial solutions of the equation

$$f'(u) = \lambda u \qquad [1]$$

for some $\lambda \in \mathbb{R}$ (nonlinear eigenvalue problem). Ljusternik and Schnirelman have replaced the dimension of the vector spaces occurring in the minimum–maximum principle for eigenvalues by the concept of "category" of a closed set $A$ in a topological space $X$. An early success of their approach was the existence of three geometrically distinct closed geodesics without self-intersections on any compact surface of genus zero. In 1960, their theory has been extended to infinite-dimensional manifolds and to other measures of the "size" of a set than the category, allowing many theoretical developments as well as various applications to nonlinear differential equations.

## Ljusternik–Schnirelman Category

Let $X$ be a topological space (e.g., a normed vector space, or a differentiable manifold, or a metric space), and $A$ a closed subset of $X$. The category of $A$ in $X$, $\mathrm{cat}_X(A)$, is the least integer $k$ such that $A$ can be written as $\bigcup_{j=1}^{k} A_j$, with $A_j$ closed and contractible in $X$, that is, continuously deformable in $X$ into a single point. If no such $k$ exists, one sets $\mathrm{cat}_X(A) = +\infty$. We write $\mathrm{cat}(X)$ for $\mathrm{cat}_X(X)$. For example, if $X$ is contractible (in itself), $\mathrm{cat}(X) = 1$. This is the case for any normed space $X$. For the hypersphere, $\mathrm{cat}_{\mathbb{R}^n}(S^{n-1}) = 1$, but $\mathrm{cat}(S^{n-1}) = 2$.

The Ljusternik–Schnirelman category satisfies the following properties, which are not too difficult to prove. If $A, B \subset X$ are closed,

1. $\mathrm{cat}_X(A) = 0$ if and only if $A = \emptyset$;
2. if $A \subset B, \mathrm{cat}_X(A) \leq \mathrm{cat}_X(B)$;
3. $\mathrm{cat}_X(A \cup B) \leq \mathrm{cat}_X(A) + \mathrm{cat}_X(B)$;
4. if $\eta : [0,1] \times X \to X$ is a continuous deformation of $X (\eta(0, A) = A), \mathrm{cat}_X(A) \leq \mathrm{cat}_X(\eta(1, A))$; and
5. if $X$ is a finite-dimensional manifold and $A \subset X$ is compact, there is a neighborhood $B$ of $A$ such that $\mathrm{cat}_X(B) = \mathrm{cat}_X(A)$.

Computing or even estimating the category of a given set is in general difficult, requiring techniques of algebraic topology. In particular, one can show that, for the $n$-torus $\mathbb{T}^n = S^1 \times S^1 \times \cdots \times S^1$ ($n$ times), $\mathrm{cat}(\mathbb{T}^n) = n+1$, and for the $n$-dimensional projective space $\mathbb{P}^n = S^n/\mathbb{Z}^2$, obtained by identifying the antipodal points of $S^n, \mathrm{cat}(\mathbb{P}^n) = n+1$. It is clear that a set of category $p$ must contain at least $p$ points. If $X$ is connected, any compact subset of category $p+1$ has (topological) dimension larger or equal to $p$.

## Ljusternik–Schnirelman Minimax Method

The Ljusternik–Schnirelman category of $M$ provides a lower bound for the number of critical points of a smooth function $f$ on suitable finite-dimensional manifolds $M$. Namely, if $M$ is a compact Riemannian $C^2$-manifold without boundary, any $f \in C^2(M, \mathbb{R})$ has at least $\mathrm{cat}(M)$ distinct critical points, with critical values

$$c_k = \inf_{A \in \mathcal{A}_k} \sup_{u \in A} f(u) \quad (1 \leq k \leq \mathrm{cat}(M)) \qquad [2]$$

where

$$\mathcal{A}_k = \{A \subset M : A \text{ closed}, \mathrm{cat}_M(A) \geq k\} \quad (1 \leq k \leq \mathrm{cat}(M)) \qquad [3]$$

A fundamental technique in the proof is a deformation lemma along the trajectories of the gradient system associated to $f$ (method of steepest descent). If $\nabla f$ denotes the gradient of $f$ in the Riemannian structure of $M$, the Cauchy problem for the gradient system

$$\frac{\mathrm{d}\eta}{\mathrm{d}t} = -\nabla f(\eta), \quad \eta(0) = u \qquad [4]$$

has a unique globally defined continuous solution $\eta(t,u)$, which is such that

$$f(\eta(1,u)) - f(u) = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} f(\eta(t,u))\,\mathrm{d}t = -\int_0^1 \|\nabla f(\eta(t,u))\|^2\,\mathrm{d}t \qquad [5]$$

Notice that, by property (4) of the category, each deformation by $\eta$ of a set in $\mathcal{A}_j$ remains in $\mathcal{A}_j$. For $c \in \mathbb{R}$, define

$$f^c := \{u \in M : f(u) \leq c\} \\ K_c := \{u \in M : \nabla f(u) = 0,\ f(u) = c\} \qquad [6]$$

From [5] it follows that given $c \in \mathbb{R}$ and an open neighborhood $U_c$ of $K_c$, one has $\eta(1, f^{c+\varepsilon} \setminus U_c) \subset f^{c-\varepsilon}$ for all sufficiently small $\varepsilon > 0$. This implies that if $c := c_j = c_{j+1} = \cdots = c_{j+q}$ for some $q \geq 0$, then $\mathrm{cat}_M(K_c) \geq q+1$. Assume, by contradiction, that $\mathrm{cat}_M(K_c) \leq q$, let $U_c$ be an open neighborhood of $K_c$ such that $\mathrm{cat}_M(U_c) = \mathrm{cat}_M(K_c)$ ($U_c = \emptyset$ if $q = 0$), $\varepsilon > 0$ such that $\eta(1, f^{c+\varepsilon} \setminus U_c) \subset f^{c-\varepsilon}$, and $A \in \mathcal{A}_{j+q}$ such that $\sup_A f \leq c + \varepsilon$, that is, $A \subset f^{c+\varepsilon}$. Then

$$\mathrm{cat}_M(\eta(1, A \setminus U_c)) \geq \mathrm{cat}_M(A \setminus U_c) \geq \mathrm{cat}_M(A) - \mathrm{cat}_M(U) \geq j$$

giving the contradiction $c \leq \sup_{\eta(1,A)} f \leq c - \varepsilon$.

Notice that, for each $j, c_j = \inf\{c \in \mathbb{R} : \mathrm{cat}_M(f^c) \geq j\}$, which shows that the $c_j$ are precisely those levels of $f$ where $\mathrm{cat}_M(f^c)$ changes. The presence of critical values is detected by changes in the topology of the

sublevel sets $f^c$ when $c$ varies, a common feature of many techniques for finding critical points of functions.

A direct consequence is that for each even $f \in C^2(\mathbb{R}^n, \mathbb{R})$, system [1] has at least $n$ pairs of solutions $(u, -u)$ with $\|u\| = 1$. Indeed, the solutions of [1] are the critical points of $f$ on $S^{n-1}$. As $f$ takes the same values at antipodal points, it is well defined on the projective space $\mathbb{P}^{n-1}$, and $\mathrm{cat}(\mathbb{P}^{n-1}) = n$.

The Ljusternik–Schnirelman theorem can be extended to the $C^1$-situation. The category of $M$ gives a lower bound for the number of critical points of $f$ on the closed manifold $M$. If $\mathrm{Crit}(M)$ denotes the minimum of the number of critical points of all $C^1$-functions on $M$, so that $\mathrm{Crit}(M) \geq \mathrm{cat}(M)$, an interesting question is to estimate the gap $\mathrm{Crit}(M) - \mathrm{cat}(M)$. For $M$ closed connected, $\mathrm{Crit}(M) \leq \dim(M) + 1$ (Takens). If $\mathrm{Crit}(M) = 2$, $M$ is homeomorphic to a sphere, so that the equality $\mathrm{Crit}(S) = \mathrm{cat}(S)$ for homotopy spheres is equivalent to Poincaré's conjecture! Manifolds with $\mathrm{Crit}(M) = \mathrm{cat}(M) + 1$ are known, but not with $\mathrm{Crit}(M) > \mathrm{cat}(M) + 1$.

## Ljusternik–Schnirelman Theory in Infinite-Dimensional Manifolds

The main difficulty in extending the results of the previous section to functions defined on infinite-dimensional manifolds lies in the lack of compactness. J T Schwartz and Palais have shown that such an extension is possible for functions $f$ satisfying on $M$ a compactness property (allowing an infinite-dimensional deformation lemma), now referred to as the Palais–Smale condition: each sequence $(u_k)$ with $(f(u_k))$ bounded and $\lim_{k\to\infty} \nabla f(u_k) = 0$ has a convergent subsequence. Such a condition can be localized at level $c$ by replacing the boundedness of $(f(u_k))$ by $\lim_{k\to\infty} f(u_k) = c$. The infinite-dimensional extension of Ljusternik–Schnirelman's theorem goes as follows: Let $M$ be an infinite-dimensional Riemannian (or even Finsler) connected complete manifold of class $C^1$ without boundary. Any $f \in C^1(M, \mathbb{R})$ bounded from below and satisfying Palais–Smale condition has at least $\mathrm{cat}(M)$ distinct critical points.

A simple application can be given to the periodic solutions of period $T$ ($T$-periodic solutions) of Lagrangian systems

$$u'' + \nabla V(u) = h(t) \qquad [7]$$

where $V \in C^1(\mathbb{R}^n, \mathbb{R})$, $2\pi$-periodic in each component $u_j (1 \leq j \leq n)$, $h$ is continuous, $T$-periodic and has mean value $\bar{h}$ equal to zero. By the least action principle, the $T$-periodic solutions of [7] are the critical points of the action functional

$$f(u) = \int_0^T \left[ \frac{\|u'(t)\|^2}{2} - V(u(t)) + h(t)u(t) \right] \mathrm{d}t$$

on the Hilbert space $H_T^1$ obtained by completion of the space of $T$-periodic $C^1$ functions for the norm associated with the inner product

$$\langle u, v \rangle := \int_0^T u(t) \cdot v(t)\, \mathrm{d}t + \int_0^T u'(t) \cdot v'(t)\, \mathrm{d}t$$

It follows easily from condition $\bar{h} = 0$ that $f$ is bounded from below and that $f(u + 2\pi e^j) = f(u)$ for all $u \in H_T^1$, with $e^j$ the $j$th unit vector in $\mathbb{R}^n (1 \leq j \leq n)$. Consequently, we can see $f$ as defined on the Riemannian manifold $\mathbb{T}^n \times \widetilde{H_T^1}$, where $\widetilde{H_T^1} = \{u \in H_T^1 : \bar{u} = 0\}$. It is easy to show that $\mathrm{cat}(\mathbb{T}^n \times \widetilde{H_T^1}) = \mathrm{cat}(\mathbb{T}^n) = n + 1$ and that $f$ satisfies Palais–Smale condition on $\mathbb{T}^n \times \widetilde{H_T^1}$. Consequently, system [7] has at least $n + 1$ geometrically distinct $T$-periodic solutions. The same result holds for the more general systems

$$Mu'' + Au + \nabla F(u) = h(t)$$

occurring in the theory of multipoint Josephson junctions or in space discretizations of the sine-Gordon equation. In particular, the classical forced pendulum equation

$$u'' + a \sin u = h(t)$$

has at least two geometrically distinct $T$-periodic solutions when $h$ is $T$-periodic and $\bar{h} = 0$, a result first proved, in a different way, by Mawhin and Willem.

Another way to study nonlinear eigenvalue problems of the form

$$f'(u) = \lambda g'(u)$$

in a Hilbert or a suitable reflexive Banach space $X$ is based upon a Rayleigh–Ritz approximation through a sequence of finite-dimensional problems, where the classical theory is applied. Conditions upon $f, g \in C^1(X, \mathbb{R})$ are given, generalizing Ljusternik–Schnirelman's ones, which ensure the existence of infinitely many solutions. Again, some compactness is needed to justify the limit process, and expressed by some assumptions upon $f$ and $g$ too lengthy to be reproduced here. The following application is exemplary. Let $\Omega \subset \mathbb{R}^N$ be a bounded domain and $X = W_0^{1,p}(\Omega)$, $p > 1$, be the Sobolev space of functions $u : \Omega \to \mathbb{R}$ obtained as the completion of the smooth functions with compact support

in $\Omega$ for the norm $\|u\|_{1,p} = (\int_\Omega \|\nabla u\ \ (x)\|^p\,dx)^{1/p}$. Define the functionals $f$ and $g$ on $W_0^{1,p}(\Omega)$ by

$$f(u) = \int_\Omega \|\nabla u(x)\|^p dx, \quad g(u) = \int_\Omega |u(x)|^p dx$$

The critical points of $f$ on $\{u \in X : g(u) = 1\}$ correspond to the nontrivial solutions of the Dirichlet eigenvalue problem

$$\Delta_p u = \lambda |u|^{p-2} u \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega \qquad [8]$$

for the $p$-Laplacian operator $\Delta_p$ defined by

$$\Delta_p u(x) := \nabla \cdot \left( \|\nabla u(x)\|^{p-2} \nabla u(x) \right)$$

which occurs in the modelization of various problems in a porous medium. An eigenvalue is any $\lambda \in \mathbb{R}$ such that problem [8] has a nontrivial solution. The Ljusternik–Schnirelman technique implies the existence of a sequence of eigenvalues going to infinity, with the usual minimax characterization. When $N = 1$, direct computations show that this sequence gives all eigenvalues, but the problem remains open for $N \geq 2$. The corresponding forced problem

$$\Delta_p u - \lambda |u|^{p-2} u = h(x) \text{ in } \Omega, \quad u = 0 \text{ on } \partial\Omega$$

is always solvable (although not uniquely) when $\lambda$ is not an eigenvalue, but solvability conditions at the higher eigenvalues (Fredholm alternative) remain almost *terra incognita*.

## Index Theories and Critical Points of Symmetric Functionals on a Banach Space

Closely related to the Ljusternik–Schnirelman category is the concept of index associated to the action of a compact topological group $G$ on a normed space $X$, that is, to a continuous map $G \times X \to X, [g, u] \mapsto gu$ such that $1 \cdot u = u, (gh)u = g(hu), \ u \mapsto gu$ is linear. The action is isometric if $\|gu\| = \|u\|$, $A \subset X$ is invariant if $gA = A$ for all $g \in G$, $f : X \to \mathbb{R}$ is invariant if $f \circ g = f$ for all $g \in G$, and $h : X \to X$ is equivariant if $g \circ h = h \circ g$ for each $g \in G$. Let $\text{Fix}\, G = \{u \in X : gu = u \text{ for all } g \in G\}$. The aim of an index is to measure the size of invariant sets. Explicitly, an index theory associates to each closed invariant subset $A$ of $X$ a non-negative (possibly infinite) integer G-ind($A$), its $G$-index, such that

1. G-ind$(A) = 0$ if and only if $A = \emptyset$;
2. if $R : A \to B$ is equivariant and continuous, G-ind$(A) \leq$ G-ind$(B)$;
3. G-ind$(A \cup B) \leq$ G-ind$(A)$ + G-ind$(B)$; and
4. if $A$ is compact, there is a closed invariant neighborhood $U$ of $A$ such that G-ind$(U) =$ G-ind$(A)$.

A first example of index is Krasnosel'skii's genus or $\mathbb{Z}_2$-index which corresponds to the action $0 \cdot u = u, 1 \cdot u = -u$ of $G = \mathbb{Z}_2$. The invariant sets are the ones symmetric with respect to the origin and $\mathbb{Z}_2$-ind$(A)$ is defined by $\mathbb{Z}_2$-ind$(\emptyset) = 0$ and, for $A \neq \emptyset$, as the smallest integer $k$ such that there exists an odd $h \in C(A, \mathbb{R}^k \setminus \{0\})$. A consequence of the Borsuk–Ulam theorem in algebraic topology is that any symmetric bounded neighborhood of the origin in $\mathbb{R}^n$ has $\mathbb{Z}_2$-index equal to $n$. Furthermore, for a compact $A \subset \mathbb{R}^n \setminus \{0\}$ symmetric with respect to the origin, and $\widetilde{A} = A/\mathbb{Z}_2$ ($A$ with antipodal points identified), one has $\mathbb{Z}_2$-ind$(A) = \text{cat}_{\mathbb{R}^n \setminus \{0\}}(\widetilde{A})$.

A second example, the $S^1$-index, is important in the study of periodic solutions of autonomous Hamiltonian systems. $S^1$-ind$(\emptyset) = 0$ and for a nonempty closed invariant $A \subset X$, $S^1$-ind$(A)$ is defined as the smallest integer $k$ such that there exists a positive integer $n$ and $h \in C(A, \mathbb{C}^k \setminus \{0\})$ with $h \circ g = g^n \circ h$ for all $g \in S^1$. A Borsuk–Ulam-type theorem for $S^1$-equivariant mappings implies that if $Z$ is a finite-dimensional invariant subspace of $X$ such that $\text{Fix}\, S^1 \cap Z = \{0\}$ and $D$ is an open bounded invariant neighborhood of 0 in $Z$, then $S^1$-ind$(\partial D) = (1/2)\dim Z$.

As the category of a Banach space $X = 1$, the classical Ljusternik–Schnirelman approach does not provide any information about the multiplicity of the unconstrained critical points of $f \in C^1(X, \mathbb{R})$. If $f$ is invariant under the action on $X$ of a compact group $G$ and satisfies Palais–Smale condition, a Ljusternik–Schnirelman minimax method associated to a G-index provides multiplicity results for unconstrained critical points. Letting

$$\mathcal{A}_j = \{A \subset X : A \text{ is compact, invariant, and G-ind}(A) \geq j\}$$
$$c_j = \inf_{A \in \mathcal{A}_j} \sup_A f \quad (j = 1, 2, \ldots)$$

one shows as in classical Ljusternik–Schnirelman theory that if $c := c_j = c_{j+1} = \cdots = c_{j+q}$ for some $j$ and some $q \geq 0$, then $G$-ind$(K_c) \geq q + 1$. The proof uses an equivariant deformation lemma.

## $\mathbb{Z}^2$- and $S^1$-Invariant Functionals

In the case of the $\mathbb{Z}^2$-action, the following multiplicity result holds for possibly unbounded even $f \in C^1(X, \mathbb{R})$ satisfying the Palais–Smale condition and having the mountain pass geometry: if $Y \cap \{u \in X : f(u) \geq 0\}$ is bounded for each finite-dimensional subspace $Y$ of $X$,

$f(0) = 0$, and $f(u) \geq a > 0$ on $\partial B(r)$, then $f$ has infinitely many couples of critical points. As an application, the semilinear Dirichlet problem

$$\begin{aligned} \Delta u + \lambda u + |u|^{p-1}u = 0 \text{ in } \Omega \\ u = 0 \text{ on } \partial\Omega \end{aligned} \qquad [9]$$

has infinitely many solutions when $\Omega \subset \mathbb{R}^N$ is bounded, $1 < p < (N+2)/(N-2)$, and $\lambda < \lambda_1$, the smallest eigenvalue of $-\Delta$ with Dirichlet boundary conditions. The corresponding energy functional, defined on $W_0^{1,2}(\Omega)$ by

$$f(u) = \int_\Omega \left[ \frac{\|\nabla u(x)\|^2}{2} - \lambda \frac{|u(x)|^2}{2} - \frac{|u(x)|^{p+1}}{p+1} \right] \mathrm{d}x$$

satisfies the Palais–Smale condition. This condition fails in the critical case where $p = (N+2)/(N-2)$, at least at some levels $c$, and this lack of compactness creates both difficulties and interesting phenomena. This situation, which occurs in many important problems of geometry and physics (harmonic maps, Yang–Mills connections, Yamabe problem, equations of constant mean curvature, closed geodesics problems, etc.), reveals indeed, in physical terms, "phase transitions" or "particle creations" at the levels where the Palais–Smale condition fails. In the special case of eqn [9] with $p = (N+2)/(N-2)$, if $N \geq 4$, a positive solution exists when $\lambda \in [0, \lambda_1]$, and, if $N = 3$, the same is true for $\lambda \in [\lambda^*, \lambda_1]$ and some $\lambda^* \in [0, \lambda_1]$, with the optimal value $\lambda^* = \lambda_1/4$ when $\Omega$ is a ball. For $N \geq 4$, [9] has at least cat$(\Omega)$ nontrivial solutions when $\lambda \in [0, \lambda^{**}]$ for some $\lambda^{**} < \lambda_1$. Such a lack of compactness, which can also occur for eqn [9] in $\mathbb{R}^N$ (nonlinear Schrödinger equation), is associated to the invariance of $f$ with respect to the action of some noncompact group, coming, for example, from scale or gauge invariance. P L Lions' concentration–compactness method is useful to analyze those problems.

The following multiplicity theorem holds for an $S^1$-invariant $f \in C^1(X, R)$ satisfying Palais–Smale condition. Let Fix$(S^1) = \{0\}$ and $Z$ be a closed invariant vector subspace of $X$ of positive finite dimension. If $f$ is bounded from below, $f(u) \leq c < 0$ whenever $u \in Z$ and $\|u\| = r$, and $f(0) \geq 0$ for $u \in$ Fix$(S^1) \cap (f')^{-1}(0)$, then $f$ has at least dim $Z/2$ distinct $S^1$-orbits of critical points of $f$ with critical values less or equal to $c$. This abstract theorem provides multiplicity results for the periodic solutions (closed orbits) of autonomous Hamiltonian systems in $\mathbb{R}^{2n}$

$$Ju' + \nabla H(u) = 0 \qquad [10]$$

where $J$ is the symplectic matrix, $H \in C^1(\mathbb{R}^{2n}, \mathbb{R})$, and $c \in \mathbb{R}$ is such $\nabla H(u) \neq 0$ for $u \in H^{-1}(c)$. If $H^{-1}(c)$ bounds a strictly convex compact set $C$ such that $B[r] \subset C \subset B[R]$ for some $0 < r < R < \sqrt{2}r$, then [10] has at least $n$ closed orbits on $H^{-1}(c)$. The problem is reduced to finding the critical points of a suitable dual action functional acting on some space $X$ of $2\pi$-periodic functions having mean value zero. The $S^1$-action on $X$ is defined by time translations $[\tau, u] \mapsto u_\tau = u(\cdot + \tau)$ for all $\gamma = e^{i\tau} \in S^1$. One takes, in the abstract result above, $Z = \{(\cos t)e + (\sin t)Je : e \in \mathbb{R}^{2n}\}$, so that dim $Z = 2n$. The complete proof is quite involved, and, although some improvements of Ekeland–Lasry conditions have been obtained, the problem remains open to know if some pinching condition of the energy surface between spheres or ellipsoids is necessary.

## Some Extensions

When dealing with unbounded functionals, it may be convenient to replace the Ljusternik–Schnirelman category $\mathrm{cat}_X(A)$ by a relative category $\mathrm{cat}_{X,Y}(A)$ with respect to a closed subset $Y$ where, in the covering of $A$ occurring in the classical definition, a set $A_0 \supset Y$ is added, which is continuously deformable in $X$ into a subset of $Y$ in such a way that points of $Y$ remain in $Y$ during the deformation. Clearly $\mathrm{cat}_{X,\emptyset}(A) = \mathrm{cat}_X(A)$. This allows us to prove, under some restrictions on the coefficients and the period, the existence of at least four periodic solutions for the double pendulum with periodic forcing of mean value zero. The classical Ljusternik–Schnirelman category gives at least three periodic solutions without restrictions, and the question of their necessity to obtain four solutions is open.

The relative category also gives a simpler proof of Conley–Zehnder's version of the Arnol'd conjecture (the existence of at least $2n+1$ geometrically distinct 1-periodic solutions for the Hamiltonian system

$$Ju' + \nabla H(t, u) = 0$$

with $H$ 1-periodic in each variable), under minimal regularity assumptions upon $H$. The general conjecture, namely that the minimum number of fixed points of all Hamiltonian symplectomorphisms of a closed symplectic manifold $M$ is larger than the minimum number of critical points of smooth functions $f$ on $M$, remains open.

In another direction, a Ljusternik–Schnirelman theory for functionals defined on closed convex sets of a Banach space has been developed, which is specially well suited for the study of the Plateau problem for minimal surfaces, for surfaces of constant mean curvature, as well as for variational inequalities.

*See also:* Bifurcations of Periodic Orbits; Compact Groups and Their Representations; Floer Homology; Ginzburg–Landau Equation; Inequalities in Sobolev Spaces; Minimal Submanifolds; Minimax Principle in the Calculus of Variations; Saddle Point Problems; Sine-Gordon Equation; Spectral Theory for Linear Operators.

## Further Reading


Ambrosetti A (1992) *Critical Point and Nonlinear Variational Problems*, Cours de la Chaire Lagrange. Paris: Société Mathématique de France.

Cornea O, Lupton G, Oprea J, and Tanré D (2003) *Ljusternik–Schnirelman Category*. Providence: American Mathematical Society.

Courant R and Hilbert D (1953–62) *Methods of Mathematical Physics*, vol. 2. New York: Wiley-Interscience.

Ekeland I (1990) *Convexity Methods in Hamiltonian Mechanics*. Berlin: Springer.

Ekeland I and Ghoussoub N (2002) Selected new aspects of the calculus of variations in the large. *Bulletin of the American Mathematical Society (NS)* 39: 207–265.

Fučik S, Nečas J, Souček J, and Souček V (1973) *Spectral Analysis of Nonlinear Operators*. Berlin: Springer.

Ghoussoub N (1993) *Duality and Pertubation Methods in Critical Point Theory*. Cambridge: Cambridge University Press.

Gould SH (1966) *Variational Methods for Eigenvalue Problems*, 2nd edn. Toronto: Toronto University Press.

Hofer H and Zehnder E (1994) *Symplectic Invariants and Hamiltonian Dynamics*. Basel: Birkhäuser.

Krasnosel'skii MA (1963) *Topological Methods in the Theory of Nonlinear Integral Equations*. Oxford: Pergamon.

Ljusternik LA (1966) *The Topology of the Calculus of Variations in the Large*. Providence: American Mathematical Society.

Ljusternik LA and Schnirelman LG (1930) *Méthodes topologiques dans les problèmes variationnels* (Russian). Moscow: Trudy Scient. Invest. Inst. Math. Mech. (French translation (1934) *Actualités scientifiques et industrielles* no. 188. Paris: Hermann).

Mawhin J and Willem M (1989) *Critical Point Theory and Hamiltonian Systems*. New York: Springer.

Palais RS (1970) Critical point theory and the minimax principle. In: *Proceedings of Symposia in Pure Mathematics*, vol. 15, pp. 115–132. Providence: American Mathematical Society.

Rabinowitz P (1986) *Minimax Methods in Critical Point Theory with Applications to Differential Equations*. Providence: American Mathematical Society.

Schwartz JT (1969) *Nonlinear Functional Analysis*. New York: Gordon and Breach.

Struwe M (1996) *Variational Methods*, 2nd edn. Berlin: Springer.

Willem M (1996) *Minimax Theorems*. Basel: Birkhäuser.

Zeidler E (1985) *Nonlinear Functional Analysis* III. New York: Springer.


# Localization for Quasiperiodic Potentials

**S Jitomirskaya**, University of California at Irvine, Irvine, CA, USA



## Introduction

Discrete Schrödinger operators with quasiperiodic potentials are operators acting on $\ell^2(\mathbb{Z}^d)$ and defined by

$$H_\lambda = \Delta + \lambda V \qquad [1]$$

where $\Delta$ is the lattice tight-binding Laplacian

$$\Delta(n,m) = \begin{cases} 1, & \text{dist}\,(n,m) = 1 \\ 0, & \text{otherwise} \end{cases}$$

and $V(n,m) = V_n\delta(n,m)$ is a potential given by $V_n = f(T_1^{n_1}\cdots T_d^{n_d}\theta), \theta \in \mathbb{T}^b$, where $T_i\theta = \theta + \omega_i$, and $\omega$ is an incommensurate vector. In certain cases $\Delta$ may also be replaced by a long-range Laplacian $L(n,m) = L(n-m)$ with $L(n) \to 0$ sufficiently fast. The questions of interest in the study of quasiperiodic and other ergodic operators are the nature and structure of the spectrum, behavior of the eigenfunctions, and the quantum dynamics: properties of the time evolution $\Psi_t = \mathrm{e}^{itH}\Psi_0$ of an initially localized wave packet $\Psi_0$.

Of particular importance is the phenomenon of Anderson localization which is usually referred to the property of having pure point spectrum with exponentially decaying eigenfunctions. A stronger property of dynamical localization (see the section "Dynamical localization") indicates the insulator behavior, while ballistic transport, which for $d = 1$ follows from the absolutely continuous spectrum, indicates the metallic behavior.

Operators with ergodic potentials always have spectra (and pure point spectra, understood as closures of the set of eigenvalues) constant for a.c. realization of the potential. The individual eigenvalues however depend very sensitively on the phase. Moreover, the pure point spectrum of operators with ergodic potentials never contains isolated eigenvalues, so pure point spectrum in such models is dense in a certain closed set. An easy example of an operator with dense pure point spectrum is $H_\infty$ which is operator [1] with $\lambda^{-1} = 0$, or pure diagonal. It has a complete set of eigenfunctions, characteristic functions of lattice points, with eigenvalues $V_j$. $H_\lambda$ may be viewed as a perturbation of $H_\infty$ for small $\lambda^{-1}$. However, since $V_j$ are dense, small denominators $(V_i - V_j)^{-1}$ make any

perturbation theory difficult, for example, requiring intricate KAM-type schemes.

Various methods developed for the Anderson model (where $V_n$ are i.i.d.r.v.'s) such as Fröhlich–Spencer multiscale analysis and its enhancements, or Aizenman–Molchanov method, do not work for quasiperiodic potentials as, among other reasons, quasiperiodicity does not allow for nice perturbations. The situation here is more difficult and the theory is far less developed than for the random case. With a few exceptions, the results are confined to the one-dimensional setting, and also the case of one frequency ($b=1$) has been much better understood than that of higher frequencies.

One might expect that $H_\lambda$ with $\lambda$ small can be treated as a perturbation of $H_0=\Delta$, and therefore have absolutely continuous spectrum. It is not the case though for random potentials in $d=1$, where Anderson localization holds for all $\lambda$. The same is expected for random potentials in $d=2$ (but not higher). Moreover, in one-dimensional case, there is strong evidence (numerical, analytical, as well as rigorous) that even models with very mild stochasticity in the underlying dynamics (and sufficiently nice sampling functions) have point spectrum for all values of $\lambda$, like in the random case (e.g., $V_n=\lambda f(n^\sigma\alpha+\theta)$, for any $\sigma>1$). At the same time, for quasiperiodic potentials, one can in many cases show absolutely continuous spectrum for $\lambda$ small as well as pure point spectrum for $\lambda$ large (see below), and therefore there is a metal–insulator transition in the coupling constant. It is an interesting question whether quasiperiodic potentials are the only ones with metal–insulator transition in 1D.

## Perturbative and Nonperturbative Approaches

It is probably fair to say that much of the theory of qusiperiodic operators has been first developed around the almost-Mathieu operator, which is

$$H_{\lambda,\omega,\theta}=\Delta+\lambda f(\theta+n\omega) \qquad [2]$$

acting on $\ell^2(\mathbb{Z})$, with $f:\mathbb{T}\to\mathbb{T}; f(\theta)=\cos(2\pi\theta)$. Several KAM-type approaches, starting with the pioneering work of Dinaburg–Sinai in 1975, were developed, in 1980s and 1990s, for this or similar models in both large and small coupling regimes. Of those, the most robust and detailed is the reducibility result of Eliasson (1998) that settled the case of small couplings for sufficiently regular potentials.

The common feature of those perturbative approaches is that, besides all of them being rather intricate multistep procedures, they rely extensively on eigenvalue and eigenfunction parametrization and perturbation arguments.

The common feature of the perturbative results in the quasiperiodic setting is that they typically provide no explicit estimates on how large (or small) the parameter $\lambda$ should be, and, more importantly, $\lambda$ clearly depends on $\omega$ at least through the constants in the Diophantine characterization of $\omega$.

In contrast, the nonperturbative results allow effective (in many cases even optimal) and, most importantly, independent of $\omega$, estimates on $\lambda$. The latter property (uniform in $\omega$ estimates on $\lambda$) has been often taken as a definition of a nonperturbative result.

Recently developed nonperturbative methods are also quite different from the perturbative ones in that they do not employ multiscale schemes: usually only a few (from one to three) sufficiently large scales are involved, do not use the eigenvalue parametrization, and rely instead on direct estimates of the Green's function. They are also significantly less involved, technically. One may think that in these latter respects they resemble the Aizenman–Molchanov method for random localization. It is, however, a superficial similarity, as, on the technical side, they are still closer to and do borrow certain ideas from the multiscale analysis proofs of localization.

## Lyapunov Exponents

Here for simplicity we consider the quasiperiodic case, although the definition of the Lyapunov exponents and some of the mentioned facts apply more generally to the one-dimensional ergodic case.

Let $d=1$. For an energy $E\in\mathbb{R}$ the Lyapunov exponent $\gamma(E)$ is defined as

$$\gamma(E)=\lim_{n\to\infty}\frac{\int_0^1 \ln\|M_k(\theta,E)\|\mathrm{d}\,\theta}{k} \qquad [3]$$

where

$$M_k(\theta,E)=\prod_{n=k-1}^{0}\begin{pmatrix}E-\lambda f(\omega n+\theta) & -1\\ 1 & 0\end{pmatrix}$$

is the $k$-step transfer matrix for the eigenvalue equation $H\Psi=E\Psi$.

In physics literature, positivity of the Lyapunov exponent is often taken as an implicit definition of localization, as Lyapunov exponent is often called the inverse localization length. Thus, we will be interested in the regime when Lyapunov exponents are positive for all energies in a certain interval intersecting the spectrum. If this condition holds for all $E\in\mathbb{R}$, there is no absolutely continuous

component in the spectrum for all $\theta$. Positivity of Lyapunov exponents, however, does not imply localization or exponential decay of eigenfunctions (in particular, neither for the Liouville $\omega$ nor for the resonant $\theta \in \mathbb{T}^b$).

Nonperturbative methods, at least in their original form, stem to a large extent from estimates involving the Lyapunov exponents and exploiting their positivity.

The general theme of the results on positivity of $\gamma(E)$, as suggested by perturbation arguments, is that the Lyapunov exponents are positive for large $\lambda$. This subject has had a rich history. The strongest result in this general context up to date is the following theorem (Bourgain 2003):

**Theorem 1** *Let $f$ be a nonconstant real-analytic function on $\mathbb{T}^b$, and $H$ given by [1]. then, for $\lambda > \lambda(f)$, we have $\gamma(E) > (1/2)\ln\lambda$ for all $E$ and all incommensurate vectors $\omega$.*

### Corollaries of Positive Lyapunov Exponents

**The almost-Mathieu operator** On one hand the almost-Mathieu operator, while simple looking, seems to represent most of the nontrivial properties expected to be encountered in the more general case. On the other hand it has a very special feature: the duality (essentially a Fourier) transform maps $H_\lambda$ to $H_{4/\lambda}$; hence $\lambda = 2$ is the self-dual point. Aubry and Andre in 1980, conjectured that for this model, for irrational $\omega$ a sharp metal–insulator transition in the coupling constant $\lambda$ occurs at the critical value of coupling $\lambda = 2$: the spectrum is pure point for $\lambda > 2$ and purely absolutely continuous for $\lambda < 2$. This conjecture was modified based on later discoveries of singular-continuous spectrum in this context for frequencies or phases with certain arithmetic properties. The modified conjecture stated pure point spectrum for Diophantine $\omega$ and a.e. $\theta$ for $\lambda > 2$ and pure absolutely continuous spectrum for $\lambda < 2$ for all $\omega$,$\theta$. The spectrum at $\lambda = 2$ is singular continuous for all $\omega$ and a.e. $\theta$(this follows from a combination of works by Gordon, Jitomirskaya, Last, Simon Avila, and Krikoryan).

As with the KAM methods, the almost-Mathieu operator was the first model where the positivity of Lyapunov exponents was effectively exploited (Jitomirskaya 1999):

**Theorem 2** *Suppose $\omega$ is Diophantine and $\gamma(E, \omega) > 0$ for all $E \in [E_1, E_2]$. Then the almost-Mathieu operator has Anderson localization in $[E_1, E_2]$ for a.e. $\theta$.*

The condition on $\theta$ can be made explicit (arithmetic) and close to optimal. This, combined with the mentioned results on the Lyapunov exponents, critical value $\lambda = 2$, and duality, gives the following description in the Diophantine case:

**Corollary 3** *The almost-Mathieu operator $H_{\omega,\lambda,\theta}$ has*

1° *for $\lambda > 2$, Diophantine $\omega \in \mathbb{R}$ and almost every $\theta \in \mathbb{R}$, only pure point spectrum with exponentially decaying eigenfunctions.*
2° *for $\lambda = 2$, all $\omega \notin \mathbb{Q}$, and a.e. $\theta \in \mathbb{R}$ purely singular-continuous spectrum.*
3° *for $\lambda < 2$, Diophantine $\omega \in \mathbb{R}$ and a.e. $\theta \in \mathbb{R}$, purely absolutely continuous spectrum.*

Precise arithmetic descriptions of $\omega, \theta$ are available. Thus, the Aubry–Andre conjecture is settled at least for almost all $\omega, \theta$. One should mention, however, that while 1° can be made optimal by existing methods, both 2° and 3° are expected to hold for all $\theta$ and all $\omega \notin \mathbb{Q}$, and such extension remains a challenging problem (see Simon (2000)).

The method in the above work, while so far the only nonperturbative method available allowing precise arithmetic conditions, uses some specific properties of the cosine. It extends to certain other situations, for example, quasiperiodic operators arising from Bloch electrons in a perpendicular magnetic field, where the lattice is triangular or has next-nearest-neighbor interactions. However, it does not extend easily to the multifrequency or even general analytic potentials. A much more robust method was developed by Bourgain–Goldstein (2000), which allowed them to extend (a measure-theoretic version of) the above localization result to the general real analytic as well as the multifrequency case. Note that essentially no results were previously available for the multifrequency case, even perturbative.

**Theorem 4** *Let $f$ be nonconstant real analytic on $\mathbb{T}^b$ and $H$ given by [2]. Suppose $\gamma(E, \omega) > 0$ for all $E \in [E_1, E_2]$ and a.e. $\omega \in \mathbb{T}^b$. Then for any $\theta$, $H$ has Anderson localization in $[E_1, E_2]$ for a.e. $\omega$.*

Combining this with Theorem 1, Bourgain (2003) obtained that for $\lambda > \lambda(f)$, $H$ as above satisfies Anderson localization for a.e. $\omega$. Those results were recently extended by S Klein to potentials belonging to certain Gevrey classes. One very important ingredient of this method is the theory of semialgebraic sets that allows one to obtain polynomial algebraic complexity bounds for certain "exceptional" sets. Combined with measure estimates coming from the large deviation analysis of $(1/n)\ln\|M_n(\theta)\|$ (using subharmonic function theory and involving approximate Lyapunov exponents),

this theory provides necessary information on the geometric structure of those exceptional sets. Such algebraic complexity bounds also exist for the almost-Mathieu operator and are actually sharp albeit trivial in this case due to the specific nature of the cosine.

Further corollaries of positive Lyapunov exponents for analytic sampling functions $f$ and $b=1$ include Hölder regularity of the integrated density of states, zero-dimensionality of spectral measures for all $\omega, \theta$, almost Lipshitz continuity of spectral gaps, continuity of measure of the spectrum (in frequency), and vanishing of lower transport exponents for all $\omega, \theta$. Some weaker statements are available for $b>1$ or $f$ belonging to certain Gevrey classes.

## Without Lyapunov Exponents

While having led to significant advances, Lyapunov exponents have obvious limitations, as any method based on them is restricted to one-dimensional nearest-neighbor Laplacians. It turns out that the above methods can be extended to obtain nonperturbative results in certain quasi-one-dimensional situations where Lyapunov exponents do not exist. For example, nonperturbative localization results extend to the strip (of arbitrary dimension).

The following nonperturbative theorem deals with the case of small coupling:

**Theorem 5** *Let $H$ be an operator [2], where $f$ is real analytic on $\mathbb{T}$ and $\omega$ is Diophantine. then, for $\lambda < \lambda(f)$, $H$ has purely absolutely continuous spectrum for a.e. $\theta$.*

We note that an analog of this theorem does not hold in the multifrequency case (see next section). The results of this type are obtained by a method (developed by Bourgain and Jitomirskaya in 2000–02) that studies large deviations for the quantities of the form $(1/n)\ln|\det(H-E)_\Lambda|$ and path-determinant expansion for the matrix elements of the resolvent. Those techniques apply also to certain other situations with long-range Laplacians, for example, the kicked-rotor model. Theorem 5 is a result on nonperturbative localization in disguise as it was obtained using duality from a localization theorem for a dual model which has in general a long-range Laplacian and a cosine potential, and was in turn obtained by an extension of the method of Jitomirskaya (1999). A certain measure-theoretic version of it allowing nonlocal Laplacians but leading only to continuous spectrum is also available (see Bourgain (2004)).

## Multidimensional Case: $d>1$

As mentioned above, there are very few results in the multidimensional lattice case ($d>1$). Essentially, the only result that existed before the recent developments was a perturbative theorem – an extension by Chulaevsky–Dinaburg of Sinai's method to the case of operator [1] on $\ell^2(\mathbb{Z}^d)$ with $V_n = \lambda f(n\cdot\omega), \omega\in\mathbb{R}^d$, where $f$ is a cos-type function on $\mathbb{T}$. This also holds nonperturbatively for any real-analytic $f$ (see Bourgain (2004)). Note that since $b=1$, this avoids most serious difficulties and is therefore significantly simpler than the general multidimensional case. We therefore have:

**Theorem 6** *For any $\epsilon>0$ there is $\lambda(f,\epsilon)$, and, for $\lambda>\lambda(f,\epsilon), \Omega(\lambda,f)\subset\mathbb{T}^d$ with $\mathrm{mes}(\Omega)<\epsilon$, so that for $\omega\notin\Omega$, operator [1] with $V_n$ as above has Anderson localization.*

This should be confronted with the following theorem of Bourgain:

**Theorem 7** *Let $d=2$ and $f(\theta)=\cos 2\pi\theta$ in $H=H_\omega$ defined as above. Then for any $\lambda$ measure of $\omega$ s.t. $H_\omega$ has some continuous spectrum is positive.*

Therefore, for large $\lambda$ there will be both $\omega$ with complete localization as well as those with at least some continuous spectrum. This shows that nonperturbative results do not hold in general in the multidimensional case! Perturbative results, however, had been obtained, see next section.

A similar (in fact, dual) situation is observed for one-dimensional multifrequency ($d=1; b>1$) case at small disorder. One has, by duality:

**Theorem 8** *Let $H$ be given by [2] with $\theta,\omega\in\mathbb{T}^b$ and $f$ real analytic on $\mathbb{T}^b$. Then for any $\epsilon>0$ there is $\lambda(f,\epsilon)$ s.t. for $\lambda<\lambda(f,\epsilon)$ there is $\Omega(\lambda,f)\subset\mathbb{T}^b$ with $\mathrm{mes}(\Omega)<\epsilon$ so that for $\omega\notin\Omega$, $H$ has purely absolutely continuous spectrum.*

And also

**Theorem 9** *Let $d=1, b=2$ and $f$ be a trigonometric polynomial on $\mathbb{T}^2$ with a nondegenerate maximum. Then for any $\lambda$, measure of $\omega$ s.t. $H_\omega$ has some point spectrum, dense in a set of positive measure, is positive.*

Therefore, unlike the $b=1$ case (see Theorem 5), nonperturbative results do not hold for absolutely continuous spectrum at small disorder.

## Perturbative Localization by Nonperturbative Methods

While the above demonstrates the limitations of the nonperturbative results, the nonperturbative

methods have been applied to significantly simplify the proofs and obtain new perturbative results that previously had been completely beyond reach.

Many such applications, that are outside the scope of this article, are described in Bourgain (2004). In particular, new results on the construction of quasiperiodic solutions in Melnikov problems and nonlinear PDEs, obtained by using certain ideas developed for nonperturbative quasiperiodic localization (e.g., the theory of semialgebraic sets), are presented there. Other results in this group contain localization for the skew-shift model by Bourgain–Goldstein–Schlag, almost periodicity for the quantum kicked-rotor model by Bourgain and Bourgain–Jitomirskaya, and localization for potentials in higher Gevrey classes by S Klein.

The main goal in a nonperturbative method is to obtain exponential off-diagonal decay for the matrix elements of the Green's function of box-restricted operators along with subexponential bounds on the distance from the spectrum of such box restrictions to a given energy. From that result one can obtain localization through elimination of energy via an argument involving complexity bounds on semialgebraic sets (see Bourgain (2004)).

A nonperturbative way to achieve the desired Green's function estimates uses Cramer's rule to represent the matrix elements of the resolvent. Then, in the one-dimensional (in space) case it is often possible to obtain the estimates from the positivity of Lyapunov exponents: uniformly for the numerator, and from large deviation bounds for the subharmonic functions for the denominator. This is done in one step for a sufficiently large scale (see the subsection "Corollaries of positive Lyapunov exponents")

A perturbative way consists of establishing the desired estimates in a multiscale scheme: namely, the estimates are proved outside a set of parameters of (subexponentially) decaying (in the size of the box) measure. Moreover, this set should be shown to have a semialgebraic description, in order to make possible sublinear upper bounds on the number of times a trajectory of a given phase (under the underlying rotation or other ergodic transformation of the torus) hits the "forbidden" set. This, plus certain subharmonic function arguments, allows passage to a larger scale through a repeated use of the resolvent identity.

An application that is most relevant to the current article is localization for a "true" $d > 1$ situation. The best currently available result is the following very recent theorem (Bourgain 2005):

**Theorem 10** *Let $d = b$ and let $f$ be real analytic on $\mathbb{T}^d$ such that for all $i = 1, \dots, d$ and $(\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d) \in \mathbb{T}^{d-1}$, the map*

$$\theta_i \mapsto v(\theta_1, \dots, \theta_i, \dots, \theta_d)$$

*is a nonconstant function of $\theta_i \in \mathbb{T}$. Then for any $\epsilon > 0$ there is $\lambda(f, \epsilon)$ s.t. for $\lambda > \lambda(f, \epsilon)$ there is $\Omega(\lambda, f) \subset \mathbb{T}^d$ with $\mathrm{mes}(\Omega) < \epsilon$ so that for $\omega \notin \Omega$ operator [1] with $V_n = \lambda f(n_1\omega_1, n_2\omega_2)$ has Anderson (and dynamical) localization.*

This result was obtained previously, for $d = 2$ only, by Bourgain, Goldstein, and Schlag. There were some serious purely arithmetic difficulties that prevented an extension of this result to higher dimensions. In the previous results on localization there were two major steps: estimations on the Green's function for fixed energy and elimination of energy. The main difficulty in the multidimensional case lies in establishing the sublinear bound described above, that enters in the first step. It is for this bound that an arithmetic condition on $\omega$ was needed. The condition used was to guarantee that the number of $(n_1, n_2) \in [1, N]^2$ such that $(n_1\omega_1, n_2\omega_2)(\mathrm{mod}\, \mathbb{Z}^2) \in S$ is bounded from above by $N^\alpha$ for some $\alpha < 1$, uniformly for all semialgebraic sets $S$ of degree $D$, with $D'/D = o(1/N)$ and with the measure of all horizontal and vertical sections $S_x$ satisfying $\log \mathrm{mes} S_x = o(\log 1/N)$. This condition roughly means that too many points close to an algebraic curve of a bounded degree would force it to oscillate more than it should. Such a statement is essentially two dimensional and not extendable to $d \geq 3$. In Theorem 10, Bourgain circumvents it by using from the beginning the theory of semialgebraic sets to eliminate energy and the translation variable to get conditions on $\omega$ (that depend on the potential) already in the first step.

## Dynamical Localization

Anderson localization does not in itself guarantee absence of quantum transport, or nonspread of an initially localized wave packet, as characterized, for example, by boundedness in time of moments of the position operator. This was first observed in del Rio *et al.* (1996), where a rather artificial example of coexistence of exponential localization and quantum transport was constructed. However, such phenomena also happen in models of interest to physicists such as the random dimer model. Considering for simplicity the second moment

$$\langle x^2 \rangle_T = \frac{1}{T} \int_0^T \sum_n |\Psi_t(n)|^2 n^2 \, \mathrm{d}t$$

we will say that $H$ exhibits dynamical localization if $\langle x^2 \rangle_T <$ const. We will say that the family $\{H_\theta\}_{\theta \in \mathbb{T}^b}$ exhibits strong dynamical localization if $\int_{\mathbb{T}^b} \mathrm{d}\theta \, \sup_t \langle x^2 \rangle_t <$ const. We note that the results mentioned below will hold with more restrictive

definitions of dynamical localization (involving the higher moments of the position operator) as well. Dynamical localization implies pure point spectrum by RAGE theorem so it is a strictly stronger notion.

It turns out that nonperturbative methods allow for such dynamical upgrades as well. For the almost-Mathieu operator, strong dynamical localization holds throughout the regime of localization. It was shown by Bourgain and Jitomirskaya that in Theorems 4 and 6 as well as some other localization results, dynamical localization also holds (see Bourgain (2004)). However, methods that require elimination of certain frequencies based on implicit conditions currently do not provide sufficient information to obtain strong (i.e., averaged) dynamical localization, like what was done in the almost-Mathieu case.

## Quasiperiodic Localization and Cantor Spectrum

A remarkable feature of quasiperiodic operators with $b = d = 1$ is their tendency to have Cantor spectrum. In particular, it was conjectured that all almost-Mathieu operators (for all nonzero couplings and all irrational frequencies) have Cantor spectrum. This conjecture became known as the Ten Martini problem. In a significant recent development (Puig 2004), it was shown that for Diophantine frequencies Cantor structure of the spectrum follows from localization for phase $\theta = 0$, with corresponding eigenvalues being the boundaries of noncollapsed gaps. The key idea here is that for energies dual to eigenvalues of $H_0$, corresponding to localized eigenfunctions, the rotation number of the transfer-matrix cocycle is of the form $k\omega(\mathrm{mod}\mathbb{Z})$, thus they are the ends of the gaps (possibly collapsed). However, a collapsed gap in this case would correspond to reducibility of the system to the identity which can be shown to contradict the simplicity of pure point spectrum for the dual model. Since those energies form a dense subset of the spectrum the result follows. The same idea works, thus establishing Cantor spectrum, for potentials that are generic in certain sense. Localization also played an important role in the final proof of the Ten Martini conjecture, for all irrationals (Avila and Jitomirskaya 2005). It can be shown that proving localization for a large set of phases allows one to conclude reducibility of the transfer-matrix cocycle for the dual model, for a large set of energies, and this in turn can be shown to contradict the presence of an interval in the spectrum.

## Acknowledgement

The research work of this author was partially supported by NSF grant DMS-0300974.

*See also:* Multiscale Approaches; Quantum Hall Effect; Quasiperiodic Systems; Schrödinger Operators; Stability Theory and KAM.

## Further Reading

Avila A and Jitomirskaya S (2005) The ten martini problem. *Preprint*.

Bourgain J (2003) Positivity and continuity of the Lyapunov exponent for shifts on $\mathbb{T}^d$ with arbitrary frequency vector and real analytic potential. *Preprint*.

Bourgain J (2004) Green's Function Estimates for Lattice Schrödinger Operators and Applications. *Annals of Mathematics Studies*, vol. 158. x+173 pp. Princeton, NJ: Princeton University Press.

Bourgain J (2005) Anderson localization for quasi-periodic lattice Schrödinger operators on $\mathbb{Z}^d$, $d$ arbitrary. *Preprint*.

Bourgain J and Goldstein M (2000) On nonperturbative localization with quasiperiodic potential. *Annals of Mathematics* 152: 835–879.

del Rio R, Jitomirskaya S, Last Y, and Simon B (1996) Operators with singular continuous spectrum: IV. *Journal d'Analyse Mathematique* 69: 153–200.

Eliasson LH (1998) Reducibility and point spectrum for linear quasiperiodic skew products. Doc. Math, Extra volume ICM 1998 II, 779–787.

Jitomirskaya S (1999) Metal–insulator transition for the almost Mathieu operator. *Annals of Mathematics* 150: 1159–1175.

Last Y (2005) Spectral theory of Sturm–Liouville operators on infinite intervals: a review of recent developments. In: Sturm–Liouville Theory, pp. 99–120. Basel: Birkhauser.

Puig J (2004) Cantor spectrum for the almost Mathieu operator. *Communications in Mathematical Physics* 244: 297–309.

Simon B (2000) Schrödinger operators in the twenty-first century. In: *Mathematical Physics 2000*. pp. 283–288. London: Imperial College.

# Loop Quantum Gravity

**C Rovelli**, Université de la Méditerranée et Centre de Physique Théorique, Marseilles, France



## Introduction

Loop quantum gravity (LQG) is a mathematical formalism that defines a tentative quantum theory of spacetime. Equally, the formalism provides a description of the gravitational field in regimes in which its quantum properties cannot be neglected. The distinctive feature of LQG is to be a quantum field theory consistent with general relativity.

According to general relativity, the physical fields that form the world do not live on a background spacetime. Rather, these fields make up spacetime themselves ("background independence"). Accordingly, the quanta of a quantum field theory compatible with this principle – the *s*-knots described below – do not live on a background spacetime: rather, they themselves form physical spacetime.

This physical idea is realized in the formalism by the gauge invariance under active diffeomorphisms of the manifold on which the fields are originally defined ("diffeomorphism invariance"). Such gauge invariance renders the localization of the field's excitations on the manifold physically irrelevant.

LQG implements these physical motivations by merging two traditional lines of thinking in theoretical physics. The first is the long-standing idea that gauge fields are naturally understood in terms of variables associated to lines (holonomies of the gauge connection, Wilson loops, Faraday lines, ...). This idea can be traced to Faraday's initial intuition that gave birth to modern field theory: physical fields are real entities formed by lines. The second is the background-independent canonical or covariant quantization of general relativity developed by following the ideas of Wheeler, DeWitt, and Hawking. Each of these two lines of research has encountered serious obstructions, but the two turn out to solve each others' difficulties: the formulation in terms of holonomies renders the old ill-defined background-independent quantum gravity well defined; conversely, background independence cures the divergences associated to the Wilson loop basis.

The formalism of LQG can be separated into two parts. A kinematics, describing the quantum properties of space, and a dynamics, describing its evolution. Here we outline the LQG kinematics, and we give only the main result of the LQG dynamics.

LQG can be extended to include standard matter couplings such as fermions and Yang–Mills fields. It finds numerous applications, for instance, in early cosmology, astrophysics and black hole thermodynamics (*see* Black Hole Mechanics, Quantum Cosmology).

So far no empirical evidence supports the physical correctness of this – nor of any other – tentative theory of quantum gravity.

## General Relativity in Canonical Form

Classical general relativity is the field theory describing the gravitational field and the structure of physical spacetime. It is a well-established physical theory, strongly supported empirically.

In its Riemannian version, the theory can be written in canonical form in terms of two fields on a three-dimensional (3D) manifold $\Sigma$ with coordinates $x^a (a = 1, 2, 3)$: a 2-form $E = E^a \epsilon_{abc}\, dx^a\, dx^b$, called the "triad field" and a 1-form $A = A_a\, dx^a$, called the "gravitational connection" ($\epsilon_{abc}$ is the totally antisymmetric tensor density). Both take values in the su(2) algebra, and they satisfy the three "constraint" equations

$$\mathcal{G} = D_a E^a = 0 \qquad [1]$$

$$C_a = \text{tr}[F_{ab} E^a] = 0 \qquad [2]$$

$$C = \text{tr}[F_{ab} E^a E^b] = 0 \qquad [3]$$

$D_a$ is the SU(2) covariant derivative defined by the connection $A$, $F_{ab}$ is the SU(2) curvature of $A$, and the trace is on su(2).

$E$ and $A$ are canonically conjugate: their Poisson brackets are $\{E^a(x), A_b(y)\} = 8\pi G c^{-3} \delta^a_b \delta^3(x, y)$; where $G$ is the Newton constant, $c$ is the speed of light, $\delta^a_b$ is the Kronecker delta, and $\delta^3(x, y)$ is the Dirac-delta on $\Sigma$, which is a scalar density in $x$. The Poisson brackets of $\mathcal{G}$ with the fields define their SU(2) gauge transformations: $E$ transforms in the adjoint representation and $A$ transforms as a connection. The Poisson brackets of $C_a$ (more precisely, of an appropriate linear combination of $C_a$ and $\mathcal{G}$) with the fields determine their transformation under a diffeomorphism of $\Sigma$: $E$ transforms as a 2-form and $A$ as a 1-form. The Poisson brackets of $C$ with the fields generate their coordinate time evolution. If the $t$ derivatives of the fields $E(x^a, t)$ and $A(x^a, t)$ are given by their Poisson brackets with (the 3D integral of) $C$, then (assuming that the determinant $E = \sqrt{\det \text{tr}[E^a E^b]}$ does not vanish) the metric field

$g^{00} = 1$, $g^{a0} = 0$, $g^{ab} = \mathrm{tr}[E^a E^b]/E$ is a general solution of the Riemannian Einstein equations in a fixed gauge.

The physical Lorentzian theory can be obtained in this formalism in two ways. Either by adding an appropriate term to eqn [3], or by taking $A$ in sl(2,C) and satisfying a suitable reality condition. (For more details, *see* Canonical General Relativity.)

## Spin Network and *s*-Knot States

LQG can be defined as a Schrödinger quantization of the canonical formalism described above. The space of the quantum states is defined as a Hilbert space $\mathcal{K}$ of Schrödinger wave functionals $\Psi[A]$ of the gravitational connection. The nontrivial aspect of this construction is the definition of a scalar product invariant under the two kinematical gauge invariances of the theory: the local SU(2) and the diffeomorphisms transformations generated by the constraints [1] and [2]. The state space $\mathcal{K}$ is defined as follows (*see* Quantum Geometry and its Applications for an essentially equivalent construction).

Given an su(2) connection $A$ and an oriented path $\gamma : s \in [0,1] \to x^a(s) \in \Sigma$, recall that the "holonomy" $U[A,\gamma]$ of $A$ along $\gamma$ is the element of SU(2) defined by

$$\frac{\mathrm{d}}{\mathrm{d}s} U[A,\gamma](s) + \dot{\gamma}^a(s) A_a(\gamma(s)) U[A,\gamma](s) = 0 \qquad [4]$$

$$U[A,\gamma](0) = \mathbb{1}, \qquad U[A,\gamma] = U[A,\gamma](1) \qquad [5]$$

where $\dot{\gamma}^a(s) \equiv \mathrm{d}x^a(s)/\mathrm{d}s$ is the tangent to the path. The solution of this equation is usually written in the form

$$U[A,\gamma] = \mathcal{P}\mathrm{e}^{\int_\gamma A} \qquad [6]$$

where the path ordered $\mathcal{P}$ is understood as acting on the power series expansion of the exponential.

Let $\mathcal{A}$ be the space of the smooth connections $A$ on $\Sigma$. (For technical reasons, it is convenient to consider smooth fields $A$ defined everywhere in $\Sigma$ except at most at a finite number of points, and the group *Diff** of the "extended diffeomorphisms" defined by the continuous invertible maps $\phi : \Sigma \to \Sigma$ that are smooth everywhere in $\Sigma$ except at most at a finite number of points.) A graph $\Gamma$ is an ordered collection of smooth oriented paths, $\gamma_l$, denoted as links, with $l = 1, \ldots, L$, where the links overlap only at their endpoints, called nodes. Given a graph $\Gamma$ and a smooth, Haar-integrable complex function $f : \boldsymbol{U} \in (\mathrm{SU}(2))^L \mapsto f(\boldsymbol{U}) \in C$, the couple $(\Gamma, f)$ defines the ("cylindrical") functional of $A$

$$\Psi_{\Gamma,f}[A] = f(\boldsymbol{U}[A,\Gamma]) \qquad [7]$$

$$\boldsymbol{U}[A,\Gamma] \equiv (U[A,\gamma_1], \ldots, U[A,\gamma_L]) \qquad [8]$$

Let $\mathcal{L}$ be the linear space of all functionals $\Psi_{\Gamma,f}[A]$, for all $\Gamma$ and $f$. $\mathcal{L}$ is dense (in an appropriate sense) in the space of all continuous functionals on $\mathcal{A}$.

An SU(2) and *Diff** invariant scalar product can be defined in $\mathcal{L}$ as follows. If two functionals $\Psi_{\Gamma,f}[A]$ and $\Psi_{\Gamma,g}[A]$ are defined by the same graph $\Gamma$, define

$$\langle \Psi_{\Gamma,f} | \Psi_{\Gamma,g} \rangle \equiv \int \mathrm{d}\boldsymbol{U}\, \overline{f(\boldsymbol{U})}\, g(\boldsymbol{U}) \qquad [9]$$

where $\mathrm{d}\boldsymbol{U}$ is the Haar measure on $(\mathrm{SU}(2))^L$. The extension to functionals defined on different graphs is obtained by observing that $(\Gamma, f)$ and $(\Gamma', f')$ define the same functional if $\Gamma$ contains $\Gamma'$ and $f$ is independent of the variables in $\Gamma$ but not in $\Gamma'$. It follows that any two given functionals $\Psi_{\Gamma',f'}$ and $\Psi_{\Gamma'',g''}$ can be written as functionals $\Psi_{\Gamma,f}$ and $\Psi_{\Gamma,g}$ with the same graph $\Gamma$, where $\Gamma$ is obtained from the union of $\Gamma'$ and $\Gamma''$. Using this, the scalar product [9] is defined for any two functionals in $\mathcal{L}$:

$$\langle \Psi_{\Gamma',f'} | \Psi_{\Gamma'',g''} \rangle \equiv \langle \Psi_{\Gamma,f} | \Psi_{\Gamma,g} \rangle \qquad [10]$$

Standard completion in the Hilbert norm defines the kinematical Hilbert space $\mathcal{K}$ of LQG. $\mathcal{L}$ is dense in $\mathcal{K}$ and defines the Gelfand triple $\mathcal{L} \subset \mathcal{K} \subset \mathcal{L}^*$. $\mathcal{K}$ carries a natural unitary representation of the group of local SU(2) representations and a natural unitary representation $U_\phi$ of the group of the extended diffeomorphism of $\Sigma$. These two properties are nontrivial; they represent the main physical motivation for the definition of the scalar product. The SU(2)-invariant subspace of $\mathcal{K}$ is a proper subspace $\mathcal{K}_0$.

An orthonormal basis in $\mathcal{K}_0$ can be defined using the Peter–Weyl theorem. The basis states are labeled by a graph $\Gamma$, by the assignment of a nonvanishing spin $j_\gamma$ to each link $\gamma \in \Gamma$ and by the assignment of a basis element $i_n$ in the space of the intertwiners (invariant tensors in the tensor product of the representations space of the adjacent links) at each node $n$ of $\Gamma$. The triple $S = (\Gamma, j_\gamma, i_n)$ is called an imbedded *spin network*. The quantum state $\Psi_S[A] = \langle A|S\rangle$ in $\mathcal{K}_0$ labeled by the spin network $S = (\Gamma, j_\gamma, i_n)$ is the cylindrical function obtained by contracting the representation matrices of the holonomies $U(A,\gamma)$, in the representations $j_\gamma$, with the invariant tensors at the nodes.

The diffeomorphism-invariant state space $\mathcal{K}_{\mathrm{diff}}$ is the SU(2) and diffeomorphism invariant subspace of $\mathcal{L}^*$. It is the (closure of the) image of the map $P_{\mathrm{diff}} : \mathcal{L} \to \mathcal{L}^*$ defined by

$$(P_{\mathrm{diff}}\Psi)(\Psi') = \sum_{\Psi'' = U_\phi \Psi} \langle \Psi'', \Psi' \rangle \quad \forall \Psi, \Psi' \in \mathcal{K} \qquad [11]$$

The sum is over all states $\Psi''$ in $\mathcal{L}$ for which there exists a diffeomorphism $\phi$ such that $\Psi'' = U_\phi\Psi$; this is a finite sum. The scalar product on this image is naturally defined by

$$\langle P_{\text{diff}}\Psi_S, P_{\text{diff}}\Psi_{S'}\rangle_{\mathcal{K}_{\text{diff}}} \equiv (P_{\text{diff}}\Psi_S)(\Psi_{S'}) \qquad [12]$$

The space $\mathcal{K}_{\text{diff}}$ obtained in this manner is separable.

The images $|s\rangle = P_{\text{diff}}|S\rangle$ of the spin network states are called $s$-knot states. They span $\mathcal{K}_{\text{diff}}$. They are determined only by the diffeomorphism equivalence class $s$ of the spin network $S$. Namely, by an abstract (non-imbedded) knotted graph, colored with spins and intertwiners. These colored knots are called $s$-knots or abstract spin networks. The $s$-knot states have a straightforward physical interpretation as quantum excitations of space, discussed below.

## Operators and Quanta of Space

The state space defined above carries a quantum representation of classical observables of general relativity. The classical quantity $U[A,\gamma]$, a function of the field variable $A$, acts naturally as a multiplicative operator on $\mathcal{K}$. Thus, $\mathcal{K}$ provides a Schrödinger functional representation $\Psi[A]$ of quantum gravity, which diagonalizes the (holonomy of the) gravitational connection. The two constraints [1] and [2] generate SU(2) gauge and diffeomorphism transformations on $A$. The corresponding transformations on the Schrödinger functional states $\Psi[A]$ are given by the unitary representations mentioned above. The quantum implementation of the two constraint equations [1] and [2], following Dirac's theory of constrained quantum systems, is the requirement of invariance under these transformations. The space $\mathcal{K}_{\text{diff}}$ is the solution to these requirement.

The triad field operator $E$ can be defined only if suitably smeared. Since $E$ is a 2-form, its geometrically natural smearing is with a 2D surface. (The 1-form field $A$ is smeared over a line in $U[A,\gamma]$.) Given a finite 2D surface $\mathcal{S}: \sigma = (\sigma^1, \sigma^2) \mapsto x^a(\sigma) \in \Sigma$, the smeared field

$$E[\mathcal{S}] = \int_{\mathcal{S}} E = \int \mathrm{d}^2\sigma\, \epsilon_{abc} \frac{\partial x^a}{\partial\sigma^1}\frac{\partial x^b}{\partial\sigma^2} E^c(x(\sigma)) \qquad [13]$$

is quantized by the functional derivative operator

$$E[\mathcal{S}] \equiv -\mathrm{i}\hbar\frac{8\pi G}{c^3}\int \mathrm{d}^2\sigma\, \epsilon_{abc}\frac{\partial x^a}{\partial\sigma^1}\frac{\partial x^b}{\partial\sigma^2}\frac{\delta}{\delta A_c(x(\sigma))} \qquad [14]$$

This operator is well defined on $\mathcal{K}$ and the quantum operators $E[\mathcal{S}]$ and $U[A,\gamma]$ define a linear representation of the Poisson algebra of the corresponding classical quantities. Thus, they define a quantization of the kinematics of general relativity. Notice that in a general covariant quantum field theory field operators can be well defined even if smeared on low-dimensional regions, while in conventional quantum field theory, these operators need to be smeared over 3D or 4D regions.

A simple calculation shows that if $\mathcal{S}$ and $\gamma$ intersect once,

$$E_v[\mathcal{S}]U[A,\gamma] = \pm\mathrm{i}\hbar\frac{8\pi G}{c^3}U[A,\gamma_1]vU[A,\gamma_2] \qquad [15]$$

where $v \in \mathrm{su}(2)$, we have written $E_v = \mathrm{tr}[vE]$, $\gamma_{1,2}$ are the two paths into which $\gamma$ is partitioned by the surface, and the sign is determined by the relative orientation of $\mathcal{S}$ and $\gamma$. More generally, $E[\mathcal{S}]U[A,\gamma]$ is a sum of one such term per intersection between $\mathcal{S}$ and $\gamma$.

Composite operators can be constructed in terms of these operators. In particular, using standard formulas in classical general relativity, the area of the surface $\mathcal{S}$ can be written as a Riemann sum

$$A[\mathcal{S}] = \lim_{N\to\infty}\sum_n \sqrt{\mathrm{tr}[E(\mathcal{S}_n)E(\mathcal{S}_n)]} \qquad [16]$$

where $\mathcal{S}_n, n = 1, \ldots, N$, is a Riemann partition of the surface. A straightforward calculation based on eqn [15] shows that, if $\mathcal{S}$ cuts $n$ links of a spin network carrying spins $(j_1 \ldots j_n) = \boldsymbol{j}$, then the spin network state $|S\rangle$ is an eigenstate of $A[\mathcal{S}]$ with eigenvalue

$$A_{\boldsymbol{j}} = \frac{8\pi\hbar G}{c^3}\sum_{i=1,n}\sqrt{j_i(j_i+1)} \qquad [17]$$

where $j_i = 1/2, 1, 3/2, 2, \ldots$ These are therefore discrete eigenvalues of the area. All eigenvalues of the area operator $A[\mathcal{S}]$ are real and discrete and $A[\mathcal{S}]$ is a self-adjoint operator. Similar results are obtained for the volume operator. This gets a discrete contribution for each node of a spin network.

These spectral properties of the area and volume operators determine the physical interpretation of the spin network states: the nodes of the spin network represent quanta of space with quantized volume; the nodes are connected by links representing quanta of surface with quantized area. The graph $\Gamma$ determines the adjacency relations between the individual quanta of space; the intertwiners $i_n$ are volume quantum numbers; the spins $j_\gamma$ are area quantum numbers.

The interpretation carries over to the $s$-nodes, which represent the same quantum excitations of space, up to its manifold coordinatization, which is physically irrelevant because of the gauge invariance under

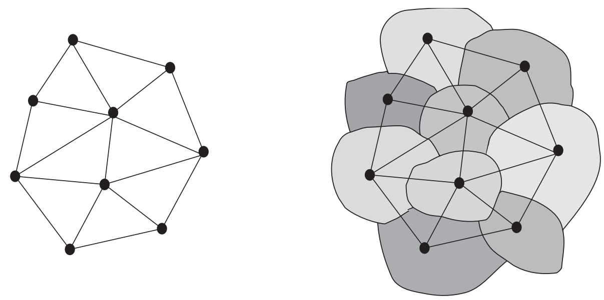


**Figure 1** The graph of an *s*-knot, namely an abstract spinfoam, and the set of quanta of space it represents. Each node $n$ of the graph defines a quantum of space. The associated intertwiner $i_n$ is the corresponding volume quantum number. Two quanta of space are adjacent if the corresponding nodes are linked. A link $\gamma$ cuts the elementary surface separating the two quanta and its spin $j_\gamma$ is the area quantum number of this surface.

diffeomorphisms of $\Sigma$. An $s$-knot state $|s\rangle$ with $N$ nodes represents a quantum excitation of space with $N$ quanta of space adjacent to one another according to the connectivity of $\Gamma$ (see Figure 1).

Notice that the quantum states $|s\rangle$ do not represent quantum excitations living in the physical space: they represent quantum excitations of the physical space. For instance, the state $|0\rangle$ defined by the empty graph does not represent an "empty" physical space, but the absence of any physical space. A generic quantum state of the physical space is represented by a normalizable linear superposition of these discrete quantized spacetimes (*see* Knot Invariants and Quantum Gravity).

In a nongeneral covariant context, the kinematical quantization predictions of quantum theory (such as the quantization of the angular momentum) are obtained from the spectral properties of operators that represent measurements at a given time. In the general covariant Hamiltonian formalism, the corresponding kinematical quantization predictions are given by spectral properties of "partial observables" operators, which in general are not gauge invariant in the sense of Dirac. Area and volume are partial observables of this kind. Their spectra are therefore interpreted as physical predictions of LQG (up to an overall numerical factor, called the Immirzi parameter, which is obtained in certain variants of the theory).

## Dynamics

The dynamics of the theory is obtained in terms of a "Hamiltonian constraint" operator $C$ that quantizes the constraint [3]. Different variants of the operator $C$, and of its Lorentzian version, have been constructed. The operator is defined via a suitable regularization procedure. The description of these constructions exceeds the scope of this article, and we limit ourself here to mentioning the main result and a few general comments.

The main result of the LQG dynamics is that $C$ turns out to be well defined and ultraviolet-finite when restricted to $\mathcal{K}_{\text{diff}}$. Finiteness holds also when standard matter couplings, such as Yang–Mills fields and fermions, are added.

The reason for this finiteness can be understood as a consequence of the discrete nature of space implied by the spectral properties of the geometric operators described above. The limit in which the ultraviolet cutoff, introduced to regulate $C$, is removed turns out to be trivial on the diffeomorphism-invariant states in $\mathcal{K}_{\text{diff}}$. This is because this limit probes the short-distance regime, but there is no *physical* (gauge-invariant) short distance, in a theory in which geometry turns out to be quantized at the Plank scale. Since the physical states in $\mathcal{K}_{\text{diff}}$ define a physical geometry only at scales larger than the Planck scale $\hbar G c^{-3}$, the "short-distance" modes in the coordinate manifold $\Sigma$ turn out to be pure gauge. This interplay between quantum field-theoretical and general-relativistic physics is the distinctive character of LQG.

Finally, we sketch the formal structure that dynamics can take in the general covariant Hamiltonian formalism of LQG. The operator $C$ defines a linear operator $P \sim \delta(C)$, usually (improperly) denoted the "projector," which sends states in $\mathcal{K}_{\text{diff}}$ into the kernel of $C$, formed by the generalized $\mathcal{K}_{\text{diff}}$ vectors that solve the Wheeler–De Witt equation $C\Psi = 0$ (*see* Wheeler–De Witt Theory). Matrix elements of $P$ are interpreted as transition amplitudes between quantum states of space.

Physical predictions for processes that take place in a finite spacetime region $\mathcal{R}$ can be obtained, in principle, as follows. One considers a state $|\Psi\rangle$ representing the result of the measurement of partial observables of the 3D boundary of a spacetime region $\mathcal{R}$. $|\Psi\rangle$ codes the nonrelativistic notions of initial, boundary and final conditions. Then $\langle 0|P|\Psi\rangle$ can be interpreted as a relative probability amplitude associated to this result. A formal expansion of this amplitude in powers of $C$ generates a spinfoam sum (*see* Spin Foams) that can be understood as the "quantum gravity sum over histories" in $\mathcal{R}$.

A systematic technique for computing physical transition amplitudes from the background-independent and nonperturbative formalism of LQG has not yet been developed.

*See also:* BF Theories; Black Hole Mechanics; Canonical General Relativity; Knot Invariants and Quantum Gravity; Knot Theory and Physics; Quantum Cosmology; Quantum Dynamics in Loop Quantum Gravity; Quantum Geometry and its Applications; Spin Foams; Wheeler–De Witt Theory.

## Further Reading

Ashtekar A and Lewandowski J (2004) Background independent quantum gravity: a status report. *Classical and Quantum Gravity* 21: R53.
Rovelli C (1998) Loop quantum gravity. Living Reviews in Relativity (*electronic journal*). (http://www.livingreviews.org/Articles/Volume1/1998-1rovelli).
Rovelli C (2004) *Quantum Gravity*. Cambridge: Cambridge University Press.
Rovelli C and Smolin L (1990) Loop space representation for quantum general relativity. *Nuclear Physics B* 331: 80.
Rovelli C and Smolin L (1995) Discreteness of area and volume in quantum gravity. *Nuclear Physics B* 442: 593.
Rovelli C and Smolin L (1995) Discreteness of area and volume in quantum gravity – erratum. *Nuclear Physics B* 456: 734.
Smolin L (2002) *Three Roads to Quantum Gravity*. New York: Perseus Books Group.
Thiemann T *Modern Canonical Quantum General Relativity*. Cambridge: Cambridge University Press (to appear).


# Lorentzian Geometry


**P E Ehrlich**, University of Florida, Gainesville, FL, USA
**S B Kim**, Chonnam National University, Gwangju, South Korea




## Introduction

Einstein's (1916) use of differential geometry as an essential tool in his theory of general relativity has long been a motivation for the study of Lorentzian geometry. More recently, the influential monographs of R Penrose (1972) and of S Hawking and G Ellis (1973), the latter still cited by some as the Bible of general relativity, so fascinated differential geometers that Lorentzian geometry took its place alongside of global Riemannian geometry as a worldwide research area.

Let $M$ be a smooth $n$-dimensional manifold, $n \geq 2$, with a countable basis. A Lorentz metric $g = < , >$ on $M$ is a symmetric nondegenerate (0, 2) tensor field on $M$ of index $(-, +, \ldots, +)$. The existence of such a tensor field implies that $M$ admits a (non-oriented) line field; hence, some compact manifolds like $S^2$ do not admit such metrics. A nonzero tangent vector $v$ in $TM$ is then timelike (resp., nonspacelike, null, spacelike) according to whether $g(v,v) < 0$ (resp., $\leq 0$, $= 0$, $> 0$). A Lorentzian manifold $(M, g)$ is a pair consisting of a smooth manifold together with a choice of Lorentz metric. In this article, we use the convention that a spacetime $(M, g)$ is a Lorentzian manifold together with a choice of time orientation, that is, a continuous timelike vector field $X$ on $M$. Then a tangent vector $v$ based at $p$ may be consistently defined to be future (resp., past) directed if $g(X(p), v) < 0$ (resp., $> 0$). (Some authors also require that $(M, g)$ be space oriented.) If a Lorentzian manifold happens not to be time orientable, then a 2-fold covering manifold with the induced pullback metric will be time orientable. Also basic are the notations $p \ll q$ (resp., $p \leq q$) if there is a future-directed timelike (resp., nonspacelike) curve from $p$ to $q$ and the corresponding chronological (resp., causal) future of $p$ given by $I^+(p) = \{q \in M;\ p \ll q\}$ and $J^+(p) = \{q \in M;\ p \leq q\}$.

For a Riemannian manifold $(N, g_0)$, the Riemannian distance function

$$d_0 : N \times N \to [0, +\infty) \qquad [1]$$

given by $d_0(p, q) = \inf\{L(c);\ c : [0, 1] \to N$ is a piecewise smooth curve with $c(0) = p$ and $c(1) = q\}$. A fundamental result in global Riemannian geometry is the celebrated Hopf–Rinow theorem.

**Hopf–Rinow Theorem** *For any Riemannian manifold* $(N, g_0)$, *the following conditions are equivalent*:

(i) *metric completeness*: $(N, d_0)$ *is a complete metric space*;
(ii) *geodesic completeness*: *for any* $v$ *in* $TN$, *the geodesic* $c_v(t)$ *in* $N$ *with initial condition* $c_v'(0) = v$ *is defined for all values of an affine parameter* $t$;
(iii) *for some point* $p$ *in* $N$, *the exponential map* $\exp_p$ *is defined on all of* $T_pN$;
(iv) *finite compactness*: *every subset* $K$ *of* $N$ *that is* $d_0$ *bounded has compact closure*.
*Moreover, if any one of* (i)–(iv) *holds, then* $(N, g_0)$ *also satisfies*
(v) *minimal geodesic connectedness*: *given any* $p$, $q$ *in* $N$, *there exists a smooth geodesic segment* $c : [0, 1] \to N$ *with* $c(0) = p, c(1) = q$ *and* $L(c) = d_0(p, q)$.

A Riemannian metric for a smooth manifold is then said to be complete if it satisfies any of the above properties (i) through (iv). The Heine–Borel property of basic topology implies (via (iv)) that all Riemannian metrics for a compact manifold are automatically complete and many of the examples studied in basic Riemannian geometry are complete.

Also, if Riem($N$) denotes the space of all Riemannian metrics for a smooth manifold $N$, both geodesic completeness (property (ii) above) and geodesic incompleteness (the failure of property (ii) to hold for all geodesics) are $C^0$ stable properties on Riem($N$), that is, given a complete (resp., incomplete) metric $g$ for $N$, there exists an open neighborhood U($g$) of $g$ in Riem($N$) in the Whitney $C^0$ fine topology such that all Riemannian metrics $h$ in U($g$) are complete (resp., incomplete).

For spacetimes $(M, g)$, however, many basic examples furnished by general relativity fail to be geodesically complete and compactness of the underlying smooth manifold $M$ does not imply that the given Lorentz metric $g$ (let alone all Lorentz metrics for $M$) are complete. Also, the stability of geodesic completeness and incompleteness is more complicated than in the Riemannian case, necessitating concepts like pseudoconvex geodesic systems and disprisonment as studied by Beem and Parker. To summarize, for spacetimes and their associated Lorentzian distance functions, no naive analogs for the Hopf–Rinow theorem are valid. Under additional hypotheses, geodesic completeness may be guaranteed. Marsden noted that a compact spacetime with a homogenous Lorentz metric is geodesically complete. Then Carriere showed that a compact spacetime whose curvature tensor vanishes is geodesically complete. Later Kamishima (assuming constant curvature) and then Romero and Sanchez more generally showed that a compact Lorentzian manifold which admits a timelike Killing field is geodesically complete.

At any point $p$ in a given spacetime, emanating from $p$ are three families of geodesics: timelike, spacelike, and null. It was hoped in the 1960s that possibly continuity arguments could be obtained for different types of geodesic completeness. However, a series of examples showed by the mid-1970s that timelike geodesic completeness, null geodesic completeness, and spacelike geodesic completeness are logically inequivalent. (Here, a given geodesic is said to be complete if it may be extended to be defined for all values of an affine parameter.) Nomizu and Ozeki for Riemannian manifolds showed that any given Riemannian metric $g_0$ for the smooth manifold $N$ could be made geodesically complete by making a conformal change of metric $\Omega g_0$, where $\Omega : N \to (0, +\infty)$ is a smooth function. Especially in general relativity, such conformal changes are natural because the causal character of tangent vectors and curves (and hence of the basic causality conditions) are preserved. For spacetimes while generally nonspacelike geodesic completeness could not be produced by conformal changes, for some subclasses of spacetimes, such as the strongly causal ones, it was possible with a global conformal change.

For a large class of spacetimes, the warped or multiwarped products (originally inspired by several cosmological models in general relativity and a basic construction from Riemannian geometry), explicit integral criterion involving the warping functions have been given for timelike or null geodesic completeness. Several early examples of this type of result are discussed in Beem *et al.* (1996, pp. 111–112).

## Lorentz Distance and the Nonspacelike Cutlocus

For an arbitrary, not necessarily complete, Riemannian manifold $(N, g_0)$, the Riemannian distance function given in eqn [1] is continuous, the metric topology induced by $d_0$ coincides with the given manifold topology, and $d_0(p, q)$ is finite for all $p, q$ in $N$. Now, for an arbitrary spacetime $(M, g)$, and $p, q$ in $M$, if there is no future-directed nonspacelike curve from $p$ to $q$, set $d(p, q) = 0$; if there is such a curve, let

$$d(p, q) = \sup\{L(c);\ c : [0, 1] \to (M, g) \text{ is a piecewise smooth future-directed nonspacelike curve with } c(0) = p \text{ and } c(1) = q\} \quad [2]$$

(Unlike the Riemannian case, [2] does not bound $d(p, q)$ from above by $L(c)$ for any selected curve $c$ and hence the Lorentz distance may assume the value $+\infty$.)

This then defines what some authors term the "Lorentzian distance function"

$$d = d(g) : M \times M \to [0, +\infty] \quad [3]$$

and other authors term "proper time." It is linked to the causal structure of the given spacetime since

$$d(p, q) > 0 \text{ iff } q \text{ is in } I^+(p) \quad [4]$$

and in place of the triangle inequality for the Riemannian distance function, a reverse triangle inequality holds:

$$\text{if } p \le r \le q, \text{ then } d(p, q) \ge d(p, r) + d(r, q) \quad [5]$$

Also in the context of eqn [2], a future-directed nonspacelike curve $c : [0, 1] \to M$ from $c(0) = p$ to $c(1) = q$ is defined to be maximal if $L(c) = d(p, q)$. Corresponding to the Riemannian theory, a maximal nonspacelike curve turns out to be a smooth null or timelike geodesic segment.

As mentioned earlier, geodesic completeness is generally not a natural requirement to place on a spacetime. But what emerges from [4] in place of Riemannian completeness is an interplay between the causal properties of the given spacetime and the continuity (and other properties) of the Lorentzian distance function (cf. Beem *et al.* (1996, chapter 4)). At the extreme of totally vicious spacetimes, the Lorentz distance is always $+\infty$. Less drastically, if $(M, g)$ contains a closed timelike curve passing through $p$, then $d(p,q) = +\infty$ for all $q$ in $J^+(p)$. Also, certain cosmological models contain pairs of points at infinite distance. In general, Lorentzian distance is only lower semicontinuous. Adding upper semicontinuity forces a distinguishing spacetime to be causally continuous. A spacetime is chronological iff $d(p,p) = 0$ for all $p$ in $M$. At the other extreme from totally vicious spacetimes are globally hyperbolic spacetimes, which share many properties somewhat analogous to complete Riemannian manifolds. The Lorentzian distance function of a globally hyperbolic spacetime is both continuous and finite valued. (Indeed, a strongly causal spacetime is globally hyperbolic iff all Lorentz metrics $g'$ in the conformal class $C(M, g)$ also have finite-valued distance functions $d(g')$.) Second, corresponding to property (v) of the Hopf–Rinow Theorem, these spacetimes all satisfy maximal nonspacelike geodesic connectability: given any $p$, $q$ in $M$ with $p \le q$, there exists a future nonspacelike geodesic segment $c : [0,1] \to M$ with $c(0) = p$, $c(1) = q$ and $L(c) = d(p,q)$.

A basic concept from the calculus of variations is that of a pair of conjugate points along a geodesic segment $c : [0,a] \to (M,g)$. A smooth vector field $J(t)$ along $c$ is said to be a "Jacobi field" if $J$ satisfies the Jacobi differential equation

$$J'' + R(J,\, c')c' = 0 \qquad [6]$$

where $R$ denotes the curvature tensor. Then $c(t)$, $c(s)$ are said to be conjugate points along $c$ if there exists a nonzero Jacobi field $J$ along $c$ with $J(t) = J(s) = 0$. Much of the basic comparison techniques in global Riemannian geometry involving lengths of geodesics in manifolds satisfying curvature inequalities, such as the "Rauch comparison theorems," the "Toponogov triangle comparison theorem," and volume comparison theorems, were first obtained through Jacobi field techniques (cf. Petersen (1998) for a contemporary account). Later, Riccati equation techniques became more popular (cf. Karcher (1989)). For spacetimes, especially in the globally hyperbolic case, analogous results have been obtained for nonspacelike geodesic segments, with a key breakthrough in 1979 being Harris's version of the "Toponogov triangle comparison theorem" for timelike geodesic triangles in globally hyperbolic spacetimes. The Raychaudhuri equation used earlier in general relativity corresponds for spacetimes to this passage in the Riemannian setting from the Jacobi equation to the Riccati equation. The basic conjugate point theory and the Morse index theory for an arbitrary timelike or null geodesic segment in a general spacetime are reasonably close to the earlier Riemannian theory, if vector fields of the form $J(t) = f(t)\beta'(t)$ are accounted for in the case of a null geodesic segment $\beta : [0,1] \to (M,g)$. But spacelike geodesics and conjugate points are more problematic, as was first established using symplectic techniques by Helfer in 1994. More recently, progress has been made in applying important ideas of Gromov (1999) for Riemannian manifolds to the spacetime context (cf. Noldus (2004) for an example).

Inspired by fundamental concepts in global Riemannian geometry, Beem and Ehrlich in 1979 introduced the concept of nonspacelike cut point, again most tractable for globally hyperbolic spacetimes. Let $\gamma : [0,a) \to (M,g)$ be a future-inextendible, future-directed nonspacelike geodesic in an arbitrary spacetime. Define

$$t_0 = \sup\{t \in [0,\, a);\; d(\gamma(0),\, \gamma(t)) = L(\gamma|_{[0,\,t]})\} \qquad [7]$$

(If there is a closed timelike curve through $\gamma(0)$, then $d(\gamma(0), \gamma(0)) = +\infty$ and $t_0$ will not exist. If $\gamma$ is a nonspacelike geodesic ray and hence $d(\gamma(0), \gamma(t)) = L(\gamma|_{[0,t]})$ for all $t$, then $t_0 = a$.) However, if $0 < t_0 < a$, then $\gamma(t_0)$ is said to be the future nonspacelike cut point of $p = \gamma(0)$ along $\gamma$. For general spacetimes, it may be shown that:

1. for $0 < s < t < t_0$, that $\gamma|_{[s,t]}$ is the unique maximal nonspacelike geodesic in all of $(M,g)$ between $\gamma(s)$ and $\gamma(t)$;
2. $\gamma|_{[0,t]}$ is maximal for all $t$ with $0 \le t \le t_0$; and
3. for all $t$ with $t_0 < t < a$, there is a longer nonspacelike curve in $(M,g)$ than $\gamma|_{[0,t]}$ between $\gamma(0)$ and $\gamma(t)$.

A nonspacelike cut point is a subtler concept than a nonspacelike conjugate point since the existence of a cut point is not necessarily captured by the behavior of families of future nonspacelike curves (or geodesics) close to the given geodesic segment $\gamma$, the basic viewpoint of the calculus of variations. But since calculus of variations arguments shows that past a nonspacelike conjugate point, longer "neighboring curves" join $\gamma(0)$ to $\gamma(t)$, the future cut point of $p = \gamma(0)$ along $\gamma$ comes no later than the first

future conjugate point to $p$ along $\gamma$ in either the timelike or null geodesic case.

In a startling result which contradicted erroneous arguments in all the standard textbooks, Margerin in 1993 gave examples to show that even for compact Riemannian manifolds, the first conjugate locus of a point (i.e., the set of all first conjugate points along all geodesics issuing from a given point) need not be closed, even though elementary arguments correctly show that the cut locus of any point (i.e., the set of all cut points along all geodesics issuing from the given point) is always closed. The timelike first conjugate locus of a point in a spacetime will generally not be closed, but because a nonspacelike geodesic in a globally hyperbolic spacetime must escape from any compact subset in finite affine parameter, the future (or past) first nonspacelike conjugate locus of any point in such a spacetime is a closed subset. In a result analogous to the Riemannian characterization, nonspacelike cut points in globally hyperbolic spacetimes may be characterized as follows: let $q=\gamma(t_0)$ be the future cut point of $p=\gamma(0)$ along the timelike (resp., null) geodesic segment $\gamma$ from $p$ to $q$. Then either one of both of the following conditions hold: (1) $q$ is the first future conjugate point to $p$ along $\gamma$, or (2) there exist at least two maximal timelike (resp., null) geodesic segments from $p$ to $q$.

Now given $p$ in an arbitrary spacetime $(M, g)$, the future timelike (resp., null) cut locus of $p$ is defined to be the set of all timelike (resp., null) cut points along all future timelike (resp., null) geodesics issuing from $p$ and the future nonspacelike cut locus of $p$ is defined as the union of the future timelike and null cut loci. Employing alternatives (1) and (2) in the preceeding paragraph, it may be shown for globally hyperbolic spacetimes that the null and nonspacelike cut loci are closed subsets of $M$.

The null cut locus has a privileged status by virtue of a phenomena not encountered for Riemannian manifolds. Under a conformal change of back-ground spacetime metric, null geodesics remain null pregeodesics (i.e., may be reparametrized to be null geodesics in the deformed Lorentz metric) while such deformations fail to preserve timelike or spacelike geodesics, or to preserve geodesics in the Riemannian case. Even though null conjugate points along a null geodesic will not remain invariant under conformal change of spacetime metric, it is remarkable that elementary arguments involving the spacetime distance function show that global conformal diffeomorphisms do preserve null cut points and hence the null cut locus of any point.

## Geodesic Incompleteness and the Lorentzian Splitting Theorem

In global Riemannian geometry, an important concept is that of a geodesic ray. In a complete Riemannian manifold $(N, g_0)$, a unit geodesic $c:[0,+\infty)\to(N, g_0)$ is said to be a (geodesic) ray if $d_0(c(0), c(t))=t$ for all $t\geq 0$. By the triangle inequality, $c(t)$ is minimal between every pair of its points. By making a limit construction, it may be shown that for each $p$ in $N$, there exists a geodesic ray $c(t)$ with $c(0)=p$. An allied concept is that of a (geodesic) line $c:\mathbb{R}\to(N, g_0)$; here $d_0(c(t), c(s))=|t-s|$ for all $t, s$ is required, that is, $c$ is minimal between every pair of its points. The existence of a line is much stronger than the existence of a ray. If $(N, g_0)$ has positive Ricci curvature everywhere, then $(N, g_0)$ contains no lines despite the fact that it contains a ray issuing from each point. A helpful tool in this setting is the compactness of sets of tangent vectors of the form

$$\{w\in T_pN;\ g_0(w, w)=1\} \qquad [8]$$

for any $p$ in $N$; hence, any infinite sequence of tangent vectors based at $p$ automatically has a convergent subsequence.

For spacetimes, geodesic completeness cannot generally be assumed. Yet a future nonspacelike geodesic ray $\gamma:[0,b)\to(M,g)$ may be defined to be a future-directed, future-inextendible nonspacelike geodesic with $d(\gamma(0),\gamma(t))=L(\gamma|_{[0,t]})$ for all $t$ in $[0,b)$. The reverse triangle inequality implies that $\gamma$ is maximal between any pair of its points. Similarly, a nonspacelike geodesic line $\gamma:(a,b)\to(M,g)$ is a past- and future-inextendible nonspacelike geodesic with $d(\gamma(t),\gamma(s))=L(\gamma|_{[t,s]})$ for all $s$, $t$. Hence, $\gamma$ is maximal between any pair of its points. If nonspacelike geodesic completeness is assumed, $a=-\infty$ and $b=+\infty$ above. Constructions here are more delicate than in the Riemannian case because the sets

$$\{v\in T_pM;\ g(v, v)=-1\} \qquad [9]$$

of unit timelike tangent vectors, while closed in the tangent space, are noncompact. Despite this technicality, using the limit curve machinery of general relativity in place of the compactness in [8], it has been shown that a strongly causal spacetime admits a past and future nonspacelike geodesic ray issuing from every point (cf. Beem *et al.* (1996, chapter 8)). (If the spacetime is not nonspacelike geodesically complete, these rays will not necessarily be past or future complete.) As in the Riemannian case, the existence of a complete line is a stronger geometric condition. For that reason, in 1977 Beem and Ehrlich introduced the concept of a spacetime causally disconnected by a compact set $K$ and

showed that a strongly causal spacetime which is causally disconnected by a compact set contains a nonspacelike geodesic line which intersects the compact set. (Again, unless the spacetime is non-spacelike geodesically complete, this line need not be future or past complete.)

A pattern common to many results in global Riemannian geometry especially since the 1950s is the following: the existence of a complete Riemannian metric on a smooth manifold which also satisfies a global curvature inequality implies a topological or geometric conclusion. A celebrated early example from the 1950s and 1960s, obtained by separate results of Rauch, Berger, and Klingenberg, is the topological sphere theorem.

**Topological Sphere Theorem** *Suppose* $(N, g_0)$ *is a complete, simply connected Riemannian n-manifold whose sectional curvatures satisfy* $1/4 < K \leq 1$. *Then N is homeomorphic to* $S^n$.

By contrast, for spacetimes, the assumption of geodesic completeness is generally unwarranted. Here is an example of one of the celebrated singularity theorems of general relativity, published in 1970 as originally stated:

**Hawking–Penrose Singularity Theorem** *No spacetime* $(M, g)$ *of dimension* $n \geq 3$ *can satisfy all of the following three requirements together*:

(i) $(M, g)$ *contains no closed timelike curves*;
(ii) *Every inextendible nonspacelike geodesic in* $(M, g)$ *contains a pair of conjugate points*; *and*
(iii) *There exists a future- or past-trapped set S in* $(M, g)$.

This theorem may be reinterpreted more akin to the Riemannian pattern above as follows: suppose $(M, g)$ is a chronological spacetime of dimensions $n \geq 3$ which satisfies the timelike convergence condition ($\mathrm{Ric}(v, v) \geq 0$ for all timelike tangent vectors) and the generic condition (every inextendible nonspacelike geodesic contains a point which has some appropriate nonzero sectional curvature). If $(M, g)$ contains a future- or past-trapped set, then $(M, g)$ is nonspacelike geodesically incomplete. Hence, this result models the pattern: global curvature inequalities (reflecting the physical assumptions that gravity is assumed to be attractive and every inextendible nonspacelike geodesic experiences tidal acceleration) and a further physical or geometric assumption (the first and third conditions) implies the existence of an incomplete timelike or null geodesic.

An influential concept in global Riemannian geometry formulated during the 1960s and 1970s is that of curvature rigidity, which first became widely known through the introduction to the text Cheeger and Ebin (1975). The above statement of the "sphere theorem" contains one hypothesis that the sectional curvature is strictly greater than 1/4. In curvature rigidity, the hypothesis of strict inequality is relaxed to include the possibility of equality as well, and then one tries to show that either the old conclusion is still valid, or if it fails, it fails in an isometric (hence "rigid") manner. Thus in the example of the sphere theorem, if the sectional curvature is now allowed to satisfy $1/4 \leq K \leq 1$, then either the given Riemannian manifold remains homeomorphic to the $n$-sphere, or if not, it is isometric to a Riemannian symmetric space of rank 1.

Already in an article in 1970, Geroch had expressed the opinion that most spacetimes should be nonspacelike geodesically incomplete and also that a spacetime should fail to be nonspacelike geodesically incomplete only under special circumstances. Apparently by the early 1980s, S T Yau had formulated the idea that timelike geodesic incompleteness of spacetimes ought to display a curvature rigidity. In the paragraph following the statement of the Hawking–Penrose singularity theorem, there are two curvature conditions mentioned – the timelike convergence condition and the generic condition. Now the timelike convergence condition already allows for the case of equality (i.e., zero timelike Ricci curvature) in its formulation; hence, curvature rigidity here would imply dropping the generic condition that each inextendible nonspacelike geodesic contains a point of nonzero sectional curvatures as a hypothesis. This notion seems first to have been published by Yau's Ph.D. student R Bartnik in 1988 as follows:

**Conjecture** *Let* $(M, g)$ *be a spacetime of dimension* $\geq 3$ *which*

(i) *contains a compact Cauchy surface and*
(ii) *satisfies the timelike convergence condition* $Ric(v, v) \geq 0$ *for all timelike* $v$.

*Then either* $(M, g)$ *is timelike geodesically incomplete, or* $(M, g)$ *splits isometrically as a product* $(|\mathbb{R} \times V, -\mathrm{d}t^2 + h)$ *where* $(H, h)$ *is a compact Riemannian manifold.*

This conjecture has been proven in many cases with the following proof scheme. From the physical or geometric assumptions made, produce an inextendible nonspacelike geodesic line. Further, prove that the line happens to be timelike rather than null. Then if the spacetime were timelike geodesically complete, it would contain a complete

timelike line. But then the desired splitting may be obtained using the Lorentzian splitting theorem.

**Lorentzian Splitting Theorem** *Let* $(M, g)$ *be a spacetime of dimension* $\geq 3$ *which satisfies each of the following conditions*:

(i) $(M, g)$ *is either globally hyperbolic or timelike geodesically complete*;
(ii) $(M, g)$ *satisfies the timelike convergence condition*; *and*
(iii) $(M, g)$ *contains a complete timelike line*.

*Then* $(M, g)$ *splits isometrically as a product* $(\mathbb{R} \times V, -\mathrm{d}t^2 + h)$ *where* $(H, h)$ *is a complete Riemannian manifold*.

This result, which corresponds to obtaining the spacetime analog of a celebrated splitting theorem of Cheeger and Gromoll for lines in complete Riemannian manifolds of non-negative Ricci curvature, published in 1971, was posed as a problem by S T Yau in a problem list stemming from the conference Special Year in Differential Geometry held at the Institute for Advanced Study in Princeton during the 1979–80 academic year. Early progress was made using maximal hypersurface methods by Gerhardt in 1983, Bartnik in 1984, and Galloway in 1984. Then in 1985, Beem, Ehrlich, Markvorsen, and Galloway introduced the methodology of employing the Busemann function of the complete timelike line, motivated by techniques from Riemannian geometry, and succeeded in obtaining a splitting under the hypothesis of global hyperbolicity and everywhere nonpositive timelike sectional curvatures. In separate publications, Eschenburg and Galloway extended the result to the desired curvature hypothesis of nonnegative timelike Ricci curvatures. Finally, Newman in 1990 achieved the originally desired goal of obtaining the splitting under the assumption of timelike geodesic completeness, rather than global hyperbolicity. This is a more delicate setting, since timelike geodesic completeness does not imply maximal nonspacelike geodesic connectability, a fairly basic geometric tool in many standard constructions. But the idea emerged with Newman's solution that the existence of a timelike geodesic line or segment in a nonglobally hyperbolic spacetime implies an adequate level of control in a tubular neighborhood of the given line to enable the proof to work. Galloway and Horta in 1996 published a much simplified working out of these concepts. A fuller exposition of these developments may be found in Beem *et al.* (1996, chapter 14). In addition, in 2000, Galloway published a version of the splitting theorem for a null maximal geodesic line.

## Two-Dimensional Spacetimes

Two-dimensional spacetimes, sometimes termed Lorentz surfaces, are especially tractable because given $(M, g)$ with $\dim M = 2$, then $(M, -g)$ is also a spacetime. Hence, it may be shown that any Lorentzian 2-manifold $(M, g)$ homeomorphic to $\mathbb{R}^2$ may be made geodesically complete (not just nonspacelike geodesically complete) by a conformal change of metric. Also, any simply connected two-dimensional Lorentzian manifold is strongly causal. In Weinstein (1996), an extensive study is made of Lorentz surfaces generally and particularly, of a conformal boundary for such surfaces first given by Kulkarni in 1985.

One of the prettiest classical results linking the geometry and topology of a Riemannian surface is the Gauss–Bonnet theorem. Let $(N, g_0)$ be a Riemannian manifold of dimension 2 and let $P$ be a polygonal subregion with piecewise smooth bounding curves $c_i$, $1 \leq i \leq k$. Let $K$ denote the Gauss curvature of $(N, g_0)$ and $\kappa$ the geodesic curvature of the smooth curves $c_i$ (which vanishes if $c_i$ happens to be a geodesic). If $\alpha_i$ denote the corresponding interior angles between the successive boundary curves $c_i$ and $c_{i+1}$, then the Gauss–Bonnet formula over $P$ is

$$\int_P\int K\,\mathrm{d}A + \int_{\partial P} \kappa\,\mathrm{d}s + \sum_i (\pi - \alpha_i) = 2\pi \qquad [10]$$

By considering a triangulation of $N$ itself and summing up the corresponding terms in [10], it follows for a compact oriented Riemannian manifold $(N, g_0)$ of dimension 2 that

$$\int_N\int K\,\mathrm{d}A = 2\pi\chi_{(N)} \qquad [11]$$

where $\chi_{(N)}$ denotes the Euler characteristic. Also lurking in the background here is a formula for computing the angle between unit tangent vectors $v$, $w$ as

$$\cos\theta = g_0(v, w) \qquad [12]$$

In the spacetime setting, different versions of a Gauss–Bonnet formula for subregions of a two-dimensional spacetime $(M, g)$ corresponding to [10] have been given in 1974 by Helzer and in 1984 by Birman and Nomizu. First, the angle computation is a bit trickier for spacetimes than in the Riemannian case; eqn [12] has to be replaced by techniques which use the hyperbolic functions $\cosh u$ and $\sinh u$ to define the angle $u$ (sometimes called the "hyperbolic angle") between two unit vectors and

then to allow for null vectors. Birman and Nomizu obtained an analog of [10] assuming that the boundary curves for $P$ are successive smooth unit timelike curves:

$$\int_{\partial P} \kappa \, ds - \int_P \int K \, dA + \sum_i \theta_i = 0$$

Helzer in his formulation allows the different boundary curves to be either unit timelike, unit spacelike or null separately. Since the only compact, orientable smooth surface which admits a spacetime metric is the 2-torus, which has zero Euler characteristic, the Riemannian formula [11] above translates into the uniform constraint on the Gauss curvature of the spacetime:

$$\int_M \int K \, dA = 0$$

*See also:* General Relativity: Overview; Geometric Analysis and General Relativity; Pseudo-Riemannian Nilpotent Lie Groups; Spacetime Topology, Causal Structure and Singularities.

## Further Reading


Beem J, Ehrlich P, and Easley K (1996) Global Lorentzian Geometry. In: *Marcel Dekker Pure and Applied Mathematics*, vol. 202, 2nd edn. New York: Dekker.

Cheeger J and Ebin D (1975) Comparison Theorems in Riemannian Geometry, *North Holland Mathematical Library*. Amsterdam: North-Holland.

Einstein A (1916) Die Grundlage der allgemeinen Relativitätstheorie. *Annalen der Physik* 49: 769–822.

Gromov M (1999) *Metric Structures for Riemannian and Non-Riemannian Spaces*, Birkhauser Progress in Mathematics, vol. 152. Boston: Birkhauser.

Hawking S and Ellis G (1973) *The Large Scale Structure of Spacetime*. Cambridge: Cambridge University Press.

Karcher H (1989) Riemannian comparison constructions. In: Chern SS (ed.) *Global Differential Geometry*, Mathematical Association of America Studies in Mathematics, vol. 27, pp. 170–222. Washington, DC: MAA.

Kriele M (1999) *Spacetime: Foundations of General Relativity and Differential Geometry*, Springer Lecture Notes in Physics (Monographs), vol. 59. Heidelberg: Springer.

Noldus J (2004) The limit space of a Cauchy sequence of globally hyperbolic spacetimes. *Classical Quantum Gravity* 21: 851–874.

O'Neill B (1983) *Semi-Riemannian Geometry with Applications to Relativity*, Academic Press Pure and Applied Mathematics. New York: Academic Press.

Penrose R (1972) *Techniques of Differential Topology in Relativity*. Society for Industrial and Applied Mathematics, Regional Conference Series in Applied Mathematics, vol. 7. Philadelphia: SIAM.

Petersen P (1998) *Riemannian Geometry*, Springer Verlag Graduate Texts in Mathematics, vol. 171. New York: Springer.

Sachs R and Wu H (1977) *General Relativity for Mathematicians*, Springer Verlag Graduate Texts in Mathematics, vol. 48. New York: Springer.

Weinstein T (1996) *An Introduction to Lorentz Surfaces*, de Gruyter Expositions in Mathematics, vol. 22. Berlin: de Gruyter.


# Lyapunov Exponents and Strange Attractors

**M Viana**, IMPA, Rio de Janeiro, Brazil



## Lyapunov Exponents

The Lyapunov exponents of a sequence $\{A^n, n \geq 1\}$ of square matrices of dimension $d \geq 1$ are the values of

$$\lambda(v) = \limsup_{n \to \infty} \frac{1}{n} \log \|A^n \cdot v\| \qquad [1]$$

over all nonzero vectors $v \in \mathbb{R}^d$. For completeness, set $\lambda(0) = -\infty$. It is easy to see that $\lambda(cv) = \lambda(v)$ and $\lambda(v + v') \leq \max\{\lambda(v), \lambda(v')\}$ for any nonzero scalar $c$ and any vectors $v$, $v'$. It follows that, given any constant $a$, the set of vectors satisfying $\lambda(v) \leq a$ is a vector subspace. Consequently, there are at most $d$ Lyapunov exponents, henceforth denoted by $\lambda_1 < \cdots < \lambda_{k-1} < \lambda_k$, and there exists a filtration $F^1 < \cdots < F^{k-1} < F^k = \mathbb{R}^d$ into vector subspaces, such that

$$\lambda(v) = \lambda_i \quad \text{for all } v \in F_i \setminus F_{i-1}$$

and every $i = 1, \ldots, k$ (write $F_0 = \{0\}$). In particular, the largest exponent is

$$\lambda_k = \limsup_{n \to \infty} \frac{1}{n} \log \|A^n\| \qquad [2]$$

One calls $\dim F_i - \dim F_{i-1}$ the multiplicity of each Lyapunov exponent $\lambda_i$.

There are corresponding notions for continuous families of matrices $A^t, t \in (0, \infty)$, taking the limit as $t$ goes to $\infty$ in the relations [1] and [2]. The theories for the two types of families, discrete and continuous, are analogous and so at each point in what follows we refer to either one or the other.

## Lyapunov Stability

Consider the linear differential equation

$$\dot{v}(t) = B(t) \cdot v(t) \qquad [3]$$

where $B(t)$ is a bounded function with values in the space of $d \times d$ matrices, defined for all $t \in \mathbb{R}$. The theory of differential equations ensures that there exists a fundamental matrix $A^t, t \in \mathbb{R}$, such that

$$v(t) = A^t \cdot v_0$$

is the unique solution of [3] with initial condition $v(0) = v_0$.

If the Lyapunov exponents of the family $A^t, t > 0$, are all negative then the trivial solution $v(t) \equiv 0$ is asymptotically stable, and even exponentially stable. The stability theorem of Lyapunov asserts that, under an additional regularity condition, stability is still valid for nonlinear perturbations

$$w(t) = B(t) \cdot w + F(t, w) \qquad [4]$$

with $\|F(t,w)\| \leq \text{const.}\|w\|^{1+c}, c > 0$. That is, the trivial solution $w(t) \equiv 0$ is still exponentially asymptotically stable.

The regularity condition means, essentially, that the limit in [1] does exist, even if one replaces vectors $v$ by elements $v_1 \wedge \cdots \wedge v_l$ of any $l$th exterior power of $\mathbb{R}^d, 1 \leq l \leq d$. By definition, the norm of an $l$-vector $v_1 \wedge \cdots \wedge v_l$ is the volume of the parallelepiped determined by the vectors $v_1, \ldots, v_k$. This condition is usually tricky to check in specific situations. However, the multiplicative ergodic theorem of V I Oseledets asserts that, for very general matrix-valued stationary random processes, regularity is an almost sure property. This result sets the foundation for the modern theory of Lyapunov exponents. We are going to discuss the precise statement of the theorem in the slightly broader setting of linear cocycles, or vector bundle morphisms.

## Linear Cocycles

Let $\mu$ be a probability measure on some space $M$ and $f: M \to M$ be a measurable transformation that preserves $\mu$. Let $\pi: \mathcal{E} \to M$ be a finite-dimensional vector bundle, endowed with a Riemannian metric $\|\cdot\|_x$ on each fiber $\mathcal{E}_x = \pi^{-1}(x)$. Let $\mathcal{A}: \mathcal{E} \to \mathcal{E}$ be a linear cocycle over $f$. What we mean by this is that

$$\pi \circ \mathcal{A} = f \circ \pi$$

and the action $A(x): \mathcal{E}_x \to \mathcal{E}_{f(x)}$ of $\mathcal{A}$ on each fiber is a linear isomorphism. Notice that the action of the $n$th iterate $\mathcal{A}^n$ is given by

$$A^n(x) = A(f^{n-1}(x)) \cdots A(f(x)) \cdot A(x)$$

for every $n \geq 1$.

Assume the function $\log^+ \|A(x)\|_x$ is $\mu$-integrable:

$$\log^+ \|A(x)\|_x \in L^1(\mu) \qquad [5]$$

(we write $\log^+ \phi = \log \max\{\phi, 1\}$, for any $\phi > 0$). It is clear that the sequence of functions $a_n(x) = \log \|A^n(x)\|_x$ satisfies

$$a_{m+n}(x) \leq a_m(x) + a_n(f^m(x))$$

for every $m$, $n$, and $x$. It follows from J Kingman's subadditive ergodic theorem that the limit

$$\lim_{n\to\infty} \frac{1}{n} a_n(x)$$

exists for $\mu$-almost all $x$. In view of [2], this means that the largest Lyapunov exponent $\lambda_k(x)$ of the sequence $A^n(x), n \geq 1$ is a limit, and not just a lim sup, at almost every point.

## Multiplicative Ergodic Theorem

The Oseledets theorem states that the same holds for all Lyapunov exponents. Namely, for $\mu$-almost every $x \in M$ there exists $k = k(x) \in \{1, \ldots, d\}$, a filtration

$$F_x^1 < \cdots < F_x^{k-1} < F_x^k = \mathcal{E}_x$$

and numbers $\lambda_1(x) < \cdots < \lambda_k(x)$ such that

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n(x)\|_x = \lambda_i(x) \qquad [6]$$

for all $v \in F_x^i \setminus F_x^{i-1}$ and $i \in \{1, \ldots, k\}$.

The Lyapunov exponents $\lambda_i(x)$, and their number $k(x)$, are measurable functions of $x$ and they are constant on orbits of the transformation $f$. In particular, if the measure $\mu$ is ergodic then $k$ and the $\lambda_i$ are constant on a full $\mu$-measure set of points. The subspaces $F_x^i$ also depend measurably on the point $x$ and are invariant under the linear cocycle:

$$A(x) \cdot F_x^i = F_{f(x)}^i$$

It is in the nature of things that, usually, these objects are not defined everywhere and they depend discontinuously on the base point $x$.

When the transformation $f$ is invertible, one obtains a stronger conclusion, by applying the previous kind of result also to the inverse of the cocycle. Namely, assuming that $\log^+ \|A^{-1}\|$ is also in $L^1(\mu)$, one gets that there exists a decomposition

$$\mathcal{E}_x = E_x^1 \oplus \cdots \oplus E_x^k$$

defined at almost every point and such that $A(x) \cdot E_x^i = E_{f(x)}^i$ and

$$\lim_{n \to \pm\infty} \frac{1}{n} \log \|A^n(x)\|_x = \lambda_i(x) \qquad [7]$$

for all $v \in E_x^i$ different from zero and all $i \in \{1, \ldots, k\}$. These Oseledets subspaces $E_x^i$ are related to the subspaces $F_x^i$ through

$$F_x^j = \bigoplus_{i=1}^{j} E_x^i$$

Hence, $\dim E_x^i = \dim F_x^i - \dim F_x^{i-1}$ is the multiplicity of the Lyapunov exponent $\lambda_i(x)$.

The angles between any two Oseledets subspaces decay subexponentially along orbits of $f$:

$$\lim_{n \to \pm\infty} \frac{1}{n} \log \text{angle}\left(E_{f^n(x)}^i, E_{f^n(x)}^j\right) = 0$$

for every $i \neq j$ and almost every point. These facts imply the regularity condition mentioned previously and, in particular,

$$\lim_{n \to \pm\infty} \frac{1}{n} \log |\det A^n(x)| = \sum_{i=1}^{k} \lambda_i(x) \dim E_x^i \qquad [8]$$

Consequently, for cocycles with values in $\mathrm{SL}(d, \mathbb{R})$, the sum of all Lyapunov exponents, counted with multiplicity, is identically zero.

As we are dealing with almost certain properties, we may generally restrict the vector bundle to some full measure subset over which it is trivial. Then each fiber $\mathcal{E}_x$ is identified with the space $\mathbb{R}^d$, and we may think of $A(x)$ as a $d \times d$ matrix. Then $A_n(x) = A(f^n(x))$ is a stationary random process relative to $(f, \mu)$. Thus, in this context it is no serious restriction to view a linear cocycle as a stationary random process with values in the linear group $\mathrm{GL}(d, \mathbb{R})$ of invertible $d \times d$ matrices.

Furthermore, given any such random process $A_n, n \geq 0$, one may consider its normalization $B_n = A_n / |\det A_n|$. The Lyapunov exponents of the two random processes $A_n, n \geq 0$, and $B_n, n \geq 0$, differ by the time average

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} \log |\det A_j(x)|$$

of the determinant. The Birkhoff ergodic theorem ensures that the time average is well defined almost everywhere, as long as the function $\log |\det A|$ is in $L^1(\mu)$; this is the case, for instance, if both $\log^+ \|A^{\pm 1}\|$ are integrable. This relates the general case to random processes with values in the special linear group $\mathrm{SL}(d, \mathbb{R})$ of $d \times d$ matrices with determinant $\pm 1$.

The Oseledets theorem was extended by D Ruelle to certain linear cocycles in infinite dimensions. He assumes that the $A(x)$ are compact operators on a Hilbert space $H$ and $\log^+ \|A\|$ is in $L^1(\mu)$. The conclusion is the same as in finite dimensions, except that the filtration

$$\cdots < F_x^i < \cdots < F_x^1 = H$$

may involve infinitely many subspaces, and the Lyapunov exponents may be $-\infty$. There is also a version for cocycles over invertible transformations, where one assumes each $A(x)$ to be invertible and the sum of a unitary operator with a compact operator, such that both $\log \|A^{\pm}\|$ are integrable. The conclusion is that there exists an Oseledets decomposition $H = E_x^1 \oplus \cdots \oplus E_x^i \oplus \cdots$ at almost every point, with finitely or countably many factors.

## Random Matrices

Relation [8] implies that, for $\mathrm{SL}(d, \mathbb{R})$ cocycles, if there is only one Lyapunov exponent (with full multiplicity) then it must be zero. When this happens, the theory contains no information on the behavior of the iterates $A^n(x) \cdot v$, apart from the fact that there is no exponential growth nor decay of their norms. Thus, the question naturally arises under which conditions is there more than one Lyapunov exponent or, equivalently, under which conditions is the largest Lyapunov exponent strictly positive.

This problem was first addressed by H Furstenberg for products of independent random variables, corresponding to the following class of linear cocycles. Let $\nu$ be a probability measure on the group $G = \mathrm{GL}(d, \mathbb{R})$. Consider $M = G^{\mathbb{N}}$ and $\mu = \nu^{\mathbb{N}}$ (or $M = G^{\mathbb{Z}}$ and $\mu = \nu^{\mathbb{Z}}$), and let $f: M \to M$ be the shift map

$$f\left((\alpha_j)_j\right) = (\alpha_{j+1})_j$$

It is clear that $\mu$ is invariant and also ergodic for the transformation $f$. Consider the cocycle $\mathcal{A}: \mathcal{E} \to \mathcal{E}$ defined by $\mathcal{E} = M \times \mathbb{R}^d$ and

$$A\left((\alpha_j)_j\right) \cdot v = \alpha_0 \cdot v$$

Clearly,

$$A^n\left((\alpha_j)_j\right) \cdot v = \alpha_{n-1} \cdots \alpha_1 \alpha_0 \cdot v$$

Corresponding to the hypothesis of the multiplicative ergodic theorem, assume that $\log^+ \|\alpha\|$ (and $\log^+ \|\alpha^{-1}\|$) are $\nu$-integrable functions of the matrix $\alpha$.

Furstenberg's theorem states that if the closed group $G(\nu)$ generated by the support of $\nu$ is

noncompact and strongly irreducible in $\mathbb{R}^d$ then the largest Lyapunov exponent of the cocycle $\mathcal{A}$ is strictly positive. Strong irreducibility means that there exists no finite union of subspaces of $\mathbb{R}^d$ that is invariant under all elements of the group. Improvements, extensions, and alternative proofs have been obtained by several authors since then.

Especially, Y Guivarc'h and A Raugi provided conditions under which there are exactly $d$ distinct Lyapunov exponents or, in other words, the multiplicity of every Lyapunov exponent is equal to 1. A matrix semigroup has the contraction property if there exists a sequence of elements $h_n$ and a probability measure on the projective space of $\mathbb{R}^d$ that gives zero weight to any projective subspace, such that the images $(h_n)_* m$ of $m$ under the $h_n$ converge to a Dirac mass in the projective space. They proved that if the closed semigroup $H(\nu)$ generated by the support of the probability $\nu$ is strongly irreducible and has the contraction property then the largest Lyapunov exponent has multiplicity 1. Applying this to the exterior powers of the cocycle, one obtains sufficient conditions for simplicity of the other Lyapunov exponents as well.

This statement has been improved by I Ya Gol'dsheid and G A Margulis, who formulated the hypotheses in terms of the algebraic closure $\tilde{G}(\nu)$ of the semigroup $H(\nu)$. They assumed that $\tilde{G}(\nu)$ has the contraction property and the connected component of the identity inside $\tilde{G}(\nu)$ is irreducible in $\mathbb{R}^d$, meaning that its elements do not have any common invariant subspace. Then the largest Lyapunov exponent is simple.

## Schrödinger Cocycles

The one-dimensional discrete Schrödinger equation is the second-order difference equation

$$-(u_{n+1} + u_{n-1}) + V_n u_n = E u_n \qquad [9]$$

derived from the stationary Schrödinger equation in dimension 1 by space discretization. Here the energy E is a constant and $V_n = V(f^n(\theta))$, where the potential $V(\cdot)$ is a bounded scalar function and $f: M \to M$ is a transformation preserving some probability measure $\mu$ on $M$. In what follows, we take $\mu$ to be ergodic. Equation [9] may be rewritten as a first-order relation,

$$\begin{pmatrix} u_{n+1} \\ v_{n+1} \end{pmatrix} = \begin{pmatrix} V_n - E & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} u_n \\ v_n \end{pmatrix}$$

Hence, it may also be interpreted as a linear cocycle $\mathcal{A}$ over $f$, where the vector bundle is $\mathcal{E} = M \times \mathbb{R}^2$ and

$$A(\theta) = \begin{pmatrix} V(\theta) - E & -1 \\ 1 & 0 \end{pmatrix} \qquad [10]$$

takes values in $\mathrm{SL}(\mathbb{R}, 2)$. By ergodicity, the Lyapunov exponents are essentially independent of the base point $\theta$. Let $\lambda(E)$ denote the largest exponent: by the relation [8], the other one is $-\lambda(E)$.

The Lyapunov exponent $\lambda(E)$ is related to the spectral theory of the linear operators $\mathcal{L}_\theta$,

$$(\mathcal{L}_\theta u)_n = -(u_{n+1} + u_{n-1}) + V_n u_n$$

on the space $\ell^2(\mathbb{Z})$ of complex square-integrable sequences $u_n$, $n \in \mathbb{Z}$. These are bounded Hermitian operators and so the spectra are compact subsets of $\mathbb{R}$. Using the assumption that $\mu$ is ergodic, one can prove that the spectrum $\mathrm{spec}(\mathcal{L}_\theta)$ is constant almost everywhere. If the transformation $f$ is minimal, the spectrum is even independent of the point $\theta$. Moreover, for all energies,

$$\lambda(E) \geq \text{const. } \mathrm{dist}(E, \mathrm{spec}(\mathcal{L}_\theta))$$

In particular, $\lambda(E)$ is always positive on the complement of the spectrum.

A fundamental problem (Anderson localization) is to decide when the spectrum is pure-point. This is reasonably well understood for a few classes of base dynamics only, for example, the very chaotic systems such as Bernoulli and Markov processes (random potentials) or uniformly hyperbolic maps and flows, or the irrational rotations on the $d$-dimensional torus (quasiperiodic potentials). In the latter case, the results are more complete when there is only one frequency ($d = 1$). It was shown by K Ishii and by L Pastur that if $\lambda(E)$ is positive for almost all values of $E$ in some Borel set then the absolutely continuous part of the spectrum is essentially disjoint from that set. The converse is also true (due to S Kotani). Thus, checking that $\lambda(E)$ is positive is an important step towards proving localization.

A very general criterion for positivity of the Lyapunov exponent was obtained by Kotani. Namely, he proved that if the potential is not deterministic then $\lambda(E)$ is positive for almost all $E$. In particular, for nondeterministic potentials the absolutely continuous spectrum is empty, almost surely. In simple terms, the hypothesis means that from the values of the potential for negative $n$ one cannot determine the values for positive $n$. More formally, one calls the potential deterministic if every $V_n, n \geq 0$ is almost everywhere a measurable function of $\{V_n : n \leq 0\}$. For instance, quasiperiodic potentials are deterministic, whereas Bernoulli potentials are not.

## Subharmonicity Method

Let $\mathbb{D}^m$ be the set of complex vectors $(z_1, \ldots, x_m) \in \mathbb{C}^m$ such that $|z_j| \le 1$ for all $j$ and let $\mathbb{T}^m$ be the subset defined by $|z_j| = 1$ for all $j$. Let $f : \mathbb{T}^m \to \mathbb{T}^m$ and $A : \mathbb{T}^m \to \mathrm{SL}(d, \mathbb{R})$ be continuous maps that admit holomorphic extensions to the interior of $\mathbb{D}^m$ with $f(0) = 0$. Assume that $f$ preserves the natural (Haar) measure $\mu$ on $\mathbb{T}^m$. Let

$$\lambda(A, \mu) = \int_{\mathbb{T}^m} \lambda(z) \mathrm{d}\mu$$

where $\lambda(z)$ denotes the largest Lyapunov exponent for the cocycle defined by $A$ over $f$. It also follows from the subadditive ergodic theorem that

$$\lambda(A, \mu) = \lim \frac{1}{n} \int_{\mathbb{T}^m} \log \|A^n(z)\| \mathrm{d}\mu$$

M Herman observed that, since the function $\log \|A^n(z)\|$ is plurisubharmonic on $\mathbb{D}^m$, one may use the maximum principle to conclude that

$$\frac{1}{n} \int_{\mathbb{T}^m} \log \|A^n(z)\| \mathrm{d}\mu \ge \frac{1}{n} \log \|A^n(0)\|$$

Then, taking the limit when $n \to \infty$ one obtains that

$$\lambda(A, \mu) \ge \rho(A) \qquad [11]$$

where $\rho(A)$ denotes the spectral radius of the matrix $A(0)$. Starting from this observation, he developed a very effective method for bounding Lyapunov exponents from below, that received several applications and extensions, in particular, to the theory of Schrödinger cocycles with quasiperiodic potentials.

The best-known application is the following bound for integrated Lyapunov exponents of two-dimensional cocycles. Let $f : M \to M$ be a continuous transformation on a compact metric space, preserving some probability measure $\mu$, and $A : M \to \mathrm{SL}(2, \mathbb{R})$ be a continuous map. For each fixed $\theta$, let $AR_\theta$ be the cocycle obtained by multiplying $A(x)$, at every point $x$, by the rotation of angle $\theta$. Herman proved that

$$\frac{1}{2\pi} \int \lambda(AR_\theta, \mu) \mathrm{d}\theta \ge \int_M N(x) \, \mathrm{d}\mu$$

(A Avila and J Bochi later showed that the equality holds) where

$$N(x) = \log \frac{\|A(x)\| + \|A(x)^{-1}\|}{2}$$

Apart from the exceptional case when $A$ acts by rotation at every point in the support of $\mu$, the right-hand side of the inequality is positive, and so the Lyapunov exponent of the cocycle $AR_\theta$ is positive for many values of $\theta$.

## Nonuniform Hyperbolicity

The prototypical example of a linear cocycle is the derivative of a smooth transformation on a manifold. More precisely, let $M$ be a finite-dimensional manifold and $f : M \to M$ be a diffeomorphism, that is, a bijective smooth map whose derivative $Df(x)$ depends continuously on $x$ and is an isomorphism at every point. Let $\mathcal{E} = TM$ be the tangent bundle to the manifold and $\mathcal{A} = Df$ be the derivative. If $M$ is compact or, more generally, if the norms of both $Df$ and its inverse are bounded, then the hypothesis in Oseledets theorem is automatically satisfied for any $f$-invariant probability $\mu$. Lyapunov exponents yield deep geometric information on the dynamics of the diffeomorphism, especially when they do not vanish. For most results that we mention in the sequel, one needs the derivative $Df$ to be Hölder continuous:

$$\|Df(x) - Df(y)\| \le \text{const.}\, d(x, y)^{\mathrm{c}}$$

Let $E^{\mathrm{s}}_x$ be the sum of the Oseledets subspaces corresponding to negative Lyapunov exponents. Pesin's stable manifold theorem states that there exists a family of embedded disks $W^{\mathrm{s}}_{\mathrm{loc}}(x)$ tangent to $E^{\mathrm{s}}_x$ at almost every point and such that the orbit of every $y \in W^{\mathrm{s}}_{\mathrm{loc}}(x)$ is exponentially asymptotic to the orbit of $x$. This lamination $\{W^{\mathrm{s}}(x)\}$ is invariant, in the sense that

$$f(W^{\mathrm{s}}(x)) \subset W^{\mathrm{s}}(f(x))$$

and has an "absolute continuity" property. There are analogous results for the sum $E^{\mathrm{u}}_x$ of the Oseledets subspaces corresponding to positive Lyapunov exponents.

The entropy of a partition $\mathcal{P}$ of $M$ is defined by

$$h_\mu(f, \mathcal{P}) = \lim_{n \to \infty} \frac{1}{n} H_\mu(\mathcal{P}^n)$$

where $\mathcal{P}^n$ is the partition into sets of the form $P = P_0 \cap f^{-1}(P_1) \cap \cdots \cap f^{-n}(P_n)$ with $P_j \in \mathcal{P}$ and

$$H_\mu(\mathcal{P}^n) = \sum_{P \in \mathcal{P}^n} -\mu(P) \log \mu(P)$$

The Kolmogorov–Sinai entropy $h_\mu(f)$ of the system is the supremum of $h_\mu(f, \mathcal{P})$ over all partitions $\mathcal{P}$ with finite entropy. The Ruelle–Margulis inequality says that $h_\mu(f)$ is bounded above by the average sum of the positive Lyapunov exponents. A major result of the theory, Pesin's entropy formula, asserts that if the invariant measure $\mu$ is smooth (e.g., a volume element) then the two invariants coincide:

$$h_\mu(f) = \int \left( \sum_{j=1}^{k} \lambda_j^+ \right) \mathrm{d}\mu$$

A complete characterization of the invariant measures for which the entropy formula is true was given by F Ledrappier and L S Young.

The invariant measure $\mu$ is called hyperbolic if all Lyapunov exponents are nonzero at almost every point. Hyperbolic measures are exact dimensional: the pointwise dimension

$$d(x) = \lim_{r \to 0} \frac{\log \mu(B_r(x))}{\log r}$$

exists at almost every point, where $B_r(x)$ is the neighborhood of radius $r$ around $x$. This fact was proved by L Barreira, Ya Pesin, and J Schmeling. Note that it means that the measure $\mu(B_r(x))$ of neighborhoods scales as $r^{d(x)}$ when the radius $r$ is small.

Another remarkable feature of hyperbolic measures, proved by A Katok, is that periodic motions are dense in their supports. More than that, assuming the measure is nonatomic, there exist Smale horseshoes $H_n$ with topological entropy arbitrarily close to the entropy $h_\mu(f)$ of the system. In this context, the topological entropy $h(f, H_n)$ may be defined as the exponential rate of growth,

$$\lim_{k \to \infty} \frac{1}{k} \log \#\{x \in H_n : f^k(x) = x\}$$

of the number of periodic points on $H_n$.

## Generic Systems

Given any area-preserving diffeomorphism on any surface $M$, one may find another whose first derivative is arbitrarily close to the initial one and which has Lyapunov exponents identically zero at almost every point, or else is globally uniformly hyperbolic (Anosov). This surprising fact was discovered by R Mañé, and a complete proof was given by J Bochi. Uniform hyperbolicity means that the tangent bundle admits a $Df$-invariant splitting

$$TM = E^{\mathrm{s}} \oplus E^{\mathrm{u}}$$

such that the line bundle $E^{\mathrm{s}}$ is uniformly contracted and $E^{\mathrm{u}}$ is uniformly expanded by the derivative. It is well known that Anosov diffeomorphisms can only occur if the surface is the torus $\mathbb{T}^2$.

In fact, the theorem of Mañé–Bochi is stronger: for a residual subset (a countable intersection of open dense sets) of all once-differentiable area-preserving diffeomorphisms on any surface, either the Lyapunov exponents vanish almost everywhere or the diffeomorphism is Anosov. This shows that zero Lyapunov exponents are actually quite common for surface diffeomorphisms that are only once-differentiable. Moreover, this theorem has been extended to diffeomorphisms on manifolds with arbitrary dimension, in a suitable formulation, by J Bochi and M Viana.

However, this phenomenon should be specific to systems with low differentiability. Indeed, already for Hölder-continuous linear cocycles over chaotic transformations it is known that vanishing Lyapunov exponents can only occur with infinite codimension. That is, unless the cocycle satisfies an infinite number of independent constraints, there exists some positive exponent. By "chaotic" we mean here that the invariant probability $\mu$ of the base transformation is assumed to be hyperbolic and to have local product structure: it is locally equivalent to a product of two measures, respectively, along stable and unstable sets.

Under additional assumptions, one can even prove that all Lyapunov exponents have multiplicity 1 outside an infinite-codimension subset. This follows from extensions of the Guivarc'h–Raugi criterion for certain linear cocycles over chaotic transformations, obtained by A Avila, C Bonatti, and M Viana.

## Strange Attractors

This expression was coined by D Ruelle and F Takens in their celebrated study on the nature of fluid turbulence. E Hopf and also L D Landau and E M Lifshitz had suggested that turbulent motion arises from the existence in the phase space of invariant tori carrying quasiperiodic flows with large number of frequencies. Ruelle and Takens observed that dissipative systems such as viscous fluids do not generally have such quasiperiodic tori, and concluded that turbulence must be credited to a different mechanism: the presence of some "strange" attractor.

While they did not propose a precise definition, two main features were mentioned:

1. *Complex geometry*: a strange attractor is not reduced to an equilibrium point or a periodic solution of the system and, generally, should have a fractal structure.
2. *Chaotic dynamics*: solutions accumulating on the attractor should be sensitive to their initial states.

As more examples were found, it became apparent that the above two features do not always come together. This led to two types of definitions in the literature, depending on whether one emphasizes the geometry or the dynamics. We adopt the second point of view, and propose to define the strange attractor as one carrying an invariant ergodic physical measure which has some positive Lyapunov exponent. The notion of physical measure will be

defined near the end. The condition on the Lyapunov exponent ensures that the dynamics near the attractor is (exponentially) sensitive to the initial states.

## Lorenz-Like Attractors

The uniformly hyperbolic attractors introduced by S Smale provided an interesting class of examples of strange attractors, both chaotic and fractal. Perhaps more striking, given that they originated from a concrete problem in fluid dynamics, were the strange attractors introduced by E N Lorenz. The Lorenz system of differential equations,

$$\begin{aligned} \dot{x} &= -\sigma x + \sigma y, & \sigma &= 10 \\ \dot{y} &= rx - y - xz, & r &= 28 \\ \dot{z} &= xy - bz, & b &= 8/3 \end{aligned} \qquad [12]$$

was derived from Lord Rayleigh's model for thermal convection, by Fourier expansion of the stream function and temperature, and truncation of all but three modes. Lorenz observed that its solutions depend sensitively on their initial states. Consequently, predictions based on the numerical integration of the equations may turn out to be very inaccurate, given that the initial data obtained from experimental measurements are never completely precise. This remarkable observation brought the issue of predictability in deterministic systems to a whole new light and motivated intense investigation of this and many other chaotic systems.

The dynamical behavior of the eqns [12] was first interpreted through certain geometric models where the presence of strange attractors, both chaotic and fractal, could be proved rigorously. It was much harder to prove that the original eqns [12] themselves have such an attractor. This was achieved just a few years ago, by W Tucker, by means of a computer-assisted rigorous argument. At about the same time, a mathematical theory of Lorenz-like attractors in three-dimensional space was developed by C Morales, M J Pacifico, and E Pujals. In particular, this theory shows that uniformly hyperbolic attractors and Lorenz-like attractors are the only ones which are robust under all small modifications of the vector field.

## Hénon-Like Attractors

Starting from the work of Lorenz, many models of strange attractors have been found and described to some extent, often related to concrete problems. From a mathematical point of view, it is usually hard to give even a rough description of the dynamics in the chaotic regime. However, this was especially successful for the family of strange attractors introduced by M Hénon. He considered a very simple nonlinear system, particularly suited for numerical experimentation: the transformation

$$f(x, y) = (1 - ax^2 + by, x) \qquad [13]$$

where $a$ and $b$ are constant parameters. In a breakthrough, M Benedicks and L Carleson were able to prove that, for a set of parameter values with positive probability, this transformation has some nonhyperbolic attractor such that the orbits accumulating on it are sensitive to the starting point. The system [13] is also a model for many other situations, including the phenomenon of creation of homoclinic motions as parameters unfold, and the conclusions of Benedicks and Carleson have been extended to such situations, starting from the work of L Mora and M Viana.

Moreover, a detailed theory of Hénon-like attractors has been developed by M Benedicks, M Viana, D Wang, L S Young, and other authors. It follows from this theory that these attractors carry an invariant ergodic probability measure $\mu$ which describes the statistical behavior of almost all trajectories $f^j(x)$, $j \geq 1$, that accumulate the attractor:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \varphi(f^j(x)) = \int \varphi \, d\mu$$

for any continuous function $\varphi$. This property implies that, despite the fact that it is supported on a zero-volume set, the measure $\mu$ is, in some sense, physically observable. For this reason, one calls it a physical measure. In other words, time averages along typical orbits in the domain of attraction coincide with the space averages determined by the probability $\mu$. Another property with physical relevance is that $\mu$ is the zero-noise limit of the stationary measures associated to the Markov chains obtained by adding random noise to $f$. One says that the system $(f, \mu)$ is stochastically stable.

*See also:* Chaos and Attractors; Dissipative Dynamical Systems of Infinite Dimension; Ergodic Theory; Fractal Dimensions in Dynamics; Generic Properties of Dynamical Systems; Gravitational *N*-Body Problem (Classical); Homoclinic Phenomena; Hyperbolic Dynamical Systems; Lagrangian Dispersion (Passive Scalar); Nonequilibrium Statistical Mechanics: Interaction between Theory and Numerical Simulations; Random Dynamical Systems; Synchronization of Chaos.

## Further Reading

Barreira L and Pesin Ya (2002) Lyapunov Exponents and Smooth Ergodic Theory, *Univ. Lecture Series*, vol. 23. American Mathematical Society.

Bochi J and Viana M (2004) Lyapunov exponents: how frequently are dynamical systems hyperbolic? In: *Modern Dynamical Systems and Applications*, pp. 271–297. Cambridge: Cambridge University Press.

Bonatti C, Díaz LJ, and Viana M (2004) *Dynamics Beyond Uniform Hyperbolicity: A Global Geometric and Probabilistic Perspective*, *Encyclopedia of Mathematical Sciences*, vol. 102. Springer.

Eckmann J-P and Ruelle D (1985) Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics* 57: 617–656.

Gol'dsheid IYa and Margulis GA (1989) Lyapunov indices of a product of random matrices. *Uspekhi Mat. Nauk*. 44: 13–60.

Mañé R (1987) *Ergodic Theory and Differentiable Dynamics*. Springer.

Spencer T (1990) Ergodic Schrödinger operators *Analysis, et cetera* 623–637. Academic Press.

Viana M (2000) What's new on Lorenz strange attractors? *Math. Intelligencer* 22: 6–19.

# M

## Macroscopic Fluctuations and Thermodynamic Functionals

**G Jona-Lasinio**, Università di Roma "La Sapienza," Rome, Italy



### Introduction

There is no theory so far of irreversible processes that is of the same generality as equilibrium statistical mechanics and presumably it may not exist. While in equilibrium the Gibbs distribution provides all the information and no equation of motion has to be solved, the dynamics plays the major role in nonequilibrium. The theory illustrated below refers to stationary states that are not restricted to being close to equilibrium, and for a wide class of models it can be shown to be exact. In this case one begins to see the appearance of some general principles.

In equilibrium statistical mechanics, there is a well-defined relationship, established by Boltzmann, between the probability of a state and its entropy. This fact was exploited by Einstein to study thermodynamic fluctuations. When we are out of equilibrium, for example, in a stationary state of a system in contact with two reservoirs, it is not completely clear how to define thermodynamic quantities such as the entropy or the free energy. One possibility is to use fluctuation theory to define their nonequilibrium analogs. In fact in this way, extensive quantities can be obtained, although not necessarily simply additive due to the presence of long-range correlations which seem to be a rather generic feature of nonequilibrium. This possibility has been pursued in recent years leading to a considerable number of interesting results. One can recognize two main lines.

1. Exact calculations in simplified models. This is well exemplified by the work of Derrida *et al.* (2002).
2. A general treatment of a class of continuous time Markov chains for which the simplified models provide examples. This is the point of view developed by Bertini *et al.* (2002, 2004).

Both approaches have been very effective and of course give the same results when a comparison is possible. The second approach seems to encompass a wide class of systems and has the advantage of leading to equations which apply to very different situations. This is the point of view we shall adopt in the following. The question whether there are alternative more natural ways of defining nonequilibrium entropies or free energies is, for the moment, open.

### Boltzmann–Einstein Formula

The Boltzmann–Einstein theory of equilibrium thermodynamic fluctuations, as described for example in the book *Physique Statistique* by Landau–Lifshitz, states that the probability of a fluctuation from equilibrium in a macroscopic region of fixed volume $V$ is proportional to $\exp\{V\Delta S/k\}$, where $\Delta S$ is the variation of entropy density in the region calculated along a reversible transformation creating the fluctuation and $k$ is the Boltzmann constant.

This formula was derived by Einstein simply by inverting the Boltzmann relationship between entropy and probability. He considered this relationship as a phenomenological definition of the probability of a state.

Einstein theory refers to fluctuations from an equilibrium state, that is from a stationary state of a system isolated or in contact with reservoirs characterized by the same chemical potentials so that there is no flow of heat, electricity, chemical substances, etc., across the system. When in contact with reservoirs, $\Delta S$ is the variation of the total entropy (system + reservoirs) which, for fluctuations of constant volume and temperature, is equal to $-\Delta F/T$, where $\Delta F$ is the variation of the free energy of the system and $T$ the temperature. In the following, we refer to $\Delta F/T$, our main object of study, as the entropy and use the letter $S$ for it but no confusion should arise.

The important question we address is then: what happens if the system is stationary but not in equilibrium, that is, flows of physical quantities are present due to external fields and/or different chemical potentials at the boundaries? To start with it is not always clear whether a closed macroscopic dynamical description is possible. If the system admits such a description of the kind provided by hydrodynamic

equations, a fact which can be rigorously established in simplified models, a reasonable goal is to find an explicit connection between time-independent thermodynamic quantities (e.g., the entropy) and dynamical macroscopic properties (e.g., transport coefficients). As we shall see, the study of large fluctuations provides such a connection. It leads in fact to a dynamical theory of the entropy which is shown to satisfy a Hamilton–Jacobi equation (HJE) in infinitely many variables requiring the transport coefficients as input. Its solution is straightforward in the case of homogeneous equilibrium states and highly nontrivial in stationary nonequilibrium states (SNSs). In the first case we recover a well-known relationship widely used in the physical and physico-chemical literature. There are several one-dimensional models, where the HJE reduces to a nonlinear ordinary differential equation which, even if it cannot be solved explicitly, leads to the important conclusion that the nonequilibrium entropy is a nonlocal functional of the thermodynamic variables. This implies that correlations over macroscopic scales are present. The existence of long-range correlations is probably a generic feature of SNSs and more generally of situations where the dynamics is not time-reversal invariant. As a consequence if we divide a system into two subsystems, the entropy is not necessarily simply additive.

The first step toward the definition of a nonequilibrium entropy is the study of fluctuations in macroscopic evolutions described by hydrodynamic equations. In a dynamical setting, a typical question one may ask is the following: what is the most probable trajectory followed by the system in the spontaneous emergence of a fluctuation or in its relaxation to an equilibrium or a stationary state? To answer this question, one first derives a generalized Boltzmann–Einstein formula from which the most probable trajectory can be calculated by solving a variational principle. The entropy is related to the logarithm of the probability of such a trajectory and satisfies the HJE associated to the variational principle.

For states near equilibrium, an answer to this type of questions was given by Onsager and Machlup in 1953. The Onsager–Machlup theory gives the following result under the assumption of time reversibility of the microscopic dynamics. In the situation of a linear hydrodynamic equation and small fluctuations, that is, close to equilibrium, the most probable creation and relaxation trajectories of a fluctuation are time reversals of one another. This conclusion holds also in nonlinear hydrodynamic regimes and without the assumption of small fluctuations. This follows from the study of concrete models. In SNSs, on the other hand, time-reversal invariance is broken and the creation and relaxation trajectories of a fluctuation are not time reversals of one another.

In the following we refer to boundary-driven stationary nonequilibrium states, for example, a thermodynamic system in contact with reservoirs characterized by different temperatures and chemical potentials, but there is no difficulty in including an external field acting in the bulk.

## Microscopic and Macroscopic Dynamics

We consider many-body systems in the limit of infinitely many degrees of freedom. The basic general assumption of the theory is Markovian evolution. Microscopically, we assume that the evolution is described by a Markov process $X_\tau$ which represents the state of the system at time $\tau$. This hypothesis probably is not so restrictive, because the dynamics of Hamiltonian systems interacting with thermostats finally is also reduced to the analysis of a Markov process. Several examples are discussed in the literature. To be more precise, $X_\tau$ represents the set of variables necessary to specify the state of the microscopic constituents interacting among themselves and with the reservoirs. The SNS is described by a stationary, that is, invariant with respect to time shifts, probability distribution $P_{\mathrm{st}}$ over the trajectories of $X_\tau$.

Macroscopically, the usual interpretation of Markovian evolution is that the time derivatives of thermodynamic variables $\dot{\rho}_i$ at a given instant of time depend only on the $\rho_i$'s and the affinities (thermodynamic forces) $\partial S/\partial \rho_i$ at the same instant of time. Our next assumption can then be formulated as follows: the system admits a macroscopic description in terms of density fields which are the local thermodynamic variables. For simplicity of notation, we assume that there is only one thermodynamic variable (e.g., $\rho$, the density). The evolution of the field $\rho = \rho(t,u)$, where $t$ and $u$ are the macroscopic time and space coordinates (see below), is given by diffusion-type hydrodynamic equations of the form

$$\begin{aligned}\partial_t \rho &= \tfrac{1}{2}\nabla\cdot(D(\rho)\nabla\rho)\\ &= \tfrac{1}{2}\sum_{1\le i,j\le d} \partial_{u_i}\big(D_{i,j}(\rho)\partial_{u_j}\rho\big)\\ &= \mathcal{D}(\rho)\end{aligned} \qquad [1]$$

The interaction with the reservoirs appears as boundary conditions to be imposed on solutions of [1]. We assume that there exists a unique stationary solution $\overline{\rho}$ of [1], that is, a profile $\overline{\rho}(u)$, which satisfies the appropriate boundary conditions and is such that $\mathcal{D}(\overline{\rho}) = 0$. This holds if the diffusion matrix $D_{i,j}(\rho)$ in [1] is strictly elliptic, namely there exists a constant $c > 0$ such that $D(\rho) \ge c$ (in matrix sense).

These equations derive from the underlying microscopic dynamics through an appropriate

scaling limit in which the microscopic time and space coordinates $\tau, x$ are rescaled as follows: $t = \tau/N^2, u = x/N$, where $N$ represents the linear size of the system. For lattice systems, $N$ is an integer. The hydrodynamic equation [1] represents a law of large numbers with respect to the probability measure $P_{st}$ conditioned on an initial state $X_0$. The initial conditions for [1] are determined by $X_0$. Of course, many microscopic configurations give rise to the same value of $\rho(0, u)$. In general, $\rho = \rho(t, u)$ is an appropriate limit of a local observable $\rho_N(X_\tau)$ as the number $N$ of degrees of freedom diverges.

The hypothesis of Markovian evolution is also the basis of the 1931 Onsager's theory of irreversible processes near equilibrium. Onsager, however, did not rely on any microscopic model and assumed, near the equilibrium, linear hydrodynamic equations or regression equations as he called them. His equations, ignoring space dependence, were of the form

$$\dot{\rho}_i = -\sum_i D_{ij}\rho_j \qquad [2]$$

The diffusion matrix $D$ is related to Onsager transport matrix $\chi$ and the entropy by the relationship

$$D = \chi s \qquad [3]$$

where the elements of $s$ are $\partial^2 S/\partial\rho_i\partial\rho_j$. The matrix $\chi$ is defined by the relationship between flows and affinities

$$\dot{\rho}_i = -\sum_j \chi_{ij}\frac{\partial S}{\partial \rho_j} \qquad [4]$$

The indices $ij$ here label different thermodynamic variables. The matrix $\chi$ is symmetric, a property known as Onsager reciprocity. Equations [2] and [3] follow by developing the entropy near an equilibrium state, that is, by taking a quadratic expression as an approximation. The minus sign in eqn [4] is due to our convention in which the entropy has the same sign as the free energy.

Equation [3] permits to reconstruct the entropy from the knowledge of the coefficients $D$ and $\chi$ and has been widely used especially in physical chemistry. In SNSs, eqn [3] is replaced by a Hamilton–Jacobi-type equation for the entropy.

## Dynamical Boltzmann–Einstein Formula

The basic assumption is that the stationary ensemble $P_{st}$ admits a principle of large deviations describing the fluctuations of the thermodynamic variables appearing in the hydrodynamic equation. This means the following. The probability that for large $N$, the evolution of the random variable $\rho_N$ deviates from the solution of the hydrodynamic equation and is close to some trajectory $\hat{\rho}(t)$ is exponentially small and of the form

$$\begin{aligned} P_{st}(\rho_N(X_{N^2t}) \sim \hat{\rho}(t), t \in [t_1, t_2]) &\approx e^{-N^d[S(\hat{\rho}(t_1)) + J_{[t_1,t_2]}(\hat{\rho})]} \\ &= e^{-N^d I_{[t_1,t_2]}(\hat{\rho})} \end{aligned} \qquad [5]$$

where $d$ is the dimensionality of the system, $J(\hat{\rho})$ is a functional which vanishes if $\hat{\rho}(t)$ is a solution of [1] and $S(\hat{\rho}(t_1))$ is the entropy cost to produce the initial density profile $\hat{\rho}(t_1)$. We normalize $S$ so that $S(\bar{\rho}) = 0$. Therefore, $J(\hat{\rho})$ represents the extra cost necessary to follow the trajectory $\hat{\rho}(t)$. Finally, $\rho_N(X_{N^2t}) \sim \hat{\rho}(t)$ means closeness in some metric and $\approx$ denotes logarithmic equivalence as $N \to \infty$. Equation [5] is the dynamical generalization of the Boltzmann–Einstein formula. Experience with many models justifies this assumption.

To understand how [5] leads to a dynamical theory of the entropy, we discuss its properties under time reversal. Let us denote by $\theta$ the time inversion operator defined by $\theta X_\tau = X_{-\tau}$. The probability measure $P^*_{st}$ describing the evolution of the time-reversed process $X^*_\tau$ is given by the composition of $P_{st}$ and $\theta^{-1}$, that is,

$$\begin{aligned} P^*_{st}(X^*_\tau &= \phi_\tau, \tau \in [\tau_1, \tau_2]) \\ &= P_{st}(X_\tau = \phi_{-\tau}, \tau \in [-\tau_2, -\tau_1]) \end{aligned} \qquad [6]$$

Let $L$ be the generator of the microscopic dynamics. We remind that $L$ induces the evolution of observables (functions on the state space) according to the equation $\partial_\tau E_{X_0}[f(X_\tau)] = E_{X_0}[(Lf)(X_\tau)]$, where $E_{X_0}$ stands for the expectation with respect to $P_{st}$ conditioned on the initial state $X_0$.

The time-reversed dynamics, that is, the dynamics which inverts the direction of the fluxes through the system, for example, heat flows under this dynamics from lower to higher temperatures, is generated by the adjoint $L^*$ of $L$ with respect to the invariant measure $\mu$:

$$E^\mu[fLg] = E^\mu[(L^*f)g] \qquad [7]$$

The measure $\mu$, which is the same for both processes, is a distribution over the configurations of the system and formally satisfies $\mu L = 0$. The expectation with respect to $\mu$ is denoted by $E^\mu$ and $f, g$ are observables. We note that the probability $P_{st}$, and therefore $P^*_{st}$, depends on the invariant measure $\mu$. The finite-dimensional distributions of $P_{st}$ are in fact given by

$$\begin{aligned} P_{st}(X_{\tau_1} &= \phi_{\tau_1}, \ldots, X_{\tau_n} = \phi_{\tau_n}) \\ &= \mu(\phi_{\tau_1})\, p_{\tau_2-\tau_1}(\phi_{\tau_1} \to \phi_{\tau_2}) \cdots p_{\tau_n-\tau_{n-1}}(\phi_{\tau_{n-1}} \to \phi_{t\tau_n}) \end{aligned} \qquad [8]$$

where $p_\tau(\phi_1 \to \phi_2)$ is the transition probability. According to [6] the finite-dimensional distributions of $P^*_{\rm st}$ are

$$P^*_{\rm st}\left(X^*_{\tau_1} = \phi_{\tau_1}, \ldots, X^*_{\tau_n} = \phi_{\tau_n}\right)$$
$$= \mu(\phi_{\tau_1})\, p^*_{\tau_2-\tau_1}(\phi_{\tau_1} \to \phi_{\tau_2}) \cdots p^*_{\tau_n-\tau_{n-1}}(\phi_{\tau_{n-1}} \to \phi_{t\tau_n})$$
$$= \mu(\phi_{\tau_n}) p_{\tau_n-\tau_{n-1}}(\phi_{\tau_n} \to \phi_{\tau_{n-1}}) \cdots p_{\tau_2-\tau_1}(\phi_{\tau_2} \to \phi_{\tau_1}) \quad [9]$$

In particular, the transition probabilities $p_\tau(\phi_1 \to \phi_2)$ and $p_\tau{}^*(\phi_1 \to \phi_2)$ are related by

$$\mu(\phi_1)\, p_\tau(\phi_1 \to \phi_2) = \mu(\phi_2)\, p_\tau{}^*(\phi_2 \to \phi_1) \quad [10]$$

This relationship reduces to the well-known detailed balance condition if $p_\tau(\phi_1 \to \phi_2) = p_\tau{}^*(\phi_1 \to \phi_2)$.

We require that also the evolution generated by $L^*$ admits a hydrodynamic description, that we call the adjoint hydrodynamics, which, however, is not necessarily of the same form as [1]. In fact, we consider models in which the adjoint hydrodynamics is nonlocal in space.

In order to avoid confusion, we emphasize that what is usually called an equilibrium state for a reversible dynamics, as distinguished from an SNS, corresponds to the special case $L^* = L$, that is, the detailed balance principle holds. In such a case, $P_{\rm st}$ is invariant under time reversal and the two hydrodynamics coincide.

We now derive a first consequence of our assumptions, that is, the relationship between the functionals $I$ and $I^*$ associated to the dynamics $L$ and $L^*$ by [5]. From eqn [6], it follows that

$$I^*_{[t_1, t_2]}(\hat\rho) = I_{[-t_2, -t_1]}(\theta\hat\rho) \quad [11]$$

with obvious notations. More explicitly, this equation reads

$$S(\hat\rho(t_1)) + J^*_{[t_1, t_2]}(\hat\rho) = S(\hat\rho(t_2)) + J_{[-t_2, -t_1]}(\theta\hat\rho) \quad [12]$$

where $\hat\rho(t_1), \hat\rho(t_2)$ are the initial and final points of the trajectory and $S(\hat\rho(t_i))$ the entropies associated with the creation of the fluctuations $\hat\rho(t_i)$ starting from the SNS. The functional $J^*$ vanishes on the solutions of the adjoint hydrodynamics. To compute $J^*$, it is necessary to know the entropy $S$.

We consider now the following physical situation. The system is macroscopically in the stationary state $\bar\rho$ at $t = -\infty$, but at $t = 0$ we find it in the state $\rho$. We want to determine the most probable trajectory followed in the spontaneous creation of this fluctuation. According to [5], this trajectory is the one that minimizes $J$ among all trajectories $\hat\rho(t)$ connecting $\bar\rho$ to $\rho$ in the time interval $[-\infty, 0]$. From [12], recalling that $S(\bar\rho) = 0$, we have that

$$J_{[-\infty, 0]}(\hat\rho) = S(\rho) + J^*_{[0, \infty]}(\theta\hat\rho) \quad [13]$$

The right-hand side is minimal if $J^*_{[0,\infty]}(\theta\hat\rho) = 0$, that is, if $\theta\hat\rho$ is a solution of the adjoint hydrodynamics. The existence of such a relaxation solution is due to the fact that the stationary solution $\bar\rho$ is attractive also for the adjoint hydrodynamics. We have therefore the following consequences:

> In a SNS the spontaneous emergence of a macroscopic fluctuation takes place most likely following a trajectory which is the time reversal of the relaxation path according to the adjoint hydrodynamics.

This implies that the entropy is related to $J$ by

$$S(\rho) = \inf_{\hat\rho} J_{[-\infty, 0]}(\hat\rho) \quad [14]$$

where the minimum is taken over all trajectories $\hat\rho(t)$ connecting $\bar\rho$ to $\rho$.

We note that the reversibility of the microscopic process $X_\tau$, which we call microscopic reversibility, is not needed in order to deduce the Onsager–Machlup result (i.e., that the trajectory which creates the fluctuation is the time reversal of the relaxation trajectory). In fact, Onsager–Machlup result holds if and only if the hydrodynamics coincides with the adjoint hydrodynamics, which we call macroscopic reversibility. Indeed, it is possible to construct microscopic nonreversible models, $L \neq L^*$, in which the hydrodynamics and the adjoint hydrodynamics coincide.

Spontaneous fluctuations, including Onsager–Machlup time-reversal symmetry, have been observed in stochastically perturbed reversible electronic devices. In nonreversible systems, an asymmetry between the emergence and the relaxation of fluctuations has been observed. The above discussion provides the explanation.

## The Hamilton–Jacobi Equation and Its Consequences

We assume that the functional $J$ has a density (which plays the role of a Lagrangian), that is,

$$J_{[t_1, t_2]}(\hat\rho) = \int_{t_1}^{t_2} {\rm d}t \mathcal{L}(\hat\rho(t), \partial_t\hat\rho(t)) \quad [15]$$

Let us introduce the Hamiltonian $\mathcal{H}(\rho, H)$ as the Legendre transform of $\mathcal{L}(\rho, \partial_t\rho)$, that is,

$$\mathcal{H}(\rho, H) = \sup_\xi \{\langle \xi, H\rangle - \mathcal{L}(\rho, \xi)\} \quad [16]$$

where $\langle \cdot, \cdot \rangle$ denotes integration with respect to the macroscopic space coordinates $u$.

Noting that $\mathcal{H}(\bar{\rho}, 0) = 0$, the Hamilton–Jacobi equation associated to [14] is

$$\mathcal{H}\left(\rho, \frac{\delta S}{\delta \rho}\right) = 0 \qquad [17]$$

This is an equation for the functional derivative $C(\rho) = \delta S/\delta\rho$, but not all the solutions of the equation $\mathcal{H}(\rho, C(\rho)) = 0$ are the derivatives of some functional. Of course, only those which are the derivative of a functional are relevant for us.

We now specify the Hamilton–Jacobi equation [17] for boundary-driven lattice gases. For models with purely diffusive hydrodynamics [1], we expect a quadratic large deviation functional of the form

$$J_{[t_1,t_2]}(\hat{\rho}) = \frac{1}{2}\int_{t_1}^{t_2} \mathrm{d}t \big\langle \nabla^{-1}(\partial_t \hat{\rho} - \mathcal{D}(\rho)), \chi(\hat{\rho})^{-1}\, \nabla^{-1}(\partial_t \hat{\rho} - \mathcal{D}(\rho))\big\rangle \qquad [18]$$

where $\mathcal{D}(\rho)$ is the right-hand side of the hydrodynamic equation [1], and by $\nabla^{-1} f$ we mean a vector field whose divergence equals $f$. The form [18], which can be derived for several models, is expected to be very general: the functional $J(\hat{\rho})$ measures how much $\hat{\rho}$ differs from a solution of the hydrodynamics [1]. The matrix $\chi(\rho) = \chi(\rho)$ with $\chi(\rho)$ has the same role in our more general context, as the Onsager matrix in [4]. This form of $J$ is also typical for diffusion processes described by finite-dimensional Langevin equations (Freidlin–Wentzell theory).

In this case, the Lagrangian $\mathcal{L}$ is quadratic in $\partial_t \hat{\rho}(t)$ and the associated Hamiltonian is given by

$$\mathcal{H}(\rho, H) = \tfrac{1}{2}\langle \nabla H, \chi(\rho)\nabla H\rangle + \langle H, \mathcal{D}(\rho)\rangle \qquad [19]$$

so that the Hamilton–Jacobi equation [17] takes the form

$$\frac{1}{2}\left\langle \nabla\frac{\delta S}{\delta\rho}, \chi(\rho)\nabla\frac{\delta S}{\delta\rho}\right\rangle + \left\langle \frac{\delta S}{\delta\rho}, \mathcal{D}(\rho)\right\rangle = 0 \qquad [20]$$

As is well known in mechanics, the Hamilton–Jacobi equation has many solutions and we must give a criterion to select the correct one. The criterion which the correct solution has to satisfy is that it must be a Lyapunov function with respect to the unique stationary state.

It is a simple calculation to show that eqn [3] follows from HJE, if we look for a solution which is a local function of $\rho$. This is the right choice in equilibrium where correlations over macroscopic distances are not expected if the microscopic forces are short range.

Out of equilibrium, it has been shown by direct calculation that for a special model, the symmetric simple exclusion, the entropy is a nonlocal function of the thermodynamic variables, that is, space correlations extend to macroscopic distances. This result can be derived in a simple way from HJE as we will discuss later.

Lattice gases which do not conserve the number of particles do not give rise in general to a purely diffusive hydrodynamics but rather to a reaction diffusion equation. In this case, the large deviation functional will not have the quadratic form [18] and also the HJE will not be quadratic. An example in which particles can be created and destroyed is the so-called Kawasaki–Glauber dynamics. In this case, HJE has exponential nonlinearities.

## Nonequilibrium Fluctuation Dissipation Relation

We now derive a twofold generalization of the celebrated fluctuation dissipation relationship: it is valid in nonequilibrium states and in nonlinear regimes.

Such a relationship will hold provided the rate function $J^*$ of the time-reversed process is of the form [18] with $\mathcal{D}$ replaced by $\mathcal{D}^*$, the adjoint hydrodynamics,

$$\partial_t \rho = \mathcal{D}^*(\rho) \qquad [21]$$

with the same boundary conditions as [1].

If $J^*$ has the form

$$J^*_{[t_1,t_2]}(\hat{\rho}) = \frac{1}{2}\int_{t_1}^{t_2} \mathrm{d}t \big\langle (\nabla^{-1}(\partial_t \hat{\rho} - \mathcal{D}^*(\hat{\rho})), \chi(\hat{\rho})^{-1}\nabla^{-1}(\partial_t \hat{\rho} - \mathcal{D}^*(\hat{\rho}))\big\rangle \qquad [22]$$

by taking the variation of eqn [12], we get

$$\mathcal{D}(\rho) + \mathcal{D}^*(\rho) = \nabla\cdot\left(\chi(\rho)\nabla\frac{\delta S}{\delta\rho}\right) \qquad [23]$$

This relation can be verified explicitly for the nonequilibrium zero-range process which we discuss later and holds for several other models. It is also easy to check that the linearization of [23] around the stationary profile $\bar{\rho}$ yields a fluctuation dissipation relationship which reduces to the usual one in equilibrium.

The fluctuation dissipation relation [23] can be used to obtain the adjoint hydrodynamics from $\mathcal{D}(\rho)$ and $\delta S/\delta\rho$; the first is usually known and the second can be calculated from the Hamilton–Jacobi equation.

## H Theorem

We show that the functional $S$ is decreasing along the solutions of both the hydrodynamic equation [1] and the adjoint hydrodynamics

$$\partial_t \rho = \mathcal{D}^*(\rho) = \nabla\cdot\left(\chi(\rho)\nabla\frac{\delta S}{\delta\rho}\right) - \mathcal{D}(\rho) \qquad [24]$$

Let $\rho(t)$ be a solution of [1] or [24]; by using the Hamilton–Jacobi equation [20], we get

$$\frac{\mathrm{d}}{\mathrm{d}t}S(\rho(t)) = \left\langle \frac{\delta S}{\delta \rho}(\rho(t)), \partial_t \rho(t) \right\rangle$$
$$= -\frac{1}{2}\left\langle \nabla \frac{\delta S}{\delta \rho}(\rho(t)), \chi(\rho(t)) \nabla \frac{\delta S}{\delta \rho}(\rho(t)) \right\rangle$$
$$\leq 0 \qquad [25]$$

In particular, we have that $(\mathrm{d}/\mathrm{d}t)S(\rho(t)) = 0$ if and only if $(\delta S/\delta\rho)(\rho(t)) = 0$.

We remark that the right-hand side of [25] vanishes in the stationary state, that is, there is no internal entropy production due to the evolution. On the other hand, there is a steady entropy production due to the differences in the chemical potentials of the reservoirs. This is not discussed in this article.

### Decomposition of Hydrodynamics

There is a structural property of hydrodynamics which follows from the HJE. The hydrodynamic equation can be decomposed as the sum of a gradient vector field and a vector field $\mathcal{A}$ orthogonal to it in the metric induced by the operator $K^{-1}$, where $Kf = -\nabla \cdot (\chi(\rho)\nabla f)$, namely

$$\mathcal{D}(\rho) = \frac{1}{2}\nabla \cdot \left(\chi(\rho)\nabla \frac{\delta S}{\delta \rho}\right) + \mathcal{A}(\rho) \qquad [26]$$

with

$$\left\langle K\frac{\delta S}{\delta \rho}, K^{-1}\mathcal{A}(\rho) \right\rangle = \left\langle \frac{\delta S}{\delta \rho}, \mathcal{A}(\rho) \right\rangle = 0$$

Similarly, using the fluctuation dissipation relationship [23] for the adjoint hydrodynamics, we have

$$\mathcal{D}^*(\rho) = \frac{1}{2}\nabla \cdot \left(\chi(\rho)\nabla \frac{\delta S}{\delta \rho}\right) - \mathcal{A}(\rho) \qquad [27]$$

Since $\mathcal{A}$ is orthogonal to $\delta S/\delta\rho$, it does not contribute to the entropy production. The vector field $\mathcal{A}$ is odd under time reversal like a magnetic force.

Both terms of the decomposition vanish in the stationary state, that is, when $\rho = \bar{\rho}$. Whereas in equilibrium the hydrodynamics is the gradient flow of the entropy $S$, the term $\mathcal{A}(\rho)$ is characteristic of nonequilibrium states. Note that, for small fluctuations $\rho \approx \bar{\rho}$, small differences in the chemical potentials at the boundaries, $\mathcal{A}(\rho)$ becomes a second-order quantity and Onsager theory is a consistent approximation.

Equation [26] is interesting because it separates the dissipative part of the hydrodynamic evolution associated to the thermodynamic force $\delta S/\delta\rho$ and provides therefore an important physical information. Notice that the thermodynamic force $\delta S/\delta\rho$ appears linearly in the hydrodynamic equation even when this is nonlinear in the macroscopic variables.

In general, the two terms of the decomposition [26] are nonlocal in space even if $\mathcal{D}$ is a local function of $\rho$. This is the case for the simple exclusion process discussed later. Furthermore while the form of the hydrodynamic equation does not depend explicitly on the chemical potentials, $\delta S/\delta\rho$ and $\mathcal{A}$ do.

To understand how the decomposition [26] arises microscopically, let us consider a stochastic lattice gas. Let

$$L = \tfrac{1}{2}(L + L^*) + \tfrac{1}{2}(L - L^*) \qquad [28]$$

be its Markov generator, where $L^*$ is the adjoint of $L$ with respect to the invariant measure, namely the generator of the time-reversed microscopic dynamics. The term $L - L^*$ behaves like a Liouville operator, that is, it is anti-Hermitian and, in the scaling limit, produces the term $\mathcal{A}$ in the hydrodynamic equation. This can be verified explicitly in the boundary-driven zero-range model introduced in the next section.

Since the adjoint generator can be written as $L^* = (L + L^*)/2 - (L - L^*)/2$, the adjoint hydrodynamics must be of the form [27]. In particular, if the microscopic generator is self-adjoint, we get $\mathcal{A} = 0$ and thus $\mathcal{D}(\rho) = \mathcal{D}^*(\rho)$. On the other hand, it may happen that microscopic nonreversible processes, namely for which $L \neq L^*$, can produce macroscopic reversible hydrodynamics if $L - L^*$ does not contribute to the hydrodynamic limit.

The decompositions [26] and [27] remind of the electrical conduction in the presence of a magnetic field. Consider the motion of electrons in a conductor: a simple model is given by the effective equation

$$\dot{\boldsymbol{p}} = -e\left(\boldsymbol{E} + \frac{1}{mc}\boldsymbol{p} \wedge \boldsymbol{H}\right) - \frac{1}{\tau}\boldsymbol{p} \qquad [29]$$

where $\boldsymbol{p}$ is the momentum, $e$ the electron charge, $\boldsymbol{E}$ the electric field, $\boldsymbol{H}$ the magnetic field, $m$ the mass, $c$ the velocity of the light, and $\tau$ the relaxation time. The dissipative term $\boldsymbol{p}/\tau$ is orthogonal to the Lorentz force $\boldsymbol{p} \wedge \boldsymbol{H}$. We define time reversal as the transformation $\boldsymbol{p} \mapsto -\boldsymbol{p}, \boldsymbol{H} \mapsto -\boldsymbol{H}$. The adjoint evolution is given by

$$\dot{\rho} = e\left(\boldsymbol{E} + \frac{1}{mc}\boldsymbol{p} \wedge \boldsymbol{H}\right) - \frac{1}{\tau}\boldsymbol{p} \qquad [30]$$

where the signs of the dissipation and the electromagnetic force transform in analogy to [26] and [27].

Let us consider in particular the Hall effect where we have conduction along a rectangular plate immersed in a perpendicular magnetic field $H$ with a potential difference across the longer side. The magnetic field determines a potential difference across the other side of the plate. In our setting on the contrary, it is the difference in chemical potentials at the boundaries that introduces in the equations a "magnetic-like" term. There is therefore a kind of equivalence between certain externally applied fields and driving the system at the boundaries.

### Minimum Dissipation Principle

In 1931 Onsager formulated, within his near equilibrium theory, a variational principle which shows that the hydrodynamic evolution minimizes at each instant of time a quadratic functional of $\dot{\rho}$. He called this the "minimum dissipation principle." We now show that the decomposition of the previous subsection leads to a natural exact generalization of this principle. We want to construct a functional of the variables $\rho$ and $\dot{\rho}$ such that the Euler equation associated to the vanishing of the first variation under arbitrary changes of $\dot{\rho}$ is the hydrodynamic equation [1]. We define the "dissipation function"

$$F(\rho, \dot{\rho}) = \left\langle (\dot{\rho} - \mathcal{A}(\rho)), K^{-1}(\dot{\rho} - \mathcal{A}(\rho)) \right\rangle \qquad [31]$$

and the functional

$$\begin{aligned} \Phi(\rho, \dot{\rho}) &= \dot{S}(\rho) + F(\rho, \dot{\rho}) \\ &= \left\langle \frac{\delta S}{\delta \rho}, \dot{\rho} \right\rangle + \langle (\dot{\rho} - \mathcal{A}(\rho)), \\ &\quad K^{-1}(\dot{\rho} - \mathcal{A}(\rho)) \rangle \end{aligned} \qquad [32]$$

which generalize the corresponding Onsager's definitions (Onsager 1931a, b). The operator $K$ has been defined in the previous subsection.

It is easy to verify that

$$\delta_{\dot{\rho}} \Phi = 0 \qquad [33]$$

is equivalent to the hydrodynamic equation [1]. Furthermore, a simple calculation gives

$$F|_{\dot{\rho} = \mathcal{D}(\rho)} = \frac{1}{4} \left\langle \nabla \frac{\delta S}{\delta \rho}, \chi(\rho) \nabla \frac{\delta S}{\delta \rho} \right\rangle \qquad [34]$$

that is, $2F$ on the hydrodynamic trajectories equals the entropy production rate as in Onsager's near equilibrium approximation.

The dissipation function for the adjoint hydrodynamics is obtained by changing the sign of $\mathcal{A}$ in [31].

### Entropy and Optimal Control

There is an interesting interpretation of the entropy as a minimal cost to produce a fluctuation by externally acting on the system. The idea is to show that there exists a cost function which on the optimal control trajectory coincides with the entropy difference with respect to the stationary state.

We add an external perturbation $v$ to the hydrodynamic equation

$$\partial_t \rho = \tfrac{1}{2} \nabla \cdot (D(\rho) \nabla \rho) + v = \mathcal{D}(\rho) + v \qquad [35]$$

We want to choose $v$ so as to drive, with minimal cost, the system from its stationary state $\bar{\rho}$ to an arbitrary state $\rho$. A simple cost function is

$$\frac{1}{2} \int_{t_1}^{t_2} \mathrm{d}s \langle v(s), K^{-1}(\rho(s)) v(s) \rangle \qquad [36]$$

where $\rho(s)$ is the solution of [35] and we recall that $K(\rho)f = -\nabla \cdot (\chi(\rho) \nabla f)$. More precisely, given $\rho(t_1) = \bar{\rho}$, we want to drive the system to $\rho(t_2) = \rho$ by an external field $v$ which minimizes [36]. This is a standard problem in control theory. Let

$$\mathcal{V}(\rho) = \inf \frac{1}{2} \int_{t_1}^{t_2} \mathrm{d}s \langle v(s), K^{-1}(\rho(s)) v(s) \rangle \qquad [37]$$

where the infimum is taken with respect to all fields $v$ which drive the system to $\rho$ in an arbitrary time interval $[t_1, t_2]$. The optimal field $v$ can be obtained by solving the Bellman equation which reads

$$\min_{v} \left\{ \frac{1}{2} \langle v, K^{-1}(\rho) v \rangle - \left\langle \mathcal{D}(\rho) + v, \frac{\delta \mathcal{V}}{\delta \rho} \right\rangle \right\} = 0 \qquad [38]$$

It is easy to express the optimal $v$ in terms of $\mathcal{V}$; we get

$$v = K \frac{\delta \mathcal{V}}{\delta \rho} \qquad [39]$$

Hence, [38] now becomes

$$\frac{1}{2} \left\langle \frac{\delta \mathcal{V}}{\delta \rho}, K(\rho) \frac{\delta \mathcal{V}}{\delta \rho} \right\rangle + \left\langle \mathcal{D}(\rho), \frac{\delta \mathcal{V}}{\delta \rho} \right\rangle = 0 \qquad [40]$$

By identifying the cost functional $\mathcal{V}(\rho)$ with $S(\rho)$, eqn [40] coincides with the Hamilton–Jacobi equation [20].

By inserting the optimal $v$ [39] in [35] and identifying $\mathcal{V}$ with $S$, we get that the optimal trajectory $\rho(t)$ solves the time-reversed adjoint hydrodynamics, namely

$$\partial_t \rho = -\mathcal{D}^*(\rho) \qquad [41]$$

The trajectory of the spontaneous emergence of a fluctuation coincides therefore with the trajectory of minimal cost for the optimal control. The optimal field $v$ does not depend on the nondissipative part $\mathcal{A}$ of the hydrodynamics.

## Models

The general theory will now be illustrated by briefly describing models where it has been successfully applied. We consider examples of different nature in order to emphasize the generality and flexibility of the point of view developed in the previous section.

We have chosen three examples in which the theory is used in different ways. The first one, the zero-range process, can be solved in a simple way so that the theory can be verified in detail. In the second one, the symmetric simple exclusion, we derive from the HJE a nonlinear ordinary differential equation first obtained by Derrida, Lebowitz, and Speer through a direct rather complex calculation. This equation implies the nonlocality of the entropy in the SNS of this model. The third model, the Kawasaki–Glauber dynamics, provides the illustration of two aspects. Nonlocality of the entropy, that is, long-range correlations, can appear in isolated equilibrium states if the microscopic dynamics is not time-reversal invariant. This means that long-range correlations as a signature of time-reversal violation are not restricted to SNSs. The second aspect to be underlined is the effectiveness of the HJE in a more complex case: in fact in this model, the number of particles is not conserved which leads to a very complicated structure of the HJE.

As a general comment, we emphasize that dynamics microscopically different but leading to the same macroscopic description, in particular the same hydrodynamics and large deviation functional, are indistinguishable for the theory which is purely macroscopic.

### Zero Range

We consider the so-called zero-range process which models a nonlinear diffusion of a lattice gas. The model is described by a positive integer variable $\eta_\tau(x)$ representing the number of particles at site $x$ and time $\tau$ of a finite lattice which for simplicity we assume one dimensional. The particles jump with rates $g(\eta(x))$ to one of the nearest-neighbor sites $x+1, x-1$ with probability 1/2. The function $g(k)$ is nondecreasing and $g(0)=0$. We assume that our system interacts with two reservoirs of particles in positions $N$ and $-N$ with rates $p_+$ and $p_-$, respectively. This model can be solved exactly and the previous theory can be checked in full detail.

Let us introduce the macroscopic coordinates, time $t=\tau/N^2$ and space $u=x/N$. To describe the macroscopic dynamics, we introduce the empirical density

$$\rho_N(t,u) = \frac{1}{N}\sum_{x=-N}^{N} \eta_{N^2 t}(x)\delta(u-x/N) \qquad [42]$$

where $\delta(u-x/N)$ is the Dirac $\delta$. One can prove that in the limit $N \to \infty$, the empirical density [42] tends in probability to a continuous function $\rho_t(u)$, which satisfies the following hydrodynamic equation:

$$\partial_t \rho = \tfrac{1}{2}\Delta\phi(\rho) = \mathcal{D}(\rho) \qquad [43]$$

where $\phi(\rho)$ can be explicitly defined in terms of the rates $g(\eta)$. The boundary conditions for [43] are $\phi(\rho(t, \pm 1)) = p_\pm$.

The adjoint hydrodynamics is

$$\partial_t \rho = \frac{1}{2}\Delta\phi(\rho) - \alpha\nabla\left\{\frac{\phi(\rho)}{\lambda(u)}\right\} = \mathcal{D}^*(\rho) \qquad [44]$$

with

$$\lambda(u) = \frac{p_+ - p_-}{2}u + \frac{p_+ + p_-}{2}$$

and

$$\alpha = \frac{p_+ - p_-}{2}$$

The boundary conditions for [44] are the same as for [43]. The second term on the right-hand side of [44] is proportional to the difference of the chemical potentials and produces an inversion of the particle flux. The action functionals $J(\hat{\rho})$ and $J^*(\hat{\rho})$ for this model have been computed and have the form [18] and [22], respectively, with $\chi(\rho)=\phi(\rho)$. The entropy $S(\rho)$ can be easily computed directly from the expression of the invariant measure which is of product type and is known explicitly:

$$S(\rho) = \int_{-1}^{1} \mathrm{d}u \left[\rho(u)\log\frac{\phi(\rho(u))}{\lambda(u)} - \log\frac{Z(\phi(\rho(u)))}{Z(\lambda(u))}\right] \qquad [45]$$

where

$$Z(\phi) = 1 + \sum_{k=1}^{\infty}\frac{\phi^k}{g(1)\cdots g(k)}$$

It is easy to verify that it solves the HJE. Due to the special zero-range character of the interaction in this model, there are no long-range correlations in nonequilibrium states.

### Simple Exclusion

The simple exclusion process is a model of a lattice gas with an exclusion principle: a particle can move to a neighboring site, with rate 1/2 for each side, only if this is empty. We consider again a one-dimensional case and we denote by $\eta_x(\tau) \in \{0,1\}$ the number of particles at the site $x$ at (microscopic) time $\tau$. The system is in contact with particle reservoirs at the boundaries $\pm N$ where a particle is created with rates $p_\pm$ if the boundary site is empty and is destroyed $1-p_\pm$ if it is occupied. In contrast to the zero-range model, the invariant measure carries long-range correlations making the entropy nonlocal.

The hydrodynamic equation for the simple exclusion process can be derived as for the zero-range process; in fact, it is easier in this case because a simple computation leads directly to a closed equation for the empirical density which is defined as in [42] except that the variable $\eta$ now takes only the values 0 or 1. We find that the limiting density evolves according to the linear heat equation

$$\partial_t \rho(t,u) = \tfrac{1}{2}\Delta\rho(t,u) = \mathcal{D}(\rho) \qquad [46]$$

with boundary conditions

$$\rho(t,\pm 1) = \frac{p_\pm}{1+p_\pm} = \rho_\pm$$

In this case, the density of particles $\rho$ takes values in [0,1]. We use the HJE to calculate the entropy. For this model, we have $\chi(\rho) = \rho(1-\rho)$. We show that the solution of the HJE for $S(\rho)$ (which is a functional derivative equation) can be reduced to the solution of an ordinary differential equation.

The Hamilton–Jacobi equation for the simple exclusion process is

$$\left\langle \nabla\frac{\delta S}{\delta\rho}, \rho(1-\rho)\nabla\frac{\delta S}{\delta\rho}\right\rangle + \left\langle \frac{\delta S}{\delta\rho}, \Delta\rho\right\rangle = 0 \qquad [47]$$

We look for a solution of the form

$$\frac{\delta S}{\delta\rho(u)} = \log\frac{\rho(u)}{1-\rho(u)} - \phi(u;\rho) \qquad [48]$$

for some functional $\phi(u;\rho)$ to be determined satisfying the boundary conditions

$$\phi(\pm 1) = \log\frac{\rho_\pm}{1-\rho_\pm}$$

in the space variable. The first term on the right-hand side is the derivative of the equilibrium entropy, that is for boundary conditions $\rho_- = \rho_+$.

Inserting [48] into [47], we get (note that $\rho - e^\phi/(1+e^\phi)$ vanishes at the boundary)

$$\begin{aligned}
0 &= -\left\langle \nabla\left(\log\frac{\rho}{1-\rho} - \phi\right), \rho(1-\rho)\nabla\phi\right\rangle \\
&= -\langle\nabla\rho,\nabla\phi\rangle + \left\langle \rho(1-\rho), (\nabla\phi)^2\right\rangle \\
&= -\left\langle \nabla\left(\rho - \frac{e^\phi}{1+e^\phi}\right), \nabla\phi\right\rangle \\
&\quad - \left\langle \left(\rho - \frac{e^\phi}{1+e^\phi}\right)\left(\rho - \frac{1}{1+e^\phi}\right), (\nabla\phi)^2\right\rangle \\
&= \left\langle \left(\rho - \frac{e^\phi}{1+e^\phi}\right), \left(\Delta\phi + \frac{(\nabla\phi)^2}{1+e^\phi} - \rho(\nabla\phi)^2\right)\right\rangle
\end{aligned}$$

We obtain a nontrivial solution of the Hamilton–Jacobi if we solve the following ordinary differential equation, corresponding to the vanishing of the right side of the scalar product, which relates the functional $\phi(u) = \phi(u;\rho)$ to $\rho$:

$$\begin{aligned}
&\frac{\Delta\phi(u)}{[\nabla\phi(u)]^2} + \frac{1}{1+e^{\phi(u)}} = \rho(u), \quad u \in (-1,1) \\
&\phi(\pm 1) = \log\frac{\rho_\pm}{1-\rho_\pm}
\end{aligned} \qquad [49]$$

It is clear that $\phi$ is a nonlocal functional of $\rho$. A computation shows that the derivative of the functional

$$S(\rho) = \int du\Big\{\rho\log\rho + (1-\rho)\log(1-\rho) + (1-\rho)\phi - \log(1+e^\phi) + \log\frac{\nabla\phi}{\nabla\bar{\rho}}\Big\}$$

is given by [48] when $\phi(u;\rho)$ solves [49].

### Kawasaki–Glauber Dynamics

The model consists of particles on a lattice evolving according to two basic dynamical processes:

1. a particle can move to a neighboring site if this is empty as in the simple exclusion and
2. a particle can disappear in an occupied site or be created if this is empty, the rate depending on the nearby configuration.

The first process is conservative while the second is not.

As before the object of our study is the empirical density [42]. It is possible to show that as $N$ goes to infinity, $\rho(t,u)$ is a solution of

$$\partial_t\rho = \tfrac{1}{2}\Delta\rho + B(\rho) - D(\rho) \qquad [50]$$

with

$$B(\rho) = E_{\nu_\rho}(c(\eta)(1-\eta(0))) \qquad [51]$$

$$D(\rho) = E_{\nu_\rho}(c(\eta)\eta(0)) \qquad [52]$$

where $\nu_\rho$ is the Bernoulli product distribution with parameter $\rho$. Typically, $B(\rho)$ and $D(\rho)$ are polynomials in $\rho$. For this model we consider equilibrium states so that we can take periodic boundary conditions. An equilibrium state corresponds to a density $\bar{\rho}$ which is the solution of the equation $B(\rho) = D(\rho)$ and gives a minimum of the potential $V(\rho) = \int^\rho [D(\rho') - B(\rho')]\mathrm{d}\rho'$. We admit potentials with several minima. The Hamiltonian associated to the large deviation functional for this model is not quadratic:

$$\mathcal{H}(\rho, H) = \int \mathrm{d}u \left\{ \frac{1}{2} H\Delta\rho + \frac{1}{2}(\nabla H)^2 \rho(1-\rho) - B(\rho)(1 - \exp H) - D(\rho) \times (1 - \exp(-H)) \right\} \quad [53]$$

where $H$ has the role of the conjugate momentum. The Hamilton–Jacobi equation

$$\mathcal{H}\left(\rho, \frac{\delta S}{\delta \rho}\right) = 0 \quad [54]$$

is therefore very complicated but can be solved by successive approximations using as an expansion parameter $\rho - \bar{\rho}$, where $\bar{\rho}$ is a solution of $B(\rho) = D(\rho)$ that is a stationary solution of hydrodynamics. For $\rho = \bar{\rho}$, we have $\delta S/\delta\rho = 0$. We are looking for an approximate solution of [54] of the form

$$S(\rho) = \frac{1}{2} \int \mathrm{d}u \int \mathrm{d}v (\rho(u) - \bar{\rho}) k(u,v) (\rho(v) - \bar{\rho}) + o(\rho - \bar{\rho})^2 \quad [55]$$

The kernel $k(u,v)$ is the inverse of the density correlation function $c(u,v)$.

$$\int c(u,y) k(y,v)\, \mathrm{d}y = \delta(u - v) \quad [56]$$

By inserting [55] in [54], one can show that $k(u,v)$ satisfies the following equation:

$$\tfrac{1}{2}\bar{\rho}(1-\bar{\rho})\Delta_u k(u,v) - b_0 k(u,v) - \tfrac{1}{2}\Delta_u \delta(u-v) + (d_1 - b_1)\delta(u-v) = 0 \quad [57]$$

where

$$b_1 = B'(\rho)|_{\rho=\bar{\rho}}, \qquad d_1 = D'(\rho)|_{\rho=\bar{\rho}}$$

and

$$b_0 = B(\bar{\rho}) = D(\bar{\rho}) = d_0 \quad [58]$$

If the entropy is a local functional of the density, $k(u,v)$ must be of the form $k(u,v) = f(\bar{\rho})\delta(u-v)$ which inserted in [57] gives

$$f(\bar{\rho}) = [\bar{\rho}(1-\bar{\rho})]^{-1} \quad [59]$$

and

$$b_0[\bar{\rho}(1-\bar{\rho})]^{-1} - (d_1 - b_1) = 0 \quad [60]$$

Therefore if $b_0, b_1, d_1$ do not satisfy the last equation, the entropy cannot be a local functional of the density. It can be shown that in this case time-reversal invariance is violated and the adjoint hydrodynamics is different from [50]. This calculation supports the conjecture that macroscopic correlations are a generic feature of equilibrium states of nonreversible lattice gases.

*See also:* Interacting Particle Systems and Hydrodynamic Equations; Interacting Stochastic Particle Systems; Nonequilibrium Statistical Mechanics (Stationary): Overview; Quantum Central-Limit Theorems.

## Further Reading


Bertini L, De Sole A, Gabrielli D, Jona–Lasinio G, and Landim C (2002) Macroscopic fluctuation theory for stationary non-equilibrium states. *Journal of Statistical Physics* 107: 635–675.

Bertini L, De Sole A, Gabrielli D, Jona–Lasinio G, and Landim C (2004) Minimum dissipation principle in stationary non-equilibrium states. *Journal of Statistical Physics* 116: 831–841.

Bertini L, De Sole A, Gabrielli D, Jona-Lasinio G, and Landim C (2005) Nonequilibrium current fluctuations in stochastic lattice gases. To appear in *Journal of Statistical Physics*, arXiv: cond-mat/0506664. See also references therein.

Derrida B, Lebowitz JL, and Speer ER (2002) Large deviation of the density profile in the steady state of the open symmetric simple exclusion process. *Journal of Statistical Physics* 107: 599–634.

Freidlin MI and Wentzell AD (1998) *Random Perturbations of Dynamical Systems*. New York: Springer.

Kipnis C and Landim C (1999) *Scaling Limits of Interacting Particle Systems*. Berlin: Springer.

Landau L and Lifshitz E (1967) *Physique Statistique*. Moscow: MIR.

Lanford OE (1973) In: Lenard A (ed.) *Entropy and Equilibrium States in Classical Statistical Mechanics*, Lecture Notes in Physics, vol. 20. Berlin: Springer.

Luchinsky DG and McClintock PVE (1997) Irreversibility of classical fluctuations studied in analogue electrical circuits. *Nature* 389: 463–466.

Onsager L (1931a) Reciprocal relations in irreversible processes. I. *Physical Review* 37: 405–426.

Onsager L (1931b) Reciprocal relations in irreversible processes. II. *Physical Review* 38: 2265–2279.

Onsager L and Machlup S (1953) Fluctuations and irreversible processes. *Physical Review* 91: 1505–1512 and 1512–1515.

Spohn H (1991) *Large Scale Dynamics of Interacting Particles*. Berlin: Springer.

# Magnetic Resonance Imaging

**C L Epstein**, University of Pennsylvania, Philadelphia, PA, USA
**F W Wehrli**, University of Pennsylvania, Philadelphia, PA, USA



## Introduction

Nuclear magnetic resonance (NMR) is a subtle quantum-mechanical phenomenon that, through magnetic resonance imaging (MRI), has played a major role in the revolution in medical imaging over the last 30 years. Before being conceived for use in imaging, NMR was employed by chemists to do spectroscopy, and it remains a very important technique for determining the structure of complex chemical compounds like proteins. In this article we explain how NMR is used to create an image of a three-dimensional object. Scant attention is paid to both NMR spectroscopy, and the quantum description of NMR. Those seeking a more complete introduction to these subjects should consult the article Nuclear Magnetic Resonance in this Encyclopedia, as well as the monographs of Abragam (1983) or Ernst *et al.* (1987), for spectroscopy, and that of Callaghan (1993) for imaging. All three books consider the quantum-mechanical description of these phenomena. Comprehensive discussions of MRI can be found in Bernstein *et al.* (2004) and Haacke *et al.* (1999), and a historical appreciation of the development of MRI is given in Wehrli (1995).

## The Bloch Equation

We begin with the Bloch phenomenological equation, which provides a model for the interactions between applied magnetic fields and the nuclear spins in the objects under consideration. This is a macroscopic averaged model that describes the interaction of aggregates of spins, called isochromats, with applied magnetic fields. An isochromat is a collection of "like" spins, which is spatially large on the atomic scale, but very small on the scale of the variations present in the applied magnetic fields. Spins are alike if they belong to the same species and are in the same chemical environment. There may be several different classes of spins, but, in this article, it is assumed that they are noninteracting and so it suffices to consider each separately. Heretofore, we suppose that there is a single class of like spins. The distribution of isochromats for these spins is described macroscopically by the spin density function, which we denote by $\rho(x,y,z)$. In most medical applications, one is imaging the distribution of spins arising from hydrogen protons in water molecules.

The state of the isochromat at spatial location $(x,y,z)$ is given by a 3-vector:

$$\boldsymbol{M}(x,y,z) = (m_1(x,y,z), m_2(x,y,z), m_3(x,y,z))$$

which is interpreted as the magnetic moment per unit volume. It is an ensemble mean of the quantum dipoles caused by the spins within the isochromat. In most applications of NMR to imaging, the applied magnetic field is described as the sum of a large, time-independent field, $\boldsymbol{B}_0(x,y,z)$, and smaller time-dependent fields, $\boldsymbol{B}'(x,y,z;t)$. In the presence of a static field, thermal fluctuations cause the nuclear spins to slightly prefer an orientation aligned with the field. Using the Boltzmann distribution, one obtains that the nuclear paramagnetic susceptibility of water protons is given by

$$\chi = \frac{\hbar^2\gamma^2}{4k_B T} \qquad [1]$$

here $\hbar$ is Planck's constant, $k_B$ the Boltzmann's constant, and $T$ the absolute temperature, (see Levitt (2001)). The constant $\gamma$ is called the gyromagnetic (or magnetogyric) ratio. For a proton,

$$\gamma \approx 2\pi \times 42.5764 \times 10^6 \text{ rad s}^{-1}\,\text{T}^{-1} \qquad [2]$$

For water molecules at room temperature, $\chi \approx 3.6 \times 10^{-9}$.

If the sample is held stationary in the field $\boldsymbol{B}_0$ for a sufficiently long time, then the spins become polarized and a bulk magnetic moment appears; this is called the equilibrium magnetization:

$$\boldsymbol{M}_0(x,y,z) = \chi\rho(x,y,z)\boldsymbol{B}_0(x,y,z) \qquad [3]$$

The Bloch equation describes the evolution of $\boldsymbol{M}$ under the influence of the applied field $\boldsymbol{B} = \boldsymbol{B}_0 + \boldsymbol{B}'$:

$$\frac{\mathrm{d}\boldsymbol{M}(x,y,z;t)}{\mathrm{d}t} = \gamma\boldsymbol{M}(x,y,z;t) \times \boldsymbol{B}(x,y,z;t) - \frac{1}{T_2}\boldsymbol{M}^{\perp}(x,y,z;t) + \frac{1}{T_1}(\boldsymbol{M}_0(x,y,z) - \boldsymbol{M}^{\parallel}(x,y,z;t)) \qquad [4]$$

Here $\times$ is the vector cross-product, $\boldsymbol{M}^{\perp}(x,y,z;t)$ the component of $\boldsymbol{M}(x,y,z;t)$ perpendicular to $\boldsymbol{B}_0(x,y,z)$ (called the transverse component), and $\boldsymbol{M}^{\parallel}$ the component of $\boldsymbol{M}$ parallel to $\boldsymbol{B}_0$ (called the longitudinal component). For hydrogen protons in other molecules, the gyromagnetic ratio is expressed in the form $(1-\sigma)\gamma$. The coefficient $\sigma$ is called the

nuclear shielding; it is typically between $-10^{-4}$ and $+10^{-4}$. The difference in the nuclear shielding causes a shift in the resonance frequency by $\gamma\sigma$.

The second and third terms in eqn [4] are relaxation terms. They provide a phenomenological model for the averaged interactions of the spins with one another and their environment. The coefficient $1/T_1(x, y, z)$ is the spin lattice relaxation rate; it describes the rate at which the magnetization returns to equilibrium. The coefficient $1/T_2(x, y, z)$ is the spin–spin relaxation rate; it describes the rate at which the transverse components of $\boldsymbol{M}$ decay. The physical processes causing these relaxation phenomena are different and so are the rates themselves, with $T_2$ less than $T_1$. The relaxation rates largely depend on the localized thermal fluctuations of the molecules and provide a useful contrast mechanism in MR imaging. Spin–spin relaxation occurs very rapidly in solids (<1 ms) and, therefore, we usually assume that we are imaging liquid-like materials such as water protons in soft mammalian tissues. In this case, $T_2$ takes values in the 40 ms to 4 s range. Notice that this model does not include any explicit interaction between isochromats at different spatial locations. A variety of such interactions exist, but, at least in liquid-like materials, they lead only to small corrections in the Bloch equation model. A derivation of the Bloch equation from the Schrödinger equation can be found in Abragam (1983) and Slichter (1990). For coupled systems, the Bloch equation formalism breaks down and a full quantum-mechanical treatment is necessary (*see* Nuclear Magnetic Resonance and Ernst *et al.* (1983)).

Much of the analysis in NMR imaging amounts to understanding the behavior of solutions to eqn [4] with different choices of $\boldsymbol{B}$. We now consider some important special cases. The simplest case occurs if $\boldsymbol{B}$ has no time-dependent component; then this equation predicts that the sample becomes polarized with the transverse part of $\boldsymbol{M}$ decaying as $e^{-t/T_2}$, and the longitudinal component approaching the equilibrium magnetization, $\boldsymbol{M}_0$, as $1 - e^{-t/T_1}$. To simplify the subsequent discussion, we assume that the field $\boldsymbol{B}_0$ is homogeneous with $\boldsymbol{B}_0 = (0, 0, b_0)$. If $\boldsymbol{B} = \boldsymbol{B}_0$ and we omit the relaxation terms (set $T_1 = T_2 = \infty$ in [4]), then an initial magnetization $\boldsymbol{M}(x, y, z; 0)$ simply precesses about $\boldsymbol{B}_0$ at angular frequency $\omega_0 = \gamma b_0 : \boldsymbol{M}(x, y, z; t) = U(t)\boldsymbol{M}(x, y, z; 0)$, with

$$U(t) = \begin{bmatrix} \cos\omega_0 t & -\sin\omega_0 t & 0 \\ \sin\omega_0 t & \cos\omega_0 t & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad [5]$$

The frequency $\omega_0$ is called the Larmor frequency; this precession of $\boldsymbol{M}$ about the axis of $\boldsymbol{B}_0$ is the resonance phenomenon referred to as NMR. In typical medical imaging systems, $b_0$ is between 1 and 3 T and the corresponding resonance frequency is between 40 and 120 MHz.

Typically, the field $\boldsymbol{B}$ takes the form

$$\boldsymbol{B} = \boldsymbol{B}_0 + \tilde{\boldsymbol{G}} + \boldsymbol{B}_1 \quad [6]$$

where $\tilde{\boldsymbol{G}}$ is a gradient field and $\boldsymbol{B}_1$ is a radio-frequency (RF) field. Usually, the gradient fields are "piecewise time-independent" fields, small relative to $\boldsymbol{B}_0$. By piecewise time-independent field, we mean a collection of static fields that, in the course of the experiment, are turned on and off. The $\boldsymbol{B}_1$ component is a time-dependent RF field, nominally at right angles to $\boldsymbol{B}_0$. It is usually taken to be spatially homogeneous, with time dependence of the form

$$\boldsymbol{B}_1(t) = U(t)\begin{pmatrix} \alpha(t) \\ \beta(t) \\ 0 \end{pmatrix} \quad [7]$$

The functions $\alpha$ and $\beta$ define an envelope that modulates the time-harmonic field, $[\cos\omega_0 t, \sin\omega_0 t, 0]$. They are supported in a finite interval $[t_0, t_1]$, that is, the $\boldsymbol{B}_1$ field is "turned on" for a finite period of time. The change in the state of the magnetization between $t_0$ and $t_1$ is called the RF excitation. It may be spatially dependent.

In light of [5] it is convenient to introduce the rotating reference frame. We replace $\boldsymbol{M}$ with $\boldsymbol{m}$, where $\boldsymbol{m}(x, y, z; t) = U(t)^{-1}\boldsymbol{M}(x, y, z; t)$. It is a classical result of Larmor, that if $\boldsymbol{M}$ satisfies [4], then $\boldsymbol{m}$ satisfies

$$\begin{aligned}\frac{d\boldsymbol{m}(x, y, z; t)}{dt} &= \gamma\boldsymbol{m}(x, y, z; t) \times \boldsymbol{B}_{\text{eff}}(x, y, z; t) \\ &\quad - \frac{1}{T_2}\boldsymbol{m}^{\perp}(x, y, z; t) + \frac{1}{T_1}(\boldsymbol{M}_0(x, y, z) \\ &\quad - \boldsymbol{m}^{\parallel}(x, y, z; t))\end{aligned} \quad [8]$$

where

$$\boldsymbol{B}_{\text{eff}} = U(t)^{-1}\boldsymbol{B} - \left(0, 0, \frac{\omega_0}{\gamma}\right)$$

As $\tilde{\boldsymbol{G}}$ is much smaller than $\boldsymbol{B}$ and quasistatic, it turns out that one can ignore the components of $\tilde{\boldsymbol{G}}$ orthogonal to $\boldsymbol{B}_0$. Indeed, in imaging applications, one usually assumes that the components of $\tilde{\boldsymbol{G}}$ depend linearly on $(x, y, z)$ with the $\hat{z}$-component given by $\langle (x, y, z), (g_1, g_2, g_3)\rangle$. The constant vector $\boldsymbol{G} = (g_1, g_2, g_3)$ is called the gradient vector. With $\boldsymbol{B}_0 = (0, 0, b_0)$ and $\boldsymbol{B}_1$ given by [7], we see that $\boldsymbol{B}_{\text{eff}}$ can be taken to equal $(0, 0, \langle (x, y, z), \boldsymbol{G}\rangle) + (\alpha, \beta, 0)$.

In the remainder of this article, we assume that $\boldsymbol{B}_{\text{eff}}$ takes this form.

If $\boldsymbol{G}=0$ and $\beta \equiv 0$, then the solution operator for Bloch's equation, without relaxation terms, is

$$V(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta(t) & \sin\theta(t) \\ 0 & -\sin\theta(t) & \cos\theta(t) \end{bmatrix} \quad [9]$$

where

$$\theta(t) = \int_0^t \alpha(s)\,\mathrm{d}s \quad [10]$$

This is simply a rotation about the $x$-axis through the angle $\theta(t)$. If $\boldsymbol{B}_1 \neq 0$ for $t \in [0,\tau]$, then the magnetization is rotated through the angle $\theta(\tau)$. Thus, RF excitation can be used to move the magnetization out of its equilibrium state. As we shall soon see, this is crucial for obtaining a measurable signal. Note that the equilibrium magnetization is a tiny perturbation of the very large field $\boldsymbol{B}_0$ and is, therefore, in practice not directly measurable. Only the precessional motion of the transverse components of $\boldsymbol{M}$ produces a measurable signal. More general $\boldsymbol{B}_1$ fields, that is, with both $\alpha$ and $\beta$ nonzero, have more complicated effects on the magnetization. In general, the angle between $\boldsymbol{M}$ and $\boldsymbol{M}_0$ at the conclusion of the RF excitation is called the flip angle.

If, on the other hand, $\boldsymbol{B}_1=0$ and $\boldsymbol{G}_l=(0,0,l(x,y,z))$, where $l(\cdot)$ is a function, then $V$ depends on $(x,y,z)$, and is given by

$$V(x,y,z;t) = \begin{bmatrix} \cos\gamma l(x,y,z)t & -\sin\gamma l(x,y,z)t & 0 \\ \sin\gamma l(x,y,z)t & \cos\gamma l(x,y,z)t & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad [11]$$

This is precession about $\boldsymbol{B}_0$ at an angular frequency that depends on the local field strength $b_0 + l(x,y,z)$. If both $\boldsymbol{B}_1$ and $\tilde{\boldsymbol{G}}$ are simultaneously nonzero, then, starting from equilibrium, the solution of the Bloch equation, at the conclusion of the RF pulse, has a nontrivial spatial dependence. In other words, the flip angle becomes a function of the spatial variables. We return to this in a later section.

## A Basic Imaging Experiment

With these preliminaries, we can describe the basic measurements in magnetic resonance imaging. When exposed to $\boldsymbol{B}_0$, the sample becomes polarized at a rate determined by $T_1$. Once the sample is polarized, a $\boldsymbol{B}_1$-field, of the form given in [7] (with $\beta \equiv 0$), is turned on for a finite time $\tau$. This is called an RF excitation. For the purposes of this discussion, we suppose that the time is chosen so that $\theta(\tau)=90°$, see eqn [10]. As $\boldsymbol{B}_0$ and $\boldsymbol{B}_1$ are spatially homogeneous, the magnetization vectors within the object remain parallel throughout the RF excitation. At the conclusion of the RF excitation, $\boldsymbol{M}$ is orthogonal to $\boldsymbol{B}_0$.

After the RF is turned off, the vector field $\boldsymbol{M}(x,y,z;t)$ precesses about $\boldsymbol{B}_0$, in phase with the angular velocity $\omega_0$. The transverse component of $\boldsymbol{M}$ decays exponentially. If we normalize the time so that $t=0$ corresponds to the conclusion of the RF pulse, then, in the laboratory frame,

$$\boldsymbol{M}(x,y,z;t) = \frac{\chi\omega_0\rho(x,y,z)}{\gamma}\Big[\mathrm{e}^{-t/T_2}\cos\omega_0 t, \; \mathrm{e}^{-t/T_2}\sin\omega_0 t, (1-\mathrm{e}^{-t/T_1})\Big] \quad [12]$$

Recall Faraday's law: a changing magnetic field induces an electromotive force (EMF) in a loop of wire according to the relation

$$\mathrm{EMF}_{\text{loop}} \propto \frac{\mathrm{d}\Phi_{\text{loop}}}{\mathrm{d}t} \quad [13]$$

Here $\Phi_{\text{loop}}$ denotes the flux of the field through the loop of wire (*see* Introductory Articles: Electromagnetism). The transverse components of $\boldsymbol{M}$ are a rapidly varying magnetic field, which, according to Faraday's law, induce a current in a loop of wire. In fact, by placing several such loops close to the sample we can measure a signal of the form

$$S_0(t) = \frac{\chi\omega_0^2\mathrm{e}^{\mathrm{i}\omega_0 t}}{\gamma}\int_{\text{sample}} \rho(x,y,z)\mathrm{e}^{-t/T_2(x,y,z)} \times b_{1\text{rec}}(x,y,z)\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \quad [14]$$

Here $b_{1\text{rec}}(x,y,z)$ quantifies the sensitivity of the detector to the precessing magnetization located at $(x,y,z)$. From $S_0(t)$ we easily obtain a measurement of the integral of the function $\rho b_{1\text{rec}}$. By using a carefully designed detector, $b_{1\text{rec}}$ can be taken to be a constant, and therefore we can determine the total spin density within the object of interest. For the rest of this article, we assume that $b_{1\text{rec}}$ is a constant. Note that the size of the measured signal is proportional to $\omega_0^2$, which is, in turn, proportional to $\|\boldsymbol{B}_0\|^2$. This explains, in part, why it is so useful to have a very strong $\boldsymbol{B}_0$-field. Though even with a 1.5 T magnet, the measured signal is only in the microwatt range (see Hoult and Lauterbur (1979) and Edelstein *et al.* (2004)).

Suppose that, at the end of the RF excitation, we turn on the gradient $\tilde{\boldsymbol{G}}$. As the magnetic field $\boldsymbol{B}=\boldsymbol{B}_0+\tilde{\boldsymbol{G}}$ now has a nontrivial spatial dependence, the precessional frequency of the spins, which equals

$\gamma\|\boldsymbol{B}\|$, also has a spatial dependence. In fact, assuming that $T_2$ is spatially independent, it follows from [11] that the measured signal would now be given by

$$S_G(t) \approx \frac{\chi b_{1\text{rec}}\omega_0^2 e^{-t/T_2} e^{i\omega_0 t}}{\gamma} \times \int_{\text{sample}} \rho(x,y,z) e^{2\pi i\langle(x,y,z),\boldsymbol{k}\rangle}\, dx\, dy\, dz \qquad [15]$$

Up to a constant, $e^{-i\omega_0 t}\, e^{-t/T_2} S_G(t)$ is simply the Fourier transform of $\rho$ at $\boldsymbol{k} = -t\gamma\boldsymbol{G}/2\pi$. By sampling in time and using a variety of different gradient vectors, we can sample the three-dimensional Fourier transform of $\rho$ in a neighborhood of 0. This suffices to reconstruct an approximation to $\rho$. In medical applications, $T_2$ is spatially dependent, which, as described later in the section "Contrast and resolution," provides a useful contrast mechanism.

Imagine that we collect samples of $\hat{\rho}(\boldsymbol{k})$ on a rectangular grid

$$\left\{ (j_x\Delta k_x, j_y\Delta k_y, j_z\Delta k_z): -\frac{N_x}{2} \le j_x \le \frac{N_x}{2}, -\frac{N_y}{2} \le j_y \le \frac{N_y}{2}, -\frac{N_z}{2} \le j_z \le \frac{N_z}{2} \right\}$$

Since we are sampling in the Fourier domain, the Nyquist sampling theorem implies that the sample spacing determines the spatial field of view from which we can reconstruct an artifact-free image: in order to avoid aliasing artifacts, the support of $\rho$ must lie in a rectangular region with side lengths $[\Delta k_x^{-1}, \Delta k_y^{-1}, \Delta k_z^{-1}]$, see Haacke *et al.* (1999), Epstein (2003), and Barrett and Myers (2004). In typical medical applications, the support of $\rho$ is much larger in one dimension than the others, and so it turns out to be impractical to use the simple data collection technique described above. Instead, the RF excitation takes place in the presence of nontrivial gradient fields, which allows for a spatially selective excitation: the magnetization in one region of space obtains a transverse component, while that in the complementary region is left in the equilibrium state. In this way, we can collect data from an essentially two-dimensional slice. This is described in the next section.

## Selective Excitation

As remarked above, practical imaging techniques do not excite all the spins in an object and directly measure samples of the three-dimensional Fourier transform. Rather, the spins lying in a slice are excited and samples of the two-dimensional Fourier transform are then measured. This process is called selective excitation and may be accomplished by applying the RF excitation with a gradient field turned on. With this arrangement, the strength of the static field, $\boldsymbol{B}_0 + \tilde{\boldsymbol{G}}$, varies with spatial position, hence the response to the RF excitation does as well. Suppose that $\tilde{\boldsymbol{G}} = (0, 0, \langle(x,y,z), \boldsymbol{G}\rangle)$ and set $f = [2\pi]^{-1}\gamma\langle(x,y,z), \boldsymbol{G}\rangle$. This is called the offset frequency, as it is the amount by which the local resonance frequency differs from the resonance frequency $\omega_0$ of the $\boldsymbol{B}_0$-field. The result of a selective RF excitation is described by a magnetization profile $\boldsymbol{m}^{\text{pr}}(f)$, which is a unit 3-vector-valued function of the offset frequency. A typical case would be

$$\boldsymbol{m}^{\text{pr}}(f) = \begin{cases} [0, 0, 1] & \text{for } f \notin [f_0, f_1] \\ [\sin\theta, 0, \cos\theta] & \text{for } f \in [f_0, f_1] \end{cases} \qquad [16]$$

The magnetization is flipped through an angle $\theta$, in regions of space where the offset frequency lies in the interval $[f_0, f_1]$ and is left in the equilibrium state otherwise.

Typically, the excitation step takes a few milliseconds and is much shorter than either $T_1$ or $T_2$; therefore, one generally uses the Bloch equation, without relaxation, in the discussion of selective excitation. In the rotating reference frame, the Bloch equation, without relaxation, takes the form

$$\frac{d\boldsymbol{m}(f;t)}{dt} = \begin{bmatrix} 0 & 2\pi f & -\gamma\beta \\ -2\pi f & 0 & \gamma\alpha \\ \gamma\beta & -\gamma\alpha & 0 \end{bmatrix} \boldsymbol{m}(f;t) \qquad [17]$$

The problem of designing a selective pulse is nonlinear. Indeed, the selective excitation problem can be rephrased as a classical inverse-scattering problem: one seeks a function $\alpha(t) + i\beta(t)$ with support in an interval $[t_0, t_1]$ so that, if $\boldsymbol{m}(f;t)$ is the solution to (17) with $\boldsymbol{m}(f;t_0) = [0, 0, 1]$, then $\boldsymbol{m}(f;t_1) = \boldsymbol{m}^{\text{pr}}(f)$. If one restricts attention to flip angles close to 0, then there is a simple linear model that can be used to find approximate solutions.

If the flip angle is close to zero, then $m_3 \approx 1$ throughout the excitation. Using this approximation, we derive the low-flip-angle approximation to the Bloch equation, without relaxation:

$$\frac{d(m_1 + im_2)}{dt} = -2\pi i f(m_1 + im_2) + i\gamma(\alpha + i\beta) \qquad [18]$$

From this approximation, we see that

$$\alpha(t) + i\beta(t) \approx \frac{\mathscr{F}(m_1^{\text{pr}} + im_2^{\text{pr}})(t)}{\gamma i} \quad \text{where } \mathscr{F}(h)(t) = \int_{-\infty}^{\infty} h(f) e^{-2\pi i f t}\, df \qquad [19]$$

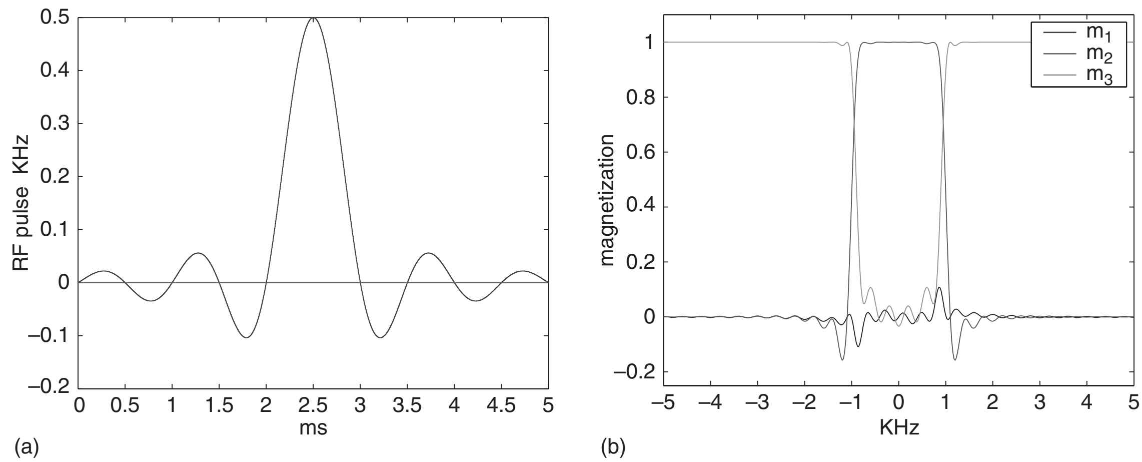



**Figure 1** A selective 90° pulse and profile designed using the linear approximation. (a) Profile of a 90° sinc-pulse. (b) The magnetization profile produced by the pulse in (a).

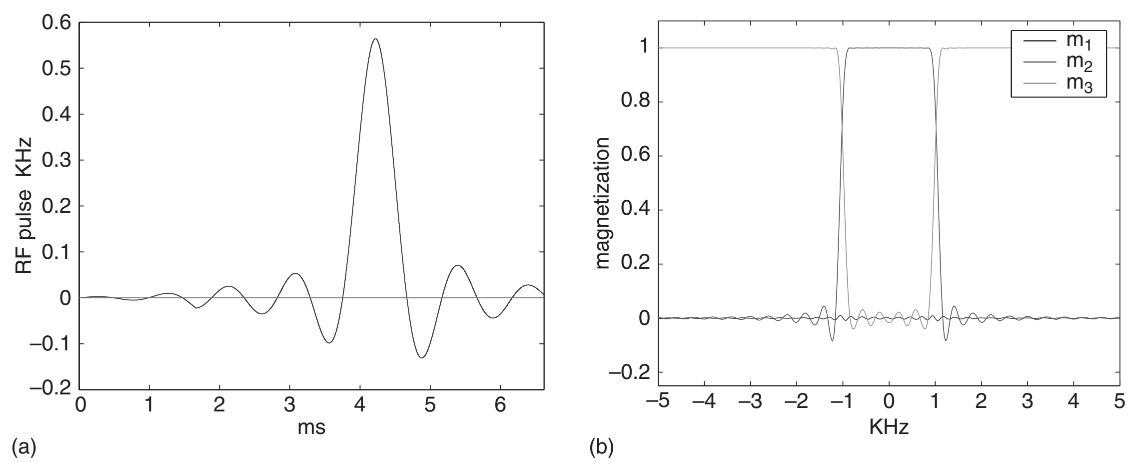



**Figure 2** A selective 90° pulse and profile designed using the inverse scattering approach. (a) Profile of a 90° inverse-scattering pulse. (b) The magnetization profile produced by the pulse in (a).

For an example such as in [16], $\theta$ close to zero, and $f_0 = -f_1$, we obtain

$$\alpha + \mathrm{i}\beta \approx \frac{\mathrm{i}\sin\theta \sin f_1 t}{\pi\gamma t} \qquad [20]$$

A pulse of this sort is called a sinc-pulse. A sinc-pulse is shown in **Figure 1a**, the result of applying it in **Figure 1b**. A more accurate pulse can be designed using the Shinnar–Le Roux algorithm (see Pauly *et al.* (1991) and Shinnar and Leigh (1989)), or the inverse scattering approach (see Epstein (2004)). An inverse-scattering 90°-pulse is shown in **Figure 2a** and the response in **Figure 2b**.

## Spin-Warp Imaging

In an earlier section we showed how NMR measurements could be used to measure the three-dimensional Fourier transform of $\rho$. In this section, we consider a more practical technique, that of measuring the two-dimensional Fourier transform of a "slice" of $\rho$. Applying a selective RF pulse, as described in the previous section, we can flip the magnetization in a region of space $z_0 - \Delta z < z < z_0 + \Delta z$, while leaving it in the equilibrium state outside a slightly larger region. Observing that a signal near the resonance frequency is only produced by isochromats whose magnetization has a nonzero transverse component, we can now measure samples of the two-dimensional Fourier transform of the function

$$\bar{\rho}_{z_0}(x, y) = \frac{1}{2\Delta z}\int_{z_0-\Delta z}^{z_0+\Delta z_0} \rho(x, y, z)\,\mathrm{d}z \qquad [21]$$

If $\Delta z$ is sufficiently small then $\bar{\rho}_{z_0}(x, y) \approx \rho(x, y, z_0)$.

In order to be able to use the fast Fourier transform (FFT) algorithm to do the reconstruction, it is very useful to sample $\widehat{\bar{\rho}}_{z_0}$ on a uniform grid. To that end, we use the gradient fields as follows: after the RF excitation we apply a gradient field of the form $\boldsymbol{G}_{\text{ph}} = (0, 0, -g_2 y + g_1 x)$ for a certain period of time $T_{\text{ph}}$. This is called a phase encoding gradient. At the conclusion of the phase encoding gradient, the transverse components of the magnetization from the excited spins has the form

$$\boldsymbol{m}^{\|}(x, y) \propto \mathrm{e}^{-2\pi \mathrm{i}(k_y y - k_x x)} \bar{\rho}_{z_0}(x, y) \qquad [22]$$

where $(k_x, k_y) = [2\pi]^{-1} \gamma T_{\text{ph}}(-g_1, g_2)$. At time $T_{\text{ph}}$, we turn off the $y$-component of $\boldsymbol{G}_{\text{ph}}$ and reverse the polarity of the $x$-component. At this point, we begin to measure the signal. We get samples of $\widehat{\bar{\rho}}(k, k_y)$ where $k$ varies from $-k_{x\,\max}$ to $k_{x\,\max}$. By repeating this process with the strength of the $y$-phase encoding gradient being stepped through a sequence of uniformly spaced values, $g_2 \in \{n\Delta g_y\}$, and collecting samples at a uniformly spaced set of times, we collect the set of samples

$$\left\{ \widehat{\bar{\rho}}_{z_0}(m\Delta k_x, n\Delta k_y): -\frac{N_x}{2} \le m \le \frac{N_x}{2}, -\frac{N_y}{2} \le n \le \frac{N_y}{2} \right\} \qquad [23]$$

The gradient $\boldsymbol{G}_{\text{fr}} = (0, 0, -g_1 x)$, left "on" during signal acquisition, is called a frequency encoding gradient. While there is no difference, mathematically, between the phase encoding and frequency encoding steps, there are significant practical differences. This approach to sampling is known as spin-warp imaging; it was introduced in Edelstein *et al.* (1980). The steps of this experiment are summarized in a pulse sequence timing diagram, shown in **Figure 3**. This graphical representation for the steps followed in a magnetic resonance imaging experiment is ubiquitous in the literature.

To avoid aliasing artifacts, the sample spacings $\Delta k_x$ and $\Delta k_y$ must be chosen so that the excited portion of the sample is contained in a region of size $\Delta k_x^{-1} \times \Delta k_y^{-1}$. This is called the field of view or FOV. Since we can only collect the signal for a finite period of time, the Fourier transform $\widehat{\bar{\rho}}(k_x, k_y)$ is sampled at frequencies lying in a rectangle with vertices $(\pm k_{x\,\max}, \pm k_{y\,\max})$, where

$$k_{x\,\max} = \frac{N_x \Delta k_x}{2}, \quad k_{y\,\max} = \frac{N_y \Delta k_y}{2} \qquad [24]$$

The maximum frequencies sampled effectively determine the resolution available in the reconstructed image. Heuristically, this resolution limit equals half the shortest measured wavelength:

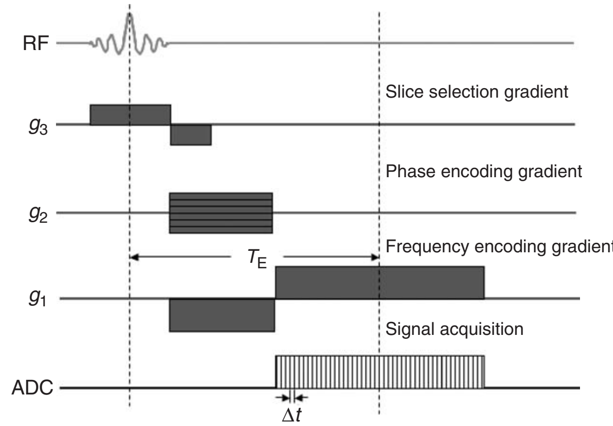



**Figure 3** Pulse timing diagram for spin-warp imaging. During the positive lobe of the frequency encoding gradient, the analog-to-digital converter (ADC) collects samples of the signal produced by the rotating transverse magnetization.

$$\Delta x \approx \frac{1}{2k_{x\,\max}} = \frac{FOV_x}{N_x}$$
$$\Delta y \approx \frac{1}{2k_{y\,\max}} = \frac{FOV_y}{N_y} \qquad [25]$$

Whether one can actually resolve objects of this size in the reconstructed image depends on other factors such as the available contrast and the signal-to-noise ratio (SNR). We consider these factors in the final sections.

## Signal-to-Noise Ratio

At a given spatial resolution, image quality is largely determined by SNR and the contrast between the different materials making up the imaging object. SNR in MRI is defined as the voxel signal amplitude divided by the noise standard deviation. The noise in the NMR signal, in general, is Gaussian distributed with zero mean. Ignoring contributions from quantization, for example, due to limitations of the analog-to-digital converter, the noise voltage of the signal can be ascribed to random thermal fluctuations in the receive circuit (see Edelstein (1986)). The variance is given by

$$\sigma^2_{\text{thermal}} = 4k_{\text{B}} TR\Delta\nu \qquad [26]$$

where $k_{\text{B}}$ is Boltzmann's constant, $T$ the absolute temperature, $R$ the effective resistance (resulting from both receive coil, $R_{\text{c}}$ and object, $R_{\text{o}}$), and $\Delta\nu$ the receive bandwidth. Both $R_{\text{c}}$ and $R_{\text{o}}$ are frequency dependent, with $R_{\text{c}} \propto \omega^{1/2}$, and $R_{\text{o}} \propto \omega$. Their relative contributions to overall circuit resistance depend in a complicated manner on coil geometry, and the imaging object's shape, size, and conductivity

(see Chen and Hoult (1989)). Hence, at high magnetic field, and for large objects, as in most medical applications, the resistance from the object dominates and the noise scales linearly with frequency. Since the signal is proportional to $\omega^2$, in MRI, the SNR increases in proportion to the field strength.

As the reconstructed image is complex valued, it is customary to display the magnitude rather than the real component. This, however, has some consequences on the noise properties. In regions where the signal is much larger than the noise, the Gaussian approximation is valid. However, in regions where the signal is low, rectification causes the noise to assume a Raleigh distribution. Mean and standard deviation can be calculated from the joint probability distribution:

$$P(N_\mathrm{r}, N_\mathrm{i}) = \frac{1}{2\pi\sigma^2}\mathrm{e}^{-(N_\mathrm{r}^2+N_\mathrm{i}^2)/2\sigma^2} \qquad [27]$$

where $N_\mathrm{r}$ and $N_\mathrm{i}$ are the noise in the real and imaginary channels, respectively. When the signal is large compared to noise, one finds that the variance $\sigma_m^2 = \sigma^2$. In the other extreme of nearly zero signal, one obtains for the mean:

$$\widehat{S} = \sigma\sqrt{\pi/2} \cong 1.253\sigma \qquad [28]$$

and, for the variance:

$$\sigma_m^2 = 2\sigma^2(1-\pi/4) \cong 0.655\sigma^2 \qquad [29]$$

Of particular practical significance is the SNR dependence on the imaging parameters. The voxel noise variance is reduced by the total number of samples collected during the data acquisition process, that is,

$$\sigma_m^2 = \sigma_\mathrm{thermal}^2/N \qquad [30]$$

where $N = N_x N_y$ in a two-dimensional spin-warp experiment. Incorporating the contributions to thermal noise variance, other than bandwidth, into a constant

$$u = 4k_\mathrm{B}TR \qquad [31]$$

we obtain for the noise variance:

$$\sigma_m^2 = \frac{u\Delta\nu}{N_x N_y N_\mathrm{avg}} \qquad [32]$$

Here $N_\mathrm{avg}$ is the number of signal averages collected at each phase encoding step. We obtain a simple formula for SNR per voxel of volume $\Delta V$:

$$\mathrm{SNR} = C\tilde{\rho}\,\Delta V\sqrt{\frac{N_x N_y N_\mathrm{avg}}{u\Delta\nu}} = C\tilde{\rho}\,\Delta x\,\Delta y\,d_z\sqrt{\frac{N_x N_y N_\mathrm{avg}}{u\Delta\nu}} \qquad [33]$$

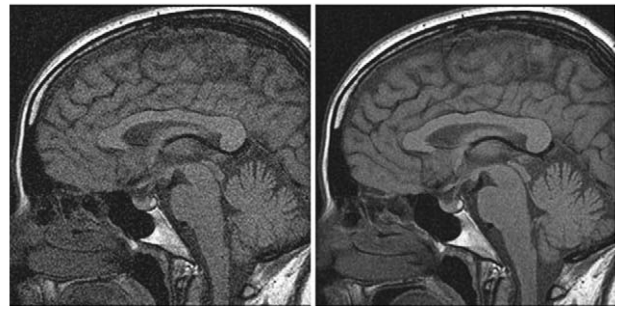



**Figure 4** $T_1$-weighted sagittal images through the midline of the brain: Image (b) has twice the SNR of image (a), showing improved conspicuity of small anatomic and low-contrast detail. The two images were acquired at 1.5 T field strength using two-dimensional spin-warp acquisition and identical scan parameters, except for $N_\mathrm{avg}$, which was 1 in (a) and 4 in (b).

where $\Delta x, \Delta y$ are defined in [25], $d_z$ is the thickness of the slab selected by the slice-selective RF pulse, and $\tilde{\rho}$ denotes the spin density weighted by effects determined by the (spatially varying) relaxation times $T_1$ and $T_2$ and the pulse sequence timing parameters. Figure 4 shows two images of the human brain obtained from the same anatomic location but differing in SNR.

## Contrast and Resolution

The single most distinctive feature of MRI is its extraordinarily large innate contrast. For two soft tissues, it can be on the order of several hundred percent. By comparison, contrast in X-ray imaging is a consequence of differences in the attenuation coefficients for two adjacent structures and is typically on the order of a few percent.

We have seen in the preceding sections that the physical principles underlying MRI are radically different from those of X-ray computed tomography, in that the signal elicited is generated by the spins themselves in response to an external perturbation. The contrast between two regions, $A$ and $B$, with signals $S_A$ and $S_B$, respectively, is defined as

$$C_{AB} = \frac{S_A - S_B}{S_A} \qquad [34]$$

If the only contrast mechanism were differences in the proton spin density of various tissues, then contrast would be on the order of 5–20%. In reality, it can be several hundred percent. The reason for this discrepancy is that the MR signal is acquired under nonequilibrium conditions. At the time of excitation, the spins have typically not recovered from the effect of the previous cycle's RF pulses, nor

is the signal usually detected immediately after its creation.

Typically, in spin-warp imaging, a spin-echo is detected as a means to alleviate spin coherence losses from static field inhomogeneity. A spin-echo is the result of applying an RF pulse that has the effect of taking $(m_1, m_2, m_3)$ to $(m_1, -m_2, -m_3)$. As such a pulse effects a 180° rotation of the $\hat{z}$-axis, it is also called a $\pi$-pulse. If, after such a pulse, the spins continue to evolve in the same environment then, following a certain period of time, the transverse components of the magnetization vectors throughout the sample become aligned. Hence a pulse of this type is also called a refocusing pulse. The time when all the transverse components are rephased is called the echo time, $T_E$.

The spin-echo signal amplitude for an RF pulse sequence $\pi/2 - \tau - \pi - \tau$, repeated every $T_R$ seconds, is approximately given by

$$S(t = 2\tau) \approx \rho(1 - e^{-T_R/T_1})e^{-T_E/T_2} \qquad [35]$$

This is a good approximation as long as $T_E << T_R$ and $T_2 << T_R$, in which case the transverse magnetization decays essentially to zero between successive pulse sequence cycles. In eqn [35], $\rho$ is voxel spin density and the echo time $T_E = 2\tau$. Empirically, it is known that tissues differ in at least one of the intrinsic quantities, $T_1, T_2$, or $\rho$. It, therefore, suffices to acquire images in such a manner that contrast is sensitive to one particular parameter. For example, a "$T_2$-weighted" image would be acquired with $T_E \sim T_2$ and $T_R >> T_1$ and, similarly, a "$T_1$-weighted" image with $T_R < T_1$ and $T_E << T_2$, with $T_1, T_2$ representing typical tissue proton relaxation times. Figure 5 shows two images obtained with the same scan parameters except for $T_R$ and $T_E$ illustrating the fundamentally different image contrasts that are achievable.

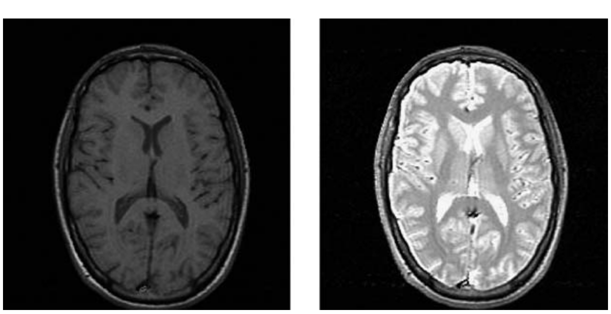
It is noteworthy that object visibility is not just determined by the contrast between adjacent structures but is also a function of the noise. It is, therefore, useful to define the contrast-to-noise ratio as

$$\mathrm{CNR}_{AB} = \frac{C_{AB}}{\sigma_{\mathrm{eff}}} \qquad [36]$$

where $\sigma_{\mathrm{eff}}$ is the effective standard deviation of the signal. Finally, it may be useful to reconstruct parametric images in which the pixel signal values represent any one of the intrinsic parameters. A $T_2$-image can be computed from eqn [35], for example, either analytically from two image data sets acquired with two different echo times, or from a series of $T_E$ values, obtained from a Carr–Purcell spin-echo train, using regression techniques (*see* Nuclear Magnetic Resonance and Haacke *et al.* (1999)).

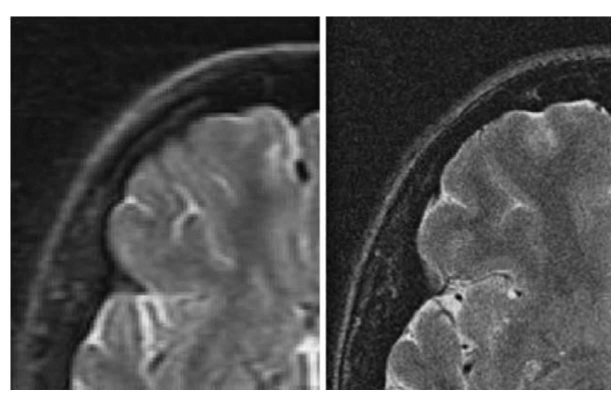
We have previously shown that the limiting resolution is given by $k_{\max}$, the largest spatial frequency sampled, see [25]. In reality, however, the actual resolution is always lower. For example, spin–spin ($T_2$) relaxation causes the signal to decay during the acquisition. In spin-warp imaging, this causes the high spatial frequencies to be further attenuated.

A further consequence of finite sampling is a ringing or Gibbs artifact that is most prominent at sharp intensity discontinuities. In practice, these artifacts are mitigated by applying an appropriate apodizing filter to the data. Figure 6 shows a portion of a brain image obtained at two different resolutions. In Figure 6b, the total ***k***-space area covered was 16 times larger than for the acquisition of the image in a). Artifacts from finite sampling and blurring of fine detail such as cortical blood vessels are clearly visible in the low-resolution image. SNR, according to eqn [33], is reduced in the latter image by a factor of 4.



(a) (b)

**Figure 5** Dependence of image contrast on pulse sequence timing parameters: (a) $T_1$-weighted; (b) proton density-weighted.



(a) (b)

**Figure 6** Effect of ***k***-space coverage on spatial resolution in axial image of the brain: the field of view in both images was 20 cm and all scan parameters were the same except that (a) was acquired with $N_x = N_y = 128$ and (b) with $N_x = N_y = 512$.

*See also:* Nuclear Magnetic Resonance; Stochastic Resonance.

## Further Reading


Abragam A (1983) *Principles of Nuclear Magnetism*. Oxford: Clarendon.

Barrett HH and Myers KJ (2004) *Foundations of Image Science*. Hoboken: Wiley.

Bernstein MA, King KF, and Zhou XJ (2004) *Handbook of MRI Pulse Sequences*. London: Elsevier Academic Press.

Callaghan PT (1993) *Principles of Nuclear Magnetic Resonance Microscopy*. Oxford: Clarendon.

Chen C-N and Hoult DI (1989) *Biomedical Magnetic Resonance Technology*. Bristol: Adam Hilger.

Edelstein WA, Glover GH, Hardy C, and Redington R (1986) The intrinsic signal-to-noise ratio in NMR imaging. *Magnetic Resonance in Medicine* 3: 604–618.

Edelstein WA, Hutchinson JM, Johnson JM, and Redpath T (1980) Spin warp NMR imaging and applications to human whole-body imaging. *Physics in Medicine and Biology* 25: 751–756.

Epstein CL (2003) *Introduction to the Mathematics of Medical Imaging*. Upper Saddle River, NJ: Prentice-Hall.

Epstein CL (2004) Minimum power pulse synthesis via the inverse scattering transform. *Journal of Magnetic Resonance* 167: 185–210.

Ernst R, Bodenhausen G, and Wokaun A (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Oxford: Clarendon.

Haacke EM, Brown RW, Thompson MR, and Venkatesan R (1999) *Magnetic Resonance Imaging*. New York: Wiley-Liss.

Hoult D and Lauterbur PC (1979) The sensitivity of the zeugmatographic experiment involving human samples. *Journal of Magnetic Resonance* 34: 425–433.

Levitt MH (2001) *Spin Dynamics, Basics of Nuclear Magnetic Resonance*. Chichester: Wiley.

Pauly J, Le Roux P, Nishimura D, and Macovski A (1991) Parameter relations for the Shinnar–Le Roux selective excitation pulse design algorithm. *IEEE Transactions on Medical Imaging* 10: 53–65.

Shinnar M and Leigh J (1989) The application of spinors to pulse synthesis and analysis. *Magnetic Resonance in Medicine* 12: 93–98.

Slichter CP (1990) *Principles of Magnetic Resonance, 3rd enl. and upd. ed.*, Springer Series in Solid-State Sciences, vol. 1. Berlin–New York: Springer.

Wehrli FW (1995) From NMR diffraction and zeugmatography to modern imaging and beyond. In: *Progress in Nuclear Magnetic Resonance Spectroscopy* 28: 87–135.


# Magnetohydrodynamics

**C Le Bris**, CERMICS – ENPC, Champs Sur Marne, France



## The Basic Modeling

Magnetohydrodynamics (MHD) is the study of the interaction of (electro-) magnetic fields and conducting fluids. When a conducting fluid (e.g., a liquid metal, a weakly ionized gas, or a plasma) is placed within a magnetic field, two coupling phenomena appear: the electric currents modify the magnetic field, and the Lorentz forces due to the magnetic field modify the motion of the fluid. At the mathematical level, two sets of equations, very different in nature, are involved. The usual description of the hydrodynamics phenomena is most often that provided by the continuum mechanics for fluids, while the description of electromagnetic phenomena essentially proceeds from the Maxwell equations.

Either category of equations can be declined in a variety of models. The coupling between the two categories might also be accounted for at different levels of accuracy. For the sake of conciseness in such an expository survey, it is neither desirable nor doable to present all the possible set of equations and their possible coupling. The difficulty stems from the incredibly large spectrum of physical phenomena where MHD plays a role. A list of such phenomena includes

- astrophysical and geophysical applications (modeling of stars in the galactic field, of pulsars, of solar spots, of the flows in the earth's core, ...),
- advanced "terrestrial" applications such as the magnetic confinement of plasmas in controlled fusion, MHD propulsion engines for rockets, and
- industrial applications in the engineering world (electromagnetic pumping, metal forming, aluminum electrolysis, and many other metallurgical applications).

Due to this variety of physical situations, no unified setting can be presented with a satisfactory degree of details. We therefore mostly concentrate throughout this article on the MHD of conducting fluids that are homogeneous, incompressible, viscous, and Newtonian. This is often the case of liquid metals in many industrial processes. The equations manipulated will first be given in their most general form and then immediately adapted to the above context. For other contexts, the modeling follows the same pattern, but other variants of the general equations must be employed. The bibliography of this article contains such general information.

### The Hydrodynamics Description

The usual description for fluids follows from continuum mechanics. In this setting, the governing equation is the equation for the conservation of momentum

$$\frac{\partial(\rho \boldsymbol{u})}{\partial t} + \operatorname{div}(\rho \boldsymbol{u} \otimes \boldsymbol{u}) - \operatorname{div}\tau = \boldsymbol{f} \qquad [1]$$

where $\rho$ denotes the density of the fluid, $\boldsymbol{u}$ its velocity, $\tau$ the stress tensor, and $\boldsymbol{f}$ the density of volumic (or per unit volume) body forces applied to the fluid. For incompressible viscous Newtonian fluids, the stress/velocity relation reads

$$\tau = \eta(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^{\mathrm{T}}) - p\mathrm{Id} \qquad [2]$$

together with the constraint

$$\operatorname{div} \boldsymbol{u} = 0 \qquad [3]$$

on the velocity. Here, $\eta$ denotes the viscosity of the fluid, $p$ the pressure, and $A^{\mathrm{T}}$ denotes the transpose matrix of the matrix $A$. A third usual assumption is that the incompressible fluid is in addition homogeneous, that is,

$$\rho = \overline{\rho} = \text{constant} \qquad [4]$$

Equations [1]–[4] lead to the equations for conservation of momentum in the case of incompressible homogeneous viscous Newtonian fluid, that is, the incompressible Navier–Stokes equations

$$\begin{aligned} \overline{\rho}\frac{\partial \boldsymbol{u}}{\partial t} + \overline{\rho}\boldsymbol{u} \cdot \nabla \boldsymbol{u} - \eta \Delta \boldsymbol{u} + \nabla p = \boldsymbol{f} \\ \operatorname{div} \boldsymbol{u} = 0 \end{aligned} \qquad [5]$$

These equations are supplied with initial and boundary conditions on the velocity $\boldsymbol{u}$. At initial time, the velocity is assumed to be known $\boldsymbol{u}(t=0,\cdot) = \boldsymbol{u}_0$ on the whole domain occupied by the fluid $\Omega$, a domain that is supposed here not to vary in time (see, nevertheless, the section "The industrial production of aluminum" for a different setting). On the other hand, the boundary conditions on the boundary $\partial\Omega$ of $\Omega$ can be of various forms. For simplicity, the boundary is supposed regular, so that its unitary outward normal $\boldsymbol{n}_{\partial\Omega}$ can be unambiguously defined. The standard choice is to set Dirichlet conditions on the velocity $\boldsymbol{u} = \boldsymbol{u}_{\text{given}}$. In the following, we will assume for simplicity that the boundary condition is the homogeneous Dirichlet boundary condition $\boldsymbol{u} = 0$, as a superposition of the nonpenetration condition $\boldsymbol{u} \cdot \boldsymbol{n}_{\partial\Omega} = 0$ and the no-slip boundary condition $\boldsymbol{u} \times \boldsymbol{n}_{\partial\Omega} = 0$. One can also impose alternative boundary conditions, for example, involving the pressure.

### The Electromagnetic Description

Classical electromagnetism is described by the Maxwell equations. For the sake of consistency, we recall here that these are:

**The Maxwell–Ampère equation**

$$-\frac{\partial \boldsymbol{D}}{\partial t} + \operatorname{curl} \boldsymbol{H} = \boldsymbol{j} \qquad [6]$$

**The Maxwell–Coulomb equation**

$$\operatorname{div} \boldsymbol{D} = \rho_{\mathrm{c}} \qquad [7]$$

**The Maxwell–Faraday equation**

$$\frac{\partial \boldsymbol{B}}{\partial t} + \operatorname{curl} \boldsymbol{E} = 0 \qquad [8]$$

**The Maxwell–Gauss equation**

$$\operatorname{div} \boldsymbol{B} = 0 \qquad [9]$$

In the above equations, the three-dimensional vector fields $\boldsymbol{D}$, $\boldsymbol{B}$, $\boldsymbol{E}$, $\boldsymbol{H}$ denote the electric and magnetic inductions, and the electric and magnetic fields, respectively. On the other hand, the three-dimensional vector field $\boldsymbol{j}$ denotes the current density, and the scalar field $\rho_{\mathrm{c}}$ denotes the charge density. Inside an electrically conducting medium, the standard assumption of perfect medium consists in assuming the following relations:

$$\begin{aligned} \boldsymbol{D} &= \varepsilon \boldsymbol{E} \\ \boldsymbol{H} &= \frac{1}{\mu}\boldsymbol{B} \end{aligned} \qquad [10]$$

often called "constitutive laws," where $\varepsilon$ and $\mu$, respectively, denote the (electric) permittivity and the (magnetic) permeability of the medium. In the simple isotropic homogeneous case, both these parameters are scalar and constant. They are often expressed as

$$\begin{aligned} \varepsilon &= \varepsilon_{\mathrm{r}}\varepsilon_0 \\ \mu &= \mu_{\mathrm{r}}\mu_0 \end{aligned} \qquad [11]$$

where $\varepsilon_0, \mu_0$ are the permittivity and the permeability of the vaccum (that satisfy $\varepsilon_0\mu_0 = 1/c^2$, with $c$ denoting the speed of light), and $\varepsilon_{\mathrm{r}}$, $\mu_{\mathrm{r}}$ are the permittivity and the permeability relative to vaccum, or relative permittivity and relative permeability.

When collecting [6]–[9], together with [10], [11], one obtains the following general system of

Maxwell equations in a continuum (dielectric) medium:

$$
\begin{aligned}
-\frac{\partial(\varepsilon \boldsymbol{E})}{\partial t} + \operatorname{curl}\left(\frac{1}{\mu}\boldsymbol{B}\right) &= \boldsymbol{j} \\
\operatorname{div}(\varepsilon \boldsymbol{E}) &= \rho_{\mathrm{c}} \\
\frac{\partial \boldsymbol{B}}{\partial t} + \operatorname{curl} \boldsymbol{E} &= 0 \\
\operatorname{div} \boldsymbol{B} &= 0
\end{aligned} \qquad [12]
$$

This system is supplied with initial conditions on the fields $\boldsymbol{B}$ and $\boldsymbol{E}$. On the other hand, boundary conditions might be necessary when the equations are restricted to a bounded domain. The latter question, quite delicate, is postponed until next section.

### The MHD Coupling

For coupling systems [5] and [12], a threefold task is in order.

On the one hand, the body force term in [5] needs to be made precise, and this is completed by setting

$$\boldsymbol{f} = \boldsymbol{j} \times \boldsymbol{B} + \boldsymbol{f}_{\mathrm{ext}} \qquad [13]$$

The first term in the right-hand side is the Lorentz force, consequence of the electric current $\boldsymbol{j}$ running within the magnetic field $\boldsymbol{B}$, a force that influences the motion, along the velocity field $\boldsymbol{u}$, of the particles of the conducting fluid. The second term is due to possible external forces. A typical case for such forces is that of the gravity forces

$$\boldsymbol{f}_{\mathrm{ext}} = \bar{\rho}\boldsymbol{g} \qquad [14]$$

On the other hand, in order to be a mathematically closed system, the Maxwell system [12] needs to be complemented by Ohm's law, another type of constitutive relation, like [10], that now relates the current density $\boldsymbol{j}$ with the other fields. When dealing with MHD phenomena, Ohm's law most often reads in the form

$$\boldsymbol{j} = \sigma(\boldsymbol{E} + \boldsymbol{u} \times \boldsymbol{B}) \qquad [15]$$

where $\sigma$ denotes the electric conductivity of the fluid. The second term of [15] explicitly accounts for the deviation of the lines of electric current by the hydrodynamics flow. In some oversimplified situations, it can be neglected, leading to Ohm's law in the more usual form $\boldsymbol{j} = \sigma \boldsymbol{E}$, that is also valid for solid media. Most of the times the term $\boldsymbol{u} \times \boldsymbol{B}$ contains crucial information, and thus is not neglected.

System [5]–[12] now reads

$$
\begin{aligned}
\bar{\rho}\frac{\partial \boldsymbol{u}}{\partial t} + \bar{\rho}\boldsymbol{u}\cdot\nabla\boldsymbol{u} - \eta\Delta\boldsymbol{u} + \nabla p &= \boldsymbol{j}\times\boldsymbol{B} + \boldsymbol{f}_{\mathrm{ext}} \\
\operatorname{div} \boldsymbol{u} &= 0 \\
-\frac{\partial(\varepsilon \boldsymbol{E})}{\partial t} + \operatorname{curl}\left(\frac{1}{\mu}\boldsymbol{B}\right) &= \boldsymbol{j} \\
\operatorname{div} \boldsymbol{E} &= \frac{1}{\varepsilon}\rho_{\mathrm{c}} \\
\frac{\partial \boldsymbol{B}}{\partial t} + \operatorname{curl} \boldsymbol{E} &= 0 \\
\operatorname{div} \boldsymbol{B} &= 0 \\
\boldsymbol{j} &= \sigma(\boldsymbol{E} + \boldsymbol{u}\times\boldsymbol{B})
\end{aligned} \qquad [16]
$$

A third task is then in order.

Apart from the constitutive laws [10] and Ohm's law [15], the specificity of the Maxwell equations for conducting fluids, as opposed to the same equations written, for example, in the vacuum, resides in the possible need for supplying the system with *ad hoc* boundary conditions. Indeed, in their most general form, the Maxwell equations are valid in the whole physical space $\mathbb{R}^3$. On the other hand, as the goal here is to simulate an MHD fluid that most often occupies only a bounded domain $\Omega$ in $\mathbb{R}^3$, there is the need to adequately define the simulation domain.

A first possibility is to set the Maxwell equations in the whole space, while solving the hydrodynamics equation on the domain $\Omega$ occupied by the fluid. Regarding only the Maxwell equations [12], this seems to be the method of choice. But then there is the need for an extension of Ohm's law [15] outside the fluid domain. Notice indeed that $\boldsymbol{u}$ appears in [15]. In addition to this, the fact that the physical confinement device for the fluid is then embedded in the domain where the Maxwell equations are set may be the source of various difficulties, as such a device is often delicate to model and treat. Therefore, alternative tracks may be followed.

A second possibility is to restrict the Maxwell equation to a bounded domain. In turn, this option divides in two: taking as the domain for the Maxwell equations that occupied by the fluid, or choosing a domain larger than $\Omega$. We cannot discuss this choice without loss of generality, and refer the reader to the literature (see e.g., Gerbeau *et al.* (2005)). In either situation, boundary conditions are needed. We only consider the former for the sake of brevity.

A standard choice for the boundary conditions for [12] is the following:

$$
\begin{aligned}
\boldsymbol{E}\times\boldsymbol{n}_{\partial\Omega} &= \boldsymbol{k}\times\boldsymbol{n}_{\partial\Omega} \\
\boldsymbol{B}\cdot\boldsymbol{n}_{\partial\Omega} &= q
\end{aligned} \qquad [17]
$$

where $\boldsymbol{k}$ and $q$, respectively, are given vector and scalar functions on the boundary.

A fact that needs to be emphasized is that it is not so easy to design accurate boundary conditions, that is, evaluations of $\boldsymbol{k}$ or $g$, especially because accurate experimental measures of magnetic quantities are often delicate to obtain, especially in industrial environments.

### A Commonly Used Simplified MHD Coupling

For the terrestrial MHD applications that are the focus of the present article, a commonly used assumption is to neglect the first term $\partial(\varepsilon \boldsymbol{E})/\partial t$, often called the displacement current, in the Maxwell–Ampère equation [6], that is the first equation of [12] or the third of [16] above. Then system [16] can be reorganized, eliminating $\boldsymbol{E}$ and $\boldsymbol{j}$, and leaving aside the Maxwell–Faraday equation [8], Ohm's law [15], and the Maxwell–Coulomb equation [7]. The latter equations amount to defining, respectively, $\boldsymbol{E}$ from $\boldsymbol{B}, \boldsymbol{j}$ from $\boldsymbol{E}$ and $\boldsymbol{B}$, and $\rho_c$ from $\boldsymbol{E}$. One is left with the following system with the triple of unknown fields $(\boldsymbol{u}, p, \boldsymbol{B})$

$$\bar{\rho}\frac{\partial \boldsymbol{u}}{\partial t} + \bar{\rho}\boldsymbol{u}\cdot\nabla\boldsymbol{u} - \eta\Delta\boldsymbol{u} + \nabla p = \frac{1}{\mu}\operatorname{curl}\boldsymbol{B}\times\boldsymbol{B} + \boldsymbol{f}_{\text{ext}}$$
$$\operatorname{div}\boldsymbol{u} = 0 \qquad [18]$$

$$\frac{\partial \boldsymbol{B}}{\partial t} + \operatorname{curl}\left(\frac{1}{\sigma}\operatorname{curl}\frac{1}{\mu}\boldsymbol{B}\right) = \operatorname{curl}(\boldsymbol{u}\times\boldsymbol{B})$$
$$\operatorname{div}\boldsymbol{B} = 0$$

Correspondingly, the initial conditions are now only on the pair $(\boldsymbol{u}, \boldsymbol{B})$. Regarding the boundary conditions on $\boldsymbol{B}$, they can be derived from [17] using, for example, a homogeneous Dirichlet boundary condition on $\boldsymbol{u}$:

$$\operatorname{curl}\boldsymbol{B}\times\boldsymbol{n}_{\partial\Omega} = \tilde{\boldsymbol{k}}\times\boldsymbol{n}_{\partial\Omega}$$
$$\boldsymbol{B}\cdot\boldsymbol{n}_{\partial\Omega} = q \qquad [19]$$

Other simplifications of system [16] can be adopted, such as steady-state approximations. In particular, it is often considered that electromagnetic phenomena have characteristic times that are so short in comparison with the characteristic time of hydrodynamics phenomena that the Maxwell equations in their stationary form may be coupled to the time-dependent hydrodynamics equations, such as [5]. We refer to the "Further reading" section for further information along these lines (see e.g., Gerbeau *et al.* (2005)).

## The Mathematical Nature of the Equations

With a view to understand the mathematical nature of systems [16] and [18], we first briefly recall some mathematical facts concerning hydrodynamics, before focusing on the coupling with electromagnetics.

Regarding the incompressible Navier–Stokes equation, we recall that the state of the art of the mathematical knowledge heavily depends on the dimension of the ambient space. In dimension 2, solutions are unique and regular (they are said to be strong), for regular enough data of course. Unfortunately, as the focus is here on MHD and electromagnetism is fundamentally a three-dimensional phenomenon, only the three-dimensional case for the Navier–Stokes equation is relevant. Now, in the context of the Navier–Stokes equations alone, only the existence of weak solutions for large times, and the existence and uniqueness of strong solutions for small times are known. Whether or not there exists a unique strong solution for all time (of course again for sufficiently regular data) is an open problem, of outstanding difficulty, (see Temam 1995).

In the coupled setting examined here, there is no reason to expect a better situation. At best, one may hope for the same situation as that for the uncoupled case (Navier–Stokes equations alone). Regarding the existence and uniqueness of solutions, a commonly used strategy is that of regularization: the Cauchy problem is studied for regularized data, and then one passes to the limit in the regularization. In this latter step, the linear terms cause no difficulty, since they pass to the limit only using weak convergence. On the other hand, the main concern is always the treatment of the nonlinear terms, which require strong convergence. Here, for the Navier–Stokes equation in the MHD setting, the additional difficulty stems from the presence of the nonlinear term $\boldsymbol{j}\times\boldsymbol{B}$ on the right-hand side. The mathematical treatment of this nonlinear term calls for a compactness argument, which in turn requires obtaining some information on the fields $\boldsymbol{j}$ and $\boldsymbol{B}$, and their derivatives, from the Maxwell equations. In this respect, the situation is radically different for system [16] and for system [18]. Likewise, these two systems behave differently regarding the other nonlinear term of electromagnetic nature, namely $\boldsymbol{u}\times\boldsymbol{B}$ in Ohm's law, or $\operatorname{curl}(\boldsymbol{u}\times\boldsymbol{B})$ on the right-hand side of the equation in $\boldsymbol{B}$, respectively.

### The Hyperbolic Variant

Due to the presence of the Maxwell equations [12] in their general form, that is a hyperbolic form,

system [16] is indeed very difficult, from the standpoint of mathematical analysis.

In order to realize this, it suffices to recall that the first step in the proof of the existence of solution to such a system of equations is to write down an *a priori* energy estimate. It is a simple manipulation on [16] to show that, formally, a solution to [16] satisfies

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega}\bar{\rho}|\boldsymbol{u}|^2 + \eta\int_{\Omega}|\nabla\boldsymbol{u}|^2 = \int_{\Omega}(\boldsymbol{j}\times\boldsymbol{B})\cdot\boldsymbol{u} \qquad [20]$$

multiplying the Navier–Stokes equation by $\boldsymbol{u}$ and integrating over the domain $\Omega$, while, on the other hand,

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega}\varepsilon|\boldsymbol{E}|^2 + \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega}\frac{1}{\mu}|\boldsymbol{B}|^2 = -\int_{\Omega}\boldsymbol{j}\cdot\boldsymbol{E} \qquad [21]$$

multiplying the Maxwell–Ampère equation by $-\boldsymbol{E}$, the Maxwell–Faraday equation by $(1/\mu)\boldsymbol{B}$, integrating over $\Omega$, and summing up the two. Next, the right-hand side of [21] can be modified, accounting for Ohm's law:

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega}\varepsilon|\boldsymbol{E}|^2 + \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega}\frac{1}{\mu}|\boldsymbol{B}|^2 = -\int_{\Omega}\frac{1}{\sigma}|\boldsymbol{j}|^2 - \int_{\Omega}(\boldsymbol{j}\times\boldsymbol{B})\cdot\boldsymbol{u} \qquad [22]$$

Summing up [20] and [22] yields the energy estimate:

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega}\left(\bar{\rho}|\boldsymbol{u}|^2 + \varepsilon|\boldsymbol{E}|^2 + \frac{1}{\mu}|\boldsymbol{B}|^2\right) + \int_{\Omega}\frac{1}{\sigma}|\boldsymbol{j}|^2 + \eta\int_{\Omega}|\nabla\boldsymbol{u}|^2 = 0 \qquad [23]$$

Notice that, in the above, we set the external forces and all boundary conditions to zero, for the sake of simplicity.

Estimate [23] clearly indicates that we dispose of $L^\infty([0,T],L^2(\Omega))$ bounds on the vector fields $\boldsymbol{E}$ and $\boldsymbol{B}$ together with an $L^2([0,T]\times\Omega)$ bound on the current $\boldsymbol{j}$, and with the (classical) $L^\infty([0,T], L^2(\Omega))\cap L^2([0,T],H^1(\Omega))$ bounds on the velocity $\boldsymbol{u}$. In addition, $\operatorname{div}\boldsymbol{B}$ and, when assuming $\rho_c$ bounded, $\operatorname{div}\boldsymbol{E}$ are bounded in $L^\infty([0,T]\times\Omega)$. Unfortunately, these bounds do not allow for passing to the limit in the nonlinear term $\boldsymbol{j}\times\boldsymbol{B}$ on the right-hand side of the Navier–Stokes equation. In addition, there seems to be no way of deriving further energy estimates on system [16] that would provide with more *a priori* regularity on the fields $\boldsymbol{E},\boldsymbol{B}$, and $\boldsymbol{j}$. To date, system [16] presents an unsolved mathematical difficulty.

### The Parabolic Variant

On the other hand, system [18] is radically different in mathematical nature, because the Maxwell equations then reduce to a parabolic-type equation. The same manipulations as above, in order to establish *a priori* estimates on the solution of [18], now lead to

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\Omega}\left(\bar{\rho}|\boldsymbol{u}|^2 + \frac{1}{\mu}|\boldsymbol{B}|^2\right) + \int_{\Omega}\frac{1}{\sigma}\left|\operatorname{curl}\left(\frac{1}{\mu}\boldsymbol{B}\right)\right|^2 + \eta\int_{\Omega}|\nabla\boldsymbol{u}|^2 = 0 \qquad [24]$$

which, together with the divergence-free constraint on $\boldsymbol{B}$, yields $L^\infty([0,T],L^2(\Omega))\cap L^2([0,T],H^1(\Omega))$ bounds on both the velocity $\boldsymbol{u}$ and the magnetic field $\boldsymbol{B}$. These bounds now allow for passing to the limit in the terms $\operatorname{curl}\boldsymbol{B}\times\boldsymbol{B}$ and $\operatorname{curl}(\boldsymbol{u}\times\boldsymbol{B})$ on the right-hand side of the equations. This being established, the rest of the mathematical analysis is straightforward, and a theorem of existence and uniqueness of solutions can be proved. Like in the case of the Navier–Stokes equations alone, we have (in dimension 3) the existence of a global-in-time weak solution (i.e., for any $T,\boldsymbol{u}$ and $\boldsymbol{B}$ both $L^\infty([0,T],L^2(\Omega))\cap L^2([0,T],H^1(\Omega))$ satisfying the divergence-free constraint). No uniqueness of this weak solution is known. On the other hand, for sufficiently regular data, we have the existence of a local-in-time strong solution (i.e., for $T$ sufficiently small, $\boldsymbol{u}$ and $\boldsymbol{B}$ both $L^\infty([0,T],H^1(\Omega))\cap L^2([0,T], H^2(\Omega))$, and uniqueness of this strong solution in the class of weak solutions as long as it exists. We refer to Sermange and Temam, (1983) and Gerbeau *et al.* (2005).

At this stage, it is to be remarked that there is a formal similarity, at first sight at least, between the parabolic form of the Maxwell equations, namely

$$\begin{aligned}\frac{\partial\boldsymbol{B}}{\partial t} + \operatorname{curl}\operatorname{curl}\boldsymbol{B} &= \operatorname{curl}\boldsymbol{h}\\ \operatorname{div}\boldsymbol{B} &= 0\end{aligned} \qquad [25]$$

and the incompressible Navier–Stokes equation [5]. Note that indeed the curl operator in the first equation of [25] can be replaced by (minus) the Laplacian operator $-\Delta$, since $\operatorname{div}\boldsymbol{B}=0$. Actually, this formal similarity cannot be translated into mathematical arguments, simply because there is no pressure in [25]. In other terms, the divergence-free constraint $\operatorname{div}\boldsymbol{B}=0$ simply propagates in time in [25] (note that the right-hand side $\operatorname{curl}\boldsymbol{h}$ is also

divergence-free by construction), while on the other hand $\operatorname{div} \boldsymbol{u} = 0$ is enforced as a constraint in [5], the pressure playing the role of a Lagrange multiplier that adjusts itself in time in order to allow for $\boldsymbol{u}$ to be divergence-free.

Of course, as in the purely hydrodynamics case, much more can be said on the equations than simply establishing the existence and uniqueness of solutions. For instance, the long time limit of the solutions can be studied, etc.... For this and other issues, we refer to the "Further reading" section (Duvaut and Lions 1972a, b, Sermange and Temam 1983, Gerbeau *et al.* 2005).

## Numerical Issues

We concentrate again on system [18]. It is illustrative to mention that this system, when written in nondimensional variables, reads

$$\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u} \nabla \boldsymbol{u} - \frac{1}{Re} \Delta \boldsymbol{u} + \nabla p = S \operatorname{curl} \boldsymbol{B} \times \boldsymbol{B} + \boldsymbol{f}_{\text{ext}}$$
$$\operatorname{div} \boldsymbol{u} = 0$$

$$\frac{\partial \boldsymbol{B}}{\partial t} + \frac{1}{Re_{\text{mag}}} \operatorname{curl}(\operatorname{curl} \boldsymbol{B}) = \operatorname{curl}(\boldsymbol{u} \times \boldsymbol{B})$$
$$\operatorname{div} \boldsymbol{B} = 0$$

where $S$ is the coupling parameter, $Re$ is the (hydrodynamic) Reynolds number, and $Re_{\text{mag}}$ denotes the magnetic Reynolds number.

As expected, the numerical simulation of a system such as [18] superposes the difficulties of the hydrodynamics simulation of incompressible viscous fluids, and those faced when simulating the parabolic form of the Maxwell equations. Therefore, the goal is to efficiently combine the techniques employed to overcome either of them.

For incompressible fluid mechanics, the method of choice is the finite-element method for the discretization of differential operators in space. A typical discretization of eqn [5], called the "mixed" finite-element method, makes use of a pair of finite elements, one for the velocity, and one for the pressure. Other possibilities exist, that amount more or less in eliminating one unknown in a first stage and calculating the second one as a postprocessing task. The mixed formulation in the pair of unknowns $(\boldsymbol{u}, p)$ is however the most employed method to date, at least in the present setting. The finite-element space for the velocity is taken richer than that for the pressure: a possibility is, for example, to take the degree of the finite element for the velocity equal to the degree of the finite element for the pressure plus one. The heuristics for this is the fact that the velocity is derived twice in [5] while the pressure is only derived once. Of course, a mathematical ground for this is available, and a key issue is the "inf–sup" condition (also compatibility condition, or stability condition) that dictates the possible choice for finite-elements pairs, so that problem [5] is well posed at the discrete level. Typically, $Q2$ finite elements for the velocity can be combined with (continuous) $Q1$ finite elements for the pressure. An alternative choice is to ignore the inf–sup condition, adopting, for example, $Q1$ finite elements for both fields $\boldsymbol{u}$ and $p$, but this requires for a so-called stabilized formulation of [5] at the discrete level. The "Further reading" section provides details on the broad variety of techniques available in the field: Quarteroni and Valli (1997), Gerbeau *et al.* (2005).

On the other hand, the parabolic equation on $\boldsymbol{B}$ in [18] may be discretized with the same finite elements as those used for the velocity. The enforcement of the divergence constraint $\operatorname{div} \boldsymbol{B} = 0$ at the discrete level deserves some attention. Recall indeed that at the continuous level the divergence-free constraint is spontaneously propagated by the equation. At the discrete level, a crucial role in this respect is played by the weak formulation of the parabolic equation and an *ad hoc* account for the boundary condition [17].

For the sake of completeness, let us mention that an alternative strategy to the use of the finite elements that have been mentioned above (and that are called Lagrangian finite elements), is to use "edge elements." In some sense, the use of such elements simplifies the treatment of the boundary conditions [17], since they are very well adapted to their mathematical nature.

Note also that, in the vein of what is done for purely hydrodynamics flow simulations, stabilized finite-elements techniques have been developed for the MHD system [18], that allow for a discretization of the three unknown fields $(\boldsymbol{u}, p, \boldsymbol{B})$ over the same finite elements, for example, $Q1$.

When coupling the two discrete formulations for simulating the whole system [18], two main strategies can be adopted: one can either treat each of the two equations separately, independently describing the propagation of $\boldsymbol{u}$ and $\boldsymbol{B}$ forward in time, or one can address directly the coupled system of equations, describing the propagation of $\boldsymbol{u}$ and $\boldsymbol{B}$ in parallel.

The first option aims in particular at obtaining in the end small algebraic systems. An instance of such

a segregated algorithm reads, formally and setting all constants to unity for simplicity,

$$\begin{aligned}
&\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^n}{\Delta t}+\boldsymbol{u}^n\cdot\nabla\boldsymbol{u}^{n+1}-\Delta\boldsymbol{u}^{n+1}+\nabla p^{n+1}\\
&\quad=\operatorname{curl}\boldsymbol{B}^n\times\boldsymbol{B}^n+f_{\mathrm{ext}}\\
&\operatorname{div}\boldsymbol{u}^{n+1}=0\\
&\frac{\boldsymbol{B}^{n+1}-\boldsymbol{B}^n}{\Delta t}+\operatorname{curl}\operatorname{curl}\boldsymbol{B}^{n+1}\\
&\quad=\operatorname{curl}\left(\boldsymbol{u}^n\times\boldsymbol{B}^{n+1}\right)\\
&\operatorname{div}\boldsymbol{B}^{n+1}=0
\end{aligned}\qquad[26]$$

At each time step, the two independent subsystems are solved, providing with $\boldsymbol{u}^{n+1}$ and $\boldsymbol{B}^{n+1}$ for the next time step. The difficulty is that it is not possible, with such segregated algorithms, to reproduce the energy estimate [24] at the discrete level. Note that, at the continuous level, the estimate [24] is based upon a proper cancelation of the term $\int_\Omega(\boldsymbol{j}\times\boldsymbol{B})\cdot\boldsymbol{u}$ present on the two right-hand sides. Such a cancelation basically stems for a nonlinear interplay that cannot be present in a segregated iteration. Consequently, some spurious energy is created in the system simply by an inadequate iteration between the two equations. More precisely, the scheme obtained is at best only conditionally stable, that is, stable for small enough time steps, a condition that might be prohibitive when it is needed to simulate the MHD coupling over large times.

On the other hand, the other option consists in attacking the full system [18] directly:

$$\begin{aligned}
&\frac{\boldsymbol{u}^{n+1}-\boldsymbol{u}^n}{\Delta t}+\boldsymbol{u}^n\cdot\nabla\boldsymbol{u}^{n+1}-\Delta\boldsymbol{u}^{n+1}+\nabla p^{n+1}\\
&\quad=\operatorname{curl}\boldsymbol{B}^{n+1}\times\boldsymbol{B}^n+f_{\mathrm{ext}}\\
&\operatorname{div}\boldsymbol{u}^{n+1}=0\\
&\frac{\boldsymbol{B}^{n+1}-\boldsymbol{B}^n}{\Delta t}+\operatorname{curl}\operatorname{curl}\boldsymbol{B}^{n+1}\\
&\quad=\operatorname{curl}\left(\boldsymbol{u}^{n+1}\times\boldsymbol{B}^n\right)\\
&\operatorname{div}\boldsymbol{B}^{n+1}=0
\end{aligned}\qquad[27]$$

Note that $\boldsymbol{B}^{n+1}$ is present in the equation yielding $\boldsymbol{u}^{n+1}$, while conversely $\boldsymbol{u}^{n+1}$ is present in that yielding $\boldsymbol{B}^{n+1}$. Then the coupled system admits at the discrete level an energy estimate analogous to the energy estimate [24], and the scheme is much more stable than the previous one, and even unconditionally stable. The price to pay is that the system is, at the algebraic level, of very large size. Being sparse, it may however be treated, for example, via a GMRES-type iterative solver.

Let us make a final remark on these numerical issues. In the whole generality, the numerical simulation of viscous fluids raises the question of large Reynolds numbers, that is, the question of the difficulties encountered in the numerical approximation for viscosities $\eta$ small with respect to the other dimensionalized parameters of the problem (density, velocity, and dimension of the domain). For such small viscosities, the flow becomes turbulent rather than laminar, and the broad range of length and energy scales in the flow turns out to be too difficult to capture numerically. A commonly used technique that is resorted to in such difficult cases is the turbulence modeling. Schematically, an averaged, or homogenized, model is derived on the basis of the Navier–Stokes equation, with the help of simplifying hypotheses, for example, in the form of closure relations. The quality of the simulation of the averaged model, and its relation to the true flow, heavily depends on these simplifying assumptions, which are in turn based upon a very deep understanding on the various physical phenomena at play. In the context of MHD flows, the situation is not clear, regarding such assumptions. It seems that there are no well-established models for turbulent MHD to date, at least from a rigorous viewpoint. In the absence of those, only a direct simulation of the Navier–Stokes equation seems possible.

## The Industrial Production of Aluminium

A prototypical example of an application of MHD to the industrial context is the production of aluminum in electrolysis cells. The numerical simulation of the process involves the simulation of the evolution of two layers of nonmiscible incompressible viscous fluids, separated by an interface, and covered by a free surface. A schematic description of an industrial cell indeed is the following. An electric current of $10^5$ A, or more, runs through two horizontal layers of conducting fluids: a bath of aluminum oxide above, and a layer of liquid aluminum below. The aluminum is produced by the reduction of the aluminum oxide, a reaction that only occurs at a temperature where aluminum is liquid. The high magnetic field induced by such a huge current produces in turn high Lorentz forces that influence the motion of either fluid. A key issue in the modeling, as well as in the technological control of the cell, is to understand the motion of the interface separating the two fluids. In a rough picture, this interface may be seen as a mobile

cathode, moving below a fixed anode. The equations describing the interior of the cell are basically of the type [18], with an important modification though: one needs to account for the presence of two fluids. They read:

$$\frac{\partial(\rho \boldsymbol{u})}{\partial t} + \operatorname{div}(\rho \boldsymbol{u} \otimes \boldsymbol{u}) - \operatorname{div}(\eta(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T))$$
$$= -\nabla p + \rho \boldsymbol{g} + \frac{1}{\mu} \operatorname{curl} \boldsymbol{B} \times \boldsymbol{B}$$
$$\operatorname{div} \boldsymbol{u} = 0$$
$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho \boldsymbol{u}) = 0 \qquad [28]$$
$$\frac{\partial \boldsymbol{B}}{\partial t} + \operatorname{curl}\left(\frac{1}{\mu\sigma} \operatorname{curl} \boldsymbol{B}\right) = \operatorname{curl}(\boldsymbol{u} \times \boldsymbol{B})$$
$$\operatorname{div} \boldsymbol{B} = 0$$

where $\boldsymbol{g}$ denotes the gravity field, we recall, and are supplied with the boundary conditions

$$\boldsymbol{u} = 0$$
$$\frac{1}{\mu\sigma} \operatorname{curl} \boldsymbol{B} \times \boldsymbol{n}_{\partial\Omega} = \boldsymbol{k} \times \boldsymbol{n}_{\partial\Omega} \qquad [29]$$
$$\boldsymbol{B}.\boldsymbol{n}_{\partial\Omega} = q$$

As opposed to [18], the density $\rho$ in [28] is no longer the constant $\overline{\rho}$, but is only piecewise constant, that is, constant in each (moving) subdomain occupied by each fluid. Likewise, the viscosity $\eta$, and the conductivity $\sigma$ are taken constant in each fluid, but with different values from one fluid to the other. While the density and the viscosity are only slightly different, the conductivity varies from many orders of magnitude, a discrepancy which ends up in some numerical stiffness of the equations. On the other hand, the permeability $\mu$ can be considered as constant throughout the domain, within a good level of approximation.

Mathematically, system [28] is an order of magnitude more difficult than [18]. We refer to Lions (1996) and Gerbeau and LeBris (1997) for some mathematical ingredients. A first major difficulty stems from the fact that the domain occupied by the fluids is no longer fixed. Notice that this difficulty already arises when simulating the MHD of one conducting fluid with a free surface. A second major difficulty is the discontinuity of the physical parameters at the interface, which causes a loss of regularity at the interface for the solution fields. The best result known to date is the existence of a global-in-time weak solution to [28]. Both mathematical difficulties above of course have significant numerical counterparts. A notable issue in such a simulation is how to handle the motion of the free interface, while ensuring that each fluid remains of constant mass (or volume) throughout the simulation. One of the most efficient method in such a context, introduced three decades ago, is the arbitrary-Lagrangian Eulerian (ALE) method. We refer to Brackbill and Pracht (1973) and Gerbeau *et al.* (2003a, b, 2005).

Apart from the direct numerical attack of system [28], which carries significant analytical and geometrical nonlinearities, there is the possibility, in particular in the industrial context, to derive a set of linearized equations at the vicinity of some equilibrium configuration of the system. This track has been extensively followed in the past and provides information that efficiently complement those provided by the much more satisfactory, but also more costly, nonlinear approach.

*See also:* Compressible Flows: Mathematical Theory; Computational Methods in General Relativity: The Theory; Fluid Mechanics: Numerical Methods; Newtonian Fluids and Thermohydraulics; Partial Differential Equations: Some Examples; Stability of Flows; Symmetric Hyperbolic Systems and Shock Waves; Topological Knot Theory and Macroscopic Physics.

## Further Reading

Bossavit A (1998) *Computational Electromagnetism. Variational Formulations, Complementarity, Edge elements*. San Diego: Academic Press.

Brackbill JU and Pracht WE (1973) An implicit, almost lagrangian algorithm for magnetohydrodynamics. *Journal of Computational Physics* 13: 455–4882.

Davidson PA (2001) *An Introduction to Magnetohydrodynamics*. Cambridge: Cambridge University Press.

Descloux J, Flueck M, and Romerio MV (1991) Modelling for instabilities in Hall Héroult cells: mathematical and numerical aspects. *Magnetohydrodyn. Process Metall.* 107–110.

Duvaut G and Lions J-L (1972a) *Les inéquations en mécanique et en physique*. Paris: Dunod.

Duvaut G and Lions J-L (1972b) Inéquations en thermoélasticité et magnétohydrodynamique. *Archives for Rational and Mechanical Analysis* 46: 241–279.

Gerbeau JF and LeBris C (1997) Existence of solution for a density-dependent magnetohydrodynamic equation. *Advances in Differential Equations* 2(3): 427–452.

Gerbeau JF, LeBris C, and Lelièvre T (2003a) Modeling and simulation of the industrial production of aluminium: the nonlinear approach. *Computers & Fluids* 33: 801–814.

Gerbeau JF, LeBris C, and Lelièvre T (2003b) Simulations of MHD flows with moving interfaces. *Journal of Computational Physics* 184(1): 163–191.

Gerbeau JF, LeBris C, and Lelièvre T (2005) Mathematical methods for the magnetohydrodynamics of liquid metals. In: *Methods and Industrial Applications*, Oxford University Press (to appear).

Hughes WF and Young FJ (1966) *The Electromagnetodynamics of Fluids*. New York: Wiley.

LaCamera AF, Ziegler DP, and Kozarek RL (1992) Magnetohydrodynamics in the Hall Héroult process, an overview. *Light Metals* 1179–1186.

Lions JL (1969) Quelques méthodes de résolution des problèmes aux limites non linéaires. *Etudes Mathématiques*. Paris: Dunod.

Lions PL (1996) *Mathematical Topics in Fluid Mechanics: Volume 1, Incompressible Models*. New York: Oxford University Press.
Moreau RJ (1991) *Magnetohydrodynamics*. Dordrecht: Kluwer Academics Publishers.
Quarteroni A and Valli A (1997) *Numerical Approximation of Partial Differential Equations*. Berlin: Springer.
Sele T (1977) Instabilities of the metal surface in electrolytic cells. *Light Metals* 1: 7–24.
Sermange M and Temam R (1983) Some mathematical questions related to the MHD equations. *Communications on Pure and Applied Mathematics* XXXVI: 635–664.
Temam R (1984) Navier–Stokes equations. *Theory and Numerical Analysis*, Studies in Mathematics and its Applications, vol. 2. Amsterdam: North-Holland.
Temam R (1995) Navier–Stokes Equations and Nonlinear Functional Analysis. *CBMS-NSF Regional Conference Series in Applied Mathematics*. Philadelphia: SIAM.

# Malliavin Calculus

**A B Cruzeiro**, University of Lisbon, Lisbon, Portugal



## Introduction

Malliavin calculus was initiated in 1976 with the work by P Malliavin (1978) and is essentially an infinite-dimensional differential calculus on the Wiener space. Its initial goal was to give conditions ensuring that the law of a random variable has a density with respect to Lebesgue measure as well as estimates for this density and its derivatives. When the random variables are solutions of stochastic differential equations (SDEs), these densities are heat kernels and Malliavin used Hörmander-type assumptions on the corresponding operators, thus providing a probabilistic proof of a Hörmander-type theorem for hypoelliptic operators.

The theory was much developed in the 1980s by Stroock, Bismut, and Watanabe, among others (the reader is referred to Nualart (1995) and Malliavin (1997)). In recent years, Malliavin calculus had great success in probabilistic numerical methods, mainly in the field of stochastic finance (Malliavin and Thalmaier 2005). However, the theory has also been applied to other fields of mathematics and physics, notably in statistical mechanics and statistical hydrodynamics (*see* Stochastic Hydrodynamics). In addition, one should remember that Wiener measure can be viewed as an "imaginary time" (but well-defined) counterpart of Feynman's "measure" for quantum systems. A stochastic calculus of variations for Wiener functionals could not be irrelevant to the path-integral approach to quantum theory.

Another field of application worth mentioning is the study of representations of stochastic oscillatory integrals with quadratic phase function and their stationary phase estimation. For this, complexification of the Wiener space must be properly defined (Malliavin and Taniguchi (1997)).

In order to give a flavor of what Malliavin calculus is all about, let us consider a second-order differential operator in $\mathbb{R}^d$ of the form

$$\mathcal{A} = \frac{1}{2}\sum_{i,j=1} a^{ij}\partial^2_{i,j} + \sum_i b^i \partial_i$$

with smooth bounded coefficients and such that the matrix $a$ is symmetric and non-negative, admitting a square root $\sigma$. The corresponding Cauchy value problem consists in finding a smooth solution $u(t,x)$ of

$$\frac{\partial u}{\partial t} = \mathcal{A}u, \quad u(0,.) = \phi(.) \qquad [1]$$

Then there exists a transition probability function $p(t,x,.)$ such that

$$u(t,x) = \int_{\mathbb{R}^d} \phi(y) p(t,x,\mathrm{d}y)$$

When $p(t,x,\mathrm{d}y) = p(t,x,y)\mathrm{d}y$, the function $p$ is the heat kernel associated to the operator $\mathcal{A}$, and from eqn [1] one may deduce Focker–Planck's equation for $p$.

Since Kolmogorov we know that it is possible to associate with such a second-order operator a stochastic family of curves like a deterministic flow is associated with a vector field. This stochastic family is a Markov process, $\xi_x(t)$, which is adapted to the increasing family $\mathcal{P}_\tau, \tau \in [0,1]$, of sigma-fields generated by the past events, that is, $u(\tau) \in \mathcal{P}_\tau$ for every $\tau$.

Itô calculus allows us to write the SDE satisfied by $\xi$:

$$\mathrm{d}\xi(t) = \sigma(\xi_x(t))\,\mathrm{d}W(t) + b(\xi_x(t))\,\mathrm{d}t, \quad \xi_x(0) = x \qquad [2]$$

where $W(t)$ stands for $\mathbb{R}^d$-valued Brownian motion (*see* Stochastic Differential Equations). Then $p$ is the image of the Wiener measure $\mu$ (the law of Brownian motion), namely $p(t,x,.) = \mu \circ \xi_x^{-1}(t)(.)$ and we have the representation

$$u(t,x) = E_\mu(\phi(\xi_x(t)))$$

The following criterion for absolute continuity of measures in finite dimensions holds:

**Lemma** *If $\gamma$ is a probability measure on $\mathbb{R}^d$ and, for every $f \in C_b^\infty$,*

$$\left| \int \partial_i f \, \mathrm{d}\gamma \right| \leq c_i \|f\|_\infty$$

*where $c_i$, $i = 1, \ldots, d$, are constants, then $\gamma$ is absolutely continuous with respect to Lebesgue measure.*

Now one can think about Wiener measure as an infinite (actually continuous) product of finite-dimensional Gaussian measures. Considering the toy model of the above-mentionned situation in one dimension, we replace Wiener measure by $\mathrm{d}\gamma(x) = (1/\sqrt{2\pi})\mathrm{e}^{-x^2/2}\,\mathrm{d}x$ and look at the process at a fixed time as a function $g$ on $\mathbb{R}$. In order to apply the lemma and study the law of $g$, one would write

$$\int (f' \circ g)\,\mathrm{d}\gamma = \int \frac{(f \circ g)'}{g'}\,\mathrm{d}\gamma$$

and then integrate by parts to obtain $\int (f \circ g)\rho\,\mathrm{d}\gamma$. A simple computation shows that $\rho(x) = (g'' + xg')/(g')^2$, and, in particular, that the nondegeneracy of the derivative of $g$ plays a role in the existence of the density.

To work with functionals on the Wiener space, one needs an infinite-dimensional calculus. Of course, other (Gateaux, Fréchet) calculi on infinite-dimensional settings are already available but the typical functionals we are dealing with, solutions of SDEs, are not continuous with respect to the underlying topology, nor even defined at every point, but only almost everywhere. Malliavin calculus, as a Sobolev differential calculus, requires very little regularity, given that there is no Sobolev imbedding theory in infinite dimensions.

## Differential Calculus on the Wiener Space

We restrict ourselves to the classical Wiener space, although the theory may be developed in abstract Wiener spaces, in the sense of Gross. For a description of this theory as well as of Segal's model developed in the 1950s for the needs of quantum field theory, the reader is referred to Malliavin (1997).

Let $\mathcal{H}$ be the Cameron–Martin space, $\mathcal{H} = \{h : [0,1] \to \mathbb{R}^d$ such that $\dot{h}$ is square integrable and $h(t) = \int_0^t \dot{h}(\tau)\mathrm{d}\tau\}$, which is a separable Hilbert space with scalar product $< h_1, h_2 > = \int_0^1 \dot{h}_1(\tau).\dot{h}_2(\tau)\mathrm{d}\tau$. The classical Wiener measure will be denoted by $\mu$; it is realized on the Banach space $X$ of continuous paths on the time interval [0,1] starting from zero at time zero, a space where $\mathcal{H}$ is densely imbedded. In finite dimensions, Lebesgue measure can be characterized by its invariance under the group of translations. In infinite dimensions there is no Lebesgue measure and this invariance must be replaced by quasi-invariance for translations of Wiener measures (Cameron–Martin admissible shifts). We recall that, if $h \in \mathcal{H}$, Cameron–Martin theorem states that

$$E_\mu(F(\omega + h)) = E_\mu\left(F(\omega)\exp\left(\int_0^1 \dot{h}(\tau)\,\mathrm{d}\omega(\tau) - \frac{1}{2}\int_0^1 |\dot{h}(\tau)|^2\,\mathrm{d}\tau\right)\right)$$

where $\mathrm{d}\omega$ denotes Itô integration.

For a cylindrical "test" functional $F(\omega) = f(\omega(\tau_1), \ldots, \omega(\tau_m))$, where $f \in C_b^\infty(\mathbb{R}^m)$ and $0 \leq \tau_1 \leq \cdots \leq \tau_m \leq 1$, the derivative operator is defined by

$$D_\tau F(\omega) = \sum_{k=1}^m \mathbf{1}_{\tau < \tau_k} \partial_k f(\omega(\tau_1), \ldots, \omega(\tau_m)) \qquad [3]$$

This operator is closed in $W_{2,1}(X; \mathbb{R})$, the completion of the space of cylindrical functionals with respect to the Sobolev norm

$$||F||_{2,1} = E_\mu ||F||^2 + E_\mu \int_0^1 |D_\tau F|^2\,\mathrm{d}\tau$$

Define $F$ to be $\mathcal{H}$-differentiable at $\omega \in X$ when there exists a linear operator $\nabla F(\omega)$ such that, for all $h \in \mathcal{H}$,

$$F(\omega + h) - F(\omega) = \langle \nabla F(\omega), h \rangle + o(||h||_H) \quad \text{as } ||h|| \to 0$$

Then $D_\tau$ disintegrates the derivative in the sense that

$$D_h F(\omega) \equiv \langle \nabla F(\omega), h \rangle = \int_0^1 D_\tau F(\omega).\dot{h}(\tau)\,\mathrm{d}\tau \qquad [4]$$

Higher ($r$)-order derivatives, as $r$-linear functionals, can be considered as well in suitable Sobolev spaces.

Denote by $\delta$ the $L^2_\mu$ adjoint of the operator $\nabla$, that is, for a process $u : X \to \mathcal{H}$ in the domain of $\delta$, the divergence $\delta(u)$ is characterized by

$$E_\mu(F\delta(u)) = E_\mu\left(\int_0^1 D_\tau F.\dot{u}(\tau)\,\mathrm{d}\tau\right) \qquad [5]$$

For an elementary process $u$ of the form $u(\tau) = \sum_j F_j(\tau \wedge \tau_j)$, where the $F_j$ are smooth random variables and the sum is finite, the divergence is

$$\delta(u) = \sum_j F_j \omega(\tau_j) - \sum_j \int_0^{\tau_j} D_\tau F_j\,\mathrm{d}\tau$$

The characterization of the domain of $\delta$ is delicate, since both terms in this last expression are not independently closable. It can be shown that $W_{1,2}(X;H)$ is in the domain of $\delta$ and that the following "energy" identity holds:

$$E_\mu(\delta(u))^2 = E_\mu||u||_H^2 + E_\mu \int_0^1 \int_0^1 D_\tau \dot{u}_\sigma . D_\sigma \dot{u}_\tau \, d\sigma \, d\tau$$

Notice that when $u$ is adapted to $\mathcal{P}_\tau$, Cameron–Martin–Girsanov theorem implies that the divergence coincides with Itô stochastic integral $\int_0^1 \dot{u}(\tau)\, d\omega(\tau)$ and, in this adapted case, the last term of the energy identity vanishes. We recover the well-known Itô isometry which is at the foundation of the construction of this integral. When the process is not adapted, the divergence turns out to coincide with a generalization of Itô integral, first defined by Skorohod.

The relation [5] is an integration-by-parts formula with respect to the Wiener measure $\mu$, one of the basic ingredients of Malliavin calculus. This formula is easily generalized when the base measure is absolutely continuous with respect to $\mu$.

Considering all functionals of the form $\mathcal{P}(\omega) = Q(\omega(\tau_1), \ldots, \omega(\tau_m))$ with $Q$ a polynomial on $\mathbb{R}^d$, the Wiener chaos of order $n$, $\mathcal{C}_n$, is defined as $\mathcal{C}_n = \mathcal{P}_n \bigotimes \mathcal{P}_{n-1}^\perp$, where $\mathcal{P}_n$ denote the polynomials on $X$ of degree $\leq n$. The Wiener-chaos decomposition $L_\mu^2(X) = \bigoplus_{n=0}^\infty \mathcal{C}_n$ holds. Denoting by $\Pi_n$ the orthogonal projection onto the chaos of order $n$, we have

$$\left\langle \nabla \left( \prod_{n+1} F \right), h \right\rangle = \prod_n (\langle \nabla F, h \rangle)$$

The derivative $D_u$ corresponds to the annihilation operator $A(u)$ and the divergence $\delta(u)$ to the creation operator $A^+(u)$ on bosonic Fock spaces.

An important result, known as the Clark–Bismut–Ocone formula, states that any functional $F \in W_{1,2}(X;\mathbb{R})$ can be represented as

$$F = E_\mu(F) + \int_0^1 E_\tau(D_\tau F)\, d\omega(\tau)$$

where $E_\tau$ denotes the conditional expectation with respect to the events prior to time $\tau$ (or, for short, the past $\mathcal{P}_\tau$ of $\tau$).

The Ornstein–Uhlenbeck generator (or minus number operator) is defined by $\mathcal{L}F = -\delta\nabla F$. On cylindrical functionals $F(\omega) = f(\omega(\tau_1), \ldots, \omega(\tau_m))$, it has the form

$$\mathcal{L}F(\omega) = \sum_{i,j} \sigma_i \wedge \sigma_j \partial_{i,j}^2 f(\omega(\tau_1), \ldots, \omega(\tau_m)) - \sum_j \omega(\tau_j) \partial_j f(\omega(\tau_1), \ldots, \omega(\tau_m))$$

where $i,j$ denote multi-dimensional ($d$) indexes.

As a multiplicative operator on the Wiener-chaos decomposition $\mathcal{L}F = -\sum_n n\Pi_n F$. It is the generator of a positive $\mu$-self-adjoint semigroup, the Ornstein–Uhlenbeck semigroup, formally given by $T_t F = \sum_n e^{-nt} \Pi_n F$. Another familiar representation of this semigroup is Mehler formula,

$$T_t F(\omega) = E_\mu \left( F\left( e^{-t}\omega + \sqrt{(1 - e^{-2t}} v \right) d\mu(v) \right)$$

Considering the map $X \to \mathbb{R}^m, \omega \to (\omega(\tau_1), \ldots, \omega(\tau_m))$, the image of this operator is the Ornstein–Uhlenbeck generator (corresponding to the Langevin equation) on $\mathbb{R}^m$ with Euclidean metric defined by the matrix $\tau_i \wedge \tau_j$.

The fundamental theorem concerning existence of the density laws of Wiener functionals is the following:

**Theorem** *Let $F$ be an $\mathbb{R}^d$-valued Wiener functional such that $F^i$ and $\mathcal{L}F^i$ belong to $L_\mu^4$ for every $i = 1, \ldots, d$. If the covariance matrix*

$$\langle \nabla F^i, \nabla F^j \rangle_{\mathcal{H}}$$

*is almost surely invertible, then the law of $F$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$.*

Under more regularity assumptions, smoothness of the density is also derived. On the other hand, the integrability assumptions on $\mathcal{L}$ can be replaced by integrability of the second derivatives, due to Krée–Meyer inequalities on the Wiener space.

We remark that, although equivalent, the initial formulation (Malliavin 1978) of Malliavin calculus was different, relying on the construction of the two-parameter process associated to $\mathcal{L}$ and on its properties. In the early 1980s, the theory was elaborated, the main applications being the study of heat kernels (cf., e.g., Stroock (1981), Ikeda and Watanabe (1989), and Bismut (1984)). Starting from an SDE [2], it is possible to apply these techniques to obtain existence and smoothness of the transition probability function $p(t,x,y)$ if the vector fields $Z_i = \sum_j \sigma^{ij}(\partial/\partial x_j)$ together with their Lie brackets generate the tangent space for "sufficientely many" (in terms of probability) paths. These results shed a new light on Hörmander theorem for partial differential equations.

## Quasi-Sure Analysis

Quasi-sure analysis is a refinement of classical probability theory and, generally speaking, replaces the fact that, due to Sobolev imbedding theorems, functions in finite dimensions belonging to Sobolev

classes are in fact smooth. We work in classical probability up to sets of probability zero; in quasi-sure analysis negligible sets are smaller and are those of capacity zero. This is the class of sets which are not charged by any measure of finite energy.

Under a nondegenerate map, Wiener measure and more general Gaussian measures may be disintegrated through a co-area formula. This principle, developed by Malliavin and co-authors (cf. Malliavin (1997) and references therein), implies that a property which is true quasi-surely will also hold true almost surely under conditioning by such a map. One can use this principle to study finer properties of SDEs. It was also used in M P Malliavin and P Malliavin (1990) to transfer properties from path to loop groups (*see* Measure on Loop Spaces). A pinned Brownian motion, for example, is well defined in quasi-sure analysis. It is possible to treat anticipative problems using quasi-sure analysis by solving the adapted problem after restriction of the solution to the finite-codimensional manifold which describes the anticipativity. These methods have also been applied to the computation of Lyapunov exponents of stochastic dynamical systems (Imkeller 1998). With a geometry of finite-codimensional manifolds of Wiener spaces well established, it is reasonable to think about applications to cases where such submanifolds correspond to level surfaces of invariant quantities for infinite-dimensional dynamical systems (cf. Cipriano (1999) for an example of such a situation in hydrodynamics).

The $(p,r)$-capacity of an open subset $O$ of the Wiener space is defined by

$$\mathrm{cap}_{p,r}(O) = \inf\{\|\phi\|^p_{W_{2r}}, \phi \geq 0, \phi \geq 1 \ \mu\text{-a.s. on } O\}$$

and, for a general set $B$, $\mathrm{cap}_{p,r}(B) = \inf\{\mathrm{cap}_{p,r}(O) : B \subset O, O \text{ open}\}$. A set is said to be slim if all its $(p,r)$-capacities are zero. For $\Phi \in W_\infty$, the space of functionals with every Malliavin derivative belonging to all $L^p_\mu$, there exists a redefinition of $\Phi$, denoted by $\Phi^*$, which is smooth and defined on the complement of a slim set.

Following Airault and Malliavin (1988), let $G \in W_\infty(X;\mathbb{R}^d)$ be of maximal rank and nondegenerate in the sense that the inverse of

$$(\det \Phi)^2(\omega) = \det(\langle \nabla\Phi^i(\omega), \nabla\Phi^j(\omega)\rangle)$$

belongs to $W_\infty$. Then for every functional $G \in W_\infty$, the measures $\mu \circ \Phi^{-1}$ and $(G\mu)\circ\Phi^{-1}$ are absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$ and have $C^\infty$ Radon–Nikodym derivatives. If

$$\rho(\lambda) = \frac{\mathrm{d}\mu\circ\Phi^{-1}}{\mathrm{d}\lambda} \quad \text{and} \quad \rho_G(\lambda) = \frac{\mathrm{d}(G\mu)\circ\Phi^{-1}}{\mathrm{d}\lambda}$$

the function $\lambda \to \rho_G(\lambda)/\rho(\lambda)$ will be smooth in the open set $\mathcal{O} = \{\lambda : \rho(\lambda) > 0\}$.

For every $\lambda \in \mathcal{O}$, it is possible to define (up to slim sets) a submanifold of the Wiener space of codimension $d$, $\mathcal{S}_\lambda = (\Phi^*)^{-1}(\lambda)$, as well as a measure $\mu_{\mathcal{S}}$ satisfying

$$\int_{\mathcal{S}_\lambda} G^* \,\mathrm{d}\mu_{\mathcal{S}}(\omega) = E^{\Phi(\omega)=\lambda}(G) = \frac{\rho_G(\lambda)}{\rho(\lambda)}$$

for every $G \in W_\infty$. This measure does not charge slim sets.

The area measure $\aleph$ on the submanifold $\mathcal{S}_\lambda$ is defined by

$$\int F^*\,\mathrm{d}\aleph = \rho(\lambda)\int F^*(\omega)\,\det(\langle\nabla\Phi^i(\omega), \nabla\Phi^j(\omega)\rangle)^{1/2}\,\mathrm{d}\mu_{\mathcal{S}}(\omega)$$

The following co-area formula on the Wiener space

$$\int_X f(\Phi(\omega))F(\omega)(\det\Phi)(\omega)\,\mathrm{d}\mu(\omega) = \int_{\mathbb{R}^d} f(\lambda)\int_{\mathcal{S}_\lambda} F^*(\omega)\,\mathrm{d}\aleph(\omega)\,\mathrm{d}\lambda$$

was proved in Airault and Malliavin (1988).

## Calculus of Variations in a Non-Euclidean Setting

Let $M$ be a $d$-dimensional compact Riemannian manifold with metric $\mathrm{d}s^2 = \sum_{i,j} g_{i,j}\,\mathrm{d}m^i\,\mathrm{d}m^j$. The Laplace–Beltrami operator is expressed in the local chart by

$$\Delta_M = g^{i,j}\frac{\partial^2 f}{\partial m^i \partial m^j} - g^{i,j}\Gamma^k_{i,j}\frac{\partial f}{\partial m^k}$$

where $\Gamma^k_{i,j}$ are the Christoffel symbols associated with the Levi-Civita connection. The corresponding Brownian motion $p_w$ is locally expressed as a solution of the SDE:

$$\mathrm{d}p^i(t) = a^{i,j}(p(t))\,\mathrm{d}W_j(t) - \tfrac{1}{2}g^{j,k}\Gamma^i_{j,k}(p(t))\,\mathrm{d}t$$

with $p(0) = m_0 \in M$ and where $a = \sqrt{g}$. Its law on the space of paths $\mathbb{P}(M) = \{p : [0,1] \to M, p \text{ continuous}, p(0) = m_0\}$ will be denoted by $\nu$.

How can we develop differential calculus and geometry on the space $\mathbb{P}(M)$? An infinite-dimensional local chart approach is delicate, due to the difficulty of finding an atlas in which the changes of charts preserve the measures. A possibility, developed in Cruzeiro and Malliavin (1996), consists in replacing the local chart approach by the Cartan-like methodology of moving frames. The canonical moving

frame in this framework is provided by Itô stochastic parallel transport. Nevertheless, a new difficulty arises: the parallel transport will not be differentiable in the Cameron–Martin sense described before.

Recall that a frame above $m$ is a Euclidean isometry $r:\mathbb{R}^d \to T_m(M)$ onto the tangent space. $O(M)$ denotes the collection of all frames above $M$ and $\pi(r)=m$ the canonical projection. $O(M)$ can be viewed as a parallelized manifold for there exist canonical differential forms $(\theta,\omega)$ realizing for every $r$ an isomorphism between $T_r(O(M))$ and $\mathbb{R}^d \times \mathrm{so}(d)$.

If $A_\alpha, \alpha=1,\ldots,d$, denote the horizontal vector fields, which are defined by $<\theta, A_\alpha> = \varepsilon_\alpha$, $<\omega, A_\alpha> = 0$, where $\varepsilon_\alpha$ are the vectors of the canonical basis of $\mathbb{R}^d$, then the horizontal Laplacian in $O(M)$ is the operator

$$\Delta_{O(M)} = \sum_{\alpha=1}^{d} A_\alpha^2$$

and we have $\Delta_{O(M)}(fo\pi) = (\Delta_M f)o\pi$. With the Laplacians on $M$ and on $O(M)$ inducing two probability measures, the canonical projection realizes an isomorphism between the corresponding probability spaces.

The Stratonovich SDE

$$\mathrm{d}r_\omega = \sum_\alpha A_\alpha(r_\omega)o\,\mathrm{d}\omega^\alpha, \quad r_\omega(0) = r_0$$

with $\pi(r_0)=m_0$ defines the lifting to $O(M)$ of the Itô parallel transport along the Brownian curve and we write $t^p_{\tau\leftarrow 0} r_0 = r_\omega(\tau)$. Itô map was defined by Malliavin as the map $I: X \to \mathbb{P}(M)$ given by

$$I(\omega)(\tau) = \pi(r_\omega(\tau))$$

This map is a.s. bijective and we have $\nu = \mu \circ I^{-1}$; therefore, it provides an isomorphism of measures from the curved path space to the "flat" Wiener space.

For a cylindrical functional $F = f(p(\tau_1),\ldots,p(\tau_m))$ on $\mathbb{P}(M)$, the derivatives are defined by

$$D_{\tau,\alpha}F(p) = \sum_{k=1}^{m} \mathbf{1}_{\tau<\tau_k}(t^p_{0\leftarrow\tau_k}(\partial_k F)|\varepsilon_\alpha)$$

The derivative operator is closable in a suitable Sobolev space.

It would be reasonable to think that the differentiable structure considered in the Wiener space would be conserved through the isomorphism $I$ and that the tangent space of $\mathbb{P}(M)$ would consist of transported vectors from the tangent space to $X$, namely Cameron–Martin vectors. Let us take a map $Z_p(\tau) \in T_{p(\tau)}(M)$ such that $z(\tau) = t^p_{0\leftarrow\tau} Z_p(\tau)$ belongs to the Cameron–Martin space $\mathcal{H}$.

In order to transfer derivatives to the Wiener space, we need to differentiate the Itô map. We have (Cruzeiro and Malliavin (1996)):

**Theorem** *The Jacobian matrix of the flow $r_0 \to r_\omega(\tau)$ is given by the linear map $J_{\omega,\tau} = (J^1_{\omega,\tau}, J^2_{\omega,\tau}) \in \mathrm{GL}(\mathbb{R}^d \times \mathrm{so}(d))$ defined by the system of Stratonovich SDE's*

$$\mathrm{d}_\tau J^1_{\omega,\tau} = \sum_{\alpha=1}^{d} \left(J^1_{\omega,\tau}\right)_\alpha o\,\mathrm{d}\omega_\alpha(\tau)$$

$$\mathrm{d}_\tau J^2_{\omega,\tau} = \sum_{\alpha=1}^{d} \Omega\left(J^1_{\omega,\tau}, \varepsilon_\alpha\right) o\,\mathrm{d}\omega_\alpha(\tau)$$

*where $\Omega$ denotes the curvature tensor of the underlying manifold read on the frame bundle.*

From this result we can deduce the behavior of the derivatives transferred to the Wiener space, a result whose origin is due to B Driver. We have, for a "vector field" $Z_p(\tau)$ on $\mathbb{P}(M)$ as above,

$$(D_Z F)oI = D_\xi(FoI)$$

with $\xi$ solving

$$\mathrm{d}\xi(\tau) = \dot{z}(\tau)\,\mathrm{d}\tau + \rho o\,\mathrm{d}\omega(\tau)$$

$$\mathrm{d}\rho(\tau) = \Omega(o\,\mathrm{d}\omega(\tau), z(\tau))$$

The process $\xi$ is no longer Cameron–Martin space valued. Nevertheless, it satisfies an SDE with an antisymmetric diffusion coefficient (given by the curvature) and therefore, by Levy's theorem, it still corresponds to a transformation of the Wiener space that leaves the measure quasi-invariant. We extend, accordingly, the notion of tangent space in the Wiener space to include processes of the form $\mathrm{d}\xi^\alpha = a^\alpha_\beta\,\mathrm{d}\omega^\beta + c^\alpha\,\mathrm{d}\tau$, with $a^\alpha_\beta + a^\beta_\alpha = 0$. These were called "tangent processes" in Cruzeiro and Malliavin (1996).

Another important consequence of the last theorem is the integration-by-parts formula in the curved setting, initially proved by Bismut (1984):

$$E_\nu(D_Z F) = E_\mu\left((FoI)\int_0^1 [\dot{z} + \tfrac{1}{2}\mathrm{Ricci}(z)]\,\mathrm{d}\omega(\tau)\right)$$

where Ricci is the Ricci tensor of $M$ read on the frame bundle.

## Some Applications

We already mentioned that Malliavin calculus has been applied to various domains connected with physics. We shall describe here some of its relations with elementary quantum mechanics.

Feynman gave a path space formulation of quantum theory whose fundamental tool is the concept of transition element of a functional $F(\omega)$ between any two $L^2$-states $\psi_s$ and $\phi_u$, for paths $\omega$ defined on a time interval $[s, u]$:

$$
\begin{aligned}
<F>_S &\equiv <\phi|F|\psi>_S \\
&= \iiint_\Omega \psi_s(x) \exp\left(\frac{\mathrm{i}}{\hbar} S_L(\omega, u-s)\right) \\
&\quad \times F(\omega)\bar{\phi}_u(z)\mathcal{D}\omega \, \mathrm{d}x \, \mathrm{d}z
\end{aligned} \qquad [6]
$$

This is a shorthand for the time discretization version along broken paths $\omega$ interpolating linearly between point $x_j = \omega(t_j)$, $t_j = j(u-s)/N$, $j = 0, 1, \ldots, N$. In [6] $\hbar$ is Planck's constant and $S = S_L$ denotes the action functional with Lagrangian $L$ of the underlying classical system. For a particle with mass $m$ in a scalar potential $V$ on the real line,

$$
S_L(\omega, u-s) = \int_s^u \left(\frac{m}{2}\dot{\omega}^2(\tau) - V(\omega(\tau))\right) \mathrm{d}\tau \qquad [7]
$$

The "$\mathcal{D}\omega$" of [6] is used as a Lebesgue measure, although there is no such thing in infinite dimensions. More generally, the construction of measures or integrals on the various path spaces required for general quantum systems is still nowadays a field of investigation.

When $F = 1$ and $\bar{\phi}_u$ (the complex conjugate of $\phi_u$) reduces to a Dirac mass at $z$, [6] is the path-integral representation of the solution $\psi(x, u)$ of the initial-value problem in $L^2$:

$$
\begin{aligned}
\mathrm{i}\hbar \frac{\partial \psi}{\partial u} &= H\psi \\
\psi(x, s) &= \psi_s(x)
\end{aligned} \qquad [8]
$$

where $H = -(\hbar^2/2)\Delta + V$ and when $S_L$ is as in [7]. Feynman's framework is time symmetric on $I$: when $\psi_s = \delta_x$ (still for $F = 1$), [6] provides a path-integral representation of the solution of the final-value problem for $\bar{\phi}(z, s)$.

According to Feynman, "it would be possible to use the integration-by-parts formula

$$
\left\langle \frac{\delta F}{\delta \omega(s)} \right\rangle = -\frac{\mathrm{i}}{\hbar} \left\langle F \frac{\delta S}{\delta \omega(s)} \right\rangle \qquad [9]
$$

as a starting point to define the laws of quantum mechanics" (Feynman and Hibbs 1965, p. 173). The functional derivative corresponds to variations of the underlying paths in directions $\delta\omega$ and

$$
\delta F = \int \frac{\delta F}{\delta \omega(s)} \delta\omega(s) \, \mathrm{d}s
$$

to an $L^2$ analog of [4].

Its first consequence, when $F = 1$, is the path space counterpart of Newton's law, in the elementary case [7],

$$
<m\ddot{\omega}>_{S_L} = -<\nabla V(\omega)>_{S_L} \qquad [10]
$$

where the left-hand side involves a time discretization of the second derivative. When $F(\omega) = \omega(t)$, Feynman obtains the path space version of Heisenberg commutation relation between position and momentum observables:

$$
\begin{aligned}
&\left\langle \omega(t) \frac{\omega(t) - \omega(t-\epsilon)}{\epsilon} \right\rangle_{S_L} - \left\langle \frac{\omega(t+\epsilon) - \omega(t)}{\epsilon} \omega(t) \right\rangle \\
&= \mathrm{i}\frac{\hbar}{m}
\end{aligned} \qquad [11]
$$

and from this the crucial fact that "quantum mechanical paths are very irregular. However, these irregularities average out over a reasonable length of time to produce a reasonable drift or average velocity" (Feynman and Hibbs 1965, p. 177).

A probabilistic interpretation (cf. Cruzeiro and Zambrini (1991)) of Feynman's calculus uses (Bernstein) diffusion processes solving the SDE

$$
\mathrm{d}z(t) = \left(\frac{\hbar}{m}\right)^{1/2} \mathrm{d}W(t) + \frac{\hbar}{m} \nabla \log \eta(z(t), t) \, \mathrm{d}t \qquad [12]
$$

where the drift stems from a positive solution of the Euclidean version of the above final-value problem for $\bar{\phi}$,

$$
\begin{aligned}
\hbar \frac{\partial \eta}{\partial t} &= H\eta \\
\eta(x, u) &= \eta_u(x)
\end{aligned} \qquad [13]
$$

For any regular function $f$, we can make sense of the "continuous limit"

$$
\begin{aligned}
\mathbb{D} f(z(t), t) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} E_t[&f(z(t+\epsilon), t+\epsilon) \\
&- f(z(t), t))]
\end{aligned} \qquad [14]
$$

where $E_t$ denotes conditional expectation with respect to the past $\mathcal{P}_t$ and check, indeed, that

$$
\mathbb{D} z(t) = \frac{\hbar}{m} \nabla \log \eta(z(t), t)
$$

is Feynman's "reasonable drift." Using Feynman–Kac formula, one shows that the diffusions [12] have laws which are absolutely continuous with respect to the Wiener measure of parameter $\hbar/m$, with Radon–Nikodym density given by

$$
\rho(z) = \frac{\eta(z(u), u)}{\eta(z(s), s)} \exp\left(-\frac{1}{\hbar} \int_0^1 V(z(\tau)) \, \mathrm{d}\tau\right)
$$

We can, therefore, use Malliavin calculus on the path space of these diffusions and the associated integration-by-parts formula to make sense of [9] and all its consequences.

The probabilistic counterpart of the time symmetry of Feynman's framework is interesting: Heisenberg's original argument to deny the existence of quantum trajectories (1927) was that any position can be associated with two velocities. Feynman's interpretation [11] and the definition [14] suggest that this has to do with a past or future conditioning at time $t$. Indeed, there is another description of diffusions $z(t)$ with respect to a family of future $\sigma$-fields, using the Euclidean version of the initial-value problem for $\psi$, underlying [6]. Another drift built on the model of the drift in [12] results, and Feynman's commutation relation [11] becomes rigorous (without, of course, the factor $i$).

We refer to Cruzeiro and Zambrini (1991) for a development of this approach using Malliavin calculus.

*See also:* Euclidean Field Theory; Functional Integration in Quantum Physics; Measure on Loop Spaces; Stochastic Differential Equations; Stochastic Hydrodynamics.

## Further Reading

Airault H and Malliavin P (1988) Intégration géometrique sur l' espace de Wiener. *Bulletin des Sciences Mathématiques* 112(2): 3–52.

Bismut JM (1984) Large Deviations and the Malliavin Calculus. *Birkhäuser Progress in Mathematics* 45.

Cipriano F (1999) The two-dimensional Euler equation: a statistical study. *Communications in Mathematical Physics* 201(1): 139–154.

Cruzeiro AB and Malliavin P (1996) Renormalized differential geometry on path space: structural equation, curvature. *Journal of Functional Analysis* 139: 119–181.

Cruzeiro AB and Zambrini JC (1991) Malliavin calculus and Euclidean quantum mechanics. I. Functional Calculus. *Journal of Functional Analysis* 96: 62–95.

Feynman RP and Hibbs AR (1965) Quantum Mechanics and Path Integrals. *International Series in Pure and Applied Physics.* New York: McGraw-Hill.

Ikeda N and Watanabe S (1989) *Stochastic Differential Equations and Diffusion Processes*. North-Holland Mathematical Library, vol. 24. Amsterdam–New York–Oxford: North-Holland.

Imkeller P (1998) The smoothness of laws of random flags and Oseledets spaces of linear stochastic differential equations. *Potential Analysis* 9: 321–349.

Malliavin P (1978) Stochastic calculus of variations and hypoelliptic operators. *Proceedings of the International Symposium on Stochastic Differential Equations*, *Kyoto 1976*, pp. 195–263. Wiley.

Malliavin P (1997) Stochastic Analysis. *Springer Verlag Grund. der Math. Wissen*. vol. 313. Berlin: Springer.

Malliavin MP and Malliavin P (1990) Integration on loop groups. I. Quasi-invariant measures. *Journal of Functional Analysis* 93: 207–237.

Malliavin P and Taniguchi S (1997) Analytic functions, Cauchy formula, and stationary phase on a real abstract Wiener space. *Journal of Functional Analysis* 143: 470–528.

Malliavin P and Thalmaier A (2005) *Stochastic Calculus of Variations in Mathematical Finance*. Berlin: Springer.

Nualart D (1995) The Malliavin Calculus and Related Topics, *Springer Verlag Problem and Its Applications*. New York–Berlin–Heidelberg: Springer.

Stroock DW (1981) The Malliavin Calculus and its application to second order parabolic differential equations I (II). *Mathematical Systems Theory* 14: 25–65 and 141–171.

**Marsden–Weinstein Reduction** *see* Cotangent Bundle Reduction; Poisson Reduction; Symmetry and Symplectic Reduction

**Maslov Index** *see* Optical Caustics; Semiclassical Spectra and Closed Orbits; Stationary Phase Approximation

# Mathai–Quillen Formalism

**S Wu**, University of Colorado, Boulder, CO, USA



## Introduction

Characteristic classes play an essential role in the study of global properties of vector bundles. Particularly important is the Euler class of real orientable vector bundles. A de Rham representative of the Euler class (for tangent bundles) first appeared in Chern's generalization of the Gauss–Bonnet theorem to higher dimensions. The representative is the Pfaffian of the curvature, whose cohomology class does not depend on the choice of connections. The Euler class of a vector bundle is also the obstruction to the existence of a nowhere-vanishing section. In fact, it is the Poincaré dual of the zero set of any section which intersects the zero section transversely. In the case of tangent bundles, it counts (algebraically) the zeros of a vector field on the manifold. That this is equal to the Euler characteristic number is known as the Hopf theorem. Also significant is the Thom class of a vector bundle: it is the Poincaré dual of the zero section in the total space. It induces, by a cup product, the Thom isomorphism between the cohomology of the base space and that of the total space with compact vertical support. Thom isomorphism also exists and plays an important role in $K$-theory.

Mathai and Quillen (1986) obtained a representative of the Thom class by a differential form on the total space of a vector bundle. Instead of having a compact support, the form has a nice Gaussian peak near the zero section and exponentially decays along the fiber directions. The pull-back of Mathai–Quillen's Thom form by any section is a representative of the Euler class. By scaling the section, one obtains an interpolation between the Pfaffian of the curvature, which distributes smoothly on the manifold, and the Poincaré dual of the zero set, which localizes on the latter. This elegant construction proves to be extremely useful in many situations, from the study of Morse theory, analytic torsion in mathematics to the understanding of topological (cohomological) field theories in physics.

In this article, we begin with the construction of Mathai–Quillen's Thom form. We also consider the case with group actions, with a review of equivariant cohomology and then Mathai–Quillen's construction in this setting. Next, we show that much of the above can be formulated as a "field theory' on a superspace of one fermionic dimension. Finally, we present the interpretation of topological field theories using the Mathai–Quillen formalism.

## Mathai–Quillen's Construction

### Berezin Integral and Supertrace

Let $V$ be an oriented real vector space of dimension $n$ with a volume element $\nu \in \wedge^n V$ compatible with the orientation. The "Berezin integral" of a form $\omega \in \wedge^* V^*$ on $V$, denoted by $\int^{\mathrm{B}} \omega$, is the pairing $\langle \nu, \omega \rangle$. Clearly, only the top degree component of $\omega$ contributes. For example, if $\sigma \in \wedge^2 V^*$ is a 2-form, then

$$\int^{\mathrm{B}} \mathrm{e}^{\sigma} = \begin{cases} \left\langle \nu, \dfrac{\sigma^{\wedge(n/2)}}{(n/2)!} \right\rangle, & \text{if } n \text{ is even} \\ 0, & \text{if } n \text{ is odd} \end{cases}$$

If $V$ has a Euclidean metric $(\cdot\,,\cdot)$, then $\nu$ is chosen to be of unit norm. If $\Sigma \in \mathrm{End}(V)$ is skew-symmetric, then $(1/2)(\cdot\,, \Sigma\,\cdot)$ is a 2-form and, if $n$ is even, the Pfaffian of $\Sigma$ is

$$\mathrm{Pf}(\Sigma) = \int^{\mathrm{B}} \exp\left(\frac{1}{2}(\cdot\,, \Sigma\,\cdot)\right)$$

The Berezin integral can be defined on elements in a graded tensor product $\wedge^* V^* \hat{\otimes} A$, where $A$ is any $\mathbb{Z}_2$-graded commutative algebra. For example, if we consider the identity operator $\boldsymbol{x} = \mathrm{id}_V$ as a $V$-valued function on $V$, then $\mathrm{d}\boldsymbol{x}$ is a 1-form on $V$ valued in $V$, and $(\mathrm{d}\boldsymbol{x}, \cdot)$ is a 1-form valued in $V^*$. Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis of $V$ and write $\boldsymbol{x} = x^i e_i$, where $x^i$ are the coordinate functions on $V$. We let

$$u(\boldsymbol{x}) = \frac{(-1)^{n(n+1)/2}}{(2\pi)^{n/2}} \int^{\mathrm{B}} \exp\left(-\frac{1}{2}(\boldsymbol{x}, \boldsymbol{x}) - (\mathrm{d}\boldsymbol{x}, \cdot)\right)$$

The integrand is in $\Omega^*(V) \hat{\otimes} \wedge^* V^*$. The result is

$$u(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}, \boldsymbol{x})\right) \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n \quad [1]$$

a Gaussian $n$-form whose (usual) integration on $V$ is 1.

Let $\mathrm{Cl}(V)$ be the Clifford algebra of $V$. For any orthonormal basis $\{e_i\}$, let $\gamma^i$ be the corresponding generators of $\mathrm{Cl}(V)$ and let $\gamma = e_i \otimes \gamma^i \in V \otimes \mathrm{Cl}(V)$. For any $\omega \in \wedge^k V^*$, we have

$$\omega(\gamma, \ldots, \gamma) = \frac{1}{k!} \omega_{i_1 \cdots i_k} \gamma^{i_1} \cdots \gamma^{i_k} \in \mathrm{Cl}(V)$$

If $n$ is even, the Clifford algebra has a unique $\mathbb{Z}_2$-graded irreducible spinor representation $S(V) = S^+(V) \oplus S^-(V)$. For any element $a \in \mathrm{Cl}(V)$, the

supertrace is $\operatorname{str} a = \operatorname{tr}_{S^+(V)} a - \operatorname{tr}_{S^-(V)} a$. If $\Sigma \in \operatorname{End}(V)$ is skew-symmetric, then

$$\operatorname{str}\exp\left(\frac{\sqrt{-1}}{4}(\gamma, \Sigma\gamma)\right) = \hat{A}(\Sigma)^{-1/2}\operatorname{Pf}(\Sigma)$$

where

$$\hat{A}(\Sigma) = \det\left(\frac{\Sigma/2}{\sinh(\Sigma/2)}\right)$$

More generally, supertrace can be defined on $\operatorname{Cl}(V)\hat{\otimes} A$ for any $\mathbb{Z}_2$-graded commutative algebra $A = A^+ \oplus A^-$. If $\Sigma$ is skew-symmetric and $\alpha \in V^* \otimes A^-$, then

$$\operatorname{str}\exp\left(\frac{\sqrt{-1}}{4}(\gamma, \Sigma\gamma) + \left(\frac{\sqrt{-1}}{2}\right)^{1/2}\alpha(\gamma)\right) = \hat{A}(\Sigma)^{-1/2}\int^{\mathrm{B}} \exp\left(\frac{1}{2}(\cdot\,, \Sigma\,\cdot) + \alpha\right) \qquad [2]$$

## Representatives of the Euler and Thom Classes

Let $M$ be a smooth manifold and let $\pi : E \to M$ be an oriented real vector bundle of rank $r$. Suppose $E$ has a Euclidean structure $(\cdot\,,\cdot)$ and $\nabla$ is a compatible connection. The curvature $R \in \Omega^2(M, \operatorname{End}(E))$ is skew-symmetric, and hence $(\cdot\,, R\,\cdot) \in \Omega^2(M, \wedge^2 E^*)$. A de Rham representative of the Euler class of $E$ is

$$e_\nabla(E) = \frac{1}{(2\pi)^{r/2}}\int^{\mathrm{B}} \exp\left(\frac{1}{2}(\cdot\,, R\,\cdot)\right) = \operatorname{Pf}\left(\frac{R}{2\pi}\right) \qquad [3]$$

Here, the Berezin integration is fiberwise in $E$: it is the pairing between the integrand and the unit section $\nu$ of the trivial line bundle $\wedge^r E$ that is consistent with the orientation of $E$. The de Rham cohomology class of [3] is independent of the choice of $(\cdot\,,\cdot)$ or $\nabla$.

Let $s$ be a section of $E$. Following Berline *et al.* (1992) and Zhang (2001), we consider

$$\mathfrak{s}_{\nabla,s} = \tfrac{1}{2}(s, s) + (\nabla s, \cdot) + \tfrac{1}{2}(\cdot\,, R\,\cdot) \qquad [4]$$

a differential form on $M$ valued in $\wedge^* E^*$. Mathai–Quillen's representative of the Euler class is

$$e_{\nabla,s}(E) = \frac{(-1)^{r(r+1)/2}}{(2\pi)^{r/2}}\int^{\mathrm{B}} \mathrm{e}^{-\mathfrak{s}_{\nabla,s}} \qquad [5]$$

One can show that $e_{\nabla,s}(E)$ is closed and that as $\beta$ varies, the cohomology class of $e_{\nabla,\beta s}(E)$ does not change. By taking $\beta \to 0$, the de Rham class of $e_{\nabla,s}(E)$ is equal to that of $e_\nabla(E)$ when $r$ is even. The form $e_{\nabla,\beta s}(E)$ provides a continuous interpolation between [3] and the limit as $\beta \to \infty$, when the form is concentrated on the zero locus of the section $s$. In fact, the Euler class is the Poincaré dual to the homology class represented by $s^{-1}(0)$. Hence, if $n \geq m$ and if $\omega \in \Omega^{n-m}(M)$ is closed, we have

$$\int_M \omega \wedge e_{\nabla,s}(E) = \int_{s^{-1}(0)} \omega \qquad [6]$$

when $s$ intersects the zero section transversely.

To obtain Mathai–Quillen's representative of the Thom class, we consider the pullback of $E$ to $E$ itself. The bundle $\pi^* E \to E$ has a tautological section $\boldsymbol{x}$. Applying [5] to this setting, we get

$$\tau_\nabla(E) = \frac{(-1)^{r(r+1)/2}}{(2\pi)^{r/2}}\int^{\mathrm{B}} \exp\left(-\frac{1}{2}(\boldsymbol{x}, \boldsymbol{x}) - (\nabla \boldsymbol{x}, \cdot) - \frac{1}{2}(\cdot\,, R\,\cdot)\right) \qquad [7]$$

where $(\cdot\,,\cdot), \nabla$, and $R$ are understood to be the pullbacks to $\pi^* E$. This is a closed form on the total space of $E$. Moreover, its restriction to each fiber is the Gaussian form [1]. The cohomology groups of differential forms with exponential decay along the fibers are isomorphic to those with compact vertical support or the relative cohomology groups $H^*(E, E\backslash M)$. Here $M$ is identified with its image under the inclusion $i : M \to E$ by the zero section. Under the above isomorphism, the cohomology class represented by $\tau_\nabla(E)$ coincides with the Thom class $\tau(E) = i_* 1 \in H^r(E, E\backslash M)$ defined topologically. For any section $s \in \Gamma(E)$, we have $e_{\nabla,s}(E) = s^* \tau_\nabla(E)$.

## Character Form of the Thom Class in *K*-Theory

Let $E = E^+ \oplus E^-$ be a $\mathbb{Z}_2$-graded vector bundle over $M$. The spaces $\Omega^*(M, E)$, $\Gamma(\operatorname{End}(E))$ and $\Omega^*(M)\hat{\otimes}\Gamma(\operatorname{End}(E))$ are also $\mathbb{Z}_2$-graded. The action of $\alpha \hat{\otimes} T \in \Omega^*(M)\hat{\otimes}\Gamma(\operatorname{End}(E))$ on $\beta \otimes s \in \Omega^*(M, E)$ is

$$\alpha \hat{\otimes} T : \beta \otimes s \mapsto (-1)^{|T|\,|\beta|}(\alpha \wedge \beta) \otimes (Ts)$$

The supertrace of $A \in \Gamma(\operatorname{End}(E))$ is $\operatorname{str} A = \operatorname{tr}_{E^+} A - \operatorname{tr}_{E^-} A$; it extends $\Omega^*(M)$-linearly to $\operatorname{str} : \Omega^*(M)\hat{\otimes}\Gamma(\operatorname{End}(E)) \to \Omega^*(M)$. Let $\nabla$ be a connection on $E$ preserving the grading. $\nabla$ is an odd operator on $\Omega^*(M, E)$. If $L \in \Gamma(\operatorname{End}(E)^-)$ is odd, then $D = \nabla + L$ is called a "superconnection" on $E$; the "curvature" $D^2 = R + \nabla L + L^2 \in (\Omega^*(M) \otimes \Gamma(\operatorname{End}(E)))^+$ is even. With the superconnection, the Chern character of the virtual vector bundle $E^+ \ominus E^-$ can be represented by

$$\operatorname{ch}_{\nabla,L}(E^+, E^-) = \operatorname{str}\exp\left(\frac{\sqrt{-1}}{2\pi}D^2\right) \qquad [8]$$

It is a closed form on $M$ and its de Rham cohomology class is independent of the choice of $\nabla$ or $L$. If $L$ is invertible everywhere on $M$ and the eigenvalues of $\sqrt{-1}L^2$ are negative, then [8] is exact:

$$\mathrm{ch}_{\nabla,L}(E^+,E^-) = -\frac{\sqrt{-1}}{2\pi}\,\mathrm{d}\int_1^\infty \mathrm{str}\left(\exp\left(\frac{\sqrt{-1}}{2\pi}(\nabla+\beta L)^2\right)L\right)\mathrm{d}\beta$$

Now let $E$ be an oriented real vector bundle of rank $r=2m$ over $M$ with a Euclidean structure $(\cdot\,,\cdot)$. Suppose further that $E$ has a spin structure. The associated spinor bundle $S(E)=S^+(E)\oplus S^-(E)$ is a graded complex vector bundle over $M$. For any section $s\in\Gamma(E)$, let $c(s)\in\Gamma(\mathrm{End}(E)^-)$ be the Clifford multiplication on $E$. Then for any $s,s'\in\Gamma(E)$, we have $\{c(s),c(s')\}=-2(s,s')$. Given a connection $\nabla$ on $E$ preserving $(\cdot\,,\cdot)$, the induced spinor connection $\nabla^S$ on $S(E)$ preserves the grading. If $R$ is the curvature of $\nabla$, that of $\nabla^S$ is $R^S=-(1/4)\,(\gamma,R\gamma)$, where $\gamma$ is now a section of $E\otimes\mathrm{Cl}(E)$. For any $s\in\Gamma(E)$, consider the superconnection

$$D_s=\nabla^S+\left(\frac{\pi}{\sqrt{-1}}\right)^{1/2}c(s)$$

The Chern character form [8] of $S^+(E)\ominus S^-(E)$ is, using [2],

$$\mathrm{ch}_{\nabla,s}(S^+(E),S^-(E))=(-1)^m\hat{A}\left(\frac{R}{2\pi}\right)^{-1/2}e_{\nabla,s}(E) \qquad [9]$$

where $e_{\nabla,s}(E)$ is given by [5]. In cohomology groups, [9] reduces to

$$\mathrm{ch}(S^+(E))-\mathrm{ch}(S^-(E))=(-1)^m\hat{A}(E)^{-1/2}e(E)$$

If $M$ is noncompact and the norm of $s$ increases rapidly away from $s^{-1}(0)$, then both sides of [9] are differential forms that decay rapidly away from $s^{-1}(0)$ and can represent cohomology classes of such. As before, we take the pullback $\pi^*E$ with the tautological section $\boldsymbol{x}$. Then [9] becomes

$$\mathrm{ch}_\nabla(\pi^*S^+(E),\pi^*S^-(E)) = (-1)^m\pi^*\hat{A}\left(\frac{R}{2\pi}\right)^{-1/2}\tau_\nabla(E) \qquad [10]$$

where $\tau_\nabla(E)$ is given by [7]. Both sides of [10] are forms on $E$ that decays exponentially in the fiber directions; hence, it descends to an equality in $H^*(E,E\backslash M)$. In the relative $K$-group $K(E,E\backslash M)$, the pair $\pi^*S^\pm(E)$ with the isomorphism $c(\boldsymbol{x})$ away form the zero section is, up to a factor of $(-1)^m$, the $K$-theoretic Thom class $i_!1\in K(E,E\backslash M)$. Therefore, [10] reduces to the well-known formula

$$\mathrm{ch}(i_!1)=\pi^*\hat{A}(E)^{-1/2}i_*1$$

in cohomology groups $H^*(E,E\backslash M)$. The refinement [10] as an equality of differential forms is due to Mathai and Quillen (1986). In fact, this is how [7] was derived originally.

## Equivariant Cohomology and Equivariant Vector Bundles

### Equivariant Cohomology

Let $G$ be a compact Lie group with Lie algebra $\mathfrak{g}$. Fixing a basis $\{e_a\}$ of $\mathfrak{g}$, the structure constants are given by $[e_a,e_b]=t^c_{ab}e_c$. Let $\{\vartheta^a\}$ and $\{\varphi^a\}$ be the dual bases of $\mathfrak{g}^*$ generating the exterior algebra $\wedge(\mathfrak{g}^*)$ and the symmetric algebra $S(\mathfrak{g}^*)$, respectively. The Weil algebra is $W(\mathfrak{g})=\wedge(\mathfrak{g}^*)\hat{\otimes}S(\mathfrak{g}^*)$. We define a grading on $W(\mathfrak{g})$ by specifying $\deg\vartheta^a=1,\deg\varphi^a=2$. The contraction $\iota_a$ and the exterior derivative $d$ are two odd derivations on $W(\mathfrak{g})$ defined by

$$\begin{aligned}\iota_a\vartheta^b&=\delta^b_a, & \iota_a\varphi^b&=0\\ \mathrm{d}\vartheta^a&=-\tfrac{1}{2}t^a_{bc}\vartheta^b\vartheta^c+\varphi^a, & \mathrm{d}\varphi^a&=-t^a_{bc}\vartheta^b\varphi^c\end{aligned} \qquad [11]$$

The Lie derivative is $L_a=\{\iota_a,d\}$. These operators satisfy the usual (anti-)commutation relations

$$d^2=0,\qquad L_a=\{\iota_a,d\},\qquad [L_a,d]=0 \qquad [12]$$

$$\begin{aligned}\{\iota_a,\iota_b\}=0,\qquad [L_a,\iota_b]=t^c_{ab}\iota_c,\\ [L_a,L_b]=t^c_{ab}L_c\end{aligned} \qquad [13]$$

The cohomology of $(W(\mathfrak{g}),d)$ is trivial.

If $G$ acts smoothly on a manifold $M$ on the left, let $V_a$ be the vector field generated by the Lie algebra element $-e_a\in\mathfrak{g}$. Then, $[V_a,V_b]=t^c_{ab}V_c$. Denote $\iota_a=\iota_{V_a}$ and $L_a=L_{V_a}$, acting on $\Omega^*(M)$. In the Weil model of equivariant cohomology, one considers the graded tensor product $W(\mathfrak{g})\hat{\otimes}\Omega^*(M)$, on which the operators

$$\begin{aligned}\tilde{\iota}_a&=\iota_a\hat{\otimes}1+1\hat{\otimes}\iota_a\\ \tilde{d}&=d\hat{\otimes}1+1\hat{\otimes}d\\ \tilde{L}_a&=L_a\hat{\otimes}1+1\hat{\otimes}L_a\end{aligned}$$

act and satisfy the same relations [12] and [13]. An element $\omega\in W(\mathfrak{g})\hat{\otimes}\Omega^*(M)$ is "basic" if it satisfies $\iota_a\omega=0, L_a\omega=0$ for all indices $a$. Let $\Omega^*_G(M)=(W(\mathfrak{g})\hat{\otimes}\Omega^*(M))_{\mathrm{bas}}$ be the set of such. Elements of $\Omega^*_G(M)$ are equivariant differential forms on $M$. The operator $\tilde{d}$ preserves $\Omega^*_G(M)$ and its cohomology groups $H^*_G(M)$ are the equivariant cohomology groups of $M$. They are

isomorphic to the singular cohomology groups of $EG \times_G M$ with real coefficients.

The BRST model of Kalkman (1993) is obtained by applying an isomorphism $\sigma = e^{\vartheta^a \otimes \iota_a}$ of $W(\mathfrak{g}) \hat{\otimes} \Omega^*(M)$. The operators become

$$\sigma \circ \tilde{\iota}_a \circ \sigma^{-1} = \iota_a \hat{\otimes} 1$$
$$\sigma \circ \tilde{d} \circ \sigma^{-1} = \tilde{d} - \varphi^a \hat{\otimes} \iota_a + \vartheta^a \hat{\otimes} L_a$$
$$\sigma \circ \tilde{L}_a \circ \sigma^{-1} = \tilde{L}_a$$

The subspace of basic forms in the Weil model becomes

$$\sigma(\Omega^*_G(M)) = (S(\mathfrak{g}^*) \otimes \Omega^*(M))^G$$

This is precisely the Cartan model of equivariant cohomology, in which the exterior differential is

$$\tilde{d}' = 1 \otimes d - \varphi^a \otimes \iota_a$$

If $P$ is a principal $G$-bundle over a base space $B$, we can form an associated bundle $P \times_G M \to B$. Choose a connection on $P$ and let $\Theta = \Theta^a e_a \in \Omega^1(P) \otimes \mathfrak{g}$, $\Phi = \Phi^a e_a \in \Omega^2(P) \otimes \mathfrak{g}$ be the connection and curvature forms, respectively. The components $\Theta^a, \Phi^a$ satisfy the same relations [11]. Replacing $\vartheta^a, \varphi^a$ by $\Theta^a, \Phi^a$, we have a homomorphism that maps $\omega \in W(\mathfrak{g}) \otimes \Omega^*(M)$ to $\hat{\omega} \in \Omega^*(P \times M)$. If $\omega$ is basic, then so is $\hat{\omega}$, and the latter descends to a form $\bar{\omega}$ on $P \times_G M$. Furthermore, the operator $\tilde{d}$ on $\Omega^*_G(M)$ descends to $d$ on $\Omega^*(P \times_G M)$. Thus, we get the Chern–Weil homomorphisms $\Omega^*_G(M) \to \Omega^*(P \times_G M)$ and $H^*_G(M) \to H^*(P \times_G M)$. For example, the vector space $\mathbb{R}^r$ has an obvious $SO(r)$ action. The Gaussian $r$-form [1] is invariant under $SO(r)$ and can be extended to an $SO(r)$-equivariant closed $r$-form, called the "universal Thom form." Let $E$ be an orientable real vector bundle $E$ of rank $r$ with a Euclidean structure. $E$ determines a principal $SO(r)$-bundle $P$; the associated bundle $P \times_{SO(r)} \mathbb{R}^r$ is $E$ itself. By applying the Chern–Weil homomorphism to this setting, we get a closed $r$-form on $E$. This is another construction of the Thom form [7] by Mathai and Quillen (1986). Further information of equivariant cohomology can be found there, and in Berline *et al.* (1992) and Guillemin and Sternberg (1999).

### Equivariant Vector Bundles

Recall that a connection on a vector bundle $E \to M$ determines, for any $k \geq 0$, a differential operator

$$\nabla : \Omega^k(M, E) \to \Omega^{k+1}(M, E)$$

The curvature $R = \nabla^2 \in \Omega^2(M, \mathrm{End}(E))$ satisfies the Bianchi identity $\nabla R = 0$. If the connection preserves a Euclidean structure on $E$, then $R$ is skew-symmetric.

If a Lie group $G$ acts on $M$ and the action can be lifted to $E$, then $G$ also acts on the spaces $\Gamma(E)$ and $\Omega^*(M, E)$. As before, the Lie derivatives $L_a$ on these spaces are the infinitesimal actions of $-e_a \in \mathfrak{g}$. We choose a $G$-invariant connection on $E$. The "moment" of the connection $\nabla$ under the $G$-action is $\mu_a = L_a - \nabla_{V_a}$ acting on $\Gamma(E)$. In fact, $\mu_a$ is a section of $\mathrm{End}(E)$, or $\mu \in \Gamma(\mathrm{End}(E)) \otimes \mathfrak{g}^*$. If a Euclidean structure on $E$ is preserved by both the connection and the $G$-action, then $\mu_a$ is skew-symmetric. On $\Omega^*(M, E)$, we have

$$L_a = \{\iota_a, \nabla\} + \mu_a$$
$$\iota_a R = \nabla \mu_a, \qquad L_a \mu_b = t^c_{ab} \mu_c$$
$$[\mu_a, \mu_b] = t^c_{ab} \mu_c + R_{ab}$$

where $R_{ab} = R(V_a, V_b) \in \Gamma(\mathrm{End}(E))$.

On the graded tensor product $W(\mathfrak{g}) \hat{\otimes} \Omega^*(M, E)$, the contraction $\tilde{\iota}_a$ and the Lie derivative $\tilde{L}_a$ act and satisfy [13]. In the Weil model, equivariant differential forms on $M$ with values in $E$ are the basic elements in $W(\mathfrak{g}) \hat{\otimes} \Omega^*(M, E)$, which form a subspace $\Omega^*_G(M, E) = (W(\mathfrak{g}) \hat{\otimes} \Omega^*(M, E))_{\mathrm{bas}}$. The "equivariant covariant derivative" is

$$\tilde{\nabla} = d \hat{\otimes} 1 + 1 \hat{\otimes} \nabla + \vartheta^a \hat{\otimes} \mu_a \qquad [14]$$

One checks that $\{\iota_a, \tilde{\nabla}\} = \tilde{L}_a$ and hence $\tilde{\nabla}$ preserves the basic subspace $\Omega^*_G(M, E)$. The equivariant curvature $\tilde{R} = \tilde{\nabla}^2$ is

$$\tilde{R} = R - \vartheta^a \nabla \mu_a + \varphi^a \mu_a + \tfrac{1}{2} \vartheta^a \vartheta^b R_{ab} \qquad [15]$$

It satisfies the equivariant Bianchi identity $\tilde{\nabla} \tilde{R} = 0$. Equivariant characteristic forms are invariant polynomials of $\tilde{R}$. They are equivariantly closed and their equivariant cohomology classes do not depend on the choice of the $G$-invariant connection. Hence, they represent the equivariant characteristic classes of $E$ in $H^*_G(M)$.

For the BRST model, we use a similar isomorphism $\sigma = e^{\vartheta^a \otimes \iota_a}$ on $W(\mathfrak{g}) \hat{\otimes} \Omega^*(M, E)$. The operators become

$$\sigma \circ \tilde{\iota}_a \circ \sigma^{-1} = \iota_a \hat{\otimes} 1$$
$$\sigma \circ \tilde{\nabla} \circ \sigma^{-1} = \tilde{\nabla} - \varphi^a \hat{\otimes} \iota_a + \vartheta^a \hat{\otimes} L_a$$
$$\sigma \circ \tilde{L}_a \circ \sigma^{-1} = \tilde{L}_a$$

and the basic subspace turns into

$$\sigma(\Omega^*_G(M, E)) = (S(\mathfrak{g}^*) \otimes \Omega^*(M, E))^G$$

This is the Cartan model, which can be found in Berline *et al.* (1992). The equivariant covariant derivative is

$$\tilde{\nabla}' = 1 \otimes \nabla - \varphi^a \otimes \iota_a$$

The equivariant curvature is $\tilde{R}' = (\tilde{\nabla}')^2 = R + \varphi^a \mu_a$ and the characteristic forms are defined similarly.

Let $P \to B$ be a principal $G$-bundle with a connection $\Theta$. Following [14], the bundle $P \times E \to P \times M$ has a connection

$$\hat{\nabla} = d \otimes 1 + 1 \otimes \nabla + \Theta^a \otimes \mu_a$$

It descends to a connection $\bar{\nabla}$ on the vector bundle $P \times_G E \to P \times_G M$. The map $\tilde{\nabla} \mapsto \bar{\nabla}$ can be considered as the analog of the Chern–Weil homomorphism for connections. There is also a homomorphism $\Omega^*_G(M, E) \to \Omega^*(P \times_G M, P \times_G E)$, which commutes with the covariant derivatives $\tilde{\nabla}, \bar{\nabla}$. The curvature $\bar{R} = \bar{\nabla}^2$ is the image of the equivariant curvature $\tilde{R}$. Consequently, the equivariant characteristic forms descend to those of $P \times_G E \to P \times_G M$ by the usual Chern–Weil homomorphism.

Now let $E = E^+ \oplus E^-$ be a graded vector bundle over $M$ with a $G$-action preserving all the structures. We have the $\Omega^*_G(M)$-linear supertrace map str: $\Omega^*_G(M) \hat{\otimes} \Gamma(\mathrm{End}(E)) \to \Omega^*_G(M)$. If $\nabla$ is a $G$-invariant connection on $E$ preserving the grading and if $L \in \Gamma(\mathrm{End}(E)^-)^G$ is odd and $G$-invariant, then $\tilde{D} = \tilde{\nabla} + L$ is an "equivariant superconnection." The equivariant counterpart of [8] is

$$\mathrm{ch}_{\tilde{\nabla}, L}(E^+, E^-) = \mathrm{str} \exp\left(\frac{\sqrt{-1}}{2\pi} \tilde{D}^2\right) \in \Omega^*_G(M)$$

representing the equivariant Chern character of $E^+ \ominus E^-$ in $H^*_G(M)$.

### Representatives of the Equivariant Euler and Thom Classes

Consider an oriented real vector bundle $E \to M$ of rank $r$ with a Euclidean structure $(\cdot, \cdot)$. Choose a connection $\nabla$ on $E$ preserving $(\cdot, \cdot)$. We assume that a Lie group $G$ acts on $M$ and that the action can be lifted to $E$ preserving all the structures on $E$. We use the Weil model; the constructions in the Cartan model are similar. For any $\alpha \in \Omega^k_G(M, E)$ and $\beta \in \Omega^l_G(M, E)$, we obtain $(\alpha, \wedge \beta) \in \Omega^{k+l}_G(M)$ by taking the wedge product of forms as well as the pairing in $E$. The Berezin integral of $\omega \in \Omega^*_G(M, \wedge^* E^*)$ along the fibers of $E$ is $\int^{\mathrm{B}} \omega = \langle \nu, \omega \rangle \in \Omega^*_G(M)$. Here, $\nu$ is the unit section of the canonically trivial determinant line bundle $\wedge^r E$, compatible with the orientation of $E$. The equivariant Euler form

$$e_{\tilde{\nabla}}(E) = \frac{1}{(2\pi)^{r/2}} \int^{\mathrm{B}} \exp\left(\frac{1}{2}(\cdot, \tilde{R} \cdot)\right) = \mathrm{Pf}\left(\frac{\tilde{R}}{2\pi}\right) \quad [16]$$

is equivariantly closed. It represents the equivariant Euler class $e_G(E) \in H^*_G(M)$.

Given a $G$-invariant section $s \in \Gamma(E)^G$, the equivariant counterpart of [4] is

$$\mathcal{S}_{\tilde{\nabla}, s} = \tfrac{1}{2}(s, s) + (\tilde{\nabla} s, \cdot) + \tfrac{1}{2}(\cdot, \tilde{R} \cdot) \quad [17]$$

and that of Mathai–Quillen's Euler form [5] is

$$e_{\tilde{\nabla}, s}(E) = \frac{(-1)^{r(r+1)/2}}{(2\pi)^{r/2}} \int^{\mathrm{B}} \mathrm{e}^{-\mathcal{S}_{\tilde{\nabla}, s}} \quad [18]$$

It is also equivariantly closed, and its equivariant cohomology class is $e_G(E)$. The equivariant extension of Mathai–Quillen's Thom form [7] is

$$\tau_{\tilde{\nabla}}(E) = \frac{(-1)^{r(r+1)/2}}{(2\pi)^{r/2}} \int^{\mathrm{B}} \exp\left(-\frac{1}{2}(\boldsymbol{x}, \boldsymbol{x}) - (\tilde{\nabla}\boldsymbol{x}, \cdot) - \frac{1}{2}(\cdot, \tilde{R} \cdot)\right) \quad [19]$$

where $\boldsymbol{x}$ is the ($G$-invariant) tautological section of $\pi^* E \to E$.

Finally, $G$ acts on the (graded) spinor bundle $S(E)$. Using the equivariant superconnection

$$\tilde{D}_s = \tilde{\nabla}^S + \left(\frac{\pi}{\sqrt{-1}}\right)^{1/2} c(s)$$

[9] generalizes to

$$\mathrm{ch}_{\tilde{\nabla}, s}(S^+(E), S^-(E)) = (-1)^m \hat{A}\left(\frac{\tilde{R}}{2\pi}\right)^{-1/2} e_{\tilde{\nabla}, s}(E)$$

Now apply the construction to the bundle $\pi^* E \to E$ and its tautological section $\boldsymbol{x}$. The pair $\pi^* S^{\pm}(E)$ with an odd bundle map $c(\boldsymbol{x})$ determines, up to a factor of $(-1)^m$, the Thom class $i_! 1_G$ in the equivariant $K$-group $K_G(E, E \backslash M)$. The equivariant analog of [10] descends to

$$\mathrm{ch}_G(i_! 1_G) = \pi^* \hat{A}_G(E)^{-1/2} i_* 1_G$$

in equivariant cohomology.

## Superspace Formulation

### Mathai–Quillen Formalism and the Superspace $\mathbb{R}^{0|1}$

Let $\mathbb{R}^{0|1}$ be the superspace with one fermionic coordinate $\theta$ but no bosonic coordinates. The translation on $\mathbb{R}^{0|1}$ is generated by $D = \partial/\partial\theta$, which satisfies $\{D, D\} = 0$. We consider a sigma model on $\mathbb{R}^{0|1}$ whose target space is an (ordinary) smooth manifold $M$ of dimension $n$. A map $X: \mathbb{R}^{0|1} \to M$ can be written as $X(\theta) = x + \sqrt{-1}\theta\psi$. Here, $x = X|_{\theta=0} \in M$ and $\psi = -\sqrt{-1} DX|_{\theta=0} \in T_x M$; the latter is fermionic. Under the translation

$\theta \mapsto \theta + \epsilon$, $x$ and $\psi$ vary according to the supersymmetry transformations

$$\begin{aligned} \delta x &= \epsilon D X|_{\theta=0} = \sqrt{-1}\epsilon\psi \\ \delta\psi &= \epsilon D(DX)|_{\theta=0} = 0 \end{aligned} \qquad [20]$$

Clearly, $\delta^2 = 0$, which is also a consequence of $D^2 = 0$.

For any $p$-form $\omega \in \Omega^p(M)$, we have an observable

$$\mathcal{O}_\omega(X) = \frac{1}{p!} X^*\omega(D, \ldots, D)|_{\theta=0}$$

In local coordinates,

$$\omega = \frac{1}{p!}\omega_{i_1\cdots i_p}(x)\, \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_p}$$

and

$$\mathcal{O}_\omega(x, \psi) = \frac{\sqrt{-1}^p}{p!}\omega_{i_1\cdots i_p}(x)\psi^{i_1}\cdots\psi^{i_p}$$

Using $C(\cdot)$ to denote the set of function(al)s on a space, we can identify $C(\mathrm{Map}(\mathbb{R}^{0|1}, M))$ with $\Omega^*(M)$. Under [20], $\delta\mathcal{O}_\omega(X) = \epsilon\mathcal{O}_{d\omega}(X)$. So, $\mathcal{O}_\omega(X)$ is invariant under supersymmetry if and only if $\omega$ is closed. The cohomology of $\delta$ is the de Rham cohomology of $M$. Consider the measure $[\mathrm{d}X] = [\mathrm{d}x][\mathrm{d}\psi]$. In local coordinates, $[\mathrm{d}x] = \mathrm{d}x^1\cdots\mathrm{d}x^n$ is the standard (bosonic) measure and $[\mathrm{d}\psi] = \mathrm{d}\psi^1\cdots\mathrm{d}\psi^n$ is a fermionic measure such that

$$\int [\mathrm{d}\psi](-1)^{n(n-1)/2}\,\psi^1\cdots\psi^n = 1$$

For any $\omega \in \Omega^n(M)$, the superfield integral $\int [\mathrm{d}X]\mathcal{O}_\omega(X)$ is equal to the usual integral $\int_M \omega$ if the latter exists.

Let $E \to M$ be a real vector bundle of rank $r$ with an inner product $(\cdot,\cdot)$, and let $\nabla$ be a compatible connection whose curvature is $R$. Consider a theory whose fields are $X \in \mathrm{Map}(\mathbb{R}^{0|1}, M)$ and a fermionic section $\Xi \in \Gamma(X^*E)$. Let $\mathcal{D} = (X^*\nabla)_D$ be the covariant derivative along $D$ in the pullback bundle $X^*E \to \mathbb{R}^{0|1}$. Then, $\chi = \Xi|_{\theta=0} \in E_x$ is fermionic and $f = \mathcal{D}\Xi|_{\theta=0} \in E_x$ is bosonic.

Given a fixed section $s \in \Gamma(E)$, we write a superspace action

$$\begin{aligned} S_{\mathrm{MQ}}[X, \Xi] &= \int_{\mathbb{R}^{0|1}} \mathrm{d}\theta(\Xi, \tfrac{1}{2}\mathcal{D}\Xi + \sqrt{-1}s \circ X) \\ &= \tfrac{1}{2}(f, f) + \sqrt{-1}(f, s) - (\nabla_\psi s, \chi) \\ &\quad + \tfrac{1}{4}(\chi, R(\psi, \psi)\chi) \end{aligned} \qquad [21]$$

It is automatically supersymmetric. Performing the Gaussian integral over $f$ and replacing $\chi$ by $-\sqrt{-1}\chi$, we get

$$\int [\mathrm{d}\Xi]\mathrm{e}^{-S_{\mathrm{MQ}}[X,\Xi]} = \frac{\sqrt{-1}^r}{(2\pi)^{r/2}} \int [\mathrm{d}\chi]\,\mathrm{e}^{-S_{\mathrm{MQ}}[x,\psi,\chi]} \qquad [22]$$

where

$$\begin{aligned} &S_{\mathrm{MQ}}[x, \psi, \chi] \\ &\quad = \tfrac{1}{2}(s, s) - \sqrt{-1}(\chi, \nabla_\psi s) - \tfrac{1}{4}(\chi, R(\psi, \psi)\chi) \end{aligned} \qquad [23]$$

When $r$ is even, [22] is equal to $\mathcal{O}_{e(\nabla,s)(E)}(X)$, where $e(\nabla, s)(E)$ is given by [5]. Furthermore, for any closed form $\omega$ on $M$, the expectation value

$$\langle \mathcal{O}_\omega(X) \rangle = \int [\mathrm{d}X][\mathrm{d}\Xi]\mathcal{O}_\omega(X)\,\mathrm{e}^{-S_{\mathrm{MQ}}[X,\Xi]} \qquad [24]$$

is equal to [6].

### Equivariant Cohomology and Gauged Sigma Model on $\mathbb{R}^{0|1}$

Suppose $G$ is a Lie group and $P$ is a principal $G$-bundle over $\mathbb{R}^{0|1}$. Since $\theta$ is nilpotent, we can choose a "trivialization" of $P$ such that the connection and curvature are $A \in \Omega^1(\mathbb{R}^{0|1}) \otimes \mathfrak{g}$ and $F \in \Omega^2(\mathbb{R}^{0|1}) \otimes \mathfrak{g}$, respectively. ($\mathfrak{g}$ is the Lie algebra of $G$.) In components, $c = \sqrt{-1}\iota_D A \in \mathfrak{g}$ is fermionic and $\phi = -(\sqrt{-1}/2)\iota_D^2 F \in \mathfrak{g}$ is bosonic. The space of connections $\mathcal{A}$ is the set of pairs $(c, \phi)$. Under $\theta \mapsto \theta + \epsilon$,

$$\begin{aligned} \delta c &= \epsilon\left(\phi + \frac{\sqrt{-1}}{2}[c, c]\right) \\ \delta\phi &= \sqrt{-1}\epsilon[c, \phi] \end{aligned} \qquad [25]$$

Thus, the algebra $C(\mathcal{A})$ is isomorphic to the Weil algebra $W(\mathfrak{g})$ and $\delta$ corresponds to the differential $d$ in [11]. This relation between gauge theory on a fermionic space and the Weil algebra can be found in Blau and Thompson (1997).

With a trivialization of $P$, the group of gauge transformation $\mathcal{G}$ can be identified with $\mathrm{Map}(\mathbb{R}^{0|1}, G)$. Any group element is of the form $\hat{g} = g\mathrm{e}^{\sqrt{-1}\xi\theta}$, with $g = \hat{g}|_{\theta=0} \in G$ and $\xi = \sqrt{-1}\iota_D\hat{g}^*\varpi \in \mathfrak{g}$ (fermionic), where $\varpi$ is the Maurer–Cartan form on $G$. The action of $\hat{g}$ is $A \mapsto A' = \mathrm{Ad}_{\hat{g}}(A - \hat{g}^*\varpi)$, or $c \mapsto c' = \mathrm{Ad}_g(c - \xi)$ and $\phi \mapsto \phi' = \mathrm{Ad}_g\phi$. By choosing $\xi = c$, we obtained a new trivialization, called the "Wess–Zumino gauge," in which $c' = 0$. The residual gauge redundancy is $G$, and $\mathcal{A}/\mathcal{G} = \mathfrak{g}/\mathrm{Ad}_G$. The Wess–Zumino gauge is not preserved by the translation on $\mathbb{R}^{0|1}$ unless we define $\delta'$ by composing $\delta$ with a suitable (infinitesimal) gauge transformation. If so, then $\delta'\phi = 0$.

Suppose $M$ is a manifold with a left $G$-action. As before, let $\{e_a\}$ be a basis of $\mathfrak{g}$ and let the vector field $V_a$ be the infinitesimal action of $-e_a$. In the gauged sigma model, we include another field $X \in \Gamma(P \times_G M)$. With a trivialization of $P$, we can identify $X$ with a map

$X : \mathbb{R}^{0|1} \to M$. The covariant derivative is given by $\nabla X = \mathrm{d}X - A^a V_a, \mathcal{D}X = \nabla_D X$. Let $x = X|_{\theta=0} \in M$ and $\psi = -\sqrt{-1}\mathcal{D}X|_{\theta=0} \in T_x M$. Then the supersymmetric transformations are

$$\begin{aligned} \delta x^i &= \sqrt{-1}\epsilon\left(\psi^i - c^a V_a^i\right) \\ \delta\psi^i &= -\epsilon\left(\phi^a V_a^i + \sqrt{-1}c^j V_{a,j}^i\right) \end{aligned} \qquad [26]$$

In the Wess–Zumino gauge, the transformations simplify to $\delta' x = \sqrt{-1}\epsilon\psi, \delta'\psi = -\epsilon\phi^a V_a$.

The observables form the $\mathcal{G}$-invariant part of the space $C(\mathcal{A} \times \mathrm{Map}(\mathbb{R}^{0|1}, M))$. For any $\omega \in \Omega^p(M)$, we have

$$\begin{aligned} \mathcal{O}_\omega(X, A) &= \frac{1}{p!}\omega(\mathcal{D}X, \ldots, \mathcal{D}X)|_{\theta=0} \\ &= \frac{\sqrt{-1}^p}{p!}\omega_{i_1\cdots i_p}(x)\psi^{i_1}\cdots\psi^{i_p} \end{aligned} \qquad [27]$$

$\mathcal{O}_\omega(X, A)$ is gauge covariant: $\mathcal{O}_\omega(X, A) \mapsto \mathcal{O}_{g^*\omega}(X, A)$, and the set of gauge-invariant observables is thus identified with $(S(\mathfrak{g}^*) \times \Omega^*(M))^G$. Moreover, since

$$\begin{aligned} \delta\mathcal{O}_\omega(X, A) = \epsilon(&\mathcal{O}_{\mathrm{d}\omega}(X, A) - \sqrt{-1}c^a\mathcal{O}_{L_a\omega}(X, A) \\ &-\sqrt{-1}\phi^a\mathcal{O}_{\iota_a\omega}(X, A)) \end{aligned}$$

$\delta$ corresponds to the differential $\tilde{d}'$ in BRST model.

Let $E \to M$ be an equivariant vector bundle and let $\nabla$ be a $G$-invariant connection with curvature $R$ and moment $\mu$. Any $s \in \Gamma(E)^G$ defines a section of $P \times_G E \to P \times_G M$, still denoted by $s$. Consider a theory with superfields $X \in \Gamma(P \times_G M)$ and $\Xi \in \Gamma(X^*(P \times_G E))$ (fermionic). Let $\mathcal{D}$ be the covariant derivative of the pullback connection. With a trivialization of $P$, we put $\chi = \Xi|_{\theta=0} \in E_x$ (fermionic) and $f = \mathcal{D}\Xi|_{\theta=0} \in E_x$ (bosonic). The equivariant extension of [21] is

$$S_{\mathrm{MQ}}[X, \Xi, A] = \int_{\mathbb{R}^{0|1}} \mathrm{d}\theta(\Xi, \tfrac{1}{2}\mathcal{D}\Xi + \sqrt{-1}s \circ X)$$

Similar to [22], we get, in the Wess–Zumino gauge,

$$\int [\mathrm{d}\Xi]\mathrm{e}^{-S_{\mathrm{MQ}}[X,\Xi,A]} = \frac{\sqrt{-1}^r}{(2\pi)^{r/2}}\int [\mathrm{d}\chi]\mathrm{e}^{-S_{\mathrm{MQ}}[x,\psi,\phi,\chi]} \qquad [28]$$

where

$$\begin{aligned} S_{\mathrm{MQ}}&[x, \psi, \phi, \chi] \\ &= \frac{1}{2}(s, s) - \sqrt{-1}(\chi, \nabla_\psi s) \\ &\quad -\frac{1}{4}(\chi, R(\psi, \psi)\chi) - \frac{\sqrt{-1}}{2}(\chi, \phi^a\mu_a\chi) \end{aligned} \qquad [29]$$

When $r$ is even, [28] is equal to $\mathcal{O}_{\tilde{e}(\nabla, s)}(X, A)$, where $\tilde{e}(\nabla, s)$ is given by [18].

### The Atiyah–Jeffrey Formula

Given the $G$-action on $M$, for any $x \in M$, there is a linear map $C_x : \mathfrak{g} \to T_x M$ defined by $C_x(e_a) = V_a(x)$. With an invariant inner product $(\cdot, \cdot)$ on $\mathfrak{g}$ and an invariant Riemannian metric on $M$, the adjoint of $C_x$ is $C_x^\dagger : T_x M \to \mathfrak{g}$, that is, $C^\dagger \in \Omega^1(M) \otimes \mathfrak{g}$. If $G$ acts on $M$ freely, then $C_x$ is injective and $(C^\dagger C)_x$ is invertible for all $x \in M$. The projection $M \to \bar{M} = M/G$ is a principal $G$-bundle. It has a connection such that the horizontal subspace is the orthogonal compliment of the $G$-orbits. The connection 1-form is $\Theta = (C^\dagger C)^{-1}C^\dagger$, whereas the curvature is $\Phi = (C^\dagger C)^{-1}$ $\mathrm{d}C^\dagger$ on horizontal vectors.

Let $\omega$ be an equivariant form on $M$. Suppose $G$ acts on $M$ freely, then $\omega$ descends to a form $\omega$ on $\bar{M}$. We look for a gauge-invariant, supersymmetric quantity $\Upsilon(X, A)$ such that

$$\begin{aligned} \frac{1}{\mathrm{vol}(\mathcal{G})}&\int [\mathrm{d}X][\mathrm{d}A]\mathcal{O}_\omega(X, A)\Upsilon(X, A) \\ &= \int [\mathrm{d}\bar{X}]\mathcal{O}_{\bar{\omega}}(\bar{X}) \end{aligned} \qquad [30]$$

Mathematically, $\Upsilon$ corresponds to a closed equivariant form $\upsilon$ on $M$ such that

$$\frac{1}{\mathrm{vol}(G)}\int_{\phi\in\mathfrak{g}}[\mathrm{d}\phi]\int_M \omega(\phi) \wedge \upsilon(\phi) = \int_{\bar{M}} \bar{\omega}$$

which is [30] in the Wess–Zumino gauge. In fact, $\upsilon$ is distribution valued in the sense of Kumar and Vergne (1993) and can be understood as an equivariant homology cycle, as in Austin and Braam (1995).

Let $P$ be a $G$-bundle over $\mathbb{R}^{0|1}$ with a connection and let $\mathrm{Ad}\,P = P \times_G \mathfrak{g} \to \mathbb{R}^{0|1}$ be the adjoint bundle. Consider a (bosonic) superfield $\Lambda \in \Gamma(\mathrm{Ad}P)$. Set $\lambda = \Lambda|_{\theta=0}$ (bosonic) and $\eta = -\sqrt{-1}\mathcal{D}\Lambda|_{\theta=0}$ (fermionic). Choosing a trivialization of $P$, $\lambda$ and $\eta$ are both in $\mathfrak{g}$. Under $\theta \mapsto \theta + \epsilon$, they transform as

$$\begin{aligned} \delta\lambda &= \sqrt{-1}\epsilon(\eta + [c, \lambda]) \\ \delta\eta &= \epsilon([\phi, \lambda] - \sqrt{-1}[c, \eta]) \end{aligned} \qquad [31]$$

The superspace action

$$S_{\mathrm{CMR}}[X, \Lambda, A] = \sqrt{-1}\int_{\mathbb{R}^{0|1}} \mathrm{d}\theta(\Lambda, C^\dagger\mathcal{D}X)$$

is invariant under [25], [26], and [31] and, under the Wess–Zumino gauge, it is

$$\begin{aligned} S_{\mathrm{CMR}}&[x, \psi, \phi, \eta, \lambda] \\ &= -\sqrt{-1}(\eta, C^\dagger\psi) - \sqrt{-1}(\lambda, \mathrm{d}C^\dagger(\psi, \psi)) \\ &\quad + (\lambda, C^\dagger C\phi) \end{aligned} \qquad [32]$$

If $G$ acts on $M$ freely, then

$$\Upsilon(X, A) = \int [\mathrm{d}\Lambda] \mathrm{e}^{-S_{\mathrm{CMR}}[X, \Lambda, A]} \qquad [33]$$

satisfies [30]. The factor $\Upsilon(X, A)$ in [30] is called "projection" in Cordes *et al.* (1996).

Let $E \to M$ be a $G$-equivariant vector bundle with a fixed $G$-invariant connection $\nabla$, moment $\mu$, and an invariant section $s$. Consider the superspace action

$$S_{\mathrm{AJ}}[X, \Xi, \Lambda, A] = S_{\mathrm{MQ}}[X, \Xi, A] + S_{\mathrm{CMR}}[X, \Lambda, A]$$

In the Wess–Zumino gauge and after the Gaussian integral over $f$, it becomes the Atiyah–Jeffrey action

$$S_{\mathrm{AJ}}[x, \psi, \phi, \chi, \eta, \lambda] = S_{\mathrm{MQ}}[x, \psi, \phi, \chi] + S_{\mathrm{CMR}}[x, \psi, \phi, \eta, \lambda] \qquad [34]$$

If $s$ intersect the zero section transversely and $G$ acts on $s^{-1}(0)$ freely, then $s^{-1}(0)/G$ is smooth and

$$\int_{s^{-1}(0)/G} \bar{\omega} = \int [\mathrm{d}x][\mathrm{d}\psi][\mathrm{d}\phi][\mathrm{d}\chi][\mathrm{d}\eta][\mathrm{d}\lambda] \times \mathcal{O}_{\omega}(x, \psi, \phi) \mathrm{e}^{-S_{\mathrm{AJ}}[x, \phi, \phi, \chi, \eta, \lambda]} \qquad [35]$$

for any closed equivariant form $\omega$ on $M$. Equation [35] is the formula of Atiyah and Jeffrey (1990) and of Witten (1988a) in an infinite-dimensional setting. When $s^{-1}(0)/G$ is not smooth, the right-hand side of [35] can be regarded as a definition of the left-hand side.

It is often convenient to add to $S_{\mathrm{AJ}}$ another term

$$\Delta S[X, \Lambda, A] = -\frac{1}{4} \int_{\mathbb{R}^{0|1}} ([\iota_D^2 F, \Lambda], \mathcal{D}\Lambda) = \frac{\sqrt{-1}}{2} (\phi, [\eta, \eta]) + \frac{1}{2} ([\phi, \lambda], [\phi, \lambda]) \qquad [36]$$

Since [36] is $\delta$-exact and no new field is added, the integral [35] does not change if $\Delta S$ is added to $S_{\mathrm{AJ}}$.

## Applications to Cohomological Field Theories

We now apply the Mathai–Quillen construction formally to a number of cases in which both the rank of the vector bundle and the dimension of the base space are infinite. Thus, the (bosonic and fermionic) integrals in [24] or [35] become path integrals in quantum mechanics or quantum field theory.

### Supersymmetric Quantum Mechanics

Let $(M, g)$ be a Riemannian manifold and $LM = \mathrm{Map}(S^1, M)$, the loop space. At each point $u \in LM$, which is a map $u : S^1 \to M$, the tangent space is $T_u LM = \Gamma(u^* TM)$. In particular, $\dot{u} = \mathrm{d}u/\mathrm{d}t$, where $t$ is a parameter on $S^1$, is a tangent vector at $u$ and $u \mapsto \dot{u}$ is a vector field on $LM$. For any Morse function $h$ on $M$, $s(u) = \dot{u} + (\mathrm{grad}\, h) \circ u$ is another vector field on $LM$.

Vector fields on $LM$ can be identified as sections of the bundle $\mathrm{ev}^* TM \to S^1 \times LM$, where $\mathrm{ev} : S^1 \times LM \to M$ is the evaluation map. The Levi-Civita connection $\nabla$ on $TM$ pulls back to a connection on $\mathrm{ev}^* TM$ and the covariant derivatives along $LM$ define a natural connection $\nabla^{LM}$ on $T(LM)$. For example, for any tangent vector $V \in T_u LM = \Gamma(u^* TM)$, we have $\nabla_V^{LM} s(u) = \nabla_t^u V + (\nabla_V \mathrm{grad}\, h) \circ u$, where $\nabla^u$ is the pullback connection on $u^* TM$. The Riemann curvature tensor $R$ on $M$ determines that on $LM$.

The (infinite-dimensional) analog of [22] is

$$\int [\mathrm{d}u][\mathrm{d}\psi][\mathrm{d}\chi] \exp\left( - \int_{S^1} \mathrm{d}t\, L[u, \psi, \chi] \right) \qquad [37]$$

where $\psi, \chi \in T_u LM = \Gamma(u^* TM)$ are fermionic and

$$L[u, \psi, \chi] = \tfrac{1}{2} g(\dot{u} + \mathrm{grad}\, h, \dot{u} + \mathrm{grad}\, h) - \sqrt{-1} g(\chi, \nabla_t^u \psi + \nabla_\psi \mathrm{grad}\, h) - \tfrac{1}{4} g(\chi, R(\psi, \psi)\chi) \qquad [38]$$

Here and below, factors of $\sqrt{-1}$ and $2\pi$ in [22] are absorbed in the path-integral measure. [38] is, up to a total derivative, the Lagrangian of the Euclidean $N = 2$ supersymmetric quantum mechanics on $M$. The partition function [37] is equal to Euler characteristic number of $LM$ or $M$, which can be confirmed by an (exact) stationary-phase calculation.

### Topological Sigma Model

Let $\Sigma$ be a Riemann surface with complex structure $\varepsilon$ and let $(M, \omega)$ be a symplectic manifold with a compatible almost-complex structure $J$. Let $\mathcal{E}$ be a vector bundle over $\mathrm{Map}(\Sigma, M)$ so that the fiber over $u$ is $\mathcal{E}_u = \Gamma(u^* TM \times T^* \Sigma)$. For any $u \in \mathrm{Map}(\Sigma, M)$, $\mathrm{d}u \in \mathcal{E}_u$ and $u \mapsto \mathrm{d}u$ is a section of $\mathcal{E}$. The pullback of the Levi-Civita connection on $TM$, tensored with a connection on $T^*\Sigma$, defines a connection on $\mathcal{E}$.

The vector bundle to which we apply the Mathai–Quillen formalism is the antiholomorphic part $\mathcal{E}^{01}$ of $\mathcal{E}$. The fiber over $u \in \mathrm{Map}(\Sigma, M)$ is $\mathcal{E}_u^{01} = \Gamma((u^* TM \otimes T^* \Sigma)^{01})$. The sub-bundle $\mathcal{E}^{01}$ has a connection $\nabla^{01}$ via projection from $\mathcal{E}$. $\mathcal{E}^{01}$ has a natural section $s : u \mapsto \bar{\partial}_J u = (1/2)(\mathrm{d}u + J \circ \mathrm{d}u \circ \varepsilon)$. Solutions to the equation $\bar{\partial}_J u = 0$ are pseudoholomorphic (or $J$-holomorphic) curves; let $\mathcal{M} = s^{-1}(0)$ be the space of such curves. Its (virtual) dimension is

$$\dim \mathcal{M} = \tfrac{1}{2} \chi(\Sigma) \dim M + 2 c_1(u^* TM) \qquad [39]$$

Along any $V \in T_u\mathrm{Map}(\Sigma, M) = \Gamma(u^*TM)$, the covariant derivative of $s = \bar{\partial}_J$ is calculated in Wu (1995):

$$\nabla^{01}_V(\bar{\partial}_J) = \tfrac{1}{2}(\nabla^u V + J \circ \nabla^u V \circ \varepsilon) + \tfrac{1}{4}\nabla_V J \circ (\mathrm{d}u \circ \varepsilon + J \circ \mathrm{d}u) \qquad [40]$$

where $\nabla^u$ is the pullback connection on $u^*TM$.

To write the Mathai–Quillen formalism for the bundle $\mathcal{E}^{01} \to \mathrm{Map}(\Sigma, M)$, we let $\psi \in \Gamma(u^*TM)$ and $\chi \in \Gamma((u^*TM \otimes T^*\Sigma)^{01})$ be fermionic fields. Equation [23] becomes the Lagrangian

$$\begin{aligned} L[u,\psi,\chi] = {} & \tfrac{1}{2}\|\mathrm{d}u\|^2 + \tfrac{1}{2}(\mathrm{d}u, J \circ \mathrm{d}u \circ \varepsilon) \\ & - \sqrt{-1}(\chi, \nabla^u\psi + (\nabla_\psi J) \circ \mathrm{d}u \circ \varepsilon) \\ & - \tfrac{1}{8}(\chi, (R(\psi,\psi) - \tfrac{1}{2}(\nabla_\psi J)^2)\chi) \end{aligned} \qquad [41]$$

It is precisely the Lagrangian of the topological sigma model of Witten (1988b). Here, the pairing $(\cdot\,,\cdot)$ is induced by the Riemannian metric $\omega(\cdot\,, J\,\cdot)$ on $M$ and a metric on $\Sigma$ that is compatible with $\varepsilon$. The second term in [41], integrated over $\Sigma$, is equal to $\int_\Sigma u^*\omega = \langle[\omega], u_*[\Sigma]\rangle$.

For any differential form $\alpha \in \Omega^p(M)$, let $\mathcal{O}_\alpha(u,\psi)$ be the observable obtained from $\mathrm{ev}^*\alpha \in \Omega^p(\Sigma \times \mathrm{Map}(\Sigma, M))$ by identifying $\Omega^*(\mathrm{Map}(\Sigma, M))$ with $C(\mathrm{Map}(\mathbb{R}^{0|1}, \mathrm{Map}(\Sigma, M)))$. If $\alpha$ is closed and $\gamma \in H_q(\Sigma)$ is a homology cycle, then $W_{\alpha,\gamma}(u,\psi) = \int_\gamma \mathcal{O}_\alpha(u,\psi)$ is identified with a closed $(p-q)$-form on $\mathrm{Map}(\Sigma, M)$. For closed $\alpha_i \in \Omega^{p_i}(M)$ and $\gamma_i \in H_{q_i}(\Sigma)(1 \le i \le r)$, the expectation values

$$\left\langle \prod_{i=1}^r W_{\alpha_i,\gamma_i} \right\rangle = \int [\mathrm{d}u][\mathrm{d}\psi][\mathrm{d}\chi] \prod_{i=1}^r W_{\alpha_i,\gamma_i}(u,\psi)\mathrm{e}^{-S[u,\psi,\chi]} \qquad [42]$$

are the Gromov–Witten invariants of $(M,\omega)$. Moreover, [42] is nonzero only if $\sum_{i=1}^r (p_i - q_i) = \dim \mathcal{M}$.

### Topological Gauge Theory

Let $M$ be a compact, oriented 4-manifold, $G$, a compact, semisimple Lie group, and $P \to M$, a principal $G$-bundle. Denote by $\mathcal{A}$ the space of connections on $P$ and $\mathcal{G}$, the group of gauge transformations. The Lie algebra of $\mathcal{G}$ is $\mathrm{Lie}(\mathcal{G}) = \Gamma(\mathrm{ad}\,P) = \Omega^0(M, \mathrm{ad}\,P)$. At $A \in \mathcal{A}$, the tangent space is $T_A\mathcal{A} = \Omega^1(M, \mathrm{ad}\,P)$. Both spaces have inner products if we choose an invariant inner product $(\cdot\,,\cdot)$ on the Lie algebra $\mathfrak{g}$ of $G$ and a Riemannian metric $g$ on $M$. The infinitesimal action of $\mathcal{G}$ on $\mathcal{A}$ is $C = \nabla_A : \mathrm{Lie}(\mathcal{G}) \to T_A\mathcal{A}$.

With a Riemannian metric, any 2-form on $M$ decomposes into self-dual and anti-self-dual parts: $\Omega^2(M) = \Omega^2_+(M) \oplus \Omega^2_-(M)$. We consider a trivial vector bundle $\mathcal{E} \to \mathcal{A}$ whose fiber is $\Omega^2_+(M, \mathrm{ad}\,P)$. $\mathcal{G}$ acts on $\mathcal{E}$ and the bundle is $\mathcal{G}$-equivariant. The trivial connection on $\mathcal{E}$ is $\mathcal{G}$-invariant; the moment is given by $\phi \in \Gamma(\mathrm{ad}\,P) : \chi \in \Omega^2_+(M, \mathrm{ad}\,P) \mapsto [\phi, \chi]$. The bundle $\mathcal{E}$ has a natural section $s : A \in \mathcal{A} \mapsto F_A^+$, the self-dual part of the curvature. Its derivative along $V \in \Omega^1(M, \mathrm{ad}\,P) = T_A\mathcal{A}$ is $L_V s = (\nabla_A V)^+$. The section $s$ is $\mathcal{G}$-invariant, the zero set $s^{-1}(0)$ is the space of anti-self-dual connections, and the quotient $\mathcal{M} = s^{-1}(0)/\mathcal{G}$ is the instanton moduli space. Its (virtual) dimension is

$$\dim \mathcal{M} = 4\check{h}(\mathfrak{g})k(P) - \tfrac{1}{2}\dim G(\chi(M) + \sigma(M))$$

where $\check{h}(\mathfrak{g})$ is the dual Coxeter number of $\mathfrak{g}$ and

$$k(P) = -\frac{1}{4\check{h}(\mathfrak{g})}\langle p_1(\mathrm{Ad}P), [M]\rangle \in \mathbb{Z}$$

is the instanton number of $P$.

We proceed with the Mathai–Quillen interpretation of Atiyah and Jeffrey (1990). Let $\psi \in \Omega^1(M, \mathrm{ad}\,P)$, $\chi \in \Omega^2_+(M, \mathrm{ad}\,P)$, $\eta \in \Gamma(\mathrm{ad}\,P)$ be fermionic fields and $\phi, \lambda \in \Gamma(\mathrm{ad}\,P)$, bosonic fields. The combination of [34] and [36] is given by the Lagrangian

$$\begin{aligned} & L[A,\psi,\phi,\chi,\eta,\lambda] \\ & = \frac{1}{2}\|F_A^+\|^2 + (\phi, \nabla_A^\dagger \nabla_A \lambda) \\ & \quad - \sqrt{-1}(\eta, \nabla_A\psi) - \sqrt{-1}(\chi, \nabla_A\psi) \\ & \quad - \sqrt{-1}(\lambda, [\psi,\psi]) \\ & \quad + \frac{\sqrt{-1}}{2}(\phi, [\chi,\chi] + [\eta,\eta]) - \frac{1}{2}\|[\phi,\lambda]\|^2 \end{aligned} \qquad [43]$$

Here, $(\cdot\,,\cdot)$ is the pairing induced by a Riemannian metric on $M$ and an invariant inner product on $\mathfrak{g}$. With an additional topological term proportional to $(F_A, \wedge F_A)$, [43] is the Lagrangian of topological gauge theory of Witten (1988a).

There is a tautological connection on the $G$-bundle $\mathcal{A} \times P \to \mathcal{A} \times M$. It is invariant under the $\mathcal{G}$-action. Identifying $\Omega^*(\mathcal{A})$ with $C(\mathrm{Map}(\mathbb{R}^{0|1}, \mathcal{A}))$ and using the Cartan model, the $\mathcal{G}$-equivariant curvature is $\mathcal{F} = F_A + \sqrt{-1}\psi + \phi$. For any homology cycle $\gamma \in H_q(M)$,

$$W_\gamma(A,\psi,\phi) = \frac{1}{4\check{h}(\mathfrak{g})}\int_\gamma (\mathcal{F}, \wedge\mathcal{F}) \qquad [44]$$

corresponds to a closed $\mathcal{G}$-equivariant form on $\mathcal{A}$. For $\gamma_i \in H_{q_i}(M)(1 \le i \le r)$, the expectation values

$$\left\langle \prod_{i=1}^r W_{\gamma_i} \right\rangle = \frac{1}{\mathrm{vol}(\mathcal{G})}\int [\mathrm{d}A][\mathrm{d}\psi][\mathrm{d}\phi][\mathrm{d}\chi][\mathrm{d}\eta][\mathrm{d}\lambda] \times \prod_{i=1}^r W_{\gamma_i}(A,\psi,\phi)\mathrm{e}^{-S[A,\psi,\phi,\chi,\eta,\lambda]} \qquad [45]$$

are, up to a factor of $|Z(G)|$, Donaldson invariants of $M$. Moreover, [45] is nonzero only if $\sum_{i=1}^{r}(4-q_i)=\dim \mathcal{M}$.

Other cohomological field theories can also be understood or constructed by the Mathai–Quillen formalism. Of such we mention only the topological field theories of abelian and nonabelian monopoles in Labastida and Mariño (1995), which are related to the Seiberg–Witten invariants.

*See also:* Characteristic Classes; Donaldson–Witten Theory; Equivariant Cohomology and the Cartan Model; *K*-Theory; Topological Quantum Field Theory: Overview; Topological Sigma Models.

## Further Reading


Atiyah MF and Jeffrey LC (1990) Topological Lagrangians and cohomology. *Journal of Geometry and Physics* 7: 119–138.

Austin DM and Braam PJ (1995) Equivariant homology. *Mathematical Proceedings of the Cambridge Philosophical Society* 118: 125–139.

Berline N, Getzler E, and Vergne M (1992) *Heat Kernels and Dirac Operators*. Berlin: Springer.

Blau M and Thompson G (1997) Aspects of $N_T \geq 2$ topological gauge theories and D-branes. *Nuclear Physics B* 492: 545–590.

Cordes S, Moore M, and Ramgoolam S (1996) *Lectures on 2D Yang–Mills Theory, Equivariant Cohomology and Topological Field Theories (Les Houches, 1994)*, pp. 505–682. Amsterdam: North-Holland.

Guillemin VW and Sternberg S (1999) *Supersymmetry and Equivariant de Rham Theory*. Berlin: Springer.

Kalkman J (1993) BRST model for equivariant cohomology and representatives for the equivariant Thom class. *Communications in Mathematical Physics* 153: 447–463.

Kumar S and Vergne M (1993) Equivariant cohomology with generalized coefficients. *Astérisque* 215: 109–204.

Labastida JMF and Mariño M (1995a) A topological Lagrangian for monopoles on four-manifolds. *Physics Letters B* 351: 146–152.

Labastida JMF and Mariño M (1995b) Non-abelian monopoles on four-manifolds. *Nuclear Physics B* 448: 373–395.

Mathai V and Quillen D (1986) Superconnections, Thom classes, and equivariant differential forms. *Topology* 25: 85–100.

Witten E (1988a) Topological quantum field theory. *Communications in Mathematical Physics* 117: 353–386.

Witten E (1988b) Topological sigma models. *Communications in Mathematical Physics* 118: 411–449.

Wu S (1995) On the Mathai–Quillen formalism of topological sigma models. *Journal of Geometry and Physics* 17: 299–309.

Zhang W (2001) Lectures on Chern–Weil Theory and Witten Deformations. *Nankai Tracts in Mathematics*, vol. 4. Singapore: World Scientific.


# Mathematical Knot Theory

**L Boi**, EHESS, Paris, France



## Fundamental Concepts of the Topological Theory of Knots and Links

The first known discovery relating to knots as mathematical objects was made by Gauss around 1833 in a note that refers to the knotting together of closed curves. This investigation originated in his work on electromagnetic theory that led him to compute inductance in a system of two linked circular wires. In this note he had given an analytic formula for the linking number of a pair of knotted curves. This number is a combinatorial topological invariant (it is an integer number). Moreover, one can now show that this number is invariant under Reidemeister moves (discussed in a later section). The linking coefficient can be generalized for the case of $p$- and $q$-dimensional manifolds in $\boldsymbol{R}^{p+q+1}$. The formula for the parametrized curves $\gamma_1(t)$ and $\gamma_2(t)$ with radius vectors $r_1(t)$, $r_2(t)$ is given by the following formula:

$$\mathrm{lk}(\gamma_1,\gamma_2)=\frac{1}{4\pi}\int_{\gamma 1}\int_{\gamma 2}\frac{(r_1-r_2,\mathrm{d}r_1,\mathrm{d}r_2)^3}{|r_1-r_2|} \qquad [1]$$

The linking coefficient allows us to distinguish some two component links. Another approach to the link coefficient is that involving Seifert surfaces. (On this subject, see the section "Isotopies, Reidemeister moves, torus knots, and the linking number.")

A systematic study of knots in $\boldsymbol{R}^3$, however, was only begun in the second half of the nineteenth century by Tait and his followers. They were motivated by Kelvin's theory of atoms modeled on knotted vortex tubes of ether. It was expected that physical and chemical properties of various atoms could be expressed in terms of properties of knots such as the knot invariants. Even though Kelvin's theory did not work, the theory of knots grew as a subfield of combinatorial and algebraic topology. Recently, new invariants of knots have been discovered and they have led to the solution of long-standing problems in knot theory. Surprising connections between the theory of knots and statistical mechanics, quantum groups, and quantum field theory are emerging. Moreover, knot theory has been shown to be intimately connected with many problems in physics, chemistry, and biology.

Tait classified the knots in terms of the crossing number of a regular projection. A regular projection of a knot on a plane is an orthogonal projection of

the knot such that, at any crossing in the projection, exactly two strands intersect transversely. He made a number of observations about some general properties of knots which have come to be known as the "Tait conjectures." In its simplest form, the classification problem for knots can be stated as follows. Given a projection of a knot, is it possible to decide in finitely many steps if it is equivalent to an unknot. This question was answered affirmatively by W Haken in 1961. (For details, see Burde and Zieschang (1985)).

## General Notions and Definitions

Let $M$ be a closed orientable 3-manifold. A smooth embedding of $S^1$ in $M$ is called a knot in $M$. A link in $M$ is a finite collection of disjoint knots. The number of disjoint knots in a link is called the number of components of the link. Thus, a knot can be considered as a link with one component. Two links $L$, $L'$ in $M$ are said to be equivalent if there exists a smooth orientation-preserving automorphism $f: M \to M$ such that $f(L) = L'$. For links with two or more components, we require $f$ to preserve a fixed given ordering of the components. Such a function $f$ is called an ambient isotopy and $L$ and $L'$ are called ambient isotopic. Here, we shall take $M$ to be $S^3 \cong \boldsymbol{R}^3 \cup \{\infty\}$ and simply write "a link" instead of "a link in $S^3$." The diagrams of links are drawn as links in $\boldsymbol{R}^3$. A link diagram of $L$ is a plane projection with crossings marked as over or under. The simplest combinatorial invariant of a knot $K$ is the crossing number $c(K)$. It is defined as the minimum number of crossings in any projection of the knot $K$. The classification of knots up to crossing number 17 is now known. The crossing numbers of some special families of knots are known; however, the question of finding the crossing number of an arbitrary knot is still unanswered. Another combinatorial invariant of a knot $K$ that is easy to define is the unknotting number $u(K)$. It is defined as the minimum number of crossing changes in any projection of the knot $K$ which makes it into a projection of the unknot. Upper and lower bounds for $u(K)$ are known for any knot $K$. An explicit formula for $u(K)$ for a family of knots called torus knots, conjectured by Milnor nearly 40 years ago, has been proved recently by a number of different methods. The 3-manifold $S^3 \backslash K$ is called the knot complement of $K$. The fundamental group $\pi_1(S^3 \backslash K)$ of the knot complement is an invariant of the knot $K$. It is called the fundamental group of the knot and is denoted by $\pi_1(K)$. Equivalent knots have homeomorphic complements and conversely. However, this result does not extend to links. (For details and a proof, see Manturov (2004), chapter 4).

## The Fundamental Group of Knots and Its Role in Topology

For a better understanding of the above considerations, we need to introduce briefly the important concept of fundamental group in topology. The fundamental group plays an essential role in topology; it is involved in the entire technical apparatus of the subject, and likewise in all applications of topological methods. In fact, for low-dimensional manifolds (i.e., of dimension 2 or 3) the fundamental group underlies all nontrivial topological facts.

Classical knot theory is concerned with the space $S^3 \backslash K = M$, an open 3-manifold. There is a natural embedding of the torus $T^2$ in $M$, namely as the boundary of small tubular neighborhood of the knot $K$. Similarly, for a link we obtain a disjoint union of 2-tori in $M$. The principal topological invariant of a knot $K$ is the fundamental group $\pi_1(M)$ of the complement $M$ of $K$, with distinguished subgroup the natural image of $\pi_1(T^2), T \in M^2$, with the obvious standard basis. The classical theorem of Papakyriakopoulos of the 1950s asserts that a knot is equivalent to the trivial one if and only if $\pi_1(M)$ is abelian. It was known by Haken in the early 1960s that there is an algorithm for deciding whether or not any knot is equivalent to the trivial knot. However, while it appears to have been established (by Waldhausen and others in the 1960s and 1970s) that two knots are topologically equivalent if and only if the corresponding fundamental groups with labeled abelian subgroups are isomorphic, the existence of an appropriate algorithm for deciding such equivalence remains an open question. The complexity of the knot group $\pi_1(M)$ has led to the search for more effectively computable invariants to distinguish knots and links. (On this subject, see the section "Polynomial invariants of knots and links.")

Starting with the oriented diagram of the knot or link $K$ on the plane, one calculates in the standard manner (see Crowell and Fox (1963) and Neuwirth (1965)) a presentation of the group $\pi_1(M)$ of the knot $(M = S^3 \backslash K)$, obtaining one generator for the edge of the diagram of a trefoil knot and a pair of relations for each crossing. Since one relation of each such pair simply equates the pair of generators corresponding to the edges forming the upper branch of the crossing, the presentation reduces immediately to the standard one involving the same number of generators and relations. The 2-complex

$L$ with exactly one 0-cell, and with 1-cells labeled by generators and 2-cells labeled by the relations, is then a deformation retract of $M$. Lifting to the universal cover we obtain a boundary operator on a complex of free $\mathbf{Z}[\pi_1]$-modules, which takes the form of a square matrix with entries from this group ring, and it is this matrix that is related to some differentiation as follows. Denoting the generators by $a_i$ and relators by $r_j$, one defines the operator $\partial_{ai}$ by

$$\partial_{a_i}(a_j) = \delta_{ij}$$
$$\partial_{a_i}(bc) = \partial_{a_i}(b) + b\partial_{a_i}(c)$$

the matrix in question then has entries $q_{ij}$ given by

$$q_{ij} = \partial_{a_i}(r_j)$$

Mapping each generator $a_i$ to $t$, we obtain a complex of modules over the ring of integer Laurent polynomials, with boundary operator the corresponding square matrix now with Laurent polynomials as entries. The determinant of this matrix turns out to be zero, and the highest common factor of its cofactors, after multiplication by a suitable power of $t$, turns out to be just the Alexander polynomials $A(t)$.

Let us say a bit more using a little different notation on this question. Let $A_q(K)$ and $J_q(K)$ be the Alexander polynomial and the Jones polynomial, respectively. One of the earliest problems in knot theory was: to what extent does the topological type $\boldsymbol{X}$ of the complementary space $X = S^3 \setminus K$ and/or the isomorphism class $\boldsymbol{G}$ of its fundamental group $G(K) = \pi_1(X, x_0)$ suffice to classify knots? The trefoil knot is the simplest example of nontrivial knot, so it seems remarkable that, not long after the discovery of the fundamental group of a topological space, Max Dehn (1914) succeeded in proving that the trefoil knot and its mirror image had isomorphic groups, but their knot types were distinct. Dehn's (ingenious) proof was the beginning of a long story, with many contributions which reduced repeatedly the number of distinct knot types that could have homeomorphic complements and/or isomorphic groups, until it was finally proved, quite recently, that (1) $\boldsymbol{X}$ determines $K$ and (2) if $K$ is prime, then $\boldsymbol{G}$ determines $K$ up to unoriented equivalence. Thus, there are at most four distinct oriented prime knot types which have the same knot group.

The knot group $\boldsymbol{G}$ is finitely presented; however, it is infinite, torsion-free, and (if $K$ is not the unknot) nonabelian. Its isomorphism class is in general not easily understood via a direct attack on the problem. In such circumstances, the obvious thing to do is to pass to the abelianized group, but unfortunately $G/[G,G] \cong H_1(X; \mathbf{Z})$ is infinite cyclic for all knots, so it is of no use in distinguishing knots. Passing to the covering space $X$ that belongs to $[G,G]$, we note that there is a natural action of the cyclic group $G/[G,G]$ on $\tilde{}X$ via covering translations. The action makes the homology group $H_1(\tilde{}X; \mathbf{Z})$ into a $\mathbf{Z}[q, q^{-1}]$-module, where $q$ is the generator of $G/[G,G]$. This module turns out to be finitely generated. It is the famous Alexander module. While the ring $\mathbf{Z}[q, q^{-1}]$ is not a principal ideal domain (PID), relevant aspects of the theory of modules over a PID apply to $H_1(\tilde{}X; \mathbf{Z})$. In particular, it splits as a direct sum of cyclic module, the first nontrivial one being $\mathbf{Z}[q, q^{-1}]/A_q(K)$. Thus, $A_q(K)$ is the generator of the "order ideal," and the smallest nontrivial torsion coefficient in the module $H_1(\tilde{}X)$. In particular, $A_q(K)$ is very clearly an invariant of the knot group.

We remark that when a knot is replaced by its mirror image (i.e., the orientation on $S^3$ is reversed), the Alexander and Jones polynomials $A_q(K)$ and $J_q(K)$ go over to $A_{q-1}(K)$ and $J_{q-1}(K)$, respectively. As noted earlier, $A_q(K)$ is invariant under such a change, but from the simplest example, the trefoil knot, we see that $J_q(K)$ is not. Now recall that $\boldsymbol{G}$ does not change under changes in the orientation of $S^3$. This simple argument shows that $J_q(K)$ cannot be a group invariant! Thus, it seems interesting indeed to ask about the underlying topology behind the Jones polynomial.

## Isotopies, Reidemeister Moves, Torus Knots, and the Linking Number

Because each knot is a smooth embedding of $S^1$ in $\boldsymbol{R}^3$, it can be arbitrarily closely approximated by an embedding of a closed broken line in $\boldsymbol{R}^3$. Here we mean a good approximation such that after a very small smoothing (in the neighborhood of all vertices) we obtain a knot from the same isotopy class. However, generally this might not be the case.

**Definition 1** An embedding of a disjoint union of $n$ closed broken lines in $R^3$ is called a polygonal n-component link. A polygonal knot is a polygonal one-component link.

**Definition 2** A link is called tame if it is isotopic to a polygonal link and wild otherwise.

All $C^1$-smooth knots are tame. In the sequel, all knots are taken to be smooth, hence, tame.

**Definition 3** Two polygonal links are isotopic if one of them can be transformed to the other by means of an iterated sequence of elementary isotopies and reverse transformations. The

elementary isotopy, generally, is assumed to be a replacement of an edge with two edges provided that the triangle has no intersection points with other edges of the link.

It can be proved that the isotopy of smooth links corresponds to that of polygonal links; the proof is technically complicated. Like smooth links, polygonal links admit planar diagrams with overcrossings and undercrossings, having such a diagram one can restore the link up to isotopy.

**Definition 4** By a planar isotopy of a smooth-link planar diagram we mean a diffeomorphism of the plane onto itself not changing the combinatorial structure of the diagram.

Obviously, planar isotopy is an isotopy, that is, it does not change the link isotopy type in $\boldsymbol{R}^3$.

**Theorem 1** (Reidemeister) *Two diagrams* $D_1$ *and* $D_2$ *of smooth links generate isotopic links if and only if* $D_1$ *can be transformed into* $D_2$ *by using a finite sequence of planar isotopy and the three Reidemeister moves* $\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2, \boldsymbol{\Omega}_3$.

**Theorem 2** *Suppose that* $D$ *and* $D'$ *are regular diagrams of two knots (or links)* $K$ *and* $K'$, *respectively. Then* $K \approx K' \Leftrightarrow D \approx D'$.

We may conclude from the above theorems that the problem of equivalence of knots, in essence, is just a problem of the equivalence of regular diagrams. Therefore, a knot (or link) invariant may be thought of as a quantity that remains unchanged when we apply any one of the Reidemeister moves to a regular diagram.

Knots and links embedded in $\boldsymbol{R}^3$ can be considered as curves (families of curves) in 2-surfaces, where the latter surfaces are standardly embedded in $\boldsymbol{R}^3$. In this section we shall briefly show that all knots and links can be obtained in this manner.

Consider a handle surface $S_g$ standardly embedded in $\boldsymbol{R}^3$ and a curve (knot) $K$ in it. We can now ask the following question: which knot isotopy classes can appear for a fixed $g$? First, let us note that for $g=0$ there exists only one knot embeddable in $S^2$, namely the unknot. The case $g=1$ (torus, torus knots) gives us some interesting information. Consider the torus as a Cartesian product $S^1 \times S^1$ with coordinates $\phi, \varphi \in [0, 2\pi]$, where $2\pi$ is identified with 0. In two dimensions, the torus can be illustrated as a square with opposite sides identified. Let us embed this torus standardly in $\boldsymbol{R}^3$; more precisely,

$$(\phi, \varphi) \to ((R + r\cos\varphi)\cos\phi, (R + r\cos\varphi)\sin\phi, r\sin\varphi) \quad [2]$$

Here $R$ is the outer radius of the torus, $r$ the small radius ($r < R$), $\phi$ the longitude, and $\varphi$ the meridian. For the classification of torus knots we shall need the classification of isotopy classes of nonintersecting curves in $T^2$: obviously, two curves isotopic in $T^2$ are isotopic in $\boldsymbol{R}^3$. Without loss of generality, we can assume the considered closed curve to pass through the point $(0,0) = (2\pi, 2\pi)$. It can intersect the edges of the square several times. In addition, assume all these intersections to be transverse. Let us calculate separately the algebraic number of intersections with horizontal edges and those with vertical edges. Here, passing through the right edge or through the upper edge is said to be positive; that through the left or the lower edge is negative. Thus, for each curve of such type we obtain a pair of integer numbers. So, each torus knot passes $p$ times the longitude of the torus, and $q$ times its meridian, where $\mathrm{GCD}(p,q) = 1$. It is easy to see that for any coprime $p$ and $q$ such a curve exists: one can just take the geodesic line $\{q\phi - p\varphi = 0 \,(\mathrm{mod}\, 2\pi)\}$. Let us denote the torus knot by $T(p,q)$. So, in order to classify torus knots, one should consider pairs of coprime numbers $p, q$ and see which of them can be isotopic in the ambient space $\boldsymbol{R}^3$. The simplest case is when either $p$ or $q$ equals 1. The next simplest example of a pair of coprime numbers is $p=3$, $q=2$ (or $p=2$, $q=3$). In each of these cases we obtain the trefoil knot. Let us state the following important result.

**Theorem 3** *For any coprime integers* $p$ *and* $q$, *the tori* $(p,q)$ *and* $(q,p)$ *are isotopic.*

**Proof** For a proof of this theorem, see Rolfsen (1990). Note that the $(p,q)$ torus knot in one full torus is just the $(q,p)$ torus knot in the other one. Thus, mapping one full torus to the other one, we obtain an isotopy of $(p,q)$ and $(q,p)$ torus knots. This homotopy of full tori can be expressed as a continuous process in $S^3$. Indeed, torus knots of type $(p,q)$ can be represented by a series of planar diagrams. Moreover, it is possible to demonstrate a way of coding a knot (link) as a ($p$-strand) braid closure.

Analogously to the case of torus knots, one can define torus links which are links embedded into the torus standardly embedded in $\boldsymbol{R}^3$. We know the construction of torus knots. So, in order to draw a torus link, one should take a torus knot $K \supset T$ (one can assume that it is represented by a straight linear curve defined by the equation $q\phi - p\varphi = 0 \,(\mathrm{mod}\, 2\pi)$ and add to the torus $T$ some closed nonintersecting simple curves; each curve should be nonintersecting and should not intersect $K$. Thus, these curves should be embedded in $T \backslash K$, that is, in the open

cylinder. Each curve on the cylinder is either contractible or passes the longitude of the cylinder once. So, each curve in $T\backslash K$ is either contractible inside $T\backslash K$, or "parallel" to $K$ inside $T$, that is, isotopic to the curve given by the equation $q\phi - p\varphi = \varepsilon \pmod{2\pi}$ inside $T\backslash K$. Thus, the following theorem holds.

**Theorem 4** *Each torus knot is isotopic to the disconnected sum of a trivial link and a link that is represented by a set of parallel torus knots of the same type* $(p, q)$.

As we already know, a link invariant is a function defined on links that is invariant under isotopies. We shall represent links by using their planar diagrams. According to the Reidemeister theorem, in order to prove the invariance of some function on links, it is sufficient to check this invariance under the three Reidemeister moves. First, let us consider the simplest integer-valued invariant of two-component links. Let $L$ be a link consisting of two oriented components $A$ and $B$ and let $L'$ be the planar diagram of $L$. Consider those crossings of the diagram $L'$ where the component $A$ goes over the component $B$. There are two possible types of such crossings with respect to the orientation. For each positive crossing we assign the number $(+1)$, for each negative crossing we assign the number $(-1)$. Let us summarize these numbers along all crossings where the component $A$ goes over the component $B$. Thus, we obtain some integer number and, in fact, this number is invariant under Reidemeister moves. The so-obtained link invariant is called linking coefficient.

## Polynomial Invariants of Knots and Links

By changing a link diagram at one crossing we can obtain three diagrams corresponding to links $L_+$, $L_-$, and $L_0$ which are identical except for this crossing. In the 1920s, Alexander gave an algorithm for computing a polynomial invariant $\Delta_K(t)$ (a Laurent polynomial in $t$) of a knot $K$, called the Alexander polynomial, by using its projection on a plane. He also gave its topological interpretation as an annihilator of a certain cohomology module associated to the knot $K$. In the 1960s, Conway defined his polynomial invariant and gave its relation to the Alexander polynomial. This polynomial is called the Alexander–Conway polynomial. The Alexander–Conway polynomial of an oriented link $L$ is denoted by $\nabla_L(z)$ or simply by $\nabla(z)$ when $L$ is fixed. We denote the corresponding polynomials of $L_+$, $L_-$, and $L_0$ by $\nabla_+$, $\nabla_-$, and $\nabla_0$, respectively. The Alexander–Conway polynomial is uniquely determined by the following axioms.

**Axiom 1** Let $L$ and $L'$ be two oriented links which are ambient isotopic. Then

$$\nabla_{L'}(z) = \nabla_L(z) \qquad [3]$$

**Axiom 2** Let $S'$ be the standard unknotted circle embedded in $S^3$. It is usually referred to as the unknot and is denoted by $O$. Then

$$\nabla_O(z) = 1 \qquad [4]$$

**Axiom 3** The polynomial satisfies the following skein relation:

$$\nabla_+(z) - \nabla_-(z) = z\nabla_0(z) \qquad [5]$$

We note that the original Alexander polynomial $\Delta_L$ is related to the Alexander–Conway polynomial of an oriented link $L$ by the relation

$$\Delta_L(t) = \nabla_L(t^{1/2} - t^{-1/2}) \qquad [6]$$

In the 1980s, Jones discovered his polynomial invariant $V_L(t)$, called the Jones polynomial, while studying von Neumann algebras and gave its interpretation in terms of statistical mechanics. A state model for the Jones polynomial was then given by Kauffman (1987) using his bracket polynomial. These new polynomial invariants have led to the proofs of most of the Tait conjectures. The Jones polynomial $V_K(t)$ of $K$ is a Laurent polynomial in $t$, which is uniquely determined by a simple set of properties similar to the axioms for the Alexander–Conway polynomials. More generally, the Jones polynomial can be defined for any oriented link $L$ as a Laurent polynomial in $t^{1/2}$, so that reversing the orientation of all components of $L$ leaves $V_L$ unchanged. In particular, $V_K$ does not depend on the orientation of the knot $K$. For a fixed link, we denote the Jones polynomial simply by $V$. Recall that there are three standard ways to change a link diagram at a crossing point. The Jones polynomial is characterized by the following properties:

1. Let $L$ and $L'$ be two oriented links which are ambient isotopic. Then
$$V_{L'}(t) = V_L(t) \qquad [7]$$
2. Let $O$ denote the unknot. Then
$$V_O(t) = 1 \qquad [8]$$
3. The polynomial satisfies the following skein relation:
$$t^{-1}V_+ - tV_- = (t^{1/2} - t^{-1/2})V_0 \qquad [9]$$

An important property of the Jones polynomial that is not shared by the Alexander–Conway polynomial is its ability to distinguish between a knot and its mirror image. More precisely, we have the following result. Let $K_m$ be the mirror image of the knot $K$. Then

$$V_{K_m}(t) = V_K(t-1) \qquad [10]$$

Since the Jones polynomial is not symmetric in $t$ and $t^{-1}$, it follows that in general

$$V_{K_m}(t) \neq V_K(t) \qquad [11]$$

We note that a knot is called amphicheiral (achiral in biochemistry) if it is equivalent to its mirror image. We shall use the simpler biochemistry term. So, a knot that is not equivalent to its mirror image is called chiral. The condition expressed by [11] is sufficient but not necessary for chirality of a knot. The Jones polynomial did not resolve the following conjecture by Tait concerning chirality: if the crossing number of a knot is odd, then it is chiral. However, it has been demonstrated recently that a 15-crossing knot provides a counterexample to the chirality conjecture.

## New Invariants and Their Applications in Mathematical Physics

There was an interval of nearly 60 years between the discovery of the Alexander polynomial and the Jones polynomial. Since then a number of polynomials and other invariants of knots and links have been found. A particularly interesting one is the two-variable polynomial generalizing $V$, called the HOMFLY polynomial (name formed from the initials of authors of the article (Freyd *et al.* 1985) and denoted by $P$. The HOMFLY polynomial $P(\alpha, z)$ satisfies the following skein relation:

$$\alpha^{-1}P_+ - \alpha P_- = zP_0 \qquad [12]$$

Both the Jones polynomial $V$ and the Alexander–Conway polynomial $\nabla_L$ are special cases of the HOMFLY polynomial. The precise relations are given by the following theorem.

**Theorem 5** *Let L be an oriented link. Then the polynomials $P_L, V_L$, and $\nabla_L$ satisfy the following relations*:

$$V_L(t) = P_L(t, t^{1/2} - t^{-1/2}) \text{ and } \nabla_L(z) = P_L(1, z) \qquad [13]$$

After defining his polynomial invariant, Jones also established the relation of some knot invariants with statistical mechanical models. Since then this has become a very active area of research. By constructing a typical statistical mechanics model – the star–triangle relations of the Yang–Baxter equations are an example of such model – one obtains a state model for the Alexander or the Jones polynomial of a knot, by associating to the knot a statistical system, whose partition function

$$Z_K := \sum E_K(s)\omega(s) \qquad [14]$$

gives the corresponding polynomial. (For details, see Jones (1989)). In the function above, $\omega = F(X, S) \to \mathbf{R}$ is a weight function and the sum is taken over all states $s \in F(X, S)$. The energy $E_k$ of the system $(X, S)$ is a functional,

$$E_k : F(X, S) \to \mathbf{R}, k \in K \qquad [15]$$

where the subscript $k \in K$ indicates the dependence of energy on the set $K$ of auxiliary parameters, such as temperature, pressure, etc.

However, these statistical models did not provide a geometrical or topological interpretation of the polynomial invariant. Such an interpretation was provided by Witten (1989) by applying ideas from quantum field theory to the Chern–Simons Lagrangian. In fact, Witten's model allows us to consider the knot and link invariants in any compact 3-manifold $M$.

## Vassiliev Invariants and the Space of All Knots: New Generalizations of Knot Theory

An entirely new collection of knot invariants, which arose out of techniques pioneered by Arnold in singularity theory, has been introduced by V A Vassiliev in the 1990s. The knot invariants, like the Alexander polynomial, associate a knot with some sort of mathematical quantity. A Vassiliev invariant, on the other hand, is an invariant that satisfies a set of conditions. In this sense, all the invariants introduced above – the Jones polynomial, the HOMFLY and the Kauffman polynomial, the Conway polynomial, and the Alexander polynomial – can all be shown to be Vassiliev invariants. However, not all the knot invariants are Vassiliev invariants, for instance, the signature of a knot is not a Vassiliev invariant. The new Vassiliev invariants have a solid basis in a very interesting new topology, where one studies not a single knot, but a space of all knots. Vassiliev's knot invariants are rational numbers. They lie in vector space $V_i$ of dimension $d_i$, $i = 1, 2, 3, \ldots$, with invariants in $V_i$ having "order" $i$. These invariants are built from different families of crossing changes.

Considering that Vassiliev's invariants require introducing an important conceptual change, shifting our attention from the knot $K$, which is the image of $S^1$ under an embedding $\phi: S^1 \to S^3$, to the embedding $\phi$ itself. A knot type $K$ thus becomes an equivalence class $\{\phi\}$ of embeddings of $S^1$ into $S^3$. The space of all such equivalence classes of embeddings is disconnected, with a component for each smooth knot type. In this way, one passes from embeddings to smooth maps, thereby admitting maps which have various types of singularities. Let $^{\sim}M$ be the space of all smooth maps from $S^1$ to $S^3$. This space is connected and contains all knot types. Our space will remain connected and will contain all knot types if we place two mild restrictions on our maps. Let $M$ denote the collection of all $\phi \in {}^{\sim}M$ such that $\phi(S^1)$ passes through a fixed point $\alpha$ and is tangent to a fixed direction at $\alpha$. The space $M$ has some interesting properties, the main one being that it can be approximated by certain affine spaces, and these affine spaces contain representatives of all knot types. The walls between distinct chambers in $M$ constitute the discriminant $\Sigma$, that is, $\Sigma = \{\phi \in M \mid \phi\}$ has a multiple point or a place where its derivative vanishes or other singularities. The space $M - \Sigma$ is our space of all knots.

The additive properties of the Alexander and Jones polynomials have a very attractive interpretation in terms of Vassiliev invariants. By a result of Bar-Natan, all coefficients of the Alexander polynomial are Vassiliev invariants (see Bar-Natan (1995)). The same can be said of the Jones polynomial, as proved by a theorem of Birman and Lin (1993). There is an attractive formula due to Kontsevich expressing all Vassiliev invariants analytically in terms of multiple integrals, assuming that the knot or link diagram comes with some generic Morse function (e.g., the projection of the planar diagram on the $y$-axis). Moreover, from the work of Kontsevich it follows that it is possible to give a purely combinatorial characterization of all Vassiliev invariants (other than the one mentioned above) by associating to an oriented knot $K$ in $R^3$ (given via coordinates $z = z(t)(= x(t) + iy(t)), t)$ a chord diagram, which is just a circle with $2k$ distinct points labeled $P_j$, $Q_j$, $j = 1, 2, \ldots, k$, marked on it, and by imposing certain relations on the free abelian group freely generated by all chord diagrams.

**Theorem 6** *Let $V_K(t)$ be the Jones polynomial of a knot $K$. Let $V_K(q)$ be the infinite series obtained from $V_K(t)$ by substituting $e^q (= 1 + q + q^2/2! + \cdots = \sum_{n=0}^{\infty} q^n/n!)$ for $t$. So we may write*

$$V_K(q) = b_0 + b_{1q} + b_2 q^2 + \cdots$$

*Then $J_m(K) = b_m$ is a Vassiliev invariant induced by the Jones polynomial of order (at most) $m$.*

The structure and significance of the HOMFLY and Kauffman polynomials can be interpreted in the language of Vassiliev invariants, which are invariants of finite type. The notion of finite type is of extraordinary significance in studying these invariants. One reason for this is the following basic lemma:

**Lemma 7** *If a graph $G$ (an embedded 4-valent graph) has exactly $k$ nodes, then the value of a Vassiliev invariant $v_k$ of type $k$ on $G$, $v_k(G)$, is independent of the embedding of $G$.*

Let us show briefly this important result. Suppose $V$ is any invariant of oriented links taking values in some abelian group. This $V$ can be extended to be an invariant of singular links in the following way (Kauffman 2001): a singular link is an immersion of simple closed curves in $S^3$ with finitely many transverse double-points. These self-intersections are required to remain transverse in any isotopy demonstrating the equivalence of such singular links. If the definition of $V$ has been extended over singular links with $n - 1$ double points, define it on a singular link $L_\times$ with $n$ singularities by

$$V(L_{¥}) = V(L_+) - V(L_-)$$

where $V(L_\times)$, $V(L_+)$, and $V(L_-)$ are identical except near a point where they form a node. Note that $V(L_+)$ and $V(L_-)$ each has $n - 1$ double points. Then $V$ is called a Vassiliev invariant of order $n$, or an invariant of finite type $n$, if $V(L) = 0$ for every $L$ with $n + 1$ or more singularities. Recall the Alexander–Conway polynomial invariant, $\nabla_L(z) \in \mathbf{Z}[z]$, of oriented links defined by $\nabla_{\text{unknot}}(z) = 1$ and

$$\nabla_{L_+}(z) - \nabla_{L_-}(z) = z\nabla_{L_0}(z)$$

Extend this over singular links by the above method. Then if $L_\times$ is a link with $r$ singularities, $\nabla_{L_\times}(z) = z\nabla_{L_0}(z)$, where $L_0$ is a link with $r - 1$ singularities. Thus, by induction on $r$, if $L$ has $r$ singularities then $\nabla_L(z)$ has a factor of $z'$. This implies at once that the coefficient of $z^n$ in the Conway polynomial of a link is a Vassiliev invariant of order $n$. Now suppose one considers the HOMFLY polynomial and makes the substitution $(l, m) = (it^{N/2}, i(t^{-1/2} - t^{1/2}))$. The characterizing skein relation becomes

$$t^{N/2}P(L_+) - t^{-N/2}P(L_-) = (t^{1/2} - t^{-1/2})P(L_0)$$

Note that this becomes the Jones polynomial when $N = 2$. Now make the further substitution $t = \exp x$. Here $\exp x$ should be thought of as the classical power series expansion. Of course, $\exp(x/2)$ and

$\exp(-x/2)$ have power series expansions; the power series can be multiplied and added to give another power series. Thus, $P(L)$ has a power series expansion in powers of $x$. It follows immediately that $P(L_+)-P(L_-)=xS(x)$ for some power series $S(x)$. Hence, the proof used for the Conway polynomial shows at once that the coefficient of $x^n$ in the power series expansion of $P(L)$ is a Vassiliev invariant of order $n$.

All present studies of Vassiliev invariants clearly indicate a major role of these invariants in the future developments of knot theory and topological quantum field theories. Many questions in knot theory remain open, nevertheless, in future it will, very likely be one of the most fruitful and beautiful subjects of research in mathematics and in mathematical physics. Knot theory also attracts attention from the fact that it is revealing new astounding and profound links between geometry, algebra, and topology.

*See also:* Finite-Type Invariants; The Jones Polynomial; Knot Invariants and Quantum Gravity; Knot Theory and Physics; Kontsevich Integral; String Topology: Homotopy and Geometric Perspectives; Topological Knot Theory and Macroscopic Physics; Topological Quantum Field Theory: Overview.

## Further Reading

Alexander JW (1923) Topological invariants of knots and links. *Transactions of the American Mathematical Society* 20: 257–306.

Atiyah M (1990) *The Geometry and Physics of Knots*. Cambridge: Cambridge University Press.

Birman JS and Lin X-S (1993) Knot polynomials and Vassiliev's invariants. *Inventiones Mathematicae* 111: 225–270.

Burde G and Zieschang H (1985) Knots. *Studies in Mathematics*, vol. 5. Berlin: Walter de Gruyter.

Crowell RH and Fox RH (1963) *Introduction to Knot Theory*. Toronto: Ginn & Company.

De La Harpe P, Kervaire M, and Weber Cl (1986) On the Jones Polynomial. *L'Enseignement Mathématique* 32: 271–335.

Dehn M (1914) Die beiden Kleeblattschlingen. *Mathematische Annalen* 75: 402–413.

Frayd R *et al.* (1985) A new polynomial invariant of knots and links. In: Freyd P, Yetter D, Hoste J, Lickorish WBR, Millett K, and Ocneau A (eds.) *Bulletin of the American Mathematical Society (NS)* 12: 239–246.

Jones VFR (1985) A polynomial invariant for knots via von Neuman algebras. *Bulletin of the American Mathematical Society* 12: 103–111.

Kauffman LH (2001) *Knots and Physics*. Singapore: World Scientific.

Kawauchi A (1996) *A survey of Knot Theory*. Boston: Birkhäuser.

Lickorish WBR and Millet K (1987) A polynomial invariant for knots and links. *Topology* 26: 107–141.

Manturov V (2004) *Knot Theory*. Boca Raton, FL: Chapman and Hall/CRC.

Murasugi K (1996) *Knot Theory and Its Applications*. Boston: Birkhäuser.

Neuwirth LP (1965) *Knot Groups*. Ann. Math. Studies, vol. 56. Princeton: Princeton University Press.

Reidemeister K (1932) *Knotentheorie*. Berlin: Springer.

Rolfsen D (1990) Knots and Links. *Math. Lecture Series*. Berkeley: Publish or Perish.

Vassiliev VA (1990) Cohomology of knot spaces. In: Arnold VI (ed.) *Theory of Singularities and Its Applications*, Advances in Soviet Mathematics, vol. 1, pp. 23–70. Providence, RI: American Mathematical Society.

Witten E (1989) Quantum field theory and the Jones polynomial. *Communications in Mathematical Physics* 121: 351–399.

**Matrix Product States** *see* Finitely Correlated States

**Mean Curvature Flow** *see* Geometric Flows and the Penrose Inequality

# Mean Field Spin Glasses and Neural Networks

**A Bovier**, Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany



## Introduction and Models

Rarely has a paper with a simple title as "A solvable model of a spin glass" had such a tremendous impact on both physics and mathematics as the seminal paper of 1972 by Sherrington and Kirkpatrick, which introduced what is now known as the Sherrington–Kirkpatrick (SK) mean-field spin glass model. As solvable as it might have appeared to the authors, it was soon found that the heuristic solution, based on the so-called replica method, was physically unacceptable. The reason was a tacit assumption, now known as replica symmetry, that proved unfounded. Several years later, Giorgio Parisi provided an ingenious way out through his continuous replica symmetry-breaking scheme, that presented a solution that, through its complexity and intrinsic beauty, both stunned and fascinated the community. Unraveling the mysteries involved in this solution has presented a challenge and driving force for the last three decades of mathematical statistical mechanics, while the use of the method in theoretical physics opened the path to solving a wide variety of problems not only in the theory of disordered magnets, but also in neural networks and combinatorial optimization. In this article the focus is on the mathematical results obtained in the study of this and a number of related models.

## Mean-Field Models

Mean-field models have played an important role in statistical mechanics by providing simple, solvable models in which some of the complex phenomena, such as phase transitions, could be studied and understood. For example, the Curie–Weiss model of a ferromagnet describes $N$ spin variables $\sigma_i$ (taking values $\pm 1$) in interaction. The simplifying assumption compared to more realistic models, such as the Ising model, is to ignore the spatial structure of the model and allow all spins to interact with each other with equal strength. This yields to a Hamiltonian function of the form

$$H_N(\sigma) = -\frac{J}{N}\sum_{i,j=1}^{N}\sigma_i\sigma_j + h\sum_{i=1}^{N}\sigma_i \qquad [1]$$

where $J$ is a coupling constant and $h$ a magnetic field. This from of the interaction implies that the Hamiltonian is in fact just a function of the empirical magnetization $m_N(\sigma) = N^{-1}\sum_{i=1}\sigma_i$, and this allows one to use methods from the theory of large deviations to analyze rather easily the corresponding Gibbs measures

$$\mu_{\beta,N}(\sigma) \equiv \frac{e^{-\beta H_N(\sigma)}}{Z_{\beta,N}} \qquad [2]$$

## The SK Model

This model was a straightforward attempt to introduce a mean-field version of models with randomly interacting spins. The interest in such models arose from the discovery of certain alloys of ferromagnets and conductors (e.g., AuFe and CuMn) that had been found to exhibit very unusual magnetic properties. Ruderman and co-workers had proposed that in these models the magnetic ions with magnetic moments $S_i$ and $S_j$ located at the points $x_i$ and $x_j$ would interact via an exchange interaction of the form

$$\frac{\cos(k_f(x_i - x_j))}{|x_i - x_j|^3} S_i \cdot S_j$$

Since the positions of the magnetic ions in the alloy are random, the signs of their interaction would be oscillatory. Anderson proposed a simplified model, in the spirit of the Ising model, where spins taking values $\pm 1$ located on a regular lattice would interact via nearest-neighbor couplings $J_{ij}$ modeled as i.i.d. random variables uniformly distributed on an interval $[-J, J]$. In the spirit of the Curie–Weiss model, Sherrington and Kirkpatrick then proposed the mean-field model where any two spins would interact via i.i.d. Gaussian random variables $J_{ij}$ of mean zero and variance one. The SK Hamiltonian is thus given by

$$H_N^{\mathrm{SK}}(\sigma) \equiv -\frac{J}{\sqrt{N}}\sum_{1\le i<j\le N} J_{ij}\sigma_i\sigma_j + h\sum_{i=1}^{N}\sigma_i \qquad [3]$$

where the normalization is chosen to ensure that the variance of $H_N$ is an extensive quantity. Although the two Hamiltonians superficially look similar, the main feature that allows one to solve the Curie–Weiss model is absent in the SK model: there is no way to write the Hamiltonian as a function of macroscopic variable(s) such as the magnetization. This implies that all methods known to solve the Curie–Weiss model fail here. The approach used systematically in the physics literature to overcome

this difficulty is to try to compute the mean free energy $f_{\beta,N} \equiv -(1/\beta N)$, $\mathbb{E}\ln Z_{\beta,N}$ using the formal identity $\ln x = \lim_{q\downarrow 0} q^{-1}(x^q - 1)$. For $q \in N$, one easily sees that (putting $h = 0$)

$$\mathbb{E}Z^q_{\beta,N} = \sum_{\sigma^1,\ldots,\sigma^q} \exp\left(\frac{\beta^2 J^2}{2} N \sum_{a,b=1}^{q} \sum_{i<j} \sigma_i^a \sigma_i^b \sigma_j^a \sigma_j^b\right)$$

This expression looks already more like the partition function of an ordinary mean-field model, and the computation with standard methods seemed feasible. However, passage to the limit $q\downarrow 0$ remains a highly risky enterprise, and it took the genius of Parisi to develop an approach that provided at least a physically meaningful and convincing answer. The replica method being dealt with elsewhere in this encyclopedia, this approach is not explained any further here, although we will explain the nature of the result in the light of recent rigorous work later on.

## Site Disordered Models

The difficulties encountered with the random-bond interactions led readily to proposals of mean-field models that were closer to the Curie–Weiss model – from the point of view that they allowed the Hamiltonian to be written as a function of macroscopic variables. The most important of these models was introduced by Figotin and Pastur. Here the disorder was introduced as an $M$-dimensional vector $\xi_i$ for each site $i$. The components of this vector are usually taken as i.i.d. random variables $\xi_i^\mu$ taking values $\pm 1$ with equal probability. One can then introduce $M$-dimensional vectors as macroscopic variables that generalize the magnetization with components

$$m_N^\mu(\sigma) \equiv N^{-1} \sum_{i=1}^{N} \xi_i^\mu \sigma_i$$

The Hamiltonian can then be written as

$$H_N(\sigma) = -N \sum_{\mu=1}^{M} \left(m_N^\mu(\sigma)\right)^2 = -\frac{1}{N} \sum_{i,j=1}^{N} \sigma_i \sigma_j \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu$$

These models were indeed found to be solvable with tools similar to those used in the Curie–Weiss case; however, they proved disappointing in that the solution did not show the characteristic features expected in a spin glass. In fact, it turns out the these models behave very much like a mean-field ferromagnet, except that as they display not just two equilibrium states at low temperatures, but $2M$ of them, concentrated on spin configurations $\sigma$ for which $m_N(\sigma)$ takes values close to one of the values $\pm m^*(\beta) e_\mu$, where $e_\mu$ is the $\mu$-unit vector in $\mathbb{R}^M$ and $m^*(\beta)$ solves the equation $m = \tanh(\beta m)$ known from the Curie–Weiss model. This model might have been forgotten, had it not been rediscovered in 1982 by Hopfield in the context of neural networks. Hopfield realized that if $\sigma_i$ are interpreted as the activation states ("firing" and "not firing") of neurons in the brain, the form of the interaction in this model is exactly the one proposed earlier by Hebb for synaptic interaction between neurons having "learned" the $M$ "patterns" $\xi^\mu$ in the past. He went on to interpret $H_N(\sigma)$ as the Lyapounov function of the retrieval algorithm by which the brain would recognize the learned pattern. Naturally, the fact the the configurations $\xi^\mu$ are minima of $H_N$ then implies the functioning of the algorithm. The important observation of Hopfield was that, based on numerical experiments, the algorithm failed when $M$ became too large. In fact, he observed a breakdown of the memory if $M \geq 0.14N$. This meant that the interesting asymptotics in this model required to consider $M$ as an increasing function of $N$. This regime was not covered by large-deviation-type results and an intensive program to investigate this model was initiated. Again, the replica method could be employed and yielded a very rich structure of the model, including an explanation of the findings of Hopfield. These models also turned out to be an important starting point for the rigorous analysis.

## Gaussian Processes and Derrida's Models

While the models discussed so far were motivated from the point of view of randomly interacting spins, Derrida had the consequential idea to view the Hamiltonian of such a model simply as a random process indexed by the set of all spin configuration. In the case of the SK model, this process was, moreover, a Gaussian process and thus characterized entirely by its mean and variance. For $h = 0$ we see that

$$\mathbb{E}H_N^{\mathrm{SK}}(\sigma)H_N^{\mathrm{SK}}(\sigma') = \frac{N}{2}\left(r_N(\sigma,\sigma')\right)^2 - \frac{1}{2} r_N(\sigma,\sigma')$$

where $r_N(\sigma,\sigma') \equiv N^{-1}\sigma_i\sigma_i'$ is usually called the overlap. This opened the view to a much larger class of models. In particular, the simplest model from this perspective corresponds to taking $H_N(\sigma)$ as a process of i.i.d. random variables. Derrida called this the *random-energy* model (REM). He also noted

that it could be seen as the limit if a sequence of the so-called $p$-spin SK models corresponding to the covariance of the Hamiltonian being $N(r_N(\sigma,\sigma'))^p$. On the other hand, Derrida observed that another class of models could be defined that were easier to analyze while exhibiting much of the complex properties of the SK model. These are obtained by choosing the covariance not as a function of the overlap (resp. the Hamming distance), but of a ultra-metric distance related to $d_N(\sigma,\sigma') \equiv N^{-1}(\inf\{i:\sigma_i \neq \sigma'_i\} - 1)$. These models, called generalized Random-Energy Models (GREM) were analyzed by Derrida and Gardner in the 1980s and are now the only models where the full predictions of the Parisi theory can be rigorously justified. This is discussed in some detail later.

## Further Models and Applications

There is a wealth of problems that can be interpreted in terms of disordered mean-field models, and which may be analyzed using methods developed here. Some of the most notable ones that have received more attention lately include: the perceptron, a feed-forward neural network was analyzed first by Gardner using the replica method. Very recently, Shcherbina and Tirozzi gave a rigorous justification of this result. The p-satisfiability problem is an important problem in computer science that also can be analyzed with the replica method. Rigorous results are still very limited. The number partitioning problem can be formulated as a random-energy model. Also, the most famous problem in combinatorial optimization, the traveling salesman problem, can be solved heuristically with the replica method. Another emerging field are applications to coding theory.

## Formulation of the Problem

Given a model, that is, a Hamiltonian function defined as a random process, the ultimate goal is to describe the asymptotic properties of the corresponding Gibbs measure, ideally identifying a (random or deterministic) limiting measure, as a function of the temperature, $\beta^{-1}$, and other parameters, such as the magnetic field $h$.

The first steps in this direction concerns global properties:

- Does the ground-state energy density,

$$\lim_{N\uparrow\infty} \max_{\sigma\in\mathcal{S}_N} H_N(\sigma)$$

converge (in what sense?) and what is the limit?

- What is the limit of the free energy

$$f_{\beta,N} \equiv \frac{-1}{\beta N} \ln Z_{\beta,N}$$

It has been noted in the mid-1990s that such quantities are usually self-averaging, for example, in the sense that

$$\lim_{N\uparrow\infty} \left(f_{\beta,N} - \mathbb{E} f_{\beta,N}\right) = 0, \quad \text{a.s.}$$

due to the concentration of measure phenomenon. However, until very recently, the existence of the limits was considered an open problem in most of the models described above. Guerra and Toninelli (2002) discovered that a clever use of comparison inequalities for convex functions of Gaussian processes allows one to prove *a priori* the existence of limits at least in the case of models based on Gaussian processes (SK, GREM). The main task is the computation of the values of the limit.

If the free energy is known as a function of sufficiently many parameters, one can frequently compute a number of correlation functions that characterize the limiting measure as well. What one should compute is somewhat model dependent.

## Geometry of Gibbs Measures and Multi-Overlap Distributions

The problem of satisfactorly describing the asymptotic geometric properties of random Gibbs measures on $\{-1,1\}$ is rendered difficult as the symmetries of the problem make the use of local topologies seem unattractive. A reasonable way of solving this problem is as follows. Let $D_N$ be a distance on $\mathcal{S}_N$ normalized so that $\max_{\sigma,\tau\in\mathcal{S}_N} D_N(\sigma,\tau) = 1$. Then consider the mass distribution around any fixed point $\sigma$,

$$m_\sigma(x) \equiv \mu_{\beta,N}(D_N(\sigma,\sigma') \leq x)$$

and construct the biased empirical average

$$\mathcal{K}_{\beta,N} \equiv \sum_{\sigma\in\mathcal{S}_N} \mu_{\beta,N}(\sigma)\delta_{m_\sigma(\cdot)}$$

The set of distributions of these random measures is compact (with respect to the weak topology) and thus we can expect to construct limits. The law of $\mathcal{K}_{\beta,N}$ is fully determined by the family of averaged distributions of the distances between $n$ independent copies of $\sigma$ drawn from the Gibbs measures,

$$\mathbb{E}\mu_{\beta,N}^{\otimes n}(D_N(\sigma^1,\sigma^2),\ldots,D_N(\sigma^{n-1},\sigma^n))$$

In the SK models, one chooses

$$D_N(\sigma,\tau) = 1 - \frac{1}{N}\sum_i \sigma_i \tau_i$$

so that these quantities can be expressed as distributions of the overlaps $(1/N)\sum_i \sigma_i \tau_i$, between $n$ "replica" spin variables. In the GREM models, it is natural to chose as distance the lexicographic distance used in the construction of the models. In this case, the limits of $\mathcal{K}_{\beta,N}$ can be constructed explicitly and it was shown that they can be expressed in terms of the size-biased empirical family size distribution of a certain continuous state branching process via a model-dependent time change. Since this plays a key rôle not only in the GREMs but in other models as well, we will go into some detail to elucidate this structure.

In the GREM models, it can be shown that the quantity $\mathcal{K}_{\beta,N}$ converges (weakly in law) to the corresponding $\mathcal{K}$ obtained from a time change of the family of measures $\theta_t$, namely

$$\theta_t^{m_\beta} \equiv \theta_{\ln m(t) - \ln m(0)}$$

where $m$ is a nondecreasing function that can be computed explicitly. Namely, if $\mathbb{E}X_\sigma X_\tau = A(\mathrm{d}_N(\sigma,\tau))$, and $\bar{a}$ denotes the right-derivative of the concave hull of $A$, then

$$m(x) = \min\left(\beta^{-1}\sqrt{2\ln 2}/\sqrt{\bar{a}(x)}, 1\right)$$

As explained below, similar results are expected in the SK models.

## Neveu's Process and Random Genealogies

The random structure of the limiting Gibbs measures of the GREM models (and presumably also the SK models, even though this is not proven) can be traced to a continuous-state branching process introduced by Neveu, and an induced associated random genealogy on the unit interval. Let $Z_t$ be a time-homogeneous continuous-time Markov process with state space $\mathbb{R}_+$ characterized by the Laplace transform of its transition kernel

$$\mathbb{E}(\mathrm{e}^{-\lambda Z_t} | Z_0 = a) = \exp\left(-a\lambda^{\mathrm{e}^{-t}}\right)$$

Based on this process, construct a two-parameter process $Z(t,a)$ with the property that, for any $a, b > 0$, the processes $Z(\cdot, a)$ and $Z(\cdot, a+b) - Z(\cdot, a)$ are independent and have the same laws as $Z_t$ with initial conditions $a$, resp. $b$. It follows that $Z(t,\cdot)$ is a stable subordinator with exponent $\mathrm{e}^{-t}$. Now let $\theta_t(a) \equiv Z(t,a)/Z(t,1)$, as a function on $[0,1]$, $\theta_t$ being a random probability distribution function (of pure point type). Any such family $\mu_t$ of distributions defines in a natural way a genealogical structure on [0, 1]. Define the ancestor of $\alpha \in [0,1]$ at time $t < 1$ to be $a_t(\alpha) \equiv \theta_t(\theta_1^{-1}(\alpha))$, where $\theta^{-1}$ is the right-continuous inverse of the nondecreasing function $\theta$.

We say that, for $\alpha, \alpha' \in [0,1], q(\alpha,\alpha') = t$ if and only if $t = \sup(s : a_s(\alpha) = a_s(\alpha'))$. It is easy to see that $1-q$ defines an ultra-metric distance. We can associate with this the distribution size of the offspring of an ancestor at time $t$, $m_\alpha(t) = |\alpha' : q(\alpha,\alpha') \le t|$, and its size-biased empirical distribution

$$\mathcal{K} \equiv \int_0^1 \mathrm{d}\alpha\, \delta_{m_\alpha(\cdot)}$$

## Interpolation Methods and Guerra's Integral Representation

Among the very important tools for the analysis of Gaussian models in particular have been the interpolation methods that allow one to compare functions of processes with different covariance. While these methods go back to early work on Gaussian processes (Slepian, Kahane), they have been employed with remarkable success in the present context. Mostly, they consist in introducing an interpolating Hamiltonian $H^t(\sigma) \equiv \sqrt{t}H(\sigma) + \sqrt{1-t}K(\sigma)$, where $K$ is a reference process that has certain desired properties. Given any function $F$ of the process (e.g., the free energy of the model), one then represents

$$F(H) = F(K) + \int_0^1 \mathrm{d}t \frac{\mathrm{d}}{\mathrm{d}t} F(H^t)$$

Often the derivative on the right-hand side can be controlled rather well, for example, because of some obvious positivity properties.

**Example 1** (Guerra and Toninelli). Choose

$$K(\sigma) = \frac{1}{\sqrt{M}}\sum_{i<j=1}^{M} J'_{ij}\sigma_i\sigma_j + \frac{1}{\sqrt{N-M}}\sum_{i<j=M+1}^{N} J'_{ij}\sigma_i\sigma_j$$

and consider the free energy $F(H^t) = f^t_{\beta,N}$. Then, first $F(H^0_N) = F(H_M) + F(H_{N-M})$. On the other hand,

$$\frac{\mathrm{d}}{\mathrm{d}t}F(H^t_N) = -\frac{1}{2N}\mu^t_{\beta,N}\left(\sum_{i<j=1}^{N}\frac{\sigma_i\sigma_j J_{ij}}{tN} - \sum_{i<j=1}^{M}\frac{J'_{ij}\sigma_i\sigma_j}{\sqrt{(1-t)M}} + \sum_{i<j=M+1}^{N}\frac{J'_{ij}\sigma_i\sigma_j}{\sqrt{(1-t)(N-M)}}\right)$$

A key tool to be used at this stage is the so-called Gaussian integration by parts formula, $\mathbb{E}gf(g) = \mathbb{E}f'(g)$. Applied here, this gives

$$\frac{\mathrm{d}}{\mathrm{d}t}F(H_N^t) = \frac{\beta}{4N^2}\mu_{\beta,N}^{t,\otimes 2}\left(\left(\sum_{i=1}^N \sigma_i\sigma_i'\right)^2 - \frac{N}{M}\left(\sum_{i=1}^M \sigma_i\sigma_i'\right)^2 - \frac{N}{N-M}\left(\sum_{i=M+1}^N \sigma_i\sigma_i'\right)^2\right) \geq 0$$

This proves superadditivity of $N\mathbb{E}f_{\beta,N}$,

$$N\mathbb{E}f_{\beta,N} \geq M\mathbb{E}f_{\beta,M} + (N-M)\mathbb{E}f_{\beta,N-M}$$

which, in turn, implies convergence of $\mathbb{E}f_{\beta,N}$ to a limit $\mathbb{E}f_\beta$. Moreover, standard concentration of measure estimates show then that $f_{\beta,N}$ also converges almost surely.

**Example 2** (Guerra, Aizenman–Sims–Starr). A more complicated application of the interpolation method allows one to relate the free energy to Parisi's solution. This was first found by Guerra (2003), but a different, and in some sense more intuitive formulation, was given later by Aizenman *et al.* (2003). It is based on the following construction. We consider a centered Gaussian process $H_N(\sigma)$ on $\mathcal{S}_N$ with covariance given by $Ng(R_N(\sigma,\sigma'))$ for some even convex function $g:[-1,1] \to [0,1]$. Let us take $F(H_N) = \ln \mathbb{E}_\sigma \mathrm{e}^{\beta H_N(\sigma)}$ (the *a priori* expectation $\mathbb{E}_\sigma$ need not be symmetric, but may incorporate a magnetic field). Before using comparison, we now want to go to a larger space. For this, introduce some set $\mathcal{A}$ equipped with some positive-definite quadratic form $q$, normalized such that $q_{\alpha,\alpha} = 1$, and $|q_{\alpha,\alpha'}| \leq 1$, $\forall_{\alpha,\alpha' \in \mathcal{A}}$. Let $\mathbb{P}_\alpha$ denote some probability measure on $\mathcal{A}$. Now introduce a centered Gaussian process $\kappa_\alpha$ on $\mathcal{A}$, independent of $H_N$, whose covariance is given by $\mathbb{E}\kappa_\alpha\kappa_{\alpha'} = r(q_{\alpha,\alpha'}) \equiv q_{\alpha,\alpha'}g'(q_{\alpha,\alpha'}) - g(q_{\alpha,\alpha'})$. Define

$$G(H_N + \sqrt{N}\kappa) = \ln\left(\mathbb{E}_\sigma\mathbb{E}_\alpha\,\mathrm{e}^{-\beta(H_N(\sigma)+\sqrt{N}\kappa_\alpha)}\right)$$

Obviously, $G(H_N,\kappa) = F(H_N) + \widetilde{F}(\kappa)$, where $\widetilde{F}(\kappa) = \ln(\mathbb{E}_\alpha\,\mathrm{e}^{-\beta\sqrt{N}\kappa_\alpha})$. The amazing idea is now to compare the process $(H_N + \kappa)$ with another process $\eta_{\sigma,\alpha}$ whose covariance is a linear function of $R_N(\sigma)$ (this is in some sense a Slepian's process), and that otherwise is smaller than the covariance of $(H_N + \kappa)$; to wit

$$\mathbb{E}\eta_{\alpha,\sigma}\eta_{\alpha',\sigma'} = R_N(\sigma,\sigma')g'(q_{\alpha,\,\alpha'})$$

By these choices of covariances, one has that for $x \in [-1,1]$, $y \in [0,1]$, since $g$ is even and convex,

$$g(x) + yg'(y) - g(y) \geq xg'(y)$$

It is an immediate consequence of Kahane's theorem, respectively the same interpolation argument given above, that

$$\mathbb{E}G(H_N + \kappa) \leq \mathbb{E}G(\eta)$$

which translates into

$$\mathbb{E}F(H_N) \leq \mathbb{E}G(\eta) - \mathbb{E}\widetilde{F}(\kappa)$$

It is clear that we can optimize this bound by choosing $\mathcal{A}, q$, and $P_\alpha$. Of course, the difficulty would be to find such a minimum. A first simplification of this optimization problem is to consider instead of the deterministic structure of $\mathbb{P}$ and $q$ random-probability measures on the space of probability measures and quadratic forms on $\mathcal{A}$, to average over the preceding equation with respect to their laws, and then take the infimum over all such random structures. This gives a (still incalculable) bound that Aizenman *et al.* (2003) have shown to be asymptotically sharp, that is, they showed that

$$\lim_{N\uparrow\infty}\mathbb{E}F(H_N) = \lim_{N\uparrow\infty}\inf_{\mathcal{A},\,\mu}\mathbb{E}_\mu(\mathbb{E}G(\eta) - \mathbb{E}\widetilde{F}(\kappa))$$

where $\mu$ is short for all probability measures on the space of $(\mathbb{P}_\alpha, q_{\alpha,\alpha'})$ on $\mathcal{A}$ (called "random overlap structures"(rosts) in Aizenman *et al.* (2003)). Guerra's bound consists in restricting the infimum to a class of rosts where the bound is calculable 'explicitly'. Maybe unsurprisingly, this is exactly the class of asymptotic models that have already arisen in the GREMs. In fact, we set $\mathcal{A} = [0,1]$, $\mathcal{M} \equiv \{m:[0,1] \to [0,1]$, non-decreasing$\}$, let $q$ be the random genealogical distance associated to the family of measures $\theta_t^m$, and let $\mathbb{P}_\alpha$ be the probability measure on $\mathcal{A}$ whose distribution function is $\theta_1^m(\alpha)$. Then Guerra's bound states that

$$\lim_{N\uparrow\infty}\mathbb{E}F(H_N) \leq \lim_{N\uparrow\infty}\inf_{m\in\mathcal{M}}\mathbb{E}G(\eta) - \mathbb{E}\widetilde{F}(\kappa)$$

where the expectations relate to all random quantities involved. By self-averaging, the same result holds almost surely. The right-hand side of this equation is known as (a particular formulation of) the famous Parisi solution. In fact, define the function $f(q,y)$ as the solution of the nonlinear partial differential equation

$$\partial_q f + \frac{1}{2}\left(\partial_y^2 f + m(q)(\partial_y f)^2\right) = 0$$

with final conditions

$$f(1,q) = \ln\cosh\beta y$$

These equations can be solved by elementary means in the case when $m$ is a step function. It turns out that, for given $m$,

$$\mathbb{E}G(\eta) - \mathbb{E}\widetilde{F}(\kappa)) = f(0, h, m, \beta) - \frac{\beta^2}{2}\int_0^1 qm(q)\,\mathrm{d}r(q)$$

where $h = \beta^{-1}\cosh^{-1}(\mathbb{E}_\sigma\sigma_1)$. This solution was originally obtained using the replica method. The preceding construction gives, at the least, a clear mathematical meaning to the objects involved. In particular, the notion of "ultra-metric zero-dimensional matrices," appears now to be equivalent to ultra-metric structures on the unit interval.

In a recent paper, Talagrand (2003) has proven that converse inequality is also true in the preceding equation, confirming that Parisi's solution yields the correct free energy in a large class of models of the SK type.

## Ghirlanda–Guerra Relations

The appearance of a universal probabilistic structure in the asymptotics of these models may appear surprising. A partial explanation can be found in a set of remarkable identities between multi-overlap distributions that has been discovered first by Ghirlanda and Guerra (1998) in the context of SK models. If $\mu_{\beta,N}^{\otimes n}$ denotes the $n$-fold product Gibbs measure, the Ghirlanda–Guerra relations assert a recursion relation of the form

$$\mathbb{E}\mu_{\beta,N}^{\otimes n+1}\Big(D_N(\sigma^{n+1}, \sigma^k) \le t|\mathcal{B}_n\Big)$$
$$= \frac{1}{n}\sum_{\ell \ne k}\mathbb{E}\mu_{\beta,N}^{\otimes n}\Big(D_N(\sigma^\ell, \sigma^k) \le t|\mathcal{B}_n\Big)$$
$$+ \frac{1}{n}\mathbb{E}\mu_{\beta,N}^{\otimes 2}\big(D_N(\sigma^1, \sigma^2) \le t|\mathcal{B}_n\big) + o(1)$$

These relations hold generically for Gaussian mean-field models, with $D_N$ being the distance through which the covariance is defined. The proof of these relations is based on Gaussian integration-by-parts formulas, and concentration of measure inequalities. In the case of the GREM models, where $D_N$ is ultra-metric, these recursions are sufficient to determine all $n$-replica overlap distributions in terms of the 2-replica distribution. On the other hand, the set of $n$-replica overlap distributions determines the law of the process $\mathcal{K}$ and thus the geometry of the Gibbs measure. In particular, they leave time changes of Neveu's process as the only candidates for limit processes. In the case of the SK models, the same does not hold *a priori*, since the Hamming distance is not an ultra-metric. However, since the Parisi solution is correct, this suggests very strongly that asymptotically the overlap distances are almost surely (with respect to the Gibbs measure) ultra-metric. Then, the Ghirlanda–Guerra identities also imply that the geometry of the Gibbs measures is described by the same structure.

## From Mean-Field to Lattice Models

One of the widely discussed issues in the theory of spin glasses is to what extent the results of mean-field theory are relevant for lattice models. This issue has been addressed elsewhere in this encyclopedia by Newman and Stein. Here, we will only mention a recent result of Franz and Toninelli (2004) that shows that the free energy of the SK model can be represented as the limit of the free energy of lattice models when the range of the interaction tends to zero while their strength tends to zero in an appropriate way (the so-called Kac models). This still leaves open many finer questions, but hints to the fact that mean-field theory bears at least some relevance for realistic spin glasses.

*See also:* Short-Range Spin Glasses: The Metastate Approach; Spin Glasses.

## Further Reading


Aizenman M, Sims R, and Starr SL (2003) An extended variational principle for the SK spin-glass model. *Physical Review B* 6821: 4403.

Amit DJ (1989) *Modelling Brain Function*. Cambridge: Cambridge University Press.

Bovier A (2001) *Statistical Mechanics of Disordered Systems*, MaphySto Lecture Notes, Aarhus University, Aarhus.

Bovier A and Picco P (eds.) (1998) *Mathematical Aspects of Spin Glasses and Neural Networks*, vol. 41. Boston: Birkhäuser.

Ghirlanda S and Guerra F (1998) General properties of overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity. *Journal of Physics A* 31: 9149–9155.

Guerra F (2002) Broken replica symmetry bounds in the mean field spin glass model. *Communications in Mathematical Physics* 233: 1–12.

Guerra F and Toninelli FL (2002) The thermodynamic limit in mean field spin glass models. *Communications in Mathematical Physics* 23: 71–79.

Mézard, Parisi G, and Virasoro MA (1988) *Spin Glass Theory and Beyond*. Singapore: World Scientific.

Ruelle D (1987) A mathematical reformulation of Derrida's REM and GREM. *Communications in Mathematical Physics* 108 (suppl. 2): 225–239.

Sherrington D and Kirkpatrick S (1972) Solvable model of a spin glass. *Physics Review Letters* 35: 1792–1796.

Talagrand M (2003) *Spin Glasses: A Challenge for Mathematicians*. Berlin: Springer.

Talagrand M (2003) The Parisi formula. *Annals of Mathematics* (in press – 2005).

Franz S and Toninelli FL (2004) Finite-range spin glasses in the Kac limit: free energy and local observables. *Journal of Physics A* 37: 7433–7446.

# Measure on Loop Spaces

**H Airault**, Université de Picardie, Amiens, France



## Introduction

Loop spaces have been considered for their geometric interest (Freed Daniel 1988) where the space of based loops on a compact Lie group is endowed with a Kählerian structure; see also the survey by L Gross (1988). The harmonic analysis on loop groups, developed by Pressley and Segal, is reviewed by Hsu (1997). Loop groups have also an impact in string theory (Bowick and Rajeev 1987). They are related to Yang–Mills theory (Levy 2003). A presentation of the history of measure on infinite-dimensional spaces has been given by P Malliavin (see Malliavin (1992) and references therein). The main problem is the construction of measures on the loop space which have quasi-invariance property. This has implications in representation theory (Neretin 1994, Jones 1995). Here we mainly concentrate on the nonlinear stochastic point of view and its interference with geometry. The geometrical study of the space of closed curves over a compact Riemannian manifold $M$, that is, the loop space over $M$, was initiated by Marston Morse in 1932. The loop space is itself a manifold where one can define a Laplace–Beltrami operator. A diffusion process can be considered on this manifold. Wiener defined the Brownian loop by the Fourier series

$$u(\tau) = \sum_{k \geq 1} \frac{\sin k\pi\tau}{k} G_k \qquad [1]$$

where the $G_k$ are independent normal variables. The time evolution of the Wiener loop and the extension of the theory to the case of a compact Riemannian manifold of finite dimension has been considered by Airault and Malliavin (1996, and references therein). The Brownian loop evolutes in the time parameter $t$ as a Brownian sheet where the independent random variables $G_k$ are function of $t$.

Starting from the zero loop, one obtains at time $t$, a random loop, and the law of this loop gives a measure on the loop space. A construction of this measure with functional analysis on infinite-dimensional manifold was done by Gaveau and Mazet (1979). The tools of stochastic analysis are important to the subject. The loop space of continuous maps from the circle to the multiplicative group of complex numbers has a group structure, hence the term "loop group." On the loop group, we consider the multiplicative Brownian motion starting at one point of the circle and conditioned to come back at this point at time $s$. It defines a probability measure on the loop group. One can also consider the set of continuous maps from the circle to the set of complex numbers of modulus equal to 1. The loop group is the space of continuous closed paths on a Lie group. More generally, on a Riemannian manifold $M$, the Brownian motion on $M$ defines a Wiener measure on the loops over $M$. To go from the path space to the loop space, an important tool is the quasisure analysis in infinite dimension. The quasisure analysis was developed by Airault and Malliavin (1996, and references therein) to obtain disintegrations of the Wiener measure and they have used this tool in 1992 to construct measures on the loop group. The main problems are:

1. The construction of heat kernel measures and the existence of a Brownian motion on the loop space, the existence of pinned Wiener measures obtained as the law of Brownian motions conditioned on the loops.
2. The quasi-invariance of these transition probability measures under translation, or multiplication if we have a multiplicative structure, or under the infinitesimal action of suitable vector fields. For the path space over the $n$-dimensional Euclidean space $R^n$, the Cameron–Martin theorem (1944) ensures the existence of a density which shows the quasi-invariance of the Wiener measure under translations. For the quasi-invariance, an important fact is the choice of the metric on the Cameron–Martin space. In the case of the Wiener measure, one considers the paths of finite energy, $\int_0^1 |h'(s)|^2 \mathrm{d}s < +\infty$. This corresponds to the metric "1." P Malliavin (1989, and references therein) discussed the case of metrics $\alpha$ with $1/2 < \alpha < 1$.
3. To define the "good" Cameron subspace, that is, find the vector fields that yield integration-by-parts formulas. The question occurs whether the Cameron–Martin space depends on time. For the loop space, it has been proved by Driver (2003) that it is not the case. A time evolution of the tangent Cameron–Martin space could appear eventually.
4. The determination of the support of the measures (e.g., the Wiener measure) is carried by the set of Hölder functions of order $1/2 - \epsilon$.
5. The absolute continuity of the measures with respect to each other.

## The Construction of Heat Measures on the Loop Space and Their Quasi-Invariance

The construction of measures giving a solution to the infinite-dimensional heat equation as well as the study of the quasi-invariance of the Wiener measure on the path space was started extensively in the work by Bismut, followed by Gross (1998), then by Aida and Elworthy (1995) where the loop group is a suitable manifold to extend to infinite-dimensional manifolds the log-Sobolev inequalities, by Malliavin and Malliavin (1992, and references therein) where the measures on the path space and the path group have been studied. Consider a compact Lie group $G$ with unit $e$ and let $\mathcal{G}$ be its Lie algebra. From the $\mathcal{G}$-valued Brownian motion, one can construct a family of measures $(\mu_t^e)_{t\geq 0}$ on the path space. These measures $\mu_t^e$ are the images of the Wiener measure on $\mathcal{G}$ through the Ito map

$$\mathrm{d}g_x(\tau) = \sqrt{t}\, g_x(\tau)\mathrm{d}x(\tau) \quad \text{with } g_x(0) = e \qquad [2]$$

The convolution of two measures $\mu_t^e$ and $\mu_{t'}^e$ is equal to $\mu_{t+t'}^e$. By choosing the initial value of the path randomly distributed according to the Haar measure on $G$, it defines a family of measures $(\mu_t)_{t\geq 0}$ on the path space with

$$\int f(\gamma)\, \mu_t(\mathrm{d}\gamma) = \int \mathrm{d}g \int f(g\gamma)\mu_t^e(\mathrm{d}\gamma)$$

The Laplacian on the path group is defined by

$$(\Delta_P f)(g) = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[\int f(g\gamma)\mu_\epsilon(\mathrm{d}\gamma) - f(g)\right]$$

The heat equation is valid for the measures $(\mu_t)_{t\geq 0}$ on the paths,

$$\frac{\partial}{\partial t}\int f(g)\mu_t(\mathrm{d}g) = \int (\Delta_P f)(g)\mu_t(\mathrm{d}g)$$

Moreover, there is a quasi-invariance density $k_{g_0}(g)$ defined on the path group ($g_0$ and $g$ are paths with values in $G$) such that

$$\mu_t(g_0 A) = \int_A k_{g_0}(g)\mu_t(\mathrm{d}g)$$

where $g_0 A$ is the translated on the left of the subset $A$ in the path space over $G$. This is a generalization to the path space of the classical Cameron–Martin theorem. Then, one can consider the loop space. The free loop space is the set of continuous maps $g$ from $[0,1]$ to $G$ such that $g(0)=g(1)$, and the loop space with a base point is the set of maps such that $g(0)=g(1)=m$ is fixed. One can define the pinned Brownian motion on the group $G$ to obtain the pinned Wiener measures $(\mu_t^{L_e})_{t\geq 0}$ on the loop group (Malliavin and Malliavin 1992, Driver and Srimurthy 2001). Denote by $p_t(g)$ the solution of the heat equation on the group $G$. Let $g$ be a map from $[0,1]$ to the finite-dimensional Lie group $G$. For $\tau_1, \tau_2, \ldots, \tau_n \in [0,1]$, consider the evaluations of the map $g$, $g_{\tau_1}, g_{\tau_2}, \ldots, g_{\tau_n} \in G$, Let $f$ be a real function defined on $G$ and denote by $\mathrm{d}g$ the Haar measure on $G$. The measure $\mu_t^{L_e}$ on the loop group is given by

$$\begin{aligned}\int f(g_{\tau_1}, g_{\tau_2}, \ldots, g_{\tau_n})\,\mathrm{d}\mu_t^{L_e}(g) &= \int f(g_1, g_2, \ldots, g_n) p_{t\tau_1}(g_1) p_{t(\tau_2-\tau_1)}(g_1^{-1}g_2)\cdots \\ &\times p_{t(\tau_n-\tau_{n-1})}(g_{n-1}^{-1}g_n) p_{t(1-\tau_n)}(g_n)\,\mathrm{d}g_1\cdots \mathrm{d}g_n\end{aligned}$$

From $\mu_t^{L_e}$, one defines a measure $\mu_t^L$ on the free loops by taking the mean over $G$ as

$$\int f(\gamma)\mu_t^L(\mathrm{d}\gamma) = \int_G \mathrm{d}g \int f(g\gamma)\mu_t^{L_e}(\mathrm{d}\gamma)$$

The quasi-invariance property for the pinned Wiener measure was proved by Malliavin and Malliavin (1992).

When the measures $(\mu_t^L)_{t\geq 0}$ are obtained by conditioning and quasisure analysis, we have heat kernel measures. The case of heat kernel measures defined on the loop group has been studied by Airault and Malliavin by disintegrating the measures on the path space and using the quasisure analysis. The Laplacian on the loop group is defined as it has been for the Laplacian on the path space,

$$(\Delta_L f)(g) = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left[\int f(gg_1)\mu_\epsilon^L(\mathrm{d}g_1) - f(g)\right]$$

but now the heat equation has a Kac's potential $\Phi_t$ defined on the loops. On the loop group, the heat equation is

$$\frac{\partial}{\partial t}\int f(l)\mu_t^L(\mathrm{d}l) = \int [(\Delta_L f)(l) + \Phi_t(l) f(l)]\mu_t^L(\mathrm{d}l) \qquad [3]$$

where

$$\Phi_t(l) = \frac{1}{t^2}\left\|\int_0^1 \mathrm{d}l(s)l(s)^{-1}\right\|_{\mathcal{G}}^2 - 2\frac{\mathrm{d}}{\mathrm{d}t}\log p_t(e) - \frac{1}{t}\dim\mathcal{G}$$

The case of the circle, $G = R/2\pi Z$, is interesting. The law of the functional

$$\int_0^1 \mathrm{d}l(s)l(s)^{-1}$$

is given in Airault and Malliavin (1996, and references therein). Moreover, the study of the heat

measures over the loop group of $R/2\pi Z$ brings new identities on the classical Jacobi theta function

$$p_t(\theta) = 1 + 2\sum_{n\geq 1}\cos(n\theta)\,e^{-n^2t/2} \quad \text{at } \theta = 0$$

Let

$$c_t = -2\frac{d}{dt}\log p_t(0) - \frac{1}{t}$$

The following system of differential equations is given by Airault–Malliavin (1996, and references therein):

$$c_t = -\frac{1}{t^2}\sum_{n\in Z} a_n(t)$$

$$\frac{d}{dt}a_n(t) = \frac{1}{2}c_t a_n(t) + \frac{2\pi^2 n^2}{t^2}a_n(t)$$

To pass from path space to loop space, it is convenient to use the "tubular chart" introduced by Gross and the quasisure analysis developed by Airault–Malliavin. Let $\Phi:\gamma\to\gamma(1)\gamma(0)^{-1}$ from the path space to the group $G$; then the free loop space over $G$ is $\Phi^{-1}(e)$. There exists a neighborhood $V$ of the neutral of $G$ such that $\Phi^{-1}(V)$ is diffeomorphic to $V\times L(G)$, the product of $V$ with the loop space over $G$. With this diffeomorphism, one can disintegrate the measures on the path space and obtain the measures on the loop space. The Cameron–Martin formula on the path space of the group $G$ is obtained from the Cameron–Martin formula for the Wiener space and the Ito's map. Let $\gamma$ be a differentiable path with finite energy on $G$, that is,

$$\int_0^1 \left\|\gamma(s)^{-1}\frac{d}{ds}\gamma(s)\right\|_{\mathcal{G}}^2 < +\infty$$

it holds

$$\int f(\gamma g)\mu_t(dg) = \int f(g)k_\gamma(g)\mu_t(dg)$$

Let us denote by $(|)_{\mathcal{G}}$ the Euclidean scalar product on the Lie algebra $\mathcal{G}$; then the density is given by

$$k_\gamma(g) = \exp\left[\frac{1}{t}\int_0^1\left(\gamma(s)^{-1}\frac{d}{ds}\gamma(s)|dg(s)g(s)^{-1}\right)_{\mathcal{G}} - \frac{1}{2t}\int_0^1\left\|\gamma(s)^{-1}\frac{d}{ds}\gamma(s)\right\|_{\mathcal{G}}^2 ds\right]$$

The previous approach relies on the heat equation on the loop space. Thus, the metric on the Cameron–Martin loop or path space is important. The problem of quasi-invariance for metrics $\alpha$ with $1/2<\alpha<1$ relates to the random series

$$u_\alpha(\tau) = \sum_{k\geq 1}\frac{\sin k\pi\tau}{k^\alpha}G_k \qquad [4]$$

where the $G_k$ are independent normal variables. Driver (2003) solved the problem for $1/2<\alpha<1$ by Riemannian geometry in infinite dimension. The Ricci curvature appears in the integration-by-parts formulas on the loop space. The case of the metric 1/2 is out of reach. Fang (1999) calculated the Ricci curvature of the loop manifold for metrics $\alpha>1/2$ and showed that when $\alpha\to 1/2$, these Ricci curvatures tend to a limit. Another presentation of the problem is that of Pickrell (1987), where he obtains a family of quasi-invariant measures on Grassmannians.

Given a family of measures $(\mu_t)_{t\geq 0}$ on the path space of a Riemannian manifold, one defines a heat operator as a family $(\mathcal{L}_t)_{t\geq 0}$ of operators depending upon $t\in[0,+\infty[$ such that

$$\int \mathcal{L}_t F\,d\mu_t = \frac{d}{dt}\int F\,d\mu_t \qquad [5]$$

where $F$ is a function defined on the path space. The heat equation with a potential as [3] gives an example of a heat operator. Heat operators have been constructed for the path space over $R^n$ by Airault–Malliavin, obtaining, after an integration by parts on the path space, a heat operator of first order. This introduces the notion of dilatation vector fields on the path space. In the case of the flat Wiener space, to each point $x$ in the path space is associated the dilatation vector field $Y$ such that $(Yf)(x) = (x|(\text{grad}\,f)(x))$. This gives a rescaling of the Wiener measure under dilatations. This idea has been exploited by Mancino (1999), who extended the method to free loop groups.

## Integration-by-Parts Formulas

The Cameron–Martin space plays the role of the tangent space to the Wiener space. The integration-by-parts formulas are an infinitesimal version of the Cameron–Martin quasi-invariance property. Let $G$ be a compact Lie group or any product of $R^n$ by a compact Lie group. For a vector field $z$, the differentiation on the right $\partial_z^{\text{right}}$ and differentiation on the left $\partial_z^{\text{left}}$ are given by

$$\partial_z^{\text{left}}F(p) = \lim_{\epsilon\to 0}\frac{F(\exp(\epsilon z)p) - F(p)}{\epsilon}$$

and

$$\partial_z^{\text{right}} F(p) = \lim_{\epsilon \to 0} \frac{F(p \exp(\epsilon z)) - F(p)}{\epsilon}$$

The operator $\partial_z^{\text{right}}$ commutes with the translation on the left, for a translation $\tau_h^{\text{left}}$, then $\partial_z^{\text{right}}(F\, o\, \tau_h^{\text{left}}) = (\partial_z^{\text{right}} F) o \tau_h^{\text{left}}$ and vice versa for $\partial_z^{\text{left}}$.

For the measures on the path space or loop space, the problem is to prove the integration-by-parts formulas. On the path spaces on $G$, let $\mu_{P_e}$ be the Wiener measure on the set of paths starting from $e$, there exists a density $k_z$ such that $E[\exp(ck_z)]$ is finite and

$$\int_{P_e(G)} \partial_z^{\text{left}} F(g)\, d\mu_{P_e}(g) = \int_{P_e(G)} F(g) k_z\, d\mu_{P_e}(g)$$

The density $k_z$ is defined on the path space by

$$k_z(g) = \int_0^1 < g(t) z'(t) g(t)^{-1}, d\omega(t) >$$

This was proved by a number of authors (see, e.g., Pickrell (1987) and, in a geometrical context, Cruzeiro and Malliavin (1996)).

The existence of a density for the differentiation on the left is valid for any Lie group. This is not true for the differentiation on the right. If $G$ is noncompact or is not the product of $R^n$ by a compact Lie group, the existence of $k_z$ is not proved on the right. This comes from the fact that the map Ad defined on the path group as a parallel transport does not preserve the Cameron–Martin subspace. In the case where $G$ is not a product of a flat space by a compact Lie group, the Cameron space, which is a kind of "tangent space" to the infinite-dimensional loop manifold, is not closed under the Lie bracket of vector fields.

The integration-by-parts formulas are obtained with the stochastic calculus of variation. On a group $G$, consider $Y_1, Y_2, \ldots, Y_p$, $p$ independent left-invariant vector fields. Let $\mathcal{G}$ be the Lie algebra of $G$. The second-order differential operator $\triangle = \sum_{j=1}^p Y_j^2$ defines a left-invariant diffusion $g_\omega(t)$ on the group $G$ with the stochastic equation $dg_\omega(t) = g_\omega(t)\left[\sum_k (Y_k)_e o\, d\omega^k\right]$ where $(\omega_k)$ are independent Brownian motions on the Euclidean space $\mathcal{G}$. In the work by Malliavin and Malliavin (1992, and references therein), the stochastic calculus of variation is done with the right-invariant connection on the Lie group by setting

$$\phi^{\text{right}} = \frac{d}{d\epsilon}_{|\epsilon=0} (g_{\omega+\epsilon h}) o g_\omega^{-1}$$

where $h$ is a differentiable function of $t$ with values in the Lie algebra $\mathcal{G}$, with finite energy $\int_0^1 |h'(s)|^2\, ds < +\infty$. By taking the derivative with respect to $\epsilon$ in the Stratonovitch equation

$$g^\epsilon(t)^{-1} o\, dg^\epsilon(t) = d\omega(t) + \epsilon h'(t)\, dt$$

and letting $\epsilon = 0$, it turns out that $\phi^{\text{right}}$ is a differentiable function of $t$ and its derivative is given by

$$\frac{d}{dt} \phi^{\text{right}}(t) = g_\omega(t) h'(t) g_\omega(t)^{-1}$$

The situation is not the same for

$$\phi^{\text{left}} = \frac{d}{d\epsilon_{|\epsilon=0}} g_\omega^{-1} o(g_{\omega+\epsilon h})$$

where $d\phi^{\text{left}}(t)$ is a stochastic differential. This generalizes to an arbitrary Riemannian manifold using a coupling of connections (see Airault and Malliavin (1996), and references therein). The construction of the appropriate Cameron subspace, that is, the choice of the infinitesimal action of vector fields on the measure, is of importance. In the commutative case of the path space over $R^n$, the classical Cameron–Martin subspace of paths $h$ such that $\int_0^1 |h'(s)|^2\, ds < +\infty$ is time invariant. To define the vector fields acting on the path (or loop) space over $M$, it is necessary to consider the geometry of the manifold $M$. The infinitesimal transformations which preserve the Riemannian metric are called Riemannian connections. In the case where $M$ is a group, the natural connections are those defined by the parallelism on the group. For a Riemannian manifold, Driver proved the existence of integration-by-parts formulas for the measures on the path space of $M$ when $M$ is endowed with a torsion skew-symmetric connection. The Levi-Civita connection, since it is torsionless, is of course a Driver (2003) connection. If the connection is not skew-symmetric, then two coupled connections permit study of the $\epsilon$-variation or "reduced variation" of a path, and one obtains a Cameron–Martin formula on the path and on the loop space of the Riemannian manifold $M$ (Fang 1999). The method of reduced variation can be used to obtain the integration-by-parts formulas over path and loop spaces. Another approach to the quasi-invariance problem, using two-parameter processes, has been provided by Norris (1995).

## The Support of the Measures and Absolute Continuity with Respect to Each Other

Given a Riemannian manifold $M$, let $(\mu_t)_t$ be the heat kernel measures on the path space of $M$ and let $(\rho_t)_t$ be heat kernel measures on the loop space of $M$; the question arises whether $\rho_t$ is absolutely continuous

with respect to $\mu_t$. For a connected compact Lie group $G$, consider the path and loop groups on $G$. The pinned Wiener measure on the loop group is defined as the law of a $G$-valued Brownian motion starting at $e$ and conditioned to end at $e$, and the heat kernel measure is the endpoint distribution of Brownian motion on the loop group.

It has been shown (Driver and Srimurthy 2001) that the heat kernel measure is absolutely continuous with respect to the pinned Wiener measure, and that the Radon–Nikodym derivative is bounded. This proof relies on the heat formula with a potential [3], which is satisfied by the heat kernel measure. They give a new proof of this heat formula. When the group $G$ is simply connected, Aida and Driver (2000) prove that the heat kernel measure over a based loop group, constructed by using the Brownian motion is equivalent to the Brownian bridge measure over a based loop group. When $G$ is the circle, the Radon–Nikodym derivative of the heat kernel measure with respect to the pinned Wiener measure can be calculated in terms of the Jacobi theta function (Driver and Srimurthy 2001). On the loop space of $R^n$, at time $t$, the two measures, "heat kernel" and "pinned Wiener" are the same.

*See also:* Abelian and Nonabelian Gauge Theories Using Differential Forms; Lie Groups: General Theory; Malliavin Calculus; Path Integrals in Noncommutative Geometry.

## Further Reading


Aida S and Driver BK (2000) Equivalence of heat kernel measure and pinned Wiener measure on loop groups. *Comptes Rendus de l'Academie des Sciences, Paris, I* 331: 709–712.

Aida, Shigeki, Elworthy, and David (1995) Differential calculus on path and loop spaces. Logarithmic Sobolev inequalities on path spaces. *Comptes Rendus de l'Academie des Sciences Paris Serie I Mathematics* 321(1): 97–102.

Airault H and Malliavin P (1996) Integration by parts formulas and dilatation vector fields on elliptic probability spaces. *Probability Theory and Related Fields* 106: 447–494.

Bowick MJ and Rajeev SG (1987) The holomorphic geometry of closed bosonic string theory and Diff$S^1/S^1$. *Nuclear Physics B* 293: 348–384.

Cruzeiro AB and Malliavin P (1996) Renormalized differential geometry on path space: structural equation, curvature. *Journal of Functional Analysis* 139: 119–181.

Driver BK (2003) *Heat Kernels Measures and Infinite Dimensional Analysis. Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces*, Paris, Contemp. Math. vol. 338, pp. 101–141. Providence, RI: American Mathematical Society.

Driver BK and Srimurthy VK (2001) Absolute continuity of heat kernel measure with pinned Wiener measure on loop groups. *Ann. Probab.* 29(2): 691–723.

Fang S (1999) Integration by parts for heat measures over loop groups. *J. Math. Pures Appl.* 78(9): 877–894.

Freed DS (1988) The geometry of loop groups. *Journal of Differential Geometry* 28: 223–276.

Gaveau B and Mazet E (1979) Diffusions sur des variétés de chemins. *Comptes Rendus de l'Academie des Sciences, Paris* 289: 643–645.

Gross L (1998) Harmonic functions on loop groups, Séminaire Bourbaki, vol. 1997–1998, Astérisque No 252, Exp. No. 846, 5, 271–286.

Hsu EP (1997) Analysis on path and loop spaces. In: Hsu EP and Varadhan SRS (eds.) *IAS/Park City Mathematics Series* 5. Princeton: American Mathematical Society and Institute for Advanced Study.

Jones V (1995) Fusion en algèbres de von Neumann et groupes de lacets (A. Wassermann). Séminaire Bourbaki, vol. 1994/95. Astérisque No. 237 (1996), Exp. No. 800, 5, 251–273.

Levy T (2003) Yang–Mills measures on compact surfaces. *Mem. Amer Math. Soc.* 166: 790.

Malliavin P (1989) Hypoellipticity in infinite dimensions. In: Pinsky M (ed.) *Diffusion Process and Related Problems in Analysis*. Chicago: Birkhauser.

Malliavin MP and Malliavin P (1992) Integration on loop groups III. Asymptotic Peter–Weyl orthogonality. *Journal of Functional Analysis* 108: 13–46.

Mancino ME (1999) Dilatation vector fields on the loop group. *Journal of Functional Analysis* 166(1): 130–147.

Neretin Y (1994) Some remarks of quasi-invariant actions of loop groups and the group of the diffeomorphisms of the circle. *Communication in Mathematical Physics* 164: 599–626.

Norris JR (1995) Twisted sheets. *Journal of Functional Analysis* 132: 273–334.

Pickrell D (1987) Measures on infinite-dimensional Grassmann manifolds. *Journal of Functional Analysis* 70(2): 323–356.


# Metastable States

**S Shlosman**, Université de Marseille, Marseille, France



## Introduction

The theory of metastability studies the states of the matter which "should not be there," but which still can be observed, albeit for only a short time. One example is water, cooled below the zero temperature. This supercool water can stay liquid, but not for a long time, and it then freezes abruptly. Such states are called metastable. They are not equilibrium states; at negative temperatures the only equilibrium state of water is ice. Physically, these metastable states are produced from the equilibrium states by slowly changing the external parameters, such as the temperature (or magnetic field): one takes, for example, water (extremely purified) at low positive temperature, $T > 0$, and then lowers the

temperature slowly to negative values $T < 0$. Thus, the family of metastable states, $\mathfrak{s}_T, T < 0$, should be thought as a continuation of the family $\mathfrak{s}_T, T > 0$ of equilibrium states through the point of phase transition $T_c = 0$, at which critical temperature these states cease to exist as equilibrium states.

Below we will present rigorous results, which validate the above picture for the case of the 2D Ising model. They are contained in Schonmann and Shlosman (1998). The relevant external parameter in this case will be the magnetic field, $h$.

It turns out that the lifetime of metastable states is determined by the quantities given by the Wulff construction.

## Equilibrium States and Dynamics

Let us denote the set $\{-1, +1\}^{\mathbb{Z}^2}$ of the Ising model configurations $\sigma$ by $\Omega$. Two configurations are specially relevant, the one with all spins –1 and the one with all spins +1. We will use the simple notation $-$ and $+$ to denote them.

Observables are just functions on $\Omega$. Local observables are those which depend only on the values of finitely many spins.

We will consider the formal Hamiltonian

$$H_h(\sigma) = -\sum_{x,y \text{ n.n.}} \sigma(x)\sigma(y) - h\sum_x \sigma(x) \qquad [1]$$

where $h \in \mathbb{R}^1$ is the external field and $\sigma \in \Omega$ is a generic configuration. We define, for each set $\Lambda \subset\subset \mathbb{Z}^2$ and each boundary condition $\xi \in \Omega$,

$$H_{\Lambda,\xi,h}(\sigma) = -\sum_{\substack{x,y \text{ n.n.}\\ x,y\in\Lambda}} \sigma(x)\sigma(y) - \sum_{\substack{x,y \text{ n.n.}\\ x\in\Lambda, y\notin\Lambda}} \sigma(x)\xi(y) - h\sum_{x\in\Lambda}\sigma(x)$$

The "grand canonical Gibbs measure" in $\Lambda$ with boundary condition $\xi$ under external field $h$ and at temperature $T$ is defined on $\Omega_\Lambda$ as

$$\mu_{\Lambda,\xi,T,h}(\sigma) = Z^{-1}_{\Lambda,\xi,T,h} \exp(-\beta H_{\Lambda,\xi,h}(\sigma))$$

where $\beta = T^{-1}$, and the partition function $Z_{\Lambda,\xi,T,h}$ is a normalization, chosen such that $\mu_{\Lambda,\xi,T,h}(\Omega_\Lambda) = 1$. The equilibrium states are obtained by taking the thermodynamic limit $\lim_{\Lambda\to\mathbb{Z}^2} \mu_{\Lambda,\xi,T,h}$. We will be interested in the states

$$\mu_{\pm,T,h} = \lim_{\Lambda\to\mathbb{Z}^2} \mu_{\Lambda,\pm,T,h}$$

corresponding to ($\pm$)-boundary conditions. If $h \neq 0$, then $\mu_{-,T,h} = \mu_{+,T,h}$, so it will be denoted simply by $\mu_{T,h}$. If $h = 0$, the same is true if the temperature is larger than or equal to a critical value $T_c = T_c$, and is false for $T < T_c$, in which case one says that there is phase coexistence. The measure $\mu_{+,T,0} \equiv \mu_{+,T}$ is called the $(+)$-phase, and $\mu_{-,T}-$ the $(-)$-phase.

For an observable $f$ we will denote by $\langle f \rangle_*$ its expected value in the state $\mu_*$, that is, the integral $\int f \mathrm{d}\mu_*$. In particular, the spontaneous magnetization $m^*(T)$ equals by definition to $\langle\sigma(0)\rangle_{+,T}$.

Next, we need to supply the Ising model with the time evolution. For this we will use the Glauber dynamics. It is a Markov process on $\Omega$, whose generator, $L$, acts on a generic local observable $f$ as

$$(Lf)(\sigma) = \sum_{x\in\mathbb{Z}^2} c(x,\sigma)(f(\sigma^x) - f(\sigma))$$

where $\sigma^x$ is the configuration obtained from $\sigma$ by flipping the spin at the site $x$ to the opposite value, and $c(x,\sigma)$ is the rate of the flip of the spin at the site $x$ when the system is in the state $\sigma$. In words, one can say that the dynamics proceeds as follows: at every site $x$ the spin $\sigma(x)$ is flipped randomly, independently of all others, with the rate $c(x,\sigma)$, where $\sigma$ is the current configuration. Common examples are "metropolis dynamics":

$$c_h(x,\sigma) = \exp(-\beta(\Delta_x H_h(\sigma))^+)$$

or "heat bath dynamics":

$$c_h(x,\sigma) = [1 + \exp(\beta\Delta_x H_h(\sigma))]^{-1}$$

Here $(a)^+ = \max\{a, 0\}$, and $\Delta_x H_h(\sigma) = H_h(\sigma^x) - H_h(\sigma)$. The spin flip system thus obtained will be denoted by $(\sigma^\xi_{T,h;t})_{t\geq 0}$, where $\xi$ is the initial configuration at time $t = 0$. If this initial configuration is selected at random according to a probability measure $\nu$, then the resulting process is denoted by $(\sigma^\nu_{T,h;t})_{t\geq 0}$. It is known that the Gibbs measures are invariant with respect to the stochastic Ising models. Moreover,

$$\sigma^-_{T,h;t} \to \mu_{-,T,h}, \sigma^+_{T,h;t} \to \mu_{+,T,h}, \quad \text{as } t\to\infty$$

We will be interested in the case when $h$ is positive, though small. Then there is only one invariant state, $\mu_{+,T,h}$, so the state $\mu_{-,T,h}$ is equal to $\mu_{+,T,h}$, and $\sigma^-_{T,h;t} \to \mu_{+,T,h}$, as $t\to\infty$. (One should intuitively think about the state $\sigma^-_{T,h;t}$ for $t$ small as the supercooled but liquid water, thinking about the state $\mu_{+,T,h}$ to be ice.) We want to control the convergence of the temporal state $\sigma^-_{T,h;t}$ to the equilibrium, $\mu_{+,T,h}$, and to see, if possible, that during some (long) initial time the state $\sigma^-_{T,h;t}$ looks very similar to the $(-)$-phase $\mu_{-,T}$, while after some time threshold it changes suddenly and looks quite similar to the state $\mu_{+,T,h}$. It turns out that all the above features can indeed be established rigorously.

If one starts to simulate the above dynamics on a computer, then the picture observed would be the following: one would see that droplets of the (+)-phase are created in the midst of (−)-phase droplets, which are there for a while, and then disappear. That process goes on for a while, until a big enough (+)-droplet is born; this one then starts to grow and eventually fills up all the display.

## The Life Span of Metastable States

Let us define the "critical time exponent" $\lambda_c = \lambda_c(T)$ by

$$\lambda_c = \frac{w_\tau}{12m^*(T)T} \qquad [2]$$

where $w_\tau = w_{\tau_T}$ is the value of the surface energy of the Wulff curve of our 2D Ising model at the temperature $T$:

$$w_\tau = \mathcal{W}_\tau(\mathfrak{W}_\tau)$$

Suppose now that $T < T_c, h > 0$. Let $\nu$ be either the (−)-phase $\mu_{-,T}$ or $\delta_{\{\sigma=-\}}$. (In fact, any $\nu$ "between" these two states would go.) Then the following happens.

1. If $0 < \lambda < \lambda_c$, then for each $n \in \{1, 2, \ldots\}$ and for each local observable $f$,

$$\mathbb{E}\left(f\left(\sigma^\nu_{T,h;t=\exp\{\lambda/h\}}\right)\right) = \sum_{j=0}^{n-1} b_j(f)h^j + O(h^n) \qquad [3]$$

   where

$$b_j(f) = \lim_{h\to 0-} \frac{\mathrm{d}^j\langle f\rangle_{-,T,h}}{\mathrm{d}h^j}$$

   (We stress that in the last relation we are using the Gibbs states corresponding to the negative values of the magnetic field.) In particular,

$$\mathbb{E}\left(\sigma^\nu_{T,h;t=\exp\{\lambda/h\}}(0)\right) = -m^*(T) + O(h) \qquad [4]$$

2. If $\lambda > \lambda_c$, then for any finite positive $C$ there is a finite positive $C_1$ such that for every local observable $f$,

$$\left|\mathbb{E}\left(f\left(\sigma^\nu_{T,h;t=\exp\{\lambda/h\}}\right)\right) - \langle f\rangle_{T,h}\right| \le C_1\|f\| \exp\left\{-\frac{C}{h}\right\} \qquad [5]$$

The relation [3] implies that the family of nonequilibrium states $\langle\cdot\rangle^\nu_{T,h;\lambda}, h > 0$, defined for every local observable $f$ by

$$\langle f\rangle^\nu_{T,h;\lambda} = \mathbb{E}\left(f\left(\sigma^\nu_{T,h;t=\exp\{\lambda/h\}}\right)\right)$$

is a $\mathcal{C}^\infty$-continuation of the curve $\{\langle\cdot\rangle_{-,T,h}, h \le 0\}$ of equilibrium states. This is true for every $0 < \lambda < \lambda_c$ and every $\nu$ as above. The states $\langle\cdot\rangle^\nu_{T,h;\lambda}$ are the "metastable states" we are looking for. The relations [3] and [4] should be interpreted in the sense that before the time $\exp\{\lambda_c/h\}$ our temporal state is still "liquid," while [5] means that after the time $\exp\{\lambda_c/h\}$ freezing happens. So one can think about the quantity $\exp\{\lambda_c/h\}$ as being the life span of the metastable state.

This theorem was obtained in Schonmann and Shlosman (1998). Let us explain the heuristics behind it. It has two ingredients. The first one is that the transition to the equilibrium is going via creation of droplets of the (+)-phase. The second one is that once such a droplet is created by a thermal fluctuation, with the size exceeding a certain critical value, it does not die out, but grows further, with a speed $v$ of the order of $h$. (This second belief can be expected to be correct only in dimension 2.) Let us see how these two hypotheses can give us the right answer. To get to the equilibrium we have to overcome the energy barrier, by creating a large droplet of the (+)-phase. Subcritical droplets are constantly created by thermal fluctuations in the metastable phase, but they tend to shrink. On the other hand, once a supercritical droplet is created due to a larger fluctuation, it will grow and drive the system to the stable phase. Indeed, the energy $\Phi(\mathfrak{m})$ of an $\mathfrak{m}$-shaped droplet of the (+)-phase in the sea of (−)-phase equals $\mathcal{W}_\tau(\mathfrak{m}) - 2m^*(T)h\,\mathrm{vol}(\mathfrak{m})$. For small $\mathfrak{m}$ the functional $\Phi(\mathfrak{m})$ decreases as $\mathfrak{m}$ shrinks, while for large $\mathfrak{m}$ the functional $\Phi(\mathfrak{m})$ decreases as $\mathfrak{m}$ grows. Its saddle point $\mathfrak{m}_{sdl}$ is precisely the Wulff shape. Since the minimal height of the barrier is $\Phi(\mathfrak{m}_{sdl})$, one predicts the rate of creation of a critical droplet with center at a given place to be

$$\exp\left\{-\frac{\Phi(\mathfrak{m}_{sdl})}{T}\right\} = \exp\left\{-\frac{w_\tau}{4m^*(T)T}\right\}$$

Comparing with [2], we see that we miss the correct answer

$$\exp\left\{-\frac{w_\tau}{12m^*(T)T}\right\}$$

by a factor of 1/3. The reason for that is the following. Note that we are concerned with an infinite system, and we are observing it through a

local function $f$, which depends on the spins in a finite set supp$(f)$. For us, the system will have relaxed to equilibrium once supp$(f)$ is covered by a big droplet of the $(+)$-phase, which appeared spontaneously somewhere and then grew, as discussed above. We want to estimate how long we have to wait for the probability of such an event to be close to 1. If we suppose that the radius of the supercritical droplet grows with a speed $v$, then we can see that the region in spacetime, where a droplet which covers supp$(f)$ at time $t$ could have appeared, is, roughly speaking, a cone with vertex in supp$(f)$ and which has as base the set of points which have time coordinate 0 and are at most at distance $tv$ from supp$(f)$. The volume of such a cone is of the order of $(vt)^2t$. The order of magnitude of the relaxation time, $t_{\text{rel}}$, at which the region supp$(f)$ starts to be covered by a large droplet can now be obtained by solving the equation

$$(vt_{\text{rel}})^2 t_{\text{rel}} \exp\left\{-\frac{\Phi(\mathfrak{m}_{\text{sdl}})}{T}\right\} \sim 1$$

This gives us what we want:

$$t_{\text{rel}} \sim v^{-2/3} \ \exp\left\{\frac{1}{3}\frac{\Phi(\mathfrak{m}_{\text{sdl}})}{T}\right\}$$

*See also:* Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Large Deviations in Equilibrium Statistical Mechanics; Wulff Droplets.

## Further Reading

Schonmann RH and Shlosman S (1998) Wulff droplets and the metastable relaxation of the kinetic Ising models. *Communications in Mathematical Physics* 194: 389–462.

# Minimal Submanifolds

**T H Colding and W P Minicozzi II**, University of New York, New York, NY, USA



## Introduction

Soap films, soap bubbles, and surface tension were extensively studied by the Belgian physicist and inventor (the inventor of the stroboscope) Joseph Plateau in the first half of the nineteenth century. At least since his studies, it has been known that the right mathematical model for soap films are minimal surfaces – the soap film is in a state of minimum energy when it is covering the least possible amount of area. Minimal surfaces and equations like the minimal surface equation have served as mathematical models for many physical problems.

The field of minimal surfaces dates back to the publication in 1762 of Lagrange's famous memoir "Essai d'une nouvelle méthode pour déterminer les maxima et les minima des formules intégrales indéfinies." Euler had already, in a paper published in 1744, discussed minimizing properties of the surface now known as the catenoid, but he only considered variations within a certain class of surfaces. In the almost one-quarter of a millennium that has past since Lagrange's memoir, the subject of minimal surfaces has remained a vibrant area of research and there are many reasons why. The study of minimal surfaces was the birthplace of regularity theory. It lies on the intersection of nonlinear elliptic PDE, geometry, topology, and general relativity.

In what follows we give a quick tour through many of the classical results in the field of minimal submanifolds, starting at the definition.

The field of minimal surfaces remains extremely active and has very recently seen major developments that have solved many longstanding open problems and conjectures; for more on this, see the expanded version of this survey (Colding and Minicozzi II, 2005). See also the recent surveys (Meeks III and Perez 2004, Perez 2005), and the expository article (Colding and Minicozzi II 2003).

Throughout this survey, we refer to Colding and Minicozzi II (1999) for references unless otherwise noted.

## Part 1. Classical and Almost Classical Results

Let $\Sigma \subset \boldsymbol{R}^n$ be a smooth $k$-dimensional submanifold (possibly with boundary) and $C_0^\infty(N\Sigma)$ the space of all infinitely differentiable, compactly supported, normal vector fields on $\Sigma$. Given $\Phi$ in $C_0^\infty(N\Sigma)$, consider the one-parameter variation

$$\Sigma_{t,\Phi} = \{x + t\,\Phi(x) | x \in \Sigma\} \qquad [1]$$

The so-called first variation formula of volume is the equation (integration is with respect to d(vol)

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \mathrm{Vol}(\Sigma_{t,\Phi}) = \int_\Sigma \langle \Phi, \boldsymbol{H} \rangle \qquad [2]$$

where $\boldsymbol{H}$ is the mean curvature (vector) of $\Sigma$. (When $\Sigma$ is noncompact, then $\Sigma_{t,\Phi}$ in [2] is replaced by

$\Gamma_{t,\Phi}$, where $\Gamma$ is any compact set containing the support of $\Phi$.) The submanifold $\Sigma$ is said to be a "minimal" submanifold (or just minimal) if

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \mathrm{Vol}(\Sigma_{t,\Phi}) = 0 \quad \text{for all } \Phi \in C_0^\infty(N\Sigma) \qquad [3]$$

or, equivalently by [2], if the mean curvature $\boldsymbol{H}$ is identically zero. Thus, $\Sigma$ is minimal if and only if it is a critical point for the volume functional. (Since a critical point is not necessarily a minimum, the term "minimal" is misleading, but it is time honored. The equation for a critical point is also sometimes called the Euler–Lagrange equation.)

Suppose now, for simplicity, that $\Sigma$ is an oriented hypersurface with unit normal $\boldsymbol{n}_\Sigma$. We can then write a normal vector field $\Phi \in C_0^\infty(N\Sigma)$ as $\Phi = \phi \boldsymbol{n}_\Sigma$, where function $\phi$ is in the space $C_0^\infty(\Sigma)$ of infinitely differentiable, compactly supported functions on $\Sigma$. Using this, a computation shows that if $\Sigma$ is minimal, then

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}t^2}\right|_{t=0} \mathrm{Vol}(\Sigma_{t,\phi \boldsymbol{n}_\Sigma}) = -\int_\Sigma \phi\, L_\Sigma \phi \qquad [4]$$

where

$$L_\Sigma \phi = \Delta_\Sigma \phi + |A|^2 \phi \qquad [5]$$

is the second variational (or Jacobi) operator. Here, $\Delta_\Sigma$ is the Laplacian on $\Sigma$ and $A$ is the second fundamental form. So $|A|^2 = \kappa_1^2 + \kappa_2^2 + \cdots + \kappa_{n-1}^2$, where $\kappa_1, \ldots, \kappa_{n-1}$ are the principal curvatures of $\Sigma$ and $\boldsymbol{H} = (\kappa_1 + \cdots + \kappa_{n-1})\, \boldsymbol{n}_\Sigma$. A minimal submanifold $\Sigma$ is said to be stable if

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}t^2}\right|_{t=0} \mathrm{Vol}(\Sigma_{t,\Phi}) \geq 0 \quad \text{for all } \Phi \in C_0^\infty(N\Sigma) \qquad [6]$$

Integrating by parts in [4], we see that stability is equivalent to the so-called stability inequality

$$\int |A|^2 \phi^2 \leq \int |\nabla \phi|^2 \qquad [7]$$

More generally, the "Morse index" of a minimal submanifold is defined to be the number of negative eigenvalues of the operator $L$. Thus, a stable submanifold has Morse index zero.

### The Gauss Map

Let $\Sigma^2 \subset \boldsymbol{R}^3$ be a surface (not necessarily minimal). The Gauss map is a continuous choice of a unit normal $\boldsymbol{n}: \Sigma \to \boldsymbol{S}^2 \subset \boldsymbol{R}^3$. Observe that there are two choices of such a map $\boldsymbol{n}$ and $-\boldsymbol{n}$ corresponding to a choice of orientation of $\Sigma$. If $\Sigma$ is minimal, then the Gauss map is an (anti) conformal map since the eigenvalues of the Weingarten map are $\kappa_1$ and $\kappa_2 = -\kappa_1$. Moreover, for a minimal surface

$$|A|^2 = \kappa_1^2 + \kappa_2^2 = -2\,\kappa_1\,\kappa_2 = -2\,K_\Sigma \qquad [8]$$

where $K_\Sigma$ is the Gauss curvature. It follows that the area of the Gauss map is a multiple of the total curvature.

### Minimal Graphs

Suppose that $u: \Omega \subset \boldsymbol{R}^2 \to \boldsymbol{R}$ is a $C^2$ function. The graph of $u$

$$\mathrm{Graph}_u = \{(x, y, u(x, y)) \mid (x, y) \in \Omega\} \qquad [9]$$

has area

$$\begin{aligned} \mathrm{Area}(\mathrm{Graph}_u) &= \int_\Omega |(1, 0, u_x) \times (0, 1, u_y)| \\ &= \int_\Omega \sqrt{1 + u_x^2 + u_y^2} \\ &= \int_\Omega \sqrt{1 + |\nabla u|^2} \end{aligned} \qquad [10]$$

and the (upward pointing) unit normal is

$$\boldsymbol{n} = \frac{(1, 0, u_x) \times (0, 1, u_y)}{|(1, 0, u_x) \times (0, 1, u_y)|} = \frac{(-u_x, -u_y, 1)}{\sqrt{1 + |\nabla u|^2}} \qquad [11]$$

Therefore, for the graphs $\mathrm{Graph}_{u+t\eta}$ where $\eta|\partial\Omega = 0$, we get that

$$\mathrm{Area}(\mathrm{Graph}_{u+t\eta}) = \int_\Omega \sqrt{1 + |\nabla u + t\,\nabla \eta|^2} \qquad [12]$$

Hence

$$\frac{\mathrm{d}}{\mathrm{d}t_{t=0}} \mathrm{Area}(\mathrm{Graph}_{u+t\eta}) = \int_\Omega \frac{\langle \nabla u, \nabla \eta \rangle}{\sqrt{1 + |\nabla u|^2}} = -\int_\Omega \eta\, \mathrm{div}\left(\frac{\nabla u}{\sqrt{1 + |\nabla u|^2}}\right) \qquad [13]$$

It follows that the graph of $u$ is a critical point for the area functional if and only if $u$ satisfies the divergence form equation

$$\mathrm{div}\left(\frac{\nabla u}{\sqrt{1 + |\nabla u|^2}}\right) = 0 \qquad [14]$$

Next we want to show that the graph of a function on $\Omega$ satisfying the minimal surface equation, that is, satisfying [14], is not just a critical point for the area functional but is actually area minimizing amongst surfaces in the cylinder $\Omega \times \boldsymbol{R} \subset \boldsymbol{R}^3$. To show this, extend first the unit normal $\boldsymbol{n}$ of the graph in [11] to a vector field, still

denoted by $\boldsymbol{n}$, on the entire cylinder $\Omega \times \boldsymbol{R}$. Let $\omega$ be the 2-form on $\Omega \times \boldsymbol{R}$ given that for $X, Y \in \boldsymbol{R}^3$

$$\omega(X, Y) = \det(X, Y, \boldsymbol{n}) \qquad [15]$$

An easy calculation shows that

$$\mathrm{d}\omega = \frac{\partial}{\partial x}\left(\frac{-u_x}{\sqrt{1+|\nabla u|^2}}\right) + \frac{\partial}{\partial y}\left(\frac{-u_y}{\sqrt{1+|\nabla u|^2}}\right) = 0 \qquad [16]$$

since $u$ satisfies the minimal surface equation. In sum, the form $\omega$ is closed and, given any $X$ and $Y$ at a point $(x, y, z)$,

$$|\omega(X, Y)| \leq |X \times Y| \qquad [17]$$

where equality holds if and only if

$$X, Y \subset T_{(x,y,u(x,y))}\mathrm{Graph}_u \qquad [18]$$

Such a form $\omega$ is called a "calibration." From this, we have that if $\Sigma \subset \Omega \times \boldsymbol{R}$ is any other surface with $\partial\Sigma = \partial\,\mathrm{Graph}_u$, then by Stokes' theorem since $\omega$ is closed,

$$\mathrm{Area}(\mathrm{Graph}_u) = \int_{\mathrm{Graph}_u} \omega = \int_\Sigma \omega \leq \mathrm{Area}(\Sigma) \qquad [19]$$

This shows that $\mathrm{Graph}_u$ is area minimizing among all surfaces in the cylinder and with the same boundary. If the domain $\Omega$ is convex, the minimal graph is absolutely area minimizing. To see this, observe first that if $\Omega$ is convex, then so is $\Omega \times \boldsymbol{R}$ and hence the nearest point projection $P: \boldsymbol{R}^3 \to \Omega \times \boldsymbol{R}$ is a distance nonincreasing Lipschitz map that is equal to the identity on $\Omega \times \boldsymbol{R}$. If $\Sigma \subset \boldsymbol{R}^3$ is any other surface with $\partial\Sigma = \partial\,\mathrm{Graph}_u$, then $\Sigma' = P(\Sigma)$ has $\mathrm{Area}(\Sigma') \leq \mathrm{Area}(\Sigma)$. Applying [19] to $\Sigma'$, we see that $\mathrm{Area}(\mathrm{Graph}_u) \leq \mathrm{Area}(\Sigma')$ and the claim follows.

If $\Omega \subset \boldsymbol{R}^2$ contains a ball of radius $r$, then, since $\partial B_r \cap \mathrm{Graph}_u$ divides $\partial B_r$ into two components at least one of which has area at most equal to $(\mathrm{Area}(S^2)/2)r^2$, we get from [19] the crude estimate

$$\mathrm{Area}(B_r \cap \mathrm{Graph}_u) \leq \frac{\mathrm{Area}(\boldsymbol{S}^2)}{2} r^2 \qquad [20]$$

When the domain $\Omega$ is convex, it is not hard to see that the minimal graph is absolutely area minimizing.

Very similar calculations to the ones above show that if $\Omega \subset \boldsymbol{R}^{n-1}$ and $u: \Omega \to \boldsymbol{R}$ is a $C^2$ function, then the graph of $u$ is a critical point for the area functional if and only if $u$ satisfies [14]. Moreover, as in [19], the graph of $u$ is actually area minimizing. Consequently, as in [20], if $\Omega$ contains a ball of radius $r$, then

$$\mathrm{Vol}(B_r \cap \mathrm{Graph}_u) \leq \frac{\mathrm{Vol}(\boldsymbol{S}^{n-1})}{2} r^{n-1} \qquad [21]$$

### The Maximum Principle

The first variation formula, [2], showed that a smooth submanifold is a critical point for area if and only if the mean curvature vanishes. We will next derive the weak form of the first variation formula which is the basic tool for working with "weak solutions" (typically, stationary varifolds). Let $X$ be a vector field on $\boldsymbol{R}^n$. We can write the divergence $\mathrm{div}_\Sigma X$ of $X$ on $\Sigma$ as

$$\begin{aligned}\mathrm{div}_\Sigma X &= \mathrm{div}_\Sigma X^{\mathrm{T}} + \mathrm{div}_\Sigma X^{\mathrm{N}} \\ &= \mathrm{div}_\Sigma X^{\mathrm{T}} + \langle X, \boldsymbol{H}\rangle\end{aligned} \qquad [22]$$

where $X^{\mathrm{T}}$ and $X^{\mathrm{N}}$ are the tangential and normal projections of $X$. In particular, we get that, for a minimal submanifold,

$$\mathrm{div}_\Sigma X = \mathrm{div}_\Sigma X^{\mathrm{T}} \qquad [23]$$

Moreover, from [22] and Stokes' theorem, we see that $\Sigma$ is minimal if and only if for all vector fields $X$ with compact support and vanishing on the boundary of $\Sigma$,

$$\int_\Sigma \mathrm{div}_\Sigma X = 0 \qquad [24]$$

The key point is that [24] makes sense as long as we can define the divergence on $\Sigma$. As a consequence of [24], we will show the following proposition:

**Proposition 1** *$\Sigma^k \subset \boldsymbol{R}^n$ is minimal if and only if the restrictions of the coordinate functions of $\boldsymbol{R}^n$ to $\Sigma$ are harmonic functions.*

**Proof** Let $\eta$ be a smooth function on $\Sigma$ with compact support and $\eta|\partial\Sigma = 0$, then

$$\begin{aligned}\int_\Sigma \langle \nabla_\Sigma \eta, \nabla_\Sigma x_i\rangle &= \int_\Sigma \langle \nabla_\Sigma \eta, e_i\rangle \\ &= \int_\Sigma \mathrm{div}_\Sigma(\eta e_i)\end{aligned} \qquad [25]$$

From this, the claim follows easily. □

Recall that if $\Xi \subset \boldsymbol{R}^n$ is a compact subset, then the smallest convex set containing $\Xi$ (the convex hull, $\mathrm{Conv}(\Xi)$) is the intersection of all half-spaces containing $\Xi$. The maximum principle forces a compact minimal submanifold to lie in the convex hull of its boundary (this is the "convex hull property"):

**Proposition 2** *If $\Sigma^k \subset \boldsymbol{R}^n$ is a compact minimal submanifold, then $\Sigma \subset \mathrm{Conv}(\partial\Sigma)$.*

**Proof** A half-space $H \subset \boldsymbol{R}^n$ can be written as

$$H = \{x \in \boldsymbol{R}^n | \langle x, e \rangle \le a\} \qquad [26]$$

for a vector $e \in \boldsymbol{S}^{n-1}$ and constant $a \in \boldsymbol{R}$. By Proposition 1, the function $u(x) = \langle e, x \rangle$ is harmonic on $\Sigma$ and hence attains its maximum on $\partial\Sigma$ by the maximum principle. □

Another application of [23], with a different choice of vector field $X$, gives that for a $k$-dimensional minimal submanifold $\Sigma$

$$\Delta_\Sigma |x - x_0|^2 = 2 \operatorname{div}_\Sigma (x - x_0) = 2k \qquad [27]$$

Later, we will see that this formula plays a crucial role in the monotonicity formula for minimal submanifolds.

The argument in the proof of the convex hull property can be rephrased as saying that as we translate a hyperplane towards a minimal surface, the first point of contact must be on the boundary. When $\Sigma$ is a hypersurface, this is a special case of the strong maximum principle for minimal surfaces:

**Lemma 1** *Let $\Omega \subset \boldsymbol{R}^{n-1}$ be an open connected neighborhood of the origin. If $u_1, u_2 : \Omega \to \boldsymbol{R}$ are solutions of the minimal surface equation with $u_1 \le u_2$ and $u_1(0) = u_2(0)$, then $u_1 \equiv u_2$.*

Since any smooth hypersurface is locally a graph over a hyperplane, Lemma 1 gives a maximum principle for smooth minimal hypersurfaces.

Thus far, the examples of minimal submanifolds have all been smooth. The simplest nonsmooth example is given by a pair of planes intersecting transversely along a line. To get an example that is not even immersed, one can take three half-planes meeting along a line with an angle of $2\pi/3$ between each adjacent pair.

## Monotonicity and the Mean-Value Inequality

Monotonicity formulas and mean-value inequalities play a fundamental role in many areas of geometric analysis.

**Proposition 3** *Suppose that $\Sigma^k \subset \boldsymbol{R}^n$ is a minimal submanifold and $x_0 \in \boldsymbol{R}^n$; then for all $0 < s < t$,*

$$t^{-k} \operatorname{Vol}(B_t(x_0) \cap \Sigma) - s^{-k} \operatorname{Vol}(B_s(x_0) \cap \Sigma) = \int_{(B_t(x_0) \setminus B_s(x_0)) \cap \Sigma} \frac{|(x - x_0)^N|^2}{|x - x_0|^{k+2}} \qquad [28]$$

Notice that $(x - x_0)^N$ vanishes precisely when $\Sigma$ is conical about $x_0$, that is, when $\Sigma$ is invariant under dilations about $x_0$. As a corollary, we get the following:

**Corollary 1** *Suppose that $\Sigma^k \subset \boldsymbol{R}^n$ is a minimal submanifold and $x_0 \in \boldsymbol{R}^n$; then the function*

$$\Theta_{x_0}(s) = \frac{\operatorname{Vol}(B_s(x_0) \cap \Sigma)}{\operatorname{Vol}(B_s \subset \boldsymbol{R}^k)} \qquad [29]$$

*is a nondecreasing function of $s$. Moreover, $\Theta_{x_0}(s)$ is constant in $s$ if and only if $\Sigma$ is conical about $x_0$.*

Of course, if $x_0$ is a smooth point of $\Sigma$, then $\lim_{s \to 0} \Theta_{x_0}(s) = 1$. We will later see that the converse is also true; this will be a consequence of the Allard regularity theorem.

The monotonicity of area is a very useful tool in the regularity theory for minimal surfaces – at least when there is some *a priori* area bound. For instance, this monotonicity and a compactness argument allow one to reduce many regularity questions to questions about minimal cones (this was a key observation of W Fleming in his work on the Bernstein problem; see the section "The theorems of Bernstein and Bers").

Arguing as in Proposition 3, we get a weighted monotonicity:

**Proposition 4** *If $\Sigma^k \subset \boldsymbol{R}^n$ is a minimal submanifold, $x_0 \in \boldsymbol{R}^n$, and $f$ is a function on $\Sigma$, then*

$$t^{-k} \int_{B_t(x_0) \cap \Sigma} f - s^{-k} \int_{B_s(x_0) \cap \Sigma} f = \int_{(B_t(x_0) \setminus B_s(x_0)) \cap \Sigma} f \frac{|(x - x_0)^N|^2}{|x - x_0|^{k+2}} + \frac{1}{2} \int_s^t \tau^{-k-1} \times \int_{B_\tau(x_0) \cap \Sigma} (\tau^2 - |x - x_0|^2) \Delta_\Sigma f \, d\tau \qquad [30]$$

We get immediately the following mean-value inequality for the special case of non-negative subharmonic functions:

**Corollary 2** *Suppose that $\Sigma^k \subset \boldsymbol{R}^n$ is a minimal submanifold, $x_0 \in \boldsymbol{R}^n$, and $f$ is a non-negative subharmonic function on $\Sigma$; then*

$$s^{-k} \int_{B_s(x_0) \cap \Sigma} f \qquad [31]$$

*is a nondecreasing function of $s$. In particular, if $x_0 \in \Sigma$, then for all $s > 0$,*

$$f(x_0) \le \frac{\int_{B_s(x_0) \cap \Sigma} f}{\operatorname{Vol}(B_s \subset \boldsymbol{R}^k)} \qquad [32]$$

## Rado's Theorem

One of the most basic questions is what does the boundary $\partial\Sigma$ tell us about a compact minimal submanifold $\Sigma$? We have already seen that $\Sigma$ must lie in the convex hull of $\partial\Sigma$, but there are many other theorems of this nature. One of the first theorems is a beautiful result of Rado which says that if $\partial\Sigma$ is a graph over the boundary of a convex set in $\mathbf{R}^2$, then $\Sigma$ is also graph (and hence embedded). The proof of this uses basic properties of nodal lines for harmonic functions.

**Theorem 1** *Suppose that $\Omega \subset \mathbf{R}^2$ is a convex subset and $\sigma \subset \mathbf{R}^3$ is a simple closed curve which is graphical over $\partial\Omega$. Then any minimal disk $\Sigma \subset \mathbf{R}^3$ with $\partial\Sigma = \sigma$ must be graphical over $\Omega$ and hence unique by the maximum principle.*

**Proof** (Sketch). The proof is by contradiction, so suppose that $\Sigma$ is such a minimal disk and $x \in \Sigma$ is a point where the tangent plane to $\Sigma$ is vertical. Consequently, there exists $(a, b) \neq (0, 0)$ such that

$$\nabla_\Sigma(ax_1 + bx_2)(x) = 0 \qquad [33]$$

By Proposition 1, $ax_1 + bx_2$ is harmonic on $\Sigma$ (since it is a linear combination of coordinate functions). The local structure of nodal sets of harmonic functions (see, e.g., Colding and Minicozzi II (1999)) then gives that the level set

$$\{y \in \Sigma | ax_1 + bx_2(y) = ax_1 + bx_2(x)\} \qquad [34]$$

has a singularity at $x$ where at least four different curves meet. If two of these nodal curves were to meet again, then there would be a closed nodal curve which must bound a disk (since $\Sigma$ is a disk). By the maximum principle, $ax_1 + bx_2$ would have to be constant on this disk and hence constant on $\Sigma$ by unique continuation. This would imply that $\sigma = \partial\Sigma$ is contained in the plane given by [34]. Since this is impossible, we conclude that all of these curves go to the boundary without intersecting again.

In other words, the plane in $\mathbf{R}^3$ given by [34] intersects $\sigma$ in at least four points. However, since $\Omega \subset \mathbf{R}^2$ is convex, $\partial\Omega$ intersects the line given by [34] in exactly two points. Finally, since $\sigma$ is graphical over $\partial\Omega$, $\sigma$ intersects the plane in $\mathbf{R}^3$ given by [34] in exactly two points, which gives the desired contradiction. □

## The Theorems of Bernstein and Bers

A classical theorem of S Bernstein from 1916 says that entire (i.e., defined over all of $\mathbf{R}^2$) minimal graphs are planes. This remarkable theorem of Bernstein was one of the first illustrations of the fact that the solutions to a nonlinear PDE, like the minimal surface equation, can behave quite differently from solutions to a linear equation.

**Theorem 2** *If $u : \mathbf{R}^2 \to \mathbf{R}$ is an entire solution to the minimal surface equation, then $u$ is an affine function.*

**Proof** (Sketch). We will show that the curvature of the graph vanishes identically; this implies that the unit normal is constant and, hence, the graph must be a plane. The proof follows by combining two facts. First, the area estimate for graphs [20] gives

$$\text{Area}(B_r \cap \text{Graph}_u) \leq 2\pi r^2 \qquad [35]$$

This quadratic area growth allows one to construct a sequence of non-negative logarithmic cutoff functions $\phi_j$ defined on the graph with $\phi_j \to 1$ everywhere and

$$\lim_{j\to\infty} \int_{\text{Graph}_u} |\nabla\phi_j|^2 = 0 \qquad [36]$$

Moreover, since graphs are area minimizing, they must be stable. We can therefore use $\phi_j$ in the stability inequality [7] to get

$$\int_{\text{Graph}_u} \phi_j^2 |A|^2 \leq \int_{\text{Graph}_u} |\nabla\phi_j|^2 \qquad [37]$$

Combining these gives that $|A|^2$ is zero, as desired. □

Rather surprisingly, this result very much depended on the dimension. The combined efforts of E De Giorgi, F J Almgren Jr., and J Simons finally gave:

**Theorem 3** *If $u : \mathbf{R}^{n-1} \to \mathbf{R}$ is an entire solution to the minimal surface equation and $n \leq 8$, then $u$ is an affine function.*

However, in 1969, E Bombieri, De Giorgi, and E Giusti constructed entire nonaffine solutions to the minimal surface equation on $\mathbf{R}^8$ and an area-minimizing singular cone in $\mathbf{R}^8$. In fact, they showed that for $m \geq 4$, the cones

$$C_m = \{(x_1, \ldots, x_{2m}) \mid x_1^2 + \cdots + x_m^2 = x_{m+1}^2 + \cdots + x_{2m}^2\} \subset \mathbf{R}^{2m} \qquad [38]$$

are area minimizing (and obviously singular at the origin).

In contrast to the entire case, exterior solutions of the minimal graph equation, that is, solutions

on $\mathbf{R}^2 \backslash B_1$, are much more plentiful. In this case, L Bers proved that $\nabla u$ actually has an asymptotic limit:

**Theorem 4** *If $u$ is a $C^2$ solution to the minimal surface equation on $\mathbf{R}^2 \backslash B_1$, then $\nabla u$ has a limit at infinity (i.e., there is an asymptotic tangent plane).*

Bers' theorem was extended to higher dimensions by L Simon:

**Theorem 5** *If $u$ is a $C^2$ solution to the minimal surface equation on $\mathbf{R}^n \backslash B_1$, then either*

(i) *$|\nabla u|$ is bounded and $\nabla u$ has a limit at infinity or*
(ii) *all tangent cones at infinity are of the form $\Sigma \times \mathbf{R}$ where $\Sigma$ is singular.*

Bernstein's theorem has had many other interesting generalizations, some of which will be discussed later.

## Simons Inequality

In this section, we recall a very useful differential inequality for the Laplacian of the norm squared of the second fundamental form of a minimal hypersurface $\Sigma$ in $\mathbf{R}^n$ and illustrate its role in *a priori* estimates. This inequality, originally due to J Simons, is:

**Lemma 2** *If $\Sigma^{n-1} \subset \mathbf{R}^n$ is a minimal hypersurface, then*

$$\Delta_\Sigma |A|^2 = -2|A|^4 + 2|\nabla_\Sigma A|^2 \geq -2|A|^4 \qquad [39]$$

An inequality of the type [39] on its own does not lead to pointwise bounds on $|A|^2$ because of the nonlinearity. However, it does lead to estimates if a "scale-invariant energy" is small. For example, H Choi and Schoen used [39] to prove:

**Theorem 6** *There exists $\epsilon > 0$ so that if $0 \in \Sigma \subset B_r(0)$ with $\partial\Sigma \subset \partial B_r(0)$ is a minimal surface with*

$$\int |A|^2 \leq \epsilon \qquad [40]$$

*then*

$$|A|^2(0) \leq r^{-2} \qquad [41]$$

## Heinz's Curvature Estimate for Graphs

One of the key themes in minimal surface theory is the usefulness of *a priori* estimates. A basic example is the curvature estimate of E Heinz for graphs. Heinz's estimate gives an effective version of the Bernstein's theorem; namely, letting the radius $r_0$ go to infinity in [42] implies that $|A|$ vanishes, thus giving Bernstein's theorem.

**Theorem 7** *If $D_{r_0} \subset \mathbf{R}^2$ and $u : D_{r_0} \to \mathbf{R}$ satisfies the minimal surface equation, then for $\Sigma = \mathrm{Graph}_u$ and $0 < \sigma \leq r_0$*

$$\sigma^2 \sup_{D_{r_0-\sigma}} |A|^2 \leq C \qquad [42]$$

**Proof** (Sketch). Observe first that it suffices to prove the estimate for $\sigma = r_0$, that is, to show that

$$|A|^2(0, u(0)) \leq C r_0^{-2} \qquad [43]$$

Recall that minimal graphs are automatically stable. As in the proof of Theorem 2, the area estimate for graphs [20] allows us to use a logarithmic cutoff function in the stability inequality [7] to get that

$$\int_{B_{r_1} \cap \mathrm{Graph}_u} |A|^2 \leq \frac{C}{\log(r_0/r_1)} \qquad [44]$$

Taking $r_0/r_1$ sufficiently large, we can then apply Theorem 6 to get [43]. □

## Embedded Minimal Disks with Area Bounds

In the early 1980s, Schoen and Simon extended the theorem of Bernstein to complete simply connected embedded minimal surfaces in $\mathbf{R}^3$ with quadratic area growth. A surface $\Sigma$ is said to have quadratic area growth if for all $r > 0$, the intersection of the surface with the ball in $\mathbf{R}^3$ of radius $r$ and center at the origin is bounded by $Cr^2$ for a fixed constant $C$ independent of $r$.

**Theorem 8** *Let $0 \in \Sigma^2 \subset B_{r_0} = B_{r_0}(x) \subset \mathbf{R}^3$ be an embedded simply connected minimal surface with $\partial\Sigma \subset \partial B_{r_0}$. If $\mu > 0$ and either*

$$\mathrm{Area}(\Sigma) \leq \mu r_0^2 \quad \text{or} \quad \int_\Sigma |A|^2 \leq \mu \qquad [45]$$

*then for the connected component $\Sigma'$ of $B_{r_0/2}(x_0) \cap \Sigma$ with $0 \in \Sigma'$ we have*

$$\sup_{\Sigma'} |A|^2 \leq C r_0^{-2} \qquad [46]$$

*for some $C = C(\mu)$.*

The result of Schoen–Simon was generalized by Colding–Minicozzi to quadratic area growth for intrinsic balls (this generalization played an important role in analyzing the local structure of embedded minimal surfaces):

**Theorem 9** *Given a constant $C_I$, there exists $C_P$ so that if $\mathcal{B}_{2r_0} \subset \Sigma \subset \mathbf{R}^3$ is an embedded minimal disk satisfying either*

$$\mathrm{Area}(\mathcal{B}_{2r_0}) \leq C_I r_0^2 \quad \text{or} \quad \int_{\mathcal{B}_{2r_0}} |A|^2 \leq C_I \qquad [47]$$

*then*

$$\sup_{\mathcal{B}_s} |A|^2 \leq C_P s^{-2} \quad [48]$$

As an immediate consequence, letting $r_0 \to \infty$ gives Bernstein-type theorems for embedded simply connected minimal surfaces with either bounded density or finite total curvature. Note that Enneper's surface is simply connected but neither flat nor embedded; this shows that embeddedness is essential for these estimates. Similarly, the catenoid shows that the surface being simply connected is essential. The catenoid is the minimal surface in $\boldsymbol{R}^3$ given by

$$\{(\cosh s \cos t, \cosh s \sin t, s) | s, t \in \boldsymbol{R}\} \quad [49]$$

## Stable Minimal Surfaces

It turns out that stable minimal surfaces have *a priori* estimates. Since minimal graphs are stable, the estimates for stable surfaces can be thought of as generalizations of the earlier estimates for graphs. These estimates have been widely applied and are particularly useful when combined with existence results for stable surfaces (such as the solution of the Plateau problem). The starting point for these estimates is that, as we saw in [4], stable minimal surfaces satisfy the stability inequality

$$\int |A|^2 \phi^2 \leq \int |\nabla \phi|^2 \quad [50]$$

We will mention two such estimates. The first is R Schoen's curvature estimate for stable surfaces:

**Theorem 10** *There exists a constant C so that if* $\Sigma \subset \boldsymbol{R}^3$ *is an immersed stable minimal surface with trivial normal bundle and* $\mathcal{B}_{r_0} \subset \Sigma \backslash \partial \Sigma$, *then*

$$\sup_{\mathcal{B}_{r_0 - \sigma}} |A|^2 \leq C \sigma^{-2} \quad [51]$$

The second is an estimate for the area and total curvature of a stable surface is due to Colding–Minicozzi; for simplicity, we will state only the area estimate:

**Theorem 11** *If* $\Sigma \subset \boldsymbol{R}^3$ *is an immersed stable minimal surface with trivial normal bundle and* $\mathcal{B}_{r_0} \subset \Sigma \backslash \partial \Sigma$, *then*

$$\text{Area}(\mathcal{B}_{r_0}) \leq 4\pi r_0^2 / 3 \quad [52]$$

As mentioned, we can use [52] to bound the energy of a cutoff function in the stability inequality and, thus, bound the total curvature of sub-balls. Combining this with the curvature estimate of Theorem 6 gives Theorem 10. Note that the bound [53] is surprisingly sharp; even when $\Sigma$ is a plane, the area is $\pi r_0^2$.

## Regularity Theory

In this section, we survey some of the key ideas in classical regularity theory, such as the role of monotonicity, scaling, $\epsilon$-regularity theorems (such as Allard's theorem) and tangent cone analysis (such as Almgren's refinement of Federer's dimension reducing). We refer to the book by Morgan (1995) for a more detailed overview and a general introduction to geometric measure theory.

The starting point for all of this is the monotonicity of volume for a minimal $k$-dimensional submanifold $\Sigma$. Namely, Corollary [1] gives that the density

$$\Theta_{x_0}(s) = \frac{\text{Vol}(B_s(x_0) \cap \Sigma)}{\text{Vol}(B_s \subset \boldsymbol{R}^k)} \quad [53]$$

is a monotone nondecreasing function of $s$. Consequently, we can define the density $\Theta_{x_0}$ at the point $x_0$ to be the limit as $s \to 0$ of $\Theta_{x_0}(s)$. It also follows easily from monotonicity that the density is semicontinuous as a function of $x_0$.

### $\epsilon$-Regularity and the Singular Set

An $\epsilon$-regularity theorem is a theorem giving that a weak (or generalized) solution is actually smooth at a point if a scale-invariant energy is small enough there. The standard example is the Allard regularity theorem:

**Theorem 12** *There exists* $\delta(k, n) > 0$ *such that if* $\Sigma \subset \boldsymbol{R}^n$ *is a k-rectifiable stationary varifold (with density at least one a.e.),* $x_0 \in \Sigma$, *and*

$$\Theta_{x_0} = \lim_{r \to 0} \frac{\text{Vol}(B_r(x_0) \cap \Sigma)}{\text{Vol}(B_r \subset \boldsymbol{R}^k)} < 1 + \delta \quad [54]$$

*then* $\Sigma$ *is smooth in a neighborhood of* $x_0$.

Similarly, the small total curvature estimate of Theorem 6 may be thought of as an $\epsilon$-regularity theorem; in this case, the scale-invariant energy is $\int |A|^2$.

As an application of the $\epsilon$-regularity theorem, Theorem [12], we can define the singular set $\mathcal{S}$ of $\Sigma$ by

$$\mathcal{S} = \{x \in \Sigma | \Theta_x \geq 1 + \delta\} \quad [55]$$

It follows immediately from the semicontinuity of the density that $\mathcal{S}$ is closed. In order to bound the size of the singular set (e.g., the Hausdorff measure), one combines the $\epsilon$-regularity with simple covering arguments.

This preliminary analysis of the singular set can be refined by doing a so-called tangent cone analysis.

### Tangent Cone Analysis

It is not hard to see that scaling preserves the space of minimal submanifolds of $\boldsymbol{R}^n$. Namely, if $\Sigma$ is minimal, then so is

$$\Sigma_{y,\lambda} = \{y + \lambda^{-1}(x - y) | x \in \Sigma\} \qquad [56]$$

(To see this, simply note that this scaling multiplies the principal curvatures by $\lambda$.) Suppose now that we fix the point $y$ and take a sequence $\lambda_j \to 0$. The monotonicity formula bounds the density of the rescaled solution, allowing us to extract a convergent subsequence and limit. This limit, which is called a "tangent cone" at $y$, achieves equality in the monotonicity formula and, hence, must be homogeneous (i.e., invariant under dilations about $y$).

The usefulness of tangent cone analysis in regularity theory is based on two key facts. For simplicity, we illustrate these when $\Sigma \subset \boldsymbol{R}^n$ is an area-minimizing hypersurface. First, if any tangent cone at $y$ is a hyperplane $\boldsymbol{R}^{n-1}$, then $\Sigma$ is smooth in a neighborhood of $y$. This follows easily from the Allard regularity theorem since the density at $y$ of the tangent cone is the same as the density at $y$ of $\Sigma$. The second key fact, known as "dimension reducing," is due to Almgren and is a refinement of an argument of Federer. To state this, we first stratify the singular set $\mathcal{S}$ of $\Sigma$ into subsets

$$\mathcal{S}_0 \subset \mathcal{S}_1 \subset \cdots \subset \mathcal{S}_{n-2} \qquad [57]$$

where we define $\mathcal{S}_i$ to be the set of points $y \in \mathcal{S}$ so that any linear space contained in any tangent cone at $y$ has dimension at most $i$. (Note that $\mathcal{S}_{n-1} = \emptyset$ by Allard's theorem.) The dimension reducing argument then gives that

$$\dim(\mathcal{S}_i) \leq i \qquad [58]$$

where dimension means the Hausdorff dimension. In particular, the solution of the Bernstein problem then gives codimension-7 regularity of $\Sigma$, that is, $\dim(\mathcal{S}) \leq n - 8$.

## Part 2. Constructing Minimal Surfaces

Thus far, we have mainly dealt with regularity and *a priori* estimates but have ignored questions of existence. In this part, we survey some of the most useful existence results for minimal surfaces. The following section gives an overview of the classical Plateau problem. Next, we recall the classical Weierstrass representation, including a few modern applications, and the Kapouleas desingularization method. Then we deal with producing area-minimizing surfaces and questions of embeddedness. Finally, we recall the min–max construction for producing unstable minimal surfaces and, in particular, doing so while controlling the topology and guaranteeing embeddedness.

## The Plateau Problem

The following fundamental existence problem for minimal surfaces is known as the Plateau problem: given a closed curve $\Gamma$, find a minimal surface with boundary $\Gamma$. There are various solutions to this problem depending on the exact definition of a surface (parametrized disk, integral current, $\boldsymbol{Z}_2$ current, or rectifiable varifold). We shall consider the version of the Plateau problem for parametrized disks; this was solved independently by J Douglas and T Rado. The generalization to Riemannian manifolds is due to C B Morrey.

**Theorem 13** *Let $\Gamma \subset \boldsymbol{R}^3$ be a piecewise $C^1$ closed Jordan curve. Then there exists a piecewise $C^1$ map $u$ from $D \subset \boldsymbol{R}^2$ to $\boldsymbol{R}^3$ with $u(\partial D) \subset \Gamma$ such that the image minimizes area among all disks with boundary $\Gamma$.*

The solution $u$ to the Plateau problem above can easily be seen to be a branched conformal immersion. R Osserman proved that $u$ does not have true interior branch points; subsequently, R Gulliver and W Alt showed that $u$ cannot have false branch points either.

Furthermore, the solution $u$ is as smooth as the boundary curve, even up to the boundary. A very general version of this boundary regularity was proved by S Hildebrandt; for the case of surfaces in $\boldsymbol{R}^3$, recall the following result of J C C Nitsche:

**Theorem 14** *If $\Gamma$ is a regular Jordan curve of class $C^{k,\alpha}$ where $k \geq 1$ and $0 < \alpha < 1$, then a solution $u$ of the Plateau problem is $C^{k,\alpha}$ on all of $\bar{D}$.*

## The Weierstrass Representation

The classical Weierstrass representation (see Osserman (1986)) takes holomorphic data (a Riemann surface, a meromorphic function, and a holomorphic 1-form) and associates a minimal surface in $\boldsymbol{R}^3$. To be precise, given a Riemann surface $\Omega$, a meromorphic function $g$ on $\Omega$, and a holomorphic 1-form $\phi$ on $\Omega$, then we

get a (branched) conformal minimal immersion $F:\Omega\to\boldsymbol{R}^3$ by

$$F(z)=\mathrm{Re}\int_{\zeta\in\gamma_{z_0,z}}\left(\frac{1}{2}(g^{-1}(\zeta)-g(\zeta)),\right.\\ \left.\frac{\mathrm{i}}{2}(g^{-1}(\zeta)+g(\zeta)),1\right)\phi(\zeta)\qquad[59]$$

Here, $z_0\in\Omega$ is a fixed base point and the integration is along a path $\gamma_{z_0,z}$ from $z_0$ to $z$. The choice of $z_0$ changes $F$ by adding a constant. In general, the map $F$ may depend on the choice of path (and hence may not be well defined); this is known as "the period problem." However, when $g$ has no zeros or poles and $\Omega$ is simply connected, then $F(z)$ does not depend on the choice of path $\gamma_{z_0,z}$.

Two standard constructions of minimal surfaces from Weierstrass data are

$$g(z)=z,\ \phi(z)=\mathrm{d}z/z,\ \Omega=\mathbf{C}\backslash\{0\}\quad\text{giving a catenoid}\qquad[60]$$

$$g(z)=\mathrm{e}^{\mathrm{i}z},\ \phi(z)=\mathrm{d}z,\ \Omega=\mathbf{C}\ \text{giving a helicoid}\qquad[61]$$

The Weierstrass representation is particularly useful for constructing immersed minimal surfaces. Typically, it is rather difficult to prove that the resulting immersion is an embedding (i.e., is 1–1), although there are some interesting cases where this can be done. For the first modern example, D Hoffman and Meeks proved that the surface constructed by Costa was embedded; this was the first new complete finite topology properly embedded minimal surface discovered since the classical catenoid, helicoid, and plane. This led to the discovery of many more such surfaces (see Rosenberg (1992) for more discussion).

## Area-Minimizing Surfaces

Perhaps the most natural way to construct minimal surfaces is to look for ones which minimize area, for example, with fixed boundary, or in a homotopy class, etc. This has the advantage that often it is possible to show that the resulting surface is embedded. We mention a few results along these lines.

The first embeddedness result, due to Meeks and Yau, shows that if the boundary curve is embedded and lies on the boundary of a smooth mean convex set (and it is null-homotopic in this set), then it bounds an embedded least area disk.

**Theorem 15** (Meeks III and Yau 1982). *Let $M^3$ be a compact Riemannian 3-manifold whose boundary is mean convex and let $\gamma$ be a simple closed curve in $\partial M$ which is null-homotopic in $M$; then $\gamma$ is bounded by a least area disk and any such least area disk is properly embedded.*

Note that some restriction on the boundary curve $\gamma$ is certainly necessary. For instance, if the boundary curve was knotted (e.g., the trefoil), then it could not be spanned by any embedded disk (minimal or otherwise). Prior to the work of Meeks and Yau, embeddedness was known for extremal boundary curves in $\boldsymbol{R}^3$ with small total curvature by the work of R Gulliver and J Spruck.

If we instead fix a homotopy class of maps, then the two fundamental existence results are due to Sacks–Uhlenbeck and Schoen–Yau (with embeddedness proved by Meeks–Yau and Freedman–Hass–Scott, respectively):

**Theorem 16** *Given $M^3$, there exist conformal (stable) minimal immersions $u_1,\ldots,u_m:\boldsymbol{S}^2\to M$ which generate $\pi_2(M)$ as a $\mathbf{Z}[\pi_1(M)]$ module. Furthermore,*

(i) *if $u:\boldsymbol{S}^2\to M$ and $[u]_{\pi_2}\neq 0$, then* $\mathrm{Area}(u)\geq\min_i\mathrm{Area}(u_i)$,
(ii) *each $u_i$ is either an embedding or a 2–1 map onto an embedded two-sided $\boldsymbol{RP}^2$.*

**Theorem 17** *If $\Sigma^2$ is a closed surface with genus $g>0$ and $i_0:\Sigma\to M^3$ is an embedding which induces an injective map on $\pi_1$, then there is a least area embedding with the same action on $\pi_1$.*

## The Min–Max Construction of Minimal Surfaces

Variational arguments can also be used to construct higher index (i.e., nonminimizing) minimal surfaces using the topology of the space of surfaces. There are two basic approaches:

1. Applying Morse theory to the energy functional on the space of maps from a fixed surface $\Sigma$ to $M$.
2. Doing a min–max argument over families of (topologically nontrivial) sweep-outs of $M$.

The first approach has the advantage that the topological type of the minimal surface is easily fixed; however, the second approach has been more successful at producing embedded minimal surfaces. We will highlight a few key results below but refer to Colding and De Lellis (2003) for a thorough treatment.

Unfortunately, one cannot directly apply Morse theory to the energy functional on the space of maps from a fixed surface because of a lack of compactness (the Palais–Smale condition C does not hold).

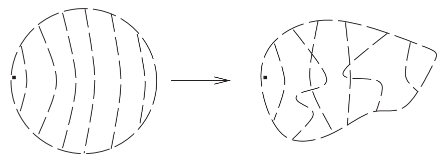


**Figure 1** A one-parameter family of curves on a 2-sphere which induces a map $F: \boldsymbol{S}^2 \to \boldsymbol{S}^2$ of degree 1. First published in Surveys in Differential Geometry, volume IX, in 2004, published by International Press.

To get around this difficulty, Sacks–Uhlenbeck introduce a family of perturbed energy functionals which do satisfy condition C and then obtain minimal surfaces as limits of critical points for the perturbed problems:

**Theorem 18** *If $\pi_k(M) \neq 0$ for some $k > 1$, then there exists a branched immersed minimal 2-sphere in $M$ (for any metric).*

The basic idea of constructing minimal surfaces via min–max arguments and sweep-outs goes back to Birkhoff, who developed it to construct simple closed geodesics on spheres. In particular, when $M$ is a topological 2-sphere, we can find a one-parameter family of curves starting and ending at point curves so that the induced map $F: S^2 \to S^2$ (see Figure 1) has nonzero degree. The min–max argument produces a nontrivial closed geodesic of length less than or equal to the longest curve in the initial one-parameter family. A curve-shortening argument gives that the geodesic obtained in this way is simple.

J Pitts applied a similar argument and geometric measure theory to get that every closed Riemannian 3-manifold has an embedded minimal surface (his argument was for dimensions up to seven), but he did not estimate the genus of the resulting surface. Finally, F Smith (under the direction of L Simon) proved (see Colding and De Lellis (2003)):

**Theorem 19** *Every metric on a topological 3-sphere $M$ admits an embedded minimal 2-sphere.*

The main new contribution of Smith was to control the topological type of the resulting minimal surface while keeping it embedded.

## Part 3. Some Applications of Minimal Surfaces

In this part, we discuss very briefly a few applications of minimal surfaces. As mentioned in the introduction, there are many to choose from and we have selected just a few.

### The Positive-Mass Theorem

The (Riemannian version of the) positive-mass theorem states that an asymptotically flat 3-manifold $M$ with non-negative scalar curvature must have positive mass. The Riemannian manifold $M$ here arises as a maximal spacelike slice in a $(3+1)$-dimensional spacetime solution of Einstein's equations.

The asymptotic flatness of $M$ arises because the spacetime models an isolated gravitational system and hence is a perturbation of the vacuum solution outside a large compact set. To make this precise, suppose for simplicity that $M$ has only one end; $M$ is then said to be asymptotically flat if there is a compact set $\Omega \subset M$ so that $M \backslash \Omega$ is diffeomorphic to $\boldsymbol{R}^3 \backslash B_R(0)$ and the metric on $M \backslash \Omega$ can be written as

$$g_{ij} = \left(1 + \frac{\mathcal{M}}{2|x|}\right)^4 \delta_{ij} + p_{ij} \qquad [62]$$

where

$$|x|^2|p_{ij}| + |x|^3|Dp_{ij}| + |x|^4|D^2p_{ij}| \le C \qquad [63]$$

The constant $\mathcal{M}$ is the so-called mass of $M$. Observe that the metric $g_{ij}$ is a perturbation of the metric on a constant-time slice in the Schwarzschild spacetime of mass $\mathcal{M}$; that is to say, the Schwarzschild metric has $p_{ij} \equiv 0$.

A tensor $h$ is said to be $O(|x|^{-p})$ if $|x|^p|h| + |x|^{p+1}|Dh| \le C$. For example, an easy calculation shows that

$$\begin{aligned} g_{ij} &= (1 + 2\mathcal{M}/|x|)\,\delta_{ij} + O(|x|^{-2}) \\ \sqrt{g} &\equiv \sqrt{\det g_{ij}} = 1 + 3\mathcal{M}|x|^{-1} + O(|x|^{-2}) \end{aligned} \qquad [64]$$

The positive-mass theorem states that the mass $\mathcal{M}$ of such an $M$ must be non-negative:

**Theorem 20** (Schoen and Yau 1979). *With $M$ as above, $\mathcal{M} \ge 0$.*

There is a rigidity theorem as well which states that the mass vanishes only when $M$ is isometric to $\boldsymbol{R}^3$:

**Theorem 21** (Schoen and Yau 1979). *If $|\nabla^3 p_{ij}| = O(|x|^{-5})$ and $\mathcal{M} = 0$ in Theorem 20, then $M$ is isometric to $\mathbf{R}^3$.*

We will give a very brief overview of the proof of Theorem 20, showing in the process where minimal surfaces appear.

**Proof** (Sketch). The argument will be by contradiction, so suppose that the mass is negative. It is not hard to prove that the slab between two parallel

planes is mean convex. That is, we have the following:

**Lemma 3** *If $\mathcal{M} < 0$ and $M$ is asymptotically flat, then there exist $R_0, h > 0$ so that for $r > R_0$ the sets*

$$C_r = \{|x|^2 \le r^2, -h \le x_3 \le h\} \qquad [65]$$

*have strictly mean-convex boundary.*

Since the compact set $C_r$ is mean convex, we can solve the Plateau problem to get an area-minimizing (and hence stable) surface $\Gamma_r \subset C_r$ with boundary

$$\partial\Gamma_r = \{|x|^2 = r^2, x_3 = h\} \qquad [66]$$

Using the disk $\{|x|^2 \le r^2, x_3 = h\}$ as a comparison surface, we get uniform local area bounds for any such $\Gamma_r$. Combining these local area bounds with the *a priori* curvature estimates for minimizing surfaces, we can take a sequence of $r$'s going to infinity and find a subsequence of $\Gamma_r$'s that converge to a complete area-minimizing surface

$$\Gamma \subset \{-h \le x_3 \le h\} \qquad [67]$$

Since $\Gamma$ is pinched between the planes $\{x_3 = \pm h\}$, the estimates for minimizing surfaces implies that (outside a large compact set) $\Gamma$ is a graph over the plane $\{x_3 = 0\}$ and hence has quadratic area growth and finite total curvature. Moreover, using the form of the metric $g_{ij}$, we see that $|\nabla u|$ decays like $|x|^{-1}$ and

$$\int_{\sigma_s} k_g = (2\pi s + O(1))(s^{-1} + O(s^{-2})) = 2\pi + O(s^{-1}) \qquad [68]$$

where $\sigma_s = \{x_1^2 + x_2^2 = s^2\} \cap \Gamma$ and $k_g$ is the geodesic curvature of $\sigma_s$ (as a curve in $\Gamma$).

To get the contradiction, one combines stability of $\Gamma$ with the positive scalar curvature of $M$ to see that no such $\Gamma$ could have existed. ($M$ was assumed only to have non-negative scalar curvature. However, a "rounding off" argument shows that the metric on $M$ can be perturbed to have positive scalar curvature outside of a compact set and still have negative mass.) Namely, substituting the Gauss equation into the stability inequality (this is the stability inequality in a general 3-manifold; see Colding and Minicozzi II (1999)) gives

$$\int_\Gamma (|A|^2/2 + \text{Scal}_M - K_\Sigma)\phi^2 \le \int_\Gamma |\nabla\phi|^2 \qquad [69]$$

Since $\Gamma$ has quadratic area growth, we can choose a sequence of (logarithmic) cutoff functions in [69] to get

$$0 < \int_\Sigma (|A|^2/2 + \text{Scal}_M) \le \int_\Sigma K_\Sigma < \infty \qquad [70]$$

since $K_\Sigma$ may not be positive, we also used that $\Gamma$ has finite total curvature. Moreover, we used that $\text{Scal}_M$ is positive outside a compact set to see that the first integral in [70] was positive. Finally, substituting [70] into the Gauss–Bonnet formula gives that $\int_{\sigma_s} k_g$ is strictly less than $2\pi$ for $s$ large, contradicting [68].

### Black holes

Another way that minimal surfaces enter into relativity is through black holes. Suppose that we have a three-dimensional time slice $M$ in a $(3+1)$-dimensional spacetime. For simplicity, assume that $M$ is totally geodesic and hence has non-negative scalar curvature. A closed surface $\Sigma$ in $M$ is said to be trapped if its mean curvature is everywhere negative with respect to its outward normal. Physically, this means that the surface emits an outward shell of light whose surface area is decreasing everywhere on the surface. The existence of a closed trapped surface implies the existence of a black hole in the spacetime.

Given a trapped surface, we can look for the outermost trapped surface containing it; this outermost surface is called an apparent horizon. It is not hard to see that an apparent horizon must be a minimal surface and, moreover, a barrier argument shows that it must be stable. Since $M$ has non-negative scalar curvature, stability in turn implies that it must be diffeomorphic to a sphere. See, for instance, Bray (2002) for references to some results on black holes, horizons, etc.

## Constant Mean Curvature Surfaces

At least since the time of Plateau, minimal surfaces have been used to model soap films. This is because the mean curvature of the surface models the surface tension and this is essentially the only force acting on a soap film. Soap bubbles, on other hand, enclose a volume and thus the pressure gives a second counterbalancing force. It follows easily that these two forces are in equilibrium when the surface has constant mean curvature (cmc).

For the same reason, cmc surfaces arise in the isoperimetric problem. Namely, a surface that minimizes surface area while enclosing a fixed volume must have cmc. It is not hard to see that such an isoperimetric surface in $\mathbf{R}^n$ must be a round sphere. There are two interesting partial converses to this. First, by a theorem of Hopf, any cmc 2-sphere in $\mathbf{R}^3$ must be round. Second, using the maximum principle ("the method of moving planes"), Alexandrov showed that any closed embedded cmc hypersurface in $\mathbf{R}^n$ must be a round sphere. It turned out, however, that not every closed immersed cmc surface is round. The

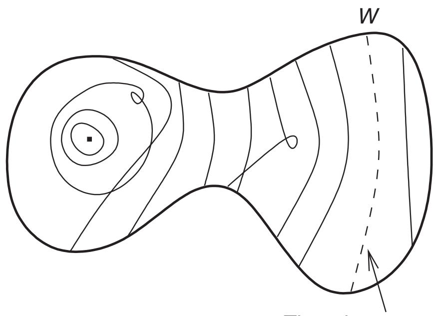



**Figure 2** The sweep-out, the min–max surface, and the width *W*. First published in the *Journal of the American Mathematical Society* in 2005, published by the American Mathematical Society.

first examples were immersed cmc tori constructed by H Wente. Kapouleas constructed many new examples, including closed higher-genus cmc surfaces.

Many of the techniques developed for studying minimal surfaces generalize to general cmc surfaces.

### Finite Extinction for Ricci Flow

We close this article by indicating how minimal surfaces can be used to show that on a homotopy 3-sphere the Ricci flow becomes extinct in finite time (see Colding and Minicozzi II (2005) and Perelman (2003) for details).

Let $M^3$ be a smooth closed orientable 3-manifold and let $g(t)$ be a one-parameter family of metrics on $M$ evolving by the Ricci flow, so

$$\partial_t g = -2\mathrm{Ric}_{M_t} \qquad [71]$$

In an earlier section, we saw that there is a natural way of constructing minimal surfaces on many 3-manifolds and that comes from the min–max argument where the minimal of all maximal slices of sweep-outs is a minimal surface. The idea is then to look at how the area of this min–max surface changes under the flow. Geometrically, the area measures a kind of width of the 3-manifold and as we will see for certain 3-manifolds (those, like the 3-sphere, whose prime decomposition contains no aspherical factors), the area becomes zero in finite time corresponding to the solution becoming extinct in finite time.

Fix a continuous map $\beta: [0,1] \to C^0 \cap L_1^2(S^2, M)$ where $\beta(0)$ and $\beta(1)$ are constant maps so that $\beta$ is in the nontrivial homotopy class $[\beta]$ (such $\beta$ exists when $M$ is a homotopy 3-sphere). We define the width $W = W(g, [\beta])$ by

$$W(g) = \min_{\gamma \in [\beta]} \max_{s \in [0,1]} \mathrm{Energy}(\gamma(s)) \qquad [72]$$

The next theorem gives an upper bound for the derivative of $W(g(t))$ under the Ricci flow which forces the solution $g(t)$ to become extinct in finite time.

**Theorem 22** *Let $M^3$ be a homotopy 3-sphere equipped with a Riemannian metric $g = g(0)$. Under the Ricci flow, the width $W(g(t))$ satisfies*

$$\frac{\mathrm{d}}{\mathrm{d}t} W(g(t)) \le -4\pi + \frac{3}{4(t+C)} W(g(t)) \qquad [73]$$

*in the sense of the limsup of forward difference quotients. Hence, $g(t)$ must become extinct in finite time.*

The $4\pi$ in [73] comes from the Gauss–Bonnet theorem and the 3/4 comes from the bound on the minimum of the scalar curvature that the evolution equation implies. Both of these constants matter whereas the constant $C$ depends on the initial metric and the actual value is not important.

To see that [73] implies finite extinction time, rewrite [73] as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(W(g(t))(t+C)^{-3/4}\right) \le -4\pi(t+C)^{-3/4} \qquad [74]$$

and integrate to get

$$(T+C)^{-3/4} W(g(T)) \le C^{-3/4} W(g(0)) - 16\pi\left[(T+C)^{1/4} - C^{1/4}\right] \qquad [75]$$

Since $W \ge 0$ by definition and the right-hand side of [75] would become negative for $T$ sufficiently large, we get the claim.

As a corollary of this theorem we get finite extinction time for the Ricci flow.

**Corollary 3** *Let $M^3$ be a homotopy 3-sphere equipped with a Riemannian metric $g = g(0)$. Under the Ricci flow $g(t)$ must become extinct in finite time.*

## Acknowledgments

The authors were partially supported by NSF Grants DMS 0104453 and DMS 0405695.

*See also:* Black Hole Mechanics; Calibrated Geometry and Special Lagrangian Submanifolds; Geometric Analysis and General Relativity; Geometric Measure Theory; Leray–Schauder Theory and Mapping Degree; Ljusternik–Schnirelman Theory; Singularities of the Ricci Flow.

## Further Reading

Bray H (2002) Black holes, geometric flows, and the Penrose inequality in general relativity. *Notices of the American Mathematical Society* 49(11): 1372–1381.

Colding TH and De Lellis C (2003) *The Min–Max Construction of Minimal Surfaces*, *Surveys in Differential Geometry*, *vol. 8*, Lectures on Geometry and Topology held in honor of Calabi, Lawson, Siu, and Uhlenbeck at Harvard University, May 3–5,

2002, Sponsored by the Journal of Differential Geometry, pp. 75–107 (math.AP/0303305).
Colding TH and Minicozzi WP II (1999) *Minimal Surfaces*. Courant Lecture Notes in Math., vol. 4.
Colding TH and Minicozzi WP II (2003) Disks that are double spiral staircases. *Notices of the American Mathematical Society* 50(3): 327–339.
Colding TH and Minicozzi WP II (2005) Estimates for the extinction time for the Ricci flow on certain 3-manifolds and a question of Perelman. *Journal of the American Mathematical Society* 18(3): 561–569 (math.AP/0308090).
Colding TH and Minicozzi WP, II (2005) *Minimal submanifolds*. Preprint.
Lawson HB (1980) *Lectures on Minimal Submanifolds*, vol. I. Berkeley: Publish or Perish.
Meeks W III and Perez J (2004) Conformal properties in classical minimal surface theory. In: Grigor'yan A and Yau ST (eds.) *Eigenvalues of Laplacians and Other Geometric Operators*, Surveys in Differential Geometry IX, pp. 275–335. Somerville, MA: International Press.
Meeks WH and Yau ST (1980) Topology of three-dimensional manifolds and the embedding problems in minimal surface theory. *Annals of Mathematics* 112(3): 441–484.
Meeks W III and Yau ST (1982) The classical Plateau problem and the topology of three dimensional manifolds. *Topology* 21: 409–442.
Morgan F (1995) *Geometric Measure Theory. A Beginner's Guide*, 2nd edn. San Diego, CA: Academic Press.
Osserman R (1986) *A Survey of Minimal Surfaces*, 2nd edn. New York: Dover.
Perelman G Finite extinction time for the solutions to the Ricci flow on certain three-manifolds, math.DG/0307245.
Perez J (2005) Limits by rescalings of minimal surfaces: minimal laminations, curvature decay and local pictures. In: *Moduli Spaces of Properly Embedded Minimal Surfaces*, Notes for the Workshop. Palo Alto, CA: American Institute of Mathematics. (http://www.ugr.es/~jperez/papers/notes.pdf)
Rosenberg H (1992) Some recent developments in the theory of properly embedded minimal surfaces in $\boldsymbol{R}^3$, Seminare Bourbaki 1991/92, Asterisque No. 206, pp. 463–535.
Schoen R and Yau ST (1979) On the proof of the positive mass conjecture in general relativity. *Communications in Mathematical Physics* 65(1): 45–76.

# Minimax Principle in the Calculus of Variations

**A Abbondandolo**, Università di Pisa, Pisa, Italy



## Introduction

When studying a functional $f$ on an infinite-dimensional function space $X$, one is often interested in finding critical points which are not local minima. A simple yet powerful method to detect those critical points is the minimax method. The idea consists in detecting some complexity in the topology of $X$, or in the structure of the sublevels of $f$, to find a class $\Gamma$ of subsets of $X$ which somehow reveals such a topological complexity, and to show that the number

$$c := \inf_{\gamma \in \Gamma} \sup_{x \in \gamma} f(x)$$

is finite (even if the functional may be unbounded above and below). If the class $\Gamma$ is positively invariant under the action of the negative-gradient flow of $f$, and if a suitable compactness assumption known as the Palais–Smale condition holds, $c$ is proved to be a critical value of $f$. Quite remarkably, the minimax method also works when no topological complexity is present, but the negative-gradient flow of $f$ exhibits some kind of rigidity.

In this article we shall describe these ideas, starting from the simplest minimax result, the "mountain-pass theorem." We will show how to apply the minimax method by discussing the existence question of solutions of a nonlinear elliptic boundary value problem, of closed geodesics on compact manifolds, and of closed characteristics on compact energy hypersurfaces.

## The Mountain-Pass Theorem

Let us start by considering the following familiar fact. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth coercive function (i.e., its sublevels have compact closure). If a sublevel $\{f < a\}$ is not connected – say $\{f < a\} = A \cup B$, with $A, B$ disjoint open sets – then $f$ has a critical point $x$ at level

$$f(x) = c := \inf_{\gamma \in \Gamma} \max_{u \in \gamma} f(u) \geq a$$

where $\Gamma$ is the class of all continuous curves in $\mathbb{R}^n$ with one end point in $A$ and the other in $B$. More figuratively: if there are two valleys, then there must be a mountain pass. Let us examine a possible proof.

First notice that any curve in the class $\Gamma$ will have to cross the level $\{f = a\}$, so $c \geq a$. If by contradiction $c$ is not a critical value of $f$, by the compactness of the sublevels there is some $\epsilon > 0$ such that $|\nabla f| \geq \epsilon$ on $\{c - \epsilon \leq f \leq c + \epsilon\}$. Then the negative-gradient flow of $f$, that is, the solution of

$$\partial_t \phi(t, u) = -\nabla f(\phi(t, u)), \quad \phi(0, u) = u$$

pulls the sublevel $\{f \le c+\epsilon\}$ down into the sublevel $\{f \le c-\epsilon\}$ in finite time $2/\epsilon$. Indeed, if $\phi([0,t],u) \subset \{c-\epsilon \le f \le c+\epsilon\}$, then the inequalities

$$\begin{aligned} 2\epsilon &\ge f(u) - f(\phi(t,u)) \\ &= -\int_0^t \frac{\mathrm{d}}{\mathrm{d}s} f(\phi(s,u))\,\mathrm{d}s \\ &= \int_0^t |\nabla f(\phi(s,u))|^2\,\mathrm{d}s \ge \epsilon^2 t \end{aligned}$$

imply that $t \le 2/\epsilon$. By definition of $c$, we can find a continuous curve $\gamma \in \Gamma$ which is contained in $\{f \le c+\epsilon\}$. But then the curve $\gamma' := \phi(2/\epsilon, \gamma)$ still has one end point in $A$, the other one in $B$, and lies in $\{f \le c-\epsilon\}$, contradicting the definition of $c$.

If we try to generalize this result to functions defined on an infinite-dimensional real Hilbert space $H$, we encounter difficulties due to lack of compactness. Indeed, a continuous function on an infinite-dimensional Hilbert space can never have compact sublevels (with respect to the norm topology). If we look back at the proof, we see that we have used coercivity to guarantee that if the level set $\{f = c\}$ contains no critical points, then $\nabla f$ is bounded away from zero on the strip $\{c-\epsilon \le f \le c+\epsilon\}$, for some small $\epsilon > 0$. A natural idea is then to replace the coercivity assumption by a condition implying the latter fact.

**Definition** Let $f : H \to \mathbb{R}$ be a continuously differentiable function on a real Hilbert space $H$. A sequence $(u_h) \subset H$ is said a Palais–Smale sequence if $f(u_h)$ is bounded and $Df(u_h)$ tends to zero. The function $f$ is said to satisfy the Palais–Smale condition if every Palais–Smale sequence has a converging subsequence.

The Palais–Smale condition readily implies the statement above. Assuming also that $f$ is twice continuously differentiable, the negative-gradient flow of $f$ (a well-defined local flow because $\nabla f$ is continuously differentiable) pulls the sublevel $\{f \le c+\epsilon\}$ down into $\{f \le c-\epsilon\}$ in finite time. These observations lead to the following:

**Theorem** (Mountain pass). *Let $f$ be a twice continuously differentiable function on a real Hilbert space $H$, satisfying the Palais–Smale condition. Assume that a sublevel $\{f < a\}$ is not connected, and let $A$, $B$ be two disjoint open sets such that $A \cup B = \{f < a\}$. Then $f$ has a critical point $x$ at level*

$$f(x) = c := \inf_{\gamma\in\Gamma} \max_{u\in\gamma} f(u) \ge a$$

*where $\Gamma$ is the class of all continuous curves in $H$ with one end point in $A$ and the other one in $B$.*

If we are even more ambitious, and we wish to consider functions defined on a real Banach space $E$, we also encounter the problem of not having a gradient vector field. Indeed, the differential of $f$ at $x$, $Df(x)$, is an element of the dual space $E^*$, but in this case we have no inner product on $E$ by which we can represent $Df(x)$ as the product by some vector of $E$. This problem can be overcome by the notion of a pseudogradient vector field. In fact, it can be proved that if $f$ is continuously differentiable on $E$, then there exists a locally Lipschitz vector field $V$ defined on the complement of the critical points of $f$, such that

$$\begin{aligned} &\|V(u)\| < \min\{\|Df(u)\|, 1\} \\ &Df(u)[V(u)] > \tfrac{1}{2}\min\{\|Df(u)\|, 1\}\|Df(u)\| \end{aligned}$$

In other words, even if there is no direction of steepest increase for $f$, we do have directions along which the increase of $f$ is steep enough, and these directions can be selected in a locally Lipschitz way. Notice that pseudogradients are useful also in the case of a continuously differentiable function on a Hilbert space: in this case the gradient of $f$ is just continuous, so it does not generate a flow. The Palais–Smale condition, as stated above, makes perfect sense on the Banach space $E$ (with the only difference that now $Df(u_h)$ tends to zero in the dual norm of $E^*$), and the mountain-pass theorem holds for functions of class $C^1$ on a Banach space.

Actually, the fact that the domain of $f$ has a vector structure is not relevant in this statement, and the mountain-pass theorem holds also for functions defined on connected infinite-dimensional manifolds. Since the essential feature is to dispose of a pseudogradient vector field, the right level of generality is to consider a Banach manifold $M$ (i.e., a manifold modeled on a Banach space) endowed with a complete Finsler structure (i.e., a Banach norm on each tangent space of $M$, varying in a suitably regular way, inducing a complete distance on $M$).

## A Nonlinear Elliptic Boundary-Value Problem

Let us consider a typical application of the mountain-pass theorem to a semilinear elliptic boundary-value problem. Let $\Omega$ be a smooth bounded domain in $\mathbb{R}^n$, and for $\lambda \in \mathbb{R}$, $p > 2$, consider the problem

$$\begin{aligned} -\Delta u &= \lambda u + u|u|^{p-2} \quad &\text{in } \Omega \\ u &= 0 \quad &\text{on } \partial\Omega \end{aligned} \qquad [1]$$

Let $0 < \lambda_1 < \lambda_2 \le \lambda_3 \le \cdots$ be the eigenvalues of the Laplace operator $-\Delta$, with domain $H^2 \cap H_0^1(\Omega)$, the Sobolev space of $L^2$-functions on $\Omega$ with weak first

two derivatives in $L^2$, vanishing on $\partial\Omega$. We claim that, if $n=2$, or if $n \geq 3$ and $2 < p < 2^* := 2n/(n-2)$, then problem [1] with $\lambda < \lambda_1$ has a nontrivial solution.

By elliptic regularity, the solutions of [1] are precisely the critical points of the functional

$$\mathcal{E}(u) = \frac{1}{2}\int_\Omega \left(|\nabla u(x)|^2 - \lambda u(x)^2\right) \mathrm{d}x - \frac{1}{p}\int_\Omega |u(x)|^p \, \mathrm{d}x$$

We recall that $H_0^1(\Omega)$ continuously embeds into $L^p(\Omega)$, for every $p < +\infty$ if $n=2$, for every $p \leq 2^*$ if $n \geq 3$. So the functional $\mathcal{E}$ is well defined, and actually continuously differentiable, on $H_0^1(\Omega)$, a Hilbert space with the inner product

$$\langle u, v\rangle_{H_0^1(\Omega)} = \int_\Omega \nabla u(x) \cdot \nabla v(x) \, \mathrm{d}x$$

Since $p > 2$, near zero the quadratic part of the functional $\mathcal{E}$ dominates over the part with the $L^p$-norm. By the Rayleigh characterization of the first eigenvalue of the Laplacian,

$$\lambda_1 = \min_{u \in H_0^1(\Omega)\setminus\{0\}} \frac{\int_\Omega |\nabla u(x)|^2 \, \mathrm{d}x}{\int_\Omega u(x)^2 \, \mathrm{d}x}$$

the assumption $\lambda < \lambda_1$ implies that the quadratic part of $\mathcal{E}$ is positive definite. So we can find a small $\rho > 0$ such that

$$a := \inf_{\|u\|_{H_0^1(\Omega)} = \rho} \mathcal{E}(u) > 0$$

On the other hand, the fact that $p > 2$ implies that

$$\lim_{\mu \to +\infty} \mathcal{E}(\mu u) = -\infty$$

for every $u \neq 0$. Therefore, the sublevel $\{\mathcal{E} < a\}$ is not connected, and if we can prove the Palais–Smale condition, the mountain-pass theorem will imply the existence of a critical point $u$ with $\mathcal{E}(u) \geq a > 0$, i.e., a nontrivial solution of [1].

In order to prove the Palais–Smale condition, notice that the expression for the differential of $\mathcal{E}$,

$$D\mathcal{E}(u)[v] = \int_\Omega \nabla u(x) \cdot \nabla v(x) \, \mathrm{d}x - \int_\Omega \left(\lambda u(x) + |u(x)|^{p-2} u(x)\right) v(x) \, \mathrm{d}x$$

and the compactness of the embedding of $H_0^1(\Omega)$ into $L^p(\Omega)$ for $p < 2^*$ imply that the gradient of $\mathcal{E}$ has the form

$$\nabla\mathcal{E}(u) = u + K(u) \qquad [2]$$

where $K : H_0^1(\Omega) \to H_0^1(\Omega)$ is a compact map, that is, it maps bounded sets into precompact ones. It is readily seen that when $\nabla\mathcal{E}$ has such a form, bounded Palais–Smale sequences are compact. Thus, it is enough to show that every Palais–Smale sequence is bounded. But this follows from the identity

$$p\mathcal{E}(u) - D\mathcal{E}(u)[u] = \left(\frac{p}{2} - 1\right) \int_\Omega \left(|\nabla u(x)|^2 - \lambda u(x)^2\right) \mathrm{d}x$$

together with the fact that the right-hand side term defines an equivalent norm on $H_0^1(\Omega)$, because $p > 2$ and $\lambda < \lambda_1$. This concludes the proof.

Actually, using the maximum principle one could show that under the same assumptions, problem [1] has a solution which is positive in $\Omega$.

When $n \geq 3$ and $p = 2^* = 2n/(n-2)$, the functional $f$ still exhibits a mountain-pass geometry, but the Palais–Smale condition fails. In fact, the embedding of $H_0^1(\Omega)$ into $L^{2^*}(\Omega)$ is not compact, so the map $K$ appearing in [2] is not compact, and bounded Palais–Smale sequences need not have a converging subsequence. We recall that the noncompactness of the embedding of $H_0^1(\Omega)$ into $L^{2^*}(\Omega)$ is due to the fact that the quotient

$$S(u) = \frac{\int_\Omega |\nabla u(x)|^2 \, \mathrm{d}x}{\left(\int_\Omega |u(x)|^{2^*} \mathrm{d}x\right)^{2/2^*}}$$

is invariant under rescaling $u \mapsto u_\mu(x) = u(\mu x)$.

When $\lambda = 0$, the Pohožaev identity – an integral formula obtained by multiplying the equation by $x \cdot \nabla u(x)$ – can be used to prove that problem [1] has no nontrivial solutions, when $\Omega$ is a star-shaped domain other than the whole $\mathbb{R}^n$.

When $\lambda \neq 0$, the presence in the functional of an $L^2$-norm – which rescales differently – breaks the symmetry, and the existence of nontrivial solutions is again possible. Indeed, Brezis and Nirenberg have shown that problem [1] with $p = 2^*$ has a nontrivial solution provided that $n \geq 4$ and $0 < \lambda < \lambda_1$, or $n = 3$ and $\lambda^* < \lambda < \lambda_1$, for some $\lambda^* \in [0, \lambda_1]$ depending on the domain $\Omega$.

The proof is based on the fact that there is a certain threshold $s > 0$, related to the best Sobolev constant obtained by taking the infimum of $S(u)$ over all $u \in H_0^1$ (the domain is irrelevant here), below which the Palais–Smale condition holds. That is, every sequence $(u_h)$ such that $\mathcal{E}(u_h)$ converges to some $b$ less than $s$, and $D\mathcal{E}(u_h)$ tends to zero, is compact. The proof of the mountain-pass theorem shows that the Palais–Smale condition is needed only at the minimax level $c$. In order to conclude, it is then enough to show that $c < s$. The value of $c$ can be estimated by using the fact that the

infimum of the quotient $S$ over functions on the whole $\mathbb{R}^n$ is attained at the family of functions

$$u^*(x) = \left( \frac{\epsilon^2 n(n-2)}{(\epsilon^2 + |x|^2)^2} \right)^{(n-2)/4}$$

which are then solutions of [1] with $p = 2^*, \lambda = 0$, and $\Omega = \mathbb{R}^n$.

Another way to break the symmetry is to keep $\lambda = 0$ but to consider domains with a rich topology. For instance, Bahri and Coron have shown that if $\Omega$ is a domain with some nonzero singular homology group $H_k(\Omega; \mathbb{Z}_2), k \geq 1$, then problem [1] with $p = 2^*$ and $\lambda = 0$ has a positive solution.

Elliptic equations having nonlinearities with the critical exponent $2^*$ arise naturally in some geometric problems. Consider a manifold $M$ of dimension $n \geq 3$, with a metric $g$ having scalar curvature $k$. The Yamabe problem calls for finding a metric $g_0$, conformally equivalent to $g$, having constant scalar curvature. If $g_0 = u^{4/(n-2)}g$, where the positive function $u$ gives the conformal factor, one finds that $u$ must solve the equation

$$-\frac{4(n-1)}{n-2} \Delta_g u = -ku + k_0 u|u|^{2^*-2}$$

where $\Delta_g$ is the Laplace–Beltrami operator associated with the metric $g$, and the constant $k_0$ is the scalar curvature of $g_0$. Again, the corresponding functional satisfies the Palais–Smale condition only below a certain threshold (actually, the same number $s$ as seen earlier; this because the lack of compactness is due to local concentration phenomena, and the metric structure of the whole ambient becomes irrelevant). The task is then to show that the minimax level is below that threshold or, equivalently, that a certain best Sobolev constant for $(M, g)$ is less than the corresponding constant for $\mathbb{R}^n$ with the flat metric (the latter constant is again the infimum of $S(u)$). This fact was proved by Aubin in the case $n \geq 6$ or $(M, g)$ not locally conformally flat. Schoen has then treated the remaining case, by means of the positive-mass theorem, a deep result in differential geometry.

## A General Minimax Principle

Let us consider again a twice continuously differentiable function $f$ on a real Hilbert space $H$. The vector field

$$V(u) = \frac{\nabla f(u)}{\sqrt{1 + \|\nabla f(u)\|^2}}$$

has the same nice properties of the gradient vector field of $f$, but in addition it is bounded. The advantage is that the flow of $-V$ is globally defined. When talking about the negative-gradient flow of $f$, we will actually refer to such a flow. It will also be useful to dispose of a negative-gradient flow truncated below level b. This is the flow of the vector field $-V_b$, where

$$V_b(u) = \varphi(f(u))V(u)$$

with $\varphi$ a smooth function on $\mathbb{R}$ which is identically zero on $[-\infty, b]$, then increases up to reaching the value 1, and afterwards remains constantly equal to 1. This truncated negative-gradient flow keeps the points in the sublevel $\{f \leq b\}$ fixed, and behaves as the negative-gradient flow above $b$ (except the fact that trajectories slow down as the value of $f$ approaches $b$).

After these preliminaries, let us consider again the characterization of the critical level $c$ appearing in the mountain-pass theorem. This critical level was obtained as the infimum over a certain class $\Gamma$ of sets $\gamma$ – the curves with end points in different components of $\{f < a\}$ – of the maximum of $f$ over $\gamma$. But if we look back at the proof, we realize that the fact that these sets were curves was not essential. The important feature was that the negative-gradient flow $\phi(t, \cdot)$ mapped a set of the class $\Gamma$ into a set still belonging to the class $\Gamma$, for $t \geq 0$. This observation leads to the following general minimax theorem, due to Palais:

**Theorem** (General minimax). *Let $f$ be a twice continuously differentiable function on a real Hilbert space $H$, satisfying the Palais–Smale condition. Let $\Gamma$ be a class of subsets of $H$ which is positively invariant under the action of the negative-gradient flow $\phi$ of $f$ (possibly truncated below level $b$): that is, if the set $\gamma$ belongs to $\Gamma$, then the set $\phi(t, \gamma)$ belongs to $\Gamma$ for all $t \geq 0$. Then, if the number*

$$c := \inf_{\gamma \in \Gamma} \sup_{u \in \gamma} f(u)$$

*is finite (and larger than $b$), then $c$ is a critical value of $f$.*

The proof goes along the same lines of the proof of the mountain-pass theorem: if $c$ is not a critical value of $f$, the (possibly truncated) negative-gradient flow $\phi(t_0, \cdot)$ pulls a sublevel $\{f \leq c + \epsilon\}$ down into the sublevel $\{f \leq c - \epsilon\}$ (with $c - \epsilon > b$), for some large $t_0$, by the Palais–Smale condition. Then we achieve a contradiction choosing a set $\gamma \in \Gamma$ on which $f$ does not exceed $c + \epsilon$, and noticing that $\phi(t_0, \gamma)$ is a set which still belongs to the class $\Gamma$, by positive invariance, and on which $f$ does not exceed $c - \epsilon$.

As we shall see in the last section, the possibility of working with a truncated negative-gradient flow

(assuming in this case that $c > b$) makes the application of this theorem easier. Again, an analogous result holds for continuously differentiable functions on Banach spaces, or more generally on Banach manifolds with a complete Finsler structure.

Trivial classes $\Gamma$ are the class of all points in $H$, and the class consisting of the single set $H$, yielding to the infimum and the supremum of $f$, respectively. More interesting classes are constructed by fixing a topological space $X$ and considering the images of all continuous maps $h: X \to H$ belonging to a certain relative homotopy class.

## Closed Geodesics on Compact Manifolds

A typical application of the general minimax theorem is Birkhoff proof of the existence of a closed geodesic on the sphere $S^2$, endowed with an arbitrary metric $g$. Closed geodesics are precisely the critical points of the energy functional

$$\mathcal{S}(x) = \frac{1}{2}\int_0^1 g(\dot{x}(t), \dot{x}(t))\,\mathrm{d}t$$

on the Hilbert manifold $H^1(\mathbb{T}, S^2)$ consisting of all one-periodic loops on $S^2$ of Sobolev regularity $H^1$ (here $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ denotes the circle parametrized by $[0,1]$). This functional satisfies the Palais–Smale condition and it is bounded below, but its minima are just the trivial constant loops, on which $\mathcal{S} = 0$.

Let us use angle coordinates $(\theta, \varphi)$ on $S^2$, $-\pi/2 \leq \theta \leq \pi/2, 0 \leq \varphi \leq 2\pi$ ($\theta$ is the latitude, $\varphi$ the longitude). A (suitably regular) map $h: S^2 \to S^2$ induces a curve in $H^1(\mathbb{T}, S^2)$ parametrized by $\theta$: the value of this curve at $\theta \in [-\pi/2, \pi/2]$ is the loop $t \mapsto h(\theta, 2\pi t)$. It is a curve that joins two constant loops. Let $\Gamma$ be the set of curves in $H^1(\mathbb{T}, S^2)$ which are obtained by maps $h: S^2 \to S^2$ of topological degree 1. This class is clearly positively invariant under the action of the negative-gradient flow of $\mathcal{S}$ (as of every homotopy fixing the constant loops).

If we can show that the minimax level

$$c := \inf_{\gamma \in \Gamma} \sup_{u \in \gamma} \mathcal{S}(x)$$

is positive, we will get a positive critical value of $\mathcal{S}$ by the general minimax theorem, hence a nontrivial closed geodesic. By considering the fact that loops with small energy also have a small diameter, it is easy to construct a homotopy on $\{\mathcal{S} < a\}$, for some small $a > 0$, which shrinks every loop to a point. If $h: S^2 \to S^2$ determines a curve $\gamma$ with $\max_{x \in \gamma} \mathcal{S}(x) < a$, composition with this homotopy yields to a homotopy of $h$ to a map whose image is a curve in $S^2$. A further homotopy then shows that the map $h$ is homotopic to a constant, which is impossible if $h$ has degree 1. This shows that $c \geq a > 0$, concluding the proof.

Actually, Ljusternik and Fet have proved that every compact manifold $M$ has a nontrivial closed geodesic. Indeed, if $M$ has nonzero fundamental group, it is enough to minimize $\mathcal{S}$ on some nontrivial homotopy class of loops. Otherwise, the fact that $M$ is a compact manifold implies that some homotopy group $\pi_{k+1}(M), 1 \leq k < \dim M$, does not vanish. A construction similar to the one described above then allows to associate with every noncontractible map $h: S^{k+1} \to M$ a map $u: (B^k, \partial B^k) \to (H^1(\mathbb{T}, M), \{\mathcal{S} = 0\})$ which is not homotopically trivial (here $B^k$ denotes the closed unit ball in $\mathbb{R}^k$, and the notation means that $u$ maps the boundary of the ball $B^k$ into the set of constant loops). Taking a minimax over the set of images of the maps $u$ associated with every noncontractible map $h: S^{k+1} \to M$ yields to the desired critical point of $\mathcal{S}$ with positive energy.

It is conjectured that every compact manifold has infinitely many closed geodesics. Morse theory allows to prove this fact for the vast majority of manifolds, but not for the spheres. Bangert and Franks have established the existence of infinitely many geodesics on $S^2$ by proving that every area-preserving homeomorphism of the open disk with two fixed points must have infinitely many periodic points. Proving the existence of infinitely many closed geodesics on higher-dimensional spheres is a challenging open problem.

## A Rigidity Property of a Certain Class of Maps

It is important that the class $\Gamma$ in the general minimax theorem is only required to be invariant under the action of the negative-gradient flow, and not, say, under the action of any continuous homotopy on which the function $f$ is nonincreasing. Indeed, too many undesirable things can be done on an infinite-dimensional Hilbert space by arbitrary continuous maps, whereas the maps arising from our negative-gradient flow might show some rigidity, forcing them to behave as maps on finite-dimensional spaces.

Let us clarify this point by considering the following example, due to Benci and Rabinowitz. It may sound a bit artificial at this moment (simpler examples could be built), but we will find it useful in the next section. Assume that our Hilbert space is

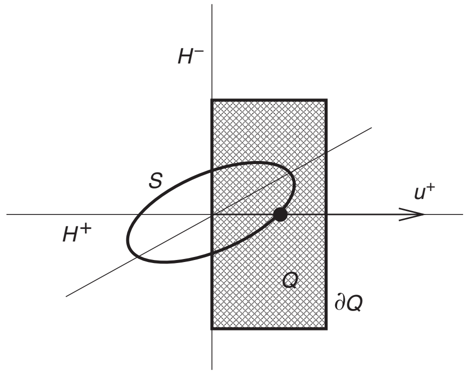



**Figure 1** The sets $S, Q, \partial Q$.

endowed with an orthogonal splitting $H = H^- \oplus H^+$, fix a unit vector $u^+$ in $H^+$, and consider the sets

$$S = \{u \in H^+ | \ \|u\| = \rho\}$$
$$Q = \{u + \lambda u^+ \,|\, u \in H^-, \ \|u\| \leq \sigma, \ 0 \leq \lambda \leq \tau\}$$
$$\partial Q = \{u + \lambda u^+ \in Q \,|\, \lambda \in \{0, \tau\} \text{ or } \|u\| = \sigma\}$$

for some positive numbers $\rho, \sigma, \tau$ such that $\tau > \rho$. The latter inequality implies that the intersection $Q \cap S$ is not empty (see Figure 1).

If the linear subspace $H^-$ is finite dimensional, a simple argument involving the topological degree shows the following fact: the image of any continuous map $h : Q \to H$ which is the identity on $\partial Q$ has nonempty intersection with $S$.

When $H^-$ is infinite dimensional, this fact is not true anymore. Indeed, it is not difficult to see that the set $Q$ is homeomorphic to an infinite-dimensional closed ball $B$, by a homeomorphism $\psi$ mapping $\partial Q$ onto the infinite-dimensional sphere $\partial B$. If $B$ is the closed ball of an infinite-dimensional Hilbert space, for instance, the space $\ell_2$ of all square-summable sequences $(x_h)$ endowed with the norm $|x|_2 = (\sum_{h=0}^{\infty} |x_h|^2)^{1/2}$, the continuous map

$$g(x_0, x_1, x_2, \ldots) = \left(\sqrt{1 - |x|_2^2}, x_0, x_1, x_2, \ldots\right)$$

maps $B$ into $\partial B$ and is a shift operator on $\partial B$. In particular, it is a continuous map on $B$ without fixed points, and it can be used to define a map $h : B \to \partial B$ which is the identity on $\partial B$, by setting

$$h(x) = \mu(x)x + (1 - \mu(x))g(x)$$
$$\text{with } \mu(x) \geq 1 \text{ such that } |h(x)|_2 = 1$$

Conjugation by the homeomorphism $\psi$ produces a continuous map from $Q$ to $\partial Q$, which is the identity on $\partial Q$, providing us with the desired counterexample.

In other terms, when $H^-$ is infinite dimensional, the sets $\partial Q$ and $S$ can be unlinked by means of a continuous map. The situation changes if we restrict the class of maps $h : Q \to H$ to those of the form

$$h(u) = u + K(u) \qquad [3]$$

where $K$ is a continuous compact map. In this case, indeed, the argument for a finite-dimensional $H^-$ can be applied, by replacing the topological degree by the Leray–Schauder degree (which is invariant precisely with respect to homotopies of the form above), and one proves that $\partial Q$ and $S$ cannot be unlinked by means of continuous maps of this form.

## Closed Characteristics on Compact Energy Hypersurfaces

Consider $\mathbb{R}^{2n}$ with coordinates $(p_1, \ldots, p_n, q_1, \ldots, q_n)$, endowed with the standard symplectic form

$$\omega := \mathrm{d}p \wedge \mathrm{d}q = \sum_{j=1}^{n} \mathrm{d}p_j \wedge \mathrm{d}q_j$$

Let $\Sigma$ be a compact connected hypersurface in $\mathbb{R}^{2n}$. The restriction of $\omega$ to the tangent space $T_x\Sigma$ has a one-dimensional kernel, which varies smoothly with $x$. In other words, there is a smooth line bundle

$$\mathcal{L}_\Sigma := \{(x, u) \in T\Sigma \,|\, \omega(u, v) = 0 \ \forall v \in T_x\Sigma\}$$

over $\Sigma$. We wish to discuss the classical problem of finding a closed characteristic for $\mathcal{L}_\Sigma$, that is, a closed curve everywhere tangent to $\mathcal{L}_\Sigma$.

This geometric problem has a dynamical interpretation. Indeed, let $H$ be a smooth real function on $\mathbb{R}^{2n}$ such that $\Sigma$ is the inverse image of the regular value 1. The function $H$ – the Hamiltonian – generates a vector field $X_H$ on $\mathbb{R}^{2n}$ by the formula

$$\omega(X_H(x), u) = -DH(x)[u], \quad \forall u \in \mathbb{R}^{2n}$$

or, equivalently,

$$X_H(x) = J\nabla H(x), \quad \text{with } J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$$

The Hamiltonian vector field $X_H$ is tangent to $\Sigma$ and belongs to $\mathcal{L}_\Sigma$. Therefore, the hypersurface $\Sigma$ is invariant for the flow of $X_H$, and the flow orbits are precisely the characteristics. So finding a closed characteristic on $\Sigma$ is equivalent to finding a periodic orbit of $X_H$ with energy $H = 1$.

Up to changing the Hamiltonian, we may assume that all the values in an interval $]1 - \delta_0, 1 + \delta_0[$ are regular for $H$, and that the corresponding level sets $\Sigma_\eta := \{H = \eta\}$ are all connected (hence diffeomorphic to $\Sigma = \Sigma_1$). We would like to sketch Hofer and Zehnder's proof of the fact that there is a dense set of values $\eta \in ]1 - \delta_0, 1 + \delta_0[$ for which $\Sigma_\eta$ admits a closed characteristic.

This proof is based on the fact that the one-periodic orbits of $X_H$ are critical points of the action functional

$$\mathcal{A}_H(x) = \int_{\mathbb{T}} x^*(p\,\mathrm{d}q - H\,\mathrm{d}t) \\ = \frac{1}{2}\int_0^1 \dot{x}(t)\cdot Jx(t)\,\mathrm{d}t - \int_0^1 H(x(t))\,\mathrm{d}t$$

on the space of loops $x:\mathbb{T}\to\mathbb{R}^{2n}$.

Clearly, it is enough to show that for every $\delta > 0$ there is a closed characteristic on some $\Sigma_\eta$ with $|\eta - 1| < \delta$. We can take advantage of the fact that we are free to change the Hamiltonian, as long as it has the level sets $\Sigma_\eta, |\eta-1|<\delta$. Denoting by $B$ the bounded component of the complement of $\{1-\delta \le H \le 1+\delta\}$, we may assume that $B$ contains the origin. We can modify $H$ in such a way that $H$ vanishes identically on $B$, then it grows, parametrizing all the hypersurfaces $\Sigma_\eta, |\eta-1|<\delta$, in a strictly increasing way, then it remains constant in a large ball, and finally it smoothly switches to the quadratic form $(3/2)\pi|x|^2$. By choosing $H$ in this way, one can ensure that all the constant orbits and all the one-periodic orbits which do not lie on $\Sigma_\eta$ for some $|\eta-1|<\delta$ have non-positive action. So it is enough to prove that the functional $\mathcal{A}_H$ has a positive critical value.

Using the Fourier series decomposition

$$x(t) = \sum_{k\in\mathbb{Z}} \mathrm{e}^{2\pi k t J}\hat{x}_k, \quad \hat{x}_k \in \mathbb{R}^{2n}$$

one sees that the quadratic part of the action functional has the form

$$\int_0^1 \dot{x}(t)\cdot Jx(t)\,\mathrm{d}t = 2\pi\sum_{k\in\mathbb{Z}} k|\hat{x}_k|^2 \qquad [4]$$

so it is positive on an infinite-dimensional linear space, negative on an infinite-dimensional linear space, and null on the $2n$-dimensional space spanned by the constant loops. The specific form of [4] suggests to choose as domain of the action functional the Sobolev space $H^{1/2}(\mathbb{T},\mathbb{R}^{2n})$, the space of square-integrable one-periodic curves $x$ in $\mathbb{R}^{2n}$ with

$$\|x\|^2_{H^{1/2}} := |\hat{x}_0|^2 + 2\pi\sum_{k\in\mathbb{Z}}|k||\hat{x}_k|^2 < +\infty$$

This is indeed a Hilbert norm on $H^{1/2}(\mathbb{T},\mathbb{R}^{2n})$. The functional $\mathcal{A}_H$ is smooth on this space, and its gradient takes the form

$$\nabla\mathcal{A}_H(x) = Lx + K(x) \qquad [5]$$

where $L$ is the self-adjoint Fredholm operator representing the quadratic form [4] with respect to the $H^{1/2}$-Hilbert product, and $K$ is a compact map. A gradient of the form [5] again implies that bounded Palais–Smale sequences are compact. The Palais–Smale condition then follows from the fact that the Hamiltonian $H$ is quadratic outside a large ball, and has no one-periodic orbits there (the large orbits are all periodic, but their period is 2/3).

Consider the splitting $H^{1/2}(\mathbb{T},\mathbb{R}^{2n}) = H^- \oplus H^+$, with

$$H^- = \{x \mid \hat{x}_k = 0 \text{ for } k > 0\} \\ H^+ = \{x \mid \hat{x}_k = 0 \text{ for } k \le 0\}$$

Let $S$, $Q$, and $\partial Q$ be the sets defined in the previous section, with

$$u^+(t) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{2\pi t J}u_0, \quad u_0\in\mathbb{R}^{2n},\ |u_0| = 1$$

and constants $\rho,\sigma,\tau$ to be determined. Since the quadratic form [4] is positive on $H^+$ and the Hamiltonian $H$ vanishes near the origin, we can find a small $\rho > 0$ such that

$$\inf_{x\in S}\mathcal{A}_H(x) > 0$$

The fact that the quadratic form [4] is seminegative on $H^-$ and the behavior of $H(x)$ for large $|x|$ imply that if $\sigma$ and $\tau$ are suitably large (in particular $\tau > \rho$), then

$$\sup_{x\in\partial Q}\mathcal{A}_H(x) \le 0$$

Let $\Gamma$ be the set of all images of maps

$$h: Q \to H^{1/2}(\mathbb{T},\mathbb{R}^{2n})$$

which are the identity on $\partial Q$ and are of the form

$$h(x) = \mathrm{e}^{\alpha(x)L}(x + K(x)) \qquad [6]$$

with $\alpha$ a continuous real-valued function, and $K$ a continuous compact map. This class of maps is more general than the one considered in the previous section, but the fact that $\mathrm{e}^{\alpha L}$ commutes with the projections onto $H^-$ and $H^+$ ensures that $\partial Q$ and $S$ cannot be unlinked even inside this class. Therefore, any $\gamma\in\Gamma$ has nonempty intersection with $S$, so

$$c := \inf_{\gamma\in\Gamma}\sup_{x\in\gamma}\mathcal{A}_H(x) \ge \inf_{x\in S}\mathcal{A}_H(x) > 0$$

We would like to apply the general minimax theorem, and conclude that $c$ is the desired positive critical value.

The number $c$ being clearly finite, it is enough to show that $\Gamma$ is positively invariant under the action of the negative-gradient flow $\phi$ of $\mathcal{A}_H$, truncated below level 0. Let $\gamma = h(Q)\in\Gamma$ and $t\ge 0$. Then $\phi(t,\gamma)$ is the image of $Q$ by the map $\phi(t,h(\cdot))$. This

map is the identity on $\partial Q$ because $\partial Q$ lies in $\{\mathcal{A}_H \leq 0\}$ and $\phi$ is truncated below level 0. It is of the form [6] because by [5] the truncated negative-gradient flow of $\mathcal{A}_H$ has the form

$$\phi(t,x) = \mathrm{e}^{-\theta(t,x)L}(x + K(t,x))$$

for some continuous function $0 \leq \theta(t,x) \leq t$ and for some continuous compact map $K$. This concludes the proof.

This result was refined by Struwe, who proved the existence of a closed characteristic on $\Sigma_\eta$ for almost every $\eta$, in the sense of the Lebesgue measure. We could try to use the abundance of closed characteristics on energy levels near $\Sigma$ to get the existence of one on $\Sigma$ by taking a limit. But this process produces a closed characteristic on $\Sigma$ only if we can bound the periods of the approximating closed orbits, otherwise a more general invariant set results. Actually, Ginzburg, Herman, and Gürel have produced examples of compact hypersurfaces without any closed characteristic.

As conjectured by Weinstein and proved by Viterbo, closed characteristics always exist on contact-type compact hypersurfaces (i.e., hypersurfaces $\Sigma$ on which the restriction of $\omega$ is the differential of a 1-form $\lambda$ such that $\lambda \wedge \mathrm{d}\lambda \wedge \cdots \wedge \mathrm{d}\lambda$ is a volume form). In this case, one should even expect a multiplicity result. For hypersurfaces which bound a strictly convex set in $\mathbb{R}^{2n}$, for instance, the existence of $n$ closed characteristics is conjectured. The best result so far is due to Long, who could prove the existence of $[n/2]+1$ of them. Hofer, Wysocki, and Zehnder have proved that, when $n=2$, there are either two or infinitely many closed characteristics (for a generic contact-type hypersurface diffeomorphic to $S^3$), by using the already mentioned theorem by Franks on periodic points of area-preserving homeomorphisms of the disk. Proving an analogous result for $n \geq 3$ is an intriguing open problem.

*See also:* Contact Manifolds; Floer Homology; Hamilton–Jacobi Equations and Dynamical Systems: Variational Aspects; Image Processing: Mathematics; Inequalities in Sobolev Spaces; Leray–Schauder Theory and Mapping Degree; Ljusternik–Schnirelman Theory; Saddle Point Problems.

## Further Reading


Abbondandolo A (2001) *Morse Theory for Hamiltonian Systems*. Pitman Research Notes in Mathematics, vol. 425. London: Chapman and Hall.

Ambrosetti A and Malchiodi A (2005) *Perturbation Methods and Semilinear Elliptic Problems* $\boldsymbol{R}^n$. Birkhäuser.

Aubin T (1998) *Some Nonlinear Problems in Riemannian Geometry*. New York: Springer.

Chang KC (1993) *Infinite-dimensional Morse Theory and Multiple Solution Problems*. Boston: Birkhäuser.

Chang KC (2005) *Methods in Nonlinear Analysis*. Springer.

Ghoussoub N (1993) *Duality and Perturbation Methods in Critical Point Theory*. Cambridge: Cambridge University Press.

Hofer H and Zehnder E (1994) *Symplectic Invariants and Hamiltonian Dynamics*. Basel: Birkhäuser.

Klingenberg W (1982) *Riemannian Geometry*. Berlin: Walter de Gruyter & Co.

Mawhin J and Willem M (1989) *Critical Point Theory and Hamiltonian Systems*. New York: Springer.

Rabinowitz PH (1986) *Minimax Methods in Critical Point Theory with Applications to Differential Equations*. Regional Conference Series in Mathematics. Providence, RI: American Mathematical Society.

Schechter M (1999) *Linking Methods in Critical Point Theory*. Boston: Birkhäuser.

Struwe M (2000) *Variational Methods*, 3rd edn. Berlin: Springer-Verlag.

Willem M (1996) *Minimax Theorems*. Boston: Birkhäuser.


# Mirror Symmetry: A Geometric Survey

**R P Thomas**, Imperial College, London, UK



## Introduction

Mirror symmetry was discovered in the late 1980s by physicists studying superconformal field theories (SCFTs). One way to produce SCFTs is from closed string theory; in the Riemannian (rather than Lorentzian) theory the string's world line gives a map of a Riemannian 2-manifold into the target with an action which is conformally invariant, so the 2-manifold can be thought of as a Riemann surface with a complex structure. Making sense of the infinities in the quantum theory (supersymmetry and anomaly cancelation) forces the target to be 10-dimensional – Minkowski space times by a 6-manifold $X$ – and $X$ to be (to first order) Ricci flat and so to have holonomy in SU(3). That is $X$ is a Calabi–Yau 3-fold $(X,\Omega,\omega)$. So SCFTs come from $\sigma$-models (mapping Riemann surfaces into Calabi–Yau 3-folds) but, it turns out, in two different ways – the A-model and the B-model. Deformations of the SCFT and either $\sigma$-model are isomorphic, so over an open set the two coincide. Thus, it was natural to conjecture that almost all of the relevant SCFTs came from geometry – from an A or B $\sigma$-model. In particular,

the A-model of a Calabi–Yau $X$ should, therefore, give the same SCFT as the B-model on another Calabi–Yau $\check{X}$. It turns out then that the A-model on $\check{X}$ should also be isomorphic to the B-model on $X$; thus, mirror symmetry should give an involution on a Calabi–Yau 3-folds. (The full picture is slightly more complicated – it involves large complex structure limits, multiple mirrors and flops.) By studying the SCFTs, Greene and Plesser predicted the mirror of the simplest Calabi–Yau 3-fold, the quintic in $\mathbb{P}^4$, and mirror symmetry was born.

Topological observables, that is, certain path integrals over the space of all maps, can be calculated by the semiclassical approximation as integrals over the space of classical minima – (anti) holomorphic curves in the Calabi–Yau (these minimize volume in a fixed homology class). From the zero homology class we get the constant maps – points in $X$ – and so integrals over $X$. In some cases, by Poincaré duality, these can be thought of as intersections of cycles; we think of the string world sheet lying at a point of intersection. When the world sheet has a nontrivial homology class, it allows more general "intersections" where the cycles need not intersect but are connected by a holomorphic curve, giving a perturbation of the usual intersection product on cohomology called quantum cohomology. Namely, there is a contribution $(a.\beta)(b.\beta)(c.\beta)\mathrm{e}^{\int_\beta \omega}$ to the quantum triple product $a.b.c$ of three 4-cycles $a, b, c \in H^{1,1} \cong H^2 \cong H_4$ from each holomorphic curve $\beta$ (of genus 0, in the 0-loop approximation to the physics) in $X$ of area $\int_\beta \omega$ (where $\omega$ is the Kähler form). The A-model correlation functions can be determined from these data; the B-model computation involves no such quantum correction and can be computed purely in terms of integrals over cycles ("periods") and their derivatives (discussed in the next section). So it is in some sense easier and, in a historic tour-de-force, was calculated by Candelas *et al.* (1991) for the Greene–Plesser mirror of the quintic. Comparing with the A-model computation on the quintic gave remarkable predictions about the number of holomorphic rational curves on the quintic. These were way beyond mathematical capabilities at the time, and sparked enormous mathematical interest. The predictions (and more) have now been proved to be true by Givental and Lian–Liu–Yau, while mirror symmetry has begun to be understood geometrically. But, in some sense, the mathematical reason for the relationship between the Yukawa couplings and the quantum cohomology of the mirror is still a little mysterious; it is the hardest part of mirror symmetry to see in the geometry, yet for the physics it was the easiest and the first prediction.

We survey, nonchronologically, some of the geometry of mirror symmetry as it is now understood, mainly in dimension $n=3$. For the many topics omitted, the reader should consult the Further Reading section.

## The Geometric Setup

A Calabi-Yau 3-fold $(X, \Omega, \omega)$ is a Kähler manifold $(X, \omega)$ with a holomorphic trivialization $\Omega$ of its canonical bundle

$$K_X = \Lambda^3_{\mathbb{C}} T^* X$$

(i.e., a nowhere-vanishing holomorphic volume form, locally $\mathrm{d}z_1 \wedge \mathrm{d}z_2 \wedge \mathrm{d}z_3$), and $b_1(X)=0$. It follows that the Hodge numbers $h^{0,2}, h^{0,1}$ vanish, and so $H^2(X,\mathbb{C}) = H^{1,1}$ and $H^3(X,\mathbb{R}) \cong H^{2,1} + H^{3,0}$. By Yau's theorem the Kähler metric can be changed within its $H^2(X,\mathbb{R})$ cohomology class to a unique Ricci-flat Kähler metric; equivalently, $\Omega$ is parallel, so the induced metric on $K_X$ is flat. Roughly speaking, mirror symmetry swaps the symplectic or Kähler structure $\omega$ on $X$ with the complex structure (encoded in $\Omega$, up to scaling by $\mathbb{C}^*$) on the (conjectural) mirror $\check{X}$. Kähler deformations are unobstructed, forming an open set $\mathcal{K}_X$ in $H^2(X,\mathbb{R})$. Its closure $\overline{\mathcal{K}}_X$ is sometimes extended by adding the Kähler cones of all birational models of $X$ to give Kawamata's movable cone. This is because the work of Aspinwall, Greene, Morrison, and Witten suggested that all birational models of $X$ are indistinguishable in string theory and so are all mirrors of $\check{X}$, corresponding to a different choice of $(1,1)$-form $\omega$ which is a Kähler form on one model only. $\mathcal{K}_X$ is also complexified by including in the A-model data any "B-field" $B \in H^2(X, \mathbb{R}/\mathbb{Z})$, and divided by holomorphic automorphisms of $X$, to give a moduli space of complex dimension $h^{1,1}(X)$. Deformations of complex structure are also unobstructed by the nontrivial Bogomolov–Tian–Todorov theorem; thus, they form a smooth space with tangent space

$$H^1(T\check{X}) \underset{\simeq}{\xrightarrow{\lrcorner\Omega}} H^1(\Lambda^2 T^* \check{X}) = H^{2,1}(\check{X})$$

(Given a deformation of complex structure, the above isomorphism takes the $H^{2,1}$-component of the derivative of the $(3,0)$-form $\Omega$.) So, for the moduli spaces to match up, we get the first and simplest prediction of mirror symmetry:

$$h^{1,1}(X) = h^{2,1}(\check{X}) \quad \text{and} \quad h^{2,1}(X) = h^{1,1}(\check{X}) \qquad [1]$$

This is where mirror symmetry gets its name, the above relation making the Hodge diamonds of $X$ and $\check{X}$ mirror images of each other.

As the complexified Kähler cone is a tube domain, it has natural partial complex compactifications (due to Looijenga, and suggested in the context of mirror symmetry by Morrison (1993)). The simplest case is where we ignore the movable cone and automorphisms and assume that there is an integral basis $e_1,\ldots,e_n$ of both $\overline{\mathcal{K}}_X$ and $H^2(X,\mathbb{Z})$/torsion. The complexified Kahler moduli space is then

$$\mathcal{K}_X^{\mathbb{C}} := H^2(X,\mathbb{R})/H^2(X,\mathbb{Z}) + \mathrm{i}\mathcal{K}_X = \{B + \mathrm{i}\omega\}$$

with natural coordinates $x_i, y_i \geq 0$ pulled back from the first and second factors, respectively, induced by the $e_i$. $x_i$ is multivalued with integer periods, so

$$z_i = \exp(2\pi\mathrm{i}(x_i + \mathrm{i}y_i)) \qquad [2]$$

is a well-defined holomorphic coordinate, giving an isomorphism to the product of $n$ punctured unit disks in $\mathbb{C}$:

$$\mathcal{K}_X^{\mathbb{C}} \cong (\Delta^*)^n = \{(z_i) : 0 < |z_i| \leq 1\} \subset (\mathbb{C}^*)^n$$

The compactification $\Delta^n$ comes from adding in the origins in the disks, which we reach by going to infinity (in various directions) in $\mathcal{K}_X^{\mathbb{C}}$. We call the point $(0,\ldots,0) \in \Delta^n$ the large Kahler limit point (LKLP) in this case. Moving along the ray generated by $\sum k_i e_i \in \mathcal{K}_X, k_i \geq 0$, complexifies in the holomorphic structure [2] to give the analytic curve

$$z_i^{k_j} = z_j^{k_i}, \quad \forall i,j \qquad [3]$$

in $\mathcal{K}_X^{\mathbb{C}}$. For $k_i \in \mathbb{Q}\ \forall i$, this extends to a complete curve in the compactification. Without loss of generality, we can assume that $k_i$ are integers with no common factor; then the link of the curve winds around the LKLP $(0,\ldots,0) \in \Delta^n$ with winding number

$$\begin{aligned}(k_1,\ldots,k_n) &\in \pi_1(H^2(X,\mathbb{R})/H^2(X,\mathbb{Z}) + \mathrm{i}\mathcal{K}_X) \\ &= H^2(X,\mathbb{Z}) = \mathbb{Z}.e_1 \oplus \cdots \oplus \mathbb{Z}.e_n\end{aligned}$$

This is because multiplying the ray $\mathbb{R}.\Sigma k_i e_i \in \mathcal{K}_X$ by i gives the direction $\mathbb{R}.\Sigma k_i e_i$ in the space $H^2(X,\mathbb{R})/H^2(X,\mathbb{Z})$ of B-fields, with the given winding number. For $k_i$ not rational we get an analytic mess; the direction in the space of B-fields does not close up to give a circle.

There is no obvious mirror to these rays since we consider $\Omega$ only up to scale. So, mirror symmetry predicts an isomorphism between $\mathcal{K}_X^{\mathbb{C}}$ and the moduli space $\mathcal{M}_{\check{X}}$ of complex structures on $\check{X}$, and a distinguished limit in $\mathcal{M}_{\check{X}}$, the large complex structure limit point (LCLP), the mirror of the LKLP $(0,\ldots,0) \in \Delta^n$ above. Morrison has given a rigorous definition of LCLPs and the canonical coordinates on $\mathcal{M}_{\check{X}}$ dual to the $z_i$ on $\mathcal{K}_X^{\mathbb{C}}$; see the section Monodromy around the LCLP. The holomorphic curves in $(\Delta)^n$ described above, corresponding to rational rays of Kähler forms, give degenerations of (the complex structure on) $\check{X}$ to the LCLP whose monodromy is discussed in this article (see "Lagrangian Torus Fibrations").

LCLPs play a vital role in mirror symmetry; in fact, mirror symmetry is really a statement about LCLPs and families of Calabi–Yau manifolds near LCLPs. Most predictions only really hold near or at the LCLP, and the complex structure moduli space only looks like $\Delta^n$ near the LCLP. For instance, manifolds can have many LCLPs and accordingly many mirrors. This also explains one obvious paradox – that rigid Calabi–Yau manifolds, those with no complex structure deformations, $h^{2,1} = 0$, and so no LCLP, can have no mirror, since a Kähler (or symplectic) manifold has $h^2 = h^{1,1} \neq 0$.

The first predicted refinement of [1] is, as discussed in the introduction, that the variation of Hodge structure (VHS) on $\check{X}$ should be describable in terms of Gromov–Witten invariants of $X$. Here VHS is governed by how the ray $\mathbb{C}.\Omega_t = H^{3,0}(\check{X}_t)$ sits inside $H^3(\check{X}_t,\mathbb{C})$ as the complex structure on $\check{X}_t$ varies, parametrized by $t \in \mathcal{M}_{\check{X}}$. By Poincaré duality, it is sufficient to know how $\Omega_t$ pairs with $H_3(\check{X})$, that is, to compute the period integrals

$$\int_{A_i} \Omega_t, \quad i = 1,\ldots,2k = 2h^{2,1} + 2$$

where $A_i$ form a basis of $H_3(\check{X},\mathbb{Z})$. (In fact we can choose the $A_i$ to be a symplectic basis, $A_i.A_j = \delta_{i+k,j}$, and then knowledge of only the periods of the first $k$ $A_i$ suffices, locally in moduli space.) These periods determine $\Omega_t$ and so the Yukawa coupling

$$H^1(T\check{X}_t)^{\otimes 3} \xrightarrow{\cup} H^3(\Lambda^3 T\check{X}_t) \xrightarrow{\lrcorner\Omega_t^{\otimes 2}} H^3(K_{\check{X}_t}) \cong \mathbb{C} \qquad [4]$$

On $X$, we get the cubic form on $H^2(X)$ described earlier in terms of numbers of rational curves in $X$. These numbers are in fact independent of the almost-complex structure on $X$ (as long as it is compatible with the symplectic form $\omega$), and, therefore, give the symplectic invariants of Gromov and Witten. The cubic form depends on $\omega = \omega_t$ as it moves in $\mathcal{K}_{X_t}$ (or in $\mathcal{K}_{X_t}^{\mathbb{C}}$, replacing $\omega_t$ by $-\mathrm{i}(B_t + \mathrm{i}\omega_t)$). Under the predicted local isomorphism $\mathcal{K}_X^{\mathbb{C}} \cong \mathcal{M}_{\check{X}}$ near the LKLP and LCLP, the equality of these cubic forms gives the predictions of number of rational curves in $X$ mentioned in the introduction. This has been carried out, and the predictions checked rigorously, in quite some generality, for instance for mirror pairs produced by Batyrev's toric methods.

There is, of course, a flat connection, the Gauss–Manin connection on the bundle over $\mathcal{M}_{\check{X}}$ with fiber $H^3(\check{X}_t, \mathbb{C})$ over $t \in \mathcal{M}_{\check{X}}$, given by the local system $H^3(\check{X}_t, \mathbb{Z}) \subset H^3(\check{X}_t, \mathbb{C})$. As mirror to this, Dubrovin has shown how to put a flat connection on the bundles with fibers $H^2(X_t)$ and $H^{\mathrm{ev}}(X_t)$ using Gromov–Witten invariants.

## Homological Mirror Symmetry

Building on the work of Witten, Kontsevich (1995) proposed a remarkable conjecture that purported to explain mirror symmetry, all the more surprising because it appeared to have little to do with what was thought to be mirror symmetry at the time. The conjecture is now reasonably well understood, while the link to Gromov–Witten invariants and Yukawa couplings is more mysterious, although it is known how both data should be encoded in the conjecture.

Kontsevich proposed that mirror symmetry should be explained by a (noncanonical) equivalence of triangulated categories between the derived Fukaya category $D^{\mathcal{F}}(X)$ of $(X, \omega)$ and the bounded derived category of coherent sheaves $D^{\mathrm{b}}(\check{X})$ on its mirror $\check{X}$. This second category consists of chain complexes of holomorphic bundles, with quasi-isomorphisms (maps of chain complexes which induce isomorphisms on cohomology) formally inverted, that is, decreed to be isomorphisms. For zero B-field the first category should be constructed from Lagrangian submanifolds $L \subset X$ carrying flat unitary connections $A$. That is, $L$ is middle- (three-) dimensional, and

$$\omega|_L \equiv 0, \qquad F_A = 0$$

For $B \neq 0$, this needs modifying to $F_A + 2\pi \mathrm{i} B.\mathrm{id} = 0$ (so, in particular, we require that $L$ satisfies $[B|_L] = 0 \in H^2(L, \mathbb{R}/\mathbb{Z})$). There are also various technical conditions such as the choice of a relative spin structure, the Maslov class of $L$ must vanish (i.e., the map $(\Omega|_L/\mathrm{vol}_L) : L \to \mathbb{C}^*$ has winding number zero) and we pick a grading on $L$ (a choice of logarithm of this map). Morphisms are defined by Floer cohomology $HF^*$ of Lagrangian submanifolds; roughly speaking, this assigns a vector space to each intersection point (the homomorphisms between the fibers of the two unitary bundles carried by the Lagrangians at this point), made into a chain complex by a certain counting of holomorphic disks between intersection points. In-depth work by Fukaya–Oh–Ohta–Ono shows that this gives the structure of an $A^\infty$-category which can then be "derived" into a triangulated category in a formal way by taking "twisted cochains." The construction is still very technical and difficult to calculate with, but the key points are that we get a category depending only on the symplectic structure, that certain "unobstructed" Lagrangian submanifolds give objects of this category, and that Hamiltonian isotopic unobstructed Lagrangian submanifolds give isomorphic objects.

Since the introduction of D-branes there is a physical interpretation of this conjecture in terms of open string theory; the objects of the two categories are boundary conditions for open strings, and morphisms correspond to strings beginning on one object and ending on the other. So, for instance, intersections of Lagrangians give morphisms corresponding to constant strings at the intersection point, while the Floer differential gives instanton tunneling corrections.

One paradox this formulation immediately sheds light on concerns automorphisms on both sides of mirror symmetry. While symplectomorphisms of $(X, \omega)$ are abundant, there are few holomorphic automorphisms of a Calabi–Yau $\check{X}$. The former induce autoequivalences of $D^{\mathcal{F}}(X)$; Kontsevich's suggestion is that as a mirror to this there should be an autoequivalence of $D^{\mathrm{b}}(\check{X})$; this need not be induced by an automorphism of $\check{X}$. Motivated by this, groups of autoequivalences of derived categories of sheaves of Calabi–Yau manifolds have now been found that were predicted by mirror symmetry; a few are mentioned below. Thus, homological mirror symmetry suggests that an SCFT is equivalent to a triangulated category, and the ambiguities in geometrizing an SCFT (finding a Calabi–Yau of which it is a $\sigma$-model) are seen in the category – not all automorphisms come from an automorphism of a Calabi–Yau (e.g., Calabi–Yau manifolds $\check{X}$ with equivalent derived categories give multiple mirrors to $X$), and not all appropriate categories need even come from a Calabi–Yau. Supporting this suggestion, Bondal–Orlov and Bridgeland have shown that indeed birational Calabi–Yau manifolds $\check{X}$ have equivalent derived categories.

Finally, Kontsevich explained how deformation theory of the categories should involve derived morphisms on the product from the diagonal (thought of as a Lagrangian in the A-model, its structure sheaf as a coherent sheaf in the B-model) to itself, giving quantum cohomology in the A-model and Hodge structure in the B-model. For instance, the holomorphic disks used to compute the Floer cohomology of the diagonal on the product $X \times X$ give holomorphic rational curves on $X$. So, one should be able to see some parts of "classical" mirror symmetry.

Below, as we describe more of the geometry of mirror symmetry that has emerged since Kontsevich's conjecture, we will mention at each stage how his conjecture fits in with it.

## The Strominger–Yau–Zaslow Conjecture

To recover more geometry from Kontsevich's conjecture, there are some obvious objects of $D^{\mathrm{b}}(\check{X})$ that reflect the geometry of $\check{X}$ – the structure sheaves $\mathcal{O}_p$ of points $p \in \check{X}$. Calculating their self-Homs, $\mathrm{Ext}^*(\mathcal{O}_p, \mathcal{O}_p) \cong \Lambda^* T_p\check{X} \cong \Lambda^*\mathbb{C}^3 \cong H^*(T^3, \mathbb{C})$, shows that if they are mirror to Lagrangians $L$ in $X$ (with flat connections $A$ on them) then we must have

$$HF^*((L, A), (L, A)) \cong H^*(T^3, \mathbb{C})$$

as graded vector spaces. Since the left-hand side is, modulo instanton corrections, $H^*(L, \mathbb{C})^{\oplus r}$, where $r$ is the rank of the bundle carried by $L$, this suggests that the mirror should be $L \cong T^3$ with a flat U(1) connection $A$ over it. There are reasons why the Floer cohomology of such an object should not be quantum corrected, and so be isomorphic to $\mathrm{Ext}^*(\mathcal{O}_p, \mathcal{O}_p)$.

For any Lagrangian $L$, the symplectic form gives an isomorphism between $T^*L$ and its normal bundle $N_L$; thus, Lagrangian tori have trivial normal bundles, and locally one can fiber $X$ by them. Thus, one might hope that $X$ is fibered by Lagrangian tori, and the mirror $\check{X}$ is (at least over the locus of smooth tori) the dual fibration. This is because the set of flat U(1) connections on a torus is naturally the dual torus.

This is the kind of philosophy that led to the Strominger–Yau–Zaslow (SYZ) conjecture (Strominger *et al.* 1996), although Strominger *et al.* were working with physical D-branes, and not Kontsevich's conjecture. Therefore, their D-branes are not the "topological D-branes" of Kontsevich, but those minimizing some action. That is, instead of holomorphic bundles in the B-model, we deal with bundles with a compatible connection satisfying an elliptic partial differential equation (PDE) (e.g., the Hermitian–Yang–Mills equations (HYM), or some perturbation thereof); instead of Lagrangian submanifolds up to Hamiltonian isotopy in the A-model, we consider special Lagrangians (sLags) (see eqn [5]). The SYZ conjecture is that a Calabi–Yau $X$ should admit a sLag torus fibration, and that the mirror $\check{X}$ should admit a fibration which is dual, in some sense.

A sLag is a Lagrangian submanifold of a Calabi–Yau manifold $X$ satisfying the further equation that the unit norm complex function (phase)

$$\frac{\Omega|_L}{\mathrm{vol}_L} = \mathrm{e}^{\mathrm{i}\theta} = \text{constant} \qquad [5]$$

(So, sLags have Maslov class zero, in particular.) This equation uses the complex structure on $X$ as well as the symplectic structure, and the resulting Ricci-flat metric of Yau, to define a metric on $L$ and so its Riemannian volume form $\mathrm{vol}_L$. SLags are calibrated by $\mathrm{Re}(\mathrm{e}^{-\mathrm{i}\theta}\Omega)$ and so minimize volume in their homology class. This is similar to the HYM equations on the mirror $\check{X}$, which are defined on holomorphic bundles on the complex manifold $\check{X}$ via a Kähler form $\omega$, and minimize the Yang–Mills action. The Donaldson–Uhlenbeck–Yau theorem states that for holomorphic bundles that are polystable (defined using $[\omega]$, this is true for the generic bundle), there is a unique compatible HYM connection. Thus, modulo stability, HYM connections are in one-to-one correspondence with holomorphic bundles. A similar correspondence is conjectured, and proved in some special cases, by Thomas and Yau, for (special) Lagrangians: that modulo issues of stability (which can be formulated precisely), sLags are in one-to-one correspondence with Lagrangian submanifolds up to Hamiltonian isotopy. That is, there should be a unique sLag in the Hamiltonian isotopy class of a Lagrangian if and only if it is stable. Currently, only the uniqueness part of this conjecture has been worked out, but, in principle at least, we do not lose much by considering only Lagrangian torus fibrations.

The SYZ conjecture is thought to hold only near the LCLPs and LKLPs of $X$ and $\check{X}$; away from these, the sLag fibers may start to cross. According to Joyce, the discriminant locus of the fibration on $X$ is expected to be a codimension one ribbon graph in a base $S^3$ near the limit points, while the discriminant locus of the dual fibration $\check{X}$ may be different – that is, the smooth parts of the fibration and its dual are compactified in different ways. In the limit of moving to the limit points, however, both discriminant loci shrink onto the same codimension-two graph. In this limit, the fibers shrink to zero size, so that $X$ (with its Ricci-flat metric) tends, in the Gromov–Hausdorff sense, to its base $S^3$ (with a singular metric). This formal picture has been made precise in two dimensions, for $K3$-surfaces, by Gross and Wilson. The limiting picture suggests that if we are only interested in topological or Lagrangian torus fibrations then we might hope for codimension-two discriminant loci, and such fibrations might make sense well away from limit points. Gross and Ruan carry this out in examples such as the quintic and its mirror, and makes sense of dualizing the fibration by dualizing monodromy around the discriminant locus

and specifying a canonical compactification over the discriminant locus. This gives the correct topology for toric varieties and their mirrors, and flips the Hodge numbers [1], for instance. Approaching the LCLP in a different way (in the example of eqn [3] this corresponds to altering the rational numbers $k_i$) can give a different graph and different fibration on $X$; the dual fibration can then be a topologically different manifold, giving a different birational model of the mirror $\check{X}$.

We focus only on Lagrangian fibrations, as they are better behaved and understood. We can expect them to be $C^\infty$ fibrations with codimension-two discriminant loci, for instance. Below we see how to put a complex structure on the smooth part of the fibration, but extending this over the compactification is much harder and will involve "instanton corrections" coming from holomorphic disks. Fukaya (2005) has beautiful conjectures about this that will explain a great deal more of mirror symmetry, but they will not be discussed here.

## Lagrangian Torus Fibrations

If $(X^{2n}, \omega) \xrightarrow{\pi} B^n$ is a *smooth* Lagrangian fibration with compact fibers, then the fibration is naturally an affine bundle of torus groups (i.e., a bundle of groups once we pick a Lagrangian 0-section – an identity in each fiber), and the base $B$ inherits a natural integral affine structure: it looks like a vector space $V$ with an integral structure $V \cong \Lambda \otimes_{\mathbb{Z}} \mathbb{R}$ up to translation by elements of $V$. This is the classical theory of action-angle variables. $T_b^*B$ acts on the fiber $X_b = \pi^{-1}(b)$: by pullback and contraction with the symplectic form, $\sigma \in T_b^*B$ gives a vector field $\underline{\sigma}$ tangent to $X_b$, and the time-one flow along $\underline{\sigma}$ gives the action. By compactness and smoothness of $X_b$ the kernel is a full-rank lattice $\Lambda_b \subset T_b^*B$, giving the isomorphism

$$X_b \cong T_b^*B/\Lambda_b$$

We define the integral affine structure on $B$ by specifying the integral affine functions $f$ (up to translation) to be those whose time-one flow along $\underline{df}$ is the identity (i.e., on the universal cover the time-one flow is to a section of the bundle of lattices $\Lambda$).

The situation that concerns us is where $B$ is a 3-manifold $\bar{B}$ (usually $S^3$) minus a graph; then the monodromy around the graph preserves the integral affine structure:

$$\pi_1(B) \to \mathbb{R}^3 \rtimes GL(3, \mathbb{Z}) \qquad [6]$$

A great deal of mirror symmetry can be seen from just this knowledge of the smooth locus of the fibration; in particular, Gross (1998) has shown how mild assumptions about the compactification (with singular fibers over $\bar{B}\backslash B$) are enough to determine much of the topology of $X$. The dual fibration $\check{\pi}$ should have the monodromy dual to [6], and he shows how this implies the switching of the Hodge numbers [1] by the Leray spectral sequence; the rough idea being the obvious isomorphism

$$R^i\pi_*\mathbb{R} \cong \Lambda^i TB \cong \Lambda^{3-i}T^*B \cong R^{3-i}\check{\pi}_*\mathbb{R}$$

induced by a trivialization of $\Lambda^3 TB$. That is, morally speaking, the flipping of Betti numbers arises by representing cycles by those with linear intersection with the fibers, and replacing this linear space by its annihilator in the dual torus. This also agrees with the equivalence taking Lagrangians to coherent sheaves described in the next section.

The dual fibration $\check{\pi}$ has a natural complex structure; here the affine structure is essential, as in general a tangent bundle $TB$ only has a natural almost complex structure along its 0-section. Since, up to translation, locally $B \cong V$ is a vector space, $TB \cong V \times V \cong V \otimes_{\mathbb{R}} \mathbb{C}$ has a natural complex structure which descends to

$$\check{\pi} : \check{X} = TB/\Lambda^* \to B \qquad [7]$$

Gross suggests that the B-field on $X$ should lie in the piece

$$H^1(R^1\pi_*\mathbb{R}/\mathbb{Z}) = H^1(TB/\Lambda^*)$$

of the Leray spectral sequence converging to $H^2(X, \mathbb{R}/\mathbb{Z})$. That is, it is represented by a Čech cocycle $e$ on overlaps of an open cover of $B$ with values in the dual bundle of groups $TB/\Lambda^*$. Using this to twist [7] and re-glue it via transition functions translated by $e$, we get a new complex manifold ($e$ is locally constant, so translation by $e$ is holomorphic) which we consider as mirror to $X$ with complexified form $B + i\omega$. In this way, Gross manages to match up complexified symplectic deformations of $X$ with complex structures on $\check{X}$.

## The 2-Torus

Mirror symmetry is nontrivial even for the simplest Calabi–Yau – the 2-torus. This can be written as an SYZ fibration $T^2 \xrightarrow{\pi} B = S^1$, and write $B$ as $\mathbb{R}/a\mathbb{Z}$ with its standard integral affine structure induced by $\mathbb{Z} \subset \mathbb{R}$. This trivializes $T^*B = B \times \mathbb{R}$ and the lattice $\Lambda$ in it as $B \times \mathbb{Z} \subset B \times \mathbb{R}$. So as a symplectic manifold,

$$T^2 = \frac{T^*S^1}{\Lambda} = \frac{[0, a] \times [0, 1]}{(0, p) \sim (a, p), (q, 0) \sim (q, 1)} \qquad [8]$$

with symplectic coordinates $(q,p)$ in which the symplectic form is $\omega = \mathrm{d}p \wedge \mathrm{d}q$ (so $\int_{T^2} \omega = a$). Again, the B-field, $b \in H^1(R^1\pi_*\mathbb{R}/\mathbb{Z}) = H^2(T^2, \mathbb{R}/\mathbb{Z})$, is in $H^1$ of the locally constant sections of the dual fibration.

In our trivialization $B \cong \mathbb{R}/a\mathbb{Z}$, $\Lambda^* \subset TB$ is also standard: $B \times \mathbb{Z} \subset B \times \mathbb{R}$, so the mirror has the same description as in [8] in which the complex structure is standard: $J\partial_p = \partial_q$. That is, $p + \mathrm{i}q$ gives a local holomorphic coordinate.

For nonzero B-field $b \neq 0$, twisting the dual fibration by $b$ gives

$$T^2 = \frac{T^*S^1}{\Lambda} = \frac{[0,a] \times [0,1]}{(0,p) \sim (a, b+p), (q,0) \sim (q,1)} \qquad [9]$$

again with holomorphic structure given by $p + \mathrm{i}q$ and SYZ fibration $\breve{\pi}$ being projection onto $q$. So, as a complex manifold the mirror is $\mathbb{C}$ divided by the lattice

$$\Lambda = \langle 1, b + \mathrm{i}a \rangle$$

Changing $b$ to $b+1$ does not alter this lattice, so the construction is well defined for $b \in \mathbb{R}/\mathbb{Z} \cong H^1(R^1\pi_*\mathbb{R}/\mathbb{Z})$, and we have the standard description of an elliptic curve via its period point $\tau = b + \mathrm{i}a$ in the upper half plane (as $a > 0$). Mirror symmetry has indeed swapped the complexified symplectic parameter $b + \mathrm{i}a = \int_{T^2}(b + \mathrm{i}\omega)$ for the complex structure modulus $\tau = b + \mathrm{i}a$. $\mathrm{SL}(2,\mathbb{Z})$ acts on both sides (in the standard way on $\tau$, and as symplectomorphisms modulo those isotopic to the identity on the A-side) permuting the choices of SYZ fibration. We note that in this case the fibrations are special Lagrangians in the flat metric, with no singular fibers.

Polishchuk and Zaslow have worked out in detail how Kontsevich's conjecture works in this case. The general picture for any torus fibration is an extension of the fiberwise duality that led to SYZ. Namely, Lagrangian multisections $L$ of the fibration, of degree $r$ over the base, give $r$ points on each fiber, and so $r$ flat U(1) connections on the dual fiber. The resulting $\mathrm{U}(1)^{\times r}$ connections can be glued together and twisted by the flat connection on $L$, to give a rank-$r$ vector bundle with connection on the mirror. Arinkin and Polishchuk show that in general the Lagrangian condition implies the integrability condition $F^{0,2} = 0$ of the resulting connection, giving a holomorphic structure on the bundle. Leung–Yau–Zaslow show that the special Lagrangian condition gives a perturbation of the HYM equations on the connection. Branching of sections has been dealt with by Fukaya, and requires instanton corrections from holomorphic disks. Other Lagrangians with linear intersection with the fibers can be dealt with similarly. $T^2$ is simpler because all Lagrangians with vanishing Maslov class can be isotoped into straight lines (i.e., sLags in the flat metric) with no branching. The upshot is that the slope of the sLag over the base corresponds to the slope $(\int_{T^2} c_1/\mathrm{rank}) \in [-\infty, \infty]$ of the mirror sheaf.

### The Large Complex Structure Limit

The LKLP for $T^2$ is clearly $\lim a \to \infty$. On the mirror then, the LCLP is at $\tau = b + \mathrm{i}a \to b + \mathrm{i}\infty$, the nodal torus compactifying the moduli of elliptic curves. Metrically, however, in the (Ricci-) flat metric, things look different; if we rescale to have fixed diameter, the torus collapses to the base of its SYZ fibration, and all of its fibers contract. This is an important general feature of the difference between complex and metric descriptions of LCLPs; see the description of the quintic in the next section.

We note that, as in the compactifications discussed in an earlier section, the monodromy around this LCLP is given by rotating the B-field: $b \mapsto b + 1$. This gives back the same elliptic curve, but after a monodromy diffeomorphism $T$, which, from [9], is seen to be

$$T : q \mapsto q, \ p \mapsto p + q/a$$

On $H^1(T^2) = \mathbb{Z}[\text{fiber}] \oplus \mathbb{Z}[\text{section}]$ this acts as

$$T_* = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad [10]$$

This is called a Dehn twist. Picking the 0-section $O = \{p = 0\}$ in the mirror [9] when $b = 0$, this is taken to the section

$$T(O) = \{p = q/a\}$$

and $T$ is in fact the translation by this section $T(O)$ on $T^2$, using the group structure on the fibers (now we have chosen a 0-section). Again, Gross (1998) has shown that this is a general feature of LCLPs.

If we pick a Kähler structure on this family of complex tori, $T$ turns out to be a symplectomorphism. Importantly, its mirror is not a holomorphic automorphism, but an equivalence of the derived category of coherent sheaves. As above, the section $T(O)$ corresponds to a slope-one line bundle $L$ on the mirror, and the monodromy action corresponds to

$$\otimes L : D^{\mathrm{b}} \to D^{\mathrm{b}} \qquad [11]$$

on the derived category. Again, this is a more general feature of these LCLPs, with $L$ such that $c_1(L)$ equals the symplectic form which generated

the ray along which the original LKLP was reached. In general, the SYZ fiber is the invariant cycle under $T_*$ [10], and, on the mirror, structure sheaves of points are invariant under $\otimes L$. On the cohomology of $T^2$, cupping with $\mathrm{ch}(L) = e^{c_1(L)} = 1 + c_1(L)$ has the same action [10] on $H^{\mathrm{ev}} = \mathbb{Z}(c_1(L)) \oplus \mathbb{Z}(1)$.

Notice we have used the choices of fibration and 0-section to produce the equivalence of triangulated categories and to equate the monodromy actions. Kontsevich's conjectural equivalence is not canonical, but is fixed by a choice of fibration and 0-section. In turn, a fibration should be fixed by a choice of LCLP or LKLP from the resulting collapse (in the Ricci-flat metric) onto a half-dimensional $S^n$ base. The choice of 0-section is then rather arbitrary (as monodromy about the LCLP changes it) but determines the equivalence of categories. Different choices of section give different equivalences, differing, for instance, by the monodromy transformation $\otimes L$ [11].

Another point of view is that a Lagrangian fibration and 0-section determine a group structure on the fibers and so on the Fukaya category (translating Lagrangian multisections by multiplication on each fiber). This corresponds to a choice of tensor product on the derived category of the mirror; the identity for this product is then the structure sheaf $\mathcal{O}_X$ mirror to the 0-section, and an ample line bundle is given by the action of the monodromy transformation $L = T(\mathcal{O}_X)$; $T$ then acts as $\otimes T(\mathcal{O}_X)$ [11]. Since $X$ is determined by the graded ring

$$\bigoplus_{j \gg 0} H^0_X(L^j) = \bigoplus_{j \gg 0} \mathrm{Hom}^*(\mathcal{O}_X, T^j(\mathcal{O}_X))$$

one might also try to construct $X$ purely from the 0-section $O$ and LCLP monodromy on $\check{X}$, as

$$X = \mathrm{Proj} \bigoplus_{j \gg 0} HF^*(O, T^j(O))$$

A problem is to show that $\oplus_{j \gg 0} HF^0(O, T^j(O))$ is finitely generated; a related problem is to show that, for $j \gg 0$, the above Floer homologies vanish except for $* = 0$.

We now turn to the quintic 3-folds, where we will see how to identify the (homology classes of the) 0-section and fiber in general using Hodge theory.

## The Quintic 3-Fold

The simplest Calabi–Yau 3-fold is given by the zeros $Q$ of a homogeneous quintic polynomial on $\mathbb{P}^4$, that is, an anticanonical divisor of $\mathbb{P}^4$. By adjunction, this has trivial canonical bundle, and so is Calabi–Yau. By the Lefschetz hyperplane theorem, it has $h^{1,1} = 1$, so computing its Euler number to be $e = -200$, we find that $h^{2,1} = 101$ gives its number of complex deformations. Alternatively, this can be seen by showing that all such deformations are themselves quintics, then dividing the 126-dimensional space of quintic polynomials by the 25-dimensional GL $(5, \mathbb{C})$. Thus, its mirror has one complex structure deformation and 101 Kähler classes.

Greene and Plesser prescribed the following mirror. Take the special one-dimensional family of Fermat quintics

$$Q_\lambda = \left\{ \sum_{i=0}^{4} x_i^5 - \lambda \prod_{i=0}^{4} x_i = 0 \right\} \subset \mathbb{P}^4 \qquad [12]$$

with the action of $\{(\alpha_0, \ldots, \alpha_4) \in (\mathbb{Z}/5)^5 : \prod_i \alpha_i = 1\} \cong (\mathbb{Z}/5)^4$ given by rescaling the $x_i$ by fifth roots of unity. Dividing by the diagonal $\mathbb{Z}/5$ projective stabilizer, we get a free $(\mathbb{Z}/5)^3$ action; the mirror of the quintic is any crepant ($K = \mathcal{O}$) resolution of the quotient:

$$\check{Q}_\lambda = \frac{\widehat{Q_\lambda}}{(\mathbb{Z}/5)^3}$$

Different resolutions give different Kähler cones whose union is the moveable cone; its complexification is locally isomorphic to the complex structure moduli space of $Q$. $h^{1,1}(\check{Q}_\lambda) = 101$ for any crepant resolution, and $h^{2,1}(\check{Q}_\lambda) = 1$ corresponds locally to the one complex structure deformation [12]. In fact, for $\alpha^5 = 1$, multiplying $x_0$ by $\alpha$ shows that $\check{Q}_\lambda \cong \check{Q}_{\alpha\lambda}$, and $\lambda^5$ parametrizes the complex structure moduli.

The LCLP is at $\lambda = \infty$, that is, it is the quotient of the union of hyperplanes

$$\begin{aligned} Q_\infty &= \left\{ \prod_{i=0}^{4} x_i = 0 \right\} \\ &= \{x_0 = 0\} \cup \cdots \cup \{x_4 = 0\} \end{aligned} \qquad [13]$$

This is a union of toric varieties, each with a $T^3$ action inherited from the toric $T^4$ action on $\mathbb{P}^4$. Much more generally, Batyrev's construction considers the anticanonical divisors (and even more generally, complete intersections) in toric varieties fibered over the boundary of the moment polytope, and takes as mirror the anticanonical divisor of the toric variety associated to the dual polytope. However, most of the geometry is visible in this quintic example.

Equation [13] is the analog of the nodal torus of the last section, and we emphasize again that metrically it looks nothing like this; the Ricci-flat metric collapses the $T^3$ toric fibers to the base $S^3$ (with a singular metric). General LCLPs look rather similar,

with such "as bad as possible" normal crossing singularities. Smoothing a local model (in $x_0 = 1$) $\prod_{i=1}^{4} x_i = 0$, we can see the tori in $\{\prod_{i=1}^{4} x_i = \epsilon\}$:

$$T^3 = \left\{ |x_1| = \delta_1, |x_2| = \delta_2, \right.$$
$$\left. |x_3| = \delta_3, x_4 = \frac{\epsilon}{x_1 x_2 x_3} \right\} \quad [14]$$

These are even Lagrangian in the standard symplectic form on the local model, and fiber the smoothing over the base $\{(\delta_1, \delta_2, \delta_3)\}$. It turns out that, metrically, these tori (which vanish into the normal crossings singularity at the LCLP) actually form a large part of the smooth Calabi–Yau. This enlightens the apparent paradox between the SYZ conjecture and the Batyrev construction, that is, why a vertex of the original moment polytope (corresponding to the deepest type of singularity $(0,0,0,0) \in \{\prod_{i=1}^{4} x_i = 0\}$) can be replaced by the dual three-dimensional face in the dual polytope. This was first suggested by Leung and Vafa.

Gross and Siebert (2003) exploit this to extend SYZ and Batyrev's construction to nontoric LCLP Calabi-Yau manifolds; it is only the local toric nature of the normal crossing singularities of the LCLP that they use. It seems possible that their construction will give the mirrors of all Calabi–Yau manifolds with LCLPs. Much of mirror symmetry should soon be reduced to graphs (the discriminant locus of a Lagrangian torus fibration) in spheres, and further graphs over which D-branes (such as holomorphic curves) fiber, as in recent conjectures of Kontsevich and Soibelman and Fukaya (2005). It may soon be possible to write down a triangulated category in terms of such data. The full geometric story (involving Joyce's description of sLag fibrations, for instance) is still some way off, however; we cannot even write down an explicit Ricci-flat metric on a compact Calabi–Yau.

#### Monodromy around the LCLP

As well as the SYZ torus fiber [14] we can also see a Lagrangian 0-section on the quintic and its mirror as a component of the real locus of [12] for $\lambda > 5$. Remarkably, like the torus [14], this cycle was already described and used by Candelas *et al.* (1991), long before the relevance of torus fibrations was suspected.

Gross and Ruan have been able to describe the quintic and its mirror (at least topologically or symplectically) very explicitly as a simple torus fibration over this $S^3$ with a natural integral affine structure and codimension-two graph discriminant locus (see, e.g., Gross *et al.* (2003)).

Under monodromy about $\lambda = \infty$, the 0-section is moved to another section $T(O)$, and $T$ is given by translation by $T(O)$ using the group structure on the fibers. This is the analog of the Dehn twist [10], and one can choose a basis of $H_3(\check{Q})$ (with first element the invariant cycle, the $T^3$-fiber, second element a cycle fibered over a curve in $S^3$, third fibered over a surface, and last the 0-section itself) such that

$$T_* = \begin{pmatrix} 1 & 1 & * & * \\ 0 & 1 & * & * \\ 0 & 0 & 1 & * \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad [15]$$

Like the Dehn twist [10], it turns out that $T_*$ is maximally unipotent; that is, we have in $n$-dimensions,

$$(T_* - 1)^{n+1} = 0 \quad \text{but} \quad (T_* - 1)^n \neq 0$$

Again, this is a general feature of LCLPs as formulated by Morrison (1993) as part of the definition.

This should be compared with the Lefschetz operator $L = \cup\omega$ on the cohomology of the mirror, which also satisfies $L^n \neq 0, L^{n+1} = 0$ (or, more relevantly, $\exp(L)$, which satisfies $(e^L - 1)^n \neq 0$, $(e^L - 1)^{n+1} = 0$). Their similarity was noticed by the Griffiths school working on VHS in the late 1960s! Now we know that for Calabi–Yau manifolds at an LCLP dual to an LKLP along a ray $\omega = c_1(L)$ on the mirror, they should be considered mirror operators (up to some factors of the Todd class of the underlying Calabi–Yau, to do with the relationship between the Chern character $e^\omega$ of the line bundle $L$ (see [11]) and the Riemann–Roch formula).

Both, by linear algebra of the nilpotent operator $N = \log T_* = \sum_{k=1}^{n} (T_* - 1)^k$, induce a natural filtration $W_\bullet : 0 \leq W_0 \leq \cdots \leq W_{2n} = H$ on the cohomology on which they operate (which is $H = H^n$ for $N = \log T_*$ and $H = H^{\text{ev}}$ for $N = L = \cup\omega$):

$$\begin{aligned} 0 &\leq \operatorname{im}(N^n) \leq \operatorname{im}(N^{n-1}) \cap \ker(N) \leq \cdots \\ &\leq \ker(N^{n-1}) + \operatorname{im}(N) \leq \ker(N^n) \leq H \end{aligned} \quad [16]$$

For a discussion of the construction of this monodromy weight filtration, the reader is referred to the further reading section. It plays a key role in studying degenerations of varieties and Hodge structures, in this case as we approach the LCLP. It is a beautiful result of Gross that this filtration coincides with the Leray filtration on $H^n$ induced by the fibration. That is, under Poincaré duality, the weight filtration on cycles is by the minimal dimension (over all homologous cycles) of the image in the base over which the cycle is fibered. So, the first graded piece is spanned by the invariant cycle, the $T^3$ fiber, supported over a point, and the last by the 0-section; cf. [15]. (Similarly on the mirror, the filtration for the Lefschetz operator $\cup e^\omega$ has first piece spanned by the cohomology class of a

point, which is invariant under the monodromy action $\otimes L$ of [11], etc.)

Letting $\gamma_0$ be the class of a fiber and $\gamma_1$ span $W_2/W_0$ (which is one-dimensional) over the integers, then $T_*\gamma_1 = \gamma_1 + \gamma_0$. It follows that

$$q = \exp\left(2\pi i \frac{\int_{\gamma_1}\Omega}{\int_{\gamma_0}\Omega}\right)$$

is invariant under monodromy. This is the higher-dimensional analog of the coordinate $\exp(2\pi i\tau)$ on the moduli space of elliptic curves, where $\tau$ is the period point. It is this coordinate $q$ that is mirror to the coordinate

$$\int_{\text{line}} \omega$$

on the Kähler moduli space on the mirror quintic, which allows one to compute the correspondence between VHS and Gromov–Witten invariants mentioned in the introduction.

More generally, following Morrison (1993), one can make a rigorous definition of an LCLP using features noted above extended to the case of $h^{2,1} > 0$ (see, e.g., Cox and Katz (1999). Roughly, the upshot is that $\mathcal{M}_{\check{X}}$ (of dimension $s = h^{2,1}(\check{X})$) should be compactified with $s$ divisors $(D_i)_{i=1}^s$ (parametrizing singular varieties) forming a normal crossings divisor meeting at the LCLP, with monodromies $T_i$ about them. There should be a unique (up to multiples) integral cycle $\gamma_0$ (our torus fiber) invariant under all $T_i$, and cycles $(\gamma_i)_{i=1}^s$ such that

$$\tau_i = \frac{\int_{\gamma_i}\Omega}{\int_{\gamma_0}\Omega}$$

is logarithmic at $D_i$; that is $\tau_i = (1/(2\pi i))\log(z_i)$, where $z_i$ is a local parameter for $D_i = \{z_i = 0\}$.

So, $z_i = \exp(2\pi i\tau_i)$ form local coordinates for moduli space, mirror to the polydisk coordinates [2] on $\mathcal{K}_X^{\mathbb{C}}$. The direction of approach to the LKLP in that section corresponds to the holomorphic curve $z_i^{k_j} = z_j^{k_i}$ [3] we take through the LCLP ($z_i = 0\,\forall i$), and the monodromy $\sum N_i T_i$ varies accordingly, but the corresponding weight filtration $W_\bullet$ remains constant if $k_i \neq 0\,\forall i$, by a theorem of Cattani and Kaplan.

Morrison then requires that the $(\gamma_i)_{i=0}^s$ should form an integral basis for $W_2 = W_3$ (with $\gamma_0$ a basis of $W_0 = W_1$). Finally, part definition and part conjecture, we should be able to make a choice such that they satisfy the condition $\log T_i(\gamma_j) = \delta_{ij}\gamma_0$.

Of course, as has been emphasized, Morrison's definition of an LCLP is really where the mathematics and geometry of mirror symmetry begin, and should have been the starting point of this article. But that would have required appreciable knowledge of abstract VHS that are best understood, in this context, through the new geometry of Lagrangian torus fibrations that mirror symmetry has inspired.

*See also:* AdS/CFT Correspondence; Calibrated Geometry and Special Lagrangian Submanifolds; Derived Categories; Fourier–Mukai Transform in String Theory; Geometric Analysis and General Relativity; Geometric Flows and the Penrose Inequality; Geometric Measure Theory; Geometric Phases; Number Theory in Physics; Riemann Surfaces; Several Complex Variables: Compact Manifolds; Topological Gravity, Two-Dimensional; Topological Sigma Models; WDVV Equations and Frobenius Manifolds.

## Further Reading

Candelas P, de la Ossa X, Green P, and Parkes L (1991) A pair of Calabi–Yau manifolds as an exactly soluble superconformal theory. *Nuclear Physics B* 359: 21–74.

Cox D and Katz S (1999) Mirror Symmetry and Algebraic Geometry. *Mathematical Surveys and Monographs,* vol. 68. Providence, RI: American Mathematical Society.

Fukaya K (2005) Multivalued Morse theory, asymptotic analysis and mirror symmetry. In: Mikhail Lyubich and Leon Takhtajan (eds.) Proceedings of the conference dedicated to Dennis Sullivan on his 60th birthday held at Stony Brook University, Stony Brook, NY, June 14–21, 2001. *Proceedings of Symposia in Pure Mathematics*, in Graphs and patterns in mathematics and theoretical physics, vol. 73, pp. 205–278. Providence, RI: American Mathematical Society.

Gross M (1998) Special Lagrangian fibrations I: topology. In: Ueno K, Saito M-H, and Shimizu Y (eds.) *Integrable Systems and Algebraic Geometry (Kobe/Kyoto, 1997)*, pp. 156–193. River Edge, NJ: World Scientific Publishing.

Gross M, Huybrechts D, and Joyce D (2003) Universitext. *Calabi–Yau Manifolds and Related Geometries.* Berlin: Springer.

Gross M and Siebert B (2003) Mirror symmetry via logarithmic degeneration data I. *Preprint math.AG/0309070.*

Hori K, Katz S, Klemm A, Pandharipande R, Thomas R *et al.* (2003) Mirror Symmetry. Clay Mathematics Monographs, vol. 1. Providence, RI: American Mathematical Society; Cambridge, MA: Clay Mathematics Institute.

Kontsevich M (1995) Homological Algebra of Mirror Symmetry. International Congress of Mathematicians. Zürich: Birkhäuser.

Morrison D (1993) Compactifications of moduli spaces inspired by mirror symmetry. *Astérisque* 218: 243–271.

Strominger A, Yau S-T, and Zaslow E (1996) Mirror symmetry is T-duality. *Nuclear Physics B* 479: 243–259.

Voisin C (1999) (Translated from the 1996 French original by Roger Cooke.) SMF/AMS Texts and Monographs, vol. 1. *Mirror Symmetry*. Providence, RI: American Mathematical Society.

**Modular Tensor Categories** *see* Braided and Modular Tensor Categories

# Moduli Spaces: An Introduction

**F Kirwan**, University of Oxford, Oxford, UK



An earlier version of this article was originally published in *Proceedings of the Workshop on Moduli Spaces, Oxford, 2–3 July 1998*, Eds. Kirwan F, Paycha S, Tsou S T (1998). Cairo, Egypt: Hindawi Publishing Corporation.

The concept of a moduli space has been used by mathematicians for nearly 150 years, although it was not until the 1960s that Mumford (1965) gave precise definitions of moduli spaces and methods for constructing them. The use of the word "moduli" in this context goes back to Riemann in a paper of 1857, in which he observed that an isomorphism class of compact Riemann surfaces of genus $g$ "hängt ... von $3g - 3$ stetig veränderlichen Grössen ab, welche die Moduln dieser Klasse genannt werden sollen." The idea of moduli as parameters in some sense measuring or describing the variation of geometric objects has been of fundamental importance in geometry ever since.

Moduli spaces arise naturally in classification problems in geometry, particularly in algebraic geometry (Mumford 1965, Newstead 1978, Popp 1977, Seshadri 1975, Sundaramanan 1980, Viehweg 1995). Algebraic geometry is, roughly speaking, the study of solutions of systems of polynomial equations in many variables; the solutions to such a system form an algebraic variety. A simple example of an algebraic variety is a hypersurface, consisting of the solutions to a single polynomial equation in some number of variables. We can try to classify hypersurfaces by their degree and their dimension; these are "discrete invariants" for the classification problem, but of course they do not determine hypersurfaces completely, even if we regard two hypersurfaces as equivalent when one is obtained from the other after making a change of coordinates. It is typical of classification problems in algebraic geometry (and other areas of geometry) that there are not enough discrete invariants to classify objects sufficiently finely, and this is where the concept of a moduli space arises.

In complex algebraic geometry, discrete invariants often come from topology. For example, a nonsingular complex curve (i.e., a complex algebraic variety which is a connected complex manifold of dimension 1, in other words a Riemann surface) which is projective (i.e., points have been added at infinity to make it compact) is topologically just a sphere with a number of handles attached to it; the number of handles is called the genus of the curve and is a discrete invariant. Nonsingular complex projective curves (or equivalently compact Riemann surfaces) are not classified completely by their genus $g$; they are determined by $g$ when regarded simply as topological surfaces, but the genus does not determine their complex structure when $g > 0$.

A classification problem such as this one (the classification of nonsingular complex projective curves up to isomorphism, or, equivalently, compact Riemann surfaces up to biholomorphism), can be resolved into two basic steps.

Step 1 is to find as many discrete invariants as possible (in the case of nonsingular complex projective curves the only discrete invariant is the genus).

Step 2 is to fix the values of all the discrete invariants and try to construct a "moduli space"; that is, a complex manifold (or an algebraic variety) whose points correspond in a natural way to the equivalence classes of the objects to be classified.

What is meant by "natural" here can be made precise (as we shall see shortly) given suitable notions of families of objects parametrized by base spaces and of equivalence of families. A "fine moduli space" is then a base space for a universal family of the objects to be classified (any family is equivalent to the pullback of the universal family along a unique map into the moduli space). If no universal family exists there may still be a "coarse moduli space" satisfying slightly weaker conditions, which are nonetheless strong enough to ensure that if a moduli space exists it will be unique up to canonical isomorphism.

It is often the case that not even a coarse moduli space will exist. Typically, particularly "bad" objects must be left out of the classification in order for a moduli space to exist. For example, a coarse moduli space of nonsingular complex projective curves exists (although to have a fine moduli space we must give the curves some extra structure, such as a level structure), but if we want to include singular curves (which is often important so that we can understand how nonsingular curves can degenerate to singular ones) we must leave out the so-called "unstable curves" to get a moduli space. However all nonsingular curves are stable, so the moduli space of stable curves of genus $g$ is then a compactification of the moduli space of nonsingular projective curves of genus $g$.

Moduli spaces are often constructed and studied as orbit spaces for group actions (using Mumford's geometric invariant theory or more recently ideas due to Kollár (1997) and Keel and Mori (1997); geometric invariant theoretic quotients can also often be described

naturally as symplectic reductions, and it is in this guise that many moduli spaces in physics appear. Another technique involves period maps, Torelli theorems and variations of Hodge structures, initiated by Griffiths (1984) and others. In the special case of moduli spaces of compact Riemann surfaces, Teichmüller theory can also be used (see e.g., Lehto (1987)).

**Remark 1** Recall that a compact Riemann surface (i.e., a compact complex manifold of complex dimension 1) can be thought of as a nonsingular complex projective curve, in the sense that every compact Riemann surface can be embedded in some complex projective space

$$\mathbb{P}_n = \mathbb{C}^{n+1} - \{0\}/(\text{multiplication by nonzero complex scalars})$$

as the solution space of a set of homogeneous polynomial equations. Moreover, two nonsingular complex projective curves are biholomorphic if and only if they are algebraically isomorphic. So, there is a natural identification between the moduli space of compact Riemann surfaces of genus $g$ up to biholomorphism and the moduli space of nonsingular complex projective curves up to isomorphism.

There are other situations where an "algebraic" moduli space can be naturally identified with the corresponding "complex analytic" moduli space, but this is not always the case. For example, if we consider K3 surfaces (compact complex manifolds of complex dimension 2 with first Betti number and first Chern class both zero), we find that the moduli space of all K3 surfaces has complex dimension 20, whereas the moduli spaces of algebraic K3 surfaces (which have one more discrete invariant, the degree, to be fixed) are 19-dimensional.

This problem of algebraic moduli spaces versus nonalgebraic ones is one reason why the question of classifying $n$-folds (i.e., compact complex manifolds – or, in the algebraic category, nonsingular projective varieties – of dimension $n$) becomes much harder when $n > 1$ than in the case $n = 1$ (which is the case of compact Riemann surfaces or nonsingular projective curves). Another difficulty is that families of $n$-folds can be "blown up" along families of subvarieties to produce ever more complicated families.

**Remark 2** Recall that we blow up a complex manifold $X$ along a closed complex submanifold $Y$ by removing the submanifold $Y$ from $X$ and glueing in the projective normal bundle of $Y$ in its place. We get a complex manifold $\tilde{X}$ with a holomorphic surjection $\pi : \tilde{X} \to X$ such that $\pi$ is an isomorphism over $X - Y$ and if $y \in Y$ then $\pi^{-1}(y)$ is the complex projective space associated to the normal space $T_yX/T_yY$ to $Y$ in $X$ at $y$. If $X = \mathbb{C}^{n+1}$ and $Y = \{0\}$ and we identify $\mathbb{P}_n$ with the set of one-dimensional linear subspaces of $\mathbb{C}^{n+1}$, then

$$\tilde{X} = \{(v, w) \in \mathbb{C}^{n+1} \times \mathbb{P}_n : v \in w\}$$

with $\pi(v, w) = v$.

Again this problem does not arise when $n = 1$, because blowing up a 1-fold makes no difference unless the 1-fold has singularities (in which case blowing up may help to "resolve" the singularities; for example, when we blow up the origin $\{0\}$ in $\mathbb{C}^2$, then the singular curve $C$ in $\mathbb{C}^2$ defined by $y^2 = x^3 + x^2$ is tranformed into a nonsingular curve $\tilde{C}$ with the origin in $C$ replaced by two points, corresponding to the two complex "tangent directions" in $C$ at 0).

Thus, the classification of $n$-folds when $n > 1$ requires a preliminary step before there is any hope of carrying out the two steps described above.

Step 0 (the "minimal model programme" of Mori (1987) and others): Instead of all the objects to be classified, consider only specially "good" objects, such that every object is obtained from one of these specially good objects by a sequence of blow-ups (or similar carefully prescribed operations).

How to carry out Mori's minimal model program is well understood for algebraic surfaces and 3-folds, but in higher dimensions is incomplete as yet (Kollár and Mori 1998). We shall ignore both step 0 and step 1 from now on, and concentrate on step 2, the construction of moduli spaces.

## Ingredients of a Moduli Problem

Formally before posing a moduli problem, we need to fix the category in which we are working; that is, we need to specify what we mean by "space" and "map" in the description below. If, for example, we are working in complex analytic geometry then we might take "space" to mean a complex manifold (or more generally we might allow singularities) and take "map" to mean a complex analytic map, whereas in algebraic geometry "space" might mean an algebraic variety, or a scheme, or even a stack, with "map" interpreted as a morphism of algebraic varieties (or schemes, or stacks).

Once this is fixed, the ingredients of a moduli problem are:

1. a set $A$ of objects to be classified,
2. an equivalence relation $\sim$ on $A$,
3. the concept of a family of objects in $A$ with base space $S$ (or parametrized by $S$), and sometimes
4. the concept of equivalence of families.

These ingredients must satisfy:

1. a family parametrized by a single point $\{p\}$ is just an object in $A$ (and equivalence of objects is equivalence of families over $\{p\}$) and
2. given a family $X$ parametrized by a space $S$ and a map $\phi : \tilde{S} \to S$, there is a family $\phi^* X$ parametrized by $\tilde{S}$ (the "pullback of $X$ along $\phi$"), with pullback being functorial and preserving equivalence.

In particular, for any family $X$ parametrized by $S$ and any $s \in S$, there is an object $X_s$ given by pulling back $X$ along the inclusion of $\{s\}$ in $S$. We think of $X_s$ as the object in the family $X$ whose parameter is the point $s$ in the base space $S$.

**Example 1** A family of compact Riemann surfaces parametrized by a complex manifold $S$ is a surjective holomorphic map

$$\pi : T \to S$$

from a complex manifold $T$ of (complex) dimension $\dim(T) = \dim(S) + 1$ to $S$, such that $\pi$ is proper (i.e., the inverse image $\pi^{-1}(C)$ of any compact subset $C$ of $S$ under $\pi$ is compact) and has maximal rank (i.e., its derivative is everywhere surjective). Then $\pi^{-1}(s)$ is a compact Riemann surface for each $s \in S$, and is the object in the family with parameter $s$.

The family defined by $\pi$ is an algebraic family if $\pi$ is a morphism of nonsingular complex projective varieties.

**Example 2** A family of nonsingular complex projective varieties parametrized by a nonsingular complex variety $S$ is a proper surjective morphism

$$\pi : T \to S$$

with $T$ nonsingular and $\pi$ having maximal rank. We can also allow $T$ and $S$ to be singular, but then we require an extra technical condition (that $\pi$ must be flat with reduced fibers).

In the above example, equivalence of families $\pi_1 : T_1 \to S_1$ and $\pi_2 : T_2 \to S_2$ is given by isomorphisms $f : T_1 \to T_2$ and $g : S_1 \to S_2$ such that $g \circ \pi_1 = \pi_2 \circ f$. Equivalence of families in the first example is similar.

**Definition 1** A "deformation" of a nonsingular projective variety or compact complex manifold $M$ is given by a family $\pi : T \to S$ together with an isomorphism

$$\pi^{-1}(s_0) \cong M$$

for some $s_0 \in S$.

Strictly speaking, the deformation is the germ at $s_0$ of such a $\pi$; that is, the restriction of $\pi$ over any open neighborhood of $s_0$ in $S$ determines the same deformation of $M$ as $\pi$ does.

A study of deformations leads to information about the local structure of moduli spaces. Let $\pi : X \to S$ be a deformation of a compact complex manifold $M = \pi^{-1}(s_0)$ where $s_0 \in S$. We can cover $M$ (thought of as a subset of $X$) with open subsets $W_i$ of $X$ such that there exist isomorphisms

$$h_i : W_i \to U_i \times V_i$$

where $V_i = \pi(W_i)$ is open in $S$ and $U_i = M \cap W_i$ is open in $M = \pi^{-1}(s_0)$ and the projection of $h_i$ onto $V_i$ is just $\pi : W_i \to V_i$. For each $i \neq j$, we then get a holomorphic vector field $\theta_{ij}$ on $U_i \cap U_j$ by differentiating $h_i \circ h_j^{-1}$ in the direction of any tangent vector $v \in T_{s_0}S$. These holomorphic vector fields define a 1-cocycle in the tangent sheaf $\Theta$ of $M$. This gives us the "Kodaira–Spencer map"

$$\rho_\pi : T_{s_0}S \to H^1(M, \Theta)$$

**Theorem 1** (Kuranishi). *If $M$ is a compact complex manifold, then it has a deformation $\pi : X \to S$ with $\pi^{-1}(s_0) = M$ such that*

(i) *the Kodaira–Spencer map $\rho_\pi : T_{s_0}S \to H^1(M, \Theta)$ is an isomorphism,*
(ii) *$\pi$ has the local universal property for deformations (i.e., any deformation of $M$ is locally the pullback of $\pi$ along a map $f$ into $S$),*
(iii) *if $H^0(M, \Theta) = 0$, then the map $f$ in (ii) is unique, and*
(iv) *if $H^2(M, \Theta) = 0$, then $S$ is nonsingular at $s_0$ and so $\dim S = \dim H^1(M, \Theta)$.*

This deformation $\pi$ is called the "Kuranishi deformation" of $M$ (its germ at $s_0$ is unique up to isomorphism), and $S$ is called the "Kuranishi space" of $M$.

**Example 3** A family of holomorphic (or algebraic) vector bundles over a compact Riemann surface (or nonsingular complex projective curve) $\Sigma$ is a vector bundle over $\Sigma \times S$ where $S$ is the base space (see e.g., Verdier and Le Potier (1985)). A deformation of a vector bundle $E_0$ over $\Sigma$ is then given by a vector bundle $E$ over a product $\Sigma \times S$ together with an isomorphism

$$E|_{\Sigma \times \{s_0\}} \cong E_0$$

for some $s_0 \in S$ (strictly speaking it is the germ at $s_0$ of such a family of vector bundles).

## Fine and Coarse Moduli Spaces

For definiteness, except when it is specified otherwise, let us consider moduli problems in algebraic geometry with "space" meaning algebraic variety (over some fixed field $k$ which is usually $\mathbb{C}$) and "map" meaning morphism of algebraic varieties.

**Definition 2** A "fine moduli space" for a given (algebro-geometric) moduli problem is an algebraic variety $M$ with a family $U$ parametrized by $M$ having the following (universal) property: for every family $X$ parametrized by a base space $S$, there exists a unique map $\phi: S \to M$ such that

$$X \sim \phi^* U$$

$U$ is then called a "universal family" for the given moduli problem.

Many moduli problems have no fine moduli space, but nonetheless there may be a moduli space satisfying slightly weaker conditions, called a coarse moduli space. If a fine moduli space does exist, it will automatically satisfy the conditions to be a coarse moduli space. Both fine and coarse moduli spaces, when they exist, are unique up to canonical isomorphism.

**Definition 3** A "coarse moduli space" for a given moduli problem is an algebraic variety $M$ with a bijection

$$\alpha : A/\sim \to M$$

(where as before $A$ is the set of objects to be classified up to the equivalence relation $\sim$) from the set $A/\sim$ of equivalence classes in $A$ to $M$ such that:

(i) For every family $X$ with base space $S$, the composition of the given bijection $\alpha: A/\sim \to M$ with the function

$$\nu_X : S \to A/\sim$$

which sends $s \in S$ to the equivalence class $[X_s]$ of the object $X_s$ with parameter $s$ in the family $X$, is a morphism.

(ii) When $N$ is any other variety with $\beta: A/\sim \to N$ such that for each family $X$ parametrized by a base space $S$ the composition $\beta \circ \nu_X : S \to N$ is a morphism, then

$$\beta \circ \alpha^{-1} : M \to N$$

is a morphism.

**Remark 3** For some moduli problems, a family $X$ with base space $S$ which is connected and of dimension strictly greater than zero may exist such that for some $s_0 \in S$ we have

(i) $X_s \sim X_t$ for all $s, t \in S - \{s_0\}$ and
(ii) $X_s \not\sim X_{s_0}$ for all $s \in S - \{s_0\}$.

This is the "jump phenomenon," and when it occurs we cannot construct a moduli space including the equivalence class of the object $X_{s_0}$. Typically, to construct a moduli space, some objects (often called "unstable") must be left out because of the jump phenomenon and we only get a moduli space of "stable" objects. This happens, for example, in the construction of moduli spaces of complex projective curves, if we want to include singular curves, or moduli spaces of vector bundles.

**Example 4** The Jacobian $J(\Sigma)$ of a compact Riemann surface $\Sigma$ is a fine moduli space for holomorphic line bundles (i.e., vector bundles of rank 1) of fixed degree over $\Sigma$ up to isomorphism. As a complex manifold

$$J(\Sigma) \cong \mathbb{C}^g / \Lambda$$

where $g$ is the genus of $\Sigma$ and $\Lambda$ is a lattice of maximal rank in $\mathbb{C}^g$ (in other words $J(\Sigma)$ is a complex torus). Since $J(\Sigma)$ is also a complex projective variety, it is an "abelian variety."

More precisely, $J(\Sigma)$ is the quotient of the complex vector space $H^0(\Sigma, K_\Sigma)$ of dimension $g$ by the lattice $H^1(\Sigma, Z) \cong Z^{2g}$. Here $K_\Sigma$ is the complex cotangent bundle of $\Sigma$ and $H^0(\Sigma, K_\Sigma)$ is the space of its holomorphic sections, that is, the space of holomorphic differentials on $\Sigma$. If we choose a basis $\omega_1, \ldots, \omega_g$ of holomorphic differentials and a standard basis $\gamma_1, \ldots, \gamma_{2g}$ for $H_1(\Sigma, \mathbb{Z})$ such that

$$\gamma_i . \gamma_{i+g} = 1 = -\gamma_{i+g} . \gamma_i$$

when $1 \le i \le g$ and all other intersection pairings $\gamma_i . \gamma_j$ are zero, then we can associate to $\Sigma$ the $g \times 2g$ "period matrix" $P(\Sigma)$ given by integrating the holomorphic differentials $\omega_i$ around the 1-cycles $\gamma_j$. The Jacobian $J(\Sigma)$ can then be identified with the quotient of $\mathbb{C}^g$ by the lattice spanned by the columns of this period matrix.

We can in fact always choose the basis $\omega_1, \ldots, \omega_g$ of holomorphic differentials so that the period matrix $P(\Sigma)$ is of the form

$$(I_g \; Z)$$

where $I_g$ is the $g \times g$ identity matrix. This period matrix is called a "normalized period matrix." The Riemann bilinear relations tell us that $Z$ is symmetric and its imaginary part is positive definite.

**Example 5** The moduli space $\mathcal{A}_g$ of all abelian varieties of dimension $g$ was one of the first moduli spaces to be constructed. We have

$$\mathcal{A}_g \cong \mathcal{H}_g/\mathrm{Sp}(2g;\mathbb{Z})$$

where $\mathcal{H}_g$ is Siegel's upper half space, which consists of the symmetric $g \times g$ complex matrices with positive-definite imaginary part.

**Example 6** One way to construct and study the moduli space $\mathcal{M}_g$ of compact Riemann surfaces of genus $g$ is via the "Torelli map"

$$\tau : \mathcal{M}_g \to \mathcal{A}_g$$

given by

$$\Sigma \mapsto J(\Sigma)$$

Torelli's theorem tells us that $\tau$ is injective (cf. Griffiths (1984)). Describing the image of $\mathcal{M}_g$ in $\mathcal{A}_g$ is known as the Schottky problem.

We can calculate the dimension of the moduli space $\mathcal{M}_g$ using Kuranishi theory as in the previous section: we get

$$\dim \mathcal{M}_g = \dim H^1(\Sigma, \Theta) = 3g - 3$$

for any compact Riemann surface $\Sigma$ of genus $g \geq 2$. In fact, if $M$ is any compact complex manifold and there exists a fine moduli space of complex manifolds diffeomorphic to $M$, then the moduli space is locally isomorphic near $[M]$ to the Kuranishi space near $s_0$. More often, there is only a coarse moduli space (as in the case of $\mathcal{M}_g$), and then the moduli space is locally isomorphic near $[M]$ to the quotient of the Kuranishi space by the action of the group of automorphisms of $M$.

For the Teichmüller approach to $\mathcal{M}_g$ (cf. Lehto (1987)), we consider the space of all pairs consisting of a compact Riemann surface of genus $g$ and a basis $\gamma_1, \ldots, \gamma_{2g}$ for $H_1(\Sigma, \mathbb{Z})$ as above such that

$$\gamma_i \cdot \gamma_{i+g} = 1 = -\gamma_{i+g} \cdot \gamma_i$$

if $1 \leq i \leq g$ and all other intersection pairings $\gamma_i \cdot \gamma_j$ are zero. If $g \geq 2$, this space (called Teichmüller space) is naturally homeomorphic to an open ball in $\mathbb{C}^{3g-3}$ (by a theorem of Bers). The mapping class group $\Gamma_g$ (which consists of the diffeomorphisms of the surface modulo isotopy) acts discretely on Teichmüller space, and the quotient can be identified with the moduli space $\mathcal{M}_g$. This gives us a description of $\mathcal{M}_g$ as a complex analytic space, but not as an algebraic variety.

To construct the moduli space $\mathcal{M}_g$ as an algebraic variety, we can use the fact that every compact Riemann surface of genus $g$ can be embedded canonically as a curve of degree $6(g-1)$ in a projective space of dimension $5g-6$. The use of the word "canonical" here is a rather poor pun; it refers both to the canonical line bundle (or cotangent bundle) of the Riemann surface, although here "tricanonical" would be more accurate, and also to the fact that no choices are involved, except that a choice of basis is needed to identify the projective space with the standard one $P_{5g-6}$. This enables us to identify $\mathcal{M}_g$ with the quotient of an algebraic variety by the group $\mathrm{PGL}(n+1;\mathbb{C})$. However, here we do not have a discrete group action, and to construct the quotient we must use Mumford's geometric invariant theory (see below), which was developed in the 1960s in order to provide algebraic constructions of this moduli space and others.

In fact, geometric invariant theory also provides a beautiful compactification of $\mathcal{M}_g$ known as the Deligne–Mumford (1969) compactification $\bar{\mathcal{M}}_g$. This compactification is itself a moduli space: it is the moduli space of (Deligne–Mumford) stable curves, which are complex projective curves with only nodal singularities and at most finitely many automorphisms. $\bar{\mathcal{M}}_g$ is singular but in a relatively mild way; it is the quotient of a nonsingular variety by a finite group action.

The moduli space $\mathcal{M}_{g,n}$ of nonsingular complex projective curves of genus $g$ with $n$ marked points has a similar compactification $\bar{\mathcal{M}}_{g,n}$ which is the moduli space of complex projective curves with $n$ marked nonsingular points and with only nodal singularities and finitely many automorphisms. Finiteness of the automorphism group of such a curve $\Sigma$ is equivalent to the requirement that any irreducible component of genus 0 (respectively 1) has at least 3 (respectively 1) special points, where "special" means either marked or singular in $\Sigma$.

The construction of $\mathcal{M}_g$ using the period matrices of curves and the Torelli theorem leads to a different compactification $\tilde{\mathcal{M}}_g$ of $\mathcal{M}_g$ known as the Satake (or Satake–Baily–Borel) compactification. Like the Deligne–Mumford compactification, $\tilde{\mathcal{M}}_g$ is a complex projective variety, but the boundary of $\mathcal{M}_g$ in $\tilde{\mathcal{M}}_g$ has (complex) codimension 2 for $g \geq 3$ whereas the boundary $\Delta$ of $\mathcal{M}_g$ in $\bar{\mathcal{M}}_g$ has codimension 1. Each of the irreducible components $\Delta_0, \ldots, \Delta_{[g/2]}$ of $\Delta$ is the closure of a locus of curves with exactly one node (irreducible curves with one node in the case of $\Delta_0$, and in the case of any other $\Delta_i$ the union of two nonsingular curves of genus $i$ and $g-i$ meeting at a single point). The divisors $\Delta_i$ meet transversely in $\bar{\mathcal{M}}_g$, and their intersections define a natural decomposition of $\Delta$ into connected strata which parametrize stable curves of a fixed topological type.

For a recent guide to many different aspects of the moduli spaces $\mathcal{M}_g$, see Harris and Morrison (1998).

**Example 7** Given any nonsingular complex projective variety $X$, we can study the moduli spaces of maps from curves to $X$ considered by Kontsevich. Intersection theory on these moduli spaces leads to Gromov–Witten theory and the quantum cohomology of $X$, with many applications, for example, to enumerative geometry (cf. Cox and Katz (1999), Fulton and Pandharipande (1997), Dijkgraaf *et al.* (1995)).

More precisely, if $2g-2+n>0$ then for any $\beta \in H_2(X;\mathbb{Z})$ there is a moduli space $\mathcal{M}_{g,n}(X,\beta)$ of $n$-pointed nonsingular complex projective curves $\Sigma$ of genus $g$ equipped with maps $f:\Sigma \to X$ satisfying $f_*[\Sigma]=\beta$. This moduli space has a compactification $\bar{\mathcal{M}}_{g,n}(X,\beta)$ which classifies "stable maps" of type $\beta$ from $n$-pointed curves of genus $g$ into $X$ (Fulton and Pandharipande 1997). Here, a map $f:\Sigma \to X$ from an $n$-pointed complex projective curve $\Sigma$ satisfying $f_*[\Sigma]=\beta$ is called stable if $\Sigma$ has only nodal singularities and $f:\Sigma \to X$ has only finitely many automorphisms, or equivalently every irreducible component of $\Sigma$ of genus 0 (respectively genus 1) which is mapped to a single point in $X$ by $f$ contains at least three (respectively 1) special points. The forgetful map from $\mathcal{M}_{g,n}(X,\beta)$ to $\mathcal{M}_{g,n}$ which sends $[\Sigma,p_1,\ldots,p_n,f:\Sigma\to X]$ to $[\Sigma,p_1,\ldots,p_n]$ extends to a forgetful map $\pi:\bar{\mathcal{M}}_{g,n}(X,\beta)\to\bar{\mathcal{M}}_{g,n}$ which collapses components of $\Sigma$ with genus 0 and at most two special points.

Of course, when $X$ is itself a single point, $\mathcal{M}_{g,n}(X,\beta)$ and $\bar{\mathcal{M}}_{g,n}(X,\beta)$ are simply the moduli spaces $\mathcal{M}_{g,n}$ and $\bar{\mathcal{M}}_{g,n}$. In general $\bar{\mathcal{M}}_{g,n}(X,\beta)$ has more serious singularities than $\bar{\mathcal{M}}_{g,n}$ and may indeed have many different irreducible components with different dimensions. In spite of this, $\bar{\mathcal{M}}_{g,n}(X,\beta)$ has a "virtual fundamental class" $[\bar{\mathcal{M}}_{g,n}(X,\beta)]^{\mathrm{vir}}$ lying in the expected dimension

$$3g-3+n+(1-g)\dim X+\int_\beta c_1(TX)$$

of $\bar{\mathcal{M}}_{g,n}(X,\beta)$. Gromov–Witten invariants (originally developed mainly in the case $g=0$ when $\bar{\mathcal{M}}_{g,n}(X,\beta)$ is more tractable, but now also studied when $g>0$) are obtained by evaluating cohomology classes on $\bar{\mathcal{M}}_{g,n}(X,\beta)$ against this virtual fundamental class.

## Moduli Spaces as Orbit Spaces

**Example 8** As a simple example, let us consider the moduli space of "hyperelliptic" curves of genus $g$. By a hyperelliptic curve of genus $g$, we mean a nonsingular complex projective curve $C$ with a double cover $f:C\to\mathbb{P}_1$ branched over $2g+2$ points in the complex projective line $\mathbb{P}_1$.

Let $S$ be the set of unordered sequences of $2g+2$ distinct points in $\mathbb{P}_1$, which we can identify with an open subset of the complex projective space $\mathbb{P}_{2g+2}$ by associating to an unordered sequence $a_1,\ldots,a_{2g+2}$ of points in $\mathbb{P}_1$ the coefficients of the polynomial whose roots are $a_1,\ldots,a_{2g+2}$. Then, it is not hard to construct a family $\mathcal{X}$ of hyperelliptic curves of genus $g$ with base space $S$ such that the curve parametrized by $a_1,\ldots,a_{2g+2}$ is a double cover of $\mathbb{P}_1$ branched over $a_1,\ldots,a_{2g+2}$. This family is not quite a universal family, but it does have the following two properties.

(i) The hyperelliptic curves $\mathcal{X}_s$ and $\mathcal{X}_t$ parametrized by elements $s$ and $t$ of the base space $S$ are isomorphic if and only if $s$ and $t$ lie in the same orbit of the natural action of $G=\mathrm{SL}(2;\mathbb{C})$ on $S$.
(ii) (Local universal property) Any family of hyperelliptic curves of genus $g$ is locally equivalent to the pullback of $\mathcal{X}$ along a morphism to $S$.

These properties (i) and (ii) imply that a (coarse) moduli space $M$ exists if and only if there is an "orbit space" for the action of $G$ on $S$ (Newstead 1978). Here, by an orbit space we mean a $G$-invariant morphism $\phi:S\to M$ such that every other $G$-invariant morphism $\psi:S\to M$ factors uniquely through $\phi$, and moreover $\phi^{-1}(m)$ is a single $G$-orbit for each $m\in M$. (We can think of an orbit space as the set of $G$-orbits endowed in a natural way with the structure of an algebraic variety.)

This sort of situation arises quite often in moduli problems, and the construction of a moduli space is then reduced to the construction of an orbit space. Unfortunately, such orbit spaces do not in general exist. The main problem (which is closely related to the jump phenomenon discussed above) is that there may be orbits contained in the closures of other orbits, which means that the natural topology on the set of all orbits is not Hausdorff, so this set cannot be endowed naturally with the structure of a variety. This is the situation the geometric invariant theory of Mumford (1965) attempts to deal with, telling us how to throw out certain "unstable" orbits in order to be able to construct an orbit space. For more general constructions of orbit spaces which can be used for moduli problems where geometric invariant theory may not be of use, see Keel and Mori (1997) and Kollár (1997).

**Example 9** Let $G=\mathrm{SL}(2;\mathbb{C})$ act on $(\mathbb{P}_1)^4$ via Möbius transformations on the Riemann sphere

$$\mathbb{P}_1=\mathbb{C}\cup\{\infty\}$$

Then,

$$\{(x_1, x_2, x_3, x_4) \in (\mathbb{P}_1)^4 : x_1 = x_2 = x_3 = x_4\}$$

is a single orbit which is contained in the closure of every other orbit. On the other hand, the open subset

$$\{(x_1, x_2, x_3, x_4) \in (\mathbb{P}_1)^4 : x_1, x_2, x_3, x_4 \text{ distinct}\}$$

of $(\mathbb{P}_1)^4$ has an orbit space which can be identified with

$$\mathbb{P}_1 - \{0, 1, \infty\}$$

via the cross ratio.

In order to describe Mumford's geometric invariant theory, let $X$ be a complex projective variety (i.e., a subset of a complex projective space defined by the vanishing of homogeneous polynomial equations), and let $G$ be a complex reductive group acting on $X$. We also require a "linearization" of the action; that is, an ample line bundle $L$ on $X$ and a lift of the action of $G$ to $L$. We lose very little generality in assuming that for some projective embedding $X \subseteq \mathbb{P}_n$ the action of $G$ on $X$ extends to an action on $\mathbb{P}_n$ given by a representation

$$\rho : G \to \mathrm{GL}(n+1)$$

and taking for $L$ the hyperplane line bundle on $\mathbb{P}_n$. Algebraic geometry associates to $X \subseteq \mathbb{P}_n$ its homogeneous coordinate ring

$$A(X) = \bigoplus_{k \geq 0} H^0(X, L^{\otimes k}) = \mathbb{C}[x_0, \ldots, x_n]/\mathcal{I}_X$$

which is the quotient of the polynomial ring $\mathbb{C}[x_0, \ldots, x_n]$ in $n+1$ variables by the ideal $\mathcal{I}_X$ generated by the homogeneous polynomials vanishing on $X$. Since the action of $G$ on $X$ is given by a representation $\rho : G \to \mathrm{GL}(n+1)$, we get an induced action of $G$ on $\mathbb{C}[x_0, \ldots, x_n]$ and on $A(X)$, and we can therefore consider the subring $A(X)^G$ of $A(X)$ consisting of the elements of $A(X)$ left invariant by $G$. This subring $A(X)^G$ is a graded complex algebra, and because $G$ is reductive it is finitely generated (Mumford 1965). To any finitely generated graded complex algebra we can associate a complex projective variety, and so we can define $X//G$ to be the variety associated to the ring of invariants $A(X)^G$. The inclusion of $A(X)^G$ in $A(X)$ defines a "rational" map $\phi$ from $X$ to $X//G$, but because there may be points of $X \subseteq \mathbb{P}_n$ where every $G$-invariant polynomial vanishes, this map will not in general be well defined everywhere on $X$ (i.e., it will not be a morphism).

We define the set $X^{ss}$ of "semistable" points in $X$ to be the set of those $x \in X$ for which there exists some $f \in A(X)^G$ not vanishing at $x$. Then, the rational map $\phi$ restricts to a surjective $G$-invariant morphism from the open subset $X^{ss}$ of $X$ to the quotient variety $X//G$. However, $\phi : X^{ss} \to X//G$ is still not in general an orbit space: when $x$ and $y$ are semistable points of $X$, we have $\phi(x) = \phi(y)$ if and only if the closures $\overline{O_G(x)}$ and $\overline{O_G(y)}$ of the $G$-orbits of $x$ and $y$ meet in $X^{ss}$. Topologically, $X//G$ is the quotient of $X^{ss}$ by the equivalence relation for which $x$ and $y$ in $X^{ss}$ are equivalent if and only if $\overline{O_G(x)}$ and $\overline{O_G(y)}$ meet in $X^{ss}$.

We define a "stable" point of $X$ to be a point $x$ of $X^{ss}$ with a neighbourhood in $X^{ss}$ such that every $G$-orbit meeting this neighborhood is closed in $X^{ss}$, and is of maximal dimension equal to the dimension of $G$. If $U$ is any $G$-invariant open subset of the set $X^s$ of stable points of $X$, then $\phi(U)$ is an open subset of $X//G$ and the restriction $\phi|_U : U \to \phi(U)$ of $\phi$ to $U$ is an orbit space for the action of $G$ on $U$ in the sense described above, so that it makes sense to write $U/G$ for $\phi(U)$. In particular, there is an orbit space $X^s/G$ for the action of $G$ on $X^s$, and $X//G$ can be thought of as a compactification of this orbit space.

$$\begin{array}{ccccc} X^s & \underset{\text{open}}{\subseteq} & X^{ss} & \underset{\text{open}}{\subseteq} & X \\ \downarrow & & \downarrow & & \\ X^s/G & \underset{\text{open}}{\subseteq} & X^{ss}/\sim & = & X//G \end{array}$$

**Example 10** Let us return to hyperelliptic curves of genus $g$. We have seen that the construction of a moduli space reduces to the construction of an orbit space for the action of $G = \mathrm{SL}(2; \mathbb{C})$ on an open subset $S$ of $\mathbb{P}_{2g+2}$. If we identify $\mathbb{P}_{2g+2}$ with the space of unordered sequences of $2g+2$ points in $\mathbb{P}_1$, then $S$ is the subset consisting of unordered sequences of distinct points. When the action of $G$ on $\mathbb{P}_{2g+2}$ is linearized in the obvious way, then an unordered sequence of $2g+2$ points in $\mathbb{P}_1$ is semistable if and only if at most $g+1$ of the points coincide anywhere on $\mathbb{P}_1$, and is stable if and only if at most $g$ of the points coincide anywhere on $\mathbb{P}_1$ (cf. Kirwan (1985), chapter 16). Thus, $S$ is an open subset of $\mathbb{P}^s_{2g+2}$, so an orbit space $S/G$ exists with compactification the projective variety $\mathbb{P}_{2g+2}//G$. This orbit space is then the moduli space of hyperelliptic curves of genus $g$.

Other moduli spaces (such as moduli spaces of curves and of vector bundles; see e.g., Donaldson (1984), Gieseker (1983), Mumford (1965, 1977), and Newstead (1978)) can be constructed as orbit spaces via geometric invariant theory in a similar way.

## Symplectic Reduction and Moduli Spaces of Vector Bundles

Geometric invariant theoretic quotients are closely related to the process of reduction in symplectic geometry, and thus many moduli spaces can be described as symplectic reductions.

Suppose that a compact, connected Lie group $K$ with Lie algebra $\boldsymbol{k}$ acts smoothly on a symplectic manifold $X$ and preserves the symplectic form $\omega$. Let us denote the vector field on $X$ defined by the infinitesimal action of $a \in \boldsymbol{k}$ by

$$x \mapsto a_x$$

By a moment map for the action of $K$ on $X$ we mean a smooth map

$$\mu : X \to \boldsymbol{k}^*$$

which satisfies

$$\mathrm{d}\mu(x)(\xi).a = \omega_x(\xi, a_x)$$

for all $x \in X$, $\xi \in T_x X$ and $a \in \boldsymbol{k}$. In other words, if $\mu_a : X \to \mathbb{R}$ denotes the component of $\mu$ along $a \in \boldsymbol{k}$ defined for all $x \in X$ by the pairing

$$\mu_a(x) = \mu(x).a$$

between $\mu(x) \in \boldsymbol{k}^*$ and $a \in \boldsymbol{k}$, then $\mu_a$ is a Hamiltonian function for the vector field on $X$ induced by $a$. We shall assume that all our moment maps are equivariant moment maps; that is, $\mu : X \to \boldsymbol{k}^*$ is $K$-equivariant with respect to the given action of $K$ on $X$ and the co-adjoint action of $K$ on $\boldsymbol{k}^*$.

It follows directly from the definition of a moment map $\mu : X \to \boldsymbol{k}^*$ that if the stabilizer $K_\zeta$ of any $\zeta \in \boldsymbol{k}^*$ acts freely on $\mu^{-1}(\zeta)$, then $\mu^{-1}(\zeta)$ is a submanifold of $X$ and the symplectic form $\omega$ induces a symplectic structure on the quotient $\mu^{-1}(\zeta)/K_\zeta$. With this symplectic structure, the quotient $\mu^{-1}(\zeta)/K_\zeta$ is called the Marsden–Weinstein reduction, or symplectic quotient, at $\zeta$ of the action of $K$ on $X$. We can also consider the quotient $\mu^{-1}(\zeta)/K_\zeta$ when the action of $K_\zeta$ on $\mu^{-1}(\zeta)$ is not free, but in this case it is likely to have singularities.

**Example 11** Consider the cotangent bundle $T^*Y$ of any $n$-dimensional manifold $Y$ with its canonical symplectic form $\omega$ which is given by the standard symplectic form

$$\omega = \sum_{j=1}^{n} \mathrm{d}p_j \wedge \mathrm{d}q_j \qquad [1]$$

with respect to any local coordinates $(q_1, \ldots, q_n)$ on $Y$ and the induced coordinates $(p_1, \ldots, p_n)$ on its cotangent spaces. If $Y$ is the configuration space of a classical mechanical system, then $T^*Y$ is the phase space of the system and the coordinates $p = (p_1, \ldots, p_n) \in T*_q Y$ are traditionally called the momenta of the system.

If $Y$ is acted on by a Lie group $K$, the induced action on $T^*Y$ preserves $\omega$ and there is a moment map $\mu : T^*Y \to \boldsymbol{k}^*$ whose components $\mu_a$ along $a \in \boldsymbol{k}$ are given by pairing the moment coordinates $p$ with the vector fields on $X$ induced by the infinitesimal action of $K$; that is,

$$\mu_a(p, q) = p.a_q$$

for all $q \in Y$ and $p \in T_q Y$. When $K = \mathrm{SO}(3)$ acts by rotations on $Y = \mathbb{R}^3$, then $\mu$ is the angular momentum, or moment of momentum, about the origin.

The connection with geometric invariant theory arises as follows. Let $X$ be a nonsingular complex projective variety embedded in complex projective space $\mathbb{P}_n$, and let $G$ be a complex Lie group acting on $X$ via a complex linear representation $\rho : G \to \mathrm{GL}(n+1; \mathbb{C})$. A necessary and sufficient condition for $G$ to be reductive is that it is the complexification of a maximal compact subgroup $K$ (e.g., $G = \mathrm{GL}(m; \mathbb{C})$ is the complexification of the unitary group $\mathrm{U}(m)$). By an appropriate choice of coordinates on $\mathbb{P}_n$, we may assume that $\rho$ maps $K$ into the unitary group $\mathrm{U}(n+1)$. Then, the action of $K$ preserves the Fubini–Study form $\omega$ on $\mathbb{P}_n$, which restricts to a symplectic form on $X$. There is a moment map $\mu : X \to \boldsymbol{k}^*$ defined (up to multiplication by a constant scalar factor depending on differences in convention on the normalization of the Fubini–Study form) by

$$\mu(x).a = \frac{\overline{\hat{x}}^t \rho_*(a)\hat{x}}{2\pi \mathrm{i} ||\hat{x}||^2} \qquad [2]$$

for all $a \in \boldsymbol{k}$, where $\hat{x} \in \mathbb{C}^{n+1} - \{0\}$ is a representative vector for $x \in \mathbb{P}_n$ and the representation $\rho : K \to \mathrm{U}(n+1)$ induces $\rho_* : \boldsymbol{k} \to \boldsymbol{u}(n+1)$ and dually $\rho^* : \boldsymbol{u}(n+1)^* \to \boldsymbol{k}^*$.

In this situation, we have two possible quotient constructions, giving us the geometric invariant theory quotient $X//G$ if we want to work in algebraic geometry and the symplectic reduction $\mu^{-1}(0)/K$ if we want to work in symplectic geometry. In fact, these give us the same quotient space, at least up to homeomorphism (and diffeomorphism away from the singularities). More precisely, any $x \in X$ is semistable if and only if the closure of its $G$-orbit meets $\mu^{-1}(0)$, and the inclusion of $\mu^{-1}(0)$ into $X^{\mathrm{ss}}$ induces a homeomorphism

$$\mu^{-1}(0)/K \to X//G$$

There are other quotient constructions closely related to symplectic reduction and geometric invariant

theory, which are useful when working with Kähler or hyper-Kähler manifolds.

In physics, moduli spaces are often described as symplectic reductions of infinite-dimensional symplectic manifolds by infinite-dimensional groups (although the moduli spaces themselves are usually finite-dimensional). One example is given by moduli spaces of holomorphic vector bundles, which can also be described using Yang–Mills theory (cf. Atiyah and Bott (1982)).

The Yang–Mills equations arose in physics as generalizations of Maxwell's equations. They have become important in differential and algebraic geometry formulated over arbitrary compact oriented Riemannian manifolds, and in particular over compact Riemann surfaces and higher dimensional Kähler manifolds. The fundamental theorem of Donaldson, Uhlenbeck, and Yau that a holomorphic bundle over a compact Kähler manifold admits an irreducible Hermitian Yang–Mills connection if and only if it is stable can be thought of as an infinite-dimensional illustration of the link between symplectic reduction and geometric invariant theory.

Let $M$ be a compact oriented Riemannian manifold and let $E$ be a fixed complex vector bundle over $M$ with a Hermitian metric. Recall that a connection $A$ on $E$ (or equivalently on its frame bundle) can be defined by a covariant derivative $d_A : \Omega^p_M(E) \to \Omega^{p+1}_M(E)$, where $\Omega^p_M(E)$ denotes the space of $C^\infty$-sections of $\bigwedge^p T^*M \otimes E$ (i.e., the space of $p$-forms on $M$ with values in $E$). This covariant derivative satisfies the extended Leibniz rule

$$d_A(\alpha \wedge \beta) = (d_A \alpha) \wedge \beta + (-1)^p \alpha \wedge d_A \beta$$

for $\alpha \in \Omega^p_M(E), \beta \in \Omega^q_M(E)$, and therefore is determined by its restriction $d_A : \Omega^0_M(E) \to \Omega^1_M(E)$. The Leibniz rule implies that the difference of two connections is given by an $E \otimes E^*$-valued 1-form on $M$, and hence that the space of all connections on $E$ is an infinite-dimensional affine space $\mathcal{A}$ based on the vector space $\Omega^1_M(E \otimes E^*)$. Similarly, the space of all unitary connections on $E$ (i.e., connections compatible with the Hermitian metric on $E$) is an infinite-dimensional affine space based on the space of 1-forms with values in the bundle $\mathfrak{g}_E$ of skew-adjoint endomorphisms of $E$. The Leibniz rule also implies that the composition $d_A \circ d_A : \Omega^0_M(E) \to \Omega^2_M(E)$ commutes with multiplication by smooth functions, and thus we have

$$d_A \circ d_A(s) = F_A s$$

for all $C^\infty$ sections $s$ of $E$, where $F_A \in \Omega^2_M(\mathfrak{g}_E)$ is defined to be the curvature of the unitary connection $A$. The Yang–Mills functional on the space $\mathcal{A}$ of all unitary connections on $E$ is defined as the $L^2$-norm square of the curvature, given by the integral over $M$ of the product of the function $\|F_A\|^2$ and the volume form on $M$ defined by the Riemannian metric and the orientation. The Yang–Mills equations are the Euler–Lagrange equations for this functional, given by

$$d_A * F_A = 0$$

where $d_A$ has been extended in a natural way to $\Omega^*_M(\mathfrak{g}_E)$. The gauge group $\mathcal{G}$, that is, the group of unitary automorphisms of $E$, preserves the Yang–Mills functional and the Yang–Mills equations.

If $M$ is a complex manifold, we can identify the space $\mathcal{A}^{(1,1)}$ of unitary connections on $E$ with curvature of type (1,1) with the space of holomorphic structures on $E$, by associating to a holomorphic structure $\mathcal{E}$ the unitary connection whose $(0,1)$-component is given by the $\bar{\partial}$-operator defined by $\mathcal{E}$. This space $\mathcal{A}^{(1,1)}$ is an infinite-dimensional complex subvariety of the infinite-dimensional complex affine space $\mathcal{A}$, acted on by the complexified gauge group $\mathcal{G}_c$ (the group of complex $C^\infty$ automorphisms of $E$), and two holomorphic structures are isomorphic if and only if they lie in the same $\mathcal{G}_c$-orbit.

When $(M, \omega)$ is a compact Kähler manifold, there is a $\mathcal{G}$-invariant Kähler form $\Omega$ on $\mathcal{A}$ defined by

$$\Omega(\alpha, \beta) = \frac{1}{8\pi^2} \int_M \mathrm{tr}(\alpha \wedge \beta) \wedge \omega^{n-1}$$

where $n$ is the complex dimension of $M$. The Lie algebra of $\mathcal{G}$ is the space $\Omega^0_M(\mathfrak{g}_E)$ of sections of $\mathfrak{g}_E$, and there is a moment map $\mu : \mathcal{A} \to (\Omega^0_M(\mathfrak{g}_E))^*$ for the action of $\mathcal{G}$ on $\mathcal{A}$ given by the composition of

$$A \mapsto \frac{1}{8\pi^2} F_A \wedge \omega^{n-1} \in \Omega^{2n}_M(\mathfrak{g}_E)$$

with integration over $M$. On $\mathcal{A}^{(1,1)}$ the norm square of this moment map agrees up to a constant factor with the Yang–Mills functional, which is minimized by the Hermitian Yang–Mills connections.

As in the finite-dimensional situation, for a suitable definition of stability, the moduli space of stable holomorphic bundles of topological type $E$ over $M$ (which plays the role of the geometric invariant theory quotient) can be identified with the moduli space of (irreducible) Hermitian Yang–Mills connections on $E$ (which plays the rôle of the symplectic reduction). This was proved in general for vector bundles over compact Kähler manifolds Uhlenbeck and Yau with a different proof for nonsingular complex projective varieties given by Donaldson.

Over a compact Riemann surface $M$ the situation is relatively simple, as all connections on $E$ have curvature of type $(1,1)$ and so the infinite-dimensional

complex affine space $\mathcal{A}$ can be identified with the space $\mathcal{C}$ of holomorphic structures on $E$. A moment map for the action of the gauge group on $\mathcal{A}$ is given by assigning to a connection $A \in \mathcal{A}$ its curvature $F_A \in \Omega^2_M(\mathbf{g}_E)$, and, after a suitable central constant has been added, the Hermitian Yang–Mills connections are exactly the zeros of the moment map.

A holomorphic bundle $\mathcal{E}$ over a Riemann surface $M$ is stable (respectively semistable) if $\mu(\mathcal{F}) < \mu(\mathcal{E})$ (respectively $\mu(\mathcal{F}) \leq \mu(\mathcal{E})$) for every proper subbundle $\mathcal{F}$ of $\mathcal{E}$, where

$$\mu(\mathcal{F}) = \deg(\mathcal{F})/\text{rank}(\mathcal{F})$$

When the theory of stability of holomorphic vector bundles was first introduced, Narasimhan and Seshadri proved that a holomorphic vector bundle over $M$ is stable if and only if it arises from an irreducible representation of a certain central extension of the fundamental group $\pi_1(M)$. Atiyah and Bott (1982) translated this in terms of connections to show that a holomorphic vector bundle over $M$ is stable if and only if it admits a unitary connection with constant central curvature. They deduced from this the existence of a homeomorphism between the moduli space $\mathcal{M}(n, d)$ of stable bundles of rank $n$ and degree $d$ over $M$ and the moduli space of irreducible connections with constant central curvature on a fixed $C^\infty$ bundle $E$ of rank $n$ and degree $d$ over $M$.

*See also:* BF Theories; Calibrated Geometry and Special Lagrangian Submanifolds; Cohomology Theories; Floer Homology; Gauge Theoretic Invariants of 4-Manifolds; Gauge Theory: Mathematical Applications; Geometric Measure Theory; Geometric Phases; Hamiltonian Group Actions; Instantons: Topological Aspects; Intersection Theory; Riemann Surfaces; Several Complex Variables: Basic Geometric Theory; Several Complex Variables: Compact Manifolds; Topological Gravity, Two-Dimensional; WDVV Equations and Frobenius Manifolds.

## Further Reading

Atiyah MF and Bott R (1982) The Yang–Mills equations over Riemann surfaces. *Philosophical Transactions of the Royal Society London* 308: 523–615.

Cox D and Katz S (1999) Mirror Symmetry and Algebraic Geometry. *Math Surveys and Monographs*, vol. 68. Providence, RI: American Mathematical Society.

Deligne P and Mumford D (1969) The irreducibility of the space of curves of given genus. *Publications of the Institute Hautes 'Etudes Scientifique'* 36: 75–110.

Dijkgraaf R *et al.* (eds.) (1995) *The Moduli Space of Curves*. Boston, MA: Birkhäuser.

Donaldson SK (1984) Instantons and geometric invariant theory. *Communications in Mathematical Physics* 93: 453–460.

Fulton W and Pandharipande R (1997) Notes on stable maps and quantum cohomology. In: *Algebraic Geometry*, *Santa Cruz 1995*, *Proceeding of the Symposia in Pure Mathematics* 62 vol. 2 (1997), 45–96.

Gieseker D (1983) Geometric invariant theory and applications to moduli problems. In: *Invariant Theory (Montecatini, 1982)*, Lecture Notes in Mathematics vol. 996, pp. 45–73. Berlin: Springer.

Griffiths P (1984) Topics in transcendental algebraic geometry. *Annals of Mathematical Studies*, vol. 106. Princeton: Princeton University Press.

Harris J and Morrison I (1998) Moduli of Curves. *Graduate Texts in Mathematics*, vol. 107. Berlin: Springer.

Keel S and Mori S (1997) Quotients by groupoids. *Annals of Mathematics* 145(2): 193–213.

Kirwan FC (1985) Cohomology of Quotients in Algebraic and Symplectic Geometry. *Math. Notes*. vol. 31. Princeton: Princeton University Press.

Kollár J (1997) Quotient spaces modulo algebraic groups. *Annals of Mathematics* 145(2): 33–79.

Kollár J and Mori S (1998) Birational Geometry of Algebraic Varieties. *Cambridge Tracts in Mathematics*, vol. 134. Cambridge: Cambridge University Press.

Lehto O (1987) Univalent Functions and Teichmüller Spaces. *Graduate Texts in Mathematics*. vol. 109. Berlin: Springer.

Mori S (1987) Classification of higher dimensional varieties, Algebraic Geometry, Bowdoin 1985. *Proceedings of the Symposia in Pure Mathematics* 46: 269–331.

Mumford D (1965) *Geometric Invariant Theory*, 3rd edn. with Fogarty J and Kirwin F (1994). Berlin: Springer.

Mumford D (1977) Stability of projective varieties. *L'Enseignement Mathématiques* 23: 33–110.

Newstead PE (1978) Introduction to Moduli Problems and Orbit Spaces. *Tata Institute Lecture Notes*. Berlin: Springer.

Popp H (1977) Moduli Theory and Classification Theory of Algebraic Varieties. *Springer Lecture Notes in Mathematics*, vol. 620. Berlin: Springer.

Seshadri C (1975) Theory of moduli, *Proceedings of the Symposia in Pure Mathematics* vol. 29, Algebraic Geometry. American Mathematical Society.

Sundaramanan D (1980) Moduli, Deformations and Classifications of Compact Complex Manifolds. *Pitman Research Notes in Mathematics*, vol. 45. Harlow: Longman.

Verdier J-L and Le Potier J (1985) Module des fibrés stables sur les courbes algébriques. *Progress in Mathematics*, vol. 54. Boston, MA: Birkhäuser.

Viehweg E (1995) *Quasi-Projective Moduli for Polarized Manifolds*. Berlin: Springer.

**Multicomponent Fluids** *see* Interfaces and Multicomponent Fluids

# Multi-Hamiltonian Systems

**F Magri**, Università di Milano Bicocca, Milan, Italy
**M Pedroni**, Università di Bergamo, Dalmine (BG), Italy



## Introduction

Since the late 1970s, a particular attention in the theory of integrability has been payed to systems admitting more than one Hamiltonian representation. The first examples belonged to the class of infinite-dimensional systems (i.e., partial differential equations), like the Korteweg–de Vries (KdV) equation, the Ablowitz–Kaup–Newell–Segur system, and many other soliton equations (*see* Bi-Hamiltonian Methods in Soliton Theory). It was realized soon that finite-dimensional integrable systems are also likely to possess a bi-Hamiltonian representation. Moreover, a geometric setting for the study of bi-Hamiltonian systems was established, with the introduction of the so-called bi-Hamiltonian manifolds. They are Poisson manifolds with an additional Poisson structure, fulfilling a suitable compatibility condition with the initial Poisson bracket. An important program for the study and the classification of (finite-dimensional) bi-Hamiltonian manifolds was started in the 1990s by Gelfand and Zakharevich. They pointed out that the geometry of such manifolds is extremely rich and complicated.

In this article we present the basic facts concerning the bi-Hamiltonian geometry and its relations with the theory of integrable systems, referring to Recursion Operators in Classical Mechanics in this encyclopedia for the connections with separable systems of Jacobi. In the first section we give the definitions of bi-Hamiltonian manifold and bi-Hamiltonian system, and we present some properties of the former. The next section contains three concrete examples (the Euler top, the open Toda lattice, and a stationary KdV flow) and two important classes of bi-Hamiltonian manifolds, both related to Lie algebras. This is followed by a discussion of the iterative construction of first integrals in involution for a given bi-Hamiltonian system. This procedure is particularly efficient in the case of Poisson–Nijenhuis manifolds, that is, those bi-Hamiltonian manifolds whose second Poisson structure can be obtained by composing the first one with a suitable recursion operator.

## Bi-Hamiltonian Systems

First of all, we recall some fundamental definitions from the theory of Poisson manifolds, which are the natural setting for the study of Hamiltonian systems. Let $M$ be a finite-dimensional $C^\infty$-differentiable manifold and let $C^\infty(M)$ be the space of $C^\infty$-functions from $M$ to $\mathbb{R}$. A Poisson bracket on $M$ is a skew-symmetric $\mathbb{R}$-bilinear map

$$\{\cdot,\cdot\}: C^\infty(M)\times C^\infty(M)\to C^\infty(M)$$

fulfilling the Jacobi identity

$$\{\{F,G\},H\}+\{\{H,F\},G\}+\{\{G,H\},F\}=0$$

and the Leibniz rule

$$\{FG,H\}=F\{G,H\}+\{F,H\}G$$

A Poisson manifold is a differentiable manifold endowed with a Poisson bracket. Starting from a Poisson bracket, one can introduce a tensor field $P$ of type $(2,0)$, which we consider as a map from $T^*M$ to $TM$, defined by

$$\langle \mathrm{d}G, P\,\mathrm{d}F\rangle=\{F,G\}$$

or, using coordinates on $M$, by $P^{ij}=\{x^i,x^j\}$. This tensor field is called the Poisson tensor associated with $\{\cdot,\cdot\}$. It is skew-symmetric, and its components satisfy the cyclic condition

$$P^{il}\frac{\partial P^{jk}}{\partial x^l}+P^{jl}\frac{\partial P^{ki}}{\partial x^l}+P^{kl}\frac{\partial P^{ij}}{\partial x^l}=0$$

meaning that the Schouten bracket $[P,P]$ vanishes.

On a Poisson manifold, the vector field $X_H=\{H,\cdot\}=P\,\mathrm{d}H$ is called the Hamiltonian vector field associated with $H$. In coordinates, $X_H^j=P^{ij}\partial H/\partial x^i$. The Jacobi identity is equivalent to the statement that the map $H\mapsto X_H$, assigning to a function $H$ its Hamiltonian vector field $X_H$, is a Lie algebra homomorphism:

$$X_{\{F,G\}}=[X_F,X_G] \qquad [1]$$

A Casimir function is a function $H$ such that $X_H=0$, that is, a function which is in involution with any other function on $M$. In terms of the Poisson tensor, a Casimir is a function whose differential belongs to the kernel of $P$.

The most famous class of Poisson manifolds is certainly that of symplectic manifolds. They can be seen as nondegenerate Poisson manifolds. Indeed, if a Poisson tensor $P$ is invertible, then its inverse defines a closed nondegenerate 2-form (i.e., a symplectic form). Moreover, any Poisson manifold turns out to be foliated in symplectic leaves.

Let us introduce now the bi-Hamiltonian manifolds, which can be considered as a geometric setting for the study of integrable Hamiltonian systems. A manifold $M$ endowed with two Poisson brackets, $\{\cdot,\cdot\}$ and $\{\cdot,\cdot\}'$, is said to be bi-Hamiltonian if the brackets are compatible, that is, if any linear combination (with constant coefficients) of them is still a Poisson bracket. Such a linear combination automatically satisfies all properties of a Poisson bracket except the Jacobi identity. This is fulfilled if and only if the following compatibility condition holds:

$$\begin{aligned}&\{F,\{G,H\}\}'+\{H,\{F,G\}\}'+\{G,\{H,F\}\}'\\&\quad+\{F,\{G,H\}'\}+\{H,\{F,G\}'\}\\&\quad+\{G,\{H,F\}'\}=0\end{aligned} \qquad [2]$$

for any triple $(F,G,H)$ of functions on $M$. This amounts to saying that the sum of the two Poisson brackets is also a Poisson bracket. In this case the two (compatible) Poisson brackets are said to form a Poisson pair.

There are some interesting equivalent forms of the compatibility condition [2]. First of all, in terms of the components of the Poisson tensors $P$ and $P'$, it reads

$$\begin{aligned}&P^{il}\frac{\partial(P')^{jk}}{\partial x^l}+P^{jl}\frac{\partial(P')^{ki}}{\partial x^l}+P^{kl}\frac{\partial(P')^{ij}}{\partial x^l}\\&\quad+(P')^{il}\frac{\partial P^{jk}}{\partial x^l}+(P')^{jl}\frac{\partial P^{ki}}{\partial x^l}+(P')^{kl}\frac{\partial P^{ij}}{\partial x^l}=0\end{aligned}$$

that is, the Schouten bracket $[P,P']$ vanishes. Moreover, if $X_F=P\,\mathrm{d}F$ is the Hamiltonian vector field associated with $F\in C^\infty(M)$ by means of $P$ and $Y_F=P'\,\mathrm{d}F$ is the one obtained by $P'$, the compatibility condition takes the form

$$\begin{aligned}&[X_F,Y_G]+[Y_F,X_G]=X_{\{F,G\}'}+Y_{\{F,G\}}\\&\quad\forall\,F,G\in C^\infty(M)\end{aligned} \qquad [3]$$

to be compared with [1]. Moreover, in terms of Lie derivatives we have the equivalent condition

$$L_{X_F}P'+L_{Y_F}P=0 \quad \forall\,F\in C^\infty(M) \qquad [4]$$

Now we turn our attention to special vector fields that can be selected on a bi-Hamiltonian manifold $M$. Let $P$ and $P'$ be the Poisson tensors associated with the (compatible) Poisson brackets of $M$. A vector field $X$ on $M$ is said to be bi-Hamiltonian if it is Hamiltonian with respect to both Poisson structures, that is, if there exist two functions $H_0$ and $H_1$ such that

$$X=P\,\mathrm{d}H_1=P'\,\mathrm{d}H_0 \qquad [5]$$

We will see in the following that such vector fields are likely to have a number of first integrals in involution, and thus they are good candidates for a geometric description of integrable systems. The next section is devoted to examples of bi-Hamiltonian (and multi-Hamiltonian) systems.

## Examples

The first example is the Euler top, that is, free motions of a rigid body with a fixed point. The equations of motion are

$$\dot{\Gamma}_1=\frac{I_2-I_3}{I_2I_3}\Gamma_2\Gamma_3$$

and its cyclic permutations. They define a vector field in $\mathbb{R}^3$, which is well known to be Hamiltonian with respect to the Lie–Poisson structure on the (dual of the) Lie algebra of $3\times 3$ skew-symmetric matrices. This means that

$$\dot{\Gamma}_j=\{H,\Gamma_j\},\quad j=1,2,3$$

where

$$H=\frac{1}{2}\left(\frac{{\Gamma_1}^2}{I_1}+\frac{{\Gamma_2}^2}{I_2}+\frac{{\Gamma_3}^2}{I_3}\right)$$

is the kinetic energy and the bracket $\{\cdot,\cdot\}$ is defined by $\{\Gamma_1,\Gamma_2\}=\Gamma_3$ and its cyclic permutations. Another Hamiltonian representation is given by

$$\dot{\Gamma}_j=\{K,\Gamma_j\}',\quad j=1,2,3$$

where

$$K=\tfrac{1}{2}({\Gamma_1}^2+{\Gamma_2}^2+{\Gamma_3}^2)$$

and the new bracket $\{\cdot,\cdot\}'$ is defined by $\{\Gamma_1,\Gamma_2\}'=-\Gamma_3/I_3$ and its cyclic permutations.Any linear combination of the two brackets has the form of the second one, and it is very easy to show that the Jacobi identity is satisfied for such a bracket. Therefore, the Euler top is a bi-Hamiltonian system. Let us also notice that

$$\{K,\Gamma_j\}=\{H,\Gamma_j\}'=0,\quad j=1,2,3$$

that is, $K$ is a Casimir function for the Lie–Poisson bracket and $H$ is a Casimir function for the new Poisson bracket. Hence, we have the following (recursion) relations:

$$\begin{aligned}\{K,\Gamma_j\}&=0\\\{H,\Gamma_j\}&=\{K,\Gamma_j\}'\\0&=\{H,\Gamma_j\}'\end{aligned} \qquad [6]$$

From a geometrical point of view, the situation is as follows. The symplectic leaves of $\{\cdot,\cdot\}$ are the level surfaces of $K$, that is, spheres, while the symplectic

leaves of $\{\cdot,\cdot\}'$ are the ellipsoids $H=$ constant. Their intersections are Lagrangian submanifolds for both symplectic leaves (in the compact case they are the Arnol'd–Liouville tori of the integrable systems, that in this case coincide with the trajectories).

Let us consider now the (three-particle) open Toda lattice. It consists in three particles (with masses equal to 1) moving on the line under a nearest-neighbor interaction of exponential type. The Hamiltonian is given by

$$H=\tfrac{1}{2}\left(p_1{}^2+p_2{}^2+p_3{}^2\right)+\exp(q_1-q_2)+\exp(q_2-q_3)$$

and the system is of course Hamiltonian with respect to the canonical Poisson structure of $\mathbb{R}^6$,

$$P=\begin{pmatrix} 0 & 0 & 0 & -1 & 0 & 0\\ 0 & 0 & 0 & 0 & -1 & 0\\ 0 & 0 & 0 & 0 & 0 & -1\\ 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

But the Toda vector field can also be written as $P'\,\mathrm{d}K$, where $K=p_1+p_2+p_3$ is the total momentum and

$$P'=\begin{pmatrix} 0 & 1 & 1 & -p_1 & 0 & 0\\ -1 & 0 & 1 & 0 & -p_2 & 0\\ -1 & -1 & 0 & 0 & 0 & -p_3\\ p_1 & 0 & 0 & 0 & \mathrm{e}^{(q_1-q_2)} & 0\\ 0 & p_2 & 0 & -\mathrm{e}^{(q_1-q_2)} & 0 & \mathrm{e}^{(q_2-q_3)}\\ 0 & 0 & p_3 & 0 & -\mathrm{e}^{(q_2-q_3)} & 0 \end{pmatrix}$$

is a Poisson tensor, which turns out to be compatible with $P$. The generalization to an arbitrary number of particles is straightforward. Hence, the open Toda lattice is a bi-Hamiltonian system. In the next section we will show that this property can be used to construct a maximal set of integrals of motion for the Toda lattice, which are automatically in involution.

The third example – a stationary reduction of the KdV equation – comes from the field of soliton equations. Let us recall that the first members of the KdV hierarchy are

$$\begin{aligned} \frac{\partial u}{\partial t_1}&=u_x\\ \frac{\partial u}{\partial t_3}&=\tfrac{1}{4}(u_{xxx}-6uu_x)\ (\text{KdV equation})\\ \frac{\partial u}{\partial t_5}&=\tfrac{1}{16}(u_{xxxxx}-10uu_{xxx}\\ &\quad -20u_xu_{xx}+30u^2u_x) \end{aligned}\qquad [7]$$

It is well known how to find finite-dimensional reductions for the KdV equation, giving rise to explicit solutions. Indeed, the set of singular points of a given vector field of the hierarchy is a finite-dimensional manifold which is invariant under the flows of the other vector fields, due to the fact that the flows commute. The (finite-dimensional) systems obtained by restricting the KdV hierarchy to such invariant manifolds are called the stationary reductions of KdV. Let us consider explicitly the reduction corresponding to the third vector field of the hierarchy. The set of its critical points is given by

$$u_{xxxxx}-10uu_{xxx}-20u_xu_{xx}+30u^2u_x=0\qquad [8]$$

and its dimension is 5, since we can use the values of $u, u_x, u_{xx}, u_{xxx}$, and $u_{xxxx}$ at a fixed point $x_0$ (i.e., the Cauchy data) as global coordinates. For the sake of simplicity, we set

$$\begin{aligned} &u_0=u(x_0),\quad u_1=u_x(x_0),\quad u_2=u_{xx}(x_0)\\ &u_3=u_{xxx}(x_0),\quad u_4=u_{xxxx}(x_0) \end{aligned}$$

In order to compute the reduced equations of the first flow of [7], we have to take its $x$-derivative and to use the constraint [8] and its differential consequences to eliminate all the derivatives of order higher than 4. We obtain the equations

$$\begin{aligned} &\frac{\partial u_0}{\partial t_1}=u_1,\ \frac{\partial u_1}{\partial t_1}=u_2,\ \frac{\partial u_2}{\partial t_1}=u_3,\ \frac{\partial u_3}{\partial t_1}=u_4\\ &\frac{\partial u_4}{\partial t_1}=10u_0u_3+20u_1u_2-30u_0{}^2u_1 \end{aligned}\qquad [9]$$

In the same way, for the KdV equation we get

$$\begin{aligned} \frac{\partial u_0}{\partial t_3}&=\tfrac{1}{4}(u_3-6u_0u_1)\\ \frac{\partial u_1}{\partial t_3}&=\tfrac{1}{4}(u_4-6u_0u_2-6u_1{}^2)\\ \frac{\partial u_2}{\partial t_3}&=\tfrac{1}{4}(4u_0u_3+2u_1u_2-30u_0{}^2u_1)\\ \frac{\partial u_3}{\partial t_3}&=\tfrac{1}{4}(4u_0u_4+6u_1u_3+2u_2{}^2\\ &\quad -30u_0{}^2u_2-60u_0u_1{}^2)\\ \frac{\partial u_4}{\partial t_3}&=\tfrac{1}{4}(10u_1u_4+10u_0{}^2u_3+10u_2u_3\\ &\quad -100u_0u_1u_2-60u_1{}^3-120u_0{}^3u_1) \end{aligned}\qquad [10]$$

There are two compatible Poisson structures giving a bi-Hamiltonian formulation of both systems. The corresponding Poisson tensors are

$$P=\begin{bmatrix} 0 & 0 & 0 & 2 & 0\\ 0 & 0 & -2 & 0 & -20u_0\\ 0 & 2 & 0 & 20u_0 & 20u_1\\ -2 & 0 & -20u_0 & 0 & -140u_0{}^2-20u_2\\ 0 & 20u_0 & -20u_1 & 140u_0{}^2+20u_2 & 0 \end{bmatrix}$$

and

$$P' = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 3u_0 & 6u_1 \\ -\frac{1}{2} & 0 & -3u_0 & -3u_1 & -4u_2 - 15{u_0}^2 \\ 0 & 3u_0 & 0 & u_2 + 15{u_0}^2 & u_3 + 30u_0u_1 \\ -3u_0 & 3u_1 & -u_2 - 15{u_0}^2 & 0 & \begin{matrix} u_4 - 40u_0u_2 + \\ 30{u_1}^2 - 60{u_0}^3 \end{matrix} \\ -6u_1 & 4u_2 + 15{u_0}^2 & -u_3 - 30u_0u_1 & \begin{matrix} -u_4 + 40u_0u_2 - \\ 30{u_1}^2 + 60{u_0}^3 \end{matrix} & 0 \end{bmatrix}$$

In fact, if we call $X_1$ and $X_3$ the vector fields given by [9] and [10], then the following recursion relations hold:

$$\begin{aligned} P\,\mathrm{d}H_0 &= 0 \\ X_1 = P\,\mathrm{d}H_1 &= P'\,\mathrm{d}H_0 \\ X_3 = P\,\mathrm{d}H_2 &= P'\,\mathrm{d}H_1 \\ 0 &= P'\,\mathrm{d}H_2 \end{aligned} \qquad [11]$$

where

$$\begin{aligned} H_0 &= -u_4 + 10u_0u_2 + 5{u_1}^2 - 10{u_0}^3 \\ H_1 &= \tfrac{1}{4}\left(2u_0u_4 - 2u_1u_3 + {u_2}^2 - 20{u_0}^2u_2 + 15{u_0}^4\right) \\ H_2 &= \tfrac{1}{16}\left(2u_2u_4 - 6{u_0}^2u_4 - {u_3}^2 + 12u_0u_1u_3 \right. \\ &\quad \left. -16u_0{u_2}^2 - 12{u_1}^2u_2 + 60{u_0}^3u_2 - 36{u_0}^5\right) \end{aligned}$$

Therefore, the vector fields $X_1$ and $X_3$ are bi-Hamiltonian. The geometry of this bi-Hamiltonian manifold is similar to the one of the first example. The symplectic leaves of both Poisson structures have dimension 4, and the Lagrangian foliation (given by the level submanifolds of $H_0, H_1$, and $H_2$) is contained in the intersections of such leaves. This Lagrangian foliation is called by Gelfand and Zakharevich the "axis" of the bi-Hamiltonian manifold.

We also notice that the relations [11] can be collected in the statement that the function $H(\lambda) = H_0\lambda^2 + H_1\lambda + H_2$ is a Casimir of the Poisson pencil $P_\lambda = P' - \lambda P$, that is,

$$P_\lambda\,\mathrm{d}H(\lambda) = 0$$

The importance of the stationary reductions of the KdV hierarchy lies in the fact that (as noticed in the early works on the subject) the reduced equations can be solved by means of the classical method of separation of variables. We mention that the separability of these systems is a particular instance of a general result, which is valid for quite a wide class of bi-Hamiltonian manifolds.

Next we present an important class of bi-Hamiltonian manifolds. We recall that the dual $\mathfrak{g}^*$ of a finite-dimensional Lie algebra $\mathfrak{g}$ possesses a canonical Poisson structure, called the Lie–Poisson structure. It is defined as

$$\{F, G\}(X) = \langle X, [\mathrm{d}F(X), \mathrm{d}G(X)]\rangle \qquad [12]$$

where $F, G \in C^\infty(\mathfrak{g}^*)$ and their differentials at $X \in \mathfrak{g}^*$ are seen as elements of $\mathfrak{g}$. If $X_0$ is a fixed element in $\mathfrak{g}^*$, the constant Poisson bracket

$$\{F, G\}'(X) = \langle X_0, [\mathrm{d}F(X), \mathrm{d}G(X)]\rangle \qquad [13]$$

is compatible with the Lie–Poisson bracket. In fact, the Poisson pencil $\{\cdot\,,\cdot\}_\lambda = \{\cdot\,,\cdot\} - \lambda\{\cdot\,,\cdot\}'$ is obtained from $\{\cdot\,,\cdot\}$ by applying the translation $X \mapsto X + \lambda X_0$; hence, it is a Poisson bracket for every value of the constant $\lambda$. The method of translation of the argument, due to Manakov, provides a lot of bi-Hamiltonian vector fields for this bi-Hamiltonian manifold. One has to consider an $\mathrm{Ad}^*$-invariant function on $\mathfrak{g}^*$, that is, a function $H \in C^\infty(\mathfrak{g}^*)$ such that

$$\langle X, [\mathrm{d}H(X), x]\rangle = 0 \quad \forall\, x \in \mathfrak{g},\ X \in \mathfrak{g}^*$$

It is clearly a Casimir function for the Lie–Poisson bracket, and this implies that the function $X \mapsto H(X - \lambda X_0)$ is a Casimir of the Poisson pencil. If this function can be developed as a Laurent series in $\lambda$, its coefficients $H_j$ fulfill the recursion relations

$$\{H_{j+1}, \cdot\} = \{H_j, \cdot\}' \qquad [14]$$

and thus give rise to a sequence of bi-Hamiltonian vector fields.

The last example is a generalization of the previous one. For the sake of simplicity, we consider a Lie algebra $\mathfrak{g}$ of matrices such that the trace of the product is nondegenerate, and the space $M = \mathfrak{g}^2 = \mathfrak{g} \times \mathfrak{g}$. If $F \in C^\infty(M)$, its differential at a point $(x_0, x_1)$ can be identified with the element $(\partial F/\partial x_0, \partial F/\partial x_1)$ of $M$ given by

$$\frac{\mathrm{d}}{\mathrm{d}t_{|\epsilon=0}} F(x_0 + \epsilon v_0, x_1 + \epsilon v_1) = \mathrm{tr}\left(\frac{\partial F}{\partial x_0} v_0 + \frac{\partial F}{\partial x_1} v_1\right)$$

for all $v_0, v_1 \in \mathfrak{g}$. The manifold $M$ has a three-dimensional family of pairwise compatible Poisson brackets:

$$\{F,G\}_0(x_0,x_1) = -\mathrm{tr}\left(x_0\left[\frac{\partial F}{\partial x_1},\frac{\partial G}{\partial x_1}\right]\right)$$

$$\{F,G\}_1(x_0,x_1) = \mathrm{tr}\left(x_1\left[\frac{\partial F}{\partial x_1},\frac{\partial G}{\partial x_1}\right]\right)$$

$$\{F,G\}_2(x_0,x_1) = \mathrm{tr}\left(x_0\left[\frac{\partial F}{\partial x_0},\frac{\partial G}{\partial x_0}\right] + x_1\left(\left[\frac{\partial F}{\partial x_1},\frac{\partial G}{\partial x_0}\right]+\left[\frac{\partial F}{\partial x_0},\frac{\partial G}{\partial x_1}\right]\right)\right)$$

Notice that the first two brackets restrict to the submanifolds $x_0 = \text{constant}$ and give rise to the bi-Hamiltonian structure presented in the previous example (via the identification between $\mathfrak{g}$ and $\mathfrak{g}^*$ given by the trace of the product). This example can be generalized to an arbitrary number $n$ of copies of $\mathfrak{g}$. In this case there is an $(n+1)$-dimensional family of pairwise compatible Poisson brackets, which can be shown to be Lie–Poisson brackets with respect to suitable Lie algebra structures on $\mathfrak{g}^n$. According to Reyman and Semenov–Tian–Shansky, these brackets can also be casted in the $R$-matrix formalism.

Also in this case, the Ad-invariant functions on $\mathfrak{g}$ give rise to functions in involution on our multi-Hamiltonian manifold. For example, if $H_k^{(\alpha)}$ denotes the $\lambda^k$-coefficient of $\mathrm{tr}(x_1\lambda + x_0)^\alpha$, then the recursion relations

$$\{H_k^{(\alpha)},\cdot\}_l = \{H_{k+1}^{(\alpha)},\cdot\}_{l+1}, \quad k\geq 0,\ l=0,1$$

hold, and they imply the existence of tri-Hamiltonian vector fields on $M$.

Finally, we mention that the bi-Hamiltonian structure of the stationary flow of KdV – discussed above – can be obtained as a suitable reduction of the multi-Hamiltonian structure on $\mathfrak{g}^3$, where $\mathfrak{g} = \mathfrak{sl}(2,\mathbb{R})$. A similar statement holds for the other stationary flows of the Gelfand–Dickey hierarchies.

## Iterative Properties and Integrability

In this section we show how to use the bi-Hamiltonian formulation of a given system to explain its integrability. In the cases similar to the open Toda lattice, where one of the Poisson structures is nondegenerate, one can introduce a recursion operator and employ its powers in order to generate a chain of integrals of motion in involution. In the other examples, where the bi-Hamiltonian structure is degenerate, the conserved quantities turn out to be the coefficients of Casimir functions of the Poisson pencil.

If $(M, \{\cdot,\cdot\}, \{\cdot,\cdot\}')$ is a bi-Hamiltonian manifold, we call bi-Hamiltonian hierarchy a sequence $\{H_k\}_{k\geq 0}$ of functions on $M$ fulfilling the recursion relations

$$\{\cdot, H_{k+1}\} = \{\cdot, H_k\}', \quad k\geq 0 \qquad [15]$$

In terms of Poisson tensors we have that $P\,\mathrm{d}H_{k+1} = P'\,\mathrm{d}H_k$. A bi-Hamiltonian hierarchy clearly gives rise to an infinite sequence of bi-Hamiltonian vector fields,

$$X_k = P\,\mathrm{d}H_k = P'\,\mathrm{d}H_{k-1}, \quad k\geq 1 \qquad [16]$$

The functions $H_k$ are in involution with respect to both Poisson brackets. Indeed, for $k > j$, one has

$$\{H_j,H_k\} = \{H_j,H_{k-1}\}' = \{H_{j+1},H_{k-1}\} = \cdots = \{H_k,H_j\}$$

so that $\{H_j,H_k\}=0$ for all $j,k\geq 0$, and therefore $\{H_j,H_k\}'=0$ for all $j,k\geq 0$. If $\{H_i\}_{i\geq 0}$ and $\{K_i\}_{i\geq 0}$ are two bi-Hamiltonian hierarchies, then all functions are in (bi-)involution provided that one of the two hierarchies starts from a Casimir of $\{\cdot,\cdot\}$. In fact, suppose that $H_0$ is such a Casimir. Then

$$\{H_i,K_j\} = \{H_{i-1},K_j\}' = \{H_{i-1},K_{j+1}\} = \cdots = \{H_0,K_{j+i}\} = 0$$

and

$$\{H_i,K_j\}' = \{H_{i+1},K_j\} = 0$$

We observe that these proofs of the involutivity do not use the compatibility condition [2] between the Poisson structures. The point is that this condition is important for the existence of bi-Hamiltonian hierarchies. Indeed, the problem of the existence and the construction of bi-Hamiltonian hierarchies is quite delicate. We tackle it first in the case of a particular class of bi-Hamiltonian manifolds, the so-called Poisson–Nijenhuis manifolds. In turn, they are a generalization of nondegenerate bi-Hamiltonian manifolds.

Let $(M,P,P')$ be a bi-Hamiltonian manifold such that $P$ is invertible. Then we can introduce the tensor field $N = P'P^{-1}$, which is of type $(1,1)$ and will always be dealt with as an endomorphism of the tangent bundle $TM$. This tensor field possesses some remarkable properties. First of all, its Nijenhuis torsion $T(N)$ vanishes; this means that

$$T(N)(X,Y) = [NX,NY] - N[X,Y]_N = 0$$

for any pair $(X,Y)$ of vector fields on $M$, where

$$[X,Y]_N = [NX,Y] + [X,NY] - N[X,Y]$$

Sometimes a tensor field with vanishing Nijenhuis torsion is called a recursion operator. Since $P$ defines a symplectic structure on $M$, such a bi-Hamiltonian manifold is called an $\omega N$ manifold.

The tensor field $N$ satisfies two compatibility conditions with $P$. The first one is simply the skew-symmetry of $P'$ and reads $NP = PN^*$, while the second one is a restatement of [3],

$$[X_F, X_G]_N = X_{\{F,G\}_{NP}} \quad \forall F, G \in C^\infty(M)$$

A manifold is said to be a Poisson–Nijenhuis manifold (briefly, a PN manifold) if it is endowed with a Poisson tensor $P$ and a torsionless $(1, 1)$ tensor field $N$ which are compatible, in the sense that the two above-mentioned conditions hold. We have just seen that every nondegenerate bi-Hamiltonian manifold (i.e., such that one of the two Poisson tensors is invertible) is a PN manifold. On the other hand, if $(M, P, N)$ is a PN manifold, then it can be shown that $P' = NP$ is a Poisson tensor, which is compatible with $P$. In other words, PN manifolds are particular examples of bi-Hamiltonian manifolds. Moreover, one has that $P^{(j)} = N^j P$ and $P^{(k)} = N^k P$ are, for every $j, k \geq 0$, compatible Poisson tensors.

Let us consider now a function $H_0$, on a PN manifold $(M, P, N)$, such that $N^* \,\mathrm{d}H_0 = \mathrm{d}H_1$ is exact, where $N^* : T^*M \to T^*M$ is the adjoint of the recursion operator $N$. This implies that

$$X = P\,\mathrm{d}H_1 = PN^*\,\mathrm{d}H_0 = P'\,\mathrm{d}H_0 \qquad [17]$$

is a bi-Hamiltonian vector field. By means of $N^*$ we can define the 1-forms $\alpha_j = (N^*)^j\,\mathrm{d}H_0$, which can be shown to be all closed. If they are exact, that is, $\alpha_k = \mathrm{d}H_k$, then the functions $H_k$ form a bi-Hamiltonian hierarchy and thus are in involution. This shows that on a (simply connected) PN manifold every bi-Hamiltonian vector field of the form [17], with $N^*\,\mathrm{d}H_0 = \mathrm{d}H_1$, belongs to a bi-Hamiltonian hierarchy and that its first integrals (in involution) can be iteratively constructed with the recursion operator. (The integrability of this vector field clearly depends on the number of independent integrals of motion.) Moreover, the vector field $X_k = P\,\mathrm{d}H_k = P'\,\mathrm{d}H_{k-1}$ of the hierarchy is Hamiltonian with respect to all Poisson structures $P^{(j)}$ with $j \geq k$, because $X_k = P^{(j)}\,\mathrm{d}H_{k-j}$.

The example of the Toda lattice presented earlier can be casted in the PN (more precisely, $\omega N$) framework. One can introduce the recursion operator $N$ and, in the three-particle case, one can define the third integral of motion as $\mathrm{d}J = N^*\,\mathrm{d}H$. Since $K$, $H$, and $J$ belong to a bi-Hamiltonian hierarchy, they are in involution, and this (along with their functional independency) proves the integrability of the Toda lattice.

In this example something more happens: the integrals of motion are (up to multiplicative constants) the traces of the powers of the recursion operator $N$. This is a general fact, since the vanishing of the torsion of $N$ implies that $N^*\,\mathrm{d}I_k = \mathrm{d}I_{k+1}$, where $I_k = (1/k)\mathrm{tr}\,N^k$.

Next we deal with the case where the bi-Hamiltonian manifold $(M, P, P')$ is not of the Poisson–Nijenhuis type, that is, both $P$ and $P'$ are degenerate. Let us suppose that their symplectic leaves have codimension 1. We also want to discuss in this case an iteration problem, namely the problem of constructing a bi-Hamiltonian hierarchy starting from a Casimir $H_0$ of $P$. Let us consider the Hamiltonian vector field $X_1 = P'\,\mathrm{d}H_0 = Y_{H_0}$ (using the notations introduced earlier). Thanks to the form [4] of the compatibility condition between $P$ and $P'$, we have that

$$L_{X_1} P = L_{Y_{H_0}} P = -L_{X_{H_0}} P' = 0$$

meaning that $X_1$ is an infinitesimal symmetry of $P$. Moreover, $X_1$ is tangent to the symplectic leaves of $P$, since $\langle \mathrm{d}H_0, X_1 \rangle = \langle \mathrm{d}H_0, P'\,\mathrm{d}H_0 \rangle = 0$. Under some suitable topological assumptions, we can conclude that there exists a function $H_1$ such that $X_1 = P\,\mathrm{d}H_1$, that is, $X_1$ is a bi-Hamiltonian vector field. Now the procedure can be iterated, that is, in the same way one can show that, if $X_2 = P'\,\mathrm{d}H_1 = Y_{H_1}$, then there exists a function $H_2$ such that $X_2 = P\,\mathrm{d}H_2$, and so on. Thus, one obtains a bi-Hamiltonian hierarchy $\{H_k\}_{k \geq 0}$, which can either be infinite or end with a Casimir of $P'$. In any case, the function $H(\lambda) = \sum_{k \geq 0} H_k \lambda^{-k}$ is a Casimir of the Poisson pencil $P_\lambda = P' - \lambda P$. As seen earlier, the typical situation is that the chain terminates with a Casimir $H_n$ of $P'$, where $\dim M = 2n + 1$. In other words, there is a Casimir of the Poisson pencil which is a polynomial of degree $n$ in the parameter $\lambda$.

As a general procedure for constructing bi-Hamiltonian hierarchies, one can look for the Casimir functions $H(\lambda)$ of the Poisson pencil which are deformations of Casimir functions of $P$, but it is not clear when such a deformation does exist in the case where the corank of the bi-Hamiltonian structure is greater than 2. Nevertheless, suppose that $H(\lambda) = \sum_{k \geq 0} H_k \lambda^{-k}$ is a Casimir of $P_\lambda$, that is, that $\{H_k\}_{k \geq 0}$ is a bi-Hamiltonian hierarchy. Then, for all $\lambda$, the bi-Hamiltonian vector fields $X_{k+1} = P\,\mathrm{d}H_{k+1} = P'\,\mathrm{d}H_k$ are Hamiltonian with respect to $P_\lambda$, with Hamiltonian function $H^{(k)}(\lambda) = \sum_{j=0}^{k} H_j \lambda^{k-j}$,

$$X_{k+1} = P_\lambda \mathrm{d}H^{(k)}(\lambda)$$

Therefore, the vector fields $X_k$ are not only bi-Hamiltonian, but they are Hamiltonian with respect to any Poisson bracket of the pencil.

In this article we have described some basic properties of bi-Hamiltonian systems, defined on manifolds possessing a Poisson pair. There are other important vector fields on these manifolds (more precisely, on $\omega N$ manifolds). They are called cyclic systems of Levi-Civita, and they give an intrinsic description of the separable systems of Jacobi. We refer to the article Recursion Operators in Classical Mechanics in this encyclopedia for these topics.

*See also:* Bi-Hamiltonian Methods in Soliton Theory; Classical *r*-Matrices, Lie Bialgebras, and Poisson Lie Groups; Integrable Systems and Algebraic Geometry; Integrable Systems and Recursion Operators on Symplectic and Jacobi Manifolds; Integrable Systems: Overview; Recursion Operators in Classical Mechanics; Separation of Variables for Differential Equations; Solitons and Kac–Moody Lie Algebras; Toda Lattices.

## Further Reading

Adler M, van Moerbeke P, and Vanhaecke P (2004) *Algebraic Integrability, Painlevé Geometry and Lie Algebras*. Berlin: Springer.

Arnol'd VI and Givental AB (2001) Symplectic geometry. In: Arnol'd VI *et al.* (eds.) *Encyclopedia of Mathematical Sciences, Dynamical Systems IV*, pp. 1–138. Berlin: Springer.

Babelon O, Cartier P, and Kosmann-Schwarzbach Y (eds.) (1994) *Lectures on Integrable Systems*. Singapore: World Scientific.

Błaszak M (1998) *Multi-Hamiltonian Theory of Dynamical Systems*. Berlin: Springer.

Dorfman I (1993) *Dirac Structures and Integrability of Nonlinear Evolution Equations*. Chichester: Wiley.

Gelfand IM and Zakharevich I (1993) On the local geometry of a bi-Hamiltonian structure. In: Corwin L *et al.* (eds.) *The Gelfand Mathematical Seminars 1990–1992*, pp. 51–112. Boston: Birkäuser.

Magri F, Falqui G, and Pedroni M (2003) The method of Poisson pairs in the theory of nonlinear PDEs. In: Conte R *et al.* (eds.) *Direct and Inverse Methods in Nonlinear Evolution Equations*, Lecture Notes in Physics, vol. 632, pp. 85–136. Berlin: Springer.

Magri F, Casati P, Falqui G, and Pedroni M (2004) Eight lectures on integrable systems. In: Kosmann-Schwarzbach Y *et al.* (eds.) *Integrability of Nonlinear Systems*, Lecture Notes in Physics, vol. 638, pp. 209–250. Berlin: Springer.

Olshanetsky MA, Perelomov AM, Reyman AG, and Semenov-Tian-Shansky MA (1994) Integrable systems. II. In: Arnol'd VI *et al.* (eds.) *Encyclopedia of Mathematical Sciences, Dynamical systems VII*, pp. 83–259. Berlin: Springer.

Olver PJ (1993) *Applications of Lie Groups to Differential Equations*, 2nd edn. New York: Springer.

Vaisman I (1994) *Lectures on the Geometry of Poisson Manifolds*. Basel: Birkhäuser.

# Multiscale Approaches

**A Lesne**, Université P.-M. Curie, Paris VI, Paris, France



## Introduction: Multiple-Scale and Multiscale Approaches

Multiscale, or more precisely *multiple-scale, method* is a technique of perturbation theory based on the introduction of additional rescaled variables, say time variables, formally considered as independent variables and describing each a different timescale (for the sake of simplicity, we will mainly consider a dynamic framework and timescales; all can be transposed to spatial dependences and scales). It was first developed to handle singular situations in which dynamic regimes of different characteristic scales coexist and intermingle in such a way that straightforward perturbation expansions are not uniformly convergent in time (hence of limited relevance and use) due to the so-called *secular terms* growing unbounded with time; the freedom introduced together with the extra variables indeed allows to impose conditions preventing these secular divergences and improving the convergence of the perturbation series. It yields a global perturbation solution describing *jointly* the behavior at small and large scales. This technique belongs to the far more wide-ranging class of *multiscale approaches*; these can be divided into four main subclasses:

1. *Mean-field techniques* exploiting scale separation between fast and slow components of the dynamics. The influence of the slow variables onto the fast dynamics, if any, is treated in a decoupled way within a parametric approximation, allowing an adiabatic elimination of fast variables (see the section "Slow/fast variables").
2. *Singular perturbations*, in which individual fast components ultimately give rise to slow trends and influence the large-scale features. Scale separation here breaks down at long times and multiple-scale method is then a method of choice (see the next section).
3. *Matched expansions* when regimes of different scales succeed (boundary-layer singularity; see the section "Boundary layers and matched expansions").

4. *Renormalization techniques*, in systems exhibiting some kind of universality in the relations between their behaviors at different scales, for example, scale invariance (see the section "Renormalization: an iterated multiscale approach").

We will first present the principles of multiple-scale method, detail its technical implementation on simple abstract examples and cite some typical applications. Then we will articulate this technique with more general multiscale methods in a brief overview (see the section "A brief overview of multiscale approaches"). The range of multiscale approaches and technical tools will then be illustrated and compared in the context of diffusion, Brownian motion, and transport phenomena (see the section "Summary: the exemplary case of diffusion").

## Multiple-Scale Method: Principles

### Context: Singular Perturbations and Secular Divergences

Multiple-scale methods have been developed to handle situations in which the dynamics involves a small parameter $\epsilon$ (e.g., the ratio of the masses of different subsystems, the strength of an additional interaction, the amplitude of an applied field) directly controlling the separation between the different characteristic timescales of the evolution and, specifically, such that *the behavior for* $\epsilon = 0$ *is qualitatively different from the behavior for* $\epsilon$ *small* ($\epsilon \ll 1$ but finite); in other words, when a weak influence, of strength controlled by $\epsilon \ll 1$, does not have only weak consequences. Typically, this occurs when $\epsilon$ represents the strength of a weak coupling between otherwise independent subsystems or when a vanishing value $\epsilon = 0$ changes a characteristic time, the sign of a friction coefficient, the order of the highest time derivative in case of ordinary differential equations (turning points), or the type of partial differential equations in case of spatially extended systems. Accordingly, a naive perturbative approach with respect to $\epsilon$, that is, an expansion taking as a basic approximation the behavior for $\epsilon = 0$, cannot bridge the qualitative gap with behaviors observed for $\epsilon > 0$. It thus fails to give a full account of the system evolution at all times: one speaks of *singular perturbation*.

A historical example arose in celestial mechanics, in the celebrated *nonintegrable* three-body problem, involving the Sun, a big planet and a smaller one, of respective masses $m_1, m_2 < m_1$ and $m_3 \ll m_2$. The straightforward approach would be to consider the presence of the small planet as a small perturbation of the *integrable* two-body problem for the masses $m_1$ and $m_2$. But when one tries to determine the solution as a series in powers of the mass ratio $\epsilon = m_3/m_2$, unbounded terms appear, the so-called *secular terms*, increasing without bounds as fast as $t$, hence of ill-defined order and impairing the very consistency of the perturbation approach at long times $t > 1/\epsilon$. Accordingly, the perturbation expansion is *not uniformly convergent in time*, preventing from using it to investigate asymptotics and determine the fate of the three-body system: the influence of the small planet on the motion of the bigger one, although seemingly a weak perturbation, might ultimately modify its trajectory around the Sun, at least in some resonant cases.

The origin of secular terms lies in a phenomenon of *resonance*, which is best explained on an example: the *Duffing oscillator* $\ddot{x} + x = -\epsilon x^3$ with $\epsilon \ll 1$. When looking for a solution in the form $x(t) = \sum \epsilon^n x_n(t)$, each component $x_n(t)$ has to be bounded in order to get a consistent perturbation expansion, in which the hierarchy of terms of different orders remains valid forever: $\epsilon x_{n+1}(t) \ll x_n(t)$. These components should satisfy the following sequence of equations:

$$\ddot{x}_0 + x_0 = 0, \quad \ddot{x}_1 + x_1 = -x_0^3, \quad \ldots \\ \text{(linearized operator } Lx \equiv \ddot{x} + x) \qquad [1]$$

It gives $x_0(t) = a\mathrm{e}^{\mathrm{i}t} + \text{c.c.}$, from which follows a secular contribution $(3\mathrm{i}/2)a|a|^2 t\,\mathrm{e}^{\mathrm{i}t}$ in $x_1(t)$. In general, solving perturbatively $\dot{z} = f(z, \epsilon)$ for an expansion $z(\epsilon, t) = \sum_n \epsilon^n z_n(t)$ yields a hierarchical sequence of equations of the form $\dot{z}_n = Lz_n + \varphi_n(z_0, z_1, \ldots, z_{n-1})$ for $n \geq 1$, where $L = Df(z_0, \epsilon = 0)$ comes from the linearization in $z_0$ of the unperturbed evolution law. A secular divergence arises in $z_n$ as soon as $\varphi_n$ contains an additive contribution which is an eigenvector of $L$ (part of a mathematical result known as the Fredholm alternative). The appearance of secular terms reflects a singular feature of the dynamics: the fact that the limits as $\epsilon \to 0$ and $t \to \infty$ do not commute. As a rule, such noninversion is associated with generalized secular divergences: the fast, short-term dynamics finally contributes to the slow, long-term behavior. This feature is a clue towards using multiple-scale method.

### Technical Principles

The first step is to perform rescalings leading to dimensionless variables and functions, which evidence a small control parameter $\epsilon$, related to scale separation and providing a natural parameter for a perturbation approach. The basic principle of multiple-scale method is to introduce additional *independent* time

variables $t_1, t_2, \ldots, t_n$ such that the physical situation corresponds in this extended time-variable space to the line

$$t_0 = t,\ t_1 = \epsilon t,\ t_2 = \epsilon^2 t,\ \ldots$$
$$\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\partial}{\partial t_0} + \epsilon\frac{\partial}{\partial t_1} + \epsilon^2\frac{\partial}{\partial t_2} + \cdots \quad [2]$$

It thus amounts to a perturbation expansion of the time-derivative operator. This method can be traced back to the *Lindstedt–Poincaré technique*, where the time variable $t$ is expanded according to $t = s(1 + \epsilon\omega_1 + \epsilon^2\omega_2 + \cdots)$ and the evolution described in terms of the new variable $s$ and unknown frequencies $(\omega_i)_{i\geq 1}$ to be determined self-consistently (Nayfeh 1973). By contrast, the multiple-scale approach puts on a par $t_0 = t$ and the additional variables $(t_i)_{i\geq 1}$. The perturbation approach is then carried out as usual, plugging eqn [2] for $\mathrm{d}/\mathrm{d}t$ and the expansion $z(\epsilon, t) = \sum_{n\geq 0} \epsilon^n z_n(t_0, t_1, t_2, \ldots)$ into the evolution equation and identifying term-wise the coefficients of the successive powers of $\epsilon$. The additional freedom thus introduced when considering $(t_i)_{i\geq 0}$ as independent variables will be compensated in the course of the computation, by imposing "solubility conditions" ensuring the vanishing of secular terms and the consistency of the perturbation method. In particular, it is possible to freely choose boundary conditions outside the physical line $t_1 = \epsilon t_0, \ldots, t_n = \epsilon^n t_0$. The resulting set of equations contains exactly the same information as the original one, only expressed in a different way: by construction, terms depending, say, on $t_0$, describe a fast component with no emerging slow trends that would intermix with the $t_1$-dependence; fast variables contribute only to fast modes. At the end, one restricts to the physical line, thus turning back to the single "real" variable $t$. The benefit of the method is to provide a *joint access to dependences at different scales*, now expressing as dependences onto the different time variables $t_0, t_1, \ldots, t_n$. One introduces as many new variables as necessary to circumvent secular divergences. We have implicitly supposed above that the behavior at timescale $\Delta t = \mathcal{O}(1)$ corresponds to the fastest timescale of the evolution. If it were not the case, the rescaled time variables would be $t_0 = \epsilon^{n_0} t$, $t_1 = \epsilon^{n_0+1} t, \ldots$ if the fastest timescale is $\Delta t = \mathcal{O}\ (\epsilon^{n_0})$. More general time-derivative expansion, associated with rescaled variables $t_n = \epsilon^{\alpha_n} t$ might be considered to better account for the hierarchy of characteristic timescales of the dynamics.

### Multiple-Scale Method: Abstract Examples

Let us first consider the simplest possible example $\dot{x} = a(1 + \epsilon)x$, for which the exact solution is trivially known, allowing to appreciate the validity of the multiscale approach compared to the straightforward perturbation expansion. In the latter case, one looks for a solution $x(t) = x_0(t) + \epsilon x_1(t) + \mathcal{O}(\epsilon^2)$ and identifies term-wise the powers of $\epsilon$. At order 0, $\dot{x_0} = ax_0$ yields $x_0(t) = c_0 \mathrm{e}^{at}$. At order 1, $\dot{x_1} - ax_1 = x_0(t)$ leads to a secular divergence: $x_1(t) = c_1^0 \mathrm{e}^{at} + c_0 t \mathrm{e}^{at}$. Carrying on the perturbation analysis yields the following expansion:

$$x(t) = c\ \mathrm{e}^{at}(1 + \epsilon t + \epsilon^2 t^2/2 + \cdots) \quad [3]$$

which is *not uniformly convergent*: for $t = \mathcal{O}(1/\epsilon)$, all terms are of the same magnitude. Using this recursive method to obtain a finite-order approximate solution (e.g., stopping, as here, after two steps of the perturbation method) *is only relevant at short times* $t \ll 1/\epsilon$. The straightforward perturbation analysis captures the behavior of the exact solution only if *all terms* are computed and taken into account (in less trivial examples, the straightforward perturbation series might even be divergent). In the multiple-scale approach, one introduces two rescaled variables $t_0 = t$ and $t_1 = \epsilon t$ and looks for a solution of the form $x(t) \equiv x_0(t_0, t_1, \ldots) + \epsilon x_1(t_0, t_1, \ldots) + \mathcal{O}(\epsilon^2)$. At order 0, $\partial_{t_0} x_0 = ax_0$ yields $x_0(t_0, t_1, \ldots) = c_0(t_1, \ldots)\mathrm{e}^{at_0}$. At order 1, we get $\partial_{t_0} x_1 + \partial_{t_1} x_0 = x_0 + ax_1$. The solubility condition writes $ac_0 - \partial_{t_1} c_0 = 0$, which allows as to avoid secular divergence and suppresses the artificial freedom introduced with the additional time variable $t_1$, yielding $c_0 = c\mathrm{e}^{at_1}$. The equation $(\partial_{t_0} - a)x_1 = 0$ is here superfluous, but in less simple situations, it remains at this stage a nontrivial equation for $x_1$. One thus directly gets the solution, uniformly valid at all times:

$$x(t) = c\ \mathrm{e}^{at_1}\ \mathrm{e}^{at_0} = c\ \mathrm{e}^{a(1+\epsilon)t} \quad [4]$$

As a rule in singular perturbation method, the difficulty here originates in the noncommuting limits $\epsilon \to 0$ and $t \to \infty$; indeed, denoting $y_\epsilon(t) = x_\epsilon(t)\mathrm{e}^{-at}$, one has $\lim_{t\to\infty} \lim_{\epsilon\to 0^+} y_\epsilon(t) = c$, whereas $\lim_{\epsilon\to 0^+} \lim_{t\to\infty} y_\epsilon(t) = \infty$.

Other training examples are the weakly damped linear oscillator $\ddot{x} + x = -2\epsilon\dot{x}$, solved with multiple scales $t_0 = t, t_1 = \epsilon t, t_1 = \epsilon^2 t$, or with the more specific variables $\theta = \sqrt{1 - \epsilon^2}t, \tau = \epsilon t$; the Duffing oscillator $\ddot{x} + x = -\epsilon x^3$ introduced above, whose multiple-scale resolution requires three variables $t_0 = t, t_1 = \epsilon t, t_1 = \epsilon^2 t$; and the Van der Pol oscillator $\ddot{x} + x = \epsilon(1 - x^2)\dot{x}$.

### An Illustration: Classical Lorentz Electron Gas in a Weak Field

As a less abstract, hence more convincing, illustration of the strength of multiple-scale method, let us

consider the dynamics of a classical Lorentz electron gas acted upon an external electric field (associated acceleration $\boldsymbol{a}$). This model considers the electrons as charged hard spheres whose motion results from the superimposition of a driven classical motion in the field and elastic collision on immobile scatterers (the atoms). It is implemented within a kinetic-theoretic framework, based upon a Boltzmann-like equation for the electron velocity distribution:

$$\left(\frac{\partial}{\partial t} + \boldsymbol{a}.\frac{\partial}{\partial \boldsymbol{v}}\right) f(\boldsymbol{v}, t) = -\frac{v}{\lambda} \mathcal{Q} f(\boldsymbol{v}, t) \qquad [5]$$

where $v = |\boldsymbol{v}|$, and $\lambda$ is the mean free path of the electrons. $\mathcal{Q}f = f - f_{\rm sph}$ is a projector accounting for the effect of collisions through the deviation of the distribution $f$ from spherical symmetry, namely through the discrepancy between $f$ and its isotropic counterpart $f_{\rm sph}(v) = (1/4\pi)\int f(\boldsymbol{v}, t)\mathrm{d}\hat{v}$ obtained as an average over the velocity directions $\hat{v}$. The relevant small parameter is $\epsilon = ma\lambda/kT$, measuring the ratio of the work $ma\lambda$ done by the field over the mean free path to the thermal energy $kT$ in the initial state. The condition $\epsilon \ll 1$ ensures the separation of the characteristic timescales of the two mechanisms experienced by an electron: the thermal motion and the field-induced deterministic motion. Denoting by $v_{\rm th} = \sqrt{kT/m}$ the thermal velocity of the electrons, we have indeed $\epsilon = (t_{\rm th}/t_{\rm acc})^2$, where $t_{\rm th} = \lambda v_{\rm th}$ is the mean time between two successive collisions with the scatterers and $t_{\rm acc} = \sqrt{\lambda/a}$ is the acceleration time required for the field to move the electron over the mean free path $\lambda$ starting from rest. The result of the plain weak-field expansion is to evidence its own failure: it shows that the perturbation is singular insofar as the asymptotic state will be fully dominated by the field, with no memory of the initial temperature. Multiple-scale method is here implemented with respect to the time variable, introducing new independent variables $(\tau_i)_{i>0}$ such that the physical situation corresponds to the line

$$\tau_0 = tv_{\rm th}/\lambda,\ \tau_1 = \epsilon\tau_0,\ \tau_2 = \epsilon^2\tau_0,\ \ldots,\ \tau_n = \epsilon^n\tau_0,\ \ldots \quad (\epsilon = ma\lambda/kT) \qquad [6]$$

The time-derivative expansion [2] is supplemented with an expansion of the velocity distribution:

$$f(\boldsymbol{v}, t) = \sum_{i\geq 0} \epsilon^i F^{(i)}(\boldsymbol{v}, \tau_0, \tau_1, \ldots, \tau_n, \ldots) \qquad [7]$$

The procedure is conducted as exposed in the general case. Identifying term-wise the coefficients of the expansion yields a hierarchy of equations for the $(F^{(i)})_{i\geq 1}$, each supplemented with a solubility condition preventing the appearance of secular divergences. A detailed presentation can be found in Piasecki (1993). The benefit of the multiple-scale method is to yield *jointly* the different stages of the gas evolution, starting from thermal equilibrium and switching on the field at $t = 0$:

- at times $\tau = \mathcal{O}(1)$, an initial transient with a drift velocity $\langle v_z\rangle(t) = at - C_1 at^2 v_{\rm th}/\lambda + \cdots$ in the direction of the applied field (denoting $C_1$ some numerical constant);
- at times $\tau = \mathcal{O}(1/\epsilon)$, a linear-response regime with a steady drift velocity $\langle v_z\rangle \sim a\lambda/v_{\rm th}$; and
- at times $\tau = \mathcal{O}(1/\epsilon^2)$, a long-time field-dominated heating of the gas, where the velocity distribution is no longer Maxwellian, and the kinetic energy of the electrons grows without bounds as $t^{2/3}$, whereas the drift velocity slowly vanishes asymptotically: $\langle v_z\rangle \sim (\lambda^2 a/t)^{1/3}$.

### Domains of Application of the Multiple-Scale Method

The multiple-scale method was first developed in nonlinear mechanics. It is fruitful and is even required in any instance where plain perturbation expansion is not uniformly convergent, more generally when it is necessary to account jointly for variations at different timescales: resonant wave interactions, for example, in plasmas, or in the case of oscillations with slowly varying coefficients. Multiple-timescale method was applied, around 1960, to get kinetic equations (closed equations for the one-particle distribution) from molecular dynamics (Liouville equation) for dilute gases, plasmas, or to establish a microscopic theory of Brownian motion from molecular dynamics of a hard-sphere system (see the section "Microscopic theory of Brownian motion"). In the same spirit, it allows to relate constructively different mesoscopic descriptions, for example, in the case of Brownian motion, to relate the Kramers equation for the distribution $P(\boldsymbol{r}, \boldsymbol{v}, t)$ to the Smoluchowski equation for $P(\boldsymbol{r}, t)$ (see the section "Mesoscopic theory of Brownian motion"). Other examples are the determination of transport coefficients (friction, viscosity) from kinetic description or, at macroscopic scale, the determination of eddy viscosity and eddy diffusivity (see the section "Effective diffusivity for a passively advected scalar"). A last domain of application concerns systems where relaxation processes at different scales superimpose, requiring to handle jointly different time dependences. Multiple-scale method then displays the physics of the relaxation process and its associated hierarchical structure (e.g., the application to the adiabatic piston problem discussed in this Encyclopedia by

Gruber and Lesne – *see* Adiabatic Piston; see also the section "Some typical applications").

## A Brief Overview of Multiscale Approaches

### Different Scales and Regimes

Common to all multiscale approaches is the *focus on the very existence of different scales*, exploited through the use of rescaled variables, which makes explicit the presence of a small parameter $\epsilon$ controlling the dynamics, responsible for the existence of different timescales and related to the scale separation. Technically, the first, very simple but essential, step is to replace the variables, fields, and parameters by their dimensionless counterparts. So doing, small parameters reflecting scale separation (in time, space, energies, amplitudes, ...) will naturally appear. Although it is thus possible to estimate the order of the different terms, it is to be underlined that *it gives no clue on their actual contribution to the long-term behavior*: in singular situations, precisely those where multiscale approaches have to be developed, small terms can have a noticeable influence at all scales. As illustrated in the following sections, different rescalings of variables and functions allow us to discriminate features at different scales and to capture different regimes. More specifically, the techniques to manage with the joint contributions of several regimes at different timescales depend on the way these regimes intermix. They can be:

- *superimposed* regimes, when fast and slow dependences intermingle in the evolution of the same variable. It is the framework of multiple-scale analysis. The solution writes typically $x(t, \epsilon t, \epsilon^2 t, \ldots)$; or
- *coexisting* regimes, namely a coexistence of fast and slow evolutions. One might focus either on the fast evolution and use a quasistatic approximation (or parametric approximation) for the slow evolution, either on the slow evolution and use a quasistationary approximation or an averaging of the fast evolution. The solution writes typically $[x_{\text{fast}}(t), x_{\text{slow}}(\epsilon t)]$ (or $[x_{\text{fast}}(\tau/\epsilon), x_{\text{slow}}(\tau)]$ if the observation takes place at long timescales, with a relevant time variable $\tau = \epsilon t$); or
- *successive* regimes, when initial conditions, bulk behavior and asymptotics are not of the same order with respect to $\epsilon$; this is a boundary-layer-like issue, and the solution writes typically $x_{\text{layer}}(t/\epsilon)$ for $0 \le t \le t_0$, then $x_{\text{bulk}}(t)$ for $t \ge t_0$, with $t_0 = \mathcal{O}(1)$.

Applications are innumerable; the most typical and investigated ones are the climate (from "hours" for the observed weather to "thousands of years" for eras), population dynamics, coasts and sand dunes (from "grains" to "country" scales), protein folding (the vibration of covalent bonds occurs at scale of femtoseconds, while the whole folding may require up to a few seconds), or trading markets (from seconds to years). Let us finally give two typical examples for the parameter $\epsilon$:

- *The weak-damping and high-friction limits*, best explained on an example. The damped oscillator $m\ddot{x} + \gamma\dot{x} + V'(x) = 0$ appears as an *Hamiltonian dynamics* $m\ddot{x} + V'(x) = 0$ as soon as the damping can be neglected, when the characteristic time $\theta = [m/V''(0)]^{1/2}$ of the undamped oscillator is far smaller than the damping time $\tau = m/\gamma$. The weak-damping limit is thus defined as $\epsilon \to 0$, where $\epsilon = \theta/\tau = [\gamma^2/mV''(0)]^{1/2}$. It leads to a singular behavior when investigating the asymptotics, as in the Duffing oscillator and weakly damped oscillator mentioned in the last section. On the contrary, the evolution appears as a *dissipative gradient dynamics* $\dot{x} = -V'(x)/\gamma = 0$ as soon as $\tau \ll \theta$. This leads to the high-friction limit: $\tau/\theta = [mV''(0)/\gamma^2]^{1/2} \to 0$. This example somehow reconciles conservative and dissipative dynamics, showing that they might coexist in the same system.
- *The hydrodynamic limit* involved in the derivation of hydrodynamics equations (namely incompressible Navier–Stokes equations) from kinetic Boltzmann equation. It writes $\epsilon = \lambda/L \to 0$, where $\epsilon$ is the so-called *Knudsen number*, defined as the ratio of the mean free path $\lambda$ (the average distance traveled by a fluid molecule between two successive collisions) to a characteristic spatial scale $L$ of the system (e.g., the size of an obstacle).

### Bridging the Scales: Mean-Field, Singular and Scaling Approaches

The aim of multiscale approaches is to bridge different scales, through the determination of the large-scale behavior of the solution, or by establishing a constructive relation between the initial model and an effective model at higher scale. We have mentioned in the introduction a first classification of multiscale systems and associated approaches: they might exhibit (1) scale decoupling, (2) some singularity in the relation between the different scales, or (3) scale invariance.

**Mean-field approaches** In case of scale decoupling, mean-field approaches apply. Let us briefly recall,

within its usual spatial formulation, that a mean-field approach amounts to identifying the local environment, which is *a priori* fluctuating and spatially inhomogeneous (e.g., the local magnetic field generated by neighboring spins in a spin lattice model) with the average one, expressed as a function of the average order parameter (spatial average or equivalently a statistical average in the limit as the system size tends to infinity). Mean-field approaches can be implemented either in time (*averaging*), in real space (*homogenization*, *coarse-graining*), or in phase space (*aggregation and projection techniques*).

In the present context, the best example of a mean-field approach is provided by *homogenization* procedures. They can be traced back to the method of Lagrange to solve the three-body problem. The issue is to describe the motion of a light body $B_2$ experiencing the gravitational attraction of the Sun and a heavier body $B_1$. The mass of $B_2$ is supposed to be small enough to neglect its influence on the Sun and $B_1$ (the so-called restricted three-body problem); $B_1$ will thus obey the Keplerian laws of motion. The method of Lagrange applies when $B_2$ is far more distant from the Sun than $B_1 (r_2 \gg r_1)$, which implies (due to the third law of Kepler: $\omega^2 r^3 = \text{const.}$) that the angular velocity $\omega_1$ of $B_1$ is far larger than $\omega_2$: the large body $B_1$ moves faster than $B_2$ around the Sun. In first approximation, Lagrange replaced the rapidly oscillating influence of $B_1$ on the motion of $B_2$ by the influence of a constant distribution of mass, obtained by spreading the mass $m_1$ of $B_1$ all over its orbit. The Gauss theorem thus states that this influence can be accounted for by simply adding the total mass of this distribution to the mass of the Sun. The stability of the system would follow: $B_2$ will remain trapped in the neighborhood of the pair composed with the Sun and $B_1$.

**Singular perturbations** A typical instance of singular multiscale behavior is associated with asymptotic expansions

$$x(t) = \sum_{r=0}^{n-1} \epsilon^r x_r + R_n(\epsilon, t) \qquad [8]$$

which are *not* convergent: $\lim_{n\to\infty} R_n(\epsilon, t) \neq 0$ at $\epsilon$ fixed, but $\lim_{\epsilon\to 0} \epsilon^{-n} R_n(\epsilon, t) = 0$ at fixed $n$ and $t$. Asymptotic expansions are ubiquitous in multiscale approaches: the coexistence of different timescales, superimposed and nontrivially coupled to get rise to the observed phenomenon, prevents from obtaining uniformly convergent perturbative expansions; it is only in this latter regular case that the above-mentioned mean-field approaches and homogenization techniques apply.

**Scale invariance, scaling theories and renormalization** Self-similarity and associated criticality prevent scale decoupling, but allow us to develop scaling theories and renormalization methods. In contrast to scale-separation arguments, the guiding principle is now to focus on the links relating one scale to the others (scaling transformations, renormalization transformations). The problem complexity is thus reduced in a some "transverse way," by retaining only scale-invariant features. We shall expose in the section "Renormalization: an iterated multiscale approach" further links between multiscale approaches and renormalization methods, beyond the restricted scope of scale-invariant systems: in many instances, renormalization can be seen as an iterated multiscale approach.

## Scaling Limits

Let us mention a specific instance of multiscale approach, which is associated with scaling limits. Scaling limit refers to a joint limiting procedure, in which several independent variables *jointly* converge towards given limits, with *prescribed relative behaviors*; this latter condition is a key point in the frequent case when the different limits do not commute, and we shall see later that it is an essential ingredient of renormalization methods. Let us cite two acknowledged examples:

- *The thermodynamic limit* for a system of $N$ particles in a volume $V$; it amounts to let $N \to \infty$, $V \to \infty$, while $N/V = n = \text{const.}$ (constant average number density). It is a prerequisite to derive standard thermodynamic behavior from the statistical–mechanical description; it supports the use of asymptotic results given by the law of large numbers and the central-limit theorem provided the correlations between the particles remain short-range.
- *The Boltzmann–Grad limit* for a system of $n$ hard spheres of radius $\epsilon$ per unit volume. In dimension $d$, it writes $\epsilon \to 0, n \to \infty$ (thus differing from the thermodynamic limit) while $n\epsilon^{d-1} = z$ remains constant. This limit is involved in kinetic theory as a limiting instance where the Boltzmann ansatz applies (identifying the two-particle distribution function with the product of the corresponding one-particle distributions). Indeed, the occupied volume fraction $n\epsilon^d$ tends to 0 so that recollisions and ensuing long-term correlations can be neglected (rarefied gas). On the other hand, the mean free path of a particle remains finite, so that numerous collisions and associated molecular chaos further support the Boltzmann decorrelation ansatz.

### Stochastic Multiscale Approaches

Multiscale approaches are far less developed for stochastic processes. Let us mention the case of a Markov process. Scale separation reflects in a spectral gap in the transition matrix generating the dynamics. Identification of fast and slow modes is then straightforward: slow modes are associated with quasidegenerated eigenvalues ($\lambda \approx 0$ in a time-continuous setting), whereas fast dynamics is associated with damped modes and negative eigenvalues ($\lambda < 0, |\lambda| \gg 1$) (Gaveau *et al.* 1999). A basic difficulty in extending methods developed in a deterministic context is the fact that the reduction (or projection) of a Markov process is *a priori* no longer Markovian. Closure relations and approximations should be introduced to circumvent memory effects, for example, supported by arguments of decorrelation and ensuing fast temporal self-averaging of the fast dynamics.

It is to note that the behavior upon rescaling of a stochastic process differs from the transformation of a deterministic evolution. The basic relation is the scaling upon a time rescaling $\theta = \epsilon t$ of the white noise involved in stochastic differential equations and defined from the Wiener process $W(t)$ through the relation $\mathrm{d}W(t) = \eta(t)\mathrm{d}t$. It follows from the definition $\widetilde{W}(\theta) = W(t)$ that $\mathrm{d}\widetilde{W}(\theta) = \sqrt{\epsilon}\,\mathrm{d}W(t)$. At this point, it is important to notice the difference with respect to the behavior of a plain deterministic function $\widetilde{f}(\theta) = f(t)$ for which $\mathrm{d}\widetilde{f}(\theta) = \epsilon\,\mathrm{d}f(t)$. Using the fact that $\delta(t) = \epsilon\delta(\theta)$ and the definition $\mathrm{d}\widetilde{W}(\theta) = \widetilde{\eta}(\theta)\mathrm{d}\theta$, we obtain that $\widetilde{\eta}(\theta)$ is a white noise with respect to the rescaled time $\theta$, that is, a stationary Gaussian process defined by its first two moments

$$\langle \widetilde{\eta}(\theta) \rangle = 0, \qquad \langle \widetilde{\eta}(\theta)\widetilde{\eta}(\theta') \rangle = \delta(\theta - \theta') \qquad [9]$$

## Slow/Fast Variables

### Slow/Fast Decomposition

Dynamics of systems made of many interacting elements, for example, chemical reactions, or population dynamics, typically involves far too many degrees of freedom to be handled at the level of individual units, and requires a drastic reduction to make sense of it. A natural way of reduction is based upon the phenomenology, taking as relevant degrees of freedom those describing the slow evolution observed at macroscopic scales. Scale separation between microscopic and macroscopic worlds has to be turned into a constructive and quantitative argument to achieve this reduction.

Solving this typical multiscale issue first requires *to identify and construct explicitly the slow variables*, for example, collective variables obtained through aggregation or coarse-grainings. The second step is *to eliminate or rather integrate the fast dynamics* into a closed system of effective equations describing the large-scale evolution. The closure requirement generically involves an approximation, neglecting the remaining dynamic coupling between fast and slow variables. It is precisely here that scale-separation arguments and the very choice of the slow variables are crucial, ensuring that the influence of fast dynamics is essentially accounted for in its effective or average contribution to the slow dynamics; remaining fluctuating influences can be either neglected or included in a noise term, required to be fully determined as a function of the slow variable only (otherwise the whole procedure would neither be consistent nor useful). In the following subsections, we shall briefly present the main techniques allowing to achieve this program, considering the simple abstract system:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = f(X, Y), \frac{\mathrm{d}Y}{\mathrm{d}t} = \epsilon\, g(X, Y), \quad (\epsilon \ll 1) \qquad [10]$$

Although involving only two variables for simplicity, it exhibits the typical multiscale structure: whereas $X$ varies on scales $\mathcal{O}(1)$, $Y$ appears as a slow variable of characteristic timescale $\mathcal{O}(1/\epsilon)$.

### Parametric Approximation

The preliminary step of the reduction is to get some knowledge on the fast dynamics, at least to choose the proper multiscale technique. A plain but nevertheless fruitful remark is that a parameter $p$ can always be seen as a variable that does not evolve: $\mathrm{d}p/\mathrm{d}t = 0$ in a deterministic setting, or $W_{p\to q} = \delta(p - q)$ in a stochastic one (transition probability $W$). Conversely, a slow variable can be transiently treated as a mere parameter in the fast dynamics. Supported by timescale separation, this *parametric approximation* (or *quasistatic* approximation) decouples the fast dynamics from the slow variable evolution, investigating the fast dynamics asymptotics ($t \to \infty$) while considering that the slow variable remains constant $Y(t) \equiv y$. In the following, we shall distinguish two cases: (1) the fast dynamics oscillates with a period $T \ll 1/\epsilon$, and (2) the fast dynamics relaxes to a stable equilibrium point $X^*(y)$ slaved to the slow variable.

### Amplitude Equations

A ubiquitous technique to account for *slowly modulated oscillations* has been introduced first by

Fresnel for light propagation and optical phenomena. The basic idea is to take benefit from the scale separation between the fundamental oscillation (frequency $\omega$, wavelength $\lambda = 2\pi/k$) and a superimposed slow variation of the wave amplitude

$$\mathcal{A}(\boldsymbol{r}, t) = A(\boldsymbol{r}, t)e^{i(\boldsymbol{k}.\boldsymbol{r} - \omega t)} \qquad K \equiv |\nabla A/A| \ll k, \qquad \Omega \equiv |\partial_t A/A| \ll \omega \quad [11]$$

The evolution can be rewritten in terms of the slowly varying amplitude $A$; by construction, it is ruled by terms involving the small parameter $\epsilon \sim K/k \sim \Omega/\omega \ll 1$, but the resulting equation is now devoid of small or large parameter. Such technique has been successfully applied and further developed, for example, in various situations involving *electromagnetic waves* (e.g., diffraction of Hertzian waves), in *plasma physics* (resonant interaction between electromagnetic waves and acoustic modes) and in *quantum mechanics*, to investigate the deformation of a wave packet in a potential.

### Averaging

Let us discuss further, in a general setting, the case when the fast dynamics is an oscillation of period $T$ (either linear modes as in the last subsection or a stable limit cycle). It is a context where *averaging techniques* apply. We refer to the associated entry in this Encyclopedia by Neishtatdt (see the article Averaging Methods) and only mention here the main principle: to exploit scale separation and self-averaging property of the fast dynamics to replace $X(t)$ by an average value

$$X_{\mathrm{av}}(t) = (1/T)\int_t^{T+t} X(s)\mathrm{d}s$$

The underlying idea is that averaging cancels out most of the fast variations so that $X_{\mathrm{av}}(t)$ is now slowly varying. In case when the fast dynamics is influenced by the slow variable $Y$, its value is kept constant in the averaging (see the section "Parametric approximation"). The resulting average behavior $X_{\mathrm{av}}[Y(t), t]$ is reinjected in the evolution of the slow component, leading to a closed equation,

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = \epsilon\, g(X_{\mathrm{av}}[Y(t), t], Y)$$

or rather [12]

$$\frac{\mathrm{d}\widetilde{Y}}{\mathrm{d}\tau} = g\left(\widetilde{X}_{\mathrm{av}}[\widetilde{Y}(\tau), \tau], \widetilde{Y}\right)$$

in terms of the more relevant rescaled time variable $\tau = \epsilon t$ and $\widetilde{Y}(\tau) \equiv Y(t)$. Denoting $\bar{Y}(\tau)$ the solution of this approximate equation, the validity of the averaging procedure is assessed by theorems giving conditions ensuring that $\lim_{\epsilon\to 0} \widetilde{Y}_\epsilon(\tau) = \bar{Y}(\tau)$. Note that such theorems (quite unusually) state the convergence, for a vanishing value of the perturbation parameter $\epsilon$, of the exact solutions towards the approximate one (solution of the average equations).

To conclude, let us notice that one speaks of averaging in temporal context and homogenization in spatial or spatio-temporal contexts, when averaging is performed over space; as discussed in the section "Bridging the scales: mean-field, scalar, and scaling approaches," averaging and homogenization belongs to the general class of mean-field approximations.

### Quasistationary Approximation

Let us now consider the case when the fast dynamics converges at fixed $Y$ towards a stable fixed point $X^*(Y)$. Focusing on the slow dynamics, the relevant time variable is $\tau = \epsilon t$, which turns the evolution [10] into

$$\epsilon\frac{\mathrm{d}X}{\mathrm{d}t} = f(X, Y), \qquad \frac{\mathrm{d}Y}{\mathrm{d}t} = g(X, Y) \quad [13]$$

(for the sake of simplicity, we use the same notation $X$ for both $X(t)$ and $\widetilde{X}(\tau)$). It is solved in two steps, by noticing that at lowest order in $\epsilon$, the fast dynamics reduces to the asymptotic regime $f(X, Y) = 0$, slaved to the slow variable $Y$. The corresponding stable state $X^*(Y)$ is then plugged into the slow dynamics to get a closed equation for $Y(\tau)$:

$$\frac{\mathrm{d}Y}{\mathrm{d}\tau} = g[X^*(Y), Y] \equiv G(Y) \quad [14]$$

This achieves the desired dimensional reduction. It works equally well when $X$ is a string of variables $X = (x_1, \ldots, x_N)$.

There is seemingly a paradox here, ubiquitous in many multiscale approaches: in order to determine the evolution of the slow variable $Y$, it is considered a constant! The solution lies in scale separation: the trick is to consider the ensuing approximate decoupling as an exact one (what it would be in the limit $\epsilon \to 0$). In other words, the constancy of $Y$ is considered over a time length which is long at the level of fast dynamics ($\Delta t \gg 1$), long enough for $X$ to reach its equilibrium state $X^*(Y)$, but short at the macroscopic level ($\epsilon\Delta t = \Delta\tau \ll 1$). As in the so-called "quasistatic evolutions" encountered in thermodynamics, the large-scale evolution will be composed of a continued succession of local equilibrium states: at each time $\tau, X$ takes its instantaneous equilibrium value, slaved to $Y(\tau)$. Here one speaks equivalently of *quasistationary*

*approximation*, *quasisteady-state approximation*, or *adiabatic elimination of fast variables*.

### Slow Invariant Manifolds

In the previous subsections, the decomposition between fast variables $X$ and slow variables $Y$ was given. But in practice, only the whole dynamics of the system is known and a main part of the issue is to find and construct explicitly the slow variables.

A geometrical viewpoint on the dynamics appears to be fruitful: if the system evolution is to be reducible to the evolution of a few degrees of freedom, it means that the flow essentially lives in a low-dimensional region of the phase space, which can be parametrized by these degrees of freedom up to some fuzziness of order $\mathcal{O}(\epsilon)$. Mathematical investigations have been conducted to assess this point, leading to the concept of *invariant slow manifold*: a manifold $\mathcal{M}$ of the phase space, invariant upon the dynamics and describing the slow dynamics once the system has reached it (Gorban *et al.* 2004). Starting from an arbitrary point $z_0$, the trajectory first exhibits a fast transient bringing the system state close to $\mathcal{M}$, up to some tolerance of order $\mathcal{O}(\epsilon)$, then sticks to $\mathcal{M}$. Its evolution on $\mathcal{M}$ is ruled by a reduced dynamics, far slower than the fast relaxation to $\mathcal{M}$ as soon as the system actually exhibits a timescale separation. This latter self-consistent assertion should be considered as a working hypothesis, to be validated by the explicit determination of $\mathcal{M}$ and associated reduced dynamics. This can be done numerically, by exploiting the presumed convergence property of any trajectory reaching $\mathcal{M}$ after some intrinsic transients. In other words, if the dynamics possesses a slow invariant manifold, an operational way to find $\mathcal{M}$ is to let the system evolve, starting from a sample of initial conditions, and to observe its stabilization on $\mathcal{M}$.

This framework obviously embeds the quasistationary approximation presented in the last subsection: in this case, the slow invariant manifold is $\mathcal{M}=\{z=(x,y), f(z)=0\}=\{(x^*(y),y)\}$ and the dynamics restricted to $\mathcal{M}$ is the slow dynamics $\mathrm{d}y/\mathrm{d}\tau = G[y(\tau)],\ x(\tau)=x^*[y(\tau)]$. Here the manifold is invariant upon the approximate dynamics (for all $t, f[z(t)]=0$, hence $z(t)\in\mathcal{M}$) but not upon the original one: some rigorous mathematical work has to be done to show that the actual dynamics keeps the trajectory in a proper neighborhood of $\mathcal{M}$ of width $\mathcal{O}(\epsilon)$. In other words, one has to control the discrepancy between the exact trajectory and the trajectory slaved on $\mathcal{M}$.

### Central Manifold

The notion of slow invariant manifold generalizes older results about *central manifolds*, exploited to reduce the dynamics *near a bifurcation point*. Let us consider a dynamical system $\dot{x}=f(x,\alpha)$ near a bifurcation point: in $\alpha=\alpha_c$, the fixed point $x_0$, stable for $\alpha<\alpha_c$, loses its stability. This reflects on the largest eigenvalue(s) of the stability matrix $Df(x_0,\alpha)$, namely $\lambda_1(\alpha)<0$ for $\alpha<\alpha_c, \lambda_1(\alpha)>0$ for $\alpha>\alpha_c$, and $\lambda_1(\alpha_c)=0$. The small parameter is then $\epsilon=\lambda_1$. A main result was to show that, near the bifurcation point, slow modes coincide with unstable directions and fast modes with stable directions (Haken 1996). The decomposition into slow and fast variables is ruled by the *central manifold theorem*: the solutions can be expressed in terms of the amplitudes along the eigenvectors of the null space of the dynamics at $\epsilon=0$; these amplitudes appear as the relevant *order parameters* near the bifurcation. This is referred to as the *slaving principle*. Compared to the setting presented in the subsection "Slow invariant manifolds," the slow invariant manifold $\mathcal{M}$ is given here by the central manifold.

### Projection Techniques

The methods presented in the previous subsections to eliminate fast variables and construct a reduced slow dynamics can be unified into a common framework: *Mori–Zwanzig projection techniques*. The full state $(x, y)$ of the system is projected onto the slow variable $y$ and the functions $w(x, y)$ are projected onto their conditional expectation

$$\mathcal{P}w(y) \equiv \int w(x,y)\rho(x|y)\mathrm{d}x \qquad [15]$$

The core of the method lies in the choice of conditional distribution $\rho(x\,|\,y)$, for instance, $\rho(x\,|\,y)=\delta(x-x^*(y))$ in case when there is an invariant manifold $x=x^*(y)$, or $\rho(x\,|\,y)=1/2\pi$ in case of averaging over a rapidly varying phase $x$. We refer to Givon *et al.* (2004) for a review.

### Aggregation Techniques and Coarse-Grainings

An intuitive guideline in the analysis of a multiscale dynamics is that *collective variables or coherent states coincide with slow modes*. The rationale is that numerous fast fluctuations at the level of agent dynamics self-average, so that only a slow trend is perceptible at large scale. *Aggregation methods* have been developed in this spirit to build reduced models governing the slow dynamics. Nevertheless, in generic situations, aggregation does not lead to

closed equations for the collective variables and some level of approximation has to be introduced.

Let us now consider a system of $N$ coupled degrees of freedom, $[x_i(t)]_{I=1\ldots N}$ (e.g., a system of $N$ interacting agents) evolving deterministically according to a two-scale dynamics (Auger and Bravo de la Parra 2000):

$$\epsilon \frac{\mathrm{d}x_i}{\mathrm{d}t} = f_i(x_1, \ldots, x_n) + \epsilon g_i(x_1, \ldots, x_n) \qquad [16]$$

where $f$ describes a fast evolution due to the coupling between species and $g_i$ a slow evolution due to internal mechanisms. A natural choice for the slow variable is $Y(x_1, \ldots, x_n) = \sum_i x_i$, but we shall write below the general case. The self-consistent requirement of the method is that this variable $Y$ reflect a global and slow behavior. Considering $t$ as a fast time variable, this condition amounts to require a quasistatic behavior for $Y$ at this timescale. In other words, the consistency condition requires that there exists a manifold $\mathcal{F}_y$ such that

$$\sum_{i=1}^{N} \frac{\partial Y}{\partial x_i}(x_1, \ldots, x_N) f_i(x_1, \ldots, x_N) = 0$$
$$\text{on } \mathcal{F}_y = \{Y(x_1, \ldots, x_N) = y\} \qquad [17]$$

We, moreover, assume that the fast dynamics on this manifold $\mathcal{F}_y$ leads to a stable equilibrium $(x_1^*(y), \ldots, x_N^*(y))$. We are then in a position to describe the slow evolution of the manifold itself, that is, the slow dynamics ruling the evolution of the aggregated variable $y$ for $\epsilon$ small enough:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \sum_i \frac{\partial Y}{\partial x_i}\left[x_1^*(y), \ldots, x_N^*(y)\right]$$
$$\times\, g_i\left[x_1^*(y), \ldots, x_N^*(y)\right] + \mathcal{O}(\epsilon) \qquad [18]$$

Internal support of the procedure is to check the structural stability of this resulting aggregated dynamics. Compared to the quasistationary approximation and slaving principle presented earlier, here the slow variable is not given independently but constructed as a function of the fast variables (aggregated variable). The same principles can also be implemented for discrete-time models.

*Coarse-graining* can be seen as the spatial analog of aggregation techniques developed in the phase space: the real space is split into cells considered as elementary units at macroscopic scale, and all the small-scale physics is averaged over each cell, yielding the apparent state of each unit (described by a few "coarse-grained" variables) and the effective interactions between them.

Let us cite two hydrodynamic examples. *Eddy viscosity* refers to an effective viscosity involved in coarse-grained hydrodynamics equations; the contribution of small-scale turbulent structures is accounted for in an integrated way in this parameter, hence its name. It is typically lower than bare viscosity, even possibly reaching negative values at large enough Reynolds number, that is, at low enough bare viscosities. *Cellular flows* are space-periodic flows, thus exhibiting a natural spatial scale: the coarse-graining amounts to an intrinsic homogenization over each cell of the flow.

Let us finally mention that coarse-grainings are involved in renormalization-group transformations once supplemented with the adequate rescalings (see the section "Renormalization: an iterated multiscale approach").

In conclusion, it is to note that all these various multiscale approaches are closely related and can all be expressed as a specific projection technique in the extended phase space containing both fast and slow variables. For instance, aggregation techniques replacing the fast variables $(x_1, \ldots, x_n)$ by the slow collective variable $y = Y(x_1, \ldots, x_n)$ amount to the projection technique involving the slow invariant manifold $\mathcal{M} = \{(x_1, \ldots, x_n, y) \mid y = Y(x_1, \ldots, x_n)\}$.

### Numerical Aspects

In the community of applied mathematics, multiscale methods refer specifically to numerical homogenization, involving *multigrid algorithms* as, for instance, multiscale finite-element method, multigrid Monte Carlo, multigrid optimization, or annealing. Basically, the idea of numerical homogenization is to avoid the numerical cost of using a mesh of size $h < \epsilon$, where $\epsilon$ is the scale of the smallest-scale features of the dynamics, and to use jointly:

- a fine mesh, to compute local quantities independently (hence with a parallelized program); and
- a coarse mesh, to compute global behavior using effective parameters and homogenized quantities determined in the prior fine-mesh computation.

We refer to Gorban *et al.* (2004) for a review.

## Boundary Layers and Matched Expansions

### Purposes and Principles

Multiscale approach to handle boundary layers was introduced in 1905 by Prandtl in fluid mechanics for situations where the solution of hydrodynamics equations far from the boundaries ("bulk" solution) does not match the conditions at the surface of the walls or obstacles. This typically originates in the presence of a multiplicative small factor $\epsilon$ in front of

the highest-order derivative; accordingly, the flow exhibits two different scales in space: a thin boundary layer of width controlled by $\epsilon$ and the bulk domain. The idea is to perform two different perturbation methods in the layer and in the bulk, involving a different rescaling in order to focus on and give the ruling place to either the boundary conditions or the bulk dynamics (one also speaks of *inner* and *outer* expansions). Then these parallel perturbation expansions have to be bridged into a single global continuous solution. The matching principle is to identify the asymptotic behavior on the boundary side with the boundary condition of the bulk behavior (Nayfeh 1973):

$$\lim_{r\to 0} X_{\text{bulk}}(r) = \lim_{\zeta\to\infty} X_{\text{layer}}(\zeta) \quad \text{with } \zeta = r/\epsilon \qquad [19]$$

Boundary layers of hydrodynamics have numerous analogs: *initial layers* in chemical kinetics, *skin layers* in electrodynamics and *edge layers* in solid-state physics (Nayfeh 1973). Adaptation of this technique is to be developed to determine the complete dynamics in the slow-invariant-manifold approach, matching the fast relaxation towards the manifold with the slow motion onto the manifold. Let us finally note that the matched-expansion approach can benefit in each region of all the above-mentioned multiscale techniques.

### Time Analog: Implementation for Initial Layers

We shall now work out the time analog of a boundary-layer problem on the abstract example encountered in [10], in the case when $X$ rapidly evolves to a slaved equilibrium state $X^*(Y)$ but with initial conditions $Y(0) = y_0$ and $X(0) = x_0 \neq X^*(y_0)$. Obviously, the quasistationary approximation fails to describe the initial regime and its applicability has to be reconsidered. The general principle of boundary-layer analysis, namely the recourse to two different perturbation approaches, is implemented as follows:

- For the initial regime, one solves the fast dynamics with initial conditions $X(0) = x_0$ while keeping $Y(t) \equiv y_0$; this yields an approximate solution $[X_{\text{layer}}(t), Y_{\text{layer}}(t)]$, satisfying the initial conditions and valid at short times, as long as $Y$ has not evolved.
- At longer times, the relevant variable is the rescaled time $\tau = \epsilon t$ and the quasistationary approximation described in the last section applies.

The consistency of the two perturbative approaches is ensured by the *matching conditions*

$$\begin{aligned} \lim_{\tau\to 0} X_{\text{bulk}}(\tau) &= \lim_{t\to\infty} X_{\text{layer}}(t) \\ \lim_{\tau\to 0} Y_{\text{bulk}}(\tau) &= \lim_{t\to\infty} Y_{\text{layer}}(t) \equiv y_0 \end{aligned} \qquad [20]$$

These conditions are actually satisfied since $X_{\text{bulk}}(\tau) \equiv X^*[Y_{\text{bulk}}(\tau)]$, hence $\lim_{\tau\to 0} X_{\text{bulk}}(\tau) = X^*(y_0)$ and, by definition of $X^*$ (at fixed $Y(t) \equiv y_0$), $\lim_{t\to\infty} X_{\text{layer}}(t) = X^*(y_0)$.

### Some Typical Applications

**Enzymatic catalysis** A matched singular perturbation approach is currently encountered in chemical systems, for instance, in the derivation of the *Michaelis–Menten kinetics* for a single enzyme and the *Hille cooperative kinetics* for an allosteric enzyme (Murray 2002). Denoting by $E$ the enzyme, by $S$ the substrate, by $ES$ the active complex, and by $P$ the product, the single-enzyme catalytic transformation of $S$ into $P$ is described by the following scheme:

$$\begin{aligned} & S + E \underset{k'}{\overset{k}{\rightleftharpoons}} ES \xrightarrow{k_{\text{cal}}} P + E \\ & [S] \equiv s, \quad [E] \equiv e, \quad [ES] \equiv c \end{aligned} \qquad [21]$$

where, as is well known, the enzyme is released at the end. Introducing dimensionless quantities

$$\begin{aligned} & \tilde{t} \equiv k e_0 t, \quad \tilde{s} \equiv \frac{s}{s_0}, \quad \tilde{c} \equiv \frac{c}{e_0} \\ & K_{\text{m}} \equiv \frac{k' + k_{\text{cat}}}{k}, \qquad \widetilde{K}_{\text{m}} = \frac{K_{\text{m}}}{s_0} \\ & \lambda \equiv \frac{k_{\text{cat}}}{k s_0}, \qquad \epsilon = \frac{e_0}{s_0} \end{aligned} \qquad [22]$$

the corresponding chemical kinetic equations can be written as

$$\begin{aligned} \frac{\mathrm{d}\tilde{s}}{\mathrm{d}\tilde{t}} &= -\tilde{s} + \tilde{c}(\tilde{s} + \widetilde{K}_{\text{m}} - \lambda) \\ \epsilon \frac{\tilde{c}}{\mathrm{d}\tilde{t}} &= g(\tilde{s}, \tilde{c}) \equiv \tilde{s} - \tilde{c}(\tilde{s} + \widetilde{K}_{\text{m}}) \end{aligned} \qquad [23]$$

Noticing that $\epsilon \ll 1$ (the enzyme is present in infinitesimal quantities compared to the substrate), a quasistationary approximation applies for the variable $\tilde{c}$: it means that the intermediary species $ES$ rapidly reaches a local equilibrium state $\tilde{c} = \tilde{c}^*(\tilde{s})$. This yields the substrate evolution

$$\frac{\mathrm{d}\tilde{s}}{\mathrm{d}\tilde{t}} = \frac{\lambda \tilde{s}}{\tilde{s} + \widetilde{K}_{\text{m}}} \qquad [24]$$

The initial condition is set only on the substrate: $s(0) = s_0$, that is, $\tilde{s}(0) = 1$. It yields the well-known expression of the velocity $V \equiv (\mathrm{d}s/\mathrm{d}t)_{|t=0}$ as a function of the initial substrate concentration: $V(s_0) = e_0 k_{\text{cat}} s_0 / (s_0 + K_{\text{m}})$ (with a maximal value

$V_{\max} = e_0 k_{\text{cat}}$). The quasistationary value for the complex (dimensionless) concentration $\tilde{c}^*(\tilde{s} = 1) = 1/(1 + \widetilde{K}_{\text{m}})$ at $t = 0$ obviously differs from the actual initial condition $\tilde{c}(0) = 0$: besides, it is quite foreseeable that the transients leading the complex *ES* to its stationary value cannot be described using a quasistationary approximation. At short times, the relevant time variable is the fast rescaled time $\theta = \tilde{t}/\epsilon$, leading to the equation describing the initial regime when supplemented with the actual initial condition $\tilde{c}(0) = 0, \tilde{s}(0) = 1$. The analysis is straightforwardly carried over, exactly as in the general abstract case, with a matching condition $\lim_{\theta \to \infty} \tilde{c}(\theta) = \tilde{c}(t = 0) = 1/(1 + \widetilde{K}_{\text{m}})$.

**Kinetic theory** Time-matched expansions have been developed in kinetic theory, for instance, to describe the fate of a tagged particle within a gas. In a first, short stage (*kinetic stage*) following the injection of the particle in the thermally equilibrated gas, the velocity distribution of the particle rapidly evolves due to collisions with gas molecules and associated momentum transfer. This stage lasts a few mean-free-times and it ends when the tagged-particle distribution is almost Maxwellian. Then, in a second stage (*hydrodynamic stage*), the distribution slowly relaxes towards a spatially uniform distribution, ultimately equal to the equilibrium Maxwell–Boltzmann distribution; at each time, the velocity distribution is almost Maxwellian. The particle dynamics is described at the level of its distribution function by the Boltzmann equation, and the resolution (the so-called Chapman–Enskog method) is based on the above general principles.

**The adiabatic-piston problem** A matched two-timescale perturbation approach has been developed for the *adiabatic piston problem*: an isolated cylinder filled with an ideal gas (noninteracting light particles of mass $m$) is separated in two compartments by a moving piston, of mass $M$, adiabatic in the sense that it has no internal degrees of freedom and does not conduct heat when fixed. The small parameter is the mass ratio $\epsilon = 2m/(M + m)$. It quantifies the efficiency of energy transfer between the gas particles and the piston upon elastic collisions, and the strength of the indirect coupling of the two gas compartments through the collisions of their particles with one and the same piston. The matched perturbation approach gives access both to a fast deterministic relaxation towards mechanical equilibrium, at timescales $\mathcal{O}(1)$, with no heat transfer between the compartments, and a slow fluctuation-driven evolution towards thermal equilibrium, where the heat transfer is achieved by the collision-induced coupling between the gas and the piston fluctuating motion, thus occurring at timescales $\mathcal{O}(M/m)$ (*see* Adiabatic Piston).

## Renormalization: An Iterated Multiscale Approach

It is not the place to expose or even summarize the implementation of renormalization techniques, for which we refer to the associated entries in this Encyclopedia. Here we will only stress the natural relations between renormalization group (RG) and multiscale approaches. The RG approach indeed shares many steps and guiding principles: joint rescalings, coarse-grainings and local averaging, effective parameters and effective terms, relevant and irrelevant contributions, with a focus on large-scale behavior. Moreover, far beyond the scope of the study of critical phenomena, RG has been extended into an *iterated multiscale approach* allowing to determine in a systematic and constructive way the effective equation describing the universal large-scale features and asymptotics of a multiscale system (see, e.g., Chen *et al.* (1996) and Mazzino *et al.* (2004).

It is first to be underlined that different meanings are associated with the term "renormalization," corresponding to very different statuses for the associated renormalization procedures.

A renormalized quantity can be plainly a *rescaled* quantity (normalized, dimensionless or put to the scale of the considered sample): here arises a first connection with multiscale approaches, both involving rescalings as an essential preliminary step.

A renormalized quantity can be an *effective quantity* accounting in an integrated way of complicated underlying mechanisms (e.g., the renormalized mass of a body moving in a fluid, accounting for hydrodynamic effects); here arises another central notion of multiscale approaches: effective parameters or effective equations (following, e.g., from averaging or homogenization).

Renormalization is also a mathematical technique developed first in celestial mechanics, and then mainly in quantum electrodynamics to *regularize* divergent expansions and perturbation series. It might proceed by means of *resummation*; the idea, implemented by Rayleigh in 1917, is to sum up correlations and interactions into a redefinition of the parameters. It might either rely on the introduction of a *cutoff* in the space, time, and energy scales, then accounting in an effective way of the host of contributions at smaller space and time scales $\Delta x \leq \Lambda$, $\Delta t \leq \theta$ (or, equivalently, larger momentum

and frequency scales: $k \geq 2\pi/\Lambda, \omega \geq 2\pi/\theta$) so as to take advantage of the physical cancellation of mathematical divergences. In any case, it turns the bare parameters of the original singular expansion into *renormalized parameters* and yields a renormalized regular expansion. Writing that the resulting large-scale behavior does not depend on the chosen cutoff $(\Lambda, \theta)$ yields *renormalization equations*, expressing quantitatively the very consistency of the procedure ("renormalizability" of the expansion). Renormalization provides alternative technical tools in instances treated above with the multiple-scale method. Its main advantage is its *recursive structure*: introducing a sequence $(\Lambda_n, \theta_n)_n$ of cutoffs (what is called momentum-shell RG), the whole procedure can be iterated to integrate recursively the influence of small-scale features on the asymptotic behavior, allowing as to handle situations exhibiting a hierarchy or even a continuum of scales.

Renormalization also refers to an *asymptotic analysis* allowing as to classify critical behaviors, to determine quantitatively the critical exponents and to handle the associated divergences. Indeed, the above-mentioned multiscale approaches fail near bifurcation points or critical points. In this case, scale separation is replaced by *scale invariance*. The key idea, underlying RG techniques is to shift the focus on the scaling procedure itself. The basic point is to construct a renormalization transformation, consisting in *joint coarse-grainings and rescalings*, thus relating the two models *describing the same phenomenon at different scales* (Lesne 1998); it puts forward their self-similar properties and associated scaling laws, while eliminating specific small-scale details having no consequences on the asymptotic, large-scale behavior. The set of renormalization transformations has a semigroup structure with respect to the rescaling factor (or plainly with respect to iteration) justifying to speak of RG. It generates a *flow in the space of models*, whose fixed points correspond either to trivial or to critical situations according to their stability. It can be shown that the linear analysis of the renormalization transformation around a critical fixed point gives access to the critical exponents. Moreover, this analysis allows us to split the space of models into *universality classes*, each associated to the basin of attraction of a critical fixed point. Let us emphasize that scale invariance leads to a deep change in the modeling and investigations, shifting from a "physics focusing on the prediction of amplitudes" to a "physics of the exponents," focusing on less specific, but more universal and above all, more intrinsic features.

Far more generally, RG is associated with a qualitative change in the questioning, since the study takes place in a space of models. Generalized renormalization transformation can be designed to extract not only self-similarity properties but any large-scale feature from a more microscopic model. In particular, RG can be specially designed to discriminate between essential and inessential terms in a model: the latter do not modify the asymptotics of the RG flow, meaning that they are of no consequence at large scales. In other words, generic properties of the renormalization flow in this space of models yield universal large-scale scaling properties. RG is thus essentially a multiscale approach, insofar as it only retains the relations between the different levels of descriptions, somehow ignoring the details at each given scale. *It is actually designed to capture universal features of the multiscale organization.*

## Summary: The Exemplary Case of Diffusion

### Bridging the Scales

Our aim in this section is to present the whole range of multiscale approaches in use, allowing both to bridge models devised at different scales and to predict the large-scale features of the phenomenon they account for. We choose the context of diffusion, Brownian motion, and transport phenomena, where such a bridge is essential and has been much investigated. Indeed, transport coefficients are defined through phenomenological equations; it is thus necessary to relate such macroscopic equations with smaller-scale theories, so as to get an expression of the coefficients in terms of the microscopic ingredients and to justify the validity of the phenomenological description.

The exposition in the various subsections below, following increasing scales, will mark out the pathway from reversible molecular dynamics to macroscopic diffusion equations. We shall thus come across the multiple-scale analysis of the Liouville equation describing at microscopic scales a Brownian grain suspended in a thermal bath of water molecules (see the next subsection) leading to the mesoscopic Kramers equation for the grain distribution function $P(\mathbf{r}, \mathbf{v}, t)$. Next, involving higher but still mesoscopic scales, we see that another multiple-scale analysis leads to the reduced Smoluchowski equation for its spatial distribution $P(\mathbf{r}, t)$. Random walks offer alternative mesoscopic models, involving effective diffusion coefficients in order to take into account underlying features like persistence length or other short-range correlations. Scaling limits or more systematic renormalization methods in real space allow to bridge discrete random-walk models with continuous descriptions. Another RG, based on

a path-integral formulation in the framework of field theory, allows to handle the case of self-avoiding walks with infinite memory. Homogenization is illustrated on the case of diffusion in a regular porous medium, whereas diffusion processes in fractal substrates provide a counterexample, singular enough to exhibit anomalous scaling behavior. The issue of reducing the dynamics of the diffusion process to a simpler effective one is encountered in many other macroscopic instances, among which we shall mention diffusion in a periodic medium, lending to space averaging, and advection of a passive scalar field in a two-scale velocity field, where a multiple-scale analysis yields the effective diffusivity at large scale. We shall give further technical guidelines for constructing these steps climbing from molecular up to large macroscopic scales, thus providing additional illustrations of the multiscale approaches introduced in the previous sections on more general and abstract grounds.

## Microscopic Theory of Brownian Motion

The first theoretical account of Brownian motion, namely the erratic movement of a micron-sized pollen grain suspended in a thermal bath, for example, water, dates back to 1905 and the famous paper by Einstein. It took almost 60 years before a *microscopic* theory was achieved; this theory has been further worked out using multiple-scale techniques (Cukier and Deutsch 1969). The challenge is to start from the complete deterministic reversible dynamics of the system, described within a probabilistic framework by the Liouville equation $\partial p/\partial t = Lp$ for the distribution of probability $p$ in the whole phase space (position and velocities of the grain, of mass $M$, and all water molecules, of mass $m \ll M$). The small parameter is the mass ratio $\epsilon = \sqrt{m/M}$ measuring the efficiency of the energy transfer upon collisions between the grain and the bath particles, assuming a binary interaction potential $U = \sum_i u(|\boldsymbol{r}_i - \boldsymbol{r}|)$. The Liouville operator is decomposed into $L = L_0 + \epsilon L_1$, and one introduces rescaled time variables $\tau_n = \epsilon^n t$, where $\tau_0 = t$ is the timescale of the fluid particle dynamics. Multiple-scale method is carried out according to the general scheme, leading to the so-called Kramers equation,

$$\left(\frac{\partial}{\partial t} + \boldsymbol{v}.\frac{\partial}{\partial \boldsymbol{r}}\right) P(\boldsymbol{r}, \boldsymbol{v}, t) = \zeta \frac{\partial}{\partial \boldsymbol{v}}\left[\boldsymbol{v} + \frac{kT}{M}\frac{\partial}{\partial \boldsymbol{v}}\right] P(\boldsymbol{r}, \boldsymbol{v}, t) \qquad [25]$$

where the friction coefficient is explicitly given as

$$\zeta = \frac{1}{3MkT}\int_0^\infty \langle \boldsymbol{F}_t.\boldsymbol{F}_0\rangle \, dt \quad \text{where } \boldsymbol{F}_t = e^{iL_0 t}\boldsymbol{F}_0 \text{ and } \boldsymbol{F}_0 = -\nabla_{\boldsymbol{r}} U \qquad [26]$$

We refer to the original, although very pedagogical, paper by Cukier and Deutsch (1969) for a thorough exposition and discussion of this derivation.

## Mesoscopic Theory of Brownian Motion

Multiple-scale method is also of relevance to determine the high-friction limit of the above Kramers equation. Standard perturbation technique with respect to the inverse of friction, $1/\zeta$, fails to describe the asymptotic regime: there is not enough freedom to fulfill all the solubility conditions required to avoid the appearance of secular divergences (Bocquet 1997). By contrast, multiple-scale technique yields a uniform expansion of the evolution equation still valid at long times, thus allowing to bridge two mesoscopic levels of description, namely the Kramers equation and the Smoluchowski equation for the spatial density $\rho(\boldsymbol{r}, t)$ of the Brownian particle:

$$\frac{\partial}{\partial t}\rho(\boldsymbol{r}, t) = \frac{1}{M\zeta}\frac{\partial}{\partial \boldsymbol{r}}\left(kT\frac{\partial}{\partial \boldsymbol{r}}\right)\rho(\boldsymbol{r}, t) \qquad [27]$$

Introducing dimensionless variables $\tau = tv_{th}/l, \boldsymbol{R} = \boldsymbol{r}/l, \boldsymbol{V} = \boldsymbol{v}/v_{th}$, where $l$ is the size and $v_{th} = \sqrt{kT/M}$ the thermal velocity of the grain, the relevant small parameter appears to be the dimensionless inverse of the friction coefficient, $\epsilon = v_{th}/l\zeta$; hence,

$$\epsilon\left(\frac{\partial}{\partial \tau} + \boldsymbol{V}.\frac{\partial}{\partial \boldsymbol{R}}\right) P(\boldsymbol{R}, \boldsymbol{V}, \tau) = \frac{\partial}{\partial \boldsymbol{V}}\left[\boldsymbol{V} + \frac{\partial}{\partial \boldsymbol{V}}\right] P(\boldsymbol{R}, \boldsymbol{V}, \tau) \qquad [28]$$

If the friction is high (i.e., $\epsilon \ll 1$), the velocity relaxes very rapidly towards the equilibrium Maxwell distribution, and it is then enough to describe the (slow) evolution of the spatial distribution $\rho(\boldsymbol{r}, t)$. Nevertheless, the relaxation stage is essential and accordingly the $\epsilon$-dependence is singular, as a rule when the small perturbation parameter multiplies the time derivative.

According to the general procedure exposed in the section "Multiple-scale method: principles," we introduce rescaled variables $\tau_0 = \tau, \tau_1 = \epsilon\tau, \tau_2 = \epsilon^2\tau, \ldots$ considered as independent variables and look for a solution of the Kramers equation of the form $P = P^{(0)} + \epsilon P^{(1)} + \epsilon^2 P^{(2)} + \cdots$, where the arguments of all the components $P^{(i)}$ are $(\boldsymbol{R}, \boldsymbol{V}, \tau_0, \tau_1, \tau_2, \ldots)$. Identifying term-wise the successive powers of $\epsilon$ yields

a hierarchy of equations. At order 0, we obtain $P^{(0)} = \Phi(\boldsymbol{R}, \tau_0, \tau_1, \tau_2, \dots)e^{-V^2/2}$. The following equations, for the $[P^{(i)}]_{i\geq 1}$, involve the linearized operator $\mathcal{L} = \partial_V(V + \partial_V)$. For each of them, there appears a solubility condition, requiring that none of the additive contributions in the equation is an eigenvector of $\mathcal{L}$; involving the components $P^{(j)}$ with $j < i$, it prevents the appearance of a secular divergence in $P^{(i)}$. At order[1], the solubility condition is $\partial\Phi/\partial\tau_0 = 0$, thus determining the (trivial) $\tau_0$-dependence of $P^{(0)}$. In a similar way, the solubility condition at order 2 allows to determine the $\tau_1$-dependence of $P^{(0)}$. This bridges the Kramers and Smoluchowski equations in the high-friction limit, when retaining only the first-order term in $\epsilon$. We refer to Bocquet (1997) for a pedagogical account of the derivation and discussion of its relation with the time-derivative expansion involved in the so-called Chapman–Enskog solution of the Boltzmann equation.

### Random-Walk Model and Weakly Correlated Diffusion

Random walks are discrete-time mesoscopic models, accounting for the diffusing motion of a particle through the statistical properties of its successive steps, when observed at a given timescale $\tau$. The basic model (*ideal random walk*) assumes isotropic, independent and identically distributed steps of variance $a^2$. Central-limit theorem straightforwardly gives the time dependence of the mean-square displacement $R^2(t) \equiv \langle |\boldsymbol{r}(t) - \boldsymbol{r}(0)|^2 \rangle = a^2 t/\tau$, showing that the motion is a normal diffusion, with diffusion coefficient $D = a^2/2d\tau$ in dimension $d$. It is to note (see also the next subsection) that $D$ depends $\tau$ and $a$, but in a *joint manner*. Actually, the diffusion coefficient associated with a diffusive motion observed at scale $a$ and modeled by a random walk on a lattice of parameter $a$ can be written as $D = \alpha a^2$, where the rate $\alpha$ depends on $a$ (effective rate at spatial resolution $a$): this is a sort of renormalization that accounts for the rate $\alpha(a)$ of all microsteps backward and forward of length far smaller than $a$.

In case of short-range correlations between the successive steps (namely if $\sum_{-\infty}^{\infty} |C(t)| < \infty$, where $C(t)$ is the statistical correlation function between elementary steps separated by a time length $t$), direct computations support a time-average-like result: the asymptotic behavior is still described by a normal diffusion law $R^2(t) \sim 2dD_{\text{eff}}t$, with $D_{\text{eff}} = D\sum_{-\infty}^{\infty} C(t)$. When $C(t) = e^{-t/\tau}$

$$D_{\text{eff}} = \frac{D(1 + e^{-1/\tau})}{1 - e^{-1/\tau}}$$

hence $D_{\text{eff}} \approx 2\tau D$ if $\tau \ll 1$.

### Renormalization Analysis in Case of Markovian Diffusion

Trying to bridge lattice random walks with a continuous description brings out the following difficulty: as the step size $a$ goes to 0, one has to obviously decrease the duration $\tau$ accordingly, but by what amount is not so obvious, since the walker velocity is ill-defined (it depends on the observation scale). Determination of the proper joint rescaling can be guessed from the knowledge obtained by another mean about the system; rather, it can also be obtained in a systematic way, thanks to RG methods. Let us explain the basic principle.

Let us denote by $P_{a,\tau}(x, y, t)$ the transition probability governing the random walk, namely the density of probability to jump from $x$ to $y$ in time $t$, where $x$, $y$ are restricted to the lattice $(a\boldsymbol{Z})^d$ and time to $\tau\boldsymbol{N}$. The renormalization transformation $\Phi_{k,\alpha}$ should express the consequence for $P_{a,\tau}$ of a joint rescaling of space (by a factor of $k$) and time (by a factor of $k^\alpha$). Taking into account the Markov character of the walks, we are thus led to define

$$[\Phi_{k,\alpha}P_{a,\tau}](x, y, t) \equiv k^d P_{a,\tau}(kx, ky, k^\alpha t) \quad \text{in dimension } d \qquad [29]$$

The proper value of $\alpha$ is to be determined self-consistently in order that the limit $\lim_{k\to\infty} \Phi_{k,\alpha}\, P_{a,\tau}$ exists (it is then a continuous transition probability $P^*_\alpha(x, y, t)$ defined on $\boldsymbol{R}^d \times \boldsymbol{R}^d \times \boldsymbol{R}$). The root-mean-square displacement

$$R(P,\, t) \equiv \left[\sum_{x,y} |x - y|^2 P(x,\, y,\, t)\right]^{1/2}$$

is transformed according to

$$R(\Phi_{k,\alpha}P_{a,\tau}, t) = k^{-1} R(P_{a,\tau}, k^\alpha t) \qquad [30]$$

Accordingly, it yields the diffusion law associated with the fixed point $P^*_\alpha$:

$$\text{for any } k,\ R(P^*_\alpha, t) = k^{-1} R(P^*_\alpha, k^\alpha t), \quad \text{hence } R(P^*_\alpha, t) \sim t^{1/\alpha} \qquad [31]$$

It is anomalous except if $\alpha = 2$. In the case of ideal random walks, the proper exponent leading to a nontrivial limit is $\alpha = 2$; this limit $P^*_2$ is the transition probability of a *Wiener process*:

$$W_D(x, y, t) = [4\pi dDt]^{-d/2} e^{-(x-y)^2/4dDt} \quad \text{with } D = a^2/2d\tau \qquad [32]$$

This shows that *all ideal lattice random walks belong to the same universality class*, that of the Wiener process. This approach has been fruitfully

applied to diffusion in disordered systems, the issue being to determine whether or not the disorder, accounted for as a noise term in the transition probabilities, modifies the normal diffusion law obtained in the unperturbed situation. Similar reasoning can also be implemented for self-similar anomalous diffusion processes, like *fractional Brownian motions* and *Levy flights* (Lesne 1998).

### Renormalization Analysis for Self-Avoiding Walks

Let us only mention, for the sake of completeness, the renormalization techniques developed for determining the conformational statistics of linear polymer chains, whose three-dimensional shape can be represented as the trajectory of a self-avoiding random walk. These techniques belong to the RG corpus developed in statistical mechanics for critical phase transitions, within a field-theoretic framework. A formal but exact analogy can actually be worked out between self-avoiding walks and a spin lattice system with $n \to 0$, where $n$ is the number of spin components.

The multiscale nature of the system is so marked here that it should rather be qualified as an *absence of characteristic scale*. In this respect, standard RG methods developed for critical phenomena lie at the very boundary of multiscale approaches. Scale decoupling is replaced by scale invariance, which is somehow the conjugate situation: homogeneity in real space is replaced by homogeneity in the conjugate space (space of characteristic scales). Scale invariance here reflects in the self-similar property, $R(N) \sim N^{\nu}$, relating the end-to-end distance $R$ of the chain to the number $N$ of elementary steps (the monomers), with an anomalous exponent $\nu$ (the Flory exponent $\nu \approx 3/5$ in dimension $d = 3$) originating from the infinite memory of the nonoverlapping chain. We refer to Lesne (1998) and references therein for a more detailed exposition of the concepts and techniques only alluded here.

### Effective Diffusion in a Porous Medium (Homogenization)

Describing the diffusion in a porous medium appears as a formidable task at the pore level: it would require us to account for all the boundary conditions at the border of the hollow domain $\mathcal{V} \in \mathcal{V}_0$ actually accessible to diffusion. When the pores have a finite characteristic size $a$, a homogenization approach can be developed at scales far larger than $a$. It allows to account for the slowing down of the motion due to obstacles in an effective diffusion coefficient (in plain words, the black and white medium made of matter and holes of size $a$ appears as a grey homogeneous medium at larger scales). More specifically, a diffusing tracer of random trajectory $\boldsymbol{r}(t)$ experiences a varying coefficient $D[\boldsymbol{r}(t)]$ (it equals $D$ inside the pores, whereas it vanishes in the nonaccessible region $\mathcal{V}_0 - \mathcal{V}$). The idea is to replace this fluctuating realization of the transport coefficient by its spatial average (independent of the trajectory), in what concerns macroscopic properties:

$$D_{\text{eff}} = \int_{\mathcal{V}_0} D\, n_0(\boldsymbol{r})\, \mathrm{d}^d\boldsymbol{r} = \int_{\mathcal{V}} D[\boldsymbol{r}]\, \mathrm{d}^d\boldsymbol{r}$$

$$(\text{where } n_0(\boldsymbol{r}) = 1 \text{ iff } \boldsymbol{r} \in \mathcal{V}) \qquad [33]$$

Rigorous mathematical theorems ensure that the large-scale motion can actually be described by a Fick law and associated plain diffusion equation (Bensoussan *et al.* 1978).

### Anomalous Diffusion in a Fractal Medium

The above homogenization for diffusion in a porous medium works well only if the pores have a finite characteristic size; by contrast, diffusion in a fractal substrate (e.g., a porous medium with pores of all sizes) generically leads to anomalous diffusion, associated with a time dependence of the mean-square displacement $R^2(t) \sim t^{\gamma}$ with $\gamma < 1$. In a fractal substrate, the existence of obstacles and pores of all sizes introduces spatial fluctuations at all scales and long-range correlations in the spatial dependence of $D$. This case corresponds to a critical situation and homogenization fails to give a relevant description of the macroscopic behavior, in the same way as mean-field methods fail to account for critical phase transitions. It reflects in the anomalous exponent $\gamma < 1$ of the diffusion law, that can be related to the fractal characteristics of the substrate ($\gamma = d_{\mathrm{s}}/d_{\mathrm{f}}$, where $d_{\mathrm{s}}$ is the spectral dimension and $d_{\mathrm{f}}$ the fractal dimension).

### Effective Diffusion in a Periodic Potential (Averaging Method)

In case of a periodic medium, where $D[\boldsymbol{r}(t)]$ oscillates with a small spatial period, an averaging procedure can be developed as in the subsection "Effective diffusion in a porous medium (homogenization)," to determine an effective diffusion equation accounting for the large-scale motion. Explicit computations within a multiple-scale approach yield

$$D_{\text{eff}} = \frac{1}{\langle D \rangle} \qquad [34]$$

where $\langle D \rangle$ denotes a space average over the elementary cell (Givon *et al.* 2004).

Let us rather detail the case of diffusion of a Brownian particle in a periodic potential $U$, with $U(x + L) = U(x)$ for any $x$ (restricting to dimension 1 for simplicity), at equilibrium at temperature $T$. Let $D$ be the coefficient of this particle in the

absence of the potential. At large scales $dx \gg L$, the substrate appears to be spatially uniform. The influence of the periodic bias exerted by the potential on the diffusive motion (superimposition of a modulated deterministic drift) can be described in an average way. The result is a normal diffusion with a reduced effective diffusion coefficient

$$D_{\rm eff}(U) = D \inf_{f \in C^{\infty}(LS_1)} \int_0^L |1 - f'(x)|^2 \, dm_U(x)$$

$$\text{with } dm_U(x) = \frac{e^{-U(x)/kT}\, dx}{\int_0^L e^{-U(x')/kT}\, dx'} \qquad [35]$$

where the infimum is taken over the set of smooth periodic functions of period $L$ and the average involves the equilibrium distribution $m_U$ of the particle in the potential landscape $U(\,.\,)$. So doing, one sees in particular that no oriented motion can arise at equilibrium, even if $U$ is asymmetric. The procedure extends to dimension $d$ with only technical differences.

## Effective Diffusivity for a Passively Advected Scalar

Still another fruitful implementation of multiple-scale method is encountered in the context of diffusion and transport phenomena, in the study of the advection by a given incompressible velocity field $\boldsymbol{v}(\boldsymbol{r}, t)$ of a passive scalar field $\theta(\boldsymbol{r}, t)$, for example, the density of small inert "tracer" particles advected by the fluid flow without modifying it back. We consider the case when the fluid motion can be decomposed into a large-scale, slowly varying component and a small-scale, rapidly varying fluctuation: $\boldsymbol{v}(\boldsymbol{r}, t) = \boldsymbol{U}(\boldsymbol{r}, t) + \lambda \boldsymbol{u}(\boldsymbol{r}, t)$. The parameter $\lambda$ controls the relative strength of these components. Another small parameter $\lambda$ is involved in this problem: the ratio $\epsilon = l/L \ll 1$ of the typical length scales $L$ and $l$ of $\boldsymbol{U}$ and $\boldsymbol{u}$, respectively. Here the issue is to bridge two macroscopic descriptions: the full hydrodynamic equation describing the evolution of the scalar field $\theta(\boldsymbol{r}, t)$

$$\frac{\partial}{\partial t}\theta(\boldsymbol{r}, t) + \boldsymbol{v}(\boldsymbol{r}, t).\nabla\theta(\boldsymbol{r}, t) = D\Delta\theta(\boldsymbol{r}, t) \qquad [36]$$

and a large-scale effective transport equation for an average scalar field $\theta_L(\boldsymbol{r}, t)$,

$$\frac{\partial}{\partial t}\theta_{\rm L}(\boldsymbol{r}, t) + \boldsymbol{U}(\boldsymbol{r}, t).\nabla\theta_{\rm L}(\boldsymbol{r}, t) = \frac{\partial}{\partial r_i}\left[D_{ij}^{\rm eff}\frac{\partial}{\partial r_j}(\boldsymbol{r}, t)\theta_{\rm L}(\boldsymbol{r}, t)\right] \qquad [37]$$

This procedure, amounting to account in an average way for the small-scale contributions to the complete hydrodynamic description, relies on a spatio-temporal generalization of the multiple-scale method: it involves rescaled space and time variables, $\boldsymbol{X} = \epsilon\, \boldsymbol{x}, \tau = \epsilon\, t, T = \epsilon^2\, t$ The different characteristic scales of the velocity components are directly reflected in their arguments: $\boldsymbol{u}(\boldsymbol{x}, t)$ and $\boldsymbol{U}(\boldsymbol{X}, T)$. The passive scalar field now expresses $\theta(\boldsymbol{x}, t, \boldsymbol{X}, \tau, T)$ and it is expanded as $\theta = \theta^0 + \epsilon\,\theta^1 + \epsilon^2\,\theta^2$. The standard multiple-scale procedure leads to introduce an auxiliary field $\chi$:

$$\partial_t \chi_j + [(\boldsymbol{u} + \lambda \boldsymbol{U}).\partial]\chi_j - D\,\partial^2\chi_j = -u_j \qquad [38]$$

yielding the effective diffusivity tensor (where $\langle\,\rangle$ is a space average)

$$D_{ij}^{\rm E} \equiv \frac{D_{ij}^{\rm eff} - D_{ji}^{\rm eff}}{2} = D\sum_p \langle \partial_p \chi_i \partial_p \chi_j \rangle \qquad [39]$$

Advection enhances transport, and eddy diffusivity is larger than molecular diffusivity. In realistic cases, there is a continuum of scales $u = \sum_{n=0}^{N} u_n$, where $u_n$ has a characteristic scale $l_n \sim 2^{-n} l_0$. Multiple-scale method is to be iterated into an RG analysis, achieving a recursive integration of the small and fast scales into $D^{\rm E}$ starting from the smallest and fastest ones.

## Conclusions

Multiscale approaches allow to predict large-scale behavior generated by a given model; even more, they offer constructive tools to bridge models at different scales for the same phenomenon. They provide systematic and mathematically well-controlled tools to turn faithful but intractable models into effective reduced ones, thus lying at the core of statistical mechanics, many-body dynamical systems, and, more generally, at all issues of the still-in-progress complex systems science. Indeed, in a complex system (that might be their very definition), levels are so interrelated that it is essential to investigate jointly all the scales, from elementary units up to the whole system, and its emergent properties; neither theoretical nor numerical approaches can alone consider all the levels together, showing the relevance, if not the necessity, of multiscale approaches.

Basic preliminary issues are to determine the proper elementary level, the proper collective variables, and the relevant small parameters. Let us remark that the implementation of a multiscale technique rapidly faces the fundamental issue of defining a macroscopic variable; it offers some clues, indicating that a macroscopic variable might be a

phenomenological quantity observable at our scale, a slow mode, or collective variable.

Multiscale approaches take benefit of the separation of scales involved in the different mechanisms at work in the phenomenon under consideration. The basic idea, seen above at work in various instances and different ways, is to somehow decouple the different scales and to solve several simpler single-scale problems. Any multiscale implementation actually involves, at some stage and more or less explicitly, a limiting process in which the scale separation ratio $1/\epsilon$ tends to $\infty$: this limiting process has to be carefully controlled in order that the method can be applied to real situation. Finally, to be successful, multiscale approaches should achieve a trade-off between:

- *accuracy* (minimizing the loss of information involved in the reduction or projection technique),
- *efficiency and tractability* (this is, e.g., one of the major successes of hydrodynamics)
- *robustness* of the resulting reduced model (to be checked *a posteriori*),
- *flexibility* (extending to heterogeneous systems involving different components), and
- *scope* (bridging many different levels in order to capture the whole hierarchical structure).

Let us conclude by emphasizing a much fruitful benefit of multiscale approaches: they allow to investigate structural stability of a model, in particular to evidence relevant parameters and essential mechanisms controlling large-scale features. In this respect, they lead beyond the (necessarily restricted) scope of a specific model and give an explicit account of the observer biased view, related to its scale of observation. They hence contribute to capture a more complete and controlled understanding of the real physical systems.

Finally, a note on bibliographic guide to multiscale approaches may be useful. Technical details and several applications of multiscale perturbative expansions, in particular multiple-timescale method, with references to the original papers, can be found in Nayfeh (1973). Applications of multiple-scale method, fully worked out in a very pedagogical way, can be found in the work of Cukier and Deutsch (1969), Piasecki (1993), Bocquet (1997), and Mazzino *et al.* (2004). An acknowledged reference on homogenization techniques and multiscale analysis in periodic media is Bensoussan *et al.* (1978); see also the monographs by Lochak and Meunier (1988) and Berdichersky *et al.* (1999). Two recent review papers on multiscale approaches and reduction techniques are Givon *et al.* (2004) and Gorban *et al.* (2004). Basic principles and technical aspects of scaling theories and RG approaches from a multiscale viewpoint can be found in Lesne (1998).

*See also:* Adiabatic Piston; Averaging Methods; Bifurcations in Fluid Dynamics; Boltzmann Equation (Classical and Quantum); Central Manifolds, Normal Forms; Interacting Particle Systems and Hydrodynamic Equations; Korteweg–de Vries Equation and Other Modulation Equations; Localization for Quasiperiodic Potentials; Singularity and Bifurcation Theory; Stability Problems in Celestial Mechanics; Stationary Phase Approximation; Universality and Renormalization.

## Further Reading

Auger P and Bravo de la Parra R (2000) Methods of aggregation of variables in population dynamics. *Comptes Rendus de l'Académie des Sciences, Paris, Life Sciences* 323: 665–674.

Bensoussan A, Lions JL, and Papanicolaou G (1978) *Asymptotic Analysis for Periodic Structures.* Amsterdam: North-Holland.

Berdichersky V, Jikov V, and Paparieolasu G (1999) *Homogenization.* Singapore: World Scientific.

Bocquet L (1997) High-friction limit of the Kramers equation: the multiple time-scale approach. *American Journal of Physics* 65: 140–144.

Chen LY, Goldenfeld N, and Oono Y (1996) Renormalization group and singular perturbations: multiple scales, boundary layers, and reductive perturbation theory. *Physical Review E* 54: 376–394.

Cukier RI and Deutch JM (1969) Microscopic theory of Brownian motion: the multiple-time-scale point of view. *Physical Review* 177: 240–244.

Gaveau B, Lesne A, and Schulman LS (1999) Spectral signatures of hierarchical relaxation. *Physics Letters A* 258: 222–228.

Givon D, Kupferman R, and Stuart A (2004) Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity* 17: R55–R127.

Gorban A, Karlin I, and Zinovyev A (2004) Constructive methods of invariant manifolds for kinetic problems. *Physics Reports* 396: 197–403.

Haken H (1996) Slaving principle revisited. *Physica D* 97: 95–103.

Lesne A (1998) *Renormalization Methods.* New York: Wiley.

Lochak P and Meunier C (1988) *Multiphase Averaging for Classical Systems.* Berlin: Springer.

Mazzino A, Musacchio S, and Vulpiani A (2004) Multiple-scale analysis and renormalization for preasymptotic scalar transport. *Physical Review E* 71: 011113.

Murray JD (2002) *Mathematical Biology*, 3rd edn. Berlin: Springer.

Nayfeh AH (1973) *Perturbation Methods.* New York: Wiley.

Piasecki J (1993) Time scales in the dynamics of the Lorentz electron gas. *American Journal of Physics* 61: 718–722.

# N

## Negative Refraction and Subdiffraction Imaging

**S O'Brien**, Tyndall National Institute, Cork, Republic of Ireland
**S A Ramakrishna**, Indian Institute of Technology, Kanpur, India



### Introduction

The concept of negative refraction has caused a revolution in classical optics and electromagnetic theory in the past few years (Pendry 2004, Ramakrishna 2005). If a material has negative dielectric permittivity ($\varepsilon$) and negative magnetic permeability ($\mu$) simultaneously at a given frequency $\omega$, then it can be said to have a negative refractive index defined as

$$n = -\sqrt{\varepsilon\mu} \qquad [1]$$

Several peculiar consequences of Maxwell's equations for the propagation of radiation in such a material were originally pointed out by Veselago (1968). But the lack of such natural materials failed to create much enthusiasm until recently when composite structured photonic materials have been shown to have negative refractive index (Smith *et al.* 2000, Shelby *et al.* 2001).

The question then boils down to what constitutes materials with negative $\varepsilon$ and $\mu$? Where the structure varies spatially on a scale much less than the wavelength of the incident radiation, composite electromagnetic materials can be regarded effectively as homogeneous media. A set of effective response functions: the effective permittivity, $\varepsilon_{\text{eff}}$, and the effective permeability, $\mu_{\text{eff}}$, can then be ascribed to these materials. To develop a homogeneous view of the electromagnetic properties of a medium composed of discrete atoms and molecules was the motivation for defining a permittivity $\varepsilon$ and permeability $\mu$. The simplicity provided by such a description cannot be understated. Provided the radiation cannot resolve the underlying structure, replicating the atoms of a material with structure on a larger scale therefore represents a straightforward extension of the original concept.

If we consider arrays of structures defined by a unit cell of dimensions, $d$, then our effective description of the response of the medium to electromagnetic radiation of angular frequency $\omega$ will be valid provided that

$$d \ll \lambda = 2\pi c/\omega \qquad [2]$$

This restriction ensures that the underlying structure of the medium will merely refract and not scatter the incident radiation, in which case an effective permittivity and permeability for the medium become valid. The above inequality defines the long wavelength or effective medium limit (Garland and Tanner 1978). Maxwell's equations, written in the absence of free charges and external currents,

$$\nabla \cdot \boldsymbol{D} = 0, \qquad \nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \qquad [3]$$

$$\nabla \cdot \boldsymbol{B} = 0, \qquad \nabla \times \boldsymbol{H} = \frac{\partial \boldsymbol{D}}{\partial t} \qquad [4]$$

together with the constitutive relations:

$$\boldsymbol{B}(\omega) = \mu_0\mu_{\text{eff}}(\omega)\boldsymbol{H}(\omega) \qquad [5]$$

$$\boldsymbol{D}(\omega) = \varepsilon_0\varepsilon_{\text{eff}}(\omega)\boldsymbol{E}(\omega) \qquad [6]$$

then provide us with a complete description of the electromagnetic properties of the material over the frequency range of interest. Note that the effective-medium parameters are a function of the frequency as the material polarization response depends on the time history of the applied fields (Landau *et al.* 1984). These effective parameters were then generalized to analytic complex functions to account for absorption, and to second-ranked tensors to describe anisotropic responses.

The real parts of these effective material parameters can always be negative; there is nothing fundamentally wrong about that. Provided that they are dispersive, that is, they vary as a function of frequency, and dissipative as a consequence of the famous Kramers–Kronig relations (Landau *et al.* 1984), such materials are causally possible. Simultaneously negative values of $\varepsilon_{\text{eff}}$ and $\mu_{\text{eff}}$ change the nature of electromagnetic radiation in these media.

For example, the wave vector in such isotropic media points opposite to the Poynting vector and gives rise to many new interesting effects such as modified refraction, negative Doppler shifts, etc. Such materials can support a variety of surface electromagnetic modes, which can have dramatic effects such as the possibility of a perfect lens which has unlimited image resolution (Pendry 2000) and is not subject to the traditional diffraction limit.

New artificial electromagnetic composite structures, often referred to as "meta-materials," allow us to access values of these material parameters which are not found in naturally occurring materials. We will show here how to obtain negative values of $\varepsilon_{\mathrm{eff}}$ and $\mu_{\mathrm{eff}}$ in meta-materials using a variety of resonance phenomena. Then we will look at the problem of imaging with subdiffraction resolution using negative refractive index materials.

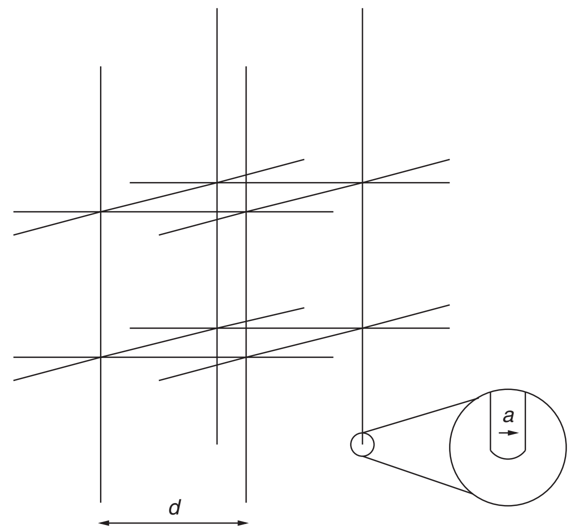



**Figure 1** A periodic structure composed of infinite conducting wires arranged in a simple cubic lattice. Provided the factor *a*/*d* is small enough, the structure responds to incident electromagnetic waves as a plasma of very heavy charged particles.

## Artificial Plasmas

From the electromagnetic viewpoint, a plasma can be represented as a medium with dielectric permittivity whose real part is negative. The Coulomb force and the finite mass of the electrons combine to give an ideal plasma a dispersion in the relative permittivity, $\tilde{\varepsilon}(\omega)$, given by

$$\tilde{\varepsilon}(\omega) = 1 - \frac{\omega_{\mathrm{p}}^2}{\omega^2} \qquad [7]$$

where the plasma frequency is defined by $\omega_{\mathrm{p}}^2 = (\rho e^2)/(\varepsilon_0 m_{\mathrm{e}})$, $\rho$ is the number density of electrons, $e$ is the electronic charge, and $m_{\mathrm{e}}$ is the electron mass. The permittivity of the plasma is negative at frequencies below the plasma frequency.

A plasma-like behavior characterizes the electron gas in the noble and alkali metals, with a plasma frequency typically at ultraviolet frequencies. Because of the presence of dissipation, at lower frequencies resistive effects dominate and the plasmons cannot be excited. To obtain materials with negative dielectric permittivity at low frequencies, a lower plasma frequency is required corresponding to more massive particles and a lower particle density $\rho$. A structure consisting of a three-dimensional lattice of very thin wires simulates a low-density plasma of very heavy charged particles and is shown in **Figure 1** (Pendry *et al.* 1998). A simple model allows us to describe the desired reduction in $\omega_{\mathrm{p}}$ in such a structure.

First consider a displacement of the electrons in the wires along one of the cubic axes. Only the wires directed along that axis are active and thus provide a lowered effective density of electrons, $\rho_{\mathrm{eff}}$, given by the area occupied by the active wires. Thus,

$$\rho_{\mathrm{eff}} = \rho \frac{\pi a^2}{d^2} \qquad [8]$$

An even more profound effect of constraining the electrons to run along thin wires is a result of the induced magnetic field which wraps the wires as the electrons are in motion. Suppose a current $I$ flows in the wires. The magnetic field is

$$H(r) = \frac{I}{2\pi R} = \frac{\rho a^2 v e}{2R} \qquad [9]$$

where $R$ is the distance from the wire center, $v$ is the electron drift velocity, and $\rho e$ is the charge density in the wire. In terms of the magnetic vector potential, the magnetic field is

$$\boldsymbol{H}(R) = \mu_0^{-1} \nabla \times \boldsymbol{A}(R) \qquad [10]$$

where

$$A(R) = \frac{\mu_0 a^2 \rho v e}{2} \ln(d/a) \qquad [11]$$

and $d$ is the lattice spacing. The importance of the divergence of the magnetic field with the wire radius as seen in eqn [9] is the contribution to the canonical electronic momentum given by $e\boldsymbol{A}$. If we neglect the variation of the fields with distance from the wire center, we can view this contribution as defining a new effective mass for the electrons given by

$$m_{\mathrm{eff}} = \frac{\mu_0 e^2 \rho}{2\pi} \ln(d/a) \qquad [12]$$

Now the effective plasma frequency for the system

$$\omega_{\mathrm{p}}^2 = \frac{\rho_{\mathrm{eff}} e^2}{\varepsilon_0 m_{\mathrm{eff}}} = \frac{2\pi c_0^2}{d^2 \ln(d/a)} \quad [13]$$

is seen to be much reduced. As an example, the plasma frequency of 1 μm aluminum wires paced by 10 mm is about 2 GHz, and the corresponding electronic effective mass is almost 15 times that of a proton! The factors of effective mass and charge density cancel leaving an expression comprising only the macroscopic system parameters. This is to be expected as a circuit analysis in terms of a capacitance and inductance can also be used to formulate the problem. However, such an approach can obscure the true nature of the problem which is encapsulated as a low-frequency plasma oscillation. Inclusion of the finite resistivity of the metal yields a finite lifetime for the plasmon excitation. Experiments have shown that a reduction in the plasma frequency of six orders of magnitude from the ultraviolet to the microwave region can be achieved in these thin-wire composites (Pendry *et al.* 1998).

## Artificial Magnetism

Although the Maxwell equations [2]–[4] are symmetric in the electric and magnetic fields, we are yet to discover a free magnetic pole. The magnetism we find in natural materials is limited to spin systems and restricts the values of $\mu_{\mathrm{eff}}$. Up to microwave frequencies, magnetic activity is common and certain insulating ferromagnets and antiferromagnetic compounds such as $MgF_2$ and $FeF_2$ can even exhibit a negative permeability at some frequencies. However, large losses can accompany the magnetic activity in these materials.

Recently, it has become clear that a wide variety of composite structures comprising resonant inclusions can display magnetic activity in the effective medium limit (Pendry *et al.* 1999). Efficient screening of AC magnetic fields can be achieved using a thin cylindrical shell of metal or superconductor. In order to obtain a large magnetic response such that the modulus of the magnetic susceptibility, $|\chi_{\mathrm{m}}| > 1$, what we require is a resonant over-screening material response. A collection of subwavelength-sized structures that exhibits such an over-screening response can constitute a negative $\mu_{\mathrm{eff}}$ material. One such resonant subwavelength structure is the so-called split-ring resonator (SRR), which can be scaled to form magnetic meta-materials from microwave to optical frequencies (Pendry *et al.* 1999, O'Brien and Pendry 2002b). An SRR

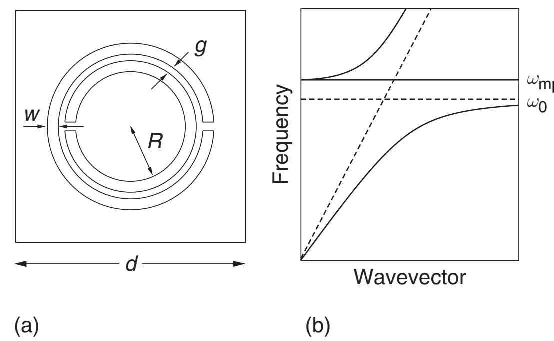



**Figure 2** (a) The split-ring resonator structure. The structure is planar with an internal radius *R*. The metal rings are of width *w* and are separated by a spacing *g*. (b) Generic dispersion relationship, $\omega$ vs. *k*, for a resonant structure with an isotropic effective permeability as in eqn [15].

structure which has been demonstrated experimentally to have a resonant magnetic response at microwave and THz frequencies is depicted in **Figure 2a** (Smith *et al.*). It comprises of two planar rings of metal on an insulating backing. The rings couple inductively to the magnetic field normal to the plane of the rings. Because of the large capacitance between the rings, the structure resonates at some frequency. Driven by the back electromotive force (emf), a large response is expected in the vicinity of the resonance frequency which is also antiphased in a small frequency range above the resonant frequency. If the SRRs are much smaller than the free-space wavelength, a collection of such SRRs would behave as a negative $\mu_{\mathrm{eff}}$ material at these frequencies.

Theoretical calculations (Pendry *et al.* 1999) assuming a nondispersive metal show that a periodic lattice of such structures is characterized by a magnetic permeability given by

$$\tilde{\mu}_{\mathrm{eff}} = 1 - \frac{f\omega^2}{\omega^2 - \omega_0^2 + \mathrm{i}\Gamma\omega} \quad [14]$$

where $f = \pi R^2/d^2$ is the filling factor,

$$\omega_0 = \sqrt{\frac{3lc^2}{\pi R^3 \ln 2w/g}} \quad [15]$$

is the resonant frequency, and the damping of the resonance is determined by the factor

$$\Gamma = \frac{2l}{\mu_0 \sigma R} \quad [16]$$

Here $d$ is the lattice spacing, $R$ is the inner radius of the ring, $w$ is the width of the rings, $l$ is the distance between adjacent planes of SRRs, and $\sigma$ is the conductance per unit length of the rings measured along the circumference. Orientation of planar SRRs

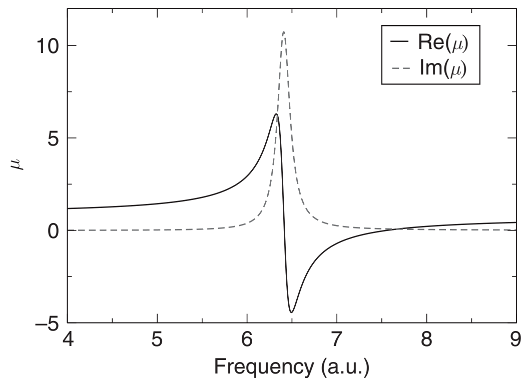



**Figure 3** (a) The generic magnetic response of the SRR structure. Re($\mu$) < 0 in a frequency band above the resonance frequency.

along all three Cartesian axes allows for the creation of an isotropic material. Figure 3 shows the generic dispersion of the $\mu(\omega)$ given by eqn [14]. A higher resistivity for the material of the SRR would broaden the resonance and the frequency region with Re($\mu$) < 0 might vanish altogether for large resistivity.

For isotropic homogeneous materials with a resonant effective permeability as in eqn [14] we can illustrate a generic dispersion relationship, $\omega$ vs. $k$, shown in Figure 2b. The solid lines represent twofold degenerate transverse modes and the dispersionless longitudinal magnetic plasmon mode at the magnetic plasmon frequency ($\omega_{\rm mp}$). The dashed lines are a band of propagating states with a linear dispersion determined by the polarizability of the SRRs and a flat band of resonant states at the magnetic resonance frequency $\omega_0$. The gap in the dispersion can be regarded as arising from the hybridization and avoided crossing of these bands. The important points to note are:

1. Wherever $\mu_{\rm eff}$ is negative there is a gap in the dispersion relationship. This is the case for $\omega_0 < \omega < \omega_{\rm mp}$, the frequency where $\mu_{\rm eff} = 0$. Only evanescent modes with imaginary wave vector exist in this region.
2. A longitudinal magnetic plasma mode, which shows no dispersion, appears at $\omega = \omega_{\rm mp}$.

An alternative approach to obtaining a nonzero magnetic susceptibility in composite media is provided by the zeroth-order transverse electric (TE) Mie resonance in dielectric particles. Ferroelectric and phonon polaritonic materials are promising candidates for providing the necessary large dielectric constants up to infrared frequencies (O'Brien and Pendry 2002a).

The high-frequency scaling properties of the SRR offer an interesting insight. The plasma-like dielectric permittivity of noble metals

$$\tilde{\varepsilon}(\omega) = (\varepsilon_1, \varepsilon_2) = \varepsilon_\infty - \frac{\omega_{\rm p}^2}{\omega(\omega + {\rm i}\gamma)} \qquad [17]$$

is essentially a large negative real number for $\omega_{\rm p} \gg \omega \gg \gamma$. For a 2D array of simplified SRRs consisting of a single conducting ring with symmetrically placed small capacitive gaps, the quasistatic effective magnetic permeability for a magnetic field applied normal to the plane of the SRR is (O'Brien and Pendry 2002b)

$$\tilde{\mu}_{\rm eff} = 1 - \frac{f'\omega^2}{\omega^2 - {\omega_0}^2 + {\rm i}\Gamma\omega} \qquad [18]$$

where $f' = L_{\rm g} f \cdot (L_{\rm g} + L_{\rm i})^{-1}$, $\Gamma = L_{\rm i}\gamma \cdot (L_{\rm g} + L_{\rm i})^{-1}$, and ${\omega_0}^2 = (L_{\rm g} + L_{\rm i})^{-1} C^{-1}$. In the above expressions, $L_{\rm g} = \mu_0 \pi R^2$ is the geometrical inductance per unit length of the structure and $C = \varepsilon_0 \tilde{\varepsilon}_{\rm s} \tau / n_{\rm c} d_{\rm c}$ is the capacitance per unit length of the structure for series connection. Here it has been assumed that the thickness of the SRR ($\tau$) is small compared to the skin depth $\delta \simeq c_0/\omega_{\rm p}$.

An additional inductive impedance in the structure, the kinetic or inertial inductance, $L_{\rm i} = 2\pi R / \varepsilon_0 \omega_{\rm p}^2 \tau = 2\mu_0 \pi R \delta^2 / \tau$, determines the effective filling fraction and damping of the resonance through the ratio of the two contributions to the total inductance. This contribution to the inductance arises from the finite electron mass and implies that simply decreasing the size of the resonators indefinitely will not result in our being able to realize a strong magnetic response at near-infrared or optical frequencies. As the dimensions of the structure are reduced that fraction of the energy of the displacement current associated with the inertial mass of the electrons increases. A finite $\gamma$ then means that dissipative losses increase. Thus, strong damping of the resonance will be avoided if the quantity $R\tau/2\delta^2$ is large. We note here that with $\delta$ equal to the London penetration depth, this ratio also determines the screening efficiency of low-frequency magnetic fields by a thin layer of superconductor. This result points to a broader similarity between the low-frequency electromagnetic properties of the superconducting condensate and those of a perfect plasma.

Other nanocomposites in addition to the SRR have been proposed which may lead to a magnetic response at optical frequencies. These include pairs of nanometer-sized metallic sticks where simultaneous electric and magnetic dipole resonances lead to a strongly dispersive effective permittivity and permeability.

## Negative Refractive Index Media

Interleaving the structures for a negative $\varepsilon_{\text{eff}}$ and $\mu_{\text{eff}}$ can create a composite with $\varepsilon_{\text{eff}} < 0$ and $\mu_{\text{eff}} < 0$ at a common frequency ($\omega$) (Smith *et al.*, Shelby *et al.* 2001), which as predicted by Veselago (1968) should give rise to a material with negative refractive index. Although this appears intuitively correct, it is actually nontrivial that the electromagnetic fields of the two composites do not interfere with each other's function (Pokrovsky and Efros 2002) and this could depend crucially on the relative placement of the two structures (Marques and Smith 2004). However, there is now overwhelming experimental and numerical evidence that such composite structures possess negative refractive index (see Ramakrishna (2005, section 6)). Now consider a medium with predominantly real $\varepsilon$ and $\mu$. For $\varepsilon > 0$ and $\mu > 0$, we have our usual optical materials. Only one of $\varepsilon$ or $\mu$ lesser than zero with the other positive would imply a medium which cannot support any propagating modes. This is a consequence of Maxwell's equations:

$$\boldsymbol{k} \cdot \boldsymbol{k} = \varepsilon(\omega)\mu(\omega)\frac{\omega^2}{c_0^2} \qquad [19]$$

which implies that only evanescently decaying waves with an imaginary component of $\boldsymbol{k}$ are possible. Common examples are ordinary metals with $\varepsilon < 0$ and $\mu > 0$. Now consider a medium with both $\varepsilon < 0$ and $\mu < 0$, or a negative refractive index medium. The Maxwell's equations for a plane time-harmonic wave $\exp[\mathrm{i}\,(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t)]$ are:

$$\boldsymbol{k} \times \boldsymbol{E} = \frac{\omega}{c}\mu(\omega)\boldsymbol{H} \qquad [20]$$

$$\boldsymbol{k} \times \boldsymbol{H} = -\frac{\omega}{c}\varepsilon(\omega)\boldsymbol{E} \qquad [21]$$

The "left-handedness" of the triad $(\boldsymbol{E}, \boldsymbol{H}, \boldsymbol{k})$ is clear from these equations for $\varepsilon(\omega), \mu(\omega) < 0$. A real refractive index means that waves propagate with the direction of energy flow given by the Poynting vector,

$$\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{H} \qquad [22]$$

opposite to the direction of the wave vector. Since the group velocity is in the direction of the energy flow, we conclude that in these left-handed materials (LHMs) the group velocity and the phase velocity are oppositely directed. The phase accumulated in propagating a distance $x$ is $\Delta\phi = -\sqrt{\varepsilon\mu}\omega/c_0 x$. Thus, the refractive index can be taken to be $n = -\sqrt{\varepsilon\mu}$, that is, a negative quantity. Mathematically, it is more reasonable to ask for the sign of the square-root to determine the wave vector given by eqn [19]. It can be shown by arguments of analytic continuity in the complex plane that the negative sign has to be chosen for propagating waves when $\mathrm{Re}(\varepsilon) < 0$ and $\mathrm{Re}(\mu) < 0$ (Ramakrishna 2005).

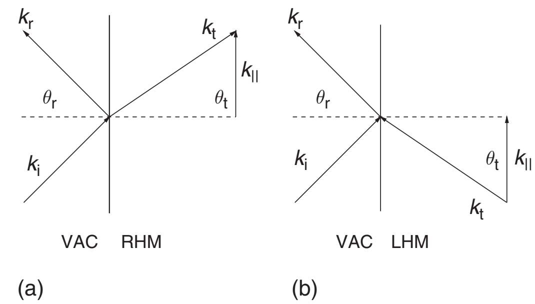



**Figure 4** Illustration of Snell's law at an interface between two media with (a) positive refractive index (VAC/RHM) and (b) negative refractive index (VAC/LHM). The arrows indicate the wave vectors and the energy flow is opposite to the wave vector in the negative index medium.

The negative refractive index has real effects on the behavior of radiation even in basic processes such as refraction. Consider an interface between vacuum and a negative refractive index medium with $n < 0$ shown in Figure 4. Continuity conditions on the electromagnetic fields at the interface require for a plane wave incident from the vacuum side at an oblique angle that the parallel wave vector $k_{\|}$ is conserved for the transmitted and reflected wave. This is the origin of Snell's law:

$$\sin(\theta_{\mathrm{i}}) = \sin(\theta_{\mathrm{r}}) = n_{-}\sin(\theta_{\mathrm{t}}) \qquad [23]$$

where $\theta_{\mathrm{i}}$, $\theta_{\mathrm{r}}$ and $\theta_{\mathrm{t}}$ are the angles of incidence, reflection, and transmission, respectively. The flow of energy across the interface determines the direction of the group velocity in the material medium as being away from the interface. Therefore, the component of the phase velocity vector normal to the interface must change sign as we pass from vacuum into the material medium. We are then forced to conclude that the ray is bent toward the same side of the surface normal as the incident wave. This picture is consistent with Snell's law with the interpretation that $n < 0 \Rightarrow \theta_{\mathrm{t}} < 0$. Figure 4 illustrates this point which has been experimentally verified by several groups (Shelby *et al.* 2001, Parazzoli *et al.* 2003, Eleftheriades *et al.* 2002).

As a direct consequence of this, it is seen that a flat slab of negative refractive medium can act as a lens as shown in Figure 5. Provided that the slab is of sufficient thickness, the refracted rays from a point source come to a focus inside the slab and upon exiting the slab the rays are redirected again such that they come to a focus on the opposite side of the slab (Veselago 1968). Veselago also predicted a negative Doppler shift in such media and an obtuse angle cone for Cerenkov radiation.

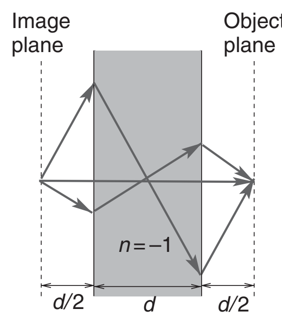



**Figure 5** Steady-state passage of rays (representing the energy flow) of light from vacuum through a slab made of a LHM with $n = -1$. The slab acts as a lens mapping a point on the image plane to a point on the object plane.

## Perfect Lens: Subwavelength Imaging

A wave analysis of the Veselago lens revealed an extremely novel aspect: it did not suffer from the diffraction limit and the image resolution could be infinite (Pendry 2000), if the negative index material were perfectly nondispersive and nonabsorbing. Before we analyze this, let us first briefly review the problem of imaging and the diffraction limit.

Any object is visible because it emits or scatters light. The problem of imaging is then concerned with reproducing the electromagnetic field distribution on a 2D object plane in the 2D image plane. If $E(x, y, 0)$ be the electric field on the object ($z = 0$) plane, the fields in free space can be decomposed into the Fourier components $k_x$ and $k_y$, and polarization defined by $\sigma$:

$$E(x, y, z; t) = \sum_{\sigma, k_x, k_y} E_\sigma(k_x, k_y) \times \exp\left[i\left(k_x x + k_y y + k_z z - \omega t\right)\right] \quad [24]$$

where

$$E_\sigma(k_x, k_y) = \int_{x,y} E_\sigma(x, y, 0)\, e^{-i(k_x x + k_y y)}\, dx\, dy \quad [25]$$

In the above expression, the source is assumed to be monochromatic of frequency $\omega$, $k_x^2 + k_y^2 + k_z^2 = \omega^2/c_0^2$, $c_0$ is the speed of light in free space, and $z$ is the optical axis. A conventional lens acts by applying a phase correction to each of the propagating components so that they reassemble to a focus at a point beyond the lens. For these components $k_z$ is real, thus a phase change is all that is required to form an image containing these components. The higher spatial details in an object, however, are described by the nonpropagating near-field components with an imaginary $k_z$ where $k_x^2 + k_y^2 > \omega^2/c^2$. A conventional lens cannot restore these components in the image plane as they decay exponentially in amplitude as one moves away from the source. Hence the resolution, $\Delta$, provided by a conventional lens is limited to those components with

$$k_x^2 + k_y^2 < \omega^2/c^2 \Rightarrow \Delta \sim \frac{2\pi c}{\omega} = \lambda \quad [26]$$

Now consider the slab of medium with $\varepsilon = -1$ and $\mu = -1$ and of thickness $d_s$. It can be shown (Pendry 2000) that the transmission and reflection coefficients are

$$\lim_{\substack{\varepsilon \to -1 \\ \mu \to -1}} \tilde{t} = \exp[-ik_z d_s] \quad [27]$$

$$\lim_{\substack{\varepsilon \to -1 \\ \mu \to -1}} \tilde{r} = 0 \quad [28]$$

respectively, where $k_z$ is the component of the wave vector normal to the interface. Thus, the slab reverses the phase advance for the propagating waves as revealed by the ray picture. Analytic continuation to imaginary wave vectors $k_z = i\kappa_z$ implies that the transmittance $\tilde{t} \to \exp(+\kappa_z d)$, that is, the slab also increases the amplitude of the evanescent waves in transmission at exactly the same rate as the rate of the decay in free space outside. Thus, each wave, propagating or evanescent, arrives at the image plane with its phase or amplitude restored exactly to the values at the object plane so as to perfectly reconstruct the image. The lens is also perfectly impedance matched and has zero reflection. These incredible properties have led the phenomenon to be called “perfect lensing.”

Note that there is no energy flux associated with purely evanescent waves, and hence the amplification obtained in the steady state corresponds to local field enhancements which would imply the presence of localized resonances. In fact, the entire mechanism of the focusing of the near-field components is due to surface modes that reside on the surfaces of these negative index materials (Ramakrishna 2005). $\varepsilon = -1$ and $\mu = -1$ are precisely the conditions for these surface modes of electric and magnetic nature, respectively. These surface plasmon resonances which are excited resonantly by the evanescent modes and the secret to the perfect lens is that all the surface modes are completely degenerate.

Although the conditions for realizing a perfect lens are easy to specify, in practice these are very difficult to meet. The requirement of negative values for $\varepsilon$ and $\mu$ implies that these quantities must disperse necessarily with frequency and be dissipative. Thus, the perfect-lens condition can only be met approximately at a single frequency. Any deviation from the ideal

conditions can then result in the excitation of slab polariton resonances which can swamp the image. The effects of absorption, which are always present, can also seriously degrade the lens performance by damping out the surface plasmon resonances (Ramakrishna 2005). Consider the transmission for the P-polarized radiation through a negative index slab:

$$\tilde{t}(k_x) = \frac{4(k_{z1}/\varepsilon_+)(k_{z2}/\varepsilon_-)\, e^{ik_{z2}d_s}}{\mathcal{D}} \quad [29]$$

where

$$\mathcal{D} = (k_{z1}/\varepsilon_+ + k_{z2}/\varepsilon_-)^2 - (k_{z1}/\varepsilon_+ - k_{z2}/\varepsilon_-)^2 e^{2ik_{z2}d_s}$$

Under the perfect-lens conditions, the first term in the denominator goes to zero for evanescent waves and the exponential in the second term decays faster than the exponential in the numerator. However, if there was a mismatch in the conditions, ($\varepsilon_+ = 1$ and $\varepsilon_- = -1 + \delta$, say) then the first term in the denominator no longer vanishes. In the large wave vector limit ($k_x \gg \omega/c_0$), the two terms in the denominator become approximately equal when

$$k_x = -\frac{1}{d_s}\ln\left|\frac{\delta}{2}\right| \quad [30]$$

thus yielding a criterion for the largest wave vector for which there is effective amplification. The dependence through the logarithm on the deviations (whether real or imaginary) from the resonant conditions underlines the fact that the perfect lens effect is indeed very sensitive. In practice, the periodicity, $d$, of the strucuture of the metamaterials comprising the negative index slab itself imposes an upper wave vector cutoff $k_c = 2\pi/d$. The material will become spatially dispersive for wave vectors $k \to k_c$, and for $k > k_c$ the very description as a homogeneous material will break down.

An important simplification of the perfect-lens conditions results when we consider a situation in which all length scales in the problem are much less than the wavelength of the light (the quasistatic approximation). Under these conditions, the electric and magnetic fields effectively decouple. If we consider the case of P-polarized fields, it can be shown (Pendry 2000) that in the quasistatic limit only the value of the permittivity is important, and there are essentially no conditions on the value of the permeability. This brings metals such as silver into the picture as the permittivity of silver becomes equal to –1 in the optical region of the spectrum and with relatively small losses (Pendry 2000). To overcome the losses, a series of refinements of the simple thin-slab picture have been proposed including dividing the lens into a series of layers and using optical amplification to act against the deleterious effects of absorption (Ramakrishna 2005).

### The Generalized Perfect-Lens Theorem

The negative refractive slab can be considered as "optical antimatter" in the sense that it cancels out the effects on radiation of the traversal through an equal amount of positive refractive index medium. This cancelation is applicable to the phase changes for the propagating modes and the amplitude changes to the evanescent modes. In fact, the focussing action can happen for more general situations where the requirement of homogeneity of the slab material can be relaxed. Now consider the more general situation where the dielectric permittivity and the magnetic permeability are arbitrary functions of the spatial coordinates:

$$\varepsilon_+ = \varepsilon(x, y), \qquad \mu_+ = \mu(x, y) \quad [31]$$

$$\varepsilon_- = -\varepsilon(x, y), \qquad \mu_- = -\mu(x, y) \quad [32]$$

corresponding to the Figure 6. We will consider the imaging axis to be the $z$-axis. Thus, we see that the system is antisymmetric with respect to the $z = d$ plane. It turns out (Pendry and Ramakrishna 2003) that such a system also transfers the image of a source placed at the $z = 0$ to the $z = 2d$ plane in the same exact sense that it includes both the propagating and evanescent components. In general, the rays in spatially varying media will not be straight lines as shown in Figure 6, but the effect of propagating through the positive medium is nullified by the negative medium. Thus, to an observer on the right-hand side, it would appear as if the region between $z = 0$ and $z = 2d$ did not exist. We will call such media with the same sense of transverse spatial variation but with opposite signs as optical complementary media, and the effect of any such pairs of complementary media on radiation is null.

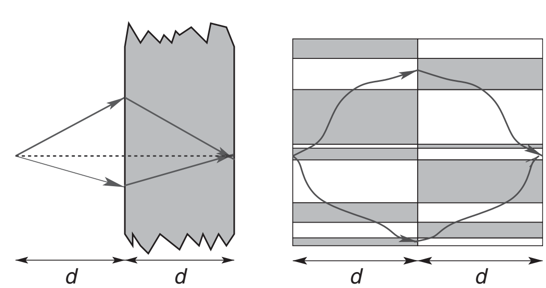



**Figure 6** A pair of complementary optical media nullify the effect of each other for the passage of light. Spatially varying positive and negative refractive indices are schematically depicted by the white or shaded regions.

The most general conditions on the permittivity and permeability tensors for such complementary behavior are:

$$\tilde{\varepsilon}_+ = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{pmatrix}$$
$$\tilde{\mu}_+ = \begin{pmatrix} \mu_{xx} & \mu_{xy} & \mu_{xz} \\ \mu_{yx} & \mu_{yy} & \mu_{yz} \\ \mu_{zx} & \mu_{zy} & \mu_{zz} \end{pmatrix} \qquad [33]$$

and

$$\tilde{\varepsilon}_- = \begin{pmatrix} -\varepsilon_{xx} & -\varepsilon_{xy} & +\varepsilon_{xz} \\ -\varepsilon_{yx} & -\varepsilon_{yy} & +\varepsilon_{yz} \\ +\varepsilon_{zx} & +\varepsilon_{zy} & -\varepsilon_{zz} \end{pmatrix}$$
$$\tilde{\mu}_- = \begin{pmatrix} -\mu_{xx} & -\mu_{xy} & +\mu_{xz} \\ -\mu_{yx} & -\mu_{yy} & +\mu_{yz} \\ +\mu_{zx} & +\mu_{zy} & -\mu_{zz} \end{pmatrix} \qquad [34]$$

and a perfect focus results whenever the two slabs of positive and negative media have such a behavior (see Pendry and Ramakrishna (2003) and Ramakrishna (2005) for the proof). This theorem clearly shows that the dependence along the $x$- and $y$-directions transverse to the imaging axis $z$ is completely irrelevant as long as the two slabs are optically complementary. As an extension, it can be shown that any system of optically complementary media will also have a perfect focus as long as the system has a plane of antisymmetry normal to the optical axis. The above effects have also been numerically verified for several such spatially varying complementary media (Pendry and Ramakrishna 2003).

### Perfect Lens in Other Geometries

The above generalized perfect-lens theorem along with a method of coordinate transformations can enable us to now generate a variety of superlenses in different geometries. In general, if we can find a geometric transformation that maps a given configuration into the geometry for the generalized slab lens, then we would have generated one more arrangement that will exhibit the property of transferring images of sources in a perfect sense. If we define the new coordinates $q_1(x, y, z)$, $q_2(x, y, z)$, and $q_3(x, y, z)$ (assumed orthogonal), then in the new frame, the material parameters and fields are given by (Ward and Pendry 1996)

$$\tilde{\varepsilon}_i = \varepsilon_i \frac{Q_1 Q_2 Q_3}{Q_i^2}, \qquad \tilde{\mu}_i = \mu_i \frac{Q_1 Q_2 Q_3}{Q_i^2} \qquad [35]$$

$$\tilde{E}_i = Q_i E_i, \qquad \tilde{H}_i = Q_i H_i \qquad [36]$$

where

$$Q_i^2 = \left(\frac{\partial x}{\partial q_i}\right)^2 + \left(\frac{\partial y}{\partial q_i}\right)^2 + \left(\frac{\partial z}{\partial q_i}\right)^2 \qquad [37]$$

Note that a distortion of space results in the change of $\varepsilon$ and $\mu$ tensors in general. Thus, in many cases, the transformed geometry would involve spatially varying (inhomogeneous) and anisotropic medium parameters.

The change in geometry can also make it possible for us to realize lenses with curved surfaces. The original slab lens maps every point on the object plane to another point on the image plane. But the size of the image is identical to that of the source. This is due to the invariance in the transverse direction and the transverse wave vector $(k_x, k_y)$ is preserved. In general, to change the size of the images, the translational symmetry would have to be broken and curved surfaces will necessarily be needed. The focussing action for the evanescent waves is crucially dependent on the near degeneracy of the surface plasmons in the case of the slab, and curved surfaces, in general, have a completely different dispersion for the surface plasmons. Thus, one should expect that inhomogeneous materials will be required for such curved lenses of negative refractive index. It can be shown (Ramakrishna 2005) that mapping the slab lens into cylindrical coordinates

$$x = r_0 e^{\ell/\ell_0} \cos\phi, \quad y = r_0 e^{\ell/\ell_0} \sin\phi, \quad z = Z \qquad [38]$$

where $\ell_0$ is some scale factor$(=1)$ generates a cylindrical annulus of inner and outer radii $a_1$ and $a_2$, respectively, with the material parameters given by

$$\begin{aligned} \varepsilon_r &= \mu_r = -1 \\ \varepsilon_\phi &= \mu_\phi = -1 \\ \varepsilon_z &= \mu_z = -1/r^2 \end{aligned} \qquad [39]$$

for the annular region. The positive material outside the annular region should vary as

$$\begin{aligned} \varepsilon_r &= \mu_r = +1 \\ \varepsilon_\phi &= \mu_\phi = +1 \\ \varepsilon_z &= \mu_z = +1/r^2 \end{aligned} \qquad [40]$$

where $r = r_0 \exp(\ell/\ell_0)$. This system transfers images in and out of the cylindrical annulus and the image of a source inside at $r = a_0$ will be formed on the surface $a_3 = a_0(a_2/a_1)^2$. Thus, there will be a magnification of the image by the factor

$$\mathcal{M} = \left(\frac{a_2}{a_1}\right)^2 \qquad [41]$$

Note that these cylindrical lenses are also short-sighted in the same manner as the slab lens. They can only focus sources from inside to the outside only when $a_1^2/a_2 < r < a_1$, and the other way around from outside to the inner world when the source is located in $a_2 < r < a_2^2/a_1$.

Similarly the transformation into spherical coordinates $(r = r_0 e^{\ell/\ell_0}, \theta, \phi)$ can be used to generate a spherical perfect lens wherein a spherical shell of negative refractive material with $\varepsilon(r) \sim -1/r$ and $\mu(r) \sim -1/r$ with arbitrary dependence along $\theta$ and $\phi$ (which could be constant too!) have the property of perfectly transferring images of sources in and out of the shell (Pendry and Ramakrishna 2003). This spherical lens also has exactly the same magnification factor given by eqn [41]. In fact, the solutions in these two cases of a cylinder and sphere can also be obtained by a more conventional electromagnetic calculation in terms of the scattering modes (Ramakrishna 2005). One can obtain even more esoteric configurations such as one or two intersecting corners of negative refracting materials that behave as perfect lenses (Pendry and Ramakrishna 2003).

### Other Approaches to Negative Refraction

There is also an approach to negative refractive materials based on loaded transmission lines (Eleftheriades *et al.* 2002), which has been implemented at radio- and microwave frequencies using lumped circuit elements. These show all the hallmarks of a negative refractive material within an effective medium approach.

Effects which can be interpreted as negative refraction have been observed in certain periodic photonic crystals (PCs) (Luo *et al.* 2003). An incident propagating plane wave from vacuum appears to undergo negative refraction inside the PC, and a slab of the PC can even work as a Veselago lens. The negative refraction in this case is a result of the curvature of the equifrequency surface and is present in spite of the right-handed nature of the propagation. In these instances, an effective permittivity and permeability cannot be easily ascribed to the crystal as the long wavelength condition is not met. It is difficult to homogenize the PC in the sense of meta-materials, and the energy transport in these PCs is very sensitive to the periodicity and the structural arrangements. Thus, it would be an over-simplification to characterize these effects in PC as merely due to an effective refractive index.

## Further Reading


Eleftheriades GV, Iyer AK, and Kremer PC (2002) Planar negative refractive index media using periodically L–C loaded transmission lines. *IEEE Transactions on Microwave Theory and Techniques* 50: 2702–2712.

Garland JC and Tanner DB (eds.) (1978) *Electrical Transport and Optical Properties of Inhomogeneous Media*. New York: American Institute of Physics.

Landau LD, Lifschitz EM, and Pitaevskii LP (1984) *Electrodynamics of Continuous Media*, 2nd edn. Oxford: Pergamon.

Luo C, Johnson SG, Joannopoulos JD, and Pendry JB (2003) Subwavelength imaging in photonic crystals. *Physical Review B* 68: 045115.

Marques R and Smith DR (2004) Comment on "Electrodynamics of Metallic Photonic Crystals and the Problem of Left-Handed Materials". *Physical Review Letters* 92: 059401.

O'Brien S and Pendry JB (2002a) Photonic band-gap effects and magnetic activity in dielectric composites. *Journal of Physics: Condensed Matter* 14: 4035–4044.

O'Brien S and Pendry JB (2002b) Magnetic activity at infrared frequencies in structured metallic photonic crystals. *Journal of Physics: Condensed Matter* 14: 6383–6394.

Parazzoli CG, Greegor RB, Li K, Koltenbah BEC, and Tanelian M (2003) Experimental verification and simulation of negative index of refraction using Snell's law. *Physical Review Letters* 90: 107401.

Pendry JB (2000) Negative refraction makes a perfect lens. *Physical Review Letters* 85: 3966–3969.

Pendry JB (2004) *Contemporary Physics* 45: 191–202.

Pendry JB, Holden AJ, Robbins DJ, and Stewart WJ (1998) Low frequency plasmons in thin-wire structures. *Journal of Physics: Condensed Matter* 10: 4785–4809.

Pendry JB, Holden AJ, Robbins DJ, and Stewart WJ (1999) Magnetism from conductors and enhanced nonlinear phenomena. *IEEE Transactions on Microwave Theory and Techniques* 47: 2075–2084.

Pendry JB and Ramakrishna SA (2003) Focusing light using negative refraction. *Journal of Physics: Condensed Matter* 15: 6345–6364.

Pokrovsky AL and Efros AL (2002) Electrodynamics of metallic photonic crystals and the problem of left-handed materials. *Physical Review Letters* 89: 093901.

Ramakrishna SA (2005) Physics of negative refractive index materials. *Reports on Progress in Physics* 68: 449–521.

Shelby RA, Smith DR, and Schultz S (2001) Experimental verification of a negative index of refraction. *Science* 292: 77–79.

Smith DR, Padilla WJ, Vier DC, Nemat-Nasser SC, and Schultz S (2000) Composite medium with simultaneously negative permeability and permittivity. *Physical Review Letters* 84: 4184–4187.

Veselago VG (1968) The electrodynamics of substances with simultaneously negative values of var $\varepsilon$ and $\mu$. *Soviet Physics–Uspekhi* 10: 509–514.

Ward AJ and Pendry JB (1996) Refraction and geometry in Maxwells equations. *Journal of Modern Optics* 43: 773–793.

# Newtonian Fluids and Thermohydraulics

**G Labrosse and G Kasperski**, Université Paris-Sud XI, Orsay, France



## Introduction

Thermohydraulics is based on the hypothesis of continuous medium. This hypothesis is easily satisfied since, for instance, a one-thousandth of $1\,\mathrm{mm}^3$ of a perfect gas at normal temperature and pressure conditions (300 K, 1 atm) contains about $2.5\times10^{13}$ molecules. Instantaneous balances are made inside a control volume fixed in the system of axes and crossed by the flows. The limit where this volume vanishes leads to the local formulation of the laws governing the flows. The flow is described by velocity $\vec{v}(\vec{r},t)$, pressure $p(\vec{r},t)$, temperature $T(\vec{r},t)$, and other fields, $\vec{r}$ being the position vector of a point $M$, and $t$ the time. The material derivative of $q(\vec{r},t)$ is

$$\frac{Dq}{Dt}\equiv\left(\frac{\partial}{\partial t}+(\vec{v}\,.\,\vec{\nabla})\right)q$$

Let $Q$ ($\vec{Q}$) be one of the scalar (vectorial) extensive quantities whose balance participates in the flow dynamics. It can be a quantity of matter, heat, impulse, or something else. Let $\Delta Q$ be the amount of $Q$ contained in the volume $\Delta\mathcal{V}$ localized around $M$, and $q(\vec{r},t)$ its local representative defined by

$$\rho(\vec{r},t)q(\vec{r},t)=\lim_{\Delta\mathcal{V}\to0}\frac{\Delta Q}{\Delta\mathcal{V}}\equiv\frac{\mathrm{d}Q}{\mathrm{d}\mathcal{V}}\qquad[1]$$

where $\rho$ is the density, similarly defined considering the case where $[Q]$ is taken as the mass $m$:

$$\rho(\vec{r},t)=\frac{\mathrm{d}m}{\mathrm{d}\mathcal{V}}\qquad[2]$$

Table 1 gives examples of $q$ quantities.

The instantaneous local balance of $Q$ reads

$$\frac{\partial}{\partial t}(\rho q)+\vec{\nabla}\cdot\left(\vec{j}_Q+\rho q\vec{v}\right)=S_Q\qquad[3]$$

where $S_Q$ stands for any possible local source of $Q$, and $\vec{j}_Q$ is the $Q$ conduction flux density. Figure 1 illustrates how these quantities allow us to evaluate the flux $d\Phi_Q=\vec{j}_Q\,.\,\mathrm{d}\vec{S}$ of $Q$ that instantaneously crosses a surface $\mathrm{d}\vec{S}$. Table 2 gathers the physical dimension of these notions for various $Q$'s.

**Table 1** Some quantities $q$. $T$ is the absolute temperature, $C_p$ the specific heat at constant pressure, and $C$ the solute mass fraction

| *Mass* | *Impulse* | *Kinetic energy* | *Heat* | *Mass fraction* |
|---|---|---|---|---|
| 1 | $\vec{v}$ | $\frac{\vec{v}^2}{2}$ | $C_pT$ | $0<C<1$ |

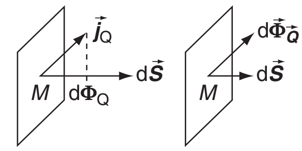



**Figure 1** $Q$ flux density and $\vec{Q}$ flux.

**Table 2** Physical dimension of fluxes, flux densities, and $\vec{\nabla}\cdot$(flux density) for some $q$ quantities

| *Q* | *q* | *Flux* | *Flux density* | $\vec{\nabla}\cdot$(*flux density*) |
|---|---|---|---|---|
| Volume | undefined | $\mathrm{m^3s^{-1}}$ | [velocity] | $\mathrm{s^{-1}}$ |
| Mass | 1 | $\mathrm{kgs^{-1}}$ | $\mathrm{kgs^{-1}\,m^{-2}}$ | $\mathrm{kgs^{-1}\,m^{-3}}$ |
| Energy, heat | $[\text{velocity}]^2$ | W | $\mathrm{Wm^{-2}}$ | $\mathrm{Wm^{-3}}$ |
| Electrical charge | Coulomb $\mathrm{kg^{-1}}$ | A | $\mathrm{Am^{-2}}$ | $\mathrm{Am^{-3}}$ |
| Impulse | [velocity] | [force] | [pressure] | [pressure] $\mathrm{m^{-1}}$ |

For $\vec{Q}$, the flux densities are second-order tensors, since $\mathrm{d}\vec{\Phi}_{\vec{Q}}=\vec{\vec{j}}_{\vec{Q}}\cdot\mathrm{d}\vec{S}$ is vectorial (Figure 1). Its balance reads

$$\frac{\partial}{\partial t}(\rho\vec{q})+\vec{\nabla}\cdot\left(\vec{\vec{j}}_{\vec{Q}}^{\,t}+\rho\vec{v}\otimes\vec{q}\right)=\vec{S}_{\vec{Q}}\qquad[4]$$

where $^t$ indicates the transposition and $\otimes$ a dyadic product. $\vec{j}_Q$ and $\vec{\vec{j}}_{\vec{Q}}$ are given later.

The governing equations of thermohydraulics are like [3] and [4]. They are completed by compatible initial and boundary conditions. The most general linear expression of the latter ones is of mixed type, for a scalar field,

$$\alpha q+\beta\left(\vec{\nabla}\cdot\hat{\boldsymbol{n}}\right)q=\gamma\quad\text{on the boundary}\qquad[5]$$

$\alpha,\beta$, and $\gamma$ being prescribed data, and $\hat{\boldsymbol{n}}$ the outward normal to the boundary. For a vectorial field, $\vec{q}$ and $\vec{\gamma}$, respectively, replace $q$ and $\gamma$. The simplest cases are Dirichlet and Neumann boundary conditions with, respectively, $\beta=0$ or $\alpha=0$.

## Governing Equations

We consider nonisothermal flows of fluids in thermodynamic conditions far from the critical point where acoustic effects are involved. The fluid is possibly a binary mixture, the simplest non-pure-fluid case where modeling does not raise conceptual difficulties. The

local composition is described by the solute (say) mass fraction,

$$C(M,t) = \lim_{\Delta\mathcal{V}\to 0} \frac{\Delta m_{\text{solute}}}{\Delta m} = \frac{\rho_{\text{solute}}}{\rho}$$

with $0 \le C \le 1$. Only thermodiffusion is treated, and the influence the solutal gradient has on the heat flux is not considered, being negligible in liquid mixtures. The coupling between the heat and species molecular transports then comes only in the solutal flux density relation

$$\vec{j}_{\text{solute}} = -\rho\kappa_C(T,C)\left[\vec{\nabla}C + C(1-C)S_T\vec{\nabla}T\right] \quad [6]$$

with $\kappa_C > 0$, and $S_T(T,C)$, the solute Soret coefficient, which is positive or negative. The order of magnitude of the Soret coefficient in the molecular solutions does not exceed few $10^{-2}\,\text{K}^{-1}$, while for colloidal solutions (ferrofluids) $|S_T|$ can be in the range 0.03–0.5 $\text{K}^{-1}$. Even if small, the induced mass fraction separation, $\Delta C \simeq S_T\Delta T$, generates a solutal buoyancy of significant dynamical influence.

### Equation of State for the Density

One must first describe the sensitivity of the density, $\rho\,(p,T,C)$, upon pressure, temperature, and mass fraction in static conditions. The pressure and temperature effective ranges, $\Delta p$ and $\Delta T$, are assumed small enough compared to their respective mean values, $p_0$ and $T_0$, for the local (at $\rho_0 \equiv \rho(p_0, T_0, C_0)$) tangent to $\rho(p,T,C)$ to be a good approximation in most cases,

$$\frac{\rho-\rho_0}{\rho_0} = \chi(p-p_0) - \alpha_T(T-T_0) + \alpha_C(C-C_0) \quad [7]$$

where

$$\chi = \frac{1}{\rho_0}\left(\frac{\partial\rho}{\partial p}\right)\bigg|_0$$

and

$$\alpha_T = -\frac{1}{\rho_0}\left(\frac{\partial\rho}{\partial T}\right)\bigg|_0, \qquad \alpha_C = \frac{1}{\rho_0}\left(\frac{\partial\rho}{\partial C}\right)\bigg|_0$$

are the compressibility, thermal, and solutal expansion positive coefficients, and $C_0$ is the solute mean mass fraction. Thermodynamic properties of some fluids are given in Table 3. Equation [7] is valid if $\chi\Delta p$, $\alpha_T\Delta T$, and $\alpha_C|\Delta C|$ are $\ll 1$. Moreover, in laboratory experiments and industrial processes, one generally has $\Delta p/p_0 \ll \Delta T/T_0$. The pressure term in [7] can thus be neglected in thermohydraulics.

**Table 3** Some values of density, thermal expansion and compressibility coefficients, specific heat at constant pressure, and sound speed at $p = 1$ atm and $T = 293$ K; in SI units

| *Fluid* | $\rho$ | $\alpha T$ | $\chi p$ | $C_p$ | $c$ |
|---|---|---|---|---|---|
| Air | 1.205 | 1 | 1 | 1005 | 344 |
| Helium | 0.167 | 1 | 1 | 5227 | 1010 |
| $CO_2$ | 1.841 | 1 | 1 | 832 | 269 |
| Water | 1000 | 0.0607 | $4.91 \times 10^{-5}$ | 4182 | 1461 |
| Glycerol | 1250 | 0.148 | $2.2 \times 10^{-6}$ | 2333 | 2044 |
| Mercury | 13579 | 0.0533 | $3.76 \times 10^{-6}$ | 1391 | 1409 |

Notice that water density exhibits a maximum around 4°C. A quadratic term in $T$ must then be added to [7].

### The Boussinesq Approximations

The parameter $\alpha_T\Delta T \ll 1$ is the primary source of thermohydraulics. Therefore, the $\vec{v}, p, T$ and $C$ fields can be expanded in series of terms of increasing power in $\alpha_T\Delta T$. The leading term of each series contains an important part of the interesting dynamics. The forthcoming equations are given in the corresponding approximation framework. They contain many simplifications, due to Boussinesq. For instance, the conductivities and diffusivities are taken as constant, as well as $C(1-C)S_T$ in eqn [6]. The next approximation step, the low-Mach model, keeps the leading compressibility and expansion effects, while discarding the associated acoustic waves. This gives access to thermo-soluto-acoustic phenomena. Expansion oscillations are indeed able to trigger, and sustain, acoustic waves provided phase agreements are fulfilled. This second-order model is not presented here.

The compliance with the criteria $\alpha_T\Delta T \ll 1$ and $\alpha_C|\Delta C| \ll 1$ must be checked case by case. The section "Steady parallel-flow model" briefly illustrates this point with an example of thermally driven flow. Furthermore, the $T$- and $C$-sensitivity of $S_T$ is an experimental fact that requires a generic approach of the problem. The $C$-sensitivity of the physical properties is generally more pronounced, nonmonotonic, for instance, over $C \in [0,1]$, than their $T$-sensitivity.

### Boussinesq Local Balances

**Mass** It reads $\partial\rho/\partial t + \vec{\nabla}\cdot(\rho\vec{v}) = 0$, or equivalently $(1/\rho)(D\rho/Dt) = -\vec{\nabla}\cdot\vec{v}$. The fluid particle density varies along its trajectory by compressibility and thermo-solutal expansion. At the leading order in $\alpha_T\Delta T$ and $\alpha_C|\Delta C|$, the latter is negligible, whereas the former is associated with acoustics effects, also negligible when the fluid velocity is much smaller

than the sound speed. The mass balance equation then reduces to

$$\vec{\nabla} \cdot \vec{v} = 0 \qquad [8]$$

Only transverse velocity waves (or shear waves) are allowed by this equation, $\vec{v} \simeq e^{i(\vec{k}\cdot\vec{r}+\omega t)}$ with $\vec{k} \cdot \vec{v} = 0$, since acoustics contributions are discarded.

**Impulse** The impulse molecular flux density is

$$\vec{\vec{j}}_{\rho\vec{v}} = p\,\vec{\vec{1}} - \mu_{\vec{v}}\left[\vec{\nabla} \otimes \vec{v} + (\vec{\nabla} \otimes \vec{v})^t\right]$$

where $\mu_{\vec{v}}$ is the impulse conductivity and $\vec{\vec{1}}$ the Kronecker tensor. A Newtonian fluid is defined as having $\mu_{\vec{v}}$ constant with respect to the rate-of-strain tensor $\vec{\nabla} \otimes \vec{v}$. The impulse balance then reads

$$\frac{\partial}{\partial t}(\rho\vec{v}) + \vec{\nabla} \cdot (\rho\vec{v} \otimes \vec{v}) + \vec{\nabla} \cdot \vec{\vec{j}}_{\rho\vec{v}} = \rho\vec{\Gamma}$$

In the source term $\rho\vec{\Gamma}$, $\vec{\Gamma} = \vec{g}$ for gravity-driven buoyant flows.

With the aforementioned approximations, the impulse balance becomes

$$\frac{D\vec{v}}{Dt} = -\frac{1}{\rho_0}\vec{\nabla}P + \frac{\rho - \rho_0}{\rho_0}\vec{g} + \nu\vec{\nabla}^2\vec{v} \qquad [9]$$

with

$$\frac{\rho - \rho_0}{\rho_0} = -\alpha_T(T - T_0) + \alpha_C(C - C_0)$$

$$\nu = \frac{\mu_{\vec{v}}}{\rho_0}$$

the impulse diffusivity, and the pressure $P = p - p_{0,h}$, $p_{0,h}$ satisfying the hydrostatic relation

$$\vec{\nabla}p_{0,h} = \rho_0\vec{g}$$

In the rotating frame of vector $\vec{\Omega}(t)$,

$$\frac{\rho - \rho_0}{\rho_0}\vec{\Omega} \wedge (\vec{\Omega} \wedge \vec{r}) + 2\vec{\Omega} \wedge \vec{v} + \frac{d\vec{\Omega}}{dt} \wedge \vec{r}$$

must be subtracted from the right-hand side of [9] and $p_{0,h}$ redefined by

$$\vec{\nabla}p_{0,h} = \rho_0\left(\vec{g} - \vec{\Omega} \wedge (\vec{\Omega} \wedge \vec{r})\right)$$

On a free surface, a particular velocity boundary condition is to be established. Let $d\vec{S} = dS\,\hat{n}$ be a surface element located around $M$. The tangential component ($\hat{t} \cdot \hat{n} = 0$) of the impulse flux across $d\vec{S}$,

$$\hat{t} \cdot d\vec{f} = \hat{t} \cdot \vec{\vec{j}}_{\rho\vec{v}} \cdot d\vec{S} = -\mu_{\vec{v}}\hat{t} \cdot \left[\vec{\nabla} \otimes \vec{v} + (\vec{\nabla} \otimes \vec{v})^t\right] \cdot d\vec{S}$$

must be continuous. Surface tension $\sigma(T, C)$ inhomogeneities make the free surface a source of impulse which diffuses in the fluid core. A flow occurs even with $\vec{\Gamma} = 0$. For the fluid located where $d\vec{S}$ points to, the velocity boundary condition on the free surface then reads

$$-\mu_{\vec{v}}\hat{t} \cdot \left[\vec{\nabla} \otimes \vec{v} + (\vec{\nabla} \otimes \vec{v})^t\right] \cdot \hat{n} = (\vec{\nabla} \cdot \hat{t})\sigma \qquad [10]$$

with

$$(\vec{\nabla} \cdot \hat{t})\sigma = \frac{\partial\sigma}{\partial T}(\vec{\nabla} \cdot \hat{t})T + \frac{\partial\sigma}{\partial C}(\vec{\nabla} \cdot \hat{t})C$$

For most fluids, $\partial\sigma/\partial T < 0$. In the Boussinesq framework $\partial\sigma/\partial T$ and $\partial\sigma/\partial C$ are constant. Equation [10] couples the impulse balance with the heat and composition ones.

**Heat** Local thermodynamic equilibrium is assumed. The molecular heat flux density is $\vec{j}_{\text{heat}} = -\mu_T\vec{\nabla}T$, with $\mu_T$ the thermal conductivity. The approximate heat balance reads

$$\frac{DT}{Dt} = \kappa_T\vec{\nabla}^2 T + S_{\text{heat}} \qquad [11]$$

where $\kappa_T = \mu_T/(\rho_0 C_p)$ is the heat diffusivity and $S_{\text{heat}}$ a possible local (Joule, radioactive,...) heat source. Thermohydraulics can simply be driven by nonuniform thermal conditions imposed along the fluid boundary, and in this article we henceforth take $S_{\text{heat}} = 0$.

**Mass fraction** Approximating [6] yields the mass fraction balance,

$$\frac{DC}{Dt} = \kappa_C\vec{\nabla}^2 C + C_0(1 - C_0)S_T\vec{\nabla}^2 T \qquad [12]$$

where $\kappa_C$ and $S_T$ are evaluated at $T_0$ and $C_0$. The normal flux condition

$$\left(\vec{\nabla}C \cdot \hat{n}\right) = -C_0(1 - C_0)S_T\left(\vec{\nabla}T \cdot \hat{n}\right)$$

is imposed on impervious boundaries.

## The Hydrostatic State

Knowing whether the fluid can be in static state with respect to its presupposed rigid container helps for a first understanding of thermohydraulic dynamics. This raises two problems: (1) the existence of this state and (2) its stability, discussed

later. Point (1) requires the fulfilment of three relations,

$$\vec{\nabla}p = \rho(p, T, C)\vec{\Gamma} \quad [13]$$

$$\frac{\partial T}{\partial t} = \kappa_T \vec{\nabla}^2 T$$
$$\frac{\partial C}{\partial t} = \kappa_C \vec{\nabla}^2 C + C_0(1 - C_0)S_T \vec{\nabla}^2 T \quad [14]$$

The curl of [13] yields

$$\vec{\nabla}\rho(p, T, C) \wedge \vec{\Gamma} + \rho(p, T, C)\vec{\nabla} \wedge \vec{\Gamma} = 0$$

which has no reason to be generically satisfied since $\rho(p, T, C)$ and $\vec{\Gamma}$ are totally uncorrelated. The hydrostatic state cannot exist if $\vec{\Gamma}$ does not derive from a scalar potential, as with

$$\vec{\Gamma} = \vec{g} - \vec{\Omega} \wedge (\vec{\Omega} \wedge \vec{r}) - \frac{\mathrm{d}\vec{\Omega}}{\mathrm{d}t} \wedge \vec{r} \quad \text{if} \quad \frac{\mathrm{d}\vec{\Omega}}{\mathrm{d}t} \neq 0$$

The Earth's rotation axis is known to precess with a period of about 26 000 years. This generates a component of 26 000 years timescale in the atmospheric, oceanic, and internal flows.

Considering now that

$$\vec{\Gamma} = -\vec{\nabla}\psi$$

the existence of a hydrostatic state only depends on the simultaneous verification of [14] and

$$\vec{\nabla}\rho(p, T, C) \wedge \vec{\nabla}\psi = 0 \quad [15]$$

Iso-$\psi$ surfaces must therefore coincide with isopycnal, isobaric, iso-$T$, and iso-$C$ surfaces since the $p$, $T$, and $C$ sensitivities of $\rho$ are uncorrelated. The compatibility of this condition with [14] is the key for concluding about the existence of the hydrostatic state. Considering again our planet as an example (forgetting about precession), the iso-$\psi$ surfaces are almost ellipsoidal. Such $T$ and $C$ distributions cannot satisfy [14]. Thus, the atmospheric and oceanic dynamics, and thermohydraulics as well, are due to a nonvanishing thermal torque, $\vec{\nabla}T \wedge \vec{\nabla}\psi$.

A free surface in hydrostatic state is isothermal and isocompositional, by eqn [10], whatever $\vec{\Gamma}$.

## Dimensionless Local Balances

In buoyancy-driven thermohydraulics, we consider four velocity scales – three of molecular origin, and the fourth is the free-fall velocity in the buoyancy,

**Table 4** Orders of magnitude of the Prandtl number for the usual fluids. Air and water are in normal conditions

| Liquid metals | Gases | Water | Oils |
|---|---|---|---|
| Several $10^{-3}$ – $10^{-2}$ | $\simeq 1$, 0.7 for air | 6.7 | >10 |

$$V_1 = \frac{\kappa_T}{L}, \quad V_2 = \frac{\kappa_C}{L}, \quad V_3 = \frac{\nu}{L}$$
$$V_4 = \sqrt{\alpha_T \Delta T g L}$$

$L$ being a fluid container size scale. Thence come the Rayleigh, Prandtl and Lewis numbers,

$$Ra = \frac{V_4^2}{V_1 V_3} = \alpha_T \Delta T \frac{gL^3}{\nu \kappa_T}$$
$$Pr = \frac{V_3}{V_1} = \frac{\nu}{\kappa_T}, \qquad Le = \frac{V_2}{V_1} = \frac{\kappa_C}{\kappa_T}$$

$Ra$ being the experimental control parameter, and $Le \ll 1$. Table 4 gives $Pr$ orders of magnitude for usual fluids. Let $V$ be the fluid velocity amplitude. The importance of the thermal, solutal, and impulse convections with respect to the corresponding diffusions is, respectively, estimated by the thermal, compositional Péclet and Reynolds numbers,

$$Pe_T = \frac{V}{V_1} = \frac{VL}{\kappa_T}, \qquad Pe_C = \frac{V}{V_2} = \frac{VL}{\kappa_C}$$
$$Re = \frac{V}{V_3} = \frac{VL}{\nu}$$

with

$$Pr = \frac{Pe_T}{Re}, \quad Le = \frac{Pe_C}{Re}, \qquad Ra = (Pe_T Re)_{|_{V=V_4}}$$

Capillary thermohydraulics introduces one velocity scale and the Marangoni number,

$$V_5 = \frac{|\Delta\sigma|}{\mu_{\vec{v}}}, \qquad Ma = \frac{V_5}{V_1} = Pe_T$$

with $\Delta\sigma = (\mathrm{d}\sigma/\mathrm{d}T)\Delta T$ in pure fluid. A small capillary number, $Ca = |\Delta\sigma|/\sigma$, indicates a weak influence of the dynamics upon the free-surface curvature.

Let $V_1$, $\Pi = \rho_0 V_1^2$, $\tau = L/V_1$, $\Delta T$ and

$$\Delta C = -C_0(1 - C_0)S_T \Delta T$$

be the velocity, pressure, time, temperature, and mass fraction scales, with

$$\Theta = \frac{T - T_0}{\Delta T} \quad \text{and} \quad \mathcal{C} = \frac{C - C_0}{\Delta C}$$

the reduced temperature and mass fraction, respectively. The other quantities, coordinates included, are similarly reduced and noted identically.

Equation [8] does not change and [9], [11] and [12] become, respectively,

$$\frac{D\vec{v}}{Dt} = -\vec{\nabla}P + Pr\left[Ra(\Theta + \Psi_B\mathcal{C})\hat{e}_z + \vec{\nabla}^2\vec{v}\right] \qquad [16]$$

$$\frac{D\Theta}{Dt} = \vec{\nabla}^2\Theta \qquad [17]$$

$$\frac{D\mathcal{C}}{Dt} = Le\,\vec{\nabla}^2\mathcal{C} - \vec{\nabla}^2\Theta \qquad [18]$$

where

$$\Psi_B = -\frac{\alpha_C \Delta C}{\alpha_T \Delta T}$$

is the buoyancy separation ratio and $\hat{e}_z = -\vec{g}/|\vec{g}|$. A $\Psi_B < 0$ ($>0$) corresponds to opposite (cooperative) thermal and solutal buoyancies. The reduced mass fraction boundary condition on impervious walls is

$$\left(\vec{\nabla}\mathcal{C}\cdot\hat{n}\right) = \left(\vec{\nabla}\Theta\cdot\hat{n}\right) \qquad [19]$$

In rotating frame, scaling $\vec{\Omega}(t)$ by $\Omega_0$, $\vec{\vec{\Omega}}(t) = \vec{\Omega}(t)/\Omega_0$,

$$Ra\ Fr\,(\Theta + \Psi_B\mathcal{C})\vec{\vec{\Omega}} \wedge (\vec{\vec{\Omega}} \wedge \vec{r}) - \frac{1}{Ek}\left(2\vec{\vec{\Omega}} \wedge \vec{v} + \frac{d\vec{\vec{\Omega}}}{dt} \wedge \vec{r}\right)$$

must be added inside the square-bracket term of [16]. The Froude and Ekman numbers appear as

$$Fr = \frac{\Omega_0^2 L}{g}, \qquad Ek = \frac{\nu}{\Omega_0 L^2}$$

The dimensionless capillarity stress condition [10] reads

$$\hat{t}\cdot\left[\vec{\nabla}\otimes\vec{v} + (\vec{\nabla}\otimes\vec{v})^t\right]\cdot\hat{n} = -Ma\left((\vec{\nabla}\cdot\hat{t})\Theta + \Psi_C(\vec{\nabla}\cdot\hat{t})\,\mathcal{C}\right) \qquad [20]$$

with

$$\Psi_C = \frac{\partial\sigma/\partial C}{\partial\sigma/\partial T}\frac{\Delta C}{\Delta T}$$

the capillarity separation ratio, and

$$Ma = \left|\frac{\partial\sigma}{\partial T}\frac{\Delta T}{\mu_{\vec{v}} V_1}\right|$$

These equations show that, in the Boussinesq framework, the flow physics does not depend on $p_0$, $T_0$, and $C_0$, except through the material properties which enter the numbers.

## Linear Stability

Given a base state $\mathcal{S} = (\vec{v}, \Theta, \mathcal{C})$, a solution of [8], [16]–[18], how does it behave in presence of an infinitesimal disturbance $(\delta\vec{v}, \delta\Theta, \delta\mathcal{C})$? Applying [8], [16]–[18] to $(\vec{v}+\delta\vec{v}, \Theta+\delta\Theta, \mathcal{C}+\delta\mathcal{C})$ and discarding the quadratic terms in perturbation provide the disturbance temporal evolution,

$$\vec{\nabla}\,.\,(\delta\vec{v}) = 0 \qquad [21]$$

$$\frac{\partial}{\partial t}\begin{pmatrix}\delta\vec{v}\\ \delta\Theta\\ \delta\mathcal{C}\end{pmatrix} = \mathcal{F} + \left(\delta\vec{v}\cdot\vec{\nabla}\right)\begin{pmatrix}\vec{v}\\ \Theta\\ \mathcal{C}\end{pmatrix} + \mathcal{A}\begin{pmatrix}\delta\vec{v}\\ \delta\Theta\\ \delta\mathcal{C}\end{pmatrix} \qquad [22]$$

where $\mathcal{F} = (-\vec{\nabla}(\delta P), 0, 0)^t$, and

$$\mathcal{A} = \begin{pmatrix}\mathcal{B}_{Pr} & Ra\,Pr\,\hat{e}_z & Ra\,Pr\,\Psi_B\,\hat{e}_z\\ 0 & \mathcal{B}_1 & 0\\ 0 & -\vec{\nabla}^2 & \mathcal{B}_{Le}\end{pmatrix} \qquad [23]$$

with $\mathcal{B}_a = -(\vec{v}\cdot\vec{\nabla}) + a\vec{\nabla}^2$. The perturbations ($\delta\vec{v}$, $\delta\Theta$, $\delta\mathcal{C}$) have the $(\vec{v}, \Theta, \mathcal{C})$ boundary conditions, but homogeneous. On a free surface, the perturbation capillary stress condition is

$$\hat{t}\cdot\left[\vec{\nabla}\otimes\delta\vec{v} + (\vec{\nabla}\otimes\delta\vec{v})^t\right]\cdot\hat{n} = -Ma\left((\vec{\nabla}\cdot\hat{t})\delta\Theta + \Psi_C(\vec{\nabla}\cdot\hat{t})\delta\mathcal{C}\right) \qquad [24]$$

Recasting [21]–[23] provides

$$\frac{\partial}{\partial t}\begin{pmatrix}\delta\vec{v}\\ \delta\Theta\\ \delta\mathcal{C}\end{pmatrix} = \mathcal{L}(\mathcal{S})\begin{pmatrix}\delta\vec{v}\\ \delta\Theta\\ \delta\mathcal{C}\end{pmatrix} \qquad [25]$$

whose solution is

$$\begin{pmatrix}\delta\vec{v}(t)\\ \delta\Theta(t)\\ \delta\mathcal{C}(t)\end{pmatrix} = e^{\mathcal{L}(\mathcal{S})t}\begin{pmatrix}\delta\vec{v}(t=0)\\ \delta\Theta(t=0)\\ \delta\mathcal{C}(t=0)\end{pmatrix} \qquad [26]$$

### Direct System

$\mathcal{L}(\mathcal{S})$ is made of $\vec{\nabla}$ acting on the initial perturbation. Conclusions about $\mathcal{S}$ stability depend on the sign of $\lambda_{max}$, the real part of the leading eigenvalue of $\mathcal{L}$ found with all the possible perturbations. There is stability if $\lambda_{max} < 0$. At $\lambda_{max} = 0$, the marginal stability, the bifurcation threshold is located at $Ra\,(Pr, Le, \Psi_B, \Psi_C, X) = Ra_c$, $Ra_c$-being the critical value of the control parameter, $X$ containing all the other parameters of the problem (container aspect ratios, etc.). The nonlinear-stability analysis in the vicinity of $Ra_c$ supplies $\xi$ in $\lambda_{max} \propto (Ra - Ra_c)^{\xi}$, which is characteristic of the bifurcation.

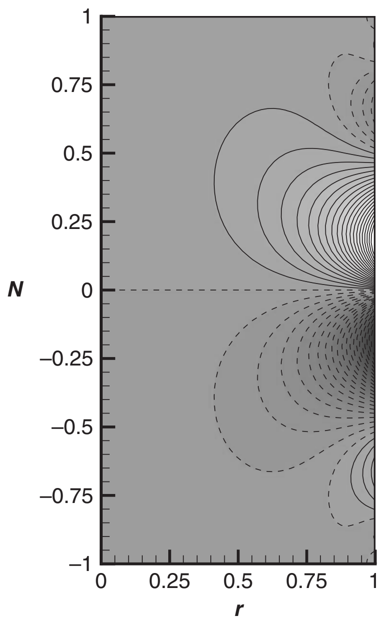



**Figure 2** Leading axisymmetric thermal adjoint eigenvector (Courtesy of O Bouizi and C Delcarte).

### Adjoint System

The leading left eigenmode complex conjugate supplies the response field of the base state to the most destabilizing punctual disturbances.

The $\mathcal{S}$ state and $\mathcal{L}$ eigenspace analytical determinations are often impossible. One must resort to specifically designed numerical tools. A numerical adjoint eigenvector is presented in Figure 2 for a ($Ma = 106$, $Pr = 10^{-2}$) side-heated cylindrical liquid bridge, with a free surface on the right and the axis on the left.

## Nonlinear Stability

When $\lambda_{\max} > 0$, the associated disturbance exponentially grows with time, until nonlinearities become essential. The flow progressively evolves from $\mathcal{S}$ towards a new state, $\mathcal{S}'$, which is a solution of [8], [16]–[18]. How can one proceed analytically to know how the nonlinearities control the bifurcation? A large number of $\mathcal{S} \rightarrow \mathcal{S}'$ bifurcations exist, with either both $\mathcal{S}, \mathcal{S}'$, steady or unsteady but with different flow structure, or one is steady and the other is not. Bifurcations can also be reversible or hysteretic, with respect to $Ra$. The symmetries of $\mathcal{S}$ play an important role and non-Boussinesq effects change the thresholds and the nature of bifurcation.

Landau's works have opened up the way to the theory of nonlinear hydrodynamic stability. The ruling equations are reduced, using an appropriate expansion method, to a set of ordinary differential equations describing the temporal evolution of amplitudes, $A_i$, $i = 1, 2, \ldots, I$, characterizing the perturbation eigenmodes,

$$\frac{\mathrm{d}A_i}{\mathrm{d}t} = \lambda_i A_i + N_i(A_j) \quad \text{for } i, j = 1, 2, \ldots, I \qquad [27]$$

where $N$ accounts for the nonlinear action of the $I$ modes on $A_i$, and the $\lambda_i$'s are the temporal growth rates coming from the linear theory. The stability of the steady solutions, $\mathrm{d}A_i/\mathrm{d}t = 0$, is determined by local analysis. With one destabilizing mode, the simplest model is $\mathrm{d}A/\mathrm{d}t = \lambda A - \alpha A|A|$, with $\alpha > 0$, constant, specific of the bifurcation. Symmetry considerations (some of them directly originate from the Boussinesq framework) may impose $\alpha = 0$, whereby the simplest model becomes $\mathrm{d}A/\mathrm{d}t = \lambda A + \beta A^3$, with $\beta$ another constant.

When the flow is weakly confined in one or two space directions, boundary effects can play a subtle dynamical role, allowing, for instance, the existence of multiple solutions, each one made of many interacting modes. A large variety of flow regimes is then observed, as steady/traveling, extended/localized wave packets, particularly in binary mixtures. Spacetime models, close to [27], such as the Ginzburg–Landau equation,

$$\frac{\partial A}{\partial t} = \lambda A + \alpha \frac{\partial^2 A}{\partial x^2} + \beta |A|^2 A$$

are derived for describing the dynamics of the wave packet envelop (of complex amplitude $A$).

## Hydrostatic State Stability

The static-state stability is analytically tractable in unbounded volume. Transverse wave (by [21]) solutions are the potentially destabilizing perturbations, with wave vector $\vec{k}$ and complex frequency $\omega$. The system [22]–[23] gets simplified, and $\mathcal{L}$ becomes algebraic upon substituting $(\mathrm{i}\vec{k}, i\omega)$ for $(\vec{\nabla}, \partial/\partial t)$. Intuitively, the quiescent state loses its stability when $\vec{\nabla}\rho(p, T, C) \cdot \vec{\nabla}\psi$ exceeds a threshold value (positive, by the dissipative effects). This analysis supplies it, together with the data of the oscillatory motions emerging at onset from the rest-state instability.

In reality, the fluid is confined to three dimensions, possibly with free surfaces, and wave solutions are no longer usable. The first approach consists in defining a simplified model confined to one dimension. The perturbations must satisfy homogeneous boundary conditions, and/or [24], and they are waves in both other space directions. The resulting problem may be analytically tractable. The stability of many quiescent-state configurations was studied, for fluid layers of infinite or very large

extension, of pure-fluid/mixtures, with/without free surface. Nonetheless, many other configurations are not yet analyzed. Two- and three-dimensional cases must be numerically treated.

## Gravitational Buoyancy Convection

Among the numberless thermal situations to analyze, research mainly favored the case where the fluid is confined in simple geometries and submitted to two distinct heating directions, $\vec{\nabla}T$ being either aligned or normal to $\vec{\Gamma}$, that is vertical or horizontal in the gravity field. Each case leads to specific thermohydraulics. The rest-state stability is the first analysis step of the former case, the first to be experimentally studied by Bénard in 1900, with a horizontal liquid layer. The latter is of more recent interest, with Batchelor's theoretical work on the parallel convective regimes of pure fluid confined in tall slot. Since then, a large amount of work has been published on those cases, tackling various confinement geometries, and involving high $Ra$ values. This problem became the paradigm of the rich spatiotemporal behaviors arising in nonlinear systems driven away from equilibrium. In binary mixtures the complexity of the dynamics increases considerably. The literature is so far practically devoid of any three-dimensional results in mixtures. Ternary mixtures have so far been only scarcely considered.

### Steady Parallel-Flow Model

This analytical approach comes from an interesting Batchelor's remark made about the vorticity but here applied to the velocity of a confined flow. "A number of flow fields are characterized by values of the magnitude of the" velocity "in the neighborhood of a certain line in the fluid which are much larger than those elsewhere," and (by $\vec{\nabla}\cdot\vec{v}=0$) "this line of necessity" is parallel to $\vec{v}$ and to the container walls.

Buoyant forces may contradict this assertion, particularly in Rayleigh–Bénard configuration with imposed temperatures. There, no parallel solution exists. Nevertheless, steady parallel flows do exist in containers. The thermally active walls (whatever they be – the largest or smallest) are either maintained at constant temperatures, or subjected to a constant heat flux. Figure 3 sketches a cross section (hereafter referred to as the vertical midplane) of such a configuration, with active (uniform heating $q$) vertical walls. The other sides are adiabatic. No rest state is allowed here. Although intrinsically three dimensional, the steady regime in

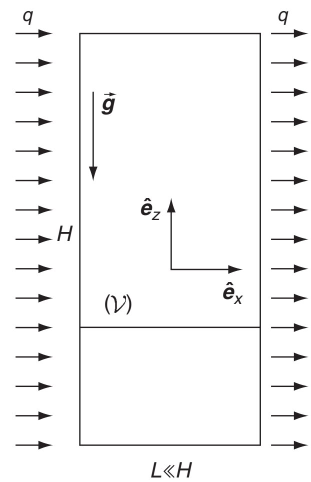



**Figure 3** Sketch of the cross section of a slender vertical container.

this cavity can be fairly well approximated as two dimensional (in the vertical midplane), and moreover mainly parallel to the active walls, in an $Ra$ range which increases with the aspect ratio, $H/L$. The influence of the horizontal sides is of limited range compared to the flow extension, $H$. The parallel flow is then the one-dimensional approximation of what occurs in the major part of the cavity. This configuration is taken with a binary mixture for illustrating an approach applicable with minor variations in other situations.

The problem becomes linear. Indeed, $\vec{v}=w(x)\hat{\boldsymbol{e}}_z$ by $\vec{\nabla}\cdot\vec{v}=0$. Taking $\Delta T=qL/\mu_T$ as temperature scale, [16]–[18] imply

$$\Theta(x,z)=G_T z+\hat{\Theta}(x),\qquad \mathcal{C}(x,z)=G_C z+\hat{\mathcal{C}}(x)$$

with $G_T, G_C$ as constants. The impulse balance is

$$\frac{\mathrm{d}^2 w}{\mathrm{d}x^2}=-Ra\left[\hat{\Theta}(x)+\Psi_\mathrm{B}\hat{\mathcal{C}}(x)\right] \qquad [28]$$

and the ruling equations

$$\begin{aligned}\frac{\mathrm{d}^4 w}{\mathrm{d}x^4}&=-Ra\left[G_T+\frac{\Psi_\mathrm{B}}{Le}(G_T+G_C)\right]w\\ wG_T&=\frac{\mathrm{d}^2\hat{\Theta}}{\mathrm{d}x^2},\qquad w(G_T+G_C)=Le\frac{\mathrm{d}^2\hat{\mathcal{C}}}{\mathrm{d}x^2}\end{aligned} \qquad [29]$$

An internal length scale is predicted, of thickness

$$\left[Ra\left(G_T+\frac{\Psi_\mathrm{B}}{Le}(G_T+G_C)\right)\right]^{-1/4}$$

By [28] and [19], the thermal flux condition yields

$$\left.\frac{\mathrm{d}^3 w}{\mathrm{d}x^3}\right|_{x=\pm 1/2}=-Ra\,(1+\Psi_\mathrm{B})$$

A last operation allows to determine $G_T$ and $G_C$. The overall heat and mass fraction balances are performed in the cavity part $(\mathcal{V})$, which is bounded by an horizontal plane located within the parallel-flow region. Since the walls are impervious, the solute is transported only across the lower boundary of $(\mathcal{V})$, through which the net vertical convective supply must be balanced, in steady regime, by vertical diffusion. The heat balance works similarly, since the walls are adiabatic or submitted to equal fluxes. Whence the relations,

$$\int_{-1/2}^{1/2} w(x)\hat{\Theta}(x)\,\mathrm{d}x = G_T$$

$$\int_{-1/2}^{1/2} w(x)\hat{C}(x)\,\mathrm{d}x = G_T + G_C$$

The steady parallel flow is determined. Its stability can be analyzed as indicated in the section "Linear stability."

Some caution must be taken for the Boussinesq approximations to be valid here, with the temperature and mass fraction increasing constantly (by $G_T, G_C$) along the direction of largest cavity extension. These gradients are at the origin of the "thermogravitational column" separation power, a device designed for the isotope separation. Extremely long columns can provide almost complete separations, with $\alpha_C|\Delta C|$ no longer $\ll 1$, and then the non-Boussinesq effects occur.

As an illustration of aforementioned notions, let us consider the $(Pr = 1,\ Le = 0.1)$ Rayleigh–Bénard–Soret (RBS) problem where horizontal solid plates of infinite extension are uniformly heated from above $(Ra < 0)$ or below $(Ra > 0)$. This configuration is simply obtained by rotating the cavity in Figure 3 by $\pm\pi/2$ with respect to $\vec{g}$ and to $(\hat{e}_x, \hat{e}_z)$. The steady parallel-flow model can lead to the right-hand side of an equation like [27] governing the time evolution of $A$, the parallel-flow amplitude,

$$\frac{\mathrm{d}A}{\mathrm{d}t} \propto A\left[Le^{-2}A^4 + \alpha\left(1 + \frac{1-r}{Le^2}\right)A^2 + \alpha^2\left(1 - \frac{r}{r_c}\right)\right] \qquad [30]$$

where

$$\alpha = \frac{315}{218}, \quad r = \frac{Ra}{720}, \quad r_c = \left[1 + \Psi_B(1 + Le^{-1})\right]^{-1}$$

Here $r_c$ is the critical value or $r$ where the rest state loses its stability towards a steady parallel flow. The roots of $\mathrm{d}A/\mathrm{d}t = 0$ are $A = A_0 = 0$, $A = A_\parallel(r, Le, \Psi_B)$, for the quiescent, convective states. Figure 4 shows that $A_0 = 0$ and the curves $A_\parallel(r)$ for several $\Psi_B(Le)$, $\varphi_1 = -(1 + Le^{-1})^{-1}$ being the $r_c$ pole. The solid (dotted) parts correspond to the stable (unstable) steady states, emerging from direct (backward) pitchfork bifurcations of the rest state at $r_c$. Saddle–node bifurcations from unstable to stable steady states are also predicted, on the dashed curve of the equation

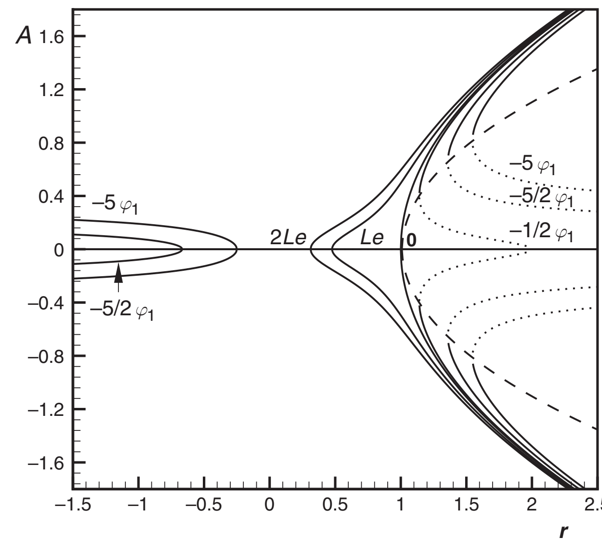



**Figure 4** Bifurcation diagram of $A_0(r)$ and $A_\parallel(r)$ for various separation ratios $\Psi_B(Le)$.

$$\widehat{A_\parallel}(r) = \pm\sqrt{\frac{\alpha}{2}}\sqrt{r - (1 + Le^2)}$$

## Fully Nonlinear Problem

Numerical tools are required for solving the system [8], [16]–[18] and analyzing the stability of the flows obtained.

**The RBS Case** Let us illustrate how the rest-state loss of stability occurs in the two-dimensional RBS case, with a $(Pr = 1,\ Le = 0.1,\ \Psi_B = -0.2)$ mixture. The flow lies in the meridian plane of an axisymmetric container with the radius/height ratio equal to 2. No-slip conditions are imposed on impervious walls; the temperature on the bottom plate is higher than on top, and the peripheral wall is adiabatic. At $t = 0$, the quiescent state is given a small random perturbation. The system evolves (Figure 5) towards a stable periodic solution via a transient regime of exponentially amplified amplitude (eqn [26]). One speaks of a Hopf bifurcation for a steady (here quiescent) state destabilization by oscillatory disturbances.

The "instantaneous" frequency (from the time running between two successive identical passes of

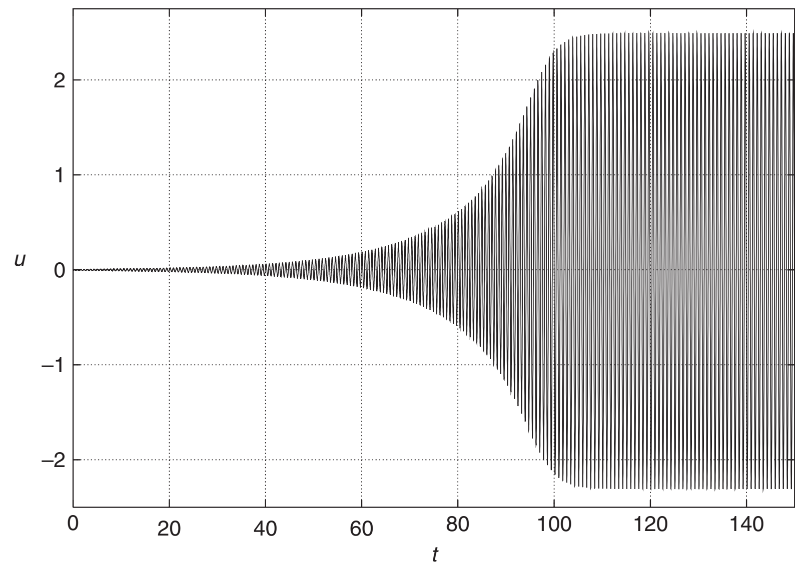



**Figure 5** Time evolution of a radial velocity nodal value for $Ra = 2600$. Reproduced from Millour, Labrosse, and Tric (2003) *Physics of Fluids* 15(10): 2791–2802, with permission from American Institute of Physics.

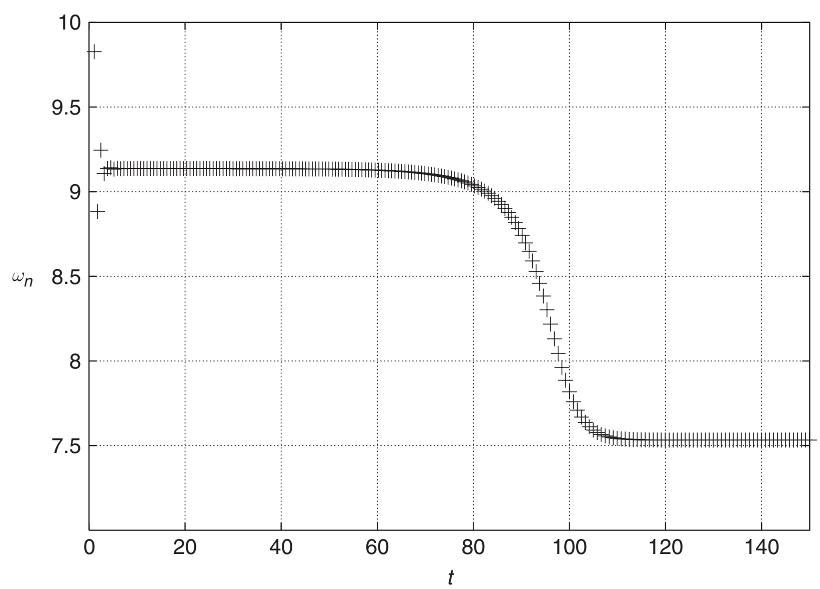



**Figure 6** Instantaneous angular frequency $\omega_n$ corresponding to **Figure 5**. Reproduced from Millour, Labrosse, and Tric (2003) *Physics of Fluids* 15(10): 2791–2802, with permission from American Institute of Physics.

the signal) evolves with time (Figure 6) from its threshold value to its nonlinearly saturated one.

Accurate determination the thresholds and identification of the associated bifurcation is possible by fitting the argument $\xi$ of $\lambda_{\max}(Ra)$ from the exponential growth of Figure 5, in the $Ra_c$ vicinity. Figure 7 shows (solid dots) $\lambda(Ra)$ measurements, and the solid line (in **Figure** 8 also) is the linear law given by the two points closest to the vanishing growth rate. The local law announced in the subsection "Direct system" is confirmed, with an exponent $\xi = 1$ for the Hopf bifurcation, and $\xi = 1/2$ for saddle–node (Figure 8) and pitchfork bifurcations.

**The Thermally Driven Cubic Cavity** All flows are obviously three dimensional. When do they possess a two-dimensional approximation? How to qualify it? Clearly, the flow that develops in the container of Figure 3 might enjoy (in a given parameter domain, $\mathcal{D}$) the mirror-reflection symmetry property about the vertical midplane. Is there a two-dimensional

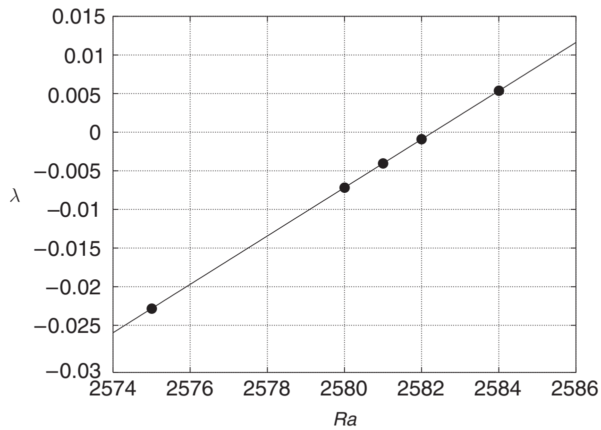



**Figure 7** Temporal growth rate, $\lambda$, of infinitesimal perturbations, in the vicinity of the Hopf bifurcation of the quiescent state. Reproduced from Millour, Labrosse, and Tric (2003) *Physics of Fluids* 15(10): 2791–2802, with permission from American Institute of Physics.

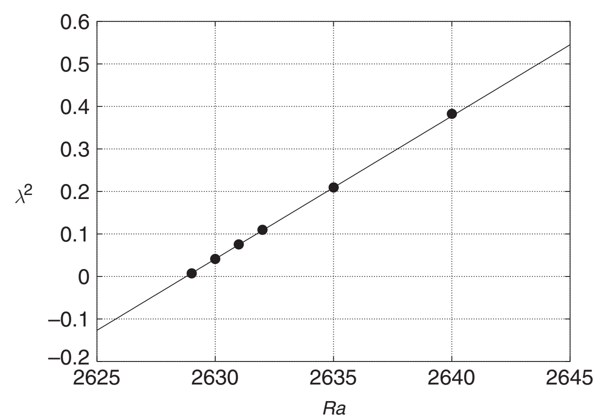



**Figure 8** Squared temporal growth rate, $\lambda^2$, of transient relaxation towards the stationary state close to the saddle–node bifurcation. Reproduced from Millour, Labrosse, and Tric (2003) *Physics of Fluids* 15(10): 2791–2802, with permission from American Institute of Physics.

approximation of the flow in this midplane? Is it able to give a correct estimate of the two-dimensional flow stability within $\mathcal{D}$, and to predict the $\mathcal{D}$ frontiers, where the mirror-reflection symmetry property ceases to be valid? Only partial answers are available so far, coming from the thermally driven cubic cavity (Figure 9).

Filled with a pure fluid, its left and right vertical plates have fixed temperatures, $T_0$ ($\Theta = 0$ at $x = 0$) and $T_0 + \Delta T$ ($\Theta = 1$ at $x = 1$), while the others are adiabatic. Any $\Delta T \neq 0$ generates a flow, possibly mirror-symmetric about the vertical (hatched) midplane, and also centrosymmetric about $\hat{e}_y$. The two-dimensional approximation was extensively analyzed, numerically, with air as a fluid. A steady flow is obtained for $Ra < Ra_{2D,c} = (1.82 \pm 0.01) \times 10^8$,

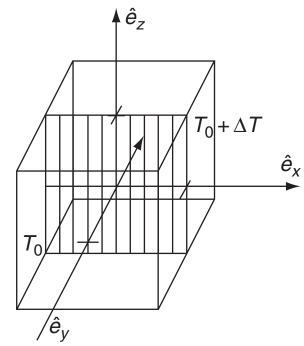



**Figure 9** Sketch of the thermally driven cubic cavity.

where an oscillatory regime appears. The numerical three-dimensional flow is steady until $Ra_{3D,c} = 3.2 \times 10^7$, where it hysteretically bifurcates towards an oscillatory regime breaking the mirror symmetry about the midplane. Let us assess the validity of the two-dimensional approximate solutions. We define dimensionless heat fluxes (Nusselt numbers) which penetrate in one of the active walls,

$$Nu(y) = \int_0^1 \left. \frac{\partial \Theta_{3D}}{\partial x} \right|_{x=0} dz$$

Three fluxes are interesting to compare: (1) in the midplane, $Nu_{mp} = Nu(y = 1/2)$, (2) globally $Nu_{3D,W} = \int_0^1 Nu(y)\, dy$, and (3) the two-dimensional approximation

$$Nu_{2D,W} = \int_0^1 \left. \frac{\partial \Theta_{2D}}{\partial x} \right|_{x=0} dz$$

Figure 10 shows how they compare themselves, as a function of $Ra$. Quantitatively, the two-

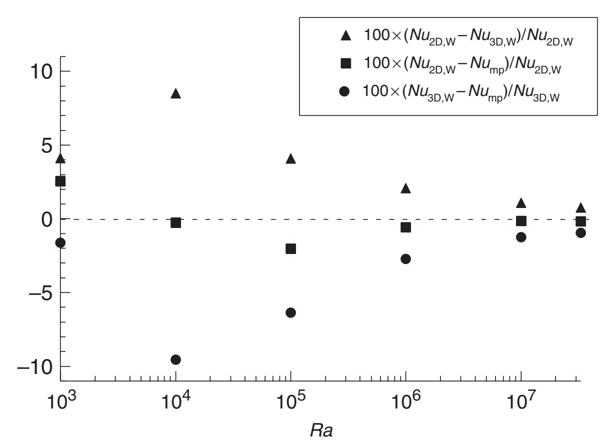



**Figure 10** Relative 2D–3D Nusselt numbers. Reproduced with permission from Tric E, Labrosse G, and Betrouni M (2000) A first incursion into the 3D structure of natural convection of air in a differentially heated cubic cavity from accurate numerical solutions. *International Journal of Heat and Mass Transfer* 43: 4043–4056. © Elsevier Ltd.

dimensional approximation is not too bad, but not qualitatively, with a nonmonotonic evolution of the discrepancies. These latter become quite negligible when the three-dimensional flow gets unsteady and paradoxically loses the symmetry property on which its two-dimensional approximation is founded.

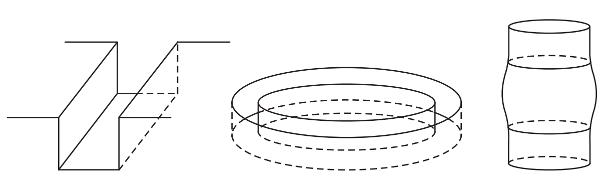
## Thermocapillary Convection

Two immiscible liquids, or a liquid and a gas, are separated by a free surface, a region of small thickness (some ten molecular sizes). From a macroscopic viewpoint, it is considered as a singular entity. Its location and geometry are part of the solutions of the governing equations, themselves supposed to satisfy [20] on the free surface. As a first iteration, the free-surface shape can be imposed, fixed, and straight often.

Numerous industrial processes involve thermocapillarity wherein thermohydraulics involves complex phenomena, such as phase-change kinetics. A relevant modeling of these situations is a research subject by itself. For thermohydraulics, some academic configurations (**Figure 11**) have retained the attention of the scientific community.

Any thermohydraulic flow transfers heat between hot and cold solid boundaries wherein heat penetrates by conduction. Consequently, the term $(\vec{\nabla} \cdot \hat{\boldsymbol{t}})\Theta$ of [20] never cancels at the solid boundary/free surface junction, as in **Figure 12**.

A nonzero vorticity is thus generated by thermocapillarity on the free surface until the wall, while flow adherence on the wall gives vorticity values of opposite sign. The problem presents therefore a vorticity singularity at the triple point. This is a deep physical and modeling problem.



(a) (b) (c)

**Figure 11** Open boat ((a) straight and (b) circular) and liquid bridge (c) configurations.

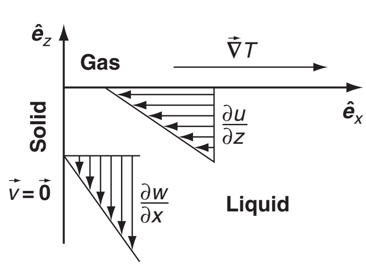



**Figure 12** Thermocapillary origin of vorticity singularity (cold wall configuration).

*See also:* Bifurcations in Fluid Dynamics; Capillary Surfaces; Compressible Flows: Mathematical Theory; Dynamical Systems and Thermodynamics; Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Fluid Mechanics: Numerical Methods; Magnetohydrodynamics; Non-Newtonian Fluids; Partial Differential Equations: Some Examples; Stability of Flows; Vortex Dynamics.

## Further Reading


Batchelor GK (1987) *An Introduction to Fluid Dynamics*. Cambridge: Cambridge University Press.

Bender CM and Orszag SA (1999) *Advanced Mathematical Methods for Scientists and Engineers I*. New York: Springer.

Bird RB, Stewart WE, and Lightfoot EN (1960) *Transport Phenomena*. New York: Wiley.

Chandrasekhar S (1961) *Hydrodynamic and Hydromagnetic Stability*. Oxford: Clarendon.

Colinet P, Legros JC, and Velarde M (2001) *Nonlinear Dynamics of Surface-Tension-Driven Instabilities*. Berlin: Wiley.

Drazin PG and Reid WH *Hydrodynamic Stability*, Cambridge Monographs on Mechanics and Applied Mathematics. Cambridge: Cambridge University Press.

Johns LE and Narayanan R (2002) *Interfacial Instability*. New York: Springer.

Koschmieder EL (1993) *Bénard Cells and Taylor Vortices*, Cambridge Monographs on Mechanics and Applied Mathematics. Cambridge: Cambridge University Press.

Kuhlmann HC (1999) Thermocapillary Convection in Models of Crystal Growth. *Springer Tracts in Modern Physics,* vol. 152. Berlin: Springer.

Labrosse G (2003) Free convection of binary liquid with variable Soret coefficient in thermogravitational column: the steady parallel base states. *Physics of Fluids* 15(9): 2694–2727.

Lücke M, Barten W, Büchel P, *et al.* (1998) Pattern formation in binary fluid convection and in systems with throughflow. In: Busse FH and Müller SC (eds.) *Evolution of Structures in Dissipative Continuous Systems*, Lecture Notes in Physics. New York: Springer.

Manneville P (1990) *Structures Dissipatives, Chaos and Turbulence*. Gif-sur-Yvette (France): Collection Aléa Saclay.

Narayanan R and Schwabe D (eds.) (2003) *Interfacial Fluid Dynamics and Transport Processes*, Lecture Notes in Physics, vol. 628. Berlin: Springer.

Platten JK and Legros JC (1984) *Convection in Liquids*. Berlin: Springer.

Turner JS (1973) *Buoyancy Effects in Fluids*, Cambridge Monographs on Mechanics and Applied Mathematics. Cambridge: Cambridge University Press.

# Newtonian Limit of General Relativity

**J Ehlers**, Max Planck Institut für Gravitationsphysik (Albert-Einstein Institut), Golm, Germany



## Introduction

The general theory of relativity (GRT) unifies special relativity theory (SRT) and Newton's theory of gravitation (NGT). SRT and NGT describe successfully large domains of physical phenomena; therefore, one would like to understand how they survive as approximations in GRT.

In GRT, spacetime is idealized as a four-dimensional Lorentz manifold whose curvature is related to the distribution of energy and momentum. In such a spacetime, the existence of the exponential map implies that the metric near any event (spacetime point) $x$ deviates from a flat metric only by terms given by the curvature there. Thus, if the gravitational tidal field, represented by the curvature tensor, is small near $x$, one may approximate the GR metric there by a flat Minkowski metric. This explains that SRT is a general local approximation to GRT. Apart from a remark at the end of the subsection "Local laws" the relation GRT $\rightarrow$ SRT will not be discussed further.

In its traditional formulation, Newton's theory differs drastically from Einstein's theory both in its spacetime structure and in its description of gravitation. The main purpose of this report is to show how NGT can nevertheless be understood as a kind of "limit" of GRT. More precisely, the structure of NGT can be viewed as a degenerate version of that of GRT, in parallel to the fact that the Galilei group can be obtained by contracting the Lorentz group.

In the next section we state the laws of GRT. We then reformulate these laws with slightly different field variables such that, besides the gravitational constant $k$, the speed of light appears via $\lambda = c^{-2}$. The resulting laws remain meaningful if $\lambda$ and/or $k$ are replaced by zero. They turn out to give a common basis for GRT, SRT, and NGT. The possibility of such a framework was indicated independently by Cartan (1923, 1924) and Friedrichs (1927) and extended by several authors; the complete formulation reviewed here was given by Ehlers (1981).

The section "Newton's theory in spacetime form" shows that the *laws* of NGT and SRT are obtained, with some additional restrictions, from the rescaled laws of GRT by putting, respectively, $\lambda = 0$ or $k = 0$. It is emphasized that Newton's theory proper is a theory only of isolated systems. Its intrinsic, four-dimensional formulation explains how the distinction between a vectorial gravitational field and inertial forces, as well as the existence of inertial frames, emerge as consequences of asymptotic flatness. These structures are lost in the so-called "Newtonian" cosmology whose dynamics is due to symmetry assumptions, whereas GR cosmology is a proper part of GRT.

The penultimate section is concerned with relations between *solutions* of GRT and NGT, and in the final section some results related to solutions are reported. They illustrate that the limit relation GRT $\rightarrow$ NGT may sometimes be inverted to get exact or approximate GR results from NGT. Approximations are related to uniform convergence in $\lambda$, as is indicated at the end of the final section.

The limit relations described here may be considered as a model for other theory relations in physics such as quantization or dequantization.

*Notation* Indices will be considered in general as "abstract" ones, characterizing the kind of objects independent of coordinate systems. Greek indices refer to spacetime, Latin ones to 3-space. Fields on spacetime will generally be taken to be smooth.

## Basic Concepts and Laws of GRT

According to GRT, spacetime is a four-dimensional manifold $M$ endowed with a Lorentzian metric $g_{\alpha\beta}$, here taken to have signature $(+ + + -)$. Any kind of matter including nongravitational fields is supposed to determine an energy tensor $T^{\alpha\beta}$. Metric and matter are interrelated by Einstein's gravitational field equation

$$R_{\alpha\beta} = \frac{8\pi k}{c^4}\left(T_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}T\right) \qquad [1]$$

In this equation, $T := T^\alpha{}_\alpha$ denotes the trace of the energy tensor, $k$ and $c$ stand for Newton's constant of gravity and the speed of light, respectively, and the Ricci tensor $R_{\alpha\beta}$ is obtained from Riemann's curvature tensor by contraction

$$R_{\alpha\beta} := R^\gamma{}_{\alpha\gamma\beta}$$

The curvature tensor is constructed from the symmetric, linear connection $\Gamma_\beta{}^\alpha{}_\gamma$ determined by the metric.

Equation [1] implies the vanishing of the covariant divergence of the energy tensor

$$T^{\alpha\beta}{}_{;\beta} = 0 \qquad [2]$$

the GRT analog of the laws of local conservation of energy and momentum.

The energy tensor depends on the kind of matter to be taken into account. In this article, only vacuum fields ($T^{\alpha\beta}=0$) and perfect fluids will be considered. For such a fluid,

$$T^{\alpha\beta} = (\rho + c^{-2}p)U^{\alpha}U^{\beta} + pg^{\alpha\beta} \qquad [3a]$$

$\rho$ and $p$ denote the mass density and the pressure, respectively, and the 4-velocity $U^{\alpha}$ is a timelike vector obeying

$$g_{\alpha\beta}U^{\alpha}U^{\beta} = -c^2 \qquad [3b]$$

If thermodynamical relations are added to specify the kind of fluid – the simplest cases are barotropic equations $p=f(\rho)$ – then eqns [1]–[3] admit a well-posed initial value problem for the fields $g_{\alpha\beta}, U^{\alpha}, \rho$.

Different matter models which could be treated in the context of this report are elastic bodies and ideal gases, but not point particles. Point particles fit into GRT even less than into electrodynamics.

## The Cartan–Friedrichs Formalism

To obtain a spacetime formulation of NGT and a limit relation ART $\rightarrow$ NGT, we recall that the metric structure of Newton's spacetime consists of a scalar $t$, absolute time, which foliates $M$ into instantaneous 3-spaces $S_t$, and Euclidean metrics $\gamma_{ab}(t)$ on these spaces. If the inverses $\gamma^{ab}(t)$ are pushed forward onto $M$ via the embeddings $S_t \rightarrow M$, a field $s^{\alpha\beta}$ on $M$ results which is assumed to be smooth. By construction,

$$s^{\alpha\beta}t_{,\beta} = 0 \qquad [4]$$

The pair $(t, s^{\alpha\beta})$ defines the "metric," that is, times and distances, in NGT.

Such a structure can arise from a Lorentzian metric, for example, the Minkowski metric $\eta_{\alpha\beta}$, by taking, component-wise, the limits

$$-c^{-2}\eta_{\alpha\beta}\,dx^{\alpha}\,dx^{\beta} \equiv dt^2 - c^{-2}\,dx^2 \underset{c\to\infty}{\longrightarrow} dt^2, \quad \eta^{\alpha\beta} \underset{c\to\infty}{\longrightarrow} s^{\alpha\beta} \qquad [5]$$

which can be interpreted geometrically as "opening up the light cones" until they degenerate into doubly covered, spacelike hyperplanes, the Newtonian $S_t$'s.

The relations [5] suggest to write the GRT laws in terms of the rescaled temporal metric ($\lambda \equiv c^{-2}$)

$$t_{\alpha\beta} := -\lambda g_{\alpha\beta} \qquad [6]$$

and to write – presently only as a change of notation – $s^{\alpha\beta}$ instead of $g^{\alpha\beta}$. Then the fields $t_{\alpha\beta}, s^{\alpha\beta}, \Gamma_{\beta}{}^{\alpha}{}_{\gamma}, T^{\alpha\beta}, \rho, p, U^{\alpha}$, called the basic fields below, and constants $k>0, \lambda>0$ satisfy the following laws:

$$t_{\alpha\gamma}s^{\gamma\beta} = -\lambda\delta^{\beta}{}_{\alpha} \qquad [7a]$$

$$t_{\alpha\beta;\gamma} = 0, \qquad s^{\alpha\beta}{}_{;\gamma} = 0 \qquad [7b]$$

$$R^{\alpha}{}_{\beta}{}^{\gamma}{}_{\delta} = R^{\gamma}{}_{\delta}{}^{\alpha}{}_{\beta} \qquad [7c]$$

$$R_{\alpha\beta} = 8\pi k\left(t_{\alpha\gamma}t_{\beta\delta} - \frac{1}{2}t_{\alpha\beta}t_{\gamma\delta}\right)T^{\gamma\delta} \qquad [7d]$$

$$T^{\alpha\beta}{}_{;\beta} = 0 \qquad [7e]$$

$$T^{\alpha\beta} = (\rho + \lambda p)U^{\alpha}U^{\beta} + ps^{\alpha\beta} \qquad [7f]$$

$$t_{\alpha\beta}U^{\alpha}U^{\beta} = 1 \qquad [7g]$$

The Lorentz signature of $g_{\alpha\beta}$ can be reexpressed thus: at each event ($\cong$ spacetime point), there exists a "timelike" vector $V^{\alpha}$, that is,

$$t_{\alpha\beta}V^{\alpha}V^{\beta} > 0 \qquad [7h]$$

and $V^{\alpha}X_{\alpha}=0$ for $X_{\alpha}\neq 0$ implies $s^{\alpha\beta}X_{\alpha}X_{\beta}>0$.

The indices in eqn [7c] are raised, here and later, by $s^{\alpha\beta}$.

Given a set of basic fields on $M$ as listed below eqn [6], the laws [7] remain meaningful for all $\lambda \geq 0$ and $k \geq 0$. If $\lambda=0$, the "metrics" $t_{\alpha\beta}$ and $s^{\alpha\beta}$ degenerate (and the pair $(t_{\alpha\beta}, s^{\alpha\beta})$ is then called a Galilei metric). Nevertheless, the definition of "timelike" will also be used in that case. Also, $X^{\alpha}$ will be said to be "spacelike" if and only if it can be written $X^{\alpha}=s^{\alpha\beta}\xi_{\beta}$ with $s^{\alpha\beta}\xi_{\alpha}\xi_{\beta}>0$. While for $\lambda>0$, some of the relations [7] are redundant, this is not so for $\lambda=0$. For example, if $\lambda=0$, the two eqns [7b] are independent and do not determine the connection $\Gamma_{\beta}{}^{\alpha}{}_{\gamma}$ uniquely, in contrast to the case $\lambda>0$. The connection will always be assumed to be symmetric.

As will be discussed below, these formulas define a framework which serves to relate GRT to NGT and special relativity (SRT). First steps to formulate such a framework have been taken independently by E Cartan and KO Friedrichs. Therefore we call the structure defined by [7] the Cartan–Friedrichs formalism (CFF). We call it a "formalism" and not a "theory" since it is of interest solely as a tool to study relations between theories.

Equations [7] remain unchanged if the basic fields and constants are rescaled according to a change of units for time, length, and mass. Here, two sets of basic fields related by such a rescaling will be considered as physically equivalent; they provide the

same relations between observables. Thus, $\lambda$ and $k$ have no physical meanings, but only their signs:

$$\lambda > 0,\ k > 0: \quad \text{GRT}$$

$$\lambda = 0,\ k > 0: \quad \text{NGT}$$

$$\lambda > 0,\ k = 0: \quad \text{SRT}$$

(The last two lines are not sufficient to specify the theories within CFF; in connection with eqn [9] and in Theorem 2 they will be completed.) For discussing limit relations between theories, it is nevertheless useful to represent physical models in different scales.

The physical interpretation of $t_{\alpha\beta}, s^{\alpha\beta}$ in terms of time and distance and that of $\Gamma_\beta{}^\alpha{}_\gamma$ through its geodesics as world lines of freely falling test particles, respectively, is the same in the three theories and can be stated in terms of the common framework CFF.

For an obvious reason, $\lambda$ may be called causality constant. Note that $\lambda$ and $k$ each occur in only one of the general laws of the theory, apart from the $\lambda$ in [7f].

The laws [7] are invariant under diffeomorphisms of the spacetime manifold. Those diffeomorphisms which map the basic fields of a solution into themselves form the symmetry group of that solution.

## Newton's Theory in Spacetime Form

### Local Laws

Remarkably, for $\lambda = 0$ and $k > 0$ the formulas [7] reproduce *almost* all the laws on which Newton's theory of spacetime coupled to Euler's fluid theory is based. This is summarized in the following:

**Theorem 1** *Let eqn [7] hold on M with* $\lambda = 0$. *Then there exists, for any event of M, a neighborhood U with coordinates* $(x^a, t)$ *such that, on U, t coincides with the absolute time,* $t_{\alpha\beta} = t_{,\alpha} t_{,\beta}$, *and on the local slices* $U \cap S_t, s^{\alpha\beta}$ *defines Euclidean metrics* $\gamma_{ab}$ *with orthonormal coordinates* $x^a, \gamma_{ab} = \delta_{ab}$. *Vectors are spacelike iff they are tangent to* $S_t$, *otherwise they are timelike. Moreover, the slices are locally geodesic with respect to the connection* $\Gamma_\beta{}^\alpha{}_\gamma$, *and the induced connection on the slices is the flat connection associated naturally to* $\gamma_{ab}$. *In addition, in the coordinate chart given by* $(x^a, t)$, *the connection components vanish except* $\Gamma_0{}^a{}_0$ *and* $\Gamma_0{}^b{}_a (= -\Gamma_0{}^a{}_b)$. *Therefore, t is an affine parameter on timelike geodesics. Further,* $U^0 = 1$, *and* $U^a = v^a$ *is the 3-velocity of the fluid. If one writes*

$$-\Gamma_0{}^a{}_0 =: g^a, \qquad -\Gamma_0{}^a{}_b =: \omega^a{}_b \tag{8a}$$

*and uses 3-vector notation with* $(g^a) = \boldsymbol{g}$, $(\omega_{23}, \omega_{31}, \omega_{12}) = \boldsymbol{\omega}$, *the timelike geodesics of* $\Gamma_\beta{}^\alpha{}_\gamma$ *are given by*

$$\ddot{\boldsymbol{x}} = \boldsymbol{g} + 2\dot{\boldsymbol{x}} \times \boldsymbol{\omega} \tag{8b}$$

$\boldsymbol{g}$ *and* $\boldsymbol{\omega}$ *satisfy*

$$\nabla \cdot \boldsymbol{\omega} = 0, \quad \nabla \times \boldsymbol{g} + 2\dot{\boldsymbol{\omega}} = 0 \tag{8c}$$

$$\nabla \times \boldsymbol{\omega} = 0,\ \nabla \cdot \boldsymbol{g} - 2\boldsymbol{\omega}^2 = -4\pi k \rho \tag{8d}$$

*and the fluid's equations of motion are*

$$\dot{\rho} + \nabla \cdot (\rho \boldsymbol{v}) = 0 \tag{8e}$$

$$\rho(\dot{\boldsymbol{v}} + \boldsymbol{v} \cdot \nabla \boldsymbol{v} - \boldsymbol{g} - 2\boldsymbol{v} \times \boldsymbol{\omega}) + \nabla p = 0 \tag{8f}$$

*A solution* $(\boldsymbol{g}, \boldsymbol{\omega}, \rho, p, \boldsymbol{v})$ *of eqns [8] on a local chart* $(x^a, t)$ *with* $t_{\alpha\beta} = \mathrm{diag}(0,0,0,1)$ *and* $s^{\alpha\beta} = \mathrm{diag}(1,1,1,0)$ *provides, via eqn [8a], the general local solution to eqns [7] for* $\lambda = 0$.

The proof consists of many, mostly elementary steps which can be gathered from Künzle (1972) and Ehlers (1981).

Given a solution to eqns (7) with $\lambda = 0$ and $k > 0$, the coordinates $x^\alpha = (x^a, t)$ referred to in the theorem are determined by the basic fields up to time-dependent Euclidean motions, time translations, and time reflections. Such a coordinate system corresponds to a rigid reference frame. As the equation of motion for freely falling particles, eqn [8b], shows, $\boldsymbol{g}$ and $\boldsymbol{\omega}$ are to be interpreted as the acceleration and rotation fields which determine, relative to a rigid frame, the combined influence of inertia and gravity on particles encoded in the spacetime connection $\Gamma_\beta{}^\alpha{}_\gamma$. (This role of a connection in NGT was recognized by E Cartan.) This interpretation is supported by the (generalized) Euler equation [8f].

As claimed above already, eqns [7] *almost* reproduce the local laws of the Newton–Euler theory. Indeed, eqns [8] are those of the Newton–Euler theory, provided $\boldsymbol{\omega}$ depends on time only. Then and only then can the coordinate freedom be used to get nonrotating rigid coordinates with respect to which $\boldsymbol{\omega} = 0$. The existence of such coordinates is indispensable for NGT since only with respect to them $-\boldsymbol{g}$ is the gradient of a potential $U$ which obeys Poisson's equation, as shown by eqns [8c] and [8d].

The preceding argument shows that the CFF, specialized to $\lambda = 0$, has to be restricted by a condition which implies $\boldsymbol{\omega} = \boldsymbol{\omega}(t)$ in order to give the local laws of NGT. One such condition is

$$R^{\alpha\beta}{}_{\gamma\delta} = 0 \tag{9}$$

as can be verified by computing the curvature tensor via eqn [8a].

Equation [9] for $\lambda = 0$ expresses that parallel transport of spacelike vectors along arbitrary spacetime curves is integrable, which corresponds to the behavior of free gyroscopes in NGT (in contrast to GRT).

Of course, eqn [9] cannot be added to the CFF since it is incompatible with GRT. If, however, the CFF with $\lambda > 0, k = 0$ is restricted by the condition [9], the spacetime and hydrodynamics of special relativity result.

### Global Laws for Isolated Systems

The laws [8] and [9] do not determine the time evolution of the basic fields. Using nonrotating coordinates we put $\boldsymbol{g} = -\nabla U$ and replace eqns [8c], [8d] by Poisson's equation

$$\Delta U = 4\pi k \rho \qquad [10]$$

In Newtonian dynamics, the potential only serves to compute forces depending instantaneously on the mass distribution. Traditionally, this is achieved by assuming $\rho$ to have spatially compact support at each time and to solve eqn [10] by

$$\phi(\boldsymbol{x}, t) = -k \int \frac{\rho(\boldsymbol{x} + \boldsymbol{y}, t)}{|\boldsymbol{y}|} \mathrm{d}^3 y \qquad [11]$$

which implies the fall-off

$$\lim_{\substack{|\boldsymbol{x}| \to \infty, \\ t = \text{const.}}} \phi(\boldsymbol{x}, t) = 0 \qquad [12]$$

($\phi$ will always be used for this solution of eqn [10]).

To relate the foregoing isolation assumptions to corresponding assumptions in GRT as far as presently possible, it seems necessary to go back to the laws [7] restricted to $\lambda = 0$ or the equivalent $(3 + 1)$ version [8] without the restriction [9].

If some global assumptions are added to eqns [8], eqns [10]–[12] can be deduced from the four-dimensional formulation. One first introduces the following two assumptions:

(1) The hypersurfaces $S_t$ of $M$ (which, for $\lambda = 0$, are the only spacelike hypersurfaces) are simply connected, complete Euclidean spaces.
(2) On each $S_t$, the support of $\rho$ is compact.

Using coordinates $(x^a, t)$ as in the last subsection, with $x^a$ now ranging on $\mathbb{R}^3$, eqns [8a] imply

$$R^{\alpha}{}_{\beta\gamma\delta} R^{\beta}{}_{\alpha}{}^{\gamma}{}_{\epsilon} = -2 \sum_{a,b} (\omega_{a,b})^2 t_{\delta\epsilon} \qquad [13]$$

Hence the sum is a 4-scalar, and since $t_{\alpha\beta}$ is covariantly constant, it is possible to require

$$R^{\alpha}{}_{\beta\gamma\delta} R^{\beta}{}_{\alpha}{}^{\gamma}{}_{\epsilon} \to 0 \quad \text{at spatial infinity} \qquad [14]$$

which expresses covariantly that $\omega_{a,b} \to 0$. Since $\boldsymbol{\omega}$ is harmonic on $S_t$ (by eqns [8c], [8d]), this in turn implies $\omega_{a,b} = 0$; thus, $\boldsymbol{\omega}$ depends on $t$ only; the asymptotic condition [14] and the local laws imply eqn [9].

We may therefore employ rigid, nonrotating coordinates, $\boldsymbol{\omega} = 0$. Then, by eqns [8a], [8c], [8d] the connection coefficients take the form

$$\Gamma_{\beta}{}^{\alpha}{}_{\gamma} = t_{,\beta} t_{,\gamma} s^{\alpha\delta} U_{,\delta} \qquad [15]$$

and

$$R^{\lambda}{}_{\alpha\mu\beta} R^{\mu}{}_{\gamma\lambda\delta} = t_{,\alpha} t_{,\beta} t_{,\gamma} t_{,\delta} \sum_{a,b} \sum_{a,b} (U_{,ab})^2 \qquad [16]$$

As before, we require

$$R^{\lambda}{}_{\alpha\mu\beta} R^{\mu}{}_{\gamma\lambda\delta} \to 0 \qquad [17]$$

and conclude $U_{,ab} \to 0$. Since the Newtonian potential $\phi$ of $\rho$ also has this fall-off and $U - \phi$ is harmonic on $S_t \cong \mathbb{R}^3$, the following conclusion can be obtained:

**Lemma 1** *The laws [8] and the global conditions* (1)–(2), [14], [17] *imply: in rigid, nonrotating coordinates, the connection*

$$\Gamma_{\beta}{}^{\alpha}{}_{\gamma} - t_{,\beta} t_{,\gamma} s^{\alpha\delta} \phi_{,\delta} = \mathring{\Gamma}_{\beta}{}^{\alpha} \qquad [18]$$

*is flat ($\phi$ according to eqn [11] is a scalar, and the $\phi$-term in eqn [18] is a tensor). In other words, $\Gamma_{\beta}{}^{\alpha}{}_{\gamma}$ is asymptotically flat since the $\phi$-term falls of as $|\boldsymbol{x}|^{-2}$.*

Because of this lemma, one can further restrict the coordinates $(x^a, t)$ by demanding $\mathring{\Gamma}_{\beta}{}^{\alpha}{}_{\gamma} = 0$. In physical terms this means: by switching to a new, "unaccelerated" frame of reference, one removes from the equations of motion a spatially homogeneous gravitational field which, in contrast to the $\phi$-term in eqn [16], is not due to matter.

The resulting coordinates are defined, up to Galilean transformations,

$$t' = \pm t + c^{0'}$$
$$x^{a'} = D^{a'}{}_{b} x^b + u^{a'} t + c^{a'}$$

where $c^{a'}, u^{a'}$ are constants and $D$ is a constant orthogonal $3 \times 3$ matrix. These coordinates are called inertial ones; with respect to them the usual laws of Newtonian mechanics hold; see [8] with $\boldsymbol{\omega} = 0$ and $U = \phi[\rho]$.

**Theorem 2** (Ehlers 1981). *The laws [7] of the CFF restricted to $\lambda = 0$ and augmented by the global and asymptotic conditions* (1)–(2), [14], [17], *provide a generally covariant, four-dimensional*

*formulation for the Newtonian theory of space, time, gravitation, and hydrodynamics.*

The possibility to split the connection $\Gamma$ into a flat part which is independent of matter and a tensorial part depending on matter and given by the vector field $g^{\alpha} = s^{\alpha\beta}\phi_{,\beta}$ (with $\phi$ from eqn [11]), arises only from supplementing the local laws [7] by the global, resp. asymptotic, conditions (1)–(2), [14], [17] stated above. The introduction of inertial coordinates is then convenient, but not necessary. In noninertial, rigid frames of reference, $\overset{\circ}{\Gamma}{}_{\beta}{}^{\alpha}{}_{\gamma}$ gives rise to inertial forces.

It should be possible to define spatial asymptotic flatness in the CFF, but that has not been done.

### Remarks on Newtonian Cosmology

In cosmology, the conditions (2) and [17] of the last subsection are not appropriate. Instead one keeps the laws [7] and adds to them eqn [9], so that with respect to nonrotating coordinates the laws [8] with $\boldsymbol{\omega} = 0$ and eqn [10] remain valid. Then, there are no longer inertial coordinate systems, and the potential $U$ is not a 4-scalar. For a slightly different approach, see Rüede and Straumann (1997).

For the purpose of this article, the term "cosmological model" will be applied to those solutions of the laws [7] and [9] which satisfy $\rho > 0$ and which have a symmetry group which acts transitively on the set of world lines representing the motion of the fluid. This strong symmetry assumption determines the time-evolution even in the "Newtonian" case $\lambda = 0$ in spite of the absence of an evolution equation for the gravitational field $\boldsymbol{g}$.

## Newtonian Limits of Families of GR Solutions

The discussion in the sections "The Cartan–Friedrichs formalism" and "Newton's theory in spacetime form" suggests the following:

**Definition 1** Let a family $\mathcal{F}(\lambda) = (t_{\alpha\beta}(\lambda), \ldots)$ of basic fields parametrized by $\lambda$, obeying the laws [7] of the CFF, be given for $0 \le \lambda < a$. We assume the underlying manifolds $M(\lambda)$ to be open submanifolds of a fixed manifold $M$ such that $M(\lambda_1) \supseteq M(\lambda_2)$ if $\lambda_1 < \lambda_2$ and $\bigcup_{\lambda} M(\lambda) = M$. Then we write

$$\lim_{\lambda \to 0} \mathcal{F}(\lambda) = \mathcal{F}(0) \qquad [19]$$

if the fields of $\mathcal{F}(\lambda)$ and their first derivatives converge pointwise to those of $\mathcal{F}(0)$.

$\mathcal{F}(0)$ is then said to be a CF limit of the sequence of ($\lambda$-rescaled) solutions $\mathcal{F}(\lambda)$ of GRT. If the fields of a $\lambda$-family of GR solutions ($\lambda > 0$) and their first derivatives converge for $\lambda \to 0$ locally uniformly, then the limit fields satisfy eqns [7]. If $\mathcal{F}(0)$ has the additional property [9], the limit is locally Newtonian.

On the basis of the section "The Cartan–Friedrichs formalism" one may conjecture that if eqn [19] holds and the $\mathcal{F}(\lambda)$ for $\lambda > 0$ are spatially asymptotically flat, $\mathcal{F}(0)$ will represent an asymptotically flat Newtonian spacetime. Examples such as Example 1 below are in agreement with this conjecture, but a general proof is not known.

**Example 1** The interior solution for a static, spherically symmetric fluid ball of constant energy density (Schwarzschild 1916) is given by

$$\mathrm{d}s^2 = \frac{\mathrm{d}r^2}{a^2} + r^2(\mathrm{d}\vartheta^2 + \sin^2\vartheta\,\mathrm{d}\varphi^2) - \frac{1}{4}(3a_0 - a)^2 c^2\,\mathrm{d}t^2$$

$$\rho = \text{const.} > 0, \qquad p = \rho c^2 \frac{a - a_0}{3a_0 - a}$$

$$U = \frac{2}{3a_0 - a}\partial_t, \quad a(r) = \left(1 - \frac{8\pi}{3}kc^{-2}r^2\rho\right)^{1/2}$$

$$a_0 = a(r_0)$$

Inserting into these expressions the parameter $\lambda = c^{-2}$ and treating $\rho$ and $r_0$ as $\lambda$-independent constants results in a $\lambda$-family with $0 \le \lambda < ((8\pi/3)kr_0^3\rho)^{-1}$. The limit solution represents a Newtonian fluid ball of constant mass density $\rho$. The Schwarzschild vacuum fields belonging to these fluid balls also have the appropriate Newtonian limits. The resulting complete spacetimes are asymptotically flat. A dimensionless small parameter which could be used instead of $\lambda$ to measure the deviation of the GR solution from its Newtonian limit is the ratio of Schwarzschild radius and the geometric radius:

$$\frac{2kM}{c^2 r_0} = \frac{8\pi}{3}\frac{k\rho r_0^2}{c^2}$$

**Example 2** A Friedmann–Lemaitre cosmological model of GR containing dust and radiation is given by

$$\mathrm{d}s^2 = R^2(t)\frac{\delta_{ab}\mathrm{d}\xi^a\mathrm{d}\xi^b}{\left(1 - (1/4)(E/c^2)\delta_{ab}\xi^a\xi^b\right)^2} - c^2\mathrm{d}t^2$$

where $R(t)$ obeys

$$\dot{R}^2 - \frac{8\pi}{3}k\left(\frac{M}{R} + \frac{S}{c^2R^2}\right) = E$$

$M$ is a mass constant, $\rho = M/R^3$ is the mass density of "dust," $S$ is an entropy constant, $\epsilon = S/R^4$ the energy density and $p = (1/3)\epsilon$ the pressure of radiation; and $E$ is a constant of dimension (speed)$^2$. The world lines of the fluid elements are given by $\xi^a = \text{const.}$ (Lagrangian comoving coordinates).

Taking $E, M, S$ constant and $\lambda = c^{-2}$ as a parameter provides a $\lambda$-family of GR models with Newtonian limit. In the limit, $t$ is the Newtonian time, and the spatial metric $R^2 \delta_{ab} d\xi^a d\xi^b$ describes an expanding Euclidian space $\mathbb{R}^3$ (if $E \leq 0$) or an open ball of radius $2R(t)$ in it (if $E > 0$). In the coordinates $(\xi^a, t)$ the connection does not have the "Newtonian" components [8a], instead its nonvanishing components are $\Gamma_0{}^a{}_b = (\dot{R}/R)\delta^a_b$. In local inertial coordinates $x^a = R\xi^a$ centered on the particle with $\xi^a = 0$ (which could be any particle because of the homogeneity of the model), the spatial metric is $d\boldsymbol{x}^2$, and the connection components are Newtonian, with $U = (2\pi/3)k\rho \boldsymbol{x}^2$ and $\Delta U = 4\pi k\rho$. In the limit, the radiation no longer influences the expansion; one gets the Newtonian dust models (eqn [9] is satisfied). The connection is, of course, not asymptotically flat. The curvature tensor $R^\alpha{}_\beta{}^\gamma{}_\delta = (4\pi/3)k\rho t_{\beta\delta} s^{\alpha\gamma}$ exhibits homogeneity and isotropy. The Gaussian sectional curvature of the 3-space at time $t$ is $K = -\lambda E/R^2$. As a dimensionless smallness parameter one can take $E/c^2$. In the "open" models, with $E \geq 0$, the coordinates $\xi^a$ cover the whole 3-manifold of fluid particles, while in the "closed" case, $E < 0$, one particle, the antipode of $\xi^a = 0$ on the 3-sphere, is not covered. That particle is missing in the Newtonian limit model. In the Newtonian case the expanding Euclidian space $\mathbb{R}^3$ can be replaced by a torus; in the GR cases this is possible only for $E = 0$.

Many examples of GR families with Newtonian limits are known (see, e.g., Ehlers (1997) and references therein). An example of a $\lambda$-family which has an almost Newtonian limit which does not satisfy eqn [9] is provided by NUT spacetimes (see Ehlers 1997), interpreted as due to a gravitomagnetic monopole (Lynden-Bell and Nouri-Zonez 1998).

## Applications and Problems

Can one construct, for a given Newtonian solution $N$, a $\lambda$-family of GR solutions which converges to $N$? Some answers are known and listed below.

U Heilig (1995) has shown: given a solution to the Euler–Poisson equations representing a stationary, rigidly rotating, self-gravitating fluid body with its surrounding gravitational field, there exists a $\lambda$-family of corresponding solutions to the Einstein–Euler system having the given solutions as its limit.

The proof is based on the fact that one can reformulate eqns [1], [2] in terms of harmonic coordinates and new dependent gravitational variables instead of $g_{\alpha\beta}$ such that the new equations given in Lottermoser (1992) are analytic in $\lambda$ and reduce, for $\lambda = 0$, to the Euler–Poisson system. In the stationary case these equations are elliptic for $\lambda \geq 0$. Using appropriate function spaces, Heilig shows, via the implicit function theorem, that a solution for $\lambda = 0$ can be extended to small, positive values of $\lambda$. Since L Lichtenstein has constructed solutions as assumed in the theorem, the existence of GR solutions follows.

The gravitational part of the system of equations referred to above is hyperbolic for $\lambda > 0$, but becomes elliptic for $\lambda = 0$, whereas the fluid equations remain hyperbolic. In spite of this difficulty Rendall (1994) has shown that $\lambda$-families of time-dependent, asymptotically flat solutions to the Einstein–Vlasov system representing gravitating systems of collisionless particles have Poisson–Vlasov limits, and that any Poisson–Vlasov solution can be so obtained.

Lottermoser (1992) succeeded in proving the existence of $\lambda$-families of solutions to the Einstein constraint equations which have Newtonian initial data as limits. Nothing seems to be known about solutions evolving from such data. Lottermoser has given an interesting discussion concerning possible extension of his work which apparently has gone unnoticed.

Rendall (1992) has defined and analyzed post-Newtonian expansions to Einstein's equations and their solvability, assuming $\lambda$-families whose $t_{\alpha\beta}, s^{\alpha\beta}$ are a few times differentiable in $\epsilon = \sqrt{\lambda}$ at $\epsilon = 0$. He found that for low orders the equations have asymptotically flat solutions, but that at order $\epsilon^8$ divergences occur for general Newtonian seed solutions. Modifications of the method to overcome these difficulties have been considered by Rendall and others; the problem is open.

In cosmology, one uses homogeneous background models and studies their perturbations. The latter are frequently based on Newtonian equations. This can perhaps be justified as follows. According to Example 2 the fields of Friedmann–Lemaitre models differ from their Newtonian limits by arbitrarily small amounts uniformly in spacetime regions where the terms involving $\lambda$ are small, that is,

$$\frac{S}{Mc^2} \ll R(t), \quad \frac{\sqrt{|E|}}{c}|\boldsymbol{x}| \ll R(t)$$

Additional conditions will be needed to ensure that Newtonian perturbations approximate relativistic ones and that gravitational wave perturbations can be neglected.

*See also:* Cosmology: Mathematical Aspects; Einstein Equations: Exact Solutions; General Relativity: Overview; Gravitational Lensing; Shock Wave Refinement of the Friedman–Robertson–Walker Metric.

## Further Reading


Cartan E (1923) Sur les variétés à connexion affine et la théorie de la relativité généraliśee. *Annales Scientifiques de l'Ecole Normale Supérieure* XL: 325–412.

Cartan E (1924) Sur les variétés à connexion affine et la théorie de la relativité généraliśee. *Annales Scientifiques de l'Ecole Normale Supérieure* XLI: 1–25.

Ehlers J (1981) Über den Newtonschen Grenzwert der Einsteinschen Gravitationstheorie. In: Nitsch J *et al.* (eds.) *Grundlagenprobleme der modernen Physik*, pp. 65–84. Mannheim: Bibliografisches Insitut Wissenschaftsverlag.

Ehlers J (1997) Examples of Newtonian limits of relativistic spacetimes. *Classical and Quantum Gravity* 14: A119–A126.

Friedrichs KO (1927) Eine invariante Formulierung des Newtonschen Gravitationsgesetzes und des Grenzüberganges vom Einsteinschen zum Newtonschen Gesetz. *Mathematische Annalen* 98: 566–575.

Heilig U (1995) On the existence of rotating stars in general relativity. *Communications in Mathematical Physics* 166: 457–493.

Künzle HP (1972) Galilei and Lorentz structures on space-time: comparison of the corresponding geometry and physics. *Annales de l'Institut Henri Poincaré* XVII: 337–362.

Lottermoser M (1992) A convergent post-Newtonian approximation for the constraint equations in general relativity. *Annales de l'Institut Henri Poincaré* 57: 279–317.

Lynden-Bell D and Nouri-Zonoz M (1998) Classical monopoles: Newton, NUT space, gravomagnetic lensing, and atomic spectra. *Reviews of Modern Physics* 70: 427–445.

Rendall AD (1992) On the definition of post-Newtonian approximations. *Proceedings of the Royal Society of London Series A* 438: 341–360.

Rendall AD (1994) The Newtonian limit for asymptotically flat solutions of the Vlasov–Einstein system. *Communications in Mathematical Physics* 163: 89–112.

Rüede Ch and Straumann N (1997) On Newton–Cartan cosmology. *Helvetica Physica Acta* 318–335.


# Noncommutative Geometry and the Standard Model

**T Schücker**, Université de Marseille, Marseille, France



## Introduction

The aim of this contribution is to explain how Connes derives the standard model of electromagnetic, weak, and strong forces from noncommutative geometry. The reader is supposed to be aware of two other derivations in fundamental physics: the derivation of the Balmer–Rydberg formula for the spectrum of the hydrogen atom from quantum mechanics and Einstein's derivation of gravity from Riemannian geometry.

At the end of the nineteenth century, new physics was discovered in atoms, namely their discrete spectra. Balmer and Rydberg succeeded to put order into the fast-growing set of experimental results with the help of a phenomenological ansatz for the frequencies $\nu$ of the spectral rays of, for example, the hydrogen atom,

$$\nu = g(n_2^q - n_1^q), \quad n_j \in \mathbb{N}, \;\; q \in \mathbb{Z}, \;\; g \in \mathbb{R} \qquad [1]$$

The integer variables $n_1$ and $n_2$ reflect the discreteness of the spectrum. On the other hand, the discrete parameter $q$ and the continuous parameter $g$ were fitted by experiment: $q = -2$ and $g = 3.289 \times 10^{15}$ Hz, the famous Rydberg constant. Later quantum mechanics was discovered and allowed to derive the Balmer–Rydberg ansatz and to constrain its parameters:

$$q = 2 \quad \text{and} \quad g = \frac{m_e}{4\pi\hbar^3}\frac{e^4}{(4\pi\epsilon_0)^2} \qquad [2]$$

in beautiful agreement with the anterior experimental fit.

## The Standard Model

We propose to introduce the standard model (*see* Standard Model of Particle Physics) in analogy with the Balmer–Rydberg formula (Table 1).

**Table 1** An analogy between atomic and particle physics elements

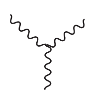
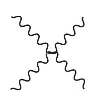
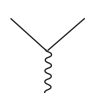
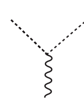
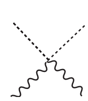
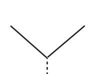
| | *Atomic physics* | *Particle physics* |
|---|---|---|
| New physics | Discrete spectra | Forces mediated by gauge bosons |
| Ansatz | $\nu = g(n_2^q - n_1^q)$ | Yang–Mills–Higgs models |
| Experimental fit | $q = -2, g = 3.289 \times 10^{15}$ Hz | Standard model |
| Underlying theory | Quantum mechanics | Noncommutative geometry |

### The Yang–Mills–Higgs Ansatz

The variables of this Lagrangian ansatz are spin-1 particles $A$, spin-(1/2) particles decomposed into left- and right-handed components $\psi = (\psi_L, \psi_R)$ and spin-0 particles $\varphi$. There are four discrete parameters, a compact real Lie group $G$, the "gauge group," and three unitary representations on complex Hilbert spaces $\mathcal{H}_L, \mathcal{H}_R$, and $\mathcal{H}_S$. The spin-1 particles come in a multiplet living in the complexified of the Lie algebra of $G$, $A \in \mathrm{Lie}(G)^{\mathbb{C}}$. The left- and right-handed spinors come in multiplets living in the Hilbert spaces, $\psi_L \in \mathcal{H}_L, \psi_R \in \mathcal{H}_R$, respectively. The (Higgs) scalar is another multiplet, $\varphi \in \mathcal{H}_S$. The Yang–Mills–Higgs Lagrangian, together with its Feynman diagrams, is spelled out in Table 2.

There are several continuous parameters: the gauge coupling $g \in \mathbb{R}_+$, the Higgs self-couplings $\lambda, \mu \in \mathbb{R}_+$, and several Yukawa couplings $g_Y \in \mathbb{C}$.

**Table 2** The Yang–Mills–Higgs Lagrangian and its Feynman diagrams

| | |
|---|---|
| $\mathcal{L}[A, \psi, \varphi] = \frac{1}{2}\mathrm{tr}(\partial_\mu A_\nu \partial^\mu A^\nu - \partial_\mu A_\nu \partial^\nu A^\mu)$ |  |
| $+g\ \mathrm{tr}(\partial_\mu A_\nu [A^\mu, A^\nu])$ |  |
| $+g^2\ \mathrm{tr}([A_\mu, A_\nu][A^\mu, A^\nu])$ |  |
| $+\bar{\psi} \partial\!\!\!/ \psi$ |  |
| $+\mathrm{i}g\bar{\psi}(\tilde{\rho}_L \oplus \tilde{\rho}_R)(A_\mu)\gamma^\mu \psi$ |  |
| $+\frac{1}{2}\partial_\mu \varphi^* \partial^\mu \varphi$ |  |
| $+\frac{1}{2}g\{(\tilde{\rho}_S(A_\mu)\varphi)^* \partial^\mu \varphi + \partial_\mu \varphi^* \tilde{\rho}_S(A_\mu)\varphi\}$ |  |
| $+\frac{1}{2}g^2(\tilde{\rho}_S(A_\mu)\varphi)^* \tilde{\rho}_S(A^\mu)\varphi$ |  |
| $+\lambda \varphi^* \varphi \varphi^* \varphi$ | |
| $-\frac{1}{2}\mu^2 \varphi^* \varphi$ |  |
| $+g_Y\ \bar{\psi}\varphi\psi + \bar{g}_Y\ \bar{\psi}\varphi^*\psi$ |  |

Let us choose $G = U(1) \ni e^{i\theta}$. Its irreducible unitary representations are all one-dimensional, $\mathcal{H} = \mathbb{C} \ni \psi$ characterized by the charge $q \in \mathbb{Z}$: $\rho(e^{i\theta})\psi = e^{iq\theta}\psi$. Then with $q_L = q_R$ and $\mathcal{H}_S = \{0\}$, we get Maxwell's theory with the photon (or gauge boson or 4-potential) $A$ coupled to the Dirac theory of a massless spinor of electric charge $q_L$ whose (relativistic) wave function is $\psi$. The gauge coupling is given by $g = e/\sqrt{\epsilon_0}$. Gauge invariance of the Yang–Mills–Higgs Lagrangian implies, via Noether's theorem, electric charge conservation in this case (*see* Symmetries and Conservation Laws).

Yang–Mills models are therefore simply nonabelian generalizations of electromagnetism where the abelian gauge group U(1) is replaced by any compact real Lie group. We insist on a compact group because all irreducible unitary representations of compact groups are finite dimensional. Finally, the Higgs scalar is added to give masses to spinors and gauge bosons via spontaneous symmetry breaking (*see* Symmetry Breaking in Field Theory).

We use compact groups and unitary representations as (discrete) parameters. One motivation is Noether's theorem and conserved quantities. The other comes from Wigner's theorem: the irreducible unitary representations of the Poincaré group are classified by mass and spin. Its orthonormal basis vectors are classified by energy–momentum and by the $z$-component of angular momentum. This theorem leads to the widely accepted definition of a particle as an orthonormal basis vector in a Hilbert space $\mathcal{H}$ carrying a unitary representation $\rho$ of a group $G$.

A precious property of the Yang–Mills–Higgs ansatz is its perturbative renormalizability necessary for fine-structure calculations like the anomalous magnetic moment of the muon.

### The Experimental Fit

Physicists have spent some 30 years and some $10^9$ Swiss Francs to distill the fit (Particle Data Group 2004):

$$G = SU(2) \times U(1) \times SU(3)/(\mathbb{Z}_2 \times \mathbb{Z}_3) \qquad [3]$$

$$\mathcal{H}_L = \bigoplus_1^3 \left[(2, \tfrac{1}{6}, 3) \oplus (2, -\tfrac{1}{2}, 1)\right] \qquad [4]$$

$$\mathcal{H}_R = \bigoplus_1^3 \left[(1, \tfrac{2}{3}, 3) \oplus (1, -\tfrac{1}{3}, 3) \oplus (1, -1, 1)\right] \qquad [5]$$

$$\mathcal{H}_S = (2, -\tfrac{1}{2}, 1) \qquad [6]$$

Here $(n_2, y, n_3)$ denotes the tensor product of an $n_2$-dimensional representation of SU(2), "(weak) isospin," an $n_3$-dimensional representation of SU(3), "color," and the one-dimensional representation of

U(1) with "hyper charge" $y$. For historical reasons, the hypercharge is an integer multiple of 1/6. This is irrelevant: in the abelian case, only the product of the hypercharge with its gauge coupling is measurable, and we do not need multivalued representations, which are characterized by noninteger, rational hypercharges. In the direct sum, we recognize the three generations of fermions, the quarks, "up, down, charm, strange, top, bottom," are SU(3) triplets, the leptons, "electron, $\mu, \tau$" and their neutrinos, are color singlets. The basis of the fermion representation space is

$$\begin{pmatrix} u \\ d \end{pmatrix}_L, \quad \begin{pmatrix} c \\ s \end{pmatrix}_L, \quad \begin{pmatrix} t \\ b \end{pmatrix}_L$$
$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$$
$$u_R, c_R, t_R, e_R, \mu_R, \tau_R$$
$$d_R, s_R, b_R,$$

The parentheses indicate isospin doublets.

The eight gauge bosons associated with su(3) are called gluons. Warning: the U(1) is not the one of electric charge; it is called hypercharge, the electric charge is a linear combination of hypercharge and weak isospin. This mixing is necessary to give electric charges to the $W$ bosons. The $W^+$ and $W^-$ are pure isospin states, while the $Z^0$ and the photon are (orthogonal) mixtures of the third isospin generator and hypercharge.

As the group $G$ contains three simple factors, there are three gauge couplings,

$$\begin{aligned} g_2 &= 0.6518 \pm 0.0003 \\ g_1 &= 0.3574 \pm 0.0001 \\ g_3 &= 1.218 \pm 0.01 \end{aligned} \qquad [7]$$

The Higgs couplings are usually expressed in terms of the $W$ and Higgs masses:

$$m_W = \tfrac{1}{2} g_2 \, v = 80.419 \pm 0.056 \text{ GeV} \qquad [8]$$

$$m_\varphi = 2\sqrt{2}\sqrt{\lambda} v > 98 \text{ GeV} \qquad [9]$$

with the vacuum expectation value $v := (1/2)\mu/\sqrt{\lambda}$. Because of the high degree of reducibility of the spin-(1/2) representations there are 27 complex Yukawa couplings. They constitute the fermionic mass matrix which contains the fermion masses and mixings:

$$\begin{aligned} m_e &= 0.510998902 \pm 0.000000021 \text{ MeV} \\ m_u &= 3 \pm 2 \text{ MeV}, \quad m_d = 6 \pm 3 \text{ MeV} \\ m_\mu &= 0.105658357 \pm 0.000000005 \text{ GeV} \\ m_c &= 1.25 \pm 0.1 \text{ GeV}, \quad m_s = 0.125 \pm 0.05 \text{ GeV} \\ m_\tau &= 1.77703 \pm 0.00003 \text{ GeV} \\ m_t &= 174.3 \pm 5.1 \text{ GeV}, \quad m_b = 4.2 \pm 0.2 \text{ GeV} \end{aligned}$$

For simplicity, we have taken massless neutrinos. Then mixing only occurs for quarks and is given by a unitary matrix, the Cabibbo–Kobayashi–Maskawa matrix

$$C_{KM} := \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \qquad [10]$$

whose matrix elements in terms of absolute values are:

$$\begin{pmatrix} 0.9750 \pm 0.0008 & 0.223 \pm 0.004 & 0.004 \pm 0.002 \\ 0.222 \pm 0.003 & 0.9742 \pm 0.0008 & 0.040 \pm 0.003 \\ 0.009 \pm 0.005 & 0.039 \pm 0.004 & 0.9992 \pm 0.0003 \end{pmatrix} \qquad [11]$$

Mathematically, the Cabibbo–Kobayashi–Maskawa matrix comes from a polar decomposition of the mass matrix. The physical meaning of the quark mixings is the following: when a sufficiently energetic $W^+$ decays into a $u$ quark, this $u$ quark is produced together with a $\bar{d}$ quark with probability $|V_{ud}|^2$, an $\bar{s}$ quark with probability $|V_{us}|^2$, and a $\bar{b}$ quark with probability $|V_{ub}|^2$.

The phenomenological success of the standard model is phenomenal: with only a handful of parameters, it reproduces correctly some millions of experimental numbers: cross sections, lifetimes, branching ratios.

## Noncommutative Geometry

Noncommutative geometry is an analytic geometry generalizing three other geometries that also had important impact on our understanding of forces and time. Let us start by briefly recalling the three forerunners (Table 3). Euclidean geometry underlies Newton's mechanics as a geometry in the space of positions. Forces are described by vectors living in the same space and the Euclidean scalar product is needed to define work and potential energy. Time is not part of geometry – it is absolute. This point of view is abandoned in special relativity unifying space and time into Minkowskian geometry. This new point of view allows to derive the magnetic

**Table 3** Four nested analytic geometries

| *Geometry* | *Force* | *Time* |
|---|---|---|
| Euclidean | $E = \int \vec{F} \cdot d\vec{x}$ | Absolute |
| Minkowskian | $\vec{E}, \epsilon_0 \Rightarrow \vec{B}, \mu_0 = \epsilon_0^{-1} c^{-2}$ | Universal |
| Riemannian | Coriolis $\leftrightarrow$ gravity | Proper, $\tau$ |
| Noncommutative | Gravity $\Rightarrow$ YMH, $\lambda = \frac{1}{3} g_2^2$ | $\Delta\tau \sim 10^{-40}$ s |

field from the electric field as a pseudoforce associated with a Lorentz boost. Although time becomes relative, one can still imagine a grid of synchronized clocks, that is, a universal time. The next generalization is "Riemannian geometry = curved spacetime." Here gravity can be viewed as the pseudoforce associated with a uniformly accelerated coordinate transformation. At the same time, universal time loses all meaning and we must content ourselves with proper time. With today's precision in time measurement, this complication of life becomes a bare necessity, for example, the global positioning system (GPS).

Our last generalization is "noncommutative geometry = curved space(time) with an uncertainty principle." As in quantum mechanics, this uncertainty principle is introduced via noncommutativity.

### Quantum Mechanics

Consider the classical harmonic oscillator. Its phase space is $\mathbb{R}^2$ with points labeled by position $x$ and momentum $p$. A classical observable is a differentiable function on phase space such as the total energy $p^2/(2m) + kx^2$. Observables can be added and multiplied, and they form the algebra $\mathcal{C}^\infty(\mathbb{R}^2)$, which is associative and commutative. To pass to quantum mechanics, this algebra is rendered noncommutative by means of a noncommutation relation for the generators $x$ and $p$: $[x,p] = \mathrm{i}\hbar 1$. Let us call $\mathcal{A}$ the resulting algebra "of quantum observables." It is still associative, and has an involution $\cdot^*$ (the adjoint or Hermitian conjugation) and a unit 1.

Of course, there is no space anymore of which $\mathcal{A}$ is the algebra of functions. Nevertheless, we talk about such a "quantum phase space" as a space that has no points or a space with an uncertainty relation. Indeed, the noncommutation relation implies Heisenberg's uncertainty relation $\Delta x \Delta p \geq \hbar/2$ and tells us that points in phase space lose all meaning; we can only resolve cells in phase space of volume $\hbar/2$, see Figure 1. To define the uncertainty $\Delta a$ for an observable $a \in \mathcal{A}$, we need a faithful representation of the algebra on a Hilbert space, that is, an injective homomorphism $\rho$ from $\mathcal{A}$ into the algebra of operators on $\mathcal{H}$. For the harmonic oscillator, this Hilbert space is $\mathcal{H} = \mathcal{L}^2(\mathbb{R})$. Its elements are the wave functions $\psi(x)$, square-integrable functions on configuration space. Finally, the dynamics is defined by the Hamiltonian, a self-adjoint observable $H = H^* \in \mathcal{A}$ via Schrödinger's equation $(\mathrm{i}\hbar\partial/\partial t - \rho(H))\psi(t,x) = 0$. Here time is an external parameter; in particular, time is not an observable. This is different in the special-relativistic setting, where Schrödinger's equation is replaced by Dirac's equation $\partial\!\!\!/\psi = 0$. Now the wave function $\psi$ is the four-component spinor consisting of left- and right-handed, particle and antiparticle wave functions. Unlike the Hamiltonian, the Dirac operator does not lie in $\mathcal{A}$, but it is still an operator on $\mathcal{H}$. In Euclidean spacetime, the Dirac operator is also self-adjoint, $\partial\!\!\!/^* = \partial\!\!\!/$.

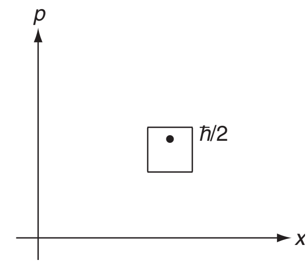



**Figure 1** The first example of noncommutative geometry.

### Spectral Triples

Noncommutative geometry (Connes 1994, 1995) does to a compact Riemannian spin manifold $M$ what quantum mechanics does to phase space. A noncommutative geometry is defined by the three purely algebraic items $(\mathcal{A}, \mathcal{H}, \partial\!\!\!/)$, called a spectral triple. $\mathcal{A}$ is a real, associative, and possibly noncommutative involution algebra with unit, faithfully represented on a complex Hilbert space $\mathcal{H}$, and $\partial\!\!\!/$ is a self-adjoint operator on $\mathcal{H}$. As the spectral triple, also the axioms linking its three items are motivated by relativistic quantum mechanics.

When $\mathcal{A} = \mathcal{C}^\infty(M)$, the functions on a Riemannian spin manifold $M$, represented on spinors $\psi$, and $\partial\!\!\!/$ is the gravitational Dirac operator, one has a spectral triple. The converse is also true when $\mathcal{A}$ is a suitable commutative algebra (Connes 1996), but the axioms make sense even when $\mathcal{A}$ is not commutative. As for quantum phase space, Connes defines a noncommutative geometry by a spectral triple whose algebra is allowed to be noncommutative and he shows how important properties like dimensions, distances, differentiation, integration, general coordinate transformations, and direct products generalize to the noncommutative setting. As a bonus, the algebraic axioms of a spectral triple, commutative or not, include discrete, that is, zero-dimensional spaces that now are naturally equipped with a differential calculus. These spaces have finite-dimensional algebras and Hilbert spaces, meaning that their algebras are just matrix algebras.

An "almost commutative geometry" is defined as a direct product of a four-dimensional commutative geometry, "ordinary spacetime," by a zero-dimensional noncommutative geometry, the "internal space." If the

latter is also commutative, for example, the ordinary two-point space, then the direct product describes a two-sheeted universe or a Kaluza–Klein space whose fifth dimension is discrete, (Madone 1995). In general, the axioms of spectral triples imply that the Dirac operator of the internal space is precisely the fermionic mass matrix.

As a generic example, here is the internal spectral triple underlying the standard model with one generation of quarks and leptons. The algebra $\mathcal{A} = \mathbb{H} \oplus \mathbb{C} \oplus M_3(\mathbb{C}) \ni (a, b, c)$ contains quaternions, that is, $2 \times 2$ matrices of the form

$$a = \begin{pmatrix} x & -\bar{y} \\ y & \bar{x} \end{pmatrix}, \quad x, y \in \mathbb{C}$$

complex numbers $b$ and complex $3 \times 3$ matrices $c$. The Hilbert space is 30-dimensional, where we count particles and antiparticles $(\cdot^c)$ separately: $\mathcal{H} = \mathcal{H}_L \oplus \mathcal{H}_R \oplus \mathcal{H}_L^c \oplus \mathcal{H}_R^c = \mathbb{C}^8 \oplus \mathbb{C}^7 \oplus \mathbb{C}^8 \oplus \mathbb{C}^7$. The representation is block-diagonal, with the four blocks

$$\rho_L(a) := \begin{pmatrix} a \otimes 1_3 & 0 \\ 0 & a \end{pmatrix}$$

$$\rho_R(b) := \begin{pmatrix} b1_3 & 0 & 0 \\ 0 & \bar{b}1_3 & 0 \\ 0 & 0 & \bar{b} \end{pmatrix} \quad [12]$$

$$\rho_L^c(b, c) := \begin{pmatrix} 1_2 \otimes c & 0 \\ 0 & \bar{b}1_2 \end{pmatrix}$$

$$\rho_R^c(b, c) := \begin{pmatrix} c & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & \bar{b} \end{pmatrix} \quad [13]$$

The internal Dirac operator (= fermionic mass matrix) contains two quark masses $m_u, m_d$ and one lepton mass $m_e$, and no mixing:

$$\mathcal{D} = \begin{pmatrix} 0 & \mathcal{M} & 0 & 0 \\ \mathcal{M}^* & 0 & 0 & 0 \\ 0 & 0 & 0 & \bar{\mathcal{M}} \\ 0 & 0 & \bar{\mathcal{M}}^* & 0 \end{pmatrix}$$

$$\mathcal{M} = \begin{pmatrix} \begin{pmatrix} m_u & 0 \\ 0 & m_d \end{pmatrix} \otimes 1_3 & 0 \\ 0 & \begin{pmatrix} 0 \\ m_e \end{pmatrix} \end{pmatrix} \quad [14]$$

These matrices look rather *ad hoc*; they are not. They define an irreducible spectral triple and, for a given algebra, there is only a finite number of such triples.

## The Spectral Action

Chamseddine and Connes (1997) generalize general relativity to noncommutative spacetimes in two strokes, kinematics and dynamics. They explicitly compute this generalization for almost commutative geometries.

*Kinematics* In noncommutative geometry, general coordinate transformations are algebra automorphisms lifted to the Hilbert space of spinors. For almost commutative geometries, these transformations are precisely general coordinate transformations of ordinary spacetime and gauge transformations. Now remember how Einstein uses the equivalence principle to produce "gravity = curvature" starting from the flat metric, which in Connes' language is the ordinary flat Dirac operator. When applied to an almost commutative geometry (Connes 1996), the equivalence principle produces again a curved metric via the ordinary coordinate transformations on $M$, while the gauge transformations applied to the fermionic mass matrix produce a new field, the Higgs scalar $\varphi$. For the example above, this field is precisely the isospin doublet, color singlet with hypercharge –1/2 of eqn [6]. Gauge transformations also apply to the ordinary Dirac operator, thereby producing the gauge fields $A$.

*Dynamics* The group of generalized coordinate transformations allowed us to construct the configuration space. In the almost commutative case it consists of Riemannian metrics, gauge fields, and Higgs scalars. We now want a dynamics on this configuration space. Of course, we want this dynamics to be invariant under the group of generalized coordinate transformations. Note that the spectrum of the Dirac operator is invariant under this group and Chamseddine and Connes (1997) define the spectral action as a regularized partition function of these eigenvalues.

On almost commutative geometries, the spectral action is equal to the Einstein–Hilbert action plus the Yang–Mills–Higgs ansatz (**Figure 2**). In other words, almost commutative geometry explains the forces mediated by gauge bosons and Higgs scalars as pseudoforces accompanying the gravitational force in the same way that Minkowskian geometry (i.e., special relativity) explains the magnetic force as a pseudoforce accompanying the electric force.

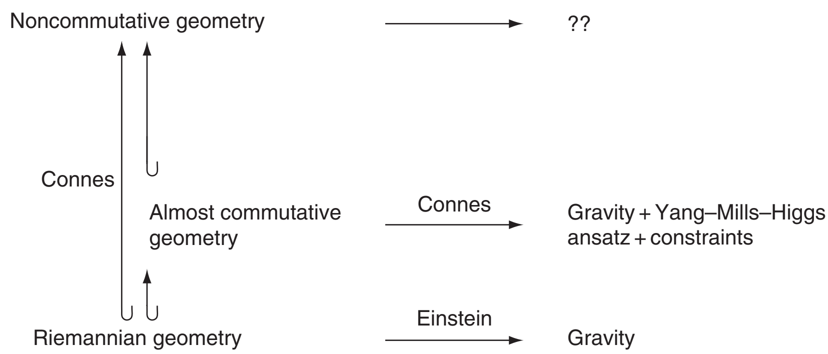



**Figure 2** Deriving the Yang–Mills–Higgs ansatz from gravity.

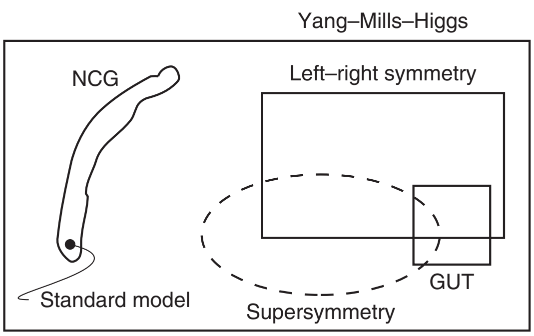



**Figure 3** Constraints inside the ansatz.

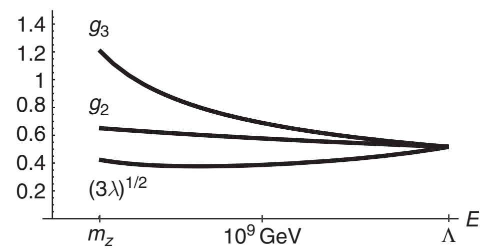



**Figure 4** Running coupling constants.

There are constraints on the discrete and continuous parameters in the Yang–Mills–Higgs ansatz deriving from the spectral action Figure 3.

In particular, if we consider only irreducible spectral triples and among them only those which produce nondegenerate fermion masses compatible with renormalization, then we only get the standard model with one generation of quarks and leptons, with a massless neutrino and with an arbitrary number of colors, and a few submodels thereof. More than one generation and neutrino masses are possible but imply reducible triples. However, in at least one generation, the neutrino must remain purely left and massless.

For the standard model with $N$ generations and $N_c$ colors, we have the constraints $g_{N_c}^2 = g_2^2 = (9/N)\lambda$ on the continuous parameters. If we put $N = N_c = 3$ and if we believe in the popular "big desert" then these constraints yield a "unification scale" $\Lambda = 10^{17}$ GeV at which the uncertainty relation in spacetime should become manifest, $\Delta\tau = \hbar/\Lambda$, and a Higgs mass of $m_\varphi = 171.6 \pm 5$ GeV for $m_t = 174.3 \pm 5.1$ GeV (see Figure 4).

It is clear that almost commutative geometries only scratch the surface of a gold mine. May we hope that a genuinely noncommutative geometry will solve our present problems with quantum field theory and quantum gravity?

*See also:* Compact Groups and Their Representations; Dirac Fields in Gravitation and Nonabelian Gauge Theory; Effective Field Theories; General Relativity: Overview; Hopf Algebras and $q$-Deformation Quantum Groups; Positive Maps on $C^*$-Algebras; Quantum Hall Effect; Standard Model of Particle Physics; Symmetries and Conservation Laws; Symmetry Breaking in Field Theory; von Neumann Algebras: Introduction, Modular Theory, and Classification Theory.

## Further Reading

Chamseddine A and Connes A (1997) The spectral action principle. *Communications in Mathematical Physics* 186: 731 (hep-th/9606001).

Connes A (1994) *Noncommutative Geometry*. San Diego: Academic Press.

Connes A (1995) Noncommutative geometry and reality. *Journal of Mathematical Physics* 36: 6194.

Connes A (1996) Gravity coupled with matter and the foundation of noncommutative geometry. *Communications in Mathematical Physics* 155: 109 (hep-th/9603053).

Gracia-Bondía JM, Várilly JC, and Figueroa H (2000) *Elements of Noncommutative Geometry*. Boston: Birkhäuser.

Kastler D (2000) Noncommutative geometry and fundamental physical interactions: the Lagrangian level. *Journal of Mathematical Physics* 41: 3867.

Landi G (1997) *An Introduction to Noncommutative Spaces and Their Geometry*, hep-th/9701078. Berlin: Springer.

Madore J (1995) *An Introduction to Noncommutative Differential Geometry and Its Physical Applications*. Cambridge: Cambridge University Press.
Martín CP, Gracia-Bondía JM, and Várilly JC (1998) The standard model as a noncommutative geometry: the low mass regime. *Physics Reports* 294: 363 (hep-th/9605001).
O'Raifeartaigh L (1986) *Group Structure of Gauge Theories*. Cambridge: Cambridge University Press.
Scheck F, Werner W, and Upmeier H (eds.) (2002) *Noncommutative Geometry and the Standard Model of Elementary Particle Physics*, Lecture Notes in Physics, vol. 596. Berlin: Springer.
Schücker T (2005) Forces from Connes' geometry. In: Bick E and Steffen F (eds.) *Topology and Geometry in Physics*, Lecture Notes in Physics, vol. 659, hep-th/0111236. Berlin: Springer.
The Particle Data Group, *Particle Physics Booklet* and http://pdg.lbl.gov

# Noncommutative Geometry from Strings

**Chong-Sun Chu**, University of Durham, Durham, UK



## Noncommutative Geometry from String Theory

The first use of noncommutative geometry in string theory appears in the work of Witten on open-string field theory where the noncommutativity is associated with the product of open-string fields. Noncommutative geometry appears in the recent development of string theory in the seminal work of Connes, Douglas, and Schwarz where they constructed and identified the compactification of Matrix theory on a noncommutative torus.

### Matrix Theory Compactification and Noncommutative Geometry

The matrix theory (M-theory) is an 11-dimensional quantum theory of gravity which is believed to underlie all superstring theories. Banks, Fischler, Shenker, and Susskind proposed that the large $N$ limit of the supersymmetric matrix quantum mechanics of $N$ D0-branes should describe the M-theory compactified on a lightlike circle. Compactification of the M-theory on a torus can be easily achieved by considering the torus as the quotient space $\mathbb{R}^d/\mathbb{Z}^d$ with the quotient conditions

$$U_i^{-1} X^j U_i = X^j + \delta_i^j 2\pi R_i, \quad i = 1, \ldots, d \qquad [1]$$

Here $R_i$ are the radii of the torus. The unitary translation generators $U_i$ generate the torus. They satisfy $U_i U_j = U_j U_i$. T-dualizing the D0 brane system, eqn [1] leads to the dual description as a $(d+1)$-dimensional supersymmetric gauge theory on the dual toroidal D-brane. A noncommutative torus $T_\theta^d$ is defined by the modified relations

$$U_i U_j = \mathrm{e}^{\mathrm{i}\theta_{ij}} U_j U_i \qquad [2]$$

where $\theta_{ij}$ specify the noncommutativity. Compactification on a noncommutative torus can be easily accommodated and leads to noncommutative gauge theory on the dual D-brane. The parameters $\theta_{ij}$ can be identified with the components $C_{-ij}$ of the 3-form potential in M-theory.

Since M-theory compactified on a circle leads to IIA string theory, the components $C_{\cdot ij}$ correspond to the Neveu–Schwarz (NS) $B$-field $B_{ij}$ in IIA string theory. The physics of the D0 brane system in the presence of an NS $B$-field can also be studied from the viewpoint of IIA string theory. This led Douglas and Hull to obtain the same result that a noncommutative field theory lives on the D-brane. Toroidally compactified IIA string theory has a T-duality group $SO(d, d; \mathbb{Z})$. The T-duality symmetry gets translated into an equivalence relation between gauge theories on the noncommutative torus: a gauge theory on the noncommutative torus $T_\theta^d$ is equivalent to that on the noncommutative torus $T_{\theta'}^d$ if their noncommutativity parameters and metrics are related by a T-duality transformation. For example,

$$\theta' = (A\theta + B)(C\theta + D)^{-1}, \quad \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in SO(d, d; \mathbb{Z}) \qquad [3]$$

It is remarkable that the T-duality acts within the field theory level, rather than mixing up the field theory modes with the string winding states and other stringy excitations. Mathematically, eqn [3] is precisely the condition for the noncommutative tori $T_\theta^d$ and $T_{\theta'}^d$ to be Morita equivalent.

### Open-String in *B*-Field

It was soon realized that the D-brane does not necessarily need to be toroidal in order to be noncommutative. A direct canonical quantization of the open-string system shows that a constant $B$-field on a D-brane leads to noncommutative geometry on the D-brane world volume. Consider an open string moving in a flat space with metric $g_{ij}$ and a constant NS $B$-field. In the presence of a D$p$ brane, the components of the $B$-field not along the brane can be gauged away; thus, the $B$-field can

have effects only in the longitudinal directions along the brane. The world-sheet (bosonic) action for this part is

$$S = \frac{1}{4\pi\alpha'}\int_\Sigma d^2\sigma \times \left(g_{ij}\partial_a x^i \partial^a x^j - 2\pi\alpha' B_{ij}\epsilon^{ab}\partial_a x^i \partial_b x^j\right) \quad [4]$$

where $i, j = 0, 1, \ldots, p$ is along the brane. It is easy to see that the boundary condition $g_{ij}\partial_\sigma x^j + 2\pi i\alpha' B_{ij}\partial_\tau x^j = 0$ at $\sigma = 0, \pi$ is not compatible with the standard canonical quantization $[x^i(\tau,\sigma), x^j(\tau,\sigma')] = 0$ at the boundary. Taking the boundary condition as constraints and performing canonical quantization, one obtains the commutation relations

$$[a^i_m, a^j_n] = mG^{ij}\delta_{m+n}, \qquad [x^i_0, p^j_0] = iG^{ij}, \qquad [x^i_0, x^j_0] = i\theta^{ij} \quad [5]$$

Here, the open-string mode expansion is

$$x^i(\tau,\sigma) = x^i_0 + 2\alpha'(p^i_0\tau - 2\pi\alpha'(g^{-1}B)^i_j p^j_0\sigma) + \sqrt{2\alpha'}\sum_{n\neq 0}\frac{e^{-in\tau}}{n} \times \left(i a^i_n \cos n\sigma - 2\pi\alpha'(g^{-1}B)^i_j a^j_n \sin n\sigma\right)$$

$G^{ij}$ and $\theta^{ij}$ are the symmetric and antisymmetric parts of the matrix $(g + 2\pi\alpha' B)^{-1ij}$:

$$G^{ij} = \left(\frac{1}{g + 2\pi\alpha' B}\, g\, \frac{1}{g - 2\pi\alpha' B}\right)^{ij}$$
$$\theta^{ij} = -(2\pi\alpha')^2\left(\frac{1}{g + 2\pi\alpha' B}\, B\, \frac{1}{g - 2\pi\alpha' B}\right)^{ij} \quad [6]$$

It follows from [5] that the boundary coordinates $x^i \equiv x^i(\tau, 0)$ obey the commutation relation

$$[x^i, x^j] = i\theta^{ij} \quad [7]$$

Relation [7] implies that the D-brane world volume, where the open-string endpoints live, is a noncommutative manifold. One may also start with the closed-string Green function and let its arguments to approach the boundary to obtain the open-string Green function

$$\langle x^i(\tau)x^j(\tau')\rangle = -\alpha' G^{ij}\ln(\tau - \tau')^2 + \frac{i}{2}\theta^{ij}\epsilon(\tau - \tau') \quad [8]$$

where $\epsilon(\tau)$ is the sign of $\tau$. From [8], one can again extract the commutator [7]. $G_{ij} = g_{ij} - (2\pi\alpha')^2 (Bg^{-1}B)_{ij}$ is called the open-string metric since it controls the short-distance behavior of open strings. In contrast, the short-distance behavior for closed strings is controlled by the closed-string metric $g_{ij}$. One may also treat the boundary $B$-term in [4] as a perturbation to the open-string conformal field theory and from which one may extract [8] from the modified operator product expansion of the open-string vertex operators.

D-branes in the Wess–Zumino–Witten model provide another example of noncommutative geometry. In this case, the background is not flat since there is a nonzero $H = dB \sim k^{-1/2}$, where $k$ is the level. Examining the vertex operator algebra, one obtains that D-branes are described by nonassociative deformations of fuzzy spheres with nonassociativity controlled by $1/k$.

### String Amplitudes and Effective Action

The effect of the $B$-field on the open-string amplitudes is simple to determine since only the $x^i_0$ commutation relation is affected nontrivially. For example, the noncommutative gauge theory can be obtained from the tree-level string amplitudes readily. For tree and one loop, the vertex operator formalism can be used. Generally, the vertex operator can be inserted at either the $\sigma = 0$ or $\sigma = \pi$ boundary, where the string has zero mode parts $x^i_0$ and $y^i_0 \equiv x^i_0 - (2\pi\alpha')^2(g^{-1}B)^i_j p^j_0$, respectively. The commutation relations are

$$[x^i_0, x^j_0] = i\theta^{ij}, \qquad [x^i_0, y^j_0] = 0, \qquad [y^i_0, y^j_0] = -i\theta^{ij} \quad [9]$$

The difference in the commutation relation for $x_0$ and $y_0$ implies that the two boundaries of the open string have opposite commutativity. This fact is not so important for tree-level calculations since one can always choose to put all the interactions at, for example, the $\sigma = 0$ boundary. Collecting all these zero mode parts of the vertex operators, one obtains a phase factor

$$e^{ip^1x_0}e^{ip^2x_0}\cdot e^{ip^Nx_0} = e^{i\sum p^a x_0}e^{-(i/2)\sum_{I<J}^N p^I\theta p^J} \quad [10]$$

where the external momenta $p^a$ are ordered cyclically on the circle, and momentum conservation has been used. The computation of the oscillator part of the amplitude is the same as in the $B = 0$ case, except that the metric $G$ is employed in the contractions. As a result, the effect of the $B$-field on the tree-level string amplitude is simply to multiply the amplitude at $B = 0$ by the phase factor and to replace the metric by the metric $G$. A generic term in the tree-level effective action simply becomes

$$\int d^{p+1}x\sqrt{-\det g}\,\mathrm{tr}\,\partial^{n_1}\Phi_1\cdots\partial^{n_k}\Phi_k \to \int d^{p+1}x\sqrt{-\det G}\,\mathrm{tr}\,\partial^{n_1}\Phi_1 * \cdots * \partial^{n_k}\Phi_k \quad [11]$$

Here the star product, also called the Moyal product, is defined by

$$(f * g)(x) = \exp\left(\mathrm{i}\frac{\theta^{\mu\nu}}{2}\frac{\partial}{\partial x_1^\mu}\frac{\partial}{\partial x_2^\nu}\right) f(x_1)g(x_2)|_{x_1=x_2} \quad [12]$$

The star product is associative and noncommutative, and satisfies $\overline{f * g} = \bar{g} * \bar{f}$ under complex conjugation. Also, for functions that vanish rapidly enough at infinity, there holds

$$\int f * g = \int g * f = \int fg \quad [13]$$

An interesting consequence of the nonlocality as expressed by the noncommutative geometry [7] is the existence of a dipole excitation whose extent is proportional to its momentum, $\Delta x = k\theta$. This relation is at the heart of the "IR/UV mixing phenomenon" (see below) of noncommutative field theory.

At one- (and higher-) loop level, the different noncommutativities for the opposite boundaries of the open string become essential and give rise to new effects. In this case nonplanar diagrams require one to put vertex operators at the two different boundaries $\sigma = 0, \pi$. A more complicated phase factor, which involves internal as well as external momentum, results. This leads to IR/UV mixing in the noncommutative quantum field theory. The different noncommutativity for the opposite boundaries of the open string [9] is the basic reason for the IR/UV mixing in the noncommutative quantum field theory. The commutation relations [5] are valid at all loops; therefore, one can use them to construct the higher-loop string amplitudes from first principles. The effect of the $B$-field on the string interaction can easily be implemented into the Reggeon vertex and the complete higher loop amplitudes in the presence of the $B$-field have been constructed.

#### Low-Energy Limit – The Seiberg–Witten Limit and the NCOS Limit

The full open-string system is still quite complicated. One may try to decouple the infinite number of massive string modes to obtain a low-energy field-theoretic description by taking the limit $\alpha' \to 0$. Since open strings are sensitive to $G$ and $\theta$, one should take the limit such that $G$ and $\theta$ are fixed. For the magnetic case $B_{0i} = 0$, Seiberg and Witten showed that this can be achieved with the following double scaling limit:

$$\alpha' \sim \epsilon^{1/2}, \quad g_{ij} \sim \epsilon \to 0 \quad [14]$$

with $B_{ij}$ and everything else kept fixed. Assuming $B$ is of rank $r$, then [6] becomes

$$G_{ij} = -(2\pi\alpha')^2 (Bg^{-1}B)_{ij}, \quad \theta^{ij} = (B^{-1})^{ij}, \quad \text{for } i,j = 1,\ldots,r \quad [15]$$

Otherwise $G_{ij} = g_{ij}, \theta^{ij} = 0$. One may also argue that the closed string decouples in this limit. As a result, in the low-energy limit a greatly simplified noncommutative Yang–Mills action $F * F$ is obtained (see below for more discussion of this field theory).

For the case of a constant electric field background, say $B_{01} \neq 0$, there is a critical electric field beyond which the open string becomes unstable and the theory does not make sense. Due to the presence of this upper bound of the electric field, one can show that there is no decoupling limit where one can reduce the string theory to a field theory on a noncommutative spacetime. However, one can consider a different scaling limit where one takes the closed-string metric scale to infinity appropriately as the electric field approaches the critical value. In this limit, all closed-string modes decouple. One obtains a novel noncritical string theory living on a noncommutative spacetime known as the noncommutative open string (NCOS).

### Noncommutative Quantum Field Theory

Field theories on noncommutative spacetime are defined by using the star product instead of the ordinary product of the fields. To illustrate the general ideas, let us consider a single real scalar field theory with the action

$$S = \int \mathrm{d}^D x \left[\frac{1}{2}\partial_\mu\phi * \partial^\mu\phi - \frac{m^2}{2}\phi * \phi - V(\phi)\right]$$
$$V(\phi) = \frac{g}{4!}\phi^{*4} \quad [16]$$

Due to the property [13], free noncommutative field theory is the same as an ordinary field theory. Treating the interaction term as a perturbation, one can perform the usual quantization and obtain the Feynman rules: the propagator is unchanged and the interaction vertex in the momentum space is given by $g$ times the phase factor

$$\exp\left(-\frac{\mathrm{i}}{2}\sum_{1\le a<b\le 4} p^a \times p^b\right) \quad [17]$$

Here $p \times q \equiv p_\mu\theta^{\mu\nu}q_\nu$. The theory is nonlocal due to the infinite order of derivatives that appear in the interaction.

### Planar and Nonplanar Diagrams

The factor [17] is cyclically symmetric but not permutation symmetric. This is analogous to the situation of an M-field theory. Using the same double-line notation as introduced by 't Hooft, one can similarly classify the Feynman diagrams of noncommutative field theory according to its genus. In particular, the total phase factor of a planar diagram behaves quite differently from that of a nonplanar diagram. It is easy to show that a planar diagram will have the phase factor

$$V_{\mathrm{p}}(p^1,\ldots,p^n) = \exp\left(-\frac{\mathrm{i}}{2}\sum_{1\le a<b\le n} p^a \times p^b\right) \qquad [18]$$

where $p^1,\ldots,p^n$ are the (cyclically ordered) external momenta of the graph. Note that the phase factor [18] is independent of the internal momenta. This is not the case for a nonplanar diagram. One can easily show that a nonplanar diagram carries an additional phase factor

$$V_{\mathrm{np}} = V_{\mathrm{p}} \exp\left(-\frac{\mathrm{i}}{2}\sum_{1\le a<b\le n} C_{ab} p^a \times p^b\right) \qquad [19]$$

where $C_{ab}$ is the signed intersection matrix of the graph, whose $ab$ matrix element counts the number of times the $a$th (internal or external) line crosses the $b$th line. The matrix $C_{ab}$ is not uniquely determined by the diagram as different ways of drawing the graph could lead to different intersections. However, the phase factor [19] is unique due to momentum conservation.

The different behaviors of the planar and nonplanar phase factors have important consequences.

1. Since the phase factor [18] is independent of the internal momenta, the divergences and renormalizability of the planar diagrams will be (simply) the same as in the commutative theory and can be handled with standard renormalization techniques. This is sharply different for the nonplanar diagrams. In fact, due to the extra oscillatory internal-momenta-dependent phase factor, one can expect the nonplanar diagrams to have an improved ultraviolet (UV) behavior. It turns out that planar and nonplanar diagrams also differ sharply in their infrared (IR) behavior due to the "IR/UV mixing effect" (see below).
2. Moreover, at high energies one can expect that noncommutative field theory will generically become planar since the nonplanar diagrams will be suppressed due to the oscillatory phase factor.
3. In the limit $\theta \to \infty$, the nonplanar sector will be totally suppressed since the rapidly oscillating phase factor will cause the nonplanar diagram to vanish upon integrating out the momenta. Thus, generically the large $\theta$ limit is analogous to the large-$N$ limit where only the planar diagrams contribute. However, these expectations do not apply for noncommutative gauge theory since one needs to include "open Wilson lines" (see below) in the construction of gauge invariant observables, and the open Wilson line grows in extent with energy and $\theta$.

### IR/UV Mixing

Due to the nonlocal nature of noncommutative field theory, there is generally a mixing of the UV and IR scales. The reason is roughly the following. Nonplanar diagrams generally have phase factors like $\exp(\mathrm{i}k\theta p)$ with $k$ a loop momentum, $p$ an external momentum. Consider a nonplanar diagram which is UV divergent when $\theta = 0$; one can expect that for very high loop momenta the phase factor will oscillate rapidly and render the integral finite. However, this is only valid for a nonvanishing external momentum $\theta p$; the infinity will come back as $\theta p \to 0$. However, this time it appears as an IR singularity. Thus, an IR divergence arises whose origin is from the UV region of the momentum integration and this is known as the IR/UV mixing phenomenon.

To be more specific, consider the $\phi^4$ scalar theory in $D = 4$ dimensions. The one-loop self-energy has a nonplanar contribution given by

$$\Gamma_{\mathrm{np}} = \frac{g}{6(2\pi)^4}\int \frac{\mathrm{d}^4 k}{k^2+m^2}\mathrm{e}^{\mathrm{i}k\theta p} \sim \frac{g}{3(4\pi^2)^2} \times(\Lambda_{\mathrm{eff}}^2 + \cdots) \qquad [20]$$

where $\Lambda_{\mathrm{eff}}^2 = (1/\Lambda^2 + (\theta p)^2)^{-1}$. One can see clearly the IR/UV mixing: $\Gamma_{\mathrm{np}}$ is UV finite as long as $\theta p \neq 0$; when $\theta p = 0$, the quadratic UV divergence is recovered, $\Gamma_{\mathrm{np}} \sim \Lambda^2$. For supersymmetric theory, one has at most logarithmic IR singularities from IR/UV mixing.

IR/UV mixing has a number of interesting consequences.

1. Due to the IR/UV mixing, noncommutative theory does not appear to have a consistent Wilsonian description since it requires that correlation functions computed at finite $\Lambda$ differ from their limiting values by terms of order $1/\Lambda$ for all values of momenta. However, this is not true for theory with IR/UV mixing. For example, the two-point function [20] at finite value of $\Lambda$ differs from its value at $\Lambda = \infty$ by the amount

$\Gamma_{\rm np}^{\Lambda} - \Gamma_{\rm np}^{\Lambda=\infty} \propto 1/(\theta p)^2$, for the range of momenta $(\theta p)^2 \ll 1/\Lambda^2$. It has been argued that the IR singularity may be associated with missing light degrees of freedom in the theory. With new degrees of freedom appropriately added, one may recover a conventional Wilsonian description. Moreover, it has been suggested to identify these degrees of freedom with the closed-string modes. However, the precise nature and origin of these degrees of freedom is not known.

2. The renormalization of the planar diagrams is straightforward; however, the situation is more subtle for the nonplanar diagrams since the IR/UV-mixed IR singularities may mix with other divergences at higher loops and render the proof of renormalizability much more difficult. IR/UV mixing renders certain large $N$ noncommutative field theory nonrenormalizable. However, for theories with a fixed set of degrees of freedom to start with, it is believed that one can have sufficiently good control of the IR divergences and prove renormalizability. An example of renormalizable noncommutative quantum field theory is the noncommutative Wess–Zumino model where IR/UV mixing is absent. However, a general proof is still lacking.
3. One can show that IR/UV mixing in timelike noncommutative theory ($\theta^{0i} \neq 0$) leads to breakdown of perturbative unitarity. For a theory without IR/UV mixing, unitarity will be respected even if the theory has a timelike noncommutativity. Theory with lightlike noncommutativity is unitary.

## Noncommutative Gauge Theory

Gauge theory on noncommutative space is defined by the action

$$S = -\frac{1}{4g^2}\int \mathrm{d}x \, \mathrm{tr}\big(F_{ij}(x) * F^{ij}(x)\big) \qquad [21]$$

where the gauge fields $A_i$ are $N \times N$ Hermitian matrices, $F_{ij}$ is the noncommutative field strength $F_{ij} = \partial_i A_j - \partial_j A_i - \mathrm{i}[A_i, A_j]_*$, and tr is the ordinary trace over $N \times N$ matrices. The theory is invariant under the star-gauge transformation

$$A_i \to g * A_i * g^\dagger - \mathrm{i}g * \partial_i g^\dagger \qquad [22]$$

where the $N \times N$ matrix function $g(x)$ is unitary with respect to the star product $g * g^\dagger = g^\dagger * g = I$. The solution is $g = \mathrm{e}_*^{\mathrm{i}\lambda}$, where $\lambda$ is Hermitian. In infinitesimal form, $\delta_\lambda A_i = \partial_i \lambda + \mathrm{i}[\lambda, A_i]_*$. The noncommutative gauge theory has $N^2$ Hermitian gauge fields. Because of the star product, the U(1) sector of the theory is not free and does not decouple from the SU($N$) factor as in the commutative case. Note that this way of defining noncommutative gauge theory does not work for other Lie groups since the star commutator generally involves the commutator as well as the anticommutator of the Lie algebra; hence, the expressions above generally involve the enveloping algebra of the underlying Lie group. With the help of the "Seiberg–Witten map" (see below), one can construct an enveloping-algebra-valued gauge theory which has the same number of independent gauge fields and gauge parameters as the ordinary Lie-algebra-valued gauge theory. However, the quantum properties of these theories are much less understood. One may also introduce certain automorphisms in the noncommutative U($N$) theory to restrict the dependence of the noncommutative space coordinates of the field configurations and obtain a notion of noncommutative theory with orthogonal and symplectic star-gauge group. However, the theory does not reduce to the standard gauge theory in the commutative limit $\theta \to 0$.

### Open Wilson Line and Gauge-Invariant Observables

One remarkable feature of noncommutative gauge theory is the mixing of noncommutative gauge transformations and spacetime translations, as can be seen from the following identity:

$$\mathrm{e}^{\mathrm{i}kx} * f(x) * \mathrm{e}^{-\mathrm{i}kx} = f(x + k\theta) \qquad [23]$$

for any function $f$. This is analogous to the situation in general relativity where translations are also equivalent to gauge transformations (general coordinate transformations). Thus, as in general relativity, there are no local gauge-invariant observables in noncommutative gauge theory. The unification of spacetime and gauge fields in noncommutative gauge theory can also be seen from the fact that derivatives can be realized as commutators, $\partial_i f \to -\mathrm{i}[\theta_{ij}^{-1} x^j, f]$, and get absorbed into the vector potential in the covariant derivative

$$D_i = \partial_i + \mathrm{i}A_i \to -\mathrm{i}\theta_{ij}^{-1} x^j + \mathrm{i}A_i \qquad [24]$$

Equation [24] clearly demonstrates the unification of spacetime and gauge fields. Note that the field strength takes the form $F_{ij} = \mathrm{i}[D_i, D_j] + \theta_{ij}^{-1}$.

The Wilson line operator for a path $C$ running from $x_1$ to $x_2$ is defined by

$$W(C) = P_* \exp\left(\mathrm{i}\int_C A\right) \qquad [25]$$

$P_*$ denotes the path ordering with respect to the star product, with $A(x_2)$ at the right. It transforms as

$$W(C) \to g(x_1) * W(C) * g(x_2)^\dagger \qquad [26]$$

In commutative gauge theory, the Wilson line operator for closed loop (or its Fourier transform) is gauge invariant. In noncommutative gauge theory, the closed Wilson loops are no longer gauge invariant. Noncommutative generalization of the gauge invariant Wilson loop operator can be constructed most readily by deforming the Fourier transform of the Wilson loop operator. It turns out that the closed loop has to open in a specific way to form an open Wilson line in order to be gauge invariant. To see this, let us consider a path $C$ connecting points $x$ and $x + l$. Using [23], it is easy to see that the operator

$$\tilde{W}(k) \equiv \int \mathrm{d}x \,\mathrm{tr}\, W(C) * \mathrm{e}^{ikx}, \quad \text{with } l^j = k_i \theta^{ij} \qquad [27]$$

is gauge invariant. Just like Wilson loops in ordinary gauge theory, these operators also constitute an overcomplete set of gauge-invariant operators parametrized by the set of curves $C$. When $\theta = 0$, $C$ becomes a closed loop and we reobtain the (Fourier transformed) usual closed Wilson loop in commutative gauge theory. Noncommutative version of the loop equation for closed Wilson loop has been constructed and involves open Wilson line. The open Wilson line is instrumental in the construction of gauge-invariant observables. An important application is in the construction of various couplings of the noncommutative D-brane to the bulk supergravity fields. The equivalence of the commutative and noncommutative couplings to the RR fields leads to the exact expression for the Seiberg–Witten map. It is remarkable that the one-loop nonplanar effective action for noncommutative scalar theory, gauge theory, as well as the two-loop effective action for scalar can be written compactly in terms of open Wilson line. Based on this result, the physical origin of the IR/UV mixing has been elucidated. One may identify the open Wilson line with the dipole excitation generically presents in noncommutative field theory and hence explain the presence of the IR/UV mixing. IR/UV mixing may also be identified with the instability associated with the closed-string exchange of the noncommutative D-branes.

### The Seiberg–Witten Map

The open string is coupled to the 1-form $A_i$ living on the D-brane through the coupling $\int_{\partial\Sigma} A$. For slowly varying fields, the effective action for this gauge potential can be determined from the $S$-matrix and is given by the Dirac–Born–Infeld (DBI) action. In the presence of a $B$-field, the discussion above (see eqn [11]) leads to the noncommutative DBI Lagrangian

$$L_{\mathrm{NCDBI}}(\hat{F}) = G_s^{-1} \mu_p \sqrt{-\det(G + 2\pi\alpha' \hat{F})} \qquad [28]$$

where $\mu_p = (2\pi)^{-p} (\alpha')^{-(p+1)/2}$ is the D-brane tension and $\hat{F}$ is the noncommutative field strength. However, one may also exploit the tensor gauge invariance on the D-brane (i.e., the string sigma model is invariant under $A \to A - \Lambda, B \to B + \mathrm{d}\Lambda$) and consider the combination $F + B$ as a whole. In this case, it is like having the open string coupled to the boundary gauge field strength $F + B$ and there is no $B$ field. One has the usual DBI Lagrangian

$$L_{\mathrm{DBI}}(F) = g_s^{-1} \mu_p \sqrt{-\det(G + 2\pi\alpha'(F + B))} \qquad [29]$$

In [28] and [29], $G_s$ and $g_s$ are the effective open-string couplings in the noncommutative and commutative descriptions. Although they look quite different, Seiberg and Witten showed that the commutative and noncommutative DBI actions are indeed equivalent if the open-string couplings are related by $g_s = G_s \sqrt{\det(g + 2\pi\alpha' B)/\det G}$ and there is a field redefinition that relates the commutative and noncommutative gauge fields. The map $\hat{A} = \hat{A}(A)$ is called the Seiberg–Witten map. Moreover, the noncommutative gauge symmetry is equivalent to the ordinary gauge symmetry in the sense that they have the same set of orbits under gauge transformation:

$$\hat{A}(A) + \hat{\delta}_{\hat{\lambda}} \hat{A}(A) = \hat{A}(A + \delta_\lambda A) \qquad [30]$$

Here $\hat{A}_i$ and $\hat{\lambda}$ are, respectively, the noncommutative gauge field and noncommutative gauge transformation parameter, and $A_i$ and $\lambda$ are, respectively, the ordinary gauge field and ordinary transformation parameter. The map between $\hat{A}_i$ and $A_i$ is called the Seiberg–Witten map. Equation [30] can be solved only if the transformation parameter $\hat{\lambda} = \hat{\lambda}(\lambda, A)$ is field dependent. The Seiberg–Witten map is characterized by the Seiberg–Witten differential equation

$$\delta \hat{A}_i(\theta) = \tfrac{1}{4} \delta\theta^{kl} \Big[ \hat{A}_k * (\partial_l \hat{A}_i + \hat{F}_{li}) + (\partial_l \hat{A}_i + \hat{F}_{li}) * \hat{A}_k \Big] \qquad [31]$$

An exact solution for the Seiberg–Witten map can be written down with the help of the open Wilson

line. For the case of U(1) with constant $F$, we have the exact solution $\hat{F} = (1 + F\theta)^{-1}F$.

That there is a field redefinition that allows one to write the effective action in terms of different fields with different gauge symmetries may seem puzzling at first sight. However, it has a clear physical origin in terms of the string world sheet. In fact, there are different possible schemes to regularize the short-distance divergence on the world sheet. One can show that the Pauli–Villars regularization gives the commutative description, while the point-splitting regularization gives the noncommutative description. Since theories defined by different regularization schemes are related by a coupling-constant redefinition, this implies that the commutative and noncommutative descriptions are related by a field redefinition, because the couplings on the world sheet are just the spacetime fields.

Despite this formal equivalence, the physics of the noncommutative theories is generally quite different from the commutative case. First, it is clear that generally the Seiberg–Witten map may take nonsingular configurations to singular configurations. Second, the observables one is interested in are also generally different. Moreover, the two descriptions are generally good for different regimes: the conventional gauge theory description is simpler for small $B$ and the noncommutative description is simpler for large $B$.

### Perturbative Gauge Theory Dynamics

The noncommutative gauge symmetry [22] can be fixed as usual by employing the Faddeev–Popov procedure, resulting in Feynman rules that are similar to the conventional gauge theory. The important difference is that now the structure constants in the phase factors [18] and [19] should be amended. It turns out that the nonplanar U($N$) diagrams contribute (only) to the U(1) part of the theory. As a result, unlike the commutative case, the U(1) part of the theory is no longer decoupled and free. Noncommutative gauge theory is one-loop renormalizable. The $\beta$-function is determined solely by the planar diagrams and, at one loop, is given by

$$\beta(g) = -\frac{22}{3}\frac{Ng^3}{16\pi^2} \quad \text{for } N \geq 1 \qquad [32]$$

Note that the $\beta$-function is independent of $\theta$; the noncommutative U(1) is asymptotically free and does not reduce to the commutative theory when $\theta \to 0$. Noncommutative theory beyond the tree level is generally not smooth in the limit $\theta \to 0$. Discontinuity of this kind was also noted for the Chern–Simon system.

Gauge anomalies can be similarly discussed and satisfy the noncommutative generalizations of the Wess–Zumino consistency conditions. In $d = 2n$ dimensions, the anomaly involves the combination $\mathrm{tr}(T^{a_1}T^{a_2}\cdots T^{a_{n+1}})$ rather than the usual symmetrized trace, since the phase factor is not permutation symmetric. As a result, the usual cancellation of the anomaly does not work and is the main obstacle to the construction of noncommutative chiral gauge theory.

There are a number of interesting features to mention for the IR/UV mixing in noncommutative gauge theory.

1. IR/UV mixing generically yields pole-like IR singularities. Despite the appearance of IR poles, gauge invariance of the theory is not endangered.
2. One can show that only the U(1) sector is affected by IR/UV mixing.
3. As a result of IR/UV mixing, noncommutative U(1) photons polarized in the noncommutative plane will have different dispersion relations from those which are not. Strange as it is, this is consistent with gauge invariance.

## Noncommutative Solitons, Instantons and D-Branes

Solitons and instantons play important roles in the nonperturbative aspects of field theory. The nonlocality of the star product gives noncommutative field theory a stringy nature. It is remarkable that this applies to the nonperturbative sector as well. Solitons and instantons in the noncommutative gauge theory amazingly reproduce the properties of D-branes in the string.

### GMS Solitons

Derrick's theorem says that commutative scalar field theories in two or higher dimensions do not admit any finite-energy classical solution. This follows from a simple scaling argument, which will fail when the theory becomes noncommutative since noncommutativity introduces a fixed length scale $\sqrt{\theta}$. Noncommutative solitons in pure scalar theory can be easily constructed in the limit $\theta = \infty$. For example, consider a $(2+1)$-dimensional single scalar theory with a potential $V$ and noncommutativity $\theta^{12} = \theta$. In the limit $\theta = \infty$, the potential term dominates and the noncommutative solitons are determined by the equation

$$\partial V/\partial \phi = 0 \qquad [33]$$

Equation [33] can be easily solved in terms of projectors. Assuming $V$ has no linear term, the general soliton (up to unitary equivalence) is

$$\phi = \sum \lambda_i P_i \qquad [34]$$

where $\lambda_i$ are the roots of $V'(\lambda) = 0$ and $P_i$ is a set of orthogonal projectors. For real scalar field theory, the sum is restricted to real roots only. These solutions are known as the Gopakumar–Minwalla–Strominger (GMS) solitons. A simple example of a projector is given by $P = |0\rangle\langle 0|$, which corresponds to a Gaussian profile in the $x^1, x^2$ plane with width $\sqrt{\theta}$. The soliton continues to exist until $\theta$ decreases below a certain critical $\theta_c$.

New solutions can be generated from known ones using the so-called solution-generating technique. If $\phi$ is a solution of [33], then

$$\phi' = T^\dagger \phi T \qquad [35]$$

is also a solution provided that $TT^\dagger = 1$. In an infinite-dimensional Hilbert space, $T$ is not necessarily unitary, that is, $T^\dagger T \neq 1$. In this case, $T$ is said to be a partial isometry. The new solution $\phi'$ is different from $\phi$ since they are not related by a global transformation of basis.

### Tachyon Condensation and D-Branes

A beautiful application of the noncommutative soliton is in the construction of D-branes as solitons of the tachyon field in noncommutative open-string theory. For the bosonic string theory, one may consider it to be a space-filling D25 brane. Integrating out the massive-string modes leads to an effective action for the tachyon and the massless gauge field $A_\mu$. It should be remarked that, contrary to the pure scalar case, noncommutative solitons can be constructed exactly for finite $\theta$ in a system with gauge and scalar fields. Although the detailed form of the effective action is unknown, one has enough confidence to say what the true vacuum configuration is according to the Sen conjecture. One can then apply the solution-generating technique to generate new soliton solutions. In this manner, with a $B$-field of rank $2k$, one can construct solutions which are localized in $\mathbb{R}^{2k}$ and represent a D$(25-2k)$ brane. This is supported by the matching of the tension and the spectrum of fluctuations around the soliton configuration. Similar ideas can also be applied to construct D-branes in type II string theory. Again the starting point is an unstable brane configuration with tachyon field(s). There are two types of unstable D-branes: non-BPS D$p$ branes ($p$ odd in IIA theory and $p$ even in IIB theory) and BPS branes–antibranes D$p$–$\overline{\text{D}p}$ systems. A similar analysis allows one to identify the noncommutative soliton with the lower-dimensional BPS D-branes which arises from tachyon condensation.

One main motivation for studying tachyon condensation in open-string theory is the hope that open-string theory may provide a fundamental nonperturbative formulation of string theory. It may not be too surprising that D-branes can be obtained in terms of open-string fields. However, to describe closed strings and NS branes in terms of open-string degrees of freedom remains an obstacle.

### Noncommutative Instanton and Monopoles

Instantons on noncommutative $\mathbb{R}^4_\theta$ can be readily constructed using the Atiyah–Drinfeld–Hitchin–Manin (ADHM) formalism by modifying the ADHM constraints with a constant additive term. The result is that the self-dual (resp. anti-self-dual) instanton moduli space depends only on the anti-self-dual (resp. self-dual) part. The construction goes through even in the U(1) case. Consider a self-dual $\theta$; the ADHM constraints for the self-dual instanton are the same as in the commutative case, and there is no nonsingular solution. On the other hand, the ADHM constraints for the anti-self-dual instanton get modified and admit nontrivial solutions. This noncommutative instanton solution is nonsingular with size $\sqrt{\theta}$. The noncommutative instanton represents a D($p$–4) brane within a D$p$ brane. The ADHM constraints are just the D-flatness condition for the D-brane world-volume gauge theory. The additive constant to the ADHM constraints also has a simple interpretation as a Fayet–Iliopolous parameter which appears in the presence of a $B$-field. Although the ADHM method does not give a self-dual instanton, a direct construction can be applied to obtain non-ADHM self-dual instantons. Recall that the gauge field strength can be written as $F_{ij} = \mathrm{i}[D_i, D_j] + \theta^{-1}_{ij}$, where $D_i$ is given by the function on the right-hand side of [24]. Thus, a simple self-dual solution can be constructed with

$$D_i = \mathrm{i}\theta^{-1}_{ij} T^\dagger x^j T \qquad [36]$$

where $T$ is a partial isometry which satisfies $TT^\dagger = 1$, but $T^\dagger T = 1 - P$ is not necessarily the identity. It is clear that $P$ is a projector. The field strength

$$F_{ij} = \theta^{-1}_{ij} P \qquad [37]$$

is self-dual and has instanton number $n$ where $n$ is the rank of the projector.

On noncommutative $\mathbb{R}^3$ (say $\theta^{12} = \theta$), BPS monopoles satisfy the Bogomolny equation:

$$\nabla_i \Phi = \pm B_i, \quad i = 1, 2, 3 \qquad [38]$$

and can be obtained by solving the Nahm equation

$$\partial_z T_i = \epsilon_{ijk} T_j T_k + \delta_{i3}\theta \qquad [39]$$

$T_i$ are $k \times k$ Hermitian matrices depending on an auxiliary variable $z$ and $k$ gives the charge of the monopole. Noncommutativity modifies the Nahm equation with a constant term, which can be absorbed by a constant shift of the generators. Therefore, unlike the case of instanton, the monopole moduli space is not modified by noncommutative deformation. The Nahm construction has a clear physical meaning in string theory. The monopole (electric charge) can be interpreted as a D-string (fundamental string) ending on a D3 brane. One can also suspend $k$ D-string between a collection of $N$ parallel D3 braness; this would correspond to a charge $k$ monopole in a Higgsed U($N$) gauge theory. The matrices $X^i$ correspond to the matrix transverse coordinates of the D-strings which lie within the D3 branes.

## Further Topics

Finally, in the following some further topics of interest are discussed briefly.

1. The noncommutative geometry discussed here is of canonical type. Other deformations exist, for example, kappa-deformation and fuzzy sphere which are of the Lie-algebra type, and quantum group deformation which is a quadratic-type deformation: $x^i x^j = q^{-1} \hat{R}^{ij}_{kl} x^k x^l$, whose consistency is guaranteed by the Yang–Baxter equation. It is interesting to see whether these noncommutative geometries arise from string theory. Another natural generalization is to consider noncommutative geometry of superspace. A simple example is to consider the fermionic coordinates to be deformed with the nonvanishing relation

$$\{\theta^\alpha, \theta^\beta\} = C^{\alpha\beta} \qquad [40]$$

   where $C^{\alpha\beta}$ are constants. It has been shown that [40] arises in certain Calabi–Yau compactification of type IIB string theory in the presence of RR background. The deformation [40] reduces the number of supersymmetries by half. Therefore, it is called $\mathcal{N} = 1/2$ supersymmetry. The noncommutativity [40] can be implemented on the superspace $(y^i, \theta^\alpha, \bar{\theta}^{\dot{\alpha}})$ as a star product for the $\theta^\alpha$'s. Unlike the bosonic deformation which involves an infinite number of higher derivatives, the star product for [40] stops at order $C^2$ due to the Grassmannian nature of the fermionic coordinates. Field theory with $\mathcal{N} = 1/2$ supersymmetry is local and differs from the ordinary $\mathcal{N} = 1$ theory by only a small number of supersymmetry breaking terms. The $N = 1/2$ Wess–Zumino model is renormalizable if extra $F$ and $F^3$ terms are added to the original Lagrangian, where $F$ is the auxiliary field. The $N = 1/2$ gauge theory is also renormalizable.
2. Integrability of a theory provides valuable information beyond the perturbative level. An integrable field theory is characterized by an infinite number of conserved charges in involution. It is natural to ask whether integrability is preserved by noncommutative deformation. Noncommutative integrable field theories have been constructed. In the commutative case, Ward has conjectured that all $(1+1)$- and $(2+1)$-dimensional integrable systems can be obtained from the four-dimensional self-dual Yang–Mills equation by reduction. Validity of the noncommutative version of the Ward conjecture has been confirmed so far. It will be interesting to see whether it is true in general.
3. Locality and Lorentz symmetry form the cornerstones of quantum field theory and standard model physics of particles. Noncommutative field theory provides a theoretical framework where one can discuss effects of nonlocality and Lorentz symmetry violation. Possible phenomenological signals have been investigated (mostly at the tree level) and a bound has been placed on the extent of noncommutativity. A proper understanding and better control of the IR/UV mixing remains the crux of the problem. Noncommutative geometry may also be relevant for cosmology and inflation.
4. Like the standard AdS/CFT correspondence, the noncommutative gauge theory should also have a gravity-dual description. The supergravity background can be determined by considering the decoupling limit of D-branes with an NS $B$-field background. However, since the noncommutative gauge theory does not permit any conventional local gauge-invariant observable, the usual AdS/CFT correspondence that relates field theory correlators with bulk interaction does not seem to apply. It has been argued that generic properties such as the relation between length and momentum for open Wilson lines

can be seen from the gravity side. A more precise understanding of the duality map is called for.

*See also:* Brane Construction of Gauge Theories; Deformation Quantization; Gauge Theories from Strings; Noncommutative Tori, Yang–Mills, and String Theory; Positive Maps on $C^*$-Algebras; Solitons and Other Extended Field Configurations; String Field Theory; Superstring Theories.

## Further Reading


Chu CS and Ho PM (1999) Noncommutative open string and D-brane. *Nuclear Physics B* 550: 151.

Connes A (1994) *Noncommutative Geometry*. New York: Academic Press Inc.

Connes A, Douglas MR, and Schwarz A (1998) Noncommutative geometry and matrix theory: compactification on tori. *JHEP* 9802: 003.

Douglas MR and Hull CM (1998) D-branes and the noncommutative torus. *JHEP* 9802: 008.

Douglas MR and Nekrasov NA (2001) Noncommutative field theory. *Reviews of Modern Physics* 73: 977.

Harvey JA (2001) Komaba lectures on noncommutative solitons and D-branes. arXiv:hep-th/0102076.

Konechny A and Schwarz A (2002) Introduction to M(atrix) theory and noncommutative geometry. *Physics Reports* 360: 353.

Nekrasov NA (2000) Trieste lectures on solitons in noncommutative gauge theories. arXiv:hep-th/0011095.

Polchinski J (1998) *String Theory*. Cambridge: Cambridge University Press.

Schomerus V (1999) D-branes and deformation quantization. *JHEP* 9906: 030.

Seiberg N and Witten E (1999) String theory and noncommutative geometry. *JHEP* 9909: 032.

Sen A (2004) Tachyon dynamics in open string theory. arXiv:hepth/0410103.

Szabo RJ (2003) Quantum field theory on noncommutative spaces. *Physics Reports* 378: 207.

Taylor W (2000) M(atrix) theory: matrix quantum mechanics as a fundamental theory. *Reviews of Modern Physics* 73: 419.


# Noncommutative Tori, Yang–Mills, and String Theory

**A Konechny**, Rutgers, The State University of New Jersey, Piscataway, NJ, USA



## Introduction

Noncommutative tori are historically among the oldest and by now the most developed examples of noncommutative spaces. Noncommutative Yang–Mills theory can be obtained from string theory. This connection led to a cross-fertilization of research in physics and mathematics on Yang–Mills theory on noncommutative tori. One important result stemming from that work is the link between T-duality in string theory and Morita equivalence of associative algebras. In this article, we give an overview of the basic results in the differential geometry of noncommutative tori. Yang–Mills theory on noncommutative tori, the duality induced by Morita equivalence and its link with T-duality are discussed. The noncommutative Nahm transform for instantons is introduced.

## Noncommutative Tori

### The Algebra of Functions

The basic notions of noncommutative differential geometry were introduced and illustrated on the example of a two-dimensional noncommutative torus by Connes (1980). To define an algebra of functions on a $d$-dimensional noncommutative torus, consider a set of linear generators $U_{\boldsymbol{n}}$ labeled by $\boldsymbol{n} \in \mathbb{Z}^d$ – a $d$-dimensional vector with integral entries. The multiplication is defined by the formula

$$U_{\boldsymbol{n}} U_{\boldsymbol{m}} = e^{\pi i n_j \theta^{jk} m_k} U_{\boldsymbol{n}+\boldsymbol{m}} \qquad [1]$$

where $\theta^{jk}$ is an antisymmetric $d \times d$ matrix, and summation over repeated indices is assumed. We further extend the multiplication from finite linear combinations to formal infinite series $\sum_{\boldsymbol{n}} C(\boldsymbol{n}) U_{\boldsymbol{n}}$ where the coefficients $C(\boldsymbol{n})$ tend to zero faster than any power of $\|\boldsymbol{n}\|$. The resulting algebra constitutes the algebra of smooth functions on a noncommutative torus and will be denoted by $T_\theta^d$. Sometimes for brevity we will omit the dimension label $d$ in the notation of the algebra. We introduce an involution $*$ in $T_\theta^d$ by the rule $U_{\boldsymbol{n}}^* = U_{-\boldsymbol{n}}$. The elements $U_{\boldsymbol{n}}$ are assumed to be unitary with respect to this involution, that is, $U_{\boldsymbol{n}}^* U_{\boldsymbol{n}} = U_{-\boldsymbol{n}} U_{\boldsymbol{n}} = 1 \equiv U_0$. One can further introduce a norm and take an appropriate completion of the involutive algebra $T_\theta^d$ to obtain the $C^*$-algebra of functions on a noncommutative torus. For our purposes, the norm structure will not be important. A canonically normalized trace on $T_\theta^d$ is introduced by specifying

$$\mathrm{tr}\, U_{\boldsymbol{n}} = \delta_{\boldsymbol{n},0} \qquad [2]$$

### Projective Modules

According to the general approach to noncommutative geometry, finitely generated projective modules

over the algebra of functions are natural analogs of vector bundles. Throughout this article, when speaking of a projective module, we will assume a finitely generated left projective module.

A free module $(T^d_\theta)^N$ is equipped with a $T^d_\theta$-valued Hermitian inner product $\langle .,. \rangle_{T_\theta}$ defined by the formula

$$\langle (a_1,\ldots,a_N),(b_1,\ldots,b_N)\rangle_{T_\theta} = \sum_{i=1}^{N} a_i^* b_i \qquad [3]$$

A projective module $E$ is by definition a direct summand of a free module. Thus, it inherits the inner product $\langle .,. \rangle_{T_\theta}$. Consider the endomorphisms of the module $E$, that is, linear mappings $E \to E$ commuting with the action of $T^d_\theta$. These endomorphisms form an associative unital algebra denoted $\mathrm{End}_{T_\theta}E$. A decomposition $(T^d_\theta)^N = E \oplus E'$ determines an endomorphism $P:(T^d_\theta)^N \to (T^d_\theta)^N$ that projects $(T^d_\theta)^N$ onto $E$. The algebra $\mathrm{End}_{T_\theta}E$ can then be identified with a subalgebra of $\mathrm{Mat}_N(T^d_{-\theta})$ – the endomorphisms of the free module $(T^d_\theta)^N$. The latter has a canonical trace that is the composition of the matrix trace with the trace specified in [2]. By restriction, it gives rise to a canonical trace tr on $\mathrm{End}_{T_\theta}E$. The same embedding also provides a canonical involution on $\mathrm{End}_{T_\theta}E$ by a composition of the matrix transposition and the involution $*$ on $T^d_\theta$.

A large class of examples of projective modules over noncommutative tori are furnished by the so-called Heisenberg modules. They are constructed as follows. Let $G$ be the direct sum of $\mathbb{R}^p$ and an abelian finitely generated group, and let $G^*$ be its dual group. In the most general situation $G = \mathbb{R}^p \times \mathbb{Z}^q \times F$ where $F$ is a finite group. Then $G^* \cong \mathbb{R}^p \times T^q \times F^*$.

Consider the linear space $\mathcal{S}(G)$ of functions on $G$ decreasing at infinity faster than any power. We define operators $U_{(\gamma,\tilde\gamma)}:\mathcal{S}(G)\to\mathcal{S}(G)$ labeled by a pair $(\gamma,\tilde\gamma)\in G\times G^*$ acting as follows:

$$(U_{(\gamma,\tilde\gamma)}f)(x) = \tilde\gamma(x)f(x+\gamma) \qquad [4]$$

One can check that the operators $U_{(\gamma,\tilde\gamma)}$ satisfy the commutation relations

$$U_{(\gamma,\tilde\gamma)}U_{(\mu,\tilde\mu)} = \tilde\mu(\gamma)\tilde\gamma^{-1}(\mu)U_{(\mu,\tilde\mu)}U_{(\gamma,\tilde\gamma)} \qquad [5]$$

If $(\gamma,\tilde\gamma)$ run over a $d$-dimensional discrete subgroup $\Gamma \subset G\times G^*, \Gamma \cong \mathbb{Z}^d$, then formula [4] defines a module over a $d$-dimensional noncommutative torus $T^d_\theta$ with

$$\exp(2\pi i\theta_{ij}) = \tilde\gamma_i(\gamma_j)\tilde\gamma_j^{-1}(\gamma_i) \qquad [6]$$

for a given basis $(\gamma_i,\tilde\gamma_i)$ of the lattice $\Gamma$. This module is projective if $\Gamma$ is such that $G\times G^*/\Gamma$ is compact. If that is the case, then the projective $T^d_\theta$-module at hand is called a Heisenberg module and denoted by $E_\Gamma$.

Heisenberg modules play a special role. If the matrix $\theta_{ij}$ is irrational in the sense that at least one of its entries is irrational, then any projective module over $T^d_\theta$ can be represented as a direct sum of Heisenberg modules. In that sense, Heisenberg modules can be used as building blocks to construct an arbitrary module.

### Connections

Next we would like to define connections on a projective module over $T^d_\theta$. To this end, let us first define a Lie algebra of shifts $L_\theta$ acting on $T^d_\theta$ by specifying a basis consisting of derivations $\delta_j: T^d_\theta \to T^d_\theta, j=1,\ldots,d$ satisfying

$$\delta_j(U_{\boldsymbol n}) = 2\pi i n_j U_{\boldsymbol n} \qquad [7]$$

These derivations span a $d$-dimensional abelian Lie algebra that we denote by $L_\theta$.

A connection on a module $E$ over $T^d_\theta$ is a set of operators $\nabla_X: E\to E, X\in L_\theta$, depending linearly on $X$ and satisfying

$$[\nabla_X, U_{\boldsymbol n}] = \delta_X(U_{\boldsymbol n}) \qquad [8]$$

where $U_{\boldsymbol n}$ are operators $E\to E$ representing the corresponding generators of $T^d_\theta$. In the standard basis [7], this relation reads as

$$[\nabla_j, U_{\boldsymbol n}] = 2\pi i n_j U_{\boldsymbol n} \qquad [9]$$

The curvature of the connection $\nabla_X$ defined as the commutator $F_{XY} = [\nabla_X,\nabla_Y]$ is an exterior 2-form on the adjoint vector space $L^*_\theta$ with values in $\mathrm{End}_{T^d_\theta}E$.

### *K*-Theory: Chern Character

The $K$-groups of a noncommutative torus coincide with those for commutative tori:

$$K_0(T^d_\theta) \cong \mathbb{Z}^{2^{d-1}} \cong K_1\left(T^d_\theta\right)$$

The Chern character of a projective module $E$ over a noncommutative torus $T^d_\theta$ can be defined as

$$\mathrm{ch}(E) = \mathrm{tr}\,\exp\left(\frac{F}{2\pi i}\right) \in \Lambda^{\mathrm{even}}\left(L^*_\theta\right) \qquad [10]$$

where $F$ is the curvature form of a connection on $E, \Lambda^{\mathrm{even}}(L^*_\theta)$ is the even part of the exterior algebra of $L^*_\theta$ and tr is the canonical trace on $\mathrm{End}_{T^d_\theta}E$. This

mapping gives rise to a noncommutative Chern character

$$\mathrm{ch} : K_0\left(T_\theta^d\right) \to \Lambda^{\mathrm{even}}\left(L_\theta^*\right) \qquad [11]$$

The component $\mathrm{ch}_0(E) = \mathrm{tr}\,\mathbf{1} \equiv \dim(E)$ is called the dimension of the module $E$.

A distinctive feature of the noncommutative Chern character [11] is that its image does not consist of integral elements, that is, there is no lattice in $L_\theta^*$ that generates the image of the Chern character. However, there is a different integrality statement that replaces the commutative one. Consider a basis in $L_\theta^*$ in which the derivations corresponding to basis elements satisfy [7]. Denote the exterior forms corresponding to the basis elements by $\alpha^1, \ldots, \alpha^d$. Then an arbitrary element of $\Lambda(L_\theta^*)$ can be represented as a polynomial in the anticommuting variables $\alpha^i$. Next let us consider the subset $\Lambda^{\mathrm{even}}(\mathbb{Z}^d) \subset \Lambda^{\mathrm{even}}(L_\theta^*)$ that consists of polynomials in $\alpha^j$ having integer coefficients. It was proved by Elliott that the Chern character is injective and its range on $K_0(T_\theta^d)$ is given by the image of $\Lambda^{\mathrm{even}}(\mathbb{Z}^d)$ under the action of the operator

$$\exp\left(-\frac{1}{2}\frac{\partial}{\partial\alpha^j}\theta^{jk}\frac{\partial}{\partial\alpha^k}\right)$$

This fact implies that the $K$-group $K_0(T_\theta^d)$ can be identified with the additive group $\Lambda^{\mathrm{even}}(\mathbb{Z}^d)$.

The $K$-theory class $\mu(E) \in \Lambda^{\mathrm{even}}(\mathbb{Z}^d)$ of a module $E$ can be computed from its Chern character by the formula

$$\mu(E) = \exp\left(\frac{1}{2}\frac{\partial}{\partial\alpha^j}\theta^{jk}\frac{\partial}{\partial\alpha^k}\right)\mathrm{ch}(E) \qquad [12]$$

Note that the anticommuting variables $\alpha^i$ and the derivatives $\partial/\partial\alpha^j$ satisfy the anticommutation relation $\{\alpha^i, \partial/\partial\alpha^j\} = \delta^i_j$.

The coefficients of $\mu(E)$ standing in front of monomials in $\alpha^i$ are integers to which we will refer as the topological numbers of the module $E$. These numbers can also be interpreted as numbers of D-branes of a definite kind although in noncommutative geometry it is difficult to talk about branes as geometrical objects wrapped on torus cycles.

One can show that for noncommutative tori $T_\theta^d$ with irrational matrix $\theta_{ij}$ the set of elements of $K_0(T_\theta^d)$ that represent a projective module (i.e., the positive cone) consist exactly of the elements of positive dimension. Moreover, if $\theta_{ij}$ is irrational, any two projective modules which represent the same element of $K_0(T_\theta^d)$ are isomorphic; that is, the projective modules are essentially specified in this case by their topological numbers.

The complex differential geometry of noncommutative tori and its relation with mirror symmetry is discussed in Polishchuk and Schwarz (2003).

## Yang–Mills Theory on Noncommutative Tori

Let $E$ be a projective module over $T_\theta^d$. We call a Yang–Mills field on $E$ a connection $\nabla_X$-compatible with the Hermitian structure, that is, a connection satisfying

$$\langle\nabla_X\xi, \eta\rangle_{T_\theta} + \langle\xi, \nabla_X\eta\rangle_{T_\theta} = \delta_X(\langle\xi, \eta\rangle_{T_\theta}) \qquad [13]$$

for any two elements $\xi, \eta \in E$. Given a positive-definite metric on the Lie algebra $L_\theta$, we can define a Yang–Mills functional

$$S_{\mathrm{YM}}(\nabla_i) = \frac{V}{4g^2_{\mathrm{YM}}} g^{ik}g^{jl}\mathrm{tr}(F_{ij}F_{kl}) \qquad [14]$$

Here $g^{ij}$ stands for the metric tensor in the canonical basis [7], $V = \sqrt{|\det g|}, g_{\mathrm{YM}}$ is the Yang–Mills coupling constant, tr stands for the canonical trace on $\mathrm{End}_{T_\theta}E$ discussed above, and summation over repeated indices is assumed. Compatibility with the Hermitian structure [13] can be shown to imply the positive definiteness of the functional $S_{\mathrm{YM}}$. The extrema of this functional are given by the solutions to the Yang–Mills equations

$$g^{ki}[\nabla_k, F_{ij}] = 0 \qquad [15]$$

A gauge transformation in the noncommutative Yang–Mills theory is specified by a unitary endomorphism $Z \in \mathrm{End}_{T_\theta}E$, that is, an endomorphism satisfying $ZZ^* = Z^*Z = 1$. The corresponding gauge transformation acts on a Yang–Mills field as

$$\nabla_j \mapsto Z\nabla_j Z^* \qquad [16]$$

The Yang–Mills functional [14] and the Yang–Mills equations [15] are invariant under these transformations.

It is easy to see that Yang–Mills fields whose curvature is a scalar operator, that is, $[\nabla_i, \nabla_j] = \sigma_{ij}\cdot\mathbf{1}$ with $\sigma_{ij}$ a real-number-valued tensor, solve the Yang–Mills equations [15]. A characterization of modules admitting a constant curvature connection and a description of the moduli spaces of constant curvature connections (i.e., the space of such connections modulo gauge transformations) is reviewed in Konechny and Schwarz (2002). Another interesting class of solutions to the Yang–Mills equations is instantons (see below).

As in the ordinary field theory, one can construct various extensions of the noncommutative Yang–Mills theory [14] by adding other fields. To obtain a

supersymmetric extension of [14], one needs to add a number of endomorphisms $X_I \in \mathrm{End}_{T_\theta} E$ that play the role of bosonic scalar fields in the adjoint representation of the gauge group and a number of odd Grassmann parity endomorphisms $\psi_i^\alpha \in \Pi\mathrm{End}_{T_\theta} E$ endowed with an SO($d$)-spinor index $\alpha$. The latter ones are analogs of the usual fermionic fields.

In string theory, one considers a maximally supersymmetric extension of the Yang–Mills theory [14]. In this case, the supersymmetric action depends on $10 - d$ bosonic scalars $X_I, I = d, \ldots, 9$, and the fermionic fields can be collected into an SO(9, 1) Majorana–Weyl spinor multiplet $\psi^\alpha, \alpha = 1, \ldots, 16$. The maximally supersymmetric Yang–Mills action takes the form

$$S_{\mathrm{SYM}} = \frac{V}{4g^2}\mathrm{tr}\Big(F_{\mu\nu}F^{\mu\nu} + [\nabla_\mu, X_I][\nabla^\mu, X^I] + [X_I, X_J][X^I, X^J] - 2\psi^\alpha \sigma^\mu_{\alpha\beta}[\nabla_\mu, \psi^\beta] - 2\psi^\alpha \sigma^I_{\alpha\beta}[X_I, \psi^\beta]\Big) \quad [17]$$

Here the curvature indices $F_{\mu\nu}, \mu, \nu = 0, \ldots, d-1$, are assumed to be contracted with a Minkowski signature metric, and $\sigma^A_{\alpha\beta}$ are blocks of the ten-dimensional $32 \times 32$ gamma-matrices

$$\Gamma_A = \begin{pmatrix} 0 & \sigma_A^{\alpha\beta} \\ (\sigma_A)_{\alpha\beta} & 0 \end{pmatrix}, \quad A = 0, \ldots, 9$$

This action is invariant under two kinds of supersymmetry transformations denoted by $\delta_\epsilon, \tilde{\delta}_\epsilon$ and defined as

$$\begin{aligned} &\delta_\epsilon \psi = \tfrac{1}{2}(\sigma^{jk}F_{jk}\epsilon + \sigma^{jI}[\nabla_j, X_I]\epsilon + \sigma^{IJ}[X_I, X_J]\epsilon) \\ &\delta_\epsilon \nabla_j = \epsilon\sigma_j\psi, \quad \delta_\epsilon X_J = \epsilon\sigma_J\psi \\ &\tilde{\delta}_\epsilon\psi = \epsilon, \quad \tilde{\delta}_\epsilon\nabla_j = 0, \quad \tilde{\delta}_\epsilon X_J = 0 \end{aligned} \quad [18]$$

where $\epsilon$ is a constant 16-component Majorana–Weyl spinor. Of particular interest for string theory applications are solutions to the equations of motion corresponding to [17] that are invariant under some of the above supersymmetry transformations. Further discussion can be found in Konechny and Schwarz (2002).

## Morita Equivalence

The role of Morita equivalence as a duality transformation in noncommutative Yang–Mills theory was elucidated by Schwarz (1998). We will adopt a definition of Morita equivalence for noncommutative tori which can be shown to be essentially equivalent to the standard definition of strong Morita equivalence. We will say that two noncommutative tori $T^d_\theta$ and $T^d_{\hat{\theta}}$ are Morita equivalent if there exists a $(T^d_\theta, T^d_{\hat{\theta}})$-bimodule $Q$ and a $(T^d_{\hat{\theta}}, T^d_\theta)$-bimodule $P$ such that

$$Q \otimes_{T_{\hat{\theta}}} P \cong T_\theta, \qquad P \otimes_{T_\theta} Q \cong T_{\hat{\theta}} \quad [19]$$

where $T_\theta$ on the right-hand side is considered as a $(T_\theta, T_\theta)$-bimodule and analogously for $T_{\hat{\theta}}$. (It is assumed that the isomorphisms are canonical.) Given a $T_\theta$-module $E$ one obtains a $T_{\hat{\theta}}$-module $\hat{E}$ as

$$\hat{E} = P \otimes_{T_\theta} E \quad [20]$$

One can show that this mapping is functorial. Moreover, the bimodule $Q$ provides us with an inverse mapping $Q \otimes_{T_{\hat{\theta}}} \hat{E} \cong E$.

We further introduce a notion of gauge Morita equivalence (originally called "complete Morita equivalence") that allows one to transport connections along with the mapping of modules [20]. Let $L$ be a $d$-dimensional commutative Lie algebra. We say that the $(T^d_{\hat{\theta}}, T^d_\theta)$ Morita equivalence bimodule $P$ establishes a gauge Morita equivalence if it is endowed with operators $\nabla^P_X, X \in L$ that determine a constant curvature connection simultaneously with respect to $T^d_\theta$ and $T^d_{\hat{\theta}}$, that is, satisfy

$$\begin{aligned} &\nabla^P_X(ea) = (\nabla^P_X e)a + e(\delta_X a) \\ &\nabla^P_X(\hat{a}e) = \hat{a}(\nabla^P_X e) + (\hat{\delta}_X \hat{a})e \\ &[\nabla^P_X, \nabla^P_Y] = 2\pi i \sigma_{XY} \cdot \mathbf{1} \end{aligned} \quad [21]$$

Here $\delta_X$ and $\hat{\delta}_X$ are standard derivations on $T_\theta$ and $T_{\hat{\theta}}$, respectively. In other words, we have two Lie algebra homomorphisms

$$\delta : L \to L_\theta, \qquad \hat{\delta} : L \to L_{\hat{\theta}} \quad [22]$$

If a pair $(P, \nabla^P_X)$ specifies a gauge $(T_\theta, T_{\hat{\theta}})$-equivalence bimodule, then there exists a correspondence between connections on $E$ and connections on $\hat{E}$. The connection $\hat{\nabla}_X$ on $\hat{E}$ corresponding to a given connection $\nabla_X$ on $E$ is defined as

$$\nabla_X \mapsto \hat{\nabla}_X = 1 \otimes \nabla_X + \nabla^P_X \otimes 1 \quad [23]$$

More precisely, an operator $1 \otimes \nabla_X + \nabla^P_X \otimes 1$ on $P \otimes_{\mathbb{C}} E$ descends to a connection $\hat{\nabla}_X$ on $\hat{E} = P \otimes_{T_\theta} E$. It is straightforward to check that under this mapping gauge equivalent connections go to gauge equivalent ones,

$$Z^\dagger \nabla_X Z = \hat{Z}^\dagger \hat{\nabla}_X \hat{Z}$$

where $\hat{Z} = 1 \otimes Z$ is the endomorphism of $\hat{E} = P \otimes_{T_\theta} E$ corresponding to $Z \in \mathrm{End}_{T^d_\theta} E$.

The curvatures of $\hat{\nabla}_X$ and $\nabla_X$ are connected by the formula

$$F_{XY}^{\nabla} = \hat{F}_{XY}^{\nabla} + \mathbf{1}\sigma_{XY} \qquad [24]$$

which in particular shows that constant curvature connections go to constant curvature ones.

Since noncommutative tori are labeled by an antisymmetric $d \times d$ matrix $\theta$, gauge Morita equivalence establishes an equivalence relation on the set of such matrices. To describe this equivalence relation, consider the action $\theta \mapsto h\theta = \hat{\theta}$ of $\mathrm{SO}(d, d|\mathbb{Z})$ on the space of antisymmetric $d \times d$ matrices by the formula

$$\hat{\theta} = (M\theta + N)(R\theta + S)^{-1} \qquad [25]$$

where the $d \times d$ matrices $M$, $N$, $R$, $S$ are such that the matrix

$$h = \begin{pmatrix} M & N \\ R & S \end{pmatrix} \qquad [26]$$

belongs to the group $\mathrm{SO}(d, d|\mathbb{Z})$. The above action is defined whenever the matrix $A \equiv R\theta + S$ is invertible. One can prove that two noncommutative tori $T_\theta^d$ and $T_{\hat{\theta}}$ are gauge Morita equivalent if and only if the matrices $\theta$ and $\hat{\theta}$ belong to the same orbit of the $\mathrm{SO}(d, d|\mathbb{Z})$ action [25].

The duality group $\mathrm{SO}(d, d|\mathbb{Z})$ also acts on the topological numbers of moduli $\mu \in \Lambda^{\mathrm{even}}(\mathbb{Z}^d)$. This action can be shown to be given by a spinor representation constructed as follows. First note that the operators $a^i = \alpha^i, b_i = \partial/\partial\alpha^i$ act on $\Lambda(\mathbb{R}^d)$ and give a representation of the Clifford algebra specified by the metric with signature $(d, d)$. The group $\mathrm{O}(d, d|\mathbb{C})$ can thus be regarded as a group of automorphisms acting on the Clifford algebra generated by $a^i, b_j$. Denote the latter action by $W_h$ for $h \in \mathrm{O}(d, d|\mathbb{C})$. One defines a projective action $V_h$ of $\mathrm{O}(d, d|\mathbb{C})$ on $\Lambda(\mathbb{R}^d)$ according to

$$V_h a^i V_h^{-1} = W_{h^{-1}}(a^i), \qquad V_h b_j V_h^{-1} = W_{h^{-1}}(b_j)$$

This projective action can be restricted to yield a double-valued spinor representation of $\mathrm{SO}(d, d|\mathbb{C})$ on $\Lambda(\mathbb{R}^d)$ by choosing a suitable bilinear form on $\Lambda(\mathbb{R}^d)$. The restriction of this representation to the subgroup $\mathrm{SO}(d, d|\mathbb{Z})$ acting on $\Lambda^{\mathrm{even}}(\mathbb{Z}^d)$ gives the action of Morita equivalence on the topological numbers of moduli.

The mapping [23] preserves the Yang–Mills equations of motion [15]. Moreover, one can define a modification of the Yang–Mills action functional [14] in such a way that the values of the functionals on $\nabla_X$ and $\hat{\nabla}_X$ coincide up to an appropriate rescaling of coupling constants. The modified action functional has the form

$$S_{\mathrm{YM}} = \frac{V}{4g^2} \mathrm{tr}(F_{jk} + \Phi_{jk} \cdot \mathbf{1})(F^{jk} + \Phi^{jk} \cdot \mathbf{1}) \qquad [27]$$

where $\Phi^{jk}$ is a scalar-valued tensor that can be thought of as some background field. Adding this term will allow us to compensate for the curvature shift by adopting the transformation rule

$$\Phi_{XY} \mapsto \Phi_{XY} - \sigma_{XY}$$

Note that the new action functional [27] has the same equations of motion [15] as the original one.

To show that the functional [27] is invariant under gauge Morita equivalence, one has to take into account two more effects. Firstly, the values of trace change by a factor $c = \dim(\hat{E})(\dim(E))^{-1}$ as $\hat{\mathrm{tr}}\,\hat{X} = c\,\mathrm{tr}\,X$. Secondly, the identification of $L_\theta$ and $L_{\hat{\theta}}$ is established by means of some linear transformation $A_j^k$, the determinant of which will rescale the volume $V$. Both effects can be absorbed into an appropriate rescaling of the coupling constant.

One can show that the curvature tensor, the metric tensor, the background field $\Phi_{ij}$, and the volume element $V$ transform according to

$$\begin{aligned} F_{ij}^{\hat{\nabla}} &= A_i^k F_{kl}^{\nabla} A_j^l + \sigma_{ij} \\ \hat{g}_{ij} &= A_i^k g_{kl} A_j^l \\ \hat{\Phi}_{ij} &= A_i^k \Phi_{kl} A_j^l - \sigma_{ij} \\ \hat{V} &= V |\det A| \end{aligned} \qquad [28]$$

where $A = R\theta + S$ and $\sigma = -RA^t$. The action functional [27] is invariant under the gauge Morita equivalence if the coupling constant transforms according to

$$\hat{g}_{\mathrm{YM}}^2 = g_{\mathrm{YM}}^2 |\det A|^{1/2} \qquad [29]$$

Supersymmetric extensions of Yang–Mills theory on noncommutative tori were shown to arise within string theory essentially in two situations. In the first case, one considers compactifications of the (BFSS or IKKT) matrix model of M-theory (Connes *et al.* 1998). A discussion regarding the connection between T-duality and Morita equivalence in this case can be found in Seiberg and Witten (1999, section 7). Noncommutative gauge theories on tori can also be obtained by taking the so-called Seiberg–Witten zero slope limit in the presence of a Neveu–Schwarz $B$-field background (Seiberg and Witten 1999). The emergence of noncommutative geometry in this limit is discussed in this article. Below we give some details on the relation between T-duality and Morita equivalence in this approach. Consider a number of D$p$-branes wrapped on $T^p$ parametrized by

coordinates $x^i \sim x^i + 2\pi r$ with a closed-string metric $G_{ij}$ and a $B$-field $B_{ij}$. The SO$(p, p|\mathbb{Z})$ T-duality group is represented by the matrices

$$T = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad [30]$$

that act on the matrix

$$E = \frac{r^2}{\alpha'}(G + 2\pi\alpha' B)$$

by a fractional transformation

$$T : E \mapsto E' = (aE + b)(cE + d)^{-1} \qquad [31]$$

The transformed metric and $B$-field are obtained by taking, respectively, the symmetric and antisymmetric parts of $E'$. The string coupling constant is transformed as

$$T : g_s \mapsto g'_s = \frac{g_s}{(\det(cE + d))^{1/2}} \qquad [32]$$

The zero slope limit of Seiberg and Witten is obtained by taking

$$\alpha' \sim \sqrt{\epsilon} \to 0, \qquad G_{ij} \sim \epsilon \to 0 \qquad [33]$$

Sending the closed-string metric to zero implies that the $B$-field dominates in the open-string boundary conditions. In the limit [33], the compactification is parametrized in terms of open-string moduli

$$g_{ij} = -(2\pi\alpha')^2 (BG^{-1}B)_{ij}$$
$$\theta^{ij} = \frac{1}{2\pi r^2}(B^{-1})^{ij} \qquad [34]$$

which remain finite. One can demonstrate that $\theta^{ij}$ is a noncommutativity parameter for the torus and the low-energy effective theory living on the D$p$-brane is a noncommutative maximally supersymmetric gauge theory with a coupling constant

$$G_s = g_s \left(\frac{\det g}{\det G}\right)^{1/4} \qquad [35]$$

From the transformation law [31], it is not hard to derive the transformation rules for the moduli [34] in the limit [33],

$$T : g \mapsto g' = (a + b\theta)g(a + b\theta)^t$$
$$T : \theta \mapsto \theta' = (c + d\theta)(a + b\theta)^t \qquad [36]$$

Furthermore, the effective gauge theory becomes a noncommutative Yang–Mills theory [17] with a coupling constant

$$(g_{\rm YM})^{-2} = \frac{(\alpha')^{(3-p)/2}}{(2\pi)^{p-2} G_s}$$

which goes to a finite limit under [33] provided one simultaneously scales $g_s$ with $\epsilon$ as

$$g_s \sim \epsilon^{(3-p+k)/4}$$

where $k$ is the rank of $B_{ij}$. The limiting coupling constant $g_{\rm YM}$ transforms under the T-duality [31], [32] as

$$T : g_{\rm YM} \mapsto g'_{\rm YM} = g_{\rm YM}(\det(a + b\theta))^{1/4} \qquad [37]$$

We see that the transformation laws [31] and [37] have the same form as the corresponding transformations in [25], [28], [29] provided one identifies matrix [26] with matrix [30] conjugated by

$$T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

The need for conjugation reflects the fact that in the BFSS M(atrix) model in the framework of which the Morita equivalence was originally considered, the natural degrees of freedom are D0 branes versus D$p$ branes considered in the above discussion of T-duality.

One can further check that the gauge field transformations following from gauge Morita equivalence match with those induced by the T-duality. It is worth stressing that in the absence of a $B$-field background the effective action based on the square of the gauge field curvature is not invariant under T-duality.

## Instantons on Noncommutative $T^4_\theta$

Consider a Yang–Mills field $\nabla_X$ on a projective module $E$ over a noncommutative 4-torus $T^4_\theta$. Assume that the Lie algebra of shifts $L_\theta$ is equipped with the standard Euclidean metric such that the metric tensor in the basis [7] is given by the identity matrix. The Yang–Mills field $\nabla_i$ is called an instanton if the self-dual part of the corresponding curvature tensor is proportional to the identity operator,

$$F^+_{jk} \equiv \tfrac{1}{2}\left(F_{jk} + \tfrac{1}{2}\epsilon_{jkmn}F^{mn}\right) = i\omega_{jk} \cdot \mathbf{1} \qquad [38]$$

where $\omega_{jk}$ is a constant matrix with real entries. An anti-instanton is defined the same way by replacing the self-dual part with the anti-self-dual one.

One can define a noncommutative analog of Nahm transform for instantons (Astashkevich *et al.* 2000) that has properties very similar to those of the ordinary (commutative) one. To that end, consider a triple $(\mathcal{P}, \nabla_i, \hat{\nabla}_i)$ consisting of a (finite projective) $(T^4_\theta, T^4_{\hat{\theta}})$-bimodule $\mathcal{P}$, $T^4_\theta$-connection $\nabla_i$ and $T^4_{\hat{\theta}}$-connection $\hat{\nabla}_i$ that satisfy the following properties. The connection $\nabla_i$ commutes with the $T_{\hat{\theta}}$-action on $\mathcal{P}$ and the connection $\hat{\nabla}_i$ with that of $T_\theta$. The commutators $[\nabla_i, \nabla_j], [\hat{\nabla}_i, \hat{\nabla}_j], [\nabla_i, \hat{\nabla}_j]$ are proportional to the identity operator

$$[\nabla_i, \nabla_j] = \omega_{ij} \cdot \mathbf{1}$$
$$[\hat{\nabla}_i, \hat{\nabla}_j] = \hat{\omega}_{ij} \cdot \mathbf{1} \qquad [39]$$
$$[\nabla_i, \hat{\nabla}_j] = \sigma_{ij} \cdot \mathbf{1}$$

The above conditions mean that $\mathcal{P}$ is a $T^8_{\theta\oplus(-\hat{\theta})}$-module and $\nabla_i \oplus \hat{\nabla}_i$ is a constant curvature connection on it. In addition, we assume that the tensor $\sigma_{ij}$ is nondegenerate.

For a connection $\nabla^E$ on a right $T^4_\theta$-module $E$, we define a Dirac operator $D = \Gamma^i(\nabla^E_i + \nabla_i)$ acting on the tensor product

$$(E \otimes_{T_\theta} \mathcal{P}) \otimes S$$

where $S$ is the SO(4) spinor representation space and $\Gamma^i$ are four-dimensional Dirac gamma-matrices. The space $S$ is $\mathbb{Z}_2$-graded: $S = S^+ \oplus S^-$ and $D$ is an odd operator so that we can consider

$$D^+ : (E \otimes_{T_\theta} \mathcal{P}) \otimes S^+ \to (E \otimes_{T_\theta} \mathcal{P}) \otimes S^-$$
$$D^- : (E \otimes_{T_\theta} \mathcal{P}) \otimes S^- \to (E \otimes_{T_\theta} \mathcal{P}) \otimes S^+$$

A connection $\nabla^E_i$ on a $T^4_\theta$-module $E$ is called $\mathcal{P}$-irreducible if there exists a bounded inverse to the Laplacian

$$\Delta = \sum_i \left(\nabla^E_{\mathrm{i}} + \nabla_i\right)\left(\nabla^E_{\mathrm{i}} + \nabla_i\right)$$

One can show that if $\nabla^E$ is a $\mathcal{P}$-irreducible instanton, then $\ker D^+ = 0$ and $D^- D^+ = \Delta$. Denote by $\hat{E}$ the closure of the kernel of $D^-$. Since $D^-$ commutes with the $T^4_{\hat{\theta}}$-action on $(E \otimes_{T_\theta} \mathcal{P}) \otimes S^-$ the space $\hat{E}$ is a right $T^4_{\hat{\theta}}$-module. One can prove that this module is finite projective. Let $P : (E \otimes_{T_\theta} \mathcal{P}) \otimes S^- \to \hat{E}$ be a Hermitian projector. Denote by $\nabla^{\hat{E}}$ the composition $P \circ \hat{\nabla}$. One can show that $\nabla^{\hat{E}}$ is a Yang–Mills field on $\hat{E}$.

The noncommutative Nahm transform of a $\mathcal{P}$-irreducible instanton connection $\nabla^E$ on $E$ is defined to be the pair $(\hat{E}, \nabla^{\hat{E}})$. One can further show that $\nabla^{\hat{E}}$ is an instanton.

*See also:* Electroweak Theory; Hopf Algebras and *q*-Deformation Quantum Groups; Noncommutative Geometry from Strings; Quantum Group Differentials, Bundles and Gauge Theory; Quantum Hall Effect; String Field Theory; von Neumann Algebras: Introduction, Modular Theory, and Classification Theory.

## Further Reading

Astashkevich A, Nekrasov N, and Schwarz A (2000) On noncommutative Nahm transform. *Communications in Mathematical Physics* 211: 167–182.

Connes A (1980) C* algèbres et géométrie differentielle. *Comptes Rendus Hebdomaclaires des Seances l'Academie des Sciences, Paris Ser. A-B* 290.

Connes A (1994) *Noncommutative Geometry*. Academic Press.

Connes A, Douglas MR, and Schwarz A (1998) Noncommutative geometry and Matrix theory: compactification on tori. *Journal of High Energy Physics* 02: 003.

Douglas MR and Nekrasov N (2001) Noncommutative field theory. *Reviews of Modern Physics* 73: 977–1029.

Konechny A and Schwarz A (2002) Introduction to M(atrix) theory and noncommutative geometry. *Physics Reports* 360: 353–465.

Li H (2004) Strong Morita equivalence of higher-dimensional noncommutative tori. *Journal für die Reineund Angewandte Mathematik* 576: 167–180.

Polishchuk A and Schwarz A (2003) Categories of holomorphic vector bundles on noncommutative two-tori. *Communications in Mathematical Physics* 236: 135–159.

Rieffel MA (1982) Morita equivalence for operator algebras. In: Kadison RV (ed.) *Operator Algebras and Applications*, vol. 38, Proc. Symp. Pure Math. Providence, RI: American Mathematical Society.

Rieffel MA (1988) Projective modules over higher-dimensional non-commutative tori. *Canadian Journal of Mathematics* 40(2): 257–338.

Rieffel MA (1990) Noncommutative tori – a case study of non-commutative differentiable manifolds. *Contemporary Mathematics* 105: 191–211.

Rieffel MA and Schwarz A (1999) Morita equivalence of multidimensional noncommutative tori. *International Journal of Mathematics* 10(2): 289.

Schwarz A (1998) Morita equivalence and duality. *Nuclear Physics B* 534: 720–738.

Seiberg N and Witten E (1999) String theory and noncommutative geometry. *Journal of High Energy Physics* 9909: 032.

Szabo RJ (2003) Quantum field theory on noncommutative spaces. *Physics Reports* 378: 207.

# Nonequilibrium Statistical Mechanics (Stationary): Overview

**G Gallavotti**, Università di Roma "La Sapienza," Rome, Italy



## Nonequilibrium

Systems in stationary nonequilibrium are mechanical systems subject to nonconservative external forces and to thermostat forces which forbid indefinite increase of the energy and allow reaching statistically stationary states. A system $\Sigma$ is described by the positions and velocities of its $n$ particles $\boldsymbol{X}, \dot{\boldsymbol{X}}$, with the particle positions confined to a finite volume container $C_0$.

If $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ are the particle positions in a Cartesian inertial system of coordinates, the equations of motion are determined by their masses $m_i > 0$, $i = 1, \ldots, n$, by the potential energy of

interaction $V(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \equiv V(\boldsymbol{X})$, by the external nonconservative forces $\boldsymbol{F}_i(\boldsymbol{X}, \boldsymbol{\Phi})$, and by the thermostat forces $-\boldsymbol{\vartheta}_i$ as

$$m_i \ddot{\boldsymbol{x}}_i = -\partial_{\boldsymbol{x}_i} V(\boldsymbol{X}) + \boldsymbol{F}_i(\boldsymbol{X}; \boldsymbol{\Phi}) - \boldsymbol{\vartheta}_i, \quad i = 1, \ldots, n \qquad [1]$$

where $\boldsymbol{\Phi} = (\varphi_1, \ldots, \varphi_q)$ are strength parameters on which the external forces depend. All forces and potentials will be supposed smooth, that is, analytic, in their variables aside from possible impulsive elastic forces describing shocks, and with the property $\boldsymbol{F}(\boldsymbol{X}; 0) = 0$. The impulsive forces are allowed here to model possible shocks with the walls of the container $C_0$ or between hard core particles.

A thermostat is a "reservoir" which may consist of one or more infinite systems which are asymptotically in thermal equilibrium and are separated by boundary surfaces from each other as well as from the system: with the latter, they interact through short-range conservative forces, see Figure 1.

The reservoirs occupy infinite regions of the space outside $C_0$, for example, sectors $C_a \subset \mathbb{R}^3, a = 1, 2 \ldots,$ in space and their particles are in a configuration which is typical of an equilibrium state at temperature $T_a$. This means that the empirical probability of configurations in each $C_a$ is Gibbsian with some temperature $T_a$. In other words, the frequency with which a configuration $(\dot{\boldsymbol{Y}}, \boldsymbol{Y} + \boldsymbol{r})$ occurs in a region $\Lambda + \boldsymbol{r} \subset C_a$ while a configuration $(\dot{\boldsymbol{W}}, \boldsymbol{W} + \boldsymbol{r})$ occurs outside $\Lambda + \boldsymbol{r}$ (with $\boldsymbol{Y} \subset \Lambda, \boldsymbol{W} \cap \Lambda = \emptyset$) averaged over the translations $\Lambda + \boldsymbol{r}$ of $\Lambda$ by $\boldsymbol{r}$ (with the restriction that $\Lambda + \boldsymbol{r} \subset C_a$) is

$$\underset{\boldsymbol{r}+\Lambda \subset C_a}{\text{average}}(f_{\Lambda+\boldsymbol{r}}[(\dot{\boldsymbol{Y}}, \boldsymbol{Y} + \boldsymbol{r}); \dot{\boldsymbol{W}}, \boldsymbol{W} + \boldsymbol{r}]) = \frac{\mathrm{e}^{-\beta_a\left((1/2m_a)|\dot{\boldsymbol{Y}}|^2 + V_a(\boldsymbol{Y}|\boldsymbol{W})\right)}}{\text{normalization}} \qquad [2]$$

Here $m_a$ is the mass of the particles in the $a$th reservoir and $V_a(\boldsymbol{Y}|\boldsymbol{W})$ is the energy of the short-range potential between pairs of particles in $\boldsymbol{Y} \subset C_a$ or with one point in $\boldsymbol{Y}$ and one in $\boldsymbol{W}$. Since the configurations in the system and in the thermostats are not random, [2] should be considered as an "empirical" probability in the sense that it is the frequency density of the events $\{(\dot{\boldsymbol{Y}}, \boldsymbol{Y} + \boldsymbol{r}); \boldsymbol{W} + \boldsymbol{r}\}$: in other words, the configurations $\boldsymbol{\omega}_a$ in the reservoirs should be "typical" in the sense of probability theory of distributions which are asymptotically Gibbsian.

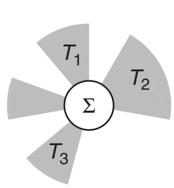



**Figure 1** A symbolic drawing of the container $C_0$ for the system $\Sigma$ and of the surrounding regions containing the particles acting as thermostats at temperatures $T_1, T_2, \ldots$.

The property of being "thermostats" means that [2] remains true for all times, if initially satisfied.

Mathematically, there is a problem at this point: the latter property is either true or false, but a proof of its validity seems out of reach of the present techniques except in very simple cases. Therefore, here we follow an intuitive approach and assume that such thermostats exist and, actually, that any configuration which is typical of a stationary state of an infinite size system of interacting particles in the $C_a$'s, with physically reasonable microscopic interactions, satisfies the property [2].

The above thermostats are examples of "deterministic thermostats" because, together with the system, they form a deterministic dynamical system. They are called "Hamiltonian thermostats" and are often considered as the most appropriate models of "physical thermostats."

A closely related thermostat model is obtained by assuming that the particles outside the system are not in a given configuration but they have a probability distribution whose conditional distributions satisfy [2] initially. Also in this case, it is necessary to assume that [2] remains true for all times, if initially satisfied. Such thermostats are examples of "stochastic thermostats" because their action on the system depends on random variables $\boldsymbol{\omega}_a$ which are the initial configurations of the particles belonging to the thermostats.

Other kinds of stochastic thermostats are "collision rules" with the container boundary $\partial C_0$ of $\Sigma$: every time a particle collides with $\partial C_0$ it is reflected with a momentum $\boldsymbol{p}$ in $\mathrm{d}^3\boldsymbol{p}$ that has a probability distribution proportional to $\mathrm{e}^{-\beta_a(1/2m)\boldsymbol{p}^2}\mathrm{d}^3\boldsymbol{p}$ where $\beta_a, a = 1, 2, \ldots$ depends on which boundary portion (labeled by $a = 1, 2, \ldots$) the collision takes place and $T_a = (k_B \beta_a)^{-1}$ and its "temperature" if $k_B$ is Boltzmann's constant. Which $\boldsymbol{p}$ is actually chosen after each collision is determined by a random variable $\boldsymbol{\omega} = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots)$.

The distinction between stochastic and deterministic thermostats ultimately rests on what we call "system." If reservoirs or the randomness generators are included in the system, then the system becomes deterministic (possibly infinite); and finite deterministic thermostats can also be regarded as simplified models for infinite reservoirs, see the section "Heat, temperature, and entropy production."

It is also possible, and convenient, to consider "finite deterministic thermostats." In the latter case, $\boldsymbol{\vartheta}$ is a force only depending upon the configuration of the $n$ particles $v$ of $\Sigma$ in their finite container $C_0$.

Examples of finite deterministic reservoirs are forces obtained by imposing a nonholonomic constraint via some *ad hoc* principle like the *Gauss principle*. For instance, if a system of particles driven by a force $\boldsymbol{G}_i \stackrel{\text{def}}{=} -\partial_{\boldsymbol{x}_i} V(\boldsymbol{X}) + \boldsymbol{F}_i(\boldsymbol{X})$ is enclosed in a box $C_0$ and $\boldsymbol{\vartheta}$ is a thermostat enforcing an anholonomic constraint $\psi(\dot{\boldsymbol{X}}, \boldsymbol{X}) \equiv 0$ via Gauss' principle, then

$$\boldsymbol{\vartheta}_i(\dot{\boldsymbol{X}}, \boldsymbol{X}) = \left[\frac{\sum_j \dot{\boldsymbol{x}}_j \cdot \partial_{\boldsymbol{x}_j}\psi(\dot{\boldsymbol{X}}, \boldsymbol{X}) + (1/m)\boldsymbol{G}_j \cdot \partial_{\dot{\boldsymbol{x}}_j}\psi(\dot{\boldsymbol{X}}, \boldsymbol{X})}{\sum_j \frac{1}{m}(\partial_{\dot{\boldsymbol{x}}_j}\psi(\dot{\boldsymbol{X}}, \boldsymbol{X}))^2}\right] \times \partial_{\dot{\boldsymbol{x}}_i}\psi(\dot{\boldsymbol{X}}, \boldsymbol{X}) \qquad [3]$$

Gauss' principle says that the force which needs to be added to the other forces $\boldsymbol{G}_i$ acting on the system minimizes

$$\sum_i \frac{(\boldsymbol{G}_i - m_i \boldsymbol{a}_i)^2}{m_i}$$

given $\dot{\boldsymbol{X}}, \boldsymbol{X}$, among all accelerations $\boldsymbol{a}_i$ which are compatible with the constraint $\psi$.

It should be kept in mind that the only known examples of mathematically treatable thermostats modeled by infinite reservoirs are cases in which the thermostat particles are either noninteracting particles or linear (i.e., noninteracting) oscillators. For simplicity stochastic or infinite thermostats will not be considered here and we restrict attention to finite deterministic systems.

In general, in order that a force $\boldsymbol{\vartheta}$ can be considered a deterministic "thermostat force" a further property is necessary: namely that the system evolves according to [1] towards a stationary state. This means that for all initial particle configurations $(\dot{\boldsymbol{X}}, \boldsymbol{X})$, except possibly for a set of zero phase-space volume, any smooth function $f(\dot{\boldsymbol{X}}, \boldsymbol{X})$ evolves in time so that, if $S_t(\dot{\boldsymbol{X}}, \boldsymbol{X})$ denotes the configuration into which the initial data evolve in time $t$ according to [1], then the limit

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(S_t(\dot{\boldsymbol{X}}, \boldsymbol{X}))\, \mathrm{d}t = \int f(z)\mu(\mathrm{d}z) \qquad [4]$$

exists and is independent of $(\dot{\boldsymbol{X}}, \boldsymbol{X})$. The probability distribution $\mu$ is then called the SRB distribution for the system. The maps $S_t$ will have the group property $S_t \cdot S_{t'} = S_{t+t'}$ and the SRB distribution $\mu$ will be invariant under time evolution.

It is important to stress that the requirement that the exceptional configurations form just a set of zero phase volume (rather than a set of zero probability with respect to another distribution, singular with respect to the phase volume) is a strong assumption and it should be considered an axiom of the theory: it corresponds to the assumption that the initial configuration is prepared as a typical configuration of an equilibrium state, which, by the classical equidistribution axiom of equilibrium statistical mechanics, is a typical configuration with respect to the phase volume.

For this reason, the SRB distribution is said to describe a "stationary nonequilibrium state" of the system. The SRB distribution depends on the parameters on which the forces acting on the system depend, for example, $|C_0|$ (volume), $\boldsymbol{\Phi}$ (strength of the forcings), $\{\beta_a^{-1}\}$ (temperatures), etc. The collection of SRB distributions obtained by letting the parameters vary defines a "nonequilibrium ensemble."

In the stochastic case, the distribution $\mu$ is required to be invariant in the sense that it can be regarded as a marginal distribution of an invariant distribution for the larger (deterministic) system formed by the thermostats and the system itself.

For more details, the reader is referred to Evans and Morriss (1990), Ruelle (1999), and Eckmann *et al.* (1999).

## Nonequilibrium Thermodynamics

The key problem of nonequilibrium statistical mechanics is to derive a macroscopic "nonequilibrium thermodynamics" in a way similar to the derivation of equilibrium thermodynamics from equilibrium statistical mechanics.

The first difficulty is that nonequilibrium thermodynamics is not well understood. For instance, there is no (agreed upon) definition of entropy of a nonequilibrium stationary state, while it should be kept in mind that the effort to find the microscopic interpretation of equilibrium entropy, as defined by Clausius, was a driving factor in the foundations of equilibrium statistical mechanics.

The importance of entropy in classical equilibrium thermodynamics rests on the implication of universal, parameter-free relations which follow from its existence (e.g., $\partial_V(1/T) \equiv \partial_U(p/T)$ if $U$ is the internal energy, $T$ the absolute temperature, and $p$ the pressure of a simple homogeneous material).

Are there universal relations among averages of observables with respect to SRB distributions?

The question has to be posed for systems "really" out of equilibrium, that is, for $\boldsymbol{\Phi} \neq 0$ (see [1]): in fact, there is a well-developed theory of the derivatives with respect to $\boldsymbol{\Phi}$ of averages of

observables evaluated at $\Phi = 0$. The latter theory is often called, and here we shall do so as well, "classical nonequilibrium thermodynamics" or "near-equilibrium thermodynamics" and it has been quite successfully developed on the basis of the notions of equilibrium thermodynamics, paying particular attention to the macroscopic evolution of systems described by macroscopic continuum equations of motion.

"Stationary nonequilibrium statistical mechanics" will indicate a theory of the relations between averages of observables with respect to SRB distributions. Systems so large that their volume elements can be regarded as being in locally stationary nonequilibrium states could also be considered. This would extend the familiar "local equilibrium states" of classical nonequilibrium thermodynamics: however, they are not considered here. This means that we shall not attempt to find the macroscopic equations regulating the time evolution of continua locally in nonequilibrium stationary states but we shall only try to determine the properties of their "volume elements" assuming that the timescale for the evolution of large assemblies of volume elements is slow compared to the timescales necessary to reach local stationarity.

For more details, the reader is referred to de Groot and Mazur (1984), Lebowitz (1993), Ruelle (1999, 2000), Gallavotti (1998, 2004), and Goldstein and Lebowitz (2004).

## Chaotic Hypothesis

In equilibrium statistical mechanics, the ergodic hypothesis plays an important conceptual role as it implies that the motions of ergodic systems have an SRB statistics and that the latter coincides with the Liouville distribution on the energy surface.

An analogous role has been proposed for the "chaotic hypothesis," which states that the

> *motion of a chaotic system, developing on its attracting set, can be regarded as an Anosov flow.*

This means that the attracting sets of chaotic systems, physically defined as systems with at least one positive Lyapunov exponent, can be regarded as smooth surfaces on which motion is highly unstable:

1. Around every point, a curvilinear coordinate system can be established which has three planes, varying continuously with $x$, which are covariant (i.e., they are coordinate planes at a point $x$ which are mapped, by the evolution $S_t$, into the corresponding coordinate planes around $S_t x$).
2. The planes are of three types, "stable," "unstable," and "marginal," with respective positive dimensions $d_s, d_u$, and 1: infinitesimal lengths on the stable plane and on the unstable plane of any point contract at exponential rate as time proceeds towards the future or towards the past. The length along the marginal direction neither contracts nor expands (i.e., it varies around the initial value staying bounded away from 0 and $\infty$): its tangent vector is parallel to the flow. In cases in which time evolution is discrete, and determined by a map $S$, the marginal direction is missing.
3. The contraction over a time $t$, positive for lines on the stable plane and negative for those on the unstable plane, is exponential, i.e. lengths are contracted by a factor uniformly bounded by $Ce^{-\kappa|t|}$ with $C, \kappa > 0$.
4. There is a dense trajectory.

It has to be stressed that the chaotic hypothesis concerns physical systems: mathematically, it is very easy to find dynamical systems for which it does not hold, at least as easy as it is to find systems in which the ergodic hypothesis does not hold (e.g., harmonic lattices or blackbody radiation). However, if suitably interpreted, the ergodic hypothesis leads, even for these systems, to physically correct results (the specific heats at high temperature, the Raleigh–Jeans distribution at low frequencies). Moreover, the failures of the ergodic hypothesis in physically important systems have led to new scientific paradigms (like quantum mechanics from the specific heats at low temperature and Planck's law).

Since physical systems are almost always not Anosov systems, it is very likely that probing motions in extreme regimes will make visible the features that distinguish Anosov systems from non-Anosov systems, much as it happens with the ergodic hypothesis.

The interest of the hypothesis is to provide a framework in which properties like the existence of an SRB distribution is *a priori* guaranteed, together with an expression for it which can be used to work with formal expressions of the averages of the observables: the role of Anosov systems in chaotic dynamics is similar to the role of harmonic oscillators in the theory of regular motions. They are the paradigm of chaotic systems, as the harmonic oscillators are the paradigm of order. Of course, the hypothesis is only a beginning and one has to learn how to extract information from it, as it was the case with the use of the Liouville distribution, once the ergodic hypothesis guaranteed that it was the

appropriate distribution for the study of the statistics of motions in equilibrium situations.

For more details, the reader is referred to Ruelle (1976), Gallavotti and Cohen (1995), Ruelle (1999), Gallavotti (1998), and Gallavotti *et al.* (2004).

## Heat, Temperature, and Entropy Production

The amount of *heat* $\dot{Q}$ that a system produces while in a stationary state is naturally identified with the work that the thermostat forces $\boldsymbol{\vartheta}$ perform per unit time

$$\dot{Q} = \sum_i \boldsymbol{\vartheta}_i \cdot \dot{\boldsymbol{x}}_i \qquad [5]$$

A system may be in contact with several reservoirs: in models, this will be reflected by a decomposition

$$\boldsymbol{\vartheta} = \sum \boldsymbol{\vartheta}^{(a)}(\dot{\boldsymbol{X}}, \boldsymbol{X}) \qquad [6]$$

where $\boldsymbol{\vartheta}^{(a)}$ is the force due to the $a$th thermostat and depends on the coordinates of the particles which are in a region $\Lambda_a \subseteq C_0$ of a decomposition $\cup_{a=1}^{m} \Lambda_a = C_0$ of the container $C_0$ occupied by the system ($\Lambda_a \cap \Lambda_{a'} = \emptyset$ if $a \neq a'$).

From several studies based on simulations of finite thermostatted systems of particles arose the proposal to consider the average of the phase-space contraction $\sigma^{(a)}(\dot{\boldsymbol{X}}, \boldsymbol{X})$ due to the $a$th thermostat

$$\sigma^{(a)}(\dot{\boldsymbol{X}}, \boldsymbol{X}) \stackrel{\text{def}}{=} \sum_j \partial_{\dot{x}_j} \cdot \boldsymbol{\vartheta}_j^{(a)}(\dot{\boldsymbol{X}}, \boldsymbol{X}) \qquad [7]$$

and to identify it with the *rate of entropy creation* in the $a$th thermostat.

Another key notion in thermodynamics is the temperature of a reservoir; in the infinite deterministic thermostat case, of the section "Nonequilibrium," it is defined as $(k_{\text{B}}\beta_a)^{-1}$ but in the finite deterministic thermostats considered here it needs to be defined. If there are $m$ reservoirs with which the system is in contact, one sets

$$\sigma_+^{(a)} \stackrel{\text{def}}{=} \langle \sigma^{(a)}(\dot{\boldsymbol{X}}, \boldsymbol{X}) \rangle \equiv \int \sigma^{(a)}(\dot{\boldsymbol{X}}, \boldsymbol{X})\, \mu(\mathrm{d}\dot{\boldsymbol{X}}\, \mathrm{d}\boldsymbol{X})$$
$$\dot{Q}_a \stackrel{\text{def}}{=} \sum_i \boldsymbol{\vartheta}_i^{(a)} \cdot \dot{\boldsymbol{x}}_i \qquad [8]$$

where $\mu$ is the SRB distribution describing the stationary state. It is natural to define the absolute temperature of the $a$th thermostat to be

$$T_a = \frac{\langle \dot{Q}_a \rangle}{k_{\text{B}} \sigma_+^{(a)}} \qquad [9]$$

It is not clear that $T_a > 0$: this happens in a rather general class of models and it would be desirable, for the interpretation that is proposed here, that it could be considered a property to be added to the requirements that the forces $\boldsymbol{\vartheta}^{(a)}$ be thermostat models.

An important class of thermostats for which the property $T_a > 0$ holds can be described as follows. Imagine $N$ particles in a container $C_0$ interacting via a potential $V_0 = \sum_{i<j} \varphi(\boldsymbol{q}_i - \boldsymbol{q}_j) + \sum_j V'(\boldsymbol{q}_j)$ (where $V'$ models external conservative forces like obstacles, walls, gravity, ...) and, furthermore, interacting with $M$ other systems $\Sigma_a$, of $N_a$ particles of mass $m_a$, in containers $C_a$ contiguous to $C_0$. The latter will model $M$ parts of the system in contact with thermostats at temperatures $T_a, a = 1, \ldots, M$.

The coordinates of the particles in the $a$th system $\Sigma_a$ will be denoted $\boldsymbol{x}_j^a, j = 1, \ldots, N_a$, and they will interact with each other via a potential $V_a = \sum_{i,j}^{N_a} \varphi_a(\boldsymbol{x}_i^a - \boldsymbol{x}_j^a)$. Furthermore, there will be an interaction between the particles of each thermostat and those of the system via potentials $W_a = \sum_{i=1}^{N} \sum_{j=1}^{N_a} w_a(\boldsymbol{q}_i - \boldsymbol{x}_j^a), a = 1, \ldots, M$.

The potentials will be assumed to be either hard core or nonsingular potentials and the external $V'$ is supposed to be at least such that it forbids the existence of obvious constants of motion.

The temperature of each $\Sigma_a$ will be defined by the total kinetic energy of its particles, that is, by $K_a = \sum_{j=1}^{N_a} (1/2) m_a (\dot{\boldsymbol{x}}_j^a)^2 \stackrel{\text{def}}{=} (3/2) N_a k_{\text{B}} T_a$: the particles of the $a$th thermostat will be kept at constant temperature by further forces $\boldsymbol{\vartheta}_j^a$. The latter are defined by imposing via a Gaussian constraint that $K_a$ is a constant of motion (see [3] with $\psi \equiv K_a$). This means that the equations of motion are

$$\begin{aligned} m\ddot{\boldsymbol{q}}_j &= -\partial_{\boldsymbol{q}_j}\left( V_0(\boldsymbol{Q}) + \sum_{a=1}^{N_a} W_a(\boldsymbol{Q}, \boldsymbol{x}^a) \right) \\ m_a \ddot{\boldsymbol{x}}_j^a &= -\partial_{\boldsymbol{x}_j^a}\left( V_a(\boldsymbol{x}^a) + W_a(\boldsymbol{Q}, \boldsymbol{x}^a) \right) - \boldsymbol{\vartheta}_j^a \end{aligned} \qquad [10]$$

and an application of Gauss' principle yields

$$\boldsymbol{\vartheta}_j^a = \frac{L_a - \dot{V}_a}{3 N_a k_{\text{B}} T_a} \dot{\boldsymbol{x}}_j^a \stackrel{\text{def}}{=} \alpha^a \dot{\boldsymbol{x}}_j^a$$

where $L_a$ is the work per unit time done by the particles in $C_0$ on the particles of $\Sigma_a$ and $V_a$ is their potential energy.

In this case, the partial divergence $\sigma^a \equiv (3N_a - 1)\alpha^a$ is, up to a constant factor $(1- \ (1/3N_a))$,

$$\sigma^a = \frac{L_a}{k_{\text{B}} T_a} - \frac{\dot{V}_a}{k_{\text{B}} T_a}$$

and it will make [9] identically satisfied with $T_a > 0$ because $L_a$ can be naturally interpreted as heat $Q_a$ ceded, per unit time, by the particles in $C_0$ to the subsystem $\Sigma_a$ (hence to the $a$th thermostat because the temperature of $\Sigma_a$ is constant), while the

derivative of $V_a$ will not contribute to the value of $\sigma_+^a$. The phase-space contraction rate is, neglecting the total derivative terms (and $O(N_a^{-1})$),

$$\sigma_{\text{true}}(\dot{X}, X) = \sum_{a=1}^{N_a} \frac{\dot{Q}_a}{k_B T_a} \qquad [11]$$

where the subscript "true" is to remind that an additive total derivative term distinguishes it from the complete phase-space contraction.

**Remarks**

(i) The above formula provides the motivation of the name "entropy creation rate" attributed to the phase-space contraction $\sigma$. Note that in this way the definition of entropy creation is "reduced" to the equilibrium notion because what is being defined is the entropy increase of the thermostats which have to be considered in equilibrium. No attempt is made here to define neither the entropy of the stationary state nor the notion of temperature of the nonequilibrium system in $C_0$ (the $T_a$ are temperatures of the $\Sigma_a$, not of the particles in $C_0$). This is an important point as it leaves open the possibility of envisaging the notion of "local equilibrium" which becomes necessary in the approximation (not considered here) in which the system is regarded as a continuum.

(ii) In the above model, another viewpoint is possible: that is, to consider the system to consist of only the $N$ particles in $C_0$ and the $M$ systems $\Sigma_a$ to be thermostats. From this point of view, it can be considered a model of a system subject to thermostats. The Gibbs distribution characterizing the infinite thermostats of the section "Nonequilibrium" becomes in this case the constraint that the kinetic energies $K_a$ are constants, enforced by the Gaussian forces. In the new viewpoint, the appropriate definition should be simply the right-hand side (RHS) of [11], i.e. the work per unit time done by the forces of the system on the thermostats divided by the temperature of the thermostats. This suggests a different and general definition of entropy creation rate, applying also to thermostats that are often considered "more physical" and that needs to be further investigated. In the example [10] the new definition differs from the phase space contraction rate by a total time derivative, i.e. rather trivially for the purposes of the following.

For more details, the reader is referred to Evans and Morriss (1990), Gallavotti and Cohen (1995), Ruelle (1996, 1997), and Gallavotti (2004).

## Thermodynamic Fluxes and Forces

Nonequilibrium stationary states depend upon external parameters $\varphi_j$ like the temperatures $T_a$ of the thermostats or the size of the force parameters $\mathbf{\Phi} = (\varphi_1, \ldots, \varphi_q)$, see [1]. Nonequilibrium thermodynamics is well developed at "low forcing": strictly speaking, this means that it is widely believed that we understand the properties of the derivatives of the averages of observables with respect to the external parameters if evaluated at $\varphi_j = 0$. Important notions are the notions of thermodynamic fluxes $J_i$ and of thermodynamic forces $\varphi_i$; hence, it seems important to extend such notions to nonequilibrium systems (i.e., $\mathbf{\Phi} \neq \mathbf{0}$).

A possible extension could be to define the thermodynamic flux $J_i$ associated with a force $\varphi_i$ as $J_i = \langle \partial_{\varphi_i} \sigma \rangle_{\text{SRB}}$ where $\sigma(X, \dot{X}; \mathbf{\Phi})$ is the volume contraction per unit time. This definition seems appropriate in several concrete cases that have been studied and it is appealing for its generality.

An interesting example is provided by the model of thermostatted system in [10]: if the container of the system is a box with periodic boundary conditions, one can imagine to add an extra constant force $\boldsymbol{E}$ acting on the particles in the container. Imagining the particles to be charged by a charge $e$ and regarding such force as an electric field, the first equation in [10] is modified by the addition of a term $e\boldsymbol{E}$.

The constraints on the thermostat temperatures imply that $\sigma$ depends also on $\boldsymbol{E}$: in fact, if $\boldsymbol{J} = e\sum_j \dot{\boldsymbol{q}}_j$ is the electric current, energy balance implies $\dot{U}_{\text{tot}} = \boldsymbol{E} \cdot \boldsymbol{J} - \sum_a (L_a - \dot{V}_a)$ if $U_{\text{tot}}$ is the sum of all kinetic and potential energies. Then, the phase-space contraction

$$\sum_a \frac{L_a - \dot{V}_a}{T_a}$$

can be written, to first order in the temperature variations $\delta T_a$ with respect to a common value $T_a = T$, as

$$-\sum_a \frac{L_a - \dot{V}_a}{T} \frac{\delta T_a}{T} + \frac{\boldsymbol{E} \cdot \boldsymbol{J} - \dot{U}_{\text{tot}}}{T}$$

hence $\sigma_{\text{true}}$, see [11], is

$$\sigma_{\text{true}} = \frac{\boldsymbol{E} \cdot \boldsymbol{J}}{k_B T} - \sum_a \frac{\dot{Q}_a}{k_B T} \frac{\delta T_a}{T} \qquad [12]$$

The definition and extension of the conjugacy between thermodynamic forces and fluxes is compatible with the key results of classical nonequilibrium thermodynamics, at least as far as Onsager

reciprocity and Green–Kubo's formulas are concerned. It can be checked that if the equilibrium system is reversible, that is, if there is an isometry $I$ on phase space which anticommutes with the evolution ($IS_t = S_{-t}I$ in the case of continuous-time dynamics $t \to S_t$ or $IS = S^{-1}I$ in the case of discrete-time dynamics $S$), then, shortening $(\dot{X}, X)$ into $x$,

$$L_{ij} \stackrel{\text{def}}{=} \partial_{\Phi_i} J_j|_{\mathbf{\Phi}=0} = \partial_{\Phi_i} \langle \partial_{\Phi_j} \sigma(x; \mathbf{\Phi}) \rangle_{\text{SRB}}|_{\mathbf{\Phi}=0} = \partial_{\Phi_j} J_i|_{\mathbf{\Phi}=0}$$
$$= L_{ji} = \frac{1}{2} \int_{-\infty}^{\infty} \langle \partial_{\Phi_j} \sigma(S_t x; \mathbf{\Phi}) \partial_{\Phi_i} \sigma(x; \mathbf{\Phi}) \rangle_{\text{SRB}}|_{\mathbf{\Phi}=0} \, dt \quad [13]$$

The $\sigma(x; \mathbf{\Phi})$ plays the role of "Lagrangian" generating the duality between forces and fluxes. The extension of the duality just considered might be of interest in situations in which $\mathbf{\Phi} \neq 0$.

For more details, the reader is referred to de Groot and Mazur (1984), Gallavotti (1996), and Gallavotti and Ruelle (1997).

## Fluctuations

As in equilibrium, large statistical fluctuations of observables are of great interest and already there is, at the moment, a rather large set of experiments dedicated to the analysis of large fluctuations in stationary states out of equilibrium.

If one defines the dimensionless phase-space contraction

$$p(x) = \frac{1}{\tau} \int_0^\tau \frac{\sigma(S_t x)}{\sigma_+} \, dt \quad [14]$$

(see also [11]), then there exists $p^* \geq 1$ such that the probability $P_\tau$ of the event $p \in [a, b]$ with $[a, b] \subset (-p^*, p^*)$ has the form

$$P_\tau(p \in [a, b]) = \text{const. } e^{\tau \max_{p \in [a,b]} \zeta(p) + O(1)} \quad [15]$$

with $\zeta(p)$ analytic in $(-p^*, p^*)$. The function $\zeta(p)$ can be conveniently normalized to have value 0 at $p = 1$ (i.e., at the average value of $p$).

Then, in Anosov systems which are reversible and dissipative (see the previous section), a general symmetry property, called the "fluctuation theorem" and reflecting the reversibility symmetry, yields the parameterless relation

$$\zeta(-p) = \zeta(p) - p\sigma_+ \quad p \in (-p^*, p^*) \quad [16]$$

This relation is interesting because it has no free parameters; in other words, it is universal for reversible dissipative Anosov systems. In connection with the flux–force duality in the previous section, it can be checked to reduce to the Green–Kubo formula and to Onsager reciprocity, see [13], in the case in which the evolution depends on several fields $\mathbf{\Phi}$ and $\mathbf{\Phi} \to 0$ (of course the relation becomes trivial as $\mathbf{\Phi} \to 0$ because $\sigma_+ \to 0$ and to obtain the result one has first to divide both sides by suitable powers of the fields $\mathbf{\Phi}$).

A more informal (but imprecise) way of writing [15] and [16] is

$$\frac{P_\tau(p)}{P_\tau(-p)} = e^{\tau p \sigma_+ + O(1)}, \quad \text{for all } p \in (-p^*, p^*) \quad [17]$$

where $P_\tau(p)$ is the probability density of $p$. An obvious but interesting consequence of [17] is that

$$\langle e^{-\tau p \sigma_+} \rangle_{\text{SRB}} = 1$$

in the sense that $(1/\tau) \log \langle e^{-\tau p \sigma_+} \rangle_{\text{SRB}} \xrightarrow[\tau \to \infty]{} 0$.

Occasionally, systems with singularities have to be considered. In such cases, the relation [16] may change in the sense that the function $\zeta(p)$ may not be analytic: in such cases, one expects that the relation holds in the largest analyticity interval symmetric around the origin. In Anasov systems and also various cases considered in the literature, such interval appears to contain the interval $(-1, 1)$.

Note that in the theory of fluctuations of the time averages $p$ we can replace $\sigma$ by any other *bounded* quantity which is a total time derivative: hence, in the example discussed above, it can be replaced by $\sigma_{\text{true}}$, see [12], which has a natural physical meaning.

It is important to remark that the above fluctuation relation is the first representative of several consequences of the reversibility and chaotic hypotheses. For instance, given $F_1, \ldots, F_n$ arbitrary observables which are (say) odd under time reversal $I$ (i.e., $F(Ix) = -F(x)$) and given $n$ functions $t \in [-\tau/2, \tau/2] \to \varphi_j(t), j = 1, \ldots, n$, one can ask which is the probability that $F_j(S_t x)$ "closely follows" the "pattern" $\varphi_j(t)$ and at the same time

$$\frac{1}{\tau} \int_0^\tau \frac{\sigma(S_\theta x)}{\sigma_+} \, d\theta$$

has value $p$. Then calling $P_\tau(F_1 \sim \varphi_1, \ldots, F_n \sim \varphi_n, p)$ the probability of this event, which we write in the imprecise form corresponding to [17] for simplicity, and defining $I\varphi_j(t) \stackrel{\text{def}}{=} -\varphi_j(-t)$, it is

$$\frac{P_\tau(F_1 \sim \varphi_1, \ldots, F_n \sim \varphi_n, p)}{P_\tau(F_1 \sim I\varphi_1, \ldots, F_n \sim I\varphi_n, -p)} = e^{\tau \sigma_+ p}$$
$$p \in (-p^*, p^*) \quad [18]$$

which is remarkable because it is parameterless and at the same time surprisingly independent of the choice of the observables $F_j$. The relation [18] has far-reaching consequences: for instance, if $n = 1$ and $F_1 = \partial_{\Phi_i} \sigma(x; \mathbf{\Phi})$ the relation [18] has been used to derive the mentioned Onsager reciprocity and Green–Kubo's formulas at $\mathbf{\Phi} = 0$.

Equation [18] can be read as follows: the probability that the observables $F_j$ follow given evolution patterns $\varphi_j$ conditioned to entropy creation rate $p\sigma_+$ is the same that they follow the time-reversed patterns if conditioned to entropy creation rate $-p\sigma_+$. In other words, to change the sign of time, it is just sufficient to reverse the sign of entropy creation rate, no "extra effort" is needed.

For more details, the reader is referred to Sinai (1972, 1994), Evans *et al.* (1993), Gallavotti and Cohen (1995), Gallavotti (1996, 1999), Gallavotti and Ruelle (1997), Gallavotti *et al.* (2004), and Bonetto *et al.* (2005).

## Fractal Attractors, Pairing, and Time Reversal

Attracting sets (i.e., sets which are the closure of attractors) are fractal in most dissipative systems. However, the chaotic hypothesis assumes that fractality can be neglected. Apart from the very interesting cases of systems close to equilibrium, in which the closure of an attractor is the whole phase space (under the chaotic hypothesis, i.e., if the system is Anosov), hence not fractal, serious problems arise in preserving validity of the fluctuation theorem.

The reason is very simple: if the attractor closure is smaller than phase space, then it is to be expected that time reversal will change the attractor into a repeller disjoint from it. Thus, even if the chaotic hypothesis is assumed, so that the attracting set $\mathcal{A}$ can be considered a smooth surface, the motion on the attractor will not be time-reversal symmetric (as its time-reversal image will develop on the repeller). One can say that an attracting set with dimension lower than that of phase space in a time-reversible system corresponds to a spontaneous breakdown of time-reversal symmetry.

It has been noted however that there are classes of systems, forming a large set in the space of evolutions depending on a parameter $\Phi$, in which geometric reasons imply that if beyond a critical value $\Phi_c$ the attracting set becomes smaller than phase space, then a map $I_P$ is generated mapping the attractor $\mathcal{A}$ into the repeller $\mathcal{R}$, and vice versa, such that $I_P^2$ is the identity on $\mathcal{A}\cup\mathcal{R}$ and $I_P$ commutes with the evolution: therefore, the composition $I\cdot I_P$ is a time-reversal symmetry (i.e., it anticommutes with evolution) for the motions on the attracting set $\mathcal{A}$ (as well as on the repeller $\mathcal{R}$).

In other words, the time-reversal symmetry in such systems "cannot be broken": if spontaneous breakdown occurs (i.e., $\mathcal{A}$ is not mapped into itself under time reversal $I$), a new symmetry $I_P$ is spawned and $I\cdot I_P$ is a new time-reversal symmetry (an analogy with the spontaneous violation of time reversal in quantum theory, where time reversal $T$ is violated but $TCP$ is still a symmetry: so $T$ plays the role of $I$ and $CP$ that of $I_P$).

Thus, a fluctuation relation will hold for the phase-space contraction of the motions taking place on the attracting set for the class of systems with the geometric property mentioned above (technically, the latter is called "axiom $C$" property).

This is interesting but it still is quite far from being checkable even in numerical experiments. There are nevertheless systems in which a "pairing property" also holds: this means that, considering the case of discrete-time maps $S$, the Jacobian matrix $\partial_x S(x)$ has $2N$ eigenvalues that can be labeled, in decreasing order, $\lambda_N(x),\ldots,\lambda_{(1/2)N}(x),\ldots,\lambda_1(x)$, with the remarkable property that $(1/2)(\lambda_{N-j}(x)+\lambda_j(x))\stackrel{\text{def}}{=}\alpha(x)$ is $j$-independent. In such systems, a relation can be established between phase-space contractions in the full phase space and on the surface of the attracting set: the fluctuation theorem for the motion on the attracting set can therefore be related to the properties of the fluctuations of the total phase-space contraction measured on the attracting set (which includes the contraction transversal to the attracting set) and if $2M$ is the attracting set dimension and $2N$ is the total dimension of phase space it is, in the analyticity interval $(-p^*,p^*)$ of the function $\zeta(p)$,

$$\zeta(-p)=\zeta(p)-p\frac{M}{N}\sigma_+ \qquad [19]$$

which is an interesting relation. It is however very difficult to test in mechanical systems because in such systems it seems very difficult to make the field so high to see an attracting set thinner than the whole phase space and still observe large fluctuations.

For more details, the reader is referred to Dettman and Morriss (1996) and Gallavotti (1999).

## Nonequilibrium Ensembles and Their Equivalence

Given a chaotic system, the collection of the SRB distributions associated with the various control parameters (volume, density, external forces, ...) forms an "ensemble" describing the possible stationary states of the system and their statistical properties.

As in equilibrium, one can imagine that the system can be described equivalently in several ways at least when the system is large ("in the

thermodynamic" or "macroscopic limit"). In nonequilibrium, equivalence can be quite different and more structured than in equilibrium because one can imagine to change not only the control parameters but also the thermostatting mechanism.

It is intuitive that a system may behave in the same way under the influence of different thermostats: the important phenomenon being the extraction of heat and not the way in which it is extracted from the system. Therefore, one should ask when two systems are "physically equivalent," that is, when the SRB distributions associated with them give the same statistical properties for the same observables, at least for the very few observables which are macroscopically relevant. The latter may be a few more than the usual ones in equilibrium (temperature, pressure, density, etc.) and include currents, conducibilities, viscosities, etc., but they will always be very few compared to the (infinite) number of functions on phase space.

As an example, consider a system of $N$ interacting particles (say hard spheres) of mass $m$ moving in a periodic box $C_0$ of side $L$ containing a regular array of spherical scatterers (a basic model for electrons in a crystal) which reflect particles elastically and are arranged so that no straight line exists in $C_0$ which avoids the obstacles (to eliminate obvious constants of motion). An external field $E\boldsymbol{u}$ acts also along the $\boldsymbol{u}$-direction: hence, the equations of motion are

$$m\ddot{\boldsymbol{x}}_i = \boldsymbol{f}_i + E\boldsymbol{u} - \boldsymbol{\vartheta}_i \qquad [20]$$

where $\boldsymbol{f}_i$ are the interparticle forces and those between scatterers and particles, and $\boldsymbol{\vartheta}_i$ are the thermostatting forces. The following thermostat models have been considered:

1. $\boldsymbol{\vartheta}_i = \nu\dot{\boldsymbol{x}}_i$ (viscosity thermostat),
2. immediately after elastic collision with an obstacle the velocity is rescaled to a prefixed value $\sqrt{3k_{\mathrm{B}}Tm^{-1}}$ for some $T$ (Drude's thermostat),
3. $\boldsymbol{\vartheta}_i = (\boldsymbol{E}\cdot\sum\dot{\boldsymbol{x}}_i)/\sum_i\dot{\boldsymbol{x}}_i^2$ (Gauss' thermostat).

The first two are not reversible. At least not manifestly such, because the natural time reversal, that is, change of velocity sign, is not a symmetry (there might be however more hidden, hitherto unknown, symmetries which anticommute with time evolution). The third is reversible and time reversal is just the change of the velocity sign. The third thermostat model generates a time evolution in which the total kinetic energy $K$ is constant.

Let $\mu'_\nu, \mu''_T, \mu'''_K$ be the SRB distributions for the system in a container $C_0$ with volume $|C_0| = L^3$ and density $\rho = N/L^3$ fixed. Imagine to tune the values of the control parameters $\nu, T, K$ in such a way that $\langle\text{kinetic energy}\rangle_\mu = \mathcal{E}$, with the same $\mathcal{E}$ for $\mu = \mu'_\nu, \mu''_T, \mu'''_K$ and consider a local observable $F(\dot{\boldsymbol{X}}, \boldsymbol{X}) > 0$ depending only on the coordinates of the particles located in a region $\Lambda \subset C_0$. Then a reasonable conjecture is that

$$\lim_{\substack{L\to\infty\\ N/L^3=\rho}} \frac{\langle F\rangle_{\mu'_\nu}}{\langle F\rangle_{\mu''_T}} = \lim_{\substack{L\to\infty\\ N/L^3=\rho}} \frac{\langle F\rangle_{\mu'_\nu}}{\langle F\rangle_{\mu'''_T}} = 1 \qquad [21]$$

if the limits are taken at fixed $F$ (hence at fixed $\Lambda$ while $L\to\to\infty$). The conjecture is an open problem: it illustrates, however, the kind of questions arising in nonequilibrium statistical mechanics.

For more details, the reader is referred to Evans and Sarman (1993), Gallavotti (1999), and Ruelle (2000).

## Outlook

The subject is (clearly) at a very early stage of development.

1. The theory can be extended to stochastic thermostats quite satisfactorily, at least as far as the fluctuation theorem is concerned.
2. Remarkable works have appeared on the theory of systems which are purely Hamiltonian and (therefore) with thermostats that are infinite: unfortunately, the infinite thermostats can be treated, so far, only if their particles are "free" at infinity (either free gases or harmonic lattices).
3. The notion of entropy turns out to be extremely difficult to extend to stationary states and there are even doubts that it could be actually extended. Conceptually, this is certainly a major open problem.
4. The statistical properties of stationary states out of equilibrium are still quite mysterious and surprising: some exactly solvable models have appeared recently, and attempts have been made at unveiling the deep reasons for their solubility and at deriving from them general guiding principles.
5. Numerical simulations have given a strong impulse to the subject; in fact, one can even say that they created it: introducing the model of thermostat as an extra microscopic force acting on the particles and providing the first reliable results on the properties of systems out of equilibrium. Simulations continue to be an essential part of the effort of research on the field.
6. Approach to stationarity leads to many important questions: is there a Lyapunov function measuring the distance between an evolving state and the stationary state towards which it evolves? In other words, can one define an

analogous of Boltzmann's *H*-function? About this question there have been proposals and the answer seems affirmative, but it does not seem that it is possible to find a universal, system-independent, such function (search for it is related to the problem of defining an entropy function for stationary states: its existence is at least controversial, see the sections "Nonequilibrium thermodynamics" and "Chaotic hypothesis").

7. Studying nonstationary evolution is much harder. The problem arises when the control parameters (force, volume, ...) change with time and the system "undergoes a process." As an example one can ask the question of how irreversible is a given irreversible process in which the initial state $\mu_0$ is a stationary state at time $t=0$, and the external parameters $\mathbf{\Phi}_0$ start changing into functions $\mathbf{\Phi}(t)$ of $t$ and tend to a limit $\mathbf{\Phi}_\infty$ as $t\to\infty$. In this case, the stationary distribution $\mu_0$ starts changing and becomes a function $\mu_t$ of $t$ which is not stationary but approaches another stationary distribution $\mu_\infty$ as $t\to\infty$. The process is, in general, irreversible and the question is how to measure its "degree of irreversibility": for simplicity we restrict attention to very special processes in which the only phenomenon is heat production because the container does not change volume and the energy also remains constants, so that the motion can be described at all times as taking place on a fixed energy surface. A natural quantity $\mathcal{I}$ associated with the evolution from an initial stationary state to a final stationary state through a change in the control parameters can be defined as follows. Consider the distribution $\mu_t$ into which $\mu_0$ evolves in time $t$, and consider also the SRB distribution $\mu_{\mathbf{\Phi}(t)}$ corresponding to the control parameters "frozen" at the value at time $t$, that is, $\mathbf{\Phi}(t)$. Let the phase-space contraction, when the forces are "frozen" at the value $\mathbf{\Phi}(t)$, be $\sigma_t(x)=\sigma(x;\mathbf{\Phi}(t))$. In general $\mu_t\neq\mu_{\mathbf{\Phi}(t)}$. Then,

$$\mathcal{I}(\{\mathbf{\Phi}(t)\},\mu_0,\mu_\infty)\stackrel{\text{def}}{=}\int_0^\infty(\mu_t(\sigma_t) - \mu_{\mathbf{\Phi}(t)}(\sigma_t))^2\,\mathrm{d}t \qquad [22]$$

can be called the degree of irreversibility of the process: it has the property that in the limit of infinitely slow evolution of $\mathbf{\Phi}(t)$, for example, if $\mathbf{\Phi}(t)=\mathbf{\Phi}_0+(1-\mathrm{e}^{-\gamma\kappa t})\Delta$ (a quasistatic evolution on timescale $\gamma^{-1}\kappa^{-1}$ from $\mathbf{\Phi}_0$ to $\mathbf{\Phi}_\infty=\mathbf{\Phi}_0+\Delta$), the irreversibility degree $\mathcal{I}_\gamma\xrightarrow[\gamma\to 0]{}0$ if (as in the case of Anosov evolutions, hence under the chaotic hypothesis) the approach to a stationary state is exponentially fast at fixed external forces $\mathbf{\Phi}$. The quantity $\mathcal{I}$ is a time scale which could be interpreted as the time needed for the process to exhibit its irreversible nature.

The entire subject is dominated by the initial insights of Onsager on classical nonequilibrium thermodynamics, which concern the properties of the infinitesimal deviations from equilibrium (i.e., averages of observables differentiated with respect to the control parameters $\mathbf{\Phi}$ and evaluated at $\mathbf{\Phi}=0$). The present efforts are devoted to studying properties at $\mathbf{\Phi}\neq 0$. In this direction, the classical theory provides certainly firm constraints (like Onsager reciprocity or Green–Kubo relations or fluctuation–dissipation theorem) but at a technical level, it gives little help to enter the terra incognita of nonequilibrium thermodynamics of stationary states.

For more details, the reader is referred to Kurchan (1998), Lebowitz and Spohn (1999), Maes (1999), Eckmann *et al.* (1999), Bonetto *et al.* (2000, 2005), Eckmann and Young (2005), Derrida *et al.* (2001), Bertini *et al.* (2001), Evans and Morriss (1990), Evans *et al.* (1993), Goldstein and Lebowitz (2004), and Gallavotti (2004).

*See also:* Adiabatic Piston; Chaos and Attractors; Ergodic Theory; Lie, Symplectic, and Poisson Groupoids and Their Lie Algebroids; Macroscopic Fluctuations and Thermodynamic Functionals; Nonequilibrium Statistical Mechanics: Dynamical Systems Approach; Quantum Dynamical Semigroups; Random Dynamical Systems.

## Further Reading


Bertini L, De Sole A, Gabrielli D, Jona G, and Landim C (2001) Fluctuations in stationary nonequilibrium states of irreversible processes. *Physical Review Letters* 87: 040601.

Bonetto F, Gallavotti G, Giuliani A, and Zamponi F (2005) Chaotic hypothesis, fluctuation theorem and singularities, mp_Ar Xiv 05-257, cond-mat/0507672.

Bonetto F, Lebowitz JL, and Rey-Bellet L (2000) Fourier's law: a challenge to theorists. In: Streater R (ed.) *Mathematical Physics 2000*, pp. 128–150. London: Imperial College Press.

de Groot S and Mazur P (1984) *Non-Equilibrium Thermodynamics*, (reprint). New York: Dover.

Derrida B, Lebowitz JL, and Speer E (2001) Free energy functional for nonequilibrium systems: an exactly solvable case. *Physical Review Letters*.

Dettman C and Morriss GP (1996) Proof of conjugate pairing for an isokinetic thermostat. *Physical Review E* 53: 5545–5549.

Eckmann JP, Pillet CA, and Rey Bellet L (1999) Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. *Communications in Mathematical Physics* 201: 657–697.

Eckmann JP and Young LS (2005) Nonequilibrium energy profiles for a class of 1D models. *Communications in Mathematical Physics*.

Evans DJ, Cohen EGD, and Morriss G (1993) Probability of second law violations in shearing steady flows. *Physical Review Letters* 70: 2401–2404.

Evans DJ and Morriss GP (1990) *Statistical Mechanics of Nonequilibrium Fluids*. New York: Academic Press.

Evans DJ and Sarman S (1993) Equivalence of thermostatted nonlinear responses. *Physical Review E* 48: 65–70.
Gallavotti G (1996) Chaotic hypothesis: Onsager reciprocity and fluctuation–dissipation theorem. *Journal of Statistical Physics* 84: 899–926.
Gallavotti G (1998) Chaotic dynamics, fluctuations, non-equilibrium ensembles. *Chaos* 8: 384–392.
Gallavotti G (1999) *Statistical Mechanics*. Berlin: Springer.
Gallavotti G (2004) Entropy production in nonequilibrium stationary states: a point of view. *Chaos* 14: 680–690.
Gallavotti G, Bonetto F, and Gentile G (2004) Aspects of the ergodic, qualitative and statistical theory of motion. pp. 1–434. Berlin: Springer.
Gallavotti G and Cohen EGD (1995) Dynamical ensembles in non-equilibrium statistical mechanics. *Physical Review Letters* 74: 2694–2697.
Gallavotti G and Ruelle D (1997) SRB states and nonequilibrium statistical mechanics close to equilibrium. *Communications in Mathematical Physics* 190: 279–285.
Goldstein S and Lebowitz J (2004) On the (Boltzmann) entropy of nonequilibrium systems. *Physica D* 193: 53–66.
Kurchan J (1998) Fluctuation theorem for stochastic dynamics. *Journal of Physics A* 31: 3719–3729.
Lebowitz JL (1993) Boltzmann's entropy and time's arrow. *Physics Today* (September): 32–38.
Lebowitz JL and Spohn H (1999) The Gallavotti–Cohen fluctuation theorem for stochastic dynamics. *Journal of Statistical Physics* 95: 333–365.
Maes C (1999) The fluctuation theorem as a Gibbs property. *Journal of Statistical Physics* 95: 367–392.
Ruelle D (1976) A measure associated with axiom A attractors. *American Journal of Mathematics* 98: 619–654.
Ruelle D (1996) Positivity of entropy production in nonequilibrium statistical mechanics. *Journal of Statistical Physics* 85: 1–25.
Ruelle D (1997) Entropy production in nonequilibrium statistical mechanics. *Communications in Mathematical Physics* 189: 365–371.
Ruelle D (1999) Smooth dynamics and new theoretical ideas in non-equilibrium statistical mechanics. *Journal of Statistical Physics* 95: 393–468.
Ruelle D (2000) A remark on the equivalence of isokinetic and isoenergetic thermostats in the thermodynamic limit. *Journal of Statistical Physics* 100: 757–763.
Sinai YaG (1972) Gibbs measures in ergodic theory. *Russian Mathematical Surveys* 166: 21–69.
Sinai YaG (1994) *Topics in Ergodic Theory*. Princeton Mathematical Series, vol. 44. Princeton: Princeton University Press.

# Nonequilibrium Statistical Mechanics: Dynamical Systems Approach

**P Buttà and C Marchioro**, Università di Roma "La Sapienza," Rome, Italy



## Time Evolution of Infinite-Particle Systems

A preliminary problem in the rigorous study of nonequilibrium statistical mechanics is to give a precise sense to the time evolution of infinitely extended systems. In fact, statistical mechanics deals with systems composed by a very large number of bodies (of the order of $10^{23}$) and studies the properties of such systems which are related to their large number of degrees of freedom. Mathematically, this aspect is stressed by introducing the so-called "thermodynamical limit," that is, by defining and analyzing systems with infinite degrees of freedom. For particle systems, the problem can be formulated in the following way. A phase point of the system is an infinite sequence $\{(\boldsymbol{x}_i, \boldsymbol{v}_i)\}_{i\in\mathbb{N}}$ of the positions and velocities of the particles, and its time evolution is characterized by the solutions of the Newton equations:

$$m\ddot{x}_i(t) = \sum_{j\in\mathbb{N}:j\neq i} \boldsymbol{F}\big(\boldsymbol{x}_i(t) - \boldsymbol{x}_j(t)\big), \quad i \in \mathbb{N} \qquad [1]$$

where $m$ is the mass of each particle, $\boldsymbol{F}(\boldsymbol{x}) = -\nabla\Phi(\boldsymbol{x})$, and $\Phi$ is a two-body potential. Equation [1] must be completed by the initial data $\{(\boldsymbol{x}_i(0), \boldsymbol{v}_i(0))\}_{i\in\mathbb{N}}$. The time evolution of a phase point implies in a natural way the time evolution of functions on the phase space, which are the observables to be compared with experiments.

The existence of a solution to eqn [1] is not obvious, because the classical theorem of existence and uniqueness for the Cauchy problem of the Newton equations depends on the number of degrees of freedom of the system. The main difficulty is that *a priori* the time evolution can bring infinitely many particles in a bounded region within a finite time, so that the right-hand side of eqn [1] becomes meaningless. Without any hypothesis on the initial conditions, this can happen, as shown by the following simple example. Consider a system of free (noninteracting) particles moving on the real line with initial conditions $x_i = i$, $v_i = -i$, $i \in \mathbb{N}$. It is clear that at time $t = 1$ all the particles are at the origin. To forbid this "collapse," we must restrict the allowed initial conditions, but we cannot be too drastic. For instance, we could surely avoid these pathologies by choosing the initial velocities uniformly bounded and the initial distribution of particles locally finite. But the set of such data is exceptional with respect to the Gibbs state (as it can be easily shown using that, at equilibrium, the velocities are independent identically distributed Gaussian variables). In conclusion, we must construct the dynamics for initial conditions which are chosen in a set sufficiently large to

be the support of states of interest from a thermodynamical point of view.

The difficulty of the problem increases with the spatial dimension $d$, as it is shown by the following example. Let the potential $\Phi$ be smooth enough and short range and assume that, initially, the velocities and the density are bounded, that is,

$$\sup_i |\boldsymbol{v}_i| < \infty, \qquad \sup_{\mu\in\mathbb{R}^d, R>1} \frac{N(X;\mu,R)}{R^d} < \infty \qquad [2]$$

where $X = \{(\boldsymbol{x}_i, \boldsymbol{v}_i)\}_{i\in\mathbb{N}}$ is the particle configuration and $N(X;\mu,R)$ is the number of particles in the ball of radius $R$, centered at $\mu$. If $V(t)$ denotes the modulus of the maximal velocity carried by the particle during the time $[0,t]$ and $X(t)$ the evolved configuration, the conservation of the particles number yields

$$N(X(t);\mu,R_0) \le N(X(0);\mu,R(t)) \le \text{const.}\, R(t)^d \qquad [3]$$

where

$$R(t) = R_0 + \int_0^t \mathrm{d}s V(s) \qquad [4]$$

On the other hand, $V(s)$ is controlled by the force, which turns out to be bounded by $\sup_\mu N(X(s);\mu,r)$, where $r > 0$ is the range of the potential. By virtue of eqns [3] and [4], we arrive at the integral inequality:

$$R(t) \le R_0 + \text{const.}\; t + \text{const.} \int_0^t \mathrm{d}s R(s)^d \qquad [5]$$

which is solvable globally in time only if $d = 1$.

In the case of interest, from a thermodynamical point of view, we also need to allow fluctuations of the density and velocities, which add further difficulties. The existence, uniqueness, and locality of the motion has been solved in dimension $d = 1$ for almost all relevant interactions (Lanford 1968, Dobrushin and Fritz 1977), and in dimension $d = 2$ for interactions not too singular at the origin (Fritz and Dobrushin 1977). (This does not cover, for instance, the hard-core interactions, where it is still an open problem to investigate whether the dynamics evolves toward a close-packing situation.) Finally, in dimension $d = 3$, the result has recently been proved only for bounded, non-negative, finite-range interactions (Caglioti *et al.* 2000).

We state the result for the three-dimensional case. Let the interaction $\Phi$ depend only on the mutual distance, be twice differentiable, positive in the origin and, for the moment, also non-negative and compactly supported. We assume that the initial data have bounded local energies and densities, with at most logarithmic divergences in velocities and densities. More precisely, we define

$$Q(X;\mu,R) = \sum_{i\in\mathbb{N}} \chi(|\boldsymbol{x}_i - \mu| \le R) \times \left[\frac{mv_i^2}{2} + \frac{1}{2}\sum_{j:j\ne i}\Phi(\boldsymbol{x}_i - \boldsymbol{x}_j) + 1\right] \qquad [6]$$

where $\chi(A)$ denotes the characteristic function of the set $A$ so that eqn [6] gives the energy and density contained in a ball centered at $\mu$ with radius $R$. Define

$$Q_\alpha(X) = \sup_\mu \sup_{R:R>\phi_\alpha(\mu)} \frac{Q(X;\mu,R)}{R^3} \qquad [7]$$

where $\alpha > 0$ and

$$\phi_\alpha(\boldsymbol{x}) \doteq \log^\alpha(e + |\boldsymbol{x}|), \quad \boldsymbol{x}\in\mathbb{R}^3 \qquad [8]$$

We denote by $\mathcal{X}_\alpha$ the set of the phase points $X$ such that $Q_\alpha(X) < \infty$. It is possible to prove that for any $\alpha \ge 1/3, \mathcal{X}_\alpha$ has full measure with respect to any Gibbs measure.

We define the partial dynamics $t \mapsto X^{(n)}(t)$ as the solutions to eqn [1] obtained by neglecting all the particles which are initially outside the ball of radius $n$ and centered at the origin.

**Theorem** *If $X \in \mathcal{X}_\alpha$ there exists a unique flow $X \to X(t) \in \mathcal{X}_{(3/2)\alpha}$ satisfying eqn [1] with $X(0) = X$. Moreover, the partial dynamics locally converges to $X(t)$ as $n \to \infty$.*

The result has been extended to bounded superstable long-range interactions. The (nontrivial) proof is based on several steps: we introduce a mollified version on the local energy and study its evolution in time under the partial dynamics. The energy conservation allows us to prove that the local energy grows at most as the cube of the maximal velocity. On the other hand, a suitable time average allows us to control the maximal velocity via the local energy in an appropriate way. The result is achieved by letting $n \to \infty$.

## Long-Time Behavior

Existence and locality of the dynamics is only a first, preliminary, step. The next and much more subtle question concerns the asymptotic (in time) and the statistical properties of the motion. Here, the main problem is the absence of simple but nontrivial models. Let us explain this point by a comparison with the situation in equilibrium statistical mechanics. In this case, even the simpler model, the free-particle system, exhibits all the relevant

thermodynamical properties of real systems away from the critical regime. In fact, the effort is often reduced to rigorously proving that the real systems away from the critical region behave as a free-particle system. The presence of the interaction is instead essential to describe phase transitions.

In the case of nonequilibrium statistical mechanics there are very few solvable models (free particles, chain of oscillators, hard-core system in one dimension), and typically they do not catch the essential properties of the real systems. For example, let us consider a system which is close to equilibrium and ask whether it converges to the corresponding Gibbs state. Two possible mechanisms usually come together: the dispersive properties of the matter (by which perturbations "escape" to infinity) and the mixing properties (by which perturbations are "spread" and disappear). The former is present also in the free-particle system, being responsible of its ergodic properties. The latter requires a deep analysis of the dynamics of interacting-particle systems and it is too difficult to be analyzed except in rare cases.

We just mention the case of systems with instantaneous interaction, which are simple enough to be studied but nevertheless exhibit a nontrivial long-time behavior. We recall in particular the famous Sinai's billiard: a particle moving freely in a two-dimensional torus except for elastic collisions with the boundary of a convex obstacle. As proved by Sinai (1970), this system has strong ergodic properties. Sinai's billiard can be proved to be equivalent to the "Lorentz gas" in which the obstacles are dislocated in a periodic way. Bunimovich and Sinai (1981) proved that when the obstacles are close enough to each other, the diffusive (weak) limit of the particle motion is the Wiener process. This remarkable result gives a rigorous derivation of Brownian motion from a Hamiltonian system.

More recently, similar questions have been investigated in the case of a charged particle subject to a constant electric field and interacting with a medium described by a particle system. Several rigorous results have been obtained on this subject. We only recall those by Boldrighini and Soloveitchik (1995, 1997). In the context of a simplified model, the asymptotic motion of the charged particle is described as a drift plus a Brownian motion, and the Einstein relation between the drift and the diffusion constant is established.

## Mean-Field Limit

The validity of any model is related to some approximation limit. In statistical mechanics, we encounter one of the most important ones, the "thermodynamical limit," used to stress the effect of large number of particles. Here we briefly discuss the "mean-field limit." For the kinetic, Boltzmann–Grad limit, *see* Boltzmann Equation (Classical and Quantum) and Kinetic Equations.

We consider $N$ particles of mass $m$ mutually interacting via the force $\boldsymbol{F}$. The equations of motion are

$$\begin{cases} m\ddot{\boldsymbol{x}}_i(t) = \sum_{j=1,\ldots,N:j\neq i} \boldsymbol{F}(\boldsymbol{x}_i(t) - \boldsymbol{x}_j(t)) \\ (\boldsymbol{x}_i(0), \dot{\boldsymbol{x}}_i(0)) = (\boldsymbol{x}_i, \boldsymbol{v}_i) \end{cases} \qquad i = 1, \ldots, N \tag{9}$$

We consider a system with $N$ very large, the mass $m$ of each particle very small, and the interaction very weak. An interesting situation arises when the quantities $N$, $m$, and $\boldsymbol{F}$ are linked by the relations

$$m = \frac{M}{N}, \qquad \boldsymbol{F} = \frac{\boldsymbol{G}}{N^2} \tag{10}$$

for some function $\boldsymbol{G}$. Of course, $M$ is the total mass of the system.

We are interested in investigating the limit $N \to \infty$. We assume that the initial data are chosen in a way that the empirical measure $N^{-1}\sum_i \delta_{\boldsymbol{x}_i}\delta_{\boldsymbol{v}_i}$ weakly converges (as $N \to \infty$) to the absolutely continuous measure $f_0(\boldsymbol{x}, \boldsymbol{v})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{v}$ with some smooth density $f_0(\boldsymbol{x}, \boldsymbol{v})$. We ask whether at some positive time $t > 0$ the empirical measure $N^{-1}\sum_i \delta_{\boldsymbol{x}_i(t)}\delta_{\boldsymbol{v}_i(t)}$ weakly converges to $f(\boldsymbol{x}, \boldsymbol{v}, t)\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{v}$ with a density $f(\boldsymbol{x}, \boldsymbol{v}, t)$ satisfying some limiting evolution equation.

Formally, it is easy to find this equation: by the Liouville theorem, a continuous medium in which each point moves under the action of an acceleration field behaves as an incompressible fluid. The continuity equation becomes

$$\begin{gathered} \partial_t f(\boldsymbol{x}, \boldsymbol{v}, t) + \boldsymbol{v} \cdot \nabla_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{v}, t) + \boldsymbol{E} \cdot \nabla_{\boldsymbol{v}} f(\boldsymbol{x}, \boldsymbol{v}, t) = 0 \\ f(\boldsymbol{x}, \boldsymbol{v}, 0) = f_0(\boldsymbol{x}, \boldsymbol{v}) \end{gathered} \tag{11}$$

where

$$\boldsymbol{E}(\boldsymbol{x}, t) = \int_{\mathbb{R}^3} \mathrm{d}\boldsymbol{y}\, \boldsymbol{G}(\boldsymbol{x} - \boldsymbol{y})\rho(\boldsymbol{y}, t) \tag{12}$$

$$\rho(\boldsymbol{x}, t) = \int_{\mathbb{R}^3} \mathrm{d}\boldsymbol{v}\, f(\boldsymbol{x}, \boldsymbol{v}, t) \tag{13}$$

This equation can be studied by following the characteristics, for which it suffices to look at the pair of functions

$$(\boldsymbol{x}, \boldsymbol{v}) \mapsto (\boldsymbol{X}(\boldsymbol{x}, \boldsymbol{v}, t), \boldsymbol{V}(\boldsymbol{x}, \boldsymbol{v}, t)), \qquad f_0(\boldsymbol{x}, \boldsymbol{v}) \mapsto f(\boldsymbol{x}, \boldsymbol{v}, t)$$

where $(\boldsymbol{x}, \boldsymbol{v}) \in \mathbb{R}^3 \times \mathbb{R}^3$ and $t \in \mathbb{R}$, solutions of

$$\begin{aligned} &\dot{\boldsymbol{X}}(\boldsymbol{x}, \boldsymbol{v}, t) = \boldsymbol{V}(\boldsymbol{x}, \boldsymbol{v}, t),\ \dot{\boldsymbol{V}}(\boldsymbol{x}, \boldsymbol{v}, t) = \boldsymbol{E}(\boldsymbol{x}, t) \\ &\boldsymbol{X}(\boldsymbol{x}, \boldsymbol{v}, 0) = \boldsymbol{x},\ \boldsymbol{V}(\boldsymbol{x}, \boldsymbol{v}, 0) = \boldsymbol{v} \\ &f(\boldsymbol{X}(\boldsymbol{x}, \boldsymbol{v}, t), \boldsymbol{V}(\boldsymbol{x}, \boldsymbol{v}, t), t) = f_0(\boldsymbol{x}, \boldsymbol{v}) \end{aligned} \qquad [14]$$

This is a weak formulation of eqn [11], in the sense that any smooth solution to eqn [11] satisfies eqn [14], but this last equation in meaningful also for nonsmooth functions. This is a weak version of the Vlasov equation and its measure solutions will play an important role in the sequel.

Equations [11]–[14] are called Vlasov equations, after Vlasov, who first introduced them in plasma physics. They have a Hamiltonian structure and conserve several quantities: the total mass, the total energy, the Liouville measure d$\boldsymbol{x}$ d$\boldsymbol{v}$, and in general each moment of this measure.

The existence and uniqueness of the solutions has been studied in many papers. Two cases have to be considered, depending on whether the total mass

$$M = \int_{\mathbb{R}^6} \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{v}\, f_0(\boldsymbol{x}, \boldsymbol{v}) \qquad [15]$$

is finite or not. We start with the first case. If the interaction $\mathbf{G}$ is bounded, the analysis is easy. On the other hand, in plasma physics one deals with the Coulomb interaction, which is singular at the origin. In this case (where eqn [11] is usually called the Vlasov–Poisson equation), existence and uniqueness can still be proved, but it is not straightforward, especially in dimension $d = 3$. The case with the complete Lorentz force, also taking into account the relativistic effect, is much more difficult.

For infinite total mass, the problem has been solved recently in three (or lower) dimensions for bounded, non-negative, finite-range interactions, and in two dimensions for singular Helmholtz interactions.

Another way to relate the Vlasov equation with the particle systems is to consider the usual transition from microscopic to macroscopic evolutions based on a separation between microscopic and macroscopic scales. Moreover, the force between the particles is due to a long-range pair interaction of the Kac type, in which the range parameter tends to infinity as the ratio $\varepsilon^{-1}$ between the macro and the micro spatial scale: $\boldsymbol{F}(\boldsymbol{x}_i - \boldsymbol{x}_j) = \varepsilon^{2d+1}\mathbf{G}(\varepsilon\boldsymbol{x}_i - \varepsilon\boldsymbol{x}_j)$. Finally, the mass of the particles is proportional to $\varepsilon^d$: $m = \varepsilon^d$. After rescaling space and time by a factor $\varepsilon$, in the macroscopic variables $(\tau, \boldsymbol{r}) = (\varepsilon t, \varepsilon \boldsymbol{x})$, the equations of motion (eqn [9]) become

$$\frac{\mathrm{d}\boldsymbol{r}_i}{\mathrm{d}\tau^2} = \sum_{j:j\neq i} \varepsilon^d \mathbf{G}(\boldsymbol{r}_i - \boldsymbol{r}_j) \qquad [16]$$

Then eqn [14] is the limiting equation as $\varepsilon \to 0$.

## Other Models

We mention another model of larger interest. We introduce it in the simplest formulation, leaving possible generalizations to the reader.

We consider an infinite chain of anharmonic oscillators, with Hamiltonian $H$ given by

$$H(q, p) = \sum_{i\in\mathbb{Z}} \left[ \frac{p_i^2}{2m} + a q_i^4 + b \sum_{j:|i-j|=1} (q_i - q_j)^2 + c q_i^2 + d \right] \qquad [17]$$

where $q_i$, $p_i \in \mathbb{R}$, $a \geq 0$, $b, c, d > 0$.

When $a = 0$, it reduces to the well-known chain of harmonic oscillators, which is integrable and widely studied in the literature.

The time evolution defined by the Hamiltonian in eqn [17] exists and it is unique for initial data chosen in a set large enough to be the support of any reasonable thermodynamic (equilibrium or nonequilibrium) state. This can be achieved by proving integral inequalities for the "Lyapunov function"

$$L(q, p) = \sup_{i\in\mathbb{Z}} \left[ \frac{p_i^2}{2m} + a q_i^4 + d \right] \frac{1}{|i| + 1}$$

It is interesting to note that uniqueness holds only in a class of data such that the position of the $i$th oscillator does not increase too much as $|i| \to \infty$. For example, besides the stationary solution $q_i(t) = 0, i \in \mathbb{Z}$, we can construct a different solution corresponding to the same initial conditions $q_i(0) = 0$, $p_i(0) = 0$, $i \in \mathbb{Z}$. In fact, by imposing $q_0(t) = t^2$ and $q_i(t) = q_{-i}(t)$, we can solve recursively the equations of motion and obtain a nonzero solution $q_i(t)$, which however increases superexponentially as $|i| \to \infty$.

The Hamiltonian dynamical systems (classical or quantum) are surely quite faithful descriptions of real systems, but they are too difficult to study. Mainly it is not known how to prove good dynamical mixing for deterministic evolutions with many degrees of freedom. Therefore, stochastic evolutions have been introduced to model the real systems. More precisely, one renounces a full description of the microscopic dynamics, introducing simplified models where the effects of the

"hidden degrees of freedom" are taken into account by adding suitable stochastic forces. Many useful results have been obtained, which show that these stochastic model systems exhibit a macroscopic behavior much closer to that observed in nature. The main criticism concerns the role of stochasticity, which in these models is introduced *ab initio*. In other words, if one believes that the statistical properties of the deterministic motion on the small scale determine the collective behavior of systems with many degrees of freedom, then these properties do have to be proved for a true understanding of nonequilibrium phenomena.

*See also:* Adiabatic Piston; Boltzmann Equation (Classical and Quantum); Fourier Law; Kinetic Equations; Nonequilibrium Statistical Mechanics (Stationary): Overview; Nonequilibrium Statistical Mechanics: Interaction between Theory and Numerical Simulations.

## Further Reading


Boldrighini C and Soloveitchik M (1995) Drift and diffusion for a mechanical system. *Probability Theory and Related Fields* 103: 349–379.

Boldrighini C and Soloveitchik M (1997) On the Einstein relation for a mechanical system. *Probability Theory and Related Fields* 107: 493–515.

Bunimovich LA and Sinai YG (1981) Statistical properties of Lorentz gas with periodic configuration of scatterers. *Communications in Mathematical Physics* 78: 479–497.

Caglioti E, Marchioro C, and Pulvirenti M (2000) Non-equilibrium dynamics of three-dimensional infinite particle systems. *Communications in Mathematical Physics* 215: 25–43.

Cornfeld IP, Fomin SV, and Sinai YG (1982) Ergodic theory. In: Artin M *et al.* (eds.) *Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences)*, vol. 245. New York: Springer.

Dobrushin RL and Fritz J (1977) Non-equilibrium dynamics of one-dimensional infinite particle systems with a hard-core interaction. *Communications in Mathematical Physics* 55: 275–292.

Fritz J and Dobrushin RL (1977) Non-equilibrium dynamics of two-dimensional infinite particle systems with a singular interaction. *Communications in Mathematical Physics* 57: 67–81.

Lanford OE III (1968) Classical mechanics of one-dimensional systems with infinitely many particles. I An existence theorem. *Communications in Mathematical Physics* 9: 176–191.

Lanford Oscar E, III (1975) Time evolution of large classical systems. In: Ehlers J, Hepp K, and Weidendmüller HA (eds.) *Dynamical Systems, Theory and Applications (Recontres, Battelle Res. Inst., Seattle, WA, 1974)*, Lecture Notes in Physics, vol. 38, pp. 1–111. Berlin: Springer.

Neunzert H (1984) An introduction to the nonlinear Boltzmann–Vlasov equation. In: Dold A and Eckmann B (eds.) *Kinetic Theories and the Boltzmann Equation (Montecatini, 1981)*, Lecture Notes in Mathematics, vol. 1048, pp. 60–110. Berlin: Springer.

Sinai TG (1970) Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards. (Russian) *Uspehi Mat. Nauk* 25: 141–192.

Spohn H (1991) Large Scale Dynamics of Interacting Particles, *Texts and Monographs in Physics*. Berlin: Springer.

Szász D (ed.) (2000) Hard Ball Systems and the Lorentz Gas. *Encyclopædia of Mathematical Sciences*, vol. 101. Berlin: Springer.


# Nonequilibrium Statistical Mechanics: Interaction between Theory and Numerical Simulations

**R Livi**, Università di Firenze, Sesto Fiorentino, Italy



## Introduction

Nonequilibrium statistical mechanics concerns a wide range of fundamental problems and applications. Perturbative methods are quite effective for approaching weakly nonlinear problems, usually relying upon effective coarse-grained equations. The attempt of obtaining a microscopic description of genuine nonlinear problems demands the combined use of theoretical methods and numerical simulations. The proprotypic case is the numerical experiment performed by Fermi, Pasta, and Ulam in 1955. As we discuss in the following section, the main questions, which had inspired this experiment, remained without an answer for a long time, while new puzzling problems emerged. Despite its apparent failure, the Fermi–Pasta–Ulam (FPU) experiment represents a remarkable example in the history of science of how a good guess may be the source of many fruitful achievements. Part of them are discussed in the section on energy relaxation in nonlinear chains, where we summarize the present understanding of the very slow relaxation mechanism, characterizing the dynamics of nonlinear chains of oscillators, like the FPU model, at low energies. Next, we report one further success of the interplay between theory and numerics, that is, the formulation of a generalized fluctuation–dissipation relation for stationary processes. Finally, we survey the main achievements concerning the study of anomalous transport properties in low-dimensional systems. In particular, we focus our attention on the heat conduction in nonlinear lattices. Lacking a general hydrodynamic theory, also in this case computer simulations and theoretical arguments have greatly

contributed to clarify the general scenario, unveiling surprising aspects, which, up to a few years ago, were completely unexpected.

## The Numerical Experiment by Fermi, Pasta, and Ulam

The impressive progress of electronic technology during World War II made possible the design of the first digital computers. The equally impressive budgets for their production and maintenance could only be justified by their employment in classified military research. Nonetheless, some of the outstanding scientists involved in these researches, like E Fermi, immediately realized the great potential of these new machines for tackling also some fundamental problems in basic science.

Fermi had in his mind a crucial and still open physical problem. In 1914 the Dutch physicist P Debye had suggested that the finiteness of thermal conductivity in crystals should be due to the nonlinear forces acting among the constituent atoms. Forty years later a microscopic theory of transport processes, including nonlinear effects, was still lacking. Actually, technical difficulties prevented a theoretical approach based on analytic methods. Numerical integration of the equations of motion by a digital machine appeared to Fermi as an effective way for tackling this problem. In collaboration with the mathematician S Ulam and the physicist J Pasta, Fermi used MANIAC 1 (a prototype digital computer installed at Los Alamos National Laboratories, USA) for integrating the dynamical equations of the simplest mathematical model of an anharmonic crystal: a chain of $N$ harmonic oscillators, coupled by nonlinear forces. Its Hamiltonian reads

$$H = \sum_{i=1}^{N} \frac{p_i^2}{2m} + \frac{\omega^2}{2}(q_{i+1} - q_i)^2 + \frac{\alpha}{3}(q_{i+1} - q_i)^3 + \frac{\beta}{4}(q_{i+1} - q_i)^4 \qquad [1]$$

where $\omega$ is the harmonic frequency, while $\alpha$ and $\beta$ are the positive coupling constants of the nonlinear terms. The integer space index $i$ labels the oscillators along the chain, while $q_i$ and $p_i$ are the displacement from the equilibrium position and the momentum of the $i$th oscillator, respectively. The potential energy is the general form taken by any nonlinear interaction potential, when expanded, up to fourth order, around its equilibrium position. This choice guarantees the boundedness of trajectories for any finite energy.

Accordingly, the model contains the minimal basic ingredients, needed for testing the conjecture about the finiteness of thermal conductivity.

The equations of motion

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial H}{\partial q_i} \qquad [2]$$

were integrated numerically by an algorithm, where space and time derivatives were approximated by proper finite-difference expressions.

The choice of the initial conditions was motivated by a further basic question concerning Fermi and his collaborators. In fact, they aimed at verifying also a common belief that had never been proved rigorously: in an isolated mechanical system with many degrees of freedom (i.e., made of a large number of oscillators), a generic nonlinear interaction among them should eventually yield equilibrium through "thermalization" of the energy. On the basis of physical intuition, nobody would object to this expectation if the mechanical system would start its evolution from an initial state very close to thermodynamic equilibrium. Nonetheless, the same should be observed by considering an initial state, where energy is supplied to a small subset of oscillatory modes of the crystal. At variance with a finite system of linear oscillators, where each initially excited mode keeps its energy constant, nonlinear terms should make the energy flow towards all oscillatory modes, until thermal equilibrium is eventually reached. Thermalization corresponds to energy equipartition among all the modes. This statement has to be interpreted in a statistical sense: the time averages of the energies contained in the modes converge to the same constant value. But if this was the case, one further fundamental aspect concerning the evolution towards thermodynamic equilibrium could be checked. In the formulation of his transport equation, L Boltzmann had conjectured that thermodynamic irreversibility can emerge from microscopic reversible dynamics (which is the case of eqns [2]). The paradoxical implication of Boltzmann's conjecture was pointed out by H Poincaré, who had proved that any isolated Hamiltonian system necessarily evolves towards an almost-recurrent dynamics. This is manifestly incompatible with the second law of thermodynamics, which implies that thermodynamic systems, in the absence of a supplied energy flux, have to evolve irreversibly towards their equilibrium state. In this perspective, the FPU numerical experiment was intended to test also if and how equilibrium is approached by a relatively large number of nonlinearly coupled oscillators, obeying the classical

laws of Newtonian mechanics. Furthermore, the measurement of the time interval needed for approaching the equilibrium state, that is, the "relaxation time" of the chain of oscillators, would have provided an indirect determination of thermal conductivity. In fact, according to elementary kinetic theory, the relaxation time, $\tau_r$, represents an estimate of the timescale of energy exchanges inside the crystal: Debye's argument predicts that thermal conductivity $\kappa$ is proportional to the specific heat at constant volume of the crystal, $C_v$, and inversely proportional to $\tau_r$, in formulas $\kappa \propto C_v/\tau_r$.

Fermi, Pasta, and Ulam considered relatively short chains, up to 64 oscillators – a size that already challenged the limits of the computational power of MANIAC 1. They imposed fixed boundary conditions (i.e., the particles at the chain boundaries interact with infinite mass walls) and the energy was initially stored just in one of the long-wavelength oscillatory modes.

A very surprising and unexpected scenario showed up. Contrary to any intuition, the energy did not flow to the higher modes, but was exchanged only among a small number of long-wavelength modes, before flowing back almost exactly to the initial state, thus yielding a recurrent behavior.

Although nonlinearities were at work, neither a tendency towards thermalization, nor a mixing rate of the energy could be identified. The dynamics exhibited regular features very close to those of an integrable system.

Fermi guessed that they were facing a very important result, but he was also quite disappointed by the difficulties in finding a convincing explanation. This lacking, he had decided not to publish the results in a scientific review, which remained confined into a Los Alamos report for almost one decade. In fact, he died in 1955, the same year of publication of the report.

The results were finally published in 1965, in a volume containing his collected papers (Fermi *et al.* 1965), and they immediately raised a renewed interest in the scientific community. Despite the failure in answering all the questions that had been raised, the FPU numerical experiment represents a crucial scientific achievement, which determined many subsequent scientific progresses. The implications about nonequilibrium will be widely discussed in the following sections. Here, we want to conclude by mentioning the important developments, inspired by the FPU experiment, that led to the discovery of solitons by Zabusky and Kruskal in 1965.

## Slow and Fast Energy Relaxation in Nonlinear Chains

The results of the FPU numerical experiment indicate that the energy initially supplied to long-wavelength oscillatory (Fourier) modes remains localized for a very long time in a small subset of long-wavelength modes. This time can be exceedingly larger than any typical timescale of the model (e.g., $\omega^{-1}$, i.e., the inverse of the harmonic frequency in [1]). An explanation of this apparently bizarre scenario has been tackled by combining theoretical approaches with numerical studies. A complete account of the many contributions in this direction being beyond the scope of this text, we shall summarize the two main lines along which this problem has been considered.

### The Resonance-Overlap Criterion

The almost-recurrent behavior of single-mode excitations studied in the FPU experiment can be explained by the resonance-overlap criterion, introduced in 1959 by the Russian scientist B Chirikov. Moreover, this criterion provides a quantitative estimate of the value of the energy density, above which the regular motion observed in the FPU experiment should be definitely lost.

In order to provide the reader with an illustration of this criterion, we have to introduce a few simple mathematical ingredients.

The Hamiltonian [1] can be rewritten in terms of linear normal Fourier coordinates, $(Q_k(t), P_k(t))$, as follows:

$$H = \frac{1}{2}\sum_k \left(P_k^2 + \omega_k^2 Q_k^2\right) + \alpha V_3(\{Q_k\}) + \beta V_4(\{Q_k\}) \qquad [3]$$

Here, we have used the shorthand notation $V_n(\{Q_k\})$ for the lengthy explicit expressions, in the new set of coordinates, of the nonlinear potentials of [1].

Without prejudice of generality, we can impose periodic boundary conditions to the FPU chain: the frequency of the $k$th normal mode is given by the expression $\omega_k = 2\sin(\pi k/N)$. The coupling constants $\alpha$ and $\beta$ control the energy exchange among the normal modes, due to nonlinear interactions.

For the sake of space, we give here a brief sketch of Chirikov's criterion for the FPU $\beta$-model (this model amounts to take $\alpha = 0$ in [3], i.e., to exclude the cubic part of the nonlinear potential).

By making reference to the initial conditions of the FPU experiment, we can consider a single excited mode, so that the Hamiltonian [3] can be

approximated by the expression in action-angle variables

$$H = H_0 + \beta H_1 \approx \omega_k J_k + \frac{\beta}{2N}(\omega_k J_k)^2 \qquad [4]$$

Here, $J_k = \omega_k Q_k^2$ is the action variable. In practice, this amounts to approximate the original Hamiltonian by the sum of the harmonic and nonlinear self-energy of the initially excited mode. In this framework, $H_0$ and $H_1$ are the unperturbed (integrable) Hamiltonian and the perturbation, respectively. Indeed, if the energy is initially attributed to mode $k$, the following relations hold: $\omega_k J_k \approx H_0 \approx E$. By the approximated Hamiltonian [4], one can compute the nonlinear correction to the linear frequency $\omega_k$, giving the renormalized frequency $\omega_k^{\mathrm{r}}$:

$$\omega_k^{\mathrm{r}} = \frac{\partial H}{\partial J_k} = \omega_k + \frac{\beta}{N}\omega_k^2 J_k = \omega_k + \Omega_k \qquad [5]$$

For $N \gg k$ one has

$$\Omega_k \approx \frac{\beta H_0 k}{N^2} \qquad [6]$$

The distance between two primary resonances, in the harmonic limit, is given by the expression

$$\Delta\omega_k = \omega_{k+1} - \omega_k \approx N^{-1} \qquad [7]$$

Consistently with [6], the last approximation is valid only for small wave number ($k \ll N$), that is, long-wavelength modes.

The "resonance overlap" criterion amounts to compare this distance with the frequency shift. In formulas:

$$\Omega_k \approx \Delta\omega_k \qquad [8]$$

This equation allows to obtain also an estimate of the "critical" energy density, $\epsilon_{\mathrm{c}}$, above which sizeable chaotic regions develop and a fast diffusion takes place in phase space:

$$\epsilon_{\mathrm{c}} = \left(\frac{H_0}{N}\right)_{\mathrm{c}} \approx \frac{1}{\beta k} \qquad [9]$$

with $k = O(1) \ll N$. Below $\epsilon_{\mathrm{c}}$, primary resonances are weakly coupled and determine a slow-relaxation process to energy equipartition. Above $\epsilon_{\mathrm{c}}$, due to "primary resonance" overlap, fast relaxation to equipartition sets in (Izrailev and Chirikov 1966).

This prediction was verified numerically later by Chirikov *et al.* (1973). The presence of a critical energy density can be tested by measuring the evolution of the finite time-averaged quantity $\bar{E}_k(t) = t^{-1}\int_0^t E_k(\tau)\mathrm{d}\tau$, where $E_k = (P_k^2 + \omega_k^2 Q_k^2)/2$ is the harmonic energy of the $k$th mode. For energy densities much smaller than $\epsilon_{\mathrm{c}}, \bar{E}_k(t)$ exhibits an extremely slow relaxation towards the equipartition condition, $\bar{E}_k = \text{constant}$. Conversely, for $\epsilon > \epsilon_{\mathrm{c}}$ such a condition is rapidly approached on a relatively short timescale. The slow relaxation below $\epsilon_{\mathrm{c}}$ can be traced back to the overlap of higher-order resonances: its typical timescale has been found to be inversely proportional to a power of the energy density (Shepelyansky 1997).

## Energy-Equipartition Thresholds

The first paper reporting evidence of the existence of an energy threshold in chains of coupled anharmonic oscillators had already been published in 1970 by Bocchieri *et al.* (1970). This pioneering numerical experiment concerned a chain of oscillators coupled through a Lennard-Jones interatomic potential. The Italian group observed an energy threshold, separating a high-energy thermalized regime from a regular dynamics regime at low energies (like the one observed by Fermi, Pasta, and Ulam). The main point raised by this experiment concerns the consequences on ergodic theory: the ordered motion observed in the low-energy regime seems to violate ergodicity, although the model is known to be chaotic at any energy.

This is quite a delicate and widely debated issue for its statistical implications. Actually, as we have mentioned in the previous section, also Fermi, Pasta, and Ulam expected that a nonlinear dynamical system, made of a large number of degrees of freedom, should naturally evolve towards equilibrium. Further confirmations to the seminal paper by Bocchieri and co-workers came from more refined numerical experiments, showing that, for sufficiently high energies, regular behaviors disappear, while equipartition among the Fourier modes sets in rapidly. Later on, the presence of the energy threshold was characterized by introducing an appropriate entropy, $S = -\sum_k p_k \ln p_k$ with $p_k = \langle E_k(t)/E \rangle$, which counts the number of effective Fourier modes involved in the dynamics: at equipartition, this entropy is maximal (Livi *et al.* 1985).

Nowadays, we know that the approach to equipartition below and above the energy threshold is a matter of timescales, which turn out to be very different in the two regimes. For instance, the analytic estimate of the maximum Lyapunov exponent $\lambda$ of the FPU $\beta$-model (Casetti *et al.* 1995) has definitely pointed out that there is a threshold value of the energy density, $\epsilon_{\mathrm{T}}$, at which its dependence on $\epsilon$ changes drastically:

$$\lambda(\epsilon) \sim \begin{cases} \epsilon^{1/4} & \text{if } \epsilon \gg \epsilon_{\mathrm{T}}; \\ \epsilon^2 & \text{if } \epsilon \ll \epsilon_{\mathrm{T}}. \end{cases} \qquad [10]$$

This implies that the typical relaxation time, that is $\lambda^{-1}$, may become exceedingly large for very small values of $\epsilon$ below $\epsilon_T$. It is worth stressing that this result holds in the thermodynamic limit, thus indicating that the presence of $\epsilon_T$ is statistically relevant.

A more controversial scenario emerges from the studies of the relaxation dynamics for specific classes of initial conditions. When a few long-wavelength modes are initially excited, regular motion may persist over times much longer than $\lambda^{-1}$ (De Luca *et al.* 1995). The excitation of small-wavelength modes yields an even more complex scenario: solitary wave dynamics is observed, followed by slow relaxation to equipartition. It is also worth mentioning that some regular features of the dynamics persist even at high energies. As we shall discuss in the section "Heat transport," such regularities still play a crucial role in determining energy transport mechanisms, although they do not affect significantly the equilibrium statistical properties of the FPU model at high energies.

## The Generalized Fluctuation–Dissipation Theorem

Another fundamental problem of nonequilibrium statistical mechanics concerns the possibility of establishing a fluctuation–dissipation theorem, generalizing the relation valid for equilibrium conditions. In fact, on this basis one might develop a large-deviation formalism, aiming at the identification of an explicit nonequilibrium statistical measure, analogous to the equilibrium Boltzmann–Gibbs measure. Recently, some relevant progresses in this direction have been made.

A crucial numerical experiment, which attracted the attention on the problem of formulating a generalized fluctuation–dissipation relation for stationary flows, was performed at the beginning of the 1990s (Evans *et al.* 1993). Stationary conditions for momentum transport were obtained in the shear flow of a fluid contained between moving walls. The reversibility of the microscopic dynamics yields the heuristic fluctuation relation:

$$\frac{1}{t}\ln\frac{\Pr(\bar{R}_t = A)}{\Pr(\bar{R}_t = -A)} = -A \qquad [11]$$

where $\Pr(\bar{R}_t = A)$ is the probability that the average entropy production rate, $\bar{R}_t$, along a trajectory segment of duration $t$, takes the value $A$. For sufficiently large values of $t$, this relation was confirmed by numerical analysis.

Gallavotti and Cohen (1995a,b) proved a theorem meant to put on a rigorous mathematical basis eqn [11], that is, the proposed extension to nonequilibrium steady states of the equilibrium fluctuation–dissipation theorem. This theorem concerns the phase-space contraction rate of the dynamics, which equals the entropy production rate in the case of particle systems, whose internal energy is a constant of the motion. The proof of the theorem is based on restrictive hypotheses, which include the existence of an average non-vanishing phase-space contraction rate, the time-reversal invariance of the dynamics and a strong form of chaos (the dynamics is assumed to be of the Anosov type, that is, smooth and uniformly hyperbolic). Nonetheless, the prediction of the theorem, that is,

$$\frac{1}{t}\ln\frac{\Pi_t(p)}{\Pi_t(-p)} = D\langle\sigma\rangle p \qquad [12]$$

is expected to hold much more generally. Here $\Pi_t(p)$ is the probability that a fluctuation variable takes the value $p$. The theorem proved by Gallavotti and Cohen states that $\Pi_t(p)$ has to satisfy the large deviation relation [12], where $\sigma$ is the average phase-space contraction rate over a trajectory segment of duration $t$ and $D$ is a suitable constant. It must be pointed out that the rigorous derivation of this relation provided strong motivations for investigating its validity and generality in many other contexts. The first numerical experiment, where almost all the constituent hypotheses of the Gallavotti–Cohen theorem were satisfied, was performed by Bonetto *et al.* (1997). They studied a Lorentz gas (massive pointlike noninteracting particles bouncing elastically on circular scatterers displaced on a regular lattice without free horizon) of charged particles moving in an uniform external electric field. Numerical simulations were found to be in very good agreement with [11] and [12] (which, in this case, refer to the same quantity). One further test of the fluctuation–dissipation relation was later performed for a different setup (Lepri *et al.* 1998). The FPU $\beta$-model is put in contact at its boundaries with thermal heat baths of different temperatures $T_+$ and $T_-(T_+ > T_-)$. Numerical simulations have been performed for sufficiently large applied thermal gradients, which guarantee sizeable effects of fluctuations, suitable for verifying a relation like [11]. It is worth noticing that many of the constituent hypotheses of the Gallavotti–Cohen theorem are not valid for this setup, but eqn [12] is still expected to hold, although in this case it does not refer to the entropy production rate. Nonetheless, the extension [11] of the fluctuation–dissipation theorem can be tested, thanks to the following useful relation,

between the heat flux $j$ and the entropy production rates, $\zeta_{\pm}$, at the chain boundaries:

$$\langle\zeta_+\rangle + \langle\zeta_-\rangle = j\left(\frac{1}{T_-} - \frac{1}{T_+}\right) \qquad [13]$$

This can be interpreted as a balance relation for the global entropy production. In fact, according to the principles of irreversible thermodynamics, the local rate of entropy production $\sigma$ in the bulk is given by

$$\sigma(x) = j\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1}{T(x)}\right) \qquad [14]$$

By integrating this equation, one straightforwardly obtains the previous one, which then applies to the entropy production from the heat baths. Careful numerical simulations show that stationary conditions are found to hold over a wide range of temperatures and gradients. Equation [13] indicates that the heat flux is equivalent to the entropy production rate, apart from a multiplicative constant which depends on the amplitude of the applied field.

Let us define the finite-time average of the global heat flux

$$J_t = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{t}\int_0^t \mathrm{d}\tau j_i(\tau) \qquad [15]$$

The normalization of this quantity can be obtained by computing the asymptotic average value

$$J_\infty = \lim_{t\to\infty} J_t \qquad [16]$$

The quantity of statistical interest is the normalized finite-time average global heat flux

$$z = \frac{J_\tau}{J_\infty} \qquad [17]$$

Accordingly, the fluctuation–dissipation relation in this case takes the form:

$$\ln\frac{P_\tau(z)}{P_\tau(-z)} = \tau z j\left(\frac{1}{T_-} - \frac{1}{T_+}\right) \qquad [18]$$

The conjecture that such a relation might be valid in this case has been confirmed by numerical analysis. It is worth stressing that, in this out-of-equilibrium setup, the probability distribution, $P_\tau(z)$, is not Gaussian and exhibits a peculiar asymmetric shape. Nonetheless, for increasing values of $\tau$, the asymmetry progressively reduces, while $P_\tau(z)$ approaches a Gaussian shape. This observation indicates that, in this case, large fluctuations deviate from the typical statistics of independent events.

It should be mentioned that generalized fluctuation–dissipation relations, like those discussed in this section, have been successfully checked in many other situations, where the hypotheses of the Gallavotti–Cohen theorem did not apply. The "robustness" of relations such as [11] and [12] indicates that a more general theory may be possible.

## Heat Transport

The validity of Debye's conjecture about the necessity of nonlinear forces for obtaining a finite heat conductivity in crystals still remained an open problem after the unsuccessful FPU numerical experiment. The setup, described in the previous section for testing the generalized fluctuation–dissipation relation in the FPU chain, can be used also for tackling the verification of this conjecture. Actually, the thermal conductivity, $\kappa$, of a chain of oscillators can be measured from the Fourier's law

$$J_Q = -\kappa\nabla T(x) \qquad [19]$$

where $J_Q$ is the heat current and $\nabla T(x)$ is the temperature gradient.

This problem was solved analytically for a chain of $N$ harmonic oscillators (Rieder *et al.* 1967). The bulk of the chain is found to reach thermal equilibrium conditions at the average temperature $T = (T_+ + T_-)/2$, corresponding to a constant temperature profile. Only at the chain boundaries the harmonic chain exhibits a steep temperature gradient. This implies that the heat current is proportional to the temperature difference, rather than to the temperature gradient, thus violating Fourier's law. Accordingly, a harmonic chain, made of $N$ oscillators, in contact with two heat reservoirs at different temperatures, exhibits anomalous transport properties and the effective thermal conductivity is found to diverge in the infinite-chain limit as $\kappa \sim N$. This peculiar behavior is a consequence of the integrability of the harmonic chain dynamics. Actually, the Fourier modes propagate with finite velocity through the harmonic chain, so that any energy injected from the hot reservoir flows ballistically to the cold one, rather than diffusing, as required for the validity of [19]. It is worth stressing that any integrable system should exhibit a similar scenario. This is the case of the equal-mass hard sphere gas in one dimension and of the Toda chain, where the harmonic potential $(\omega^2/2)(q_{i+1} - q_i)^2$ is replaced by the nonlinear expression

$$a\exp[-b(q_{i+1} - q_i)]$$

In the former case, integrability and ballistic propagation are straightforward consequences of

the conservation laws, inherent elastic collisions between hard spheres. In the latter model, the normal nonlinear modes, called "Toda solitons," are responsible for such anomalous behavior.

Debye's conjecture should be modified accordingly: nonintegrability of the equations of motion has to be invoked as a necessary property for explaining heat transport in real solids. Let us observe that the FPU model is known not to be integrable and it is expected to be a good candidate for confirming Debye's conjecture, at least in its fully chaotic regime. Careful and extended numerical simulations have shown that the FPU chain maintains anomalous properties (Lepri *et al.* 1997). In particular, the thermal conductivity, $\kappa$, is found to diverge in the infinite chain limit as

$$\kappa \sim N^{\gamma} \quad [20]$$

with $\gamma \approx 2/5$. This value agrees with independent analytic estimates (e.g., see Lepri *et al.* (2003)), although renormalization arguments indicate that one should rather find $\gamma = 1/3$ (Narayan and Ramaswamy 2002). This discrepancy could be due to the peculiar features associated with the presence of a quartic nonlinearity in the FPU problem and also to the fact that in the FPU chain heat can be transported only through longitudinal oscillations. Anyway, this is still an open problem, which requires further theoretical advances to be solved.

In a more general perspective, the main outcome of these numerical studies indicates that a power-law divergence like [20] is found in all one-dimensional nonintegrable models. This general feature must be attributed to the combined effect of low-space dimensionality, with energy and momentum conservation. In such a situation, fluctuations are strongly constrained, so that the evolution of long-wavelength hydrodynamic modes is not sufficiently damped, to be ruled by diffusion (which is a necessary ingredient for the validity of [19]). It must be stressed that these numerical investigations have strongly revived the interest for this problem. In particular, they have also stimulated new theoretical efforts for explaining the power-law divergence of transport coefficients in $d = 1$. One of the main achievements of these theoretical approaches is that the power-law divergence turns to a logarithmic one in $d = 2$, while the divergence should disappear in $d \geq 3$. Despite the difficulty of performing the necessary large-scale simulations for such systems in $d > 1$, it seems that numerics essentially agree with such predictions.

One can find normal transport properties even in $d = 1$, if suitable models are considered. For instance, momentum conservation can be broken by adding to the Hamiltonian [1] a local interaction potential, $U(q_i)$, which breaks translation invariance, thus restoring finite heat conductivity (e.g., see Casati *et al.* 1984). The exception to this case is the harmonic chain with the addition of a local harmonic potential: in this case the dynamics is still integrable and there are as many conserved quantities as degrees of freedom. A further peculiar case is represented by the rotator model in $d = 1$, which is known to be nonintegrable. Its Hamiltonian contains the interaction potential $\epsilon[1 - \cos(q_{i+1} - q_i)]$, replacing the algebraic potentials of the FPU chain. Anyway, such a Hamiltonian still guarantees momentum conservation, since the nearest-neighbor form of the interaction is maintained. Notice that, for small oscillations around the equilibrium position, also the rotator potential admits a Taylor-series expansion, whose first three terms correspond to quadratic, cubic, and quartic contributions, as in the FPU chain. Nonetheless, at variance with the FPU problem, the potential of the rotator model is bounded also from above. Numerical investigations (Giardina *et al.* 2000) have shown that for any finite energy density and for a sufficiently long finite time, some previously oscillating rotators start to rotate, due to local energy fluctuations, that allow to overtake the potential barrier. These dynamical configurations typically appear in the form of spatially localized, synchronous rotating clusters. Their time evolution is characterized by an intermittent behavior: they are eventually reabsorbed by lattice fluctuations and may reappear afterwards at other lattice positions. In this way they play the role of scattering centers for hydrodynamic modes. It must be pointed out that such a qualitative argument is not sufficient for explaining the onset of a genuine diffusive behavior, compatible with the validity of Fourier's law. A hydrodynamic theory, still to be developed, could provide a more convincing insight on these results.

It is worth concluding this section by mentioning that the overall scenario described above is confirmed by numerical studies, relying upon a different approach, based on equilibrium measurements. Actually, the linear response theory by Green and Kubo (see Kubo (1985)) provides an alternative, but essentially equivalent, definition of the thermal conductivity, according to the expression

$$\kappa = \frac{1}{K_B T^2} \lim_{t \to \infty} \lim_{N \to \infty} \frac{1}{N} \int_0^t d\tau \langle J(\tau) J(0) \rangle \quad [21]$$

The crucial quantity to be computed numerically is the heat-flux time-correlation function $C_J(\tau) = \langle J(\tau)J(0)\rangle$, where $\langle\ \rangle$ represents the thermodynamic equilibrium average. In practice, numerical simulations can be performed for a chain of $N$ oscillators in contact with boundary heat reservoirs at the same temperature $T = T_+ = T_-$. The presence of anomalous transport coefficients can be singled out by analyzing the long-time behavior of $C_J(\tau)$. It has to decay at least as $\tau^{-(1+\varepsilon)}$, with $\varepsilon > 0$ to yield a finite heat conductivity. In one-dimensional models exhibiting the power-law divergence [20] one rather finds

$$C_J(\tau) \sim \tau^{-1+\gamma} \qquad [22]$$

where the positive exponent $\gamma$ is the same appearing in [20]. This relation between space and time exponents can be easily explained, by considering that space and time variables depend linearly on each other through a proportionality constant, which is the velocity of sound in the lattice. Since $0 < \gamma < 1$, the anomalous behavior observed in out-of-equilibrium conditions is recovered.

One major problem in performing proper numerical studies concerns the control over finite-size effects, which demands a consistent increase of the integration time with the system size. This may yield very extended and expensive computations, mainly when very slow relaxation processes set in. This is the case of the low-energy regime originally studied by FPU in their pioneering computer simulations. Numerical analysis indicates that in this regime the expected behavior of $C_J(\tau)$, reported in eqn [22], sets in after a crossover time $t_c$, which increases, for decreasing energy density $\epsilon$, as $t_c \approx \epsilon^{-2}$. This seems to be compatible with the studies described earlier.

We conclude this section by pointing out that this result also contributes significantly to clarify one of the basic questions raised by the FPU numerical experiment.

*See also:* Dynamical Systems and Thermodynamics; Ergodic Theory; Fourier Law; Gravitational *N*-Body Problem (Classical); Lyapunov Exponents and Strange Attractors; Nonequilibrium Statistical Mechanics: Dynamical Systems Approach.

## Further Reading

Bocchieri P, Scotti A, Bearzi B, and Loinger A (1970) Anharmonic chain with Lennard–Jones interaction. *Physical Review A* 2: 2013.

Bonetto F, Gallavotti G, and Garrido P (1997) Chaotic principle: an experimental test. *Physica D* 105: 226.

Casati G, Ford J, Vivaldi F, and Visscher WM (1984) One-dimensional classical many-body system having a normal thermal conductivity. *Physical Review Letters* 52: 1861.

Casetti L, Livi R, and Pettini M (1995) Gaussian model for chaotic instability of Hamiltonian flows. *Physical Review Letters* 74: 375.

Chirikov BV, Izrailev FM, and Tayursky VA (1973) Numerical experiments on the statistical behaviour of dynamical systems with a few degrees of freedom. *Computational Physics Communications* 5: 11–16.

De Luca J, Lichtenberg AJ, and Lieberman MA (1995) Time scale to ergodicity in the Fermi–Pasta–Ulam system. *Chaos* 5: 283.

Evans DJ, Cohen EGD, and Morriss GP (1993) Probability of second law violations in shearing steady state. *Physical Review Letters* 71: 2401.

Fermi E, Pasta JR, and Ulam S (1965) Studies of nonlinear problems. In: *Collected Works of E. Fermi*, vol. 2, p. 978. Chicago: University of Chicago Press.

Gallavotti G and Cohen EGD (1995a) Dynamical ensembles in stationary states. *Journal of Statistical Physics* 80: 931.

Gallavotti G and Cohen EGD (1995b) Dynamical ensembles in nonequilibrium statistical mechanics. *Physical Review Letters* 74: 2694.

Giardiná C, Livi R, Politi A, and Vassalli M (2000) Finite thermal conductivity in 1D lattices. *Physical Review Letters* 84: 2144.

Izrailev FM and Chirikov BV (1966) Statistical properties of a nonlinear string. *Soviet Physics Doklady* 11: 30.

Kubo R, Toda M, and Hashitsume N (1985) *Statistical Physics II*. Berlin: Springer.

Lepri S, Livi R, and Politi A (1997) Heat conduction in chains of nonlinear oscillators. *Physical Review Letters* 78: 1896.

Lepri S, Livi R, and Politi A (1998) Energy transport in anharmonic lattices close to and far from equilibrium. *Physica D* 119: 140.

Lepri S, Livi R, and Politi A (2003) Thermal conduction in classical low-dimensional lattices. *Physics Reports* 377: 1.

Livi R, Pettini M, Ruffo S, Sparpaglione M, and Vulpiani A (1985) *Physical Review A* 31: 1039, 2740.

Narayan O and Ramaswamy S (2002) Anomalous heat conduction in one-dimensional momentum-conserving systems. *Physical Review Letters* 89: 200601.

Rieder Z, Lebowitz JL, and Lieb E (1967) Properties of a harmonic crystal in a stationary nonequilibrium state. *Journal of Mathematical Physics* 8: 1073.

Shepelyansky D (1997) Low energy chaos in the Fermi–Pasta–Ulam problem. *Nonlinearity* 10: 1331.

# Nonlinear Schrödinger Equations

**M J Ablowitz**, University of Colorado, Boulder, CO, USA
**B Prinari**, Università degli Studi di Lecce, Lecce, Italy



## Historical Background

### Ginzburg–Landau Equations

Nonlinear Schrödinger (NLS) equations have become one of the most important nonlinear systems studied in mathematics and physics. Actually, one can find the essence of NLS equations in the early work of Ginzburg and Landau (1950) and Ginzburg (1956) in their study of the macroscopic theory of superconductivity, and also of Ginzburg and Pitaevskii (1958), who subsequently investigated the theory of superfluidity.

By minimizing the free energy of a superconductor near the superconducting transition, Ginzburg and Landau arrived at what are now called the Ginzburg–Landau equations:

$$\frac{1}{2m}\left(-\mathrm{i}\hbar\nabla - \frac{e}{c}\boldsymbol{A}\right)^2\psi + \alpha\psi + \beta|\psi|^2\psi = 0 \qquad [1]$$

$$\boldsymbol{J} = -\frac{\mathrm{i}e\hbar}{mc}[\psi^*\nabla\psi - \psi\nabla\psi^*] - \frac{e^2}{mc}|\psi|^2\boldsymbol{A} \qquad [2]$$

where $\alpha, \beta$ are phenomenological parameters, $\boldsymbol{A}$ the electromagnetic vector potential, and $\psi^*$ denotes complex conjugate of $\psi$. The first equation determines the field $\psi$ based on the applied magnetic field. The second equation provides the superconducting current $\boldsymbol{J}$.

The equation describing the behavior of superfluid helium near the transition point in the stationary case derived in Ginzburg and Pitaevskii (1958) is completely analogous to eqn [1] in the phenomenological theory of superconductivity.

Equation [1] contains all the ingredients of the NLS equations which are discussed below. However, it was not until the 1960s that the wide physical importance of NLS equation became evident. The next section discusses how the NLS equation historically first appeared in the context of nonlinear optics.

### Nonlinear Optics: Self-Focusing of Optical Beams in Nonlinear Media

In the mid-1960s, Chiao *et al.* (1964) and Talanov (1964) investigated the conditions under which an electromagnetic beam can produce its own dielectric waveguide and propagate without spreading. This is a reflection of the phenomenon of self-focusing. In fact, self-focusing of optical beams may occur in materials whose dielectric constant increases with field intensity. In the general situation, a beam of uniform intensity in a dielectric broadens due to diffraction. However, the refractive index of many physically important materials (the so-called Kerr materials, such as silica) depends on the field intensity as follows:

$$n = n_0 + n_2|E|^2 + \cdots$$

If the term $n_2|E|^2$ is large enough, the critical angle for total internal reflection at the beam's boundary can be greater than the angular divergence due to diffraction; thus, spreading does not occur as a result of diffraction. As a consequence, a beam above a certain critical power level is trapped and does not spread.

In a remarkable contribution, Kelley (1965) observed, using computational methods (years before computational methods became easy to implement and, consequently, so popular) that when the self-focusing effect due to the increase in the nonlinear index is not compensated by diffraction, there is a buildup in intensity of part of the beam as a function of the distance in the direction of propagation. Consequently, the intensity of the self-focused regions tended to become "anomalously large," that is, a singularity appeared to develop.

Consider as starting equation the electromagnetic wave equation in the presence of nonlinearities derived earlier by Chiao *et al.* (1964):

$$\nabla^2\boldsymbol{E} - \frac{\epsilon_0}{c^2}\partial_t^2\,\boldsymbol{E} - \frac{\epsilon_2}{c^2}\partial_t^2\,(E^2\boldsymbol{E}) = 0 \qquad [3]$$

where $\epsilon_2|E|^2 \ll 1$. One assumes a linearly polarized wave of frequency $\omega$, propagating along the $z$-axis, so that

$$\boldsymbol{E} = \tfrac{1}{2}(\mathcal{E}\mathrm{e}^{\mathrm{i}(kz-\omega t)} + \mathrm{c.c.})\widehat{e}$$

where c.c. denotes complex conjugation, $k = \epsilon_0^{1/2}\omega/c$, the factor $\exp(\mathrm{i}kz - \omega t)$ represents the propagating part, that is, the "carrier," of the wave, and $\mathcal{E}$ is the slowly varying part. Substituting the above expression for $\boldsymbol{E}$ into eqn [3], neglecting the third-harmonic term and the term $\partial_z^2\mathcal{E}$ from $\nabla^2\boldsymbol{E}$ (assuming it to be small), yields

$$2\mathrm{i}k\partial_z\mathcal{E} + \left(\partial_x^2 + \partial_y^2\right)\mathcal{E} + \frac{3}{4}k^2\frac{\epsilon_2}{\epsilon_0}|\mathcal{E}|^2\mathcal{E} = 0 \qquad [4]$$

or, with a suitable rescaling of the dependent and independent variables ($\mathcal{E} \to \psi/((3/4)k^2\epsilon_2/\epsilon_0)^{1/2}$, $z \to 2kz$),

$$i\partial_z\psi + \nabla_\perp^2\psi + 2|\psi|^2\psi = 0 \qquad [5]$$

which is the NLS equation in standard nondimensional form.

It should be remarked here that the name NLS equation for equations of the form of [5] is natural due to the formal analogy with the Schrödinger equation in quantum mechanics:

$$i\partial_t\psi + \nabla^2\psi + V\psi = 0 \qquad [6]$$

If one sets $V = 2|\psi|^2$ in eqn [6], the result is the NLS equation. In the context of quantum mechanics, a nonlinear potential arises in the "mean-field" description of interacting particles.

Modifications of [6] also arise as mean-field descriptions of Bose–Einstein condensates which is of keen interest in physics (see Pethick and Smith (2002) and references therein). The normalized equation is

$$i\partial_t\psi - \nabla^2\psi + \left(V(x,y) + 2|\psi|^2\right)\psi = 0 \qquad [7]$$

where $V$ is an external potential. This is generally referred to as the Gross–Pitaevskii equation.

Talanov (1965) (see also Zakharov *et al.* (1971)) investigated the behavior of stationary light beams in a self-focusing nonlinear medium and found that for a purely cubic nonlinearity, "collapse" of the beam can take place. The proof that there is a singularity in eqn [5] is remarkably straightforward. This is discussed in the section "Wave collapse." In order to avoid wave collapse, other physical effects (e.g., saturable nonlinearity or dissipation) are required.

### Universal Character of the NLS Equation

It turns out that almost any dispersive, energy-preserving system gives rise, in an appropriate limit, to the NLS equation. For instance, one can derive the NLS from other physically significant equations such as the Klein–Gordon equation

$$u_{tt} - u_{xx} + u + ku^3 = 0$$

and the Korteweg–de Vries (KdV) equation

$$u_t + 6uu_x + u_{xxx} = 0$$

Actually, the NLS equation provides a "canonical" description for the envelope dynamics of a quasi-monochromatic plane wave (the carrier wave) propagating in a weakly nonlinear dispersive medium when dissipative processes are negligible.

Indeed, consider a scalar nonlinear wave equation written symbolically as

$$L(\partial_t, \nabla)u + G(u) = 0$$

where $L$ is a linear differential operator with constant coefficients and $G$ a nonlinear function of $u$ and its derivatives. For a real, small-amplitude solution of magnitude $\epsilon \ll 1$, the nonlinear effects can first be neglected, and the equation admits approximate monochromatic wave solutions

$$u = \epsilon\psi e^{i(\boldsymbol{k}\cdot\boldsymbol{x} - \omega t)} + \text{c.c.} \qquad [8]$$

with small amplitude $\epsilon|\psi|$. Substituting [8] into the linear equation, one can find that the frequency $\omega$ and the wave vector $\boldsymbol{k}$ are related by the dispersion relation

$$L(-i\omega, i\boldsymbol{k}) = 0$$

Let

$$\omega = \omega(\boldsymbol{k})$$

be one of the solutions of the previous equation. Suppose one is interested in a solution $\psi$ which is not constant, but slowly varying in space and time. This has the interpretation of $\boldsymbol{k}$ having a "sideband" wave vector and $\omega$ a "sideband" frequency. More precisely, restricting discussion, for simplicity, to the $(1+1)$-dimensional case, the slowly varying amplitude assumption corresponds to letting

$$\psi(x,t) = \psi(X,T) = \psi_0 e^{i(Kx - \Omega t)}$$

where $X = \epsilon x$ and $T = \epsilon t$. Note that $K = \epsilon k$ and $\Omega = \epsilon\omega$ are sometimes referred to as the sideband wave number and frequency, respectively, because they correspond to a deviation from the central wave number $k$ and central frequency $\omega$. Looking at these deviations from the point of view of operators, whereby $\omega \to i\partial_t$, $k \to -i\partial_x$ and $\Omega \to i\partial_T$, $K \to -i\partial_X$, one has

$$\omega_{\text{tot}} \sim \omega + \epsilon\Omega = \omega + i\epsilon\partial_T$$
$$k_{\text{tot}} \sim k + \epsilon K = k - i\epsilon\partial_X$$

Then $\omega(k)$ can be expanded in a Taylor series around the central wave number as

$$\omega(k - i\epsilon\partial_X) \sim \omega(k) - i\epsilon\omega'\partial_X - \epsilon^2\frac{\omega''}{2}\partial_X^2 + \cdots$$

Therefore,

$$\omega_{\text{tot}}(k)\psi \sim [\omega(k) + i\epsilon\partial_T]\psi$$
$$\sim \left(\omega(k) - i\epsilon\omega'\partial_X - \epsilon^2\frac{\omega''}{2}\partial_X^2\right)\psi$$

which shows that, to the leading order,

$$\mathrm{i}\epsilon\left(\frac{\partial\psi}{\partial T}+\omega'\frac{\partial\psi}{\partial X}\right)+\epsilon^2\frac{\omega''}{2}\frac{\partial^2\psi}{\partial X^2}=0 \qquad [9]$$

In the moving frame $\xi = X - \omega'(k)T$, $\tau = \epsilon T \equiv \epsilon^2 t$, eqn [9] transforms to

$$\epsilon^2\left(\mathrm{i}\psi_\tau+\frac{\omega''}{2}\psi_{\xi\xi}\right)=0$$

which is the linear Schrödinger equation with the canonical $\omega''(k)/2$ coefficient. On the other hand, if one considers rather general conservative nonlinear wave problems with leading quadratic or cubic nonlinearity, asymptotic analysis (e.g., multiple scale analysis which yields the so-called Stokes–Poincaré frequency shift) shows that a wave solution of the form

$$u(x,t)=\epsilon\psi(\tau)\mathrm{e}^{\mathrm{i}(kx-\omega t)}+\mathrm{c.c.}$$

with $\tau = \epsilon^2 t$ has $\psi(\tau)$ satisfying

$$\mathrm{i}\frac{\partial\psi}{\partial\tau}+n|\psi|^2\psi=0 \qquad [10]$$

where the constant coefficient $n$ depends on the particular equation under study. It should be remarked here that cubic nonlinearity yields an $O(\epsilon^3)$ contribution, which is balanced by a slow timescale of order $\epsilon^2$. Putting the linear and nonlinear effects together (i.e., eqns [9] and [10]) implies that an NLS equation of the form

$$\mathrm{i}\frac{\partial\psi}{\partial t}+\frac{\omega''}{2}\frac{\partial^2\psi}{\partial\xi^2}+n|\psi|^2\psi=0$$

naturally arises. The NLS equation is viewed as a "universal" equation as it generically governs the slowly varying envelope of a monochromatic wave train (see also Benney and Newell (1969)).

## Physical Applications

The nonlinear propagation of wave packets is governed by NLS-type systems in several different branches of scientific and technological applications, beyond what has been mentioned earlier. Some of these applications are discussed below.

### NLS equation in Water Waves

The NLS equation in the context of small-amplitude water waves was derived by Zakharov (1968) (infinite depth) and Benney and Roskes (1969) (finite depth). The procedure for deriving the NLS equation from the Euler–Bernoulli equations of fluid dynamics in one horizontal direction will now be discussed, under the assumption of small-amplitude waves and deep water. The interested reader can also find the details of the derivation in Ablowitz and Clarkson (2006). The relevant equations are

$$\phi_{xx}+\phi_{zz}=0, \qquad -\infty<z<\epsilon\eta(x,t) \qquad [11]$$

$$\phi_z=0, \qquad z\to-\infty \qquad [12]$$

$$\phi_t+\frac{\epsilon}{2}\left(\phi_x^2+\phi_z^2\right)+g\eta=0, \qquad z=\epsilon\eta \qquad [13]$$

$$\eta_t+\epsilon\eta_x\phi_x=\phi_z, \quad z=\epsilon\eta \qquad [14]$$

where $\phi$ is the velocity potential of an ideal (i.e., incompressible, irrotational, and inviscid) fluid, $\eta(x,t)$ is the free surface of the fluid, which is to be found, in addition to $\phi(x,z;t)$.

Equation [11] expresses the ideal nature of the fluid; the condition [12] expresses the requirement that there is no vertical flow at infinity; and eqn [13] is the Bernoulli equation of energy conservation. Finally, eqn [14] is a kinematic condition stating that no flow occurs transverse to the free surface.

At the free boundary, for small amplitudes, one can expand $\phi = \phi(t,x,\epsilon\eta)$ for $\epsilon \ll 1$ as

$$\phi=\phi(t,x,0)+\epsilon\eta\phi_z(t,x,0)+\frac{(\epsilon\eta)^2}{2}\phi_{zz}(t,x,0)+\cdots$$

and similarly for the derivatives. Second, one introduces slow temporal and spatial scales (one expects the slowly varying envelope of the wave to depend on slow variables $X=\epsilon x$, $Z=\epsilon z$, $T=\epsilon t$). Finally, because of the quadratic nonlinearity one expects second harmonics to be generated; hence,

$$\phi=\left(A\mathrm{e}^{\mathrm{i}\Theta+|k|z}+\mathrm{c.c.}\right)+\epsilon\left(A_2\mathrm{e}^{2\mathrm{i}\Theta+2|k|z}+\mathrm{c.c.}+\bar{\phi}\right)$$
$$\eta=\left(B\mathrm{e}^{\mathrm{i}\Theta}+\mathrm{c.c.}\right)+\epsilon\left(B_2\mathrm{e}^{2\mathrm{i}\Theta}+\mathrm{c.c.}+\bar{\eta}\right)$$

where $A$, $A_2$, $\bar{\phi}$ depend on $X$, $Z$, $T$ and $B$, $B_2$, $\bar{\eta}$ depend on $X, T$ ($\bar{\phi}$ and $\bar{\eta}$ are mean contributions, which are real) and $\Theta = kx - \omega t$ with the dispersion relation $\omega^2 = g|k|$. Substituting this ansatz into the equations, one obtains from the order-$\epsilon^2$ terms

$$2\mathrm{i}\omega A_\tau-\left(\frac{v_{\mathrm{g}}^2}{2\omega}A_{\xi\xi}+\frac{2k^4}{\omega}|A|^2A\right)=0 \qquad [15]$$

where $v_{\mathrm{g}} = \omega'(k) = g/2\omega$ is the group velocity and the new variables $\tau = \epsilon T$, $\xi = X - v_{\mathrm{g}}T$.

Equation [15] is the typical formulation of the (1 + 1)-dimensional NLS equation found in water wave theory for large depth.

In the section "NLS in nonlinear optics," a special solution to (a rescaled version of) eqn [15], namely a soliton solution, is discussed in the

context of nonlinear optics. It should be remarked here that the coefficients of both terms $A_{\xi\xi}$ and $|A|^2A$ have the same sign. This is necessary for a decaying soliton solution to exist (see, e.g., Lighthill (1965)).

### NLS in Nonlinear Optics

The NLS equation also describes self-compression and self-modulation of electromagnetic wave packets in weakly nonlinear media. Hasegawa and Tappert (1973a, b) first derived the NLS equation in the context of fiber optics. Light-wave propagation in a fiber is mainly affected by: (1) group velocity dispersion (GVD), that is, the frequency dependence of the group velocity originating from the refractive index of the fiber and (2) fiber nonlinearity (the so-called Kerr effect), originating from the dependence of the refractive index on the intensity of the optical pulse. In the presence of GVD and Kerr nonlinearity, the refractive index is expressed as

$$n(\omega, E) = n_0(\omega) + n_2|E|^2 \qquad [16]$$

where $\omega$ and $E$ represent the frequency and electric field of the light wave, respectively, $n_0(\omega)$ is the frequency-dependent linear refractive index, and the constant $n_2$, referred to as the Kerr coefficient, is "small" but can have significant impact since the nonlinear effects accumulate over long distances. Normally, the electric field is modulated into a slowly varying amplitude of a carrier wave:

$$E(z,t) = \mathcal{E}(z,t)\mathrm{e}^{\mathrm{i}(k_0 z - \omega_0 t)} + \mathrm{c.c.} \qquad [17]$$

where $z$ denotes the distance along the fiber, $t$ the time, $k_0 = k_0(\omega_0)$ the wave number, $\omega_0$ the frequency, and $\mathcal{E}(z,t)$ the envelope of the electromagnetic field.

A Taylor series expansion of the dispersion relation (see also the section "Universal character of the NLS equation")

$$k(\omega, E) = \frac{\omega}{c}(n_0(\omega) + n_2|E|^2)$$

around the carrier frequency $\omega = \omega_0$ yields

$$k - k_0 = k'(\omega_0)(\omega - \omega_0) + \frac{k''(\omega_0)}{2}(\omega - \omega_0)^2 + \frac{\omega_0 n_2}{c}|E|^2 \qquad [18]$$

where the prime represents derivative with respect to $\omega$ and $k_0 = k(\omega_0)$. Replacing $k - k_0$ and $\omega - \omega_0$ by their Fourier operator equivalents, $\mathrm{i}\partial_z$ and $\mathrm{i}\partial_t$ resp., using $k - k_0 = (\omega/c)n_0(\omega)$ and letting eqn [18] operate on $\mathcal{E}$ yields

$$\mathrm{i}\left(\frac{\partial \mathcal{E}}{\partial z} + k_0'(\omega_0)\frac{\partial \mathcal{E}}{\partial t}\right) - \frac{k_0''(\omega_0)}{2}\frac{\partial^2 \mathcal{E}}{\partial t^2} + \nu|\mathcal{E}|^2\mathcal{E} = 0 \qquad [19]$$

where $\nu = \omega_0 n_2 / cA_{\mathrm{eff}}$, with $A_{\mathrm{eff}}$ being the effective cross-section area of the fiber (the factor $1/A_{\mathrm{eff}}$ comes from a more detailed derivation which takes into account the finite size of the fiber; the factor $1/A_{\mathrm{eff}}$ is needed in order to account for the variation of field intensity in the cross section of the fiber). Note that $k_0'(\omega_0) = 1/v_{\mathrm{g}}$, where $v_{\mathrm{g}}$ represents the group velocity of the wave train. Introducing dimensionless variables $t' = t_{\mathrm{ret}}/t_*, z' = z/z_*, q = \mathcal{E}/\sqrt{P_*}$ yields the NLS equation

$$\mathrm{i}\frac{\partial q}{\partial z'} + \frac{\mathrm{sgn}(-k_0''(\omega_0))}{2}\frac{\partial^2 q}{\partial t'^2} + |q|^2 q = 0 \qquad [20]$$

where $t_*, P_*$ are the characteristic time and power, respectively, and $t_{\mathrm{ret}} = t - k_0'(\omega_0)z = t - z/v_g$, $z_* = 1/\nu P_*$, with the constraint that the "nonlinear length" is balanced by the linear dispersion time, that is, $t_* = (z_*| - k''(\omega_0)|)^{1/2}$.

There are two cases of physical interest depending on the sign of $k_0''$. The so-called focusing case occurs when $k_0'' < 0$; this is called "anomalous" dispersion. The defocusing case obtains when the dispersion is "normal": $k_0'' > 0$.

Now write eqn [20] in the form

$$\mathrm{i}q_t + q_{xx} \pm 2|q|^2 q = 0 \qquad [21]$$

with $\pm$ corresponding to the focusing (+) and defocusing (–) case, respectively. The focusing NLS equation admits special solutions called "bright" solitons (solutions that are traveling localized "humps"). A pure one-soliton solution in the focusing (+) case has the form

$$q(x,t) = \eta\,\mathrm{sech}[\eta(x + 2\xi t - x_0)]\,\mathrm{e}^{-\mathrm{i}\Theta} \qquad [22]$$

where $\Theta = \xi x + (\xi^2 - \eta^2)t + \Theta_0$. The parameters $\xi$ and $\eta$ are such that $\lambda = \xi/2 + \mathrm{i}\eta/2$ is an eigenvalue from the inverse scattering transform analysis.

The defocusing (–) NLS equation does not admit solitons that decay at infinity. However, it does admit soliton solutions which have a nontrivial background intensity (called "dark" and "gray" solitons). A dark-soliton solution has the form

$$q(x,t) = \eta\,\tanh(\eta x)\,\mathrm{e}^{-2\mathrm{i}\eta^2 t} \qquad [23]$$

Note that $q \to \pm\eta$ as $x \to \pm\infty$. A gray-soliton solution is

$$q(x,t) = \eta\Big[1 - B^2\,\mathrm{sech}^2(\eta B(x - x_0))\Big]^{1/2}\mathrm{e}^{\mathrm{i}\phi(x,t)} \qquad [24]$$

with

$$\phi(x,t) = -\eta^2(2-B^2)t + \eta\sqrt{1-B^2}x + \tan^{-1}\left(\frac{B\tanh(\eta Bx)}{\sqrt{1-B^2}}\right) + \phi_0$$

and $|B| < 1$. Note that as $B \to 1^-$, the gray soliton becomes a dark soliton, taking $\phi_0 = -\pi/2$.

Recall that the solutions [23] and [24] can be allowed to travel uniformly by making a Galilean transformation, that is, taking into account that if $q_1(x,t)$ is a solution of [21], then so is

$$q_2(x,t) = q_1(x - vt, t)\,\mathrm{e}^{\mathrm{i}(kx-\omega t)}$$

with $k = -v$ and $\omega = -k^2/2$.

It should also be remarked that Ablowitz *et al.* (1997) have shown that, in quadratically nonlinear optical materials, more complicated NLS-type equations arise. These equations are analogous to the finite-depth multidimensional nonlocal NLS-type systems derived in the context of water waves by Benney and Roskes (1967) and later by Davey and Stewartson (1974).

### Optical Communications

Hasegawa and Tappert (1973) first suggested using solitons as the "bit" format for transmission of information in optical fiber systems. Motivated by this, in 1980, scientists at Bell Laboratories observed solitons (described by the NLS equation) in optical fibers (Mollenauer *et al.* 1980). The development of optical amplifiers (erbium-doped amplifiers) in the mid-1980s provided a mechanism to compensate fiber loss, and this permitted the transmission of information entirely optically over long distances. With damping and amplification included (see, e.g., Hasegawa and Kodama (1995)), the NLS equation [20] takes the form

$$\mathrm{i}\frac{\partial q}{\partial z} + \frac{\mathrm{sgn}(-k_0''(\omega_0))}{2}\frac{\partial^2 q}{\partial t^2} + g(z)|q|^2 q = 0 \qquad [25]$$

where $g(z) = a_0^2 \exp(-2\Gamma z/z_a), 0 < z < z_a$, and periodically extended thereafter, and $a_0^2$ is determined by

$$< g > = \frac{1}{z_a}\int_0^{z_a} g(z/z_a)\mathrm{d}z = 1$$

with $z_a = l_a/z_*, l_a$ being the amplifier length. Remarkably, asymptotic analysis ($z_a \ll 1$) shows that, to leading order, $q(z,t)$ still satisfies the NLS equation [20].

Amplifiers, however, introduce small amounts of noise to the system, which causes the temporal position of the soliton to fluctuate (cf. Gordon and Haus (1986)) and thus limits the distance signals can be reliably transmitted to. Soliton control mechanisms were introduced in the early 1990s in order to deal with these difficulties (cf. Mecozzi *et al.* (1991) and Kodama and Hasegawa (1992)).

By the mid-1990s, the development of all optical transmission systems began to take great advantage of wavelength-division-multiplexing (WDM), that is, the simultaneous transmission of multiple signals in different frequency (or equivalently wavelength) "channels" (Hasegawa 2000). However, it was found that a serious problem affected WDM systems. Namely, the interactions of solitons traveling at different velocities cause resonant amplifier-induced instabilities in adjacent frequency channels (four-wave mixing (Mamyshev and Mollenauer 1996, Ablowitz *et al.* 1996)). In order to avoid these instabilities, researchers developed and analyzed dispersion-managed (DM) transmission systems (cf. Hasegawa (2000)). In a DM transmission system, the fiber is composed of alternating sections of positive (normal) and negative (anomalous) dispersion fibers. The (dimensionless) NLS equation that governs this phenomenon is

$$\mathrm{i}\frac{\partial q}{\partial z} + \frac{d(z)}{2}\frac{\partial^2 q}{\partial t^2} + g(z)|q|^2 q = 0 \qquad [26]$$

where $d(z)$ is usually taken to be a periodic, large, rapidly varying function of the form $d(z) = \delta_a + \Delta(z)$, with $|\Delta(z)| \gg 1$ and having zero average in the period $z_a$ (generally the same as that of the amplifier). In fact, asymptotic analysis of [26] yields a nonlocal NLS-type equation (Gabitov and Turitsyn 1996, Ablowitz and Biondini 1998). It has also been shown that eqn [26] admits various types of optical pulses, such as DM solitons (Ablowitz and Biondini 1998), and quasilinear modes (Ablowitz *et al.* 2001).

### NLS Equation in Other Settings

Many other interesting applications of the NLS equations exist in such different areas of physics as magnetic spin waves (see, e.g., the work by Zvezdin and Popkov (1983) and also by Kalinikos *et al.* (1997)), plasma physics (cf. the work by Zakharov (1972) on collapse of Langmuir waves), other areas of fluid dynamics, etc. (the interested reader can find an overview in the monograph by Ablowitz (1981)).

## Mathematical Framework

Mathematically, the NLS equation had attained broad significance since it is integrable via

inverse-scattering transform (IST), admits multisoliton solutions, has an infinite number of conserved quantities, and possesses many other interesting properties. Some of these are discussed below.

### The Inverse-Scattering Transform

The IST method allows one to linearize a large class of nonlinear evolution equations and can be considered as a nonlinear version of the Fourier transform. An essential prerequisite of IST method is the association of the nonlinear evolution equation with a pair of linear problems (Lax pair), a linear eigenvalue problem, and a second associated linear problem, such that the given equation results as a compatibility condition between them. A key research breakthrough on NLS systems appeared in 1972, in the papers of Zakharov and Shabat (1972, 1973), who first analyzed the scalar NLS equation in the form

$$\mathrm{i}q_t = q_{xx} \pm 2|q|^2 q \qquad [27]$$

($\pm$ correspond to the focusing/defocusing case, respectively) and found the associated Lax pair

$$v_x = \begin{pmatrix} -\mathrm{i}k & q \\ \mp q^* & \mathrm{i}k \end{pmatrix} v \qquad [28]$$

$$v_t = \begin{pmatrix} 2\mathrm{i}k^2 \mp \mathrm{i}|q|^2 & -2kq - \mathrm{i}q_x \\ \pm 2kq^* \mp \mathrm{i}q_x^* & -2\mathrm{i}k^2 \pm \mathrm{i}|q|^2 \end{pmatrix} v \qquad [29]$$

where $v(x,t)$ is a two-component vector. The compatibility of [28] and [29] yields eqn [27], assuming that the eigenvalue parameter $k$ is constant in time (so that [27] is often said to be isospectral).

The solution of the initial-value problem of a nonlinear evolution equation by IST proceeds in three steps, as follows:

1. *the forward problem* – the transformation of the initial data from the original "physical" variables to the transformed "scattering" variables;
2. *time dependence* – the evolution of the transformed data according to simple, explicitly solvable evolution equations; and
3. *the inverse problem* – the recovery of the evolved solution in the original variables from the evolved solution in the transformed variables.

The implementation of steps 1–3 described above is more concretely carried out as follows. The initial (Cauchy) datum $q(x,0)$ for eqn [27] is mapped into scattering data $S(k,0)$ (comprising, in general, discrete eigenvalues and associated normalization constants, and reflection coefficients) by means of eqn [28]. The data $S(k,0)$ are evolved via eqn [29] to get $S(k,t)$ at an arbitrary time $t > 0$. Finally, by employing the methods of inverse scattering, eqn [28] allows one to reconstruct the evolved solution $q(x,t)$ from $S(k,t)$.

One can easily note the "formal" resemblance to the well-known method of Fourier transform for linear differential equations.

There is considerable literature on the subject and the interested reader is encouraged to consult, for instance, some of the following references: Ablowitz and Segur (1981), Calogero and Degasperis (1982), Novikov *et al.* (1984), Ablowitz and Clarkson (1991), Ablowitz *et al.* (2004).

### Linear Stability Analysis

Consider a special solution of eqn [27] in the focusing (+sign) case: $q = a\exp(-2\mathrm{i}a^2 t)$. If this solution is perturbed as

$$q(x,t) = a\mathrm{e}^{2\mathrm{i}a^2 t}(1 + \epsilon(x,t))$$

where $|\epsilon| \ll 1$, it is found that $\epsilon$ satisfies the condition

$$\mathrm{i}\epsilon_t = \epsilon_{xx} + 2a^2(\epsilon + \epsilon^*)$$

On the periodic spatial domain $0 < x < L, \epsilon$ has the Fourier expansion

$$\epsilon(x,t) = \sum_{-\infty}^{\infty} \hat{\epsilon}_n(t)\mathrm{e}^{\mathrm{i}\mu_n x}$$

where

$$\mu_n = \frac{2\pi n}{L} \qquad [30]$$

Assuming a solution of the form

$$\begin{pmatrix} \hat{\epsilon}_n \\ \hat{\epsilon}^*_{-n} \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \mathrm{e}^{\mathrm{i}\sigma_n t}$$

one finds that $\sigma_n$ satisfies

$$\sigma_n^2 = \mu_n^2(\mu_n^2 - 4a^2) \qquad [31]$$

It then turns out that when $aL/\pi < n$ the system is unstable. Note that there are only a finite number of unstable modes (i.e., for fixed $a$, $L$, sufficiently high mode numbers $n$ will not satisfy the above inequality). In the context of water waves, this corresponds to the famous experimental and theoretical result by Benjamin and Feir that the Stoke's water wave is unstable. Later, Benney and Roskes (1969) showed that all periodic wave solutions of the generalized nonlocal NLS equation resulting from water waves in $(2+1)$-dimensions are unstable. Also, in $(2+1)$-dimensions soliton solutions are unstable to weak transverse modulations.

### Wave Collapse

The equation

$$i\psi_t + \Delta\psi + |\psi|^2\psi = 0, \qquad \boldsymbol{x} \equiv (x, y) \in \mathbb{R}^2 \qquad [32]$$

has the following conserved quantities:

$$P = \int |\psi|^2 \, \mathrm{d}\boldsymbol{x}$$

$$\boldsymbol{M} = \int \psi \nabla \psi \, \mathrm{d}\boldsymbol{x}$$

$$H = \int \left[ |\nabla\psi|^2 - \frac{1}{2}|\psi|^4 \right] \mathrm{d}\boldsymbol{x}$$

that is, mass (power), momentum, and energy (Hamiltonian) are conserved. Remarkably, Talanov (1965) showed that eqn [32] satisfies the following equation:

$$\frac{\partial^2 V}{\partial t^2} = 8H \qquad [33]$$

where

$$V = \int (x^2 + y^2)|\psi|^2 \, \mathrm{d}x \, \mathrm{d}y$$

Equation [33] is also known as the "virial" theorem. Hence, it follows that

$$V = 4Ht^2 + c_1 t + c_2$$

and if $H < 0$ initially, then a singularity in eqn [32] results since $V$ must be positive. Actually, one can further show (see, e.g., C Sulem and P L Sulem (1999), and references therein) that there exists a time $t^*$ such that

$$\int |\nabla\psi|^2 \, \mathrm{d}\boldsymbol{x}$$

becomes infinite as $t \to t^*$, which in turn implies that $\psi$ also becomes infinite as $t \to t^*$ (blowup in finite time).

Note also that for the more general equation

$$i\psi_t + \Delta_d\psi + |\psi|^{2\sigma}\psi = 0, \qquad \boldsymbol{x} \in \mathbb{R}^d$$

where $\Delta_d$ is the $d$-dimensional Laplacian, one has the following types of solutions:

- *Supercritical* ($\sigma d > 2$): the solution blows up.
- *Critical* ($\sigma d = 2$): blowup can occur or global solution can exist.
- *Subcritical* ($\sigma d < 2$): global solutions exist.

### Vector NLS Systems

In many applications vector NLS (VNLS) systems are the key governing equations. Physically, the VNLS arise under conditions similar to those described by NLS with the additional proviso that there are multiple wave trains moving nearly with the same group velocities (Roskes 1976). Importantly, VNLS also models systems where the field has more than one component. For example, in optical fibers and waveguides, the propagating electric field has two components transverse to the direction of propagation. The nondimensional system

$$i q_z^{(1)} = q_{xx}^{(1)} + 2\left(|q^{(1)}|^2 + |q^{(2)}|^2\right) q^{(1)} \qquad [34a]$$

$$i q_z^{(2)} = q_{xx}^{(2)} + 2\left(|q^{(1)}|^2 + |q^{(2)}|^2\right) q^{(2)} \qquad [34b]$$

is an asymptotic model which governs the propagation of the electric field in a waveguide, where $z$ is the normalized distance along the waveguide and $x$ a transversal spatial coordinate. It was first examined by Manakov (1974) (see also Anastassiou *et al.* (1999) and Soljačić *et al.* (2003)). Subsequently, this system was derived as a key model for light-wave propagation in optical fibers. More precisely, in optical fibers with constant birefringence (i.e., constant phase and group velocities as a function of distance) Menyuk (1987) has shown that the two polarization components of the electromagnetic field $\mathcal{E} = (u, v)^{\mathrm{T}}$ which are orthogonal to the direction of propagation, $z$, along the fiber asymptotically satisfy the following nondimensional equations (assuming anomalous dispersion):

$$i(u_z + \delta u_t) + \tfrac{1}{2} u_{tt} + (|u|^2 + \alpha|v|^2)u = 0 \qquad [35a]$$

$$i(v_z - \delta v_t) + \tfrac{1}{2} v_{tt} + (\alpha|u|^2 + |v|^2)u = 0 \qquad [35b]$$

where $\delta$ represents the group velocity "mismatch" between the $u$, $v$ components of the electromagnetic field, $\alpha$ is a constant that depends on the polarization properties of the fiber, $z$ the distance along the fiber, and $t$ a retarded temporal frame. In deriving eqn [35], it is assumed that the electromagnetic field is slowly varying (as in the scalar problem); certain nonlinear (four-wave mixing) terms are neglected in the derivation of eqn [35], because the light wave is rapidly varying due to large, but constant, linear birefringence. In this context, birefringence means that the phase and group velocities of the electromagnetic wave in each polarization component are different. In a communications environment, due to the distances involved (hundreds to thousands of kilometers), the polarization properties evolve rapidly and randomly as the light wave evolves along the propagation distance, $z$. Not only does the birefringence evolve, but it does so randomly, and on a scale much faster than the distances required for

communication transmission (birefringence polarization changes on a scale of 10–100 m). In this case, the relevant nonlinear equation is eqn [35] above, but with $\delta = 0$ and $\alpha = 1$. Indeed, this is the integrable VNLS equation first derived by Manakov (1974).

It should be remarked that the VNLS equation [34] and its generalization to an arbitrary number of components,

$$\mathrm{i}\boldsymbol{q}_t = \boldsymbol{q}_{xx} \pm 2\|\boldsymbol{q}\|^2\boldsymbol{q} \qquad [36]$$

where $\boldsymbol{q}$ is an $N$-component vector and $\|\cdot\|$ is the Euclidean norm, are integrable by the IST. One has to suitably extend the analysis discussed earlier in this article (cf. e.g., Ablowitz *et al.* (2004)).

### Discrete NLS Systems

Both the NLS and the VNLS equations discussed above admit integrable discretizations which, besides being used as the basis for constructing numerical schemes for the continuous counterparts, also have physical applications as discrete systems.

A natural discretization of NLS [27] is the following:

$$\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}q_n = \frac{1}{h^2}(q_{n+1} - 2q_n + q_{n-1}) \pm |q_n|^2(q_{n+1} + q_{n-1}) \qquad [37]$$

which is referred to as the integrable discrete NLS (IDNLS). It is an $O(h^2)$ finite-difference approximation of [27] which is integrable via the IST and has soliton solutions on the infinite lattice (Ablowitz and Ladik 1975, 1976). Note that if the nonlinear term in [37] is changed to $2|q_n|^2 q_n$, the equation, which is often called the discrete NLS (DNLS) equation, is apparently no longer integrable. It should be remarked that the (apparently nonintegrable) DNLS equation arises in many important physical contexts.

Correspondingly, one can consider the discretization of VNLS given by the following system:

$$\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{q}_n = \frac{1}{h^2}(\boldsymbol{q}_{n+1} - 2\boldsymbol{q}_n + \boldsymbol{q}_{n-1}) \pm \|\boldsymbol{q}_n\|^2(\boldsymbol{q}_{n+1} + \boldsymbol{q}_{n-1}) \qquad [38]$$

where $\boldsymbol{q}_n$ is an $N$-component vector. Equation [38] for $\boldsymbol{q}_n = \boldsymbol{q}(nh)$ in the limit $h \to 0, nh = x$ gives VNLS [36]. The discrete vector NLS system [38] is also integrable (Ablowitz *et al.* 1999, Tsuchida *et al.* 1999). The interested reader can find further details in Ablowitz *et al.* (2004).

*See also:* Boundary-Value Problems for Integrable Equations; Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Evolution Equations: Linear and Nonlinear; Ginzburg–Landau Equation; Integrable Systems and Discrete Geometry; Integrable Systems: Overview; Partial Differential Equations: Some Examples; Riemann–Hilbert Methods in Integrable Systems; Schrödinger Operators.

## Further Reading


Ablowitz MJ and Biondini G (1998) Multiscale pulse dynamics in communication systems with strong dispersion management. *Optics Letters* 23: 1668–1670.

Ablowitz MJ, Biondini G, and Blair S (1997) Multi-dimensional propagation in non-resonant $\chi^{(2)}$ materials. *Physics Letters A* 236: 520–524.

Ablowitz MJ, Biondini G, Chakravarty S, Jenkins RB, and Sauer JR (1996) Four-wave mixing in wavelength-division multiplexed soliton systems: damping and amplification. *Optics Letters* 21: 1646.

Ablowitz MJ and Clarkson PA, *Nonlinear Waves, Solitons and Symmetries*, London Mathematical Society Lecture Notes Series. Cambridge: Cambridge University Press (to be published).

Ablowitz MJ and Clarkson PA (1991) *Solitons Nonlinear Evolution Equations and Inverse Scattering*. London Mathematical Society Lecture Notes Series, vol. 149. Cambridge: Cambridge University Press.

Ablowitz MJ, Hirooka T, and Biondini G (2001) Quasi-linear optical pulses in strongly dispersion managed transmission systems. *Optics Letters* 26: 459–461.

Ablowitz MJ and Ladik JF (1975) Nonlinear differential-difference equations. *Journal of Mathematical Physics* 16: 598–603.

Ablowitz MJ and Ladik JF (1976) Nonlinear differential-difference equations and Fourier analysis. *Journal of Mathematical Physics* 17: 1011–1018.

Ablowitz MJ, Ohta Y, and Trubatch AD (1999) On discretizations of the vector nonlinear Schrödinger equation. *Physics Letters A* 253: 253–287.

Ablowitz MJ, Prinari B, and Trubatch AD (2004) *Continuous and Discrete Nonlinear Schrödinger Systems,* London Mathematical Society Lecture Notes Series, vol. 302. Cambridge: Cambridge University Press.

Ablowitz MJ and Segur H (1981) Solitons and the inverse scattering transform. *Soceity for Industrial and Applied Mathematics* 4.

Anastassiou C, Segev M, Steiglitz K, Giordmaine JA, Mitchell M *et al.* (1999) Energy-exchange interactions between colliding vector solitons. *Physical Review Letters* 83: 2332.

Benney DJ and Newell AC (1967) The propagation of nonlinear wave envelopes. *Journal of Mathematical Physics* 46: 133–139.

Benney DJ and Roskes GJ (1969) Wave instabilities. *Studies in Applied Mathematics* 48: 377–385.

Calogero F and Degasperis A (1982) *Spectral Transform and Solitons I.* Amsterdam: North-Holland.

Chiao RY, Garmire E, and Townes CH (1964) Self-trapping of optical beams. *Physical Review Letters* 15: 479–482.

Davey A and Stewartson K (1974) On three-dimensional packets of surface waves. *Proceedings of the Royal Society of London Series A* 338: 101–110.

Gabitov I and Turitsyn S (1996) Averaged pulse dynamics in a cascaded transmission system with passive dispersion compensation. *Optics Letters* 21: 327.

Ginzburg VL (1956) On the macroscopic theory of superconductivity. *Soviet Physics – JETP* 2: 589 (russian: *Journal of Experimental and Theoretical Physics* USSR 29: (1955) 748–761.).

Ginzburg VL and Pitaevskii LP (1958) On the theory of superfluidity. *Soviet Physics – JETP* 7: 858–861 (russian: *Journal of Experimental and Theoretical Physics*. USSR 34: 1240–1245).

Gordon JP and Haus HA (1986) Random walk of coherently amplified solitons in optical fiber transmission. *Optics Letters* 11: 665.
Gross EP (1961) Structure of quantized vortex. *Nuovo Cimento* 20: 454.
Hasegawa A (ed.) (2000) *Massive WDM and TDM Soliton Transmission Systems*. Dordrecht: Kluwer Academic.
Hasegawa A and Kodama Y (1995) *Solitons in Optical Communications*. Oxford: Oxford University Press.
Hasegawa A and Tappert F (1973a) Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers I. Anomalous dispersion. *Applied Physics Letters* 23: 142.
Hasegawa A and Tappert F (1973b) Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers II. Normal dispersion. *Applied Physics Letters* 23: 171.
Kalinikos BA, Kovshikov NG, and Patton CE (1997) Decay-free microwave envelope soliton pulse trains in yittrium iron garnet thin films. *Physical Review Letters* 78: 2827–2830.
Kelley PL (1965) Self-focusing of optical beams. *Physical Review Letters* 15: 1005–1008.
Kodama Y and Hasegawa A (1992) Generation of asymptotically stable optical solitons and suppression of the Gordon–Haus effect. *Optics Letters* 17: 31.
Landau LD and Ginzburg VL (1950) *Journal of Experimental and Theoretical Physics USSR* 20: 1064.
Lighthill MJ (1965) Contribution to the theory of waves in nonlinear dispersive media. *Journal of the Institute for Mathematics and Its Applications* 1: 269–306.
Mamyshev PV and Mollenauer LF (1996) Pseudo-phase-matched four-wave mixing in soliton wavelength-division multiplexed transmission. *Optics Letters* 21: 396.
Manakov SV (1974) On the theory of two-dimensional stationary self-focusing of electromagnetic waves. *Soviet Physics – JETP* 38: 248–253.
Marcuse D, Menyuk CR, and Wai PKA (1997) Applications of the Manakov-pmd equations to studies of signal propagation in fibers with randomly-varying birefringence. *Journal of Lightwave Technology* 15: 1735–1745.
Mecozzi A, Moores JD, Haus HA, and Lai Y (1991) Soliton transmission control. *Optics Letters* 16: 1841.
Menyuk CR (1987) Nonlinear pulse propagation in birefringent optical fibers. *IEEE Journal of Quantum Electronics* 23: 174–176.
Mollenauer LF, Stolen LF, and Gordon JP (1980) Experimental observation of picoseconds pulse narrowing and solitons in optical fibers. *Physical Review Letters* 45: 1095.
Novikov SP, Manakov SV, Pitaevskii LP, and Zakharov VE (1984) *Theory of Solitons. The Inverse Scattering Method*. New York: Plenum.
Pethick CJ and Smith H (2002) *Bose–Einstein Condensation in Dilute Gases*. Cambridge: Cambridge University Press.
Pitaevskii LP (1961) *Soviet Physics JETP* 13: 451 (russian: *Zhurnal Ekspermentalnoi i Teoreticheskoi Fiziki* 40: 646.).
Roskes GJ (1976) Some nonlinear multiphase interactions. *Studies in Applied Mathematics* 55: 231.
Soljačić M, Steiglitz K, Sears SM, Segev M, Jakubowski MH *et al.* (2003) Collisions of two solitons in an arbitrary number of coupled nonlinear Schrödinger equations. *Physical Review Letters* 90(25): 254102.
Sulem C and Sulem PL (1999) The Nonlinear Schrödinger Equation – Self-Focusing and Wave Collapse. *Applied Mathematical Sciences*, vol. 139. Springer.
Talanov VI (1964) *Radiophysics* 7: 254.
Talanov VI (1965) Self-focusing of wave beams in nonlinear media. *Soviet Physics – JETP Letters* 109: 138.
Tsuchida T, Ujino H, and Wadati M (1999) Integrable semi-discretization of the coupled nonlinear Schrödinger equations. *Journal of Physics A: Mathematical and General* 32: 2239–2262.
Zakharov VE (1968) Stability of periodic waves of finite amplitude on the surface of a deep fluid. *Soviet Physics Journal of Applied Mechanics and Technical Physics* 4: 190–194.
Zakharov VE (1972) Collapse of Langmuir waves. *Soviet Physics – JETP* 35: 908–914.
Zakharov VE and Shabat AB (1972) Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media. *Soviet Physics – JETP* 34: 62–69.
Zakharov VE and Shabat AB (1973) Interaction between solitons in a stable medium. *Soviet Physics – JETP* 37: 823–828.
Zakharov VE, Sobolev VV, and Synakh VC (1971) Behavior of light beams in nonlinear media. *Soviet Physics – JETP* 33: 77–81.
Zvezdin AK and Popkov AF (1983) Contribution to the nonlinear theory of magnetostatic spin waves. *Soviet Physics – JETP* 57: 350–355.

# Non-Newtonian Fluids

**C Guillopé**, Université Paris XII – Val de Marne, Créteil, France



## Introduction

The flow of a fluid, liquid or gas, is described by three conservation laws, the conserved physical quantities being the mass, the linear momentum, and the energy, and by constitutive equations. The constitutive equations are specific to each fluid, and link deformations to stresses.

A fluid is said to be Newtonian if it satisfies the simplest constitutive equation, which gives the stress tensor $\sigma$ as a linear function of the rate of deformation tensor $\boldsymbol{D} = (1/2)(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^{\mathrm{T}})$, namely

$$\sigma = (\lambda \,\mathrm{tr}\, \boldsymbol{D} - p)\boldsymbol{I} + 2\eta \boldsymbol{D} \qquad [1]$$

where $\boldsymbol{u}$ is the fluid velocity, $p$ is the hydrostatic pressure ($p \geq 0$), and $\lambda$ and $\eta$ are the Lamé viscosity coefficients of the fluid, satisfying $\eta \geq 0$ and $\lambda + 2\eta/3 \geq 0$. The superscript T designates the transpose operation, the abbreviation "tr" the trace operator of a tensor, and $\boldsymbol{I}$ the unit tensor. Water and glycerin are examples of Newtonian liquids.

Non-Newtonian fluids are fluids for which the behavior is not described by eqn [1]. Silicone oils, polymers (melted or in solution), egg yolks, and blood are examples of non-Newtonian liquids. Other examples include liquid crystals, rubbers, suspensions, paints, etc.

In the following we shall first describe flows which show Newtonian or non-Newtonian behaviors. Then we shall describe the requirements a constitutive equation needs to satisfy to be considered, introducing the notions of continuum mechanics we need. After giving the most commonly used constitutive equations, we will give a few ideas about the mathematical study of the set of equations, and their numerical study, in the particular case of viscoelastic fluids.

Numerous kinds of materials are already known to exist, and more might exist in the future. This report, however, will be limited to the most commonly materials used nowadays, which are polymers, liquid crystals and polymeric liquids crystals, and paints. Moreover, we shall only consider isothermal flows, even though temperature might be an important parameter in experiments or in industry, because in particular most theoretical or numerical studies concern isothermal problems.

Non-Newtonian fluids will always be liquids, and we shall use the terms liquid or fluid indifferently.

## Non-Newtonian Behaviors

We describe a few experiments to show how differently both types of fluids, Newtonian or non-Newtonian, might react in some experimental situations. We also give some mechanical explanation when possible.

### Shear Thinning or Shear Thickening

In a Poiseuille experiment, where a fluid flows in a tube under the action of a pressure drop, the volumetric flow rate of a Newtonian fluid is inversely proportional to the constant fluid viscosity. Under the same pressure-drop condition, a polymer melt flows much faster out of the tube, which means that there is a decreasing apparent viscosity with increasing shear rate: this is referred to as shear thinning effect. Other fluids might exhibit the opposite behavior and flow out of the tube more slowly: this is called the shear thickening effect.

### Rod Climbing

When a rotating rod is inserted in a beaker filled with a Newtonian fluid, it is observed that the liquid near the rotating rod is pushed outwards by centrifugal force and that a dip on the surface of the liquid near the rod results. On the contrary, if we make the same experiment with a polymer, the fluid climbs along the rod. Moreover, for comparable rotation speed, the difference in behaviors might be quantitatively considerable. This is explained by totally different pressure repartitions in both fluids, Newtonian or non-Newtonian: in particular, the pressure in the polymer along the rod is much larger than that along the beaker, so that this pressure difference fights the centrifugal force; this is in contrast with the situation in a Newtonian fluid.

### Extrudate Swell

If a fluid is forced to flow from a large reservoir out of a circular tube of small diameter, the swell at the exit is much larger for a polymer solution than for a Newtonian fluid. A polymer flowing out of a die might also show a delayed die well, which means that the swell is not at the exit but on the jet at a certain distance of the exit. The explanation of this phenomenon is not unique: it is due partly to memory effects (the fluid remembers its former shape, the one in the reservoir), partly to the release of normal stresses, to interfacial forces, compressibility, viscous heating, and the complicated flow near the die exit.

### Difference in Normal Stresses

In a shearing flow of a Newtonian fluid, the two normal stress differences are both zero, whereas for a polymer the first normal stress difference might be very large, the second one being nearly zero. These differences in stresses in shearing flow might be a partial answer to the extrudate swell and to rod climbing experienced by polymers.

### Presence of a Yield Stress

Some materials, when subjected to shear stress, flow only after a critical value is attained. Such fluids are referred to as Bingham fluids: some cements, slurries, paints, and biological fluids might exhibit such a behavior. It is actually a well-known property of paints: if put in large quantities on a vertical wall, the paint will flow, whereas if put as a very thin film on the same wall, the paint will not flow, but stay in place, and dry to form a nice colored covering.

### Preferred Orientation of the Particles of Fluid

Fluids with properties as above, Newtonian or non-Newtonian, are isotropic in nature, even though they are constituted of atoms, or of long chains of material. They are the same everywhere, optically,

magnetically, or electrically. Some fluids, liquid crystals, or polymeric liquid crystals in particular, have remarkable properties of nonanisotropy, being able to orient themselves, on average, along a particular direction: this is the nematic phase, which is used in many devices (screens for clocks, hand calculators, and cell phones), because the average orientation may be changed by applying an electric field. Other phases of liquid crystals include smectic A, C, and C* phases, where one sees a preferred orientation (tilted for C phases) of the fluid, and also a layer-like structure. As an example, let us mention discotic nematic liquid crystals, which are precursors for carbon-based materials, such as fibers, composites, and films, which possess excellent mechanical and thermal properties. Sails for race sailing boats are made of Kevlar, which is one of these new materials with remarkable properties.

## Modeling

The flowing fluid will be described by its (Eulerian) velocity at time $t$ and position $x$, say $u(x,t)$, for $x$ belonging to the domain of the flow $\Omega$ and the time $t$ to $\mathbb{R}_+$, by its mass density $\rho(x,t)$, its pressure $p(x,t)$ ($p > 0$ defined up to an additive constant), and its stress $\sigma(x,t)$ – which is a symmetric tensor.

The partial differential equations describing the flow are satisfied in the domain of the flow and read as follows:

$$\begin{aligned} &\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho u) = 0 \\ &\rho\left(\frac{\partial u}{\partial t} + (u \cdot \nabla)u\right) = \operatorname{div}\sigma + f \end{aligned} \qquad [2]$$

where $f$ denotes some external forces applied to the fluid. These equations describe the conservation of mass and the conservation of linear momentum. To close the system, we need a constitutive equation for the stress $\sigma$ as well as initial conditions and boundary conditions.

Moreover, most non-Newtonian fluids are practically incompressible in most regions of the flow, so that we shall only consider this case: the first equation in [2] is replaced by condition $\operatorname{div} u = 0$ in the domain of the flow.

### Notions of Continuum Mechanics

At time $t$, a body $S$ occupies a region $\Omega_t$ of the Euclidean space $\mathcal{E}_3$, called the configuration at time $t$, of the body. Points $p$ of $S$ are called material points or particles of fluids. The configuration $\Omega_t$ is assumed to be regular in the following sense: $\Omega_t$ is closed, its interior is connected and dense everywhere, its boundary is piecewise regular, $\mathcal{C}^0$ at least.

A mapping $\Phi : \Omega_0 \longrightarrow \Omega_t$ is a deformation if $\Phi$ is a bijection from $\Omega_0$ onto $\Omega_t$ and is a $\mathcal{C}^1$–diffeomorphism from the interior of $\Omega_0$ onto the interior of $\Omega_t$, with positive Jacobian.

The motion of a body $S$ is given by a set of deformations $\Pi(t,t') : \Omega_t \longrightarrow \Omega_{t'}$, satisfying

$$\Pi(t,t) = \mathrm{Id}, \quad \Pi(t'',t) = \Pi(t'',t') \circ \Pi(t',t)$$

The trajectory of the material point which is in $X$ at $t_0$ is the set

$$\{\Pi(t,t_0)(X), t \geq t_0\}$$

A body is said to be rigid if the deformation $\Pi(t,t')$ is an isometry for all times $t$ and $t'$. A material point $p$ is said to be attached to the rigid body $S$ if the body $p \cup S$ is rigid.

The motion of a fluid might be described in terms of the Lagrangian coordinates $X \in \Omega_0$ of each particle of fluid: $\Omega_0$ is called the reference configuration and is the fixed configuration occupied by the body of fluid at the time of reference, say $t_0$. The motion of the fluid might also by described in terms of the Eulerian coordinates $x = \chi(X,t)$, which represent the position of a particle at time $t$ which has position $X$ at $t_0$. The Lagrangian and Eulerian coordinates of the same particle of fluid are linked by the differential equation

$$\begin{aligned} &\dot{\chi}(X,t) = u((X,t),t), \quad \text{for } t \geq t_0 \\ &\chi(X,t_0) = X \end{aligned}$$

For defining the constitutive equations, we shall use a few tensors that we define now. The deformation gradient is defined by $\boldsymbol{F}(X,t) = \partial\chi(X,t)/\partial X$, and the right Cauchy–Green tensor by $\boldsymbol{C} = \boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}$ (also called Cauchy strain). To define relative tensors, we denote by $\chi = \chi_t(x,s)$ the position at time $s \leq t$ of the material point, which is at $x$ at time $t$. The relative tensors are defined in the following way:

- the relative deformation gradient $\boldsymbol{F}_t(s) = \nabla\chi_t(x,s)$,
- the relative right Cauchy–Green tensor $\boldsymbol{C}_t(s) = \boldsymbol{F}_t^{\mathrm{T}}(s)\boldsymbol{F}_t(s)$, and
- the relative Finger tensor $\boldsymbol{C}_t(s)^{-1}$.

Note that the rate of deformation tensor is obtained as the time derivative of the relative Cauchy strain tensor:

$$\boldsymbol{D} = \frac{1}{2}\frac{\partial \boldsymbol{C}_t(s)}{\partial s}\Big|_{s=t}$$

### Principle of Objectivity and Frame Invariance

A frame of reference is defined in the spacetime $\mathcal{E}_3 \times \mathbb{R}$ attached to the observer by giving a chronology and a system of reference. The chronology is a timescale, which will be assumed to be the same for all observers. The system of reference is a set of at least four points attached to a rigid body (this is the observer), which are not coplanar.

The constitutive equation needs to satisfy the principle of frame invariance and of frame indifference (or objectivity), which means that the equation does not depend on rigid motions of the observer. In the mathematical framework, it means that the equation has to be invariant under a change of orthonormal frame of reference $x^* = Q(t)x$, where $Q(t)$ is an orthogonal tensor: the transformed equation has to have the same expression, and also to be frame indifferent. We define a scalar quantity $\varphi$, a vector field $u$, or a tensor field $\tau$, as being frame indifferent if, under the change of variables $x^* = Q(t)x$, they satisfy the relations $\varphi(x,t) = \varphi^*(x,t)$, $u(x,t) = Q(t)^{\mathrm{T}} u^*(x^*,t)$, and $\tau(x,t) = Q(t)^{\mathrm{T}} \tau^*(x^*,t) Q(t)$, respectively.

The velocity gradient $\nabla u$ is not frame indifferent, but its symmetric part is. The vorticity, which is the antisymmetric part $W = (\nabla u - \nabla u^{\mathrm{T}})/2$ of the velocity gradient, satisfies the equation $\dot{W} = Q^{\mathrm{T}} W^* Q - Q^{\mathrm{T}} \dot{Q}$, where the dot denotes the convective derivative $\mathrm{d}/\mathrm{d}t = \partial/\partial t + (u \cdot \nabla)$.

Note that the convective derivative of a scalar function $\varphi$ is frame indifferent, which means that

$$\frac{\partial \varphi}{\partial t} + (u \cdot \nabla)\varphi = \frac{\partial \varphi^*}{\partial t} + (u^* \cdot \nabla_*)\varphi^*$$

but the convective derivative of a vector or a tensor is not frame indifferent.

It can be easily checked that the derivative

$$\frac{\mathcal{D}_0 \tau}{\mathcal{D}t} = \frac{\mathrm{d}\tau}{\mathrm{d}t} + \tau W - W\tau \qquad [3]$$

of a (frame-indifferent) tensor $\tau$ is frame indifferent, which means that

$$\frac{\mathcal{D}_0 \tau}{\mathcal{D}t} = Q^{\mathrm{T}} \frac{\mathcal{D}_0^* \tau^*}{\mathcal{D}t} Q$$

To obtain another frame-indifferent derivative of a tensor $\tau$, we need to start with the expression [3], to which we may add other terms containing frame-indifferent quantities, for example, combinations of $\tau$ and $D$. A derivative which is often considered is the Oldroyd derivative, as introduced by Oldroyd in 1958:

$$\frac{\mathcal{D}_a \tau}{\mathcal{D}t} = \frac{\mathrm{d}\tau}{\mathrm{d}t} + \tau W - W\tau - a(D\tau + \tau D) \qquad [4]$$

where $a$ is a real parameter, chosen in the interval $[-1, 1]$. (This restriction on $a$ is necessary for viscometric reasons, and obtained when simple flows, such as Couette or Poiseuille flows, are studied.)

The case $a = 1$ corresponds to the upper convected derivative, and the case $a = -1$ to the lower convected derivative. The case $a = 0$ refers to the corotational or Jaumann derivative. Derivatives corresponding to cases $a = -1, 0$, or $1$ might actually be obtained by derivating $\tau$ in a frame fixed locally to the body of fluid, and which rotates and/or deforms with the body. Moreover, we shall see that the derivatives corresponding to $a = 1$ or $-1$ have very simple integral expressions.

### Constitutive Equations

The constitutive equation of a non-Newtonian fluid is a nonlinear relationship between the stress tensor and objective variables depending on the flow, such as the pressure, the rate of deformation, frame-indifferent derivatives of such quantities, etc.

Analogously to the constitutive equation for an incompressible Newtonian fluid, we may also write the stress tensor in the form $\sigma = -pI + \tau$. The extra stress tensor $\tau$ could be either a function of objective variables, which characterize the flow, or defined by a differential equation or by an integral equation. The point here is to model the fact that the fluid might have some elasticity or some memory, or might experience, for example, yield stress or orientational properties.

**Shear dependent viscosity fluids** A very simple generalization of the incompressible Newtonian fluid consists in making the viscosity dependent on the rate of deformation tensor, $\eta = \eta(D)$. This generalization has been introduced by O A Ladyzhenskaya in 1970 and, if the function is chosen properly, this model reproduces the behavior of existing fluids, at least in certain parts of their flow. For power-law fluids, the viscosity depends on the second invariant $\mathbb{I}_{\mathrm{D}} = (1/2)\mathrm{tr}\, D^2$ of the symmetric tensor $D$ (the first invariant $\mathrm{tr}\, D$ is zero because of incompressibility), and reads as

$$\eta(D) = \eta_0 + m\mathbb{I}_{\mathrm{D}}^{n-1} \qquad [5]$$

where $\eta_0 \geq 0$, $m > 0$, and $n \geq 0$. If $n = 1$, we recover the Newtonian case, whereas for $n < 1$ this equation describes a shear thinning fluid, and for $n > 1$ a shear thickening fluid. The power law is not valid for $\mathbb{I}_{\mathrm{D}}$

close to 0, so that the Carreau–Yasuda law is preferred:

$$\frac{\eta - \eta_\infty}{\eta_0 - \eta_\infty} = \left(1 + (\lambda \mathbb{I}_D)^{2\alpha}\right)^{(n-1)/(2\alpha)} \qquad [6]$$

where $\eta_0$ is the zero-shear rate viscosity, $\eta_\infty$ is the infinite-shear rate viscosity, $\lambda$ a time constant, $n$ a dimensionless power-law index, $n \geq 0$, and $\alpha > 0$ a parameter (generally equal to 1 for a monomolecular polymer).

**Oldroyd models and related models** Oldroyd models are differential models built with one of the Oldroyd derivatives, and are very commonly used for polymer solutions or melts. The stress tensor is given as a solution of a differential equation in the following way:

$$\tau + \lambda_1 \frac{\mathcal{D}_a \tau}{\mathcal{D}t} + \boldsymbol{g}(\tau, \boldsymbol{D}) = 2\eta\left(\boldsymbol{D} + \lambda_2 \frac{\mathcal{D}_a \boldsymbol{D}}{\mathcal{D}t}\right) \qquad [7]$$

where $\lambda_1 > 0$ is a relaxation time, $\lambda_2$ is a retardation time, $0 \leq \lambda_2 < \lambda_1$, and $\boldsymbol{g}(\tau, \boldsymbol{D})$ is a tensor-valued function, constrained to certain restrictions due to objectivity, and which is at least quadratic.

The Johnson–Segalman model has $\boldsymbol{g} = 0$, and $-1 \leq a \leq 1$. Other models of differential type often suppose the parameter $a$ to be 1, because it has been noticed that with $a$ close to 1 the model is able to reproduce some experimental behavior, whereas for $a = -1$ or close to $-1$, the model does not work at all. Among the models with $a = 1$, the following ones are fairly popular: the model of Phan-Thien and Tanner has $\boldsymbol{g}(\tau, \boldsymbol{D}) = \alpha\tau \operatorname{tr}\tau$, where $\alpha$ is a constant; this model can be generalized by defining $\boldsymbol{g}(\tau, \boldsymbol{D}) = \alpha\tau^2 + \beta\tau, \alpha$ and $\beta$ being functions of the trace of $\tau$ and of its determinant; the model of Giesekus is the particular case where $\alpha$ is a constant and $\beta = 0$. The Oldroyd eight-constant model is given by

$$\boldsymbol{g}(\tau, \boldsymbol{D}) = \mu_0(\operatorname{tr}\tau)\boldsymbol{D} + \nu_1 \operatorname{tr}(\tau\boldsymbol{D})\,\boldsymbol{I} + \mu_2\boldsymbol{D}^2 + \nu_2 \operatorname{tr}(\boldsymbol{D}^2)\,\boldsymbol{I}$$

where $\mu_0, \nu_1, \mu_2$, and $\nu_2$ are constants.

In [7], the limit case $\lambda_2 = 0$ corresponds to Maxwell's type models, where there is no Newtonian viscosity, while the case $\lambda_2 > 0$ corresponds to the Jeffreys' type models. The cases where $a = 1$ and $\boldsymbol{g} = 0$, are often considered in mathematical or numerical studies: this is the upper convected Maxwell (UCM) model for $\lambda_2 = 0$, and the Oldroyd B model for $\lambda_2 > 0$.

The parameters $\lambda_1, \lambda_2$, and $\eta$ might also depend on $\mathbb{I}_D$: such a model where the upper convected derivative $(a = 1)$ is chosen is referred to as the White–Metzner model, and reads as follows:

$$\tau + \lambda_{\mathbb{I}} \frac{\mathcal{D}_1 \tau}{\mathcal{D}t} = 2\left(\eta_{\mathbb{I}}\boldsymbol{D} + \eta_\infty\left(\boldsymbol{D} + \lambda_{\mathbb{I}} \frac{\mathcal{D}_1 \boldsymbol{D}}{\mathcal{D}t}\right)\right)$$

where $\eta_\infty$ is also the Newtonian viscosity.

**Integral equations** Other constitutive equations for viscoelastic fluids include integral equations. Actually, some differential equations have integral counterparts: this is the case for the differential equations associated with the upper or lower convected frame-indifferent derivatives. For the upper convected derivative $(a = 1)$, the extra stress is given by the integral expression

$$\tau(x, t) = 2\eta\frac{\lambda_2}{\lambda_1}\boldsymbol{D}(x, t) + 2\eta\frac{\lambda_1 - \lambda_2}{\lambda_1^2} \times \int_{-\infty}^{t} \mathrm{e}^{-(t-s)/\lambda_1}(\nabla_X x)\,\boldsymbol{D}(X, s)(\nabla_X x)^{\mathrm{T}}\,\mathrm{d}s$$

where $X$ is the position, at time $s$, of the point which is at $x$ at time $t$. A similar expression might be obtained for the lower convected derivative.

A very common integral equation is the K–BKZ equation (introduced independently by Kaye and Bernstein, Kearsley, and Zapas in 1962–63). In a simplified form, the extra-stress tensor is given as the integral of a combination of the relative Cauchy strain tensor $\boldsymbol{C}_t$ and its inverse:

$$\tau(x, t) = 2\int_{-\infty}^{t} G(t - s)\left[\frac{\partial\mathcal{W}(I_1, I_2)}{\partial I_1}\boldsymbol{C}_t^{-1}(s) - \frac{\partial\mathcal{W}(I_1, I_2)}{\partial I_2}\boldsymbol{C}_t(s)\right]\mathrm{d}s$$

where $I_1 = \operatorname{tr}\boldsymbol{C}_t^{-1}(s)$ and $I_2 = \operatorname{tr}\boldsymbol{C}_t(s)$. The function $G$ is a given kernel, and $\mathcal{W}$ a given scalar potential. The upper convected Maxwell model is obtained from the K–BKZ model by setting $\mathcal{W}(I_1, I_2) = I_1$ and $G(s) = (\lambda_1\lambda_2/2)\,\mathrm{e}^{-\lambda_1 s}$.

**Models issued from kinetic theories or micro–macro models** Polymeric fluids could also be modeled by coupling a macroscopic viewpoint – the one of continuum mechanics, as described above – and a microscopic viewpoint. A polymer is, in general, made of long chains of molecules. Rather than trying to represent the polymer behavior by a sophisticated constitutive equation, one describes the mean behavior of the molecules by using their microscopic description.

To take an example, we consider a dilute solution of polymer, where each chain of polymer is modeled as a collection of dumbbells, each of them consisting of two beads connected by a spring. The configuration of the spring, namely its length and orientation, is described by a random vector field $Q \in \mathbb{R}^3$. The dumbbells are convected and stretched by the flow.

The probability $\psi(x, Q, t)\,\mathrm{d}Q$ of finding a dumbbell with a configuration $Q$ at $(x, t)$ is governed by a Fokker–Planck equation:

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} + \mathrm{div}_Q((\nabla u)Q\psi) = \frac{2}{\zeta}\ \mathrm{div}_Q((\nabla_Q \mathcal{W})\psi) + \frac{2kT}{\zeta}\Delta_Q\psi$$

where $\zeta$ is the friction coefficient of the dumbbell beads, $T$ the temperature, and $k$ the Planck constant, and $\mathcal{W}$ the spring potential. The extra stress is given by the constitutive equation

$$\tau = \lambda \int (\nabla_Q \mathcal{W} \otimes Q)\psi(x, Q, t)\,\mathrm{d}Q$$

The simplest potential is the linear one (also called Hookean potential) $\mathcal{W}(Q) = H|Q|^2$, where $|Q|$ is the length of $Q$, and $H$ the elasticity constant. In fact, in the case of the Hookean potential, this set of equations is equivalent to the Oldroyd B model. Another potential corresponds to finitely extendable nonlinear elastic (FENE) chain of dumbbells,

$$\mathcal{W}(Q) = -\frac{HQ_0^2}{2}\log\left(1 - \frac{|Q|^2}{Q_0^2}\right)$$

for $|Q| \leq Q_0$, and gives the FENE model, for which there is no macroscopic constitutive equation known.

We have only made here a short incursion in these micro–macro models: research is in progress, both analytical and numerical (Öttinger 1996, Suen *et al.* 2002, Keunings 2004).

**Liquid crystals and polymeric liquid crystals** As an example, we present the constitutive equations for a uniaxial nematic liquid crystal.

In the theory of Leslie and Ericksen, established in the 1960s and the 1970s, the stress tensor is given as a function of the orientation unit vector $n$, through the Oseen–Frank elastic energy,

$$2\mathcal{W}(n, \nabla n) = \kappa_1(\mathrm{div} n)^2 + \kappa_2(n \cdot \mathrm{curl}\, n)^2 + \kappa_3|n \times \mathrm{curl}\, n|^2$$

where $\kappa_1 > 0, \kappa_2 > 0$, and $\kappa_3 > 0$ are the three basic modes (splay, twist, and bend, respectively). The extra stress tensor is precisely given by the relation

$$\begin{aligned}\tau = & -(\nabla n)^{\mathrm{T}}\frac{\partial \mathcal{W}}{\partial \nabla n} + \alpha_1(n \cdot \boldsymbol{D}n)n \otimes n \\ & + \alpha_2 \boldsymbol{N} \otimes n + \alpha_3 n \otimes \boldsymbol{N} \\ & + \alpha_4 \boldsymbol{D} + \alpha_5 \boldsymbol{D}n \otimes n + \alpha_6 n \otimes\end{aligned}$$

where $\boldsymbol{N} = \dot{n} - \boldsymbol{W}n$ is the corotational derivative of the director, and $\alpha_i, i = 1, \ldots, 6$, the six Leslie viscosity coefficients.

The director satisfies a differential equation derived from continuum mechanics,

$$\rho_1 \ddot{n} = G + g + \mathrm{div}\pi$$

where $\rho_1$ is the moment of inertia per unit volume, $G$ the external director body force (torque per unit volume), $\pi$ the director stress tensor, and $g$ the intrinsic director body force. Precisely,

$$g = \lambda n - (\nabla n)\beta - \frac{\partial \mathcal{W}}{\partial n} - \gamma_1 \boldsymbol{N} - \gamma_2 \boldsymbol{D}n$$

$$\pi = n \otimes \beta + \frac{\partial \mathcal{W}}{\partial \nabla n}$$

where $\beta$ is a Lagrange multiplier vector, and $\lambda = -\gamma_2/\gamma_1$ is the reactive parameter, with $\gamma_1 = \alpha_3 - \alpha_2$ the rotational viscosity, and $\gamma_2 = \alpha_6 - \alpha_5 = \alpha_3 + \alpha_2$ the irrotational torque coefficient.

Polymeric liquid crystals might have other variables entering in the modeling, such as order parameters, order tensors, etc.

Because of the complexity of modeling, most studies concern either very simple flows, such as Couette or Poiseuille flows, or steady flows, or flows for which the coefficients satisfy specific relationships.

Reports about earlier studies, theoretical as well as numerical, can be found in Coron *et al.* (1991), and references therein. The study of polymeric liquid crystals, or of the smectic phase of liquid crystals is at its very early stage and one could look into it in specialized journals, such as the *Journal of Non-Newtonian Fluid Mechanics*, or *see* Liquid Crystals.

**Yield stress fluids** Bingham materials have the property of flowing only when the stress magnitude is greater than a critical value, and being a solid otherwise. Precisely, in the simplest and the most widely used model, the Bingham model, the extra stress tensor $\tau$ is given by the relations

$$\begin{aligned}&\tau = 2\eta\boldsymbol{D} + \tau_*\frac{\boldsymbol{D}}{\mathbb{I}_{\mathrm{D}}} \quad &\text{if } \mathbb{I}_{\mathrm{D}} \neq 0 \\ &|\tau| \leq \tau_* \quad &\text{if } \mathbb{I}_{\mathrm{D}} = 0\end{aligned} \qquad [8]$$

where $\tau_* > 0$ is the yield limit. The Bingham model is generalized in taking the viscosity $\eta$ to be a function of the shear stress: $\eta$ is given by the relation

$$\eta = 1 + 2\left(\frac{\tau_*}{\mathbb{I}_{\mathrm{D}}}\right)^{1/2}$$

for the Casson law, and by the power law [5] for the Herschel–Bulkley model.

The mathematical study was started by Duvaut and Lions (1976), and regained interest recently (Malek and Rajagopal 2005), especially in relation with other recent studies in polymeric liquids.

## Theoretical and Numerical Problems for Viscoelastic Flows

The mathematical study of viscoelastic fluid flows amounts to studying systems of partial differential equations, which all include either the incompressible Euler equation or the incompressible Navier–Stokes equation as particular cases. In particular, it means that the results obtained from such a study are similar to the ones obtained for Euler or Navier–Stokes equations, and, because of the complexity of the system, the results are expected to be qualitatively as good, actually more often less good, than for these equations. For example, the existence of weak three-dimensional solutions to the Navier–Stokes system is known, while for non-Newtonian flows, this result will be true only in very specific cases. Moreoever, when a result is not known for the Navier–Stokes problem, such as the uniqueness of solution for all data in a three-dimensional problem, there is no hope something similar could be proved for non-Newtonian fluid flows.

As an example, we consider the case of Johnson–Segalman fluids, which are described by constitutive equation [7] with $\boldsymbol{g} = 0$. Recall that the limit case $\lambda_2 = 0$ corresponds to the purely elastic case, and $\lambda_2 = \lambda_1$ to the purely Newtonian case. Equation [7] is coupled with the equations of motion:

$$\begin{aligned} \rho \frac{\mathrm{d}u}{\mathrm{d}t} + \nabla p &= \operatorname{div} \tau + f \\ \operatorname{div} u &= 0 \end{aligned} \qquad [9]$$

Equations [7] and [9] have to be solved in the domain of the flow, which might be the whole space $\mathbb{R}^3$ (or $\mathbb{R}$ or $\mathbb{R}^2$ in case of symmetries), or a domain $\Omega$, bounded or not, in $\mathbb{R}^n$, $n = 1, 2,$ or 3. These equations are supplemented by appropriate boundary conditions and initial conditions for the velocity $u$ and the extra stress $\tau$ (no boundary condition on $\tau$ is needed if the homogeneous nonslip boundary condition $u = 0$ is chosen).

We first make explicit the Newtonian contribution to the stress by setting $\tau = \tau^{\mathrm{s}} + \tau^{\mathrm{p}}$ and $\tau^{\mathrm{s}} = 2\eta_{\mathrm{s}}\boldsymbol{D}$. The differential equation for $\tau^{\mathrm{p}}$ is then

$$\tau^{\mathrm{p}} + \lambda_1 \frac{\mathcal{D}_a \tau^{\mathrm{p}}}{\mathcal{D}t} = 2\eta_{\mathrm{p}}\boldsymbol{D}$$

where $\eta_{\mathrm{p}} = (1 - \lambda_2/\lambda_1)\eta$ is the so-called polymeric viscosity, $\eta_{\mathrm{s}} = (\lambda_2/\lambda_1)\eta$ the so-called Newtonian viscosity (or solvent viscosity).

We then use nondimensional variables, so as to make explicit the characteristic parameters, which the flow depends on. The non-Newtonian fluid considered in this model will always be homogeneous: its density $\rho$ is a constant independent of $x$ and $t$. The dimensional variables are now asterisked. We define quantities which are characteristic of the flow: a length $L$, a velocity magnitude $U$, a stress magnitude $T$, a force magnitude $F$, and a pressure $P$. We operate the change of variables and functions $x = x^*/L$, $u = u^*/U$, $t = Ut^*/L$, and also introduce the nondimensional functions

$$\tau = \frac{\tau^*}{T}, \quad p = \frac{p^*}{P}, \quad f = \frac{f^*}{F}$$

After choosing the parameters $T$, $P$, and $F$ in an appropriate way, namely $T = P = \eta U/L$, and $F = \eta U/L^2$, we obtain the following system

$$\begin{aligned} Re \frac{\mathrm{d}u}{\mathrm{d}t} + \nabla p &= (1 - \omega)\Delta u + \operatorname{div} \tau + f \\ \operatorname{div} u &= 0 \\ We \frac{\mathcal{D}_a \tau}{\mathcal{D}t} + \tau &= 2\omega \boldsymbol{D} \end{aligned} \qquad [10]$$

Here the three nondimensional parameters which the flow depends on are the usual Reynolds number $Re = \rho_0 UL/\eta$ and two other numbers: the Weissenberg number $We = \lambda U/L$ measures the elasticity per unit time (sometimes also called the Deborah number), and the parameter $\omega = \eta_{\mathrm{p}}/\eta$ is the ratio of elastic viscosity to total viscosity ($\omega = 0$ corresponds to the Newtonian case, while $\omega = 1$ corresponds to the purely elastic case).

System [10] couples a transport equation (the equation for the stress $\tau$), and either a Navier–Stokes type equation when $\omega < 1$, or a Euler type equation when $\omega = 1$ (for the velocity $u$). This system is not hyperbolic, parabolic, or elliptic.

Maxwell's type models ($\omega = 1$) display two striking phenomena. First, the Cauchy problem (with initial data) can present Hadamard instabilities, that is, instabilities to short waves. It means, in particular, that the Cauchy problem is not well posed in any good class but analytic. Moreover, the partial differential system for Maxwell's type steady flows may experience a change of type, analogous to the situation in gas dynamics, if the "Mach number" $Re\,We$ is larger than 1.

Jeffreys' type models ($\omega < 1$), because of the presence of a Newtonian viscosity, do not exhibit such phenomenon, but their study does not enter in

the theory of parabolic equations either, the type of the system being composite.

Problems of interest for rheologists, as well as for mathematicians, include in particular the high Weissenberg asymptotics, the high Weissenberg boundary layers, the singularity of flows near a reentrant corner, and the stability of flows.

We give a few details about stability questions. Instabilities are seen in experimental extrusion of melted polymers from a pipe: melt fracture designates different phenomena appearing at different stages of the experiment, when the speed of the extrusion is increased, such as sharkskin instability, slight distortions of the extrudate, large distortions and wavyness of the extrudate. One may distinguish two kinds of instabilities. First, constitutive instabilities are associated with nonmonotonicity of constitutive functions and loss of evolutionary property of the equations of motion. Other kinds of instabilities are close to classical hydrodynamic instabilities at increasing *Re*. Note that in viscoelastic flows the *Re* is usually very small, and might even be set to zero in some studies.

Other mathematical questions for system [10] include existence of weak solutions (for the very special case of Oldroyd model with the Jaumann derivative where ($a=0$) in [5]), existence of regular solutions defined on some time interval, depending on the magnitude of the data, and existence of regular solutions for all times. Other studies concern the existence, uniqueness, and stability of steady solutions. Another field of study is the numerical simulation of such flows.

In summary, there have been numerous computations made in the field of steady or unsteady viscoelastic fluids, and especially models using continuum mechanics. Standard test problems include the cavity-driven flow, flows inside a 4 : 1 contraction, extrusion flows, flows between eccentric cylinders, and flows in "wiggly" pipes. As mentioned already, the type of the sytem of partial differential equations is composite, neither elliptic nor hyperbolic. The numerical codes have to take into account the precise nature of the set of partial differential equations, so as to be able to obtain noncatastrophic results. One of the main challenges has been to deal with the high-*We* problem: with increasing *We*, the results would become totally incoherent, and the numerical algorithms would diverge.

Nowadays, with the power of computers increasing, molecular simulations of flows are proposed, using the macro–micro modeling mentioned above. Also, simulations of flows of colloidal suspensions and reacting flows have been undertaken with success.

*See also:* Compressible Flows: Mathematical Theory; Fluid Mechanics: Numerical Methods; Incompressible Euler Equations: Mathematical Theory; Interfaces and Multicomponent Fluids; Inviscid Flows; Liquid Crystals; Newtonian Fluids and Thermohydraulics; Partial Differential Equations: Some Examples; Stability of Flows; Stochastic Hydrodynamics; Viscous Incompressible Fluids: Mathematical Theory.

## Further Reading

Baranger J, Guillopé C, and Saut J-C (1996) Mathematical analysis of differential models for viscoelastic fluids. In: Piau J-M and Agassant J-F (eds.) *Rheology for Polymer Melt Processing*, pp. 199–236. Amsterdam: Elsevier.

Bird RB, Armstrong RC, and Hassager O (1987a) *Dynamics of Polymeric Liquids. Volume 1: Fluid Mechanics*, 2nd edn. New York: Wiley-Interscience.

Bird RB, Curtiss CF, Armstrong RC, and Hassager O (1987b) *Dynamics of Polymeric Liquids. Volume 2. Kinetic Theory*, 2nd edn. New York: Wiley-Interscience.

Coron J-M, Ghidaglia J-M, and Hélein F (eds.) (1991) *Nematics: Mathematical and Physical Aspects*, NATO Series, Series C Mathematical and Physical Sciences, Dordrecht: Kluwer.

de Gennes P-G and Prost P (1995) *The Physics of Liquid Crystals*, The International Series of Monographs on Physics, vol. 83, 2nd edn. Oxford: Oxford University Press.

Doi M and Edwards SF (1988) *The Theory of Polymer Dynamics*, The International Series of Monographs on Physics, vol. 73. Oxford: Oxford University Press.

Duvaut G and Lions J-L (1976) *Inequalities in Mechanics and Physics*, Springer Grundlehren, vol. 219. Berlin: Springer.

Joseph DD (1990) *Fluid Dynamics of Viscoelastic Liquids*, Applied Math Sciences, vol. 84. Berlin: Springer.

Keunings R (2004) Micro–macro methods for the multiscale simulation of viscoelastic flow using molecular models of kinetic theory. In: Binding DM and Walters K (eds.) *Rheology Reviews 2004*. British Society of Rheology, pp. 67–98.

Malek J and Rajagopal KR (2005) Mathematical issues concerning the Navier–Stokes equations and some of their generalizations. In: Dafermos C and Feireisl E (eds.) *Handbook of Differential Equations. Evolutionary Equations: Volume 2*. Amsterdam: North-Holland.

Öttinger HC (1996) *Stochastic Processes in Polymeric Fluids*. Berlin: Springer.

Renardy M (2000) Current issues in non-Newtonian flows: a mathematical perspective. *Journal of Non-Newtonian Fluid Mechanics* 90: 243–259.

Renardy M, Hrusa WJ, and Nohel JA (1987) *Mathematical Problems in Viscoelasticity*, Pitman Monographs and Surveys in Pure and Applied Mathematics, vol. 35. Harlow: Longman Scientific and Technical.

Suen JKC, Joo YL, and Armstrong RC (2002) Molecular orientation effects in viscoelasticity. *Annual Review of Fluid Mechanics* 34: 417–444.

Tanner RI and Walters K (1998) *Rheology: An Historical Perspective*, Rheology Series, vol. 9. Amsterdam: Elsevier.

# Nonperturbative and Topological Aspects of Gauge Theory

**R W Jackiw**, Massachusetts Institute of Technology, Cambridge, MA, USA



## Introduction

Classical fields that enter a classical field theory provide a mapping from the "base" manifold on which they are defined (space or spacetime) to a "target" space over which they range. The base and target spaces, as well as the map, may possess nontrivial topological features, which affect the fixed-time description and the temporal evolution of the fields, thereby influencing the physical reality that these fields describe. Quantum fields of a quantum field theory are operator-valued distributions whose relevant topological properties are obscure. Nevertheless, topological features of the corresponding classical fields are important in the quantum theory for a variety of reasons: (1) Quantized fields can undergo local (spacetime-dependent) transformations (gauge transformations, coordinate diffeomorphisms) that involve classical functions whose topological properties determine the allowed quantum field theoretic structures. (2) One formulation of the quantum field theory uses a functional integral over classical fields, and classical topological features become relevant. (3) Semiclassical (WKB) approximations to the quantum theory rely on classical dynamics, and again classical topology plays a role in the analysis.

Topological effects of gauge fields in quantum theory were first appreciated by Dirac in his study of the quantum mechanics for (hypothetical) magnetic point monopoles. Although here one is not dealing with a field theory, the consequences of his analysis contain many features that were later encountered in field theory models.

The Lorentz equations of motion for a charged ($e$) massive ($M$) particle in a monopole magnetic field ($\boldsymbol{B}=m\boldsymbol{r}/r^3$) are unexceptional,

$$\dot{\boldsymbol{r}} = \frac{\boldsymbol{p}}{M} \qquad [1a]$$

$$\dot{\boldsymbol{p}} = \frac{e}{M}\boldsymbol{p}\times\boldsymbol{B} \quad (c=1) \qquad [1b]$$

and completely determine classical dynamics. But knowledge of the Lagrangian $L$ and of the action $I$ – the time integral of $L$: $I=\int \mathrm{d}t\, L$ – is further needed for quantum mechanics, either in its functional integral formulation or in its Hamiltonian formulation, which requires the canonical momentum $\boldsymbol{\pi} \equiv \partial L/\partial \dot{\boldsymbol{r}}$. The Lorentz-force action is expressed in terms of the vector potential $\boldsymbol{A}$, $\boldsymbol{B}=\nabla\times\boldsymbol{A}$: $I_{\mathrm{Lorentz}}=e\int \mathrm{d}t\,\dot{\boldsymbol{r}}\cdot\boldsymbol{A}=e\int \mathrm{d}\boldsymbol{r}\cdot\boldsymbol{A}$. The magnetic monopole vector potential is necessarily singular because $\nabla\cdot\boldsymbol{B}=4\pi m\delta^3(\boldsymbol{r})\neq 0$. The singularity (Dirac string) can be moved, but not removed, by gauge transformations, which also are singular, and do not leave the Lorentz action invariant. Noninvariance of the action can be tolerated provided its change is an integral multiple of $2\pi$, since the functional integrand involves $\exp(\mathrm{i}I)$ (with $\hbar=1$). The quantal requirement, which is not seen in the equations of motion, is met when

$$eg = N/2 \qquad [2]$$

The topological background to this (Dirac) quantization condition is the fact that $\Pi_1(\mathrm{U}(1))$ is the group of integers, that is, the map of the unit circle into the gauge group, here U(1), is classified by integers.

Further analysis shows that only point magnetic sources can be incorporated in particle quantum mechanics, which is governed by the particle Hamiltonian $H=\boldsymbol{p}^2/2M$ (magnetic fields do no work and are not seen in $H$). Quantum Lorentz equations are regained by commutation with $H$: $\dot{\boldsymbol{r}}=\mathrm{i}[H,\boldsymbol{r}]$, $\dot{\boldsymbol{p}}=\mathrm{i}[H,\boldsymbol{p}]$, provided

$$\mathrm{i}[r^i, r^j] = 0 \qquad [3a]$$

$$\mathrm{i}[p^i, r^j] = \delta^{ij} \qquad [3b]$$

$$\mathrm{i}[p^i, p^j] = -e\varepsilon^{ijk}B^k \qquad [3c]$$

But [3c] implies that the Jacobi identity is obstructed by magnetic sources $\nabla\cdot\boldsymbol{B}\neq 0$.

$$\tfrac{1}{2}\varepsilon^{ijk}[p^i,[p^j,p^k]] = e\nabla\cdot\boldsymbol{B} \qquad [4]$$

This obstruction is better understood by examining the unitary operator $U(\boldsymbol{a})\equiv\exp(\mathrm{i}\boldsymbol{a}\cdot\boldsymbol{p})$, which according to [3b] implements finite translations of $\boldsymbol{r}$ by $\boldsymbol{a}$. The commutator algebra [3] and the failure of the Jacobi identity [4] imply that these operators do not associate. Rather one finds

$$U(\boldsymbol{a}_1)(U(\boldsymbol{a}_2)U(\boldsymbol{a}_3)) = \mathrm{e}^{\mathrm{i}\Phi}(U(\boldsymbol{a}_1)U(\boldsymbol{a}_2))U(\boldsymbol{a}_3) \qquad [5]$$

where $\Phi=e\int \mathrm{d}^3x\,\nabla\cdot\boldsymbol{B}$ is the total flux emerging from the tetrahedron formed from the three vectors $\boldsymbol{a}_i$ with vertex at $\boldsymbol{r}$ (see Figure 1). But quantum mechanics realized by linear operators acting on a Hilbert space requires that operator multiplication

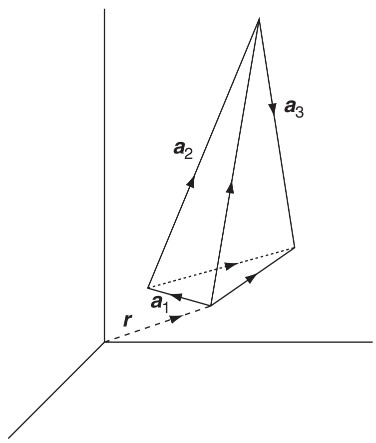



**Figure 1** Tetrahedron pierced by magnetic flux that obstructs associativity.

be associative. This can be achieved, in spite of [5], provided $\Phi$ is an integral multiple of $2\pi$, hence invisible in the exponent. This then needs that (1) $\nabla \cdot \boldsymbol{B}$ be localized at points, so that the volume integral of $\nabla \cdot \boldsymbol{B}$ retain integrality for arbitrary $\boldsymbol{a}_i$ and (2) the strengths of the localized poles obey Dirac quantization. The points at which $\nabla \cdot \boldsymbol{B}$ is localized can now be removed from the manifold and the Jacobi identity is regained. The above argument, which rederives Dirac's quantization, makes no reference to gauge variance of magnetic potentials.

In the remainder we shall discuss related phenomena for selected gauge field theories in four, three, and two dimensions that describe actual physical events occurring in nature. We shall encounter in generalized form, analogs to the above quantum mechanical system.

Some definitions and notational conventions: Nonabelian gauge potentials $A^a_\mu$ carry a spacetime index ($\mu$) (metric tensor $g_{\mu\nu} = \mathrm{diag}(1, -1, \ldots)$) and an adjoint group index (a). When contracted with anti-Hermitian matrices $T_a$ that represent the group's Lie algebra (structure constants $f_{ab}{}^c$)

$$[T_a, T_b] = f_{ab}{}^c T_c \qquad [6]$$

they become Lie algebra-valued.

$$A_\mu \equiv A^a_\mu T_a \qquad [7]$$

Gauge transformations transform $A_\mu$ by group elements $U$:

$$A_\mu \to A^U_\mu \equiv U^{-1} A_\mu U + U^{-1} \partial_\mu U \qquad [8a]$$

For infinitesimal gauge transformations, $U \approx I + \lambda$, $\lambda \equiv \lambda^a T_a$; this leads to the covariant derivative $D_\mu$:

$$\begin{aligned} A_\mu &\to A_\mu + \partial_\mu \lambda + [A_\mu, \lambda] \equiv A_\mu + D_\mu \lambda \\ A^a_\mu &\to A^a_\mu + \partial_\mu \lambda^a + f_{bc}{}^a A^b_\mu \lambda^c \equiv A^a_\mu + (D_\mu \lambda)^a \end{aligned} \qquad [8b]$$

(In a quantum field theory, $A_\mu$ becomes an operator but the gauge transformations $U, \lambda$ remain c-number functions.) The field strength $F_{\mu\nu}$ given by

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu] \qquad [9a]$$

is also given by

$$[D_\mu, D_\nu] \ldots = [F_{\mu\nu}, \ldots] \qquad [9b]$$

(coupling strength $g$ has been scaled to unity). The definition [9] implies the Bianchi identity

$$D_\mu F_{\nu\omega} + D_\omega F_{\mu\nu} + D_\nu F_{\omega\mu} = 0 \qquad [10]$$

Here $F_{\mu\nu}$ is gauge covariant

$$F_{\mu\nu} \to F_{\mu\nu}{}^U = U^{-1} F_{\mu\nu} U \qquad [11a]$$

or, infinitesimally,

$$F_{\mu\nu} \to F_{\mu\nu} + [F_{\mu\nu}, \lambda] \qquad [11b]$$

In the gauge invariant Yang–Mills action $I_{\mathrm{YM}}$, the Yang–Mills Lagrange density $\mathcal{L}_{\mathrm{YM}}$ is integrated over the base space,

$$\begin{aligned} \mathcal{L}_{\mathrm{YM}} &= \tfrac{1}{2} \operatorname{tr} F^{\mu\nu} F_{\mu\nu} \\ I_{\mathrm{YM}} &= \int \mathcal{L}_{\mathrm{YM}} = \frac{1}{2} \int \operatorname{tr} F^{\mu\nu} F_{\mu\nu} \end{aligned} \qquad [12]$$

The trace is evaluated with the convention

$$\operatorname{tr} T_a T_b = -\tfrac{1}{2} \delta_{ab} \qquad [13]$$

and henceforth there is no distinction between upper and lower group indices. The Euler–Lagrange condition for stationarizing $I_{\mathrm{YM}}$ gives the Yang–Mills equation

$$D_\mu F^{\mu\nu} = 0 \qquad [14a]$$

Should sources $J^\mu$ be present, [14a] becomes

$$D_\mu F^{\mu\nu} = J^\nu \qquad [14b]$$

and $J^\mu$ must be covariantly conserved:

$$\begin{aligned} D_\nu J^\nu = D_\nu D_\mu F^{\mu\nu} &= -\tfrac{1}{2} [D_\mu, D_\nu] F^{\mu\nu} \\ &= -\tfrac{1}{2} [F_{\mu\nu}, F^{\mu\nu}] = 0 \end{aligned} \qquad [15]$$

All this is a nonabelian generalization of familiar Maxwell electrodynamics.

## Gauge Theories in Four Dimensions

Gauge theories in four-dimensional spacetime are at the heart of the standard particle physics model. Their topological features have physical consequences and merit careful study.

### Yang–Mills Theory

In four dimensions, we define nonabelian electric $\boldsymbol{E}^a$ and magnetic $\boldsymbol{B}^a$ fields,

$$E^{ia} = F^a_{0i}, \qquad B^{ia} = -\tfrac{1}{2}\varepsilon^{ijk}F^a_{jk} \qquad [16]$$

Canonical analysis and quantization is carried out in the Weyl gauge ($A^a_0 = 0$), where the Lagrangian and Hamiltonian (energy) densities read

$$\mathcal{L}_{\rm YM} = \tfrac{1}{2}(\boldsymbol{E}^a\cdot\boldsymbol{E}^a - \boldsymbol{B}^a\cdot\boldsymbol{B}^a) \qquad [17]$$

$$\mathcal{H}_{\rm YM} = \tfrac{1}{2}(\boldsymbol{E}^a\cdot\boldsymbol{E}^a + \boldsymbol{B}^a\cdot\boldsymbol{B}^a) \qquad [18]$$

The first term is kinetic, with $\boldsymbol{E}^a = -\partial_t \boldsymbol{A}^a$ also functioning as the (negative) canonical momentum $\boldsymbol{\pi}^a$, conjugate to the canonical variable $\boldsymbol{A}^a$; the second magnetic term gives the potential. In the Weyl gauge, the theory remains invariant against time-independent gauge transformations. The time component of equation [14] (Gauss law) is absent (because there is no $A^a_0$ to vary); rather it is imposed as a fixed-time constraint on the canonical variables $\boldsymbol{E}^a$ and $\boldsymbol{A}^a$. This regains the Gauss law:

$$(\boldsymbol{\mathcal{D}}\cdot\boldsymbol{E})^a = 0 \quad \text{(in the absence of sources)} \qquad [19a]$$

In the quantum theory $\boldsymbol{\mathcal{D}}\cdot\boldsymbol{E}$ annihilates “physical” states. Explicitly, in a functional Schrödinger representation, where states are functionals of the canonical fixed-time variable $\boldsymbol{A}|\Psi\rangle \to \Psi(\boldsymbol{A})$, [19a] requires

$$\left(\boldsymbol{\mathcal{D}}\cdot\frac{\delta}{\delta\boldsymbol{A}}\right)^a \Psi(\boldsymbol{A}) = 0 \qquad [19b]$$

that is, physical states must be invariant against infinitesimal gauge transformation, or equivalently, against gauge transformations that are homotopic (continuously deformable) to the identity (the so-called “small” gauge transformations)

$$\Psi(\boldsymbol{A} + \boldsymbol{\mathcal{D}}\lambda) = \Psi(\boldsymbol{A}) \qquad [20]$$

But homotopically nontrivial gauge transformation functions that cannot be deformed to the identity (the so-called “large” gauge transformations) may be present. Their effect is not controlled by Gauss’ law, and must be discussed separately.

Fixed-time gauge transformation functions depend on the spatial variable $\boldsymbol{r}: U(\boldsymbol{r})$. For a topological classification, we require that $U$ tend to a constant at large $r$. Equivalently, we compactify the base space $R^3$ to $S^3$. Thus, the gauge functions provide a mapping from $S^3$ into the relevant gauge group $G$, and for nonabelian compact gauge groups such mappings fall into disjoint homotopy classes labeled by an integer winding number $n$: $\Pi^3(G) = Z$. Gauge functions $U_n$ belonging to different classes cannot be deformed into each other; only those in the “zero” class are deformable to the identity. An analytic expression for the winding number $\omega(U)$ is

$$\omega(U) = \frac{1}{24\pi^2}\int \mathrm{d}^3x\, \varepsilon^{ijk}\,\mathrm{tr}(U^{-1}\partial_i U U^{-1}\partial_j U U^{-1}\partial_k U) \qquad [21]$$

This is a most important topological entity for gauge theories in four-dimensional spacetime, that is, in 3-space, and we shall meet it again in a description of gauge theories in three-dimensional spacetime, that is, on a plane. Various features of $\omega$ expose its topological character: (1) $\omega(U)$ does not involve a metric tensor, yet it is diffeomorphism invariant. (2) $\omega(U)$ does not change under local variations of $U$:

$$\begin{aligned}\delta\omega(U) &= \frac{1}{8\pi^2}\int \mathrm{d}^3x\, \partial_i \varepsilon^{ijk}\,\mathrm{tr}(U^{-1}\delta U U^{-1}\partial_j U U^{-1}\partial_k U)\\ &= \frac{1}{8\pi^2}\int \mathrm{d}S^i\, \varepsilon^{ijk}\,\mathrm{tr}(U^{-1}\delta U U^{-1}\partial_j U U^{-1}\partial_k U)\\ &= 0\end{aligned} \qquad [22]$$

The last integral is over the surface (at infinity) bounding the base space and vanishes for localized variations $\delta U$. In fact, the entire $\omega(U)$, not only its variation, can be presented as a surface integral, but this requires parametrizing the group element $U$ on $R^3$. For example, for SU(2),

$$\begin{aligned}&U = \exp\lambda, \quad \lambda = \lambda^a\sigma^a/2\mathrm{i} \quad (\boldsymbol{\sigma} \equiv \text{Pauli matrices})\\ &\omega(U) = \frac{1}{16\pi^2}\int \mathrm{d}S^i\, \varepsilon^{ijk}\varepsilon_{abc}\hat{\lambda}^a\partial_j\hat{\lambda}^b\partial_k\hat{\lambda}^c(\sin|\lambda| - |\lambda|)\\ &|\lambda| \equiv \sqrt{\lambda^a\lambda^a}, \qquad \hat{\lambda}^a \equiv \lambda^a/|\lambda|\end{aligned} \qquad [23]$$

Specifically, with $|\lambda| \underset{r\to\infty}{\longrightarrow} 2\pi n$ (so that $U \underset{r\to\infty}{\longrightarrow} \pm I$), $\omega(U) = -n$. As befits a topological entity, $\omega(U)$ is determined by global (here large distance) properties of $U$.

Since all gauge transformations, small and large, are symmetry operations for the theory, [20] should be generalized to

$$\Psi(\boldsymbol{A}^{U_n}) = \mathrm{e}^{\mathrm{i}n\theta}\Psi(\boldsymbol{A}) \qquad [24]$$

where $\theta$ is an universal constant. Thus, Yang–Mills quantum states behave as Bloch waves in a periodic lattice, with large gauge transformations playing the role of lattice translations and the Yang–Mills vacuum angle $\theta$ playing the role of the Bloch momentum. This is further understood by noting that the profile of the potential energy density, $\frac{1}{2}\boldsymbol{B}^a\cdot\boldsymbol{B}^a$ possesses a periodic structure symbolically depicted in Figure 2.

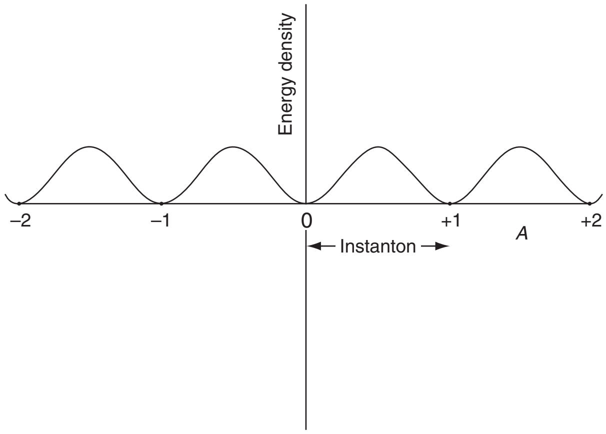



**Figure 2** Schematic for energy periodicity of Yang–Mills fields.

Thanks to Gauss' law, potentials $\boldsymbol{A}$ that differ by small gauge transformations are identified, while those differing by large gauge transformations give rise to the periodicity. Zero energy troughs correspond to pure gauge vector potentials in different homotopy classes $n$: $\boldsymbol{A} = -U_n^{-1}\nabla U_n$.

The $\theta$ angle (Bloch momentum) arises from quantum tunneling in $\boldsymbol{A}$ space. Usually, in field theory tunneling is suppressed by infinite energy barriers. (This gives rise to spontaneous symmetry breaking.) However, in Yang–Mills theory there are paths in field space that avoid such barriers. Quantum tunneling paths are exhibited in a semiclassical approximation by identifying classical motion in imaginary time (Euclidean space) that interpolates between classically degenerate vacua and possesses finite action.

In Yang–Mills theory, continuation to imaginary time, $x^0 \to ix^4$, places a factor of i on $\boldsymbol{E}^a$. Zero (Euclidean) energy is maintained when $\boldsymbol{E}^a = \pm \boldsymbol{B}^a$, or with covariant notation in Euclidean space,

$$\tfrac{1}{2}\varepsilon^{\mu\nu\alpha\beta}F_{\alpha\beta} \equiv {}^*F^{\mu\nu} = \pm F^{\mu\nu} \qquad [25]$$

Euclidean finite action field configurations that satisfy [25] are called self-dual or anti-self-dual instantons. By virtue of the Bianchi identity [10], instantons also solve the field equation [14a] in Euclidean space. Since the Euclidean action may also be written as

$$I_{\mathrm{YM}} = \frac{1}{4}\int \mathrm{d}^4x\, \mathrm{tr}(F^{\mu\nu} \pm {}^*F^{\mu\nu})(F_{\mu\nu} \pm {}^*F_{\mu\nu}) \mp \frac{1}{2}\int \mathrm{d}^4x\, \mathrm{tr}\,{}^*F^{\mu\nu}F_{\mu\nu} \qquad [26]$$

and the first term vanishes for instantons, we see that instantons are characterized by the last term, the Chern–Pontryagin index,

$$\mathcal{P} \equiv -\frac{1}{16\pi^2}\int \mathrm{d}^4x\, \mathrm{tr}({}^*F^{\mu\nu}F_{\mu\nu}) = -\frac{1}{32\pi^2}\int \mathrm{d}^4x\, \varepsilon^{\mu\nu\alpha\beta}\,\mathrm{tr}(F_{\alpha\beta}F_{\mu\nu}) \qquad [27]$$

This again is an important topological entity:

1. The diffeomorphism invariant $\mathcal{P}$ does not involve the metric tensor.
2. $\mathcal{P}$ is insensitive to local variations of $A_\mu$,

$$\delta\mathcal{P} = -\frac{1}{8\pi^2}\int \mathrm{d}^4x\, \mathrm{tr}({}^*F^{\mu\nu}\delta F_{\mu\nu}) = -\frac{1}{4\pi^2}\int \mathrm{d}^4x\, \mathrm{tr}({}^*F^{\mu\nu}D_\mu\delta A_\nu) = \frac{1}{4\pi^2}\int \mathrm{d}^4x\, \mathrm{tr}(D_\mu{}^*F^{\mu\nu}\delta A_\nu) = 0 \qquad [28]$$

3. $\mathcal{P}$ may be presented as a surface integral owing to the formula

$$\tfrac{1}{4}\mathrm{tr}\,{}^*F^{\mu\nu}F_{\mu\nu} = \partial_\mu K^\mu \qquad [29]$$

$$K^\mu \equiv \varepsilon^{\mu\nu\alpha\beta}\,\mathrm{tr}\left(\tfrac{1}{2}A_\alpha\partial_\beta A_\gamma + \tfrac{1}{3}A_\alpha A_\beta A_\gamma\right) \qquad [30]$$

where $K^\mu$ is the Chern–Simons current,

$$\mathcal{P} = -\frac{1}{4\pi^2}\int \mathrm{d}S_\mu K^\mu \qquad [31]$$

The integral [31] is over the base space boundary, $S^3$. The Chern–Pontryagin index of any gauge field configuration with finite (Euclidean) action (not only instantons) is quantized. This is because finite action requires $F_{\mu\nu}$ to vanish at large distances; equivalently, $A_\mu \to U^{-1}\partial_\mu U$. Using this in [30] renders [31] as

$$\mathcal{P} = \frac{1}{24\pi^2}\int \mathrm{d}S_\mu \varepsilon^{\mu\alpha\beta\gamma} \times \mathrm{tr}(U^{-1}\partial_\alpha U U^{-1}\partial_\beta U U^{-1}\partial_\gamma U) \qquad [32]$$

which is the same as [20] and, for the same reason, is given by an integer $[\Pi^3(G) = Z]$. Alternatively, for instantons in the (Euclidean) Weyl gauge ($A_4 = 0$), which interpolate as $x^4$ passes from $-\infty$ to $+\infty$ between degenerate, classical vacua $A_i = 0$ and $A_i = -U^{-1}\nabla_i U$, $\mathcal{P}$ becomes

$$\mathcal{P} = \frac{1}{4\pi^2}\int \mathrm{d}x^4\,\mathrm{d}^3x\,(\partial_4K^4 + \nabla\cdot\boldsymbol{K}) = \frac{1}{4\pi^2}\int \mathrm{d}^3x\, K^4\,|_{x^4=\infty} = \frac{1}{24\pi^2}\int \mathrm{d}^3x\,\varepsilon^{ijk}\,\mathrm{tr}(U^{-1}\partial_i U U^{-1}\partial_j U U^{-1}\partial U) = \omega(U) \qquad [33]$$

We have assumed that the potentials decrease at large arguments sufficiently rapidly so that the gradient term in the first integrand does not contribute. This rederivation of [32] relies on the "motion" of an instanton between vacuum configurations of different winding numbers.

An explicit 1-instanton SU(2) solution ($\mathcal{P}=1$) is

$$A_\mu = \frac{-2i}{(x-\xi)^2+\rho^2}\sigma_{\mu\nu}x^\nu \qquad [34]$$

(Upon reinserting the coupling constant $g$, which has been scaled to unity, the field profiles acquire the factor $g^{-1}$.) In [34], $\sigma_{\mu\nu} \equiv (1/4i)(\sigma^\dagger_\mu\sigma_\nu - \sigma^\dagger_\nu\sigma_\mu), \sigma_\mu \equiv (-i\sigma, I)$. $\xi$ is the "location" of the instanton, $\rho$ is its "size," and there are three more implicit parameters fixing the gauge, for a total of eight parameters that are needed to specify a single SU(2) instanton. One can show that there exist $N$ instanton/anti-instanton solutions ($\mathcal{P}=N/-N$) and in SU(2) they depend on $8N$ parameters. From [26] we see that at fixed $N$, instantons minimize the (Euclidean) action. Explicit formulas exist for the most general $N=2$ solution, while for $N\geq 3$ explicit formulas exhibit only $5N+7$ parameters. But algorithms have been found that construct the most general $8N$-parameter instantons. The 1-instanton solution is unchanged by SO(5) rotations, the maximal compact subgroup of the SO(5, 1) conformal invariance group for the Euclidean 4-space Yang–Mills equation [14a].

The Chern–Pontryagin index also appears in the Yang–Mills quantum action, for the following reason. Since all physical states respond to gauge transformations $U_n$ with the universal phase $n\theta$ [24], physical states may be presented in factorized form,

$$\Psi(\boldsymbol{A}) = e^{i\mu W(\boldsymbol{A})\theta}\Phi(\boldsymbol{A}) \qquad [35]$$

where $\Phi(\boldsymbol{A})$ is invariant against all gauge transformations, small and large, while the phase response is carried by $W(\boldsymbol{A})$,

$$W(\boldsymbol{A}^{U_n}) = W(\boldsymbol{A}) + n \qquad [36]$$

An explicit expression for $W(\boldsymbol{A})$ is given by $-(1/4\pi^2)\int d^3x\, K^0$, where $K^0$ is the time (fourth) component of $K^\mu$, with dependence on the fourth variable suppressed, that is, $K^0$ is defined on 3-space,

$$W(A) = -\frac{1}{4\pi^2}\int d^3x\, \varepsilon^{ijk}\,\mathrm{tr}\left(\tfrac{1}{2}A_i\partial_j A_k + \tfrac{1}{3}A_iA_jA_k\right) \qquad [37]$$

The gauge transformation properties of $W(A)$ are

$$\begin{aligned} W(\boldsymbol{A}^U) &= W(\boldsymbol{A}) + \frac{1}{8\pi^2}\int d^3x\,\varepsilon^{ijk}\partial_i\,\mathrm{tr}(\partial_j U U^{-1}A_k) \\ &\quad + \frac{1}{24\pi^2}\int d^3x\,\varepsilon^{ijk}\mathrm{tr}(U^{-1}\partial_i U\; U^{-1}\partial_j U\; U^{-1}\partial_k U) \end{aligned} \qquad [38]$$

The middle surface term does not contribute for well-behaved $\boldsymbol{A}$; the last term is again $\omega(U)$, the winding number of the gauge transformation $U$. Thus, [36] is verified.

The universal gauge-varying phase $e^{i\theta W(A)}$, which multiplies all gauge-invariant functional states, may be removed at the expense of subtracting from the action

$$\theta\int d^4x\,\partial_t W(\boldsymbol{A}) = -\frac{\theta}{4\pi^2}\int d^4x\,\partial_t K^0 = \theta\mathcal{P}$$

(as in [33]). Thus, the Yang–Mills quantum action extends [12] to

$$I_{\mathrm{YM}}{}^{\mathrm{quantum}} = \int d^4x\,\mathrm{tr}\left(\frac{1}{2}F^{\mu\nu}F_{\mu\nu} + \frac{\theta}{16\pi^2}{}^*F^{\mu\nu}F_{\mu\nu}\right) \qquad [39]$$

The additional Chern–Pontryagin term in [35] does not contribute to equations of motion, but it is needed to render all physical states invariant against all gauge transformations, large and small. With this transformation, one sees that the $\theta$-angle is a Lorentz invariant, but CP noninvariant effect. Evidently, specifying a classical gauge theory requires fixing a group; a quantized gauge theory is specified by a group and a $\theta$-angle, which arises from topological properties of the gauge theory. The energy eigenvalues depend on $\theta$, and distinct $\theta$'s correspond to distinct theories.

Note that the reasoning leading to [24] and [39] relies on exact quantum-mechanical arguments, while the instanton-based tunneling discussion is semiclassical.

### Adding Fermions

When fermions couple to the gauge fields, the previously described topological effects are modified by action of the chiral anomaly. Dirac fields, either noninteracting but quantized, or unquantized but interacting with a gauge potential through a covariantly conserved current $J^\mu_a$, $\mathcal{L}_\mathrm{I} = -J^\mu_a A^a_\mu$, also possess a chiral current $j^\mu_5 = \bar\psi\gamma^\mu\gamma_5\psi$, which satisfies

$$\partial_\mu j^\mu_5 = 2m\,i\bar\psi\gamma_5\,\psi \qquad [40]$$

Here $m$ is the mass, if any, of the fermions. $j_5^\mu$ is conserved for massless fermions, which therefore enjoy a chiral symmetry: $\psi \to e^{i\alpha\gamma_5}\psi$. However, when the interacting fermions are quantized, there arises correction to [40]; this is the chiral anomaly:

$$\partial_\mu \langle j_5^\mu \rangle_A = 2im\langle \bar{\psi}\gamma_5\psi\rangle_A + C\,{}^*F^{\mu\nu\,a}F^a_{\mu\nu} \qquad [41]$$

$C$ is determined by the fermion quantum numbers and coupling strengths. (For a single charged ($e$) fermion and a U(1) gauge potential, $C = e^2/8\pi^2$.) $\langle | \rangle_A$ signifies the fermionic vacuum matrix element in the presence of $A_\mu$. The modified equation [41] indicates that even in the massless limit chiral symmetry remains broken due to the anomaly, which arises with quantized fermions.

$\langle j_5^\mu \rangle_A$ may also be presented as

$$\langle j_5^\mu \rangle_A = \mathrm{tr}\,\gamma_5\gamma^\mu\langle \psi\bar{\psi}\rangle_A \qquad [42]$$

In Euclidean space $\langle \psi\bar{\psi}\rangle_A$ is the coincident-point limit of the resolvent $R(x,y;\mu)$ for the Dirac equation,

$$R(x,y;\mu) = \sum_\epsilon \frac{\psi_\epsilon(x)\,\psi_\epsilon^\dagger(y)}{\epsilon + i\mu} \qquad [43]$$

Here $\psi_\epsilon$ is an eigenfunction of the massless, Euclidean Dirac operator in the presence of the gauge field $A_\mu$,

$$i\,\gamma^\mu\,(\partial_\mu + A_\mu)\psi_\epsilon = \epsilon\psi_\epsilon \qquad [44]$$

The coincident-point limit is singular, so $R$ must be regulated: $R \to R - R_{\mathrm{Reg}}$ (we do not specify the regularization procedure). It then follows that

$$\begin{aligned}\partial_\mu\langle j_5^\mu\rangle &= 2i\mu\sum_\epsilon \frac{\psi_\epsilon^\dagger(x)\gamma_5\,\psi_\epsilon(x)}{\epsilon + i\mu} - \mathrm{tr}\,\gamma_5\,\gamma^\mu\partial_\mu R_{\mathrm{Reg}} \\ &= 2i\mu\sum_\epsilon \frac{\psi_\epsilon^\dagger(x)\gamma_5\psi_\epsilon(x)}{\epsilon + i\mu} + C\,{}^*F^{\mu\nu\,a}F^a_{\mu\nu}\end{aligned} \qquad [45]$$

The first term on the right-hand side is the (Euclidean space) analog of the mass term in [40] or [41], while the second survives even after the regulators are removed, giving the anomaly $\mathrm{tr}\,{}^*F^{\mu\nu}F_{\mu\nu}$.

The anomaly formula [41], or more explicitly [45], is also the local form of the Atiyah–Singer index theorem, which follows after [45] is integrated over all space: The left-hand side integrates to zero. The integral of the first term on the right-hand side, $\int dx\,\psi_\epsilon^*\gamma_5\psi_\epsilon$, vanishes for $\epsilon \neq 0$ by orthogonality, because $\gamma_5\psi_\epsilon$ is an eigenfunction of [44] with eigenvalue $-\epsilon$. Only zero modes contribute to the $\epsilon$ sum since these can be chosen to be eigenfunctions of $\gamma_5$, $n_\pm$ of them satisfying $\psi_0 = \pm\gamma_5\psi_0$. For a single multiplet, the normalizations work out so that

$$n_+ - n_- = \frac{1}{16\pi^2}\int d^4x\,\mathrm{tr}\,{}^*F^{\mu\nu}F_{\mu\nu} \qquad [46]$$

The result that the (signed) number of zero modes is the Chern–Pontryagin index is an instance of the Atiyah–Singer theorem. (In specific applications, one can frequently show that $n_+$ or $n_-$ vanishes.) It, therefore, follows that in the background field of instantons, the Euclidean Dirac equation possesses zero modes.

Another viewpoint on the chiral anomaly arises within the functional integral formulation, where the exponentiated action is constructed from unquantized fields, over which the functional integration is performed. Here the classical action retains chiral symmetry $\psi \to e^{i\alpha\gamma_5}\psi$, but the Grassmann fermion measure $d\psi d\bar{\psi}$, once it is properly regularized, looses chiral invariance and acquires the anomaly,

$$d\psi d\bar{\psi} \to d\psi d\bar{\psi}\,\exp iC\int d^4x\,\alpha\,\mathrm{tr}\,{}^*F^{\mu\nu}F_{\mu\nu} \qquad [47]$$

Evidently, the chiral anomaly involves the gauge-theoretic topological entity, the Chern–Pontryagin density. Not unexpectantly, the anomaly phenomenon affects significantly the topological properties of the gauge theory that are connected to $\mathcal{P}$ and were described previously.

When there is (at least) one massless fermion coupling to the Yang–Mills fields, the Yang–Mills $\theta$-angle looses physical relevance. This is because a chiral transformation that redefines the massless Dirac field does not modify the classical action, but owing to the chiral noninvariance of the functional measure, [47], an anomaly term is induced in the (effective) quantum action. The strength of this induced term can be fixed so that it cancels the $\theta$-term in [39]. Since field redefinition cannot affect physics, the elimination of the $\theta$-term indicates that it had no physical relevance in the first place. In particular, energy eigenvalues no longer depend on $\theta$.

An alternate argument for the same conclusion is based on the functional determinant that arises when the functional integral is performed over the massless Dirac field: $\det[\gamma^\mu(\partial_\mu + A_\mu)]$. The semiclassical tunneling analysis of the $\theta$-angle is based on instantons, but in the presence of instantons the Dirac equation has a zero mode [46]. Consequently the determinant vanishes, tunneling is suppressed and so is the $\theta$-angle.

However, in the standard model for particle physics, there are no massless fermions, so the presence of the $\theta$-angle entails the following physical consequences. The tunneling amplitude $\Gamma$ in leading semiclassical approximation is determined by the Euclidean action, namely the continuation of $iI_{\mathrm{YM}}$ in

[39] to imaginary time. This results in the same expression except that the topological $\theta$-term acquires a factor of i. Only the 1-instanton and anti-instanton give the dominant contribution,

$$\Gamma \propto \cos\theta\, \mathrm{e}^{-8\pi^2/g^2} \qquad [48]$$

where the coupling constraint $g$ has been reinserted; the proportionality constant has not been computed, owing to infrared divergences. (Higher-instanton-number configurations contribute at an exponentially subdominant order and have thus far played no role in physics.) The tunneling leads to baryon decay, but fortunately at an exponentially small rate. More useful is the fact that instanton tunneling gives semiclassical evidence for the removal of an unwanted chiral U(1) Goldstone symmetry, which would be present in the standard model if the chiral anomaly did not interfere. Furthermore, the chiral anomaly facilitates the decay of the neutral pion to two photons; a process forbidden by other apparent chiral symmetries of the standard model, which in fact are modified by the chiral anomaly. Gauge fields in four dimensions must interact with anomaly free currents. This necessitates a precise adjustment of fermion content and charges so that the anomaly coefficients (analogs of "C" in [41]) vanish for currents coupled to gauge fields. Finally, $\theta$ provides a tantalizing source of CP violation in the strong-interaction sector of the standard model. But no experimental signal (e.g., neutron electric dipole moment) for this effect has been seen. At present, we do not know what mechanism is responsible for keeping $\theta$ vanishingly small.

These are the physical consequences of topological effects in four-dimensional gauge theories. Although they have provided experimentalists with only a few numbers to measure (e.g., $\pi^0 \to 2\gamma$ decay amplitude, prediction of anomaly-free arrangements of quarks and leptons in families), they have added enormously to our appreciation of the complexities of quantized gauge theories.

That chiral anomalies are an obstruction to consistent gauge interactions can be established within perturbation theory. A similar, but nonperturbative effect is seen in an SU(2) gauge theory with $N$ Weyl fermion ($\gamma_5\psi = \pm\psi$) SU(2) doublets, which lead upon functional integration to $\det[\gamma^\mu(\partial_\mu + A_\mu)]^{N/2}$. But because $\Pi^4(\mathrm{SU}(2)) = Z_2$, there exists a single homotopy class of gauge transformations which are not deformable to the identity. One shows that the determinant changes sign when such a gauge transformation is performed. Thus, the theory is ill-defined for odd $N$. Consistent SU(2) gauge theories must possess an even number of Weyl fermion doublets, but such models have not found a place in physical theory.

### Adding Bosons

Instantons are finite-action solutions to classical equations continued to imaginary time; they provide a semiclassical description of quantum-mechanical tunneling. A field theory may also possess finite-energy, time-independent (static) solutions to the real-time equations of motion. When these solutions are stable for topological reasons, they are called "solitons." Solitons give semiclassical evidence for the existence in the quantum field theory of a particle sector disjoint from the particles obtained by quantizing field fluctuations around the vacuum state. The soliton particles are heavy for weak coupling $g$. (Their energy is $O(1/g^2)$; the field profiles are $O(1/g)$.) They do not decay owing to the conservation of "charges" that do not arise from Noether's theorem but are topological.

Yang–Mills theory does not possess soliton solutions (except in five-dimensional spacetime, where the static solitons are just the four-dimensional instantons discussed previously). However, when a gauge theory, based on a simple group is coupled to a scalar field that undergoes symmetry breaking to U(1), soliton solutions exist. These are the 't Hooft–Polyakov magnetic monopoles, found in a SU(2) gauge theory with scalar fields in the adjoint representation, as well as various generalizations. The topological consideration that arises here concerns finite energy of the static, scalar field multiplet $\varphi$, which in the Weyl gauge is

$$E(\varphi) = \int \mathrm{d}^3 \boldsymbol{x} \Big( |(\boldsymbol{D}\varphi)^a \cdot (\boldsymbol{D}\varphi)^a|^2 + V(\varphi) \Big) \qquad [49]$$

$V$ is non-negative and possesses non trivial symmetry breaking zeroes. On the sphere $S^2$ at spatial infinity, $\varphi$ must tend to such a zero. Thus, the fields belong to $G/H$, where $G$ is the gauge group and $H$ the unbroken subgroup. For the 't Hooft–Polyakov monopole these are SU(2) and U(1), respectively, and the scalar field provides a mapping of the sphere at infinity $S^2$ to $S^2 \approx \mathrm{SU}(2)/\mathrm{U}(1)$.

One now considers $\Pi^2(S^2) = \Pi^2(\mathrm{SU}(2)/\mathrm{U}(1)) = \Pi^1(\mathrm{U}(1)) = Z$, and one shows that the magnetic flux is determined by the winding number. Hence, the magnetic charge is quantized. Explicitly, the electromagnetic U(1) gauge field is given by

$$\begin{aligned} f_{\mu\nu} &\equiv \hat{\varphi}^a\, {}^*F^a_{\mu\nu} - \varepsilon_{abc}\hat{\varphi}^a (D_\mu\hat{\varphi})^b (D_\nu\hat{\varphi})^c \\ &= \partial_\mu a_\nu - \partial_\nu a_\mu \\ a_\mu &\equiv \hat{\varphi}^a A^a_\mu - \cos\alpha\,\partial_\mu\beta \end{aligned} \qquad [50]$$

where $\hat{\varphi}^a$ is the unit isovector, parametrized as $\hat{\varphi}^a = (\sin\alpha\cos\beta, \sin\alpha\sin\beta, \cos\alpha)$. The manifestly conserved magnetic current

$$j_m^\mu = \partial_\nu {}^*f^{\mu\nu} \qquad [51a]$$

is rearranged to read

$$j_m^\mu = -\tfrac{1}{2}\varepsilon^{\mu\alpha\beta\gamma}\varepsilon_{abc}\partial_\alpha\hat{\varphi}^a\partial_\beta\hat{\varphi}^b\partial_\gamma\hat{\varphi}^c \qquad [51b]$$

and is nonvanishing because $\varphi^a$ possesses zeroes, where $\partial_\alpha\hat{\varphi}^a$ acquires localized singularities. The magnetic charge

$$m = -\frac{1}{4\pi}\int d^3x\, j_m^0 = \frac{1}{4\pi}\int d^3x\, \nabla\cdot \boldsymbol{b} \qquad [52]$$

($b^i = \mathrm{U}(1)$ magnetic field: $-\frac{1}{2}\varepsilon^{ijk}f_{jk} = {}^*f^{i0}$) is given by the topological entity (Kronecker index of the mapping)

$$\begin{aligned} m &= \frac{1}{8\pi}\int d^3x\, \varepsilon^{ijk}\varepsilon_{abc}\partial_i\hat{\varphi}^a\partial_j\hat{\varphi}^b\partial_k\hat{\varphi}^c \\ &= \frac{1}{8\pi}\int dS^i\, \varepsilon^{ijk}\varepsilon_{abc}\hat{\varphi}^a\partial_j\hat{\varphi}^b\partial_k\hat{\varphi}^c \\ &= -\frac{1}{4\pi}\int dS^i\, \varepsilon^{ijk}\partial_j\cos\alpha\,\partial_k\beta \end{aligned} \qquad [53]$$

which readily evaluates the integer winding number.

The theory also supports charged magnetic monopole solutions called "dyons." Here the profiles involve time-periodic gauge potentials, where the time variation is just a gauge transformation $\partial_t A_\mu = D_\mu\lambda$. (Gauge-equivalent, static expressions have slow large-distance fall-off, which is removed by the time-dependent gauge function.) For dyons, the integer valued Chern–Pontryagin index, with the integration taken over all space and in time over the dyon period, reproduces the magnetic monopole strength.

Regrettably, these fascinating structures are not found in nature. Nor do they arise in the standard model, whose structure group is not simple, although speculative grand unified models, with simple $G$ and $H = \mathrm{SU}(3)\times\mathrm{U}(1)$, would support magnetic monopoles and dyons. While challenged physically, the magnetic monopole phenomena have produced extensive and interesting mathematical analysis.

## Gauge Theories in Two Dimensions

Two-dimensional gauge theories have only a few physical applications; edge states of the planar quantum Hall effect can be described by excitations moving on a line. However, the abelian model with fermions is useful in that it provides a very accurate reflection of topological behavior in the physically important four-dimensional theory.

### Abelian Gauge Theory

Take the spatial interval to be $[-L, L]$. Homotopically nontrivial gauge transformations satisfy $\lambda(L) - \lambda(-L) = 2\pi n$ ($\Pi^1 U(1) = Z$). States $\Psi(A)$ of the free gauge theory that satisfy Gauss' law and respond with a $\theta$-angle are

$$\begin{aligned} \Psi(A) &= \exp\frac{i\theta}{2\pi}\int dx\, A \\ \Psi(A+\partial\lambda) &= e^{in\theta}\Psi(A) \end{aligned} \qquad [54]$$

In this model, $\theta$ has the interpretation of a constant background electric field $\mathcal{E} = -\theta/2\pi$,

$$\begin{aligned} E\Psi(A) &= \mathcal{E}\Psi(A), \quad E \equiv F_{01} \\ i\frac{\delta}{\delta A}\Psi(A) &= -\frac{\theta}{2\pi}\Psi(A) \end{aligned} \qquad [55]$$

This also gives the energy eigenvalue:

$$\frac{1}{2}\int dx\, E^2\Psi(A) = \frac{1}{2}\int dx\, \mathcal{E}^2\Psi(A) \qquad [56]$$

The phase may be removed by adding to the Lagrangian $-(\theta/2\pi)\int dx\, \partial_t A$; equivalently, the action becomes

$$I_{\mathrm{EM}}^{\mathrm{quantum}} = \int d^2x\left(-\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{\theta}{4\pi}\varepsilon^{\mu\nu}F_{\mu\nu}\right) \qquad [57a]$$

which apart from a constant is also given by a formula with the background field:

$$I_{\mathrm{EM}}^{\mathrm{quantum}} = \frac{1}{2}\int dx (E+\mathcal{E})^2 \qquad [57b]$$

Because of gauge invariance, there is only one state, annihilated by $E$ and carrying energy $\frac{1}{2}\int dx\, \mathcal{E}^2$. Distinct $\theta$ (different $\mathcal{E}$) correspond to distinct theories.

We recognize in [57a] the two-dimensional Chern–Pontryagin density, contributing a total derivative to the action,

$$\mathcal{P} = \frac{1}{4\pi}\int d^2x\, \varepsilon^{\mu\nu}F_{\mu\nu} \qquad [58]$$

the Chern–Simons current, whose divergence is $\mathcal{P}$,

$$K^\mu = \frac{1}{2\pi}\varepsilon^{\mu\nu}A_\nu \qquad [59]$$

and the Chern–Simons term, which carries the phase of $\Psi$

$$\int dx\, K^0 = \frac{1}{2\pi}\int dx\, A \qquad [60]$$

For Euclidean-space gauge potentials, which are given at large distance by the pure gauge $2\pi n \tan^{-1} y/x$, $\mathcal{P} = n$. All this is just as in the four-dimensional theory, except there are no instantons and no tunneling.

### Adding Fermions

The addition of massless fermions to the U(1) gauge theory results in the Schwinger model of massless quantum electrodynamics in two-dimensional spacetime. The equation of motion becomes

$$\partial_\mu F^{\mu\nu} = J^\nu \qquad [61]$$

with the vector current constructed from the Dirac fields as $J^\mu = \bar{\psi}\gamma^\mu\psi$. This current remains conserved in the quantized version because it couples to the gauge field. But the axial vector current $j_5^\mu = \bar{\psi}\gamma^\mu\gamma_5\psi$ acquires an anomaly that involves the Chern–Pontryagin density in [58],

$$\partial_\mu j_5^\mu = \frac{1}{2\pi}\varepsilon^{\mu\nu}F_{\mu\nu} \qquad [62]$$

The model is readily solved, and shows no $\theta$-angle (background field) dependence in physical quantities. The solution is directly obtained by combining [61] with [62] into a second-order differential equation and using the matrix identity of two-dimensional Dirac (= Pauli) matrices: $\varepsilon^{\mu\nu}\gamma_\nu\gamma_5 = \gamma^\mu$. It follows that

$$\left(\Box + \frac{1}{\pi}\right)E = 0 \qquad [63]$$

So the theory describes a free massive photon (mass squared $= 1/\pi$ in units of $\hbar$ and the coupling constant, which have been scaled to unity), with no sign of a $\theta$-angle (background field).

However, in parallel with four-dimensional behavior, the model with massive fermions regains a $\theta$ dependence in the particles' energy spectrum; a result that is established perturbatively, because a complete solution is not available.

Note that in the Schwinger model, the gauge particle ("photon") acquires a mass, even though local gauge invariance is preserved. This happens essentially for topological/anomaly reasons. Such topological mass generation is met again in three dimensions.

### Adding Bosons

Scalar electrodynamics with a negative mass squared term in $(3+1)$-dimensional spacetime leads to the Higgs mechanism and short-range interactions due to the massive photons. In $(1+1)$ spacetime dimensions, the model possesses instantons – scalar and gauge field profiles that solve the imaginary-time equations of motion – labeled by $\Pi^1(\mathrm{U}(1)) = Z$. These disorder the Higgs condensate so that the force between charged particles remains long-range, like in the positive mass-squared case. This is a vivid example of how excitations arising from nontrivial topological issues significantly effect physical content.

## Gauge Theories in Three Dimensions

Gauge theories on three-dimensional spacetime, that is, evolving on a plane, have physical application to planar phenomena, like the quantum Hall effect. Also, the high-temperature limit of four-dimensional field theories is governed by the corresponding field theory in three Euclidean dimensions.

In three (more generally, odd) dimensions, there are no Chern–Pontryagin quantities, no Chern–Simon currents, no axial vector currents or anomalies (there is no $\gamma_5$ matrix). These are replaced by odd-dimensional entities that can modify Yang–Mills dynamics.

### Yang–Mills and Other Gauge Theories

Using the three-index Levi-Civita tensor, one can construct a gauge-covariant, covariantly conserved vector, which can be added to the Yang–Mills equation. Thus, [14] can be modified to

$$D_\mu F^{\mu\nu} + \frac{m}{2}\varepsilon^{\mu\alpha\beta}F_{\alpha\beta} = J^\nu \qquad [64a]$$

or, equivalently, in terms of the dual-field strength ${}^*F^\mu \equiv \frac{1}{2}\varepsilon^{\mu\alpha\beta}F_{\alpha\beta}$,

$$\varepsilon^{\mu\nu\alpha}D_\mu {}^*F_\alpha + m {}^*F^\nu = J^\nu \qquad [64b]$$

For dimensional balance, $m$ carries dimension of mass. Indeed, in the source-free case [64] implies

$$(D^\alpha D_\alpha + m^2){}^*F_\mu = \varepsilon_{\mu\alpha\beta}[{}^*F^\alpha, {}^*F^\beta] \qquad [65]$$

This shows that excitations are massive, even though local gauge invariance is preserved. Otherwise, as in the Dirac monopole case, the equations of motion are unexceptional.

However, for the quantum theory we need the action, whose variation produces the mass term in [64]. This is just the Chern–Simons term $W(A)$ in [37], multiplied by $-8\pi^2 m$ and now defined on $(2+1)$-dimensional spacetime:

$$I_{\mathrm{CS}} = 2m\int \mathrm{d}^3x\, \varepsilon^{\alpha\beta\gamma}\mathrm{tr}\left(\tfrac{1}{2}A_\alpha\partial_\beta A_\gamma + \tfrac{1}{3}A_\alpha A_\beta A_\gamma\right) \qquad [66]$$

Everything holds also in the abelian theory; the last term in [66] is then absent.

In this model, the mass is generated by a topological mechanism since $I_{CS}$ possesses the usual attributes for a topological entity: it is diffeomorphisms invariant without a metric tensor; when the potentials are appropriately parametrized, it is given by a surface term. (In the abelian case, the appropriate parametrization is in terms of Clebsch decomposition, $A_\mu = \partial_\mu\theta + \alpha\partial_\mu\beta$.) Most importantly, in the nonabelian theory [66] changes by $8\pi^2 mn$ with three-dimensional gauge transformations carrying winding number $n$. Hence, for consistency of the nonabelian quantum theory, $m$ must be quantized as $n/4\pi$ (in units of $\hbar$ and the coupling constant, which have been scaled to unity). All this is a clear field-theoretic analog to the quantum mechanics of the Dirac monopole, and just as for the magnetic monopole, a Hamiltonian argument for quantizing $m$ can be constructed, as an alternative to the above action-based derivation.

The time component of [64] relates the electric and magnetic fields to the charge density:

$$\boldsymbol{D}\cdot\boldsymbol{E} - mB = \rho \qquad [67]$$

In the abelian case, the first term involves a total derivative and its spatial integral vanishes, leaving a formula that identifies magnetic flux with a total charge. At low energy, the mass term dominates the conventional kinetic term in [64], and the flux–charge relation becomes a local field-current identity,

$$m^*F^\nu \approx J^\nu \qquad [68]$$

These formulas have made Chern–Simons-modified gauge theories relevant to issues in condensed matter physics, for example, the quantum Hall effect. In the abelian case, $m$ need not be quantized.

### Adding Fermions

Three-dimensional Dirac matrices are minimally realized by $2\times 2$ Pauli matrices. As a consequence, a mass term is not parity invariant; also, there is no $\gamma_5$ matrix, since the product of the three Dirac (= Pauli) matrices is proportional to $I$. While there are no chiral anomalies, there is the so-called parity anomaly: integrating a single doublet of massless SU(2) fermions one obtains $\Delta(\boldsymbol{A}) \equiv \det[\gamma^\mu(\mathrm{i}\partial_\mu + \boldsymbol{A}_\mu)]$, which should preserve parity and gauge invariance.

Since there are no anomalies in current divergences, $\Delta(\boldsymbol{A})$ is certainly invariant against infinitesimal gauge transformations. But for finite gauge transformations (categorized by $\Pi^3(\mathrm{SU}(2) = Z)$ one finds that $\Delta(\boldsymbol{A})$ is not invariant: when the gauge transformation belongs to an odd-numbered homotopy class, $\Delta(\boldsymbol{A})$ changes sign. To regain gauge invariance, one must either work with an even number of fermion doublets or, if only one doublet (more generally, odd number) is to be used, one must add to the gauge Lagrangian a parity-violating Chern–Simons term with half the correctly quantized coefficient, to neutralize the gauge non-invariance of $\Delta(\boldsymbol{A})$.

Alternatively, $\Delta(\boldsymbol{A})$ can be regularized in a gauge-invariant manner. But this requires massive, Pauli–Villars regulator fields, which produce a parity-violating expression for $\Delta(A)$. One cannot avoid the parity anomaly.

### Adding Bosons

There are a variety of bosonic field models that one may consider: Abelian or nonabelian; with conventional kinetic term or supplemented by the Chern–Simons topological mass; or, for low energy, no kinetic term but only the Chern–Simons term, as in [68]. Abelian charged Bose fields in a Maxwell theory lead to vortex solitons, based on $\Pi^1(\mathrm{U}(1)) = Z$. These are just the instantons of the $(1+1)$-dimensional bosonic gauge theory discussed previously. With Maxwell kinematics there are no charged vortices, but these appear when the Chen–Simons mass is added; see [67]. Pure Chern–Simons kinematics, with no Maxwell term, can produce completely integrable soliton equations (Liouville, Toda) when the Bose field dynamics is appropriately chosen.

## Conclusion

Topological effects in field theory are associated with the infinities and regularization that beset quantum field theories. These give rise to the chiral anomaly, parity anomaly (and scale symmetry anomalies, not discussed here). Yet the anomalies themselves are finite quantities that have topological significance (Atiyah–Singer, Chern–Pontryagin, Chern–Simons). This paradoxical pairing has not been understood. Nor can we explain why the anomalies interfere in a topological manner with symmetries associated with masslessness.

Although the range of topological effects in gauge theory is large, and even larger in non-gauge theories (sigma models, Skyrme models) the relevance to actual fundamental physics is confined to the $\theta$-angle phenomenon, which is analyzed accurately and abstractly by reference to $\Pi^3(G)$ and to the interplay with fermions through the chiral anomaly. Instantons are relevant only to an approximate, semiclassical discussion. Although after much mathematical work, general instanton configurations are well understood, only the 1-instanton solution enjoys physical significance.

Other topological entities that fascinate are either nonexistent in fundamental physics or are relevant to

condensed matter physics (vortices, Chern–Simons effects). But here too, we note that the fundamental equation of condensed matter physics – the many-body Schrödinger equation – carries no evident topological structure. Only the phenomenological equations, which replace the fundamental one, give rise to topological intricacies.

## Acknowledgment

This work is supported in part by funds provided by the US Department of Energy (DOE) under cooperative research agreement DE-FC02-94ER40818.

*See also:* Abelian and Nonabelian Gauge Theories Using Differential Forms; Abelian Higgs Vortices; Anomalies; BF Theories; Gauge Theories from Strings; Gauge Theory: Mathematical Applications; Quantum Field Theory: A Brief Introduction; Seiberg–Witten Theory.

## Further Reading

Adler SL (1970) Perturbation theory anomalies. In: Deser S, Grisaru M, and Pendleton H (eds.) *Lectures on Elementary Particles and Quantum Field Theory*, vol. 1, pp. 3–164. Cambridge: MIT Press.
Adler SL (2005) Anomalies to all orders (ArXiv: hep-th/0405040). In: 't Hooft G (ed.) *Fifty Years of Yang–Mills Theory*. Singapore: Word Scientific.
Bertlmann RA (1996) *Anomalies in Quantum Field Theory*. Oxford: Clarendon.
Coleman S (1985) Classical lumps and their quantum descendants and the uses of instantons. In: Coleman S *Aspects of Symmetry*, pp. 185–350. Cambridge: Cambridge University Press.
Fujikawa K and Suzuki H (2004) *Path Integrals and Quantum Anomalies*. Oxford: Oxford University Press.
Jackiw R (1977) Quantum meaning of classical field theory. *Reviews of Modern Physics* 49: 681–706.
Jackiw R (1979) Introduction to the Yang–Mills quantum theory. *Reviews of Modern Physics* 52: 661–673.
Jackiw R (1985) Field theoretic investigations in current algebra and topological investigations in quantum gauge theories. In: Treiman S, Jackiw R, Zumino B, and Witten E (eds.) *Current Algebra and Anomalies*, pp. 81–359. Princeton: Princeton University Press; Singapore: World Scientific.
Jackiw R (1995) *Diverse Topics in Theoretical and Mathematical Physics*. Singapore: World Scientific.
Jackiw R (2005) Fifty years of Yang–Mills theory and our moments of triumph (ArXiv: physics/0403109). In: 't Hooft G (ed.) *Fifty Years of Yang–Mills Theory*. Singapore: World Scientific.
Jackiw R and Pi S-Y (1992) Chern–Simons solitons. *Progress of Theoretical Physics* 107: 1–40.
Rajaraman R (1982) *Solitons and Instantons*. Amsterdam: North-Holland.
Shifman M (1994) *Instantons in Gauge Theories*. Singapore: World Scientific.
't Hooft G (1976) Symmetry breaking through Bell–Jackiw anomalies. *Physical Review Letters* 37: 8–11.
Weinberg S (1996) *The Quantum Theory of Fields*. vol. II, chs. 22 and 23. Cambridge: Cambridge University Press.

# Normal Forms and Semiclassical Approximation

**D Bambusi**, Universitá di Milano, Milan, Italy



## Introduction

Quantum mechanics was born at the beginning of the twentieth century with the quantization rules for the harmonic oscillator and for the hydrogen atom. Such rules were almost immediately extended to more general systems by the so-called Bohr–Sommerfeld quantization rule: "the actions of the classical system can assume only those values which are integer multiples of $\hbar$." However, the actions are defined only in some special situations and, moreover, at the present time the Schrödinger equation is the paradigm of quantum mechanics. A question naturally arises: is there any relation between the eigenvalues of the Schrödinger operator and the numbers obtained by Bohr–Sommerfeld quantization rule (when available)?

According to common wisdom, the "Bohr–Sommerfeld numbers" are a first approximation to the eigenvalues of the Schrödinger operator in the so-called semiclassical limit. However, precise mathematical results on the subject were obtained only in the 1980s and a good understanding of the problem has been achieved only recently. In particular it is now clear how to compute higher-order corrections to the eigenvalues: this is done through suitable normal form procedures.

In the present article we will discuss the above questions for the case of perturbed harmonic oscillators, a case which, on the one hand, is physically relevant and, on the other, is well understood. We will only briefly discuss the quantization of perturbations of integrable systems.

## A Statement

On $L^2(\mathbb{R}^n)$, consider the Schrödinger operator

$$\hat{H} = -\frac{\hbar^2}{2}\Delta + V \qquad [1]$$

where $\Delta$ is the $n$-dimensional Laplacian and $V$ is a smooth real potential having an absolute nondegenerate minimum at the origin. We are interested in

the eigenvalues of [1] close to zero. Introduce coordinates adapted to the normal modes, namely such that

$$V(x) = \sum_{i=1}^{n} \frac{\omega_i^2 x_i^2}{2} + O(\|x\|^3)$$

Assume

(H1) *Nonresonance*: There exist $\gamma > 0$ and $\tau \in \mathbb{R}$ such that, for any $k \in \mathbb{Z}^n - \{0\}$ one has

$$|\omega \cdot k| \geq \frac{\gamma}{|k|^\tau} \qquad [2]$$

(H2) $V(x) > 0$ for $x \neq 0$, and

$$\liminf_{|x|\to\infty} V(x) > 0$$

(H3) $V \in C^\infty(\mathbb{R}^n)$ and for any $r \geq 0$ there exists $C_r$ such that

$$\left|\frac{\partial^{|\alpha|} V}{\partial x^\alpha}(x)\right| \leq C_{|\alpha|}\langle x\rangle^m, \quad \forall \alpha \in \mathbb{N}^n$$

where we used the notation $\langle x\rangle := (1 + \|x\|^2)^{1/2}$.

**Theorem 1** *Assume that* (H1)–(H3) *hold. Then, for any positive N, M there exist positive constants* $\hbar_{N,M}, \epsilon_{N,M}, C^1_{N,M}, C^2_{N,M}$, *and a smooth function*

$$\mathcal{Z}_{N,M}(I_1, \ldots, I_n; \hbar)$$

*such that,* $\forall 0 < \epsilon \leq \epsilon_{N,M}$ *and* $0 < \hbar \leq \hbar_{N,M}\epsilon$, *the eigenvalues of* [1] *in* $[0, \epsilon)$ *have the representation*

$$\lambda_k = \left(k + \tfrac{1}{2}\right) \cdot \omega\hbar + \mathcal{Z}_{N,M}\left(\left(k + \tfrac{1}{2}\right)\hbar; \hbar\right) + R_{NM}(k, \hbar), \quad k \in \mathbb{N}^n,\ k_j \geq 1 \qquad [3]$$

*where*

$$|R_{NM}(k, \hbar)| \leq C^1_{N,M}\epsilon^N + C^2_{N,M}\left(\frac{\hbar}{\epsilon}\right)^M$$

*More precisely, for any* $k \in \mathbb{N}^n$ *such that*

$$\left(k + \tfrac{1}{2}\right) \cdot \omega\hbar + \mathcal{Z}_{N,M}\left(\left(k + \tfrac{1}{2}\right)\hbar; \hbar\right) \in [0, \epsilon) \qquad [4]$$

*there exists an eigenvalue* $\lambda_k \in [0, \epsilon)$ *for which* [3] *holds, and vice versa, for any eigenvalue in* $[0, \epsilon)$ *there exists a k satisfying* [3] *and* [4]. *The function* $\mathcal{Z}_{N,M}(I_1, \ldots, I_n; 0)$ *coincides with the classical Birkhoff normal form of the system computed up to order N.*

The proof of the theorem is constructive, in the sense that it provides an algorithm allowing to construct explicitly, by elementary operations, the function $\mathcal{Z}_{N,M}$. One could choose $\epsilon = \epsilon(\hbar) = \hbar^\delta$ with some positive $\delta < 1$, obtaining a simpler statement valid for the eigenvalues in $[0, \hbar^\delta)$. It is also possible to weaken the nonresonance condition (H1) to the condition $\omega \cdot k \neq 0$ for $k \in \mathbb{Z}^n - \{0\}$.

A theorem very close to [1] was proved by Sjöstrand (1992) by a method different from the one that will be presented here (see also Graffi and Paul (1987)). In the analytic or Gevrey case (recall that a $C^\infty$ function $f(x)$ is Gevrey in some domain if there exist constants $C, \sigma$ such that, for all multi-indexes $\alpha \in \mathbb{N}^n$ one has

$$\left|\frac{\partial^{|\alpha|} f}{\partial x^\alpha}\right| \leq C^{|\alpha|}(\alpha!)^\sigma$$

in the whole domain), the error can be reduced to be exponentially small with the parameters (Bambusi *et al.* 1999). Previous results dealing with compact perturbations of the harmonic oscillator were obtained by Bellissard and Vittot (1990). It is possible to deal also with the resonant case in which (H1) is violated. In this case the spectrum of the complete system is qualitatively different from the spectrum of the harmonic one. As discussed later, the normal form allows one to compute the main qualitative differences.

## Birkhoff Normal Form

In this section we recall the procedure leading to classical Birkhoff normal form, whose quantization leads to the proof of Theorem 5.

### Birkhoff's Theorem

The operator [1] is the quantization of the classical Hamiltonian

$$\sum_{i=1}^{n} \frac{\xi_i^2}{2} + V(x) \qquad [5]$$

Denote

$$H_0(\xi, x) := \sum_{j=1}^{n} \omega_j I_j, \quad I_j := \frac{\xi_j^2 + \omega_j^2 x_j^2}{2\omega_j} \qquad [6]$$

then we have

**Theorem 2** *For any positive integer* $N \geq 2$ *there exist a neighborhood* $\mathcal{U}_N$ *of the origin and a canonical transformation* $\mathcal{T}_N : \mathbb{R}^{2n} \supset \mathcal{U}_N \to \mathbb{R}^{2n}$ *which puts the system* [5] *in Birkhoff normal form up to order N, namely such that*

$$H \circ \mathcal{T}_N = H_0 + Z^N + R_N \qquad [7]$$

*where* $Z^N$ *Poisson-commutes with* $H_0$, *namely* $\{H_0; Z^N\} \equiv 0$ *and* $R_N$ *is small*, *that is*,

$$|R_N(\xi, x)| \leq C_N \|(\xi, x)\|^{N+1} \quad [8]$$

*Moreover*, *if the frequencies are nonresonant*, *namely*

$$\omega \cdot k \neq 0, \quad \forall k \in \mathbb{Z}^n \backslash \{0\} \quad [9]$$

*the function* $Z^N$ *depends on the actions* $I_j$ *only*. We recall that the Poisson bracket of two functions $f$ and $g$ is defined by

$$\{f; g\} := \sum_{j=1}^{n} \left( \frac{\partial f}{\partial \xi_j} \frac{\partial g}{\partial x_j} - \frac{\partial f}{\partial x_j} \frac{\partial g}{\partial \xi_j} \right) = -\{g; f\}$$

and coincides with the Lie derivative of $g$ with respect to the Hamiltonian vector field of $f$.

**Remark 1** In the case where the frequencies fulfill (H1) and the potential $V$ is analytic (or of Gevrey class) the remainder can be reduced to be exponentially small with $\|(\xi, x)\|$.

### Scheme of the Proof

Make the rescaling $\xi = \epsilon \xi'$, $x = \epsilon x'$. In terms of the primed variables, the Hamiltonian of the system [5] takes the form

$$H_\epsilon(\xi', x') = H_0(\xi', x') + \epsilon W(x') \quad [10]$$

with

$$W(x') := \frac{V(\epsilon x') - \epsilon^2 \sum_{j=1}^{n} \omega_j^2 (x_j')^2 / 2}{\epsilon^3} = W_3(x') + \epsilon W_4(x') + \cdots \quad [11]$$

and $W_l$ is the Taylor polynomial of order $l$ of $V$. In what follows we will omit primes from the scaled variables.

Given an auxiliary Hamiltonian $\chi_3$, denote by $\Phi_t^3$ the flow of the corresponding Hamiltonian vector field. We construct $\chi_3$ so that $H_\epsilon \circ \Phi_\epsilon^3$ is in normal form up to order $\epsilon^2$.

**Remark 2** Given a $C^\infty$ function $g$ one has $g \circ \Phi_\epsilon^3 \sim \sum_{l=0}^{\infty} \epsilon^l g_l$, with

$$g_0 := g, \qquad g_l = \frac{1}{l} \{\chi_3; g_{l-1}\}, \quad l \geq 1 \quad [12]$$

where $\sim$ denotes the fact that the left-hand side is asymptotic to the right-hand side (a precise definition appears later in the article). If both $g$ and $\chi_3$ are analytic then the series of $g \circ \Phi_\epsilon^3$ can be shown to converge in a neighborhood of the origin. Using [12] to compute $H_\epsilon \circ \Phi_\epsilon^3$, we get

$$H_\epsilon \circ \Phi_\epsilon^3 = H_0 + \epsilon[W_3 + \{\chi_3; H_0\}] + O(\epsilon^2)$$

So $H_\epsilon \circ \Phi_\epsilon^3$ is in normal form up to $O(\epsilon^2)$ provided $\chi_3$ fulfills the so-called homological equation:

$$W_3 + \{\chi_3; H_0\} = Z_3 \quad [13]$$

where the unknown function $Z_3$ has to be in normal form. Note that, since the operator

$$\chi \mapsto \{\chi; H_0\}$$

maps linearly polynomials of degree $l$ into polynomials of degree $l$, eqn [13] can be interpreted as a linear equation in the finite-dimensional space of polynomials of degree 3 in the phase-space variables.

**Lemma 1** *The homological equation* [13] *admits a solution* $(\chi_3, Z_3)$.

**Proof** Introduce the canonical coordinates $(\zeta, \eta)$ by

$$\zeta_j := \frac{1}{\sqrt{2}} \left( \frac{\xi_j}{\sqrt{\omega_j}} + \mathrm{i} x_j \sqrt{\omega_j} \right)$$
$$\eta_j := \frac{1}{\mathrm{i}\sqrt{2}} \left( \frac{\xi_j}{\sqrt{\omega_j}} - \mathrm{i} x_j \sqrt{\omega_j} \right) \quad [14]$$

In these variables the unperturbed Hamiltonian $H_0$ reads $H_0 = \sum_{j \geq 1} \mathrm{i}\omega_j \zeta_j \eta_j$ and $W_3$ is transformed in a different polynomial, again of third order. The important fact is that in these coordinates the eigenvectors of the linear operator $\{H_0; .\}$ are the monomials

$$\zeta^k \eta^l \equiv \zeta_1^{k_1} \cdots \zeta_n^{k_n} \eta_1^{l_1} \cdots \eta_n^{l_n}$$

Indeed, one has $\{H_0; \zeta^k \eta^l\} = \mathrm{i}\omega \cdot (k - l) \zeta^k \eta^l$. As a consequence, writing

$$W_3(\zeta, \eta) = \sum_{k,l} C_{k,l} \zeta^k \eta^l$$

one can define the resonant set

$$\mathcal{R} := \{(k, l) : \omega \cdot (k - l) = 0\}$$

and

$$Z_3(\zeta, \eta) := \sum_{k,l \in \mathcal{R}} C_{k,l} \zeta^k \eta^l$$
$$\chi_3(\zeta, \eta) := \sum_{k,l \notin \mathcal{R}} \frac{C_{k,l}}{\mathrm{i}\omega \cdot (k - l)} \zeta^k \eta^l \quad [15]$$

Going back to the original variables, one has the solution of the homological equation. □

**Definition 1** The function $Z_3$ solving [13] will be called the resonant part of $W_3$ and will be denoted by $\langle W_3 \rangle$.

Using the function $\chi_3$, one can transform the Hamiltonian to the form

$$H_0 + \epsilon Z_3 + \epsilon^2 R_3$$

**Remark 3** Equation [12] allows to construct directly the Taylor expansion of $R_3$ in terms of the Taylor expansion of $W$ and of its Poisson brackets with $\chi_3$.

Iterating the construction (which however slightly changes due to the presence of $Z_3$), one gets the proof of Theorem 2.

**Remark 4** In the nonresonant case $\omega \cdot (k - l) = 0$ implies that $k = l$; therefore, the resonant part of a polynomial is the sum of monomials of the form

$$\zeta^k \eta^k = I_1^{k_1} \cdots I_n^{k_n}$$

that is, it is a function of the actions only. Moreover, in this case one has $Z_3 = 0$, while in general $Z_4 \neq 0$.

## Some Symbolic Calculus

To understand how to quantize the procedure of Birkhoff normal form, we consider the classical–quantum correspondence. It is well known that there are different procedures in order to associate an operator with a classical observable. Here we concentrate on the Weyl quantization rule.

To a function $f \in \mathcal{S}(\mathbb{R}^{2n})$ (Schwartz class), we associate an operator $\hat{f}$ acting on functions $\psi \in \mathcal{S}(\mathbb{R}^n)$, which is defined by

$$[\hat{f}\psi](x) := \frac{1}{(2\pi\hbar)^n} \int_{\mathbb{R}^n \times \mathbb{R}^n} f\left(\frac{x+y}{2}, \xi\right) \times e^{\frac{i(x-y)\cdot\xi}{\hbar}} \psi(y)\, dy\, d\xi \qquad [16]$$

**Definition 2** The operator [16] is called the Weyl quantization of $f$ and in turn $f$ is called the symbol of $\hat{f}$.

Using the method of oscillatory integrals, the Weyl quantization rule can be extended to much more general observables $f$. We recall that, roughly speaking, the method of oscillatory integrals consists in giving meaning to a formal expression of the form [16] by using successive integration by parts (see, e.g., Martinez (2001)).

**Definition 3** A function $f \in C^\infty(\mathbb{R}^{2n})$ will be called a smooth symbol of class $S(\langle z \rangle^m)$ if, for any $r \geq 0$, there exists $C_r$ such that

$$\left| \frac{\partial^{|\alpha|} f}{\partial z^\alpha}(z) \right| \leq C_{|\alpha|} \langle z \rangle^m, \quad \forall \alpha \in \mathbb{N}^{2n}$$

Where $\langle z \rangle$ is as defined earlier.

It is useful to extend such a definition to functions explicitly depending also on $\hbar$. This can be done in a straightforward way by asking the constants $C_r$ to be independent of $\hbar$ in a neighborhood of the origin. Different classes of symbols can also be defined, but for our purpose this class is enough.

**Theorem 3** *Let $f \in S(\langle z \rangle^m)$, $m \in \mathbb{R}$, and $\psi \in \mathcal{S}(\mathbb{R}^n)$; then the formal expression [16] is a well-defined oscillatory integral.*

**Example 1** Under Weyl quantization rule, one has

$$\hat{\xi}_j = i\hbar\partial_{x_j}, \quad \hat{x}_j = x_j \quad \text{(multiplication operator)}$$
$$\widehat{\xi_j x_j} = \tfrac{1}{2}(\hat{\xi}_j \hat{x}_j + \hat{x}_j \hat{\xi}_j)$$

**Definition 4** A sequence $(f_j)_{j\geq 0}$ with $f_j \in S(\langle z \rangle^m)$ will be called the asymptotic expansion of $f \in S(\langle z \rangle^m)$ if, for any integer $N$, there exist two positive constants $C_N, \hbar_N$ such that

$$f = \sum_{j=0}^{N} \hbar^j f_j + R_N$$

with $|R_N(z, \hbar)| \leq C_N \hbar^{N+1} \langle z \rangle^m$, and $\hbar \in (0, \hbar_N)$.

The key point for the quantization of the normal form procedure is the following.

**Theorem 4** *Let $f \in S(\langle z \rangle^{m_1})$ and $g \in S(\langle z \rangle^{m_2})$; then there exists a unique $F \in S(\langle z \rangle^{m_1+m_2})$ such that*

$$\hat{F} = \hat{f}\hat{g} \quad (\textit{operator product!})$$

*moreover, one has*

$$F = \exp\left(\frac{i\hbar}{2}(\partial_x \cdot \partial_\eta - \partial_y \cdot \partial_\xi)\right) \times (f(x,\xi) g(y,\eta))|_{y=x,\, \eta=\xi} \qquad [17]$$

*Finally, $F$ admits an asymptotic expansion in $\hbar$ which coincides with the formal expansion of* [17].

The proof is obtained by using eqn [16] to write down an expression for $\hat{f}\hat{g}\psi$ and obtain a formula for $F$. Then, one shows that the formula is well defined and therefore the result is not formal.

**Definition 5** In the above context, the symbol $G$ of

$$\frac{i}{\hbar}\left[\hat{f}; \hat{g}\right] =: \hat{G}$$

will be called the "Moyal bracket" of $f$ and $g$ and will be denoted by $\{f; g\}_M$.

By formula [17], one has in particular

$$\{f; g\}_M = \{f; g\} + \hbar^2 \Delta_1(f, g) + O(\hbar^4) \qquad [18]$$

where

$$\Delta_1(f,g) = -\frac{1}{24}\left(\frac{\partial^3 f}{\partial\xi^3}\frac{\partial^3 g}{\partial x^3} - 3\frac{\partial^3 f}{\partial\xi^2\partial x}\frac{\partial^3 g}{\partial x^2\partial\xi} + 3\frac{\partial^3 f}{\partial\xi\partial x^2}\frac{\partial^3 g}{\partial x\partial\xi^2} - \frac{\partial^3 f}{\partial x^3}\frac{\partial^3 g}{\partial\xi^3}\right)$$

where we used a vector notation for the derivatives. If either $f$ or $g$ are polynomials of degree $\leq 2$, then

$$\{f;g\}_{\mathrm{M}} = \{f;g\} \qquad [19]$$

Given a self-adjoint operator $A$ and a smooth function $G:\mathbb{R}\to\mathbb{R}$, it is well known how to define by spectral theorem the operator $G(A)$. Suppose now that $A=\hat{f}$ for some symbol $f$. In general, one has $G(\hat{f})\neq\widehat{G\circ f}$. However, by symbolic calculus (i.e., using eqn [17]), one has:

**Lemma 2** *Denote* $I_j(x,\xi)=(\omega_j^2 x_j^2+\xi_j^2)/2\omega_j$. *Then, for any positive integer* $k$ *there exists a function* $F_k(I_j,\hbar)$ *such that*

$$\widehat{(I_j)^k} = F_k(\hat{I}_j,\hbar)$$

*where the right-hand side is defined by spectral calculus. Moreover,* $F_k$ *can be computed explicitly by the recursion formula* $F_{k+1}=I_jF_k+F_{k-1}\hbar^2(k^2-k+1)/4$.

As a consequence of this fact and of the fact that $[\hat{I}_j,\hat{I}_l]=0$, one has that the Weyl quantization of a polynomial function of the actions is a function of the action operators.

## Semiclassical Normal Form

Let $\chi$ be a smooth symbol such that $\hat{\chi}$ is self-adjoint, and consider the group of unitary operators $X_\epsilon := \exp((\mathrm{i}\epsilon/\hbar)\hat{\chi})$. Let $g$ be a smooth symbol; apply the unitary transformation $X_\epsilon$ to $\hat{g}$, namely compute $X_\epsilon\hat{g}X_\epsilon^{-1}$. Noting that (on a suitable domain)

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}(X_\epsilon\hat{g}X_\epsilon^{-1}) = X_\epsilon\frac{\mathrm{i}}{\hbar}[\hat{\chi};\hat{g}]X_\epsilon^{-1}$$

one has (formally!) the expansion of $X_\epsilon\hat{g}X_\epsilon^{-1}$ in $\epsilon$:

$$X_\epsilon\hat{g}X_\epsilon^{-1} = \sum_{l\geq 0}\epsilon^l\widehat{g_{q,l}}$$

where

$$\widehat{g_{q,0}} := \hat{g}, \quad \widehat{g_{q,l}} = \frac{1}{l}\frac{\mathrm{i}}{\hbar}[\hat{\chi};\widehat{g_{q,l-1}}],\ l\geq 1 \qquad [20]$$

(Such a series can be interpreted as an asymptotic expansion provided one restricts the domain at each step of the approximation.) Equivalently, the symbol of $X_\epsilon\hat{g}X_\epsilon^{-1}$ is formally given by $\sum_l \epsilon_l g_{q,l}$ with

$$g_{q,0} := g, \quad g_{q,l} := \frac{1}{l}\{\chi; g_{q,l-1}\}_{\mathrm{M}},\ l\geq 1 \qquad [21]$$

from which one sees a remarkable similitude with the classical equation. Moreover, [21] converges to [12] when $\hbar\to 0$.

Applying the unitary transformation generated by $\hat{\chi}$ to the Hamiltonian operator $\hat{H}_\epsilon$ (cf. eqn [10]), one has $X_\epsilon\hat{H}_\epsilon X_\epsilon^{-1}=\widehat{H_q^1}$ with

$$H_q^1 = H_0 + \epsilon[W_3 + \{\chi;H_0\}_{\mathrm{M}}] + O(\epsilon^2) \qquad [22]$$

$$\equiv H_0 + \epsilon[W_3 + \{\chi;H_0\}] + O(\epsilon^2) \qquad [23]$$

where we used the fact that $H_0$ is a quadratic polynomial, so that [19] holds. It is thus clear that Lemma 1 allows to solve also the quantum homological equation appearing in this context and to determine the symbol of the operator generating the unitary transformation putting the Hamiltonian operator in normal form up to corrections of order $\epsilon^2$. Moreover, one can compute in terms of Moyal brackets (of polynomials!) the expansion of the symbol of the new remainder and of the normal form. Iterating the construction, one generates a well-defined semiclassical normal form of the quantum system.

**Example 2** Denote by $Z_{q,l}, l=1,2\ldots,$ the term added to the semiclassical normal form at the $l$th step of the iterative construction. Explicitly, the first terms are given by

$$Z_{q,1} = \langle W_3\rangle = Z_3 \qquad [24]$$

$$Z_{q,2} = \langle W_4\rangle + \tfrac{1}{2}\langle\{\chi_3;W_3\}_{\mathrm{M}}\rangle + \tfrac{1}{2}\langle\{\chi_3;Z_3\}_{\mathrm{M}}\rangle \qquad [25]$$

$$Z_{q,3} = \langle W_5\rangle + \langle\{\chi_4;Z_3\}_{\mathrm{M}}\rangle + \tfrac{1}{3}\langle\{\chi_3;H_2\}_{\mathrm{M}}\rangle + \tfrac{1}{2}\langle\{\chi_3;W_{3,1}\}_{\mathrm{M}}\rangle + \langle\{\chi_3;W_4\}_{\mathrm{M}}\rangle \qquad [26]$$

where, according to Definition 1, $\langle\,.\,\rangle$ is the resonant part of its argument, $\chi_j$ is (formally) the symbol of the operator generating the $j$th unitary transformation, and

$$H_2 := \tfrac{1}{2}\{\chi_3;Z_3-W_3\}_{\mathrm{M}}, \quad W_{3,1} := \{\chi_3;W_3\}_{\mathrm{M}}$$

Note that all the Moyal brackets involved contain polynomials of degree at most 4, so that they can be computed exactly using formula [18] which in this case does not contain corrections of order $\hbar^4$.

The problem in making previous construction rigorous is that all the series involved are in general

divergent. Moreover, it is not possible to show that the remainders appearing when truncating such series are small in a reasonable sense. Nevertheless, it is possible, using the tools of microlocal analysis, to show that the semiclassical normal form contains essentially all the information on the part of the spectrum close to zero.

The precise relation between the spectrum of the original Hamiltonian and the spectrum of the semiclassical normal form is captured by the following definition.

Let $H_1(\epsilon, \hbar), H_2(\epsilon, \hbar)$ be two families of self-adjoint operators; set $\mathrm{Spec}_\epsilon(H_{1,2}) := \mathrm{Spec}(H_{1,2}) \cap [0, \epsilon)$.

**Definition 6** We say that

$$\mathrm{Spec}_\epsilon(H_1) = \mathrm{Spec}_\epsilon(H_2) \bmod (\epsilon^\infty + (\hbar/\epsilon)^\infty)$$

if for any $N, M > 0$ there exist $C^1_{N,M}$ and $C^2_{N,M}$ such that for any $\lambda_1 \in \mathrm{Spec}_\epsilon(H_1)$ there exists $\lambda_2 \in \mathrm{Spec}_\epsilon(H_2)$ such that $\lambda_1 = \lambda_2 + R_{N,M}$ with

$$|R_N| \le C^1_{N,M}\epsilon^N + C^2_{N,M}(\hbar/\epsilon)^M \qquad [27]$$

and conversely. Equation [27] has to hold for any couple $(\hbar, \epsilon)$ with $\epsilon$ and $(\hbar/\epsilon)$ small enough.

**Theorem 5** *Assume* (H2) *and* (H3); *assume also*: (H1′) *There exist* $\gamma > 0$ *and* $\tau \in \mathbb{R}$ *such that, for any* $k \in \mathbb{Z}^n$, *one has*

$$\text{either } \omega \cdot k = 0 \quad \text{or } |\omega \cdot k| \ge \frac{\gamma}{|k|^\tau} \qquad [28]$$

*Then there exists a polynomial function* $\mathcal{Z}_q$ *such that one has*

$$\mathrm{Spec}_\epsilon(\hat{H}) = \mathrm{Spec}_\epsilon(\hat{H}_0 + \widehat{\mathcal{Z}_q}) \bmod \left(\epsilon^\infty + \left(\frac{\hbar}{\epsilon}\right)^\infty\right) \qquad [29]$$

*The polynomial* $\mathcal{Z}_q$ *coincides with the semiclassical normal form defined at the beginning of the section.*

**Scheme of the proof** It consists of six steps. (1) Make the unitary transformation $(U\psi)(x) := \epsilon^{n/4}\psi(\epsilon^{1/2}x)$ which transforms the Hamiltonian operator [1] into the Weyl quantized of $\epsilon H_\epsilon := \epsilon(H_0 + \epsilon^{1/2}W)$, but a Weyl quantization where $\hbar$ is substituted by $\hbar' := \hbar/\epsilon$. (2) Make a cutoff of $H_\epsilon$, namely, fix $R$ and consider a smooth function $t$ such that $t(s) \equiv 1$ for $|s| \le R$, $t(s) \equiv 0$ for $|s| \ge 2R$, define $a(x, \xi) := W(x)t(\|(\xi, x)\|)$. (3) Compare the spectrum of the Hamiltonian $\hat{H}_\epsilon$ with the spectrum of $H^t := H_0 + \epsilon a$. By microlocal analysis, one has that, in any fixed bounded interval such spectra coincide modulo $\hbar^\infty$ (see, e.g., Martinez (2001)). (4) Rescale back the variables, namely apply the transformation $U^{-1}_\epsilon$ to $H^t$. (5) Apply the normal form algorithm to the so-obtained Hamiltonian showing that all the series involved are convergent in suitable norms. (6) Use again microlocal analysis to show that the spectrum of the semiclassical normal form coincides with the spectrum of the normalized operator with compactly supported symbol. □

**Remark 5** Fix an arbitrary $1 > \delta > 0$ and link $\epsilon$ to $\hbar$ by $\epsilon := \hbar^\delta$. Then one obtains a simplified statement according to which the spectrum of [1] in $[0, \hbar^\delta]$ coincides modulo $\hbar^\infty$ with the spectrum of $\hat{H}_0 + \widehat{\mathcal{Z}}_q$ in the same interval.

**Remark 6** In the case where the frequencies are nonresonant one has that the symbol of the normal form depends on the actions only. By Lemma 2 one has that also the quantization of the normal form is a function of the action operators only (explicitly computable), and therefore the spectrum of the normal form is given by a quantization formula as claimed in Theorem 1.

## The Resonant Case

In the case where the frequencies are nonresonant, due to the particular structure of the normal form, one obtains a very precise information on the spectrum. In the case where there are some resonances, the situation is more difficult. In order to illustrate what happens we concentrate on the completely resonant case, that is, the case where all the frequencies are integer multiples of a single fundamental frequency $\nu$.

In this case, the eigenvalues of $\hat{H}_0$ form a subset of $\mathbb{N}\hbar\nu + (1/2)|\omega|\hbar$ and are degenerate. One expects the nonlinear part to break such a degeneracy and to transform each eigenvalue in a small band. One can use the normal form to study the structure of the so-obtained band. To this end, the most relevant contribution is due to the first nonvanishing term of the normal form. For the sake of definiteness, we assume that this is the term of order 4, namely $\mathcal{Z}_4$. Denote

$$N := \mathcal{Z}_4\big|_{H_0^{-1}(1)}, \qquad B(E) \equiv \left[E - \tfrac{1}{3}\nu\hbar, E + \tfrac{1}{3}\nu\hbar\right]$$

**Theorem 6** *Fix* $1 > \gamma_1 > 1/2$, *then, provided* $\hbar$ *is small enough, one has*

$$\mathrm{Spec}(\hat{H}) \cap (\hbar, \hbar^{\gamma_1}) \subset \bigcup_{E \in \mathrm{Spec}(\hat{H}_0)} B(E) \qquad [30]$$

*Moreover, denote by*

$$E + \lambda_1(E, \hbar) \le \cdots \le E + \lambda_m(E, \hbar) \qquad [31]$$

*the eigenvalues of $\hat{H}$ in $B(E)$ counted with multiplicity, then*

$$\lambda_1(E,\hbar) = E^2 \mathrm{Min}\, N + E^2(O(\hbar/E) + O(E^{1/2})) \quad [32]$$

*and similarly*

$$E^2\lambda_m(E,\hbar) = \mathrm{Max}\ N + E^2(O(\hbar/E) + O(E^{1/2})) \quad [33]$$

This statement is due to Bambusi, Charles, and Tagliaferro (see Bambusi 2004); for previous results, see Vũ Ngọc (1998).

Equation [30] shows that the spectrum has a band structure, while eqns [32] and [33] allow one to compute the minimum and the maximum of each band.

The idea of the proof is as follows. First forget high-order terms of the normal form, whose effect is included in the error terms. Then, due to the commutation property of the normal form with $\hat{H}_0$, one has that $\mathcal{Z}_4$ restricts to an operator acting on the eigenspaces of $\hat{H}_0$. On the classical side, one has that by Marsden–Weinstein procedure $\mathcal{Z}_4$ defines a classical Hamiltonian system on the manifold obtained by symplectic reduction of the original phase space. By the methods of geometric quantization, it turns out that the quantum operator acting on an eigenspace of $\hat{H}_0$ is a Toeplitz operator whose principal symbol is exactly the above reduced classical Hamiltonian. Then, the proof follows by classical properties of Toeplitz operators.

We point out that results of this kind are useful in the computation of the molecular spectra (Michel and Zhilinskii 2001, Zhilinskii 2001).

## Quantization of KAM Tori

In this section we present a result on the quantization of KAM tori. It allows one to construct part of the spectrum of a close-to-integrable system.

We recall that a classical Hamiltonian system with $n$ degrees of freedom is said to be integrable if it has $n$ integrals of motion independent and in involution. If the energy surface is compact, then, by Arnol'd–Liouville theorem there exists a canonical transformation $\mathcal{T}_0 : \mathbb{R}^n \times \mathbb{T}^n \supset D \times \mathbb{T}^n \to \mathbb{R}^{2n}$ introducing action-angle variables, namely such that, denoting by $K_0$ the original integrable Hamiltonian, $K_0 \circ \mathcal{T}_0$ is independent of the angles $\phi \in \mathbb{T}^n$. Here, $D$ is an open bounded domain.

Consider now a close-to-integrable analytic Hamiltonian system, namely a Hamiltonian system with Hamiltonian

$$K = K_0 + \epsilon K_1$$

where $\epsilon$ is a small parameter. We assume that, denoting again by $\mathcal{T}_0$ the canonical transformation introducing action-angle variables for the system $K_0$, one has that both $K_0 \circ \mathcal{T}_0$ and $K_1 \circ \mathcal{T}_0$ are real analytic on $D \times \mathbb{T}^n$. Then, the KAM theory applies. To state the corresponding result, denote by $D_0 \subset D$ a domain whose closure is contained in $D$.

**Theorem 7** *Assume that $\forall I \in D$ one has*

$$\det\left(\frac{\partial^2 (K_0 \circ \mathcal{T}_0)}{\partial I^2}\right) \neq 0 \quad [34]$$

*then there exists a positive constant $\epsilon_*$ and, for any $\epsilon$ with $|\epsilon| < \epsilon_*$, there exists a Gevrey canonical transformation $\mathcal{T}_\epsilon : D_0 \times \mathbb{T}^n \to \mathbb{R}^{2n}$ and a Cantor set $D_\epsilon \subset D_0$ with the following properties:*

$$K \circ \mathcal{T}_\epsilon = Z(I) + R(I,\phi,\epsilon) \quad [35]$$

*where $R(I,\phi,\epsilon)$ vanishes at infinite order on $D_\epsilon$, that is, for any multi-index $\alpha$ there exists $C_{|\alpha|}$ such that one has*

$$\left|\frac{\partial^{|\alpha|} R}{\partial (I,\phi)^\alpha}(I,\phi,\epsilon)\right| \leq C_{|\alpha|} \exp\left(-\frac{c}{|I - D_\epsilon|^\rho}\right) \quad [36]$$

*with a suitable $\rho > 0$ and $|I - D_\epsilon|$ denoting the distance from $D_\epsilon$. Moreover, as $\epsilon$ tends to zero, the measure of $D_\epsilon$ tends to the measure of $D_0$.*

A particular consequence is that the set $D_\epsilon$ is foliated in invariant tori. From the proof, it also turns out that the motion on each torus is quasiperiodic with frequencies fulfilling the assumption (H1) stated earlier. Moreover, the tori are linearly stable and even more: they are stable in an exponential sense (namely, a solution starting $O(\mu)$ close to a torus takes at least a time $O(\exp(c/\mu^\rho))$ to double its distance from the torus).

Quantizing the normalizing transformation $\mathcal{T}_\epsilon$ by using the theory of Fourier integral operators, one can also put the quantum Hamiltonian in a suitable normal form which allows to deduce some spectral information on the system.

To fix ideas we restrict to the case where $K$ is a natural system, namely it has the form (3.1), and is close to integrable in the above sense. Fix two parameters $E_1 < E_2$; assume (1) that $K^{-1}([-\infty, E_2 + \delta])$ is compact for some positive $\delta$ and (2) that the domain $D_0$ can be constructed in such a way that $\mathcal{T}_0 : D_0 \times \mathbb{T}^n \to K_0^{-1}([E_0, E_1])$ is a bijection and, moreover, the KAM condition [34] holds. Denote by $\theta \in \mathbb{Z}^n$ the Maslov class of the tori of $K_0$ (see, e.g., Lazutkin (1993)) and, having fixed some $0 < \sigma < 1$, define the set of indexes

$$\mathcal{I} := \{k \in \mathbb{Z}^n : |D_\epsilon - \hbar(k + \theta/4)| \leq \hbar^\sigma\} \quad [37]$$

**Theorem 8** *There exist positive constants $\hbar_*$, $c$, $C$, and $\sigma < 1$, and a function $K_q : D_0 \times (0, \hbar_*) \to \mathbb{R}$ with the following property: for any $k \in \mathcal{I}$ there exists at least one eigenvalue of $\hat{K}$ in the interval*

$$\left[Z_q(\hbar(k+\theta/4),\hbar) - \mathrm{C}e^{-c/\hbar^\rho},\right.$$
$$\left. Z_q(\hbar(k+\theta/4),\hbar) + \mathrm{C}e^{-c/\hbar^\rho}\right] \quad [38]$$

One can also show that a large part of the spectrum is constructed in this way. This is obtained by comparing the semiclassical estimate of the number of eigenvalues in $[E_1, E_2]$ to the number of eigenvalues thus constructed.

Theorem 8 is due to Popov (2000); the quantization of KAM tori was initiated by Lazutkin and widely developed by Colin de Verdière, who obtained a result similar to Theorem 8 for the case where $K$ is $C^\infty$ and describes the geodesic flow on a compact Riemannian manifold (Colin de Verdière 1977).

*See also:* Central Manifolds, Normal Forms; *h*-Pseudodifferential Operators and Applications; Optical Caustics; Quantum Mechanics: Foundations; Schrödinger Operators; Stationary Phase Approximation.

## Further Reading

Bambusi D (2004) Semiclassical normal forms. In: *Multiscale Methods in Quantum Mechanics*, pp. 23–39. Trends in Mathematics. Boston, MA: Birkhäuser.

Bambusi D, Graffi S, and Paul T (1999) Normal forms and quantization formulae. *Communications in Mathematical Physics* 207: 173–195.

Bellissard J and Vittot M (1990) Heisenberg's picture and noncommutative geometry of the semiclassical limit in quantum mechanics. *Annales de l'Institut Henri Poincaré. Physique Théorique* 52: 175–235.

Colin de Verdière Y (1977) Quasi Modes sur les variétés Riemanniennes. *Inventiones Mathematichaes* 3: 15–52.

Graffi S and Paul T (1987) The Schrödinger equation and canonical perturbation theory. *Communications in Mathematical Physics* 108: 25–40.

Lazutkin V (1993) *KAM Theory and Semiclassical Approximations to Eigenfunctions*. Berlin: Springer.

Martinez A (2001) *An Introduction to Semiclassical and Microlocal Analysis*. New York: Springer.

Michel L and Zhilinskii BI (2001) Symmetry, invariants, and topology. I–III. *Physics Reports* 341: 11–84, 175–264.

Popov G (2000) Invariant tori effective stability and quasimodes with exponentially small error terms I–II. *Annales Henri Poincaré* 1: 223–248, 249–279.

Robert D (1987) *Autour de l'approximation semiclassique*. Basel: Birkhäuser.

Sjöstrand J (1992) Semi-excited levels in non-degenerate potential wells. *Asymptotic Analysis* 6: 29–43.

Vũ Ngọc S (1998) *Sur le spectre des systèmes complètement intégrables semi-classiques avec singularités*. Ph.D. thesis, Institut Fourier.

Zhilinskii BI (2001) Symmetry, invariants, and topology. II: *Physics Reports* 341: 85–171.

# *N*-Particle Quantum Scattering

**D R Yafaev**, Université de Rennes, Rennes, France



## Introduction

The present article relies heavily on Quantum Mechanical Scattering Theory in this Encyclopedia and can be considered as its continuation. We use here freely the notation and results discussed in this article.

An important problem of scattering theory concerns the Schrödinger $H$ operator of $N$, $N \geq 3$, interacting particles. Since the potential energy of pair interactions between particles depends on their relative positions only, it does not tend to zero at infinity in the configuration space of a system, even if the center-of-mass motion is removed. This is qualitatively different from the two-particle case. It turns out that asymptotically (for large times $t \to +\infty$ or $t \to -\infty$) an $N$-particle system splits up into clusters,

$$\mathcal{C}_1 \cup \cdots \cup \mathcal{C}_n = \{1, \ldots, N\}, \quad \mathcal{C}_k \cap \mathcal{C}_l = \emptyset \text{ if } k \neq l \quad [1]$$

Particles from the same cluster $\mathcal{C}_k$, $k = 1, \ldots, n$, form a bound state, and different clusters do not interact with each other. In particular, if $n = 1$ and $\mathcal{C}_1 = \{1, 2, \ldots, N\}$, then we have a bound state of the system. In another extreme case $n = N$, all particles are free. The asymptotic evolution determined by clusters $\mathcal{C}_1, \ldots, \mathcal{C}_n$ where $n \geq 2$, and bound states of all these clusters is called a scattering channel. Physically it is natural to expect that the list of all such channels is exhaustive, that is, no other scattering process is possible. This statement is called asymptotic completeness.

We emphasize that an $N$-particle system may be in different scattering states as $t \to +\infty$ and $t \to -\infty$ and different rearrangement processes are possible. For example, a three-particle system may asymptotically consist of free particles or a pair of particles may be in a bound state, whereas the third particle may be asymptotically free. If particles are free at both $-\infty$ and $+\infty$, then one speaks about elastic scattering; we have a capture if particles free at $-\infty$ form a bound state of a couple after the interaction; an opposite process, when a bound state at $-\infty$ gives three free particles, is known as a breakup. It is also possible that a bound state of one couple yields a bound state of

another pair (a rearrangement) or a bound state of a couple transforms into another bound state of the same couple (an excitation). All these processes are described by the scattering operator. On the contrary, if the whole system forms a bound state at $-\infty$ (i.e., $n=1$), then it remains in the same state for all $t$.

As far as monographic literature on $N$-particle scattering is concerned, we mention Dereziński and Gérard (1997), Faddeev (1965), Reed and Simon (1979), and Yafaev (2000).

## Setting the Scattering Problem

Let us recall the definition of the $N$-particle Schrödinger operator (Hamiltonian)

$$\boldsymbol{H} = \boldsymbol{H}_0 + \boldsymbol{V} \qquad [2]$$

If the configuration space of each particle is $\mathbb{R}^d$, then the operator $\boldsymbol{H}$ acts in the space $L_2(\mathbb{R}^{dN})$. The operator of kinetic energy (the "unperturbed" Hamiltonian) is

$$\boldsymbol{H}_0 = -\sum_{j=1}^{N}(2m_j)^{-1}\Delta_{x_j} \qquad [3]$$

where $x_j$ and $m_j$ are the position and mass of the particle labeled by $j$. The operator of potential energy of pair interactions of particles (the perturbation) $\boldsymbol{V}$ is the operator of multiplication by the function

$$V(x) = \sum_{i<j} V^{ij}(x_j - x_i), \quad i,j = 1,\ldots,N \qquad [4]$$

Set $\alpha=(ij)$, $x^\alpha = x_j - x_i$. It is assumed that the functions $V^\alpha(x^\alpha)$ tend to zero sufficiently rapidly as $|x^\alpha|\to\infty$ in $\mathbb{R}^d$. However, the function $V(x)\not\to 0$ as $|x|\to\infty$ in $\mathbb{R}^{dN}$ if at least one of the distances $|x_i - x_j|$ between particles remains bounded. This difficulty is manifest even for two particles ($N=2$), but in this case it disappears if the motion of the center of mass of the system is removed.

This means the following. Let the subspace $X^{\mathrm{cm}}$ of $\mathbb{R}^{dN}$ be distinguished by the condition

$$\sum_{j=1}^{N} m_j x_j = 0 \qquad [5]$$

and let $X_{\mathrm{cm}}$ be the orthogonal complement to $X^{\mathrm{cm}}$ in the space $\mathbb{R}^{dN}$ endowed with the scalar product

$$\langle x, y\rangle = 2\sum_{j=1}^{N} m_j \langle x_j, y_j\rangle_{\mathbb{R}^d} \qquad [6]$$

Then

$$L_2(\mathbb{R}^{dN}) = L_2(X_{\mathrm{cm}}) \otimes L_2(X^{\mathrm{cm}})$$

Denote by $x_{\mathrm{cm}}$, $x^{\mathrm{cm}}$ the orthogonal projections of $x\in\mathbb{R}^{dN}$ on the subspaces $X_{\mathrm{cm}}$, $X^{\mathrm{cm}}$, respectively, so that $x=(x_{\mathrm{cm}}, x^{\mathrm{cm}})$. Clearly, the vector $x_{\mathrm{cm}}$ has components

$$x_{\mathrm{cm}} = M^{-1}\sum_{j=1}^{N} m_j x_j, \quad M = \sum_{j=1}^{N} m_j$$

Let $T(p)$, $(T(p)f)(x_1,\ldots,x_N) = f(x_1+p,\ldots,x_N+p)$, be the operator of common translations of particles. The operator $\boldsymbol{H}$ commutes with $T(p)$, that is, $T(p)\boldsymbol{H} = \boldsymbol{H}T(p)$, for all $p\in\mathbb{R}^d$. It follows that

$$\boldsymbol{H} = K\otimes I + I\otimes H, \quad K = -(2M)^{-1}\Delta_{x_{\mathrm{cm}}} \qquad [7]$$

where $K$ is the kinetic energy operator of the center-of-mass motion.

The operator

$$H = H_0 + V \qquad [8]$$

acts in the space $\mathcal{H} = L_2(X^{\mathrm{cm}})$. Here $V$ is again the operator of multiplication by function [4]. The precise form of the differential operator $H_0$ depends on the choice of coordinates in $X^{\mathrm{cm}}$. For example, if $N=2$ and $x = x_2 - x_1$, then $H_0 = -(2m)^{-1}\Delta_x$ where $m = m_1 m_2 (m_1+m_2)^{-1}$. In the case $N=3$, a natural choice of coordinates in $X^{\mathrm{cm}}$ is given by one of the three sets of Jacobi variables:

$$x^{12} = x_2 - x_1$$

$$x_{12} = x_3 - (m_1+m_2)^{-1}(m_1x_1 + m_2x_2)$$

and similarly for $x^{13}, x_{13}$ and $x^{23}, x_{23}$. In coordinates $x^\alpha, x_\alpha$ the operator of kinetic energy is determined by the formula

$$H_0 = -(2m_\alpha)^{-1}\Delta_{x_\alpha} - (2m^\alpha)^{-1}\Delta_{x^\alpha}$$

where, for example,

$$(m^{12})^{-1} = m_1^{-1} + m_2^{-1}, \quad m_{12}^{-1} = (m_1+m_2)^{-1} + m_3^{-1}$$

If $N=2$, then $V(x)\to 0$ as $|x|\to\infty, x\in X^{\mathrm{cm}}$, but this is no longer true for $N\geq 3$. According to eqn [7] the spectral and scattering theories for the operator $\boldsymbol{H}$ reduce to those for the operator $H$. However, for $N\geq 3$, this reduction is not really helpful.

Let us now consider a breakup $a = \{\mathcal{C}_1,\ldots,\mathcal{C}_n\}$ of an $N$-particle system into clusters $\mathcal{C}_1,\ldots,\mathcal{C}_n$, $1\leq n =: \#(a)\leq N$ satisfying conditions [1]. If interactions between different clusters are neglected, we obtain the operator

$$H_a = H_0 + V^a, \quad V^a = \sum_{l=1}^{n}\sum_{\alpha\in\mathcal{C}_l} V^\alpha \qquad [9]$$

In particular, $H_a = H_0$ if $\#(a) = N$ and $H_a = H$ if $\#(a)=1$. Let the operator of common translations

of particles from the same cluster be defined by the equation

$$(T_a(p_1,\ldots,p_n)f)(x_1,\ldots,x_N) = f(x'_1,\ldots,x'_N)$$

where $x'_j = x_j + p_l$ if $j \in \mathcal{C}_l$. The operator $H_a$ commutes with the operators $T_a(p_1,\ldots,p_n)$ for all vectors $p_1,\ldots,p_n \in \mathbb{R}^d$. Let the subspace $X^a$ be determined by the condition

$$\sum_{j\in\mathcal{C}_l} m_j x_j = 0, \quad l = 1,\ldots,n$$

and let $X_a$ be the orthogonal complement to $X^a$ in $X^{\text{cm}}$ with respect to scalar product [6]. Clearly, $\dim X^a = (N - \#(a))d$, $\dim X_a = (\#(a) - 1)d$. Then the space $\mathcal{H}$ splits into the tensor product

$$L_2(X^{\text{cm}}) = L_2(X_a) \otimes L_2(X^a) \qquad [10]$$

In what follows, $x_a$ and $x^a$ are the orthogonal projections of $x \in X^{\text{cm}}$ on the subspaces $X_a$ and $X^a$, respectively. The "external" variable $x_a = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$, where

$$\boldsymbol{x}_l = M_l^{-1} \sum_{j\in\mathcal{C}_l} m_j x_j, \quad M_l = \sum_{j\in\mathcal{C}_l} m_j$$

describes positions of centers of masses of the clusters. The "internal" variable $x^a$ is the set of numbers $x_j - \boldsymbol{x}_l$ for all $j \in \mathcal{C}_l$ and all $l = 1,\ldots,n$. Of course, for each $l$ only $|\mathcal{C}_l| - 1$ ($|\mathcal{C}_l|$ is the number of particles in a cluster $\mathcal{C}_l$) of variables $x_j - \boldsymbol{x}_l$ are independent. Set

$$K_a = -\Delta_{x_a} = -\sum_{l=1}^{n} (2M_l)^{-1} \Delta_{\boldsymbol{x}_l}$$

and

$$H^a = -\Delta_{x^a} + V^a$$

Then

$$H_a = K_a \otimes I + I \otimes H^a$$

Note that eigenvalues $\lambda^{a,n}$ of the operator $H^a$ are sums over $l = 1,\ldots,n$ of eigenvalues of the operators

$$H(\mathcal{C}_l) = H_0(\mathcal{C}_l) + \sum_{\alpha\in\mathcal{C}_l} V^\alpha$$

describing each cluster. Similarly, eigenfunctions $\psi^{a,n}$ of $H^a$ are products of eigenfunctions of these operators. We usually write $\boldsymbol{a}$ instead of a couple $\{a, n\}$. In the following, the index $\boldsymbol{a}$ labels all cluster decompositions with $\#(a) \geq 2$. The eigenvalues $\lambda^{\boldsymbol{a}}$ of the operators $H^a$ ($\lambda^a = 0$ if $\#(a) = N$) are called thresholds of the Schrödinger operator [8]. If all functions $V^\alpha(x^\alpha) \to 0$ as $|x^\alpha| \to \infty$, then the essential spectrum of the operator $H$ consists of the interval $[\lambda_0, \infty)$, where

$$\lambda_0 = \min_{\boldsymbol{a}} \lambda^{\boldsymbol{a}}$$

(the Hunziker–Van Winter–Zhislin theorem). Moreover, the eigenvalues of the operator $H$ may accumulate at its thresholds only.

The fundamental result of scattering theory for the $N$-particle Schrödinger operator can be formulated as follows. Let $\mathsf{P}^{\boldsymbol{a}}$ be the orthogonal projection in $L_2(X^a)$ on the subspace $\mathcal{H}_a^{(p)}$ spanned by all eigenvectors $\psi^{a,n}$ of $H^a$, and let $\mathsf{P}_{\boldsymbol{a}} = I \otimes \mathsf{P}^{\boldsymbol{a}}$, where the tensor product is defined by eqn [10]. Then $\mathsf{P}_{\boldsymbol{a}}$ commutes with the operator $H_a$. Set also $K_0 = H_0$, $\mathsf{P}_0 = I$. Suppose that for all $\alpha$

$$|V^\alpha(x^\alpha)| \leq C(1 + |x^\alpha|)^{-\rho}, \quad \rho > 1 \qquad [11]$$

(the short-range assumption). Then, for all $\boldsymbol{a}$, the wave operators

$$W_{\boldsymbol{a}}^{\pm} = W^{\pm}(H, H_a; \mathsf{P}_{\boldsymbol{a}}) = \operatorname*{s-lim}_{t\to\pm\infty} e^{iHt} e^{-iH_a t} \mathsf{P}_{\boldsymbol{a}}$$

exist and are isometric on the ranges $\operatorname{Ran} \mathsf{P}_{\boldsymbol{a}}$ of projections $\mathsf{P}_{\boldsymbol{a}}$. The subspaces $\operatorname{Ran} W_{\boldsymbol{a}}^{\pm}$ are mutually orthogonal, and scattering is asymptotically complete:

$$\bigoplus_{\boldsymbol{a}} \operatorname{Ran} W_{\boldsymbol{a}}^{\pm} = \mathcal{H}^{(\text{ac})}$$

The singular continuous spectrum of $H$ is empty, so the absolutely continuous subspace $\mathcal{H}^{(\text{ac})}$ of the operator $H$ can be replaced by $\mathcal{H} \ominus \mathcal{H}^{(p)}$, where $\mathcal{H}^{(p)}$ is spanned by all eigenvectors of $H$.

These results can be reformulated in terms of scattering theory in a couple of spaces. Suppose that, for every $a$, eigenvectors $\psi^{a,n}$ are normalized and orthogonal if the corresponding eigenvalues $\lambda^{a,n}$ coincide. Let us introduce an auxiliary space

$$\hat{\mathcal{H}} = \bigoplus_{\boldsymbol{a}} \mathcal{H}_{\boldsymbol{a}}, \quad \mathcal{H}_{\boldsymbol{a}} = \mathcal{H}_a = L_2(X_a) \qquad [12]$$

and an auxiliary operator

$$\hat{H} = \bigoplus_{\boldsymbol{a}} K_{\boldsymbol{a}}, \quad K_{\boldsymbol{a}} = K_a + \lambda^{\boldsymbol{a}} \qquad [13]$$

in this space. Here and below, the sums are taken over all $\boldsymbol{a}$. We define an identification $\hat{J} : \hat{\mathcal{H}} \to \mathcal{H}$ by the relations

$$\hat{J} = \sum_{\boldsymbol{a}} J^{\boldsymbol{a}}, \quad J^{\boldsymbol{a}} f_{\boldsymbol{a}} = f_{\boldsymbol{a}} \otimes \psi^{\boldsymbol{a}} \qquad [14]$$

where the tensor product is the same as in [10]. In particular, $J^0 = I$. Since $H_a J^{\boldsymbol{a}} = J^{\boldsymbol{a}} K_{\boldsymbol{a}}$, the wave operators $W^{\pm}(H, \hat{H}; \hat{J})$ exist and are isometric and complete, that is,

$$\operatorname{Ran} W^{\pm}(H, \hat{H}; \hat{J}) = \mathcal{H}^{(\text{ac})}$$

Thus, for states orthogonal to eigenvectors of $H$, evolution of an $N$-particle system decomposes

asymptotically into a sum of evolutions which are "free" in external variables $x_a$ and are determined by eigenvalues and eigenfunctions of the Hamiltonians $H^a$ in internal variables $x^a$. To be more precise, we have that, for all $f \in \mathcal{H}^{(\mathrm{ac})}$ and $t \to \pm\infty$,

$$\exp(-\mathrm{i}Ht)f = \sum_a \exp(-\mathrm{i}K_a t) f_a^{\pm} \otimes \psi^a + o(1) \qquad [15]$$

where

$$f_a^{\pm} = W^{\pm}(H, K_a; J^a)^* f$$

and the term $o(1)$ tends to zero in $\mathcal{H}$. The wave operator $W^{\pm}(H, K_a; J^a)$ describes the scattering channel where a system of $N$ interacting particles splits up asymptotically (for $t \to \pm\infty$) into noninteracting clusters $\mathcal{C}_1, \ldots, \mathcal{C}_n$, $n \geq 2$, and particles from the same cluster $\mathcal{C}_l$ are in the bound state (if there are more than one particle in $\mathcal{C}_l$) given by the function $\psi^a(x^a)$. Somewhat loosely speaking, this implies that the continuous spectrum of the operator $H$ consists of branches starting from all its thresholds.

Note that the scattering problem can equivalently be formulated without the separation of center-of-mass motion. In this case, a trivial decomposition with $\#(a) = 1$ should be added, and the set of thresholds of the operator $\boldsymbol{H}$ includes eigenvalues of the operator $H$.

The existence of the wave operators and their isometricity can be obtained by the Cook method. Only the asymptotic completeness is a difficult mathematical problem. It can be solved within the framework of the smooth method, which requires a study of boundary values of resolvents as the spectral parameter $z$ approaches the continuous spectrum or, equivalently, a study of a large-time behavior of evolution operators.

The scattering operator

$$\boldsymbol{S} = W^+(H, \hat{H}; \hat{J})^* W^-(H, \hat{H}; \hat{J})$$

is unitary on the space $\hat{\mathcal{H}}$ and commutes with the operator $\hat{H}$. Its component $\boldsymbol{S}_{ab} : \mathcal{H}_b \to \mathcal{H}_a$ describes a process where a system in a state $\boldsymbol{b}$ as $t \to -\infty$ goes over in a state $\boldsymbol{a}$ as $t \to +\infty$. Diagonalizing the operator $\hat{H}$ by a unitary operator $\hat{F}, (\hat{F}\hat{H}f)(\lambda) = \lambda(\hat{F}f)(\lambda), \lambda > \lambda_0$, we obtain the scattering matrix $S(\lambda)$ defined by the equation $(\hat{F}\boldsymbol{S}f)(\lambda) = S(\lambda)(\hat{F}f)(\lambda)$. In its turn, the scattering matrix is also a matrix operator with components $S_{ab}(\lambda)$. For $N \geq 3$, the structure of the scattering matrix is essentially more complicated than for $N = 2$. This is discussed in some detail in the next section.

## Resolvent Equations for Three-Particle Systems

Let the Hamiltonian $\boldsymbol{H}$ be defined by eqns [2]–[4], where $N = 3$, and let the configuration space of each particle be $\mathbb{R}^d$, $d \geq 3$. The operator $H$ acts in the space $\mathcal{H} = L_2(X^{\mathrm{cm}})$, where the subspace $X^{\mathrm{cm}} \subset \mathbb{R}^{3d}$ is distinguished by condition [5]. Let $R_0(z) = (H_0 - z)^{-1}, R(z) = (H - z)^{-1}$. Since $V(x)$ does not tend to 0 as $|x| \to \infty, x \in X^{\mathrm{cm}}$, in the three-particle case, the resolvent equation

$$R(z) = R_0(z) - R_0(z) V R(z) \qquad [16]$$

is not Fredholm even for $\operatorname{Im} z \neq 0$.

To overcome this difficulty, Faddeev (1965) derived a system of equations for components of the resolvent. The entries of this system are constructed in terms of three Hamiltonians

$$H_\alpha = H_0 + V^\alpha$$

$\alpha = (12), (13), (23)$, containing only one pair interaction each, and their resolvents $R_\alpha(z) = (H_\alpha - z)^{-1}$. Let us write down the resolvent equation for each pair $H_\alpha, H$

$$R(z) = R_\alpha(z) - R_\alpha(z) \sum_{\beta \neq \alpha} V^\beta R(z)$$

We multiply it by $|V^\alpha|^{1/2}$ and set

$$\boldsymbol{r}_\alpha^0(z) = |V^\alpha|^{1/2} R_\alpha(z), \quad \boldsymbol{r}_\alpha(z) = |V^\alpha|^{1/2} R(z)$$
$$\boldsymbol{t}_{\alpha,\alpha}(z) = 0, \quad \boldsymbol{t}_{\alpha,\beta}(z) = |V^\alpha|^{1/2} R_\alpha(z) (V^\beta)^{1/2}$$

where $(V^\beta)^{1/2} = V^\beta |V^\beta|^{-1/2}$. This yields a system of equations

$$\boldsymbol{r}_\alpha(z) = \boldsymbol{r}_\alpha^0(z) - \sum_{\beta \neq \alpha} \boldsymbol{t}_{\alpha,\beta}(z) \boldsymbol{r}_\beta(z) \qquad [17]$$

for the operators $\boldsymbol{r}_\alpha(z)$. Note that the resolvent $R(z)$ can be recovered from its components $\boldsymbol{r}_\alpha(z)$ by the formula

$$R(z) = R_0(z) - R_0(z) \sum_\alpha (V^\alpha)^{1/2} \boldsymbol{r}_\alpha(z)$$

It is convenient to rewrite eqn [17] in the matrix notation

$$\boldsymbol{r}(z) = \boldsymbol{r}^0(z) - \boldsymbol{t}(z) \boldsymbol{r}(z) \qquad [18]$$

where $\boldsymbol{r}^0(z) = \{\boldsymbol{r}_\alpha^0(z)\}, \boldsymbol{r}(z) = \{\boldsymbol{r}_\alpha(z)\}$ are the "vector" operators in the three-component space $L_2^{(3)}(X^{\mathrm{cm}})$ and $\boldsymbol{t}(z) = \{\boldsymbol{t}_{\alpha,\beta}(z)\}$ is the "matrix" operator in this space.

The advantage of eqn [17] compared to [16] is that the operators $\boldsymbol{t}_{\alpha,\beta}(z)$ are compact for $\operatorname{Im} z \neq 0$. This can be deduced from the fact that the product $V^\alpha(x^\alpha) V^\beta(x^\beta)$, where $\alpha \neq \beta$ tends to 0 as

$|x| \to \infty, x \in X^{cm}$, provided that $V^\alpha(x^\alpha) \to 0$ as $|x^\alpha| \to \infty$ for all $\alpha$. Moreover, the homogeneous equation [17] has only a trivial solution. Indeed, if for some $z$ with $\mathrm{Im}\, z \neq 0$

$$f_\alpha = -\sum_{\beta \neq \alpha} \boldsymbol{t}_{\alpha,\beta}(z) f_\beta \qquad [19]$$

then the function

$$u = \sum_\alpha (V^\alpha)^{1/2} f_\alpha$$

satisfies the equation $u = -R_0(z)Vu$. Since the operator $H$ is self-adjoint, this implies that $u = 0$ and hence $f_\alpha = 0$ for all $\alpha$. According to the Fredholm alternative, eqns [17] for $\boldsymbol{r}_\alpha(z)$ or [18] for $\boldsymbol{r}(z)$ can be solved if $\mathrm{Im}\, z \neq 0$, that is,

$$\boldsymbol{r}(z) = (I + \boldsymbol{t}(z))^{-1} \boldsymbol{r}^0(z) \qquad [20]$$

This equation allows one to deduce the existence of necessary boundary values of the "sandwiched" resolvent $R(z)$ from similar results for the resolvents $R_\alpha(z)$ of the "two-particle" operators $H_\alpha$. In its turn, $R_\alpha(z)$ can be expressed in terms of the resolvent $R^\alpha(z)$ of the operator $H^\alpha$ acting in the space $L_2(\mathbb{R}^d)$. Indeed, in the "mixed" representation $(\xi_\alpha, x^\alpha)$, where the Fourier transform in the variable $x_\alpha$ is performed and the variable $\xi_\alpha$ is dual to $x_\alpha$, we have

$$(R_\alpha(z)f)(\xi_\alpha, x^\alpha) = (R^\alpha(z - (2m_\alpha)^{-1}|\xi_\alpha|^2)f) \times (\xi_\alpha, x^\alpha) \qquad [21]$$

The passage to the limit $\mathrm{Im}\, z \to 0$ requires that assumption [11] be satisfied for $\rho > 2$. Moreover, we have to suppose that the operators $H^\alpha$ do not have the so-called zero-energy resonances as well as eigenvalues embedded in the continuous spectrum. Then the operator functions $\langle x^\alpha \rangle^{-l} R^\alpha(z) \langle x^\alpha \rangle^{-l}, l > 1$, $\langle x^\alpha \rangle = (1 + |x^\alpha|^2)^{1/2}$, are analytic in the complex plane cut along $[0, \infty)$, they have poles only at the points $\lambda^{\alpha,n}$, and are continuous up to the cut, the point $z = 0$ included. In particular, it follows from eqn [21] that, if the operators $H^\alpha$ do not have negative eigenvalues, then the operator functions $\langle x^\alpha \rangle^{-l} R_\alpha(z) \langle x^\alpha \rangle^{-l}, l > 1$, are also analytic in the complex plane cut along $[0, \infty)$ and are continuous up to the cut.

The next result is of genuinely three-particle nature and is crucial for the study of the operator $\boldsymbol{t}(z)$. The operator functions $\langle x^\alpha \rangle^{-l} R_0(z) \langle x^\beta \rangle^{-l}, \alpha \neq \beta, l > 1$, are continuous in norm up to the cut along $[0, \infty)$.

Now it follows from eqn [20] that the operator-valued functions $\boldsymbol{r}_\alpha(z)|V^\alpha|^{1/2}$ are continuous up to the cut $(0, \infty)$ except points $\lambda \in (0, \infty)$, where the homogeneous equation [19] for $z = \lambda \pm \mathrm{i}0$ has a nontrivial solution. The set $\mathcal{N} = \mathcal{N}_+ \cup \mathcal{N}_-$ of such points $\lambda \in (0, \infty)$ is closed and has Lebesgue measure zero. In particular, the operators $\langle x^\alpha \rangle^{-l}$, $l > 1$, are $H$-smooth on any compact subinterval of $\Lambda = (0, \infty) \backslash \mathcal{N}$. Therefore, the smooth method of scattering theory can be directly applied. It yields the existence and completeness of the wave operators $W_\pm(H, H_0)$. In this case, three-particles are necessarily asymptotically free.

"Two-particle" channels of scattering arise if the operators $H^\alpha$ have negative eigenvalues. To simplify notation, we assume that every $H^\alpha$ has exactly one eigenvalue $\lambda^\alpha < 0$. Moreover, it is supposed that the corresponding eigenfunction $\psi^\alpha(x^\alpha)$ tends to zero sufficiently rapidly as $|x^\alpha| \to \infty$. Analytically, the appearance of new channels is due to new singularities of the resolvents. Indeed, in this case

$$R^\alpha(z) = (\lambda^\alpha - z)^{-1} P^\alpha + \hat{R}^\alpha(z)$$

where the function $\hat{R}^\alpha(z)$ is analytic and continuous up to the cut in the complex plane cut along $[0, \infty)$. It follows from eqn [21] that in this case the resolvent $R_\alpha(z)$ contains the additional term

$$((2m_\alpha)^{-1}|\xi_\alpha|^2 + \lambda^\alpha - z)^{-1} \otimes P^\alpha$$

which is analytic only in the complex plane cut along $[\lambda^\alpha, \infty)$. To take these terms into account, system [17] should be further rearranged. This yields the following result. Let us set

$$\mathsf{G}_{\alpha 0} = \langle x^\alpha \rangle^{-l}(I - P_\alpha), \quad \mathsf{G}_{\alpha 1} = \langle x_\alpha \rangle^{l} (J^\alpha)^* \sum_{\beta \neq \alpha} V^\beta \qquad [22]$$

Then, for all $\alpha, \beta, i, j = 0, 1$, a suitable $l > 1$ and $\lambda_0 = \min\{\lambda^\alpha\}$, the operator functions $\mathsf{G}_{\alpha i} R(z) \mathsf{G}^*_{\beta j}$ are norm continuous as $z$ approaches the cut $(\lambda_0, \infty)$ at the points of $\Lambda = (\lambda_0, \infty) \backslash \mathcal{N}$, where $\mathcal{N}$ is again a closed set of measure zero. In particular, the operators $\mathsf{G}_{\alpha 0}$ and $\mathsf{G}_{\alpha 1}$ are $H$-smooth on any compact subinterval of $\Lambda$.

In the multichannel case, to fit scattering for the Hamiltonian $H$ into the framework of smooth theory, it is convenient to reformulate the result in terms of scattering theory in a couple of spaces. Let the space $\hat{\mathcal{H}}$, the operator $\hat{H}$, and the identification $\hat{J}$ be defined by eqns [12], [13], and [14], respectively, where the index $\boldsymbol{a}$ takes four values $\boldsymbol{a} = 0, \alpha$ and $\alpha = (12), (13), (23)$. One, further, needs to introduce auxiliary identifications

$$J^0 = I - \sum_\alpha P_\alpha$$

and

$$\hat{J} = J^0 \oplus \bigoplus_\alpha J^\alpha$$

The $H$- (and $\hat{H}$-) smoothness of operators [22] imply that the wave operators

$$W^{\pm}(H, \hat{H}; \hat{\boldsymbol{J}}) \quad \text{and} \quad W^{\pm}(\hat{H}, H; \hat{\boldsymbol{J}}^*)$$

exist.

The operators $W^{\pm}(H, \hat{H}; \hat{\boldsymbol{J}})$ are isometric because

$$\operatorname*{s-lim}_{|t| \to \infty} P_\alpha \exp(-\mathrm{i}H_0 t) = 0 \qquad [23]$$

and the operators $P_\alpha P_\beta$ are compact for $\alpha \neq \beta$. Using that the operator

$$\hat{\boldsymbol{J}}\hat{\boldsymbol{J}}^* - I = \sum_{\alpha \neq \beta} P_\alpha P_\beta$$

is compact (whereas $\hat{J}\hat{J}^* - I$ is not), we see that the operators $W^{\pm}(\hat{H}, H; \hat{\boldsymbol{J}}^*)$ are also isometric. Finally, we remark that, by eqn [23],

$$W^{\pm}(H, \hat{H}; \hat{\boldsymbol{J}}) = W^{\pm}(H, \hat{H}, \hat{J})$$

This implies the asymptotic completeness.

Let us discuss properties of the scattering matrix in the one-channel case where the pair operators $H^\alpha$ do not have negative eigenvalues. The scattering matrix $S(\lambda): L_2(\mathbb{S}^{2d-1}) \to L_2(\mathbb{S}^{2d-1})$, $\lambda > 0$, is of course a unitary operator, but in contrast to the two-particle case the operator $S(\lambda) - I$ is not compact because its kernel contains the Dirac functions $\delta(\xi_\alpha - \xi'_\alpha)$. Nevertheless, the structure of its singularities can be explicitly described. Actually, let $S_\alpha(\lambda)$ be the "two-particle" scattering matrix for the pair $H_0, H_\alpha$. Then

$$S(\lambda) = S_{12}(\lambda) S_{23}(\lambda) S_{13}(\lambda) \tilde{S}(\lambda)$$

where the operator $\tilde{S}(\lambda) - I$ is compact.

The approach described briefly in this section relies on a kind of an advanced perturbation theory where the free problem is determined by the set of all sub-Hamiltonians. Its generalization to the case of an arbitrary number of particles meets with numerous difficulties. A different, nonperturbative, approach which works well for any number of particles will be discussed in the next section.

A purely time-dependent method in three-particle scattering is exposed in Enss (1983).

## Nonperturbative Approach

Now $N$ and $d$ are arbitrary. In the nonperturbative approach (see Graf (1990), Sigal and Soffer (1989), and Yafaev (1993)) the operators $H$ and $H_0$ as well as the Hamiltonians of all subsystems are treated on an equal basis. It is supposed that all pair potentials satisfy condition [11]. No assumptions on subsystems are required.

The starting point of this approach is the limiting-absorption principle, which claims that the operator $\langle x \rangle^{-l}$, $x \in X^{\text{cm}}$, for $l > 1/2$ is $H$-smooth on any compact interval $\Lambda$ not containing the thresholds and eigenvalues of $H$. Its proof relies on the Mourre commutator method (see Cycon *et al.* (1987)). To be more precise, it is deduced from the following estimate:

$$\mathrm{i}([H, A]f, f) \geq c\|f\|^2, \quad c = c(\lambda) > 0$$
$$f \in E(\Lambda_\lambda)\mathcal{H} \qquad [24]$$

for the commutator of $H$ with the generator of translations

$$A = -\mathrm{i} \sum_j (x_j \partial_j + \partial_j x_j)$$

Here $x_j$ are coordinates of $x \in X^{\text{cm}}$ in some orthonormal (with respect to scalar product [6]) basis in $X^{\text{cm}}$, $\lambda$ is neither a threshold nor an eigenvalue of the operator $H$ and $\Lambda_\lambda$ is a sufficiently small interval. Very roughly speaking, the Mourre estimate [24] means that, similarly to the two-particle case, the observable

$$(A e^{-\mathrm{i}Ht} f, e^{-\mathrm{i}Ht} f)$$

is a strictly increasing function of $t$ for all $f \in \mathcal{H}^{(\text{ac})}$.

The limiting-absorption principle implies that the singular continuous spectrum of the operator $H$ is empty, but it is not sufficient for scattering theory. If the limiting-absorption principle were true for the critical value $l = 1/2$, then it would imply asymptotic completeness. Unfortunately, the operator $\langle x \rangle^{-1/2}$ is definitely not smooth even with respect to the free operator $H_0$. However, by introducing an auxiliary differential operator we can fix this problem. This leads to the radiation estimates. These estimates look differently in different regions of the configuration space. Choose any cluster decomposition $a = (\mathcal{C}_1, \ldots, \mathcal{C}_n)$. The radiation estimate morally implies that the motion of a system is asymptotically free in the variable $x_a$ (describing the relative motion of clusters) in the region where particles from each cluster $\mathcal{C}_l$, $l = 1, \ldots, n$, are close to each other compared to distances between different clusters. On the contrary, this motion is very complicated in the variable $x^a$ pertaining to bound states of different clusters. In particular, the radiation estimate is the same as for the two-particle case in the "free" region where all particles are far from each other.

To be more precise, let $\nabla_a = \nabla_{x_a}$ be the gradient in the variable $x_a$ and let $\nabla_a^\perp$,

$$(\nabla_a^\perp u)(x) = (\nabla_a u)(x) - |x_a|^{-2} \langle (\nabla_a u)(x), x_a \rangle x_a$$

be its orthogonal projection in $X_a$ on the subspace orthogonal to the vector $x_a$. Let $\chi_a$ be the

characteristic function of a closed cone $Y_a \subset X^{\rm cm}$ satisfying the condition $Y_a \cap X_b = \emptyset$ for all $b$ such that $X_a \not\subset X_b$. Then the operator

$$G_a = \chi_a \langle x \rangle^{-1/2} \nabla_a^{\perp}$$

is $H$-smooth on $\Lambda$.

A proof of the radiation estimates is based on the consideration of the commutator of $H$ with some differential operator $M = -{\rm i}\sum (m^{(j)}\partial_j + \partial_j m^{(j)})$, where $m^{(j)} = \partial m/\partial x_j$. Here $m$ (it depends on $a$) is a specially constructed function satisfying the following properties:

1. $m(x)$ is homogeneous (for $|x| \geq 1$) of order 1;
2. for any $b$ it does not depend on $x^b$ in some conical neighborhood of the subspace $X_b$;
3. $m(x)$ is convex; and
4. $m(x) = \mu_a |x_a|, \mu_a \geq 1$, on support of the function $\chi_a$.

Note that we can set $m(x) = |x|$ in the case of the operator $H_0$.

Due to properties (1) and (2) the commutator $[V, M]$ is a short-range function (estimated by $\langle x \rangle^{-1-\varepsilon}$ for $\varepsilon > 0$). Due to properties (3) and (4) the commutator $[H_0, M] \geq c G_a^* G_a, c > 0,$ up to short-range terms. The estimate

$$[H, M] \geq c G_a^* G_a - c_1 \langle x \rangle^{-1-\varepsilon}$$

implies that the operator $G_a$ is $H$-smooth on $\Lambda$.

The main difficulty in the $N$-particle problem is that pair potentials $V^{\alpha}(x^{\alpha})$ do not tend to zero as $|x| \to \infty$. The idea of the proof of asymptotic completeness is to introduce auxiliary wave operators such that "effective" perturbations are decaying functions. This requires a suitable smooth partition of unity. Moreover, it is convenient to choose auxiliary identifications as first-order differential operators rather than operators of multiplication. Unfortunately, although such identifications allow one to "kill" directions where the potentials $V^{\alpha}(x^{\alpha})$ do not tend to zero, their commutators with the operator $H_0$ have coefficients decaying at infinity only as $|x|^{-1}$.

Thus, we introduce differential operators

$$M_a = -{\rm i}\sum \left( m_a^{(j)} \partial_j + \partial_j m_a^{(j)} \right)$$

with coefficients $m_a^{(j)} = \partial m_a / \partial x_j$. The functions $m_a$ satisfy properties (1), (2) formulated above and

5. $m_a(x) = 0$ in some conical neighborhoods of the subspaces $X_b$ such that $X_a \not\subset X_b$. To put it differently, $m_a(x) = 0$ in some conical neighborhood of the subspace where $x_i = x_j$ for some $i, j$ belonging to different clusters $\mathcal{C}_1, \ldots, \mathcal{C}_n$.

Let the operator $H_a$ be defined by eqn [9]. Given the limiting-absorption principle and the radiation estimates, we first check the existence of auxiliary wave operators

$$W^{\pm}(H, H_a; M_a E_a(\Lambda))$$

and

$$W^{\pm}(H_a, H; M_a E(\Lambda)) \qquad [25]$$

Here we use that according to (5) coefficients of the differential operator $(V - V^a)M_a$ are, under assumption [11], short-range (in the configuration space $X^{\rm cm}$). By property (2), the function $[V^a, M_a]$ is also short-range. Thus, the operator $VM_a - M_a V^a$ can be taken into account by the limiting-absorption principle. The commutator $[H_0, M_a]$ factorizes into a product of $H_a$- and $H$-smooth operators according to the radiation estimates.

Similar arguments show that, for $\sum_a m_a = m$ and $M = \sum_a M_a$ (the sums here are taken over all possible breakups of the $N$-particle system), the wave operator (observable)

$$W^{\pm}(H, H; \pm M E(\Lambda)) \qquad [26]$$

also exists. Moreover, it can be easily achieved that $m(x) \geq 1$. Then it follows from the Mourre estimate that operator [26] is positive definite on the subspace $E(\Lambda)\mathcal{H}$ and hence its range coincides with this subspace. It means that for all $f \in E(\Lambda)\mathcal{H}$

$$\lim_{t \to \pm\infty} \| \exp(-{\rm i}Ht)f - M \exp(-{\rm i}Ht) g^{\pm} \| = 0 \qquad [27]$$

if $f = W^{\pm}(H, H; ME(\Lambda)) g^{\pm}$.

The existence of wave operators [25] implies that for any $g^{\pm} = E(\Lambda) g^{\pm}$ and $g_a^{\pm} = W^{\pm}(H_a, H; M_a E(\Lambda)) g^{\pm}$

$$\lim_{t \to \pm\infty} \| M \exp(-{\rm i}Ht) g^{\pm} - \sum_a \exp(-{\rm i}H_a t) g_a^{\pm} \| = 0 \qquad [28]$$

Combining eqns [27] and [28], we see that $\exp(-{\rm i}Ht)f$ decomposes asymptotically into simpler evolutions $\exp(-{\rm i}H_a t) g_a^{\pm}$. This is one of the equivalent formulations of asymptotic completeness and leads to eqn [15].

Finally, we note that eqn [15] can be rewritten as

$$\exp(-{\rm i}Ht)f = \sum_a \exp({\rm i}\Phi_a(x_a, t))(2{\rm i}t)^{-d_a/2} \times \hat{f}_a^{\pm}(x_a/(2t)) \psi^a(x^a) + o(1) \qquad [29]$$

where $t \to \pm\infty$, $d_a = \dim X_a$, $\hat{f}_a^{\pm}$ is the Fourier transform of $f_a^{\pm}$ and

$$\Phi_a(x_a, t) = x_a^2(4t)^{-1} - \lambda^a t \qquad [30]$$

## Long-Range Interactions: New Channels

The multiparticle problem acquires a long-range character if pair potentials decay as Coulomb potentials or slower. Similarly to the two-particle problem, for long-range potentials the definition of wave operators should be naturally modified. As in the short-range case, only the asymptotic completeness is a really difficult mathematical problem. Assume that pair potentials satisfy condition

$$|(\partial^\kappa V^\alpha)(x^\alpha)| \le C(1 + |x^\alpha|)^{-\rho-|\kappa|}, \quad \rho > \sqrt{3} - 1$$

for all $|\kappa| \le \kappa_0$ and sufficiently large $\kappa_0$. Then only phase factors in eqn [29] should be modified. Actually, instead of eqn [30] we should set

$$\Phi_a(x_a, t) = x_a^2(4t)^{-1} - \lambda^a t - t\int_0^1 V_a(sx_a, 0)\, ds$$

where $V_a(x) = V(x) - V^a(x)$ and as usual $x = (x_a, x^a)$. As shown in Dereziński (1993), with this definition of wave operators, the asymptotic completeness holds.

On the contrary, if pair potentials decay slower than $|x|^{-1/2}$, then the traditional picture of scattering breaks down (see Yafaev (1996)). Actually, a three-particle system might have additional scattering channels intermediary between the channel where three particles are asymptotically free and the channels where a couple of particles form a bound state. In these additional channels, the bound state of a couple of particles depends on a position of the third particle, and it is destroyed asymptotically.

*See also:* Quantum Mechanical Scattering Theory; Schrödinger Operators.

## Further Reading


Cycon H, Froese R, Kirsh W, and Simon B (1987) *Schrödinger Operators*, Texts and Monographs in Physics. Berlin: Springer.

Dereziński J (1993) Asymptotic completeness of long-range quantum systems. *Annals of Mathematics* 138: 427–473.

Dereziński J and Gérard C (1997) *Scattering Theory of Classical and Quantum N Particle Systems*. Berlin: Springer.

Enss V (1983) Completeness of three-body quantum scattering. In: Blanchard P and Streit L (eds.) *Dynamics and Processes*, Springer Lecture Notes in Mathematics, vol. 1031, pp. 62–88.

Faddeev LD (1965) *Mathematical Aspects of the Three Body Problem in Quantum Scattering Theory*, Israel Program of Sci. Transl.

Graf GM (1990) Asymptotic completeness for *N*-body short-range quantum systems: a new proof. *Communications in Mathematical Physics* 132: 73–101.

Reed M and Simon B (1979) *Methods of Modern Mathematical Physics III*. New York: Academic Press.

Sigal IM and Soffer A (1987) The *N*-particle scattering problem: asymptotic completeness for short-range systems. *Annals of Mathematics* 126: 35–108.

Yafaev DR (1993) Radiation conditions and scattering theory for *N*-particle Hamiltonians, *Communications in Mathematical Physics* 154: 523–554.

Yafaev DR (1996) New channels of scattering for three-body quantum systems with long-range potentials, *Duke Mathematical Journal* 82: 553–584.

Yafaev DR (2000) *Scattering Theory: Some Old and New Problems*, Springer Lecture Notes in Mathematics, vol. 1735.


# Nuclear Magnetic Resonance


**P T Callaghan**, Victoria University of Wellington, Wellington, New Zealand




## Introduction

The existence of nuclear spin and its associated magnetism was first suggested by Wolfgang Pauli in 1924, a conjecture based on the fine details of atomic spectra, the so-called hyperfine structure. The interaction of this nuclear magnetism with an external magnetic field was predicted to result in a finite number of discrete energy levels known as the Zeeman structure. However, the first direct excitation of transitions between nuclear Zeeman levels was by Isador Rabi in 1933, using radiofrequency (RF) waves in an atomic beam apparatus. In 1945, Felix Bloch and co-workers at Stanford, and Edward Purcell and co-workers at MIT, performed the first nuclear magnetic resonance (NMR) experiments in condensed matter, with the RF response of the hydrogen nucleus (proton) being directly detected.

The early prospects for this new technique were limited to precise measurements of magnetic fields and nuclear magnetic moments. However, three transformational discoveries intervened to set NMR on a course that would result in initially unimaginable contributions to physics, chemistry,

engineering, medicine, geology, food science, and biochemistry. In 1950, it was found that atomic nuclei at different sites of a molecular orbital had slightly different resonant frequencies, a phenomenon known as "chemical shift." In the a same year, Erwin Hahn discovered the spin echo, thus opening the possibility that multiple RF pulse trains could be used to remove unwanted nuclear spin interactions while being used to manipulate spin coherences with exquisite resolution. In addition, in 1951, using this spin echo, Herbert Gutowsky and Charles Slichter revealed a hitherto unobserved scalar spin–spin interaction between nuclei, mediated by the molecular orbital electrons.

The discovery of the chemical shift and the scalar coupling would immediately revolutionize chemistry. Further discoveries of nuclear quadrupole interactions and through-space dipolar interactions would add to the capacity of NMR to provide insight regarding structure and order in the solid and liquid crystalline state. But the spin echo would provide a platform for new advances in science in every one of the six decades following the discovery of NMR in 1945. These were successively diffusion and flow NMR, multidimensional NMR, magnetic resonance imaging, protein structure NMR, *ex situ* NMR, and quantum computing NMR.

## Resonant Excitation and Detection

In quantum-mechanical language, the Zeeman Hamiltonian $H$ for a nuclear spin experiencing a magnetic field $\boldsymbol{B}_0$ along the laboratory $z$-axis may be written as

$$H = -\gamma B_0 I_z \qquad [1]$$

$\gamma$ being the (nuclear) gyromagnetic ratio while $I_z$ is the operator for the $z$-component of angular momentum, with eigenvalues $m\hbar$, $m$ lying in the range $-I, -I+1, \ldots, I$. $I$ is the angular momentum quantum number, being either integer or half-integer. From the Schrödinger equation, it can be seen that the eigenkets of $H$ precess about the $z$-axis at a rate $\gamma B_0$, the frequency corresponding to the energy difference between the $2I+1$ Zeeman levels. For convenience, we shall take the eigenvalues of $I_z$ to be simply $m$, dropping the factor $\hbar$, and leading to a Hamiltonian expressed in frequency rather than in energy units.

Resonant excitation between the Zeeman levels is achieved by the application of an RF ($\omega$) magnetic field of amplitude $2B_1$ linearly polarized normal to $\boldsymbol{B}_0$ such that the total Hamiltonian becomes

$$H = -\gamma B_0 I_z - 2\gamma B_1 \cos \omega t I_x \qquad [2]$$

This excitation is easily applied by means of a transversely oriented antenna coil, the same coil generally being used to detect the nuclear spin response. In the frame of reference rotating about $\boldsymbol{B}_0$ at $\omega$, the Hamiltonian transforms to

$$H = -\gamma\left(B_0 - \frac{\omega}{\gamma}\right) I_z - \gamma B_1 I_x - \gamma B_1 \exp(\mathrm{i}2\omega t I_z) I_x \exp(-\mathrm{i}2\omega t I_z) \qquad [3]$$

At resonance, $\omega = \omega_0 = \gamma B_0$. The last term in eqn [3] averages to zero and may be neglected (the Heisenberg condition) provided $\omega \gg \gamma B_1$, that is, $B_0 \gg B_1$. Given $B_0$ of the order of tesla and $B_1$ of the order of millitesla, this condition is easily satisfied. Hence, from the perspective of the rotating frame, the spins at resonance see only the static magnetic interaction $\gamma B_1 I_x$, so that application of this resonant RF field causes spins to nutate about the rotating frame $x$-axis at a rate $\gamma B_1$. Thus, by application of RF pulses of different duration, and phases, one may produce arbitrary reorientation of the spins about various axes in the rotating frame.

With the spin system disturbed from equilibrium, the NMR "signal" is detected via the subsequent free precession, and usually via the same antenna coil used for resonant excitation, Semiclassically, the phenomenon may be pictured as follows. RF excitation nutates an initial $z$-magnetization into the transverse plane of the rotating frame. Such transverse magnetization corresponds the laboratory frame to a magnetization precessing at the Larmor frequency, thus inducing an oscillating emf in the receiver coil. In the next section, we see how to describe this phenomenon in the language of quantum mechanics.

Typically, NMR is performed using the nuclei of common atoms in organic molecules, ($^1$H, $^2$H, $^{13}$C, $^{15}$N, $^{19}$F, $^{31}$P) although for inorganic matter a wider class of nuclei are available. Of all these, the proton is most abundant and most sensitive, having the highest gyromagnetic ratio, $\gamma$, of all stable nuclei.

## The Quantum Statistics of the Spin Ensemble

The nuclear Zeeman energy in typically available laboratory magnetic fields, $\gamma B_0 \hbar$, is many orders of magnitude smaller than the Boltzmann energy, $k_B T$, except at millikelvin temperatures. At room temperature in thermal equilibrium, the fractional difference in populations between the Zeeman levels

is normally very small, for example, for protons, about $10^{-5}$. Of course, the total number of spins available may be very large, for example, on the order of $10^{20}$.

The signal in magnetic resonance is detected as a collective effect of the large ensemble of nuclear spins. The natural language of quantum statistics is that of the density matrix, $\rho$; the time-dependent expectation value for any observable represented by an operator $O$ is then, $\mathrm{tr}(O\rho(t))$, the diagonal sum of the product of $O$ and $\rho$. The time evolution of the density matrix is given by the Liouville equation

$$\mathrm{i}\frac{\partial \rho}{\partial t} = [H, \rho] \qquad [4]$$

where $[\,,]$ is a commutator. For a constant Hamiltonian, this equation gives

$$\rho(t) = \exp(\mathrm{i}Ht)\rho(0)\exp(-\mathrm{i}Ht) \qquad [5]$$

Physical solutions to the density matrix (Liouville space) are $(2I+1)^2$ (square) matrices formed in the $(2I+1)$-dimensional angular momentum eigenbasis. Generally, we may write the density matrix in a representation of irreducible tensor operators. One very convenient representation is the set formed by taking products of spin operators. For example, in the case of spin-1/2 where Liouville space is $2^2$-dimensional, we may write

$$\rho(t) = \tfrac{1}{2}I + a_x I_x + a_y I_y + a_z I_z \qquad [6]$$

where $I$ is the identity operator. The operators $I_x$ and $I_y$ provide the off-diagonal elements of $\rho$ and define the degree of phase coherence in the ensemble, while the operator $I_z$ defines the degree to which the diagonal elements differ, thus defining the polarization. $a_x$ and $a_y$ give the amount of "one-quantum coherence" in the ensemble while $a_z$ gives the polarization. In thermal equilibrium $a_x = a_y = 0$, and the spin ensemble exists in a state of pure longitudinal polarization given, in the high-temperature approximation, $\gamma B_0 \hbar << k_B T$, by

$$\rho_{\mathrm{eqbm}}(0) \approx \frac{1}{(2I+1)}I + \frac{\gamma\hbar B_0}{(2I+1)k_B T}I_z \qquad [7]$$

This is the starting point for all NMR experiments (Figure 1).

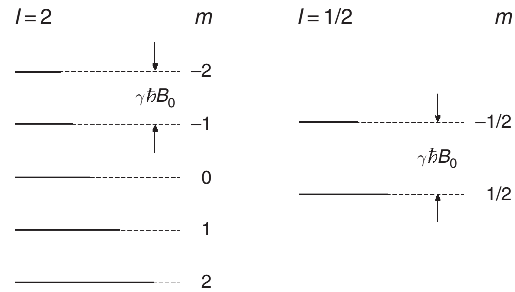



**Figure 1** Schematic Zeeman levels for the case $I = 2$ *and* $I = 1/2$. The bold lines indicate the relative population in each state in thermal equilibrium.

Consider then the detection of precession via the Faraday induction. The size of the signal observed will be proportional to the size of the transverse magnetization $M = \mathrm{tr}[(I_x + \mathrm{i}I_y)\rho(t)]$ present in the rotating frame, this magnetization producing an induced emf with real and imaginary components because of the capacity of heterodyne receivers to detect quadrature phase. In the laboratory frame, the detected signal has a prefactor of $\gamma B_0$ reflecting the Faraday induction, which, taken together with the dependence of the initial equilibrium magnetization on $\gamma B_0$, gives an overall NMR sensitivity $(\gamma B_0)^2$, helping to explain in part why high magnetic fields are advantageous. Take the simple example for $I = 1/2$, where a single 90° resonant RF pulse is applied to the spin system, subsequent free precession occurring under the Zeeman Hamiltonian. The density matrix at detection is

$$\begin{aligned}\rho(t) &= \exp(\mathrm{i}\omega_0 t I_z)\exp\left(\mathrm{i}\frac{\pi}{2}I_x\right)\rho_{\mathrm{eqbm}}(0)\\ &\quad\times\exp\left(-\mathrm{i}\frac{\pi}{2}I_x\right)\exp(-\mathrm{i}\omega_0 t I_z)\\ &= \exp(\mathrm{i}\omega_0 t I_z)\exp\left(\mathrm{i}\frac{\pi}{2}I_x\right)a_{\mathrm{eqbm}}I_z\\ &\quad\times\exp\left(-\mathrm{i}\frac{\pi}{2}I_x\right)\exp(-\mathrm{i}\omega_0 t I_z)\\ &= \exp(\mathrm{i}\omega_0 t I_z)a_{\mathrm{eqbm}}I_y\exp(-\mathrm{i}\omega_0 t I_z)\\ &= a_{\mathrm{eqbm}}I_y\cos(\omega_0 t) + a_{\mathrm{eqbm}}I_x\sin(\omega_0 t)\end{aligned} \qquad [8]$$

Noting $\mathrm{tr}(I_x^2) = \mathrm{tr}(I_y^2) = \mathrm{tr}(I_z^2) = (1/3)(2I+1)I(I+1)$ and $\mathrm{tr}(I_\alpha I_\beta) = 0$, the signal may easily be calculated as $S(t) : a_{\mathrm{eqbm}}\exp(\mathrm{i}\omega_0 t)$, corresponding, upon Fourier transformation, to a unique frequency at $\omega_0$. Note that a basis consisting of products of angular momentum operators are easy to handle since all evolution properties follow from the usual angular momentum commutation algebra.

The spin echo pulse scheme of Figure 2 is one of the most important in NMR. It allows one to refocus dephasing effects caused by inhomogeneous broadening, for example, due to the heterogeneity of the magnetic field across the sample. Rewriting the density matrix equation in the rotating frame, replacing the Zeeman precession

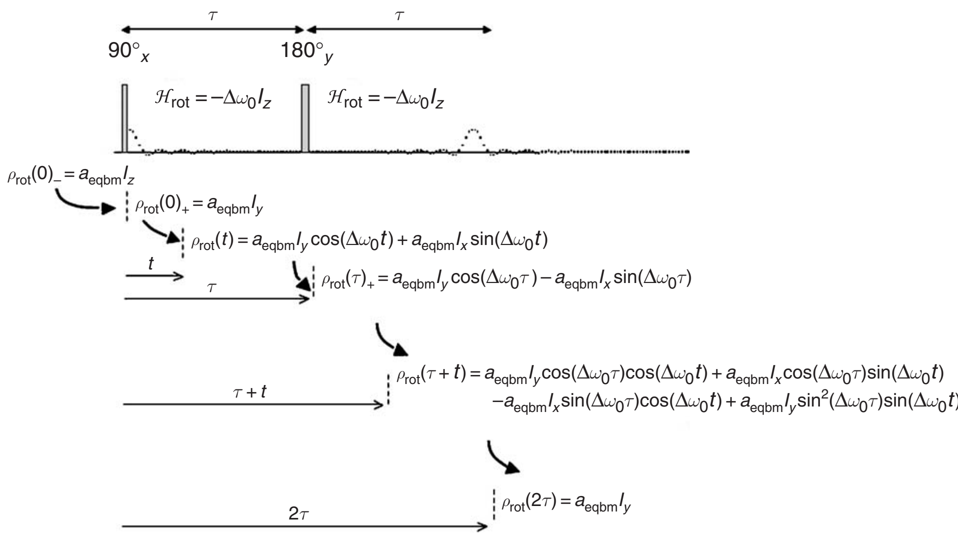



**Figure 2** Spin echo pulse scheme showing the evolution of the density matrix.

by its residual offset, and accounting for both RF pulses,

$$\begin{aligned}\rho_{\rm rot}(2\tau) &= \exp(\mathrm{i}\Delta\omega_0\tau I_z)\exp(\mathrm{i}\pi I_y)\exp(\mathrm{i}\Delta\omega_0\tau I_z)\\ &\quad\times \exp\left(\mathrm{i}\frac{\pi}{2}I_x\right)\rho_{\rm eqbm}(0)\exp\left(-\mathrm{i}\frac{\pi}{2}I_x\right)\\ &\quad\times \exp(-\mathrm{i}\Delta\omega_0\tau I_z)\exp(-\mathrm{i}\pi I_y)\\ &\quad\times \exp(-\mathrm{i}\Delta\omega_0\tau I_z)\\ &= a_{\rm eqbm} I_y \end{aligned} \quad [9]$$

Details of the density matrix evolution are given in Figure 2. The inversion pulse has the effect of completely reversing all the phase shifts that occur during the first interval, resulting in an echo signal when the two time periods are equal. Note the use of nested operators representing the successive influences of RF pulses (assumed to be ideal rotations) and Hamiltonian evolutions. The overall influence of the RF pulses is to render the effective Hamiltonian zero in this case.

This echo sequence (and its equivalent multiple RF train, the Carr–Purcell–Meiboom–Gill sequence) allows one to remove the effect of magnetic field inhomogeneities so as to investigate the underlying homogeneous broadening and associated signal damping.

## Spin Relaxation

The free precession of nuclear spins does not continue indefinitely. Ultimately the off-diagonal elements of the density matrix lose phase coherence while the diagonal elements gradually return to their thermal equilibrium state, two processes known, respectively, as $T_2$ (spin–spin) and $T_1$ (spin–lattice) relaxation. The rate of relaxation depends on interactions between the spins themselves and between the spins and their thermal environment. The process of $T_1$ relaxation requires fluctuations that induce transitions between the Zeeman levels. Clearly the relevant quantum-mechanical operators must possess a nonzero matrix element coupling the Zeeman levels, and the frequency of those fluctuations must match the energy gap spacing. Predominant in causing such relaxation in diamagnetic environments are the internuclear dipolar interactions, while in paramagnetic environments, dipolar interactions between nuclear and electronic spins are effective. One simple way of representing these processes is by the spectral density function, the Fourier power transform of their fluctuations, dipolar interactions causing spin–lattice relaxation due to fluctuations at $\omega_0$ and $2\omega_0$. For a fluctuating interaction with correlation time, $\tau_c$, that spectral density may approximate a Lorentzian of the form

$$J(\omega) = \frac{\tau_c}{1 + \omega^2\tau_c^2} \quad [10]$$

Thus, as the rate of molecular motions varies, due to the influence of temperature on $\tau_c$, the $T_1$ relaxation rate will be a maximum when $\omega_0\tau_c = 1$. Both solids ($\omega_0\tau_c \gg 1$) and liquids ($\omega_0\tau_c \ll 1$) have long $T_1$ relaxation times while soft solids or complex liquids may have faster relaxation. $T_1$ relaxation manifests as an exponential return to equilibrium values of longitudinal magnetization. Typical vales range from hundreds of milliseconds to hours, and the need to re-establish equilibrium between repetitions of the experiment can severely limit signal averaging

and hence available signal-to-noise ratios. Note that $T_1$ relaxation occurs by stimulated emission. Spontaneous emission is effectively absent from nuclear spin systems owing to the long-radiation wavelength.

The case of $T_2$ (spin–spin) relaxation is inherently more complex. First, the definition of "loss of phase coherence" depends on the particular RF pulse sequence employed. Second, the simple perturbation theory description applied to $T_1$ relaxation only works in the fast motion limit, where the $T_2$ relaxation rate may be shown to depend on spectral density terms not only at $\omega_0$ and $2\omega_0$ but also $\omega = 0$. In consequence, $T_2 \ll T_1$. $T_2$ relaxation is sensitive to static components. These static components may dominate in soft solids and solids. Indeed, any term in the Hamiltonian which spreads spin phases, and which cannot be recovered by means of a judicious RF pulse train, will contribute to $T_2$ relaxation. Suppose the effective frequency distribution causing dephasing is described by an ensemble second moment $<\Delta\omega^2>$, and exhibits fluctuations about a mean of zero with correlation time, $\tau_c$. Then we may identify two limiting cases: in the slow motion limit $<\Delta\omega^2>^{1/2}\ \tau_c \gg 1$, the decay of the detected magnetization is Gaussian, and given by a factor $\exp(-1/2 <\Delta\omega^2> t^2)$. In solids, the proton $T_2$ relaxation may take place in a few tens of microseconds. In the fast motion limit $<\Delta\omega^2>^{1/2}\ \tau_c \ll 1$, the decay of the detected magnetization is exponential, and given by a factor $\exp(-<\Delta\omega^2>\tau_c t)$. Liquid state $T_2$ values approach $T_1$ under extreme narrowing conditions.

## The Details of the Nuclear Spin Hamiltonian

Atomic nuclei interact with their environment, with surrounding electrons, and with other nuclear spins. It is precisely this feature that provides such a sensitive probe of material structure and dynamics. For a material immersed in a steady magnetic field $\boldsymbol{B}_0$ along the laboratory $z$-axis, the Hamiltonian for the $i$th nuclear spin can be written

$$H = -\gamma B_0 I_{iz} - \boldsymbol{I}_i.\underline{\underline{S}}.\boldsymbol{B}_0 + \sum_j J\,\boldsymbol{I}_i.\boldsymbol{I}_j + \sum_j \boldsymbol{I}_i.\underline{\underline{D}}.\boldsymbol{I}_j + \boldsymbol{I}_i.\underline{\underline{Q}}.\boldsymbol{I}_i \qquad [11]$$

It is the variety of the terms in the nuclear spin Hamiltonian that imparts power to NMR. The first is the nuclear Zeeman interaction with the applied magnetic field. In modern laboratory superconducting magnets, this interaction can be as large as 1000 MHz, although in earth field applications it can be as small as 2.5 kHz. Given that the sensitivity and resolution of NMR generally improve with increasing magnetic field, the range of 100–1000 MHz is typically the operating regime of choice. All other terms in the nuclear spin Hamiltonian are smaller and thus act as first-order perturbations only, projecting their quantum operators into the zeroth-order Zeeman eigenbasis, the quantum frame of the operator $I_z$. Because several of the terms in $H$ depend on the orientation of the local nuclear environment (e.g., the molecular orbital) with respect to the magnetic field, these terms will fluctuate in the presence of reorientational motions. By the Heisenberg uncertainty principle, fluctuations faster in frequency than the size of the Hamiltonian contribution, expressed in frequency units, will result in an averaging to the mean, a phenomenon known as "motional averaging."

The term $-\boldsymbol{I}_i.\underline{\underline{S}}.\boldsymbol{B}_0$ is the chemical shift that occurs for nuclei in molecular atoms, or the knight shift for nuclei in metals. It is typically a few ppm to several 100 ppm (i.e., 100's Hz to 10 kHz), depending on the nucleus. $\underline{\underline{S}} = \gamma\underline{\underline{\sigma}}$ is a tensor whose principal axes (1, 2, 3) are associated with the local symmetry axis of the molecular orbital (bond) in the vicinity of the nucleus. For a liquid state molecule tumbling rapidly and isotropically, only the averaged trace of $\underline{\underline{\sigma}}, \sigma_i = (1/3)(\sigma_{11} + \sigma_{22} + \sigma_{33})$ survives under motional averaging, giving a fixed frequency shift $-\sigma_i \gamma B_0 I_{iz}$. However, in a solid-state environment, the remaining terms also contribute to the anisotropic chemical shift

$$H_{CS} = -\sigma_i \gamma B_0 I_{iz} - \tfrac{1}{2}(3\cos^2\beta - 1) \times (\sigma_{33} - \sigma_i)\gamma B_0 I_{iz} \qquad [12]$$

where $\beta$ is the polar angle between the magnetic field and the principal axis (the axis "3").

The scalar coupling term, $\sum_j J\boldsymbol{I}_i.\boldsymbol{I}_j$ causes each ($i$th spin) energy level to be sensitive to the quantum states of the neighboring $j$-spins, the coupling constant $J$ being typically tens to hundreds of hertz for nearby spins, but reducing rapidly with greater distance in the molecular orbital. Note that the operator $\sum_j J\boldsymbol{I}_i.\boldsymbol{I}_j$ is nondiagonal in the zeroth-order representation, but provided that the chemical shift between the $I$ and $j$ spins is larger than the coupling frequency (known in chemistry as an AX spin system), the operator reduces to $\sum_j JI_{iz}I_{jz}$ the effect being to split the $i$-spin resonance in to a multiplet, depending on the state of the nearby $j$-spin. For $m$ identical nearby $j$-spins, the multiplet bears a simple

binomial relationship to $m$, allowing one to "read" this number directly. The combination of chemical shift and scalar coupling information is of profound importance in identifying molecular structure in chemistry.

The terms $\sum_j \boldsymbol{I}_i.\underline{\underline{D}}.\boldsymbol{I}_j$ and $\boldsymbol{I}_i.\underline{\underline{Q}}.\boldsymbol{I}_i$ are, respectively, the through-space dipolar interaction, $H_\mathrm{D}$, and the nuclear quadrupole interaction, $H_\mathrm{Q}$, the latter being nonzero only for nuclear spin quantum numbers $I \geq 1/2$, for example, $^2$H. These interactions, projected into the zeroth-order Zeeman frame, for the dipole–dipole interaction, are

$$H_\mathrm{D} = \frac{\mu_0\hbar}{4\pi}\sum_{j>i}\frac{\gamma_i\gamma_j}{r_{ij}^3}\frac{1}{2}\left(1-3\cos^2\theta_{ij}\right) \times \left(3I_{iz}I_{jz} - \boldsymbol{I}_i.\boldsymbol{I}_j\right) \qquad [13]$$

where $r_{ij}$ is the internuclear distance and $\theta_{ij}$ is the angle made by the internuclear vector with the magnetic field direction; while, for the quadrupole interaction

$$H_\mathrm{Q} = \frac{3eV_{ZZ}Q}{4I(2I-1)\hbar}\frac{1}{2}(1-3\cos\theta_{ZZ}) \times \left(3I_z^2 - I(I+1)\right) \qquad [14]$$

where $Q$ is the nuclear quadrupole moment, $V_{ZZ}$ is the electric field gradient (assuming axial symmetry) and $\theta_{ZZ}$ is the angle made by the principal axis of that gradient with the magnetic field direction. For protons in organic matter, the internuclear dipole interaction strength is on the order of 100 kHz, a similar strength being found for the quadrupole interaction of deuterons. However, in the liquid state, these orientation-dependent interactions fluctuate so rapidly that they are typically motionally averaged to zero. Nonetheless, their fluctuations do contribute to the relaxation process.

Liquid-state NMR can result in exceptionally high-resolution (sub-Hz) spectra, if care is taken to adjust the magnetic field harmonics (shims) to produce a highly uniform Zeeman field across the sample. The last contribution of residual inhomogeneities to line broadening can often be removed by gently spinning the sample about its axis at a rate of a few tens of hertz.

## The Evolution Domain, Multiple RF Pulses, and Multidimensional NMR

Having seen the complexity of the spin Hamiltonian, one may envisage experiments where the spin coherences evolve in a much more complicated manner. To this end, consider the case of a molecular liquid two-spin (AX) system coupled via the scalar spin–spin interaction. In first-order perturbation theory, we may represent the simple two-spin Hamiltonian (in the rotating frame of the averaged Larmor frequency) as

$$\begin{aligned} H_\mathrm{rot} &= -\sigma_1\gamma B_0 I_{1z} - \sigma_2\gamma B_0 I_{2z} + J I_{1z}I_{2z} \\ &= -\omega_1 I_{1z} - \omega_2 I_{2z} + J I_{1z}I_{2z} \end{aligned} \qquad [15]$$

We now write down the density matrix in the rotating frame following a single $90^\circ_x$ RF pulse ($I_x$),

$$\begin{aligned} \rho(t) &= \exp(\mathrm{i}\omega_1 t I_{1z} + \mathrm{i}\omega_2 t I_{2z} + \mathrm{i}J I_{1z}I_{2z}t) \\ &\quad \times \exp\left(\mathrm{i}\frac{\pi}{2}I_x\right) a_\mathrm{eqbm}(I_{1z}+I_{2z}) \exp\left(-\mathrm{i}\frac{\pi}{2}I_x\right) \\ &\quad \times \exp(-\mathrm{i}\omega_1 t I_{1z} - \mathrm{i}\omega_2 t I_{2z} - \mathrm{i}J I_{1z}I_{2z}t) \\ &= \exp(\mathrm{i}\omega_1 t I_{1z} + \mathrm{i}\omega_2 t I_{2z} + \mathrm{i}J I_{1z}I_{2z}t) a_\mathrm{eqbm}(I_{1y}+I_{2y}) \\ &\quad \times \exp(-\mathrm{i}\omega_1 t I_{1z} - \mathrm{i}\omega_2 t I_{2z} - \mathrm{i}J I_{1z}I_{2z}t) \\ &= \exp(\mathrm{i}\omega_1 t I_{1z} + \mathrm{i}\omega_2 t I_{2z}) a_\mathrm{eqbm} \\ &\quad \times \left(\left(I_{1y}+I_{2y}\right)\cos\left(\tfrac{1}{2}Jt\right) + 2(I_{1z}I_{2x} + I_{1x}I_{2z}) \right. \\ &\quad \left. \times \sin\left(\tfrac{1}{2}Jt\right)\right) \exp(-\mathrm{i}\omega_1 t I_{1z} - \mathrm{i}\omega_2 t I_{2z}) \\ &= a_\mathrm{eqbm}\left(\begin{aligned} &(I_{1y}\cos\omega_1 t + I_{2y}\cos\omega_2 t \\ &\quad + I_{1x}\sin\omega_1 t + I_{2x}\sin\omega_2 t)\cos\left(\tfrac{1}{2}Jt\right) \\ &\quad + 2\left(I_{1z}I_{2x}\cos\omega_2 t - I_{1z}I_{2y}\sin\omega_2 t\right. \\ &\quad \left. + I_{1x}I_{2z}\cos\omega_1 t - I_{1y}I_{2z}\sin\omega_1 t\right) \\ &\quad \times \sin\left(\tfrac{1}{2}Jt\right) \end{aligned}\right) \end{aligned} \qquad [16]$$

Detection in the rotating frame with $I_x + \mathrm{i}I_y$ gives a signal

$$S(t) \sim a_\mathrm{eqbm}(\exp(\mathrm{i}\omega_1 t) + \exp(\mathrm{i}\omega_2 t))\cos\left(\tfrac{1}{2}Jt\right) \qquad [17]$$

Fourier transformation with respect to $t$ yields a spectrum corresponding to two spectral lines at $\omega_1$ and $\omega_2$, each split into a doublet of two sidebands separated by $J$.

Notice that it is easier to follow the evolution of the density matrix by simply writing down a time sequence of behaviors under the influence of the successive Hamiltonians. Where simultaneous terms in the Hamiltonians commute, the order of their operation may be set at will. Thus, the above example becomes

$$\begin{aligned} I_{1z} + I_{2z} &\xrightarrow{\frac{\pi}{2}I_x} I_{1y} + I_{2y} \xrightarrow{JI_{1z}I_{2z}t} \left(I_{1y}+I_{2y}\right)\cos\left(\tfrac{1}{2}Jt\right) \\ &\quad + 2(I_{1z}I_{2x} + I_{1x}I_{2z})\sin\left(\tfrac{1}{2}Jt\right) \\ &\xrightarrow{\omega_1 t I_{1z} + \omega_2 t I_{2z}} (I_{1y}\cos\omega_1 t + I_{2y}\cos\omega_2 t \\ &\quad + \mathrm{i}I_{1x}\sin\omega_1 t + \mathrm{i}I_{2x}\sin\omega_2 t)\cos\left(\tfrac{1}{2}Jt\right) \\ &\quad + 2\left(I_{1z}I_{2x}\cos\omega_2 t - \mathrm{i}I_{1z}I_{2y}\sin\omega_2 t\right. \\ &\quad \left. + I_{1x}I_{2z}\cos\omega_1 t - \mathrm{i}I_{1y}I_{2z}\sin\omega_1 t\right)\sin\left(\tfrac{1}{2}Jt\right) \end{aligned} \qquad [18]$$

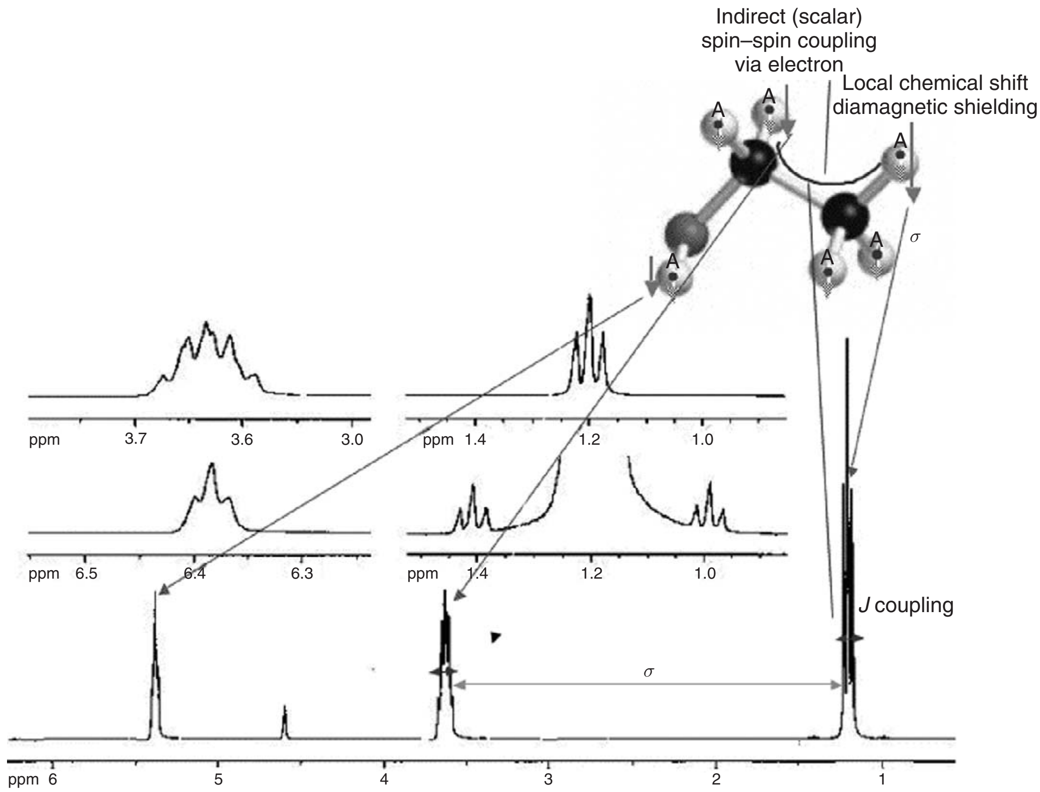



**Figure 3** The proton NMR spectrum of ethanol showing three major peaks, separated by chemical shift, each split into multiplets arising from nearby protons via the scalar coupling.

Now consider a two RF pulse scheme as shown in Figure 4, each RF pulse being $90°_x$. The evolution is

$$\begin{aligned}
&I_{1z}+I_{2z}\xrightarrow{\frac{\pi}{2}I_x} I_{1y}+I_{2y}\xrightarrow{\omega_1 t_1 I_{1z}+\omega_2 t_1 I_{2z}-JI_{1z}I_{2z}t_1}\\
&\quad\big(I_{1y}\cos\omega_1 t_1+I_{2y}\cos\omega_2 t_1+I_{1x}\sin\omega_1 t_1\\
&\qquad +I_{2x}\sin\omega_2 t_1\big)\cos\left(\tfrac{1}{2}Jt_1\right)+2\big(I_{1z}I_{2x}\cos\omega_2 t_1\\
&\qquad -I_{1z}I_{2y}\sin\omega_2 t_1+I_{1x}I_{2z}\cos\omega_1 t_1\\
&\qquad -I_{1y}I_{2z}\sin\omega_1 t_1\big)\sin\left(\tfrac{1}{2}Jt_1\right)\\
&\quad\xrightarrow{\frac{\pi}{2}I_x}\big(-I_{1z}\cos\omega_1 t_1-I_{2z}\cos\omega_2 t_1+I_{1x}\sin\omega_1 t_1\\
&\qquad +I_{2x}\sin\omega_2 t_1\big)\cos\left(\tfrac{1}{2}Jt_1\right)\\
&\qquad +2\big(I_{1y}I_{2x}\cos\omega_2 t_1+I_{1y}I_{2z}\sin\omega_2 t_1\\
&\qquad\qquad +I_{1x}I_{2y}\cos\omega_1 t_1+I_{1z}I_{2y}\sin\omega_1 t_1\big)\sin\left(\tfrac{1}{2}Jt_1\right)\\
&\quad\xrightarrow[\text{Keeping only observable magnetization}]{\omega_1 t_2 I_{1z}+\omega_2 t_2 I_{2z}+JI_{1z}I_{2z}t_2}\\
&\qquad\big(I_{1x}\sin\omega_1 t_1\cos\omega_1 t_2+I_{2x}\sin\omega_2 t_1\cos\omega_2 t_2\big)\\
&\qquad\times\cos\left(\tfrac{1}{2}Jt_1\right)\cos\left(\tfrac{1}{2}Jt_2\right)+\big(I_{1x}\sin\omega_2 t_1\sin\omega_1 t_2\\
&\qquad +I_{2x}\sin\omega_1 t_1\sin\omega_2 t_2\big)\times\sin\left(\tfrac{1}{2}Jt_1\right)\sin\left(\tfrac{1}{2}Jt_2\right)
\end{aligned}\quad [19]$$

If the idealized experiment is performed with two independent time dimensions $t_1$ and $t_2$, then detection in the rotating frame over the $t_2$ period with $I_x + iI_y$ gives a signal (restricting our attention to the quadrant of positive frequencies)

$$\begin{aligned}
S(t_1,t_2)\sim\, &a_{\mathrm{eqbm}}\big(\exp(i\omega_1 t_1)\exp(i\omega_1 t_2)+\exp(i\omega_2 t_1)\\
&\times\exp(i\omega_2 t_2)\big)\cos\left(\tfrac{1}{2}Jt_1\right)\cos\left(\tfrac{1}{2}Jt_2\right)\\
&+a_{\mathrm{eqbm}}\big(\exp(i\omega_2 t_1)\exp(i\omega_1 t_2)\\
&+\exp(i\omega_1 t_1)\exp(i\omega_2 t_2)\big)\\
&\times\sin\left(\tfrac{1}{2}Jt_1\right)\sin\left(\tfrac{1}{2}Jt_2\right)
\end{aligned}\quad [20]$$

When Fourier transformed in two dimensions with respect to $t_1$ and $t_2$, the pattern shown in Figure 5 results. Remarkably, while the diagonal spectrum is the same pair of doublets seen in the figure, this two-dimensional spectrum contains off-diagonal antiphase peaks for scalar-coupled sites where magnetization transfer has occurred.

The idea of performing NMR in two or more dimensions was first proposed by Jean Jeener in 1971. The example outlined above, correlation spectroscopy (COSY), is just one of an array of coherence transfer experiments using multiple RF pulse trains and time domain evolution of the spin ensemble. Notice that in the COSY experiment, $t_1$ is an evolution dimension during which no detection of NMR signal occurs, while $t_2$ is the detection domain.

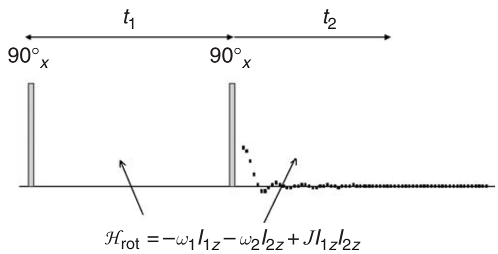



**Figure 4** RF pulse scheme used for COSY experiment.

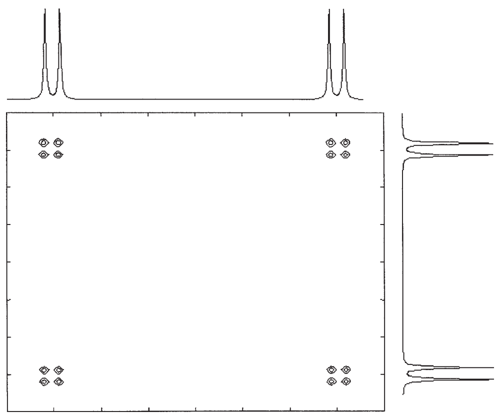


**Figure 5** Schematic COSY (modulus) spectrum for an AX spin system. Not that the (antiphase) off-diagonal peaks indicate $J$-couplings between chemical-shift-separated spins.

The effect of the evolution is indelibly imprinted in the spin system density matrix allowing later recall of vital information concerning the interactions present in the spin Hamiltonian. The COSY experiment allows one to determine which spins are coupled via their molecular orbital electrons. Other multidimensional methods that rely on dipole–dipole relaxation effects, such as NOESY, determine which spin sites have "through-space" proximity.

The use of two- and higher-dimensional methods has allowed the NMR spectra of biological macromolecules to be unraveled, with COSY methods used for spectral assignment of amino acid units, and NOESY methods used to determine any close proximities of amino acids otherwise well separated in the sequence. Such distance information has allowed the reconstruction of protein conformations by NMR.

The second RF pulse of Figure 4 also generates a state of the density matrix, $I_{1y}I_{2x}$ known as a double quantum coherence, and, in the simple COSY experiment, lost to observable magnetization. Other RF pulse schemes can take advantage of this state, converting it via suitable "coherence pathways" into an observable. For a detailed summary of these various NMR phenomena, readers are referred to the book by Ernst *et al.* (1987).

## Solid-State NMR

As with $J$ couplings, dipolar interactions and quadrupole interactions ($I > 1/2$) are bilinear in the spin operators and can be used to generate various higher-order coherence pathways in NMR experiments. Unlike the simple spin–spin coupling, they have an angular dependence. In solids, these interactions may broaden the NMR resonance line by tens to hundreds of kilohertz. In the case where a probe nucleus is located at a known site in the material (often achieved by deuteron labeling), these Hamiltonian terms may contribute important information about structure, and especially orientational anisotropy. For example, the quadrupole interaction for the spin-1 deuteron (see eqns [11] and [14]) depends as $P_2(\cos\theta_{ZZ}) = (1/2)(3\cos\theta_{ZZ} - 1)$ on the angle between the external magnetic field and the electric field gradient (generally associated with the local molecular orbital or bond direction, and taken here to be axially symmetric). Note that the first-order contribution of the quadrupole interaction leads to an unequal separation of the $m = 1, 0, -1$ Zeeman energy levels, resulting in a doublet NMR spectrum, for any particular orientation, $\theta_{ZZ}$. Such a unique orientation might be found in a single crystal, or in a nematic liquid crystalline state. For a polycrystalline material, however, the NMR spectrum has a contribution from all orientations, leading to a characteristic powder pattern. The details of $^2H$ spectral distributions may be used to characterize the degree of orientational order in solids and soft, anisotropic matter.

For $^1H$, $^{13}C$, and other spin-1/2 nuclei, dipolar interactions (with a wide distribution of spin spacings and internuclear vector orientation) may severely broaden the NMR spectrum in the solid state (see eqns [11] and [13]). Such interactions, along with quadrupole interactions for nuclei with $I > 1/2$, may be significantly reduced by modulating the effective dipolar Hamiltonian at a rate faster than its strength in frequency units. Two methods are available, one (magic angle spinning or MAS) relying on the angular terms in eqns [13] and [14], and the other (multiple pulse line narrowing) on the spin terms. The MAS technique relies on spinning the sample rapidly about at angle oriented at 54.4° to the magnetic field, such that the average value of $P_2(\cos\theta_{ij})$ becomes its projection along this spinning axis, while the projection of the spinning axis residual is $P_2(\cos 54.4°) \approx 0$. Multiple pulse methods rely on a successive reorientation of the spin system such that the effective dipolar Hamiltonian that results from the application of the nested evolution operators is rendered close to zero.

In practice, MAS techniques work best with $^{13}C$ NMR where the moderate $^{1}H$–$^{13}C$ dipolar interactions may be removed with achievable spinning speeds (a few tens of kilohertz). Furthermore, the larger proton magnetization ($\gamma_{^1H}/\gamma_{^{13}C} \approx 4$) can be transferred to the $^{13}C$ nuclei via Hartman–Hahn cross-polarization thus significantly enhancing sensitivity. Such methodology is referred to as CPMAS NMR.

The real art of solid-state NMR is in removing the unwanted dipolar or quadrupolar interactions, but leaving specific interactions of interest. This may be achieved by including in the MAS experiment, specific combinations of pulses which recouple selected spins. Some of the most sophisticated experiments in modern NMR are to be found in this domain of application.

## Conclusion

NMR provides exceptional structural information concerning molecules, biomolecules as well as molecular assemblies, liquid crystals, soft solids, and solids. In addition, the method provides unique information concerning molecular dynamics, through both relaxation methods and the direct measurement of diffusion or flow. One spectacular application of NMR concerns its use in imaging, achieved by giving the Larmor frequency a spatial tag through the use of deliberately inhomogeneous magnetic fields. This topic is covered in the article on Magnetic Resonance Imaging.

*See also:* Magnetic Resonance Imaging.

## Further Reading


Abragam A (1961) *Principles of Nuclear Magnetism*. Oxford: Clarendon.

Bloch F, Hansen WW, and Packard M (1946) Nuclear induction. *Physical Review* 70: 474.

Ernst RR, Bodenhausen G, and Wokaun A (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Oxford: Clarendon.

Goldman M (1991) *Quantum Description of High-Resolution NMR in Liquids*. New York: Oxford University Press.

Hahn EL (1950) Spin echoes. *Physical Review* 80: 580.

Jeener J (1971) *Ampere International Summer School*. Yugoslavia: Basko Polje.

McNeil EB, Slichter CP, and Gutowsky HS (1951) Slow beats in 19F nuclear spin echoes. *Physical Review* 84: 1245.

Mehring M (1983) *Principles of High Resolution NMR in Solids*. Berlin: Springer.

Purcell EM, Torrey HC, and Pound RV (1946) Resonance absorption by nuclear magnetic moments in a solid. In: *Physical Review* 69: 37.

Rabi II and Cohen VW (1933) Measurement of nuclear spin by the method of molecular beams: the nuclear spin of sodium. *Physical Review* 46: 707.

Schmidt-Rohr K and Spiess H-W (1994) *Multi-Dimensional Solid State NMR and Polymers*. London: Academic Press.

Slichter CP (1963) *Principles of Magnetic Resonance*. New York: Harper and Row.


# Number Theory in Physics

**M Marcolli**, Max-Planck-Institut für Mathematik, Bonn, Germany



Several fields of mathematics have closely been associated to physics: this has always been the case for the theory of differential equations. In the early twentieth century, with the advent of general relativity and quantum mechanics, topics such as differential and Riemannian geometry, operator algebras, and functional analysis, or group theory also developed a close relation to physics. In the 1990s, mostly through the influence of string theory, algebraic geometry also began to play a major role in this interaction. Recent years have seen an increasing number of results suggesting that number theory also is beginning to play an essential part on the scene of contemporary theoretical and mathematical physics. Conversely, ideas from physics, mostly from quantum field theory and string theory, have started to influence work in number theory.

In describing significant occurrences of number theory in physics, we will, on the one hand, restrict our attention to quantum physics, while, on the other, we will assume a somewhat extensive definition of number theory that will allow us to include arithmetic algebraic geometry. The territory is vast and an extensive treatment would go beyond the size limits imposed by the encyclopedia. The choice of topics represented here inevitably reflects the limited knowledge, particular interests, and bias of the author. Very useful references, collecting a lot of material on number theory and physics, are the proceedings of the Les Houches conference in 2003 (Beilinson and Manin 1986), as well as the two volumes of a previous Les Houches conference on number theory and physics, which took place in 1989, published by Springer in 1990 and 1992. A number theory and physics database is presently maintained online by M R Watkins.

In the following, we have organized the material by topics in number theory that have so far made an appearance in physics, and for each we briefly describe the relevant context and results. This singles out many themes. We first discuss a class of functions that occur in physics and their special values that are of great number-theoretic importance. This includes the dilogarithm, the polylogarithms and multiple polylogarithms, and the multiple zeta values. We also discuss the most important symmetry groups of number theory, the Galois groups, and occurrences in physics of some forms of Galois theory. We then discuss how techniques from the arithmetic geometry of algebraic varieties, especially Arakelov geometry, play a role in string theory. Finally, we discuss briefly the theory of motives and outline its possible relation to quantum physics. From the physics point of view, it seems that the most promising directions in which number-theoretic tools have come to play a crucial role are to be found mostly in the realm of rational conformal field theories and of noncommutative geometry, as well as in certain aspects of string theory.

Among the topics that are very relevant to this theme, but that will not be touched upon in this article, there are important subjects such as the theory of "arithmetic quantum chaos," the use of methods of random matrix theory applied to the study of zeros of zeta functions, or mirror symmetry and its connection to modular forms. The interested reader can find such topics treated in other articles of this encyclopedia and in the references mentioned above (*see* Quantum Ergodicity and Mixing of Eigenfunctions; Random Matrix Theory in Physics; Mirror Symmetry: a Geometric Survey).

## Dilogarithm, Multiple Polylogarithms, Multiple Zeta Values

The dilogarithm is defined as

$$Li_2(z) = \int_z^0 \frac{\log(1-t)}{t}\,dt = \sum_{n=1}^{\infty} \frac{z^n}{n^2}$$

It satisfies the functional equation $Li_2(z) + Li_2(1-z) = Li_2(1) - \log(z)\log(1-z)$, where $Li_2(1) = \zeta(2)$, for $\zeta(s)$ the Riemann zeta function. A variant is given by the Rogers dilogarithm $L(x) = Li_2(x) + (1/2)\log(x)\log(1-x)$. For more details, see Zagier's paper (Julia *et al.* 2005, vol. II).

The polylogarithms are similarly defined by the series $Li_k(z) = \sum_{n\geq 1} z^n/n^k$. In quantum electrodynamics, there are corrections to the value of the gyromagnetic ratio, in powers of the fine structure constant. The correction terms that are known exactly involve special values of the zeta function such as $\zeta(3), \zeta(5)$ and values of polylogarithms such as $Li_4(1/2)$. The series defining the polylogarithm function $Li_s(z) = \sum_{n\geq 1} z^n/n^s$ converges absolutely for all $s \in \mathbb{C}$ and $|z| < 1$ and has analytic continuation to $z \in \mathbb{C} \setminus [1,\infty)$. The Fermi–Dirac and Bose–Einstein distributions are expressed in terms of the polylogarithm function as

$$\int_0^\infty \frac{x^s}{e^{x-\mu} \pm 1}\,dx = -\Gamma(s+1) Li_{1+s}(\pm e^\mu)$$

The multiple polylogarithms are functions defined by the expressions

$$Li_{s_1,\ldots,s_r}(z_1, z_2, \ldots, z_r) = \sum_{n_1>n_2>\cdots>n_r>0} \frac{z_1^{n_1} z_2^{n_2} \cdots z_r^{n_r}}{n_1^{s_1} n_2^{s_2} \cdots n_r^{s_r}} \qquad [1]$$

By analytic continuation, the functions $Li_{s_1,\ldots,s_r}(z_1, z_2, \ldots, z_r)$ are defined for all complex $s_i$ and for $z_i$ in the complement of the cut $[1,\infty)$ in the complex plane. Multiple zeta values of weight $k$ and depth $r$ are given by the expressions

$$\zeta(k_1, \ldots, k_r) = \sum_{n_1>n_2>\cdots>n_r>0} \frac{1}{n_1^{k_1} \cdots n_r^{k_r}} \qquad [2]$$

with $k_i \in \mathbb{N}$ and $k_1 \geq 2$. These satisfy many combinatorial identities and nontrivial relations over $\mathbb{Q}$. For an informative overview on the subject, see Cartier (2002). Notice that, for the sums in [1] and [2], a different summation convention can also be found in the literature.

### Conformal Field Theories and the Dilogarithm

There is a relation between the torsion elements in the algebraic $K$-theory group $K_3(\mathbb{C})$ and rational conformally invariant quantum field theories in two dimensions (see Nahm (2005)). There is, in fact, a map, given by the dilogarithm, from torsion elements in the Bloch group (closely related to the algebraic $K$-theory) to the central charges and scaling dimensions of the conformal field theories.

This correspondence arises by considering sums of the form

$$\sum_{m\in\mathbb{N}^r} \frac{q^{Q(m)}}{(q)_m} \qquad [3]$$

where $(q)_m = (q)_{m_1} \cdots (q)_{m_r}, (q)_{m_i} = (1-q)(1-q^2)\cdots(1-q^{m_i})$ and $Q(m) = m^t Am/2 + bm + h$ has rational coefficients. Such sums are naturally obtained from considerations involving the partition function of a bosonic rational conformal field theory (CFT). In

particular, [3] can define a modular function only if all the solutions of the equation

$$\sum_j A_{ij} \log(x_j) = \log(1 - x_i) \qquad [4]$$

determine elements of finite order in an extension $\hat{B}(\mathbb{C})$ of the Bloch group, which accounts for the fact that the logarithm is multivalued. The Rogers dilogarithm gives a natural group homomorphism $(2\pi i)^2 L : \hat{B}(\mathbb{C}) \to \mathbb{C}/\mathbb{Z}$, which takes values in $\mathbb{Q}/\mathbb{Z}$ on the torsion elements. These values give the conformal dimensions of the fields in the theory.

### Feynman Graphs

Multiple zeta values appear in perturbative quantum field theory. D Kreimer (2000) developed a connection between knot theory and a class of transcendental numbers, such as multiple zeta values, obtained by quantum field-theoretic calculations as counterterms generated by corresponding Feynman graphs. Broadhurst and Kreimer (1997) identified Feynman diagrams with up to nine loops whose corresponding counterterms give multiple zeta values up to weight 15. Recently, Kreimer showed some deep analogies between residues of quantum fields and variations of mixed Hodge–Tate structures associated to polylogarithms.

Testing predictions about the standard model of elementary particles, in the hope of detecting new physics, requires developing effective computational methods handling the huge number of terms involved in any such calculation, that is, efficient algorithms for the expansion of higher transcendental functions to a very high order. The interesting fact is that abstract number-theoretic objects, such as multiple zeta values and multiple polylogarithms, appear naturally in this context (cf., e.g., Moch *et al.* (2002)). The explicit recursive algorithms are based on Hopf algebras and produce expansions of nested finite or infinite sums involving ratios of gamma functions and $Z$-sums (Euler–Zagier sums), which naturally generalize multiple polylogarithms and multiple zeta values. Such sums typically arise in the calculation of multiscale multiloop integrals. The algorithms are designed to recursively reduce the $Z$-sums involved to simpler ones with lower weight or depth.

## Galois Theory

Given a number field $\mathbb{K}$, which is an algebraic extension of $\mathbb{Q}$ of some degree $[\mathbb{K}:\mathbb{Q}] = n$, there is an associated fundamental symmetry group, given by the absolute Galois group $\mathrm{Gal}(\bar{\mathbb{K}}/\mathbb{K})$, where $\bar{\mathbb{K}}$ is an algebraic closure of $\mathbb{K}$. Even in the case of $\mathbb{Q}$, the absolute Galois group $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ is a very complicated object, far from being fully understood.

One can consider an easier symmetry group, which is the abelianization of the absolute Galois group. This corresponds to considering the field $\mathbb{K}^{ab}$, the "maximal abelian extension" of $\mathbb{K}$, which has the property that

$$\mathrm{Gal}(\mathbb{K}^{ab}/\mathbb{K}) = \mathrm{Gal}(\bar{\mathbb{K}}/\mathbb{K})^{ab}$$

The Kronecker–Weber theorem shows that for $\mathbb{K} = \mathbb{Q}$ the maximal abelian extension can be identified with the cyclotomic field (generated by all roots of unity), $\mathbb{Q}^{ab} = \mathbb{Q}^{\mathrm{cycl}}$, and the Galois group is identified with $\mathrm{Gal}(\mathbb{Q}^{ab}/\mathbb{Q}) \cong \hat{\mathbb{Z}}^*$, where $\hat{\mathbb{Z}}^* = \mathbb{A}_f^*/\mathbb{Q}_+^*$. In general, for other number fields, one has the "class field theory isomorphism"

$$\theta : \mathrm{Gal}(\mathbb{K}^{ab}/\mathbb{K}) \xrightarrow{\simeq} C_{\mathbb{K}}/D_{\mathbb{K}}$$

where $C_{\mathbb{K}} = \mathbb{A}_{\mathbb{K}}^*/\mathbb{K}^*$ is the group of idele classes and $D_{\mathbb{K}}$ the connected component of the identity in $C_{\mathbb{K}}$. In general, however, one does not have an explicit description of the generators of the maximal abelian extension $\mathbb{K}^{ab}$ and the action of the Galois group. This is the content of the explicit class field theory problem, Hilbert's 12th problem. In addition to the Kronecker–Weber case, a complete answer is known in the case of imaginary quadratic fields $\mathbb{K} = \mathbb{Q}(\sqrt{-d})$, with $d > 1$ a positive integer. In this case generators are obtained by evaluating modular functions at a point $\tau$ in the upper-half plane such that $\mathbb{K} = \mathbb{Q}(\tau)$ and the Galois action is described explicitly through the group of automorphisms of the modular field, through Shimura reciprocity. For a survey of the explicit class field theory problem and the case of imaginary quadratic fields, see Stevenhagen (2001).

As we mentioned above, understanding the structure of the absolute Galois group $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ is a fundamental question in number theory. Grothendieck described, in his famous proposal "Esquisse d'un programme," how to obtain an action of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ on an essentially combinatorial object, the set of "dessins d'enfants." These are connected graphs (on a surface) such that the complement of the graph is a union of open cells and the vertices have two different markings, with the properties that adjacent vertices have opposite markings. Such objects arise by considering the projective line $\mathbb{P}^1$ minus three points. Any finite cover of $\mathbb{P}^1$ branched only over $\{0, 1, \infty\}$ gives an algebraic curve defined over $\bar{\mathbb{Q}}$. The dessin is the inverse image under the covering map of the segment $[0, 1]$ in $\mathbb{P}^1$. The absolute Galois group $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ acts on the data of the curve and the covering map, hence on the set of

dessins. A theorem of Bielyi shows that, in fact, all algebraic curves defined over $\bar{\mathbb{Q}}$ are obtained as coverings of the projective line ramified only over the points $\{0, 1, \infty\}$. This has the effect of realizing the absolute Galois group as a subgroup of outer automorphisms of the profinite fundamental group of the projective line minus three points. For a general reference on the subject, see Schneps (1994).

A different type of Galois symmetry of great arithmetic significance is "motivic" Galois theory. This will be discussed later in the section dedicated to motives, where we discuss a surprising occurrence in the context of perturbative quantum field theory and renormalization.

## Quantum Statistical Mechanics and Class Field Theory

In quantum statistical mechanics, one considers an algebra of observables, which is a unital $C^*$-algebra $\mathcal{A}$ with a time evolution $\sigma_t$. States are given by linear functionals $\varphi : \mathcal{A} \to \mathbb{C}$ satisfying $\varphi(1) = 1$ and positivity $\varphi(x^* x) \geq 0$. Equilibrium states $\varphi$ at inverse temperature $\beta$ satisfy the Kubo–Martin–Schwinger (KMS) condition, namely, for all $x, y \in \mathcal{A}$ there exists a bounded holomorphic function $F_{x,y}(z)$ on the strip $0 < \Im(z) < \beta$, which extends continuously to the boundary, such that for all $t \in \mathbb{R}$

$$F_{x,y}(t) = \varphi(x\sigma_t(y))$$

and

$$F_{x,y}(t + \mathrm{i}\beta) = \varphi(\sigma_t(y)x) \qquad [5]$$

Cases of number-theoretic interest arise when one considers the noncommutative space of commensurability classes of $\mathbb{Q}$-lattices up to scaling as algebra of observables, with a natural time evolution determined by the covolume, as shown in the paper *Quantum Statistical Mechanics of* $\mathbb{Q}$*-Lattices* of Connes–Marcolli (Julia *et al.* 2005, vol. I). A $\mathbb{Q}$-lattice in $\mathbb{R}^n$ consists of a pair $(\Lambda, \phi)$ of a lattice $\Lambda \subset \mathbb{R}^n$ together with a homomorphism of abelian groups $\phi : \mathbb{Q}^n/\mathbb{Z}^n \longrightarrow \mathbb{Q}\Lambda/\Lambda$. Two $\mathbb{Q}$-lattices are commensurable, $(\Lambda_1, \phi_1) \sim (\Lambda_2, \phi_2)$, iff $\mathbb{Q}\Lambda_1 = \mathbb{Q}\Lambda_2$ and $\phi_1 = \phi_2 \bmod \Lambda_1 + \Lambda_2$.

**The Bost–Connes system** The quantum statistical mechanical system considered by Bost and Connes (1995) corresponds to the case of one-dimensional $\mathbb{Q}$-lattices. The partition function of the system is the Riemann zeta function $\zeta(\beta)$. The system has spontaneous symmetry breaking at $\beta = 1$, with a single KMS state for all $0 < \beta \leq 1$. For $\beta > 1$, the extremal equilibrium states are parametrized by the embeddings of $\mathbb{Q}^{\mathrm{cycl}}$ in $\mathbb{C}$ with a free transitive action of the idele class group $C_{\mathbb{Q}}/D_{\mathbb{Q}} = \hat{\mathbb{Z}}^*$. At zero temperature, the evaluation of $\mathrm{KMS}_\infty$ states on elements of a rational subalgebra intertwines the action of $\hat{\mathbb{Z}}^*$ by automorphisms of $(\mathcal{A}, \sigma_t)$ with the action of $\mathrm{Gal}(\mathbb{Q}^{ab}/\mathbb{Q})$ on the values of the states. This recovers the explicit class field theory of $\mathbb{Q}$ from a physical perspective.

**Noncommutative space of adele classes** The algebra $\mathcal{A}$ of the Bost–Connes system is the noncommutative algebra of functions $f(r, \rho)$, for $\rho \in \hat{\mathbb{Z}}$ and $r \in \mathbb{Q}^*$ such that $r\rho \in \hat{\mathbb{Z}}$, with the convolution product

$$f_1 * f_2(r, \rho) = \sum_{s \in \mathbb{Q}^* : s\rho \in \hat{\mathbb{Z}}} f_1(rs^{-1}, s\rho) f_2(s, \rho) \qquad [6]$$

and the adjoint $f^*(r, \rho) = \overline{f(r^{-1}, r\rho)}$. According to the general philosophy of Connes style noncommutative geometry, it is the algebra of coordinates of the noncommutative space defined by the "bad quotient" $\mathrm{GL}_1(\mathbb{Q}) \setminus (\mathbb{A}_f \times \{\pm 1\})$ – a noncommutative version of the zero-dimensional Shimura variety $\mathrm{Sh}(\mathrm{GL}_1, \{\pm 1\}) = \mathrm{GL}_1(\mathbb{Q}) \setminus (\mathrm{GL}_1(\mathbb{A}_f) \times \{\pm 1\})$. Its "dual system" (in the sense of Connes's duality of type III and type II factors) is obtained by taking the crossed product by the time evolution. It gives the algebra of coordinates of the noncommutative space defined by the quotient $\mathbb{A}/\mathbb{Q}^*$. This is the noncommutative space of "adele classes" used by Connes in his spectral realization of the zeros of the Riemann zeta function.

**The $\mathrm{GL}_2$-system** A generalization of the Bost–Connes system was introduced by Connes and Marcolli in the paper *Quantum Statistical Mechanics of* $\mathbb{Q}$*-Lattices* (Julia *et al.* 2005). This corresponds to the case of two-dimensional $\mathbb{Q}$-lattices. The partition function is the product $\zeta(\beta)\zeta(\beta - 1)$. The system in this case has two phase transitions, with no KMS states for $\beta \leq 1$. For $\beta > 2$, the extremal KMS states are parametrized by the invertible $\mathbb{Q}$-lattices, namely, those for which $\phi$ is an isomorphism. The algebra $\mathcal{A}$ has an arithmetic structure given by a rational algebra of unbounded multipliers. This rational algebra contains modular functions and Hecke operators. At zero temperature, extremal KMS states can be evaluated on these multipliers. Symmetries of $(\mathcal{A}, \sigma_t)$ are realized in part by endomorphisms (as in the theory of superselection sectors) and the symmetry group acting on low-temperature KMS states is the group of automorphisms of the modular field $\mathrm{GL}_2(\mathbb{A}_f)/\mathbb{Q}^*$. For a generic set of extremal $\mathrm{KMS}_\infty$ states, evaluation at the rational algebra intertwines this action with the action on the values of an embedding of the modular field as a subfield of $\mathbb{C}$.

**The complex multiplication system** In the case of an imaginary quadratic field $\mathbb{K} = \mathbb{Q}(\tau)$, an analogous construction is possible. A one-dimensional $\mathbb{K}$-lattice is a pair $(\Lambda, \phi)$ of a finitely generated $\mathcal{O}$-submodule $\Lambda$ of $\mathbb{C}$, with $\Lambda\mathbb{K} = \mathbb{K}$, and a homomorphism of $\mathcal{O}$-modules $\phi : \mathbb{K}/\mathcal{O} \to \mathbb{K}\Lambda/\Lambda$. Two $\mathbb{K}$-lattices are commensurable iff $\mathbb{K}\Lambda_1 = \mathbb{K}\Lambda_2$ and $\phi_1 = \phi_2 \bmod \Lambda_1 + \Lambda_2$. Connes *et al.* (Preprint 2005) constructed a quantum statistical mechanical system describing the noncommutative space of commensurability classes of one-dimensional $\mathbb{K}$-lattices up to scale. The partition function is the Dedekind zeta function $\zeta_{\mathbb{K}}(\beta)$. The system has a phase transition at $\beta = 1$ with a unique KMS state for higher temperatures and extremal KMS states parametrized by the invertible $\mathbb{K}$-lattices at lower temperatures. There is a rational subalgebra induced by the rational structure of the $GL_2$-system (one-dimensional $\mathbb{K}$-lattices are also two-dimensional $\mathbb{Q}$-lattices with compatible notions of commensurability). The symmetries of the system are given by the idele class group $\mathbb{A}^*_{\mathbb{K},f}/\mathbb{K}^*$. The action is partly realized by endomorphisms corresponding to the possible presence of a nontrivial class group (for class number $>1$). The values of extremal $KMS_\infty$ states on the rational subalgebra intertwine the action of the idele class group with the Galois action on the values. This fully recovers the explicit class field theory for imaginary quadratic fields.

#### Conformal Field Theory and the Absolute Galois Group

Moore and Seiberg considered data associated to any rational conformal field theory, consisting of matrices, obtained as monodromies of some holomorphic multi-valued functions on the relevant moduli spaces, satisfying polynomial equations. Under reasonable hypotheses, the coefficients of the Moore–Seiberg matrices are algebraic numbers. This allows for the presence of interesting arithmetic phenomena. Through the Chern–Simons/Wess–Zumino–Witten correspondence, it is possible to construct three-dimensional topological field theories from solutions to the Moore–Seiberg equations.

On the arithmetic side, Grothendieck proposed in his "Esquisse d'un programme" the existence of a Teichmüller tower given by the moduli spaces $M_{g,n}$ of Riemann surfaces of arbitrary genus $g$ and number of marked points $n$, with maps defined by operations such as cutting and pasting of surfaces and forgetting marked points, all encoded in a family of fundamental groupoids. He conjectured that the whole tower can be reconstructed from the first two levels, providing, respectively, generators and relations. He called this a "game of Lego–Teichmüller." He also conjectured that the absolute Galois group acts by outer automorphisms on the profinite completion of the tower. The basic building blocks of the tower are provided by "pairs of pants," that is, by projective lines minus three points.

This leads to a conjectural relation between the Moore–Seiberg equations and this Grothendieck–Teichmüller setting (cf. Degiovanni 1994) according to which solutions of the Moore–Seiberg equations provide projective representations of the Teichmüller tower, and the action of the absolute Galois group $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ corresponds to the action on the coefficients of the Moore–Seiberg matrices.

Rational conformal field theories are, in general, one of the most promising sources of interactions between number theory and physics, involving interesting Galois actions, modular forms, Brauer groups, and complex multiplication. Some fundamental work in this direction was done by, for example, Borcherds and Gannon.

## Arithmetic Algebraic Geometry

In this section we describe occurrences in physics of various aspects of the arithmetic geometry of algebraic varieties.

#### Arithmetic Calabi–Yau

In the context of type II string theory, compactified on Calabi–Yau 3-folds, Greg Moore considered certain black hole solutions and a resulting dynamical system given by a differential equation in the corresponding moduli. The fixed points of these equations determine certain "black hole attractor varieties." In the case of varieties obtained from a product of elliptic curves or of a K3 surface and an elliptic curve, the attractor equation singles out an arithmetic property: the elliptic curves have complex multiplication. The class number of the corresponding imaginary quadratic field counts U-duality classes of black holes with the same area. Other results point to a relation between the arithmetic properties of Calabi–Yau 3-folds and conformal field theory. For instance, it was shown by Schimmrigk that, in certain cases, the algebraic number field defined via the fusion rules of a conformal field theory as the field defined by the eigenvalues of the integer-valued fusion matrices

$$\phi_i * \phi_j = (N_i)_j^k \phi_k$$

can be recovered from the Hasse–Weil $L$-function of the Calabi–Yau. An interesting case is provided by the Gepner model associated with the Fermat quintic Calabi–Yau 3-fold.

### Arakelov Geometry

For $\mathbb{K}$ a number field and $O_{\mathbb{K}}$ its ring of integers, a smooth proper algebraic curve $X$ over $\mathbb{K}$ determines a smooth minimal model $X_{O_{\mathbb{K}}}$, which defines an arithmetic surface $\mathcal{X}_{O_{\mathbb{K}}}$ over Spec($O_{\mathbb{K}}$). The closed fiber $X_{\wp}$ of $\mathcal{X}_{O_{\mathbb{K}}}$ over a prime $\wp \in O_{\mathbb{K}}$ is given by the reduction mod $\wp$.

When Spec($O_{\mathbb{K}}$) is "compactified" by adding the Archimedean primes, one can correspondingly enlarge the group of divisors on the arithmetic surface by adding formal real linear combinations of irreducible "closed vertical fibers at infinity." Such fibers are only treated as formal objects. The main idea of Arakelov geometry is that it is sufficient to work with "infinitesimal neighborhood" $X_{\alpha}(\mathbb{C})$ of these fibers, given by the Riemann surfaces obtained from the equation defining $X$ over $\mathbb{K}$ under the embeddings $\alpha : \mathbb{K} \hookrightarrow \mathbb{C}$ that constitute the Archimedean primes. Arakelov developed a consistent intersection theory on arithmetic surfaces, by computing the contribution of the Archimedean primes to the intersection indices using Hermitian metrics on these Riemann surfaces and the Green function of the Laplacian.

A general introduction to the subject of Arakelov geometry can be found in Lang (1988). Manin (1991) showed that these Green functions can be computed in terms of geodesics in a hyperbolic 3-manifold that has the Riemann surface $X_{\alpha}(\mathbb{C})$ as its conformal boundary at infinity.

**The Polyakov measure** A first application to physics of methods of Arakelov geometry was an explicit formula obtained by Beilinson and Manin (1986) for the Polyakov bosonic string measure in terms of Faltings's height function at algebraic points of the moduli space of curves.

The partition function for the closed bosonic string has a perturbative expansion $Z = \sum_{g \geq 0} Z_g$, with

$$Z_g = e^{\beta(2-2g)} \int_{\Sigma} e^{-S(x,\gamma)} Dx D\gamma \qquad [7]$$

written in terms of a compact Riemann surface $\Sigma$ of genus $g$, maps $x : \Sigma \to \mathbb{R}^d$, and metrics $\gamma$ on $\Sigma$. The classical action is of the form

$$S(x, \gamma) = \int_{\Sigma} d^2 z \sqrt{|\gamma|} \gamma^{ab} \partial_a x^{\mu} \partial_b x^{\mu} \qquad [8]$$

Using the invariance of the classical action with respect to the semidirect product of diffeomorphisms of $\Sigma$ and the conformal group, the integral is reduced (in the critical dimension $d = 26$ where the conformal anomaly cancels) to a zeta regularized determinant of the Laplacian for the metric on $\Sigma$ and an integration over the moduli space $M_g$ of genus $g$ algebraic curves. Beilinson and Manin gave an explicit formula for the resulting Polyakov measure on $M_g$ using results of Faltings on Arakelov geometry of arithmetic surfaces. In particular, their argument uses essentially the properties of the Faltings metrics on the invertible sheaves $d(L)$ given by the "multiplicative Euler characteristics" of sheaves $L$ of relative 1-forms. For a suitable choice of bases $\{\phi_j\}$ and $\{w_j\}$ of differentials and quadratic differentials, the formula for the Polyakov measure is then of the form (up to a multiplicative constant)

$$d\pi_g = |\det B|^{-18} (\det \Im \tau)^{-13} W_1 \wedge \bar{W}_1 \wedge \cdots \wedge W_{3g-3} \wedge \bar{W}_{3g-3} \qquad [9]$$

with $\tau$ in the Siegel upper-half space, $B_{ij} = \int_{a_i} \phi_j$, and the $W_j$ given by the images of the basis $w_j$ under the Kodaira–Spencer isomorphism.

**Holography** In the case of the elliptic curve $X_q(\mathbb{C}) = \mathbb{C}^*/q^{\mathbb{Z}}$, a formula of Alvarez-Gaume, Moore, and Vafa gives the operator product expansion of the path integral for bosonic field theory as

$$g(z,1) = \log \left( |q|^{B_2(\log|z|/\log|q|)/2} |1-z| \times \prod_{n=1}^{\infty} |1-q^n z| \, |1-q^n z^{-1}| \right) \qquad [10]$$

where $B_2$ is the second Bernoulli polynomial. Expression [8] is in fact the Arakelov Green function on $X_q(\mathbb{C})$ (cf. Lang (1988)).

Using this and analogous results for higher genus Riemann surfaces, Manin and Marcolli (2001) showed that the result of Manin (1991) on Arakelov and hyperbolic geometry can be rephrased in terms of the AdS/CFT correspondence, or holography principle. Expression [8] can then be written as a combination of terms involving geodesic lengths in the Euclidean BTZ black hole.

In the case of higher genus curves, the Arakelov Green function on a compact Riemann surface, which is related to the two-point correlation function for bosonic field theory, can be expressed in terms of the semiclassical limit of gravity (the geodesic propagator) on the bulk space of Euclidean versions of asymptotically $AdS_{2+1}$ black holes introduced by K Krasnov.

## Motives

There are several cohomology theories for algebraic varieties: de Rham, Betti, étale cohomology. de Rham

and Betti are related by the period isomorphism, and comparison isomorphisms relate Étale and Betti cohomology. In the smooth projective case, they have the expected properties of Poincaré duality, Künneth isomorphisms, etc. Moreover, Étale cohomology provides interesting $\ell$-adic representations of $\mathrm{Gal}(\bar{k}/k)$. In order to understand what type of information, such as maps or operations can be transferred from one to another cohomology, Grothendieck introduced the idea of the existence of a "universal cohomology theory" with realization functors to all the known cohomology theories for algebraic varieties. He called this the theory of "motives." Properties that can be transferred between different cohomology theories are those that exist at the motivic level. A short introduction to motives can be found in Serre (1992).

The first constructions of a category of motives proposed by Grothendieck covers the case of smooth projective varieties. The corresponding motives form a $\mathbb{Q}$-linear abelian category of "pure motives." Roughly, objects are varieties and morphisms are "correspondences" given by algebraic cycles in the product, modulo a suitable equivalence relation. The category also contains Tate objects generated by $\mathbb{Q}(1)$, which is the inverse of the pure motive $H^2(\mathbb{P}^1)$. Grothendieck's standard conjectures imply that the category of pure motives is equivalent to the category of representations $\mathrm{Rep}_G$ of a "motivic Galois group," which in the case of pure motives is proreductive. The subcategory of pure Tate motives has as motivic Galois group the multiplicative group $\mathbb{G}_m$. The situation is more complicated for "mixed motives," for which constructions were only very recently proposed (e.g., in the work of Voevodsky). These provide a universal cohomology theory for more general classes of algebraic varieties. Mixed Tate motives are the subcategory generated by the Tate objects. There is again a motivic Galois group. For mixed motives it is an extension of a proreductive group by a prounipotent group, with the proreductive part coming from pure motives and the prounipotent part from the presence of a weight filtration on mixed motives. The multiple zeta values appear as periods of mixed Tate motives.

### Renormalization and Motivic Galois Theory

A manifestation of motivic Galois groups in physics arises in the context of the Connes–Kreimer theory of perturbative renormalization (for an introduction to this topic, *see* Hopf Algebra Structure of Renormalizable Quantum Field Theory). In fact, according to the Connes–Kreimer theory, the Bogoliubov–Parasiuk–Hepp–Zimmerman (BPHZ) renormalization scheme with dimensional regularization and minimal subtraction can be formulated mathematically in terms of the Birkhoff factorization

$$\gamma(z) = \gamma_-(z)^{-1}\gamma_+(z) \qquad [11]$$

of loops in a prounipotent Lie group $G$, which is the group of characters of the Hopf algebra of Feynman graphs. Here, the loop $\gamma$ is defined on a small punctured disk around the critical dimension $D$, $\gamma_+$ is holomorphic in a neighborhood of $D$, and $\gamma_-$ is holomorphic in the complement of $D$ in $\mathbb{P}^1(\mathbb{C})$. The renormalized value is given by $\gamma_+(D)$ and the counterterms by $\gamma_-(z)$.

The paper of Connes and Marcolli *Renormalization, the Riemann–Hilbert Correspondence*, and *Motivic Galois Theory* in volume II of Julia *et al.* (2005) shows that the data of the Birkhoff factorization are equivalently described in terms of solutions to a certain class of differential systems with irregular singularities. This is obtained by writing the terms in the Birkhoff factorization as time-ordered exponentials, and then using the fact that

$$\mathrm{T}e^{\int_a^b \alpha(t)\,\mathrm{d}t} := 1 + \sum_{n=1}^{\infty} \int_{a\le s_1\le\cdots\le s_n\le b} \alpha(s_1)\cdots\alpha(s_n)\,\mathrm{d}s_1\cdots\mathrm{d}s_n$$

is the value $g(b)$ at $b$ of the unique solution $g(t)\in G$ with value $g(a)=1$ of the differential equation $\mathrm{d}g(t) = g(t)\alpha(t)\,\mathrm{d}t$.

The singularity types are specified by physical conditions, such as the independence of the counterterms on the mass scale. These conditions are expressed geometrically through the notion of $G$-valued "equisingular connections" on a principal $\mathbb{C}^*$-bundle $B$ over a disk $\Delta$, where $G$ is the prounipotent Lie group of characters of the Connes–Kreimer Hopf algebra of Feynman graphs. The "equisingularity" condition is the property that such a connection $\omega$ is $\mathbb{C}^*$-invariant and that its restrictions to sections of the principal bundle that agree at $0\in\Delta$ are mutually equivalent, in the sense that they are related by a gauge transformation by a $G$-valued $\mathbb{C}^*$-invariant map regular in $B$; hence, they have the same type of (irregular) singularity at the origin.

The classification of equivalence classes of these differential systems via the Riemann–Hilbert correspondence and differential Galois theory yields a Galois group $U^* = U \rtimes \mathbb{G}_m$, where $U$ is prounipotent, with Lie algebra the free graded Lie algebra with one generator $e_{-n}$ in each degree $n\in\mathbb{N}$. The group $U^*$ is identified with the motivic Galois group of mixed Tate motives over the cyclotomic ring $\mathbb{Z}[e^{2\pi i/N}]$, for $N=3$ or $N=4$, localized at $N$.

### Speculations on Arithmetical Physics

In a lecture written for the 25th Arbeitstagung in Bonn, Y Manin presented intriguing connections between arithmetic geometry (especially Arakelov geometry) and physics. The theme is also discussed in Manin (1989). These considerations are based on a philosophical viewpoint according to which fundamental physics might, like adeles, have Archimedean (real or complex) as well as non-Archimedean ($p$-adic) manifestations. Since adelic objects are more fundamental and often simpler than their Archimedean components, one can hope to use this point of view in order to carry over some computation of physical relevance to the non-Archimedean side where one can employ number-theoretic methods.

**Adelic physics?** Some of the results mentioned in the previous sections seem to lend themselves well to this adelic interpretation. The quantum statistical mechanics of $\mathbb{Q}$-lattices relies fundamentally on adeles and it admits generalizations to systems associated to other algebraic varieties (Shimura varieties) that have an adelic description and adelic groups of symmetries. The result on the Polyakov measure also has an adelic flavor, as it uses essentially the Archimedean component of the Faltings height function. The latter is in fact a product of contributions from all the Archimedean and non-Archimedean places of the field of definition of algebraic points in the moduli space, so that one can expect that there would be an adelic Polyakov measure, of which one normally sees the Archimedean side only. The Freund–Witten adelic product formula for the Veneziano string amplitude fits in the same context, with $p$-adic amplitudes

$$B_p(\alpha, \beta) = \int_{\mathbb{Q}_p} |x|_p^{\alpha-1} |1-x|_p^{\beta-1}\, dx$$

and $B_\infty(\alpha,\beta)^{-1} = \prod_p B_p(\alpha,\beta)$ (cf. Varadarajan (2004)).

**Adelic physics and motives** A similar adelic philosophy was taken up by other authors, who proposed ways of introducing non-Archimedean and adelic geometries in quantum physics. A recent survey is given in Varadarajan (2004). For instance, Volovich (1995) proposed spacetime models based on cohomological realizations of motives, with étale topology "interpolating" between a proposed non-Archimedean geometry at the Planck scale and Euclidean geometry at the macroscopic scale. In this viewpoint, motivic $L$-functions appear as partition functions and actions of motivic Galois groups govern the dynamics.

*See also:* Hopf Algebra Structure of Renormalizable Quantum Field Theory; Mirror Symmetry: A Geometric Survey; Quantum Ergodicity and Mixing of Eigenfunctions; Random Matrix Theory in Physics; Regularization for Dynamical Zeta Functions.

## Further Reading


Beilinson A and Manin Yu (1986) The Mumford form and the Polyakov measure in string theory. *Communications in Mathematical Physics* 107(3): 359–376.

Bost JB and Connes A (1995) Hecke algebras, type III factors and phase transitions with spontaneous symmetry breaking in number theory. *Selecta Mathematica (N.S.)* 1(3): 411–457.

Broadhurst DJ and Kreimer D (1997) Association of multiple zeta values with positive knots via Feynman diagrams up to 9 loops. *Physics Letters* 393(3–4): 403–412.

Cartier P (2002) Fonctions polylogarithmes, nombres polyzetas et groupes pro-unipotents, Séminaire Bourbaki, Vol. 2000/2001. Astérisque No. 282 (2002), Exp. No. 885, viii, 137–173.

Connes A, Marcolli M, and Ramachandran N (2005) KMS states and complex multiplication. Preprint (to appear in *Selecta Mathematica*), arXiv math.OA/0501424.

Degiovanni P (1994) Moore and Seiberg equations, topological field theories and Galois theory. In: Leila S (ed.) *The Grothendieck Theory of Dessins D'enfants*. Cambridge: Cambridge University Press.

Julia BL, Moussa P, and Vanhove P (eds.) (2005) *Frontiers in Number Theory, Physics and Geometry*, vols. I, II, Papers from the Meeting held in Les Houches, March 9–21, 2003. Berlin: Springer.

Kreimer D (2000) *Knots and Feynman Diagrams*. Cambridge Lecture Notes in Physics, vol. 13. Cambridge: Cambridge University Press.

Lang S (1988) *Introduction to Arakelov Theory*. Berlin: Springer.

Manin Yu (1989) Reflections on arithmetical physics. In: Dita P and Georgescu V (eds.) *Conformal Invariance and String Theory*, pp. 293–303. Boston: Academic Press.

Manin Yu (1991) Three-dimensional hyperbolic geometry as $\infty$-adic Arakelov geometry. *Inventiones Mathematicae* 104(2): 223–243.

Manin Yu and Marcolli M (2001) Holography principle and arithmetic of algebraic curves. *Advanced Theoretical and Mathematical Physics* 5(3): 617–650.

Moch S, Uwer P, and Weinzierl S (2002) Nested sums, expansion of transcendental functions, and multiscale multiloop integrals. *Journal of Mathematical Physics* 43(6): 3363–3386.

Moore G (1998) Arithmetic and attractors, hep-th/9807087.

Nahm W (2005) Conformal field theory and torsion elements of the bloch group. In: Julia BL, Moussa P, and Vanhove P (eds.) *Frontiers in Number Theory, Physics and Geometry*, vol. I. Berlin: Springer.

Schneps L (ed.) (1994) *The Grothendieck Theory of Dessins d'enfants*. Cambridge: Cambridge University Press.

Serre JP (1992) Motifs, in Journées Arithmétiques, 1989 (Luminy, 1989). Astérisque No. 198–200, (1991), 11, 333–349.

Stevenhagen P (2001) Hilbert's 12th problem, complex multiplication and Shimura reciprocity. In: *Class Field Theory – Its Centenary and Prospect* (Tokyo, 1998), 161–176, Adv. Stud. Pure Math., 30, Math. Soc. Japan, Tokyo, 2001.

Varadarajan VS (2004) Arithmetic quantum physics: why, what and whether. *Proceedings of the Steklov Institute of Mathematics* 245: 258–265.

Volovich IV (1995) From $p$-adic strings to étale strings. *Proceedings of the Steklov Institute of Mathematics* 203(3): 37–42.

# O

## Operads

**J Stasheff**, Lansdale, PA, USA



### Introduction

An operad is an abstraction of a family of composable functions of $n$ variables for various $n$, useful for the "bookkeeping" and applications of such families. Operads are particularly important and useful in categories with a good notion of "homotopy," where they play a key role in organizing hierarchies of higher homotopies, reflecting their original use as a tool in homotopy theory, especially for studying (iterated) loop spaces. For several years now, operads have become increasingly important in mathematical physics, especially in string field theory, where they organize the terms of higher order in perturbed actions, and in deformation quantization.

The major focus of this article will be on operads as they are relevant to mathematical physics, but will also include some background material from homotopy theory, where they originated. A borderland where homotopy theory and cohomological physics overlap is the world of differential graded vector spaces, including those of differential forms, ghosts, anti-ghosts, etc., sometimes lumped together as BRST theory. Here, as elsewhere in contemporary mathematical physics, the flow has been in both directions – sometimes physicists have discovered or reinvented known mathematics but finding new applications, at other times physics has suggested new concepts for mathematicians to develop further. In the case of operads, they have provided general structure for varieties of algebras, some of which are novel types contributed by physicists.

For a reasonably up-to-date introduction and survey, consider Markl *et al.* (2002), although there have been many developments since then. Two particularly important original works are Boardman and Vogt (1973) and May (1972).

### Definitions and Examples

The term "operad" is due to May, building on work of Stasheff and of Boardman–Vogt. The most fundamental example of an operad is the endomorphism operad $\mathcal{E}nd_X := \{\mathrm{Map}(X^n, X)\}_{n\geq 1}$ where, for a set or topological space $X$, $\{\mathrm{Map}(X^n, X)\}$ means the set or space of functions or continuous functions from the $n$-fold product of $X$ with itself to $X$, together with the operations

$$\circ_i : Map(X^n, X) \times Map(X^m, X) \longrightarrow Map(X^{n+m-1}, X)$$

given, for $1 \leq i \leq n$, by

$$(f \circ_i g)(x_1, \ldots, x_{m+n-1}) \\ = f(x_1, \ldots, x_{i-1}, g(x_i, \ldots, x_{i+m-1}), x_{i+m}, \ldots)$$

In the endomorphism operad $\mathcal{E}nd_X$, there are easily discovered relations involving iterated $\circ_i$-operations and the symmetric group $\Sigma_n$ actions on the $X^n$s. For example,

$$(f \circ_i g) \circ_j h = f \circ_i (g \circ_{j-i+1} h) \\ \text{for } i \leq j \leq i + m - 1$$

if $g$ is a function of $m$ variables, since only the name of the position for the insertion is changed.

An operad $(\mathcal{O}, \circ_i)$ consists of a collection $\{\mathcal{O}(n)\}_{n\geq 1}$ of objects and maps $\circ_i : \mathcal{O}(n) \times \mathcal{O}(m) \to \mathcal{O}(n + m - 1)$ for $m, n \geq 1$ and $i \leq n$ satisfying the relations manifest in the example $\mathcal{E}nd_X$.

May's original definition corresponds to simultaneous insertions into all possible positions of inputs into $f \in Map(X^n, X)$. In most examples, the structures are "manifest" without appeal to the technical definitions.

It helps to see graphic examples of operads, particularly ones relevant for physics. Two kinds that are particularly important are the tree operads and the little disks (or cubes) operads.

Let $\mathcal{T}(n)$ be the set of planar trees with one root and $n$ leaves labeled (arbitrarily) 1 through $n$. The collection $\mathcal{T} = \{\mathcal{T}(n)\}_{n\geq 1}$ of sets of trees forms an operad by grafting the root of $g$ to the leaf of $f$ labeled $i$, as in **Figure 1**, where the leaves are assumed labeled in order from left to right. **Figure 1** can be interpreted as portraying the $\circ_4$ result of inserting a 3-linear operation into a 5-linear one.

The little $n$-disks operad $\mathcal{D}_n = \{\mathcal{D}_n(j)\}_{j\geq 1}$ where $\mathcal{D}_n(j)$ consists of an ordered collection of $j$ $n$-disks

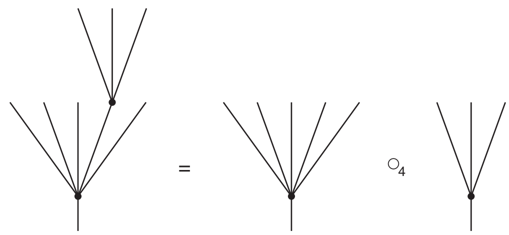



**Figure 1** Grafting with the leaves numbered from left to right.

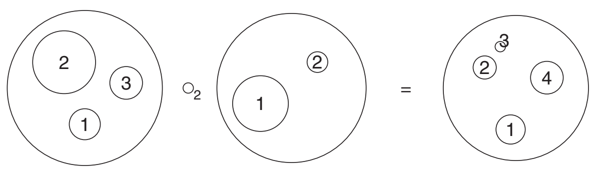



**Figure 2** The little 2-disks operad.

embedded in the standard $n$-dimensional unit disk $D^n$ with disjoint interiors, the embedding being of the form $az + b$ with $0 < a \in \mathbb{R}$. The operations are given as indicated in Figure 2.

Just as group theory without representations is rather sterile, so are operads best appreciated by their representations known as (varieties of) algebras, especially algebras with higher homotopies.

An algebra $A$ over an operad $\mathfrak{P}$ "is" a map of operad $\mathfrak{P} \to \mathcal{E}nd_A$. This is just a compact way of saying that an algebra $A$ has a coherent system of maps $\mathfrak{P}(n) \times A^n \to A$. Much of this article will speak in terms of such algebras with the corresponding operad being understood.

## Operads in Homotopy Theory

A major motivation for the development of operads was the desire to have a homotopy-invariant characterization of based loop spaces and iterated loop spaces. Precisely such coherent systems of higher homotopies provided the answers. For based loop spaces, the operad in question $\mathcal{K} = \{K_n\}_{n\geq 1}$ consists of the polytopes known as "associahedra." The usual product of based loops is only homotopy associative.

If we fix a specific associating homotopy and consider the five ways of parenthesizing the product of four loops, there results a pentagon whose edges correspond to a path of loops (Figure 3).

From the leftmost vertex to the rightmost, consider the two paths of loops across the top or around the bottom. By further adjustment of parameters, the pentagon can be filled in by a family of such paths.

The associahedron $K_n$ can be described as a convex polytope with one vertex for each way of associating $n$ ordered variables, that is, ways of inserting parentheses in a meaningful way in a word

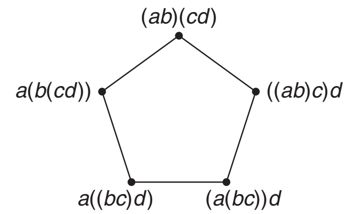



**Figure 3** The associahedron $K_4$.

of $n$ letters. The edges correspond to a single application of an associating homotopy. More generally, the cellular structure of the associahedra is well described by planar rooted trees, the vertices corresponding to binary trees and so forth (see Figure 4).

For $K_5$, see Figure 5 or a rotatable image available at http://igd.univ-lyon1.fr/~ chapoton/stasheff.html. The facets are all products of two associahedra of lower dimension and specific imbeddings can be given to play the role of the $\circ_i$ operations as in an operad.

An $A_\infty$-space is a space $Y$ which admits a coherent family of maps

$$m_n : K_n \times Y^n \to Y$$

so that they make $Y$ an algebra over the operad (without $\Sigma_n$-actions) $\mathcal{K} = \{K_n\}_{n\geq 1}$.

The main result by Stasheff is: A connected space Y (of the homotopy type of a CW-complex) has the homotopy type of a based loop space $\Omega X$ for some X if and only if Y is an $A_\infty$-space.

Homotopy characterization of iterated loop spaces $\Omega^n X_n$ for some space $X_n$ required the full power of the theory of operads with the symmetries.

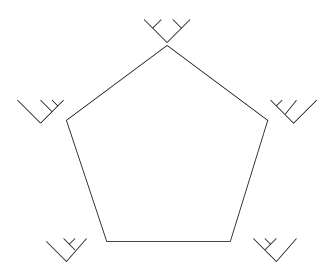


**Figure 4** $K_4$ with vertices labeled by trees.

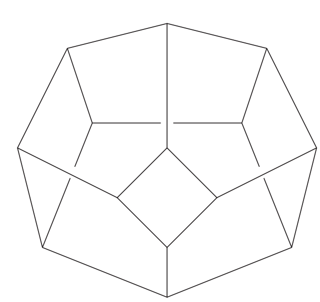


**Figure 5** The associahedron $K_5$.

An early motivation for the invention of a theory of operads was the consideration of infinite loop spaces, that is, a sequence of spaces $X_n$ such that each $X_n$ is homotopy equivalent to $\Omega X_{n+1}$.

Although introduced originally in the category of topological spaces, operads were available almost immediately for differential graded (dg) vector spaces, also known as chain complexes. In physics, the differential is often called a BRST operator, a term that should be reserved for a special kind of dg algebra, see below.

## Operads in Algebra

The $\circ_i$ notation first appeared in Gerstenhaber's study of the algebraic structure of the Hochschild cohomology of an associative algebra, about the same time as the construction of the associahedra where the operations were given in a less convenient notation. Recall the Hochschild cohomology of an associative algebra $A$ is the homology of the complex $\mathrm{Hom}(A^{\otimes n}, A)$ with the coboundary given as follows (all signs below are indicated as $\pm$, any of the standard references will specify conventions and signs): for $f \in \mathrm{Hom}(A^{\otimes n}, A)$ and $g \in \mathrm{Hom}(A^{\otimes m}, A)$ let

$$f \circ g = \Sigma_1^n \pm f \circ_i g \qquad [1]$$

Gerstenhaber then defines his bracket as $[f, g] = f \circ g \pm g \circ f$. With hindsight, he realized that the Hochschild coboundary can be written as

$$\delta h = [m, h] \qquad [2]$$

where $m : A \otimes A \to A$ is the multiplication. Moreover, the associativity of $m$ is equivalent to

$$[m, m] = 0 \qquad [3]$$

### $A_\infty$-Algebras

In the setting of graded vector spaces $V = \oplus_{r \in \mathbb{Z}} V^r$, there are two conventions for defining $A_\infty$-algebras, which differ by a shift in grading. We adopt the physics convention so that $A$ here is the suspension of that considered in the original papers. The cellular chains of the associahedra form the $A_\infty$-operad, providing the following definition.

**Definition 1** $A_\infty$-algebra (Strong homotopy associative algebra). Let $A$ be a $\mathbb{Z}$-graded vector space $A = \oplus_{r \in \mathbb{Z}} A^r$ and suppose that there exists a collection of degree 1 multilinear maps

$$\mathfrak{m} := \{m_k : A^{\otimes k} \to A\}_{k \geq 1}$$

$(A, \mathfrak{m})$ is called an $A_\infty$-algebra when the multilinear maps $m_k$ satisfy the following relations:

$$\sum_{p+q=n+1} \sum_{i=1}^{p} \pm m_p \circ_i m_q = 0 \qquad [4]$$

with an appropriate set of signs for $n \geq 1$.

A weak $A_\infty$-algebra consists of a collection of degree 1 multilinear maps

$$\mathfrak{m} := \{m_k : A^{\otimes k} \to A\}_{k \geq 0}$$

satisfying the above relations, but for $n \geq 0$ and in particular with $k, l \geq 0$.

**Remark 1** The "weak" version is fairly new, inspired by physics, where $m_0 : \mathbb{C} \to A$, regarded as an element $m_0(1) \in A$, is related to what physicists refer to as a "background." The augmented relation then implies that $m_0(1)$ is a cycle, but $m_1 m_1$ need no longer be 0, rather

$$m_1 m_1 = \pm\, m_2(m_0 \otimes 1) \pm\, m_2(1 \otimes m_0) \qquad [5]$$

Just as associativity was captured by the equation $[m, m] = 0$, so the defining relations of the definition of an $A_\infty$-algebra are captured by

$$[\mathfrak{m}, \mathfrak{m}] = 0 \qquad [6]$$

Decades later it was realized that considering $T^c A = \Sigma A^{\otimes n}$ as a coalgebra with

$$\Delta(a_1 \otimes \cdots \otimes a_n) = \Sigma_{p+q} \pm (a_1 \otimes \cdots \otimes a_p) \otimes (a_{p+1} \otimes \cdots \otimes a_n)$$

we then have an isomorphism

$$\Sigma \mathrm{Hom}(A^{\otimes n}, A) \simeq \mathrm{Coder}(T^c A)$$

Here Coder is the space of all coderivations of $T^c A$. The Gerstenhaber bracket is indeed the "intrinsic" commutator bracket of coderivations via the above isomorphism. As such, it satisfies a graded version of the Jacobi identity; after a shift in grading from the original one of Hochschild, the Hochschild cochain complex forms a dg Lie algebra.

### $L_\infty$-Algebras

Since an ordinary Lie algebra $\mathfrak{g}$ is regarded as ungraded, the defining bracket is regarded as skew-symmetric. For dg Lie algebras and $L_\infty$-algebras, we need graded symmetry, which refers to symmetry with signs determined by the grading. The basic operation is

$$\tau : x \otimes y \mapsto (-1)^{|x||y|} y \otimes x \qquad [7]$$

Also we adopt the convention that tensor products of graded functions or operators have the signs built

in; for example, $(f \otimes g)(x \otimes y) = (-1)^{|g||x|} f(x) \otimes g(y)$. By decomposing each permutation as a product of transpositions, there is then defined the sign of a permutation of $n$ graded elements, for example, for any $c_i \in V, 1 \le i \le n$, and any $\sigma \in \mathfrak{S}_n$, the permutation of $n$ graded elements is defined by

$$\sigma(c_1, \ldots, c_n) = (-1)^{\epsilon(\sigma)}(c_{\sigma(1)}, \ldots, c_{\sigma(n)}) \qquad [8]$$

The sign $(-1)^{\epsilon(\sigma)}$ is often referred to as the Koszul sign of the permutation.

**Definition 2** (Graded symmetry). A graded symmetric multilinear map of a graded vector space $V$ to itself is a linear map $f: V^{\otimes n} \to V$ such that for any $c_i \in V, 1 \le i \le n$, and any $\sigma \in \mathfrak{S}_n$ (the permutation group of $n$ elements), the relation

$$f(c_1, \ldots, c_n) = (-1)^{\epsilon(\sigma)} f(c_{\sigma(1)}, \ldots, c_{\sigma(n)}) \qquad [9]$$

holds.

**Definition 3** By a $(k, l)$-unshuffle of $c_1, \ldots, c_n$ with $n = k + l$ is meant a permutation $\sigma$ such that for $i < j \le k$, we have $\sigma(i) < \sigma(j)$ and similarly for $k < i < j \le k + l$. We denote the subset of $(k, l)$-unshuffles in $\mathfrak{S}_{k+l}$ by $\mathfrak{S}_{k,l}$ and by $\mathfrak{S}_{k+l=n}$, the union of the subsets $\mathfrak{S}_{k,l}$ with $k + l = n$. Similarly, a $(k_1, \ldots, k_i)$-unshuffle means a permutation $\sigma \in \mathfrak{S}_n$ with $n = k_1 + \cdots + k_i$ such that the order is preserved within each block of length $k_1, \ldots, k_i$. The subset of $\mathfrak{S}_n$ consisting of all such unshuffles we denote by $\mathfrak{S}_{k_1,\ldots,k_i}$.

**Definition 4** $L_\infty$-algebra (Strong homotopy Lie algebra). Let $L$ be a graded vector space and suppose that a collection of degree 1 graded symmetric linear maps $\mathfrak{l} := \{l_k : L^{\otimes k} \to L\}_{k \ge 1}$ is given. $(L, \mathfrak{l})$ is called an $L_\infty$-algebra iff the maps satisfy the following relations:

$$\sum_{\sigma \in \mathfrak{S}_{k+l=n}} (-1)^{\epsilon(\sigma)} l_{1+l}(l_k(c_{\sigma(1)}, \ldots, c_{\sigma(k)}), \\ [c_{\sigma(k+1)}, \ldots, c_{\sigma(n)}) = 0 \qquad [10]$$

for $n \ge 1$.

A weak $L_\infty$-algebra consists of a collection of degree 1 graded symmetric linear maps $\mathfrak{l} := \{l_k : L^{\otimes k} \to L\}_{l \ge 0}$ satisfying the above relations, but for $n \ge 0$ and with $k, l \ge 0$.

**Remark 2** The alternate definition in which the summation is over all permutations, rather than just unshuffles, requires the inclusion of appropriate coefficients involving factorials.

Just as an $A_\infty$-algebra can be described as a coderivation of $T^c A$, similarly an $L_\infty$-algebra $L$ can be described as a coderivation on $S^c L$, the symmetric subcoalgebra of $T^c A$.

The operad of Lie algebras was defined rather late, although it was earlier implicit in the work of Fred Cohen. It is defined as the homology $H_{n-1}(\mathrm{Config}(\mathbb{R}^2, n))$ for $n \ge 1$, where $\mathrm{Config}(\mathbb{R}^2, n)$ denotes the configuration space of ordered $n$-tuples of distinct points in $\mathbb{R}^2$. Equivalently, the configurations can be thought of as the centers of the little 2-disks. The open disks being contractible to their centers, this is a suboperad of the full homology $H_\bullet(\mathcal{D}_2)$.

Just as a Lie algebra is obtained from an associative algebra using the commutator as bracket and, inversely, a Lie algebra gives rise to its universal enveloping associative algebra, an $L_\infty$-algebra can be obtained from an $A_\infty$-algebra by $n$-variable analogs of commutators and there is a universal enveloping $A_\infty$-algebra of a given $L_\infty$-algebra.

### Open–Closed Homotopy Algebras

Open–closed string field theory suggests interaction between an $L_\infty$-algebra $\mathcal{H}_c$ and an $A_\infty$-algebra $\mathcal{H}_o$ including a strong homotopy representation of $\mathcal{H}_c$ on $\mathcal{H}_o$ by strong homotopy derivations. Here is the formal definition:

**Definition 5** Let $\mathcal{H} = \mathcal{H}_o \oplus \mathcal{H}_c$ be a graded vector space and $(\mathcal{H}_c, \mathfrak{l})$ be a weak $L_\infty$-algebra. Consider a collection of multilinear maps

$$\mathfrak{n} := \{n_{k,l} : (\mathcal{H}_o)^{\otimes k} \otimes (\mathcal{H}_c)^{\otimes l} \to \mathcal{H}_o\}_{k,l \ge 0}$$

each of which is graded symmetric on $(\mathcal{H}_c)^{\otimes l}$. We denote the collection also by $\mathfrak{n}$. We call $(\mathcal{H}, \mathfrak{n}, \mathfrak{l})$ a (partial) open-closed homotopy algebra (OCHA) when $\mathfrak{n}$ satisfies the following relations (up to some factorial coefficients):

$$0 = \sum_{k,l \ge 0} \sum_{p=0}^{m-k} \sum_{\sigma \in \mathfrak{S}_n} \pm n_{m+1-k,n-l}(o_1, \ldots, o_p, \\ n_{k,l}(o_{p+1}, \ldots, o_{p+k}; c_{\sigma(1)}, \ldots, c_{\sigma(l)}), \\ o_{p+k+1}, \ldots, o_m; c_{\sigma(l+1)}, \ldots, c_{\sigma(n)}) \\ + \sum_{\sigma \in \mathfrak{S}_n} \sum_{l=1}^{n} \pm n_{m,n+1-l}(o_1, \ldots, o_m; \\ l_l(c_{\sigma(1)}, \ldots, c_{\sigma(l)}), c_{\sigma(l+1)}, \ldots, c_{\sigma(n)}) \qquad [11]$$

### Other Algebras of Interest

The Hochschild complex also has a graded product (without invoking the shift) known as the cup product. Except for the signs and the grading, the bracket and the product satisfy the Leibniz rule of a Poisson algebra on the cohomology; the result is

axiomatized as a "Gerstenhaber algebra." However, on the cochain complex, the Lie bracket and the associative product are compatible only up to homotopy.

This naturally raises the issue of an operad for strong homotopy Gerstenhaber algebras. The operad $\mathcal{G}$ for Gerstenhaber algebras is the homology of the little disks operad, $H_\bullet(\mathcal{D}_2)$. But now we have choices: in addition to relaxing the Leibniz rule up to homotopy, the bracket could be relaxed to be part of an $L_\infty$-algebra and/or the product could be relaxed to be part of an $A_\infty$-algebra. The choice which is now known as the $G_\infty$-operad is defined in terms of a procedure which works for what are known as quadratic operads, indicating they have generators in $\mathcal{O}(2)$ and relations in $\mathcal{O}(3)$: the corresponding $\mathcal{O}_\infty$ has "dual" relations. For example, this gives the classical Koszul duality between Lie and commutative associative algebras. The $\mathcal{G}_\infty$-operad can also be described as the "minimal model" of $\mathcal{G}$ in the sense of Markl.

Another alternative is to consider just the "brace" operations, originally introduced by Kadeishvili and later independently by Getzler, but described in the Hochschild complex setting by Gerstenhaber–Voronov. Together with the cup product, these determine an operad denoted $\mathcal{HG}$ which acts on the Hochschild complex; there is an operad map from $\mathcal{G}_\infty$ to $\mathcal{HG}$, hence $\mathcal{G}_\infty$ also acts on the Hochschild complex. Finally, Tamarkin showed that $\mathcal{G}_\infty$ is quasi-isomorphic to the dg operad of singular chains on the little disks operad, thus providing one of several proofs of what had been a conjecture by Deligne.

Algebras with invariant inner products $<-,->$ are of considerable importance in mathematics and especially in mathematical physics; invariance means $<a, bc> = <ab, c>$ or $<a, [b, c]> = <[a, b], c>$ in, respectively, the associative or the Lie case (with appropriate signs in the graded case). Using the inner product, $n$-ary operations $A^{\otimes n} \to A$ can be converted to operations $A^{\otimes n+1} \to \mathbb{C}$ of which we can require cyclic symmetry. To handle such algebras, there is a notion of "cyclic operad." In terms of trees, the transition is to take a rooted tree and then regard the root edge as just another leaf. This point of view corresponds to an essential symmetry for particle interactions.

## Operads in Mathematical Physics

One reason for the explosive development of operad theory in the 1990s was the introduction of operadic structures in field theories, for example, conformal field theories (CFTs) and string field theories (SFTs). These operadic structures were directly related to the moduli spaces of Riemann surfaces with punctures or boundaries (or other decorations) in these physical theories.

Two special "higher-homotopy algebras" have been emphasized because they are particularly important in mathematical physics: $A_\infty$ for open-string field theory and $L_\infty$ for closed-string field theory and for deformation quantization. Open–closed string field theory combines $A_\infty$-algebra and $L_\infty$-algebra in a particular way known as an OCHA.

The operad for $L_\infty$-algebras is given a very nice and physically relevant geometric interpretation in terms of a real compactification of the moduli space of Riemann spheres with punctures, while for OCHAs, there is a real compactification of the moduli space of Riemann disks with punctures on the boundary or in the interior (bulk). Thus, this operad can be regarded as obtained from a moduli space of configurations of points (punctures) in the disk by compactifying the moduli spaces by adding boundary strata where two (or more) points (punctures) collide. Points on the boundary strata can be visualized as "bubble trees" of disks and/or spheres, see Figure 6. Alternatively, the little disks operad can be regarded as being obtained by "decorating" the points with little disks, while for OCHAs there is also a basic half-disk decorated with little disks in the bulk and little half-disks for the boundary points. The corresponding colored operad is Voronov's "Swiss-cheese operad." "Colored" refers to the fact that disks can be inserted into half-disks but not vice versa. Compare trees with two "colors" of edges with grafts restricted to ones which match colors.

### On-Shell versus Off-Shell

In cohomological physics, the "on-shell" states or observables are usually given by the cohomology with respect to an internal differential, which in physics is called the BRST differential or BRST operator, though originally this meant the Chevalley–Eilenberg differential associated to the action of the Lie algebra of

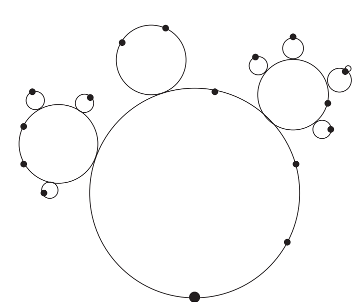


**Figure 6** Bubble tree for circle configurations.

gauge symmetries of a physical theory. The generators of the Chevalley–Eilenberg cochain complex are known as "ghosts". On-shell subspaces of algebras which are not closed under the product of the larger "off-shell" algebra are called "open" algebras by physicists. Quite generally, this situation gives rise to an algebra over an appropriate operad. A special case involves a differential graded algebra $A$ and a linear imbedding $H(A) \subset A$. The (co)homology is in turn a graded algebra (with 0 as differential), but inherits a higher-homotopy structure so that cohomology and original algebra are equivalent.

In the associative case, the inheritance is a result of Kadeishvili:

*Let* $(A,\ d)$ *be a differential graded associative or* $A_\infty$*-algebra, then the homology* $H(A)$ *inherits the structure of an* $A_\infty$*-algebra.*

Even if the original algebra $A$ is strictly associative, the inherited $A_\infty$-structure generally has nontrivial operations $m_i$.

Analogous results hold for $L_\infty$-algebras and others. It is the $L_\infty$-version that is relevant for closed-string field theory (CSFT). Zwiebach showed the quantum theory of covariant closed strings has an action defined in terms of an infinite chain of string field products. The genus-0 (tree level) string field algebra is an $L_\infty$-algebra inherited from the off-shell state space modeled by the Batalin–Vilkovisky (BV) construction. The higher-order brackets provide higher-order correlation or $n$-point functions which play a crucial role in the extended Lagrangian of the theory.

### Batalin–Fradkin–Vilkovisky and Batalin–Vilkovisky Constructions

The constructions of Batalin–Fradkin–Vilkovisky (BFV) for constrained Hamiltonian systems and of Batalin–Vilkovisky (BV) for Lagrangians with symmetries are important examples of $L_\infty$-structures derived from "open" algebra settings, though the $L_\infty$-structures were recognized quite a while after the constructions.

The BFV setting is that of a symplectic manifold $W$ with a family of constraints, that is, a family of functions $\phi^\alpha \in C^\infty(W)$. The constraints are called "first class" if the ideal they generate is closed under the Poisson bracket. The vector space spanned by the constraints will in general be an open algebra; the structure of the bracket is given by structure functions, rather than structure constants. The zero locus of all the constraints forms the constraint surface $V$. In the first-class case, the constraints are in involution and determine a foliation $\mathcal{F}$ of $V$. If the space of leaves $V/\mathcal{F}$ is a manifold, it would be considered the true physical space and the physical observables would be functions in $C^\infty(V/\mathcal{F})$. BFV construct a differential graded Poisson algebra such that the cohomology in degree 0 agrees with $C^\infty(V/\mathcal{F})$ when that makes sense and, in the regular case, the rest of the cohomology is that of the differential forms along the leaves of the foliation. The BFV differential is a deformation of the Chevalley–Eilenberg/BRST differential and can be constructed most efficiently by the same techniques used in proving Kadeishivili's inheritance theorem. Crucially, it is an inner derivation with respect to the Poisson bracket. After the fact, an $L_\infty$-structure can be observed in the extended algebra.

For a Lagrangian with symmetries, BV develop a similar construction, the main difference being that there is no Poisson bracket initially, but one is constructed by adjoining "anti-fields" as conjugate to the fields but of ghost degree –1 and the differential of an anti-field being the Euler–Lagrange expression for the corresponding field. Then, as in the Hamiltonian case, ghosts and anti-ghosts, etc. are adjoined and the construction proceeds in a parallel fashion.

### Deformation Quantization

Once algebras over an operad $\mathfrak{P}$ are considered, it is natural to consider also morphisms of such algebras over a fixed $\mathfrak{P}$.

From a homotopy point of view, the appropriate maps need not respect the operad structure strictly but only up to higher homotopy; indeed, there is a related operad to define such maps. For $L_\infty$-algebras, such $L_\infty$-maps play a key role in deformation quantization. That refers to deformation of the commutative multiplication of a Poisson algebra in the direction of the Poisson bracket; that is, to first order, the deformation is given by the bracket.

More generally, for any associative algebra $A$ with multiplication $m$, one considers formal deformations

$$a \star b = m(a,b) + tm_1(a,b) + t^2 m_2(a,b) + \cdots \quad [12]$$

where each $m_i \in \mathrm{Hom}(A \otimes A, A)$. The associativity of $\star$ provides a sequence of constraints on the $m_i$. In particular, $m_1$ must be a Hochschild cocycle and the obstruction to the existence of $m_2$ is a class in the Hochschild cohomology of degree 3. In fact, the primary obstruction is represented by $[m_1, m_1]$. If it is cohomologous to zero, that fact identifies candidates for $m_2$, that is,

$$[m_1, m_1] = \pm 2[m, m_2] \quad [13]$$

or, using the notation $d = [m,]$,

$$dm_2 - 1/2[m_1, m_1] = 0 \quad [14]$$

once known as the integrability equation but now, more frequently, as a Maurer–Cartan equation. For a Poisson algebra, the Poisson bracket is a Hochschild cocycle but in general a full deformation need not exist. However, for the algebra $A$ of smooth functions on a Poisson (e.g., symplectic) manifold $M$, Kontsevich showed that such a full formal deformation does exist.

The guiding philosophy is that deformations are controlled by a dg Lie or $L_\infty$-algebra $L$, unique up to $L_\infty$-homotopy equivalence. Therefore, the obstructions can be computed in any of the equivalent dg Lie algebras. Moreover, the structure of the obstructions is known sufficiently so that if there is an equivalent dg Lie algebra with $d$ in fact zero, then all the obstructions to deformation quantization vanish. The key to Kontsevich's proof was the construction of an $L_\infty$-map, inducing an isomorphism in cohomology, from the Lie algebra of polyvector fields on $\mathbb{R}^d$ with the Schouten bracket and $d=0$ to the Lie algebra of multidifferential operators on $A=C^\infty(\mathbb{R}^d)$ regarded as a subalgebra of the Hochschild cochain complex for $A$ with the Gerstenhaber bracket.

### BV Algebras

In addition to their construction of a differential graded Gerstenhaber algebra (a differential graded commutative algebra with a compatible Poisson bracket of degree 1), BV introduced a new mathematical structure, adding a second-order differential operator $\Delta$ relating the commutative product and the bracket. The operator $\Delta$ is a derivation of the bracket and of square zero. Moreover,

$$[a,b] = \Delta(ab) - \Delta(a)b \pm a\Delta(b) \qquad [15]$$

so that the failure of $\Delta$ to be a derivation of the product is given by the bracket.

The definition of a BV algebra is then a Gerstenhaber algebra with such an operator, though alternative definitions exist in which $\Delta$ and the product are primary and the bracket is defined by the above equation. From the operadic/higher-homotopy point of view, one can then go on to consider $BV_\infty$ algebras.

Recall that $A_\infty$-algebras and $L_\infty$-algebras (among others) can be characterized by an "inner" coderivation $d=[m,\ ]$ of square zero on an appropriate "standard" construction. In the context of BV algebras, where the bracket is more commonly written as {,}, the classical action is an element $S_0$ such that $\{S_0,S_0\}=0$ or, equivalently, $d=\{S_0,\ \}$ is of square zero. The quantum analog $S$ is a perturbation of $S_0$ and satisfies instead

$$\{S,S\} = \Delta S \qquad [16]$$

This was originally called the "master equation," but now is increasingly referred to as a "Maurer–Cartan" equation.

### Insertion Operads

There is another class of operads illustrated by trees (and more generally graphs) with a very different sort of "composition," namely insertion of one graph into another. The most directly relevant to physics is the kind of insertion used by Connes and Kreimer in their Hopf algebra constructed for renormalization of Feynman diagrams. For example, consider all finite graphs with exactly two external edges and internal numbered edges. Given two graphs $\Gamma_1,\Gamma_2$, define $\Gamma_1\circ_i\Gamma_2$ by cutting edge $i$ of $\Gamma_1$ and identifying the dangling edges with the two external edges of $\Gamma_2$.

For planar trees, yet another insertion operad is obtained by Chapoton, isolating a part of a structure due to Kontsevich, in which a small neighborhood of a vertex of the second planar tree is removed and the dangling edges are attached to a vertex of the first tree by entering through the angles between the edges at that vertex (Figure 7).

Inside the $\mathcal{HG}$-operad is the operad $\mathcal{B}race$ for an abstract brace algebra (forgetting the cup product), first described as such by Chapoton using the insertion operations of Kontsevich and Soibelman.

### $A_\infty$-Categories

Also of importance for applications to mathematical physics is the notion of an $A_\infty$-category, first made explicit by Fukaya and now playing a major role in string D-brane theory and homological mirror symmetry. The D-branes are the objects of the $A_\infty$-category and the open strings with boundaries on two (possibly equal) D-branes $B_1,B_2$ are the morphisms from $B_1$ to $B_2$. The operations $m_i$ are defined only on tuples $(a_1,\ldots,a_i)$ of "composable" morphisms (e.g., strings).

### PROPs

While an operad is an abstraction of a family of composable functions of $n$ variables for various $n$, a PROP is an abstraction of a family of functions in

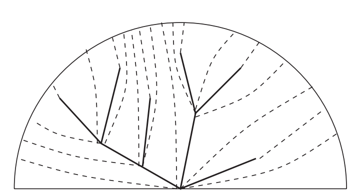


**Figure 7** Angles determined by edges with leaves extended to the semicircle.

Hom($A^{\otimes p}, A^{\otimes q}$) for all $p$ and $q$. Now the relevant images are graphs with $p$ input legs and $q$ output legs with composition being defined by grafting output legs of one graph to inputs of another. Feynman diagrams are the obvious example in physics or, in conformal field theory, tubular neighborhoods of such graphs, which is to say, Riemann surfaces with boundary circles: $p$ as inputs and $q$ as outputs.

*See also:* Algebraic Approach to Quantum Field Theory; Batalin–Vilkovisky Quantization; Constrained Systems; Deformations of the Poisson Bracket on a Symplectic Manifold; Deformation Quantization; Deformation Theory; Hopf Algebra Structure of Renormalizable Quantum Field Theory; String Field Theory.

## Further Reading

Batalin IA and Fradkin ES (1983) A generalized canonical formalism and quantization of reducible gauge theories. *Physics Letters B* 122: 157–164.

Batalin IA and Vilkovisky GS (1981) Gauge algebra and quantization. *Physics Letters B* 102: 27–31.

Boardman JM and Vogt RM (1973) *Homotopy Invariant Algebraic Structures on Topological Spaces*, Lecture Notes in Mathematics, vol. 347. Berlin–New York: Springer.

Connes A and Kreimer D (2000) Renormalization in quantum field theory and the Riemann–Hilbert problem, I: the Hopf algebra structure of graphs and the main theorem. *Communications in Mathematical Physics* 210: 249–273, (hep-th/9912092).

Fradkin ES and Vilkovisky GS (1975) Quantization of relativistic systems with constraints. *Physics Letters B* 55: 224–226.

Fukaya K (1997) Floer homology, $A_\infty$-categories and topological field theory. In: Andersen J, Dupont J, Pertersen H, and Swan A (eds.) *Geometry and Physics*, Lecture Notes in Pure and Applied Mathematics, Notes by Seidel P, vol. 184, pp. 9–32. New York: Marcel Dekker, Inc.

Gerstenhaber M (1963) The cohomology structure of an associative ring. *Annals of Mathematics* 78: 267–288.

Getzler E (1995) Operads and moduli spaces of genus 0 Riemann surfaces. In: Dijkgraaf R, Faber C, and van der Geer (eds.) *The Moduli Space of Curves*, Progr. Math., vol. 129, pp. 199–230. Boston: Birkhäuser.

Getzler E and Jones JDS (1993) *n*-algebras and Batalin–Vilkovisky algebras. *Preprint*.

Getzler E and Jones JDS (1994) Operads, homotopy algebra and iterated integrals for double loop spaces. *Preprint*, Department of Mathematics, MIT; Department of Mathematics Northwestern University, March 1994, hep-th/9403055.

Getzler E and Kapranov M (1994) Modular operads. *Compositio Mathematica* 110: 65–126 (dg-ga/9408003).

Hinich V and Schechtman V (1993) Homotopy Lie algebras. *Advanced Studies in Soviet Mathematics* 16: 1–18.

Lada T and Markl M (1995) Strongly homotopy Lie algebras. *Communications in Algebra* 2147–2161 (hep-th/9406095).

Lada T and Stasheff JD (1993) Introduction to sh Lie algebras for physicists. *International Journal of Theoretical Physics* 32: 1087–1103.

Markl M, Shnider S, and Stasheff J (2002) *Operads in Algebra, Topology and Physics,* Mathematical Surveys and Monographs, vol. 96, MR 2003f: 18011. Providence, RI: American Mathematical Society.

May JP (1972) *The Geometry of Iterated Loop Spaces*, Lecture Notes in Mathematics, vol. 271. Springer.

Stasheff J (1963) Homotopy associativity of *H*-spaces, I, II. *Transactions of the American Mathematical Society* 108: 293–312, 313–327.

Voronov AA (1999) *The Swiss Cheese Operad*, Contemp. Math. vol. 239, pp. 365–373. Providence, RI: American Mathematical Society.

Zwiebach B (1993) Closed string field theory: quantum action and the Batalin–Vilkovisky master equation. *Nuclear Physics B* 390: 33–152.

Zwiebach B (1998) Oriented open–closed string theory revisited. *Annals of Physics* 267: 193–248 (hep-th/9705241).

# Operator Product Expansion in Quantum Field Theory

**H Osborn**, University of Cambridge, Cambridge, UK



## Introduction

The operator product expansion (OPE) provides an algebraic structure in quantum field theory. In a sense it supercedes or rather transcends the equal-time commutation relations, which provide the traditional starting point for the canonical quantization of any quantum field theory. The essential idea is that for any two local operator quantum fields at spacetime points $x_1, x_2$ their product may be expressed in terms of a series of other local quantum fields at a point $x$, which may be identified with $x_1$ or $x_2$, times c-number coefficient functions which depend on $x_1 - x_2$. The set of operators which may appear depends on the particular quantum field theory and must of course be in accord with any requirements of conserved quantum numbers. The coefficient functions depend on $x_1 - x_2$ in a fashion which depends on the dimensions of the various operators involved, at least up to renormalization group corrections. The most singular contributions are those for the operators appearing in the OPE with lowest scale dimension. From a phenomenological point of view, only the first few terms in the OPE are of relevance. However, theoretically, especially for conformal field theories, it is desirable to know the full expansion to all orders in powers of $x_1 - x_2$ in such a way that the operator product may

be replaced by the full expansion in appropriate correlation functions. We first discuss the OPE for free theories and then the interacting case.

## Free Field Theory

The OPE is most straightforward in free field theory when it almost reduces to a Taylor series expansion. For a simple free massless scalar field $\phi(x)$ then in four dimensions we may write

$$\phi(x)\phi(0) = \frac{C}{x^2} + :\phi(x)\phi(0): \qquad [1]$$

where :: denotes normal ordering (moving all annihilation operators to the right of creation operators) and $C$ is just a normalization numerical constant (for canonical normalization $C = 1/4\pi^2$). The $1/x^2$ term proportional to the identity operator reflects the leading singular behavior at short distances of $\phi(x)\phi(0)$, the power being determined by $\phi$ having dimension 1. For the normal-ordered term we may expand in terms of an infinite set of local operators by using the Taylor expansion

$$:\phi(x)\phi(0): = \sum_{n=0}^{\infty} \frac{1}{n!} x^{\mu_1} \cdots x^{\mu_n} :\partial_{\mu_1} \cdots \partial_{\mu_n}\phi(0)\phi(0): \qquad [2]$$

where the operator appearing in the $n$th term has dimension $n+2$. Manifestly at short distances only the leading terms are relevant. Equation [1] also provides a point splitting definition of the local composite operator $:\phi^2(0):$ in terms of limit of $\phi(x)\phi(0)$ as $x \to 0$ after subtraction of the singular $C/x^2$ term.

The OPE can be easily generalized to composite operators defined by normal ordering. For $:\phi^2:$ we have, by applying Wick's theorem,

$$:\phi^2(x)::\phi^2(0): = \frac{2C^2}{x^4} + \frac{4C}{x^2} :\phi(x)\phi(0): + :\phi^2(x)\phi^2(0): \qquad [3]$$

where Taylor series expansion may be applied to both $:\phi(x)\phi(0):$ and also $:\phi^2(x)\phi^2(0):$ to give an infinite sequence of local operators of increasing dimensions.

The expansion in terms of local operators may be reordered. For instance, from [1] we may write, using $\partial^2\phi = 0$,

$$\phi(x)\phi(0) = \frac{C}{x^2} + \left(1 + \tfrac{1}{2}x^\mu\partial_\mu + \tfrac{1}{4}x^\mu x^\nu \partial_\mu\partial_\nu + \tfrac{1}{16}x^2\partial^2\right) :\phi^2(0): - \tfrac{1}{2}x^\mu x^\nu T_{\mu\nu} + \mathrm{O}(x^3) \qquad [4]$$

where

$$T_{\mu\nu} = :\partial_\mu\phi\partial_\nu\phi: -\tfrac{1}{4}\eta_{\mu\nu} :\partial\phi\cdot\partial\phi: \qquad [5]$$

is the energy–momentum tensor. In [4], and also in a similar context subsequently, we define $\partial :\phi^2(0): = \partial_y :\phi^2(y):|_{y=0}$. The expansion [4] provides a point splitting definition of $T_{\mu\nu}$ and also demonstrates that many operators appearing in the OPE are expressible in terms of overall derivatives of lower-dimension operators. We may also note that without further input there is an ambiguity in the definition of $T_{\mu\nu}$ of the form

$$T_{\mu\nu} \sim T_{\mu\nu} + a(\partial_\mu\partial_\nu - \tfrac{1}{4}\eta_{\mu\nu}\partial^2) :\phi^2: \qquad [6]$$

In a conformal theory, however, we require $a = -1/6$.

## Interacting Theories

The OPE becomes an essential tool in the context of interacting quantum field theories. For renormalizable quantum field theories various results can be proved to all orders in the standard perturbative expansion and are naturally assumed to be properties of the complete theory. In interacting theories we may no longer use normal ordering to define composite operators which, in general, have anomalous dimensions. The coefficient functions appearing in the OPE also gain perturbative corrections but these are constrained by renormalization group (RG) Callan–Symanzik equations.

Again if we consider the simplest case of a massless scalar theory as above but now with a renormalized coupling constant $g$ the leading terms in the expansion of $\phi(x)\phi(0)$ are of the form (here we assume a $Z_2$ symmetry under $\phi \to -\phi$, otherwise the operator $\phi$ would be expected to appear in the OPE)

$$\phi(x)\phi(0) = \frac{C(g,\mu^2x^2)}{x^2} + D(g,\mu^2x^2)\phi^2(0) + \cdots \qquad [7]$$

where $\mu$ is an arbitrary renormalization scale. This arbitrariness is reflected in the RG equation

$$\left(\mu\frac{\partial}{\partial\mu} + \beta(g)\frac{\partial}{\partial g} + 2\gamma_\phi(g)\right) C(g,\mu^2x^2) = 0 \qquad [8]$$

At a fixed point $\beta(g_*) = 0$ this equation may be solved with an arbitrary choice of normalization to give $C(g_*,\mu^2x^2) = (\mu^2x^2)^{-\gamma_\phi(g_*)}$, which corresponds to the fields $\phi$ having a modified scale dimension $1 + \gamma_\phi(g_*)$. In a similar fashion the coefficient $D(g,\mu^2x^2)$ in [7] satisfies

$$\left(\mu\frac{\partial}{\partial\mu} + \beta(g)\frac{\partial}{\partial g} + 2\gamma_\phi(g) - \gamma_{\phi^2}(g)\right) \times D(g,\mu^2x^2) = 0 \qquad [9]$$

where it is necessary to introduce a new anomalous dimension function $\gamma_{\phi^2}(g)$ related to the composite

operator $\phi^2$. Although it is natural to label the operator as $\phi^2$ its definition in terms of the elementary field $\phi$ is essentially only as given in terms of the OPE [9]. At a fixed point again $D(g_*, \mu^2 x^2) = k(\mu^2 x^2)^{-\gamma_\phi(g_*)+(1/2)\gamma_{\phi^2}(g_*)}$, where the coefficient $k$ is determined by the scale of the three-point function $\langle\phi(x)\phi(y)\phi^2(0)\rangle$. In asymptotically free theories the RG equations show that at short distances the coefficient functions tend to those of free field theory but with calculable logarithmic corrections. More generally, for a set of operators $\{O_i\}$ the OPE has the form

$$O_i(x)O_j(0) \sim \frac{1}{(x^2)^p}\sum_k C_{ijk}(g,\mu^2x^2)O_k(0) \qquad [10]$$

where $p$ is determined by the free scale dimensions of the $O_i$ and

$$\begin{aligned}\left(\mu\frac{\partial}{\partial\mu}+\beta(g)\frac{\partial}{\partial g}\right)&C_{ijk}(g,\mu^2x^2)\\ =\sum_n\big(&\gamma_{kn}(g)C_{ijn}(g,\mu^2x^2)-\gamma_{in}(g)C_{njk}(g,\mu^2x^2)\\ &-\gamma_{jn}(g)C_{ink}(g,\mu^2x^2)\big)\end{aligned} \qquad [11]$$

with $\gamma_{in}(g)$ the anomalous dimension matrix arising from the mixing of composite operators.

An important aspect of the OPE is that the coefficient functions may be calculated perturbatively, essentially by applying the OPE in some suitable correlation function. Essentially the OPE provides a factorization between short-distance UV singularities and nonperturbative effects. In a Feynman graph the short distances in an operator product correspond to the large-momentum behavior and power-counting theorems allow a factorization up to calculable logarithmic corrections. A detailed analysis depends on the detailed technicalities of the proofs of renormalization to all orders of perturbation theory.

The coefficient functions in the OPE should be independent of any infrared or nonperturbative long-distance effects (such as confinement in QCD). However, the operators which appear in the OPE, such as $\phi^2$ above, may have nonzero expectation values which are absent to all orders in perturbation theory.

## Perturbative Example

The general considerations can be illustrated by considering a scalar field theory to lowest order in a perturbative expansion. We consider a four dimensional theory with a single scalar field and a potential $V(\phi) = \frac{1}{2}m^2\phi^2 + \frac{1}{24}g\phi^4$. Using dimensional regularization $m^2$, as well as $g$, is treated as a coupling with an associated $\beta$-function $\gamma_{\phi^2}(g)m^2$. With a mass term the operator $\phi^2$ mixes with the identity operator so that

$$\begin{aligned}(\mathcal{D}+\gamma_{\phi^2}(g))\langle\phi^2(0)\rangle &= -\gamma_{\phi^2 I}(g)m^2\\ \mathcal{D} = \mu\frac{\partial}{\partial\mu}+\beta(g)&\frac{\partial}{\partial g}+\gamma_{\phi^2}(g)m^2\frac{\partial}{\partial m^2}\end{aligned} \qquad [12]$$

where $\gamma_{\phi^2 I}$ reflects the mixing. At one loop order we have

$$\beta(g)=\frac{3g^2}{16\pi^2},\quad \gamma_{\phi^2}(g)=\frac{g}{16\pi^2},\quad \gamma_{\phi^2 I}(g)=\frac{1}{8\pi^2} \qquad [13]$$

and we may also set $\gamma_\phi(g)=0$. In this case in the operator product expansion (7) the coefficient $C$ also depends on $m^2x^2$ and the RG equations [8] and [9] are now modified to include the effects of mixing

$$\begin{aligned}&\mathcal{D}C(g,m^2x^2,\mu^2x^2) = m^2x^2\gamma_{\phi^2 I}(g)D(g,\mu^2x^2)\\ &(\mathcal{D}-\gamma_{\phi^2}(g))D(g,\mu^2x^2)=0\end{aligned} \qquad [14]$$

From lowest order perturbation theory with [13], and using [14] to include all orders in $g\ln\mu^2x^2$, we have in this approximation

$$\begin{aligned}C(g,&m^2x^2,\ \mu^2x^2)\\ &=\frac{1}{4\pi^2}-\frac{2m^2x^2}{g}\left(1+\frac{3g}{32\pi^2}\ln\mu^2x^2\right)^{2/3}\\ &\quad\times\left(\left(1+\frac{3g}{32\pi^2}\ln\mu^2x^2\right)^{-1/3}-1\right)\\ D(g,&\mu^2x^2)=\left(1+\frac{3g}{32\pi^2}\ln\mu^2x^2\right)^{1/3}\end{aligned} \qquad [15]$$

The operator product expansion then reproduces the small $x$ behavior of the two point function $\langle\phi(x)\phi(0)\rangle$ at one loop, expanding $C, D$ to first order in $g$, if we take

$$\langle\phi^2(0)\rangle = -\frac{m^2}{8\pi^2}\ln\frac{\mu}{m}+O(g) \qquad [16]$$

which is in accord with [12]. If $m^2<0$ the symmetry $\phi\leftrightarrow-\phi$ is broken and it is necessary to shift the field $\phi=\nu+f$, with $\nu^2=-6m^2/g$ and the field $f$ has a mass $m_f$ with $m_f^2=-2m^2$. The operator product expansion [7] with the same coefficient functions as in [15] remains valid. The two point function $\langle\phi(x)\phi(0)\rangle$, which includes a nonperturbative term $\nu^2$, is again reproduced for small $x$ at one loop now if

$$\langle\phi^2(0)\rangle = -\frac{6m^2}{g}-\frac{m^2}{2\pi^2}\ln\frac{\mu}{m_f}+O(g) \qquad [17]$$

but in this case it is necessary to expand $D(g,\mu^2x^2)$ to $O(g^2)$ as a consequence of the leading $1/g$ term in [17]. Note that both [16] and [17] contain the nonperturbative dependance on $\ln m$ and $\ln m_f$ which is present in the two point function.

## Conformal Field Theories

When the $\beta$-function vanishes and a quantum field theory enjoys conformal invariance the operator product expansion is a potentially convergent expansion. It is natural to restrict to conformal quasiprimary operators which do not mix with lower scale dimensions under conformal transformations. If we consider, for instance, two scalar operators $\phi$ with scale dimension $\Delta_\phi$ then the OPE has the generic form

$$\phi(x)\phi(0) = \frac{1}{x^{2\Delta_\phi}} + \sum_I C_{\phi\phi O^I} \frac{1}{(x^2)^{1/2}(2\Delta_\phi - \Delta^I + \ell)} \times C^{(\ell)}_{\Delta^I}(x,\partial)^{\mu_1,\ldots,\mu_\ell} O^I_{\mu_1,\ldots,\mu_\ell}(0) \qquad [18]$$

where there is a sum over quasiprimary operators $O^I_{\mu_1,\ldots,\mu_\ell}$ with scale dimension $\Delta^I$ and spin $\ell$, so they are symmetric traceless tensors of rank $\ell$. In the first term in [18] the coefficient is chosen to be 1 by a choice of normalization. The coefficients $C_{\phi\phi O^I}$, with a standard normalization for $O^I$, are then determined by the coefficients of the corresponding three-point functions involving $\phi\phi$ and $O^I$. In [18] $C^{(\ell)}_{\Delta^I}$ are differential operators which sum up the contributions of all derivatives or descendants of the quasiprimary operator $O^I$. They can be explicitly given in terms of an integral representation, for any spacetime dimension, where the scale is fixed by requiring for the leading term $C^{(\ell)}_{\Delta^I}(x,0)^{\mu_1,\ldots,\mu_\ell} = x^{\mu_1}\cdots x^{\mu_\ell}$ – traces. The spectrum of operators which appear is obviously a property of the particular conformal field theory.

## Ward Identities

If the theory has a symmetry with corresponding conserved currents then there are Ward identities which constrain the OPE of fields with the conserved current. For a current $J_{\mu a}$ then we have, in $d$ dimensions, the singular contribution in the OPE is given by

$$J_{\mu a}(x)O(0) \sim -\frac{1}{S_d}\frac{x_\mu}{(x^2)^{(1/2)d}} t_a O(0) \qquad [19]$$

where $t_a$ are a set of matrix generators corresponding to the symmetry acting on the fields $O$ and $S_d$ is the volume of the unit $(d-1)$-dimensional sphere, $S_4 = 2\pi^2$. For a conserved current there are no anomalous dimensions and the coefficient in [19], which depends on the normalization for the current $J_{\mu a}$, is chosen so that $[Q_a, O(0)] = -t_a O(0)$ with $Q_a$ the charge formed from $J_{\mu a}$. For the energy–momentum tensor the operator there is an analogous result. We consider the simpler case of a conformal theory when the energy–momentum tensor is both conserved and traceless and

$$T_{\mu\nu}(x)O(0) \sim A_{\mu\nu}(x)O(0) + B_{\mu\nu\lambda}(x)\partial^\lambda O(0) + \cdots \qquad [20]$$

where $A_{\mu\nu}(x) = \mathrm{O}(x^{-d})$ and $B_{\mu\nu\lambda}(x) = \mathrm{O}(x^{-d+1})$. As a distribution $A_{\mu\nu}(x)$ is ambiguous up to terms proportional to $\delta^d(x)$. If $\Delta$ is the scale dimension of $O$ and $s_{\mu\nu}$ are the Lorentz spin generators acting on $O$ the Ward identities then give

$$\begin{aligned}\partial^\mu A_{\mu\nu}(x) &= \left(\frac{\Delta}{d}\eta_{\nu\lambda} + C_{\nu\lambda} + \tfrac{1}{2}s_{\nu\lambda}\right)\partial^\lambda\delta^d(x)\\ A_{\mu\mu}(x) &= C_{\mu\mu}\delta^d(x)\\ \partial^\mu B_{\mu\nu\lambda}(x) &= -\eta_{\nu\lambda}\delta^d(x)\end{aligned} \qquad [21]$$

where $C_{\mu\nu}$ is a constant tensor reflecting the arbitrariness in $A_{\mu\nu}$, it is immaterial as far as Ward identities are concerned. We may choose

$$\frac{\Delta}{d}\eta_{\nu\lambda} + C_{\nu\lambda} = 0 \qquad [22]$$

(If desired, we might also take $A'_{\mu\nu}(x) = A_{\mu\nu}(x) + (1/2)s_{\mu\nu}\delta^d(x)$ in which case $\partial_\mu A'_{\mu\nu}(x) = 0, A'_{[\mu\nu]}(x) = (1/2)s_{\mu\nu}\delta^d(x)$ but such an antisymmetric piece seems unnatural). In general there is no unique form for $A_{\mu\nu}(x)$, as a consequence of the freedom of choice for $C_{\nu\lambda}$ in [21]. However, for a scalar field $O$ we must have, for $x \neq 0$,

$$\begin{aligned}A_{\mu\nu}(x) &= \frac{\Delta}{d-1}\frac{1}{S_d}\left(\eta_{\mu\nu} - d\frac{x_\mu x_\nu}{x^2}\right)\frac{1}{(x^2)^{(1/2)d}}\\ &= -\frac{\Delta}{(d-1)(d-2)}\frac{1}{S_d}\partial_\mu\partial_\nu\frac{1}{(x^2)^{(1/2)d-1}}\end{aligned} \qquad [23]$$

with the overall scale determined by [21].

For the operator product of the current $J_{\mu a}$ with itself there is an additional term proportional to the identity operator of the form

$$J_{\mu a}(x)J_{\nu b}(0) \sim C_J\delta_{ab}\left(\eta_{\mu\nu} - 2\frac{x_\mu x_\nu}{x^2}\right)\frac{1}{x^{2(d-1)}} \qquad [24]$$

where the coefficient $C_J$, which determines the scale of the two-point function for $J_{\mu a}$, is well defined since the normalization of the current is determined through the Ward identity. A similar result also holds for the operator product of the energy–momentum tensor with itself, with an overall coefficient $C_T$. In general, we may also write for the operator product of two scalar fields $O$:

$$O(x)O(0) \sim C_O\frac{1}{x^{2\Delta}} - \frac{C_O}{C_T S_d}\frac{d\Delta}{d-1}\frac{1}{(x^2)^{\Delta-(1/2)d+1}} \times x^\mu x^\nu T_{\mu\nu}(0) \qquad [25]$$

neglecting other contributions. The contribution of the energy–momentum tensor does not therefore introduce any new coefficient.

## Two Dimensions

In two dimensions the OPE plays an essential role in the discussion of conformal field theories. For a Euclidean metric it is natural to use complex variables $z$ and $\bar{z}$. The energy–momentum tensor in this case reduces to a chiral field $T(z)$ and its conjugate $\bar{T}(\bar{z})$. For the operator product with a chiral field $\phi(z)$ with scale dimension $\Delta$,

$$T(z)\phi(0) \sim \frac{\Delta}{z^2}\phi(0) + \frac{1}{z}\phi'(0) \qquad [26]$$

and, for the operator product of $T$ with itself,

$$T(z)T(0) \sim \frac{c}{2z^4} + \frac{2}{z^2}T(0) + \frac{1}{z}T'(0) \qquad [27]$$

Here $c$ is the Virasoro central charge, which plays a critical role in the discussion of two-dimensional conformal field theories, it is given by the two-point function which follows from [27], $\langle T(z)T(0)\rangle = (1/2)cz^{-4}$.

In simple rational conformal field theories the operators are organized into conformal blocks by the infinite-dimensional extended conformal symmetry in two dimensions. This allows the full spectrum of operators and their dimensions to be determined and in consequence complete results for the OPE to be found in many cases.

## Further Remarks

The OPE reflects the locality properties of quantum field theories and can be extended without difficulty to curved space backgrounds. For a product $\phi(x)\phi(0)$, the separation $x^2$ may be replaced by a biscalar at $x$ and 0 but it is necessary to include in the OPE contributions involving the background Riemann tensor as well as the operator fields present in flat space. There is also a generalization of the OPE for superfields on superspace.

At a fundamental level although the OPE can be derived to all orders in perturbation theory the contribution of nonperturbative effects such as instantons to the coefficients is not entirely clear. Issues of associativity have yet to be fully analyzed.

There are also important applications to the phemenonological analysis of QCD when assumptions about the OPE and saturation of sum rules can lead to results for the vacuum expectation value of gauge-invariant operators such as $F^{\mu\nu}F_{\mu\nu}$.

*See also:* Boundary Conformal Field Theory; Effective Field Theories; Quantum Chromodynamics; Renormalization: General Theory; Renormalization: Statistical Mechanics and Condensed Matter; Two-Dimensional Models.

## Further Reading

Cardy J (1987) Anisotropic corrections to correlation functions in finite size systems. *Nuclear Physics B* 290: 355.

Collins JC (1984) *Renormalization: An Introduction to Renormalization, The Renormalization Group and the Operator-Product Expansion.* Cambridge: Cambridge University Press.

David F (1984) On the ambiguity of composite operators, I.R. renormalons and the status of the operator product expansion. *Nuclear Physics B* 234: 237.

Di Francesco P, Mathieu P, and Sénéchal D (1997) *Conformal Field Theory.* New York: Springer.

Erdmenger J and Osborn H (1996) Conserved currents and the energy–momentum tensor in conformally invariant theories for general dimensions. *Nuclear Physics B* 483: 431.

Kadanoff LP (1969) Operator algebra and the determination of critical indices. *Physical Review Letters* 23: 1430.

Novikov VA, Shifman MA, Voinshtein AJ, and Zakharov VI (1985) Wilson's Operator product expansion: can it fail? *Nuclear Physics* B249: 445.

Wilson KG (1961) Non-Lagrangian models of current algebra. *Physical Review* 179: 1499.

Wilson KG (1971) Renormalization group and strong interactions. *Physical Review D* 3: 1818.

# Optical Caustics

**A Joets**, Université Paris-Sud, Orsay, France



## Introduction

Optical caustics are the bright forms created by the focalization, natural or artificial, of light (Figure 1). Special caustic points, called focuses, are produced by stigmatic optical systems in order to visualize objects. However, there are no special conditions for producing usual caustics. Every congruence of rays always generates a caustic, more or less intricate.

Caustics have been observed and described since a long time, tracing back to antiquity. The name itself was coined after the Greek root "kausticos" meaning burning and expressing that a high energy density is produced by ray focalization at a caustic point. Conceptually, they appeared in the literature as "evolutes," "envelopes," "centers of curvature," "focals," etc. However, these different approaches, often too restricted, were unable to clarify the

**Figure 1** Optical caustics may be produced by reflection (on window glasses) or by refraction (through the wavy surface of a swimming pool). Here the light source, the Sun, has some angular extension and the caustic appears somewhat blurred.

general properties of caustics, for instance, their classification in generic types. This difficult question was solved only recently in the framework of the singularity theory which appeared in the second half of the twentieth century (Whitney 1955, Thom 1956). Caustics are now understood as physical realizations of Lagrangian singularities, and they are often called optical singularities or optical catastrophes.

The aim of this introductory article is to show in which sense caustics can be understood as singularities, and to present their main properties.

## The Physical Phenomenon

Caustics are usually observed by interposing a screen on the ray trajectories and their trace in the screen forms a set of bright curves called "fold" ($A_2$). Across the fold, the number of rays passing through a given point jumps by $\pm 2$. Two fold curves may join at some point forming there a tip called cusp ($A_3$). A simple example is provided by the nephroid that one sees in a cup of coffee when the light is reflected off the cylindrical sides. In the three-dimensional (3D) space, the folds form surfaces and the cusps form curves (Figure 2). For particular positions of the screen, three other types of caustics may be observed: the swallowtail ($A_4$), the meeting point of two cusp lines; the elliptic umbilic ($D_4^-$), the meeting point of three cusp lines; and the hyperbolic umbilic ($D_4^+$) where a cusp line tangentially meets a fold surface (Figure 2). These five caustic types are generic in the sense that any other type of caustic point is unstable and decomposes into these generic caustic points under small perturbations. The perfect focus is an example of a nongeneric caustic point, obtained by imposing a special symmetry. The natural focusing of light, as in gravitational optics, produces only generic caustics. A caustic point is then a generalized focus. The caustic surface is a complex surface in the 3D physical space, generally self-intersecting and possessing singular lines $A_3$ ending at singular points $A_4, D_4^-$, or $D_4^+$.

At the scale of the wavelength of the light, the caustics have a more complex structure. Instead of well-defined surfaces, lines and points, one observes a system of interference fringes concentrated in the vicinity of the geometrical caustic. Each type of caustic point has its own diffraction pattern (also called diffraction catastrophe) (Figure 3). These interference systems are easily produced, for instance, by focusing a coherent laser beam by a

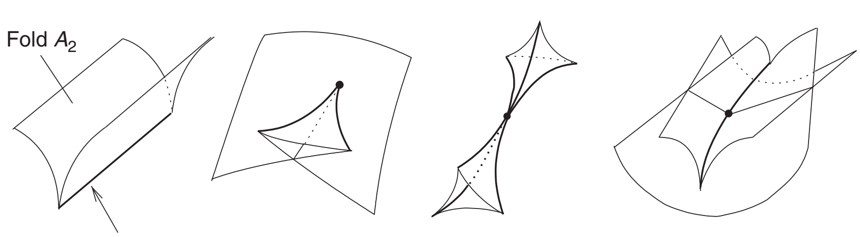



**Figure 2** The five generic types of caustics of the 3D space.

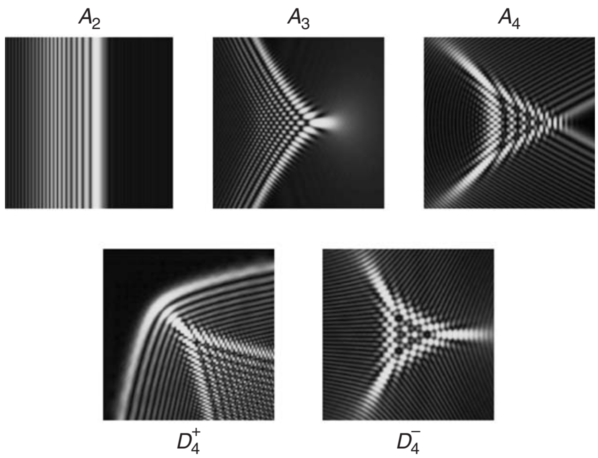



**Figure 3** Interference fringes produced by the five generic caustics of the 3D space (numerical simulation).

corrugated glass or by a water droplet. An important feature is revealed by Gouy's experiment, in which bright and dark fringes are inverted when the rays are forced to pass through a focus (Guillemin and Sternberg 1977). The experiment shows that the wave undergoes a phase shift of $\pi/2$ when the associated ray passes through a caustic point.

So, caustics are fundamental objects of both the geometrical optics and the wave optics.

## Modeling Caustics

Because of the presence of a caustic, a congruence of rays generally presents intersecting rays. At the points of intersection, the coordinates $q_1, q_2, q_3$ of the physical space $R^3$ are unable to distinguish the various intersecting rays and they do not constitute a convenient system of coordinates. It is then interesting to construct an abstract space in which the rays are represented by nonintersecting curves. The initial congruence is recovered by projecting the abstract space into the physical one. All the models use this type of construction in which the properties of the caustics are deduced from those of the projection.

### Caustics as Envelopes of Rays

In this geometrical modeling, each ray is labeled by two parameters $r_1, r_2$, for instance, the coordinates on the initial wave front $W$. A third coordinate $r_3$ specifies the points along the ray, for instance, by assigning their distance to $W$. Taken together, these three coordinates represent the congruence of rays, and define a 3D space, the source space $M = \{r_1, r_2, r_3\}$. By construction, the rays in $M$ do not intersect. The coordinates $(q_1, q_2, q_3)$ of the current point $P \in R^3$ along each ray depend differentiably on the coordinates $(r_1, r_2, r_3)$ and define a "projection" $f:(r_1, r_2, r_3) \mapsto (q_1, q_2, q_3)$ from the source space $M$ into the physical space $R^3$.

The caustic points correspond to the envelope of the rays. At a caustic point $P$, the energy density flowing along the rays becomes infinite, since the small volume delimited by neighboring rays is shrunk into a small surface at $P$. This behavior may be simply expressed with the help of the projection $f$: the rank rk of the derivative $Df$ is equal to 2 at the point representing $P$ in $M$. This motivates the following definition. Given a map $f: M \to N$, a point $x \in M$ is said to be critical (or singular) if the rank of the derivative $Df$ is less than the maximal possible value min(dim $M$, dim $N$). Here, $\dim M = \dim N = 3$, and a critical point is a point where rk $< 3$. The set $\Sigma \subset M$ of the critical points is called the singular set. The caustic $C$ is the image of the singular set: $C = f(\Sigma)$. One also says that the caustic points are the critical values of $f$.

In practice, the derivative $Df$ is expressed by the Jacobian matrix $J = \partial(q_1, q_2, q_3)/\partial(r_1, r_2, r_3)$ and the singular set $\Sigma$ is defined by solving the equation

$$\det(J) = 0 \qquad [1]$$

If this equation permits one to express explicitly one coordinate, say $r_3$, as a function of the other two, the caustic surface $C$ is found in parametric form: $q_1 = q_1(r_1, r_2, r_3(r_1, r_2))$, etc. For a homogeneous medium, equation [1] is of second degree in $r_3$ and the caustic is composed of two sheets which meet at the umbilic points $D_4$.

Equation [1] gives all caustic points independently of their nature, that is, it does not distinguish between $A_2$, $A_3$, $A_4$, $D_4^-$, and $D_4^+$. A refinement allows one to recognize different types of caustic points. One defines the Thom–Boardman class $\Sigma^i$ as the points in $M$ where $Df$ has a kernel of dimension $i$. Then one defines inductively the class $\Sigma^{i,\ldots,j,k}$ as the class $\Sigma^k$ of the restriction of $f$ to $\Sigma^{i,\ldots,j}$. Thus, $\Sigma^0$ represents the regular points (noncaustic points), $\Sigma^{1,0}$ the fold points $A_2$, $\Sigma^{1,1,0}$ the cusp points $A_3$, $\Sigma^{1,1,1,0}$ the swallow-tail points $A_4$, and $\Sigma^{2,0}$ the umbilics $D_4$ (hyperbolic or elliptic). Altogether, the classes $\Sigma^I$, $I \neq 0$, form the singular set $\Sigma$.

The Thom–Boardman classes constitute a simple and powerful tool for computing the structure of a caustic. Each class is obtained by canceling some functional determinants associated with the map $f$ or with its restriction to some class. However, the method presents the weakness of ignoring the special nature of a set of rays: its Lagrangian character. As a consequence, it is unable, for instance, to distinguish between $D_4^-$ and $D_4^+$.

### Caustics as Lagrangian Singularities

As for mechanics, the natural framework for geometrical optics is a phase space: the cotangent space $T^*R^3 = \{p_i, q_i\}$ of the configuration space $R^3 = \{q_i\}$. The phase space is characterized by its symplectic structure, that is, the differential 2-form $\omega = \sum_i \mathrm{d}p_i \wedge \mathrm{d}q_i$, which is nondegenerate and closed ($\mathrm{d}\omega = 0$).

A set of rays in the phase space is defined by specifying the wave vector (or momentum) $\boldsymbol{p}$ at each point $\boldsymbol{q}$ of the congruence. In the simple case where only one ray passes through each point, one has $\boldsymbol{p} = \nabla S$, where $S$ is the optical length $\int n\,\mathrm{d}s$ and $n$ the refractive index. In other words, $\boldsymbol{p}$ is the differential of the optical length. The wave vector $\boldsymbol{p}$ is tangent to the ray and orthogonal to the (geometrical) wave front $S = \text{const}$. The eikonal equation shows that its modulus is $n$. As a direct consequence of the relation $\boldsymbol{p} = \nabla S$, the symplectic form annihilates identically for these $\boldsymbol{p}$. However, in general, because of the presence of the caustics, one must not expect to have $\boldsymbol{p} = \nabla S$ for some function $S$. Nevertheless, it is possible to keep the more general property to annihilate $\omega$. This motivates the definition of a Lagrangian submanifold: a submanifold $L \subset T^*R^3$ of dimension 3 (that is, half of the dimension of the phase space) on which the symplectic form vanishes: $\omega|_L = 0$. Every congruence of rays is described by a Lagrangian submanifold. The Lagrangian submanifold plays the same role as the source space in the preceding section. The role of the projection $f$ is played by the natural projection $\pi$ from the phase space into the configuration space $\pi(\boldsymbol{p}, \boldsymbol{q}) = \boldsymbol{q}$, or more precisely to its restriction to $L$: $f = \pi|_L$. It is called a Lagrangian map (or Lagrangian projection) and it is again a map between two spaces of the same dimension (here 3). When $L$ is given by an embedding $\iota : L \to T^*R^3$, one has $f = \pi \circ \iota$. A caustic is then defined as the set of critical values of a Lagrangian map.

There exist two remarkable results showing that a Lagrangian submanifold may be described in terms of functions or of families of functions. As a consequence, caustics are not directly related to the singularities of maps but, more particularly, to the singularities of functions.

**",1,5,3,0,0pc,0pc,0pc,0pc>Generating function of a Lagrangian submanifold** The 3D Lagrangian submanifold $L \subset \{p_i, q_i\}$ is locally defined by three coordinates $p_\alpha (\alpha \in A)$ and $q_\beta (\beta \in B)$ depending on the three other ones $p_\beta$ and $q_\alpha$: $p_\alpha = p_\alpha(q_\alpha, p_\beta)$, $q_\beta = q_\beta(q_\alpha, p_\beta)$. One can show that this may be done in such a way that each conjugate pair $(q_i, p_i)$ gives exactly one independent variable and one dependent variable. Formally: $A \cup B = \{1, 2, 3\}$, $A \cap B = \emptyset$.

In fact, introducing the function $S(q_\alpha, p_\beta) = \int \langle p, \mathrm{d}q \rangle - \langle q_\beta, p_\beta \rangle (\langle , \rangle$ denotes the scalar product), the local equation for $L$ takes a more simple form:

$$q_\beta = -\frac{\partial S}{\partial p_\beta}, \qquad p_\alpha = \frac{\partial S}{\partial q_\alpha} \qquad [2]$$

The function $S$ is well defined, since, by the definition of a Lagrangian submanifold $\int \langle p, \mathrm{d}q \rangle$ is locally path independent: it depends only on its end points. $S$ is called a (local) generating function.

Formula [2] generalizes $\boldsymbol{p} = \nabla S$, to which it reduces when $B = \emptyset$, that is, for nonintersecting rays.

**",1,5,3,0,0pc,0pc,0pc,0pc>Generating family and optical catastrophes** Formula [2] may be rewritten in an interesting way. Taking the $|B|$ variables $p_\beta$ as internal parameters $x$ and $q = (q_\alpha, q_\beta)$ as external parameters, we construct a function $F$ of $x$ parametrized by $q$: $F(x, q) = S(q_\alpha, x) + \langle q_\beta, x \rangle$. Now the Lagrangian submanifold $L$ is defined by

$$L = \left\{ (q, p) : \exists x : \frac{\partial F}{\partial x} = 0,\ p = \frac{\partial F}{\partial q} \right\}$$

$F$ is called the generating family. The first equation $\partial F / \partial x = 0$ determines the rays passing through the fixed external parameter $\boldsymbol{q} \in R^3$. The second one distinguishes these rays according to their wave vector $\boldsymbol{p}$. Each ray corresponds to a critical point (i.e., an extremum) of $F$ considered as a function of $x$. At a caustic point, two infinitely close rays are converging and $F$ then presents a degenerate critical point. So the generating-family technique links the caustics to the theory of singularities of functions depending on some parameters, that is, to the catastrophe theory (Thom 1969). Caustics are also called optical catastrophes.

The generating families are not uniquely defined, even locally. In optics, one may always take for $F$ the equivalent family "optical length" $d$, considered as a function defined on the initial wave front $W$ (this is discussed in the following).

### Caustics as the Locus of Wave Front Singularities

There exists a remarkable duality linking rays and wave fronts. As a consequence, the caustic points (i.e., Lagrangian singularities) are related to singularities of wave fronts (i.e., Legendrian singularities). A typical wave front $W$ may possess only two types of singularities: cuspidal curves and swallow-tail points. During the motion of $W$, governed by the eikonal equation, the cuspidal curves generate surfaces, and

swallow tails generate curves. These surfaces are exactly the fold surfaces of the caustic $C$ and the curves are the cusp lines of $C$. The point singularities of the caustic, that is, the swallow tails and the umbilics, correspond to bifurcations of the instantaneous wave front, at certain moments of its motion.

### Caustics as Short Wave Asymptotic

The fine observation of the optical caustics shows that they never appear as the well-defined surfaces given by the geometrical optics, but rather as diffraction patterns concentrated around these surfaces. So wave optics is the natural framework to account for this fundamental feature. One exploits the fact that the wave number $k = 2\pi/\lambda$ ($\lambda$: wavelength of the light) is a large parameter. This short-wave approximation permits the use of powerful expansion techniques and clarifies the relation with the geometrical optics viewpoint, formally obtained for $k$ tending to infinity.

**The stationary phase** In the most simple model, the Huygens–Fresnel principle, the amplitude $U(P)$ of the optical field may be evaluated by adding the secondary disturbances emitted from the points $Q$ of some initial wave front $W$:

$$U(P) = c \iint_W \frac{\mathrm{e}^{ikd}}{d} G \, \mathrm{d}s \qquad [3]$$

where $d$ is the distance $QP$. $G$ is the inclination factor, a smooth function defined on $W$ and $c$ some prefactor. For simplicity, $G$ and $n$ (the refractive index) are assumed to be constant. Defining $a = cG/d$, formula [3] appears as an integral of the form $\int a(y)\mathrm{e}^{ik\phi(y)}\mathrm{d}y$. This type of integral may be evaluated for large $k$ by the method of stationary phase. The principal contributions are due to points where the phase $\phi$ is stationary: $\nabla\phi = 0$. For wave optics, $\phi$ is the length $PQ$, considered as a function of $Q$ and parametrized by $P$. The stationary condition means that $PQ$ is normal to $W$, that is, it represents a ray of geometrical optics. The function $PQ$ is a generating family in the sense of the discussion earlier.

If no stationary points exist, that is, if $P$ is in the shadow, the integral is $O(k^{-N})$ for any $N$. Otherwise, and if the critical points are not degenerate, the phase stationary method gives (Guillemin and Sternberg 1977):

$$U(P) = \frac{2\pi}{k} \sum_{\text{rays } PQ} \mathrm{e}^{(1-\sharp)\pi\mathrm{i}/2} \times \frac{a(Q)\mathrm{e}^{\mathrm{i}kd}}{|(1-\mu_1 d)(1-\mu_2 d)|^{1/2}} + O(k^{-2}) \qquad [4]$$

where $\mu_1^{-1}$ and $\mu_2^{-1}$ are the two principal radii of curvature at $Q \in W$, and $\sharp$ the number of caustic points (also called focal points) along the ray $PQ$.

In the stationary-phase approach, the caustic $C$, locus of centers of curvature of $W$, appears as an obstacle in constructing asymptotics, since formula [4] diverges when $d\mu_i \to 1$, that is, when $P$ tends to $C$. It is, nevertheless, remarkable that $C$ also appears explicitly when [4] is valid, via the $\mu_i$'s and $\sharp$. In particular, the term $\mathrm{e}^{-\sharp\mathrm{i}\pi/2}$, applied in the case of a focus ($\sharp = 2$), accounts for the phase shift of $\pi$ observed in Gouy's experiment.

**Asymptotics on caustics** Uniform asymptotic formulas, valid also on the caustic, need a more complex theoretical framework, for instance, Maslov's theory, presented here in a necessarily simplified version (see Maslov and Fedoriuk (1981) for more detail).

The starting point is the equation of wave optics, that is, the Helmholtz equation

$$(\Delta + k^2 n^2)U = 0 \qquad [5]$$

where the refractive index $n$ generally varies from point to point. For $k \to \infty$, one looks for an asymptotic solution in the (tentatively) form:

$$U(P) = \mathrm{e}^{\mathrm{i}kS(q_1,q_2,q_3)} \sum_{j=0}^{\infty} (\mathrm{i}k)^{-j} \varphi_j(q_1,q_2,q_3) \qquad [6]$$

Inserting this form in eqn [5] one obtains the eikonal equation (or characteristic equation) for the phase $S$:

$$(\nabla S)^2 = n^2$$

and an infinite series of equations for the amplitudes $\varphi_j$, called the transport equations. One knows that the Cauchy problem for the eikonal equation may be reduced to the integration of the corresponding Cauchy problem for the Hamilton system (or bicharacteristic system):

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{\partial H}{\partial p} = 2p, \qquad \frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial H}{\partial q} = \nabla n^2$$

where $H = \langle p, p \rangle - n^2$. Its solutions, the bicharacteristics $q(t,\xi)$, $p(t,\xi)$ are parametrized by the "time" $t$ and the 2D parameter $\xi$ parametrizing the points on the initial wave front $W$. The bicharacteristics form a 3D Lagrangian submanifold $L$ in the phase space $\{p_i, q_i\}$ and one recovers the preceding situation. Assuming $L$ to be simply connected, one defines a global phase function $S$ on $L$ by formula $S(t,\xi) = \int \langle p, \mathrm{d}q \rangle$.

In a domain $\Omega_j \subset L$ not containing the singular set and in which the coordinates $t, \xi$ are in a one-to-one correspondence with the physical coordinates, $S$

becomes a function of $q_i$. Using the transport equation, one finds the leading term of the asymptotic solution (with accuracy to $k^{-1}$) in the following form:

$$U(P) = (K(\Omega_j)\varphi)(q) = \sqrt{\left|\frac{d\sigma}{dq_i}\right|}e^{ikS(q_1,q_2,q_3)}\varphi(q_1,q_2,q_3) \qquad [7]$$

where $d\sigma$ and $dq_i$, respectively, represent the measures on the Lagrangian submanifold and on the physical space. The amplitude $\varphi$ depends on the initial conditions. Formula [7] defines a precanonical operator $K(\Omega_j)$. It has the same form as [4], with the same drawback to diverge near the singular set $\Sigma$, where $dq_i = 0$.

In a domain $\Omega_j$ containing the singular set, $L$ is locally parametrized by mixed coordinates $q_\alpha, p_\beta$. The basic idea is then, roughly speaking, to carry out a Fourier transform $F_k$ with respect to these $p_\beta$ (in fact, a variant of the usual Fourier transform, in which the parameter $k$ appears in the prefactor and in the phase term). This leads one to consider, instead of $\mathcal{L} = \Delta + k^2n^2$, the operator $\hat{\mathcal{L}} = F_k\mathcal{L}F_k^{-1}$, and instead of $U$, the unknown function $V = F_kU$. In this Fourier space, $V$ may be found in the same way as $U$ was found in the preceding case, with $S$ replaced here by the local generating function $S_j(q_\alpha, p_\beta) = S - \langle q_\beta, p_\beta\rangle$. Coming back to the real space by $F_k^{-1}$, one obtains (with the same accuracy):

$$U(P) = (K(\Omega_j)\varphi)(q) = F_k^{-1}\left[\sqrt{\left|\frac{d\sigma}{dp_\beta dq_\alpha}\right|}e^{ikS_j(q_\alpha,p_\beta)}\varphi(q_\alpha,p_\beta)\right] \qquad [8]$$

There is no divergence in this local solution. So local short-wave asymptotics may be found everywhere, even on the caustic where they have a more complex form than the form [6] or [7].

**Global asymptotics and Maslov's index** The global asymptotic solution is obtained by formally gluing the local solutions by a partition of unity $\Sigma e_j = 1$ subordinate to a covering $\{\Omega_j\}$ of $L$. However there is a difficulty. The representations of the same precanonical operator in different local coordinates $q_\alpha$, $p_\beta$, even not containing the singular set, agree only up to a constant multiplier $e^{i\pi m/2}$, where the integer $m$ is the number of negative eigenvalues of some matrix. One is led to multiply every precanonical operator by a convenient phase factor $e^{-i\pi\gamma/2}$, where $\gamma \in Z_4$ is called Maslov's index. The coherency of the phase factor in different domains is realized by using the important property of $\Sigma$ to be co-oriented. Thus, $\gamma$ counts the number of passages of an oriented path on $L$ from the negative side of $\Sigma$ to its positive side, minus the number of passages in the opposite sense. Maslov's index is locally constant and jumps by $\pm 1$ only across the singular set $\Sigma$. The global canonical operator is now formally defined as $K = \Sigma_j e^{-i\pi\gamma_j/2}K(\Omega_j)e_j$.

Finally, the canonical operator $K$ is well defined only if it is independent of the $\{\Omega_j\}$ and $e_j$ used for its definition. This possibility is expressed (in the case of a simply connected $L$) by the following property, intrinsically attached to $L$: the Maslov index cancels on every closed loop. So the only obstruction for global asymptotics is the nontriviality of the characteristic class defined by Maslov's index and not the caustic.

The central object of the caustic modeling is then the projection of the submanifold representing the rays ($M$ or $L$) into the physical space. The possibility to reduce this projection to some normal form is the key result for the local classification of caustics.

## Local Classification of Caustics

### Equivalence, Stability, and Genericity

In order to distinguish different types of singularities, one has to define an equivalence relation. Two Lagrangian maps $f_i : T^*M_i \supset L_i \to M_i$ $(i = 1, 2)$, are said to be Lagrange equivalent if there is a diffeomorphism $h : T^*M_1 \to T^*M_2$ preserving both the symplectic and the fiber structures, and sending $L_1$ to $L_2$. In fact, only the local problem of classification makes sense, and one considers, instead of Lagrangian maps, germs of Lagrangian maps. A map germ is a map locally defined, that is, defined in an infinitely small neighborhood around a point (depending on the germ). The notion of Lagrange equivalence is extended to the germs. A Lagrangian singularity is then the Lagrange equivalence class of a germ at a critical point. Each equivalence class represents a type of Lagrangian singularity, that is, a type of caustic point.

The example of the perfect focus point shows that there exist singularities which are totally unstable. In this sense, they correspond to idealized situations not physically realizable, and they have to be disregarded. Conversely, stable singularities resist under the action of small perturbations. They correspond to Lagrangian germs for which all neighboring germs are Lagrange equivalent (not necessarily at the same point, but near the point considered).

Now the important question is: do the stable germs represent the generality? In the best case, stable germs form a dense open set. This means that every germ may be approximated by stable germs. In this case, one says that the stable germs are generic.

Stability and genericity are disctinct notions. It turns out that they coincide for low values of the dimension $n$ of the “physical space” ($n < 6$), but they may disagree at higher dimensions.

### Classification of Stable Caustics

The fundamental result of the theory is the local classification of Lagrangian singularities (Arnol’d 1972). With the help of the generating families, the study of Lagrangian singularities is reduced to the study of singularities of families of functions. More precisely, at a singular point, every stable Lagragian map is equivalent to one of the following maps, given by their generating function $S$ and by their generating family $F$:

$$
\begin{aligned}
A_2: \quad & S = p_1^3 \\
& F = x^3 + q_1 x \\
A_3: \quad & S = \pm p_1^4 + q_2 p_1^2 \\
& F = \pm x^4 + q_1 x^2 + q_2 x \\
A_4: \quad & S = p_1^5 + q_2 p_1^3 + q_3 p_1^2 \\
& F = x^5 + q_1 x^3 + q_2 x^2 + q_3 x \\
D_4^{\pm}: \quad & S = p_1^3 \pm p_1 p_2^2 + q_3 p_1^2 \\
& F = x_1^2 x_2 \pm x_2^3 + q_1 x_2^2 + q_2 x_2 + q_3 x_1
\end{aligned}
$$

These polynomial functions are called normal forms. The stable singularities are generic. In other words, every other type of singularity is destroyed by infinitely small perturbations and gives a set of singularities belonging to the list. The five generic caustics have been observed and experimentally studied in detail (Berry and Upstill 1980, Nye 1999).

By inserting the normal forms $S$ in a short-wave asymptotic, one obtains the diffraction patterns associated with the five caustic types (**Figure 3**). They generalize the Airy function which corresponds to the fold singularity.

The normal forms describe at once the geometry of the caustics and the interference systems around them.

### Codimension, Corank, Multiplicity, and Index

Lagrangian singularities are also characterized by some numbers. They have a codimension $c$ equal to the difference between the dimension of the physical space and their dimension: $c(A_2) = 1$, $c(A_3) = 2$, $c(A_4) = c(D_4^{\pm}) = 3$. They have a corank ck, equal to the difference between the dimension of the space and the rank of the Lagrangian map: $\mathrm{ck}(A_2) = \mathrm{ck}(A_3) = \mathrm{ck}(A_4) = 1$, $\mathrm{ck}(D_4^{\pm}) = 2$. The corank is the number of internal parameters of the generating family $F$. They also have a multiplicity $\mu$, which is the number of nondegenerate critical points of $F$, that is, the number of rays coinciding at the singularity. In the 3D space, one has $\mu = c + 1$: $\mu(A_2) = 2$, $\mu(A_3) = 3$, $\mu(A_4) = \mu(D_4^{\pm}) = 4$.

Short-wave asymptotics near the caustic present remarkable scaling properties (Berry and Upstill 1980). In particular, the amplitude $|U(P)|$ increases like $k^{\delta}$ as $k \to \infty$. The number $\delta$ depends only on the type of the singularity and it is called the singularity index. The more “degenerate” the singularities, the larger the index, and then the brighter the caustic point: $\delta(A_2) = 1/6 < \delta(A_3) = 1/4 < \delta(A_4) = 3/10 < \delta(D_4^{\pm}) = 1/3$.

## Global Organization of Caustics

The global properties of caustics are less understood than the local ones. There is, nevertheless, an interesting result concerning specifically the caustics in the 3D space (Chekanov 1986). Given a Lagrangian map $f: L \to R^3$, the Euler characteristic $\chi(\Sigma)$ of the singular set $\Sigma \subset L$ and the number $\sharp D_4(-1/2)$ of umbilics of index –1/2 are related by the formula

$$\chi(\Sigma) + 2\sharp D_4(-1/2) = 0 \qquad [9]$$

At an umbilic point $T$, $\Sigma$ is locally a cone with vertex at $T$. The index is defined according to the relative positions of the following elements: the 2D plane $\Pi = \ker f$, the cusp lines $A_3 \subset \Sigma$ passing through $T$, and the characteristic line $l$ which represents the ray at $T$. If $l$ and $A_3$ are separated by $\Pi$, the index is equal to $+1/2$, and to $-1/2$ in the other case. The index of an elliptic umbilic is always equal to $-1/2$.

The validity of Chekanov’s formula [9] requires that $L$ lies on a hypersurface $E$ of the phase space, convex with respect to the wave vectors. The characteristics are the orthocomplements of $E$. In this framework, the singularities are called optical singularities, because such an $E$ is always defined in geometrical optics by the eikonal equation. All Lagrangian singularities can be realized as optical singularities. Chekanov’s formula has been experimentally checked (Joets and Ribotta 1996).

The Chekanov relation has an important consequence on the caustic bifurcations (also called metamorphoses or perestroikas), that is, the generic transformations modifying the topology of a caustic depending on one parameter. Among the 11 possible caustic bifurcations, considered as bifurcations of general Lagrangian singularities, four of them cannot be realized as bifurcations of optical Lagrangian singularities. So Chekanov’s relation reduces the number of optical metamorphoses to seven.

## Extensions

### Caustics in Spaces of Higher Dimension

The local classification of Lagrangian singularities has been extended in spaces of higher dimension. For $n=4$, in addition to the preceding ones, two new singularities appear: the butterfly $A_5$ and the parabolic umbilic $D_5$. For $n=5$, in addition to $A_6$ and $D_6^{\pm}$, one has a new type of umbilic: $E_6$. However, in higher dimensions, the classification becomes more complex. In addition to stable singularities, like those of the series $A_i$, $D_i$, $E_i$, one encounters unstable generic singularities which depend on arbitrary parameters (moduli). Despite this difficulty, there exists a classification of generic Lagrangian singularities up to the dimension $n=10$.

The Maslov index has been extended in spaces of higher dimension and has led to the discovery of invariants associated with particular types of singularities (Vassilyev 1988). These invariants control the number of some types of singularities. For instance, in dimension $n=4$, the number of $A_5$ (taking account of sign) is equal to zero.

### Symmetrical Caustics

Another extension consists in imposing some constraint, for instance, a symmetry (Janeczko and Roberts 1993). Symmetrical caustics are not merely the symmetrized usual caustics. Many of them result from the stabilization of unstable singularities of higher codimension by the symmetry. For example, in the 3D space, the butterfly $A_5$ is unstable, but the symmetrical butterfly is a generic singularity in the class of Lagrangian singularities having the mirror symmetry.

### Nonoptical Caustics

Caustics, as locus of focalization, are not restricted to the usual optics. They are also observed in electronic optics or in gravitational optics and the preceding results apply to these waves. They also appear in nonelectromagnetic waves, for instance, acoustic waves, seismic waves, etc. Propagation always generates caustics.

Optical caustics are now understood as Lagrangian singularities and, as singularities, their interest is not restricted to optics. They became indispensable for understanding other domains of mathematical physics, for instance, the variational calculus, the classical mechanics, the Hamilton–Jacobi equations, the control theory, the field theory, etc.

*See also:* Billiards in Bounded Convex Domains; Normal Forms and Semiclassical Approximation; Stationary Phase Approximation; Singularity and Bifurcation Theory.

## Further Reading

Arnol'd VI (1972) Normal forms for functions near degenerate critical points, the Weyl groups $A_k, D_k, E_k$ and Lagrangian singularities. *Functional Analysis and its Applications* 6: 254–272.

Arnol'd VI (1983) Singularities of systems of rays. *Uspekhi Matematicheskykh Nauk* 38: 77–147.

Arnol'd VI (1990) *Singularities of Caustics and Wave Fronts,* Math. Appl. (Soviet Series), vol. 62. Dordrecht: Kluwer Academic.

Arnol'd VI, Gusein-Zade SM, and Varchenko AN (1985) *Singularities of Differentiable Maps*, Volume I, Monographs in Mathematics, vol. 82. Boston: Birkhäuser.

Bennequin D (1986) *Caustique Mystique*, Siéminaire Bourbaki, 37e année, no. 634, S.M.F., *Astérisque* 133–134, pp. 19–56.

Berry MV and Upstill C (1980) Catastrophe optics: morphologies of caustics and their diffraction patterns. In: Wolf E (ed.) *Progress in Optics XVIII*, pp. 257–346. Amsterdam: North-Holland Publishing Company.

Chekanov Yu V (1986) Caustics in geometrical optics. *Functional Analysis and its Applications* 6: 223–226.

Guillemin V and Sternberg S (1977) *Geometric Asymptotics*, Mathematical Surveys and Monographs, vol. 14. Providence: American Mathematical Society.

Janeczko S and Roberts M (1993) Classification of symmetric caustics II: caustic equivalence. *Journal of the London Mathematical Society* 48: 178–192.

Joets A and Ribotta R (1995) Structure of caustics studied using the global theory of singularities. *Europhysics Letters* 29: 593–598.

Joets A and Ribotta R (1996) Experimental determination of a topological invariant in a pattern of optical singularities. *Physical Review Letters* 77: 1755–1758.

Kravtsov Yu A and Orlov Yu I (1993) *Caustics, Catastrophes and Wave Fields*, Wave Phenomena, vol. 15. Berlin: Springer.

Maslov VP and Fedoriuk MV (1981) Semi-classical approximation in quantum mechanics. In: *Mathematical Physics and Applied Mathematics*, vol. 7. Dordrecht: D Reidel Publishing Company.

Nye JF (1999) *Natural Focusing and Fine Structure of Light*. Bristol: Institute of Physics Publishing.

Thom R (1956) Les singularités des applications différentiables. *Annales de l'Institut Fourier* 6: 43–87.

Thom R (1969) Topological models in biology. *Topology* 8: 313–335.

Vassilyev VA (1988) Lagrange and Legendre characteristic classes. *Advanced Studies in Contemporary Mathematics*, vol. 3. New York: Gordon and Breach Science Publishers.

Whitney H (1955) On singularities of mappings of Euclidean spaces. I. Mappings of the plane into the plane. *Annals of Mathematics* 62: 374–410.

# Optimal Cloning of Quantum States

**M Keyl**, Università di Pavia, Pavia, Italy



## Introduction

According to the well-known "no-cloning theorem" (Wootters and Zurek 1982) perfect copying of quantum information is impossible, that is, there is no machine which takes a quantum system in an unknown state as input and produces two systems of the same kind, such that none of them is distinguishable from the input by a statistical experiment. In this qualitative form, however, the theorem is not very useful, because in the presence of noise classical information cannot be copied perfectly as well. Therefore, the crucial point is that even under ideal conditions the errors produced in the clones cannot be made arbitrarily small. The best we can hope for is to find an optimal cloning device which makes these errors as small as possible.

More generally, we can consider cloning devices, which take as input a certain number, $N$, of identically prepared systems, and produce a larger number, $M$, of systems as output. Again, the cloning task is to make the output state resemble as much as possible a state of $M$ systems all prepared in the same state as the inputs. This variant of the problem is of interest as a "quantum amplifier." It also has a better chance of reasonable success than a cloning device operating on single-input systems: in the limit of many-input systems, the device can make a good statistical estimate of the input density matrix and hence produce arbitrarily good clones.

### Figures of Merit

To get a precise mathematical description of the problem, let us consider a one-particle Hilbert space $\mathcal{H}$ (which is assumed to be finite dimensional, $\mathcal{H} = \mathbb{C}^d$, if nothing else is explicitly stated) and the algebras $\mathcal{B}(\mathcal{H}^{\otimes N}), \mathcal{B}(\mathcal{H}^{\otimes M})$ of (bounded) operators on the $N$-fold, respectively $M$-fold, tensor product of $\mathcal{H}$. A quantum operation which takes $N$ particles as input and produces $M$ output particles is then described, in the Heisenberg picture, by a completely positive, unital map (a completely positive, unital and normal map if $\mathcal{H}$ is infinite dimensional):

$$T : \mathcal{B}(\mathcal{H}^{\otimes M}) \to \mathcal{B}(\mathcal{H}^{\otimes N}) \quad [1]$$

while the Schrödinger picture representation is given in terms of the (pre-)dual of $T$, that is,

$$T_* : \mathcal{B}_*(\mathcal{H}^{\otimes N}) \to \mathcal{B}_*(\mathcal{H}^{\otimes M}) \quad [2]$$

where $\mathcal{B}_*(\,\cdot\,)$ denotes the space of trace-class operators. Hence, if $T$ operates on input systems in the (joint) state $\rho^{\otimes N}$, the output systems (i.e., the "clones") are in the state $T_*(\rho^{\otimes N})$. We will call each such $T$ a cloning map.

Now our aim is to find an operation $T$ such that the output state $T_*(\rho^{\otimes N})$ approximates the product state $\rho^{\otimes M}$ as well as possible. The quality of the approximation is measured by a distance function $\delta$ on the convex set $\mathcal{S}(\mathcal{H}^{\otimes M}) \subset \mathcal{B}_*(\mathcal{H}^{\otimes M})$ of density operators on $\mathcal{H}^{\otimes M}$ and, since it is impossible to minimize $\delta(T_*(\rho^{\otimes N}), \rho^{\otimes M})$ for all $\rho$ simultaneously, we are looking only for the worst case. Hence, the quality of a cloning map $T$ is measured by a figure of merit of the form

$$\Delta_{X,\delta}(T) = \sup_{\rho \in X} \delta\big(T_*(\rho^{\otimes N}), \rho^{\otimes M}\big) \quad [3]$$

Here $X \subset \mathcal{S}(\mathcal{H})$ is a set of "preferred" density operators whose role will be explained in the next section. An optimal cloning device is described by a cloning map $\hat{T}$ which minimizes $\Delta_{X,\delta}$, that is,

$$\Delta_{X,\delta}(\hat{T}) \le \Delta_{X,\delta}(T) \quad [4]$$

should hold for each cloning map $T$.

### The Preferred Set of States

The set $X \subset \mathcal{S}(\mathcal{H})$ of density operators introduced in the last equation describe *a priori* knowledge about the one-particle input state $\rho$; for example, if we want to clone only signal states $\rho_1, \ldots, \rho_k$ used to transmit classical information through a quantum channel, the choice for $X$ is $\{\rho_1, \ldots, \rho_k\}$. Other possibilities include: $X = S(\mathcal{H})$ if nothing is known about $\rho$, the set of pure states, the states in the "equatorial plane" of the Bloch sphere, or Gaussian states if $\mathcal{H}$ is infinite dimensional. Each different choice for $X$ leads to a different variant of the cloning problem, and we will summarize the most relevant cases treated in the literature in the section "Examples."

A different kind of *a priori* knowledge is *a priori* measures, that is, instead of knowing that all possible input states lie in a special set $X$, we know for each measurable set $X \subset \mathcal{S}(\mathcal{H})$ the probability $\mu(X)$ for $\rho \in X$. Such a situation typically arises when we are trying to clone states of systems which

originate from a source with known characteristics. In this case, we can use mean errors,

$$\bar{\Delta}_{\mu,\delta}(T) = \int_{\mathcal{S}(\mathcal{H})} \delta\big(T^*(\rho^{\otimes N}),\ \rho^{\otimes M}\big)\mu(\mathrm{d}\rho) \qquad [5]$$

as a figure of merit. Sometimes these are easier to compute than maximal errors as in eqn [3]. Often, however, $\Delta$ leads to stronger results than $\bar{\Delta}$, therefore we will concentrate our discussion on maximal rather than mean errors.

### The Distance Measure

The remaining freedom in eqn [3] is the distance measure $\delta$ and there are mainly two physically different choices: we can either check the quality of each clone separately or we can test, in addition, the correlations between output systems. The most common choice for a figure of merit for the first type is given by (where $\hat{\mathrm{tr}}_j$ denotes partial trace over all but the $j$th tensor factor)

$$\Delta_1(T) = \sup_{\rho\in X,j} \big|1 - \mathcal{F}\big(\hat{\mathrm{tr}}_j T_*(\rho^{\otimes N}),\ \rho\big)\big| \qquad [6]$$

Here $\mathcal{F}(\rho,\sigma)$ denotes the (quadratic) fidelity of $\rho$ and $\sigma$, that is,

$$\mathcal{F}(\rho,\sigma) = \mathrm{tr}\left(\left(\rho^{1/2}\sigma\rho^{1/2}\right)^{1/2}\right)^2 \qquad [7]$$

and the supremum is taken over all $\rho\in X$ and $j=1,\ldots,N$. $\Delta_1$ measures the worst one-particle error of the output state $T^*(\sigma^{\otimes N})$, and we will refer to it in the following as the local error. If we are interested in correlations too, we have to choose

$$\Delta_{\mathrm{all}}(T) = \sup_{\rho\in X} \big|1 - \mathcal{F}\big(T_*(\rho^{\otimes N}), \rho^{\otimes M}\big)\big| \qquad [8]$$

$\Delta_{\mathrm{all}}$ measures again a "worst-case" error, but now of the full output with respect to $M$ uncorrelated copies of the input $\rho$. We will call it the global error. Alternative figures of merit arise if we replace the fidelity in eqns [6] and [8] by other distance measures like the trace norm, the Hilbert–Schmidt norm, or the relative entropy. If $X$ consists only of pure states, the operations $T$ which minimize $\Delta_1$ or $\Delta_{\mathrm{all}}$ are usually not altered by such different choices. If $X$ is a set of mixed states, however, the correct choice is unclear and might depend on the precise physical context (there is, in particular, no reason to prefer fidelities).

## General Properties

Before we consider more special examples in the next section, let us discuss some general properties of the figure of merit $\Delta_{X,\delta}$ from eqn [3] and the corresponding optimization problem.

### Existence of Solutions

If the distance measure $\delta$ is continuous in the first argument, the optimization problem [4] has a solution, that is, optimal cloning machines exist: the set $\mathcal{T}$ of cloning maps [1] is compact and the quantity $\Delta_{X,\delta}$ is – as a supremum over continuous functions – lower-semicontinuous. Hence, the statement follows from the fact that a lower-semicontinuous function on a compact set always admits a minimizer.

This argument can be generalized to the infinite-dimensional case, if we choose the set $\mathcal{T}$ of allowed cloning maps more carefully (the restriction to normal channels proposed above is most probably not sufficient for this purpose) and if we equip it with an appropriate topology. The latter should be weak enough for $\mathcal{T}$ to be compact, and strong enough for $\Delta_{X,\delta}$ to be lower-semicontinuous. A typical choice is the weak*-topology arising from an embedding of $\mathcal{T}$ into the dual of a Banach space (such that we can apply the Banach–Alaoglu Theorem). Detailed studies in this direction are, however, not yet available.

### Covariant Cloning Maps

To solve the optimization problem [4] is a difficult and, in many cases, impossible task. However, it can be simplified significantly if $X$ and $\delta$ admit a nontrivial symmetry group. Hence, consider again a distance $\delta$ which is continuous and convex in its first argument and a closed subgroup $G$ of the group $U(d)$ of unitary operators on $\mathcal{H}=\mathbb{C}^d$, such that

$$UXU^* \subset X, \quad \delta\big(U^{\otimes M}\rho U^{\otimes M*}, U^{\otimes M}\sigma U^{\otimes M*}\big) = \delta(\rho,\sigma) \qquad [9]$$

hold for all $U\in G$ and $\rho,\sigma\in\mathcal{S}(\mathcal{H}^{\otimes M})$. Then $\Delta_{X,\delta}$ is invariant under the induced $G$ action on the set $\mathcal{T}$ of cloning maps, that is,

$$\Delta_{X,\delta}(\tau_U T) = \Delta_{X,\delta}(T) \quad \text{with } (\tau_U T)(A) = U^{\otimes N} T(U^{\otimes M*} A U^{\otimes M}) U^{\otimes N*} \qquad [10]$$

holds for all $U\in G$ and all $T\in\mathcal{T}$. Convexity of $\Delta_{X,\delta}$ in $T$ implies (with the Haar measure $\mu_{\mathrm{H}}$ on $G$)

$$\Delta_{X,\delta}(\bar{T}) \le \Delta_{X,\delta}(T), \quad \text{with } \bar{T} = \int_G \tau_U(T)\mu_{\mathrm{H}}(\mathrm{d}U) \qquad [11]$$

for all $T$. Hence, we can replace each cloning map by its group average $\bar{T}$ without sacrificing the

quality of the clones. This implies that $\bar{T}$ is optimal if $T$ is, and, since $\bar{T}$ is $G$-covariant,

$$\tau_U(\bar{T}) = \bar{T} \quad \forall U \in G \qquad [12]$$

we can conclude, together with the arguments from the last section, that the optimization problem [4] always admits covariant solutions. Similarly, we can show that permutation invariant (sometimes called "symmetric") solutions exist, that is, cloners which do not prefer a particular clone or a particular input system.

This is a very useful result, because the set of covariant and permutation-invariant $T$ is much smaller than the set of all cloning maps, and it can be parametrized in terms of irreducible representations of $G$ and the permutation group. In particular, the case $G = U(d)$ (such a $T$ is often called "universal" because it does not prefer any direction in the Hilbert space $\mathcal{H}$) leads to quite general solutions.

### Relationships with Quantum State Estimation

If a procedure to estimate the input state $\rho$ from a measurement on the $N$-fold system in the joint state $\rho^{\otimes N}$ is given, there is a simple way to produce a cloning machine: we just have to take the estimate $\hat{\rho}$ for the density matrix $\rho$ and prepare $M > N$ systems in the state $\hat{\rho}^{\otimes M}$. If $X$ is finite and estimation (which in this case is called hypothesis testing) is done in terms of a positive operator valued measure $(E_\sigma)_{\sigma \in X}$, $E_\sigma \in \mathcal{B}(\mathcal{H}^{\otimes N})$, the probability to get the estimate $\sigma \in X$ when the input is in the state $\rho^{\otimes N}$ is given by $\mathrm{tr}(E_\sigma \rho^{\otimes N})$. Hence, the cloning map derived from this estimation scheme is given by

$$\tilde{E}_*(\rho^{\otimes N}) = \sum_{\sigma \in X} \mathrm{tr}\left(E_\sigma \rho^{\otimes N}\right)\sigma^{\otimes M} \qquad [13]$$

A generalization to arbitrary $X$ is straightforward, but requires the use of measure theory. It is easy to see that the cloning map $\tilde{E}$ from eqn [13] is in general not optimal, in particular if $M$ is only slightly bigger than $N$. However, $\tilde{E}$ has the interesting feature that $\Delta_{X,\delta}(\tilde{E})$ depends only on the number of input systems, $N$, but not on the number of clones, $M$, we want to produce. This observation leads immediately to the conjecture that $\tilde{E}$ becomes optimal in the limit $M \to \infty$. A general proof is currently not available, in those cases, however, where optimal cloner and estimater can be explicitly calculated for all $N$ and $M$ (i.e., the cases treated in the sections "Universal pure-state cloning" and "Phase-covariant pure-state cloning") the conjecture is true. A more detailed discussion of this problem together with information about its current status can be found on the web at http://www.imaph.tu-bs.de/qi/problems/problems-html.

## Examples

In this section, we will discuss concrete examples that arise from different choices of the distance measure $\delta$ and the set $X$ of preferred states.

### Universal Pure-State Cloning

The most frequently discussed case arises if $X$ is the set of pure states, that is, the input states are pure, but otherwise unknown. Under this condition, it is sufficient to consider the symmetric part $\mathcal{H}_+^{\otimes N}$ of the tensor product $\mathcal{H}^{\otimes N}$, and only cloning maps $T : \mathcal{B}(\mathcal{H}^{\otimes M}) \to \mathcal{B}(\mathcal{H}_+^{\otimes N})$, because only this part affects the local or the global error. A complete solution for arbitrary $N$, $M$ and all finite-dimensional Hilbert spaces is available for $\Delta_{\mathrm{all}}$ in Werner (1998) and for $\Delta_1$ in Keyl and Werner (1999). Both cases admit the same (surprisingly simple) unique solution

$$\hat{T}_*(\sigma) = \frac{d[N]}{d[M]} S_M \left(\sigma \otimes \mathbb{1}^{\otimes(M-N)}\right) S_M \qquad [14]$$

where $S_M$ is the projection onto the symmetric tensor product $\mathcal{H}_+^{\otimes M}$ and $d[M]$ denotes the dimension of $\mathcal{H}_+^{\otimes M}$. To derive these results, the group-theoretic methods sketched in the section "Covariant cloning maps" are used. The fact that global and local figures of merit are minimized by the same cloning map is surprising and a special feature of pure-state cloning. It implies that correlations and entanglement between the clones does not matter at all.

### Phase-Covariant Pure-State Cloning

Consider a fixed basis $|j\rangle$, $j = 0, \ldots, d-1$, in $\mathcal{H}$ and let $X$ be the set of states given by

$$\psi = |0\rangle + \sum_{j=1}^{d-1} \mathrm{e}^{\mathrm{i}\phi_j}|j\rangle \qquad [15]$$

where the $\phi_j$ denote arbitrary phases. Obviously, this set is invariant under the set of all unitaries which are diagonal in the given basis (i.e., a maximal torus in $U(d)$). Using the methods outlined in the section "Covariant cloning maps," the corresponding cloning problem is (almost) completely solved in Buscemi *et al.* (2005). For arbitrary $d = \dim \mathcal{H}$, $N$ and all $M = N + dk$, with $k \in \mathbb{N}$ a

cloning map which minimizes global as well as local errors is given in terms of the unitary

$$\hat{U}:\mathcal{H}_+^{\otimes N}\to\mathcal{H}_+^{\otimes M},\hat{U}|n_0,\ldots,n_d\rangle = |n_0+k,\ldots,n_d+k\rangle \qquad [16]$$

where $|n_1,\ldots,n_d\rangle, n_j\in\mathbb{N}$, denotes the number basis of $\mathcal{H}^{\otimes N}$ associated with the distinguished basis $|j\rangle$ of $\mathcal{H}$.

### Cloning Finitely Many States

If $X$ is a finite set of pure states, a general solution is not available, but there are several important partial results. The easiest situation arises if the elements of $X$ are mutually orthogonal pure states. In this case, ideal cloning is possible in terms of an appropriately chosen unitary. If the states are linearly independent but nonorthogonal, ideal cloning is possible as well if we consider probabilistic cloning machines (Duan and Guo 1998); that is, there is a nonvanishing probability that the machine fails and does not produce any clones at all (this means $T$ is not unital). Optimal cloning (with deterministic operations) of two nonorthogonal qubit states $\rho_j=|\psi_j\rangle\langle\psi_j|$, $j=1,2$, is considered for all $N$, $M$ in (Bruß *et al.* (1998) and Chefles and Barnett (1999)) (using averaged global fidelity as the figure of merit). The crucial observation in this case is that the optimal clones are pure, that is, $T_*(\rho_j^{\otimes N})=|\Psi_j\rangle\langle\Psi_j|$ and that the $\Psi_j$ lie in the subspace spanned by the (unattainable) ideal clones $\psi_j^{\otimes M}$.

### Universal Mixed-State Cloning

$X=\mathcal{S}(\mathcal{H})$ means that absolutely nothing is known *a priori* about the input state $\rho$. If the distance measure $\delta$ is $U(d)$ and permutation invariant (which is the case for all possible choices discussed in the section "The distance measure") the analysis from the section "Covariant cloning maps" shows that a universal and symmetric minimizer exists. An explicit solution, however, is not known, and even the physically most appropriate choice for $\delta$ is unclear. In contrast to the pure-state case, this is a serious question, because the set of optimal cloners is, in this case, much more sensitive to changes in $\delta$. In particular, correlations among the clones become crucial, and it is very likely that local and global figures of merit lead to very different solutions. To emphasize this difference, an operation which minimizes only local errors is sometimes called "broadcasting," rather than cloning. A related problem with (at least) partial solutions ("purification") will be discussed in the section "Purification."

### Cloning of Gaussian States

If the Hilbert space is infinite dimensional, the restriction to a reasonable small set $X$ of preferred states is crucial, because otherwise the search for minimizers becomes hopeless. A physically relevant class with nice mathematical properties are Gaussian states and in particular coherent states. Cloning of the latter has been studied in Cerf *et al.* (2005) for the case $N=1$ (and $M$ arbitrary). As in the section "Covariant cloning maps," it can be shown that the search for optimal cloners can be restricted to those which are covariant with respect to phase space translations. This simplifies the problem significantly and leads to the result that the global error is minimized by Gaussian cloning maps, while in the local case the best cloner is non-Gaussian.

### Asymmetric Cloning

In all examples discussed up to now, we have considered symmetric cloners, that is, the quality of all clones is measured with equal weight. Alternatively, we can look for asymmetric cloners which produce clones with different quality and ask for the trade-off between them. This problem was first discussed in Cerf (2000) and later in Iblisdir *et al.* (2005). It can be regarded as a constraint optimization problem, where the error of the first $M'<M$ clones should be minimized under the constraint that the error of the rest is bounded by a fixed value. In Iblisdir *et al.* (2005), it is conjectured that for pure input states and local errors the optimal solution to this problem is given by

$$T_*(\sigma) = V^*\left(\sigma\ \otimes \mathbb{1}^{\otimes(M-N)}\right)V \qquad [17]$$

where $V$ is a linear combination of projections in the commutant of $\{U^{\otimes N}\,|\,U\in U(\mathcal{H})\}$. This conjecture is true (at least) for qubits in the case $1\to n+1$ and $1\to 1+n$.

## Related Problems

Instead of cloning, we can also try to approximate other impossible machines by channels which operate on multiple inputs. To this end, we only have to replace the figure of merit [6] by

$$\Delta_{1,\beta}(T) = \sup_{\rho\in X,j}\left|1-\mathcal{F}\left(\hat{\mathrm{tr}}_j T_*(\rho^{\otimes N}),\ \beta(\rho)\right)\right| \qquad [18]$$

where $\beta:\mathcal{S}(\mathcal{H})\to\mathcal{S}(\mathcal{H})$ is a (possibly nonlinear) functional which describes the task we want to approximate. The generalization $\Delta_{\mathrm{all},\beta}$ of $\Delta_{\mathrm{all}}$ can be given similarly. If $\beta$ has the appropriate continuity and symmetry properties, the discussion in the section "General properties" applies completely, that is, we can assume covariance and permutation

invariance, and we can consider operations which use state estimation in an intermediate step.

### Purification

Consider $N$ quantum systems, all originally prepared in the same pure state $\sigma$, and then subsequently exposed to the same (known) decoherence process, described by a depolarizing channel $R$. The task of purification is to produce $M$ output systems which approximate the original pure input state as well as possible. Hence, the corresponding figure of merit arises with $X = \{R(\sigma) \mid \sigma \text{ pure}\}$ and $\beta(\rho) = R^{-1}(\rho)$. This problem is discussed for qubits in Cirac *et al.* (1999), Keyl and Werner (2001) and D'Ariano *et al.* (2005). The optimal purifier can be given explicitly for all $N$, $M$ in terms of irreducible SU(2) representations. Surprisingly, it turns out that the output purity can be improved even if the number of outputs, $M$, is larger than the number of available input systems, $N$ (although $N$ should be large enough). If we measure purity in terms of local errors, it can be shown that, in the limit $N \to \infty$, perfectly purified qubits can be produced at an infinite rate (i.e., the number of output systems per input system can become infinite). However, we have to pay for this result with extremely large correlations between the output systems. Therefore, the global error does not disappear asymptotically, if we insist on a nonvanishing rate.

### Universal Not

"Universal not" (UNOT) is an operation which sends each pure state $\sigma$ to its orthocomplement. This is a positive but not a completely positive operation. Hence, it cannot be performed by any physical device. However, we can try to approximate it by a cloning map $T$ operating on $N$ input systems. The corresponding figure of merit [18] arises if $X$ is the set of pure states and $\beta(\rho) = \mathbb{1} - \rho$. In Bužek *et al.* (1999), it is shown that the optimal solution to this problem (for all $N$ and $M$) is to estimate and reprepare as described in the section "Relationships with quantum state estimation." Approximating UNOT is, therefore, significantly more difficult than (pure-state) cloning, where the optimal solution is always (for finite $M$) better than estimation.

*See also:* Channels in Quantum Information Theory; Compact Groups and Their Representations; Positive Maps on $C^*$-algebras.

## Further Reading


Bruß D, DiVincenzo DP, Ekert A *et al.* (1998) Optimal universal and state-dependent cloning. *Physical Review A* 57(4): 2368–2378.

Buscemi F, D'Ariano GM, and Macchiavello C (2005) Economical phase-covariant cloning of qudits. *Physical Review A* 71: 042327.

Bužek V, Hillery M, and Werner RF (1999) Optimal manipulations with qubits: universal-not gate. *Physical Review A* 60(4): R2626–R2629.

Cerf NJ (2000) Asymmetric quantum cloning machines. *Journal of Modern Optics* 47: 187–209.

Cerf NJ, Krueger O, Navez P, Werner RF, and Wolf MM (2005) Non-Gaussian cloning of quantum coherent states is optimal. *Physical Review Letters* 95: 070501.

Chefles A and Barnett SM (1999) Strategies and networks for state-dependent quantum cloning. *Physical Review A* 60: 000136.

Cirac JI, Ekert AK, and Macchiavello C (1999) Optimal purification of single qubits. *Physical Review Letters* 82: 4344–4347.

D'Ariano GM, Macchiavello C, and Perinotti P (2005) Superbroadcasting of mixed states. *Physical Review Letters* 95: 060503.

Duan L-M and Guo G-C (1998) Probabilistic cloning and identification of linearly independent quantum states. *Physical Review Letters* 80: 4999–5002.

Iblisdir S, Acin A, and Gisin N (2005) Generalised asymmetric quantum cloning, quant-ph/0505152.

Keyl M and Werner RF (1999) Optimal cloning of pure states, testing single clones. *Journal of Mathematical Physics* 40: 3283–3299.

Keyl M and Werner RF (2001) The rate of optimal purification procedures. *Annales Henri Poincaré* 2: 1–26.

Some open problems in quantum information theory, http://www.imaph.tu-bs.de/qi/problems/problems_.html.

Werner RF (1998) Optimal cloning of pure states. *Physical Review A* 58: 980–1003.

Wootters WK and Zurek WH (1982) A single quantum cannot be cloned. *Nature* 299: 802–803.


# Optimal Transportation

**Y Brenier**, Université de Nice Sophia Antipolis, Nice, France



The purpose of this article is to introduce some of the main ideas of optimal transportation theory. A lot more can be found in Villani's book (Villani 2003), in a somewhat similar spirit. Supplementary information is also available in Ambrosio *et al.* (2005), Evans and Gangbo (1999), and Rüschendorf and Rachev (1990).

## Transportation Maps

Let us start by a rather abstract definition:

**Definition 1** Let $X$ and $Y$ be two topological spaces with Borel probability measures $\alpha$ and $\beta$, respectively. We say that a Borel map $T : X \to Y$ is a transportation map between $(X, \alpha)$ and $(Y, \beta)$ if, for each Borel subset $A$ of $Y$,

$$\int_{T(x)\in A} \alpha(\mathrm{d}x) = \int_{y\in A} \beta(\mathrm{d}y)$$

It is customary to say that $T$ pushes forward $\alpha$ to $\beta$, or to say that $\beta$ is the image of $\alpha$ by $T$. An abstract measure-theoretic result asserts that there is always such a transportation map $T$, as soon as $\alpha$ has no atom (i.e., the $\alpha$ measure of any point $x \in X$ is zero).

A more concrete situation is when $X = \bar{\Omega}_0, Y = \bar{\Omega}_1$, where $\Omega_0$ and $\Omega_1$ are two smooth bounded open subsets of the $d$-dimensional Euclidean space $\mathbf{R}^d$. In such a case, a classical result, due to Moser and improved by Dacorogna and Moser (1990), reads:

**Theorem 1** *Let $\Omega_0$ and $\Omega_1$ be two smooth bounded open sets in $\mathbf{R}^d$. Let $\rho_0 > 0$ and $\rho_1 > 0$ be two smooth functions on $\mathbf{R}^d$ such that*

$$\int_{\Omega_0} \rho_0(x)\mathrm{d}x = \int_{\Omega_1} \rho_1(x)\mathrm{d}x = 1$$

*Then there is a smooth transportation map $T$ between $(\bar{\Omega}_0, \rho_0(x)\mathrm{d}x)$ and $(\bar{\Omega}_1, \rho_0(y)\mathrm{d}y)$. Furthermore, $T$ is an orientation-preserving diffeomorphism and solves the Jacobian equation:*

$$\rho_1(T(x))\det(DT(x)) = \rho_0(x), \quad \forall x \in \Omega_0 \qquad [1]$$

## Transportation Maps with Convex Potentials

An important property of Moser's construction, which we did not state, is the possibility of prescribing the restriction of $T$ along the boundary $\partial\Omega_0$. If one does not care about this latter property, one can improve Theorem 1 as follows (Caffarelli 1992):

**Theorem 2** *Assume further that $\Omega_1$ is a uniformly strictly convex set. Then, there is a transportation map $T$ with a smooth convex potential, namely*

$$T(x) = D\Phi(x), \quad \forall x \in \Omega_0$$

*for some smooth convex function $\Phi$ defined on $\mathbf{R}^d$ and strictly convex on $\Omega_0$. In addition, among all Borel maps $T$ transporting $(\bar{\Omega}_0, \rho_0(x)\mathrm{d}x)$ to $(\bar{\Omega}_1, \rho_1(y)\mathrm{d}y)$, $D\Phi$ is the unique map that minimizes*

$$\inf_T \int_{\mathbf{R}^d} |T(x) - x|^2 \rho_0(x)\mathrm{d}x \qquad [2]$$

*where $|\cdot|$ denotes the Euclidean norm on $\mathbf{R}^d$.*

Because of its characterization, $T = D\Phi$ is often called the "optimal transportation map" with respect to the "transportation cost" [2]. Notice that, because of the Jacobian equation [1], $\Phi$ automatically is a classical solution to the Monge–Ampère equation:

$$\rho_1(D\Phi(x))\det(D^2\Phi(x)) = \rho_0(x), \quad \forall x \in \Omega_0 \qquad [3]$$

(The Monge–Ampère equation is a famous geometric PDE, related to the seeking of hypersurfaces with prescribed Gaussian curvature.) The main gain with respect to Moser's construction is the property that the optimal map $T$ has, at each $x \in \Omega_0$, a Jacobian matrix $DT(x) = D^2\Phi(x)$ which is a positive-definite symmetric matrix. This property has been first exploited by McCann (1997) and later by many authors (see Villani (2003), for many references) to prove a large series of geometric and functional inequalities. A very fine example can be found in Barthe (1998). Let us just consider, as an elementary illustration, a short and sharp proof of the isoperimetric inequality using the optimal transportation map.

## A Proof of the Isoperimetric Inequality Using Optimal Transportation Maps

Let us recall the isoperimetric inequality:

**Theorem 3** *Let $\Omega$ be a smooth bounded open subset in $\mathbf{R}^d$. Then*

$$|\partial\Omega| \geq \mathrm{d}|B_1|^{1/d}|\Omega|^{1-1/d}$$

*holds true where $B_1$ is the unit ball in $\mathbf{R}^d, |\Omega|$ and $|\partial\Omega|$, respectively, denote the $d$-dimensional volume of $\Omega$ and the $(d-1)$-dimensional Hausdorff measure of the boundary $\partial\Omega$. In addition, the inequality becomes an equality if and only if $\Omega$ is a ball.*

To prove this result, let us define densities:

$$\rho_0(x) = \frac{1}{|\Omega|}, \quad x \in \Omega$$
$$\rho_1(y) = \frac{1}{|B_1|}, \quad y \in B_1$$

and consider the associated optimal transportation map $D\Phi$ from $(\bar{\Omega}_0, \rho_0(x)\mathrm{d}x)$ to $(\bar{\Omega}_1, \rho_0(y)\mathrm{d}y)$. From the Monge–Ampère equation,

$$\rho_1(D\Phi(x))\det(D^2\Phi(x)) = \rho_0(x)$$

we get:

$$\det(D^2\Phi(x)) = \frac{|B_1|}{|\Omega|}, \quad x \in \Omega \qquad [4]$$

Since the range of $D\Phi$ on $\Omega$ is the unit ball $B_1$, we have

$$I = \int_{\partial\Omega} D\Phi(x) \cdot n(x)\mathrm{d}\sigma(x) \leq \int_{\partial\Omega} \mathrm{d}\sigma(x) = |\partial\Omega|$$

where $n(x)$ and $\mathrm{d}\sigma(x)$ respectively, denote the outward unit normal and the $(d-1)$-dimensional

Hausdorff measure along $\partial\Omega$. Using the divergence theorem, we also have:

$$I = \int_\Omega \Delta\Phi(x)\mathrm{d}x$$

where $\Delta\Phi(x) = \mathrm{trace}(D^2\Phi(x))$ is the Laplacian of $\Phi$. From the geometric mean inequality, we know that, for any symmetric matrix $A \geq 0$,

$$(\det A)^{1/d} \leq 1/d \text{ trace } (A)$$

holds true, with equality if and only if $A$ is equal to the identity matrix multiplied by a non-negative scalar factor. Thus,

$$I \geq d\int_\Omega (\det(D^2\Phi(x))^{1/d}\mathrm{d}x$$
$$= d|\Omega|^{1-1/d}|B_1|^{1/d}$$

(because of [4]). So, we have obtained the isoperimetric inequality:

$$|\partial\Omega| \geq \mathrm{d}|B_1|^{1/d}|\Omega|^{1-1/d}$$

Let us now consider the case when this inequality becomes an equality. Then, necessarily, for each $x \in \Omega, A = D^2\Phi(x)$ satisfies $\det A = (\mathrm{trace}(A)/d)^d$ and, therefore, must be the identity matrix multiplied by a scalar factor $\lambda > 0$, possibly depending on $x$. Because of [4], the determinant of $D^2\Phi(x)$ is constant over $\Omega$. Thus, $\lambda > 0$ must be constant. It follows that $D\Phi(x) = \lambda(x - a)$, for some point $a$ in $\boldsymbol{R}^d$. Therefore, $\Omega$ must be the ball centered at $a$ of radius $1/\lambda$.

## Monge's Optimal Transportation Problem

Theorem 2 is one of the numerous avatars of the so-called optimal transportation theory that goes back to Monge's mass transfer problem which addressed in 1781 the 'mémoire sur la théorie des déblais et des remblais' and was completely renewed by Kantorovich in the 1940s (see e.g., Rüschendorf and Rachev (1990) for instance). Let us quote a typical result, similar to Theorem 2, but without regularity assumptions on the data (see Brenier and Caffarelli (1992)):

**Theorem 4** *Let $\rho_0$ be a non-negative Lebesgue integrable function on $\boldsymbol{R}^d$, such that*

$$\int_{\boldsymbol{R}^d} \rho_0(x)\mathrm{d}x = 1$$

*Then for any Borel probability measure $\rho_1(\mathrm{d}y)$ with compact support on $\boldsymbol{R}^d$, there is a unique map $T$ transporting $\rho_0(x)\mathrm{d}x$ to $\rho_1(\mathrm{d}y)$, which minimizes*

$$\int_{\boldsymbol{R}^d} |T(x) - x|^2\rho_0(x)\mathrm{d}x$$

*where $|\cdot|$ denotes the Euclidean norm on $\boldsymbol{R}^d$. In addition, there is a Lipschitz continuous convex function $\Phi$ defined on $\boldsymbol{R}^d$ such that $T(x) = D\Phi(x)$ for $\rho_0$ almost every $x \in \boldsymbol{R}^d$, which implies:*

$$\int_{\boldsymbol{R}^d} f(D\Phi(x))\rho_0(x)\mathrm{d}x = \int_{\boldsymbol{R}^d} f(y)\rho_1(\mathrm{d}y)$$

*for all continuous functions $f$ on $\boldsymbol{R}^d$.*

Theorem 2, which can be interpreted as a regularity result with respect to Theorem 4, is the main output of Caffarelli's regularity theory for transportation maps with convex potentials (Caffarelli 1992). Caffarelli's analysis starts by a proof that $\Phi$ actually is a weak solution of the Monge–Ampère equation [3] in the sense of Alexandrov and is strictly convex. Then, Caffarelli shows that $D^2\Phi$ is Hölder continuous, as soon as $\rho_0$ and $\rho_1$ are Hölder continuous.

Notice that the convexity assumption for $\Omega_1$ is crucial to insure the regularity of the convex potential. Caffarelli provided counter-examples when $\Omega_1$ is made of two separate balls attached together by a sufficiently thin pipe.

Surprisingly enough, results such as Theorem 4 are related to concrete applications in, for example, astrophysics, image processing, etc. (Frisch *et al.* 2002, Haker and Tannenbaum 2003).

## The Kantorovich Optimal Transportation Problem

The Monge optimal transportation problem can be solved using the Kantorovich duality method, based on the key concept of "generalized transportation maps," also called "transportation plans" or "doubly stochastic measures." The abstract definition is:

**Definition 2** Let $X$ and $Y$ be two topological spaces with Borel probability measures $\alpha$ and $\beta$, respectively. We say that a Borel probability measure $\mu$ on $X \times Y$ is a generalized transportation map, or a transportation plan, if its marginals are, respectively, $\alpha$ and $\beta$, namely

$$\begin{aligned}\int_{x\in A, y\in Y} \mu(\mathrm{d}x, \mathrm{d}y) &= \int_{x\in A} \alpha(\mathrm{d}x) \\ \int_{x\in X, y\in B} \mu(\mathrm{d}x, \mathrm{d}y) &= \int_{y\in B} \beta(\mathrm{d}y)\end{aligned} \qquad [5]$$

for all Borel subsets $A$ and $B$ of $X$ and $Y$, respectively.

The Monge–Kantorovich (MK) optimal transportation problem amounts, given a "transportation

cost," that is, a continuous function $c: X \times Y \to \boldsymbol{R}$, to find a minimizer for

$$I_{\mathrm{MK}} = \inf_{\mu} \int c(x, y)\mu(\mathrm{d}x, \mathrm{d}y) \qquad [6]$$

where $\mu$ is subject to be a transportation plan between $(X, \alpha)$ and $(Y, \beta)$. Notice that this problem is convex (and can be seen as an infinite-dimensional linear program) and its dual problem can be easily computed (using, e.g., Rockafellar's theorem in convex analysis and assuming, for simplicity, that both $X$ and $Y$ are compact).

**Theorem 5** *We have*

$$I_{\mathrm{MK}} = \sup_{a,b} \left\{ \int a(x)\alpha(\mathrm{d}x) + \int b(y)\beta(\mathrm{d}y) \right\} \qquad [7]$$

*where $(a, b)$ is any pair of continuous functions, defined on X and Y, respectively, and subject to*:

$$a(x) + b(y) \le c(x, y), \quad \forall x \in X, \forall y \in Y$$

Of course, each transportation map T, in the sense of Definition 1, can be seen as a transportation plan $\mu$ in the Kantorovich framework, just by setting

$$\mu(\mathrm{d}x, \mathrm{d}y) = \delta(y - T(x))\alpha(\mathrm{d}x)$$

which means

$$\int_{x \in A,\, y \in B} \mu(\mathrm{d}x, \mathrm{d}y) = \int_{x \in A, T(x) \in B} \alpha(\mathrm{d}x)$$

for all Borel subsets $A$ and $B$ of $X$ and $Y$, respectively. Then, we have

$$\int c(x, y)\mu(\mathrm{d}x, \mathrm{d}y) = \int c(x, T(x))\alpha(\mathrm{d}x)$$

So, the MK problem can be seen as a "relaxed" version of the "classical" optimal transportation problem à *la* Monge:

$$I_{\mathrm{M}} = \inf_{T} \int c(x, T(x))\alpha(\mathrm{d}x) \qquad [8]$$

where $T$ is subject to be a transportation map between $(X, \alpha)$ and $(Y, \beta)$. Indeed, we have $I_{\mathrm{MK}} \le I_{\mathrm{M}}$. It turns out that, in many important situations, there is no gap between these two values, which makes the MK problem a perfectly convenient convex substitute for the original, nonconvex, Monge transportation problem. This is, in particular, the case of the situation considered in Theorem 4, when the cost function is just

$$c(x, y) = |x - y|^2$$

or, more generally, $c(x, y) = k(x - y)$, where $k$ is a uniformly strictly convex function. A typical result is:

**Theorem 5** *Let $\rho_0$ be a non-negative Lebesgue integrable function on $\boldsymbol{R}^d$, with unit integral, and $\rho_1(\mathrm{d}y)$ be a Borel probability measure with compact support on $\boldsymbol{R}^d$. Let $k$ be a uniformly strictly convex function on $\boldsymbol{R}^d$. Then the MK problem*

$$I_{\mathrm{MK}} = \inf_{\mu} \int k(y - x)\mu(\mathrm{d}x, \mathrm{d}y)$$

*where $\mu$ is subject to be a transportation plan between $\rho_0(x)\mathrm{d}x$ and $\rho_1(\mathrm{d}y)$ on $\boldsymbol{R}^d$, has a unique solution of form*

$$\mu(\mathrm{d}x, \mathrm{d}y) = \delta(y - T(x))\alpha(\mathrm{d}x)$$

*where $T$ is the unique minimizer of the Monge problem*:

$$I_{\mathrm{M}} = \inf_{T} \int k(T(x) - x)\rho_0(x)\mathrm{d}x$$

*among all transportation maps $T$ between $\rho_0(x)\mathrm{d}x$ and $\rho_1(\mathrm{d}y)$ on $\boldsymbol{R}^d$. In addition $I_{\mathrm{MK}} = I_{\mathrm{M}}$.*

**Proof for Theorem 5 (Sketch)** For simplicity, we assume that $\rho_0$ and $\rho_1$ are both compactly supported in a ball $B$ in $\boldsymbol{R}^d$ and we limit ourselves to the simplest cost function $k(x) = |x|^2/2$. We first denote by $M$ the set of all Borel regular probability measures $\nu$ on $B \times B$ having $\rho_0(x)\mathrm{d}x$ and $\rho_1(\mathrm{d}y)$ as marginals, which means

$$\int_{B \times B} f(x)\nu(\mathrm{d}x, \mathrm{d}y) = \int_B f(x)\rho_0(x)\mathrm{d}x$$
$$\int_{B \times B} f(y)\nu(\mathrm{d}x, \mathrm{d}y) = \int_B f(y)\rho_1(\mathrm{d}y)$$

for all continuous functions $f$ on $\boldsymbol{R}^d$. From Theorem 7, we deduce:

$$\max_{\nu \in M} \int_{B \times B} x \cdot y \nu(\mathrm{d}x, \mathrm{d}y) = \inf \int_B [\Phi(x)\rho_0(x) + \Psi(x)\rho_1(x)]\mathrm{d}x$$

where the infimum is taken over all pairs $(\Phi, \Psi)$ of continuous functions on $B$ satisfying

$$\Phi(x) + \Psi(y) \ge x \cdot y, \quad \forall x \in B, \forall y \in B$$

Then, it can be established that the infimum is attained by a pair $(\Phi, \Psi)$ such that $\Phi$ is the restriction of a Lipschitz continuous convex function defined on $\boldsymbol{R}^d$, and for $\rho_0(x)\mathrm{d}x$ almost every point of $\boldsymbol{R}^d$, $\Psi$ coincides with the Legendre–Fenchel transform of $\Phi$,

$$\mathrm{LF}(\Phi)(y) = \sup_{x \in \boldsymbol{R}^d} (x \cdot y - \Phi(x))$$

Moreover, if $\nu = \nu_{\text{opt}} \in M$ maximizes $\int_{B\times B} x \cdot y\nu(\mathrm{d}x, \mathrm{d}y)$, then

$$\Phi(x) + \Psi(y) = x \cdot y$$

holds for $\nu_{\text{opt}}$-almost every $(x, y) \in \boldsymbol{R}^d \times \boldsymbol{R}^d$. Using well-known properties of the Legendre–Fenchel transform in convex analysis, one deduces that $\nu_{\text{opt}}$ is necessarily of the form

$$\nu_{\text{opt}}(\mathrm{d}x, \mathrm{d}y) = \delta(y - D\Phi(x))\rho_0(x)\,\mathrm{d}x$$

which implies

$$\int_{\boldsymbol{R}^d\times\boldsymbol{R}^d} f(y)\nu_{\text{opt}}(\mathrm{d}x, \mathrm{d}y) = \int_{\boldsymbol{R}^d} f(D\Phi(x))\rho_0(x)\,\mathrm{d}x$$

for all continuous functions $f$ on $\boldsymbol{R}^d$ and achieves the proof since the second marginal of $\nu_{\text{opt}}$ is $\rho_1(\mathrm{d}y)$.

## The Wasserstein Distance

Optimal transportation theory is strongly related to the geometric analysis of probability measures. For simplicity, let us just consider the space Prob($B$) of all Borel probability measures $\rho$ supported by some fixed ball $B$ in $\boldsymbol{R}^d$. This space is compact for the weak topology of measures. An equivalent definition of this topology is provided by the distance $d$, naturally attached to the MK problem:

$$\mathrm{d}(\rho_0, \rho_1) = \inf_{\nu}\left(\int_{B\times B} |x - y|^2\mu(\mathrm{d}x, \mathrm{d}y)\right)^{1/2} \qquad [9]$$

where $\mu$ is subject to be a transportation plan between $\rho_0$ and $\rho_1$ on $B$. (Of course, more general convex functions $k$ can be used to define the cost function.) It has become popular to call this distance as Wasserstein distance (or its generalizations for various $k$). It turns out that Prob($B$) equipped with this distance has a formal Riemannian structure (Otto 2001, Ambrosio *et al.* 2005). For instance, given two probability measures $\rho_0(x)\mathrm{d}x$ and $\rho_1(x)\mathrm{d}x$, we can define a "shortest path" $t \to \rho(t, \cdot) \in$ Prob($B$) such that $\rho(0) = \rho_0$, $\rho(1) = \rho_1$, just by setting:

$$\rho(t, \mathrm{d}x) = \int_B \delta(a + (D\Phi(a) - a)t - x)\rho_0(a)\mathrm{d}a, \quad \forall t \in [0, 1]$$

where $D\Phi$ is the optimal transportation map between $\rho_0$ and $\rho_1$ on $B$. This idea, which is somewhat related to the geometric analysis of hydrodynamics and various concepts of generalized flows Arnol'd and Khesin 1998, Brenier, was successfully used by McCann (1997) and Otto (2001). In particular, the concept of convexity along these geodesic paths on Prob($B$) has been pointed out by McCann (1997) to be a crucial tool for new proofs of geometric and functional inequalities. Otto, and other contributors (see Ambrosio *et al.* (2005) for a comprehensive discussion), observed that many important parabolic or dissipative evolution PDEs can be described as "gradient flows" (or "steepest descent") of such functionals, with respect to the Wasserstein metric.

## Further Reading


Ambrosio L, Gigli N, and Savaré G (2005) *Gradient Flows*, ETH Lectures in Mathematics. Birkhäuser.

Arnol'd VI and Khesin B (1998) *Topological Methods in Hydrodynamics*. Berlin: Springer.

Barthe F (1998) *Inventiones Mathematicae* 134: 335–361.

Brenier Y Polar factorization and monotone rearrangement of vector-valued functions.

Brenier Y (1991) Minimal geodesics on groups of volume-preserving maps. *Communications on Pure and Applied Mathematics* 44: 375–417.

Brenier Y (1999) Minimal geodesics on groups of volume-preserving maps. *Communications on Pure and Applied Mathematics* 52: 411–452.

Caffarelli L (1992) Boundary regularity of maps with convex potentials. *Communications on Pure and Applied Mathematics* 45: 1141–1151.

Cullen M, Norbury J, and Purser J (1991) Generalised Lagrangian solutions for atmospheric and oceanic flows. *SIAM Journal of Applied Mathematics* 51: 20–31.

Dacorogna B and Moser J (1990) *Ann. Inst. H. Poincaré Anal. NL* 7: 1–26.

Evans C and Gangbo W (1999) *Mem. Amer. Math. Soc.* 137: 653.

Frisch U, Matarrese S, Mohayaee R, and Sobolevski A (2002) *Nature* 417: 260–262.

Haker S and Tannenbaum A (2003) *New Approach to Monge–Kantorovich with Applications to Computer Vision and Image Processing*, IMA Series on Applied Mathematics. Springer.

McCann R (1997) *Advances in Mathematics* 128: 153–179.

Otto F (2001) *CPDEs* 26: 101–174.

Rüschendorf L and Rachev ST (1990) *Journal of Multivariate Analysis* 32: 48–54.

Rüschendorf L and Rachev ST (1998) *Mass Transportation Problems*. Berlin: Springer.

Villani C (2003) *Topics in Optimal Transportation*, AMS Graduate Studies in Mathematics.

# Ordinary Special Functions

**W Van Assche**, Katholieke Universiteit Leuven, Leuven, Belgium



## Introduction

The exponential function, the logarithm, the trigonometric functions, and various other functions are often used in mathematics and physics. They are transcendental functions in the sense that they cannot be obtained by a finite number of operations as a solution of an algebraic (polynomial) equation. Typically, they are obtained by a Taylor series expansion. Many other higher transcendental functions arise in mathematical physics, often as solutions of differential equations. A precise knowledge of the behavior of such functions, their relation with other functions, addition, multiplication and composition properties, representations as an infinite series, or as an integral, often shed a lot of light onto the problem in which they arise. If they are sufficiently useful to a large audience, then they usually get a name and they will be called special functions. In what follows, we describe a few of these special functions of one variable, but clearly this is just a tip of the iceberg. Many other special functions exist and we refer to the classical tables of Abramowitz and Stegun (1964) and the Bateman manuscript project (Erdélyi *et al.* 1953–55) for more special functions. Nowadays, there have been numerous $q$-extensions of special functions (*see* q-Special Functions).

## Gamma and Beta Function

The gamma function is defined by

$$\Gamma(z) = \int_0^\infty t^{z-1} \mathrm{e}^{-t}\, \mathrm{d}t, \quad \Re z > 0. \qquad [1]$$

It satisfies the functional equation $\Gamma(z+1) = z\Gamma(z)$ and since $\Gamma(1) = 1$ we have $\Gamma(n+1) = n!$ for $n \in \mathbb{N}$. The gamma function therefore extends the factorial function for integers to complex numbers. The functional equation

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin \pi z} \qquad [2]$$

allows to continue the gamma function analytically to $\Re z < 0$ and the gamma function becomes an analytic function in the complex plane, with a simple pole at 0 and at all the negative integers. The residue of $\Gamma(z)$ at $z = -n$ is equal to $(-1)^n/n!$. Legendre's duplication formula is

$$\Gamma(2z) = \frac{2^{2z-1}}{\sqrt{\pi}} \Gamma(z)\Gamma(z+1/2) \qquad [3]$$

from which one can obtain the special value $\Gamma(1/2) = \sqrt{\pi}$. Finally, two useful infinite product representations are

$$\Gamma(z) = \lim_{n\to\infty} \frac{n!n^z}{z(z+1)\cdots(z+n)}$$

and

$$\frac{1}{\Gamma(z)} = z\mathrm{e}^{\gamma z} \prod_{n=1}^{\infty} \left( (1+z/n)\mathrm{e}^{-z/n} \right)$$

where $\gamma$ is Euler's constant:

$$\gamma = \lim_{n\to\infty} \left( \sum_{k=1}^{n} \frac{1}{k} - \log n \right) = 0.577\,215\,664\,9\ldots \qquad [4]$$

The beta function is a function of two variables given by

$$\mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\, \mathrm{d}t$$
$$\Re x > 0, \Re y > 0 \qquad [5]$$

Clearly it satisfies $\mathrm{B}(x,y) = \mathrm{B}(y,x)$ and it is related to the gamma function by

$$\mathrm{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \qquad [6]$$

The gamma and beta function are quite useful in probability theory. One of the most common probability distributions on the positive real line is the gamma distribution

$$\Pr(X \le x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x \mathrm{e}^{-t/\beta} t^{\alpha-1}\, \mathrm{d}t, \quad x \ge 0$$

The case $\alpha = 3/2$ is the Maxwell–Boltzmann distribution. The most common probability distribution on the interval [0, 1] is the beta distribution

$$\Pr(Y \le x) = \frac{1}{\mathrm{B}(\alpha,\beta)} \int_0^x t^{\alpha-1}(1-t)^{\beta-1} \mathrm{d}t$$

where $0 \le x \le 1$.

The psi function is the logarithmic derivative of the gamma function

$$\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)} \qquad [7]$$

It is meromorphic with simple poles at 0 and at the negative integers. Special values are $\psi(1) = -\gamma$ and

$$\psi(n+1) = \sum_{k=1}^{n} \frac{1}{k} - \gamma$$

where $\gamma$ is Euler's constant. These can be obtained from the functional equation

$$\psi(z) = \psi(z+1) - \frac{1}{z}$$

## Bessel Functions

Bessel's differential equation is

$$x^2 y'' + xy' + (x^2 - \nu^2)y = 0 \qquad [8]$$

where derivatives are with respect to $x$ and $\nu$ is a complex number. This differential equation has a regular singularity at $x = 0$ and an irregular singularity at $x = \infty$. The standard method of finding a solution in the neighborhood of a regular singularity gives the solution

$$J_\nu(x) = (x/2)^\nu \sum_{k=0}^{\infty} \frac{(-x^2/4)^k}{k!\Gamma(k+\nu+1)}$$

and $J_{-\nu}(x)$ is another solution (if $\nu \neq 0$). The function $J_\nu$ is called the "Bessel function of the first kind" and $\nu$ is the "order" of the Bessel function. The series $x^{-\nu} J_\nu(x)$ is an entire function of the variable $x$. The function

$$Y_\nu(x) = \frac{J_\nu(x)\cos(\nu\pi) - J_{-\nu}(x)}{\sin(\nu\pi)}$$

is also a solution of Bessel's differential equation and is known as the "Bessel function of the second kind of order $\nu$." Two other solutions that are often used are

$$H_\nu^{(1)}(x) = J_\nu(x) + \mathrm{i}Y_\nu(x)$$
$$H_\nu^{(2)}(x) = J_\nu(x) - \mathrm{i}Y_\nu(x)$$

which are the first and second "Hankel functions."

Bessel functions appear if one solves the wave equation in cylindrical or spherical coordinates, using separation of variables. The Helmholtz equation $\nabla^2 F + k^2 F = 0$ in cylindrical coordinates $\rho, \phi, z$ is

$$\frac{\partial^2 F}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial F}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 F}{\partial \phi^2} + \frac{\partial^2 F}{\partial z^2} + k^2 F = 0$$

and if we look for a solution of the form $f(\rho)g(\phi)h(z)$, then this leads to a differential equation for $f$ of the form

$$\frac{\mathrm{d}^2 f}{\mathrm{d}\rho^2} + \frac{1}{\rho}\frac{\mathrm{d}f}{\mathrm{d}\rho} + [k^2 - a^2 - (\nu/\rho)^2]f = 0$$

where $a$ and $\nu$ are separation constants. The general solution is $f(\rho) = Z_\nu(\rho(k^2 - a^2))$, where $Z_\nu$ is any of the Bessel functions given higher or linear combinations of them. In spherical coordinates $r$, $\theta$, $\phi$ the Helmholtz equation is

$$\frac{\partial^2 F}{\partial r^2} + \frac{2}{r}\frac{\partial F}{\partial r} + \frac{1}{r^2}\frac{\partial^2 F}{\partial \theta^2} + \frac{\cot\theta}{r^2}\frac{\partial F}{\partial \theta} + \frac{1}{r^2 \sin^2 \phi}\frac{\partial^2 F}{\partial \phi^2} + k^2 F = 0$$

and for a solution of the form $f(r)g(\theta)h(\phi)$ one obtains a differential equation for $f$ of the form

$$\frac{1}{r}\frac{\mathrm{d}^2(rf)}{\mathrm{d}r^2} + [k^2 - \nu(\nu+1)/r^2]f = 0$$

with general solution $f(r) = Z_{\nu+(1/2)}(kr)/\sqrt{r}$.

Bessel functions have very simple differentiation formulas:

$$[z^\nu J_\nu(z)]' = z^\nu J_{\nu-1}(z)$$
$$[z^{-\nu} J_\nu(z)]' = -z^{-\nu} J_{\nu+1}(z)$$

The first formula can be seen as a lowering operation, the second as a raising operation. Some integral representations are

$$J_\nu(z) = \frac{(z/2)^\nu}{\sqrt{\pi}\,\Gamma(\nu+1/2)} \int_0^\pi \sin^{2\nu}\theta \cos(z\cos\theta)\mathrm{d}\theta$$

or

$$J_\nu(z) = \frac{(z/2)^\nu}{\sqrt{\pi}\,\Gamma(\nu+1/2)} \int_{-1}^{1} (1-x^2)^{\nu-1/2} \cos zx \,\mathrm{d}x$$

which hold for $\Re\nu > -1/2$. For real $\nu$ the Bessel function $J_\nu$ has infinitely many real zeros, and when $\nu > -1$, then all the zeros are real. All the zeros are simple (except possibly at the origin). Each of the functions $J_\nu(z), Y_\nu(z), H_\nu^{(1)}(z)$, or $H_\nu^{(2)}(z)$ satisfies the recurrence relation

$$za_{\nu-1}(z) + za_{\nu+1}(z) = 2\nu a_\nu(z)$$

and the differential–recurrence relation

$$a_{\nu-1}(z) - a_{\nu-1}(z) = 2a'_\nu(z)$$

### Modified Bessel Functions

The modified Bessel equation is

$$x^2y'' + xy' - (x^2 + \nu^2)y = 0 \quad [9]$$

Clearly $J_\nu(\mathrm{i}x)$ is a solution of this equation. The "modified Bessel function of the first kind" is defined as

$$I_\nu(x) = \mathrm{e}^{-\nu\pi\mathrm{i}/2} J_\nu(x\mathrm{e}^{\pi\mathrm{i}/2}), \quad -\pi < \arg x \leq \pi/2 \quad [10]$$

so that

$$I_\nu(x) = (x/2)^\nu \sum_{k=0}^{\infty} \frac{(x/2)^{2k}}{k!\Gamma(\nu + k + 1)}$$

If $\nu$ is not an integer, then $I_\nu(x)$ and $I_{-\nu}(x)$ are two linearly independent solutions of [9], and when $\nu = n$ is an integer one has $I_n(x) = I_{-n}(x)$. The "modified Bessel function of the second kind" is defined by

$$K_\nu(x) = \frac{\pi}{2\sin\nu\pi}[I_{-\nu}(x) - I_\nu(x)]$$

Some special cases of modified Bessel functions are

$$I_{1/2}(x) = \sqrt{\frac{2}{\pi x}}\sinh x$$
$$I_{-1/2}(x) = \sqrt{\frac{2}{\pi x}}\cosh x$$

and

$$K_{1/2}(x) = \sqrt{\frac{\pi}{2x}}\mathrm{e}^{-x}$$

One has the integral representation

$$K_\nu(z) = \int_0^\infty \mathrm{e}^{-z\cosh x}\cosh\nu x\,\mathrm{d}x$$

and

$$I_\nu(z) = \frac{(z/2)^\nu}{\sqrt{\pi}\,\Gamma(\nu + 1/2)}\int_{-1}^{1}(1 - x^2)^{\nu - 1/2}\mathrm{e}^{\pm zx}\,\mathrm{d}x$$

whenever $\Re\nu > -1/2$. The "Airy functions" are given by

$$\mathrm{Ai}(z) = \frac{\sqrt{z}}{3}\left[I_{-1/3}(\zeta) - I_{1/3}(\zeta)\right] = \frac{\sqrt{z/3}}{\pi}K_{1/3}(\zeta)$$
$$\mathrm{Bi}(z) = \sqrt{z/3}\left[I_{-1/3}(\zeta) + I_{1/3}(\zeta)\right]$$

where $\zeta = 2z^{2/3}/3$. They are both a solution of Airy's differential equation

$$y''(z) - zy(z) = 0$$

## Hypergeometric Series

A power series $\sum_{n=0}^{\infty} c_n z^n$ is said to be hypergeometric when the ratio $c_{n+1}/c_n$ is a rational function of the index $n$. Most series that one finds in calculus textbooks are hypergeometric series and some of them define important special functions. When

$$\frac{c_{n+1}}{c_n} = \frac{(n + a_1)(n + a_2)\cdots(n + a_p)}{(n + b_1)(n + b_2)\cdots(n + b_q)(n + 1)}$$

then we write the corresponding series as

$$ {}_pF_q\left(\left.\begin{matrix} a_1, a_2, \ldots, a_p \\ b_1, b_2, \ldots, b_q \end{matrix}\right| z\right) = \sum_{n=0}^{\infty} \frac{(a_1)_n(a_2)_n\cdots(a_p)_n}{(b_1)_n(b_2)_n\cdots(b_q)_n}\frac{z^n}{n!} \quad [11]$$

where $(a)_n = a(a + 1)(a + 2)\cdots(a + n - 1)$, with $(a)_0 = 1$, is the rising factorial or Pochhammer symbol. When $p$ and $q$ are small, one also uses the notation ${}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; z)$ where a semicolon (;) is used to separate the parameters in the numerator from the parameters in the denominator and also to separate the parameters from the variable $z$. Some special cases are:

- the exponential series

$$ {}_0F_0(-; -; z) = \sum_{n=0}^{\infty}\frac{z^n}{n!} = \exp(z)$$

- the geometric series

$$ {}_1F_0(1; -; z) = \sum_{n=0}^{\infty} z^n = \frac{1}{1 - z}$$

- the binomial series

$$ {}_1F_0(-\alpha; -; -z) = \sum_{n=0}^{\infty}\binom{\alpha}{n}z^n = (1 + z)^\alpha$$

- the logarithmic function

$$ {}_2F_1(1, 1; 2; z) = \sum_{n=0}^{\infty}\frac{z^n}{n + 1} = -\frac{1}{z}\log(1 - z)$$

- the Bessel function

$$(z/2)^\nu {}_0F_1(-; \nu + 1; -z^2/4) = \Gamma(\nu + 1)J_\nu(z)$$

For generic values of the parameters, we see that the hypergeometric series converges everywhere in the complex plane when $q \geq p$, it converges for $|z| < 1$ when $p = q + 1$, and for $p > q + 1$ it is only defined at

$z = 0$. When one of the numerator parameters is a negative integer, say $a_1 = -m$, then the series is terminating and defines a polynomial of degree $m$. None of the denominator parameters is allowed to be a negative integer $-m$, unless there is a numerator parameter which is a negative integer $-k$ with $k < m$. For $q \geq p$, the hypergeometric series therefore defines an entire function which is the corresponding hypergeometric function. For $p = q + 1$, the hypergeometric series only converges in the open unit disk, but sometimes it can be continued analytically to a larger domain in the complex plane. The analytic continuation of the hypergeometric series is then called the hypergeometric function. Take for example the geometric series, then it is clear that the hypergeometric series converges in the open unit disk, but the corresponding hypergeometric function is defined in the whole complex plane with a simple pole at $z = 1$. The logarithmic function $-\log(1 - z)$ has a hypergeometric series in the open unit disk, but it can be continued analytically to the complex plane with a cut along $[1, \infty)$ and a branch point at $z = 1$.

### Gauss Hypergeometric Function

The most famous hypergeometric function is the Gauss hypergeometric function defined for $|z| < 1$ by the hypergeometric series

$$ {}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n n!} z^n \qquad [12] $$

which is often denoted by $F(a, b; c; z)$. It is a solution of the hypergeometric equation

$$ z(1 - z)y''(z) + [c - (a + b + 1)z]y'(z) - aby(z) = 0 \qquad [13] $$

and this solution is regular at $z = 0$. Obviously, ${}_2F_1(a, b; \; c; z) = {}_2F_1(b, a; c; z)$. The six functions ${}_2F_1(a \pm 1, b; \; c; z)$, ${}_2F_1(a, b \pm 1; c; z)$, and ${}_2F_1(a, b; c \pm 1; z)$ are called contiguous to ${}_2F_1(a, b; c; z)$ and there are 15 linear relations (with coefficients which are linear functions of $z$) between ${}_2F_1(a, b; c; z)$ and any two contiguous functions. Two of these relations are

$$ (2a - c - az + bz)F(a, b; c; z) + (c - a)F(a - 1, b; c; z) + a(z - 1)F(a + 1, v; c; z) = 0 $$

and

$$ c(a - (c - b)z)F(a, b; c; z) - ac(1 - z)F(a + 1, b; c; z) + (c - a)(c - b)zF(a, b; c + 1; z) = 0 $$

Euler gave the integral representation

$$ {}_2F_1(a, b; c; z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c - b)} \int_0^1 \frac{x^{b-1}(1 - x)^{c-b-1}}{(1 - zx)^a} \mathrm{d}x \qquad [14] $$

for $\Re c > 0$ and $\Re b > 0$. This allows to find the analytic continuation from the open unit disk to the complex plane. A useful result is the Gauss summation formula

$$ {}_2F_1(a, b; c; 1) = \frac{\Gamma(c)\Gamma(c - a - b)}{\Gamma(c - a)\Gamma(c - b)} \quad \Re(c - a - b) > 0 $$

The special case for a terminating series is known as the Chu–Vandermonde sum

$$ {}_2F_1(-n, a; c; 1) = \frac{(c - a)_n}{(c)_n} $$

Pfaff's transformation is

$$ {}_2F_1(a, b; c; z) = (1 - z)^{-a} {}_2F_1\left(a, c - b; c; \frac{z}{z - 1}\right) $$

and Euler's transformation is

$$ {}_2F_1(a, b; c; z) = (1 - z)^{c-a-b} {}_2F_1(c - a, c - b; c; z) $$

### Confluent Hypergeometric Function

The hypergeometric series ${}_1F_1(a; c; z)$ defines an entire function in the complex plane and satisfies the differential equation

$$ zy''(z) + (c - z)y'(z) - ay(z) = 0 \qquad [15] $$

This hypergeometric series (and the differential equation) are formally obtained from ${}_2F_1(a, b; c; z/b)$ by letting $b \to \infty$, which gives a confluence of two of the singularities at $z = \infty$. This is the reason why the differential equation [15] is known as the confluent hypergeometric equation. The solution

$$ \Phi(a, c; z) = {}_1F_1(a; c; z) \qquad [16] $$

is called a confluent hypergeometric function, and a second linearly independent solution of [15] is $z^{1-c}\Phi(c - a + 1, 2 - c; z)$. The function

$$ \Psi(a, c; z) = \frac{\Gamma(1 - c)}{\Gamma(a - c + 1)} \Phi(a, c; z) + \frac{\Gamma(c - 1)}{\Gamma(a)} z^{1-c} \Phi(a - c + 1, 2 - c; z) \qquad [17] $$

is therefore also a solution of eqn [15]. The following integral representations hold:

$$\Phi(a,c;z)=\frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)}\int_0^1 \mathrm{e}^{zx}x^{a-1}(1-x)^{c-a-1}\mathrm{d}x$$

whenever $\Re c > \Re a > 0$, and

$$\Psi(a,c;z)=\frac{1}{\Gamma(a)}\int_0^\infty \mathrm{e}^{-zx}x^{a-1}(1+x)^{c-a-1}\mathrm{d}x$$

whenever $\Re a > 0$.

The "Whittaker functions" are defined as

$$M_{\lambda,\mu}(z)=\mathrm{e}^{-z/2}z^{c/2}\Phi(a,c;z)$$
$$W_{\lambda,\mu}(z)=\mathrm{e}^{-z/2}z^{c/2}\Psi(a,c;z)$$

with $\lambda=-a+c/2$ and $\mu=(c-1)/2$. They are a solution of the Whittaker equation

$$y''(z)+\left(-\frac{1}{4}+\frac{\lambda}{z}+\frac{1-4\mu^2}{4z^2}\right)y(z)=0$$

The "parabolic cylinder functions" are also confluent hypergeometric functions. They are given by

$$\begin{aligned}D_\nu(z)&=2^{\nu/2}\mathrm{e}^{-z^2/4}\Psi(-\nu/2,1/2;z^2/2)\\&=2^{(\nu-1)/2}\mathrm{e}^{-z^2/4}z\Psi((1-\nu)/2,3/2;z^2/2)\end{aligned}$$

When $\nu$ is a non-negative integer, one finds Hermite polynomials

$$H_n(z)=2^{n/2}\mathrm{e}^{z^2/2}D_n(\sqrt{2}z)$$

## Classical Orthogonal Polynomials

A family of polynomials $\{p_n(x), n\in\mathbb{N}\}$, where $p_n$ has degree $n$, is orthogonal on the real line if there is a positive measure $\mu$ on the real line for which

$$\int_{\mathbb{R}} p_n(x)p_m(x)\mathrm{d}\mu(x)=h_n\delta_{m,n} \qquad [18]$$

Usually the measure $\mu$ is absolutely continuous, in which case $\mathrm{d}\mu(x)=w(x)\,\mathrm{d}x$ with $w$ a non-negative density function on the real line, or $\mu$ is discrete and supported on a finite or at most countable set. Any family of orthogonal polynomials satisfies a "three-term recurrence relation"

$$xp_n(x)=A_np_{n+1}(x)+B_np_n(x)+C_np_{n-1}(x) \qquad [19]$$

with $C_nA_{n-1}>0$ for every $n\geq 1$. For the monic polynomials $P_n(x)=p_n(x)/k_n$, with $k_n=1/(A_0A_1A_2\cdots A_{n-1})$ this relation becomes

$$P_{n+1}(x)=(x-b_n)P_n(x)-a_n^2P_{n-1}(x)$$

with $b_n=B_n$ and $a_n^2=A_{n-1}C_n$. This recurrence relation gives rise to a tridiagonal matrix

$$J=\begin{pmatrix} b_0 & a_1 & 0 & 0 & 0 & 0\\ a_1 & b_1 & a_2 & 0 & 0 & 0\\ 0 & a_2 & b_2 & a_3 & 0 & 0\\ 0 & 0 & a_3 & b_3 & a_4 & 0\\ 0 & 0 & 0 & a_4 & \ddots & \ddots\\ 0 & 0 & 0 & 0 & \ddots & \end{pmatrix}$$

which is formally symmetric and which is called the "Jacobi matrix." The spectral measure of this operator, acting on $\ell_2(\mathbb{N})$, is equal to the orthogonality measure $\mu$ whenever this symmetric operator can be extended to a self-adjoint operator. If this is not possible in a unique way – a situation which can occur for unbounded operators only – then every self-adjoint extension of $J$ gives rise to a spectral measure which can be used for the orthogonality conditions [18]. In this case, there are infinitely many positive measures which can be used in the orthogonality relations and all these measures have the same moments

$$m_n=\int_{\mathbb{R}} x^n\,\mathrm{d}\mu(x)$$

Some families of orthogonal polynomials have additional properties which are quite useful in many practical and physical applications, such as the following:

- The derivatives $p_n'$ are again a family of orthogonal polynomials (Hahn property).
- The polynomials $p_n$ satisfy a second-order linear differential equation of the form

  $$\sigma(x)y''(x)+\tau y'(x)=\lambda_ny(x)$$

  where $\sigma$ is a polynomial of degree at most 2, $\tau$ is a polynomial of degree 1, both independent of $n$, and $\lambda_n$ is a real number (Bochner property).
- The polynomials can be obtained by a Rodrigues formula

  $$w(x)p_n(x)=C_n\frac{\mathrm{d}^n}{\mathrm{d}x^n}(w(x)\sigma^n(x))$$

  where $w$ is a non-negative function and $\sigma$ a polynomial of degree at most 2 (Hildebrand property).

There are three families of orthogonal polynomials on the real line which have these three properties, and

each of these three properties characterizes these families. These are the Hermite polynomials, the Laguerre polynomials, and the Jacobi polynomials. In a more general situation when the orthogonality relation is described by a linear functional and the functional is not required to be positive, one has an additional family of Bessel polynomials. The densities $w(x)$ for these families all satisfy a first-order differential equation $[\sigma(x)w(x)]' = \tau(x)w(x)$, where $\sigma$ is a polynomial of degree at most 2 and $\tau$ a polynomial of degree 1. This equation is known as the "Pearson equation."

### Hermite Polynomials

Hermite polynomials $H_n(x)$ are orthogonal with respect to the normal density $w(x) = e^{-x^2}$:

$$\int_{-\infty}^{\infty} H_n(x)H_m(x)e^{-x^2}\,dx = 2^n n!\delta_{n,m}$$

Observe that the density satisfies $w' = -2xw$ so that $\sigma = 1$ and $\tau(x) = -2x$. The recurrence relation is

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x)$$

and the polynomials satisfy the second-order differential equation

$$y''(x) - 2xy'(x) + 2ny(x) = 0$$

The functions $h_n(x) = e^{-x^2/2}H_n(x)$ satisfy the differential equation

$$h_n''(x) + (2n + 1 - x^2)h_n(x) = 0$$

The derivatives satisfy $H_n'(x) = 2nH_{n-1}(x)$ (lowering operation) and one also has $[e^{-x^2}H_n(x)]' = -e^{-x^2}H_{n+1}(x)$ (raising operation). The Rodrigues formula is

$$e^{-x^2}H_n(x) = (-1)^n \frac{d^n}{dx^n} e^{-x^2}$$

The polynomials can be written as a hypergeometric series

$$H_n(x) = (2x)^n {}_2F_0(-n/2, -(n-1)/2; -; -1/x^2)$$

or alternatively as

$$H_n(x) = n! \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{(-1)^k (2x)^{n-2k}}{k!(n-2k)!}$$

Their generating function is

$$\sum_{n=0}^{\infty} H_n(x)\frac{t^n}{n!} = \exp(2xt - t^2)$$

Hermite polynomials are relevant for the analysis of the quantum harmonic oscillator, and the lowering and raising operators there correspond to creation and annihilation.

### Laguerre Polynomials

Laguerre polynomials $L_n^\alpha(x)$ are for $\alpha > -1$ orthogonal with respect to the gamma density $w(x) = x^\alpha e^{-x}$ on $[0, \infty)$:

$$\int_0^{\infty} L_n^\alpha(x)L_m^\alpha(x)x^\alpha e^{-x}\,dx = \frac{\Gamma(n+\alpha)}{n!}\delta_{m,n}$$

The Pearson equation is $[xw]' = (\alpha + 1 - x)w$ so that $\sigma(x) = x$ and $\tau(x) = \alpha + 1 - x$. The recurrence relation is

$$\begin{aligned}(n+1)L_{n+1}^\alpha(x) \\ &= (2n + \alpha + 1 - x)L_n^\alpha(x) - (n+\alpha)L_{n-1}^\alpha(x)\end{aligned}$$

and the differential equation is

$$xy''(x) + (\alpha + 1 - x)y'(x) + ny(x) = 0$$

The functions $\ell_n(x) = x^{\alpha/2}e^{-x/2}L_n^\alpha(x)$ satisfy

$$(x\ell_n')' + \left(n + \frac{\alpha+1}{2} - \frac{x}{4} - \frac{\alpha^2}{4x}\right)\ell_n = 0$$

Differentiation has the effect that

$$\left[L_n^\alpha(x)\right]' = -L_{n-1}^{\alpha+1}(x)$$

and

$$\left[x^\alpha e^{-x}L_n^\alpha(x)\right]' = (n+1)x^{\alpha-1}e^{-x}L_{n+1}^{\alpha-1}(x)$$

The Rodrigues formula is

$$x^\alpha e^{-x}L_n^\alpha(x) = \frac{1}{n!}\frac{d^n}{dx^n}[x^{n+\alpha}\,e^{-x}]$$

The hypergeometric expression is

$$n!L_n^\alpha(x) = (\alpha+1)_n {}_1F_1(-n; \alpha+1; x)$$

and the generating function is

$$\sum_{n=0}^{\infty} L_n^\alpha(x)t^n = (1-t)^{-\alpha-1}\exp\left(\frac{xt}{t-1}\right)$$

Laguerre polynomials occur as eigenfunctions of the hydrogen atom.

### Jacobi Polynomials

Jacobi polynomials $P_n^{(\alpha,\beta)}(x)$ are orthogonal for the beta density $w(x) = (1-x)^\alpha(1+x)^\beta$ on $[-1,1]$ whenever $\alpha > -1$ and $\beta > -1$:

$$\int_{-1}^{1} P_n^{(\alpha,\beta)}(x)P_m^{(\alpha,\beta)}(x)(1-x)^\alpha(1+x)^\beta\,\mathrm{d}x = \frac{2^{\alpha+\beta+1}}{2n+\alpha+\beta+1}\frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{\Gamma(n+\alpha+\beta+1)}\delta_{n,m}$$

The Pearson equation is $[(1-x^2)w]' = [\beta-\alpha-(\alpha+\beta+2)x]w$ and the differential equation is

$$(1-x^2)y''(x) + [\beta-\alpha-(\alpha+\beta+2)x]y'(x) + n(n+\alpha+\beta+1)y(x) = 0$$

Differentiation has the effect

$$\left[P_n^{(\alpha,\beta)}(x)\right]' = (n+\alpha+\beta+1)/2P_{n-1}^{(\alpha+1,\beta+1)}(x)$$

and

$$\left[(1-x)^\alpha(1+x)^\beta P_n^{(\alpha,\beta)}(x)\right]' = -2(n+1)(1-x)^{\alpha-1}(1+x)^{\beta-1}P_{n+1}^{(\alpha-1,\beta-1)}(x)$$

The Rodrigues formula is

$$(1-x)^\alpha(1+x)^\beta P_n^{(\alpha,\beta)}(x) = \frac{(-1)^n}{2^n n!}\frac{\mathrm{d}^n}{\mathrm{d}x^n}\left[(1-x)^{n+\alpha}(1+x)^{n+\beta}\right]$$

In terms of hypergeometric series, one has

$$P_n^{(\alpha,\beta)}(x) = \frac{(\alpha+1)_n}{n!} \times {}_2F_1\left(\begin{matrix}-n, n+\alpha+\beta+1\\ \alpha+1\end{matrix}\,\middle|\,\frac{1-x}{2}\right)$$

Observe that one has $P_n^{(\alpha,\beta)}(-x) = (-1)^n P_n^{(\beta,\alpha)}(x)$.

Special cases of the Jacobi polynomials are as follows:

- The "Legendre polynomials" $P_n(x) = P_n^{(0,0)}(x)$. They appear when the Laplacian is separated in spherical coordinates as functions of the polar angle $\theta$, for which $x = \cos\theta$.
- The "Chebyshev polynomials" of the first kind

  $$T_n(x) = P_n^{(-1/2,-1/2)}(x)/P_n^{(-1/2,-1/2)}(1)$$

  and of the second kind

  $$U_n(x) = (n+1)P_n^{(1/2,1/2)}(x)/P_n^{(1/2,1/2)}(1)$$

  These functions are more easily written by using the change of variable $x = \cos\theta$ and then $T_n(\cos\theta) = \cos n\theta$ and $U_n(\cos\theta) = \sin(n+1)\theta/\sin\theta$.
- The "Gegenbauer polynomials" or ultraspherical polynomials are Jacobi polynomials with equal parameters:

  $$C_n^\lambda(x) = (2\lambda)_n/(\lambda+1/2)_n P_n^{(\lambda-1/2,\lambda-1/2)}(x)$$

Gegenbauer polynomials are involved in the angular or spatial part of the wave function of physical systems in a central potential in both position and momentum space, and in the spatial part of the wave function of hydrogenic systems in momentum space, as well as in the eigenfunctions of several quantum-mechanical potentials, such as the relativistic harmonic oscillator.

### Other Classical Orthogonal Polynomials

Instead of restricting attention to the differential operator $D = \mathrm{d}/\mathrm{d}x$, one can also use the (forward) difference operator $\Delta$ for which $\Delta f(x) = f(x+1) - f(x)$, the divided difference operator $\Delta_\lambda$ for which $\Delta_\lambda f(x) = \Delta f(\lambda(x))/\Delta\lambda(x)$ with a quadratic function $\lambda$, or certain $q$-difference operators and look for orthogonal polynomials that satisfy difference equations in the variable $x$. Together with the three-term recurrence relation (in the degree $n$), one then has families of polynomials satisfying a bispectral problem. For the difference operator and the divided difference operator, this gives several important families of orthogonal polynomials which all have a hypergeometric representation. These hypergeometric polynomials are usually listed in a table, and each level indicates the number of parameters and/or the order of the hypergeometric function. This table is known as Askey's table and is given in Figure 1. The extension with $q$-difference operators involves basic hypergeometric series and $q$-extensions of classical orthogonal polynomials.

"Charlier polynomials" $C_n(x;a)$ are orthogonal with respect to the Poisson distribution

$$\sum_{k=0}^{\infty} C_n(k;a)C_m(k;a)\frac{a^k}{k!} = \mathrm{e}^a/a^n\delta_{n,m}$$

The recurrence relation is

$$aC_{n+1}(x;a) + (x-n-a)C_n(x;a) + nC_{n-1}(x;a) = 0$$

and the second-order difference equation is

$$ay(x+1) + (n-x-a)y(x) + xy(x-1) = 0$$

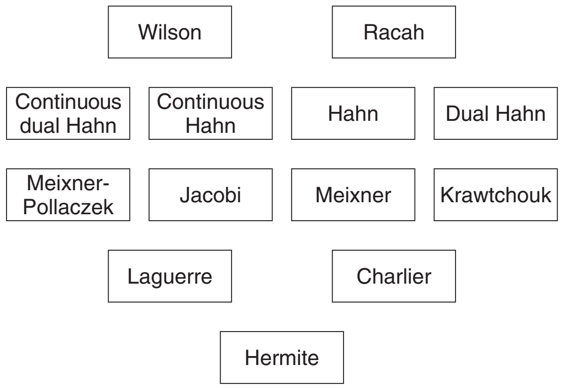



**Figure 1** Askey's table.

The forward difference operator has the effect $\Delta C_n(x;a) = -n/aC_{n-1}(x;a)$ and the backward difference operator $\nabla f(x) = f(x) - f(x-1)$ has the effect $\nabla[a^x/x!C_n(x;a)] = a^x/x!C_{n+1}(x;a)$. The hypergeometric representation is $C_n(x;a) = {}_2F_0(-n-x;\ -;\ -1/a)$. Observe that the variable $x$ appears as a parameter of the hypergeometric series.

"Krawtchouk polynomials" $K_n(x;p,N)$ are orthogonal with respect to the binomial distribution:

$$\sum_{k=0}^{N} K_n(k;p,N)K_m(k;p,N)\binom{N}{k}p^k(1-p)^{N-k} = \frac{(-1)^n n!}{(-N)_n}p^{-n}(1-p)^n\delta_{n,m}$$

where $N$ is a positive integer and $0 < p < 1$. They are given by $K_n(x;p,N) = {}_2F_1(-n,\ -x;\ -N;1/p)$ and correspond to Meixner polynomials for which the parameter $\beta$ is a negative integer.

"Meixner polynomials" $m_n(x;\beta,c)$ are orthogonal with respect to the negative binomial distribution (Pascal distribution)

$$\sum_{k=0}^{\infty} m_n(k;\beta,c)m_j(k;\beta,c)\frac{(\beta)_k c^k}{k!} = \frac{n!}{c^n(\beta)_n(1-c)^\beta}\delta_{n,j}$$

where $\beta > 0$ and $0 < c < 1$. They are given by $m_n(x;\beta,c) = {}_2F_1(-n,\ -x;\beta;1-1/c)$.

"Meixner–Pollaczek polynomials" $P_n^\lambda(x;\phi)$ are orthogonal on $(-\infty,\infty)$:

$$\int_{-\infty}^{\infty} P_m^\lambda(x;\phi)P_n^\lambda(x;\phi)e^{(2\phi-\pi)x}|\Gamma(\lambda+ix)|^2\,dx = \frac{2\pi\Gamma(n+2\lambda)}{(2\sin\phi)^{2\lambda}n!}\delta_{m,n}$$

where $\lambda > 0$ and $0 < \phi < \pi$. The appropriate difference operator $\delta$ has an imaginary shift $\delta f(x) = f(x+i/2) - f(x-i/2)$ and one has $\delta P_n^\lambda(x;\phi) = 2\sin\phi P_{n-1}^{\lambda+1/2}(x;\phi)$. They are given by

$$P_n^\lambda(x;\phi) = \frac{(2\lambda)_n}{n!}e^{in\phi}{}_2F_1\left(\begin{matrix}-n,\lambda+ix\\2\lambda\end{matrix}\middle|1-e^{-2i\phi}\right)$$

"Hahn and dual Hahn polynomials" are orthogonal on a finite set of points. Hahn polynomials are given by

$$Q_n(x;\alpha,\beta,N) = {}_3F_2\left(\begin{matrix}-n,n+\alpha+\beta+1,-x\\\alpha+1,-N\end{matrix}\middle|1\right)$$

and their orthogonality is with respect to a hypergeometric distribution on $\{0,1,\ldots,N\}$. The appropriate difference operator is the (forward) difference operator $\Delta$. They are related to the $3-j$ symbols or Wigner coefficients that arise when considering angular momenta in two quantum systems. Dual Hahn polynomials are given by

$$R_n(\lambda(x);\gamma,\delta,N) = {}_3F_2\left(\begin{matrix}-n,-x,x+\gamma+\delta+1\\\gamma+1,-N\end{matrix}\middle|1\right)$$

where $\lambda(x) = x(x+\gamma+\delta+1)$. They are obtained from the Hahn polynomials by interchanging the roles of $n$ and $x$. They are orthogonal on the set $\{\lambda(0),\lambda(1),\ldots,\lambda(N)\}$. The appropriate difference operator is the divided difference operator which acts on $f$ as $\Delta f(\lambda(x))/\Delta\lambda(x)$.

"Continuous Hahn and dual Hahn polynomials" are orthogonal on the real line. The continuous Hahn polynomials are

$$p_n(x;a;b;c;d) = i^n\frac{(a+c)_n(a+d)_n}{n!} \times {}_3F_2\left(\begin{matrix}-n,n+a+b+c+d-1,a+ix\\a+c,a+d\end{matrix}\middle|1\right)$$

and the appropriate difference operator is the difference operator $\delta$ with imaginary shift. The continuous dual Hahn polynomials are

$$S_n(x^2;a,b,c) = (a+b)_n(a+c)_n \times {}_3F_2\left(\begin{matrix}-n,a+ix,a-ix\\a+b,a+c\end{matrix}\middle|1\right)$$

and the appropriate difference operator is the divided difference operator which acts on $f$ as $\delta f(x^2)/\delta x^2$.

"Wilson polynomials" are the most general system of hypergeometric polynomials satisfying a bispectral problem. All the other classical orthogonal polynomials can be obtained from them by taking

appropriate parameters or as limiting cases. They are given by

$$\frac{W_n(x^2;a,b,c,d)}{(a+b)_n(a+c)_n(a+d)_n} = {}_4F_3\left(\begin{matrix}-n, n+a+b+c+d-1, a+\mathrm{i}x, a-\mathrm{i}x\\ a+b, a+c, a+d\end{matrix}\middle| 1\right)$$

and for $\Re(a,b,c,d) > 0$ (with nonreal parts appearing in conjugate pairs) they are orthogonal on the positive real line with respect to the weight function

$$w(x) = \left|\frac{\Gamma(a+\mathrm{i}x)\Gamma(b+\mathrm{i}x)\Gamma(c+\mathrm{i}x)\Gamma(d+\mathrm{i}x)}{\Gamma(2\mathrm{i}x)}\right|$$

"Racah polynomials" can be obtained from Wilson polynomials when the parameters are such that one of $a+b$, $a+c$, or $a+d$ is a negative integer $-N$. They are given by

$$R_n(\lambda(x);\alpha,\beta,\gamma,\delta) = {}_4F_1\left(\begin{matrix}-n, n+\alpha+\beta+1, -x, x+\gamma+\delta+1\\ \alpha+1, \beta+\delta+1, \gamma+1\end{matrix}\middle| 1\right)$$

where $\alpha+1=-N$ or $\beta+\delta+1=-N$ or $\gamma+1=-N$, and $N$ is a non-negative integer. They are orthogonal on the finite set $\{\lambda(0), \lambda(1), \ldots, \lambda(N)\}$, where $\lambda(x) = x(x+\gamma+\delta+1)$. They arise as $6-j$ symbols in the coupling of three angular momenta.

*See also:* Combinatorics: Overview; Compact Groups and their Representations; Integrable Systems: Overview; Painlevé Equations; q-Special Functions; Random Matrix Theory in Physics; Separation of Variables for Differential Equations.

## Further Reading


Abramowitz M and Stegun IA (1964) *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards Applied Mathematics Series, vol. 55 (reprinted 1984). New York: Dover.

Andrews GE, Askey R, and Roy R (1999) *Special Functions*, Encyclopedia of Mathematics and Its Applications, vol. 71. Cambridge: Cambridge University Press.

Bailey WN (1935) *Generalized Hypergeometric Series*, Cambridge Mathematical Tract, vol. 32. Cambridge: Cambridge University Press.

Erdélyi A, Magnus W, Oberhettinger F, and Tricomi FG (1953–1955) *Higher Transcendental Functions*, Bateman Manuscript Project, vols. 1–3. New York: McGraw-Hill.

Gradshteyn IS and Ryzhik IM (1965) *Table of Integrals, Series, and Products*, chs 8–9, pp. 904–1080. New York: Academic Press.

Koekoek R and Swarttouw R (1998) *The Askey-Scheme of Hypergeometric Orthogonal Polynomials and Its q-Analogue*. Reports of the faculty of Technical Mathematics and Informatics, no. 98–17, Delft University of Technology.

Lozier D, Olver F, Clark C, and Boisvert R *Digital Library of Mathematical Functions*, http://dlmf.nist.gov.

Nikiforov AF and Uvarov VB (1988) *Special Functions of Mathematics Physics*. Basel: Birkhäuser.

Nikiforov AF, Suslov SK, and Uvarov VB (1991) *Classical Orthogonal Polynomials of a Discrete Variable*, Springer Series in Computational Physics. Berlin: Springer.

Szegö G (1939) *Orthogonal Polynomials*, American Mathematical Society Colloquium Publications XXIII, 4th edn., 1975. Providence, RI: American Mathematical Society.

Watson GN (1922) *A Treatise on the Theory of Bessel Functions*. Cambridge: Cambridge University Press.

# P

## Painlevé Equations

**N Joshi**, University of Sydney, Sydney, NSW, Australia



### Introduction

The Painlevé equations $P_I$–$P_{VI}$ are six classical second-order ordinary differential equations that appear widely in modern physical applications. Their conventional forms (governing $y(x)$ with derivatives $y' = dy/dx$, $y'' = d^2y/dx^2$) are:

$$P_I: \quad y'' = 6y^2 + x$$

$$P_{II}: \quad y'' = 2y^3 + xy + \alpha$$

$$P_{III}: \quad y'' = \frac{y'^2}{y} - \frac{y'}{x} + \frac{1}{x}(\alpha y^2 + \beta) + \gamma y^3 + \frac{\delta}{y}$$

$$P_{IV}: \quad y'' = \frac{y'^2}{2y} + \frac{3}{2}y^3 + 4xy^2 + 2(x^2 - \alpha)y + \frac{\beta}{y}$$

$$P_V: \quad y'' = \left(\frac{1}{2y} + \frac{1}{y-1}\right)y'^2 - \frac{y'}{x} + \frac{(y-1)^2}{x^2}\left(\alpha y + \frac{\beta}{y}\right) + \frac{\gamma y}{x} + \frac{\delta y(y+1)}{y-1}$$

$$P_{VI}: \quad y'' = \frac{1}{2}\left(\frac{1}{y} + \frac{1}{y-1} + \frac{1}{y-x}\right)y'^2 - \left(\frac{1}{x} + \frac{1}{x-1} + \frac{1}{y-x}\right)y' + \frac{y(y-1)(y-x)}{x^2(x-1)^2}\left\{\alpha + \frac{\beta x}{y^2} + \frac{\gamma(x-1)}{(y-1)^2} + \frac{\delta x(x-1)}{(y-x)^2}\right\}$$

where $\alpha$, $\beta$, $\gamma$, $\delta$ are constants. They were identified and studied by Painlevé and his school in their search for ordinary differential equations (in the class $y'' = R(x, y, y')$, where $R$ is rational in $y'$, $y$ and analytic in $x$) that define new transcendental functions. Painlevé focussed his search on equations that possess what is now known as the Painlevé property: that all solutions are single-valued around all movable singularities (a singularity is "movable" if its location changes with initial conditions).

For the Painlevé equations, all movable singularities are poles. For $P_I$ and $P_{II}$, all solutions are meromorphic functions. However, the solutions of each of the remaining equations have other singularities called "fixed" singularities, with locations that are determined by the singularities of the coefficient functions of the equation. $P_{III}$–$P_{VI}$ have a fixed singularity at $x = \infty$. $P_{III}$ and $P_V$ have additional fixed singularities at $x = 0$, and $P_{VI}$ has them at $x = 0$ and 1. Although each solution of $P_{III}$–$P_{VI}$ is single-valued around a movable singularity, it may be multivalued around a fixed singularity.

Painlevé's school considered canonical classes of ordinary differential equations equivalent under linear fractional transformations of $y$ and $x$. Of the fifty canonical classes of equations they found, all except six were found to be solvable in terms of already known functions. These six lead to the Painlevé equations $P_I$–$P_{VI}$ as their canonical representatives.

A resurgence of interest in the Painlevé equations came about from the observation (due to Ablowitz and Segur) that they arise as similarity reductions of well-known integrable partial differential equations (PDEs), or soliton equations, such as the Korteweg–de Vries equation, the sine-Gordon equation, and the self-dual Yang–Mills equations.

As this connection suggests, the Painlevé equations possess many of the special properties that are commonly associated with soliton equations. They have associated linear problems (i.e., Lax pairs) for which they act as compatibility conditions. There exist special transformations (called Bäcklund transformations) mapping a solution of one equation to a solution of another Painlevé equation (or the same equation with changed parameters). There exist Hamiltonian forms that are related to existence of tau-functions, that are analytic everywhere except at the fixed singularities. They also possess multilinear forms (or Hirota forms) that are satisfied by tau-functions. In the following subsections, for conciseness, we give examples of these properties for the first or second Painlevé equations and briefly indicate differences, in any, with other Painlevé equations.

## Complex Analytic Structure of Solutions

Consider the two-(complex-)parameter manifold of solutions of a Painlevé equation. Each solution is globally determined by two initial values given at a regular point of the solution. However, the solution can also be determined by two pieces of data given at a movable pole. The location $x_0$ of such a pole provides one of the two free parameters. The other free parameter occurs as a coefficient in the Laurent expansion of the solution in a domain punctured at $x_0$. For $P_I$, the Laurent expansion of a solution at a movable singularity $x_0$ is

$$y(x) = \frac{1}{(x-x_0)^2} + \frac{x_0}{10}(x-x_0)^2 + \frac{1}{6}(x-x_0)^3 + c_I(x-x_0)^4 + \cdots \quad [1]$$

where $c_I$ is arbitrary. This second free parameter is normally called a "resonance parameter." For $P_{II}$, the Laurent expansion of a solution at a movable singularity $x_0$ is

$$y(x) = \frac{\pm 1}{(x-x_0)} + \frac{\mp x_0}{6}(x-x_0) + \frac{\mp 1 - \alpha}{4}(x-x_0)^2 + c_{II}(x-x_0)^3 + \cdots \quad [2]$$

where $c_{II}$ is arbitrary. The symmetric solution of $P_I$ that has a pole at the origin and corresponding resonance parameter $c_I = 0$ has a distribution of poles in the complex $x$-plane shown in Figure 1. (This figure was obtained by searching for zeros of truncated Taylor expansions of the tau-function $\tau_I$ described in the section "Bäcklund and Miura transformations." One hundred and sixty numerical zeros are shown. The two pairs of closely spaced zeros near the imaginary axis (between $8 < \pm\Im x < 12$) may be numerical artifacts. We used the command NSolve to 32 digits in MATHEMATICA4.)

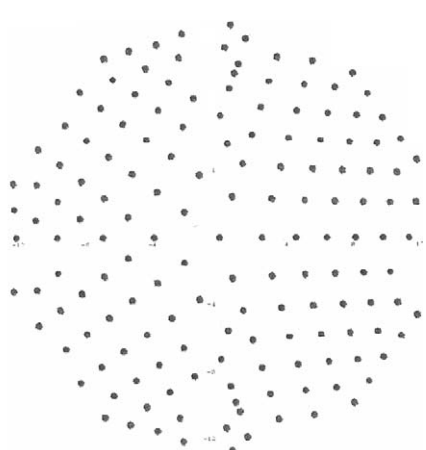


**Figure 1** Poles of a symmetric solution of $P_I$ in the complex $x$-plane, with a pole at the origin and zero corresponding resonance parameter, i.e., $x_0 = 0$, $c_I = 0$.

The rays of symmetry evident in Figure 1 reflect discrete symmetries of $P_I$. The solutions of $P_I$ and $P_{II}$ are invariant under the respective discrete symmetries,

$$P_I: \quad y_n(x) = e^{2\pi i n/5} y(e^{4\pi i n/5} x), \quad n = \pm 1, \pm 2$$

$$P_{II}: \quad y_n(x) = e^{\pi i n/3} y(e^{2\pi i n/3} x), \quad \alpha \mapsto e^{-\pi i n}\alpha \quad n = \pm 1, \pm 2, 3$$

The rays of angle $2\pi n/5$ for $P_I$ and $\pi n/3$ for $P_{II}$ related to these symmetries play special roles in the asymptotic behaviors of the corresponding solutions for $|x| \to \infty$.

## Linear Problems

The Painlevé equations are regarded as completely integrable because they can be solved through an associated system of linear equations (Jimbo and Miwa 1981).

$$\frac{d\varphi}{d\zeta} = L(x,\zeta)\varphi \quad [3a]$$

$$\frac{d\varphi}{dx} = M(x,\zeta)\varphi \quad [3b]$$

The compatibility condition, that is,

$$L_x - M_\zeta + [L, M] = 0 \quad [4]$$

is equivalent to the corresponding Painlevé equation. The matrices $L$, $M$ for $P_I$ and $P_{II}$ are listed below:

$$P_I: \quad L_I(x,\zeta) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\zeta^2 + \begin{pmatrix} 0 & y \\ 4 & 0 \end{pmatrix}\zeta + \begin{pmatrix} -z & y^2 + x/2 \\ -4y & z \end{pmatrix}$$

$$M_I(x,\zeta) = \begin{pmatrix} 0 & 1/2 \\ 0 & 0 \end{pmatrix}\zeta + \begin{pmatrix} 0 & y \\ 2 & 0 \end{pmatrix}$$

$$\text{where } z = y', \quad z' = 6y^2 + x$$

$$P_{II}: \quad L_{II}(x,\zeta) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\zeta^2 + \begin{pmatrix} 0 & u \\ -2z/u & 0 \end{pmatrix}\zeta + \begin{pmatrix} z + x/2 & -uy \\ -2(\vartheta + zy)/u & -(z + x/2) \end{pmatrix}$$

$$M_{II}(x,\zeta) = \begin{pmatrix} 1/2 & 0 \\ 0 & -1/2 \end{pmatrix}\zeta + \begin{pmatrix} 0 & u/2 \\ -z/u & 0 \end{pmatrix}$$

$$\text{where } u' = -uy, \; z = y' - y^2 - x/2$$

$$\vartheta := \frac{1}{2} - \alpha$$

Alternative linear problems also exist for each equation. For example, for $P_{II}$, an alternative choice of $L$ and $M$ is (Flaschka and Newell 1980):

$$P_{II}: \quad L_{II'}(x,\zeta) = \begin{pmatrix} -4\,i & 0 \\ 0 & 4\,i \end{pmatrix}\zeta^2 + \begin{pmatrix} 0 & 4y \\ 4y & 0 \end{pmatrix}\zeta + \begin{pmatrix} -i(x+2y^2) & 2\,iy' \\ -2\,iy' & i(x+2y^2) \end{pmatrix} + \frac{\alpha}{\zeta}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$M_{II'}(x,\zeta) = \begin{pmatrix} -i\zeta & y \\ y & i\zeta \end{pmatrix}$$

The matrix $L$ for each Painlevé equation is singular at a finite number of points $a_i(x)$ in the $\zeta$-plane. For the above choices of $L$ for $P_I$ and $P_{II}$, the point $\zeta = \infty$ is clearly a singularity. For $L_{II'}$, the origin $\zeta = 0$ is also a singularity. The analytic continuation of a fundamental matrix of solutions $\Phi$ around $a_i$ gives a new solution $\widetilde{\Phi}$ which must be related to the original solution: $\widetilde{\Phi} = \Phi A$. $A$ is called the monodromy matrix and its trace and determinant are called the monodromy data. In general, the data will change with $x$. However, eqn [4] ensures that the monodromy data remain constant in $x$. For this reason, the system [3] is called an isomonodromy problem.

## Bäcklund and Miura Transformations

Bäcklund transformations are those that map a solution of a Painlevé equation with one choice of parameter to a solution of the same equation with different parameters. For $P_I$ no such transformation is known. For $P_{II}$, there is one Bäcklund transformation. Let $y = y(x;\alpha)$ denote a solution of $P_{II}$ with parameter $\alpha$. Then $\tilde{y} = y(x;\alpha - 1)$, which solves $P_{II}$ with parameter $\alpha - 1$, is given by

$$\tilde{y} := -y + \frac{\alpha - \frac{1}{2}}{y' - y^2 - x/2} \quad \text{if } \alpha \neq 1/2 \qquad [5]$$

If $\alpha = 1/2$, then $y' = y^2 + x/2$ and $\tilde{y} = -y$ (see the next section for this case). Combined with the symmetry $y \mapsto -y, \alpha = -\alpha$, we can write down another version of this Bäcklund transformation which maps $y$ to $\hat{y} = y(x;\alpha + 1)$:

$$\hat{y} := -y - \frac{\alpha + \frac{1}{2}}{y' + y^2 + x/2}, \quad \text{if } \alpha \neq -\frac{1}{2} \qquad [6]$$

If we parametrize $\alpha$ by $c + n$ for arbitrary $c$, and denote the solution for corresponding parameter as $y_n$, we can write a difference equation relating $y_{n-1}$ and $y_{n+1}$ (by eliminating $y'$ from the two transformations $\tilde{y}$, $\hat{y}$) as

$$\frac{c + \frac{1}{2} + n}{y_{n+1} + y_n} + \frac{c - \frac{1}{2} + n}{y_{n-1} + y_n} + 2y_n^2 + x = 0$$

This is an example of a discrete Painlevé equation (called "alternate" $dP_I$ in the literature). In such a discrete Painlevé equation, $x$ is fixed while $n$ varies. Another lesser known Bäcklund transformation for $P_{II}$ is

$$y' - y^2 - \frac{x}{2} - \delta\, v^2 = 0 \qquad [7]$$

$$v' + y\, v = 0 \qquad [8]$$

between $P_{II}$ with $\alpha = 1/2$ and

$$v'' + \delta\, v^3 + \frac{x}{2}\, v = 0$$

which can be scaled (take $v(x) = y(\sqrt{2}x)/\sqrt{\sqrt{2}\delta}$) to the usual form of $P_{II}$ with $\alpha = 0$.

Miura transformations are those that map a solution of a Painlevé equation to another equation in the 50 canonical types classified by Painlevé's school. If $y$ is a solution of $P_{II}$ with parameter $\alpha \neq 1/2$, then

$$(2\alpha - 1)w = 2(y' - y^2 - x/2), \quad y = \frac{1 - w'}{2w}$$

maps between $P_{II}$ and

$$w'' = \frac{(w')^2}{2\,w} - (2\alpha - 1)\,w^2 - xw - \frac{1}{2\,w}$$

which represents the 34th canonical class in the Painlevé classification listed in Ince (1927).

The Painlevé equations do not possess continuous symmetries other than Bäcklund and Miura transformations described here. However, they do possess discrete symmetries described in the section "Complex analytic structure of solutions."

## Classical Special Solutions

Painlevé showed that there can be no explicit first integral that is rational in $y$ and $y'$ for his eponymous equations. It is known that this statement can be extended to say that no such algebraic first integral exists. But the question whether the Painlevé equations define new transcendental functions remained open until recently.

Form a class of functions consisting of those satisfying linear second-order differential equations, such as the Airy, Bessel, and hypergeometric functions, as well as rational, algebraic, and exponential functions. Extend this class to include arithmetic operations, compositions under such functions, and

solutions of linear equations with these earlier functions as coefficients. Members of this class are called classical functions. For general values of the constants $\alpha, \beta, \gamma, \delta$, it is now known (Umemura 1990, Umemura and Watanabe 1997) that the six Painlevé equations cannot be solved in terms of classical functions. However, there are special values of the constant parameters $\alpha$, $\beta$, $\gamma$, $\delta$ for which classical functions do solve the Painlevé equations. Each Painlevé equation, except $P_I$, has special solutions given by classical functions when the parameters in the Painlevé equation take on special values. For $P_{II}$, with $\alpha = 1/2$ we have the special integral

$$I_{1/2} \equiv y' - y^2 - \frac{x}{2} = 0 \qquad [9]$$

which, modulo $P_{II}$ with $\alpha = 1/2$, satisfies the relation

$$\left(\frac{\mathrm{d}}{\mathrm{d}x} + 2y\right) I_{1/2} = 0$$

The Riccati eqn [9] can be linearized via $y = -\psi'/\psi$ to yield

$$\psi'' + \frac{x}{2}\psi = 0$$

which gives

$$\psi(x) = a\,\mathrm{Ai}(-2^{-1/3}x) + b\,\mathrm{Bi}(-2^{-1/3}\,x)$$

for arbitrary constants $a$ and $b$, that is, the well-known Airy function solutions of $P_{II}$. Iterations of the Bäcklund transformations $\tilde{y}$ and $\hat{y}$, [5]–[6] give further classical solutions in terms of Airy functions for the case when $\alpha = (2N+1)/2$ for integer $N$.

Similarly, there is a sequence of rational solutions of the family of equations $P_{II}$ with $\alpha = N$, for integer $N$, if we iterate the Bäcklund transformations $\tilde{y}, \hat{y}$ by starting with the trivial solution $y \equiv 0$ for the case $\alpha = 0$. For example, for $\alpha = 1$, we have $\hat{y} = -1/x$. The transformations [7]–[8] give a mapping that shows that this family of rational solutions and the above family of Airy-type solutions of $P_{II}$ both exist for the cases when $\alpha$ is half-integer and when it is integer.

## Hamiltonians and Tau-Functions

Each Painlevé equation has a Hamiltonian form. For $P_I$ and $P_{II}$, these can be found by integrating each equation after multiplying by $y'$. These give

$$P_I: \qquad \frac{y'^2}{2} = 2y^3 + xy - \int^x y(\xi)\mathrm{d}\xi + E_I$$

$$P_{II}: \qquad \frac{y'^2}{2} = \frac{y^4}{2} + \frac{x}{2}y^2 - \frac{1}{2}\int^x y(\xi)^2\mathrm{d}\xi + \alpha\, y + E_{II}$$

where $E_{II}$ and $E_{II}$ are constants. We choose canonical variables $q_1(t) = y(x), p_1(t) = y'(x)$, where $t = x$. Furthermore, for $P_I$, we take

$$q_2(t) = x, \qquad p_2(t) = \int^x y(\xi)\mathrm{d}\xi$$

and the Hamiltonian

$$H_I := \frac{p_1^2}{2} - 2{q_1}^3 - q_2 q_1 + p_2$$

so that the Hamiltonian equations of motion $\dot{q}_i = \partial H/\partial p_i$ and $\dot{p}_i = -\partial H/\partial q_i$ are satisfied. For $P_{II}$, we take

$$q_2(t) = x/2, \qquad p_2(t) = \int^x y(\xi)^2\mathrm{d}\xi$$

and the Hamiltonian

$$H_{II} := \frac{{p_1}^2}{2} - \frac{{q_1}^4}{2} - q_2{q_1}^2 + \frac{1}{2}p_2 - \alpha q_1$$

We note that these Hamiltonians govern systems with two degrees of freedom and each is conserved. However, no explicit second conserved quantity is known (see comments on first integrals in the last section).

Painlevé's viewpoint of the transcendental solutions of the Painlevé equations as natural generalizations of elliptic functions also led him to search for entire functions that play the role of theta functions in this new setting. He found that analogous functions could be defined which have only zeros at the locations of the movable singularities of the Painlevé transcendents. These functions are now commonly known as tau-functions (also denoted $\tau$-functions). For $P_I$ and $P_{II}$, the corresponding tau-functions are entire functions (i.e., they are analytic everywhere in the complex $x$-plane). However, for the remaining Painlevé equations, they are singular at the fixed singularities of the respective equation.

For $P_I$, all movable singularities of $P_I$ are double poles of strength unity (see eqn [1]). Therefore, the function given by

$$P_I: \quad \tau_I(x) = \exp\left(-\int^x\int^s y(t)\mathrm{d}t\mathrm{d}s\right)$$

has Taylor expansion with leading term $(x - x_0)$. In other words, $\tau_I(x)$ is analytic at all the poles of the corresponding solution of $P_I$. Since $y(x)$ has no other singularity (other than at infinity), $\tau_I(x)$ must be analytic everywhere in the complex $x$-plane. Differentiation and substitution of $P_I$ shows that $\tau_I(x)$ satisfies the fourth-order equation

$$P_I: \quad \tau_I^{(4)}(x)\tau_I(x) = 4\tau_I'(x)\tau_I^{(3)}(x) - 3\tau_I''(x)^2 - x\tau_I(x)^2$$

Note that this equation is bilinear in $\tau$ and its derivatives. Such bilinear, or in general, multilinear, equations are called Hirota-type forms of the Painlevé equations. The special nature of such equations is most simply expressed in terms of the Hirota $D(\equiv D_x)$ operator, an antisymmetric differential operator defined here on products of functions of $x$:

$$D^n f \cdot g = (\partial_\xi - \partial_\eta)^n f(\xi) g(\eta)|_{\xi=\eta=x}$$

Notice that

$$D^2 \tau \cdot \tau = \tau\tau'' - \tau'^2,$$

$$D^4 \tau \cdot \tau = \tau\tau^{(4)} - 4\tau'\tau^{(3)} + 3\tau''^2$$

Hence the equation satisfied by $\tau_{\rm I}(x)$ can be rewritten more succinctly as

$$(D^4 + x)\tau_{\rm I} \cdot \tau_{\rm I} = 0$$

For $P_{\rm II}$, a generic solution $y(x)$ has movable simple poles of residue $\pm 1$ (see eqn [2]). Painlevé pointed out that if we square the function $y(x)$, multiply by $-1$ and integrate twice, we obtain a function with Taylor expansion with leading term $(x - x_0)$. However, the square is not invertible and to construct an invertible mapping to entire functions, we need two $\tau$-functions. We denote these by $\tau(x)$ and $\sigma(x)$:

$$P_{\rm II}: \qquad \tau_{\rm II}(x) = \exp\left(-\int^x \int^s y(t)^2 {\rm d}t{\rm d}s\right)$$

$$\sigma_{\rm II}(x) = y(x)\tau_{\rm II}(x)$$

The equations satisfied by these tau-functions are

$$P_{\rm II}: \qquad \tau''(x)\tau(x) = \tau'(x)^2 - \sigma(x)^2$$

$$\sigma''(x)\tau(x)^2 = 2\tau(x)\tau'(x)\sigma'(x) - \tau'(x)^2\tau(x) + \sigma(x)^3 + x\tau(x)^2\sigma(x) + \alpha\tau(x)^3$$

## Hierarchies

Each Painlevé equation is associated with at least one infinite sequence of ordinary differential equations (ODEs) indexed by order. These sequences are called hierarchies and arise from symmetry reductions of PDE hierarchies that are associated with soliton equations.

Define the operator $\mathcal{L}_n\{v(z)\}$ (the Lenard recursion operator) recursively by

$$\frac{\rm d}{{\rm d}z}\mathcal{L}_{n+1}\{v\} = \left(\frac{{\rm d}^3}{{\rm d}z^3} + 4v\frac{\rm d}{{\rm d}z} + 2v'\right)\mathcal{L}_n\{v\}$$

$$\mathcal{L}_1\{v\} = v$$

where primes denote $z$-derivatives. Note that

$$\mathcal{L}_2\{v\} = v'' + 3v^2$$

$$\mathcal{L}_3\{v\} = v^{(4)} + 10vv'' + 5v'^2 + 10v^3$$

This operator is intimately related to the Korteweg–de Vries equation. (It was first discovered as a method of generating the infinite number conservation laws associated with this soliton equation.)

The scaling $v(z) = \lambda y(\mu x)$, with $\lambda = (-2)^{1/3}$, $\mu = (-2)^{-1/3}$, shows that the case $n = 2$ of the sequence of ODEs defined recursively by

$$\mathcal{L}_n\{v\} = z$$

is $P_{\rm I}$. Hence this is called the first Painlevé hierarchy.

A second Painlevé hierarchy is given recursively by

$$\left(\frac{\rm d}{{\rm d}x} + 2y\right)\mathcal{L}_n\{y' - y^2\} = xy + \alpha_n, \quad n \geq 1$$

where $\alpha_n$ are constants.

Each Painlevé equation may arise as a reduction of more than one PDE. Since different soliton equations have different hierarchies, this means that more than one hierarchy may be associated with each Painlevé equation.

*See also:* Bäcklund Transformations; Integrable Discrete Systems; Integrable Systems: Overview; Isomonodromic Deformations; Ordinary Special Functions; Riemann–Hilbert Methods in Integrable Systems; Riemann–Hilbert Problem; Solitons and Kac–Moody Lie Algebras; Two-Dimensional Ising Model; WDVV Equations and Frobenius Manifolds.

## Further Reading


Flaschka H and Newell AC (1980) Monodromy- and spectrum-preserving deformations. *Communications in Mathematical Physics* 76: 65–116.

Ince EL (1927) *Ordinary Differential Equations*, London: Longmans, Green and Co. (Reprinted in 1956 – New York: Dover).

Jimbo M and Miwa T (1981) Monodromy preserving deformation of linear ordinary differential equations with rational coefficients. II. *Physica D* 2: 407–448.

Umemura H (1990) Second proof of the irreducibility of the first differential equation of Painlevé. *Nagoya Mathematical Journal* 117: 125–171.

Umemura H and Watanabe H (1997) Solutions of the second and fourth Painlevé equations. I. *Nagoya Mathematical Journal* 148: 151–198.

# Partial Differential Equations: Some Examples

**R Temam**, Indiana University, Bloomington, IN, USA



## Introduction

Many physical laws are mathematically expressed in terms of partial differential equations (PDEs); this is, for instance, the case in the realm of classical mechanics and physics of the laws of conservation of angular momentum, mass, and energy.

The object of this short article is to provide an overview and make a few comments on the set of PDEs appearing in classical mechanics, which is tremendously rich and diverse. From the mathematical point of view the PDEs appearing in mechanics range from well-understood PDEs to equations which are still at the frontier of sciences as far as their mathematical theory is concerned. The mathematical theory of PDEs deals primarily with their "well-posedness" in the sense of Hadamard. A well-posed PDE problem is a problem for which existence and uniqueness of solutions in suitable function spaces and continuous dependence on the data have been proved.

For simplicity, let us restrict ourselves to space dimension 2. Several interesting and important PDEs are of the form

$$a\frac{\partial^2 u}{\partial x^2} + b\frac{\partial^2 u}{\partial x \partial y} + c\frac{\partial^2 u}{\partial y^2} = 0 \qquad [1]$$

Here $a$, $b$, $c$ may depend on $x$ and $y$ or they may be constants, and then eqn [1] is linear: they may also depend on $u$, $\partial u/\partial x$, and $\partial u/\partial y$, in which case the equation is nonlinear.

Such an equation is

- elliptic when (where) $b^2 - 4ac < 0$,
- hyperbolic when (where) $b^2 - 4ac > 0$,
- parabolic when (where) $b^2 - 4ac = 0$.

Among the simplest linear equations, we have the elliptic equation

$$\Delta u = 0 \qquad [2]$$

which governs the following phenomena: equation for the potential or stream function of plane, incompressible irrotational fluids; equation for some potential in linear elasticity, or the equation for the temperature in suitable conditions (stationary case; see below for the time-dependent case).

Another eqn of the form [1] is the hyperbolic equation

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 0 \qquad [3]$$

which governs, for example, linear acoustics in one dimension (sound pipes) or the propagation of an elastic wave along an elastic string.

A third equation of type [1] is the linear parabolic equation

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0 \qquad [4]$$

also called the heat equation, which governs, under appropriate circumstances, the temperature ($u(x,t) =$ temperature at $x$ at time $t$).

All these equations are well understood from the mathematical viewpoint and many well-posedness results are available. A fundamental difference between eqns [2], [3], and [4] is that for [2] and [4] the solution is as smooth as allowed by the data (forcing terms, boundary data not mentioned here), whereas the solutions of [3] usually present some discontinuities corresponding to the propagation of a wave or wave front.

A considerable jump of complexity occurs if we consider the equation of transonic flows in which

$$\begin{aligned} a &= \left(1 - \frac{1}{v^2}\left(\frac{\partial u}{\partial x}\right)^2\right) \\ b &= -\frac{2}{v^2}\frac{\partial u}{\partial x}\frac{\partial u}{\partial y} \\ c &= 1 - \frac{1}{v^2}\left(\frac{\partial u}{\partial y}\right)^2 \end{aligned} \qquad [5]$$

where $v = v(x,y)$ is the local speed of sound. This is a mixed second-order equation: it is elliptic in the subsonic region where $M < 1, M$ the Mach number being the ratio of the velocity

$$|\text{grad } u| = \left(\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2\right)^{1/2}$$

to the local velocity of sound $v = v(x,y)$; eqn [1] (with [5]) is hyperbolic in the supersonic region, where $M > 1$ and parabolic on the sonic line $M = 1$. Essentially no result of well-posedness is available for this problem, and it is not even totally clear what are the boundary conditions that one should associate to [1]–[5] to obtain a well-posed problem.

Intermediate mathematical situations are encountered with the Navier–Stokes and Euler equations, which govern the motion of fluids in the viscous and inviscid cases, respectively. A number of mathematical results are available for these equations (*see* Compressible Flows: Mathematical Theory, Incompressible Euler Equations: Mathematical Theory, Viscous Incompressible Fluids: Mathematical Theory, Inviscid Flows); but other questions are still open, including the famous Clay prize problem, which is: to show that the solutions of the (viscous, incompressible) Navier–Stokes equations, in space dimension three, remain smooth for all time, or to exhibit an example of appearance of singularity. A prize of US$ 1 million will be awarded by the Clay Foundation for the solution of this problem.

For compressible fluids, the Navier–Stokes equations expressing conservation of angular momentum and mass read

$$\rho\left(\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u}\right) - \mu\Delta\boldsymbol{u} + \nabla p - (\lambda+\mu)\nabla(\nabla\cdot\boldsymbol{u}) = 0 \qquad [6]$$

$$\frac{\partial \rho}{\partial t} + \nabla(\rho\boldsymbol{u}) = 0 \qquad [7]$$

Here $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x}, t)$ is the velocity at $\boldsymbol{x}$ at time $t$, $p = p(\boldsymbol{x}, t)$ the pressure, $\rho$ the density; $\lambda, \mu$ are viscosity coefficients, $\mu > 0, 3\lambda + 2\mu \geq 0$. When $\mu = \lambda = 0$, we obtain the Euler equation (*see* Compressible Flows: Mathematical Theory). If the fluid is incompressible and homogeneous, then the density is constant, $\rho = \rho_0$ and

$$\nabla\cdot\boldsymbol{u} = 0 \qquad [8]$$

so that eqn [8] replaces eqn [7] and eqn [6] simplifies accordingly.

Finally, let us mention still different nonlinear PDEs corresponding to nonlinear wave phenomena, namely the Korteweg–de Vries (*see* Korteweg–de Vries Equation and Other Modulation Equations)

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + \frac{\partial^3 u}{\partial x^3} = 0 \qquad [9]$$

and the nonlinear Schrödinger equation (*see* Nonlinear Schrödinger Equations)

$$\frac{\partial u}{\partial z} + i\beta\frac{\partial^2 A}{\partial t^2} - \mathrm{i}\,\gamma|A|^2 A + \alpha A = 0 \qquad [10]$$

$\alpha, \gamma > 0$. These equations are very different from eqns [1]–[8] and are reasonably well understood from the mathematical point of view; they produce and describe the amazing physical wave phenomenon known as the soliton (*see* Solitons and Kac–Moody Lie Algebras).

This article is based on the Appendix of the book by Miranville and Temam quoted below, with the authorization of Cambridge University Press.

*See also:* Compressible Flows: Mathematical Theory; Elliptic Differential Equations: Linear Theory; Evolution Equations: Linear and Nonlinear; Fluid Mechanics: Numerical Methods; Fractal Dimensions in Dynamics; Image Processing: Mathematics; Incompressible Euler Equations: Mathematical Theory; Integrable Systems and the Inverse Scattering Method; Interfaces and Multicomponent Fluids; Inviscid Flows; Korteweg–de Vries Equation and Other Modulation Equations; Leray–Schauder Theory and Mapping Degree; Magnetohydrodynamics; Newtonian Fluids and Thermohydraulics; Nonlinear Schrödinger Equations; Solitons and Kac–Moody Lie Algebras; Stochastic Hydrodynamics; Symmetric Hyperbolic Systems and Shock Waves; Viscous Incompressible Fluids: Mathematical Theory; Non-Newtonian Fluids.

## Further Reading

Brezis H and Browder F (1998) Partial differential equations in the 20th century. *Advances in Mathematics*, 135: 76–144.

Evans LC (1998) *Partial Differential Equations*. Providence, RI: American Mathematical Society.

Miranville A and Temam R (2001) *Mathematical Modelling in Continuum Mechanics*. Cambridge: Cambridge University Press.

**Path Integral Methods** *see* Functional Integration in Quantum Physics; Feynman Path Integrals

# Path Integrals in Noncommutative Geometry

**R Léandre**, Université de Bourgogne, Dijon, France



## Introduction

Let us recall that there are basically two algebraic infinite-dimensional distribution theories:

- The first one is white-noise analysis (Hida *et al.* 1993, Berezansky and Kondratiev 1995), and uses Fock spaces and the algebra of creation and annihilation operators.
- The second one is the noncommutative differential geometry of Connes (1988) and uses the entire cyclic complex.

If we disregard the differential operations, these two distribution theories are very similar. Let us recall quickly their background on geometrical examples. Let $V$ be a compact Riemannian manifold and $E$ a Hermitian bundle on it. We consider an elliptic Laplacian $\Delta_E$ acting on sections $\omega$ of this bundle. We consider the Sobolev space $H_k, k > 0$, of sections $\omega$ of $E$ such that:

$$\int_V \left\langle \left(\Delta_E^k + 1\right)\omega, \omega \right\rangle \mathrm{d}m_V < \infty \qquad [1]$$

where $\mathrm{d}m_V$ is the Riemannian measure on $V$ and $\langle\,,\rangle$ the Hermitian structure on $V$. $H_{k+1}$ is included in $H_k$ and the intersection of all $H_k$ is nothing other than the space of smooth sections of the bundle $E$, by the Sobolev embedding theorem.

Let us quickly recall Connes' distribution theory: let $\alpha(n)$ be a sequence of real strictly positive numbers. Let

$$\sigma = \sum \sigma_n \qquad [2]$$

where $\sigma_n$ belongs to $H_k^{\otimes n}$ with the Hilbert structure naturally inherited from the Hilbert structure of $H_k$. We put, for $C > 0$,

$$\|\sigma\|_{1,C,k} = \sum C^n \alpha(n) \|\sigma_n\|_{H_k^{\otimes n}} \qquad [3]$$

The set of $\sigma$ such that $\|\sigma\|_{1,C,k} < \infty$ is a Banach space called $\mathrm{Co}_{C,k}$. The space of Connes functionals $\mathrm{Co}_{\infty-}$ is the intersection of these Banach spaces for $C > 0$ and $k > 0$ endowed with its natural topology. Its topological dual $\mathrm{Co}_{-\infty}$ is the space of distributions in Connes' sense.

**Remark** We do not give the original version of the space of Connes where tensor products of Banach algebras appear but we use here the presentation of Jones and Léandre (1991).

Let us now quickly recall the theory of distributions in the white-noise sense. The main tools are Fock spaces. We consider interacting Fock spaces (Accardi and Bożejko (1998)) constituted of $\sigma$ written as in [2] such that

$$\|\sigma\|_{2,C,k}^2 = \sum C^n \alpha(n)^2 \|\sigma_n\|_{H_k^{\otimes n}}^2 < \infty \qquad [4]$$

The space of white-noise functionals $\mathrm{WN}_{\infty-}$ is the intersection of these interacting Fock spaces $\Lambda_{k,C}$ for $C > 0, k > 0$. Its topological dual $\mathrm{WN}_{-\infty}$ is called the space of white-noise distributions.

Traditionally, in white-noise analysis, one considers in [2] the case where $\sigma_n$ belongs to the symmetric tensor product of $H_k$ endowed with its natural Hilbert structure. We get a symmetric Fock space $\Lambda^{\mathrm{s}}_{C,k}$ and another space of white-noise distributions $\mathrm{WN}_{\mathrm{s},-\infty}$. The interest in considering symmetric Fock spaces, instead of interacting Fock spaces, arises from the characterization theorem of Potthoff–Streit. For the sake of simplicity, let us consider the case where $\alpha(n) = 1$. If $\omega$ if a smooth section of $E$, we can consider its exponential $\exp[\omega] = \sum n!^{-1}\omega^{\otimes n}$. If we consider an element $\Phi$ of $\mathrm{WN}_{\mathrm{s},-\infty}$, $\langle\Phi, \exp[\omega]\rangle$ satisfies two natural conditions:

1. $|\langle\Phi, \exp[\omega]\rangle| \le C\exp[C\|\omega\|_{H_k}^2]$ for some $k > 0$.
2. $z \to \langle\Phi, \exp[\omega_1 + z\omega_2]\rangle$ is entire.

The Potthoff–Streit theorem states the opposite: a functional which sends a smooth section of $V$ into a Hilbert space and which satisfies the two previous requirements defines an element of $\mathrm{WN}_{\mathrm{s},-\infty}$ with values in this Hilbert space. Moreover, if the functional depends holomorphically on a complex parameter, then the distribution depends holomorphically on this complex parameter as well.

The Potthoff–Streit theorem allows us to define flat Feynman path integrals as distributions. It is the opposite point of view, from the traditional point of view of physicists, where generally path integrals are defined by convergence of the finite-dimensional lattice approximations. Hida–Streit have proposed replacing the approach of physicists by defining path integrals as infinite-dimensional distributions, and by using Wiener chaos. Getzler was the first who thought of replacing Wiener chaos by other functionals on path spaces, that is, Chen iterated integrals. In this article, we review the recent

developments of path integrals in this framework. We will mention the following topics:

- infinite-dimensional volume element
- Feynman path integral on a manifold
- Bismut–Chern character and path integrals
- fermionic Brownian motion

The reader who is interested in various rigorous approaches to path integrals should consult the review of Albeverio (1996).

## Infinite-Dimensional Volume Element

Let us recall that the Lebesgue measure does not exist generally as a measure in infinite dimensions. For instance, the Haar measure on a topological group exists if and only if the topological group is locally compact. Our purpose in this section is to define the Lebesgue measure as a distribution.

We consider the set $C^\infty(M;N)$ of smooth maps $x(.)$ from a compact Riemannian manifold $M$ into a compact Riemannian manifold $N$ endowed with its natural Fréchet topology. $S$ is the generic point of $M$ and $x$ the generic point of $N$. We would like to say that the law of $x(S_i)$ for a finite set of $n$ different points $S_i$ under the formal Lebesgue measure $\mathrm{d}D(x(.))$ on $C^\infty(M;N)$ is the product law of $n\,\mathrm{d}m_N$ (This means that the Lebesgue measure on $C^\infty(M;N)$ is a cylindrical measure). Let us consider a smooth function $\sigma_n$ from $(M\times N)^n$ into $C$. We introduce the associated functional $F(\sigma_n)(x(.))$ on $C^\infty(M;N)$:

$$F(\sigma_n)(x(.)) = \int_{M^n} \sigma_n(S_1,\ldots,S_n,x(S_1),\ldots,x(S_n))\mathrm{d}m_{M^n} \qquad [5]$$

If we use formally the Fubini formula, we get

$$\int_{C^\infty(M;N)} F(\sigma_n)(x(.))\mathrm{d}D(x(.)) = \int_{M^n\times N^n} F(S_1,\ldots,S_n,x_1,\ldots,x_n)\mathrm{d}m_{M^n\times N^n} \qquad [6]$$

We will interpret this formal remark in the framework of the distribution theories of the introduction. We consider $V = M\times N$ and $E$ the trivial complex line bundle endowed with the trivial metric and $\alpha(n)=1$. We can define the associated algebraic spaces $\mathrm{Co}_{-\infty}$ and $\mathrm{WN}_{-\infty}$ and we can extend to $\mathrm{Co}_{\infty-}$ and $\mathrm{WN}_{\infty-}$ the map $F$ of [5]. $F$ sends elements of $\mathrm{Co}_{\infty-}$ and $\mathrm{WN}_{\infty-}$ into the set of continuous bounded maps of $C^\infty(M;N)$ where we can extend [6]. We obtain:

**Theorem 1** $\sigma\to\int_{C^\infty(M;N)}F(\sigma)(x(.))\mathrm{d}D(x(.))$ *defines an element of* $\mathrm{Co}_{-\infty}$ *or* $\mathrm{WN}_{-\infty}$.

## Feynman Path Integral on a Manifold

Let us introduce the flat Brownian motion $s\to B(s)$ in $\boldsymbol{R}^d$ starting from 0. It has formally the Gaussian law

$$Z^{-1}\exp\left[-\frac{1}{2}\int_0^1\left|\frac{\mathrm{d}}{\mathrm{d}s}B(s)\right|^2\mathrm{d}s\right]\mathrm{d}D(B(.))$$

where $\mathrm{d}D(B(.))$ is the formal Lebesgue measure on finite-energy paths starting from 0 in $\boldsymbol{R}^d$ (the partition function $Z$ is infinite!). Let $N$ be a compact Riemannian manifold of dimension $d$ endowed with the Levi–Civita connection. The stochastic parallel transport on semimartingales for the Levi-Civita connection exists almost surely (Ikeda and Watanabe 1981). Let us introduce the Laplace–Beltrami operator $\Delta_N$ on $N$ and the Eells–Elworthy–Malliavin equation starting from $x$ (Ikeda and Watanabe 1981):

$$\mathrm{d}x_s(x) = \tau_s(x)\mathrm{d}B(s) \qquad [7]$$

where $B(.)$ is a Brownian motion in $T_x(M)$ starting from 0 and $s\to\tau_s(x)$ is the stochastic parallel transport associated to the solution. $s\to x_s(x)$ is called the Brownian motion on $N$. The heat semigroup associated to $\Delta_N$ satisfies $\exp[-t\Delta_N]f(x)=E[f(x_t(x))]$ for $f$ continuous on $N$. Formally, there is a Jacobian which appears in the transformation of the formal path integral which governs $B(.)$ into the formal path integral which governs $x_.(x)$

$$\mathrm{d}\mu_x(1) = Z_x^{-1}\exp[-I(x_.(x))/2]\mathrm{d}D(x_.(x)) \qquad [8]$$

It was shown by B DeWitt, in a formal way, that the action in [8] is not the energy of the path and that there are some counter-terms in the action where the scalar curvature $K$ of $N$ appears (see Andersson and Driver (1999) and Sidorova *et al.* (2004) for rigorous results). In order to describe Feynman path integrals, we perform, as it is classical in physics, analytic continuation on the semigroup and on the "measure" $\mathrm{d}\mu_x(1)$ such that we get a distribution $\mathrm{d}\mu_x(\alpha)$ which depends holomorphically on $\alpha$, $\mathrm{Re}\,\alpha\geq 0$.

In order to return to the formalism of the introduction, we consider $V=N$, $E$ the trivial complex line bundle and the symmetric Fock space and $\alpha(n)=1$. To $\sigma_n/n!$ belonging to $H_k^{\otimes\mathrm{sym}\,n}$ we associate the functional on $P(N)$, the smooth path space on $N$:

$$F(\sigma_n/n!)(x(.)) = \int_{\Delta_n}\sigma_n(x(s_1),\ldots,x(s_n))\mathrm{d}s_1\cdots\mathrm{d}s_n \qquad [9]$$

where $\Delta_n$ is the $n$-dimensional simplex of $[0,1]^n$ constituted of times $0<s_1<\cdots<s_n<1$ (Léandre (2003)). We remark that $F$ maps $\mathrm{WN}_{s,\infty-}$ into the set of bounded continuous functionals on $P(N)$. We

introduce an element $h$ of $L^2(N)$. The map which to $\omega$, a smooth function on $N$, associates $\exp[\alpha(\Delta_N + \omega)]h(\mathrm{Re}\,\alpha \le 0)$ satisfies the requirements (1) and (2) of the introduction and depends holomorphically on $\alpha$. This defines by the Potthoff–Streit theorem a distribution $\Phi_\alpha$ which depends holomorphically on $\alpha$, $\mathrm{Re}\,\alpha \le 0$ with values in $L^2(N)$. By uniqueness of analytic continuation, we obtain:

**Theorem 2** *If $P_x(N)$ is the space of smooth paths starting from $x$ in $N$, we have*

$$\langle \Phi_\alpha, \sigma \rangle = \left\{ x \to \int_{P_x(N)} F(\sigma) h(x(1)) \mathrm{d}\mu_x(-\alpha) \right\} \quad [10]$$

Instead of taking functions, we can consider as bundle $E$ the space of complex 1-forms on $N$. We then consider Chen (1973) iterated integrals:

$$F(\sigma_n)(x(.)) = \int_{\Delta_n} \langle \sigma_n(x(s_1), \ldots, x(s_n)), \mathrm{d}x(s_1), \ldots, \mathrm{d}x(s_n) \rangle \quad [11]$$

such that $F$ maps $\mathrm{WN}_{s,\infty-}$ into the set of measurable maps on $P(N)$. These maps are generally not bounded. Namely,

$$F(\exp[\omega]) = \exp\left[\int_0^1 \langle \omega(x(s)), \mathrm{d}x(s) \rangle\right] \quad [12]$$

instead of $\exp[\int_0^1 \omega(x(s))\mathrm{d}s]$ in the previous case. By using the Cameron–Martin–Girsanov–Maruyama formula and Kato perturbation theory, we get an analog of Theorem 2 for Chen iterated integrals, but for $\mathrm{Re}\,\alpha < 0$, because we have to deal with a perturbation of $\Delta_N$ by a drift when we want to check (1) and (2). The interest of this formalism is that the parallel transport belongs in some sense to the domain of the distribution and that we get the flat Feynman path integral from the curved one by using an analog of [7].

## Bismut–Chern Character and Path Integrals

Since we are concerned in this part with index theory, we replace the free path space of $N$ by the free smooth loop space $L(N)$. We consider the case where $V = N$ is a compact oriented Riemannian spin manifold and $E = E_- \oplus E_+$. $E_-$ is the bundle of complexified odd forms and $E_+$ is the bundle of complexified even forms. To $\sigma_n = n!^{-1}(\omega_1 + \omega_1^1) \otimes \cdots \otimes (\omega_n + \omega_n^1)$, we associate the even Chen (1973) iterated integral

$$F(\sigma_n) = \int_{\Delta_n} \left(\omega_1(\mathrm{d}x(s_1), .) + \omega_1^1 \mathrm{d}s_1\right) \wedge \cdots \wedge \left(\omega_n(\mathrm{d}x(s_n), .) + \omega_n^1 \mathrm{d}s_n\right) \quad [13]$$

where $s \to x(s)$ is a smooth loop in $N$, $\omega_i$ is of odd degree and $\omega_i^1$ is of even degree. Let us recall that even forms on the free loop space commute. $F(\sigma_n)$ is built from even forms on the free loop space, which commute. This explains why we have to consider the symmetric Fock space. Therefore, if $\sigma$ belongs to $\mathrm{WN}_{s,\infty-}$, then $F(\sigma) = \sum F^{2r}(\sigma)$, where $F^{2r}(\sigma)$ is a measurable form on $L(N)$ of degree $2r$ (see Jones and Léandre (1991) for an analogous statement in the stochastic context).

Let us explain why the free loop space is important in this context. Let $\mathrm{d}\nu_x(1)$ be the law of the Brownian bridge on $N$ starting from $x$ and coming back at $x$ at time 1: this is the law of the Brownian motion $x_.(x)$ subject to return in time 1 at its departure. Let $p_t(x, y)$ be the heat kernel associated with $x_t(x)$: the law of $x_t(x)$ is namely $p_t(x, y)\,\mathrm{d}m_N(y)$ (Ikeda and Watanabe 1981). We consider the Bismut–Høegh–Krohn measure on the continuous free loop space $L_0(N)$:

$$\mathrm{d}P = p_1(x, x)\mathrm{d}x \otimes \mathrm{d}\nu_x(1) \quad [14]$$

This satisfies

$$\mathrm{tr}[\exp[-s_1\Delta_N] f_1 \cdots f_n \exp[-(1 - s_n)\Delta_N]] = \int_{L_0(N)} f_1(x(s_1)) \cdots f_n(x(s_n)) \mathrm{d}P \quad [15]$$

(We are interested in the trace of the heat semigroup instead of the heat semigroup itself unlike in the previous section.)

Since $N$ is spin, we can consider the spin bundle $\mathrm{Sp} = \mathrm{Sp}_+ \oplus \mathrm{Sp}_-$ on it, the Clifford bundle Cl on it with its natural $\mathbb{Z}/2\mathbb{Z}$ gradation (Gilkey 1995). Let us recall that the Clifford algebra acts on the spinors. A form $\omega$ can be associated with an element $\tilde{\omega}$ of the Clifford bundle (Gilkey 1995). We consider the Brownian loop $x(.)$ associated to the Bismut–Høegh–Krohn measure. If $s < t$, we can define the stochastic parallel transport $\tilde{\tau}_{s,t}$ from $x(t)$ to $x(s)$ (we identify a loop to a path from $[0, 1]$ into $N$ with the same end values). We remark that with the notations of [13]

$$\int_{\Delta_n} \tilde{\tau}_{0,s_1}(\tilde{\omega}_1(\mathrm{d}x(s_1)) + \tilde{\omega}_1 \mathrm{d}s_1)\tilde{\tau}_{s_1,s_2} \ldots \times \tilde{\tau}_{s_{n-1},s_n}(\tilde{\omega}_n(\mathrm{d}x(s_n)) + \tilde{\omega}_n^1 \mathrm{d}s_n)\tilde{\tau}_{s_n,1} = A \quad [16]$$

is a random almost surely defined even element of the Clifford bundle over $x(0)$. Acting on $\mathrm{Sp}(x(0))$, it thus preserves the gradation. We consider its supertrace $\mathrm{tr}_s A = \mathrm{tr}_{\mathrm{Sp}_+} A - \mathrm{tr}_{\mathrm{Sp}_-} A$. This becomes a random variable on $L_0(N)$. We introduce the scalar curvature $K$ of the Levi–Civita connection on $N$, whose introduction arises from the Lichnerowicz formula given the square of the Dirac operator in terms of the horizontal

Laplacian on the spin bundle (Gilkey 1995). We consider the expression $\int_{L_0(N)} \exp[-\int_0^1 K(x(s)\,\mathrm{d}s/8]$ $\mathrm{tr}_s A\,\mathrm{d}P$. This expression can be extended to $\mathrm{WN}_{s,\infty-}$ and therefore defines an element Wi of $\mathrm{WN}_{s,-\infty}$ called by Getzler (Léandre 2002) the Witten current.

Bismut has introduced a Hermitian bundle $\xi$ on $M$. He deduces a bundle $\xi_\infty$ on $L(N)$: the fiber on a loop $x(.)$ is the space of smooth sections along the loop of $\xi$. We can suppose that $\xi$ is a sub-bundle given by a projector $p$ of a trivial bundle. We can suppose that the Hermitian connection on $\xi$ is the projection connection $A = p\mathrm{d}p$ such that its curvature is $R = p\mathrm{d}p \wedge p\mathrm{d}p$. Bismut (1985, 1987) has introduced the Bismut–Chern character:

$$\mathrm{Ch}(\xi_\infty) = \mathrm{tr}\left( \int_{\Delta_n} (A\mathrm{d}x(s_1) - R\mathrm{d}s_1) \wedge \cdots \wedge (A\mathrm{d}x(s_n) - R\mathrm{d}s_n) \right) \qquad [17]$$

$\mathrm{Ch}(\xi_\infty)$ is a collection of even forms equal to $F(\sigma(\xi))$, where $\sigma(\xi)$ belongs to $\mathrm{WN}_{s,\infty-}$. We obtain:

**Theorem 3** *Let us consider the index* $\mathrm{Ind}(D_\xi)$ *of the Dirac operator on N with auxiliary bundle* $\xi$ (Hida *et al.* 1993). *We have*

$$\langle \mathrm{Wi}, \sigma(\xi)\rangle = \mathrm{Ind}\ D_\xi \qquad [18]$$

The proof arises from the Lichnerowicz formula, the matricial Feynman–Kac formula, and the decomposition of the solution of a stochastic linear equation into the sum of iterated integrals.

By using the Potthoff–Streit theorem, we can do the analytic continuation of [18], as is suggested by the path-integral interpretation of Atiyah (1985) or Bismut (1985, 1987) of [18], motivated by the Duistermaat–Heckman or Berline–Vergne localization formulas on the free loop space. For this, these authors consider the Atiyah–Witten even form on the free loop space given by $I(x(.)) = \int_{S^1} |(\mathrm{d}/\mathrm{d}s)x(s)|^2 \mathrm{d}s + \mathrm{d}X_\infty$, where $\mathrm{d}X_\infty$ is the exterior derivative of the Killing form $X_\infty$ which to a vector $X(.)$ on the loop associates $\langle X_\infty, X(.)\rangle = \int_{S^1} \langle X(s), \mathrm{d}x(s)\rangle$. We should obtain the heuristic formula

$$\langle \mathrm{Wi}, \sigma\rangle = Z^{-1} \int_{L(M)} F(\sigma) \wedge \exp\left[-\frac{1}{2} I(x(.))\right] \qquad [19]$$

We refer to Léandre (2002) for details.

Let us remark that Bismut (1987) and Léandre (2003) has continued his formal considerations to the case of the index theorem for a family of Dirac operators. We consider a fibration $\pi: N \to B$ of compact manifolds. Bismut replaces [19] by an integral of forms on the set of loops of $N$ which project to a given loop of $B$. Bismut remarks that this integration in the fiber is related to filtering theory in stochastic analysis.

## Fermionic Brownian Motion

Alvarez-Gaumé has given a supersymmetric proof of the index theorem: the path representation of the index of the Dirac operator involves infinite-dimensional Berezin integrals, while in the previous section only integrals of forms on the free loop space were concerned. Rogers (1987) has given an interpretation of the work of Alvarez-Gaumé, which begins with the study of fermionic Brownian motion. Let us interpret the considerations of Rogers (1987) in this framework.

We consider $\mathbf{C}^d$. $H$ is the space of $L^2$-maps from $[0,1]$ into $\mathbf{C}^d$. We denote such a path by $\phi(s) = (\phi_1(s), \ldots, \phi_d(s))$, where $\phi_i(s) = q_i(s) + \sqrt{-1}p_i(s)$. $p_i(s)$ is the $i$th momentum and $q_i(s)$ the $i$th position. We denote by $\hat{\Lambda}(H)$ the fermionic Fock space associated with $H$.

We introduce the bilinear antisymmetric form on $H$:

$$\Omega(\phi^1, \phi^2) = \sqrt{-1} \sum_{i=1}^d \int_0^1 -p_i^1(s)\,\mathrm{d}q_i^2(s) + p_i^2(s)\,\mathrm{d}q_i^1(s) \qquad [20]$$

and we consider the formal expression $\exp[\Omega] = \sum_{n=0}^\infty n!^{-1}\Omega^{\wedge n}$. We define a state on $\hat{\Lambda}^2(H)$ by $\omega(\phi^1 \wedge \phi^2) = \Omega(\phi^1, \phi^2)$. We put $\hat{\phi}_i(s) = 1_{[0,s]} + \sqrt{-1}1_{[0,s]}$ where we take the $i$th coordinate in $\mathbf{C}^d$. We obtain, if $s_1 < s_2$,

$$\omega(\hat{\phi}_i(s_1) \wedge \hat{\phi}_j(s_2)) = -\sqrt{-1}\delta_{i,j} \qquad [21]$$

where $\delta_{i,j}$ is the Kronecker symbol. We change the sign if $s_2 > s_1$ and we write 0 if $s_1 = s_2$.

We consider the finite-dimensional space Pol of fermionic polynomials on $\mathbf{C}^d$. Pol is endowed with a suitable norm, and we consider $\mathrm{Pol}^{\otimes n}$ endowed with the induced norm. We consider a formal series $\sigma = \sum \sigma_n$, where $\sigma_n$ belongs to $\mathrm{Pol}^{\otimes n}$. In order to simplify the treatment, we suppose that our fermionic polynomials do not contain constant terms. We introduce the following Banach norm:

$$\|\sigma\|_C = \sum \frac{C^n}{n!}\|\sigma_n\| \qquad [22]$$

We obtain the notion of Connes space $\mathrm{Co}_{\infty-}$ in this simpler context: $\sigma$ belongs to $\mathrm{Co}_{\infty-}$ if $\|\sigma\|_C < \infty$ for all C. If $\sigma_n = P_1 \otimes \cdots \otimes P_n$, we associate

$$F(\sigma_n) = \int_{\Delta_n} P_1(\hat{\phi}(s_1)) \wedge \cdots \wedge P_n(\hat{\phi}(s_n))\,\mathrm{d}s_1 \cdots \mathrm{d}s_n \qquad [23]$$

$F$ can be extended in an injective continuous map from $\mathrm{Co}_{\infty-}$ into $\hat{\Lambda}(H)$. By using [21], we get:

**Theorem 4** $\exp[\Omega]$ *is a distribution in the sense of Connes.*

We have only to use the formula [21] and

$$\langle \exp[\Omega], \phi^1 \wedge \phi^2 \cdots \wedge \phi^{2n} \rangle = \mathrm{Pf}\{\omega(\phi^i \wedge \phi^j)\} \qquad [24]$$

and to estimate the obtained Pfaffians when $n \to \infty$.

Theorem 4 allows us to give a rigorous interpretation of the fermionic Feynman–Kac formula of Rogers (1987). We refer to Roepstorff (1994) for details.

$\exp[\Omega]$ should give a rigorous interpretation to the Gaussian Berezin integral with formal density $\exp[\sqrt{-1}\int_0^1 \sum p_i(s)\, \mathrm{d}q_i(s)]$.

*See also:* Equivariant Cohomology and the Cartan Model; Feynman Path Integrals; Functional Integration in Quantum Physics; Hopf Algebras and $q$-Deformation Quantum Groups; Index Theorems; Measure on Loop Spaces; Positive Maps on $C^*$-Algebras; Stationary Phase Approximation; Stochastic Differential Equations; Supermanifolds; Supersymmetric Quantum Mechanics.

## Further Reading


Accardi L and Bożejko M (1998) Interacting Fock spaces and Gaussianization of probability measures. *Infinite Dimensional Analysis, Quantum Probability and Related Topics* 1: 663–670.

Albeverio S (1996) Wiener and Feynman path integrals and their applications. In: Masani PR (eds.) *Norbert Wiener Centenary Congress*, Proc. Symp. Appl. Math. vol. 52, pp. 163–194. Providence, RI: American Mathematical Society.

Andersson L and Driver B (1999) Finite dimensional approximation to Wiener measure and path integral formulas on manifolds. *Journal of Functional Analysis* 165: 430–498.

Atiyah M (1985) Circular symmetry and stationary phase approximation. In: *Colloque en l'honneur de L. Schwartz*, vol. 131, pp. 43–59. Paris: Asterisque.

Bismut JM (1985) Index theorem and equivariant cohomology on the loop space. *Communications in Mathematical Physics* 98: 213–237.

Berezansky YM and Kondratiev YO (1995) *Spectral Methods in Infinite-Dimensional Analysis*, vols. I, II. Dordrecht: Kluwer.

Bismut JM (1986) Localization formula, superconnections and the index theorem for families. *Communications in Mathematical Physics* 103: 127–166.

Bismut JM (1987) Filtering equation, equivariant cohomology and the Chern character. In: Seneor R (ed.) *Proc. VIIIth Int. Cong. Math. Phys.*, pp. 17–56. Singapore: World Scientific.

Chen KT (1973) Iterated path integrals of differential forms and loop space homology. *Annals of Mathematics* 97: 213–237.

Connes A (1988) Entire cyclic cohomology of Banach algebras and character of $\theta$-summable Fredholm modules. *K-Theory* 1: 519–548.

Cuntz J (2001) Cyclic theory, bivariant K-theory and the bivariant Chern–Connes character. In: *Cyclic Homology in Noncommutative Geometry*, pp. 2–69. Encyclopedia of Mathematical Sciences, 121. Heidelberg: Springer.

Gilkey P (1995) *Invariance Theory, the Heat Equation and the Atiyah–Singer Theorem*. Boca Raton: CRC Press.

Hida T, Kuo HH, Potthoff J, and Streit L (1993) *White Noise: An Infinite Dimensional Calculus*. Dordrecht: Kluwer.

Ikeda N and Watanabe S (1981) *Stochastic Differential Equations and Diffusion Processes*. Amsterdam: North-Holland.

Jones JDS and Léandre R (1991) $L^p$ Chen forms on loop spaces. In: Barlow M and Bingham N (eds.) *Stochastic Analysis*, pp. 104–162. Cambridge: Cambridge University Press.

Léandre R (2002) White noise analysis, filtering equation and the index theorem for families. In: Heyer H and Saitô (eds.) *Infinite Dimensional Harmonic Analysis* (to appear).

Léandre R (2003) Theory of distributions in the sense of Connes–Hida and Feynman path integral on a manifold. *Inf. Dim. Anal. Quant. Probab. Rel. Top.* 6: 505–517.

Roepstorff G (1994) *Path Integral Approach to Quantum Physics. An Introduction*. Heidelberg: Springer.

Rogers A (1987) Fermionic path integration and Grassmann Brownian motion. *Communications in Mathematical Physics* 113: 353–368.

Sidorova N, Smolyanov O, von Weizsaecker H, and Wittich O (2004) The surface limit of Brownian motion in tubular neighborhood of an embedded Riemannian manifold. *Journal of Functional Analysis* 206: 391–413.

Szabo R (2000) Equivariant cohomology and localization of path integrals in physics, *Lecture Notes in Physics* M63. Berlin: Springer.


# Peakons

**D D Holm**, Imperial College, London, UK



## Introduction

Peakons are singular solutions of the dispersionless Camassa–Holm (CH) shallow-water wave equation in one spatial dimension. These are reviewed in the context of asymptotic expansions and Euler–Poincaré (EP) variational principles. The dispersionless CH equation generalizes to the EPDiff equation (defined subsequently in this article), whose singular solutions are peakon wave fronts in higher dimensions. The reduction of these singular solutions of CH and EPDiff to canonical Hamiltonian dynamics on lower-dimensional sets may be understood, by realizing that their solution ansatz is a momentum map, and momentum maps are Poisson.

Camassa and Holm (1993) discovered the "peakon" solitary traveling-wave solution for a shallow-water wave:

$$u(x,t) = c\mathrm{e}^{-|x-ct|/\alpha} \qquad [1]$$

whose fluid velocity $u$ is a function of position $x$ on the real line and time $t$. The peakon traveling wave

moves at a speed equal to its maximum height, at which it has a sharp peak (jump in derivative). Peakons are an emergent phenomenon, solving the initial-value problem for a partial differential equation (PDE) derived by an asymptotic expansion of Euler's equations using the small parameters of shallow-water dynamics. Peakons are nonanalytic solitons, which superpose as

$$u(x,t) = \sum_{a=1}^{N} p_a(t)\mathrm{e}^{-|x-q_a(t)|/\alpha} \qquad [2]$$

for sets $\{p\}$ and $\{q\}$ satisfying canonical Hamiltonian dynamics. Peakons arise for shallow-water waves in the limit of zero linear dispersion in one dimension. Peakons satisfy a PDE arising from Hamilton's principle for geodesic motion on the smooth invertible maps (diffeomorphisms) with respect to the $H^1$ Sobolev norm of the fluid velocity. Peakons generalize to higher dimensions, as well. We explain how peakons were derived in the context of shallow-water asymptotics and describe some of their remarkable mathematical properties.

## Shallow-Water Background for Peakons

Euler's equations for irrotational incompressible ideal fluid motion under gravity with a free surface have an asymptotic expansion for shallow-water waves that contains two small parameters, $\epsilon$ and $\delta^2$, with ordering $\epsilon \geq \delta^2$. These small parameters are $\epsilon = a/h_0$ (the ratio of wave amplitude to mean depth) and $\delta^2 = (h_0/l_x)^2$ (the squared ratio of mean depth to horizontal length, or wavelength). Euler's equations are made nondimensional by introducing $x = l_x x'$ for horizontal position, $z = h_0 z'$ for vertical position, $t = (l_x/c_0)t'$ for time, $\eta = a\eta'$ for surface elevation, and $\varphi = (gl_x a/c_0)\varphi'$ for velocity potential, where $c_0 = \sqrt{gh_0}$ is the mean wave speed and $g$ is the constant gravity. The quantity $\sigma = \sigma'/(h_0 \rho c_0^2)$ is the dimensionless Bond number, in which $\rho$ is the mass density of the fluid and $\sigma'$ is its surface tension, both of which are taken to be constants. After dropping primes, this asymptotic expansion yields the nondimensional Korteweg–de Vries (KdV) equation for the horizontal velocity variable $u = \varphi_x(x,t)$ at "linear" order in the small dimensionless ratios $\epsilon$ and $\delta^2$, as the left-hand side of

$$u_t + u_x + \frac{3\epsilon}{2}uu_x + \frac{\delta^2}{6}(1-3\sigma)u_{xxx} = \mathrm{O}(\epsilon\delta^2) \qquad [3]$$

Here, partial derivatives are denoted using subscripts, and boundary conditions are $u = 0$ and $u_x = 0$ at spatial infinity on the real line. The famous $\mathrm{sech}^2(x-t)$ traveling-wave solutions (the solitons) for KdV [3] arise in a balance between its (weakly) nonlinear steepening and its third-order linear dispersion, when the quadratic terms in $\epsilon$ and $\delta^2$ on its right-hand side are neglected.

In eqn [3], a normal-form transformation due to Kodama (1985) has been used to remove the other possible quadratic terms of order $O(\epsilon^2)$ and $O(\delta^4)$. The remaining quadratic correction terms in the KdV equation [3] may be collected at order $O(\epsilon\delta^2)$. These terms may be expressed, after introducing a "momentum variable,"

$$m = u - \nu\delta^2 u_{xx} \qquad [4]$$

and neglecting terms of cubic order in $\epsilon$ and $\delta^2$, as

$$m_t + m_x + \frac{\epsilon}{2}(um_x + bmu_x) + \frac{\delta^2}{6}(1-3\sigma)u_{xxx} = 0 \qquad [5]$$

In the momentum variable $m = u - \nu\delta^2 u_{xx}$, the parameter $\nu$ is given by Dullin *et al.* (2001):

$$\nu = \frac{19 - 30\sigma - 45\sigma^2}{60(1-3\sigma)} \qquad [6]$$

Thus, the effects of $\delta^2$-dispersion also enter the nonlinear terms. After restoring dimensions in eqn [5] and rescaling velocity $u$ by $(b+1)$, the following "$b$-equation" emerges,

$$m_t + c_0 m_x + um_x + bmu_x + \Gamma u_{xxx} = 0 \qquad [7]$$

where $m = u - \alpha^2 u_{xx}$ is the dimensional momentum variable, and the constants $\alpha^2$ and $\Gamma/c_0$ are squares of length scales. When $\alpha^2 \to 0$, one recovers KdV from the $b$-equation [7], up to a rescaling of velocity. Any value of the parameter $b \neq -1$ may be achieved in eqn [7] by an appropriate Kodama transformation (Dullin *et al.* 2001).

As already emphasized, the values of the coefficients in the asymptotic analysis of shallow-water waves at quadratic order in their two small parameters only hold, modulo the Kodama normal-form transformations. Hence, these transformations may be used to advance the analysis and thereby gain insight, by optimizing the choices of these coefficients. The freedom introduced by the Kodama transformations among asymptotically equivalent equations at quadratic order in $\epsilon$ and $\delta^2$ also helps to answer the perennial question, "Why are integrable equations so ubiquitous when one uses asymptotics in modeling?"

### Integrable Cases of the *b*-equation [7]

The cases $b = 2$ and $b = 3$ are special values for which the $b$-equation becomes a completely integrable Hamiltonian system. For $b = 2$, eqn [7]

specializes to the integrable CH equation of Camassa and Holm (1993). The case $b = 3$ in [7] recovers the integrable equation of Degasperis and Procesi (1999) (henceforth DP equation). These two cases exhaust the integrable candidates for [7], as was shown using Painlevé analysis. The $b$-family of eqns [7] was also shown in Mikhailov and Novikov (2002) to admit the symmetry conditions necessary for integrability, only in the cases $b = 2$ for CH and $b = 3$ for DP.

The $b$-equation [7] with $b = 2$ was first derived in Camassa and Holm (1993) by using asymptotic expansions directly in the Hamiltonian for Euler's equations governing inviscid incompressible flow in the shallow-water regime. In this analysis, the CH equation was shown to be bi-Hamiltonian and thereby was found to be completely integrable by the inverse-scattering transform (IST) on the real line. Reviews of IST may be found, for example, in Ablowitz and Clarkson (1991), Dubrovin (1981), and Novikov *et al.* (1984). For discussions of other related bi-Hamiltonian equations, see Degasperis and Procesi (1999).

Camassa and Holm (1993) also discovered the remarkable peaked soliton (peakon) solutions of [1], [2] for the CH equation on the real line, given by [7] in the case $b = 2$. The peakons arise as solutions of [7], when $c_0 = 0$ and $\Gamma = 0$ in the absence of linear dispersion. Peakons move at a speed equal to their maximum height, at which they have a sharp peak (jump in derivative). Unlike the KdV soliton, the peakon speed is independent of its width ($\alpha$). Periodic peakon solutions of CH were treated in Alber *et al.* (1999). There, the sharp peaks of periodic peakons were associated with billiards reflecting at the boundary of an elliptical domain. These billiard solutions for the periodic peakons arise from geodesic motion on a triaxial ellipsoid, in the limit that one of its axes shrinks to zero length.

Before Camassa and Holm (1993) derived their shallow-water equation, a class of integrable equations existed, which was later found to contain eqn [7] with $b = 2$. This class of integrable equations was derived using hereditary symmetries in Fokas and Fuchssteiner (1981). However, eqn [7] was not written explicitly, nor was it derived physically as a shallow-water equation and its solution properties for $b = 2$ were not studied before Camassa and Holm (1993). (See Fuchssteiner (1996) for an insightful history of how the shallow-water equation [7] in the integrable case with $b = 2$ relates to the mathematical theory of hereditary symmetries.)

Equation [7] with $b = 2$ was recently re-derived as a shallow-water equation by using asymptotic methods in three different approaches in Dullin *et al.* (2001), in Fokas and Liu (1996), and also in Johnson [2002]. All the three derivations used different variants of the method of asymptotic expansions for shallow-water waves in the absence of surface tension. Only the derivation in Dullin *et al.* (2001) used the Kodama normal-form transformations to take advantage of the nonuniqueness of the asymptotic expansion results at quadratic order.

The effects of the parameter $b$ on the solutions of eqn [7] were investigated in Holm and Staley (2003), where $b$ was treated as a bifurcation parameter, in the limiting case when the linear dispersion coefficients are set to $c_0 = 0$ and $\Gamma = 0$. This limiting case allows several special solutions, including the peakons, in which the two nonlinear terms in eqn [7] balance each other in the "absence" of linear dispersion.

## Peakons: Singular Solutions without Linear Dispersion in One Spatial Dimension

Peakons were first found as singular soliton solutions of the completely integrable CH equation. This is eqn [7] with $b = 2$, now rewritten in terms of the velocity as

$$\begin{aligned} u_t + c_0 u_x + 3uu_x + \Gamma u_{xxx} \\ = \alpha^2(u_{xxt} + 2u_x u_{xx} + uu_{xxx}) \end{aligned} \qquad [8]$$

Peakons were found in Camassa and Holm (1993) to arise in the absence of linear dispersion. That is, they arise when $c_0 = 0$ and $\Gamma = 0$ in CH [8]. Specifically, peakons are the individual terms in the peaked $N$-soliton solution of CH [8] for its velocity

$$u(x,t) = \sum_{b=1}^{N} p_b(t) e^{-|x - q_b(t)|/\alpha} \qquad [9]$$

in the absence of linear dispersion. Each term in the sum is a soliton with a sharp peak at its maximum, hence the name "peakon." Expressed using its momentum, $m = (1 - \alpha^2 \partial_x^2)u$, the peakon velocity solution [9] of dispersionless CH becomes a sum over a delta functions, supported on a set of points moving on the real line. Namely, the peakon velocity solution [9] implies

$$m(x,t) = 2\alpha \sum_{b=1}^{N} p_b(t)\delta(x - q_b(t)) \qquad [10]$$

because of the relation $(1 - \alpha^2\partial_x^2)e^{-|x|/\alpha} = 2\alpha\delta(x)$. These solutions satisfy the $b$-equation [7] for any value of $b$, provided $c_0 = 0$ and $\Gamma = 0$.

Thus, peakons are "singular momentum solutions" of the dispersionless $b$-equation, although

they are not stable for every value of $b$. From numerical simulations (Holm and Staley 2003), peakons are conjectured to be stable for $b > 1$. In the integrable cases $b=2$ for CH and $b=3$ for DP, peakons are stable singular soliton solutions. The spatial velocity profile $e^{-|x|/\alpha}/2\alpha$ of each separate peakon in [9] is the Green's function for the Helmholtz operator on the real line, with vanishing boundary conditions at spatial infinity. Unlike the KdV soliton, whose speed and width are related, the width of the peakon profile is set by its Green's function, independently of its speed.

### Integrable Peakon Dynamics of CH

Substituting the peakon solution ansatz [9] and [10] into the dispersionless CH equation

$$m_t + um_x + 2mu_x = 0, \quad m = u - \alpha^2 u_{xx} \qquad [11]$$

yields Hamilton's canonical equations for the dynamics of the discrete set of peakon parameters $q_a(t)$ and $p_a(t)$:

$$\dot{q}_a(t) = \frac{\partial h_N}{\partial p_a} \quad \text{and} \quad \dot{p}_a(t) = -\frac{\partial h_N}{\partial q_a} \qquad [12]$$

for $a = 1, 2, \ldots, N$, with Hamiltonian given by (Camassa and Holm 1993):

$$h_N = \frac{1}{2}\sum_{a,\, b=1}^{N} p_a p_b e^{-|q_a - q_b|/\alpha} \qquad [13]$$

Thus, one finds that the points $x = q^a(t)$ in the peakon solution [9] move with the flow of the fluid velocity $u$ at those points, since $u(q^a(t), t) = \dot{q}^a(t)$. This means the $q^a(t)$ are Lagrangian coordinates. Moreover, the singular momentum solution ansatz [10] is the Lagrange-to-Euler map for an invariant manifold of the dispersionless CH equation [11]. On this finite-dimensional invariant manifold for the PDE [11], the dynamics is canonically Hamiltonian.

With Hamiltonian [13], the canonical equations [12] for the $2N$ canonically conjugate peakon parameters $p_a(t)$ and $q_a(t)$ were interpreted in Camassa and Holm (1993) as describing "geodesic motion" on the $N$-dimensional Riemannian manifold whose co-metric is $g^{ij}(\{q\}) = e^{-|q_i - q_j|/\alpha}$. Moreover, the canonical geodesic equations arising from Hamiltonian [13] comprise an integrable system for any number of peakons $N$. This integrable system was studied in Camassa and Holm (1993) for solutions on the real line, and in Alber *et al.* (1999) and Mckean and Constantin (1999) and references therein, for spatially periodic solutions.

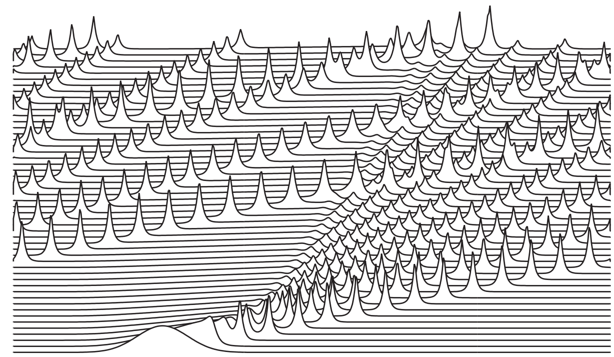


**Figure 1** A smooth localized (Gaussian) initial condition for the CH equation breaks up into an ordered train of peakons as time evolves (the time direction being vertical). The peakon train eventually wraps around the periodic domain, thereby allowing the leading peakons to overtake the slower emergent peakons from behind in collisions that cause phase shifts as discussed in Camassa and Holm (1993). Courtesy of Staley M.

Being a completely integrable Hamiltonian soliton equation, the continuum CH equation [8] has an associated isospectral eigenvalue problem, discovered in Camassa and Holm (1993) for any values of its dispersion parameters $c_0$ and $\Gamma$. Remarkably, when $c_0 = 0$ and $\Gamma = 0$, this isospectral eigenvalue problem has a purely "discrete" spectrum. Moreover, in this case, each discrete eigenvalue corresponds precisely to the time-asymptotic velocity of a peakon. This discreteness of the CH isospectrum in the absence of linear dispersion implies that only the singular peakon solutions [10] emerge asymptotically in time, in the solution of the initial-value problem for the dispersionless CH equation [11]. This is borne out in numerical simulations of the dispersionless CH equation [11], starting from a smooth initial distribution of velocity (Fringer and Holm 2001, Holm and Staley 2003).

**Figure 1** shows the emergence of peakons from an initially Gaussian velocity distribution and their subsequent elastic collisions in a periodic one-dimensional domain. This figure demonstrates that singular solutions dominate the initial-value problem and, thus, that it is imperative to go beyond smooth solutions for the CH equation; the situation is similar for the EPDiff equation.

### Peakons as Mechanical Systems

Being governed by canonical Hamiltonian equations, each $N$-peakon solution can be associated with a mechanical system of moving particles. Calogero (1995) further extended the class of mechanical systems of this type. The $r$-matrix approach was applied to the Lax pair formulation of the $N$-peakon system for CH by Ragnisco and Bruschi (1996), who also pointed out the connection

of this system with the classical Toda lattice. A discrete version of the Adler–Kostant–Symes factorization method was used by Suris (1996) to study a discretization of the peakon lattice, realized as a discrete integrable system on a certain Poisson submanifold of gl($N$) equipped with an $r$-matrix Poisson bracket. Beals *et al.* (1999) used the Stieltjes theorem on continued fractions and the classical moment problem for studying multipeakon solutions of the CH equation. Generalized peakon systems are described for any simple Lie algebra by Alber *et al.* (1999).

#### Pulsons: Generalizing the Peakon Solutions of the Dispersionless $b$-Equation for Other Green's Functions

The Hamiltonian $h_N$ in eqn [13] depends on the Green's function for the relation between velocity $u$ and momentum $m$. However, the singular momentum solution ansatz [10] is "independent" of this Green's function. Thus, as discovered in Fringer and Holm (2001), the singular momentum solution ansatz [10] for the dispersionless equation

$$m_t + um_x + 2mu_x = 0, \quad \text{with } u = g * m \qquad [14]$$

provides an invariant manifold on which canonical Hamiltonian dynamics occurs, for any choice of the Green's function $g$ relating velocity $u$ and momentum $m$ by the convolution $u = g * m$.

The fluid velocity solutions corresponding to the singular momentum ansatz [10] for eqn [14] are the "pulsons". Pulsons are given by the sum over $N$ velocity profiles determined by the Green's function $g$, as

$$u(x, t) = \sum_{a=1}^{N} p_a(t) g(x, q_a(t)) \qquad [15]$$

Again for [14], the singular momentum ansatz [10] results in a finite-dimensional invariant manifold of solutions, whose dynamics is canonically Hamiltonian. The Hamiltonian for the canonical dynamics of the $2N$ parameters $p_a(t)$ and $q_a(t)$ in the "pulson" solutions [15] of eqn [14] is

$$h_N = \frac{1}{2} \sum_{a,\, b=1}^{N} p_a p_b \, g(q_a, q_b) \qquad [16]$$

Again, for the pulsons, the canonical equations for the invariant manifold of singular momentum solutions provide a phase-space description of geodesic motion, this time with respect to the co-metric given by the Green's function $g$. Mathematical analysis and numerical results for the dynamics of these pulson solutions are given in Fringer and Holm (2001). These results describe how the collisions of pulsons [15] depend upon their shape.

#### Compactons in the $1/\alpha^2 \to 0$ Limit of CH

As mentioned earlier, in the limit that $\alpha^2 \to 0$, the CH equation [8] becomes the KdV equation. In contrast, when $1/\alpha^2 \to 0$, CH becomes the Hunter–Zheng equation (Hunter and Zheng 1994):

$$(u_t + uu_x)_{xx} = \tfrac{1}{2}\left(u_x^2\right)_x$$

This equation has "compacton" solutions, whose collision dynamics was studied numerically and put into the present context in Fringer and Holm (2001). The corresponding Green's function satisfies $-\partial_x{}^2 g(x) = 2\delta(x)$, so it has the triangular shape, $g(x) = 1 - |x|$ for $|x| < 1$, and vanishes otherwise, for $|x| \geq 1$. That is, the Green's function in this case has compact support, hence the name "compactons" for these pulson solutions, which as a limit of the integrable CH equations are true solitons, solvable by IST.

#### Pulson Solutions of the Dispersionless $b$-Equation

Holm and Staley (2003) give the pulson solutions of the traveling-wave problem and their elastic collision properties for the dispersionless $b$-equation:

$$m_t + um_x + bmu_x = 0, \quad \text{with } u = g * m \qquad [17]$$

with any (symmetric) Green's function $g$ and for any value of the parameter $b$. Numerically, pulsons and peakons are both found to be stable for $b > 1$ (Holm and Staley 2003). The reduction to "noncanonical" Hamiltonian dynamics for the invariant manifold of singular momentum solutions [10] of the other integrable case $b = 3$ with peakon Green's function $g(x, y) = \mathrm{e}^{-|x-y|/\alpha}$ is found in Degasperis and Procesi (1999) and Degasperis *et al.* (2002).

### Euler–Poincaré Theory in More Dimensions

#### Generalizing the Peakon Solutions of the CH Equation to Higher Dimensions

In Holm and Staley (2003), weakly nonlinear analysis and the assumption of columnar motion in the variational principle for Euler's equations are found to produce the two-dimensional generalization of the dispersionless CH equation [11]. This generalization is the EP equation (Holm *et al.* 1998a, b) for the Lagrangian consisting of the kinetic energy:

$$\ell = \frac{1}{2} \int \left[ |\boldsymbol{u}|^2 + \alpha^2 (\operatorname{div} \boldsymbol{u})^2 \right] \mathrm{d}x \, \mathrm{d}y \qquad [18]$$

in which the fluid velocity $\boldsymbol{u}$ is a two-dimensional vector. Evolution generated by kinetic energy in

Hamilton's principle results in geodesic motion, with respect to the velocity norm $\|\boldsymbol{u}\|$, which is provided by the kinetic-energy Lagrangian. For ideal incompressible fluids governed by Euler's equations, the importance of geodesic flow was recognized by Arnol'd (1966) for the $L^2$ norm of the fluid velocity. The EP equation generated by any choice of kinetic-energy norm without imposing incompressibility is called "EPDiff," for "Euler–Poincaré equation for geodesic motion on the diffeomorphisms." EPDiff is given by (Holm *et al.* 1998a):

$$\left(\frac{\partial}{\partial t} + \boldsymbol{u} \cdot \nabla\right)\boldsymbol{m} + \nabla \boldsymbol{u}^{\mathrm{T}} \cdot \boldsymbol{m} + \boldsymbol{m}(\operatorname{div} \boldsymbol{u}) = 0 \qquad [19]$$

with momentum density $\boldsymbol{m} = \delta\ell/\delta\boldsymbol{u}$, where $\ell = (1/2)\|\boldsymbol{u}\|^2$ is given by the kinetic energy, which defines a norm in the fluid velocity $\|\boldsymbol{u}\|$, yet to be determined. By design, this equation has no contribution from either potential energy or pressure. It conserves the velocity norm $\|\boldsymbol{u}\|$ given by the kinetic energy. Its evolution describes geodesic motion on the diffeomorphisms with respect to this norm (Holm *et al.* 1998a).

An alternative way of writing the EPDiff equation [19] in either two or three dimensions is

$$\frac{\partial}{\partial t}\boldsymbol{m} - \boldsymbol{u} \times \operatorname{curl} \boldsymbol{m} + \nabla(\boldsymbol{u} \cdot \boldsymbol{m}) + \boldsymbol{m}(\operatorname{div} \boldsymbol{u}) = 0 \qquad [20]$$

This form of EPDiff involves all three differential operators: curl, gradient, and divergence. For the kinetic-energy Lagrangian $\ell$ given in [18], which is a norm for "irrotational" flow (with $\operatorname{curl} \boldsymbol{u} = 0$), we have the EPDiff equation [19] with momentum $\boldsymbol{m} = \delta\ell/\delta\boldsymbol{u} = \boldsymbol{u} - \alpha^2\nabla(\operatorname{div} \boldsymbol{u})$.

EPDiff [19] may also be written intrinsically as

$$\frac{\partial}{\partial t}\frac{\delta\ell}{\delta\boldsymbol{u}} = -\mathrm{ad}^*_{\boldsymbol{u}}\frac{\delta\ell}{\delta\boldsymbol{u}} \qquad [21]$$

where $\mathrm{ad}^*$ is the $L^2$ dual of the ad-operation (commutator) for vector fields (see Arnol'd and Khesin (1998) and Marsden and Ratiu (1999) for additional discussions of the beautiful geometry underlying this equation).

### Reduction to the Dispersionless CH Equation in One Dimension

In one dimension, the EPDiff equations [19]–[21] with Lagrangian $\ell$ given in [18] simplify to the dispersionless CH equation [11]. The dispersionless limit of the CH equation appears, because potential energy and pressure have been ignored.

### Strengthening the Kinetic-Energy Norm to Allow for Circulation

The kinetic-energy Lagrangian [18] is a norm for irrotational flow, with $\operatorname{curl} \boldsymbol{u} = 0$. However, inclusion of rotational flow requires the kinetic-energy norm to be strengthened to the $H^1_\alpha$ norm of the velocity, defined as

$$\begin{aligned}\ell &= \frac{1}{2}\int\left[|\boldsymbol{u}|^2 + \alpha^2(\operatorname{div} \boldsymbol{u})^2 + \alpha^2(\operatorname{curl} \boldsymbol{u})^2\right]\mathrm{d}x\,\mathrm{d}y \\ &= \frac{1}{2}\int\left[|\boldsymbol{u}|^2 + \alpha^2|\nabla\boldsymbol{u}|^2\right]\mathrm{d}x\,\mathrm{d}y = \frac{1}{2}\|\boldsymbol{u}\|^2_{H^1_\alpha}\end{aligned} \qquad [22]$$

Here, we assume boundary conditions that give no contributions upon integrating by parts. The corresponding EPDiff equation is [19] with $\boldsymbol{m} \equiv \delta\ell/\delta\boldsymbol{u} = \boldsymbol{u} - \alpha^2\Delta\boldsymbol{u}$. This expression involves inversion of the familiar Helmholtz operator in the (nonlocal) relation between fluid velocity and momentum density. The $H^1_\alpha$ norm $\|\boldsymbol{u}\|^2_{H^1_\alpha}$ for the kinetic energy [22] also arises in three dimensions for turbulence modeling based on Lagrangian averaging and using Taylor's hypothesis that the turbulent fluctuations are "frozen" into the Lagrangian mean flow (Foias *et al.* 2001).

### Generalizing the CH Peakon Solutions to *n* Dimensions

Building on the peakon solutions [9] for the CH equation and the pulsons [15] for its generalization to other traveling-wave shapes in Fringer and Holm (2001), Holm and Staley (2003) introduced the following measure-valued singular momentum solution ansatz for the $n$-dimensional solutions of the EPDiff equation [19]:

$$\boldsymbol{m}(\boldsymbol{x}, t) = \sum_{a=1}^{N}\int \boldsymbol{P}^a(s,t)\delta(\boldsymbol{x} - \boldsymbol{Q}^a(s,t))\,\mathrm{d}s \qquad [23]$$

These singular momentum solutions, called "diffeons," are vector density functions supported in $\mathbb{R}^n$ on a set of $N$ surfaces (or curves) of codimension $(n-k)$ for $s \in \mathbb{R}^k$ with $k < n$. They may, for example, be supported on sets of points (vector peakons, $k = 0$), one-dimensional filaments (strings, $k = 1$), or two-dimensional surfaces (sheets, $k = 2$) in three dimensions.

Figure 2 shows the results for the EPDiff equation when a straight peakon segment of finite length is created initially moving rightward (East). Because of propagation along the segment in adjusting to the condition of zero speed at its ends and finite speed in its interior, the initially straight segment expands outward as it propagates and curves into a peakon "bubble."

Figure 3 shows an initially straight segment whose velocity distribution is exponential in the transverse

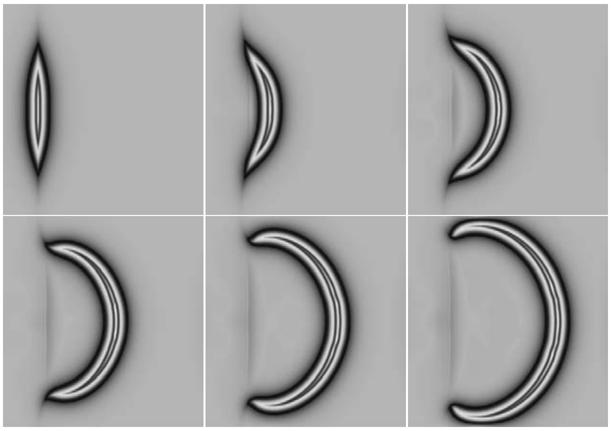


**Figure 2** A peakon segment of finite length is initially moving rightward (east). Because its speed vanishes at its ends and it has fully two-dimensional spatial dependence, it expands into a peakon "bubble" as it propagates. (The various shades indicate different speeds. Any transverse slice will show a wave profile with a maximum at the center of the wave, which falls exponentially with distance away from the center.)

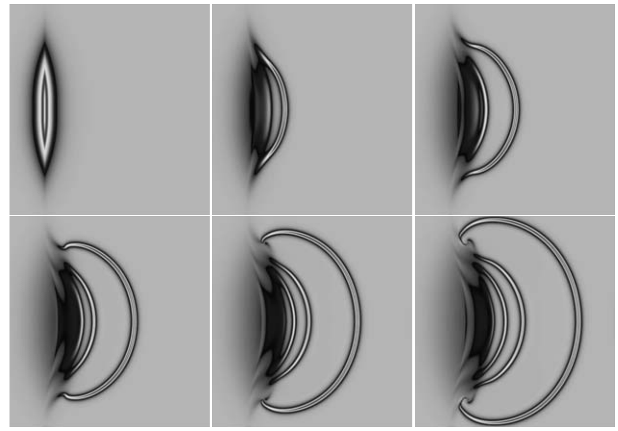


**Figure 3** An initially straight segment of velocity distribution whose exponential profile is wider than the width $\alpha$ for the peakon solution breaks up into a train of curved peakon "bubbles," each of width $\alpha$. This example illustrates the emergent property of the peakon solutions in two dimensions.

direction, but is wider than $\alpha$ for the peakon solution. This initial-velocity distribution evolves under EPDiff to separate into a train of curved peakon "bubbles," each of width $\alpha$. This example illustrates the emergent property of the peakon solutions in two dimensions. This phenomenon is observed in nature, for example, as trains of internal wave fronts in the South China Sea (Liu *et al.* 1998).

Substitution of the singular momentum solution ansatz [23] into the EPDiff equation [19] implies the following integro-partial-differential equations (IPDEs) for the evolution of the parameters $\{\boldsymbol{P}\}$ and $\{\boldsymbol{Q}\}$:

$$\begin{aligned}\frac{\partial}{\partial t}\boldsymbol{Q}^a(s,t) &= \sum_{b=1}^{N}\int \boldsymbol{P}^b(s',t)\, G\Big(\boldsymbol{Q}^a(s,t) - \boldsymbol{Q}^b(s',t)\Big)\,\mathrm{d}s' \\ \frac{\partial}{\partial t}\boldsymbol{P}^a(s,t) &= -\sum_{b=1}^{N}\int \Big(\boldsymbol{P}^a(s,t)\cdot\boldsymbol{P}^b(s',t)\Big) \times \frac{\partial}{\partial \boldsymbol{Q}^a(s,t)} G\Big(\boldsymbol{Q}^a(s,t) - \boldsymbol{Q}^b(s',t)\Big)\,\mathrm{d}s'\end{aligned} \quad [24]$$

Importantly for the interpretation of these solutions, the coordinates $s \in \mathbb{R}^k$ turn out to be Lagrangian coordinates. The velocity field corresponding to the momentum solution ansatz [23] is given by

$$\begin{aligned}\boldsymbol{u}(\boldsymbol{x},t) &= G * \boldsymbol{m} \\ &= \sum_{b=1}^{N}\int \boldsymbol{P}^b(s',t)\, G\Big(\boldsymbol{x} - \boldsymbol{Q}^b(s',t)\Big)\,\mathrm{d}s'\end{aligned} \quad [25]$$

for $\boldsymbol{u} \in \mathbb{R}^n$. When evaluated along the curve $\boldsymbol{x} = \boldsymbol{Q}^a(s,t)$, this velocity satisfies

$$\begin{aligned}\boldsymbol{u}(\boldsymbol{Q}^a(s,t),t) &= \sum_{b=1}^{N}\int \boldsymbol{P}^b(s',t) \times G\Big(\boldsymbol{Q}^a(s,t) - \boldsymbol{Q}^b(s',t)\Big)\,\mathrm{d}s' \\ &= \frac{\partial \boldsymbol{Q}^a(s,t)}{\partial t}\end{aligned} \quad [26]$$

Consequently, the lower-dimensional support sets defined on $\boldsymbol{x} = \boldsymbol{Q}^a(s,t)$ and parametrized by coordinates $s \in \mathbb{R}^k$ move with the fluid velocity. This means that the $s \in \mathbb{R}^k$ are Lagrangian coordinates. Moreover, eqns [24] for the evolution of these support sets are canonical Hamiltonian equations:

$$\frac{\partial}{\partial t}\boldsymbol{Q}^a(s,t) = \frac{\delta H_N}{\delta \boldsymbol{P}^a}, \qquad \frac{\partial}{\partial t}\boldsymbol{P}^a(s,t) = -\frac{\delta H_N}{\delta \boldsymbol{Q}^a} \quad [27]$$

The corresponding Hamiltonian function $H_N : (\mathbb{R}^n \times \mathbb{R}^n)^N \to \mathbb{R}$ is

$$H_N = \frac{1}{2}\iint \sum_{a,b=1}^{N}\Big(\boldsymbol{P}^a(s,t)\cdot\boldsymbol{P}^b(s',t)\Big) \times G(\boldsymbol{Q}^a(s,t), \boldsymbol{Q}_j(s',t))\,\mathrm{d}s\,\mathrm{d}s' \quad [28]$$

This is the Hamiltonian for geodesic motion on the cotangent bundle of a set of curves $\boldsymbol{Q}^a(s,t)$ with respect to the metric given by $G$. This dynamics was investigated numerically in Holm and Staley (2003) which can be referred to for more details of the solution properties. One important result found "numerically" in Holm and Staley (2003) is that only codimension-1 singular momentum solutions

appear to be stable under the evolution of the EPDiff equation. Thus,

*Stability for codimension-1 solutions: the singular momentum solutions of EPDiff are stable, as points on the line (peakons), as curves in the plane (filaments, or wave fronts), or as surfaces in space (sheets).*

Proving this stability result analytically remains an outstanding problem. The stability of peakons on the real line is proven in Constantin and Strauss (2000).

### Reconnections in Oblique Overtaking Collisions of Peakon Wave Fronts

Figures 4 and 5 show results of oblique wave front collisions producing reconnections for the EPDiff equation in two dimensions. Figure 4 shows a single oblique overtaking collision, as a faster expanding peakon wave front overtakes a slower one and reconnects with it at the collision point. Figure 5 shows a series of reconnections involving the oblique overtaking collisions of two trains of curved peakon filaments, or wave fronts.

### The Peakon Reduction is a Momentum Map

As shown in Holm and Marsden (2004), the singular solution ansatz [23] is a momentum map from the cotangent bundle of the smooth embeddings of lower-dimensional sets $\mathbb{R}^s \in \mathbb{R}^n$, to the dual of the Lie algebra of vector fields defined on these sets. (Momentum maps for Hamiltonian dynamics are reviewed in Marsden and Ratiu (1999), for example.) This geometric feature underlies the remarkable reduction properties of the EPDiff equation, and it also explains why the reduced equations must be Hamiltonian on the invariant manifolds of the singular solutions; namely, because

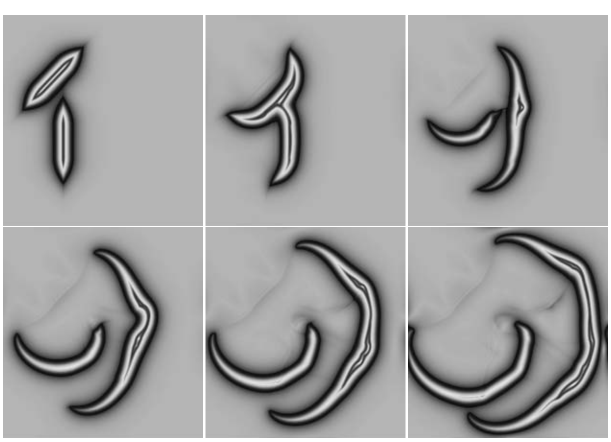


**Figure 4** A single collision is shown involving reconnection as the faster peakon segment initially moving southeast along the diagonal expands, curves, and obliquely overtakes the slower peakon segment initially moving rightward (east). This reconnection illustrates one of the collision rules for the strongly two-dimensional EPDiff flow.

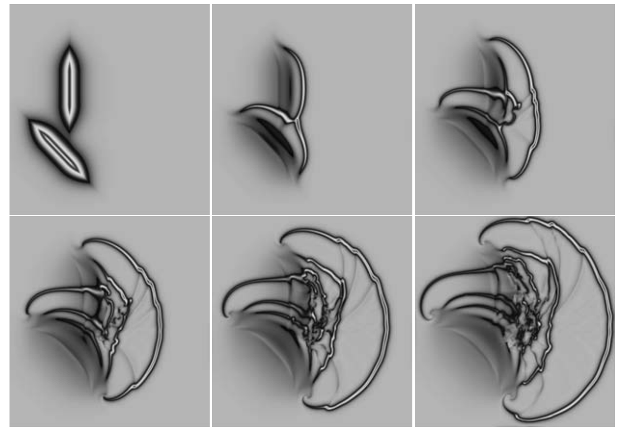


**Figure 5** A series of multiple collisions is shown involving reconnections as the faster wider peakon segment initially moving northeast along the diagonal expands, breaks up into a wave train of peakons, each of which propagates, curves, and obliquely overtakes the slower wide peakon segment initially moving rightward (east), which is also breaking up into a train of wave fronts. In this series of oblique collision, the now-curved peakon filaments exchange momentum and reconnect several times.

momentum maps are Poisson maps. This geometric feature also underlies the singular momentum solution [23] and its associated velocity [25] which generalize the peakon solutions, both to higher dimensions and to arbitrary kinetic-energy metrics. The result that the singular solution ansatz [23] is a momentum map helps to organize the theory, to explain previous results, and to suggest new avenues of exploration.

## Acknowledgment

The author is grateful to R Camassa, J E Marsden, T S Ratiu, and A Weinstein for their collaboration, help, and inspiring discussions over the years. He also thanks M F Staley for providing the figures obtained from his numerical simulations in the collaborations. US DOE provided partial support, under contract W-7405-ENG-36 for Los Alamos National Laboratory, and Office of Science ASCAR/AMS/MICS.

*See also:* Hamiltonian Systems: Obstructions to Integrability; Integrable Systems: Overview; Wave Equations and Diffraction.

## Further Reading

Ablowitz MJ and Clarkson PA (1991) *Solitons, Nonlinear Evolution Equations and Inverse Scattering*. Cambridge: Cambridge University Press.

Ablowitz MJ and Segur H (1981) *Solitons and the Inverse Scattering Transform*. Philadelphia: SIAM.

Alber M, Camassa R, Fedorov Y, Holm D, and Marsden JE (1999) On billiard solutions of nonlinear PDEs. *PhysicsLetters A* 264: 171–178.

Alber M, Camassa R, Fedorov Y, Holm D, and Marsden JE (2001) The complex geometry of weak piecewise smooth solutions of integrable nonlinear PDE's of shallow water and Dym type. *Communications in Mathematical Physics* 221: 197–227.

Alber M, Camassa R, Holm D, and Marsden JE (1994) The geometry of peaked solitons and billiard solutions of a class of integrable PDEs. *Letters in Mathematical Physics* 32: 137–151.

Alber MS, Camassa R, and Gekhtman M (2000) On billiard weak solutions of nonlinear PDE's and Toda flows. *CRM Proceedings and Lecture Notes* 25: 1–11.

Arnol'd VI (1966) Sur la géométrie differentielle des groupes de Lie de dimenson infinie et ses applications à l'hydrodynamique des fluids parfaits. *Annales de l'Institut Fourier, Grenoble* 16: 319–361.

Arnol'd VI and Khesin BA (1998) *Topological Methods in Hydrodynamics*. Springer: New York.

Beals R, Sattinger DH, and Szmigielski J (1999) Multi-peakons and a theorem of Stietjes. *Inverse Problems* 15: L1–4.

Beals R, Sattinger DH, and Szmigielski J (2000) Multipeakons and the classical moment problem. *Advances in Mathematics* 154: 229–257.

Beals R, Sattinger DH, and Szmigielski J (2001) Peakons, strings, and the finite Toda lattice. *Communications in Pure and Applied Mathematics* 54: 91–106.

Calogero F (1995) An integrable Hamiltonian system. *Physics Letters A* 201: 306–310.

Calogero F and Francoise J-P (1996) A completely integrable Hamiltonian system. *Journal of Mathematics* 37: 2863–2871.

Camassa R and Holm DD (1993) An integrable shallow water equation with peaked solitons. *Physical Review Letters* 71: 1661–1664.

Camassa R, Holm DD, and Hyman JM (1994) A new integrable shallow water equation. *Advances in Applied Mechanics* 31: 1–33.

Constantin A and Strauss W (2000) Stability of peakons. *Communications on Pure and Applied Mathematics* 53: 603–610.

Degasperis A and Procesi M (1999) Asymptotic integrability. In: Degasperis A and Gaeta G (eds.) *Symmetry and Perturbation Theory*, pp. 23–37. Singapore: World Scientific.

Degasperis A, Holm DD, and Hone ANW (2002) A new integrable equation with peakon solutions. *Theoretical and Mathematical Physics* 133: 1463–1474.

Dubrovin B (1981) Theta functions and nonlinear equations. *Russian Mathematical Surveys* 36: 11–92.

Dubrovin BA, Novikov SP, and Krichever IM (1985) Integrable systems. *I. Itogi Nauki i Tekhniki. Sovr. Probl. Mat. Fund. Naprav*. 4. VINITI (Moscow) (Engl. transl. (1989) *Encyclopaedia of Mathematical Sciences*, vol. 4. Berlin: Springer).

Dullin HR, Gottwald GA, and Holm DD (2001) An integrable shallow water equation with linear and nonlinear dispersion. *Physical Review Letters* 87: 194501–04.

Dullin HR, Gottwald GA, and Holm DD (2003) Camassa–Holm, Korteweg–de Vries-5 and other asymptotically equivalent equations for shallow water waves. *Fluid Dynamics Research* 33: 73–95.

Dullin HR, Gottwald GA, and Holm DD (2004) On asymptotically equivalent shallow water wave equations. *Physica D* 190: 1–14.

Foias C, Holm DD, and Titi ES (2001) The Navier–Stokes-alpha model of fluid turbulence. *Physica D* 152: 505–519.

Fokas AS and Fuchssteiner B (1981) Bäcklund transformations for hereditary symmetries. *Nonlinear Analysis Transactions of the American Mathematical Society* 5: 423–432.

Fokas AS and Liu QM (1996) Asymptotic integrability of water waves. *Physical Review Letters* 77: 2347–2351.

Fuchssteiner B (1996) Some tricks from the symmetry-toolbox for nonlinear equations: generalization of the Camassa–Holm equation. *Physica D* 95: 229–243.

Fringer O and Holm DD (2001) Integrable vs. nonintegrable geodesic soliton behavior. *Physica D* 150: 237–263.

Holm DD (2005) The Euler–Poincaré variational framework for modeling fluid dynamics. In: Montaldi J and Ratiu T (eds.) *Geometric Mechanics and Symmetry: The Peyresq Lectures*, pp. 157–209. London Mathematical Society Lecture Notes Series 306. Cambridge: Cambridge University Press.

Holm DD and Marsden JE (2004) Momentum maps and measure-valued solutions (peakons, filaments and sheets) for the EPDiff equation. In: Marsden JE and Ratiu TS (eds.) *The Breadth of Symplectic and Poisson Geometry, A Festshcrift for Alan Weinstein*, pp. 203–235, Progress in Mathematics, vol. 232. Boston: Birkhäuser.

Holm DD, Marsden JE, and Ratiu TS (1998a) The Euler–Poincaré equations and semidirect products with applications to continuum theories. *Advances in Mathematics* 137: 1–81.

Holm DD, Marsden JE, and Ratiu TS (1998b) Euler–Poincaré models of ideal fluids with nonlinear dispersion. *Physical Review Letters* 349: 4173–4177.

Holm DD and Staley MF (2003) Nonlinear balance and exchange of stability in dynamics of solitons, peakons, ramps/cliffs and leftons in a 1+1 nonlinear evolutionary PDE. *Physical Letters A* 308: 437–444.

Holm DD and Staley MF (2003) Wave structures and nonlinear balances in a family of evolutionary PDEs. *SIAM Journal of Applied Dynamical Systems* 2(3): 323–380.

Holm DD and Staley MF (2004) Interaction dynamics of singular wave fronts in nonlinear evolutionary fluid equations (in preparation).

Hunter JK and Zheng Y (1994) On a completely integrable hyperbolic variational equation. *Physica D* 79: 361–386.

Johnson RS (2002) Camassa–Holm, Korteweg–de Vries models for water waves. *Journal of Fluid Mechanics* 455: 63–82.

Kodama Y (1985) On integrable systems with higher order corrections. *Physics Letters A* 107: 245–249.

Kodama Y (1985a) Normal forms for weakly dispersive wave equations. *Physics Letters A* 112: 193–196.

Kodama Y (1987) On solitary-wave interaction. *Physics letters A* 123: 276–282.

Liu AK, Chang YS, Hsu M-K, and Liang NK (1998) Evolution of nonlinear internal waves in the east and south China Seas. *Journal of Geophysical Research* 103: 7995–8008.

Marsden JE and Ratiu TS (1999) *Introduction to Mechanics and Symmetry*. Texts in Applied Mathematics, 2nd edn. vol. 17, Berlin: Springer.

McKean HP and Constantin A (1999) A shallow water equation on the circle. *Communications on Pure and Applied Mathematics* 52: 949–982.

Mikhailov AV and Novikov VS (2002) Perturbative symmetry approach. *Journal of Physics A* 35: 4775–4790.

Novikov SP, Manakov SV, Pitaevski LP, and Zakharov VE (1984) *Theory of Solitons. The Inverse Scattering Method*, Comtemporary Soviet Mathematics. Consultants Bureau (translated from Russian). New York: Plenum.

Ragnisco O and Bruschi M (1996) Peakons, *r*-matrix and Toda lattice. *Physica A* 228: 150–159.

Suris YB (1996) A discrete time peakons lattice. *Physics Letters A* 217: 321–329.

**Penrose Inequality** *see* Geometric Flows and the Penrose Inequality

# Percolation Theory

**V Beffara**, Ecole Normale Supérieure de Lyon, Lyon, France
**V Sidoravicius**, IMPA, Rio de Janeiro, Brazil



## Introduction

Percolation as a mathematical theory was introduced by Broadbent and Hammersley (1957), as a stochastic way of modeling the flow of a fluid or gas through a porous medium of small channels which may or may not let gas or fluid pass. It is one of the simplest models exhibiting a phase transition, and the occurrence of a critical phenomenon is central to the appeal of percolation. Having truly applied origins, percolation has been used to model the fingering and spreading of oil in water, to estimate whether one can build nondefective integrated circuits, and to model the spread of infections and forest fires. From a mathematical point of view, percolation is attractive because it exhibits relations between probabilistic and algebraic/topological properties of graphs.

To make the mathematical construction of such a system of channels, take a graph $\mathcal{G}$ (which originally was taken as $\mathbb{Z}^d$), with vertex set $\mathcal{V}$ and edge set $\mathcal{E}$, and make all the edges independently open (or passable) with probability $p$ or closed (or blocked) with probability $1-p$. Write $P_p$ for the corresponding probability measure on the set of configurations of open and closed edges – that model is called bond percolation. The collection of open edges thus forms a random subgraph of $\mathcal{G}$, and the original question stated by Broadbent was whether the connected component of the origin in that subgraph is finite or infinite.

A path on $\mathcal{G}$ is a sequence $v_1, v_2, \ldots$ of vertices of $\mathcal{G}$, such that for all $i \geq 1$, $v_i$ and $v_{i+1}$ are adjacent on $\mathcal{G}$. A path is called open if all the edges $\{v_i, v_{i+1}\}$ between successive vertices are open. The infiniteness of the cluster of the origin is equivalent to the existence of an unbounded open path starting from the origin.

There is an analogous model, called "site percolation," in which all edges are assumed to be passable, but the vertices are independently open or closed with probability $p$ or $1-p$, respectively. An open path is then a path along which all vertices are open. Site percolation is more general than bond percolation in the sense that the existence of a path for bond percolation on a graph $\mathcal{G}$ is equivalent to the existence of a path for site percolation on the covering graph of $\mathcal{G}$. However, site percolation on a given graph may not be equivalent to bond percolation on any other graph.

All graphs under consideration will be assumed to be connected, locally finite and quasitransitive. If $A, B \subset \mathcal{V}$, then $A \leftrightarrow B$ means that there exists an open path from some vertex of $A$ to some vertex of $B$; by a slight abuse of notation, $u \leftrightarrow v$ will stand for the existence of a path between sites $u$ and $v$, that is, the event $\{u\} \leftrightarrow \{v\}$. The open cluster $C(v)$ of the vertex $v$ is the set of all open vertices which are connected to $v$ by an open path:

$$C(v) = \{u \in \mathcal{V} : u \leftrightarrow v\}$$

The central quantity of the percolation theory is the percolation probability:

$$\theta(p) := P_p\{\mathbf{0} \leftrightarrow \infty\} = P_p\{|C(\mathbf{0})| = \infty\}$$

The most important property of the percolation model is that it exhibits a phase transition, that is, there exists a threshold value $p_c \in [0,1]$, such that the global behavior of the system is substantially different in the two regions $p < p_c$ and $p > p_c$. To make this precise, observe that $\theta$ is a nondecreasing function. This can be seen using Hammersley's joint construction of percolation systems for all $p \in [0,1]$ on $\mathcal{G}$: let $\{U(v), v \in \mathcal{V}\}$ be independent random variables, uniform in [0,1]. Declare $v$ to be $p$-open if $U(v) \leq p$, otherwise it is declared $p$-closed. The configuration of $p$-open vertices has the distribution $P_p$ for each $p \in [0,1]$. The collection of $p$-open vertices is nondecreasing in $p$, and therefore $\theta(p)$ is nondecreasing as well. Clearly, $\theta(0) = 0$ and $\theta(1) = 1$ (**Figure 1**).

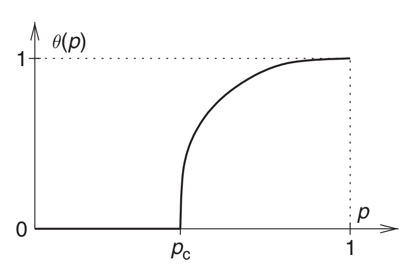



**Figure 1** The behavior of $\theta(p)$ around the critical point (for bond percolation).

The critical probability is defined as

$$p_c := p_c(\mathcal{G}) = \sup\{p : \theta(p) = 0\}$$

By definition, when $p < p_c$, the open cluster of the origin is $P_p$-a.s. finite; hence, all the clusters are also finite. On the other hand, for $p > p_c$ there is a strictly positive $P_p$-probability that the cluster of the origin is infinite. Thus, from Kolmogorov's zero–one law it follows that

$$P_p\{|C(v)| = \infty \text{ for some } v \in \mathcal{V}\} = 1 \quad \text{for } p > p_c$$

Therefore, if the intervals $[0, p_c)$ and $(p_c, 1]$ are both nonempty, there is a phase transition at $p_c$.

Using a so-called Peierls argument it is easy to see that $p_c(\mathcal{G}) > 0$ for any graph $\mathcal{G}$ of bounded degree. On the other hand, Hammersley proved that $p_c(\mathbb{Z}^d) < 1$ for bond percolation as soon as $d \geq 2$, and a similar argument works for site percolation and various periodic graphs as well. But for some graphs $\mathcal{G}$, it is not so easy to show that $p_c(\mathcal{G}) < 1$. One says that the system is in the subcritical (resp. supercritical) phase if $p < p_c$ (resp. $p > p_c$).

It was one of the most remarkable moments in the history of percolation when Kesten (1980) proved, based on results by Harris, Russo, Seymour and Welsh, that the critical parameter for bond percolation on $\mathbb{Z}^2$ is equal to 1/2. Nevertheless, the exact value of $p_c(\mathcal{G})$ is known only for a handful of graphs, all of them periodic and two dimensional – see below.

## Percolation in $\mathbb{Z}^d$

The graph on which most of the theory was originally built is the cubic lattice $\mathbb{Z}^d$, and it was not before the late twentieth century that percolation was seriously considered on other kinds of graphs (such as Cayley graphs), on which specific phenomena can appear, such as the coexistence of multiple infinite clusters for some values of the parameter $p$. In this section, the underlying graph is thus assumed to be $\mathbb{Z}^d$ for $d \geq 2$, although most of the results still hold in the case of a periodic $d$-dimensional lattice.

### The Subcritical Regime

When $p < p_c$, all open clusters are finite almost surely. One of the greatest challenges in percolation theory has been to prove that $\chi(p) := E_p\{|C(v)|\}$ is finite if $p < p_c$ ($E_p$ stands for the expectation with respect to $P_p$). For that one can define another critical probability as the threshold value for the finiteness of the expected cluster size of a fixed vertex:

$$p_T(\mathcal{G}) := \sup\{p : \chi(p) < \infty\}$$

It was an important step in the development of the theory to show that $p_T(\mathcal{G}) = p_c(\mathcal{G})$. The fundamental estimate in the subcritical regime, which is a much stronger statement than $p_T(\mathcal{G}) = p_c(\mathcal{G})$, is the following:

**Theorem 1** (Aizenman and Barsky, Menshikov). *Assume that $\mathcal{G}$ is periodic. Then for $p < p_c$ there exist constants $0 < C_1, C_2 < \infty$, such that*

$$P_p\{|C(v)| \geq n\} \leq C_1\, e^{-C_2 n}$$

The last statement can be sharpened to a "local limit theorem" with the help of a subadditivity argument: for each $p < p_c$, there exists a constant $0 < C_3(p) < \infty$, such that

$$\lim_{n\to\infty} -\frac{1}{n}\log P_p\{|C(v)| = n\} = C_3(p)$$

### The Supercritical Regime

Once an infinite open cluster exists, it is natural to ask how it looks like, and how many infinite open clusters exist. It was shown by Newman and Schulman that for periodic graphs, for each $p$, exactly one of the following three situations prevails: if $N \in \mathbb{Z}_+ \cup \{\infty\}$ is the number of infinite open clusters, then $P_p(N = 0) = 1$, or $P_p(N = 1) = 1$, or $P_p(N = \infty) = 1$.

Aizenman, Kesten, and Newman showed that the third case is impossible on $\mathbb{Z}^d$. By now several proofs exist, perhaps the most elegant of which is due to Burton and Keane, who prove that indeed there cannot be infinitely many infinite open clusters on any amenable graph. However, there are some graphs, such as regular trees, on which coexistence of several infinite clusters is possible.

The geometry of the infinite open cluster can be explored in some depth by studying the behavior of a random walk on it. When $d = 2$, the random walk is recurrent, and when $d \geq 3$ is a.s. transient. In all dimensions $d \geq 2$, the walk behaves diffusively, and the "central limit theorem" and the "invariance principle" were established in both the annealed and quenched cases.

**Wulff droplets** In the supercritical regime, aside from the infinite open cluster, the configuration contains finite clusters of arbitrary large sizes. These large finite open clusters can be thought of as droplets swimming in the areas surrounded by an infinite open cluster. The presence at a particular location of a large finite cluster is an event of low probability, namely, on $\mathbb{Z}^d, d \geq 2$, for $p > p_c$, there exist positive constants $0 < C_4(p), C_5(p) < \infty$, such that

$$C_4(p) \leq -\frac{1}{n^{(d-1)/d}}\log P_p\{|C(v)| = n\} \leq C_5(p)$$

for all large $n$. This estimate is based on the fact that the occurrence of a large finite cluster is due to a surface effect. The typical structure of the large finite cluster is described by the following theorem:

**Theorem 2** *Let $d \geq 2$, and $p > p_c$. There exists a bounded, closed, convex subset $W$ of $\mathbb{R}^d$ containing the origin, called the* normalized *Wulff crystal of the Bernoulli percolation model, such that, under the conditional probability $P_p\{\cdot\, | n^d \leq |C(0)| < \infty\}$, the random measure*

$$\frac{1}{n^d}\sum_{x\in C(\mathbf{0})} \delta_{x/n}$$

*(where $\delta_x$ denotes a Dirac mass at $x$) converges weakly in probability toward the random measure $\theta(p)\mathbb{1}_W(x-M)\,dx$ (where $M$ is the rescaled center of mass of the cluster $C(\mathbf{0})$). The deviation probabilities behave as $\exp\{-cn^{d-1}\}$ (i.e., they exhibit large deviations of surface order; in dimensions 4 and more it holds up to re-centering).*

This result was proved in dimension 2 by Alexander et al. (1990), and in dimensions 3 and more by Cerf (2000).

### Percolation Near the Critical Point

**Percolation in Slabs** The main macroscopic observable in percolation is $\theta(p)$, which is positive above $p_c$, 0 below $p_c$, and continuous on $[0,1]\backslash\{p_c\}$. Continuity at $p_c$ is an open question in the general case; it is known to hold in two dimensions (cf. below) and in high enough dimension (at the moment $d \geq 19$ though the value of the critical dimension is believed to be 6) using lace expansion methods. The conjecture that $\theta(p_c) = 0$ for $3 \leq d \leq 18$ remains one of the major open problems.

Efforts to prove that led to some interesting and important results. Barsky, Grimmett, and Newman solved the question in the half-space case, and simultaneously showed that the slab percolation and half-space percolation thresholds coincide. This was complemented by Grimmett and Marstrand showing that

$$p_c(\text{slab}) = p_c(\mathbb{Z}^d)$$

**Critical exponents** In the subcritical regime, exponential decay of the correlation indicates that there is a finite correlation length $\xi(p)$ associated to the system, and defined (up to constants) by the relation

$$P_p(0 \leftrightarrow nx) \approx \exp\left(-\frac{n\varphi(x)}{\xi(p)}\right)$$

where $\varphi$ is bounded on the unit sphere (this is known as Ornstein–Zernike decay). The phase transition can then also be defined in terms of the divergence of the correlation length, leading again to the same value for $p_c$; the behavior at or near the critical point then has no finite characteristic length, and gives rise to scaling exponents (conjecturally in most cases).

The most usual critical exponents are defined as follows, if $\theta(p)$ is the percolation probability, $C$ the cluster of the origin, and $\xi(p)$ the correlation length:

$$\frac{\partial^3}{\partial p^3}E_p[|C|^{-1}] \approx |p-p_c|^{-1-\alpha}$$
$$\theta(p) \approx (p-p_c)_+^{\beta}$$
$$\chi^f(p) := E_p[|C|\mathbb{1}_{|C|<\infty}] \approx |p-p_c|^{-\gamma}$$
$$P_{p_c}[|C| = n] \approx n^{-1-1/\delta}$$
$$P_{p_c}[x \in C] \approx |x|^{2-d-\eta}$$
$$\xi(p) \approx |p-p_c|^{\nu}$$
$$P_{p_c}[\text{diam}(C) = n] \approx n^{-1-1/\rho}$$
$$\frac{E_p[|C|^{k+1}\mathbb{1}_{|C|<\infty}]}{E_p[|C|^{k}\mathbb{1}_{|C|<\infty}]} \approx |p-p_c|^{-\Delta}$$

These exponents are all expected to be universal, that is, to depend only on the dimension of the lattice, although this is not well understood at the mathematical level; the following scaling relations between the exponents are believed to hold:

$$2-\alpha = \gamma + 2\beta = \beta(\delta+1), \quad \Delta = \delta\beta,\ \gamma = \nu(2-\eta)$$

In addition, in dimensions up to $d_c = 6$, two additional hyperscaling relations involving $d$ are strongly conjectured to hold:

$$d\rho = \delta + 1, \qquad d\nu = 2 - \alpha$$

while above $d_c$ the exponents are believed to take their mean-field value, that is, the ones they have for percolation on a regular tree:

$$\alpha = -1,\ \beta = 1,\ \gamma = 1,\ \delta = 2$$
$$\eta = 0,\ \nu = \tfrac{1}{2},\ \rho = \tfrac{1}{2},\ \Delta = 2$$

Not much is known rigorously on critical exponents in the general case. Hara and Slade (1990) proved that mean field behavior does happen above dimension 19, and the proof can likely be extended to treat the case $d \geq 7$. In the two-dimensional case on the other hand, Kesten (1987) showed that, assuming that the exponents $\delta$ and $\rho$ exist, then so do $\beta, \gamma, \eta$, and $\nu$, and they satisfy the scaling and hyperscaling relations where they appear.

**The incipient infinite cluster** When studying long-range properties of a critical model, it is useful to have an object which is infinite at criticality, and such is not the case for percolation clusters. There are two ways to condition the cluster of the origin to

be infinite when $p = p_c$: The first one is to condition it to have diameter at least $n$ (which happens with positive probability) and take a limit in distribution as $n$ goes to infinity; the second one is to consider the model for parameter $p > p_c$, condition the cluster of 0 to be infinite (which happens with positive probability) and take a limit in distribution as $p$ goes to $p_c$. The limit is the same in both cases; it is known as the incipient infinite cluster.

As in the supercritical regime, the structure of the cluster can be investigated by studying the behavior of a random walk on it, as was suggested by de Gennes; Kesten proved that in two dimensions, the random walk on the incipient infinite cluster is subdiffusive, that is, the mean square displacement after $n$ steps behaves as $n^{1-\varepsilon}$ for some $\varepsilon > 0$.

The construction of the incipient infinite cluster was done by Kesten (1986) in two dimensions, and a similar construction was performed recently in high dimension by van der Hofstad and Jarai (2004).

## Percolation in Two Dimensions

As is the case for several other models of statistical physics, percolation exhibits many specific properties when considered on a two-dimensional lattice: duality arguments allow for the computation of $p_c$ in some cases, and for the derivation of *a priori* bounds for the probability of crossing events at or near the critical point, leading to the fact that $\theta(p_c) = 0$. On another front, the scaling limit of critical site percolation on the two-dimensional triangular lattice can be described in terms of Stochastic Loewner evolutions (SLE) processes.

### Duality, Exact Computations, and RSW Theory

Given a planar lattice $\mathcal{L}$, define two associated graphs as follows. The dual lattice $\mathcal{L}'$ has one vertex for each face of the original lattice, and an edge between two vertices if and only if the corresponding faces of $\mathcal{L}$ share an edge. The star graph $\mathcal{L}^*$ is obtained by adding to $\mathcal{L}$ an edge between any two vertices belonging to the same face ($\mathcal{L}^*$ is not planar in general; $(\mathcal{L}, \mathcal{L}^*)$ is commonly known as a matching pair). Then, a result of Kesten is that, under suitable technical conditions,

$$p_c^{\text{bond}}(\mathcal{L}) + p_c^{\text{bond}}(\mathcal{L}') = p_c^{\text{site}}(\mathcal{L}) + p_c^{\text{site}}(\mathcal{L}^*) = 1$$

Two cases are of particular importance: the lattice $\mathbb{Z}^2$ is isomorphic to its dual; the triangular lattice $\mathcal{T}$ is its own star graph. It follows that

$$p_c^{\text{bond}}(\mathbb{Z}^2) = p_c^{\text{site}}(\mathcal{T}) = \tfrac{1}{2}$$

The only other critical parameters that are known exactly are $p_c^{\text{bond}}(\mathcal{T}) = 2\sin(\pi/18)$ (and hence also $p_c^{\text{bond}}$ for $\mathcal{T}'$, i.e., the hexagonal lattice) and $p_c^{\text{bond}}$ for the bow-tie lattice which is a root of the equation $p^5 - 6p^3 + 6p^2 + p - 1 = 0$. The value of the critical parameter for site percolation on $\mathbb{Z}^2$ might, on the other hand, never be known; it is even possible that it is "just a number" without any other signification.

Still using duality, one can prove that the probability, for bond percolation on the square lattice with parameter $p = 1/2$, that there is a connected component crossing an $(n+1) \times n$ rectangle in the longer direction is exactly equal to 1/2. This and clever arguments involving the symmetry of the lattice lead to the following result, proved independently by Russo and by Seymour and Welsh and known as the RSW theorem:

**Theorem 3** (Russo 1978, Seymour and Welsh 1978). *For every $a, b > 0$ there exist $\eta > 0$ and $n_0 > 0$ such that for every $n > n_0$, the probability that there is a cluster crossing an $\lfloor na \rfloor \times \lfloor nb \rfloor$ rectangle in the first direction is greater than $\eta$.*

The most direct consequence of this estimate is that the probability that there is a cluster going around an annulus of a given modulus is bounded below independently of the size of the annulus; in particular, almost surely there is some annulus around 0 in which this happens, and that is what allows to prove that $\theta(p_c) = 0$ for bond percolation on $\mathbb{Z}^2$ (**Figure 2**).

### The Scaling Limit

RSW-type estimates give positive evidence that a scaling limit of the model should exist; it is indeed essentially sufficient to show convergence of the crossing probabilities to a nontrivial limit as $n$ goes to infinity. The limit, which should depend only on the ratio $a/b$, was predicted by Cardy using conformal field theory methods. A celebrated result of Smirnov is the proof of Cardy's formula in the case of site percolation on the triangular lattice $\mathcal{T}$:

**Theorem 4** (Smirnov (2001)). *Let $\Omega$ be a simply connected domain of the plane with four points $a$, $b$, $c$, $d$ (in that order) marked on its boundary. For every $\delta > 0$, consider a critical site-percolation*

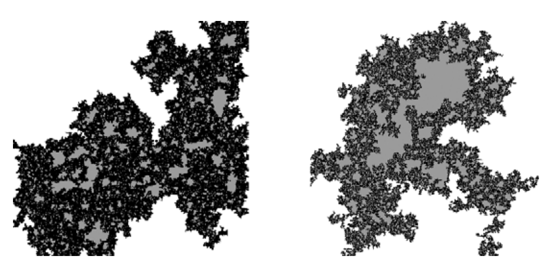


**Figure 2** Two large critical percolation clusters in a box of the square lattice (first: bond percolation, second: site percolation).

*model on the intersection of* $\Omega$ *with* $\delta\mathcal{T}$ *and let* $f_\delta(ab, cd; \Omega)$ *be the probability that it contains a cluster connecting the arcs ab and cd. Then*:

(i) $f_\delta(ab, cd; \Omega)$ *has a limit* $f_0(ab, cd; \Omega)$ *as* $\delta \to 0$;

(ii) *the limit is conformally invariant, in the following sense: if* $\Phi$ *is a conformal map from* $\Omega$ *to some other domain* $\Omega' = \Phi(\Omega)$, *and maps* $a$ *to* $a', b$ *to* $b', c$ *to* $c'$ *and* $d$ *to* $d'$, *then* $f_0(ab, cd; \Omega) = f_0(a'b', c'd'; \Omega')$; *and*

(iii) *in the particular case when* $\Omega$ *is an equilateral triangle of side length* 1 *with vertices a, b and c, and if d is on (ca) at distance* $x \in (0, 1)$ *from c, then* $f_0(ab, cd; \Omega) = x$.

Point (iii) in particular is essential since it allows us to compute the limiting crossing probabilities in any conformal rectangle. In the original work of Cardy, he made his prediction in the case of a rectangle, for which the limit involves hypergeometric functions; the remark that the equilateral triangle gives rise to nicer formulae is originally due to Carleson.

To precisely state the convergence of percolation to its scaling limit, define the random curve known as the percolation exploration path (see Figure 3) as follows: In the upper half-plane, consider a site-percolation model on a portion of the triangular lattice and impose the boundary conditions that on the negative real half-line all the sites are open, while on the other half-line the sites are closed. The exploration curve is then the common boundary of the open cluster spanning from the negative half-line, and the closed cluster spanning from the positive half-line; it is an infinite, self-avoiding random curve in the upper half-plane.

As the mesh of the lattice goes to 0, the exploration curve then converges in distribution to the trace of an SLE process, as introduced by Schramm, with parameter $\kappa = 6$ – see Figure 4. The limiting curve is not simple anymore (which corresponds to the existence of pivotal sites on large critical percolation clusters), and it has Hausdorff dimension 7/4. For more details on SLE processes, see, for example, the related entry in the present volume.

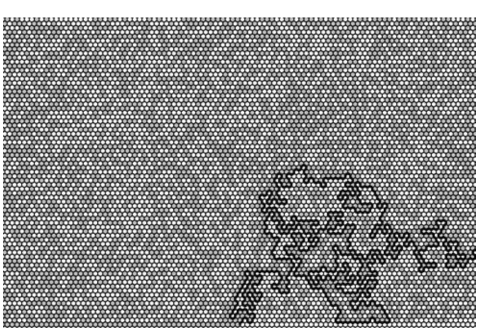


Figure 3 A percolation exploration path. Figure courtesy Schramm O (2000) Scaling limits of loop-erased random walks and uniform spanning trees. *Israel Journal of Mathematics* 118: 221–228.

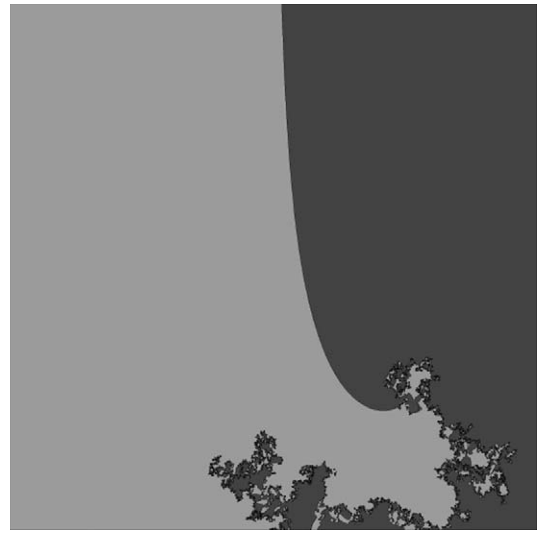


Figure 4 An SLE process with parameter $\kappa = 6$ (infinite time, with the driving process stopped at time 1).

As an application of this convergence result, one can prove that the critical exponents described in the previous section do exist (still in the case of the triangular lattice), and compute their exact values, except for $\alpha$, which is still listed here for completeness:

$$\left[\alpha = -\frac{2}{3}\right], \beta = \frac{5}{36}, \gamma = \frac{43}{18}, \delta = \frac{91}{5}$$
$$\eta = \frac{5}{24}, \nu = \frac{4}{3}, \rho = \frac{48}{5}, \Delta = \frac{91}{36}$$

These exponents are expected to be universal, in the sense that they should be the same for percolation on any two-dimensional lattice; but at the time of this writing, this phenomenon is far from being understood on a mathematical level.

The rigorous derivation of the critical exponents for percolation is due to Smirnov and Werner (2001); the dimension of the limiting curve was obtained by Beffara (2004).

## Other Lattices and Percolative Systems

Some modifications or generalizations of standard Bernoulli percolation on $\mathbb{Z}^d$ exhibit an interesting behavior and as such provide some insight into the original process as well; there are too many mathematical objects which can be argued to be percolative in some sense to give a full account of all

of them, so the following list is somewhat arbitrary and by no means complete.

### Percolation on Nonamenable Graphs

The first modification of the model one can think of is to modify the underlying graph and move away from the cubic lattice; phase transition still occurs, and the main difference is the possibility for infinitely many infinite clusters to coexist. On a regular tree, such is the case whenever $p \in (p_c, 1)$, the first nontrivial example was produced by Grimmett and Newman as the product of $\mathbb{Z}$ by a tree: there, for some values of $p$ the infinite cluster is unique, while for others there is coexistence of infinitely many of them. The corresponding definition, due to Benjamini and Schramm, is then the following: if $N$ is as above the number of infinite open clusters,

$$p_u := \inf\{p : P_p(N = 1) = 1\} \geq p_c$$

The main question is then to characterize graphs on which $0 < p_c < p_u < 1$.

A wide class of interesting graphs is that of Cayley graphs of infinite, finitely generated groups. There, by a simultaneous result by Häggström and Peres and by Schonmann, for every $p \in (p_c, p_u)$ there are $P_p$-a.s. infinitely many infinite cluster, while for every $p \in (p_u, 1]$ there is only one – note that this does not follow from the definition since new infinite components could appear when $p$ is increased. It is conjectured that $p_c < p_u$ for any Cayley graph of a nonamenable group (and more generally for any quasitransitive graph with positive Cheeger constant), and a result by Pak and Smirnova is that every infinite, finitely generated, nonamenable group has a Cayley graph on which $p_c < p_u$; this is then expected not to depend on the choice of generators. In the general case, it was recently proved by Gaboriau that if the graph $\mathcal{G}$ is unimodular, transitive, locally finite, and supports nonconstant harmonic Dirichlet functions (i.e., harmonic functions whose gradient is in $\ell^2$), then indeed $p_c(\mathcal{G}) < p_u(\mathcal{G})$.

For reference and further reading on the topic, the reader is advised to refer to the review paper by Benjamini and Schramm (1996), the lecture notes of Peres (1999), and the more recent article of Gaboriau (2005).

### Gradient Percolation

Another possible modification of the original model is to allow the parameter $p$ to depend on the location; the porous medium may for instance have been created by some kind of erosion, so that there will be more open edges on one side of a given domain than on the other. If $p$ still varies smoothly, then one expects some regions to look subcritical and others to look supercritical, with interesting behavior in the vicinity of the critical level set $\{p = p_c\}$. This particular model was introduced by Sapoval *et al.* (1978) under the name of gradient percolation (see Figure 5).

**Figure 5** Gradient percolation in a square. In black is the cluster spanning from the bottom side of the square.

The control of the model away from the critical zone is essentially the same as for usual Bernoulli percolation, the main question being how to estimate the width of the phase transition. The main idea is then the same as in scaling theory: if the distance between a point $v$ and the critical level set is less than the correlation length for parameter $p(v)$, then $v$ is in the phase transition domain. This of course makes sense only asymptotically, say in a large $n \times n$ square with $p(x, y) = 1 - y/n$ as is the case in the figure: the transition then is expected to have width of order $n^a$ for some exponent $a > 0$.

### First-Passage Percolation

First-passage percolation (also known as Eden or Richardson model) was introduced by Hammersley and Welsh (1965) as a time-dependent model for the passage of fluid through a porous medium. To define the model, with each edge $e \in \mathcal{E}(\mathbb{Z}^d)$ is associated a random variable $T(e)$, which can be interpreted as being the time required for fluid to flow along $e$. The $T(e)$ are assumed to be independent non-negative random variables having common distribution $F$. For any path $\pi$ we define the passage time $T(\pi)$ of $\pi$ as

$$T(\pi) := \sum_{e \in \pi} T(e)$$

The first passage time $a(x, y)$ between vertices $x$ and $y$ is given by

$$a(x,y) = \inf\{T(\pi) : \pi \text{ a path from } x \text{ to } y\}$$

and we can define

$$W(t) := \{x \in \mathbb{Z}^d : a(0,x) \le t\}$$

the set of vertices reached by the liquid by time $t$. It turns out that $W(t)$ grows approximately linearly as time passes, and that there exists a nonrandom limit set $B$ such that either $B$ is compact and

$$(1-\varepsilon)B \subseteq \frac{1}{t}\widetilde{W}(t) \subseteq (1+\varepsilon)B, \text{ eventually a.s.}$$

for all $\epsilon > 0$, or $B = \mathbb{R}^d$, and

$$\{x \in \mathbb{R}^d : |x| \le K\} \subseteq \frac{1}{t}\widetilde{W}(t), \text{ eventually a.s.}$$

for all $K > 0$. Here $\widetilde{W}(t) = \{z + [-1/2, 1/2]^d : z \in W(t)\}$.

Studies of first-passage percolation brought many fascinating discoveries, including Kingman's celebrated subadditive ergodic theorem. In recent years interest has been focused on study of fluctuations of the set $\widetilde{W}(t)$ for large $t$. In spite of huge effort and some partial results achieved, it still remains a major task to establish rigorously conjectures predicted by Kardar–Parisi–Zhang theory about shape fluctuations in first passage percolation.

### Contact Processes

Introduced by Harris and conceived with biological interpretation, the contact process on $\mathbb{Z}^d$ is a continuous-time process taking values in the space of subsets of $\mathbb{Z}^d$. It is informally described as follows: particles are distributed in $\mathbb{Z}^d$ in such a way that each site is either empty or occupied by one particle. The evolution is Markovian: each particle disappears after an exponential time of parameter 1, independently from the others; at any time, each particle has a possibility to create a new particle at any of its empty neighboring sites, and does so with rate $\lambda > 0$, independently of everything else.

The question is then whether, starting from a finite population, the process will die out in finite time or whether it will survive forever with positive probability. The outcome will depend on the value of $\lambda$, and there is a critical value $\lambda_c$, such that for $\lambda \le \lambda_c$ process dies out, while for $\lambda > \lambda_c$ indeed there is survival, and in this case the shape of the population obeys a shape theorem similar to that of first-passage percolation.

The analogy with percolation is strong, the corresponding percolative picture being the following: in $\mathbb{Z}_+^{d+1}$, each edge is open with probability $p \in (0,1)$, and the question is whether there exists an infinite oriented path $\pi$ (i.e., a path along which the sum of the coordinates is increasing), composed of open edges. Once again, there is a critical parameter customarily denoted by $\boldsymbol{p}_c$, at which no such path exists (compare this to the open question of the continuity of the function $\theta$ at $p_c$ in dimensions $3 \le d \le 18$). This variation of percolation lies in a different universality class than the usual Bernoulli model.

### Invasion Percolation

Let $X(e) : e \in \mathcal{E}$ be independent random variables indexed by the edge set $\mathcal{E}$ of $\mathbb{Z}^d$, $d \ge 2$, each having uniform distribution in $[0,1]$. One constructs a sequence $C = \{C_i, i \ge 1\}$ of random connected subgraphs of the lattice in the following iterative way: the graph $C_0$ contains only the origin. Having defined $C_i$, one obtains $C_{i+1}$ by adding to $C_i$ an edge $e_{i+1}$ (with its outer lying end-vertex), chosen from the outer edge boundary of $C_i$ so as to minimize $X(e_{i+1})$. Still very little is known about the behavior of this process.

An interesting observation, relating $\theta(p_c)$ of usual percolation with the invasion dynamics, comes from CM Newman:

$$\theta(p_c) = 0 \Leftrightarrow P\{x \in C\} \to 0 \text{ as } |x| \to \infty$$

## Further Remarks

For a much more in-depth review of percolation on lattices and the mathematical methods involved in its study, and for the proofs of most of the results we could only point at, we refer the reader to the standard book of Grimmett (1999); another excellent general reference, and the only place to find some of the technical graph-theoretical details involved, is the book of Kesten (1982). More information in the case of graphs that are not lattices can be found in the lecture notes of Peres (1999).

For curiosity, the reader can refer to the first mention of a problem close to percolation, in the problem section of the first volume of the *American Mathematical Monthly* (problem 5, June 1894, submitted by D V Wood).

*See also:* Determinantal Random Fields; Stochastic Loewner Evolutions; Two-Dimensional Ising Model; Wulff Droplets.

## Further Reading


Alexander K, Chayes JT, and Chayes L (1990) The Wulff construction and asymptotics of the finite cluster distribution for two-dimensional Bernoulli percolation. *Communications in Mathematical Physics* 131: 1–50.

Beffara V (2004) Hausdorff dimensions for $SLE_6$. *Annals of Probability* 32: 2606–2629.

Benjamini I and Schramm O (1996) Percolation beyond $\mathbb{Z}^d$, many questions and a few answers. *Electronic Communications in Probability* 1(8): 71–82 (electronic).

Broadbent SR and Hammersley JM (1957) Percolation processes, I and II. *Mathematical Proceedings of the Cambridge Mathematical Society* 53, pp. 629–645.

Cerf R (2000) Large deviations for three dimensional supercritical percolation. *Astérisque, SMF* vol. 267.

Gaboriau D (2005) Invariant percolation and harmonic Dirichlet functions. *Geometric and Functional Analysis* (in print).

Grimmett G (1999) Percolation. *Grundlehren der Mathematischen Wissenschaften*, 2nd edn, vol. 321. Berlin: Springer.

Hammersley JM and Welsh DJA (1965) First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory. In: *Proc. Internat. Res. Semin*, Statist. Lab., Univ. California, Berkeley, CA, pp. 61–110. New York: Springer.

Hara T and Slade G (1990) Mean-field critical behaviour for percolation in high dimensions. *Communications in Mathematical Physics* 128: 333–391.

Kesten H (1980) The critical probability of bond percolation on the square lattice equals 1/2. *Communications in Mathematical Physics* 74: 41–59.

Kesten H (1982) *Percolation Theory for Mathematicians, Progress in Probability and Statistics*, vol. 2. Boston, MA: Birkhäuser.

Kesten H (1986) The incipient infinite cluster in two-dimensional percolation. *Probability Theory and Related Fields* 73: 369–394.

Kesten H (1987) Scaling relations for 2D-percolation. *Communications in Mathematical Physics* 109: 109–156.

Peres Y (1999) *Probability on Trees: An Introductory Climb*, Lectures on Probability Theory and Statistics (Saint-Flour, 1997), Lecture Notes in Math, vol. 1717, pp. 193–280. Berlin: Springer.

Russo L (1978) A note on percolation. *Zeitschift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 43: 39–48.

Sapoval B, Rosso M, and Gouyet J (1985) The fractal nature of a diffusion front and the relation to percolation. *Journal de Physique Lettres* 46: L146–L156.

Seymour PD and Welsh DJA (1978) Percolation probabilities on the square lattice. *Annals of Discrete Mathematics* 3: 227–245.

Smirnov S (2001) Critical percolation in the plane: conformal invariance, Cardy's formula, scaling limits. *Comptes-rendus de l'Académie des Sciences de Paris, Série I Mathématiques* 333: 239–244.

Smirnov S and Werner W (2001) Critical exponents for two-dimensional percolation. *Mathematical Research Letters* 8: 729–744.

van der Hofstad R and Járai AA (2004) The incipient infinite cluster for high-dimensional unoriented percolation. *Journal of Statistical Physics* 114: 625–663.


# Perturbation Theory and Its Techniques

**R J Szabo**, Heriot-Watt University, Edinburgh, UK



## Introduction

There are several equivalent formulations of the problem of quantizing an interacting field theory. The list includes canonical quantization, path-integral (or functional) techniques, stochastic quantization, "unified" methods such as the Batalin–Vilkovisky formalism, and techniques based on the realizations of field theories as low-energy limits of string theory. The problem of obtaining an exact nonperturbative description of a given quantum field theory is most often a very difficult one. Perturbative techniques, on the other hand, are abundant, and common to all of the quantization methods mentioned above is that they admit particle interpretations in this formalism.

The basic physical quantities that one wishes to calculate in a relativistic $(d+1)$-dimensional quantum field theory are the $S$-matrix elements

$$S_{ba} = {}_{\text{out}}\langle \psi_b(t)|\psi_a(t)\rangle_{\text{in}} \qquad [1]$$

between in and out states at large positive time $t$. The scattering operator $\mathsf{S}$ is then defined by writing [1] in terms of initial free-particle (descriptor) states as

$$S_{ba} =: \langle \psi_b(0)|\mathsf{S}|\psi_a(0)\rangle \qquad [2]$$

Suppose that the Hamiltonian of the given field theory can be written as $H = H_0 + H'$, where $H_0$ is the free part and $H'$ the interaction Hamiltonian. The time evolutions of the in and out states are governed by the total Hamiltonian $H$. They can be expressed in terms of descriptor states which evolve in time with $H_0$ in the interaction picture and correspond to free-particle states. This leads to the Dyson formula

$$\mathsf{S} = \mathrm{T}\exp\left(-\mathrm{i}\int_{-\infty}^{\infty} \mathrm{d}t\, H_I(t)\right) \qquad [3]$$

where T denotes time ordering and $H_I(t) = \int \mathrm{d}^d x \mathcal{H}_{\text{int}}(x,t)$ is the interaction Hamiltonian in the interaction picture, with $\mathcal{H}_{\text{int}}(x,t)$ the interaction Hamiltonian density, which deals with essentially free fields. This formula expresses $\mathsf{S}$ in terms of interaction-picture operators acting on free-particle states in [2] and is the first step towards Feynman perturbation theory.

For many analytic investigations, such as those which arise in renormalization theory, one is interested instead in the Green's functions of the quantum field theory, which measure the response of the system to an external perturbation. For definiteness, let us consider a free real scalar field theory in $d+1$ dimensions with Lagrangian density

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 + \mathcal{L}_{\rm int} \qquad [4]$$

where $\mathcal{L}_{\rm int}$ is the interaction Lagrangian density which we assume has no derivative terms. The interaction Hamiltonian density is then given by $\mathcal{H}_{\rm int} = -\mathcal{L}_{\rm int}$. Introducing a real scalar source $J(x)$, we define the normalized "partition function" through the vacuum expectation values,

$$Z[J] = \frac{\langle 0|\mathsf{S}[J]|0\rangle}{\langle 0|\mathsf{S}[0]|0\rangle} \qquad [5]$$

where $|0\rangle$ is the normalized perturbative vacuum state of the quantum field theory given by (4) (defined to be destroyed by all field annihilation operators), and

$$\mathsf{S}[J] = \mathrm{T}\exp\left(\mathrm{i}\int \mathrm{d}^{d+1}x(\mathcal{L}_{\rm int} + J(x)\phi(x))\right) \qquad [6]$$

from the Dyson formula. This partition function is the generating functional for all Green's functions of the quantum field theory, which are obtained from [5] by taking functional derivatives with respect to the source and then setting $J(x)=0$. Explicitly, in a formal Taylor series expansion in $J$ one has

$$Z[J] = \sum_{n=0}^{\infty}\frac{\mathrm{i}^n}{n!}\prod_{i=1}^{n}\int \mathrm{d}^{d+1}x_i\, J(x_i)\, G^{(n)}(x_1,\ldots,x_n) \qquad [7]$$

whose coefficients are the Green's functions

$$G^{(n)}(x_1,\ldots,x_n) := \frac{\langle 0|\mathrm{T}[\exp\left(\mathrm{i}\int \mathrm{d}^{d+1}x\mathcal{L}_{\rm int}\right)\phi(x_1)\cdots\phi(x_n)]|0\rangle}{\langle 0|\mathrm{T}\exp\left(\mathrm{i}\int \mathrm{d}^{d+1}x\mathcal{L}_{\rm int}\right)|0\rangle} \qquad [8]$$

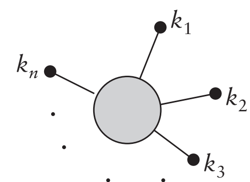
It is customary to work in momentum space by introducing the Fourier transforms

$$\tilde{J}(k) = \int \mathrm{d}^{d+1}x\, \mathrm{e}^{\mathrm{i}k\cdot x}J(x)$$
$$\tilde{G}^{(n)}(k_1,\ldots,k_n) = \prod_{i=1}^{n}\int \mathrm{d}^{d+1}x_i\, \mathrm{e}^{\mathrm{i}k_i\cdot x_i}G^{(n)}(x_1,\ldots,x_n) \qquad [9]$$

in terms of which the expansion [7] reads

$$Z[J] = \sum_{n=0}^{\infty}\frac{\mathrm{i}^n}{n!}\prod_{i=1}^{n}\int \frac{\mathrm{d}^{d+1}k_i}{(2\pi)^{d+1}}\tilde{J}(-k_i) \times \tilde{G}^{(n)}(k_1,\ldots,k_n) \qquad [10]$$

The generating functional [10] can be written as a sum of Feynman diagrams with source insertions. Diagrammatically, the Green's function is an infinite series of graphs which can be represented symbolically as

$\tilde{G}^{(n)}(k_1,\ldots,k_n) =$  [11]


where the $n$ external lines denote the source insertions of momenta $k_i$ and the bubble denotes the sum over all Feynman diagrams constructed from the interaction vertices of $\mathcal{L}_{\rm int}$.

This procedure is, however, rather formal in the way that we have presented it, for a variety of reasons. First of all, by Haag's theorem, it follows that the interaction representation of a quantum field theory does not exist unless a cutoff regularization is introduced into the interaction term in the Lagrangian density (this regularization is described explicitly below). The addition of this term breaks translation covariance. This problem can be remedied via a different definition of the regularized Green's functions, as we discuss below. Furthermore, the perturbation series of a quantum field theory is typically divergent. The expansion into graphs is, at best, an asymptotic series which is Borel summable. These shortcomings will not be emphasized any further in this article. Some mathematically rigorous approaches to perturbative quantum field theory can be found in the bibliography.

The Green's functions can also be used to describe scattering amplitudes, but there are two important differences between the graphs [11] and those which appear in scattering theory. In the present case, external lines carry propagators, that is, the free-field Green's functions

$$\Delta(x-y) = \langle 0|\,\mathrm{T}[\phi(x)\,\phi(y)]\,|0\rangle = \left\langle x\middle|\left(\Box + m^2\right)^{-1}\middle|y\right\rangle = \int \frac{\mathrm{d}^{d+1}p}{(2\pi)^{d+1}}\frac{\mathrm{i}}{p^2 - m^2 + \mathrm{i}\epsilon}\mathrm{e}^{-\mathrm{i}p\cdot(x-y)} \qquad [12]$$

where $\epsilon \to 0^+$ regulates the mass shell contributions, and their momenta $k_i$ are off-shell in general ($k_i^2 \neq m^2$). By the LSZ theorem, the S-matrix element

is then given by the multiple on-shell residue of the Green's function in momentum space as

$$\langle k'_1,\ldots,k'_n|\,\mathsf{S}-\mathbb{1}|k_1,\ldots,k_l\rangle = \lim_{\substack{k'_1,\ldots,k'_n\to m^2\\ k_1,\ldots,k_l\to m^2}} \prod_{i=1}^{n}\frac{1}{\mathrm{i}\sqrt{c'_i}}\left(k_i'^2-m^2\right)\prod_{j=1}^{l}\frac{1}{\mathrm{i}\sqrt{c_j}}\left(k_j^2-m^2\right) \times \tilde{G}^{(n+m)}\left(-k'_1,\ldots,-k'_n,k_1,\ldots,k_l\right) \quad [13]$$

where $\mathrm{i}c'_i, \mathrm{i}c_j$ are the residues of the corresponding particle poles in the exact two-point Green's function.

This article deals with the formal development and computation of perturbative scattering amplitudes in relativistic quantum field theory, along the lines outlined above. Initially we deal only with real scalar field theories of the sort [4] in order to illustrate the concepts and technical tools in as simple and concise a fashion as possible. These techniques are common to most quantum field theories. Fermions and gauge theories are then separately treated afterwards, focusing on the methods which are particular to them.

## Diagrammatics

The pinnacle of perturbation theory is the technique of Feynman diagrams. Here we develop the basic machinery in a quite general setting and use it to analyze some generic features of the terms comprising the perturbation series.

### Wick's Theorem

The Green's functions [8] are defined in terms of vacuum expectation values of time-ordered products of the scalar field $\phi(x)$ at different spacetime points. Wick's theorem expresses such products in terms of normal-ordered products, defined by placing each field creation operator to the right of each field annihilation operator, and in terms of two-point Green's functions [12] of the free-field theory (propagators). The consequence of this theorem is the Haffnian formula

$$\langle 0|\,\mathrm{T}[\phi(x_1)\cdots\phi(x_n)]|0\rangle = \begin{cases} 0 & n=2k-1 \\ \displaystyle\sum_{\pi\in S_{2k}}\prod_{i=1}^{k}\langle 0|\,\mathrm{T}\left[\phi(x_{\pi(2i-1)})\,\phi(x_{\pi(2i)})\right]|0\rangle & n=2k \end{cases} \quad [14]$$

The formal Taylor series expansion of the scattering operator $\mathsf{S}$ may now be succinctly summarized into a diagrammatic notation by using Wick's theorem. For each spacetime integration $\int \mathrm{d}^{d+1}x_i$ we introduce a vertex with label $i$, and from each vertex there emanate some lines corresponding to field insertions at the point $x_i$. If the operators represented by two lines appear in a two-point function according to [14], that is, they are contracted, then these two lines are connected together. The $\mathsf{S}$ operator is then represented as a sum over all such Wick diagrams, bearing in mind that topologically equivalent diagrams correspond to the same term in $\mathsf{S}$. Two diagrams are said to have the same pattern if they differ only by a permutation of their vertices. For any diagram $\mathfrak{D}$ with $n(\mathfrak{D})$ vertices, the number of ways of interchanging vertices is $n(\mathfrak{D})!$. The number of diagrams per pattern is always less than this number. The symmetry number $S(\mathfrak{D})$ of $\mathfrak{D}$ is the number of permutations of vertices that give the same diagram. The number of diagrams with the pattern of $\mathfrak{D}$ is then $n(\mathfrak{D})!/S(\mathfrak{D})$.

In a given pattern, we write the contribution to $\mathsf{S}$ of a single diagram $\mathfrak{D}$ as

$$\frac{1}{n(\mathfrak{D})!}\,{:}\theta(\mathfrak{D}){:}$$

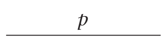
where the combinatorial factor comes from the Taylor expansion of $\mathsf{S}$, the large colons denote normal ordering of quantum operators, and ${:}\theta(\mathfrak{D}){:}$ contains spacetime integrals over normal-ordered products of the fields. Then all diagrams with the pattern of $\mathfrak{D}$ contribute ${:}\theta(\mathfrak{D}){:}/S(\mathfrak{D})$ to $\mathsf{S}$. Only the connected diagrams $\mathfrak{D}_r, r\in\mathbb{N}$ (those in which every vertex is connected to every other vertex) contribute and we can write the scattering operator in a simple form which eliminates contributions from all disconnected diagrams as

$$\mathsf{S} = {:}\exp\left(\sum_{r=1}^{\infty}\frac{\theta(\mathfrak{D}_r)}{S(\mathfrak{D}_r)}\right){:} \quad [15]$$

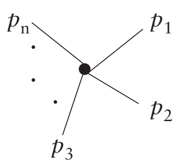
### Feynman Rules

Feynman diagrams in momentum space are defined from the Wick diagrams above by dropping the labels on vertices (and also the symmetry factors $S(\mathfrak{D})^{-1}$), and by labeling the external lines by the momenta of the initial and final particles that the corresponding field operators annihilate. In a spacetime interpretation, external lines

represent on-shell physical particles while internal lines of the graph represent off-shell virtual particles ($k^2 \neq m^2$). Physical particles interact via the exchange of virtual particles. An arbitrary diagram is then calculated via the Feynman rules:




$$= \int \frac{d^{d+1}p}{(2\pi)^{d+1}} \frac{i}{p^2 - m^2 + i\epsilon}$$

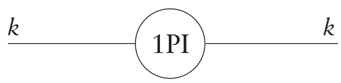



$$= ig\,(2\pi)^{d+1}\,\delta^{(d+1)}(p_1 + \cdots + p_n) \qquad [16]$$

for a monomial interaction $\mathcal{L}_{\text{int}} = (g/n!)\phi^n$.

## Irreducible Green's Functions

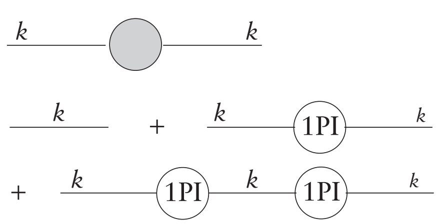
A one-particle irreducible (1PI) or proper Green's function is given by a sum of diagrams in which each diagram cannot be separated by cutting one internal line. In momentum space, it is defined without the overall momentum conservation delta-function factors and without propagators on external lines. For example, the two particle 1PI Green's function




$$=: \Sigma(k) \qquad [17]$$

is called the self-energy. If $G(k)$ is the complete two-point function in momentum space, then one has

$G(k) :=$




$$= \frac{i}{k^2 - m^2 - \Sigma(k)} \qquad [18]$$

and thus it suffices to calculate only 1PI diagrams.

The 1PI effective action, defined by the Legendre transformation $\Gamma[\phi] := -i \ln Z[J] - \int d^{d+1}x J(x)\phi(x)$ of [5], is the generating functional for proper vertex functions and it can be represented as a functional of only the vacuum expectation value of the field $\phi$, that is, its classical value. In the semiclassical (WKB) approximation, the one-loop effective action is given by

$$\begin{aligned}
\Gamma[\phi] &= S[\phi] + \frac{i\hbar}{2} \mathrm{Tr} \ln(1 + \Delta V''[\phi]) + O(\hbar^2) \\
&= S[\phi] + i\hbar \sum_{n=1}^{\infty} \frac{(-1)^n}{2n} \\
&\quad \times \prod_{i=1}^{n} \int d^{d+1}x_i \Delta(x_i - x_{i+1}) V''[\phi(x_{i+1})] \\
&\quad + O(\hbar^2)
\end{aligned} \qquad [19]$$

where we have denoted $S[\phi] = \int d^{d+1}x\mathcal{L}$ and $V[\phi] = -\mathcal{L}_{\text{int}}$, and for each term in the infinite series we define $x_{n+1} := x_1$. The first term in [19] is the classical contribution and it can be represented in terms of connected tree diagrams. The second term is the sum of contributions of one-loop diagrams constructed from $n$ propagators $-i\Delta(x-y)$ and $n$ vertices $-iV''[\phi]$. The expansion may be carried out to all orders in terms of connected Feynman diagrams, and the result of the above Legendre transformation is to select only the one-particle irreducible diagrams and to replace the classical value of $\phi$ by an arbitrary argument. All information about the quantum field theory is encoded in this effective action.

## Parametric Representation

Consider an arbitrary proper Feynman diagram $\mathfrak{D}$ with $n$ internal lines and $v$ vertices. The number, $\ell$, of independent loops in the diagram is the number of independent internal momenta in $\mathfrak{D}$ when conservation laws at each vertex have been taken into account, and it is given by $\ell = n + 1 - v$. There is an independent momentum integration variable $k_i$ for each loop, and a propagator for each internal line as in [16]. The contribution of $\mathfrak{D}$ to a proper Green's function with $r$ incoming external momenta $p_i$, with $\sum_{i=1}^{r} p_i = 0$, is given by

$$\begin{aligned}
\tilde{I}_{\mathfrak{D}}(p) &= \frac{V(\mathfrak{D})}{S(\mathfrak{D})} \prod_{i=1}^{n} \int \frac{d^{d+1}k_i}{(2\pi)^{d+1}} \frac{i}{k_i^2 - m^2 + i\epsilon} \\
&\quad \times \prod_{j=1}^{v} (2\pi)^{d+1} \delta^{(d+1)}\left(P_j - K_j\right)
\end{aligned} \qquad [20]$$

where $V(\mathfrak{D})$ contains all contributions from the interaction vertices of $\mathcal{L}_{\text{int}}$, and $P_j$ (resp. $K_j$) is the sum of incoming external momenta $p_{l_j}$ (resp. internal momenta $k_{l_j}$) at vertex $j$ with respect to a fixed chosen orientation of the lines of the graph. After resolving the delta-functions in terms of independent internal loop momenta $k_1, \ldots, k_\ell$ and dropping the overall momentum conservation

delta-function along with the symmetry and vertex factors in [20], one is left with a set of momentum space integrals

$$I_{\mathfrak{D}}(p) = \prod_{i=1}^{\ell} \int \frac{\mathrm{d}^{d+1} k_i}{(2\pi)^{d+1}} \prod_{j=1}^{n} \frac{\mathrm{i}}{a_j(k,p) + \mathrm{i}\epsilon} \qquad [21]$$

where $a_j(k,p)$ are functions of both the internal and external momenta.

It is convenient to exponentiate propagators using the Schwinger parametrization

$$\frac{\mathrm{i}}{a_j + \mathrm{i}\epsilon} = \int_0^{\infty} \mathrm{d}\alpha_j \, \mathrm{e}^{\mathrm{i}\alpha_j(a_j + \mathrm{i}\epsilon)} \qquad [22]$$

and after some straightforward manipulations one may write the Feynman parametric formula

$$\prod_{j=1}^{n} \frac{\mathrm{i}}{a_j(k,p) + \mathrm{i}\epsilon} = (n-1)! \prod_{j=1}^{n} \int_0^1 \mathrm{d}\alpha_j \frac{\delta\left(1 - \sum_j \alpha_j\right)}{D_{\mathfrak{D}}(k;\alpha,p)^n} \qquad [23]$$

where $D_{\mathfrak{D}}(k;\alpha,p) := \sum_j \alpha_j [a_j(k,p) + \mathrm{i}\epsilon]$ is generically a quadratic form

$$D_{\mathfrak{D}}(k;\alpha,p) = \frac{1}{2} \sum_{i,j=1}^{\ell} k_i \cdot Q_{ij}(\alpha) k_j + \sum_{i=1}^{\ell} L_i(p) \cdot k_i + \lambda(p^2) \qquad [24]$$

The positive symmetric matrix $Q_{ij}$ is independent of the external momenta $p_l$, invertible, and has nonzero eigenvalues $Q_1, \ldots, Q_\ell$. The vectors $L_i^\mu$ are linear combinations of the $p_j^\mu$, while $\lambda(p^2)$ is a function of only the Lorentz invariants $p_i^2$. After some further elementary manipulations, the loop diagram contribution [21] may be written as

$$I_{\mathfrak{D}}(p) = (n-1)! \prod_{j=1}^{n} \int_0^1 \mathrm{d}\alpha_j \prod_{i=1}^{\ell} \frac{1}{Q_i(\alpha)^2} \int \frac{\mathrm{d}^{d+1} k_i}{(2\pi)^{d+1}} \delta\left(1 - \sum_j \alpha_j\right) \times \left(\frac{1}{2} \sum_i k_i^2 + \lambda(p^2) - \frac{1}{2} \sum_{i,j} L_i(p) \cdot Q^{-1}(\alpha)_{ij} L_j(p)\right)^{-n} \qquad [25]$$

Finally, the integrals over the loop momenta $k_i$ may be performed by Wick-rotating them to Euclidean space and using the fact that the combination of $\ell$ integrations in $\mathbb{R}^{d+1}$ has $O((d+1)\ell)$ rotational invariance. The contribution from the entire Feynman diagram $\mathfrak{D}$ thereby reduces to the calculation of the parametric integrals:

$$I_{\mathfrak{D}}(p) = \frac{\Gamma\left(n - \frac{(d+1)\ell}{2}\right)}{(2\pi)^{\frac{(d+1)\ell}{2}} \mathrm{i}^{d\ell}} \prod_{j=1}^{n} \int_0^1 \mathrm{d}\alpha_j \prod_{i=1}^{\ell} \frac{1}{Q_i(\alpha)^2} \times \frac{\delta\left(1 - \sum_j \alpha_j\right)}{\left(\lambda(p^2) - \frac{1}{2} \sum_{i,j} L_i(p) \cdot Q^{-1}(\alpha)_{ij} L_j(p)\right)^{\frac{n-(d+1)\ell}{2}}} \qquad [26]$$

where $\Gamma(s)$ is the Euler gamma-function.

## Regularization

The parametric representation [26] is generically convergent when $2n - (d+1)\ell > 0$. When divergent, the infinities arise from the lower limits of integration $\alpha_j \to 0$. This is just the parametric representation of the large-$k$ divergence of the original Feynman amplitude [20]. Such ultraviolet divergences plague the very meaning of a quantum field theory and must be dealt with in some way. We will now quickly tour the standard methods of ultraviolet regularization for such loop integrals, which is a prelude to the renormalization program that removes the divergences (in a renormalizable field theory). Here we consider regularization simply as a means of justification for the various formal manipulations that are used in arriving at expressions such as [26].

### Momentum Cutoff

Cutoff regularization introduces a mass scale $\Lambda$ into the quantum field theory and throws away the Fourier modes of the fields for spatial momenta $\boldsymbol{k}$ with $|\boldsymbol{k}| > \Lambda$. This regularization spoils Lorentz invariance. It is also nonlocal. For example, if we restrict to a hypercube in momentum space, so that $|k_i| < \Lambda$ for $i = 1, \ldots, d$, then

$$\int_{|\boldsymbol{k}| > \Lambda} \frac{\mathrm{d}^d \boldsymbol{k}}{(2\pi)^d} \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}} = \prod_{i=1}^{d} \frac{\sin(\Lambda x^i)}{\pi x^i}$$

which is a delta-function in the limit $\Lambda \to \infty$ but is nonlocal for $\Lambda < \infty$. The regularized field theory is finite order by order in perturbation theory and depends on the cutoff $\Lambda$.

#### Lattice Regularization

We can replace the spatial continuum by a lattice $\mathfrak{L}$ of rank $d$ and define a Lagrangian on $\mathfrak{L}$ by

$$L_{\mathfrak{L}} = \frac{1}{2}\sum_{i\in \mathrm{S}(\mathfrak{L})} \dot{\phi}_i^2 + J\sum_{\langle i,j\rangle\in \mathrm{L}(\mathfrak{L})} \phi_i\phi_j + \sum_{i\in \mathrm{S}(\mathfrak{L})} V(\phi_i) \quad [27]$$

where $\mathrm{S}(\mathfrak{L})$ is the set of sites $i$ of the lattice on each of which is situated a time-dependent function $\phi_i$, and $L_{\mathfrak{L}}$ is the collection of links connecting pairs $\langle i,j\rangle$ of nearest-neighbor sites $i,j$ on $\mathfrak{L}$. The regularized field theory is now local, but still has broken Lorentz invariance. In particular, it suffers from broken rotational symmetry. If $\mathfrak{L}$ is hypercubic with lattice spacing $a$, that is, $\mathfrak{L} = (\mathbb{Z}a)^d$, then the momentum cutoff is at $\Lambda = a^{-1}$.

#### Pauli–Villars Regularization

We can replace the propagator $\mathrm{i}(k^2 - m^2 + \mathrm{i}\epsilon)^{-1}$ by $\mathrm{i}(k^2 - m^2 + \mathrm{i}\epsilon)^{-1} + \mathrm{i}\sum_{j=1}^{N} c_j(k^2 - M_j^2 + \mathrm{i}\epsilon)^{-1}$, where the masses $M_j \gg m$ are identified with the momentum cutoff as $\min\{M_j\} = \Lambda \to \infty$. The mass-dependent coefficients $c_j$ are chosen to make the modified propagator decay rapidly as $(k^2)^{-N-1}$ at $k \to \infty$, which gives the $N$ equations $(m^2)^i + \sum_j c_j(M_j^2)^i = 0, i = 0, 1, \ldots, N-1$. This regularization preserves Lorentz invariance (and other symmetries that the field theory may possess) and is local in the following sense. The modified propagator can be thought of as arising through the alteration of the Lagrangian density [4] by $N$ additional scalar fields $\varphi_j$ of masses $M_j$ with

$$\mathcal{L}_{\mathrm{PV}} = \tfrac{1}{2}\partial_\mu\phi\partial^\mu\phi - \tfrac{1}{2}m^2\phi^2 + \sum_{j=1}^{N}\left(\tfrac{1}{2}\partial_\mu\varphi_j\partial^\mu\varphi_j - \tfrac{1}{2}M_j^2\varphi_j^2\right) + \mathcal{L}_{\mathrm{int}}[\Phi] \quad [28]$$

where $\Phi := \phi + \sum_j \sqrt{c_j}\varphi_j$. The contraction of the $\Phi$ field thus produces the required propagator. However, the $c_j$'s as computed above are generically negative numbers and so the Lagrangian density [28] is not Hermitian (as $\Phi \neq \Phi^\dagger$). It is possible to make [28] formally Hermitian by redefining the inner product on the Hilbert space of physical states, but this produces negative-norm states. This is no problem at energy scales $E \ll M_j$ on which the extra particles decouple and the negative probability states are invisible.

#### Dimensional Regularization

Consider a Euclidean space integral $\int \mathrm{d}^4k(k^2 + a^2)^{-r}$ arising after Wick rotation from some loop diagram in $(3+1)$-dimensional scalar field theory. We replace this integral by its $D$-dimensional version

$$\int \frac{\mathrm{d}^D k}{(k^2 + a^2)^r} = \frac{\pi^{D/2}(a^2)^{D/2-r}}{(r-1)!}\Gamma\left(r - \frac{D}{2}\right) \quad [29]$$

This integral is absolutely convergent for $D < 2r$. We can analytically continue the result of this integration to the complex plane $D \in \mathbb{C}$. As an analytic function, the only singularities of the Euler function $\Gamma(z)$ are poles at $z = 0, -1, -2, \ldots$. In particular, $\Gamma(z)$ has a simple pole at $z = 0$ of residue 1. If we write $D = 4 + \epsilon$ with $|\epsilon| \to 0$, then the integral [29] is proportional to $\Gamma(r - 2 - \epsilon/2)$ and $\epsilon$ plays the role of the regulator here. This regularization is Lorentz invariant (in $D$ dimensions) and is distinguished as having a dimensionless regularization parameter $\epsilon$. This parameter is related to the momentum cutoff $\Lambda$ by $\epsilon^{-1} = \ln(\Lambda/m)$, so that the limit $\epsilon \to 0$ corresponds to $\Lambda \to \infty$.

#### Infrared Divergences

Thus far we have only considered the ultraviolet behavior of loop amplitudes in quantum field theory. When dealing with massless particles ($m = 0$ in [4]) one has to further worry about divergences arising from the $k \to 0$ regions of Feynman integrals. After Wick rotation to Euclidean momenta, one can show that no singularities arise in a given Feynman diagram as some of its internal masses vanish provided that all vertices have superficial degree of divergence $d + 1$, the external momenta are not exceptional (i.e., no partial sum of the incoming momenta $p_i$ vanishes), and there is at most one soft external momentum. This result assumes that renormalization has been carried out at some fixed Euclidean point. The extension of this property when the external momenta are continued to physical on-shell values is difficult. The Kinoshita–Lee–Nauenberg theorem states that, as a consequence of unitarity, transition probabilities in a theory involving massless particles are finite when the sum over all degenerate states (initial and final) is taken. This is true order by order in perturbation theory in bare quantities or if minimal subtraction renormalization is used (to avoid infrared or mass singularities in the renormalization constants).

### Fermion Fields

We will now leave the generalities of our pure scalar field theory and start considering the extensions of our previous considerations to other types of particles. Henceforth we will primarily deal with the case of $(3+1)$-dimensional spacetime. We begin by indicating how the rudiments of perturbation

theory above apply to the case of Dirac fermion fields. The Lagrangian density is

$$\mathcal{L}_F = \overline{\psi}(\mathrm{i}\partial\!\!\!/ - m)\psi + \mathcal{L}' \quad [30]$$

where $\psi$ are four-component Dirac fermion fields in $3+1$ dimensions, $\overline{\psi} := \psi^\dagger\gamma^0$ and $\partial\!\!\!/ = \gamma^\mu\partial_\mu$ with $\gamma^\mu$ the generators of the Clifford algebra $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$. The Lagrangian density $\mathcal{L}'$ contains couplings of the Dirac fields to other field theories, such as the scalar field theories considered previously.

Wick's theorem for anticommuting Fermi fields leads to the Pfaffian formula

$$\langle 0|\mathrm{T}[\psi(1)\cdots\psi(n)]|0\rangle = \begin{cases} 0, \; n = 2k-1 \\ \dfrac{1}{2^k k!}\sum_{\pi\in S_{2k}} \mathrm{sgn}(\pi) \\ \times \prod_{i=1}^{k} \langle 0|\mathrm{T}[\psi(\pi(2i-1))\psi(\pi(2i))]|0\rangle \\ n = 2k \end{cases} \quad [31]$$

where for compactness we have written in the argument of $\psi(i)$ the spacetime coordinate, the Dirac index, and a discrete index which distinguishes $\psi$ from $\overline{\psi}$. The nonvanishing contractions in [31] are determined by the free-fermion propagator

$$\begin{aligned}\Delta_F(x-y) &= \langle 0|\mathrm{T}[\psi(x)\overline{\psi}(y)]|0\rangle \\ &= \left\langle x\middle|(\mathrm{i}\partial\!\!\!/ - m)^{-1}\middle|y\right\rangle \\ &= \mathrm{i}\int \frac{\mathrm{d}^4p}{(2\pi)^4}\frac{p\!\!\!/ + m}{p^2 - m^2 + \mathrm{i}\epsilon}\mathrm{e}^{-\mathrm{i}p\cdot(x-y)}\end{aligned} \quad [32]$$

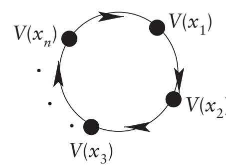
Perturbation theory now proceeds exactly as before. Suppose that the coupling Lagrangian density in [30] is of the form $\mathcal{L}' = \overline{\psi}(x)V(x)\psi(x)$. Both the Dyson formula [3] and the diagrammatic formula [15] are formally the same in this instance. For example, in the formal expansion in powers of $\int \mathrm{d}^4x\mathcal{L}'$, the vacuum-to-vacuum amplitude (the denominator in [5]) will contain field products of the form

$$\prod_{i=1}^{n}\int \mathrm{d}^4x_i\langle 0|\mathrm{T}[\overline{\psi}(x_i)V(x_i)\psi(x_i)]|0\rangle$$

which correspond to fermion loops. Before applying Wick's theorem, the fields must be rearranged as

$$\mathrm{tr}\prod_{i=1}^{n} V(x_i)\psi(x_i)\overline{\psi}(x_{i+1})$$

(with $x_{n+1} := x_1$), where tr is the $4\times 4$ trace over spinor indices. This reordering introduces the familiar minus sign for a closed fermion loop, and one has




$$\begin{aligned}= (-)\prod_{i=1}^{n}\int \mathrm{d}^4x_i \\ \times\, \mathrm{tr}\prod_{j=1}^{n-1}\Delta_F(x_j - x_{j+1}) \\ \times\, V(x_{j+1})\,\Delta_F(x_{j+1} - x_{j+2})\end{aligned} \quad [33]$$

Feynman rules are now described as follows. Fermion lines are oriented to distinguish a particle from its corresponding antiparticle, and carry both a four-momentum label $p$ as well as a spin polarization index $r = 1, 2$. Incoming fermions (resp. antifermions) are described by the wave functions $u_p^{(r)}$ (resp. $\overline{v}_p^{(r)}$), while outgoing fermions (resp. antifermions) are described by the wave functions $\overline{u}_p^{(r)}$ (resp. $v_p^{(r)}$). Here $u_p^{(r)}$ and $v_p^{(r)}$ are the classical spinors, that is, the positive and negative-energy solutions of the Dirac equation $(p\!\!\!/ - m)u_p^{(r)} = (p\!\!\!/ + m)v_p^{(r)} = 0$. Matrices are multiplied along a Fermi line, with the head of the arrow on the left. Closed fermion loops produce an overall minus sign as in [33], and the multiplication rule gives the trace of Dirac matrices along the lines of the loop. Unpolarized scattering amplitudes are summed over the spins of final particles and averaged over the spins of initial particles using the completeness relations for spinors

$$\sum_{r=1,2} u_p^{(r)}\overline{u}_p^{(r)} = p\!\!\!/ + m, \quad \sum_{r=1,2} v_p^{(r)}\overline{v}_p^{(r)} = p\!\!\!/ - m \quad [34]$$

leading to basis-independent results. Polarized amplitudes are computed using the spinor bilinears $\overline{u}_p^{(r)}\gamma^\mu u_p^{(s)} = \overline{v}_p^{(r)}\gamma^\mu v_p^{(s)} = 2p^\mu\delta^{rs}, \overline{u}_p^{(r)}u_p^{(s)} = -\overline{v}_p^{(r)}v_p^{(s)} = 2m\,\delta^{rs}$, and $\overline{u}_p^{(r)}v_p^{(s)} = 0$.

When calculating fermion loop integrals using dimensional regularization, one utilizes the Dirac algebra in $D$ dimensions

$$\begin{aligned}\gamma^\mu\gamma_\mu &= \eta^\mu_\mu = D \\ \gamma_\mu p\!\!\!/\gamma^\mu &= (2-D)p\!\!\!/ \\ \gamma^\mu p\!\!\!/ k\!\!\!/\gamma_\mu &= 4p\cdot k + (D-4)p\!\!\!/ k\!\!\!/ \\ \gamma^\mu p\!\!\!/ k\!\!\!/ q\!\!\!/\gamma_\mu &= -2q\!\!\!/ k\!\!\!/ p\!\!\!/ - (D-4)p\!\!\!/ k\!\!\!/ q\!\!\!/\end{aligned} \quad [35]$$

$$\mathrm{tr}\mathbb{1} = 4, \mathrm{tr}\gamma^{\mu_1}\cdots\gamma^{\mu_{2k-1}} = 0, \mathrm{tr}\gamma^\mu\gamma^\nu = 4\eta^{\mu\nu}$$

$$\mathrm{tr}\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma = 4(\eta^{\mu\nu}\eta^{\rho\sigma} - \eta^{\mu\rho}\eta^{\nu\sigma} + \eta^{\mu\sigma}\eta^{\nu\rho})$$

Specific to $D=4$ dimensions are the trace identities

$$\begin{aligned} \mathrm{tr}\gamma^5 &= \mathrm{tr}\gamma^\mu\gamma^\nu\gamma^5 = 0, \\ \mathrm{tr}\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma\gamma^5 &= -4\mathrm{i}\epsilon^{\mu\nu\rho\sigma} \end{aligned} \qquad [36]$$

where $\gamma^5 := \mathrm{i}\gamma^0\gamma^1\gamma^2\gamma^3$. Finally, loop diagrams evaluated with the fermion propagator [32] require a generalization of the momentum space integral [29] given by

$$\begin{aligned} &\int \frac{\mathrm{d}^D k}{(2\pi)^D} \frac{1}{(k^2 + 2k\cdot p + a^2 + \mathrm{i}\epsilon)^r} \\ &= \frac{\mathrm{i}(-\pi)^{D/2}\Gamma\left(r - \frac{D}{2}\right)}{(2\pi)^D (r-1)!} \frac{1}{(a^2 - p^2 + \mathrm{i}\epsilon)^{r-D/2}} \end{aligned} \qquad [37]$$

From this formula we can extract expressions for more complicated Feynman integrals which are tensorial, that is, which contain products of momentum components $k^\mu$ in the numerators of their integrands, by differentiating [37] with respect to the external momentum $p^\mu$.

## Gauge Fields

The issues we have dealt with thus far have interesting difficulties when dealing with gauge fields. We will now discuss some general aspects of the perturbation expansion of gauge theories using as prototypical examples quantum electrodynamics (QED) and quantum chromodynamics (QCD) in four spacetime dimensions.

### Quantum Electrodynamics

Consider the QED Lagrangian density

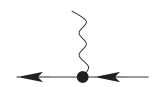
$$\begin{aligned} \mathcal{L}_{\mathrm{QED}} &= -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ &\quad + \tfrac{1}{2}\mu^2 A_\mu A^\mu + \overline{\psi}(\mathrm{i}\partial\!\!\!/ - eA\!\!\!/ - m)\psi \end{aligned} \qquad [38]$$

where $A_\mu$ is a U(1) gauge field in $3+1$ dimensions and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is its field strength tensor. We have added a small mass term $\mu \to 0$ for the gauge field, which at the end of calculations should be taken to vanish in order to describe real photons (as opposed to the soft photons described by [38]). This is done in order to cure the infrared divergences generated in scattering amplitudes due to the masslessness of the photon, that is, the long-range nature of the electromagnetic interaction. The Bloch–Nordsieck theorem in QED states that infrared divergences cancel for physical processes, that is, for processes with an arbitrary number of undetectable soft photons.

Perturbation theory proceeds in the usual way via the Dyson formula, Wick's theorem, and Feynman diagrams. The gauge field propagator is given by

$$\begin{aligned} &\langle 0|\mathrm{T}\left[A_\mu(x)A_\nu(y)\right]|0\rangle \\ &= \langle x|\left[\eta_{\mu\nu}\left(\Box + \mu^2\right) - \partial_\mu\partial_\nu\right]^{-1}|y\rangle \\ &= \mathrm{i}\int \frac{\mathrm{d}^4 p}{(2\pi)^4} \frac{-\eta_{\mu\nu} + \frac{p_\mu p_\nu}{\mu^2}}{p^2 - \mu^2 + \mathrm{i}\epsilon} \mathrm{e}^{-\mathrm{i}p\cdot(x-y)} \end{aligned} \qquad [39]$$

and is represented by a wavy line. The fermion–fermion–photon vertex is


$\mu$ $= -\mathrm{i}e\gamma_\mu$ [40]

An incoming (resp. outgoing) soft photon of momentum $k$ and polarization $r$ is described by the wave function $e_\mu^{(r)}(k)$ (resp. $e_\mu^{(r)}(k)^*$), where the polarization vectors $e_\mu^{(r)}(k), r = 1, 2, 3$ solve the vector field wave equation $(\Box + \mu^2)A_\mu = \partial_\mu A^\mu = 0$ and obey the orthonormality and completeness conditions

$$\begin{aligned} e^{(r)}(k)^* \cdot e^{(s)}(k) &= -\delta^{rs} \\ \sum_{r=1}^{3} e_\mu^{(r)}(k) e_\nu^{(r)}(k)^* &= -\eta_{\mu\nu} + \frac{k_\mu k_\nu}{\mu^2} \end{aligned} \qquad [41]$$

along with $k \cdot e^{(r)}(k) = 0$. All vector indices are contracted along the lines of the Feynman graph. All other Feynman rules are as previously.

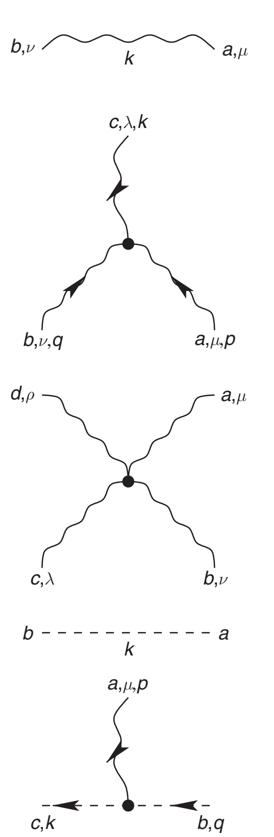
### Quantum Chromodynamics

Consider nonabelian gauge theory in $3+1$ dimensions minimally coupled to a set of fermion fields $\psi^A, A = 1, \ldots, N_f$, each transforming in the fundamental representation of the gauge group $G$ whose generators $T^a$ satisfy the commutation relations $[T^a, T^b] = f^{abc}T^c$. The Lagrangian density is given by

$$\begin{aligned} \mathcal{L}_{\mathrm{QCD}} &= -\frac{1}{4}F^a_{\mu\nu}F^{a\,\mu\nu} + \frac{1}{2\alpha}\left(\partial_\mu A^{a\,\mu}\right)^2 + \partial_\mu\overline{\eta}\,D^\mu\eta \\ &\quad + \sum_{A=1}^{N_f} \overline{\psi}^A(\mathrm{i}D\!\!\!/ - m_A)\psi^A \end{aligned} \qquad [42]$$

where $F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc}A^b_\mu A^c_\nu$ and $D_\mu = \partial_\mu + \mathrm{i}eR(T^a)A^a_\mu$, with $R$ the pertinent representation of $G$ ($R(T^a)_{bc} = f^a_{bc}$ for the adjoint representation and $R(T^a) = T^a$ for the fundamental representation). The first term is the Yang–Mills Lagrangian density, the second term is the covariant gauge-fixing term, and the third term contains the Faddeev–Popov ghost fields $\eta$ which transform in the adjoint representation of the gauge group.

Feynman rules are straightforward to write down and are given in **Figure 1** where wavy lines




$$-\mathrm{i}\,\delta^{ab}\,\frac{\eta_{\mu\nu}+(\alpha-1)\,\dfrac{k_\mu k_\nu}{k^2}}{k^2+\mathrm{i}\epsilon}$$

$$e\,f^{abc}\left[\eta_{\lambda\mu}(p_\nu-k_\nu)+\eta_{\mu\nu}\,(q_\lambda-p_\lambda)+\eta_{\nu\lambda}(k_\mu-q_\mu)\right]$$

$$-\mathrm{i}\,e^2\left[f^{eab}\,f^{ecd}\,(\eta_{\mu\lambda}\,\eta_{\nu\rho}-\eta_{\mu\rho}\,\eta_{\nu\lambda}) + f^{eac}\,f^{ebd}\,(\eta_{\mu\nu}\,\eta_{\lambda\rho}-\eta_{\mu\rho}\,\eta_{\nu\lambda})+f^{ead}\,f^{ebc}\,(\eta_{\mu\nu}\,\eta_{\lambda\rho}-\eta_{\nu\rho}\,\eta_{\mu\lambda})\right]$$

$$\frac{\mathrm{i}\,\delta^{ab}}{k^2+\mathrm{i}\epsilon}$$

$$e k_\mu\, f^{abc}$$

**Figure 1** Feynman rules.

represent gluons and dashed lines represent ghosts. Feynman rules for the fermions are exactly as before, except that now the vertex [40] is multiplied by the color matrix $T^a$. All color indices are contracted along the lines of the Feynman graph. Color factors may be simplified by using the identities

$$\begin{aligned}&\operatorname{Tr} R^a\, R^b = \frac{\dim R}{\dim G}\, C_2(R)\delta^{ab}, \quad R^a\, R^a = C_2(R)\\ &R^a\, R^b\, R^a = \left(C_2(R) - \frac{1}{2}\, C_2(G)\right) R^b\end{aligned} \qquad [43]$$

where $R^a := R(T^a)$ and $C_2(R)$ is the quadratic Casimir invariant of the representation $R$ (with value $C_2(G)$ in the adjoint representation). For $G = \mathrm{SU}(N)$, one has $C_2(G) = N$ and $C_2(N) = (N^2 - 1)/2N$ for the fundamental representation.

The cancellation of infrared divergences in loop amplitudes of QCD is far more delicate than in QED, as there is no analog of the Bloch–Nordsieck theorem in this case. The Kinoshita–Lee–Nauenberg theorem guarantees that, at the end of any perturbative calculation, these divergences must cancel for any appropriately defined physical quantity. However, at a given order of perturbation theory, a physical quantity typically involves both virtual and real emission contributions that are separately infrared divergent. Already at two-loop level these divergences have a highly intricate structure. Their precise form is specified by the Catani color-space factorization formula, which also provides an efficient way of organizing amplitudes into divergent parts, which ultimately drop out of physical quantities, and finite contributions.

The computation of multigluon amplitudes in nonabelian gauge theory is rather complicated when one uses polarization states of vector bosons. A much more efficient representation of amplitudes is provided by adopting a helicity (or circular polarization) basis for external gluons. In the spinor–helicity formalism, one expresses positive and negative-helicity polarization vectors in terms of massless Weyl spinors $|k^\pm\rangle := \frac{1}{2}(\mathbb{1} \pm \gamma_5)u_k = \frac{1}{2}(\mathbb{1} \pm \gamma_5)v_k$ through

$$e^\pm_\mu(k;q) = \pm\frac{\langle q^\mp|\gamma_\mu|k^\mp\rangle}{\sqrt{2}\langle q^\mp|k^\pm\rangle} \qquad [44]$$

where $q$ is an arbitrary null reference momentum which drops out of the final gauge-invariant amplitudes. The spinor products are crossing symmetric, antisymmetric in their arguments, and satisfy the identities

$$\left\langle k_i^-|k_j^+\right\rangle\left\langle k_j^+|k_i^-\right\rangle = 2k_i\cdot k_j$$
$$\left\langle k_i^-\middle|k_j^+\right\rangle\langle k_l^-|k_r^+\rangle = \langle k_i^-|k_r^+\rangle\left\langle k_l^-|k_j^+\right\rangle + \langle k_i^-|k_l^+\rangle\left\langle k_j^-|k_r^+\right\rangle \quad [45]$$

Any amplitude with massless external fermions and vector bosons can be expressed in terms of spinor products. Conversely, the spinor products offer the most compact representation of helicity amplitudes which can be related to more conventional amplitudes described in terms of Lorentz invariants. For loop amplitudes, one uses a dimensional regularization scheme in which all helicity states are kept four dimensional and only internal loop momenta are continued to $D = 4 + \epsilon$ dimensions.

## Computing Loop Integrals

At the very heart of perturbative quantum field theory is the problem of computing Feynman integrals for multiloop scattering amplitudes. The integrations typically involve serious technical challenges and for the most part are intractable by straightforward analytical means. We will now survey some of the computational techniques that have been developed for calculating quantum loop amplitudes which arise in the field theories considered previously.

### Asymptotic Expansion

In many physical instances one is interested in scattering amplitudes in certain kinematical limits. In this case one may perform an asymptotic expansion of multiloop diagrams whose coefficients are typically nonanalytic functions of the perturbative expansion parameter $\hbar$. The main simplification which arises comes from the fact that the expansions are done before any momentum integrals are evaluated. In the limits of interest, Taylor series expansions in different selected regions of each loop momentum can be interpreted in terms of subgraphs and co-subgraphs of the original Feynman diagram.

Consider a Feynman diagram $\mathfrak{D}$ which depends on a collection $\{Q_i\}$ of large momenta (or masses), and a collection $\{m_i, q_i\}$ of small masses and momenta. The prescription for the large-momentum asymptotic expansion of $\mathfrak{D}$ may be summarized in the diagrammatic formula

$$\lim_{Q\to\infty}\mathfrak{D}(Q;m,q) = \sum_{\mathfrak{d}\subset\mathfrak{D}}(\mathfrak{D}/\mathfrak{d})(m,q)\star\left(\mathcal{T}_{\{m_\mathfrak{d},q_\mathfrak{d}\}}\mathfrak{d}\right)(Q;m_\mathfrak{d},q_\mathfrak{d}) \quad [46]$$

where the sum runs through all subgraphs $\mathfrak{d}$ of $\mathfrak{D}$ which contain all vertices where a large momentum enters or leaves the graph and is one-particle irreducible after identifying these vertices. The operator $\mathcal{T}_{\{m_\mathfrak{d},q_\mathfrak{d}\}}$ performs a Taylor series expansion before any integration is carried out, and the notation $(\mathfrak{D}/\mathfrak{d})\star(\mathcal{T}_{\{m_\mathfrak{d},q_\mathfrak{d}\}}\mathfrak{d})$ indicates that the subgraph $\mathfrak{d}\subset\mathfrak{D}$ is replaced by its Taylor expansion in all masses and external momenta of $\mathfrak{d}$ that do not belong to the set $\{Q_i\}$. The external momenta of $\mathfrak{d}$ which become loop momenta in $\mathfrak{D}$ are also considered to be small. The loop integrations are then performed only after all these expansions have been carried out. The diagrams $\mathfrak{D}/\mathfrak{d}$ are called co-subgraphs.

The subgraphs become massless integrals in which the scales are set by the large momenta. For instance, in the simplest case of a single large momentum $Q$ one is left with integrals over propagators. The co-subgraphs may contain small external momenta and masses, but the resulting integrals are typically much simpler than the original one. A similar formula is true for large-mass expansions, with the vertex conditions on subdiagrams replace by propagator conditions. For example, consider the asymptotic expansion of the two-loop double bubble diagram (Figure 2) in the region $q^2 \ll m^2$, where $m$ is the mass of the inner loop. The subgraphs (to the right of the stars) are expanded in all external momenta including $q$ and reinserted into the fat vertices of the co-subgraphs (to the left of the stars). Once such asymptotic expansions are carried out, one may attempt to reconstruct as much information as possible about the given scattering amplitude

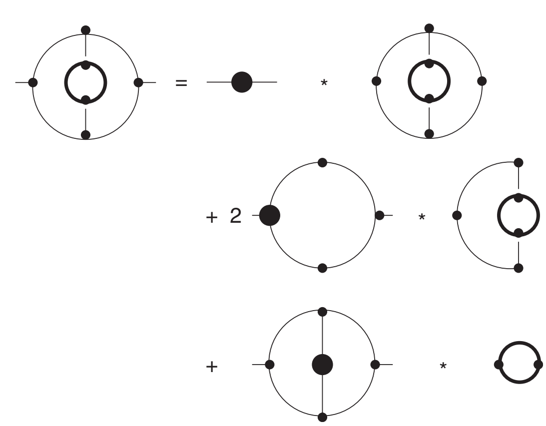



**Figure 2** Asymptotic expansion of the two-loop double bubble diagram.

by using the method of Padé approximation which requires knowledge of only part of the expansion of the diagram. By construction, the Padé approximation has the same analytic properties as the exact amplitude.

### Brown–Feynman Reduction

When considering loop diagrams which involve fermions or gauge bosons, one encounters tensorial Feynman integrals. When these involve more than three distinct denominator factors (propagators), they require more than two Feynman parameters for their evaluation and become increasingly complicated. The Brown–Feynman method simplifies such higher-rank integrals and effectively reduces them to scalar integrals which typically require fewer Feynman parameters for their evaluation.

To illustrate the idea behind this method, consider the one-loop rank-3 tensor Feynman integral

$$J^{\mu\nu\lambda} = \int \frac{\mathrm{d}^D k}{(2\pi)^D} \times \frac{k^\mu k^\nu k^\lambda}{k^2(k^2-\mu^2)(q-k)^2((k-q)^2+\mu^2)(k^2+2k\cdot p)} \quad [47]$$

where $p$ and $q$ are external momenta with the mass-shell conditions $p^2=(p-q)^2=m^2$. By Lorentz invariance, the general structure of the integral [47] will be of the form

$$J^{\mu\nu\lambda} = a^{\mu\nu}p^\lambda + b^{\mu\nu}q^\lambda + c^\mu s^{\nu\lambda} + c^\nu s^{\mu\lambda} \quad [48]$$

where $a^{\mu\nu}, b^{\mu\nu}$ are tensor-valued functions and $c^\mu$ a vector-valued function of $p$ and $q$. The symmetric tensor $s^{\mu\nu}$ is chosen to project out components of vectors transverse to both $p$ and $q$, i.e., $p_\mu s^{\mu\nu} = q_\mu s^{\mu\nu} = 0$, with the normalization $s_\mu{}^\mu = D-2$. Solving these constraints leads to the explicit form

$$s^{\mu\nu} = \eta^{\mu\nu} - \frac{m^2 q^\mu q^\nu + q^2 p^\mu p^\nu - (p\cdot q)(q^\mu p^\nu + p^\mu q^\nu)}{m^2 q^2 - (p\cdot q)^2} \quad [49]$$

To determine the as yet unknown functions $a^{\mu\nu}, b^{\mu\nu}$ and $c^\mu$ above, we first contract both sides of the decomposition [48] with $p^\mu$ and $q^\mu$ to get

$$\begin{aligned} 2p_\lambda J^{\mu\nu\lambda} &= 2m^2 a^{\mu\nu} + 2(p\cdot q)b^{\mu\nu} \\ 2q_\lambda J^{\mu\nu\lambda} &= 2(p\cdot q)a^{\mu\nu} + 2q^2 b^{\mu\nu} \end{aligned} \quad [50]$$

Inside the integrand of [47], we then use the trivial identities

$$\begin{aligned} 2k\cdot p &= \left(k^2 + 2k\cdot p\right) - k^2 \\ 2q\cdot k &= k^2 + q^2 - (k-q)^2 \end{aligned} \quad [51]$$

to write the left-hand sides of [50] as the sum of rank-2 Feynman integrals which, with the exception of the one multiplied by $q^2$ from [51], have one less denominator factor. This formally determines the coefficients $a^{\mu\nu}$ and $b^{\mu\nu}$ in terms of a set of rank-2 integrations. The vector function $c^\mu$ is then found from the contraction

$$J^{\mu\nu}{}_\nu = p_\nu a^{\mu\nu} + q_\nu b^{\mu\nu} + (D-2)c^\mu \quad [52]$$

This contraction eliminates the $k^2$ denominator factor in the integrand of [47] and produces a vector-valued integral. Solving the system of algebraic equations [50] and [52] then formally determines the rank-3 Feynman integral [47] in terms of rank-1 and rank-2 Feynman integrals. The rank-2 Feynman integrals thus generated can then be evaluated in the same way by writing a decomposition for them analogous to [48] and solving for them in terms of vector-valued and scalar-valued Feynman integrals. Finally, the rank-1 integrations can be solved for in terms of a set of scalar-valued integrals, most of which have fewer denominator factors in their integrands.

Generally, any one-loop amplitude can be reduced to a set of basic integrals by using the Passarino–Veltman reduction technique. For example, in supersymmetric amplitudes of gluons any tensor Feynman integral can be reduced to a set of scalar integrals, that is, Feynman integrals in a scalar field theory with a massless particle circulating in the loop, with rational coefficients. In the case of $\mathcal{N}=4$ supersymmetric Yang–Mills theory, only scalar box integrals appear.

### Reduction to Master Integrals

While the Brown–Feynman and Passarino–Veltman reductions are well suited for dealing with one-loop diagrams, they become rather cumbersome for higher-loop computations. There are other more powerful methods for reducing general tensor integrals into a basis of known integrals called master integrals. Let us illustrate this technique on a scalar example. Any scalar massless two-loop Feynman integral can be brought into the form

$$I(p) = \int \frac{\mathrm{d}^D k}{(2\pi)^D} \int \frac{\mathrm{d}^D k'}{(2\pi)^D} \prod_{j=1}^{t} \Delta_j^{-l_j} \prod_{i=1}^{q} \Sigma_i^{n_i} \quad [53]$$

where $\Delta_j$ are massless scalar propagators depending on the loop momenta $k, k'$ and the external momenta $p_1, \ldots, p_n$, and $\Sigma_i$ are scalar products of a loop momentum with an external momentum or of the two loop momenta. The topology of the corresponding Feynman diagram is uniquely determined by specifying the set $\Delta_1, \ldots, \Delta_t$ of $t$ distinct

propagators in the graph, while the integral itself is specified by the powers $l_j \geq 1$ of all propagators, by the selection $\Sigma_1, \ldots, \Sigma_q$ of $q$ scalar products and by their powers $n_i \geq 0$.

The integrals in a class of diagrams of the same topology with the same denominator dimension $r = \sum_j l_j$ and same total scalar product number $s = \sum_i n_i$ are related by various identities. One class follows from the fact that the integral over a total derivative with respect to any loop momentum vanishes in dimensional regularization as

$$\int \frac{\mathrm{d}^D k}{(2\pi)^D} \frac{\partial J(k)}{\partial k^\mu} = 0$$

where $J(k)$ is any tensorial combination of propagators, scalar products and loop momenta. The resulting relations are called integration-by-parts identities and for two-loop integrals can be cast into the form

$$\int \frac{\mathrm{d}^D k}{(2\pi)^D} \int \frac{\mathrm{d}^D k'}{(2\pi)^D} v^\mu \frac{\partial f(k,k',p)}{\partial k^\mu} = 0 = \int \frac{\mathrm{d}^D k}{(2\pi)^D} \int \frac{\mathrm{d}^D k'}{(2\pi)^D} v^\mu \frac{\partial f(k,k',p)}{\partial k'^\mu} \qquad [54]$$

where $f(k,k',p)$ is a scalar function containing propagators and scalar products, and $v^\mu$ is any internal or external momentum. For a graph with $\ell$ loops and $n$ independent external momenta, this results in a total of $\ell(n+\ell)$ relations.

In addition to these identities, one can also exploit the fact that all Feynman integrals [53] are Lorentz scalars. Under an infinitesimal Lorentz transformation $p^\mu \to p^\mu + \delta p^\mu$, with $\delta p^\mu = p^\nu \delta\epsilon^\mu_\nu$, $\delta\epsilon^\mu_\nu = -\delta\epsilon^\nu_\mu$, one has the invariance condition $I(p+\delta p) = I(p)$, which leads to the linear homogeneous differential equations

$$\sum_{i=1}^{n} \left( p_i^\nu \frac{\partial}{\partial p_{i\mu}} - p_i^\mu \frac{\partial}{\partial p_{i\nu}} \right) I(p) = 0 \qquad [55]$$

This equation can be contracted with all possible antisymmetric combinations of $p_{i\mu} p_{j\nu}$ to yield linearly independent Lorentz invariance identities for (53).

Using these two sets of identities, one can either obtain a reduction of integrals of the type (53) to those corresponding to a small number of simpler diagrams of the same topology and diagrams of simpler topology (fewer denominator factors), or a complete reduction to diagrams with simpler topology. The remaining integrals of the topology under consideration are called irreducible master integrals. These momentum integrals cannot be further reduced and have to be computed by different techniques. For instance, one can apply a Mellin–Barnes transformation of all propagators given by

$$\frac{1}{(k^2+a)^l} = \frac{1}{(l-1)!} \int_{-\mathrm{i}\infty}^{\mathrm{i}\infty} \frac{\mathrm{d}z}{2\pi \mathrm{i}} \frac{a^z}{(k^2)^{l+z}} \Gamma(l+z)\Gamma(-z) \qquad [56]$$

where the contour of integration is chosen to lie to the right of the poles of the Euler function $\Gamma(l+z)$ and to the left of the poles of $\Gamma(-z)$ in the complex $z$-plane. Alternatively, one may apply the negative-dimension method in which $D$ is regarded as a negative integer in intermediate calculations and the problem of loop integration is replaced with that of handling infinite series. When combined with the above methods, it may be used to derive powerful recursion relations among scattering amplitudes. Both of these techniques rely on an explicit integration over the loop momenta of the graph, their differences occurring mainly in the representations used for the propagators.

The procedure outlined above can also be used to reduce a tensor Feynman integral to scalar integrals, as in the Brown–Feynman and Passarino–Veltman reductions. The tensor integrals are expressed as linear combinations of scalar integrals of either higher dimension or with propagators raised to higher powers. The projection onto a tensor basis takes the form [53] and can thus be reduced to master integrals.

## String Theory Methods

The realizations of field theories as the low-energy limits of string theory provides a number of powerful tools for the calculation of multiloop amplitudes. They may be used to provide sets of diagrammatic computational rules, and they also work well for calculations in quantum gravity. In this final part we shall briefly sketch the insights into perturbative quantum field theory that are provided by techniques borrowed from string theory.

### String Theory Representation

String theory provides an efficient compact representation of scattering amplitudes. At each loop order there is only a single closed string diagram, which includes within it all Feynman graphs along with the contributions of the infinite tower of massive string excitations. Schematically, at one-loop order, the situation is as shown in Figure 3. The terms arising from the heavy string modes are removed by taking the low-energy limit in which all external momenta lie well below the energy scale set by the string tension. This limit picks out the regions of integration in the string diagram corresponding to particle-like graphs, but with different diagrammatic rules.

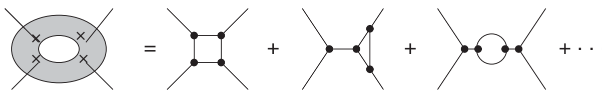


**Figure 3** String theory representation at one-loop order.

Given these rules, one may formulate a purely field-theoretic framework which reproduces them. In the case of QCD, a key ingredient is the use of a special gauge originally derived from the low-energy limit of tree-level string amplitudes. This is known as the Gervais–Neveu gauge and it is defined by the gauge-fixing Lagrangian density

$$\mathcal{L}_{\mathrm{GN}} = -\frac{1}{2}\mathrm{Tr}\left(\partial_\mu A^\mu - \frac{\mathrm{i}e}{\sqrt{2}}A_\mu A^\mu\right)^2 \qquad [57]$$

This gauge choice simplifies the color factors that arise in scattering amplitudes. The string theory origin of gauge theory amplitudes is then most closely mimicked by combining this gauge with the background field gauge, in which one decomposes the gauge field into a classical background field and a fluctuating quantum field as $A_\mu = A_\mu^{\mathrm{cl}} + A_\mu^{\mathrm{qu}}$, and imposes the gauge-fixing condition $D_\mu^{\mathrm{cl}} A^{\mathrm{qu}\mu} = 0$, where $D_\mu^{\mathrm{cl}}$ is the background field covariant derivative evaluated in the adjoint representation of the gauge group. This hybrid gauge is well suited for computing the effective action, with the quantum part describing gluons propagating around loops and the classical part describing gluons emerging from the loops. The leading loop momentum behavior of one-particle irreducible graphs with gluons in the loops is very similar to that of graphs with scalar fields in the loops.

### Supersymmetric Decomposition

String theory also suggests an intimate relationship with supersymmetry. For example, at tree level, QCD is effectively supersymmetric because a multi-gluon tree amplitude contains no fermion loops, and so the fermions may be taken to lie in the adjoint representation of the gauge group. Thus, pure gluon tree amplitudes in QCD are identical to those in supersymmetric Yang–Mills theory. They are connected by supersymmetric Ward identities to amplitudes with fermions (gluinos) which drastically simplify computations. In supersymmetric gauge theory, these identities hold to all orders of perturbation theory.

At one-loop order and beyond, QCD is not supersymmetric. However, one can still perform a supersymmetric decomposition of a QCD amplitude for which the supersymmetric components of the amplitude obey the supersymmetric Ward identities. Consider, for example, a one-loop multigluon scattering amplitude. The contribution from a fermion propagating in the loop can be decomposed into the contribution of a complex scalar field in the loop plus a contribution from an $\mathcal{N}=1$ chiral supermultiplet consisting of a complex scalar field and a Weyl fermion. The contribution from a gluon circulating in the loop can be decomposed into contributions of a complex scalar field, an $\mathcal{N}=1$ chiral supermultiplet, and an $\mathcal{N}=4$ vector supermultiplet comprising three complex scalar fields, four Weyl fermions and one gluon all in the adjoint representation of the gauge group. This decomposition assumes the use of a supersymmetry-preserving regularization.

The supersymmetric components have important cancellations in their leading loop momentum behavior. For instance, the leading large loop momentum power in an $n$-point 1PI graph is reduced from $|k|^n$ down to $|k|^{n-2}$ in the $\mathcal{N}=1$ amplitude. Such a reduction can be extended to any amplitude in supersymmetric gauge theory and is related to the improved ultraviolet behavior of supersymmetric amplitudes. For the $\mathcal{N}=4$ amplitude, further cancellations reduce the leading power behavior all the way down to $|k|^{n-4}$. In dimensional regularization, $\mathcal{N}=4$ supersymmetric loop amplitudes have a very simple analytic structure owing to their origins as the low-energy limits of superstring scattering amplitudes. The supersymmetric Ward identities in this way can be used to provide identities among the nonsupersymmetric contributions. For example, in $\mathcal{N}=1$ supersymmetric Yang–Mills theory one can deduce that fermion and gluon loop contributions are equal and opposite for multi-gluon amplitudes with maximal helicity violation.

### Scattering Amplitudes in Twistor Space

The scattering amplitude in QCD with $n$ incoming gluons of the same helicity vanishes, as does the amplitude with $n-1$ incoming gluons of one helicity and one gluon of the opposite helicity for $n \geq 3$. The first nonvanishing amplitudes are the maximal helicity violating (*MHV*) amplitudes involving $n-2$ gluons of one helicity and two gluons of the opposite helicity. Stripped of the momentum conservation delta-function and the group theory factor, the tree-level amplitude for a pair of gluons of negative helicity is given by

$$\mathcal{A}(k) = e^{n-2}\langle k_r^- | k_s^+ \rangle \prod_{i=1}^{n} \langle k_i^- | k_{i+1}^+ \rangle^{-1} \qquad [58]$$

This amplitude depends only on the holomorphic (negative chirality) Weyl spinors. The full MHV amplitude (with the momentum conservation delta-function) is invariant under the conformal group $\mathrm{SO}(4,2) \cong \mathrm{SU}(2,2)$ of four-dimensional

Minkowski space. After a Fourier transformation of the positive-chirality components, the complexification $SL(4, \mathbb{C})$ has an obvious four-dimensional representation acting on the positive- and negative-chirality spinor products. This representation space is isomorphic to $\mathbb{C}^4$ and is called twistor space. Its elements are called twistors.

Wave functions and amplitudes have a known behavior under the $\mathbb{C}^\times$-action which rescales twistors, giving the projective twistor space $\mathbb{CP}^3$ or $\mathbb{RP}^3$ according to whether the twistors are complex valued or real valued. The Fourier transformation to twistor space yields (due to momentum conservation) the localization of an MHV amplitude to a genus-0 holomorphic curve $\mathbb{CP}^1$ of degree 1 in $\mathbb{CP}^3$ (or to a real line $\mathbb{RP}^1 \subset \mathbb{RP}^3$). It is conjectured that, generally, an $\ell$-loop amplitude with $p$ gluons of positive helicity and $q$ gluons of negative helicity is supported on a holomorphic curve in twistor space of degree $q + \ell - 1$ and genus $\leq \ell$. The natural interpretation of this curve is as the world sheet of a string. The perturbative gauge theory may then be described in terms of amplitudes arising from the couplings of gluons to a string. This twistor string theory is a topological string theory which gives the appropriate framework for understanding the twistor properties of scattering amplitudes. This framework has been used to analyze MHV tree diagrams and one-loop $\mathcal{N} = 4$ supersymmetric amplitudes of gluons.

*See also:* Constructive Quantum Field Theory; Dispersion Relations; Effective Field Theories; Gauge Theories from Strings; Hopf Algebra Structure of Renormalizable Quantum Field Theory; Perturbative Renormalization Theory and BRST; Quantum Chromodynamics; Renormalization: General Theory; Scattering, Asymptotic Completeness and Bound States; Scattering in Relativistic Quantum Field Theory: Fundamental Concepts and Tools; Stationary Phase Approximation; Supersymmetric Particle Models.

## Further Reading

Bern Z, Dixon L, and Kosower DA (1996) Progress in one-loop QCD computations. *Annual Review of Nuclear and Particle Science* 46: 109–148.

Brown LM and Feynman RP (1952) Radiative corrections to Compton scattering. *Physical Review* 85: 231–244.

Cachazo F and Svrček P (2005) Lectures on twistor strings and perturbative Yang–Mills theory. Preprint arXiv:hep-th/0504194.

Catani S (1998) The singular behaviour of QCD amplitudes at two-loop order. *Physics Letters B* 427: 161–171.

Chetyrkin KG and Tkachov FV (1981) Integration by parts: the algorithm to calculate beta functions in 4 loops. *Nuclear Physics B* 192: 159–204.

Dixon L (1996) Calculating scattering amplitudes efficiently. In: Soper DE *QCD and Beyond*, pp. 539–584. Singapore: World Scientific.

Fleischer J, Jegerlehner F, and Tarasov OV (2000) Algebraic reduction of one-loop Feynman graph amplitudes. *Nuclear Physics B* 566: 423–440.

Gehrmann T and Remiddi E (2000) Differential equations for two-loop four-point functions. *Nuclear Physics B* 580: 485–518.

Halliday IG and Ricotta RM (1987) Negative dimensional integrals 1: Feynman graphs. *Physics Letters B* 193: 241–246.

Itzykson C and Zuber J-B (1980) *Quantum Field Theory*. New York: McGraw-Hill.

Mangano ML and Parke SJ (1991) Multiparton amplitudes in gauge theories. *Physics Reports* 200: 301–367.

Magnen J and Seneor R (1979) Expansion and summability methods in constructive field theory. *Mathematical Problems in Theoretical Physics, Proc. Lausanne* 1979: 217–223.

Passarino G and Veltman MJG (1979) One-loop corrections for $e^+e^-$ annihilation into $\mu^+\mu^-$ in the Weinberg model. *Nuclear Physics B* 160: 151–207.

Peskin ME and Schroeder DV (1995) *An Introduction to Quantum Field Theory*. Reading, MA: Addison-Wesley.

Pokorski S (1987) *Gauge Field Theories*. Cambridge: Cambridge University Press.

Steinhauser M (2002) Results and techniques of multiloop calculations. *Physics Reports* 364: 247–357.

Steinmann O (1995) Axiomatic approach to perturbative quantum field theory. *Annales de Poincare en Physique Theorique*. 63: 399–409.

Weinberg S (1995) *The Quantum Theory of Fields, Volume 1: Foundations*. Cambridge: Cambridge University Press.

# Perturbative Renormalization Theory and BRST

**K Fredenhagen**, Universität Hamburg, Hamburg, Germany
**M Dütsch**, Universität Zürich, Zürich, Switzerland



## Main Problems in the Perturbative Quantization of Gauge Theories

Gauge theories are field theories in which the basic fields are not directly observable. Field configurations yielding the same observables are connected by a gauge transformation. In the classical theory, the Cauchy problem is well posed for the observables, but in general not for the nonobservable gauge-variant basic fields, due to the existence of time-dependent gauge transformations.

Attempts to quantize the gauge-invariant objects directly have not yet been completely satisfactory. Instead, one modifies the classical action by adding a gauge-fixing term such that standard techniques of perturbative quantization can be applied and such that the dynamics of the gauge-invariant classical fields is not changed. In perturbation theory, this

problem shows up already in the quantization of the free gauge fields (see the section "Quantization of free gauge fields"). In the final (interacting) theory the physical quantities should be independent on how the gauge fixing is done ("gauge independence").

Traditionally, the quantization of gauge theories is mostly analyzed in terms of path integrals (e.g., by Faddeev and Popov), where some parts of the arguments are only heuristic. In the original treatment of Becchi, Rouet, and Stora (cf. also Tyutin) (which is called "BRST-quantization"), a restriction to purely massive theories was necessary; the generalization to the massless case by Lowenstein's method is cumbersome.

The BRST quantization is based on earlier work of Feynman, Faddeev, and Popov (introduction of "ghost fields"), and of Slavnov. The basic idea is that after adding a term to the Lagrangian which makes the Cauchy problem well posed but which is not gauge-invariant one enlarges the number of fields by infinitesimal gauge transformations ("ghosts") and their duals ("anti-ghosts"). One then adds a further term to the Lagrangian which contains a coupling of the anti-ghosts and ghosts. The BRST transformation acts as an infinitesimal gauge transformation on the original fields and on the gauge transformations themselves and maps the anti-ghosts to the gauge-fixing terms. This is done in such a way that the total Lagrangian is invariant and that the BRST transformation is nilpotent. The hard problem in the perturbative construction of gauge theories is to show that BRST symmetry can be maintained during renormalization (see the section on perturbative renormalization). By means of the "quantum action principle" of Lowenstein (1971) and Lam (1972, 1973) a cohomological classification of anomalies was worked out (an overview is given, e.g., in the book of Piguet and Sorella (1995)). For more details, *see* BRST Quantization.

The BRST quantization can be carried out in a transparent way in the framework of algebraic quantum field theory (AQFT, *see* Algebraic Approach to Quantum Field Theory). The advantage of this formulation is that it allows one to separate the three main problems of perturbative gauge theories:

1. the elimination of unphysical degrees of freedom,
2. positivity (or "unitarity"), and
3. the problem of infrared divergences.

In AQFT, the procedure is the following: starting from an algebra of all local fields, including the unphysical ones, one shows that after perturbative quantization the algebra admits the BRST transformation as a graded nilpotent derivation. The algebra of observables is then defined as the cohomology of the BRST transformation. To solve the problem of positivity, one has to show that the algebra of observables, in contrast to the algebra of all fields, has a nontrivial representation on a Hilbert space. Finally, one can attack the infrared problem by investigating the asymptotic behavior of states. The latter problem is nontrivial even in quantum electrodynamics (since an electron is accompanied by a "cloud of soft photons") and may be related to confinement in quantum chromodynamics.

The method of BRST quantization is by no means restricted to gauge theories, but applies to general constrained systems. In particular, massive vector fields, where the masses are usually generated by the Higgs mechanism, can alternatively be treated directly by the BRST formalism, in close analogy to the massless case (cf. the section on quantization of free gauge fields).

## Local Operator BRST Formalism

In AQFT, the principal object is the family of operator algebras $\mathcal{O} \to \mathcal{A}(\mathcal{O})$ (where $\mathcal{O}$ runs, e.g., through all double cones in Minkowski space), which fulfills the Haag–Kastler axioms (cf. Algebraic Approach to Quantum Field Theory). To construct these algebras, one considers the algebras $\mathcal{F}(\mathcal{O})$ generated by all local fields including ghosts $u$ and anti-ghosts $\tilde{u}$. Ghosts and anti-ghosts are scalar fermionic fields. The algebra gets a $\mathbb{Z}_2$ grading with respect to even and odd ghost numbers, where ghosts get ghost numbers +1 and anti-ghosts ghost number –1. The BRST transformation $s$ acts on these algebras as a $\mathbb{Z}_2$-graded derivation with $s^2 = 0$, $s(\mathcal{F}(\mathcal{O})) \subset \mathcal{F}(\mathcal{O})$, and $s(F^*) = -(-1)^{\delta_F} s(F)^*$, $\delta_F$ denoting the ghost number of $F$.

The observables should be $s$-invariant and may be identified if they differ by a field in the range of $s$. Since the range $\mathcal{A}_{00}$ of $s$ is an ideal in the kernel $\mathcal{A}_0$ of $s$, the algebra of observables is defined as the quotient

$$\mathcal{A} := \mathcal{A}_0 / \mathcal{A}_{00} \qquad [1]$$

and the local algebras $\mathcal{A}(\mathcal{O}) \subset \mathcal{A}$ are the images of $\mathcal{A}_0 \cap \mathcal{F}(\mathcal{O})$ under the quotient map $\mathcal{A}_0 \to \mathcal{A}$.

To prove that $\mathcal{A}$ admits a nontrivial representation by operators on a Hilbert space, one may use the BRST operator formalism (Kugo and Ojima 1979, Dütsch and Fredenhagen 1999): one starts from a representation of $\mathcal{F}$ on an inner-product space $(\mathcal{K}, \langle \cdot\,, \cdot \rangle)$ such that $\langle F^*\phi, \psi \rangle = \langle \phi, F\psi \rangle$

and that $s$ is implemented by an operator $Q$ on $\mathcal{K}$, that is,

$$s(F) = [Q, F] \qquad [2]$$

with $[\cdot,\cdot]$ denoting the graded commutator, such that $Q$ is symmetric and nilpotent. One may then construct the space of physical states as the cohomology of $Q$, $\mathcal{H} := \mathcal{K}_0/\mathcal{K}_{00}$, where $\mathcal{K}_0$ is the kernel and $\mathcal{K}_{00}$ the range of $Q$. The algebra of observables now has a natural representation $\pi$ on $\mathcal{H}$:

$$\pi([A])[\phi] := [A\phi] \qquad [3]$$

(where $A \in \mathcal{A}_0, \phi \in \mathcal{K}_0, [A] := A + \mathcal{A}_{00}, [\phi] := \phi + \mathcal{K}_{00}$). The crucial question is whether the scalar product on $\mathcal{H}$ inherited from $\mathcal{K}$ is positive definite.

In free quantum field theories $(\mathcal{K}, \langle\cdot,\cdot\rangle)$ can be chosen in such a way that the positivity can directly be checked by identifying the physical degrees of freedom (see next section). In interacting theories (see the section on perturbative construction of gauge theories), one may argue in terms of scattering states that the free BRST operator on the asymptotic fields coincides with the BRST operator of the interacting theory. This argument, however, is invalidated by infrared problems in massless gauge theories. Instead, one may use a stability property of the construction.

Namely, let $\tilde{\mathcal{F}}$ be the algebra of formal power series with values in $\mathcal{F}$, and let $\tilde{\mathcal{K}}$ be the vector space of formal power series with values in $\mathcal{K}$. $\tilde{\mathcal{K}}$ possesses a natural inner product with values in the ring of formal power series $\mathbb{C}[[\lambda]]$, as well as a representation of $\tilde{\mathcal{F}}$ by operators. One also assumes that the BRST transformation $\tilde{s}$ is a formal power series $\tilde{s} = \sum_n \lambda^n s_n$ of operators $s_n$ on $\mathcal{F}$ and that the BRST operator $\tilde{Q}$ is a formal power series $\tilde{Q} = \sum_n \lambda^n Q_n$ of operators on $\mathcal{K}$. The algebraic construction can then be done in the same way as before, yielding a representation $\tilde{\pi}$ of the algebra of observables $\tilde{\mathcal{A}}$ by endomorphisms of a $\mathbb{C}[[\lambda]]$ module $\tilde{\mathcal{H}}$, which has an inner product with values in $\mathbb{C}[[\lambda]]$.

One now assumes that at $\lambda = 0$ the inner product is positive, in the sense that

(Positivity)

$$\begin{aligned} &\text{(i) } \langle\phi,\phi\rangle \geq 0 \quad \forall \phi \in \mathcal{K} \text{ with } Q_0\phi = 0, \text{ and} \\ &\text{(ii) } Q_0\phi = 0 \wedge \langle\phi,\phi\rangle = 0 \Longrightarrow \phi \in Q_0\mathcal{K} \end{aligned} \qquad [4]$$

Then the inner product on $\tilde{\mathcal{H}}$ is positive in the sense that for all $\tilde{\phi} \in \tilde{\mathcal{H}}$ the inner product with itself, $\langle\tilde{\phi},\tilde{\phi}\rangle$, is of the form $\tilde{c}^*\tilde{c}$ with some power series $\tilde{c} \in \mathbb{C}[[\lambda]]$, and $\tilde{c} = 0$ iff $\tilde{\phi} = 0$.

This result guarantees that, within perturbation theory, the interacting theory satisfies positivity, provided the unperturbed theory was positive and BRST symmetry is preserved.

## Quantization of Free Gauge Fields

The action of a classical free gauge field $A$,

$$\begin{aligned} S_0(A) &= -\frac{1}{4}\int \mathrm{d}x\, F^{\mu\nu}(x)F_{\mu\nu}(x) \\ &= \frac{1}{2}\int \mathrm{d}k \hat{A}_\mu(k)^* M^{\mu\nu}(k)\hat{A}_\nu(k) \end{aligned} \qquad [5]$$

(where $F^{\mu\nu} := \partial^\mu A^\nu - \partial^\nu A^\mu$ and $M^{\mu\nu}(k) := k^2 g^{\mu\nu} - k^\mu k^\nu$) is unsuited for quantization because $M^{\mu\nu}$ is not invertible: due to $M^{\mu\nu}k_\mu = 0$, it has an eigenvalue 0. Therefore, the action is usually modified by adding a Lorentz-invariant gauge-fixing term: $M^{\mu\nu}$ is replaced by $M^{\mu\nu}(k) + \lambda k^\mu k^\nu$, where $\lambda \in \mathbb{R}\setminus\{0\}$ is an arbitrary constant. The corresponding Euler–Lagrange equation reads

$$\Box A^\mu - (1-\lambda)\partial^\mu\partial_\nu A^\nu = 0 \qquad [6]$$

For simplicity, let us choose $\lambda = 1$, which is referred to as Feynman gauge. Then the algebra of the free gauge field is the unital $\star$-algebra generated by elements $A^\mu(f), f \in \mathcal{D}(\mathbb{R}^4)$, which fulfill the relations:

$$f \mapsto A^\mu(f) \quad \text{is linear} \qquad [7]$$

$$A^\mu(\Box f) = 0 \qquad [8]$$

$$A^\mu(f)^* = A^\mu(\bar{f}) \qquad [9]$$

$$[A^\mu(f), A^\nu(g)] = \mathrm{i}g^{\mu\nu}\int \mathrm{d}x\,\mathrm{d}y\, f(x)D(x-y)g(y) \qquad [10]$$

where $D$ is the massless Pauli–Jordan distribution.

This algebra does not possess Hilbert space representations which satisfy the microlocal spectrum condition, a condition which in particular requires the singularity of the two-point function to be of the so-called Hadamard form. It possesses, instead, representations on vector spaces with a nondegenerate sequilinear form, for example, the Fock space over a one-particle space with scalar product

$$\langle\phi,\psi\rangle = (2\pi)^{-3}\int \frac{\mathrm{d}^3\boldsymbol{p}}{2|\boldsymbol{p}|}\overline{\phi^\mu(p)}\psi_\mu(p)|_{p^0=|\boldsymbol{p}|} \qquad [11]$$

Gupta and Bleuler characterized a subspace of the Fock space on which the scalar product is semidefinite; the space of physical states is then obtained

by dividing out the space of vectors with vanishing norm.

After adding a mass term

$$\frac{m^2}{2}\int \mathrm{d}x A_\mu(x)A^\mu(x)$$

to the action [5], it seems to be no longer necessary to add also a gauge-fixing term. The fields then satisfy the Proca equation

$$\partial_\mu F^{\mu\nu} + m^2 A^\nu = 0 \qquad [12]$$

which is equivalent to the equation $(\Box + m^2)A^\mu = 0$ together with the constraint $\partial_\mu A^\mu = 0$. The Cauchy problem is well posed, and the fields can be represented in a positive-norm Fock space with only physical states (corresponding to the three physical polarizations of $A$). The problem, however, is that the corresponding propagator admits no power-counting renormalizable perturbation series.

The latter problem can be circumvented in the following way: for the algebra of the free quantum field, one takes only the equation $(\Box + m^2)A^\mu = 0$ into account (or, equivalently, one adds the gauge-fixing term $(1/2)(\partial_\mu A^\mu)^2$ to the Lagrangian) and goes over from the physical field $A^\mu$ to

$$B^\mu := A^\mu + \frac{\partial^\mu \phi}{m} \qquad [13]$$

where $\phi$ is a real scalar field, to the same mass $m$ where the sign of the commutator is reversed ("bosonic ghost field" or "Stückelberg field"). The propagator of $B^\mu$ yields a power-counting renormalizable perturbation series; however, $B^\mu$ is an unphysical field. One obtains four independent components of $B$ which satisfy the Klein–Gordon equation. The constraint $0 = \partial_\mu A^\mu = \partial_\mu B^\mu + m\phi$ is required for the expectation values in physical states only. So, quantization in the case $m > 0$ can be treated in analogy with [8]–[10] by replacing $A^\mu$ by $B^\mu$, the wave operator by the Klein–Gordon operator $(\Box + m^2)$ in [8], and $D$ by the corresponding massive commutator distribution $\Delta_m$ in [10]. Again, the algebra can be nontrivially represented on a space with indefinite metric, but not on a Hilbert space.

One can now use the method of BRST quantization in the massless as well as in the massive case. One introduces a pair of fermionic scalar fields (ghost fields) $(u, \tilde{u})$. $u, \tilde{u}$, and (for $m > 0$) $\phi$ fulfill the Klein–Gordon equation to the same mass $m \geq 0$ as the vector field $B$. The free BRST transformation reads

$$\begin{aligned} s_0(B^\mu) &= \mathrm{i}\partial^\mu u, \qquad s_0(\phi) = \mathrm{i}mu \\ s_0(u) &= 0, \qquad s_0(\tilde{u}) = -\mathrm{i}(\partial_\nu B^\nu + m\phi) \end{aligned} \qquad [14]$$

(see, e.g., Scharf (2001)). It is implemented by the free BRST charge

$$Q_0 = \int_{x^0=\text{const.}} \mathrm{d}^3x j_0^{(0)}(x^0, \boldsymbol{x}) \qquad [15]$$

where

$$j_\mu^{(0)} := (\partial_\nu B^\nu + m\phi)\partial_\mu u - \partial_\mu(\partial_\nu B^\nu + m\phi)u \qquad [16]$$

is the free BRST current, which is conserved. (The interpretation of the integral in [15] requires some care.) $Q_0$ satisfies the assumptions of the (local) operator BRST formalism, in particular it is nilpotent and positive [4]. Distinguished representatives of the equivalence classes $[\phi] \in \mathrm{Ke}\, Q_0/\mathrm{Ra}\, Q_0$ are the states built up only from the three spatial (two transversal for $m = 0$, respectively) polarizations of $A$.

## Perturbative Renormalization

The starting point for a perturbative construction of an interacting quantum field theory is Dyson's formula for the evolution operator in the interaction picture. To avoid conflicts with Haag's theorem on the nonexistence of the interaction picture in quantum field theory, one multiplies the interaction Lagrangian $\mathcal{L}$ with a test function $g$ and studies the local $S$-matrix,

$$S(g\mathcal{L}) = 1 + \sum_{n=1}^{\infty} \frac{\mathrm{i}^n}{n!}\int \mathrm{d}x_1 \cdots \mathrm{d}x_n g(x_1)\cdots g(x_n) \times T(\mathcal{L}(x_1)\cdots\mathcal{L}(x_n)) \qquad [17]$$

where $T$ denotes a time-ordering prescription. In the limit $g \to 1$ (adiabatic limit), $S(g\mathcal{L})$ tends to the scattering matrix. This limit, however, is plagued by infrared divergences and does not always exist. Interacting fields $F_{g\mathcal{L}}$ are obtained by the Bogoliubov formula:

$$F_{g\mathcal{L}}(x) = \frac{\delta}{\delta h(x)}\Big|_{h=0} S(g\mathcal{L})^{-1} S(g\mathcal{L} + hF) \qquad [18]$$

The algebraic properties of the interacting fields within a region $\mathcal{O}$ depend only on the interaction within a slightly larger region (Brunetti and Fredenhagen 2000), hence the net of algebras in the sense of AQFT can be constructed in the adiabatic limit without the infrared problems (this is called the "algebraic adiabatic limit").

The construction of the interacting theory is thus reduced to a definition of time-ordered products of fields. This is the program of causal perturbation theory (CPT), which was developed by Epstein and Glaser (1973) on the basis of previous work by Stückelberg and Petermann (1953) and Bogoliubov

and Shirkov (1959). For simplicity, we describe CPT only for a real scalar field. Let $\varphi$ be a classical real scalar field which is not restricted by any field equation. Let $\mathcal{P}$ denote the algebra of polynomials in $\varphi$ and all its partial derivatives $\partial^a\varphi$ with multi-indices $a \in \mathbb{N}_0^4$. The time-ordered products $(T_n)_{n\in\mathbb{N}}$ are linear and symmetric maps $T_n : (\mathcal{P} \otimes \mathcal{D}(\mathbb{R}^4))^{\otimes n} \to L(\mathcal{D})$, where $L(\mathcal{D})$ is the space of operators on a dense invariant domain $\mathcal{D}$ in the Fock space of the scalar free field. One often uses the informal notation

$$T_n(g_1F_1 \otimes \cdots \otimes g_nF_n) = \int \mathrm{d}x_1 \cdots \mathrm{d}x_n T_n(F_1(x_1), \ldots, F_n(x_n)) \times g_1(x_1) \cdots g_n(x_n) \qquad [19]$$

where $F_j \in \mathcal{P}$, $g_j \in \mathcal{D}(\mathbb{R}^4)$.

The sequence $(T_n)$ is constructed by induction on $n$, starting with the initial condition

$$T_1\left(\prod_j \partial^{a_j}\varphi(x)\right) =: \prod_j \partial^{a_j}\phi(x) : \qquad [20]$$

where the right-hand side is a Wick polynomial of the free field $\phi$. In the inductive step the requirement of causality plays the main role, that is, the condition that

$$T_n(f_1 \otimes \cdots \otimes f_n) = T_k(f_1 \otimes \cdots \otimes f_k) \times T_{n-k}(f_{k+1} \otimes \cdots \otimes f_n) \qquad [21]$$

if

$$(\operatorname{supp} f_1 \cup \cdots \cup \operatorname{supp} f_k) \cap ((\operatorname{supp} f_{k+1} \cup \cdots \cup \operatorname{supp} f_n) + \bar{V}_-) = \emptyset$$

(where $\bar{V}_-$ is the closed backward light cone). This condition expresses the composition law for evolution operators in a relativistically invariant and local way. Causality determines $T_n$ as an operator-valued distribution on $\mathbb{R}^{4n}$ in terms of the inductively known $T_l$, $l < n$, outside of the total diagonal $\Delta_n := \{(x_1, \ldots, x_n) \mid x_1 = \cdots = x_n\}$, that is, on test functions from $\mathcal{D}(\mathbb{R}^{4n} \setminus \Delta_n)$.

Perturbative renormalization is now the extension of $T_n$ to the full test function space $\mathcal{D}(\mathbb{R}^{4n})$. Generally, this extension is nonunique. In contrast to other methods of renormalization, no divergences appear, but the ambiguities correspond to the finite renormalizations that persist after removal of divergences by infinite counter terms. The ambiguities can be reduced by (re-)normalization conditions, which means that one requires that certain properties which hold by induction on $\mathcal{D}(\mathbb{R}^{4n} \setminus \Delta_n)$ are maintained in the extension, namely:

- (N0) a bound on the degree of singularity near the total diagonal;
- (N1) Poincaré covariance;
- (N2) unitarity of the local $S$-matrix;
- (N3) a relation to the time-ordered products of subpolynomials;
- (N4) the field equation for the interacting field $\varphi_{g\mathcal{L}}$ [18];
- (AWI) the "action Ward identity" (Stora 2002, Dütsch and Fredenhagen 2003): $\partial^\mu T(\cdots F_l(x) \cdots) = T(\cdots \partial^\mu F_l(x) \cdots)$. This condition can be understood as the requirement that physics depends on the action only, so total derivatives in the interaction Lagrangian can be removed; and
- further symmetries, in particular in gauge theories, Ward identities expressing BRST invariance. A universal formulation of all symmetries which can be derived from the field equation in classical field theory is the "master Ward identity" (which presupposes (N3) and (N4)) (Boas and Dütsch 2002, Dütsch and Fredenhagen 2003); see next section.

The problem of perturbative renormalization is to construct a solution of all these normalization conditions. Epstein and Glaser have constructed the solutions of (N0)–(N3). Recently, the conditions (N4) and (AWI) have been included. The master Ward identity cannot always be fulfilled, the obstructions being the famous "anomalies" of perturbative quantum field theory.

## Perturbative Construction of Gauge Theories

In the case of a purely massive theory, the adiabatic limit $S = \lim_{g\to 1} S(g\mathcal{L})$ exists (Epstein and Glaser 1976), and one may adopt a formalism due to Kugo and Ojima (1979), who use the fact that in these theories the BRST charge $Q$ can be identified with the incoming (free) BRST charge $Q_0$ [15]. For the scattering matrix $S$ to be a well-defined operator on the physical Hilbert space of the free theory, $\mathcal{H} = \mathrm{Ke}\, Q_0 / \mathrm{Ra}\, Q_0$, one then has to require

$$\lim_{g\to 1} [Q_0, T((g\mathcal{L})^{\otimes n})]|_{\ker Q_0} = 0 \qquad [22]$$

This is the motivation for introducing the condition of "perturbative gauge invariance" (Dütsch *et al.* 1993, 1994); see Scharf (2001)): according to this condition, there should exist a Lorentz

vector $\mathcal{L}_1^\nu \in \mathcal{P}$ associated with the interaction $\mathcal{L}$, such that

$$[Q_0, T_n(\mathcal{L}(x_1)\cdots\mathcal{L}(x_n)] = \mathrm{i}\sum_{l=1}^{n} \partial_\nu^{x_l} T_n(\mathcal{L}(x_1)\cdots\mathcal{L}_1^\nu(x_l)\cdots\mathcal{L}(x_n)) \qquad [23]$$

This is a somewhat stronger condition than [22] but has the advantage that it can be formulated independently of the adiabatic limit. The condition [22] (or perturbative gauge invariance) can be satisfied for tree diagrams (i.e., the corresponding requirement in classical field theory can be fulfilled). In the massive case, this is impossible without a modification of the model; the inclusion of additional physical scalar fields (corresponding to Higgs fields) yields a solution. It is gratifying that, by making a polynomial ansatz for the interaction $\mathcal{L} \in \mathcal{P}$, perturbative gauge invariance [23] for tree diagrams, renormalizability (i.e., the mass dimension of $\mathcal{L}$ is $\leq 4$), and some obvious requirements (e.g., the Lorentz invariance) determine $\mathcal{L}$ to a far extent. In particular, the Lie-algebraic structure needs not to be put in, as it can be derived in this way (Stora 1997, unpublished). Including loop diagrams (i.e., quantum effects), it has been proved that (N0)–(N2) and perturbative gauge invariance can be fulfilled to all orders for massless SU($N$) Yang–Mills theories.

Unfortunately, in the massless case, it is unlikely that the adiabatic limit exists and, hence, an $S$-matrix formalism is problematic. One should better rely on the construction of local observables in terms of couplings with compact support. However, then the selection of the observables [1] has to be done in terms of the BRST transformation $\tilde{s}$ of the interacting fields.

For the corresponding BRST charge, one makes the ansatz

$$\tilde{Q} = \int \mathrm{d}^4x\, \tilde{j}^\mu_{g\mathcal{L}}(x) b_\mu(x), \quad \mathcal{L} = \sum_{n\geq 1} \mathcal{L}_n \lambda^n \qquad [24]$$

where $(b_\mu)$ is a smooth version of the $\delta$-function characterizing a Cauchy surface and $\tilde{j}^\mu_{g\mathcal{L}}$ is the interacting BRST-current [18] (where $\tilde{j}_\mu = \sum_n j_\mu^{(n)}\, \lambda^n$ ($j_\mu^{(n)} \in \mathcal{P}$) is a formal power series with $j_\mu^{(0)}$ given by [16]). (Note that there is a volume divergence in this integral, which can be avoided by a spatial compactification. This does not change the abstract algebra $\mathcal{F}_\mathcal{L}(\mathcal{O})$.) A crucial requirement is that $\tilde{j}^\mu_{g\mathcal{L}}$ is conserved in a suitable sense. This condition is essentially equivalent to perturbative gauge invariance and hence its application to classical field theory determines the interaction $\mathcal{L}$ in the same way, and in addition the deformation $j^{(0)} \to \tilde{j}_{g\mathcal{L}}$. The latter also gives the interacting BRST charge and transformation, $\tilde{Q}$ and $\tilde{s}$, by [24] and [2]. The so-obtained $\tilde{Q}$ is often nilpotent in classical field theory (and hence this holds also for $\tilde{s}$). However, in QFT conservation of $\tilde{j}_{g\mathcal{L}}$ and $\tilde{Q}^2 = 0$ requires the validity of additional Ward identities, beyond the condition of perturbative gauge invariance [23]. All the necessary identities can be derived from the master Ward identity

$$T_{n+1}(A, F_1, \ldots, F_n) = -\sum_{k=1}^{n} T_n(F_1, \ldots, \delta_A F_k, \ldots, F_n) \qquad [25]$$

where $A = \delta_A S_0$ with a derivation $\delta_A$. The master Ward identity is closely related to the quantum action principle which was formulated in the formalism of generating functionals of Green's functions. In the latter framework, the anomalies have been classified by cohomological methods. The vanishing of anomalies of the BRST symmetry is a selection criterion for physically acceptable models.

In the particular case of QED, the Ward identity

$$\partial_\mu^y T(j^\mu(y) F_1(x_1)\cdots F_n(x_n)) = \mathrm{i}\sum_{j=1}^{n} \delta(y - x_j) \times T\big(F_1(x_1)\cdots(\theta F_j)(x_j)\cdots F_n(x_n)\big) \qquad [26]$$

for the Dirac current $j^\mu := \bar{\psi}\gamma^\mu\psi$, is sufficient for the construction, where $(\theta F) := \mathrm{i}(r - s)F$ for $F = \psi^r \bar{\psi}^s B_1 \cdots B_l$ ($B_1, \ldots, B_l$ are nonspinorial fields) and $F_1, \ldots, F_n$ run through all subpolynomials of $\mathcal{L} = j^\mu A_\mu$, (N0)–(N4) and [26] can be fulfilled to all orders (Dütsch and Fredenhagen, 1999).

*See also:* Algebraic Approach to Quantum Field Theory; Axiomatic Quantum Field Theory; Batalin–Vilkovisky Quantization; BRST Quantization; Constrained Systems; Indefinite Metric; Perturbation Theory and its Techniques; Quantum Chromodynamics; Quantum Field Theory: A Brief Introduction; Quantum Fields with Indefinite Metric: Non-Trivial Models; Renormalization: General Theory; Renormalization: Statistical Mechanics and Condensed Matter; Standard Model of Particle Physics.

## Further Reading

Becchi C, Rouet A, and Stora R (1975) Renormalization of the abelian Higgs–Kibble model. *Communications in Mathematical Physics* 42: 127.

Becchi C, Rouet A, and Stora R (1976) Renormalization of gauge theories. *Annals of Physics (NY)* 98: 287.

Bogoliubov NN and Shirkov DV (1959) *Introduction to the Theory of Quantized Fields*. New York: Interscience Publishers Inc.

Brunetti R and Fredenhagen K (2000) Microlocal analysis and interacting quantum field theories: renormalization on physical backgrounds. *Communications in Mathematical Physics* 208: 623.

Dütsch M, Hurth T, Krahe K, and Scharf G (1993) Causal construction of Yang–Mills theories. I. *N. Cimento A* 106: 1029.

Dütsch M, Hurth T, Krahe K, and Scharf G (1994) Causal construction of Yang–Mills theories. II. *N Cimento A* 107: 375.

Dütsch M and Fredenhagen K (1999) A local (perturbative) construction of observables in gauge theories: the example of QED. *Communications in Mathematical Physics* 203: 71.

Dütsch M and Boas F-M (2002) The Master Ward identity. *Reviews of Mathematical Physics* 14: 977–1049.

Dütsch M and Fredenhagen K (2003) The master Ward identity and generalized Schwinger–Dyson equation in classical field theory. *Communications in Mathematical Physics* 243: 275.

Dütsch M and Fredenhagen K (2004) Causal perturbation theory in terms of retarded products, and a proof of the action Ward identity. *Reviews in Mathematical Physics* 16: 1291–1348.

Epstein H and Glaser V (1973) *Annals Institut Henri Poincaré A* 19: 211.

Epstein H and Glaser V (1976) Adiabatic limit in perturbation theory. In: Velo G and Wightman AS (eds.) *Renormalization Theory*, pp. 193–254.

Henneaux M and Teitelboim C (1992) *Quantization of Gauge Systems*. Princeton: Princeton University Press.

Kugo T and Ojima I (1979) Local covariant operator formalism of nonabelian gauge theories and quark confinement problem. *Supplement of the Progress of Theoritical Physics* 66: 1.

Lam Y-MP (1972) Perturbation Lagrangian theory for scalar fields – Ward–Takahashi identity and current algebra. *Physics Reviews D* 6: 2145.

Lam Y-MP (1973) Equivalence theorem on Bogoliubov–Parasiuk–Hepp–Zimmermann – renormalized Lagrangian field theories. *Physics Reviews D* 7: 2943.

Lowenstein JH (1971) Differential vertex operations in Lagrangian field theory. *Communications in Mathematical Physics* 24: 1.

Piguet O and Sorella S (1995) Algebraic Renormalization: Perturbative Renormalization, Symmetries and Anomalies, Lecture Notes in Physics. Berlin: Springer.

Scharf G (1995) *Finite Quantum Electrodynamics. The Causal Approach*, 2nd edn. Berlin: Springer.

Scharf G (2001) *Quantum Gauge Theories – A True Ghost Story*. New York: Wiley.

Stora R (2002) Pedagogical experiments in renormalized perturbation theory. Contribution to the conference. *Theory of Renormalization and Regularization*. Germany: Hesselberg.

Stückelberg ECG and Petermann A (1953) La normalisation des constantes dans la theorie des quanta. *Helvetica Physica Acta* 26: 499–520.

Weinberg S (1996) *The Quantum Theory of Fields*. Cambridge: Cambridge University Press.

# Phase Transition Dynamics

**A Onuki**, Kyoto University, Kyoto, Japan



## Introduction

When an external parameter such as the temperature $T$ is changed, physical systems in a homogeneous state often become unstable and tend to an ordered phase with broken symmetry. The growth of new order takes place with coarsening of domains or defect structures on mesoscopic spatial scales much longer than the microscopic molecular scale. Such ordering processes are ubiquitously observed in many systems such as ferromagnetic (spin) systems, solid alloys, and fluids. Historically, structural ordering and phase separation in solid alloys have been one of the central problems in metallurgy (Cahn 1961). These are highly nonlinear and far-from-equilibrium processes and have been studied as challenging subjects in condensed matter physics, polymer science, and metallurgy (Gunton *et al.* 1983, Binder 1991, Bray 1994, Onuki 2002). Here a short review on phase ordering is given on the basis of prototype mathematical models, which can be a starting point to understand the real complex problems.

## Phase Ordering in Nonconserved Systems

Let us consider phase ordering in a system with a scalar spacetime-dependent variable $\psi(\boldsymbol{r}, t)$. If its space integral is not conserved in time, it is called the nonconserved order parameter, representing magnetization, electric polarization, etc. After appropriate scaling of time $t$, space $\boldsymbol{r}$, and $\psi$, the simplest dynamic equation reads

$$\frac{\partial}{\partial t}\psi = \nabla^2\psi - \tau\psi - \psi^3 + h + \theta \qquad [1]$$

The coefficient $\tau$ is related to the temperature by $\tau = A(T - T_c)$, where $A$ is a constant and $T_c$ is the critical temperature. The constant $h$ is also an externally controllable parameter, proportional to the applied magnetic field for the ferromagnetic case. The last term is the Markovian Gaussian random noise needed when eqn [1] is treated as a Langevin (stochastic differential) equation. In physics its stochastic property is usually expressed as

$$\langle\theta(\boldsymbol{r}, t)\theta(\boldsymbol{r}', t')\rangle = 2\varepsilon\delta(\boldsymbol{r} - \boldsymbol{r}')\delta(t - t') \qquad [2]$$

where $\varepsilon$ represents the strength of the noise (proportional to the temperature before the scaling). In the presence of $\theta$, the variable $\psi$ is a random variable, whose probability distribution $P(\{\psi\}, t)$

obeys the Fokker–Planck equation. The equilibrium (steady) distribution is given by

$$P_{\mathrm{eq}}\{\psi\} = \text{const. } \exp(-F\{\psi\}/\varepsilon) \qquad [3]$$

where

$$F = \int \mathrm{d}\boldsymbol{r}\left[\frac{\tau}{2}\psi^2 + \frac{1}{4}\psi^4 + \frac{1}{2}|\nabla\psi|^2 - h\psi\right] \qquad [4]$$

is the so-called Ginzburg–Landau free energy. Using $F$ we rewrite eqn [1] in a standard form of the Langevin equation,

$$\frac{\partial}{\partial t}\psi = -\frac{\delta F}{\delta \psi} + \theta \qquad [5]$$

In equilibrium $\psi$ consists of the average $\psi_e$ and the deviation $\delta\psi$, where the latter is a Gaussian fluctuation in the limit of small $\varepsilon$. If $\tau > 0$ and $h = 0$, we obtain $\psi_e = 0$. If $\tau < 0$ and $h = 0$, there are two minima $\psi_e = \pm|\tau|^{1/2}$. These two states can coexist in equilibrium with a planar interface separating them at $h = 0$. If its normal is along the $x$-axis, the interface solution is of the form

$$\psi(x) = |\tau|^{1/2}\tanh(|\tau|^{1/2}x/\sqrt{2}) \qquad [6]$$

which tends to $\pm|\tau|^{1/2}$ as $x \pm \infty$ and satisfies

$$\delta F/\delta\psi = (\tau + \psi^2)\psi - \mathrm{d}^2\psi/\mathrm{d}x^2 = 0 \qquad [7]$$

It is well known that the fluctuations of $\psi$ are increasingly enhanced near the critical point. The renormalization group theory shows how the equilibrium distribution $P_{\mathrm{eq}}\{\psi\}$ in eqn [3] depends on the upper cutoff wave number $\Lambda$ of $\psi$, where we suppose that $\psi$ consists of the Fourier components $\psi_{\boldsymbol{k}}$ with $k < \Lambda$ (Onuki 2002). In our phase-ordering problem the shortest relevant spatial scale is the interface width of the order of the thermal correlation length $\xi$ at the final temperature. Therefore, near criticality, we may assume that the thermal fluctuations with wave numbers larger than $\xi^{-1}$ have been eliminated in the model (or $\Lambda \sim \xi^{-1}$ at the starting point).

## Domain Growth

Thermodynamic instability occurs when $\tau$ is changed from a positive value $\tau_{\mathrm{i}}$ to a negative value $\tau_{\mathrm{f}}$ at $t = 0$. We here assume $h = 0$. We set $\tau_{\mathrm{f}} = -1$ using the scaling. At long wavelengths $k < 1$, small plane wave fluctuations with wave vector $\boldsymbol{k}$ grow exponentially as

$$\psi_{\boldsymbol{k}}(t) \sim \exp[(1 - k^2)t] \qquad [8]$$

with the growth rate largest at $k = 0$. This suggests that the nonlinear term in eqn [1] becomes crucial after a transient time. Numerically obtained snapshots of the subsequent $\psi(\boldsymbol{r}, t)$ are shown in Figure 1

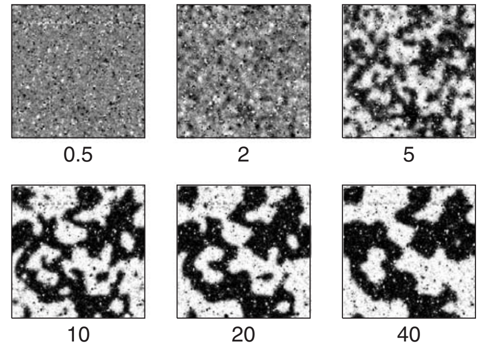



**Figure 1** Time evolution of $\psi$ in model [1] in 2D with system length = 128. The numbers are the times after quenching. Noise is added, but is not essential for large patterns or in the late stage. Reproduced with permission from Onuki A (2002) *Phase Transition Dynamics.* Cambridge, UK: Cambridge University Press.

in two dimensions (2D), where we can see the coarsening of the patterns. The characteristic domain size $\ell(t)$ grows algebraically as

$$\ell(t) \sim t^a \qquad [9]$$

where $a = 1/2$ is known for the model [1]. Scattering experiments detect the time-dependent correlation

$$g(r, t) = \langle \delta\psi(\boldsymbol{r} + \boldsymbol{r}_0, t)\delta\psi(\boldsymbol{r}_0, t)\rangle \qquad [10]$$

$$S(k, t) = \int \mathrm{d}\boldsymbol{r} g(r, t)\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}} \qquad [11]$$

where $S(k, t)$ is called the structure factor. We assume the translational invariance and the spatial isotropy after the thermal average $\langle\cdots\rangle$. If $\tau_i \gg 1$, the quartic term in $F$ is negligible, leading to the initial structure factor

$$S(k, 0) \cong \varepsilon/(\tau_{\mathrm{i}} + k^2) \qquad [12]$$

which is produced by the thermal fluctuations. However, when the domain size $\ell(t)$ much exceeds the microscopic length (lattice constant), the following scaling behavior emerges:

$$g(r, t) = G(r/\ell(t)) \qquad [13]$$

$$S(k, t) = \ell(t)^d Q(\ell(t)k) \qquad [14]$$

where $d$ is the space dimensionality and $G(x)$ and $Q(x)$ are the scaling functions of order unity for $x \sim 1$. The correlation on the scale of $\ell(t)$ in eqn [13] arises from large-scale domain structures, while eqn [14] is simply its Fourier transformation. The maximum of the structure factor grows as $\ell(t)^d$. When $\varepsilon \ll 1$, however, there can be a well-defined initial stage in which $S(k, t)$ grows exponentially at long wavelengths.

We may explain the roles of the terms on the right-hand side of eqn [1] in phase ordering in a simple manner.

1. The linear term $-\tau\psi$ triggers instability for $\tau < 0$.
2. The nonlinear term $-\psi^3$ gives rise to saturation of $\psi$ into $\pm 1$. To see this, we neglect $\nabla^2\psi$ and $\theta$ to have $\partial\psi/\partial t = (1-\psi^2)\psi$ for $\tau = -1$. This equation is solved to give

$$\psi(t) = \psi_0 \Big/ \sqrt{\psi_0^2 + (1-\psi_0^2)\mathrm{e}^{-2t}} \qquad [15]$$

where $\psi_0 = \psi(0)$ is the initial value. Thus, $\psi \to 1$ for $\psi_0 > 0$ and $\psi \to -1$ for $\psi_0 < 0$ as $t \to \infty$.
3. The gradient term limits the instability only in the long wavelength region $k < 1$ in the initial stage (see eqn [8]) and creates the interfaces in the late stage (see eqn [7]).
4. The noise term $\theta$ is relevant only in the early stage where $\psi$ is still on the order of the initial thermal fluctuations. The range of the early stage is of order 1 for $\varepsilon \gtrsim 1$, but weakly grows as $\ln(1/\varepsilon)$ for $\varepsilon \ll 1$. The noise term can be neglected once the fluctuations much exceed the thermal level.
5. If $h$ is a small positive number, it favors growth of regions with $\psi \cong 1$.

### Interface Dynamics

At long times $t \gg 1$ domains with typical size $\ell(t)$ are separated by sharp interfaces and the thermal noise is negligible. Allowing the presence of a small positive $h$, we may approximate the free energy $F$ as

$$F = \sigma S(t) - 2hV_+(t) + \text{const.} \qquad [16]$$

where $\sigma$ is a constant (surface tension), $S(t)$ is the surface area, and $V_+(t)$ is the volume of the regions with $\psi \cong 1$. In this stage the interface velocity $v_{\text{int}} = \boldsymbol{v}_{\text{int}} \cdot \boldsymbol{n}$ is given by the Allen–Cahn formula (Allen and Cahn 1979):

$$v_{\text{int}} = -\mathcal{K} + (2/\sigma)h \qquad [17]$$

The normal unit vector $\boldsymbol{n}$ is from a region with $\psi \cong 1$ to a region with $\psi \cong -1$. The $\mathcal{K}$ is the sum of the principal curvatures $1/R_1 + 1/R_2$ in 3D. This equation can be derived from eqn [1]. If the interface position $\boldsymbol{r}_a$ moves to $\boldsymbol{r}_a + \delta\zeta\boldsymbol{n}$ infinitesimally, the surface area changes by $\delta S = \int \mathrm{d}a\mathcal{K}\delta\zeta$, where $\int \mathrm{d}a \cdots$ denotes the surface integral. Therefore, $F$ in eqn [16] changes in time as

$$\frac{\mathrm{d}F}{\mathrm{d}t} = \int \mathrm{d}a(\sigma\mathcal{K} - 2h)v_{\text{int}} \le 0 \qquad [18]$$

which is non-negative-definite owing to eqn [17]. Furthermore, we may draw three results from eqn [17].

1. If we set $v_{\text{int}} \sim \ell(t)/t$ and $\mathcal{K} \sim 1/\ell(t)$, we obtain $a = 1/2$ in the growth law [9].
2. In phase ordering under very small positive $h$, the balance $1/\ell(t) \sim h/\sigma$ yields the crossover time $t_h \sim h^{-2}$. For $t < t_h$ the effect of $h$ is small, while for $t > t_h$ the region with $\psi \cong 1$ becomes predominant.
3. A spherical droplet with $\psi \cong 1$ evolves as

$$\frac{\partial R}{\partial t} = -\frac{2}{R} + \frac{2h}{\sigma} \qquad [19]$$

from which the critical radius is determined as

$$R_{\mathrm{c}} = \sigma/h \qquad [20]$$

A droplet with $R > R_{\mathrm{c}}(R < R_{\mathrm{c}})$ grows (shrinks).

We mention a statistical theory of interface dynamics at $h = 0$ by Ohta (1982). There, a smooth subsidiary field $u(\boldsymbol{r}, t)$ is introduced to represent surfaces by $u = \text{const}$. The differential geometry is much simplified in terms of such a field. The two-phase boundaries are represented by $u = 0$. If all the surfaces follow $v_{\text{int}} = -\mathcal{K}$ in eqn [17] in the whole space, $u$ obeys

$$\frac{\partial}{\partial t}u = \Big[\nabla^2 - \sum_{ij} n_i n_j \nabla_i \nabla_j\Big]u \qquad [21]$$

where $\nabla_i = \partial/\partial x_i$ and $n_i = \nabla_i u/|\nabla u|$. This equation becomes a linear diffusion equation if $n_i n_j \nabla_i \nabla_j$ is replaced by $d^{-1}\delta_{ij}\nabla^2$. Then $u$ can be expressed in terms of its initial value and the correlation function of $\psi(\boldsymbol{r}, t)(\cong u(\boldsymbol{r}, t)/|u(\boldsymbol{r}, t)|$ in the late stage) is calculated in the form of eqn [13] with

$$G(x) = \frac{2}{\pi}\sin^{-1}\left[\exp\left(-\frac{1}{8(1-1/d)}x^2\right)\right] \qquad [22]$$

which excellently agrees with simulations.

## Spinodal Decomposition in Conserved Systems

The order parameter $\psi$ can be a conserved variable such as the density or composition in fluids or alloys. With the same $F$ in eqn [4], a simple dynamic model in such cases reads

$$\frac{\partial}{\partial t}\psi = \nabla^2 \frac{\delta F}{\delta\psi} - \nabla \cdot \boldsymbol{j}^R \qquad [23]$$

Here $\boldsymbol{j}^R$ is the random current characterized by

$$\left\langle j_\alpha^R(\boldsymbol{r}, t) j_\beta^R(\boldsymbol{r}', t')\right\rangle = 2\varepsilon\delta_{\alpha\beta}\delta(\boldsymbol{r} - \boldsymbol{r}')\delta(t - t') \qquad [24]$$

which ensures the equilibrium distribution [3] of $\psi$. However, the noise $\boldsymbol{j}^R$ is negligible in late-stage phase separation as in the nonconserved case. Note that $h$ in the conserved case is the chemical potential

conjugate to $\psi$ and, if it is homogeneous, it vanishes in the dynamic equation [23]. In experiments the average order parameter

$$M = \langle \psi \rangle = \int d\boldsymbol{r} \psi(\boldsymbol{r})/V \qquad [25]$$

is used as a control parameter instead of $h$, where the integral is within the system with volume $V$. If there is no flux from outside, $M$ is constant in time. Here the instability occurs below the so-called spinodal $M^2 < 1/3(M^2 < |\tau|/3$ for general $\tau < 0)$. In fact, small fluctuations with wave vector $\boldsymbol{k}$ grow exponentially as

$$\psi_{\boldsymbol{k}}(t) \sim \exp[k^2(1 - 3M^2 - k^2)t] \qquad [26]$$

right after the quenching as in eqn [8]. The growth rate is largest at an intermediate wave number $k = k_m$ with

$$k_m = [(1 - 3M^2)/2]^{1/2} \qquad [27]$$

This behavior and the exponential growth of the structure factor have been observed in polymer mixtures where the parameter $\varepsilon$ in eqn [3] or [12] is expected to be small (Onuki 2002). In late-stage coarsening the peak position of $S(k, t)$ decreases in time as

$$k_m(t) \sim 2\pi/\ell(t) \qquad [28]$$

in terms of the domain size $\ell(t)$. The growth exponent in eqn [9] is given by 1/3 for the simple model [23] (see eqn [33] below).

Figure 2 shows the patterns after quenching in 2D. For $M = 0$ the two phases are symmetric and the patterns are bicontinuous, while for $M \neq 0$ the minority phase eventually appears as droplets in the percolating region of the majority phase.

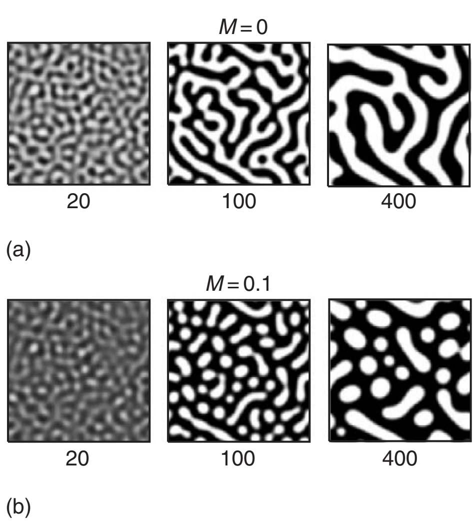



**Figure 2** Time evolution of $\psi$ in model [23] in 2D with system length = 128 without thermal noise: (a) $M = 0$ and (b) $M = 0.1$. The numbers are the times after quenching. Reproduced with permission from Onuki A (2002) *Phase Transition Dynamics.* Cambridge, UK: Cambridge University Press.

### Interface Dynamics

Interface dynamics in the conserved case is much more complicated than in the nonconserved case, because the coarsening can proceed only through diffusion. Long-distance correlations arise among the domains and the interface velocity cannot be written in terms of the local quantities like the curvature. As a simple example, we give the counterpart of eqn [19]. In 3D a spherical droplet with $\psi \cong 1$ appears in a nearly homogeneous matrix with $\psi = M$ far from the droplet. The droplet radius $R$ is then governed by (Lifshitz and Slyozov 1961)

$$\frac{\partial}{\partial t} R = D\left(\frac{\Delta}{R} - \frac{2d_0}{R^2}\right) \qquad [29]$$

where $\Delta = (M + 1)/2$ is called the supersaturation, while $D$ and $d_0$ are constants (equal to 2 and $\sigma/8$, respectively, after the scaling). The critical radius is written as

$$R_c = 2d_0/\Delta \qquad [30]$$

The general definition of the supersaturation is

$$\Delta = \left(M - \psi_{cx}^{(2)}\right) \Big/ \left(\psi_{cx}^{(1)} - \psi_{cx}^{(2)}\right) \qquad [31]$$

Here the equilibrium values of $\psi$ are written as $\psi_{cx}^{(1)}$ and $\psi_{cx}^{(2)}$ and $M$ is supposed to be slightly different from $\psi_{cx}^{(2)}$.

Lifshitz and Slyozov (1961) analyzed domain coarsening in binary AB alloys when the volume fraction $q$ of the A-rich domains is small. They noticed that the supersaturation $\Delta$ around each domain decreases in time with coarsening. That is, the A component atoms in the B-rich matrix are slowly absorbed onto the growing A-rich domains, while a certain fraction of the A-rich domains disappear. Thus, $q(t)$ and $\Delta(t)$ both depend on time, but satisfy the conservation law

$$q(t) + \Delta(t) = \Delta(0) = (M + 1)/2 \qquad [32]$$

With this overall constraint, they found the asymptotic late-stage behavior

$$\ell(t) \sim \Delta(t)^{-1} \sim t^{1/3} \qquad [33]$$

where $\ell(t)$ is the average droplet radius. Notice that this behavior is consistent with the droplet equation [29], where each term is of order $R/t \sim t^{-2/3}$.

## Nucleation

In metastable states the free energy is at a local minimum but not at the true minimum. Such states

are stable for infinitesimal fluctuations, but rare spatially localized fluctuations, called critical nuclei, can continue to grow, leading to macroscopic phase ordering (Onuki 2002, Debenedetti 1996). The birth of a critical droplet is governed by the Boltzmann factor $\exp(-F_c/k_BT)$ at finite temperatures, where $F_c$ is the free energy needed to create a critical droplet and $k_BT$ is the thermal energy with $k_B$ being the Boltzmann constant. In this section we explicitly write $k_BT$, but we may scale $\psi$ and space such that $\tau = -1$ at the final temperature.

### Droplet Free Energy and Experiments

In the nonconserved case we prepare a spin-down state with $\psi \cong -1$ in the time region $t < 0$ and then apply a small positive field $h$ at $t = 0$. For $t > 0$ a spin-up droplet with radius $R$ requires a free energy change

$$F(R) = 4\pi\sigma R^2 - \frac{8\pi}{3}hR^3 \qquad [34]$$

The first term is the surface free energy and the second term is the bulk decrease due to $h$. The critical radius $R_c$ in eqn [20] gives the maximum of $F(R)$ given by

$$F_c = \frac{4\pi}{3}\sigma R_c^2 \qquad [35]$$

In fact, $F'(R) = \partial F(R)/\partial R$ is written as

$$F'(R) = 8\pi\sigma(R - R^2/R_c) \qquad [36]$$

In conserved systems such as fluids or alloys, we lower the temperature slightly below the coexistence curve with the average order parameter $M$ held fixed. We again obtain the droplet free energy [34], but

$$h = (\sigma/2d_0)\Delta \qquad [37]$$

in terms of the (initial) supersaturation $\Delta = \Delta(0)$. Let the equilibrium values $\psi_{cx}^{(1)}$ and $\psi_{cx}^{(2)}$ in the two phases be written as $\pm A(T_c - T)^\beta$ with $A$ and $\beta$ being constants ($\beta \cong 1/3$ as $T \to T_c$). For each given $M$, we define the coexistence temperature $T_{cx}$ by $M = \psi_{cx}^{(2)} = -A(T_c - T_{cx})^\beta$. In nucleation experiments the final temperature $T$ is slightly below $T_{cx}$ and $\delta T \equiv T_{cx} - T$ is a positive temperature increment. For small $\delta T$ we find

$$\Delta \cong \frac{\beta}{2}\delta T/(T_c - T_{cx}) \qquad [38]$$

### Droplet Size Distribution and Nucleation Rate

In a homogeneous metastable matrix, droplets of the new phase appear as rare thermal fluctuations. We describe this process by adding a thermal noise term to the droplet equation [19] or [29]. The droplet size distribution $n(R,t)$ then obeys the Fokker–Planck equation

$$\frac{\partial}{\partial t}n = \frac{\partial}{\partial R}\mathcal{L}(R)\left[\frac{\partial}{\partial R} + \frac{F'(R)}{k_BT}\right]n \qquad [39]$$

Here $n(R,t)\mathrm{d}R$ denotes the droplet number density in the range $[R, R+\mathrm{d}R]$. We determine the kinetic coefficient $\mathcal{L}(R)$ such that

$$v(R) \equiv -\mathcal{L}(R)F'(R)/k_BT \qquad [40]$$

is the right-hand side of eqn [19] or [29]. It is equal to $\partial R/\partial t$ when the thermal noise is neglected. Thus, $\mathcal{L}(R) \propto R^{-2}$ or $R^{-3}$ for the nonconserved or conserved case. The second derivative $(\partial/\partial R)\mathcal{L}(R)(\partial/\partial R)$ in eqn [39] stems from the thermal noise and is negligible for $R - R_c \gtrsim 1$ in 3D (Onuki 2002). Hence, for $R - R_c \gtrsim 1$, the droplets follow the deterministic equation [19] or [29] and $n$ obeys

$$\frac{\partial}{\partial t}n = -\frac{\partial}{\partial R}[v(R)n] \qquad [41]$$

In **Figure 3**, we plot the solution of eqn [39] for the conserved case with $F_c/k_BT = 17.4$ (Onuki 2002). The time is measured in units of $1/\Gamma_c$, which is the timescale of a critical droplet defined by

$$\Gamma_c = (\partial v(R)/\partial R)_{R=R_c} \qquad [42]$$

We notice $\Gamma_c \propto R_c^{-3}$ from eqn [29] so $\Gamma_c$ is small. The initial distribution is given by

$$n(R,0) = n_0 \exp(-4\pi\sigma R^2/k_BT) \qquad [43]$$

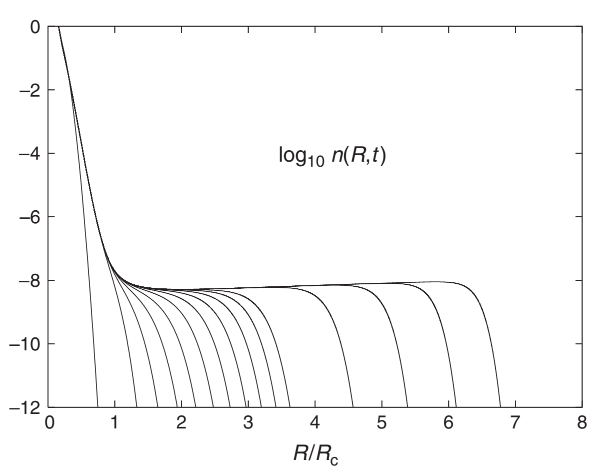



**Figure 3** Time evolution of the droplet size distribution $n(R,t)$ on a semilogarithmic scale as a solution of eqn [39] in the 3D conserved case. The first 11 curves correspond to the times at $\Gamma_c t = 0, 1, \ldots$ and 10. The last four curves are those at $\Gamma_c t = 15, 20, 25,$ and 30. Reproduced with permission from Onuki A (2002) *Phase Transition Dynamics.* Cambridge, UK: Cambridge University Press.

with $n_0$ being a constant number density. This form has been observed in computer simulations as the droplet size distribution on the coexistence curve ($h = 0$). Figure 3 indicates that $n(R, t)$ tends to a steady solution $n_s(R)$ which satisfies

$$\mathcal{L}(R)\left[\frac{\partial}{\partial R} + \frac{F'(R)}{k_B T}\right] n_s = -I \quad [44]$$

where $I$ is a constant. Imposing the condition $n_s(R) \to 0$ as $R \to \infty$, we integrate the above equation as

$$n_s(R) = I \int_R^\infty \mathrm{d}R_1 \frac{1}{\mathcal{L}(R_1)} \exp\left[\frac{F(R_1) - F(R)}{k_B T}\right] \quad [45]$$

For $R - R_c \gg 1$ we may replace $F(R_1) - F(R)$ by $F'(R)(R_1 - R)$ in the integrand of eqn [45] to obtain

$$n_s(R) \cong I/v(R) \quad [46]$$

which also follows from eqn [41]. Thus

$$n_s(R)\mathrm{d}R = I\,\mathrm{d}t \quad (\mathrm{d}R = v(R)\mathrm{d}t) \quad [47]$$

This means that $I$ is the nucleation rate of droplets with radii larger than $R_c$ emerging per unit volume and per unit time. Furthermore, as $R \to 0$, we require $n_s(R) \to n_0 = \text{const.}$ in eqn [43] so that

$$n_0 = I \int_0^\infty \mathrm{d}R_1 \frac{1}{\mathcal{L}(R_1)} \exp\left[\frac{F(R_1)}{k_B T}\right] \quad [48]$$

where the integrand becomes maximum around $R_c$. Using the expansion $F(R) = F_c + F''(R_c)\ (R - R_c)^2/2 + \cdots$, we obtain the famous formula for the nucleation rate

$$I = I_0\ \exp(-F_c/k_B T) \quad [49]$$

$$= I_0\ \exp(-C_0/\Delta^2) \quad [50]$$

where the coefficient $I_0$ is of order $n_0 \Gamma_c$. The second line holds in the 3D conserved case. Here, $C_0 \sim 10^{-3}$ typically and $I_0$ is a very large number in units of $\mathrm{cm}^{-3}\,\mathrm{s}^{-1}$, say, $10^{30}$. Then the exponential factor in $I$ changes abruptly from a very small to a very large number with only a slight increase of $\Delta$ at small $\Delta \ll 1$. For example, if $C_0/\Delta^2 = 50$, $I$ is increased by $\exp(100\delta\Delta/\Delta)$ with a small increase of $\Delta$ to $\Delta + \delta\Delta$. This factor can be of order $10^3$ even for $\delta\Delta/\Delta = 0.05$. Unless very close to criticality, simple metastable fluids become opaque suddenly with increasing $\Delta$ or $\delta T$ at a rather definite cloud point. In near-critical fluids, however, $I_0$ itself becomes small ($\propto \xi^{-6}$) such that the cloud point considerably depends on the experimental timescale (observation time).

## Remarks

The order parameter can be a scalar, a vector as in the Heisenberg spin system, a tensor as in liquid crystals, and a complex number as in superfluids and superconductors. In phase ordering a crucial role is played by topological singularities like interfaces in the scalar case and vortices in the complex number case. Furthermore, a rich variety of phase transition dynamics can be explained if the order parameter is coupled to other relevant variables in the free energy and/or in the dynamic equations. We mention couplings to velocity field in fluids, electrostatic field in charged systems, and elastic field in solids. Phase ordering can also be influenced profoundly by external fields such as electric field or shear flow.

*See also:* Reflection Positivity and Phase Transitions; Renormalization: Statistical Mechanics and Condensed Matter; Statistical Mechanics of Interfaces; Topological Defects and Their Homotopy Classification.

## Further Reading

Allen SM and Cahn JW (1979) Microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening. *Acta Metallurgica* 27: 1085.

Binder K (1991) Spinodal decomposition. In: Cohen RW, Haasen P, and Kramer EJ (eds.) *Material Sciences and Technology*, vol. 5. Weinheim: VCH.

Bray AJ (1994) Theory of phase-ordering kinetics. *Advances in Physics* 43: 357.

Cahn JW (1961) On spinodal decomposition. *Acta Metallurgica* 9: 795.

Debenedetti PG (1996) *Metastable Liquids*. Princeton: Princeton University.

Gunton JD, San Miguel M, and Sani PS (1983) The dynamics of first-order phase transitions. In: Domb C and Lebowitz JL (eds.) *Phase Transition and Critical Phenomena*, vol. 8. London: Academic Press.

Lifshitz IM and Slyozov VV (1961) The kinetics of precipitation from supersaturated solid solutions. *Journal of Physics and Chemistry of Solids* 19: 35.

Ohta T, Jasnow D, and Kawasaki K (1982) Universal scaling in the motion of random interfaces. *Physical Review Letters* 49: 1223.

Onuki A (2002) *Phase Transition Dynamics*. Cambridge: Cambridge University Press.

# Phase Transitions in Continuous Systems

**E Presutti**, Università di Roma “Tor Vergata,” Rome, Italy



## Introduction

Many aspects of our everyday life, from weather to boiling water for a cup of coffee, involve heat exchanges and variations of pressure and, as a result, a phase transition. The general theory behind these phenomena is thermodynamics, which studies fluids and macroscopic bodies under these and more general transformations.

In the simple case of a one-component substance, the behavior under changes of temperature $T$ and pressure $P$ is described, according to the Gibbs phase rule, by a phase diagram such as the one in Figure 1. The curves in the $(T, P)$ plane, distinguish regions where the substance is in its solid, liquid, and gas phases. Thus, in an experiment where we vary the pressure and temperature moving along a line which crosses a transition curve, we observe an abrupt and dramatic change at the crossing, when the system changes phase. As already stated, everyday life is an active source of examples of such phenomena.

The picture is “far from innocent”, it states that air, liquid, and solid are not different elements of nature, as for long believed, but just different aspects of the same thing: substances are able to adapt to different external conditions in dramatically different ways. What properties of intermolecular forces are responsible for such astonishing behavior? The question has been extensively studied and it is the argument of the present article, where it will be discussed in the framework of statistical mechanics for continuous systems. Before entering into the matter, let us mention two basic motivations.

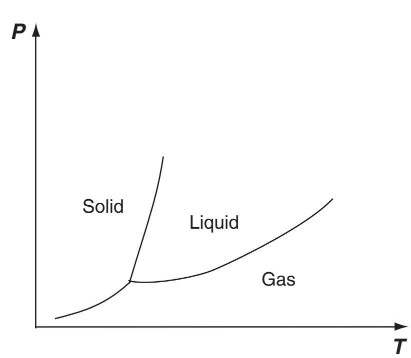



**Figure 1** Phase diagram of a one-component substance.

As always, there is a “fundamental theory” aspect; in the specific case it is the attempt for an atomistic theory able to describe also macroscopic phenomena, thus ranging from the angstrom to the kilometer scales. From an engineering point of view, the target is, for instance, to understand why and when a substance is an insulator, or a conductor or, maybe, a superconductor, and, more importantly, how should we change its microscopic interactions to produce such effects: this opens the way to technologies which are indeed enormously affecting our life.

## Phase Transitions and Statistical Mechanics

The modern theory of statistical mechanics is based upon the Gibbs hypothesis. In a classical (i.e., not quantum) framework, the macroscopic states are described by probability measures on a particle configuration phase space. The equilibrium states are then selected by the Gibbs prescription, which requires that the probability of observing a configuration which has energy $E$ should be proportional to $e^{-\beta E}$, where $\beta = 1/kT$, $k$ is the Boltzmann constant, and $T$ the absolute temperature. These are the “Gibbs measures” and the purpose of statistical mechanics is to study their properties. A prerequisite for the success of the theory is compatibility with the principles of thermodynamics, the theory should then be able to explain the origin of the various phase diagrams and in particular to determine the circumstances under which phase transitions appear.

The theory, commonly called DLR, after Dobrushin, Lanford, and Ruelle, who, in the 1960s, contributed greatly to its foundations, has solid mathematical basis. Its main success is a rigorous proof of consistency with thermodynamics, which is derived under the only assumption that surface effects are negligible, a condition which is mathematically achieved by studying the system in a “thermodynamic limit,” where the region containing the system invades the whole space.

In the thermodynamic limit, the equilibrium states can no longer be defined by the Gibbs prescription, because the energy of configurations in the whole space, being extensive, is typically infinite. The problem has been solved by first proving convergence of the finite-volume Gibbs measures in the thermodynamic limit. After defining the limit states, called “DLR states,” as the equilibrium states of the

infinite systems, it is proved that the DLR states can be directly characterized (i.e., without using limit procedures) as the solutions of a set of equations, the "DLR equations," which generalize the finite-volume Gibbs prescription.

In terms of DLR states, the mathematical meaning of phase transitions becomes very clear and sharp. The starting point is the proof that the physical property that intensive variables in a pure phase have negligible fluctuations is verified by all the DLR measures which are in a special class, thus selected by this property, and which are therefore interpreted as "pure phases." All the other DLR measures are proved to be mixtures, that is, general convex combinations, of the pure DLR states. Thus, in the DLR theory, the system is in a single phase when there is only one DLR state, at the given values of the thermodynamic parameters (e.g., temperature and chemical potential), while the system is at a phase transition if there are several distinct DLR states.

While the theory beautifully clarifies the meaning of phase transitions, it does not say whether the phenomenon really occurs! This is maybe the main open problem in equilibrium statistical mechanics. A general proof of existence of phase diagrams is needed, which should at least capture the basic property behind the Gibbs phase rule, namely that in most of the space (of thermodynamic parameters) there is a single phase, with rare exceptions where several phases coexist. A more refined result should then indicate that coexistence occurs only on regular surfaces of positive codimension.

There is, however, a general result of existence of the gaseous phase, with a proof of uniqueness of DLR measures when temperature is large and density low. Coexistence of phases is much less understood at a general level, but results for particular classes of models exist, for instance, in lattice systems at low temperatures. The prototype is the ferromagnetic Ising model in two or more dimensions, where indeed the full diagram has been determined, see Figure 2. The transition curve is the segment $\{0 \le T \le T_c, h = 0\}$, in the $(T, h)$ plane, $h$ being the magnetic field. In the upper-half plane, there is a single phase with positive magnetization, in the lower one with a negative value; at $h = 0$, positive and negative magnetization states can coexist, if the temperature is lower than the critical value $T_c$. Correspondingly, there are, simultaneously, a positive and a distinctly negative DLR state, which describe the two phases.

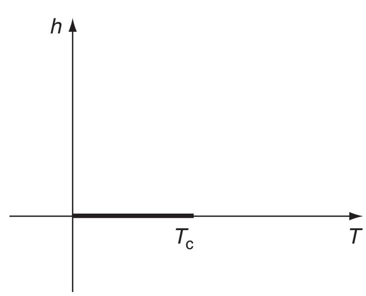



**Figure 2** Phase diagram of the Ising ferromagnet.

An analogous result is missing for systems of particles in the continuum, but there has been recent progress on the analysis of the liquid–vapor branch of the phase diagram, and the issue will be the main focus of this article.

## Sensitive Dependence on Boundary Conditions

Phase transitions describe exceptional regimes where the system is in a critical state; this is why they are so interesting and difficult to study. As in chaotic systems, criticality corresponds to a "butterfly effect," which, in a statistical-mechanics setting means changing far-away boundary conditions. Such changes affect the neighbors, which in turn influence their neighbors, and so on. In general, the effect decays with the distance but, at phase transition, it provokes an avalanche which propagates throughout the system reaching all its points. Its occurrence is not at all obvious, if we remember the stochastic nature of the theory. The domino effect described above can in fact, at each step, be subverted by stochastic fluctuations. The latter, in the end, may completely hide the effect of changing the boundary conditions. This is an instance of a competition between energy and entropy which is the ruling phenomenon behind phase transitions.

This intuitive picture also explains the relevance of space dimensionality. In a many-dimensional space, the influence of the boundary conditions has clearly many more ways to percolate, in contrast to the one-dimensional case, where in fact there is a general result on the uniqueness of DLR measures and therefore absence of phase transitions, for short-range interactions. For pair potentials, "short" means that the interaction energy between two molecules, respectively at $r$ and $r'$, decays as $|r - r'|^{-\alpha}$, $\alpha > 2$. There are results on the converse, namely on the presence of phase transitions when the above condition is not satisfied, mainly for lattice systems, but with partial extensions also to continuous systems. One-dimensional and long-range cases are not the main focus of this article, and the issue will not be discussed further here.

## Ising Model

In order to make the previous ideas quantitative, let us first describe the simple case of the Ising model. Ising spin configurations are collections $\{\sigma(x), x \in \mathbb{Z}^d\}$ of $\sigma(x) \in \{\pm 1\}$ magnetic moments called spins. In the nearest-neighbor case, the interaction between two spins is $-J\sigma(x)\sigma(y)$, $J > 0$, if $x$ and $y$ are nearest neighbors on $\mathbb{Z}^d$, or is vanishing otherwise. There are, therefore, two ground states, one with all spins equal to +1 and the other one with all spins equal to –1. Since the Gibbs probability of higher energies vanishes as the temperature goes to zero, these are interpreted as the equilibrium states at temperature $T = 0$.

If $T > 0$, configurations with larger energy will appear, even though depressed by the Gibbs factor, but their occurrence is limited if $T$ is small. In fact, in the ferromagnetic Ising model at zero magnetic field, dimensions $d \geq 2$, and low enough temperature, it has been proved that there are two distinct DLR measures, one called positive and the other negative. The typical configurations in the positive measure are mainly made by positive spins and, in such an "ocean of positive spins" there are rare and small islands of negative spins. The same situation, but with the positive and negative spins interchanged, occurs in the negative DLR state.

The selection of one of these two states can be made by choosing the positive or the negative boundary conditions, which shows how a surface effect, namely putting the boundary spins equal to 1 or –1, has a volume effect, as most of the spins in the system follow the value indicated by the boundary values. Again, this is more and more striking as we note that each spin is random, yet a strong, cooperative effect takes over and controls the system.

The original proof due to Peierls exploits the spin-flip symmetry of the Ising interaction, but it has subsequently been extended to a wider class of systems on the lattice, in the general framework of the "Pirogov–Sinai theory." This theory studies the low-temperature perturbations of ground states and it applies to many lattice systems, proving the existence of a phase transition and determining the structure of the phase diagram in the low-temperature region. The theory, however, does not cover continuous systems, where the low-temperature regime is essentially not understood, with the notable exception of the Widom and Rowlinson model.

## Two Competing Species in the Continuum

The simplest version of the Widom and Rowlinson model has two types of particles, red and black, which are otherwise identical. Particles are massive points and the only interaction is a hard-core interaction among different colors, namely a red and a black particle cannot be closer than $2R_0, R_0 > 0$ being the hard-core radius.

The order parameter for the phase transition is the particle color. For large values of the chemical potential, and thus large densities, there are two states, one essentially red, the other black, while, if the density is low, the colors "are not separated" and there is a unique state. The proof of the statement starts by dividing the particles of a configuration into clusters, each cluster made by a maximal connected component, where two particles are called connected when their mutual distance is $< 2R_0$. Then, in each cluster, all particles have the same color (because of the hard-core exclusion between black and red), and the color is either black or red, with equal probability.

The question of phase transition is then related to cluster percolation, namely the existence of clusters which extend to infinity. If this occurs, then the influence of fixing the color of a particle may propagate infinitely far away, hence the characteristic "sensitive dependence phenomenon" of phase transitions. Percolation and hence phase transitions have been proved to exist in the positive and negative states, if the density is large and, respectively, small. The above argument is a more recent version of the original proof by Ruelle, which goes back to the 1970s.

The key element for the appearance of the phase transition is the competition between two different components, so that the analysis is not useful in explaining the mechanisms for coexistence in the case of identical particles, which are considered in the following.

## Coarse Graining Transformations

The Peierls argument in Ising systems does not seem to extend to the continuum, certainly not in a trivial way. The ground states, in fact, will not be as simple as the constant configurations of a lattice system; they will instead be periodic or quasiperiodic configurations with a complicated dependence on the particle interactions. The typical fluctuations when we raise the temperature above zero have a much richer and complex structure and are correspondingly more difficult to control. Closeness to the ground states at nonzero temperature, as described in the Ising model, would prove the spontaneous breaking of the Euclidean symmetries and the existence of a crystalline phase. The question is, of course, of great interest, but it looks far beyond the reach of our present mathematical techniques.

The simpler Ising picture should instead reappear at the liquid–vapor coexistence line. Looking at the fluid on a proper spatial scale, we should in fact see a density that is essentially constant, except for small and rare fluctuations. Its value will differ in the liquid and in the gaseous states, $\rho_{\mathrm{gas}} < \rho_{\mathrm{liq}}$. Therefore, density is an order parameter for the transition and plays the role of the spin magnetization in the Ising picture.

There are general mathematical techniques developed to translate these ideas into proofs, they involve "coarse graining," "block spin transformations," and "renormalization group" procedures. The starting point is to ideally divide the space into cells. Their size should be chosen to be much larger than the typical microscopic distance between molecules, to depress fluctuations of the particle density in a cell. To study the probability distribution of the latter, we integrate out all the other degrees of freedom. After such a coarse graining, we are left with a system of spins on a lattice, the lattice sites labeling the cells (also called blocks) and each spin (also called block spin) giving the value of the density of particles in the corresponding cell. Translated into the language of block spins, the previous physical analysis of the state of the fluid suggests that most probably, in each block the density is approximately equal to either $\rho_{\mathrm{liq}}$ or $\rho_{\mathrm{gas}}$, and the same in different blocks, except in the case of small and rare fluctuations. If we represent the probability distribution of the block spins in terms of a Gibbs measure (as always possible if the system is in a bounded region), the previous picture is compatible with a new Hamiltonian with a single spin (one-body) potential which favors the two values $\rho_{\mathrm{liq}}$ and $\rho_{\mathrm{gas}}$ and an attractive interaction between spins which suppresses changes from one to the other. A new effective low temperature should finally dampen the fluctuations.

Thus, after coarse graining, the system should be in the same universality class as of the low-temperature Ising model, and we may hope, in this way, to extend to the liquid–vapor branch of the phase diagram the Pirogov–Sinai theory of low-temperature lattice systems. In particular, as in the Ising model, we will then be able to select the liquid or the vapor phases by the introduction of suitable boundary conditions.

The conditional tense arises because the computation of the coarse graining transformation is in general very difficult, if not impossible, to carry out, but there is a class of systems where it has been accomplished. These are systems of identical point particles in $\mathbb{R}^d$, $d \geq 2$, which interact with "special" two- and four-body potentials, having finite range and which can be chosen to be rotation and translation invariant; their specific form will be described later. For such systems, the above coarse graining picture works and it has been proved that in a "small" region of the temperature–chemical potential plane, there is a part of the curve where two distinct phases coexist, while elsewhere in the neighborhood, the phase is unique.

The ideas behind the choice of the Hamiltonian go back to van der Waals, and the Ginzburg–Landau theory, which are milestones in the theory of phase transitions, while the mathematics of variational problems also enters here in an important way. These are briefly discussed in the next sections.

## The van der Waals Liquid–Vapor Transition

Let us then do a step backwards and recall the van der Waals theory of the liquid–vapor transition. As typical intermolecular forces have a strong repulsive core and a rather long attractive tail, in a continuum, mesoscopic approximation of the system will be described by a free-energy functional of the type

$$F(\rho) = \int_\Lambda f^0_{\beta,\lambda}(\rho(r))\mathrm{d}r - \frac{1}{2}\int_{\Lambda\times\Lambda} J(r,r')\rho(r)\rho(r')\mathrm{d}r\,\mathrm{d}r' \qquad [1]$$

where $\rho = \{\rho(r), r \in \Lambda\}$ is the particles density and $\Lambda$ the region where the system is confined, which, for simplicity, is taken here as a torus in $\mathbb{R}^d$, consisting of a cube with periodic boundary conditions. The term $-J(r,r')\rho(r)\rho(r')$, $J(r,r') \geq 0$, is the energy due to the attractive tail of the interaction, which is periodic in $\Lambda$; $f^0_{\beta,\lambda}(\rho) = f^0_{\beta,0}(\rho) - \lambda\rho$ is the free-energy density due to the short, repulsive part of the interaction, $\lambda$ being the chemical potential.

As noted later, [1] can be rigorously derived by a coarse graining transformation; it will be used to build a bridge between the van der Waals theory and the previous block spin analysis of the liquid–vapor phase transition. Let us take for the moment [1] as a primitive notion. By invoking the second principle of thermodynamics, the equilibrium states can be found by minimizing the free-energy functional. Supposing $J$ to be translation invariant, that is, $J(r,r') = J(r+a, r'+a)$, $r$, $r'$, $a \in \mathbb{R}^d$, and calling $\alpha = \int J(r,r')\mathrm{d}r'$ the intensity of $J$, we can rewrite $F(\rho)$ as

$$F(\rho) = \int_\Lambda \left\{ f^0_{\beta,\lambda}(\rho(r)) - \frac{\alpha\rho(r)^2}{2} \right\} \mathrm{d}r + \frac{1}{4}\int_{\Lambda\times\Lambda} J(r,r')[\rho(r) - \rho(r')]^2 \mathrm{d}r\,\mathrm{d}r' \qquad [2]$$

This shows that the minimizer must have $\rho(r)$ constant (so that the second integral is minimized) and equal to any value which minimizes the function $\{f^0_{\beta,\lambda}(\rho) - \alpha\rho^2/2\}$. By thermodynamic principles, the free energy $f^0_{\beta,\lambda}(\rho)$ is convex in $\rho$, but, if $\alpha$ is large enough, the above expression is not convex and, by properly choosing the value of $\lambda$, the minimizers are no longer unique, hence the van der Waals phase transition.

## Kac Potentials

The analogy between the above analysis of [2] and the previous heuristic study of the fluid based on coarse graining is striking. As customary in continuum theory, each mesoscopic point $r$ should be regarded as representative of a cell containing many molecules. Then the functional $F(\rho)$ can be interpreted as the effective Hamiltonian after coarse graining. The role of the one-body term is played in [2] by the curly bracket, which selects two values of $\rho$ (its minimizers, to be identified with $\rho_{\text{liq}}$ and $\rho_{\text{gas}}$); the attractive two-body potential is then related to the last term in [2], as it suppresses the variations of $\rho$. The analogy clearly suggests a strategy for a rigorous proof of phase transitions in the continuum, an approach which has been and still is actively pursued. It will be discussed briefly in the sequel.

The first rigorous derivation of the van der Waals theory in a statistical-mechanics setting goes back to the 1960s and to Kac, who proposed a model where the particle pair interaction is

$$-\alpha\gamma^d\, e^{-\gamma|q_i-q_j|} + \text{hard core}, \quad \gamma, \alpha > 0 \qquad [3]$$

The phase diagram of such systems, after the thermodynamic limit, can be quite explicitly determined in the limit $\gamma \to 0$, where it has been proved to converge to the van der Waals phase diagram, under a proper choice of $f^0_{\beta,\lambda}(\cdot)$ in [1].

The characteristic features of the first term in [3] are: (1) very long range, which scales as $\gamma^{-1}$, and (2) very small intensity, which scales as $\gamma^d$, so that the total intensity of the potential, defined as the integral over the second position, is independent of $\gamma$. The additional hard-core term (which imposes that any two particles cannot get closer than $2R_0, R_0 > 0$ being the hard-core radius) is to ensure stability of matter, that is, to avoid collapse of the whole system on an infinitesimally small region, as it would happen if only the attractive part of the interaction were present.

Derivation of the van der Waals theory has been proved for a general class of Kac potentials, where the exponential term in [3] is replaced by functions whose dependence on $\gamma$ has the same scaling properties as mentioned above (in (1) and (2)), while the hard core can be replaced by suitably repulsive interactions.

The proof, in the version proposed by Lebowitz and Penrose, uses coarse graining and shows that the effective Hamiltonian is well approximated by the van der Waals functional [1], when $\gamma$ is small, while the effective temperature scales as $\gamma^d$. The approximation becomes exact in the limit $\gamma \to 0$, where it reduces the computation of the partition function to the analysis of the minima and the ground states of an effective Hamiltonian which, in the limit $\gamma \to 0$, is exactly the van der Waals functional.

A true proof of phase transitions requires instead to keep $\gamma > 0$ fixed (instead of letting $\gamma \to 0$) and thus to control the difference of the effective Hamiltonian after coarse graining and the van der Waals functional, which is the effective Hamiltonian, but only in the actual limit $\gamma \to 0$. In general, there is no symmetry between the two ground states, unlike in the Ising case where they are related by spin flip, and the Pirogov–Sinai theory thus enters into play. The framework in fact is exactly similar, with the lattice Hamiltonian replaced by the functional and low temperatures by small $\gamma$ (recall that the effective temperature scales as $\gamma^d$). The extension of the theory to such a setting, however, presents difficulties and success has so far been only partial.

## A Model for Phase Transitions in the Continuum

The problem is twofold: to have a good control of (1) the limit theory and (2) the perturbations induced by a nonzero value of the Kac parameter $\gamma$. The former falls in the category of variational problems for integral functionals, whose prototype is the Ginzburg–Landau free energy

$$F^{\text{gl}}(\rho) = \int \{w(\rho) + |\nabla\rho|^2\}\, dr \qquad [4]$$

which can be regarded as an approximation of [2] with $w$ equal to the curly bracket in [2] and $J$ replaced by a $\delta$-function. Minimization problems for this and similar functionals have been widely analyzed in the context of general variational problems theory and partial differential equations (PDEs), and the study of the limit theory can benefit from a vast literature on the subject. The analysis of the corrections due to small $\gamma$ is, however, so far quite limited. To implement the Pirogov–Sinai strategy, we need, in the case of the interaction [3],

a very detailed knowledge of the system without the Kac part of the interaction and with only hard cores. This, however, is so far not available when the particle density is near to close-packing (i.e., the maximal density allowed by the hard-core potential). Replacing hard cores by other short-range repulsive interactions does not help either, and this seems the biggest obstacle to the program.

The difficulty, however, can be avoided by replacing the hard-core potential by a repulsive many-body (more than two) Kac potential, which ensures stability as well. The class of systems covered by the approach is characterized by Hamiltonian of the form

$$H_{\gamma,\lambda}(q) = \int_{\mathbf{R}^d} e_\lambda(\phi_\gamma(r))\mathrm{d}r \qquad [5]$$

where $e_\lambda(\phi)$ is a polynomial of the scalar field variable $\phi$, a specific example being

$$e_\lambda(\phi) = \frac{\phi^4}{4!} - \frac{\phi^2}{2} - \lambda\phi \qquad [6]$$

This form of the Hamiltonian is familiar from Euclidean field theories. In these theories, the free distribution of the field is Gaussian; in our case, however, the field $\phi = \phi_\gamma(r)$ is a function of the particle configurations $q = (q_i, i = 1, \ldots, n)$:

$$\phi_\gamma(r) = j_\gamma * q(r) = \sum_{i=1}^{n} j_\gamma(r, q_i)$$
$$j_\gamma(r, r') = \gamma^d j(\gamma r, \gamma r') \qquad [7]$$

where $j(r, r')$ is a translation-invariant, symmetric transition probability kernel. Thus, $\phi_\gamma(r)$ is a nonnegative variable which has the meaning of a local density at $r$, weighted by the Kac kernel $j_\gamma(r, r')$.

## Contours and Phase Indicators

The dependence on $\gamma$ yields the scaling properties characteristic of the Kac potentials and [5] may be regarded as a generalized Kac Hamiltonian, which, in the polynomial case of [6], involves up to four-body Kac potentials. The phase diagram of the model, after taking first the thermodynamic limit and then the limit $\gamma \to 0$, is determined by the free-energy functional

$$F(\rho) = \int \left\{ e_\lambda(j * \rho(r)) - \frac{S(\rho(r))}{\beta} \right\} \mathrm{d}r \qquad [8]$$

$$S(\rho) = -\rho(\log\rho - 1) \qquad [9]$$

where [8] is taken to be defined on a torus (to avoid convergence problems of the integral), and $j = j_\gamma, \gamma = 1$.

Exploiting the concavity of the entropy $S(\rho)$, it is proved that the minimizers of $F(\cdot)$ are constant functions with the constants minimizing

$$f_{\lambda,\beta}(u) = e_\lambda(u) - \frac{S(u)}{\beta}, \quad u \geq 0 \qquad [10]$$

In the case of [6], to which we restrict in the sequel, for any $\beta > (3/2)^{3/2}$ there is $\lambda_\beta$ so that $f_{\lambda_\beta,\beta}(u)$ is double-well with two minimizers, $\rho_{\mathrm{gas}} < \rho_{\mathrm{liq}}$ (dependence on $\beta$ is omitted).

To "recognize" the densities $\rho_{\mathrm{gas}}$ and $\rho_{\mathrm{liq}}$ in a particle configuration, we use coarse graining and introduce two partitions of $\mathbb{R}^d$ into cubes $C^{(\ell_\mp,\gamma)}$. The cubes $C^{(\ell_-,\gamma)}$ of the first partition have side $\ell_{-,\gamma}$ proportional to $\gamma^{-1+\alpha}, \alpha > 0$ suitably small; those of the second one have length $\ell_{+,\gamma}$ proportional to $\gamma^{-1-\alpha}$; they are chosen so that each cube $C^{(\ell_+,\gamma)}$ is union of cubes $C^{(\ell_-,\gamma)}$. Notice that the small cubes have side much smaller than the interaction range (for small $\gamma$), while the opposite is true for the large cubes.

Given a particle configuration $q$, we say that a point $r$ is in the liquid phase and write $\Theta(r; q) = 1$, if

$$\left| \frac{|q \sqcap C^{(\ell_-,\gamma)}|}{\ell^d_{-\gamma}} - \rho_{\mathrm{liq}} \right| \leq \gamma^a, \quad a > 0 \text{ suitably small} \qquad [11]$$

for any small cube $C^{(\ell_-,\gamma)}$ contained either in $C_r^{(\ell_+,\gamma)}$ or in the cubes $C^{(\ell_+,\gamma)}$ contiguous to $C_r^{(\ell_+,\gamma)}$: $|q \sqcap C^{(\ell_-,\gamma)}|$ is referred to as the number of particles of $q$ in $C^{(\ell_-,\gamma)}$, and $C_r^{(\ell_+,\gamma)}$ as the large cube which contains $r$.

Thus, $\Theta(r; q) = 1$ if the local particle density is constantly close to $\rho_{\mathrm{liq}}$ in a large region around $r$. Defining $\Theta(r; q) = -1$ if the above holds with $\rho_{\mathrm{gas}}$ instead of $\rho_{\mathrm{liq}}$ and setting $\Theta(r; q) = 0$ in all the other cases, we then have a phase indicator $\Theta(r; q)$, which identifies, for all particle configurations, which spatial regions should be attributed to the liquid and gas phases. The connected components of the complementary region are called contours and the definition of $\Theta(r; q)$ has been structured in such a way that liquid and gas are always separated by a contour. The liquid phase will then be represented by a measure which gives large probability to configurations having mostly $\Theta = 1$, while the gas phase by configurations with mostly $\Theta = -1$.

This is quite similar to the Ising picture and, as in the Ising model, the existence of a phase transition follows from a Peierls estimate that contours have small probability. In fact, if there are few contours, the phase imposed on the boundaries of the region

where the system is observed percolates inside, invading most of the space. Thus, boundary conditions select the phase in the whole volume. The absence of the short-range potential, which was the hard-core interaction in [3], and hence the absence of all the difficulties which originate from it, allow one to carry through successfully the Pirogov–Sinai program and prove Peierls estimates on contours and, hence, the existence of a phase transition. In particular, the statistical weight of a contour is estimated by first relating the computation to one involving the functional [8] and then computing its value on density profiles compatible with the existence of the given contour. This part of the problem needs variational analysis for [8], with constraints and benefits of a vast literature on the subject.

The phase transition is very sharp, as shown by the following ideal experiment. Having fixed $\beta > (3/2)^{3/2}$, let $\lambda$ vary in a (suitably) small interval $[\lambda_\beta - \delta, \lambda_\beta + \delta], \delta > 0$, centered around the mean-field critical value $\lambda_\beta$. We consider the system in a large region with, for instance, boundary conditions $\Theta = -1$ (i.e., forcing the gas phase) and fix $\gamma$ small enough. At $\lambda = \lambda_\beta - \delta$, the system has $\Theta = -1$ in most of the domain, and this persists when we increase $\lambda$ till a critical value, $\lambda_{\beta,\gamma}$, close to, but not the same as $\lambda_\beta$. For $\lambda > \lambda_{\beta,\gamma}, \Theta = 1$ in most of the domain, except for a small layer around the boundaries. The analogous picture holds if we choose boundary conditions $\Theta = 1$, and $\lambda = \lambda_{\beta,\gamma}$ is the only value of the chemical potential where the system is sensitive to the boundary conditions and both phases can be produced by the right boundary conditions. The fact that the actual value $\lambda_{\beta,\gamma}$ differs from $\lambda_\beta$, is characteristic of the Pirogov–Sinai approach and enlightens the delicate nature of the proofs.

## Some Related Problems

In this concluding section, two important related problems, which have not been mentioned so far, are discussed.

A natural question, after proving a phase transition, is to describe how two phases coexist, once forced to be simultaneously present in the system. This can be achieved, for instance, by suitable boundary conditions (typically positive and negative on the top and bottom of the spatial domain) or by imposing a total density (or magnetization in the case of spins) intermediate between those of the pure phases. There will then be an interface separating the two phases with a corresponding surface tension and the geometry will be determined by the solution of a variational problem and given by the Wulff shape.

Can statistical mechanics explain and describe the phenomenon? Important progress has been made recently on the subject in the case of lattice systems at low temperatures. The question has also been widely studied at the mesoscopic level, in the context of variational problems for Ginzburg and Landau and many other functionals. Therefore, all the ingredients of further development of the theory in this direction are now present.

We have so far discussed only classical systems; a few words about extensions to the quantum case are now in order. In the range of values of temperatures and densities where the liquid–vapor transition occurs, the quantum effects are not expected to be relevant. Referring to the case of bosons, and away from the Bose condensation regime (and for system with Boltzmann statistics as well), the quantum delocalization of particles caused by the indeterminacy principle should essentially disappear after macroscopic coarse graining, and the block-spin variables should again behave classically, even though their underlying constituents are quantal. If this argument proves correct, then progress along these lines may be expected in near future.

*See also:* Cluster Expansion; Ergodic Theory; Finite Group Symmetry Breaking; Pirogov–Sinai Theory; Reflection Positivity and Phase Transitions; Statistical Mechanics and Combinatorial Problems; Statistical Mechanics of Interfaces; Symmetry Breaking in Field Theory; Two-Dimensional Ising Model.

## Further Reading

Dal Maso G (1993) *An Introduction to Γ-Convergence.* Birkhäuser.

Lebowitz JL, Mazel A, and Presutti E (1999) Liquid–vapor phase transitions for systems with finite range interactions. *Journal of Statistical Physics* 94: 955–1025.

Ruelle D (1969) *Statistical Mechanics. Rigorous Results.* Benjamin.

Sinai YaG (1982) *Theory of Phase Transitions: Rigorous Results.* Pergamon Press (co-edition with Akademiai Kiadó, Budapest).

# Pirogov–Sinai Theory

**R Kotecký**, Charles University, Prague, Czech Republic, and the University of Warwick, UK



## Introduction

Pirogov–Sinai theory is a method developed to study the phase diagrams of lattice models at low temperatures. The general claim is that, under appropriate conditions, the phase diagram of a lattice model is, at low temperatures, a small perturbation of the zero-temperature phase diagram designed by ground states. The treatment can be generalized to cover temperature driven transitions with coexistence of ordered and disordered phases.

## Formulation of the Main Result

### Setting

Refraining first from full generality, we formulate the result for a standard class of lattice models with finite spin state and finite-range interaction. We will mention different generalizations later.

We consider *classical lattice models* on the $d$-dimensional hypercubic lattice $\mathbb{Z}^d$ with $d \geq 2$. A *spin configuration* $\sigma = (\sigma_x)_{x \in \mathbb{Z}^d}$ is an assignment of a spin with values in a finite set $S$ to each lattice site $x \in \mathbb{Z}^d$; the configuration space is $\Omega = S^{\mathbb{Z}^d}$. For $\sigma \in \Omega$ and $\Lambda \subset \mathbb{Z}^d$, we use $\sigma_\Lambda \in \Omega_\Lambda = S^\Lambda$ to denote the restriction $\sigma_\Lambda = \{\sigma_x; x \in \Lambda\}$.

The Hamiltonian is given in terms of a collection of interaction potentials $(\Phi_A)$, where $\Phi_A$ are real functions on $\Omega$, depending only on $\sigma_x$ with $x \in A$, and $A$ runs over all finite subsets of $\mathbb{Z}^d$. We assume that the potential is *periodic* with *finite range of interactions*. Namely, $\Phi_{A'}(\sigma') = \Phi_A(\sigma)$ whenever $A$ and $\sigma$ are related to $A'$ and $\sigma'$ by a translation from $(a\mathbb{Z})^d$ for some fixed integer $a$ and there exists $R \geq 1$ such that $\Phi_A \equiv 0$ for all $A$ with diameter exceeding $R$.

Without loss of generality (possibly multiplying the number $a$ by an integer and increasing $R$), we may assume that $R = a$.

The Hamiltonian $H_\Lambda(\sigma|\eta)$ in $\Lambda$ with boundary conditions $\eta \in \Omega$ is then given by

$$H_\Lambda(\sigma|\eta) = \sum_{A \cap \Lambda \neq \emptyset} \Phi_A(\sigma_\Lambda \vee \eta_{\Lambda^c}) \qquad [1]$$

where $\sigma_\Lambda \vee \eta_{\Lambda^c} \in \Omega$ is the configuration $\sigma_\Lambda$ extended by $\eta_{\Lambda^c}$ on $\Lambda^c$. The *Gibbs state in* $\Lambda$ *under boundary conditions* $\eta \in \Omega$ (and with Hamiltonian $H$) is the probability $\mu_\Lambda(\cdot|\eta)$ on $\Omega_\Lambda$ defined by

$$\mu_\Lambda(\{\sigma_\Lambda\}|\eta) = \frac{\exp\{-\beta H_\Lambda(\sigma|\eta)\}}{Z(\Lambda|\eta)} \qquad [2]$$

with the *partition function*

$$Z(\Lambda|\eta) = \sum_{\sigma_\Lambda} \exp\{-\beta H_\Lambda(\sigma|\eta)\} \qquad [3]$$

We use $\mathcal{G}(H)$ to denote the set of all *periodic Gibbs states with Hamiltonian H* defined on $\Omega$ by means of the Dobrushin–Lanford–Ruelle (DLR) equations.

### Ground-State Phase Diagram and the Removal of Degeneracy

A periodic configuration $\sigma \in \Omega$ is called a (*periodic*) *ground state* of a Hamiltonian $H = (\Phi_A)$ if

$$H(\tilde{\sigma}; \sigma) = \sum_A (\Phi_A(\tilde{\sigma}) - \Phi_A(\sigma)) \geq 0 \qquad [4]$$

for every finite perturbation $\tilde{\sigma} \neq \sigma$ of $\sigma$ ($\tilde{\sigma}$ differs from $\sigma$ at a finite number of lattice sites). We use $g(H)$ to denote the set of all periodic ground states of $H$. For every configuration $\sigma \in g(H)$, we define the *specific energy* $e_\sigma(H)$ by

$$e_\sigma(H) = \lim_{n \to \infty} \frac{1}{|V_n|} \sum_{A \cap V_n \neq \emptyset} \Phi_A(\sigma) \qquad [5]$$

(with $V_n$ denoting a cube consisting of $n^d$ lattice sites).

To investigate the phase diagram, we will consider a parametric class of Hamiltonians around a fixed Hamiltonian $H^{(0)}$ with a finite set of periodic ground states $g(H^{(0)}) = \{\sigma_1, \ldots, \sigma_r\}$. Namely, let $H^{(0)}$, $H^{(1)}, \ldots,$ and $H^{(r-1)}$ be Hamiltonians determined by potentials $\Phi^{(0)}, \Phi^{(1)}, \ldots,$ and $\Phi^{(r-1)}$, respectively, and consider the $(r-1)$-parametric set of Hamiltonians $H_{\boldsymbol{t}} = H^{(0)} + \sum_{\ell=1}^{r-1} t_\ell H^{(\ell)}$ with $\boldsymbol{t} = (t_1, \ldots, t_{r-1}) \in \mathbb{R}^{r-1}$. Using a shorthand $e_m(H) = e_{\sigma_m}(H)$, and introducing the vectors $\boldsymbol{e}(H) = (e_1(H), \ldots, e_r(H))$ and $\boldsymbol{h}(\boldsymbol{t}) = \boldsymbol{e}(H_{\boldsymbol{t}}) - \min_m e_m(H_{\boldsymbol{t}})$, we notice that for each $\boldsymbol{t} \in \mathbb{R}^{r-1}$, the vector $\boldsymbol{h}(\boldsymbol{t}) \in \partial Q_r$, the boundary of the positive octant in $\mathbb{R}^r$. A crucial assumption for such a parametrization $H_{\boldsymbol{t}}$ to yield a meaningful phase diagram is the *condition of removal of degeneracy*: we assume that $g(H^{(0)} + H^{(\ell)}) \subsetneqq g(H^{(0)}), \ell = 1, \ldots, r-1$, and that the vectors $\boldsymbol{e}(H^{(\ell)}), \ell = 1, \ldots, r-1$, are linearly independent.

In particular, its immediate consequence is that the mapping $R^{r-1} \ni \boldsymbol{t} \mapsto \boldsymbol{h}(\boldsymbol{t}) \in \partial Q_r$ is a bijection. This fact has a straightforward interpretation in terms of *ground-state phase diagram*. Viewing the

phase diagram (at zero temperature) as a partition of the parameter space into regions $K_g$ with a given set $g \subset g(H^{(0)})$ of ground states – "coexistence of zero-temperature phases from $g$" – the above bijection means that the region $K_g$ is the preimage of the set

$$Q_g = \{\boldsymbol{h} \in \partial Q_r | h_m = 0 \text{ for } \sigma_m \in g \text{ and } h_m > 0 \text{ otherwise}\} \quad [6]$$

The partition of the set $\partial Q_r$ has a natural hierarchical structure implied by the fact that $\overline{Q_{g_1}} \cap \overline{Q_{g_2}} = \overline{Q_{g_1 \cup g_2}}$ ($\overline{Q_g}$ is the closure of $Q_g$). Namely, the origin $\{0\} = Q_{g(H^{(0)})}$ is the intersection of $r$ positive coordinate axes $\overline{Q_{\{\sigma_{\tilde{m}}, \tilde{m} \neq m\}}}, m = 1, \ldots, r$; each of those half-lines is an intersection of $r-1$ two-dimensional quarter-planes with boundaries on positive coordinate axes, etc., up to $(r-1)$-dimensional planes $\overline{Q_{\{\sigma_m\}}}, m = 1, \ldots, r$. This hierarchical structure is thus inherited by the partition of the parameter space $\mathbb{R}^{r-1}$ into the regions $K_g$. The phase diagrams with such regular structure are sometimes said to satisfy the Gibbs phase rule.

We can thus summarize in a rather trivial conclusion that the condition of removal of degeneracy implies that the ground-state phase diagram obeys the Gibbs phase rule. The task of the Pirogov–Sinai theory is to provide means for proving that this remains true, at least in a neighborhood of the origin of parameter space, also for small nonzero temperatures. To achieve this, we need an effective control of excitation energies.

### Peierls Condition

A crucial assumption for the validity of the Pirogov–Sinai theory is a lower bound on energy of excitations of ground states – the Peierls condition.

In spite of the fact that for a study of phase diagram we consider a parametric set of Hamiltonians whose set of ground states may differ, it is useful to introduce the Peierls condition with respect to a single fixed collection $G$ of *reference configurations* (eventually, it will be identified with the ground states of the Hamiltonian $H^{(0)}$). Let thus a fixed set $G$ of periodic configurations $\{\sigma_1, \ldots, \sigma_r\}$ be given. Again, without loss of generality, we may assume that the periodicity of all configurations $\sigma_m \in G$ is $R$.

Before formulating the Peierls condition, we have to introduce the notion of contours. Consider the set of all *sampling cubes* $C(x) = \{y \in \mathbb{Z}^d \| y_i - x_i| \leq R \text{ for } 1 \leq i \leq d\}, x \in \mathbb{Z}^d$. A *bad cube* of a configuration $\sigma \in \Omega$ is a sampling cube $C$ for which $\sigma_C$ differs from $\sigma_m$ restricted to $C$ for every $\sigma_m \in G$. The boundary $B(\sigma)$ of $\sigma$ is the union of all bad cubes of $\sigma$. If $\sigma_m \in G$ and $\sigma$ is its finite perturbation (differing from $\sigma_m$ on a finite set of lattice sites), then, necessarily, $B(\sigma)$ is finite. A *contour* of $\sigma$ is a pair $\gamma = (\Gamma, \sigma_\Gamma)$, where $\Gamma$ (the *support* of the contour $\gamma$) is a connected component of $B(\sigma)$ (and $\sigma_\Gamma$ is the restriction of $\sigma$ on $\Gamma$). Here, the connectedness of $\Gamma$ means that it cannot be split into two parts whose (Euclidean) distance is larger than 1. We use $\partial(\sigma)$ to denote the set of all contours of $\sigma$, $B(\sigma) = \bigcup_{\gamma \in \partial(\sigma)} \Gamma$.

Consider a configuration $\sigma^\gamma$ such that $\gamma$ is its unique contour. The set $\mathbb{Z}^d \backslash \Gamma$ has one infinite component to be denoted $\operatorname{Ext} \gamma$ and a finite number of finite components whose union will be denoted $\operatorname{Int} \gamma$. Observing that the configuration $\sigma^\gamma$ coincides with one of the states $\sigma_m \in G$ on every component of $\mathbb{Z}^d \backslash B(\sigma)$, each of those components can be labeled by the corresponding $m$. Let $q$ be the label of $\operatorname{Ext} \gamma$, we say that $\gamma$ is a $q$-contour, and let $\operatorname{Int}_m \gamma$ be the union of all components of $\operatorname{Int} \gamma$ labeled by $m, m = 1, \ldots, r$.

Defining the "energy" $\Psi(\gamma)$ of a $q$-contour $\gamma$ by the equation

$$\Psi(\gamma) = H(\sigma^\gamma; \sigma_q) + e_q(H)|\Gamma| - \sum_{m=1}^{r} (e_m(H) - e_q(H))|\operatorname{Int}_m \gamma| \quad [7]$$

the *Peierls condition* with respect to the set $G$ of reference configurations is an assumption of the existence of $\rho > 0$ such that

$$\Psi(\gamma) \geq (\rho + \min_m e_m(H))|\Gamma| \quad [8]$$

for any contour of any configuration $\sigma$ that is a finite perturbation of $\sigma_q \in G$.

Notice that if $G = g(H)$, the sum on the right-hand side of [7] vanishes.

### Phase Diagram

The main claim of the Pirogov–Sinai theory provides, for $\beta$ sufficiently large, a construction of regions $\mathcal{K}_g(\beta)$ of the parameter space characterized by the coexistence of phases labeled by configurations $\sigma_m \in g$. This is done similarly as for the ground-state phase diagram discussed earlier by constructing a homeomorphism $\boldsymbol{t} \mapsto \boldsymbol{a}(\boldsymbol{t})$ from a neighborhood of the origin of the parameter space to a neighborhood of the origin of $\partial Q_r$ that provides the phase diagram (actually, the function $\boldsymbol{a}(\boldsymbol{t})$ will turn out to be just a perturbation of $\boldsymbol{h}(\boldsymbol{t})$ with errors of order $e^{-\beta}$).

Before stating the result, however, we have to clarify what exactly is meant by existence of phase $m$ for a given Hamiltonian $H$. Roughly speaking, it is the existence of a periodic extremal Gibbs state $\mu_m \in \mathcal{G}(H)$, whose typical configurations do not differ too much from the ground-state configuration $\sigma_m$. In more technical terms, the existence of such a state is provided once we prove a

suitable bound, for the finite-volume Gibbs state $\mu_\Lambda(\{\sigma_\Lambda\}|\sigma_m)$ under the boundary conditions $\sigma_m$, on the probability that a fixed point in $\Lambda$ is encircled by a contour from $\partial\sigma$. If this is the case, we say that the *phase m is stable*. It turns out that such a bound is actually an integral part of the construction of metastable free energies $f_m(\boldsymbol{t})$ yielding the homeomorphism $\boldsymbol{t}\mapsto \boldsymbol{a}(\boldsymbol{t})$. In this way, we get the main claim formulated as follows:

**Theorem 1** *Consider a parametric set of Hamiltonians* $H_t = H^{(0)} + \sum_{\ell=1}^{r-1} t_\ell H^{(\ell)}$ *with periodic finite-range interactions satisfying the condition of removal of degeneracy as well as the Peierls condition with respect to the reference set* $G = g(H^{(0)})$. *Let* $d \geq 2$ *and let* $\beta$ *be sufficiently large. Then there exists a homeomorphism* $\boldsymbol{t}\mapsto \boldsymbol{a}(\boldsymbol{t})$ *of a neighborhood* $V_\beta$ *of the origin of the parameter space* $R^{r-1}$ *onto a neighborhood* $U_\beta$ *of the origin of* $\partial Q_r$ *such that, for any* $\boldsymbol{t} \in V_\beta$, *the set of all stable phases is* $\{m \in \{1,\ldots,r\} | a_m(\boldsymbol{t}) = 0\}$.

The Peierls condition can be actually assumed only for the Hamiltonian $H^{(0)}$ inferring its validity for $H_t$ on a sufficiently small neighborhood $V_\beta$.

Notice also that the result can be actually stated not as a claim about phase diagram in a space of parameters, but as a statement about stable phases of a fixed Hamiltonian $H$. Namely, for a Hamiltonian $H$ satisfying Peierls condition with respect to a reference set $G$, one can assure the existence of parameters $a_m$ labeled by elements from $G$ such that the set of extremal periodic Gibbs states of $H$ consists of all those $m$-phases for which $a_m = 0$.

### Construction of Metastable Free Energies

An important part of the Pirogov–Sinai theory is an actual construction of the *metastable free energies* – a set of functions $f_m(\boldsymbol{t}), m = 1,\ldots,r$, that provide the homeomorphism $\boldsymbol{a}(\boldsymbol{t})$ by taking $a_m(\boldsymbol{t}) = f_m(\boldsymbol{t}) - \min_{\bar m} f_{\bar m}(\boldsymbol{t})$.

We start with a *contour representation of partition function* $Z(\Lambda|\sigma_q)$. Considering, for each contributing configuration $\sigma$, the collection $\partial(\sigma)$ of its contours, we notice that, in addition to the fact that different contours $\gamma, \gamma' \in \partial(\sigma)$ have disjoint supports, $\Gamma \cap \Gamma' = \emptyset$, the contours from $\partial(\sigma)$ have to satisfy the *matching conditions*: if $C$ is a connected component of $\mathbb{Z}^d \setminus \bigcup_{\gamma\in\partial}\Gamma$, then the restrictions of the spin configurations $\sigma^\gamma$ to $C$ are the same for all contours $\gamma \in \partial(\sigma)$ with $\mathrm{dist}(\Gamma, C) = 1$. In other words, the contours touching $C$ induce the same label on $C$. Let us observe that there is actually one-to-one correspondence between configurations $\sigma$ that coincide with $\sigma_q$ on $\Lambda^c$ and collections $\mathcal{M}(\Lambda, q)$ of contours $\partial$ in $\Lambda$ satisfying the matching condition, and such that the external among them are $q$-contours. Here, a contour $\gamma \in \partial$ is called an *external contour* in $\partial$ if $\Gamma \subset \mathrm{Ext}\,\gamma'$ for all $\gamma' \in \partial$ different from $\gamma$.

With this observation and using $\Lambda_m(\partial)$ to denote the union of all components of $\Lambda \setminus \bigcup_{\gamma\in\partial}\Gamma$ with label $m$, we get

$$Z(\Lambda|\sigma_q) = \sum_{\partial\in\mathcal{M}(\Lambda,q)} \prod_m \mathrm{e}^{-\beta e_m(H)|\Lambda_m(\gamma)|} \prod_{\gamma\in\partial} \mathrm{e}^{-\beta\Psi(\gamma)} \qquad [9]$$

Usefulness of such contour representations stems from an expectation that, for a stable phase $q$, contours should constitute a suppressed excitation and one should be able to use cluster expansions to evaluate the behavior of the Gibbs state $\mu_q$. However, the direct use of the cluster expansion on [9] is trammeled by the presence of the energy terms $\mathrm{e}^{-\beta e_m(H)|\Lambda_m(\partial)|}$ and, more seriously, by the requirement that the contour labels match.

Nevertheless, one can rewrite the partition function in a form that does not involve any matching condition. Namely, considering first a sum over mutually external contours $\partial^{\mathrm{ext}}$ and resumming over collections of contours which are contained in their interiors without touching the boundary (being thus prevented to "glue" with external contours), we get

$$Z(\Lambda|\sigma_q) = \sum_{\partial^{\mathrm{ext}}} \mathrm{e}^{-\beta e_q(H)|\mathrm{Ext}|} \times \prod_{\gamma\in\partial^{\mathrm{ext}}} \left\{ \mathrm{e}^{-\beta\Psi(\gamma)} \prod_m Z^{\mathrm{dil}}(\mathrm{Int}_m\gamma|\sigma_m) \right\} \qquad [10]$$

Here the sum goes over all collections of compatible external $q$-contours in $\Lambda$, $\mathrm{Ext} = \mathrm{Ext}_\Lambda(\partial^{\mathrm{ext}}) = \bigcap_{\gamma\in\partial^{\mathrm{ext}}} (\mathrm{Ext}\,\gamma \cap \Lambda)$, and the partition function $Z^{\mathrm{dil}}(\Lambda|\sigma_q)$ is defined by [9] with $\mathcal{M}(\Lambda, q)$ replaced by $\mathcal{M}^{\mathrm{dil}}(\Lambda, q) \subset \mathcal{M}(\Lambda, q)$, the set of all those collections whose external countours $\gamma$ are such that $\mathrm{dist}\,(\Gamma, \Lambda^c) > 1$. Multiplying now each term by

$$1 = \prod_{\gamma\in\partial^{\mathrm{ext}}} \prod_m \frac{Z^{\mathrm{dil}}(\mathrm{Int}_m\,\gamma|\sigma_q)}{Z^{\mathrm{dil}}(\mathrm{Int}_m\,\gamma|\sigma_q)} \qquad [11]$$

we get

$$Z(\Lambda|\sigma_q) = \sum_{\partial^{\mathrm{ext}}} \mathrm{e}^{-\beta e_q(H)|\mathrm{Ext}|} \times \prod_{\gamma\in\partial^{\mathrm{ext}}} \left( \mathrm{e}^{-\beta e_q(H)|\Gamma|} w_q(\gamma) Z^{\mathrm{dil}}(\mathrm{Int}\,\gamma|\sigma_q) \right) \qquad [12]$$

where $w_q(\gamma)$ is given by

$$w_q(\gamma) = \mathrm{e}^{-\beta\Psi(\gamma)}\,\mathrm{e}^{\beta e_q(H)|\Gamma|} \prod_m \frac{Z^{\mathrm{dil}}(\mathrm{Int}_m\,\gamma|\sigma_m)}{Z^{\mathrm{dil}}(\mathrm{Int}_m\,\gamma|\sigma_q)} \qquad [13]$$

Observing that a similar expression is valid for $Z^{\mathrm{dil}}(\Lambda|\sigma_q)$ (with an appropriate restriction on the sum over external contours $\partial^{ext}$) and proceeding by induction, we eventually get the representation

$$Z(\Lambda|\sigma_q) = \mathrm{e}^{-\beta e_q(H)|\Lambda|} \sum_{\partial \in \mathcal{C}(\Lambda,q)} \prod_{\gamma \in \partial} w_q(\gamma) \qquad [14]$$

where $\mathcal{C}(\Lambda, q)$ denotes the set of all collections of nonoverlapping $q$-contours in $\Lambda$. Clearly, the sum on the right-hand side is exactly of the form needed to apply cluster expansion, provided the contour weights satisfy the necessary convergence assumptions.

Even though this is not necessarily the case, there is a way to use this representation. Namely, one can artificially change the weights to satisfy the needed bound, for example, by modifying them to the form

$$w'_q(\gamma) = \min\left(w_q(\gamma), \mathrm{e}^{-\tau|\Gamma|}\right) \qquad [15]$$

with a suitable constant $\tau$. The modified partition function

$$Z'(\Lambda|\sigma_q) = \mathrm{e}^{-\beta e_q(H)|\Lambda|} \sum_{\partial \in \mathcal{C}(\Lambda,q)} \prod_{\gamma \in \partial} w'_q(\gamma) \qquad [16]$$

can then be controlled by cluster expansion allowing to *define*

$$f_q(H) = -\frac{1}{\beta} \lim_{|\Lambda|\to\infty} \frac{1}{|\Lambda|} \log Z'(\Lambda|\sigma_q) \qquad [17]$$

This is the *metastable free energy* corresponding to the phase $q$. Applying the cluster expansion to the logarithm of the sum in [16], we get $|f_q(H) - e_q(H)| \leq \mathrm{e}^{-\tau/2}$. The metastable free energy corresponds to taking the ground state $\sigma_q$ and its excitations as long as they are sufficiently suppressed. Once $w_q(\gamma)$ exceeds the weight $\mathrm{e}^{-\tau|\Gamma|}$ (and the contour would have been actually preferred), we suppress it "by hand." The point is that if the phase $q$ is stable, this never happens and $w'_q(\gamma) = w_q(\gamma)$ for *all* $q$-contours $\gamma$. This is the idea behind the use of the function $f_q(H)$ as an indicator of the stability of the phase $q$ by taking

$$a_q(\boldsymbol{t}) = f_q(H_{\boldsymbol{t}}) - \min_m f_m(H_{\boldsymbol{t}}) \qquad [18]$$

Of course, the difficult point is to actually prove that the stability of phase $q$ (i.e., the fact that $a_q(\boldsymbol{t}) = 0$) indeed implies $w'_q(\gamma) = w_q(\gamma)$ for all $\gamma$. The crucial step is to prove, by induction on the diameter of $\Lambda$ and $\gamma$, the following three claims (with $\epsilon = 2\mathrm{e}^{-\tau/2}$):

1. *If $\gamma$ is a $q$-contour with $a_q(\boldsymbol{t})\,\mathrm{diam}\,\Gamma \leq \tau/4$, then* $w'_q(\gamma) = w_q(\gamma)$.
2. *If* $a_q(\boldsymbol{t})\,\mathrm{diam}\,\Lambda \leq \tau/4$, *then* $Z(\Lambda|\sigma_q) = Z'(\Lambda|\sigma_q) \neq 0$ *and*

$$\left|Z(\Lambda|\sigma_q)\right| \geq \mathrm{e}^{-f_q(H_t)|\Lambda| - \epsilon|\partial\Lambda|} \qquad [19]$$

3. *If $m \in G$, then*

$$|Z(\Lambda|\sigma_m)| \leq \mathrm{e}^{-\min_q f_q(H_t)|\Lambda|}\mathrm{e}^{\epsilon|\partial\Lambda|} \qquad [20]$$

A standard example illuminating the perturbative construction of the metastable free energies and showing the role of entropic contributions is the Blume–capel model. It is defined by the Hamiltonian

$$H_\Lambda(\sigma) = -J\sum_{\langle x,y\rangle}(\sigma_x - \sigma_y)^2 - \lambda\sum_{x\in\Lambda}\sigma_x^2 - h\sum_{x\in\Lambda}\sigma_x \qquad [21]$$

with spins $\sigma_x \in \{-1, 0, 1\}$. Taking into account only the lowest-order excitations, we get:

$$\tilde{f}_\pm(\lambda, h) = -\lambda \mp h - \frac{1}{\beta}\mathrm{e}^{-\beta(2d-\lambda\pm h)}$$

(sea of pluses or minuses with a single spin flip $\pm \to 0$) and

$$\tilde{f}_0(\lambda, h) = -\frac{1}{\beta}\mathrm{e}^{-\beta(2d+\lambda)}\left(\mathrm{e}^{\beta h} + \mathrm{e}^{-\beta h}\right)$$

(sea of zeros with a single spin flip either $0 \to +$ or $0 \to -$)
Since these functions differ from full metastable free energies $f_\pm(\lambda, h), f_0(\lambda, h)$ by terms of higher order ($\sim\mathrm{e}^{-(4d-2)\beta}$), the real phase diagram differs in this order from the one constructed by equating the functions $\tilde{f}_\pm(\lambda, h)$ and $\tilde{f}_0(\lambda, h)$. It is particularly interesting to inspect the origin, $\lambda = h = 0$. It is only the phase 0 that is stable there at all small temperatures since

$$f_0(0,0) \sim -\frac{2}{\beta}\mathrm{e}^{-\beta 2d} < f_\pm(0,0) \sim -\frac{1}{\beta}\mathrm{e}^{-\beta 2d} \qquad [22]$$

The only reason why the phase 0 is favored at this point with respect to phases + and − is that there are two excitations of order $\mathrm{e}^{-2d\beta}$ for the phase 0, while there is only *one* such excitation for + or −. The entropy of the lowest-order contribution to $f_0(0,0)$ is overweighting the entropy of the contribution to $f_\pm(0,0)$ of the same order.

## Applications

Several applications, stemming from the Pirogov–Sinai theory, are based on the fact that, due to the cluster expansion, we have quite accurate description of the model in finite volume.

One class of applications concerns various problems featuring interfaces between coexisting phases. To be able to transform the problem into a study of the random boundary line separating the two phases, one needs a precise cluster expansion formula for partition functions in volumes occupied by those phases. In the situation with no symmetry

between the phases, the use of the Pirogov–Sinai theory is indispensable.

Another interesting class of applications concerns the behavior of the system with periodic boundary conditions. It is based on the fact that the partition function $Z_{T_N}$ on a torus $T_N$ consisting of $N^d$ sites can be, again with the help of the cluster expansions, explicitly and very accurately evaluated in terms of metastable free energies,

$$\left| Z_{T_N} - \sum_{q=1}^{r} e^{-\beta f_q(H) N^d} \right| \leq \exp\{-\beta \min_m f_m(H) N^d - b\beta N\} \qquad [23]$$

with a fixed constant $b$. This formula (and its generalization to the case of complex parameters) allows us to obtain various results concerning the behavior of the model in finite volumes.

### Finite-Size Effects

Considering as an illustration a perturbation of the Ising model, so that it does not have the $\pm$ symmetry any more (and the value $h_t(\beta)$ of external field at which the phase transition between plus and minus phase occurs is not known), we can pose a natural question that has an importance for correct interpretation of simulation data. Namely, what is the asymptotic behavior of the magnetization $m_N^{\text{per}}(\beta, h) = \mu_{T_N}(1/\Lambda \sum_{x\in\Lambda} \sigma_x)$ on a torus? In the thermodynamic limit, the magnetization $m_\infty^{\text{per}}(\beta, h)$ displays, as a function of $h$, a discontinuity at $h = h_t(\beta)$. For finite $N$, we get a *rounding* of the discontinuity – the jump is smoothed. What is the shift of a naturally chosen finite-volume transition point $h_t(N)$ with respect to the limiting value $h_t$? The answer can be obtained with the help of [23] once sufficient care is taken to use the freedom in the definition of the metastable free energies $f_+(h)$ and $f_-(h)$ to replace them with a sufficiently smooth version allowing an approximation of the functions $f_\pm(h)$ around limiting point $h_t$ in terms of their Taylor expansion.

As a result, in spite of the asymmetry of the model, the finite-volume magnetization $m_N^{\text{per}}(\beta, h)$ has a universal behavior in the neighborhood of the transition point $h_t$. With suitable constants $m$ and $m_0$, we have

$$m_N^{\text{per}}(\beta, h) \sim m_0 + m \tanh\{N^d \beta m (h - h_t)\} \qquad [24]$$

Choosing the inflection point $h_{\max}(N)$ of $m_N^{\text{per}}(\beta, h)$ as a natural finite-volume indicator of the occurence of the transition, one can show that

$$h_{\max}(N) = h_t + \frac{3\chi}{2\beta^2 m^3} N^{-2d} + O(N^{-3d}) \qquad [25]$$

### Zeros of Partition Functions

The full strength of the formula [23] is revealed when studying the zeros of the partition function $Z_{T_N}(z)$ as a polynomial in a complex parameter $z$ entering the Hamiltonian of the model. To be able to use the theory in this case, one has to extend the definitions of the metastable free energies to complex values of $z$. Indeed, the construction still goes through, now yielding genuinely complex, contour models $w_\pm$ with the help of an inductive procedure. Notice that no analytic continuation is involved. An analog of [23] is still valid,

$$\left| Z_{T_N}(z) - \sum_{m=1}^{r} e^{-\beta f_m(z) N^d} \right| \leq \exp\{-\beta \min_m \Re f_m(z) N^d - b\beta N\} \qquad [26]$$

Using [26], it is not difficult to convince oneself that the loci of zeros can be traced down to the phase coexistence lines. Indeed, on the line of the coexistence of two phases $\Re f_m = \Re f_q$, the partition function $Z_{T_N}(z)$ is approximated by $e^{-\beta f N^d}(e^{-\beta \Im f_m N^d} + e^{-\beta \Im f_n N^d})$. The zeros of this approximation are thus given by the equations

$$\begin{aligned} &\Re f_m = \Re f_n < \Re f_\ell \quad \text{for all } \ell \neq m, n \\ &\beta N^d(\Im f_m - \Im f_n) = \pi \bmod 2\pi \end{aligned} \qquad [27]$$

The zeros of the full partition function $Z_{T_N}(z)$ can be proved to be exponentially close, up to a shift of order $\mathcal{O}(e^{-\beta b N})$, to those of the discussed approximation.

Briefly, the zeros of $Z_{T_N}(z)$ asymptotically concentrate on the phase coexistence curves with the density $(1/2\pi)\beta N^d |(d/dz)(f_m - f_n)|$.

## Bibliographical Remarks and Generalizations

The original works Pirogov and Sinai (1975, 1976) and Sinai (1982) introduced an analog of the weights $w'_q(\gamma)$ and parameters $a_q(H)$ as a fixed point of a suitable mapping on a Banach space. The inductive definition used here was introduced in Kotecký and Preiss (1983) and Zahradník (1984). The *completeness of phase diagram* – the fact that the stable phases exhaust the set of all periodic extremal Gibbs states was first proved in Zahradník (1984). Extension to complex parameters was first considered in Gawędzki *et al.* (1987) and Borgs and Imbrie (1989). For a review of the standard Pirogov–Sinai theory, see Sinai (1982) and Slawny (1987).

Application of Pirogov–Sinai theory for finite-size effects was studied in Borgs and Kotecký (1990) and

general theory of zeros of partition functions is presented in Biskup *et al.* (2004).

The basic statement of the Pirogov–Sinai theory yielding the construction of the full phase diagram has been extended to a large class of models. Let us mention just few of them (with rather incomplete references):

1. *Continuous spins.* The main difficulty in these models is that one has to deal with contours immersed in a sea of fluctuating spins (Dobrushin and Zahradník 1986, Borgs and Waxler 1989).
2. *Potts model.* An example of a system a transition in temperature with the coexistence of the low-temperature ordered and the high-temperature disordered phases. Contour reformulation is employing contours between ordered and disordered regions (Bricmont *et al.* 1985, Kotecký *et al.* 1990). The treatment is simplified with help of Fortuin–Kasteleyn representation (Laanait *et al.* 1991).
3. *Models with competing interactions.* ANNNI model, microemulsions. Systems with a rich phase structure (Dinaburg and Sinai 1985).
4. *Disordered systems.* An example is a proof of the existence of the phase transition for the three-dimensional random field Ising model (Bricmont and Kupiainen 1987, 1988) using a renormalization group version of the Pirogov–Sinai theory first formulated in Gawędzki *et al.* (1987).
5. *Quantum lattice models.* A class of quantum models that can be viewed as a quantum perturbation of a classical model. With the help of Feynman–Kac formula these are rewritten as a $(d+1)$-dimensional classical model that is, in its turn, treated by the standard Pirogov–Sinai theory (Datta *et al.* 1996, Borgs *et al.* 1996).
6. *Continuous systems.* Gas of particles in continuum interacting with a particular potential of Kac type. Pirogov–Sinai theory is used for a proof of the existence of the phase transitions after a suitable discretisation (Lebowitz *et al.* 1999).

*See also:* Cluster Expansion; Falicov–Kimball Model; Phase Transitions in Continuous Systems; Quantum Spin Systems.

## Further Reading

Biskup M, Borgs C, Chayes JT, and Kotecký R (2004) Partition function zeros at first-order phase transitions: Pirogov–Sinai theory. *Journal of Statistical Physics* 116: 97–155.

Borgs C and Imbrie JZ (1989) A unified approach to phase diagrams in field theory and statistical mechanics. *Communications in Mathematical Physics* 123: 305–328.

Borgs C and Kotecký R (1990) A rigorous theory of finite-size scaling at first-order phase transitions. *Journal of Statistical Physics* 61: 79–119.

Borgs C, Kotecký R, and Ueltschi D (1996) Low temperature phase diagrams for quantum perturbations of classical spin systems. *Communications in Mathematical Physics* 181: 409–446.

Borgs C and Waxler R (1989) First order phase transitions in unbounded spin systems: construction of the phase diagram. *Communications in Mathematical Physics* 126: 291–324.

Bricmont J, Kuroda T, and Lebowitz J (1985) First order phase transitions in lattice and continuum systems: extension of Pirogov–Sinai theory. *Communications in Mathematical Physics* 101: 501–538.

Bricmont J and Kupiainen A (1987) Lower critical dimensions for the random field Ising model. *Physical Review Letters* 59: 1829–1832.

Bricmont J and Kupiainen A (1988) Phase transition in the 3D random field Ising model. *Communications in Mathematical Physics* 116: 539–572.

Datta N, Fernández R, and Fröhlich J (1996) Low-temperature phase diagrams of quantum lattice systems. I. Stability for quantum perturbations of classical systems with finitely-many ground states. *Helv. Phys. Acta* 69: 752–820.

Dinaburg EL and Sinaï YaG (1985) An analysis of ANNNI model by Peierls contour method. *Communications in Mathematical Physics* 98: 119–144.

Dobrushin RL and Zahradník M (1986) Phase diagrams of continuous lattice systems. In: Dobrushin RL (ed.) *Math. Problems of Stat. Physics and Dynamics*, pp. 1–123. Dordrecht: Reidel.

Gawędzki K, Kotecký R, and Kupiainen A (1987) Coarse-graining approach to first-order phase transitions. *Journal of Statistical Physics* 47: 701–724.

Kotecký R, Laanait L, Messager A, and Ruiz J (1990) The *q*-state Potts model in the standard Pirogov–Sinai theory: surface tensions and Wilson loops. *Journal of Statistical Physics* 58: 199–248.

Kotecký R and Preiss D (1983) An inductive approach to PS theory, *Proc. Winter School on Abstract Analysis, Suppl. ai Rend. del Mat. di Palermo*.

Lebowitz JL, Mazel A, and Presutti E (1999) Liquid–vapor phase transitions for systems with finite range interactions. *Journal of Statistical Physics* 94: 955–1025.

Laanait L, Messager A, Miracle-Solé S, Ruiz J, and Shlosman SB (1991) Interfaces in the Potts model I: Pirogov–Sinai theory of the Fortuin–Kasteleyn representation. *Communications in Mathematical Physics* 140: 81–91.

Pirogov SA and Sinai YaG (1975) Phase diagrams of classical lattice systems (Russian). *Theoretical and Mathematical Physics* 25(3): 358–369.

Pirogov SA and Sinai YaG (1976) Phase diagrams of classical lattice systems. Continuation (Russian). *Theoretical and Mathematical Physics* 26(1): 61–76.

Sinai YaG (1982) *Theory of Phase Transitions: Rigorous Results.* New York: Pergamon.

Slawny J (1987) Low temperature properties of classical lattice systems: phase transitions and phase diagrams. In: Domb C and Lebowitz JL (eds.) *Phase Transitions and Critical Phenomena*, vol. 11, pp. 127–205. New York: Academic Press.

Zahradník M (1984) An alternate version of Pirogov–Sinai theory. *Communications in Mathematical Physics* 93: 559–581.

# Point-Vortex Dynamics

**S Boatto**, IMPA, Rio de Janeiro, Brazil
**D Crowdy**, Imperial College, London, UK



## Introduction

Vortices have a long fascinating history. Descartes wrote in his *Le Monde*:

> ...que tous les mouvements qui se font au Monde sont en quelque façon circulaire: c'est à dire que, quand un corps quitte sa place, il entre toujours en celle d'un autre, et celui-ci en celle d'un autre, et ainsi de suite jusques au dernier, qui occupe au même instant le lieu délaissé par le premier.

In particular, Descartes thought of vortices to model the dynamics of the solar system, as reported by W W R Ball (1940):

> Descartes' physical theory of the universe, embodying most of the results contained in his earlier and unpublished *Le Monde*, is given in his *Principia*, 1644,... He assumes that the matter of the universe must be in motion, and that the motion must result in a number of vortices. He stated that the sun is the center of an immense whirlpool of this matter, in which the planets float and are swept round like straws in a whirlpool of water.

Descartes' theory was later on recused by Newton in his *Principia* in 1687. Few centuries later, W Thomson (1867) the later Lord Kelvin, made use of vortices to formulate his atomic theory: each atom was assumed to be made up of vortices in a sort of ideal fluid. In 1878–79 the American physicist A M Mayer conducted a few experiments with needle magnets placed on floating pieces of cork in an applied magnetic field, as toy models for studying atomic interactions and forms (Mayer 1878, Aref *et al.* 2003). In 1883 inspired by Mayer experiments, J J Thomson combined W Thomson's atomic theory with H von Helmholtz's point-vortex theory (Helmholtz 1858): he thought as the electrons were point vortices inside a positively charged shell (see Figure 1), the vortices being located at the vertices of regular parallelograms and investigated about the stability of such structures (see Thomson (1883, section 2.1)). The vortex-atomic theory survived for quite a few years up to Rutherford's experiments proved that atoms have quite a different structure! Before continuing this historical/modeling overview, let's address the following question:

> *what is a vortex and, more specifically, what is a point-vortex?*

Roughly speaking, following Descartes, a vortex is an entity which makes particles move along circular-like orbits. Examples are the cyclones and anticyclones in the atmosphere (see Figure 3). Mathematically speaking, let $\boldsymbol{u} = (u, v, w) \in \mathbb{R}^3$ be a velocity field, the associated vorticity field $\omega$ is defined to be

$$\omega = \nabla \wedge \boldsymbol{u} \qquad [1]$$

In this article we are considering exclusively inviscid flows which are also incompressible, that is,

$$\nabla \cdot \boldsymbol{u} = 0 \qquad [2]$$

and have constant density $\rho$, which we normalize to be equal to 1 ($\rho = 1$). In two dimensions, a point-vortex field is the simplest of all vorticity fields: it can be thought as an entity where the vorticity field is concentrated into a point. In other words, point vortices are singularities of the vorticity field! Then, in the plane the vorticity field associated to a system of $N$ point vortices is

$$\omega(\boldsymbol{r}) = \sum_{\alpha=1}^{N} \Gamma_\alpha \delta(\boldsymbol{r} - \boldsymbol{r}_\alpha) \qquad [3]$$

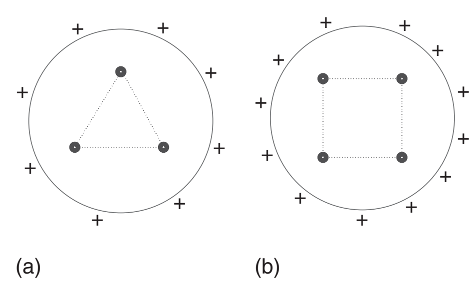



**Figure 1** Thomson atomic model: (a) atom with three electrons and (b) atom with four electrons. From Thomson JJ (1883) *A Treatise on the Motion of Vortex Rings*. New York: Macmillan and Thomson JJ (1904) *Electricity and Matter*. Westmister: Archibald Constable.

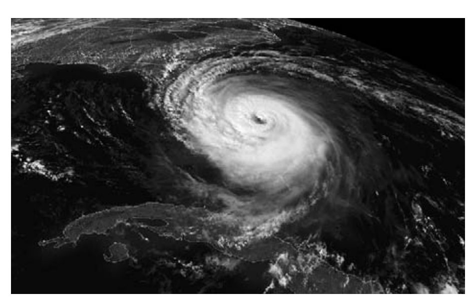


**Figure 2** Hurricane Jeanne. Reproduced with permission from the National Oceanic and Atmospheric Administration (NOAA) (www.noaanews.noaa.gov).

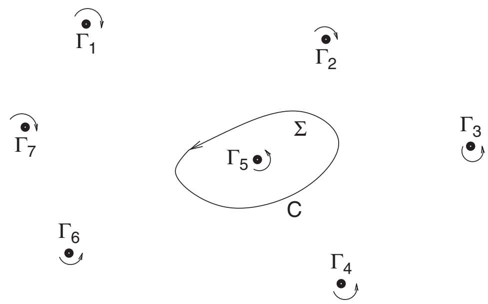



**Figure 3** Cyclones and anticyclones in the atmosphere. Reproduced from Boatto S and Cabrel HE, *SIAM Journal of Applied Mathematics* 64:216–230 (2003). With the permission of SIAM.

where $\Gamma_\alpha$, $\alpha = 1, \ldots, N$, is a constant and corresponds to the vorticity (or circulation) of the $\alpha$-vortex, situated at $\boldsymbol{r}_\alpha$. In fact by definition, the circulation around a curve $C$ delimiting a region $\Sigma$ with boundary $C$,

$$\Gamma_C = \oint_C \boldsymbol{u} \cdot \mathrm{d}\boldsymbol{s} = \iint_\Sigma (\nabla \wedge \boldsymbol{u}) \cdot \boldsymbol{n}\, \mathrm{d}A = \iint_\Sigma \omega \qquad [4]$$

where we have used Stokes' theorem to bring in the vorticity. Then if the region contains only the $\alpha$th point vortex, we obtain

$$\Gamma_C = \iint_\Sigma \omega \cdot \mathrm{d}\boldsymbol{A} = \Gamma_\alpha \qquad [5]$$

by eqn [3]. A positive (resp. negative) sign of $\Gamma_\alpha$ indicates that the corresponding point vortex induces an anticlockwise (resp. clockwise) particle motion, see **Figure 4a**)). Is there an analog of a point-vortex system for a three-dimensional flow?

Yes, and this brings in the analogy between vortex lines and magnetic field lines that Mayer used in his experiments with floating magnets. In fact, in three dimensions, the notion of a point vortex can be extended to that one of a straight vortex line (see **Figure 4b**), where, by definition, a vortex line is a curve that is tangent to the vorticity vector $\omega$ at each of its point. In this context we would like to mention the beautiful experiments of Yarmchuck–Gordon–Packard on vortices in superfluid helium. They observed the formation of stable polygonal configurations of identical vortices, quite similar to the ones observed by Mayer with his magnets (see **Figures 5 and 1**).

One would like to understand how such configurations form and to give a theoretical account about their stability. In order to answer these questions we have to first be able to describe the dynamics of a system of point vortices from a mathematical point of view.

## Evolution Equations

Can point vortices be viewed as "discrete" (or localized) solutions of Euler equation in two dimensions? Let us consider the Euler equation

$$\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u} \cdot \nabla \boldsymbol{u} = -\nabla p + \boldsymbol{f} \qquad [6]$$

where $p$ is the pressure, $\boldsymbol{f} = -\nabla U$ is a conservative force, and restrict our attention to the two-dimensional setting, for example, vortex dynamics on the plane (or a sphere). Then it is immediate that by taking the curl of eqn [6] we obtain the evolution equation of the vorticity, that is,

$$\frac{\partial \omega}{\partial t} + \boldsymbol{u} \cdot \nabla \omega = 0, \quad \text{or} \quad \frac{D\omega}{Dt} = 0 \qquad [7]$$

where the operator $D/Dt = \partial/\partial t + \boldsymbol{u} \cdot \nabla$ is called the material derivative and describes the evolution along the flow lines. It follows from eqn [7] that in two dimensions the vorticity is conserved as it is transported along the flow lines. Then a natural question arises: supposing the vorticity field $\omega$ is known, is it possible to deduce the velocity field $\boldsymbol{u}$ generating $\omega$? Or in other words, is it possible to solve the system of eqns [1]–[2]? It is immediate to see that in general the solution is not unique, if some boundary conditions are not specified (see Marchioro and Pulvirenti (1993)). Furthermore, as already observed by Kirchhoff in 1876 (Boatto and Cabral 2003), in two dimensions we can recast the fluid equations [1]–[2] into a Hamiltonian formalism. In fact, notice that on the plane $\boldsymbol{u} = (\dot{x}, \dot{y})$ and eqn [2] is still satisfied if we represent the velocity components as

**Figure 4** (a) Advected by the velocity field of one point vortex, a test particle follows a circular orbit, with a speed proportional to the absolute value of the vortex circulation and inversely proportional to the square of its distance from the vortex. (b) Straight vortex lines.

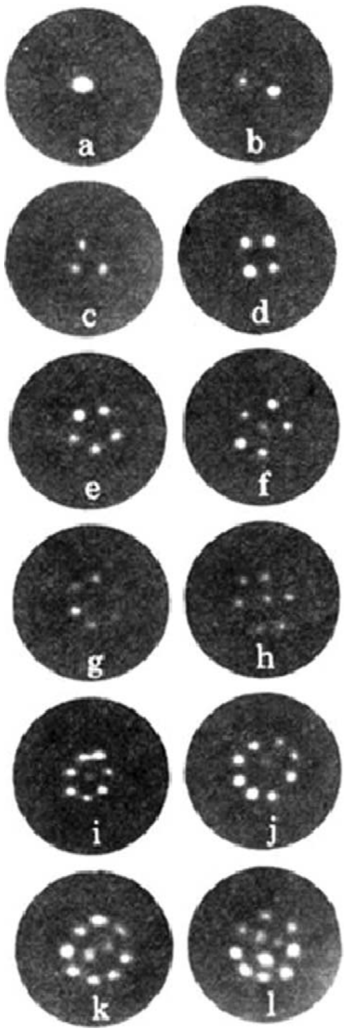



**Figure 5** Photographs of vortex configurations in a rotated sample of superfluid helium with 1,...,11 vortices. Reprinted figure with permission from Yarmchuk EJ, Gordon MJV, and Packard RE (1979) Observation of stationary vortices arrays in rotating superfluid Helium. *Physical Review Letters* 43(3): 214–217. Copyright (1979) by the American Physical Society.

$$\dot{x} = \frac{\partial \Psi}{\partial y}, \qquad \dot{y} = -\frac{\partial \Psi}{\partial x} \qquad [8]$$

that is, by means of $\Psi$, called the stream function. Formally, $\Psi$ plays the rôle of a Hamiltonian for the pair of conjugate variables $(x, y)$ and it is used to describe the dynamics of a test particle, located at $(x, y)$ and advected by the flow. By substituting [8] into [1], we obtain

$$\Delta\Psi(\boldsymbol{r}) = \omega(\boldsymbol{r}) \qquad [9]$$

that is, a Poisson equation with $\omega$ as a source term. Then, once we specify the vorticity field, by inverting [9] we obtain the stream function $\Psi$ to be

$$\Psi(\boldsymbol{r}) = \int G(\boldsymbol{r}, \boldsymbol{r}')\omega(\boldsymbol{r}')\,\mathrm{d}\boldsymbol{r}' \qquad [10]$$

where $G(\boldsymbol{r}, \boldsymbol{r}')$ is the Green's function, solution of the equation $\Delta G(x, y) = -\delta(x, y)$. The Green's function both for the plane and the sphere is (Marchioro and Pulvirenti 1993)

$$G(\boldsymbol{r}, \boldsymbol{r}') = -\frac{1}{4\pi}\log\|\boldsymbol{r} - \boldsymbol{r}'\|^2 \qquad [11]$$

where $\|\boldsymbol{r} - \boldsymbol{r}'\|^2 = (x - x')^2 + (y - y')^2$. By [10], once we specify the vorticity field $\omega(\boldsymbol{r})$ we can compute $\Psi$, and by replacing it into [8] the velocity field becomes

$$\boldsymbol{u}(\boldsymbol{r}) = \int K(\boldsymbol{r}, \boldsymbol{r}')\omega(\boldsymbol{r}')\,\mathrm{d}\boldsymbol{r}' \qquad [12]$$

where $K(\boldsymbol{r}, \boldsymbol{r}') = -(\boldsymbol{r} - \boldsymbol{r}')^{\perp}/[2\pi\|\boldsymbol{r} - \boldsymbol{r}'\|^2]$ and it represents the velocity field generated by a point vortex of intensity one, located at $\boldsymbol{r}'$. Then by considering the vorticity field generated by point vortices, eqn [3], together with eqn [11], eqn [10] becomes

$$\Psi(\boldsymbol{r}) = -\frac{1}{4\pi}\int \log\|\boldsymbol{r} - \boldsymbol{r}'\|^2 \left(\sum_{\alpha=1}^{N}\Gamma_\alpha \delta(\boldsymbol{r}' - \boldsymbol{r}_\alpha)\right)\mathrm{d}\boldsymbol{r}'$$
$$= -\frac{1}{4\pi}\sum_{\alpha=1}^{N}\Gamma_\alpha \log\|\boldsymbol{r} - \boldsymbol{r}_\alpha\|^2 \qquad [13]$$

Equation [13] describes together with [8], the dynamics of a test particle at a point $\boldsymbol{r} = (x, y)$ in the plane. Analogously, it can be shown that the dynamics of a systems of point vortices in the plane is given by the equations

$$\Gamma_\alpha \frac{\mathrm{d}x_\alpha}{\mathrm{d}t} = \frac{\partial H_v}{\partial y_\alpha}, \qquad \Gamma_\alpha \frac{\mathrm{d}y_\alpha}{\mathrm{d}t} = -\frac{\partial H_v}{\partial x_\alpha} \qquad [14]$$

where $(q_\alpha, p_\alpha) = (x_\alpha, \Gamma_\alpha y_\alpha)$, $\alpha = 1, \ldots, N$, is a pair of conjugate variables and $H_v$ is the generalization of the stream function $\Psi$ (eqn [13]):

$$H_v = -\frac{1}{4\pi}\sum_{\substack{\alpha,\beta=1\\ \alpha\neq\beta}}^{N}\Gamma_\alpha\Gamma_\beta \log\|\boldsymbol{r}_\alpha - \boldsymbol{r}_\beta\|^2 \qquad [15]$$

Notice that the vortex Hamiltonian $H_v$ (eqn [15]) is an autonomous Hamiltonian and, as we will discuss in the first subsection, it provides a good Lyapunov-like function to study stability properties of some vortex configurations. Moreover, $H_v$ is invariant with respect to rotations and translations, then by the Noether theorem there are other first integrals of motion, that is,

$$L = \sum_{k=1}^{N}\Gamma_k \parallel \boldsymbol{x}_k \parallel^2, \qquad M_x = \sum_{k=1}^{N}\Gamma_k x_k,$$
$$M_y = \sum_{k=1}^{N}\Gamma_k y_k$$

expressing, respectively, the conservation of angular momentum, $L$, and linear momentum, $\boldsymbol{M} = (M_x, M_y)$, on the plane. We shall denote with $M$ the magnitude of $\boldsymbol{M}$ (i.e., $M = \|\boldsymbol{M}\|$). Furthermore, by introducing the Poisson bracket

$$[f, g] = \sum_{\alpha=1}^{N} \left( \frac{\partial f}{\partial q_\alpha} \frac{\partial g}{\partial p_\alpha} - \frac{\partial f}{\partial p_\alpha} \frac{\partial g}{\partial q_\alpha} \right) = \sum_{\alpha=1}^{N} \frac{1}{\Gamma_\alpha} \left( \frac{\partial f}{\partial x_\alpha} \frac{\partial g}{\partial y_\alpha} - \frac{\partial f}{\partial y_\alpha} \frac{\partial g}{\partial x_\alpha} \right)$$

we can construct three integrals in involution out of the four conserved quantities $L$, $M_x, M_y$, and $H_v$. These are $L, M_x^2 + M_y^2$ and $H_v$: in fact,

$$[H_v, L] = 0, \qquad \left[H_v, M_x^2 + M_y^2\right] = 0, \qquad \left[L, M_x^2 + M_y^2\right] = 0$$

It is then possible to reduce the system of equations from $N$ to $N - 2$ degrees of freedom. A Hamiltonian system with $N$ degrees of freedom is integrable whenever there are $N$ independent integrals of motion in involution. It follows that a vortex system with $N \leq 3$ is integrable, whereas the system of equations of four identical vortices has been shown by Ziglin to be nonintegrable in the sense that there are no other first integrals analytically depending on the coordinates and circulations, and functionally independent of $L, H_v, M_x, M_y$ (see Ziglin (1982)). The following, however, has been shown:

1. Let $K = \sum_{\alpha=1}^{N} k_\alpha$ be the total vorticity, $\boldsymbol{M} = (M_x, M_y)$ the total momentum and $M = \|\boldsymbol{M}\|$. Then, as shown by Aref and Stremler (1999), if $K = 0$ and $M = 0$, $N$-vortex problem [16] is integrable.
2. A system of four identical vortices (i.e., $k_\alpha = k$ for $\alpha = 1, \ldots, 4$) can undergo periodic or quasiperiodic motion for special initial conditions (see Khanin (1981) *Russian Math. Surveys* 36: 231; Aref and Pomphrey (1982) *Proc. R. Soc. Lond. A* 380: 359–387). More specifically, the motion of a system of four identical vortices can be periodic, quasiperiodic, or chaotic depending on the symmetry of the initial configuration. In fact, every vortex configuration that belongs to the subspace of symmetric configurations – $x_\alpha = -x_{\alpha+2}$ and $y_\alpha = y_{\alpha+2}, \alpha = 1, 2$ – gives rise to an integrable vortex motion.

We have that up to two vortices, the motion is almost always periodic and the orbits are circles; the only exception being the case for which $k_2 = -k_1$, when the circles degenerate into straight lines. Thus, a configuration of two point vortices is always a relative equilibrium configuration, that is, there exists a specific reference frame in which the two vortices are at rest. If the vortices are identical ($\Gamma_1 = \Gamma_2 = \Gamma$), the motion is synchronous with frequency $\Omega = \Gamma/\pi$ and the vortices share the same circular orbit (see Figure 6a). If the vortices are not identical and have vorticities of different magnitudes (say $|\Gamma_1| > |\Gamma_2|$), their motion is still synchronous and periodic, with frequency $\Omega = (\Gamma_1 + \Gamma_2)/(2\pi)$, and the vortices move on different circular orbits (with $r_2 < r_1$) both centered at the center of vorticity. Note that for both cases, identical and nonidentical vortices, we can view the vortex dynamics in a co-rotating frame where the vortices are simply at rest.

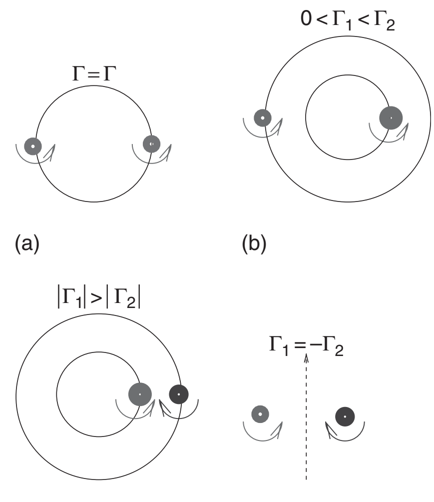



**Figure 6(a–d)** For $N = 2$ the vortex dipole exhibits a synchronous and the orbits are in general circular orbits, with the exception of the case (d) for which $\Gamma_1 = -\Gamma_2$ and the circular orbit degenerates into a line (or a circle of infinite radius).

For three vortices we can have periodic and quasiperiodic motion, depending on the initial conditions, and for four vortices we can have periodic, quasiperiodic, or weakly chaotic motion.

### Remarks

(i) The nonintegrability of the 4-vortex system was also proved for configurations of nonidentical vortices. Koiller and Carvalho (1989) gave an analytical proof for $\Gamma_1 = -\Gamma_2$ and $\Gamma_3 = \Gamma_4 = \epsilon, 0 \ll \epsilon \ll 1$. Moreover, Castilla *et al.* (1993) considered the case: $\Gamma_1 = \Gamma_2 = \Gamma_3 = 1$ and $\Gamma_4 = \epsilon$.

(ii) Due to the translational and rotational symmetries of $H_v$, there are some analogies between the $N$-vortex problem and the $N$-body problem, especially for what concerns configurations of relative equilibria (see Albouy (1996) and Glass (2000)). A relative equilibrium is a vortex (or mass) configuration that moves without change of shape or form, that is, a configuration which is steadily

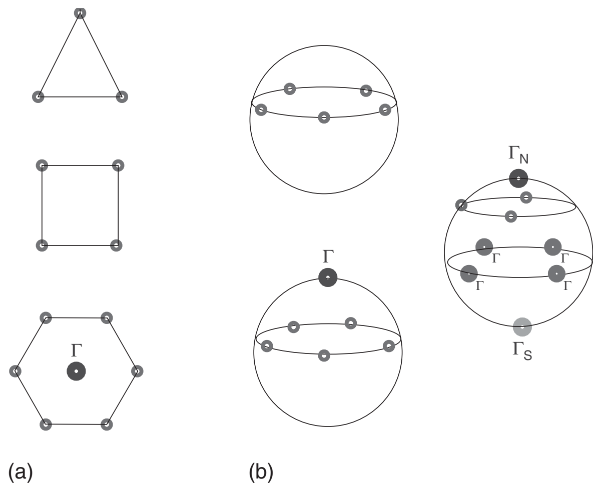



**Figure 7** Polygonal configuration of vortices: (a) planar configurations and (b) configurations of vortex rings on a sphere, with and without polar vortices.

rotating or translating. A few examples are vortex polygons (see **Figure 7**) like the ones studied by Thomson, Mayer, Yarmchuk–Gordon–Packard, Boatto–Cabral (2003), Cabral–Schmidt (1999/2000), Dritschel–Polvani (1993), Lim–Montaldi–Roberts (2001), Sakajo (2004). For an exhaustive review on relative equilibria of vortices, see the article by Aref *et al.* (2003). We shall discuss stability of polygonal vortex configuration in the following subsection.

(iii) As shown by Kimura (1999) in a beautiful geometrical formalism, on the unit sphere ($S^2$) and on the Hyperbolic plane ($H^2$), the vortex Hamiltonians [15] are

$$H_v = -\frac{1}{4\pi}\sum_{\alpha\neq\beta}^{N}\Gamma_\alpha\Gamma_\beta\log(1-\cos\rho_{\alpha\beta}) \quad \text{on } S^2$$

$$H_v = -\frac{1}{4\pi}\sum_{\alpha\neq\beta}^{N}\Gamma_\alpha\Gamma_\beta\log\frac{\cosh\rho_{\alpha\beta}-1}{\cosh\rho_{\alpha\beta}+1} \quad \text{on } H^2$$

where

$$\cos\rho_{\alpha\beta} = \cos\theta_\alpha\cos\theta_\beta + \sin\theta_\alpha\sin\theta_\beta\cos(\phi_\alpha-\phi_\beta) \quad \text{on } S^2$$

$$\cosh\rho_{\alpha\beta} = \cosh\theta_\alpha\cosh\theta_\beta + \sinh\theta_\alpha\sinh\theta_\beta\cos(\phi_\alpha-\phi_\beta) \quad \text{on } H^2$$

On $S^2, \theta_\alpha$ and $\phi_\alpha$ are, respectively, the co-latitude and the longitude of the $\alpha$-vortex, $\alpha = 1, \ldots, N$. We can define canonical variables $q_\alpha$ and $p_\alpha$ on $S^2$ and $H^2$, respectively, as

$$q_\alpha = \Gamma_\alpha\cos\theta_\alpha, \qquad p_\alpha = \phi_\alpha \quad \text{on } S^2$$
$$q_\alpha = \Gamma_\alpha\cosh\theta_\alpha, \qquad p_\alpha = \phi_\alpha \quad \text{on } H^2$$

Montaldi *et al.* (2002) studied vortex dynamics on a cylindrical surface, and Soulière and Tokieda (2002) considered vortex dynamics on surfaces with symmetries.

(iv) As we shall see in the section on point vortex motion, it is sometimes useful to employ the complex analysis formalism. Then the variables of interest are $z_\alpha = x_\alpha + \mathrm{i}y_\alpha, \alpha = 1, \ldots, N$, and its conjugate $\bar{z}_\alpha$, the Hamiltonian [15] takes the form

$$H_v = -\frac{1}{2\pi}\sum_{\alpha\neq\beta}\Gamma_\alpha\Gamma_\beta\log|z_\alpha - z_\beta|$$

and the equations of motions become

$$\dot{z}_\alpha = \frac{\mathrm{i}}{2\pi}\sum_{\beta\neq\alpha,\beta=1}^{N}\Gamma_\beta\frac{z_\alpha - z_\beta}{|z_\alpha - z_\beta|^2}, \quad \alpha = 1, \ldots, N \quad [16]$$

(v) Equation [14] can we rewritten in a more compact form as

$$\frac{\mathrm{d}X}{\mathrm{d}t} = J\nabla_X H_v \quad [17]$$

where

$$X = (q_1, \ldots, q_N, p_1, \ldots, p_N)$$
$$\nabla_X = \left(\frac{\partial}{\partial q_1}, \ldots, \frac{\partial}{\partial q_N}, \frac{\partial}{\partial p_1}, \ldots, \frac{\partial}{\partial p_N}\right)$$
$$J = \begin{pmatrix} O & \mathbb{I} \\ -\mathbb{I} & O \end{pmatrix}$$

$\mathbb{I}$ being the $N \times N$ identity matrix.

(vi) How close is the point-vortex model to the original Euler equation? Point-vortex systems represent discrete solutions of the Euler equation in a "weak" sense – see both the book and the article by Marchioro and Pulvirenti (1993, 1994). These authors proved that the Euler dynamics is "similar" to the vortex dynamics in which the vortices are localized in very small regions, and the vortex intensities are the total vorticities associated to such small regions. In particular, let us consider a vorticity field with compact support on a family of $\epsilon$-balls, that is,

$$\omega^\epsilon = \sum_{i=1}^{N}\omega_i^\epsilon$$

with support of $\omega_i^\epsilon$ contained in the ball of center $x_i$ (independent of $\epsilon$) and radius $\epsilon$. Furthermore let us assume that

$$\int_{|\boldsymbol{r}-\boldsymbol{r}_i|\leq\epsilon}\omega_i^\epsilon\,\mathrm{d}\boldsymbol{r} = \Gamma_i$$

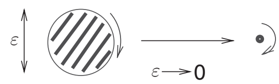



**Figure 8** In the limit $\epsilon \to 0$, the dynamics of the center of vorticity of a vortex $\epsilon$-ball is approximated by the dynamics of a point vortex.

with the $\gamma_i$ independent of $\epsilon$. Then in the limit $\epsilon \to 0$ the dynamics of the center of vorticity $B_\epsilon(t) = \int r\omega_\epsilon(r, t)\,\mathrm{d}r$, of a given $\epsilon$-ball, "converges" to the motion of a single point vortex (see Figure 8). This result is important to illustrate as vortex systems provide both a useful heuristic tool in the analysis of the general properties of the solutions of Euler's equations (Poupaud 2002, Schochet 1995), and a useful starting point for the construction of practical algorithms for solving equations in specific situations. In particular, it provides a theoretical justification to the vortex method previously introduced by Carnevale *et al.* (1992). These authors constructed a numerical algorithm to study turbulence decaying in two dimensions. Their vortex method greatly simplifies fluid simulations as basically it relies on a discretization of the fluid into circular patches. The dynamics of patches is given by the centers of vorticity, which interact as a point-vortex system, endowed with a rule dictating how patches merge (see Figure 9).

### Stability of a Vortex Ring

As mentioned in the Introduction section, the study of vortex relative equilibria has a long history. Kelvin showed that steadily rotating patterns of identical vortices arise as solutions of a variational problem in which the interaction energy (vortex Hamiltonian) is minimized subject to the constraint that the angular impulse be maintained (see Aref (2003). In 1883, while studying and modeling the atomic structure, J J Thomson investigated the linear stability of co-rotating point vortices in the plane. In particular, his interest was in configurations of identical vortices equally spaced along the circumference of a circle, that is, located at the vertices of a regular polygon (see Figure 7). He proved that for six or fewer vortices the polygonal configurations are stable, while for seven vortices – the Thomson heptagon – he erroneously concluded that the configuration is slightly unstable. It took more than a century to make some progresses on this problem. D G Dritschel (1985) succeeded in solving the heptagon mystery for what concerns its linear stability analysis, leaving open the nonlinear stability question: he proved that the Thomson heptagon is neutrally stable and that for eight or more vortices the corresponding polygonal configurations are linearly unstable. Later on in 1993, Polvani and Dritschel (1993) generalized the techniques used in Dritschel (1985) to study the linear stability of a "latitudinal" ring of point vortices on the sphere, as a function of the number $N$ of vortices in the ring, and of the ring's co-latitude $\theta$ (see Figure 10). They proved that polygonal configurations are more unstable on the sphere than in the plane. In particular, they showed that at the pole, for $N < 7$ the configuration is stable, for $N = 7$ it is neutrally stable and for $N > 7$ it is unstable. By means of the energy momentum method (Marsden–Meyer–Weistein reduction), J E Marsden and S Pekarsky (1998) studied the nonlinear stability analysis for the integrable case of polygonal configurations of three vortices of arbitrary vorticities ($\Gamma_1, \Gamma_2$ and $\Gamma_3$) on the sphere, leaving open the stability analysis for nonintegrable vortex systems ($N > 3$). In 1999 H E Cabral and D S Schmidt completed the linear and nonlinear stability analysis at once for polygonal configurations in the plane. In 2003 Boatto and Cabral studied the nonlinear stability of a ring of vortices on the sphere, as a function of the number of vortices $N$ and the ring colatitude $\theta$.

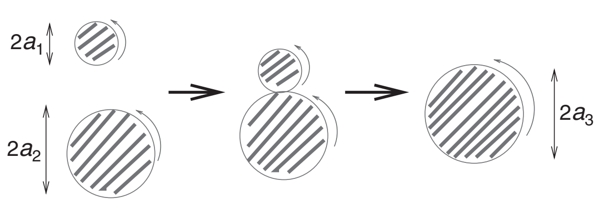



**Figure 9** In Carnevale *et al.* (1992) the fluid is modeled by a dilute vortex gas with density $\rho$ and typical radius $a$. The dynamics is governed by the point-vortex dynamics of the disk centers, each disk corresponding to a point vortex of intensity $\Gamma = \pi\xi_{ext}a^2$, where $\xi_{ext}$ plays the role of a vorticity density. Two vortices or radius $a_1$ and $a_2$ merge when their center-to-center distance is less or equal to the sum of their radii, $a_1 + a_1$. Then a new vortex is created and its radius $a_3$ is given by $a_3 = (a_1^4 + a_2^4)^{1/4}$.

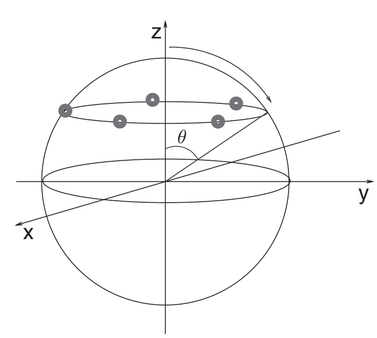



**Figure 10** Latitudinal ring of vortices. Reproduced with permission from Boatto S and Cabral HE *SIAM Journal of Applied Mathematics* 64: 216–230 (2003).

Boatto and Simó (2004) generalized the stability analysis to the case of a ring with polar vortices and of multiple rings, the key idea being, as we shall discuss in this section, the structure of the Hessian of the Hamiltonian.

*How to infer about linear and nonlinear stability of steadily rotating configurations?*

Let us restrict the discussion to a polygonal ring of identical vortices on a sphere as illustrated in **Figure 7** (Boatto and Cabral 2003, Boatto and Simó 2004). The reasoning is easily generalized for the planar case. The case of multiple rings is discussed in great detail in Boatto and Simó (2004). A polygonal ring is a relative equilibrium of coordinates $X(t) = (q_1(t), \ldots, q_N(t), p_1(t), \ldots, p_N(t))$, where

$$\begin{aligned} q_\alpha(t) &= \phi_\alpha(t) = \omega t + \phi_{o\alpha} \\ p_\alpha(t) &= p_o = \Gamma \cos\theta_o \quad \alpha = 1, \ldots, N \end{aligned} \qquad [18]$$

$\omega = (N-1)p_o/r_o^2$, $r_o = \sqrt{1 - p_o^2/\Gamma^2}$, $\phi_{o\alpha}$ and $\theta_{o\alpha} = \theta_o$ being the initial longitude and co-latitude of the $\alpha$th vortex.

**Theorem 1** (Spherical case) (Boatto and Simó 2004). *The relative equilibrium* [18] *is* (*linearly and nonlinearly*) *stable if*

$$\begin{aligned} &-4(N-1)(11-N) + 24(N-1)r_o^2 \\ &\quad + 2N^2 + 1 + 3(-1)^N < 0 \end{aligned} \qquad [19]$$

*and it is unstable if the inequality is reversed.*

**Remarks**

(i) By Theorem 1 a vortex polygon, of $N$ point vortices, is stable for $0° \le \theta_o \le \theta_o^*$ and $(180° - \theta_o^*) \le \theta_o \le 180°$, where $\theta_o^* = \arcsin(r_o^*)$ and

$$r_o^{*2} < \frac{7-N}{4} \qquad \text{for } N \text{ odd}$$

$$r_o^{*2} < -\frac{N^2 - 8N + 8}{4(N-1)} \qquad \text{for } N \text{ even}$$

where $r_o^* = \sin\theta_o^*$.

(ii) Theorem 1 includes at once the results of Thomson (1883), Dritschel (1985), and Polvani and Dritschel (1993) (and other authors who have been working in the area (Aref *et al.* 2003)). We recover the planar case by setting $r_o = 0$ in eqn [19], deducing that stability is guaranteed for $N < 7$.

To prove Theorem 1 it is useful to consider the Hamiltonian equations as in eqn [17]. The first step is to make a change of reference frame: view the dynamics in a frame co-rotating with the relative equilibrium configuration. In the co-rotating reference system, the Hamiltonian takes the form

$$\tilde{H} = H + \omega M$$

where $M$ is the momentum of the system, and $H$ and $\omega$ are, respectively, the Hamiltonian and the rotational frequency of the relative equilibrium in the original frame of reference. In the new reference frame, the relative equilibrium becomes an equilibrium, $X^*$, and the standard techniques can be used to study its stability.

To study linear stability, the relevant equation is

$$\frac{d\Delta X}{dt} = JS\Delta X \qquad [20]$$

where $X = X^* + \Delta X$, and $S$ is the Hessian of $\tilde{H}$ evaluated at the equilibrium $X^*$. Then linear (or spectral) stability is deduced by studying the eigenvalues of the matrix $JS$ (spectral stability). For nonlinear stability we make use of a sufficient stability criterion due to Dirichlet (1897) (see G Lejeune Dirichlet (1897). *Werke*, vol. 2, Berlin, pp. 5–8; Boatto and Cabral (2003) and references therein).

**Theorem 2** *Let $X^*$ be an equilibrium of an autonomous system of ordinary differential equations*

$$\frac{dX}{dt} = f(X), \qquad \Omega \subset \mathbb{R}^{2N} \qquad [21]$$

*that is, $f(X^*) = 0$. If there exists a positive* (*or negative*) *definite integral* **F** *of the system* [21] *in a neighborhood of the equilibrium $X^*$, then $X^*$ is stable.*

In our case the Hamiltonian itself is an integral of motion. Then by studying definiteness of its Hessian, $S$, evaluated at $X^*$, we infer minimal stability intervals in $\theta$ and $N$. Details are given in Boatto and Cabral (2003) and Boatto and Simó (2004). The proof is mainly based on the following considerations:

1. Since $S$ is a symmetric matrix it is diagonalizable, that is, there exists an orthogonal matrix $C$ such that $C^T S C = D$, where $D$ is a diagonal matrix, $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$. Furthermore, the matrix $C$ can be chosen to leave invariant the symplectic form (equivalently $J = C^T J C$). Then by the canonical change of variables $Y = C^T X$ eqn [20] becomes

$$\frac{d\Delta Y}{dt} = JD\Delta Y \qquad [22]$$

where $Y = (\tilde{q}_1, \ldots, \tilde{q}_N, \tilde{p}_1, \ldots, \tilde{p}_N)$ and $(\tilde{q}_j, \tilde{p}_j)$, $j = 1, \ldots, N$, are pairs of conjugate variables. Equation [22] can be rewritten as

$$\frac{d^2 \Delta\tilde{q}_j}{dt^2} = -\lambda_j \lambda_{j+N} \Delta\tilde{q}_j, \quad j = 1, \ldots, N$$

2. When evaluated at the equilibrium $X^*$, the Hessian $S$ takes the block structure

$$\tilde{S} = \left( \begin{array}{c|c} Q & \mathbf{O} \\ \hline \mathbf{O} & P \end{array} \right)$$

where the matrices $Q$ and $P$ are symmetric circulant matrices, that is, $(N \times N)$ matrices of the form

$$A = \begin{pmatrix} a_1 & a_2 & \ldots & a_N \\ a_N & a_1 & \ldots & a_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ a_2 & a_3 & \ldots & a_1 \end{pmatrix} \qquad [23]$$

Circulant matrices are of special interest to us because we can easily compute their eigenvalues and eigenvectors for all $N$. In fact, it is immediate to show that:

**Lemma 3** *All circulant matrices [23] have eigenvalues*

$$\lambda_j = \sum_{k=1}^{N} a_k r_j^{k-1}, \quad j = 1, \ldots, N$$

*and corresponding eigenvectors* $v_j = (1, r_j, \ldots, r_j^{N-1})^{\mathrm{T}}, j = 1, \ldots, N,$ *where* $r_j = \exp(2\pi(j-1)/N)$ *are solutions of* $r^N = 1$.

### Passive Tracers in the Velocity Fields of $N$ Point Vortices: The Restricted $(N+1)$-Vortex Problem

The terminology "restricted $(N+1)$-vortex problem" is used in analogy with celestial mechanics literature, when one of the vorticities is taken to be zero. The zero-vorticity vortex does not affect the dynamics of the remaining $N$-vortices. For this reason, it is said to be passively advected by the flow of the remaining $N$-vortices and in the fluid mechanics literature the terminology "passive tracer" is also employed. The tracer dynamics is given by the Hamiltonian equations [8]. Notice that in general the Hamiltonian $\Psi$ is time dependent, through the vortex variables $\boldsymbol{r}_j, j = 1, \ldots, N$, that is,

$$\Psi(\boldsymbol{r}, t) = \Psi(\boldsymbol{r}, \boldsymbol{r}_1(t), \ldots, \boldsymbol{r}_N(t))$$

and $(q, p) = (x, y)$ play the role of conjugate canonical variables. There is an extensive literature on the subject both from theoretical (see, e.g., Boatto and Simó (2004) and Newton (2001)) and an experimental (van Heijst 1993, Ottino 1990) point of view. As discussed in the previous section, there are some vortex configurations, such as the polygonal ones, for which vortices undergo a periodic circular motion. Then by viewing the dynamics in a reference frame co-rotating with the vortices the tracer Hamiltonian is manifestly time independent and, therefore, integrable – since it reduces to a Hamiltonian of one degree of freedom. In such an occurrence, tracer trajectories form a web of homoclinic and heteroclinic orbits. An interesting theoretical problem is to study how the tracer transport properties (i.e., existence of barriers to transport, diffusion etc.) are affected by perturbing the polygonal vortex configuration, that is, by introducing in $\Psi$ a "genuine" time dependence (periodic, quasi-periodic, or chaotic) (see, e.g., Boatto and Pierrehumbert (1999), Rom-Kedar, Leonard and Wiggins (1990), Kuznetsov and Zaslavsky (2000), and Newton (2001)). Furthermore, in the lab experiments, color dyes, which monitor the flow velocity field, are often used as the experimental equivalent of tracer particles. In this context we would like to stress the striking resemblance between theoretical particle trajectories, deduced from point vortex dynamics, and the actual dye visualizations observed by van Heijst and Flor for vortex dipoles in a stratified fluid (see **Figures 11 and 12**) (van Heijst 1993). Similarly, tripolar structures have been observed both in lab experiments (see **Figure 13**) and in nature (see **Figure 14**). Recently, the Danish group of Jansson–Haspang–Jensen–Hersen–Bohr has observed beautiful rotating polygons, such as squares and pentagons, on a fluid surface in the presence of a rotating cylinder (see **Figure 15**).

## Point Vortex Motion with Boundaries

In comparison with the extensive literature on point vortex motion in unbounded domains, the study of point vortex motion in the presence of walls is modest. There is, however, a general theory for such problems, and some recent new developments in this area have resulted in a versatile tool for analyzing point vortex motion with boundaries. Newton (Newton 2001) contains a chapter on point vortex motion with boundaries and also features a detailed bibliography. The reader is referred there for standard treatments; here, we focus on more recent developments of the mathematical theory.

### The Method of Images

When point vortices move around in bounded domains, it is clear that the motion is subject to the constraint that no fluid should penetrate any of

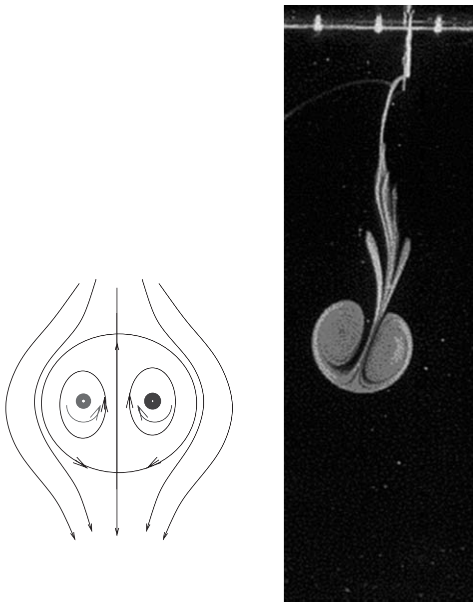


**Figure 11** Test-particle trajectories: on the left, theoretical trajectories, from the point-vortex model; on the right, a top view of a laboratory experiment in stratified flows. Reproduced from van Heijst GJF and Flor JB (1989) Dipole formation and collisions in a stratified fluid. *Nature* 340: 212–215, with permission from Nature Publishing Group.

the boundary walls of the domain. If $\boldsymbol{n}$ denotes the local normal to the boundary walls, the boundary condition on the velocity field $\boldsymbol{u}$ is therefore $\boldsymbol{u} \cdot \boldsymbol{n} = 0$ everywhere on the walls. Another way to say the same thing is that all the walls must be streamlines so that the streamfunction, $\psi$ say, must be constant on any boundary wall.

A classical approach to bounded vortex motion is the celebrated method of images – a rather special technique limited to cases where the domain of interest has certain geometrical symmetries so that an appropriate distribution of image vorticity can be ascertained, essentially by inspection. This image vorticity is placed in nonphysical regions of the plane in order to satisfy the boundary conditions that the walls act as impenetrable barriers for the flow.

The simplest example is the motion of a single vortex next to a straight plane wall of infinite extent. Suppose the wall is along $y = 0$ in an $(x, y)$-plane and that the fluid occupies the upper-half plane. If a circulation-$\Gamma$ vortex is at the complex position $z_0 = x_0 + \mathrm{i}y_0$, the solution for the streamfunction is

$$\psi(z, \bar{z}) = -\frac{\Gamma}{2\pi}\log\left|\frac{z - z_0}{z - \overline{z_0}}\right| \qquad [24]$$

where $z = x + \mathrm{i}y$. This has a single logarithmic singularity in the upper-half plane at $z = z_0$

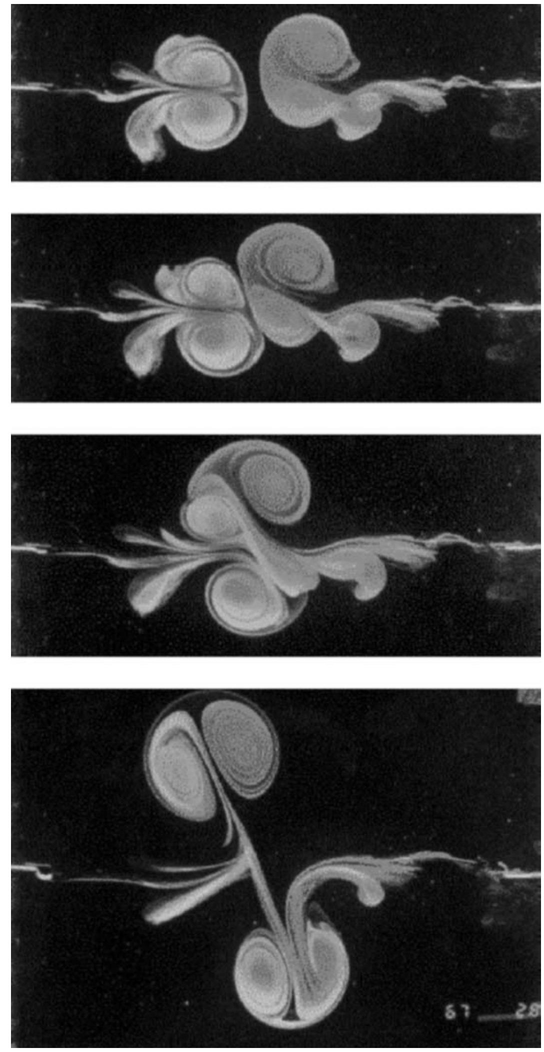


**Figure 12** A frontal collision of two dipoles as observed in a stratified fluid: after a so called “partner-exchange” two new dipoles are formed. Reproduced from van Heijst GJF and Flor JB (1989) Dipole formation and collisions in a stratified fluid. *Nature* 340: 212–215, with permission from Nature Publishing Group.

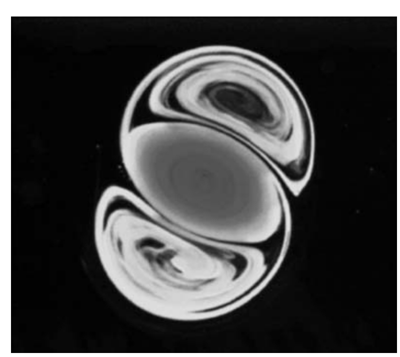


**Figure 13** A tripolar vortex structure as observed in a rotating stratified fluid. Reproduced from van Heijst GJF, Kloosterziel RC, and Williams CWM (1991) Laboratory experiments on the tripolar vortex in a rotating fluid. *Journal of Fluid Mechanics* 225: 301–331, with permission from Cambridge University Press.

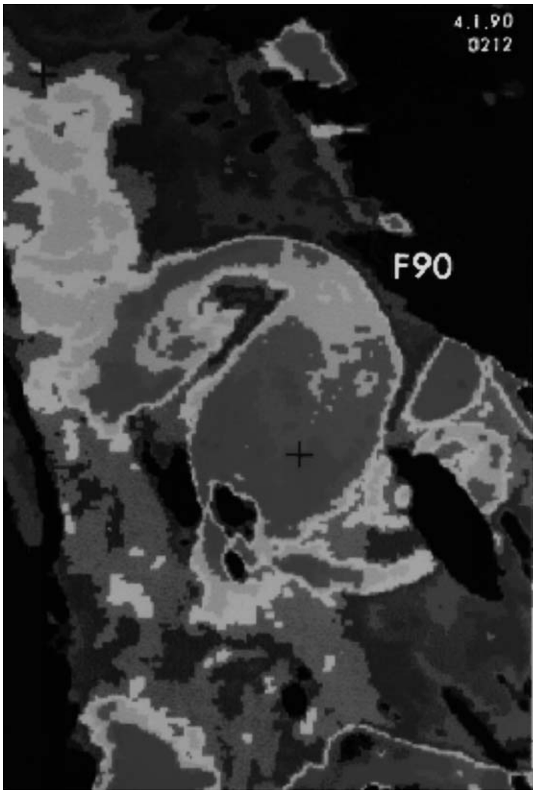



**Figure 14** Infrared image taken by NOAA11 satellite on January 4 1990 (0212 UT) shows a tripolar structure in the Bay of Biscay. The central part of the tripole measures about 50–70 km and rotates clockwise, whereas the two satellite vortices rotate anticlockwise. The dipoles persisted for a few days before it fell apart. Reproduced from Pingree RD and Le Cann B, Anticyclonic Eddy X91 in the Southern Bay of Biscay, *Journal of Geophysical Research*, 97: 14353–14362, May 1991 to February 1992. Copyright (1992) American Geophysical Union. Reproduced/modified by permission of American Geophysical Union.

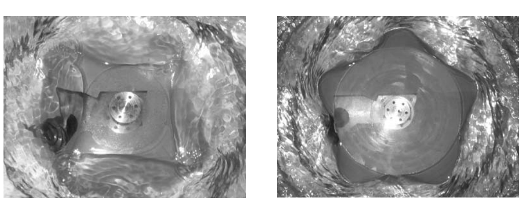


**Figure 15** The free surface of a rotating fluid will, due to the centrifugal force, be pressed radially outward. If the flow is driven by rotating the bottom plate, the axial symmetry can break spontaneously and the surface can take the shape of a rigidly rotating polygon. With water Jansson–Haspang–Jensen–Hersen–Bohr have observed polygons with up to six corners. The rotation speed of the polygons does not coincide with that of the plate, but it is often mode-locked, such that the polygon rotates by one corner for each complete rotation of the plate. Reproduced from Jansson TRN, Haspang M, Jensen KH, Hersen P, and Bohr T (2005) Rotating polygons on a fluid surface. *Preprint*, with permission from T Bohr.

(corresponding to the point vortex) and it is easily checked that $\psi=0$ on $y=0$. Therefore, no fluid penetrates the wall. Equation [24] can be written as

$$\psi(z,\bar{z}) = -\frac{\Gamma}{2\pi}\log|z-z_0| + \frac{\Gamma}{2\pi}\log|z-\overline{z_0}| \qquad [25]$$

which is the sum of the streamfunction due to a point vortex of circulation $\Gamma$ at $z_0=x_0+iy_0$ and another, one imagines, of circulation $-\Gamma$ at $\overline{z_0}=x_0-iy_0$. In this case, the image vortex distribution is simple: it is just the second vortex sitting at the reflected point in the wall. The method of images can be applied to flows in other regions bounded by straight line segments (e.g., wedge regions of various angles (Newton 2001)).

A variant of the method of images is the Milne–Thomson circle theorem relevant to planar flow around a circular cylinder. Given a complex potential $w(z)$ with the required singularities in the fluid region exterior to the cylinder, but failing to satisfy the boundary condition that the surface of the cylinder is a streamline, this theorem says that the correct potential $W(z)$ is

$$W(z) = w(z) + \overline{w}(a^2/z) \qquad [26]$$

where $a$ is the cylinder radius and $\overline{w}(z)$ is the conjugate function to $w(z)$. It is easy to verify that the imaginary part of $W(z)$, that is, the streamfunction, is zero on $|z|=a$. The second term, $\overline{w}(a^2/z)$, produces the required distribution of image vorticity inside the cylinder. A famous example is the Föppl vortex pair which is the simplest model of the trailing vortices shed in the wake of a circular aerofoil traveling at uniform speed.

### Kirchhoff–Routh–Lin Theory

The most important general mathematical tool for point vortex motion in bounded planar regions is the Hamiltonian approach associated with the names of Kirchhoff (1876) and Routh (1881), who developed the early theory. It is now known that the problem of $N$-vortex motion in a simply connected domain is a Hamiltonian dynamical system. Moreover, the Hamiltonian has simple transformation properties when a given flow domain of interest is mapped conformally to another – a result originally due to Routh. A formula for the Hamiltonian can be built from knowledge of the instantaneous Green's function associated with motion of the point vortex in the simply connected domain $D$. In fact, [24] is precisely the relevant Green's function when $D$ is the upper-half plane.

Much later, in 1941, Lin (1941a) extended these general results to the case of multiply connected fluid regions. To visualize such a region, think of a bounded region of the plane containing fluid but also a finite number of impenetrable islands whose boundaries act as barriers for the fluid motion. If the islands are infinitely thin, they can be thought of as straight wall segments immersed in the flow (see later examples). Lin (1941b) showed that both the Hamiltonian structure, and the transformation properties of the Hamiltonian under conformal mapping, are preserved in the multiply connected case.

### Lin's Special Green's Function

Since Lin's result subsumes the earlier simply connected studies, we now outline the key results as presented in Lin (1941a). Consider a fluid region $D$, with outer boundary $C_0$ and $M$ enclosed islands each having boundaries $\{C_j | j = 1, \ldots, M\}$. Lin introduced a special Green's function $G(x, y; x_0, y_0)$ satisfying the following properties:

1. the function

$$g(x, y; x_0, y_0) = -G(x, y; x_0, y_0) - \frac{1}{2\pi}\log r_0 \qquad [27]$$

is harmonic with respect to $(x, y)$ throughout the region $D$ including at the point $(x_0, y_0)$. Here, $r_0 = \sqrt{(x - x_0)^2 + (y - y_0)^2}$;

2. if $\partial G/\partial n$ is the normal derivative of $G$ on a curve then

$$G(x, y; x_0, y_0) = A_k, \quad \text{on } C_k,\ k = 1, \ldots, M$$
$$\oint_{C_k} \frac{\partial G}{\partial n}\, \mathrm{d}s = 0, \quad k = 1, \ldots, M \qquad [28]$$

where $\mathrm{d}s$ denotes an element of arc and $\{A_k\}$ are constants;

3. $G(x, y; x_0, y_0) = 0$ on $C_0$.

Flucher and Gustafsson (1997) refer to this $G$ as the hydrodynamic Green's function. (In fact, it coincides with the modified Green's function arising in abstract potential theory – a function that is dual to the usual first-type Green's function that equals zero on all the domain boundaries.) On the use of $G$, Lin established the following two key results:

**Theorem 4** *If $N$ vortices of strengths $\{\Gamma_k | k = 1, \ldots, N\}$ are present in an incompressible fluid at the points $\{(x_k, y_k) | k = 1, \ldots, N\}$ in a general multiply connected region $D$ bounded by fixed boundaries, the stream function of the fluid motion is given by*

$$\psi(x, y; x_k, y_k) = \psi_0(x, y) + \sum_{k=1}^{N} \Gamma_k G(x, y; x_k, y_k) \qquad [29]$$

*where $\psi_0(x, y)$ is the streamfunction due to outside agencies and is independent of the point vortex positions.*

**Theorem 5** *For the motion of vortices of strengths $\{\Gamma_k | k = 1, \ldots, N\}$ in a general region $D$ bounded by fixed boundaries, there exists a Kirchhoff–Routh function $H(\{x_k, y_k\})$, depending on the point vortex positions, such that*

$$\Gamma_k \frac{\mathrm{d}x_k}{\mathrm{d}t} = \frac{\partial H}{\partial y_k}, \qquad \Gamma_k \frac{\mathrm{d}y_k}{\mathrm{d}t} = -\frac{\partial H}{\partial x_k} \qquad [30]$$

*where $H(\{x_k, y_k\})$ is given by*

$$H(\{x_k, y_k\}) = \sum_{k=1}^{N} \Gamma_k \psi_0(x_k, y_k) + \sum_{\substack{k_1, k_2 = 1 \\ k_1 > k_2}}^{N} \Gamma_{k_1}\Gamma_{k_2} G(x_{k_1}, y_{k_1}; x_{k_2}, y_{k_2}) - \frac{1}{2}\sum_{k=1}^{N} \Gamma_k^2 g(x_k, y_k; x_k, y_k) \qquad [31]$$

*In rescaled coordinates $(x_k, \Gamma_k y_k)$, [30] is a Hamiltonian system in canonical form. For historical reasons, $H$ is often called the Kirchhoff–Routh path function. Analyzing the separate contributions to the path function [31] is instructive: the first term is the contribution from flows imposed from outside (e.g., background flows and round-island circulations), the second term is the "free-space" contribution (it is the relevant Hamiltonian when no boundaries are present) while the third term encodes the effect of the boundary walls (or, the effect of the "image vorticity" distribution discussed earlier).*

Lin (1941a) went on to show that, with the Hamiltonian in some $D$ given by $H$ in [31], the Hamiltonian relevant to vortex motion in another domain obtained from $D$ by a conformal mapping $z(\zeta)$ consists of [31] with some simple extra additive contributions dependent only on the derivative of the map $z(\zeta)$ evaluated at the point vortex positions.

Flucher and Gustafsson (1997) also introduce the Robin function $\mathcal{R}(x_0, y_0)$ defined as the regular part of the above hydrodynamic Green's function evaluated at the point vortex. Indeed, $\mathcal{R}(x_0, y_0) \equiv g(x_0, y_0; x_0, y_0)$, where $g$ is defined in [27]. An interesting fact is that, for single-vortex motion in a simply connected domain, $\mathcal{R}(x_0, y_0)$ satisfies the quasilinear elliptic Liouville equation everywhere in

$D$ with the boundary condition that it becomes infinite everywhere on the boundary of $D$.

By combining the Kirchhoff–Routh theory with conformal mapping theory, many interesting problems can be studied. What happens, for example, if there is a gap in the wall of **Figure 16**? In recent work, Johnson and McDonald (2005) show that if the vortex starts off, far from the gap, at a distance of less than half the gap width from the wall, then it will eventually penetrate the gap. Otherwise, it will dip towards the gap but not go through it. The trajectories are shown in **Figure 17**.

Unfortunately, Lin did not provide any explicit analytical expressions for $G$ in the multiply connected case. This has limited the applicability of his theory beyond fluid regions that are anything other than simply and doubly connected. Recently, however, Lin's theory has recently been brought to implementational fruition by Crowdy and Marshall (2005a), who, up to conformal mapping, have derived explicit formulas for the hydrodynamic Green's function in multiply connected fluid regions of arbitrary finite connectivity. Their approach makes use of elements of classical function theory dating back to the work of Poincaré, Schottky, and Klein (among others). This allows new problems involving bounded vortex motion to be tackled. For example, the motion of a single vortex around multiple circular islands has been studied in Crowdy and Marshall (2005b), thereby extending recent work on the two-island problem (Johnson and McDonald 2005). If the wall in **Figure 17** happens to have two (or more) gaps, then the fluid region is multiply connected. The two-gap (doubly connected) case was recently solved by Johnson and McDonald (2005) using Schwarz–Christoffel maps combined with elements of elliptic function theory (see **Figure 18**). Crowdy and Marshall have solved the problem of an arbitrary number of gaps in a wall by exploiting the new general theory presented in Crowdy and Marshall (2005a,b) (and related works by the authors). The case of a wall with three gaps represents a triply connected fluid region and the critical vortex trajectory is plotted in **Figure 19**.

Point vortex motion in bounded domains on the surface of a sphere has received scant attention in

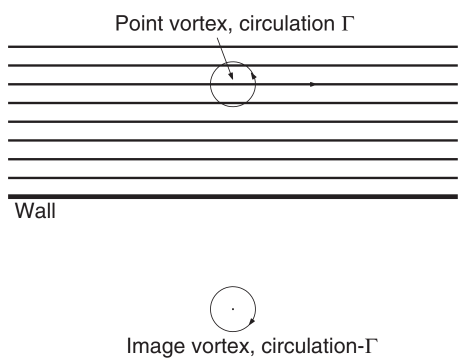



**Figure 16** The motion of a point vortex near an infinite straight wall. The vortex moves, at constant speed, maintaining a constant distance from the wall. Other possible trajectories are shown; they are all straight lines parallel to the wall. The motion can be thought of as being induced by an opposite-circulation "image" vortex at the reflected point in the wall.

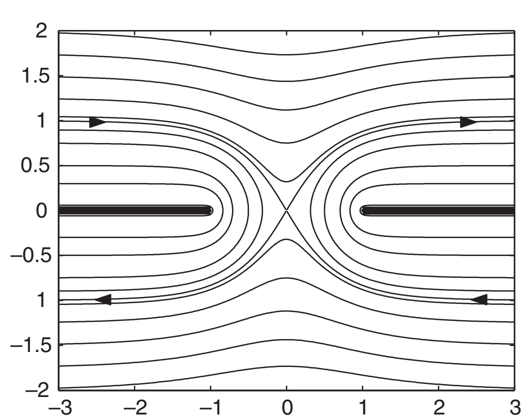



**Figure 17** Distribution of point vortex trajectories near a wall with a single gap of length 2. There is a critical trajectory which, far from the gap, is unit distance from the wall.

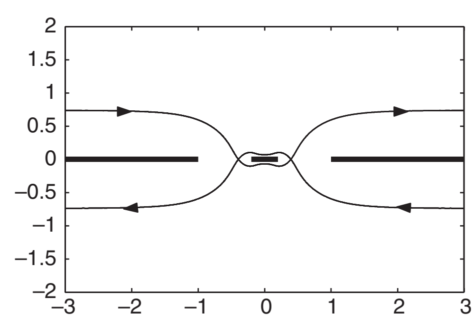



**Figure 18** The critical trajectory when there are two symmetric gaps in a wall. The fluid region is now doubly connected. This problem is solved in Johnson and McDonald (2005) and Crowdy and Marshall (2005).

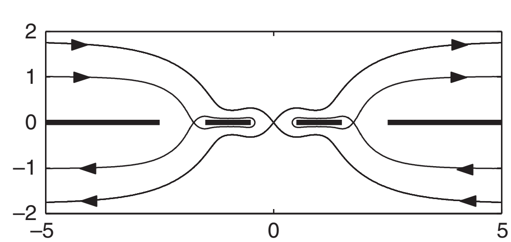



**Figure 19** The critical vortex trajectories when there are three gaps in the wall. This time the fluid region is triply connected. This problem is solved in Crowdy and Marshall (2005) using the general methods in Crowdy and Marshall (2005).

the literature, although Kidambi and Newton (2000) and Newton (2001) have recently made a contribution. Such paradigms are clearly relevant to planetary-scale oceanographic flows in which oceanic eddies interact with topography such as ridges and land masses and deserve further study.

## Further Reading


Albouy A (1996) The symmetric central configurations of four equal masses. *Contemporary Mathematics* 198: 131–135.

Aref H and Stremler MA (1999) Four-vortex motion with zero total circulation and impulse. *Physics of Fluids* 11(12): 3704–3715.

Aref H, Newton PK, Stremler MA, Tokieda T, and Vainchtein DL (2003) Vortex crystals. *Advances in Applied Mathematics* 39: 1–79.

Ball WWR (1940) *A Short Account of the History of Mathematics*, 12th edn. London: MacMillan.

Boatto S and Pierrehumbert RT (1999) Dynamics of a passive tracer in the velocity field of four identical point-vortices. *Journal of Fluid Mechanics* 394: 137–174.

Boatto S and Cabral HE (2003) Nonlinear stability of a latitudinal ring of point vortices on a non-rotating sphere. *SIAM Journal of Applied Mathematics* 64: 216–230.

Boatto S and Simó C (2004) Stability of latitudinal vortex rings with polar vortices. Mathematical Physics Preprint Archive (mp_arc) 04-67.

Cabral HE and Schmidt DS (1999/00) Stability of relative equilibria in the problem of $N + 1$ vortices. *SIAM Journal of Mathematical Analysis* 31(2): 231–250.

Carnevale GF, McWilliams JC, Pomeau Y, Weiss JB, and Young WR (1991) Evolution of vortex statistics in two-dimensional turbulence. *Physical Review Letters* 66(21): 2735–2737.

Carnevale GF, McWilliams JC, Pomeau Y, Weiss JB, and Young WR (1992) Rates, pathways, and end states of nonlinear evolution in decaying two-dimensional. *Physics of Fluids A* 4(6): 1314–1316.

Castilla MSAC, Moauro V, Negrini P, and Oliva WM (1993) The four positive vortices problem – regions of chaotic behavior and non-integrability. *Annales de l'Institut Henri Poincaré. Section A. Physique Theorique* 59(1): 99–115.

Crowdy DG and Marshall JS (2005a) Analytical formulae for the Kirchhoff–Routh path function in multiply connected domains. *Proceedings of the Royal Society A* 461: 2477–2501.

Crowdy DG and Marshall JS (2005b) The motion of a point vortex around multiple circular islands. *Physics of Fluids* 17: 560–602.

Dritschel DG (1985) The stability and energetics of co-rotating uniform vortices. *Journal of Fluid Mechanics* 157: 95–134.

Flucher M and Gustafsson B (1997) *Vortex Motion in Two Dimensional Hydrodynamics*. Royal Institute of Technology Report No. TRITA-MAT-1997-MA-02.

Glass K (2000) Symmetry and bifurcations of planar configurations of the *N*-body and other problems. *Dynamics and Stability of Systems* 15(2): 59–73.

van Heijst GJF and Flor JB (1989) Dipole formation and collisions in a stratified fluid. *Nature* 340: 212–215.

van Heijst GJF, Kloosterziel RC, and Williams CWM (1991) Laboratory experiments on the tripolar vortex in a rotating fluid. *Journal of Fluid Mechanics* 225: 301–331.

van Heijst GJF (1993) Self-organization of two-dimensional flows. *Nederlands Tijdschrift voor Natuurkunde* 59: 321–325 (http://www.fluid.tue.nl).

von Helmholtz H (1858) On the integrals of the hydrodynamical equations which express vortex motion. *Philosophical Magazine* 4(33): 485–512.

Jansson TRN, Haspang M, Jensen KH, Hersen P, and Bohr T (2005) Rotating polygons on a fluid surface. *Preprint*.

Johnson ER and McDonald NR (2005) Vortices near barriers with multiple gaps. *Journal of Fluid Mechanics* 531: 335–358.

Kidambi R and Newton PK (2000) Point vortex motion on a sphere with solid boundaries. *Physical Fluids* 12: 581–588.

Kimura Y (1999) Vortex motion on surfaces with constant curvature. *Proceedings of the Royal Society of London A* 455: 245–259.

Kirchhoff G (1876) *Vorlesunger über mathematische Physik, Mechanik*. Leipzig.

Koiller J and Carvalho SP (1989) Non-integrability of the 4-vortex system: analytical proof. *Communications in Mathematical Physics* 120(4): 643–652.

Kuznetsov L and Zavlasky GM (2000) Passive tracer transport in three-vortex flow. *Physical Review A* 61(4): 3777–3792.

Lim C, Montaldi J, and Roberts M (2001) Relative equilibria of point vortices on the sphere. *Physica D* 148: 97–135.

Lin CC (1941a) On the motion of vortices in two dimensions. I. Existence of the Kirchhoff–Routh function. *Proceedings of the National Academy of Sciences* 27(12): 570–575.

Lin CC (1941b) On the motion of vortices in two dimensions. II. Some further investigations on the Kirchhoff–Routh function. *Proceedings of the National Academy Sciences* 27(12): 575–577.

Marchioro C and Pulvirenti M (1993) Vortices and localization in Euler flows. *Communications in Mathematical Physics* 154: 49–61.

Marchioro C and Pulvirenti M (1994) *Mathematical Theory of Incompressible Non-viscous Fluids*. vol. 96, AMS. New York: Springer.

Mayer AM (1878) Floating magnetics. *Nature* 17: 487–488.

Mayer AM (1878) *Scientific American* 2045–2047.

Mayer AM (1878) On the morphological laws of the configurations formed by magnets floating vertically and subjected to the attraction of a superposed magnetic. *American Journal of Science* 16: 247–256.

Montaldi J, Soulière A, and Tokieda T (2002) Vortex dyanmics on a cylinder. *SIAM Journal of Applied Dynamical Systems* 2(3): 417–430.

Newton PK (2001) *The N-Vortex Problem. Analytical Techniques*. New York: Springer.

Ottino JM (1990) *The Kinematics of Mixing: Stretching, Chaos and Transport*. Cambridge: Cambridge University Press.

Pingree RD and Le Cann B (1992) Anticyclonic Eddy X91 in the Southern Bay of Biscay. *Journal of Geophysical Research* 97: 14353–14362.

Polvani LM and Dritschel DG (1993) Wave and vortex dynamics on the surface of a sphere. *Journal of Fluid Mechanics* 255: 35–64.

Poupaud F (2002) Diagonal defect measures, adhesion dynamics and Euler equation. *Methods and Applications of Analysis* 9(4): 533–562.

Rom-Kedar V, Leonard A, and Wiggins S (1990) An analytical study of transport, mixing and chaos in an unsteady vortical flow. *Journal of Fluid Mechanics* 214: 347–394.

Routh E (1881) Some applications of conjugate functions. *Proceedings of the London Mathematical Society* 12: 73–89.

Sakajo T (2004) Transition of global dynamics of a polygonal vortex ring on a sphere with pole vortices. *Physica D* 196: 243–264.

Schochet S (1995) The weak vorticity formulation of the 2-D Euler equations and concentration-cancellation. *Communications in Partial Differential Equations* 20(5&6): 1077–1104.

Soulière A and Tokieda T (2002) Periodic motion of vortices on surfaces with symmetries. *Journal of Fluid Mechanics* 460: 83–92.

Thomson JJ (1883) *A Treatise on the Motion of Vortex Rings.* New York: Macmillan.

Thomson JJ (1904) *Electricity and Matter.* Westmister Archibald Constable.

Thomson W (1867) On vortex atoms. *Proceedings of the Royal Society of Edinburgh* 6: 94–105.

Yarmchuk EJ, Gordon MJV, and Packard RE (1979) Observation of stationary vortices arrays in rotating superfluid helium. *Physical Review Letters* 43(3): 214–217.

Ziglin SL (1982) Quasi-periodic motions of vortex systems. *Physica D* 4: 261–269 (addendum to K M Khanin).

**Poisson Lie Groups** *see* Classical *r*-Matrices, Lie Bialgebras, and Poisson Lie Groups

# Poisson Reduction

**J-P Ortega**, Université de Franche-Comté, Besançon, France
**T S Ratiu**, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland



## Introduction

The Poisson reduction techniques allow the construction of new Poisson structures out of a given one by combination of two operations: "restriction" to submanifolds that satisfy certain compatibility assumptions and passage to a "quotient space" where certain degeneracies have been eliminated. For certain kinds of reduction, it is necessary to pass first to a submanifold and then take a quotient. Before making this more explicit, we introduce the notations that will be used in this article. All manifolds in this article are finite dimensional.

### Poisson Manifolds

A "Poisson manifold" is a pair $(M, \{\cdot\,,\cdot\})$, where $M$ is a manifold and $\{\cdot\,,\cdot\}$ is a bilinear operation on $C^\infty(M)$ such that $(C^\infty(M), \{\cdot\,,\cdot\})$ is a Lie algebra and $\{\cdot\,,\cdot\}$ is a derivation (i.e., the Leibniz identity holds) in each argument. The pair $(C^\infty(M), \{\cdot\,,\cdot\})$ is also called a "Poisson algebra." The functions in the center $\mathcal{C}(M)$ of the Lie algebra $(C^\infty(M), \{\cdot\,,\cdot\})$ are called "Casimir functions." From the natural isomorphism between derivations on $C^\infty(M)$ and vector fields on $M$, it follows that each $h \in C^\infty(M)$ induces a vector field on $M$ via the expression $X_h = \{\cdot\,, h\}$, called the "Hamiltonian vector field" associated to the "Hamiltonian function" $h$. The triplet $(M, \{\cdot\,,\cdot\}, h)$ is called a "Poisson dynamical system." Any Hamiltonian system on a symplectic manifold is a Poisson dynamical system relative to the Poisson bracket induced by the symplectic structure. Given a Poisson dynamical system $(M, \{\cdot\,,\cdot\}, h)$, its "integrals of motion" or "conserved quantities" are defined as the centralizer of $h$ in $(C^\infty(M), \{\cdot\,,\cdot\})$ that is, the subalgebra of $(C^\infty(M), \{\cdot\,,\cdot\})$ consisting of the functions $f \in C^\infty(M)$ such that $\{f, h\} = 0$. Note that the terminology is justified since, by Hamilton's equations in Poisson bracket form, we have $\dot{f} = X_h[f] = \{f, h\} = 0$, that is, $f$ is constant on the flow of $X_h$. A smooth mapping $\varphi : M_1 \to M_2$, between the two Poisson manifolds $(M_1, \{\cdot\,,\cdot\}_1)$ and $(M_2, \{\cdot\,,\cdot\}_2)$, is called "canonical" or "Poisson" if for all $g$, $h \in C^\infty(M_2)$ we have $\varphi^*\{g, h\}_2 = \{\varphi^* g, \varphi^* g\}_1$. If $\varphi : M_1 \to M_2$ is a smooth map between two Poisson manifolds $(M_1, \{\cdot\,,\cdot\}_1)$ and $(M_2, \{\cdot\,,\cdot\}_2)$, then $\varphi$ is a Poisson map if and only if $T\varphi \circ X_{h\circ\varphi} = X_h \circ \varphi$ for any $h \in C^\infty(M_2)$, where $T\varphi : TM_1 \to TM_2$ denotes the tangent map (or derivative) of $\varphi$.

Let $(S, \{\cdot\,,\cdot\}^S)$ and $(M, \{\cdot\,,\cdot\}^M)$ be two Poisson manifolds such that $S \subset M$ and the inclusion $i_S : S \hookrightarrow M$ is an immersion. The Poisson manifold $(S, \{\cdot\,,\cdot\}^S)$ is called a "Poisson submanifold" of $(M, \{\cdot\,,\cdot\}^M)$ if $i_S$ is a canonical map. An immersed submanifold $Q$ of $M$ is called a "quasi-Poisson submanifold" of $(M, \{\cdot\,,\cdot\}^M)$ if for any $q \in Q$, any open neighborhood $U$ of $q$ in $M$, and any $f \in C^\infty(U)$ we have $X_f(i_Q(q)) \in T_q i_Q(T_q Q)$, where $i_Q : Q \hookrightarrow M$ is the inclusion and $X_f$ is the Hamiltonian vector field of $f$ on $U$ with respect to the Poisson bracket of $M$ restricted to $U$. If $(S, \{\cdot\,,\cdot\}^S)$ is a Poisson submanifold of $(M, \{\cdot\,,\cdot\}^M)$, then there is no other bracket $\{\cdot\,,\cdot\}'$ on $S$ making the inclusion $i : S \hookrightarrow M$ into a canonical map. If $Q$ is a quasi-Poisson submanifold of $(M, \{\cdot\,,\cdot\})$, then there exists a unique Poisson structure $\{\cdot\,,\cdot\}^Q$ on $Q$ that makes it into a Poisson submanifold of $(M, \{\cdot\,,\cdot\})$ but this Poisson structure may be different from the given one on $Q$. Any Poisson submanifold is quasi-Poisson but the converse is not true in general.

### The Poisson Tensor and Symplectic Leaves

The derivation property of the Poisson bracket implies that for any two functions $f, g \in C^\infty(M)$, the value of the bracket $\{f, g\}(z)$ at an arbitrary point $z \in M$ (and therefore $X_f(z)$ as well) depends on $f$ only through $\boldsymbol{d}f(z)$ which allows us to define a contravariant antisymmetric 2-tensor $B \in \Lambda^2(T^*M)$, called the "Poisson tensor," by $B(z)(\alpha_z, \beta_z) = \{f, g\}(z)$, where $\boldsymbol{d}f(z) = \alpha_z \in T_z^*M$ and $\boldsymbol{d}g(z) = \beta_z \in T_z^*M$. The vector bundle map $B^\sharp : T^*M \to TM$ over the identity naturally associated to $B$ is defined by $B(z)(\alpha_z, \beta_z) = \langle \alpha_z, B^\sharp(\beta_z) \rangle$. Its range $D := B^\sharp(T^*M) \subset TM$ is called the "characteristic distribution" of $(M, \{\cdot\,,\cdot\})$ since $D$ is a generalized smooth integrable distribution. Its maximal integral leaves are called the "symplectic leaves" of $M$ for they carry a symplectic structure that makes them into Poisson submanifolds. As integral leaves of an integrable distribution, the symplectic leaves $\mathcal{L}$ are "initial submanifolds" of $M$, that is, the inclusion $i : \mathcal{L} \hookrightarrow M$ is an injective immersion such that for any smooth manifold $P$, an arbitrary map $g : P \to \mathcal{L}$ is smooth if and only if $i \circ g : P \to M$ is smooth.

## Poisson Reduction

### Canonical Lie Group Actions

Let $(M, \{\cdot\,,\cdot\})$ be a Poisson manifold and let $G$ be a Lie group acting canonically on $M$ via the map $\Phi : G \times M \to M$. An action is called "canonical" if for any $h \in G$ and $f, g \in C^\infty(M)$, one has

$$\{f \circ \Phi_h, g \circ \Phi_h\} = \{f, g\} \circ \Phi_h$$

If the $G$-action is free and proper, then the orbit space $M/G$ is a smooth regular quotient manifold. Moreover, it is also a Poisson manifold with the Poisson bracket $\{\cdot\,,\cdot\}^{M/G}$, uniquely characterized by the relation

$$\{f, g\}^{M/G}(\pi(m)) = \{f \circ \pi, g \circ \pi\}(m) \qquad [1]$$

for any $m \in M$ and where $f, g : M/G \to \mathbb{R}$ are two arbitrary smooth functions. This bracket is appropriate for the reduction of Hamiltonian dynamics in the sense that if $h \in C^\infty(M)^G$ is a $G$-invariant smooth function on $M$, then the Hamiltonian flow $F_t$ of $X_h$ commutes with the $G$-action, so it induces a flow $F_t^{M/G}$ on $M/G$ that is Hamiltonian on $(M/G, \{\cdot\,,\cdot\}^{M/G})$ for the reduced Hamiltonian function $[h] \in C^\infty(M/G)$ defined by $[h] \circ \pi = h$.

If the Poisson manifold $(M, \{\cdot\,,\cdot\})$ is actually symplectic with form $\omega$ and the $G$-action has an associated momentum map $\boldsymbol{J} : M \to \mathfrak{g}^*$, then the symplectic leaves of $(M/G, \{\cdot\,,\cdot\}^{M/G})$ are given by the spaces $(M^c_{\mathcal{O}_\mu} := G \cdot \boldsymbol{J}^{-1}(\mu)^c/G, \omega^c_{\mathcal{O}_\mu})$, where $\boldsymbol{J}^{-1}(\mu)^c$ is a connected component of the fiber $\boldsymbol{J}^{-1}(\mu)$ and $\omega^c_{\mathcal{O}_\mu}$ is the restriction to $M^c_{\mathcal{O}_\mu}$ of the symplectic form $\omega_{\mathcal{O}_\mu}$ of the symplectic orbit reduced space $M_{\mathcal{O}_\mu}$ (*see* Symmetry and Symplectic Reduction). If, additionally, $G$ is compact, $M$ is connected, and the momentum map $\boldsymbol{J}$ is proper, then $M^c_{\mathcal{O}_\mu} = M_{\mathcal{O}_\mu}$.

In the remainder of this section, we characterize the situations in which new Poisson manifolds can be obtained out of a given one by a combination of restriction to a submanifold and passage to the quotient with respect to an equivalence relation that encodes the symmetries of the bracket.

**Definition 1** Let $(M, \{\cdot\,,\cdot\})$ be a Poisson manifold and $D \subset TM$ a smooth distribution on $M$. The distribution $D$ is called "Poisson" or "canonical," if the condition $\boldsymbol{d}f|_D = \boldsymbol{d}g|_D = 0$, for any $f, g \in C^\infty(U)$ and any open subset $U \subset P$, implies that $\boldsymbol{d}\{f, g\}|_D = 0$.

Unless strong regularity assumptions are invoked, the passage to the leaf space of a canonical distribution destroys the smoothness of the quotient topological space. In such situations, the Poisson algebra of functions is too small and the notion of presheaf of Poisson algebras is needed. *See* Singularity and Bifurcation Theory for more information on singularity theory.

**Definition 2** Let $M$ be a topological space with a presheaf $\mathcal{F}$ of smooth functions. A presheaf of Poisson algebras on $(M, \mathcal{F})$ is a map $\{\cdot\,,\cdot\}$ that assigns to each open set $U \subset M$ a bilinear operation $\{\cdot\,,\cdot\}_U : \mathcal{F}(U) \times \mathcal{F}(U) \to \mathcal{F}(U)$ such that the pair $(\mathcal{F}(U), \{\cdot\,,\cdot\}_U)$ is a Poisson algebra. A presheaf of Poisson algebras is denoted as a triple $(M, \mathcal{F}, \{\cdot\,,\cdot\})$. The presheaf of Poisson algebras $(M, \mathcal{F}, \{\cdot\,,\cdot\})$ is said to be "nondegenerate" if the following condition holds: if $f \in \mathcal{F}(U)$ is such that $\{f, g\}_{U \cap V} = 0$, for any $g \in \mathcal{F}(V)$ and any open set of $V$, then $f$ is constant on the connected components of $U$.

Any Poisson manifold $(M, \{\cdot\,,\cdot\})$ has a natural presheaf of Poisson algebras on its presheaf of smooth functions that associates to any open subset $U$ of $M$ the restriction $\{\cdot\,,\cdot\}|_U$ of $\{\cdot\,,\cdot\}$ to $C^\infty(U) \times C^\infty(U)$.

**Definition 3** Let $P$ be a topological space and $\mathcal{Z} = \{S_i\}_{i \in I}$ a locally finite partition of $P$ into smooth manifolds $S_i \subset P, i \in I$, that are locally closed topological subspaces of $P$ (hence their manifold topology is the relative one induced by $P$). The pair $(P, \mathcal{Z})$ is called a "decomposition" of $P$ with "pieces" in $\mathcal{Z}$, or a "decomposed space," if the following "frontier condition" holds:

**Condition** (**DS**) If $R, S \in \mathcal{Z}$ are such that $R \cap \bar{S} \neq \emptyset$, then $R \subset \bar{S}$. In this case, we write $R \preceq S$. If, in addition, $R \neq S$ we say that $R$ is incident to $S$ or that it is a boundary piece of $S$ and write $R \prec S$.

**Definition 4** Let $M$ be a differentiable manifold and $S \subset M$ a decomposed subset of $M$. Let $\{S_i\}_{i \in I}$

be the pieces of this decomposition. The topology of $S$ is not necessarily the relative topology as a subset of $M$. Then $D \subset TM|_S$ is called a "smooth distribution" on $S$ adapted to the decomposition $\{S_i\}_{i \in I}$, if $D \cap TS_i$ is a smooth distribution on $S_i$ for all $i \in I$. The distribution $D$ is said to be "integrable" if $D \cap TS_i$ is integrable for each $i \in I$.

In the situation described by the previous definition and if $D$ is integrable, the integrability of the distributions $D_{S_i} := D \cap TS_i$ on $S_i$ allows us to partition each $S_i$ into the corresponding maximal integral manifolds. Thus, there is an equivalence relation on $S_i$ whose equivalence classes are precisely these maximal integral manifolds. Doing this on each $S_i$, we obtain an equivalence relation $D_S$ on the whole set $S$ by taking the union of the different equivalence classes corresponding to all the $D_{S_i}$. Define the quotient space $S/D_S$ by

$$S/D_S := \bigcup_{i \in I} S_i/D_{S_i}$$

and let $\pi_{D_S}: S \to S/D_S$ be the natural projection.

### The Presheaf of Smooth Functions on $S/D_S$

Define the presheaf of smooth functions $C^\infty_{S/D_S}$ on $S/D_S$ as the map that associates to any open subset $V$ of $S/D_S$ the set of functions $C^\infty_{S/D_S}(V)$ characterized by the following property: $f \in C^\infty_{S/D_S}(V)$ if and only if for any $z \in V$ there exists $m \in \pi^{-1}_{D_S}(V)$,$U_m$ open neighborhood of $m$ in $M$, and $F \in C^\infty(U_m)$ such that

$$f \circ \pi_{D_S}|_{\pi^{-1}_{D_S}(V) \cap U_m} = F|_{\pi^{-1}_{D_S}(V) \cap U_m} \qquad [2]$$

$F$ is called a "local extension" of $f \circ \pi_{D_S}$ at the point $m \in \pi^{-1}_{D_S}(V)$. When the distribution $D$ is trivial, the presheaf $C^\infty_{S/D_S}$ coincides with the presheaf of Whitney smooth functions $C^\infty_{S,M}$ on $S$ induced by the smooth functions on $M$.

The presheaf $C^\infty_{S/D_S}$ is said to have the $(D, D_S)$-local extension property when the topology of $S$ is stronger than the relative topology and, at the same time, the local extensions of $f \circ \pi_{D_S}$ defined in [2] can always be chosen to satisfy

$$\mathbf{d}F(n)|_{D(n)} = 0 \quad \text{for any } n \in \pi^{-1}_{D_S}(V) \cap U_m$$

$F$ is called a "local $D$-invariant extension" of $f \circ \pi_{D_S}$ at the point $m \in \pi^{-1}_{D_S}(V)$. If $S$ is a smooth embedded submanifold of $M$ and $D_S$ is a smooth, integrable, and regular distribution on $S$, then the presheaf $C^\infty_{S/D_S}$ coincides with the presheaf of smooth functions on $S/D_S$ when considered as a regular quotient manifold.

The following definition spells out what we mean by obtaining a bracket via reduction.

**Definition 5** Let $(M, \{\cdot\,,\cdot\})$ be a Poisson manifold, $S$ a decomposed subset of $M$, and $D \subset TM|_S$ a Poisson-integrable generalized distribution adapted to the decomposition of $S$. Assume that $C^\infty_{S/D_S}$ has the $(D, D_S)$-local extension property. Then $(M, \{\cdot\,,\cdot\}, D, S)$ is said to be "Poisson reducible" if $(S/D_S, C^\infty_{S/D_S}, \{\cdot\,,\cdot\}^{S/D_S})$ is a well-defined presheaf of Poisson algebras where, for any open set $V \subset S/D_S$, the bracket $\{\cdot\,,\cdot\}^{S/D_S}_V : C^\infty_{S/D_S}(V) \times C^\infty_{S/D_S}(V) \to C^\infty_{S/D_S}(V)$ is given by

$$\{f, g\}^{S/D_S}_V(\pi_{D_S}(m)) := \{F, G\}(m)$$

for any $m \in \pi^{-1}_{D_S}(V)$ for local $D$-invariant extensions $F$,$G$ at $m$ of $f \circ \pi_{D_S}$ and $g \circ \pi_{D_S}$, respectively.

**Theorem 1** *Let $(M, \{\cdot\,,\cdot\})$ be a Poisson manifold with associated Poisson tensor $B \in \Lambda^2(T^*M)$, $S$ a decomposed space, and $D \subset TM|_S$ a Poisson-integrable generalized distribution adapted to the decomposition of $S$ (see Definitions 4 and 1). Assume that $C^\infty_{S/D_S}$ has the $(D, D_S)$-local extension property. Then $(M, \{\cdot\,,\cdot\}, D, S)$ is Poisson reducible if for any $m \in S$*

$$B^\sharp(\Delta_m) \subset \left[\Delta^S_m\right]^\circ \qquad [3]$$

*where $\Delta_m := \{\mathbf{d}F(m) | F \in C^\infty(U_m), \mathbf{d}F(z)|_{D(z)} = 0,$ for all $z \in U_m \cap S$, and for any open neighborhood $U_m$ of $m$ in $M\}$ and $\Delta^S_m := \{\mathbf{d}F(m) \in \Delta_m | F|_{U_m \cap V_m}$ is constant for an open neighborhood $U_m$ of $m$ in $M$ and an open neighborhood $V_m$ of $m$ in $S\}$.*

If $S$ is endowed with the relative topology, then $\Delta^S_m := \{\mathbf{d}F(m) \in \Delta_m | F|_{U_m \cap V_m}$ is constant for an open neighborhood $U_m$ of $m$ in $M\}$.

### Reduction by Regular Canonical Distributions

Let $(M, \{\cdot\,,\cdot\})$ be a Poisson manifold and $S$ an embedded submanifold of $M$. Let $D \subset TM|_S$ be a sub-bundle of the tangent bundle of $M$ restricted to $S$ such that $D_S := D \cap TS$ is a smooth, integrable, regular distribution on $S$ and $D$ is canonical.

**Theorem 2** *With the above hypotheses, $(M, \{\cdot\,,\cdot\}, D, S)$ is Poisson reducible if and only if*

$$B^\sharp(D^\circ) \subset TS + D \qquad [4]$$

## Applications of the Poisson Reduction Theorem

### Reduction of Coisotropic Submanifolds

Let $(M, \{\cdot\,,\cdot\})$ be a Poisson manifold with associated Poisson tensor $B \in \Lambda^2(T^*M)$ and $S$ an immersed smooth submanifold of $M$. Denote by $(TS)^\circ := \{\alpha_s \in T^*_sM | \langle \alpha_s, v_s \rangle = 0,$ for all $s \in S,$ $v_s \in T_sS\} \subset T^*M$ the

conormal bundle of the manifold $S$; it is a vector sub-bundle of $T^*M|_S$. The manifold $S$ is called "coisotropic" if $B^\sharp((TS)^\circ) \subset TS$. In the physics literature, coisotropic submanifolds appear sometimes under the name of "first-class constraints." The following are equivalent:

1. $S$ is coisotropic;
2. if $f \in C^\infty(M)$ satisfies $f|_S \equiv 0$, then $X_f|_S \in \mathfrak{X}(S)$;
3. for any $s \in S$, any open neighborhood $U_s$ of $s$ in $M$, and any function $g \in C^\infty(U_s)$ such that $X_g(s) \in T_sS$, if $f \in C^\infty(U_s)$ satisfies $\{f, g\}(s) = 0$, it follows that $X_f(s) \in T_sS$;
4. the subalgebra $\{f \in C^\infty(M) \mid f|_S \equiv 0\}$ is a Poisson subalgebra of $(C^\infty(M), \{\cdot, \cdot\})$.

The following proposition shows how to endow the coisotropic submanifolds of a Poisson manifold with a Poisson structure by using the reduction theorem 1.

**Proposition 1** *Let $(M, \{\cdot, \cdot\})$ be a Poisson manifold with associated Poisson tensor $B \in \Lambda^2(T^*M)$. Let $S$ be an embedded coisotropic submanifold of $M$ and $D := B^\sharp((TS)^\circ)$. Then*

(i) *$D = D \cap TS = D_S$ is a smooth generalized distribution on $S$.*
(ii) *$D$ is integrable.*
(iii) *If $C^\infty_{S/D_S}$ has the $(D, D_S)$-local extension property, then $(M, \{\cdot, \cdot\}, D, S)$ is Poisson reducible.*

Coisotropic submanifolds usually appear as the level sets of integrals in involution. Let $(M, \{\cdot, \cdot\})$ be a Poisson manifold with Poisson tensor $B$ and let $f_1, \ldots, f_k \in C^\infty(M)$ be $k$ smooth functions in involution, that is, $\{f_i, f_j\} = 0$, for any $i, j \in \{1, \ldots, k\}$. Assume that $\mathbf{0} \in \mathbb{R}^k$ is a regular value of the function $F := (f_1, \ldots, f_k) : M \to \mathbb{R}^k$ and let $S := F^{-1}(0)$. Since for any $s \in S$, $\operatorname{span}\{\mathbf{d}f_1(s), \ldots, \mathbf{d}f_k(s)\} \subset (T_sS)^\circ$ and the dimensions of both sides of this inclusion are equal, it follows that $\operatorname{span}\{\mathbf{d}f_1(s), \ldots, \mathbf{d}f_k(s)\} = (T_sS)^\circ$. Hence, $B^\sharp(s)((T_sS)^\circ) = \operatorname{span}\{X_{f_1}(s), \ldots, X_{f_k}(s)\}$ and $B^\sharp(s)\ ((T_sS)^\circ) \subset T_sS$ by the involutivity of the components of $F$. Consequently, $S$ is a coisotropic submanifold of $(M, \{\cdot, \cdot\})$.

### Cosymplectic Submanifolds and Dirac's Constraints Formula

The Poisson reduction theorem 2 allows us to define Poisson structures on certain embedded submanifolds that are not Poisson submanifolds.

**Definition 6** Let $(M, \{\cdot, \cdot\})$ be a Poisson manifold and let $B \in \Lambda^2(T^*M)$ be the corresponding Poisson tensor. An embedded submanifold $S \subset M$ is called cosymplectic if

(i) $B^\sharp((TS)^\circ) \cap TS = \{0\}$,
(ii) $T_sS + T_s\mathcal{L}_s = T_sM$,

for any $s \in S$ and $\mathcal{L}_s$ the symplectic leaf of $(M, \{\cdot, \cdot\})$ containing $s \in S$.

The cosymplectic submanifolds of a symplectic manifold $(M, \omega)$ are its symplectic submanifolds. Cosymplectic submanifolds appear in the physics literature under the name of "second-class constraints."

**Proposition 2** *Let $(M, \{\cdot, \cdot\})$ be a Poisson manifold, $B \in \Lambda^2(T^*M)$ the corresponding Poisson tensor, and $S$ a cosymplectic submanifold of $M$. then, for any $s \in S$,*

(i) *$T_s\mathcal{L}_s = (T_sS \cap T_s\mathcal{L}_s) \oplus B^\sharp(s)((T_sS)^\circ)$, where $\mathcal{L}_s$ is the symplectic leaf of $(M, \{\cdot, \cdot\})$ that contains $s \in S$.*
(ii) *$(T_sS)^\circ \cap \ker B^\sharp(s) = \{0\}$.*
(iii) *$T_sM = B^\sharp(s)((T_sS)^\circ) \oplus T_sS$.*
(iv) *$B^\sharp((TS)^\circ)$ is a sub-bundle of $TM|_S$ and hence $TM|_S = B^\sharp((TS)^\circ) \oplus TS$.*
(v) *The symplectic leaves of $(M, \{\cdot, \cdot\})$ intersect $S$ transversely and hence $S \cap \mathcal{L}$ is an initial submanifold of $S$, for any symplectic leaf $\mathcal{L}$ of $(M, \{\cdot, \cdot\})$.*

**Theorem 3** (The Poisson structure of a cosymplectic submanifold). *Let $(M, \{\cdot, \cdot\})$ be a Poisson manifold, $B \in \Lambda^2(T^*M)$ the corresponding Poisson tensor, and $S$ a cosymplectic submanifold of $M$. Let $D := B^\sharp\ ((TS)^\circ) \subset TM|_S$. Then,*

(i) *$(M, \{\cdot, \cdot\}, D, S)$ is Poisson reducible.*
(ii) *The corresponding quotient manifold equals $S$ and the reduced bracket $\{\cdot, \cdot\}^S$ is given by*

$$\{f, g\}^S(s) = \{F, G\}(s) \qquad [5]$$

*where $f, g \in C^\infty_{S,M}(V)$ are arbitrary and $F, G \in C^\infty(U)$ are local $D$-invariant extensions of $f$ and $g$ around $s \in S$, respectively.*
(iii) *The Hamiltonian vector field $X_f$ of an arbitrary function $f \in C^\infty_{S,M}(V)$ is given either by*

$$Ti \circ X_f = X_F \circ i \qquad [6]$$

*where $F \in C^\infty(U)$ is a local $D$-invariant extension of $f$ and $i : S \hookrightarrow M$ is the inclusion, or by*

$$Ti \circ X_f = \pi_S \circ X_{\overline{F}} \circ i \qquad [7]$$

*where $\overline{F} \in C^\infty(U)$ is an arbitrary local extension of $f$ and $\pi_S : TM|_S \to TS$ is the projection induced by the Whitney sum decomposition $TM|_S = B^\sharp((TS)^\circ) \oplus TS$ of $TM|_S$.*
(iv) *The symplectic leaves of $(S, \{\cdot, \cdot\}^S)$ are the connected components of the intersections $S \cap \mathcal{L}$, where $\mathcal{L}$ is a symplectic leaf of $(M, \{\cdot, \cdot\})$. Any*

*symplectic leaf of $(S,\{\cdot,\cdot\}^S)$ is a symplectic submanifold of the symplectic leaf of $(M,\{\cdot,\cdot\})$ that contains it.*

(v) *Let $\mathcal{L}_s$ and $\mathcal{L}_s^S$ be the symplectic leaves of $(M,\{\cdot,\cdot\})$ and $(S,\{\cdot,\cdot\}^S)$, respectively, that contain the point $s\in S$. Let $\omega_{\mathcal{L}_s}$ and $\omega_{\mathcal{L}_s^S}$ be the corresponding symplectic forms. Then $B^\sharp(s)((T_sS)^\circ)$ is a symplectic subspace of $T_s\mathcal{L}_s$ and*

$$B^\sharp(s)((T_sS)^\circ) = \left(T_s\mathcal{L}_s^S\right)^{\omega_{\mathcal{L}_s}(s)} \qquad [8]$$

*where $(T_s\mathcal{L}_s^S)^{\omega_{\mathcal{L}_s}(s)}$ denotes the $\omega_{\mathcal{L}_s}(s)$-orthogonal complement of $T_s\mathcal{L}_s^S$ in $T_s\mathcal{L}_s$.*

(vi) *Let $B_S\in\Lambda^2(T^*S)$ be the Poisson tensor associated to $(S,\{\cdot,\cdot\}^S)$. Then*

$$B_S^\sharp = \pi_S\circ B^\sharp|_S\circ\pi_S^* \qquad [9]$$

*where $\pi_S^*: T^*S\to T^*M|_S$ is the dual of $\pi_S: TM|_S\to TS$.*

The "Dirac constraints formula" is the expression in coordinates for the bracket of a cosymplectic submanifold. Let $(M,\{\cdot,\cdot\})$ be an $n$-dimensional Poisson manifold and let $S$ be a $k$-dimensional cosymplectic submanifold of $M$. Let $z_0$ be an arbitrary point in $S$ and $(U,\overline{\kappa})$ a submanifold chart around $z_0$ such that $\overline{\kappa}=(\overline{\varphi},\overline{\psi}): U\to V_1\times V_2$, where $V_1$ and $V_2$ are two open neighborhoods of the origin in two Euclidean spaces such that $\overline{\kappa}(z_0)=(\overline{\varphi}(z_0),\overline{\psi}(z_0))=(0,0)$ and

$$\overline{\kappa}(U\cap S) = V_1\times\{0\} \qquad [10]$$

Let $\overline{\varphi}=:(\overline{\varphi}^1,\ldots,\overline{\varphi}^k)$ be the components of $\overline{\varphi}$ and define $\widehat{\varphi}^1:=\overline{\varphi}^1|_{U\cap S},\ldots,\widehat{\varphi}^k:=\overline{\varphi}^k|_{U\cap S}$. Extend $\widehat{\varphi}^1,\ldots,\widehat{\varphi}^k$ to $D$-invariant functions $\varphi^1,\ldots,\varphi^k$ on $U$. Since the differentials $\boldsymbol{d}\widehat{\varphi}^1(s),\ldots,\boldsymbol{d}\widehat{\varphi}^k(s)$ are linearly independent for any $s\in U\cap S$, we can assume (by shrinking $U$ if necessary) that $\boldsymbol{d}\varphi^1(z),\ldots,\boldsymbol{d}\varphi^k(z)$ are also linearly independent for any $z\in U$. Consequently, $(U,\kappa)$ with $\kappa:=(\varphi^1,\ldots,\varphi^k,\psi^1,\ldots,\psi^{n-k})$ is a submanifold chart for $M$ around $z_0$ with respect to $S$ such that, by construction,

$$\boldsymbol{d}\varphi^1(s)|_{B^\#(s)((T_sS)^\circ)} = \cdots = \boldsymbol{d}\varphi^k(s)|_{B^\#(s)((T_sS)^\circ)} = 0$$

for any $s\in U\cap S$. This implies that for any $i\in\{1,\ldots,k\}$, $j\in\{1,\ldots,n-k\}$, and $s\in S$

$$\{\varphi^i,\psi^j\}(s) = \boldsymbol{d}\varphi^i(s)\big(X_{\psi^j}(s)\big) = 0$$

since $\boldsymbol{d}\psi^j(s)\in(T_sS)^\circ$ by [10] and hence

$$X_{\psi^j}(s)\in B^\#(s)((T_sS)^\circ) \qquad [11]$$

Additionally, since the functions $\varphi^1,\ldots,\varphi^k$ are $D$-invariant, by [6], it follows that

$$X_{\varphi^1}(s) = X_{\widehat{\varphi}^1}(s)\in T_sS,\ldots,X_{\varphi^k}(s) = X_{\widehat{\varphi}^k}(s)\in T_sS$$

for any $s\in S$. Consequently, $\{X_{\varphi^1}(s),\ldots,X_{\varphi^k}(s), X_{\psi^1}(s),\ldots,X_{\psi^{n-k}}(s)\}$ spans $T_s\mathcal{L}_s$ with

$$\{X_{\varphi^1}(s),\ldots,X_{\varphi^k}(s)\}\subset T_sS\cap T_s\mathcal{L}_s$$

and

$$\{X_{\psi^1}(s),\ldots,X_{\psi^{n-k}}(s)\}\subset B^\#(s)((T_sS)^\circ)$$

By Proposition 2(i),

$$\operatorname{span}\{X_{\varphi^1}(s),\ldots,X_{\varphi^k}(s)\} = T_sS\cap T_s\mathcal{L}_s$$

and

$$\operatorname{span}\{X_{\psi^1}(s),\ldots,X_{\psi^{n-k}}(s)\} = B^\#(s)((T_sS)^\circ)$$

Since $\dim(B^\#(s)((T_sS)^\circ))=n-k$ by Proposition 2(iii), it follows that $\{X_{\psi^1}(s),\ldots,X_{\psi^{n-k}}(s)\}$ is a basis of $B^\#(s)((T_sS)^\circ)$.

Since $B^\#(s)((T_sS)^\circ)$ is a symplectic subspace of $T_s\mathcal{L}_s$ by Theorem 3(v), there exists some $r\in\mathbb{N}$ such that $n-k=2r$ and, additionally, the matrix $C(s)$ with entries

$$C^{ij}(s) := \{\psi^i,\psi^j\}(s),\quad i,j\in\{1,\ldots,n-k\}$$

is invertible. Therefore, in the coordinates $(\varphi^1,\ldots,\varphi^k,\psi^1,\ldots,\psi^{n-k})$, the matrix associated to the Poisson tensor $B(s)$ is

$$B(s) = \begin{pmatrix} B_S(s) & 0 \\ 0 & C(s)\end{pmatrix}$$

where $B_S\in\Lambda^2(T^*S)$ is the Poisson tensor associated to $(S,\{\cdot,\cdot\}^S)$. Let $C_{ij}(s)$ be the entries of the matrix $C(s)^{-1}$.

**Proposition 3** (Dirac formulas). *In the coordinate neighborhood $(\varphi^1,\ldots,\varphi^k,\psi^1,\ldots,\psi^{n-k})$ constructed above and for $s\in S$ we have, for any $f,g\in C^\infty_{S,M}(V)$:*

$$X_f(s) = X_F(s) - \sum_{i,j=1}^{n-k}\{F,\psi^i\}(s)C_{ij}(s)X_{\psi^j}(s) \qquad [12]$$

and

$$\{f,g\}^S(s) = \{F,G\}(s) - \sum_{i,j=1}^{n-k}\{F,\psi^i\}(s)C_{ij}(s)\{\psi^j,G\}(s) \qquad [13]$$

*where $F,G\in C^\infty(U)$ are arbitrary local extensions of $f$ and $g$, respectively, around $s\in S$.*

*See also:* Classical *r*-Matrices, Lie Bialgebras, and Poisson Lie Groups; Cotangent Bundle Reduction; Graded Poisson Algebras; Symmetry and Symplectic Reduction; Hamiltonian Group Actions; Lie, Symplectic, and Poisson Groupoids and their Lie Algebroids; Singularity and Bifurcation Theory.

## Further Reading


Abraham R and Marsden JE (1978) *Foundations of Mechanics*, 2nd edn. Reading, MA: Addison–Wesley.

Casati P and Pedroni M (1992) Drinfeld–Sokolov reduction on a simple Lie algebra from the bi-Hamiltonian point of view. *Letters in Mathematical Physics* 25(2): 89–101.

Castrillón-López M and Marsden JE (2003) Some remarks on Lagrangian and Poisson reduction for field theories. *Journal of Geometry and Physics* 48: 52–83.

Cendra H, Marsden JE, and Ratiu TS (2003) Cocycles, compatibility, and Poisson brackets for complex fluids. In: Capriz G and Mariano P (eds.) *Advances in Multifield Theories of Continua with Substructures*, Memoirs, pp. 51–73. Aarhus: Aarhus Univ.

Faybusovich L (1991) Hamiltonian structure of dynamical systems which solve linear programming problems. *Physica D* 53: 217–232.

Faybusovich L (1995) A Hamiltonian structure for generalized affine-scaling vector fields. *Journal of Nonlinear Science* 5(1): 11–28.

Gotay MJ, Nester MJ, and Hinds G (1978) Presymplectic manifolds and the Dirac–Bergmann theory of constraints. *Journal of Mathematical Physics* 19: 2388–2399.

Krishnaprasad PS and Marsden JE (1987) Hamiltonian structure and stability for rigid bodies with flexible attachments. *Archives for Rational and Mechanical Analysis* 98: 137–158.

Lewis D, Marsden JE, Montgomery R, and Ratiu TS (1986) The Hamiltonian structure for dynamic free boundary problems. *Physica D* 18: 391–404.

Lu J-H and Weinstein A (1990) Poisson Lie groups, dressing transformations and Bruhat decompositions. *Journal of Differential Geometry* 31: 510–526.

Marsden JE and Ratiu TS (1986) Reduction of Poisson manifolds. *Letters in Mathematical Physics* 11: 161–169.

Marsden JE and Ratiu TS (2003) *Introduction to Mechanics and Symmetry*, second printing; 1st edn. (1994), Texts in Applied Mathematics, 2nd edn., vol. 17. New York: Springer.

Ortega J-P and Ratiu TS (1998) Singular reduction of Poisson manifolds. *Letters in Mathematical Physics* 46: 359–372.

Ortega J-P and Ratiu TS (2003) *Momentum Maps and Hamiltonian Reduction*. Progress in Math. vol. 222. Boston: Birkhäuser.

Pedroni M (1995) Equivalence of the Drinfeld–Sokolov reduction to a bi-Hamiltonian reduction. *Letters in Mathematical Physics,* 35(4): 291–302.

Sundermeyer K (1982) *Constrained Dynamics*. Lecture Notes in Physics, vol. 169. New York: Springer.

Weinstein A (1983) The local structure of Poisson manifolds. *Journal of Differential Geometry* 18: 523–557.

Weinstein A (1985) The local structure of Poisson manifolds – errata and addenda. *Journal of Differential Geometry* 22(2): 255.

Zaalani N (1999) Phase space reduction and Poisson structure. *Journal of Mathematical Physics* 40: 3431–3438.


# Polygonal Billiards

**S Tabachnikov**, Pennsylvania State University, University Park, PA, USA



## Mechanical Examples. Unfolding Billiard Trajectories

The billiard system inside a polygon $P$ has a very simple description: a point moves rectilinearly with the unit speed until it hits a side of $P$; there it instantaneously changes its velocity according to the rule "the angle of incidence equals the angle of reflection," and continues the rectilinear motion. If the point hits a corner, its further motion is not defined. (*see* Billiards in Bounded Convex Domains). From the point of view of the theory of dynamical systems, polygonal billiards provide an example of parabolic dynamics in which nearby trajectories diverge with subexponential rate.

One of the motivations for the study of polygonal billiards comes from the mechanics of elastic particles in dimension 1. For example, consider the system of two point-masses $m_1$ and $m_2$ on the positive half-line $x \geq 0$. The collision between the points is elastic, that is, the energy and momentum are conserved. The reflection off the left endpoint of the half-line is also elastic: if a point hits the "wall" $x = 0$, its velocity changes sign. The configuration space of this system is the wedge $0 \leq x_1 \leq x_2$. After the rescaling $\bar{x}_i = \sqrt{m_i} x_i$, $i = 1, 2$, this system identifies with the billiard inside a wedge with the angle measure $\arctan \sqrt{m_1/m_2}$.

Likewise, the system of two elastic point-masses on a segment is the billiard system in a right triangle; a system of a number of elastic point-masses on the positive half-line or a segment is the billiard inside a multidimensional polyhedral cone or a polyhedron, respectively. The system of three elastic point-masses on a circle has three degrees of freedom; one can reduce one by assuming that the center of mass of the system is fixed. The resulting two-dimensional system is the billiard inside an acute triangle with the angles

$$\arctan\left(m_i\sqrt{\frac{m_1+m_2+m_3}{m_1 m_2 m_3}}\right), \quad i = 1, 2, 3$$

For comparison, the more realistic system of elastic balls identifies with the billiard system in a domain with nonflat boundary components.

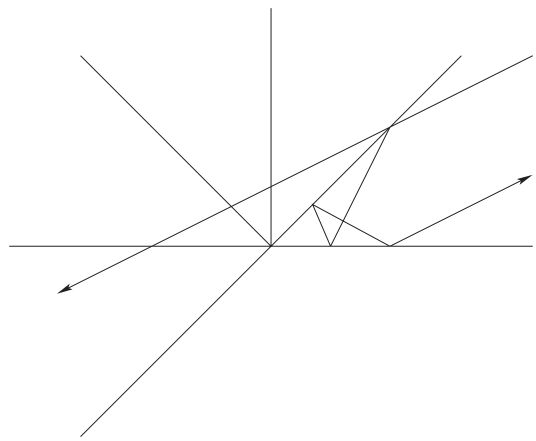


**Figure 1** Unfolding a billiard trajectory in a wedge.

A useful elementary method of study is unfolding: instead of reflecting the billiard trajectory in the sides of the polygon, reflect the polygon in the respective side and unfold the billiard trajectory to a straight line. This method yields an upper bound

$$\left\lceil \frac{\pi}{\arctan \sqrt{m_1/m_2}} \right\rceil$$

for the number of collisions in the system of two point-masses $m_1$ and $m_2$ on the positive half-line. Likewise, the number of collisions for any number of elastic point-masses on the positive half-line is bounded above by a constant depending on the masses only. Similar results are known for systems of elastic balls (Figure 1).

Similarly, one studies the billiard inside the unit square. Unfolding the square yields a square grid in the plane, acted upon by the group of parallel translations $2\mathbf{Z} \oplus 2\mathbf{Z}$. Factorizing by this group action yields a torus, and the billiard flow in a given direction becomes a constant flow on the torus. If the slope is rational, then all orbits are periodic, and if the slope is irrational, then all orbits are dense and the billiard flow is ergodic. Its metric entropy is equal to zero. Periodic trajectories of the billiard in a square come in bands of parallel ones. Let $f(\ell)$ be the number of such bands of length not greater than $\ell$. Then, $f(\ell)$ equals the number of coprime lattice points inside the circle of radius $\ell$, that is, $f(\ell)$ has quadratic growth in $\ell$.

## Periodic Trajectories

The simplest example of a periodic orbit in a polygonal billiard is the 3-periodic Fangano trajectory in an acute triangle: it connects the bases of the three altitudes of the triangle and has minimal perimeter among inscribed triangles. The Fagnano trajectory belongs to a band of 6-periodic ones. It is not known whether every acute triangle has other periodic trajectories.

For a right triangle, one has the following result: almost every (in the sense of the Lebesgue measure) billiard trajectory that leaves a leg in the perpendicular direction returns to the same leg in the same direction and is therefore periodic. A similar existence result holds for polygons whose sides have only two directions.

In general, not much is known about the existence of periodic billiard trajectories in polygons. Conjecturally, every polygon has one, but this is not known even for all obtuse triangles. Recently, R Schwartz proved that every obtuse triangle with the angles not exceeding 100° has a periodic billiard path. This work substantially relies on a computer program, McBilliards, written by Schwartz and Hooper.

If an arbitrary small perturbation of the vertices of a billiard polygon leads to a perturbation of a periodic billiard trajectory, but not to its destruction, then this trajectory is called stable. Label the sides of the polygon $1, 2, \ldots, k$. Then a periodic trajectory is coded by the word consisting of the labels of the consecutively visited sides. An even-periodic trajectory is stable if and only if the numbers in the respective word can be partitioned in pairs of equal numbers, so that the number from each pair appears once at an even position, and once at an odd one. As a consequence, if the angles of a polygon are independent over the rational numbers, then every periodic billiard trajectory in it is stable.

## Complexity of Billiard Trajectories

The encoding of billiard trajectories by the consecutively visited sides of the billiard polygon provides a link between billiard and symbolic dynamics. For a billiard $k$-gon $P$, denote by $\Sigma$ the set of words in letters $1, 2, \ldots, k$ corresponding to billiard trajectories in $P$, and let $\Sigma_n$ be the set of such words of length $n$.

One has a general theorem: *the topological entropy of the billiard flow is zero*. This implies that a number of quantities, associated with a polygonal billiard, grow slower than exponentially, as functions of $n$: the cardinality $|\Sigma_n|$, the number of strips of $n$-periodic trajectories, the number of generalized diagonals with $n$ links (i.e., billiard trajectories that start and end at corners of the billiard polygon), etc. Conjecturally, all these quantities have polynomial growth in $n$.

The complexity of the billiard in a polygon is defined as the function $p(n) = |\Sigma_n|$. Likewise, one may consider the billiard trajectories in a given direction $\theta$ and define the corresponding complexity $p_\theta(n)$.

In the case of a square, one modifies the encoding using only two symbols, say, 0 and 1, to indicate that a trajectory reflects in a horizontal or a vertical side, respectively. If $\theta$ is a direction with an irrational slope, then $p_\theta(n) = n + 1$. This is a classical result by Hedlund and Morse. The sequences with complexity $p(n) = n + 1$ are called Sturmian; this is the smallest complexity of aperiodic sequences. A generalization for multidimensional cubes and parallelepipeds, due to Yu Baryshnikov, is known.

For a $k$-gon $P$, let $N$ be the least common denominator of its $\pi$-rational angles and $s$ be the number of its distinct $\pi$-irrational angles. Then,

$$p_\theta(n) \leq kNn\left(1 + \frac{n}{2}\right)^s$$

Concerning billiard trajectories in all directions, one has a lower bound for complexity: $p(n) \geq cn^2$ for a constant $c$ depending on the polygon. A similar estimate holds for a $d$-dimensional polyhedron with the exponent 2 replaced by $d$.

## Rational Polygons and Flat Surfaces

The only class of polygons for which the billiard dynamics is well understood are rational one, the polygons satisfying the property that the angles between all pairs of sides are rational multiples of $\pi$.

Let $P$ be a simply connected (without holes) rational $k$-gon with angles $\pi m_i/n_i$, where $m_i$ and $n_i$ are coprime integers. The reflections in the sides of $P$ generate a subgroup of the group of isometries of the plane. Let $G(P) \subset \mathrm{O}(2)$ consist of the linear parts of the elements of this group. Then, $G(P)$ is the dihedral group $D_N$ consisting of $2N$ elements. When a billiard trajectory reflects in a side of $P$, its direction changes by the action of the group $G(P)$, and the orbit of a generic direction $\theta \neq k\pi/N$ on the unit circle consists of $2N$ points.

The phase space of the billiard flow is the unit tangent bundle $P \times S^1$. Let $M_\theta$ be the subset of points whose projection to $S^1$ belongs to the orbit of $\theta$ under $G(P) = D_N$. Then, $M_\theta$ is an invariant surface of the billiard flow in $P$. The surface $M_\theta$ is obtained from $2N$ copies of $P$ by gluing their sides according to the action of $D_N$. This oriented compact surface depends only on the polygon $P$, but not on the choice of $\theta$, and may be denoted by $M$. The directional billiard flows $F_\theta$ on $M$ in directions $\theta$

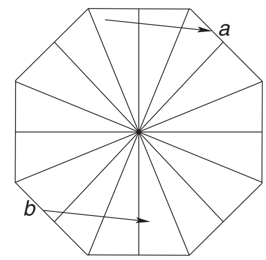



**Figure 2** The invariant surface for a right triangle with acute angle $\pi/8$ has genus 2.

are obtained, one from another, by rotations. The genus of $M$ is given by the formula

$$1 + \frac{N}{2}\left(k - 2 - \sum \frac{1}{n_i}\right)$$

For example, if $P$ is a right triangle with an acute angle $\pi/8$, then $M$ is a surface of genus 2 (Figure 2).

The cases when $M$ is a torus are as follows: the angles of $P$ are all of the form $\pi/n_i$, where $n_i$ are equal, up to permutations, to

$$(3, 3, 3), (2, 4, 4), (2, 3, 6), (2, 2, 2, 2)$$

and the respective polygons are an equilateral triangle, an isosceles right triangle, a right triangle with an acute angle $\pi/6$, and a square. All these polygons tile the plane.

The billiard flow on the surface $M$ has saddle singularities at the points obtained from the vertices of $P$. The surface $M$ inherits a flat metric from $P$ with a finite number of cone-type singularities, corresponding to the vertices of $P$, with cone angles multiples of $2\pi$ (Figure 3).

A flat surface $M$ is a compact smooth surface with a distinguished finite set of points $\Sigma$. On $M \setminus \Sigma$, one has coordinate charts $v = (x, y)$ such that the transition functions on the overlaps are of the form

$$v \to v + c \quad \text{or} \quad v \to -v + c$$

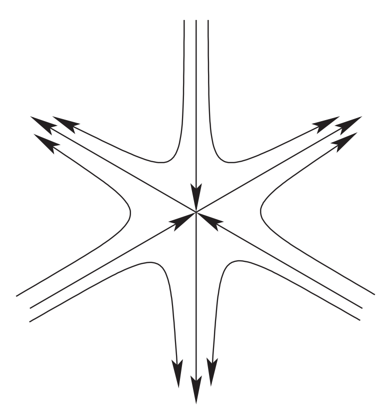


**Figure 3** A cone singularity for the flow on an invariant surface.

In particular, one may talk about directions on a flat surface.

The group PSL$(2, \boldsymbol{R})$ acts on the space of flat structures. From the point of view of complex analysis, a flat surface is a Riemann surface with a holomorphic quadratic differential; the set of cone points $\Sigma$ corresponds to the zeros of the quadratic differential. Not every flat surface is associated with a polygonal billiard.

Concerning ergodicity, one has the theorem of Kerckhoff, Masur, and Smillie: given a flat surface of genus not less than 2, for almost all directions $\theta$ (in the sense of the Lebesgue measure), the flow $F_\theta$ is uniquely ergodic. Furthermore, the Hausdorff dimension of the set of angles $\theta$ for which ergodicity fails does not exceed 1/2, and this bound is sharp. As a consequence, the billiard flow on the invariant surface is uniquely ergodic for almost all directions. Another corollary: *there is a dense $G_\delta$ subset in the space of polygons consisting of polygons for which the billiard flow is ergodic. If a billiard polygon admits approximation by rational polygons at a superexponentially fast rate, then the billiard flow in it is ergodic.*

Concerning periodic orbits, one has the following theorem due to H Masur: given a flat surface of genus not less than 2, there exists a dense set of angles $\theta$ such that $F_\theta$ has a closed trajectory. As a consequence, for any rational billiard polygon, there is a dense set of directions each with a periodic orbit. Furthermore, periodic points are dense in the phase space of the billiard flow in a rational polygon.

Similarly to the case of a square, let $f(\ell)$ be the number of strips of periodic trajectories of length not greater than $\ell$ in a rational polygon $P$. By a theorem of H Masur, there exist constants $c$ and $C$ such that for sufficiently large $\ell$ one has: $c\ell^2 < f(\ell) < C\ell^2$, and likewise for flat surfaces.

There is a class of flat surfaces, called Veech (or lattice) surfaces, for which more refined results are available. The groups of affine transformations of a flat surface determine a subgroup in $SL(2, \boldsymbol{R})$. If this subgroup is a lattice in $SL(2, \boldsymbol{R})$, then the flat surface is called a Veech surface. Similarly, one defines a Veech rational polygon. For example, regular polygons and isosceles triangles with equal angles $\pi/n$ are Veech. All acute Veech triangles are described.

For a Veech surface, one has the following Veech dichotomy: for any direction $\theta$, either the flow $F_\theta$ is minimal or its every leaf is closed (unless it is a saddle connection, i.e., a segment connecting cone points). For a Veech surface (and polygon), the quadratic bounds for the counting function $f(\ell)$ become quadratic asymptotics: $f(\ell)/\ell^2$ has a limit as $\ell \to \infty$. The value of this limit is expressed in arithmetical terms.

A generic flat surface also has quadratic asymptotics. The value of the limit depends only on the stratum of the Teichmuller space that contains this surface. These values are known, due to Eskin, Masur, Okunkov, and Zorich. Since a generic flat surface does not correspond to a rational polygon, this result does not immediately apply to polygonal billiards. However, quadratic asymptotics are established for rectangular billiards with barriers.

Note, in conclusion, a close relation of billiards in rational polygons and interval exchange transformations; the reduction of the former to the latter is a particular case of the reduction of the billiard flow to the billiard ball map. On an invariant surface $M$ of the billiard flow, consider a segment $I$, perpendicular to the directional flow. Since "the width of a beam" is an invariant transversal measure for the constant flow, the first return map to $I$ is a piecewise orientation preserving isometry, that is, an interval exchange transformation.

## Acknowledgment

This work was partially supported by NSF.

*See also:* Billiards in Bounded Convex Domains; Ergodic Theory; Fractal Dimensions in Dynamics; Generic Properties of Dynamical Systems; Holomorphic Dynamics; Hyperbolic Billiards; Riemann Surfaces.

## Further Reading

Burago D, Ferleger S, and Kononenko A (2000) *A Geometric Approach to Semi-Dispersing Billiards. Hard Ball Systems and the Lorentz Gas*, pp. 9–27. Berlin: Springer.

Chernov N and Markarian R *Theory of Chaotic Billiards* (to appear).

Galperin G, Stepin A, and Vorobets Ya (1992) Periodic billiard trajectories in polygons: generating mechanisms. *Russian Mathematical Surveys* 47(3): 5–80.

Gutkin E (1986) Billiards in polygons. *Physica D* 19: 311–333.

Gutkin E (1996) Billiards in polygons: survey of recent results. *Journal of Statistical Physics* 83: 7–26.

Gutkin E (2003) Billiard dynamics: a survey with the emphasis on open problems. *Regular and Chaotic Dynamics* 8: 1–13.

Katok A and Hasselblatt B (1995) *Introduction to the Modern Theory of Dynamical Systems*. Cambridge: Cambridge University Press.

Kozlov V and Treshchev D (1991) *Billiards. A Genetic Introduction to the Dynamics of Systems with Impacts*. Providence: American Mathematical Society.

Masur H and Tabachnikov S (2002) *Rational Billiards and Flat Structures. Handbook of Dynamical Systems*, vol. 1A, pp. 1015–1089. Amsterdam: North-Holland.

Sinai Ya (1976) *Introduction to Ergodic Theory*. Princeton: Princeton University Press.

Smillie J (2000) *The Dynamics of Billiard Flows in Rational Polygons*, Encyclopaedia of Mathematical Sciences, vol. 100, pp. 360–382. Berlin: Springer.

Tabachnikov S (1995) *Billiards, Société Math*. de France, Panoramas et Syntheses, No 1.

Tabachnikov S (2005) *Geometry and Billiards*. American Mathematical Society.

# Positive Maps on C*-Algebras

**F Cipriani**, Politecnico di Milano, Milan, Italy



## Introduction

The theme of positive maps on $^*$-algebras and other ordered vector spaces, dates back to the Perron–Frobenius theory of matrices with positive entries, the Shur's product of matrices, the study of doubly stochastic matrices describing discrete-time random walks and the behavior of limits of powers of positive matrices in ergodic theory.

A long experience proved that far-reaching generalizations of the above situations have to be considered in various fields of mathematical physics and that $C^*$-algebras, their positive cones, and other associated ordered vector spaces provide a rich unifying framework of functional analysis to treat them.

It is the scope of this note to review some of the basic aspects both of the general theory and of the applications.

In the next section we briefly recall the definitions of $C^*$-algebras and their positive cones. However, throughout this article we refer to C*-Algebras and their Classification and von Neumann Algebras: Introduction, Modular Theory and Classification Theory as sources of the definitions and general properties of the objects of these operator algebras. We then introduce positive maps, illustrate their general properties, and discuss some relevant classes of them. The correspondence between states and representations is described next, as well as the appearance of vector, normal and non-normal states in applications. We then illustrate the structure of completely positive maps and their relevance in mathematical physics. Finally, we describe the relevance of the class of completely positive maps to understand the structure of nuclear $C^*$-algebras.

## Positive Cones in $C^*$-Algebras

A $C^*$-algebra $A$ is a complex Banach algebra with a conjugate-linear involution $a \mapsto a^*$ such that $\|a^*a\| = \|a\|^2$ for all $a \in A$.

When $A$ has a unit $1_A$, the spectrum $\mathrm{Sp}(a)$ of an element $a$ is the subset of all complex numbers $\lambda$ such that $a - \lambda \cdot 1_A$ is not invertible in $A$. When $A$ is realized as a subalgebra of some $\mathcal{B}(\mathcal{H})$, and this is always possible, the set $\mathrm{Sp}(a)$ coincides with the spectrum of the bounded operator $a$ on the Hilbert space $\mathcal{H}$.

The involution determines the self-adjoint part $A_h := \{a \in A: a = a^*\}$ of $A$, a real subspace such that $A = A_h + \mathrm{i}A_h$. A self-adjoint element $a$ of $A$ satisfies $\mathrm{Sp}(a) \subseteq \mathbb{R}$ and, if $k \geq 0$, one has $\|a\| \leq k$ if and only if $\mathrm{Sp}(a) \subseteq [-k, k]$.

The involution determines another important subset of $A$: $A_+ := \{a^*a: a \in A\}$. This subset of $A_h$ is closed in the norm topology of $A$ and contains the sums of its elements as well as their multiples by positive scalars: in other words, it is a closed convex cone. From a spectral point of view, one has the following characterization: a self-adjoint element $a$ belongs to $A_+$ if and only if its spectrum is positive $\mathrm{Sp}(a) \subset [0, +\infty)$. It is this property that allows us to call $A_+$ the positive cone of $A$ and its elements positive. If it exists, a unit $1_A$ in $A$ is always positive and a Hermitian element $a$ is positive if and only if $\|1_A - a/\|a\|\,\| \leq 1$.

The continuous functional calculus in $A$ allows to write any self-adjoint element of $A_h$ as a difference of elements of $A_+$: $A_h = A_+ - A_+$. Moreover, $A_+ \cap (-A_+) = \{0\}$ and the decomposition $a = b - c$ of a self-adjoint element $a$ as difference of positive elements $b$ and $c$ is unique provided one requires that $bc = cb = 0$. In this case, it is called the orthogonal decomposition.

The cone $A_+$ determines an underlying structure of order space on $A$: for $a, b \in A$ one says that $a$ is less than or equal to $b$, in symbols $a \leq b$, if and only if $b - a \in A_+$. In particular, $a \geq 0$ just means that $a$ is positive.

Another fundamental characterization of the positive cone is the following: a self-adjoint element $a = a^*$ is positive if and only if there exists an element $b$ in $A$ such that $a = b^2$. Moreover, among the elements $b$ with this property, there exists one and only one which is positive, the square root of $a$. Some examples of positive cones are provided in the following.

**Example 1** By a fundamental result of I M Gelfand, a commutative $C^*$-algebra $A$ is isomorphic to the $C^*$-algebra $C_0(X)$ of all complex continuous functions vanishing at infinity on a locally compact Hausdorff topological space $X$. The algebraic operations have the usual pointwise meaning and the norm is the uniform one. The constant function 1 represents the unit precisely when $X$ is compact. The positive cone $C_0(X)_+$ coincides with that of the positive continuous functions in $C_0(X)$.

**Example 2** Finite dimensional $C^*$-algebras $A$ are classified as finite sums $M_{n_1}(\mathbb{C}) \oplus M_{n_2}(\mathbb{C}) \oplus \cdots \oplus M_{n_k}(\mathbb{C})$ of full matrix algebras $M_{n_i}(\mathbb{C})$. An element

$a_1 \oplus a_2 \oplus \cdots \oplus a_k$ is positive if and only if the matrices $a_i$ have positive eigenvalues.

**Example 3** When a C*-algebra $A \subseteq \mathcal{B}(\mathcal{H})$ is represented as a self-adjoint closed algebra of operators on a Hilbert space $\mathcal{H}$, its positive elements are those which have non-negative spectrum.

## Positive Maps on $C^*$-Algebras

Among the various relevant classes of maps between $C^*$-algebras, we are going to consider the following ones, whose properties are connected with the underlying structures of ordered vector spaces.

**Definition 1** Given two C*-algebras $A$ and $B$, a map $\phi : A \longrightarrow B$ is called positive if $\phi(A_+) \subseteq B_+$. In other words, a map is positive if and only if it transforms the positive elements of $A$ into positive elements of $B$:

$$a \in A \Rightarrow \phi(a^*a) \in B_+ \qquad [1]$$

If $A$ and $B$ have units, the map is called unital provided $\phi(1_A) = 1_B$.

### Morphisms and Jordan Morphisms

A $^*$-morphism between C*-algebras $\phi : A \longrightarrow B$ is positive; in fact, $\phi(a^*a) = \phi(a)^*\phi(a) \geq 0$.

This also the case for Jordan $^*$-morphism, the linear maps satisfying $\phi(a^*) = \phi(a)^*$ and $\phi(\{a, b\}) = \{\phi(a), \phi(b)\}$, where $\{a, b\} = ab + ba$ denotes the Jordan product. In fact, if $a = a^*$ then $\phi(a^2) = \phi(a)^2$ is positive.

### Shur's Product of Matrices

Let $A \in M_n(\mathbb{C})$ be a positive matrix and define a linear map $\phi : M_n(\mathbb{C}) \longrightarrow M_n(\mathbb{C})$ through the Shur's product of matrices: $\phi_A(B) := [A_{ij}B_{ij}]_{i,j=1}^n$. Since the Shur's product of positive matrices is positive too (i.e., the positive cone of $M_n(\mathbb{C})$ is a semigroup under matrix product), the above map is positive.

### Positive-Definite Function on Groups

Positive maps also arise naturally in harmonic analysis. Let $G$ be a locally compact topological group with identity $e$ and left Haar's measure $m$. Let $p : G \longrightarrow \mathbb{C}$ be a continuous positive-definite function on $G$. This just means that for all $n \geq 1$ and all $s_1, \ldots, s_n \in G$, the matrix $\{p(s_i^{-1}s_j)\}_{i,j=1}^n$ belongs to the positive cone of $M_n(\mathbb{C})$: $\sum_{i,j=1}^n p(s_i^{-1}s_j)\overline{\alpha}_i\alpha_j \geq 0$ for all $\alpha_1, \ldots, \alpha_n$. Such functions are necessarily bounded with $\|p\|_\infty \leq p(e)$, so that an operator $\phi : L^1(G, m) \longrightarrow L^1(G, m)$ is well defined by pointwise multiplication: $\phi(f)(s) := p(s)f(s)$. This map extends to a positive map $\phi : C^*(G) \longrightarrow C^*(G)$, which is unital when $p(e) = 1$, on the full group $C^*$-algebra $C^*(G)$. When $G$ is amenable, this algebra coincides with reduced C*-algebra $C_r(G)$ so that, if $G$ is also unimodular (as is the case if $G$ is compact), the positive elements can be approximated by positive-definite functions in $L^1(G, m)$ and the positivity of $\phi$ follows exactly as in the previous example.

### Positive Maps in Commutative $C^*$-Algebras

Positive maps $\phi : C_0(Y) \longrightarrow C_0(X)$ between commutative $C^*$-algebras have the following structure: $\phi(a)(x) = \int_Y k(x, \mathrm{d}y)a(y), a \in C_0(Y)$. Here the kernel $x \mapsto k(x, \cdot)$ is a continuous map from $X$ to the space of positive Radon measures on $Y$. In case $X$ and $Y$ are compact, the map is unital provided $k(x, \cdot)$ is a probability measure for each $x \in X$. In fact, for a fixed $x \in X$, the map $a \mapsto \phi(a)(x)$ is a positive linear functional from $C_0(Y)$ to $\mathbb{C}$ and Riesz's theorem guarantees that it can be represented by a positive Radon measure on $Y$.

In probability theory, one-parameter semigroups $\phi_t \circ \phi_s = \phi_{t+s}$ of positive maps $\phi_t : C_0(X) \longrightarrow C_0(X)$ such that $\phi_t(1) \leq 1$ for all $t \geq 0$, are called Markovian semigroups (conservative, if the maps are unital). They represent the expectation at time $t > 0$ of Markovian stochastic processes on $X$. In this case, the time-dependent kernel $k(t, x, \cdot)$ represents the distribution probability at time $t$ of a particle starting in $x \in X$ at time $t = 0$.

These kinds of maps arise also in potential theory, where the dependence of the solution $\phi(a)$ of a Dirichlet problem on a bounded domain $\Omega$, with nice boundary $\partial\Omega$, upon the continuous boundary data $a \in C(\partial\Omega)$ gives rise to a linear unital map $\phi : C(\partial\Omega) \longrightarrow C(\Omega \cup \partial\Omega)$, whose positivity and unitality translates the "maximum principle" for harmonic functions. When $\Omega$ is the unit disk, $k$ is the familiar Poisson's kernel.

### Continuity and Algebraic Properties of Positive Maps

Since the order structure of a $C^*$-algebra $A$ is defined by its positive cone $A_+$, positive maps are

1. real: $\phi(a^*) = \phi(a)^*$ and
2. order preserving: $\phi(a) \leq \phi(b)$ whenever $a \leq b$.

From this follows an important interplay between positivity and continuity:

a positive map $\phi : A \longrightarrow B$
between $C^*$-algebras is continuous

In case $A$ has a unit, this follows by the fact that $\phi$ is order preserving and that, for self-adjoint $a$, one has

$-\|a\|1_A \le a \le +\|a\|1_A$, so that $-\|a\|\phi(1_A) \le \phi(a) \le +\|a\|\phi(1_A)$ and then $\|\phi(a)\| \le \|\phi(1_A)\| \cdot \|a\|$. In general, splitting $a = b + ic$ as a combination of Hermitian elements $b$ and $c$, as $\|b\| \le \|a\|$ and $\|c\| \le \|a\|$, one obtains

$$\begin{aligned}\|\phi(a)\| &\le \|\phi(b)\| + \|\phi(c)\| \\ &\le \|\phi(1_A)\|(\|b\| + \|c\|) \\ &\le 2\|\phi(1_A)\| \cdot \|a\|\end{aligned}$$

The second general result concerning positivity and continuity is the following:

Let $\phi : A \longrightarrow B$ be a linear map between $C^*$-algebras with unit such that $\phi(1_A) = 1_B$; then $\phi$ is positive if and only if $\|\phi\| = 1$.

The result relies, among other things, on the generalized Schwarz inequality for unital positive maps on normal elements,

$$\phi(a^*)\phi(a) \le \phi(a^*a), \quad a^*a = aa^*$$

These results may be used to reveal the strong interplay between the algebraic, continuity and positivity properties of maps:

Let $\phi : A \longrightarrow B$ be an invertible linear map between unital $C^*$-algebras such that $\phi(1_A) = 1_B$. The following properties are equivalent:

1. $\phi$ is Jordan isomorphism,
2. $\phi$ is an isometry, and
3. $\phi$ is an order isomorphism ($\phi$ and $\phi^{-1}$ are order preserving).

The above conclusions can be strengthened if, instead of individual maps, continuous groups of maps are considered.

Let $t \mapsto \alpha_t$ be a strongly continuous, one-parameter group of maps of a unital $C^*$-algebra $A$ and assume that $\alpha_t(1_A) = 1_A$ for all $t \in \mathbb{R}$. The following properties are equivalent:

1. $\alpha_t$ is a $^*$-automorphism of $A$ for all $t \in \mathbb{R}$,
2. $\|\alpha_t\| \le 1$ for all $t \in \mathbb{R}$, and
3. $\alpha_t$ is positive for all $t \in \mathbb{R}$.

An analogous result holds true for $w^*$-continuous groups on abelian, or factors, von Neumann algebras.

## States on $C^*$-Algebras

A state on a $C^*$-algebra $A$ is a positive functional $\phi : A \longrightarrow \mathbb{C}$ of norm 1:

- $\phi(a^*a) \ge 0$ for all $a \in A$, and
- $\|\phi\| = 1$.

As $\mathbb{C}$ is a $C^*$-algebra, when $A$ is unital, a state on it is just a unital positive map:

- $\phi(a^*a) \ge 0$ for all $a \in A$, and
- $\phi(1_A) = 1$.

States for which $\phi(ab) = \phi(ba)$ are called tracial states.

States constitute a distinguished class of positive maps, both from a mathematical viewpoint and for application to mathematical physics. We will see below that states are deeply connected to representations of $C^*$-algebras (*see* $C^*$-Algebras and their Classification).

### States on Commutative $C^*$-Algebras

Since this is a subcase of positive maps in commutative $C^*$-algebras we only add a comment. As far as a $C^*$-algebra represents observable quantities of a physical system, states carry our actual knowledge about the system itself. The smallest $C^*$-sub-algebra $\{f(a): f \in C_0(\mathbb{R})\}$ of $A$ containing a given self-adjoint element $a \in A$, representing a certain observable quantity, is isomorphic to the algebra $C(\mathrm{Sp}(a))$ of continuous functions on the spectrum of $a$. A state on $A$ induces, by restriction, a state on $C(\mathrm{Sp}(a))$, which, by the Riesz representation theorem, is associated to a probability measure $\mu_a$ on $\mathrm{Sp}(a)$ through the formula

$$\phi(f(a)) = \int_{\mathrm{Sp}(a)} f(x)\mu_a(\mathrm{d}x)$$

Since $\mathrm{Sp}(a)$ represents the possible values of the observable associated to $a$, $\mu_a$ represents the distribution of these values when the physical state of the system is represented by $\phi$.

### Vector States and Density Matrices

In case $A$ is acting on a Hilbert space $h$, $A \subseteq \mathcal{B}(h)$, each unit vector $\xi \in h$ gives rise to a vector state $\phi_\xi(a) = (\xi|a\xi)$. In the quantum-mechanical description of a finite system, as far as observables with discrete spectrum are concerned, one can assume $A$ to be the $C^*$-algebra $\mathcal{K}(h)$ of compact operators on the Hilbert space $h$. In this case every state is a convex superposition of vector states, in the sense that it can be represented by the formula

$$\phi(a) = \mathrm{tr}(\rho a)/\mathrm{tr}(\rho), \quad a \in \mathcal{K}(h)$$

for a suitable density matrix $\rho$, that is, a positive, compact operator with finite trace. In quantum statistical mechanics, the grand canonical Gibbs equilibrium state of a finite system at inverse temperature $\beta$ and chemical potential $\mu$, with Hamiltonian $H$ and number operator $N$, is of the above type

$$\phi_{\beta,\mu}(a) = \mathrm{tr}(\mathrm{e}^{-\beta K}a)/\mathrm{tr}(\mathrm{e}^{-\beta K})$$

where $K = H - \mu N$, and the spectrum of $H$ is assumed to be discrete and such that $e^{-\beta K}$ is trace-class. For infinite systems, $A$ is a quasilocal $C^*$-algebra generated by a net $\{A_\alpha\}_\alpha$ of $C^*$-subalgebras describing observables referred to finite-volume regions. Infinite-volume equilibrium states on $A$ can then be obtained as thermodynamic limits of finite-volume Gibbs equilibrium states of the above type.

### Normal and Singular States

When observables with continuous spectrum have to be considered and one chooses the algebra $\mathcal{B}(h)$ of all bounded operators, the above formula, although still meaningful, does not describe all states on $\mathcal{B}(h)$ but only the important subclass of the normal ones. To this class, which can be considered on any von Neumann algebra $\mathcal{M}$, belong states $\phi$ which are $\sigma$-weakly continuous functionals. Equivalently, these are the states such that for all increasing net $a_\alpha \in \mathcal{M}_+$ with least upper bound $a \in \mathcal{M}_+, \phi(a)$ is least upper bound of the net $\phi(a_\alpha)$.

In general, each state $\phi$ on a von Neumann algebra $\mathcal{M}$ splits as a sum of a maximal normal piece and a singular one. Singular traces appear in noncommutative geometry as very useful tools to get back local objects from spectral ones via the familiar principle that local properties of functions depend on the asymptotics of their Fourier coefficients.

This is best illustrated on a compact, Riemannian $n$-manifold $M$ by the formula

$$\int_M f\, \mathrm{d}m = c_n \cdot \tau_\omega(M_f |D|^{-n})$$

which expresses the Riemannian integral of a nice function $f$ in terms of the Dirac operator $D$ acting on the Hilbert space of square-integrable spinors, the multiplication operator $M_f$ by $f$, and the singular Dixmier tracial state $\tau_\omega$ on $\mathcal{B}(H)$. Here the compactness of $M$ implies the compactness of the operator $M_f|D|^{-n}$ and $\tau_\omega$ is a limiting procedure depending only on the asymptotic behavior of the eigenvalues of $M_f|D|^{-n}$. Similar formulas are valid on self-similar fractals as well as on quasiconformal manifolds. Local index formulas represent cyclic cocycles in Connes' spectral geometry (*see* Noncommutative Geometry and the Standard Model; Noncommutative Geometry from Strings; Path-Integrals in Noncommutative Geometry).

### States and Representations: The GNS Construction

A fundamental tool in studying a $C^*$-algebra $A$ are its representations. These are morphisms of $C^*$-algebras $\pi : A \longrightarrow \mathcal{B}(\mathcal{H})$ from $A$ to the algebra of all bounded operators on some Hilbert space $\mathcal{H}$. There is a symbiotic appearance of states and representations on $C^*$-algebras. In fact, given a representation $\pi : A \longrightarrow \mathcal{B}(\mathcal{H})$, one easily constructs states on $A$ by unit vectors $\xi \in H$ by

$$\phi_\xi(a) = (\xi | \pi(a)\xi)$$

In fact, one checks that $\phi_\xi(a^*a) = (\xi|\pi(a^*a)\,\xi) = (\xi|\pi(a^*)\pi(a)\,\xi) = \|\pi(a)\,\xi\|^2 \geq 0$ and, at least if a unit exists, that $\phi_\xi(1_A) = \|\xi\|^2 = 1$.

A fundamental construction due to Gelfand, Naimark, and Segal allows to associate a representation to each state in such a way that each state is a vector state for a suitable representation.

"Let $\omega$ be a state over the $C^*$-algebra $A$. It follows that there exists cyclic representation $(\pi_\omega, \mathcal{H}_\omega, \xi_\omega)$ of $A$ such that

$$\omega(a) = (\xi_\omega | \pi_\omega(a)\xi_\omega)$$

Moreover, the representation is unique up to unitary equivalence. It is called the canonical cyclic representation of $A$ associated with $\omega$."

The positivity property of the state allows to introduce the positive-semidefinite scalar product $\langle a|b\rangle = \omega(a^*b)$ on the vector space $A$. Moreover, its kernel $\mathcal{I}_\omega = \{a \in A\colon \omega(a^*a) = 0\}$ is a left-ideal of $A$: in fact, if $a \in A$ and $b \in \mathcal{I}_\omega$ then $\omega((ba)^*(ba)) \leq \|a\|^2\omega(b^*b) = 0$. This allows to define, on the quotient pre-Hilbert space $A/\mathcal{I}_\omega$, an action of the elements $a \in A$: $\pi_\omega(a)(b + \mathcal{I}_\omega) := ab + \mathcal{I}_\omega$. It is the extension of this action to the Hilbert space completion $\mathcal{H}_\omega$ of $A/\mathcal{I}_\omega$ that gives the representation associated to $\omega$. When $A$ has a unit, the cyclic vector $\xi_\omega$ with the stated properties is precisely the image of $1_A + \mathcal{I}_\omega$. By definition, the cyclicity of the representation amounts to check that $\pi_\omega(A)\xi_\omega$ is dense in $\mathcal{H}_\omega$.

## Completely Positive Maps

In a sense, the order structure of a $C^*$-algebra $A$ is better understood through the sequence of $C^*$-algebras $A \otimes M_n(\mathbb{C}) \cong M_n(A)$, obtained as tensor products of $A$ and full matrix algebras $M_n(\mathbb{C})$. For example, $C^*$-algebras are matrix-ordered vector spaces as $\alpha^*(M_m(A))_+\alpha \subseteq (M_n(A))_+$ for all matrices $\alpha \in M_{m\times n}(\mathbb{C})$.

In this respect, one is naturally led to consider stronger notion of positivity:

"A map $\phi : A \longrightarrow B$ is called $n$-positive if its extension

$$\phi \otimes 1_n : A \otimes M_n(\mathbb{C}) \longrightarrow B \otimes M_n(\mathbb{C})$$
$$(\phi \otimes 1_n)[a_{i,j}]_{i,j} = [\phi(a_{i,j})]_{i,j}$$

is positive and completely positive (CP map for short) if this happens for all $n$."

Equivalently, $n$-positive means that $\sum_{i,j=1}^{n} b_i^* \phi \times (a_i^* a_j) b_j \geq 0$ for all $a_1, \ldots, a_n \in A$ and $b_1, \ldots, b_n \in b$. In particular, if $\phi$ is $n$-positive then it is $k$-positive for all $k \leq n$. Many positive maps we considered are in fact CP maps:

1. morphisms of $C^*$-algebras are CP maps;
2. positive maps $\phi : A \longrightarrow B$ are automatically CP maps provided $A, B$ or both are commutative and states are, in particular, CP maps; and
3. an important class of CP maps is the following. A norm one projection $\varepsilon : A \longrightarrow B$, from a $C^*$-algebra $A$ onto a $C^*$-subalgebra $B$, is a contraction such that $\varepsilon(b) = b$ for all $b \in B$. It can be proved that these maps satisfy $\varepsilon(bac) = b\varepsilon(a)c$ for all $a \in A$ and $b, c \in B$ and for this reason they are called conditional expectations. This property then implies that they are CP maps.

However, the identity map from a $C^*$-algebra $A$ into its opposite $A^\circ$ is positive but not 2-positive unless $A$ is commutative, the transposition $a \mapsto a^t$ in $M_n(\mathbb{C})$ is positive and not 2-positive if $n \geq 2$ and, for all $n$, there exist $n$-positive maps which are not $(n+1)$-positive.

### CP Maps in Mathematical Physics

In several fields of application, the transition of a state of a system into another state can be described by a completely positive map $\phi : A \longrightarrow B$ between $C^*$-algebras: for any given state $\omega$ of $B$, $\omega \circ \phi$ is then a state of $A$.

1. In the theory of quantum communication processes (*see* Channels in Quantum Information Theory; Optimal Cloning of Quantum States; Source Coding in Quantum Information Theory; Capacity for Quantum Information), for example, $B$ and $A$ represent the input and output systems, respectively, $\omega$ the signal to be transmitted, $\omega \circ \phi$ the received signal, and $\phi$ the system of transmission, called the channel.
2. In quantum probability and in the theory of quantum open systems, continuous semigroups of CP maps (*see* Quantum Dynamical Semigroups) describe dissipative time evolutions of a system due to interaction with an external one (heat bath).
3. In the theory of measurement in quantum mechanics, an observable can be described by a positive-operator-valued (POV) measure $M$ which assigns a positive element $m(E)$ in a $C^*$-algebra $A$ to each Borel subset $E$ of a topological space $X$. For each $a \in C_0(X)$, one can define its integral $\phi(f) := \int_X f \, dE$ as an element of $A$. The map $\phi : C_0(X) \longrightarrow A$, called the observation channel, is then a CP map.
4. Another field of mathematical physics in which CP maps play a distinguished role is in the construction and application of the quantum dynamical entropy, an extension of the Kolmogorov–Sinai entropy of measure preserving transformations (*see* Quantum Entropy). When dealing with a noncommutative dynamical system $(\mathcal{M}, \alpha, \tau)$ in which $\tau$ is a normal trace state on a finite von Neumann algebra $\mathcal{M}$, the Connes–Størmer entropy $h_\tau(\alpha)$ is defined through the consideration of an entropy functional $H_\tau(N_1, \ldots, N_k)$ of finite-dimensional von Neumann subalgebras $N_1, \ldots, N_k \subset \mathcal{M}$. To extend the definition to more general $C^*$-algebras and states on them, one has to face the fact that $C^*$-algebras may have no nontrivial $C^*$-subalgebras. To circumvent the problem A Connes, H Narnhofer, and W Thirring (CNT) introduced an entropy functional $H(\gamma_1, \ldots, \gamma_k)$ associated to a set $\gamma_i : A_i \longrightarrow A$ of CP maps (finite channels) from finite-dimensional $C^*$-algebras $A_i$ into $A$. This led to the CNT entropy $h_\omega(\alpha)$ of a noncommutative dynamical system $(A, \alpha, \omega)$, where $\omega$ is a state on $A$ and $\alpha$ is an automorphism or a CP map preserving it: $\omega \circ \alpha = \omega$.

### CP Maps and Continuity

Since for an element $a \in A$ of a unital $C^*$-algebra, one has $\|a\| \leq 1$ precisely when

$$\begin{pmatrix} 1 & a \\ a^* & 1 \end{pmatrix}$$

is positive in $M_2(A)$, it follows that

2-positive unital maps are contractive

Unital 2-positive maps satisfy, in particular, the generalized Schwarz inequality for all $a \in A$,

$$\phi(a^*)\phi(a) \leq \phi(a^* a)$$

In particular,

"CP maps are completely bounded as $\sup_n \|\phi \otimes 1_n\| = \|\phi(1_A)\|$ and completely contractive if they are unital. Conversely unital, completely contractive maps are CP maps."

### CP Maps and Matrix Algebras

When the domain or the target space of a map are matrix algebras, one has the following equivalences concerning positivity. Let $[e_{i,j}]_{i,j}$ denote the standard

matrix units in $M_n(\mathbb{C})$ and $\phi : M_n(\mathbb{C}) \longrightarrow B$ into a $C^*$-algebra $B$. The following conditions are equivalent:

1. $\phi$ is a CP map,
2. $\phi$ is $n$-positive, and
3. $[\phi(e_{i,j})]_{i,j}$ is positive in $M_n(B)$.

Associating to a linear map $\phi : A \longrightarrow M_n(\mathbb{C})$, the linear functional $s_\phi : M_n(A) \longrightarrow \mathbb{C}$ by $s_\phi([a_{i,j}]) := \sum_{i,j} \phi(a_{i,j})_{i,j}$, one has the following equivalent properties:

1. $\phi$ is a CP map,
2. $\phi$ is $n$-positive,
3. $s_\phi$ is positive, and
4. $s_\phi$ is positive on $A_+ \otimes M_n(\mathbb{C})_+$.

#### Stinspring Representation of CP Maps

CP maps are relatively easy to handle, thanks to the following dilation result due to W F Stinspring. It describes a CP map as the compression of a morphism of $C^*$-algebras.

Let $A$ be a unital $C^*$-algebra and $\phi : A \longrightarrow \mathcal{B}(\mathcal{H})$ a linear map. Then $\phi$ is a CP map if and only if it has the form

$$\phi(a) = V^* \pi(a) V$$

for some representation $\pi : A \longrightarrow \mathcal{B}(\mathcal{K})$ on a Hilbert space $\mathcal{K}$, and some bounded linear map $V : \mathcal{H} \longrightarrow \mathcal{K}$. If $A$ is a von Neumann algebra and $\phi$ is normal then $\pi$ can be taken to be normal. When $A = \mathcal{B}(\mathcal{H})$ and $\mathcal{H}$ is separable, one has, for some $b_n \in \mathcal{B}(\mathcal{H})$,

$$\phi(a) = \sum_{n=1}^{\infty} b_n^* a b_n$$

The proof of this result is reminiscent of the GNS construction for states and its extension, by G Kasparov, to $C^*$-modules is central in bivariant $K$-homology theory.

Despite the above satisfactory result, one should be aware that positive but not CP maps are much less understood and only for maps on very low dimensional matrix algebras do we have a definitive classification. To have an idea of the intricacies of the matter, one may consult Størmer (1963).

#### Positive Semigroups on Standard Forms of von Neumann Algebras and Ground State for Physical Hamiltonians

The above result allows one to derive the structure of generators of norm-continuous dynamical semigroups in terms of dissipative operators.

Strongly continuous positive semigroups, which are KMS symmetric with respect to a KMS state $\omega$ of a given automorphism group of a $C^*$-algebra $A$, can be analyzed as positive semigroups in the standard representation $(\mathcal{M}, \mathcal{H}, \mathcal{P}, J)$ (*see* Tomita–Takesaki Modular Theory) of the von Neumann algebra $\mathcal{M} := \pi_\omega(A)''$. A semigroup on $A$ gives rise to a corresponding $w^*$-continuous positive semigroup on $\mathcal{M}$ and to a strongly continuous positive semigroup on the ordered Hilbert space $(\mathcal{H}, \mathcal{P})$ of the standard form. In the latter framework, one can develop an infinite-dimensional, noncommutative extension of the classical Perron–Frobenius theory for matrices with positive entries. This applies, in particular, to semigroups generated by physical Hamiltonians and has been used to prove existence and uniqueness of the ground state for bosons and fermions systems in quantum field theory (one may consult Gross (1972)).

## Nuclear $C^*$-Algebras and Injective von Neumann Algebras

The nonabelian character of the product in $C^*$-algebras may prevent the existence of nontrivial morphisms between them, while one may have an abundance of CP maps. For example, there are no nontrivial morphisms from the algebra of compact operators to $\mathbb{C}$, but there exist sufficiently many states to separate its elements. A much more well-behaved category of $C^*$-algebras is obtained by considering CP maps as morphisms. This is true, in particular, for nuclear $C^*$-algebras: those for which any tensor product $A \otimes B$ with any other $C^*$-algebra $B$ admits a unique $C^*$-cross norm (*see* $C^*$-Algebras and their Classification). The intimate relation between this class of algebras and CP maps is illustrated by the following characterization:

1. $A$ is nuclear;
2. the identity map of $A$ is a pointwise limit of CP maps of finite rank;
3. the identity map of $A$ can be approximately factorized, $\lim_\alpha (T_\alpha \circ S_\alpha) a \to a$ for all $a \in A$, through matrix algebras and nets of CP maps $S_\alpha : A \to M_n(\mathbb{C}), T_\alpha : M_n(\mathbb{C}) \to A$.

A second important relation between nuclear $C^*$-algebras and CP maps emerges in connection to the lifting problem.

"Let $A$ be a nuclear $C^*$-algebra and $J$ a closed two-sided ideal in a $C^*$-algebra $B$. Then every CP map $\phi : A \to B/J$ can be lifted to a CP map $\phi' : A \to B$. In other words, $\phi$ factors through $B$ by the quotient map $q : B \to B/J$: $\phi = q \circ \phi$."

This and related results are used to prove that the Brown–Douglas–Fillmore $K$-homology invariant Ext($A$) is a group for separable, nuclear $C^*$-algebras.

Our last basic result, due to W Arveson, about CP maps concerns the extension problem.

> "Let $A$ be a unital $C^*$-algebra and $N$ a self-adjoint closed subspace of $A$ containing the identity. Then every CP map $\phi: N \to \mathcal{B}(\mathcal{H})$ from $N$ into a type I factor $\mathcal{B}(\mathcal{H})$ can be extended to a CP map $\phi: A \to \mathcal{B}(\mathcal{H})$."

This result can be restated by saying that type I factors are injective von Neumann algebras. It may suggest how the notion of a completely positive map plays a fundamental role along Connes' proof of one culminating result of the theory of von Neumann algebras, namely the fact that the class of injective von Neumann algebras coincides with the class of approximately finite-dimensional ones (*see* von Neumann Algebras: Introduction, Modular Theory and Classification Theory).

*See also:* Capacity for Quantum Information; $C^*$-Algebras and Their Classification; Channels in Quantum Information Theory; Noncommutative Geometry and the Standard Model; Noncommutative Geometry from Strings; Optimal Cloning of Quantum States; Path Integrals in Noncommutative Geometry; Quantum Dynamical Semigroups; Quantum Entropy; Source Coding in Quantum Information Theory; Tomita–Takesaki Modular Theory; von Neumann Algebras: Introduction, Modular Theory, and Classification Theory.

## Further Reading


Bratteli O and Robinson DW (1987) *Operator Algebras and Quantum Statistical Mechanics 1*, 2nd edn., 505 pp. Berlin: Springer; New York: Heidelberg.

Bratteli O and Robinson DW (1997) *Operator Algebras and Quantum Statistical Mechanics 2*, 2nd edn., 518 pp. Berlin: Springer; New York: Heidelberg.

Connes A (1994) *Noncommutative Geometry*, 661 pp. San Diego, CA: Academic Press.

Davies EB (1976) *Quantum Theory of Open Systems*, 171 pp. London: Academic Press.

Gross L (1972) Existence and uniqueness of physical ground states. *Journal of Functional Analysis* 10: 52–109.

Lance EC (1995) *Hilbert C*-Modules*, 130 pp. London: Cambridge University Press.

Ohya M and Petz D (1993) *Quantum Entropy and Its Use*, 335 pp. New York: Academic Press.

Paulsen VI (1996) *Completely Bounded Maps and Dilations*, 187 pp. Harlow: Longman Scientific-Technical.

Pisier G (2003) *Introduction to Operator Space Theory*, 479 pp. London: Cambridge University Press.

Størmer E (1963) Positive linear maps of operator algebras. *Acta Mathematica* 110: 233–278.

Takesaki M (2004a) Theory of Operator Algebras I, Second printing of the 1st edn., 415 pp. Berlin: Springer.

Takesaki M (2004b) *Theory of Operator Algebras II*, 384 pp. Berlin: Springer.

Takesaki M (2004c) Theory of Operator Algebras III, 548 pp. Berlin: Springer.


# Pseudo-Riemannian Nilpotent Lie Groups

**P E Parker**, Wichita State University, Wichita, KS, USA



## Nilpotent Lie Groups

While not much had been published on the geometry of nilpotent Lie groups with a left-invariant Riemannian metric till around 1990, the situation is certainly better now; see the references in Eberlein (2004). However, there is still very little that is conspicuous about the more general pseudo-Riemannian case. In particular, the two-step nilpotent groups are nonabelian and as close as possible to being abelian, but display a rich variety of new and interesting geometric phenomena (Cordero and Parker 1999). As in the Riemannian case, one of many places where they arise naturally is as groups of isometries acting on horospheres in certain (pseudo-Riemannian) symmetric spaces. Another is in the Iwasawa decomposition $G = KAN$ of semisimple groups with the Killing metric tensor, which need not be (positive or negative) definite even on $N$. Here, $K$ is compact and $A$ is abelian.

An early motivation for this study was the observation that there are two nonisometric pseudo-Riemannian metrics on the Heisenberg group $H_3$, one of which is flat. This is a strong contrast to the Riemannian case in which there is only one (up to positive homothety) and it is *not* flat. This is not an anomaly, as we now well know.

While the idea of more than one timelike dimension has appeared a few times in the physics literature, both in string/M-theory and in brane-world scenarios, essentially all work to date assumes only one. Thus, all applications so far are of Lorentzian or definite nilpotent groups. Guediri and co-workers led the Lorentzian studies, and most of their results stated near the end of the section "Lorentzian groups" concern a major, perennial interest in relativity: the (non)existence of closed timelike geodesics in compact Lorentzian manifolds.

Others have made use of nilpotent Lie groups with left-invariant (positive or negative) definite metric tensors, such as Hervig's (2004) constructions of black hole spacetimes from solvmanifolds (related to solvable groups: those with Iwasawa decomposition $G = AN$), including the so-called BTZ constructions. Definite groups and their applications, already having received thorough surveys elsewhere, most notably those of Eberlein, are not included here.

Although the geometric properties of Lie groups with left-invariant definite metric tensors have been studied extensively, the same has not occurred for indefinite metric tensors. For example, while the paper of Milnor (1976) has already become a classic reference, in particular for the classification of positive-definite (Riemannian) metrics on three-dimensional Lie groups, a classification of the left-invariant Lorentzian metric tensors on these groups became available only in 1997. Similarly, only a few partial results in the line of Milnor's study of definite metrics were previously known for indefinite metrics. Moreover, in dimension 3, there are only two types of metric tensors: Riemannian (definite) and Lorentzian (indefinite). But in higher dimensions, there are many distinct types of indefinite metrics while there is still essentially only one type of definite metric. This is another reason why this area has special interest now.

The list in "Further reading" at the end of this article consists of general survey articles and a select few of the more historically important papers. Precise bibliographical information for references merely mentioned or alluded to in this article may be found in those. The main, general reference on pseudo-Riemannian geometry is O'Neill's (1983) book. Eberlein's (2004) article covers the Riemannian case. At this time, there is no other comprehensive survey of the pseudo-Riemannian case. One may use Cordero and Parker (1999) and Guediri (2003) and their reference lists to good advantage, however.

## Inner Product and Signature

By an inner product on a vector space $V$ we shall mean a nondegenerate, symmetric bilinear form on $V$, generally denoted by $\langle\, , \rangle$. In particular, we *do not* assume that it is positive definite. It has become customary to refer to an ordered pair of non-negative integers $(p, q)$ as the signature of the inner product, where $p$ denotes the number of positive eigenvalues and $q$ the number of negative eigenvalues. Then nondegeneracy means that $p + q = \dim V$. Note that there is no real geometric difference between $(p, q)$ and $(q, p)$; indeed, O'Neill gives handy conversion procedures for this and for the other major sign variant (e.g., curvature) (see O'Neill (1983, pp. 92 and 89, respectively)).

A Riemannian inner product has signature $(p, 0)$. In view of the preceding remark, one might as well regard signature $(0, q)$ as also being Riemannian, so that "Riemannian geometry is that of definite metric tensors." Similarly, a Lorentzian inner product has either $p = 1$ or $q = 1$. In this case, both sign conventions are used in relativistic theories with the proviso that the "1" axis is always timelike.

If neither $p$ nor $q$ is 1, there is no physical convention. We shall say that $v \in V$ is *timelike* if $\langle v, v \rangle > 0$, *null* if $\langle v, v \rangle = 0$, and *spacelike* if $\langle v, v \rangle < 0$. (In a Lorentzian example, one may wish to revert to one's preferred relativistic convention.) We shall refer to these collectively as the *causal type* of a vector (or of a curve to which a vector is tangent).

Considering indefinite inner products (and metric tensors) thus greatly expands one's purview, from one type of geometry (Riemannian), or possibly two (Riemannian and Lorentzian), to a total of $\lfloor (p + q)/2 \rfloor + 1$ distinctly different types of geometries on the same underlying differential manifolds.

## Rise of 2-Step Groups

Throughout, $N$ will denote a connected (and simply connected, usually), nilpotent Lie group with Lie algebra $\mathfrak{n}$ having center $\mathfrak{z}$. We shall use $\langle\, , \rangle$ to denote either an inner product on $\mathfrak{n}$ or the induced left-invariant pseudo-Riemannian (indefinite) metric tensor on $N$.

For all nilpotent Lie groups, the exponential map $\exp : \mathfrak{n} \to N$ is surjective. Indeed, it is a diffeomorphism for simply connected $N$; in this case, we shall denote the inverse by log.

One of the earliest papers on the Riemannian geometry of nilpotent Lie groups was Wolf (1964). Since then, a few other papers about general nilpotent Lie groups have appeared, including Karidi (1994) and Pauls (2001), but the area has not seen a lot of progress.

However, everything changed with Kaplan's (1981) publication. Following this paper and its successor (Kaplan 1983), almost all subsequent work on the left-invariant geometry of nilpotent groups has been on two-step groups.

Briefly, Kaplan defined a new class of nilpotent Lie groups, calling them *of Heisenberg type*. This was soon abbreviated to $H$-type, and has since been called also as Heisenberg-like and (unfortunately) "generalized Heisenberg." (Unfortunate, because that term was already in use for another class, not all of which are of $H$-type.) What made them so

compelling was that (almost) everything was explicitly calculable, thus making them the next great test bed after symmetric spaces.

**Definition 1** We say that $N$ (or $\mathfrak{n}$) is 2-step nilpotent when $[\mathfrak{n},\mathfrak{n}] \subseteq \mathfrak{z}$. Then $[[\mathfrak{n},\mathfrak{n}],\mathfrak{n}]=0$ and the generalization to $k$-step nilpotent is clear:

$$[[\cdots[[[\mathfrak{n},\mathfrak{n}],\mathfrak{n}],\mathfrak{n}]\cdots],\mathfrak{n}] = 0$$

with $k+1$ copies of $\mathfrak{n}$ (or $k$ nested brackets, if you prefer).

It soon became apparent that $H$-type groups comprised a subclass of 2-step groups; for a nice, modern proof see Berndt *et al.* (1995). By around 1990, they had also attracted the attention of the spectral geometry community, and Eberlein produced the seminal survey (with important new results) from which the modern era began. (It was published in 1994 (Eberlein 1994), but the preprint had circulated widely since 1990.) Since then, activity around 2-step nilpotent Lie groups has mushroomed; see the references in Eberlein (2004).

Finally, turning to pseudo-Riemannian nilpotent Lie groups, with perhaps one or two exceptions, all results so far have been obtained only for 2-step groups. Thus, the remaining sections of this article will be devoted almost exclusively to them.

The Baker–Campbell–Hausdorff formula takes on a particularly simple form in these groups:

$$\exp(x)\exp(y) = \exp\left(x + y + \tfrac{1}{2}[x,y]\right) \qquad [1]$$

**Proposition 1** *In a pseudo-Riemannian 2-step nilpotent Lie group, the exponential map preserves causal character. Alternatively, one-parameter subgroups are curves of constant causal character.*

Of course, one-parameter subgroups need not be geodesics.

### Lattices and Completeness

We shall need some basic facts about lattices in $N$. In nilpotent Lie groups, a lattice is a discrete subgroup $\Gamma$ such that the homogeneous space $M=\Gamma\backslash N$ is compact. Here we follow the convention that a lattice acts on the left, so that the coset space consists of left cosets and this is indicated by the notation. Other subgroups will generally act on the right, allowing better separation of the effects of two simultaneous actions.

Lattices do not always exist in nilpotent Lie groups.

**Theorem 1** *The simply connected, nilpotent Lie group $N$ admits a lattice if and only if there exists a basis of its Lie algebra $\mathfrak{n}$ for which the structure constants are rational.*

Such a group is said to have a rational structure, or simply to be rational.

A *nilmanifold* is a (compact) homogeneous space of the form $\Gamma\backslash N$, where $N$ is a connected, simply connected (rational) nilpotent Lie group and $\Gamma$ is a lattice in $N$. An *infranilmanifold* has a nilmanifold as a finite covering space. They are commonly regarded as a noncommutative generalization of tori, the Klein bottle being the simplest example of an infranilmanifold that is not a nilmanifold.

We recall the result of Marsden from O'Neill (1983).

**Theorem 2** *A compact, homogeneous pseudo-Riemannian space is geodesically complete.*

Thus, if a rational $N$ is provided with a bi-invariant metric tensor $\langle\,,\rangle$, then $M$ becomes a compact, homogeneous pseudo-Riemannian space which is therefore complete. It follows that $(N,\langle\,,\rangle)$ is itself complete. In general, however, the metric tensor is not bi-invariant and $N$ need not be complete.

For 2-step nilpotent Lie groups, things work nicely as shown by this result first published by Guediri.

**Theorem 3** *On a 2-step nilpotent Lie group, all left-invariant pseudo-Riemannian metrics are geodesically complete.*

No such general result holds for 3- and higher-step groups, however.

## 2-Step Groups

In the Riemannian (positive-definite) case, one splits $\mathfrak{n}=\mathfrak{z}\oplus\mathfrak{v}=\mathfrak{z}\oplus\mathfrak{z}^{\perp}$, where the superscript denotes the orthogonal complement with respect to the inner product $\langle\,,\rangle$. In the general pseudo-Riemannian case, however, $\mathfrak{z}\oplus\mathfrak{z}^{\perp}\neq\mathfrak{n}$. The problem is that $\mathfrak{z}$ might be a degenerate subspace; that is, it might contain a null subspace $\mathfrak{U}$ for which $\mathfrak{U}\subseteq\mathfrak{U}^{\perp}$.

It turns out that this possible degeneracy of the center causes the essential differences between the Riemannian and pseudo-Riemannian cases. So far, the only general success in studying groups with degenerate centers was in Cordero and Parker (1999) where an adapted Witt decomposition of $\mathfrak{n}$ was used together with an involution $\iota$ exchanging the two null parts.

Observe that if $\mathfrak{z}$ is degenerate, the null subspace $\mathfrak{U}$ is well defined invariantly. We shall use a decomposition

$$\mathfrak{n} = \mathfrak{z}\oplus\mathfrak{v} = \mathfrak{U}\oplus\mathfrak{Z}\oplus\mathfrak{V}\oplus\mathfrak{E} \qquad [2]$$

in which $\mathfrak{z} = \mathfrak{U} \oplus \mathfrak{Z}$ and $\mathfrak{v} = \mathfrak{V} \oplus \mathfrak{E}$, $\mathfrak{U}$ and $\mathfrak{V}$ are complementary null subspaces, and $\mathfrak{U}^\perp \cap \mathfrak{V}^\perp = \mathfrak{Z} \oplus \mathfrak{E}$. Although the choice of $\mathfrak{V}$ is not well defined invariantly, once a $\mathfrak{V}$ has been chosen then $\mathfrak{Z}$ and $\mathfrak{E}$ are well defined invariantly. Indeed, $\mathfrak{Z}$ is the portion of the center $\mathfrak{z}$ in $\mathfrak{U}^\perp \cap \mathfrak{V}^\perp$, and $\mathfrak{E}$ is its orthocomplement in $\mathfrak{U}^\perp \cap \mathfrak{V}^\perp$. This is a Witt decomposition of $\mathfrak{n}$ given $\mathfrak{U}$, easily seen by noting that $(\mathfrak{U} \oplus \mathfrak{V})^\perp = \mathfrak{Z} \oplus \mathfrak{E}$, adapted to the special role of the center in $\mathfrak{n}$.

We shall also need to use an involution $\iota$ that interchanges $\mathfrak{U}$ and $\mathfrak{V}$ and which reduces to the identity on $\mathfrak{Z} \oplus \mathfrak{E}$ in the Riemannian (positive-definite) case. (The particular choice of such an involution is not significant.) It turns out that $\iota$ is an isometry of $\mathfrak{n}$ which does not integrate to an isometry of $N$. The adjoint with respect to $\langle\,,\rangle$ of the adjoint representation of the Lie algebra $\mathfrak{n}$ on itself is denoted by $\mathrm{ad}^\dagger$.

**Definition 2** The linear mapping

$$j : \mathfrak{U} \oplus \mathfrak{Z} \to \mathrm{End}(\mathfrak{V} \oplus \mathfrak{E})$$

is given by

$$j(a)x = \iota\, \mathrm{ad}_x^\dagger\, \iota a$$

Formulas for the connection and curvatures, and explicit forms for many examples, may be found in Cordero and Parker (1999). It turns out there is a relatively large class of flat spaces, a clear distinction from the Riemannian case in which there are none.

Let $x, y \in \mathfrak{n}$. Recall that homaloidal planes are those for which the numerator $\langle R(x,y)y, x\rangle$ of the sectional curvature formula vanishes. This notion is useful for degenerate planes tangent to spaces that are not of constant curvature.

**Definition 3** A submanifold of a pseudo-Riemannian manifold is flat if and only if every plane tangent to the submanifold is homaloidal.

**Theorem 4** *The center $Z$ of $N$ is flat.*

**Corollary 1** *The only $N$ of constant curvature are flat.*

The degenerate part of the center can have a profound effect on the geometry of the whole group.

**Theorem 5** *If $[\mathfrak{n}, \mathfrak{n}] \subseteq \mathfrak{U}$ and $\mathfrak{E} = \{0\}$, then $N$ is flat.*

Among these spaces, those that also have $\mathfrak{Z} = \{0\}$ (which condition itself implies $[\mathfrak{n}, \mathfrak{n}] \subseteq \mathfrak{U}$) are fundamental, with the more general ones obtained by making nondegenerate central extensions. It is also easy to see that the product of any flat group with a nondegenerate abelian factor is still flat.

This is the best possible result in general. Using weaker hypotheses in place of $\mathfrak{E} = \{0\}$, such as $[\mathfrak{V}, \mathfrak{V}] = \{0\} = [\mathfrak{E}, \mathfrak{E}]$, it is easy to construct examples which are not flat.

**Corollary 2** *If* $\dim Z \geq \lceil n/2 \rceil$, *then there exists a flat metric on $N$.*

Here $\lceil r \rceil$ denotes the least integer greater than or equal to $r$ and $n = \dim N$.

Before continuing, we pause to collect some facts about the condition $[\mathfrak{n}, \mathfrak{n}] \subseteq \mathfrak{U}$ and its consequences.

**Remark 1** Since it implies $j(z) = 0$ for all $z \in \mathfrak{Z}$, this latter is possible with no pseudo-Euclidean de Rham factor, unlike the Riemannian case. (On the other hand, a pseudo-Euclidean de Rham factor is characterized in terms of the Kaplan-Eberlein map $j$ whenever the center is nondegenerate.)

Also, it implies $j(u)$ interchanges $\mathfrak{V}$ and $\mathfrak{E}$ for all $u \in \mathfrak{U}$ if and only if $[\mathfrak{V}, \mathfrak{V}] = [\mathfrak{E}, \mathfrak{E}] = \{0\}$. Examples are the Heisenberg group and the groups $H(p, 1)$ for $p \geq 2$ with null centers.

Finally, we note that it implies that, for every $u \in \mathfrak{U}$, $j(u)$ maps $\mathfrak{V}$ to $\mathfrak{V}$ if and only if $j(u)$ maps $\mathfrak{E}$ to $\mathfrak{E}$ if and only if $[\mathfrak{V}, \mathfrak{E}] = \{0\}$.

**Proposition 2** *If $j(z) = 0$ for all $z \in \mathfrak{Z}$ and $j(u)$ interchanges $\mathfrak{V}$ and $\mathfrak{E}$ for all $u \in \mathfrak{U}$, then $N$ is Ricci flat.*

**Proposition 3** *If $j(z) = 0$ for all $z \in \mathfrak{Z}$, then $N$ is scalar flat. In particular, this occurs when $[\mathfrak{n}, \mathfrak{n}] \subseteq \mathfrak{U}$.*

Much like the Riemannian case, we would expect that $(N, \langle\,,\rangle)$ should in some sense be similar to flat pseudo-Euclidean space. This is seen, for example, via the existence of totally geodesic subgroups (Cordero and Parker 1999). (O'Neill (1983, ex. 9, p. 125) has extended the definition of totally geodesic to degenerate submanifolds of pseudo-Riemannian manifolds.)

**Example 1** For any $x \in \mathfrak{n}$ the one-parameter subgroup $\exp(tx)$ is a geodesic if and only if $x \in \mathfrak{z}$ or $x \in \mathfrak{U} \oplus \mathfrak{E}$. This is essentially the same as the Riemannian case, but with some additional geodesic one-parameter subgroups coming from $\mathfrak{U}$.

**Example 2** Abelian subspaces of $\mathfrak{V} \oplus \mathfrak{E}$ are Lie subalgebras of $\mathfrak{n}$, and give rise to complete, flat, totally geodesic abelian subgroups of $N$, just as in the Riemannian case. Eberlein's construction is valid in general, and shows that if $\dim \mathfrak{V} \oplus \mathfrak{E} \geq 1 + k + k \dim \mathfrak{z}$, then every nonzero element of $\mathfrak{V} \oplus \mathfrak{E}$ lies in an abelian subspace of dimension $k + 1$.

**Example 3** The center $Z$ of $N$ is a complete, flat, totally geodesic submanifold. Moreover, it determines a foliation of $N$ by its left translates, so each leaf is flat and totally geodesic, as in the Riemannian

case. In the pseudo-Riemannian case, this foliation in turn is the orthogonal direct sum of two foliations determined by $\mathfrak{U}$ and $\mathfrak{Z}$, and the leaves of the $\mathfrak{U}$-foliation are also null. All these leaves are complete.

There is also the existence of $\dim \mathfrak{z}$ independent first integrals, a familiar result in pseudo-Euclidean space, and the geodesic equations are completely integrable; in certain cases (mostly when the center is nondegenerate), one can obtain explicit formulas. Unlike the Riemannian case, there are flat groups (nonabelian) which are isometric to pseudo-Euclidean spaces (abelian).

**Theorem 6** *If* $[\mathfrak{n}, \mathfrak{n}] \subseteq \mathfrak{U}$ *and* $\mathfrak{E} = \{0\}$, *then* $N$ *is geodesically connected. Consequently, so is any nilmanifold with such a universal covering space.*

Thus, these compact nilmanifolds are much like tori. This is also illustrated by the computation of their period spectrum.

### Isometry Group

The main new feature is that when the center is degenerate, the isometry group can be strictly larger in a significant way than when the center is nondegenerate (which includes the Riemannian case).

Letting $\mathrm{Aut}(N)$ denote the automorphism group of $N$ and $I(N)$ the isometry group of $N$, set $O(N) = \mathrm{Aut}(N) \cap I(N)$. In the Riemannian case, $I(N) = O(N) \ltimes N$, the semidirect product where $N$ acts as left translations. We have chosen the notation $O(N)$ to suggest an analogy with the pseudo-Euclidean case in which this subgroup is precisely the (general, including reflections) pseudo-orthogonal group. According to Wilson (1982), this analogy is good for any nilmanifold (not necessarily 2-step).

To see what is true about the isometry group in general, first consider the (left-invariant) splitting of the tangent bundle $TN = \mathfrak{z}N \oplus \mathfrak{v}N$.

**Definition 4** Denote by $I^{\mathrm{spl}}(N)$ the subgroup of the isometry group $I(N)$ which preserves the splitting $TN = \mathfrak{z}N \oplus \mathfrak{v}N$. Further, let $I^{\mathrm{aut}}(N) = O(N) \ltimes N$, where $N$ acts by left translations.

**Proposition 4** *If* $N$ *is a simply connected, 2-step nilpotent Lie group with left-invariant metric tensor, then* $I^{\mathrm{spl}}(N) \leq I^{\mathrm{aut}}(N)$.

There are examples to show that $I^{\mathrm{spl}} < I^{\mathrm{aut}}$ is possible when $\mathfrak{U} \neq \{0\}$.

When the center is degenerate, the relevant group analogous to a pseudo-orthogonal group may be larger.

**Proposition 5** *Let* $\tilde{O}(N)$ *denote the subgroup of* $I(N)$ *which fixes* $1 \in N$. *Then* $I(N) \cong \tilde{O}(N) \ltimes N$, *where* $N$ *acts by left translations.*

The proof is obvious from the definition of $\tilde{O}$. It is also obvious that $O \leq \tilde{O}$. Examples show that $O < \tilde{O}$, hence $I^{\mathrm{aut}} < I$, is possible when the center is degenerate.

Thus, we have three groups of isometries, not necessarily equal in general: $I^{\mathrm{spl}} \leq I^{\mathrm{aut}} \leq I$. When the center is nondegenerate ($\mathfrak{U} = \{0\}$), the Ricci transformation is block-diagonalizable and the rest of Kaplan's proof using it now also works.

**Corollary 3** *If the center is nondegenerate, then* $I(N) = I^{\mathrm{spl}}(N)$ *whence* $\tilde{O}(N) \cong O(N)$.

In the next few results, we use the phrase "a subgroup isometric to" a group to mean that the isometry is also an isomorphism of groups.

**Proposition 6** *For any* $N$ *containing a subgroup isometric to the flat three-dimensional Heisenberg group,*

$$I^{\mathrm{spl}}(N) < I^{\mathrm{aut}}(N) < I(N)$$

Unfortunately, this class does not include our flat groups in which $[\mathfrak{n}, \mathfrak{n}] \subseteq \mathfrak{U}$ and $\mathfrak{E} = \{0\}$. However, it does include many groups that do not satisfy $[\mathfrak{n}, \mathfrak{n}] \subseteq \mathfrak{U}$, such as the simplest quaternionic Heisenberg group.

**Remark 2** A direct computation shows that on this flat $H_3$ with null center, the only Killing fields with geodesic integral curves are the nonzero scalar multiples of a vector field tangent to the center.

**Proposition 7** *For any* $N$ *containing a subgroup isometric to the flat* $H_3 \times \mathbb{R}$ *with null center,*

$$I^{\mathrm{spl}}(N) < I^{\mathrm{aut}}(N) < I(N)$$

Many of our flat groups in which $[\mathfrak{n}, \mathfrak{n}] \subseteq \mathfrak{U}$ and $\mathfrak{E} = \{0\}$ have such a subgroup isometrically embedded, as in fact do many others which are not flat.

### Lattices and Periodic Geodesics

In this subsection, we assume that $N$ is rational and let $\Gamma$ be a lattice in $N$.

Certain tori $T_F$ and $T_B$ provide the model fiber and the base for a submersion of the coset space $\Gamma \backslash N$. This submersion may not be pseudo-Riemannian in the usual sense, because the tori may be degenerate. We began the study of periodic geodesics in these compact nilmanifolds, and obtained a complete calculation of the period spectrum for certain flat spaces.

To the compact nilmanifold $\Gamma\backslash N$ we may associate two flat (possibly degenerate) tori.

**Definition 5** Let $N$ be a simply connected, two-step nilpotent Lie group with lattice $\Gamma$ and let $\pi : \mathfrak{n} \to \mathfrak{v}$ denote the projection. Define

$$T_{\mathfrak{z}} = \mathfrak{z}/(\log \Gamma \cap \mathfrak{z})$$
$$T_{\mathfrak{v}} = \mathfrak{v}/\pi(\log \Gamma)$$

Observe that $\dim T_{\mathfrak{z}} + \dim T_{\mathfrak{v}} = \dim \mathfrak{z} + \dim \mathfrak{v} = \dim \mathfrak{n}$.

Let $m = \dim \mathfrak{z}$ and $n = \dim \mathfrak{v}$. It is a consequence of a theorem of Palais and Stewart that $\Gamma\backslash N$ is a principal $T^m$-bundle over $T^n$. The model fiber $T^m$ can be given a geometric structure from its closed embedding in $\Gamma\backslash N$; we denote this geometric $m$-torus by $T_F$. Similarly, we wish to provide the base $n$-torus with a geometric structure so that the projection $p_B : \Gamma\backslash N \twoheadrightarrow T_B$ is the appropriate generalization of a pseudo-Riemannian submersion (O'Neill 1983) to (possibly) degenerate spaces. Observe that the splitting $\mathfrak{n} = \mathfrak{z} \oplus \mathfrak{v}$ induces splittings $TN = \mathfrak{z}N \oplus \mathfrak{v}N$ and $T(\Gamma\backslash N) = \mathfrak{z}(\Gamma\backslash N) \oplus \mathfrak{v}(\Gamma\backslash N)$, and that $p_{B*}$ just mods out $\mathfrak{z}(\Gamma\backslash N)$. Examining O'Neill's definition, we see that the key is to construct the geometry of $T_B$ by defining

$$p_{B*} : \mathfrak{v}_\eta(\Gamma\backslash N) \to T_{p_B(\eta)}(T_B) \quad \text{for each } \eta \in \Gamma\backslash N \text{ is an isometry} \qquad [3]$$

and

$$\nabla^{T_B}_{p_{B*}x} p_{B*}y = p_{B*}(\pi \nabla_x y) \quad \text{for all } x, y \in \mathfrak{v} = \mathfrak{V} \oplus \mathfrak{E} \qquad [4]$$

where $\pi : \mathfrak{n} \to \mathfrak{v}$ is the projection. Then the rest of the usual results will continue to hold, provided that sectional curvature is replaced by the numerator of the sectional curvature formula at least when elements of $\mathfrak{V}$ are involved:

$$\langle R_{T_B}(p_{B*}x, p_{B*}y)\, p_{B*}y, p_{B*}x\rangle = \langle R_{\Gamma\backslash N}(x, y)y, x\rangle + \tfrac{3}{4}\langle [x, y], [x, y]\rangle \qquad [5]$$

Now $p_B$ will be a pseudo-Riemannian submersion in the usual sense if and only if $\mathfrak{U} = \mathfrak{V} = \{0\}$, as is always the case for Riemannian spaces.

In the Riemannian case, Eberlein showed that $T_F \cong T_{\mathfrak{z}}$ and $T_B \cong T_{\mathfrak{v}}$. In general, $T_B$ is flat only if $N$ has a nondegenerate center or is flat.

**Remark 3** Observe that the torus $T_B$ may be decomposed into a topological product $T_E \times T_V$ in the obvious way. It is easy to check that $T_E$ is flat and isometric to $(\log \Gamma \cap \mathfrak{E})\backslash \mathfrak{E}$, and that $T_V$ has a linear connection not coming from a metric and not flat in general. Moreover, the geometry of the product is "twisted" in a certain way. It would be interesting to determine which tori could appear as such a $T_V$ and how.

**Theorem 7** *Let $N$ be a simply connected, 2-step nilpotent Lie group with lattice $\Gamma$, a left-invariant metric tensor, and tori as above. The fibers $T_F$ of the (generalized) pseudo-Riemannian submersion $\Gamma\backslash N \twoheadrightarrow T_B$ are isometric to $T_{\mathfrak{z}}$. If in addition the center $Z$ of $N$ is nondegenerate, then the base $T_B$ is isometric to $T_{\mathfrak{v}}$.*

We recall that elements of $N$ can be identified with elements of the isometry group $I(N)$: namely, $n \in N$ is identified with the isometry $\phi = L_n$ of left translation by $n$. We shall abbreviate this by writing $\phi \in N$.

**Definition 6** We say that $\phi \in N$ translates the geodesic $\gamma$ by $\omega$ if and only if $\phi\gamma(t) = \gamma(t + \omega)$ for all $t$. If $\gamma$ is a unit-speed geodesic, we say that $\omega$ is a period of $\phi$.

Recall that unit speed means that $|\dot\gamma| = |\langle \dot\gamma, \dot\gamma\rangle|^{1/2} = 1$. Since there is no natural normalization for null geodesics, we do not define periods for them. In the Riemannian case and in the timelike Lorentzian case in strongly causal spacetimes, unit-speed geodesics are parametrized by arclength and this period is a translation distance. If $\phi$ belongs to a lattice $\Gamma$, it is the length of a closed geodesic in $\Gamma\backslash N$.

In general, recall that if $\gamma$ is a geodesic in $N$ and if $p_N : N \twoheadrightarrow \Gamma\backslash N$ denotes the natural projection, then $p_N\gamma$ is a periodic geodesic in $\Gamma\backslash N$ if and only if some $\phi \in \Gamma$ translates $\gamma$. We say *periodic* rather than *closed* here because in pseudo-Riemannian spaces it is possible for a null geodesic to be closed but not periodic. If the space is geodesically complete or Riemannian, however, then this does not occur; the former is in fact the case for our 2-step nilpotent Lie groups. Further, recall that free homotopy classes of closed curves in $\Gamma\backslash N$ correspond bijectively with conjugacy classes in $\Gamma$.

**Definition 7** Let $\mathcal{C}$ denote either a nontrivial, free homotopy class of closed curves in $\Gamma\backslash N$ or the corresponding conjugacy class in $\Gamma$. We define $\wp(\mathcal{C})$ to be the set of all periods of periodic unit-speed geodesics that belong to $\mathcal{C}$.

In the Riemannian case, this is the set of lengths of closed geodesics in $\mathcal{C}$, frequently denoted by $\ell(\mathcal{C})$.

**Definition 8** The *period spectrum* of $\Gamma\backslash N$ is the set

$$\mathrm{spec}_\wp(\Gamma\backslash N) = \bigcup_{\mathcal{C}} \wp(\mathcal{C})$$

where the union is taken over all nontrivial, free homotopy classes of closed curves in $\Gamma\backslash N$.

In the Riemannian case, this is the length spectrum $\mathrm{spec}_\ell(\Gamma\backslash N)$.

**Example 4** Similar to the Riemannian case, we can compute the period spectrum of a flat torus $\Gamma\backslash\mathbb{R}^m$, where $\Gamma$ is a lattice (of maximal rank, isomorphic to $\mathbb{Z}^m$). Using calculations in an analogous way as for finding the length spectrum of a Riemannian flat torus, we easily obtain

$$\mathrm{spec}_\wp(\Gamma\backslash\mathbb{R}^m) = \{|g| \neq 0 \mid g \in \Gamma\}$$

It is also easy to see that the nonzero d'Alembertian spectrum is related to the analogous set produced from the dual lattice $\Gamma^*$, multiplied by factors of $\pm 4\pi^2$, almost as in the Riemannian case.

As in this example, simple determinacy of periods of unit-speed geodesics helps make calculation of the period spectrum possible purely in terms of $\log\Gamma \subseteq \mathfrak{n}$.

For the rest of this subsection, we assume that $N$ is a simply connected, two-step nilpotent Lie group with left-invariant pseudo-Riemannian metric tensor $\langle\,,\rangle$. Note that non-null geodesics may be taken to be of unit speed. Most non-identity elements of $N$ translate some geodesic, but not necessarily one of unit speed.

For our special class of flat 2-step nilmanifolds, we can calculate the period spectrum completely.

**Theorem 8** *If* $[\mathfrak{n},\mathfrak{n}] \subseteq \mathfrak{U}$ *and* $\mathfrak{E} = \{0\}$, *then* $\mathrm{spec}_\wp(M)$ *can be completely calculated from* $\log\Gamma$ *for any* $M = \Gamma\backslash N$.

Thus, we see again just how much these flat, two-step nilmanifolds are like tori. All periods can be calculated purely from $\log\Gamma \subseteq \mathfrak{n}$, although some will not show up from the tori in the fibration.

**Corollary 4** $\mathrm{spec}_\wp(T_B)$ (*respectively*, $T_F$) *is* $\cup_{\mathcal{C}}\wp(\mathcal{C})$ *where the union is taken over all those free homotopy classes* $\mathcal{C}$ *of closed curves in* $M = \Gamma\backslash N$ *that* do not (*respectively*, do) *contain an element in the center of* $\Gamma \cong \pi_1(M)$, *except for those periods arising only from unit-speed geodesics in* $M$ *that project to null geodesics in both* $T_B$ *and* $T_F$.

We note that one might consider using this to assign periods to some null geodesics in the tori $T_B$ and $T_F$.

When the center is nondegenerate, we obtain results similar to Eberlein's. Here is part of them.

**Theorem 9** *Assume* $\mathfrak{U} = \{0\}$. *Let* $\phi \in N$ *and write* $\log\phi = z^* + e^*$. *Assume* $\phi$ *translates the unit-speed geodesic* $\gamma$ *by* $\omega > 0$. *Let* $z'$ *denote the component of* $z^*$ *orthogonal to* $[e^*,\mathfrak{n}]$ *and set* $\omega^* = |z' + e^*|$. *Let* $\dot\gamma(0) = z_0 + e_0$. *Then*

(i) $|e^*| \leq \omega$. *In addition,* $\omega < \omega^*$ *for timelike (spacelike) geodesics with* $\omega z_0 - z'$ *timelike (spacelike), and* $\omega > \omega^*$ *for timelike (spacelike) geodesics with* $\omega z_0 - z'$ *spacelike (timelike);*

(ii) $\omega = |e^*|$ *if and only if* $\gamma(t) = \exp(te^*/|e^*|)$ *for all* $t \in \mathbb{R}$; *and*

(iii) $\omega = \omega^*$ *if and only if* $\omega z_0 - z'$ *is null.*

Although $\omega^*$ need not be an upper bound for periods as in the Riemannian case, it nonetheless plays a special role among all periods, as seen in (iii) above, and we shall refer to it as the distinguished period associated with $\phi \in N$. When the center is definite, for example, we do have $\omega \leq \omega^*$.

Now the following definitions make sense at least for $N$ with a nondegenerate center.

**Definition 9** Let $\mathcal{C}$ denote either a nontrivial, free homotopy class of closed curves in $\Gamma\backslash N$ or the corresponding conjugacy class in $\Gamma$. We define $\wp^*(\mathcal{C})$ to be the distinguished periods of periodic unit-speed geodesics that belong to $\mathcal{C}$.

**Definition 10** The *distinguished period spectrum* of $\Gamma\backslash N$ is the set

$$\mathcal{D}\mathrm{spec}_\wp(\Gamma\backslash N) = \bigcup_{\mathcal{C}} \wp^*(\mathcal{C})$$

where the union is taken over all nontrivial, free homotopy classes of closed curves in $\Gamma\backslash N$.

Then we get this result:

**Corollary 5** *Assume the center is nondegenerate. If* $\mathfrak{n}$ *is nonsingular, then* $\mathrm{spec}_\wp(T_B)$ (*respectively*, $T_F$) *is precisely the period spectrum (respectively, the distinguished period spectrum) of those free homotopy classes* $\mathcal{C}$ *of closed curves in* $M = \Gamma\backslash N$ *that do not (respectively, do) contain an element in the center of* $\Gamma \cong \pi_1(M)$, *except for those periods arising only from unit-speed geodesics in* $M$ *that project to null geodesics in both* $T_B$ *and* $T_F$.

## Conjugate Loci

This is the only general result on conjugate points.

**Proposition 8** *Let* $N$ *be a simply connected, 2-step nilpotent Lie group with left-invariant metric tensor* $\langle\,,\rangle$, *and let* $\gamma$ *be a geodesic with* $\dot\gamma(0) = a \in \mathfrak{z}$. *If* $\mathrm{ad}^\dagger_\bullet a = 0$, *then there are no conjugate points along* $\gamma$.

In the rest of this subsection, we assume that the center of $N$ is nondegenerate.

For convenience, we shall use the notation $J_z = \mathrm{ad}^\dagger_\bullet z$ for any $z \in \mathfrak{z}$. (Since the center is

nondegenerate, the involution $\iota$ may be omitted.) We follow Ciatti (2000) for this next definition. As in the Riemannian case, one might as well make 2-step nilpotency part of the definition since it effectively is so anyway.

**Definition 11** $N$ is said to be of *pseudoH-type* if and only if

$$J_z^2 = -\langle z, z\rangle I$$

for any $z \in \mathfrak{z}$.

Complete results on conjugate loci have been obtained only for these groups (Jang *et al.* 2005). For example, using standard results from analytic function theory, one can show that the conjugate locus is an analytic variety in $N$. This is probably true for general two-step groups, but the proof we know works only for pseudo$H$-type.

**Definition 12** Let $\gamma$ denote a geodesic and assume that $\gamma(t_0)$ is conjugate to $\gamma(0)$ along $\gamma$. To indicate that the multiplicity of $\gamma(t_0)$ is $m$, we shall write $\text{mult}_{\text{cp}}(t_0) = m$. To distinguish the notions clearly, we shall denote the multiplicity of $\lambda$ as an eigenvalue of a specified linear transformation by $\text{mult}_{\text{ev}}\lambda$.

Let $\gamma$ be a geodesic with $\gamma(0) = 1$ and $\dot\gamma(0) = z_0 + x_0 \in \mathfrak{z} \oplus \mathfrak{v}$, respectively, and let $J = J_{z_0}$. If $\gamma$ is not null, we may assume that $\gamma$ is normalized so that $\langle\dot\gamma,\dot\gamma\rangle = \pm 1$. As usual, $\mathbb{Z}^*$ denotes the set of all integers with 0 removed.

**Theorem 10** *Under these assumptions, if $N$ is of pseudoH-type, then*:

(i) *if $z_0 = 0$ and $x_0 \neq 0$, then $\gamma(t)$ is conjugate to $\gamma(0)$ along $\gamma$ if and only if $\langle x_0, x_0\rangle < 0$ and*

$$-\frac{12}{t^2} = \langle x_0, x_0\rangle$$

*in which case* $\text{mult}_{\text{cp}}(t) = \dim \mathfrak{z}$;

(ii) *if $z_0 \neq 0$ and $x_0 = 0$, then $\gamma(t)$ is conjugate to $\gamma(0)$ along $\gamma$ if and only if $\langle z_0, z_0\rangle > 0$ and*

$$t \in \frac{2\pi}{|z_0|}\mathbb{Z}^*$$

*in which case* $\text{mult}_{\text{cp}}(t) = \dim \mathfrak{v}$.

**Theorem 11** *Let $\gamma$ be such a geodesic in a pseudoH-type group $N$ with $z_0 \neq 0 \neq x_0$.*

(i) *If $\langle z_0, z_0\rangle = \alpha^2$ with $\alpha > 0$, then $\gamma(t_0)$ is conjugate to $\gamma(0)$ along $\gamma$ if and only if*

$$t_0 \in \frac{2\pi}{\alpha}\mathbb{Z}^* \cup \mathbb{A}_1 \cup \mathbb{A}_2$$

*where*

$$\mathbb{A}_1 = \left\{ t \in \mathbb{R} \,\middle|\, \langle x_0, x_0\rangle \frac{\alpha t}{2}\cot\frac{\alpha t}{2} = \langle\dot\gamma,\dot\gamma\rangle \right\}$$

*and*

$$\mathbb{A}_2 = \left\{ t \in \mathbb{R} \,\middle|\, \alpha t = \frac{\langle x_0, x_0\rangle}{\langle\dot\gamma,\dot\gamma\rangle + \langle z_0, z_0\rangle}\sin\alpha t \right\}$$

*when* $\dim\mathfrak{z} \geq 2$

*If $t_0 \in (2\pi/\alpha)\mathbb{Z}^*$, then*

$$\text{mult}_{\text{cp}}(t_0) = \begin{cases} \dim\mathfrak{v} - 1 & \textit{if } \langle\dot\gamma,\dot\gamma\rangle + \langle z_0, z_0\rangle \neq 0 \\ \dim\mathfrak{n} - 2 & \textit{if } \langle\dot\gamma,\dot\gamma\rangle + \langle z_0, z_0\rangle = 0 \end{cases}$$

*If $t_0 \notin (2\pi/\alpha)\mathbb{Z}^*$, then*

$$\text{mult}_{\text{cp}}(t_0) = \begin{cases} 1 & \textit{if } t_0 \in \mathbb{A}_1 - \mathbb{A}_2 \\ \dim\mathfrak{z} - 1 & \textit{if } t_0 \in \mathbb{A}_2 - \mathbb{A}_1 \\ \dim\mathfrak{z} & \textit{if } t_0 \in \mathbb{A}_1 \cap \mathbb{A}_2 \end{cases}$$

(ii) *If $\langle z_0, z_0\rangle = -\beta^2$ with $\beta > 0$, then $\gamma(t_0)$ is a conjugate point along $\gamma$ if and only if $t_0 \in \mathbb{B}_1 \cup \mathbb{B}_2$ where*

$$\mathbb{B}_1 = \left\{ t \in \mathbb{R} \,\middle|\, \langle x_0, x_0\rangle \frac{\beta t}{2}\coth\frac{\beta t}{2} = \langle\dot\gamma,\dot\gamma\rangle \right\}$$

*and*

$$\mathbb{B}_2 = \left\{ t \in \mathbb{R} \,\middle|\, \beta t = \frac{\langle x_0, x_0\rangle}{\langle\dot\gamma,\dot\gamma\rangle + \langle z_0, z_0\rangle}\sinh\beta t \right\}$$

*when* $\dim\mathfrak{z} \geq 2$

*The multiplicity is*

$$\text{mult}_{\text{cp}}(t_0) = \begin{cases} 1 & \textit{if } t_0 \in \mathbb{B}_1 - \mathbb{B}_2 \\ \dim\mathfrak{z} - 1 & \textit{if } t_0 \in \mathbb{B}_2 - \mathbb{B}_1 \\ \dim\mathfrak{z} & \textit{if } t_0 \in \mathbb{B}_1 \cap \mathbb{B}_2 \end{cases}$$

(iii) *If $\langle z_0, z_0\rangle = 0$, then $\gamma(t_0)$ is a conjugate point along $\gamma$ if and only if*

$$t_0^2 = -\frac{12}{\langle x_0, x_0\rangle}$$

*and* $\text{mult}_{\text{cp}}(t_0) = \dim\mathfrak{z} - 1$.

This covers all cases for a pseudo$H$-type group with a center of any dimension.

Some results on other two-step groups and examples (including pictures in dimension 3) may be found in the references cited in Jang *et al.* (2005). When the groups are not pseudo$H$-type, however, complete results are available only when the center is one dimensional. Guediri (2004) has results in the timelike Lorentzian case.

## Lorentzian Groups

Not too long ago, only a few partial results in the line of Milnor's study of definite metrics were known for indefinite metrics (Barnet 1989, Nomizu 1979), and they were Lorentzian.

Guediri (2003) and others have made special study of Lorentzian two-step groups, partly because of their relevance to general relativity, where they can be used to provide interesting and important (counter)examples. Special features of Lorentzian geometry frequently enable them to obtain much more complete and explicit results than are possible in general.

For example, Guediri (2003) was able to provide a complete and explicit integration of the geodesic equations for Lorentzian 2-step groups. This includes the case of a degenerate center, which only required extremely careful handling through a number of cases. He also paid special attention to the existence of closed timelike geodesics, reflecting the relativistic concerns.

As usual, $N$ denotes a connected and simply connected 2-step nilpotent Lie group. For the rest of this section, we assume that the left-invariant metric tensor is Lorentzian. Whenever a lattice is mentioned, we also assume that the group is rational.

**Proposition 9** *If the center is degenerate, then no timelike geodesic can be translated by a central element.*

Thus, there can be no closed timelike geodesics parallel to the center in any nilmanifold obtained from such an $N$.

**Theorem 12** *If the center is Lorentzian, then $\Gamma\backslash N$ contains no timelike or null closed geodesics for any lattice $\Gamma$.*

To handle degenerate centers, three refined notions for nonsingular are used: *almost*, *weakly*, and *strongly* nonsingular. The precise definitions involve an adapted Witt decomposition (as in the general pseudo-Riemannian case, but a rather different one here) and are quite technical, as is typical. We refer to Guediri (2003) for details.

**Theorem 13** *If $N$ is weakly nonsingular, then no timelike geodesic can be translated by an element of $N$.*

**Corollary 6** *If $N$ is flat, then no timelike geodesic can be translated by a non-identity element.*

**Corollary 7** *If $N$ is flat, then $\Gamma\backslash N$ contains no closed timelike geodesics for any lattice $\Gamma$.*

**Corollary 8** *If $N$ is weakly nonsingular, then $\Gamma\backslash N$ contains no closed timelike geodesic.*

**Corollary 9** *If $N = H_{2k+1}$ is a Lorentzian Heisenberg group with degenerate center, then $\Gamma\backslash N$ contains no closed timelike geodesic.*

Guediri also has the only non-Riemannian results so far about the phenomenon Eberlein called "in resonance." Roughly speaking, this occurs when the eigenvalues of the map $j$ have rational ratios. (The Lorentzian case actually requires a slightly more complicated condition when the center is degenerate.)

**Theorem 14** *If $N$ is almost nonsingular, then $N$ is in resonance if and only if every geodesic of $N$ is translated by some element of $N$.*

*See also:* Classical Groups and Homogeneous Spaces; Einstein Equations: Exact Solutions; Lorentzian Geometry.

## Further Reading

Barnet F (1989) On Lie groups that admit left-invariant Lorentz metrics of constant sectional curvature. *Illinois Journal of Mathematics* 33: 631–642.

Berndt J, Tricerri F, and Vanhecke L (1995) *Generalized Heisenberg Groups and Damek-Ricci Harmonic Spaces*, LNM 1598. Berlin: Springer.

Ciatti P (2000) Scalar products on Clifford modules and pseudo-*H*-type Lie algebras. *Annali di Matematica Pura ed Applicata* 178: 1–31.

Cordero LA and Parker PE (1999) Pseudoriemannian 2-step nilpotent Lie groups, Santiago-Wichita. *Preprint* DGS/CP4, (arXiv:math.DG/9905188).

Eberlein P (1994) Geometry of 2-step nilpotent groups with a left-invariant metric. *Annales Scientifique de l'École Normale Supérieure* 27: 611–660.

Eberlein P (2004) Left-invariant geometry of Lie groups. *Cubo* 6: 427–510. (See also http://www.math.unc.edu/faculty/pbe.)

Guediri M (2003) Lorentz geometry of 2-step nilpotent Lie groups. *Geometriae Dedicata* 100: 11–51.

Guediri M (2004) The timelike cut locus and conjugate points in Lorentz 2-step nilpotent Lie groups. *Manuscripta Mathematica* 114: 9–35.

Hervig S (2004) Einstein metrics: homogeneous solvmanifolds, generalised Heisenberg groups and black holes. *Journal of Geometry and Physics* 52: 298–312.

Jang C, Parker PE, and Park K (2005) Pseudo*H*-type 2-step nilpotent Lie groups. *Houston Journal of Mathematics* 31: 765–786 (arXiv:math.DG/0307368).

Kaplan A (1981) Riemannian nilmanifolds attached to Clifford modules. *Geometriae Dedicata* 11: 127–136.

Kaplan A (1983) On the geometry of Lie groups of Heisenberg type. *Bulletin of the London Mathematical Society* 15: 35–42.

Karidi R (1994) Geometry of balls in nilpotent Lie groups. *Duke Mathematical Journal* 74: 301–317.

Milnor J (1976) Curvatures of left-invariant metrics on Lie groups. *Advances in Mathematics* 21: 293–329.

Nomizu K (1979) Left-invariant Lorentz metrics on Lie groups. *Osaka Mathematical Journal* 16: 143–150.
O'Neill B (1983) *Semi-Riemannian Geometry*. New York: Academic Press.
Pauls SD (2001) The large scale geometry of nilpotent Lie groups. *Communications in Analysis and Geometry* 9: 951–982.
Wilson EN (1982) Isometry groups on homogeneous manifolds. *Geometriae Dedicata* 12: 337–346.
Wolf JA (1964) Curvature in nilpotent Lie groups. *Proceedings of the American Mathematical Society* 15: 271–274.

# Q

## $q$-Special Functions

**T H Koornwinder**, University o
f Amsterdam, Amsterdam, The Netherlands



### Introduction

In this article we give a brief introduction to $q$-special functions, that is, $q$-analogs of the classical special functions. Here $q$ is a deformation parameter, usually $0 < q < 1$, where $q = 1$ is the classical case. The deformation is such that the calculus simultaneously deforms to a $q$-calculus involving $q$-derivatives and $q$-integrals. The main topics to be treated are $q$-hypergeometric series, with some selected evaluation and transformation formulas, and some $q$-hypergeometric orthogonal polynomials, most notably the Askey–Wilson polynomials. In several variables, we discuss Macdonald polynomials associated with root systems, with most emphasis on the $A_n$ case. The rather new theory of elliptic hypergeometric series gets some attention. While much of the theory of $q$-special functions keeps $q$ fixed, some of the deeper aspects with number-theoretic and combinatorial flavor emphasize expansion in $q$. Finally, we indicate applications and interpretations in quantum groups, Chevalley groups, affine Lie algebras, combinatorics, and statistical mechanics.

#### Conventions

$q \in \mathbb{C}\backslash\{1\}$ in general, but $0 < q < 1$ in all infinite sums and products.
$n, m, N$ will be non-negative integers unless mentioned otherwise.

### $q$-Hypergeometric Series

#### Definitions

For $a, q \in \mathbb{C}$ the $q$-shifted factorial $(a;q)_k$ is defined as a product of $k$ factors:

$$(a;q)_k := (1-a)(1-aq)\cdots(1-aq^{k-1}) \quad (k \in \mathbb{Z}_{>0}); \quad (a;q)_0 := 1 \qquad [1]$$

If $|q| < 1$ this definition remains meaningful for $k = \infty$ as a convergent infinite product:

$$(a;q)_\infty := \prod_{j=0}^{\infty}(1-aq^j) \qquad [2]$$

We also write $(a_1, \ldots, a_r; q)_k$ for the product of $r$ $q$-shifted factorials:

$$(a_1, \ldots, a_r; q)_k := (a_1;q)_k \ldots (a_r;q)_k \quad (k \in \mathbb{Z}_{\geq 0} \text{ or } k = \infty) \qquad [3]$$

A $q$-hypergeometric series is a power series (for the moment still formal) in one complex variable $z$ with power series coefficients which depend, apart from $q$, on $r$ complex upper parameters $a_1, \ldots, a_r$ and $s$ complex lower parameters $b_1, \ldots, b_s$ as follows:

$$\begin{aligned} {}_r\phi_s\left[\begin{matrix} a_1, \ldots, a_r \\ b_1, \ldots, b_s \end{matrix}; q, z\right] &= {}_r\phi_s(a_1, \ldots, a_r; b_1, \ldots, b_s; q, z) \\ &:= \sum_{k=0}^{\infty} \frac{(a_1, \ldots, a_r; q)_k}{(b_1, \ldots, b_s; q)_k (q;q)_k} \\ &\quad \times \left((-1)^k q^{(1/2)k(k-1)}\right)^{s-r+1} z^k \quad (r, s \in \mathbb{Z}_{\geq 0}) \end{aligned} \qquad [4]$$

Clearly the above expression is symmetric in $a_1, \ldots, a_r$ and symmetric in $b_1, \ldots, b_s$. On the right-hand side of [4], we have that

$$\frac{(k+1)\text{th term}}{k\text{th term}} = \frac{(1-a_1q^k)\cdots(1-a_rq^k)(-q^k)^{s-r+1}z}{(1-b_1q^k)\cdots(1-b_sq^k)(1-q^{k+1})} \qquad [5]$$

is rational in $q^k$. Conversely, any rational function in $q^k$ can be written in the form of the right-hand side of [5]. Hence, any series $\sum_{k=0}^{\infty} c_k$ with $c_0 = 1$ and $c_{k+1}/c_k$ rational in $q^k$ is of the form of a $q$-hypergeometric series [4].

In order to avoid singularities in the terms of [4], we assume that $b_1, \ldots, b_s \neq 1, q^{-1}, q^{-2}, \ldots$. If, for some $i, a_i = q^{-n}$, then all terms in the series [4] with $k > n$ will vanish. If none of the $a_i$ is equal to $q^{-n}$

and if $|q| < 1$, then the radius of convergence of the power series [4] equals $\infty$ if $r < s+1$, 1 if $r = s+1$, and 0 if $r > s+1$.

We can view the $q$-shifted factorial as a $q$-analog of the shifted factorial (or Pochhammer symbol) by the limit formula

$$\lim_{q\to 1}\frac{(q^a;q)_k}{(1-q)^k} = (a)_k := a(a+1)\cdots(a+k-1) \quad [6]$$

Hence the $q$-binomial coefficient

$$\begin{bmatrix} n \\ k \end{bmatrix}_q := \frac{(q;q)_n}{(q;q)_k(q;q)_{n-k}} \quad (n,k\in\mathbb{Z}, n\geq k\geq 0) \quad [7]$$

tends to the binomial coefficient for $q \to 1$:

$$\lim_{q\to 1}\begin{bmatrix} n \\ k \end{bmatrix}_q = \binom{n}{k} \quad [8]$$

and a suitably renormalized $q$-hypergeometric series tends (at least formally) to a hypergeometric series as $q \uparrow 1$:

$$\lim_{q\uparrow 1} {}_{r+r'}\phi_{s+s'}\left[\begin{matrix} q^{a_1},\ldots,q^{a_r},c_1,\ldots,c_{r'} \\ q^{b_1},\ldots,q^{b_s},d_1,\ldots,d_{s'}\end{matrix};q,(q-1)^{1+s-r}z\right]$$
$$= {}_rF_s\left(\begin{matrix} a_1,\ldots,a_r \\ b_1,\ldots,b_s\end{matrix};\frac{(c_1-1)\cdots(c_{r'}-1)z}{(d_1-1)\cdots(d_{s'}-1)}\right) \quad [9]$$

At least formally, there are limit relations between $q$-hypergeometric series with neighboring $r, s$:

$$\lim_{a_r\to\infty} {}_r\phi_s\left[\begin{matrix} a_1,\ldots,a_r \\ b_1,\ldots,b_s\end{matrix};q,\frac{z}{a_r}\right] = {}_{r-1}\phi_s\left[\begin{matrix} a_1,\ldots,a_{r-1} \\ b_1,\ldots,b_s\end{matrix};q,z\right] \quad [10]$$

$$\lim_{b_s\to\infty} {}_r\phi_s\left[\begin{matrix} a_1,\ldots,a_r \\ b_1,\ldots,b_s\end{matrix};q,b_s z\right] = {}_r\phi_{s-1}\left[\begin{matrix} a_1,\ldots,a_r \\ b_1,\ldots,b_{s-1}\end{matrix};q,z\right] \quad [11]$$

A terminating $q$-hypergeometric series $\sum_{k=0}^{n} c_k z^k$ rewritten as $z^n \sum_{k=0}^{n} c_{n-k} z^{-k}$ yields another terminating $q$-hypergeometric series, for instance:

$$\begin{aligned} &{}_{s+1}\phi_s\left[\begin{matrix} q^{-n},a_1,\ldots,a_s \\ b_1,\ldots,b_s\end{matrix};q,z\right] \\ &= (-1)^n q^{-(1/2)n(n+1)}\frac{(a_1,\ldots,a_n;q)_n}{(b_1,\ldots,b_s;q)_n} z^n \\ &\times {}_{s+1}\phi_s\left[\begin{matrix} q^{-n},q^{-n+1}b_1^{-1},\ldots,q^{-n+1}b_s^{-1} \\ q^{-n+1}a_1^{-1},\ldots,q^{-n+1}a_s^{-1}\end{matrix}; q,\frac{q^{n+1}b_1\cdots b_s}{a_1\cdots a_s z}\right] \end{aligned} \quad [12]$$

Often, in physics and quantum groups related literature, the following notation is used for $q$-number, $q$-factorial, and $q$-Pochhammer symbol:

$$[a]_q := \frac{q^{(1/2)a}-q^{-(1/2)a}}{q^{1/2}-q^{-1/2}} \qquad [k]_q! := \prod_{j=1}^{k}[j]_q$$
$$([a]_q)_k := \prod_{j=0}^{k-1}[a+j]_q \quad (k\in\mathbb{Z}_{\geq 0}) \quad [13]$$

For $q \to 1$, these symbols tend to their classical counterparts without the need for renormalization. They are expressed in terms of the standard notation [1] as follows:

$$\begin{aligned} [k]_q! &= q^{-(1/4)k(k-1)}\frac{(q;q)_k}{(1-q)^k} \\ ([a]_q)_k &= q^{-(1/2)k(a-1)}\,q^{-(1/4)k(k-1)}\frac{(q^a;q)_k}{(1-q)^k} \end{aligned} \quad [14]$$

### Special Cases

For $s = r-1$, formula [4] simplifies to

$$\begin{aligned} &{}_r\phi_{r-1}\left[\begin{matrix} a_1,\ldots,a_r \\ b_1,\ldots,b_{r-1}\end{matrix};q,z\right] \\ &= \sum_{k=0}^{\infty}\frac{(a_1,\ldots,a_r;q)_k}{(b_1,\ldots,b_{r-1};q)_k\,(q;q)_k}z^k \end{aligned} \quad [15]$$

which has radius of convergence 1 in the nonterminating case. The case $r = 2$ of [15] is the $q$-analog of the Gauss hypergeometric series.

#### $q$-Binomial series

$${}_1\phi_0(a;-;q,z) = \sum_{k=0}^{\infty}\frac{(a;q)_k z^k}{(q;q)_k} = \frac{(az;q)_\infty}{(z;q)_\infty}$$
(if series is not terminating, then $|z| < 1$) [16]

#### $q$-Exponential series

$$\begin{aligned} e_q(z) &:= {}_1\phi_0(0;-;q,z) \\ &= \sum_{k=0}^{\infty}\frac{z^k}{(q;q)_k} = \frac{1}{(z;q)_\infty} \quad (|z|<1) \end{aligned} \quad [17]$$

$$\begin{aligned} E_q(z) &:= {}_0\phi_0(-;-;q,-z) = \sum_{k=0}^{\infty}\frac{q^{(1/2)k(k-1)}z^k}{(q;q)_k} \\ &= (-z;q)_\infty = \left(e_q(-z)\right)^{-1} \quad (z\in\mathbb{C}) \end{aligned} \quad [18]$$

$$\begin{aligned} \varepsilon_q(z) &:= {}_1\phi_1(0;-q^{1/2};q^{1/2},-z) \\ &= \sum_{k=0}^{\infty}\frac{q^{(1/4)k(k-1)}}{(q;q)_k}z^k \quad (z\in\mathbb{C}) \end{aligned} \quad [19]$$

**Jackson's $q$-Bessel functions**

$$J_\nu^{(1)}(x;q) := \frac{(q^{\nu+1};q)_\infty}{(q;q)_\infty}\left(\frac{1}{2}x\right)^\nu \times {}_2\phi_1\left[\begin{matrix}0,0\\ q^{\nu+1}\end{matrix};q,-\frac{1}{4}x^2\right] \quad (0<x<2) \qquad [20]$$

$$J_\nu^{(2)}(x;q) := \frac{(q^{\nu+1};q)_\infty}{(q;q)_\infty}\left(\frac{1}{2}x\right)^\nu {}_0\phi_1\left[\begin{matrix}-\\ q^{\nu+1}\end{matrix};q,-\frac{1}{4}q^{\nu+1}x^2\right] = \left(-\frac{1}{4}x;q\right)_\infty J_\nu^{(1)}(x;q) \quad (x>0) \qquad [21]$$

$$J_\nu^{(3)}(x;q) := \frac{(q^{\nu+1};q)_\infty}{(q;q)_\infty}\left(\frac{1}{2}x\right)^\nu \times {}_1\phi_1\left[\begin{matrix}0\\ q^{\nu+1}\end{matrix};q,\frac{1}{4}qx^2\right] \quad (x>0) \qquad [22]$$

See [90] for the orthogonality relation for $J_\nu^{(3)}(x;q)$.

If $\exp_q(z)$ denotes one of the three $q$-exponentials [17]–[19], then $(1/2)(\exp_q(\mathrm{i}x)+\exp_q(-\mathrm{i}x))$ is a $q$-analog of the cosine and $-(1/2)\mathrm{i}(\exp_q(\mathrm{i}x) - \exp_q(-\mathrm{i}x))$ is a $q$-analog of the sine. The three $q$-cosines are essentially the case $\nu=-1/2$ of the corresponding $q$-Bessel functions [20]–[22], and the three $q$-sines are essentially the case $\nu=1/2$ of $x$ times the corresponding $q$-Bessel functions.

### *q*-Derivative and *q*-Integral

The $q$-derivative of a function $f$ given on a subset of $\mathbb{R}$ or $\mathbb{C}$ is defined by

$$(D_qf)(x) := \frac{f(x)-f(qx)}{(1-q)x} \quad (x\neq 0,\ q\neq 1) \qquad [23]$$

where $x$ and $qx$ should be in the domain of $f$. By continuity, we set $(D_qf)(0):=f'(0)$, provided $f'(0)$ exists. If $f$ is differentiable on an open interval $I$, then

$$\lim_{q\uparrow 1}(D_qf)(x) = f'(x) \quad (x\in I) \qquad [24]$$

For $a\in\mathbb{R}\backslash\{0\}$ and a function $f$ given on $(0,a]$ or $[a,0)$, we define the $q$-integral by

$$\int_0^a f(x)\,\mathrm{d}_qx := a(1-q)\sum_{k=0}^\infty f(aq^k)\,q^k = \sum_{k=0}^\infty f(aq^k)\,(aq^k-aq^{k+1}) \qquad [25]$$

provided the infinite sum converges absolutely (e.g., if $f$ is bounded). If $F(a)$ is given by the left-hand side of [25], then $D_qF=f$. The right-hand side of [25] is an infinite Riemann sum. For $q\uparrow 1$ it converges, at least formally, to $\int_0^a f(x)\,\mathrm{d}x$.

For nonzero $a,b\in\mathbb{R}$ we define

$$\int_a^b f(x)\,\mathrm{d}_qx := \int_0^b f(x)\,\mathrm{d}_qx - \int_0^a f(x)\,\mathrm{d}_qx \qquad [26]$$

For a $q$-integral over $(0,\infty)$, we have to specify a $q$-lattice $\{aq^k\}_{k\in\mathbb{Z}}$ for some $a>0$ (up to multiplication by an integer power of $q$):

$$\int_0^{a.\infty} f(x)\,\mathrm{d}_qx := a(1-q)\sum_{k=-\infty}^\infty f(aq^k)\,q^k = \lim_{n\to\infty}\int_0^{q^{-n}a} f(x)\,\mathrm{d}_qx \qquad [27]$$

### The *q*-Gamma and *q*-Beta Functions

The $q$-gamma function is defined by

$$\Gamma_q(z) := \frac{(q;q)_\infty(1-q)^{1-z}}{(q^z;q)_\infty} \quad (z\neq 0,-1,-2,\ldots) \qquad [28]$$

$$= \int_0^{(1-q)^{-1}} t^{z-1}E_q(-(1-q)qt)\,\mathrm{d}_qt \quad (\Re z>0) \qquad [29]$$

Then

$$\Gamma_q(z+1) = \frac{1-q^z}{1-q}\Gamma_q(z) \qquad [30]$$

$$\Gamma_q(n+1) = \frac{(q;q)_n}{(1-q)^n} \qquad [31]$$

$$\lim_{q\uparrow 1}\Gamma_q(z) = \Gamma(z) \qquad [32]$$

The $q$-beta function is defined by

$$B_q(a,b) := \frac{\Gamma_q(a)\Gamma_q(b)}{\Gamma_q(a+b)} = \frac{(1-q)\,(q,q^{a+b};q)_\infty}{(q^a,q^b;q)_\infty} \quad (a,b\neq 0,-1,-2,\ldots) \qquad [33]$$

$$= \int_0^1 t^{b-1}\frac{(qt;q)_\infty}{(q^at;q)_\infty}\,\mathrm{d}_qt \quad (\Re b>0,\ a\neq 0,-1,-2,\ldots) \qquad [34]$$

### The *q*-Gauss Hypergeometric Series

**$q$-Analog of Euler's integral representation**

$${}_2\phi_1(q^a,q^b;q^c;q,z) = \frac{\Gamma_q(c)}{\Gamma_q(a)\Gamma_q(c-b)}\int_0^1 t^{b-1}\frac{(tq;q)_\infty}{(tq^{c-b};q)_\infty} \times \frac{(tzq^a;q)_\infty}{tz;q)_\infty}\,\mathrm{d}_qt \quad (\Re b>0,\ |z|<1) \qquad [35]$$

By substitution of [25], formula [35] becomes a transformation formula:

$$ {}_2\phi_1(a,b;c;q,z) = \frac{(az;q)_\infty}{(z;q)_\infty}\frac{(b;q)_\infty}{(c;q)_\infty}\,{}_2\phi_1(c/b,z;az;q,b) \quad [36] $$

Note the mixing of argument $z$ and parameters $a,b,c$ on the right-hand side.

**Evaluation formulas in special points**

$$ {}_2\phi_1(a,b;c;q,c/(ab)) = \frac{(c/a,c/b;q)_\infty}{(c,c/(ab);q)_\infty} \quad (|c/(ab)|<1) \quad [37] $$

$$ {}_2\phi_1(q^{-n},b;c;q,cq^n/b) = \frac{(c/b;q)_n}{(c;q)_n} \quad [38] $$

$$ {}_2\phi_1(q^{-n},b;c;q,q) = \frac{(c/b;q)_n\,b^n}{(c;q)_n} \quad [39] $$

**Two general transformation formulas**

$$ {}_2\phi_1\left[\begin{matrix}a,b\\c\end{matrix};q,z\right] = \frac{(az;q)_\infty}{(z;q)_\infty}\,{}_2\phi_2\left[\begin{matrix}a,c/b\\c,az\end{matrix};q,bz\right] \quad [40] $$

$$ = \frac{(abz/c;q)_\infty}{(z;q)_\infty}\,{}_2\phi_1\left[\begin{matrix}c/a,c/b\\c\end{matrix};q,\frac{abz}{c}\right] \quad [41] $$

**Transformation formulas in the terminating case**

$$ {}_2\phi_1\left[\begin{matrix}q^{-n},b\\c\end{matrix};q,z\right] = \frac{(c/b;q)_n}{(c;q)_n}\,{}_3\phi_2\left[\begin{matrix}q^{-n},b,q^{-n}bc^{-1}z\\q^{1-n}bc^{-1},0\end{matrix};q,q\right] \quad [42] $$

$$ = (q^{-n}bc^{-1}z;q)_n\,{}_3\phi_2\left[\begin{matrix}q^{-n},cb^{-1},0\\c,qcb^{-1}z^{-1}\end{matrix};q,q\right] \quad [43] $$

$$ = \frac{(c/b;q)_n}{(c;q)_n}b^n\,{}_3\phi_1\left[\begin{matrix}q^{-n},b,qz^{-1}\\q^{1-n}bc^{-1}\end{matrix};q,\frac{z}{c}\right] \quad [44] $$

**Second order $q$-difference equation**

$$ z(q^c-q^{a+b+1}z)(D_q^2u)(z) + \left(\frac{1-q^c}{1-q} - \left(q^b\frac{1-q^a}{1-q} + q^a\frac{1-q^{b+1}}{1-q}\right)z\right)(D_qu)(z) - \frac{1-q^a}{1-q}\frac{1-q^b}{1-q}u(z) = 0 \quad [45] $$

Some special solutions of [45] are:

$$ u_1(z) := {}_2\phi_1(q^a,q^b;q^c;q,z) \quad [46] $$

$$ u_2(z) := z^{1-c}{}_2\phi_1(q^{1+a-c},q^{1+b-c};q^{2-c};q,z) \quad [47] $$

$$ u_3(z) := z^{-a}{}_2\phi_1(q^a,q^{a-c+1};q^{a-b+1};q,q^{-a-b+c+1}z^{-1}) \quad [48] $$

They are related by:

$$ u_1(z) + \frac{(q^a,q^{1-c},q^{c-b};q)_\infty}{(q^{c-1},q^{a-c+1},q^{1-b};q)_\infty} \times \frac{(q^{b-1}z,q^{2-b}z^{-1};q)_\infty}{(q^{b-c}z,q^{c-b+1}z^{-1};q)_\infty}u_2(z) = \frac{(q^{1-c},q^{a-b+1};q)_\infty}{(q^{1-b},q^{a-c+1};q)_\infty} \times \frac{(q^{a+b-c}z,q^{c-a-b+1}z^{-1};q)_\infty z^a}{(q^{b-c}z,q^{c-b+1}z^{-1};q)_\infty}u_3(z) \quad [49] $$

## Summation and Transformation Formulas for ${}_r\phi_{r-1}$ Series

An ${}_r\phi_{r-1}$ series [15] is called "balanced" if $b_1\ldots b_{r-1} = qa_1\ldots a_r$ and $z=q$, and the series is called "very well-poised" if $qa_1 = a_2b_1 = a_3b_2 = \cdots = a_rb_{r-1}$ and $qa_1^{1/2} = a_2 = -a_3$. The following more compact notation is used for very well-poised series:

$$ {}_rW_{r-1}(a_1;a_4,a_5,\ldots,a_r;q,z) := {}_r\phi_{r-1}\left[\begin{matrix}a_1,qa_1^{1/2},-qa_1^{1/2},a_4,\ldots,a_r\\a_1^{1/2},-a_1^{1/2},qa_1/a_4,\ldots,qa_1/a_r\end{matrix};q,z\right] \quad [50] $$

Below only a few of the most important identities are given. See Gasper and Rahman (2004) for many more. An important tool for obtaining complicated identities from more simple ones is Bailey's Lemma, which can moreover be iterated (Bailey chain), see Andrews (1986, ch.3).

**The $q$-Saalschütz sum for a terminating balanced ${}_3\phi_2$**

$$ {}_3\phi_2\left[\begin{matrix}a,b,q^{-n}\\c,q^{1-n}abc^{-1}\end{matrix};q,q\right] = \frac{(c/a,c/b;q)_n}{(c,c/(ab);q)_n} \quad [51] $$

**Jackson's sum for a terminating balanced ${}_8W_7$**

$$ {}_8W_7(a;b,c,d,q^{n+1}a^2/(bcd),q^{-n};q,q) = \frac{(qa,qa/(bc),qa/(bd),qa/(cd);q)_n}{(qa/b,qa/c,qa/d,qa/(bcd);q)_n} \quad [52] $$

**Watson's transformation of a terminating $_8W_7$ into a terminating balanced $_4\phi_3$**

$$
{}_8W_7\left(a;b,c,d,e,q^{-n};q,\frac{q^{n+2}a^2}{bcde}\right) = \frac{(qa,qa/(de);q)_n}{(qa/d,qa/e;q)_n} \times {}_4\phi_3\left[\begin{matrix} q^{-n},d,e,qa/(bc) \\ qa/b,qa/c,q^{-n}de/a \end{matrix};q,q\right] \qquad [53]
$$

**Sears' transformation of a terminating balanced $_4\phi_3$**

$$
{}_4\phi_3\left[\begin{matrix} q^{-n},a,b,c \\ d,e,f \end{matrix};q,q\right] = \frac{(e/a,f/a;q)_n}{(e,f;q)_n}a^n\,{}_4\phi_3\left[\begin{matrix} q^{-n},a,d/b,d/c \\ d,q^{1-n}a/e,q^{1-n}a/f \end{matrix};q,q\right] \qquad [54]
$$

By iteration and by symmetries in the upper and in the lower parameters, many other versions of this identity can be found. An elegant comprehensive formulation of all these versions is as follows.

Let $x_1x_2x_3x_4x_5x_6 = q^{1-n}$. Then the following expression is symmetric in $x_1,x_2,x_3,x_4,x_5,x_6$:

$$
\frac{q^{(1/2)n(n-1)}(x_1x_2x_3x_4,x_1x_2x_3x_5,x_1x_2x_3x_6;q)_n}{(x_1x_2x_3)^n} \times {}_4\phi_3\left[\begin{matrix} q^{-n},x_2x_3,x_1x_3,x_1x_2 \\ x_1x_2x_3x_4,x_1x_2x_3x_5,x_1x_2x_3x_6 \end{matrix};q,q\right] \qquad [55]
$$

Similar formulations involving symmetry groups can be given for other transformations, see Van der Jeugt and Srinivasa Rao (1999).

**Bailey's transformation of a terminating balanced $_{10}W_9$**

$$
{}_{10}W_9\left(a;b,c,d,e,f,\frac{q^{n+2}a^3}{bcdef},q^{-n};q,q\right) = \frac{(qa,qa/(ef),(qa)^2/(bcde),(qa)^2/(bcdf);q)_n}{(qa/e,qa/f,(qa)^2/(bcdef),(qa)^2/(bcd);q)_n} \times {}_{10}W_9\left(\frac{qa^2}{bcd};\frac{qa}{cd},\frac{qa}{bd},\frac{qa}{bc},e,f,\frac{q^{n+2}a^3}{bcdef},q^{-n};q,q\right) \qquad [56]
$$

### Rogers–Ramanujan Identities

$$
{}_0\phi_1(-;0;q,q) = \sum_{k=0}^{\infty}\frac{q^{k^2}}{(q;q)_k} = \frac{1}{(q,q^4;q^5)_\infty} \qquad [57]
$$

$$
{}_0\phi_1(-;0;q,q^2) = \sum_{k=0}^{\infty}\frac{q^{k(k+1)}}{(q;q)_k} = \frac{1}{(q^2,q^3;q^5)_\infty} \qquad [58]
$$

### Bilateral Series

Definition [1] can be extended by

$$
(a;q)_k := \frac{(a;q)_\infty}{(aq^k;q)_\infty} \quad (k\in\mathbb{Z}) \qquad [59]
$$

Define a bilateral $q$-hypergeometric series by the Laurent series

$$
{}_r\psi_s\left[\begin{matrix} a_1,\ldots,a_r \\ b_1,\ldots,b_s \end{matrix};q,z\right] = {}_r\psi_s(a_1,\ldots,a_r;b_1,\ldots,b_s;q,z) := \sum_{k=-\infty}^{\infty}\frac{(a_1,\ldots,a_r;q)_k}{(b_1,\ldots,b_s;q)_k}\left((-1)^k q^{(1/2)k(k-1)}\right)^{s-r} z^k
$$
$$
(a_1,\ldots,a_r,b_1,\ldots,b_s\neq 0,\ s\geq r) \qquad [60]
$$

The Laurent series is convergent if $|b_1\ldots b_s/(a_1\ldots a_r)| < |z|$ and moreover, for $s=r$, $|z|<1$.

**Ramanujan's $_1\psi_1$ summation formula**

$$
{}_1\psi_1(b;c;q,z) = \frac{(q,c/b,bz,q/(bz);q)_\infty}{(c,q/b,z,c/(bz);q)_\infty} \quad (|c/b| < |z < 1) \qquad [61]
$$

This has as a limit case

$$
{}_0\psi_1(-;c;q,z) = \frac{(q,z,q/z;q)_\infty}{(c,c/z;q)_\infty} \quad (|z| > |c|) \qquad [62]
$$

and as a further specialization the Jacobi triple product identity

$$
\sum_{k=-\infty}^{\infty}(-1)^k q^{(1/2)k(k-1)} z^k = (q,z,q/z;q)_\infty \quad (z\neq 0) \qquad [63]
$$

which can be rewritten as a product formula for a theta function:

$$
\theta_4(x;q) := \sum_{k=-\infty}^{\infty}(-1)^k q^{k^2} e^{2\pi ikx} = \prod_{k=1}^{\infty}(1-q^{2k}) \times\left(1-2q^{k-1}\cos(2\pi x)+q^{4k-2}\right) \qquad [64]
$$

## *q*-Hypergeometric Orthogonal Polynomials

Here we discuss families of orthogonal polynomials $\{p_n(x)\}$ which are expressible as terminating $q$-hypergeometric series $(0<q<1)$ and for which either (1) $P_n(x):=p_n(x)$ or (2) $P_n(x):=p_n$

$((1/2)(x+x^{-1}))$ are eigenfunctions of a second-order $q$-difference operator, that is,

$$A(x)\,P_n(qx) + B(x)\,P_n(x) + C(x)\,P_n(q^{-1}x) = \lambda_n\,P_n(x) \quad [65]$$

where $A(x), B(x)$, and $C(x)$ are independent of $n$, and where the $\lambda_n$ are the eigenvalues. The generic cases are the four-parameter classes of "Askey–Wilson polynomials" (continuous weight function) and $q$-Racah polynomials (discrete weights on finitely many points). They are of type (2) (quadratic $q$-lattice). All other cases can be obtained from the generic cases by specialization or limit transition. In particular, one thus obtains the generic three-parameter classes of type (1) (linear $q$-lattice). These are the big $q$-Jacobi polynomials (orthogonality by $q$-integral) and the $q$-Hahn polynomials (discrete weights on finitely many points).

#### Askey–Wilson Polynomials

**Definition as $q$-hypergeometric series**

$$p_n(\cos\theta) = p_n(\cos\theta; a,b,c,d\,|\,q) := \frac{(ab,ac,ad;q)_n}{a^n}\,{}_4\phi_3\left[\begin{matrix} q^{-n}, q^{n-1}abcd, ae^{i\theta}, ae^{-i\theta} \\ ab, ac, ad \end{matrix}; q, q\right] \quad [66]$$

This is symmetric in $a,b,c,d$.

**Orthogonality relation** Assume that $a,b,c,d$ are four reals, or two reals and one pair of complex conjugates, or two pairs of complex conjugates. Also assume that $|ab|, |ac|, |ad|, |bc|, |bd|, |cd| < 1$. Then

$$\int_{-1}^{1} p_n(x)p_m(x)w(x)\,dx + \sum_k p_n(x_k)\,p_m(x_k)\omega_k = h_n\,\delta_{n,m} \quad [67]$$

where

$$2\pi\sin\theta\,w(\cos\theta) = \left|\frac{(e^{2i\theta};q)_\infty}{(ae^{i\theta}, be^{i\theta}, ce^{i\theta}, de^{i\theta};q)_\infty}\right|^2 \quad [68]$$

$$h_0 = \frac{(abcd;q)_\infty}{(q,ab,ac,ad,bc,bd,cd;q)_\infty}$$

$$\frac{h_n}{h_0} = \frac{1-abcdq^{n-1}}{1-abcdq^{2n-1}} \times \frac{(q,ab,ac,ad,bc,bd,cd;q)_n}{(abcd;q)_n} \quad [69]$$

and the $x_k$ are the points $(1/2)(eq^k + e^{-1}q^{-k})$ with $e$ any of the $a,b,c,d$ of absolute value $>1$; the sum is over the $k \in \mathbb{Z}_{\geq 0}$ with $|eq^k| > 1$. The $\omega_k$ are certain weights which can be given explicitly. The sum in [67] does not occur if moreover $|a|, |b|, |c|, |d| < 1$.

A more uniform way of writing the orthogonality relation [67] is by the contour integral

$$\frac{1}{2\pi i}\oint_C p_n\left(\frac{1}{2}(z+z^{-1})\right)p_m\left(\frac{1}{2}(z+z^{-1})\right) \times \frac{(z^2, z^{-2};q)_\infty}{(az, az^{-1}, bz, bz^{-1}, cz, cz^{-1}, dz, dz^{-1};q)_\infty}\frac{dz}{z} = 2h_n\delta_{n,m} \quad [70]$$

where $C$ is the unit circle traversed in positive direction with suitable deformations to separate the sequences of poles converging to zero from the sequences of poles diverging to $\infty$.

The case $n=m=0$ of [70] or [67] is known as the Askey–Wilson integral.

***q*-Difference equation**

$$A(z)P_n(qz) - \big(A(z)+A(z^{-1})\big)P_n(z) + A(z^{-1})P_n(q^{-1}z) = (q^{-n}-1)(1-q^{n-1}abcd)P_n(z) \quad [71]$$

where $P_n(z) = p_n(\frac{1}{2}(z+z^{-1}))$ and $A(z) = (1-az)(1-bz)(1-cz)\,(1-dz)/((1-z^2)(1-qz^2))$

**Special cases** These include the continuous $q$-Jacobi polynomials (two parameters), the continuous $q$-ultraspherical polynomials (symmetric one-parameter case of continuous $q$-Jacobi), the Al-Salam-Chihara polynomials (Askey–Wilson with $c=d=0$), and the continuous $q$-Hermite polynomials (Askey–Wilson with $a=b=c=d=0$).

#### Continuous *q*-Ultraspherical Polynomials

**Definitions as finite Fourier series and as special Askey–Wilson polynomial**

$$C_n(\cos\theta;\beta\,|\,q) := \sum_{k=0}^{n}\frac{(\beta;q)_k(\beta;q)_{n-k}}{(q;q)_k(q;q)_{n-k}}\,e^{i(n-2k)\theta} \quad [72]$$

$$= \frac{(\beta;q)_n}{(q;q)_n}p_n(\cos\theta;\beta^{1/2}, q^{1/2}\beta^{1/2}, -\beta^{1/2}, -q^{1/2}\beta^{1/2}\mid q) \quad [73]$$

Orthogonality relation $(-1 < \beta < 1)$

$$\frac{1}{2\pi}\int_0^{\pi} C_n(\cos\theta;\beta,q)C_m(\cos\theta;\beta,q)\left|\frac{(e^{2i\theta};q)_\infty}{(\beta e^{2i\theta};q)_\infty}\right|^2 d\theta = \frac{(\beta,q\beta;q)_\infty}{(\beta^2,q;q)_\infty}\frac{1-\beta}{1-\beta q^n}\frac{(\beta^2;q)_n}{(q;q)_n}\delta_{n,m} \quad [74]$$

*q*-Difference equation

$$A(z)P_n(qz) - \left(A(z)+A(z^{-1})\right)P_n(z) + A(z^{-1})P_n(q^{-1}z) = (q^{-n}-1)(1-q^n\beta^2)P_n(z) \quad [75]$$

where $P_n(z) = C_n(\frac{1}{2}(z+z^{-1});\beta\,|\,q)$ and $A(z) = (1-\beta z^2)(1-q\beta z^2)/((1-z^2)(1-qz^2))$.

Generating function

$$\frac{(\beta e^{i\theta}z,\beta e^{-i\theta}z;q)_\infty}{(e^{i\theta}z,e^{-i\theta}z;q)_\infty} = \sum_{n=0}^{\infty} C_n(\cos\theta;\beta|q)z^n \quad (|z|<1, 0\le\theta\le\pi, -1<\beta<1) \quad [76]$$

Special case: the continuous *q*-Hermite polynomials

$$H_n(x|q) = (q;q)_n\, C_n(x;0|q) \quad [77]$$

Special cases: the Chebyshev polynomials

$$C_n(\cos\theta;q|q) = U_n(\cos\theta) := \frac{\sin((n+1)\theta)}{\sin\theta} \quad [78]$$

$$\lim_{\beta\uparrow 1}\frac{(q;q)_n}{(\beta;q)_n}C_n(\cos\theta;\beta|q) = T_n(\cos\theta) := \cos(n\theta)\ (n>0) \quad [79]$$

### *q*-Racah Polynomials

Definition as *q*-hypergeometric series $(n = 0, 1, \ldots, N)$

$$R_n(q^{-y}+\gamma\delta q^{y+1};\alpha,\beta,\gamma,\delta|q) := {}_4\phi_3\left[\begin{matrix} q^{-n},\alpha\beta q^{n+1},q^{-y},\gamma\delta q^{y+1} \\ q\alpha,q\beta\delta,q\gamma\end{matrix};q,q\right] \quad (\alpha,\beta\delta \text{ or } \gamma = q^{-N-1}) \quad [80]$$

Orthogonality relation

$$\sum_{y=0}^{N} R_n(q^{-y}+\gamma\delta q^{y+1})R_m(q^{-y}+\gamma\delta q^{y+1})\omega_y = h_n\delta_{n,m} \quad [81]$$

where $\omega_y$ and $h_n$ can be explicitly given.

### Big *q*-Jacobi Polynomials

Definition as *q*-hypergeometric series

$$P_n(x) = P_n(x;a,b,c;q) := {}_3\phi_2\left[\begin{matrix} q^{-n},q^{n+1}ab,x \\ qa,qc\end{matrix};q,q\right] \quad [82]$$

Orthogonality relation

$$\int_{qc}^{qa} P_n(x)P_m(x)\frac{(a^{-1}x,c^{-1}x;q)_\infty}{(x,bc^{-1}x;q)_\infty}d_qx = h_n\,\delta_{n,m}, \quad (0<a<q^{-1}, 0<b<q^{-1}, c<0) \quad [83]$$

where $h_n$ can be explicitly given.

*q*-Difference equation

$$A(x)P_n(qx) - (A(x)+C(x))P_n(x) + C(x)P_n(q^{-1}x) = (q^{-n}-1)(1-abq^{n+1})P_n(x) \quad [84]$$

where $A(x) = aq(x-1)(bx-c)/x^2$ and $C(x) = (x-qa)(x-qc)/x^2$

Limit case: Jacobi polynomials $P_n^{(\alpha,\beta)}(x)$

$$\lim_{q\uparrow 1} P_n(x;q^\alpha,q^\beta,-q^{-1}d;q) = \frac{n!}{(\alpha+1)_n}P_n^{(\alpha,\beta)}\left(\frac{2x+d-1}{d+1}\right) \quad [85]$$

Special case: the little *q*-Jacobi polynomials

$$p_n(x;a,b;q) = (-b)^{-n}q^{-(1/2)n(n+1)} \times \frac{(qb;q)_n}{(qa;q)_n}P_n(qbx;b,a,0;q) \quad [86]$$

$$= {}_2\phi_1(q^{-n},q^{n+1}ab;qa;q,qx) \quad [87]$$

which satisfy orthogonality relation (for $0 < a < q^{-1}$ and $b < q^{-1}$)

$$\int_0^1 p_n(x;a,b;q)p_m(x;a,b;q)\frac{(qx;q)_\infty}{(qbx;q)_\infty}x^{\log_q a}d_qx = \frac{(q,qab;q)_\infty}{(qa,qb;q)_\infty}\frac{(1-q)(qa)^n}{1-abq^{2n+1}}\frac{(q,qb;q)_n}{(qa,qab;q)_n}\delta_{n,m} \quad [88]$$

Limit case: Jackson's third *q*-Bessel function (see [22])

$$\lim_{N\to\infty} p_{N-n}(q^{N+k};q^\nu,b;q) = \frac{(q;q)_\infty}{(q^{\nu+1};q)_\infty}q^{-\nu(n+k)} \times J_\nu^{(3)}(2q^{(1/2)(n+k)};q) \quad (\nu > -1) \quad [89]$$

by which [88] tends to the orthogonality relation for $J_\nu^{(3)}(x;q)$:

$$\sum_{k=-\infty}^{\infty} J_\nu^{(3)}(2q^{(1/2)(n+k)};q)\,J_\nu^{(3)}(2q^{(1/2)(m+k)};q)q^k = \delta_{n,m}q^{-n} \quad (n,m\in\mathbb{Z}) \qquad [90]$$

### q-Hahn Polynomials

Definition as $q$-hypergeometric series

$$Q_n(x;\alpha,\beta,N;q) := {}_3\phi_2\left[\begin{matrix} q^{-n}, q^{n+1}\alpha\beta, x \\ q\alpha, q^{-N}\end{matrix}; q, q\right] \quad (n=0,1,\ldots,N) \qquad [91]$$

Orthogonality relation

$$\sum_{y=0}^{N} Q_n(q^{-y})Q_m(q^{-y})\frac{(q\alpha,q^{-N};q)_y(q\alpha\beta)^{-y}}{(q^{-N}\beta^{-1},q;q)_y} = h_n\delta_{n,m} \qquad [92]$$

where $h_n$ can be explicitly given.

### Stieltjes–Wigert Polynomials

Definition as $q$-hypergeometric series

$$S_n(x;q) = \frac{1}{(q;q)_n}\,{}_1\phi_1\left[\begin{matrix} q^{-n} \\ 0\end{matrix}; q, -q^{n+1}x\right] \qquad [93]$$

The orthogonality measure is not uniquely determined:

$$\int_0^\infty S_n(q^{1/2}x;q)S_m(q^{1/2}x;q)w(x)\,\mathrm{d}x = \frac{1}{q^n(q;q)_n}\delta_{n,m},$$

where, for instance

$$w(x) = \frac{q^{1/2}}{\log(q^{-1})(q,-q^{1/2}x,-q^{1/2}x^{-1};q)_\infty} \quad \text{or} \quad \frac{q^{1/2}}{\sqrt{2\pi\log(q^{-1})}}\exp\left(-\frac{\log^2 x}{2\log(q^{-1})}\right) \qquad [94]$$

### Rahman–Wilson Biorthogonal Rational Functions

The following functions are rational in their first argument:

$$R_n\left(\tfrac{1}{2}(z+z^{-1});a,b,c,d,e\right) := {}_{10}W_9(a/e;q/(be),q/(ce), q/(de),az,a/z,q^{n-1}abcd,q^{-n};q,q) \qquad [95]$$

They satisfy the biorthogonality relation

$$\frac{1}{2\pi\mathrm{i}}\oint_C R_n\left(\frac{1}{2}(z+z^{-1});a,b,c,d,e\right) \times R_m\left(\frac{1}{2}(z+z^{-1});a,b,c,d,\frac{q}{abcde}\right)w(z)\frac{\mathrm{d}z}{z} = 2h_n\delta_{n,m} \qquad [96]$$

where the contour $C$ is as in [70], and where

$$w(z) = \frac{(z^2,z^{-2},abcdez,abcde/z;q)_\infty}{(az,a/z,bz,b/z,cz,c/z,dz,d/z,ez,e/z;q)_\infty} \qquad [97]$$

$$h_0 = \frac{(bcde,acde,abde,abce,abcd;q)_\infty}{(q,ab,ac,ad,ae,bc,bd,be,cd,ce,de;q)_\infty} \qquad [98]$$

and $h_n/h_0$ can also be given explicitly. For $ab=q^{-N}, n,m\in\{0,1,\ldots,N\}$, there is a related discrete biorthogonality of the form

$$\sum_{k=0}^{N} R_n\left(\frac{1}{2}(aq^k+a^{-1}q^{-k});a,b,c,d,e\right) \times R_m\left(\frac{1}{2}(aq^k+a^{-1}q^{-k});a,b,c,d,\frac{q}{abcde}\right)w_k = 0 \quad (n\neq m) \qquad [99]$$

## Identities and Functions Associated with Root Systems

### η-Function Identities

Let $R$ be a root system on a Euclidean space of dimension $l$. Then Macdonald (1972) generalizes Weyl's denominator formula to the case of an affine root system. The resulting formula can be written as an explicit expansion in powers of $q$ of

$$\prod_{n=1}^{\infty}\left((1-q^n)^l\prod_{\alpha\in R}(1-q^ne^\alpha)\right)$$

which expansion takes the form of a sum over a lattice related to the root system. For root system $A_1$ this reduces to Jacobi's triple product identity [63]. Macdonald's formula implies a similar expansion in powers of $q$ of $\eta(q)^{l+|R|}$, where $\eta(q)$ is "Dedekind's $\eta$-function" $\eta(q) := q^{1/24}(q;q)_\infty$.

### Constant Term Identities

Let $R$ be a reduced root system, $R^+$ the positive roots, and $k\in\mathbb{Z}_{>0}$. Macdonald conjectured the second equality in

$$\frac{\int_T\prod_{\alpha\in R^+}(e^{-\alpha};q)_k(qe^\alpha;q)_k\,\mathrm{d}x}{\int_T\mathrm{d}x} = \mathrm{CT}\left(\prod_{\alpha\in R^+}\prod_{i=1}^{k}(1-q^{i-1}e^{-\alpha})(1-q^ie^\alpha)\right) = \prod_{i=1}^{l}\begin{bmatrix} kd_i \\ k\end{bmatrix}_q \qquad [100]$$

where $T$ is a torus determined by $R$, CT means the constant term in the Laurent expansion in $e^{\alpha}$, and the $d_i$ are the degrees of the fundamental invariants of the Weyl group of $R$. The conjecture was extended for real $k > 0$, for several parameters $k$ (one for each root length), and for root system $BC_n$, where Gustafson's five-parameter $n$-variable analog of the Askey–Wilson integral ([70] for $n = 0$) settles:

$$\int_{[0,2\pi]^n} |\Delta(e^{i\theta_1}, \ldots, e^{i\theta_n})|^2 \frac{d\theta_1 \ldots \theta_n}{(2\pi)^n} = 2^n n! \times \prod_{j=1}^{n} \frac{(t, t^{n+j-2}abcd; q)_\infty}{(t^j, q, abt^{j-1}, act^{j-1}, \ldots, cdt^{j-1}; q)_\infty} \quad [101]$$

where

$$\Delta(z) := \prod_{1 \le i < j \le n} \frac{(z_i z_j, z_i/z_j; q)_\infty}{(tz_i z_j, tz_i/z_j; q)_\infty} \times \prod_{j=1}^{n} \frac{(z_j^2; q)_\infty}{(az_j, bz_j, cz_j, dz_j; q)_\infty} \quad [102]$$

Further extensions were in Macdonald's conjectures for the quadratic norms of Macdonald polynomials associated with root systems (see the subsection "Macdonald–Koornwinder polynomials"), and finally proved by Cherednik.

### Macdonald Polynomials for Root System $A_{n-1}$

Let $n \in \mathbb{Z}_{>0}$. We work with partitions $\lambda = (\lambda_1, \ldots, \lambda_n)$ of length $\le n$, where $\lambda_1 \ge \cdots \ge \lambda_n \ge 0$ are integers. On the set of such partitions, we take the partial order $\lambda \le \mu \Rightarrow \lambda_1 + \cdots + \lambda_n = \mu_1 + \cdots + \mu_n$ and $\lambda_1 + \cdots + \lambda_i \le \mu_1 + \cdots + \mu_i$ ($i = 1, \ldots, n-1$). Write $\lambda < \mu$ iff $\lambda \le \mu$ and $\lambda \ne \mu$. The monomials are $z^\alpha = z_1^{\alpha_1} \ldots z_n^{\alpha_n}$ ($\alpha_1, \ldots, \alpha_n \in \mathbb{Z}_{\ge 0}$). For $\lambda$ a partition the symmetrized monomials $m_\lambda(z)$ and the Schur functions $s_\lambda(z)$ are defined by:

$$m_\lambda(z) := \sum_\alpha z^\alpha \quad \text{(sum over all distinct permutations } \alpha \text{ of } (\lambda_1, \ldots, \lambda_n)) \quad [103]$$

$$s_\lambda(z) := \frac{\det(z_i^{\lambda_j + n - j})_{i,j=1,\ldots,n}}{\det(z_i^{n-j})_{i,j=1,\ldots,n}} \quad [104]$$

We integrate a function over the torus $T := \{z \in \mathbb{C}^n \mid |z_1| = \cdots = |z_n| = 1\}$ as

$$\int_T f(z)\, dz := \frac{1}{(2\pi)^n} \times \int_0^{2\pi} \cdots \int_0^{2\pi} f(e^{i\theta_1}, \ldots, e^{i\theta_n}) d\theta_1 \ldots d\theta_n \quad [105]$$

**Definition** For $\lambda$ a partition and for $0 \le t \le 1$, the (analytically defined) Macdonald polynomial $P_\lambda(z) = P_\lambda(z; q, t)$ is of the form

$$P_\lambda(z) = P_\lambda(z; q, t) = m_\lambda(z) + \sum_{\mu < \lambda} u_{\lambda,\mu} m_\mu(z) \quad (u_{\lambda,\mu} \in \mathbb{C})$$

such that for all $\mu < \lambda$

$$\int_T P_\lambda(z)\, \overline{m_\mu(z)}\, \Delta(z)\, dz = 0$$

where

$$\Delta(z) = \Delta(z; q, t) := \prod_{i \ne j} \frac{(z_i z_j^{-1}; q)_\infty}{(tz_i z_j^{-1}; q)_\infty} \quad [106]$$

**Orthogonality relation**

$$\frac{1}{n!} \int_T P_\lambda(z)\, \overline{P_\mu(z)}\, \Delta(z)\, dz = \prod_{i<j} \frac{(q^{\lambda_i - \lambda_j} t^{j-i}, q^{\lambda_i - \lambda_j + 1} t^{j-i}; q)_\infty}{(q^{\lambda_i - \lambda_j} t^{j-i+1}, q^{\lambda_i - \lambda_j + 1} t^{j-i-1}; q)_\infty} \delta_{\lambda,\mu} \quad [107]$$

***q*-Difference equation**

$$\sum_{i=1}^{n} \prod_{j \ne i} \frac{tz_i - z_j}{z_i - z_j} \tau_{q,z_i} P_\lambda(z; q, t) = \left( \sum_{i=1}^{n} q^{\lambda_i} t^{n-i} \right) P_\lambda(z; q, t) \quad [108]$$

where $\tau_{q,z_i}$ is the $q$-shift operator: $\tau_{q,z_i} f(z_1, \ldots, z_n) := f(z_1, \ldots, qz_i, \ldots, z_n)$. See (Macdonald 1995, ch. VI, §3) for the full system of $q$-difference equations.

**Special value**

$$P_\lambda(1, t, \ldots, t^{n-1}; q, t) = \prod_{i=1}^{n} t^{(i-1)\lambda_i} \times \prod_{i<j} \frac{(tq^{j-i}; q)_{\lambda_i - \lambda_j}}{(q^{j-i}; q)_{\lambda_i - \lambda_j}} \quad [109]$$

**Restriction of number of variables**

$$P_{\lambda_1, \lambda_2, \ldots, \lambda_{n-1}, 0}(z_1, \ldots, z_{n-1}, 0; q, t) = P_{\lambda_1, \lambda_2, \ldots, \lambda_{n-1}}(z_1, \ldots, z_{n-1}; q, t) \quad [110]$$

**Homogeneity**

$$P_{\lambda_1, \ldots, \lambda_n}(z; q, t) = z_1 \ldots z_n P_{\lambda_1 - 1, \ldots, \lambda_n - 1}(z; q, t) \quad (\lambda_n > 0) \quad [111]$$

**Self-duality** Let $\lambda, \mu$ be partitions.

$$\frac{P_\lambda(q^{\mu_1}t^{n-1}, q^{\mu_2}t^{n-2}, \ldots, q^{\mu_n}; q, t)}{P_\lambda(t^{n-1}, t^{n-2}, \ldots, 1; q, t)} = \frac{P_\mu(q^{\lambda_1}t^{n-1}, q^{\lambda_2}t^{n-2}, \ldots, q^{\lambda_n}; q, t)}{P_\mu(t^{n-1}, t^{n-2}, \ldots, 1; q, t)} \quad [112]$$

**Special cases and limit relations**
*Continuous q-ultraspherical polynomials* (see [72]):

$$P_{m,n}(re^{i\theta}, re^{-i\theta}; q, t) = \frac{(q;q)_{m-n}}{(t;q)_{m-n}} r^{m+n} \times C_{m-n}(\cos\theta; t \mid q) \quad [113]$$

*Symmetrized monomials* (see [103]):

$$P_\lambda(z; q, 1) = m_\lambda(z) \quad [114]$$

*Schur functions* (see [104]):

$$P_\lambda(z; q, q) = s_\lambda(z) \quad [115]$$

*Hall–Littlewood polynomials* (see Macdonald (1995), ch. III):

$$P_\lambda(z; 0, t) = P_\lambda(z; t) \quad [116]$$

*Jack polynomials* (see Macdonald (1995), §VI.10):

$$\lim_{q\uparrow 1} P_\lambda(z; q, q^a) = P_\lambda^{(1/a)}(z) \quad [117]$$

**Algebraic definition of Macdonald polynomials** Macdonald polynomials can also be defined algebraically. We work now with partitions $\lambda$ $(\lambda_1 \geq \lambda_2 \geq \cdots \geq 0)$ of arbitrary length $l(\lambda)$, and with symmetric polynomials in arbitrarily many variables $x_1, x_2, \ldots,$ which can be canonically extended to symmetric functions in infinitely many variables $x_1, x_2, \ldots$. The $r$th power sum $p_r$ and the symmetric functions $p_\lambda$ are formally defined by

$$p_r = \sum_{i\geq 1} x_i^r, \quad p_\lambda = p_{\lambda_1} p_{\lambda_2} \ldots \quad [118]$$

Put

$$z_\lambda := \prod_{i\geq 1} i^{m_i} m_i! \quad \text{where } m_i = m_i(\lambda) \text{ is the number of parts of } \lambda \text{ equal to } i. \quad [119]$$

Define an inner product $\langle\,,\rangle_{q,t}$ on the space of symmetric functions such that

$$\langle p_\lambda, p_\mu \rangle_{q,t} = \delta_{\lambda,\mu} z_\lambda \prod_{i=1}^{l(\lambda)} \frac{1 - q^{\lambda_i}}{1 - t^{\lambda_i}} \quad [120]$$

For partitions $\lambda, \mu$ the partial ordering $\lambda \geq \mu$ means now that $\sum_{j\geq 1} \lambda_j = \sum_{j\geq 1} \mu_j$ and $\lambda_1 + \cdots + \lambda_i \geq \mu_1 + \cdots + \mu_i$ for all $i$. The Macdonald polynomial $P_\lambda(x; q, t)$ can now be algebraically defined as the unique symmetric function $P_\lambda$ of the form $P_\lambda = \sum_{\mu\leq\lambda} u_{\lambda,\mu} m_\mu$ $(u_{\lambda,\mu} \in \mathbb{C}, u_{\lambda,\lambda} = 1)$ such that

$$\langle P_\lambda, P_\mu \rangle_{q.t} = 0 \quad \text{if } \lambda \neq \mu \quad [121]$$

If $l(\lambda) \leq n$, then the newly defined $P_\lambda(x)$ with $x_{n+1} = x_{n+2} = \cdots = 0$ coincides with $P_\lambda(x; q, t)$ defined analytically, and the new inner product is a constant multiple (depending on $n$) of the old inner product.

**Bilinear sum**

$$\sum_\lambda \frac{1}{\langle P_\lambda, P_\lambda \rangle_{q,t}} P_\lambda(x; q, t) P_\lambda(y; q, t) = \prod_{i,j\geq 1} \frac{(tx_i y_j; q)_\infty}{(x_i y_j; q)_\infty} \quad [122]$$

**Generalized Kostka numbers** The Kostka numbers $K_{\lambda,\mu}$ occurring as expansion coefficients in $s_\lambda = \sum_\mu K_{\lambda,\mu} m_\mu$ were generalized by Macdonald to coefficients $K_{\lambda,\mu}(q,t)$ occurring in connection with Macdonald polynomials, see Macdonald (1995, §VI.8). Macdonald's conjecture that $K_{\lambda,\mu}(q,t)$ is a polynomial in $q$ and $t$ with coefficients in $\mathbb{Z}_{\geq 0}$ was fully proved in Haiman (2001).

### Macdonald–Koornwinder Polynomials

Macdonald (2000, 2001) also introduced Macdonald polynomials associated with an arbitrary root system. For root system $BC_n$ this yields a three-parameter family which can be extended to the five-parameter Macdonald–Koornwinder (M–K) polynomials (Koornwinder 1992). They are orthogonal with respect to the measure occurring in [101] with $\Delta(z)$ given by [102]. The M–K polynomials are $n$-variable analogs of the Askey–Wilson polynomials. All polynomials just discussed tend, for $q \uparrow 1$, to Jacobi polynomials associated with root systems.

Macdonald conjectured explicit expressions for the quadratic norms of the Macdonald polynomials associated with root systems and of the M–K polynomials. These were proved by Cherednik by considering these polynomials as Weyl group symmetrizations of non-invariant polynomials

which are related to double affine Hecke algebras (see Macdonald (2003)).

## Elliptic Hypergeometric Series

Let $p, q \in \mathbb{C}, |p|, |q| < 1$. Define a modified Jacobi theta function by

$$\theta(x;p) := (x, p/x; p)_\infty \ (x \neq 0) \qquad [123]$$

and the elliptic shifted factorial by

$$(a;q,p)_k := \theta(a;p)\theta(aq;p)\ldots\theta(aq^{k-1};p) \quad (k \in \mathbb{Z}_{>0}),\ (a;q,p)_0 := 1 \qquad [124]$$

$$(a_1,\ldots,a_r;q,p)_k := (a_1;q,p)_k \ldots (a_r;q,p)_k \qquad [125]$$

where $a, a_1, \ldots, a_r \neq 0$. For $q = e^{2\pi i\sigma}$, $p = e^{2\pi i\tau}$ ($\Im\tau > 0$), and $a \in \mathbb{C}$ we have

$$\frac{\theta(ae^{2\pi i\sigma(x+\sigma^{-1})}; e^{2\pi i\tau})}{\theta(ae^{2\pi i\sigma x}; e^{2\pi i\tau})} = 1$$
$$\frac{\theta(ae^{2\pi i\sigma(x+\tau\sigma^{-1})}; e^{2\pi i\tau})}{\theta(ae^{2\pi i\sigma x}; e^{2\pi i\tau})} = -a^{-1}q^{-x} \qquad [126]$$

A series $\sum_{k=0}^{\infty} c_k$ with $c_{k+1}/c_k$ being an elliptic (i.e., doubly periodic meromorphic) function of $k$ considered as a complex variable is called an elliptic hypergeometric series. In particular, define the ${}_rE_{r-1}$ theta hypergeometric series as the formal series

$${}_rE_{r-1}(a_1,\ldots,a_r;b_1,\ldots,b_{r-1};q,p;z) := \sum_{k=0}^{\infty} \frac{(a_1,\ldots,a_r;q,p)_k}{(b_1,\ldots,b_{r-1};q,p)_k} \frac{z^k}{(q;q,p)_k} \qquad [127]$$

It has $g(k) := c_{k+1}/c_k$ with

$$g(x) = \frac{z\theta(a_1q^x;p)\ldots\theta(a_rq^x;p)}{\theta(q^{x+1};p)\,\theta(b_1q^x;p)\ldots\theta(b_{r-1}q^x;p)}$$

By [126], $g(x)$ is an elliptic function with periods $\sigma^{-1}$ and $\tau\sigma^{-1}$ ($q = e^{2\pi i\sigma}, p = e^{2\pi i\tau}$) if the balancing condition $a_1 \ldots a_r = qb_1 \ldots b_{r-1}$ is satisfied.

The ${}_rV_{r-1}$ very well-poised theta hypergeometric series (a special ${}_rE_{r-1}$) is defined, in case of argument 1, as:

$${}_rV_{r-1}(a_1;a_6,\ldots,a_r;q,p) := \sum_{k=0}^{\infty} \frac{\theta(a_1q^{2k};p)}{\theta(a_1;p)} \frac{(a_1,a_6,\ldots,a_r;q,p)_k}{(qa_1/a_6,\ldots,qa_1/a_r;q,p)_k} \times \frac{q^k}{(q;q,p)_k} \qquad [128]$$

The series is called balanced if $a_6^2 \ldots a_r^2 = a_1^{r-6}q^{r-4}$. The series terminates if, for instance, $a_r = q^{-n}$.

#### Elliptic Analog of Jackson's ${}_8W_7$ Summation

$${}_{10}V_9(a;b,c,d,q^{n+1}a^2/(bcd),q^{-n};q,p) = \frac{(qa,qa/(bc),qa/(bd),qa/(cd);q,p)_n}{(qa/b,qa/c,qa/d,qa/(bcd);q,p)_n} \qquad [129]$$

#### Elliptic Analog of Bailey's ${}_{10}W_9$ Transformation

$${}_{12}V_{11}\left(a;b,c,d,e,f,\frac{q^{n+2}a^3}{bcdef},q^{-n};q,p\right) = \frac{(qa,qa/(ef),(qa)^2/(bcde),(qa)^2/(bcdf);q,p)_n}{(qa/e,qa/f,(qa)^2/(bcdef),(qa)^2/(bcd);q,p)_n} \times {}_{12}V_{11}\left(\frac{qa^2}{bcd};\frac{qa}{cd},\frac{qa}{bd},\frac{qa}{bc},e,f,\frac{q^{n+2}a^3}{bcdef},q^{-n};q,p\right) \qquad [130]$$

Suitable ${}_{12}V_{11}$ functions satisfy a discrete biorthogonality relation which is an elliptic analog of [99].

#### Ruijsenaars' elliptic gamma function

$$\Gamma(z;q,p) := \prod_{j,k=0}^{\infty} \frac{1-z^{-1}q^{j+1}p^{k+1}}{1-zq^jp^k} \qquad [131]$$

which is symmetric in $p$ and $q$. Then

$$\Gamma(qz;q,p) = \theta(z;p)\Gamma(z;q,p)$$
$$\Gamma(q^nz;q,p) = (z;q,p)_n\Gamma(z;q,p) \qquad [132]$$

## Applications

#### Quantum Groups

A specific quantum group is usually a Hopf algebra which is a $q$-deformation of the Hopf algebra of functions on a specific Lie group or, dually, of a universal enveloping algebra (viewed as Hopf algebra) of a Lie algebra. The general philosophy is that representations of the Lie group or Lie algebra also deform to representations of the quantum group, and that special functions associated with the representations in the classical case deform to $q$-special functions associated with the representations in the quantum case. Sometimes this is straightforward, but often new subtle phenomena occur.

The representation-theoretic objects which may be explicitly written in terms of $q$-special functions include matrix elements of representations with respect to specific bases (in particular spherical elements), Clebsch–Gordan coefficients and Racah coefficients. Many one-variable $q$-hypergeometric functions have found interpretation in some way in connection with a quantum analog of a three-dimensional Lie group (generically the Lie group

SL(2, $\mathbb{C}$) and its real forms). Classical by now are: little $q$-Jacobi polynomials interpreted as matrix elements of irreducible representations of $SU_q(2)$ with respect to the standard basis; Askey–Wilson polynomials similarly interpreted with respect to a certain basis not coming from a quantum subgroup; Jackson's third $q$-Bessel functions as matrix elements of irreducible representations of $E_q(2)$; $q$-Hahn polynomials and $q$-Racah polynomials interpreted as Clebsch–Gordan coefficients and Racah coefficients, respectively, for $SU_q(2)$.

Further developments include: Macdonald polynomials as spherical elements on quantum analogs of compact Riemannian symmetric spaces; $q$-analogs of Jacobi functions as matrix elements of irreducible unitary representations of $SU_q(1,1)$; Askey–Wilson polynomials as matrix elements of representations of the SU(2) dynamical quantum group; an interpretation of discrete ${}_{12}V_{11}$ biorthogonality relations on the elliptic U(2) quantum group.

Since the $q$-deformed Hopf algebras are usually presented by generators and relations, identities for $q$-special functions involving noncommuting variables satisfying simple relations are important for further interpretations of $q$-special functions in quantum groups, for instance:

**$q$-Binomial formula with $q$-commuting variables**

$$(x+y)^n = \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix}_q y^{n-k} x^k \quad (xy = qyx) \qquad [133]$$

**Functional equations for $q$-exponentials with $xy = qyx$**

$$\begin{aligned} e_q(x+y) &= e_q(y) e_q(x) \\ E_q(x+y) &= E_q(x) E_q(y) \end{aligned} \qquad [134]$$

$$\begin{aligned} e_q(x+y-yx) &= e_q(x) e_q(y) \\ E_q(x+y+yx) &= E_q(y) E_q(x) \end{aligned} \qquad [135]$$

### Various Algebraic Settings

**Classical groups over finite fields (Chevalley groups)** $q$-Hahn polynomials and various kinds of $q$-Krawtchouk polynomials have interpretations as spherical and intertwining functions on classical groups ($GL_n, SO_n, Sp_n$) over a finite field $F_q$ with respect to suitable subgroups, see Stanton (1984).

**Affine Kac–Moody algebras (see Lepowsky (1982))** The Rogers–Ramanujan identities [57], [58] and some of their generalizations were interpreted in the context of characters of representations of the simplest affine Kac–Moody algebra $A_1^{(1)}$. Macdonald's generalization of Weyl's denominator formula to affine root systems has an interpretation as an identity for the denominator of the character of a representation of an affine Kac–Moody algebra.

### Partitions of Positive Integers

Let $n$ be a positive integer, $p(n)$ the number of partitions of $n$, $p_N(n)$ the number of partitions of $n$ into parts $\le N$, $p_{\text{dist}}(n)$ the number of partitions of $n$ into distinct parts, and $p_{\text{odd}}(n)$ the number of partitions of $n$ into odd parts. Then, Euler observed:

$$\frac{1}{(q;q)_\infty} = \sum_{n=0}^{\infty} p(n) q^n \quad \frac{1}{(q;q)_N} = \sum_{n=0}^{\infty} p_N(n) q^n \qquad [136]$$

$$\begin{aligned} (-q;q)_\infty &= \sum_{n=0}^{\infty} p_{\text{dist}}(n) q^n \\ \frac{1}{(q;q^2)_\infty} &= \sum_{n=0}^{\infty} p_{\text{odd}}(n) q^n \end{aligned} \qquad [137]$$

and

$$(-q;q)_\infty = \frac{1}{(q;q^2)_\infty}, \qquad p_{\text{dist}}(n) = p_{\text{odd}}(n) \qquad [138]$$

The Rogers–Ramanujan identity [57] has the following partition-theoretic interpretation: the number of partitions of $n$ with parts differing at least 2 equals the number of partitions of $n$ into parts congruent to 1 or 4 (mod 5). Similarly, [58] yields: the number of partitions of $n$ with parts larger than 1 and differing at least 2 equals the number of partitions of $n$ into parts congruent to 2 or 3 (mod 5).

The left-hand sides of the Rogers–Ramanujan identities [57] and [58] have interpretations in the "hard hexagon model," see Baxter (1982). Much further work has been done on Rogers–Ramanujan-type identities in connection with more general models in statistical mechanics. The so-called "fermionic expressions" do occur.

*See also:* Combinatorics: Overview; Eight Vertex and Hard Hexagon Models; Hopf Algebras and $q$-Deformation Quantum Groups; Integrable Systems: Overview; Ordinary Special Functions; Solitons and Kac–Moody Lie Algebras.

## Further Reading

Andrews GE (1986) q-Series: Their Development and Application in Analysis, Number Theory, Combinatorics, Physics, and Computer Algebra, *CBMS Regional Conference Series in Mathematics*, vol. 66. Providence, RI: American Mathematical Society.

Andrews GE, Askey R, and Roy R (1999) *Special Functions*. Cambridge: Cambridge University Press.
Andrews GE and Eriksson K (2004) *Integer Partitions*. Cambridge: Cambridge University Press.
Baxter RJ (1982) *Exactly Solved Models in Statistical Mechanics*. London: Academic Press.
Gasper G and Rahman M (2004) *Basic Hypergeometric Series*, 2nd edn. Cambridge: Cambridge University Press.
Haiman M (2001) Hilbert schemes, polygraphs and the Macdonald positivity conjecture. *Journal of the American Mathematical Society* 14: 941–1006.
Koekoek R and Swarttouw RF (1998) *The Askey-Scheme of Hypergeometric Orthogonal Polynomials and Its q-Analogue. Report 98-17*, Faculty of Technical Mathematics and Informatics, Delft University of Technology.
Koornwinder TH (1992) Askey–Wilson polynomials for root systems of type *BC*. In: Richards DStP (ed.) *Hypergeometric Functions on Domains of Positivity, Jack Polynomials, and Applications*, Contemp. Math., vol. 138, pp. 189–204. Providence, RI: American Mathematical Society.
Koornwinder TH (1994) Compact quantum groups and $q$-special functions. In: Baldoni V and Picardello MA (eds.) *Representations of Lie Groups and Quantum Groups*, Pitman Research Notes in Mathematics Series, vol. 311, pp. 46–128. Harlow: Longman Scientific & Technical.
Koornwinder TH (1997) Special functions and $q$-commuting variables. In: Ismail MEH, Masson DR, and Rahman M (eds.) *Special Functions, q-Series and Related Topics*, Fields Institute Communications, vol. 14, pp. 131–166. Providence, RI: American Mathematical Society.
Lepowsky J (1982) Affine Lie algebras and combinatorial identities. In: Winter DJ (ed.) *Lie Algebras and Related Topics*, Lecture Notes in Math., vol. 933, pp. 130–156. Berlin: Springer.
Macdonald IG (1972) Affine root systems and Dedekind's $\eta$-function. *Inventiones Mathematicae* 15: 91–143.
Macdonald IG (1995) *Symmetric Functions and Hall Polynomials*, 2nd edn. Oxford: Clarendon.
Macdonald IG (2000, 2001) Orthogonal polynomials associated with root systems. *Séminaire Lotharingien de Combinatoire* 45: Art. B45a.
Macdonald IG (2003) *Affine Hecke Algebras and Orthogonal Polynomials*. Cambridge: Cambridge University Press.
Stanton D (1984) Orthogonal polynomials and Chevalley groups. In: Askey RA, Koornwinder TH, and Schempp W (eds.) *Special Functions: Group Theoretical Aspects and Applications*, pp. 87–128. Dordrecht: Reidel.
Suslov SK (2003) *An Introduction to Basic Fourier Series*. Dordrecht: Kluwer Academic Publishers.
Van der Jeugt J and Srinivasa Rao K (1999) Invariance groups of transformations of basic hypergeometric series. *Journal of Mathematical Physics* 40: 6692–6700.
Vilenkin NJ and Klimyk AU (1992) *Representation of Lie Groups and Special Functions*, vol. 3. Dordrecht: Kluwer Academic Publishers.

# Quantum 3-Manifold Invariants

**C Blanchet**, Université de Bretagne-Sud, Vannes, France
**V Turaev**, IRMA, Strasbourg, France



## Introduction

The idea to derive topological invariants of smooth manifolds from partition functions of certain action functionals was suggested by A Schwarz (1978) and highlighted by E Witten (1988). Witten interpreted the Jones polynomial of links in the 3-sphere $S^3$ as a partition function of the Chern–Simons field theory. Witten conjectured the existence of mathematically defined topological invariants of 3-manifolds, generalizing the Jones polynomial (or rather its values in complex roots of unity) to links in arbitrary closed oriented 3-manifolds. A rigorous construction of such invariants was given by N Reshetikhin and V Turaev (1989) using the theory of quantum groups. The Witten–Reshetikhin–Turaev invariants of 3-manifolds, also called the "quantum invariants," extend to a topological quantum field theory (TQFT) in dimension 3.

## Ribbon and Modular Categories

The Reshetikhin–Turaev approach begins with fixing suitable algebraic data, which are best described in terms of monoidal categories. Let $\mathcal{C}$ be a monoidal category (i.e., a category with an associative tensor product and unit object $\mathbb{1}$). A "braiding" in $\mathcal{C}$ assigns to any objects $V, W \in \mathcal{C}$ an invertible morphism $c_{V,W} : V \otimes W \to W \otimes V$ such that, for any $U, V, W \in \mathcal{C}$,

$$c_{U,V\otimes W} = (\mathrm{id}_V \otimes c_{U,W})(c_{U,V} \otimes \mathrm{id}_W)$$
$$c_{U\otimes V,W} = (c_{U,W} \otimes \mathrm{id}_V)(\mathrm{id}_U \otimes c_{V,W})$$

A "twist" in $\mathcal{C}$ assigns to any object $V \in \mathcal{C}$ an invertible morphism $\theta_V : V \to V$ such that, for any $V, W \in \mathcal{C}$,

$$\theta_{V\otimes W} = c_{W,V}\, c_{V,W}(\theta_V \otimes \theta_W)$$

A "duality" in $\mathcal{C}$ assigns to any object $V \in \mathcal{C}$ a "dual" object $V^* \in \mathcal{C}$, and evaluation and co-evaluation morphisms $d_V : V^* \otimes V \to \mathbb{1}$, $b_V : \mathbb{1} \to V \otimes V^*$ such that

$$(\mathrm{id}_V \otimes d_V)(b_V \otimes \mathrm{id}_V) = \mathrm{id}_V$$
$$(d_V \otimes \mathrm{id}_{V^*})(\mathrm{id}_{V^*} \otimes b_V) = \mathrm{id}_{V^*}$$

The category $\mathcal{C}$ with duality, braiding, and twist is ribbon, if for any $V \in \mathcal{C}$,

$$(\theta_V \otimes \mathrm{id}_{V^*})b_V = (\mathrm{id}_V \otimes \theta_{V^*})b_V$$

For an endomorphism $f: V \to V$ of an object $V \in \mathcal{C}$, its trace "$\mathrm{tr}(f) \in \mathrm{End}_{\mathcal{C}}(\mathbb{1})$" is defined as

$$\mathrm{tr}(f) = d_V c_{V,V^*}((\theta_V f) \otimes \mathrm{id}_{V^*})b_V : \mathbb{1} \to \mathbb{1}$$

This trace shares a number of properties of the standard trace of matrices, in particular, $\mathrm{tr}(fg) = \mathrm{tr}(gf)$ and $\mathrm{tr}(f \otimes g) = \mathrm{tr}(f)\mathrm{tr}(g)$. For an object $V \in \mathcal{C}$, set

$$\dim(V) = \mathrm{tr}(\mathrm{id}_V) = d_V c_{V,V^*}(\theta_V \otimes \mathrm{id}_{V^*})b_V$$

Ribbon categories nicely fit the theory of knots and links in $S^3$. A link $L \subset S^3$ is a closed one-dimensional submanifold of $S^3$. (A manifold is closed if it is compact and has no boundary.) A link is oriented (resp. framed) if all its components are oriented (resp. provided with a homotopy class of nonsingular normal vector fields). Given a framed oriented link $L \subset S^3$ whose components are labeled with objects of a ribbon category $\mathcal{C}$, one defines a tensor $\langle L \rangle \in \mathrm{End}_{\mathcal{C}}(\mathbb{1})$. To compute $\langle L \rangle$, present $L$ by a plane diagram with only double transversal crossings such that the framing of $L$ is orthogonal to the plane. Each double point of the diagram is an intersection of two branches of $L$, going over and under, respectively. Associate with such a crossing the tensor $(c_{V,W})^{\pm 1}$ where $V, W \in \mathcal{C}$ are the labels of these two branches and $\pm 1$ is the sign of the crossing determined by the orientation of $L$. We also associate certain tensors with the points of the diagram where the tangent line is parallel to a fixed axis on the plane. These tensors are derived from the evaluation and co-evaluation morphisms and the twists. Finally, all these tensors are contracted into a single element $\langle L \rangle \in \mathrm{End}_{\mathcal{C}}(\mathbb{1})$. It does not depend on the intermediate choices and is preserved under isotopy of $L$ in $S^3$. For the trivial knot $O(V)$ with framing 0 and label $V \in \mathcal{C}$, we have $\langle O(V) \rangle = \dim(V)$.

Further constructions need the notion of a tangle. An (oriented) tangle is a compact (oriented) one-dimensional submanifold of $\boldsymbol{R}^2 \times [0,1]$ with endpoints on $\boldsymbol{R} \times 0 \times \{0,1\}$. Near each of its endpoints, an oriented tangle $T$ is directed either down or up, and thus acquires a sign $\pm 1$. One can view $T$ as a morphism from the sequence of $\pm 1$'s associated with its bottom ends to the sequence of $\pm 1$'s associated with its top ends. Tangles can be composed by putting one on top of the other. This defines a category of tangles $\mathcal{T}$ whose objects are finite sequences of $\pm 1$'s and whose morphisms are isotopy classes of framed oriented tangles. Given a ribbon category $\mathcal{C}$, we can consider $\mathcal{C}$-labeled tangles, that is, (framed oriented) tangles whose components are labeled with objects of $\mathcal{C}$. They form a category $\mathcal{T}_{\mathcal{C}}$. Links appear here as tangles without endpoints, that is, as morphisms $\emptyset \to \emptyset$. The link invariant $\langle L \rangle$ generalizes to a functor $\langle \cdot \rangle : \mathcal{T}_{\mathcal{C}} \to \mathcal{C}$.

To define 3-manifold invariants, we need modular categories (Turaev 1994). Let $\boldsymbol{k}$ be a field. A monoidal category $\mathcal{C}$ is $\boldsymbol{k}$-additive if its Hom sets are $\boldsymbol{k}$-vector spaces, the composition and tensor product of the morphisms are bilinear, and $\mathrm{End}_{\mathcal{C}}(\mathbb{1}) = \boldsymbol{k}$. An object $V \in \mathcal{C}$ is simple if $\mathrm{End}_{\mathcal{C}}(V) = \boldsymbol{k}$. A modular category is a $\boldsymbol{k}$-additive ribbon category $\mathcal{C}$ with a finite family of simple objects $\{V_\lambda\}_\lambda$ such that (1) for any object $V \in \mathcal{C}$ there is a finite expansion $\mathrm{id}_V = \sum_i f_i g_i$ for certain morphisms $g_i : V \to V_{\lambda_i}$, $f_i : V_{\lambda_i} \to V$ and (2) the $S$-matrix $(S_{\lambda,\mu})$ is invertible over $\boldsymbol{k}$ where $S_{\lambda,\mu} = \mathrm{tr}(c_{V_\lambda,V_\mu} c_{V_\mu,V_\lambda})$. Note that $S_{\lambda,\mu} = \langle H(\lambda,\mu) \rangle$ where $H(\lambda,\mu)$ is the oriented Hopf link with framing 0, linking number +1, and labels $V_\lambda$, $V_\mu$.

Axiom (1) implies that every simple object in $\mathcal{C}$ is isomorphic to exactly one of $V_\lambda$. In most interesting cases (when there is a well-defined direct summation in $\mathcal{C}$), this axiom may be rephrased by saying that $\mathcal{C}$ is finite semisimple, that is, $\mathcal{C}$ has a finite set of isomorphism classes of simple objects and all objects of $\mathcal{C}$ are direct sums of simple objects. A weaker version of the axiom (2) yields premodular categories.

The invariant $\langle \cdot \rangle$ of links and tangles extends by linearity to the case where labels are finite linear combinations of objects of $\mathcal{C}$ with coefficients in $\boldsymbol{k}$. Such a linear combination $\Omega = \sum_\lambda \dim(V_\lambda) V_\lambda$ is called the Kirby color. It has the following sliding property: for any object $V \in \mathcal{C}$, the two tangles in Figure 1 yield the same morphism $V \to V$. Here, the dashed line represents an arc on the closed component labeled by $\Omega$. This arc can be knotted or linked with other components of the tangle (not shown in the figure).

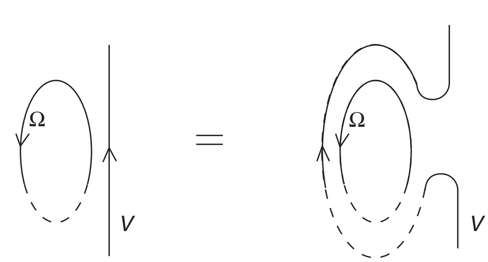



**Figure 1** Sliding property.

## Invariants of Closed 3-Manifolds

Given an embedded solid torus $g: S^1 \times D^2 \hookrightarrow S^3$, where $D^2$ is a 2-disk and $S^1 = \partial D^2$, a 3-manifold can be built as follows. Remove from $S^3$ the interior of $g(S^1 \times D^2)$ and glue back the solid torus $D^2 \times S^1$ along $g|_{S^1 \times S^1}$. This process is known as "surgery." The resulting 3-manifold depends only on the isotopy class of the framed knot represented by $g$. More generally, a surgery on a framed link $L = \cup_{i=1}^m L_i$ in $S^3$ with $m$ components yields a closed oriented 3-manifold $M_L$. A theorem of W Lickorish and A Wallace asserts that any closed connected oriented 3-manifold is homeomorphic to $M_L$ for some $L$. R Kirby proved that two framed links give rise to homeomorphic 3-manifolds if and only if these links are related by isotopy and a finite sequence of geometric transformations called Kirby moves. There are two Kirby moves: adjoining a distant unknot $O^\varepsilon$ with framing $\varepsilon = \pm 1$, and sliding a link component over another one as in Figure 1.

Let $L = \cup_{i=1}^m L_i \subset S^3$ be a framed link and let $(b_{i,j})_{i,j=1,\ldots,m}$ be its linking matrix: for $i \neq j, b_{i,j}$ is the linking number of $L_i, L_j$, and $b_{i,i}$ is the framing number of $L_i$. Denote by $e_+$ (resp. $e_-$) the number of positive (resp. negative) eigenvalues of this matrix. The sliding property of modular categories implies the following theorem. In its statement, a knot $K$ with label $\Omega$ is denoted by $K(\Omega)$.

**Theorem 1** *Let $\mathcal{C}$ be a modular category with Kirby color $\Omega$. Then $\langle O^1(\Omega)\rangle \neq 0, \langle O^{-1}(\Omega)\rangle \neq 0$ and the expression*

$$\tau_{\mathcal{C}}(M_L) = \langle O^1(\Omega)\rangle^{-e_+} \langle O^{-1}(\Omega)\rangle^{-e_-} \langle L_1(\Omega), \ldots, L_m(\Omega)\rangle$$

*is invariant under the Kirby moves on L. This expression yields, therefore, a well-defined topological invariant $\tau_{\mathcal{C}}$ of closed connected oriented 3-manifolds.*

Several competing normalizations of $\tau_{\mathcal{C}}$ exist in the literature. Here, the normalization used is such that $\tau_{\mathcal{C}}(S^3) = 1$ and $\tau_{\mathcal{C}}(S^1 \times S^2) = \sum_\lambda (\dim(V_\lambda))^2$. The invariant $\tau_{\mathcal{C}}$ extends to 3-manifolds with a framed oriented $\mathcal{C}$-labeled link $K$ inside by

$$\begin{aligned}\tau_{\mathcal{C}}&(M_L, K) \\ &= \langle O^1(\Omega)\rangle^{-e_+} \langle O^{-1}(\Omega)\rangle^{-e_-} \langle L_1(\Omega), \ldots, L_m(\Omega), K\rangle\end{aligned}$$

## Three-Dimensional TQFTs

A three-dimensional TQFT $\boldsymbol{V}$ assigns to every closed oriented surface $X$ a finite-dimensional vector space $\boldsymbol{V}(X)$ over a field $\boldsymbol{k}$ and assigns to every cobordism $(M, X, Y)$ a linear map $\boldsymbol{V}(M) = \boldsymbol{V}(M, X, Y): \boldsymbol{V}(X) \to \boldsymbol{V}(Y)$. Here, a "cobordism" $(M, X, Y)$ between surfaces $X$ and $Y$ is a compact oriented 3-manifold $M$ with $\partial M = (-X) \amalg Y$ (the minus sign indicates the orientation reversal). A TQFT has to satisfy axioms which can be expressed by saying that $\boldsymbol{V}$ is a monoidal functor from the category of surfaces and cobordisms to the category of vector spaces over $\boldsymbol{k}$. Homeomorphisms of surfaces should induce isomorphisms of the corresponding vector spaces compatible with the action of cobordisms. From the definition, $\boldsymbol{V}(\emptyset) = \boldsymbol{k}$. Every compact oriented 3-manifold $M$ is a cobordism between $\emptyset$ and $\partial M$ so that $\boldsymbol{V}$ yields a "vacuum" vector $\boldsymbol{V}(M) \in \mathrm{Hom}(\boldsymbol{V}(\emptyset), \boldsymbol{V}(\partial M)) = \boldsymbol{V}(\partial M)$. If $\partial M = \emptyset$, then this gives a numerical invariant $\boldsymbol{V}(M) \in \boldsymbol{V}(\emptyset) = \boldsymbol{k}$.

Interestingly, TQFTs are often defined for surfaces and 3-cobordisms with additional structure. The surfaces $X$ are normally endowed with Lagrangians, that is, with maximal isotropic subspaces in $H_1(X; \boldsymbol{R})$. For 3-cobordisms, several additional structures are considered in the literature: for example, 2-framings, $p_1$-structures, and numerical weights. All these choices are equivalent. The TQFTs requiring such additional structures are said to be "projective" since they provide projective linear representations of the mapping class groups of surfaces.

Every modular category $\mathcal{C}$ with ground field $\boldsymbol{k}$ and simple objects $\{V_\lambda\}_\lambda$ gives rise to a projective three-dimensional TQFT $\boldsymbol{V}_{\mathcal{C}}$. It depends on the choice of a square root $\mathcal{D}$ of $\sum_\lambda (\dim(V_\lambda))^2 \in \boldsymbol{k}$. For a connected surface $X$ of genus $g$,

$$\boldsymbol{V}_{\mathcal{C}}(X) = \mathrm{Hom}_{\mathcal{C}}\left(\mathbb{1}, \bigoplus_{\lambda_1,\ldots,\lambda_g} \bigotimes_{r=1}^{g} (V_{\lambda_r} \otimes V_{\lambda_r}^*)\right)$$

The dimension of this vector space enters the Verlinde formula

$$\dim_{\boldsymbol{k}}(\boldsymbol{V}_{\mathcal{C}}(X)) \cdot 1_{\boldsymbol{k}} = \mathcal{D}^{2g-2} \sum_\lambda (\dim(V_\lambda))^{2-2g}$$

where $1_{\boldsymbol{k}} \in \boldsymbol{k}$ is the unit of the field $\boldsymbol{k}$. If $\mathrm{char}(\boldsymbol{k}) = 0$, then this formula computes $\dim_{\boldsymbol{k}}(\boldsymbol{V}_{\mathcal{C}}(X))$. For a closed connected oriented 3-manifold $M$ with numerical weight zero, $\boldsymbol{V}_{\mathcal{C}}(M) = \mathcal{D}^{-b_1(M)-1} \tau_{\mathcal{C}}(M)$, where $b_1(M)$ is the first Betti number of $M$.

The TQFT $\boldsymbol{V}_{\mathcal{C}}$ extends to a vaster class of surfaces and cobordisms. Surfaces may be enriched with a finite set of marked points, each labeled with an object of $\mathcal{C}$ and endowed with a tangent direction. Cobordisms may be enriched with ribbon (or fat) graphs whose edges are labeled with objects of $\mathcal{C}$ and whose vertices are labeled with appropriate intertwiners. The resulting TQFT, also denoted $\boldsymbol{V}_{\mathcal{C}}$, is nondegenerate in the sense that, for any surface $X$, the vacuum vectors in $\boldsymbol{V}(X)$ determined by all $M$

with $\partial M = X$ span $\mathbf{V}(X)$. A detailed construction of $\mathbf{V}_{\mathcal{C}}$ is given in Turaev (1994).

The two-dimensional part of $\mathbf{V}_{\mathcal{C}}$ determines a "modular functor" in the sense of G Segal, G Moore, and N Seiberg.

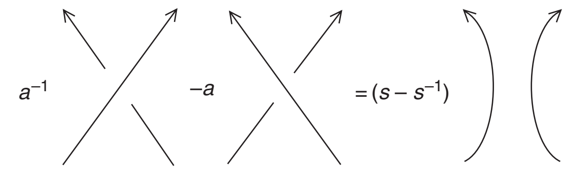



**Figure 2** The Homfly relation.

## Constructions of Modular Categories

The universal enveloping algebra $U\mathfrak{g}$ of a (finite-dimensional complex) simple Lie algebra $\mathfrak{g}$ admits a deformation $U_q\mathfrak{g}$, which is a quasitriangular Hopf algebra. The representation category $\mathrm{Rep}(U_q\mathfrak{g})$ is $\mathbf{C}$-linear and ribbon. For generic $q \in \mathbf{C}$, this category is semisimple. (The irreducible representations of $\mathfrak{g}$ can be deformed to irreducible representations of $U_q\mathfrak{g}$.) For $q$, an appropriate root of unity, a certain subquotient of $\mathrm{Rep}(U_q\mathfrak{g})$ is a modular category with ground field $\boldsymbol{k} = \mathbf{C}$. For $\mathfrak{g} = \mathrm{sl}_2(\mathbf{C})$, it was pointed out by Reshetikhin and Turaev; the general case involves the theory of tilting modules. The corresponding 3-manifold invariant $\tau$ is denoted $\tau_q^{\mathfrak{g}}$. For example, if $\mathfrak{g} = \mathrm{sl}_2(\mathbf{C})$ and $M$ is the Poincaré homology sphere (obtained by surgery on a left-hand trefoil with framing –1), then (Le 2003)

$$\tau_q^{\mathfrak{g}}(M) = (1-q)^{-1} \sum_{n \geq 0} q^n (1 - q^{n+1}) \times (1 - q^{n+2}) \cdots (1 - q^{2n+1})$$

The sum here is finite since $q$ is a root of unity.

There is another construction (Le 2003) of a modular category associated with a simple Lie algebra $\mathfrak{g}$ and certain roots of unity $q$. The corresponding quantum invariant of 3-manifolds is denoted $\tau_q^{P\mathfrak{g}}$. (Here, it is normalized so that $\tau_q^{P\mathfrak{g}}(S^3) = 1$.) Under mild assumptions on the order of $q$, we have $\tau_q^{\mathfrak{g}}(M) = \tau_q^{\mathfrak{g}}(M)\tau'(M)$ for all $M$, where $\tau'(M)$ is a certain Gauss sum determined by $\mathfrak{g}$, the homology group $H = H_1(M)$ and the linking form $\mathrm{Tors}\, H \times \mathrm{Tors}\, H \to \mathbf{Q}/\mathbf{Z}$.

A different construction derives modular categories from the category of framed oriented tangles $\mathcal{T}$. Given a ring $K$, a bigger category $K[\mathcal{T}]$ can be considered whose morphisms are linear combinations of tangles with coefficients in $K$. Both $\mathcal{T}$ and $K[\mathcal{T}]$ have a natural structure of a ribbon monoidal category.

The skein method builds ribbon categories by quotienting $K[\mathcal{T}]$ using local "skein" relations, which appear in the theory of knot polynomials (the Alexander–Conway polynomial, the Homfly polynomial, and the Kauffman polynomial). In order to obtain a semisimple category, one completes the quotient category with idempotents as objects (the Karoubi completion). Choosing appropriate skein relations, one can recover the modular categories derived from quantum groups of series $A$, $B$, $C$, $D$. In particular, the categories determined by the series $A$ arise from the Homfly skein relation shown in Figure 2 where $a, s \in K$. The categories determined by the series $B$, $C$, $D$ arise from the Kauffman skein relation.

The quantum invariants of 3-manifolds and the TQFTs associated with $\mathrm{sl}_N$ can be directly described in terms of the Homfly skein theory, avoiding the language of ribbon categories (W Lickorish, C Blanchet, N Habegger, G Masbaum, P Vogel for $\mathrm{sl}_2$ and Y Yokota for all $\mathrm{sl}_N$).

## Unitarity

From both physical and topological viewpoints, one is mainly interested in Hermitian and unitary TQFTs (over $\boldsymbol{k} = \mathbf{C}$). A TQFT $\mathbf{V}$ is Hermitian if the vector space $\mathbf{V}(X)$ is endowed with a nondegenerate Hermitian form $\langle .,. \rangle_X : \mathbf{V}(X) \otimes_{\mathbb{C}} \mathbf{V}(X) \to \mathbb{C}$ such that:

1. the form $\langle .,. \rangle_X$ is natural with respect to homeomorphisms and multiplicative with respect to disjoint union and
2. for any cobordism $(M, X, Y)$ and any $x \in \mathbf{V}(X)$, $y \in \mathbf{V}(Y)$,

$$\langle \mathbf{V}(M, X, Y)(x), y \rangle_Y = \langle x, \mathbf{V}(-M, Y, X)(y) \rangle_X$$

If $\langle .,. \rangle_X$ is positive definite for every $X$, then the Hermitian TQFT is "unitary." Note two features of Hermitian TQFTs. If $\partial M = \emptyset$, then $\mathbf{V}(-M) = \overline{\mathbf{V}(M)}$. The group of self-homeomorphisms of any $X$ acts in $\mathbf{V}(X)$ preserving the form $\langle .,. \rangle_X$. For a unitary TQFT, this gives an action by unitary matrices.

The three-dimensional TQFT derived from a modular category $\mathcal{V}$ is Hermitian (resp. unitary) under additional assumptions on $\mathcal{V}$ which are discussed briefly. A "conjugation" in $\mathcal{V}$ assigns to each morphism $f : V \to W$ in $\mathcal{V}$ a morphism $\bar{f} : W \to V$ so that

$$\bar{\bar{f}} = f, \quad \overline{f + g} = \bar{f} + \bar{g} \quad \text{for any } f, g : V \to W$$
$$\overline{f \otimes g} = \bar{f} \otimes \bar{g} \quad \text{for any morphisms } f, g \text{ in } \mathcal{C}$$
$$\overline{f \circ g} = \bar{g} \circ \bar{f} \quad \text{for any morphisms } f : V \to W,\ g : W \to V$$

One calls $\mathcal{V}$ Hermitian if it is endowed with conjugation such that

$$\overline{\theta_V} = (\theta_V)^{-1}, \quad \overline{c_{V,W}} = (c_{V,W})^{-1}$$
$$\overline{b_V} = d_V c_{V,V^*}(\theta_V \otimes \mathbf{1}_{V^*})$$
$$\overline{d_V} = (\mathbf{1}_{V^*} \otimes \theta_V^{-1}) c_{V^*,V}^{-1} b_V$$

for any objects $V$, $W$ of $\mathcal{V}$. A Hermitian modular category $\mathcal{V}$ is unitary if $\mathrm{tr}(f\bar{f}) \geq 0$ for any morphism $f$ in $\mathcal{V}$. The three-dimensional TQFT, derived from a Hermitian (resp. unitary) modular category, has a natural structure of a Hermitian (resp. unitary) TQFT.

The modular category derived from a simple Lie algebra $\mathfrak{g}$ and a root of unity $q$ is always Hermitian. It may be unitary for some $q$. For simply laced $\mathfrak{g}$, there are always such roots of unity $q$ of any given sufficiently big order. For non-simply-laced $\mathfrak{g}$, this holds under certain divisibility conditions on the order of $q$.

## Integral Structures in TQFTs

The quantum invariants of 3-manifolds have one fundamental property: up to an appropriate rescaling, they are algebraic integers. This was first observed by H Murakami, who proved that $\tau_q^{\mathrm{sl}_2}(M)$ is an algebraic integer, provided the order of $q$ is an odd prime and $M$ is a homology sphere. This extends to an arbitrary closed connected oriented 3-manifold $M$ and an arbitrary simple Lie algebra $\mathfrak{g}$ as follows (Le 2003): for any sufficiently big prime integer $r$ and any primitive $r$th root of unity $q$,

$$\tau_q^{P\mathfrak{g}}(M) \in \mathbf{Z}[q] = \mathbf{Z}[\exp(2\pi i/r)] \qquad [1]$$

This inclusion allows one to expand $\tau_q^{P\mathfrak{g}}(M)$ as a polynomial in $q$. A study of its coefficients leads to the Ohtsuki invariants of rational homology spheres and further to perturbative invariants of 3-manifolds due to T Le, J Murakami, and T Ohtsuki (see Ohtsuki (2002)). Conjecturally, the inclusion [1] holds for nonprime (sufficiently big) $r$ as well. Connections with the algebraic number theory (specifically modular forms) were studied by D Zagier and R Lawrence.

It is important to obtain similar integrality results for TQFTs. Following P Gilmer, fix a Dedekind domain $D \subset \mathbf{C}$ and call a TQFT $\mathbf{V}$ almost $D$-integral if it is nondegenerate and there is $d \in \mathbf{C}$ such that $d\mathbf{V}(M) \in D$ for all $M$ with $\partial M = \emptyset$. Given an almost-integral TQFT $\mathbf{V}$ and a surface $X$, we define $S(\hat{X})$ to be the $D$-submodule of $\mathbf{V}(X)$, generated by all vacuum vectors for $X$. This module is preserved under the action of self-homeomorphisms of $X$. It turns out that $S(X)$ is a finitely generated projective $D$-module and $\mathbf{V}(X) = S(X) \otimes_D \mathbf{C}$. A cobordism $(M, X, Y)$ is targeted if all its connected components meet $Y$ along a nonempty set. In this case, $\mathbf{V}(M)(S(X)) \subset S(Y)$. Thus, applying $S$ to surfaces and restricting $\tau$ to targetet cobordisms, we obtain an "integral version" of $\mathbf{V}$. In many interesting cases, the $D$-module $S(X)$ is free and its basis may be described explicitly. A simple Lie algebra $\mathfrak{g}$ and a primitive $r$th (in some cases 4$r$th) root of unity $q$ with sufficiently big prime $r$ give rise to an almost $D$-integral TQFT for $D = \mathbf{Z}[q]$.

## State-Sum Invariants

Another approach to three-dimensional TQFTs is based on the theory of 6$j$-symbols and state sums on triangulations of 3-manifolds. This approach introduced by V Turaev and O Viro is a quantum deformation of the Ponzano–Regge model for the three-dimensional lattice gravity. The quantum 6$j$-symbols derived from representations of $U_q(\mathrm{sl}_2\mathbf{C})$ are $\mathbf{C}$-valued rational functions of the variable $q_0 = q^{1/2}$

$$\begin{vmatrix} i & j & k \\ l & m & n \end{vmatrix} \qquad [2]$$

numerated by 6-tuples of non-negative integers $i$, $j$, $k$, $l$, $m$, $n$. One can think of these integers as labels sitting on the edges of a tetrahedron (see Figure 3). The 6$j$-symbol admits various equivalent normalizations and we choose the one which has full tetrahedral symmetry. Now, let $q_0 \in \mathbf{C}$ be a primitive 2$r$th root of unity with $r \geq 2$. Set $I = \{0, 1, \ldots, r-2\}$. Given a labeled tetrahedron $T$ as in Figure 3 with $i, j, k, l, m, n \in I$, the 6$j$-symbol [2] can be evaluated at $q_0$ and we can obtain a complex number denoted $|T|$. Consider a closed three-dimensional manifold $M$ with triangulation $t$. (Note that all 3-manifolds can be triangulated.) A coloring of $M$ is a mapping $\varphi$ from the set $\mathrm{Edg}(t)$ of the edges of $t$ to $I$. Set

$$|M| = (\sqrt{2r}/(q_0 - q_0^{-1}))^{-2a} \sum_{\varphi} \prod_{e \in \mathrm{Edg}(t)} \langle \varphi(e) \rangle \prod_T |T^{\varphi}|$$

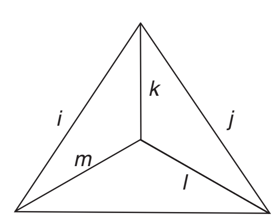



**Figure 3** Labeled tetrahedron.

where $a$ is the number of vertices of $t$, $\langle n \rangle = (-1)^n (q_0^n - q_0^{-n})/(q_0 - q_0^{-1})$ for any integer $n$, $T$ runs over all tetrahedra of $t$, and $T^\varphi$ is $T$ with the labeling induced by $\varphi$. It is important to note that $|M|$ *does not depend on the choice of t and thus yields a topological invariant of M*.

The invariant $|M|$ is closely related to the quantum invariant $\tau_q^{\mathfrak{g}}(M)$ for $\mathfrak{g} = \mathrm{sl}_2(\mathbf{C})$. Namely, $|M|$ is the square of the absolute value of $\tau_q^{\mathfrak{g}}(M)$, that is, $|M| = |\tau_q^{\mathfrak{g}}(M)|^2$. This computes $|\tau_q^{\mathfrak{g}}(M)|$ inside $M$ without appeal to surgery. No such computation of the phase of $\tau_q^{\mathfrak{g}}(M)$ is known.

These constructions generalize in two directions. First, they extend to manifolds with boundary. Second, instead of the representation category of $U_q(\mathrm{sl}_2\mathbf{C})$, one can use an arbitrary modular category $\mathcal{C}$. This yields a three-dimensional TQFT, which associates to a surface $X$ a vector space $|X|_\mathcal{C}$, and to a 3-cobordism $(M, X, Y)$ a homomorphism $|M|_\mathcal{C} : |X|_\mathcal{C} \to |Y|_\mathcal{C}$, (see Turaev (1994)). When $X = Y = \emptyset$, this homomorphism is multiplication $\mathbf{C} \to \mathbf{C}$ by a topological invariant $|M|_\mathcal{C} \in \mathbf{C}$. The latter is computed as a state sum on a triangulation of $M$ involving the 6*j*-symbols associated with $\mathcal{C}$. In general, these 6*j*-symbols are not numbers but tensors so that, instead of their product, one should use an appropriate contraction of tensors. The vectors in $V(X)$ are geometrically represented by trivalent graphs on $X$ such that every edge is labeled with a simple object of $\mathcal{C}$ and every vertex is labeled with an intertwiner between the three objects labeling the incident edges. The TQFT $|\cdot|_\mathcal{C}$ is related to the TQFT $\boldsymbol{V} = \boldsymbol{V}_\mathcal{C}$ by $|M|_\mathcal{C} = |\boldsymbol{V}(M)|^2$. Moreover, for any closed oriented surface $X$,

$$\begin{aligned}|X|_\mathcal{C} &= \mathrm{End}(\boldsymbol{V}(X)) = \boldsymbol{V}(X) \otimes (\boldsymbol{V}(X))^* \\ &= \boldsymbol{V}(X) \otimes \boldsymbol{V}(-X)\end{aligned}$$

and for any three-dimensional cobordism $(M, X, Y)$,

$$\begin{aligned}|M|_\mathcal{C} = \boldsymbol{V}(M) \otimes \boldsymbol{V}(-M) &: \boldsymbol{V}(X) \otimes \boldsymbol{V}(-X) \\ &\to \boldsymbol{V}(Y) \otimes \boldsymbol{V}(-Y)\end{aligned}$$

J Barrett and B Westbury introduced a generalization of $|M|_\mathcal{C}$ derived from the so-called spherical monoidal categories (which are assumed to be semisimple with a finite set of isomorphism classes of simple objects). This class includes modular categories and a most interesting family of (unitary monoidal) categories arising in the theory of subfactors (see Evans and Kawahigashi (1998) and Kodiyalam and Sunder (2001)). Every spherical category $\mathcal{C}$ gives rise to a topological invariant $|M|_\mathcal{C}$ of a closed oriented 3-manifold $M$. (It seems that this approach has not yet been extended to cobordisms.)

Every monoidal category $\mathcal{C}$ gives rise to a double (or a center) $Z(\mathcal{C})$, which is a braided monoidal category (see Majid (1995)). If $\mathcal{C}$ is spherical, then $Z(\mathcal{C})$ is modular. Conjecturally, $|M|_\mathcal{C} = \tau_{Z(\mathcal{C})}(M)$. In the case where $\mathcal{C}$ arises from a subfactor, this has been recently proved by Y Kawahigashi, N Sato, and M Wakui.

The state sum invariants above are closely related to spin networks, spin foam models, and other models of quantum gravity in dimension $2 + 1$ (see Baez (2000) and Carlip (1998)).

*See also:* Axiomatic Approach to Topological Quantum Field Theory; Braided and Modular Tensor Categories; Chern–Simons Models: Rigorous Results; Finite-type Invariants of 3-Manifolds; Large-*N* and Topological Strings; Schwarz-Type Topological Quantum Field Theory; Topological Quantum Field Theory: Overview; von Neumann Algebras: Subfactor Theory.

## Further Reading


Baez JC (2000) *An Introduction to Spin Foam Models of BF Theory and Quantum Gravity, Geometry and Quantum Physics*, Lecture Notes in Physics, No. 543, pp. 25–93. Berlin: Springer.

Bakalov B and Kirillov A Jr. (2001) *Lectures on Tensor Categories and Modular Functors*. University Lecture Series, vol. 21. Providence, RI: American Mathematical Society.

Blanchet C, Habegger N, Masbaum G, and Vogel P (1995) Topological quantum field theories derived from the Kauffman bracket. *Topology* 34: 883–927.

Carlip S (1998) *Quantum Gravity in* 2 + 1 *Dimensions,*, Cambridge Monographs on Mathematical Physics Cambridge: Cambridge University Press

Carter JS, Flath DE, and Saito M (1995) *The Classical and Quantum 6j-Symbols*. Mathematical Notes, vol. 43. Princeton: Princeton University Press.

Evans D and Kawahigashi Y (1998) *Quantum Symmetries on Operator Algebras,* Oxford Mathematical Monographs, Oxford Science Publications. New York: The Clarendon Press, Oxford University Press.

Kauffman LH (2001) *Knots and Physics*, 3rd edn., Series on Knots and Everything, vol. 1. River Edge, NJ: World Scientific.

Kerler T and Lyubashenko V (2001) *Non-Semisimple Topological Quantum Field Theories for 3-Manifolds with Corners*. Lecture Notes in Mathematics, vol. 1765. Berlin: Springer.

Kodiyalam V and Sunder VS (2001) *Topological Quantum Field Theories from Subfactors*. Research Notes in Mathematics, vol. 423. Boca Raton, FL: Chapman and Hall/CRC Press.

Le T (2003) Quantum invariants of 3-manifolds: integrality, splitting, and perturbative expansion. *Topology and Its Applications* 127: 125–152.

Lickorish WBR (2002) *Quantum Invariants of 3-Manifolds. Handbook of Geometric Topology*, pp. 707–734. Amsterdam: North-Holland.

Majid S (1995) *Foundations of Quantum Group Theory*. Cambridge: Cambridge University Press

Ohtsuki T (2002) *Quantum Invariants. A Study of Knots, 3-Manifolds, and Their Sets,* Series on Knots and Everything, vol. 29. River Edge, NJ: World Scientific.

Turaev V (1994) *Quantum Invariants of Knots and 3- Manifolds*. de Gruyter Studies in Mathematics, vol. 18. Berlin: Walter de Gruyter.

# Quantum Calogero–Moser Systems

**R Sasaki**, Kyoto University, Kyoto, Japan



## Introduction

Calogero–Moser (C–M) systems are multiparticle (i.e., finite degrees of freedom) dynamical systems with long-range interactions. They are integrable and solvable at both classical and quantum levels. These systems offer an ideal arena for interplay of many important concepts in mathematical/theoretical physics: to name a few, classical and quantum mechanics, classical and quantum integrability, exact and quasi-exact solvability, addition of discrete (spin) degrees of freedom, quantum Lax pair formalism, supersymmetric quantum mechanics, crystallographic root systems and associated Weyl groups and Lie algebras, noncrystallographic root systems, and Coxeter groups or finite reflection groups. The quantum integrability or solvability of C–M systems does not depend on such known solution mechanisms as Yang–Baxter equations, quantum *R*-matrix or Bethe ansatz for the quantum systems. In fact, quantum C–M systems provide a good material for pondering about quantum integrability.

## Quantum (Liouville) Integrability

The classical Liouville theorem for an integrable system consists of two parts. Let us consider Hamiltonian dynamics of finite degrees of freedom $N$ with coordinates $q=(q_1,\ldots,q_N)$ and conjugate momenta $p=(p_1,\ldots,p_N)$ equipped with Poisson brackets $\{q_j,p_k\}=\delta_{jk},\{q_j,q_k\}=\{p_j,p_k\}=0$. The first part is the existence of a set of independent and involutive $\{K_j,K_k\}=0$ conserved quantities $\{K_j\}$ as many as the degrees of freedom $(j=1,\ldots,N)$. The second part asserts that the generating function of the canonical transformation for the action-angle variables can be constructed from the conserved quantities via quadrature. In other words, the second part, that is, the reducibility to the action-angle variables is the integrability. The quantum counterpart of the first half is readily formulated: that is, the existence of a set of independent and mutually commuting (involutive) $[K_j,K_k]=0$ conserved quantities $\{K_j\}$ as many as the degrees of freedom. (This does not necessary imply, however, that they are well defined in a proper Hilbert space.) The definition of the quantum integrability should come as a second part, which is yet to be formulated. It is clear that the quantum Liouville integrability does not imply the complete determination of the eigenvalues and eigenfunctions. Such systems would be called exactly solvable. This can be readily understood by considering any (autonomous) degree-1 Hamiltonian system, which, by definition, is Liouville integrable at the classical and quantum levels. However, it is known that the number of excatly solvable degree-1 Hamiltonians are very limited. What would be the quantum counterpart of the "transformation to action-angle variables by quadrature"? Could it be better formulated in terms of a path integral? Many questions remain to be answered. The quantum C–M systems, an infinite family of exactly solvable multiparticle Hamiltonians, would shed some light on the problem of quantum integrability, in addition to their own beautiful structure explored below.

Throughout this article, the dependence on Planck's constant, $\hbar$, is shown explicitly to distinguish the quantum effects.

## Simplest Cases (Based on $A_{r-1}$ Root System)

The simplest example of a C–M system consists of $r$ particles of equal mass (normalized to unity) on a line with pairwise $1/(\text{distance})^2$ interactions described by the following Hamiltonian:

$$\hat{\mathcal{H}}=\frac{1}{2}\sum_{j=1}^{r}p_j^2+g(g-\hbar)\sum_{j<k}^{r}\frac{1}{(q_j-q_k)^2} \qquad [1]$$

in which $g$ is a real positive coupling constant. Here $q=(q_1,\ldots,q_r)$ are the coordinates and $p=(p_1,\ldots,p_r)$ are the conjugate canonical momenta obeying the canonical commutation relations: $[q_j,p_k]=i\hbar\delta_{jk},[q_j,q_k]=[p_j,p_k]=0,\quad j,k=1,\ldots,r$. The Heisenberg equations of motion are $\dot{q}_j=(i/\hbar)[\hat{\mathcal{H}},q_j]=p_j,\ \ddot{q}_j=\dot{p}_j=(i/\hbar)[\hat{\mathcal{H}},p_j]=2g(g-\hbar)\sum_{k\neq j}1/(q_j-q_k)^3$. The repulsive $1/(\text{distance})^2$ potential cannot be surmounted classically or quantum mechanically, and the relative position of the particles on the line is not changed during the time evolution. Classically, it means that if a motion starts at a configuration $q_1>q_2>\cdots>q_r$, then the inequalities remain valid throughout the time evolution. At the quantum level, the wave functions vanish at the boundaries, and the configuration space can be naturally limited to $q_1>q_2>\cdots>q_r$ (the principal Weyl chamber).

Similar integrable quantum many-particle dynamics are obtained by replacing the inverse square potential in [1] by the trigonometric

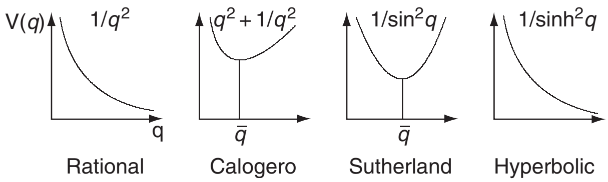



**Figure 1** Four different types of quantum C–M potentials.

(hyperbolic) counterpart (see Figure 1) $1/(q_j - q_k)^2 \to a^2/\sinh^2 a(q_j - q_k)$, in which $a > 0$ is a real parameter. The $1/\sin^2 q$ potential case (the Sutherland system) corresponds to the $1/(\text{distance})^2$ interaction on a circle of radius $1/2a$, see Figure 2. A harmonic confining potential $\omega^2 \sum_{j=1}^r q_j^2/2$ can be added to the rational Hamiltonian [1] without breaking the integrability (the Calogero system, see Figure 1). At the classical level, the trigonometric (hyperbolic) and rational C–M systems are obtained from the elliptic potential systems (with the Weierstrass $\wp$ function) as the degenerate limits: $\wp(q_1 - q_2) \to a^2/\sinh^2 a(q_1 - q_2) \to 1/(q_1 - q_2)^2$, namely as one (two) period(s) of the $\wp$ function tends to infinity.

It is remarkable that these equations of motion can be expressed in a matrix form (Lax pair): $\mathrm{i}/\hbar[\hat{\mathcal{H}}, L] = \mathrm{d}L/\mathrm{d}t = LM - ML = [L, M] \Leftrightarrow$ Heisenberg equation of motion, in which $L$ and $M$ are given by

$$L = \begin{pmatrix} p_1 & \frac{\mathrm{i}g}{q_1-q_2} & \cdots & \frac{\mathrm{i}g}{q_1-q_r} \\ \frac{\mathrm{i}g}{q_2-q_1} & p_2 & \cdots & \frac{\mathrm{i}g}{q_2-q_r} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathrm{i}g}{q_r-q_1} & \frac{\mathrm{i}g}{q_r-q_2} & \cdots & p_r \end{pmatrix}$$

$$M = \begin{pmatrix} m_1 & -\frac{\mathrm{i}g}{(q_1-q_2)^2} & \cdots & -\frac{\mathrm{i}g}{(q_1-q_r)^2} \\ -\frac{\mathrm{i}g}{(q_2-q_1)^2} & m_2 & \cdots & -\frac{\mathrm{i}g}{(q_2-q_r)^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\mathrm{i}g}{(q_r-q_1)^2} & -\frac{\mathrm{i}g}{(q_r-q_2)^2} & \cdots & m_r \end{pmatrix} \quad [2]$$

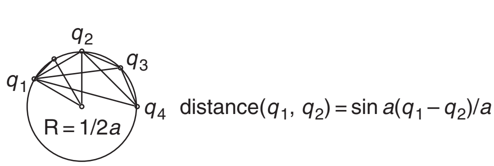



**Figure 2** Sutherland potential is $1/(\text{distance})^2$ interaction on a circle. The large-radius limit, $a \to 0$, gives the rational potential.

The diagonal element $m_j$ of $M$ is given by $m_j = \mathrm{i}g \sum_{k \neq j} 1/(q_j - q_k)^2$. The matrix $M$ has a special property $\sum_{j=1}^r M_{jk} = \sum_{k=1}^r M_{jk} = 0$, which ensures the quantum conserved quantities as the total sum of powers of Lax matrix $L$: $[\hat{\mathcal{H}}, K_n] = 0$, $K_n \equiv \mathrm{Ts}(L^n) = \sum_{j,k} (L^n)_{jk}$, $(n = 1, 2, 3, \ldots)$, $[K_n, K_m] = 0$. It should be stressed that the trace of $L^n$ is not conserved because of the noncommutativity of $q$ and $p$. The Hamiltonian is equivalent to $K_2, \hat{\mathcal{H}} \propto K_2 +$ const. In other words, the Lax matrix $L$ is like a "square root" of the Hamiltonian. The quantum equations of motion for the Sutherland and hyperbolic potentials are again expressed by Lax pairs if the following replacements are made: $1/(q_j - q_k) \to a \coth a(q_j - q_k)$ in $L$ and $1/(q_j - q_k)^2 \to a^2/\sinh^2 a(q_j - q_k)$ in $M$. The quantum conserved quantities are obtained in the same manner as above for the systems with the trigonometric and hyperbolic interactions.

The main goal here is to find all the eigenvalues $\{\mathcal{E}\}$ and eigenfunctions $\{\psi(q)\}$ of the Hamiltonians with the rational, Calogero, Sutherland, and hyperbolic potentials: $\hat{\mathcal{H}}\psi(q) = \mathcal{E}\psi(q)$. The momentum operator $p_j$ acts as differential operators $p_j = -\mathrm{i}\hbar\partial/\partial q_j$. For example, for the rational model Hamiltonian [1], the eigenvalue equation reads

$$\left[-\frac{\hbar^2}{2}\sum_{j=1}^{r}\frac{\partial^2}{\partial q_j^2} + g(g-\hbar)\sum_{j<k}^{r}\frac{1}{(q_j - q_k)^2}\right]\psi(q) = \mathcal{E}\psi(q) \quad [3]$$

which is a second-order Fuchsian differential equation for each variable $\{q_j\}$ with a regular singularity at each hyperplane $q_j = q_k$ whose exponents are $g/\hbar, 1 - g/\hbar$. Any solution $\psi$ of [3] is regular at all points, except for those on the union of hyperplanes $q_j = q_k$. Since the structure of the singularity is the same for the other three types of potentials, the same assertion for the regularity and singularity of the solution $\psi$ holds for these cases, too. For the trigonometric (Sutherland) case, there are other singularities at $q_j - q_k = l\pi/a$, $l \in \mathbf{Z}$, due to the periodicity of the potential. As is clear from the shape of the potentials, see Figure 1, the rational and hyperbolic Hamiltonians have only continuous spectra, whereas the Calogero and Sutherland Hamiltonians have only discrete spectra.

The integrability or more precisely the triangularity of the quantum C–M Hamiltonian was first discovered by Calogero for particles on a line with inverse square potential plus a confining harmonic force and by Sutherland for the particles on a circle

with the trigonometric potential. Later, classical integrability of the models in terms of Lax pairs was proved by Moser. Olshanetsky and Perelomov showed that these systems were based on $A_{r-1}$ root systems, that is, $q_j - q_k = \alpha \cdot q$, and $\alpha$ is one of the root vectors of $A_{r-1}$ root system [13]. They also introduced generalizations of the C–M systems based on any root system including the noncrystallographic ones.

As shown by Heckman–Opdam and Sasaki and collaborators, quantum C–M systems with degenerate potentials (i.e., the rational potentials with/without harmonic force, the hyperbolic, and the trigonometric potentials), based on any root system can be formulated and solved universally. To be more precise, the rational and Calogero systems are integrable for all root systems, the crystallographic and noncrystallographic. The hyperbolic and trigonometric (Sutherland) systems are integrable for any crystallographic root system. The universal formulas for the Hamiltonians, Lax pairs, ground state wave functions, conserved quantities, the triangularity, the discrete spectra for the Calogero and Sutherland systems, the creation and annihilation operators, etc., are equally valid for any root system. This will be shown in the next section. Some rudimentary facts of the root systems and reflections are summarized in the appendix.

## Universal Formalism

A C–M system is a Hamiltonian dynamical systems associated with a root system $\Delta$ of rank $r$, which is a set of vectors in $\mathbf{R}^r$ with its standard inner product. A brief review of the properties of the root systems and the associated reflections together with explicit realizations of all the classical root systems will be found in the appendix.

### Factorized Hamiltonian

The Hamiltonian for the quantum C–M system can be written in terms of a pre-potential $W(q)$ in a "factorized form":

$$\mathcal{H} = \frac{1}{2}\sum_{j=1}^{r}\left(p_j - \mathrm{i}\frac{\partial W(q)}{\partial q_j}\right)\left(p_j + \mathrm{i}\frac{\partial W(q)}{\partial q_j}\right) \qquad [4]$$

The pre-potential is a sum over positive roots:

$$W(q) = \sum_{\alpha\in\Delta_+} g_\alpha \ln|w(\alpha\cdot q)| + \left(-\frac{\omega}{2}q^2\right) \qquad [5]$$

The real positive coupling constants $g_\alpha$ are defined on orbits of the corresponding Coxeter group, that is, they are identical for roots in the same orbit. That is, for the simple Lie algebra cases, one coupling constant, $g_\alpha = g$, for all roots in simply laced models and two independent coupling constants, $g_\alpha = g_{\mathrm{L}}$ for long roots and $g_\alpha = g_{\mathrm{S}}$ for short roots, in non-simply laced models. The function $w(u)$ and the other functions $x(u)$ and $y(u)$ appearing in the Lax pair [10], [11] are listed in Table 1 for each type of degenerate potentials. The dynamics of the prepotentials $W(q)$ (eqn [5]) has been discussed by Dyson from a different point of view (random-matrix model). The above factorized Hamiltonian [4] consists of an operator part $\hat{\mathcal{H}}$, which is the Hamiltonian in the usual definition (see the Hamiltonians in the previous section, e.g., [1]), and a constant $\mathcal{E}_0$ which is the ground-state energy, $\mathcal{H} = \hat{\mathcal{H}} - \mathcal{E}_0$. The factorized Hamiltonian [4] also arises within the context of supersymmetric quantum mechanics.

**Table 1** Functions appearing in the prepotential and Lax pair

| *Potential* | $w(u)$ | $x(u)$ | $y(u)$ |
|---|---|---|---|
| Rational | $u$ | $1/u$ | $-1/u^2$ |
| Hyperbolic | $\sinh au$ | $a\coth au$ | $-a^2/\sinh^2 au$ |
| Trigonometric | $\sin au$ | $a\cot au$ | $-a^2/\sin^2 au$ |

The pre-potential and the Hamiltonian are invariant under reflection of the phase space variables in the hyperplane perpendicular to any root $W(s_\alpha(q)) = W(q), \mathcal{H}(s_\alpha(p), s_\alpha(q)) = \mathcal{H}(p,q), \forall\alpha\in\Delta$, with $s_\alpha$ defined by [12]. The above Coxeter (Weyl) invariance is the only (discrete) symmetry of the C–M systems. The main problem is, as in the $A_{r-1}$ case, to find all the eigenvalues $\{\mathcal{E}\}$ and eigenfunctions $\{\psi(q)\}$ of the above Hamiltonian $\mathcal{H}\psi(q) = \mathcal{E}\psi(q)$.

For any root system and for any choice of potential, the C–M system has a hard repulsive potential $\sim 1/(\alpha\cdot q)^2$ near the reflection hyperplane $H_\alpha = \{q\in\mathbf{R}^r, \alpha\cdot q = 0\}$. The C–M eigenvalue equation is a second-order Fuchsian differential equation with regular singularities at each reflection hyperplane $H_\alpha$ and those arising from the periodicity in the case of the Sutherland potential. Near the reflection hyperplane $H_\alpha$, the solution behaves as follows:

$$\psi \sim (\alpha\cdot q)^{g_\alpha/\hbar}(1 + \text{regular terms}), \quad \text{or}$$
$$\psi \sim (\alpha\cdot q)^{1-g_\alpha/\hbar}(1 + \text{regular terms})$$

The former solution is chosen for the square integrability. Because of the singularities, the configuration space is restricted to the principal Weyl chamber PW or the principal Weyl alcove $\mathrm{PW_T}$ for the trigonometric potential (see Figure 3): $\mathrm{PW} = \{q\in\mathbf{R}^r \mid \alpha\cdot q > 0, \alpha\in\Pi\}, \mathrm{PW_T} = \{q\in\mathbf{R}^r \mid \alpha\cdot q > 0,$

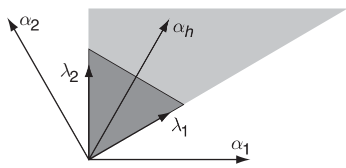



**Figure 3** Simple roots, the highest root, fundamental weights, and the principal Weyl alcove (grey) and the principal Weyl chamber (light grey, extending to infinity) in a two-dimensional root system.

$\alpha \in \Pi, \alpha_h \cdot q < \pi/a\}$, ($\Pi$: set of simple roots, see the appendix). Here $\alpha_h$ is the highest root.

### Ground-State Wave Function and Energy

One straightforward outcome of the factorized Hamiltonian [4] is the universal ground-state wave function, which is given by

$$\begin{aligned}\Phi_0(q) &= e^{W(q)/\hbar} \\ &= \prod_{\alpha\in\Delta_+} |w(\alpha\cdot q)|^{g_\alpha/\hbar}\left(\times e^{-(\omega/2\hbar)q^2}\right) \\ \mathcal{H}\Phi_0(q) &= 0\end{aligned} \qquad [6]$$

The exponential factor $e^{-(\omega/2\hbar)q^2}$ exists only for the Calogero systems. The ground-state energy, that is, the constant part of $\mathcal{H} = \hat{\mathcal{H}} - \mathcal{E}_0$, has a universal expression for each potential:

$$\mathcal{E}_0 = \begin{cases} 0 & \text{rational} \\ \omega\left(\hbar r/2 + \sum_{\alpha\in\Delta_+} g_\alpha\right) & \text{Calogero}\end{cases}$$
$$\mathcal{E}_0 = 2a^2\rho^2 \times \begin{cases} -1 & \text{hyperbolic} \\ 1 & \text{Sutherland}\end{cases} \qquad [7]$$

where $\rho = 1/2\sum_{\alpha\in\Delta_+} g_\alpha \alpha$ is called a "deformed Weyl vector." Obviously, $\Phi_0(q)$ is square integrable in the configuration spaces for the Calogero and Sutherland systems and not square integrable for the rational and hyperbolic potentials.

### Excited States, Triangularity, and Spectrum

Excited states of the C–M systems can be easily obtained as eigenfunctions of a differential operator $\tilde{\mathcal{H}}$ obtained from $\mathcal{H}$ by a similarity transformation:

$$\begin{aligned}\tilde{\mathcal{H}} &= e^{-W/\hbar}\mathcal{H}e^{W/\hbar} \\ &= \frac{-1}{2}\sum_{j=1}^{r}(\hbar^2\partial^2/\partial q_j^2 + 2\hbar\partial W/\partial q_j \partial/\partial q_j)\end{aligned}$$

The eigenvalue equation for $\tilde{\mathcal{H}}$, $\tilde{\mathcal{H}}\Psi_\mathcal{E} = \mathcal{E}\Psi_\mathcal{E}$, is then equivalent to that of the original Hamiltonian, $\mathcal{H}\Psi_\mathcal{E} e^W = \mathcal{E}\Psi_\mathcal{E} e^W$. Since all the singularities of the Fuchsian differential equation $\mathcal{H}\psi(q) = \mathcal{E}\psi(q)$ are contained in the ground-state wave function $e^W$, $\Psi_\mathcal{E}$ must be regular at finite $q$, including all the reflection boundaries. As for the rational and hyperbolic potentials, the energy eigenvalues are only continuous. For the rational case, the eigenfunctions are multivariable generalization of Bessel functions.

**Calogero systems** The similarity-transformed Hamiltonian $\tilde{\mathcal{H}}$ reads

$$\begin{aligned}\tilde{\mathcal{H}} &= \hbar\omega q\cdot\frac{\partial}{\partial q} - \frac{\hbar^2}{2}\sum_{j=1}^{r}\frac{\partial^2}{\partial q_j^2} \\ &\quad - \hbar\sum_{\alpha\in\Delta_+}\frac{g_\alpha}{\alpha\cdot q}\alpha\cdot\frac{\partial}{\partial q}\end{aligned} \qquad [8]$$

which maps a Coxeter-invariant polynomial in $q$ of degree $d$ to another of degree $d$. Thus, the Hamiltonian $\tilde{\mathcal{H}}$ (8) is lower-triangular in the basis of Coxeter-invariant polynomials and the diagonal elements have values as $\hbar\omega \times$ degree, as given by the first term. Independent Coxeter-invariant polynomials exist at the degrees $f_j$ listed in Table 2: $f_j = 1 + e_j$, $j = 1, \ldots, r$, where $\{e_j\}$, $j = 1, \ldots, r$, are the exponents of $\Delta$.

The eigenvalues of the Hamiltonian $\mathcal{H}$ are $\hbar\omega N$ with $N$ a non-negative integer. $N$ can be expressed as $N = \sum_{j=1}^r n_j f_j$, $n_j \in \mathbf{Z}_+$, and the degeneracy of the eigenvalue $\hbar\omega N$ is the number of partitions of $N$. It is remarkable that the coupling constant dependence appears only in the ground-state energy $\mathcal{E}_0$. This is a deformation of the isotropic harmonic oscillator confined in the principal Weyl chamber. The eigenpolynomials are generalization of multivariable Laguerre (Hermite) polynomials. One immediate consequence of this spectrum is the periodicity of the quantum motion. If a system has a wave function $\psi(0)$ at $t = 0$, then at $t = T = 2\pi/\omega$ the system has physically the same wave function as $\psi(0)$, that is, $\psi(T) = e^{-i\mathcal{E}_0 T/\hbar}\psi(0)$. The same assertion holds at the classical level, too.

**Table 2** The degrees $f_j$ in which independent Coxeter-invariant polynomials exist

| $\Delta$ | $f_j = 1 + e_j$ | $\Delta$ | $f_j = 1 + e_j$ |
|---|---|---|---|
| $A_r$ | $2, 3, 4, \ldots, r+1$ | $E_8$ | 2, 8, 12, 14, 18, 20, 24, 30 |
| $B_r$ | $2, 4, 6, \ldots, 2r$ | $F_4$ | 2, 6, 8, 12 |
| $C_r$ | $2, 4, 6, \ldots, 2r$ | $G_2$ | 2, 6 |
| $D_r$ | $2, 4, \ldots, 2r-2, r$ | $I_2(m)$ | 2, $m$ |
| $E_6$ | 2, 5, 6, 8, 9, 12 | $H_3$ | 2, 6, 10 |
| $E_7$ | 2, 6, 8, 10, 12, 14, 18 | $H_4$ | 2, 12, 20, 30 |

**Sutherland Systems** The periodicity of the trigonometric potential dictates that the wave function should be a Bloch factor $e^{2ia\mu\cdot q}$ (where $\mu$ is a weight) multiplied by a Fourier series in terms of simple roots. The basis of the Weyl invariant wave functions is specified by a dominant weight $\lambda=\sum_{j=1}^{r} m_j\lambda_j$, $m_j\in\mathbf{Z}_+$, $\phi_\lambda(q)\equiv\sum_{\mu\in O_\lambda} e^{2ia\mu\cdot q}$, where $O_\lambda$ is the orbit of $\lambda$ by the action of the Weyl group: $O_\lambda=\{g(\lambda)\,|\,g\in G_\Delta\}$. The set of functions $\{\phi_\lambda\}$ has an order $\succ$, $|\lambda|^2>|\lambda'|^2\Rightarrow\phi_\lambda\succ\phi_{\lambda'}$. The similarity-transformed Hamiltonian $\tilde{\mathcal{H}}$ given by

$$\tilde{\mathcal{H}}=-\frac{\hbar^2}{2}\sum_{j=1}^{r}\frac{\partial^2}{\partial q_j^2}-a\hbar\sum_{\alpha\in\Delta_+}g_\alpha\cot(a\alpha\cdot q)\alpha\cdot\frac{\partial}{\partial q}\quad[9]$$

is lower-triangular in this basis: $\tilde{\mathcal{H}}\phi_\lambda=2a^2(\hbar^2\lambda^2+2\hbar\rho\cdot\lambda)\phi_\lambda+\sum_{|\lambda'|<|\lambda|}c_{\lambda'}\phi_{\lambda'}$. That is, the eigenvalue is $\mathcal{E}=2a^2(\hbar^2\lambda^2+2\hbar\rho\cdot\lambda)$ or $\mathcal{E}+\mathcal{E}_0=2a^2(\hbar\lambda+\rho)^2$. Again, the coupling constant dependence comes solely from the deformed Weyl vector $\rho$. This spectrum is a deformation of the spectrum corresponding to the free motion with momentum $2\hbar a\lambda$ in the principal Weyl alcove. The corresponding eigenfunction is called a generalized Jack polynomial or Heckman–Opdam's Jacobi polynomial. For the rank-2 $(r=2)$ root systems, $A_2$, $B_2\cong C_2$ and $I_2(m)$ (the dihedral group), the complete set of eigenfunctions are known explicitly.

### Quantum Lax Pair and Quantum Conserved Quantities

The universal Lax pair for C–M systems is given in terms of the representations of the Coxeter (Weyl) group in stead of the Lie algebra. The Lax operators without spectral parameter for the rational, trigonometric, and hyperbolic potentials are

$$L(p,q)=p\cdot\hat{H}+X(q)$$
$$X(q)=\mathrm{i}\sum_{\alpha\in\Delta_+}g_\alpha(\alpha\cdot\hat{H})x(\alpha\cdot q)\hat{s}_\alpha\quad[10]$$

$$M(q)=\frac{\mathrm{i}}{2}\sum_{\alpha\in\Delta_+}g_\alpha\alpha^2y(\alpha\cdot q)(\hat{s}_\alpha-I)\quad[11]$$

where $I$ is the identity operator and $\{\hat{s}_\alpha|\alpha\in\Delta\}$ are the reflection operators of the root system. They act on a set of $\mathbf{R}^r$ vectors, $\mathcal{R}=\{\mu^{(k)}\in\mathbf{R}^r\,|\,k=1,\ldots,d\}$, permuting them under the action of the reflection group. The vectors in $\mathcal{R}$ form a basis for the representation space $\mathbf{V}$ of dimension $d$. The matrix elements of the operators $\{\hat{s}_\alpha\,|\,\alpha\in\Delta\}$ and $\{\hat{H}_j\,|\,j=1,\ldots,r\}$ are defined as follows: $(\hat{s}_\alpha)_{\mu\nu}=\delta_{\mu,s_\alpha(\nu)}=\delta_{\nu,s_\alpha(\mu)}$, $(\hat{H}_j)_{\mu\nu}=\mu_j\delta_{\mu\nu}$, $\alpha\in\Delta$, $\mu$, $\nu\in\mathcal{R}$. The form of the functions $x,y$ depends on the chosen potential as given in Table 1. Then the equations of motion can be expressed in a matrix form $\mathrm{d}L/\mathrm{d}t=\mathrm{i}/\hbar[\mathcal{H},L]=[L,M]$. The operator $M$ satisfies the relation $\sum_{\mu\in\mathcal{R}}M_{\mu\nu}=\sum_{\nu\in\mathcal{R}}M_{\mu\nu}=0$, which is essential for deriving quantum conserved quantities as the total sum (Ts) of all the matrix elements of $L^n$: $K_n=\mathrm{Ts}(L^n)\equiv\sum_{\mu,\nu\in\mathcal{R}}(L^n)_{\mu\nu}$, $[\mathcal{H},K_n]=0$, $[K_m,K_n]=0$, $n,m=1,\ldots$ In particular, the power 2 is universal to all the root systems, and the quantum Hamiltonian is given by $\mathcal{H}\propto K_2+$ const. As in the affine Toda molecule systems, a Lax pair with a spectral parameter can also be introduced universally for all the above potentials. The Dunkl operators, or the commuting differential–difference operators are also used to construct quantum conserved quantities for some root systems. This method is essentially equivalent to the universal Lax operator formalism. As the Lax operators do not contain the Planck's constant, the quantum Lax pair is essentially of the same form as the classical Lax pair. The difference between the trace (tr) and the total sum (Ts) vanishes as $\hbar\to 0$.

**Lax pair for Calogero systems** The quantum Lax pair for the Calogero systems is obtained from the universal Lax pair [10] by replacement $L\to L^\pm=L\pm\mathrm{i}\omega Q$, $Q\equiv q\cdot\hat{H}$, which correspond to the creation and annihilation operators of a harmonic oscillator. The equations of motion are rewritten as $\mathrm{d}L^\pm/\mathrm{d}t=\mathrm{i}/\hbar[\mathcal{H},L^\pm]=[L^\pm,M]\pm\mathrm{i}\omega L^\pm$. Then $\mathcal{L}^\pm=L^\pm L^\mp$ satisfy the Lax type equation $\mathrm{d}\mathcal{L}^\pm/\mathrm{d}t=\mathrm{i}/\hbar[\mathcal{H},\mathcal{L}^\pm]$, giving rise to conserved quantities $\mathrm{Ts}(\mathcal{L}^\pm)^n$, $n=1,2,\ldots$ The Calogero Hamiltonian is given by $\mathcal{H}\propto\mathrm{Ts}(\mathcal{L}^\pm)$.

All the eigenstates of the Calogero Hamiltonian $\mathcal{H}$ with eigenvalues $\hbar\omega N$, $N=\sum_{j=1}^{r}n_jf_j$, $n_j\in\mathbf{Z}_+$, are simply constructed in terms of $L^\pm$: $\prod_{j=1}^{r}(B_{f_j}^+)^{n_j}e^W$. Here the integers $\{f_j\}$, $j=1,\ldots,r$, are listed in Table 2. The creation operators $B_{f_j}^+$ and the corresponding annihilation operators $B_{f_j}^-$ are defined by $B_{f_j}^\pm=\mathrm{Ts}(L^\pm)^{f_j}$, $j=1,\ldots,r$. They are Hermitian conjugate to each other $(B_{f_j}^\pm)^\dagger=B_{f_j}^\mp$ with respect to the standard Hermitian inner product of the states defined in PW. They satisfy commutation relations $[\mathcal{H},B_k^\pm]=\pm\hbar k\omega B_k^\pm$, $[B_k^+,B_l^+]=[B_k^-,B_l^-]=0$, $k,l\in\{f_j\,|\,j=1,\ldots,r\}$. The ground state is annihilated by all the annihilation operators $B_{f_j}^-e^W=0$, $j=1,\ldots,r$.

## Further Developments

### Rational Potentials: Superintegrability

The systems with the rational potential have a remarkable property: superintegrability. A rational C–M system based on a rank-$r$ root system has $2r-1$

independent conserved quantities. Roughly speaking, they are of the form $K_n = \mathrm{Ts}(L^n), J_m = \mathrm{Ts}(QL^m), Q \equiv q \cdot \hat{H}$, among which only $r$ are involutive. At the classical level, superintegrability can be characterized as algebraic linearizability. Since a commutator of any conserved quantities is again a conserved quantity, these conserved quantities form a nonlinear algebra called a quadratic algebra. It can be considered as a finite-dimensional analog of the W-algebra appearing in certain conformal field theory.

#### Quantum vs Classical Integrability

In C–M systems, the classical and quantum integrability are very closely related. The quantum discrete spectra of the Calogero and the Sutherland systems are, as shown above, expressed in terms of the coupling constant $(\omega, g)$ and the exponents or the weights of the corresponding root systems. Namely, they are integral multiples of coupling constants. The corresponding classical systems with the potential $V(q) = (1/2)\sum_{j=1}^{r}(\partial W(q)/\partial q_j)^2$ share many remarkable properties. As is clear from Figure 1, they always have an equilibrium position. The equilibrium positions $(\bar{q})$ are described by the zeros of a classical orthogonal polynomial; the Hermite polynomial ($A$-type Calogero), the Laguerre polynomial ($B, C, D$-type Calogero), the Chebyshev polynomial ($A$-type Sutherland) and the Jacobi polynomial ($B, C, D$-type Sutherland). For the exceptional root systems, the corresponding polynomials were not known for a long time. The minimum energy of the classical potential $V(q)$ at the equilibrium is the quantum ground-state energy $\lim_{\hbar\to 0}\mathcal{E}_0$ itself. It is also an integral multiple of coupling constants for both Calogero and Sutherland cases. Near a classical equilibrium, a multiparticle dynamical system is always reduced to a system of coupled harmonic oscillators. For Calogero systems, the eigenfrequencies of these small oscillations are, in fact, exactly the same as the quantum eigenfrequencies, $\omega f_j = \omega(1 + e_j)$. For Sutherland systems, the classical eigenfrequencies are the same as the $o(\hbar)$ part of the quantum spectra corresponding to all the fundamental weights $\lambda_j$: $2a^2\lambda_j \cdot \rho$. Moreover, the eigenvalues of various Lax matrices $L$ and $M$ at the equilibrium take many "interesting values." These results provide ample explicit examples of the general theorem on the quantum–classical correspondence formulated by Loris–Sasaki.

#### Spin Models

For any root system $\Delta$ and an irreducible representation $\mathcal{R}$ of the Coxeter (Weyl) group $G_\Delta$, a spin C–M system can be defined for each of the potentials: rational, Calogero, hyperbolic and Sutherland. For each member $\mu$ of $\mathcal{R}$, to be called a "site," a vector space $\mathbf{V}_\mu$ is associated whose element is called a "spin." The dynamical variables are those of the particles $\{q_j, p_j\}$ and the spin exchange operators $\{\hat{\mathcal{P}}_\alpha\}$ $(\alpha \in \Delta)$ which exchange the spins at the sites $\mu$ and $s_\alpha(\mu)$. For each $\Delta$ and $\mathcal{R}$ a spin exchange model can be defined by "freezing" the particle degrees of freedom at the equilibrium point of the corresponding classical potential $\{q, p\} \to \{\bar{q}, 0\}$. These are generalization of Haldane–Shastry model for Sutherland potentials and that of Polychronakos for the Calogero potentials. Universal Lax pair operators for both spin C–M systems and spin exchange models are known and conserved quantities are constructed.

#### Integrable Deformations

C–M systems allow various integrable deformations at the classical and/or quantum levels. One of the well-known deformations is the so-called "relativistic" C–M system or the Ruijsenaars–Schneider (R–S) system. For degenerate potentials, they are integrable both at the classical and quantum levels. The classical quantities of the R–S systems at equilibrium exhibit many interesting properties, too. The equilibrium positions are described by the zeros of certain deformation of the above-mentioned classical polynomials. The frequencies of small oscillations are also related to the exact quantum spectrum, and they can be expressed as coupling constant times the ($q$-) integers.

Inozemtsev models are classically integrable multiparticle dynamical systems related to C–M systems based on classical root systems $(A, B, C, D)$ with additional $q^6$ (rational) or $\sin^2 2q$ (trigonometric) potentials. Their quantum versions are not exactly solvable in contrast to the C–M or R–S systems, although there is some evidence of their Liouville integrability (without a proper Hilbert space). Quantum Inozemtsev systems can be deformed to be a widest class of quasi-exactly solvable multiparticle dynamical systems. They possess a form of higher-order supersymmetry for which the method of prepotential is also useful.

### Appendix: Root Systems

Some rudimentary facts of the root systems and reflections are recapitulated here. The set of roots $\Delta$ is invariant under reflections in the hyperplane perpendicular to each vector in $\Delta$. In other words, $s_\alpha(\beta) \in \Delta, \forall \alpha, \beta \in \Delta$, where

$$s_\alpha(\beta) = \beta - (\alpha^\vee \cdot \beta)\alpha, \quad \alpha^\vee \equiv 2\alpha/|\alpha|^2 \qquad [12]$$

The set of reflections $\{s_\alpha \mid \alpha \in \Delta\}$ generates a group $G_\Delta$, known as a Coxeter group, or finite reflection group. The orbit of $\beta \in \Delta$ is the set of root vectors resulting from the action of the Coxeter group on it. The set of positive roots $\Delta_+$ may be defined in terms of a vector $U \in \mathbf{R}^r$, with $\alpha \cdot U \neq 0, \forall \alpha \in \Delta$, as the roots $\alpha \in \Delta$ such that $\alpha \cdot U > 0$. Given $\Delta_+$, there is a unique set of $r$ simple roots $\Pi = \{\alpha_j \mid j = 1, \ldots, r\}$ defined such that they span the root space and the coefficients $\{a_j\}$ in $\beta = \sum_{j=1}^r a_j \alpha_j$ for $\beta \in \Delta_+$ are all non-negative. The highest root $\alpha_h$, for which $\sum_{j=1}^r a_j$ is maximal, is then also determined uniquely. The subset of reflections $\{s_\alpha \mid \alpha \in \Pi\}$ in fact generates the Coxeter group $G_\Delta$. The products of $s_\alpha$, with $\alpha \in \Pi$, are subject solely to the relations $(s_\alpha s_\beta)^{m(\alpha,\beta)} = 1$, $\alpha, \beta \in \Pi$. The interpretation is that $s_\alpha s_\beta$ is a rotation in some plane by $2\pi/m(\alpha, \beta)$. The set of positive integers $m(\alpha, \beta)$ (with $m(\alpha, \alpha) = 1, \forall \alpha \in \Pi$) uniquely specifies the Coxeter group. The weight lattice $P(\Delta)$ is defined as the $\mathbf{Z}$-span of the fundamental weights $\{\lambda_j\}$, defined by $\alpha_j^\vee \cdot \lambda_k = \delta_{jk}, \forall \alpha_j \in \Pi$.

The root systems for finite reflection groups may be divided into two types: crystallographic and noncrystallographic. Crystallographic root systems satisfy the additional condition $\alpha^\vee \cdot \beta \in \mathbf{Z}, \forall \alpha, \beta \in \Delta$. The remaining noncrystallographic root systems are $H_3, H_4$, whose Coxeter groups are the symmetry groups of the icosahedron and four-dimensional 600-cell, respectively, and the dihedral group of order $2m, \{I_2(m) | m \geq 4\}$.

The explicit examples of the classical root systems, that is, $A, B, C$, and $D$ are given below. For the exceptional and noncrystallographic root systems, the reader is referred to Humphrey's book. In all cases, $\{\boldsymbol{e}_j\}$ denotes an orthonormal basis in $\mathbf{R}^r$.

1. $A_{r-1}$: This root system is related with the Lie algebra $su(r)$.

$$\Delta = \bigcup_{1 \leq j \leq k \leq r} \{\pm(\boldsymbol{e}_j - \boldsymbol{e}_k)\}, \qquad \prod = \bigcup_{j=1}^{r-1} \{\boldsymbol{e}_j - \boldsymbol{e}_{j+1}\} \tag{13}$$

2. $B_r$: This root system is associated with Lie algebra $so(2r+1)$. The long roots have $(\text{length})^2 = 2$ and short roots have $(\text{length})^2 = 1$:

$$\Delta = \bigcup_{1 \leq j \leq k \leq r} \{\pm \boldsymbol{e}_j \pm \boldsymbol{e}_k\} \cup_{j=1}^r \{\pm \boldsymbol{e}_j\}, \qquad \prod = \bigcup_{j=1}^{r-1} \{\boldsymbol{e}_j - \boldsymbol{e}_{j+1}\} \cup \{\boldsymbol{e}_r\} \tag{14}$$

3. $C_r$: This root system is associated with Lie algebra $sp(2r)$. The long roots have $(\text{length})^2 = 4$ and short roots have $(\text{length})^2 = 2$:

$$\Delta = \bigcup_{1 \leq j \leq k \leq r} \{\pm \boldsymbol{e}_j \pm \boldsymbol{e}_k\} \cup_{j=1}^r \{\pm 2\boldsymbol{e}_j\}, \qquad \prod = \bigcup_{j=1}^{r-1} \{\boldsymbol{e}_j - \boldsymbol{e}_{j+1}\} \cup \{2\boldsymbol{e}_r\} \tag{15}$$

4. $D_r$: This root system is associated with Lie algebra $so(2r)$:

$$\Delta = \bigcup_{1 \leq j \leq k \leq r} \{\pm \boldsymbol{e}_j \pm \boldsymbol{e}_k\}, \qquad \prod = \bigcup_{j=1}^{r-1} \{\boldsymbol{e}_j - \boldsymbol{e}_{j+1}\} \cup \{\boldsymbol{e}_{r-1} + \boldsymbol{e}_r\} \tag{16}$$

*See also:* Calogero–Moser–Sutherland Systems of Nonrelativistic and Relativistic Type; Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Functional Equations and Integrable Systems; Integrable Discrete Systems; Integrable Systems in Random Matrix Theory; Integrable Systems: Overview; Isochronous Systems; Toda Lattices.

## Further Reading

Calogero F (1971) Solution of the one-dimensional $N$-body problem with quadratic and/or inversely quadratic pair potentials. *Journal of Mathematical Physics* 12: 419–436.

Calogero F (2001) *Classical Many-Body Problems Amenable to Exact Treatments*. New York: Springer.

Dunkl C (2001) *Orthogonal Polynomials of Several Variables*. Cambridge: Cambridge University Press.

Humphreys JE (1990) *Reflection Groups and Coxeter Groups*. Cambridge: Cambridge University Press.

Macdonald IG (1995) *Symmetric Functions and Hall Polynomials*, 2nd edn. Oxford University Press.

Moser J (1975) Three integrable Hamiltonian systems connected with isospectral deformations. *Advances in Mathematics* 16: 197–220.

Olshanetsky MA and Perelomov AM (1983) Quantum integrable systems related to Lie algebras. *Physics Reports* 94: 313–404.

Ruijsenaars SNM (1999) Systems of Calogero–Moser Type. *CRM Series in Mathematical Physics* 1: 251–352. Springer.

Sasaki R (2001) *Quantum Calogero–Moser Models: Complete Integrability for All the Root Systems*. Proceedings Quantum Integrable Models and Their Applications, 195–240. World Scientific.

Sasaki R (2002) *Quantum vs Classical Calogero–Moser Systems*. NATO ARW Proceedings, Elba, Italy.

Stanley R (1989) Some combinatorial properties of Jack symmetric function. *Adv. Math.* 77: 76–115.

Sutherland B (1972) Exact results for a quantum many-body problem in one dimension. II. *Physical Review A* 5: 1372–1376.

# Quantum Central-Limit Theorems

**A F Verbeure**, Institute for Theoretical Physics, KU Leuven, Belgium



## Introduction

Statistical physics deals with systems with many degrees of freedom and the problems concern finding procedures for the extraction of relevant physical quantities for these extremely complex systems. The idea is to find relevant reduction procedures which map the complex systems onto simpler, tractable models at the price of introducing elements of uncertainty. Therefore, probability theory is a natural mathematical tool in statistical physics. Since the early days of statistical physics, in classical (Newtonian) physical systems, it is natural to model the observables by a collection of random variables acting on a probability space. Kolmogorovian probability techniques and results are the main tools in the development of classical statistical physics. A random variable is usually considered as a measurable function with expectation given as its integral with respect to a probability measure. Alternatively, a random variable can also be viewed as a multiplication operator by the associated function. Different random variables commute as multiplication operators, and one speaks of a commutative probabilistic model.

Now, looking at genuine quantum systems, in many cases the procedure mentioned above leads to commutative probabilistic models, but there exist the realms of physics where quantum noncommutative probabilistic concepts are unavoidable. Typical examples of such areas are quantum optics, low-temperature solid-state physics and ground-state physics such as quantum field theory. During the last 50 years physicists have developed more or less heuristic methods to deal with, for example, manifestations of fluctuations of typical quantum nature. In the last 30 years, mathematical foundations of such theories were also formulated, and a notion of quantum probability was launched as a branch of mathematical physics and mathematics (Cushen and Hudson 1971, Fannes and Quaegebeur 1983, Quaegebeur 1984, Hudson 1973, Giri and von Waldenfels 1978).

The aim of this article is to review briefly a few selected rigorous results concerning noncommutative limit theorems. This choice is made not only because of the author's interest but also for its close relation to concrete problems in statistical physics where one aims at understanding the macroscopic phenomena on the basis of the microscopic structure. A precise definition or formulation of a microscopic and a macroscopic system is of prime importance. The so-called algebraic approach of dynamical systems (Brattelli and Robinson 1979 and 2002) offers the necessary generality and mathematical framework to deal with classical and quantum, microscopic and macroscopic, finite and infinite systems. The observables of any system are assumed to be elements of an ($C^*$- or von Neumann) algebra $\mathcal{A}$, and the physical states are given by positive linear normalized functionals $\omega$ of $\mathcal{A}$, mapping the observables on their expectation values.

A common physicist's belief is that the macroscopic behavior of an idealized infinite system is described by a reduced set of macroscopic quantities (Sewell 1986). Some examples of these are the average densities of particles, energy, momentum, magnetic moment, etc. Analogously as the microscopic quantities, the macroscopic observables should be elements of an algebra, and macroscopic states of the system should be states on this algebra. The main problem is to construct the precise mathematical procedures to go from a given microscopic system to its macroscopic systems.

A well-known macroscopic system is the one given by the algebra of the observables at infinity (Lanford and Ruelle 1969) containing the spacial averages of local micro-observables, that is, for any local observable $A$ one considers the observable

$$A_\omega = \omega_\lim_{V\to\infty} \frac{1}{V}\int_V \mathrm{d}x\, \tau_x A$$

where $V$ is any finite volume in $\mathbb{R}^\nu$ and $\tau_x$ the translation over $x \in \mathbb{R}^\nu$, and where $\omega_\lim$ is the weak operator limit in the microstate $\omega$. The limits $A_\omega$ obtained correspond to the law of large numbers in probability. The algebra generated by these limit observables $\mathcal{A}_\omega = \{A_\omega \mid A \in \mathcal{A}\}$ is an abelian algebra of observables of a macroscopic system. This algebra can be identified with an algebra with pointwise product of measurable functions for some measure or macroscopic state.

The content of this review is to describe an analogous mapping from micro to macro but for a different type of scaling, namely the scaling of fluctuations. For any local observable $A \in \mathcal{A}$, one considers the limit

$$\lim_V \frac{1}{V^{1/2}}\int_V \mathrm{d}x(\tau_x A - \omega(\tau_x A)) \equiv F(A)$$

The problem consists in characterizing the $F(A)$ as an operator on a Hilbert space, called fluctuation

operator, and to specify the algebraic character of the set of all of these.

Based on this quantum central-limit theorem, one notes that not all locally different microscopic observables always yield different fluctuation operators. Hence the central-limit theorem realizes a well-defined procedure of coarse graining or reduction procedure which is handled by the mathematical notion of an equivalence relation on the microscopic observables yielding the same fluctuation operator.

In the following sections we discuss the preliminaries, the basic results about normal and abnormal fluctuations. Three model-independent applications are also discussed. In this review, we omit the properties of the so-called modulated fluctuations.

One should remark that we discuss only fluctuations in space. One can also consider timelike fluctuations. The theory of fluctuation operators for these has not been explicitly worked out so far. However, it is clear that for normal fluctuations the clustering properties of the time correlation functions will play a crucial role. On the other hand, typical properties of the structure of this fluctuation algebra may come up.

Another point which one has to stress is that all systems, which are treated in this review, are quasilocal systems. Other systems, for example, fermion systems, are note treated. But, in particular, fermion systems share many properties of quasilocality, and many of the results mentioned hold true also for fermion systems.

## Preliminaries

### Quantum Lattice Systems

Although all results we review can be extented to continuous or more general systems, modulo some technicalities, we limit ourself to quasilocal quantum dynamical lattice systems.

We consider the quasilocal algebra built on a $\nu$-dimensional lattice $\mathbb{Z}^\nu$. Let $\mathcal{D}(\mathbb{Z}^\nu)$ be the directed set of finite subsets of $\mathbb{Z}^\nu$ where the direction is the inclusion. With each point $x \in \mathbb{Z}^\nu$ we associate an algebra ($C^*$- or von Neumann algebra) $\mathcal{A}_x$, all copies of an algebra $\mathcal{A}$. For all $\Lambda \in \mathcal{D}(\mathbb{Z}^\nu)$, the tensor product $\otimes_{x\in\Lambda}\mathcal{A}_x$ is denoted by $\mathcal{A}_\Lambda$. We take $\mathcal{A}$ to be nuclear, then there exists a unique $C^*$-norm on $\mathcal{A}_\Lambda$. Every copy $\mathcal{A}_x$ is naturally embedded in $\mathcal{A}_\Lambda$. The family $\{\mathcal{A}_\Lambda\}_{\Lambda\in\mathcal{D}(\mathbb{Z}^\nu)}$ has the usual relations of locality and isotony:

$$[\mathcal{A}_{\Lambda_1}, \mathcal{A}_{\Lambda_2}] = 0 \quad \text{if } \Lambda_1 \cap \Lambda_2 = \emptyset \qquad [1]$$

$$\mathcal{A}_{\Lambda_1} \subseteq \mathcal{A}_{\Lambda_2} \quad \text{if } \Lambda_1 \subseteq \Lambda_2 \qquad [2]$$

Denote by $\mathcal{A}_\mathrm{L}$ all local observables, that is,

$$\mathcal{A}_\mathrm{L} = \bigcup_\Lambda \mathcal{A}_\Lambda$$

This algebra is naturally equipped with a $C^*$-norm $\|\cdot\|$ and its closure

$$\mathcal{B} = \overline{\mathcal{A}_\mathrm{L}}$$

is called a quasilocal $C^*$-algebra and considered as the microscopic algebra of observables of the system. Typical examples are spin systems where $\mathcal{A} = M_n$ is the $n \times n$ complex matrix algebra. In this case, every state $\omega$ of $\mathcal{B}$ is then locally normal, that is, there exists a family of density matrices $\{\rho_\Lambda \mid \Lambda \in \mathcal{D}(\mathbb{Z}^\nu)\}$ such that

$$\omega(A) = \mathrm{tr}\,\rho_\Lambda A \quad \text{for all } A \in \mathcal{A}_\Lambda$$

An important group of $*$-automorphisms of $\mathcal{B}$ is the group of space translations $\{\tau_x, x \in \mathbb{Z}^\nu\}$:

$$\tau_x : A_y \in \mathcal{A}_y \to \tau_x A_y = A_{x+y} \in \mathcal{A}_{y+x}$$

for all $A \in \mathcal{A}$.

Note that the quasilocal algebra $\mathcal{B}$ is asymptotically abelian for space translations: that is, for all $A, B \in \mathcal{B}$

$$\lim_{|x|\to\infty} \|[A, \tau_x B]\| = 0$$

A state $\omega$ of $\mathcal{B}$ represents a physical state of the system, assigning to every observable $A$ its expectation value $\omega(A)$. Therefore, this setting can be viewed as the quantum analog of the classical probabilistic setting. Sequences of random variables or observables can be constructed by considering an observable and its translates, that is, $\tau_x(A)_{x\in\mathbb{Z}^\nu}$ is a noncommutative random field. If a state $\omega$ is translation invariant, that is, $\omega \circ \tau_x = \omega$ for all $x$, then all $\tau_x(A)$ are identically distributed random variables. The mixing property of the random field is then expressed by the spatial correlations tending to zero:

$$\omega\big(\tau_x(A)\tau_y(B)\big) - \omega(\tau_x(A))\omega\big(\tau_y(B)\big) \to 0 \qquad [3]$$

if $|x - y| \to \infty$.

One of the basic limit theorems of probability theory is the weak law of large numbers. In this noncommutative setting the law of large numbers is translated into the problem of the convergence of space averages of an observable $A \in \mathcal{B}$. A first result was given by the mean ergodic theorem of von Neumann (1929). In Brattelli and Robinson (1979, 2002) one finds the following theorem: if the state $\omega$ is space translation invariant and mixing (see [3]) then for all $A, B$, and $C$ in $\mathcal{B}$

$$\lim_{\Lambda\to\mathbb{Z}^\nu} \omega\left(A \frac{1}{|\Lambda|}\left(\sum_{x\in\Lambda} \tau_x(B)\right) C\right) = \omega(AC)\omega(B) \qquad [4]$$

That is, in the GNS (Gelfand–Naimark–Segal) representation of the state $\omega$, the sequence $S_\Lambda(B) = 1/|\Lambda| \sum_{x\in\Lambda} \tau_x B$ converges weakly to a multiple of the identity: $S(B) \equiv \omega(B)\mathbb{1}$. This theorem, called the mean ergodic theorem, characterizes the class of states yielding a weak law of large numbers. Clearly, these limits $\{S(A)|A \in \mathcal{B}\}$ form a trivial abelian algebra of macroscopic observables.

Now we go a step further and consider space fluctuations. Define the local fluctuation of an observable $A$ in a homogeneous (spatial invariant) state $\omega$ by

$$F_\Lambda(A) = \frac{1}{|\Lambda|^{1/2}} \sum_{x\in\Lambda} (\tau_x A - \omega(A)) \qquad [5]$$

The problem is to give a rigorous meaning to $\lim F_\Lambda(A)$ for $\Lambda$ tending to $\mathbb{Z}^\nu$ in the sense of extending boxes. When does such a limit exist? What are the properties of the fluctuations or the limits $F(A) = \lim F_\Lambda(A)$, etc.? Again, the $F(A)$ are macroscopic variables of the microsystem.

Already we remark the following: if $A, B$ are strictly local elements, $A, B \in \mathcal{A}_\mathrm{L}$, then

$$\sum_{y\in\mathbb{Z}^\nu} [A, \tau_y B] \in \mathcal{A}_\mathrm{L}$$

and an easy computation yields, by [4],

$$\begin{aligned} &\text{weak}\lim_\Lambda [F_\Lambda(A), F_\Lambda(B)] \\ &= \text{weak}\lim_\Lambda \frac{1}{|\Lambda|} \sum_{x\in\Lambda} \tau_x \left( \sum_{y\in\Lambda} [A, \tau_{y-x} B] \right) \\ &= \text{weak}\lim_\Lambda \frac{1}{|\Lambda|} \sum_{x\in\Lambda} \tau_x \left( \sum_{y\in\mathbb{Z}^\nu} [A, \tau_y B] \right) \\ &= \sum_{y\in\mathbb{Z}^\nu} \omega([A, \tau_y B]) \equiv \mathrm{i}\sigma(A, B)\mathbb{1} \end{aligned}$$

that is, if the $F(A)$ and $F(B)$ limits do exist, then

$$[F(A), F(B)] = \mathrm{i}\sigma(A, B)\mathbb{1} \qquad [6]$$

This property indicates that fluctuations should have the same commutation relations as boson fields. If fluctuations can be characterized as macroscopic observables, they must satisfy the canonical commutation relations (CCRs). Therefore, in the next section we introduce the essentials on CCR representations.

### CCR Representations

We present the abstract Weyl CCR $C^*$-algebra. More details can be found in Brattelli and Robinson (1979, 2002) and in particular in Manuceau *et al.* (1973), where the case of a real test function space $(H, \sigma)$ with a possibly degenerate symplectic form $\sigma$ is treated. Hence, $H$ is a real vector space and $\sigma$ a bilinear, antisymmetric form on $H$.

Denote by $W(H, \sigma)$ the complex vector space generated by the functions $W(f)$, $f \in H$, defined by

$$W(f) : H \to \mathbb{C} : g \to W(f)g = \begin{cases} 0 & \text{if } f \neq g \\ 1 & \text{if } f = g \end{cases}$$

$W(H, \sigma)$ becomes an algebra with unit $W(0)$ for the product

$$W(f)W(g) = W(f+g)\mathrm{e}^{-(\mathrm{i}/2)\sigma(f,g)}; \quad f, g \in H$$

and a $^*$-algebra for the involution

$$W(f) \to W(f)^* = W(-f)$$

It becomes a $C^*$-algebra $C^*(H, \sigma)$ following the construction of Verbeure and Zagrebnov (1992). A linear functional $\omega$ of a $C^*$-algebra $C^*(H, \sigma)$ is called a state if $\omega(I) = 1$ and $\omega(A^*A) \geq 0$ for all $A \in C^*(H, \sigma)$ and $I = W(0)$. Every state gives rise to a representation through the GNS construction (Brattelli and Robinson 1979, 2002). In particular, $\omega$ is a state if for any choice of $A = \sum_j c_j W(f_j)$ we have

$$\sum_{j^k} c_j \bar{c}_k \omega\big(W(f_j - f_j)\big)\mathrm{e}^{-\mathrm{i}\sigma(f_j, f_k)} \geq 0$$

$$\omega(W(0)) = 1$$

A remark about the special case that $\sigma$ is degenerate is in order. Denote by $H_0$ the kernel of $\sigma$:

$$H_0 = \{f \in H | \sigma(f, g) = 0 \text{ for all } g \in H\}$$

If $H = H_0 \oplus H_1$ with $\sigma_1$ a nondegenerate symplectic form on $H_1$ and $\sigma_1$ equal to the restriction of $\sigma$ to $H_1$, we have that $C^*(H, \sigma)$ is a tensor product:

$$C^*(H, \sigma) = C^*(H_0, 0) \otimes C^*(H_1, \sigma_1)$$

Note that $C^*(H_0, 0)$ is abelian and that each positive-definite normalized functional $\varphi$,

$$\varphi : h \in H_0 \to \varphi(W(h))$$

defines a state $\omega(W(h)) = \varphi(W(h))$ on $C^*(H_0, 0)$.

Let $\xi$ be any character of the abelian additive group $H$, then the map $\tau_\xi$,

$$\tau_\xi W(f) = \xi(f) W(f)$$

extends to a $^*$-automorphism of $C^*(H, \sigma)$. Let $s$ be a positive symmetric bilinear form on $H$ such that for all $f, g \in H$:

$$\tfrac{1}{4}|\sigma(f, g)|^2 \leq s(f, f)\, s(g, g) \qquad [7]$$

and let $\omega_{s,\xi}$ be the linear functional on $C^*(H,\sigma)$ given by

$$\omega_{s,\xi}(W(h)) = \xi(h)\mathrm{e}^{-(1/2)s(h,h)} \qquad [8]$$

then it is straightforward (Brattelli and Robinson 1979, 2002) to check that $\omega_{s,\xi}$ is a state on $C^*(H,\sigma)$. All states of the type [8] are called quasifree states on the CCR algebra $C^*(H,\sigma)$.

A state $\omega$ of $C^*(H,\sigma)$ is called a regular state if, for all $f,g \in H$, the map $\lambda \in \mathbb{R} \to \omega(W(\lambda f + g))$ is continuous. The regularity property of a state yields the existence of a Bose field as follows. Let $(\mathcal{H},\pi,\Omega)$ be the GNS representation (Brattelli and Robinson 1979, 2002) of the state $w$, then the regularity of $w$ implies that there exists a real linear map $b: H \to \mathcal{L}(\mathcal{H})$ (linear operators on $\mathcal{H}$) such that $\forall f \in H$: $b(f)^* = b(f)$ and

$$\pi(W(f)) = \exp(\mathrm{i}b(x))$$

The map $b$ is called the Bose field satisfying the Bose field commutation relations:

$$[b(f), b(g)] = \mathrm{i}\sigma(f,g) \qquad [9]$$

Note that the Bose fields are state dependent. Note also already that if $\xi$ is a continuous character of $H$, then any quasifree state [8] is a regular state guaranteeing the existence of a Bose field.

## Normal Fluctuations

In this section we develop the theory of normal fluctuations for $\nu$-dimensional quantum lattice systems with a quasilocal structure (see the section "Quantum lattice systems") and for technical simplicity we assume that the local $C^*$-algebra $\mathcal{A}_x, x \in \mathbb{Z}^\nu$, are copies of the matrix algebra $M_n(\mathbb{C})$ of $n \times n$ complex matrices. Most of the results stated can be extended to the case where $\mathcal{A}_x$ is a general $C^*$-algebra (Goderis *et al.* 1989, 1990, Goderis and Vets 1989).

We consider a physical system $(\mathcal{B},\omega)$ where $\omega$ is a translation-invariant state of $\mathcal{B}$, that is, $\omega \circ \tau_x = \omega$ for all $x \in \mathbb{Z}^\nu$. Later on we extend the situation to a $C^*$-dynamical system $(\mathcal{B},\omega,\alpha_t)$ and analyze the properties of the dynamics $\alpha_t$ under the central limit.

For any local $A$ we introduced its local fluctuation in the state $\omega$ of the system:

$$F_\Lambda(A) = \frac{1}{|\Lambda|^{1/2}} \sum_{x\in\Lambda} (\tau_x A - \omega(A)) \qquad [10]$$

The main problem is to give a rigorous mathematical meaning to the limits

$$\lim_{\Lambda\to\infty} F_\Lambda(A) \equiv F(A)$$

where the limit is taken for any increasing $\mathbb{Z}^\nu$-absorbing sequence $\{\Lambda\}_\Lambda$ of finite volumes of $\mathbb{Z}^\nu$. The limits $F(A)$ are called the macroscopic fluctuation operators of the system $(\mathcal{B},\omega)$.

Already earlier work (Cushen and Hudson 1971, Sewell 1986) suggested that the fluctuations behave like bosons. We complete this idea by proving that one gets a well-defined representation of a CCR $C^*$-algebra of fluctuations uniquely defined by the original system $(\mathcal{B},\omega)$.

Denote by $\mathcal{A}_{\mathrm{L,sa}}$ and $\mathcal{B}_{\mathrm{sa}}$ the real vector space of the self-adjoint elements of $\mathcal{A}_{\mathrm{L}}$, respectively, $\mathcal{B}$.

**Definition 1** An observable $A \in \mathcal{B}_{\mathrm{sa}}$ satisfies the central-limit theorem if

(i) $\lim_\Lambda \omega(F_\Lambda(A)^2) \equiv s_\omega(A,A)$ exists and is finite, and
(ii) $\lim_\Lambda \omega(\mathrm{e}^{\mathrm{i}tF_\Lambda(A)}) = \mathrm{e}^{(-t/2)^2 s_\omega(A,A)}$ for all $t \in \mathbb{R}$.

Clearly, our definition coincides with the notion in terms of characteristic functions, for classical systems ($\mathcal{A}$ abelian) equivalent with the notion of convergence in distribution. For quantum systems, there does not exist a standard notion of "convergence in distribution." Only the concept of expectations is relevant. This does not exclude the notion of central-limit theorem in terms of the moments, which is the analog of the moment problem (Giri and von Waldenfels 1978).

**Definition 2** The system $(\mathcal{B},\omega)$ is said to have normal fluctuations if $\omega$ is translation invariant and if

(i) $\forall A,B \in \mathcal{A}_{\mathrm{L}}$

$$\sum_{x\in\mathbb{Z}^\nu} |\omega(A\tau_x B) - \omega(A)\omega(B)| < \infty$$

(ii) the central-limit theorem holds for all $A \in \mathcal{A}_{\mathrm{L,sa}}$.

Note that (i) implies that the state $\omega$ is mixing for space translations. Also by (i), one can define a sesquilinear form on $\mathcal{A}_{\mathrm{L}}$:

$$\begin{aligned}\langle A,B\rangle_\omega &= \lim_\Lambda \omega(F_\Lambda(A^*)F_\Lambda(B)) \\ &= \sum (\omega(A^*\tau_x B) - \omega(A^*)\omega(B))\end{aligned}$$

and denote

$$\begin{aligned}s_\omega(A,B) &= \mathrm{Re}\langle A,B\rangle_\omega \\ \sigma_\omega(A,B) &= 2\,\mathrm{Im}\langle A,B\rangle_\omega\end{aligned}$$

For $A,B \in \mathcal{A}_{\mathrm{L,sa}}$ one has

$$\sigma_\omega(A,B) = -\mathrm{i}\sum_{x\in\mathbb{Z}^\nu} \omega([A,\tau_x B]) \qquad [11]$$

$$s_\omega(A,A) = \langle A,A\rangle_\omega \qquad [12]$$

Clearly, $(\mathcal{A}_{\mathrm{L,sa}},\sigma_\omega)$ is a symplectic space and $s_\omega$ a non-negative symmetric bilinear form on $\mathcal{A}_{\mathrm{L,sa}}$.

Following the discussion in the section "CCR representations" we get a natural CCR $C^*$-algebra $C^*(\mathcal{A}_{L,sa}, \sigma_\omega)$ defined on this symplectic space. The following theorem is an essential step in the construction of a macroscopic physical system of fluctuations of the microsystem $(\mathcal{B}, \omega)$.

**Theorem 1** *If the system* $(\mathcal{B}, \omega)$ *has normal fluctuations, then the limits* $\{\lim_\Lambda \omega(e^{iF_\Lambda(A)}) = \exp((-1/2)s_\omega(A,A)), A \in \mathcal{A}_L\}$ *define a quasifree state* $\tilde{\omega}$ *on the CCR* $C^*$*-algebra* $C^*(\mathcal{A}_{L,sa}, \sigma_\omega)$ *by*

$$\tilde{\omega}(W(A)) = \exp\left(-\tfrac{1}{2}s_\omega(A,A)\right)$$

**Proof** The proof is clear from the definition [8] if one can prove that the positivity condition [7] holds. But the latter follows readily from

$$\begin{aligned}\tfrac{1}{4}|\sigma_\omega(A,B)|^2 &= \lim_\Lambda |\mathrm{Im}\,\omega(F_\Lambda(A)F_\Lambda(B))|^2 \\ &\leq \lim_\Lambda \omega(F_\Lambda(A)^2)\omega(F_\Lambda(B)^2) \\ &= s_\omega(A,A)s_\omega(B,B)\end{aligned}$$

by Schwarz inequality. □

This theorem indicates that the quantum-mechanical alternative for (classical) Gaussian measures are quasifree states on CCR algebras. However, the following basic question arises: is it possible to take the limits of products of the form

$$\lim_\Lambda \omega\left(e^{iF_\Lambda(A)}e^{iF_\Lambda(B)}\cdots\right)$$

and, if they exist, do they preserve the CCR structure? Clearly, this is a typical noncommutative problem.

Using the following general bounds: for $C^* = C$ and $D^* = D$ norm-bounded operators one has

$$\begin{aligned}&\left\|e^{i(C+D)} - e^{iC}\right\| \leq \|D\| \\ &\left\|[e^{iC}, e^{iD}]\right\| \leq \|[C,D]\| \\ &\left\|e^{i(C+D)} - e^{iC}e^{iD}\right\| \leq \tfrac{1}{2}\|[C,D]\|\end{aligned}$$

and by the expansion of the exponential function one proves easily that

$$\begin{aligned}\lim_\Lambda &\left\|e^{iF_\Lambda(A)}e^{iF_\Lambda(B)} - e^{i(F_\Lambda(A)+F_\Lambda(B))}\right. \\ &\left.\times\, e^{-(1/2)[F_\Lambda(A),F_\Lambda(B)]}\right\| = 0\end{aligned} \qquad [13]$$

if $A$ and $B$ are one-point observables, that is, if $A, B \in \mathcal{A}_{\{0\}}$. For general local elements the proof is somewhat more technical and can be based on a Bernstein-like argument (for details see Goderis and Vets (1989)). The property [13] can be seen as a Baker–Campbell–Hausdorff formula for fluctuations.

From [13], the mean ergodic theorem, and Theorem 1 we get:

**Theorem 2** *If the system* $(\mathcal{B}, \omega)$ *has normal fluctuations then for* $A, B \in \mathcal{A}_{L,sa}$:

$$\begin{aligned}\lim_\Lambda &\,\omega\left(e^{iF_\Lambda(A)}e^{iF_\Lambda(B)}\right) \\ &= \exp\left\{-\frac{1}{2}s_\omega(A+B, A+B) - \frac{i}{2}\sigma_\omega(A,B)\right\} \\ &= \tilde{\omega}(W(A)W(B))\end{aligned}$$

*with* $\tilde{\omega}$ *a quasifree state on the CCR algebra* $C^*(\mathcal{A}_{L,sa}^{\sigma_\omega})$.

Theorems 1 and 2 describe completely the topological and analytical aspects of the quantum central-limit theorem under the condition of normal fluctuations (Definition 2). In fact, the quantum central limit yields, for every microphysical system $(\mathcal{B}, \omega)$, a macrophysical system $(C^*(\mathcal{A}_{L,sa}, \sigma_\omega), \tilde{\omega})$ defined by the CCR $C^*$-algebra of fluctuation observables $C^*(\mathcal{A}_{L,sa}, \sigma_\omega)$ in the representation defined by the quasifree state $\tilde{\omega}$. As the state $\tilde{\omega}$ is a quasifree state, it is a regular state, that is, the map $\lambda \in \mathbb{R} \to \tilde{\omega}(W(\lambda A + B))$ is continuous. From in section "CCR representations" we know that this regularity property yields the existence of a Bose field, that is, there exists a real linear map

$$F: A \in \mathcal{A}_{L,sa} \to F(A)$$

where $F(A)$ is a self-adjoint operator on the GNS representation space $\tilde{\mathcal{H}}$ of $\tilde{\omega}$, such that for all $A, B \in \mathcal{A}_{L,sa}$:

$$[F(A), F(B)] = i\sigma_\omega(A,B)$$

Moreover, if one has a complex structure J on $(\mathcal{A}_{L,sa}, \sigma_\omega)$ such that $J^2 = -\mathbb{1}$ and for all $A, B \in \mathcal{A}_{L,sa}$:

$$\begin{aligned}&\sigma_\omega(JA, B) = -\sigma_\omega(A, JB) \\ &\sigma_\omega(A, JB) > 0\end{aligned}$$

then one defines the boson creation and annihilation operators

$$F^\pm(A) = \frac{1}{\sqrt{2}}(F(A) \mp iF(JA))$$

satisfying the usual boson commutation relations

$$[F^-(A), F^+(B)] = \sigma_\omega(A, JB) + i\sigma_\omega(A,B)$$

Finally, it is straightforward, nevertheless important, to remark that Theorems 1 and 2 hold true if the linear space of local observables $\mathcal{A}_{L,sa}$ is replaced by any of its subspaces. Some of them can have greater physical importance than others. This means that the quantum central-limit theorems can realize several macrophysical systems of fluctuations. But all of them are Bose field systems.

It is also important to remark that these results end up in giving a probabilistic canonical basis of the canonical commutation relations.

Now we analyze the notion of coarse graining due to the quantum central limit. Consider on $\mathcal{A}_L$ the sesquilinear form (see [11], [12]) again

$$\langle A, B\rangle_\omega = \sum_{x\in\mathbb{Z}^\nu} (\omega(A^*\tau_x B) - \omega(A)\omega(B)) = s_\omega(A,B) + \mathrm{i}\sigma_\omega(A,B) \qquad [14]$$

This form defines a topology on $\mathcal{A}_L$ which is not comparable with the operator topologies induced by $\omega$. In fact, this form is not closable in the weak, strong, ultraweak, or ultrastrong operator topologies.

We call $A$ and $B$ in $\mathcal{A}_L$ equivalent, denoted by $A \sim B$ if $\langle A-B, A-B\rangle_\omega = 0$. Clearly, this defines an equivalence relation on $\mathcal{A}_L$. The property of coarse graining is mathematically characterized by the following: for all $A, B \in \mathcal{A}_{L,sa}$ the relation $A \sim B$ is equivalent with $F(A) = F(B)$. Suppose first that $F(A) = F(B)$, then

$$[W(A), W(B)] = 0$$

hence $\sigma_\omega(A,B) = 0$. Therefore, from Theorem 1:

$$1 = \tilde{\omega}(W(A)W(B)^*) = \tilde{\omega}(W(A)W(-B)) = \tilde{\omega}(W(A-B)) = \exp\left(-\tfrac{1}{2}s_\omega(A-B, A-B)\right)$$

and from [12] and [14]: $\langle A-B, A-B\rangle_\omega = 0$. The converse is equally straightforward.

From this property, it follows immediately that, for example, the action of the translation group is trivial or that $F(\tau_x A) = F(A)$ for all $x \in \mathbb{Z}^\nu$. Therefore, the map $F: \mathcal{A}_{L,sa} \to C^*(\mathcal{A}_{L,sa}, \sigma_\omega)$ is not injective. This expresses the physical phenomenon of coarse graining and gives a mathematical signification of the fluctuations being macroscopic observables.

In the above, we have constructed the new macroscopic physical system of quantum fluctuations for any microsystem with the property of normal fluctuations (see Definition 2). The main problem remains: when the microsystem does have normal fluctuations. We end this section with the formulation of a general sufficient clustering condition for the microstate $\omega$ in order that the microsystem $(\mathcal{B}, \omega)$ has normal fluctuations.

Let $\Lambda, \Lambda' \in \mathcal{D}(\mathbb{Z}^\nu)$ and $\omega$ a translation invariant state, denote

$$\alpha^\omega(\Lambda, \Lambda') = \sup_{\substack{A\in\mathcal{A}_\Lambda:\|A\|=1\\ B\in\mathcal{A}_{\Lambda'}:\|B\|=1}} |\omega(AB) - \omega(A)\omega(B)|$$

The cluster function $\alpha_N^\omega(d)$ is defined by

$$\alpha_N^\omega(d) = \sup\{\alpha^\omega(\Lambda,\Lambda') : d(\Lambda,\Lambda') \geq d \text{ and } \max(|\Lambda|, |\Lambda'|) \leq N\}$$

where $N, d \in \mathbb{R}^+$ and $d(\Lambda, \Lambda')$ is the Euclidean distance between $\Lambda$ and $\Lambda'$. It is obvious that

$$\alpha_N^\omega(d) \leq \alpha_N^\omega(d') \quad \text{if } d \geq d'$$
$$\alpha_N^\omega(d) \leq \alpha_{N'}^\omega(d) \quad \text{if } N \leq N'$$

The clustering condition is expressed by the following scaling law:

$$\exists\delta > 0: \quad \lim_{N\to\infty} N^{1/2}\alpha_N^\omega\left(N^{1/2\nu-\delta}\right) = 0 \qquad [15]$$

or, equivalently,

$$\exists\delta > 0: \quad \lim_{N\to\infty} N^{\nu+\delta}\alpha^\omega_{N^{2(\nu+\delta)}}(N) = 0 \qquad [16]$$

Note that this condition implies that

$$\sum_{x\in\mathbb{Z}^\nu} \alpha_N^\omega(|x|) < \infty$$

that is, that the function $\alpha_N^\omega(\cdot)$ is an $L^1(\mathbb{Z}^\nu)$-function for all $N$. In fact, this condition corresponds to the uniform mixing condition in the commutative (classical) central-limit theorem (see, e.g., Ibragimov and Linnick (1971)). This condition can also be called the modulus of decoupling. Product states, for example, equilibrium states of mean-field systems are uniformly clustering with $\alpha^\omega(d) = 0$ for $d > 0$.

The normality of the fluctuations of the microsystem $(\mathcal{B}, \omega)$ for product states is proved and extensively studied in Goderis *et al.* (1989), and for states satisfying the condition [15] or [16] in Goderis and Vets (1989). In the latter case, the proofs are very technical and based on a generalization of the well-known Bernstein argument (Ibragimov and Linnick 1971) of the classical central-limit theorem to the noncommutative situation. A refinement of these arguments can be found in Goderis *et al.* (1990). For the sake of formal self-consistency we formulate the theorem:

**Theorem 3** (Central-limit theorem) *Take the microsystem $(\mathcal{B}, \omega)$ such that $\omega$ is lattice translation invariant and satisfies the clustering condition [15]; then the system has normal fluctuations for all elements of the vector space of local observables $\mathcal{A}_{L,sa}$.* □

In Goldshtein (1982) a noncommutative central-limit theorem is derived using similar techniques. The main difference, however, is its strictly local character, namely for one local operator separately. The conditions depend on the spectral properties of the operator. It excludes a global approach resulting in a CCR algebra structure.

Even for quantum lattice systems, it is not straightforward to check whether a state satisfies

the degree of mixing as expressed in conditions [15]–[16]. Clearly, one expects the condition to hold for equilibrium states at high enough temperatures. For quantum spin chains, a theorem analogous with Theorem 3 under weaker conditions than [15] is proved for example, in Matsui (2003).

So far we have reviewed the quantum central-limit theorem for physical $C^*$-spin systems $(\mathcal{B}, \omega)$ with normal fluctuations.

Now we extend the physical system to a $C^*$-dynamical system $(\mathcal{B}, \omega, \alpha_t)$ (Brattelli and Robinson 1979, 2002) and we investigate the properties of the dynamics $\alpha_t$ under the central limit. As usual, the dynamics is supposed to be of the short-range type in order to guarantee the norm limit:

$$\alpha_t(\cdot) = n - \lim_{\Lambda} \mathrm{e}^{\mathrm{i}tH_\Lambda} \cdot \mathrm{e}^{-\mathrm{i}tH_\Lambda}$$

and space homogeneous $\alpha_t \cdot \tau_x = \tau_x \cdot \alpha_t, \forall t \in \mathbb{R}, \forall x \in \mathbb{Z}^\nu$. We suppose that the state $\omega$ is both space as time translation invariant. Moreover, we assume that the state $\omega$ satisfies the mixing condition [15] for normal fluctuations.

In [10] we defined, for every local $A \in \mathcal{A}_{\mathrm{L,sa}}$, the local fluctuation $F_\Lambda(A)$ and obtained a clear meaning of $F(A) = \lim_\Lambda F_\Lambda(A)$ from the central-limit theorem. Now we are interested in the dynamics of the fluctuations $F(A)$. Clearly, for all $A \in \mathcal{A}_{\mathrm{L,sa}}$ and all finite $\Lambda$:

$$\alpha_t F_\Lambda(A) = F_\Lambda(\alpha_t A) \qquad [17]$$

and one is tempted to define the dynamics $\tilde{\alpha}_t$ of the fluctuations in the $\Lambda$-limit by the formula

$$\tilde{\alpha}_t F(A) = F(\alpha_t A) \qquad [18]$$

Note, however, that in general $\alpha_t A$ is not a local element of $\mathcal{A}_{\mathrm{L,sa}}$. It is unclear whether the central limit of elements of the type $\alpha_t A$, with $A \in \mathcal{A}_{\mathrm{L,sa}}$ exists or not and hence whether one can give a meaning to $F(\alpha_t A)$. Moreover, if $F(\alpha_t A)$ exists, it remains to prove that $(\tilde{\alpha}_t)_t$ defines a weakly continuous group of $*$-automorphisms on the fluctuation CCR algebra $\tilde{\mathcal{M}} = C^*(\mathcal{A}_{\mathrm{L,sa}}, \sigma_\omega)''$ (the von Neumann algebra generated by the $\tilde{\omega}$-representation of $C^*(\mathcal{A}_{\mathrm{L,sa}}, \sigma_\omega)$). All this needs a proof. In Goderis *et al.* (1990), one finds the proof of the following basic theorem about the dynamics.

**Theorem 4** *Under the conditions on the dynamics* $\alpha_t$ *and on the state* $\omega$ *expressed above, the limit* $F(\alpha_t A) = \lim_\Lambda F_\Lambda(\alpha_t A)$ *exists as a central limit as in Theorem 2, and the maps* $\tilde{\alpha}_t$ *defined by* [18] *extend to a weakly continuous one-parameter group of* $*$*-automorphisms of the von Neumann algebra* $\tilde{\mathcal{M}}$. *The quasifree state* $\tilde{\omega}$ *is* $\tilde{\alpha}_t$*-invariant (time invariant).*

This theorem yields the existence of a dynamics $\tilde{\alpha}_t$ on the fluctuations algebra and shows that it is of the quasifree type

$$\tilde{\alpha}_t F(A) = F(\alpha_t A)$$

where $F(A)$ is a representation of a Bose field in a quasifree state $\tilde{\omega}$, the noncommutative version of a Gaussian distribution. In physical terms, it also means that any microdynamics $\alpha_t$ induces a linear process on the level of its fluctuations.

We can conclude that on the basis of the Theorems 3 and 4 the quantum central-limit theorem realized a map from the microdynamical system $(\mathcal{B}, \omega, \alpha_t)$ to a macrodynamical system $(C^*(\mathcal{A}_{\mathrm{L,sa}}, \sigma_\omega), \tilde{\omega}, \tilde{\alpha}_t)$ of the quantum fluctuations. The latter system is a quasifree Boson system.

Note that, contrary to the central-limit theorem, the law of large numbers [4] maps local observables to their averages forming a trivial commutative algebra of macro-observables. The macrodynamics is mapped to a trivial dynamics as well. Therefore, the consideration of law of large numbers does not allow one to observe genuine quantum phenomena. On the other hand, on the level of the fluctuations, macroscopic quantum phenomena are observable.

## Abnormal Fluctuations

The results about normal fluctuations in the last section contain two essential elements. On the one hand, the central limit has to exist. The condition in order that this occurs is the validity of the cluster condition ([15] or [16]) guaranteeing the normality of the fluctuations. On the other hand, there is the reconstruction theorem, identifying the CCR algebra representation of the fluctuation observables or operators in the quasifree state, which is denoted by $\tilde{\omega}$.

The cluster condition is in general not satisfied for systems with long-range correlations, for example, for equilibrium states at low temperatures with phase transitions. It is a challenging question to also study in this case the existence of fluctuations operators and, if they exist, to study their mathematical structure. Here we detect structures other than the CCR structure, other states or distributions different from quasifree states, etc.

Progress in the elucidation of all these questions started with a detailed study of abnormal fluctuations in the harmonic and anharmonic crystal models (Verbeure and Zagrebnov 1992, Momont *et al.* 1997). More general Lie algebras are obtained than the Heisenberg Lie algebra of the CCR algebra, and more general states $\tilde{\omega}$ or quantum distributions

are computed beyond quasifree states, which is the case for normal fluctuations.

Abnormal fluctuations turn up, if one has an ergodic state $\omega$ with long-range correlations. We have in mind continuous (second-order) phase transitions, then typically, for example, the heat capacity or some more general susceptibilities diverge at critical points or lines. This means that normally scaled (with the factor $|\Lambda|^{-1/2}$) fluctuations of some observables diverge. This is equivalent with the divergence of sums of the type

$$\sum_{x\in\mathbb{Z}^\nu}(\omega(A\tau_x A)-\omega(A)^2)$$

for some local observable $A$.

In order to deal with these situations, we rescale the local fluctuations. One determines a scaling index $\delta_A \in (-1/2, 1/2)$, depending on the observable $A$, such that the abnormally scaled local fluctuations

$$F_\Lambda^{\delta_A} = |\Lambda|^{-\delta_A} F_\Lambda(A)$$

with $F_\Lambda(A)$ as in [10], yield a nontrivial characteristic function: $\forall t \in \mathbb{R}$,

$$\lim_\Lambda \omega_\Lambda(\mathrm{e}^{\mathrm{i}tF_\Lambda^{\delta_A}(A)}) \equiv \phi_A(t) \qquad [19]$$

where we limit our discussion to states $\omega_\Lambda$ local Gibbs states. The index $\delta_A$ is a measure for the abnormality of the fluctuation of $A$. Note that $\delta_A = -1/2$ yields a triviality and that $\delta_A = 1/2$ would lead to a law of large numbers (theory of averages). Observe also that in general the characteristic function $\phi_A$ or the corresponding state $\tilde{\omega}$ need not be Gaussian or quasifree.

In the physics literature, one describes the long-range order by means of the asymptotic form of the connected two-point function in terms of the critical exponent $\eta$

$$\omega_\Lambda(a\tau_x A) - \omega_\Lambda(A)^2 \simeq 0\left(\frac{1}{|x|^{\nu-2+\eta}}\right); \quad |x| \to \infty \qquad [20]$$

Our scaling index $\delta_A$ is related to the critical exponent $\eta$ by the straightforward relation

$$\eta = 2 - 2\nu\delta_A$$

As stated above, the index $\delta_A$ is determined by the existence of the central limit and explicitly computed in several model calculations, for example, Verbeure and Zagrebnov (1992), and for equilibrium states. Apart from the strong model dependence, the indices also depend strongly on the chosen boundary conditions. This fact draws a new light on the universality of the critical exponents.

Suppose now that the indices $\delta_A$ are determined by the existence of the central limit [19]. The next problem is to find out whether also in these cases a reconstruction theorem, comparable to, for example, Theorem 2, can be proved giving again a mathematical meaning to the limits

$$\lim_\Lambda F_\Lambda^{\delta_A}(A) \equiv F^{\delta_A}(A) \qquad [21]$$

as operators, in general unbounded, on a Hilbert space.

Here we develop a proof of the Lie algebra character of the abnormal fluctuations under the conditions: (1) the $\delta$-indices are determined by the existence of the variances (second moments), and (2) the existence of the third moments (for more details see, e.g., Momont *et al.* (1997)).

Consider a local algebra, namely an $n$-dimensional vector space $\mathcal{G}$ with basis $\{v_i\}_{i=1,\ldots,n}$ and product

$$v_j \cdot v_k \equiv [v_j, v_k] = \sum_{\ell=1}^{n} c_{jk}^\ell v_\ell \qquad [22]$$

with structure constants $c_{jk}^\ell$ satisfying

$$c_{jk}^\ell + c_{kj}^\ell = 0$$
$$\sum_r (c_{ij}^r c_{rk}^s + c_{jk}^r c_{ri}^s + c_{ki}^r c_{rj}^s) = 0$$

Consider the concrete Lie algebra basis of operators in $\mathcal{A}_{\{0\}}$

$$\{L_0 = i\mathbb{1}, L_1, \ldots, L_m\}, \quad m < \infty$$

such that $L_j^* = -L_j$, $j = 0, 1, \ldots, m$ and $\omega(L_j) = \lim_\Lambda \omega_\Lambda(L_j) = 0$ for $j > 0$. Clearly, $\omega_\Lambda(L_0) = i$ for all $\Lambda$, and the $\{L_i\}$ satisfy eqn [22]. Because of the special choices of $L_0$ one has $c_{ok}^\ell = c_{ko}^\ell = 0$ and $c_{jk}^o = -\mathrm{i}\lim_\Lambda \omega_\Lambda([L_j, L_k])$. We consider now the fluctuations of these generators and we are looking for a characterization of the Lie algebra of the fluctuations if any.

For a translation-invariant local state $\omega_\Lambda$, $\Lambda \subseteq \mathbb{Z}^\nu$, such that $\omega = \lim_\Lambda \omega_\Lambda$ is mixing, define the local fluctuations, for $j = 1, \ldots, m$,

$$F_j^{\delta_j}, \Lambda = \frac{1}{|\Lambda|^{1/2+\delta_j}} \sum_{x\in\Lambda} \left(\tau_x L_j - \omega_\Lambda(L_j)\right) \qquad [23]$$

and for notational convenience, take

$$F_{0,\Lambda} = \mathrm{i}\mathbb{1}$$

Now we formulate the conditions for our purposes.

**Condition A** We assume that the parameters $\delta_j$ are determined by the existence of the finite and nontrivial variances: for all $j = 1, \ldots, m$,

$$0 < \lim_\Lambda \omega_\Lambda\left((F_{j,\Lambda}^{\delta_j})^2\right) < \infty \qquad [24]$$

After reordering, take $1/2 > \delta_1 \geq \delta_2 \geq \cdots \geq \delta_m > -1/2$.

**Condition B** Assume that all third moments are finite, that is,

$$\lim_{\Lambda} \left| \omega_\Lambda \left( F^{\delta_j}_{j,\Lambda} F^{\delta_k}_{k,\Lambda} F^{\delta_\ell}_{\ell,\Lambda} \right) \right| < \infty$$

We have in mind, that the $\omega_\Lambda$'s are Gibbs states for some local Hamiltonians with some specific boundary conditions. The limit $\Lambda \to \mathbb{Z}^\nu$ may depend very strongly on these boundary conditions, in the sense that they are visible in the values of the indices $\delta_j$ (see, e.g., Verbeure and Zagrebnov (1992)). If for some $j \geq 1$, the corresponding $\delta_j = 0$ then the operator $L_j$ has a normal fluctuation operator

$$F^{\delta_j}_j = \lim_{\Lambda} F^{\delta_j}_{j,\Lambda} \qquad [25]$$

where the limit is understood in the sense of Condition A, namely a finite nontrivial variance. If, for some $j \geq 1$, the corresponding $\delta_j \neq 0$, then the fluctuation [25] is called an abnormal fluctuation operator. In order to satisfy Condition A, it happens sometimes that $\delta_j$ has to be chosen negative (see, e.g., Verbeure and Zagrebnov (1992)). In this case, it is reasonable to limit our discussion to the situation that all $\delta_j > -1/2$.

On the basis of Condition A, the limit set $\{F^{\delta_j}_j\}_{j=0,\ldots,m}$ of fluctuation operators generates a Hilbert space $\mathcal{H}$ with scalar product

$$\left( F^{\delta_j}_j, F^{\delta_k}_k \right) = \lim_{\Lambda} \omega_\Lambda \left( (F^{\delta_j}_{j,\Lambda})^* F^{\delta_k}_{k,\Lambda} \right) \qquad [26]$$

On the basis of Condition B, the fluctuation operators are defined as multiplication operators of the Hilbert space $\mathcal{H}$. Note that the Conditions A and B are not sufficient to obtain a characteristic function. However, they are sufficient to obtain the notion of fluctuation operator. Now we proceed to clarify the Lie algebra character of these fluctuation operators on $\mathcal{H}$.

Consider the Lie product of two local fluctuations for a finite $\Lambda$, one gets

$$\left[ F^{\delta_j}_{j,\Lambda}, F^{\delta_k}_{k,\Lambda} \right] = \sum_{\ell=0}^{m} c^\ell_{jk}(\Lambda) F^{\delta_\ell}_{\ell,\Lambda} \qquad [27]$$

with

$$c^\ell_{jk}(\Lambda) = \frac{c^\ell_{jk}}{|\Lambda|^{1/2+\delta_j+\delta_k-\delta_\ell}}, \quad \ell = 1, \ldots, m$$

$$c^0_{jk}(\Lambda) = |\Lambda|^{-\delta_j-\delta_k} \sum_{\ell=0}^{m} c^\ell_{jk} \omega_\Lambda (F^{\delta_\Lambda}_{\ell,\Lambda})$$

It is an easy exercise to check that the $\{c^\ell_{jk}(\Lambda)\}$ are the structure coefficients of a Lie algebra $\mathcal{G}(\Lambda)$. Hence, by considering local fluctuations, one constructs a map from the Lie algebra $\mathcal{G}$ onto the Lie algebra $\mathcal{G}(\Lambda)$ by a nontrivial change of the structure constants. When the transformed structure constants approach a well-defined limit, a new nonisomorphic Lie algebra might appear. The limit algebra $\mathcal{G}(\mathbb{Z}^\nu)$, called the contracted one of the original one $\mathcal{G}$ is always nonsemisimple. This contraction is a typical Inönü–Wigner contraction (Inönü and Wigner 1953). About the limit algebra $\mathcal{G}(\mathbb{Z}^\nu)$, the following results are obtained (see Momont *et al.* (1997)):

$$\lim_{\Lambda} c^\ell_{jk}(\Lambda) = \begin{cases} 0 & \text{if } \frac{1}{2} + \delta_j + \delta_k - \delta_\ell > 0 \\ c^\ell_{jk} & \text{if } \ldots\ldots\ldots\ldots\ldots = 0 \\ 0 & \text{if } \ldots\ldots\ldots\ldots\ldots < 0 \end{cases} \qquad [28]$$

It is interesting to distinguish a number of special cases:

1. If all fluctuations are normal, one recovers the Heisenberg algebra of the canonical commutation relations with the right symplectic form $\sigma_\omega$.
2. If $1/2 + \delta_j + \delta_k - \delta_\ell > 0$ for all $j, k, \ell$ one obtains an abelian Lie algebra of fluctuations.
3. One gets the richest structure if $1/2 + \delta_j + \delta_k - \delta_\ell = 0$ for all $j, k, \ell$ or for some of them. One notes a phenomenon of scale invariance, the $c^\ell_{jk}(\Lambda)$ are $\Lambda$-independent. Algebras different from the CCR algebra are observed. A particularly interesting case turns up if $\delta_j = -\delta_k \neq 0$, that is, one of the indices is negative, for example, $\delta_j < 0$, the corresponding fluctuation $F^{\delta_j}_j$ shows a property of space squeezing, and then $\delta_k > 0$, the fluctuation $F^{\delta_k}_k$ expresses the property of space dilation. These phenomena are observed and computed in several models (see, e.g., Verbeure and Zagrebnov (1992)). This yields in particular a microscopic explanation of the phenomenon of squeezing (squeezed states and all that) in quantum optics. We refer also to the section "Spontaneous symmetry breaking" for this phenomenon as being the basis of the construction of the Goldstone normal modes of the Goldstone particle appearing in systems showing spontaneous symmetry breakdown.

## Some Applications

The notion of fluctuation operator as presented above, and the mathematical structure of the algebra of fluctuations have been tested in several soluble models. Many applications of this theory of quantum fluctuations can be found in the list of references. Here we are not entering into the details

of any model, but we limit ourselves to mention three applications which are of a general nature and totally model independent.

### Conservation of the KMS Property under the Transition from Micro to Macro

Suppose that we start with a micro-dynamical system $(\mathcal{B}, \omega, \alpha_t)$ with normal fluctuations, that is, we are in the situation as treated in the section "Normal fluctuations." Hence, we know that the quantum central-limit theorem maps the system $(\mathcal{B}, \omega, \alpha_t)$ onto the macrodynamical system $(C^*(\mathcal{A}_{\mathrm{L,sa}}, \sigma_\omega), \tilde{\omega}, \tilde{\alpha}_t)$ of quantum fluctuations.

If the microstate $\omega$ is $\alpha_t$-time invariant ($\omega \cdot \alpha_t = \omega$ for all $t \in \mathbb{R}$), then it also follows readily that the macrostate $\tilde{\omega}$ is $\tilde{\alpha}_t$-time invariant (see Theorem 4, i.e., $\tilde{\omega} \cdot \tilde{\alpha}_t = \tilde{\omega}$ for all $t \in \mathbb{R}$).

A less trivial question to pose is: suppose that the microstate $\omega$ is an equilibrium state for the micro-dynamics $\alpha_t$, is then the macrostate $\tilde{\omega}$ also an equilibrium state for the macrodynamics $\tilde{\alpha}_t$ of the fluctuations? In Goderis *et al.* (1990) this question is answered positively in the following more technical sense: if $\omega$ is an $\alpha_t$-KMS state of $\mathcal{B}$ at inverse temperature $\beta$, then $\tilde{\omega}$ is an $\tilde{\alpha}_t$-KMS state at the same temperature.

This property proves that the notion of equilibrium is preserved under the operation of coarse graining induced by the central-limit theorem. This statement constitutes a proof of one of the basic assumptions of the phenomenological theory of Onsager about small oscillations around equilibrium.

This result also yields a contribution to the discussion whether or not quantum systems should be described at a macroscopic level by classical observables. The result above states that the macroscopic fluctuation observables behave classically if and only if they are time invariant. In other words, it can only be expected *a priori* that conserved quantities behave classically. In principle, other observables follow a quantum dynamics.

### Linear Response Theory

In particular, in the study of equilibrium states (KMS states) a standard procedure is to perturb the system and to study the response of the system as a function of the perturbation. The response elucidates many, if not all, of the properties of the equilibrium state.

Technically, one considers a perturbation of the dynamics by adding a term to the Hamiltonian. One expands the perturbed dynamics in terms of the perturbation and the unperturbed dynamics. It is often argued that when the perturbation is small, one can limit the study of the response to the first-order term in the perturbation in the corresponding Dyson expansion. This is the basis of what is called the "linear response theory of Kubo."

A long-term debate is going on about the validity of the linear response theory. The question is how to understand from a microscopic point of view the validity of the response theory being linear or not. One must realize that the linear response theory actually observed in macroscopic systems seems to have a significant range of validity beyond the criticism being expressed about it.

Here we discuss the main result of the paper (Goderis *et al.* 1991) in which contours are sketched for the exactness of the response being linear.

We assume:

1. that the microdynamics $\alpha_t$ is the norm-limit of the local dynamics $\alpha_t^\Lambda = \mathrm{e}^{\mathrm{i}tH_\Lambda} \cdot \mathrm{e}^{-\mathrm{i}tH_\Lambda}$, where $H_\Lambda$ contains only standard finite-range interactions (as in the section "Normal fluctuations");
2. that the $\omega_\Lambda$ are states such that $\omega = \lim_\Lambda \omega_\Lambda$ is a state which is time and space translation invariant; and
3. that $\omega$ satisfies the cluster condition [15] or [16].

From the time invariance of the state, one has a Hamiltonian GNS representation of the dynamics: $\alpha_t = \mathrm{e}^{\mathrm{i}tH} \cdot \mathrm{e}^{-\mathrm{i}tH}$. On the basis of Theorem 4, one has the dynamics $\tilde{\alpha}_t$ of the fluctuation algebra $C^*(\mathcal{A}_{\mathrm{L,sa}}, \sigma_\omega)$ in the state $\tilde{\omega}$. This GNS representation yields a Hamiltonian representation for $\tilde{\alpha}_t$:

$$\tilde{\alpha}_t = \mathrm{e}^{\mathrm{i}t\tilde{H}} \cdot \mathrm{e}^{-\mathrm{i}t\tilde{H}}$$

Now take any local perturbation $P \in \mathcal{A}_{\mathrm{L,sa}}$ of $\alpha_t$, namely

$$\alpha_{t,\Lambda}^P = \mathrm{e}^{\mathrm{i}t(H + F_\Lambda(P))} \cdot \mathrm{e}^{-\mathrm{i}t(H + F_\Lambda(P))}$$

where $F_\Lambda(P)$ is the local fluctuation of $P$ in $\omega$. Then one proves the following central-limit theorem (Goderis *et al.* 1991): for all $A$ and $B$ in $\mathcal{A}_{\mathrm{L,sa}}$, one has the perturbed dynamics

$$\tilde{\alpha}_t^P = \mathrm{e}^{\mathrm{i}t(\tilde{H} + F(P))} \cdot \mathrm{e}^{-\mathrm{i}t(\tilde{H} + F(P))}$$

of the fluctuation algebra in the sense of [18]:

$$\tilde{\alpha}_t^P F(A) = \lim_\Lambda F(\alpha_{t,\Lambda}^P(A))$$

This proves the existence and the explicit form of the perturbed dynamics lifted to the level of the fluctuations. In particular, one has

$$\lim_\Lambda \omega_\Lambda \left( \alpha_{t,\Lambda}^P (F_\Lambda(A)) \right) = \tilde{\omega}(\tilde{\alpha}_t^P F(A))$$

This is nothing but the existence of the relaxation function of Kubo but lifted to the level of the fluctuations and instead of dealing with strictly local observables here one considers fluctuations.

Assume, furthermore, that the state $\omega$ is an $(\alpha_t, \beta)$-KMS state; then one derives readily Kubo's famous formula of his linear response theory:

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\omega}(\tilde{\alpha}_t^P F(A)) = \mathrm{i}\tilde{\omega}([F(P), \tilde{\alpha}_t F(A)])$$

which shows full linearity in the perturbation observable $P$. Kubo's formula arises as the central limit of the microscopic response to the dynamics perturbed by a fluctuation observable. We remark that if $\omega$ is an equilibrium state, then the right-hand side of the formula above can be expressed in terms of the Duhamel two-point function, which is the common way of doing in linear response theory.

### Spontaneous Symmetry Breaking

SSB is one of the basic phenomena accompanying collective phenomena, such as phase transitions in statistical mechanics, or specific ground states in field theory. SSB goes back to the Goldstone theorem. There are many different situations to consider, for example, in the case of short-range interactions, it is typical that SSB yields a dynamics which remains symmetric, whereas for long-range interactions SSB also breaks the symmetry of the dynamics. However, in all cases the physics literature predicts the appearance of a particular particle, namely the Goldstone boson, to appear as a result of SSB. The theory of fluctuation operators allows the construction of the canonical coordinates of this particle. The most general result can be found in Michoel and Verbeure (2001). We sketch the essentials in two cases, namely for systems of long-range interactions (mean fields) and for systems with short-range interactions.

**Long-range (mean-field) interactions** Here we give explicitly the example of the strong-coupling BCS model in one dimension ($\nu = 1$). The microscopic algebra of observables is $\mathcal{B} = \otimes_i (M_2)_i$, where $M_2$ is the algebra of $2 \times 2$ complex matrices. The local Hamiltonian of the models is given by

$$H_N = \epsilon \sum_{i=-N}^{N} \sigma_i^z - \frac{1}{2N+1} \sum_{i,j=-N}^{N} \sigma_i^+ \sigma_j^-$$
$$0 < \epsilon < \tfrac{1}{2}$$

where $\sigma^z, \sigma^\pm$ are the usual $2 \times 2$ Pauli matrices. In the thermodynamic limit, the KMS equation has the following product state solutions: $\omega_\lambda = \otimes_i \operatorname{tr}\rho_\lambda$, where

$$\rho_\lambda = \frac{\mathrm{e}^{-\beta h_\lambda}}{\operatorname{tr} \mathrm{e}^{-\beta h_\lambda}}, \quad \lambda = \operatorname{tr} \rho_\lambda \sigma^- = \omega_\lambda(\sigma^-)$$
$$h_\lambda = \epsilon\sigma^z - \lambda\sigma^+ - \lambda\sigma^-$$

Note that $\lambda = \operatorname{tr}\rho_\lambda\sigma^-$ is a nonlinear equation for $\lambda$ whose solutions determine the density matrix $\rho_\lambda$. This equation always has the solution $\lambda = 0$, describing the so-called normal phase. For $\beta > \beta_c$, with $th\beta_c\epsilon = 2\epsilon$, one has a solution $\lambda \neq 0$, describing the superconducting phase. Remark that if $\lambda$ is a solution, then also $\lambda e^{i\phi}$ for all $\phi$ is a solution as well. It is clear that $H_N$ is invariant under the continuous gauge transformation automorphism group $\mathcal{G} = \{\gamma_\varphi \mid \varphi \in [0, 2\pi]\}$ of $\mathcal{B}$:

$$\gamma_\varphi(\sigma_i^+) = e^{-i\varphi}\sigma_i^+$$

Hence $\mathcal{G}$ is a symmetry group. On the other hand: $\omega_\lambda(\gamma_\varphi(\sigma_i^+)) = e^{-i\varphi}\omega_\lambda(\sigma_i^+) \neq \omega_\lambda(\sigma_i^+)$. The gauge group $\mathcal{G}$ is spontaneously broken. Remark also that the gauge transformations are implemented locally by the charges

$$Q_N = \sum_{j=-N}^{N} \sigma_i^z, \quad \text{i.e.,} \ \gamma_\varphi(\sigma_i^+) = \mathrm{e}^{-\mathrm{i}\varphi Q_N}\sigma_i^+\mathrm{e}^{\mathrm{i}\varphi Q_N}$$

and $\sigma^z$ is the symmetry generator density. As the states $\omega_\lambda$ are product states, all fluctuations are normal (see the section "Normal fluctuations"). One considers the local operators

$$Q = \frac{|\lambda|^2}{\mu^2}\sigma^z + \frac{\epsilon}{\mu^2}(\lambda\sigma^+ + \bar{\lambda}\sigma^-)$$
$$P = \frac{\mathrm{i}}{\mu}(\lambda\sigma^+ - \bar{\lambda}\sigma^-)$$

where $\mu = (\epsilon^2 + |\lambda|^2)^{1/2}$. Note that $P$ is essentially the order parameter operator, that is, the operator $P$ is breaking the symmetry:

$$\frac{\mathrm{d}}{\mathrm{d}\varphi}\omega_\lambda(\gamma_\varphi(A)) \neq 0, \quad \omega_\lambda(A) = 0$$

On the other hand, $Q$ is essentially the generator of the symmetry $\sigma^z$ normalized to zero, that is, $\omega_\lambda(Q) = 0$.

Michoel and Verbeure (2001) proved in detail that the fluctuations $F(Q)$ and $F(P)$ form a canonical pair

$$[F(Q), F(P)] = \mathrm{i}\frac{4|\lambda|^2}{\mu}$$

and that they behave, under the time evolution, as harmonic oscillator coordinates oscillating with a

frequency equal to $2\mu$. This frequency is called a plasmon frequency. Moreover, the variances are

$$\tilde{\omega}_\lambda(F(Q)^2) = \frac{|\lambda|^2}{\mu^2} = \tilde{\omega}_\lambda(F(P)^2)$$

This means that these coordinates vanish or disappear if $\lambda = 0$. The coordinates $F(Q)$ and $F(P)$ are the canonical coordinates of a particle appearing only if there is spontaneous symmetry breakdown. They are the canonical coordinates of the Goldstone boson, which arise if SSB occurs.

**Short-range interactions** An analogous result, as for long-range interactions, can be derived for systems with short-range interactions. However, in this case we have equilibrium states with poor cluster properties. We are now in the situation as described in the "Abnormal Fluctuations" section. Also in this case we have the phenomenon of SSB, which shows the appearance of a Goldstone particle. Also in this case one is able to construct its canonical coordinates. The details of this construction can be found in Michoel and Verbeure (2001). Here we give a heuristic picture of this construction.

Consider again a microsystem $(\mathcal{B}, \omega, \alpha_t)$ and let $\gamma_s$ be a strongly continuous one-parameter symmetry group of $\alpha_t$ which is locally generated by $Q_\Lambda = \sum_{x \in \Lambda} q_x$. SSB amounts to find an equilibrium (KMS) or ground state $\omega$ which breaks the symmetry, that is, there exists a local observable $A \in \mathcal{A}_{\mathrm{L,sa}}$ such that for $s \neq 0$ holds: $\omega(\gamma_s(A)) \neq \omega(A)$ and $\alpha_t \gamma_s = \gamma_s \alpha_t$. This is equivalent to

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\omega(\gamma_s(A))\right|_{s=0} = \lim_\Lambda \omega([Q_\Lambda, A]) = c \neq 0$$

with $c$ a constant.

Now we turn this equation into a relation for fluctuations. Using space translation invariance of the state, one gets

$$\lim_\Lambda \frac{1}{|\Lambda|}\omega\left(\left[\sum_{x\in\Lambda}(q_x - \omega(q))\sum_{y\in\Lambda}(\tau_x A - \omega(A))\right]\right) = c$$

We now use another consequence of the Goldstone theorem, namely that SSB implies poor clustering properties for the order parameter $A$, that is, in the line of what is done in the last section, we assume that the lack of clustering is expressed by the existence of a positive index $\delta$ such that

$$\lim_\Lambda \omega\left(\frac{1}{|\Lambda|^{1+2\delta}}\left(\sum_{x\in\Lambda}(\tau_x A - \omega(A))\right)^2\right)$$

is nontrivial and finite. This means that the fluctuation $F^\delta(A)$ exists. Then we get

$$\lim_\Lambda \omega\left(\left[\frac{1}{|\Lambda|^{1/2-\delta}}\sum_{x\in\Lambda}(q_x - \omega(q)), \frac{1}{|\Lambda|^{1/2+\delta}}\sum_{y\in\Lambda}(\tau_x A - \omega(A))\right]\right) = c$$

Hence

$$\tilde{\omega}\big(\big[F^{-\delta}(q), F^\delta(A)\big]\big) = c$$

which for equilibrium states $\omega$, turns into the operator equation for fluctuations

$$[F^{-\delta}(q), F^\delta(A)] = c\mathbb{1}$$

In other words, one obtains a canonical pair $(F^{-\delta}(q), F^\delta(A))$ of normal coordinates of the collective Goldstone mode.

Note that the long-range correlation of the order-parameter operator (positive $\delta$) is exactly compensated by a squeezing, described by the negative index $-\delta$, for the fluctuation operator of the local generator of the broken symmetry. This result can also be expressed as typical for SSB, namely that the symmetry is not completely broken, but only partially. More detailed information about all this is found in Michoel and Verbeure (2001).

*See also:* Algebraic Approach to Quantum Field Theory; Large Deviations in Equilibrium Statistical Mechanics; Macroscopic Fluctuations and Thermodynamic Functionals; Quantum Phase Transitions; Quantum Spin Systems; Symmetry Breaking in Field Theory; Tomita–Takesaki Modular Theory.

## Further Reading


Brattelli O and Robinson DW (1979) *Operator Algebras and Quantum Statistical Mechanics*, vol. I. New York–Heidelberg–Berlin: Springer.

Brattelli O and Robinson D (2002) *Operator Algebras and Quantum Statistical Mechanics,* vol. II. New York–Heidelberg–Berlin: Springer.

Cushen CD and Hudson RL (1971) A quantum mechanical central limit theorem. *Journal of Applied Probability* 8: 454–469.

Fannes M and Quaegebeur J (1983) Central limits of product mappings between CAR-algebras. *Publications of the Research Institute for Mathematical Studies Kyoto* 19: 469–491.

Giri N and von Waldenfels W (1978) An algebraic version of the central limit theorem. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte gebiete* 42: 129–134.

Goderis D, Verbeure A, and Vets P (1989) Non-commutative central limits. *Probability and Related Fields* 82: 527–544.

Goderis D, Verbeure A, and Vets P (1990) Dynamics of fluctuations for quantum lattice systems. *Communications in Mathematical Physics* 128: 533–549.

Goderis D, Verbeure A, and Vets P (1991) About the exactness of the linear response theory. *Communications in Mathematical Physics* 136: 265–583.
Goderis D and Vets P (1989) Central limit theorem for mixing quantum systems and the CCR-algebra of fluctuations. *Communications in Mathematical Physics* 122: 249.
Goldshtein BG (1982) A central limit theorem of non-commutative probability theory. *Theory of Probability and its Applications* 27: 703.
Hudson RL (1973) A quantum mechanical central limit theorem for anti-commuting observables. *Journal of Applied Probability* 10: 502–509.
Ibragimov IA and Linnick YuV (1971) *Independent and stationary sequences of random variables.* Groningen: Wolters-Noordhoff.
Inönü E and Wigner EP (1953) On the contraction groups and their representations. *Proceedings of the National Academy of Sciences, USA* 39: 510–524.
Lanford DE and Ruelle D (1969) Observables at infinity and states with short-range correlations in statistical mechanics. *Journal of Mathematical Physics* 13: 194.
Manuceau J, Sirugue M, Testard D, and Verbeure A (1973) The smallest $C^*$-algebra for canonical commutation relations. *Communications in Mathematical Physics* 32: 231.
Matsui T (2003) On the algebra of fluctuations in quantum spin chains. *Annales Henri Poincaré* 4: 63–83.
Michoel T and Verbeure A (2001) Goldstone boson normal coordinates. *Communications in Mathematical Physics* 216: 461–490.
Momont B, Verbeure A, and Zagrebnov VA (1997) Algebraic structure of quantum fluctuations. *Journal of Statistical Physics* 89: 633–653.
Quaegebeur J (1984) A non-commutative central limit theorem for CCR-algebras. *Journal of Functional Analysis* 57: 1–20.
Sewell GL (1986) *Quantum theory of collective phenomena.* Oxford: Oxford University Press.
Verbeure A and Zagrebnov VA (1992) Phase transitions and algebra of fluctuation operators in an exactly soluble model of a quantum anharmonic crystal. *J. Stat. Phys.* 69: 329–359.

# Quantum Channels: Classical Capacity

**A S Holevo**, Steklov Mathematical Institute, Moscow, Russia



## The Definition

A numerical measure of the ability of a classical or quantum information processing system (for definiteness, one speaks of a *communication channel*) to transmit information expressible as a text message (called "classical information" as distinct from quantum information). It is equal to the least upper bound for rates of the asymptotically perfect transmission of classical information through the system, when the transmission time tends to infinity, and arbitrary pre- and post-processing (encoding and decoding) are allowed at the input and the output of the system. Typically, for rates exceeding the capacity, not only the asymptotically perfect transmission is impossible, but the error probability with arbitrary encoding–decoding scheme tends to 1, so that the capacity has a nature of a threshold parameter.

## From Classical to Quantum Information Theory

A central result of the classical information theory is the *Shannon coding theorem*, giving an explicit expression to the capacity in terms of the maximal mutual information between the input and the output of the channel. The issue of the information capacity of quantum communication channels arose soon after the publication of the pioneering papers by Shannon and goes back to the classical works of Gabor, Brillouin, and Gordon, asking for fundamental physical limits on the rate and quality of information transmission. This work laid a physical foundation and raised the question of consistent quantum treatment of the problem. Important steps in this direction were made in the early 1970s when a quantum probabilistic framework for this type of problem was created and the conjectured upper bound for the classical capacity of quantum channel was proved. A long journey to the quantum coding theorem culminated in 1996 with the proof of achievability of the upper bound (the Holevo–Schumacher–Westmoreland theorem; see Holevo (1998) for a detailed historical survey). Moreover, it was realized that quantum channel is characterized by the whole spectrum of capacities depending on the nature of the information resources and the specific protocols used for the transmission. To a great extent, this progress was stimulated by an interplay between the quantum communication theory and quantum information ideas related to more recent development in quantum computing. This new age of quantum information science is characterized by emphasis on the new possibilities (rather than restrictions) opened by the quantum nature of the information processing agent. On the other hand, the question of information capacity is important for the theory of quantum computer, particularly in connection with quantum error-correcting codes, communication and algorithmic complexity, and a number of other important issues.

## The Quantum Coding Theorem

In the simplest and most basic memoryless case, the information processing system is described by the sequence of block channels,

$$\Phi^{\otimes n} = \underbrace{\Phi \otimes \cdots \otimes \Phi}_{n}, \quad n = 1, \ldots$$

of $n$ parallel and independent uses of a channel $\Phi$, $n$ playing the role of transmission time (Holevo 1998). More generally, one can consider memory channels given by open dynamical systems with a kind of ergodic behavior and the limit where the transmission time goes to infinity (Kretschmann and Werner 2005).

Restricting to the memoryless case, encoding is given by a mapping of classical messages $x$ from a given codebook of size $N$ into states (density operators) $\rho_x^{(n)}$ in the input space $\mathcal{H}_1^{\otimes n}$ of the block channel $\Phi^{\otimes n}$, and decoding – by an observable $M^{(n)}$ in the output space $\mathcal{H}_2^{\otimes n}$, that is, a family $\{M_y^{(n)}\}$ of operators constituting a resolution of the identity in $\mathcal{H}_2^{\otimes n}$:

$$M_y^{(n)} \geq 0, \qquad \sum_y M_y^{(n)} = I$$

Here $y$ plays the role of outcomes of the whole decoding procedure involving both the quantum measurement at the output and the possible classical information post-processing. Then the diagram for the classical information transmission is

$$x \to \underbrace{\rho_i^{(n)}}_{\substack{\text{input}\\ \text{state}}} \to \underbrace{\Phi^{\otimes n}[\rho_x^{(n)}]}_{\substack{\text{output}\\ \text{state}}} \stackrel{M^{(n)}}{\to} y$$

The such-described encoding and decoding constitute a quantum block code of length $n$ and size $N$ for the memoryless channel. The conditional probability of obtaining an outcome $y$ provided the message $x$ was sent for a chosen block code is given by the statistical formula

$$p^{(n)}(y|x) = \mathrm{tr}\, \Phi^{\otimes n}[\rho_x^{(n)}] M_y^{(n)}$$

and the error probability for the code is just $\max_x (1 - p^{(n)}(x \mid x))$.

Denoting by $p_e(n, N)$ the infimum of the error probability over all codes of length $n$ and size $N$, the classical capacity $C(\Phi)$ of the memoryless channel is defined as the least upper bound of the rates $R$ for which $\lim_{n\to\infty} p_e(n, 2^{nR}) = 0$.

Let $\Phi$ be a quantum channel from the input to the output quantum systems, assumed to be finite dimensional. The *coding theorem* for the classical capacity says that

$$C(\Phi) = \lim_{n\to\infty} \frac{1}{n} C_\chi(\Phi^{\otimes n}) \qquad [1]$$

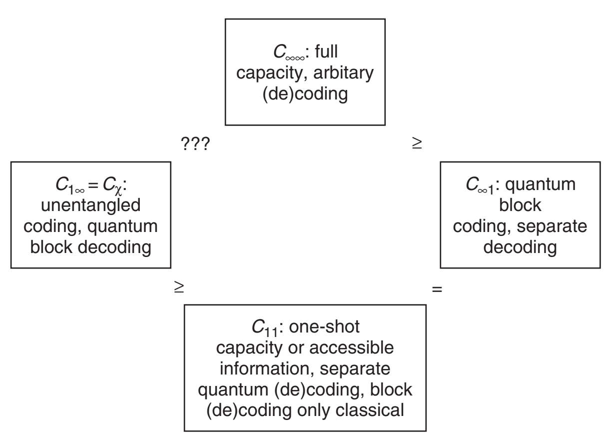
where

$$C_\chi(\Phi) = \max\left\{ H\left(\sum_x p_x \Phi[\rho_x]\right) - \sum_x p_x H(\Phi[\rho_x]) \right\} \qquad [2]$$

$H(\rho) = -\mathrm{tr}\, \rho \log_2 \rho$ is the binary von Neumann entropy, and the maximum is taken over all probability distributions $\{p_x\}$ and collections of density operators $\{\rho_x\}$ in $\mathcal{H}_1$.

## The Variety of Capacities

This basic definition and the formulas [1], [2] generalize the definition of the Shannon capacity and the coding theorem for classical memoryless channels. For quantum channel, there are several different capacities because one may consider sending different kinds (classical or quantum) of information, restrict the admissible coding and decoding operations, and/or allow the use of additional resources, such as shared entanglement, forward or backward communication, leading to really different quantities (Bennett *et al.* 2004). Few of these resources (such as feedback) also exist for classical channels but usually influence the capacity less dramatically (at least for memoryless channels). Restricting to the transmission of classical information with no additional resources, one can distinguish at least four capacities (Bennett and Shor 1998), according to whether, for each block length $n$, one is allowed to use arbitrary entangled quantum operations on the full block of input (resp. output) systems, or if, for each of the parallel channels, one has to use a separate quantum encoding (resp. decoding), and combine these only by classical pre- (resp. post-) processing:




The full capacity $C_{\infty\infty}$ is just the classical capacity $C(\Phi)$ given by [1]. That $C_{1\infty}$ coincides with the

quantity $C_\chi(\Phi)$ given by [2] is the essential content of the HSW theorem, from which [1] is obtained by additional blocking. Since $C_\chi$ is apparently superadditive, $C_\chi(\Phi_1 \otimes \Phi_2) \geq C_\chi(\Phi_1) + C_\chi(\Phi_2)$, one has $C_{\infty\infty} \geq C_\chi$. It is still not known whether the quantity $C_\chi(\Phi)$ is in fact additive for all channels, which would imply the equalities here. Additivity of $C_\chi(\Phi)$ would have the important physical consequence – it would mean that using entangled input states does not increase the classical capacity of quantum channel. While such a result would be very much welcome, giving a single-letter expression for the classical capacity, it would call for a physical explanation of asymmetry between the effects of entanglement in encoding and decoding procedures. Indeed, the inequality in the lower left is known to be strict sometimes (Holevo 1998), which means that entangled decodings *can* increase the classical capacity. There is even an intermediate capacity between $C_{11}$ and $C_{1\infty}$ obtained by restricting the quantum block decodings to adaptive ones (Shor 2002). The additivity of the quantity $C_\chi$ for *all* channels is one of the central open problems in quantum information theory; it was shown to be equivalent to several other important open problems, notably (super)additivity of the entanglement of formation and additivity of the minimal output entropy (Shor 2004).

For infinite-dimensional quantum processing systems, one needs to consider the input constraints such as the power constraint for bosonic Gaussian channels. The definition of the classical capacity and the capacity formula are then modified by introducing the constraint in a way similar to the classical theory (Holevo 1998, Holevo and Werner 2001). Another important extension concerns multiuser quantum information processing systems and their capacity regions (Devetak and Shor 2003).

*See also:* Capacities Enhanced by Entanglement; Capacity for Quantum Information; Channels in Quantum Information Theory; Entanglement Measures.

## Further Reading


Bennett CH and Shor PW (1998) Quantum information theory. *IEEE Transactions on Information Theory* 44: 2724–2742.

Bennett CH, Devetak I, Shor PW, and Smolin JA Inequality and separation between assisted capacities of quantum channel, e-print quant-ph/0406086.

Devetak I and Shor PW The capacity of quantum channel for simultaneous transmission of classical and quantum information, e-print quant-ph/0311131.

Holevo AS (1998) Quantum Coding Theorems. *Russian Math. Surveys* vol. 53. pp. 1295–1331, e-print quant-ph/9808023.

Holevo AS (2000) Coding theorems of quantum information theory. In: Grigoryan A, Fokas A, Kibble T, and Zegarlinski B (eds.) *Proc. XIII ICMP*, pp. 415–422. London: International Press of Boston.

Holevo AS and Werner RF (2001) Evaluating capacities of bosonic Gaussian channels. *Physical Review A* 63: 032312 (e-print quant-ph/9912067).

Kretschmann D and Werner RF Quantum channels with memory, e-print quant-ph/0502106.

Shor PW The adaptive classical capacity of a quantum channel, or information capacity of 3 symmetric pure states in three dimensions, e-print quant-ph/0206058.

Shor PW (2004) Equivalence of additivity questions in quantum information theory. *Communications in Mathematical Physics* 246: 4334–4340 (e-print quant-ph/0305035).


# Quantum Chromodynamics

**G Sterman**, Stony Brook University, Stony Brook, NY, USA



## Introduction

Quantum chromodynamics, or QCD, as it is normally called in high-energy physics, is the quantum field theory that describes the strong interactions. It is the SU(3) gauge theory of the current standard model for elementary particles and forces, $SU(3) \times SU(2)_L \times U(1)$, which encompasses the strong, electromagnetic, and weak interactions. The symmetry group of QCD, with its eight conserved charges, is referred to as color SU(3). As is characteristic of quantum field theories, each field may be described in terms of quantum waves or particles.

Because it is a gauge field theory, the fields that carry the forces of QCD transform as vectors under the Lorentz group. Corresponding to these vector fields are the particles called "gluons," which carry an intrinsic angular momentum, or spin, of 1 in units of $\hbar$. The strong interactions are understood as the cumulative effects of gluons, interacting among themselves and with the quarks, the spin-1/2 particles of the Dirac quark fields.

There are six quark fields of varying masses in QCD. Of these, three are called "light" quarks, in a sense to be defined below, and three "heavy." The light quarks are the up ($u$), down ($d$), and strange ($s$), while the heavy quarks are the charm ($c$), bottom ($b$),

and top ($t$). Their well-known electric charges are $e_f = 2e/3(u, c, t)$ and $e_f = -e/3(d, s, b)$, with $e$ the positron charge. The gluons interact with each quark field in an identical fashion, and the relatively light masses of three of the quarks provide the theory with a number of approximate global symmetries that profoundly influence the manner in which QCD manifests itself in the standard model.

These quark and gluon fields and their corresponding particles are enumerated with complete confidence by the community of high-energy physicists. Yet, none of these particles has ever been observed in isolation, as one might observe a photon or an electron. Rather, all known strongly interacting particles are colorless; most are "mesons," combinations with the quantum numbers of a quark $q$ and a antiquark $\bar{q}'$, or "baryons" with the quantum numbers of (possibly distinct) combinations of three quarks $qq'q''$. This feature of QCD, that its underlying fields never appear as asymptotic states, is called "confinement." The very existence of confinement required new ways of thinking about field theory, and only with these was the discovery and development of QCD possible.

## The Background of QCD

The strong interactions have been recognized as a separate force of nature since the discovery of the neutron as a constituent of atomic nuclei, along with the proton. Neutrons and protons (collectively, nucleons) possess a force, attractive at intermediate distances and so strong that it overcomes the electric repulsion of the protons, each with charge $e$. A sense of the relative strengths of the electromagnetic and strong interactions may be inferred from the typical distance between mutually repulsive electrons in an atom, $\sim 10^{-8}$ cm, and the typical distance between protons in a nucleus, of order $10^{-13}$ cm.

The history that led up to the discovery of QCD is a fascinating one, beginning with Yukawa's 1935 theory of pion exchange as the source of the forces that bind nuclei, still a useful tool for low-energy scattering. Other turning points include the creation of nonabelian gauge theories by Yang and Mills in 1954, the discovery of the quantum number known as strangeness, the consequent development of the quark model, and then the proposal of color as a global symmetry. The role of pointlike constituents in hadrons was foreshadowed by the identification of electromagnetic and weak currents and the analysis of their quantum-mechanical algebras. Finally, the observation of "scaling" in deep-inelastic scattering, which we will describe below, made QCD, with color as a local symmetry, the unique explanation of the strong interactions, through its property of asymptotic freedom.

## The Lagrangian and Its Symmetries

The QCD Lagrangian may be written as

$$\mathcal{L} = \sum_{f=1}^{n_f} \bar{q}_f \left( \mathrm{i}\, \not{D}[A] - m_f \right) q_f - \frac{1}{2} \operatorname{tr}\left[ F_{\mu\nu}^2(A) \right] - \frac{\lambda}{2} (B_a(A))^2 + \bar{c}_b \left[ \frac{\delta B_b(A)}{\delta \alpha_a} \right] c_a \qquad [1]$$

with $\not{D}[A] = \gamma \cdot \partial + \mathrm{i} g_s \gamma \cdot A$ the covariant derivative in QCD. The $\gamma^\mu$ are the Dirac matrices, satisfying the anticommutation relations, $[\gamma^\mu, \gamma^\nu]_+ = 2g^{\mu\nu}$. The SU(3) gluon fields are $A^\mu = \sum_{a=1}^{8} A_a^\mu T_a$, where $T_a$ are the generators of SU(3) in the fundamental representation. The field strengths $F_{\mu\nu}[A] = \partial_\mu A_\nu - \partial_\nu A_\mu + \mathrm{i} g_s [A_\mu, A_\nu]$ specify the three- and four-point gluon couplings of nonabelian gauge theory. In QCD, there are $n_f = 6$ flavors of quark fields, $q_f$, with conjugate $\bar{q}_f = q_f^\dagger \gamma^0$.

The first two terms in the expression [1] make up the classical Lagrangian, followed by the gauge-fixing term, specified by a (usually, but not necessarily linear) function $B_a(A)$, and the ghost Lagrangian. The ghost (anti-ghost) fields $c_a(\bar{c}_a)$ carry the same adjoint index as the gauge fields.

The classical QCD Lagrangian before gauge fixing is invariant under the local gauge transformations

$$\begin{aligned} A'_\mu(x) &= \frac{\mathrm{i}}{g_s} \partial_\mu \Omega(x) \Omega^{-1}(x) + \Omega(x) A'_\mu(x) \Omega^{-1}(x) \\ &= A_\mu(x) - \partial_\mu \delta\alpha(x) \\ &\quad + \mathrm{i} g_s \left[ \delta\alpha(x), A_\mu(x) \right] + \cdots \\ \psi'_i(x) &= \Omega(x)_{ij} \psi_j(x) = \psi_i(x) \\ &\quad + \mathrm{i} g_s \delta\alpha(x)_{ij} \psi_j(x) + \cdots \\ \delta\alpha(x) &= \sum_{a=1}^{8} \delta\alpha_a(x) T_a \end{aligned} \qquad [2]$$

The full QCD action including gauge-fixing and ghost terms is also invariant under the Bechi, Rouet, Stora, Tyutin (BRST) transformations with $\delta\xi$ an anticommuting variable.

$$\begin{aligned} \delta A_{\mu,a} &= \left( \delta_{ab} \partial_\mu + g A_{\mu c} f_{abc} \right) c_b \delta\xi \\ \delta c_a &= -\tfrac{1}{2} g C_{abc}\, c_b\, c_c \delta\xi, \quad \delta\bar{c} = \lambda B_a \delta\xi \\ \delta\psi_i &= \mathrm{i} g [T_b]_{ij}\, c_b \psi_j \end{aligned} \qquad [3]$$

with $f_{abc}$ the SU(3) structure constants. The Jacobian of these transformations is unity.

In addition, neglecting masses of the light quarks, $u$, $d$, and $s$, the QCD Lagranian has a class of global flavor and chiral symmetries, the latter connecting left- and right-handed components of the quark fields, $\psi_{\mathrm{L,R}} \equiv (1/2)(1 \mp \gamma_5)\psi$,

$$\psi'(x) = \mathrm{e}^{\mathrm{i}\alpha\gamma_5^P}\, \psi(x), \quad P = 0, 1 \qquad [4]$$

Here, power $P = 0$ describes phase, and $P = 1$ chiral, transformations. Both transformations can be extended to transformations among the light flavors, by letting $\psi$ become a vector, and $\alpha$ an element in the Lie algebra of SU($M$), with $M = 2$ if we take only the $u$ and $d$ quarks, and $M = 3$ if we include the somewhat heavier strange quark. These symmetries, not to be confused with the local symmetries of the standard model, are strong isospin and its extension to the "eightfold way," which evolved into the (3-)quark model of Gell–Mann and Zweig. The many successes of these formalisms are automatically incorporated into QCD.

## Green Functions, Phases, and Gauge Invariance

In large part, the business of quantum field theory is to calculate Green functions,

$$G_n(x_1 \dots x_n) = \langle 0|T(\Phi_1(x_1) \dots \Phi_i(x_i) \dots \Phi_n(x_n))|0\rangle \qquad [5]$$

where $T$ denotes time ordering. The $\Phi_i(x)$ are elementary fields, such as $A$ or $q_f$, or composite fields, such as currents like $J^\mu = \bar{q}_f \gamma^\mu q_f$. Such a Green function generates amplitudes for the scattering of particles of definite momenta and spin, when in the limit of large times the $x_i$-dependence of the Green function is that of a plane wave. For example, we may have in the limit $x_i^0 \rightarrow \infty$,

$$G_n(x_1 \dots x_n) \rightarrow \phi_i(p, \lambda)\, \mathrm{e}^{\mathrm{i}p \cdot x_i} \langle (p, \lambda)|T(\Phi_1(x_1) \dots \Phi_{i-1}(x_{i-1})\Phi_{i+1}(x_{i+1}) \dots \Phi_n(x_n))|0\rangle \qquad [6]$$

where $\phi_i(p, \lambda)$ is a solution to the free-field equation for field $\Phi_i$, characterized by momentum $p$ and spin $\lambda$. (An inegral over possible momenta $p$ is understood.) When this happens for field $i$, the vacuum state is replaced by $|(p, \lambda)\rangle$, a particle state with precisely this momentum and spin; when it occurs for all fields, we derive a scattering ($S$)-matrix amplitude. In essence, the statement of confinement is that Green functions with fields $q_f(x)$ never behave as plane waves at large times in the past or future. Only Green functions of color singlet composite fields, invariant under gauge transformations, are associated with plane wave behavior at large times.

Green functions remain invariant under the BRST transformations [3], and this invariance implies a set of Ward identities

$$\frac{\delta}{\delta\xi(z)} \sum_{i=1}^{n} \langle 0|T(\Phi_1(x_1) \dots \delta_{\mathrm{BRS}}\Phi_i(x_i) \dots \Phi_n(x_n))|0\rangle = 0 \qquad [7]$$

The variation of the anti-ghost as in [3] is equivalent to an infinitesimal change in the gauge-fixing term; variations in the remaining fields all cancel single-particle plane wave behavior in the corresponding Green functions. These identities then ensure the gauge invariance of the perturbative $S$-matrix, a result that turns out to be useful despite confinement.

To go beyond a purely perturbative description of QCD, it is useful to introduce a set of nonlocal operators that are variously called nonabelian phases, ordered exponentials, and Wilson lines,

$$U_C(z, y) = P \exp\left[-\mathrm{i}g_\mathrm{s} \int_y^z \mathrm{d}x^\mu A_\mu(x)\right] \qquad [8]$$

where $C$ is some self-avoiding curve between $y$ and $z$. The $U$'s transform at each end linearly in nonabelian gauge transformations $\Omega(x)$ at that point,

$$U'_C(z, y) = \Omega(z) U_C(z, y) \Omega^{-1}(y) \qquad [9]$$

Especially interesting are closed curves $C$, for which $z = y$. The phases about such closed loops are, like their abelian counterparts, sensitive to the magnetic flux that they enclose, even when the field strengths vanish on the curve.

## QCD at the Shortest and Longest Distances

Much of the fascination of QCD is its extraordinary variation of behavior at differing distance scales. Its discovery is linked to asymptotic freedom, which characterizes the theory at the shortest scales. Asymptotic freedom also suggests (and in part provides) a bridge to longer distances.

Most analyses in QCD begin with a path-integral formulation in terms of the elementary fields $\Phi_a = q_f \dots$,

$$G_n(x_i, (z_j, y_j)) = \int \left[\prod_{a=q,\bar{q},G,c,\bar{c}} \mathcal{D}\Phi_a\right] \prod_i \Phi_i(x_i) \times \prod_j U_{C_j}(z_j, y_j)\, \mathrm{e}^{\mathrm{i}S_{\mathrm{QCD}}} \qquad [10]$$

with $S_{\mathrm{QCD}}$ the action. Perturbation theory keeps only the kinetic Lagrangian, quadratic in fields, in the exponent, and expands the potential terms in the coupling. This procedure produces Feynman diagrams, with vertices corresponding to the cubic and quartic terms in the QCD Lagrangian [1].

Most nonperturbative analyses of QCD require studying the theory on a Euclean, rather than Minkowski space, related by an analytic continuation in the times $x^0, y^0, z^0$ in $G_n$ from real to imaginary values. In Euclidean space, we find, for example,

classical solutions to the equations of motion, known as instantons, that provide nonperturbative contributions to the path integral. Perhaps the most flexible nonperturbative approach approximates the action and the measure at a lattice of points in four-dimensional space. For this purpose, integrals over the gauge fields are replaced by averages over "gauge links," of the form of eqn [8] between neighboring points.

Perturbation theory is most useful for processes that occur over short timescales and at high relative energies. Lattice QCD, on the other hand, can simulate processes that take much longer times, but is less useful when large momentum transfers are involved. The gap between the two methods remains quite wide, but between the two they have covered enormous ground, enough to more than confirm QCD as the theory of strong interactions.

### Asymptotic Freedom

QCD is a renormalizable field theory, which implies that the coupling constant $g$ must be defined by its value at a "renormalization scale," and is denoted $g(\mu)$. Usually, the magnitude of $\alpha_s(\mu) \equiv g^2/4\pi$, is quoted at $\mu = m_Z$, where it is $\sim 0.12$. In effect, $g(\mu)$ controls the amplitude that connects any state to another state with one more or one fewer gluon, including quantum corrections that occur over timescales from zero up to $\hbar/\mu$ (if we measure $\mu$ in units of energy). The QCD Ward identities mentioned above ensure that the coupling is the same for both quarks and gluons, and indeed remains the same in all terms in the Lagrangian, ensuring that the symmetries of QCD are not destroyed by renormalization.

Quantum corrections to gluon emission are not generally computable directly in renormalizable theories, but their dependence on $\mu$ is computable, and is a power series in $\alpha_s(\mu)$ itself,

$$\mu^2 \frac{d\alpha_s(\mu)}{d\mu^2} = -b_0 \frac{\alpha_s^2(\mu)}{4\pi} - b_1 \frac{\alpha_s^3(\mu)}{(4\pi)^2} + \cdots \equiv \beta(\alpha_s) \quad [11]$$

where $b_0 = 11 - 2n_f/3$ and $b_1 = 2(31 - 19n_f/3)$. The celebrated minus signs on the right-hand side are associated with both the spin and self-interactions of the gluons.

The solution to this equation provides an expression for $\alpha_s$ at any scale $\mu_1$ in terms of its value at any other scale $\mu_0$. Keeping only the lowest-order, $b_0$, term, we have

$$\alpha_s(\mu_1) = \frac{\alpha_s(\mu_0)}{1 + (b_0/4\pi)\ln\left(\mu_1^2/\mu_0^2\right)} = \frac{4\pi}{b_0 \ln\left(\mu_1^2/\Lambda_{QCD}^2\right)} \quad [12]$$

where in the second form, we have introduced $\Lambda_{QCD}$, the scale parameter of the theory, which embodies the condition that we get the same coupling at scale $\mu_1$ no matter which scale $\mu_0$ we start from. Asymptotic freedom consists of the observation that at larger renormalization masses $\mu$, or correspondingly shorter timescales, the coupling weakens, and indeed vanishes in the limit $\mu \to \infty$. The other side of the coin is that over longer times or lower momenta, the coupling grows. Eventually, near the pole at $\mu_1 = \Lambda_{QCD}$, the lowest-order approximation to the running fails, and the theory becomes essentially nonperturbative. Thus, the discovery of asymptotic freedom suggested, although it certainly does not prove, that QCD is capable of producing very strong forces, and confinement at long distances. Current estimates of $\Lambda_{QCD}$ are $\sim 200$ MeV.

### Spontaneous Breaking of Chiral Symmetry

The number of quarks and their masses is an external input to QCD. In the standard model masses are provided by the Higgs mechanism, but in QCD they are simply parameters. Because the standard model has chosen several of the quarks to be especially light, QCD incorporates the chiral symmetries implied by eqn [4] (with $P = 1$). In the limit of zero quark masses, these symmetries becomes exact, respected to all orders of perturbation theory, that is, for any finite number of gluons emitted or absorbed.

At distances on the order to $1/\Lambda_{QCD}$, however, QCD cannot respect chiral symmetry, which would require each state to have a degenerate partner with the opposite parity, something not seen in nature. Rather, QCD produces, nonperturbatively, nonzero values for matrix elements that mix right- and left-handed fields, such as $\langle 0|\bar{u}_L u_R|0\rangle$, with $u$ the up-quark field. Pions are the Goldstone bosons of this symmetry, and may be thought of as ripples in the chiral condensate, rotating it locally as they pass along. The observation that these Goldstone bosons are not exactly massless is due to the "current" masses of the quarks, their values in $\mathcal{L}_{QCD}$. The (chiral perturbation theory) expansion in these light-quark masses also enables us to estimate them quantitatively: $1.5 \leq m_u \leq 4$ MeV, $4 \leq m_d \leq 8$ MeV, and $80 \leq m_s \leq 155$ MeV. These are the light quarks, with masses smaller than $\Lambda_{QCD}$. (Like $\alpha_s$, the masses are renormalized; these are quoted from Eidelman (2004) with $\mu = 2$ GeV.) For comparison, the heavy quarks have masses $m_c \sim 1$–$1.5$ GeV, $m_b \sim 4$–$4.5$ GeV, and $m_t \sim 180$ GeV (the giant among the known elementary particles).

Although the mechanism of the chiral condensate (and in general other nonperturbative aspects of

QCD) has not yet been demonstrated from first principles, a very satisfactory description of the origin of the condensate, and indeed of much hadronic structure, has been given in terms of the attractive forces between quarks provided by instantons. The actions of instanton solutions provide a dependence $\exp[-8\pi^2/g_s^2]$ in Euclidean path integrals, and so are characteristically nonperturbative.

### Mechanisms of Confinement

As described above, confinement is the absence of asymptotic states that transform nontrivially under color transformations. The full spectrum of QCD, however, is a complex thing to study, and so the problem has been approached somewhat indirectly. A difficulty is the same light-quark masses associated with approximate chiral symmetry. Because the masses of the light quarks are far below the scale $\Lambda_{QCD}$ at which the perturbative coupling blows up, light quarks are created freely from the vacuum and the process of "hadronization," by which quarks and gluons form mesons and baryons, is both nonperturbative and relativistic. It is therefore difficult to approach in both perturbation theory and lattice simulations.

Tests and studies of confinement are thus normally formulated in truncations of QCD, typically with no light quarks. The question is then reformulated in a way that is somewhat more tractable, without relativistic light quarks popping in and out of the vacuum all the time. In the limit that its mass becomes infinite compared to the natural scale of fluctuations in the QCD vacuum, the propagator of a quark becomes identical to a phase operator, [8], with a path $C$ corresponding to a constant velocity. This observation suggests a number of tests for confinement that can be implemented in the lattice theory. The most intuitive is the vacuum expectation value of a "Wilson loop," consisting of a rectangular path, with sides along the time direction, corresponding to a heavy quark and antiquark at rest a distance $R$ apart, and closed at some starting and ending times with straight lines. The vacuum expectation value of the loop then turns out to be the exponential of the potential energy between the quark pair, multiplied by the elapsed time,

$$\left\langle 0 \middle| P \exp\left[-\mathrm{i}g_s \oint_C A_\mu(x)\,\mathrm{d}x^\mu\right] \middle| 0 \right\rangle = \exp(-V(R)T/\hbar) \qquad [13]$$

When $V(R) \propto R$ ("area law" behavior), there is a linearly rising, confining potential. This behavior, not yet proven analytically yet well confirmed on the lattice, has an appealing interpretation as the energy of a "string," connecting the quark and antiquark, whose energy is proportional to its length.

Motivation for such a string picture was also found from the hadron spectrum itself, before any of the heavy quarks were known, and even before the discovery of QCD, from the observation that many mesonic ($\bar{q}q'$) states lie along "Regge trajectories," which consist of sets of states of spin $J$ and mass $m_J^2$ that obey a relation

$$J = \alpha' m_J^2 \qquad [14]$$

for some constant $\alpha'$. Such a relation can be modeled by two light particles ("quarks") revolving around each other at some constant (for simplicity, fixed nonrelativistic) velocity $v_0$ and distance $2R$, connected by a "string" whose energy per unit length is a constant $\rho$.

Suppose the center of the string is stationary, so the overall system is at rest. Then neglecting the masses, the total energy of the system is $M = 2R\rho$. Meanwhile, the momentum density per unit length at distance $r$ from the center is $v(r) = (r/R)v_0$, and the total angular momentum of the system is

$$J = 2\rho v_0 \int_0^R \mathrm{d}r\, r^2 = \frac{2\rho v_0}{3} R^2 = \frac{v_0}{6\rho} M^2 \qquad [15]$$

and for such a system, [14] is indeed satisfied. Quantized values of angular momentum $J$ give quantized masses $m_J$, and we might take this as a sort of "Bohr model" for a meson. Indeed, string theory has its origin in related consideration in the strong interactions.

Lattice data are unequivocal on the linearly rising potential, but it requires further analysis to take a lattice result and determine what field configurations, stringlike or not, gave that result. Probably the most widely accepted explanation is in terms of an analogy to the Meissner effect in superconductivity, in which type II superconductors isolate magnetic flux in quantized tubes, the result of the formation of a condensate of Cooper pairs of electrons. If the strings of QCD are to be made of the gauge field, they must be electric ($F^{\mu 0}$) in nature to couple to quarks, so the analogy postulates a "dual" Meissner effect, in which electric flux is isolated as the result of a condensate of objects with magnetic charge (producing nonzero $F^{ij}$). Although no proof of this mechanism has been provided yet, the role of magnetic fluctuations in confinement has been widely investigated in lattice simulations, with encouraging results. Of special interest are magnetic field configurations, monopoles or vortices, in the $Z_3$ center of SU(3), $\exp[\mathrm{i}\pi k/3]I_{3\times 3}$, $k = 0, 1, 2$. Such configurations, even when localized, influence closed gauge loops [13] through the nonabelian Aharonov–Bohm effect. Eventually, of course, the role of light quarks must be crucial for any complete

description of confinement in the real world, as emphasized by Gribov.

Another related choice of closed loop is the "Polyakov loop," implemented at finite temperature, for which the path integral is taken over periodic field configurations with period $1/T$, where $T$ is the temperature. In this case, the curve $C$ extends from times $t=0$ to $t=1/T$ at a fixed point in space. In this formulation it is possible to observe a phase transition from a confined phase, where the expectation is zero, to a deconfined phase, where it is nonzero. This phase transition is currently under intense experimental study in nuclear collisions.

## Using Asymptotic Freedom: Perturbative QCD

It is not entirely obvious how to use asymptotic freedom in a theory that should (must) have confinement. Such applications of asymptotic freedom go by the term perturbative QCD, which has many applications, not the least as a window to extensions of the standard model.

### Lepton Annihilation and Infrared Safety

The electromagnetic current, $J_\mu = \sum_f e_f \bar{q}_f \gamma_\mu q_f$, is a gauge-invariant operator, and its correlation functions are not limited by confinement. Perhaps, the simplest application of asymptotic freedom, yet of great physical relevance, is the scalar two-point function,

$$\pi(Q) = \frac{-\mathrm{i}}{3} \int \mathrm{d}^4 x\, \mathrm{e}^{-\mathrm{i}Q\cdot x} \langle 0 | \mathrm{T}(J^\mu(0) J_\mu(x)) | 0 \rangle \qquad [16]$$

The imaginary part of this function is related to the total cross section for the annihilation process $\mathrm{e}^+\mathrm{e}^- \to$ hadrons in the approximation that only one photon takes part in the reaction. The specific relation is $\sigma_{\mathrm{QCD}} = (e^4/Q^2)\, \mathrm{Im}\, \pi(Q^2)$, which follows from the optical theorem, illustrated in Figure 1. The perturbative expansion of the function $\pi(Q)$ depends, in general, on the mass scales $Q$ and the quark masses $m_f$ as well as on the strong coupling $\alpha_s(\mu)$ and on the renormalization scale $\mu$. We may also worry about the influence of other, truly nonperturbative scales, proportional to powers of $\Lambda_{\mathrm{QCD}}$. At large values of $Q^2$, however, the situation simplifies greatly, and dependence on all scales below $Q$ is suppressed by powers of $Q$. This may be expressed in terms of the operator product expansion,

$$\begin{aligned} &\langle 0 | T(J^\mu(0) J_\mu(x)) | 0 \rangle \\ &\quad = \sum_{O_I} (x^2)^{-3+d_I/2} C_I(x^2\mu^2, \alpha_s(\mu)) \\ &\quad \times \langle 0 | O_I(0) | 0 \rangle \end{aligned} \qquad [17]$$

where $d_I$ is the mass dimension of operator $O_I$, and where the dimensionless coefficient functions $C_I$ incorporate quantum corrections. The sum over operators begins with the identity ($d_I = 0$), whose coefficient function is identified with the sum of quantum corrections in the approximation of zero masses. The sum continues with quark mass corrections, which are suppressed by powers of at least $m_f^2/Q^2$, for those flavors with masses below $Q$. Any QCD quantity that has this property, remaining finite in perturbation theory when all particle masses are set to zero, is said to be "infrared safe."

The effects of quarks whose masses are above $Q$ are included indirectly, through the couplings and masses observed at the lower scales. In summary, the leading power behavior of $\pi(Q)$, and hence of the cross section, is a function of $Q$, $\mu$, and $\alpha_s(\mu)$ only. Higher-order operators whose vacuum matrix elements receive nonperturbative corrections include the "gluon condensate," identified as the product $\alpha_s(\mu) G_{\alpha\beta} G^{\alpha\beta} \propto \Lambda_{\mathrm{QCD}}^4$.

Once we have concluded that $Q$ is the only physical scale in $\pi$, we may expect that the right choice of the renormalization scale is $\mu = Q$. Any observable quantity is independent of the choice of renormalization scale, $\mu$, and neglecting quark masses, the chain rule gives

$$\mu \frac{\mathrm{d}\sigma(Q/\mu, \alpha_s(\mu))}{\mathrm{d}\mu} = \mu \frac{\partial \sigma}{\partial \mu} + 2\beta(\alpha_s) \frac{\partial \sigma}{\partial \alpha_s} = 0 \qquad [18]$$

which shows that we can determine the beta function directly from the perturbative expansion of the cross section. Defining $a \equiv \alpha_s(\mu)/\pi$, such a perturbative calculation gives

$$\begin{aligned} \mathrm{Im}\pi(Q^2) = \frac{3}{4\pi} \sum_f e_f^2 \Bigg(1 + a + a^2 \Bigg(1.986 \\ - 0.115 n_f - (b_0/4\pi) \ln \frac{Q^2}{\mu^2}\Bigg)\Bigg) \end{aligned} \qquad [19]$$

with $b_0$ as above. Now, choosing $\mu = Q$, we see that asymptotic freedom implies that when $Q$ is large, the total cross section is given by the lowest order,

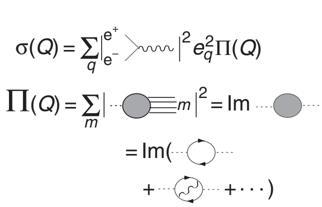


**Figure 1** First line: schematic relation of lowest order $\mathrm{e}^+\mathrm{e}^-$ annihilation to sum over quarks $q$, each with electric charge $e_q$. Second line: perturbative unitarity for the current correlation function $\pi(Q)$.

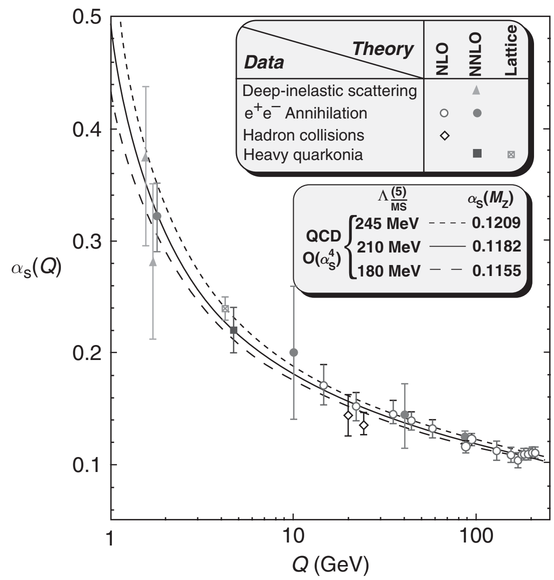



**Figure 2** Experimental variation of the strong coupling with scales. Reproduced from Bethke S (2004) Alpha(s) at Zinnowitz. *Nuclear Physics Proceedings Supplements* 135: 345–352, with permission from Elsevier.

plus small and calculable QCD corrections, a result that is borne out in experiment. Comparing experiment to an expression like [19], one can measure the value of $\alpha_s(Q)$, and hence, with eqn [12], $\alpha_s(\mu)$ for any $\mu \gg \Lambda_{\mathrm{QCD}}$. Figure 2 shows a recent compilation of values of $\alpha_s$ from this kind of analysis in different experiments at different scales, clearly demonstrating asymptotic freedom.

## Factorization, Scaling, and Parton distributions

One step beyond vacuum matrix elements of currents are their expectation values in single-particle states, and here we make contact with the discovery of QCD, through scaling. Such expectations are relevant to the class of experiments known as deep-inelastic scattering, in which a high-energy electron exchanges a photon with a nucleon target. All QCD information is contained in the tensor matrix element

$$W_N^{\mu\nu}(p,q) \equiv \frac{1}{8\pi}\sum_\sigma \int d^4x\, e^{-iq\cdot x}\langle p,\sigma|J^\mu(0)J^\nu(x)|p,\sigma\rangle \quad [20]$$

with $q$ the momentum transfer carried by the photon, and $p,\sigma$ the momentum and spin of the target nucleon, $N$. This matrix element is not infrared safe, since it depends in principle on the entire history of the nucleon state. Thus, it is not accessible to direct perturbative calculation.

Nevertheless, when the scattering involves a large momentum transfer compared to $\Lambda_{\mathrm{QCD}}$, we may expect a quantum-mechanical incoherence between the scattering reaction, which occurs (by the uncertainty principle) at short distances, and the forces that stabilize the nucleon. After all, we have seen that the latter, strong forces, should be associated with long distances. Such a separation of dynamics, called factorization, can be implemented in perturbation theory, and is assumed to be a property of full QCD. Factorization is illustrated schematically in Figure 3. Of course, short and long distances are relative concepts, and the separation requires the introduction of a so-called factorization scale, $\mu_F$, not dissimilar to the renormalization scale described above. For many purposes, it is convenient to choose the two equal, although this is not required.

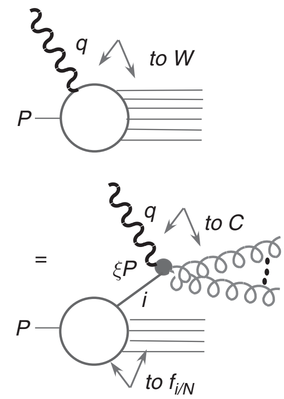



**Figure 3** Schematic depiction of factorization in deep-inelastic scattering.

The expression of factorization for deep-inelastic scattering is

$$W_N^{\mu\nu}(p,q) = \sum_{i=q_f,\bar{q}_f,G}\int_x^1 d\xi\, C_i^{\mu\nu}(\xi p, q, \mu_F, \alpha_s(\mu_F)) \times f_{i/N}(\xi,\mu_F) \quad [21]$$

where the functions $C_i^{\mu\nu}$ (the coefficient functions) can be computed as an expansion in $\alpha_s(\mu_F)$, and describe the scattering of the "partons," quarks, and gluons, of which the target is made. The variable $\xi$ ranges from unity down to $x \equiv -q^2/2p\cdot q > 0$, and has the interpretation of the fractional momentum of the proton carried by parton $i$. (Here $-q^2 = Q^2$ is positive.) The parton distributions $f_{i/N}$ can be defined in terms of matrix elements in the nucleon, in which the currents are replaced by quark (or antiquark or gluon) fields, as

$$f_{q/N}(x,\mu) = \frac{1}{4\pi}\int_{-\infty}^{\infty} d\lambda\, e^{-i\lambda x p^+} \times \langle p,\sigma|\bar{q}(\lambda n)U_n(n\lambda,0)n\cdot\gamma q(0)|p,\sigma\rangle \quad [22]$$

$n^{\mu}$ is a light-like vector, and $U_n$ a phase operator whose path $C$ is in the $n$-direction. The dependence of the parton distribution on the factorization scale is through the renormalization of the composite operator consisting of the quark fields, separated along the light cone, and the nonabelian phase operator $U_n(n\lambda, 0)$, which renders the matrix element gauge invariant by eqn [9]. By combining the calculations of the $C$'s and data for $W_N^{\mu\nu}$, we can infer the parton distributions, $f_{i/N}$. Important factorizations of a similar sort also apply to some exclusive processes, including amplitudes for elastic pion or nucelon scattering at large momentum transfer.

Equation [21] has a number of extraordinary consequences. First, because the coefficient function is an expansion in $\alpha_s$, it is natural to choose $\mu_F^2 \sim Q^2 \sim p \cdot q$ (when $x$ is of order unity). When $Q$ is large, we may approximate $C_i^{\mu\nu}$ by its lowest order, which is first order in the electromagnetic coupling of quarks to photons, and zeroth order in $\alpha_s$. In this approximation, dependence on $Q$ is entirely in the parton distributions. But such dependence is of necessity weak (again for $x$ not so small as to produce another scale), because the $\mu_F$ dependence of $f_{i/N}(\xi, \mu_F)$ must be compensated by the $\mu_F$ dependence of $C_i^{\mu\nu}$, which is order $\alpha_s$. This means that the overall $Q$ dependence of the tensor $W_N^{\mu\nu}$ is weak for $Q$ large when $x$ is moderate. This is the scaling phenomenon that played such an important role in the discovery of QCD.

### Evolution: Beyond Scaling

Another consequence of the factorization [21], or equivalently of the operator definition [22], is that the $\mu_F$-dependence of the coefficient functions and the parton distributions are linked. As in the lepton annihilation cross section, this may be thought of as due to the independence of the physically observable tensor $W_N^{\mu\nu}$ from the choice of factorization and renormalization scales. This implies that the $\mu_F$-dependence of $f_{i/N}$ may be calculated perturbatively since it must cancel the corresponding dependence in $C_i$. The resulting relation is coventionally expressed in terms of the "evolution equations,"

$$\mu \frac{\mathrm{d}f_{a/N}(x,\mu)}{\mathrm{d}\mu} = \sum_c \int_x^1 \mathrm{d}\xi \; P_{ac}(x/\xi, \alpha_s(\mu)) f_{c/N}(\xi, \mu) \qquad [23]$$

where $P_{ac}(\xi)$ are calculable as power series, now known up to $\alpha_s^3$. This relation expands the applicability of QCD from scales where parton distributions can be inferred directly from experiment, to arbitrarily high scales, reachable in accelerators under construction or in the imagination, or even on the cosmic level.

At very high energy, however, the effective values of the variable $x$ can become very small and introduce new scales, so that eventually the evolution of eqn [23] fails. The study of nuclear collisions may provide a new high-density regime for QCD, which blurs the distinction between perturbative and nonperturbative dynamics.

### Inclusive Production

Once we have evolution at our disposal, we can take yet another step, and replace electroweak currents with any operator from any extension of QCD, in the standard model or beyond, that couples quarks and gluons to the particles of as-yet unseen fields. Factorization can be extended to these situations as well, providing predictions for the production of new particles, $F$ of mass $M$, in the form of factorized inclusive cross sections,

$$\begin{aligned} &\sigma_{AB\to F(M)}(M, p_A, p_B) \\ &\quad = \sum_{i,j=q_f\bar{q}_f,G} \int \mathrm{d}\xi_a \, \mathrm{d}\xi_b f_{i/A}(\xi_a,\mu) f_{j/B}(\xi_b,\mu) \\ &\quad\quad \times H_{ij\to F(M)}(x_a p_A, x_b p_B, M, \mu, \alpha_s(\mu)) \end{aligned} \qquad [24]$$

where the functions $H_{ij\to F}$ may be calculated perturbatively, while the $f_{i/A}$ and $f_{j/B}$ parton distributions are known from a combination of lower-energy observation and evolution. In this context, they are said to be "universal," in that they are the same functions in hadron–hadron collisions as in the electron–hadron collisions of deep-inelastic scattering. In general, the calculation of hard-scattering functions $H_{ij}$ is quite nontrivial beyond lowest order in $\alpha_s$. The exploration of methods to compute higher orders, currently as far as $\alpha_s^2$, has required extraordinary insight into the properties of multidimensional integrals.

The factorization method helped predict the observation of the $W$ and $Z$ bosons of electroweak theory, and the discovery of the top quark. The extension of factorization from deep-inelastic scattering to hadron production is nontrivial; indeed, it only holds in the limit that the velocities, $\beta_i$, of the colliding particles approach the speed of light in the center-of-momentum frame of the produced particle. Corrections to the relation [24] are then at the level of powers of $\beta_i - 1$, which translates into inverse powers of the invariant mass(es) of the produced particle(s) $M$. Factorizations of this sort do not apply to low-velocity collisions. Arguments for this

result rely on relativistic causality and the uncertainty principle. The creation of the new state happens over timescales of order $1/M$. Before that well-defined event, the colliding particles are approaching at nearly the speed of light, and hence cannot affect the distributions of each others' partons. After the new particle is created, the fragments of the hadrons recede from each other, and the subsequent time development, when summed over all possible final states that include the heavy particle, is finite in perturbation theory as a direct result of the unitarity of QCD.

### Structure of Hadronic Final States

A wide range of semi-inclusive cross sections are defined by measuring properties of final states that depend only on the flow of energy, and which bring QCD perturbation theory to the threshold of nonperturbative dynamics. Schematically, for a state $N = |k_1 \dots k_N\rangle$, we define $\mathcal{S}(N) = \sum_i s(\Omega_i) k_i^0$, where $s(\Omega)$ is some smooth function of directions. We generalize the $e^+e^-$ annihilation case above, and define a cross section in terms of a related, but highly nonlocal, matrix element,

$$\frac{\mathrm{d}\sigma(Q)}{\mathrm{d}\mathcal{S}} \equiv \sigma_0 \int \mathrm{d}^4 x\, \mathrm{e}^{-\mathrm{i}Q\cdot x} \left\langle 0 \middle| J^\mu(0) \times \delta\left( \int \mathrm{d}^2\Omega s(\Omega)\mathcal{E}(\Omega) - \mathcal{S} \right) J_\mu(x) \middle| 0 \right\rangle \qquad [25]$$

where $\sigma_0$ is a zeroth-order cross section, and where $\mathcal{E}$ is an operator at spatial infinity, which measures the energy flow of any state in direction $\Omega$: $\mathcal{E}(\Omega) |k_1 \dots k_N\rangle = (1/Q) \sum_i k_i^0 \delta^2(\Omega - \Omega_i)$. This may seem a little complicated, but like the total annihilation cross section, the only dimensional scale on which it depends is $Q$. The operator $\mathcal{E}$ can be defined in a gauge-invariant manner, through the energy–momentum tensor for example, and has a meaning independent of partonic final states. At the same time, this sort of cross section may be implemented easily in perturbation theory, and like the total annihilation cross section, it is infrared safe. To see why, notice that when a massless ($k^2 = 0$) particle decays into two particles of momenta $xk$ and $(1-x)k$ ($0 \le x \le 1$), the quantity $\mathcal{S}$ is unchanged, since the sum of the new energies is the same as the old. This makes the observable $\mathcal{S}(N)$ insensitive to processes at low momentum transfer.

For the case of leptonic annihilation, the lowest-order perturbative contribution to energy flow requires no powers of $\alpha_s$, and consists of an oppositely moving quark and antiquark pair. Any measure of energy flow that includes these configurations will dominate over correlations that require $\alpha_s$ corrections. As a result, QCD predicts that in most leptonic annihilation events, energy will flow in two back-to-back collimated sets of particles, known as "jets." In this way, quarks and gluons are observed clearly, albeit indirectly.

With varying choices of $\mathcal{S}$, many properties of jets, such as their distributions in invariant mass, and the probabilities and angular distributions of multijet events, and even the energy dependence of their particle multiplicities, can be computed in QCD. This is in part because hadronization is dominated by the production of light quarks, whose production from the vacuum requires very little momentum transfer. Paradoxically, the very lightness of quarks is a boon to the use of perturbative methods. All these considerations can be extended to hadronic scattering, and jet and other semi-inclusive properties of final states also computed and compared to experiment.

## Conclusions

QCD is an extremely broad field, and this article has hardly scratched the surface. The relation of QCD-like theories to supersymmetric and string theories, and implications of the latter for confinement and the computation of higher-order perturbative amplitudes, have been some of the most exciting developments of recent years. As another example, we note that the reduction of the heavy-quark propagator to a nonabelian phase, noted in our discussion of confinement, is related to additional symmetries of heavy quarks in QCD, with many consequences for the analysis of their bound states. Of the bibliography given below, one may mention the four volumes of Shifman (2001, 2002), which communicate in one place a sense of the sweep of work in QCD.

Our confidence in QCD as the correct description of the strong interactions is based on a wide variety of experimental and observational results. At each stage in the discovery, confirmation, and exploration of QCD, the mathematical analysis of relativistic quantum field theory entered new territory. As is the case for gravity or electromagnetism, this period of exploration is far from complete, and perhaps never will be.

*See also:* AdS/CFT Correspondence; Aharonov–Bohm Effect; BRST Quantization; Current Algebra; Dirac Operator and Dirac Field; Euclidean Field Theory; Effective Field Theories; Electroweak Theory; Lattice Gauge Theory; Operator Product Expansion in Quantum Field Theory; Perturbation Theory and its Techniques; Perturbative Renormalization Theory and BRST; Quantum Field Theory: A Brief Introduction; Random

Matrix Theory in Physics; Renormalization: General Theory; Scattering in Relativistic Quantum Field Theory: Fundamental Concepts and Tools; Scattering, Asymptotic Completeness and Bound States; Seiberg–Witten Theory; Standard Model of Particle Physics.

## Further Reading


Bethke S (2004) $\alpha_s$ at Zinnowitz, 2004. *Nuclear Physics Proceeding Supplements* 135: 345–352.

Brodsky SJ and Lepage P (1989) Exclusive processes in quantum chromodynamics. In: Mueller AH (ed.) *Perturbative Quantum Chromodynamics*. Singapore: World Scientific.

Collins JC, Soper DE, and Sterman G (1989) Factorization. In: Mueller AH (ed.) *Perturbative Quantum Chromodynamics*. Singapore: World Scientific.

Dokshitzer Yu L and Kharzeev DE (2004) Gribov's conception of quantum chromodynamics. *Annual Review of Nuclear and Particle Science* 54: 487–524.

Dokshitzer Yu L, Khoze V, Troian SI, and Mueller AH (1988) QCD coherence in high-energy reactions. *Reviews of Modern Physics* 60: 373.

Eidelman S *et al.* (2004) Review of particle physics. *Physics Letters B* 592: 1–1109.

Ellis RK, Stirling WJ, and Webber BR (1996) QCD and Collider Physics, *Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology*, vol. 8. Cambridge: Cambridge University Press.

Greensite J (2003) The confinement problem in lattice gauge theory. *Progress in Particle and Nuclear Physics* 51: 1.

Mandelstam S (1976) Vortices and quark confinement in nonabelian gauge theories. *Physics Reports* 23: 245–249.

Muta T (1986) *Foundations of Quantum Chromodynamics*. Singapore: World Scientific.

Neubert H (1994) Heavy quark symmetry. *Physics Reports* 245: 259–396.

Polyakov AM (1977) Quark confinement and topology of gauge groups. *Nuclear Physics B* 120: 429–458.

Schafer T and Shuryak EV (1998) Instantons and QCD. *Reviews of Modern Physics* 70: 323–426.

Shifman M (ed.) (2001) *At the Frontier of Particle Physics: Handbook of QCD*, vols. 1–3. River Edge, NJ: World Scientific.

Shifman M (ed.) (2002) *At the Frontier of Particle Physics: Hand book of QCD*, vol. 4. River Edge, NJ: World Scientific.

Sterman G (1993) *An Introduction to Quantum Field Theory*. Cambridge: Cambridge University Press.

't Hooft G (1977) On the phase transition towards permanent quark confinement (1978). *Nuclear Physics B* 138: 1.

't Hooft G (ed.) (2005) *Fifty Years of Yang–Mills Theories*. Hackensack: World Scientific.

Weinberg S (1977) The problem of mass. *Transactions of the New York Academy of Science* 38: 185–201.

Wilson KG (1974) Confinement of quarks. *Physical Review D* 10: 2445–2459.


# Quantum Cosmology

**M Bojowald**, The Pennsylvania State University, University Park, PA, USA



## Introduction

Classical gravity, through its attractive nature, leads to a high curvature in important situations. In particular, this is realized in the very early universe where in the backward evolution energy densities are growing until the theory breaks down. Mathematically, this point appears as a singularity where curvature and physical quantities diverge and the evolution breaks down. It is not possible to set up an initial-value formulation at this place in order to determine the further evolution.

In such a regime, quantum effects are expected to play an important role and to modify the classical behavior such as the attractive nature of gravity or the underlying spacetime structure. Any candidate for quantum gravity thus allows us to reanalyze the singularity problem in a new light which implies the tests of the characteristic properties of the respective candidate. Moreover, close to the classical singularity, in the very early universe, quantum modifications will give rise to new equations of motion which turn into Einstein's equations only on larger scales. The analysis of these equations of motion leads to new classes of early universe phenomenology.

The application of quantum theory to cosmology presents a unique problem with not only mathematical but also many conceptual and philosophical ramifications. Since by definition there is only one universe which contains everything accessible, there is no place for an outside observer separate from the quantum system. This eliminates the most straightforward interpretations of quantum mechanics and requires more elaborate, and sometimes also more realistic, constructions such as decoherence. From the mathematical point of view, this situation is often expected to be mirrored by a new type of theory which does not allow one to choose initial or boundary conditions separately from the dynamical laws. Initial or boundary conditions, after all, are meant to specify the physical system prepared for observations which is impossible in cosmology. Since we observe only one universe, the expectation goes, our theories should finally present us with only

one, unique solution without any freedom for further conditions. This solution then contains all the information about observations as well as observers. Mathematically, this is an extremely complicated problem which has received only scant attention. Equations of motion for quantum cosmology are usually of the type of partial differential or difference equations such that new ingredients from quantum gravity are needed to restrict the large freedom of solutions.

## Minisuperspace approximation

In most investigations, the problem of applying full quantum gravity to cosmology is simplified by a symmetry reduction to homogeneous or isotropic geometries. Originally, the reduction was performed at the classical level, leaving in the isotropic case only one gravitational degree of freedom given by the scale factor $a$. Together with homogeneous matter fields, such as a scalar $\phi$, there are then only finitely many degrees of freedom which one can quantize using quantum mechanics. The classical Friedmann equation for the evolution of the scale factor, depending on the spatial curvature $k = 0$ or $\pm 1$, is then quantized to the Wheeler–DeWitt equation, commonly written as

$$\left(\frac{1}{9}\ell_{\mathrm{P}}^{4} a^{-x}\frac{\partial}{\partial a}a^{x}\frac{\partial}{\partial a} - ka^{2}\right)\psi(a,\phi) = -\frac{8\pi G}{3}a\hat{H}_{\mathrm{matter}}(a)\psi(a,\phi) \qquad [1]$$

for the wave function $\psi(a,\phi)$. The matter Hamiltonian $\hat{H}_{\mathrm{matter}}(a)$, such as

$$\hat{H}_{\mathrm{matter}}(a) = -\frac{1}{2}\hbar^{2}a^{-3}\frac{\partial^{2}}{\partial\phi^{2}} + a^{3}V(\phi) \qquad [2]$$

is left unspecified here, and $x$ parametrizes factor ordering ambiguities (but not completely). The Planck length $\ell_{\mathrm{P}} = \sqrt{8\pi G\hbar}$ is defined in terms of the gravitational constant $G$ and the Planck constant $\hbar$.

The central conceptual issue then is the generality of effects seen in such a symmetric model and its relation to the full theory of quantum gravity. This is completely open in the Wheeler–DeWitt form since the full theory itself is not even known. On the other hand, such relations are necessary to value any potential physical statement about the origin and early history of the universe. In this context, symmetric situations thus present models, and the degree to which they approximate full quantum gravity remains mostly unknown. There are examples, for instance, of isotropic models in anisotropic but still homogeneous models, where a minisuperspace quantization does not agree at all with the information obtained from the less symmetric model. However, often those effects already have a classical analog such as instability of the more symmetric solutions. A wider investigation of the reliability of models and when correction terms from ignored degrees of freedom have to be included has not been done yet.

With candidates for quantum gravity being available, the current situation has changed to some degree. It is then not only possible to reduce classically and then simply use quantum mechanics, but also perform at least some of the reduction steps at the quantum level. The relation to models is then much clearer, and consistency conditions which arise in the full theory can be made certain to be observed. Moreover, relations between models and the full theory can be studied to elucidate the degree of approximation. Even though new techniques are now available, a detailed investigation of the degree of approximation given by a minisuperspace model has not been completed due to its complexity.

This program has mostly been developed in the context of loop quantum gravity, where the specialization to homogeneous models is known as loop quantum cosmology. More specifically, symmetries can be introduced at the level of states and basic operators, where symmetric states of a model are distributions in the full theory, and basic operators are obtained by the dual action on those distributions. In such a way, the basic representation of models is not assumed but derived from the full theory where it is subject to much stronger consistency conditions. This has implications even in homogeneous models with finitely many degrees of freedom, despite the fact that quantum mechanics is usually based on a unique representation if the Weyl operators $e^{isq}$ and $e^{itp}$ for the variables $q$ and $p$ are represented weakly continuously in the real parameters $s$ and $t$.

The continuity condition, however, is not necessary in general, and so inequivalent representations are possible. In quantum cosmology this is indeed realized, where the Wheeler–DeWitt representation assumes that the conjugate to the scale factor, corresponding to extrinsic curvature of an isotropic slice, is represented through a continuous Weyl operator, while the representation derived for loop quantum cosmology shows that the resulting operator is not weakly continuous. Furthermore, the scale factor has a continuous spectrum in the Wheeler–DeWitt representation but a discrete spectrum in the loop representation. Thus, the underlying geometry

of space is very different, and also evolution takes a new form, now given by a difference equation of the type

$$\begin{aligned}&(V_{\mu+5}-V_{\mu+3})\mathrm{e}^{\mathrm{i}k}\psi_{\mu+4}(\phi)\\&\quad-(2+k^2)(V_{\mu+1}-V_{\mu-1})\psi_{\mu}(\phi)\\&\quad+(V_{\mu-3}-V_{\mu-5})\mathrm{e}^{-\mathrm{i}k}\psi_{\mu-4}(\phi)\\&\quad=-\tfrac{4}{3}\pi G\ell_{\mathrm{P}}^{2}\hat{H}_{\mathrm{matter}}(\mu)\psi_{\mu}(\phi)\end{aligned}\qquad[3]$$

in terms of volume eigenvalues $V_\mu=(\ell_{\mathrm{P}}^2|\mu|/6)^{3/2}$. For large $\mu$ and smooth wave functions, one can see that the difference equation reduces to the Wheeler–DeWitt equation with $|\mu|\propto a^2$ to leading order in derivatives of $\psi$. At small $\mu$, close to the classical singularity, however, both equations have very different properties and lead to different conclusions. Moreover, the prominent role of difference equations leads to new mathematical problems.

This difference equation is not simply obtained through a discretization of [1], but derived from a constraint operator constructed with methods from full loop quantum gravity. It is, thus, to be regarded as more fundamental, with [1] emerging in a continuum limit. The structure of [3] depends on the properties of the full theory such that its qualitative analysis allows conclusions for full quantum gravity.

## Applications

Traditionally, quantum cosmology has focused on three main conceptual issues:

- the fate of classical singularities,
- initial conditions and the "prediction" of inflation (or other early universe scenarios), and
- arrow of time and the emergence of a classical world.

The first issue consists of several subproblems since there are different aspects to a classical singularity. Often, curvature or energy densities diverge and one can expect quantum gravity to provide a natural cutoff. More importantly, however, the classical evolution breaks down at a singularity, and quantum gravity, if it is to cure the singularity problem, has to provide a well-defined evolution which does not stop. Initial conditions are often seen in relation to the singularity problem since early attempts tried to replace the singularity by choosing appropriate conditions for the wave function at $a=0$. Different proposals then lead to different solutions for the wave function, whose dependence on the scalar $\phi$ can be used to determine its probability distribution such as that for an inflaton. Since initial conditions often provide special properties early on, the combination of evolution and initial conditions has been used to find a possible origin of an arrow of time.

### Singularities

While classical gravity is based on spacetime geometry and thus metric tensors, this structure is viewed as emergent only at large scales in canonical quantum gravity. A gravitational system, such as a whole universe, is instead described by a wave function which, at best, yields expectation values for a metric. The singularity problem thus takes a different form since it is not metrics which need to be continued as solutions to Einstein's field equations but the wave function describing the quantum system. In the strong curvature regime around a classical singularity, one does not expect classical geometry to be applicable, such that classical singularities may just be a reflection of the breakdown of this picture, rather than a breakdown of physical evolution. Nevertheless, the basic feature of a singularity as presenting a boundary to the evolution of a system equally applies to the quantum equations. One can thus analyze this issue, using new properties provided by the quantum evolution.

The singularity issue is not resolved in the Wheeler–DeWitt formulation since energy densities, with $a$ being a multiplication operator, diverge and the evolution does not continue anywhere beyond the classical singularity at $a=0$. In some cases one can formally extend the evolution to negative $a$, but this possibility is not generic and leaves open what negative $a$ means geometrically. This is different in the loop quantization: here, the theory is based on triad rather than metric variables. There is thus a new sign factor corresponding to spatial orientation, which implies the possibility of negative $\mu$ in the difference equation. The equation is then defined on the full real line with the classical singularity $\mu=0$ in the interior. Outside $\mu=0$, we have positive volume at both sides, and opposite orientations. Using the difference equation, one can then see that the evolution does not break down at $\mu=0$, showing that the quantum evolution is singularity free.

For the example [3] shown here, one can follow the evolution, for instance, backward in internal time $\mu$, starting from initial values for $\psi$ at large positive $\mu$. By successively solving for $\psi_{\mu-4}$, the wave function at lower $\mu$ is determined. This goes on in this manner only until the coefficient $V_{\mu-3}-V_{\mu-5}$ of $\psi_{\mu-4}$ vanishes, which is the case if and only if $\mu=4$.

The value $\psi_0$ of the wave function exactly at the classical singularity is thus not determined by initial data, but one can easily see that it completely drops out of the evolution. In fact, the wave function at all negative $\mu$ is uniquely determined by initial values at positive $\mu$. Equation [3] corresponds to one particular ordering, which in the Wheeler–DeWitt case is usually parametrized by the parameter $x$ (although the particular ordering obtained from the continuum limit of [3] is not contained in the special family [1]). Other nonsingular orderings exist, such as that after symmetrizing the constraint operator, in which case the coefficients never become 0.

In more complicated systems, this behavior is highly nontrivial but still known to be realized in a similar manner. It is not automatic that the internal time evolution does not continue since even in isotropic models one can easily write difference equations for which the evolution breaks down. That the most natural orderings imply nonsingular evolution can be taken as a support of the general framework of loop quantum gravity. It should also be noted that the mechanism described here, providing essentially a new region beyond a classical singularity, presents one mechanism for quantum gravity to remove classical singularities, and so far the only known one. Nevertheless, there is no claim that the ingredients have to be realized in any nonsingular scenario in the same manner. Different scenarios can be imagined, depending on how quantum evolution is understood and what the interpretation of nonsingular behavior is. It is also not claimed that the new region is semiclassical in any sense when one looks at it at large volume. If the initial values for the wave function describe a semiclassical wave packet, its evolution beyond the classical singularity can be deformed and develop many peaks. What this means for the re-emergence of a semiclassical spacetime has to be investigated in particular models, and also in the context of decoherence.

## Initial Conditions

Traditional initial conditions in quantum cosmology have been introduced by physical intuition. The main mathematical problem, once such a condition is specified in sufficient detail, then is to study well-posedness, for instance, for the Wheeler–DeWitt equation. Even formulating initial conditions generally, and not just for isotropic models, is complicated, and systematic investigations of the well-posedness have rarely been undertaken. An exception is the historically first such condition, due to DeWitt, that the wave function vanishes at parts of minisuperspace, such as $a=0$ in the isotropic case, corresponding to classical singularities. This condition, unfortunately, can easily be seen to be ill posed in anisotropic models where in general the only solution vanishes identically. In other models, $\lim_{a\to 0}\psi(a)$ does not even exist. Similar problems of the generality of conditions arise in other scenarios. Most well known are the no-boundary and tunneling proposal where initial conditions are still imposed at $a=0$, but with a nonvanishing wave function there.

This issue is quite different for difference equations since at first the setup is less restrictive: there are no continuity or differentiability conditions for a solution. Moreover, oscillations that become arbitrarily rapid, which can be responsible for the nonexistence of $\lim_{a\to 0}\psi(a)$, cannot be supported on a discrete lattice. It can then easily happen that a difference equation is well posed, while its continuum limit with an analogous initial condition is ill posed. One example are the dynamical initial conditions of loop quantum cosmology which arise from the dynamical law in the following way: the coefficients in [3] are not always nonzero but vanish if and only if they are multiplied with the value of the wave function at the classical singularity $\mu=0$. This value thus decouples and plays no role in the evolution. The instance of the difference equation that would determine $\psi_0$, for example, the equation for $\mu=4$ in the backward evolution, instead implies a condition on the previous two values, $\psi_4$ and $\psi_8$, in the example. Since they have already been determined in previous iteration steps, this translates to a linear condition on the initial values chosen. We thus have one example where indeed initial conditions and the evolution follow from only one dynamical law, which also extends to anisotropic models. Without further conditions, the initial-value problem is always well posed, but may not be complete, in the sense that it results in a unique solution up to norm. Most of the solutions, however, will be rapidly oscillating. In order to guarantee the existence of a continuum approximation, one has to add a condition that these oscillations are suppressed in large volume regimes. Such a condition can be very restrictive, such that the issue of well-posedness appears in a new guise: nonzero solutions do exist, but in some cases all of them may be too strongly oscillating.

In simple cases, one can use generating function techniques advantageously to study oscillating solutions, at least if oscillations are of alternating nature between two subsequent levels of the difference equation. The idea is that a generating function $G(x)=\sum_n \psi_n x^n$ has a stronger pole at $x=-1$ if $\psi_n$

is alternating compared to a solution of constant sign. Choosing initial conditions which reduce the pole order thus implies solutions with suppressed oscillations. As an example, we can look at the difference equation

$$\psi_{n+1} + \frac{2}{n}\psi_n - \psi_{n-1} = 0 \qquad [4]$$

whose generating function is

$$G(x) = \frac{\psi_1 x + \psi_0(1 + 2x(1 - \log(1 - x)))}{(1 + x)^2} \qquad [5]$$

The pole at $x = -1$ is removed for initial values $\psi_1 = \psi_0(2\log 2 - 1)$ which corresponds to nonoscillating solutions. In this way, analytical expressions can be used instead of numerical attempts which would be sensitive to rounding errors. Similarly, the issue of finding bounded solutions can be studied by continued fraction methods. This illustrates how an underlying discrete structure leads to new questions and the application of new techniques compared to the analysis of partial differential equations which appear more commonly.

## More General Models

Most of the time, homogeneous models have been studied in quantum cosmology since even formulating the Wheeler–DeWitt equation in inhomogeneous cases, the so-called midisuperspace models, is complicated. Of particular interest among homogeneous models is the Bianchi IX model since it has a complicated classical dynamics of chaotic behavior. Moreover, through the Belinskii–Khalatnikov–Lifschitz (BKL) picture, the Bianchi IX mixmaster behavior is expected to play an important role even for general inhomogeneous singularities. The classical chaos then indicates a very complicated approach to classical singularities, with structure on arbitrarily small scales.

On the other hand, the classical chaos relies on a curvature potential with infinitely high walls, which can be mapped to a chaotic billiard motion. The walls arise from the classical divergence of curvature, and so quantum effects have been expected to change the picture, and shown to do so in several cases.

Inhomogeneous models (e.g., the polarized Gowdy models) have mostly been studied in cases where one can reformulate the problem as that of a massless free scalar on flat Minkowski space. The scalar can then be quantized with familiar techniques in a Fock space representation, and is related to metric components of the original model in rather complicated ways. Quantization can thus be performed, but transforming back to the metric at the operator level and drawing conclusions is quite involved. The main issue of interest in the recent literature has been the investigation of field theory aspects of quantum gravity in a tractable model. In particular, it turns out that self-adjoint Hamiltonians, and thus unitary evolution, do not exist in general.

Loop quantizations of inhomogeneous models are available even in cases where a reformulation such as a field theory on flat space does not exist, or is not being made use of to avoid special gauges. This is quite valuable in order to see if specific features exploited in reformulations lead to artifacts in the results. So far, the dynamics has not been investigated in detail, even though conclusions for the singularity issue can already be drawn.

From a physical perspective, it is most important to introduce inhomogeneities at a perturbative level in order to study implications for cosmological structure formation. On a homogeneous background, one can perform a mode decomposition of metric and matter fields and quantize the homogeneous modes as well as amplitudes of higher modes. Alternatively, one can first quantize the inhomogeneous system and then introduce the mode decomposition at the quantum level. This gives rise to a system of infinitely many coupled equations of infinitely many variables, which needs to be truncated, for example, for numerical investigations. At this level, one can then study the question to which degree a given minisuperspace model presents a good approximation to the full theory, and where additional correction terms should be introduced. It also allows one to develop concrete models of decoherence, which requires a “bath” of many weakly interacting degrees of freedom usually thought of as being provided by inhomogeneities in cosmology, and an understanding of the semiclassical limit.

## Interpretations

Due to the complexity of full gravity, investigations without symmetry assumptions or perturbative approximations usually focus on conceptual issues. As already discussed, cosmology presents a unique situation for physics since there cannot be any outside observer. While this fact has already implications on the interpretation of observations at the classical level, its full force is noticed only in quantum cosmology. Since some traditional interpretations of quantum mechanics require the role of

observers outside the quantum system, they do not apply to quantum cosmology.

Sometimes, alternative interpretations such as Bohm theory or many-world scenarios are championed in this situation, but more conventional relational pictures are most widely adopted. In such an interpretation, the wave function yields relational probabilities between degrees of freedom rather than absolute probabilities for measurements done by an outside observer. This has been used, for instance, to determine the probability of the right initial conditions for inflation, but it is marred by unresolved interpretational issues and still disputed. These problems can be avoided by using effective equations, in analogy to an effective action, which modify classical equations on small scales. Since the new equations are still of classical type, that is, differential equations in coordinate time, no interpretational issues arise at least if one stays in semiclassical regimes. In this manner, new inflationary scenarios motivated from quantum cosmology have been developed.

In general, a relational interpretation, though preferable conceptually, leads to technical complications since the situation is much more involved and evolution is not easy to disentangle. In cosmology, one often tries to single out one degree of freedom as internal time with respect to which evolution of other degrees of freedom is measured. In homogeneous models, one can simply take the volume as internal time, such as $a$ or $\mu$ earlier, but in full no candidate is known. Even in homogeneous models, the volume is not suitable as internal time to describe a possible recollapse. One can use extrinsic curvature around such a point, but then one has to understand what changing the internal time in quantum cosmology implies, that is, whether evolution pictures obtained in different internal time formulations are equivalent to each other.

There are thus many open issues at different levels, which, strictly speaking, do not apply only to quantum cosmology but to all of physics. After all, every physical system is part of the universe, and thus a potential ingredient of quantum cosmology. Obviously, physics works well in most situations without taking into account its being part of one universe. Similarly, much can be learned about a quantum universe if only some degrees of freedom of gravity are considered as in mini- or midisuperspace models. In addition, complicated interpretational issues, as important as they are for a deep understanding of quantum physics, do not prevent the development of physical applications in quantum cosmology, just as they did not do so in the early stages of quantum mechanics.

*See also:* Canonical General Relativity; Cosmology: Mathematical Aspects; Loop Quantum Gravity; Quantum Geometry and its Applications; Spacetime Topology, Causal Structure and Singularities; Wheeler–De Witt Theory.

## Further Reading

Bojowald M (2001a) Absence of a singularity in loop quantum cosmology. *Physical Review Letters* 86: 5227–5230 (gr-qc/0102069).

Bojowald M (2001b) Dynamical initial conditions in quantum cosmology. *Physical Review Letters* 87: 121301 (gr-qc/0104072).

Bojowald M (2003) Initial conditions for a universe. *General Relativity and Gravitation* 35: 1877–1883 (gr-qc/0305069).

Bojowald M (2005) Loop quantum cosmology. *Living Reviews in Relativity* (to appear).

Bojowald M and Morales-Técotl HA (2004) Cosmological applications of loop quantum gravity. In: *Proceedings of the Fifth Mexican School (DGFM): The Early Universe and Observational Cosmology,* Lecture Notes in Physics, vol. 646, pp. 421–462. Berlin: Springer (gr-qc/0306008).

DeWitt BS (1967) Quantum theory of gravity. I. The canonical theory. *Physical Review* 160: 1113–1148.

Giulini D, Kiefer C, Joos E, Kupsch J, Stamatescu IO *et al.* (1996) *Decoherence and the Appearance of a Classical World in Quantum Theory*. Berlin: Springer.

Hartle JB (2003) What connects different interpretations of quantum mechanics? quant-ph/0305089.

Hartle JB and Hawking SW (1983) Wave function of the universe. *Physical Review D* 28: 2960–2975.

Kiefer C (2005) Quantum cosmology and the arrow of time. In: *Proceedings of the Conference DICE2004*, Piombino, Italy, September 2004 (gr-qc/0502016).

Kuchař KV (1992) Time and interpretations of quantum gravity. In: Kunstatter G, Vincent DE, and Williams JG (eds.) *Proceedings of the 4th Canadian Conference on General Relativity and Relativistic Astrophysics*. Singapore: World Scientific.

McCabe G (2005) The structure and interpretation of cosmology: Part II. The concept of creation in inflation and quantum cosmology. *Studies in History and Philosophy of Modern Physics* 36: 67–102 (gr-qc/0503029).

Vilenkin A (1984) Quantum creation of universes. *Physical Review D* 30: 509–511.

Wiltshire DL (1996) An introduction to quantum cosmology. In: Robson B, Visvanathan N, and Woolcock WS (eds.) *Cosmology: The Physics of the Universe*, pp. 473–531. Singapore: World Scientific (gr-qc/0101003).

# Quantum Dynamical Semigroups

**R Alicki**, University of Gdańsk, Gdańsk, Poland



## Introduction

With a given quantum system we associate a Hilbert space $\mathcal{H}$ such that pure states of the system are represented by normalized vectors $\psi$ in $\mathcal{H}$ or equivalently by one-dimensional projections $|\psi\rangle\langle\psi|$, whereas mixed states are given by density matrices $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|, p_j > 0, \sum_j p_j = 1$, that is, positive trace-1 operators and observables are identified with self-adjoint operators $A$ acting on $\mathcal{H}$. The mean value of an observable $A$ at a state $\rho$ is given by the following expression:

$$< A >_\rho = \mathrm{tr}(\rho A) \qquad [1]$$

The time evolution of the isolated system is determined by the self-adjoint operator $H$ (Hamiltonian) corresponding to the energy of the system. The infinitesimal change of state of the isolated system can be written as

$$\begin{aligned} \psi(t+dt) &= \psi(t) - \mathrm{i}H\mathrm{d}t\psi(t), \quad \text{or} \\ \rho(t+\mathrm{d}t) &= \rho(t) - \mathrm{i}\mathrm{d}t[H,\rho] \end{aligned} \qquad [2]$$

what leads to a reversible purity preserving unitary dynamics $\psi(t) = \mathrm{e}^{-\mathrm{i}tH}\psi, \rho(t) = \mathrm{e}^{-\mathrm{i}tH}\rho\mathrm{e}^{\mathrm{i}tH}$. We use the notation $[A,B] \equiv AB - BA, \{A,B\} = AB + BA$ and put $\hbar \equiv 1$. An interaction with environment leads to irreversible changes of the density matrix transforming, in general, pure states into mixed ones. Such a process can be modeled phenomenologically by a transition map $V : \mathcal{H} \mapsto \mathcal{H}$ leading to

$$\rho(t+\mathrm{d}t) = \rho(t) + \mathrm{d}tV\rho V^* - \mathrm{d}t\frac{1}{2}\{V^*V,\rho\} \qquad [3]$$

Combining Hamiltonian dynamics with several irreversible processes governed by a family of transition operators $\{V_j\}$ we obtain the following formal evolution equation in the Schrödinger picture (quantum Markovian master equation)

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\rho(t) &= \mathcal{L}\rho(t) \\ &= -\mathrm{i}[H,\rho(t)] + \frac{1}{2}\sum_{j\in I}\left([V_j,\rho(t)V_j^*] + [V_j\rho(t),V_j^*]\right) \\ &= D\rho(t) + \rho(t)D^* + \Phi\rho(t) \end{aligned} \qquad [4]$$

with the initial condition $\rho(0) = \rho$. Here $D = -\mathrm{i}H - (1/2)\sum_j V_j^* V_j$, $\Phi\rho = \sum_{j\in I} V_j\rho V_j^*$, and $I$ is a certain countable set of indices. Assume for the moment that the Hilbert space $\mathcal{H} = \mathbb{C}^n$. Then the eqn [4] is always meaningful and its solution is given in terms of the exponential $\rho(t) = \Lambda(t)\rho \equiv \mathrm{e}^{t\mathcal{L}}\rho$. The linear map $\Phi$ is a general completely positive map on matrices, which preserves the positivity of $\rho$ and $\Phi \otimes \mathbb{I}_d$ preserves positivity of $nd \times nd$ matrices for arbitrary $d = 1,2,3,\ldots$ A useful Dyson-type expansion

$$\begin{aligned} \mathrm{e}^{t\mathcal{L}}\rho = \mathcal{W}(t)\rho + \sum_{k=1}^{\infty}\int_0^t dt_k \int_0^{t_k} dt_{k-1}\cdots \\ \times \int_0^{t_2} dt_1 \mathcal{W}(t-t_k)\Phi\mathcal{W}(t_k - t_{k-1}) \\ \times \Phi\cdots\Phi\mathcal{W}(t_1)\rho \end{aligned} \qquad [5]$$

with $\mathcal{W}(t)\rho \equiv W(t)\rho W(t)^*, W(t) = \mathrm{e}^{tD}$ shows that $\Lambda(t)$ is also completely positive. It is often convenient to describe quantum evolution in terms of observables (Heisenberg picture)

$$\begin{aligned} < A >_{\rho(t)} &= \mathrm{tr}\left((\mathrm{e}^{t\mathcal{L}}\rho)A\right) \\ &= \mathrm{tr}\left(\rho\,\mathrm{e}^{t\mathcal{L}^*}A\right) = < A(t) >_\rho \end{aligned} \qquad [6]$$

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}A(t) &= \mathcal{L}^*A(t) \\ &= \mathrm{i}[H,A(t)] + \frac{1}{2}\sum_{j\in I}\left(V_j^*[A(t),V_j] + [V_j^*,A(t)]V_j\right) \\ &= D^*\rho(t) + \rho(t)D + \Phi^*\rho(t) \end{aligned} \qquad [7]$$

with the initial condition $A(0) = A$, completely positive $\Phi^*A = \sum_{j\in I} V_j^* A V_j$ and the corresponding Dyson expansion.

The solutions of eqns [4] and [7] are given in terms of dynamical semigroups. Their general mathematical properties and particular examples will be reviewed in this article. Various methods of derivation of master equations for open quantum systems from the underlying Hamiltonian dynamics of composed systems will also be presented.

## Semigroups and Their Generators

For standard quantum-mechanical models it is convenient to define quantum dynamical semigroup in the Schrödinger picture as a one-parameter family $\{\Lambda(t); t \geq 0\}$ of linear and bounded maps acting on the Banach space of trace-class operators $\mathcal{T}(\mathcal{H})$ equipped with the norm $\|\sigma\|_1 = \mathrm{tr}(\sigma\sigma^*)^{1/2}$ and satisfying the following conditions:

1. *Composition (semigroup) law*

$$\Lambda(t)\Lambda(s) = \Lambda(t+s), \quad \text{for all } t,s \geq 0 \qquad [8]$$

2. *Complete positivity*

$$\Lambda(t) \otimes \mathbb{I}_d \text{ is positive on } \mathcal{T}(\mathcal{H} \otimes \mathbb{C}^d) \text{ for all } d = 1, 2, 3, \ldots \text{ and } t \geq 0 \quad [9]$$

3. *Conservativity* (*trace preservation*)

$$\mathrm{tr}(\Lambda(t)\rho) = \mathrm{tr}(\rho), \quad \text{for all } \rho \in \mathcal{T}(\mathcal{H}) \quad [10]$$

4. *Continuity* (*in a weak sense*)

$$\lim_{t \to 0} \mathrm{tr}(A\Lambda(t)\rho) = \mathrm{tr}(A\rho) \text{ for all } \rho \in \mathcal{T}(\mathcal{H}),\ A \in \mathcal{B}(\mathcal{H}) \quad [11]$$

From a general theory of one-parameter semigroups on Banach spaces it follows that under the conditions (1)–(4) $\Lambda(t)$ is a one-parameter strongly continuous semigroup of contractions on $\mathcal{T}(\mathcal{H})$ uniquely characterized by a generally unbounded but densely defined semigroup generator $\mathcal{L}$ with the domain $\mathrm{dom}(\mathcal{L}) \subset \mathcal{T}(\mathcal{H})$ such that for any $\rho \in \mathrm{dom}(\mathcal{L})$

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) = \mathcal{L}\rho(t), \quad \rho(t) = \Lambda(t)\rho \quad [12]$$

One can show that for $\lambda > 0$ the resolvent $R(\lambda) = (\lambda\mathbb{I} - \mathcal{L})^{-1}$ can be extended to a bounded operator satisfying $\|R(\lambda)\| \leq \lambda^{-1}$ and, therefore, the following formula makes sense:

$$\lim_{n \to \infty} \left(\mathbb{I} - \frac{t}{n}\mathcal{L}\right)^{-n} \rho = \Lambda(t)\rho, \quad \text{for all } \rho \in \mathcal{T}(\mathcal{H}) \quad [13]$$

Under the additional assumption that the generator $\mathcal{L}$ is bounded (and hence everywhere defined) Gorini, *et al.* (1976) and Lindblad (1976) proved that eqns [4] and [7] with bounded $H, V_j$ and $\sum_j V_j V_j^*$ provide the most general form of $\mathcal{L}$. The choice of $H$ and $V_j$ is not unique and the sum over $j$ can be replaced by an integral. In the case of $n$-dimensional Hilbert space we can always choose the form of eqn [4] with at most $n^2 - 1$ $V_j$'s. Sometimes the structure [4] is hidden as for the following useful example of the relaxation process to a fixed density matrix $\rho_0$ with the rate $\mu > 0$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) = \mu(\rho_0 - \rho(t)) \quad [14]$$

The general structure of an unbounded $\mathcal{L}$ is not known. However, the formal expressions [4] and [7] with possibly unbounded $D$ and $V_j$ are meaningful under the following conditions:

- the operator $D$ generates a strongly continuous contracting semigroup $\{\mathrm{e}^{tD}; t \geq 0\}$ on $\mathcal{H}$;
- $\mathrm{dom}(V_j) \supseteq \mathrm{dom}(D)$, for all $j$;
- $\langle\phi, D\psi\rangle + \langle D\phi, \psi\rangle + \sum_j \langle V_j\phi, V_j\psi\rangle = 0$, for all $\phi, \psi \in \mathrm{dom}(D)$.

We can solve eqn [4] in terms of a minimal solution. Defining by $Z$ the generator of the contracting semigroup $\rho \mapsto \mathrm{e}^{tD}\rho\mathrm{e}^{tD^*}$ and denoting by $J$ the completely positive (unbounded) map $\rho \mapsto \sum_{j \in I} V_j \rho V_j^*$, one can show that for any $\lambda > 0$, $J(\lambda\mathbb{I} - Z)^{-1}$ possesses a unique bounded completely positive extension denoted by $A_\lambda$ with $\|A_\lambda\| \leq 1$. Hence, for any $0 \leq r < 1$ there exists a strongly continuous, completely positive and contracting semigroup $\Lambda^{(r)}(t)$ with the resolvent explicitly given by

$$R^{(r)}(\lambda) = (\lambda\mathbb{I} - Z)^{-1} \sum_{k=0}^{\infty} r^k A_\lambda^k \quad [15]$$

As $\|R^{(r)}(\lambda)\| \leq 1$ the limit $\lim_{r \to 1} R^{(r)}(\lambda) = R(\lambda)$, where $R(\lambda)$ is the resolvent of the semigroup $\Lambda(t)$ satisfying (1), (2), and (4) and called the minimal solution of the eqn [4]. The minimal solution need not be a unique solution or conservative (generally $\mathrm{tr}\,\rho(t) \leq \mathrm{tr}\,\rho(0)$ and for any other solution $\rho'(t) \geq \rho(t)$). There exist useful sufficient conditions for conservativity, an example of a sufficient and necessary condition is the following: $A_\lambda^n \to 0$ strongly as $n \to \infty$ for all $\lambda > 0$ (Chebotarev and Fagnola 1988).

### Examples

**Bloch equation** The simplest two-level system can be described in terms of spin operators $S_k = (1/2)\sigma_k, k = 1, 2, 3$, where $\sigma_k$ are Pauli matrices. The most general master equation of the form [4] can be written as (Alicki and Fannes 2001, Ingarden *et al.* 1997)

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -\mathrm{i}\sum_{k=1}^{3} h_k[S_k, \rho] + \frac{1}{2}\sum_{k,l=1}^{3} a_{kl}\{[S_k\rho, S_l] + [S_k, \rho S_l]\} \quad [16]$$

where $h_k \in \mathbb{R}$ and $[a_{kl}]$ is a $3 \times 3$ complex, positively defined matrix. Introducing the magnetization vector $M_k(t) = \mathrm{tr}(\rho(t)S_k)$, we obtain the following Bloch equation used in the magnetic resonance theory:

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{M}(t) = \boldsymbol{h} \times (\boldsymbol{M}(t) - \boldsymbol{M}_0) - \mathbb{F}(\boldsymbol{M}(t) - \boldsymbol{M}_0) \quad [17]$$

where the tensor $\mathbb{F}$ (real, symmetric, and positive $3 \times 3$ matrix) and the vector $\boldsymbol{M}_0$ are functions of $[a_{kl}]$. In particular, complete positivity implies the following inequalities for the inverse relaxation times $\gamma_1, \gamma_2, \gamma_3$ (eigenvalues of $\mathbb{F}$):

$$\gamma_k \geq 0, \quad \gamma_1 + \gamma_2 \geq \gamma_3, \quad \gamma_3 + \gamma_1 \geq \gamma_2, \quad \gamma_2 + \gamma_3 \geq \gamma_1 \quad [18]$$

**Damped and pumped harmonic oscillator** The quantum master equation for a linearly damped and pumped harmonic oscillator with frequency $\omega$ and the damping (pumping) coefficient $\gamma_{\downarrow}(\gamma_{\uparrow})$ has form

$$\frac{d\rho}{dt} = -i\omega[a^*a,\rho] + \frac{\gamma_{\downarrow}}{2}([a\rho,a^*]+[a,\rho a^*]) + \frac{\gamma_{\uparrow}}{2}([a^*\rho,a]+[a^*,\rho a]) \quad [19]$$

where $a^*, a$ are creation and annihilation operators satisfying $[a,a^*]=1$. Taking diagonal elements $p_n = <n,\rho n>$ in the "particle number" basis $a^*a|n>=n|n>$, $n=0,1,2,\ldots$, which evolve independently of the off-diagonal elements, one obtains the birth and death process,

$$\frac{dp_n}{dt} = \gamma_{\downarrow}(n+1)p_{n+1} + \gamma_{\uparrow}np_{n-1} - \big(\gamma_{\downarrow}n + \gamma_{\uparrow}(n+1)\big)p_n \quad [20]$$

It is convenient to use the Heisenberg picture and find an explicit solution in terms of Weyl unitary operators $W(z) = \exp[(i/\sqrt{2})(za+\bar{z}a^*)]$,

$$\Lambda^*(t)W(z) = \exp\left\{-\frac{|z|^2}{4}\frac{\gamma_{\downarrow}}{\gamma_{\downarrow}-\gamma_{\uparrow}}\left(1-e^{-(\gamma_{\downarrow}-\gamma_{\uparrow})t}\right)\right\}W(z(t)) \quad [21]$$

where $z(t)=\exp\{-(i\omega+\frac{1}{2}(\gamma_{\downarrow}-\gamma_{\uparrow}))t\}, t\geq 0$. For $\gamma_{\downarrow}>\gamma_{\uparrow}$ the solution of eqn [19] always tends to the stationary Gibbs state

$$\rho_{\beta} = Z^{-1}e^{-\beta\omega a^*a}, \qquad Z = \mathrm{tr}e^{-\beta\omega a^*a}$$
$$\beta = \frac{1}{\omega}\ln(\gamma_{\downarrow}/\gamma_{\uparrow}) \quad [22]$$

**Quasifree semigroups** The previous example is the simplest instance of the dynamical semigroups for noninteracting bosons and fermions which are completely determined on the single-particle level. Such systems are defined by a single-particle Hilbert space $\mathcal{H}_1$ and a linear map $\mathcal{H}_1 \ni \phi \mapsto a^*(\phi)$ into creation operators satisfying canonical commutation or anticommutation relations (CCRs or CARs, respectively) for bosons and fermions, respectively

$$[a(\psi),a^*(\phi)]_{\pm} = <\psi,\phi>$$
$$[A,B]_{\pm} \equiv AB - (\pm 1)BA \quad [23]$$

In all expressions containing $(\pm)$, sign $(+)$ refers to bosons and $(-)$ to fermions.

Consider a nonhomogeneous evolution equation on the trace-class operators $\sigma \in \mathcal{T}(\mathcal{H}_1)$:

$$\frac{d\sigma}{dt} = -i[H_1,\sigma] - \frac{1}{2}\{(\Gamma_{\downarrow}-(\pm)\Gamma_{\uparrow}),\sigma\} + \Gamma_{\uparrow} \quad [24]$$

with a single-particle Hamiltonian $H_1$ and a damping (pumping) positive operator $\Gamma_{\downarrow}(\Gamma_{\uparrow}) \geq 0$. The operators $H_1, \Gamma_{\downarrow}$, and $\Gamma_{\uparrow}$ need not be bounded provided $-iH_1 - (1/2)\{(\Gamma_{\downarrow} - (\pm)\Gamma_{\uparrow})$ generates a (contracting in the fermionic case) semigroup $\{T(t); t\geq 0\}$ on $\mathcal{H}_1$ and the formal solution of eqn [24]

$$\sigma(t) = T(t)\sigma(0)T^*(t) + Q(t)$$
$$\text{where } Q(t) = \int_0^t T(s)\Gamma_{\uparrow}T^*(s)ds \quad [25]$$

is meaningful. We can now define the quasifree dynamical semigroup for the many-particle system described by the Fock space $\mathcal{F}_{\pm}(\mathcal{H}_1)$ (Alicki and Lendi 1987, Alicki and Fannes 2001). The simplest definition involves Heisenberg evolution of the ordered monomials in $a^*(\psi_j)$ and $a(\phi_j)$:

$$\Lambda^*(t)a^*(\psi_1)\cdots a^*(\psi_m)a(\phi_1)\cdots a(\phi_n) = \sum_P \epsilon^{\pm}\mathrm{Det}_{\pm}\big[<\psi_{j_k},Q(t)\phi_{i_l}>\big]_{k,l=1,2,\ldots,r} \times a^*(T^*(t)\psi_{\alpha_1})\cdots a^*(T^*(t)\psi_{\alpha_{m-r}}) \times a\big(T^*(t)\phi_{\beta_1}\big)\cdots a\big(T^*(t)\phi_{\beta_{n-r}}\big) \quad [26]$$

The sum is taken over all partitions $\{(j_1,\ldots,j_r)(\alpha_1,\ldots,\alpha_{m-r})\}, \{(i_1,\ldots,i_r)(\beta_1,\ldots,\beta_{n-r}\}$ such that $j_1<j_2<\cdots<j_r$, $\alpha_1<\alpha_2,\cdots<\cdots<\alpha_{m-r}$, $i_1<i_2<\cdots<r_r$, $\beta_1<\beta_2\cdots<\alpha_{n-r}$; $\epsilon^+\equiv 1$, $\epsilon^-$ is a product of signatures of the permutations $\{1,2,\ldots,m\}\mapsto\{j_1,\ldots,j_r,\alpha_1,\ldots,\alpha_{m-r}\}, \{1,2,\ldots,n\}\mapsto\{i_1,\ldots,i_r,\beta_1,\ldots,\beta_{n-r}\}$; a permanent $\mathrm{Det}_+$ is taken for bosons, a determinant $\mathrm{Det}_-$ for fermions.

Introducing an orthonormal basis $\{e_k\}$ in $\mathcal{H}_1$ and using the notation $a^*(e_k)\equiv a_k^*$, we can write a formal master equation for density matrices on the Fock space corresponding to eqn [26]:

$$\frac{d\rho}{dt} = -i[H_F,\rho] + \frac{1}{2}\sum_{k,l}\Big(\Gamma_{\downarrow}^{kl}\big([a_k\rho,a_l^*] + [a_k,\rho a_l^*]\big) + \Gamma_{\uparrow}^{kl}\big([a_k^*\rho,a_l]+[a_k^*,\rho a_l]\big)\Big) \quad [27]$$

Again, formally,

$$H_F = \sum_{k,l} <e_k,H_1e_l>a_k^*a_l$$
$$\Gamma_{\downarrow}^{kl} = <e_k,\Gamma_{\downarrow}e_l>, \qquad \Gamma_{\uparrow}^{kl} = <e_k,\Gamma_{\uparrow}e_l> \quad [28]$$

Often the formulas [27], [28] are not well-defined, but replacing the (infinite) matrices by (distribution-valued) integral kernels, sums by integrals, and $a_k^*, a_l$ by quantum fields, we can obtain meaningful objects.

Quasifree dynamical semigroups find applications in the theory of unstable particles, quantum linear

optics, solid-state physics, quantum information theory, etc. (Alicki and Lendi 1987, Sewell 2002).

## Ergodic Properties

Dynamical semigroups which possess stationary states satisfying $\mathcal{L}\rho_0 = 0$ are of particular interest, for example, in the description of relaxation processes toward equilibrium states (Frigerio 1977, Spohn 1980, Alicki and Lendi 1987). The dynamical semigroup $\{\Lambda(t)\}$ with a stationary state $\rho_0$ is called ergodic if

$$\lim_{t\to\infty} \Lambda(t)\rho = \rho_0, \quad \text{for any initial } \rho \qquad [29]$$

For the case of finite-dimensional $\mathcal{H}$ at least one stationary state always exists. If, moreover, it is strictly positive, $\rho_0 > 0$, then we have the following sufficient condition of ergodicity:

$$\{V_j; j\in I\}' \equiv \{A; A\in\mathcal{B}(\mathcal{H}), [A, V_j] = 0, j\in I\} = \mathbb{C}\mathbf{1} \qquad [30]$$

Open systems interacting with heat baths at the temperature $T$ are described by the semigroups with generators [4] of the special form

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) = -\mathrm{i}[H,\rho(t)] + \frac{1}{2}\sum_{\omega_j\geq 0}\Big\{\Big([V_j,\rho(t)V_j^*] + [V_j\rho(t), V_j^*]\Big) + \mathrm{e}^{-\beta\omega_j}\Big([V_j^*,\rho(t)V_j] + [V_j^*\rho(t), V_j]\Big)\Big\} \qquad [31]$$

where

$$\beta = \frac{1}{k_{\mathrm{B}}T}, \qquad [H, V_j] = -\omega_j V_j \qquad [32]$$

The Gibbs state $\rho_\beta = Z^{-1}\,\mathrm{e}^{-\beta H}$ is a stationary state for eqn [31] and the condition $\{V_j, V_j^*; j\in I\}' = \mathbb{C}\mathbf{1}$ implies ergodicity (return to equilibrium). Moreover, the matrix elements of $\rho$ diagonal in $H$-eigenbasis transform independently of the off-diagonal ones and satisfy the Pauli master equation

$$\frac{\mathrm{d}p_k}{\mathrm{d}t} = \sum_l (a_{kl}p_l - a_{lk}p_k) \qquad [33]$$

with the detailed balance condition $a_{kl}\,\mathrm{e}^{-\beta E_l} = a_{lk}\,\mathrm{e}^{-\beta E_k}$, where $E_k$ are eigenvalues of $H$.

Define the new Hilbert space $\mathcal{L}^2(\mathcal{H},\rho_\beta)$ as a completion of $\mathcal{B}(\mathcal{H})$ with respect to the scalar product $(A,B)_{\rho_\beta} \equiv \mathrm{tr}(\rho_\beta A^* B)$. The semigroup's generators in the Heisenberg picture corresponding to eqn [31] are normal operators in $\mathcal{L}^2(\mathcal{H},\rho_\beta)$ with the Hamiltonian part $\mathrm{i}[H,\cdot\,]$ being the anti-Hermitian one (automatically for bounded $\mathcal{L}^*$, and for unbounded one under technical conditions concerning domains). This allows spectral decomposition of $\mathcal{L}^*$ and a proper definition of damping rates for the obtained eigenvectors. The normality condition is one of the possible definitions of quantum detailed balance. The other, based on the time-reversal operation, often coincides with the previous one for important examples.

Interesting examples of nonergodic dynamical semigroups are given for open systems consisting of $N$ identical particles with Hamiltonians $H^{(N)}$ and operators $V_j^{(N)}$ invariant with respect to particles permutations. Then the commutant $\{H^{(N)}, V_j^{(N)}, j\in I\}'$ contains an abelian algebra generated by projections on irreducible tensors corresponding to Young tables.

## From Hamiltonian Dynamics to Semigroups

One of the main tasks in the quantum theory of open systems is to derive master equations [4] from the model of a "small" open system $S$ interacting with a "large" reservoir $R$ at a certain reference state $\omega_R$ (Davies 1976, Spohn 1980, Alicki and Lendi 1987, Breurer and Petruccione 2002, Garbaczewski and Olkiewicz 2002). Starting with the total Hamiltonian $H_\lambda = H_S\otimes\mathbf{1}_R + \mathbf{1}_S\otimes H_R + \lambda\sum_\alpha S_\alpha\otimes R_\alpha$, where $S_\alpha = S_\alpha^*, R_\alpha = R_\alpha^*, \mathrm{tr}(\omega_R R_\alpha) = 0$, and $\lambda$ is a coupling constant, we define the reduced dynamics of $S$ by

$$\rho(t) = \Lambda^{(\lambda)}(t)\rho = \mathrm{tr}_R\big(U_\lambda(t)\rho\otimes\omega_R U_\lambda^*(t)\big) \qquad [34]$$

with $U_\lambda(t) = \exp(-\mathrm{i}tH_\lambda)$. Here $\mathrm{tr}_R$ denotes a partial trace over $R$ defined in terms of an arbitrary basis $\{e_k\}$ of $R$ by the formula $<\phi,(\mathrm{tr}_R A)\phi> = \sum_k <\phi\otimes e_k, A\phi\otimes e_k>$. Generally, $\Lambda^{(\lambda)}(t+s)\neq\Lambda^{(\lambda)}(t)\Lambda^{(\lambda)}(s)$, but dynamical semigroups can provide good approximations in important cases.

### Weak-Coupling Limit

Under the conditions of sufficiently fast decay of multitime correlation functions constructed from the observables $R_\alpha$ at the state $\omega_R$, one can prove that for small coupling constant $\lambda$ the exact dynamical map $\Lambda^{(\lambda)}(t)$ can be approximated by the dynamical semigroup corresponding to the following master equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) = -\mathrm{i}[H,\rho(t)] + \frac{\lambda^2}{2}\sum_{\alpha\beta}\sum_{\omega\in Sp} C_{\alpha\beta}(\omega) \times\Big([V_\omega^\alpha,\rho(t)V_\omega^{\beta*}] + [V_\omega^\alpha\rho(t), V_\omega^{\beta*}]\Big) \qquad [35]$$

where $H = H_S + \lambda^2 \sum_{\alpha\beta} \sum_{\omega\in Sp} K_{\alpha\beta}(\omega) V_\omega^{\alpha*} V_\omega^\beta$ is a renormalized Hamiltonian, $\sum_{\omega\in Sp}$ denotes the sum over eigenfrequencies of $-[H,\cdot]$, $e^{itH} S_\alpha e^{-itH} = \sum_{\omega\in Sp} V_\omega^\alpha e^{-i\omega t}$ and

$$\int_0^\infty e^{i\omega t}\mathrm{tr}\big(\omega_R e^{itH_R} R_\alpha e^{-itH_R} R_\beta\big)dt = \tfrac{1}{2} C_{\alpha\beta}(\omega) + iK_{\alpha\beta}(\omega) \quad [36]$$

The rigorous derivation involves van Hove or weak coupling limit, $\lambda \to 0$, with $\tau = \lambda^2 t$ kept fixed.

It follows from the Bochner theorem that the matrix $[C_{\alpha\beta}(\omega)]$ is positively defined and therefore by its diagonalization we can convert eqn [35] into the standard form [4]. If the reservoir's state $\omega_R$ is an equilibrium state (Kubo–Martin–Schwinger state) then $C_{\alpha\beta}(-\omega) = e^{-\omega/k_B T}\ C_{\beta\alpha}(\omega)$ and therefore eqn [35] can be written in a form [31]. Moreover, transition probabilities $a_{kl}$ from eqn [33] coincide with those obtained using the "Fermi golden rule."

### Low-Density Limit

If the reservoir can be modeled by a gas of noninteracting particles (bosons or fermions) at low density $\nu$, we can derive the following master equation which approximates an exact dynamics [34] in the low-density limit ($\nu \to 0$, with $\tau = \nu t$ kept fixed)

$$\frac{d}{dt}\rho(t) = -i[H,\rho(t)] + \nu\pi \sum_{\omega\in S} \int_{\mathbb{R}^6} d^3\boldsymbol{p}\, d^3\boldsymbol{p}' G(\boldsymbol{p}) \times \delta\big(E_{p'} - E_p + \omega\big)\big([T_\omega(\boldsymbol{p},\boldsymbol{p}'),\rho(t)T_\omega(\boldsymbol{p},\boldsymbol{p}')^*] + [T_\omega(\boldsymbol{p},\boldsymbol{p}')\rho(t), T_\omega(\boldsymbol{p},\boldsymbol{p}')^*]\big) \quad [37]$$

Here $H$ is a renormalized Hamiltonian of the system S, $e^{itH} T e^{-itH} = \sum_{\omega\in S} T_\omega e^{-i\omega t}$, $T$ is a $T$-matrix describing the scattering process involving $S$ and a single particle, $T = V\Omega_+$, where $V$ is a particle-system potential and $\Omega_+$ is a Møller operator. $T_\omega(\boldsymbol{p},\boldsymbol{p}')$ denotes the integral kernel corresponding to $T_\omega$ expressed in terms of momenta of the bath particle, $E_p$ the kinetic energy of a particle, and $G(\boldsymbol{p})$ its probability distribution in the momentum space. If $G(\boldsymbol{p}) \sim \exp(-E_p/k_B T)$ and microreversibility conditions, $E_p = E_{-p}$ and $T_\omega(-\boldsymbol{p},-\boldsymbol{p}') = T_\omega(\boldsymbol{p}',\boldsymbol{p})$, hold, then eqn [37] satisfies the quantum detailed-balance condition with the stationary Gibbs state $\rho_\beta$, $\beta = 1/k_B T$.

## Entropy and Purity

The relative entropy $S(\rho\,|\,\sigma) = \mathrm{tr}(\rho\ln\rho - \rho\ln\sigma)$ is monotone with respect to any trace-preserving completely positive map $\Lambda$, that is, $S(\Lambda\rho|\Lambda\sigma) \le S(\rho|\sigma)$. Hence, for the quantum dynamical semigroup $\Lambda(t)$ with the stationary state $\rho_0$ we obtain the following relation for the von Neumann entropy $S(\rho) = -\mathrm{tr}(\rho\ln\rho)$:

$$\frac{d}{dt}S(\rho(t)) = -\frac{d}{dt}S(\rho(t)|\rho_0) - \frac{d}{dt}\mathrm{tr}(\rho(t)\ln\rho_0) \quad [38]$$

where $-(d/dt)S(\rho(t)\,|\,\rho_0) \ge 0$ is an entropy production and the second term describes entropy exchange with environment (Spohn 1980, Alicki and Lendi 1987).

Bistochastic dynamical semigroups preserve the maximally mixed state, that is, $\mathcal{L}(\mathbf{1}) = 0$. For them, the von Neumann entropy does not decrease and the purity $\mathrm{tr}\,\rho^2$ never increases (Streater 1995). Two important classes of master equations, used to describe decoherence, yield bistochastic dynamical semigroups:

$$\frac{d}{dt}\rho(t) = -i[H,\rho(t)] - \sum_j [A_j,[A_j,\rho(t)]],\ \ A_j = A_j^* \quad [39]$$

$$\frac{d}{dt}\rho(t) = -i[H,\rho(t)] + \int_M \mu(d\alpha)(U(\alpha)\rho(t)U^*(\alpha) - \rho(t)) \quad [40]$$

where $U(\alpha)$ are unitary and $\mu(\cdot)$ is a (positive) measure on $M$.

## Itô–Schrödinger Equations

Up to technical problems in the case of unbounded operators, the master equation [4] is completely equivalent to the following stochastic differential equation (in Itô form):

$$d\psi(t) = -iH\psi(t)\,dt - \frac{1}{2}\sum_{j\in I} V_j^* V_j \psi(t)\,dt - i\sum_{j\in I} V_j\psi(t)dX_j(t) \quad [41]$$

where $X_j(t)$ are arbitrary statistically independent stochastic processes with independent increments (continuous or jump processes) such that the expectation $E(dX_j(t)\,dX_k(t)) = \delta_{jk}\,dt$. Equation [41] should be understood as an integral equation involving stochastic Itô integrals with respect to $\{X_j(t)\}$ computed according to the Itô rule: $dX_j(t)\,dX_k(t) = \delta_{jk}\,dt$. Taking the average $\rho(t) = E(|\psi(t)><\psi(t)|)$ one can show, using the Itô rule, that $\rho(t)$ satisfies eqn [4]. For numerical

applications, it is convenient to use the nonlinear version of eqn [41] for the normalized stochastic vector $\phi(t) = \psi(t)/\|\psi(t)\|$, which can be easily derived from eqn [41] (Breurer and Petruccione 2002).

Introducing quantum noises, for example, quantum Brownian motions defined in terms of bosonic or fermionic fields and satisfying suitable quantum Itô rules one can develop the theory of noncommutative stochastic differential equations (NSDE) (Hudson and Parthasarathy 1984). Both, eqn [41] and NSDE, provide examples of unitary dilations – (physically singular) mathematical constructions of the environment $R$ and the $R$–$S$ coupling which exactly reproduce dynamical semigroups as reduced dynamics [34].

## Algebraic Formalism

In order to describe open systems in thermodynamical limit (e.g., infinite spin systems) or systems in the quantum field theory one needs the formalism based on $C^*$ or von Neumann algebras. In the $C^*$-algebraic language, by dynamical semigroup (in the Heisenberg picture) we mean a family $\{T(t); t \geq 0\}$ of linear maps on the unital $C^*$-algebra $\mathcal{A}$ satisfying the following conditions: (1) complete positivity, (2) $T(t)T(s) = T(t+s)$, (3) weak (or strong) continuity, and (4) $T(t)\mathbf{1} = \mathbf{1}$. Assuming the existence of a faithful stationary state $\omega = \omega \circ T(t)$ on $\mathcal{A}$, one can use a Gelfand–Naimark–Segal (GNS) representation $\pi_\omega(\mathcal{A})$ of $\mathcal{A}$ in terms of bounded operators on the suitable Hilbert space $\mathcal{H}_\omega$ with the cyclic and separating vector $\Omega$ satisfying $\omega(A) = <\Omega, \pi_\omega(A)\Omega>$ for all $A \in \mathcal{A}$. Then the dynamical semigroup can be defined on the von Neumann algebra $\mathcal{M}$ (obtained by a weak closure of $\pi_\omega(\mathcal{A})$) as $\hat{T}(t)\pi_\omega(A) \equiv \pi_\omega(T(t)A)$. The Kadison inequality valid even for 2-positive bounded maps $\Lambda$ on $\mathcal{A}$

$$\Lambda(AA^*) \geq \Lambda(A)\Lambda(\mathbf{1})\Lambda(A^*) \qquad [42]$$

implies that $\omega([T(t)A]^* T(t)A) \leq \omega(A^*A)$, which allows one to extend the dynamical semigroup to the contracting semigroup $\tilde{T}(t)[\pi_\omega(A)\Omega] \equiv [\pi_\omega(T(t)A)]\Omega$ on the GNS Hilbert space $\mathcal{H}_\omega$. Typically, one tries to define the semigroup in terms of the proper limiting procedures $T(t) = \lim_{n\to\infty} T_n(t)$, where $T_n(t)$ is well defined on $\mathcal{A}$. However, the limit may not exist as an operator on $\mathcal{A}$ but can be well defined on the von Neumann algebra $\mathcal{M}$. If not, the contracting semigroup on $\mathcal{H}_\omega$ may still be a useful object.

Although there exists a rich ergodic theory of dynamical semigroups for the special types of von Neumann algebras, the most difficult problem of constructing physically relevant semigroups for generic infinite systems remains unsolved (Majewski and Zegarliński 1996, Garbaczewski and Olkiewicz 2002).

## Nonlinear Dynamical Semigroups

The reduced description of many-body classical or quantum systems in terms of single-particle states (probability distributions, wave functions, or density matrices) leads to nonlinear dynamics (e.g., Boltzmann, Vlasov, Hartree, or Hartree–Fock equations) (Spohn 1980, Garbaczewski and Olkiewicz 2002). A large class of nonlinear evolution equations for single-particle density matrices $\rho$ can be written as Alicki and Lendi (1987)

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \mathcal{L}[\rho]\rho \qquad [43]$$

where $\sigma \mapsto \mathcal{L}[\sigma]$ is a map from density matrices to semigroup generators of the type [4]. Under certain technical conditions the solution of eqn [43] exists and defines a nonlinear dynamical semigroup – a family $\{\Gamma(t); t \geq 0\}$ of maps on the set of density matrices satisfying the composition law $\Gamma(t+s) = \Gamma(t)\Gamma(s)$.

A simple example is provided by an open $N$-particle system with the total Hamiltonian invariant with respect to particle permutations. The Markovian approximation combined with the mean-field method leads to a nonlinear dynamical semigroup which preserves purity and for initial pure states is governed by the nonlinear Schrödinger equation with the following structure:

$$\begin{aligned}\frac{\mathrm{d}\psi}{\mathrm{d}t} &= -\mathrm{i}(h + NU(\psi))\psi \\ &\quad + \frac{N}{2}\sum_j \Big( <\psi, V_j^*\psi> V_j\psi \\ &\quad - <\psi, V_j\psi> V_j^*\psi \Big)\end{aligned} \qquad [44]$$

Here $h$ is a single-particle Hamiltonian, $U(\psi)$ a Hartree potential, and $V_j$ are single-particle operators describing collective dissipation.

*See also:* Boltzmann Equation (Classical and Quantum); Channels in Quantum Information Theory; Evolution Equations: Linear and Nonlinear; Kinetic Equations; Nonequilibrium Statistical Mechanics (Stationary): Overview; Positive Maps on $C^*$-Algebras; Quantum Error Correction and Fault Tolerance; Quantum

Mechanical Scattering Theory; Stochastic Differential Equations.

## Further Reading


Alicki R and Fannes M (2001) *Quantum Dynamical Systems.* Oxford: Oxford University Press.

Alicki R and Lendi K (1987) *Quantum Dynamical Semigroups and Applications*. LNP, vol. 286. Berlin: Springer.

Breurer H-P and Petruccione F (2002) *Theory of Open Quantum Systems.* Oxford: Oxford University Press.

Chebotarev AM and Fagnola F (1998) Sufficient conditions for conservativity of quantum dynamical semigroups. *Journal of Functional Analysis* 153: 382–404.

Davies EB (1976) *Quantum Theory of Open Systems.* London: Academic Press.

Frigerio A (1977) Quantum dynamical semigroups and approach to equilibrium. *Letters in Mathematical Physics* 2: 79–87.

Garbaczewski P and Olkiewicz R (eds.) (2002) *Dynamics of Dissipation.* LNP, vol. 597. Berlin: Springer.

Gorini V, Kossakowski A, and Sudarshan ECG (1976) Completely positive dynamical semigroups of *n*-level systems. *Journal of Mathematical Physics* 17: 821–825.

Hudson R and Parthasarathy KR (1984) Quantum Itô's formula and stochastic evolutions. *Communications in Mathematical Physics* 93: 301–323.

Ingarden RS, Kossakowski A, and Ohya M (1997) *Information Dynamics and Open Systems.* Dordrecht: Kluwer.

Lindblad G (1976) On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics* 48: 119–130.

Majewski WA and Zegarliński B (1996) Quantum stochastic dynamics II. *Reviews in Mathematical Physics* 8: 689–713.

Sewell G (2002) *Quantum Mechanics and Its Emergent Macrophysics.* Princeton: Princeton University Press.

Spohn H (1980) Kinetic equations from Hamiltonian dynamics. *Review of Modern Physics* 52: 569–616.

Streater RF (1995) *Statistical Dynamics.* London: Imperial College Press.


# Quantum Dynamics in Loop Quantum Gravity

**H Sahlmann**, Universiteit Utrecht, Utrecht, The Netherlands



## Introduction

In general relativity, the metric is a dynamic entity, there is no preferred notion of time, and the theory is invariant under diffeomorphisms. Therefore, one expects the concept of dynamics to be very different from that in mechanical or special relativistic systems. Indeed, in a canonical formulation, the diffeomorphism symmetry manifests itself through the appearance of constraints (*see* Constrained Systems). In particular, in the absence of boundaries, the Hamiltonian turns out to be a linear combination of them. Thus, the dynamics is completely encoded in the constraints.

To quantize such a system following Dirac, one has to define operators corresponding to the constraints on an auxiliary Hilbert space. Solutions to the quantum dynamics are then vectors that are annihilated by all the constraint operators. Technical complications can arise, and the solutions might not lie in the auxiliary Hilbert space but in an appropriately chosen dual.

Physical observables on the other hand are associated with operators on the auxiliary space that commute with the constraints or, equivalently, operators that act within the space of solutions.

Since the solutions of the quantum dynamics will not depend on any sort of time parameter in an explicit way, they cannot be readily interpreted as a (quantum) spacetime history. The conceptual questions related to this are known as the "problem of time" in quantum gravity.

We should mention that there is a proposal – consistent discretizations – that allows us to eliminate constraints, at the expense of a discretization of the classical theory and dynamical specification of Lagrange multipliers. Application of this technique to gravity is currently under study.

Loop quantum gravity (LQG) (*see* Loop Quantum Gravity) is based on the choice of a canonical pair $(A_a, E^b)$ of an SU(2) connection and an su(2)-valued vector density. The constraints come in three classes:

$$G_i[A,E](x) = 0, \qquad V_a[A,E](x) = 0, \qquad C[A,E](x) = 0$$

the Gauss, vector, and scalar constraints, respectively.

Before giving some detail about the quantization of the constraints and their solutions, we should mention that there exists an analogous classical formulation in terms of complex (self-dual) variables. The quantization in that formulation faces serious technical obstacles, but in the case of positive cosmological constant an elegant formal solution to all the constraints – the Kodama state – is known. It is related to the Chern–Simons action on the spatial slice.

As said before, strictly speaking, implementing the dynamics comprises quantizing and satisfying all the constraints. Here we will however focus on $C$ since it is the most challenging, and most closely related to standard dynamics in that it generates changes under timelike deformations of the Cauchy surface $\Sigma$ on which the canonical formulation is based.

The quantum solutions of the other constraints, linear combinations of $s$-knots, lie in a Hilbert space $\mathcal{K}_{\text{diff}}$ which is part of the dual of the kinematical Hilbert space $\mathcal{K}$ of the theory. For details on these solutions as well as some basic definitions that will be used without comment below (*see* Loop Quantum Gravity). Since $s$-knots are labeled, among other things, by a diffeomorphism equivalence class of a graph, relations to knot theory are emerging at this level (*see* Knot Invariants and Quantum Gravity).

It is important to note that $C$ does not Poisson-commute with the diffeomorphism constraints. Therefore, in the quantum theory it does matter in which order the constraints are solved. It turns out that on the quantum solutions to the other constraints, the scalar constraint can be defined by introducing a regulator, and stays well defined even when the regulator is removed. This ultraviolet finiteness on $\mathcal{K}_{\text{diff}}$ can be intuitively understood from the diffeomorphism invariance of its elements: There is no problematic short-distance regime since the states do not contain any scale at all.

In the following we will briefly review the implementation of the scalar constraint in LQG and comment on some ramifications and open questions.

## The Scalar Constraint Operator

In the Lorentzian theory the scalar constraint $C$ is the sum of the scalar constraint $C^{\text{E}}$ of the Euclidean theory:

$$C^{\text{E}} = (\det q)^{-1/2}\,\text{tr}(F_{ab}[E^a, E^b])$$

a second term of a similar form, but with the curvature $F$ of the connection $A$ replaced by the curvature associated to a certain triad $e$, and possibly matter terms. In the following we will just discuss $C^{\text{E}}$, the other terms can be handled in a similar fashion.

There appear to be a number of obstacles to the quantization of $C^{\text{E}}$: for one, the inverse of the determinant would likely be ill defined, as the volume operator – essentially a quantization of $\int(\det q)^{1/2}$ – has a large kernel. In addition, there are no well-defined operators corresponding to $F$ and $E$ evaluated at points. Rather, only holonomies $h_e[A]$ of $A$ along curves $e$ and certain functionals of $E$ are well defined as operators. These issues can however be dealt with in an elegant way as follows.

The first step is to absorb the determinant factor into a Poisson bracket,

$$C^{\text{E}} = \frac{2}{\kappa}\epsilon^{abc}\,\text{tr}(F_{ab}\{A_c, V\})$$

where $V$ is the volume of the spatial slice $\Sigma$. Then one approximates the curvature by (identity minus) the holonomy around a small loop. In the present case one finds that for a small tetrahedron $\Delta$ with base point $v$, one can approximate

$$\begin{aligned} C^{\text{E}}_{\Delta}(N) &:= 2\kappa^{-1}\int_{\Delta} N\,\text{tr}(F\wedge\{A,V\}) \\ &\approx -\frac{2}{3\kappa}N(v)\epsilon^{ijk}\,\text{tr}(h_{\alpha_{ij}}h_{s_k}\{h_{s_k}^{-1}, V\}) \quad [1] \end{aligned}$$

where (see **Figure 1a**)) the $s_i$ are edges of $\Delta$ incident at $v$ and the $\alpha_{ij}$ loops around the faces of $\Delta$ incident at $v$.

This suggests how to define an operator $\widehat{C}^{\text{E}}_{\Gamma}$ that acts on cylindrical functions on a given graph $\Gamma$: one chooses a triangulation adapted to the graph and quantizes the $C^{\text{E}}_{\Delta}(N)$ (where $\Delta$ is a tetrahedron of this triangulation) using the right-hand side of [1] – holonomies are quantized by the holonomy operators of the quantum theory, $V$ by the volume operator $\widehat{V}$, and the Poisson bracket by the corresponding commutator divided by $i\hbar$. To be more precise, the triangulation is chosen such that the $s_k$ in [1] are part of $\Gamma$, and the operators corresponding to the $h_\alpha$ are creating new edges that connect the endpoints of the $s_k$ (see **Figure 1b**).

Still this is not sufficient, since the definition of $\widehat{C}^{\text{E}}_{\Gamma}$ depends quite heavily on the choice of the triangulation, and there is no natural way to choose one. Furthermore, there is no choice that would guarantee

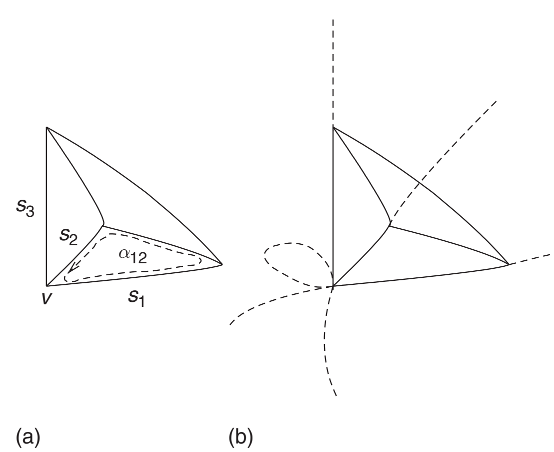



**Figure 1** (a) A tetrahedron $\Delta$ and its labeling of edges and loops. (b) A tetrahedron $\Delta$ adapted to the edges (dashed lines) of a graph $\Gamma$.

that the $\widehat{C}^{E}_{\Gamma}$ for different $\Gamma$ are consistent in the sense that they correspond to the action of the same operator $\widehat{C}^{E}$ on two different cylindrical subspaces. Here, the diffeomorphism invariance of the theory comes to the rescue: a well-defined operator largely free of ambiguities can be obtained by letting the operators above act (by duality) on $\mathcal{K}_{\text{diff}}$ to give elements in $\mathcal{K}^*$. When acting on diffeomorphism-invariant states, the ambiguities in the definition of the triangulations can be eliminated, and the operators $\widehat{C}^{E}_{\Gamma}$ for different $\Gamma$ are consistent and together define an operator $\widehat{C}^{E}(N)$. Roughly speaking, for a diffeomorphism-invariant state, it does not matter anymore where on the graph the endpoints of the $s_k$ lie and how they are connected to form the loops $\alpha$. The final picture looks as follows: for each s-knot $s$, the operator gives a sum of contributions, one for each vertex of $s$, that is, $\widehat{C}^{E}(N)s = \sum_v \widehat{C_v}(N)s$. The terms in this sum are not diffeomorphism invariant. Their evaluation on a spin network $S$ is of the form

$$(\widehat{C_v}s)[S] = \sum_{s'} c(s')N(x(v))s'[S] \qquad [2]$$

where the $s'$ are $s$-knots that differ from $s$ by the addition or deletion of certain edges, and corresponding changes in coloring (by $\pm 1/2$) and intertwiners. As an example, Figure 2 schematically depicts the action on a trivalent vertex. The point $x(v)$ on which $N$ is evaluated in the above formula gets determined as follows: the evaluation $s'[S]$ is zero unless the graph $\Gamma$ on which $S$ is based is an element in the diffeomorphism equivalence class on which $s'$ is based. $x(v)$ is the position of the vertex $v$ in this element of the equivalence class. Because of this $x(v)$, the action of $\widehat{C}^{E}(N)$ is not diffeomorphism invariant.

Similar techniques give a quantization $\widehat{C}$ of the full constraint. The solutions to the constraint can be determined as the vectors $\psi \in \mathcal{K}_{\text{diff}}$ that are annihilated by $\widehat{C}$ in the sense that $(\widehat{C}(N)\psi)[f] = 0$ for all functions $N$ and elements $f$ of $\mathcal{K}$. The solutions are more or less explicitly known; however, the task of interpreting them is a hard one and remains an object of current research.

It should be mentioned that, strictly speaking, one can arrive at several slightly different versions of the

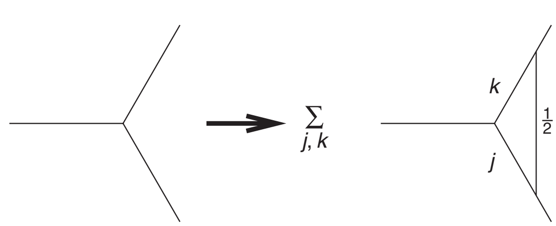



**Figure 2** A schematic rendering of the action of the operator $\widehat{C}_v$ for a trivalent vertex.

constraint operator along the lines sketched above. The quantization ambiguities include changes in the power of the volume operator and the spin quantum number that the constraint creates or annihilates. An interesting check on these quantizations would be to inspect the algebra of constraint operators for anomalies. In the present situation, this can only be carried out to a certain extent, because $\widehat{C}$ is defined on diffeomorphism-invariant states. The Poisson bracket between two scalar constraints is proportional to a diffeomorphism constraint, and indeed it turns out that in the quantum theory the commutator of two scalar constraint operators vanishes for quantizations as described above. In that sense they are ambiguity free; however, this criterion is not strong enough to distinguish between the candidates.

Recently, a slightly different strategy has been proposed, which, if successfully implemented, would eliminate some of the questions regarding the constraint algebra. The idea is to combine the constraints $C(N)$ for different lapse functions $N$ into one master constraint

$$M = \int_{\Sigma} (\det q)^{-1/2} C^2 \, \mathrm{d}^3 x$$

$M$ is manifestly diffeomorphism invariant and could replace all the noncommuting constraints $C(N)$, hence simplifying the constraint algebra considerably.

The interpretation of the solutions of all the constraints hinges on the construction of observables for the theory. This is already a difficult task in the classical theory, and thus even more so after quantization. Though there is no general solution to this problem available, interesting proposals are being studied.

Finally, it should be said that the quantization of the scalar constraint can be used to obtain a picture that resembles more the standard time evolution in quantum field theory. The (formal) power series expansion of the projector

$$P = \prod_{x \in \Sigma} \delta(\widehat{C}(x)) = \int D[N] \exp\left[ \mathrm{i} \int_{\Sigma} N(x)\widehat{C}(x) \right]$$

onto the kernel of $\widehat{C}$ can be described by a spin foam model (*see* Spin Foams).

For further information on the subject of this article see the references: Thiemann (to appear), Rovelli (2004), and Ashtekar and Lewandowski (2004) for general reviews on LQG (with a systematic exposition of a large class of quantizations of the scalar constraint and their solutions in Ashtekar and Lewandowski (2004)); Thiemann (1998) for a seminal work on the quantization of the scalar constraint; Rovelli (1999) and Reisenberger and Rovelli (1997) on the connection to spin foam models; Di Bartolo *et al.* (2002) on

consistent discretizations; Kodama (1990) and Freidel and Smolin (2004) on the Kodama state; and Thiemann (2003) on the master constraint program.

*See also:* Constrained Systems; Knot Invariants and Quantum Gravity; Loop Quantum Gravity; Quantum Geometry and its Applications; Spin Foams; Wheeler–De Witt Theory.

## Further Reading


Ashtekar A and Lewandowski J (2004) Background independent quantum gravity: a status report. *Classical and Quantum Gravity* 21: R53.

Di Bartolo C, Gambini R, and Pullin J (2002) Canonical quantization of constrained theories on discrete space-time lattices. *Classical Quantum Gravity* 19: 5275.

Freidel L and Smolin L (2004) The linearization of the Kodama state. *Classical and Quantum Gravity* 21: 3831.

Kodama H (1990) Holomorphic wave function of the universe. *Physical Review D* 42: 2548.

Reisenberger MP and Rovelli C (1997) "Sum over surfaces" form of loop quantum gravity. *Physical Review D* 56: 3490.

Rovelli C (1999) The projector on physical states in loop quantum gravity. *Physical Review D* 59: 104015.

Rovelli C (2004) *Quantum Gravity*, Cambridge Monographs in Mathematical Physics. Cambridge: Cambridge University Press.

Thiemann T (1998) Quantum spin dynamics (QSD). *Classical and Quantum Gravity* 15: 839.

Thiemann T (2003) The Phoenix project: master constraint programme for loop quantum gravity, arXiv:gr-qc/0305080.

Thiemann T (2006) *Modern Canonical Quantum General Relativity*, Cambridge University Press (to appear).


# Quantum Electrodynamics and Its Precision Tests


**S Laporta**, Università di Parma, Parma, Italy
**E Remiddi**, Università di Bologna, Bologna, Italy




## Introduction

Quantum electrodynamics (QED) describes the interaction of the electromagnetic field (EMF) with charged particles. Any physical particle interacts, directly or indirectly, with any other particle (including itself); in the case of the electron, however, at low and medium energy (say, up to a few GeV) the interaction with the EMF is by and far the most important, so that QED describes with great precision the dynamics of the electron, and at the same time the electron provides with the most stringent tests of QED currently available.

In the various sections of this article we will discuss, in the following order, the origin of QED, the structure of the radiative corrections, the application of QED to various bound states problems (the hydrogen-like atoms, the muonium, and positronium) and the anomalous magnetic moments of the leptons (the muon and the electron).

## Origin of QED

The origin of QED can ideally be traced back to the very beginning of quantum mechanics, the blackbody formula by M Planck (1900), which was soon understood as pointing to a discretization of the energy and momentum associated to the EMF into quanta of light or photons (Einstein 1905).

The quantization of the EMF was first worked out by P Jordan, within the article (1926) by M Born, W Heisenberg, and P Jordan (usually referred to as the *Dreimännerarbeit*) and then in the paper "The quantum theory of emission and absorption of radiation" by PAM Dirac, commonly considered the beginning of the so-called second quantization formalism.

In the subsequent year (1928) Dirac published the famous equation for the relativistic electron, from which it was immediately deduced, on a firmer basis, that the electron has spin 1/2, that its spin gyromagnetic ratio (the ratio between spin and associated magnetic moment in suitable dimensionless units; see below for more details) is twice the value predicted by classical physics (a result expressed as $g_e = 2$) and that the levels of atomic hydrogen with the same principal quantum number $n$ are not fully degenerate, as in the nonrelativistic limit, but do possess the so-called fine structure splitting. In particular, the energy of the $n = 2$ levels splits into two values, one value for $2P_{3/2}$ states with total angular momentum $J = 3/2$ and another value for the states $2S_{1/2}$ and $2P_{1/2}$, which have $J = 1/2$; note that the $2S_{1/2}$ and $2P_{1/2}$ states are still degenerate.

Very soon it was realized that Dirac's equation also requires that each particle must be accompanied by its antiparticle, with exactly the same mass and opposite charge. The antiparticle of the electron, the positron, was indeed discovered by C Anderson

(1932), establishing Dirac's equation as one of the cornerstones of theoretical physics.

All the ingredients needed for the evaluation of the perturbative corrections to the QED theory (usually called radiative corrections) were already present at that moment, but radiative corrections were not systematically investigated for several years, due perhaps to the length and difficulty of the calculations and the absence of important disagreements between theoretical predictions and experimental results.

The situation changed in 1947, when two experiments were carried out, measuring the energy difference between the $2^2S_{1/2}$ and $2^2P_{1/2}$ levels of the hydrogen atom and the gyromagnetic ratios of the electron.

Lamb and Retherford (1947), by using the "great wartime advances in microwaves techniques," succeeded in establishing that in the hydrogen atom "the $2^2S_{1/2}$ state is higher than the $2^2P_{1/2}$ by about 1000 Mc/sec.," while (as observed above) according to the Dirac theory the two states are expected to have exactly the same energy. Subsequent refinements of the experiment (Triebwasser *et al.* 1953) gave for the difference (now referred to as Lamb shift) the value $1057.77 \pm 0.10$ MHz, with a relative error $1 \times 10^{-4}$.

The authors of the second 1947 experiment (Kusch and Foley 1947) measured the frequencies associated with the Zeeman splitting of two different states of gallium, finding an inconsistency with the theoretical values of the gyromagnetic ratios of the electron. More exactly, write the magnetic moments $\boldsymbol{\mu}_L, \boldsymbol{\mu}_S$ associated to the (dimensionless) orbital and spin angular momenta $\boldsymbol{L}, \boldsymbol{S}$ of the electron as

$$\boldsymbol{\mu}_L = -g_L \frac{e\hbar}{2m_e c}\boldsymbol{L}, \quad \boldsymbol{\mu}_S = -g_S \frac{e\hbar}{2m_e c}\boldsymbol{S} \qquad [1]$$

where $(-e)$ is the charge of the electron $(e > 0)$, $m_e$ its mass, $c$ the speed of light and $g_L, g_S$, respectively, the orbital and spin gyromagnetic ratios; the Dirac theory then predicts $g_L = 1$ and $g_S = 2$, while the results of Kusch and Foley (1947) gave a discrepancy which could be accounted for by taking $g_S = 2.00229 \pm 0.00008$ and $g_L = 1$, or alternatively $g_S = 2$ and $g_L = 0.99886 \pm 0.00004$. In modern notation the first conjecture can be rewritten as

$$g_S = g_e = 2(1 + a_e), \quad a_e = 0.001145 \pm 0.00004 \qquad [2]$$

where $a_e$ is the anomalous magnetic moment (or magnetic anomaly) of the electron.

The need of explaining the two experimental results gave rise to a rapid development of covariant perturbation theory (which replaced the previous noncovariant "old fashioned" perturbation theory) and of the renormalization theory, which liberated the perturbative expansion from the divergences plaguing the older approach, opening the path to the evaluation of radiative corrections and to the great success of precision predictions of QED.

The formalism improved quickly, evolving in the more general quantum field theory (QFT) approach; three of the main contributors were Sin-Itiro Tomonaga, Julian Schwinger, and Richard P Feynman, awarded a few years later (1965) the Nobel price "for their fundamental work in quantum electrodynamics, with deep-ploughing consequences for the physics of elementary particles." QFT was then successfully used for describing the weak interactions in the electroweak model and later on also for the strong interactions theory, dubbed quantum chromodynamics (or QCD, in analogy with the popular QED acronym). For more details and references to original works, the reader is invited to look at any treatise on QED or QFT, such as, for instance, Weinberg (1995).

Initially, the Lamb shift was perhaps more important than the electron magnetic anomaly both for the establishment of renormalization theory and as a test of QED, but in the following years it was supplanted by the latter as a precision test of QED.

In 1947 the "best values" for some fundamental constants were indeed

$$\begin{aligned} c &= (2.99776 \pm 0.00004) \times 10^{10}\,\text{cm s}^{-1} \\ R_\infty &= \frac{m_e c^2 \alpha^2}{2hc} = 109737.303 \pm 0.017\,\text{cm}^{-1} \\ 1/\alpha &= 137.030 \pm 0.016 \end{aligned} \qquad [3]$$

where $R_\infty$ is the Rydberg constant for infinite mass, $h$ the Planck constant, and $\alpha$ the fine structure constant (let us observe here in passing that $R_\infty$ was and is still known much better than the separate values of $m_e, \alpha$, and $h$ entering in its definition); for comparison, the current (2005) values for $c$ and $R_\infty$ are

$$\begin{aligned} c &= 299792458\,\text{m s}^{-1} \\ R_\infty &= 109737.31568525(73)\,\text{cm}^{-1} \end{aligned} \qquad [4]$$

where the value of $c$ is exact (it is in fact the definition of the meter), and the relative error in $R_\infty$ is $6.6 \times 10^{-12}$ (the value of $\alpha$ will be discussed later). The measurement of the Lamb shift, repeated several times, gave results in nice agreement with the original value, and for several years it was providing either a test of QED or a precise value for $\alpha$. But the Lamb shift is the energy difference between the metastable level $2S_{1/2}$ (whose lifetime

is about 1/7 s) and the $2P_{1/2}$ level, which has a lifetime of about 1.596 ns or a natural linewidth of 99.7 MHz. Such a large linewidth poses a strong intrinsic limitation to the precision attainable in the measure of the Lamb shift, which is just ten times larger; as a matter of fact, that precision could never reach the $1 \times 10^{-6}$ relative error level, while in the meantime the relative precision in $a_e$ reached the $10^{-9}$ range, replacing the Lamb shift in the role of the leading quantity in high-precision QED.

## The Structure of Radiative Corrections

For obvious space problems we can only superficially sketch here the lines along which the perturbative expansion of QED leading to the evaluation of radiative corrections can be built, considering for simplicity only the photon and the electron. One can start from a QED Lagrangian, formally similar to the classical Lagrangian, involving the electron field and the vector potentials of the electromagnetic (or photon) field. The theory is a gauge theory (its physical content should not change if a gradient is added to the vector potentials); it is further an abelian gauge theory as the EMF does not interact directly with itself.

The QED Lagrangian is separated into a free part and an interaction part. From the free part, one derives the wave functions of the free-particle states and the corresponding time-evolution operators (free Green's functions or propagators; let us just recall here that to obtain a convenient photon propagator one has to break the gauge invariance by adding to the Lagrangian a suitable gauge-breaking term), while the interaction part of the Lagrangian gives the "interaction vertices" of the theory.

Aim of the theory is to build the Green's function for the various processes in the presence of the interaction; from these Green's functions, one then derives all the physical quantities of interest.

With the free propagators and the interaction vertices, one generates the perturbative expansion of the Green's functions. The result, namely the contributions to the perturbative expansion (or radiative corrections), can be depicted in terms of Feynman graphs: they consist of various particle lines joined in the interaction vertices, with external lines corresponding to the initial and final particles and internal lines corresponding to intermediate or virtual particle states. Each graph stands for an integral on the momenta of all the intermediate states, each vertex implying among other things an interaction constant, which is $(-e)$ in the case of electron QED, and a $\delta$-function imposing the conservation of the momenta at that vertex. For each process, the Feynman graphs are naturally classified by the total number of the interaction vertices they contain. In the simplest graphs for a given process (the so-called tree graphs) the $\delta$-functions at the vertices make the integrations trivial; but when the number of vertices increases, closed loops of virtual particle states appear, whose evaluation quickly becomes extremely demanding. In QED, each loop gives an extra factor $(-e)^2$ with respect to the tree graph; it is customary to express it in terms of $(\alpha/\pi) = (e/2\pi)^2$, so that the resulting power of $(\alpha/\pi)$ corresponds to the number of internal loops. The typical QED prediction for a physical quantity is then expressed as a series of powers of the fine structure constant $\alpha$ (and of its logarithm in bound-state problems). As $\alpha$ is small ($\alpha \simeq 1/137$), and the first coefficients of the expansions are usually of the order of 1, a small number of terms in the expansion is in general sufficient to match the precision of the available experimental data.

But the number of different graphs for a given number of loops grows quickly with the number of the loops; in turn, each graph consists in general of a great number of terms and the loop integrations become prohibitively difficult when the number of loops increases, so that the evaluation of radiative corrections proved to be one of the major computational challenges of theoretical physics. As a matter of fact, it prompted the development of computer programs (Veltman 1999) for processing the huge algebraic expressions usually encountered, and of many sophisticated numerical and analytical techniques for performing the loop integrations.

It should be further mentioned here that Feynman graphs written by naively following the above sketched rules are often mathematically ill-defined, taking the form of nonconvergent integrals on the loop momenta. A regularization procedure is needed to give an unambiguous meaning to all the integrals; currently the most powerful regularization is the continuous dimensional regularization scheme, in which the loop integrations are carried out in $d$ continuous dimensions, with $d$ unspecified; renormalization counter-terms are also evaluated in the same scheme, and the physical quantities are recovered in the $d \to 4$ limit (unrenormalized loop integrals and renormalization counterterms are usually singular as powers of $1/(d-4)$ in the $d \to 4$ limit, but all those divergences cancel out in the physical combinations of interest).

QED describes the main interaction of the charged leptons (e, $\mu$, and $\tau$) which have, however, weak interactions as well. Strictly speaking, pure

QED processes do not exist; it is an essential feature of QFT that any existing particle can contribute to the Feynman graphs for any process, when the approximation is pushed to a sufficiently high degree. In particular the photon, which is the main carrier of the QED interaction, is directly coupled also to the strongly interacting particles (the resulting contributions are referred to as "hadronic vacuum polarization" effects).

The precision tests of QED are then to be necessarily searched for in those phenomena where non-QED contributions are presumably small and which involve quantities already well known independently of QED itself. But such high-precision quantities are not always available, and as QED is known better than the rest of physics, very often it is taken to be correct by assumption, and used as a tool for extracting or measuring some of the non-QED quantities relevant to various physical processes.

In any case, as QED predictions are expressed in terms of the fine structure constant $\alpha$, a determination of $\alpha$ independent of QED is needed; without it, the most precise predictions of QED would simply become measures of $\alpha$ and not tests of the theory.

Finally, it is to be recalled that, ironically, the problem of the convergence of the expansion in powers of $\alpha$ is still open, even if it is commonly accepted that convergence problems will matter only for precisions and corresponding perturbative orders (say at order $1/\alpha \simeq 137$) absolutely out of reach of present experimental and computational possibilities, involving further extremely high energies, where the other fundamental interactions are expected to be as important as QED, so that it would be meaningless to consider only QED.

In the following we will discuss only the QED predictions for bound states and the anomalous magnetic moments of $\mu$ and e.

## The Bound States

A very good review of the current status of the theory of hydrogen-like atoms can be found in Eides *et al.* (2001), to which we refer for more details and citation of the original papers. The starting point for studying the bound-state problem in QED is the scattering amplitude of two charged particles, predicted by perturbative QED (pQED) as a (formal) series expansion in powers of $\alpha$. In the static limit $v \to 0$, where $v$ is the relative velocity of the two particles, some of the pQED terms behave as $\alpha/v$, so that the naive expansion in $\alpha$ becomes meaningless. Fortunately, it is relatively easy to identify the origin of those terms (which are essentially due to the Coulomb interaction between the two charges) and to devise techniques for their resummation. Among them, one can quote the Bethe–Salpeter equation, formally very elegant and complete but difficult to use in practice. A great progress has been achieved by the NRQED (nonrelativistic QED) approach, which is a nonrelativistic theory designed to reproduce the full QED scattering amplitude in the nonrelativistic limit by the *ad hoc* definition, *a posteriori*, of a suitable effective Hamiltonian. The Hamiltonian is then divided into a part containing the Coulomb interaction, which is treated exactly and which gives rise to the bound states, and all the rest, to be treated perturbatively. The power of the NRQED approach was further boosted by the continuous dimensional regularization technique of Feynman graph integrals.

Traditionally, the results are expressed in terms of the energies of the bound states, but as in practice the precise measurements concern the transition frequencies between various levels, it is customary to express any energy contribution to some level, say $\Delta E$, also in terms of the associated frequency $\nu = (\Delta E)/h$, where $h$ is the Planck constant.

### The Hydrogen-Like Atoms

Quite in general, a hydrogen-like atom consists of a single electron bound to a positively charge particle, which is a proton for the hydrogen atom, a deuteron nucleus for deuterium, a Helium nucleus for an $He^+$ ion, a $\mu^+$ meson for muonium, or a positron for positronium. Even if QED alone is not sufficient to treat the dynamical properties of the nuclei, their strong interactions can be described by introducing suitable form factors and a few phenomenological parameters; weak interactions could be treated perturbatively, but are not yet required at the precision levels achieved so far.

The QED results for the hydrogen-like atoms can be expressed in terms of the mass $M$ of the positive particle and of its charge $Ze$ (of course $Z = 1$ for hydrogen). When the electron mass $m_e$ is smaller then $M$ (which is always the case, except the positronium case) one can take as a starting point the QED electron moving in the external field of the positive particle, and treat all the other aspects of the relativistic two-body problem (the so-called recoil effects) perturbatively in $m_e/M$.

Neglecting the spin of the positive particle, the energy levels of the hydrogen-like atom are identified by the usual principal quantum number $n$, the orbital angular momentum $l$ (with the convention of writing $S, P, D, \ldots$ instead of $l = 0$, $l = 1$, $l = 2, \ldots$) and $j$, the total angular momentum including the spin of the electron. It turns out that the bound

levels consist of very many contributions of different kinds; dropping quantum number indices for simplicity, the energy levels can be written as an expression of the form

$$E = -\frac{m_e c^2 (Z\alpha)^2}{2}\left(\frac{m_r}{m_e}\right) \times \left[\frac{1}{n^2} + (Z\alpha)^2 f_4 + (Z\alpha)^4 f_6 + \cdots\right] + \Delta E_{rad} + \Delta E_{rec} + \Delta E_{nucl} + \cdots \qquad [5]$$

Let us observe that it is convenient to write explicitly the $Z$ factors even when $Z = 1$ for a better bookkeeping of the various corrections. As usual, $m_r$ is the reduced mass of the electron, $m_r = m_e M/(m_e + M)$ the mass of the nucleus being $M$; the first term in the square bracket, $1/n^2$, the familiar Balmer term, is by and far the dominant one, giving for the $n = 1$ level in the $Z = 1$ case an energy of about 13.6 eV or a corresponding frequency of $3.3 \times 10^{15}$ Hz. The other terms in the square bracket, $f_4$ and $f_6$, are known coefficients (depending also on the small parameter $m_e/M; f_4$ is essentially the fine structure).

The term $\Delta E_{rad}$, is the bulk of the radiative QED corrections; it can be written as a multiple expansion on $(Z\alpha), \alpha$ and $L = \ln[1/(Z\alpha)^2]$, which turns out to have the following explicit form:

$$\Delta E_{rad} = m_e c^2 (Z\alpha)^4 \frac{1}{n^3}\left\{\left(\frac{\alpha}{\pi}\right)\left[A_{41}L + A_{40} + (Z\alpha)A_{50} + (Z\alpha)^2\left(A_{62}L^2 + A_{61}L + A_{60}\right) + \cdots\right] + \left(\frac{\alpha}{\pi}\right)^2\left[B_{40} + (Z\alpha)B_{50} + (Z\alpha)^2\left(B_{63}L^3 + B_{62}L^2 + B_{61}L + B_{60}\right) + \cdots\right] + \left(\frac{\alpha}{\pi}\right)^3[C_{40} + (Z\alpha)C_{50} + \cdots] + \cdots\right\} \qquad [6]$$

The first index of the coefficients refers to the power of $(Z\alpha)$, the second to the power of $L$; as a rule, there are three powers of $(Z\alpha)$ due to the normalization of the wave function and one power of $(Z\alpha)$ for each interaction with the nucleus (in the leading term of eqn [5] one must subtract two powers of $(Z\alpha)$ due to the long-range nature of the Coulomb interaction), while the terms in $L = \ln[1/(Z\alpha)^2]$ are related to the infrared divergences of the scattering amplitude, with the binding energy acting as infrared cutoff. The $A$-coefficients refers to order $(\alpha/\pi)$ or one-loop virtual correction (we do not distinguish here between one-loop self-mass and vacuum-polarization contribution, as usually done in the literature), the $B$-coefficients to two loops, etc. The coefficients are pure numbers, entirely determined within QED, even if their actual calculation is an extremely demanding task. One of the first results obtained in 1947 was $A_{41} = (4/3)\delta_{l0}$, contributing to the $2S$ but not to the $2P$ states (quite in general, most corrections are much bigger for $l = 0$ states than for higher-angular-momentum states), which is sufficient to give the right order of magnitude of the $(2S_{1/2}–2P_{1/2})$ Lamb shift (about 1000 MHz). The other coefficients are now known, thanks to the strenuous and continued efforts (Eides *et al.* 2001) since then, which is impossible to refer properly here in any detail. The current frontier of the theoretical calculation (around the dots in the previous formula) corresponds to 8–9 total powers of $(\alpha/\pi)$ and $(Z\alpha)$ or some kHz for the $1S$ state.

The next term in eqn [5], $\Delta E_{rec}$ contains contributions of order $m_e c^2 (Z\alpha)^5 (m_e/M)$ or smaller (some care must be done for classifying the contributions of order $m_e/M$, which can be accounted for by proper use of $m_r$ rather than $m_e$ and genuine $m_e/M$ contributions), and are sufficiently known for practical purposes; the same is true for many other contributions discussed in Eides *et al.* (2001) and skipped in eqn [5]. A troublesome contribution comes however from $\Delta E_{nucl}$; at leading order, one has

$$\Delta E_{nucl} = \frac{2(Z\alpha)^4 mc^2}{3n^3}\left(\frac{mcR_p}{\hbar}\right)^2 \delta_{l0}$$

where $R_p$ is the so-called root-mean-square charge radius of the proton, which is not well known experimentally (in the literature, there are indeed two direct measurements, $R_p = 0.805(11)$ fm and $R_p = 0.862(12)$ fm, in poor agreement with each other; a new independent measurement is strongly needed).

**The hyperfine splitting** The effect of the interaction of the electron with the spin of the positive particle introduces the so-called hyperfine splitting of all the levels. The order of magnitude of the hyperfine splitting of the $1S$ state is given by the Fermi energy

$$E_F = \frac{4}{3} m_e c^2 (Z\alpha)^4 g_p \frac{m_e}{m_p}$$

where $g_p \simeq 5.586$ is the $g$-factor of the proton, which gives $\simeq 1.42$ GHz. It was dubbed hyperfine because it is smaller than the fine structure terms by the factor $m_e/m_p$. Many classes of corrections can be worked out, with patterns similar to those of the previous subsection, and also in this case the nuclear contributions (this time mainly due to the theoretically unknown magnetic form factor and the

so-called polarizability of the proton) prevent from obtaining predictions with an error less than 1 kHz (or a relative precision better than $1 \times 10^{-6}$).

**The comparison with the experiments** Experimen- Experimentally, one measures transition frequencies among the various levels. For many years the precision record was given by the hyperfine splitting of the ground states of hydrogen $\nu_{\mathrm{hfs}}(1S)$ was measured long ago (see Hellwig *et al.* (1970) and Essen *et al.* (1971)),

$$\nu_{\mathrm{hfs}}(1S) = 1\,420\,405.751\,766\,7(9)\,\mathrm{kHz} \qquad [7]$$

with a relative error $6 \times 10^{-13}$. The current record in the optical range is the value of the $(1S\text{–}2S)$ hydrogen transition frequency, obtained by means of two-photon Doppler-free spectroscopy Niering *et al.* (2000),

$$\nu(1S\text{–}2S) = 2\,466\,061\,413\,187.103(46)\,\mathrm{kHz} \qquad [8]$$

with a relative precision $1.9 \times 10^{-14}$; other optical transitions, such as $(2S\text{–}8D)$, $(2S\text{–}12D)$ are measured with precision of about $1 \times 10^{-11}$.

The measurement of the Lamb shift was repeated several times, with results in nice agreement with the original value, such as Lundeen and Pipkin (1986), 1057.845(9) MHz. The most precise value, $1057.8514 \pm 0.0019$ MHz was given in Palchikov *et al.* (1985) (the result depends, however, on the theoretical value of the lifetime, and should be changed into $1057.8576 \pm 0.0021$ according to subsequent analysis (see Karshenboim (1996)). The experimental $(2S_{1/2}\text{–}2P_{1/2})$ Lamb shift was also obtained as the difference between the measured fine structure separation $(2P_{3/2}\text{–}2S_{1/2})$ and the theoretical value of the $(2P_{3/2}\text{–}2P_{1/2})$ frequency, and the radiative corrections $\Delta E_{\mathrm{rad}}$ to any level are now referred to as the Lamb shift of that level.

As a somewhat deceiving conclusion, the wonderful experimental results of eqns [7] and [8] cannot be used as a high-precision test of the theory or to obtain precise values of many fundamental constants, as the theoretical calculations depend, unfortunately, on hadronic quantities which are not known accurately. Combining theoretical predictions, the above transitions and Lamb shift data, and the available values of $\alpha$ and $m_e/m_p$, one can indeed obtain a measure of $R_p$ ($R_p = 0.883 \pm 0.014$, according to Melnikov and van Ritbergen (2000)) and the value of $R_\infty$ already quoted above.

### Muonium

The muonium is the bound state of a positive $\mu^+$ meson and an electron. At variance with the proton, the $\mu^+$ lepton has no strong interactions, the $\mu^+e^-$ system can be studied theoretically within pure QED, with the weak interactions giving a known and small perturbation. Further, the ratio of the masses $m_e/m_\mu \simeq 4.8 \times 10^{-3}$ is small, so that the external field approximation holds. However, the $\mu$ is unstable (lifetime $\simeq 2.2\,\mu$s), which makes experiments more difficult to carry out. The best measured quantity is the hyperfine splitting of the 1S ground state (see Liu *et al.* (1999))

$$\nu_{\mathrm{hfs}}(\mu e, 1S) = 4\,463\,302\,765(53)\,\mathrm{Hz}$$

with a relative precision of $12 \times 10^{-9}$. The theoretical treatment is similar to the case of hydrogen, with the important advantage that nuclear interactions are absent and everything can be evaluated within QED, so that the bulk of the contribution is given by a formula with the structure of eqn [6]. But the prediction depends, in any case, on the $m_e/m_\mu$ mass, which is not known with the required precision. Indeed, a recent theoretical calculation (Czarnecki *et al.* 2002) (which includes also a contribution of 0.233(3) kHz from hadronic vacuum polarization) gives 4 463 302 680(510) (30)(220) Hz, where the first (and biggest) error comes from $m_e/m_\mu$, the second from $\alpha$, and the third is the theoretical error (an estimation of higher-order contributions not yet evaluated).

### Positronium

The positronium is the bound state of an electron and a positron. Theoretically, it is an ideal system to study, as it can be described entirely within QED, without any unknown parameter of non-QED origin. As the masses of the two constituents, positron and electron, are strictly equal, the reduced mass of the system is exactly equal to half of the electron mass, $m_r = m_e/2$, and the energy scale of the bound states is half of $R_\infty$.

At variance with the muonium case, the external field approximation is not valid, so that positronium must be treated with the full two-body bound-state machinery of QFT, of which it provides an excellent test (Karshenboim 2004).

Experimentally, radioactive positron sources are available, so that positronium is easier to produce than muonium. It is, however, unstable; states with total spin $S$ equal 0 (also called parapositronium states) annihilate into an even number (mainly two) of gammas, and states with $S = 1$ (orthopositronium) into an odd number (mainly three) of gammas, with short lifetimes (which make precise measurements difficult). Further, as positronium is the lightest

atom, Doppler-broadening effects are very important, reducing the precision of spectroscopical measurements.

**Positronium decay rates** There has been a long-time discrepancy between theory and experiment in decay rate of ground-state orthopositronium, which prompted thorough theoretical investigations looking for errors in the calculations or flaws in the formalism, but it turned out that the flaw was on the experimental side. The current theoretical prediction for the ground state $S=1$ decay is (Adkins *et al.* 2002)

$$\Gamma(1S,\text{ortho}) = \Gamma_0\left\{1 + A\left(\frac{\alpha}{\pi}\right) + \frac{1}{3}\alpha^2\ln\alpha + B\left(\frac{\alpha}{\pi}\right)^2 - \frac{3\alpha^3}{2\pi}\ln^2\alpha + C\frac{\alpha^3}{\pi}\ln\alpha + \cdots\right\} = 7.039979(11)\,\mu\text{s}^{-1}$$

where $\Gamma_0 = 2(\pi^2 - 9)m_e\alpha^6/(9\pi) = 7.2111670(1)$, $A = -10.286606(10), B = 45.06(26), C = -5.517$, in nice agreement with the less precise experimental result of Karshenboim (2004, ref. 38) 7.0404 (10)(8) $\mu\text{s}^{-1}$. As a curiosity, the coefficients $A$, $B$ above are among the greatest coefficients so far appeared in QED radiative corrections.

The agreement between theory and experiment for the ground-state parapositronium decay rate has always been good; the current status of Karshenboim (2004, ref. 41) is 7990.9(1.7) $\mu\text{s}^{-1}$ for the experimental result and of Karshenboim (2004, ref. 43) 7989.64(2) $\mu\text{s}^{-1}$ for the theoretical prediction.

**Positronium levels** The quantum number structure of the levels is similar to muonium, with the important difference, however, that the hyperfine splitting (which in hydrogen or muonium is small because it is proportional to the ratio of the masses of the two components) is in fact of the same order as the fine structure. The theoretical evaluation of the energy levels provides a very stringent check of QED and of the overall treatment of the bound-state problem. Corrections have been evaluated, typically, up to order $mc^2\alpha^7$. The best-known quantities are the ground state (hyper)fine splitting, experimental value (Ritter *et al.* 1984) 203.38910(74) GHz ($3.6\times10^{-6}$ relative error), theoretical (Karshenboim 2004) 203.3917(6), and the 1*S*–2*S* transition for orthopositronium, experiment (Fee *et al.* 1993) 1 233 607 216.4 (3.2) MHz, theory 1 233 607 222.2(6). The general agreement is good; the precisions achieved are, however, not yet sufficient to allow a determination of $R_\infty$ or $\alpha$ competitive with other measurements.

## The Anomalous Magnetic Moments of Leptons

The precision of the measurements requires, for both the $e$ and $\mu$ leptons, to also take into account graphs with contributions from the other leptons as virtual intermediate states and those of hadronic and weak origin. Quite in general, if the mass of the virtual particle, say $m_v$, is smaller than the mass of the external lepton, say $m_l$, one can have an $\ln(m_l/m_v)$ behavior of the contributions; that is the case of the virtual electron contributions to the muon magnetic anomaly $a_\mu$, which can be enhanced by powers of $\ln(m_\mu/m_e)$. In the opposite case, $m_v > m_l$, the contribution has the behavior $(m_l/m_v)^2$; that is the case of the $(m_\mu/m_\tau)^2$ contributions to $a_\mu$ from $\tau$ loops and of the $(m_e/m_\mu)^2$ contributions from $\mu$ loops to th electron magnetic anomaly, $a_e$. As strong and weak interactions are in general associated with heavy-mass particles, they are expected to be more important for $a_\mu$ than $a_e$; further, a given heavy particle contribution to $a_e$ is smaller by a factor $(m_e/m_\mu)^2$ than the corresponding contribution to $a_\mu$.

### The Magnetic Anomaly $a_\mu$ of the $\mu$

The $a_\mu$ has been reviewed in Passera (2005). The present (2005) world average experimental value is

$$a_\mu(\text{exp}) = 116\,592\,080(60)\times10^{-11}$$

with a relative error $0.5\times10^{-6}$.

Theoretically, one can write

$$a_\mu = a_\mu(\text{QED}) + a_\mu(\text{had}) + a_\mu(\text{EW}) \qquad [9]$$

where the three terms stand for the contributions from pure QED, strong interacting hadrons and electroweak interactions. In turn, one can expand $a_\mu(\text{QED})$ in powers of $\alpha$ as

$$a_\mu(\text{QED}) = \sum_l C_l\left(\frac{\alpha}{\pi}\right)^l = \sum_l\left[A_1^{(l)} + A_2^{(l)}\left(\frac{m_\mu}{m_e}\right) + A_2^{(l)}\left(\frac{m_\mu}{m_\tau}\right) + A_3^{(l)}\left(\frac{m_\mu}{m_e},\frac{m_\mu}{m_\tau}\right)\right]\left(\frac{\alpha}{\pi}\right)^l \qquad [10]$$

The coefficients $A_1^{(l)}$ involve only the photon and the external lepton as virtual states, are identically the same as in $a_e$; they are known up to $l=4$ included (but, strictly speaking, the contribution of $A_1^{(4)}$ is smaller than the experimental error of $a_\mu$) and will be discussed later for the electron. The $A_2^{(l)}(m_\mu/m_e)$ are very large, being enhanced by powers of $\ln(m_\mu/m_e)$, and are required and known up to $l=5$; $A_2^{(l)}(m_\mu/m_\tau)$ starts with $A_2^{(2)}(m_\mu/m_\tau)\simeq$

$1/45(m_\mu/m_\tau)^2$, contributing $4.2\times10^{-11}$ to $a_\mu$, so that the $A_2^{(l)}(m_\mu/m_\tau)$ with higher values of $l$ are not needed. $A_3^{(l)}(m_\mu/m_e,\ m_\mu/m_\tau)$, finally, starts from $l=3$, and gives a negligible contribution $0.7\times10^{-11}$. Summing up, one finds $C_1=1/2$, $C_2=0.765\,857\,410(27)$ (the error is from the experimental errors in the lepton masses) $C_3=24.050\,509\,64(43)$, $C_4=131.011(8)$, and $C_5=677(40)$. As already observed, the coefficients are large due to the presence of $\ln(m_\mu/m_e)$ factors. The last term $C_5$ contributes $4.6(0.3)\times10^{-11}$ to $a_\mu$, and the total QED contribution is

$$a_\mu(\text{QED}) = 116\,584\,718.8(0.3)(0.4)\times10^{-11}$$

where the first error is due to the uncertainties in the coefficients $C_2, C_3$, and $C_5$ and the second from the value of $\alpha$ coming from atom interferometry measurements (see below).

The hadronic contributions are of two kinds, those due to vacuum polarization, $a_\mu$(vac.pol), which can be evaluated by sound theoretical methods by using existing experimental data, and those due to light-by-light hadronic scattering, $a_\mu$(lbl), whose evaluation relies on much less firmer grounds and are entirely model-dependent. The value of $a_\mu$(vac.pol) varies slightly among the various authors (see Passera (2005) for reference to original work), let us take as a typical value $a_\mu(\text{vac.pol})=6834(92)\times10^{-11}$ (based on $e^+e^-$ scattering data and including also first-order radiative corrections). The model-dependent value of the light-by-light contribution changed several times in the years (also in sign!) but now there is a general consensus that it should be positive; let us take, somewhat arbitrarily, $a_\mu(\text{lbl})=136(25)\times10^{-11}$, so that the total hadronic contribution becomes

$$a_\mu(\text{had}) = 6970(92)\times10^{-11}$$

The electroweak contribution, finally, is

$$a_\mu(\text{EW}) = 154(2)\times10^{-11}$$

which accounts for a one-loop purely weak contribution and a two-loop electromagnetic and weak contribution, which turns out to be very large ($-42\times10^{-11}$) for the presence of logarithms in the masses (the error is due to the uncertainty in the Higgs boson mass).

Summing up, eqn [9] gives $a_\mu=116\,591\,842(92)\times10^{-11}$, so that

$$a_\mu(\text{exp}) - a_\mu = 138(60)(90)\times10^{-11}$$

The substantial agreement can be considered to be a good overall check of QED and electroweak interactions. But another attitude is often adopted in the scientific community: the validity of QED and electroweak models is taken for granted, and a disagreement, if any, is considered to be an indication of new physics. To obtain significant information in that direction, however, the experimental and the theoretical errors (dominated in turn by the experimental error in $e^+e^-$ scattering data) should be significantly reduced.

### The Magnetic Anomaly $a_e$ of the Electron

Experimentally, one has the 1987 value (Kinoshita 2005, ref. 1).

$$a_e(\text{exp}) = 1\,159\,652\,188.4(4.3)\times10^{-12} \qquad [11]$$

with a relative error $3.7\times10^{-9}$ and the preliminary Harvard (2004) measurement (Kinoshita 2005, ref. 3).

$$a_e(\text{Harvard}) = 1\,159\,652\,180.86(0.57)\times10^{-12} \qquad [12]$$

with $0.5\times10^{-9}$ relative error, that is, an increase in precision by a factor 7.

Theoretically, eqns [9] and [10] apply also to the electron; given the smallness of the electron mass, the relevant terms up to the precision of the experimental data are

$$\begin{aligned} a_e &= A_1^{(1)}\left(\frac{\alpha}{\pi}\right) + A_1^{(2)}\left(\frac{\alpha}{\pi}\right)^2 + A_1^{(3)}\left(\frac{\alpha}{\pi}\right)^3 \\ &\quad + A_1^{(4)}\left(\frac{\alpha}{\pi}\right)^4 + \cdots + A_2^{(2)}\left(\frac{m_e}{m_\mu}\right)\left(\frac{\alpha}{\pi}\right)^2 \\ &\quad + a_e(\text{had}) + a_e(\text{EW}) \end{aligned} \qquad [13]$$

The explicit calculation gives

$$A_1^{(1)} = \frac{1}{2} \quad \text{(Passera 2005, ref. 1)}$$

$$\begin{aligned} A_1^{(2)} &= \frac{197}{144} + \frac{1}{12}\pi^2 - \frac{1}{2}\pi^2\ln 2 + \frac{3}{4}\zeta(3) \\ &= -0.328\,478\,965\,579\ldots \\ &\qquad \text{(Passera 2005, ref. 17)} \end{aligned}$$

$$\begin{aligned} A_1^{(3)} &= \frac{83}{72}\pi^2\zeta(3) - \frac{215}{24}\zeta(5) \\ &\quad + \frac{100}{3}\left[a_4 + \frac{1}{24}\ln^4 2 - \frac{1}{24}\pi^2\ln^2 2\right] \\ &\quad - \frac{239}{2160}\pi^4 + \frac{139}{18}\zeta(3) - \frac{298}{9}\pi^2\ln 2 \\ &\quad + \frac{17101}{810}\pi^2 + \frac{28259}{5184} \\ &= 1.181\,241\,456\ldots \quad \text{(Laporta and Remiddi 1996)} \end{aligned}$$

$$A_1^{(4)} = -1.7283(35) \quad \text{(Kinoshita 2005)}$$

and

$$A_2^{(2)}\left(\frac{m_e}{m_\mu}\right)\left(\frac{\alpha}{\pi}\right)^2 \simeq \frac{1}{45}\left(\frac{m_e}{m_\mu}\right)^2\left(\frac{\alpha}{\pi}\right)^2 \simeq 2.72\times 10^{-12}$$
$$a_e(\text{had}) = 1.67(0.02)\times 10^{-12}$$
$$a_e(\text{EW}) = 0.03\times 10^{-12} \qquad [14]$$

For obtaining a meaningful prediction, one needs now a precise value of $\alpha$. The most precise value available at present is that of Passera (2005, ref. 49)

$$\alpha^{-1}(\text{aif}) = 137.036\,000\,3(10)$$

with relative error $7\times 10^{-9}$, obtained by the atom interferometry method (which is independent of QED, depending only on the kinematics of the Doppler effect). With that value of $\alpha$, the theoretical prediction for $a_e$ becomes

$$a_e = 1\,159\,652\,175.9(8.5)(0.1)10^{-12}$$

where the first error comes from $\alpha$ and the second from $C_4$; conversely, one can use the QED prediction for $a_e$ and $a_e$(Harvard) for obtaining $\alpha$; one obtains in that way

$$\alpha^{-1}(\text{QED}, a_e) = 137.035\,999\,708(12)(67)$$

where the first uncertainty is from $C_4$ and the second from the experiment. We see that theory and experiment are in good agreement.

As a concluding remark, another independent and more precise (or analytic!) evaluation of $C_4$ contribution would be welcome. The five-loop term is not known; but as $(\alpha/\pi)^5 \approx 0.07\times 10^{-12}$, if $C_5$ is, say, not greater than 2, its contribution to $a_e$ becomes equal to the contribution of the error $\Delta C_4$ of $C_4$ and is not yet required to match the current precision of $a_e(\exp)$. The ultimate theoretical limit, the error of the hadronic contribution, $\Delta a_e(\text{had}) = 0.02\times 10^{-12}$, is still smaller, corresponding to a change $\Delta C_4 = 0.0007$ of $C_4$ or $\Delta C_5 = 0.3$ of $C_5$.

*See also:* Abelian and Nonabelian Gauge Theories Using Differential Forms; Anomalies; Effective Field Theories; Electroweak Theory; Quantum Field Theory: A Brief Introduction; Standard Model of Particle Physics.

## Further Reading


Adkins GS, Fell RN, and Sapirstein J (2002) Two-loop correction to the orthopositronium decay rate. *Annals of Physics (NY)* 295: 136.

Czarnecki A, Eidelman SI, and Karshenboim SG (2002) Muonium hyperfine structure and hadronic effects. *Physical Review D* 65: 053004.

Dirac PAM (1927) The quantum theory of emission and absorption of radiation. *Proceedings of the Royal Society A* 114: 243.

Eides MI, Grotch H, and Shelyuto VA (2001) Theory of light hydrogenlike atoms. *Physics Reports* 342: 63 (arXiv:hep-ph/0002158).

Essen L, Donaldson RW, Bangman MJ, and Hope EG (1971) Frequency of the hydrogen maser. *Nature* 229: 110.

Fee MS *et al.* Measurement of the positronium $1^3s_1$–$2^3s_1$ interval by continuous-wave two-photon excitation. *Physical Review Letters* 70: 1397.

Hellwig H, Vessot RFC, Levine MW, Zitzewitz PW, Allan DW, and Glaze DJ (1970) Measurement of the unperturbed hydrogen hyperfine transition frequency. *IEEE Transactions on Instrument Measurement* IM19: 200.

Karshenboim SG (1996) Leading logarithmic corrections and uncertainty of muonium hyperfine splitting calculations. *Z. Physics D* 36: 11.

Karshenboim SG (2004) Precision study of positronium: testing bound state QED theory. *International Journal of Modern Physics A* 19: 3879 (e-print hep-ph/0310099).

Kinoshita T (2005) Improved $\alpha^4$ term of the electron anomalous magnetic moment (hep-ph/0507249).

Kusch P and Foley HM (1947) Precision measurements of the ratio of the atomic 'g values' in the $^2P_{3/2}$ and $^2P_{1/2}$ states of gallium. *Physical Review* 72: 1256(L).

Lamb WE Jr. and Retherford RC (1947) Fine structure of the hydrogen atom by a microwave method. *Physical Review* 72: 241.

Laporta S and Remiddi E (1996) The analytical value of the electron $(g-2)$ at order $\alpha^3$ in QED. *Physics Letters B* 379: 283.

Liu W *et al.* (1999) High precision measurements of the ground state hyperfine structure interval of muonium and of the muon magnetic moment. *Physical Review Letters* 82: 711.

Lundeen SR and Pipkin FM (1986) Separated oscillatory field measurement of Lamb shift in atomic hydrogen, $n=2$. *Metrologia* 22(1): 9.

Melnikov K and van Ritbergen T (2000) Three-loop slope of the Dirac form factor and the 1S Lamb shift in hydrogen. *Physical Review Letters* 84: 1673 (hep-ph/9911277).

Niering M *et al.* (2000) Measurement of the hydrogen 1s–2s transition frequency by phase coherent comparison with a microwave cesium fountain clock. *Physics Review Letters* 84: 5496.

Pal'chikov VG, Sokolov YL, and Yakovlev VP (1985) Measurement of the Lamb shift in H, $n=2$. *Metrologia* 21: 99.

Passera M (2005) The standard model prediction of the muon anomalous magnetic moment. *Journal of Physics G* 31: R75 (hep-ph/0411168).

Ritter MW, Egan PO, Hughes VW, and Woodle KA (1984) Precision determination of the hyperfine-structure interval in the ground state of positronium, V. *Physical Review A* 30: 1331.

Triebwasser S, Dayhoff ES, and Lamb WE Jr. (1953) Fine structure of the hydrogen atom, V. *Physical Review* 89: 98.

Veltman MJC (1999) *Nobel Lecture* (http://nobelprize.org/physics/laureates/1999/veltman-lecture.pdf).

Weinberg S (1995) *The Quantum Theory of Fields*. Cambridge: Cambridge University Press.

# Quantum Entropy

**D Petz**, Budapest University of Technology and Economics, Budapest, Hungary



In the past 50 years, entropy has broken out of thermodynamics and statistical mechanics and invaded communication theory, ergodic theory mathematical statistics, and even the social and life sciences. The favorite subjects of entropy concern macroscopic phenomena, irreversibility, and incomplete knowledge. In the strictly mathematical sense entropy is related to the asymptotics of probabilities or concerns the asymptotic behavior of probabilities.

This review is organized as follows. First the history of entropy is discussed generally and then we concentrate on the von Neumann entropy again somewhat historically following the work of von Neumann. Umegaki's quantum relative entropy is discussed both in case of finite systems and in the setting of $C^*$-algebras. An axiomatization is presented. To show physical applications of the concept of entropy, the statistical thermodynamics is reviewed in the setting of spin chains. The relative entropy shows up in the asymptotic theory of hypothesis testing and data compression.

## General Introduction to Entropy: From Clausius to von Neumann

The word "entropy" was created by Rudolf Clausius and it appeared in his work *Abhandlungen über die mechanische Wärmetheorie* published in 1864. The word has a Greek origin, its first part reminds us of "energy" and the second part is from "tropos," which means "turning point." Clausius' work is the foundation stone of classical thermodynamics. According to Clausius, the change of entropy of a system is obtained by adding the small portions of heat quantity received by the system divided by the absolute temperature during the heat absorption. This definition is satisfactory from a mathematical point of view and gives nothing other than an integral in precise mathematical terms. Clausius postulated that the entropy of a closed system cannot decrease, which is generally referred to as the second law of thermodynamics.

The concept of entropy was really clarified by Ludwig Boltzmann. His scientific program was to deal with the mechanical theory of heat in connection with probabilities. Assume that a macroscopic system consists of a large number of microscopic ones, we simply call them particles. Since we have ideas of quantum mechanics in mind, we assume that each of the particles is in one of the energy levels $E_1 < E_2 < \cdots < E_m$. The number of particles in the level $E_i$ is $N_i$, so $\sum_i N_i = N$ is the total number of particles. A macrostate of our system is given by the occupation numbers $N_1, N_2, \ldots, N_m$. The energy of a macrostate is $E = \sum_i N_i E_i$. A given macrostate can be realized by many configurations of the $N$ particles, each of them at a certain energy level $E_i$. These configurations are called microstates. Many microstates realize the same macrostate. We count the number of ways of arranging $N$ particles in $m$ boxes (i.e., energy levels) such that each box has $N_1, N_2, \ldots, N_m$ particles. There are

$$\binom{N}{N_1, N_2, \ldots, N_m} := \frac{N!}{N_1! N_2! \ldots N_m!} \qquad [1]$$

such ways. This multinomial coefficient is the number of microstates realizing the macrostate $(N_1, N_2, \ldots, N_m)$ and it is proportional to the probability of the macrostate if all configurations are assumed to be equally likely. Boltzmann called [1] the thermodynamical probability of the macrostate, in German "thermodynamische Wahrscheinlichkeit," hence the letter $W$ was used. Of course, Boltzmann argued in the framework of classical mechanics and the discrete values of energy came from an approximation procedure with "energy cells."

If we are interested in the thermodynamic limit $N$ increasing to infinity, we use the relative numbers $p_i := N_i/N$ to label a macrostate and, instead of the total energy $E = \sum_i N_i E_i$, we consider the average energy pro particle $E/N = \sum_i p_i E_i$. To find the most probable macrostate, we wish to maximize [1] under a certain constraint. The Stirling approximation of the factorials gives

$$\frac{1}{N}\log\binom{N}{N_1, N_2, \ldots, N_m} = H(p_1, p_2, \ldots, p_m) + O(N^{-1}\log N) \qquad [2]$$

where

$$H(p_1, p_2, \ldots, p_m) := \sum_i -p_i \log p_i \qquad [3]$$

If $N$ is large then the approximation [2] yields that instead of maximizing the quantity [1] we can maximize [3]. For example, maximizing [3] under the constraint $\sum_i p_i E_i = e$, we get

$$p_i = \frac{e^{-\lambda E_i}}{\sum_j e^{-\lambda E_j}} \qquad [4]$$

where the constant $\lambda$ is the solution of the equation

$$\sum_i E_i \frac{e^{-\lambda E_i}}{\sum_j e^{-\lambda E_j}} = e$$

Note that the last equation has a unique solution if $E_1 < e < E_m$, and the distribution [4] is now known as the discrete Maxwell–Boltzmann law.

Let $p_1, p_2, \ldots, p_n$ be the probabilities of different outcomes of a random experiment. According to Shannon, the expression [1] is a measure of our ignorance prior to the experiment. Hence it is also the amount of information gained by performing the experiment. The quantity [1] is maximum when all the $p_i$'s are equal. In information theory, logarithms with base 2 are used and the unit of information is called bit (from binary digit). As will be seen below, an extra factor equal to Boltzmann's constant is included in the physical definition of entropy.

The comprehensive mathematical formalism of quantum mechanics was first presented in the famous book *Mathematische Grundlagen der Quantenmechanik* published in 1932 by Johann von Neumann. In the traditional approach to quantum mechanics, a physical system is described in a Hilbert space: observables correspond to self-adjoint operators and statistical operators are associated with the states. In fact, a statistical operator describes a mixture of pure states. Pure states are really the physical states and they are given by rank-1 statistical operators, or equivalently by rays of the Hilbert space.

von Neumann associated an entropy quantity to a statistical operator in 1927 and the discussion was extended in his book (von Neumann 1932). His argument was a gedanken experiment on the grounds of phenomenological thermodynamics. Let us consider a gas of $N(\gg 1)$ molecules in a box. Suppose that the gas behaves like a quantum system and is described by a statistical operator $\omega$ which is a mixture $\sum_i \lambda_i |\varphi_i\rangle\langle\varphi_i|$, where $|\varphi_i\rangle \equiv \varphi_i$ are orthogonal state vectors. We may take $\lambda_i N$ molecules in the pure state $\varphi_i$ for every $i$. The gedanken experiment gave

$$S\left(\sum_i \lambda_i |\varphi_i\rangle\langle\varphi_i|\right) = \sum_i \lambda_i S(|\varphi_i\rangle\langle\varphi_i|) - \kappa \sum_i \lambda_i \log \lambda_i \qquad [5]$$

where $\kappa$ is Boltzmann's constant and $S$ is certain thermodynamical entropy quantity (relative to the fixed temperature and molecule density).

After this, von Neumann showed that $S(|\varphi\rangle\langle\varphi|)$ is independent of the state vector $|\varphi\rangle$, so that

$$S\left(\sum_i \lambda_i |\varphi_i\rangle\langle\varphi_i|\right) = -\kappa \sum_i \lambda_i \log \lambda_i \qquad [6]$$

up to an additive constant, which could be chosen to be 0 as a matter of normalization. Equation [6] is von Neumann's celebrated entropy formula; it has a more elegant form

$$S(\omega) = \kappa \operatorname{tr} \eta(\omega) \qquad [7]$$

where the state $\omega$ is identified with the corresponding statistical operator, and $\eta : \mathbb{R}^+ \to \mathbb{R}$ is the continuous function $\eta(t) = -t \log t$.

von Neumann solved the maximization problem for $S(\omega)$ under the constraint $\operatorname{tr} \omega H = e$. This means the determination of the ensemble of maximal entropy when the expectation of the energy operator $H$ is a prescribed value $e$. It is convenient to rephrase his argument in terms of conditional expectations. $H = H^*$ is assumed to have a discrete spectrum and we have a conditional expectation $E$ determined by the eigenbasis of $H$. If we pass from an arbitrary statistical operator $\omega$ with $\operatorname{tr} \omega H = e$ to $E(\omega)$, then the entropy is increasing, on the one hand, and the expectation of the energy does not change, on the other, so the maximizer should be searched among the operators commuting with $H$. In this way we are (and von Neumann was) back to the classical problem of statistical mechanics treated at the beginning of this article. In terms of operators, the solution is in the form

$$\frac{\exp(-\beta H)}{\operatorname{tr} \exp(-\beta H)} \qquad [8]$$

which is called Gibbs state today.

## The von Neumann Entropy

von Neumann was aware of the fact that statistical operators form a convex set whose extreme points are exactly the pure states. He also knew that entropy is a concave functional, so

$$S\left(\sum_i \lambda_i \omega_i\right) \geq \sum_i \lambda S(\omega_i) \qquad [9]$$

for any convex combination. To determine the entropy of a statistical operator, he used the Schatten decomposition, which is an orthogonal extremal decomposition in our present language. For a statistical operator $\omega$ there are many ways to write it in the form

$$\omega = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$$

if we do not require the state vectors to be orthogonal. The geometry of the statistical operators, that is, the state space, allows many extremal decompositions and among them there is a unique orthogonal one if the spectrum of $\omega$ is not

degenerate. Nonorthogonal pure states are essentially nonclassical. They are between identical and completely different. Jaynes recognized in 1956 that from the point of view of information the Schatten decomposition is optimal. He proved that

$$S(\omega) = \sup\left\{-\sum_i \lambda_i \log \lambda_i\colon\ \omega = \sum_i \lambda_i \omega_i\right\} \qquad [10]$$

where the supremum is over all convex combinations $\omega = \sum_i \lambda_i \omega_i$ statistical operators. This is Jaynes contribution to the von Neumann entropy. By the way, formula [10] may be used to define von Neumann entropy for states of an arbitrary $C^*$-algebra whose states cannot be described by statistical operators.

Certainly the highlight of quantum entropy theory in the 1970s was the discovery of subadditivity. This property is formulated in a tripartite system whose Hilbert space $\mathcal{H}$ is a tensor product $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$. A statistical operator $\omega_{ABC}$ admits several reduced densities, $\omega_{AB}, \omega_B, \omega_{BC}$, and others. The strong subadditivity is the inequality due to Lieb and Ruskai in 1973:

**Theorem 1**

$$S(\omega_{ABC}) + S(\omega_B) \le S(\omega_{AB}) + S(\omega_{BC}) \qquad [11]$$

The strong subadditivity inequality [11] is conveniently rewritten in terms of the relative entropy. For statistical operators $\rho$ and $\omega$,

$$S(\rho \| \omega) = \operatorname{tr} \rho(\log \rho - \log \omega) \qquad [12]$$

if $\operatorname{supp} \rho \le \operatorname{supp} \omega$, otherwise $S(\rho\|\omega) = +\infty$. The relative entropy expresses statistical distinguishability and therefore it decreases under stochastic mappings:

$$S(\rho\|\omega) \ge S(\mathcal{E}(\rho)\|\mathcal{E}(\omega)) \qquad [13]$$

for a completely positive trace-preserving mapping $\mathcal{E}$.

The strong subadditivity is equivalent to

$$S(\omega_{AB}, \varphi \otimes \omega_B) \le S(\omega_{ABC}, \varphi \otimes \omega_{BC}) \qquad [14]$$

where $\varphi$ is any state on $B(\mathcal{H}_A)$ of finite entropy. This inequality is a consequence of monotonicity of the relative entropy, since $\omega_{AB} = \mathcal{E}(\omega_{ABC})$ and $\varphi \otimes \omega_B = \mathcal{E}(\varphi \otimes \omega_{BC})$, where $\mathcal{E}$ is the partial trace over $\mathcal{H}_C$. Clearly, the equality in [11] is equivalent to equality in [14].

**Theorem 2** *The equality holds in [11] if and only if there is an orthogonal decomposition* $p_B \mathcal{H}_B = \bigoplus_n \mathcal{H}^{\mathrm{L}}_{nB} \otimes \mathcal{H}^{\mathrm{R}}_{nB}, p_B = \operatorname{supp} \omega_B$, *such that the density operator of* $\omega_{ABC}$ *satisfies*

$$\omega_{ABC} = \sum_n \omega_B(p_n) \omega^{\mathrm{L}}_n \otimes \omega^{\mathrm{R}}_n \qquad [15]$$

*where* $\omega^{\mathrm{L}}_n \in B(\mathcal{H}_A) \otimes B(\mathcal{H}^{\mathrm{L}}_{nB})$ *and* $\omega^{\mathrm{R}}_n \in B(\mathcal{H}^{\mathrm{R}}_{nB}) \otimes B(\mathcal{H}_C)$ *are density operators and* $p_n \in B(\mathcal{H}_B)$ *are the orthogonal projections* $\mathcal{H}_B \to \mathcal{H}^{\mathrm{L}}_{nB} \otimes \mathcal{H}^{\mathrm{R}}_{nB}$.

## Quantum Relative Entropy

The quantum relative entropy is an information measure representing the uncertainty of a state with respect to another state. Hence it indicates a kind of distance between the two states. The formal definition [12] is due to Umegaki.

Now we approach quantum relative entropy axiomatically. Our crucial postulate includes the notion of conditional expectation. Let us recall that in the setting of operator algebras conditional expectation (or projection of norm 1) is defined as a positive unital idempotent linear mapping onto a subalgebra.

Now we list the properties of the relative entropy functional which will be used in an axiomatic characterization:

1. *Conditional expectation property*. Assume that $\mathcal{A}$ is a subalgebra of $\mathcal{B}$ and there exists a projection of norm 1 $E$ of $\mathcal{B}$ onto $\mathcal{A}$, such that $\varphi \circ E = \varphi$. Then for every state $\omega$ of $\mathcal{B}$ $S(\omega, \varphi) = S(\omega|\mathcal{A}, \varphi|\mathcal{A}) + S(\omega, \omega \circ E)$ holds.
2. *Invariance property*. For every automorphism $\alpha$ of $\mathcal{B}$ we have $S(\omega, \varphi) = S(\omega \circ \alpha, \varphi \circ \alpha)$.
3. *Direct sum property*. Assume that $\mathcal{B} = \mathcal{B}_1 \oplus \mathcal{B}_2$. Let $\varphi_{12}(a \oplus b) = \lambda\varphi_1(a) + (1-\lambda)\varphi_2(b)$ and $\omega_{12}(a \oplus b) = \lambda\omega_1(a) + (1-\lambda)\omega_2(b)$ for every $a \in \mathcal{B}_1$, $b \in \mathcal{B}_2$ and some $0 < \lambda < 1$. Then $S(\omega_{12}, \varphi_{12}) = \lambda S\ (\omega_1, \varphi_1) + (1-\lambda)S(\omega_2, \varphi_2)$.
4. *Nilpotence property*. $S(\varphi, \varphi) = 0$.
5. *Measurability property*. The function $(\omega, \varphi) \mapsto S(\omega, \varphi)$ is measurable on the state space of the finite dimensional $C^*$-algebra $\mathcal{B}$ (when $\varphi$ is assumed to be faithful).

**Theorem 3** *If a real valued functional* $R(\omega, \varphi)$ *defined for faithful states* $\varphi$ *and arbitrary states* $\omega$ *of finite quantum systems shares the properties [1]–[5], then there exists a constant* $c \in \mathbb{R}$ *such that*

$$R(\omega, \varphi) = c \operatorname{Tr} D_\omega(\log D_\omega - \log D_\varphi)$$

The relative entropy may be defined for linear functionals of an arbitrary $C^*$-algebra. The general definition may go through von Neumann algebras, normal states and the relative modular operator. Another possibility is based on the monotonicity. Let $\omega$ and $\varphi$ be states of a $C^*$-algebra $\mathcal{A}$. Consider finite-dimensional algebras $\mathcal{B}$ and completely positive unital mappings $\alpha: \mathcal{B} \to \mathcal{A}$. Then the supremum

of the relative entropies $S(\omega \circ \alpha \| \varphi \circ \alpha)$ (over all $\alpha$) can be defined as $S(\omega \| \varphi)$.

**Theorem 4** *The relative entropy of states of $C^*$-algebras shares the following properties.*

(i) $(\omega, \varphi) \mapsto S(\omega \| \varphi)$ *is convex and weakly lower-semicontinuous.*
(ii) $\|\varphi - \omega\|^2 \leq 2S(\omega, \varphi)$.
(iii) *For a unital Schwarz map* $\alpha : \mathcal{A}_0 \to \mathcal{A}_1$ *the relation* $S(\omega \circ \alpha \| \varphi \circ \alpha) \leq S(\omega \| \varphi)$ *holds.*

Property (iii) is Uhlmann's monotonicity theorem, which we have already applied above.

The relative entropy appears in many concepts and problems in the area of quantum information theory (Nielsen and Chuang 2000, Schumacher and Westmoreland 2002).

## Statistical Thermodynamics

Let an infinitely extended system of quantum spins be considered in the simple cubic lattice $L = \mathbb{Z}^\nu$, where $\nu$ is a positive integer. The observables confined to a lattice site $x \in \mathbb{Z}^\nu$ form the self-adjoint part of a finite-dimensional $C^*$-algebra $\mathcal{A}_x$ which is a copy of the matrix algebra $M_d(\mathbb{C})$. It is assumed that the local observables in any bounded region $\Lambda \subset \mathbb{Z}^\nu$ are those of the finite quantum system

$$\mathcal{A}_\Lambda = \bigotimes_{x \in \Lambda} \mathcal{A}_x$$

It follows from the definition that for $\Lambda \subset \Lambda'$ we have $\mathcal{A}_{\Lambda'} = \mathcal{A}_\Lambda \otimes \mathcal{A}_{\Lambda' \setminus \Lambda}$, where $\Lambda' \setminus \Lambda$ is the complement of $\Lambda$ in $\Lambda'$. The algebra $\mathcal{A}_\Lambda$ and the subalgebra $\mathcal{A}_\Lambda \otimes \mathbb{C} I_{\Lambda' \setminus \Lambda}$ of $\mathcal{A}_{\Lambda'}$ have identical structure and we identify the element $A \in \mathcal{A}_\Lambda$ with $A \otimes I_{\Lambda' \setminus \Lambda}$ in $\mathcal{A}_{\Lambda'}$. If $\Lambda \subset \Lambda'$ then $\mathcal{A}_\Lambda \subset \mathcal{A}_{\Lambda'}$ and it is said that $\mathcal{A}_\Lambda$ is isotonic with respect to $\Lambda$. The definition also implies that if $\Lambda_1$ and $\Lambda_2$ are disjoint then elements of $\mathcal{A}_{\Lambda_1}$ commute with those of $\mathcal{A}_{\Lambda_2}$. The quasilocal $C^*$-algebra $\mathcal{A}$ is the norm completion of the normed algebra $\mathcal{A}_\infty = \cup_\Lambda \mathcal{A}_\Lambda$, the union of all local algebras $\mathcal{A}_\Lambda$ associated with bounded (finite) regions $\Lambda \subset \mathbb{Z}^\nu$.

We denote by $a_x$ the element of $\mathcal{A}_x$ corresponding to $a \in \mathcal{A}_0 (x \in \mathbb{Z}^\nu)$. It follows from the definition that the algebra $\mathcal{A}_\infty$ consists of linear combinations of terms $a^{(1)}_{x_1} \cdots a^{(k)}_{x_k}$ where $x_1, \ldots, x_k$ and $a^{(1)}, \ldots, a^{(k)}$ run through $\mathbb{Z}^\nu$ and $\mathcal{A}_0$, respectively. We define $\gamma_x$ to be the linear transformation

$$a^{(1)}_{x_1} \cdots a^{(k)}_{x_k} \longmapsto a^{(1)}_{x_1+x} \cdots a^{(k)}_{x_k+x}$$

$\gamma_x$ corresponds to the space translation by $x \in \mathbb{Z}^\nu$ and it extends to an automorphism of $\mathcal{A}$. Hence $\gamma$ is a representation of the abelian group $\mathbb{Z}^\nu$ by automorphisms of the quasilocal algebra $\mathcal{A}$. Clearly, the covariance condition

$$\gamma_x(\mathcal{A}_\Lambda) = \mathcal{A}_{\Lambda+x}$$

holds, where $\Lambda + x$ is the space-translate of the region $\Lambda$ by the displacement $x$.

Having described the kinematical structure of lattice systems, we turn to the dynamics. The local Hamiltonian $H(\Lambda)$ is taken to be the total potential energy between the particles confined to $\Lambda$. This energy may come from many-body interactions of various orders. Most generally, we assume that there exists a global function $\Phi$ such that for any finite subsystem $\Lambda$ the local Hamiltonian takes the form

$$H(\Lambda) = \sum_{X \subset \Lambda} \Phi(X) \qquad [16]$$

Each $\Phi(X)$ represents the interaction energy of the particles in $X$. Mathematically, $\Phi(X)$ is a self-adjoint element of $\mathcal{A}_X$ and $H(\Lambda)$ will be a self-adjoint operator in $\mathcal{A}_\Lambda$. We restrict our discussion to translation-invariant interactions, which satisfy the additional requirement

$$\gamma_x(\Phi(X)) = \Phi(X + x)$$

for every $x \in \mathbb{Z}^\nu$ and every region $X \subset \mathbb{Z}^\nu$. An interaction $\Phi$ is said to be of finite range if $\Phi(\Lambda) = 0$ when the cardinality (or diameter) of $\Lambda$ is large enough, $d(\Lambda) \geq d_\Phi$. The infimum of such numbers is called the range of $\Phi$.

If $\varphi$ is a state of the quasilocal algebra $\mathcal{A}$ then it will induce a state $\varphi_\Lambda$ on $\mathcal{A}(\Lambda)$, the finite system comprising the spin in the bounded region $\Lambda$ of $\mathbb{Z}^\nu$. The (local) energy, entropy, and free energy of this finite system are given by the following formulas:

$$\begin{aligned} E_\Lambda(\varphi) &:= \mathrm{tr}_\Lambda \omega_\Lambda H(\Lambda) \\ S_\Lambda(\varphi) &:= -\mathrm{tr}_\Lambda \omega_\Lambda \log \omega_\Lambda \\ F^\beta_\Lambda(\varphi) &:= E_\Lambda(\varphi) - \frac{1}{\beta} S_\Lambda(\varphi) \end{aligned} \qquad [17]$$

Here $\omega_\Lambda$ denotes the density of $\varphi_\Lambda$ with respect to the trace $\mathrm{tr}_\Lambda$ of $\mathcal{A}_\Lambda$, and $\beta$ denotes the inverse temperature. The functionals $E_\Lambda$, $S_\Lambda$, and $F^\beta_\Lambda$ are termed local. It is rather obvious that all three local functionals are continuous if the weak* topology is considered on the state space of the quasilocal algebra. The energy is affine, the entropy is concave and consequently, the free energy is a convex functional.

The free energy functional $F^\beta_\Lambda$ is minimized by the Gibbs state (see [8] with $H = H(\Lambda)$), and the minimum value is given by

$$-\frac{1}{\beta} \log \mathrm{tr}_\Lambda e^{-\beta H(\Lambda)} \qquad [18]$$

Our aim is to explain this variational principle after the thermodynamic limit is performed.

The thermodynamic limit "$\Lambda$ tends to infinity" may be taken along lattice parallelepipeds. Let $a \in \mathbb{Z}^\nu$ with positive coordinates and define

$$\Lambda(a) = \{x \in \mathbb{Z}^\nu:\ 0 \le x_i < a_i,\ i = 1, 2, \ldots, \nu\} \quad [19]$$

When $a \to \infty$, $\Lambda(a)$ tends to infinity in a manner suitable for the study of thermodynamic limit: the boundary of the parallelepipeds is getting more and more negligible compared with the volume. The notion of limit in the sense of van Hove makes this idea more precise and physically more satisfactory. For the sake of simplicity, we restrict ourselves to thermodynamic limit along parallelepipeds.

Denoting by $|\Lambda|$ the volume of $\Lambda$ (or the number of points in $\Lambda$), we may define the global energy, entropy, and free energy functionals of translationally invariant states to be

$$e(\varphi) := \lim_{\Lambda \to \infty} E_\Lambda(\varphi)/|\Lambda| \quad [20]$$

$$s(\varphi) := \lim_{\Lambda \to \infty} S_\Lambda(\varphi)/|\Lambda| \quad [21]$$

$$f^\beta(\varphi) := \lim_{\Lambda \to \infty} F^\beta_\Lambda(\varphi)/|\Lambda| \quad [22]$$

The existence of the limit in [21] is guaranteed by the strong subadditivity of entropy, while that of the limits in [20] and [22] is assumed if the interaction is suitably tempered, as it certainly does if the interaction is of finite range.

**Theorem 5** *If $\varphi$ is a translationally invariant state of the quasilocal algebra $\mathcal{A}$, then the limit [21] exists and*

$$s(\varphi) = \inf\{S_{\Lambda(a)}(\varphi)/|\Lambda(a)|:\ a \in \mathbb{Z}^\nu_+\} \quad [23]$$

*Moreover, the von Neumann entropy density functional $\varphi \mapsto s(\varphi)$ is affine and upper-semicontinuous when the state space is endowed with the weak* topology.*

*Let $\Phi$ be an interaction of finite range. Then the thermodynamic limit [20] exists and the energy density is given by*

$$e(\varphi) = \varphi(E_\Phi) \quad \text{and} \quad E_\Phi = \sum_{0 \in \Lambda} \frac{\Phi(\Lambda)}{|\Lambda|}$$

*Furthermore, $e(\varphi)$ is an affine weak* continuous functional of $\varphi$.*

*It follows that the free energy density $f(\varphi)$ exists and it is an affine lower-semicontinuous function of the translation-invariant state $\varphi$.*

*For $0 < \beta < \infty$ the thermodynamic limit*

$$\lim_{\Lambda \to \infty} \frac{1}{|\Lambda|} \log \mathrm{tr}_\Lambda \mathrm{e}^{-\beta H(\Lambda)} \equiv p(\beta, \Phi)$$

*exists.*

In accordance with the lattice-gas interpretation of our model, the global quantity $p$ is termed pressure.

In the treatment of quantum spin systems, the set $\mathfrak{S}_\gamma$ of all translation-invariant states is essential. The global entropy functional $s$ is a continuous affine function on $\mathfrak{S}_\gamma$ and physically it is a macroscopic quantity which does not have microscopic (i.e., local) counterpart. Indeed, the local entropy functional is not an observable because it is not affine on the (local) state space. The local internal energy $E_\Lambda(\varphi)$ is microscopic observable and the energy density functional $e$ of $\mathfrak{S}_\gamma$ is the corresponding global extensive quantity.

As an analog of the variational principle for finite quantum systems, the global free-energy functional $f_\beta$ attains an absolute minimum at a translationally invariant state, and the minimum value of $f^\beta$ is equal to the thermodynamic limit of the canonical free-energy densities of the local finite systems. In the next theorem, this global variational principle will be formulated in a slightly different but equivalent way.

**Theorem 6** *When $\Phi$ is an interaction of finite range, then*

$$p(\beta, \Phi) = \sup\{s(\omega) - \beta e(\omega)\}$$

*holds, when the supremum is over all translationally invariant states $\omega$ on $\mathcal{A}$.*

The minimizers of the right-hand side are called *equilibrium states* and they have several different characterizations.

## Asymptotical Properties

We keep the notation of the previous section but we consider one-dimensional chains, $\nu = 1$. Let $\omega$ be translation-invariant state on $\mathcal{A}$ and we fix a positive number $\varepsilon < 1$. We have in our mind that $\varepsilon$ is small and say that a sequence of projection $Q_n \in \mathcal{A}_{[1,n]}$ is of high probability if $\omega(Q_n) \ge 1 - \varepsilon$. The size of $Q_n$, the cardinality of a maximal pairwise orthogonal family of projections contained in $Q_n$, is given by $\mathrm{tr}_n Q_n$. (The subscript $n$ in $\mathrm{tr}_n$ indicates that the algebraic trace functional on $\mathcal{A}_n$ is meant here.) The theorem below says that the entropy density of $\omega$ governs asymptotically the rank of the high-probability projections.

**Theorem 7** *Assume that $\omega$ is an ergodic translation-invariant state of $\mathcal{A}$. Then the limit relation*

$$\lim_{n \to \infty} \frac{1}{n} \inf\{\log \mathrm{tr}_n Q_n\} = s(\omega)$$

*holds, when the infimum is over all projections $Q_n \in \mathcal{A}_{[1,n]}$ such that $\omega_n(Q_n) \ge 1 - \varepsilon$.*

This result is strongly related to data compression. When $\omega$ is interpreted as a stationary quantum source (with possible memory), then efficient and reliable data compression needs a subspace of small dimension and the range of $Q_n$ can play this role. The entropy density is the maximal rate of reliable compression.

It is interesting that one can impose further requirements on the high-probability projections and the statement of the theorem remains true.

1. The partial trace of $Q_{n+1}$ over $\mathcal{A}_{n+1}$ is $Q_n$;
2. $e^{n(s-\varepsilon)} \le \operatorname{tr} Q_n \le e^{n(s+\varepsilon)}$ if $n$ is large enough; and
3. if $q \le Q_n$ is a minimal projection (in $\mathcal{A}_{[1,n]}$), then $\omega(q) \le e^{-n(s-\varepsilon)}$ if $n$ is large enough.

In (2) and (3) $s$ stands for $s(\omega)$. Let $D_n$ be the density matrix of the restriction of $\omega$ to $\mathcal{A}_{[1,n]}$. It follows that for an eigenvalue $\lambda$ of $Q_n D_n Q_n$ the inequality

$$s - \varepsilon \le -\frac{\log \lambda}{n}$$

holds.

From the point of view of data compression, it is important if the sequence $Q_n \in \mathcal{A}_{[1,n]}$ works universally for many states. Indeed, in this case the compression algorithm can be universal for several quantum sources.

**Theorem 8** *Let* $R > 0$. *There is a projection* $Q_n \in \mathcal{A}_{[1,n]}$ *such that*

$$\limsup_n \frac{1}{n} \log \operatorname{tr} Q_n \le R \qquad [24]$$

*and for any ergodic state* $\omega$ *on* $\mathcal{A}$ *such that* $s(\omega) < R$ *the relation*

$$\lim_n \omega(Q_n) = 1 \qquad [25]$$

*holds.*

In the simplest quantum hypothesis testing problem, one has to decide between two states of a system. The state $\rho_0$ is the null hypothesis and $\rho_1$ is the alternative hypothesis. The problem is to decide which hypothesis is true. The decision is performed by a two-valued measurement $\{T, I - T\}$, where $0 \le T \le I$ is an observable. $T$ corresponds to the acceptance of $\rho_0$ and $I - T$ corresponds to the acceptance of $\rho_1$. $T$ is called a test. When the measurement value is 0, the hypothesis $\rho_0$ is accepted, otherwise the alternative hypothesis $\rho_1$ is accepted. The quantity $\alpha[T] = \operatorname{tr}\rho_0(I - T)$ is interpreted as the probability that the null hypothesis is true but the alternative hypothesis is accepted. This is the error of the first kind. Similarly, $\beta[T] = \operatorname{tr}\rho_1 T$ is the probability that the alternative hypothesis is true but the null hypothesis is accepted. It is called the error of the second kind.

Now we fix a formalism for an asymptotic theory of the hypothesis testing. Suppose that a sequence $(\mathcal{H}_n)$ of Hilbert spaces is given, $(\rho_0^{(n)})$ and $(\rho_1^{(n)})$ are density matrices on $\mathcal{H}_n$. The typical example we have in mind is $\rho_0^{(n)} = \rho_0 \otimes \rho_0 \otimes \cdots \otimes \rho_0$ and $\rho_1^{(n)} = \rho_1 \otimes \rho_1 \otimes \cdots \otimes \rho_1$. A positive contraction $T_n \in \mathcal{B}(\mathcal{H}_n)$ is considered as a test on a composite system. Now the errors of the first and second kind depend on $n$: $\alpha_n[T_n] = \operatorname{tr}\rho_0^{(n)}(I - T_n)$ and $\beta_n[T_n] = \operatorname{tr}\rho_1^{(n)} T_n$.

Set

$$\beta^*(n, \varepsilon) = \inf\{\operatorname{tr}\rho_1^{(n)} A_n\} \qquad [26]$$

where the infimum is over all $A_n \in \mathcal{B}(\mathcal{H}_n)$ such that $0 \le A_n \le I$ and $\operatorname{tr}\rho_0^{(n)}(I - A_n) \le \varepsilon$. In other words, this is the infimum of the error of the second kind when the error of the first kind is at most $\varepsilon$. The importance of this quantity is in the customary approach to hypothesis testing.

The following result is the quantum Stein lemma.

**Theorem 9** *In the above setting, the relation*

$$\lim_{n\to\infty} \frac{1}{n} \log \beta^*(n, \varepsilon) = -S(\rho_0 \| \rho_1)$$

*holds for every* $0 < \varepsilon < 1$.

## Bibliographic Notes

A book about entropy in several contexts is Greven *et al.* (2003). The role of von Neumann in the mathematization of the quantum entropy is described in Petz *et al.* (2001); this work contains also his gedanken experiment.

The first proof of the strong subadditivity appeared in Lieb and Ruskai (1973) and a didactical elementary approach is in Nielsen and Petz (2005). The structure of the case of equality was obtained in Hayden *et al.* (2004). The quantum relative entropy was introduced by Umegaki in 1962 and their properties and its axiomatization are contained in the monograph by Ohya and Petz (1993). The monotonicity of the relative entropy is called Uhlmann's theorem, see Uhlmann (1977) and Ohya and Petz (1993).

The rigorous and comprehensive treatment of quantum lattice systems was one of the early successes of the algebraic approach to quantum statistical thermodynamics. The subject is well summarized in Bratteli and Robinson (1981). The book by Sewell (1986) contains more physics and has less in mathematical technicalities. For many interesting entropy results concerning mean field systems, see, for example, Ragio and Werner (1991).

The high probability subspace theorem is due to Ohya and Petz – Petz (1992) and Ohya and Petz (1993) for product states – and was extended to

some algebraic and Gibbs states by Hiai and Petz. The application to data compression was first observed by Schumacher (1995). The chained property of the high-probability subspaces was studied in Bjelaković *et al.* (2003) and the universality is from Kaltchenko and Yang (2003).

A weak form of the quantum Stein lemma was proved in Hiai and Petz (1991) and the stated form is due to Nagaoka and Ogawa (2000). An extension to the case where $\rho_0^{(n)}$ is not a product was given in Bjelaković and Siegmund-Schultze (2004).

Other surveys about quantum entropy are Petz (1992) and Schumacher and Westmoreland (2002).

*See also:* Asymptotic Structure and Conformal Infinity; Capacities Enhanced by Entanglement; Channels in Quantum Information Theory; Entropy and Qualitative Transversality; Positive Maps on $C^*$-Algebras; von Neumann Algebras: Introduction, Modular Theory and Classification Theory; von Neumann Algebras: Subfactor Theory.

## Further Reading


Bjelaković I, Krüger T, Siegmund-Schultze R, and Szkoła A (2003) Chained typical subspaces – a quantum version of Breiman's therem. *Preprint*.

Bjelaković I and Siegmund-Schultze R (2004) An ergodic theorem for the quantum relative entropy. *Communications in Mathematical Physics* 247: 697–712.

Bratteli O and Robinson DW (1981) *Operator Algebras and Quantum Statistical Mechanics. 2. Equilibrium States. Models in Quantum Statistical Mechanics*, Texts and Monographs in Physics, (2nd edn., 1997). Berlin: Springer.

Greven A, Keller G, and Warnecke G (2003) *Entropy*. Princeton and Oxford: Princeton University Press.

Hayden P, Jozsa R, Petz D, and Winter A (2004) Structure of states which satisfy strong subadditivity of quantum entropy with equality. *Communications in Mathematical Physics* 246: 359–374.

Hiai F and Petz D (1991) The proper formula for relative entropy and its asymptotics in quantum probability. *Communications in Mathematical Physics* 143: 99–114.

Kaltchenko A and Yang E-H (2003) Universal compression of ergodic quantum sources. *Quantum Information and Computation* 3: 359–375.

Lieb EH and Ruskai MB (1973) Proof of the strong subadditivity of quantum mechanical entropy. *Journal of Mathematical Physics* 14: 1938–1941.

Nielsen MA and Petz D (2005) A simple proof of the strong subadditivity inequality. *Quantum Information and Computation* 5: 507–513.

Ogawa T and Nagaoka H (2000) Strong converse and Stein's lemma in quantum hypothesis testing. *IEEE Transactions on Information Theory* 46: 2428–2433.

Ohya M and Petz D (1993) *Quantum Entropy and Its Use*, Texts and Monographs in Physics, (2nd edn., 2004). Berlin: Springer.

Petz D (1992) Entropy in quantum probability. In: Accardi L (ed.) *Quantum Probability and Related Topics VII*, pp. 275–297. Singapore: World Scientific.

Petz D (2001) Entropy, von Neumann and the von Neumann entropy. In: Rédei M and Stöltzner M (eds.) *John von Neumann and the Foundations of Quantum Physics*. Dordrecht: Kluwer.

Raggio GA and Werner RF (1991) The Gibbs variational principle for inhomogeneous mean field systems. *Helvetica Physica Acta* 64: 633–667.

Schumacher B (1995) Quantum coding. *Physical Review A* 51: 2738–2747.

Schumacher B and Westmoreland MD (2002) Relative entropy in quantum information theory. In: *Quantum Computation and Information*, (Washington, DC, 2000), Contemp. Math. vol. 305, pp. 265–289. Providence, RI: American Mathematical Society.

Sewell GL (1986) *Quantum Theory of Collective Phenomena*, New York: Clarendon.

Uhlmann A (1977) Relative entropy and the Wigner–Yanase–Dyson–Lieb concavity in an interpolation theory. *Communications in Mathematical Physics* 54: 21–32.

von Neumann J (1932) *Mathematische Grundlagen der Quantenmechanik*. Berlin: Springer. (In English: von Neumann J, *Mathematical Foundations of Quantum Mechanics*. Princeton: Princeton University Press.)


# Quantum Ergodicity and Mixing of Eigenfunctions

**S Zelditch**, Johns Hopkins University, Baltimore, MD, USA



Quantum ergodicity and mixing belong to the field of quantum chaos, which studies quantizations of "chaotic" classical Hamiltonian systems. The basic question is: how does the chaos of the classical dynamics impact on the eigenvalues/eigenfunctions of the quantum Hamiltonian $\hat{H}$ and on long-time dynamics generated by $\hat{H}$?

These problems lie at the foundations of the semiclassical limit, that is, the limit as the Planck constant $\hbar \to 0$ or the energy $E \to \infty$. More generally, one could ask what impact any dynamical feature of a classical mechanical system (e.g., complete integrability, KAM, and ergodicity) has on the eigenfunctions and eigenvalues of the quantization.

Over the last 30 years or so, these questions have been studied rather systematically by both mathematicians and physicists. There is an extensive literature comparing classical and quantum dynamics of model systems, such as comparing the geodesic flow and wave group on a compact (or finite-volume) hyperbolic surface, or comparing classical and quantum billiards on the Sinai billiard or the Bunimovich stadium, or comparing the

discrete dynamical system generated by a hyperbolic torus automorphism and its quantization by the metaplectic representation. As these models indicate, the basic problems and phenomena are richly embodied in simple, low-dimensional examples in much the same way that two-dimensional toy statistical mechanical models already illustrate complex problems on phase transitions. The principles established for simple models should apply to far more complex systems such as atoms and molecules in strong magnetic fields.

The conjectural picture which has emerged from many computer experiments and heuristic arguments on these simple model systems is roughly that there exists a length scale in which quantum chaotic systems exhibit universal behavior. At this length scale, the eigenvalues resemble eigenvalues of random matrices of large size and the eigenfunctions resemble random waves. A small sample of the original physics articles suggesting this picture is Berry (1977), Bohigas *et al.* (1984), Feingold and Peres (1986), and Heller (1984).

This article reviews some of the rigorous mathematical results in quantum chaos, particularly those on eigenfunctions of quantizations of classically ergodic or mixing systems. They support the conjectural picture of random waves up to two moments, that is, on the level of means and variances. A few results also exist on higher moments in very special cases. But from the mathematical point of view, the conjectural links to random matrices or random waves remain very much open at this time. A key difficulty is that the length scale on which universal behavior should occur is far below the resolving power of any known mathematical techniques, even in the simplest model problems. The main evidence for the random matrix and random wave connections comes from numerous computer experiments of model cases in the physics literature. We will not review numerical results here, but to get a well-rounded view of the field, it is important to understand the computer experiments (see, e.g., Bäcker *et al.* (1998a, b) and Barnett (2005)).

The model quantum systems that have been most intensively studied in mathematical quantum chaos are Laplacians or Schrödinger operators on compact (or finite-volume) Riemannian manifolds, with or without boundary, and quantizations of symplectic maps on compact Kähler manifolds. Similar techniques and results apply in both settings, so for the sake of coherence we concentrate on the Laplacian on a compact Riemannian manifold with "chaotic" geodesic flow and only briefly allude to the setting of "quantum maps." Additionally, two main kinds of methods are in use: (1) methods of semiclassical (or microlocal) analysis, which apply to general Laplacians (and more general Schrödinger operators), and (2) methods of number theory and automorphic forms, which apply to arithmetic models such as arithmetic hyperbolic manifolds or quantum cat maps. Arithmetic models are far more "explicitly solvable" than general chaotic systems, and the results obtained for them are far sharper than the results of semiclassical analysis. This article is primarily devoted to the general results on Laplacians obtained by semiclassical analysis; *see* Arithmetic Quantum Chaos for results by J Marklov. For background on semiclassical analysis, see Heller (1984).

## Wave Group and Geodesic Flow

The model quantum Hamiltonians we will discuss are Laplacians $\Delta$ on compact Riemannian manifolds $(M, g)$ (with or without boundary). The classical phase space in this setting is the cotangent bundle $T^*M$ of $M$, equipped with its canonical symplectic form $\sum_i \mathrm{d}x_i \wedge \mathrm{d}\xi_i$. The metric defines the Hamiltonian

$$H(x, \xi) = |\xi|_g = \sqrt{\sum_{ij=1}^{n} g^{ij}(x)\xi_i\xi_j}$$

on $T^*M$, where

$$g_{ij} = g\left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\right)$$

$[g^{ij}]$ is the inverse matrix to $[g_{ij}]$. We denote the volume density of $(M, g)$ by dVol and the corresponding inner product on $L^2(M)$ by $\langle f, g\rangle$. The unit (co-) ball bundle is denoted $B^*M = \{(x, \xi) : |\xi| \leq 1\}$.

The Hamiltonian flow $\Phi^t$ of $H$ is the geodesic flow. By definition, $\Phi^t(x, \xi) = (x_t, \xi_t)$, where $(x_t, \xi_t)$ is the terminal tangent vector at time $t$ of the unit speed geodesic starting at $x$ in the direction $\xi$. Here and below, we often identify $T^*M$ with the tangent bundle $TM$ using the metric to simplify the geometric description. The geodesic flow preserves the energy surfaces $\{H = E\}$ which are the co-sphere bundles $S^*_E M$. Due to the homogeneity of $H$, the flow on any energy surface $\{H = E\}$ is equivalent to that on the co-sphere bundle $S^*M = \{H = 1\}$. (This homogeneity could be broken by adding a potential $V \in C^\infty(M)$ to form a semiclassical Schrödinger operator $-\hbar^2\Delta + V$, whose underlying Hamiltonian flow is generated by $|\xi|_g^2 + V(x)$.) *See* h-Pseudodifferential Operators and Applications.

The quantization of the Hamiltonian $H$ is the square root $\sqrt{\Delta}$ of the positive Laplacian

$$\Delta = -\frac{1}{\sqrt{g}}\sum_{i,j=1}^{n}\frac{\partial}{\partial x_i}g^{ij}g\frac{\partial}{\partial x_j}$$

of $(M,g)$. Here, $g = \det[g_{ij}]$. We choose to work with $\sqrt{\Delta}$ rather than $\Delta$ since the former generates the wave

$$U_t = \mathrm{e}^{\mathrm{i}t\sqrt{\Delta}}$$

which is the quantization of the geodesic flow $\Phi^t$.

By the last statement we mean that $U_t$ is related to $\Phi^t$ in several essentially equivalent ways:

1. singularities of waves, that is, solutions $U_t\psi$ of the wave equation, propagate along geodesics;
2. $U_t$ is a Fourier integral operator (= quantum map) associated to the canonical relation defined by the graph of $\Phi^t$ in $T^*M \times T^*M$; and
3. Egorov's theorem holds.

We only define the latter since it plays an important role in studying eigenfunctions. As with any quantum theory, there is an algebra of observables on the Hilbert space $L^2(M, \mathrm{dvol}_g)$ which quantizes $T^*M$. Here, $\mathrm{dvol}_g$ is the volume form of the metric. The algebra is that $\Psi^*(M)$ of pseudodifferential operators $\psi DO$'s of all orders, though we often restrict to the subalgebra $\Psi^0$ of $\psi\, DO$'s of order zero. We denote by $\Psi^m(M)$ the subspace of pseudodifferential operators of order $m$. The algebra is defined by constructing a quantization Op from an algebra of symbols $a \in S^m(T^*M)$ of order $m$ (polyhomogeneous functions on $T^*M\backslash 0$) to $\Psi^m$. The map Op is not unique. In the reverse direction is the symbol map $\sigma_A : \Psi^m \to S^m(T^*M)$ which takes an operator $\mathrm{Op}(a)$ to the homogeneous term $a_m$ of order $m$ in $a$.

Egorov's theorem for the wave group concerns the conjugations

$$\alpha_t(A) := U_t A U_t^*, \quad A \in \Psi^m(M) \qquad [1]$$

Such a conjugation defines the quantum evolution of observables in the Heisenberg picture, and, since the early days of quantum mechanics, it was known to correspond to the classical evolution

$$V_t(a) := a \circ \Phi^t \qquad [2]$$

of observables $a \in C^\infty(S^*M)$. Egorov's theorem is the rigorous version of this correspondence: it states that $\alpha_t$ defines an order-preserving automorphism of $\Psi^*(M)$, that is, $\alpha_t(A) \in \Psi^m(M)$ if $A \in \Psi^m(M)$, and that

$$\sigma_{U_t A U_t^*}(x,\xi) = \sigma_A(\Phi^t(x,\xi)) := V_t(\sigma_A), \quad (x,\xi) \in T^*M\backslash 0 \qquad [3]$$

This formula is almost universally taken to be the definition of quantization of a flow or map in the physics literature.

The key difficulty in quantum chaos is that it involves a comparison between long-time dynamical properties of $\Phi^t$ and $U_t$ through the symbol map and similar classical limits. The classical dynamics defines the "principal symbol" behavior of $U_t$ and the "error" $U_t A U_t^* - Op(\sigma_A \circ \Phi^t)$ typically grows exponentially in time. This is just the first example of a ubiquitous "exponential barrier" in the subject.

## Eigenvalues and Eigenfunctions of $\Delta$

The eigenvalue problem on a compact Riemannian manifold

$$\Delta\varphi_j = \lambda_j^2\varphi_j, \quad \langle\varphi_j, \varphi_k\rangle = \delta_{jk}$$

is dual under the Fourier transform to the wave equation. Here, $\{\varphi_j\}$ is a choice of orthonormal basis of eigenfunctions, which is not unique if the eigenvalues have multiplicities $>1$. The individual eigenfunctions are difficult to study directly, and so one generally forms the spectral projections kernel,

$$E(\lambda, x, y) = \sum_{j:\lambda_j \le \lambda}\varphi_j(x)\varphi_j(y) \qquad [4]$$

Semiclassical asymptotics is the study of the $\lambda \to \infty$ limit of the spectral data $\{\varphi_j, \lambda_j\}$ or of $E(\lambda, x, y)$. The (Schwartz) kernel of the wave group can be represented in terms of the spectral data by

$$U_t(x,y) = \sum_j \mathrm{e}^{\mathrm{i}t\lambda_j}\varphi_j(x)\varphi_j(y)$$

or equivalently as the Fourier transform $\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}t\lambda}\,\mathrm{d}E(\lambda, x, y)$ of the spectral projections. Hence, spectral asymptotics is often studied through the large-time behavior of the wave group.

The link between spectral theory and geometry, and the source of Egorov's theorem for the wave group, is the construction of a parametrix (or WKB formula) for the wave kernel. For small times $t$, the simplest is the Hadamard parametrix,

$$U_t(x,y) \sim \int_0^\infty \mathrm{e}^{\mathrm{i}\theta(r^2(x,y)-t^2)}\sum_{k=0}^\infty U_k(x,y)\theta^{((d-3)/2)-k}\,\mathrm{d}\theta \quad (t < \mathrm{inj}(M,g)) \qquad [5]$$

where $r(x,y)$ is the distance between points, $U_0(x,y) = \Theta^{-1/2}(x,y)$ is the volume 1/2-density, $\mathrm{inj}(M,g)$ is the injectivity radius, and the higher Hadamard coefficients are obtained by solving transport equations along geodesics. The parametrix is asymptotic to the wave kernel in the sense of

smoothness, that is, the difference of the two sides of [5] is smooth. The relation [5] may be iterated using $U_{tm} = U_t^m$ to obtain a parametrix for long times. This is obviously complicated and not necessarily the best long-time parametrix construction, but it illustrates again the difficulty of a long-time analysis.

### Weyl Law and Local Weyl Law

A fundamental and classical result in spectral asymptotics is Weyl's law on counting eigenvalues:

$$N(\lambda) = \#\{j : \lambda_j \leq \lambda\} = \frac{|B_n|}{(2\pi)^n}\mathrm{Vol}(M, g)\lambda^n + O(\lambda^{n-1}) \qquad [6]$$

Here, $|B_n|$ is the Euclidean volume of the unit ball and $\mathrm{Vol}(M, g)$ is the volume of $M$ with respect to the metric $g$. An equivalent formula which emphasizes the correspondence between classical and quantum mechanics is

$$\mathrm{tr}E_\lambda = \frac{\mathrm{Vol}(|\xi|_g \leq \lambda)}{(2\pi)^n} \qquad [7]$$

where Vol is the symplectic volume measure relative to the natural symplectic form $\sum_{j=1}^n \mathrm{d}x_j \wedge \mathrm{d}\xi_j$ on $T^*M$. Thus, the dimension of the space where $H = \sqrt{\Delta}$ is $\leq \lambda$ is asymptotically the volume where its symbol $|\xi|_g \leq \lambda$.

The remainder term in Weyl's law is sharp on the standard sphere, where all geodesics are periodic, but is not sharp on $(M, g)$ for which the set of periodic geodesics has measure zero (Duistermaat–Guillemin, Ivrii) (*see* Semiclassical Spectra and Closed Orbits). When the set of periodic geodesics has measure zero (as is the case for ergodic systems), one has

$$N(\lambda) = \#\{j : \lambda_j \leq \lambda\} = \frac{|B_n|}{(2\pi)^n}\mathrm{Vol}(M, g)\lambda^n + o(\lambda^{n-1}) \qquad [8]$$

The remainder is then of smaller order than the derivative of the principal term, and one then has asymptotics in shorter intervals:

$$N([\lambda, \lambda+1]) = \#\{j : \lambda_j \in [\lambda, \lambda+1]\} = n\frac{|B_n|}{(2\pi)^n}\mathrm{Vol}(M, g)\lambda^{n-1} + o(\lambda^{n-1}) \qquad [9]$$

Physicists tend to write $\lambda \sim \hbar^{-1}$ and to average over intervals of this width. Then mean spacing between the eigenvalues in this interval is $\sim C_n \mathrm{Vol}(M, g)^{-1} \times \lambda^{-(n-1)}$, where $C_n$ is a constant depending on the dimension.

An important generalization is the "local Weyl law" concerning the traces $\mathrm{tr}AE(\lambda)$, where $A \in \Psi^m(M)$. It asserts that

$$\sum_{\lambda_j \leq \lambda} \langle A\varphi_j, \varphi_j \rangle = \frac{1}{(2\pi)^n}\int_{B^*M} \sigma_A \mathrm{d}x\, \mathrm{d}\xi\ \lambda^n + O(\lambda^{n-1}) \qquad [10]$$

There is also a pointwise local Weyl law:

$$\sum_{\lambda_j \leq \lambda} |\varphi_j(x)|^2 = \frac{1}{(2\pi)^n}|B^n|\lambda^n + R(\lambda, x) \qquad [11]$$

where $R(\lambda, x) = O(\lambda^{n-1})$ uniformly in $x$. Again, when the periodic geodesics form a set of measure zero in $S^*M$, one could average over the shorter interval $[\lambda, \lambda+1]$. Combining the Weyl and local Weyl law, we find the surface average of $\sigma_A$ is a limit of traces:

$$\omega(A) := \frac{1}{\mu(S^*M)}\int_{S^*M} \sigma_A\, \mathrm{d}\mu = \lim_{\lambda \to \infty} \frac{1}{N(\lambda)} \sum_{\lambda_j \leq \lambda} \langle A\varphi_j, \varphi_j \rangle \qquad [12]$$

Here, $\mu$ is the "Liouville measure" on $S^*M$, that is, the surface measure $\mathrm{d}\mu = \mathrm{d}x\, \mathrm{d}\xi / \mathrm{d}H$ induced by the Hamiltonian $H = |\xi|_g$ and by the symplectic volume measure $\mathrm{d}x\, \mathrm{d}\xi$ on $T^*M$.

### Problems on Asymptotics Eigenfunctions

Eigenfunctions arise in quantum mechanics as stationary states, that is, states $\psi$ for which the probability measure $|\psi(t, x)|^2 \mathrm{dvol}$ is constant in time where $\psi(t, x) = U_t \psi(x)$ is the evolving state. This follows from the fact that

$$U_t \varphi_k = \mathrm{e}^{\mathrm{i}t\lambda_k} \varphi_k \qquad [13]$$

and that $|\mathrm{e}^{\mathrm{i}t\lambda_k}| = 1$. They are the basic modes of the quantum system. One would like to know the behavior as $\lambda_j \to \infty$ (or $\hbar \to 0$ in the semiclassical setting) of invariants such as:

1. matrix elements $\langle A\varphi_j, \varphi_j \rangle$ of observables in this state;
2. transition elements $\langle A\varphi_i, \varphi_j \rangle$ between states;
3. size properties as measured by $L^p$ norms $\|\varphi_j\|_{L^p}$;
4. value distribution as measured by the distribution function $\mathrm{Vol}\{x \in M : |\varphi_j(x)|^2 > t\}$; and
5. shape properties, for example, distribution of zeros and critical points of $\varphi_j$.

Let us introduce some problems which have motivated much of the work in this area.

**Problem 1** Let $\mathcal{Q}$ denote the set of "quantum limits," that is, weak* limit points of the sequence $\{\Phi_k\}$ of distributions on the classical phase space $S^*M$, defined by

$$\int_X a\, \mathrm{d}\Phi_k := \langle Op(a)\varphi_k, \varphi_k\rangle$$

where $a \in C^\infty(S^*M)$.

The set $\mathcal{Q}$ is independent of the definition of Op. It follows almost immediately from Egorov's theorem that $\mathcal{Q} \subset \mathcal{M}_I$, where $\mathcal{M}_I$ is the convex set of invariant probability measures for the geodesic flow. Furthermore, they are time-reversal invariant, that is, invariant under $(x, \xi) \to (x, -\xi)$ since the eigenfunctions are real valued.

To see this, it is helpful to introduce the linear functionals on $\Psi^0$:

$$\rho_k(A) = \langle Op(a)\varphi_k, \varphi_k\rangle \qquad [14]$$

We observe that $\rho_k(I) = 1$, $\rho_k(A) \geq 0$ if $A \geq 0$, and that

$$\rho_k\left(U_t A U_t^*\right) = \rho_k(A) \qquad [15]$$

Indeed, if $A \geq 0$ then $A = B^*B$ for some $B \in \Psi^0$ and we can move $B^*$ to the right-hand side. Similarly, [15] is proved by moving $U_t$ to the right-hand side and using [13]. These properties mean that $\rho_j$ is an "invariant state" on the algebra $\Psi^0$. More precisely, one should take the closure of $\Psi^0$ in the operator norm. An invariant state is the analog in quantum statistical mechanics of an invariant probability measure.

The next important fact about the states $\rho_k$ is that any weak limit of the sequence $\{\rho_k\}$ on $\Psi^0$ is an invariant probability measure on $C(S^*M)$, that is, a positive linear functional on $C(S^*M)$ rather than just a state on $\Psi^0$. This follows from the fact that $\langle K\varphi_j, \varphi_j\rangle \to 0$ for any compact operator $K$, and so any limit of $\langle A\varphi_k, \varphi_k\rangle$ is equally a limit of $\langle (A+K)\varphi_k, \varphi_k\rangle$. Hence, any limit is bounded by $\inf_K \|A+K\|$ (the infimum taken over compact operators), and for any $A \in \Psi^0, \|\sigma_A\|_{L^\infty} = \inf_K \|A+K\|$. Hence, any weak limit is bounded by a constant times $\|\sigma_A\|_{L^\infty}$ and is therefore continuous on $C(S^*M)$. It is a positive functional since each $\rho_j$, and hence any limit, is a probability measure. By Egorov's theorem and the invariance of the $\rho_k$, any limit of $\rho_k(A)$ is a limit of $\rho_k(Op(\sigma_A \circ \Phi^t))$ and hence the limit measure is invariant.

Problem 1 is thus to identify which invariant measures in $\mathcal{M}_I$ show up as weak limits of the functionals $\rho_k$ or equivalently the distributions $\mathrm{d}\Phi_k$. The weak limits reflect the concentration and oscillation properties of eigenfunctions. Here are some possibilities:

1. Normalized Liouville measure. In fact, the functional $\omega$ of [12] is also a state on $\Psi^0$ for the reason explained above. A subsequence $\{\varphi_{k_j}\}$ of eigenfunctions is considered diffuse if $\rho_{k_j} \to \omega$.
2. A periodic orbit measure $\mu_\gamma$ defined by

$$\mu_\gamma(A) = \frac{1}{L_\gamma}\int_\gamma \sigma_A\, \mathrm{d}s$$

where $L_\gamma$ is the length of $\gamma$. A sequence of eigenfunctions for which $\rho_{k_j} \to \mu_\gamma$ obviously concentrates (or strongly "scars") on the closed geodesic.
3. A finite sum of periodic orbit measures.
4. A delta-function along an invariant Lagrangian manifold $\Lambda \subset S^*M$. The associated eigenfunctions are viewed as "localizing" along $\Lambda$.
5. A more general invariant measure which is singular with respect to $\mathrm{d}\mu$.

All of these possibilities can and do happen in different examples. If $\mathrm{d}\Phi_{k_j} \to \omega$, then in particular we have

$$\frac{1}{\mathrm{Vol}(M)}\int_E |\varphi_{k_j}(x)|^2 \mathrm{dVol} \to \frac{\mathrm{Vol}(E)}{\mathrm{Vol}(M)}$$

for any measurable set $E$ whose boundary has measure zero. Interpreting $|\varphi_{k_j}(x)|^2\mathrm{dVol}$ as the probability density of finding a particle of energy $\lambda_k^2$ at $x$, this result means that the sequence of probabilities tends to uniform measure.

However, $\mathrm{d}\Phi_{k_j} \to \omega$ is much stronger since it says that the eigenfunctions become diffuse on the energy surface $S^*M$ and not just on the configuration space $M$. As an example, consider the flat torus $\mathbb{R}^n/\mathbb{Z}^n$. An orthonormal basis of eigenfunctions is furnished by the standard exponentials $\mathrm{e}^{2\pi\mathrm{i}\langle k, x\rangle}$ with $k \in \mathbb{Z}^n$. Obviously, $|\mathrm{e}^{2\pi\mathrm{i}\langle k, x\rangle}|^2 = 1$, so the eigenfunctions are already diffuse in configuration space. On the other hand, they are far from diffuse in phase space, and localize on invariant Lagrangian tori in $S^*M$. Indeed, by definition of pseudodifferential operator, $A\mathrm{e}^{2\pi\mathrm{i}\langle k, x\rangle} = a(x,k)\,\mathrm{e}^{2\pi\mathrm{i}\langle k, x\rangle}$, where $a(x,k)$ is the complete symbol. Thus,

$$\langle A\mathrm{e}^{2\pi\mathrm{i}\langle k,x\rangle}, \mathrm{e}^{2\pi\mathrm{i}\langle k,x\rangle}\rangle = \int_{\mathbb{R}^n/\mathbb{Z}^n} a(x,k)\,\mathrm{d}x \sim \int_{\mathbb{R}^n/\mathbb{Z}^n} \sigma_A\left(x, \frac{k}{|k|}\right)\mathrm{d}x$$

A subsequence $\mathrm{e}^{2\pi\mathrm{i}\langle k_j, x\rangle}$ of eigenfunctions has a weak limit if and only if $k_j/|k_j|$ tends to a limit vector $\xi_0$ in the unit sphere in $\mathbb{R}^n$. In this case, the associated

weak* limit is $\int_{\mathbb{R}^n/\mathbb{Z}^n} \sigma_A(x, \xi_0)\mathrm{d}x$, that is, the delta-function on the invariant torus $T_{\xi_0} \subset S^*M$ defined by the constant momentum condition $\xi = \xi_0$. The eigenfunctions are said to localize on this invariant torus for $\Phi^t$.

The flat torus is a model of a completely integrable system on both the classical and quantum levels. Another example is that of the standard round sphere $S^n$. In this case, the author and D Jakobson showed that absolutely any invariant measure $\nu \in \mathcal{M}_I$ can arise as a weak limit of a sequence of eigenfunctions. This reflects the huge degeneracy (multiplicities) of the eigenvalues.

On the other hand, if the geodesic flow is ergodic, one would expect the eigenfunctions to be diffuse in phase space. In the next section, we will discuss the rigorous results on this problem.

Off-diagonal matrix elements

$$\rho_{jk}(A) = \langle A\varphi_i, \varphi_j \rangle \qquad [16]$$

are also important as transition amplitudes between states. They no longer define states since $\rho_{jk}(I) = 0$, are positive, or invariant. Indeed, $\rho_{jk}(U_t A U_t^*) = \mathrm{e}^{\mathrm{i}t(\lambda_j - \lambda_k)}\rho_{jk}(A)$, so they are eigenvectors of the automorphism $\alpha_t$ of [1]. A sequence of such matrix elements cannot have a weak limit unless the spectral gap $\lambda_j - \lambda_k$ tends to a limit $\tau \in \mathbb{R}$. In this case, by the same discussion as above, any weak limit of the functionals $\rho_{jk}$ will be an eigenmeasure of the geodesic flow which transforms by $\mathrm{e}^{\mathrm{i}\tau t}$ under the action of $\Phi^t$. Examples of such eigenmeasures are orbital Fourier coefficients

$$\frac{1}{L_\gamma} \int_0^{L_\gamma} \mathrm{e}^{-\mathrm{i}\tau t} \sigma_A(\Phi^t(x, \xi))\,\mathrm{d}t$$

along a periodic orbit. Here, $\tau \in (2\pi/L_\gamma)\mathbb{Z}$. We denote by $\mathcal{Q}_\tau$ such eigenmeasures of the geodesic flow. Problem 1 has the following extension to off-diagonal elements:

**Problem 2** Determine the set $\mathcal{Q}_\tau$ of "quantum limits," that is, weak* limit points of the sequence $\{\Phi_{kj}\}$ of distributions on the classical phase space $S^*M$, defined by

$$\int_X a\,\mathrm{d}\Phi_{kj} := \langle \mathrm{O}p(a)\varphi_k, \varphi_j \rangle$$

where $\lambda_j - \lambda_k = \tau + o(1)$ and where $a \in C^\infty(S^*M)$, or equivalently of the functionals $\rho_{jk}$.

As will be discussed in the section "Quantum weak mixing," the asymptotics of off-diagonal elements depends on the weak mixing properties of the geodesic flow and not just its ergodicity.

Matrix elements of eigenfunctions are quadratic forms. More "nonlinear" problems involve the $L^p$-norms or the distribution functions of eigenfunctions. Estimates of the $L^\infty$-norms can be obtained from the local Weyl law [10]. Since the jump in the left-hand side at $\lambda$ is $\sum_{j:\lambda_j=\lambda} |\varphi_j(x)|^2$ and the jump in the right-hand side is the jump of $R(\lambda, x)$, this implies

$$\sum_{j:\lambda_j=\lambda} |\varphi_j(x)|^2 = \mathrm{O}(\lambda^{n-1}) \Longrightarrow \|\varphi_j\|_{L^\infty} = \mathrm{O}(\lambda^{\frac{n-1}{2}}) \quad [17]$$

For general $L^p$-norms, the following bounds were proved by C Sogge for any compact Riemannian manifold:

$$\frac{\|\varphi_j\|_p}{\|\varphi\|_2} = \mathrm{O}(\lambda^{\delta(p)}), \quad 2 \le p \le \infty \qquad [18]$$

where

$$\delta(p) = \begin{cases} n\left(\dfrac{1}{2} - \dfrac{1}{p}\right) - \dfrac{1}{2}, & \dfrac{2(n+1)}{n-1} \le p \le \infty \\[2ex] \dfrac{n-1}{2}\left(\dfrac{1}{2} - \dfrac{1}{p}\right), & 2 \le p \le \dfrac{2(n+1)}{n-1} \end{cases} \qquad [19]$$

These estimates are sharp on the unit sphere $S^n \subset \mathbb{R}^{n+1}$. The extremal eigenfunctions are the zonal spherical harmonics, which are the $L^2$-normalized spectral projection kernels $\Pi_N(x, x_0)/\|\Pi_N(\cdot, x_0)\|$ centered at any $x_0$. However, they are not sharp for generic $(M, g)$, and it is natural to ask how "chaotic dynamics" might influence $L^p$-norms.

**Problem 3** Improve the estimates $\|\varphi_j\|_p/\|\varphi\|_2 = O(\lambda^{\delta(p)})$ for $(M, g)$ with ergodic or mixing geodesic flow.

C Sogge and the author have proved that if a sequence of eigenfunctions attains the bounds in [17], then there must exist a point $x_0$ so that a positive measure of geodesics starting at $x_0$ in $S^*_{x_0}M$ returns to $x_0$ at a fixed time $T$. In the real analytic case, all return so $x_0$ is a perfect recurrent point. In dimension 2, such a perfect recurrent point cannot occur if the geodesic flow is ergodic; hence $\|\varphi_j\|_{L^\infty} = o(\lambda^{(n-1)/2})$ on any real analytic surface with ergodic geodesic flow. This shows that none of the $L^p$-estimates above the critical index are sharp for real analytic surfaces with ergodic geodesic flow, and the problem is the extent to which they can be improved.

The random wave model (see the section "Random waves and orthonormal bases") predicts that eigenfunctions of Riemannian manifolds with chaotic geodesic flow should have the bounds $\|\varphi_\lambda\|_{L^p} = O(1)$ for $p < \infty$ and that $\|\varphi_\lambda\|_{L^\infty} < \sqrt{\log \lambda}$. But there are

no rigorous estimates at this time close to such predictions. The best general estimate to date on negatively curved compact manifolds (which are models of chaotic geodesic flow) is just the logarithmic improvement

$$||\varphi_j||_{L^\infty} = O\left(\frac{\lambda^{n-1}}{\log \lambda}\right)$$

on the standard remainder term in the local Weyl law. This was known for compact hyperbolic manifolds from the Selberg trace formula, and similar estimates hold manifolds without conjugate points (P Bérard). The exponential growth of the geodesic flow again causes a barrier in improving the estimate beyond the logarithm. In the analogous setting of quantum "cat maps," which are models of chaotic classical dynamics, there exist arbitrarily large eigenvalues with multiplicities of the order $O(\lambda^{n-1}/\log \lambda)$; the $L^\infty$-norm of the $L^2$-normalized projection kernel onto an eigenspace of this multiplicity is of order of the square root of the multiplicity (Faure *et al.* 2003). This raises doubt that the logarithmic estimate can be improved by general dynamical arguments. Further discussion of $L^\infty$-norms, as well as zeros, will be given at the end of the next section for ergodic systems.

## Quantum Ergodicity

In this section, we discuss results on the problems stated above when the geodesic flow of $(M, g)$ is assumed to be ergodic. Let us recall that this means that Liouville measure is an ergodic measure for $\Phi^t$. This is a spectral property of the operator $V_t$ of [2] on $L^2(S^*M, \mathrm{d}\mu)$, namely that $V_t$ has 1 as an eigenvalue of multiplicity 1. That is, the only invariant $L^2$-functions (with respect to Liouville measure) are the constant functions. This implies that the only invariant sets have Liouville measure 0 or 1 and (Birkhoff's ergodic theorem) that time averages of functions are constant almost everywhere (equal to the space average).

In this case, there is a general result which originated in the work of Schnirelman and was developed into the following theorem by Zelditch, Colin de Verdière, and Sunada (manifolds without boundary), and Gérard–Leichtnam and Zelditch–Zworski (manifolds with boundary). The following discussion is based on the articles (Zelditch 1996b, c, Zelditch and Zworski 1996), which contain further references to the literature.

**Theorem 1** *Let $(M, g)$ be a compact Riemannian manifold (possibly with boundary), and let $\{\lambda_j, \varphi_j\}$ be the spectral data of its Laplacian $\Delta$. Then the geodesic flow $G^t$ is ergodic on $(S^*M, \mathrm{d}\mu)$ if and only if, for every $A \in \Psi^o(M)$, we have:*

(i) $\lim_{\lambda \to \infty} \frac{1}{N(\lambda)} \sum_{\lambda_j \le \lambda} |(A\varphi_j, \varphi_j) - \omega(A)|^2 = 0.$

(ii) $(\forall \epsilon)(\exists \delta) \limsup_{\lambda \to \infty} \frac{1}{N(\lambda)} \sum_{j \ne k : \lambda_j, \lambda_k \le \lambda |\lambda_j - \lambda_k| < \delta} |(A\varphi_j, \varphi_k)|^2 < \epsilon.$

This implies that there exists a subsequence $\{\varphi_{j_k}\}$ of eigenfunctions whose indices $j_k$ have counting density 1 for which $\langle A\varphi_{j_k}, \varphi_{j_k}\rangle \to \omega(A)$. We will call the eigenfunctions in such a sequence "ergodic eigenfunctions." One can sharpen the results by averaging over eigenvalues in the shorter interval $[\lambda, \lambda + 1]$ rather than in $[0, \lambda]$.

There is also an ergodicity result for boundary values of eigenfunctions on domains with boundary and with Dirichlet, Neumann, or Robin boundary conditions (Gérard–Leichtnam, Hassell–Zelditch, Burq). This corresponds to the fact that the billiard map on $B^*\partial M$ is ergodic.

The first statement (i) is essentially a convexity result. It remains true if one replaces the square by any convex function $\varphi$ on the spectrum of $A$,

$$\frac{1}{N(E)} \sum_{\lambda_j \le E} \varphi(\langle A\varphi_k, \varphi_k\rangle - \omega(A)) \to 0 \qquad [20]$$

Before sketching a proof, we point out a somewhat heuristic "picture proof" of the theorem. Namely, ergodicity of the geodesic flow is equivalent to the statement that Liouville measure is an extreme point of the compact convex set $\mathcal{M}_I$. In fact, it further implies that $\omega$ is an extreme point of the compact convex set $\mathcal{E}_{\mathbb{R}}$ of invariant states for $\alpha_t$ of eqn [1]; see Ruelle (1969) for background. But the local Weyl law says that $\omega$ is also the limit of the convex combination

$$\frac{1}{N(E)} \sum_{\lambda_j \le E} \rho_j$$

An extreme point cannot be written as a convex combination of other states unless all the states in the combination are equal to it. In our case, $\omega$ is only a limit of convex combinations so it need not (and does not) equal each term. However, almost all terms in the sequence must tend to $\omega$, and that is equivalent to [1].

**Sketch of Proof of Theorem 1(i)** As mentioned above, this is a convexity result and with no additional effort we can consider more general sums of the form. We then have

$$\sum_{\lambda_j \le E} \varphi(\langle A\varphi_k, \varphi_k\rangle - \omega(A)) = \sum_{\lambda_j \le E} \varphi(\langle\langle A\rangle_T - \omega(A)\varphi_k, \varphi_k\rangle) \qquad [21]$$

where

$$\langle A\rangle_T = \frac{1}{2T}\int_{-T}^{T} U_t A U_t^* \, \mathrm{d}t$$

We then apply the Peierls–Bogoliubov inequality

$$\sum_{j=1}^{n} \varphi((B\varphi_j, \varphi_j)) \le \operatorname{tr} \varphi(B)$$

with $B = \Pi_E[\langle A\rangle_T - \omega(A)]\Pi_E$ to get

$$\sum_{\lambda_j \le E} \varphi(\langle\langle A\rangle_T - \omega(A)\varphi_k, \varphi_k\rangle) \le \operatorname{tr} \varphi(\Pi_E[\langle A\rangle_T - \omega(A)]\Pi_E) \qquad [22]$$

Here, $\Pi_E$ is the spectral projection for $\hat{H}$ corresponding to the interval $[0, E]$. From the Berezin inequality we then have (if $\varphi(0) = 0$):

$$\begin{aligned}\frac{1}{N(E)} &\operatorname{tr} \varphi(\Pi_E[\langle A\rangle_T - \omega(A)]\Pi_E) \\ &\le \frac{1}{N(E)} \operatorname{tr} \Pi_E \varphi([\langle A\rangle_T - \omega(A)])\Pi_E \\ &\to \omega_E(\varphi(\langle A\rangle_T - \omega(A))), \text{ as } E \to \infty\end{aligned}$$

As long as $\varphi$ is smooth, $\varphi(\langle A\rangle_T - \omega(A))$ is a pseudodifferential operator of order zero with principal symbol $\varphi(\langle\sigma_A\rangle_T - \omega(A))$. By the assumption that $\omega_E \to \omega$ we get

$$\begin{aligned}\lim_{E\to\infty} \frac{1}{N(E)} &\sum_{\lambda_j \le E} \varphi(\langle A\varphi_k, \varphi_k\rangle - \omega(A)) \\ &\le \int_{\{H=1\}} \varphi(\langle\sigma_A\rangle_T - \omega(A)) \, \mathrm{d}\mu\end{aligned}$$

where

$$\langle\sigma_A\rangle_T = \frac{1}{2T}\int_{-T}^{T} \sigma_A \circ \Phi^t \, \mathrm{d}t$$

As $T \to \infty$ the right-hand side approaches $\varphi(0) = 0$ by the dominated convergence theorem and by Birkhoff's ergodic theorem. Since the left-hand side is independent of $T$, this implies that

$$\lim_{E\to\infty} \frac{1}{N(E)} \sum_{\lambda_j \le E} \varphi(\langle A\varphi_k, \varphi_k\rangle - \omega(A)) = 0$$

for any smooth convex $\varphi$ on Spec$(A)$ with $\varphi(0) = 0$. □

As mentioned above, the statement of Theorem 1(i) is equivalent to saying that there is a subsequence $\{\varphi_{j_k}\}$ of counting density 1 for which $\rho_{j_k} \to \omega$. The above proof does not and cannot settle the question whether there exist exceptional sparse subsequences of eigenfunctions of density zero tending to other invariant measures. To see this, we observe that the proof is so general that it applies to seemingly very different situations. In place of the distributions $\{\Phi_j\}$ we may consider the set $\mu_\gamma$ of periodic orbit measures for a hyperbolic flow on a compact manifold $X$. That is,

$$\mu_\gamma(f) = \frac{1}{T_\gamma}\int_\gamma f \quad \text{for } f \in C(X)$$

where $\gamma$ is a closed orbit and $T_\gamma$ is its period. According to the Bowen–Margulis equidistribution theorem for closed orbits of hyperbolic flows, we have

$$\frac{1}{\Pi(T)} \sum_{\gamma: T_\gamma \le T} \frac{1}{|\det(I - P_\gamma)|} \mu_\gamma \to \mu$$

where as above $\mu$ is the Liouville measure, where $P_\gamma$ is the linear Poincaré map and where $\Pi(T)$ is the normalizing factor which makes the left side a probability measure, that is, defined by the integral of 1 against the sum. An exact repetition of the previous argument shows that up to a sparse subsequence of $\gamma$'s, $\mu_\gamma \to \mu$ individually. Yet clearly, the whole sequence does not tend to $\mathrm{d}\mu$: for instance, one could choose the sequence of iterates $\gamma^k$ of a fixed closed orbit.

### Quantum Ergodicity in Terms of Operator Time and Space Averages

The first part of the result above may be reformulated as a relation between operator time and space averages.

**Definition** Let $A \in \Psi^0$ be an observable and define its time average to be:

$$\langle A\rangle := \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} U_t^* A U_t \, \mathrm{d}t$$

and its space average to be scalar operator

$$\omega(A) \cdot I$$

Here, the limit is taken in the weak operator topology (i.e., one matrix element at a time). To see what is involved, we consider matrix elements with respect to the eigenfunctions. We have

$$\left(\frac{1}{2T}\int_{-T}^{T} U_t^* A U_t \, \mathrm{d}t \varphi_i, \varphi_j\right) = \frac{\sin T(\lambda_i - \lambda_j)}{T(\lambda_i - \lambda_j)} (A\varphi_i, \varphi_j)$$

from which it is clear that the matrix element tends to zero as $T \to \infty$ unless $\lambda_i = \lambda_j$. However, there is no uniformity in the rate at which it goes to zero since the spacing $\lambda_i - \lambda_j$ could be uncontrollably small.

In these terms, Theorem 1(i) states that

$$\langle A \rangle = \omega(A)I + K, \quad \text{where} \lim_{\lambda\to\infty} \omega_\lambda(K^*K) \to 0 \qquad [23]$$

where $\omega_\lambda(A) = \operatorname{tr} E(\lambda)A$. Thus, the time average equals the space average plus a term $K$ which is semiclassically small in the sense that its Hilbert–Schmidt norm square $\|E_\lambda K\|^2_{\mathrm{HS}}$ in the span of the eigenfunctions of eigenvalue $\leq \lambda$ is $o(N(\lambda))$.

This is not exactly equivalent to Theorem 1(i) since it is independent of the choice of orthonormal basis, while the previous result depends on the choice of basis. However, when all eigenvalues have multiplicity 1, then the two are equivalent. To see the equivalence, note that $\langle A \rangle$ commutes with $\sqrt{\Delta}$ and hence is diagonal in the basis $\{\varphi_j\}$ of joint eigenfunctions of $\langle A \rangle$ and of $U_t$. Hence, $K$ is the diagonal matrix with entries $\langle A\varphi_k, \varphi_k\rangle - \omega(A)$. The condition is therefore equivalent to

$$\lim_{E\to\infty} \frac{1}{N(E)} \sum_{\lambda_j \leq E} |\langle A\varphi_k, \varphi_k\rangle - \omega(A)|^2 = 0$$

Since all the terms are positive, no cancellation is possible and this condition is equivalent to the existence of a subset $\mathcal{S} \subset \mathbb{N}$ of density 1 such that $\mathcal{Q}_\mathcal{S} := \{\mathrm{d}\Phi_k : k \in \mathcal{S}\}$ has only $\omega$ as a weak$^*$ limit point. As above, one says that the sequence of eigenfunctions is ergodic.

One could take this restatement of Theorem 1(i) as a semiclassical definition of quantum ergodicity. Two natural questions arise. First:

**Problem 4** Suppose the geodesic flow $\Phi^t$ of $(M, g)$ is ergodic on $S^*M$. Is the operator $K$ in

$$\langle A \rangle = \omega(A) + K$$

a compact operator? In this case, $\sqrt{\Delta}$ is said to be quantum uniquely ergodic (QUE). If ergodicity is not sufficient for the QUE property, what extra conditions need to be added?

Compactness would imply that $\langle K\varphi_k, \varphi_k\rangle \to 0$, hence $\langle A\varphi_k, \varphi_k\rangle \to \omega(A)$ along the entire sequence. Quite a lot of attention has been focused on this problem in the last decade. It is probable that ergodicity is not by itself sufficient for the QUE property of general Riemann manifold. For instance, it is believed that there exist modes of asymptotic bouncing ball type which concentrate on the invariant Lagrangian cylinder (with boundary) formed by bouncing ball orbits of the Bunimovich stadium (see e.g., Heller (1984) for more on such "scarring"). Further, Faure *et al.* (2003) have shown that QUE does not hold for the hyperbolic system defined by a quantum cat map on the torus. Since the methods applicable to eigenfunctions of quantum maps and of Laplacians have much in common, this negative result shows that there cannot exist a universal structural proof of QUE.

The principal positive result available at this time is the recent proof by Lindenstrauss of the QUE property for the orthonormal basis of Laplace–Hecke eigenfunctions on arithmetic hyperbolic surfaces. It is generally believed that the spectrum of the Laplace eigenvalues is of multiplicity 1 for such surfaces, so this should imply QUE completely for these surfaces. Earlier partial results on Hecke eigenfunctions are due to Rudnick–Sarnak, Wolpert, and others. For references and further discussion onf Hecke eigenfunctions, see Rudnick and Sarnak (1994) (*see* Arithmetic Quantum Chaos).

So far we have not mentioned Theorem 1(ii). In the next section, we will describe a similar but more general result for mixing systems and the relevance of (ii) will become clear. An interesting open problem is the extent to which (ii) is actually necessary for the equivalence to classical ergodicity.

**Problem 5** Converse QE: What can be said of the classical limit of a quantum ergodic system, that is, a system for which $\langle A \rangle = \omega(A) + K$, where $K$ is compact? Is it necessarily ergodic?

Very little is known on this converse problem at present. It is known that if there exists an open set in $S^*M$ filled by periodic orbits, then the Laplacian cannot be quantum ergodic (see Marklof and O'Keefe (2005) for recent results and references). But no proof exists at this time that KAM systems, which have Cantor-like positive measure invariant sets, are not quantum ergodic. It is known that there exists a positive proportion of approximate eigenfunctions (quasimodes) which localize on the invariant tori, but it has not been proved that a positive proportion of actual eigenfunctions has this localization property.

## Further Problems and Results on Ergodic Eigenfunctions

Ergodicity is also known to have an impact on the distribution of zeros. The complex zeros in Kähler phase spaces of ergodic eigenfunctions of quantum ergodic maps become uniformly distributed with respect to the Kähler volume form (Nonnenmacher–Voros, Shiffman–Zelditch). An interesting problem is whether the real analog is true:

**Problem 6** Ergodicity and equidistribution of nodal sets. Let $\mathcal{N}_{\varphi_j} \subset M$ denote the nodal set (zero set) of $\varphi_j$, and equip it with its hypersurface volume form $\mathrm{d}\mathcal{H}^{n-1}$ induced by $g$. Let $(M, g)$ have ergodic geodesic flow, and suppose that $\{\varphi_j\}$ is an ergodic

sequence of eigenfunctions. Are the following asymptotics valid?

$$\int_{\mathcal{N}_{\varphi_j}} f\, \mathrm{d}\mathcal{H}^{n-1} \sim \lambda_j \frac{1}{\mathrm{Vol}(M,g)} \int_M f\, \mathrm{dVol}$$

This is predicted by the random wave model of the section "Random waves and orthonormal bases." An equidistribution law for the complex zeros is known which gives some evidence for the validity of this limit formula. Let $(M,g)$ be a compact real analytic Riemannian manifold and let $\varphi_j^{\mathbb{C}}$ be the holomorphic extension of the real analytic eigenfunction $\varphi_j$ to the complexification $M_{\mathbb{C}}$ of $M$ (its Grauert tube). Then, if the geodesic flow is ergodic and if $\varphi_j$ is an ergodic sequence of eigenfunctions, the normalized current of integration $(1/\lambda_j)Z_{\varphi_j^{\mathbb{C}}}$ over the complex zero set of $\varphi_j^{\mathbb{C}}$ tends weakly to $(\mathrm{i}/\pi)\bar{\partial}\partial|\xi_g|$. This current is singular along the zero section.

Finally, we mention some results on $L^\infty$-norms of eigenfunctions on arithmetic hyperbolic manifolds of dimensions 2 and 3. It was proved by Iwaniec–Sarnak that the joint eigenfunctions of $\Delta$ and the Hecke operators on arithmetic hyperbolic surfaces have the upper bound $\|\varphi_j\|_\infty = O_\epsilon(\lambda_j^{5/48+\epsilon})$ for all $j$ and $\epsilon > 0$, and the lower bound $\|\varphi_j\|_\infty \geq c\sqrt{\log\log \lambda_j}$ for some constant $c > 0$ and infinitely many $j$. Rudnick and Sarnak (1994) proved that there exists an arithmetic hyperbolic manifold and a subsequence $\varphi_{j_k}$ of eigenfunctions with $\|\varphi_{j_k}\|_{L^\infty} \gg \lambda_{j_k}^{1/4}$, contradicting the random wave model predictions.

## Quantum Weak Mixing

There are parallel results on quantizations of weak-mixing geodesic flows which are the subject of this section. First we recall the classical definition: the geodesic flow of $(M,g)$ is weak mixing if the operator $V_t$ has purely continuous spectrum on the orthogonal complement of the constant functions in $L^2(S^*M, \mathrm{d}\mu)$. Hence, like ergodicity, it is a spectral property of the geodesic flow.

We have:

**Theorem 2** (Zelditch 1996c). *The geodesic flow $\Phi^t$ of $(M,g)$ is weak mixing if and only if the conditions (i) and (ii) of Theorem 1 hold and additionally, for any $A \in \Psi^o(M)$,*

$$(\forall\epsilon)(\exists\delta) \limsup_{\lambda\to\infty} \frac{1}{N(\lambda)} \sum_{\substack{j\neq k:\lambda_j,\lambda_k\leq\lambda \\ |\lambda_j-\lambda_k-\tau|<\delta}} |(A\varphi_j,\varphi_k)|^2 < \epsilon$$

$$(\forall\tau\in\mathbb{R})$$

The restriction $j \neq k$ is of course redundant unless $\tau = 0$, in which case the statement coincides with quantum ergodicity. This result follows from the general asymptotic formula, valid for any compact Riemannian manifold $(M,g)$, that

$$\frac{1}{N(\lambda)} \sum_{i\neq j,\lambda_i,\lambda_j\leq\lambda} |\langle A\varphi_i,\varphi_j\rangle|^2 \left|\frac{\sin T(\lambda_i-\lambda_j-\tau)}{T(\lambda_i-\lambda_j-\tau)}\right|^2 \sim \left\|\frac{1}{2T}\int_{-T}^{T} \mathrm{e}^{\mathrm{i}t\tau} V_t(\sigma_A)\right\|_2^2 - \left|\frac{\sin T\tau}{T\tau}\right|^2 \omega(A)^2 \qquad [24]$$

In the case of weak-mixing geodesic flows, the right-hand side tends to 0 as $T\to\infty$. As with diagonal sums, the sharper result is true where one averages over the short intervals $[\lambda, \lambda+1]$.

### Spectral Measures and Matrix Elements

Theorem 2 is based on expressing the spectral measures of the geodesic flow in terms of matrix elements. The main limit formula is

$$\int_{\tau-\varepsilon}^{\tau+\varepsilon} \mathrm{d}\mu_{\sigma_A} := \lim_{\lambda\to\infty} \frac{1}{N(\lambda)} \sum_{\substack{i,j:\ \lambda_j\leq\lambda, \\ |\lambda_i-\lambda_j-\tau|<\varepsilon}} |\langle A\varphi_i,\varphi_j\rangle|^2 \qquad [25]$$

where $\mathrm{d}\mu_{\sigma_A}$ is the spectral measure for the geodesic flow corresponding to the principal symbol of $A$, $\sigma_A \in C^\infty(S^*M, \mathrm{d}\mu)$. Recall that the spectral measure of $V_t$ corresponding to $f \in L^2$ is the measure $\mathrm{d}\mu_f$ defined by

$$\langle V_t f, f\rangle_{L^2(S^*M)} = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}t\tau}\, \mathrm{d}\mu_f(\tau)$$

The limit formula [25] is equivalent to the dual formula (under the Fourier transform):

$$\lim_{\lambda\to\infty} \frac{1}{N(\lambda)} \sum_{i,j:\lambda_j\leq\lambda} \mathrm{e}^{\mathrm{i}t(\lambda_i-\lambda_j)} |\langle A\varphi_i,\varphi_j\rangle|^2 = \langle V_t\sigma_A, \sigma_A\rangle_{L^2(S^*M)} \qquad [26]$$

The proof of [26] is to consider, for $A \in \Psi^\circ$, the operator $A_t^*A \in \Psi^\circ$ with $A_t = U_t^* A U_t$. By the local Weyl law,

$$\lim_{\lambda\to\infty} \frac{1}{N(\lambda)} \mathrm{tr}\, E(\lambda) A_t^* A = \langle V_t\sigma_A, \sigma_A\rangle_{L^2(S^*M)}$$

The right-hand-side of [25] defines a measure $\mathrm{d}m_A$ on $\mathbb{R}$ and [26] says

$$\int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}t\tau}\, \mathrm{d}m_A(\tau) = \langle V_t\sigma_A, \sigma_A\rangle_{L^2(S^*M)} = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}t\tau}\, \mathrm{d}\mu_{\sigma_A}(\tau)$$

Since weak-mixing systems are ergodic, it is not necessary to average in both indices along an ergodic subsequence:

$$\lim_{\lambda_j\to\infty}\langle A_t^* A\varphi_j,\varphi_j\rangle = \sum_j e^{it(\lambda_i-\lambda_j)}|\langle A\varphi_i,\varphi_j\rangle|^2 = \langle V_t\sigma_A,\sigma_A\rangle_{L^2(S^*M)} \qquad [27]$$

Dually, one has

$$\lim_{\lambda_j\to\infty}\sum_{i:|\lambda_i-\lambda_j-\tau|<\varepsilon}|\langle A\varphi_i,\varphi_j\rangle|^2 = \int_{\tau-\varepsilon}^{\tau+\varepsilon} d\mu_{\sigma_A} \qquad [28]$$

For QUE systems, these limit formulas are valid for the full sequence of eigenfunctions.

## Rate of Quantum Ergodicity and Mixing

A quantitative refinement of quantum ergodicity is to ask at what rate the sums in Theorem 1(i) tend to zero, that is, to establish a rate of quantum ergodicity. More generally, we consider "variances" of matrix elements. For diagonal matrix elements, we define

$$V_A(\lambda) := \frac{1}{N(\lambda)}\sum_{j:\lambda_j\le\lambda}|\langle A\varphi_j,\varphi_j\rangle-\omega(A)|^2 \qquad [29]$$

In the off-diagonal case, one may view $|\langle A\varphi_i,\varphi_j\rangle|^2$ as analogous to $|\langle A\varphi_j,\varphi_j\rangle-\omega(A)|^2$. However, the sums in [25] are double sums while those of [29] are single. One may also average over the shorter intervals $[\lambda,\lambda+1]$.

### Quantum Chaos Conjectures

First, consider off-diagonal matrix elements. One conjecture is that it is not necessary to sum in $j$ in [28]: each individual term has the asymptotics consistent with [28]. This is implicitly conjectured by Feingold–Peres (1986) (see [11]) in the form

$$|\langle A\varphi_i,\varphi_j\rangle|^2 \simeq \frac{C_A\left(\frac{E_i-E_j}{\hbar}\right)}{2\pi\rho(E)} \qquad [30]$$

where

$$C_A(\tau)=\int_{-\infty}^{\infty} e^{-i\tau t}\langle V_t\sigma_A,\sigma_A\rangle\, dt$$

In our notation, $\lambda_j=\hbar^{-1}E_j$ and $\rho(E)\,dE\sim dN(\lambda)$. There are $\sim C\lambda^{n-1}$ eigenvalues $\lambda_i$ in the interval $[\lambda_j-\tau-\epsilon,\lambda_j-\tau+\epsilon]$, so [30] states that individual terms have the asymptotics of [28].

On the basis of the analogy between $|\langle A\varphi_i,\varphi_j\rangle|^2$ and $|\langle A\varphi_j,\varphi_j\rangle-\omega(A)|^2$, it is conjectured in Feingold and Peres (1986) that

$$V_A(\lambda)\sim\frac{C_{A-\omega(A)I}(0)}{\lambda^{n-1}\operatorname{vol}(\Omega)}$$

The idea is that $\varphi_\pm=(1/\sqrt{2})(\varphi_i\pm\varphi_j)$ have the same matrix element asymptotics as eigenfunctions when $\lambda_i-\lambda_j$ is sufficiently small. But then $2\langle A\varphi_+,\varphi_-\rangle=\langle A\varphi_i,\varphi_i\rangle-\langle A\varphi_j,\varphi_j\rangle$ when $A^*=A$. Since we are taking a difference, we may replace each matrix element $\langle A\varphi_i,\varphi_i\rangle$ by $\langle A\varphi_i,\varphi_i\rangle-\omega(A)$ (and also for $\varphi_j$). The conjecture then assumes that $\langle A\varphi_i,\varphi_i\rangle-\omega(A)$ has the same order of magnitude as $\langle A\varphi_i,\varphi_i\rangle-\langle A\varphi_j,\varphi_j\rangle$. Dynamical grounds for this conjecture are given in Eckhardt *et al.* (1995). The order of magnitude is predicted by some natural random wave models, as discussed in the next section.

### Rigorous results

At this time, the strongest variance result is an asymptotic formula for the diagonal variance proved by Luo and Sarnak (2004) for special Hecke eigenfunctions on the quotient $\boldsymbol{H}^2/\mathrm{SL}(2,\mathbb{Z})$ of the upper half plane by the modular group. Their result pertains to holomorphic Hecke eigenforms, but the analogous statement for smooth Maass–Hecke eigenfunctions is expected to hold by similar methods, so we state the result as a theorem/conjecture. Note that $\boldsymbol{H}^2/\mathrm{SL}(2,\mathbb{Z})$ is a noncompact finite-area surface whose Laplacian $\Delta$ has both a discrete and a continuous spectrum. The discrete Hecke eigenfunctions are joint eigenfunctions of $\Delta$ and the Hecke operators $T_p$.

**Theorem/Conjecture 1** (Luo and Sarnak 2004). *Let $\{\varphi_k\}$ denote the orthonormal basis of Hecke eigenfunctions for $\boldsymbol{H}^2/\mathrm{SL}(2,\mathbb{Z})$. Then there exists a quadratic form $B(f)$ on $C_0^\infty(\boldsymbol{H}^2/\mathrm{SL}(2,\mathbb{Z}))$ such that*

$$\frac{1}{N(\lambda)}\sum_{\lambda_j\le\lambda}\left|\int_X f|\varphi_j|^2 d\mathrm{vol}-\frac{1}{\mathrm{Vol}(X)}\int_X f d\mathrm{Vol}\right|^2 = \frac{B(f,f)}{\lambda}+o\left(\frac{1}{\lambda}\right)$$

When the multiplier $f=\varphi_\lambda$ is itself an eigenfunction, Luo–Sarnak have shown that

$$B(\varphi_\lambda,\varphi_\lambda)=C_{\varphi_\lambda}(0)L(\tfrac12,\varphi_\lambda)$$

where $L(\frac12,\varphi_\lambda)$ is a certain $L$-function. Thus, the conjectured classical variance is multiplied by an arithmetic factor depending on the multiplier. A crucial fact in the proof is that the quadratic form B is diagonalized by the $\varphi_\lambda$.

The only rigorous result to date which is valid on general Riemannian manifolds with hyperbolic geodesic flow is the logarithmic decay:

**Theorem 3** (*Zelditch*). *For any* $(M, g)$ *with hyperbolic geodesic flow,*

$$\frac{1}{N(\lambda)}\sum_{\lambda_j\leq\lambda}|(A\varphi_j,\varphi_j)-\omega(A)|^{2p}=\frac{1}{(\log\lambda)^p}$$

The logarithm reflects the exponential blow-up in time of remainder estimates for traces involving the wave group associated to hyperbolic flows. It would be surprising if the logarithmic decay is sharp for Laplacians. However, a recent result of R Schubert shows that the estimate is sharp in the case of two-dimensional hyperbolic quantum cat maps. Hence, the estimate cannot be improved by semiclassical arguments that hold in both settings.

## Random Waves and Orthonormal Bases

We have mentioned that the random wave model provides a kind of guideline for what to conjecture about eigenfunctions of quantum chaotic system. In this final section, we briefly discuss random wave models and what they predict.

By a random wave model, one means a probability measure on a space of functions. To deal with orthonormal bases rather than individual functions, one sets a probability measure on a space of orthonormal bases, that is, on a unitary group. We denote expected values relative to a given probability measure by $\boldsymbol{E}$. We now consider some specific Gaussian models and what they predict about variances.

As a model for quantum chaotic eigenfunctions in plane domains, Berry (1977) suggested using the Euclidean random wave model at fixed energy. A rigorous version of such a model is as follows: let $\mathcal{E}_\lambda$ denote the space of (tempered) eigenfunctions of eigenvalue $\lambda^2$ of the Euclidean Laplacian $\Delta$ on $\mathbb{R}^n$. It is spanned by exponentials $e^{i\langle k,x\rangle}$ with $k\in\mathbb{R}^n, |k|=\lambda$. The infinite-dimensional space $\mathcal{E}_\lambda$ is a unitary representation of the Euclidean motion group and carries an invariant inner product. The inner product defines an associated Gaussian measure whose covariance kernel $C_\lambda(x,y)=\boldsymbol{E}f(x)\bar{f}(y)$ is the derivative at $\lambda$ of the spectral function

$$E(\lambda,x,y)=(2\pi)^{-n}\int_{|\xi|\leq\lambda}e^{i\langle x-y,\xi\rangle}\,d\xi,\quad \xi\in\mathbb{R}^n \quad [31]$$

Thus,

$$\begin{aligned}C_\lambda(x,y)&=\frac{d}{d\lambda}E(\lambda,x,y)\\&=(2\pi)^{-n}\int_{|\xi|=\lambda}e^{i\langle x-y,\xi\rangle}dS\\&=(2\pi)^{-n}\lambda^{n-1}\int_{|\xi|=1}e^{i\lambda\langle x-y,\xi\rangle}\,dS \quad [32]\end{aligned}$$

where $dS$ is the usual surface measure. With this definition, $C_\lambda(x,x)\sim\lambda^{n-1}$. In order to make $\mathbf{E}(f(x)^2)=1$ consistent with normalized eigenfunctions, we divide by $\lambda^{n-1}$ to define

$$\hat{C}_\lambda(x,y)=(2\pi)^{-n}\int_{|\xi|=1}e^{i\lambda\langle x-y,\xi\rangle}dS$$

One could express the integral as a Bessel function to rewrite this as

$$\Gamma\left(\frac{n-1}{2}\right)|\lambda|x-y||^{-(n-2)/2}J_{(n-2)/2}(\lambda|x-y|)$$

Wick's formula in this ensemble gives

$$\boldsymbol{E}\varphi(x)^2\varphi(y)^2=\frac{1}{\mathrm{Vol}(\Omega)^2}[1+2C_\lambda(x,y)^2]$$

Thus, in dimension $n$ we have

$$\begin{aligned}&\boldsymbol{E}\left[\int\!\!\int V(x)V(y)\varphi(x)^2\varphi(y)^2dxdy-\bar{V}^2\right]\\&=\frac{2}{\mathrm{Vol}(\Omega)^2}\int_\Omega\int_\Omega\hat{C}_\lambda(x,y)^2V(x)V(y)dx\,dy\\&\sim\frac{1}{\lambda^{n-1}\mathrm{Vol}(\Omega)^2}\int_\Omega\int_\Omega\frac{V(x)V(y)}{|x-y|^{n-1}}\cos(|x-y|\lambda)^2dx\,dy\end{aligned}$$

In the last line, we used the stationary-phase asymptotics

$$\begin{aligned}&(2\pi)^{-n}\lambda^{n-1}\int_{|\xi|=1}e^{i\lambda\langle x-y,\xi\rangle}dS\\&\quad\sim C_n(\lambda|x-y|)^{-(n-1)/2}\cos(|x-y|\lambda) \quad [33]\end{aligned}$$

Thus, the variances have order $\lambda^{-(n-1)}$ in dimension $n$, consistent with the conjectures in Feingold and Peres (1986) and Eckhardt *et al.* (1995).

This model is often used to obtain predictions on eigenfunctions of chaotic systems. By construction, it is tied to Euclidean geometry and only pertains directly to individual eigenfunctions of a fixed eigenvalue. It is based on the infinite-dimensional multiplicity of eigenfunctions of fixed eigenvalue of the Euclidean Laplacian on $\mathbb{R}^n$. There also exist random wave models on a curved Riemannian manifold $(M,g)$, which model individual eigenfunctions and also random orthonormal bases

(Zelditch 1996a). Thus, one can compare the behavior of sums over eigenvalues of the orthonormal basis of eigenfunctions of $\Delta$ with that of a random orthonormal basis. Instead of taking Gaussian random combinations of Euclidean plane waves of a fixed eigenvalue, one takes Gaussian random combinations $\sum_{j:\lambda_j\in[\lambda,\lambda+1]} c_j\varphi_j$ of the eigenfunctions of $(M,g)$ with eigenvalues in a short interval in the sense above. Equivalently, one takes random combinations with $\sum_j |c_j|^2 = 1$. These random waves are globally adapted to $(M,g)$. The statistical results depend on the measure of the set of periodic geodesics of $(M,g)$; thus, as discussed in Kaplan and Heller (1998), different random wave models make different predictions about off-diagonal variances.

Fix a compact Riemannian manifold $(M,g)$ and partition the spectrum of $\sqrt{\Delta}$ into the intervals $I_k = [k, k+1]$. Let $\Pi_k = E(k+1) - E(k)$ be the kernel of the spectral projections for $\sqrt{\Delta}$ corresponding to the interval $I_k$. Its kernel $\Pi_k(x,y)$ is the covariance kernel of Gaussian random combinations $\sum_{j:\lambda_j\in I_k} c_j\varphi_j$ and is analogous to $C_\lambda(x,y)$ in the Euclidean case; it is of course not the derivative $\mathrm{d}E(\lambda,x,y)$ but the difference of the spectral projector over $I_k$. We denote by $N(k)$ the number of eigenvalues in $I_k$ and put $\mathcal{H}_k = \mathrm{ran}\Pi_k$ (the range of $\Pi_k$). We define a "random" orthonormal basis of $\mathcal{H}_k$ by changing the basis of eigenfunctions $\{\varphi_j\}$ of $\Delta$ in $\mathcal{H}_k$ by a random element of the unitary group $U(\mathcal{H}_k)$ of the finite-dimensional Hilbert space $\mathcal{H}_k$. We then define a random orthonormal basis of $L^2(M)$ by taking the product over all the spectral intervals in our partition. More precisely, we define the infinite-dimensional unitary group

$$U(\infty) = \prod_{k=1}^{\infty} U(\mathcal{H}_k)$$

of sequences $(U_1, U_2, \ldots)$, with $U_k \in U(\mathcal{H}_k)$. We equip $U(\infty)$ with the product

$$\mathrm{d}\nu_\infty = \prod_{k=1}^{\infty} \mathrm{d}\nu_k$$

of the unit mass Haar measures $\mathrm{d}\nu_k$ on $U(\mathcal{H}_k)$: we then define a random orthonormal basis of $L^2(M)$ to be obtained by applying a random element $U \in U(\infty)$ to the orthonormal basis $\Phi = \{\varphi_j\}$ of eigenfunctions of $\sqrt{\Delta}$.

Assuming the set of periodic geodesics of $(M,g)$ has measure zero, the Weyl remainder results [8] and strong Szegö limit asymptotics of Guillemin–Okikiolu and Laptev–Robert–Safarov give two term asymptotics for the traces $\Pi_k A \Pi_k, (\Pi_k A \Pi_k)^2$ for any pseudodifferential operator $A$. Combining the strong Szegö asymptotics with the arguments of Zelditch (1996a), random orthonormal bases can be proved to satisfy the following variance asymptotics:

1. $\boldsymbol{E}(\sum_{j:\lambda_j\in I_k} |(AU\varphi_j, U\varphi_j) - \omega(A)|^2$
   $\sim (\omega(A^*A) - \omega(A)^2)$;

2. $\boldsymbol{E}(\sum_{i\neq j:\lambda_j,\lambda_i\in I_k} \left|\frac{\sin T(\lambda_i-\lambda_j-\tau)}{T(\lambda_i-\lambda_j-\tau)}\right|^2 |(AU\varphi_j, U\varphi_i)|^2$
   $\sim \left\{ 2\left|\frac{\sin\tau T}{\tau T}\right|^2 + \frac{1}{N(k)}\sum_{i\neq j}\left|\frac{\sin T(\lambda_i-\lambda_j-\tau)}{T(\lambda_i-\lambda_j-\tau)}\right|^2 \right\}$
   $\times(\omega(A^*A) - \omega(A)^2)$

*See also:* Arithmetic Quantum Chaos; Determinantal Random Fields; Eigenfunctions of Quantum Completely Integrable Systems; Fractal Dimensions in Dynamics; *h*-Pseudodifferential Operators and Applications; Number Theory in Physics; Regularization for Dynamical Zeta Functions; Semiclassical Spectra and Closed Orbits.

## Further Reading


Bäcker A, Schubert R, and Stifter P (1998a) Rate of quantum ergodicity in Euclidean billiards. *Physical Review E (3)* 57(5): part A 5425–5447.

Bäcker A, Schubert R, and Stifter P (1998b) Rate of quantum ergodicity in Euclidean billiards – erratum. *Physical Review E, (3)* 58(4): 5192 (math-ph/0512030).

Barnett A (2005) Asymptotic rate of quantum ergodicity in chaotic Euclidean billiards.

Berry MV (1977) Regular and irregular semiclassical wavefunctions. *Journal of Physics A* 10(12): 2083–2091.

Bohigas O, Giannoni MJ, and Schmit C (1984) Characterization of chaotic quantum spectra and universality of level fluctuation laws. *Physical Review Letters* 52(1): 1–4.

Eckhardt B, Fishman S, Keating J, Agam O, Main J *et al.* (1995) Approach to ergodocity in quantum wave functions. *Physical Review E* 52: 5893–5903.

Faure F, Nonnenmacher S, and De Bièvre S (2003) Scarred eigenstates for quantum cat maps of minimal periods. *Communications in Mathematical Physics* 239(3): 449–492.

Feingold M and Peres A (1986) Distribution of matrix elements of chaotic systems. *Physical Review A (3)* 34(1): 591–595.

Heller EJ (1984) Bound-state eigenfunctions of classically chaotic Hamiltonian systems: Scars of periodic orbits. *Physical Review Letters* 53(16): 1515–1518.

Kaplan L and Heller EJ (1998) Weak quantum ergodicity. *Physica D* 121(1–2): 1–18 (Appendix by S. Zelditch.).

Luo W and Sarnak P (2004) Quantum variance for Hecke eigenforms. *Annales Scientifiques de l'Ecole Normale Superieure* 37: 769–799.

Marklof J and O'Keefe S (2005) Weyl's law and quantum ergodicity for maps with divided phase space (with an appendix 'Converse quantum ergodicity' by Steve Zelditch). *Nonlinearity* 18(1): 277–304.

Rudnick Z and Sarnak P (1994) The behaviour of eigenstates of arithmetic hyperbolic manifolds. *Communications in Mathematical Physics* 161: 195–213.

Ruelle D (1969) *Statistical Mechanics: Rigorous Results*. New York: W. A. Benjamin.

Sarnak P (1995) Arithmetic quantum chaos. In: *The Schur Lectures* (*Tel Aviv*, *1992*), *Israel Mathematical Conference Proc*, vol. 8 (1995), pp. 183–236.
Zelditch S (1996a) A random matrix model for quantum mixing. *International Mathematics Research Notices* 3: 115–137.
Zelditch S (1996b) Quantum ergodicity of $C^*$ dynamical systems. *Communications in Mathematical Physics* 177: 507–528.
Zelditch S (1996c) Quantum Mixing. *Journal of Functional Analysis* 140: 68–86.
Zelditch S and Zworski M (1996) Ergodicity of eigenfunctions for ergodic billiards. *Communications in Mathematical Physics* 175: 673–682.

# Quantum Error Correction and Fault Tolerance

**D Gottesman**, Perimeter Institute, Waterloo, ON, Canada



## Quantum Error Correction

Building a quantum computer or a quantum communications device in the real world means having to deal with errors. Any qubit stored unprotected or one transmitted through a communications channel will inevitably come out at least slightly changed. The theory of quantum error-correcting codes (QECCs) has been developed to counteract noise introduced in this way. By adding extra qubits and carefully encoding the quantum state we wish to protect, a quantum system can be insulated to a great extent against errors.

To build a quantum computer, we face an even more daunting task: if our quantum gates are imperfect, everything we do will add to the error. The theory of fault-tolerant quantum computation tells us how to perform operations on states encoded in a QECC without compromising the code's ability to protect against errors.

In general, a QECC is a subspace of a Hilbert space designed so that any of a set of possible errors can be corrected by an appropriate quantum operation. Specifically:

**Definition 1** Let $\mathcal{H}_n$ be a $2^n$-dimensional Hilbert space ($n$ qubits), and let $C$ be a $K$-dimensional subspace of $\mathcal{H}_n$. Then $C$ is an $((n, K))$ (binary) QECC correcting the set of errors $\mathcal{E} = \{E_a\}$ iff $\exists \mathcal{R}$ s.t. $\mathcal{R}$ is a quantum operation and $(\mathcal{R} \circ E_a)(|\psi\rangle) = |\psi\rangle$ for all $E_a \in \mathcal{E}, |\psi\rangle \in C$.

$\mathcal{R}$ is called the "recovery" or "decoding" operation and serves to actually perform the correction of the state. The decoder is sometimes also taken to map $\mathcal{H}_n$ into an unencoded Hilbert space $\mathcal{H}_{\log K}$ isomorphic to $C$. This should be distinguished from the "encoding" operation which maps $\mathcal{H}_{\log K}$ into $\mathcal{H}_n$, determining the imbedding of $C$. The computational complexity of the encoder is frequently a great deal lower than that of the decoder. In particular, the task of determining what error has occurred can be computationally difficult (NP-hard, in fact), and designing codes with efficient decoding algorithms is an important task in quantum error correction, as in classical error correction.

This article will cover only binary quantum codes, built with qubits as registers, but all of the techniques discussed here can be generalized to higher-dimensional registers, or "qudits."

To determine whether a given subspace is able to correct a given set of errors, we can apply the quantum error-correction conditions (Bennett *et al.* 1996, Knill and Laflamme 1997):

**Theorem 1** *A QECC C corrects the set of errors $\mathcal{E}$ iff*

$$\langle \psi_i | E_a^\dagger E_b | \psi_j \rangle = C_{ab} \delta_{ij} \qquad [1]$$

*where $E_a, E_b \in \mathcal{E}$ and $\{|\psi_i\rangle\}$ form an orthonormal basis for C.*

The salient point in these error-correction conditions is that the matrix element $C_{ab}$ does not depend on the encoded basis states $i$ and $j$, which, roughly speaking, indicates that neither the environment nor the decoding operation learns any information about the encoded state. We can imagine the various possible errors taking the subspace $C$ into other subspaces of $\mathcal{H}_n$, and we want those subspaces to be isomorphic to $C$, and to be distinguishable from each other by an appropriate measurement. For instance, if $C_{ab} = \delta_{ab}$, then the various erroneous subspaces are orthogonal to each other.

Because of the linearity of quantum mechanics, we can always take the set of errors $\mathcal{E}$ to be a linear space: if a QECC corrects $E_a$ and $E_b$, it will also correct $\alpha E_a + \beta E_b$ using the same recovery operation. In addition, if we write any superoperator $\mathcal{S}$ in terms of its operator-sum representation $\mathcal{S}(\rho) \mapsto \sum A_k \rho A_k^\dagger$, a QECC that corrects the set of errors $\{A_k\}$ automatically corrects $\mathcal{S}$ as well. Thus, it is sufficient in general to check that the error-correction conditions hold for a basis of errors.

Frequently, we are interested in codes that correct any error affecting $t$ or fewer physical qubits. In that case, let us consider tensor products of the Pauli matrices

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$
$$Y = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad [2]$$

Define the Pauli group $\mathcal{P}_n$ as the group consisting of tensor products of $I$, $X$, $Y$, and $Z$ on $n$ qubits, with an overall phase of $\pm 1$ or $\pm \mathrm{i}$. The weight $\mathrm{wt}(P)$ of a Pauli operator $P \in \mathcal{P}_n$ is the number of qubits on which it acts as $X$, $Y$, or $Z$ (i.e., not as the identity). Then the Pauli operators of weight $t$ or less form a basis for the set of all errors acting on $t$ or fewer qubits, so a QECC which corrects these Pauli operators corrects all errors acting on up to $t$ qubits. If we have a channel which causes errors independently with probability $O(\epsilon)$ on each qubit in the QECC, then the code will allow us to decode a correct state except with probability $O(\epsilon^{t+1})$, which is the probability of having more than $t$ errors. We get a similar result in the case where the noise is a general quantum operation on each qubit which differs from the identity by something of size $O(\epsilon)$.

**Definition 2** The distance $d$ of an $((n, K))$ QECC is the smallest weight of a nontrivial Pauli operator $E \in \mathcal{P}_n$ s.t. the equation

$$\langle \psi_i | E | \psi_j \rangle = C(E)\delta_{ij} \qquad [3]$$

fails.

We use the notation $((n, K, d))$ to refer to an $((n, K))$ QECC with distance $d$. Note that for $P, Q \in \mathcal{P}_n, \mathrm{wt}(PQ) \le \mathrm{wt}(P) + \mathrm{wt}(Q)$. Then by comparing the definition of distance with the quantum error-correction conditions, we immediately see that a QECC corrects $t$ general errors iff its distance $d > 2t$. If we are instead interested in "erasure" errors, when the location of the error is known but not its precise nature, a distance $d$ code corrects $d-1$ erasure errors. If we only wish to detect errors, a distance $d$ code can detect errors on up to $d-1$ qubits.

One of the central problems in the theory of quantum error correction is to find codes which maximize the ratios $(\log K)/n$ and $d/n$, so they can encode as many qubits as possible and correct as many errors as possible. Conversely, we are also interested in the problem of setting upper bounds on achievable values of $(\log K)/n$ and $d/n$. The quantum Singleton bound (or Knill–Laflamme (1997) bound) states that any $((n, K, d))$ QECC must satisfy

$$n - \log K \ge 2d - 2 \qquad [4]$$

We can set a lower bound on the existence of QECCs using the quantum Gilbert–Varshamov bound, which states that, for large $n$, an $((n, 2^k, d))$ QECC exists provided that

$$k/n \le 1 - (d/n)\log 3 - h(d/n) \qquad [5]$$

where $h(x) = -x\log x - (1-x)\log(1-x)$ is the binary Hamming entropy. Note that the Gilbert–Varshamov bound simply states that codes at least this good exist; it does not suggest that better codes cannot exist.

## Stabilizer Codes

In order to better manipulate and discover QECCs, it is helpful to have a more detailed mathematical structure to work with. The most widely used structure gives a class of codes known as "stabilizer codes" (Calderbank *et al.* 1998, Gottesman 1996). They are less general than arbitrary quantum codes, but have a number of useful properties that make them easier to work with than the general QECC.

**Definition 3** Let $S \subset \mathcal{P}_n$ be an abelian subgroup of the Pauli group that does not contain $-1$ or $\pm\mathrm{i}$, and let $C(S) = \{|\psi\rangle \text{ s.t. } P|\psi\rangle = |\psi\rangle \forall P \in S\}$. Then $C(S)$ is a stabilizer code and $S$ is its stabilizer.

Because of the simple structure of the Pauli group, any abelian subgroup has order $2^{n-k}$ for some $k$ and can easily be specified by giving a set of $n-k$ commuting generators.

The code words of the QECC are by definition in the +1-eigenspace of all elements of the stabilizer, but an error $E$ acting on a code word will move the state into the $-1$-eigenspace of any stabilizer element $M$ which anticommutes with $E$:

$$M(E|\psi\rangle) = -EM|\psi\rangle = -E|\psi\rangle \qquad [6]$$

Thus, measuring the eigenvalues of the generators of $S$ tells us information about the error that has occurred. The set of such eigenvalues can be represented as an $(n-k)$-dimensional binary vector known as the "error syndrome." Note that the error syndrome does not tell us anything about the encoded state, only about the error that has occurred.

**Theorem 2** *Let $S$ be a stabilizer with $n-k$ generators, and let $S^{\perp} = \{E \in \mathcal{P}_n \text{ s.t. } [E, M] = 0\ \forall M \in S\}$. Then $S$ encodes $k$ qubits and has distance $d$, where $d$ is the smallest weight of an operator in $S^{\perp} \backslash S$.*

We use the notation $[[n, k, d]]$ to a refer to such a stabilizer code. Note that the square brackets specify that the code is a stabilizer code, and that the middle term $k$ refers to the number of encoded qubits, and not the dimension $2^k$ of the encoded subspace, as for the general QECC (whose dimension might not be a power of 2).

$S^\perp$ is the set of Pauli operators that commute with all elements of the stabilizer. They would therefore appear to be those errors which cannot be detected by the code. However, the theorem specifies the distance of the code by considering $S^\perp \backslash S$. A Pauli operator $P \in S$ cannot be detected by the code, but there is in fact no need to detect it, since all code words remain fixed under $P$, making it equivalent to the identity operation. A distance $d$ stabilizer code which has nontrivial $P \in S$ with $\mathrm{wt}(P) < d$ is called degenerate, whereas one which does not is nondegenerate. The phenomenon of degeneracy has no analog for classical error-correcting codes, and makes the study of quantum codes substantially more difficult than the study of classical error correction. For instance, a standard bound on classical error correction is the Hamming bound (or sphere-packing bound), but the analogous quantum Hamming bound

$$k/n \leq 1 - (t/n)\log 3 - h(t/n) \qquad [7]$$

for $[[n, k, 2t+1]]$ codes (when $n$ is large) is only known to apply to nondegenerate quantum codes (though in fact we do not know of any degenerate QECCs that violate the quantum Hamming bound).

An example of a stabilizer code is the 5-qubit code, a [[5,1,3]] code whose stabilizer can be generated by

$$\begin{array}{c} X\otimes Z\otimes Z\otimes X\otimes I \\ I\otimes X\otimes Z\otimes Z\otimes X \\ X\otimes I\otimes X\otimes Z\otimes Z \\ Z\otimes X\otimes I\otimes X\otimes Z \end{array}$$

The 5-qubit code is a nondegenerate code, and is the smallest possible QECC which corrects 1 error (as one can see from the quantum Singleton bound).

It is frequently useful to consider other representations of stabilizer codes. For instance, $P \in \mathcal{P}_n$ can be represented by a pair of $n$-bit binary vectors $(p_X \mid p_Z)$, where $p_X$ is 1 for any location where $P$ has an $X$ or $Y$ tensor factor and is 0 elsewhere, and $p_Z$ is 1 for any location where $P$ has a $Y$ or $Z$ tensor factor. Two Pauli operators $P = (p_X|p_Z)$ and $Q = (q_X|q_Z)$ commute iff $p_X \cdot q_Z + p_Z \cdot q_X = 0$. Then the stabilizer for a code becomes a pair of $(n-k) \times n$ binary matrices, and most interesting properties can be determined by an appropriate linear algebra exercise. Another useful representation is to map the single-qubit Pauli operators $I$, $X$, $Y$, $Z$ to the finite field GF(4), which sets up a connection between stabilizer codes and a subset of classical codes on four-dimensional registers.

## CSS Codes

CSS codes are a very useful class of stabilizer codes invented by Calderbank and Shor (1996), and by Steane (1996). The construction takes two binary classical linear codes and produces a quantum code, and can therefore take advantage of much existing knowledge from classical coding theory. In addition, CSS codes have some very useful properties which make them excellent choices for fault-tolerant quantum computation.

A classical $[n, k, d]$ linear code ($n$ physical bits, $k$ logical bits, classical distance $d$) can be defined in terms of an $(n-k) \times n$ binary "parity check" matrix $H$ – every classical code word $v$ must satisfy $Hv = 0$. Each row of the parity check matrix can be converted into a Pauli operator by replacing each 0 with an $I$ operator and each 1 with a $Z$ operator. Then the stabilizer code generated by these operators is precisely a quantum version of the classical error-correcting code given by $H$. If the classical distance $d = 2t+1$, the quantum code can correct $t$ bit flip ($X$) errors, just as could the classical code.

If we want to make a QECC that can also correct phase ($Z$) errors, we should choose two classical codes $C_1$ and $C_2$, with parity check matrices $H_1$ and $H_2$. Let $C_1$ be an $[n, k_1, d_1]$ code and let $C_2$ be an $[n, k_2, d_2]$ code. We convert $H_1$ into stabilizer generators as above, replacing each 0 with $I$ and each 1 with $Z$. For $H_2$, we perform the same procedure, but each 1 is instead replaced by $X$. The code will be able to correct bit flip ($X$) errors as if it had a distance $d_1$ and to correct phase ($Z$) errors as if it had a distance $d_2$. Since these two operations are completely separate, it can also correct $Y$ errors as both a bit flip and a phase error. Thus, the distance of the quantum code is at least $\min(d_1, d_2)$, but might be higher because of the possibility of degeneracy.

However, in order to have a stabilizer code at all, the generators produced by the above procedure must commute. Define the dual $C^\perp$ of a classical code $C$ as the set of vectors $w$ s.t. $w \cdot v = 0$ for all $v \in C$. Then the $Z$ generators from $H_1$ will all commute with the $X$ generators from $H_2$ iff $C_2^\perp \subseteq C_1$ (or equivalently, $C_1^\perp \subseteq C_2$). When this is true, $C_1$ and $C_2$ define an $[[n, k_1 + k_2 - n, d]]$ stabilizer code, where $d \geq \min(d_1, d_2)$.

The smallest distance-3 CSS code is the 7-qubit code, a [[7, 1, 3]] QECC created from the classical Hamming code (consisting of all sums of classical strings 1111000, 1100110, 1010101, and 1111111). The encoded $|\bar{0}\rangle$ for this code consists of the superposition of all even-weight classical code words and the encoded $|\bar{1}\rangle$ is the superposition of all odd-weight classical code words. The 7-qubit code is much studied because its properties make it particularly well suited to fault-tolerant quantum computation.

## Fault Tolerance

Given a QECC, we can attempt to supplement it with protocols for performing fault-tolerant operations. The basic design principle of a fault-tolerant protocol is that an error in a single location – either a faulty gate or noise on a quiescent qubit – should not be able to alter more than a single qubit in each block of the QECC. If this condition is satisfied, $t$ separate single-qubit or single-gate failures are required for a distance $2t + 1$ code to fail.

Particular caution is necessary, as computational gates can cause errors to propagate from their original location onto qubits that were previously correct. In general, a gate coupling pairs of qubits allows errors to spread in both directions across the coupling.

The solution is to use transversal gates whenever possible (Shor 1996). A transversal operation is one in which the $i$th qubit in each block of a QECC interacts only with the $i$th qubit of other blocks of the code or of special ancilla states. An operation consisting only of single-qubit gates is automatically transversal. A transversal operation has the virtue that an error occurring on the third qubit in a block, say, can only ever propagate to the third qubit of other blocks of the code, no matter what other sequence of gates we perform before a complete error-correction procedure.

In the case of certain codes, such as the 7-qubit code, a number of different gates can be performed transversally. Unfortunately, it does not appear to be possible to perform universal quantum computations using just transversal gates. We therefore have to resort to more complicated techniques. First we create special encoded ancilla states in a non-fault-tolerant way, but perform some sort of check on them (in addition to error correction) to make sure they are not too far off from the goal. Then we interact the ancilla with the encoded data qubits using gates from our stock of transversal gates and perform a fault-tolerant measurement. Then we complete the operation with a further transversal gate which depends on the outcome of the measurement.

## Fault-Tolerant Gates

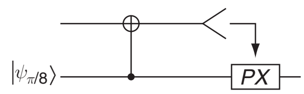
We will focus on stabilizer codes. Universal fault tolerance is known to be possible for any stabilizer code, but in most cases the more complicated type of construction is needed for all but a few gates. The Pauli group $\mathcal{P}_k$, however, can be performed transversally on any stabilizer code. Indeed, the set $S^{\perp}\backslash S$ of undetectable errors is a boon in this case, as it allows us to perform these gates. In particular, each coset $S^{\perp}/S$ corresponds to a different logical Pauli operator (with $S$ itself corresponding to the identity). On a stabilizer code, therefore, logical Pauli operations can be performed via a transversal Pauli operation on the physical qubits.

Stabilizer codes have a special relationship to a finite subgroup $\mathcal{C}_n$ of the unitary group $U(2^n)$ frequently called the "Clifford group." The Clifford group on $n$ qubits is defined as the set of unitary operations which conjugate the Pauli group $\mathcal{P}_n$ into itself; $\mathcal{C}_n$ can be generated by the Hadamard transform, the controlled-NOT (CNOT), and the single-qubit $\pi/4$ phase rotation diag(1, i). The set of stabilizer codes is exactly the set of codes which can be created by a Clifford group encoder circuit using $|0\rangle$ ancilla states.

Some stabilizer codes have interesting symmetries under the action of certain Clifford group elements, and these symmetries result in transversal gate operations. A particularly useful fact is that a transversal CNOT gate (i.e., CNOT acting between the $i$th qubit of one block of the QECC and the $i$th qubit of a second block for all $i$) acts as a logical CNOT gate on the encoded qubits for any CSS code. Furthermore, for the 7-qubit code, transversal Hadamard performs a logical Hadamard, and the transversal $\pi/4$ rotation performs a logical $-\pi/4$ rotation. Thus, for the 7-qubit code, the full logical Clifford group is accessible via transversal operations.

Unfortunately, the Clifford group by itself does not have much computational power: it can be efficiently simulated on a classical computer. We need to add some additional gate outside the Clifford group to allow universal quantum computation; a single gate will suffice, such as the single-qubit $\pi/8$ phase rotation diag(1, $\exp(i\pi/4)$). Note that this gives us a finite generating set of gates. However, by taking appropriate products, we get an infinite set of gates, one that is dense in the unitary group $U(2^n)$, allowing universal quantum computation.

The following circuit performs a $\pi/8$ rotation, given an ancilla state $|\psi_{\pi/8}\rangle = |0\rangle + \exp(i\pi/4)|1\rangle$:




Here $P$ is the $\pi/4$ phase rotation diag(1, i), and $X$ is the bit flip. The product is in the Clifford group, and is only performed if the measurement outcome is 1. Therefore, given the ability to perform fault-tolerant Clifford group operations, fault-tolerant measurements, and to prepare the encoded $|\psi_{\pi/8}\rangle$ state, we have universal fault-tolerant quantum computation. A slight generalization of the fault-tolerant measurement procedure below can be used to fault-tolerantly verify the $|\psi_{\pi/8}\rangle$ state, which is a +1 eigenstate of $PX$. Using this or another verification procedure, we can check a non-fault-tolerant construction.

## Fault-Tolerant Measurement and Error Correction

Since all our gates are unreliable, including those used to correct errors, we will need some sort of fault-tolerant quantum error-correction procedure. A number of different techniques have been developed. All of them share some basic features: they involve creation and verification of specialized ancilla states, and use transversal gates which interact the data block with the ancilla state.

The simplest method, due to Shor, is very general but also requires the most overhead and is frequently the most susceptible to noise. Note that the following procedure can be used to measure (non-fault-tolerantly) the eigenvalue of any (possibly multiqubit) Pauli operator $M$: produce an ancilla qubit in the state $|+\rangle = |0\rangle + |1\rangle$. Perform a controlled-$M$ operation from the ancilla to the state being measured. In the case where $M$ is a multiqubit Pauli operator, this can be broken down into a sequence of controlled-$X$, controlled-$Y$, and controlled-$Z$ operations. Then measure the ancilla in the basis of $|+\rangle$ and $|-\rangle = |0\rangle - |1\rangle$. If the state is a +1 eigenvector of $M$, the ancilla will be $|+\rangle$, and if the state is a –1 eigenvector, the ancilla will be $|-\rangle$.

The advantage of this procedure is that it measures just $M$ and nothing more. The disadvantage is that it is not transversal, and thus not fault tolerant. Instead of the unencoded $|+\rangle$ state, we must use a more complex ancilla state $|00\ldots0\rangle + |11\ldots1\rangle$ known as a "cat" state. The cat state contains as many qubits as the operator $M$ to be measured, and we perform the controlled-$X$, -$Y$, or -$Z$ operations transversally from the appropriate qubits of the cat state to the appropriate qubits in the data block. Since, assuming the cat state is correct, all of its qubits are either $|0\rangle$ or $|1\rangle$, the procedure either leaves the data state alone or performs $M$ on it uniformly. A +1 eigenstate in the data therefore leaves us with $|00\ldots0\rangle + |11\ldots1\rangle$ in the ancilla and a –1 eigenstate leaves us with $|00\ldots0\rangle - |11\ldots1\rangle$. In either case, the final state still tells us nothing about the data beyond the eigenvalue of $M$. If we perform a Hadamard transform and then measure each qubit in the ancilla, we get either a random even-weight string (for eigenvalue +1) or an odd-weight string (for eigenvalue –1).

The procedure is transversal, so an error on a single qubit in the initial cat state or in a single gate during the interaction will only produce one error in the data. However, the initial construction of the cat state is not fault tolerant, so a single-gate error then could eventually produce two errors in the data block. Therefore, we must be careful and use some sort of technique to verify the cat state, for instance, by checking if random pairs of qubits are the same. Also, note that a single phase error in the cat state will cause the final measurement outcome to be wrong (even and odd switch places), so we should repeat the measurement procedure multiple times for greater reliability.

We can then make a full fault-tolerant error-correction procedure by performing the above measurement technique for each generator of the stabilizer. Each measurement gives us one bit of the error syndrome, which we then decipher classically to determine the actual error.

More sophisticated techniques for fault-tolerant error correction involve less interaction with the data but at the cost of more complicated ancilla states. A procedure due to Steane uses (for CSS codes) one ancilla in a logical $|\bar{0}\rangle$ state of the same code and one ancilla in a logical $|\bar{0}\rangle + |\bar{1}\rangle$ state. A procedure due to Knill (for any stabilizer code) teleports the data qubit through an ancilla consisting of two blocks of the QECC containing an encoded Bell state $|\overline{00}\rangle + |\overline{11}\rangle$. Because the ancillas in Steane and Knill error correction are more complicated than the cat state, it is especially important to verify the ancillas before using them.

## The Threshold for Fault Tolerance

In an unencoded protocol, even one error can destroy the computation, but a fully fault-tolerant protocol will give the right answer unless multiple

errors occur before they can be corrected. On the other hand, the fault-tolerant protocol is larger, requiring more qubits and more time to do each operation, and therefore providing more opportunities for errors. If errors occur on the physical qubits independently at random with probability $p$ per gate or time step, the fault-tolerant protocol has probability of logical error for a single logical gate or time step at most $Cp^2$, where $C$ is a constant that depends on the design of the fault-tolerant circuitry (assume the QECC has distance 3, as for the 7-qubit code). When $p < p_t = 1/C$, the fault tolerance helps, decreasing the logical error rate. $p_t$ is the "threshold" for fault-tolerant quantum computation. If the error rate is higher than the threshold, the extra overhead means that errors will occur faster than they can be reliably corrected, and we are better off with an unencoded system.

To further lower the logical error rate, we turn to a family of codes known as "concatenated codes" (Aharonov and Ben-Or, Kitaev 1997, Knill *et al.* 1998). Given a code word of a particular $[[n, 1]]$ QECC, we can take each physical qubit and again encode it using the same code, producing an $[[n^2, 1]]$ QECC. We could repeat this procedure to get an $n^3$-qubit code, and so forth. The fault-tolerant procedures concatenate as well, and after $L$ levels of concatenation, the effective logical error rate is $p_t(p/p_t)^{2^L}$ (for a base code correcting 1 error). Therefore, if $p$ is below the threshold $p_t$, we can achieve an arbitrarily good error rate $\epsilon$ per logical gate or time step using only poly$(\log \epsilon)$ resources, which is excellent theoretical scaling.

Unfortunately, the practical requirements for this result are not nearly so good. The best rigorous proofs of the threshold to date show that the threshold is at least $2 \times 10^{-5}$ (meaning one error per 50,000 operations). Optimized simulations of fault-tolerant protocols suggest that the true threshold may be as high as 5%, but to tolerate this much error, existing protocols require enormous overhead, perhaps increasing the number of gates and qubits by a factor of a million or more for typical computations. For lower physical error rates, overhead requirements are more modest, particularly if we only attempt to optimize for calculations of a given size, but are still larger than one would like.

Furthermore, these calculations make a number of assumptions about the physical properties of the computer. The errors are assumed to be independent and uncorrelated between qubits except when a gate connects them. It is assumed that measurements and classical computations can be performed quickly and reliably, and that quantum gates can be performed between arbitrary pairs of qubits in the computer, irrespective of their physical proximity. Of these, only the assumption of independent errors is at all necessary, and that can be considerably relaxed to allow short-range correlations and certain kinds of non-Markovian environments. However, the effects of relaxing these assumptions on the threshold value and overhead requirements have not been well studied.

## Further Reading


Aharonov D and Ben-Or M (1999) Fault-tolerant quantum computation with constant error rate, quant-ph/9906129.

Bennett C, DiVincenzo D, Smolin J, and Wootters W (1996) Mixed state entanglement and quantum error correction. *Physical Review A* 54: 3824–3851 (quant-ph/9604024).

Calderbank AR and Shor PW (1996) Good quantum error-correcting codes exist. *Physical Review A* 54: 1098–1105 (quant-ph/9512032).

Calderbank AR, Rains EM, Shor PW, and Sloane NJA (1998) Quantum error correction via codes over GF(4). *IEEE Transactions on Information Theory* 44: 1369–1387 (quant-ph/9605005).

Gottesman D (1996) Class of quantum error-correcting codes saturating the quantum Hamming bound. *Physical Review A* 54: 1862–1868 (quant-ph/9604038).

Kitaev AY (1997) Quantum error correction with imperfect gates. In: Hirota O, Holeva AS, and Caves CM (eds.) *Quantum Communication, Computing, and Measurement (Proc. 3rd Int. Conf. of Quantum Communication and Measurement)*, pp. 181–188. New York: Plenum.

Knill E and Laflamme R (1997) A theory of quantum error-correcting codes. *Physical Review A* 55: 900–911 (quant-ph/9604034).

Knill E, Laflamme R, and Zurek WH (1998) Resilient quantum computation. *Science* 279: 342–345.

Shor PW (1996) Fault-tolerant quantum computation. In: *Proc. 35th Ann. Symp. on Fundamentals of Computer Science*, pp. 56–65. (quant-ph/9605011). Los Alamitos: IEEE Press.

Steane AM (1996) Multiple particle interference and quantum error correction. *Proceedings of the Royal Society of London A* 452: 2551–2577 (quant-ph/9601029).

# Quantum Field Theory in Curved Spacetime

**B S Kay**, University of York, York, UK



## Introduction and Preliminaries

Quantum Field Theory (QFT) in curved spacetime is a hybrid approximate theory in which quantum matter fields are assumed to propagate in a fixed classical background gravitational field. Its basic physical prediction is that strong gravitational fields can polarize the vacuum and, when time dependent, lead to pair creation just as a strong and/or time-dependent electromagnetic field can polarize the vacuum and/or give rise to pair creation of charged particles. One expects it to be a good approximation to full quantum gravity provided the typical frequencies of the gravitational background are very much less than the Planck frequency $(c^5/G\hbar)^{1/2} \sim 10^{43}\,\mathrm{s}^{-1}$ and provided, with a suitable measure for energy, the energy of created particles is very much less than the energy of the background gravitational field or of its matter sources. Undoubtedly, the most important prediction of the theory is the Hawking effect, according to which a, say spherically symmetric, classical black hole of mass $M$ will emit thermal radiation at the Hawking temperature $T = (8\pi M)^{-1}$ (here and from now on, we use Planck units where $G, c, \hbar$ and, $k$ (Boltzmann's constant) are all taken to be 1).

On the mathematical side, the need to formulate the laws and derive the general properties of QFT on nonflat spacetimes forces one to state and prove results in local terms and, as a byproduct, thereby leads to an improved perspective on flat-spacetime QFT too. It is also interesting to formulate QFT on idealized spacetimes with particular global geometrical features. Thus, QFT on spacetimes with bifurcate Killing horizons is intimately related to the Hawking effect; QFT on spacetimes with closed timelike curves is intimately related to the question whether the laws of physics permit the manufacture of a time machine.

As is standard in general relativity, a curved spacetime is modeled mathematically as a (paracompact, Hausdorff) manifold $\mathcal{M}$ equipped with a pseudo-Riemannian metric $g$ of signature $(-,+++)$ (we follow the conventions of the standard text by Misner *et al.* (1973)). We shall also assume, except where otherwise stated, our spacetime to be *globally hyperbolic*, that is, that $\mathcal{M}$ admits a *global time coordinate*, by which we mean a global coordinate $t$ such that each constant-$t$ surface is a smooth Cauchy surface, that is, a smooth spacelike 3-surface cut exactly once by each inextendible causal curve. (Without this default assumption, extra problems arise for QFT which we shall briefly mention in connection with the "time machine" question discussed later.) In view of this definition, globally hyperbolic spacetimes are clearly time-orientable and we shall assume a choice of time-orientation has been made so we can talk about the "future" and "past" directions. Modern formulations of the subject take, as the fundamental mathematical structure modeling the quantum field, a $*$-algebra $\mathcal{A}$ (with identity $I$) together with a family of local sub $*$-algebras $\mathcal{A}(\mathcal{O})$ labeled by bounded open regions $\mathcal{O}$ of the spacetime $(\mathcal{M}, g)$ and satisfying the *isotony* or *net* condition that $\mathcal{O}_1 \subset \mathcal{O}_2$ implies $\mathcal{A}(\mathcal{O}_1)$ is a subalgebra of $\mathcal{A}(\mathcal{O}_2)$ as well as the condition that whenever two bounded open regions $\mathcal{O}_1$ and $\mathcal{O}_2$ are spacelike separated, then $\mathcal{A}(\mathcal{O}_1)$ and $\mathcal{A}(\mathcal{O}_2)$ commute.

Standard concepts and techniques from algebraic quantum theory are then applicable: In particular, *states* are defined to be positive (this means $\omega(A^*A) \geq 0\ \forall A \in \mathcal{A}$) normalized (this means $\omega(I) = 1$) linear functionals on $\mathcal{A}$. One distinguishes between *pure* states and *mixed* states, only the latter being writable as nontrivial convex combinations of other states. To each state, $\omega$, the *GNS construction* associates a representation, $\rho_\omega$, of $\mathcal{A}$ on a Hilbert space $\mathcal{H}_\omega$ together with a cyclic vector $\Omega \in \mathcal{H}_\omega$ such that

$$\omega(A) = \langle \Omega | \rho_\omega(A) \Omega \rangle$$

(and the *GNS triple* $(\rho_\omega, \mathcal{H}, \Omega)$ is unique up to equivalence). There are often technical advantages in formulating things so that the $*$-algebra is a $C^*$-algebra. Then the GNS representation is as everywhere-defined bounded operators and is irreducible if and only if the state is pure. A useful concept, due to Haag, is the *folium* of a given state $\omega$ which may be defined to be the set of all states $\omega_\sigma$ which arise in the form $\mathrm{tr}(\sigma\rho_\omega(\cdot))$, where $\sigma$ ranges over the density operators (trace-class operators with unit trace) on $\mathcal{H}_\omega$.

Given a state, $\omega$, and an automorphism, $\alpha$, which preserves the state (i.e., $\omega \circ \alpha = \omega$) then there will be a unitary operator, $U$, on $H_\omega$ which *implements* $\alpha$ in the sense that $\rho_\omega(\alpha(A)) = U^{-1}\rho_\omega(A)U$ and $U$ is chosen uniquely by the condition $U\Omega = \Omega$.

On a *stationary* spacetime, that is, one which admits a one-parameter group of isometries whose integral curves are everywhere timelike, the algebra will inherit a one-parameter group (i.e., satisfying $\alpha(t_1) \circ \alpha(t_2) = \alpha(t_1 + t_2)$) of time-translation

automorphisms, $\alpha(t)$, and, given any stationary state (i.e., one which satisfies $\omega \circ \alpha(t) = \omega \ \forall t \in \boldsymbol{R}$), these will be implemented by a one-parameter group of unitaries, $U(t)$, on its GNS Hilbert space satisfying $U(t)\Omega = \Omega$. If $U(t)$ is strongly continuous so that it takes the form $e^{-iHt}$ and if the Hamiltonian, $H$, is positive, then $\omega$ is said to be a "ground state." Typically one expects *ground states* to exist and often be unique.

Another important class of stationary states for the algebra of a stationary spacetime is the class of *KMS states*, $\omega^\beta$, at inverse temperature $\beta$; these have the physical interpretation of thermal equilibrium states. In the GNS representation of one of these, the automorphisms are also implemented by a strongly continuous unitary group, $e^{-iHt}$, which preserves $\Omega$ but (in place of $H$ positive) there is a complex conjugation, $J$, on $H_\omega$ such that

$$e^{-\beta H/2}\rho_\omega(A)\Omega = J\rho_\omega(A^*)\Omega \qquad [1]$$

for all $A \in \mathcal{A}$. An attractive feature of the subject is that its main qualitative features are already present for linear field theories and, unusually in comparison with other questions in QFT, these are susceptible of a straightforward explicit and rigorous mathematical formulation. In fact, as our principal example, we give, in the next section a construction for the field algebra for the quantized real linear Klein–Gordon equation

$$(\Box_g - m^2 - V)\phi = 0 \qquad [2]$$

of mass $m$ on a globally hyperbolic spacetime $(\mathcal{M}, g)$. Here, $\Box_g$ denotes the Laplace–Beltrami operator $g^{ab}\nabla_a\partial_b$ $(= (|\det(g)|)^{-1/2}\partial_a((|\det(g)|^{1/2}g^{ab}\partial_b))$. We include a scalar external background classical field, $V$, in addition to the external gravitational field represented by $g$. In case $m$ is zero, taking $V$ to equal $R/6$, where $R$ denotes the Riemann scalar, makes the equation *conformally invariant*.

The main new feature of QFT in curved spacetime (present already for linear field theories) is that, in a general (neither flat nor stationary) spacetime there will not be any single preferred state but rather a family of preferred states, members of which are best regarded as on an equal footing with one another. It is this feature which makes the above algebraic framework particularly suitable, indeed essential, to a clear formulation of the subject. Conceptually, it is this feature which takes the most getting used to. In particular, one must realize that, as we shall explain later, the interpretation of a state as having a particular "particle content" is in general problematic because it can only be relative to a particular choice of "vacuum" state and, depending on the spacetime of interest, there may be one state or several states or, frequently, no states at all which deserve the name "vacuum" and even when there are states which deserve this name, they will often only be defined in some approximate or asymptotic or transient sense or only on some subregion of the spacetime.

Concomitantly, one does not expect global observables such as the "particle number" or the quantum Hamiltonian of flat-spacetime free-field theory to generalize to a curved spacetime context, and for this reason local observables play a central role in the theory. The quantized stress–energy tensor is a particularly natural and important such local observable and the theory of this is central to the whole subject. A brief introduction to it is given in a later section.

This is followed by a further section on the Hawking and Unruh effects and then a brief section on the problems of extending the theory beyond the "default" setting, to nonglobally hyperbolic spacetimes. Finally, we briefly mention a number of other interesting and active areas of the subject as well as issuing a few *warnings* to be borne in mind when reading the literature.

## Construction of $*$-Algebra(s) for a Real Linear Scalar Field on Globally Hyperbolic Spacetimes and Some General Theorems

On a globally hyperbolic spacetime, the classical equation [2] admits well-defined *advanced* and *retarded Green* functions (strictly bidistributions) $\Delta^A$ and $\Delta^R$ and the standard covariant quantum free real (or "Hermitian") scalar field commutation relations familiar from Minkowski spacetime free-field theory naturally generalize to the (heuristic) equation

$$[\hat{\phi}(x), \hat{\phi}(y)] = i\Delta(x, y)I$$

where $\Delta$ is the *Lichnérowicz commutator function* $\Delta = \Delta^A - \Delta^R$. Here, the "$\hat{}$" on the quantum field $\hat{\phi}$ serves to distinguish it from a classical solution $\phi$. In mathematical work, one does not assign a meaning to the field at a point itself, but rather aims to assign meaning to *smeared fields* $\hat{\phi}(F)$ for all real-valued test functions $F \in C_0^\infty(\mathcal{M})$ which are then to be interpreted as standing for $\int_\mathcal{M} \hat{\phi}(x)f(x)|\det(g)|^{1/2}\, d^4x$. In fact, it is straightforward to define a *minimal field algebra* (see below) $\mathcal{A}_{\min}$ generated by such $\hat{\phi}(F)$ which satisfy the suitably smeared version

$$[\hat{\phi}(F), \hat{\phi}(G)] = i\Delta(F, G)I$$

of the above commutation relations together with Hermiticity (i.e., $\hat{\phi}(F)^* = \hat{\phi}(F)$), the property of being a weak solution of eqn [2] (i.e., $\hat{\phi}((\Box_g - m^2 - V)F) = 0 \ \forall F \in C_0^\infty(\mathcal{M})$) and linearity in test functions. There is a technically different alternative formulation of this minimal algebra, which is known as the *Weyl algebra*, which is constructed to be the C*-algebra generated by operators $W(F)$ (to be interpreted as standing for $\exp(\mathrm{i}\int_{\mathcal{M}} \hat{\phi}(x) f(x) |\det(g)|^{1/2} \mathrm{d}^4x)$ satisfying

$$W(F_1)W(F_2) = \exp(-\mathrm{i}\Delta(F_1, F_2)/2)W(F_1 + F_2)$$

together with $W(F)^* = W(-F)$ and $W((\Box_g - m^2 - V)F) = I$. With either the minimal algebra or the Weyl algebra one can define, for each bounded open region $\mathcal{O}$, subalgebras $A(\mathcal{O})$ as generated by the $\hat{\phi}(\cdot)$ (or the $W(\cdot)$) smeared with test functions supported in $\mathcal{O}$ and verify that they satisfy the above "net" condition and commutativity at spacelike separation.

Specifying a state, $\omega$, on $\mathcal{A}_{\min}$ is tantamount to specifying its collection of $n$-point distributions (i.e., smeared $n$-point functions) $\omega(\hat{\phi}(F_1)\ldots\hat{\phi}(F_n))$. (In the case of the Weyl algebra, one restricts attention to "regular" states for which the map $F \to \omega(W(F))$ is sufficiently often differentiable on finite-dimensional subspaces of $C_0^\infty(\mathcal{M})$ and defines the $n$-point distributions in terms of derivatives with respect to suitable parameters of expectation values of suitable Weyl algebra elements.) A particular role is played in the theory by the *quasifree* states for which all the *truncated* $n$-point distributions except for $n=2$ vanish. Thus, all the $n$-point distributions for odd $n$ vanish while the four-point distribution is made out of the two-point distribution according to

$$\begin{aligned}\omega(\hat{\phi}(F_1)\hat{\phi}(F_2)\hat{\phi}(F_3)\hat{\phi}(F_4)) \\ = \omega(\hat{\phi}(F_1)\hat{\phi}(F_2))\omega(\hat{\phi}(F_3)\hat{\phi}(F_4)) \\ + \omega(\hat{\phi}(F_1)\hat{\phi}(F_3))\omega(\hat{\phi}(F_2)\hat{\phi}(F_4)) \\ + \omega(\hat{\phi}(F_1)\hat{\phi}(F_4))\omega(\hat{\phi}(F_2)\hat{\phi}(F_3))\end{aligned}$$

etc. The anticommutator distribution

$$G(F_1, F_2) = \omega(\hat{\phi}(F_1)\hat{\phi}(F_2)) + \omega(\hat{\phi}(F_2)\hat{\phi}(F_1)) \qquad [3]$$

of a quasifree state (or indeed of any state) will satisfy the following conditions (for all test functions $F, F_1, F_2$, etc.):

C1. *Symmetry*

$$G(F_1, F_2) = G(F_2, F_1)$$

C2. *Weak bisolution property*

$$\begin{aligned}G((\Box_g - m^2 - V)F_1, F_2) = 0 \\ = G(F_1, (\Box_g - m^2 - V)F_2)\end{aligned}$$

C3. *Positivity*

$$\begin{aligned}G(F, F) \geq 0 \text{ and } G(F_1, F_1)^{1/2} G(F_2, F_2)^{1/2} \\ \geq |\Delta(F_1, F_2)|\end{aligned}$$

and it can be shown that, to every bilinear functional $G$ on $C_0^\infty(\mathcal{M})$ satisfying (C1)–(C3), there is a quasifree state with two-point distribution $(1/2)(G + \mathrm{i}\Delta)$. One further declares a quasifree state to be *physically admissible* only if (for pairs of points in sufficiently small convex neighborhoods)

C4. *Hadamard condition*

$$\text{“}G(x_1, x_2) = \frac{1}{2\pi^2}\left(u(x_1, x_2)\mathrm{P}\frac{1}{\sigma} + v(x_1, x_2)\log|\sigma| + w(x_1, x_2)\right)\text{”}$$

This last condition expresses the requirement that (locally) the two-point distribution actually "is" (in the usual sense in which one says that a distribution "is" a function) a smooth function for pairs of non-null-separated points. At the same time, it requires that the two-point distribution be singular at pairs of null-separated points and locally specifies the nature of the singularity for such pairs of points with a leading "principal part of $1/\sigma$" type singularity and a subleading "$\log|\sigma|$" singularity, where $\sigma$ denotes the square of the geodesic distance between $x_1$ and $x_2$. $u$ (which satisfies $u(x_1, x_2) = 1$ when $x_1 = x_2$) and $v$ are certain smooth two-point functions determined in terms of the local geometry and the local values of $V$ by something called the *Hadamard procedure* while the smooth two-point function $w$ depends on the state. We shall omit the details. The important point is that this Hadamard condition on the two-point distribution is believed to be the correct generalization to a curved spacetime of the well-known universal short-distance behavior shared by the truncated two-point distributions of all physically relevant states for the special case of our theory when the spacetime is flat (and $V$ vanishes). In the latter case, $u$ reduces to 1, and $v$ to a simple power series $\sum_{n=0}^{\infty} v_n \sigma^n$ with $v_0 = m^2/4$, etc.

Actually, it is known (this is the content of "Kay's conjecture" which was proved by M Radzikowski in 1992) that (C1)–(C4) together imply that the two-point distribution is nonsingular at all pairs of (not necessarily close together) spacelike separated points. More important than this result itself is a reformulation of the Hadamard condition in terms of the concepts of *microlocal analysis* which Radzikowski originally introduced as a tool towards its proof.

C4′. *Wave front set* (*or microlocal*) *spectrum condition*

$$\begin{aligned}&\mathrm{WF}(G+\mathrm{i}\Delta)\\&=\{(x_1,p_1;x_2,p_2)\in T^*(\mathcal{M}\times\mathcal{M})\setminus \mathbf{0}|x_1 \text{ and } x_2\\&\quad\text{lie on a single null geodesic, } p_1 \text{ is tangent to}\\&\quad\text{that null geodesic and future pointing, and}\\&\quad p_2 \text{ when parallel transported along that null}\\&\quad\text{geodesic from } x_2 \text{ to } x_1 \text{ equals } -p_1\}\end{aligned}$$

For the gist of what this means, it suffices to know that to say that an element $(x,p)$ of the cotangent bundle of a manifold (excluding the zero section $\mathbf{0}$) is in the *wave front set*, WF, of a given distribution on that manifold may be expressed informally by saying that that distribution is singular *at* the point $x$ *in the direction* $p$. (And here the notion is applied to $G+\mathrm{i}\Delta$, thought of as a distribution on the manifold $\mathcal{M}\times\mathcal{M}$.)

We remark that generically (and, e.g., always if the spatial sections are compact and $m^2+V(x)$ is everywhere positive) the Weyl algebra for eqn [2] on a given stationary spacetime will have a unique ground state and unique KMS states at each temperature and these will be quasifree and Hadamard.

Quasifree states are important also because of a theorem of R Verch (1994, in verification of another conjecture of Kay) that (in the Weyl algebra framework) on the algebra of any bounded open region, the folia of the quasifree Hadamard states coincide. With this result one can extend the notion of physical admissibility to not-necessarily-quasifree states by demanding that, to be admissible, a state belong to the resulting common folium when restricted to the algebra of each bounded open region; equivalently, that it be a *locally normal* state on the resulting natural extension of the net of local Weyl algebras to a net of local $W^*$-algebras.

## Particle Creation and the Limitations of the Particle Concept

Global hyperbolicity also entails that the Cauchy problem is well posed for the classical field equation [2] in the sense that for every Cauchy surface, $\mathcal{C}$, and every pair $(f,p)$ of Cauchy data in $C_0^\infty(\mathcal{C})$, there exists a unique solution $\phi$ in $C_0^\infty(\mathcal{M})$ such that $f=\phi|_{\mathcal{C}}$ and $p=|\det(g)|^{1/2}g^{ab}\partial_b\phi|_{\mathcal{C}}$. Moreover, $\phi$ has compact support on all other Cauchy surfaces. Given a global time coordinate $t$, increasing towards the future, foliating $\mathcal{M}$ into a family of constant-$t$ Cauchy surfaces, $\mathcal{C}_t$, and given a choice of global timelike vector field $\tau^a$ (e.g., $\tau^a=g^{ab}\partial_b t$) enabling one to identify all the $\mathcal{C}_t$, say with $\mathcal{C}_0$, by identifying points cut by the same integral curve of $\tau^a$, a single such classical solution $\phi$ may be pictured as a family $\{(f_t,p_t): t\in\boldsymbol{R}\}$ of time-evolving Cauchy data on $\mathcal{C}_0$. Moreover, since [2] implies, for each pair of classical solutions, $\phi_1,\phi_2$, the conservation (i.e., $\partial_a j^a=0$) of the current $j^a=|\det(g)|^{1/2}g^{ab}(\phi_1\partial_b\phi_2-\phi_2\partial_b\phi_1)$, the symplectic form (on $C_0^\infty(\mathcal{C})\times C_0^\infty(\mathcal{C})$)

$$\sigma((f_t^1,p_t^1);(f_t^2,p_t^2))=\int_{\mathcal{C}_0}(f_t^1p_t^2-p_t^1f_t^2)\mathrm{d}^3x$$

will be conserved in time.

Corresponding to this picture of classical dynamics, one expects there to be a description of quantum dynamics in terms of a family of sharp-time quantum fields $(\varphi_t,\pi_t)$ on $\mathcal{C}_0$, satisfying heuristic canonical commutation relations

$$\begin{aligned}[\varphi_t(\boldsymbol{x}),\varphi_t(\boldsymbol{y})]&=0\\ [\pi_t(\boldsymbol{x}),\pi_t(\boldsymbol{y})]&=0\\ [\varphi_t(\boldsymbol{x}),\pi_t(\boldsymbol{y})]&=\mathrm{i}\delta^3(\boldsymbol{x},\boldsymbol{y})I\end{aligned}$$

and evolving in time according to the same dynamics as the Cauchy data of a classical solution. (Both these expectations are correct because the field equation is linear.) An elegant way to make rigorous mathematical sense of these expectations is in terms of a $*$-algebra with identity generated by Hermitian objects "$\sigma((\varphi_0,\pi_0);(f,p))$" ("symplectically smeared sharp-time fields at $t=0$") satisfying linearity in $f$ and $p$ together with the commutation relations

$$\begin{aligned}&[\sigma((\varphi_0,\pi_0);(f^1,p^1)),\sigma((\varphi_0,\pi_0);(f^2,p^2))]\\&\quad=\mathrm{i}\sigma((f^1,p^1);(f^2,p^2))I\end{aligned}$$

and to define (symplectically smeared) time-$t$ sharp-time fields by demanding

$$\sigma((\varphi_t,\pi_t);(f_t,p_t))=\sigma((\varphi_0,\pi_0);(f_0,p_0))$$

where $(f_t,p_t)$ is the classical time-evolute of $(f_0,p_0)$. This $*$-algebra of sharp-time fields may be identified with the (minimal) field $*$-algebra of the previous section, the $\hat\phi(F)$ of the previous section being identified with $\sigma((\varphi_0,\pi_0);(f,p))$, where $(f,p)$ are the Cauchy data at $t=0$ of $\Delta * F$. (This identification is of course many–one since $\hat\phi(F)=0$ whenever $F$ arises as $(\Box_g-m^2-V)G$ for some test function $G\in C_0^\infty(\mathcal{M})$.)

Specializing momentarily to the case of the free scalar field $(\Box-m^2)\phi=0$ $(m\neq 0)$ in Minkowski space with a flat $t=0$ Cauchy surface, the "symplectically smeared" two-point function of the usual ground state ("Minkowski vacuum state"), $\omega_0$, is given, in this formalism, by

$$\begin{aligned}&\omega_0(\sigma((\varphi,\pi);(f^1,p^1))\sigma((\varphi,\pi);(f^2,p^2)))\\&\quad=\tfrac{1}{2}(\langle f^1|\mu f^2\rangle+\langle p^1|\mu^{-1}p^2\rangle\\&\qquad+\mathrm{i}\sigma((f^1,p^1);(f^2,p^2)))\end{aligned}\qquad[4]$$

where the inner products are in the *one-particle Hilbert space* $\mathcal{H} = L^2_{\mathrm{C}}(\boldsymbol{R}^3)$ and $\mu = (m^2 - \nabla^2)^{1/2}$. The GNS representation of this state may be concretely realized on the familiar *Fock space* $\mathcal{F}(\mathcal{H})$ over $\mathcal{H}$ by

$$\rho_0(\sigma((\varphi, \pi); (f, p))) = -\mathrm{i}(\hat{a}^\dagger(\mathrm{a}) - (\hat{a}^\dagger(\mathrm{a}))^*)$$

where a denotes the element of $\mathcal{H}$:

$$\mathrm{a} = \frac{(\mu^{1/2} f + \mathrm{i}\mu^{-1/2} p)}{\sqrt{2}}$$

(we note in passing that, if we equip $\mathcal{H}$ with the symplectic form $2\,\mathrm{Im}\langle\cdot|\cdot\rangle$, then $K:(f,p)\mapsto \mathrm{a}$ is a symplectic map) and $\hat{a}^\dagger(\mathrm{a})$ is the usual smeared *creation operator* (= "$\int \hat{a}^\dagger(\boldsymbol{x})\mathrm{a}(\boldsymbol{x})\mathrm{d}^3\boldsymbol{x}$") on $\mathcal{F}(\mathcal{H})$ satisfying

$$[(\hat{a}^\dagger(\mathrm{a}^1))^*, \hat{a}^\dagger(\mathrm{a}^2)] = \langle \mathrm{a}^1 | \mathrm{a}^2 \rangle_{\mathcal{H}} I$$

The usual (smeared) *annihilation operator*, $\hat{a}(\mathrm{a})$, is $(\hat{a}^\dagger(C\mathrm{a}))^*$, where $C$ is the natural complex conjugation, $\mathrm{a} \mapsto \mathrm{a}^*$ on $\mathcal{H}$. Both of these operators annihilate the *Fock vacuum vector* $\Omega^{\mathcal{F}}$. In this representation, the one-parameter group of time-translation automorphisms

$$\alpha(t): \sigma((\varphi_0, \pi_0); (f, p)) \mapsto \sigma((\varphi_t, \pi_t); (f, p)) \quad [5]$$

is implemented by $\exp(-\mathrm{i}Ht)$ where $H$ is the second quantization of $\mu$ (i.e., the operator otherwise known as $\int \mu(k)\hat{a}^\dagger(k)\hat{a}(k)\mathrm{d}^3 k$) on $\mathcal{F}(\mathcal{H})$.

The most straightforward (albeit physically artificial) situation involving "particle creation" in a curved spacetime concerns a globally hyperbolic spacetime which, outside of a compact region, is isometric to Minkowski space with a compact region removed – that is, to a globally hyperbolic spacetime which is flat except inside a localized "bump" of curvature (see Figure 1). (One could also allow the function $V$ in [2] to be nonzero inside the bump.) On the field algebra (defined as in the previous section) of such a spacetime, there will be an "in" vacuum state (which may be identified with the Minkowski vacuum to the past of the bump) and an "out" vacuum state (which may be identified with the Minkowski vacuum to the future of the bump) and one expects, for example, the "in vacuum" to arise as a many-particle state in the GNS representation of the "out vacuum" corresponding to the creation of particles out of the vacuum by the bump of curvature.

In the formalism of this section, if we choose our global time coordinate on such a spacetime so that, say, the $t = 0$ surface is to the past of the bump and the $t = T$ surface to its future, then the single automorphism $\alpha(T)$ (defined as in [5]) encodes the overall effect of the bump of curvature on the quantum field and one can ask whether it is implemented by a unitary operator in the GNS representation of the Minkowski vacuum state [4].

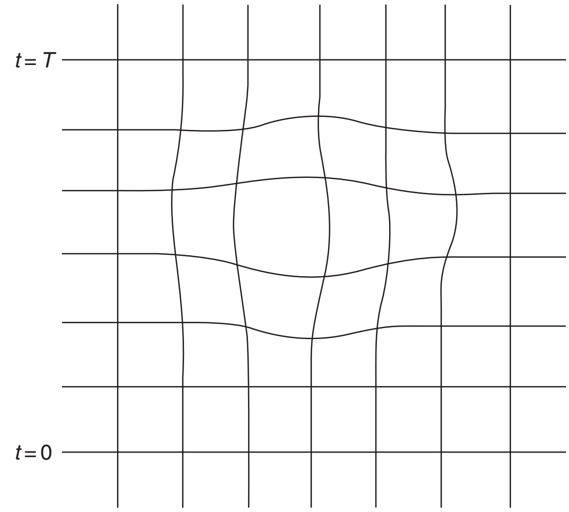



**Figure 1** A spacetime which is flat outside of a compact bump of curvature.

This question may be answered by referring to the real linear map $\mathcal{T}:\mathcal{H}\to\mathcal{H}$ which sends $\mathrm{a}_T = 2^{-1/2}(\mu^{1/2} f_T + \mathrm{i}\mu^{-1/2} p_T)$ to $\mathrm{a}_0 = 2^{-1/2}(\mu^{1/2} f_0 + \mathrm{i}\mu^{-1/2} p_0)$. By the conservation in time of $\sigma$ and the symplecticity, noted in passing above, of the map $K:(f,p)\mapsto \mathrm{a}$, this satisfies the defining relation

$$\mathrm{Im}\langle \mathcal{T}\mathrm{a}^1 | \mathcal{T}\mathrm{a}^2 \rangle = \mathrm{Im}\langle \mathrm{a}^1 | \mathrm{a}^2 \rangle$$

of a *classical Bogoliubov transformation*. Splitting $\mathcal{T}$ into its complex-linear and complex-antilinear parts by writing

$$\mathcal{T} = \alpha + \beta C$$

where $\alpha$ and $\beta$ are complex-linear operators, this relation may alternatively be expressed in terms of the pair of relations

$$\alpha^*\alpha - \bar{\beta}^*\bar{\beta} = I, \qquad \bar{\alpha}^*\bar{\beta} = \beta^*\alpha$$

where $\bar{\alpha} = C\alpha C$, $\bar{\beta} = C\beta C$.

We remark that there is an easy-to-visualize equivalent way of defining $\alpha$ and $\beta$ in terms of the analysis, to the past of the bump, into *positive- and negative-frequency parts* of complex solutions to [2] which are purely *positive frequency* to the future of the bump. In fact, if, for any element $\mathrm{a} \in \mathcal{H}$, we identify the positive-frequency solution to the Minkowski-space Klein–Gordon equation

$$\phi^{\mathrm{pos}}_{\mathrm{out}}(t, \boldsymbol{x}) = ((2\mu)^{-1/2} \exp(-\mathrm{i}\mu t)\mathrm{a})(\boldsymbol{x})$$

with a complex solution to [2] to the future of the bump, then (it may easily be seen) to the past of the bump, this same solution will be identifiable with

the (partly positive-frequency, partly negative-frequency) Minkowski-space Klein–Gordon solution

$$\phi_{\mathrm{in}}(t,\boldsymbol{x}) = \left((2\mu)^{-1/2}\exp(-\mathrm{i}\mu t)\alpha \mathrm{a}\right)(\boldsymbol{x}) + \left((2\mu)^{-1/2}\exp(\mathrm{i}\mu t)\bar{\beta}\mathrm{a}\right)(\boldsymbol{x})$$

and this could be taken to be the defining equation for the operators $\alpha$ and $\beta$.

It is then known (by a 1962 theorem of Shale) that the automorphism [5] (strictly, its Weyl algebra counterpart) will be unitarily implemented if and only if $\beta$ is a Hilbert–Schmidt operator on $\mathcal{H}$. Wald (1979, in case $m \geq 0$) and Dimock (1979, in case $m \neq 0$) have verified that this condition is satisfied in the case of our bump-of-curvature situation. In that case, if we denote the unitary implementor by $U$, we have the following results:

R1. The expectation value $\langle U\Omega | N(\mathrm{a})U\Omega\rangle_{\mathcal{F}(\mathcal{H})}$ of the number operator, $N(\mathrm{a}) = \hat{a}^{\dagger}(\mathrm{a})\hat{a}(\mathrm{a})$, where a is a normalized element of $\mathcal{H}$, is equal to $\langle \beta \mathrm{a} \mid \beta \mathrm{a}\rangle_{\mathcal{H}}$.

R2. First note that there exists an orthonormal basis of vectors, $e_i, (i = 1 \ldots \infty)$, in $\mathcal{H}$ such that the (Hilbert–Schmidt) operator $\bar{\beta}^{*}\bar{\alpha}^{*-1}$ has the canonical form $\sum_i \lambda_i \langle Ce_i | \cdot \rangle | e_i\rangle$. We then have (up to an undetermined phase)

$$U\Omega = N\exp\left(-\frac{1}{2}\sum_i \lambda_i \hat{a}^{\dagger}(e_i)\hat{a}^{\dagger}(e_i)\right)\Omega$$

where the normalization constant $N$ is chosen so that $\|U\Omega\| = 1$. This formula makes manifest that the particles are created in pairs.

We remark that, identifying elements, a, of $\mathcal{H}$ with positive-frequency solutions (below, we shall call them "modes") as explained above, result (R1) may alternatively be expressed by saying that the expectation value, $\omega_{\mathrm{in}}(N(\mathrm{a}))$, in the *in-vacuum state* of the occupation number, $N(\mathrm{a})$, of a *normalized mode*, a, to the future of the bump, is given by $\langle \beta \mathrm{a} | \beta \mathrm{a}\rangle_{\mathcal{H}}$.

This formalism and the results, (R1) and (R2) above, will generalize (at least heuristically, and sometimes rigorously – see especially the rigorous scattering-theoretic work in the 1980s by Dimock and Kay and more recently by A Bachelot and others) to more realistic spacetimes which are only asymptotically flat or asymptotically stationary. In favorable cases, one will still have notions of classical solutions which are positive frequency asymptotically towards the future/past, and, in consequence, one will have well-defined asymptotic notions of "vacuum" and "particles." Also, in, for example, cosmological, models where the background spacetime is slowly varying in time, one can define approximate *adiabatic* notions of classical positive-frequency solutions, and hence also of quantum "vacuum" and "particles" at each finite value of the cosmological time. But, at times where the gravitational field is rapidly varying, one does not expect there to be any sensible notion of "particles." And, in a rapidly time-varying background gravitational field which never settles down, one does not expect there to be any sensible particle interpretation of the theory at all. To understand these statements, it suffices to consider the $(1+0)$-dimensional Klein–Gordon equation with an external potential $V$:

$$\left(-\frac{\mathrm{d}^2}{\mathrm{d}t^2} - m^2 - V(t)\right)\phi = 0$$

which is of course a system of one degree of freedom, mathematically equivalent to the harmonic oscillator with a time-varying angular frequency $\varpi(t) = (m^2 + V(t))^{1/2}$. One could of course express its quantum theory in terms of a time-evolving Schrödinger wave function $\Psi(\varphi, t)$ and attempt to give this a particle interpretation at each time, $s$, by expanding $\Psi(\varphi, s)$ in terms of the harmonic oscillator wave functions for a harmonic oscillator with some particular choice of angular frequency. But the problem is, as is easy to convince oneself, that there is no such good choice. For example, one might think that a good choice would be to take, at time $s$, the set of harmonic oscillator wave functions with angular frequency $\varpi(s)$. (This is sometimes known as the method of "instantaneous diagonalization of the Hamiltonian.") But suppose we were to apply this prescription to the case of a smooth $V(\cdot)$ which is constant in time until time 0 and assume the initial state is the usual vacuum state. Then at some positive time $s$, the number of particles predicted to be present is the same as the number of particles predicted to be present on the same prescription at all times after $s$ for a $\hat{V}(\cdot)$ which is equal to $V(\cdot)$ up to time $s$ and then takes the constant value $V(s)$ for all later times (see Figure 2). But $\hat{V}(\cdot)$ will generically have a sharp corner in its graph (i.e., a

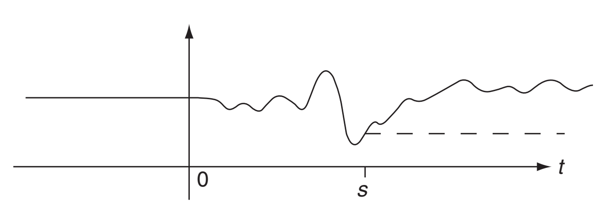



**Figure 2** Plots of $\varpi$ against $t$ for the two potentials $V$ (continuous line) and $\hat{V}$ (continuous line upto $s$ and then dashed line) which play a role in our critique of "instantaneous diagonalization."

discontinuity in its time derivative) at time $s$, and one would expect a large part of the particle production in the latter situation to be accounted for by the presence of this sharp corner – and therefore a large part of the predicted particle production in the case of $V(\cdot)$ to be spurious.

Back in $1+3$ dimensions, even where a good notion of particles is possible, it depends on the choice of time evolution, as is dramatically illustrated by the Unruh effect discussed in the relevant section.

## Theory of the Stress–Energy Tensor

To orient ideas, consider first the free (minimally coupled) scalar field, $(\Box - m^2)\phi = 0$, in Minkowski space. If one quantizes this system in the usual Minkowski-vacuum representation, then the expectation value of the *renormalized stress-energy tensor* (which in this case is the same thing as the *normal ordered stress–energy tensor*) in a vector state $\Psi$ in the Fock space will be given by the formal *point-splitting* expression

$$\begin{aligned}\langle\Psi|T_{ab}(x)\Psi\rangle &= \lim_{(x_1,x_2)\to(x,x)}\Big(\partial_a^1\partial_b^2 - \tfrac{1}{2}\eta_{ab}(\eta^{cd}\partial_c^1\partial_d^2 + m^2)\Big)\\ &\times(\langle\Psi|\rho_0(\phi(x_1)\phi(x_2))\Psi\rangle\\ &- \langle\Omega^{\mathcal{F}}|\rho_0(\phi(x_1)\phi(x_2))\Omega^{\mathcal{F}}\rangle)\end{aligned} \qquad [6]$$

where $\eta_{ab}$ is the usual Minkowski metric. A sufficient condition for the limit here to be finite and well defined would, for example, be for $\Psi$ to consist of a (normalized) finite superposition of $n$-particle vectors of form $\hat{a}^\dagger(\mathrm{a}_1),\ldots,\hat{a}^\dagger(\mathrm{a}_n)\Omega^{\mathcal{F}}$ where the smearing functions $\mathrm{a}_1,\ldots,\mathrm{a}_n$ are all $C^\infty$ elements of $\mathcal{H}$ (i.e., of $L^2_{\mathbb{C}}(\boldsymbol{R}^3)$. The reason this works is that the two-point function in such states shares the same short-distance singularity as the Minkowski-vacuum two-point function. For exactly the same reason, one obtains a well-defined finite limit if one defines the expectation value of the stress–energy tensor in any physically admissible quasifree state by the expression

$$\begin{aligned}\omega(T_{ab}(x)) &= \lim_{(x_1,x_2)\to(x,x)}\Big(\partial_a^1\partial_b^2 - \tfrac{1}{2}\eta_{ab}(\eta^{cd}\partial_c^1\partial_d^2 + m^2)\Big)\\ &\times(\omega(\phi(x_1)\phi(x_2)) - \omega_0(\phi(x_1)\phi(x_2)))\end{aligned} \qquad [7]$$

This latter point-splitting formula generalizes to a definition for the *expectation value of the renormalized stress–energy tensor* for an arbitrary physically admissible quasifree state (or indeed for an arbitrary state whose two-point function has *Hadamard form* – i.e., whose anticommutator function satisfies condition (C4)) on the minimal field algebra and to other linear field theories (including the stress tensor for a conformally coupled linear scalar field) on a general globally hyperbolic spacetime (and the result obtained agrees with that obtained by other methods, including *dimensional regularization* and *zeta-function regularization*). However, the generalization to a curved spacetime involves a number of important new features which we now briefly list (see Wald (1978) for details).

First, the subtraction term which replaces $\omega_0(\phi(x_1)\phi(x_2))$ is, in general, not the expectation value of $\phi(x_1)\phi(x_2)$ in any particular state, but rather a particular *locally constructed Hadamard two-point function* whose physical interpretation is more subtle; the renormalization is thus in general not to be regarded as a normal ordering. Second, the immediate result of the resulting limiting process will not be covariantly conserved and, in order to obtain a covariantly conserved quantity, one needs to add a particular *local geometrical correction term*. The upshot of this is that the resulting expected stress–energy tensor is covariantly conserved but possesses a (state-independent) *anomalous trace*. In particular, for a massless conformally coupled linear scalar field, one has (for all physically admissible quasifree states, $\omega$) the *trace anomaly formula*

$$\omega(T^a_a(x)) = (2880\pi^2)^{-1}\Big(C_{abcd}C^{abcd} + R_{ab}R^{ab} - \tfrac{1}{3}R^2\Big)$$

plus an arbitrary multiple of $\Box R$. In fact, in general, the thus-defined renormalized stress–energy tensor operator (see below) is only defined up to a *finite renormalization ambiguity* which consists of the addition of arbitrary multiples of the functional derivatives with respect to $g_{ab}$ of the quantities

$$I_n = \int_{\mathcal{M}} F_n(x)|\det(g)|^{1/2}\mathrm{d}^4x$$

where $n$ ranges from 1 to 4 with $F_1 = 1$, $F_2 = R$, $F_3 = R^2$, and $F_4 = R_{ab}R^{ab}$. In the Minkowski-space case, only the first of these ambiguities arises and it is implicitly resolved in the formulas [6], [7] inasmuch as these effectively incorporate the *renormalization condition* that $\omega_0(T_{ab}) = 0$. (For the same reason, the locally flat example we give below has no ambiguity.)

One expects, in both flat and curved cases, that, for test functions, $F \in C_0^\infty(\mathcal{M})$, there will exist operators $T_{ab}(F)$ which are *affiliated to* the net of

local $W^*$-algebras referred to earlier and that it is meaningful to write

$$\int_{\mathcal{M}} \omega(T_{ab}(x))F(x)|\det(g)|^{1/2}\mathrm{d}^4x = \omega(T_{ab}(F))$$

provided that, by $\omega$ on the right-hand side, we understand the extension of $\omega$ from the Weyl algebra to this net. ($T_{ab}(F)$ is however not expected to belong to the minimal algebra or be affiliated to the Weyl algebra.)

An interesting simple example of a renormalized stress–energy tensor calculation is the so-called *Casimir effect* calculation for a linear scalar field on a (for further simplicity, $(1+1)$-dimensional) timelike cylinder spacetime of radius $R$ (see Figure 3). This spacetime is globally hyperbolic and stationary and, while locally flat, globally distinct from Minkowski space. As a result, while – provided the regions $\mathcal{O}$ are sufficiently small (such as the diamond region in Figure 3) – elements $\mathcal{A}(\mathcal{O})$ of the minimal net of local algebras on this spacetime will be identifiable, in an obvious way, with elements of the minimal net of local algebras on Minkowski space, the stationary ground state $\omega_{\text{cylinder}}$ will, when restricted to such thus-identified regions, be distinct from the Minkowski vacuum state $\omega_0$. The resulting renormalized stress–energy tensor (as first pointed out in Kay (1979)), definable, once the above identification has been made, exactly as in [7]) turns out, in the massless case, to be nonzero and, interestingly, to have a (in the natural coordinates, constant) *negative energy-density* $T_{00}$. In fact, in this massless case,

$$\omega_{\text{cylinder}}(T_{ab}) = \frac{1}{24\pi R^2}\eta_{ab}$$

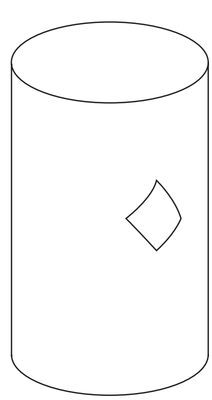


**Figure 3** The timelike cylinder spacetime of radius $R$ with a diamond region isometric to a piece of Minkowski space. See Kay (1979). Casimir effect in quantum field theory. (Original title: The Casimir effect without magic.) *Physical Review D* 20: 3052–3062. Reprinted with permission © 1979 by the American Physical Society.

## Hawking and Unruh Effects

The original calculation by Hawking (1975) concerned a model spacetime for a star which collapses to a black hole. For simplicity, we shall only discuss the spherically symmetric case (see Figure 4). Adopting a similar "mode" viewpoint to that mentioned after results (R1) and (R2) discussed earlier, the result of the calculation may be stated as follows: For a real linear scalar field satisfying [2] with $m=0$ (and $V=0$) on this spacetime, the expectation value $\omega_{\text{in}}(N(\mathrm{a}_{\varpi,\ell}))$ of the occupation number of a one-particle outgoing mode $\mathrm{a}_{\varpi,\ell}$) localized (as far as a normalized mode can be) around $\varpi$ in angular-frequency space and about retarded time $v$, and with angular momentum "quantum number" $\ell$, in the in-vacuum state (i.e., on the minimal algebra for a real scalar field on this model spacetime) $\omega_{\text{in}}$ is, at late retarded times, given by the formula

$$\omega_{\text{in}}(N(\mathrm{a}_{\varpi,\ell})) = \frac{\Gamma(\varpi,\ell)}{\exp(8\pi M\varpi)-1}$$

where $M$ is the mass of the black hole and the *absorption factor* (alternatively known as *gray-body factor*) $\Gamma(\varpi,\ell)$ is equal to the norm-squared of that part of the one-particle mode $\mathrm{a}_{\varpi,\ell}$ which, viewed as a complex positive-frequency classical solution propagating backwards in time from late retarded times, would be absorbed by the black hole. (Note the independence of the right-hand side of this formula from the retarded time, $v$.) This calculation can be understood as an application of result (R1)

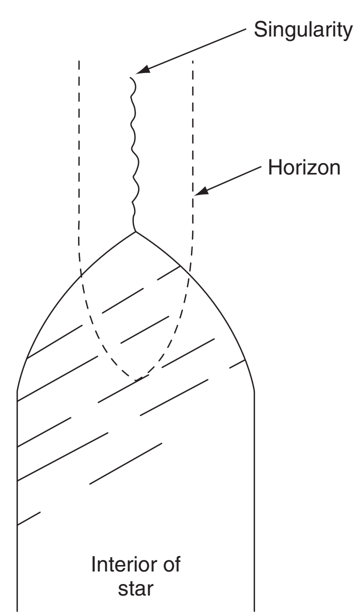



**Figure 4** The spacetime of a star collapsing to a spherical black hole.

(even though the spacetime is more complicated than one with a localized "bump of curvature" and even though the relevant overall time evolution will not be unitarily implemented, the result still applies when suitably interpreted) and the heart of the calculation is an asymptotic estimate of the relevant "$\beta$" Bogoliubov coefficient which turns out to be dependent on the geometrical optics of rays which pass through the star just before the formation of the horizon. This result suggests that the in-vacuum state is indistinguishable at late retarded times from a state of blackbody radiation at the *Hawking temperature*, $T_{\text{Hawking}} = 1/8\pi M$, in Minkowski space from a blackbody (gray body) with the same absorption factor. This was confirmed by further work by many authors. Much of that work, as well as the original result of Hawking was partially heuristic but later work by Dimock and Kay (1987), by Fredenhagen and Haag (1990), and by Bachelot (1999) and others has put different aspects of it on a rigorous mathematical footing. The result generalizes to nonzero mass and higher spin fields to interacting fields as well as to other types of black hole and the formula for the Hawking temperature generalizes to

$$T_{\text{Hawking}} = \kappa/2\pi$$

where $\kappa$ is the *surface gravity* of the black hole.

This result suggests that there is something fundamentally "thermal" about quantum fields on black-hole backgrounds and this is confirmed by a number of mathematical results. In particular, the theorems in the two papers Kay and Wald (1991) and Kay (1993), combined together, tell us that there is a unique state on the Weyl algebra for the *maximally extended Schwarzschild spacetime* (a.k.a. *Kruskal–Szekeres spacetime*) (see **Figure 5**) which is invariant under the *Schwarzschild isometry group* and whose two-point function has Hadamard form. Moreover, they tell us that this state, when restricted to a single wedge (i.e., the exterior Schwarzschild spacetime) is necessarily a KMS state at the Hawking temperature. This unique state is known as the *Hartle–Hawking–Israel state*. These results in fact apply more generally to a wide class of globally hyperbolic spacetimes with *bifurcate Killing horizons* including de Sitter space – where the unique state is sometimes called the *Euclidean* and sometimes the *Bunch–Davies* vacuum state – as well as to Minkowski space, in which case the unique state is the usual Minkowski vacuum state, the analog of the exterior Schwarzschild wedge is a so-called *Rindler wedge*, and the relevant isometry group is a one-parameter family of wedge-preserving Lorentz boosts. In the latter situation, the fact that the Minkowski vacuum state is a KMS state (at "temperature" $1/2\pi$)

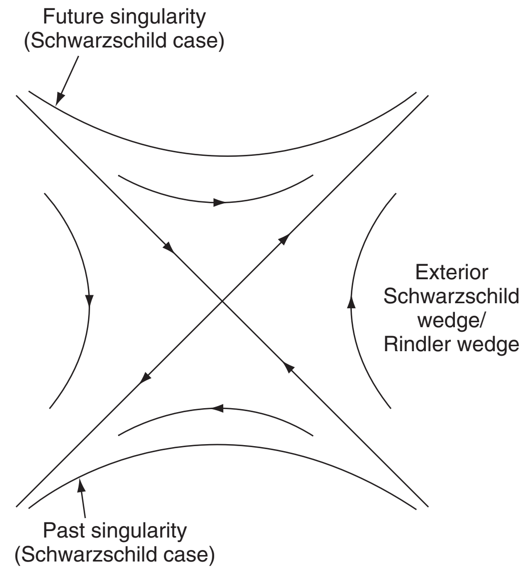



**Figure 5** The geometry of maximally extended Schwarzschild (/or Minkowski) spacetime. In the Schwarzschild case, every point represents a 2-sphere (/in the Minkowski case, a 2-plane). The curves with arrows on them indicate the Schwarzschild time evolution (/one-parameter family of Lorentz boosts). These curves include the (straight lines at right angles) event horizons (/Killing horizons).

when restricted to a Rindler wedge and regarded with respect to the time evolution consisting of the wedge-preserving one-parameter family of Lorentz boosts is known as the *Unruh effect* (1975). This latter property of the Minkowski vacuum in fact generalizes to general *Wightman QFTs* and is in fact an immediate consequence of a combination of the *Reeh–Schlieder theorem* (applied to a Rindler wedge) and the *Bisognano–Wichmann theorem* (1975). The latter theorem says that the defining relation [1] of a KMS state holds if, in [1], we identify the operator $J$ with the complex conjugation which implements wedge reflection and $H$ with the self-adjoint generator of the unitary implementor of Lorentz boosts. We remark that the Unruh effect illustrates how the concept of "vacuum" (when meaningful at all) is dependent on the choice of time evolution under consideration. Thus, the usual Minkowski vacuum is a ground state with respect to the usual Minkowski time evolution but not (when restricted to a Rindler wedge) with respect to a one-parameter family of Lorentz boosts; with respect to these, it is, instead, a KMS state.

## Nonglobally Hyperbolic Spacetimes and the "Time Machine" Question

Hawking (1992) argued that a spacetime in which a time machine gets manufactured should be modeled (see **Figure 6**) by a spacetime with an initial *globally*

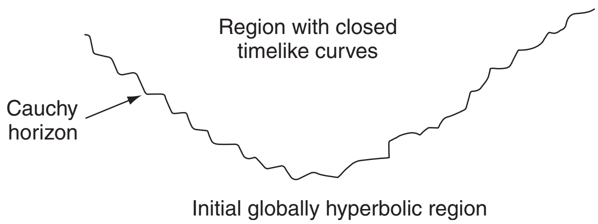



**Figure 6** The schematic geometry of a spacetime in which a time machine gets manufactured.

*hyperbolic region* with a region containing closed timelike curves to its future and such that the future boundary of the globally hyperbolic region is a *compactly generated Cauchy horizon*. On such a spacetime, Kay *et al.* (1997) proved that it is impossible for any distributional bisolution which satisfies (even a certain weakened version of) the Hadamard condition on the initial globally hyperbolic region to continue to satisfy that condition on the full spacetime – the (weakened) Hadamard condition being necessarily violated at at least one point *on* the Cauchy horizon. This result implies that, however one extends a state, satisfying our conditions (C1)–(C4), on the minimal algebra for [2] on the initial globally hyperbolic region, the expectation value of its stress–energy tensor must necessarily become singular on the Cauchy horizon. This result, together with many heuristic results and specific examples considered by many other authors appears to support the validity of the (Hawking 1992) *chronology protection conjecture* to the effect that it is impossible in principle to manufacture a time machine. However, there are potential loopholes in the physical interpretation of this result as pointed out by Visser (1997), as well as other claims by various authors that one can nevertheless violate the chronology protection conjecture. For a recent discussion on this question, we refer to Visser (2003).

## Other Related Topics and Some Warnings

There is a vast computational literature, calculating the expectation values of stress–energy tensors in states of interest for scalar and higher spin linear fields (and also some work for interacting fields) on interesting cosmological and black-hole backgrounds. QFT on de Sitter and anti-de Sitter space is a big subject area in its own right with recent renewed interest because of its relevance to *string theory* and *holography*. Also important on black-hole backgrounds is the calculation of gray-body factors, again with renewed interest because of relevance to string theory and to *brane-world* scenarios.

There are many further mathematically rigorous results on algebraic and axiomatic QFT in a curved spacetime setting, including versions of *PCT*, *spin-statistics* and *Reeh–Schlieder* theorems and also rigorous *energy inequalities* bounding the extent to which expected energy densities can be negative, etc.

There is much mathematical work controlling scattering theory on black holes, partly with a view to further elucidating the Hawking effect.

Perturbative renormalization theory of interacting quantum fields in curved spacetime is also now a highly developed subject.

Beyond QFT in a fixed curved spacetime is *semiclassical gravity* which takes into account the *back-reaction* of the expectation value of the stress–energy tensor on the classical gravitational background. There are also interesting condensed matter analogs of the Hawking effect such as *dumb holes*.

Readers exploring the wider literature, or doing further research on the subject should be aware that the word "vacuum" is sometimes used to mean "ground state" and sometimes just to mean "quasifree state." They should be cautious of attempts to define particles on Cauchy surfaces in *instantaneous diagonalization* schemes (cf. the remarks at the end of the section "Particle creation and the limitations of the particle concept"). When studying (or performing) calculations of the "expectation value of the stress–energy tensor" it is always important to ask oneself with respect to which state the expectation value is being taken. It is also important to remember to check that candidate two-point (anticommutator) functions satisfy the positivity condition (C3) discussed earlier. Typically, two-point distributions obtained via *mode sums* automatically satisfy condition (C3) (and condition (C4)), but those obtained via *image* methods do not always satisfy it. (When they do not, the presence of nonlocal spacelike singularities is often a tell-tale sign as can be inferred from Kay's conjecture/Radzikowski's theorem discussed earlier.) There are a number of apparent implicit assertions in the literature that some such two-point functions arise from "states" when of course they cannot. Some of these concern proposed analogs to the Hartle–Hawking–Israel state for the (appropriate maximal globally hyperbolic portion of the maximally extended) Kerr spacetime. That they cannot belong to states is clear from a theorem in Kay and Wald (1991) which states that there is no stationary Hadamard state on this spacetime at all. Others of them concern claimed "states" on spacetimes such as those discussed in the previous section which, if they really were states would seem to be in conflict with the chronology protection conjecture. Finally, beware states (such as the so-called $\alpha$-*vacua* of de Sitter spacetime) whose two-point

distributions violate the "Hadamard" condition (C4) and which therefore do not have a well-defined finite expectation value for the renormalized stress–energy tensor.

*See also:* AdS/CFT Correspondence; Algebraic Approach to Quantum Field Theory; Axiomatic Quantum Field Theory; Black Hole Mechanics; Bosons and Fermions in External Fields; Integrability and Quantum Field Theory; Quantum Fields with Indefinite Metric: Non-Trivial Models; Quantum Fields with Topological Defects; Quantum Geometry and Its Applications; Scattering in Relativistic Quantum Field Theory: Fundamental Concepts and Tools; Thermal Quantum Field Theory.

## Further Reading


Birrell ND and Davies PCW (1982) *Quantum Fields in Curved Space*. Cambridge: Cambridge University Press.

Brunetti R, Fredenhagen K, and Verch R (2003) The generally covariant locality principle – a new paradigm for local quantum physics. *Communications in Mathematical Physics* 237: 31–68.

DeWitt BS (1975) Quantum field theory in curved space-time. *Physics Reports* 19(6): 295–357.

Dimock J (1980) Algebras of local observables on a manifold. *Communications in Mathematical Physics* 77: 219–228.

Haag R (1996) *Local Quantum Physics*. Berlin: Springer.

Hartle JB and Hawking SW (1976) Path-integral derivation of black-hole radiance. *Physical Review D* 13: 2188–2203.

Hawking SW (1975) Particle creation by black holes. *Communications in Mathematical Physics* 43: 199–220.

Hawking SW (1992) The chronology protection conjecture. *Physical Review D* 46: 603–611.

Israel W (1976) Thermo-field dynamics of black holes. *Physics Letters A* 57: 107–110.

Kay BS (1979) Casimir effect in quantum field theory. (Original title: The Casimir effect without magic.) *Physical Review D* 20: 3052–3062.

Kay BS (1993) Sufficient conditions for quasifree states and an improved uniqueness theorem for quantum fields on space-times with horizons. *Journal of Mathematical Physics* 34: 4519–4539.

Kay BS (2000) Application of linear hyperbolic PDE to linear quantum fields in curved spacetimes: especially black holes, time machines and a new semi-local vacuum concept. *Journées Équations aux Dérivées Partielles*, Nantes, 5–9 juin 2000, GDR 1151 (CNRS): IX1–IX19. (Also available at http://www.math.sciences.univ-nantes.fr or as gr-qc/0103056.)

Kay BS, Radzikowski MJ, and Wald RM (1997) Quantum field theory on spacetimes with a compactly generated Cauchy horizon. *Communications in Mathematical Physics* 183: 533–556.

Kay BS and Wald RM (1991) Theorems on the uniqueness and thermal properties of stationary, nonsingular, quasifree states on spacetimes with a bifurcate Killing horizon. *Physics Reports* 207(2): 49–136.

Misner CW, Thorne KS, and Wheeler JA (1973) *Gravitation*. San Francisco: W.H. Freeman.

Unruh W (1976) Notes on black hole evaporation. *Physical Review D* 14: 870–892.

Visser M (2003) The quantum physics of chronology protection. In: Gibbons GW, Shellard EPS, and Rankin SJ (eds.) *The Future of Theoretical Physics and Cosmology*. Cambridge: Cambridge University Press.

Wald RM (1978) Trace anomaly of a conformally invariant quantum field in a curved spacetime. *Physical Review D* 17: 1477–1484.

Wald RM (1994) *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*. Chicago: University of Chicago Press.


# Quantum Field Theory: A Brief Introduction

**L H Ryder**, University of Kent, Canterbury, UK



By any account quantum field theory occupies a prominent place in the history of mathematical physics. This article is, however, not intended to serve as an overview of this subject, but has the more modest aim of identifying a few areas which seem to me interesting and significant.

## Historical Remarks; Second Quantization

At the time when quantum field theory was at the forefront of theoretical physics its *raison d'être* was to complete the quantum description of the sub-atomic world. Quantum mechanics had been amazingly successful in solving almost the whole of atomic physics by making explicit the quantum (wave) nature of the electron, according to the formulations of Heisenberg and Schrödinger. The introduction of the quantum idea into physics, however, by Planck in 1900 closely followed by Einstein in 1905 was the proposal of a quantum (particular) aspect of the electromagnetic field – the photon. In the mid-1920s the only force in nature to be considered was the electromagnetic interaction; this was before the theories of Yukawa and Fermi, concerning the strong and weak nuclear forces. Dirac, Heisenberg, Jordan, and others then addressed themselves to finding a formulation of quantum electrodynamics (QED) comparable in mathematical sophistication to the Heisenberg–Schrödinger formulation of quantum mechanics – which Planck's and Einstein's theories were not.

The idea that was pursued, at least in the early stages, was that the Schrödinger wave function $\psi$, taken as a wave field, should be "quantized"; Dirac

seems to have taken this as a model for photons. Jordan further proposed that electrons should be treated as the quanta of an electron field, but recognized that their fermionic nature would modify the quantization procedure. This generic idea involved what was called "second quantization" – of a field into a particle.

One of the earliest quantization rules was Bohr's condition relating to the periodic orbits of electrons in atoms, $J = \int p\, \mathrm{d}q = nh$. At the hands of Heisenberg and Dirac this became upgraded to the commutation relation

$$[q, p] = \mathrm{i}\hbar$$

where the operators $p$ and $q$ are "observables." In their papers on quantum field theory, Dirac, Jordan and Wigner, and Heisenberg introduced creation and annihilation operators which had the function, as their name implied, of creating and destroying single particles – quanta of the field. These operators obeyed the commutation rules (with $[A, B] = AB - BA$)

$$[b_r, b_s^*] = \delta_{rs}, \quad [b_r, b_s] = [b_r^*, b_s^*] = 0$$

when the field quanta were bosons, and the anti-commutation rules

$$\{b_r, b_s^*\} = \delta_{rs}, \quad \{b_r, b_s\} = \{b_r^*, b_s^*\} = 0$$

(with $\{A, B\} = AB + BA$) when the field quanta were fermions (e.g., electrons). These steps constitute second quantization, but it may be noted that the creation and annihilation operators are not observables, as $p$ and $q$ are in the Heisenberg commutation relation. In addition, the second quantization conditions do not involve Planck's constant. "First" and "second" quantization are therefore not so similar as one might like to think.

The question of what exactly is being quantized was in fact the source of some confusion. In his paper of 1927, Dirac's attention is focussed on electromagnetic radiation, but he nevertheless discusses the difference between "a light-wave and the de Broglie or Schrödinger wave associated with the light-quanta." As Dirac points out, "their intensities are to be interpreted in different ways. The number of light quanta per unit volume associated with a monochromatic light-wave equals the energy per unit volume of the wave divided by the energy $(2\pi h)\nu$ of a single light quantum. On the other hand a monochromatic de Broglie wave of amplitude $a$ (multiplied into the imaginary exponential factor) must be interpreted as representing $a^2$ light quanta per unit volume for all frequencies." There are at least two problematic issues here. First, is the Schrödinger wave function $\psi$ to be considered as a "real" field, whose quanta result in "real" particles, or is it a probability field, whose significance lies in Born's probabilistic interpretation of quantum mechanics? Born wrote in 1926, "[Einstein said that] the waves are present only to show the corpuscular light quanta the way, and he spoke in the sense of a "ghost field". This determines the probability that a light quantum, the bearer of energy and momentum, takes a certain path; however, the field itself has no energy and no momentum." This is the first problem. The second one concerns the nature of the quantization itself. Is this a quantization of field energy, or a quantization of the field itself, as a substantial entity? If the field is real, the second of these does not imply the first.

Ambiguities surrounding the idea of second quantization survived into the 1960s. Wigner is recorded as saying, in an interview in 1963, "just as we get photons by quantising the electromagnetic fields, so we should be able to get material particles by quantising the Schrödinger field." And Rosenfeld, also in an interview in 1963, said, "in some sense or other, Jordan himself took the wave function, the probability amplitude, physically more seriously than most people [did]."

It would seem we are justified in concluding that the idea of second quantization contains flaws, but an even clearer indication of the need for rethinking is provided by the story of the Dirac equation. This is a wave equation for the electron, compatible with special relativity, and taking explicit account of its spin being $(1/2)\hbar$. The equation famously had both positive- and negative-energy solutions. This potential disaster was converted by Dirac into a triumph by reinterpreting the (absence of) negative-energy solutions as (positive-energy) antiparticles – positrons, particles with positive charge but the same mass and spin as the electron. Positrons were eventually discovered by Anderson. It was later shown that the existence of antiparticles is a general feature of quantum field theory, not just a peculiarity of spin-1/2 particles. The significance of this discovery, however, is that the twin requirements of relativity and quantum theory are not compatible with a single-particle state; rather, these requirements result in a two-particle state. Thus, in some sense the requirements of relativity and quantum mechanics already start to take us down the road to a quantum theory of fields.

Quantum field theory is then constructed on the following sort of framework: "classical" theories for fields with any spin may be written down and these are quantized by reinterpreting the field variables as operators and imposing Heisenberg-type commutation relations on the field and its corresponding

"momentum" variable. So, for example, for spinless fields we have the equal-time commutation relation

$$[\phi(\boldsymbol{x}, t), \pi(\boldsymbol{y}, t)] = \mathrm{i}\hbar\delta^{(3)}(\boldsymbol{x} - \boldsymbol{y})$$

where $\pi = \partial \mathrm{L}/\partial(\partial_0\phi)$ and L is the Lagrange density. The mass and spin of particles are defined with reference to the Poincaré group (thereby incorporating special relativity) and the quantum requirement is the familiar one that physical states are represented by vectors in Hilbert space. The rest follows: as Weinberg says, "quantum field theory is the way it is because (with certain qualifications) this is the only way to reconcile quantum mechanics with special relativity."

## Renormalization

A notorious problem in quantum field theory is the occurrence of infinities. In QED, for example, the electron acquires a self-energy – and therefore a contribution to its mass – by virtue of the emission and reabsorption of virtual photons. It turns out that this self-energy is infinite – it is given by a divergent integral – even in the lowest order of perturbation theory. In the early days, this was recognized as being a serious problem, and in fact it turns out to be a generic problem in quantum field theory. It was realized by Dyson, however, that in some field theories these divergences may be dealt with by redefining a small number of parameters (e.g., in QED, the electron mass, charge, and field amplitude) so that thereafter the theory is finite to all orders of perturbation theory. Such theories are called renormalizable, and QED is a renormalizable field theory.

Some important field theories, however, are not renormalizable; an example is Fermi's theory of weak interactions. To lowest order in perturbation theory, Fermi's theory works well (e.g., in accounting for the electron spectrum in neutron beta decay), but to higher orders divergent results are obtained, which cannot be waved away by redefining a finite number of parameters; that is to say, as the order of perturbation increases, so also does the number of parameters to be redefined. Nonrenormalizable theories of this type have traditionally been regarded as highly undesirable, not to say rather nasty.

The modern view of renormalization is, however, somewhat different. The problem with nonrenormalizable theories is that, in order to calculate a physical process to all orders in perturbation theory, an infinite number of parameters must be renormalized, so the theory has no predictive power. In practice, however, we do not need to calculate to all orders in perturbation theory, since any physical process (say a scattering process or a particle decay) will only be observed at a finite energy and comparison of theory and experiment therefore only requires calculation up to a finite order of perturbation theory. So even nonrenormalizable theories are perfectly acceptable as low-energy theories. This amounts to a philosophy of effective field theories; an effective field theory is a model which holds good up to a particular energy scale, or equivalently down to a particular length scale.

An important addition to the theoretical armoury is the renormalization group. Renormalization is implemented first of all by a scheme of regularization, which enables the divergences to be exhibited explicitly. The simplest type of regularization is the introduction of a cutoff in the momentum integrals, but in modern particle physics the favored scheme is dimensional regularization. The dimensionality of the integrals in momentum space is taken to be $d = 4 - \varepsilon$ and the divergent quantities have an explicit dependence on $\varepsilon$ (which, of course, as the "real" world is approached, approaches zero). At the same time, a mass parameter $\mu$ is introduced in order to define dimensionless quantities, for example, a dimensionless coupling constant. The renormalized quantities then depend on the "bare" (unrenormalized) quantities and on $\mu$ and $\varepsilon$. The arbitrariness of $\mu$ enables a differential equation, for scattering amplitudes, for example, to be written down. While at first sight this renormalization group equation might seem to have no physical importance, in fact it gives a powerful way of studying scattering behavior at large momenta.

Most interestingly, the concept of the renormalization group also arises in condensed matter physics. Here, rather than, for example, a cutoff in momentum space, the relevant parameter is a distance scale. In the Ising model in statistical mechanics, for example, in which spins are located on a lattice, the parameter is the lattice spacing. To construct a theory that describes the physics on the macroscopic scale involves integrating out the details on the microscopic scale and one way to do this is via the "block spin" transformation originally introduced by Kadanoff. In this way the renormalization group has had a large impact in condensed matter physics, for example, in the study of critical phenomena.

## Particle Physics and Cosmology

Probably the most spectacular success of quantum field theory in the twentieth century has been in particle physics. The "standard model" accounts for the strong, electromagnetic, and weak interactions

between elementary particles with outstanding success. The interactions are generalizations of Maxwell's electrodynamics, which is invariant under a symmetry group U(1) of gauge transformations. An enlargement of this group to $SU(2) \otimes U(1)$ accounts for the unified electroweak interaction (the unification resulting from the fact that the two U(1)'s above are not exactly the same; there is some on-diagonal mixing), and the strong interactions between quarks, which binds them into hadrons, are invariant under an SU(3) group of gauge transformations. The gauge fields are the photon $\gamma$, the $W$ and $Z$ bosons (both heavy; of the order of 100 times the proton mass), and the (massless) gluons mediating the force between quarks (quantum chromodynamics, QCD). An important feature of the standard model is spontaneous symmetry breaking, which is the mechanism by which the $W$ and $Z$ particles acquire a mass (but the photon does not, and neither do the gluons). This goes by the name of the Higgs mechanism.

The quantization of the standard model is most successfully carried out using the path-integral formalism, rather than canonical quantization, and the proof of the renormalizability of the model (of nonabelian gauge theories with spontaneous symmetry breaking) was given by 't Hooft. Details of these topics are now available in many textbooks.

Confidence that this is a realistic model of elementary particles – that is to say, of quarks and leptons – depends, of course, on particular experiments and their interpretation and an important milestone on this journey was Feynman's quark–parton model of deep inelastic electron–proton scattering. The interpretation of the data required a picture of an electron scattering from an individual quark in the proton, and this in turn required a negligible interaction between quarks; in other words, that at small distances (inside the proton) the quarks are (almost) free – despite the fact that at large distances they most certainly are not! The proof, by Gross, Politzer, and Wilczek, that nonabelian gauge are indeed asymptotically free (asymptotic in momentum space, that is) was therefore an important event in helping to establish the credibility of the standard model.

A characteristic contribution of quantum field theory to our view of the physical world is its picture of the vacuum, as being populated with virtual particle–antiparticle pairs. A consequence of this is the phenomenon of vacuum polarization – that the presence of an electric charge in free space polarizes these virtual pairs. This in turns leads to the phenomenon of screening in QED, and antiscreening in QCD, SU(3) having a more complicated structure than U(1). It also leads to a nonzero (in fact, quadratically divergent!) value for the energy of the vacuum. This is in effect the contribution of the zero-point energies of all the oscillators in the Fourier expansion of the scalar field operator. In any other interaction than gravity, this zero-point energy may be ignored, but in gravity it may be expected to have observable consequences, and indeed it turns out that it plays the same role as a cosmological constant $\Lambda$, and therefore acts as an agent of acceleration, rather than deceleration, of the universe.

A final topic worth noting is one whose existence would have been inconceivable in the early days of this subject. The nonlinearity of the (nonabelian) gauge field equations and the existence of a nontrivial group space allows new types of topologically nontrivial solutions to these equations: solitons, bounces, instantons, sphalerons, and so on. Effects such as fractional spin and nonconservation of fermion number also appear, and, on the cosmological scale, domain walls and cosmic strings. There is something here for theoretical physicists of many differing interests.

*See also:* Algebraic Approach to Quantum Field Theory; Axiomatic Quantum Field Theory; BRST Quantization; Constrained Systems; Constructive Quantum Field Theory; Deformation Quantization; Electroweak Theory; Euclidean Field Theory; Exact Renormalization Group; Integrability and Quantum Field Theory; Nonperturbative and Topological Aspects of Gauge Theory; Perturbative Renormalization Theory and BRST; Quantum Chromodynamics; Quantum Electrodynamics and Its Precision Tests; Quantum Fields with Indefinite Metric: Non-Trivial Models; Quantum Fields with Topological Defects; Renormalization: General Theory; Standard Model of Particle Physics; Symmetries and Conservation Laws; Symmetries in Quantum Field Theory of Lower Spacetime Dimensions; Topological Defects and Their Homotopy Classification; Topological Quantum Field Theory: Overview; Twistors.

## Further Reading

Cao TY (1997) *Conceptual Developments of 20th Century Field Theories*. Cambridge: Cambridge University Press.

Davies P (ed.) (1989) *The New Physics*. Cambridge: Cambridge University Press.

Gross F (1993) *Relativistic Quantum Mechanics and Field Theory*. New York: Wiley.

Huang K (1998) *Quantum Field Theory*. New York: Wiley.

Itzykson C and Zuber J-B (1980) *Quantum Field Theory*. New York: McGraw-Hill.

Maggiore M (2005) *A Modern Introduction to Quantum Field Theory*. Oxford: Oxford University Press.

Rubakov V (2002) *Classical Theory of Gauge Fields*. Princeton: Princeton University Press.

Schweber SS (1994) *QED and the Men Who Made It*. Princeton: Princeton University Press.

Schwinger J (ed.) (1958) *Quantum Electrodynamics*. New York: Dover.

't Hooft G (1997) *In Search of the Ultimate Building Blocks*. Cambridge: Cambridge University Press.

Weinberg S (1995, 1996) *The Quantum Theory of Fields*, vol. 1 and 2. Cambridge: Cambridge University Press.
Wentzel G (1960) Quantum theory of fields (until 1947). In: Fierz M and Weisskopf VF (eds.) *Theoretical Physics in the Twentieth Century*, (Reprinted in Mehra J (1973) *The Physicist's Conception of Nature*. Dordrecht: Reidel.) New York: Interscience.
Zee A (2003) *Quantum Field Theory in a Nutshell*. Princeton: Princeton University Press.

# Quantum Fields with Indefinite Metric: Non-Trivial Models

**S Albeverio and H Gottschalk**, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany



## Introduction

The nonperturbative construction of quantum field models with nontrivial scattering in arbitrary dimension $d$ of the underlying Minkowski spacetime is much simpler in the framework of quantum field theory with indefinite metric than in the positive-metric case. In particular, there exist a number of solutions in the physical dimension $d = 4$, where up to now no positive-metric solutions are known. The reasons why this is so are reviewed in this article, and some examples obtained by analytic continuation from the solutions of Euclidean covariant stochastic partial differential equations (SPDEs) driven by non-Gaussian white noise are discussed.

## The Hilbert Space Structure Condition

It has been proved by F Strocchi that a quantum gauge field in a local, covariant gauge cannot act on a Hilbert space with a positive-definite inner product. But it is possible to overcome this obstacle by passing from a Hilbert space representation of the algebra of the quantum field to Krein space representations in order to preserve locality and covariance under the Poincaré group.

A Krein space $\mathcal{K}$ is an inner-product space which also is a Hilbert space with respect to some auxiliary scalar product. The relation between the inner product $\langle .,. \rangle$ and the auxiliary scalar product $(.,.)$ is given by a self-adjoint linear operator $J:\mathcal{K}\to\mathcal{K}$ with $J^2 = \mathbf{1}_{\mathcal{K}}$ and $\langle .,. \rangle = (.,J.)$. $J$ is called the metric operator. A quantum field acting on such a space is called a quantum field with indefinite metric. The formal definition is as follows.

Let $\mathcal{D}\subseteq\mathcal{K}$ be a dense linear space and $\Omega\in\mathcal{D}$ a distinguished vector (henceforth called the vacuum). Let $\mathcal{S}=\mathcal{S}(\mathbb{R}^d,\mathbb{C}^N)$ be the space of Schwartz test functions with values in $\mathbb{C}^N$. A quantum field $\phi$ by definition is a linear mapping from $\mathcal{S}$ to the linear operators on $\mathcal{D}$. One usually assumes that $\mathcal{D}$ is generated as the linear span of vectors generated by repeated application of field operators to the vacuum. The following properties should hold for the quantum field $\phi$:

1. *Temperedness*: $f_n \to f$ in $\mathcal{S} \Rightarrow \langle \Psi, \phi(f_n)\Phi\rangle \to \langle \Psi, \phi(f)\Psi\rangle\ \forall \Psi,\Phi\in\mathcal{S}$.
2. *Covariance*: There exists a weakly continuous representation $U$ of the covering of the orthochronous, proper Poincaré group $\tilde{P}_+^{\uparrow}$ by linear operators on $\mathcal{D}$ which is $J$-unitary, that is, $U^{[*]} = U^{-1}$ with $U^{[*]} = JU^*J|_{\mathcal{D}}$ and leaves $\Omega$ invariant. $\phi$ is said to be covariant with respect to $U$ and a representation $\tau$ of the covering of the orthochronous, proper Lorentz group $\tilde{L}_+^{\uparrow}$ if $U(g)\phi(f)U(g)^{-1} = \phi(f_g)$, where $f_g(x) = \tau(\Lambda)f(\Lambda^{-1}(x-a))$, $g=\{\Lambda,a\}$, $\Lambda\in\tilde{L}_+^{\uparrow}$, $a\in\mathbb{R}^d$.
3. *Spectrality*: Let $U(a)$, $a\in\mathbb{R}^d$, be the representation of the translation group and let $\sigma = \cup_{\Psi,\Phi\in\mathcal{D}}\operatorname{supp}\mathcal{F}(\langle\Psi,U(.)\Phi\rangle)$ with $\mathcal{F}$ the Fourier transform (in the sense of tempered distributions). Formally, $\sigma$ is the joint spectrum of the generators of spacetime translations $U(a)$. The spectral condition then demands that $\sigma\subseteq\bar{V}_0^+$, the closed forward light cone in energy–momentum space.
4. *Locality*: There is a decomposition $\mathbb{C}^N=\oplus_\kappa V_\kappa$ such that for each $f,h\in\mathcal{S}$ taking values in a $V_\kappa$ and having spacelike separated supports one has either $[\phi(f),\phi(h)]=0$ or $\{\phi(f),\phi(h)\}=0$, where $[.,.]$ is the commutator and $\{.,.\}$ the anticommutator.
5. *Hermiticity*: There is an involution $*$ on $\mathcal{S}$ such that $\phi(f)^{[*]}=\phi(f^*)$.

The quantum-mechanical interpretation of the inner product of two vectors in $\mathcal{K}$ as a probability amplitude, however, gets lost. It has to be restored by the construction of a physical subspace of $\mathcal{K}$ where the restriction of the inner product is non-negative. This is called the Gupter–Bleuler gauge procedure. Typically, one first considers the problem of constructing quantum fields with indefinite metric, that is, the dynamical problem is addressed. This is often followed by the construction of the physical states, which involves implementation of quantum constraints.

The vacuum expectation values (VEVs), also called Wightman functions, of the quantum field theory with indefinite metric (IMQFT) are defined as

$$W_n(f_1 \otimes \cdots \otimes f_n) = \langle \Omega, \phi(f_1) \cdots \phi(f_n)\Omega \rangle \quad f_1, \ldots, f_n \in \mathcal{S} \qquad [1]$$

An axiomatic framework for (unconstrained) IMQFT has been suggested by G Morchio and F Strocchi in terms of the Wightman functions $W_n \in \mathcal{S}'$, $n \in \mathbb{N}_0$. Previous work on the topic had been done by J Yngvason. These generalized Wightman axioms of Morchio and Strocchi replace the positivity condition on the Wightman functions by a so-called Hilbert space structure condition (HSSC): for $n \in \mathbb{N}_0$ there exist $p_n$ a Hilbert seminorm on $\mathcal{S}^{\otimes n}$ such that

$$|W_{n+m}(f \otimes h)| \leq p_n(f) p_m(h) \quad \forall n, m \in \mathbb{N}_0 \quad f \in \mathcal{S}^{\otimes n},\ h \in \mathcal{S}^{\otimes m} \qquad [2]$$

This condition makes sure that a field algebra on a Krein space with VEVs equal to the given set of Wightman functions can be constructed. The remaining axioms of the Wightman framework – temperedness, covariance, spectral condition, locality, and Hermiticity – remain the same. Clustering of Wightman functions is assumed at least for massive theories:

$$\lim_{t\to\infty} W_{n+m}(f \otimes h_{ta}) = W_n(f) W_m(h) \quad \forall n, m \in \mathbb{N}_0 \quad f \in \mathcal{S}^{\otimes n},\ h \in \mathcal{S}^{\otimes m} \qquad [3]$$

for spacelike $a \in \mathbb{R}^d$. It fails to hold in certain physical contexts where multiple vacua (also called $\Theta$-vacua) accompanied with massless Goldstone bosons occur due to spontaneous symmetry breaking.

In the original Wightman axioms, there are essentially two nonlinear axioms: positivity and clustering. Here nonlinear means that checking that condition involves more than one VEV with a given number of field operators. The cluster condition can be linearized by an operation on the Wightman functions called “truncation.” The equations

$$W_n(f_1 \otimes \cdots \otimes f_n) = \sum_{I \in \mathcal{P}^{(n)}} \prod_{\substack{\{j_1,\ldots,j_l\} \in I \\ j_1 < j_2 < \cdots < j_l}} W_n^T(f_{j_1} \otimes \cdots \otimes f_{j_l}) \qquad [4]$$

recursively define the truncated Wightman functions $W_n^T$ for $n \in \mathbb{N}$. Here $\mathcal{P}^{(n)}$ stands for the set of all partitions of $\{1, \ldots, n\}$ into disjoint, nonempty sets. Unfortunately, the positivity condition (at least when combined with nontrivial scattering) becomes highly nonlinear for truncated Wightman functions. This can be seen as one explanation why it is so difficult to find nontrivial (i.e., corresponding to nontrivial interactions) solutions to the Wightman axioms.

But it turns out that, in contrast to positivity, the HSSC is essentially linear for truncated Wightman functions.

**Theorem 1** *If there exists a Schwartz norm $\|\cdot\|$ on $\mathcal{S}$ such that $W_n^T$ is continuous with respect to $\|\cdot\|^{\otimes n}$ for $n \in \mathbb{N}$ then the associated sequence of Wightman functions $\{W_n\}$ fulfills the HSSC [2].*

Note that $\|\cdot\|^{\otimes n}$ is well defined as $\mathcal{S}$ is a nuclear space. This theorem makes it much easier to construct IMQFTs. In particular, all known solutions of the linear program for truncated Wightman functions lead to an abundance of mathematical solutions to the axioms of IMQFT, as long as the singularities of truncated Wightman functions in position and energy–momentum space do not become increasingly stronger with growing $n$. For example, the perturbative solutions to Wightman functions of Ostendorf and Steinmann provide solutions when the perturbation series is truncated at a given order.

## Relativistic Fields from Euclidean Stochastic Equations

In the classical work on constructive quantum field theory, relativistic fields in spacetime dimensions $d = 2$ and 3 have been constructed by analytic continuation from Euclidean random fields. This, in particular, has led to firm connections between quantum field theory and equilibrium statistical mechanics. Let us discuss one specific class of solutions of the axioms of IMQFT for arbitrary $d$ which also stem from random fields related to an ensemble of statistical mechanics of classical, continuous particles. Mathematically, this is connected with using random fields with Poisson distribution. As in constructive QFT, the moments, also called Schwinger functions, of the random field can be analytically continued from Euclidean imaginary time to relativistic real time. That this is possible results from an explicit calculation. Axiomatic results cannot be used, as they depend on positivity or reflection positivity in the Euclidean spacetime, respectively.

By definition, a mixing Euclidean covariant random field $\varphi$ is an almost surely linear mapping from $\mathcal{S}_{\mathbb{R}} = \mathcal{S}(\mathbb{R}^d, \mathbb{R}^N)$ to the space of real-valued

measurable functions (random variables) on some probability space that fulfills the following properties:

1. *Temperedness*: $f_n \to f$ in $\mathcal{S}_{\mathbb{R}} \Rightarrow \varphi(f_n) \xrightarrow{\mathcal{L}} \varphi(f)$.
2. *Covariance*: $\varphi(f) \overset{\mathcal{L}}{=} \varphi(f_g)\ \forall f \in \mathcal{S}_{\mathbb{R}},\ g = \{\Lambda, a\}$, $\Lambda \in \mathrm{SO}(d),\ a \in \mathbb{R}^d,\ f_g(x) = \tau(\Lambda) f(\Lambda^{-1}(x-a))$ for some continuous representation $\tau : \mathrm{SO}(d) \to \mathrm{GL}(N)$.
3. *Mixing*: $\lim_{t\to\infty} \mathbb{E}[AB_{ta}] = \mathbb{E}[A]\mathbb{E}[B]$ for all square-integrable random variables $A = A(\varphi)$, $B = B(\varphi)$, and $B_{ta} = B(\varphi_{ta})$, $\varphi_{ta}(f) = \varphi(f_{ta})\ \forall f \in \mathcal{S}_{\mathbb{R}}$, $a \in \mathbb{R}^d \setminus \{0\}$.

The mixing condition in the Euclidean spacetime plays the same role as the cluster property in the generalized Wightman axioms.

In particular, we consider random fields $\varphi$ obtained as solutions of the SPDE $D\varphi = \eta$. In this equation, $\eta$ is a noise field, that is, $\eta$ is $\tau$-covariant for some representation of $\mathrm{SO}(d)$, $\eta(f)$ has infinitely divisible probability law and $\eta(f)$, $\eta(h)$ are independent $\forall f, h \in \mathcal{S}_{\mathbb{R}}$ with $\operatorname{supp} f \cap \operatorname{supp} h = \emptyset$. $D$ is a $\tau$-covariant (i.e., $\tau(\Lambda) D \tau(\Lambda)^{-1} = D \quad \forall \Lambda \in \mathrm{SO}(d)$) partial differential operator with constant coefficients (also pseudodifferential operators $D$ could be considered). From the classification of infinitely divisible probability laws, it is known that $\eta$ essentially consists of Gaussian white noise and Poisson fields and derivatives thereof. Such a Gauss–Poisson noise field by the Bochner–Minlos theorem is characterized by its Fourier transform. Direct relations with QFT arise if one chooses

$$\mathbb{E}[\mathrm{e}^{\mathrm{i}\eta(f)}] = \exp\left\{ \int_{\mathbb{R}^d} \psi(f) - f \cdot \bar{\sigma}^2 p(-\Delta) f \,\mathrm{d}x \right\}, \quad f \in \mathcal{S}_{\mathbb{R}} \qquad [5]$$

where $\psi : \mathbb{R}^N \to \mathbb{C}$ is a Lévy function,

$$\psi(t) = \mathrm{i}a \cdot t - \frac{t \cdot \sigma^2 t}{2} + z \int_{\mathbb{R}^N \setminus \{0\}} (\mathrm{e}^{\mathrm{i}t\cdot s} - 1)\,\mathrm{d}r(s), \quad t \in \mathbb{R}^N \qquad [6]$$

Here the centered dot represents a $\tau$-invariant scalar product on $\mathbb{R}^N$, $\sigma$ a positive-semidefinite $\tau$-invariant $N \times N$ matrix, $z \geq 0$ a real number and $r$ is a $\tau$-invariant probability measure on $\mathbb{R}^n \setminus \{0\}$ with all moments. Further, $\bar{\sigma}^2_{\alpha,\beta} = (\partial^2 \psi(t)/\partial t_\alpha \partial t_\beta)|_{t=0}$, and $p : [0,\infty) \to [0,\infty)$ is a polynomial depending on $D$. If $\hat{D}^{-1}$, the Fourier-transformed inverse of $D$, exists, it can be represented by

$$\hat{D}^{-1}(k) = \frac{Q_E(k)}{\prod_{l=1}^{P} (|k|^2 + m_l^2)^{\nu_l}} \qquad [7]$$

Here $Q_E(k)$ is a complex $N \times N$ matrix with polynomial entries being $\tau$-covariant, $\tau(\Lambda) Q_E (\Lambda^{-1}k) \tau(\Lambda)^{-1} = Q_E(k) \quad \forall \Lambda \in \mathrm{SO}(d),\ \ k \in \mathbb{R}^d$. $\nu_l \in \mathbb{N}$ and $m_1 \in \mathbb{C} \setminus (-\infty, 0)$ are parameters with the interpretation of the mass spectrum $(m_1, \ldots, m_P)$ and $(\nu_1, \ldots, \nu_P)$ the dipole degrees of the related masses. We restrict ourselves to the case of positive mass spectrum where $m_l > 0$, and in this case

$$p(t) = p(t, D) = \frac{\prod_{l=1}^{P} (t + m_l^2)^{\nu_l}}{\prod_{l=1}^{P} m_l^{2\nu_l}}, \quad t > 0 \qquad [8]$$

One can show that $\varphi$ obtained as the unique solution of the SPDE $D\varphi = \eta$ is a Euclidean covariant, mixing random field. The Schwinger functions (moments) of $\varphi$ are given by

$$S_n(f_1 \otimes \cdots \otimes f_n) = E[\varphi(f_1) \cdots \varphi(f_n)], \quad f_1, \ldots, f_n \in \mathcal{S}_{\mathbb{R}} \qquad [9]$$

Now the Schwinger functions can be calculated explicitly. They are determined by the truncated Schwinger functions, cf. [4], as follows: for $n = 2$,

$$S^{\mathrm{T}}_{2,\alpha_1,\alpha_2}(x_1, x_2) = \frac{Q^{\mathrm{E}}_{2,\alpha_1,\alpha_2}(-\mathrm{i}\,\underline{\nabla}_2)}{\prod_{l=1}^{N} m_l^{2\nu_l}} \left[ \prod_{l=1}^{N} (-\Delta + m_l^2)^{-\nu_l} \right] (x_1 - x_2) \qquad [10]$$

and for $n \geq 3$

$$S^{\mathrm{T}}_{n,\alpha_1\cdots\alpha_n}(x_1, \ldots, x_n) = Q^{\mathrm{E}}_{n,\alpha_1\cdots\alpha_n}(-\mathrm{i}\,\underline{\nabla}_n) \times \int_{\mathbb{R}^d} \prod_{j=1}^{n} \left[ \prod_{l=1}^{N} (-\Delta + m_l^2)^{-\nu_l} \right] (x_j - x)\,\mathrm{d}x \qquad [11]$$

where

$$Q^{\mathrm{E}}_{n,\alpha_1\cdots\alpha_n}(-\mathrm{i}\,\underline{\nabla}_n) = C^{\beta_1\cdots\beta_n} \prod_{l=1}^{n} Q_{E,\beta_l,\alpha_l}\left(-\mathrm{i}\frac{\partial}{\partial x_l}\right) \qquad [12]$$

with

$$C_{\beta_1\cdots\beta_n} = (-\mathrm{i})^n \frac{\partial^n \psi(t)}{\partial t_{\beta_1} \cdots \partial t_{\beta_n}}\bigg|_{t=0} \qquad [13]$$

and the Einstein convention of summation and raising/lowering of indices on $\mathbb{R}^N$ with respect to the invariant inner product $\cdot$ is applied. The Schwinger functions fulfill the requirements of $\tau$-covariance, symmetry, clustering, and Hermiticity from the Osterwalder–Schrader axioms of Euclidean QFT.

While there is no known general reason why a relativistic QFT should exist for a given set of Schwinger functions, one can take advantage of the explicit formulas [10]–[13] in order to calculate the analytic continuation from Euclidean to relativistic times explicitly.

It simplifies the considerations to exclude dipole fields, that is, one assumes that $\nu_l = 1$ for $l = 1, \ldots, n$. In physical terms, the no-dipole condition guarantees that the asymptotic fields in Minkowski spacetime fulfill the Klein–Gordon equation and thus generate particles in the usual sense if applied to the vacuum. If this condition is not imposed, asymptotic fields might only fulfill a dipole equation $(\Box + m^2)^2 \phi^{\text{in/out}} = 0$ or a related hyperbolic equation of even higher order, and the particle states generated by application of such fields to the vacuum require a gauge fixing (constraints) in order to obtain a physical interpretation. Given the no-dipole condition, one obtains by expansion into partial fractions

$$\frac{1}{\prod_{l=1}^{P}(|k|^2 + m_l^2)} = \sum_{l=1}^{N} \frac{b_l}{(|k|^2 + m_l^2)} \qquad [14]$$

with $b_l \in (0, \infty)$ uniquely determined and $b_l \neq 0$. For the truncated Schwinger functions, this implies $(n \geq 3)$ that

$$\begin{aligned} S^{\mathrm{T}}_{n,\alpha_1\cdots\alpha_n}&(x_1, \ldots, x_n) \\ &= Q^{\mathrm{E}}_{n,\alpha_1\cdots\alpha_n}(-\mathrm{i}\,\underline{\nabla}_n) \sum_{l_1,\ldots,l_n=1}^{P} \\ &\quad \times \prod_{r=1}^{n} b_{l_r} \int_{\mathbb{R}^d} \prod_{j=1}^{n} (-\Delta + m_{l_j}^2)^{-1}(x - x_j)\, \mathrm{d}x \end{aligned} \qquad [15]$$

At this point, a lengthy calculation yields a representation of the functions $\int_{\mathbb{R}^d} \prod_{j=1}^{n} (-\Delta + m_j^2)^{-1} (x - x_j)\, \mathrm{d}x$ as the Fourier–Laplace transform of a distribution $\hat{W}^{\mathrm{T}}_{n,m_1,\ldots,m_n}$ that fulfills the spectral condition. This is equivalent to the statement that the analytic continuation of such functions to relativistic times yields $W^{\mathrm{T}}_{n,m_1,\ldots,m_n}$, where the latter distribution is the inverse Fourier transform of $\hat{W}^{\mathrm{T}}_{n,m_1,\ldots,m_n}$. This distribution up to a constant that can be integrated into $Q^{\mathrm{E}}$ is given by

$$\left\{ \sum_{j=1}^{n} \prod_{l=1}^{j-1} \delta^{-}_{m_l}(k_l) \frac{(-1)}{k^2 - m_j^2} \prod_{l=j+1}^{n} \delta^{+}_{m_l}(k_l) \right\} \delta\left( \sum_{l=1}^{n} k_l \right) \qquad [16]$$

Here $\delta^{\pm}_m(k) = \theta(\pm k^0)\delta(k^2 - m^2)$, where $\theta$ is the Heaviside step function and $k^2 = k^{0^2} - |\boldsymbol{k}|^2$. On the other hand, the partial differential operator $Q^{\mathrm{E}}_n$ can be analytically continued in momentum space:

$$\begin{aligned} Q^{\mathrm{M}}_n&((k_1^0, \boldsymbol{k}_1), \ldots, (k_n^0, \boldsymbol{k}_n)) \\ &= Q^{\mathrm{E}}_n((\mathrm{i}k_1^0, \boldsymbol{k}_1), \ldots, (\mathrm{i}k_n^0, \boldsymbol{k}_n)) \end{aligned} \qquad [17]$$

$k_1, \ldots, k_n \in \mathbb{R}^d$. With the definition

$$\begin{aligned} \hat{W}^{\mathrm{T}}_{2,\alpha_1\alpha_2}(k_1, k_2) &= (2\pi)^{(d+1)} \frac{Q^{\mathrm{M}}_{2,\alpha_1\alpha_2}(k_1, k_2)}{\prod_{l=1}^{N} m_l^2} \\ &\quad \times \sum_{l=1}^{N} b_l\, \delta^{-}_{m_l}(k_1)\delta(k_1 + k_2) \end{aligned} \qquad [18]$$

and

$$\begin{aligned} \hat{W}^{\mathrm{T}}_{n,\alpha_1\cdots\alpha_n}&(k_1, \ldots, k_n) \\ &= Q^{\mathrm{M}}_{n,\alpha_1\cdots\alpha_n}(k_1, \ldots, k_n) \\ &\quad \times \sum_{l_1,\ldots,l_n=1}^{N} \prod_{j=1}^{n} b_{l_j}\, \hat{W}^{\mathrm{T}}_{n,m_{l_1},\ldots,m_{l_n}}(k_1, \ldots, k_n) \end{aligned} \qquad [19]$$

the analytic continuation of Schwinger functions can be summarized as follows:

**Theorem 2** *The truncated Schwinger functions $\mathcal{S}^{\mathrm{T}}_n$ have a Fourier–Laplace representation with $\hat{W}^{\mathrm{T}}_n$ defined in eqns [18] and [19]. Equivalently, $\mathcal{S}^{\mathrm{T}}_n$ is the analytic continuation of $W^{\mathrm{T}}_n$ from purely real relativistic time to purely imaginary Euclidean time. The truncated Wightman functions $W^{\mathrm{T}}_n$ fulfill the requirements of temperedness, relativistic covariance with respect to the representation of the orthochronous, proper Lorentz group $\tilde{\tau} : \mathrm{L}^{\uparrow}_{+}(d) \to$* Gl(L), *locality, spectral property, and cluster property. Here $\tilde{\tau}$ is obtained by analytic continuation of $\tau$ to a representation of the proper complex Lorentz group over $\mathbb{C}^d$ (which contains* SO($d$) *as a real submanifold) and restriction of this representation to the real orthochronous proper Lorentz group.*

Again making use of the explicit formula in Theorem 2, the condition of Theorem 1 can be verified. This proves the existence of IMQFT models associated with the class of random fields under discussion.

**Theorem 3** *The Wightman functions defined in Theorem 2 fulfill the HSSC [2]. In particular, there exists a QFT with indefinite metric such that the Wightman functions are given as the VEVs of that IMQFT.*

## Nontrivial Scattering

Theories as described in Theorem 2 obviously have trivial scattering behavior if the noise field $\eta$ is Gaussian, that is, if, in [7], $z = 0$. In the case where there is also a Poisson component in $\eta$, that is, $z > 0$, higher-order truncated Wightman functions do not vanish and such relativistic theories have nontrivial scattering.

Before the scattering of the models can be discussed, some comments about scattering in IMQFT in general are in order. The scattering

theory in axiomatic QFT, Haag–Ruelle theory, relies on positivity. In fact, one can show that in the class of models under discussion, the LSZ asymptotic condition is violated if dipole degrees of freedom are admitted. In that case more complicated asymptotic conditions have to be used. In any case, the Haag–Ruelle theory cannot be adapted to IMQFT.

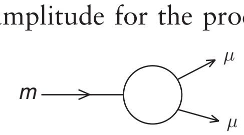
Nevertheless, asymptotic fields and states can be constructed in IMQFT if one imposes a no-dipole condition in a mathematically precise way. Then the LSZ asymptotic condition leads to the construction of mixed VEVs of asymptotic in- and out-fields with local fields. The collection of such VEVs is called the form-factor functional. After constructing this collection of mixed VEVs, one can try to check the HSSC for this functional and obtains a Krein space representation for the algebra generated by in- local and out-fields.

Following this line, asymptotic in- and out-particle states can be constructed for the given mass spectrum $(m_1, \ldots, m_P)$. If $a^{\mathrm{in/out}\dagger}_{\alpha, l}(k)$, $l = 1, \ldots, P$, denotes the creation operator for an incoming/outgoing particle with mass $m_l$, spin component $\alpha$, and energy–momentum $k$, the following scattering amplitude can be derived for $r$ incoming particles with masses $m_{l_1}, \ldots, m_{l_r}$ and $n - r$ outgoing particles with masses $m_{l_{r+1}}, \ldots, m_{l_n}$:

$$\left\langle a^{\mathrm{in}\dagger}_{\alpha_1, l_1}(k_1) \cdots a^{\mathrm{in}\dagger}_{\alpha_r, l_r}(k_r)\Omega, a^{\mathrm{out}\dagger}_{\alpha_{r+1}, l_{r+1}}(k_{r+1}) \cdots a^{\mathrm{out}\dagger}_{\alpha_n, l_n}(k_n)\Omega \right\rangle^{\mathrm{T}}$$
$$= -(2\pi)\mathrm{i}Q^{\mathrm{M}}_{\alpha_1,\ldots,\alpha_n}(-k_1, \ldots, -k_r, k_{r+1}, \ldots, k_n)$$
$$\times \prod_{j=1}^{n} \delta^{+}_{m_{l_j}}(k_j)\, \delta(K^{\mathrm{in}} - K^{\mathrm{out}}) \qquad [20]$$

$K^{\mathrm{in/out}}$ stand for the total energy–momentum of in- and out-particles, that is, $K^{\mathrm{in}} = \sum_{j=1}^{r} k_j$ and $K^{\mathrm{out}} = \sum_{j=r+1}^{n} k_j$.

Two immediate consequences can be drawn from [20]. First, choosing a model with nonvanishing Poisson part such that $C_{\beta_1\beta_2\beta_3} \neq 0$ and a differential operator $D$ containing in its mass spectrum the masses $m$ and $\mu$ with $m > 2\mu$, one gets a nonvanishing scattering amplitude for the process




[21]

even though in- and out-particle states consist of particles with well-defined sharp masses. Thus, for the incoming particle, the energy uncertainty, which for a particle at rest is proportional to the mass uncertainty, vanishes but still the particle undergoes a nontrivial decay and must have a finite decay time. This appears to be a contradiction to the energy–time uncertainty relation, which therefore seems to have an unclear status in IMQFT (i.e., in QFT including gauge fields). The origin of this inequality, which of course is experimentally very well tested, apparently has to be located in the constraints, that is, in the procedure of implementing a gauge, of the theory and not in the unconstrained IMQFT.

Second, one can replace somewhat artificially the polynomials $Q^{\mathrm{M}}_n$ in [17] by any other symmetric and relativistically covariant polynomial. If the sequence of the "new" $Q^{\mathrm{M}}_n$ is of uniformly bounded degree in any of the arguments $k_1, \ldots, k_n$, the redefined Wightman functions in [17] still fulfill the requirements of Theorem 1 and thus define a new relativistic, local IMQFT. The scattering amplitudes of such a theory are again well defined and given by [20]. For example, in the case of only one scalar particle with mass $m$, one can show that arbitrary Lorentz-invariant scattering behavior of bosonic particles can be reproduced by such theories for energies below an arbitrary maximal energy up to arbitrary precision. This kind of interpolation theorem shows that the outcome of an arbitrary scattering experiment can be reproduced within the formalism of (unconstrained) IMQFT as long as it is in agreement with the general requirements of Poincaré invariance and statistics.

## List of Symbols

| | |
|---|---|
| $\rightarrow$ | converges to |
| $\xrightarrow{\mathcal{L}}$ | convergence in law |
| $\mathbb{N}$ | set of natural numbers |
| $\mathbb{N}_0$ | set of natural numbers and zero |
| $\mathbb{R}$ | set of real numbers |
| $\mathbb{C}$ | set of complex numbers |
| $\mathbf{1}$ | identity mapping |
| $\restriction_{\mathcal{D}}$ | restricted to $\mathcal{D}$ |
| $x^0$ and $\boldsymbol{x}$ | time and spatial part of $x = (x^0, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{R}^{d-1}$ |
| $\underline{\nabla}_n$ | gradient operator on $\mathbb{R}^{dn}$ |

*See also:* Algebraic Approach to Quantum Field Theory; Euclidean Field Theory; Indefinite Metric; Perturbative Renormalization Theory and BRST; Quantum Field Theory in Curved Spacetime; Quantum Field Theory: A Brief Introduction; Stochastic Differential Equations.

## Further Reading


Albeverio S and Gottschalk H (2001) Scattering theory for quantum fields with indefinite metric. *Communications in Mathematical Physics* 216: 491–513.

Albeverio S, Gottschalk H, and Wu J-L (1997) Models of local relativistic quantum fields with indefinite metric (in all dimensions). *Communications in Mathematical Physics* 184: 509–531.

Gottschalk H (2001) On the stability of one particle states generated by quantum fields fulfilling Yang–Feldman equations. *Reports on Mathematical Physics* 47(2): 241–246.

Grothaus M and Streit L (1999) Construction of relativistic quantum fields in the framework of white noise analysis. *Journal of Mathematical Physics* 40(11): 5387.
Morchio G and Strocchi F (1980) Infrared singularities, vacuum structure and pure phases in local quantum field theory. *Ann. Inst. H. Poincaré* 33: 251.
Steinmann O (2000) *Perturbative Quantum Electrodynamics and Axiomatic Field Theory*. Berlin: Springer.
Strocchi F (1993) *Selected Topics on the General Properties of Quantum Field Theory*. Lecture Notes in Physics, vol. 51. Singapore: World Scientific.

# Quantum Fields with Topological Defects

**M Blasone and G Vitiello**, Università degli Studi di Salerno, Baronissi (SA), Italy
**P Jizba**, Czech Technical University, Prague, Czech Republic



## Introduction

The ordered patterns we observe in condensed matter and in high-energy physics are created by the quantum dynamics. Macroscopic systems exhibiting some kind of ordering, such as superconductors, ferromagnets, and crystals, are described by the underlying quantum dynamics. Even the large-scale structures in the universe, as well as the ordering in the biological systems appear to be the manifestation of the microscopic dynamics governing the elementary components of these systems. Thus, we talk of macroscopic quantum systems: these are quantum systems in the sense that, although they behave classically, some of their macroscopic features nevertheless cannot be understood without recourse to quantum theory.

The question then arises how the quantum dynamics generates the observed macroscopic properties. In other words, how it happens that the macroscopic scale characterizing those systems is dynamically generated out of the microscopic scale of the quantum elementary components (Umezawa 1993, Umezawa *et al.* 1982).

Moreover, we also observe a variety of phenomena where quantum particles coexist and interact with extended macroscopic objects which show a classical behavior, for example, vortices in superconductors and superfluids, magnetic domains in ferromagnets, dislocations and other topological defects (grain boundaries, point defects, etc.) in crystals, and so on.

We are thus also faced with the question of the quantum origin of topological defects and their interaction with quanta (Umezawa 1993, Umezawa *et al.* 1982): this is a crucial issue for the understanding of symmetry-breaking phase transitions and structure formation in a wide range of systems from condensed matter to cosmology (Kibble 1976, Zurek 1997, Volovik 2003).

Here, we will review how the generation of ordered structures and extended objects is explained in quantum field theory (QFT). We follow Umezawa (1993) and Umezawa *et al.* (1982) in our presentation. We will consider systems in which spontaneous symmetry breaking (SSB) occurs and show that topological defects originate by inhomogeneous (localized) condensation of quanta. The approach followed here is alternative to the usual one (Rajaraman 1982), in which one starts from the classical soliton solutions and then "quantizes" them, as well as to the QFT method based on dual (disorder) fields (Kleinert 1989).

In the next section we introduce some general features of QFT useful for our discussion and treat some aspects of SSB and the rearrangement of symmetry. Next we discuss the boson transformation theorem and the topological singularities of the boson condensate. We then present, as an example, a model with U(1) gauge invariance in which SSB, rearrangement of symmetry, and topological defects are present (Matsumoto *et al.* 1975a, b). There we show how macroscopic fields and currents are obtained from the microscopic quantum dynamics. The Nielsen–Olesen vortex solution is explicitly obtained as an example. The final section is devoted to conclusions.

## Symmetry and Order in QFT: A Dynamical Problem

QFT deals with systems with infinitely many degrees of freedom. The fields used for their description are operator fields whose mathematical significance is fully specified only when the state space where they operate is also assigned. This is the space of the states, or physical phase, of the system under given boundary conditions. A change in the boundary conditions may result in the transition of the system from one phase to another. For example, a change of temperature from above to below the critical temperature may induce the transition from the

normal to the superconducting phase in a metal. The identification of the state space where the field operators have to be realized is thus a physically nontrivial problem in QFT. In this respect, the QFT structure is drastically different from the one of quantum mechanics (QM). The reason is the following.

The von Neumann theorem (1955) in QM states that for systems with a finite number of degrees of freedom all the irreducible representations of the canonical commutation relations are unitarily equivalent. Therefore, in QM the physical system can only live in one single physical phase: unitary equivalence means indeed physical equivalence and thus there is no room (no representations) for physically different phases. Such a situation drastically changes in QFT where systems with infinitely many degrees of freedom are treated. In such a case, the von Neumann theorem does not hold and infinitely many unitarily inequivalent representations of the canonical commutation relations do in fact exist (Umezawa 1993, Umezawa *et al.* 1982). It is such richness of QFT that allows the description of different physical phases.

### QFT as a Two-Level Theory

In the perturbative approach, any quantum experiment or observation can be schematized as a scattering process where one prepares a set of free (noninteracting) particles (incoming particles or in-fields) which are then made to collide at some later time in some region of space (spacetime region of interaction). The products of the collision are expected to emerge out of the interaction region as free particles (outgoing particles or out-fields). Correspondingly, one has the in-field and the out-field state space. The interaction region is where the dynamics operates: given the in-fields and the in-states, the dynamics determines the out-fields and the out-states.

The incoming particles and the outgoing ones (also called quasiparticles in solid state physics) are well distinguishable and localizable particles only far away from the interaction region, at a time much before ($t=-\infty$) and much after ($t=+\infty$) the interaction time: in- and out-fields are thus said to be asymptotic fields, and for them the interaction forces are assumed not to operate (switched off).

The only regions accessible to observations are those far away (in space and in time) from the interaction region, that is, the asymptotic regions (the in- and out-regions). It is so since, at the quantum level, observations performed in the interaction region or vacuum fluctuations occurring there may drastically interfere with the interacting objects, thus changing their nature. Besides the asymptotic fields, one then also introduces dynamical or Heisenberg fields, that is, the fields in terms of which the dynamics is given. Since the interaction region is precluded from observation, we do not observe Heisenberg fields. Observables are thus solely described in terms of asymptotic fields.

Summing up, QFT is a "two-level" theory: one level is the interaction level where the dynamics is specified by assigning the equations for the Heisenberg fields. The other level is the physical level, the one of the asymptotic fields and of the physical state space directly accessible to observations. The equations for the physical fields are equations for free fields, describing the observed incoming/outgoing particles.

To be specific, let the Heisenberg operator fields be generically denoted by $\psi_{\mathrm{H}}(x)$ and the physical operator fields by $\varphi_{\mathrm{in}}(x)$. For definiteness, we choose to work with the in-fields, although the set of out-fields would work equally well. They are both assumed to satisfy equal-time canonical (anti)-commutation relations.

For brevity, we omit considerations on the renormalization procedure, which are not essential for the conclusions we will reach. The Heisenberg field equations and the free-field equations are written as

$$\Lambda(\partial)\psi_{\mathrm{H}}(x) = \mathcal{J}[\psi_{\mathrm{H}}](x) \qquad [1]$$

$$\Lambda(\partial)\varphi_{\mathrm{in}}(x) = 0 \qquad [2]$$

where $\Lambda(\partial)$ is a differential operator, $x \equiv (t, \boldsymbol{x})$ and $\mathcal{J}$ is some functional of the $\psi_{\mathrm{H}}$ fields, describing the interaction.

Equation [1] can be formally recast in the following integral form (Yang–Feldman equation):

$$\psi_{\mathrm{H}}(x) = \varphi_{\mathrm{in}}(x) + \Lambda^{-1}(\partial) * \mathcal{J}[\psi_{\mathrm{H}}](x) \qquad [3]$$

where $*$ denotes convolution. The symbol $\Lambda^{-1}(\partial)$ denotes formally the Green function for $\varphi_{\mathrm{in}}(x)$. The precise form of Green's function is specified by the boundary conditions. Equation [3] can be solved by iteration, thus giving an expression for the Heisenberg fields $\psi_{\mathrm{H}}(x)$ in terms of powers of the $\varphi_{\mathrm{in}}(x)$ fields; this is the Haag expansion in the LSZ formalism (or "dynamical map" in the language of Umezawa 1993 and Umezawa *et al.* 1982), which might be formally written as

$$\psi_{\mathrm{H}}(x) = F[x; \varphi_{\mathrm{in}}] \qquad [4]$$

(A (formal) closed form for the dynamical map is obtained in the closed time path (CTP) formalism (Blasone and Jizba 2002). Then the Haag expansion [4] is directly applicable to both equilibrium and nonequilibrium situations.)

We stress that the equality in the dynamical map [4] is a "weak" equality, which means that it must be understood as an equality among matrix elements computed in the Hilbert space of the physical particles.

We observe that mathematical consistency in the above procedure requires that the set of $\varphi_{\rm in}$ fields must be an irreducible set; however, it may happen that not all the elements of the set are known from the beginning. For example, there might be composite (bound states) fields or even elementary quanta whose existence is ignored in a first recognition. Then the computation of the matrix elements in physical states will lead to the detection of unexpected poles in the Green's functions, which signal the existence of the ignored quanta. One thus introduces the fields corresponding to these quanta and repeats the computation. This way of proceeding is called the self- consistent method (Umezawa 1993, Umezawa *et al.* 1982). Thus it is not necessary to have a one-to-one correspondence between the sets $\{\psi^j_{\rm H}\}$ and $\{\varphi^i_{\rm in}\}$, as it happens whenever the set $\{\varphi^i_{\rm in}\}$ includes composite particles.

### The Dynamical Rearrangement of Symmetry

As already mentioned, in QFT the Fock space for the physical states is not unique since one may have several physical phases, for example, for a metal the normal phase and the superconducting phase, and so on. Fock spaces describing different phases are unitarily inequivalent spaces and correspondingly we have different expectation values for certain observables and even different irreducible sets of physical quanta. Thus, finding the dynamical map involves singling out the Fock space where the dynamics has to be realized.

Let us now suppose that the Heisenberg field equations are invariant under some group $G$ of transformations of $\psi_{\rm H}$:

$$\psi_{\rm H}(x) \rightarrow \psi'_{\rm H}(x) = g[\psi_{\rm H}(x)] \qquad [5]$$

with $g \in G$. The symmetry is spontaneously broken when the vacuum state in the Fock space $\mathcal{H}$ is not invariant under the group $G$ but only under one of its subgroups (Umezawa 1993, Umezawa *et al.* 1982).

On the other hand, eqn [4] implies that when $\psi_{\rm H}$ is transformed as in [5], then

$$\varphi_{\rm in}(x) \rightarrow \varphi'_{\rm in}(x) = g'[\varphi_{\rm in}(x)] \qquad [6]$$

with $g'$ belonging to some group of transformations $G'$ and such that

$$g[\psi_{\rm H}(x)] = F[g'[\varphi_{\rm in}(x)]] \qquad [7]$$

When symmetry is spontaneously broken it is $G' \neq G$, with $G'$ the group contraction of $G$; when symmetry is not broken then $G' = G$.

Since $G$ is the invariance group of the dynamics, eqn [4] requires that $G'$ is the group under which free fields equations are invariant, that is, also $\varphi'_{\rm in}$ is a solution of [2]. Since eqn [4] is a weak equality, $G'$ depends on the choice of the Fock space $\mathcal{H}$ among the physically realizable unitarily inequivalent state spaces. Thus, we see that the (same) original invariance of the dynamics may manifest itself in different symmetry groups for the $\varphi_{\rm in}$ fields according to different choices of the physical state space. Since this process is constrained by the dynamical equations [1], it is called the dynamical rearrangement of symmetry (Umezawa 1993, Umezawa *et al.* 1982).

In conclusion, different ordering patterns appear to be different manifestations of the same basic dynamical invariance. The discovery of the process of the dynamical rearrangement of symmetry leads to a unified understanding of the dynamical generation of many observable ordered patterns. This is the phenomenon of the dynamical generation of order. The contraction of the symmetry group is the mathematical structure controlling the dynamical rearrangement of the symmetry. For a qualitative presentation see Vitiello (2001).

One can now ask which ones are the carriers of the ordering information among the system elementary constituents and how the long-range correlations and the coherence observed in ordered patterns are generated and sustained. The answer is in the fact that SSB implies the appearance of bosons (Goldstone 1961, Goldstone *et al.* 1962, Nambu and Jona-Lasinio 1961), the so-called Nambu–Goldstone (NG) modes or quanta. They manifest as long-range correlations and thus they are responsible of the above-mentioned change of scale, from microscopic to macroscopic. The coherent boson condensation of NG modes turns out to be the mechanism by which order is generated, as we will see in an explicit example in a later section.

## The "Boson Transformation" Method

We now discuss the quantum origin of extended objects (defects) and show how they naturally emerge as macroscopic objects (inhomogeneous condensates) from the quantum dynamics. At zero temperature, the classical soliton solutions are then recovered in the Born approximation. This approach is known as the "boson transformation" method (Umezawa 1993, Umezawa *et al.* 1982).

### The Boson Transformation Theorem

Let us consider, for simplicity, the case of a dynamical model involving one scalar field $\psi_{\rm H}$ and one asymptotic field $\varphi_{\rm in}$ satisfying eqns [1] and [2], respectively.

As already remarked, the dynamical map is valid only in a weak sense, that is, as a relation among matrix elements. This implies that eqn [4] is not unique, since different sets of asymptotic fields and the corresponding Hilbert spaces can be used in its construction. Let us indeed consider a c–number function $f(x)$, satisfying the $\varphi_{\rm in}$ equations of motion [2]:

$$\Lambda(\partial)f(x) = 0 \qquad [8]$$

The boson transformation theorem (Umezawa 1993, Umezawa *et al.* 1982) states that the field

$$\psi^f_{\rm H}(x) = F[x; \varphi_{\rm in} + f] \qquad [9]$$

is also a solution of the Heisenberg equation [1]. The corresponding Yang–Feldman equation takes the form

$$\psi^f_{\rm H}(x) = \varphi_{\rm in}(x) + f(x) + \Lambda^{-1}(\partial) * \mathcal{J}[\psi^f_{\rm H}](x) \qquad [10]$$

The difference between the two solutions $\psi_{\rm H}$ and $\psi^f_{\rm H}$ is only in the boundary conditions. An important point is that the expansion in [9] is obtained from that in [4] by the spacetime-dependent translation

$$\varphi_{\rm in}(x) \to \varphi_{\rm in}(x) + f(x) \qquad [11]$$

The essence of the boson transformation theorem is that the dynamics embodied in eqn [1] contains an internal freedom, represented by the possible choices of the function $f(x)$, satisfying the free-field equation [8].

We also observe that the transformation [11] is a canonical transformation since it leaves invariant the canonical form of commutation relations.

Let $|0\rangle$ denote the vacuum for the free field $\varphi_{\rm in}$. The vacuum expectation value of eqn [10] gives

$$\begin{aligned}\phi^f(x) &\equiv \langle 0|\psi^f_{\rm H}(x)|0\rangle \\ &= f(x) + \left\langle 0\middle|\left[\Lambda^{-1}(\partial) * \mathcal{J}[\psi^f_{\rm H}](x)\right]\middle|0\right\rangle \end{aligned} \qquad [12]$$

The c–number field $\phi^f(x)$ is the order parameter. We remark that it is fully determined by the quantum dynamics. In the classical or Born approximation, which consists in taking $\langle 0|\mathcal{J}[\psi^f_{\rm H}]|0\rangle = \mathcal{J}[\phi^f]$, that is, neglecting all the contractions of the physical fields, we define $\phi^f_{\rm cl}(x) \equiv \lim_{\hbar\to 0}\phi^f(x)$. In this limit, we have

$$\Lambda(\partial)\phi^f_{\rm cl}(x) = \mathcal{J}[\phi^f_{\rm cl}](x) \qquad [13]$$

that is, $\phi^f_{\rm cl}(x)$ provides the solution of the classical Euler–Lagrange equation.

Beyond the classical level, in general, the form of this equation changes. The Yang–Feldman equation [10] gives not only the equation for the order parameter, eqn [13], but also, at higher orders in $\hbar$, the dynamics of the physical quanta in the potential generated by the "macroscopic object" $\phi^f(x)$ (Umezawa 1993, Umezawa *et al.* 1982).

One can show (Umezawa 1993, Umezawa *et al.* 1982) that the class of solutions of eqn [8] which lead to topologically nontrivial (i.e., carrying a nonzero topological charge) solutions of eqn [13], are those which have some sort of singularity with respect to Fourier transform. These can be either divergent singularities or topological singularities. The first are associated to a divergence of $f(x)$ for $|\boldsymbol{x}| = \infty$, at least in some direction. Topological singularities are instead present when $f(x)$ is not single-valued, that is, it is path dependent. In both cases, the macroscopic object described by the order parameter, carries a nonzero topological charge.

### Topological Singularities and Massless Bosons

An important result is that the boson transformation functions carrying topological singularities are only allowed for massless bosons (Umezawa 1993, Umezawa *et al.* 1982).

Consider a generic boson field $\chi_{\rm in}$ satisfying the equation

$$(\partial^2 + m^2)\chi_{\rm in}(x) = 0 \qquad [14]$$

and suppose that the function $f(x)$ for the boson transformation $\chi_{\rm in}(x) \to \chi_{\rm in}(x) + f(x)$ carries a topological singularity. It is then not single-valued and thus path dependent:

$$G^+_{\mu\nu}(x) \equiv [\partial_\mu, \partial_\nu]f(x) \neq 0, \quad \text{for certain}\, \mu, \nu, x \qquad [15]$$

On the other hand, $\partial_\mu f(x)$, which is related with observables, is single-valued, that is, $[\partial_\rho, \partial_\nu]\partial_\mu f(x) = 0$. Recall that $f(x)$ is solution of the $\chi_{\rm in}$ equation:

$$(\partial^2 + m^2)f(x) = 0 \qquad [16]$$

From the definition of $G^+_{\mu\nu}(x)$ and the regularity of $\partial_\mu f(x)$, it follows, by computing $\partial^\mu G^+_{\mu\nu}(x)$, that

$$\partial_\mu f(x) = \frac{1}{\partial^2 + m^2}\partial^\lambda G^+_{\lambda\mu}(x) \qquad [17]$$

This equation and the antisymmetric nature of $G^+_{\mu\nu}(x)$ then lead to $\partial^2 f(x) = 0$, which in turn implies $m = 0$. Thus, we conclude that [15] is only compatible with massless equation for $\chi_{\rm in}$.

The topological charge is defined as

$$N_{\mathrm{T}} = \int_C \mathrm{d}l^\mu \partial_\mu f = \int_S \mathrm{d}S_\mu \epsilon^{\mu\nu\sigma}\partial_\nu\partial_\sigma f = \frac{1}{2}\int_S \mathrm{d}S^{\mu\nu} G^+_{\mu\nu} \qquad [18]$$

Here $C$ is a contour enclosing the singularity and $S$ a surface with $C$ as boundary. $N_{\mathrm{T}}$ does not depend on the path $C$ provided this does not cross the singularity. The dual tensor $G^{\mu\nu}(x)$ is

$$G^{\mu\nu}(x) \equiv -\tfrac{1}{2}\epsilon^{\mu\nu\lambda\rho}G^+_{\lambda\rho}(x) \qquad [19]$$

and satisfies the continuity equation

$$\partial_\mu G^{\mu\nu}(x) = 0 \Leftrightarrow \partial_\mu G^+_{\lambda\rho}(x) + \partial_\rho G^+_{\mu\lambda}(x) + \partial_\lambda G^+_{\rho\mu}(x) = 0 \qquad [20]$$

Equation [20] completely characterizes the topological singularity (Umezawa 1993, Umezawa *et al.* 1982).

## An Example: The Anderson–Higgs–Kibble Mechanism and the Vortex Solution

We consider a model of a complex scalar field $\phi(x)$ interacting with a gauge field $A_\mu(x)$ (Anderson 1958, Higgs 1960, Kibble 1967). The lagrangian density $\mathcal{L}[\phi(x), \phi^*(x), A_\mu(x)]$ is invariant under the global and the local U(1) gauge transformations (we do not assume a particular form for the Lagrangian density, so the following results are quite general):

$$\phi(x) \to \mathrm{e}^{\mathrm{i}\theta}\phi(x), \quad A_\mu(x) \to A_\mu(x) \qquad [21]$$

$$\phi(x) \to \mathrm{e}^{\mathrm{i}e_0\lambda(x)}\phi(x), \quad A_\mu(x) \to A_\mu(x) + \partial_\mu\lambda(x) \qquad [22]$$

respectively, where $\lambda(x) \to 0$ for $|x_0| \to \infty$ and/or $|\boldsymbol{x}| \to \infty$ and $e_0$ is the coupling constant. We work in the Lorentz gauge $\partial_\mu A^\mu(x) = 0$. The generating functional, including the gauge constraint, is (Matsumoto *et al.* 1975a, b)

$$\mathcal{Z}[J, K] = \frac{1}{\mathcal{N}}\int [\mathrm{d}A_\mu][\mathrm{d}\phi][\mathrm{d}\phi^*][\mathrm{d}B] \times \exp\left[\mathrm{i}\,\mathcal{S}[A_\mu, B, \phi]\right] \qquad [23]$$

$$\mathcal{S} = \int \mathrm{d}^4x\Big[\mathcal{L}(x) + B(x)\partial^\mu A_\mu(x) + K^*(x)\phi(x) + K(x)\phi^*(x) + J^\mu(x)A_\mu(x) + \mathrm{i}\epsilon|\phi(x) - v|^2\Big]$$

$$\mathcal{N} = \int [\mathrm{d}A_\mu][\mathrm{d}\phi][\mathrm{d}\phi^*][\mathrm{d}B] \times \exp\left[\mathrm{i}\int \mathrm{d}^4x\Big(\mathcal{L}(x) + \mathrm{i}\epsilon|\phi(x) - v|^2\Big)\right]$$

$B(x)$ is an auxiliary field which implements the gauge-fixing condition (Matsumoto *et al.* 1975a, b). Notice the $\epsilon$-term where $v$ is a complex number; its rôle is to specify the condition of symmetry breaking under which we want to compute the functional integral and it may be given the physical meaning of a small external field triggering the symmetry breaking (Matsumoto *et al.* 1975a, b). The limit $\epsilon \to 0$ must be made at the end of the computations. We will use the notation

$$\langle F[\phi]\rangle_{\epsilon,J,K} \equiv \frac{1}{\mathcal{N}}\int [\mathrm{d}A_\mu][\mathrm{d}\phi][\mathrm{d}\phi^*][\mathrm{d}B]F[\phi] \times \exp\left[\mathrm{i}\mathcal{S}[A_\mu, B, \phi]\right] \qquad [24]$$

with $\langle F[\phi]\rangle_\epsilon \equiv \langle F[\phi]\rangle_{\epsilon,J=K=0}$ and $\langle F[\phi]\rangle \equiv \lim_{\epsilon\to 0}\langle F[\phi]\rangle_\epsilon$.

The fields $\phi, A_\mu$, and $B$ appearing in the generating functional are c-number fields. In the following, the Heisenberg operator fields corresponding to them will be denoted by $\phi_{\mathrm{H}}, A_{\mathrm{H}\mu}$, and $B_{\mathrm{H}}$, respectively. Thus, the spontaneous symmetry breaking condition is expressed by $\langle 0|\phi_{\mathrm{H}}(x)|0\rangle \equiv \tilde{v} \neq 0$, with $\tilde{v}$ constant.

Since in the functional integral formalism the functional average of a given c-number field gives the vacuum expectation value of the corresponding operator field, for example, $\langle F[\phi]\rangle \equiv \langle 0|F[\phi_{\mathrm{H}}]|0\rangle$, we have $\lim_{\epsilon\to 0}\langle\phi(x)\rangle_\epsilon \equiv \langle 0|\phi_{\mathrm{H}}(x)|0\rangle = \tilde{v}$.

Let us introduce the following decompositions:

$$\phi(x) = \frac{1}{\sqrt{2}}[\psi(x) + \mathrm{i}\chi(x)]$$

$$K(x) = \frac{1}{\sqrt{2}}[K_1(x) + \mathrm{i}K_2(x)]$$

$$\rho(x) \equiv \psi(x) - \langle\psi(x)\rangle_\epsilon$$

Note that $\langle\chi(x)\rangle_\epsilon = 0$ because of the invariance under $\chi \to -\chi$.

### The Goldstone Theorem

Since the functional integral [23] is invariant under the global transformation [21], we have that $\partial\mathcal{Z}[J,K]/\partial\theta = 0$ and subsequent derivatives with respect to $K_1$ and $K_2$ lead to

$$\langle\psi(x)\rangle_\epsilon = \sqrt{2}\epsilon v\int \mathrm{d}^4y\langle\chi(x)\chi(y)\rangle_\epsilon = \sqrt{2}\epsilon v\Delta_\chi(\epsilon, 0) \qquad [25]$$

In momentum space the propagator for the field $\chi$ has the general form

$$\Delta_\chi(0, p) = \lim_{\epsilon\to 0}\left[\frac{Z_\chi}{p^2 - m_\chi^2 + \mathrm{i}\epsilon a_\chi} + (\text{continuum contributions})\right] \qquad [26]$$

Here $Z_\chi$ and $a_\chi$ are renormalization constants. The integration in eqn [25] picks up the pole contribution at $p^2 = 0$, and leads to

$$\tilde{v} = \sqrt{2}\frac{Z_\chi}{a_\chi} v \Leftrightarrow m_\chi = 0, \quad \tilde{v} = 0 \Leftrightarrow m_\chi \neq 0 \quad [27]$$

The Goldstone theorem (Goldstone 1961, Goldstone *et al.* 1962) is thus proved: if the symmetry is spontaneously broken ($\tilde{v} \neq 0$), a massless mode must exist, whose field is $\chi(x)$, that is, the NG boson mode. Since it is massless, it manifests as a long-range correlation mode. (Notice that in the present case of a complex scalar field model, the NG mode is an elementary field. In other models, it may appear as a bound state, for example, the magnon in (anti)ferromagnets.) Note that

$$\frac{\partial}{\partial v}\langle\psi(x)\rangle_\epsilon = \sqrt{2}\epsilon \int \mathrm{d}^4 y \langle\rho(x)\rho(y)\rangle_\epsilon \quad [28]$$

and because $m_\rho \neq 0$, the right-hand side of this equation vanishes in the limit $\epsilon \to 0$; therefore, $\tilde{v}$ is independent of $|v|$, although the phase of $|v|$ determines the one of $\tilde{v}$ (from eqn [25]): as in ferromagnets, once an external magnetic field is switched on, the system is magnetized independently of the strength of the external field.

### The Dynamical Map and the Field Equations

Observing that the change of variables [21] (and/or [22]) does not affect the generating functional, we may obtain the Ward–Takahashi identities. Also, using $B(x) \to B(x) + \lambda(x)$ in [23] gives $\langle\partial^\mu A_\mu(x)\rangle_{\epsilon,J,K} = 0$. One then finds the following two-point function pole structures (Matsumoto *et al.* 1975a, b):

$$\langle B(x)\chi(y)\rangle = \lim_{\epsilon\to 0}\left\{\frac{-\mathrm{i}}{(2\pi)^4}\int \mathrm{d}^4 p\, \mathrm{e}^{-\mathrm{i}p(x-y)}\frac{e_0\tilde{v}}{p^2 + \mathrm{i}\epsilon a_\chi}\right\} \quad [29]$$

$$\langle B(x)A^\mu(y)\rangle = \partial_x^\mu \frac{\mathrm{i}}{(2\pi)^4}\int \mathrm{d}^4 p\, \mathrm{e}^{-\mathrm{i}p(x-y)}\frac{1}{p^2} \quad [30]$$

$$\langle B(x)B(y)\rangle = \lim_{\epsilon\to 0}\left\{\frac{-\mathrm{i}}{(2\pi)^4}\int \mathrm{d}^4 p\, \mathrm{e}^{-\mathrm{i}p(x-y)}\frac{(e_0\tilde{v})^2}{Z_\chi} \times \left[\frac{1}{p^2 + \mathrm{i}\epsilon a_\chi} - \frac{1}{p^2}\right]\right\} \quad [31]$$

The absence of branch-cut singularities in propagators [29]–[31] suggests that $B(x)$ obeys a free-field equation. In addition, eqn [31] indicates that the model contains a massless negative-norm state (ghost) besides the NG massless mode $\chi$. Moreover, it can be shown (Matsumoto *et al.* 1975a, b) that a massive vector field $U_{\mathrm{in}}^\mu$ also exists in the theory. Note that because of the invariance $(\chi, A_\mu, B) \to (-\chi, -A_\mu, -B)$, all the other two-point functions must vanish.

The dynamical maps expressing the Heisenberg operator fields in terms of the asymptotic operator fields are found to be (Matsumoto *et al.* 1975a, b)

$$\phi_{\mathrm{H}}(x) = :\exp\left\{\mathrm{i}\frac{Z_\chi^{1/2}}{\tilde{v}}\chi_{\mathrm{in}}(x)\right\}\left[\tilde{v} + Z_\rho^{1/2}\rho_{\mathrm{in}}(x) + \mathcal{F}[\rho_{\mathrm{in}}, U_{\mathrm{in}}^\mu, \partial(\chi_{\mathrm{in}} - b_{\mathrm{in}})]\right]: \quad [32]$$

$$A_{\mathrm{H}}^\mu(x) = Z_3^{1/2}U_{\mathrm{in}}^\mu(x) + \frac{Z_\chi^{1/2}}{e_0\tilde{v}}\partial^\mu b_{\mathrm{in}}(x) + :\mathcal{F}^\mu[\rho_{\mathrm{in}}, U_{\mathrm{in}}^\mu, \partial(\chi_{\mathrm{in}} - b_{\mathrm{in}})]: \quad [33]$$

$$B_{\mathrm{H}}(x) = \frac{e_0\tilde{v}}{Z_\chi^{1/2}}[b_{\mathrm{in}}(x) - \chi_{\mathrm{in}}(x)] + c \quad [34]$$

where $:\ldots:$ denotes the normal ordering and the functionals $\mathcal{F}$ and $\mathcal{F}^\mu$ are to be determined within a particular model. In eqns [32]–[34], $\chi_{\mathrm{in}}$ denotes the NG mode, $b_{\mathrm{in}}$ the ghost mode, $U_{\mathrm{in}}^\mu$ the massive vector field, and $\rho_{\mathrm{in}}$ the massive matter field. In eqn [34] $c$ is a c-number constant, whose value is irrelevant since only derivatives of $B$ appear in the field equations (see below). $Z_3$ represents the wave function renormalization for $U_{\mathrm{in}}^\mu$. The corresponding field equations are

$$\partial^2\chi_{\mathrm{in}}(x) = 0, \qquad \partial^2 b_{\mathrm{in}}(x) = 0$$
$$(\partial^2 + m_\rho^2)\rho_{\mathrm{in}}(x) = 0 \quad [35]$$

$$(\partial^2 + m_V^2)U_{\mathrm{in}}^\mu(x) = 0, \quad \partial_\mu U_{\mathrm{in}}^\mu(x) = 0 \quad [36]$$

with $m_V{}^2 = (Z_3/Z_\chi)(e_0\tilde{v})^2$. The field equations for $B_{\mathrm{H}}$ and $A_{\mathrm{H}\mu}$ read (Matsumoto *et al.* 1975a, b)

$$\partial^2 B_{\mathrm{H}}(x) = 0, \quad -\partial^2 A_{\mathrm{H}\mu}(x) = j_{\mathrm{H}\mu}(x) - \partial_\mu B_{\mathrm{H}}(x) \quad [37]$$

with $j_{\mathrm{H}\mu}(x) = \delta\mathcal{L}(x)/\delta A_{\mathrm{H}}^\mu(x)$. One may then require that the current $j_{\mathrm{H}\mu}$ is the only source of the gauge field $A_{\mathrm{H}\mu}$ in any observable process. This amounts to impose the condition: ${}_p\langle b|\partial_\mu B_{\mathrm{H}}(x)|a\rangle_p = 0$, that is,

$$(-\partial^2)_p\langle b|A_{\mathrm{H}\mu}^0(x)|a\rangle_p = {}_p\langle b|\, j_{\mathrm{H}\mu}(x)|a\rangle_p \quad [38]$$

where $|a\rangle_p$ and $|b\rangle_p$ denote two generic physical states and $A_{\mathrm{H}}^{0\mu}(x) \equiv A_{\mathrm{H}}^\mu(x) - e_0\tilde{v} :\partial^\mu b_{\mathrm{in}}(x):$. Equations [38] are the classical Maxwell equations. The condition ${}_p\langle b|\partial_\mu B_{\mathrm{H}}(x)|a\rangle_p = 0$ leads to the Gupta–Bleuler–like condition

$$[\chi_{\mathrm{in}}^{(-)}(x) - b_{\mathrm{in}}^{(-)}(x)]|a\rangle_p = 0 \quad [39]$$

where $\chi_{\mathrm{in}}^{(-)}$ and $b_{\mathrm{in}}^{(-)}$ are the positive-frequency parts of the corresponding fields. Thus, we see that $\chi_{\mathrm{in}}$ and $b_{\mathrm{in}}$ cannot participate in any observable reaction.

This is confirmed by the fact that they are present in the $S$-matrix in the combination $(\chi_{\rm in} - b_{\rm in})$ (Matsumoto *et al.* 1975a, b). It is to be remarked, however, that the NG boson does not disappear from the theory: we shall see below that there are situations in which the NG fields do have observable effects.

### The Dynamical Rearrangement of Symmetry and the Classical Fields and Currents

From eqns [32]–[33] we see that the local gauge transformations of the Heisenberg fields

$$\begin{aligned}&\phi_{\rm H}(x) \rightarrow e^{ie_0\lambda(x)}\phi_{\rm H}(x)\\ &A^\mu_{\rm H}(x) \rightarrow A^\mu_{\rm H}(x) + \partial^\mu\lambda(x), \quad B_{\rm H}(x) \rightarrow B_{\rm H}(x)\end{aligned} \qquad [40]$$

with $\partial^2\lambda(x) = 0$, are induced by the in-field transformations

$$\begin{aligned}\chi_{\rm in}(x) &\rightarrow \chi_{\rm in}(x) + \frac{e_0\tilde{v}}{Z_\chi^{1/2}}\lambda(x)\\ b_{\rm in}(x) &\rightarrow b_{\rm in}(x) + \frac{e_0\tilde{v}}{Z_\chi^{1/2}}\lambda(x)\\ \rho_{\rm in}(x) &\rightarrow \rho_{\rm in}(x), \quad U^\mu_{\rm in}(x) \rightarrow U^\mu_{\rm in}(x)\end{aligned} \qquad [41]$$

On the other hand, the global phase transformation $\phi_{\rm H}(x) \rightarrow e^{i\theta}\phi_{\rm H}(x)$ is induced by

$$\begin{aligned}&\chi_{\rm in}(x) \rightarrow \chi_{\rm in}(x) + \frac{\tilde{v}}{Z_\chi^{1/2}}\theta f(x), \quad b_{\rm in}(x) \rightarrow b_{\rm in}(x)\\ &\rho_{\rm in}(x) \rightarrow \rho_{\rm in}(x), \quad U^\mu_{\rm in}(x) \rightarrow U^\mu_{\rm in}(x)\end{aligned} \qquad [42]$$

with $\partial^2 f(x) = 0$ and the limit $f(x) \rightarrow 1$ to be performed at the end of computations. Note that under the above transformations, the in-field equations and the $S$-matrix are invariant and that $B_{\rm H}$ is changed by an irrelevant c-number (in the limit $f \rightarrow 1$).

Consider now the boson transformation $\chi_{\rm in}(x) \rightarrow \chi_{\rm in}(x) + \alpha(x)$: in local gauge theories the boson transformation must be compatible with the Heisenberg field equations but also with the physical state condition [39]. Under the boson transformation with $\alpha(x) = \tilde{v}Z_\chi^{-1/2}\theta f(x)$ and $\partial^2 f(x) = 0$, $B_{\rm H}$ changes as

$$B_{\rm H}(x) \rightarrow B_{\rm H}(x) - \frac{e_0\tilde{v}^2}{Z_\chi}f(x) \qquad [43]$$

eqn [38] is thus violated when the Gupta–Bleuler-like condition is imposed. In order to restore it, the shift in $B_{\rm H}$ must be compensated by means of the following transformation on $U^\mu_{\rm in}$:

$$U^\mu_{\rm in}(x) \rightarrow U^\mu_{\rm in}(x) + Z_3^{-1/2}a^\mu(x), \quad \partial_\mu a^\mu(x) = 0 \qquad [44]$$

with a convenient c-number function $a^\mu(x)$. The dynamical maps of the various Heisenberg operators are not affected by [44] since they contain $U^\mu_{\rm in}$ and $B_{\rm H}$ in a combination such that the changes of $B_{\rm H}$ and of $U^\mu_{\rm in}$ compensate each other provided

$$(\partial^2 + m_V^2)a_\mu(x) = \frac{m_V^2}{e_0}\partial_\mu f(x) \qquad [45]$$

Equation [45] thus obtained is the Maxwell equation for the massive potential vector $a_\mu$ (Matsumoto *et al.* 1975a, b). The classical ground state current $j^\mu$ turns out to be

$$j^\mu(x) \equiv \langle 0|j^\mu_{\rm H}(x)|0\rangle = m_V^2\left[a^\mu(x) - \frac{1}{e_0}\partial^\mu f(x)\right] \qquad [46]$$

The term $m_V^2 a^\mu(x)$ is the Meissner current, while $(m_V^2/e_0)\partial^\mu f(x)$ is the boson current. The key point here is that both the macroscopic field and current are given in terms of the boson condensation function $f(x)$.

Two remarks are in order: first, note that the terms proportional to $\partial^\mu f(x)$ are related to observable effects, for example, the boson current which acts as the source of the classical field. Second, note that the macroscopic ground state effects do not occur for regular $f(x)(G^+_{\mu\nu}(x) = 0)$. In fact, from [45] we obtain $a_\mu(x) = (1/e_0)\partial_\mu f(x)$ for regular $f(x)$ which implies zero classical current $(j_\mu = 0)$ and zero classical field $(F_{\mu\nu} = \partial_\mu a_\nu - \partial_\nu a_\mu)$, since the Meissner and the boson current cancel each other.

In conclusion, the vacuum current appears only when $f(x)$ has topological singularities and these can be created only by condensation of massless bosons, that is, when SSB occurs. This explains why topological defects appear in the process of phase transitions, where NG modes are present and gradients in their condensate densities are nonzero (Kibble 1976, Zurek 1997).

On the other hand, the appearance of spacetime order parameter is no guarantee that persistent ground state currents (and fields) will exist: if $f(x)$ is a regular function, the spacetime dependence of $\tilde{v}$ can be gauged away by an appropriate gauge transformation.

Since, as already mentioned, the boson transformation with regular $f(x)$ does not affect observable quantities, the $S$-matrix is actually given by

$$S = :S\left[\rho_{\rm in}, U^\mu_{\rm in} - \frac{1}{m_V}\partial(\chi_{\rm in} - b_{\rm in})\right]: \qquad [47]$$

This is indeed independent of the boson transformation with regular $f(x)$:

$$\begin{aligned}S \rightarrow S' = :S\bigg[\rho_{\rm in}, U^\mu_{\rm in} &- \frac{1}{m_V}\partial(\chi_{\rm in} - b_{\rm in})\\ &+ Z_3^{-1/2}(a^\mu - \frac{1}{e_0}\partial^\mu f)\bigg]:\end{aligned} \qquad [48]$$

since $a_\mu(x) = (1/e_0)\partial_\mu f(x)$ for regular $f(x)$. However, $S' \neq S$ for singular $f(x)$: $S'$ includes the interaction of the quanta $U^\mu_{\rm in}$ and $\phi_{\rm in}$ with the classically behaving macroscopic defects (Umezawa 1993, Umezawa *et al.* 1982).

### The Vortex Solution

Below we consider the example of the Nielsen–Olesen vortex string solution. We show which one is the boson function $f(x)$ controlling the nonhomogeneous NG boson condensation in terms of which the string solution is described. For brevity, we only report the results of the computations. The detailed derivation as well as the discussion of further examples can be found in (Umezawa 1993, Umezawa *et al.* 1982).

In the present U(1) problem, the electromagnetic tensor and the vacuum current are (Umezawa 1993, Umezawa *et al.* 1982, Matsumoto *et al.* 1975a, b)

$$F_{\mu\nu}(x) = \partial_\mu a_\nu(x) - \partial_\nu a_\mu(x) = 2\pi \frac{m_V^2}{e_0} \int d^4x' \Delta_c(x-x') G^+_{\mu\nu}(x') \quad [49]$$

$$j_\mu(x) = -2\pi \frac{m_V^2}{e_0} \int d^4x' \Delta_c(x-x') \partial^\nu_{x'} G^+_{\nu\mu}(x') \quad [50]$$

respectively, and satisfy $\partial^\mu F_{\mu\nu}(x) = -j_\nu(x)$. In these equations,

$$\Delta_c(x-x') = \frac{1}{(2\pi)^4} \int d^4p\, e^{-ip(x-x')} \frac{1}{p^2 - m_V^2 + i\epsilon} \quad [51]$$

The line singularity for the vortex (or string) solution can be parametrized by a single line parameter $\sigma$ and by the time parameter $\tau$. A static vortex solution is obtained by setting $y_0(\tau,\sigma)=\tau$ and $\boldsymbol{y}(\tau,\sigma)=\boldsymbol{y}(\sigma)$, with $y$ denoting the line coordinate. $G^+_{\mu\nu}(x)$ is nonzero only on the line at $y$ (we can consider more lines but let us limit to only one line, for simplicity). Thus, we have

$$\begin{aligned} G_{0i}(x) &= \int d\sigma \frac{dy_i(\sigma)}{d\sigma} \delta^3[\boldsymbol{x} - \boldsymbol{y}(\sigma)] \quad G_{ij}(x) = 0 \\ G^+_{ij}(x) &= -\epsilon_{ijk} G_{0k}(x), \quad G^+_{0i}(x) = 0 \end{aligned} \quad [52]$$

Equation [49] shows that these vortices are purely magnetic. We obtain

$$\begin{aligned} \partial_0 f(x) &= 0 \\ \partial_i f(x) &= \frac{1}{(2\pi)^2} \int d\sigma \epsilon_{ijk} \frac{dy_k(\sigma)}{d\sigma} \partial^x_j \\ &\quad \times \int d^3p \frac{e^{i\boldsymbol{p}\cdot(\boldsymbol{x}-\boldsymbol{y}(\sigma))}}{\boldsymbol{p}^2} \end{aligned} \quad [53]$$

that is, by using the identity $(2\pi)^{-2}\int d^3p(e^{i\boldsymbol{p}\cdot\boldsymbol{x}}/\boldsymbol{p}^2) = 1/2|\boldsymbol{x}|$,

$$\nabla f(x) = -\frac{1}{2} \int d\sigma \frac{d\boldsymbol{y}_k(\sigma)}{d\sigma} \wedge \nabla_x \frac{1}{|\boldsymbol{x} - \boldsymbol{y}(\sigma)|} \quad [54]$$

Note that $\nabla^2 f(x) = 0$ is satisfied.

A straight infinitely long vortex is specified by $y_i(\sigma) = \sigma\delta_{i3}$ with $-\infty < \sigma < \infty$. The only nonvanishing component of $G^{\mu\nu}(x)$ are $G^{03}(x) = G^+_{12}(x) = \delta(x_1)\delta(x_2)$. Equation [54] gives (Umezawa 1993, Umezawa *et al.* 1982, Matsumoto 1975a, b)

$$\begin{aligned} \frac{\partial}{\partial x_1} f(x) &= \frac{1}{2} \int d\sigma \frac{\partial}{\partial x_2} [x_1^2 + x_2^2 + (x_3 - \sigma)^2]^{-1/2} \\ &= -\frac{x_2}{x_1^2 + x_2^2} \\ \frac{\partial}{\partial x_2} f(x) &= \frac{x_1}{x_1^2 + x_2^2}, \quad \frac{\partial}{\partial x_3} f(x) = 0 \end{aligned} \quad [55]$$

and then

$$f(x) = \tan^{-1}\left(\frac{x_2}{x_1}\right) = \theta(x) \quad [56]$$

We have thus determined the boson transformation function corresponding to a particular vortex solution. The vector potential is

$$\begin{aligned} a_1(x) &= -\frac{m_V^2}{2e_0} \int d^4x' \Delta_c(x-x') \frac{x_2'}{x_1'^2 + x_2'^2} \\ a_2(x) &= \frac{m_V^2}{2e_0} \int d^4x' \Delta_c(x-x') \frac{x_1'}{x_1'^2 + x_2'^2} \\ a_3(x) &= a_0(x) = 0 \end{aligned} \quad [57]$$

and the only nonvanishing component of $F_{\mu\nu}$:

$$\begin{aligned} F_{12}(x) &= -2\pi \frac{m_V^2}{e_0} \int d^4x' \Delta_c(x-x') \delta(x_1')\delta(x_2') \\ &= \frac{m_V^2}{e_0} K_0\left(m_V \sqrt{x_1^2 + x_2^2}\right) \end{aligned} \quad [58]$$

Finally, the vacuum current eqn [50] is given by

$$\begin{aligned} j_1(x) &= -\frac{m_V^3}{e_0} \frac{x_2}{\sqrt{x_1^2 + x_2^2}} K_1\left(m_V \sqrt{x_1^2 + x_2^2}\right) \\ j_2(x) &= \frac{m_V^3}{e_0} \frac{x_1}{\sqrt{x_1^2 + x_2^2}} K_1\left(m_V \sqrt{x_1^2 + x_2^2}\right) \\ j_3(x) &= j_0(x) = 0 \end{aligned} \quad [59]$$

We observe that these results are the same of the Nielsen–Olesen vortex solution. Notice that we did not specify the potential in our model but only the invariance properties. Thus, the invariance properties of the dynamics determine the characteristics of the topological solutions. The vortex solution

manifests the original U(1) symmetry through the cylindrical angle $\theta$ which is the parameter of the U(1) representation in the coordinate space.

## Conclusions

We have discussed how topological defects arise as inhomogeneous condensates in QFT. Topological defects are shown to have a genuine quantum nature. The approach reviewed here goes under the name of "boson transformation method" and relies on the existence of unitarily inequivalent representations of the field algebra in QFT.

Describing quantum fields with topological defects amounts then to properly choose the physical Fock space for representing the Heisenberg field operators. Once the boundary conditions corresponding to a particular soliton sector are found, the Heisenberg field operators embodied with such conditions contain the full information about the defects, the quanta and their mutual interaction. One can thus calculate Green's functions for particles in the presence of defects. The extension to finite temperature is discussed in Blasone and Jizba (2002) and Manka and Vitiello (1990).

As an example we have discussed a model with U(1) gauge invariance and SSB and we have obtained the Nielsen–Olesen vortex solution in terms of localized condensation of Goldstone bosons. These thus appear to play a physical role, although, in the presence of gauge fields, they do not show up in the physical spectrum as excitation quanta. The function $f(x)$ controlling the condensation of the NG bosons must be singular in order to produce observable effects. Boson transformations with regular $f(x)$ only amount to gauge transformations. For the treatment of topological defects in nonabelian gauge theories, see Manka and Vitiello (1990).

Finally, when there are no NG modes, as in the case of the kink solution or the sine-Gordon solution, the boson transformation function has to carry divergence singularity at spatial infinity (Umezawa 1993, Umezawa *et al.* 1982, Blasone and Jizba 2002). The boson transformation has also been discussed in connection with the Bäklund transformation at a classical level and the confinement of the constituent quanta in the coherent condensation domain.

For further reading on quantum fields with topological defects, see Blasone *et al.* (2006).

## Acknowledgments

The authors thank MIUR, INFN, INFM, and the ESF network COSLAB for partial financial support.

*See also:* Abelian Higgs Vortices; Algebraic Approach to Quantum Field Theory; Quantum Field Theory: A Brief Introduction; Quantum Field Theory in Curved Spacetime; Symmetries in Quantum Field Theory: Algebraic Aspects; Symmetries in Quantum Field Theory of Lower Spacetime Dimensions; Topological Defects and their Homotopy Classification.

## Further Reading

Anderson PW (1958) Coherent excited states in the theory of superconductivity: gauge invariance and the Meissner effect. *Physical Review* 110: 827–835.

Blasone M and Jizba P (2002) Topological defects as inhomogeneous condensates in quantum field theory: kinks in (1 + 1) dimensional $\lambda\psi^4$ theory. *Annals of Physics* 295: 230–260.

Blasone M, Jizba P, and Vitiello G (2006) *Spontaneous Breakdown of Symmetry and Topological Defects*, London: Imperial College Press. (in preparation).

Goldstone J (1961) Field theories with "superconductor" solutions. *Nuovo Cimento* 19: 154–164.

Goldstone J, Salam A, and Weinberg S (1962) Broken symmetries. *Physical Review* 127: 965–970.

Higgs P (1960) Spontaneous symmetry breakdown without massless bosons. *Physical Review* 145: 1156–1163.

Kibble TWB (1967) Symmetry breaking in non-abelian gauge theories. *Physical Review* 155: 1554–1561.

Kibble TWB (1976) Topology of cosmic domains and strings. *Journal of Physics A* 9: 1387–1398.

Kibble TWD (1980) Some implications of a cosmological phase transition. *Physics Reports* 67: 183–199.

Kleinert H (1989) *Gauge Fields in Condensed Matter*, vols. I & II Singapore: World Scientific.

Manka R and Vitiello G (1990) Topological solitons and temperature effects in gauge field theory. *Annals of Physics* 199: 61–83.

Matsumoto H, Papastamatiou NJ, Umezawa H, and Vitiello G (1975a) Dynamical rearrangement in Anderson–Higgs–Kibble mechanism. *Nuclear Physics B* 97: 61–89.

Matsumoto H, Papastamatiou NJ, and Umezawa H (1975b) The boson transformation and the vortex solutions. *Nuclear Physics B* 97: 90–124.

Nambu Y and Jona-Lasinio G (1961) Dynamical model of elementary particles based on an analogy with superconductivity. I. *Physical Review* 122: 345–358.

Nambu Y and Jona-Lasinio G (1961) Dynamical model of elementary particles based on an analogy with superconductivity. II. *Physical Review* 124: 246–254.

Rajaraman R (1982) *Solitons and Instantons: An Introduction to Solitons and Instantons in Quantum Field Theory*. Amsterdam: North-Holland.

Umezawa H (1993) *Advanced Field Theory: Micro, Macro and Thermal Physics*. New York: American Institute of Physics.

Umezawa H, Matsumoto H, and Tachiki M (1982) *Thermo Field Dynamics and Condensed States*. Amsterdam: North-Holland.

Vitiello G (2001) *My Double Unveiled*. Amsterdam: John Benjamins.

Volovik GE (2003) *The Universe in a Helium Droplet*. Oxford: Clarendon.

von Neumann J (1955) *Mathematical Foundation of Quantum Mechanics*. Princeton: Princeton University Press.

Zurek WH (1997) Cosmological experiments in condensed matter systems. *Physics Reports* 276: 177–221.

# Quantum Geometry and Its Applications

**A Ashtekar**, Pennsylvania State University, University Park, PA, USA
**J Lewandowski**, Uniwersyte Warszawski, Warsaw, Poland



## Introduction

In general relativity, the gravitational field is encoded in the Riemannian geometry of spacetime. Much of the conceptual compactness and mathematical elegance of the theory can be traced back to this central idea. The encoding is also directly responsible for the most dramatic ramifications of the theory: the big bang, black holes, and gravitational waves. However, it also leads one to the conclusion that spacetime itself must end and physics must come to a halt at the big bang and inside black holes, where the gravitational field becomes singular. But this reasoning ignores quantum physics entirely. When the curvature becomes large, of the order of $1/\ell_{\mathrm{Pl}}^2 = c^3/G\hbar$, quantum effects dominate and predictions of general relativity can no longer be trusted. In this "Planck regime," one must use an appropriate synthesis of general relativity and quantum physics, that is, a quantum gravity theory. The predictions of this theory are likely to be quite different from those of general relativity. In the real, quantum world, evolution may be completely nonsingular. Physics may not come to a halt and quantum theory could extend classical spacetime.

There are a number of different approaches to quantum gravity. One natural avenue is to retain the interplay between gravity and geometry but now use "quantum" Riemannian geometry in place of the standard, classical one. This is the key idea underlying loop quantum gravity. There are several calculations which indicate that the well-known failure of the standard perturbative approach to quantum gravity may be primarily due to its basic assumption that spacetime can be modeled as a smooth continuum at all scales. In loop quantum gravity, one adopts a nonperturbative approach. There is no smooth metric in the background. Geometry is not only dynamical but quantum mechanical from "birth." Its fundamental excitations turn out to be one dimensional and polymer-like. The smooth continuum is only a coarse-grained approximation. While a fully satisfactory quantum gravity theory still awaits us (in any approach), detailed investigations have been carried out to completion in simplified models – called mini- and midi-superspaces. They show that quantum spacetime does not end at singularities. Rather, quantum geometry serves as a "bridge" to another large classical spacetime.

This article will focus on structural issues from the perspective of mathematical physics. For complementary perspectives and further details, *see* Loop Quantum Gravity, Canonical General Relativity, Quantum Cosmology, Black Hole Mechanics, and Spin Foams in this Encyclopedia.

## Basic Framework

The starting point is a Hamiltonian formulation of general relativity based on spin connections (Ashtekar 1987). Here, the phase space $\mathbf{\Gamma}$ consists of canonically conjugate pairs $(A, \mathbb{P})$, where $A$ is a connection on a 3-manifold $M$ and $\mathbb{P}$ a 2-form, both of which take values in the Lie algebra su(2). Since $\mathbf{\Gamma}$ can also be thought of as the phase space of the SU(2) Yang–Mills theory, in this approach there is a unified kinematic framework for general relativity that describes gravity and the gauge theories which describe the other three basic forces of nature. The connection $A$ enables one to parallel transport chiral spinors (such as the left-handed fermions of the standard electroweak model) along curves in $M$. Its curvature is directly related to the electric and magnetic parts of the spacetime "Riemann tensor." The dual $P$ of $\mathbb{P}$ plays a double role (the dual is defined via $\int_M \mathbb{P} \wedge \omega = \int_M P \lrcorner \omega$ for any 1-form $\omega$ on $M$). Being the momentum canonically conjugate to $A$, it is analogous to the Yang–Mills electric field. But (apart from a constant), it is also an orthonormal triad (with density weight 1) on $M$ and therefore determines the positive-definite ("spatial") 3-metric, and hence the Riemannian geometry of $M$. This dual role of $P$ is a reflection of the fact that now SU(2) is the (double cover of the) group of rotations of the orthonormal spatial triads on $M$ itself rather than of rotations in an "internal" space associated with $M$.

To pass to quantum theory, one first constructs an algebra of "elementary" functions on $\mathbf{\Gamma}$ (analogous to the phase-space functions $x$ and $p$ in the case of a particle) which are to have unambiguous operator analogs. The holonomies

$$h_e(A) := \mathcal{P} \exp - \int_e A \qquad [1]$$

associated with a curve/edge $e$ on $M$ are (SU(2)-valued) configuration functions on $\mathbf{\Gamma}$. Similarly,

given a 2-surface $S$ on $M$, and an su(2)-valued (test) function $f$ on $M$,

$$P_{S,f} := \int_S \mathrm{tr}(f\,\mathbb{P}) \qquad [2]$$

is a momentum function on $\Gamma$, where tr is over the su(2) indices. (For simplicity of presentation, all fields are assumed to be smooth and curves/edges $e$ and surfaces $S$, finite and piecewise analytic in a specific sense. The extension to smooth curves and surfaces was carried out by Bacz and Sawin, Lewandowski and Thiemann, and Fleischhack. It is technically more involved but the final results are qualitatively the same.) The symplectic structure on $\Gamma$ enables one to calculate the Poisson brackets $\{h_e, P_{S,f}\}$. The result is a linear combination of holonomies and can be written as a Lie derivative,

$$\{h_e, P_{S,f}\} = \mathcal{L}_{X_{S,f}} h_e \qquad [3]$$

where $X_{S,f}$ is a derivation on the ring generated by holonomy functions, and can therefore be regarded as a vector field on the configuration space $\mathcal{A}$ of connections. This is a familiar situation in classical mechanics of systems whose configuration space is a finite-dimensional manifold. Functions $h_e$ and vector fields $X_{S,f}$ generate a Lie algebra. As in quantum mechanics on manifolds, the first step is to promote this algebra to a quantum algebra by demanding that the commutator be given by $\mathrm{i}\hbar$ times the Lie bracket. The result is a $\star$-algebra $\mathfrak{a}$, analogous to the algebra generated by operators $\exp \mathrm{i}\lambda\hat{x}$ and $\hat{p}$ in quantum mechanics. By exponentiating the momentum operators $\hat{P}_{S,f}$ one obtains $\mathfrak{W}$, the analog of the quantum-mechanical Weyl algebra generated by $\exp \mathrm{i}\lambda\hat{x}$ and $\exp \mathrm{i}\mu\hat{p}$.

The main task is to obtain the appropriate representation of these algebras. In that representation, quantum Riemannian geometry can be probed through the momentum operators $\hat{P}_{S,f}$, which stem from classical orthonormal triads. As in quantum mechanics on manifolds or simple field theories in flat space, it is convenient to divide the task into two parts. In the first, one focuses on the algebra $\mathfrak{C}$ generated by the configuration operators $\hat{h}_c$ and finds all its representations, and in the second one considers the momentum operators $\hat{P}_{S,f}$ to restrict the freedom.

$\mathfrak{C}$ is called the holonomy algebra. It is naturally endowed with the structure of an abelian $C^\star$ algebra (with identity), whence one can apply the powerful machinery made available by the Gel'fand theory. This theory tells us that $\mathfrak{C}$ determines a unique compact, Hausdorff space $\bar{\mathcal{A}}$ such that the $C^\star$ algebra of all continuous functions on $\bar{\mathcal{A}}$ is naturally isomorphic to $\mathfrak{C}$. $\bar{\mathcal{A}}$ is called the Gel'fand spectrum of $\mathfrak{C}$. It has been shown to consist of "generalized connections" $\bar{A}$ defined as follows: $\bar{A}$ assigns to any oriented edge $e$ in $M$ an element $\bar{A}(e)$ of SU(2) (a "holonomy") such that $\bar{A}(e^{-1}) = [\bar{A}(e)]^{-1}$; and, if the endpoint of $e_1$ is the starting point of $e_2$, then $\bar{A}(e_1 \circ e_2) = \bar{A}(e_1)\cdot \bar{A}(e_2)$. Clearly, every smooth connection $A$ is a generalized connection. In fact, the space $\mathcal{A}$ of smooth connections has been shown to be dense in $\bar{\mathcal{A}}$ (with respect to the natural Gel'fand topology thereon). But $\bar{\mathcal{A}}$ has many more "distributional elements." The Gel'fand theory guarantees that every representation of the $C^\star$ algebra $\mathfrak{C}$ is a direct sum of representations of the following type: the underlying Hilbert space is $\mathcal{H} = L^2(\bar{\mathcal{A}}, \mathrm{d}\mu)$ for some measure $\mu$ on $\bar{\mathcal{A}}$ and (regarded as functions on $\bar{\mathcal{A}}$) elements of $\mathfrak{C}$ act by multiplication. Since there are many inequivalent measures on $\bar{\mathcal{A}}$, there is a multitude of representations of $\mathfrak{C}$. A key question is how many of them can be extended to representations of the full algebra $\mathfrak{a}$ (or $\mathfrak{W}$) without having to introduce any "background fields" which would compromise diffeomorphism covariance. Quite surprisingly, the requirement that the representation be cyclic with respect to a state which is invariant under the action of the (appropriately defined) group Diff $M$ of piecewise-analytic diffeomorphisms on $M$ singles out a unique irreducible representation. This result was established for $\mathfrak{a}$ by Lewandowski, Okołów, Sahlmann and Thiemann, and for $\mathfrak{W}$ by Fleischhack. It is the quantum geometry analog to the seminal results by Segal and others that characterized the Fock vacuum in Minkowskian field theories. However, while that result assumes not only Poincaré invariance but also specific (namely free) dynamics, it is striking that the present uniqueness theorems make no such restriction on dynamics. The requirement of diffeomorphism invariance is surprisingly strong and makes the "background-independent" quantum geometry framework surprisingly tight.

This representation had been constructed by Ashtekar, Baez, and Lewandowski some ten years before its uniqueness was established. The underlying Hilbert space is given by $\mathcal{H} = L^2(\bar{\mathcal{A}}, \mathrm{d}\mu_o)$ where $\mu_o$ is a diffeomorphism-invariant, faithful, regular Borel measure on $\bar{\mathcal{A}}$, constructed from the normalized Haar measure on SU(2). Typical quantum states can be visualized as follows. Fix: (1) a graph $\alpha$ on $M$ (by a graph on $M$ we mean a set of a finite number of embedded, oriented intervals called edges; if two edges intersect, they do so only at one or both ends, called vertices), and (2) a smooth function $\psi$ on $[\mathrm{SU}(2)]^n$. Then, the function

$$\Psi_\gamma(\bar{A}) := \psi(\bar{A}(e_1), \ldots, \bar{A}(e_n)) \qquad [4]$$

on $\bar{\mathcal{A}}$ is an element of $\mathcal{H}$. Such states are said to be "cylindrical" with respect to the graph $\alpha$ and their space is denoted by $\mathrm{Cyl}_\alpha$. These are "typical states" in the sense that $\mathrm{Cyl} := \cup_\alpha \mathrm{Cyl}_\alpha$ is dense in $\mathcal{H}$. Finally, as ensured by the Gel'fand theory, the holonomy (or configuration) operators $\hat{h}_e$ act just by multiplication. The momentum operators $\hat{P}_{S,f}$ act as Lie derivatives: $\hat{P}_{S,f}\Psi = -\mathrm{i}\hbar\mathcal{L}_{X_{S,f}}\Psi$.

**Remark** Given any graph $\alpha$ in $M$, and a labeling of each of its edges by a nontrivial irreducible representation of SU(2) (i.e., by a nonzero half integer $j$), one can construct a finite-dimensional Hilbert space $\mathcal{H}_{\alpha,j}$, which can be thought of as the state space of a spin system "living on" the graph $\alpha$. The full Hilbert space admits a simple decomposition: $\mathcal{H} = \oplus_{\alpha,j}\mathcal{H}_{\alpha,j}$. This is called the spin-network decomposition. The geometric operators discussed in the next section leave each $\mathcal{H}_{\alpha,j}$ invariant. Therefore, the availability of this decomposition greatly simplifies the task of analyzing their properties.

## Geometric Operators

In the classical theory, $E := 8\pi G\gamma P$ has the interpretation of an orthonormal triad field (or a "moving frame") on $M$ (with density weight 1). Here, $\gamma$ is a dimensionless, strictly positive number, called the Barbero–Immirzi parameter, which arises as follows. Because of emphasis on connections, in the classical theory the first-order Palatini action is a more natural starting point than the second-order Einstein–Hilbert action. Now, there is a freedom to add a term to the Palatini action which vanishes when Bianchi identities are satisfied and therefore does not change the equations of motion. $\gamma$ arises as the coefficient of this term. In some respects $\gamma$ is analogous to the $\theta$ parameter of Yang–Mills theory. Indeed, while theories corresponding to any permissible values of $\gamma$ are related by a canonical transformation classically, quantum mechanically this transformation is not unitarily implementable. Therefore, although there is a unique representation of the algebra $\mathfrak{a}$ (or $\mathfrak{W}$), there is a one-parameter family of inequivalent representations of the algebra of geometric operators generated by suitable functions of orthonormal triads $E$, each labeled by the value of $\gamma$. This is a genuine quantization ambiguity. As with the $\theta$ ambiguity in QCD, the actual value of $\gamma$ in nature has to be determined experimentally. The current strategy in quantum geometry is to fix its value through a thought experiment involving black hole thermodynamics (see below).

The basic object in quantum Riemannian geometry is the triad flux operator $\hat{E}_{S,f} := 8\pi G\gamma\hat{P}_{S,f}$. It is self-adjoint and all its eigenvalues are discrete. To define other geometric operators such as the area operator $\hat{A}_S$ associated with a surface $S$ or a volume operator $\hat{V}_R$ associated with a region $R$, one first expresses the corresponding phase-space functions in terms of the "elementary" functions $E_{S_i,f_i}$ using suitable surfaces $S_i$ and test functions $f_i$ and then promotes $E_{S_i,f_i}$ to operators. Even though the classical expressions are typically nonpolynomial functions of $E_{S_i,f_i}$, the final operators are all well defined, self-adjoint and with purely discrete eigenvalues. Therefore, in the sense of the word used in elementary quantum mechanics (e.g., of the hydrogen atom), one says that geometry is quantized. Because the theory has no background metric or indeed any other background field, all geometric operators transform covariantly under the action of the Diff $M$. This diffeomorphism covariance makes the final expressions of operators rather simple. In the case of the area operator, for example, the action of $\hat{A}_S$ on a state $\Psi_\alpha$ [4] depends entirely on the points of intersection of the surface $S$ and the graph $\alpha$ and involves only right- and left-invariant vector fields on copies of SU(2) associated with edges of $\alpha$ which intersect $S$. In the case of the volume operator $\hat{V}_R$, the action depends on the vertices of $\alpha$ contained in $R$ and, at each vertex, involves the right- and left-invariant vector fields on copies of SU(2) associated with edges that meet at each vertex.

To display the explicit expressions of these operators, let us first define on $\mathrm{Cyl}_\alpha$ three basic operators $\hat{J}_j^{(v,e)}$, with $j \in \{1,2,3\}$, associated with the pair consisting of an edge $e$ of $\alpha$ and a vertex $v$ of $e$:

$$\hat{J}_j^{(v,e)}\Psi_\alpha(\bar{A}) = \begin{cases} \mathrm{i}\dfrac{\mathrm{d}}{\mathrm{d}t}|_{t=0}\psi_\alpha(\ldots, U_e(\bar{A})\exp(t\tau_j),\ldots) \\ \quad \text{if } e \text{ begins at } v \\ \mathrm{i}\dfrac{\mathrm{d}}{\mathrm{d}t}|_{t=0}\psi_\alpha(\ldots,\exp(-t\tau_j)U_e(\bar{A}),\ldots) \\ \quad \text{if } e \text{ ends at } v \end{cases} \qquad [5]$$

where $\tau_j$ denotes a basis in su(2) and "..." stands for the rest of the arguments of $\Psi_\alpha$ which remain unaffected. The quantum area operator $A_s$ is assigned to a finite two-dimensional submanifold $S$ in $M$. Given a cylindrical state we can always represent it in the form [4] using a graph $\alpha$ adapted to $S$, such that every edge $e$ either intersects $S$ at exactly one endpoint, or is contained in the closure $\bar{S}$, or does not intersect $\bar{S}$. For each vertex $v$ in $S$ of the graph $\alpha$, the family of edges intersecting $v$ can be divided into three classes: edges $\{e_1,\ldots,e_u\}$ lying on one side (say "above") $S$, edges $\{e_{u+1},\ldots,e_{u+d}\}$ lying

on the other side (say "below"), and edges contained in $S$. To each $v$ we assign a generalized Laplace operator

$$\Delta_{S,v} = -\eta^{ij}\left(\sum_{I=1}^{u}\hat{J}_i^{(v,e_I)} - \sum_{I=u+1}^{u+d}\hat{J}_i^{(v,e_I)}\right) \times \left(\sum_{K=1}^{u}\hat{J}_j^{(v,e_K)} - \sum_{K=u+1}^{u+d}\hat{J}_j^{(v,e_K)}\right) \quad [6]$$

where $\eta_{ij}$ stands for –1/2 the Killing form on su(2). Now, the action of the quantum area operator $\hat{A}_S$ on $\Psi_\alpha$ is defined as follows:

$$\hat{A}_S\Psi_\alpha = 4\pi\gamma\ell_{\mathrm{Pl}}^2\sum_{v\in S}\sqrt{-\Delta_{S,v}}\Psi_\alpha \quad [7]$$

The quantum area operator has played the most important role in applications. Its complete spectrum is known in a closed form. Consider arbitrary sets $j_I^{(u)}, j_I^{(d)}$, and $j_I^{(u+d)}$ of half-integers, subject to the condition

$$j_I^{(u+d)} \in \{|j_I^{(u)} - j_I^{(d)}|, |j_I^{(u)} - j_I^{(d)}| + 1, \ldots, j_I^{(u)} + j_I^{(d)}\} \quad [8]$$

where $I$ runs over any finite number of integers. The general eigenvalues of the area operator are given by:

$$a_S = 4\pi\gamma\ell_{\mathrm{Pl}}^2\sum_I\Big(2j_I^{(u)}(j_I^{(u)}+1) + 2j_I^{(d)}(j_I^{(d)}+1) - j_I^{(u+d)}(j_I^{(u+d)}+1)\Big)^{1/2} \quad [9]$$

On the physically interesting sector of SU(2)-gauge-invariant subspace $\mathcal{H}_{\mathrm{inv}}$ of $\mathcal{H}$, the lowest eigenvalue of $\hat{A}_S$ – "the area gap" – depends on some global properties of $S$. Specifically, it "knows" whether the surface is open, or a 2-sphere, or, if $M$ is a 3-torus, a (nontrivial) 2-torus in $M$. Finally, on $\mathcal{H}_{\mathrm{inv}}$, one is often interested only in the subspace of states $\Psi_\alpha$, where $\alpha$ has no edges which lie within a given surface $S$. Then, the expression of eigenvalues simplifies considerably:

$$a_S = 8\pi\gamma\ell_{\mathrm{Pl}}^2\sum_I\sqrt{j_I(j_I+1)} \quad [10]$$

To display the action of the quantum volume operator $\hat{V}_R$, for each vertex $v$ of a given graph $\alpha$, let us first define an operator $\hat{q}_v$ on $\mathrm{Cyl}_\alpha$.

$$\hat{q}_v = (8\pi\gamma\ell_{\mathrm{Pl}}^2)^3\frac{1}{48} \times \sum_{e,e',e''}\epsilon(e,e',e'')c^{ijk}\hat{J}_i^{(v,e)}\hat{J}_j^{(v,e')}\hat{J}_k^{(v,e'')} \quad [11]$$

where $e$, $e'$, and $e''$ run over the set of edges intersecting $v$, $\epsilon(e,e',e'')$ takes values $\pm1$ or 0 depending on the orientation of the half-lines tangent to the edges at $v$, $[\tau_i,\tau_j] = c^k{}_{ij}\tau_k$ and the indices are raised by the tensor $\eta_{ij}$. The action of the quantum volume operator on a cylindrical state [4] is then given by

$$\hat{V}_R\Psi_\alpha = \kappa_o\sum_{v\in R}\sqrt{|\hat{q}_v|}.\Psi_\alpha \quad [12]$$

Here, $\kappa_o$ is an overall, independent of a graph, constant resulting from an averaging.

The volume operator plays an unexpectedly important role in the definition of both the gravitational and matter contributions to the scalar constraint operator which dictates dynamics. Finally, a notable property of the volume operator is the following. Let $R(p,\epsilon)$ be a family of neighborhoods of a point $p \in M$. Then, as indicated above, $\hat{V}_{R(p,\epsilon)}\Psi_\alpha = 0$ if $\alpha$ has no vertex in the neighborhood. However, if $\alpha$ has a vertex at $p$

$$\lim_{\epsilon\to0}\hat{V}_{R(x,\epsilon)}\Psi_\alpha$$

exists but is not necessarily zero. This is a reflection of the "distributional" nature of quantum geometry.

**Remark** States $\Psi_\alpha \in \mathrm{Cyl}$ have support only on the graph $\alpha$. In particular, they are simply annihilated by geometric operators such as $\hat{A}_S$ and $\hat{V}_R$ if the support of the surface $S$ and the region $R$ does not intersect the support of $\alpha$. In this sense the fundamental excitations of geometry are one dimensional and geometry is polymer-like. States $\Psi_\alpha$, where $\alpha$ is just a "small graph," are highly quantum mechanical – like states in QED representing just a few photons. Just as coherent states in QED require an infinite superposition of such highly quantum states, to obtain a semiclassical state approximating a given classical geometry, one has to superpose a very large number of such elementary states. More precisely, in the Gel'fand triplet $\mathrm{Cyl} \subset \mathcal{H} \subset \mathrm{Cyl}^\star$, semiclassical states belong to the dual $\mathrm{Cyl}^\star$ of Cyl.

## Applications

Since quantum Riemannian geometry underlies loop quantum gravity and spin-foam models, all results obtained in these frameworks can be regarded as its applications. Among these, there are two which have led to resolutions of long-standing issues. The first concerns black hole entropy, and the second, quantum nature of the big bang.

### Black Holes

Seminal advances in fundamentals of black hole physics in the mid-1970s suggested that the entropy of large black holes is given by $S_{\mathrm{BH}} = (a_{\mathrm{hor}}/4\ell_{\mathrm{Pl}}^2)$,

where $a_{\rm hor}$ is the horizon area. This immediately raised a challenge to potential quantum gravity theories: give a statistical mechanical derivation of this relation. For familiar thermodynamic systems, a statistical mechanical derivation begins with an identification the microscopic degrees of freedom. For a classical gas, these are carried by molecules; for the black body radiation, by photons; and for a ferromagnet, by Heisenberg spins. What about black holes? The microscopic building blocks cannot be gravitons because the discussion involves stationary black holes. Furthermore, the number of microscopic states is absolutely huge: some $\exp 10^{77}$ for a solar mass black hole, a number that completely dwarfs the number of states of systems one normally encounters in statistical mechanics. Where does this huge number come from? In loop quantum gravity, this is the number of states of the "quantum horizon geometry."

The idea behind the calculation can be heuristically explained using the "It from Bit" argument, put forward by Wheeler in the 1990s. Divide the black hole horizon into elementary cells, each with one Planck unit of area, $\ell_{\rm Pl}^2$, and assign to each cell two microstates. Then the total number of states $\mathcal{N}$ is given by $\mathcal{N} = 2^n$, where $n = (a_{\rm hor}/\ell_{\rm Pl}^2)$ is the number of elementary cells, whence entropy is given by $S = \ln \mathcal{N} \sim a_{\rm hor}$. Thus, apart from a numerical coefficient, the entropy (It) is accounted for by assigning two states (Bit) to each elementary cell. This qualitative picture is simple and attractive. However, the detailed derivation in quantum geometry has several new features.

First, Wheeler's argument would apply to any 2-surface, while in quantum geometry the surface must represent a horizon in equilibrium. This requirement is encoded in a certain boundary condition that the canonically conjugate pair $(A, \mathbb{P})$ must satisfy at the surface and plays a crucial role in the quantum theory. Second, the area of each elementary cell is not a fixed multiple of $\ell_{\rm Pl}^2$ but is given by [10], where $I$ labels the elementary cells and $j_I$ can be any half-integer (such that the sum is within a small neighborhood of the classical area of the black hole under consideration). Finally, the number of quantum states associated with an elementary cell labeled by $j_I$ is not 2 but $(2j_I + 1)$.

The detailed theory of the quantum horizon geometry and the standard statistical mechanical reasoning is then used to calculate the entropy and the temperature. For large black holes, the leading contribution to entropy is proportional to the horizon area, in agreement with quantum field theory in curved spacetimes. (The subleading term $-(1/2)\ln(a_{\rm hor}/\ell_{\rm Pl}^2)$ is a quantum gravity correction to Hawking's semiclassical result. This correction, with the –1/2 factor, is robust in the sense that it also arises in other approaches.) However, as one would expect, the proportionality factor depends on the Barbero–Immirzi parameter $\gamma$ and so far loop quantum gravity does not have an independent way to determine its value. The current strategy is to determine $\gamma$ by requiring that, for the Schwarzschild black hole, the leading term agrees exactly with Hawking's semiclassical answer. This requirement implies that $\gamma$ is the root of algebraic equation and its value is given by $\gamma \approx 0.2735$. Now, quantum geometry theory is completely fixed. One can calculate entropy of other black holes, with angular momentum and distortion. A nontrivial check on the strategy is that for all these cases, the coefficient in the leading-order term again agrees with Hawking's semiclassical result.

The detailed analysis involves a number of structures of interest to mathematical physics. First, the intrinsic horizon geometry is described by a U(1) Chern–Simons theory on a punctured 2-sphere (the horizon), the level $k$ of the theory being given by $k = a_{\rm hor}/4\pi\gamma\ell_{\rm Pl}^2$. The punctures are simply the intersections of the excitations of the polymer geometry in the bulk with the horizon 2-surface. Second, because of the horizon boundary conditions, in the classical theory the gauge group SU(2) is reduced to U(1) at the horizon. At each puncture, it is further reduced to the discrete subgroup $\mathbb{Z}_k$ of U(1), sometimes referred to as a "quantum U(1) group." Third, the "surface phase space" associated with the horizon is represented by a noncommutative torus. Finally, the surface Chern–Simons theory is entirely unrelated to the bulk quantum geometry theory but the quantum horizon boundary condition requires that the spectrum of a certain operator in the Chern–Simons theory must be identical to that of another operator in the bulk theory. The surprising fact is that there is an exact agreement. Without this seamless matching, a coherent description of the quantum horizon geometry would not have been possible.

The main weakness of this approach to black hole entropy stems from the Barbero–Immirzi ambiguity. The argument would be much more compelling if the value of $\gamma$ were determined by independent considerations, without reference to black hole entropy. (By contrast, for extremal black holes, string theory provides the correct coefficient without any adjustable parameter. The AdS/CFT duality hypothesis (as well as other semiquantitative) arguments have been used to encompass certain black holes which are away from extremality. But in these cases, it is not known if the numerical coefficient is

1/4 as in Hawking's analysis.) It's primary strengths are twofold. First, the calculation encompasses all realistic black holes – not just extremal or near-extremal – including the astrophysical ones, which may be highly distorted. Hairy black holes of mathematical physics and cosmological horizons are also encompassed. Second, in contrast to other approaches, one works directly with the physical, curved geometry around black holes rather than with a flat-space system which has the same number of states as the black hole of interest.

### The Big Bang

Most of the work in physical cosmology is carried out using spatially homogeneous and isotropic models and perturbations thereon. Therefore, to explore the quantum nature of the big bang, it is natural to begin by assuming these symmetries. Then the spacetime metric is determined simply by the scale factor $a(t)$ and matter fields $\phi(t)$ which depend only on time. Thus, because of symmetries, one is left with only a finite number of degrees of freedom. Therefore, field-theoretic difficulties are bypassed and passage to quantum theory is simplified. This strategy was introduced already in the late 1960s and early 1970s by DeWitt and Misner. Quantum Einstein's equations now reduce to a single differential equation of the type

$$\frac{\partial^2}{\partial a^2}(f(a)\Psi(a,\phi)) = \text{const.}\, \hat{H}_\phi\, \Psi(a,\phi) \qquad [13]$$

on the wave function $\Psi(a,\phi)$, where $\hat{H}_\phi$ is the matter Hamiltonian and $f(a)$ reflects the freedom in factor ordering. Since the scale factor $a$ vanishes at the big bang, one has to analyze the equation and its solutions near $a = 0$. Unfortunately, because of the standard form of the matter Hamiltonian, coefficients in the equation diverge at $a = 0$ and the evolution cannot be continued across the singularity unless one introduces unphysical matter or a new principle. A well-known example of new input is the Hartle–Hawking boundary condition which posits that the universe starts out without any boundary and a metric with positive-definite signature and later makes a transition to a Lorentzian metric.

Bojowald and others have shown that the situation is quite different in loop quantum cosmology because quantum geometry effects make a qualitative difference near the big bang. As in older quantum cosmologies, one carries out a symmetry reduction at the classical level. The final result differs from older theories only in minor ways. In the homogeneous, isotropic case, the freedom in the choice of the connection is encoded in a single function $c(t)$ and, in that of the momentum/triad, in another function $p(t)$. The scale factor is given by $a^2 = |p|$. (The variable $p$ itself can assume both signs; positive if the triad is left handed and negative if it is right handed. $p$ vanishes at degenerate triads which are permissible in this approach.) The system again has only a finite number of degrees of freedom. However, quantum theory turns out to be inequivalent to that used in older quantum cosmologies.

This surprising result comes about as follows. Recall that in quantum geometry, one has well-defined holonomy operators $\hat{h}$ but there is no operator corresponding to the connection itself. In quantum mechanics, the analog would be for operators $\hat{U}(\lambda)$ corresponding to the classical functions $\exp i\lambda x$ to exist but not be weakly continuous in $\lambda$; the operator $\hat{x}$ would then not exist. Once the requirement of weak continuity is dropped, von Neumann's uniqueness theorem no longer holds and the Weyl algebra can have inequivalent irreducible representations. The one used in loop quantum cosmology is the direct analog of full quantum geometry. While the space $\mathcal{A}$ of smooth connections reduces just to the real line $\mathbb{R}$, the space $\bar{\mathcal{A}}$ of generalized connections reduces to the Bohr compactification $\bar{\mathbb{R}}_{\text{Bohr}}$ of the real line. (This space was introduced by the mathematician Harold Bohr (Nils' brother) in his theory of almost-periodic functions. It arises in the present application because holonomies turn out to be almost periodic functions of $c$.) The Hilbert space of states is thus $\mathcal{H} = L^2(\bar{\mathbb{R}}_{\text{Bohr}}, \mathrm{d}\mu_o)$ where $\mu_o$ is the Haar measure on (the abelian group) $\bar{\mathbb{R}}_{\text{Bohr}}$. As in full quantum geometry, the holonomies act by multiplication and the triad/momentum operator $\hat{p}$ via Lie derivatives.

To facilitate comparison with older quantum cosmologies, it is convenient to use a representation in which $\hat{p}$ is diagonal. Then, quantum states are functions $\Psi(p,\phi)$. But the Wheeler–DeWitt equation is now replaced by a difference equation:

$$C^+(p)\,\Psi(p+4p_o,\phi) + C^o(p)\,\Psi(p,\phi) + C^-(p)\,\Psi(p-4p_o)(\phi) = \text{const.}\, \hat{H}_\phi\Psi(p,\phi) \qquad [14]$$

where $p_o$ is determined by the lowest eigenvalue of the area operator ("area gap") and the coefficients $C^\pm(p)$ and $C^o(p)$ are functions of $p$. In a backward "evolution," given $\Psi$ at $p+4$ and $p$, such a "recursion relation" determines $\Psi$ at $p-4$, provided $C^-$ does not vanish at $p-4$. The coefficients are well behaved and nowhere vanishing, whence the evolution does not stop at any finite $p$, either in the past or in the future. Thus, near $p = 0$ this equation is drastically different from the Wheeler–DeWitt equation [13]. However, for large $p$ – that is, when the universe is large – it is well

approximated by [13] and smooth solutions of [13] are approximate solutions of the fundamental discrete equation [14] in a precise sense.

To complete quantization, one has to introduce a suitable Hilbert space structure on the space of solutions to [14], identify physically interesting operators and analyze their properties. For simple matter fields, this program has been completed. With this machinery at hand, one begins with semiclassical states which are peaked at configurations approximating the classical universe at late times (e.g., now) and evolves backwards. Numerical simulations show that the state remains peaked at the classical solution till very early times when the matter density becomes of the order of Planck density. This provides, in particular, a justification, from first principles, for the assumption that spacetime can be taken to be classical even at the onset of the inflationary era, just a few Planck times after the (classical) big bang. While one would expect a result along these lines to hold on physical grounds, technically it is nontrivial to obtain semiclassicality over such huge domains. However, in the Planck regime near the big bang, there are major deviations from the classical behavior. Effectively, gravity becomes repulsive, the collapse is halted and then the universe re-expands. Thus, rather than modifying spacetime structure just in a tiny region near the singularity, quantum geometry effects open a bridge to another large classical universe. These are dramatic modifications of the classical theory.

For over three decades, hopes have been expressed that quantum gravity would provide new insights into the true nature of the big bang. Thanks to quantum geometry effects, these hopes have been realized and many of the long-standing questions have been answered. While the final picture has some similarities with other approaches, (e.g., "cyclic universes," or pre-big-bang cosmology), only in loop quantum cosmology is there a fully deterministic evolution across what was the classical big-bang. However, so far, detailed results have been obtained only in simple models. The major open issue is the inclusion of perturbations and subsequent comparison with observations.

*See also:* Algebraic Approach to Quantum Field Theory; Black Hole Mechanics; Canonical General Relativity; Knot Invariants and Quantum Gravity; Loop Quantum Gravity; Quantum Cosmology; Quantum Dynamics in Loop Quantum Gravity; Quantum Fields Theory in Curved Spacetime; Spacetime Topology, Causal Structure and Singularities; Spin Foams; Wheeler–De Witt Theory.

## Further Reading


Ashtekar A (1987) New Hamiltonian formulation of general relativity. *Physical Review D* 36: 1587–1602.

Ashtekar A and Krishnan B (2004) Isolated and dynamical horizons and their applications. *Living Reviews in Relativity* 10: 1–78 (gr-qc/0407042).

Ashtekar A and Lewandowski L (2004) Background independent quantum gravity: a status report. *Classical Quantum Gravity* 21: R53–R152.

Bojowald M and Morales-Tecotl HA (2004) Cosmological applications of loop quantum gravity. *Lecture Notes in Physics* 646: 421–462 (also available at gr-qc/0306008).

Gambini R and Pullin J (1996) *Loops, Knots, Gauge Theories and Quantum Gravity*. Cambridge: Cambridge University Press.

Perez A (2003) Spin foam models for quantum gravity. *Classical Quantum Gravity* 20: R43–R104.

Rovelli C (2004) *Quantum Gravity*. Cambridge: Cambridge University Press.

Thiemann T (2005) *Introduction to Modern Canonical Quantum General Relativity*, Cambridge: Cambridge University Press. (draft available as gr-qc/0110034).


# Quantum Group Differentials, Bundles and Gauge Theory


**T Brzeziński**, University of Wales Swansea, Swansea, UK




## Introduction

Mathematics of classical gauge theories is contained in the theory of principal and associated vector bundles. Principal bundles describe pure gauge fields and their transformations, while the associated bundles contain matter fields. A structure group of a bundle has a meaning of a gauge group, while the base manifold is a spacetime for the theory. In this article, we review the theory of bundles in which a structure group is a quantum group and base space or spacetime might be noncommutative. To fully deal with geometric aspects, we first review differential geometry of quantum groups. Then we describe the theory of quantum principal bundles, connections on such bundles, gauge transformations, associated vector bundles and their sections. We indicate that, for a certain class of quantum principal bundles, sections of an associated bundle become vector bundles of noncommutative geometry *à la* Connes, that is,

finite projective modules. The theory is illustrated by two explicit examples that can be viewed as deformations of the classical magnetic monopole and the instanton.

## Differential Structures on Algebras

### Algebraic Conventions

Throughout this article, $A$ ($P$ etc.,) will be an associative unital complex algebra. To gain some geometric intuition the reader can think of $A$ as an algebra of continuous complex functions on a compact (Hausdorff) space $X$, $C(X)$, with product given by pointwise multiplication $fg(x) = f(x)g(x)$, and with the unit provided by a constant function $x \mapsto 1$. The algebra $C(X)$ is commutative, but, in what follows, we do not assume that $A$ is a commutative algebra. By an $A$-bimodule we mean a vector space with mutually commuting left and right actions of $A$. All modules are unital (i.e., the unit element of $A$ acts trivially). On elements, the multiplication in an algebra or an action of $A$ on a module is denoted by juxtaposition.

### Differential Calculus on an Algebra

A first-order differential calculus on $A$ is a pair $(\Omega^1(A), \mathrm{d})$, where $\Omega^1(A)$ is an $A$-bimodule and $\mathrm{d}: A \to \Omega^1(A)$ is a linear map such that:

1. for all $a, b \in A$, $\mathrm{d}(ab) = (\mathrm{d}a)b + a\mathrm{d}b$ (the Leibniz rule); and
2. every $\omega \in \Omega^1(A)$ can be written as $\omega = \sum_i a_i \mathrm{d}b_i$ for some $a_i, b_i \in A$.

Elements of $\Omega^1(A)$ are called differential 1-forms and the map d is called an exterior derivative. As a motivating example, take $A = C(X)$ and $\Omega^1(A)$ the space of 1-forms on $X$ (sections of the cotangent bundle $T^*X$), and d the usual exterior differential.
Higher-differential forms corresponding to $(\Omega^1(A), \mathrm{d})$ are defined as elements of a differential graded algebra $\Omega(A)$. This is an algebra which can be decomposed into the direct sum of $A$-bimodules $\Omega^n(A)$, that is, $\Omega(A) = A \oplus \Omega^1(A) \oplus \Omega^2(A) \oplus \cdots$. In addition to $\mathrm{d}: A \to \Omega^1(A)$, there are maps $\mathrm{d}_n: \Omega^n(A) \to \Omega^{n+1}(A)$ such that, for all $\omega_n \in \Omega^n(A)$, $\omega_k \in \Omega^k(A)$,

1. $\mathrm{d}_1 \circ \mathrm{d} = 0$ and $\mathrm{d}_{n+1} \circ \mathrm{d}_n = 0, n = 1, 2, \ldots$;
2. $\omega_n \omega_k \in \Omega^{n+k}(A)$; and
3. $\mathrm{d}_{n+k}(\omega_n \omega_k) = (\mathrm{d}_n \omega_n)\omega_k + (-1)^n \omega_n(\mathrm{d}_k \omega_k)$.

Elements of $\Omega^n(A)$ are known as "differential $n$-forms." $\Omega^n(A)$ contains all linear combinations of expressions $a_0\, \mathrm{d}a_1\, \mathrm{d}a_2 \cdots \mathrm{d}a_n$ with $a_0, \ldots, a_n \in A$. One says that $\Omega(A)$ satisfies the "density condition" if any element of $\Omega^n(A)$ is of the above form, for any $n$. To simplify notation, one writes d for $\mathrm{d}_n$.

As an example of $\Omega(A)$, take $A = C(X)$ and then the exterior algebra $\Omega(X)$ for $\Omega(A)$. The exterior algebra satisfies density condition as any $n$-form can be written as $f(x) \wedge \mathrm{d}g(x) \wedge \mathrm{d}h(x) \wedge \cdots$. The wedge product is anticommutative, but for a noncommutative algebra $A$, the anticommutativity of the product in $\Omega(A)$ cannot be generally required.

### The Universal Differential Calculus

Any algebra $A$ comes equipped with a universal differential calculus denoted by $(\Omega^1 A, \mathrm{d})$. $\Omega^1 A$ is defined as the kernel of the multiplication map, that is, $\Omega^1 A := \{\sum_i a_i \otimes b_i \in A \otimes A \mid \sum_i a_i b_i = 0\} \subseteq A \otimes A$. The derivative is defined by $\mathrm{d}(a) = 1 \otimes a - a \otimes 1$. The $n$-forms are defined as $\Omega^n A = \Omega^1 A \otimes_A \Omega^1 A \otimes_A \cdots \otimes_A \Omega^1 A$ ($n$-copies of $\Omega^1 A$). $\Omega^n A$ can be identified with a subspace of $A \otimes A \otimes \cdots \otimes A$ ($n+1$-copies of $A$) consisting of all such elements that vanish upon multiplication of any two consecutive factors. With this identification, higher derivatives read

$$\mathrm{d}\Big(\sum_i a_0^i \otimes a_1^i \otimes \cdots \otimes a_n^i\Big) = \sum_{k=0}^{n+1} \sum_i (-1)^k a_0^i \otimes a_1^i \otimes \cdots \otimes a_{k-1}^i \otimes 1 \otimes a_k^i \cdots \otimes a_n^i$$

The universal differential calculus satisfies the density condition.

This calculus captures very little (if any) of the geometry of the underlying algebra $A$, but it has the universality property, that is, any differential calculus on $A$ can be obtained as a quotient of $\Omega A$. In other words, any differential calculus $\Omega(A)$ is fully determined by a system of $A$-sub-bimodules $N_n \in A^{\otimes n+1}$ (or homogeneous ideals in the algebra $\Omega A$), so that $\Omega^n(A) = \Omega^n A / N_n$. The differentials d in $\Omega(A)$ are derived from universal differentials via the canonical projections $\pi_n: \Omega^n A \to \Omega^n(A)$.

Typical examples of algebras in quantum geometry are given by generators and relations, that is, $A = \mathbb{C}\langle x_1, \ldots, x_n\rangle / \langle R_i(x_1, \ldots, x_n)\rangle$, where $\mathbb{C}\langle x_1, \ldots, x_n\rangle$ is a free algebra on generators $x_k$ and $R_i(x_1, \ldots, x_n)$ are polynomials, so that $R_i(x_1, \ldots, x_n) = 0$ in $A$. Correspondingly, the modules $\Omega^n(A)$ are given by generators and relations. If $\Omega(A)$ satisfies the density condition, that the whole of $\Omega(A)$ must be generated by some 1-forms. The sub-bimodules $N_n$ contain relations satisfied by these generators.

### $*$-Calculi

If $A$ is a $*$-algebra, then a calculus is called a "$*$-calculus" provided $\Omega(A)$ is a graded $*$-algebra, and $\mathrm{d}(\rho^*) = (\mathrm{d}\rho)^*$, for all $\rho \in \Omega(A)$.

## Differential Structures on Quantum Groups (Hopf Algebras)

### Hopf Algebra Preliminaries

From now on, $A$ is a Hopf algebra (quantum group), with a coproduct $\Delta : A \to A \otimes A$, counit $\varepsilon : A \to \mathbb{C}$ and antipode $S$. We use Sweedler's notation $\Delta(a) = \sum a_{(1)} \otimes a_{(2)}$. We also write $A^+ = \ker \varepsilon$ (the augmentation ideal).

For any algebra $P$, the convolution product of linear maps $f, g : A \to P$ is a linear map $f * g : A \to P$, defined by $f * g(a) = \sum f(a_{(1)}) * g(a_{(2)})$. A map $f : A \to P$ is said to be convolution invertible, provided there exists $f^{-1} : A \to P$ such that $f * f^{-1} = f * f^{-1} = 1\varepsilon$.

An $A$-coaction on a comodule $V$, $\varrho : V \to V \otimes A$, is denoted by $\varrho(v) = \sum v_{(0)} \otimes v_{(1)}$. The right adjoint coaction in $A$ is a map

$$\mathrm{Ad} : A \to A \otimes A,$$
$$\mathrm{Ad}(a) = \sum a_{(2)} \otimes (Sa_{(1)})a_{(3)}$$

A subspace $B$ of $A$ is said to be "Ad-invariant" provided $\mathrm{Ad}(B) \subseteq B \otimes A$. For example, $A^+$ is such a space.

### Covariant Differential Calculi

For Hopf algebras one can study calculi that are covariant with respect to $\Delta$. For $A = \mathbb{C}[G]$ (an algebra of functions on a Lie group), this corresponds to the covariance of a differential structure on $G$ with respect to regular representations.

A first-order differential calculus $\Omega^1(A)$ on a quantum group $A$ is said to be left-covariant, if there exists a linear map $\Delta_{\mathrm{L}} : \Omega^1(A) \to A \otimes \Omega^1(A)$ (called a left coaction) such that, for all $a, b \in A$,

$$\Delta_{\mathrm{L}}(a\mathrm{d}b) = \sum a_{(1)}b_{(1)} \otimes a_{(2)}\mathrm{d}b_{(2)}$$

$\Omega^1(A)$ is called a right-covariant differential calculus if there exists a linear map $\Delta_{\mathrm{R}} : \Omega^1(A) \to \Omega^1(A) \otimes A$ (called a right coaction) such that, for all $a, b \in A$,

$$\Delta_{\mathrm{R}}(a\mathrm{d}b) = \sum a_{(1)}\mathrm{d}b_{(1)} \otimes a_{(2)}b_{(2)}$$

If $\Omega^1(A)$ is both left- and right-covariant, it is called a "bicovariant differential calculus." A bicovariant $\Omega^1(A)$ has a structure of a Hopf $A$-bimodule, that is, it is an $A$-bimodule and an $A$-bicomodule such that the coactions are compatible with actions.

The universal calculus on $A$ is bicovariant with coactions

$$\Delta_{\mathrm{R}}^{U}\Big(\sum_i a^i \otimes b^i\Big) = \sum_i a^i_{(1)} \otimes b^i_{(1)} \otimes a^i_{(2)}b^i_{(2)},$$
$$\Delta_{\mathrm{L}}^{U}\Big(\sum_i a^i \otimes b^i\Big) = \sum_i a^i_{(1)}b^i_{(1)} \otimes a^i_{(2)} \otimes b^i_{(2)}$$

Since $\Omega^1(A) = \Omega^1 A/N$ for an $A$-sub-bimodule $N \in \Omega^1 A$, the calculus $\Omega^1(A)$ is left (resp. right) covariant if and only if $\Delta_{\mathrm{L}}^{U}(N) \subseteq A \otimes N$ (resp. $\Delta_{\mathrm{R}}^{U}(N) \subseteq N \otimes A$).

### The Woronowicz Theorems

A form $\omega$ in a left-covariant differential calculus $\Omega^1(A)$ is said to be left-invariant provided $\Delta_{\mathrm{L}}(\omega) = 1 \otimes \omega$. $\Omega^1(A)$ is a free $A$-module with basis given by left-invariant forms, that is, one can choose a set of left-invariant forms $\omega^i$ such that any 1-form $\rho$ can be uniquely written as a finite sum $\rho = \sum_i a_i\omega^i$, $a_i \in A$.

The first Woronowicz theorem states that there is a one-to-one correspondence between left-covariant calculi on $A$ and right $A$-ideals $Q \subseteq A^+$. The correspondence is provided by the map

$$\kappa : A \otimes Q \to N, \qquad a \otimes q \mapsto \sum aSq_{(1)} \otimes q_{(2)}$$

where $N$ is such that $\Omega^1(A) = (\Omega^1 A)/N$. The inverse of $\kappa$ reads $\kappa^{-1}(\sum_i a^i \otimes b^i) = \sum_i a^i b^i_{(1)} \otimes b^i_{(2)}$. The map $\kappa$ induces the map $\bar{\kappa} : A^+/Q \to \Omega^1(A)$, via $\omega([a]) = [\kappa(1 \otimes a)]$ where $[-]$ denotes cosets in $A^+/Q$ and in $\Omega^1(A) = (\Omega^1 A)/N$. This establishes a one-to-one correspondence between the space $\Lambda^1 = A^+/Q$ and the space of left-invariant 1-forms in $\Omega^1(A)$. The dual space to $\Lambda^1$, that is, the space of linear functionals $\Lambda^1 \to \mathbb{C}$, is often termed a "quantum Lie algebra" or a "quantum tangent space" corresponding to a left-covariant calculus $\Omega^1(A)$. The dimension of $\Lambda^1$ is known as a dimension of $\Omega^1(A)$.

The definitions and analysis of right-covariant differential calculi are done in a symmetric manner. For a bicovariant calculus, a form $\omega$ that is both left- and right-invariant, is termed a "bi-invariant" form.

The second Woronowicz theorem states a one-to-one correspondence between bicovariant differential calculi and Ad-invariant $A$-ideals $Q \subseteq A^+$ (cf. the subsection "Hopf algebra preliminaries"). The correspondence is provided by the map $\kappa$ above. For the universal calculus, $Q$ is trivial, and hence $\Lambda^1 = A^+ = \ker(\varepsilon)$.

### Higher-order Bicovariant Calculi

Given a first-order bicovariant calculus $\Omega^1(A)$, one constructs a braiding operator, known as the

"Woronowicz braiding" $\tau : \Omega^1(A) \otimes_A \Omega^1(A) \to \Omega^1(A) \otimes_A \Omega^1(A)$ by setting $\tau(a\omega \otimes_A \eta) = a\eta \otimes_A \omega$ for all $a \in A$, and any left-invariant $\omega$ and right-invariant $\eta$, and then extending it $A$-linearly to the whole of $\Omega^1(A) \otimes_A \Omega^1(A)$. This operator satisfies the braid relation $(\mathrm{id} \otimes_A \tau) \circ (\tau \otimes_A \mathrm{id}) \circ (\mathrm{id} \otimes_A \tau) = (\tau \otimes_A \mathrm{id}) \circ (\mathrm{id} \otimes_A \tau) \circ (\tau \otimes_A \mathrm{id})$, and is invertible provided the antipode $S$ is invertible. The Woronowicz braiding is used to define symmetric forms as those invariant under $\tau$. One then defines exterior 2-forms as elements of $\Omega^1(A) \otimes_A \Omega^1(A)/\ker(\mathrm{id} - \tau)$, and introduces the wedge product. The wedge product is not in general anticommutative, but one does have $\omega \wedge \eta = -\eta \wedge \omega$ for bi-invariant $\omega, \eta$. This construction is extended to higher forms and leads to the definition of the exterior algebra $\Omega(A)$. To define exterior $n$-forms, one maps any permutation on $n$-elements to the corresponding element of the braid group generated by $\tau$ and then takes the quotient of the $n$th tensor power of $\Omega^1(A)$ by all elements corresponding to even permutations. The differential $\mathrm{d} : A \to \Omega^1(A)$ is extended to an exterior differential in the whole of $\Omega(A)$ in the following way. First, $\Omega^1(A)$ is extended by a one-dimensional $A$-bimodule generated by a form $\theta$ that is required to be bi-invariant. The resulting extended bimodule (which, in general, is not a first-order differential calculus, as $\theta$ is not necessarily of the form $\sum_i a_i \,\mathrm{d}b_i$, for some $a_i, b_i \in A$) is then determined from the relation $\mathrm{d}a = \theta a - a\theta$ for all $a \in A$. Higher exterior derivative is then defined by $\mathrm{d}\rho = \theta \wedge \rho - (-1)^n \rho \wedge \theta$, for any $\rho \in \Omega^n(A)$.

The algebra $\Omega(A)$ is a $\mathbb{Z}_2$-graded differential Hopf algebra, that is, it has a coproduct such that

$$\Delta(\omega \wedge \eta) = \sum (-1)^{|\omega_{(2)}||\eta_{(1)}|} \omega_{(1)} \wedge \eta_{(1)} \otimes \omega_{(2)} \wedge \eta_{(2)}$$

where $|\omega_{(2)}|$ etc., denotes the degree of a homogeneous component in the decomposition of $\Delta(\omega)$. Furthermore,

$$\Delta(\mathrm{d}\omega) = \sum \left( \mathrm{d}\omega_{(1)} \otimes \omega_{(2)} + (-1)^{|\omega_{(1)}|} \omega_{(1)} \otimes \mathrm{d}\omega_{(2)} \right)$$

On the 1-forms this coproduct is simply the sum $\Delta_{\mathrm{L}} + \Delta_{\mathrm{R}}$.

### Classification

There is no unique covariant differential calculus on $A$, so classification of covariant differential calculi is an important problem. For example, it is known that the quantum group $\mathrm{SU}_q(2)$ admits a left-covariant three-dimensional calculus, but there is no three-dimensional bicovariant calculus. On the other hand, there are two four-dimensional bicovariant calculi on $\mathrm{SU}_q(2)$. Differential calculi are classified for standard quantum groups such as $\mathrm{SL}_q(N)$ or $\mathrm{Sp}_q(N)$.

General classification results are based on the equivalence between the category of Hopf bimodules of a finite-dimensional Hopf algebra $A$ and that of Yetter–Drinfeld or crossed modules of $A$. These are the modules of the Drinfeld double of $A$. As a result, in the case of a finite-dimensional factorizable coquasitriangular Hopf algebra $A$ with a dual Hopf algebra $H$, the bicovariant $\Omega^1(A)$ are in one-to-one correspondence with two-sided ideals in $H^+$. If, in addition, $A$ is semisimple, then (coirreducible) calculi are in one-to-one correspondence with nontrivial irreducible representations of $H$. This can be extended to infinite-dimensional algebras, provided one works over a field of formal power series in the deformation parameter.

## Quantum Group Principal Bundles

### Quantum Principal Bundles

In classical geometry, a (topological) principal bundle is a locally compact Hausdorff space with a (continuous) free and proper action of a locally compact group (e.g., a Lie group). In terms of algebras of functions this gives rise to the following structure. $A$ is a Hopf algebra (the model is functions on a group $G$), $P$ is a right $A$-comodule algebra with a coaction $\Delta_P : P \to P \otimes A$ (the model is functions on a total space $X$). Let $B = \{b \in P \mid \Delta_P(b) = b \otimes 1\}$ be the coinvariant subalgebra (the model is functions on a base manifold $M = X/G$). Fix a bicovariant calculus $\Omega^1(A)$, with the corresponding $Q$ and $\Lambda^1 = A^+/Q$ as in the subsection "The Woronowicz theorems." Take a differential calculus $\Omega^1(P) = \Omega^1 P/N_P$ such that:

1. $\Delta_{\Omega^1 P}(N_P) \subseteq N_P \otimes A$, where for all $\sum_i p^i \otimes q^i \in \Omega^1 P$,
$$\Delta_{\Omega^1 P}\left( \sum_i p^i \otimes q^i \right) = \sum_i p^i_{(0)} \otimes q^i_{(0)} \otimes p^i_{(1)} q^i_{(1)} \in \Omega^1 P \otimes A$$
2. $\tilde{\chi}(N_P) \subseteq N_P \otimes Q$, where
$$\tilde{\chi} : \Omega^1 P \to P \otimes A^+, \quad \sum_i p^i \otimes q^i \mapsto \sum_i p^i \Delta_P(q^i) = \sum_i p^i q^i_{(0)} \otimes q^i_{(1)}$$
3. $N_B = N_P \cap \Omega^1 B$ gives rise to a differential structure $\Omega^1(B) = \Omega^1 B/N_B$ on $B$. Condition (1) ensures that $\Delta_{\Omega^1 P}$ descends to a coaction $\Delta_{\Omega^1(P)} : \Omega^1(P) \to \Omega^1(P) \otimes A$, while (2) allows for defining a map
$$\mathrm{ver} : \Omega^1(P) \to P \otimes \Lambda^1, \qquad \mathrm{ver}([\omega]) = [\tilde{\chi}(\omega)]$$

Since $B$ is a subalgebra of $P$, the $P$-bimodule

$$P\Omega^1(B)P := \left\{ \sum_i p^i(\mathrm{d}b^i)q^i | p^i, q^i \in P, b^i \in B \right\}$$

is a sub-bimodule of $\Omega^1 P$, known as horizontal forms. $P$ is called a "quantum principal bundle" over $B$ with quantum structure group $A$ and calculi $\Omega^1(A)$ and $\Omega^1(P)$ provided the following sequence;

$$0 \longrightarrow P\Omega^1(B)P \longrightarrow \Omega^1(P) \xrightarrow{\mathrm{ver}} P \otimes \Lambda^1 \longrightarrow 0$$

is exact. This definition reflects the geometric content of principal bundles, but is not restricted to any specific differential calculus. The surjectivity of ver corresponds to the freeness of the (co)action, while the condition $\ker(\mathrm{ver}) = P\Omega^1(B)P$ corresponds to identification of vertical vector fields as those that are annihilated by horizontal forms.

### The Universal Calculus Case

In the universal calculus case, both $N_P$ and $Q$ in the previous subsection are trivial, and $\mathrm{ver} = \tilde{\chi}$. Universal horizontal forms $P(\Omega^1 B)P$ coincide with the kernel of the canonical projection $P \otimes_B P \to P \otimes P$. The exactness of the sequence in the last subsection is equivalent to the requirement that the map

$$\begin{aligned} \mathrm{can} : P \otimes_B P &\to P \otimes A \\ p \otimes_B q &\mapsto p\Delta_P(q) = \sum pq_{(0)} \otimes q_{(1)} \end{aligned}$$

be bijective. In algebra, such an inclusion of algebras $B \subseteq P$ is known as a Hopf–Galois extension. Thus, a geometric notion of a quantum principal bundle with the universal calculus is the same as the algebraic notion of a Hopf–Galois extension.

If (2) in the previous subsection is replaced by stronger conditions $\tilde{\chi}(N_P) = N_P \otimes Q$ and $(N_P \cap \ker \tilde{\chi}) \subseteq P(\Omega^1 B)P$, then exactness of the sequence in the previous subsection is equivalent to the bijectivity of "can." Thus, although defined in a purely algebraic way, the notion of a Hopf–Galois extension carries deep geometric meaning. It therefore makes sense to consider primarily Hopf–Galois extensions and then specify differential structure in such a way that this stronger version of (2) is satisfied. Henceforth, unless specified otherwise, a quantum principal bundle is taken with the universal differential calculus.

### Quantum Homogeneous Bundles

Suppose that $P$ is a Hopf algebra, and that there is a Hopf algebra surjection $\pi : P \to A$. This induces a coaction of $P$ on $A$ via $\Delta_P = (\mathrm{id} \otimes \pi) \circ \Delta$, where now $\Delta$ is a coproduct in $P$. $P$ is a quantum principal $A$-bundle over the coinvariants $B$, provided $\ker \pi \subseteq B^+P$, where $B^+ = B \cap P^+$. $B$ is a left quantum homogeneous space in the sense that $\Delta(B) \subseteq P \otimes B$, and $P$ is known as a quantum homogeneous bundle. An example of this is the standard quantum 2-sphere – a quantum homogeneous space of $SU_q(2)$ (see the subsection "The Dirac $q$-monopole"). This construction reflects the classical construction of a principal bundle over a homogeneous space, since every homogeneous space of a group $G$ can be identified with a quotient $G/H$, where $H \subseteq G$ is a subgroup. Not every quantum homogeneous space can be obtained in this way (e.g., nonstandard quantum 2-spheres), as quantum groups $P$ do not have sufficiently many quantum subgroups $A$ (in a sense of Hopf algebra projections $\pi : P \to A$). To study gauge theory on general quantum homogeneous spaces, more general notion of a bundle needs to be developed (see the subsection "Generalizations of quantum principal bundles").

A general differential calculus on a quantum homogeneous bundle is specified by choosing a left-covariant calculus on $P$ with an ideal $Q_P \in P^+$ such that $(\mathrm{id} \otimes \pi) \circ \mathrm{Ad}(Q_P) \subseteq Q_P \otimes A$. A bicovariant calculus on $A$ is then given by $Q_A = \pi(Q_P)$.

### Quantum Trivial Bundles

A quantum principal bundle (with the universal differential calculus) is said to be "trivial" or "cleft" provided there exists a linear map $\Phi : A \to P$ such that

1. $\Phi(1) = 1$ (unitality);
2. $\Delta_P \circ \Phi = (\Phi \otimes \mathrm{id}) \circ \Delta$ (*colinearity* or *covariance*); and
3. $\Phi$ is convolution invertible (cf. the subsection "Hopf algebra preliminaries").

$\Phi$ is called a trivialization. In this case, $P$ is isomorphic to $B \otimes A$ as a left $B$-module and right $A$-comodule via the map $B \otimes A \to P, b \otimes a \mapsto b\Phi(a)$. In particular, an $A$-covariant (i.e., colinear) algebra map $j : A \to P$ is a trivialization (the convolution inverse of $j$ is $j \circ S$).

Based on trivial bundles, locally trivial bundles can be constructed by choosing a compatible covering of $B$ (in terms of ideals).

At this point, the reader should be warned that the notion of a trivial quantum principal bundle includes bundles which are not trivial classically (i.e., do not correspond to functions on the Cartesian product of spaces). As an example, consider the Möbius strip viewed as a $\mathbb{Z}_2$-principal bundle over the circle $S^1$. Obviously, this is not a trivial bundle (the Möbius strip is not isomorphic to

$S^1 \times \mathbb{Z}_2$). It can be shown, however, that the quantum principal bundle corresponding to the Möbius strip has a trivialization $\Phi$ in the above sense.

### Generalizations of Quantum Principal Bundles

In the case of majority of quantum homogeneous spaces, the map $\pi$ in the subsection "Quantum homogeneous bundles" is a coalgebra and right $P$-module map, but not an algebra map. Thus, the induced coaction is not an algebra map either. To cover examples like these, one needs to introduce a generalization of quantum principal bundles. Consider an algebra $P$ that is also a right comodule of a coalgebra $C$ with coaction $\Delta_P$. Define

$$B := \Big\{ b \in P | \forall p \in P, \Delta_P(bp) = b\Delta_P(p) = \sum bp_{(0)} \otimes p_{(1)} \Big\}$$

$B$ is a subalgebra of $P$. $P$ is a principal coalgebra-bundle over $B$ or $B \subseteq P$ is a coalgebra-Galois extension provided the map

$$\mathrm{can} : P \otimes_B P \to P \otimes C$$
$$p \otimes_B q \mapsto p\Delta_P(q) = \sum pq_{(0)} \otimes q_{(1)}$$

is bijective. This purely algebraic requirement induces a rich symmetry structure on $P$, given in terms of entwining, which allows one for developing various differential geometric notions such as those discussed in the next section. The lack of space does not permit us to describe this theory here.

## Connections, Gauge Transformations, Matter Fields

### Connections and Connection Forms

A "connection" in a quantum principal bundle with calculi $\Omega^1(P), \Omega^1(A)$ is a left $P$-linear map $\Pi : \Omega^1(P) \to \Omega^1(P)$ such that:

1. $\Pi \circ \Pi = \Pi$ ($\Pi$ is a projection);
2. $\ker \Pi = P\Omega^1(B)P$; and
3. $\Delta_{\Omega^1(P)} \circ \Pi = (\Pi \otimes \mathrm{id}) \circ \Delta_{\Omega^1(P)}$ (colinearity or covariance).

The exact sequence in the subsection "Quantum principal bundles" implies that $\Pi$ is a left $P$-linear projection if and only if there exists a left $P$-linear map $\sigma : P \otimes \Lambda^1 \to \Omega^1(P)$ such that $\mathrm{ver} \circ \sigma = \mathrm{id}$. Since $\sigma$ is left $P$-linear, it is fully specified by its action on $\Lambda^1$. This leads to the equivalent definition of a connection as a connection form or a gauge field, that is, a map $\omega : \Lambda^1 \to \Omega^1(P)$ such that:

1. for all $\lambda \in \Lambda^1, \mathrm{ver}(\omega(\lambda)) = 1 \otimes \lambda$; and
2. $\Delta_{\Omega^1(P)} \circ \omega = (\omega \otimes \mathrm{id}) \circ \mathrm{Ad}^1_\Lambda$ (Ad-covariance), where $\mathrm{Ad}_{\Lambda^1}$ is a projection of the adjoint coaction to $\Lambda^1$, that is, $\mathrm{Ad}_{\Lambda^1}([a]) = [\mathrm{Ad}(a)]$ (well defined, because $Q$ is Ad-invariant for a bicovariant calculus, see the subsection "The Woronowicz theorems").

The correspondence between connections and connection 1-forms is given by the formula

$$\prod(p\mathrm{d}q) = \sum pq_{(0)}\omega([q_{(1)}])$$

In the universal differential calculus case, $\Lambda^1 = A^+$, hence $\omega$ can be viewed as a map $\omega : A \to \Omega^1 P$, such that $\omega(1) = 0$. The map $\mathcal{F}_\omega : A \to \Omega^2 P$, given by $\mathcal{F}_\omega = \mathrm{d}\omega + \omega * \omega$ is called a "curvature" of $\omega$. The curvature satisfies the Bianchi identity, $\mathrm{d}\mathcal{F}_\omega = \mathcal{F}_\omega * \omega - \omega * \mathcal{F}_\omega$.

In the case of a trivial bundle with trivialization $\Phi$ and universal calculus, any linear map $\beta : A \to \Omega^1 B$ such that $\beta(1) = 0$ defines a connection 1-form

$$\omega = \Phi^{-1} * \mathrm{d}\Phi + \Phi^{-1} * \beta * \Phi$$

The corresponding curvature is $\mathcal{F}_\omega = \Phi^{-1} * \mathcal{F}_\beta * \Phi$, where $\mathcal{F}_\beta = \mathrm{d}\beta + \beta * \beta$.

In the case of a quantum homogeneous bundle with calculus determined by $Q_P \in P^+$ and $Q_A = \pi(Q_P)$ (cf. the subsection "Quantum homogeneous bundles"), a canonical connection form can be assigned to any algebra map $i : A \to P$ such that

1. $\pi \circ i = \mathrm{id}$ ($i$-splits $\pi$);
2. $\varepsilon_P \circ i = \varepsilon_A$ (co-unitality);
3. $(\mathrm{id} \otimes \pi) \circ \mathrm{Ad}_P \circ i = (i \otimes \mathrm{id}) \circ \mathrm{Ad}_A$ (Ad-covariance); and
4. $i(Q_A) \subseteq Q_P$ (differentiability).

Explicitly, $\omega([a]) = \sum (Si(a)_{(1)})\mathrm{d}i(a)_{(2)}$.

### Covariant Derivative: Strong Connections

A covariant derivative associated to a connection $\Pi$ is a map $\mathrm{D} : P \to P\Omega^1(B)P, p \mapsto \mathrm{d}p - \Pi(\mathrm{d}p)$. A covariant derivative maps elements of $P$ into horizontal forms, since $\ker \Pi = P\Omega^1(B)P$, and satisfies the Leibniz rule $\mathrm{D}(bp) = (\mathrm{d}b)p + b\mathrm{D}p$, for all $b \in B$, $p \in P$.

A connection is "strong" provided $\mathrm{D}(p) \in \Omega^1(B)P$. A covariant derivative of a strong connection is a connection on module $P$ in the sense of Connes. Furthermore, in the universal calculus case, and when $A$ has invertible antipode, the existence of strong connections leads to rich gauge theory of associated bundles (cf. the subsection "Associated bundles: matter fields"). A connection in a trivial bundle

described in the subsection "Connections and connection forms" is strong (and every strong connection in a trivial bundle is of this form). Assuming invertibility of the antipode in $A$, a canonical connection in a quantum homogeneous bundle described in that subsection is strong provided Ad-covariance (3) is replaced by conditions $(\mathrm{id}\otimes\pi)\circ\Delta\circ i=(i\otimes\mathrm{id})\circ\Delta_A$ (right covariance) and $(\pi\otimes\mathrm{id})\circ\Delta\circ i=(\mathrm{id}\otimes i)\circ\Delta_A$ (left covariance), where $\Delta$ is a coproduct in $P$, and $\Delta_A$ is a coproduct in $A$.

In the universal calculus case, the map D can be extended to a map $\mathrm{D}:\Omega^1P\to\Omega^2P$ via the formula $\mathrm{D}(\rho)=\mathrm{d}\rho+\sum\rho_{(0)}\omega(\rho_{(1)})$. Then $\mathrm{D}\circ\mathrm{D}(p)=\sum p_{(0)}\mathcal{F}_\omega(p_{(1)})$, where $\mathcal{F}_\omega$ is the curvature of $\omega$ (cf. the subsection "Connections and connection forms"). This explains the relationship between a curvature understood as the square of a covariant derivative and $\mathcal{F}_\omega$.

## Bundle Automorphisms and Gauge Transformations

A quantum bundle automorphism is a left $B$-linear right $A$-covariant (i.e., colinear) automorphism $F:P\to P$ such that $F(1)=1$. Bundle automorphisms form a group with operation $FG=G\circ F$. This group is isomorphic to the group $\mathcal{G}(P)$ of gauge transformations, that is, maps $f:A\to P$ that satisfy the following conditions:

1. $f(1_A)=1_P$ (unitality);
2. $\Delta_P\circ f=(f\otimes\mathrm{id})\circ\mathrm{Ad}$ (Ad-covariance); and
3. $f$ is convolution invertible (cf. the subsection "Hopf algebra preliminaries").

The product in $\mathcal{G}(P)$ is the convolution product (cf. the subsection "Hopf algebra preliminaries"). The group of gauge transformations acts on the space of (strong) connection forms $\omega$ via the formula

$$f\triangleright\omega=f*\omega*f^{-1}+f*\mathrm{d}f^{-1},\quad\forall f\in\mathcal{G}(P)$$

This resembles the gauge transformation law of a gauge field in the standard gauge theory. The curvature transforms covariantly as $\mathcal{F}_{f\triangleright\omega}=f*\mathcal{F}_\omega*f^{-1}$.

In the case of a trivial principal bundle, gauge transformations correspond to a change of the trivialization and can be identified with convolution-invertible maps $\gamma:A\to B$ such that $\gamma(1)=1$. A map $\beta:A\to\Omega^1B$ that induces a connection as in the subsection "Connections and connection forms" is transformed to $\gamma*\beta*\gamma^{-1}+\gamma*\mathrm{d}\gamma^{-1}$, and the curvature $\mathcal{F}_\beta\mapsto\gamma*\mathcal{F}_\beta*\gamma^{-1}$.

## Associated Bundles: Matter Fields

Given a right $A$-comodule (corepresentation) $\varrho:V\to V\otimes A$ one defines a quantum vector bundle associated to $P$ as

$$E=\left\{\sum_i v^i\otimes p^i\in V\otimes P\,\middle|\,\sum_i v^i_{(0)}\otimes p^i_{(0)}\otimes v^i_{(1)}p^i_{(1)}=\sum_i v^i\otimes p^i\otimes 1\right\}\subseteq V\otimes P$$

$E$ is a right $B$-module with product $(\sum_i v^i\otimes p^i)b=\sum_i v^i\otimes p^ib$. A right $B$-linear map $s:E\to B$ is called a section of $E$. The space of sections $\Gamma(E)$ is a left $B$-module via $(bs)(p)=bs(p)$.

The theory of associated bundles is particularly rich when $A$ has a bijective antipode and $P$ has a strong connection form $\omega$. In this case, $\Gamma(E)$ is isomorphic to the left $B$-module $\Gamma_\varrho$ of maps $\phi:V\to P$ such that $\Delta_P\circ\phi=(\phi\otimes\mathrm{id})\circ\varrho$. If $V$ is finite dimensional, then $\Gamma_\varrho$ is a finite projective $B$-module, that is, it is a module of sections of a noncommutative vector bundle in the sense of Connes. The strong connection induces a map $\nabla:\Gamma_\varrho\to\Omega^1B\otimes_B\Gamma_\varrho$, given by $\nabla(\phi)(v)=\mathrm{d}\phi(v)+\sum\phi(v_{(0)})\omega(v_{(1)})$. $\nabla$ is a connection in the sense of Connes (in a projective left $B$-module), that is, for all $b\in B,\phi\in\Gamma_\varrho$, $\nabla(b\phi)=\mathrm{d}b\otimes_B\phi+b\nabla(\phi)$.

In the case of a trivial bundle, $\Gamma_\varrho$ can be identified with the space of linear maps $V\to B$. Thus, sections of an associated bundle correspond to pullbacks of matter fields, as in the classical local gauge theory matter fields are defined as functions on a spacetime with values in a representation (vector) space of the gauge group.

## The Dirac $q$-Monopole

This is an example of a strong connection in a quantum homogeneous bundle (cf. the subsection "Quantum homogeneous bundles"). $P=\mathrm{SU}_q(2)$ is a matrix Hopf $*$-algebra with matrix of generators

$$\begin{pmatrix} a & -qc^* \\ c & a^* \end{pmatrix}$$

and relations

$$ac=qca,\qquad ac^*=qc^*a,\qquad cc^*=c^*c$$
$$a^*a+c^*c=1,\qquad aa^*+q^2cc^*=1$$

where $q$ is a real parameter. $A=\mathbb{C}[U(1)]$ is a Hopf $*$-algebra generated by unitary and group-like $u$ (i.e., $uu^*=u^*u=1,\Delta(u)=u\otimes u$). The $*$-projection $\pi:P\to A$ is defined by $\pi(a)=u$. The coinvariant subalgebra $B$ is generated by $x=cc^*$, $z=ac^*$, $z^*=ca^*$. The elements $x$ and $z$ satisfy relations

$$x^*=x,\qquad zx=q^2xz,$$
$$zz^*=q^2x(1-q^2x),\qquad z^*z=x(1-x)$$

Thus, $B$ is the algebra of functions on the standard quantum 2-sphere. A strong connection is obtained from a bicovariant $*$-map $i: A \to P$ given by $i(u^n) = a^n$ (cf. the subsections "Quantum homogeneous bundles," "Connections and connection forms," and "Covariant derivative: strong connections"). Explicitly, the connection form reads

$$\omega(u^n) = \sum_{k=0}^{n} \binom{n}{k}_{q^{-2}} c^{*k} a^{*n-k} \mathrm{d}(a^{n-k} c^k)$$
$$\omega(u^{*n}) = \sum_{k=0}^{n} q^{2k} \binom{n}{k}_{q^{-2}} a^{n-k} c^k \mathrm{d}(c^{*k} a^{*n-k})$$

where the deformed binomial coefficients are defined for any number $x$ by

$$\binom{n}{k}_x = \frac{(x^n - 1)(x^{n-1} - 1) \cdots (x^{k+1} - 1)}{(x^{n-k} - 1)(x^{n-k-1} - 1) \cdots (x - 1)}$$

There is a family $V_n, n \in \mathbb{Z}$ of one-dimensional corepresentations of $\mathbb{C}[\mathrm{U}(1)]$ with $V_n = \mathbb{C}$ and $\varrho^n(1) = 1 \otimes u^n, n \geq 0$ and $\varrho^n(1) = 1 \otimes u^{*n}, n < 0$. This leads to the family of finite projective modules $\Gamma_n = \Gamma_{\varrho^n}$ as described in the subsection "Associated bundles: matter fields." The Hermitian projectors $\boldsymbol{e}(n)$ of these modules come out as, for $n > 0$,

$$e(n)_{ij} = \sqrt{\binom{n}{i}_{q^{-2}} \binom{n}{j}_{q^{-2}}} a^{n-i} c^i c^{*j} a^{*n-j}, \quad i, j = 0, 1, \ldots, n$$
$$e(-n)_{ij} = q^{i+j} \sqrt{\binom{n}{i}_{q^{-2}} \binom{n}{j}_{q^{-2}}} c^{*i} a^{*n-i} a^{n-j} c^j, \quad i, j = 0, 1, \ldots, n$$

The $\boldsymbol{e}(n)$ describe $q$-monopoles of magnetic charge $-n$. For example, the charge-1 projector explicitly reads

$$\begin{pmatrix} 1 - x & z^* \\ z & q^2 x \end{pmatrix}$$

and reduces to the usual charge-1 Dirac monopole projector when $q = 1$. The covariant derivatives $\nabla$ are Levi-Civita or Grassmann connections in modules $\Gamma_n$ corresponding to projectors $\boldsymbol{e}(n)$.

### The $q$-Instanton

This is an example of a coalgebra bundle and the associated vector bundle, which is a deformation of an instanton (with instanton number 1). $P = \mathbb{C}[S_q^7]$ is the $*$-algebra of polynomial functions on the quantum 7-sphere. As a $*$-algebra it is defined by generators $z_1, z_2, z_3, z_4$ and relations

$$z_i z_j = q z_j z_i \quad (\text{for } i < j)$$
$$z_j^* z_i = q z_i z_j^* \quad (\text{for } i \neq j)$$
$$z_k^* z_k = z_k z_k^* + (1 - q^2) \sum_{j<k} z_j z_j^*,$$
$$\sum_{k=1} z_k z_k^* = 1$$

where $q \in \mathbb{R}$. The coaction of the $*$-Hopf algebra $A = \mathrm{SU}_q(2)$ (cf. the previous subsection) on $P$ is constructed as follows. Start with the quantum group $\mathrm{U}_q(4)$, generated by a matrix $\boldsymbol{t} = (t_{ij})_{i,j=1}^4$ and view $\mathbb{C}[S_q^7]$ as a right quantum homogeneous space of $\mathrm{U}_q(4)$ generated by the bottom row in $\boldsymbol{t}$. Thus, there is a right coaction of $\mathrm{U}_q(4)$ on $\mathbb{C}[S_q^7]$ obtained by the restriction of the coproduct in $\mathrm{U}_q(4)$. Next, project $\mathrm{U}_q(4)$ to $\mathrm{SU}_q(2)$ by a suitable coideal and a right ideal in $\mathrm{U}_q(4)$. The corresponding canonical surjection $r: \mathrm{U}_q(4) \to \mathrm{SU}_q(2)$ is a coalgebra map, characterized as a right $\mathrm{U}_q(4)$-module map by $r(t_{11} t_{22} - q t_{12} t_{21}) = 1$ and

$$r(\boldsymbol{t}) = \begin{pmatrix} \boldsymbol{u} & 0 \\ 0 & \boldsymbol{u} \end{pmatrix}, \qquad \bar{\boldsymbol{u}} = \begin{pmatrix} u_{22} & -u_{21} \\ -u_{12} & u_{11} \end{pmatrix}$$

where $\boldsymbol{u} = (u_{ij})_{i,j=1}^2$ is the matrix of generators of $\mathrm{SU}_q(2)$ (cf. the previous subsection). When applied to the coaction of $\mathrm{U}_q(4)$ on $\mathbb{C}[S_q^7]$, $r$ induces the required coaction $\Delta_P: \mathbb{C}[S_q^7] \to \mathbb{C}[S_q^7] \otimes \mathrm{SU}_q(2)$. Explicitly, the coaction comes out on generators as $\Delta_P(z_j) = \sum_i z_i \otimes r(t_{ij})$. The coaction $\Delta_P$ is not an algebra map. The coinvariant subalgebra $B$ is a $*$-algebra generated by

$$a = z_1 z_4^* - z_2 z_3^*$$
$$b = z_1 z_3 + q^{-1} z_2 z_4$$
$$R = z_1 z_1^* + z_2 z_2^*$$

The elements $a, a^*, b, b^*, R$ satisfy the following relations:

$$Ra = q^{-2} aR, \qquad Rb = q^2 bR$$
$$ab = q^3 ba, \qquad ab^* = q^{-1} b^* a$$
$$aa^* + q^2 bb^* = R(1 - q^2 R)$$
$$aa^* = q^2 a^* a + (1 - q^2) R^2$$
$$b^* b = q^4 bb^* + (1 - q^2) R$$

Hence $B$ can be understood as a deformation of the algebra of functions on the 4-sphere and is denoted by $\mathbb{C}[\Sigma_q^4]$. One can show that the map "can" in the subsection "Generalizations of quantum principal bundles" is bijective, hence there is an $\mathrm{SU}_q(2)$-coalgebra principal bundle with the total space the quantum 7-sphere $\mathbb{C}[S_q^7]$ and the base space the

quantum 4-sphere $\mathbb{C}[\Sigma_q^4]$. By abstract arguments that involve cosemisimplicity of $SU_q(2)$, one can prove that there exists a strong connection in this bundle; this is the $q$-deformed instanton field. At the time of writing this article, however, the explicit form of this connection is not known.

On the other hand, following the classical construction of an instanton, one can take the fundamental two-dimensional corepresentation $V = \mathbb{C}^2$ of $SU_q(2)$ and explicitly construct $q$-instanton projection with instanton number 1. Writing $e_1, e_2$ for the basis of $V$, the coaction $\varrho: V \to V \otimes SU_q(2)$ is given by

$$(e_j) = \sum_i e_i \otimes u_{ij}$$

The associated bundle (cf. the subsection "Associated bundles: matter fields") is a finite projective left module over $\mathbb{C}[\Sigma_q^4]$. The corresponding $q$-instanton projector comes out as

$$\begin{pmatrix} q^2 R & 0 & qa & q^2 b \\ 0 & q^2 R & qb^* & -q^3 a^* \\ qa^* & qb & 1-R & 0 \\ q^2 b^* & -q^3 a & 0 & 1-q^4 R \end{pmatrix}$$

*See also:* Bicrossproduct Hopf Algebras and Noncommutative Spacetime; Hopf Algebras and $q$-Deformation Quantum Groups; Noncommutative Tori, Yang–Mills, and String Theory.

## Further Reading


Bonechi F, Ciccoli N, and Tarlini M (2003) Noncommutative instantons and the 4-sphere from quantum groups. *Communications in Mathematical Physics* 226: 419–432.

Bonechi F, Ciccoli N, Dąbrowski L, and Tarlini M (2004) Bijectivity of the canonical map for the noncommutative instanton bundle. *Journal of Geometry and Physics* 51: 71–81.

Brzeziński T and Majid S (1993) Quantum group gauge theory on quantum spaces. *Communications in Mathematical Physics* 157: 591–638.

Brzeziński T and Majid S (1995) Quantum group gauge theory on quantum spaces – Erratum. *Communications in Mathematical Physics* 167: 235.

Brzeziński T and Majid S (2000) Quantum geometry of algebra factorisations and coalgebra bundles. *Communications in Mathematical Physics* 213: 491–521.

Calow D and Matthes R (2002) Connections on locally trivial quantum principal fibre bundles. *Journal of Geometry and Physics* 41: 114–165.

Dąbrowski L, Grosse H, and Hajac PM (2001) Strong connections and Chern–Connes pairing in the Hopf–Galois theory. *Communications in Mathematical Physics* 220: 301–331.

Hajac PM and Majid S (1999) Projective module description of the $q$-monopole. *Communications in Mathematical Physics* 206: 247–264.

Heckenberger I and Schmüdgen K (1998) Classification of bicovariant differential calculi on the quantum groups $SL_q(n+1)$ and $Sp_q(2n)$. *Journal für die Reine und Angewandte Mathematik* 502: 141–162.

Klimyk A and Schmüdgen K (1997) *Quantum Groups and Their Representations*. Berlin: Springer.

Majid S (1998) Classification of bicovariant differential calculi. *Journal of Geometry and Physics* 25: 119–140.

Majid S (1999) Quantum and braided Riemannian geometry. *Journal of Geometry and Physics* 30: 113–146.

Majid S (2000) *Foundations of Quantum Group Theory*, 1st pbk. edn. Cambridge: Cambridge University Press.

Wess J and Zumino B (1990) Covariant differential calculus on the quantum hyperplane, *Nuclear Physics B. Proceedings Supplement* 18B: 302–312.

Woronowicz SL (1989) Differential calculus on compact matrix pseudogroups (quantum groups). *Communications in Mathematical Physics* 122: 125–170.


# Quantum Hall Effect

**K Hannabuss**, University of Oxford, Oxford, UK



## Introduction

When a current flows in a thin sample with a transverse magnetic field $B$, the Lorentz force deflects the trajectories of the charge carriers, producing an excess charge on one side and a charge deficiency on the other, and creating a potential difference across the conductor perpendicular to both the direct current and the magnetic field. This is known as the Hall effect, in honour of E H Hall, who, inspired by a remark of Maxwell, first demonstrated it in thin samples of gold foil in October 1879 (Hall's subsequent measurements of the potential difference showed that the carriers could be positively or negatively charged for different materials). A schematic diagram of Hall's experiment and the lateral separation of charges is shown in Figure 1.

Equilibrium is reached when the magnetic force balances that from the potential difference $E$ due to the displaced charge. When the charge carriers are electrons, with the electron density $n$, and the electron current $J$, this gives $neE = JB$. Comparison with Ohm's law, $J = \sigma E$, gives conductance (the reciprocal of resistance) to be $\sigma = ne/B$. More generally, considering the currents and fields as

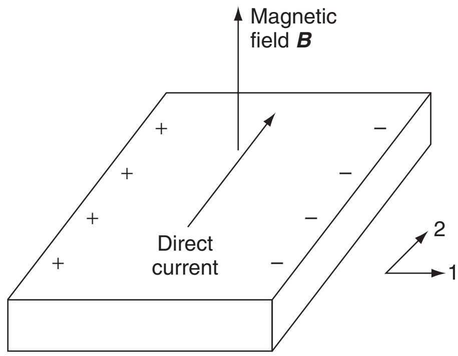



**Figure 1** Schematic diagram of charge separation in Hall's experiment.

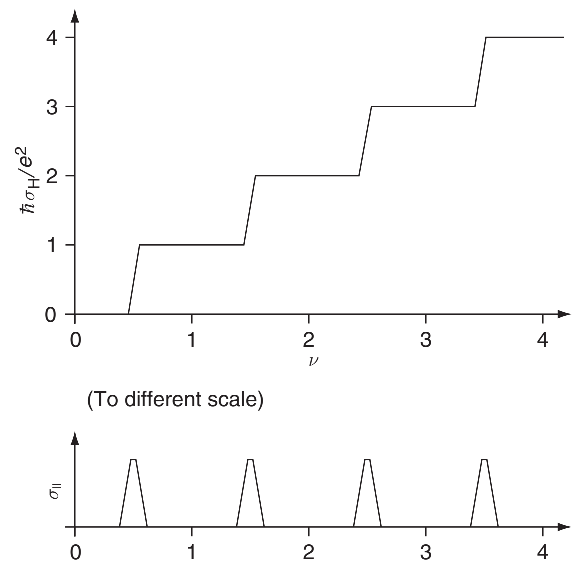



**Figure 2** Schematic diagram of the Hall and direct conductivities plotted against the filling factor $\nu$.

vectors, $\sigma$ is represented by a matrix. Rescaling by the sample thickness $\delta$, the diagonal components of $\delta\sigma$ give the direct conductivity $\sigma_{\|}$ and its off-diagonal elements give the Hall conductivity: $\sigma_H = \delta\sigma_{21}$. (For systems symmetric under 90° rotations, $\sigma_{11} = \sigma_{22}$ and $\sigma_{12} = -\sigma_{21}$.) In quantum theory, one usually works in terms of the filling fraction $\nu = n\delta h/eB$ and then $\sigma_H = \nu e^2/h$.

In 1980 von Klitzing, Dorda, and Pepper discovered that at very low temperatures in very high magnetic fields, the Hall conductivity $\sigma_H$ is quantized as integral multiples of $e^2/\hbar$, a fact known as the integer quantum hall effect (IQHE). The integer multiples were accurate to 1 part in $10^8$, and the effect was exceptionally robust against changes in the geometry of the samples and in the experimental parameters. Indeed, the unprecedented accuracy of the effect led to its adoption as the international standard for resistance in 1990.

More precisely, the Hall conductivity was no longer proportional to the filling fraction $\nu$, but the graph of $\sigma_H$ against $\nu$ displayed a sequence of jumps, as shown in Figure 2. In this figure, the conductivity has plateau at the integer multiples of $e^2/\hbar$, and jumps between them within fairly small ranges of the filling fraction. Moreover, the direct conductivity vanishes where the Hall conductivity takes its constant integral values.

These results raise numerous questions.

1. Why does the conductivity take such precise integer values, and why are they so stable under changes of the geometry and physical parameters?
2. Why does the direct conductivity vanish, except in regions where the Hall conductivity jumps between integer values, and how are such jumps possible?

Moreover, any theory must also explain why these features are not present under the more normal conditions of the classical Hall effect. The following features seem to play a role, and in the case of the first three, even in the classical effect.

1. As Hall discovered, the samples must be very thin to exhibit even the classical effect. (Nowadays they are often a surface layer between two semiconductors.)
2. The samples are macroscopic and much larger than the quantum wavelengths appearing in the problem.
3. The electric field is small enough that nonlinear effects are negligible.
4. The quantum effect appears only at a very low temperature.

The first of these suggests that we should idealize to the case where the motion of the charge carriers is restricted to a two-dimensional region, and the second that we may work in the thermodynamic limit where the conducting surface is the whole of $\boldsymbol{R}^2$. The third and fourth ensure both that the linear Ohm's law should be adequate, and also that it should be enough to consider the limiting cases of very weak electric fields and zero temperature. Multiple limits of this sort raise delicate mathematical issues. Indeed, many plausible models of the effect turn out, on careful analysis, to predict vanishing Hall conductivity.

A theoretical explanation of the quantization of the conductivity was soon suggested by Laughlin. Exploiting the apparent independence of sample geometry, he considered a cylindrical conductor where quantization followed on consideration of

the flux tubes threading it. Laughlin's choice of a particular configuration precluded investigation of the influence of changing geometry. This was soon provided by Thouless, Kohmoto, Nightingale, and de Nijs, who argued (from a lattice version of the problem) that the conductivity could be identified with the Chern character of a line bundle over a Brillouin zone (a quotient of momentum space by the action of the reciprocal crystal lattice), so that it had to be integral and the stability of the effect was a consequence of the topological nature of $\sigma$. Unfortunately, whilst suggestive, this explanation worked only under the physically implausible constraint that the magnetic flux through a crystal cell was rational, offered no explanation of the link between the Hall and direct conductivities, and, working with a periodic Hamiltonian, made no allowance for the impurities and disorder usually important in solid-state problems of this sort.

Notwithstanding these deficiencies, this model contained important insights, which inspired Bellissard to model the effect using Connes' newly developed noncommutative geometry (Bellissard 1986, Connes 1986). (Kunz produced a Hilbert space theory at about the same time, but that has been rather less influential.) Connes' work turned out to contain all the relevant concepts and tools needed to provide a good understanding of the effect, based on interpreting the conductivity as a noncommutative Chern character for a noncommutative version of the Brillouin zone. In fact, the techniques of noncommutative geometry seemed to fit the quantum Hall effect so well that this has become one of the standard examples of the theory.

Even whilst the theorists were struggling to explain the experiments, observations by Tsui, Störmer, and Gossard showed that, with suitable care, fractional Hall conductivities could also be observed, although these were far less stable than those given by integers. One, therefore, distinguishes between IQHE and the fractional quantum Hall effect (FQHE), and this survey concentrates largely on the former. One simplifying feature of the IQHE is that it seems to be comprehensible at the level of individual noninteracting electrons, whereas the FQHE certainly involves some kind of interaction and many-body theory.

This article presents an outline of the connection between noncommutative differential geometry and the IQHE, and concludes by discussing some of the approaches to the FQHE, and some other applications of noncommutative geometry and mathematical directions suggested by the theory. The sections alternate between the physical model and the mathematical abstraction from it.

There are good surveys of the area (Bellissard *et al.* 1994, McCann 1998) explaining how the mathematical model arises out of the physics, the mathematical models themselves. As well as being the standard reference for noncommutative geometry, Connes (1994) discusses the Hall effect. These resources contain good bibliographies, which may be consulted for further references.

## Electron Motion in a Magnetic Field

The following discussion restricts attention to motion in two dimensions, with electrons as the charge carriers, and no interactions between them. (The first condition is essential; the second could be relaxed a little to allow sufficiently long-lived quasi-particles.) A single free electron with mass $m$ and charge $e$ moving in the $x_1$–$x_2$ plane with a constant transverse magnetic field $\boldsymbol{B}$ in the positive $x_3$-direction, can be described by the Landau Hamiltonian

$$H_{\mathrm{L}} = |\boldsymbol{P} - e\boldsymbol{A}|^2/2m \qquad [1]$$

where $\boldsymbol{A} = \frac{1}{2}\boldsymbol{B} \times \boldsymbol{X}$ is a magnetic vector potential that gives rise to $\boldsymbol{B}$. This problem is exactly solvable by, for example, introducing $\boldsymbol{K}_{\pm} = (K_{\pm}^1, K_{\pm}^2) = \boldsymbol{P} \mp e\boldsymbol{A}$. The components of $\boldsymbol{K}_+$ and $\boldsymbol{K}_-$ commute with each other, but $[K_{\pm}^1, K_{\pm}^2] = \pm i\hbar eB$. Comparison with the harmonic oscillator shows that the energy spectrum of $H_{\mathrm{L}} = [(K_+^1)^2 + (K_+^2)^2]/2m$ is $\{(n + \frac{1}{2})\hbar eB/m \colon n \in \boldsymbol{Z}\}$. Since $H_{\mathrm{L}}$ commutes with the components of $\boldsymbol{K}_-$, each of these Landau energy levels is infinitely degenerate, and the filling fraction $\nu$ measures what proportion of states in the Landau levels are filled. The frequency $\omega_c = eB/m$ is the cyclotron frequency for classical circular orbits in the magnetic field.

The degeneracy of the Landau Hamiltonian can also be understood in terms of the magnetic translations obtained by exponentiating the connection defined by the magnetic potential $\boldsymbol{A}$: $\nabla_j = \partial_j + ieA_j/\hbar = iK_-^j/\hbar$. More precisely, we set

$$U(\boldsymbol{a}) = \exp(-i\boldsymbol{a} \cdot \nabla) = \exp(-i\boldsymbol{a} \cdot \boldsymbol{K}_-/\hbar) \qquad [2]$$

which clearly commutes with $H_{\mathrm{L}}$, expressing the translational symmetry of this model. The curvature $[\nabla_1, \nabla_2] = B$ of the connection manifests itself in the identities

$$U(\boldsymbol{a})U(\boldsymbol{b}) = e^{(1/2)i\phi}U(\boldsymbol{a} + \boldsymbol{b}) = e^{i\phi}U(\boldsymbol{b})U(\boldsymbol{a}) \qquad [3]$$

where $\phi = e\boldsymbol{B} \cdot (\boldsymbol{a} \times \boldsymbol{b})/\hbar$ measures the magnetic flux through the parallelogram spanned by $\boldsymbol{a}$ and $\boldsymbol{b}$. These show that $U$ is a projective representation of $\boldsymbol{R}^2$ with projective multiplier $\lambda(\boldsymbol{a}, \boldsymbol{b}) = \exp(\frac{1}{2}i\phi)$.

The significance of this is that, unless $\phi$ is an integer multiple of $2\pi$, $U(\boldsymbol{a})$ and $U(\boldsymbol{b})$ generate a noncommutative algebra. This replaces the commutative algebra of functions on two-dimensional momentum space and leads naturally to a noncommutative geometry.

The unembellished Landau Hamiltonian cannot describe the Hall effect without adding an electric potential $e\boldsymbol{E}\cdot\boldsymbol{X}$ to drive the current in the sample. (Alternatively, and useful for the later discussion, one could use the radiation gauge in which, instead of introducing a scalar potential, a time-dependent term is added to $\boldsymbol{A}$ so that $\boldsymbol{E}=-\partial\boldsymbol{A}/\partial t$.)

The quantum Hall effect also depends crucially on the effects of impurities in the conducting material. These can be modeled by adding a random potential $V_\omega$ with $\omega$ in a compact probability space $\Omega$ to obtain $H_\omega = H_\mathrm{L} + e\boldsymbol{E}\cdot\boldsymbol{X} + V_\omega(X)$. A continuous function $f$ on $\Omega$ can be interpreted as a random variable, and its expectation $\tau_\Omega(f)$ gives a trace on the $C^*$-algebra $C(\Omega)$ (i.e., a positive linear functional such that $\tau_\Omega(AB)=\tau_\Omega(BA)$).

Although the magnetic translations commute with $H_\mathrm{L}$, they do not generally commute with the potentials so they act on $\Omega$, but, on the other hand, the physics of a disordered system and its translates should be the same, so we assume that the probability measure and hence also $\tau_\Omega$ are invariant under magnetic translations. (As noted earlier, we work in the thermodynamic limit, where the Hall sample expands to fill $\boldsymbol{R}^2$, so we do not need to worry about translations moving the sample itself.) Then $\Omega$ with the magnetic translation action can be interpreted as the noncommutative Brillouin zone. (A space $\Omega$ can be reconstructed from the magnetic translations of the resolvents of the Hamiltonians (Bellisard *et al.* 1994).)

The current $\boldsymbol{J}$ may be defined as the functional derivative of the Hamiltonian with respect to the vector potential $\boldsymbol{A}$ or, in components, $J_k = \delta_k H = \delta H/\delta A_k$. For the Landau Hamiltonian, this gives

$$\mathrm{i}\hbar\delta_k H_\mathrm{L} = \mathrm{i}e\hbar(P_k - eA_k)/m = e[X_k, H_\mathrm{L}] \qquad [4]$$

a relation which persists for $H = H_\mathrm{L} + V(X)$ whenever the potential $V$ is independent of $\boldsymbol{A}$, so that $\delta_k H = -\mathrm{i}e[X_k, H]/\hbar = e\,\mathrm{d}X_k/\mathrm{d}t$, the charge times velocity, as one might expect. The operator functional calculus delivers a similar formula for derivations of the spectral projections of $H$. We have $\delta_k = e\partial_k/\hbar$, where, in view of the commutation relations, $\partial_k = -\mathrm{i}[X_k, \cdot\,]$ can be regarded as a momentum-space derivative, confirming that we are dealing with the differential geometry of momentum space.

We now wish to calculate the expected current $\langle J_k\rangle$, in a thermal state with chemical potential $\mu$ at inverse temperature $\beta = 1/kT$ (where $k$ is Boltzmann's constant and the temperature is $T$(kelvin). Using the Fermi–Dirac distribution, the grand canonical expectation value is

$$\langle J_k\rangle = \mathrm{tr}\left[\left(1+\mathrm{e}^{-\beta(H-\mu)}\right)^{-1} J_k\right] \qquad [5]$$

Since the quantum Hall effect occurs at low temperatures (large $\beta$) and for weak fields, we formally proceed to those limits. Then $(1+\mathrm{e}^{-\beta(H-\mu)})^{-1}$ tends to the projection $P_\mathrm{F}$ onto the states with energy less than the Fermi energy $E_\mathrm{F}$ in the absence of the electric field. The limiting expected current is, therefore, $\mathrm{tr}(P_\mathrm{F}J_k) = \mathrm{tr}(P_\mathrm{F}\delta_k H)$, where $H$ is now the Hamiltonian including the electric field (without which there would be no current).

A detailed calculation of the Hall conductivity using the Kubo–Greenwood formula shows that the conductivity matrix is actually

$$\sigma_{kj} = \mathrm{i}(e^2/\hbar)\mathrm{tr}(P_\mathrm{F}[\partial_j P_\mathrm{F}, \partial_k P_\mathrm{F}]) \qquad [6]$$

In particular, this immediately implies that the direct conductivity terms $\sigma_{jj}$ vanish, as observation suggested. The derivation of [6] requires great care, and references may be found in the surveys, but a formal argument in the next section may lend this expression some plausibility.

## The Noncommutative Geometry

The principal ingredient for noncommutative geometry is an algebra, and thus we shall now consider a class of algebras broad enough to include the physical example.

The action of the magnetic translations on $\Omega$ defines automorphisms of the $C^*$-algebra $C(\Omega)$, which permit the construction of a twisted crossed-product algebra, in which these automorphisms are represented by conjugation. Because much of the theory has been formulated with lattice approximations using $\boldsymbol{Z}^2$ rather than $\boldsymbol{R}^2$, it is useful to work more generally with a separable locally compact abelian group $G$ with continuous multiplier $\lambda$, and a homomorphism $\alpha$ to automorphisms of a $C^*$-algebra $\mathcal{A}_1$ with trace $\tau_1$, which will in practice be the commutative algebra $C(\Omega)$ with $\tau_\Omega$. The twisted crossed product $\mathcal{A} = C(\mathcal{A}_1, G, \lambda)$ can be constructed as the norm completion of the continuous compactly

supported functions from $G$ to $\mathcal{A}_1$ with the product, adjoint and norm

$$(f^*g)(x) = \int_G \lambda(y, x-y) f(y)(\alpha_y g)(x-y)\,\mathrm{d}y \quad [7]$$

$$f^*(x) = \lambda(x, -x)^{-1} f(-x)^* \quad [8]$$

$$\|f\| = \max\left\{\int_G \|f(x)\|_{\mathcal{A}_1}\,\mathrm{d}x, \int_G \|f^*(x)\|_{\mathcal{A}_1}\,\mathrm{d}x\right\} \quad [9]$$

integration being with respect to the Haar measure. The crossed-product algebra is noncommutative, both because of the action of $G$ and due to the multiplier $\lambda$. It has a trace $\tau[f] = \tau_1[f(0)]$ and, when $G = \boldsymbol{R}^2$, has derivations given by $\partial_k f = -\mathrm{i}x_k f(\boldsymbol{x})$.

As an example, consider the case of periodic potentials invariant under translation by vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. Then the group $G \cong \boldsymbol{Z}^2$ generated by $\boldsymbol{a}$ and $\boldsymbol{b}$ acts trivially on $\Omega$ and the crossed-product algebra is just a product of $\mathcal{A}_1$ and the twisted group algebra of complex-valued functions $C(\boldsymbol{C}, G, \lambda)$, generated by $U(\boldsymbol{a})$ and $U(\boldsymbol{b})$. We already noted that the algebra is commutative only when the flux $\phi \in 2\pi\boldsymbol{Z}$, in which case it is just the convolution algebra of $\boldsymbol{Z}^2$, which by Fourier transforming (effectively setting $U(\boldsymbol{a}) = \mathrm{e}^{\mathrm{i}\alpha}$ and $U(\boldsymbol{b}) = \mathrm{e}^{\mathrm{i}\beta}$) is the algebra $C(\boldsymbol{T}^2)$, with torus coordinates $\alpha$ and $\beta$. For fluxes which are rational multiples of $2\pi$ we obtain a matrix algebra, whilst irrational fluxes give an infinite-dimensional irrational rotation algebra or noncommutative torus, a standard example in noncommutative geometry.

Any $*$-representation $\rho$ of $\mathcal{A}_1$ on a Hilbert space $\mathcal{H}_\rho$ can be induced to a $*$-representation $\pi_\rho$ of the twisted crossed product on $\mathcal{H} = L^2(G, \mathcal{H}_\rho)$ by setting

$$(\pi_\rho(f)\psi)(x) = \int_G \lambda(x, y-x)^{-1} \rho(\alpha_{-x} f)(y-x)\psi(y)\,\mathrm{d}y \quad [10]$$

for $f \in \mathcal{A}$ and $\psi \in \mathcal{H}$. When $\mathcal{A}_1 = \mathrm{C}(\Omega)$, we may take $\rho$ to be a one-dimensional irreducible $*$-representations given by evaluating the function at a point $\omega \in \Omega$.

When $G = \boldsymbol{R}^2$, it is easy to construct a Fredholm module from $\pi_\rho$. The space $\mathcal{H}_2 = \mathcal{H} \otimes \boldsymbol{C}^2$ has actions $\pi_\rho$ of $\mathcal{A}$ on the first factor and of the Pauli spin matrices $\sigma_1, \sigma_2, \sigma_3$, on the second. It may be regarded as a graded module with grading operator $\sigma_3$, and

$$F = (x_1^2 + x_2^2)^{-1/2}(x_1\sigma_1 + x_2\sigma_2) \quad [11]$$

provides a Connes–Fredholm involution which anticommutes with $\sigma_3$. Detailed technical results of Connes show how to use the supertrace on $\mathcal{H}_2$ and the Dixmier trace to interpret the physically important quantities in this setting.

We now turn to the formal derivation of the key alternative expression for the conductivity. In the abstract algebraic setting, when $p \in \mathcal{A}$ is a projection in the domain of a derivation $\delta$ the derivative of $(1-p)p = 0$ gives

$$0 = \delta((1-p)p) = (1-p)\delta(p) - \delta(p)p \quad [12]$$

and then an easy calculation leads to

$$[p, [\delta p, p]] = 2p(\delta p)p - (\delta p)p^2 - p^2(\delta p) = -\delta p \quad [13]$$

In the identity for elements $a$, $b$, $c$, and $h \in \mathcal{A}$

$$\begin{aligned} &\tau([a, [b, c]]h) - \tau(c[[h, a], b]) \\ &\quad = \tau([a, [b, c]h]) + \tau([b, c[h, a]]) = 0 \end{aligned} \quad [14]$$

we set $a = c = p$ and $b = \delta p$ to obtain

$$\tau([p, [\delta p, p]]h) = \tau(p[[h, p], \delta p]) \quad [15]$$

Combining this with [12] when $\tau \circ \delta = 0$, one obtains

$$\begin{aligned} \tau(p\delta h) &= \tau(\delta(ph)) - \tau(\delta(p)h) \\ &= \tau([p, [\delta p, p]]h) = \tau(p[[h, p], \delta p]) \end{aligned} \quad [16]$$

## The Hall Conductivity and Anderson Localisation

Substituting $p = P_\mathrm{F}$ and $h = H$ in formula [16] would give the current $\mathrm{tr}(P_\mathrm{F}[[H, P_\mathrm{F}], \delta P_\mathrm{F}])$. Since $\delta_k$ is proportional to the commutator with $X_k$, it is true that $\mathrm{tr} \circ \delta_k = 0$, but, unfortunately, $P_\mathrm{F}$ need not lie in the domain of $\delta_k$, and $H$ is unbounded, further compounding the difficulties. These are serious problems, although the situation is not quite as bad as it seems. Without the electrostatic term $e\boldsymbol{E} \cdot \boldsymbol{X}$ in $H, P_\mathrm{F}$ would have been a spectral projection with which $H$ would commute, so that

$$[H, P_\mathrm{F}] = e[\boldsymbol{E} \cdot \boldsymbol{X}, P_\mathrm{F}] = eE_j[X_j, P_\mathrm{F}] = \mathrm{i}eE_j\partial_j P_\mathrm{F} \quad [17]$$

and $H$ disappears from the formula, to be replaced by $\partial_j P_\mathrm{F}$. This would give the expected current $\mathrm{i}(e^2/\hbar)\mathrm{tr}(P_\mathrm{F}[\partial_j P_\mathrm{F}, \partial_k P_\mathrm{F}])E_j$, and the conductivity matrix

$$\sigma_{kj} = \mathrm{i}(e^2/\hbar)\mathrm{tr}(P_\mathrm{F}[\partial_j P_\mathrm{F}, \partial_k P_\mathrm{F}]) \quad [18]$$

given earlier (there is no need to scale by the thickness in two dimensions).

However it is derived, this expression for the conductivity only makes sense under suitable conditions, otherwise $\mathrm{tr}(P_F[\partial_j P_F, \partial_k P_F])$ might either be undefined (because $P_F$ is not differentiable) or might not be trace class. There is a simple condition sufficient to handle both these difficulties, which also leads to an interesting physical insight. From the obvious inequality

$$0 \le \mathrm{tr}\left[P_F(\partial_1 P_F \pm \mathrm{i}\partial_2 P_F)^*(\partial_1 P_F \pm \mathrm{i}\partial_2 P_F)\right] \quad [19]$$

$$= \mathrm{tr}\left[P_F\left((\partial_1 P_F)^2 + (\partial_2 P_F)^2\right)\right] \pm \mathrm{i}\,\mathrm{tr}(P_F[\partial_1 P_F, \partial_2 P_F]) \quad [20]$$

and the fact that $1 \ge P_F$, we deduce that

$$\mathrm{tr}\left[\left((\partial_1 P_F)^2 + (\partial_2 P_F)^2\right)\right] \ge \mathrm{tr}\left[P_F\left((\partial_1 P_F)^2 + (\partial_2 P_F)^2\right)\right] \ge |\mathrm{tr}(P_F[\partial_1 P_F, \partial_2 P_F])| \quad [21]$$

Thus, if $\mathrm{tr}[((\partial_1 P_F)^2 + (\partial_2 P_F)^2)]$ exists and is finite, then our expression for the conductivity is well defined. Mathematically, this is a Sobolev type condition. To see the physical significance, we recall that $\partial_k P_F = -\mathrm{i}[X_k, P_F]$, so that the condition is equivalent to the finiteness of $\mathrm{tr}[(X_1^2 + X_2^2)P_F{}^2] - \mathrm{tr}[(X_1 P_F)^2 + (X_2 P_F)^2]$.

This condition imposes a requirement for some localization in the system (when $P_F$ is a rank-1 projection, it reduces to the requirement that the variance $\langle X_1^2 + X_2^2\rangle - \langle X_1\rangle^2 - \langle X_2\rangle^2$ be finite). This links with a much older observation of Anderson that the interference caused by impurities in a crystal, which cancel at long range, should, at smaller distances, cause localized clumping. The mathematical development of this idea by Pastur provides an appropriate tool for handling the conditions for the valdiity of the conductivity formula. The impurities generating Anderson localization are provided in this model by the random potential in the Hamiltonian. It also leads us to restrict attention to the dense subalgebra $\mathcal{A}_0$ of $f \in \mathcal{A}$, where $\tau[(\partial_1 f)^*(\partial_1 f) + (\partial_2 f)^*(\partial_2 f)] < \infty$.

## The Integral Quantum Hall Effect

Having identified the features of physical interest, we can return to the abstract algebraic description with conductivity $\mathrm{i}(e^2/\hbar)\tau(p[\partial_j p, \partial_k p])$. The key observation is that this can be interpreted as the Connes pairing between a cyclic cocycle $c_\tau$ on $\mathcal{A}_0$ and the projection $p$ whose stable equivalence class represents an element of the $C^*$-algebraic $K$-theory, $K_0(\mathcal{A})$. Such pairings give noncommutative Chern characters. The cyclic cocycle is a trilinear form defined on elements $a_0, a_1, a_2 \in \mathcal{A}_0$ by

$$c_\tau(a_0, a_1, a_2) = \tau[a_0(\delta_1 a_1 \delta_2 a_2 - \delta_2 a_1 \delta_1 a_2)] \quad [22]$$

This is easily shown to be cyclic, $c_\tau(a_0, a_1, a_2) = c_\tau(a_1, a_2, a_0)$, and to satisfy the cyclic 2-cocycle condition

$$c_\tau(a_0 a_1, a_2, a_3) - c_\tau(a_0, a_1 a_2, a_3) + c_\tau(a_0, a_1, a_2 a_3) - c_\tau(a_3 a_0, a_1, a_2) = 0 \quad [23]$$

The Hall conductivity $\sigma_{21} = \mathrm{i}c_\tau(p, p, p)e^2/\hbar$ can now be interpreted as the noncommut ative Chern character defined by the projection $p$.

This interpretation of the Hall conductivity clears the way to prove that it is integral, and there are several different routes to this.

One approach is to identify the conductivity with some kind of index which is clearly integral. Bellissard worked with the Fredholm module where, by results of Connes, the Chern character is interpreted as the index of the Fredholm operator $\pi_\rho(p)F\pi_\rho(p)$. Avron, Seiler and Simon have interpreted the conductivity as a relative index $\dim[\ker(P_F - Q_F - 1)] - \dim[\ker(Q_F - P_F - 1)]$ of the projections $P_F$ and its conjugate $Q_F = uP_F u^*$ by an off-diagonal element $u$ of $F$. This is particularly interesting as the conjugation by $u$ can be interpreted as a nonsingular gauge transformation of exactly the kind introduced by Laughlin in his original explanation of the quantum Hall effect in terms of singular flux tubes piercing a cylindrical conductor.

Xia suggested another approach rewriting $\mathcal{A}$ as a repeated crossed product with $\boldsymbol{R}$, which allows us to calculate $K_0(\mathcal{A})$, using either Connes' Thom isomorphism theorem or the Takai duality theorem for stable algebras to get

$$K_0(\mathcal{A}) = K_0\left[C(\mathcal{A}_1, G, \lambda)\right] \cong K_0(\mathcal{A}_1) \quad [24]$$

which, when $\mathcal{A}_1 = C(\Omega)$, is just $K^0(\Omega)$, leading to identification as a topological index. For the simplest case of $\Omega = \boldsymbol{T}^2$, this gives $K^0(\Omega) \cong \boldsymbol{Z}^2$. The image of $\tau$, and so also $c_\tau$, actually sits in just one component, leading to quantization of the Hall conductivity.

The two questions posed in the introduction can now be answered as follows: The Hall conductivity can be identified with a topological index which can take only integer values, and therefore does not respond to continuous changes in any of the physical parameters until the change brings the system into a region where one of the background assumptions fails, such as a breakdown in the localization condition. The same conditions also ensure that the direct current vanishes. Roughly speaking, the

plateaus occur when the Fermi energy is in a gap in the extended (nonlocalized) spectrum.

This brief overview has omitted many of the interesting features of the detailed theory, which can be found in the surveys, such as the fact that low-lying energy levels do not contribute to the conductivity, and Shubin's theorem identifying $\tau(p)$ as the integrated density of states. Harper's equation describing a discrete lattice analog of the IQHE has been a test-bed for many of the ideas, and various results were first proved in that setting. The FQHE was discovered during an unsuccessful search for a Wigner crystal phase transition, but analysis of discrete models provides strong evidence that Hall conductors have very complicated phase diagrams.

## The Fractional Quantum Hall Effect

As mentioned in the introduction, by the time IQHE had been understood theoretically, it had been found that, with appropriate care, fractional conductivities could also be observed, although they were much less precise and stable than the integer values, and the plateaus less pronounced. Although there have been many phenomenological explanations, there is as yet no mathematical understanding from quantum field theory as compelling as that for the integer effect. We shall briefly summarize some of the main lines of attack.

The first explanation, again due to Laughlin, has also provided the basis for many subsequent treatments of the problem. The wave functions of the oscillator-like Landau Hamiltonian can conveniently be represented in the Bargmann–Segal Fock space of holomorphic functions $f$ on $\boldsymbol{R}^2 \sim \boldsymbol{C}$ which are square-integrable with respect to a Gaussian measure. Incorporating the measure into the functions, these have the form $f(z)\exp(-|z|^2/2)$. Many particle wave functions are similarly realized in terms of holomorphic functions on $\boldsymbol{C}^N$, and must be antisymmetric under odd permutations of the particles to describe fermions. This quickly leads one to consider functions of the form

$$\prod_{r<s}(z_r - z_s)^k \exp\left(-\sum_j |z_j|^2/2\right) \qquad [25]$$

for odd integers $k > 0$, and their multiples by even holomorphic functions. The lowest energy where such a wave function occurs is when $k = 1$, and larger values of $k$ have the effect of dividing the Hall conductivity by $k$, which produces fractional conductivities.

Halperin suggested quite early that counterflowing currents in the interior of a sample would tend to cancel, so that most of the current would be carried near the edge of the sample. There are several mathematical derivations of this, by, for example, Macris, Martin, and Pulé, and by Fröhlich, Graf, and Walcher. The $K$-theory of the boundary and bulk of a sample can be linked by exact sequences such as those of the commutative theory (Kellendonk *et al.* 2000), and even in the IQHE boundary and bulk conductivities can be used (Schulz-Baldes *et al.* 2002).

It has been fairly clear that whilst the IQHE can already be understood in terms of the motion of a single electron, the fractional effect is a many-body cooperative effect. One attempt to simplify the description is to work with an incompressible quantum fluid, and for edge currents one should study the boundary theory of such a fluid, in which the dominant contribution to the action is a Chern–Simons term, with conductivity as a coefficient. For an annular sample, this leads, in a suitable limit, to a chiral Luttinger model on the boundary circles, which can then be tackled mathematically using the representation theory of loop groups. This leads to some elegant mathematics, including extensions to multiple coupled bands, with conductivities described by Cartan matrices, as explained in the International Congress of Mathematicians (ICM) survey (Fröhlich 1995), and in the review by Fröhlich and Studer (1993).

The theory of composite fermions provides another physical approach in which field-theoretic effects result in the electrons sharing their charges in such a way as to produce fractional charges, and there is experimental evidence of such fractional charges in studies of tunneling from one edge to another. Then the FQHE is easily understood by simply replacing the electron charge $e$ by $e/k$ in the appropriate formulas.

Susskind has suggested combining noncommutative geometry with the theory of incompressible quantum fluids, an idea taken up by Polychronakos (2001). There are intriguing mathematical parallels with work by Berest and Wilson on ideals in the Weyl algebra and the Calogero–Moser model.

## Further Developments

Bellissard and others have extended the use of noncommutative geometrical methods into other parts of solid-state theory, where they clarify a number of the physical ideas. This is particularly useful in the case of quasicrystals, which are not easily handled by the conventional methods (Bellissard *et al.* 2000). Some ideas in string theory resemble higher-dimensional analogs, and higher-dimensional versions of the quantum Hall effect have also been studied by Hu and Zhang.

Finally, we conclude with some mathematical extensions of the theory. We have seen that, for periodic systems, the noncommutative Brillouin

zone can be a noncommutative torus, and it is possible to consider noncommutative versions of Riemann surfaces of higher genera. Carey *et al.* (1998) studied the effect in a noncommutative hyperbolic geometry with a discrete group action, generalizing the action of a Fuchsian group on the unit disc. This provides a tractable example in which one has an edge (albeit rather different from the normal physical situations) and also examples of a Hall effect in higher-genus noncommutative Riemann surfaces closely related to those of Klimek and Lesznewski. Natsumé and Nest have subsequently shown that these are deformation quantizations of the commutative Riemann surface theory in the sense of Rieffel. Coverings of noncommutative Riemann surfaces, which might provide an analoge of composite fermions, have been investigated by Marcolli and Mathai (1999, 2001).

*See also:* $C^*$-Algebras and Their Classification; Chern–Simons Models: Rigorous Results; Fractional Quantum Hall Effect; Hopf Algebras and *q*-Deformation Quantum Groups; Localization for Quasiperiodic Potentials; Noncommutative Geometry and the Standard Model; Noncommutative Tori, Yang–Mills, and String Theory; Schrödinger Operators.

## Further Reading


Bellissard J (1986) *K*-theory of $C^*$-algebras, in solid state physics. In: Dorlas T, Hugenholtz NM, and Winnink M (eds.) *Statistical Mechanics and Field Theory: Mathematical Aspects*, Springer Lecture Notes in Physics vol. 257, pp. 99–156. Berlin: Springer.

Bellissard J, Herrmann DJL, and Zarrouati M (2000) Hulls of a periodic solids and gap labelling theorem. In: Baake M and Moody RV (eds.) *Directions in Mathematical Quasi-Crystals*, CRM Monograph Series vol. 13, pp. 207–258. Providence, RI: American Mathematical Society.

Bellissard J, van Elst A, and Schulz-Baldes H (1994) The noncommutative geometry of the quantum Hall effect. *Journal of Mathematical Physics* 35: 5373–5457.

Carey AL, Hannabuss KC, Mathai V, and McCann P (1998) Quantum Hall effect on the hyperbolic plane. *Communications in Mathematical Physics* 190: 629–673.

Connes A (1986) Non-commutative differential geometry. *Publications of the Institut des Hautes Etudes Scientifiques* 62: 257–360.

Connes A (1994) *Non Commutative Geometry*. San Diego: Academic Press.

Fröhlich J (1995) *The Fractional Quantum Hall Effect, Chern–Simons Theory and Integral Lattices*. Proceedings of the International Congress of Mathematicians, 1994, Zürich, 75–105. Basel: Birkhäuser Verlag.

Fröhlich J and Studer UM (1993) Gauge invariance and current algebra in nonrelativistic many-body theory. *Reviews of Modern Physics* 65: 733–802.

Kellendonk J, Richter T, and Schulz-Baldes M (2002) Edge current channels and Chern numbers in the integral quantum Hall effect. *Reviews in Mathematical Physics* 14: 87–119.

Marcolli M and Mathai V (1999) Twisted index theory on good orbifolds. I. Non-commutative Bloch theory. *Communications in Contemporary Mathematics* 1: 553–587.

Marcolli M and Mathai V (2001) Twisted index theory on good orbifolds. II. Fractional quantum numbers. *Communications in Mathematical Physics* 217: 55–87.

McCann PJ (1998) Geometry and the integer quantum Hall effect. In: Carey AL and Murray MK (eds.) *Geometric Analysis and Lie Theory in Mathematics and Physics*, pp. 122–208. Australian Mathematical Society Lecture Series. Cambridge: Cambridge University Press.

Polychronakos A (2001) Quantum Hall states as matrix Chern–Simons theory. *The Journal of High Energy Physics* 4(paper 11): 20.

Schulz-Baldes M, Kellendonk J, and Richter T (2000) Simultaneous quantization of edge and bulk Hall conductivity. *Journal of Physics A* 33: L27–L32.


# Quantum Mechanical Scattering Theory

**D R Yafaev**, Université de Rennes, Rennes, France



## Introduction

Scattering theory is concerned with the study of the large-time behavior of solutions of the time-dependent Schrödinger equation [1] for a system with a Hamiltonian $H$:

$$i\partial u/\partial t = Hu, \quad u(0) = f \qquad [1]$$

Being a part of the perturbation theory, scattering theory describes the asymptotics of $u(t)$ as $t \to +\infty$ or $t \to -\infty$ in terms of solutions of the Schrödinger equation for a "free" system with a Hamiltonian $H_0$. Of course, eqn [1] has a unique solution $u(t) = \exp(-iHt)f$, while the solution of the same equation with the operator $H_0$ and the initial data $u_0(0) = f_0$ is given by the formula $u_0(t) = \exp(-iH_0 t)f_0$. From the viewpoint of scattering theory, the function $u(t)$ has free asymptotics as $t \to \pm\infty$ if for appropriate initial data $f_0^{\pm}$ eqn [2] holds:

$$\lim_{t\to\pm\infty} \|u(t) - u_0^{\pm}(t)\| = 0 \qquad [2]$$

Here and throughout this article a relation containing the signs "±" is understood as two independent equalities. We emphasize that initial data $f_0^{\pm}$ are different for $t \to +\infty$ and $t \to -\infty$ and

$u_0^\pm(t) = \exp(-iH_0t)f_0^\pm$. Equation [2] leads to a connection between the corresponding initial data $f_0^\pm$ and $f$ given by

$$f = \lim_{t\to\pm\infty} \exp(iHt)\exp(-iH_0t)f_0^\pm \qquad [3]$$

If $f$ is an eigenvector of $H$, that is, $Hf = \lambda f$, then obviously $u(t) = e^{-i\lambda t}f$. On the contrary, if $f$ belongs to the (absolutely) continuous subspace of $H$, then necessarily $u(t)$ has the free asymptotics as $t \to \pm\infty$. This result is known as asymptotic completeness.

The Schrödinger operator $H = -\Delta + V(x)$ in the space $\mathcal{H} = L_2(\mathbb{R}^d)$ with a real potential $V$ decaying at infinity is a typical Hamiltonian of scattering theory. The operator $H$ describes a particle in an external potential $V$ or two interacting particles. Asymptotically (as $t \to +\infty$ or $t \to -\infty$), particles may either form a bound state or be free (a scattering state). Of course, a bound (scattering) state at $-\infty$ remains the same at $+\infty$. To be more precise, suppose that

$$|V(x)| \le C(1 + |x|)^{-\rho} \qquad [4]$$

where $\rho > 1$. Then relation [2] can be justified with the kinetic energy operator $H_0 = -\Delta$ playing the role of the unperturbed operator.

As discussed in Landau and Lifshitz (1965) (see also Amrein *et al.* (1977), Pearson (1988), and Yafaev (2000)), in scattering experiments one sends a beam of particles of energy $\lambda > 0$ in a direction $\omega$. Such a beam is described by the plane wave

$$\psi_0(x;\omega,\lambda) = \exp(ik\langle\omega, x\rangle), \quad \lambda = k^2 > 0$$

(which satisfies of course the free equation $-\Delta\psi_0 = \lambda\psi_0$). The scattered particles are described for large distances by the outgoing spherical wave

$$a(\hat{x},\omega;\lambda)|x|^{-(d-1)/2}\exp(ik|x|)$$

Here $\hat{x} = x|x|^{-1}$ is the direction of observation and the coefficient $a(\hat{x},\omega;\lambda)$ is known as the scattering amplitude. This means that quantum particles subject to a potential $V(x)$ are described by the solution $\psi$ of eqn [5] with asymptotics [6] at infinity:

$$-\Delta\psi + V(x)\psi = \lambda\psi \qquad [5]$$

$$\psi(x;\omega,\lambda) = \exp(ik\langle\omega,x\rangle) + a(\hat{x},\omega;\lambda)|x|^{-(d-1)/2}\exp(ik|x|) + o\left(|x|^{-(d-1)/2}\right) \qquad [6]$$

The existence of such solutions requires of course a proof. The differential scattering crosssection defined by eqn [7] gives us the part of particles scattered in a solid angle $d\hat{x}$:

$$d\sigma(\hat{x},\omega;\lambda) = |a(\hat{x},\omega;\lambda)|^2 d\hat{x} \qquad [7]$$

As discussed below, the temporal asymptotics of solutions of the time-dependent Schrödinger equation [1] are closely related to the asymptotics at large distances of solutions of the stationary Schrödinger equation [5].

## Time-Dependent Scattering Theory and Møller Operators

If $V(x) \to 0$ as $|x| \to \infty$, then the essential spectrum of the Schrödinger operator $H = -\Delta + V(x)$ covers the whole positive half-line, whereas the negative spectrum of $H$ consists of eigenvalues accumulating, perhaps, at the point zero only.

Scattering theory requires a more advanced classification of the spectrum based on measure theory. Consider a self-adjoint operator $H$ defined on domain $\mathcal{D}(H)$ in a Hilbert space $\mathcal{H}$. Let $E$ be its spectral family. Then the space $\mathcal{H}$ can be decomposed into the orthogonal sum of invariant subspaces $\mathcal{H}^{(p)}, \mathcal{H}^{(sc)}$ and $\mathcal{H}^{(ac)}$. The subspace $\mathcal{H}^{(p)}$ is spanned by eigenvectors of $H$ and the subspaces $\mathcal{H}^{(sc)}, \mathcal{H}^{(ac)}$ are distinguished by the condition that the measure $(E(X)f, f)$ (here $X \subset \mathbb{R}$ is a Borel set) is singularly or absolutely continuous with respect to the Lebesgue measure for all $f \in \mathcal{H}^{(sc)}$ or $f \in \mathcal{H}^{(ac)}$. Typically (in applications to quantum-mechanical problems) the singularly continuous part is absent, that is, $\mathcal{H}^{(sc)} = \{0\}$. We denote by $H^{(ac)}$ the restriction of $H$ on its absolutely continuous subspace $\mathcal{H}^{(ac)}$ and by $P^{(ac)}$ the orthogonal projection on this subspace. The same objects for the operator $H_0$ will be endowed with the index "0."

Equation [3] motivates the following fundamental definition. The wave, or Møller, operator $W_\pm = W_\pm(H, H_0)$ for a pair of self-adjoint operators $H_0$ and $H$ is defined by eqn [8] provided that the corresponding strong limit exists:

$$W_\pm = \operatorname*{s-lim}_{t\to\pm\infty} \exp(iHt)\exp(-iH_0t)P_0^{(ac)} \qquad [8]$$

The wave operator is isometric on $\mathcal{H}_0^{(ac)}$ and enjoys the intertwining property

$$HW_\pm = W_\pm H_0 \qquad [9]$$

Therefore, its range $\operatorname{Ran} W_\pm$ is contained in the absolutely continuous subspace $\mathcal{H}^{(ac)}$ of the operator $H$.

The operator $W_\pm(H, H_0)$ is said to be complete if eqn [10] holds:

$$\operatorname{Ran} W_\pm(H, H_0) = \mathcal{H}^{(ac)} \qquad [10]$$

It is easy to see that the completeness of $W_\pm(H, H_0)$ is equivalent to the existence of the "inverse" wave operator $W_\pm(H_0, H)$. Thus, if the wave operator $W_\pm(H, H_0)$ exists and is complete, then the operators $H_0^{(\mathrm{ac})}$ and $H^{(\mathrm{ac})}$ are unitarily equivalent. We emphasize that scattering theory studies not arbitrary unitary equivalence but only the "canonical" one realized by the wave operators.

Along with the wave operators an important role in scattering theory is played by the scattering operator defined by eqn [11] where $W_+^*$ is the operator adjoint to $W_+$:

$$S = S(H, H_0) = W_+^*(H, H_0)W_-(H, H_0) \qquad [11]$$

The operator $S$ commutes with $H_0$ and hence reduces to multiplication by the operator function $S(\lambda) = S(\lambda; H, H_0)$ in a representation of $\mathcal{H}_0^{(\mathrm{ac})}$ which is diagonal for $H_0^{(\mathrm{ac})}$. The operator $S(\lambda)$ is known as the scattering matrix. The scattering operator [11] is unitary on the subspace $\mathcal{H}_0^{(\mathrm{ac})}$ provided the wave operators $W_\pm(H, H_0)$ exist and are complete. The scattering operator $S(H, H_0)$ connects the asymptotics of the solutions of eqn [1] as $t \to -\infty$ and as $t \to +\infty$ in terms of the free problem, that is $S(H, H_0) : f_0^- \mapsto f_0^+$, where $f_0^\pm$ are the same as in eqn [2]. The scattering operator and the scattering matrix are usually of great interest in mathematical physics problems, because they connect the "initial" and the "final" characteristics of the process directly, bypassing its consideration for finite times.

The definition of the wave operators can be extended to self-adjoint operators acting in different spaces. Let $H_0$ and $H$ be self-adjoint operators in Hilbert spaces $\mathcal{H}_0$ and $\mathcal{H}$, respectively, and let "identification" $J : \mathcal{H}_0 \to \mathcal{H}$ be a bounded operator. Then the wave operator $W_\pm = W_\pm(H, H_0; J)$ for the triple $H_0, H$, and $J$ is defined by eqn [12] provided again that the strong limit there exists:

$$W_\pm = \operatorname*{s-lim}_{t\to\pm\infty} \exp(\mathrm{i}Ht)J\exp(-\mathrm{i}H_0t)P_0^{(\mathrm{ac})} \qquad [12]$$

Intertwining property [9] is preserved for wave operator [12]. This operator is isometric on $\mathcal{H}_0^{(\mathrm{ac})}$ if and only if

$$\lim_{t\to\pm\infty} \|J\exp(-\mathrm{i}H_0t)f_0\| = \|f_0\|$$

for all $f_0 \in \mathcal{H}_0^{(\mathrm{ac})}$. Since

$$\operatorname*{s-lim}_{|t|\to\infty} K\exp(-\mathrm{i}H_0t)P_0^{(\mathrm{ac})} = 0$$

for a compact operator $K$, wave operators [12] corresponding to identifications $J_1$ and $J_2$ coincide if $J_2 - J_1$ is compact or, at least, the operators $(J_2 - J_1)$ $E_0(X)$ are compact for all bounded intervals $X$. Consideration of wave operators [12] with $J \neq I$ may of course be of interest also in the case $\mathcal{H}_0 = \mathcal{H}$.

It suffices to verify the existence of limits [8] or [12] on some set dense in the absolutely continuous subspace $\mathcal{H}_0^{(\mathrm{ac})}$ of the operator $H_0$. The following simple but convenient condition for the existence of wave operators is usually called Cook's criterion. Suppose that $H_0 = H_0^{(\mathrm{ac})}$ and that the operator $J$ maps domain $\mathcal{D}(H_0)$ of the operator $H_0$ into $\mathcal{D}(H)$. Let

$$\int_0^{\pm\infty} \|(HJ - JH_0)\exp(-\mathrm{i}H_0t)f\|\mathrm{d}t < \infty$$

for all $f$ from some set $D_0 \subset \mathcal{D}(H_0)$ dense in $\mathcal{H}_0$. Then the wave operator $W_\pm(H, H_0; J)$ exists.

This result is often useful in applications since the operator $\exp(-\mathrm{i}H_0t)$ is known explicitly. For example, it works with $J = I$ for the pair

$$H_0 = -\Delta, \qquad H = H_0 + V(x) \qquad [13]$$

if $V(x)$ satisfies estimate [4] with $\rho > 1$. On the other hand, different proofs of the existence of the wave operators $W_\pm(H_0, H; J^*)$ require new mathematical tools. There are two essentially different approaches in scattering theory: the trace-class and smooth methods.

## Time-Independent Scattering Theory

The approach in scattering theory relying on definition [8] is called time dependent. An alternative possibility is to change the definition of wave operators replacing the unitary groups by the corresponding resolvents $R_0(z) = (H_0 - z)^{-1}$ and $R(z) = (H - z)^{-1}$. They are related by a simple identity

$$\begin{aligned} R(z) &= R_0(z) - R_0(z)VR(z) \\ &= R_0(z) - R(z)VR_0(z) \end{aligned} \qquad [14]$$

where $V = H - H_0$ and $\operatorname{Im} z \neq 0$. In the stationary approach in place of limits [8] one has to study the boundary values (in a suitable topology) of the resolvents as the spectral parameter $z$ tends to the real axis. An important advantage of the stationary approach is that it gives convenient formulas for the wave operators and the scattering matrix.

Let us discuss here the stationary formulation of the scattering problem for operators [13] in the Hilbert space $\mathcal{H} = L_2(\mathbb{R}^d)$ in terms of solutions of the Schrödinger equation [5]. If $V(x)$ satisfies estimate [4] with $\rho > (d+1)/2$, then for all $\lambda > 0$ and all unit vectors $\omega \in \mathbb{S}^{d-1}$, eqn [5] has the solution $\psi(x; \omega, \lambda)$ with asymptotics [6] as $|x| \to \infty$. Moreover, the scattering amplitude $a(\hat{x}, \omega; \lambda)$ belongs to the space

$L_2(\mathbb{S}^{d-1})$ in the variable $\hat{x}$ uniformly in $\omega \in \mathbb{S}^{d-1}$, and it can be expressed via $\psi(x;\omega,\lambda)$ by the formula

$$a(\theta,\omega;\lambda) = -\gamma_d(\lambda)\int_{\mathbb{R}^d} e^{-ik\langle\theta,x\rangle}V(x)\psi(x;\omega,\lambda)\,dx$$

where

$$\gamma_d(\lambda) = e^{-\pi i(d-3)/4}2^{-1}(2\pi)^{-(d-1)/2}\lambda^{(d-3)/4}$$

Let us define two sets of scattering solutions, or eigenfunctions of the continuous spectrum, by the formulas

$$\psi_-(x;\omega,\lambda) = \psi(x;\omega,\lambda) \text{ and } \psi_+(x;\omega,\lambda) = \overline{\psi(x;-\omega,\lambda)}$$

In terms of boundary values of the resolvent, the functions $\psi_\pm(\omega,\lambda)$ can be constructed by the formula

$$\psi_\pm(\omega,\lambda) = \psi_0(\omega,\lambda) - R(\lambda \mp i0)V\psi_0(\omega,\lambda) \qquad [15]$$

Obviously, functions [15] satisfy eqn [5]. Using resolvent identity [14], it is easy to derive the Lippmann–Schwinger equation

$$\psi_\pm(\omega,\lambda) = \psi_0(\omega,\lambda) - R_0(\lambda \mp i0)V\psi_\pm(\omega,\lambda)$$

for $\psi_\pm(\omega,\lambda)$. Asymptotics [6] can be deduced from the formula

$$(R_0(\lambda \pm i0)f)(x) = c_\pm(\lambda)(\Gamma_0(\lambda)f)(\pm\hat{x})|x|^{-(d-1)/2} \times \exp(\pm ik|x|) + O(|x|^{-(d+1)/2})$$

where $f \in C_0^\infty(\mathbb{R}^d)$, $c_\pm(\lambda) = \pi^{1/2}\lambda^{-1/4}e^{\mp i\pi(d-3)/4}$ and the operator $\Gamma_0(\lambda)$ defined by eqn [16] is (up to the numerical factor) the restriction of the Fourier transform $\hat{f} = \mathcal{F}_0 f$ onto the sphere of radius $\lambda^{1/2}$:

$$(\Gamma_0(\lambda)f)(\omega) = 2^{-1/2}\lambda^{(d-2)/4}\hat{f}(\lambda^{1/2}\omega), \quad \omega \in \mathbb{S}^{d-1} \qquad [16]$$

The wave operators $W_\pm(H,H_0)$ can be constructed in terms of the solutions $\psi_\pm$. Set $\xi = \lambda^{1/2}\omega$ ($\xi$ is the momentum variable), write $\psi_\pm(x,\xi)$ instead of $\psi_\pm(x;\omega,\lambda)$, and consider two transformations

$$(\mathcal{F}_\pm f)(\xi) = (2\pi)^{-d/2}\int_{\mathbb{R}^d}\overline{\psi_\pm(x,\xi)}f(x)\,dx \qquad [17]$$

(defined initially, e.g., on the Schwartz class $\mathcal{S}(\mathbb{R}^d)$) of the space $L_2(\mathbb{R}^d)$ into itself. The operators $\mathcal{F}_\pm$ can be regarded as generalized Fourier transforms, and both of them coincide with the usual Fourier transform $\mathcal{F}_0$ if $V=0$. It follows from eqns [5], [17] that under the action of $\mathcal{F}_\pm$ the operator $H$ goes over into multiplication by $|\xi|^2$, that is,

$$(\mathcal{F}_\pm Hf)(\xi) = |\xi|^2(\mathcal{F}_\pm f)(\xi)$$

Moreover, with the help of eqn [15], it can be shown that $\mathcal{F}_\pm$ is an isometry on $\mathcal{H}^{(ac)}$, it is zero on $\mathcal{H} \ominus \mathcal{H}^{(ac)}$, and its range Ran $\mathcal{F}_\pm = L_2(\mathbb{R}^d)$. This is equivalent to eqns [18]:

$$\mathcal{F}_\pm^*\mathcal{F}_\pm = P^{(ac)}, \qquad \mathcal{F}_\pm\mathcal{F}_\pm^* = I \qquad [18]$$

Hence any function $f \in \mathcal{H}^{(ac)}$ admits the expansion in the generalized Fourier integral

$$f(x) = (2\pi)^{-d/2}\int_{\mathbb{R}^d}\psi_\pm(x,\xi)(\mathcal{F}_\pm f)(\xi)\,d\xi$$

It can also be deduced from eqn [6] that the vector

$$(\mathcal{F}_\pm^* - \mathcal{F}_0^*)\exp(-i|\xi|^2t)\hat{f}$$

tends to zero as $t \to \pm\infty$ for an arbitrary $\hat{f} \in L_2(\mathbb{R}^d)$. This implies the existence of the wave operators $W_\pm = W_\pm(H,H_0)$ for pair [13] and gives the representation

$$W_\pm = \mathcal{F}_\pm^*\mathcal{F}_0 \qquad [19]$$

Completeness of $W_\pm$ follows from eqn [19] and the first equation in [18]. The second equality in [18] is equivalent to the isometricity of $W_\pm$. Formula [19] is an example of a stationary representation for the wave operator. It formally implies that

$$W_\pm\psi_0(\omega,\lambda) = \psi_\pm(\omega,\lambda)$$

which means that each wave operator establishes a one-to-one correspondence between eigenfunctions of the continuous spectrum of the operators $H_0$ and $H$.

The main ideas of the stationary approach go back to Friedrichs (1965), and Povzner. The inverse problem of reconstruction of a potential $V$ given the scattering amplitude $a$ (see eqn [6]) is treated in Faddeev (1976).

## The Trace-Class Method

Recall that the class $\mathfrak{S}_p, p \geq 1$, consists of compact operators $T$ such that the norm

$$\|T\|_p = \left(\sum_n \lambda_n^p(|T|)\right)^{1/p}, \quad |T| = (T^*T)^{1/2}$$

is finite. Eigenvalues $\lambda_n(|T|) =: s_n(T)$ of a non-negative operator $|T|$ are called singular numbers of $T$. In particular, $\mathfrak{S}_1$ is the trace class and $\mathfrak{S}_2$ is the Hilbert–Schmidt class.

The trace-class method (see Reed and Simon (1976) or Yafaev (1992) for a detailed presentation) makes no assumptions about the "unperturbed" operator $H_0$. Its basic result is the following theorem of Kato and Rosenblum. If $V = H - H_0$ belongs to the trace class

$\mathfrak{S}_1$, then the wave operators $W_\pm(H, H_0)$ exist and are complete. In particular, the operators $H_0^{(ac)}$ and $H^{(ac)}$ are unitarily equivalent. This can be considered as a far advanced extension of the H Weyl theorem, which states the stability of the essential spectrum under compact perturbations.

The condition $V \in \mathfrak{S}_1$ in the Kato–Rosenblum theorem cannot be relaxed in the framework of operator ideals $\mathfrak{S}_p$. This follows from the Weyl–von Neumann–Kuroda theorem. Let $H_0$ be an arbitrary self-adjoint operator. For any $p > 1$ and any $\varepsilon > 0$ there exists a self-adjoint operator $V$ such that $V \in \mathfrak{S}_p$, $\|V\|_p < \varepsilon$ and the operator $H = H_0 + V$ has purely point spectrum. Of course, such an operator $H$ has no absolutely continuous part. At the same time, the operator $H_0$ may be absolutely continuous. In this case, the wave operators $W_\pm(H, H_0)$ do not exist.

Although sharp in the abstract framework, the Kato–Rosenblum theorem cannot directly be applied to the theory of differential operators where a perturbation is usually an operator of multiplication and hence is not even compact. We mention its two generalizations applicable to this theory. The first, the Birman–Kato–Kreĭn theorem, claims that the wave operators $W_\pm(H, H_0)$ exist and are complete provided

$$R^n(z) - R_0^n(z) \in \mathfrak{S}_1$$

for some $n = 1, 2, \ldots$ and all $z$ with $\operatorname{Im} z \neq 0$. The second, the Birman theorem, asserts that the same is true if $\mathcal{D}(H) = \mathcal{D}(H_0)$ or $\mathcal{D}(|H|^{1/2}) = \mathcal{D}(|H_0|^{1/2})$ and

$$E(X)(H - H_0)E_0(X) \in \mathfrak{S}_1$$

for all bounded intervals $X$.

The wave operators enjoy the following property known as the Birman invariance principle. Suppose that $\varphi(H) - \varphi(H_0) \in \mathfrak{S}_1$ for a real function $\varphi$ such that its derivative $\varphi'$ is absolutely continuous and $\varphi'(\lambda) > 0$. Then the wave operators $W_\pm(H, H_0)$ exist and eqn [20] holds:

$$W_\pm(H, H_0) = W_\pm(\varphi(H), \varphi(H_0)) \qquad [20]$$

A direct generalization of the Kato–Rosenblum theorem to the operators acting in different spaces is due to Pearson. Suppose that $H_0$ and $H$ are self-adjoint operators in spaces $\mathcal{H}_0$ and $\mathcal{H}$, respectively, $J : \mathcal{H}_0 \to \mathcal{H}$ is a bounded operator and $V = HJ - JH_0 \in \mathfrak{S}_1$. Then the wave operators $W_\pm(H, H_0; J)$ and $W_\pm(H_0, H; J^*)$ exist.

Although rather sophisticated, the proof relies only on the following elementary lemma of Rosenblum. For a self-adjoint operator $H$, consider the set $\mathfrak{R} \subset \mathcal{H}^{(ac)}$ of elements $f$ such that

$$r_H^2(f) := \operatorname{ess\,sup}\, \mathrm{d}(E(\lambda)f, f)/\mathrm{d}\lambda < \infty$$

If $K : \mathcal{H} \to \mathcal{G}$ ($\mathcal{G}$ is some Hilbert space) is a Hilbert–Schmidt operator, then for all $f \in \mathfrak{R}$

$$\int_{-\infty}^{\infty} \|K \exp(-\mathrm{i}Ht)f\|^2 \,\mathrm{d}t \leq 2\pi r_H^2(f)\|K\|_2^2 \qquad [21]$$

Moreover, the set $\mathfrak{R}$ is dense in $\mathcal{H}^{(ac)}$.

The Pearson theorem allows to simplify considerably the original proofs of different generalizations of the Kato–Rosenblum theorem.

A typical application of the trace-class theory is the following result. Suppose that

$$\mathcal{H} = L_2(\mathbb{R}^d),\ H_0 = -\Delta + V_0(x),\ H = H_0 + V(x) \qquad [22]$$

where the functions $V_0$ and $V$ are real, $V_0 \in L_\infty(\mathbb{R}^d)$ and $V$ satisfies estimate [4] for some $\rho > d$. Then the wave operators $W_\pm(H, H_0)$ exist and are complete.

## The Smooth Method

The smooth method (see Kuroda (1978), Reed and Simon (1979), or Yafaev (1992), for a detailed presentation) relies on a certain regularity of the perturbation in the spectral representation of the operator $H_0$. There are different ways to understand regularity. For example, in the Friedrichs–Faddeev model $H_0$ acts as multiplication by independent variable in the space $\mathcal{H} = L_2(\Lambda; \mathfrak{N})$, where $\Lambda$ is an interval and $\mathfrak{N}$ is an auxiliary Hilbert space. The perturbation $V$ is an integral operator with sufficiently smooth kernel.

Another possibility is to use the concept of $H$-smoothness introduced by Kato. An $H$-bounded operator $K$ is called $H$-smooth if, for all $f \in \mathcal{D}(H)$,

$$\int_{-\infty}^{\infty} \|K \exp(-\mathrm{i}Ht)f\|^2 \,\mathrm{d}t \leq C\|f\|^2 \qquad [23]$$

(cf. eqns [21] and [23]). Here and below, $C$ are different positive numbers whose precise values are inessential. It is important that this definition admits equivalent reformulations in terms of the resolvent or of the spectral family. Thus, $K$ is $H$-smooth if and only if

$$\sup_{\lambda \in \mathbb{R}, \varepsilon > 0} \|K(R(\lambda + \mathrm{i}\varepsilon) - R(\lambda - \mathrm{i}\varepsilon))K^*\| < \infty$$

or if and only if

$$\sup |X|^{-1} \|KE(X)\|^2 < \infty$$

for all intervals $X \subset \mathbb{R}$.

In applications the assumption of $H$-smoothness of an operator $K$ imposes too stringent conditions on the operator $H$. In particular, the operator $H$ is necessarily absolutely continuous if kernel of $K$ is trivial. This assumption excludes eigenvalues and other singular points in the spectrum of $H$, for

example, the bottom of the continuous spectrum for the Schrödinger operator with decaying potential or edges of bands if the spectrum has the band structure. The notion of local $H$-smoothness suggested by Lavine is considerably more flexible. By definition, $K$ is called $H$-smooth on a Borel set $X \subset \mathbb{R}$ if the operator $KE(X)$ is $H$-smooth. Note that, under the assumption

$$\sup_{\lambda \in X, \varepsilon > 0} \| K(R(\lambda + \mathrm{i}\varepsilon) - R(\lambda - \mathrm{i}\varepsilon))K^* \| < \infty \qquad [24]$$

the operator $K$ is $H$-smooth on the closure of $X$.

The following Kato–Lavine theorem is simple but very useful. Suppose that

$$HJ - JH_0 = K^* K_0$$

where the operators $K_0$ and $K$ are $H_0$-smooth and $H$-smooth, respectively, on an arbitrary compact subinterval of some interval $\Lambda$. Then the wave operators

$$W_\pm(H, H_0; JE_0(\Lambda)) \text{ and } W_\pm(H_0, H; J^* E(\Lambda))$$

exist (and are adjoint to each other).

This result cannot usually be applied directly since the verification of $H_0$- and especially of $H$-smoothness may be a difficult problem. Let us briefly explain how it can be done on the example of pair [10], where the potential $V(x)$ satisfies estimate [4] for some $\rho > 1$. Let us start with the operator $H_0 = -\Delta$. Denote by $L_2^{(l)} = L_2^{(l)}(\mathbb{R}^d)$ the Hilbert space with the norm $\|f\|_l = \|\langle x \rangle^l f\|$, where $\langle x \rangle = (1 + |x|^2)^{1/2}$. Let the operator $\Gamma_0(\lambda)$ be defined by eqn [16], and let $X \subset (0, \infty)$ be some compact interval. Set $\mathfrak{N} = L_2(\mathbb{S}^{d-1})$. If $f \in L_2^{(l)}$ with $l > 1/2$, then, by the Sobolev trace theorem,

$$\begin{aligned} \|\Gamma_0(\lambda) f\|_{\mathfrak{N}} &\leq C \|f\|_l \\ \|\Gamma_0(\lambda) f - \Gamma_0(\lambda') f\|_{\mathfrak{N}} &\leq C |\lambda - \lambda'|^\alpha \|f\|_l \end{aligned} \qquad [25]$$

for an arbitrary $\alpha \leq l - 1/2, \alpha < 1$ and all $\lambda, \lambda' \in X$. These estimates imply that the function

$$(E_0(\lambda) f, f) = \int_{|\xi|^2 < \lambda} |\hat{f}(\xi)|^2 \, \mathrm{d}\xi \qquad [26]$$

is differentiable and the derivative

$$\mathrm{d}(E_0(\lambda) f, f)/\mathrm{d}\lambda = \|\Gamma_0(\lambda) f\|_{\mathfrak{N}}^2, \quad f \in L_2^{(l)}, \ l > 1/2$$

is Hölder-continuous in $\lambda > 0$ (uniformly in $f$, $\|f\|_l \leq 1$). Therefore, applying the Privalov theorem to the Cauchy integral

$$(R_0(z) f, f) = \int_0^\infty (\lambda - z)^{-1} \, \mathrm{d}(E_0(\lambda) f, f)$$

we obtain that the analytic operator function

$$\mathcal{R}_0(z) = \langle x \rangle^{-l} R_0(z) \langle x \rangle^{-l}, \quad l > 1/2$$

considered in the space $\mathcal{H}$, is continuous in norm in the closed complex plane $\mathbb{C}$ cut along $(0, \infty)$ with possible exception of the point $z = 0$. This implies $H_0$-smoothness of the operator $\langle x \rangle^{-l}, l > 1/2$, on all compact intervals $X \subset (0, \infty)$.

To obtain a similar result for the operator $H$, we proceed from the resolvent identity [14]. Let $\mathcal{R}(z) = \langle x \rangle^{-l} R(z) \langle x \rangle^{-l}$, and let $B$ be the operator of multiplication by the bounded function $(1 + |x|)^\rho V(x)$. If

$$f + \mathcal{R}_0(z) B f = 0$$

then $\psi = R_0(z) \langle x \rangle^{-l} B f$ satisfies the Schrödinger equation $H\psi = z\psi$. Since $H$ is self-adjoint, this implies that $\psi = 0$ and hence $f = 0$. Using eqn [14], we obtain that

$$\mathcal{R}(z) = (I + \mathcal{R}_0(z) B)^{-1} \mathcal{R}_0(z), \quad \operatorname{Im} z \neq 0 \qquad [27]$$

because the inverse operator here exists by the Fredholm alternative.

The operator function $(I + \mathcal{R}_0(z) B)^{-1}$ is analytic in the complex plane cut along $(0, \infty)$ with possible exception of poles (coinciding with eigenvalues of $H$) on the negative half-axis. Moreover, $(I + \mathcal{R}_0(z) B)^{-1}$ is continuous up to the cut except the set $\mathcal{N} \subset (0, \infty)$ of $\lambda$ where at least one of the homogeneous equations

$$f + \mathcal{R}_0(\lambda \pm \mathrm{i}0) B f = 0 \qquad [28]$$

has a nontrivial solution. It follows from eqn [27] that the same is true for the operator function $\mathcal{R}(z)$. It can be shown that the set $\mathcal{N}$ is closed and has the Lebesgue measure zero. Let $\Lambda = (0, \infty) \backslash \mathcal{N}$; then $\Lambda = \cup_n \Lambda_n$ where $\Lambda_n$ are disjoint open intervals. By condition [24], the operator $\langle x \rangle^{-l}, l > 1/2$, is $H$-smooth on any strictly interior subinterval of every $\Lambda_n$. Applying the Kato–Lavine theorem, we see that the wave operators $W_\pm(H, H_0; E_0(\Lambda_n))$ and $W_\pm(H_0, H; E(\Lambda_n))$ exist for all $n$. Since $E_0(\Lambda) = I$ and $E(\Lambda) = P^{(\mathrm{ac})}$, this implies the existence of $W_\pm(H, H_0)$ and $W_\pm(H_0, H)$. Thus, the wave operators $W_\pm(H, H_0)$ for pair [13] exist and are complete if estimate [4] holds for some $\rho > 1$.

Compared to the trace-class method, conditions on the perturbation $V(x)$ are less restrictive, while the class of admissible "free" problems is essentially more narrow (in eqn [22] $V_0(x)$ is an arbitrary bounded function). It is not known whether the wave operators $W_\pm(H, H_0)$ exist for all pairs [22] such that $V_0 \in L_\infty$ and $V$ satisfies [4] for some $\rho > 1$.

It is important that the smooth method allows one to prove the absence of the singular continuous spectrum. Note first that the continuity of $\mathcal{R}(z)$ implies that the operator $H$ is absolutely continuous on the subspace $E(\Lambda)\mathcal{H}$. Therefore, the singular

positive spectrum of $H$ is necessarily contained in $\mathcal{N}$. To prove that its continuous part is empty, it suffices to check that the set $\mathcal{N}$ consists of eigenvalues of the operator $H$. In terms of $u = \langle x \rangle^{-l} Bf, l = \rho/2$, eqn [28] can be rewritten as

$$u + VR_0(\lambda \pm i0)u = 0 \qquad [29]$$

Multiplying this equation by $R_0(\lambda \pm i0)u$ and taking the imaginary part of the scalar product, we see that

$$\pi\, d(E_0(\lambda)u, u)/d\lambda = \mathrm{Im}(R_0(\lambda \pm i0)u, u) = 0$$

According to eqn [26], this implies that

$$\hat{u}(\xi) = 0 \quad \text{for} \quad |\xi| = \lambda^{1/2} \qquad [30]$$

It follows from eqn [29] that

$$\psi = R_0(\lambda \pm i0)u \qquad [31]$$

that is, $\hat{\psi}(\xi) = (|\xi|^2 - \lambda \mp i0)^{-1}\hat{u}(\xi)$, is a formal (because of the singularity of the denominator) solution of Schrödinger equation [5]. Therefore, one needs only to verify that $\psi \in L_2(\mathbb{R}^d)$. Since $u \in L_2^{(l)}$, where $l = \rho/2$, this is a direct consequence of [25] and [30] if $\rho > 2$. In the general case, one uses that under assumption [30] the function $(|\xi|^2 - \lambda)^{-1}\hat{u}(\xi)$ belongs to the space $L_2^{(p)}$ for any $p < l - 1$. By virtue of condition [4] where $\rho > 1$, eqn [29] now shows that actually $u \in L_2^{(p)}$ for any $p < l + \rho - 1$. Repeating these arguments, we obtain, after $n$ steps, that $u \in L_2^{(p)}$ for any $p < l + n(\rho - 1)$. For $n$ large enough, this implies that $u \in L_2^{(p)}$ for $p > 1$, and consequently function [31] belongs to $L_2(\mathbb{R}^d)$.

Similar arguments show that eigenvalues of $H$ have finite multiplicity and do not have positive accumulation points. For the proof of boundedness of the set of eigenvalues, one uses additionally the estimate

$$\|\mathcal{R}_0(\lambda \pm i0)\| = O(\lambda^{-1/2}), \quad \lambda \to \infty \qquad [32]$$

Actually, according to Kato theorem the Schrödinger operator $H$ does not have positive eigenvalues.

There exists also a purely time-dependent approach, the Enss method (see Perry (1983)), which relies on an advanced study of the free evolution operator $\exp(-iH_0 t)$.

## The Scattering Matrix

The operator $H_0 = -\Delta$ can of course be diagonalized by the classical Fourier transform. To put it slightly differently, set

$$(F_0 f)(\lambda) = \Gamma_0(\lambda) f$$

where the operator $\Gamma_0(\lambda)$ is defined by eqn [16]. Then

$$F_0 : L_2(\mathbb{R}^d) \to L_2(\mathbb{R}_+; \mathfrak{N}), \quad \mathfrak{N} = L_2(\mathbb{S}^{d-1})$$

is a unitary operator and $(F_0 H_0 f)(\lambda) = \lambda (F_0 f)(\lambda)$.

Under assumption [4] where $\rho > 1$, the scattering operator $S$ for pair [13] is defined by eqn [11]. It is unitary on the space $\mathcal{H} = L_2(\mathbb{R}^d)$ and commutes with the operator $H_0$. It follows that $(F_0 S f)(\lambda) = S(\lambda)(F_0 f)(\lambda), \lambda > 0$, where the unitary operator $S(\lambda) : \mathfrak{N} \to \mathfrak{N}$ is known as the scattering matrix. The scattering matrix $S(\lambda)$ for the pair $H_0, H$ can be computed in terms of the scattering amplitude. Namely, $S(\lambda)$ acts in the space $L_2(\mathbb{S}^{d-1})$, and $S(\lambda) - I$ is the integral operator whose kernel is the scattering amplitude. More precisely,

$$(S(\lambda)f)(\theta) = f(\theta) + 2i\lambda^{1/2}\overline{\gamma_d(\lambda)} \int_{\mathbb{S}^{d-1}} a(\theta, \omega; \lambda) f(\omega)\, d\omega$$

In operator notation, this representation can be rewritten as

$$S(\lambda) = I - 2\pi i \Gamma_0(\lambda)(V - VR(\lambda + i0)V)\Gamma_0^*(\lambda) \qquad [33]$$

The right-hand side here is correctly defined as a bounded operator in the space $\mathfrak{N}$ and is continuous in $\lambda > 0$. Moreover, the operator $S(\lambda) - I$ is compact since $\Gamma_0(\lambda)\langle x \rangle^{-l} : \mathcal{H} \to \mathfrak{N}$ is compact for $l > 1/2$ by virtue of the Sobolev trace theorem.

It follows that the spectrum of the operator $S(\lambda)$ consists of eigenvalues of finite multiplicity, except possibly the point 1, lying on the unit circle and accumulating at the point 1 only. In the general case, eigenvalues of $S(\lambda)$ play the role of scattering phases or shifts considered often for radial potentials $V(x) = V(|x|)$.

The scattering amplitude is singular on the diagonal $\theta = \omega$ only. Moreover, this singularity is weaker for potentials with faster decay at infinity (for $\rho$ bigger). If $\rho > (d+1)/2$, then the operator $S(\lambda) - I$ belongs to the Hilbert–Schmidt class. In this case the total scattering cross section

$$\sigma(\omega; \lambda) = \int_{\mathbb{S}^{d-1}} |a(\theta, \omega; \lambda)|^2\, d\theta$$

is finite for all energies $\lambda > 0$ and all incident directions $\omega \in \mathbb{S}^{d-1}$. If $\rho > d$, then the operator $S(\lambda) - I$ belongs to the trace class. In this case, the scattering amplitude $a(\theta, \omega; \lambda)$ is a continuous function of $\theta, \omega \in \mathbb{S}^{d-1}$ (and $\lambda > 0$). The unitarity of the operator $S(\lambda)$ implies the optical theorem

$$\sigma(\omega; \lambda) = \lambda^{-1/2} \mathrm{Im}\big(\gamma_d^{-1}(\lambda) a(\omega, \omega; \lambda)\big)$$

Using resolvent identity [14], one deduces from eqn [33] the Born expansion

$$S(\lambda) = I - 2\pi \mathrm{i} \sum_{n=0}^{\infty} (-1)^n \Gamma_0(\lambda) V (R_0(\lambda + \mathrm{i}0) V)^n \Gamma_0^*(\lambda)$$

This series is norm-convergent for small potentials $V$ and according to estimate [32] for high energies $\lambda$.

## Long-Range Interactions

Potentials decaying at infinity as the Coulomb potential

$$V(x) = \gamma |x|^{-1}, \quad d \geq 3$$

or slower are called long-range. More precisely, it is required that

$$|\partial^\alpha V(x)| \leq C(1 + |x|)^{-\rho - |\alpha|}, \quad \rho \in (0, 1] \qquad [34]$$

for all derivatives of $V$ up to some order. In the long-range case, the wave operators $W_\pm(H, H_0)$ do not exist, and the asymptotic dynamics should be properly modified. It can be done in a time-dependent way either in the coordinate or momentum representations. For example, in the coordinate representation, the free evolution $\exp(-\mathrm{i}H_0 t)$ should be replaced in definition [8] of wave operators by unitary operators $U_0(t)$ defined by

$$(U_0(t)f)(x) = \exp(\mathrm{i}\,\Xi(x,t))(2\mathrm{i}t)^{-d/2} \hat{f}(x/(2t))$$

where $\hat{f}$ is the Fourier transform of $f$. For short-range potentials we can set $\Xi(x,t) = (4t)^{-1}|x|^2$. In the long-range case the phase function $\Xi(x,t)$ should be chosen as a (perhaps, approximate) solution of the eikonal equation

$$\partial\Xi/\partial t + |\nabla\Xi|^2 + V = 0$$

In particular, we can set

$$\Xi(x,t) = (4t)^{-1}|x|^2 - t \int_0^1 V(sx)\,\mathrm{d}s$$

if $\rho > 1/2$ in [34]. For the Coulomb potential,

$$\Xi(x,t) = (4t)^{-1}|x|^2 - \gamma t |x|^{-1} \ln|t|$$

(the singularity at $x = 0$ is inessential here). Thus, both in short- and long-range cases solutions of the time-dependent Schrödinger equation "live" in a region of the configuration space where $|x|$ is of order $|t|$. Long-range potentials change only asymptotic phases of these solutions.

Another possibility is a time-independent modification in the phase space. Let us consider wave operators $W_\pm(H, H_0; J)$, where $J$ is a pseudodifferential operator,

$$(Jf)(x) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \mathrm{e}^{\mathrm{i}\langle x,\xi\rangle}\, \mathrm{e}^{\mathrm{i}\Phi(x,\xi)} \zeta(x,\xi) \hat{f}(\xi)\,\mathrm{d}\xi$$

with oscillating symbol $\exp(\mathrm{i}\Phi(x,\xi))\zeta(x,\xi)$. Due to the conservation of energy, we may suppose that $\zeta(x,\xi)$ contains a factor $\psi(|\xi|^2)$ with $\psi \in C_0^\infty(0,\infty)$. Set

$$\varphi(x,\xi) = \langle x,\xi\rangle + \Phi(x,\xi)$$

The perturbation $HJ - JH_0$ is also a pseudodifferential operator, and its symbol is short-range (it is $O(|x|^{-1-\varepsilon}), \varepsilon > 0$, as $|x| \to \infty$) if $\exp(\mathrm{i}\varphi(x,\xi))$ is an approximate eigenfunction of the operator $H$ corresponding to the "eigenvalue" $|\xi|^2$. This leads to the eikonal equation

$$|\nabla_x \varphi(x,\xi)|^2 + V(x) = |\xi|^2$$

The notorious difficulty (for $d \geq 2$) of this method is that the eikonal equation does not have (even approximate) solutions such that $|\nabla_x \Phi(x,\xi)| \to 0$ as $|x| \to \infty$ and the arising error term is short-range. However, it is easy to construct functions $\varphi = \varphi_\pm$ satisfying these conditions if a conical neighborhood of the direction $\mp\xi$ is removed from $\mathbb{R}^d$. For example,

$$\Phi_\pm(x,\xi) = \pm 2^{-1} \int_0^\infty (V(x \pm \tau\xi) - V(\pm\tau\xi))\,\mathrm{d}\tau$$

if $\rho > 1/2$ in eqn [34]. Then the cutoff function $\zeta(x,\xi) = \zeta_\pm(x,\xi)$ should be homogeneous of order zero in the variable $x$ and it should be equal to zero in a neighborhood of the direction $\mp\xi$. We emphasize that now we have a couple of different identifications $J = J_\pm$.

The long-range problem is essentially more difficult than the short-range one. The limiting absorption principle remains true in this case, but its proof cannot be performed within perturbation theory. The simplest proof relies on the Mourre estimate (see Cycon *et al.* (1987)) for the commutator $\mathrm{i}[H, A]$ of $H$ with the generator of dilations

$$A = -\mathrm{i} \sum_{j=1}^{d} (x_j \partial_j + \partial_j x_j)$$

The Mourre estimate affirms that, for all $\lambda > 0$,

$$\mathrm{i}E(\Lambda_\lambda)[H, A]E(\Lambda_\lambda) \geq c(\lambda)E(\Lambda_\lambda), \quad c(\lambda) > 0 \qquad [35]$$

if $\Lambda_\lambda = (\lambda - \varepsilon, \lambda + \varepsilon)$ and $\varepsilon$ is small enough. For the free operator $H_0$, this estimate takes the form $\mathrm{i}[H_0, A] = 4H_0$ and can be regarded as a commutation relation. Estimate [35] means that the observable

$$(A\mathrm{e}^{-\mathrm{i}Ht} f,\ \mathrm{e}^{-\mathrm{i}Ht} f)$$

is a strictly increasing function of $t$ for all $f \in \mathcal{H}^{(ac)}$. The $H$-smoothness of the operator $\langle x \rangle^{-l}, l > 1/2$, is deduced from this fact by some arguments of abstract nature (they do not really use concrete forms of the operators $H$ and $A$).

However, the limiting absorption principle is not sufficient for construction of scattering theory in the long-range case, and it should be supplemented by an additional estimate. To formulate it, denote by

$$(\nabla^{\perp} u)(x) = (\nabla u)(x) - |x|^{-2} \langle (\nabla u)(x), x \rangle x$$

the orthonal projection of a vector $(\nabla u)(x)$ on the plane orthogonal to $x$. Then the operator $K = \langle x \rangle^{-1/2} \nabla^{\perp}$ is $H$-smooth on any compact $X \subset (0, \infty)$. This result is formulated as an estimate (either on the resolvent or on the unitary group of $H$), which we refer to as the radiation estimate. This estimate is not very astonishing from the viewpoint of analogy with the classical mechanics. Indeed, in the case of free motion, the vector $x(t)$ of the position of a particle is directed asymptotically as its momentum $\xi$. Regarded as a pseudodifferential operator, $\nabla^{\perp}$ has symbol $\xi - |x|^{-2} \langle \xi, x \rangle x$, which equals zero if $x = \gamma \xi$ for some $\gamma \in \mathbb{R}$. Thus, $\nabla^{\perp}$ removes the part of the phase space where a classical particle propagates. The proof of the radiation estimate is based on the inequality

$$K^* K \leq C_0 [H, \partial_r] + C_1 \langle x \rangle^{-1-\rho}, \quad \partial_r = \partial / \partial |x|$$

which can be obtained by a direct calculation. Since the integral

$$\begin{aligned} &\mathrm{i} \int_0^t ([H, \partial_r] \mathrm{e}^{-\mathrm{i}Hs} f, \, \mathrm{e}^{-\mathrm{i}Hs}) f) \, \mathrm{d}s \\ &\quad = (\partial_r \, \mathrm{e}^{-\mathrm{i}Ht} f, \, \mathrm{e}^{-\mathrm{i}Ht} f) - (\partial_r f, f) \end{aligned}$$

is bounded by $C(X) \|f\|^2$ for $f \in E(X) f$ and the operator $\langle x \rangle^{-(1+\rho)/2}$ is $H$-smooth on $X$, this implies $H$-smoothness of the operator $KE(X)$.

Calculating the perturbation $HJ_{\pm} - J_{\pm} H_0$, we see that it is a sum of two pseudodifferential operators. The first of them is short-range and thus can be taken into account by the limiting absorption principle. The symbol of the second one contains first derivatives (in the variable $x$) of the cutoff function $\zeta_{\pm}(x, \xi)$ and hence decreases at infinity as $|x|^{-1}$ only. This operator factorizes into a product of $H_0$- and $H$-smooth operators according to the radiation estimate. Thus, all wave operators $W_{\pm}(H, H_0; J_{\pm})$ and $W_{\pm}(H_0, H; J_{\pm}^*)$ exist. These operators are isometric since the operators $J_{\pm}$ are in some sense close to unitary operators. The isometricity of $W_{\pm}(H_0, H; J_{\pm}^*)$ is equivalent to the completeness of $W_{\pm}(H, H_0; J_{\pm})$.

Although the modified wave operators enjoy basically the same properties as in the short-range case, properties of the scattering matrices in the short- and long-range cases are drastically different. Here we note only that for long-range potentials, due to a wild diagonal singularity of kernel of the scattering matrix, its spectrum covers the whole unit circle.

Different aspects of long-range scattering are discussed in Dereziński and Gérard (1997), Pearson (1988), Saitō (1979), and Yafaev (2000).

*See also:* *N*-Particle Quantum Scattering; Quantum Dynamical Semigroups; Random Matrix Theory in Physics; Scattering in Relativistic Quantum Field Theory: Fundamental Concepts and Tools; Schrödinger Operators; Spectral Theory for Linear Operators.

## Further Reading

Amrein WO, Jauch JM, and Sinha KB (1977) *Scattering Theory in Quantum Mechanics*. New York: Benjamin.

Cycon H, Froese R, Kirsh W, and Simon B (1987) *Schrödinger Operators*, Texts and Monographs in Physics. Berlin: Springer.

Dereziński J and Gérard C (1997) *Scattering Theory of Classical and Quantum N Particle Systems*. Berlin: Springer.

Faddeev LD (1976) Inverse problem of quantum scattering theory II. *J. Soviet. Math.* 5: 334–396.

Friedrichs K (1965) *Perturbation of Spectra in Hilbert Space*. Providence, RI: American Mathematical Society.

Kuroda ST (1978) *An Introduction to Scattering Theory, Lecture Notes Series No. 51*. Aarhus University.

Landau LD and Lifshitz EM (1965) *Quantum Mechanics*. Pergamon.

Pearson D (1988) *Quantum Scattering and Spectral Theory*. London: Academic Press.

Perry PA (1983) *Scattering Theory by the Enss Method*. vol. 1. Harwood: Math. Reports.

Reed M and Simon B (1979) *Methods of Modern Mathematical Physics III*. New York: Academic Press.

Saitō Y (1979) *Spectral Representation for Schrödinger Operators with Long-Range Potentials*. Springer Lecture Notes in Math, vol. 727.

Yafaev DR (1992) *Mathematical Scattering Theory*. Providence, RI: American Mathematical Society.

Yafaev DR (2000) *Scattering Theory: Some Old and New Problems*, Springer Lecture Notes in Math., vol. 1735.

# Quantum Mechanics: Foundations

**R Penrose**, University of Oxford, Oxford, UK



## The Framework of Quantum Mechanics

In 1900, Max Planck initiated the quantum revolution by presenting the hypothesis that radiation is emitted or absorbed only in "quanta," each of energy $h\nu$, for frequency $\nu$ (where $h$ was a new fundamental constant of Nature). By this device, he explained the precise shape of the puzzling black-body spectrum. Then, in 1905, Albert Einstein introduced the concept of the photon, according to which light, of frequency $\nu$ would, in appropriate circumstances, behave as though it were constituted as individual particles, each of energy $h\nu$, rather than as continuous waves, and he was able to explain the conundrum posed by the photoelectric effect by this means. Later, in 1923, Prince Louis de Broglie proposed that, conversely, all particles behave like waves, the energy being Planck's $\hbar\nu$ and the momentum being $h\lambda^{-1}$, where $\lambda$ is the wavelength, which was later strikingly confirmed in a famous experiment of Davisson and Germer in 1927. Some years earlier, in 1913, Niels Bohr had used another aspect of this curious quantum "discreteness," explaining the stable electron orbits in hydrogen by the assumption that (orbital) angular momentum must be quantized in units of $\hbar(=h/2\pi)$.

All this provided a very remarkable collection of facts and concepts, albeit somewhat disjointed, explaining a variety of previously baffling physical phenomena, where a certain discreteness seemed to be entering Nature at a fundamental level, where previously there had been continuity, and where there was an overriding theme of a confusion as to whether – or in what circumstances – waves or particles provide better pictures of reality. Moreover, no clear and consistent picture of an actual "quantum-level reality" as yet seemed to arise out of all this. Then, in 1925, Heisenberg introduced his "matrix mechanics," subsequently developed into a more complete theory by Born, Heisenberg and Jordan, and then more fully by Dirac. Some six months after Heisenberg, in 1926, Schrödinger introduced his very different-looking "wave mechanics," which he subsequently showed was equivalent to Heisenberg's scheme. These became encompassed into a comprehensive framework through the transformation theory of Dirac, which he put together in his famous book *The Principles of Quantum Mechanics*, first published in 1930. Later, von Neumann set the framework on a more rigorous basis in his 1932 book, *Mathematische Grundlageen der Quantenmechanik* (later translated as *Mathematical Foundations of Quantum Mechanics*, 1955).

This formalism, now well known to physicists, is based on the presence of a quantum state $|\psi\rangle$ (Dirac's "ket" notation being adopted here). In Schrödinger's description, $|\psi\rangle$ is to evolve by unitary evolution, according to the Schrödinger equation

$$i\hbar\frac{\partial|\psi\rangle}{\partial t}=\mathtt{H}|\psi\rangle$$

where $\mathtt{H}$ is the quantum Hamiltonian. The totality of allowable states $|\psi\rangle$ constitutes a Hilbert space $\mathbf{H}$ and the Schrödinger equation provides a continuous one-parameter family of unitary transformations of $\mathbf{H}$. The letter $\boldsymbol{U}$ is used here for the "quantum-level" evolution whereby the state $|\psi\rangle$ evolves in time according to this unitary Schrödinger evolution. However, we must be careful not to demand an interpretation of this evolution similar to that which we adopt for a classical theory, such as is provided by Maxwell's equations for the electromagnetic field. In Maxwell's theory, the evolution that his equations provide is accepted as very closely mirroring the actual way in which a physically real electromagnetic field evolves with time. In quantum mechanics, however, it is a highly contentious matter how we should regard the "reality" of the unitarily evolving state $|\psi\rangle$.

One of the key difficulties resides in the fact that the world that we actually observe about us rather blatantly does not accord with such a unitarily evolving $|\psi\rangle$. Indeed, the standard way that the quantum formalism is to be interpreted is very far from the mere following of such a picture. So long as no "measurement" is deemed to have been taking place, this $\boldsymbol{U}$-evolution procedure would be adopted, but upon measurement, the state is taken to behave in a very different way, namely to "jump" instantaneously to some eigenstate $|\phi\rangle$ of the quantum operator $\boldsymbol{Q}$ which is taken to represent the measurement, with probability given by the Born rule

$$|\langle\phi|\psi\rangle|^2$$

if we assume that both $|\psi\rangle$ and $|\phi\rangle$ are normalized ($\langle\psi|\psi\rangle=1=\langle\phi|\phi\rangle$); otherwise we can express this probability simply as

$$\frac{\langle\phi|\psi\rangle\langle\psi|\phi\rangle}{\langle\psi|\psi\rangle\langle\phi|\phi\rangle}$$

(The operator $\boldsymbol{Q}$ is normally taken to be self-adjoint, so that $\boldsymbol{Q}=\boldsymbol{Q}^*$ and its eigenvalues are real, but more

generally complex eigenvalues are accommodated if we allow $\boldsymbol{Q}$ to be normal, that is, $\boldsymbol{Q}\boldsymbol{Q}^* = \boldsymbol{Q}^*\boldsymbol{Q}$. In each case we require the eigenvectors of $\boldsymbol{Q}$ to span the Hilbert space **H**.) This "evolution procedure" of the quantum state is very different from $\boldsymbol{U}$, owing both to its discontinuity and its indeterminacy. The letter $\boldsymbol{R}$ will be used for this, standing for the "reduction" of the quantum state (sometimes referred to as the "collapse of the wave function"). This strange hybrid, whereby $\boldsymbol{U}$ and $\boldsymbol{R}$ are alternated, with $\boldsymbol{U}$ holding between measurements and $\boldsymbol{R}$ holding at measurements, is the standard procedure that is pragmatically adopted in conventional quantum mechanics, and which works so marvelously well, with no known discrepancy between the theory and observation. (In his classic account, von Neumann (1932, 1955), "$\boldsymbol{R}$" is referred to as his "process I" and "$\boldsymbol{U}$" as his "process II.") However, there appears to be no consensus whatever about the relation between this mathematical procedure and what is "really" going on in the physical world. This is the kind of issue that will be of concern to us here.

## Quantum Reality

The discussion here will be given only in the Schrödinger picture, for the reason that the issues appear to be clearer with this description. In the Heisenberg picture, the state $|\psi\rangle$ does not evolve in time, and all dynamics is taken up in the time evolution of the dynamical variables. But this evolution does not refer to the evolution of specific systems, the "state" of any particular system being defined to remain constant in time. Since the Schrödinger and Heisenberg pictures are deemed to be equivalent (at least for the "normal" systems that are under consideration here), we do not lose anything substantial by sticking to Schrödinger's description, whereas there does seem to be a significant gain in understanding of what the formalism is actually telling us.

There are, however, many different attitudes that are expressed as to the "reality" of $|\psi\rangle$. (There is an unfortunate possibility of confusion here in the two uses of the word "real" that come into the discussion here. In the quantum formalism, the state is mathematically a "complex" rather than a "real" entity, whereas our present concern is not directly to do with this, but with the "ontology" of the quantum description.) According to what is commonly regarded as the standard – "Copenhagen" – interpretation of quantum mechanics (due primarily to Bohr, Heisenberg, and Pauli), the quantum state $|\psi\rangle$ is not taken as a description of a quantum-level reality at all, but merely as a description of the observer's knowledge of the of the quantum system under consideration. According to this view, the "jumping" that the quantum state undergoes is regarded as unsurprising, since it does not represent a sudden change in the reality of the situation, but merely in the observer's knowledge, as new information becomes available, when the result of some measurement becomes known to the observer. According to this view, there is no objective quantum reality described by $|\psi\rangle$. Whether or not there might be some objective quantum-level reality with some other mathematical description seems to be left open by this viewpoint, but the impression given is that there might well not be any such quantum-level reality at all, in the sense that it becomes meaningless to ask for a description of "actual reality" at quantum-relevant scales.

Of course some connection with the real world is necessary, in order that the quantum formalism can relate to the results of experiment. In the Copenhagen viewpoint, the experimenter's measuring apparatus is taken to be a classical-level entity, which can be ascribed a real ontological status. When the Geiger counter "clicks" or when the pointer "points" to some position on a dial, or when the track in the cloud chamber "becomes visible" – these are taken to be real events. The intervening description in terms of a quantum state vector $|\psi\rangle$ is not ascribed a reality. The role of $|\psi\rangle$ is merely to provide a calculational procedure whereby the different outcomes of an experiment can be assigned probabilities. Reality comes about only when the result of the measurement is manifested, not before.

A difficulty with this viewpoint is that it is hard to draw a clear line between those entities which are considered to have an actual reality, such as the experimental apparatus or a human observer, and the elemental constituents of those entities, which are such things as electrons or protons or neutrons or quarks, which are to be treated quantum mechanically and therefore, on the "Copenhagen" view, their mathematical descriptions are denied such an honored ontological status. Moreover, there is no limit to the number of particles that can partake in a quantum state. According to current quantum mechanics, the most accurate mathematical procedure for describing a system with a large number of particles would indeed be to use a unitarily evolving quantum state. What reasons can be presented for or against the viewpoint that this gives us a reasonable description of an actual reality? Can our perceived reality arise as some kind of statistical limit when very large numbers of constituents are involved?

Before entering into the more subtle and contentious issues of the nature of "quantum reality," it

is appropriate that one of the very basic mathematical aspects of the quantum formalism be addressed first. It is an accepted aspect of the quantum formalism that a state-vector such as $|\psi\rangle$ should not, in any case, be thought of as providing a unique mathematical description of a "physical reality" for the simple reason that $|\psi\rangle$ and $z|\psi\rangle$, where $z$ is any nonzero complex number, describe precisely the same physical situation. It is a common, but not really necessary, practice to demand that $|\psi\rangle$ be normalized to unity: $\langle\psi|\psi\rangle = 1$, in which case the freedom in $|\psi\rangle$ is reduced to the multiplication by a phase factor $|\psi\rangle \mapsto e^{i\theta}|\psi\rangle$. Either way, the physically distinguishable states constitute a projective Hilbert space $\mathbb{P}\mathbf{H}$, where each point of $\mathbb{P}\mathbf{H}$ corresponds to a one-dimensional linear subspace of the Hilbert space $\mathbf{H}$. The issue, therefore, is whether quantum reality can be described in terms of the points of a projective Hilbert space $\mathbb{P}\mathbf{H}$.

## Reality in Spin-1/2 Systems

As a general comment, it seems that for systems with a small number of degrees of freedom – that is, for a Hilbert space $\mathbf{H}^n$ of small finite dimension $n$ – it seems more reasonable to assign a reality to the elements of $\mathbb{P}\mathbf{H}^n$ than is the case when $n$ is large. Let us begin with a particularly simple case, where $n = 2$, and $\mathbf{H}^2$ describes the two-dimensional space of spin states of a massive particle of spin 1/2, such as an electron, proton, or quark, or suitable atom. Here we can take as an orthonormal pair of basis states $|\uparrow\rangle$ and $|\downarrow\rangle$, representing right-handed spin about the "up" and "down" directions, respectively. Clearly there is nothing special about these particular directions, so any other state of spin, of direction $|\nearrow\rangle$ say, is just as "real" as the original two. Indeed, we always find

$$|\nearrow\rangle = w|\uparrow\rangle + z|\downarrow\rangle$$

for some pair of complex numbers $z$ and $w$ (not both zero). The different possible ratios $z:w$ give us a complex plane (of $zw^{-1}$) compactified by a point at infinity (where $w = 0$) – a "Riemann sphere" – which is a realization of the complex projective 1-space $\mathbb{P}\mathbf{H}^2$.

There does indeed seem to be something "real" about the spin state of such a spin-1/2 particle or atom. We might imagine preparing the spin of a suitable spin-1/2 atom using a Stern–Gerlach apparatus (*see* Introductory Article: Quantum Mechanics) oriented in some chosen direction. The atom seems to "know" the direction of its spin, because if we measure it again in the same direction it has to be prepared to give us the answer "YES," to the second measurement, with certainty, and that direction for its spin state is the only one that can guarantee this answer. (We are, of course, considering only "ideal" measurements, for the purpose of argument.) Moreover, we could imagine that between the two measurements, some appropriate magnetic field had been introduced so as to rotate the spin direction in some very specific way, so that the spin state is now some other direction such as $|\searrow\rangle$. By rotating our second Stern–Gerlach apparatus to agree with this new direction, we must again get certainty for the YES answer, the guaranteeing of this by the rotated state seeming now to give a "reality" to this new state $|\searrow\rangle$. The quantum formalism does not allow us to ascertain an unknown direction of spin. But it does allow for us to "confirm" (or "refute") a proposed direction for the spin state, in the sense that if the proposed direction is incorrect, then there is a nonzero probability of refutation. Only the correct direction can be guaranteed to give the YES answer.

## EPR–Bohm and Bell's Theorem

For a pair of particles or atoms of spin 1/2, the issue of the "reality" of spin states becomes less clear. Consider, for example, the EPR–Bohm example (where "EPR" stands for Einstein–Podolski–Rosen) whereby an initial state of spin 0 decays into two spin-1/2 atoms, traveling in opposite directions (east E, and west W). If a suitable Stern–Gerlach apparatus is set up to measure the spin of the atom at E, finding an answer $|\searrow\rangle$, say, then this immediately ensures that the state at W is the oppositely pointing $|\nwarrow\rangle$, which can subsequently be "confirmed" by measurement at W. This, then, seems to provide a "reality" for the spin state $|\nwarrow\rangle$ at W as soon as the E measurement has been performed, but not before. Now, let us suppose that some orientation different from $\nwarrow$ had actually been set up for the measurement at W, namely that which would have given YES for the direction $\leftarrow$. This measurement can certainly give the answer YES upon encountering $|\nwarrow\rangle$ (with a certain nonzero probability, namely $(1 + \cos\theta)/2$, where $\theta$ is the angle between $\nwarrow$ and $\leftarrow$). So far, this provides us with no problem with the "reality" of the spin state of the atom at W, since it would have been $|\nwarrow\rangle$ before the measurement at W and would have "collapsed" (by the ***R***-process) to $|\leftarrow\rangle$ after the measurement. But now suppose that the measurement at W had actually been performed momentarily before the measurement at E, rather than just after it. Then there is no reason that the W-measurement would encounter $|\nwarrow\rangle$, rather than some other direction, but the result $|\leftarrow\rangle$ of the measurement at W now seems to force the state at E to be $|\rightarrow\rangle$. Indeed, the two measurements, at E and

at W, might have been spacelike separated, and because of the requirements of special relativity there would be no meaning to say which of the two measurements – at E or at W – had "actually" occurred first. One seems to obtain a different picture of "reality" depending on this ordering.

In fact, the calculations of probabilities come out the same whichever picture is used, so if one asks only for a calculational procedure for the probabilities, rather than an actual picture of quantum reality, these considerations are not problematic. But they do provide profound difficulties for any view of quantum reality that is entirely local. The difficulty is made particularly clear in a theorem due to John Bell (1964, 1966a, b) which showed that on the basis of the assumptions of local realism, there are particular relations between the conditional probabilities, which must hold in any situation of this kind; moreover, these inequalities can be violated in various situations in standard quantum mechanics. (See, most specifically, Clauser *et al.* (1969).) Several experiments that were subsequently performed (notably Aspect *et al.* (1982)) confirmed the expectations of quantum mechanics, thereby presenting profound difficulties for any local realistic model of the world. There are also situations of this kind which involve only yes/no questions, so that actual probabilities do not need to be considered, see Kochen and Specker (1967), Peres (1991), Hardy (1993), Conway and Kochen (2002). Basically: if one insists on realism, then one must give up locality. Moreover, nonlocal realistic models, consistent with the requirements of special relativity, are not easy to construct (*see* Quantum Mechanics: Generalizations), and have so far proved elusive.

## Other Aspects of Quantum Nonlocality

Problems of this kind occur even at the more elementary level of single particles, if one tries to consider that an ordinary particle wave function (position-space description of $|\psi\rangle$) might be just some kind of "local disturbance," like an ordinary classical wave. Consider the wave function spreading out from a localized source, to be detected at a perpendicular screen some distance away. The detection of the particle at any one place on the screen immediately forbids the detection of that particle at any other place on the screen, and if we are to think of this information as being transmitted as a classical signal to all other places on the screen, then we are confronted with problems of superluminary communication. Again, any "realistic" picture of this process would require nonlocal ingredients, which are difficult to square with the requirements of special relativity. (It is possible that these difficulties might be resolved within some kind of nonlocal geometry, such as that supplied by twistor theory (*see* Twistors; Twistor Theory: Some Applications); see, particularly, Penrose (2005).)

These types of issues are made even more dramatic and problematic in the procedure of "quantum teleportation," whereby the information in a quantum state (e.g., the unknown actual direction $\nearrow$ in some quantum state $|\nearrow\rangle$) can be transported from one experimenter A to another one B, by merely the sending of a small finite number of classical bits of information from A to B, where before this classical information is transmitted, A and B must each be in possession of one member of an EPR pair. More explicitly, we may suppose A (Alice) is presented with a spin-1/2 state $|\nearrow\rangle$, but is not told the direction $\nearrow$. She has in her possession another spin-1/2 state which is an EPR–Bohm partner of a spin-1/2 state in the possession of B (Bob). She combines this $|\nearrow\rangle$ with her EPR atom and then performs a measurement which distinguishes the four orthogonal "Bell states"

$$\begin{aligned}
0&: \ |\uparrow\rangle|\downarrow\rangle - |\downarrow\rangle|\uparrow\rangle \\
1&: \ |\uparrow\rangle|\uparrow\rangle - |\downarrow\rangle|\downarrow\rangle \\
2&: \ |\uparrow\rangle|\uparrow\rangle + |\downarrow\rangle|\downarrow\rangle \\
3&: \ |\uparrow\rangle|\downarrow\rangle + |\downarrow\rangle|\uparrow\rangle
\end{aligned}$$

where the first state in each product refers to her unknown state and the second refers to her EPR atom. The result of this measurement is conveyed to Bob by an ordinary classical signal, coded by the indicated numbers 0, 1, 2, 3. On receiving Alice's message, Bob takes the other member of the EPR pair and performs the following rotation on it:

0: leave alone
1: 180° about $x$-axis
2: 180° about $y$-axis
3: 180° about $z$-axis

This achieves the successful "teleporting" of $|\nearrow\rangle$ from A to B, despite the fact that only 2 bits of classical information have been signaled. It is the acausal EPR–Bohm connection that provides the transmission of "quantum information" in a classically acausal way. Again, we see the essentially nonlocal (or acausal) nature of any attempted "realistic" picture of quantum phenomena. It may be regarded as inappropriate to use the term "information" for something that is propagated acausally and cannot be directly used for signaling. It has been suggested, accordingly, that a term such as "quanglement" might be more appropriate to use for this concept; see Penrose (2002, 2004).

The preceding arguments illustrate how quantum systems involving even just a few particles can exhibit features quite unlike the ordinary behavior of classical particles. This was pointed out by Schrödinger (1935), and he referred to this key property of composite quantum systems as "entanglement." An entangled quantum state (vector) is an element of a product Hilbert space $\mathbf{H}^m \otimes \mathbf{H}^n$ which cannot be written as a tensor product of elements $|\psi\rangle|\phi\rangle$, with $|\psi\rangle \in \mathbf{H}^m$ and $|\phi\rangle \in \mathbf{H}^n$, where $\mathbf{H}^m$ refers to one part of the system and $\mathbf{H}^n$ refers to another part, usually taken to be physically widely separated from the first. EPR systems are a clear example, and we begin to see very nonclassical, effectively nonlocal behavior with entangled systems generally. A puzzling aspect of this is that the vast majority of states are indeed entangled, and the more parts that a system has, the more entangled it becomes (where the generalization of this notion to more than two parts is evident). One might have expected that "big" quantum systems with large numbers of parts ought to behave more and more like classical systems when they get larger and more complicated. However, we see that this is very far from being the case. There is no good reason why a large quantum system, left on its own to evolve simply according to $\boldsymbol{U}$ should actually resemble a classical system, except in very special circumstances. Something of the nature of the $\boldsymbol{R}$ process seems to be needed in order that classical behaviour can "emerge."

## Schrödinger's Cat

To clarify the nature of the problem we must consider a key feature of the $\boldsymbol{U}$ formalism, namely "linearity," which is supposed to hold no matter how large or complicated is the quantum system under consideration. Recall the quantum superposition principle, which allows us to construct arbitrary combinations of states

$$|\psi\rangle = w|\chi\rangle + z|\phi\rangle$$

from two given states $|\chi\rangle$ and $|\phi\rangle$. Quantum linearity tells us that if

$$|\chi\rangle \rightsquigarrow |\chi'\rangle \qquad \text{and} \qquad |\phi\rangle \rightsquigarrow |\phi'\rangle$$

where the symbol "$\rightsquigarrow$" expresses how a state will have evolved after a specified time period $T$, then

$$|\psi\rangle = w|\chi\rangle + z|\phi\rangle \rightsquigarrow |\psi'\rangle = w|\chi'\rangle + z|\phi'\rangle$$

Let us now consider how this might be applied in a particular, rather outlandish situation. Let us suppose that the $|\chi\rangle$-evolution consists of a photon going in one direction, encountering a detector, which is connected to some murderous device which kills a cat. The $|\phi\rangle$-evolution, on the other hand, consists of the photon going in some other direction, missing the detector so that the murderous device is not activated, and the cat is left alive. These two alternatives would each be perfectly plausible evolutions which might take place in the physical world. Now, by use of a beam splitter (effectively a "half-silvered mirror") we can easily arrange for the initial state of the photon to be the superposition $w|\chi\rangle + z|\phi\rangle$ of the two. Then by quantum linearity we find, as the final result, the superposed state $w|\chi'\rangle + z|\phi'\rangle$, in which the cat is in a superposition of life and death (a "Schrödinger's cat").

We note that the two individual final states $|\chi'\rangle$ and $|\phi'\rangle$ would each involve not just the cat but also its environment, fully entangled with the cat's state, and perhaps also some human observer looking at the cat. In the latter case, $|\chi'\rangle$ would involve the observer in a state of unhappily perceiving a dead cat, and $|\phi'\rangle$ happily perceiving a live one. Two of the "conventional standpoints" with regard to the measurement problem are of relevance here. According to the standpoint of environmental decoherence, the details of the environmental degrees of freedom are completely inaccessible, and it is deemed to be appropriate to construct a density matrix to describe the situation, which is a partial trace $\boldsymbol{D}$ of the quantity $|\psi\rangle\langle\psi|$, constructed by tracing out over all the environmental degrees of freedom:

$$\boldsymbol{D} = \text{trace over environment}\{|\psi\rangle\langle\psi|\}$$

The density matrix tends to be regarded as a more appropriate quantity than the ket $|\psi\rangle$ to represent the physical situation, although this represents something of an "ontology shift" from the point of view that was being held previously. Under appropriate assumptions, $\boldsymbol{D}$ may now be shown to attain a form that is close to being diagonal in a basis with respect to which the cat is either dead or alive, and then, by a second "ontology shift" $\boldsymbol{D}$ is re-read as describing a probability mixture of these two states.

According to the second "conventional standpoint" under consideration here, it is not logical to take this detour through a density-matrix description, and instead one should maintain a consistent ontology by following the evolution of the state $|\psi\rangle$ itself throughout. The "real" resulting physical state is then taken to be actually $|\psi'\rangle$, which involves the superposition of a dead and live cat. Of course this "reality" does not agree with the reality that we actually perceive, so the position is taken that a conscious mind would not actually be able to function in such a superposed condition, and would have to settle into a state of perception of either a dead cat or a live one, these two alternatives occurring with probabilities as given by the Born rule stated above. It may be argued that this conclusion depends

upon some appropriate theory of how conscious minds actually perceive things, and this appears to be lacking.

A good many physicists might argue that none of these attempts at resolution of the measurement problem is satisfactory, including "Copenhagen," although the latter at least has the advantage of offering a pragmatic, if not fully logical, stance. Such physicists might take the position that it is necessary to move away from the precise version of quantum theory that we have at present, and turn to one of its modifications. Some major candidates for modification are discussed in Quantum Mechanics: Generalizations. Most of these actually make predictions that, at some stage, would differ from those of standard quantum mechanics. So it becomes an experimental matter to ascertain the plausibility of these schemes. In addition, there are reinterpretations which do not change quantum theory's predictions, such as the de Broglie–Bohm model. In this, there are two levels of "reality," a firmer one with a particle or position-space ontology, and a secondary one containing waves which guide the behavior at the firmer level. It is clear, however, that these issues will remain the subject of debate for many years to come.

*See also:* Functional Integration in Quantum Physics; Normal Forms and Semiclassical Approximation; Quantum Mechanics: Generalizations; Twistor Theory: Some Applications [In Integrable Systems, Complex Geometry and String Theory]; Twistors.

## Further Reading

Aspect A, Grangier P, and Roger G (1982) Experimental realization of Einstein–Podolsky–Rosen–Bohm. *Gedankenexperiment*: a new violation of Bell's inequalities. *Physical Review Letters* 48: 91–94.

Bell JS (1964) On the Einstein Podolsky Rosen paradox. *Physics* 1: 195–200. Reprinted in *Quantum Theory and Measurement*, eds., Wheeler JA and Zurek WH (Princeton Univ. Press, Princeton, 1983).

Bell JS (1966a) On the problem of hidden variables in quantum theory. *Reviews in Modern Physics* 38: 447–452.

Bell JS (1966b) *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press. Reprint 1987.

Clauser JF, Horne MA, Shimony A, and Holt RA (1969) Proposed experiment to test local hidden-variable theories. *Physical Review Letters* 23: 880–884.

Conway J and Kochen S (2002) The geometry of the quantum paradoxes. In: Bertlmann RA and Zeilinger A (eds.) *Quantum [Un]speakables: From Bell to Quantum Information*, Ch. 18 (ISBN 3-540-42756-2). Berlin: Springer.

Hardy L (1993) Nonlocality for two particles without inequalities for almost all entangled states. *Physical Review Letters* 71(11): 1665.

Kochen S and Specker EP (1967) *Journal of Mathematics and Mechanics* 17: 59.

Penrose R (2002) John Bell, state reduction, and quanglement. In: Reinhold A, Bertlmann, and Zeilinger A (eds.) *Quantum [Un]speakables: From Bell to Quantum Information*, pp. 319–331. Berlin: Springer.

Penrose R (2004) *The Road to Reality: A Complete Guide to the Laws of the Universe*. London: Jonathan Cape.

Penrose R (2005) The twistor approach to space-time structures. In: Ashtekar A (ed.) *In 100 Years of Relativity; Space-Time Structure: Einstein and Beyond*. Singapore: World Scientific.

Peres A (1991) Two simple proofs of the Kochen–Specker theorem. *Journal of Physics A: Mathematical and General* 24: L175–L178.

Schrödinger E (1935) Probability relations between separated systems. *Proceedings of the Cambridge Philosophical Society* 31: 555–563.

von Neumann J (1932) *Mathematische Grundlageen der Quantenmechanik*. Berlin: Springer.

von Neumann J (1955) *Mathematical Foundations of Quantum Mechanics*. Princeton: Princeton University Press.

# Quantum Mechanics: Generalizations

**P Pearle**, Hamilton College, Clinton, NY, USA
**A Valentini**, Perimeter Institute for Theoretical Physics, Waterloo, ON, Canada



## Introduction

According to the so-called "Copenhagen Interpretation," standard quantum theory is limited to describing experimental situations. It is at once remarkably successful in its predictions, and remarkably ill-defined in its conceptual structure: what is an experiment? what physical objects do or do not require quantization? how are the states realized in nature to be characterized? how and when is the wave-function "collapse postulate" to be invoked? Because of its success, one may suspect that quantum theory can be promoted from a theory of measurement to a theory of reality. But, that requires there to be an unambiguous specification (S) of the possible real states of nature and their probabilities of being realized.

There are several approaches that attempt to achieve S. The more conservative approaches (e.g., consistent histories, environmental decoherence, many worlds) do not produce any predictions that differ from the standard ones because they do not tamper with the usual basic mathematical

formalism. Rather, they utilize structures compatible with standard quantum theory to elucidate S. These approaches, which will not be discussed in this article, have arguably been less successful so far at achieving S than approaches that introduce significant alterations to quantum theory.

This article will largely deal with the two most well-developed realistic models that reproduce quantum theory in some limit and yield potentially new and testable physics outside that limit. First, the pilot-wave model, which will be discussed in the broader context of "hidden-variables theories." Second, the continuous spontaneous localization (CSL) model, which describes wave-function collapse as a physical process. Other related models will also be discussed briefly.

Due to bibliographic space limitations, this article contains a number of uncited references, of the form "[author] in [year]." Those in the next section can be found in Valentini (2002b, 2004a,b) or at www.arxiv.org. Those in the subsequent sections can be found in Adler (2004), Bassi and Ghirardi (2003), Pearle (1999) (or in subsequent papers by these authors, or directly, at www.arxiv.org), and in Wallstrom (1994).

## Hidden Variables and Quantum Nonequilibrium

A deterministic hidden-variables theory defines a mapping $\omega = \omega(M, \lambda)$ from initial hidden parameters $\lambda$ (defined, e.g., at the time of preparation of a quantum state) to final outcomes $\omega$ of quantum measurements. The mapping depends on macroscopic experimental settings $M$, and fixes the outcome for each run of the experiment. Bell's theorem of 1964 shows that, for entangled quantum states of widely separated systems, the mapping must be nonlocal: some outcomes for (at least) one system must depend on the setting for another distant system.

In a viable theory, the statistics of quantum measurement outcomes – over an ensemble of experimental trials with fixed settings $M$ – will agree with quantum theory for some special distribution $\rho_{\mathrm{QT}}(\lambda)$ of hidden variables. For example, expectation values will coincide with the predictions of the Born rule

$$\langle \omega \rangle_{\mathrm{QT}} \equiv \int \mathrm{d}\lambda\, \rho_{\mathrm{QT}}(\lambda)\omega(M, \lambda) = \mathrm{tr}(\hat{\rho}\hat{\Omega})$$

for an appropriate density operator $\hat{\rho}$ and Hermitian observable $\hat{\Omega}$. (As is customary in this context, $\int \mathrm{d}\lambda$ is to be understood as a generalized sum.) However, given the mapping $\omega = \omega(M, \lambda)$ for individual trials, one may, in principle, consider nonstandard distributions $\rho(\lambda) \neq \rho_{\mathrm{QT}}(\lambda)$ that yield statistics outside the domain of ordinary quantum theory (Valentini 1991, 2002a). We may say that such distributions correspond to a state of quantum nonequilibrium.

Quantum nonequilibrium is characterized by the breakdown of a number of basic quantum constraints. In particular, nonlocal signals appear at the statistical level. We shall first illustrate this for the hidden-variables model of de Broglie and Bohm. Then we shall generalize the discussion to all (deterministic) hidden-variables theories.

At present there is no experimental evidence for quantum nonequilibrium in nature. However, from a hidden-variables perspective, it is natural to explore the theoretical properties of nonequilibrium distributions, and to search experimentally for the statistical anomalies associated with them.

From this point of view, quantum theory is a special case of a wider physics, much as thermal physics is a special case of a wider (nonequilibrium) physics. (The special distribution $\rho_{\mathrm{QT}}(\lambda)$ is analogous to, say, Maxwell's distribution of molecular speeds.) Quantum physics may be compared with the physics of global thermal equilibrium, which is characterized by constraints – such as the impossibility of converting heat into work (in the absence of temperature differences) – that are not fundamental but contingent on the state. Similarly, quantum constraints such as statistical locality (the impossibility of converting entanglement into a practical signal) are seen as contingencies of $\rho_{\mathrm{QT}}(\lambda)$.

### Pilot-Wave Theory

The de Broglie–Bohm "pilot-wave theory" – as it was originally called by de Broglie, who first presented it at the Fifth Solvay Congress in 1927 – is the classic example of a deterministic hidden-variables theory of broad scope (Bohm 1952, Bell 1987, Holland 1993). We shall use it to illustrate the above ideas. Later, the discussion will be generalized to arbitrary theories.

In pilot-wave dynamics, an individual closed system with (configuration-space) wave function $\Psi(X, t)$ satisfying the Schrödinger equation

$$\mathrm{i}\hbar \frac{\partial \Psi}{\partial t} = \hat{H}\Psi \qquad [1]$$

has an actual configuration $X(t)$ with velocity

$$\dot{X}(t) = \frac{J(X, t)}{|\Psi(X, t)|^2} \qquad [2]$$

where $J = J[\Psi] = J(X, t)$ satisfies the continuity equation

$$\frac{\partial |\Psi|^2}{\partial t} + \nabla \cdot J = 0 \qquad [3]$$

(which follows from [1]). In quantum theory, $J$ is the "probability current." In pilot-wave theory, $\Psi$ is an objective physical field (on configuration space) guiding the motion of an individual system.

Here, the objective state (or ontology) for a closed system is given by $\Psi$ and $X$. A probability distribution for $X$ – discussed below – completes an unambiguous specification S (as mentioned in the introduction).

Pilot-wave dynamics may be applied to any quantum system with a locally conserved current in configuration space. Thus, $X$ may represent a many-body system, or the configuration of a continuous field, or perhaps some other entity.

For example, at low energies, for a system of $N$ particles with positions $\boldsymbol{x}_i(t)$ and masses $m_i\,(i = 1, 2, \ldots, N)$, with an external potential $V$, [1] (with $X \equiv (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)$) reads

$$i\hbar \frac{\partial \Psi}{\partial t} = \sum_{i=1}^{N} -\frac{\hbar^2}{2m_i} \nabla_i^2 \Psi + V\Psi \qquad [4]$$

while [2] has components

$$\frac{d\boldsymbol{x}_i}{dt} = \frac{\hbar}{m_i} \mathrm{Im}\left(\frac{\nabla_i \Psi}{\Psi}\right) = \frac{\nabla_i S}{m_i} \qquad [5]$$

(where $\Psi = |\Psi| e^{(i/\hbar)S}$).

In general, [1] and [2] determine $X(t)$ for an individual system, given the initial conditions $X(0), \Psi(X, 0)$ at $t = 0$. For an arbitrary initial distribution $P(X, 0)$, over an ensemble with the same wave function $\Psi(X, 0)$, the evolution $P(X, t)$ of the distribution is given by the continuity equation

$$\frac{\partial P}{\partial t} + \nabla \cdot (P\dot{X}) = 0 \qquad [6]$$

The outcome of an experiment is determined by $X(0), \Psi(X, 0)$, which may be identified with $\lambda$. For an ensemble with the same $\Psi(X, 0)$, we have $\lambda = X(0)$.

**Quantum equilibrium** From [3] and [6], if we assume $P(X, 0) = |\Psi(X, 0)|^2$ at $t = 0$, we obtain $P(X, t) = |\Psi(X, t)|^2$ – the Born-rule distribution of configurations – at all times $t$.

Quantum measurements are, like any other process, described and explained in terms of evolving configurations. For measurement devices whose pointer readings reduce to configurations, the distribution of outcomes of quantum measurements will match the statistical predictions of quantum theory (Bohm 1952, Bell 1987, Dürr *et al.* 2003). Thus, quantum theory emerges phenomenologically for a "quantum equilibrium" ensemble with distribution $P(X, t) = |\Psi(X, t)|^2$ (or $\rho(\lambda) = \rho_{\mathrm{QT}}(\lambda)$).

**Quantum nonequilibrium** In principle, as we saw for general hidden-variables theories, we may consider a nonequilibrium distribution $P(X, 0) \neq |\Psi(X, 0)|^2$ of initial configurations while retaining the same deterministic dynamics [1], [2] for individual systems (Valentini 1991). The time evolution of $P(X, t)$ will be determined by [6].

As we shall see, in appropriate circumstances (with a sufficiently complicated velocity field $\dot{X}$), [6] generates relaxation $P \to |\Psi|^2$ on a coarse-grained level, much as the analogous classical evolution on phase space generates thermal relaxation. But for as long as the ensemble is in nonequilibrium, the statistics of outcomes of quantum measurements will disagree with quantum theory.

Quantum nonequilibrium may have existed in the very early universe, with relaxation to equilibrium occurring soon after the big bang. Thus, a hidden-variables analog of the classical thermodynamic "heat death of the universe" may have actually taken place (Valentini 1991). Even so, relic cosmological particles that decoupled sufficiently early could still be in nonequilibrium today, as suggested by Valentini in 1996 and 2001. It has also been speculated that nonequilibrium could be generated in systems entangled with degrees of freedom behind a black-hole event horizon (Valentini 2004a).

Experimental searches for nonequilibrium have been proposed. Nonequilibrium could be detected by the statistical analysis of random samples of particles taken from a parent population of (for example) relics from the early universe. Once the parent distribution is known, the rest of the population could be used as a resource, to perform tasks that are currently impossible (Valentini 2002b).

### *H*-Theorem: Relaxation to Equilibrium

Before discussing the potential uses of nonequilibrium, we should first explain why all systems probed so far have been found in the equilibrium state $P = |\Psi|^2$. This distribution may be accounted for along the lines of classical statistical mechanics, noting that all currently accessible systems have had a long and violent astrophysical history.

Dividing configuration space into small cells, and introducing coarse-grained quantities $\bar{P}, \overline{|\Psi|^2}$, a general argument for relaxation $\bar{P} \to \overline{|\Psi|^2}$ is based on an

analog of the classical coarse-graining $H$-theorem. The coarse-grained $H$-function

$$\bar{H} = \int dX \ \bar{P} \ln(\bar{P}/\overline{|\Psi|^2}) \qquad [7]$$

(minus the relative entropy of $\bar{P}$ with respect to $|\Psi|^2$) obeys the $H$-theorem (Valentini 1991)

$$\bar{H}(t) \leq \bar{H}(0)$$

(assuming no initial fine-grained microstructure in $P$ and $|\Psi|^2$). Here, $\bar{H} \geq 0$ for all $\bar{P}, \overline{|\Psi|^2}$ and $\bar{H} = 0$ if and only if $\bar{P} = \overline{|\Psi|^2}$ everywhere.

The $H$-theorem expresses the fact that $P$ and $|\Psi|^2$ behave like two "fluids" that are "stirred" by the same velocity field $\dot{X}$, so that $P$ and $|\Psi|^2$ tend to become indistinguishable on a coarse-grained level. Like its classical analog, the theorem provides a general understanding of how equilibrium is approached, while not proving that equilibrium is actually reached. (And of course, for some simple systems – such as a particle in the ground state of a box, for which the velocity field $\nabla S/m$ vanishes – there is no relaxation at all.) A strict decrease of $\bar{H}(t)$ immediately after $t = 0$ is guaranteed if $\dot{X}_0 \cdot \nabla(P_0/|\Psi_0|^2)$ has nonzero spatial variance over a coarse-graining cell, as shown by Valentini in 1992 and 2001.

A relaxation timescale $\tau$ may be defined by $1/\tau^2 \equiv -(d^2\bar{H}/dt^2)_0/\bar{H}_0$. For a single particle with quantum energy spread $\Delta E$, a crude estimate given by Valentini in 2001 yields $\tau \sim (1/\varepsilon)\hbar^2/m^{1/2}(\Delta E)^{3/2}$, where $\varepsilon$ is the coarse-graining length. For wave functions that are superpositions of many energy eigenfunctions, the velocity field (generally) varies rapidly, and detailed numerical simulations (in two dimensions) show that relaxation occurs with an approximately exponential decay $\bar{H}(t) \approx \bar{H}_0 e^{-t/t_c}$, with a time constant $t_c$ of order $\tau$ (Valentini and Westman 2005).

Equilibrium is then to be expected for particles emerging from the violence of the big bang. The possibility is still open that relics from very early times may not have reached equilibrium before decoupling.

## Nonlocal Signaling

We now show how nonequilibrium, if it were ever discovered, could be used for nonlocal signaling.

Pilot-wave dynamics is nonlocal. For a pair of particles $A$, $B$ with entangled wave function $\Psi(\boldsymbol{x}_A, \boldsymbol{x}_B, t)$, the velocity $\dot{\boldsymbol{x}}_A(t) = \nabla_A S(\boldsymbol{x}_A, \boldsymbol{x}_B, t)/m_A$ of $A$ depends instantaneously on $\boldsymbol{x}_B$, and local operations at $B$ – such as switching on a potential – instantaneously affect the motion of $A$. For an ensemble $P(\boldsymbol{x}_A, \boldsymbol{x}_B, t) = |\Psi(\boldsymbol{x}_A, \boldsymbol{x}_B, t)|^2$, local operations at $B$ have no statistical effect at $A$: the individual nonlocal effects vanish upon averaging over an equilibrium ensemble.

Nonlocality is (generally) hidden by statistical noise only in quantum equilibrium. If instead $P(\boldsymbol{x}_A, \boldsymbol{x}_B, 0) \neq |\Psi(\boldsymbol{x}_A, \boldsymbol{x}_B, 0)|^2$, a local change in the Hamiltonian at $B$ generally induces an instantaneous change in the marginal $p_A(\boldsymbol{x}_A, t) \equiv \int d^3\boldsymbol{x}_B P(\boldsymbol{x}_A, \boldsymbol{x}_B, t)$ at $A$. For example, in one dimension a sudden change $\hat{H}_B \to \hat{H}'_B$ in the Hamiltonian at $B$ induces a change $\Delta p_A \equiv p_A(x_A, t) - p_A(x_A, 0)$ (for small $t$) (Valentini 1991),

$$\Delta p_A = -\frac{t^2}{4m}\frac{\partial}{\partial x_A}\left(a(x_A)\int dx_B \ b(x_B) \times \frac{P(x_A, x_B, 0) - |\Psi(x_A, x_B, 0)|^2}{|\Psi(x_A, x_B, 0)|^2}\right) \qquad [8]$$

(Here $m_A = m_B = m$, $a(x_A)$ depends on $\Psi(x_A, x_B, 0)$, while $b(x_B)$ also depends on $\hat{H}'_B$ and vanishes if $\hat{H}'_B = \hat{H}_B$.) The signal is generally nonzero if $P_0 \neq |\Psi_0|^2$.

Nonlocal signals do not lead to causal paradoxes if, at the hidden-variable level, there is a preferred foliation of spacetime with a time parameter that defines a fundamental causal sequence. Such signals, if they were observed, would define an absolute simultaneity as discussed by Valentini in 1992 and 2005. Note that in pilot-wave field theory, Lorentz invariance emerges as a phenomenological symmetry of the equilibrium state, conditional on the structure of the field-theoretical Hamiltonian (as discussed by Bohm and Hiley in 1984, Bohm, Hiley and Kaloyerou in 1987, and Valentini in 1992 and 1996).

## Subquantum Measurement

In principle, nonequilibrium particles could also be used to perform "subquantum measurements" on ordinary, equilibrium systems. We illustrate this with an exactly solvable one-dimensional model (Valentini 2002b).

Consider an apparatus "pointer" coordinate $y$, with known wave function $g_0(y)$ and known (ensemble) distribution $\pi_0(y) \neq |g_0(y)|^2$, where $\pi_0(y)$ has been deduced by statistical analysis of random samples from a parent population with known wave function $g_0(y)$. (We assume that relaxation may be neglected: for example, if $g_0$ is a box ground state, $\dot{y} = 0$ and $\pi_0(y)$ is static.) Consider also a "system" coordinate $x$ with known wave function $\psi_0(x)$ and known distribution $\rho_0(x) = |\psi_0(x)|^2$. If $\pi_0(y)$ is arbitrarily narrow, $x_0$ can be measured without

disturbing $\psi_0(x)$, to arbitrary accuracy (violating the uncertainty principle).

To do this, at $t = 0$ we switch on an interaction Hamiltonian $\hat{H} = a\hat{x}\hat{p}_y$, where $a$ is a constant and $p_y$ is canonically conjugate to $y$. For relatively large $a$, we may neglect the Hamiltonians of $x$ and $y$. For $\Psi = \Psi(x, y, t)$, we then have $\partial\Psi/\partial t = -ax\partial\Psi/\partial y$. For $|\Psi|^2$ we have the continuity equation $\partial|\Psi|^2/\partial t = -ax\partial|\Psi|^2/\partial y$, which implies the hidden-variable velocity fields $\dot{x} = 0, \dot{y} = ax$ and trajectories $x(t) = x_0$, $y(t) = y_0 + ax_0t$.

The initial product $\Psi_0(x, y) = \psi_0(x)g_0(y)$ evolves into $\Psi(x, y, t) = \psi_0(x)g_0(y - axt)$. For $at \to 0$ (with $a$ large but fixed), $\Psi(x, y, t) \to \psi_0(x)g_0(y)$ and $\psi_0(x)$ is undisturbed: for small $at$, a standard quantum pointer with the coordinate $y$ would yield negligible information about $x_0$. Yet, for arbitrarily small $at$, the hidden-variable pointer coordinate $y(t) = y_0 + ax_0t$ does contain complete information about $x_0$ (and $x(t) = x_0$). This "subquantum" information will be visible to us if $\pi_0(y)$ is sufficiently narrow.

For, over an ensemble of similar experiments, with initial joint distribution $P_0(x, y) = |\psi_0(x)|^2\pi_0(y)$ (equilibrium for $x$ and nonequilibrium for $y$), the continuity equation $\partial P/\partial t = -ax\partial P/\partial y$ implies that $P(x, y, t) = |\psi_0(x)|^2\pi_0(y - axt)$. If $\pi_0(y)$ is localized around $y = 0$ ($\pi_0(y) = 0$ for $|y| > w/2$), then a standard (faithful) measurement of $y$ with result $y_{\text{meas}}$ will imply that $x$ lies in the interval ($y_{\text{meas}}/at - w/2at$, $y_{\text{meas}}/at + w/2at$) (so that $P(x, y, t) \neq 0$). Taking the simultaneous limits $at \to 0$, $w \to 0$, with $w/at \to 0$, the midpoint $y_{\text{meas}}/at \to x_0$ (since $y_{\text{meas}} = y_0 + ax_0t$ and $|y_0| \leq w/2$), while the error $w/2at \to 0$.

If $w$ is arbitrarily small, a sequence of such measurements will determine the hidden trajectory $x(t)$ without disturbing $\psi(x, t)$, to arbitrary accuracy.

### Subquantum Information and Computation

From a hidden-variables perspective, immense physical resources are hidden from us by equilibrium statistical noise. Quantum nonequilibrium would probably be as useful technologically as thermal or chemical nonequilibrium.

**Distinguishing nonorthogonal states** In quantum theory, nonorthogonal states $|\psi_1\rangle, |\psi_2\rangle$ ($\langle\psi_1|\psi_2\rangle \neq 0$) cannot be distinguished without disturbing them. This theorem breaks down in quantum nonequilibrium (Valentini 2002b). For example, if $|\psi_1\rangle, |\psi_2\rangle$ are distinct states of a single spinless particle, then the associated de Broglie–Bohm velocity fields will in general be different, even if $\langle\psi_1|\psi_2\rangle \neq 0$, and so will the hidden-variable trajectories. Subquantum measurement of the trajectories could then distinguish the states $|\psi_1\rangle, |\psi_2\rangle$.

**Breaking quantum cryptography** The security of standard protocols for quantum key distribution depends on the validity of the laws of quantum theory. These protocols would become insecure given the availability of nonequilibrium systems (Valentini 2002b).

The protocols known as BB84 and B92 depend on the impossibility of distinguishing nonorthogonal quantum states without disturbing them. An eavesdropper in possession of nonequilibrium particles could distinguish the nonorthogonal states being transmitted between two parties, and so read the supposedly secret key. Further, if subquantum measurements allow an eavesdropper to predict quantum measurement outcomes at each "wing" of a (bipartite) entangled state, then the EPR (Einstein–Podolsky–Rosen) protocol also becomes insecure.

**Subquantum computation** It has been suggested that nonequilibrium physics would be computationally more powerful than quantum theory, because of the ability to distinguish nonorthogonal states (Valentini 2002b). However, this ability depends on the (less-than-quantum) dispersion $w$ of the nonequilibrium ensemble. A well-defined model of computational complexity requires that the resources be quantified in some way. Here, a key question is how the required $w$ scales with the size of the computational task. So far, no rigorous results are known.

### Extension to All Deterministic Hidden-Variables Theories

Let us now discuss arbitrary (deterministic) theories.

**Nonlocal signaling** Consider a pair of two-state quantum systems $A$ and $B$, which are widely separated and in the singlet state. Quantum measurements of observables $\hat{\sigma}_A \equiv \boldsymbol{m}_A \cdot \hat{\boldsymbol{\sigma}}_A, \hat{\sigma}_B \equiv \boldsymbol{m}_B \cdot \hat{\boldsymbol{\sigma}}_B$ (where $\boldsymbol{m}_A, \boldsymbol{m}_B$ are unit vectors in Bloch space and $\hat{\boldsymbol{\sigma}}_A, \hat{\boldsymbol{\sigma}}_B$ are Pauli spin operators) yield outcomes $\sigma_A, \sigma_B = \pm 1$, in the ratio $1:1$ at each wing, with a correlation $\langle\hat{\sigma}_A\hat{\sigma}_B\rangle = -\boldsymbol{m}_A \cdot \boldsymbol{m}_B$. Bell's theorem shows that for a hidden-variables theory to reproduce this correlation – upon averaging over an equilibrium ensemble with distribution $\rho_{\text{QT}}(\lambda)$ – it must take the nonlocal form

$$\sigma_A = \sigma_A(\boldsymbol{m}_A, \boldsymbol{m}_B, \lambda), \qquad \sigma_B = \sigma_B(\boldsymbol{m}_A, \boldsymbol{m}_B, \lambda) \qquad [9]$$

More precisely, to obtain $\langle\sigma_A\sigma_B\rangle_{\text{QT}} = -\boldsymbol{m}_A \cdot \boldsymbol{m}_B$ (where $\langle\sigma_A\sigma_B\rangle_{\text{QT}} \equiv \int d\lambda\rho_{\text{QT}}(\lambda)\sigma_A\sigma_B$), at least one of

$\sigma_A, \sigma_B$ must depend on the measurement setting at the distant wing. Without loss of generality, we assume that $\sigma_A$ depends on $\boldsymbol{m}_B$.

For an arbitrary nonequilibrium ensemble with distribution $\rho(\lambda) \neq \rho_{\mathrm{QT}}(\lambda)$, in general $\langle \sigma_A \sigma_B \rangle \equiv \int \mathrm{d}\lambda\, \rho(\lambda) \sigma_A \sigma_B$ differs from $-\boldsymbol{m}_A \cdot \boldsymbol{m}_B$, and the outcomes $\sigma_A, \sigma_B = \pm 1$ occur in a ratio different from $1:1$. Further, a change of setting $\boldsymbol{m}_B \to \boldsymbol{m}'_B$ at $B$ will generally induce a change in the outcome statistics at $A$, yielding a nonlocal signal at the statistical level. To see this, note that, in a nonlocal theory, the "transition sets"

$$T_A(-,+) \equiv \{\lambda | \sigma_A(\boldsymbol{m}_A, \boldsymbol{m}_B, \lambda) = -1, \; \sigma_A(\boldsymbol{m}_A, \boldsymbol{m}'_B, \lambda) = +1\}$$
$$T_A(+,-) \equiv \{\lambda | \sigma_A(\boldsymbol{m}_A, \boldsymbol{m}_B, \lambda) = +1, \; \sigma_A(\boldsymbol{m}_A, \boldsymbol{m}'_B, \lambda) = -1\}$$

cannot be empty for arbitrary settings. Yet, in quantum equilibrium, the outcomes $\sigma_A = \pm 1$ occur in the ratio $1:1$ for all settings, so the transition sets must have equal equilibrium measure, $\mu_{\mathrm{QT}}[T_A(-,+)] = \mu_{\mathrm{QT}}[T_A(+,-)]$ ($\mathrm{d}\mu_{\mathrm{QT}} \equiv \rho_{\mathrm{QT}}(\lambda)\mathrm{d}\lambda$). That is, the fraction of the equilibrium ensemble making the transition $\sigma_A = -1 \to \sigma_A = +1$ under $\boldsymbol{m}_B \to \boldsymbol{m}'_B$ must equal the fraction making the reverse transition $\sigma_A = +1 \to \sigma_A = -1$. (This "detailed balancing" is analogous to the principle of detailed balance in statistical mechanics.) Since $T_A(-,+), T_A(+,-)$ are fixed by the deterministic mapping, they are independent of the ensemble distribution $\rho(\lambda)$. Thus, for $\rho(\lambda) \neq \rho_{\mathrm{QT}}(\lambda)$, in general $\mu[T_A(-,+)] \neq \mu[T_A(+,-)]$ ($\mathrm{d}\mu \equiv \rho(\lambda) d\lambda$): the fraction of the nonequilibrium ensemble making the transition $\sigma_A = -1 \to \sigma_A = +1$ will not in general balance the fraction making the reverse transition. The outcome ratio at $A$ will then change under $\boldsymbol{m}_B \to \boldsymbol{m}'_B$ and there will be an instantaneous signal at the statistical level from $B$ to $A$ (Valentini 2002a).

Thus, in any deterministic hidden-variables theory, nonequilibrium distributions $\rho(\lambda) \neq \rho_{\mathrm{QT}}(\lambda)$ generally allow entanglement to be used for nonlocal signalling (just as, in ordinary statistical physics, differences of temperature make it possible to convert heat into work).

**Experimental signature of nonequilibrium** Quantum expectations are additive, $\langle c_1 \hat{\Omega}_1 + c_2 \hat{\Omega}_2 \rangle = c_1 \langle \hat{\Omega}_1 \rangle + c_2 \langle \hat{\Omega}_2 \rangle$, even for noncommuting observables ($[\hat{\Omega}_1, \hat{\Omega}_2] \neq 0$, with $c_1$, $c_2$ real). As emphasized by Bell in 1966, this seemingly trivial consequence of the (linearity of the) Born rule $\langle \hat{\Omega} \rangle = \mathrm{tr}(\hat{\rho}\hat{\Omega})$ is remarkable because it relates statistics from distinct, "incompatible" experiments. In nonequilibrium, such additivity generically breaks down (Valentini 2004b).

Further, for a two-state system with observables $\boldsymbol{m} \cdot \hat{\boldsymbol{\sigma}}$, the "dot-product" structure of the quantum expectation $\langle \boldsymbol{m} \cdot \hat{\boldsymbol{\sigma}} \rangle = \mathrm{tr}(\hat{\rho} \boldsymbol{m} \cdot \hat{\boldsymbol{\sigma}}) = \boldsymbol{m} \cdot \boldsymbol{P}$ (for some Bloch vector $\boldsymbol{P}$) is equivalent to expectation additivity (Valentini 2004b). Nonadditive expectations then provide a convenient signature of nonequilibrium for any two-state system. For example, the sinusoidal modulation of the quantum transmission probability for a single photon through a polarizer

$$p^{+}_{\mathrm{QT}}(\Theta) = \tfrac{1}{2}(1 + \langle \boldsymbol{m} \cdot \hat{\boldsymbol{\sigma}} \rangle) = \tfrac{1}{2}(1 + P \cos 2\Theta) \qquad [10]$$

(where an angle $\theta$ on the Bloch sphere corresponds to a physical angle $\Theta = \theta/2$) will generically break down in nonequilibrium. Deviations from [10] would provide an unambiguous violation of quantum theory (Valentini 2004b).

Such deviations were searched for by Papaliolios in 1967, using laboratory photons and successive polarization measurements over very short times, to test a hidden-variables theory (distinct from pilot-wave theory) due to Bohm and Bub (1966), in which quantum measurements generate nonequilibrium for short times. Experimentally, successive measurements over timescales $\sim 10^{-13}$ s agreed with the (quantum) sinusoidal modulation $\cos^2 \Theta$ to $\lesssim 1\%$. Similar tests might be performed with photons of a more exotic origin.

## Continuous Spontaneous Localization Model (CSL)

The basic postulate of CSL is that the state vector $|\psi, t\rangle$ represents reality. Since, for example, in describing a measurement, the usual Schrödinger evolution readily takes a real state into a nonreal state, that is, into a superposition of real states (such as apparatus states describing different experimental outcomes), CSL requires a modification of Schrödinger's evolution. To the Hamiltonian is added a term which depends upon a classical randomly fluctuating field $w(\boldsymbol{x}, t)$ and a mass-density operator $\hat{A}(\boldsymbol{x}, t)$. This term acts to collapse a superposition of states, which differ in their spatial distribution of mass density, to one of these states. The rate of collapse is very slow for a superposition involving a few particles, but very fast for a superposition of macroscopically different states. Thus, very rapidly, what you see (in nature) is what you get (from the theory). Each state vector evolving under each $w(\boldsymbol{x}, t)$ corresponds to a realizable state, and a rule is given for how to associate a probability with each. In this way, an

unambiguous specification S, as mentioned in the introduction, is achieved.

### Requirements for Stochastic Collapse Dynamics

Consider a normalized state vector $|\psi, t\rangle = \sum_n \alpha_n(t)|a_n\rangle (\langle a_n|a_{n'}\rangle = \delta_{nn'})$ which undergoes a stochastic dynamical collapse process. This means that, starting from the initial superposition at $t=0$, for each run of the process, the squared amplitudes $x_n(t) \equiv |\alpha_n(t)|^2$ fluctuate until all but one vanish, that is, $x_m(\infty)=1, (x_{\neq m}(\infty)=0)$ with probability $x_m(0)$.

This may be achieved simply, assuming negligible effect of the usual Schrödinger evolution, if the stochastic process enjoys the following properties (Pearle 1979):

$$\sum_n x_n(t) = 1 \quad [11a]$$

$$\overline{x_n(t)} = x_n(0) \quad [11b]$$

$$\overline{x_n(\infty)x_m(\infty)} = 0 \quad \text{for } m \neq n \quad [11c]$$

where the overbar indicates the ensemble average at the indicated time. The only way that a sum of products of non-negative terms can vanish is for at least one term in each product to vanish. Thus, according to [11c], for each run, at least one of each pair $\{x_n(\infty), x_m(\infty)\}(n \neq m)$ must vanish. This means that at most one $x_n(\infty)$ might not vanish and, by [11a], applied at $t=\infty$, one $x_n(\infty)$ must not vanish and, in fact, must equal 1: hence, each run produces collapse. Now, let the probability of the outcome $\{x_n(\infty)=1, x_{\neq n}(\infty)=0\}$ be denoted $P_n$. Since $\overline{x_n(\infty)} = 1 \cdot P_n + \sum_{m \neq n} 0 \cdot P_m = P_n$ then, according to the Martingale property [11b], applied at $t=\infty, P_n = x_n(0)$: hence, the ensemble of runs produces the probability postulated by the usual "collapse rule" of standard quantum theory.

A (nonquantum) stochastic process which obeys these equations is the gambler's ruin game. Suppose one gambler initially possesses the fraction $x_1(0)$ of their joint wealth, and the other has the fraction $x_2(0)$. They toss a coin: heads, a dollar goes from gambler 1 to gambler 2, tails the dollar goes the other way. [11a] is satisfied since the sum of money in the game remains constant, [11b] holds because it is a fair game, and [11c] holds because each game eventually ends. Thus, gambler $i$ wins all the money with probability $x_i(0)$.

### CSL in Essence

Consider the (nonunitary) Schrödinger picture evolution equation

$$|\psi, t\rangle_w = \mathcal{T} \exp\left(-\int_0^t \mathrm{d}t' \{i\hat{H} + (4\lambda)^{-1}[w(t') - 2\lambda\hat{A}]^2\}\right)|\psi, 0\rangle \quad [12]$$

where $\hat{H}$ is the usual Hamiltonian, $w(t')$ is an arbitrary function of white noise class, $\hat{A}$ is a Hermitian operator $(\hat{A}|a_n\rangle = a_n|a_n\rangle), \lambda$ is a collapse rate parameter, $\mathcal{T}$ is the time-ordering operator and $\hbar=1$. Associated with this, the probability rule

$$P_t(w)Dw \equiv {}_w\langle\psi, t|\psi, t\rangle_w \prod_{j=0}^{t/\mathrm{d}t} \mathrm{d}w(t_j)/(2\pi\lambda/\mathrm{d}t)^{1/2} \quad [13]$$

is defined, which gives the probability that nature chooses a noise which lies in the range $\{w(t'), w(t') + \mathrm{d}w(t')\}$ for $0 \leq t' \leq t$ (for calculational purposes, time is discretized, with $t_0=0$).

Equations [12] and [13] contain the essential features of CSL, and are all that is needed to discuss the simplest collapse behavior. Set $\hat{H}=0$, so there is no competition between collapse and the usual Schrödinger evolution, and let the initial state vector be $|\psi, 0\rangle = \sum_n \alpha_n|a_n\rangle$. Equations [12] and [13] become

$$|\psi, t\rangle_w = \sum_n \alpha_n|a_n\rangle \exp\left(-(4\lambda)^{-1}\int_0^t \mathrm{d}t'[w(t') - 2\lambda a_n]^2\right) \quad [14a]$$

$$P_t(w) = \sum_n |\alpha_n|^2 \exp\left(-(2\lambda)^{-1}\int_0^t \mathrm{d}t'[w(t') - 2\lambda a_n]^2\right) \quad [14b]$$

When the unnormalized state vector in [14a] is divided by $P_t^{1/2}(w)$ and so normalized, the squared amplitudes are

$$x_n(t) = |\alpha_n|^2 \exp\left(-(2\lambda)^{-1} \times \int_0^t \mathrm{d}t'[w(t') - 2\lambda a_n]^2\right)/P_t(w)$$

which are readily shown to satisfy [11a], [11b], and [11c] in the form $\overline{x_n^{1/2}(\infty)x_m^{1/2}(\infty)} = 0 (m \neq n)$ (which does not change the argument in the last subsection, but makes for an easier calculation). Thus, [14a] and [14b] describe collapse dynamics.

To describe collapse to a joint eigenstate of a set of mutually commuting operators $\hat{A}^r$, replace $(4\lambda)^{-1}[w(t') - 2\lambda\hat{A}]^2$ in the exponent of [12] by $\sum_r (4\lambda)^{-1}[w^r(t') - 2\lambda\hat{A}^r]^2$. The interaction picture state vector in this case is [12] multiplied by $\exp(\mathrm{i}\hat{H}t)$:

$$|\psi, t\rangle_w = \mathcal{T}\exp\left(-(4\lambda)^{-1}\int_0^l \mathrm{d}t' \times \sum_r [w^r(t') - 2\lambda\hat{A}^r(t')]^2\right)|\psi, 0\rangle \quad [15]$$

where $\hat{A}^r(t') \equiv \exp(\mathrm{i}\hat{H}t')\hat{A}^r\exp(-\mathrm{i}\hat{H}t')$. The density matrix follows from [15], and [13]:

$$\hat{\rho}(t) \equiv \int P_t(w)Dw|\psi, t\rangle_{w\,w}\langle\psi, t|/P_t(w) = \mathcal{T}\exp\left(-\lambda/2\int_0^t \mathrm{d}t' \times \sum_r [\hat{A}^r_{\mathrm{L}}(t') - \hat{A}^r_{\mathrm{R}}(t')]^2\right)\hat{\rho}(0) \quad [16]$$

where $\hat{A}^r_{\mathrm{L}}(t')(\hat{A}^r_{\mathrm{R}}(t'))$ appears to the left (right) of $\hat{\rho}(0)$, and is time-ordered (time reverse-ordered). In the example described by [14], the density matrix [16] is

$$\hat{\rho}(t) = \sum_{n,m} \mathrm{e}^{-(\lambda t/2)(a_n - a_m)^2}\alpha_n\alpha_m^*|a_n\rangle\langle a_m|$$

which encapsulates the ensemble's collapse behavior.

### CSL

The CSL proposal (Pearle 1989) is that collapse is engendered by distinctions between states at each point of space, so the index $r$ of $\hat{A}^r$ in [15] becomes $\boldsymbol{x}$,

$$|\psi, t\rangle_w = \mathcal{T}\exp\left(-(4\lambda)^{-1}\int_0^t\int \mathrm{d}t'\,\mathrm{d}\boldsymbol{x}' \times [w(\boldsymbol{x}', t') - 2\lambda\hat{A}(\boldsymbol{x}', t')]^2\right)|\psi, 0\rangle \quad [17]$$

and the distinction looked at is mass density. However, one cannot make the choice $\hat{A}(\boldsymbol{x}, 0) = \hat{M}(\boldsymbol{x})$, where $\hat{M}(\boldsymbol{x}) = \sum_i m_i\hat{\xi}_i^\dagger(\boldsymbol{x})\hat{\xi}_i(\boldsymbol{x})$ is the mass-density operator ($m_i$ is the mass of the $i$th type of particle, so $m_{\mathrm{e}}, m_{\mathrm{p}}, m_{\mathrm{n}}, \ldots$ are the masses, respectively, of electrons, protons, neutrons…, and $\hat{\xi}_i^\dagger(\boldsymbol{x})$ is the creation operator for such a particle at location $\boldsymbol{x}$), because this entails an infinite rate of energy increase of particles ([23] with $a = 0$). Instead, adapting a "Gaussian smearing" idea from the Ghirardi *et al.* (1986) spontaneous localization (SL) model (see the subsection "Spontaneous localization model"), choose $\hat{A}^{\boldsymbol{x}}$ as, essentially, proportional to the mass in a sphere of radius $a$ about $\boldsymbol{x}$:

$$\hat{A}(\boldsymbol{x}, t) \equiv \mathrm{e}^{\mathrm{i}\hat{H}t}\frac{1}{(\pi a^2)^{3/4}} \times \int \mathrm{d}\boldsymbol{z}\frac{\hat{M}(\boldsymbol{z})}{m_{\mathrm{p}}}\mathrm{e}^{-(2a^2)^{-1}(\boldsymbol{x}-\boldsymbol{z})^2}\mathrm{e}^{-\mathrm{i}\hat{H}t} \quad [18]$$

The parameter value choices of SL, $\lambda \approx 10^{-16}\,\mathrm{s}^{-1}$ (according to [17] and [18], the collapse rate for protons) and $a \approx 10^{-5}$ cm are, so far, consistent with experiment (see the next subsection), and will be adopted here.

The density matrix associated with [17] is, as in [16],

$$\hat{\rho}(t) = \mathcal{T}\exp\left(-(\lambda/2)\int_0^t \mathrm{d}t'\,\mathrm{d}\boldsymbol{x}'[\hat{A}_{\mathrm{L}}(\boldsymbol{x}', t') - \hat{A}_{\mathrm{R}}(\boldsymbol{x}', t')]^2\right)\hat{\rho}(0) \quad [19]$$

which satisfies the differential equation

$$\frac{\mathrm{d}\hat{\rho}(t)}{\mathrm{d}t} = -\frac{\lambda}{2}\int \mathrm{d}\boldsymbol{x}'[\hat{A}(\boldsymbol{x}', t), [\hat{A}(\boldsymbol{x}', t), \hat{\rho}(t)]] \quad [20]$$

of Lindblad–Kossakowski form.

### Consequences of CSL

Since the state vector dynamics of CSL is different from that of standard quantum theory, there are phenomena for which the two make different predictions, allowing for experimental tests. Consider an $N$-particle system with position operators $\hat{\boldsymbol{X}}_i(\hat{\boldsymbol{X}}_i|\boldsymbol{x}\rangle = \boldsymbol{x}_i|\boldsymbol{x}\rangle)$. Substitution of $\hat{A}(\boldsymbol{x}')$ from [18] in the Schrödinger picture version of [20], integration over $\boldsymbol{x}'$, and utilization of

$$f(\boldsymbol{z})\hat{M}(\boldsymbol{z})|\boldsymbol{x}\rangle = \sum_{i=1}^N m_i f(\hat{\boldsymbol{X}}_i)\delta(\boldsymbol{z} - \hat{\boldsymbol{X}}_i)|\boldsymbol{x}\rangle$$

results in

$$\frac{\mathrm{d}\hat{\rho}(t)}{\mathrm{d}t} = -\mathrm{i}[\hat{\rho}(t), \hat{H}] - \frac{\lambda}{2}\sum_{i=1}^N\sum_{j=1}^N \frac{m_i}{m_{\mathrm{p}}}\frac{m_j}{m_{\mathrm{p}}} \times \left[\mathrm{e}^{-(4a^2)^{-1}(\hat{\boldsymbol{X}}_{\mathrm{L}i} - \hat{\boldsymbol{X}}_{\mathrm{L}j})^2} + \mathrm{e}^{-(4a^2)^{-1}(\hat{\boldsymbol{X}}_{\mathrm{R}i} - \hat{\boldsymbol{X}}_{\mathrm{R}j})^2} - 2\mathrm{e}^{-(4a^2)^{-1}(\hat{\boldsymbol{X}}_{\mathrm{L}i} - \hat{\boldsymbol{X}}_{\mathrm{R}j})^2}\right]\hat{\rho}(t) \quad [21]$$

which is a useful form for calculations first suggested by Pearle and Squires in 1994.

**Interference** Consider the collapse rate of an initial state $|\phi\rangle = \alpha_1|1\rangle + \alpha_2|2\rangle$, where $|1\rangle, |2\rangle$ describe a

clump of matter, of size $\ll a$, at different locations with separation $\gg a$. Electrons may be neglected because of their small collapse rate compared to the much more massive nucleons, and the nucleon mass difference may be neglected. In using [21] to calculate $\mathrm{d}\langle 1|\hat{\rho}(t)|2\rangle/\mathrm{d}t$, since $\exp[-(4a^2)^{-1}(\hat{\boldsymbol{X}}_i - \hat{\boldsymbol{X}}_j)^2] \approx 1$ when acting on state $|1\rangle$ or $|2\rangle$, and $\approx 0$ when $\hat{\boldsymbol{X}}_i$ acts on $|1\rangle$ and $\hat{\boldsymbol{X}}_j$ acts on $|2\rangle$, [21] yields, for $N$ nucleons, the collapse rate $\lambda N^2$:

$$\frac{\mathrm{d}\langle 1|\hat{\rho}(t)|2\rangle}{\mathrm{d}t} = -\mathrm{i}\langle 1|[\hat{\rho}(t), \hat{H}]|2\rangle - \lambda N^2\langle 1|\hat{\rho}(t)|2\rangle \quad [22]$$

If the clump undergoes a two-slit interference experiment, where the size and separation conditions above are satisfied for a time $\Delta T$, and if the result agrees with the standard quantum theory prediction to 1%, it also agrees with CSL provided $\lambda^{-1} > 100N^2\Delta T$. So far, interference experiments with $N$ as large as $\approx 10^3$ have been performed, by Nairz, Arndt, and Zeilinger in 2000. The SL value of $\lambda^{-1} \approx 10^{16}$ would be testable, that is, the quantum-predicted interference pattern would be "washed out" to 1% accuracy, if the clump were an $\approx 10^{-6}$ cm radius sphere of mercury, which contains $N \approx 10^8$ nucleons, interfered for $\Delta T = 0.01$ s. Currently envisioned but not yet performed experiments (e.g., by Marshall, Simon, Penrose, and Bouwmester in 2003) have been analyzed (e.g., by Bassi, Ippoliti, and Adler in 2004 and by Adler in 2005), which involve a superposition of a larger clump of matter in slightly displaced positions, entangled with a photon whose interference pattern is measured: these proposed experiments are still too crude to detect the SL value of $\lambda$, or the gravitationally based collapse rate proposed by Penrose in 1996 (see the next section and papers by Christian in 1999 and 2005).

**Bound state excitation** Collapse narrows wave packets, thereby imparting energy to particles. If $\hat{H} = \sum_{i=1}^{N} \hat{\boldsymbol{P}}_i^2/2m_i + \hat{V}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$, it is straightforward to calculate from [21] that

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\hat{H}\rangle \equiv \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{tr}[\hat{H}\hat{\rho}(t)] = \sum_{i=1}^{N} \frac{3\lambda\hbar^2}{4m_i a^2} \quad [23]$$

For a nucleon, the mean rate of energy increase is quite small, $\approx 3 \times 10^{-25}$ eV s$^{-1}$. However, deviations from the mean can be significantly greater.

Equation [21] predicts excitation of atoms and nuclei. Let $|E_0\rangle$ be an initial bound energy eigenstate. Expanding [21] in a power series in (bound state size/$a$)$^2$, the excitation rate of state $|E\rangle$ is

$$\Gamma \equiv \frac{\mathrm{d}\langle E|\hat{\rho}(t)|E\rangle}{\mathrm{d}t}\Big|_{t=0} = \frac{\lambda}{2a^2}\left\langle E\Big|\sum_{i=1}^{N}\frac{m_i}{m_\mathrm{p}}\hat{\boldsymbol{X}}_i\Big|E_0\right\rangle\left\langle E_0\Big|\sum_{i=1}^{N}\frac{m_i}{m_\mathrm{p}}\hat{\boldsymbol{X}}_i\Big|E\right\rangle + \mathrm{O}(\mathrm{size}/a)^4 \quad [24]$$

Since $|E_0\rangle, |E\rangle$ are eigenstates of the center-of-mass operator $\sum_{i=1}^{N} m_i\hat{\boldsymbol{X}}_i/\sum_{i=1}^{N} m_i$ with eigenvalue 0, the dipole contribution explicitly given in [24] vanishes identically. This leaves the quadrupole contribution as the leading term, which is too small to be measured at present.

However, the choice of $\hat{A}(\boldsymbol{x})$ as mass-density operator was made only after experimental indication. Let $g_i$ replace $m_i/m_\mathrm{p}$ in [21] and [24], so that $\lambda g_i^2$ is the collapse rate for the $i$th particle. Then, experiments looking for the radiation expected from "spontaneously" excited atoms and nuclei, in large amounts of matter for a long time, as shown by Collett, Pearle, Avignone, and Nussinov in 1995, Pearle, Ring, Collar, and Avignone in 1999, and Jones, Pearle, and Ring in 2004, have placed the following limits:

$$\left|\frac{g_\mathrm{e}}{g_\mathrm{p}} - \frac{m_\mathrm{e}}{m_\mathrm{p}}\right| < \frac{12m_\mathrm{e}}{m_\mathrm{p}}, \quad \left|\frac{g_\mathrm{n}}{g_\mathrm{p}} - \frac{m_\mathrm{n}}{m_\mathrm{p}}\right| < \frac{3(m_\mathrm{n} - m_\mathrm{p})}{m_\mathrm{p}}$$

**Random walk** According to [17] and [13], the center-of-mass wave packet, of a piece of matter of size $\approx a$ or smaller, containing $N$ nucleons, achieves equilibrium size $s$ in a characteristic time $\tau_s$, and undergoes a random walk through a root-mean-square distance $\Delta Q$:

$$s \approx \left[\frac{a^2\hbar}{\lambda m_\mathrm{p}N^3}\right]^{1/4}, \qquad \tau_s \approx \frac{Nm_\mathrm{p}s^2}{\hbar}$$
$$\Delta Q \approx \frac{\hbar\lambda^{1/2}t^{3/2}}{m_\mathrm{p}a} \quad [25]$$

The results in [25] were obtained by Collett and Pearle in 2003. These quantitative results can be qualitatively understood as follows.

In time $\Delta t$, the usual Schrödinger equation expands a wave packet of size $s$ to $\approx s + (\hbar/Nm_\mathrm{p}s)\Delta t$. CSL collapse, by itself, narrows the wave packet to $\approx s[1 - \lambda N^2(s/a)^2\Delta t]$. The condition of no change in $s$ is the result quoted above. $\tau_s$ is the time it takes the Schrödinger evolution to expand a wave packet near size $s$ to size $s$: $(\hbar/Nm_\mathrm{p}s)\tau_s \approx s$.

The $t^{3/2}$ dependence of $\Delta Q$ arises because this is a random walk without damping (unlike Brownian motion, where $\Delta Q \sim t^{1/2}$). The mean energy increase $\approx \lambda N \hbar^2 m_p^{-1} a^{-2} t$ of [23] implies the root-mean-square velocity increase $\approx [\lambda \hbar^2 m_p^{-2} a^{-2} t]^{1/2}$, whose product with $t$ is $\Delta Q$.

For example, a sphere of density $1\,\text{cm}^{-3}$ and radius $10^{-5}$ cm has $s \approx 4 \times 10^{-7}$ cm, $\tau_s \approx 0.6$ s and $\Delta Q \approx 5[t \text{ in days}]^{3/2}$ cm. At the low pressure of $5 \times 10^{-17}$ torr at 4.2 K reported by Gabrielse's group in 1990, the mean collision time with gas molecules is $\approx$80 min, over which $\Delta Q \approx 0.7$ mm. Thus, observation of this effect should be feasible.

### Further Remarks

It is possible to define energy for the $w(\boldsymbol{x}, t)$ field so that total energy is conserved: as the particles gain energy, the $w$-field loses energy, as shown by Pearle in 2005.

Attempts to construct a special-relativistic CSL-type model have not yet succeeded, although Pearle in 1990, 1992, and 1999, Ghirardi, Grassi, and Pearle in 1990, and Nicrosini and Rimini in 2003 have made valiant attempts. The problem is that the white noise field $w(\boldsymbol{x}, t)$ contains all wavelengths and frequencies, exciting the vacuum in lowest order in $\lambda$ to produce particles at the unacceptable rate of infinite energy/per second per cubic centimeter. Collapse models which utilize a colored noise field $w$ have a similar problem in higher orders. In 2005, Pearle suggested a quasirelativistic model which reduces to CSL in the low-speed limit.

CSL is a phenomenological model which describes dynamical collapse so as to achieve S. Besides needing decisive experimental verification, it needs identification of the $w(\boldsymbol{x}, t)$ field with a physical entity.

Other collapse models which have been investigated are briefly described below.

### Spontaneous Localization Model

The SL model of Ghirardi *et al.* (1986), although superseded by CSL, is historically important and conceptually valuable. Let $\hat{H} = 0$ for simplicity, and consider a single particle whose wave function at time $t$ is $\psi(\boldsymbol{x}, t)$. Over the next interval d$t$, with probability $1 - \lambda \mathrm{d}t$, it does not change. With probability $\lambda \mathrm{d}t$ it does change, by being "spontaneously localized" or "hit." A hit means that the new (unnormalized) wave function suddenly becomes

$$\psi(\boldsymbol{x}, t + \mathrm{d}t) = \psi(\boldsymbol{x}, t)(\pi a^2)^{-3/4}\, \mathrm{e}^{-(2a^2)^{-1}(\boldsymbol{x}-\boldsymbol{z})^2}$$

with probability

$$\lambda \mathrm{d}t\, \mathrm{d}\boldsymbol{z} \int \mathrm{d}\boldsymbol{x} |\psi(\boldsymbol{x}, t + \mathrm{d}t)|^2$$

Thus $\boldsymbol{z}$, the "center" of the hit, is most likely to be located where the wave function is large. For a single particle in the superposition described in the subsection "Interference," a single hit is overwhelmingly likely to reduce the wave function to one or the other location, with total probability $|\alpha_i|^2$, at the rate $\lambda$.

For an $N$-particle clump, it is considered that each particle has the same independent probability, $\lambda \mathrm{d}t$, of being hit. But, for the example in the subsection "Interference," a single hit on any particle in one location of the clump has the effect of multiplying the wave function part describing the clump in the other location by the tail of the Gaussian, thereby collapsing the wave function at the rate $\lambda N$.

By use of the Gaussian hit rather than a delta-function hit, SL solves the problem of giving too much energy to particles as mentioned in the subsection "CSL." By the hypothesis of independent particle hits, SL also solves the problem of achieving a slow collapse rate for a superposition of small objects and a fast collapse rate for a superposition of large objects. However, the hits on individual particles destroys the (anti-) symmetry of wave functions. The CSL collapse toward mass density eigenstates removes that problem. Also, while SL modifies the Schrödinger evolution of a wave function, it involves discontinuous dynamics and so is not described by a modified Schrödinger equation as is CSL.

## Other Models

For a single (low-energy) particle, the polar decomposition $\Psi = R\mathrm{e}^{(\mathrm{i}/\hbar)S}$ of the Schrödinger equation implies two real equations,

$$\frac{\partial R^2}{\partial t} + \boldsymbol{\nabla} \cdot \left( R^2 \frac{\boldsymbol{\nabla} S}{m} \right) = 0 \qquad [26]$$

(the continuity equation for $R^2 = |\Psi|^2$) and

$$\frac{\partial S}{\partial t} + \frac{(\boldsymbol{\nabla} S)^2}{2m} + V + Q = 0 \qquad [27]$$

where $Q \equiv -(\hbar^2/2m)\nabla^2 R/R$ is the "quantum potential." (These equations have an obvious generalisation to higher-dimensional configuration space.) In 1926, Madelung proposed that one should start from [26] and [27] – regarded as hydrodynamical equations for a classical charged fluid with mass density $mR^2$ and fluid velocity $\boldsymbol{\nabla} S/m$ – and construct $\Psi = R\mathrm{e}^{(\mathrm{i}/\hbar)S}$ from the solutions.

This "hydrodynamical" interpretation suffers from many difficulties, especially for many-body systems. In any case, a criticism by Wallstrom (1994) seems decisive: [26] and [27] (and their higher-dimensional analogs) are not, in fact, equivalent to the Schrödinger equation. For, as usually understood, the quantum wave function $\Psi$ is a single-valued and continuous complex field, which typically possesses nodes ($\Psi = 0$), in the neighborhood of which the phase $S$ is multivalued, with values differing by integral multiples of $2\pi\hbar$. If one allows $S$ in [26], [27] to be multivalued, there is no reason why the allowed values should differ by integral multiples of $2\pi\hbar$, and in general $\Psi$ will not be single-valued. On the other hand, if one restricts $S$ in [26], [27] to be single-valued, one will exclude wave functions – such as those of nonzero angular momentum – with a multivalued phase. (This problem does not exist in pilot-wave theory as we have presented it here, where $\Psi$ is regarded as a basic entity.)

Stochastic mechanics, introduced by Fényes in 1952 and Nelson (1966), has particle trajectories $\boldsymbol{x}(t)$ obeying a "forward" stochastic differential equation $\mathrm{d}\boldsymbol{x}(t) = \boldsymbol{b}(\boldsymbol{x}(t), t)\mathrm{d}t + \mathrm{d}\boldsymbol{w}(t)$, where $\boldsymbol{b}$ is a drift (equal to the mean forward velocity) and $\boldsymbol{w}$ a Wiener process, and also a similar "backward" equation. Defining the "current velocity" $\boldsymbol{v} = (1/2)(\boldsymbol{b} + \boldsymbol{b}_*)$, where $\boldsymbol{b}_*$ is the mean backward velocity, and using an appropriate time-symmetric definition of mean acceleration, one may impose a stochastic version of Newton's second law. If one assumes, in addition, that $\boldsymbol{v}$ is a gradient ($\boldsymbol{v} = \nabla S/m$ for some $S$), then one obtains [26], [27] with $R \equiv \sqrt{\rho}$, where $\rho$ is the particle density. Defining $\Psi \equiv \sqrt{\rho}\mathrm{e}^{(\mathrm{i}/\hbar)S}$, it appears that one recovers the Schrödinger equation for the derived quantity $\Psi$. However, again, there is no reason why $S$ should have the specific multivalued structure required for the phase of a single-valued complex field. It then seems that, despite appearances, quantum theory cannot in fact be recovered from stochastic mechanics (Wallstrom 1994). The same problem occurs in models that use stochastic mechanics as an intermediate step (e.g., Markopoulou and Smolin in 2004): the Schrödinger equation is obtained only for exceptional, nodeless wave functions.

Bohm and Bub (1966) first proposed dynamical wave-function collapse through deterministic evolution. Their collapse outcome is determined by the value of a Wiener–Siegel hidden variable (a variable distributed uniformly over the unit hypersphere in a Hilbert space identical to that of the state vector). In 1976, Pearle proposed dynamical wave-function collapse equations where the collapse outcome is determined by a random variable, and suggested (Pearle 1979) that the modified Schrödinger equation be formulated as an Itô stochastic differential equation, a suggestion which has been widely followed. (The equation for the state vector given here, which is physically more transparent, has its time derivative equivalent to a Stratonovich stochastic differential equation, which is readily converted to the Itô form.) The importance of requiring that the density matrix describing collapse be of the Lindblad–Kossakowski form was emphasized by Gisin in 1984 and Diosi in 1988. The stochastic differential Schrödinger equation that achieves this was found independently by Diosi in 1988 and by Belavkin, Gisin, and Pearle in separate papers in 1989 (see Ghirardi *et al.* 1990).

A gravitationally motivated stochastic collapse dynamics was proposed by Diosi in 1989 (and somewhat corrected by Ghirardi *et al.* in 1990). Penrose emphasized in 1996 that a quantum state, such as that describing a mass in a superposition of two places, puts the associated spacetime geometry also in a superposition, and has argued that this should lead to wave-function collapse. He suggests that the collapse time should be $\sim\hbar/\Delta E$, where $\Delta E$ is the gravitational potential energy change obtained by actually displacing two such masses: for example, the collapse time $\approx\hbar/(Gm^2/R)$, where the mass is $m$, its size is $R$, and the displacement is $\approx R$ or larger. No specific dynamics is offered, just the vision that this will be a property of a correct future quantum theory of gravity.

Collapse to energy eigenstates was first proposed by Bedford and Wang in 1975 and 1977 and, in the context of stochastic collapse (e.g., [11] with $\hat{A} = \hat{H}$), by Milburn in 1991 and Hughston in 1996, but it has been argued by Finkelstein in 1993 and Pearle in 2004 that such energy-driven collapse cannot give a satisfactory picture of the macroscopic world. Percival in 1995 and in a 1998 book, and Fivel in 1997 have discussed energy-driven collapse for microscopic situations.

Adler (2004) has presented a classical theory (a hidden-variables theory) from which it is argued that quantum theory "emerges" at the ensemble level. The classical variables are $N \times N$ matrix field amplitudes at points of space. They obey appropriate classical Hamiltonian dynamical equations which he calls "trace dynamics," since the expressions for Hamiltonian, Lagrangian, Poisson bracket, etc., have the form of the trace of products of matrices and their sums with constant coefficients. Using classical statistical mechanics, canonical ensemble averages of (suitably projected) products of fields are analyzed and it is argued that they obey all the properties associated with Wightman functions, from which quantum field theory, and its nonrelativistic-limit quantum mechanics, may be derived. As well as obtaining the algebra of quantum theory in this way,

it is argued that statistical fluctuations around the canonical ensemble can give rise to the behavior of wave-function collapse, of the kind discussed here, both energy-driven and CSL-type mass-density-driven collapse so that, with the latter, comes the Born probability interpretation of the algebra. The Hamiltonian needed for this theory to work is not provided but, as the argument progresses, its necessary features are delimited.

*See also*: Quantum Mechanics: Foundations.

## Further Reading


Adler SL (2004) *Quantum Theory as an Emergent Phenomenon.* Cambridge: Cambridge University Press.

Bassi A and Ghirardi GC (2003) Dynamical reduction models. *Physics Reports* 379: 257–426 (ArXiv: quant-ph/0302164).

Bell JS (1987) *Speakable and Unspeakable in Quantum Mechanics.* Cambridge: Cambridge University Press.

Bohm D (1952) A suggested interpretation of the quantum theory in terms of 'hidden' variables. I and II. *Physical Review* 85: 166–179 and 180–193.

Bohm D and Bub J (1966) A proposed solution of the measurement problem in quantum mechanics by a hidden variable theory. *Reviews of Modern Physics* 38: 453–469.

Dürr D, Goldstein S, and Zanghì N (2003) Quantum equilibrium and the role of operators as observables in quantum theory, ArXiv: quant-ph/0308038.

Ghirardi G, Rimini A, and Weber T (1986) Unified dynamics for microscopic and macroscopic systems. *Physical Review D* 34: 470–491.

Ghirardi G, Pearle P, and Rimini A (1990) Markov processes in Hilbert space and continuous spontaneous localization of systems of identical particles. *Physical Review A* 42: 78–89.

Holland PR (1993) *The Quantum Theory of Motion: An Account of the de Broglie–Bohm Causal Interpretation of Quantum Mechanics.* Cambridge: Cambridge University Press.

Nelson E (1966) Derivation of the Schrödinger equation from Newtonian mechanics. *Physical Review* 150: 1079–1085.

Pearle P (1979) Toward explaining why events occur. *International Journal of Theoretical Physics* 18: 489–518.

Pearle P (1989) Combining stochastic dynamical state-vector reduction with spontaneous localization. *Physical Review A* 39: 2277–2289.

Pearle P (1999) Collapse models. In: Petruccione F and Breuer HP (eds.) *Open Systems and Measurement in Relativistic Quantum Theory*, pp. 195–234. Heidelberg: Springer. (ArXiv: quant-ph/9901077).

Valentini A (1991) Signal-locality, uncertainty, and the subquantum *H*-theorem. I and II. *Physics Letters A* 156: 5–11 and 158: 1–8.

Valentini A (2002a) Signal-locality in hidden-variables theories. *Physics Letters A* 297: 273–278.

Valentini A (2002b) Subquantum information and computation. *Pramana – Journal of Physics* 59: 269–277 (ArXiv: quant-ph/0203049).

Valentini A (2004a) Black holes, information loss, and hidden variables, ArXiv: hep-th/0407032.

Valentini A (2004b) Universal signature of non-quantum systems. *Physics Letters A* 332: 187–193 (ArXiv: quant-ph/0309107).

Valentini A and Westman H (2005) Dynamical origin of quantum probabilities. *Proceedings of the Royal Society of London Series A* 461: 253–272.

Wallstrom TC (1994) Inequivalence between the Schrödinger equation and the Madelung hydrodynamic equations. *Physical Review A* 49: 1613–1617.


# Quantum Mechanics: Weak Measurements

**L Diósi**, Research Institute for Particle and Nuclear Physics, Budapest, Hungary



## Introduction

In quantum theory, the mean value of a certain observable $\hat{A}$ in a (pure) quantum state $|i\rangle$ is defined by the quadratic form:

$$\langle \hat{A} \rangle_i =: \langle i|\hat{A}|i\rangle \qquad [1]$$

Here $\hat{A}$ is Hermitian operator on the Hilbert space $\mathcal{H}$ of states. We use Dirac formalism. The above mean is interpreted statistically. No other forms had been known to possess a statistical interpretation in standard quantum theory. One can, nonetheless, try to extend the notion of mean for normalized bilinear expressions (Aharonov *et al.* 1988):

$$A_{\mathrm{w}} =: \frac{\langle f|\hat{A}|i\rangle}{\langle f|i\rangle} \qquad [2]$$

However unusual is this structure, standard quantum theory provides a plausible statistical interpretation for it, too. The two pure states $|i\rangle, |f\rangle$ play the roles of the prepared initial and the postselected final states, respectively. The statistical interpretation relies upon the concept of weak measurement. In a single weak measurement, the notorious decoherence is chosen asymptotically small. In physical terms, the coupling between the measured state and the meter is assumed asymptotically weak. The novel mean value [2] is called the (complex) weak value.

The concept of quantum weak measurement (Aharonov *et al.* 1988) provides particular

conclusions on postselected ensembles. Weak measurements have been instrumental in the interpretation of time-continuous quantum measurements on single states as well. Yet, weak measurement itself can properly be illuminated in the context of classical statistics. Classical weak measurement as well as postselection and time-continuous measurement are straightforward concepts leading to conclusions that are natural in classical statistics. In quantum context, the case is radically different and certain paradoxical conclusions follow from weak measurements. Therefore, we first introduce the classical notion of weak measurement on postselected ensembles and, alternatively, in time-continuous measurement on a single state. Certain idioms from statistical physics will be borrowed and certain not genuinely quantum notions from quantum theory will be anticipated. The quantum counterpart of weak measurement, postselection, and continuous measurement will be presented afterwards. The apparent redundancy of the parallel presentations is of reason: the reader can separate what is common in classical and quantum weak measurements from what is genuinely quantum.

## Classical Weak Measurement

Given a normalized probability density $\rho(X)$ over the phase space $\{X\}$, which we call the state, the mean value of a real function $A(X)$ is defined as

$$\langle A\rangle_\rho =: \int \mathrm{d}X\, A\rho \qquad [3]$$

Let the outcome of an (unbiased) measurement of $A$ be denoted by $a$. Its stochastic expectation value $E[a]$ coincides with the mean [3]:

$$E[a] = \langle A\rangle_\rho \qquad [4]$$

Performing a large number $N$ of independent measurements of $A$ on the elements of the ensemble of identically prepared states, the arithmetic mean $\bar{a}$ of the outcomes yields a reliable estimate of $E[a]$ and, this way, of the theoretical mean $\langle A\rangle_\rho$.

Suppose, for concreteness, the measurement outcome $a$ is subject to a Gaussian stochastic error of standard dispersion $\sigma > 0$. The probability distribution of $a$ and the update of the state corresponding to the Bayesian inference are described as

$$p(a) = \langle G_\sigma(a - A)\rangle_\rho \qquad [5]$$

$$\rho \to \frac{1}{p(a)} G_\sigma(a - A)\rho \qquad [6]$$

respectively. Here $G_\sigma$ is the central Gaussian distribution of variance $\sigma$. Note that, as expected, eqn [5] implies eqn [4]. Nonzero $\sigma$ means that the measurement is nonideal, yet the expectation value $E[a]$ remains calculable reliably if the statistics $N$ is suitably large.

Suppose the spread of $A$ in state $\rho$ is finite:

$$\Delta_\rho^2 A =: \langle A^2\rangle_\rho - \langle A\rangle_\rho^2 < \infty \qquad [7]$$

Weak measurement will be defined in the asymptotic limit (eqns [8] and [9]) where both the stochastic error of the measurement and the measurement statistics go to infinity. It is crucial that their rate is kept constant:

$$\sigma, N \to \infty \qquad [8]$$

$$\Delta^2 =: \frac{\sigma^2}{N} = \text{const.} \qquad [9]$$

Obviously for asymptotically large $\sigma$, the precision of individual measurements becomes extremely weak. This incapacity is fully compensated by the asymptotically large statistics $N$. In the weak measurement limit (eqns [8] and [9]), the probability distribution $p_\mathrm{w}$ of the arithmetic mean $\bar{a}$ of the $N$ independent outcomes converges to a Gaussian distribution:

$$p_\mathrm{w}(\bar{a}) \to G_\Delta\left(\bar{a} - \langle A\rangle_\rho\right) \qquad [10]$$

The Gaussian is centered at the mean $\langle A\rangle_\rho$, and the variance of the Gaussian is given by the constant rate [9]. Consequently, the mean [3] is reliably calculable on a statistics $N$ growing like $\sim\sigma^2$.

With an eye on quantum theory, we consider two situations – postselection and time-continuous measurement – of weak measurement in classical statistics.

### Postselection

For the preselected state $\rho$, we introduce postselection via the real function $\Pi(X)$, where $0 \le \Pi \le 1$. The postselected mean value of a certain real function $A(X)$ is defined by

$$_\Pi\langle A\rangle_\rho =: \frac{\langle \Pi A\rangle_\rho}{\langle \Pi\rangle_\rho} \qquad [11]$$

where $\langle \Pi\rangle_\rho$ is the rate of postselection. Postselection means that after having obtained the outcome $a$ regarding the measurement of $A$, we measure the function $\Pi$, too, in ideal measurement with random outcome $\pi$ upon which we base the following random decision. With probability $\pi$, we include the current $a$ into the statistics and we discard it

with probability $1-\pi$. Then the coincidence of $E[a]$ and ${}_\Pi\langle A\rangle_\rho$, as in eqn [4], remains valid:

$$E[a] = {}_\Pi\langle A\rangle_\rho \qquad [12]$$

Therefore, a large ensemble of postselected states allows one to estimate the postselected mean ${}_\Pi\langle A\rangle_\rho$.

Classical postselection allows introducing the effective postselected state:

$$\rho_\Pi =: \frac{\Pi\rho}{\langle\Pi\rangle_\rho} \qquad [13]$$

Then the postselected mean [11] of $A$ in state $\rho$ can, by eqn [14], be expressed as the common mean of $A$ in the effective postselected state $\rho_\Pi$:

$$_\Pi\langle A\rangle_\rho = \langle A\rangle_{\rho_\Pi} \qquad [14]$$

As we shall see later, quantum postselection is more subtle and cannot be reduced to common statistics, that is, to that without postselection. The quantum counterpart of postselected mean does not exist unless we combine postselection and weak measurement.

### Time-Continuous Measurement

For time-continuous measurement, one abandons the ensemble of identical states. One supposes that a single time-dependent state $\rho_t$ is undergoing an infinite sequence of measurements (eqns [5] and [6]) of $A$ employed at times $t=\delta t, t=2\delta t, t=3\delta t, \ldots$ . The rate $\nu =: 1/\delta t$ goes to infinity together with the mean squared error $\sigma^2$. Their rate is kept constant:

$$\sigma, \nu \to \infty \qquad [15]$$

$$g^2 =: \frac{\sigma^2}{\nu} = \text{const.} \qquad [16]$$

In the weak measurement limit (eqns [15] and [16]), the infinite frequent weak measurements of $A$ constitute the model of time-continuous measurement. Even the weak measurements will significantly influence the original state $\rho_0$, due to the accumulated effect of the infinitely many Bayesian updates [6]. The resulting theory of time-continuous measurement is described by coupled Gaussian processes [17] and [18] for the primitive function $\alpha_t$ of the time-dependent measurement outcome and, respectively, for the time-dependent Bayesian conditional state $\rho_t$:

$$d\alpha_t = \langle A\rangle_{\rho_t} dt + g\, dW_t \qquad [17]$$

$$d\rho_t = g^{-1}\left(A - \langle A\rangle_{\rho_t}\right)\rho_t\, dW_t \qquad [18]$$

Here $dW_t$ is the Itô differential of the Wiener process.

Equations [17] and [18] are the special case of the Kushner–Stratonovich equations of time-continuous Bayesian inference conditioned on the continuous measurement of $A$ yielding the time-dependent outcome value $a_t$. Formal time derivatives of both sides of eqn [17] yield the heuristic equation

$$a_t = \langle A\rangle_{\rho_t} + g\xi_t \qquad [19]$$

Accordingly, the current measurement outcome is always equal to the current mean plus a term proportional to standard white noise $\xi_t$. This plausible feature of the model survives in the quantum context as well. As for the other equation [18], it describes the gradual concentration of the distribution $\rho_t$ in such a way that the variance $\Delta_{\rho_t}A$ tends to zero while $\langle A\rangle_{\rho_t}$ tends to a random asymptotic value. The details of the convergence depend on the character of the continuously measured function $A(X)$. Consider a stepwise $A(X)$:

$$A(X) = \sum_\lambda a^\lambda P^\lambda(X) \qquad [20]$$

The real values $a^\lambda$ are step heights all differing from each other. The indicator functions $P^\lambda$ take values 0 or 1 and form a complete set of pairwise disjoint functions on the phase space:

$$\sum_\lambda P^\lambda \equiv 1 \qquad [21]$$

$$P^\lambda P^\mu = \delta_{\lambda\mu} P^\lambda \qquad [22]$$

In a single ideal measurement of $A$, the outcome $a$ is one of the $a^\lambda$'s singled out at random. The probability distribution of the measurement outcome and the corresponding Bayesian update of the state are given by

$$p^\lambda = \langle P^\lambda\rangle_{\rho_0} \qquad [23]$$

$$\rho_0 \to \frac{1}{p^\lambda} P^\lambda \rho_0 =: \rho^\lambda \qquad [24]$$

respectively. Equations [17] and [18] of time-continuous measurement are a connatural time-continuous resolution of the "sudden" ideal measurement (eqns [23] and [24]) in a sense that they reproduce it in the limit $t\to\infty$. The states $\rho^\lambda$ are trivial stationary states of the eqn [18]. It can be shown that they are indeed approached with probability $p^\lambda$ for $t\to\infty$.

## Quantum Weak Measurement

In quantum theory, states in a given complex Hilbert space $\mathcal{H}$ are represented by non-negative density operators $\hat\rho$, normalized by $\mathrm{tr}\,\hat\rho = 1$. Like the

classical states $\rho$, the quantum state $\hat{\rho}$ is interpreted statistically, referring to an ensemble of states with the same $\hat{\rho}$. Given a Hermitian operator $\hat{A}$, called observable, its theoretical mean value in state $\hat{\rho}$ is defined by

$$\langle \hat{A} \rangle_{\hat{\rho}} = \mathrm{tr}(\hat{A}\hat{\rho}) \qquad [25]$$

Let the outcome of an (unbiased) quantum measurement of $\hat{A}$ be denoted by $a$. Its stochastic expectation value $E[a]$ coincides with the mean [25]:

$$E[a] = \langle \hat{A} \rangle_{\hat{\rho}} \qquad [26]$$

Performing a large number $N$ of independent measurements of $\hat{A}$ on the elements of the ensemble of identically prepared states, the arithmetic mean $\bar{a}$ of the outcomes yields a reliable estimate of $E[a]$ and, this way, of the theoretical mean $\langle \hat{A} \rangle_{\hat{\rho}}$. If the measurement outcome $a$ contains a Gaussian stochastic error of standard dispersion $\sigma$, then the probability distribution of $a$ and the update, called collapse in quantum theory, of the state are described by eqns [27] and [28], respectively. (We adopt the notational convenience of physics literature to omit the unit operator $\hat{I}$ from trivial expressions like $a\hat{I}$.)

$$p(a) = \left\langle G_\sigma(a - \hat{A}) \right\rangle_{\hat{\rho}} \qquad [27]$$

$$\hat{\rho} \to \frac{1}{p(a)} G_\sigma^{1/2}(a - \hat{A})\hat{\rho}G_\sigma^{1/2}(a - \hat{A}) \qquad [28]$$

Nonzero $\sigma$ means that the measurement is nonideal, but the expectation value $E[a]$ remains calculable reliably if $N$ is suitably large.

Weak quantum measurement, like its classical counterpart, requires finite spread of the observable $\hat{A}$ on state $\hat{\rho}$:

$$\Delta_{\hat{\rho}}^2 \hat{A} =: \langle \hat{A}^2 \rangle_{\hat{\rho}} - \langle \hat{A} \rangle_{\hat{\rho}}^2 < \infty \qquad [29]$$

Weak quantum measurement, too, will be defined in the asymptotic limit [8] introduced for classical weak measurement. Single quantum measurements can no more distinguish between the eigenvalues of $\hat{A}$. Yet, the expectation value $E[a]$ of the outcome $a$ remains calculable on a statistics $N$ growing like $\sim\sigma^2$.

Both in quantum theory and classical statistics, the emergence of nonideal measurements from ideal ones is guaranteed by general theorems. For completeness of this article, we prove the emergence of the nonideal quantum measurement (eqns [27] and [28]) from the standard von Neumann theory of ideal quantum measurements (von Neumann 1955). The source of the statistical error of dispersion $\sigma$ is associated with the state $\hat{\rho}_M$ in the complex Hilbert space $\mathcal{L}^2$ of a hypothetic meter. Suppose $R \in (-\infty, \infty)$ is the position of the "pointer." Let its initial state $\hat{\rho}_M$ be a pure central Gaussian state of width $\sigma$; then the density operator $\hat{\rho}_M$ in Dirac position basis takes the form

$$\hat{\rho}_M = \int \mathrm{d}R \int \mathrm{d}R' G_\sigma^{1/2}(R) G_\sigma^{1/2}(R') |R\rangle\langle R'| \qquad [30]$$

We are looking for a certain dynamical interaction to transmit the "value" of the observable $\hat{A}$ onto the pointer position $\hat{R}$. To model the interaction, we define the unitary transformation [31] to act on the tensor space $\mathcal{H} \otimes \mathcal{L}^2$:

$$\hat{U} = \exp(\mathrm{i}\hat{A} \otimes \hat{K}) \qquad [31]$$

Here $\hat{K}$ is the canonical momentum operator conjugated to $\hat{R}$:

$$\exp(\mathrm{i}a\hat{K})|R\rangle = |R + a\rangle \qquad [32]$$

The unitary operator $\hat{U}$ transforms the initial uncorrelated quantum state into the desired correlated composite state:

$$\hat{\Sigma} =: \hat{U}\hat{\rho} \otimes \hat{\rho}_M \hat{U}^\dagger \qquad [33]$$

Equations [30]–[33] yield the expression [34] for the state $\hat{\Sigma}$:

$$\hat{\Sigma} = \int \mathrm{d}R \int \mathrm{d}R' G_\sigma^{1/2}(R - \hat{A})\hat{\rho}G_\sigma^{1/2} \times (R' - \hat{A}) \otimes |R\rangle\langle R'| \qquad [34]$$

Let us write the pointer's coordinate operator $\hat{R}$ into the standard form [35] in Dirac position basis:

$$\hat{R} = \int \mathrm{d}a |a\rangle\langle a| \qquad [35]$$

The notation anticipates that, when pointer $\hat{R}$ is measured ideally, the outcome $a$ plays the role of the nonideally measured value of the observable $\hat{A}$. Indeed, let us consider the ideal von Neumann measurement of the pointer position on the correlated composite state $\hat{\Sigma}$. The probability of the outcome $a$ and the collapse of the composite state are given by the following standard equations:

$$p(a) = \mathrm{tr}\left[(\hat{I} \otimes |a\rangle\langle a|)\hat{\Sigma}\right] \qquad [36]$$

$$\hat{\Sigma} \to \frac{1}{p(a)} \left[(\hat{I} \otimes |a\rangle\langle a|)\hat{\Sigma}(\hat{I} \otimes |a\rangle\langle a|)\right] \qquad [37]$$

respectively. We insert eqn [34] into eqns [36] and [37]. Furthermore, we take the trace over $\mathcal{L}^2$ of both sides of eqn [37]. In such a way, as expected, eqns [36] and [37] of ideal measurement of $\hat{R}$ yield the

earlier postulated eqns [27] and [28] of nonideal measurement of $\hat{A}$.

### Quantum Postselection

A quantum postselection is defined by a Hermitian operator satisfying $\hat{0} \leq \hat{\Pi} \leq \hat{I}$. The corresponding postselected mean value of a certain observable $\hat{A}$ is defined by

$$ {}_{\hat{\Pi}}\langle \hat{A} \rangle_{\hat{\rho}} =: \mathrm{Re}\frac{\langle \hat{\Pi}\hat{A} \rangle_{\hat{\rho}}}{\langle \hat{\Pi} \rangle_{\hat{\rho}}} \qquad [38] $$

The denominator $\langle \hat{\Pi} \rangle_{\hat{\rho}}$ is the rate of quantum postselection. Quantum postselection means that after the measurement of $\hat{A}$, we measure the observable $\hat{\Pi}$ in ideal quantum measurement and we make a statistical decision on the basis of the outcome $\pi$. With probability $\pi$, we include the case in question into the statistics while we discard it with probability $1-\pi$. By analogy with the classical case [12], one may ask whether the stochastic expectation value $E[a]$ of the postselected measurement outcome does coincide with

$$ E[a] \stackrel{?}{=} {}_{\hat{\Pi}}\langle \hat{A} \rangle_{\hat{\rho}} \qquad [39] $$

Contrary to the classical case, the quantum equation [39] does not hold. The quantum counterparts of classical equations [12]–[14] do not exist at all. Nonetheless, the quantum postselected mean ${}_{\hat{\Pi}}\langle \hat{A} \rangle_{\hat{\rho}}$ possesses statistical interpretation although restricted to the context of weak quantum measurements. In the weak measurement limit (eqns [8] and [9]), a postselected analog of classical equation [10] holds for the arithmetic mean $\bar{a}$ of postselected weak quantum measurements:

$$ p_{\mathrm{w}}(\bar{a}) \rightarrow G_{\Delta}\left(\bar{a} - {}_{\hat{\Pi}}\langle \hat{A} \rangle_{\hat{\rho}}\right) \qquad [40] $$

The Gaussian is centered at the postselected mean ${}_{\hat{\Pi}}\langle \hat{A} \rangle_{\hat{\rho}}$, and the variance of the Gaussian is given by the constant rate [9]. Consequently, the mean [38] becomes calculable on a statistics $N$ growing like $\sim\sigma^2$.

Since the statistical interpretation of the postselected quantum mean [38] is only possible for weak measurements, therefore ${}_{\hat{\Pi}}\langle \hat{A} \rangle_{\hat{\rho}}$ is called the (real) weak value of $\hat{A}$. Consider the special case when both the state $\hat{\rho} = |i\rangle\langle i|$ and the postselected operator $\hat{\Pi} = |f\rangle\langle f|$ are pure states. Then the weak value ${}_{\hat{\Pi}}\langle \hat{A} \rangle_{\hat{\rho}}$ takes, in usual notations, a particular form [41] yielding the real part of the complex weak value $A_{\mathrm{w}}$ [1]:

$$ {}_f\langle \hat{A} \rangle_i =: \mathrm{Re}\frac{\langle f|\hat{A}|i\rangle}{\langle f|i\rangle} \qquad [41] $$

The interpretation of postselection itself reduces to a simple procedure. One performs the von Neumann ideal measurement of the Hermitian projector $|f\rangle\langle f|$, then includes the case if the outcome is 1 and discards it if the outcome is 0. The rate of postselection is $|\langle f|i\rangle|^2$. We note that a certain statistical interpretation of $\mathrm{Im}\, A_{\mathrm{w}}$, too, exists although it relies upon the details of the "meter."

We outline a heuristic proof of the central equation [40]. One considers the nonideal measurement (eqns [27] and [28]) of $\hat{A}$ followed by the ideal measurement of $\hat{\Pi}$. Then the joint distribution of the corresponding outcomes is given by eqn [42]. The probability distribution of the postselected outcomes $a$ is defined by eqn [43], and takes the concrete form [44]. The constant $\mathcal{N}$ assures normalization:

$$ p(\pi, a) = \mathrm{tr}\left(\delta(\pi - \hat{\Pi}) G_{\sigma}^{1/2}(a - \hat{A}) \hat{\rho} G_{\sigma}^{1/2}(a - \hat{A})\right) \qquad [42] $$

$$ p(a) =: \frac{1}{\mathcal{N}} \int \pi p(\pi, a)\, \mathrm{d}\pi \qquad [43] $$

$$ p(a) =: \frac{1}{\mathcal{N}} \left\langle G_{\sigma}^{1/2}(a - \hat{A}) \hat{\Pi} G_{\sigma}^{1/2}(a - \hat{A}) \right\rangle_{\hat{\rho}} \qquad [44] $$

Suppose, for simplicity, that $\hat{A}$ is bounded. When $\sigma \rightarrow \infty$, eqn [44] yields the first two moments of the outcome $a$:

$$ E[a] \rightarrow {}_{\hat{\Pi}}\langle \hat{A} \rangle_{\hat{\rho}} \qquad [45] $$

$$ E[a^2] \sim \sigma^2 \qquad [46] $$

Hence, by virtue of the central limit theorem, the probability distribution [40] follows for the average $\bar{a}$ of postselected outcomes in the weak measurement limit (eqns [8] and [9]).

### Quantum Weak-Value Anomaly

Unlike in classical postselection, effective postselected quantum states cannot be introduced. We can ask whether eqn [47] defines a correct postselected quantum state:

$$ \hat{\rho}_{\hat{\Pi}}^{?} =: \mathrm{Herm}\frac{\hat{\Pi}\hat{\rho}}{\langle \hat{\Pi} \rangle_{\hat{\rho}}} \qquad [47] $$

This pseudo-state satisfies the quantum counterpart of the classical equation [14]:

$$ {}_{\hat{\Pi}}\langle \hat{A} \rangle_{\hat{\rho}} = \mathrm{tr}\left(\hat{A}\hat{\rho}_{\hat{\Pi}}^{?}\right) \qquad [48] $$

In general, however, the operator $\hat{\rho}_{\hat{\Pi}}^{?}$ is not a density operator since it may be indefinite. Therefore, eqn [47] does not define a quantum state. Equation [48] does not guarantee that the quantum weak value

${}_{\Pi}\langle\hat{A}\rangle_{\hat{\rho}}$ lies within the range of the eigenvalues of the observable $\hat{A}$.

Let us see a simple example for such anomalous weak values in the two-dimensional Hilbert space. Consider the pure initial state given by eqn [49] and the postselected pure state by eqn [50], where $\phi \in [0, \pi]$ is a certain angular parameter.

$$|i\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} e^{i\phi/2} \\ e^{-i\phi/2} \end{bmatrix} \quad [49]$$

$$|f\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} e^{-i\phi/2} \\ e^{i\phi/2} \end{bmatrix} \quad [50]$$

The probability of successful postselection is $\cos^2\phi$. If $\phi \neq \pi/2$, then the postselected pseudo-state follows from eqn [47]:

$$\hat{\rho}^{?}_{\Pi} = \frac{1}{2}\begin{bmatrix} 1 & \cos^{-1}\phi \\ \cos^{-1}\phi & 1 \end{bmatrix} \quad [51]$$

This matrix is indefinite unless $\phi = 0$, its two eigenvalues are $1 \pm \cos^{-1}\phi$. The smaller the post-selection rate $\cos^2\phi$, the larger is the violation of the positivity of the pseudo-density operator. Let the weakly measured observable take the form

$$\hat{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad [52]$$

Its eigenvalues are $\pm 1$. We express its weak value from eqns [41], [49], and [50] or, equivalently, from eqns [48] and [51]:

$${}_{f}\langle\hat{A}\rangle_{i} = \frac{1}{\cos\phi} \quad [53]$$

This weak value of $\hat{A}$ lies outside the range of the eigenvalues of $\hat{A}$. The anomaly can be arbitrarily large if the rate $\cos^2\phi$ of postselection decreases.

Striking consequences follow from this anomaly if we turn to the statistical interpretation. For concreteness, suppose $\phi = 2\pi/3$ so that ${}_{f}\langle\hat{A}\rangle_{i} = 2$. On average, 75% of the statistics $N$ will be lost in postselection. We learnt from eqn [40] that the arithmetic mean $\bar{a}$ of the postselected outcomes of independent weak measurements converges stochastically to the weak value upto the Gaussian fluctuation $\Delta$, as expressed symbolically by

$$\bar{a} = 2 \pm \Delta \quad [54]$$

Let us approximate the asymptotically large error $\sigma$ of our weak measurements by $\sigma = 10$ which is already well beyond the scale of the eigenvalues $\pm 1$ of the observable $\hat{A}$. The Gaussian error $\Delta$ derives from eqn [9] after replacing $N$ by the size of the postselected statistics which is approximately $N/4$:

$$\Delta^2 = 400/N \quad [55]$$

Accordingly, if $N = 3600$ independent quantum measurements of precision $\sigma = 10$ are performed regarding the observable $\hat{A}$, then the arithmetic mean $\bar{a}$ of the $\sim 900$ postselected outcomes $a$ will be $2 \pm 0.33$. This exceeds significantly the largest eigenvalue of the measured observable $\hat{A}$. Quantum postselection appears to bias the otherwise unbiased nonideal weak measurements.

### Quantum Time-Continuous Measurement

The mathematical construction of time-continuous quantum measurement is similar to the classical one. We consider the weak measurement limit (eqns [15] and [16]) of an infinite sequence of nonideal quantum measurements of the observable $\hat{A}$ at $t = \delta t, 2\delta t, \ldots$, on the time-dependent state $\hat{\rho}_t$. The resulting theory of time-continuous quantum measurement is incorporated in the coupled stochastic equations [56] and [57] for the primitive function $\alpha_t$ of the time-dependent outcome and the conditional time-dependent state $\hat{\rho}_t$, respectively (Diósi 1988):

$$d\alpha_t = \langle\hat{A}\rangle_{\hat{\rho}_t}\, dt + g dW_t \quad [56]$$

$$d\hat{\rho}_t = -\tfrac{1}{8}\, g^{-2}[\hat{A}, [\hat{A}, \hat{\rho}_t]]\, dt + g^{-1}\,\mathrm{Herm}\left(\hat{A} - \langle\hat{A}\rangle_{\hat{\rho}_t}\right)\hat{\rho}_t\, dW_t \quad [57]$$

Equation [56] and its classical counterpart [17] are perfectly similar. There is a remarkable difference between eqn [57] and its classical counterpart [18]. In the latter, the stochastic average of the state is constant: $E[d\rho_t] = 0$, expressing the fact that classical measurements do not alter the original ensemble if we "ignore" the outcomes of the measurements. On the contrary, quantum measurements introduce irreversible changes to the original ensemble, a phenomenon called decoherence in the physics literature. Equation [57] implies the closed linear first-order differential equation [58] for the stochastic average of the quantum state $\hat{\rho}_t$ under time-continuous measurement of the observable $\hat{A}$:

$$\frac{dE[\hat{\rho}_t]}{dt} = -\tfrac{1}{8}g^{-2}[\hat{A}, [\hat{A}, E[\hat{\rho}_t]]] \quad [58]$$

This is the basic irreversible equation to model the gradual loss of quantum coherence (decoherence) under time-continuous measurement. In fact, the very equation models decoherence under the influence of a large class of interactions, for example, with thermal reservoirs or complex environments. In

two-dimensional Hilbert space, for instance, we can consider the initial pure state $\langle i| =: [\cos\phi, \sin\phi]$ and the time-continuous measurement of the diagonal observable [59] on it. The solution of eqn [58] is given by eqn [60]:

$$\hat{A} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad [59]$$

$$E[\hat{\rho}_t] = \begin{bmatrix} \cos^2\phi & e^{-t/4g^2}\cos\phi\sin\phi \\ e^{-t/4g^2}\cos\phi\sin\phi & \sin^2\phi \end{bmatrix} \quad [60]$$

The off-diagonal elements of this density matrix go to zero, that is, the coherent superposition represented by the initial pure state becomes an incoherent mixture represented by the diagonal density matrix $\hat{\rho}_\infty$.

Apart from the phenomenon of decoherence, the stochastic equations show remarkable similarity with the classical equations of time-continuous measurement. The heuristic form of eqn [56] is eqn [61] of invariable interpretation with respect to the classical equation [19]:

$$a_t = \langle\hat{A}\rangle_{\hat{\rho}_t} + g\xi_t \quad [61]$$

Equation [57] describes what is called the time-continuous collapse of the quantum state under time-continuous quantum measurement of $\hat{A}$. For concreteness, we assume discrete spectrum for $\hat{A}$ and consider the spectral expansion

$$\hat{A} = \sum_\lambda a^\lambda \hat{P}^\lambda \quad [62]$$

The real values $a^\lambda$ are nondegenerate eigenvalues. The Hermitian projectors $\hat{P}^\lambda$ form a complete orthogonal set:

$$\sum_\lambda \hat{P}^\lambda \equiv \hat{I} \quad [63]$$

$$\hat{P}^\lambda \hat{P}^\mu = \delta_{\lambda\mu}\hat{P}^\lambda \quad [64]$$

In a single ideal measurement of $\hat{A}$, the outcome $a$ is one of the $a^\lambda$'s singled out at random. The probability distribution of the measurement outcome and the corresponding collapse of the state are given by

$$p^\lambda = \langle\hat{P}^\lambda\rangle_{\hat{\rho}_0} \quad [65]$$

$$\hat{\rho}_0 \to \frac{1}{p^\lambda}\hat{P}^\lambda\hat{\rho}_0\hat{P}^\lambda =: \hat{\rho}^\lambda \quad [66]$$

respectively. Equations [56] and [57] of continuous measurements are an obvious time-continuous resolution of the "sudden" ideal quantum measurement (eqns [65] and [66]) in a sense that they reproduce it in the limit $t \to \infty$. The states $\hat{\rho}^\lambda$ are stationary states of eqn [57]. It can be shown that they are indeed approached with probability $p^\lambda$ for $t \to \infty$ (Gisin 1984).

## Related Contexts

In addition to the two particular examples as in postselection and in time-continuous measurement, respectively, presented above, the weak measurement limit itself has further variants. A most natural example is the usual thermodynamic limit in standard statistical physics. Then weak measurements concern a certain additive microscopic observable (e.g., the spin) of each constituent and the weak value represents the corresponding additive macroscopic parameter (e.g., the magnetization) in the infinite volume limit. This example indicates that weak values have natural interpretation despite the apparent artificial conditions of their definition. It is important that the weak value, with or without postselection, plays the physical role similar to that of the common mean $\langle\hat{A}\rangle_{\hat{\rho}}$. If, between their pre- and postselection, the states $\hat{\rho}$ become weakly coupled with the state of another quantum system via the observable $\hat{A}$, their average influence will be as if $\hat{A}$ took the weak value ${}_{\hat{\Pi}}\langle\hat{A}\rangle_{\hat{\rho}}$. Weak measurements also open a specific loophole to circumvent quantum limitations related to the irreversible disturbances that quantum measurements cause to the measured state. Noncommuting observables become simultaneously measurable in the weak limit: simultaneous weak values of noncommuting observables will exist.

Literally, weak measurement had been coined in 1988 for quantum measurements with (pre- and) postselection, and became the tool of a certain time-symmetric statistical interpretation of quantum states. Foundational applications target the paradoxical problem of pre- and retrodiction in quantum theory. In a broad sense, however, the very principle of weak measurement encapsulates the trade between asymptotically weak precision and asymptotically large statistics. Its relevance in different fields has not yet been fully explored and a growing number of foundational, theoretical, and experimental applications are being considered in the literature – predominantly in the context of quantum physics. Since specialized monographs or textbooks on quantum weak measurement are not yet available, the reader is mostly referred to research articles, like the recent one by Aharonov and Botero (2005), covering many topics of postselected quantum weak values.

## Nomenclature

| | |
|---|---|
| $a$ | measurement outcome |
| $\bar{a}$ | arithmetic mean of measurement outcomes |
| $\hat{A}$ | Hermitian operator, quantum observable |
| $A(X)$ | real phase-space function |
| $E[\ldots]$ | stochastic expectation value |
| $\langle f\|\hat{A}\|i\rangle$ | matrix element |
| $\langle f\|i\rangle$ | inner product |
| $\mathcal{H}$ | Hilbert space |
| $\mathcal{L}^2$ | space of Lebesgue square-integrable complex functions |
| $p$ | probability distribution |
| tr | trace |
| $\hat{U}$ | unitary operator |
| $W_t$ | Wiener process |
| $\xi_t$ | white noise process |
| ${}_{\Pi}\langle\ldots\rangle_\rho$ | postselected mean value |
| $\hat{\rho}$ | density operator |
| $\rho(X)$ | phase-space distribution |
| $\otimes$ | direct product |
| $\dagger$ | operator adjoint |
| $\|\ldots\rangle$ | state vector |
| $\langle\ldots\|$ | adjoint state vector |
| $\langle\ldots\rangle_\rho$ | mean value |

## Further Reading

Aharonov Y, Albert DZ, and Vaidman L (1988) How the result of measurement of a component of the spin of a spin-1/2 particle can turn out to be 100. *Physical Review Letters* 60: 1351–1354.

Aharonov Y and Botero A (2005) Quantum averages of weak values. *Physical Review A* (in print).

Aharonov Y, Popescu S, Rohrlich D, and Vaidman L (1993) Measurements, errors, and negative kinetic energy. *Physical Review A* 48: 4084–4090.

Aharonov Y and Vaidman L (1990) Properties of a quantum system during the time interval between two measurements. *Physical Review A* 41: 11–20.

Belavkin VP (1989) A new equation for a continuous nondemolition measurement. *Physics Letters A* 140: 355–358.

Diósi L (1988) Continuous quantum measurement and Itô-formalism. *Physics Letters A* 129: 419–432.

Gisin N (1984) Quantum measurements and stochastic processes. *Physical Review Letters* 52: 1657–1660.

Giulini D, Joos E, Kiefer C, Kupsch J, Stamatescu IO *et al.* (1996) *Decoherence and the Appearance of a Classical World in Quantum Theory*. Berlin: Springer.

Gough J and Sobolev A (2004) Stochastic Schrödinger equations as limit of discrete filtering. *Open Systems and Information Dynamics* 11: 1–21.

Kraus K (1983) *States, Effects, and Operations: Fundamental Notions of Quantum Theory*. Berlin: Springer.

von Neumann J (1955) *Mathematical Foundations of Quantum Mechanics*. Princeton: Princeton University Press.

Nielsen MA and Chuang IL (2000) *Quantum Computation and Quantum Information*. Cambridge: Cambridge University Press.

Stratonovich R (1968) *Conditional Markov Processes and Their Application to the Theory of Optimal Control*. New York: Elsevier.

Wiseman HM (2002) Weak values, quantum trajectories, and the cavity-QED experiment on wave-particle correlation. *Physical Review A* 65: 32111–32116.

# Quantum *n*-Body Problem

**R G Littlejohn**, University of California at Berkeley, Berkeley, CA, USA



## Introduction

This article concerns the nonrelativistic quantum mechanics of isolated systems of $n$ particles interacting by means of a scalar potential, what we shall call the "quantum $n$-body problem." Such systems are described by the kinetic-plus-potential Hamiltonian,

$$H = T + V = \sum_{\alpha=1}^{n} \frac{|\boldsymbol{P}_\alpha|^2}{2m_\alpha} + V(\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n) \qquad [1]$$

where $\boldsymbol{R}_\alpha, \boldsymbol{P}_\alpha, \alpha = 1, \ldots, n$ are the positions and momenta of the $n$ particles in three-dimensional space, $m_\alpha$ are the masses, and $V$ is the potential energy. This Hamiltonian also occurs in the "classical $n$-body problem," in which $V$ is usually assumed to consist of the sum of the pairwise gravitational interactions of the particles. In this article, we shall only assume that $V$ (hence $H$) is invariant under translations, proper rotations, parity, and permutations of identical particles. The Hamiltonian $H$ is also invariant under time reversal. This Hamiltonian describes the dynamics of isolated atoms, molecules, and nuclei, with varying degrees of approximation, including the case of molecules in the Born–Oppenheimer approximation, in which $V$ is the Born–Oppenheimer potential. We shall ignore the spin of the particles, and treat the wave function $\Psi$ as a scalar. We assume that $\Psi$ is an eigenfunction of $H$, $H\Psi = E\Psi$. In practice, the value of $n$ typically ranges from 2 to several hundred. Often the cases $n = 3$ and $n = 4$ are of special interest. In this article, we shall assume that $n \geq 3$, since $n = 2$ is the trivial case of central-force motion. The quantum $n$-body problem is not to be confused with the "quantum

many-body problem," which usually refers to the quantum mechanics of large numbers of identical particles, such as the electrons in a solid.

Of particular interest is the "reduction" of the Hamiltonian [1], that is, the elimination of those degrees of freedom that can be eliminated due to the continuous symmetries of translations and rotations. A basic problem is to write down the reduced Hamiltonian and to make its analytical and geometrical properties clear. In the following we shall present this reduction in two stages, dealing first with the translations and second with the proper rotations. In each stage, we shall describe the reduction first in coordinate language and then in geometrical language. The discrete symmetries of parity, time reversal, and permutation of identical particles are handled by standard methods of group representation theory, and will not be discussed here.

There has been considerable interest in mathematical circles in recent years in the reduction of dynamical systems with symmetry, and the quantum $n$-body problem is one of the most important such systems from a physical standpoint. As such, the basic theory of the quantum $n$-body problem has received considerable attention in the physical literature going back to the birth of quantum mechanics, and continues to be of great practical importance. This article and the bibliography attempt to bridge these two centers of interest.

## Reduction by Translations: Coordinate Description

We begin with a coordinate description of the reduction of the system [1] by translations. The coordinates $(\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n)$ are coordinates on the configuration space of the system, called the "original configuration space" or OCS. The OCS is $\mathbb{R}^{3n}$. The original system has $3n$ degrees of freedom. The translation group acts on configuration space by $\boldsymbol{R}_\alpha \mapsto \boldsymbol{R}_\alpha + \xi$, for $\alpha = 1, \ldots, n$, where $\xi$ is a displacement vector. It acts on wave functions by $\Psi(\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n) \mapsto \Psi(\boldsymbol{R}_1 - \xi, \ldots, \boldsymbol{R}_n - \xi)$.

To reduce the system by translations, we perform a linear coordinate transformation on the OCS, taking us from the original vectors $(\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n)$ to a new set of $n$ vectors $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-1}, \boldsymbol{R}_{\mathrm{CM}})$, where $\boldsymbol{R}_{\mathrm{CM}}$ is the center-of-mass position,

$$\boldsymbol{R}_{\mathrm{CM}} = \frac{1}{M} \sum_{\alpha=1}^{n} m_\alpha \boldsymbol{R}_\alpha \qquad [2]$$

where $M = \sum_\alpha m_\alpha$ is the total mass of the system, and the other $n-1$ vectors of the new coordinate system, $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-1})$, are required to be translationally invariant, that is, independent linear functions of the relative particle positions $\boldsymbol{R}_\alpha - \boldsymbol{R}_\beta$. We denote the momenta conjugate to $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-1}, \boldsymbol{R}_{\mathrm{CM}})$ by $(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{n-1}, \boldsymbol{P}_{\mathrm{CM}})$, of which $\boldsymbol{P}_{\mathrm{CM}}$ turns out to be the total momentum of the system,

$$\boldsymbol{P}_{\mathrm{CM}} = \sum_{\alpha=1}^{n} \boldsymbol{P}_\alpha \qquad [3]$$

Under such a coordinate transformation, the potential energy becomes simply a function of the $n-1$ relative vectors, $V(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-1})$, whereas the kinetic energy becomes

$$T = \frac{|\boldsymbol{P}_{\mathrm{CM}}|^2}{2M} + \frac{1}{2} \sum_{\alpha,\beta=1}^{n-1} K^{\alpha\beta} \boldsymbol{p}_\alpha \cdot \boldsymbol{p}_\beta \qquad [4]$$

where $K^{\alpha\beta}$ is a symmetric tensor (the "inverse mass tensor").

The vectors $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-1})$ specify the positions of $n$ particles relative to their center of mass. As described so far, these vectors need only be independent, translationally invariant linear combinations of the particle postitions. However, it is convenient to choose them so that the inverse mass tensor becomes proportional to the identity, $K^{\alpha\beta} = (1/M)\delta_{\alpha\beta}$. An elegant way of doing this is the method of Jacobi vectors, which involves splitting the original set of particles into two nonempty subsets, which are then split into smaller subsets, etc., until only subsets of a single particle remain. The process can be represented by a tree growing downward, with the original $n$ particles as the root, and the ends of the branches at the bottom each containing one particle. Then the vectors $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-1})$ (the Jacobi vectors) are chosen to be proportional to the differences between the centers of mass of the two subsets at each splitting. With the right constants of proportionality, the kinetic energy becomes

$$T = \frac{1}{2M} |\boldsymbol{P}_{\mathrm{CM}}|^2 + \frac{1}{2M} \sum_{\alpha=0}^{n-1} |\boldsymbol{p}_\alpha|^2 \qquad [5]$$

Henceforth, we shall assume that the vectors $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-1})$ are Jacobi vectors with conjugate momenta $(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{n-1})$.

The choice of Jacobi vectors is not unique. In the first place, there is a discrete set of possible ways of splitting the original set of $n$ particles into subsets (of forming trees), each of which leads to the same form [5] of the kinetic energy. More generally, the kinetic energy [5] is invariant under transformations

$$\boldsymbol{r}'_\alpha = \sum_{\beta=1}^{n-1} Q_{\alpha\beta} \boldsymbol{r}_\beta \qquad [6]$$

where $Q_{\alpha\beta}$ is an orthogonal matrix, $Q \in O(n-1)$. Such transformations are called "kinematic rotations." The discrete choices of trees in forming the Jacobi vectors are equivalent to a discrete set of kinematic rotations $Q_{\alpha\beta}$ that map one standard choice of Jacobi vectors into the others.

Since the momentum $\boldsymbol{P}_{\mathrm{CM}}$ of the center of mass commutes with $H$, the eigenfunctions $\Psi$ of $H$ can be chosen to have the form

$$\Psi(\boldsymbol{R}_1, \ldots, \boldsymbol{R}_n) = \exp(\mathrm{i}\boldsymbol{R}_{\mathrm{CM}} \cdot \boldsymbol{P}_{\mathrm{CM}}/\hbar)\psi(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-1}) \qquad [7]$$

This causes $\psi$ to be an eigenfunction of the "translation-reduced Hamiltonian," $H_{\mathrm{tr}}\psi = E_{\mathrm{tr}}\psi$, where

$$H_{\mathrm{tr}} = \frac{1}{2M}\sum_{\alpha=0}^{n-1} |\boldsymbol{p}_\alpha|^2 + V(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-1}) \qquad [8]$$

The kinetic energy of the center of mass, $|\boldsymbol{P}_{\mathrm{CM}}|^2/2M$, has been discarded from both $H_{\mathrm{tr}}$ and $E_{\mathrm{tr}}$, which represent physically the energy of the system about its center of mass.

## Reduction by Translations: Geometrical Description

The kinetic energy $T$ in eqn [1] specifies a metric $\mathrm{d}s^2 = \sum_\alpha m_\alpha |\mathrm{d}\boldsymbol{R}_\alpha|^2$ on the OCS ($=\mathbb{R}^{3n}$). The translation group ($=\mathbb{R}^3$) acts freely on the OCS, with an action that is generated by $\boldsymbol{P}_{\mathrm{CM}}$. This action defines an orthogonal decomposition of the OCS, $\mathbb{R}^{3n} = \mathbb{R}^3 \oplus \mathbb{R}^{3n-3}$, where $\mathbb{R}^3$ is the orbit of the origin (the other orbits of the translation group action are parallel spaces), and $\mathbb{R}^{3n-3}$ is the orthogonal subspace (henceforth the "translation-reduced configuration space" or TRCS for short). The TRCS is physically the space of configurations relative to the center of mass. The vectors $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-1})$ are coordinates on the TRCS. The TRCS possesses a metric which is the projection of the metric on the OCS onto the TRCS by means of the translation group action. The metric can be projected because translations preserve the original metric (they are isometries). Jacobi vectors are Euclidean coordinates on the TRCS with respect to this metric.

The tree method of constructing Jacobi vectors can be understood in terms of certain group actions which take place as each subset of particles is split into two further subsets. The group action in question leaves the center of mass of the original subset invariant, while moving the two new subsets apart along a line. This motion in the configuration space is orthogonal to all the other group actions that are created in the process of splitting subsets of particles, including the original action of the translation group. Thus, each splitting of a subset of particles generates a three-dimensional subspace of the OCS, on which one of the $\boldsymbol{r}_\alpha$ are coordinates. The conjugate momentum $\boldsymbol{p}_\alpha$ is the generator of the group action moving the two new subsets apart. The final result is that the OCS is decomposed into $n$ orthogonal, three-dimensional subspaces, one of which contains the action of the original translation group, and the others of which represent the decomposition of the TRCS into $n-1$, three-dimensional orthogonal subspaces.

The TRCS can also be seen as a global section of a flat, trivial, principal fiber bundle created by the action of the translation group on the OCS. Alternatively, the TRCS can be seen as the quotient space, $\mathbb{R}^{3n}/\mathbb{R}^3$. The construction is fairly simple because the translation group is Abelian.

The wave function $\psi$ can be seen as a member of the Hilbert space of wave functions on the TRCS, upon which the reduced Hamiltonian $H_{\mathrm{tr}}$ of eqn [8] acts. Alternatively, it can be seen as the function obtained by restricting $\Psi$ on the OCS to the TRCS, where $\Psi$ has a dependence along the orbits of the translation group given by $\exp(\mathrm{i}\boldsymbol{R}_{\mathrm{CM}} \cdot \boldsymbol{P}_{\mathrm{CM}}/\hbar)$, that is, by an irreducible representation (irrep) of the translation group.

## Reduction by Rotations: Coordinate Description

The Hamiltonian $H_{\mathrm{tr}}$ acts on wave functions $\psi$ defined on the TRCS and has $3n-3$ degrees of freedom. Consider a coordinate transformation to eliminate further degrees of freedom due to the rotational invariance. This coordinate transformation takes us from the Jacobi vectors $\{\boldsymbol{r}_\alpha, \alpha = 1, \ldots, n-1\}$ to orientational and shape coordinates. Shape coordinates are a set of $3n-6$ coordinates $\{q^\mu, \mu = 1, \ldots, 3n-6\}$ that specify the shape of the $n$-particle system, that is, they are $3n-6$ independent functions of the interparticle distances (hence rotationally invariant). We will call the space upon which the $q^\mu$ are coordinates "shape space." For example, in the case of the three-body problem, shape space is the space of all triangles.

As for orientational coordinates, to define them it is necessary first to define a "body frame." We assume we are already given one frame, the "space frame," a fixed inertial frame. The body frame is a 3-frame attached in a conventional way to each shape of the system of particles, which rotates with the particles. The orientational coordinates, to be

denoted by $\{\theta^i, i = 1, 2, 3\}$, are three coordinates (e.g., Euler angles) specifying the SO(3) rotation that maps the space frame into the body frame. We shall write the new coordinates collectively as $\{\theta^i, q^\mu\}$.

There is a great deal of arbitrariness in the choice of a body frame, since for a given shape a body frame can be attached in many ways, the different choices being related by proper rotations. The only requirement is that the body frame should change smoothly as the shape changes. Popular choices for the body frame are the principal axis and Eckart frames.

When the potential energy is transformed to the new coordinates, it becomes a function only of the $\{q^\mu\}$, that is, of the shape. The potential can be written as $V = V(q)$. $V$ is a scalar field on shape space.

The transformation of the kinetic energy is more complicated. When the (Euclidean) metric tensor on the TRCS is transformed to orientational and shape coordinates there results a $(3n-3)\times(3n-3)$ component matrix which may be partitioned into blocks according to the coordinates $\{\theta^i, q^\mu\}$, that is, according to $3n-3 = 3+(3n-6)$. This matrix cannot be made diagonal or even block diagonal by any choice of orientational or shape coordinates, or by any choice of body frame.

The components of the metric tensor in the new coordinates are conveniently expressed in terms of three fields on shape space. The first is the moment-of-inertia tensor $\mathsf{E}$, which describes the $3\times 3$ upper block of the metric tensor. Its components are given by

$$E_{ij} = M\sum_{\alpha=1}^{n-1}\left(|\boldsymbol{r}_\alpha|^2\,\delta_{ij} - r_{\alpha i}r_{\alpha j}\right) \qquad [9]$$

The vectors and tensors in this equation can be referred either to the space frame or the body frame, but the body frame is more convenient because then the components of the vectors $r_\alpha$ are functions only of the shape coordinates $q^\mu$. Thus, the body frame components $E_{ij}$ of the moment-of-inertia tensor define a field on shape space.

The second field is the "gauge potential" $\boldsymbol{A}_\mu$, an object with $3(3n-6)$ components $A^i_\mu, i = 1, 2, 3, \mu = 1, \ldots, 3n-6$, which describes the off-diagonal blocks of the metric tensor. It is defined by

$$\boldsymbol{A}_\mu = \mathsf{E}^{-1}\left(M\sum_{\alpha=1}^{n-1}\boldsymbol{r}_\alpha \times \frac{\partial \boldsymbol{r}_\alpha}{\partial q^\mu}\right) \qquad [10]$$

in which all vectors are understood to be referred to the body frame (so the partial derivatives make sense). The gauge potential $\boldsymbol{A}_\mu$ is responsible for the "falling cat" phenomenon, in which a flexible body of zero angular momentum nevertheless manages to rotate.

The third field is the $(3n-6)\times(3n-6)$ lower block of the metric tensor on the TRCS, an object with two shape indices. It is given by

$$g_{\mu\nu} = M\sum_{\alpha=1}^{n-1}\left(\frac{\partial \boldsymbol{r}_\alpha}{\partial q^\mu}\cdot\frac{\partial \boldsymbol{r}_\alpha}{\partial q^\nu}\right) - \boldsymbol{A}_\mu\cdot\mathsf{E}\cdot\boldsymbol{A}_\nu \qquad [11]$$

where again the vectors are referred to the body frame. The notation suggests (correctly) that $g_{\mu\nu}$ is the metric tensor on shape space.

On transforming the wave function from the Jacobi vectors to coordinates $(\theta^i, q^\mu)$, it is convenient to introduce a Jacobian factor, $\psi(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-1}) = D^{1/4}\phi(\theta^i, q^\mu)$, where $D = (\det \mathsf{E})(\det g_{\mu\nu})$. This causes the new wave function $\phi$ to have the normalization

$$\int \mathrm{d}\mathsf{R}\left(\prod_{\mu=1}^{3n-6}\mathrm{d}q^\mu\right)|\phi|^2 \qquad [12]$$

where $\mathrm{d}\mathsf{R}$ is the Haar measure on the group SO(3). The factor $D$ depends only on the $q^\mu$, not the $\theta^i$. Then the Schrödinger equation can be written as $H_{\mathrm{tr}}\phi = E_{\mathrm{tr}}\phi$, where $H_{\mathrm{tr}}$ is a differential operator involving $\partial/\partial\theta^i$ and $\partial/\partial q^\mu$.

The orientational derivatives $\partial/\partial\theta^i$ in $H_{\mathrm{tr}}$ are conveniently expressed in terms of the angular momentum operator $\boldsymbol{L}$. When acting on the original wave function $\Psi$ on the OCS, the angular momentum is

$$\boldsymbol{L} = \sum_{\alpha=1}^{n}\boldsymbol{R}_\alpha \times \boldsymbol{P}_\alpha \qquad [13]$$

When this is transformed to the coordinates $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-1}, \boldsymbol{R}_{\mathrm{CM}})$, it becomes $\boldsymbol{L} = \boldsymbol{L}_{\mathrm{CM}} + \boldsymbol{L}_{\mathrm{tr}}$, where $\boldsymbol{L}_{\mathrm{CM}} = \boldsymbol{R}_{\mathrm{CM}}\times\boldsymbol{P}_{\mathrm{CM}}$, and

$$\boldsymbol{L}_{\mathrm{tr}} = \sum_{\alpha=1}^{n-1}\boldsymbol{r}_\alpha \times \boldsymbol{p}_\alpha \qquad [14]$$

Physically, $\boldsymbol{L}_{\mathrm{tr}}$ is the angular momentum of the system about the center of mass.

We shall henceforth drop the "tr" on $H_{\mathrm{tr}}, E_{\mathrm{tr}}$, and $\boldsymbol{L}_{\mathrm{tr}}$, thereby restricting attention to the energy and angular momentum about the center of mass.

The angular momentum $\boldsymbol{L}$, when acting on wave functions $\psi(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{n-1})$ on the TRCS, is a vector of differential operators involving $\partial/\partial\boldsymbol{r}_\alpha$. When these are transformed to orientational and shape coordinates, the components of $\boldsymbol{L}$ become differential operators involving only orientational derivatives, $\partial/\partial\theta^i$. There are no shape derivatives, $\partial/\partial q^\mu$, since $\boldsymbol{L}$ generates rotations, that is, changes in orientation, not shape. Thus, one can solve for the operators $\partial/\partial\theta^i$ in terms of the components of $\boldsymbol{L}$. This is true

both for the space and the body components of $\boldsymbol{L}$, although the differential operators are not the same in the two cases. The space components of $\boldsymbol{L}$ satisfy the usual angular momentum commutation relations, $[L_i, L_j] = i\hbar\epsilon_{ijk} L_k$, while the body components of satisfy $[L_i, L_j] = -i\hbar\epsilon_{ijk} L_k$ (with a minus sign relative to the space commutation relations).

Thus, the Hamiltonian can be expressed in terms of $\boldsymbol{L}$ and the shape momentum operators, $p_\mu = -i\hbar\partial/\partial q^\mu$. The result is

$$H = \frac{1}{2}\boldsymbol{L}\cdot\mathsf{E}^{-1}\cdot\boldsymbol{L} + \frac{1}{2}(p_\mu - \boldsymbol{L}\cdot\boldsymbol{A}_\mu)g^{\mu\nu}(p_\nu - \boldsymbol{L}\cdot\boldsymbol{A}_\nu) + V_2(q) + V(q) \qquad [15]$$

where all vectors are referred to the body frame, where $g^{\mu\nu}$ is the contravariant metric tensor on shape space, and where $V_2$ is given by

$$V_2 = \frac{\hbar^2}{2} D^{-1/4} \frac{\partial}{\partial q^\mu}\left(g^{\mu\nu}\frac{\partial D^{1/4}}{\partial q^\nu}\right) \qquad [16]$$

$V_2$ looks like a potential (it is a function of only $q$), hence the notation, but physically it belongs to the kinetic energy. It is sometimes called an "extrapotential." It arises from nonclassical commutators in the transformation of the kinetic energy (hence the $\hbar^2$ dependence). The first term of eqn [15] is the kinetic energy of rotation, also called the "vertical" kinetic energy, the next two terms are the remainder of the kinetic energy, somewhat imprecisely thought of as the kinetic energy of vibrations or changes in shape, also called the "horizontal kinetic energy," and the final term is the (true) potential, discussed above.

Since the Hamiltonian commutes with the angular momentum, $[H, \boldsymbol{L}] = 0$, $\phi$ can be chosen to be simultaneous eigenfunctions of $L^2$ and $L_z$ (the latter being the space component), as well as of energy. Let $\phi_{lm}$ be these eigenfunctions, where $l$ and $m$ are the quantum numbers of $L^2$ and $L_z$, respectively. Then by the transformation properties of $\phi$ under rotations, we can write

$$\phi_{lm}(\theta^i, q^\mu) = \sum_{k=-l}^{+l} \chi_{lk}(q^\mu)\, D^l_{km}(\theta^i) \qquad [17]$$

where $D$ is a standard rotation matrix and $\chi_{lk}$ are functions only of $q^\mu$. In these equations we use the phase and other standard conventions of the theory of rotations. The wave function $\chi$ is a function only of $q^\mu$ and can loosely be thought of as the wave function on shape space. It is not a scalar like $\Psi$, $\psi$, or $\phi$, but rather has $2l+1$ components indexed by $k$.

The Schrödinger equation for $\chi$ can be written as $H\chi = E\chi$, where $H$ has the same form as in eqn [15], except that now the components of the angular momentum $L_i$ are interpreted, no longer as differential operators in $\theta^i$, but as $(2l+1)\times(2l+1)$ matrices that act on the "spinor" $\chi$. These matrices are the transposes of the usual angular momentum matrices in angular momentum theory, that is, $(L_i)_{kk'} = \langle k'|L_i|k\rangle$.

This is the final form of the Schrödinger equation after all reductions by all continuous symmetries have been carried out. The fully reduced system has $3n-5$ degrees of freedom ($3n-6$ for the shape coordinates, and one for the "spinor" index $k$).

## Reduction by Rotations: Geometrical Description

The proper rotation group SO(3) acts on the OCS by $\boldsymbol{R}_\alpha \mapsto \mathsf{R}\boldsymbol{R}_\alpha$, and on the TRCS by $\boldsymbol{r}_\alpha \mapsto \mathsf{R}\boldsymbol{r}_\alpha$, where $\mathsf{R} \in SO(3)$. Rotations acting on the OCS do not commute with translations, but the action preserves the translation fibers, and thus can be projected onto the TRCS.

The action of SO(3) on the TRCS is effective but not free, that is, most orbits are diffeomorphic to SO(3), but a subset of measure zero (the "singular" orbits) are diffeomorphic to $S^2$ or a single point. Configurations of the $n$-particle system in which the particles do not lie on a line ("noncollinear shapes") have SO(3) orbits, those in which the particles do lie on a line but are not coincident have $S^2$ orbits, and the $n$-body collision (a single shape) has an orbit that is a single point. Thus, the action of SO(3) on the TRCS foliates the TRCS into a $(3n-6)$-parameter family of copies of SO(3), plus the singular orbits. If we exclude the singular orbits, then the TRCS has the structure of an SO(3) principal fiber bundle. In general, the bundle is not trivial. Shape space may be defined as the quotient space under the SO(3) action. Omitting the singular shapes, shape space is the base space of the bundle. The coordinates $q^\mu$ introduced above are coordinates on shape space. The singular shapes and orbits are physically accessible, and there are important questions regarding the behavior of the system in their neighborhood.

The definition of a body frame is equivalent to the choice of a section of the fiber bundle, generally only locally defined over some region of shape space. A configuration (a point in the TRCS) on the section defines an orientation of the $n$-particle system for the given shape, which serves as a reference orientation to which others can be referred. We think of the reference orientation as one in which the space and body frames coincide; in other orientations of the same shape, the body frame has been rotated with the body to a new orientation. The choice of the section (body frame) allows us to

impose coordinates on each (nonsingular) rotation fiber, that is, we label points on the fiber by the rotation that takes us from the section to the actual configuration in question. This is why a choice of body frame is necessary before defining orientational coordinates. Sections are only defined locally. Popular choices of body frame, such as the principal axis frame, imply multivalued sections, unless branch cuts are introduced. Orientational coordinates are simply coordinates on the group manifold SO(3), transferred to the nonsingular rotation fibers, with the group identity element mapped onto the point where the fiber intersects the section.

The metric tensor determines much of the geometry of the reduction by rotations. Since the metric on the TRCS is SO(3)-invariant, horizontal subspaces in the SO(3) fiber bundle (the TRCS minus the singular orbits) can be defined as the spaces orthogonal to the fibers (hence orthogonal to the vertical subspaces). This is a standard construction in Kaluza–Klein theories, which reappears here. Thus, the bundle has a connection, induced by the metric.

The moment-of-inertia tensor is the metric tensor restricted to a fiber, evaluated in a basis of left- (body frame) or right-invariant (space frame) vector fields on SO(3), which are transported to the fibers to create a basis of vertical vector fields.

The coordinate description of the connection is the gauge potential $\boldsymbol{A}_\mu$, in which the $\mu$ index refers to shape coordinates $q^\mu$, and the components of the 3-vector $\boldsymbol{A}$ refer to the standard set of left- or right-invariant vector fields on SO(3). The coordinate representative of the curvature 2-form is conveniently denoted by $\boldsymbol{B}_{\mu\nu}$, defined by

$$\boldsymbol{B}_{\mu\nu} = \frac{\partial \boldsymbol{A}_\nu}{\partial q^\mu} - \frac{\partial \boldsymbol{A}_\mu}{\partial q^\nu} - \boldsymbol{A}_\mu \times \boldsymbol{A}_\nu \qquad [18]$$

where it is understood that body frame components are used. Direct calculation shows that it is nonzero, hence the fiber bundle is not flat, for any value of $n \geq 3$. The curvature form $\boldsymbol{B}_{\mu\nu}$ appears in the classical equation of motion and in the quantum commutation relations.

The field $\boldsymbol{B}_{\mu\nu}$ satisfies differential equations on shape space that have the form of Yang–Mills field equations. It is interesting that the sources of this field are singularities of the monopole type, located on the singular shapes. In the case $n = 3$, the source is a single monopole located at the three-body collision, which is similar to a Dirac monopole in electromagnetic theory.

The $(3n - 6)$-dimensional horizontal subspaces of the TRCS are annihilated by three differential forms, whose values on a velocity vector of the system are the components of the classical angular momentum $\boldsymbol{L}$ (body or space components, depending on the basis of forms). Thus, horizontal motions are those for which $\boldsymbol{L} = 0$, and horizontal lifts of curves in shape space are motions of the system with vanishing angular momentum. Since angular momentum is conserved, such motions are generated by the classical equations of motion and are physically allowed. For loops in shape space, the holonomy generated by the horizontal lift is physically the rotation that a flexible body experiences when it is carried under conditions of vanishing angular momentum from an initial shape, through intermediate shapes and back to the initial shape. An example is the rotation generated by the "falling cat."

Since the metric on the TRCS is SO(3)-invariant, it may be projected onto shape space, which therefore is a Riemannian manifold in its own right. The projected metric is $ds^2 = g_{\mu\nu}\, dq^\mu\, dq^\nu$. This metric is not flat (the Riemann curvature tensor is nonzero for all values $n \geq 3$). Geodesics in shape space have horizontal lifts that are free particle motions ($V = 0$) of zero angular momentum. Conversely, such motions project onto geodesics on shape space.

A popular choice of body frame in molecular physics is the Eckart frame, which has advantages for the description of small vibrations and other purposes. The section defining the Eckart frame is a flat vector subspace of the TRCS of dimension $3n - 6$ that is orthogonal (horizontal) to a particular fiber (over an equilibrium shape) at a particular orientation.

The geometrical meaning of eqn [17] is that rotations act on a set of wave functions $\phi$ that span an irrep of SO(3) by multiplication by the representative element of the group. In standard physics notation, $l$ indexes the irrep, and $m$ indexes the basis vectors spanning the irrep. Thus, the values of these wave functions at any point on the fiber are known once their values are given at a reference point. A convenient choice for the reference point is the point on the section, and the wave functions $\chi_{lk}$ are simply the values of the $\phi_{lm}$ on this reference point (with a change of notation, $m \to k$). Thus, the wave functions $\chi_{lk}$ are properly not "wave functions on shape space," but rather wave functions on the section.

Shape space in the case $n = 3$ is homeomorphic to the region $x_3 \geq 0$ of $\mathbb{R}^3$, and in the case $n = 4$ to $\mathbb{R}^6$. A convenient tool for understanding the structure of shape space is by its foliation under the action of the kinematic rotations, eqn [5]. The kinematic rotations commute with ordinary rotations, and hence have an action on shape space. This action preserves the eigenvalues of the moment-of-inertia tensor.

## Concluding Remarks

The quantum $n$-body problem provides an interesting example in which nonabelian gauge theories find application in nonrelativistic quantum mechanics. The fields $\mathsf{E}$, $A_\mu$, and $g_{\mu\nu}$, and fields derived from them such as the curvature tensor $B_{\mu\nu}$ and the Riemann curvature tensor derived from $g_{\mu\nu}$, satisfy a complex set of differential equations on shape space that can be derived by considering the vanishing of the Riemann tensor on the TRCS. The resulting field equations are useful in perturbation theory, for example, in the study of small vibrations of a molecule. This means of constructing field equations on the base space of a bundle is standard in Kaluza–Klein theories, which are an important line of thinking in modern attempts to understand gauge field theories in particle physics.

The rotations generated by flexible bodies of vanishing angular momentum (the "falling cat") are an example of a "geometric phase," that is, a nonabelian generalization of "Berry's phase." It is interesting how the associated gauge potential $A_\mu$ in this problem plays a role in the dynamics of the $n$-particle system.

The Hamiltonian [15] is the starting point for numerous practical calculations, for example, the numerical evaluation of energy levels, cross-sections and reaction rates in molecular physics. One can compute, for example, chemical reaction rates for molecular processes in atmospheric or astrophysical contexts, where experiments would be difficult or expensive. The numerical analysis of the Hamiltonian [15] usually requires the introduction of a basis set and the processing of large matrices. Current techniques for basis set selection are not very satisfactory, and this is an area where research into wavelets and numerical analysis could have an impact.

*See also:* Bosons and Fermions in External Fields; Gravitational $N$-Body Problem (Classical); Integrable Systems: Overview.

## Further Reading


Abraham R and Marsden JE (1978) *Foundations of Mechanics*, 2nd edn. Reading, MA: Benjamin/Cummings.

Aquilanti V and Cavalli S (1986) Coordinates for molecular dynamics: orthogonal local frames. *Journal of Chemical Physics* 85: 1355–1361.

Berry MV (1984) Quantal phase factors accompanying adiabatic changes. *Proceedings of the Royal Society of London. Series A* 392: 45–57.

Coquereaux R (1988) *Riemannian Geometry, Fiber bundles, Kaluza–Klein Theories and All That*. World Scientific Lecture Notes in Physics, vol. 16. Singapore: World Scientific.

Ezra GS (1982) *Symmetry Properties of Molecules*. New York: Springer.

Iwai T (1987) A gauge theory for the quantum planar three-body problem. *Journal of Mathematical Physics* 28: 964–974.

Kupperman A (1993) A new look at symmetrized hyperspherical coordinates. *Advances in Molecular Vibrations and Collision Dynamics* vol. 2B: 117–186.

Littlejohn RG and Reinsch M (1997) Gauge fields in the separation of rotations and internal motions in the $n$-body problem. *Reviews of Modern Physics* 69: 213–275.

Mead CA (1992) The geometric phase in molecular systems. *Reviews of Modern Physics* 64: 51–85.

Mitchell KA and Littlejohn RG (2000) Kinematic orbits and the structure of the internal space for systems of five or more bodies. *Journal of Physics A* 33: 1395–1416.

Naber GL (1997) *Topology, Geometry and Gauge Fields*. New York: Springer.

Wilson EB Jr., Decius JC, and Cross PC (1955) *Molecular Vibrations: the Theory of Infrared and Raman Vibrational Spectra*. New York: McGraw-Hill.


# Quantum Phase Transitions

**S Sachdev**, Yale University, New Haven, CT, USA



## Introduction

The study of second-order phase transitions at nonzero temperatures has a long and distinguished history in statistical mechanics. Many key physical phenomena, such as the loss of ferromagnetism in iron at the Curie temperature or the critical endpoint of $CO_2$, are now understood in precise quantitative detail. This understanding began in the work of Onsager, and is based upon what may now be called the Landau–Ginzburg–Wilson theory. The content of this sophisticated theory may be summarized in a few basic principles: (1) The collective thermal fluctuations near second-order transitions can be accurately described by simple classical models, that is, quantum-mechanical effects can be entirely neglected. (2) The classical models identify an "order parameter," a collective variable which has to be treated on par with other thermodynamic variables, and whose correlations exhibit distinct behavior in the phases on either side of the transition. (3) The thermal fluctuations of the order parameter near the transition are controlled by a continuum field theory whose structure is usually completly dictated by simple symmetry considerations.

This article will not consider such nonzero temperature phase transitions, but will instead describe second-order phase transitions at the absolute zero of temperature. Such transitions are driven by quantum fluctuations mandated by the Heisenberg uncertainty principle: one can imagine moving across the quantum critical point by effectively "tuning the value of Planck's constant, $\hbar$." Clearly, quantum mechanics plays a central role at such transitions, unlike the situation at nonzero temperatures. The reader may object that absolute zero is an idealization not realized by any experimental system; hence, the study of quantum phase transitions is a subject only of academic interest. As we will illustrate below, knowledge of the zero-temperature quantum critical points of a system is often the key to understanding its finite-temperature properties, and in some cases the influence of a zero-temperature critical point can be detected at temperatures as high as ambient room temperature.

We will begin in the following section by introducing some simple lattice models which exhibit quantum phase transitions. Next the theory of the critical point in these models is based upon a natural extension of the Landau–Ginzburg–Wilson (LGW) method, and this will be presented. This section will also describe the consequences of a zero-temperature critical point on the nonzero temperature properties. Finally, we will consider more complex models in which quantum interference effects play a more subtle role, and which cannot be described in the LGW framework: such quantum critical points are likely to play a central role in understanding many of the correlated electron systems of current interest.

## Simple Models

### Quantum Ising Chain

This is a simple model of $N$ qubits, labeled by the index $j = 1, \ldots, N$. On each "site" $j$ there are two qubit quantum states $|\uparrow\rangle_j$ and $|\downarrow\rangle_j$ (in practice, these could be two magnetic states of an ion at site $j$ in a crystal). The Hilbert space therefore consists of $2^N$ states, each consisting of a tensor product of the states on each site. We introduce the Pauli spin operators, $\hat{\sigma}_j^\alpha$, on each site $j$, with $\alpha = x, y, z$:

$$\hat{\sigma}^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \hat{\sigma}^y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$
$$\hat{\sigma}^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad [1]$$

These operators clearly act on the two states of the qubit on site $j$, and the Pauli operators on different sites commute.

The quantum Ising chain is defined by the simple Hamiltonian

$$H_\mathrm{I} = -J \sum_{j=1}^{N-1} \hat{\sigma}_j^z \hat{\sigma}_{j+1}^z - gJ \sum_{j=1}^{N} \hat{\sigma}_j^x \qquad [2]$$

where $J > 0$ sets the energy scale, and $g \geq 0$ is a dimensionless coupling constant. In the thermodynamic limit ($N \to \infty$), the ground state of $H_\mathrm{I}$ exhibits a second-order quantum phase transition as $g$ is tuned across a critical value $g = g_\mathrm{c}$ (for the specific case of $H_\mathrm{I}$ it is known that $g_\mathrm{c} = 1$), as we will now illustrate.

First, consider the ground state of $H_\mathrm{I}$ for $g \ll 1$. At $g = 0$, there are two degenerate "ferromagnetically ordered" ground states

$$|\Uparrow\rangle = \prod_{j=1}^{N} |\uparrow\rangle_j, \qquad |\Downarrow\rangle = \prod_{j=1}^{N} |\downarrow\rangle_j \qquad [3]$$

Each of these states breaks a discrete "Ising" symmetry of the Hamiltonian rotations of all spins by 180° about the $x$-axis. These states are more succinctly characterized by defining the ferromagnetic moment, $N_0$, by

$$N_0 = \langle \Uparrow | \hat{\sigma}_j^z | \Uparrow \rangle = -\langle \Downarrow | \hat{\sigma}_j^z | \Downarrow \rangle \qquad [4]$$

At $g = 0$ we clearly have $N_0 = 1$. A key point is that in the thermodynamic limit, this simple picture of the ground state survives for a finite range of small $g$ (indeed, for all $g < g_\mathrm{c}$), but with $0 < N_0 < 1$. The quantum tunneling between the two ferromagnetic ground states is exponentially small in $N$ (and so can be neglected in the thermodynamic limit), and so the ground state remains 2-fold degenerate and the discrete Ising symmetry remains broken. The change in the wave functions of these states from eqn [3] can be easily determined by perturbation theory in $g$: these small $g$ quantum fluctuations reduce the value of $N_0$ from unity but do not cause the ferromagnetism to disappear.

Now consider the ground state of $H_\mathrm{I}$ for $g \gg 1$. At $g = \infty$ there is a single nondegenerate ground state which fully preserves all symmetries of $H_\mathrm{I}$:

$$\Rightarrow\rangle = 2^{-N/2} \prod_{j=1}^{N} \left( |\uparrow\rangle_j + |\downarrow\rangle_j \right) \qquad [5]$$

It is easy to verify that this state has no ferromagnetic moment $N_0 = \langle \Rightarrow | \hat{\sigma}_j^z | \Rightarrow \rangle = 0$. Further, perturbation theory in $1/g$ shows that these features of the ground state are preserved for a finite range of large $g$ values

(indeed, for all $g > g_c$). One can visualize this ground state as one in which strong quantum fluctuations have destroyed the ferromagnetism, with the local magnetic moments quantum tunneling between "up" and "down" on a timescale of order $\hbar/J$.

Given the very distinct signatures of the small $g$ and large $g$ ground states, it is clear that the ground state cannot evolve smoothly as a function of $g$. These must be at least one point of nonanalyticity as a function of $g$: for $H_I$ it is known that there is only a single nonanalytic point, and this is at the location of a second-order quantum phase transition at $g = g_c = 1$.

The character of the excitations above the ground state also undergoes a qualitative change across the quantum critical point. In both the $g < g_c$ and $g > g_c$ phases, these excitations can be described in the Landau quasiparticle scheme, that is, as superpositions of nearly independent particle-like excitations; a single well-isolated quasiparticle has an infinite lifetime at low excitation energies. However, the physical nature of the quasiparticles is very different in the two phases. In the ferromagnetic phase, with $g < g_c$, the quasiparticles are domain walls between regions of opposite magnetization:

$$|j, j+1\rangle = \prod_{k=1}^{j} |\uparrow\rangle_k \prod_{\ell=j+1}^{N} |\downarrow\rangle_\ell \qquad [6]$$

This is the exact wave function of a stationary quasiparticle excitation between sites $j$ and $j+1$ at $g = 0$; for small nonzero $g$ the quasiparticle acquires a "cloud" of further spin-flips and also becomes mobile. However its qualitative interpretation as a domain wall between the two degenerate ground states remains valid for all $g < g_c$. In contrast, for $g > g_c$, there is no ferromagnetism, and the nondegenerate paramagnetic state has a distinct quasiparticle excitation:

$$|j\rangle = 2^{-N/2}\left(|\uparrow\rangle_j - |\downarrow\rangle_j\right) \prod_{k \neq j} \left(|\uparrow\rangle_k + |\downarrow\rangle_k\right) \qquad [7]$$

This is a stationary "flipped spin" quasiparticle at site $j$, with its wave function exact at $g = \infty$. Again, this quasiparticle is mobile and applicable for all $g > g_c$, but there is no smooth connection between eqns [7] and [6].

## Coupled Dimer Antiferromagnet

This model also involves qubits, but they are now placed on the sites, $j$, of a two-dimensional square lattice. Models in this class describe the magnetic

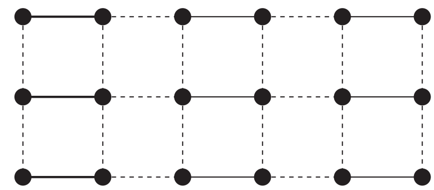


**Figure 1** The coupled dimer antiferromagnet. Qubits (i.e., $S = 1/2$ spins) are placed on the sites, the $\mathcal{A}$ links are shown as full lines, and the $\mathcal{B}$ links as dashed lines.

excitations of many experimentally important spin gap compounds.

The Hamiltonian of the dimer antiferromagnet is illustrated in Figure 1 and is given by

$$H_d = J \sum_{\langle jk \rangle \in A} \left( \hat{\sigma}_j^x \hat{\sigma}_k^x + \hat{\sigma}_j^y \hat{\sigma}_k^y + \hat{\sigma}_j^z \hat{\sigma}_k^z \right) + \frac{J}{g} \sum_{\langle jk \rangle \in B} \left( \hat{\sigma}_j^x \hat{\sigma}_k^x + \hat{\sigma}_j^y \hat{\sigma}_k^y + \hat{\sigma}_j^z \hat{\sigma}_k^z \right) \qquad [8]$$

where $J > 0$ is the exchange constant, $g \geq 1$ is the dimensionless coupling, and the set of nearest-neighbor links $\mathcal{A}$ and $\mathcal{B}$ are defined in Figure 1. An important property of $H_d$ is that it is now invariant under the full O(3) group of spin rotations under which the $\hat{\sigma}^\alpha$ transform as ordinary vectors (in contrast to the $Z_2$ symmetry group of $H_I$). In analogy with $H_I$, we will find that $H_d$ undergoes a quantum phase transition from a paramagnetic phase which preserves all symmetries of the Hamiltonian at large $g$, to an antiferromagnetic phase which breaks the O(3) symmetry at small $g$. This transition occurs at a critical value $g = g_c$, and the best current numerical estimate is $1/g_c = 0.52337(3)$.

As in the previous section, we can establish the existence of such a quantum phase transition by contrasting the disparate physical properties at large $g$ with those at $g \approx 1$. At $g = \infty$ the exact ground state of $H_d$ is

$$|\text{spin gap}\rangle = \prod_{\langle jk \rangle \in A} \frac{1}{\sqrt{2}} \left( |\uparrow\rangle_j |\downarrow\rangle_k - |\downarrow\rangle_j |\uparrow\rangle_k \right) \qquad [9]$$

and is illustrated in Figure 2. This state is nondegenerate and invariant under spin rotations, and so is a paramagnet: the qubits are paired into spin singlet valence bonds across all the $\mathcal{A}$ links.

The excitations above the ground state are created by breaking a valence bond, so that the pair of spins form a spin triplet with total spin $S = 1$ – this is illustrated in Figure 3. It costs a large energy to create this excitation, and at finite $g$ the

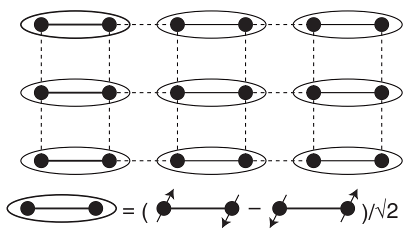



**Figure 2** The paramagnetic state of $H_d$ for $g > g_c$. The state illustrated is the exact ground state for $g = \infty$, and it is adiabatically connected to the ground state for all $g > g_c$.

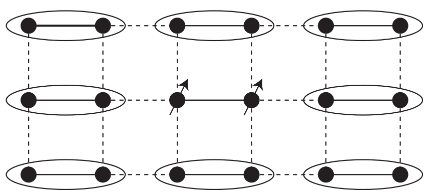


**Figure 3** The triplon excitation of the $g > g_c$ paramagnet. The stationary triplon is an eigenstate only for $g = \infty$ but it becomes mobile for finite $g$.

triplet can hop from link to link, creating a gapped "triplon" quasiparticle excitation. This is similar to the large $g$ paramagnet for $H_I$, with the important difference that each quasiparticle is now 3-fold degenerate.

At $g = 1$, the ground state of $H_d$ is not known exactly. However, at this point $H_d$ becomes equivalent to the nearest-neighbor square lattice antiferromagnet, and this is known to have antiferromagnetic order in the ground state, as illustrated in Figure 4. This state is similar to the ferromagnetic ground state of $H_I$, with the difference that the magnetic moment now acquires a staggered pattern on the two sublattices, rather than the uniform moment of the ferromagnet. Thus, in this ground state

$$\langle \mathrm{AF}|\hat{\sigma}_j^\alpha|\mathrm{AF}\rangle = N_0 \eta_j n_\alpha \qquad [10]$$

where $0 < N_0 < 1$ is the antiferromagnetic moment, $\eta_j = \pm 1$ identifies the two sublattices in Figure 4, and $n_\alpha$ is an arbitrary unit vector specifying the orientation of the spontaneous magnetic moment which breaks the O(3) spin rotation invariance of $H_d$. The excitations above this antiferromagnet are also distinct from those of the paramagnet: they are a doublet of spin waves consisting of a spatial variation in the local orientation, $n_\alpha$, of the antiferromagnetic order: the energy of this excitation vanishes in the limit of long wavelengths, in contrast to the finite energy gap of the triplon excitation of the paramagnet.

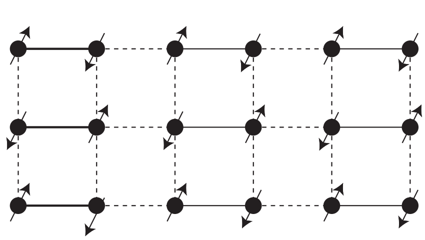


**Figure 4** Schematic of the ground state with antiferromagnetic order with $g < g_c$.

As with $H_I$, we can conclude from the distinct characters of the ground states and excitations for $g \gg 1$ and $g \approx 1$ that there must be a quantum critical point at some intermediate $g = g_c$.

## Quantum Criticality

The simple considerations of the previous section have given a rather complete description (based on the quasiparticle picture) of the physics for $g \ll g_c$ and $g \gg g_c$. We turn, finally, to the region $g \approx g_c$. For the specific models discussed in the previous section, a useful description is obtained by a method that is a generalization of the LGW method developed earlier for thermal phase transitions. However, some aspects of the critical behavior (e.g., the general forms of eqns [13]–[15]) will apply also to the quantum critical point of the section "Beyond LGW theory."

Following the canonical LGW strategy, we need to identify a collective order parameter which distinguishes the two phases. This is clearly given by the ferromagnetic moment in eqn [4] for the quantum Ising chain, and the antiferromagnetic moment in eqn [10] for the coupled dimer antiferromagnet. We coarse-grain these moments over some finite averaging region, and at long wavelengths this yields a real order parameter field $\phi_a$, with the index $a = 1, \ldots, n$. For the Ising case we have $n = 1$ and $\phi_a$ is a measure of the local average of $N_0$ as defined in eqn [4]. For the antiferromagnet, $a$ extends over the three values $x, y, z$ (so $n = 3$), and three components of $\phi_a$ specify the magnitude and orientation of the local antiferromagnetic order in eqn [10]; note the average orientation of a specific spin at site $j$ is $\eta_j$ times the local value of $\phi_a$.

The second step in the LGW approach is to write down a general field theory for the order parameter, consistent with all symmetries of the underlying model. As we are dealing with a quantum transition, the field theory has to extend over spacetime, with the temporal fluctuations representing the sum over histories in the Feynman path-integral approach. With this reasoning, the proposed partition function

for the vicinity of the critical point takes the following form:

$$\mathcal{Z}_\phi = \int \mathcal{D}\phi_a(x,\tau) \times \exp\left[-\int d^d x\, d\tau \left(\frac{1}{2}\left((\partial_\tau \phi_a)^2 + c^2(\nabla_x \phi_a)^2 + s\phi_a^2\right) + \frac{u}{4!}\left(\phi_a^2\right)^2\right)\right] \quad [11]$$

Here $\tau$ is imaginary time; there is an implied summation over the $n$ values of the index $a$, $c$ is a velocity, and $s$ and $u > 0$ are coupling constants. This is a field theory in $d+1$ spacetime dimensions, in which the Ising chain corresponds to $d=1$ and the dimer antiferromagnet to $d=2$. The quantum phase transition is accessed by tuning the "mass" $s$: there is a quantum critical point at $s=s_c$ and the $s<s_c(s>s_c)$ regions correspond to the $g<g_c(g>g_c)$ regions of the lattice models. The $s<s_c$ phase has $\langle\phi_a\rangle \neq 0$ and this corresponds to the spontaneous breaking of spin rotation symmetry noted in eqns [4] and [10] for the lattice models. The $s>s_c$ phase is the paramagnet with $\langle\phi_a\rangle=0$. The excitations in this phase can be understood as small harmonic oscillations of $\phi_a$ about the point (in field space) $\phi_a=0$. A glance at eqn [11] shows that there are $n$ such oscillators for each wave vector. These oscillators clearly constitute the $g>g_c$ quasiparticles found earlier in eqn [7] for the Ising chain (with $n=1$) and the triplon quasiparticle (with $n=3$) illustrated in **Figure 3** for the dimer antiferromagnet.

We have now seen that there is a perfect correspondence between the phases of the quantum field theory $\mathcal{Z}_\phi$ and those of the lattice models $H_I$ and $H_d$. The power of the representation in eqn [11] is that it also allows us to get a simple description of the quantum critical point. In particular, readers may already have noticed that if we interpret the temporal direction $\tau$ in eqn [11] as another spatial direction, then $\mathcal{Z}_\phi$ is simply the classical partition function for a thermal phase transition in a ferromagnet in $d+1$ dimensions: this is the canonical model for which the LGW theory was originally developed. We can now take over standard results for this classical critical point, and obtain some useful predictions for the quantum critical point of $\mathcal{Z}_\phi$. It is useful to express these in terms of the dynamic susceptibility defined by

$$\chi(k,\omega) = \frac{i}{\hbar}\int d^d x \times \int_0^\infty dt \left\langle \left[\hat{\phi}(x,t), \hat{\phi}(0,0)\right]\right\rangle_T e^{-ikx+i\omega t} \quad [12]$$

Here $\hat{\phi}$ is the Heisenberg field operator corresponding to the path integral in eqn [11], the square brackets represent a commutator, and the angular brackets an average over the partition function at a temperature $T$. The structure of $\chi$ can be deduced from the knowledge that the quantum correlators of $\mathcal{Z}_\phi$ are related by analytic continuation in time to the corresponding correlators of the classical statistical mechanics problem in $d+1$ dimensions. The latter are known to diverge at the critical point as $\sim 1/p^{2-\eta}$ where $p$ is the $(d+1)$-dimensional momentum, $\eta$ is defined to be the anomalous dimension of the order parameter ($\eta=1/4$ for the quantum Ising chain). Knowing this, we can deduce the form of the quantum correlator in eqn [12] at the zero-temperature quantum critical point

$$\chi(k,\omega) \sim \frac{1}{(c^2k^2-\omega^2)^{1-\eta/2}}, \quad T=0, \quad g=g_c \quad [13]$$

The most important property of eqn [13] is the absence of a quasiparticle pole in the spectral density. Instead, $\mathrm{Im}(\chi(k,\omega))$ is nonzero for all $\omega > ck$, reflecting the presence of a continuum of critical excitations. Thus the stable quasiparticles found at low enough energies for all $g \neq g_c$ are absent at the quantum critical point.

We now briefly discuss the nature of the phase diagram for $T>0$ with $g$ near $g_c$. In general, the interplay between quantum and thermal fluctuations near a quantum critical point can be quite complicated, and we cannot discuss it in any detail here. However, the physics of the quantum Ising chain is relatively simple, and also captures many key features found in more complex situations, and is summarized in **Figure 5**. For all $g \neq g_c$ there is a range of low temperatures ($T \lesssim |g-g_c|$) where the long time dynamics can be described using a dilute gas of thermally excited quasiparticles. Further, the

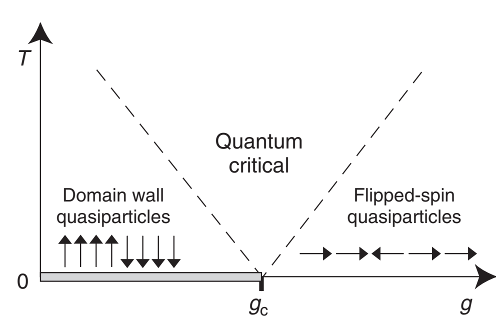



**Figure 5** Nonzero temperature phase diagram of $H_I$. The ferromagnetic order is present only at $T=0$ on the shaded line with $g<g_c$. The dashed lines at finite $T$ are crossovers out of the low-$T$ quasiparticle regimes where a quasiclassical description applies. The state sketched on the paramagnetic side used the notation $|\rightarrow\rangle_j = 2^{-1/2}(|\uparrow\rangle_j + |\downarrow\rangle_j)$ and $|\leftarrow\rangle_j = 2^{-1/2}(|\uparrow\rangle_j - |\downarrow\rangle_j)$.

dynamics of these quasiparticles is quasiclassical, although we reiterate that the nature of the quasiparticles is entirely distinct on opposite sides of the quantum critical point. Most interesting, however, is the novel quantum critical region, $T \gtrsim |g - g_c|$, where neither quasiparticle picture nor a quasiclassical description are appropriate. Instead, we have to understand the influence of temperature on the critical continuum associated with eqn [13]. This is aided by scaling arguments which show that the only important frequency scale which characterizes the spectrum is $k_B T/\hbar$, and the crossovers near this scale are universal, that is, independent of specific microscopic details of the lattice Hamiltonian. Consequently, the zero-momentum dynamic susceptibility in the quantum critical region takes the following form at small frequencies:

$$\chi(k = 0, \omega) \sim \frac{1}{T^{2-\eta}} \frac{1}{(1 - i\omega/\Gamma_R)} \quad [14]$$

This has the structure of the response of an overdamped oscillator, and the damping frequency, $\Gamma_R$, is given by the universal expression

$$\Gamma_R = \left(2 \tan \frac{\pi}{16}\right) \frac{k_B T}{\hbar} \quad [15]$$

The numerical proportionality constant in eqn. [15] is specific to the quantum Ising chain; other models also obey eqn [15] but with a different numerical value for this constant.

## Beyond LGW Theory

The quantum transitions discussed so far have turned to have a critical theory identical to that found for classical thermal transitions in $d+1$ dimensions. Over the last decade it has become clear that there are numerous models, of key physical importance, for which such a simple classical correspondence does not exist. In these models, quantum Berry phases are crucial in establishing the nature of the phases, and of the critical boundaries between them. In less technical terms, a signature of this subtlety is an important simplifying feature which was crucial in the analyses of the section "Simple models": both models had a straightforward $g \to \infty$ limit in which we were able to write down a simple, nondegenerate, ground-state wave function of the "disordered" paramagnet. In many other models, identification of the disordered phase is not as straightforward: specifying absence of a particular magnetic order is not enough to identify a quantum state, as we still need to write down a suitable wave function. Often, subtle quantum interference effects induce new types of order in the disordered state, and such effects are entirely absent in the LGW theory.

An important example of a system displaying such phenomena is the $S = 1/2$ square lattice antiferromagnet with additional frustrating interactions. The quantum degrees of freedom are identical to those of the coupled dimer antiferromagnet, but the Hamiltonian preserves the full point-group symmetry of the square lattice:

$$H_s = \sum_{j<k} J_{jk} \left( \hat{\sigma}_j^x \hat{\sigma}_k^x + \hat{\sigma}_j^y \hat{\sigma}_k^y + \hat{\sigma}_j^z \hat{\sigma}_k^z \right) + \cdots \quad [16]$$

Here the $J_{jk} > 0$ are short-range exchange interactions which preserve the square lattice symmetry, and the ellipses represent possible further multiple spin terms. Now imagine tuning all the non-nearest-neighbor terms as a function of some generic coupling constant $g$. For small $g$, when $H_s$ is nearly the square lattice antiferromagnet, the ground state has antiferromagnetic order as in **Figure 4** and eqn [10]. What is now the disordered ground state for large $g$? One natural candidate is the spin-singlet paramagnet in **Figure 2**. However, because all nearest neighbor bonds of the square lattice are now equivalent, the state in **Figure 2** is degenerate with three other states obtained by successive 90° rotations about a lattice site. In other words, the state in **Figure 2**, when transferred to the square lattice, breaks the symmetry of lattice rotations by 90°. Consequently it has a new type of order, often called valence-bond-solid (VBS) order. It is now believed that a large class of models like $H_s$ do indeed exhibit a second-order quantum phase transition between the antiferromagnetic state and a VBS state – see **Figure 6**. Both the existence of VBS order in the paramagnet, and of a second-order quantum transition, are features that are not predicted by LGW theory: these can only be

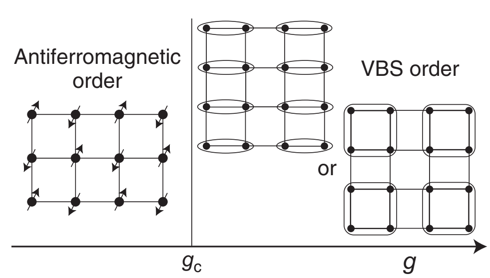



**Figure 6** Phase diagram of $H_s$. Two possible VBS states are shown: one which is the analog of **Figure 2**, and the other in which spins form singlets in a plaquette pattern. Both VBS states have a 4-fold degeneracy due to breaking of square lattice symmetry. So the novel critical point at $g = g_c$ (described by $\mathcal{Z}_z$) has the antiferromagnetic and VBS orders vanishing as it is approached from either side: this coincident vanishing of orders is generically forbidden in LGW theories.

understood by a careful study of quantum interference effects associated with Berry phases of spin fluctuations about the antiferromagnetic state. We will not enter into details of this analysis here, but will conclude our discussion by writing down the theory so obtained for the quantum critical point in Figure 6:

$$\mathcal{Z}_z = \int \mathcal{D}z_\alpha(x,\tau)\mathcal{D}A_\mu(x,\tau) \times \exp\left(-\int \mathrm{d}^2x\,\mathrm{d}\tau\left[|(\partial_\mu - \mathrm{i}A_\mu)z_\alpha|^2 + s|z_\alpha|^2 + \frac{u}{2}(|z_\alpha|^2)^2 + \frac{1}{2e^2}(\epsilon_{\mu\nu\lambda}\partial_\nu A_\lambda)^2\right]\right) \qquad [17]$$

Here $\mu, \nu, \lambda$ are spacetime indices which extend over the two spatial directions and $\tau$, $\alpha$ is a spinor index which extends over $\uparrow$, $\downarrow$, and $z_\alpha$ is complex spinor field. In comparing $\mathcal{Z}_z$ to $\mathcal{Z}_\phi$, note that the vector order parameter $\phi_a$ has been replaced by a spinor $z_\alpha$, and these are related by $\phi_a = z_\alpha^* \sigma^a_{\alpha\beta} z_\beta$, where $\sigma^a$ are the Pauli matrices. So the order parameter has fractionalized into the $z_\alpha$. A second novel property of $\mathcal{Z}_z$ is the presence of a U(1) gauge field $A_\mu$: this gauge force emerges near the critical point, even though the underlying model in eqn [16] only has simple two spin interactions. Studies of fractionalized critical theories like $\mathcal{Z}_c$ in other models with spin and/or charge excitations is an exciting avenue for further theoretical research.

*See also:* Bose–Einstein Condensates; Boundary Conformal Field Theory; Fractional Quantum Hall Effect; Ginzburg–Landau Equation; High *Tc* Superconductor Theory; Quantum Central-Limit Theorems; Quantum Spin Systems; Quantum Statistical Mechanics: Overview.

## Further Reading

Matsumoto M, Yasuda C, Todo S, and Takayama H (2002) *Physical Review* B 65: 014407.

Sachdev S (1999) *Quantum Phase Transitions*. Cambridge: Cambridge University Press.

Senthil T, Balents L, Sachdev S, Vishwanath A, and Fisher MPA, http://arxiv.org/abs/cond-mat/0312617.

# Quantum Spin Systems

**B Nachtergaele**, University of California at Davis, Davis, CA, USA



## Introduction

The theory of quantum spin systems is concerned with the properties of quantum systems with an infinite number of degrees of freedom that each have a finite-dimensional state space. Occasionally, one is specifically interested in finite systems. Among the most common examples, one has an *n*-dimensional Hilbert space associated with each site of a *d*-dimensional lattice.

A model is normally defined by describing a Hamiltonian or a family of Hamiltonians, which are self-adjoint operators on the Hilbert space, and one studies their spectrum, the eigenstates, the equilibrium states, the system dynamics, and non-equilibrium stationary states, etc.

More particularly, the term "quantum spin system" often refers to such models where each degree of freedom is thought of as a spin variable, that is, there are three basic observables representing the components of the spin, $S^1$, $S^2$, and $S^3$, and these components transform according to a unitary representation of SU(2). The most commonly encountered situation is where the system consists of $N$ spins, each associated with a fixed irreducible representation of SU(2). One speaks of a spin-$J$ model if this representation is the $(2J+1)$-dimensional one. The possible values of $J$ are $1/2, 1, 3/2, \ldots$

The spins are usually thought of as each being associated with a site in a lattice, or more generally, a vertex in a graph. In a condensed-matter-physics model, each spin may be associated with an ion in a crystalline lattice. Quantum spin systems are also used in quantum information theory and quantum computation, and show up as abstract mathematical objects in representation theory and quantum probability.

In this article we give a brief introduction to the subject, starting with a very short review of its history. The mathematical framework is sketched and the most important definitions are given. Three sections, "Symmetries and symmetry breaking," "Phase transitions," and "Dynamics," together cover the most important aspects of quantum spin systems actively pursued today.

## A Very Brief History

The introduction of quantum spin systems was the result of the marriage of two developments during the 1920s. The first was the realization that angular momentum (hence, also the magnetic moment) is quantized (Pauli 1920, Stern and Gerlach 1922) and that particles such as the electron have an intrinsic

angular momentum called spin (Compton 1921, Goudsmit and Uhlenbeck 1925).

The second development was the attempt in statistical mechanics to explain ferromagnetism and the phase transition associated with it on the basis of a microscopic theory (Lenz and Ising 1925). The fundamental interaction between spins, the so-called exchange operator which is a subtle consequence of the Pauli exclusion principle, was introduced independently by Dirac and Heisenberg in 1926. With this discovery, it was realized that magnetism is a quantum effect and that a fundamental theory of magnetism requires the study of quantum-mechanical models. This realization and a large amount of subsequent work notwithstanding, some of the most fundamental questions, such as a derivation of ferromagnetism from first principles, remain open.

The first and most important quantum spin model is the Heisenberg model, so named after Heisenberg. It has been studied intensely ever since the early 1930s and its study has led to an impressive variety of new ideas in both mathematics and physics. Here, we limit ourselves to listing only some landmark developments.

Spin waves were discovered independently by Bloch and Slater in 1930 and they continue to play an essential role in our understanding of the excitation spectrum of quantum spin Hamiltonians. In two papers published in 1956, Dyson advanced the theory of spin waves by showing how interactions between spin waves can be taken into account.

In 1931, Bethe introduced the famous Bethe ansatz to show how the exact eigenvectors of the spin-1/2 Heisenberg model on the one-dimensional lattice can be found. This exact solution, directly and indirectly, led to many important developments in statistical mechanics, combinatorics, representation theory, quantum field theory and more. Hulthén used the Bethe ansatz to compute the ground-state energy of the antiferromagnetic spin-1/2 Heisenberg chain in 1938.

In their famous 1961 paper, Lieb, Schultz, and Mattis showed that some quantum spin models in one dimension can be solved exactly by mapping them into a problem of free fermions. This paper is still one of the most cited in the field.

Robinson, in 1967, laid the foundation for the mathematical framework, which we describe in the next section. Using this framework, Araki established the absence of phase transitions at positive temperatures in a large class of one-dimensional quantum spin models in 1969.

During the more recent decades, the mathematical and computational techniques used to study quantum spin models have fanned out in many directions.

When it was realized in the 1980s that the magnetic properties of complex materials play an important role in high-$T_c$ superductivity, a variety of quantum spin models studied in the literature proliferated. This motivated a large number of theoretical and experimental studies of materials with exotic properties that are often based on quantum effects that do not have a classical analog. An example of unexpected behavior is the prediction by Haldane of the spin liquid ground state of the spin-1 Heisenberg antiferromagnetic chain in 1983. In the quest for a mathematical proof of this prediction (a quest still ongoing today), Affleck, Kennedy, Lieb, and Tasaki introduced the AKLT model in 1987. They were able to prove that the ground state of this model has all the characteristic properties predicted by Haldane for the Heisenberg chain: a unique ground state with exponential decay of correlations and a spectral gap above the ground state.

There are also particle models that are defined on a lattice, or more generally, a graph. Unlike spins, particles can hop from one site to another. These models are closely related to quantum spin systems and, in some cases, are mathematically equivalent. The best-known example of a model of lattice fermions is the Hubbard model. Such systems are not discussed further in this article.

## Mathematical Framework

Quantum spin systems present an area of mathematical physics where the demands of mathematical rigor can be fully met and, in many cases, this can be done without sacrificing the ability to include all physically relevant models and phenomena. This does not mean, however, that there are few open problems remaining. But it does mean that, in general, these open problems are precisely formulated mathematical questions.

In this section we review the standard mathematical framework for quantum spin systems, in which the topics discussed in the subsequent section can be given a precise mathematical formulation. It is possible, however, to skip this section and read the rest with only a physical or intuitive understanding of the notions of observable, Hamiltonian, dynamics, symmetry, ground state, etc.

The most common mathematical setup is as follows. Let $d \geq 1$, and let $\mathcal{L}$ denote the family of finite subsets of the $d$-dimensional integer lattice $\mathbb{Z}^d$. For simplicity we will assume that the Hilbert space of the "spin" associated with each $x \in \mathbb{Z}^d$ has the same dimension $n \geq 2$: $\mathcal{H}_{\{x\}} \cong \mathbb{C}^n$. The Hilbert space associated with the finite volume $\Lambda \in \mathcal{L}$ is then $\mathcal{H}_\Lambda = \bigotimes_{x \in \Lambda} \mathcal{H}_x$. The algebra of observables for the spin of site $x$ consists of the $n \times n$ complex matrices: $\mathcal{A}_{\{x\}} \cong M_n(\mathbb{C})$. For any

$\Lambda \in \mathcal{L}$, the algebra of observables for the system in $\Lambda$ is given by $\mathcal{A}_\Lambda = \bigotimes_{x\in\Lambda} \mathcal{A}_{\{x\}}$. The primary observables for a quantum spin model are the spin-$S$ matrices $S^1$, $S^2$, and $S^3$, where $S$ is the half-integer such that $n = 2S + 1$. They are defined as Hermitian matrices satisfying the SU(2) commutation relations. Instead of $S^1$ and $S^2$, one often works with the spin-raising and -lowering operators, $S^+$ and $S^-$, defined by the relations $S^1 = (S^+ + S^-)/2$, and $S^2 = (S^+ - S^-)/(2\mathrm{i})$. In terms of these, the SU(2) commutation relations are

$$[S^+, S^-] = 2S^3, \quad [S^3, S^\pm] = \pm S^\pm \qquad [1]$$

where we have used the standard notation for the commutator for two elements $A$ and $B$ in an algebra: $[A, B] = AB - BA$. In the standard basis $S^3$, $S^+$, and $S^-$ are given by the following matrices:

$$S^3 = \begin{pmatrix} S & & & \\ & S-1 & & \\ & & \ddots & \\ & & & -S \end{pmatrix}$$

$S^- = (S^+)^*$, and

$$S^+ = \begin{pmatrix} 0 & c_S & & & \\ & 0 & c_{S-1} & & \\ & & \ddots & \ddots & \\ & & & 0 & c_{-S+1} \\ & & & & 0 \end{pmatrix}$$

where, for $m = -S, -S+1, \ldots, S$,

$$c_m = \sqrt{S(S+1) - m(m-1)}$$

In the case $n = 2$, one often works with the Pauli matrices, $\sigma^1, \sigma^2, \sigma^3$, simply related to the spin matrices by $\sigma^j = 2S^j$, $j = 1, 2, 3$.

Most physical observables are expressed as finite sums and products of the spin matrices $S_x^j$, $j = 1, 2, 3$, associated with the site $x \in \Lambda$:

$$S_x^j = \bigotimes_{y\in\Lambda} A_y$$

with $A_x = S^j$, and $A_y = \mathbb{1}$ if $y \neq x$.

The $\mathcal{A}_\Lambda$ are finite-dimensional C*-algebras for the usual operations of sum, product, and Hermitian conjugation of matrices and with identity $\mathbb{1}_\Lambda$.

If $\Lambda_0 \subset \Lambda_1$, there is a natural embedding of $\mathcal{A}_{\Lambda_0}$ into $\mathcal{A}_{\Lambda_1}$, given by

$$\mathcal{A}_{\Lambda_0} \cong \mathcal{A}_{\Lambda_0} \otimes \mathbb{1}_{\Lambda_1\setminus\Lambda_0} \subset \mathcal{A}_{\Lambda_1}$$

The algebra of local observables is then defined by

$$\mathcal{A}_{\mathrm{loc}} = \bigcup_{\Lambda\in\mathcal{L}} \mathcal{A}_\Lambda$$

Its completion is the $C^*$-algebra of quasilocal observables, which we will simply denote by $\mathcal{A}$.

The dynamics and symmetries of a quantum spin model are described by (groups of) automorphisms of the $C^*$-algebra $\mathcal{A}$, that is, bijective linear transformations $\alpha$ on $\mathcal{A}$ that preserve the product and $^*$ operations. Translation invariance, for example, is expressed by the translation automorphisms $\tau_x$, $x \in \mathbb{Z}^d$, which map any subalgebra $\mathcal{A}_\Lambda$ to $\mathcal{A}_{\Lambda+x}$, in the natural way. They form a representation of the additive group $\mathbb{Z}^d$ on $\mathcal{A}$.

A translation-invariant interaction, or potential, defining a quantum spin model, is a map $\phi : \mathcal{L} \to \mathcal{A}$ with the following properties: for all $X \in \mathcal{L}$, we have $\phi(X) \in \mathcal{A}_X$, $\phi(X) = \phi(X)^*$, and for $x \in \mathbb{Z}^d$, $\phi(X + x) = \tau_x(\phi(X))$. An interaction is called finite range if there exists $R > 0$ such that $\phi(X) = 0$ whenever $\mathrm{diam}(X) > R$. The Hamiltonian in $\Lambda$ is the self-adjoint element of $\mathcal{A}_\Lambda$ defined by

$$H_\Lambda = \sum_{X\subset\Lambda} \phi(X)$$

For the standard Heisenberg model the interaction is given by

$$\phi(\{x, y\}) = -J\mathbf{S}_x \cdot \mathbf{S}_y, \text{ if } |x - y| = 1 \qquad [2]$$

and $\phi(X) = 0$ in all other cases. Here, $\mathbf{S}_x \cdot \mathbf{S}_y$ is the conventional notation for $S_x^1 S_y^1 + S_x^2 S_y^2 + S_x^3 S_y^3$. The magnitude of the coupling constant $J$ sets a natural unit of energy and is irrelevant from the mathematical point of view. Its sign, however, determines whether the model is ferromagnetic ($J > 0$), or antiferromagnetic ($J < 0$). For the classical Heisenberg model, where the role of $\mathbf{S}_x$ is played by a unit vector in $\mathbb{R}^3$, and which can be regarded, after rescaling by a factor $S^{-2}$, as the limit $S \to \infty$ of the quantum Heisenberg model, there is a simple transformation relating the ferro- and antiferromagnetic models (just map $\mathbf{S}_x$ to $-\mathbf{S}_x$ for all $x$ in the even sublattice of $\mathbb{Z}^d$). It is easy to see that there does not exist an automorphism of $\mathcal{A}$ mapping $\mathbf{S}_x$ to $-\mathbf{S}_x$, since that would be inconsistent with the commutation relations [1]. Not only is there no exact mapping between the ferro- and the antiferromagnetic models, their ground states and equilibrium states have radically different properties. See below for the definitions and further discussion.

The dynamics (or time evolution), of the system in finite volume $\Lambda$ is the one-parameter group of automorphisms of $\mathcal{A}_\Lambda$ given by

$$\alpha_t^{(\Lambda)}(A) = \mathrm{e}^{\mathrm{i}tH_\Lambda} A \mathrm{e}^{-\mathrm{i}tH_\Lambda}, \quad t \in \mathbb{R}$$

For each $t \in \mathbb{R}$, $\alpha_t^{(\Lambda)}$ is an automorphism of $\mathcal{A}$ and the family $\{\alpha_t^{(\Lambda)} \mid t \in \mathbb{R}\}$ forms a representation of the additive group $\mathbb{R}$.

Each $\alpha_t^{(\Lambda)}$ can trivially be extended to an automorphism on $\mathcal{A}$, by tensoring with the identity map. Under quite general conditions, $\alpha_t^{(\Lambda)}$ converges strongly as $\Lambda \to \mathbb{Z}^d$ in a suitable sense, that is, for every $A \in \mathcal{A}$, the limit

$$\lim_{\Lambda \uparrow \mathbb{Z}^d} \alpha_t^{(\Lambda)}(A) = \alpha_t(A)$$

exists in the norm in $\mathcal{A}$, and it can be shown that it defines a strongly continuous one-parameter group of automorphisms of $\mathcal{A}$. $\Lambda \uparrow \mathbb{Z}^d$ stands for any sequence of $\Lambda \in \mathcal{L}$ such that $\Lambda$ eventually contains any given element of $\mathcal{L}$. A sufficient condition on the potential $\phi$ is that there exists $\lambda > 0$ such that $\|\Phi\|_\lambda$ is finite, with

$$\|\Phi\|_\lambda = \sum_{X \ni 0} e^{\lambda |X|} \|\phi(X)\| \qquad [3]$$

Here $|\cdot|$ denotes the number of elements in $X$. One can show that, under the same conditions, $\delta$ defined on $\mathcal{A}_{\text{loc}}$ by

$$\delta(A) = \lim_{\Lambda \uparrow \mathbb{Z}^d} [H_\Lambda, A]$$

is a norm-closable (unbounded) derivation on $\mathcal{A}$ and that its closure is, up to a factor i, the generator of $\{\alpha_t \mid t \in \mathbb{R}\}$, that is, formally

$$\alpha_t = e^{it\delta}$$

For the class of $\phi$ with finite $\|\Phi\|_\lambda$ for some $\lambda > 0$, $\mathcal{A}_{\text{loc}}$ is a core of analytic vectors for $\delta$. This means that, for each $A \in \mathcal{A}_{\text{loc}}$, the function $t \mapsto \alpha_t(A)$ can be extended to a function $\alpha_z(A)$ analytic in a strip $|\text{Im } z| < a$ for some $a > 0$.

A state of the quantum spin system is a linear functional on $\mathcal{A}$ such that $\omega(A^*A) \geq 0$, for all $A \in \mathcal{A}$ (positivity), and $\omega(\mathbb{1}) = 1$ (normalization). The restriction of $\omega$ to $\mathcal{A}_\Lambda$, for each $\Lambda \in \mathcal{L}$, is uniquely determined by a density matrix, that is, $\rho_\Lambda \in \mathcal{A}_\Lambda$, such that

$$\omega(A) = \text{tr}\, \rho_\Lambda A, \quad \text{for all } A \in \mathcal{A}_\Lambda$$

where tr denotes the usual trace of matrices. $\rho_\Lambda$ is non-negative definite and of unit trace. If the density matrix is a one-dimensional projection, the state is called a vector state, and can be identified with a vector $\psi \in \mathcal{H}_\Lambda$, such that $\mathbb{C}\psi = \text{ran}\, \rho_\Lambda$.

A ground state of the quantum spin system is a state $\omega$ satisfying the local stability inequalties:

$$\omega(A^*\delta(A)) \geq 0, \quad \text{for all } A \in A_{\text{loc}} \qquad [4]$$

The states describing thermal equilibrium are characterized by the Kubo–Martin–Schwinger (KMS) condition: for any $\beta \geq 0$ (related to absolute temperature by $\beta = 1/(k_B T)$, where $k_B$ is the Boltzmann constant), $\omega$ is called $\beta$-KMS if

$$\omega(A\alpha_{i\beta}(B)) = \omega(BA), \quad \text{for all } A, B \in \mathcal{A}_{\text{loc}} \qquad [5]$$

The most common way to construct ground states and equilibrium states, namely solutions of [4] and [5], respectively, is by taking thermodynamic limits of finite-volume states with suitable boundary conditions. A ground state of the finite-volume Hamiltonian $H_\Lambda$ is a convex combination of vector states that are eigenstates of $H_\Lambda$ belonging to its smallest eigenvalue. The finite-volume equilibrium state at inverse temperature $\beta$ has density matrix $\rho_\beta$ defined by

$$\rho_\beta = Z(\Lambda, \beta)^{-1} e^{-\beta H_\Lambda}$$

where $Z(\Lambda, \beta) = \text{tr } e^{-\beta H_\Lambda}$ is called the partition function. By considering limit points as $\Lambda \to \mathbb{Z}^d$, one can show that a quantum spin model always has at least one ground state and at least one equilibrium state for all $\beta$.

In this section, the basic concepts have so far been discussed in the most standard setup. Clearly, many generalizations are possible: one can consider non-translation-invariant models; models with random potentials; the state spaces at each site may have different dimensions; instead of $\mathbb{Z}^d$ one can consider other lattices or define models on arbitrary graphs; one can allow interactions of infinite range that satisfy weaker conditions than those imposed by the finiteness of the norm [3], or restrict to subspaces of the Hilbert space by imposing symmetries or suitable hardcore conditions; and one can study models with infinite-dimensional spins. Examples of all these types of generalizations have been considered in the literature and have interesting applications.

## Symmetries and Symmetry Breaking

Many interesting properties of quantum spin systems are related to symmetries and symmetry breaking. Symmetries of a quantum spin model are realized as representations of groups, Lie algebras, or quantum (group) algebras on the Hilbert space and/or the observable algebra. The symmetry property of the model is expressed by the fact that the Hamiltonian (or the dynamics) commutes with this representation. We briefly discuss the most common symmetries.

*Translation invariance*. The translation automorphisms $\tau_x$ have already been defined on the

observable algebra of infinite quantum spin systems on $\mathbb{Z}^d$. One can also define translation automorphisms for finite systems with periodic boundary conditions, which are defined on the torus $\mathbb{Z}^d/T\mathbb{Z}^d$, where $T = (T_1, \ldots, T_d)$ is a positive integer vector representing the periods.

*Other graph automorphisms*. In general, if $G$ is a group of automorphisms of the graph $\Gamma$, and $\mathcal{H}_\Gamma = \bigotimes_{x\in\Gamma} \mathbb{C}^n$ is the Hilbert space of a system of identical spins defined on $\Gamma$, then, for each $g \in G$, one can define a unitary $U_g$ on $\mathcal{H}_\Gamma$ by linear extension of $U_g \bigotimes \varphi_x = \bigotimes \varphi_{g^{-1}(x)}$, where $\varphi_x \in \mathbb{C}^n$, for all $x \in \Gamma$. These unitaries form a representation of $G$. With the unitaries one can immediately define automorphisms of the algebra of observables: for $A \in \mathcal{A}_\Lambda$, and $U \in \mathcal{A}_\Lambda$ unitary, $\tau(A) = U^*AU$ defines an automorphism, and if $U_g$ is a group representation, the corresponding $\tau_g$ will be, too. Common examples of graph automorphisms are the lattice symmetries of rotation and reflection. Translation symmetry and other graph automorphisms are often referred to collectively as spatial symmetries.

*Local symmetries* (also called gauge symmetries). Let $G$ be a group and $u_g$, $g \in G$, a unitary representation of $G$ on $\mathbb{C}^n$. Then, $U_g = \bigotimes_{x\in\Lambda} u_g$ is a representation on $\mathcal{H}_\Lambda$. The Heisenberg model [2], for example, commutes with such a representation of SU(2). It is often convenient, and generally equivalent, to work with a representation of the Lie algebra. In that case the SU(2) invariance of the Heisenberg model is expressed by the fact that $H_\Lambda$ commutes with the following three operators:

$$S^i = \sum_{x\in\Lambda} S^i_x, \quad i = 1, 2, 3$$

Note: sometimes the Hamiltonian is only symmetric under certain combinations of spatial and local symmetries. CP symmetry is an example.

For an automorphism $\tau$, we say that a state $\omega$ is $\tau$-invariant if $\omega \circ \tau = \tau$. If $\omega$ is $\tau_g$-invariant for all $g \in G$, we say that $\omega$ is $G$-invariant.

It is easy to see that if a quantum spin model has a symmetry $G$, then the set of all ground states or all $\beta$-KMS states will be $G$-invariant, meaning that if $\omega$ is in the set, then so is $\omega \circ \tau_g$, for all $g \in G$. By a suitable averaging procedure, it is usually easy to establish that the sets of ground states or equilibrium states contain at least one $G$-invariant element.

An interesting situation occurs if the model is $G$-invariant, but there are ground states or KMS states that are not. This means that, for some $g \in G$, and some $\omega$ in the set (of ground states or KMS states), $\omega \circ \tau \neq \omega$. When this happens, one says that there is spontaneous symmetry breaking, a phenomenon that also plays an important role in quantum field theory.

The famous Hohenberg–Mermin–Wagner theorem, applied to quantum spin models, states that, as long as the interactions do not have very long range and the dimension of the lattice is 2 or less, continuous symmetries cannot be spontaneously broken in a $\beta$-KMS state for any finite $\beta$.

*Quantum group symmetries*. We restrict ourselves to one important example: the $\mathrm{SU}_q(2)$ invariance of the spin-1/2 XXZ Heisenberg chain with $q \in [0, 1]$, and with special boundary terms. The Hamiltonian of the $\mathrm{SU}_q(2)$-invariant XXZ chain of length $L$ is given by

$$H_L = \sum_{x=1}^{L-1} -\frac{1}{\Delta}\left(S^1_x S^1_{x+1} + S^2_x S^2_{x+1}\right) - \left(S^3_x S^3_{x+1} - 1/4\right) + \frac{1}{2}\sqrt{1-\Delta^{-2}}\left(S^3_{x+1} - S^3_x\right)$$

where $q \in (0, 1]$ is related to the parameter $\Delta \geq 1$ by the relation $\Delta = (q + q^{-1})/2$. When $q = 0$, $H_L$ is equivalent to the Ising chain. Thus, the XXZ model interpolates between the Ising model (the primordial classical spin system) and the isotropic Heisenberg model (the most widely studied quantum spin model). In the limit of infinite spin ($S \to \infty$), the model converges to the classical Heisenberg model (XXZ or isotropic). An interesting feature of the XXZ model are its non-translation-invariant ground states, called kink states.

In this family of models, one can see how aspects of discreteness (quantized spins) and continuous symmetry (SU(2), or quantum symmetry $\mathrm{SU}_q(2)$) are present at the same time in the quantum Heisenberg models, and the two classical limits ($q \to 0$ and $S \to \infty$) can be used as a starting point to study its properties.

Quantum group symmetry is not a special case of invariance under the action of a group. There is no group, but there is an algebra represented on the Hilbert space of each spin, for which there is a good definition of tensor product of representations, and "many" irreducible representations. In this example, the representation of $\mathrm{SU}_q(2)$ on $\mathcal{H}_{[1,L]}$ commuting with $H_L$ is generated by

$$S^3 = \sum_{x=1}^{L} \mathbb{1}_1 \otimes \cdots \otimes S^3_x \otimes \mathbb{1}_{x+1} \otimes \cdots \mathbb{1}_L$$

$$S^+ = \sum_{x=1}^{L} t_1 \otimes \cdots \otimes t_{x-1} \otimes S^+_x \otimes \mathbb{1}_{x+1} \otimes \cdots \mathbb{1}_L$$

$$S^- = \sum_{x=1}^{L} \mathbb{1}_1 \otimes \cdots \otimes S^-_x \otimes t^{-1}_{x+1} \otimes \cdots t^{-1}_L$$

where

$$t = \begin{pmatrix} q^{-1} & 0 \\ 0 & q \end{pmatrix}$$

Quantum group symmetries were discovered in exactly solvable models, starting with the spin-1/2 XXZ chain. One can exploit their representation theory to study the spectrum of the Hamiltonian in very much the same way as ordinary symmetries. The main restriction to its applicability is that the tensor product structure of the representations is inherently one-dimensional, that is, relying on an ordering from left to right. For the infinite XXZ chain the left-to-right and right-to-left orderings can be combined to generate an infinite-dimensional algebra, the quantum affine algebra $U_q(\widehat{sl_2})$.

## Phase Transitions

Quantum spin models of condensed matter physics often have interesting ground states. Not only are the ground states often a good approximation of the low-temperature behavior of the real systems that are modeled by it, and studying them is therefore useful, it is in many cases also a challenging mathematical problem. This is in contrast with classical lattice models for which the ground states are usually simple and easy to find. In more than one way, ground states of quantum spin systems display behavior similar to equilibrium states of classical spin systems at positive temperature.

The spin-1/2 Heisenberg antiferromagnet on $\Lambda \subset \mathbb{Z}^d$, with Hamiltonian

$$H_\Lambda = \sum_{x,y;\in\Lambda|x-y|=1} S_x \cdot S_y \qquad [6]$$

is a case in point. Even in the one-dimensional case ($d = 1$), and even though the model in that case is exactly solvable by the Bethe ansatz, its ground state is highly nontrivial. Analysis of the Bethe ansatz solution (which is not fully rigorous) shows that spin–spin correlation function decays to zero at infinity, but slower than exponentially (roughly as inverse distance squared). For $d = 2$, it is believed, but not mathematically proved, that the ground state has Néel order, that is, long-range antiferromagnetic order, accompanied by a spontaneous breaking of the SU(2) symmetry. Using reflection positivity, Dyson, Lieb, and Simon were able to prove the Néel order at sufficiently low temperature (large $\beta$), for $d \geq 3$ and all $S \geq 1/2$. This was later extended to the ground state for $d = 2$ and $S \geq 1$, and $d \geq 3$ and $S \geq 1/2$, that is, all the cases where Néel order is expected except $d = 2$, $S = 1/2$.

In contrast, no proof of long-range order in the Heisenberg ferromagnet at low temperature exists. This is rather remarkable since proving long-range order in the ground states of the ferromagnet is a trivial problem.

Of particular interest are the so-called quantum phase transitions. These are phase transitions that occur as a parameter in the Hamiltonian is varied and which are driven by the competing effects of energy and quantum fluctuations, rather than the balance between energy and entropy which drives usual equilibrium phase transitions. Since entropy does not play a role, quantum phase transitions can be oberved at zero temperature, that is, in the ground states.

An important example of a quantum phase transition occurs in the two-or higher-dimensional $XY$ model with a magnetic field in the $Z$-direction. It was proved by Kennedy, Lieb, and Shastry that, at zero field, this model has off-diagonal long-range order (ODLRO), and can be interpreted as a hard-core Bose gas at half-filling. It is also clear that if the magnetic field exceeds a critical value, $h_c$, the model has a simple ferromagnetically ordered ground state. There are indications that there is ODLRO for all $|h| < h_c$. However, so far there is no proof that ODLRO exists for any $h \neq 0$.

What makes the ground-state problem of quantum spin systems interesting and difficult at the same time is that ground states, in general, do not minimize the expectation value of the interaction terms in the Hamiltonian individually although, loosely speaking, the expectation value of their sum (the Hamiltonian) is minimized. However, there are interesting exceptions to this rule. Two examples are the AKLT model and the ferromagnetic XXZ model.

The wide-ranging behavior of quantum spin models has required an equally wide range of mathematical approaches to study them. There is one group of methods, however, that can make a claim of substantial generality: those that start from a representation of the partition function based on the Feynman–Kac formula. Such representations turn a $d$-dimensional quantum spin model into a $(d + 1)$-dimensional classical problem, albeit one with some special features. This technique was pioneered by Ginibre in 1968 and was quickly adopted by a number of authors to solve a variety of problems. Techniques borrowed from classical statistical mechanics have been adapted with great success to study ground states, the low-temperature phase diagram, or the high-temperature regime of quantum spin models that can be regarded as perturbations of a classical system. More recently, it was used to develop a quantum version of Pirogov–Sinai theory which is applicable to a large class of problems, including some with low-temperature phases not related by symmetry.

## Dynamics

Another feature of quantum spin systems that makes them mathematically richer than their classical couterpart is the existence of a Hamiltonian dynamics. Quite generally, the dynamics is well defined in the thermodynamic limit as a strongly continuous one-parameter group of automorphisms of the $C^*$-algebra of quasilocal observables. Strictly speaking, a quantum spin model is actually defined by its dynamics $\alpha_t$, or by its generator $\delta$, and not by the potential $\phi$. Indeed, $\phi$ is not uniquely determined by $\alpha_t$. In particular, it is possible to incorporate various types of boundary conditions into the definition of $\phi$. This approach has proved very useful in obtaining important structural results, such as the proof by Araki of the uniqueness the KMS state at any finite $\beta$ in one dimension. Another example is a characterization of equilibrium states by the energy–entropy balance inequalities, which is both physically appealing and mathematically useful: $\omega$ is a $\beta$-KMS state for a quantum spin model in the setting of the section on the mathematical framework in this article (and in fact also for more general quantum systems), if and only if the inequality

$$\beta\omega(X^*\delta(X)) \geq \omega(X^*X)\log\frac{\omega(X^*X)}{\omega(XX^*)}$$

is satisfied for all $X \in \mathcal{A}_{\text{loc}}$. This characterization and several related results were proved in a series of works by various authors (mainly Roepstorff, Araki, Fannes, Verbeure, and Sewell).

Detailed properties of the dynamics for specific models are generally lacking. One could point to the "immediate nonlocality" of the dynamics as the main difficulty. By this, we mean that, except in trivial cases, most local observables $A \in \mathcal{A}_{\text{loc}}$, become nonlocal after an arbitrarily short time, that is, $\alpha_t(A) \notin \mathcal{A}_{\text{loc}}$, for any $t \neq 0$. This nonlocality is not totally uncontrolled however. A result by Lieb and Robinson establishes that, for models with interactions that are sufficiently short range (e.g., finite range), the nonlocality propagates at a bounded speed. More precisely, under quite general conditions, there exist constants $c, v > 0$ such that, for any two local observables $A, B \in \mathcal{A}_{\{0\}}$,

$$\|[\alpha_t(A), \tau_x(B)]\| \leq 2\|A\|\|B\|\mathrm{e}^{-c(|x|-v|t|)}$$

Attempts to understand the dynamics have generally been aimed at one of the two issues: return to equilibrium from a perturbed state, and convergence to a nonequilibrium steady state in the presence of currents. Some interesting results have been obtained although much remains to be done.

## Acknowledgment

This work was supported in part by the National Science Foundation under Grant # DMS-0303316.

*See also:* Bethe Ansatz; Channels in Quantum Information Theory; Eight Vertex and Hard Hexagon Models; Exact Renormalization Group; Falicov–Kimball Model; Finitely Correlated States; High *Tc* Superconductor Theory; Hubbard Model; Pirogov–Sinai Theory; Quantum Central-Limit Theorems; Quantum Phase Transitions; Quantum Statistical Mechanics: Overview; Reflection Positivity and Phase Transitions; Symmetry and Symmetry Breaking in Dynamical Systems; Symmetry Breaking in Field Theory.

## Further Reading

Affleck I, Kennedy T, Lieb EH, and Tasaki H (1988) Valence bond ground states in isotropic quantum antiferromagnets. *Communications in Mathematical Physics* 115: 477–528.

Aizenman M and Nachtergaele B (1994) Geometric aspects of quantum spin states. *Communications in Mathematical Physics* 164: 17–63.

Araki H (1969) Gibbs states of a one dimensional quantum lattice. *Communications in Mathematical Physics* 14: 120–157.

Borgs C, Kotecký R, and Ueltschi D (1996) Low-temperature phase diagrams for quantum perturbations of classical spin systems. *Communications in Mathematical Physics* 181: 409–446.

Bratteli O and Robinson DW (1981, 1997) *Operator Algebras and Quantum Statistical Mechanics 2. Equilibrium States. Models in Quantum Statistical Mechanics*. Berlin: Springer.

Datta N, Fernández R, and Fröhlich J (1996) Low-temperature phase diagrams of quantum lattice systems. I. Stability for quantum perturbations of classical systems with finitely-many ground states. *Journal of Statistical Physics* 84: 455–534.

Dyson F, Lieb EH, and Simon B (1978) Phase transitions in quantum spin systems with isotropic and non-isotropic interactions. *Journal of Statistical Physics* 18: 335–383.

Fannes M, Nachtergaele B, and Werner RF (1992) Finitely correlated states on quantum spin chains. *Communications in Mathematical Physics* 144: 443–490.

Kennedy T (1985) Long-range order in the anisotropic quantum ferromagnetic Heisenberg model. *Communications in Mathematical Physics* 100: 447–462.

Kennedy T and Nachtergaele B (1996) *The Heisenberg model – a bibliography*. http://math.arizona.edu/~tgk/qs.html.

Kennedy T and Tasaki H (1992) Hidden symmetry breaking and the Haldane phase in $S = 1$ quantum spin chains. *Communications in Mathematical Physics* 147: 431–484.

Lieb E, Schultz T, and Mattis D (1964) Two soluble models of an antiferromagnetic chain. *Annals of Physics (NY)* 16: 407–466.

Matsui T (1990) Uniqueness of the translationally invariant ground state in quantum spin systems. *Communications in Mathematical Physics* 126: 453–467.

Mattis DC (1981, 1988) *The Theory of Magnetism. I.* Berlin: Springer.

Simon B (1993) *The Statistical Mechanics of Lattice Gases. Volume I.* Princeton: Princeton University Press.

# Quantum Statistical Mechanics: Overview

**L Triolo**, Università di Roma "Tor Vergata", Rome, Italy



## Introduction

Quantum theory actually started at the beginning of the twentieth century as a many-body theory, attempting to solve problems to which classical physics gave unsatisfactory answers.

This article aims to follow the developments of quantum statistical mechanics, hereafter called QSM, staying close to the underlying physics and sketching its methods and perspectives. The next section outlines the historical path, and the first achievements by Planck (1900) and Debye (1913); the subsequent free quantum gas theory will be recalled in the first original insights due to Fermi, Dirac (1926) and Bose, Einstein (1924–25), when many open problems began to find a coherent treatment.

In this framework, an interesting new idea appeared: the elementary units of the systems could be "particles", in the usual or in a broad meaning, a notion which includes photons, phonons, and quasiparticles of current use in condensed matter physics. The description of a classical harmonic system through independent normal modes is an example of a very fruitful use of collective variables.

The subsequent section will deal with more recent achievements, related to the properties of quantum $N$-body systems, which are fundamental for the derivation of their macroscopic behavior. In particular, the works by Dyson–Lenard and Lieb–Lebowitz on the stability of matter have to be recalled: a system made of electrons and ions has a thermodynamic behavior, thanks to the quantum nature of its constituents, where the Pauli exclusion principle plays an essential role.

We will then present relations that arise in quantum field theory, that is, from the second quantization methods; related technical and conceptual problems will also be presented briefly.

This is necessary for taking into account the recent works and perspectives, which will be considered in the last section. Here the new inputs and challenges from outstanding achievements in physics laboratories will be taken into account, referring to some exactly solvable models which help in understanding and in fixing the boundaries of approximate methods.

## The Crisis of Classical Physics: The Quantum Free Gas

Let us briefly recall some of what Lord Kelvin called the "nineteenth century clouds" over the physics of that time (1884), and the subsequent new ideas, (Gallavotti 1999).

It is well known that the classical Dulong–Petit law of specific heat of solid crystals may be derived from the model of point particles interacting through harmonic forces; the equipartition of the mean energy among the degrees of freedom implies, for $N$ particles, the linear dependence of the internal energy $U_N$ on absolute temperature $T$, hence a constant heat capacity $C_N$ ($k_B$ is the Boltzmann's constant)

$$U_N = 6 \cdot \frac{1}{2} N k_B T, \qquad C_N := \frac{\partial U_N}{\partial T} = 3 N k_B \qquad [1]$$

Experimentally this is relatively well satisfied at high temperatures but it is violated for low $T$: one observes that $U_N$ vanishes faster than linearly as $T$ goes to zero, so that $C_N$ vanishes. Moreover, the contributions to the heat capacity from the internal degrees of freedom of the molecular gases or from the free electrons in conducting solids are negligible, at room temperature: these degrees of freedom, in spite of the equipartition principle, seem frozen.

The analysis of the blackbody radiation problem from the classical point of view, that is, using equipartition among the normal modes of the electromagnetic field in the "black" cavity at temperature $T$, gives the following dependence, Rayleigh–Jeans law (1900), of the spectral energy density $u(\nu, T)$, on frequency $\nu$ and temperature $T$ ($c$ is the speed of light in vacuum):

$$u(\nu, T) = \frac{8\pi\nu^2}{c^3} k_B T \qquad [2]$$

The experimental curves for any positive $T$ show a maximum for a frequency $\nu_{\max}(T)$ which increases linearly with $T$ according to Wien's displacement law (1893). The spectral energy density decreases fast enough to zero as $\nu \to \infty$ in such a way that the overall (integrated) energy is (finite and) proportional to $T^4$, according to Stefan's law (1879); the agreement with the classical form holds for low frequencies. The analytic form of the classical $u(\nu, T)$ in [2] does not present maxima and the overall radiated energy is clearly divergent (this bad behavior for large $\nu$, present in many formulas for other models, sometimes in the corresponding "short-distance" form, is called an "ultraviolet catastrophe").

The effort by M Planck (1900) to understand the right dependence of $u$ from $\nu$ and $T$ was based on a thermodynamic argument about the possible energy–entropy relation, and on an assumption similar to the discretization rules on which the "old quantum theory" for the atomic structure is based. The electromagnetic field is represented, via Fourier analysis, as a set of infinitely many independent harmonic oscillators, two for every wave vector $\boldsymbol{k}$, to take into account the polarization. The frequency depends linearly on the wave number $k = |\boldsymbol{k}|$ (linear dispersion law), and the spacing becomes negligible for macroscopic dimensions of the cavity. The key idea for computing the partition function is the discretization of the phase space of each oscillator (of frequency $\nu = \omega/2\pi$). Putting there the adimensionalized Lebesgue measure $\mathrm{d}p\,\mathrm{d}q/h$, where $h$ is a constant with physical dimensions of an action, we consider the regions $R_E$ bounded by the constant-energy ellipses and their areas $|R_E|$, and find

$$|R_E| = \int_{R_E} \frac{\mathrm{d}p\mathrm{d}q}{h} = \frac{2\pi E}{h\omega} = \frac{E}{h\nu}$$

If these adimensional areas have integer values, that is, $E = nh\nu, n = 0, 1, 2, \ldots$, the annular region ("cell" $C_n$) between $R_{E_n}$ and $R_{E_{n+1}}$ has unit area and so we approximate the partition function with the series ($\beta = 1/(k_\mathrm{B}T)$, the ubiquitous parameter in statistical mechanics, often called "inverse temperature")

$$Z_{\mathrm{discr}} = \sum_n \exp(-\beta nh\nu) = \frac{1}{1 - \exp(-\beta h\nu)}$$

In this way, the probabilistic weight given to this cell is

$$\begin{aligned} p(C_n) &= \frac{\exp(-\beta nh\nu)}{\sum_j \exp(-\beta jh\nu)} \\ &= \exp(-\beta nh\nu)(1 - \exp(-\beta h\nu)) \end{aligned} \qquad [3]$$

A well-defined value for the constant $h$ (i.e., $h = 6.626\ldots \times 10^{-27}$ erg s, the Planck constant), combined with the usual computation for the density of states, gives a formula which quantitatively agrees with experimental data (see Figure 1)

$$u(\nu, T) = \frac{8\pi\nu^2}{c^3}\frac{h\nu}{\exp(h\nu/k_\mathrm{B}T) - 1} \qquad [4]$$

Moreover, for a certain range of parameters, that is, such that $\beta h\nu \ll 1$, there is agreement with the classical law.

The "quantum of light," introduced by Einstein in 1905 in his work on photoelectric effect, was later (1926) called photon by G N Lewis. The picture for representing the radiating system was one of a gas of noninteracting photons, carrying energy and momentum, and being continuously created and absorbed.

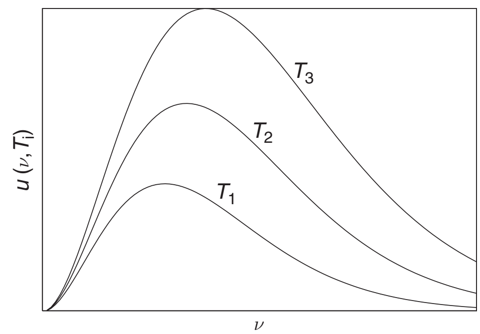



**Figure 1** Dependence of the electromagnetic energy density on $\nu$, for $T_1 < T_2 < T_3$.

A slightly different approach was used about the same time, for the problem of specific heat of crystalline solids.

The simpler model considers $N$ points on the nodes of the lattice $\mathbb{Z}^3$, in a cubic box of side $L$, and interacting through harmonic forces; similarly to the radiation problem, the system is represented by a collection of independent harmonic oscillators (normal modes), which are "quantized" as before: the corresponding quanta were called phonons (by Fraenkel, in 1932) for the role of the acoustic band of frequencies. In this simplified approach (by Debye, in 1913) the different phonons are determined by a finite set of wave vectors

$$\boldsymbol{k} = \frac{2\pi}{L}\boldsymbol{n}, \quad n_i \text{ integer}, \ i = 1, 2, 3; \ |\boldsymbol{k}| \le k_\mathrm{M}$$

where the maximal modulus $k_\mathrm{M}$ is such that the total number of different $\boldsymbol{k}$'s is $3N$ (degrees of freedom).

Moreover, the frequency–wave number relation is simplified too, extrapolating the low-frequency (acoustic) linear relation $\nu = |\boldsymbol{k}|v_0$ ($v_0$ is the sound speed). In this way, the density of states which is quadratic in the frequency, has a cutoff to zero at the maximal frequency, $\nu_\mathrm{D}$, corresponding to $|k_\mathrm{M}|$, with an associate temperature $\Theta_\mathrm{D} = h\nu_\mathrm{D}/k_\mathrm{B}$ (Debye's temperature). The expected energy $U_N$ in the canonical ensemble, after the computation of the canonical partition function, is given in term of the Debye function $D(\cdot)$:

$$\begin{aligned} U_N &= 3Nk_\mathrm{B}TD\left(\frac{\Theta}{T}\right) \\ D(y) &:= \frac{3}{y^3}\int_0^y \frac{x^3\mathrm{d}x}{\exp x - 1} \end{aligned} \qquad [5]$$

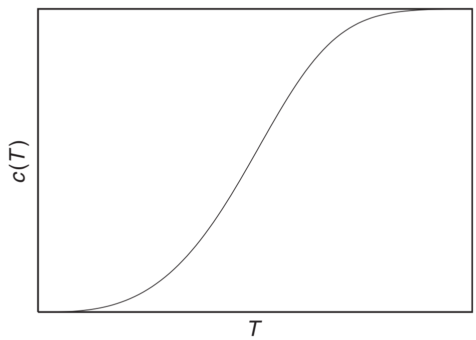



**Figure 2** The specific heat of crystal solids according to Debye.

The agreement with experimental data, for the specific heat of different materials (i.e., different $\Theta_D$), at low and high temperatures, is rather good. At low temperatures, one recovers the empirical $T^3$ behavior (see Figure 2). More careful measurements at low $T$ put into evidence, for metallic solids, the role of the conduction electrons: their contribution to the heat capacity turns out to be linear in $T$, with a coefficient such that at room temperature it is much smaller than the lattice contribution, so that a satisfactory agreement with the classical law is found.

Soon after the beginning of quantum mechanics in its modern form (1925–26), physicists considered many-particle systems, dealing initially with the simplest situations, with a relatively easy formal apparatus, yet sufficient enough to understand in the main lines the "anomalous," that is, nonclassical, behavior.

For a system of $N$ free particles in a cubic box of side $L$, quantum theory brings the labeling of the one-particle states with the wave vectors $\boldsymbol{k}$, recalling the de Broglie relation for the momentum $\boldsymbol{p} = \hbar \boldsymbol{k}$, with a possible additional spin (intrinsic angular moment) label $\sigma$

$$\boldsymbol{k} = \frac{2\pi}{L}\boldsymbol{n}, \quad \boldsymbol{n} \in \mathbb{Z}^3 \backslash \{0\}$$

and the statistics of the particles: because of indistinguishability, the wave function of several identical particles has to be symmetric (B–E, Bose–Einstein statistics) or antisymmetric (F–D, Fermi–Dirac statistics) in the exchange of the particles. This has the deep implication that no more than one fermion shares the same quantum state.

We may here recall the spin–statistics connection, which, in the framework of a local relativistic theory, states that integer spin particles are bosons, while particles with half-odd-integer spin are fermions.

As the state is completely defined by the knowledge of occupation numbers $n_{\boldsymbol{k},\sigma}$, we have the simple and relevant statement on the ground states for the $N$ spinless bosons and $N$ spin-1/2 fermion systems are described by the statement:

$$\begin{aligned} &\text{B–E system}: \quad && n_{\boldsymbol{k}} = N\delta_{\boldsymbol{k},\boldsymbol{k}_0} \\ &\text{F–D system}: \quad && n_{\boldsymbol{k},\sigma} = \mathbf{1}(|\boldsymbol{k}| \le k_F)\forall\sigma \end{aligned} \qquad [6]$$

The constant $k_F$ (Fermi wave number), or the equivalent $p_F = \hbar k_F$ and $\varepsilon_F = {p_F}^2/2m$ (Fermi momentum and energy, respectively) denotes the higher occupied level. In the continuum approximation, this implies the following relation between Fermi energy $\varepsilon_F$ and density $\rho = N/L^3$:

$$\varepsilon_F = \frac{\hbar^2}{2m}(3\pi^2\rho)^{2/3} \qquad [7]$$

Going to the positive-temperature case, the grand canonical partition function is computed by considering that occupation numbers are non-negative integers for the B–E case and just 0 or 1 for the F–D case. This implies the simple formulas, with obvious meaning of symbols and leaving more details to the vast literature (see Figure 3):

$$\langle n_{\boldsymbol{k},\sigma}\rangle_{\beta,\mu} = \frac{1}{\exp(\beta(\varepsilon_{\boldsymbol{k}} - \mu)) \pm 1}, \quad + \text{for F–D}, \ - \text{for B–E} \qquad [8]$$

It is useful to introduce the Fermi temperature $T_F = \varepsilon_F / k_B$; using some realistic data, that is, for common metals like copper, $T_F$ ranges roughly between $10^4$–$10^5$ K, that is, well above the "normal," room temperatures: the quantum nature (i.e., quantum degeneracy) of the conduction electrons, modeled as free electrons, is macroscopically visible in normal conditions.

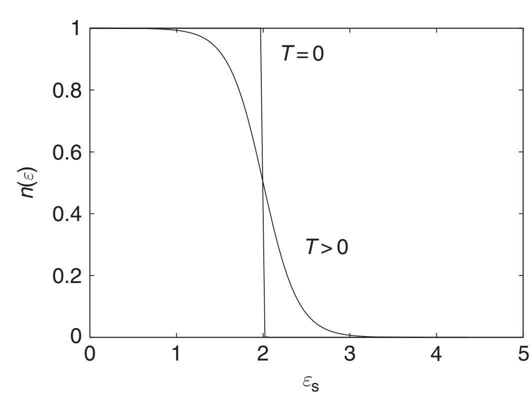



**Figure 3** Expected fermionic occupation number, for $T = 0$, $T > 0$, and $\mu = 2$.

The presence of an external field, like the periodic one given by the ionic lattice of a crystal, changes the situation in a relevant way, as the one-particle spectrum generally gets a band structure, and the allowed momenta are described in the reciprocal lattice: the Fermi sphere becomes a surface, and its structure is central for further developments.

For massive bosons, the strange superfluid features of liquid $^4$He at low temperature, that is, below the critical value 2.17 $K$, led F London, just after Kapitza's discovery in 1937, to speculate that these were related to a macroscopic occupation of the ground state (B–E condensation). A more realistic model has to take into account interaction between bosons (see last section) as the microscopic interactions in superfluid liquid $^4$He are not negligible.

## Quantum *N*-Body Properties: Second Quantization

The main step in analyzing a quantum $N$-body system is its energy spectrum, and in particular its ground state, as it may represent a good approximation of the low-temperature states: its structure, the relations with possible symmetries of the Hamiltonian, its degeneracy, the dependence of its energy on the number of particles, are further relevant questions. The last one is related to the possibility of defining a thermodynamics for the system (Ruelle 1969). As a physically very interesting example, consider a system of electrically charged particles, $N$ electrons with negative unit charge, and $K$ atoms with positive charge $z$, say, interacting through electrostatic forces; the classical Coulomb potential as a function of distance behaves badly, as it diverges at zero and decreases slowly at infinity. The first question is about the stability: thanks to the exclusion principle, for the ground-state energy $E^0_{N,K}$ an extensive estimate from below is valid:

$$E^0_{N,K} \geq -c_0(N + Kz)$$

so that a finite-volume grand partition function exists, while for the thermodynamic limit, which involves large distances, we need more, that is, charge neutrality, which allows for screening, and a fast-decreasing effective interaction.

Let us see an example (quantum spin, Heisenberg model) belonging to the class of lattice models, where the identical microscopic elements are distinguishable by their fixed positions, that is, the nodes of a lattice like $\mathbb{Z}^d$. To any site $x \in \mathbb{Z}^d$ is associated a copy $\mathcal{H}_x$ of a $(2s+1)$-dimensional Hilbert space $\mathcal{H}$, where an irreducible unitary representation of SU(2) is given, so that the nonzero values for $s$ are $1/2, 1, 3/2, \ldots$. For any $x$, the generators $S_\alpha(x), (\alpha = 1, 2, 3)$ satisfy the well-known commutation relations of the angular momentum; moreover, $\sum_\alpha {S_\alpha}^2(x) = s(s+1)\mathbf{1}$, and operators related to different sites commute. The ferromagnetic, isotropic, next-neighbors, magnetic field Hamiltonian for the finite system is

$$H_\Lambda = -J \sum_{<x,y>} \boldsymbol{S}(x) \cdot \boldsymbol{S}(y) - h \sum_x S_3(x) \qquad [9]$$

where $J$ is the positive strength of the next-neighbors coupling ($<x, y>$ means that $x$ and $y$ are next neighbors); $h$ is the intensity of the magnetic field oriented along the third axis. This model is considerably studied even now with several variants regarding possible anisotropies of the interaction, the possibly infinite range of the interaction, and the sign of $J$, for other (e.g., antiferromagnetic) couplings. Among the relevant results, the Mermin–Wagner theorem, at variance with the analogous classical spin model, states the absence of spontaneous magnetization in this zero-field model for $d = 2$ for any positive temperature; this can also be formulated as absence of symmetry breaking for this model (Fröhlich and Pfister in 1981 shed more light on this point).

As mentioned earlier, a useful mathematical tool for dealing with quantum systems of many particles or quasiparticles, is the occupation-number representation for the state of the system. The vector space for a system with an indefinite number of particles is the Fock space: it is the direct sum of all spaces with any number of particles, starting with the zero-particle, vacuum state. The operators which connect these subspaces are the creation and annihilation operators, very similar to the raising and lowering operators introduced by Dirac for the spectral analysis of the harmonic-oscillator Hamiltonian and the angular momentum, in the context of one-particle quantum theory.

It is perhaps worth sketching the action of these operators on the Fock space.

We consider spinless bosons first, as spin might easily be taken into account, if necessary. We suppose that a one-particle Hamiltonian has eigenfunctions labeled by a set of quantum numbers $\boldsymbol{k}$, say, as the wave vector for the purely kinetic one-particle Hamiltonian. Let $|n_{k_1}, n_{k_2}, \ldots, n_{k_p} >$ denote a vector state with $\sum_{i=1,\ldots,p} n_{k_i}$ particles, where $n_{k_i}$ denotes the number of particles with wave vector $\boldsymbol{k}_i, i = 1, \ldots, p; |0>$ denotes the no-particle, vacuum state. We define the creation operators $a^*_{\boldsymbol{k}}$ as follows:

$$a^*_{\boldsymbol{k}} | \ldots n_{\boldsymbol{k}}, \ldots \rangle = \sqrt{n_{\boldsymbol{k}} + 1} | \ldots, n_{\boldsymbol{k}} + 1, \ldots \rangle \qquad [10]$$

Its adjoint $a_k$ is called the annihilation operator, for its action on the vectors

$$a_k|\ldots n_k,\ldots\rangle = \sqrt{n_k}|\ldots,n_k-1,\ldots\rangle \quad [11]$$

The operator $a_k^*$ creates a new particle with that momentum: for any $\boldsymbol{k}$

$$a_k^* a_k|\ldots n_k,\ldots\rangle = n_k|\ldots n_k,\ldots\rangle$$

$$a_k^* a_k := \hat{n}_k \text{ (the occupation-number operator)}$$

The vacuum state belongs to $\mathrm{Ker}\, a_k$ for any $\boldsymbol{k}$, and the whole space is generated by application of creation operators on the vacuum state.

The following basic commutation relations, for any $\boldsymbol{k}, \boldsymbol{k}'$, are valid:

$$[a_k, a_{k'}^*] = \delta(\boldsymbol{k}, \boldsymbol{k}'), \quad [a_k, a_{k'}] = [a_k^*, a_{k'}^*] = 0 \quad [12]$$

For fermions, multiple occupancy is forbidden, so that the analogous annihilation ($\alpha_k$) and creation ($\alpha_k^*$) operators satisfy anticommutation relations:

$$[\alpha_k, \alpha_{k'}^*]_+ = \delta(\boldsymbol{k}, \boldsymbol{k}'), \quad [\alpha_k, \alpha_{k'}]_+ = [\alpha_k^*, \alpha_{k'}^*]_+ = 0 \quad [13]$$

The presence of spin is dealt by an additional spin label $\sigma$ to these symbols, and a $\delta(\sigma, \sigma')$, where necessary.

The Hamiltonian for a system of particles, say spinless bosons, in a box $\Lambda$, made of its kinetic part together with a two-body ($\lambda v(x-y)$) interaction, is written in terms of the "field operators"; if $\{\phi_k(x)\}$ are the one-particle eigenfunctions of the single-particle purely kinetic Hamiltonian for the spinless case, and their complex conjugates are $\{\phi_k^*(x)\}$, we define the fields

$$\Phi(x) = \sum_k \phi_k(x) a_k, \quad \Phi^*(x) = \sum_k \phi_k^*(x) a_k^* \quad [14]$$

So that the full Hamiltonian is given by

$$H_\Lambda = \int_\Lambda \mathrm{d}x\, \Phi^*(x) \left(\frac{-\hbar^2}{2m}\right) \Delta\Phi(x) + \lambda \int_\Lambda \mathrm{d}x \int_\Lambda \mathrm{d}y\, v(x-y) \Phi^*(x)\Phi(x)\Phi^*(y)\Phi(y) \quad [15]$$

We mention that a theoretical breakthrough in the analysis of superfluidity was made by Bogoliubov (1946), who, starting from the Hamiltonian in [15], introduced the following Hamiltonian in the momentum representation:

$$H_\Lambda = \sum_k \varepsilon_k a_k^* a_k + \frac{1}{2} \sum_{k,k',q} \hat{v}_q a_{k-q}^* a_{k'+q}^* a_k a_{k'} \quad [16]$$

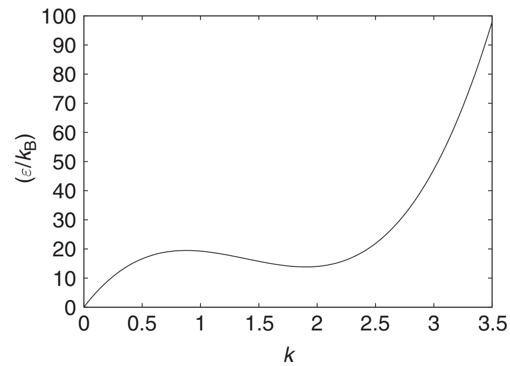



**Figure 4** Excitation spectrum for superfluids.

where $\varepsilon_k$ is the one-particle kinetic energy and $\hat{v}_q$ is the Fourier transform of the two-body potential. To study the excitation spectrum above the ground state, he introduced an approximation about the persistency of a macroscopic occupation of the ground state and a diagonalization procedure leading to new quasiparticles with a characteristic energy spectrum, linearly increasing near $|\boldsymbol{k}| = 0$, then presenting a positive minimum before the subsequent increase (see Figure 4).

## Some Mathematical Tools for Macroscopic Quantum Systems

The formal apparatus of second quantization, born in the context of the quantum field theory, brought to statistical mechanics new ideas and techniques and related difficulties. For instance, the renormalization group was conceived in the 1970s to deal both with critical phenomena (i.e., power singularities of thermodynamic quantities around the critical point) and with divergences in quantum field theory. This subject is currently being developed and applied in models of quantum statistical mechanics (QSM) (Benfatto and Gallavotti, 1995).

Another issue, which has again strong relations with quantum field theory, is the algebraic formulation of QSM. This point of view, which is well suited for the analysis of infinitely extended quantum systems, uses a unified, synthetic, and rigorous language. The procedure for passing from a finite quantum system to its infinitely extended version deserves some attention.

It is well known that, for finite quantum systems, say $N$ particles in a box $\Lambda$, an observable is represented by a self-adjoint operator $A$ on a Hilbert space $\mathcal{H}_\Lambda$, and the normalized elements $\{|\psi\rangle\}$ of this space are the pure states $\rho_\psi$ which define the expectations

$$\rho_\psi(A) := \langle\psi|A\psi\rangle$$

The mixed states (mixtures) are defined by convex combinations of pure states, the coefficients having an obvious statistical meaning.

Among the observables, the Hamiltonian plays a special role, as it generates the dynamics of the system, which evolves the pure states through the unitary group (Schrödinger picture)

$$\left|\psi(t) \geq \exp\left(-\frac{\mathrm{i}tH_\Lambda}{\hbar}\right)\right|\psi>$$

To the notion of equilibrium probability measure on the phase space of a classical system, corresponds the mixed state $\rho_{H_\Lambda,\beta}$ such that

$$\rho_{H_\Lambda,\beta}(A) := Z_{\Lambda,\beta}{}^{-1}\,\mathrm{tr}(\exp(-\beta H_\Lambda)A) \qquad [17]$$

The normalization factor $Z_{\Lambda,\beta} = \mathrm{tr}(\exp(-H_{\Lambda\beta}))$ is the canonical partition function.

Consider now the algebra $\mathcal{A}(\Lambda)$ of local observables; sending $\Lambda$ to infinity, by induction, it is possible to define the algebra $\mathcal{A}$ of quasilocal observables. The main point is a set of algebraic relations like the canonical commutation relations (CCRs) and the canonical anticommutation relations (CARs) for the creation/annihilation operators: the observables of $\mathcal{A}$, through the GNS (Gel'fand, Naimark, and Segal) construction may be represented as operators on the appropriate Hilbert spaces, depending on the chosen state; the representations, at variance with the finite case, might be inequivalent. It is possible to define the equilibrium state for the infinite system and how to insert in a natural way the possible group invariance of the system ($\mathbb{R}^d$ or $\mathbb{Z}^d$, typically), ending with characterization of the pure phases of the system as the ergodic components in the decomposition of an equilibrium state. These states have the property that coarse-grained observables have sharp values (Ruelle 1969, Sewell 2002): if $\mathrm{Av}_l(A)$ is the space average on scale $l$, that is, over boxes of side $l$, for an ergodic state $\rho$,

$$\lim_{l\to\infty} \rho([\mathrm{Av}_l(A) - \rho(A)]^2) = 0$$

Another issue which is worth mentioning is the characterization of equilibrium states through the KMS (Kubo–Martin–Schwinger) condition. The strong formal similarity between the finite-volume quantum evolution operator $\alpha_t := \exp(-\mathrm{i}tH_\Lambda/\hbar)$ and the statistical equilibrium density operator $\exp(-\beta H_\Lambda)$, leads to the identity, valid for any couple of bounded observables $A$ and $B$, using the short symbol $< \cdot >_{\beta,\Lambda}$ for the expectations with respect to the statistical operator:

$$<A_t B>_{\beta,\Lambda} = <BA_{t+\mathrm{i}\hbar\beta}>_{\beta,\Lambda} \qquad [18]$$

This relation is suitably extended for infinite size, and therefore defines a KMS state; it implies some physically relevant properties like stability with respect to local disturbances and dissipativity (Sewell 2002).

A final issue in this section concerns another formalism stemming from the Feynman path-integral formulation of quantum mechanics: here a functional integral represents the statistical equilibrium density operator $W_\beta = \exp(-\beta H)$. For a $d$-dimensional system of $N$ particles in a potential field ($X \in \mathbb{R}^{dN}$) $V = V(X) = \sum_{i<j}\phi(x_i - x_j)$ and Hamiltonian $H = -(1/2)\,\Delta + V$ the Feynman–Kac formula which, for a test function $\chi$, may be written as follows:

$$(W_\beta\chi)(X) = \int P^\beta_{X,Y}(\mathrm{d}\omega)\exp\left(-\int_0^\beta \mathrm{d}s V(\omega(s))\chi(Y)\mathrm{d}Y\right)$$

where $P^\beta_{X,Y}(\mathrm{d}\omega)$ is the Wiener measure on the space of paths $\{\omega(s), s \in [0,\beta]\}$. For details on the construction and several other related features on the treatment of the different statistics, see Glimm and Jaffe (1981).

## New Problems and Challenges

In this final section, we recall some phenomena which have been observed recently in physics laboratories, and which presumably deserve considerable efforts to overcome the heuristic level of explanation. About this last point, it is worth quoting a method that has been used to get results even without clear justifications of the underlying hypotheses, that is, the mean-field procedure. It started with the Curie–Weiss theory of magnetism and is based on the following drastic simplification: the microscopic element of the system feels an average interaction field due to other elements, indipendently of the positions of the latter. This method might provide relatively good results if the range of the interaction is very large, and in fact, a clear version with due limiting procedure was introduced by Kac, and applied by Lebowitz and Penrose in the 1960s for a microscopic derivation of van der Waals equation, and soon extended by Lieb to quantum systems.

We will briefly outline some aspects of three recent achievements of condensed matter physics for which modeling is still on the way of further progress: the B–E condensation, the high-$T_c$ superconductivity, and the fractional quantum Hall effect. The first consists in trapping an ultracold (at less than 50 μK) dilute bosonic gas, for example, $10^4$–$10^7$ atoms of $^{87}$Rb, finding experimental evidence for Bose condensation. To understand the

properties of this system, an important tool is the Gross–Pitaevskii energy functional for the condensate wave function $\Phi$,

$$E[\Phi] = \int \mathrm{d}x \left[ \frac{\hbar^2}{2m} |\nabla\Phi|^2 + V_{\mathrm{ext}}(x)|\Phi|^2 + \frac{g}{2}|\Phi|^4 \right]$$

where the quartic term represents the reduced (mean-field) interaction among particles.

The second issue, that is, the high-temperature superconductivity, certainly deserves much attention. It has been observed recently in some ceramic materials well above 100 K, and a clear model which takes into account the formation of pairs and the peculiar isotropy–anisotropy aspects of the normal conductivity and superconductivity is still lacking (Mattis 2003).

Finally, let us consider the fractional quantum Hall effect; recall that the integer version, that is, a discretization of the Hall resistivity $R_{\mathrm{H}}$ by multiples of $h/(e^2)$, finds an explanation in terms of band spectra, formation of magnetic Landau levels, and localization from surface impurities, that is, without taking into account direct interactions among electrons.

The fractional discretization of $R_{\mathrm{H}}$ (Störmer 1999) has a theoretical interpretation, in terms of subtle collective behavior of the two-dimensional semiconductor electron system: the quasiparticles which represent the excitations may behave as composite fermions or bosons, or exhibit a fractional statistics (*see* Fractional Quantum Hall Effect).

This brief excursion through these new fascinating phenomena shows the rich interplay between theory and experiments: these phenomena are a source of new ideas and suggest new models for further progress.

*See also:* Bose–Einstein Condensates; Dynamical Systems and Thermodynamics; Exact Renormalization Group; Falicov–Kimball Model; Fermionic Systems; Finitely Correlated States; Fractional Quantum Hall Effect; High $T_c$ Superconductor Theory; Hubbard Model; Quantum Phase Transitions; Quantum Spin Systems; Stability of Matter.

## Further Reading

Benfatto G and Gallavotti G (1995) *Renormalization Group*. Princeton: Princeton University Press.

Gallavotti G (1999) *Statistical Mechanics: A Short Treatise*. Berlin: Springer.

Glimm J and Jaffe A (1981) *Quantum Physics. A Functional Integral Point of View*. New York: Springer.

Landau LD and Lifschitz EN (2000) Statistical Physics, *Course of Theoretical Physics*, 3rd edn., vol 5. Parts I and II. Oxford: Butterworth-Heinemann.

Mattis DC (2003) *Statistical Mechanics Made Simple*. NJ: World Scientific.

Ruelle D (1969) *Statistical Mechanics. Rigorous Results*. New York: Benjamin.

Sewell GL (2002) *Quantum Mechanics and Its Emergent Macrophysics*. Princeton: Princeton University Press.

Sinai YaG (1982) *Theory of Phase Transitions: Rigorous Results*. Budapest: Akadémiai Kiadó.

Störmer HL (1999) Nobel lecture: the quantum Hall effect. *Reviews of Modern Physics* 71: 875–889.

Taylor PL and Heinonen O (2002) *A Quantum Approach to Condensed Matter Physics*. Cambridge: Cambridge University Press.

# Quasiperiodic Systems

**P Kramer**, Universität Tübingen, Tübingen, Germany



## Introduction: From Periodic to Quasiperiodic Systems

Periodic systems occur in many branches of physics. Their mathematical analysis was stimulated in particular by the analysis of the periodic translational symmetry of crystals. The systematic study of the compatibility between translational and crystallographic point or reflection symmetry leads to the concept of space group symmetry. Mathematical crystallography in three dimensions (3D) culminated in 1892 in the complete classification of the 230 space groups due to Fedorov and Schoenflies (see Schwarzenberger (1980, pp. 132–135). One characteristic property of periodic systems is that their Fourier transform has a pure point spectrum. Since the Fourier spectrum is experimentally accessible through diffraction experiments, it provides a main tool for the structure determination of crystals.

With quantum mechanics in the twentieth century, it became possible to describe crystal structures quantitatively as ordered systems of atomic nuclei and electrons with electromagnetic interactions. The representation theory of crystallographic space groups now opened the way to verify the

space group symmetry of atomic systems for example from the band structure of crystals. It was then believed that in physics atomic long-range order is linked to periodicity and hence to the paradigm of the 230 space groups in 3D.

Mathematical analysis beyond this paradigm started independently in various directions. Bohr (1925) studied quasiperiodic functions and their Fourier transform. He interpreted them as restrictions of periodic functions in $n$D to their values on a linear subspace of orientation irrational with respect to a lattice. Mathematical crystallography in general dimension $n > 3$, including point group symmetry, was started around 1949 in work by Hermann and by Zassenhaus (see Schwarzenberger (1980)), and completed in 1978 for $n = 4$ in Brown *et al.* (1978). A different route was taken by Penrose (1974). He constructed an aperiodic tiling (covering without gaps or overlaps) of the plane. Its tiles in two rhombus shapes provide global 5-fold point symmetry and make the tiling incompatible with any periodic lattice in 2D. The connection between Penrose's aperiodic tiling and irrational subspaces in periodic structures was made by de Bruijn (1981). He interpreted the Penrose rhombus tiling as the intersection of geometric objects from cells of a hypercubic lattice in 5D with a 2D subspace, irrational and invariant under 5-fold noncrystallographic point symmetry. Kramer and Neri (1984) embedded the icosahedral group as a point group into the hypercubic lattice in 6D and constructed a 3D irrational subspace invariant under the noncrystallographic icosahedral point group. From intersections of boundaries of the hypercubic lattice cells with this subspace, they constructed a 3D tiling of global icosahedral point symmetry with two rhombohedral tiles.

Shechtman *et al.* (1984) discovered in the system *AlMn* diffraction patterns of icosahedral point symmetry. Since icosahedral symmetry is incompatible with a lattice in 3D, they concluded that there exists atomic long-range order without a lattice. The new paradigm of quasiperiodic long-range order in quasicrystals was established and since then stimulated a broad range of theoretical and experimental research.

The interplay between the notions – (1) of crystallographic symmetry in $n$D, $n > 3$, (2) of subspaces invariant under a point group but irrational and hence incompatible with a lattice, and (3) of discrete geometric periodic objects in $n$D providing quasiperiodic tilings on these subspaces – forms the mathematical basis for a new quasiperiodic long-range order found in quasicrystals. The present-day theory of quasicrystals offers the most elaborate study of quasiperiodic systems. Therefore, we shall focus in what follows on the concepts developed in this theory.

In the following section, we briefly review basic concepts of periodic systems and lattices in $n$D, their classification in terms of point symmetry and space groups, and their cell structure. In a section on quasiperiodic point sets and functions, a quasiperiodic system is taken as a geometric object on an irrational $m$D subspace in an $n$-dimensional space and lattice. Noncrystallographic point symmetry is shown to select the irrational subspace. Next, scaling symmetry in quasiperiodic systems is demonstrated. Then, examples of quasiperiodic systems with point and scaling symmetry are given. The penultimate section discusses quasiperiodic tilings and their windows. Finally, the notion of a fundamental domain for quasiperiodic functions compatible with a tiling is illustrated.

## Concepts from Periodic Systems

A distribution $f^p(x)$ of geometric objects on Euclidean space $E^n$ (a real linear space equipped with standard Euclidean scalar product $\langle\,,\rangle$ and metric) with coordinates $x \in E^n$ is called "periodic" if it is invariant under translations $b^i$ in $n$ linearly independent directions,

$$(p) : f^p : f^p(x + b^i) = f^p(x), \quad i = 1, \ldots, n \qquad [1]$$

The set of all translations on $E^n$ forms the discrete additive abelian translation group

$$T = \left\{ b \in E^n : b = \sum_i^n m_i b^i, (m_1, \ldots, m_n) \in \mathbf{Z}^n \right\} \qquad [2]$$

Any orbit (set of all images of an initial point) under the action $T \times E^n \to E^n$ yields a lattice $\Lambda$ on $E^n$. Since $T$ acts fixpoint-free, there is a one-to-one correspondence $\Lambda \leftrightarrow T$. A fundamental domain on $E^n$ is defined as a subset of points $x \in E^n$ which contains a representative point from any orbit under $T$. Such a fundamental domain can be chosen, for example, as the unit cell of the lattice $\Lambda$ or as the Voronoi cell (eqn [5]). By eqn [1], the functional values on $E^n$ of a periodic function $f^p(x)$ are completely determined from its values on a fundamental domain of $E^n$.

Given the lattice basis $(b^1, \ldots, b^n)$ of eqn [2] in $E^n$, the vector components of the basis form the $n \times n$ basis matrix $B$ of $\Lambda$. The most general change of the basis preserving the lattice is given by acting with any element $h$ of the general linear group

$\mathrm{Gl}(n, \boldsymbol{Z})$, with integral matrix entries and determinant $\pm 1$, on the lattice basis,

$$\mathrm{Gl}(n, \boldsymbol{Z}) \ni h : B \to B' = Bh \qquad [3]$$

The crystallographic classification of inequivalent lattices in $E^n$ starts from $\mathrm{Gl}(n, \boldsymbol{Z})$. In addition to translations, it employs crystallographic point symmetry operations, (Brown *et al.* 1978, p. 9). A crystallographic point group operation of a lattice $\Lambda$ is a Euclidean isometry $g$ which belongs to a group $G \ni g$ with representations $D : G \to O(n, R)$ and $\mathcal{D} : G \to \mathrm{Gl}(n, \boldsymbol{Z})$ such that

$$G = \{g : D(g)B = B\mathcal{D}(g)\} \qquad [4]$$

The maximal crystallographic point group for given lattice $\Lambda$ is the holohedry of $\Lambda$. The group generated by $T, G$ is a space group which classifies the lattice. For finer details in the classification of space groups, we refer to Brown *et al.* (1978). For crystallography in $E^3$, this classification yields 230 space groups. Crystallography in $E^n$ is described in Schwarzenberger (1980) and in Brown *et al.* (1978) where it is elaborated for $E^4$.

From a lattice $\Lambda \in E^n$ and from the Euclidean metric, one constructs a cell structure as follows: the Voronoi cell $V(b)$, centered at a lattice point $b \in \Lambda$, known in physics as the Wigner–Seitz cell, is the set of points

$$V(b) = \{x \in E^n : |x - b| \leq |x - b'|,\ b' \in \Lambda\} \qquad [5]$$

Any Voronoi cell has a hierarchy of boundaries $X_p$ of dimension $p, 0 \leq p \leq n$ which we denote as $p$-boundaries.

The set of Voronoi cells at all lattice points form the $\Lambda$-periodic Voronoi complex of $\Lambda \in E^n$. The Voronoi cells and complexes associated with a lattice admit a notion of geometric duality. We denote dual objects by a star, $^*$. They are built from convex hulls of sets of lattice points (Kramer and Schlottmann 1989) as follows. A Voronoi $p$-boundary $X_p$ is shared by several Voronoi cells $V(b)$ and determines a set of lattice points

$$S(X_p) : \{b \in \Lambda : X_p \in V(b)\} \qquad [6]$$

The boundary dual to $X_p$ is defined as the convex hull $X^*_{(n-p)} := \mathrm{conv}\{b : b \in S(X_p)\}$. $X^*_{(n-p)}$ can be shown to be an $(n-p)$-boundary of a dual Delone cell. A Delone cell $D$ is defined as the convex hull of all lattice points whose Voronoi cells share a single vertex, called a hole of the lattice. Since these vertices fall into classes of orbits under translations, they determine translationally inequivalent classes of Delone cells $D^\alpha, D^\beta, \ldots$.

Fourier analysis applied to a periodic function $f^p(x)$ on $E^n$ reduces to an $n$-fold Fourier series. The Fourier spectrum is a pure point spectrum and the Fourier coefficients can be referred to the points of a reciprocal lattice $\Lambda^\circ$ (eqn [7]) in Fourier space $E^{\circ n}$. We denote objects belonging to this Fourier space by the index $^\circ$. The basis matrix $B^\circ$ of the reciprocal lattice $\Lambda^\circ \in E^{\circ n}$ is obtained from $B$ as the inverse transpose,

$$\langle b^{\circ i}, b^j \rangle = \delta^{ij} \leftrightarrow B^\circ = (B^{-1})^{\mathrm{T}} \qquad [7]$$

The values of the Fourier coefficients of $f^p(x)$ reduce to integrals over the fundamental domain of the lattice $\Lambda$. From eqns [4] and [7] it follows that the orthogonal representation of a point group $G$ in coordinate and in Fourier space coincides. The Fourier spectrum and its point symmetry in crystals are observed in diffraction experiments.

## Quasiperiodic Point Sets and Functions

Quasiperiodic functions are characterized from their Fourier spectrum (Bohr 1925) by

(**qp**$^\circ$) The Fourier point spectrum of a quasiperiodic function forms a $\boldsymbol{Z}$-module $M^\circ$ of rank $n, n > m$ on Fourier space $E^{\circ m}$.

A $\boldsymbol{Z}$-module of rank $n, n > m$ on $E^{\circ m}$ is defined as a set

$$M^\circ = \left\{ b^\circ : b^\circ = \sum_j^n m_j b^{\circ i},\ (m_1, \ldots, m_n) \in \boldsymbol{Z}^n \right\} \qquad [8]$$

with the $\boldsymbol{Z}$-module basis $(b^{\circ 1}, \ldots, b^{\circ n})$ linearly independent with respect to integral linear combinations. The step from a lattice $\Lambda^\circ$ to a module $M^\circ$ is nontrivial since the set of all module points becomes dense on $E^{\circ m}$. The Fourier coefficients of a quasiperiodic function are assigned to the discrete set of module points (eqn [8]).

Bohr in his analysis of quasiperiodic functions (Bohr 1925, II, pp. 111–125) shows that a general $\boldsymbol{Z}$-module $M^\circ$ of rank $n$ can be taken as the projection to a subspace $E^{\circ m}$ of dimension $m$ of a (nonunique) lattice $\Lambda^\circ \in E^{\circ n}, n > m$. It is convenient to consider in Fourier space $E^{\circ n}$ an orthogonal splitting which we denote as

$$E^{\circ n} = E^{\circ m}_{\parallel} + E^{\circ (n-m)}_{\perp},\ E^{\circ m}_{\parallel} \perp E^{\circ (n-m)}_{\perp} \qquad [9]$$

A characterization of a quasiperiodic function $f^{qp}(x)$ on coordinate space is obtained as follows. From $\Lambda^\circ$ one can construct with the help of eqn [7] the lattice $\Lambda := (\Lambda^\circ)^\circ$ reciprocal to $\Lambda^\circ$ on a coordinate space $E^n$ and associate to it via the Fourier series a quasiperiodic function on a coordinate subspace $E^m_{\parallel}$ of $E^n = E^m_{\parallel} + E^{(n-m)}_{\perp}$, equipped with a $\boldsymbol{Z}$-module $M$ (eqn [11]). As a result one finds a

characterization of a quasiperiodic function in coordinate space:

(**qp**) A quasiperiodic function $f^{qp}(x_\parallel), x_\parallel \in E^m_\parallel$ can always be interpreted as the restriction to a subspace $E^m_\parallel$ of a $\Lambda$-periodic function $f^p(x)$ on $E^n$,

$$E^n = E^m_\parallel + E^{(n-m)}_\perp, x = x_\parallel + x_\perp \\ f^p(x_\parallel + c_\perp) =: f^{qp}(x_\parallel) \qquad [10]$$

In the interpretation (**qp**) (eqn [10]), the $\boldsymbol{Z}$-modules in Fourier space (eqn [8]) and in coordinate space $E^m_\parallel$ become projections of reciprocal lattices,

$$M^\circ = \pi_\parallel(\Lambda^\circ), M = \pi_\parallel(\Lambda) \qquad [11]$$

The linear independence of the module basis enforces a splitting (eqn [9]), irrational with respect to the lattice $\Lambda^\circ \in E^{\circ n}$.

As in the classification of crystal lattices, point symmetry plays a crucial role in the classification of $\boldsymbol{Z}$-modules for quasiperiodic systems like quasicrystals. Noncrystallographic point groups $G$ (with a representation incompatible with any lattice) give rise to quasiperiodic systems as follows:

(**qp**) Given a point group $G$ with orthogonal representations $D_\parallel : G \to O(m, R), D_\perp : G \to O(n-m, R)$ such that $D_\parallel$ is incompatible with any lattice in $E^m_\parallel$, we now require in $E^n$ instead of eqn [10] a lattice $\Lambda$ with basis $B$ and a representation $\mathcal{D} : G \to \mathrm{Gl}(n, \boldsymbol{Z})$ such that

$$\begin{bmatrix} D_\parallel(G) & 0 \\ 0 & D_\perp(G) \end{bmatrix} B = B\,\mathcal{D}(G) \qquad [12]$$

Equation [12] requires that the matrix $B$ provides an irrational reduction of the representation $\mathcal{D}(G)$ into the two representations $D_\parallel(G), D_\perp(G)$. Periodic functions restricted as in the second line of eqn [10] are quasiperiodic.

For any finite group $G$, a representation $\mathcal{D}(G)$ allowing for lattice embedding can always be constructed by the technique of induced representations. Its reduction into representations $D_\parallel(G)$, $D_\perp(G)$ contained in this induced representation is obtained by standard techniques. If $D_\parallel(G)$ is noncrystallographic and inequivalent to $D_\perp(G)$, the subspace decomposition (eqn [12]) is unique.

Quasiperiodic functions compatible with tilings and their windows can be constructed from the dual cell structure (eqns [5] and [6]) of the embedding lattice (Kramer and Schlottmann 1989). Examples are given in the sections "Point symmetry in quasiperiodic systems" and "Quasiperiodic tilings and their windows".

## Scaling and Quasiperiodicity

Quasiperiodic systems lack periodicity but can have scaling symmetries originating from a non-Euclidean extension of eqn [12].

### Example 1: Scaling in the Square Lattice $Z^2$

We begin with the Fibonacci scaling on the square lattice $\boldsymbol{Z}^2$ of $E^2$. The symmetric matrix

$$h = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \in \mathrm{Gl}(2, \boldsymbol{Z}) \qquad [13]$$

has eigenvalues

$$\lambda_1 = -\tau^{-1} = -\tau + 1, \qquad \lambda_2 = \tau := (1 + \sqrt{5})/2 \qquad [14]$$

Evaluation of the orthogonal eigenvectors allows us to define a lattice basis $B = (b^1, b^2)$ and rewrite the eigenvalue equation similar to eqn [12] as

$$\begin{bmatrix} -\tau^{-1} & 0 \\ 0 & \tau \end{bmatrix} B = B \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \\ B = \begin{bmatrix} -\sqrt{\frac{-\tau+3}{5}} & \sqrt{\frac{\tau+2}{5}} \\ \sqrt{\frac{\tau+2}{5}} & \sqrt{\frac{-\tau+3}{5}} \end{bmatrix} \qquad [15]$$

This relation shows that $h$ with respect to the basis $B$ acts as a non-Euclidean point symmetry of the square lattice and generates an infinite discrete group. Equation [15] provides an orthogonal splitting $E^2 = E_\parallel + E_\perp$. The element $h$ acts on the two subspaces as a discrete linear scaling by $-\tau^{-1}, \tau$, respectively. It maps points of $\boldsymbol{Z}^2$ in $E^2$, hence also their projections to $E_\parallel$, into one another.

**Figure 1** shows the lattice basis from eqn [15]. We choose as fundamental domain of $\boldsymbol{Z}^2$ two squares $A, B$ whose boundaries are parallel or perpendicular to $E_\parallel$. A horizontal line $E_\parallel$ intersects these two squares at vertical distances varying with respect to their horizontal boundaries. The quasiperiodic restriction $f^{qp}(x_\parallel) = f^p(x_\parallel + c_\perp)$ of a $\boldsymbol{Z}^2$-periodic function $f^p(x)$ to a line $x = x_\parallel + c_\perp$ picks up varying functional values on these sections. Clearly, one needs all the values of $f^p$ on its fundamental domain in $E^2$ to obtain all the values taken by $f^{qp}$.

Scaling symmetry appears in conjunction with noncrystallographic point symmetry (cf. the following section). Combined with quasiperiodic tilings, it gives rise to a hierarchy of self-similar tilings whose tiles scale with $\tau$.

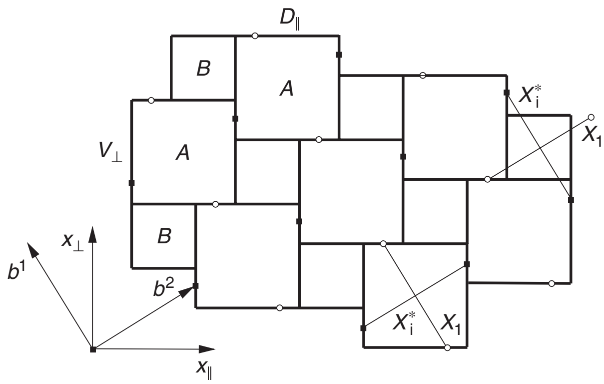



**Figure 1** The square lattice with Fibonacci scaling. Lattice points are black squares, holes white circles. The vectors $(b^1, b^2)$ indicate the lattice basis. The directions $x_\|, x_\perp$ of scalings by $-\tau^{-1}, \tau$ run horizontally and vertically, respectively. Perpendicular and parallel projections $V_\perp$ of Voronoi and $D_\|$ of Delone cells are attached to the lattice and hole points, respectively. Two different pairs of dual 1-boundaries $X_1, X_1^*$ of Voronoi and Delone squares are marked on the right. The product polytopes $X^*_{1,\|} \times X_{1,\perp}$ of their projections form two squares $A, B$ and yield a periodic tiling of $E^2$. A single pair $A, B$ forms a fundamental domain of the lattice. The characteristic functions on $A, B$ are windows for the tiles. A general quasiperiodic function $f^{qp}(x_\|)$ is the restriction of a periodic function $f^p(x)$, defined on $A, B$, to its values on a horizontal line $x = x_\| + c_\perp$. If the periodic function $f^p(x)$ on $A, B$ takes only values independent of $x_\perp$, its quasiperiodic restriction $f^{qp}(x_\|) := f^p(x_\| + c_\perp)$ to this line repeats its values on the long and short tiles $A_\|, B_\|$, respectively, of the standard Fibonacci tiling. Then $A_\|, B_\|$ form a fundamental domain for quasiperiodic functions compatible with the tiling.

## Point Symmetry in Quasiperiodic Systems

Quasiperiodic systems with noncrystallographic point symmetry provide the structure theory and physics of quasicrystals. We illustrate the general scheme (**qp**) of eqn [12] by examples of 5-fold and icosahedral point symmetry. For generalizations, see Janssen (1986).

### Example 2: 5-Fold Point Symmetry from the Root Lattice $A_4$

The $A_4$ root lattice basis in $E^4$ may be derived (Baake *et al.* 1990) from five orthonormal unit vectors $(e^1, e^2, e^3, e^4, e^5)$ in $E^5$ as

$$B = (b^1, b^2, b^3, b^4) := (e^1 - e^2, e^2 - e^3, e^3 - e^4, e^4 - e^5) \qquad [16]$$

As the generator of the cyclic group $C_5$ of 5-fold rotations, we take the cyclic permutation (12345) in cycle notation acting on the vectors $(e^1, e^2, e^3, e^4, e^5)$.

A possible choice of the basis for eqn [12] is the irrational matrix

$$B = (b^1, b^2, b^3, b^4) = \begin{bmatrix} 1 & c & c' & c' & c \\ 0 & s & s' & -s' & -s \\ 1 & c' & c & c & c' \\ 0 & s' & -s & s & -s' \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

$$c = \cos\left(\frac{2\pi}{5}\right) = \frac{\tau - 1}{2}, \quad s = \sin\left(\frac{2\pi}{5}\right) = \frac{\sqrt{\tau + 2}}{2}$$
$$c' = \cos\left(\frac{4\pi}{5}\right) = -\frac{\tau}{2}, \quad s' = \sin\left(\frac{4\pi}{5}\right) = -\frac{\sqrt{3 - \tau}}{2} \qquad [17]$$

Equation [12] for the representation of the generator (12345) of the cyclic group $C_5$ becomes

$$\begin{bmatrix} c & -s & 0 & 0 \\ s & c & 0 & 0 \\ 0 & 0 & c' & -s' \\ 0 & 0 & s' & c' \end{bmatrix} (b^1, b^2, b^3, b^4) = (b^1, b^2, b^3, b^4) \begin{bmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \qquad [18]$$

The left of eqn [18] generates two 2D inequivalent representations of 5-fold planar rotations which are incompatible with any 2D lattice.

The lattice $A_4$ in addition has a scaling symmetry with a factor $\tau$. The scaling transformation may be expressed in terms of the basis (eqn [16]) and an element $h \in Gl(4, \mathbf{Z})$ as

$$\begin{bmatrix} -\tau & 0 & 0 & 0 \\ 0 & -\tau & 0 & 0 \\ 0 & 0 & \tau^{-1} & 0 \\ 0 & 0 & 0 & \tau^{-1} \end{bmatrix} (b^1, b^2, b^3, b^4) = (b^1, b^2, b^3, b^4) \begin{bmatrix} 0 & -1 & 0 & 1 \\ 0 & -1 & -1 & 1 \\ 1 & -1 & -1 & 0 \\ 1 & 0 & -1 & 0 \end{bmatrix} \qquad [19]$$

It is easily verified that the operations of scaling and of 5-fold rotation (eqns [19] and [18]) commute with one another.

**Example 3: Icosahedral Point Symmetry from Lattices $\Lambda = \boldsymbol{Z}^6, D_6$**

The icosahedral group $G = H_3$ has two inequivalent 3D noncrystallographic representations. $H_3$ allows for an induced embedding representation $\mathcal{D}: H_3 \to \mathrm{Gl}(6, \boldsymbol{Z})$, (Kramer and Neri 1984, Kramer *et al.* 1992, Kramer and Papadopolos 1997) into a hypercubic lattice $\Lambda = \boldsymbol{Z}^6$. This representation reduces into two 3D orthogonal inequivalent irreducible noncrystallographic representations $D_\parallel : H_3 \to O(3, R), D_\perp : H_3 \to O(3, R)$. The irrational basis matrix of eqn [12] for $\Lambda = \boldsymbol{Z}^6$ becomes (Kramer *et al.* 1992, p. 185, eqn (7))

$$B = (b^1, b^2, b^3, b^4, b^5, b^6)$$

$$= \sqrt{\frac{1}{2(\tau+2)}} \begin{bmatrix} 0 & 1 & \bar{1} & \bar{\tau} & 0 & \tau \\ 1 & \tau & \tau & 0 & \bar{1} & 0 \\ \tau & 0 & 0 & 1 & \tau & 1 \\ 0 & \tau & \bar{\tau} & 1 & 0 & \bar{1} \\ \tau & \bar{1} & \bar{1} & 0 & \bar{\tau} & 0 \\ \bar{1} & 0 & 0 & \tau & \bar{1} & \tau \end{bmatrix} \qquad [20]$$

with $\bar{\tau} = -\tau, \bar{1} = -1$. The six basis vectors with components in the upper three rows span the so-called primitive icosahedral $\boldsymbol{Z}$-module associated with $D_\parallel$ in $E^3_\parallel$ in the sense of eqn [11]. In this space they point along the directions of six 5-fold axes of the icosahedron.

A second lattice in $E^6$ which admits icosahedral point symmetry is the root lattice $D_6$. The basis of this root lattice, often denoted as the $P$-lattice, is obtained from eqn [20] by a centering matrix given in Kramer *et al.* (1992, p. 185, eqn (8)). The corresponding $\boldsymbol{Z}$-module is inequivalent to the module projected from eqn [20]. The third lattice of icosahedral point symmetry in $E^6$ is $\Lambda = I := P^\circ$ reciprocal to the root lattice $D_6$. All three icosahedral modules admit (powers of) $\tau$-scaling.

## Quasiperiodic Tilings and Their Windows

Quasiperiodic sets of points arise from the general scheme (**qp**) (eqn [12]) by choosing particular periodic functions in the embedding space $E^n$, called the "windows," whose intersections with $E_\parallel$ are the quasiperiodic sets of points.

The window for the construction of a discrete quasiperiodic point set based on eqn [12] is given by the characteristic function $\chi(x_\perp)$ on the projection $V_\perp(x_\perp) := \pi_\perp(V(b))$ of the Voronoi cell (eqn [5]), attached to any lattice point $b \in \Lambda$.

**Example 4: The Quasiperiodic Fibonacci Point Set**

If in the Fibonacci system (Figure 1), one attaches to any point $b$ of the square lattice as a window the characteristic function $\chi$ of the perpendicular projection $V_\perp(b)$ of the unit square attached to $b$, the function $f^{qp}(x_\parallel)$ becomes the standard quasiperiodic Fibonacci sequence of points.

The dual cell geometry of Voronoi and Delone cells and their dual boundaries (eqns [5] and [6]) allows us to construct dual canonical quasiperiodic tilings $(\mathcal{T}, \Lambda), (\mathcal{T}^*, \Lambda)$ (Kramer and Schlottmann 1989). To this end one constructs from local projections of pairs of dual boundaries $X_{m,\parallel}, X^*_{(n-m),\perp}$ or $X^*_{m,\parallel}, X_{(n-m),\perp}$ the direct product polytopes $X_{m,\parallel} \times X^*_{(n-m),\perp}$ or $X^*_\parallel \times X_\perp$ called "klotz polytopes." The characteristic functions on these polytopes form the windows for the tiles $X_{m,\parallel}, X^*_{(n-m),\parallel}$, respectively.

**Example 5: The Quasiperiodic Fibonacci Tiling**

The Voronoi cells $V$ of the square lattice are squares centered at lattice points, the Delone cells $D$ are squares centered at the vertices of Voronoi squares. The product polytopes $X^*_{1,\perp} \times X_{1,\parallel}$ from projections of dual 1-boundaries $X^*_1, X_1$ of Delone and Voronoi squares (cf. Figure 1) become the two types of square windows $A, B$. If a parallel line section $x = x_\parallel + c_\perp$ crosses one of these squares, the tile $A_\parallel$ or $B_\parallel$ is formed. The standard Fibonacci tiling results.

**Example 6: Canonical Tilings from the Root Lattices $A_4, D_6$**

The two rhombus tiles of the planar quasiperiodic Penrose pattern (Penrose 1974) $(\mathcal{T}, A_4)$ are the projections of 2-boundaries of the Voronoi complex of the root lattice $A_4 \in E^4$ (Baake *et al.* 1990). The triangle tiles of the dual tiling $(\mathcal{T}^*, A_4)$ are shown in **Figure 2**.

They are projections of 2-boundaries from the Delone complex of the same lattice. A full analysis of dual Voronoi and Delone boundaries of the root lattice $D_6$ is given in Kramer *et al.* (1992). It leads to icosahedral tilings $(\mathcal{T}, D_6)$ and $(\mathcal{T}^*, D_6)$ of $E^3$, (Kramer *et al.* 1992, Kramer and Papadopolos 1997, Kramer and Schlottmann 1989) and to models of icosahedral quasicrystals.

## Fundamental Domains for Quasiperiodic Tilings

Canonical tilings allow us to construct quasiperiodic functions equipped with a quasiperiodic counterpart of fundamental domains or cells in crystals: assume that the tiles of a tiling $(\mathcal{T}, \Lambda)$ all are translates in $E^m$ of a finite minimal set of prototiles $(X^1, \ldots, X^r)$. Consider the class of quasiperiodic functions which

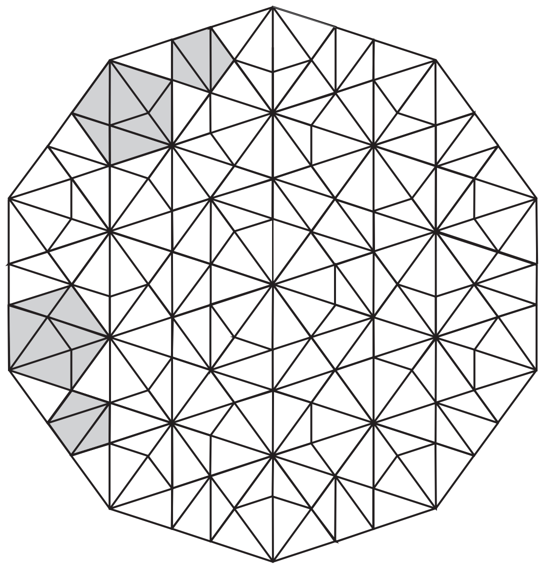


**Figure 2** A patch of the planar quasiperiodic triangle tiling $(\mathcal{T}^*, A_4)$ obtained from the root lattice $A_4 \in E^4$. The tiles are two triangles, projections of 2-boundaries from the Delone cells of $A_4$. The vertices are projections of lattice points. The 20 shaded triangles form a set of prototiles such that any other tile is a translate of one of them. The shaded set forms a fundamental domain for the tiling.

take identical values on any translate of a prototile. These values are prescribed on the finite set of prototiles in $E^m$ which define a fundamental domain for this class of quasiperiodic functions. Only this class of quasiperiodic functions is compatible with the tiling. It can be characterized in the scheme (**qp**) (eqn [12]) by $\Lambda$-periodic functions on $E^n$ whose values on the tile windows of the previous section are independent of the perpendicular coordinate. A fundamental domain for the triangle tiling $(\mathcal{T}^*, A_4)$ is given by the shaded parts in **Figure 2**. The fundamental domain property appears in relation with the theory of covering of quasiperiodic sets (see Kramer and Papadopolos (2000)).

### Example 7: Fundamental Domain for the Fibonacci Tiling

Attach to the squares $A, B$ in **Figure 1** a periodic function $f^p(x)$ with functional values independent of the perpendicular coordinate $x_\perp$ within the two squares. Consider the functional values $f^{qp}(x_\parallel) = f^p(x_\parallel + c_\perp)$ picked up on a parallel line. Clearly, these values become independent of the perpendicular coordinate of any intersection with a square $A, B$. The general prescription of values on a fundamental domain of $\Lambda \in E^2$ needed for a quasiperiodic function reduces to a prescription of its functional values in $E_\parallel$ on the fundamental domain formed by the two prototiles $A_\parallel, B_\parallel$.

## Conclusion

For quasiperiodic systems, the general construction was introduced in the section “Quasiperiodic point sets and functions”, and illustrations were given in four subsequent sections. Further reading resources are provided by the references given at the end. Here, we mention some of the many possible generalizations.

Bohr (1925) considers quasiperiodic as special cases of almost periodic systems. The module of an almost periodic function has a countable basis.

Moody (1997) discusses the notion of Meyer sets. These describe discrete sets on locally compact abelian groups and as particular cases encompass quasiperiodic systems.

Lagarias (2000) studies aperiodic sets characterized by the following properties, shared with periodic and quasiperiodic sets:

(**ap1**): inequivalent patches of points are volume bounded,
(**ap2**): pure point Fourier spectrum,
(**ap3**): linear repetitivity of patches, and
(**ap4**): self-similarity.

*See also:* Compact Groups and Their Representations; Finite Group Symmetry Breaking; Lie Groups: General Theory; Localization for Quasiperiodic Potentials; Symmetries and Conservation Laws; Symmetry and Symmetry Breaking in Dynamical Systems.

## Further Reading


Baake M, Kramer P, Schlottmann M, and Zeidler D (1990) Planar patterns with fivefold symmetry as sections of periodic structures in 4-space. *International Journal of Modern Physics B* 4: 2217–2268.

Bohr H (1925) Zur Theorie der fastperiodischen Funktionen. I. *Acta Mathematicae* 45: 29–127.

Bohr H (1925) Zur Theorie der fastperiodischen Funktionen. II. *Acta Mathematicae* 46: 101–214.

Brown R, Bülow R, Neubüser J, Wondratschek H, and Zassenhaus H (1978) *Crystallographic Groups of Four-Dimensional Space*. New York: Wiley.

de Bruijn NG (1981) Algebraic theory of Penrose’s non-periodic tilings of the plane. I. *Proceedings Koninklijke Nederlandse Akademie van Wetenschappen* 84: 39–52.

de Bruijn NG (1981) Algebraic theory of Penrose’s non-periodic tilings of the plane. II. *Proceedings Koninklijke Nederlandse Akademie van Wetenschappen* 84: 53–66.

Janssen T (1986) Crystallography of quasicrystals. *Acta Crystallographica A* 42: 261–71.

Kramer P and Neri R (1984) On periodic and non-periodic space fillings of $E^n$ obtained by projection. *Acta Crystallographica A* 40: 580–587.

Kramer P and Papadopolos Z (1997) Symmetry concepts for quasicrystals and non-commutative crystallography. In: Moody RV (ed.) *The Mathematics of Long-Range Aperiodic Order*, pp. 307–330. Dordrecht: Kluwer.

Kramer P and Papadopolos Z (eds.) (2000) *Coverings of Discrete Quasiperiodic Sets, Theory and Applications to Quasicrystals*. New York: Springer.

Kramer P, Papadopolos Z, and Zeidler D (1992) The root lattice $D_6$ and icosahedral quasicrystals. In: Frank A, Seligman TH, and Wolf KB (eds.) *Group Theory in Physics*, American Institute of Physics Conference Proceedings, vol. 266, pp. 179–200. New York.

Kramer P and Schlottmann M (1989) Dualization of Voronoi domains and klotz construction: a general method for the generation of proper space filling. *Journal of Physics A: Mathematical and General* 22: L1097–L1102.

Lagarias J (2000) The impact of aperiodic order on mathematics. *Material Science Engineering A* 294–296: 186–191.

Moody RV (1997) Meyer sets and their duals. In: Moody RV (ed.) *The Mathematics of Long-Range Aperiodic Order*, pp. 403–441. Dordrecht: Kluwer.

Penrose R (1974) The role of aesthetics in pure and applied mathematical research. *Bulletin of the Institute of Mathematics and its Applications* 10: 266–71.

Schwarzenberger RLE (1980) *N-Dimensional Crystallography*. San Francisco: Pitman.

Shechtman D, Blech I, Gratias D, and Cahn JW (1984) Metallic phase with long-range orientational order and no translational symmetry. *Physical Review Letters* 53: 1951–1953.

# Quillen Determinant

**S Scott**, King's College London, London, UK



## Determinants in Finite Dimensions

The determinant of a linear transformation $A: V \longrightarrow W$ acting between finite-dimensional complex vector spaces is an element det $A$ of a complex line $L_A$. The abstract element $\det A$ is called the *Quillen determinant* of $A$, and the complex line $L_A$ is called the determinant line of $A$. A choice of (linear) isomorphism

$$\phi : L_A \longrightarrow \mathbb{C} \qquad [1]$$

associates to det $A$ the complex number

$$\det{}_{\phi} A := \phi(\det A) \in \mathbb{C} \qquad [2]$$

which can equivalently be written as the ratio

$$\det{}_{\phi} A = \frac{\det A}{\phi^{-1}(1)} \qquad [3]$$

taken in the one-dimensional complex vector space $L_A$ relative to the canonical generator $\phi^{-1}(1)$. It is not necessarily the case that det $A$ determines a generator for $L_A$; specifically, if $\dim V = m$ and $\dim W = n$, then $\det A = 0$ if $m \neq n$ (by "fiat"), while if $m = n$, then $\det A = 0$ precisely when $A$ is not invertible. For the moment, set $m = n$.

For $k \in \{0, 1, \ldots, n\}$ the $k$th exterior power operator is defined by

$$\wedge^k A : \wedge^k V \longrightarrow \wedge^k W$$
$$\wedge^k A(v_1 \wedge v_2 \wedge \cdots \wedge v_k) := Av_1 \wedge Av_2 \wedge \cdots \wedge Av_k \qquad [4]$$

where $v_1, \ldots, v_k \in V$ and $\wedge^0 V := \mathbb{C}$ and $\wedge^0 A := 1$. When $k = n$, $\operatorname{Det} V := \wedge^n V$ and $\operatorname{Det} W := \wedge^n W$ are complex lines and the determinant line of $A$ is

$$L_A := \operatorname{Det} V^* \otimes \operatorname{Det} W \qquad [5]$$

while for any basis $\{e_1, \ldots, e_n\}$ for $V$, with dual basis $\{e_1^*, \ldots, e_n^*\}$ for $V^*$,

$$\det A := e_1^* \wedge \cdots \wedge e_n^* \otimes (\wedge^n A)(e_1 \wedge \cdots \wedge e_n) \in L_A \qquad [6]$$

There is a canonical isomorphism for $A \in \operatorname{Hom}(V, W)$, $B \in \operatorname{Hom}(U, V)$

$$L_{AB} \cong L_A \otimes L_B \qquad [7]$$

coming from the isomorphism

$$\operatorname{Det} V^* \otimes \operatorname{Det} V \longrightarrow \mathbb{C} \qquad [8]$$

defined by the canonical pairing $\operatorname{Det} V^* \times \operatorname{Det} V \to \mathbb{C}$, and this preserves the determinant elements

$$\det\,(AB) \longleftrightarrow \det A \otimes \det B \qquad [9]$$

### The Classical Determinant

When $V = W$ these constructions take on a more familiar form. Then $\phi$ can be chosen to be the canonical isomorphism [8] and evaluation on $\det A \in L_A$ outputs the classical determinant

$$\det{}_{\mathbb{C}} A = \sum_{\sigma} (-1)^{\sigma} a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} \qquad [10]$$

where the sum is over permutations of $\{1, \ldots, n\}$ and $(a_{i,j})$ is the matrix of $A$ with respect to any basis of $V$ – changing the basis may change the summands on the right-hand side of [10], but not their sum. It is fundamental that when $V = W$ the classical determinant is an intrinsic invariant of the operator $A$, independent of the choice of basis for $V$; when $V \neq W$ that is no longer so since there is then no *canonical* bilinear pairing $\operatorname{Det} V^* \times \operatorname{Det} W \longrightarrow \mathbb{C}$; the choice of a nondegenerate pairing is equivalent to a choice of $\phi$ in [1].

The identification of [10] from [6] and [8] amounts to the identity in Det $V$

$$(\wedge^n A)(e_1 \wedge \cdots \wedge e_n) = \det{}_{\mathbb{C}} A \,.\, e_1 \wedge \cdots \wedge e_n \qquad [11]$$

Since $\wedge^n(AB) = \wedge^n A \circ \wedge^n B$, [11] in turn implies the characterizing multiplicativity property of the classical determinant

$$\det{}_{\mathbb{C}}(AB) = \det{}_{\mathbb{C}}A \, . \det{}_{\mathbb{C}}B \qquad [12]$$

for $A, B \in \mathrm{End}(V)$, specializing the general fact in [7]. Similarly, the group $\mathrm{Gl}(V, \mathbb{C})$ of invertible elements of $\mathrm{End}(V)$ is identified with those $A$ with $\det_{\mathbb{C}} A \neq 0$.

The classical determinant can also be thought of in the following ways. First, the direct sum of the operators defined in [4] yields the total exterior power operator $\wedge A : \wedge V \longrightarrow \wedge V$ on the exterior algebra $\wedge V = \oplus_{k=0}^{n} \wedge^k V$ and this has trace

$$\mathrm{tr}\,(\wedge A) = \det{}_{\mathbb{C}}(I + A) \qquad [13]$$

where $I$ is the identity. Alternatively, one can do something a little more sophisticated and use the holomorphic functional calculus to define the logarithm $\log_\theta B$ of $B \in \mathrm{End}(V)$ by

$$\log_\theta B = \frac{\mathrm{i}}{2\pi} \int_{\Gamma_\theta} \log_\theta \lambda \, (B - \lambda I)^{-1} \, \mathrm{d}\lambda \qquad [14]$$

Here $\log_\theta \lambda$ is the branch of the complex logarithm defined by $\theta - 2\pi < \arg(\lambda) \leq \theta$ and $\Gamma_\theta$ is a positively oriented contour enclosing $\mathrm{spec}(B)$ but not any point of the spectral cut $R_\theta = \{r\mathrm{e}^{\mathrm{i}\theta} \mid r \geq 0\}$. Then, if $B$ is invertible,

$$\mathrm{tr}\,(\log_\theta B) = \log_\theta \det{}_{\mathbb{C}}B \qquad [15]$$

## The Fredholm Determinant

The advantage of the constructions [13] and [15] is that they extend to a restricted class of bounded linear operators on infinite-dimensional Hilbert spaces. This is consequent on the fact that both of the formulas [13] and [15] are computed as operator traces.

(Recall that a trace on a Banach algebra $\mathcal{B}$ is a linear functional $\tau : \mathcal{B} \to \mathbb{C}$ which has the property $\tau([a, b]) = 0$ for all $a$, $b$ in $\mathcal{B}$, where $[a, b] := ab - ba$ is defined by the product structure on $\mathcal{B}$. Since one can define the logarithm $\log_\theta b$ of an element $b$ of $\mathcal{B}$ with spectral cut $R_\theta$ by the formula [14], one in this case obtains a determinant $\det_{\tau,\theta}(b)$ on such elements by setting

$$\log_\theta \det{}_{\tau,\theta}(b) = \tau(\log_\theta b) \qquad [16]$$

If $a, b, ab \in \mathcal{B}$ have common spectral cuts $\theta$, the trace property of $\tau$ translates into the multiplicativity property $\det_{\tau,\theta}(ab) = \det_{\tau,\theta}(a)\det_{\tau,\theta}(b)$ via a version of the Campbell–Hausdorff formula.)

The operator trace arises as follows. Let $H$ be a complex separable Hilbert space with inner product $<,>$, let $\mathcal{C}(H)$ be the algebra of compact operators on $H$, and let

$$L_1 = \left\{ A \in \mathcal{C}(H) \mid \|A\|_1^2 := \sum_{i=1}^{\infty} \mu_i(A^*A) < \infty \right\} \qquad [17]$$

be the ideal of trace-class operators, where the sum is over the real discrete eigenvalues $\mu_i(A^*A) \searrow +0$ of the compact self-adjoint operator $A^*A$. For any orthonormal basis $\{\eta_j\}$ of $H$ the map

$$\mathrm{tr} : L_1 \longrightarrow \mathbb{C}, \qquad A \mapsto \mathrm{tr}\,(A) := \sum_j < \eta_j, A\eta_j >$$

is a trace functional on $L_1(H)$, independent of the choice of basis. Lidskii's theorem states that

$$\mathrm{tr}\,(A) = \sum_{\lambda \in \mathrm{spec}(A)} \lambda \qquad [18]$$

with the sum over the eigenvalues of $A$ counted up to algebraic multiplicity; for general trace-class operators this equality is highly nontrivial.

If $A$ is trace class, then for each non-negative integer $k$ so is each of the exterior power operators $\wedge^k A : \wedge^k H \longrightarrow \wedge^k H$, defined as in [4]. Following [13], a determinant can therefore be defined on the semigroup $I + L_1 := \{I + A \mid A \in L_1\}$ of determinant-class operators by the absolutely convergent sum

$$\det{}_{\mathrm{F}}(I + A) := \mathrm{tr}\,(\wedge A) = 1 + \sum_{k=1}^{\infty} \mathrm{tr}\,(\wedge^k A) \qquad [19]$$

On the other hand, since tr is tracial and $\log_\pi (I + A)$ defined by [14] is trace class, then according to [16], there is a determinant given on invertible determinant-class operators by

$$\log_\pi \det{}_{\mathrm{F}}(I + A) = \mathrm{tr}\,(\log_\pi (I + A)) \qquad [20]$$

which, as the left-hand side already suggests, coincides with the Fredholm determinant.

The Fredholm determinant retains the characterizing properties of the classical determinant in finite dimensions, that $\det_{\mathrm{F}} : I + L_1 \longrightarrow \mathbb{C}$ is multiplicative,

$$\det{}_{\mathrm{F}}((I + A)(I + B)) = \det{}_{\mathrm{F}}(I + A) \det{}_{\mathrm{F}}(I + B), \quad A, B \in L_1 \qquad [21]$$

and $\det_{\mathrm{F}}(I + A) \neq 0$ if and only if $I + A$ is invertible. It is, moreover, essentially unique; any other multiplicative functional on $I + L_1$ is equal to some power of the Fredholm determinant, or, equivalently, any trace on $L_1$ is a constant multiple of the operator trace. The trace property, the operator trace, and the multiplicativity of the Fredholm determinant do not, however, persist to any functional extension of the operator trace (resp. Fredholm determinant) on the

space of pseudodifferential operators of any real order acting on function spaces (fields over space-time). In quantum physics, this is a primary cause of anomalies. More precisely, determinants of differential operators arise in quantum field theories (QFTs) and string theory through the formal evaluation of their defining Feynman path integrals and the calculation of certain stable quantum numbers, which are in some sense "topological."

From the latter perspective, it is instructive to be aware also of the following, third, construction of the Fredholm determinant, which equates the existence of a nontrivial determinant to the existence of nontrivial topology of the general linear group. First, in a surprising contrast to $\mathrm{Gl}(n, \mathbb{C})$, the general linear group $\mathrm{Gl}(H)$ of an infinite-dimensional Hilbert space $H$ with the norm topology is contractible, and hence topologically trivial. By transgression properties in cohomology, this implies any vector bundle with structure group $\mathrm{Gl}(H)$ is isomorphic to the trivial bundle. In order to recapture some topology (and hence, in applications, some physics), it is necessary to reduce to certain infinite-dimensional subgroups of $\mathrm{Gl}(H)$. The most obvious one is the group $\mathrm{Gl}(\infty)$ of of invertible operators differing from the identity by an operator of finite rank. As the inductive limit of the $\mathrm{Gl}(n, \mathbb{C})$, the cohomology and homotopy groups of $\mathrm{Gl}(\infty)$ are a stable version of those of $\mathrm{Gl}(n, \mathbb{C})$. Precisely, $\mathrm{Gl}(\infty)$ is torsion free and its cohomology ring is an exterior algebra with odd degree generators, while Bott (1959) periodicity identifies $\pi_k(\mathrm{Gl}(\infty))$ to be isomorphic to $\mathbb{Z}$ if $k$ is odd and trivial if $k$ is even. Topologically, it is preferable to consider the closure of $\mathrm{Gl}(\infty)$ in $\mathrm{Gl}(H)$, which yields the group $\mathrm{Gl}_{\mathrm{cpt}}(H)$ of operators differing from the identity by a compact operator, but this is now a little "too large" for analysis and differential geometry. Given our earlier comments, there is an intermediate natural choice of the Banach Lie group $\mathrm{Gl}_1(H)$ of operators differing from the identity by a trace-class operator (in fact, there is a tower of such Schatten class groups). Moreover, the inclusions $\mathrm{Gl}(\infty) \subset \mathrm{Gl}_1(H) \subset \mathrm{Gl}_{\mathrm{cpt}}(H)$ are homotopy equivalences, and so the cohomology of $\mathrm{Gl}_1(H)$ is just the exterior algebra mentioned above

$$H^*(\mathrm{Gl}_1(H)) = \wedge(\omega_1, \omega_3, \omega_5, \ldots), \quad \deg\omega_j = 2j - 1 \qquad [22]$$

The advantage of considering $\mathrm{Gl}_1(H)$ is that precise analytical representatives for the classes $\omega_j$ can be written down:

$$\omega_j = \left(\frac{\mathrm{i}}{2\pi}\right)^j \frac{(j-1)!}{(2j-1)!}\, \Theta^{2j-1}$$

where

$$\Theta = \mathrm{tr}\,(Z^{-1}\mathrm{d}Z) \qquad [23]$$

is the 1-form on $\mathrm{Gl}_1(H)$.

This equation makes sense because the derivative $\mathrm{d}Z$ is trace class, and hence so is $Z^{-1}\mathrm{d}Z$. Now, locally $\Theta = \mathrm{d}\log \det_{\mathrm{F}}(Z)$, so that the 1-form $\omega_1$ pulled back by a path $\sigma: S^1 \to \mathrm{Gl}_1(H)$ is precisely the winding number of the curve traced out in $\mathbb{C}^*$ by the function $\det_{\mathrm{F}}(\sigma)$. In fact, this is just a special case of the Bott periodicity theorem, which tells us that the stable homotopy group $\pi_{2j-1}(\mathrm{Gl}_1(H))$ is isomorphic to $\mathbb{Z}$ and an isomorphism is defined by assigning to a map $f: S^{2j-1} \to \mathrm{Gl}_1(H)$ the integer $\int_{S^{2j-1}} f^*\omega_j \in \mathbb{Z}$ (it is not obvious *a priori* that it is an integer).

Notice that it was not necessary to have mentioned the Fredholm determinant of $Z$ at this point. Indeed, the third definition of the Fredholm determinant is to see it as the integral of the 1-form $\Theta$, define

$$\log_\pi \det_{\mathrm{F}}(\mathrm{I} + \mathrm{A}) := \int_\gamma \Theta \qquad [24]$$

where $\gamma: [0,1] \to \mathrm{Gl}_1(H)$ is any path with $\gamma(0) = I$ and $\gamma(1) = I + A$; this uses the connectedness of $\mathrm{Gl}_1(H)$ and independence of the choice of $\gamma$, as guaranteed by Bott periodicity.

Interestingly, this is closely tied in with the Atiyah–Singer index theorem for elliptic pseudodifferential operators (which in full generality uses the Bott periodicity theorem). Here, there is the following simple but quintessential version of that theorem which links it to the winding number of the determinant of the symbol of a differential operator

$$D = \sum_{\alpha| \leq m} a_\alpha(x) D_x^\alpha \qquad [25]$$

on Euclidean space $\mathbb{R}^n$ with $\alpha = (\alpha_1, \ldots, \alpha_n)$ a multi-index of non-negative integers, $|\alpha| = \alpha_1 + \cdots + \alpha_n$, and $D_x = \mathrm{i}\partial/\partial x_i$. Here $D$ acts on $C^\infty(\mathbb{R}^n, V)$ with $V$ a finite-dimensional complex vector space and the coefficients of $D$ are matrices varying smoothly with $x$ which are required to decay suitably fast, $\left|D_x^\beta a_\alpha(x)\right| = O(|x|^{-|\beta|})$ as $|x| \to \infty$. If the symbol $\sigma_D$ of $D$, defined by

$$\sigma_D(x, \xi) = \sum_{\alpha| \leq m} a_\alpha(x) \xi^\alpha \qquad [26]$$

with $\xi = (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n$, satisfies the ellipticity condition of being invertible on the $2n-1$ sphere $S^{2n-1}$ in $(x, \xi)$ space, then $D$ is a Fredholm operator. The index theorem then states

$$\mathrm{index}\,(D) = \int_{S^{2j-1}} \sigma_D^*(\omega_n)$$

the higher-dimensional analog of the winding number of the determinant.

## Fredholm Operators and Determinant Line Bundles

The operators whose determinants are considered in this article are all Fredholm operators. Recall that a linear operator $A : H_1 \to H_2$ between Hilbert spaces is Fredholm if it is invertible modulo compact operators; that is, there is a "parametrix" $Q : H_2 \to H_1$ such that $QA - I$ and $AQ - I$ are compact operators on $H_1$ and $H_2$, respectively. Equivalently, the range $A(H_1)$ of $A$ is closed in $H_2$, and the kernel $\mathrm{Ker}(A) = \{\eta \in H_1 \mid A\eta = 0\}$ and cokernel $\mathrm{Coker}(A) = H_2/A(H_1)$ of $A$ are finite dimensional. (This is equally true for Banach and Frechet spaces, we restrict our attention to Hilbert spaces for brevity.) The space $\mathsf{Fred}$ of all such Fredholm operators with the norm topology has the homotopy type of the classifying space $\mathbb{Z} \times B\mathrm{Gl}(\infty)$. The first factor parametrizes the connected components of $\mathsf{Fred}$, two Fredholm operators are in the same component if and only if they have the same index

$$\mathrm{index}\,(A) = \dim \mathrm{Ker}(A) - \dim \mathrm{Coker}(A)$$

Mostly we restrict our attention to the connected component $\mathsf{Fred}_0$ of operators of index zero. The cohomology of $\mathsf{Fred}_0 \sim B\mathrm{Gl}(\infty)$ is a polynomial ring

$$H^*(\mathsf{Fred}_0) = \mathbb{R}[\mathrm{ch}_1, \mathrm{ch}_2, \mathrm{ch}_3, \ldots]$$

whose generators may be formally realized as the even degree components of the Chern character of an infinite-dimensional bundle over $\mathsf{Fred}_0$. In fact, the generators $\omega_{2j-1}$ of $H^*(\mathrm{Gl}_1(H))$ are related to the $\mathrm{ch}_j$ through transgression, see Chern and Simons (1974). We shall be interested here in the first generator $\mathrm{ch}_1$, a transgression of the Fredholm determinant "winding number 1-form" $\omega_1$, which coincides with the real Chern class of a canonical complex line bundle $\mathrm{DET}_0 \to \mathsf{Fred}_0$. The fiber of $\mathrm{DET}_0$ at $A \in \mathsf{Fred}_0$ is the determinant line $\mathrm{Det}(A)$ of the Fredholm operator $A$, which is defined as follows (Segal 2004).

Just as for finite-rank operators (see the subsection "Determinants in finite dimensions"), the determinant of a Fredholm operator $A : H^1 \to H^2$ exists abstractly not as a number but as an element $\det A$ of a complex line $\mathrm{Det}(A)$. For simplicity, we suppose that $\mathrm{index}\,(A) = 0$. Elements of the determinant line $\mathrm{Det}(A)$ are equivalence classes $[E, \lambda]$ of pairs $(E, \lambda)$, where $E : H^1 \to H^2$ such that $A - E$ is trace class and relative to the equivalence relation $(Eq, \lambda) \sim (E, \det_{\mathrm{F}}(q)\lambda)$ for $q : H^1 \to H^1$ of determinant class and where $\det_{\mathrm{F}}(q)$ is the Fredholm determinant of $q$. Complex multiplication on $\mathrm{Det}(A)$ is defined by $\mu[A, \lambda] = [A, \mu\lambda]$. The abstract, or Quillen, determinant of $A$ is the preferred element $\det A := [A, 1]$ in $\mathrm{Det}(A)$.

Here are some essential properties of the determinant line. First, $\det A$ is nonzero if and only if $A$ is invertible. Next, quotients of abstract determinants in $\mathrm{Det}(A)$ are given by Fredholm determinants; for if $A_1 : H^1 \to H^2$, $A_2 : H^1 \to H^2$ are Fredholm operators such that $A_i - A$ are trace class, then if $A_2$ is invertible we see that $A_2^{-1}A_1$ is determinant class and hence from the definition that

$$\frac{\det(A_1)}{\det(A_2)} = \det{}_{\mathrm{F}}(A_2^{-1}A_1) \qquad [27]$$

where the quotient on the left-hand side is taken in $\mathrm{Det}(A)$. The principal functorial property of the determinant line is that given a commutative diagram with exact rows and Fredholm columns

$$\begin{array}{ccccccccc} 0 & \longrightarrow & H_1 & \longrightarrow & H_1' & \longrightarrow & H_1'' & \longrightarrow & 0 \\ & & \downarrow_A & & \downarrow_{A'} & & \downarrow_{A''} & & \\ 0 & \longrightarrow & H_2 & \longrightarrow & H_2' & \longrightarrow & H_2'' & \longrightarrow & 0 \end{array} \qquad [28]$$

then there is canonical isomorphism of complex lines

$$\mathrm{Det}(A') \cong \mathrm{Det}(A) \otimes \mathrm{Det}(A'') \qquad [29]$$

preserving the Quillen determinants $\det(A') \leftrightarrow \det(A) \otimes \det(A'')$. A consequence of this property is that given Fredholm operators $A : H_2 \to H_3$ and $B : H_1 \to H_2$, then

$$\mathrm{Det}(AB) \cong \mathrm{Det}(A) \otimes \mathrm{Det}(B)$$

with $\det(AB) \leftrightarrow \det(A) \otimes \det(B)$, generalizing the elementary property [9].

The principal context of interest for studying determinant lines is the case where one has a family $\mathcal{A} = \{A_x \mid x \in B\}$ of Fredholm operators parametrized by a manifold $B$, satisfying suitable continuity properties, and one aims to make sense of the determinant as a function $\mathcal{A} \to \mathbb{C}$. It is then of no difficulty to show that the corresponding family of determinant lines $\mathrm{DET}(\mathcal{A}) = \cup \mathrm{Det}(A_x)$ defines a complex line bundle over $B$ endowed with a canonical section $\det : B \longrightarrow \mathrm{DET}(\mathcal{A})$

assigning to $x \in B$ the Quillen determinant $\det(A_x) \in \mathrm{Det}(A_x)$ (Quillen 1985, Segal 2004). To identify the Quillen determinant section with a function on $\mathcal{A}$, we need to identify a trivialization of the line bundle $\mathrm{DET}(\mathcal{A})$, giving a global basis for the fibers. This is the same thing as giving a non(or never)vanishing section $\psi : B \to \mathrm{DET}(\mathcal{A})$, with respect to which we have the regularized determinant function (cf. [3]):

$$x \mapsto \det{}_{\psi}(A_x) := \frac{\det(A_x)}{\psi(x)} \qquad [30]$$

If $\mathcal{A}$ is trivializable, so a nonzero section exists, there will be many such sections and some extra data is needed to fix a natural choice of $\psi$.

Each of the properties mentioned above for determinant lines carries forward to determinant line bundles in a natural way. In particular, one easily deduces from [28], or from the exact sequence

$$0 \longrightarrow \mathrm{Ker}A_x \longrightarrow H_{1,x} \xrightarrow{A_x} H_{2,x} \longrightarrow \mathrm{Coker}A_x \longrightarrow 0$$

that if the kernels $\mathrm{Ker}A_x$ have constant dimension as $x$ varies, then there is a canonical isomorphism

$$\mathrm{Det}(\mathcal{A}) \cong \wedge^{\max}\mathrm{Ker}(\mathcal{A})^* \otimes \wedge^{\max}\mathrm{Coker}(\mathcal{A}) \qquad [31]$$

where $\mathrm{Ker}(\mathcal{A})$ is the finite-rank complex vector bundle over $B$ with fiber $\mathrm{Ker}A_x$, and $\mathrm{Coker}(\mathcal{A})$ similarly. The interesting feature here is that it shows the determinant bundle to be the top exterior power of the index bundle $\mathrm{Ind}(\mathcal{A}) = [\mathrm{Ker}(\mathcal{A})] - [\mathrm{Coker}(\mathcal{A})] \in K(B)$ in the even $K$-theory of $B$, and in this sense determinant theory may be seen as a particular aspect of index theory – understood is the very broadest sense; in fact, the computation of determinants is usually a considerably more complex and difficult task than computing an index.

## Determinant Bundles for Differential Operators over Manifolds

The Quillen determinant has been of particular interest in the case of families of Dirac operators. Such a family is associated to a $C^\infty$ fibration $\pi : M \longrightarrow B$ of closed boundaryless finite-dimensional Riemannian manifolds of even dimension. If there is a graded Hermitian vector bundle $\mathcal{E} = \mathcal{E}^+ \oplus \mathcal{E}^- \to M$ of Clifford modules, then from the Riemannian structure one can construct a Levi-Civita connection on the vertical tangent bundle $T(M/B)$ which can be lifted to a Clifford connection on $\mathcal{E}$; for example, the spinor connection if we have a family of spin manifolds. This data yields a smooth family of first-order elliptic differential operators $\boldsymbol{D} = \{D_x : C^\infty(M_x; \mathcal{E}_x^+) \longrightarrow C^\infty(M_x; \mathcal{E}_x^-) \mid x \in B\}$ of chiral Dirac-type, with $D_x$ a Dirac-type operator acting over the manifold $M_x = \pi^{-1}(x)$ parametrized by the fibration, along with a determinant line bundle $\mathrm{DET}(\boldsymbol{D}) \to B$ endowed with a canonical section $x \mapsto \det(D_x)$. There are various contexts in mathematics and physics in which one would like to assign to the determinant section a naturally associated smooth function (a regularized determinant) $\det_{\mathrm{reg}} : B \longrightarrow \mathbb{C}$, which can, for example, then be integrated. As discussed in the previous section, this depends on identifying a trivializing (nonzero) section of $\mathrm{DET}(\boldsymbol{D})$. For such a section to exist, the first Chern class $c_1(\mathrm{DET}(\boldsymbol{D}) \in H^2(B)$ must vanish, and this in turn can be computed as a term in the Atiyah–Singer (1984) index theorem for families. Indeed, this is clear from the formal identification [31] which here takes on a precise meaning.

The following simple example, which is the basic topological anomaly computation in string theory, may help to explain the type of computation. Let $M_x$ be a copy of $\Sigma$ a compact Riemann surface, so that $M$ is a family of surfaces parametrized by $B$. Let $T = \cup T_x$ be the vertical complex tangent line bundle on $M$, where $T_x$ is the complex tangent line bundle to $M_x$. Each fiber has an associated $\overline{\partial}$-operator $\overline{\partial}_x$ which we couple to the Hermitian bundle $\mathcal{E}_x := T_x^{\otimes m}$ for $m$ a non-negative integer. In this way, we get a family $\boldsymbol{D}_\Sigma$ of $\overline{\partial}$-operators coupled to $\mathcal{E} = T^{\otimes m}$ whose index bundle is the element $\mathrm{Ind}(\boldsymbol{D}_\Sigma) = f_!(T^{\otimes m}) \in K(B)$. The Atiyah–Singer index theorem for families in this situation coincides with the Grothendieck–Riemann–Roch theorem and this says that

$$\mathrm{ch}(f_!(T^{\otimes m})) = f_*(\mathrm{ch}(T^{\otimes m})\mathrm{Todd}(T))$$

where ch is the Chern character class and $\mathrm{Todd}(T)$ is the Todd class defined for a vector bundle $F$ whose first few terms are

$$\mathrm{Todd}(F) = 1 + \tfrac{1}{2}c_1(F) + \tfrac{1}{12}c_1(F)^2 + \cdots$$

and where $f_* : H^i(M) \to H^{i-1}(B)$ is integration over the fibers. That is, with $\xi = c_1(T)$,

$$\begin{aligned}\mathrm{ch}(f_!(T^{\otimes m})) &= f_*\big(\big(1 + m\xi + \tfrac{1}{2}m^2\xi^2 + \cdots\big) \\ &\qquad \times \big(1 + \tfrac{1}{2}\xi + \tfrac{1}{12}\xi^2 + \cdots\big)\big) \\ &= f_*\big(1 + \big(m + \tfrac{1}{2}\big)\xi + \tfrac{1}{12}\big(m^2 + m + \tfrac{1}{6}\big) \\ &\qquad \times \xi^2 + \cdots\big)\end{aligned}$$

So we have

$$c_1(f_!(T^{\otimes m})) = \tfrac{1}{12}(6m^2 + 6m + 1)f_*(\xi^2) \in H^2(B) \qquad [32]$$

But for any element of $K$-theory, $c_1(E) = c_1(\mathrm{DET}(E))$, and so the left-hand side of [32] is the first Chern class of the determinant line bundle $\mathrm{DET}(\boldsymbol{D}_\Sigma)$. If we take, in particular, $B = \mathrm{Conf}(\Sigma)$, the space of conformal classes of metrics on $\Sigma$ (or compact subsets of this space), and couple the family $\boldsymbol{D}_\Sigma$ to a background trivial real bundle of rank $d/2$, or its negative in $K$-theory, then taking $m = 1$ [32] is easily seen to be modified to

$$c_1(\boldsymbol{D}_{\Sigma,-d/2}) = \frac{(d-26)}{24} f_*(\xi^2)$$

It follows for this topological anomaly to vanish one must have background spacetime of dimension $d = 26$. The idea here is that $\mathrm{Conf}(\Sigma)$ is a configuration space for bosonic strings in $\mathbb{R}^d$ with the requirement that the determinant section of the determinant line bundle be conformally invariant, corresponding to the classical invariance of the string Lagrangian defining the string path integral from which the determinant arises. That is, in order to evaluate the path integral on the reduced configuration space, one requires a trivialization of the determinant line bundle which defines a conformally invariant regularized determinant function. The above calculation says that there is a topological obstruction to this occurring when the background space dimension differs from 26.

This is the most basic example of determinant anomaly computations, which have acquired considerably more sophisticated constructions in modern versions of string theory and QFT. One immediate deficiency in the approach explained so far is that not all anomalies are topological and so even though the first Chern class of the determinant line bundle may vanish, there may still be local and global obstructions to the existence of a determinant function with the correct symmetry properties. To be more precise, one needs to say not just that a trivialization of the determinant line bundle formally exists, but to actually be able to construct a specific preferred trivialization. For this more refined objective, one needs to know more about the differential geometry of the determinant line. One approach is to fix a canonical choice of connection and, if the determinant bundle is topologically trivial, to construct a determinant section (up to phase) using the parallel transport of the connection.

The principal contribution to such a theory was made in a remarkable four page paper by Quillen (1985) in which using zeta-function regularization he presented a construction of a metric and connection on the determinant line bundle for a family of $\bar{\partial}$-operators over a Riemann surface coupled to a holomorphic vector bundle. (This is the first paper one should read on determinant line bundles; Quillen's motivation, in fact, did not come from physics but from a problem in number theory.)

To outline this construction, which was extended to general families of Dirac-type operators in Bismut and Freed (1986), first we recall that if $\Delta$ is an invertible Laplacian-type second-order elliptic differential operator acting on the space of sections of a vector bundle over a compact manifold of dimension $n$, then it has a spectrum consisting of real discrete eigenvalues $\{\lambda\}$ forming an unbounded subset of the positive real line. The zeta function of $\Delta$ is defined in the complex half-plane $\mathrm{Re}(s) > n/2$ by

$$\zeta(\Delta, s) = \mathrm{tr}\,(\Delta^{-s}) = \sum_\lambda \lambda^{-s}, \quad \mathrm{Re}(s) > \frac{n}{2}$$

and extends to a meromorphic function of $s$ on the whole complex plane. It turns out that the extension has no pole at $s = 0$ and this means that we may define the zeta-function regularized determinant of $\Delta$ by

$$\det_\zeta(\Delta) := \exp\left(-\frac{\mathrm{d}}{\mathrm{d}s_{|s=0}} \zeta(\Delta, s)\right)$$

since $(\mathrm{d}/\mathrm{d}s)|_{s=0} \lambda^s = \log \lambda$ this formally represents a regularized product of the eigenvalues of $\Delta$. A metric is now defined on the determinant line bundle $\mathrm{DET}(\boldsymbol{D})$ by defining the norm square of the element $\det(D_x) \in \mathrm{Det}(D_x)$ by

$$\| \det(D_x) \|^2 := \det_\zeta(D_x^* D_x)$$

over the subset $B_0$ of $x \in B$ where $D_x$ is invertible. Elsewhere in $B$, one includes a factor defined by the induced $L^2$ metric in the kernel and cokernel. See Quillen (1985) and Bismut and Freed (1986) for full details.

A connection is defined by similarly constructing a regularized version of the connection we would define if we were working with finite-rank bundles. First, one includes in the data associated to the fibration $\pi : M \longrightarrow B$ defining the family of operators $\boldsymbol{D}$ a splitting of the tangent bundle $TM = T(M/B) \oplus \pi^*(TB)$. This assumption and the Riemannian geometry of the fibration yield a connection $\nabla^{(\pi)}$ defined along the fibers of the fibration. The connection form over $B_0$ is then defined by

$$\omega(x) = \mathrm{tr}_\zeta(D_x^{-1} \nabla^{(\pi)} D_x)$$

where the zeta-regularized trace $\mathrm{tr}_\zeta$ is defined on a vertical bundle endomorphism-valued 1 form $x \mapsto A_x$ on $M$ by

$$\mathrm{tr}_\zeta(A_x) := \mathrm{fp}_{s=0}\mathrm{tr}\,\left(A_x\left(D_x^* D_x\right)^{-s}\right)|^{\mathrm{mer}}$$

where the superscript indicates we are considering the meromorphically extended form, and $\mathrm{fp}_{s=0}(G(s))$ means the finite part of a meromorphic function $G$ on $\mathbb{C}$; that is, the constant term in the Laurent expansion of $G(s)$ near $s = 0$.

A theorem of Bismut and Freed, generalizing Quillen's original computation, computes the curvature $\Omega^{(\mathrm{DET}(\boldsymbol{D}))}$ of this connection to be the 2-form component in the local Atiyah–Singer families index density. This is a refined version of the topological version of that theorem which we utilized earlier; it expresses the characteristic classes on $B$ in terms of specific canonical differential forms constructed by integrating, along the fibers of the fibration, canonically defined vertical characteristic forms. More precisely, they prove the formula (Bismut and Freed 1986 and Berline *et al.* 1992)

$$\Omega^{(\mathrm{DET}(\boldsymbol{D}))} = (2\pi \mathrm{i})^{-n/2}\left(\int_{M/B} \widehat{A}(M/B)\mathrm{ch}(\mathcal{E})\right)_{[2]} \qquad [33]$$

where $(\sigma)_{[2]} \in \Omega^2(B)$ means the 2-form component of a differential form $\sigma$ on $B$. Here $\widehat{A}(M/B) = \det^{1/2}((R^{M/B}/2)/\sinh(R^{M/B}/2))$ is the vertical $\widehat{A}$-genus differential form, while $\mathrm{ch}(\mathcal{E})$ is the vertical Chern character form associated to the curvature form of the bundle $\mathcal{E}$.

This theory seems a long way from the classical theory of stable characteristic classes and the Fredholm determinant discussed in earlier sections. There are, however, interesting parallels which may guide the search for an understanding of the geometry of families of elliptic operators, of which determinants form a component. The prototypical situation where determinants arise in the quantization of gauge theory is the following. Consider the infinite-dimensional affine space $\mathcal{A}$ of connections on a complex vector bundle $E$ with structure group $G$ sitting over $S^n$ the $n$-sphere. The Lie group $G$ is assumed to be compact. For each connection $A \in \mathcal{A}$, we consider a Dirac operator $D_A : C^\infty(S^n, S^+ \otimes E) \longrightarrow C^\infty(S^n, S^- \otimes E)$, where $E$ is a Hermitian vector bundle coupled to the spinor bundles $S^\pm$. The group $\mathcal{G}$ of based gauge transformations acts on $\mathcal{A}$ and symmetry properties of conservation laws lead one to be interested in constructing a determinant function on the quotient space $\mathcal{A}/\mathcal{G}$. More precisely, $g \in \mathcal{G}$ transforms $D_A$ to $D_{g.A}$ and by equivariance the Quillen determinant section pushes down to a section of a reduced determinant line bundle over $\mathcal{A}/\mathcal{G}$. As seen earlier, the topological obstruction to realizing this determinant section as a function on $\mathcal{A}/\mathcal{G}$ can be computed from the Atiyah–Singer index theorem for families applied to the corresponding index bundle $\mathrm{Ind}(\boldsymbol{D}_{\mathcal{A}/\mathcal{G}})$ in the $K(\mathcal{A}/\mathcal{G})$ by picking out the degree-2 component in $H^2(\mathcal{A}/\mathcal{G})$ of the Chern character $\mathrm{ch}(\mathrm{Ind}(\boldsymbol{D}_{\mathcal{A}/\mathcal{G}}))$. On the other hand, it turns out that this characteristic class is the transgression of the element of $H^1(\mathcal{G}, \mathbb{Z})$ defined by the zeta-determinant trace

$$\begin{aligned}\Theta_\zeta &:= \mathrm{tr}_\zeta\left(\left(D_A^* D_{g.A}\right)^{-1} d_{\mathcal{G}}\left(D_A^* D_{g.A}\right)\right) \\ &:= \mathrm{fp}_{s=0}\mathrm{tr}\,\left(D_A^* D_{g.A}\right)^{-1} d_{\mathcal{G}}\left(D_A^* D_{g.A}\right)\left(D_A^* D_{g.A}\right)^{-s})|^{\mathrm{mer}}\end{aligned}$$

which counts the winding number of the zeta determinant $\mathcal{G} \longrightarrow \mathbb{C}^*$ defined by $\det_\zeta(D_A^* D_{g.A})$. This provides an interesting parallel of the classical theory described in the section "The Fredholm determinant." For more details of this and more advanced ideas take a look at Singer (1985). (A similar parallel holds between the topological derivation of the conformal anomaly outlined at the beginning of this section and what it called the Polyakov multiplicative anomaly formula for the zeta determinant of the Laplacian with respect to conformal changes in the metric on the surface.)

Aspects of more recent work in this direction have been the extension of the theory to manifolds with boundary, and how it encodes into the structures of topological and conformal field theories, see Segal (2004) and Mickelsson and Scott (2001), and more generally into $M$-theory (Freed and Moore 2004).

*See also:* Anomalies; Feynman Path Integrals; Index Theorems; Regularization for Dynamical $\zeta$-Functions.

## Further Reading


Alvarez O and Singer IM (1984) Gravitational anomalies and the families index theorem. *Communications in Mathematical Physics* 96: 409–417.

Atiyah MF and Singer IM (1984) Dirac operators coupled to vector potentials. *Proceedings of the National Academy of Sciences of the USA* 81: 2597–2600.

Bott R (1959) The stable homotopy of the classical groups. *Annals of Mathematics* 70: 313–337.

Berline N, Getzler E, and Vergne M (1992) Heat Kernels and Dirac Operators, *Grundlehren der Mathematischen Wissenschaften*, vol. 298. Berlin: Springer.

Bismut J-M and Freed DS (1986) The analysis of elliptic families I. *Communications in Mathematical Physics* 106: 159–176.

Chern SS and Simons J (1974) Characteristic forms and geometric invariants. *Annals of Mathematics* 99: 48–69.

Freed DS and Moore G (2004) Setting the quantum integrand of M-theory, arXiv:hep-th/0409135.
Mickelsson J and Scott S (2001) Functorial QFT, gauge anomalies and the Dirac determinant bundle. *Communications in Mathematical Physics* 219: 567–605.
Quillen D (1985) Determinants of Cauchy–Riemann opeators over a Riemann surface. *Functional Analysis and its Applications* 19: 31–34.
Segal G (2004) The definition of conformal field theory. In: Tillmann U (ed.) *Topology, Geometry and Quantum Field Theory*, pp. 421–577. Cambridge University Press. *Topological field theory*, "Stanford lecture notes," http://www.cgtp.duke.edu.
Singer IM (1985) Families of Dirac operators with applications to physics. *Asterique, hors serie*, 323–340.

**Quivers** *see* Finite-Dimensional Algebras and Quivers

# R

## Random Algebraic Geometry, Attractors and Flux Vacua

**M R Douglas**, Rutgers, The State University of New Jersey, Piscataway, NJ, USA



### Introduction

A classic question in probability theory, studied by M Kac, S O Rice, and many others, is to find the expected number and distribution of zeros or critical points of a random polynomial. The same question can be asked for random holomorphic functions or sections of bundles, and are the subject of "random algebraic geometry."

While this theory has many physical applications, in this article we focus on a variation on a standard question in the theory of disordered systems. This is to find the expected distribution of minima of a potential function randomly chosen from an ensemble, which might be chosen to model a crystal with impurities, a spin glass, or another disordered system. Now whereas standard potentials are real-valued functions, analogous functions in supersymmetric theories, such as the superpotential and the central charge, are holomorphic sections of a line bundle. Thus, one is interested in finding the distribution of critical points of a randomly chosen holomorphic section.

Two related and much-studied problems of this type are (1) the problem of finding attractor points in the sense of Ferrara, Kallosh, and Strominger, and (2) the problem of finding flux vacua as posed by Giddings, Kachru, and Polchinski. These problems involve a good deal of fascinating mathematics and are good illustrations of the general theory.

A note on general references for further reading on the subject of this article is in order. For background on random algebraic geometry and some of its other applications, as well as references in the text not listed here, consult Edelman and Kostlan (1995) and Zelditch (2001). The attractor problem is discussed in Ferrara *et al.* (1995) and Moore (2004), while IIb flux vacua were introduced in Giddings *et al.* (2002). Background on Calabi–Yau manifolds can be found in Cox and Katz (1999) and Gross *et al.* (2003).

### Elementary Random Algebraic Geometry

Let us introduce this subject with the problem of finding the expected distribution of zeros of a random polynomial,

$$f(z) = c_0 + c_1 z + \cdots + c_N z^N$$

We define a random polynomial to be a probability measure on a space of polynomials. A natural choice might be independent Gaussian measures on the coefficients,

$$d\mu\,[f] = d\mu[c_0, \ldots, c_N] = \prod_{i=0}^{N} d^2 c_i \frac{\sigma_i}{2\pi}\, e^{-|c_i|^2/2\sigma_i^2} \quad [1]$$

We still need to choose the variances. At first the most natural choice would seem to be equal variance for each coefficient, say $\sigma_i = 1/2$. We can characterize this ensemble by its two-point function,

$$\begin{aligned} G(z_1, \bar{z}_2) &\equiv \mathbb{E}[f(z_1) f^*(\bar{z}_2)] \\ &= \int d\mu[f]\, f(z_1) f^*(\bar{z}_2) \\ &= \sum_{n=0}^{N} (z_1 \bar{z}_2)^n \\ &= \frac{1 - z_1^{N+1} \bar{z}_2^{N+1}}{1 - z_1 \bar{z}_2} \end{aligned}$$

We now define $d\mu_0(z)$ to be a measure with unit weight at each solution of $f(z) = 0$, such that its integral over a region in $\mathbb{C}$ counts the expected number of zeros in that region. It can be written in terms of the standard Dirac delta function, by multiplication by a Jacobian factor,

$$d\mu_0(z) = \mathbb{E}[\delta^{(2)}(f(z)) \partial f(z) \bar{\partial} f^*(\bar{z})] \quad [2]$$

To compute this expectation value, we introduce a constrained two-point function,

$$G_{f(z)=0}(z_1, \bar{z}_2) = \frac{\mathbb{E}[\delta^{(2)}(f(z))\, f(z_1) f^*(\bar{z}_2)]}{\mathbb{E}[\delta^{(2)}(f(z))]}$$

It could be explicitly computed by using the constraint $f(z) = 0$ to solve for a coefficient $c_i$ in

the Gaussian integral, that is, projecting on the linear subspace $0 = \sum c_i z^i$. The result, in terms of $G(z_1, \bar{z}_2)$, is

$$\mathbb{E}[\delta^{(2)}(f(z))] = \frac{1}{\pi G(z, \bar{z})}$$
$$G_{f(z)=0}(z_1, \bar{z}_2) = G(z_1, \bar{z}_2) - \frac{G(z_1, \bar{z})G(z, \bar{z}_2)}{G(z, \bar{z})}$$

as can be verified by considering

$$\mathbb{E}[\delta^{(2)}(f(z))f(z)f^*(\bar{z}_2)] \propto G_{f(z)=0}(z, \bar{z}_2)$$
$$= G(z, \bar{z}_2) - \frac{G(z, \bar{z})G(z, \bar{z}_2)}{G(z, \bar{z})} = 0$$

Using this, eqn [2] can be evaluated by taking derivatives:

$$\mathrm{d}\mu_0(z) = \frac{1}{G(z, \bar{z})} \lim_{z_1, z_2 \to z} D_1 \bar{D}_2 G_z(z_1, \bar{z}_2)$$
$$= \frac{1}{\pi} \partial\bar{\partial} \log G(z, \bar{z})$$

For the constant variance ensemble eqn [2],

$$\mathrm{d}\mu_0(z) = \frac{\mathrm{d}^2 z}{\pi} \left( \frac{1}{(1 - z\bar{z})^2} - \frac{(N+1)^2 (z\bar{z})^N}{(1 - (z\bar{z})^{N+1})^2} \right) \qquad [3]$$

We see that as $N \to \infty$, the zeros concentrate on the unit circle $|z| = 1$ (Hammersley 1954).

A similar formula can be derived for the distribution of roots of a real polynomial on the real axis, using $\mathrm{d}\mu(t) = \mathbb{E}[\delta(f(t))|\mathrm{d}f/\mathrm{d}t|]$. One obtains (Kac 1943):

$$\mathrm{d}\mu_0^r(t) = \frac{\mathrm{d}t}{\pi} \sqrt{\frac{1}{(1 - t^2)^2} - \frac{(N+1)^2 t^{2N}}{(1 - t^{2N+2})^2}}$$

Integrating, one finds the expected number of real zeros of a degree $N$ random real polynomial is $E_N \sim (2/\pi) \log N$, and as $N \to \infty$ the zeros are concentrated at $t = \pm 1$.

While concentration of measure is a fairly generic property for random polynomials, it is by no means universal. Let us consider another Gaussian ensemble, with variance $\sigma_n = N!/n!(N-n)!$. This choice leads to a particularly simple two-point function,

$$G(z, \bar{z}) = (1 + z\bar{z})^N \qquad [4]$$

and the distribution of zeros

$$\mathrm{d}\mu_0 = \frac{1}{\pi} \partial\bar{\partial} \log G = \frac{N \, \mathrm{d}^2 z}{\pi (1 + z\bar{z})^2} \qquad [5]$$

Rather than concentrate the zeros, in this ensemble zeros are uniformly distributed according to the volume of the Fubini–Study (SU(2)-invariant) Kähler metric

$$\omega = \partial\bar{\partial} K, \quad K = \log(1 + z\bar{z})$$

on complex projective space $\mathbb{CP}^1$.

We can better understand the different behaviors in our two examples by focusing on a Hermitian inner product $(f, g)$ on function space, associated to the measure eqn [1] by the formal expression

$$\mathrm{d}\mu[f] = [Df] \, \mathrm{e}^{-(f, f)}$$

In making this precise, let us generalize a bit further and allow $f$ to be a holomorphic section of a line bundle $\mathcal{L}$, say $\mathcal{O}(N)$ over $\mathbb{CP}^1$ in our examples. We then choose an orthonormal basis of sections $(s_i, s_j) = \delta_{ij}$, and write

$$f \equiv \sum_i c_i s_i \qquad [6]$$

and

$$\mathrm{d}\mu[f] = \frac{1}{(2\pi)^N} \prod_{i=1}^{N} \mathrm{d}^2 c_i \, \mathrm{e}^{-|c_i|^2/2}$$

We can then compute the two-point function

$$G(z_1, \bar{z}_2) \equiv \mathbb{E}[s(z_1)s^*(\bar{z}_2)] = \sum_{i=1}^{N} s_i(z_1)s_i^*(\bar{z}_2) \qquad [7]$$

and proceed as before.

In these terms, the simplest way to describe the measure for our first example is that it follows from the inner product on the unit circle,

$$(f, g) = \oint_{|z|=1} \frac{\mathrm{d}z}{2\pi z} f^*(z) g(z)$$

Thus, we might suspect that this has something to do with the concentration of eqn [3] on the unit circle. Indeed, this idea is made precise and generalized in Shiffman and Zelditch (2003).

Our second example belongs to a class of problems in which $\mathcal{M}$ is compact and $\mathcal{L}$ positive. In this case, the space $H^0(\mathcal{M}, \mathcal{L})$ of holomorphic sections is finite dimensional, so we can take the basis to consist of all sections. Then, if $\mathcal{M}$ is in addition Kähler, we can derive all the other data from a choice of Hermitian metric $h(f, g)$ on $\mathcal{L}$. In particular, this determines a Kähler form $\omega$ as the curvature of the metric compatible connection, and thus a volume form $\mathrm{Vol}_\omega = \omega^n/n!$. We then define the inner product to be

$$(f, g) = \int_{\mathcal{M}} \mathrm{Vol}_\omega \, h(f, g)$$

Thus, the measure equation [1] and the final distribution equation [2] are entirely determined by $h$. In

these terms, the underlying reason for the simplicity of eqn [5] is that we started with the SU(2)-invariant metric $h$, so the final distribution must be invariant as well. More generally, eqn [7] is a Szegö kernel. Taking $\mathcal{L} = \mathcal{L}_1^{\otimes N}$ for $N$ large, this has a known asymptotic expansion, enabling a rather complete treatment (Zelditch 2001).

Our two examples also make the larger point that a wide variety of distributions are possible. Thus, to get convincing results, we must put in some information about the ensemble of random polynomials or sections which appear in the problem at hand.

The basic computation we just discussed can be vastly generalized to multiple variables, multipoint correlation functions, many different ensembles, and different counting problems. We will discuss the distribution of critical points of holomorphic sections below.

## The Attractor Problem

We now turn to our physical problems. Both are posed in the context of compactification of the type IIb superstring theory on a Calabi–Yau 3-fold $M$. This leads to a four-dimensional effective field theory with $N = 2$ supersymmetry, determined by the geometry of $M$.

Let us begin by stating the attractor problem mathematically, and afterwards give its physical background. We begin by reviewing a bit of the theory of Calabi–Yau manifolds. By Yau's proof of the Calabi conjecture, the moduli space of Ricci-flat metrics on $M$ is determined by a choice of complex structure on $M$, denote this $J$, and a choice of Kähler class. Using deformation theory, it can be shown that the moduli space of complex structures, denote this $\mathcal{M}_c(M)$, is locally a complex manifold of dimension $h^{2,1}(M)$. A point $J$ in $\mathcal{M}_c(M)$ picks out a holomorphic 3-form $\Omega_J \in H^{3,0}(M, \mathbb{C})$, unique up to an overall choice of normalization. The converse is also true; this can be made precise by defining the period map $\mathcal{M}_c(M) \to \mathbb{P}(H^3(M, \mathbb{Z}) \otimes \mathbb{C})$ to be the class of $\Omega$ in $H^3(M, \mathbb{Z}) \otimes \mathbb{C}$ up to projective equivalence. One can prove that the period map is injective (the Torelli theorem), locally in general and globally in certain cases such as the quintic in $\mathbb{CP}^4$.

Now, the data for the attractor problem is a charge, a class $\gamma \in H^3(M, \mathbb{Z})$. An attractor point for $\gamma$ is then a complex structure $J$ on $M$ such that

$$\gamma \in H_J^{3,0}(M, \mathbb{C}) \oplus H_J^{0,3}(M, \mathbb{C}) \qquad [8]$$

This amounts to $h^{2,1}$ complex conditions on the $h^{2,1}$ complex structure moduli, so picks out isolated points in $\mathcal{M}_c(M)$, the attractor points.

There are many mathematical and physical questions one can ask about attractor points, and it would be very interesting to have a general method to find them. As emphasized by G Moore, this is one of the simplest problems arising from string theory in which integrality (here due to charge quantization) plays a central role, and thus it provides a natural point of contact between string theory and number theory. For example, one might suspect that attractor Calabi–Yau's are arithmetic, that is, are projective varieties whose defining equations live in an algebraic number field. This can be shown to always be true for $K3 \times T^2$, and there are conjectures about when this is true more generally (Moore 2004).

A simpler problem is to characterize the distribution of attractor points in $\mathcal{M}_c(M)$. As these are infinite in number, one must introduce some control parameter. While the first idea which might come to mind is to bound the magnitude of $\gamma$, since the intersection form on $H^3(M, \mathbb{Z})$ is antisymmetric, there is no natural way to do this. A better way to get a finite set is to bound the period of $\gamma$, and consider the attractor points satisfying

$$Z_{\max}^2 \geq |Z(\gamma; z)|^2 \equiv \frac{|\int_M \gamma \wedge \Omega|^2}{\int_M \Omega \wedge \bar{\Omega}} \qquad [9]$$

As an example of the type of result we will discuss below, one can show that for large $Z_{\max}$, the density of such attractor points asymptotically approaches the Weil–Peterson volume form on $\mathcal{M}_c$.

We now briefly review the origins of this problem, in the physics of 1/2 BPS (Bogomoln'yi–Prasad–Sommerfield) black holes in $N = 2$ supergravity. We begin by introducing local complex coordinates $z^i$ on $\mathcal{M}_c(M)$. Physically, these can be thought of as massless complex scalar fields. These sit in vector multiplets of $N = 2$ supersymmetry, so there must be $h^{2,1}(M)$ vector potentials to serve as their bosonic partners under supersymmetry. These appear because the massless modes of the type IIb string include various higher rank-$p$ form gauge potentials, in particular a self-dual 4-form which we denote $C$. Self-duality means that $dC = * dC$ up to nonlinear terms, where $*$ is the Hodge star operator in ten dimensions. Now, Kaluza–Klein reduction of this 4-form potential produces $b^3(M)$ 1-form vector potentials $A_I$ in four dimensions. Given an explicit basis of 3-forms $\omega_I$ for $H^3(M, \mathbb{R}) \cap H^3(M, \mathbb{Z})$, this follows from the decomposition

$$C = \sum_{I=1}^{b_3} A_I \wedge \omega_I + \text{massive modes}$$

However, because of the self-duality relation, only half of these vector potentials are independent; the other half are determined in terms of them by four-dimensional electric–magnetic duality. Explicitly, given the intersection form $\eta_{ij}$ on $H^3 \otimes H^3$, we have

$$\mathrm{d}A_i = \eta_{ij} *_4 \mathrm{d}A_j \qquad [10]$$

where $*_4$ denotes the Hodge star in $d=4$. Thus we have $h^{2,1}+1$ independent vector potentials. One of these sits in the $N=2$ supergravity multiplet, and the rest are the correct number to pair with the complex structure moduli. We now consider 1/2 BPS black hole solutions of this four-dimensional $N=2$ theory. Choosing any $S^2$ which surrounds the horizon, we can define the charge $\gamma$ as the class in $H^3(M,\mathbb{Z})$ which reproduces the corresponding magnetic charges

$$Q_i = \frac{1}{2\pi}\int_{S^2} \mathrm{d}A_i \equiv \int_M \omega_i \wedge \gamma$$

Using eqn [10], this includes all charges.

One can show that the mass $M$ of any charged object in supergravity satisfies a BPS bound,

$$M^2 \geq |Z(\gamma;z)|^2 \qquad [11]$$

The quantity $|Z(\gamma;z)|^2$, defined in eqn [9], depends explicitly on $\gamma$, and implicitly on the complex structure moduli $z$ through $\Omega$. A 1/2 BPS solution by definition saturates this bound.

We now explain the "attractor paradox." According to Bekenstein and Hawking, the entropy of any black hole is proportional to the area of its event horizon. This area can be found by finding the black hole as an explicit solution of four-dimensional supergravity, which clearly depends on the charge $\gamma$. In fact, we must fix boundary conditions for all the fields at infinity, in particular the complex structure moduli, to get a particular black hole solution. Now, normally varying the boundary conditions varies all the data of a solution in a continuous way. On the other hand, if the entropy has any microscopic interpretation as the logarithm of the number of quantum states of the black hole, one would expect $e^S$ to be integrally quantized. Thus, it must remain fixed as the boundary conditions on complex structure moduli are varied, in contradiction with naive expectations for the area of the horizon, and seemingly contradicting Bekenstein and Hawking.

The resolution of this paradox is the attractor mechanism. Let us work in coordinates for which the four-dimensional metric takes the form

$$\mathrm{d}s^2 = -f(r)\,\mathrm{d}t^2 + \mathrm{d}r^2 + \frac{A(r)}{4\pi}\,\mathrm{d}\Omega^2_{S^2}$$

With some work, one can see that in the 1/2 BPS case, the equations of motion imply that as $r$ decreases, the complex structure moduli $z$ follow gradient flow with respect to $|Z(\gamma,z)|^2$ in eqn [11], and the area $A(r)$ of an $S^2$ at radius $r$ decreases. Finally, at the horizon, $z$ reaches a value $z_*$ at which $|Z(\gamma,z_*)|^2$ is a local minimum, and the area of the event horizon is $A=4\pi|Z(\gamma,z_*)|^2$. Since $z_*$ is determined by minimization, this area will not change under small variations of the initial $z$, resolving the paradox.

A little algebra shows that the problem of finding nonzero critical points of $|Z(\gamma,z)|^2$ is equivalent to that of finding critical points $D_iZ=0$ of the period associated to $\gamma$,

$$Z = \int_M \gamma \wedge \Omega \qquad [12]$$

usually called the central charge, with respect to the covariant derivative

$$D_i Z = \partial_i Z + (\partial_i K) Z \qquad [13]$$

Here

$$\mathrm{e}^{-K} \equiv \int \Omega \wedge \bar{\Omega} \qquad [14]$$

The mathematical significance of this rephrasing is that $K$ is a Kähler potential for the Weil–Peterson Kähler metric on $\mathcal{M}_c(M)$, with Kähler form $\omega = \partial\bar{\partial}K$, and eqn [13] is the unique connection on $H^{(3,0)}(M,\mathbb{C})$ regarded as a line bundle over $\mathcal{M}_c(M)$, whose curvature is $-\omega$. These facts can be used to show that $D_i\Omega$ provides a basis for $H^{(2,1)}(M,\mathbb{C})$, so that the critical point condition forces the projection of $\gamma$ on $H^{(2,1)}$ to vanish. This justifies our original definition eqn [8].

## Flux Vacua in IIb String Theory

We will not describe our second problem in as much detail, but just give the analogous final formulation. In this problem, a "choice of flux" is a pair of elements of $H^3(M,\mathbb{Z})$, or equivalently a single element

$$F \in H^3(M, \mathbb{Z} \oplus \tau\mathbb{Z}) \qquad [15]$$

where $\tau \in \mathcal{H} \equiv \{\tau \in \mathbb{C} | \mathrm{Im}\tau > 0\}$ is the so-called "dilaton-axion."

A flux vacuum is then a choice of complex structure $J$ and $\tau$ for which

$$F \in H^{3,0}_J(M,\mathbb{C}) \oplus H^{1,2}_J(M,\mathbb{C}) \qquad [16]$$

Now we have $h^{2,1}+h^{0,3}=h^{2,1}+1$ complex conditions on the joint choice of $h^{2,1}$ complex structure

moduli and $\tau$, so this condition also picks out special points, now in $\mathcal{M}_c \times \mathcal{H}$.

The critical point formulation of this problem is that of finding critical points of

$$W = \int \Omega \wedge F \qquad [17]$$

under the covariant derivatives eqn [13] and

$$D_\tau W = \partial_\tau W + (\partial_\tau W) Z$$

with $K$ the sum of eqn [14] and the Kähler potential $-\log \mathrm{Im}\tau$ for the metric on the upper half-plane of constant curvature –1.

This is a sort of complexified version of the previous problem and arises naturally in IIb compactification by postulating a nonzero value $F$ for a certain 3-form gauge field strength, the flux. The quantity eqn [17] is the superpotential of the resulting $N=1$ supergravity theory, and it is a standard fact in this context that supersymmetric vacua (critical points of the effective potential) are critical points of $W$ in the sense we just stated.

We can again pose the question of finding the distribution of flux vacua in $\mathcal{M}_c(M) \times \mathcal{H}$. Besides $|W|^2$, which physically is one of the contributions to the vacuum energy, we can also use the "length of the flux"

$$L = \frac{1}{\mathrm{Im}\tau} \int \mathrm{Re}\ F \wedge \mathrm{Im}\ F \qquad [18]$$

as a control parameter, and count flux vacua for which $L \leq L_{\max}$. In fact, this parameter arises naturally in the actual IIb problem, as the "orientifold three-plane charge."

What makes this problem particularly interesting physically is that it (and its analogs in other string theories) may bear on the solution of the cosmological constant problem. This begins with Einstein's famous observation that the equations of general relativity admit a one-parameter generalization,

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R = 8\pi T_{\mu\nu} + \Lambda g_{\mu\nu}$$

Physically, the cosmological constant $\Lambda$ is the vacuum energy, which in our flux problem takes the form $\Lambda = \cdots - 3|W|^2$ (the other terms are inessential for us here).

Cosmological observations tell us that $\Lambda$ is very small, of the same order as the energy of matter in the present era, about $10^{-122} M^4_{\mathrm{Planck}}$ in Planck units. However, in a generic theory of quantum gravity, including string theory, quantum effects are expected to produce a large vacuum energy, *a priori* of order $M^4_{\mathrm{Planck}}$. Finding an explanation for why the theory of our universe is in this sense nongeneric is the cosmological constant problem.

One of the standard solutions of this problem is the "anthropic solution," initiated in work of Weinberg and others, and discussed in string theory in Bousso and Polchinski (2000). Suppose that we are discussing a theory with a large number of vacuum states, all of which are otherwise candidates to describe our universe, but which differ in $\Lambda$. If the number of these vacuum states were sufficiently large, the claim that a few of these states realize a small $\Lambda$ would not be surprising. But one might still feel a need to explain why our universe is a vacuum with small $\Lambda$, and not one of the multitude with large $\Lambda$.

The anthropic argument is that, according to accepted models for early cosmology, if the value of $|\Lambda|$ were even 100 times larger than what is observed, galaxies and stars could not form. Thus, the known laws of physics guarantee that we will observe a universe with $\Lambda$ within this bound; it is irrelevant whether other possible vacuum states "exist" in any sense.

While such anthropic arguments are controversial, one can avoid them in this case by simply asking whether or not any vacuum state fits the observed value of $\Lambda$. Given a precise definition of vacuum state, this is a question of mathematics. Still, answering it for any given vacuum state is extremely difficult, as it would require computing $\Lambda$ to $10^{-122}$ precision. But it is not out of reach to argue that out of a large number of vacua, some of them are expected to realize small $\Lambda$. For example, if we could show that the number of otherwise physically acceptable vacua was larger than $10^{122}$, and that the distribution of $\Lambda$ among these was approximately uniform over the range $(-M^4_{\mathrm{Planck}}, M^4_{\mathrm{Planck}})$, we would have made a good case for this expectation. This style of reasoning can be vastly generalized and, given favorable assumptions about the number of vacua in a theory, could lead to falsifiable predictions independent of any *a priori* assumptions about the choice of vacuum state (Douglas 2003).

## Asymptotic Counting Formulas

We have just defined two classes of physically preferred points in the complex structure moduli space of Calabi–Yau 3-folds, the attractor points and the flux vacua. Both have simple definitions in terms of Hodge structure, eqn [8] and eqn [16], and both are also critical points of integral periods of the holomorphic 3-form.

This second phrasing of the problem suggests the following language. We define a random period of the holomorphic 3-form to be the period for a randomly chosen cycle in $H_3(M, \mathbb{Z})$ of the types we

just discussed (real or complex, and with the appropriate control parameters). We are then interested in the expected distribution of critical points for a random period. This brings our problem into the framework of random algebraic geometry. Before proceeding to use this framework, let us first point out some differences with the toy problems we discussed. First, while eqn [12] and eqn [17] are sums of the form eqn [6], we take not an orthonormal basis but instead a basis $s_i$ of integral periods of $\Omega$. Second, the coefficients $c_i$ are not normally distributed but instead drawn from a discrete uniform distribution, that is, correspond to a choice of $\gamma$ in $H^3(M, \mathbb{Z})$ or $F$ as in eqn [15], satisfying the bounds on $|Z|$ or $L$. Finally, we do not normalize the distribution (which is thus not a probability measure) but instead take each choice with unit weight.

These choices can of course be modified, but are made in order to answer the question, "how many attractor points (or flux vacua) sit within a specified region of moduli space?" The answer we will get is a density $\mu(Z_{\max})$ or $\mu(L_{\max})$ on moduli space, such that as the control parameter becomes large, the number of critical points within a region $R$ asymptotes to

$$\mathcal{N}(R; Z_{\max}) \sim \int_R \mu(Z_{\max})$$

The key observation is that to get such asymptotics, we can start with a Gaussian random element $s$ of $H^3(M, \mathbb{R})$ or $H^3(M, \mathbb{C})$. In other words, we neglect the integral quantization of the charge or flux. Intuitively, this might be expected to make little difference in the limit that the charge or flux is large, and in fact one can prove that this simplification reproduces the leading large $L$ or $|Z|$ asymptotics for the density of critical points, using standard ideas in lattice point counting.

This justifies starting with a two-point function like eqn [7]. While the integral periods $s_i$ of $\Omega$ can be computed in principle (and have been in many examples) by solving a system of linear PDEs, the Picard–Fuchs equations, it turns out that one does not need such detailed results. Rather, one can use the following ansatz for the two-point function,

$$\begin{aligned} G(z_1, \bar{z}_2) &= \sum_{I=1}^{b_3} \eta^{IJ} s_I(z_1) s_J^*(\bar{z}_2) \\ &= \int_M \Omega(z_1) \wedge \bar{\Omega}(\bar{z}_2) \\ &= \exp -K(z_1, \bar{z}_2) \end{aligned}$$

In words, the two-point function is the formal continuation of the Kähler potential on $\mathcal{M}_c(M)$ to independent holomorphic and antiholomorphic variables. This incorporates the quadratic form appearing in eqn [18] and can be used to count sections with such a bound.

We can now follow the same strategy as before, by introducing an expected density of critical points,

$$d\mu(z) = \mathbb{E}[\delta^{(n)}(D_i s(z)) \delta^{(n)}(\bar{D}_{\bar{i}} \bar{s}(\bar{z})) \mid \det_{1 \le i,j \le 2n} H_{ij}|] \quad [19]$$

where the "complex Hessian" $H$ is the $2n \times 2n$ matrix of second derivatives

$$H \equiv \begin{pmatrix} \partial_i \bar{D}_{\bar{j}} \bar{s}(z) & \partial_i D_j s(z) \\ \bar{\partial}_{\bar{i}} \bar{D}_{\bar{j}} \bar{s}(z) & \bar{\partial}_{\bar{i}} D_j s(z) \end{pmatrix} \quad [20]$$

(note that $\partial Ds = DDs$ at a critical point). One can then compute this density along the same lines. The holomorphy of $s$ implies that $\partial_i \bar{D}_{\bar{j}} s = \omega_{i\bar{j}} s$, which is one simplification. Other geometric simplifications follow from the fact that eqn [19] depends only on $s$ and a finite number of its derivatives at the point $z$.

For the attractor problem, using the identity

$$D_i D_j s = \mathcal{F}_{ijk} \omega^{k\bar{k}} \bar{D}_{\bar{k}} s = 0$$

from special geometry of Calabi–Yau 3-folds, the Hessian becomes trivial, and $\det H = |s|^{2n}$. One thus finds (Denef and Douglas 2004) that the asymptotic density of attractor points with large $|Z| \le Z_{\max}$ in a region $R$ is

$$\mathcal{N}(R, |Z| \le Z_{\max}) \sim \frac{2^{n+1}}{(n+1)\pi^n} Z_{\max}^{n+1} \cdot \text{vol}(R)$$

where $\text{vol}(R) = \int_R \omega^n / n!$ is the volume of $R$ in the Weil–Peterson metric. The total volume is known to be finite for Calabi–Yau 3-fold moduli spaces, and thus so is the number of attractor points under this bound.

The flux vacuum problem is complicated by the fact that $DDs$ is nonzero and thus the determinant of the Hessian does not take a definite sign, and implementing the absolute value in eqn [19] is nontrivial. The result (Douglas, *et al.* 2004) is

$$\mu(z) \sim \frac{1}{b_3! \sqrt{\det \Lambda(z)}} \int_{\mathcal{H}(z) \times \mathbb{C}} |\det(HH^* - |x|^2 \cdot \mathbf{1})| \times e^{H^t \Lambda(z)^{-1} H - |x|^2} dH\, dx$$

where $\mathcal{H}(z)$ is the subspace of Hessian matrices eqn [20] obtainable from periods at the point $z$, and $\Lambda(z)$ is a covariance matrix computable from the period data.

A simpler lower bound for the number of solutions can be obtained by instead computing the index density

$$\mu_I(z) = \mathbb{E}\left[\delta^{(n)}(D_i s)\delta^{(n)}(\bar{D}_i \bar{s}) \det_{1\le i,j\le 2n} H_{ij}\right] \qquad [21]$$

so-called because it weighs the vacua with a Morse–Witten sign factor. This admits a simple explicit formula (Ashok and Douglas 2004),

$$I_{\text{vac}}(R, L \le L_{\max}) \sim \frac{(2\pi L_{\max})^{b_3}}{\pi^{n+1} b_3!} \int_R \det(\mathcal{R} + \omega \cdot 1) \qquad [22]$$

where $\mathcal{R}$ is the $(n+1)\times(n+1)$-dimensional matrix of curvature 2-forms for the Weil–Peterson metric.

One might have guessed this density by the following reasoning. If $s$ had been a single-valued section on a compact $\mathcal{M}_c$ (it is not), topological arguments determine the total index to be $[c_{n+1}(\mathcal{L}\otimes T^*\mathcal{M})]$, and this is the simplest density constructed solely from the metric and curvatures in the same cohomology class.

It is not in general known whether this integral over Calabi–Yau moduli space is finite, though this is true in examples studied so far. One can also control $|W|^2$ as well as other observables, and one finds that the distribution of $|W|^2$ among flux vacua is to a good approximation uniform. Considering explicit examples, the prefactor in eqn [22] is of order $10^{100}$–$10^{300}$, so assuming that this factor dominates the integral, we have justified the Bousso–Polchinski solution to the cosmological constant problem in these models.

The finite $L$ corrections to these formulas can be estimated using van der Corput techniques, and are suppressed by better than the naive $L^{-1/2}$ or $|Z|^{-1}$ one might have expected. However the asymptotic formulas for the numbers of flux vacuum break down in certain limits of moduli space, such as the large complex structure limit. This is because eqn [18] is an indefinite quadratic form, and the fact that it bounds the number of solutions at all is somewhat subtle. These points are discussed at length in (Douglas *et al.* 2005).

Similar results have been obtained for a wide variety of flux vacuum counting problems, with constraints on the value of the effective potential at the minimum, on the masses of scalar fields, on scales of supersymmetry breaking, and so on. And in principle, this is just the tip of an iceberg, as the study of more or less any class of superstring vacua leads to similar questions of counting and distribution, less well understood at present. Some of these are discussed in Douglas (2003), Acharya *et al.* (2005), Denef and Douglas (2005), Blumenhagen *et al.* (2005).

*See also:* Black Hole Mechanics; Chaos and Attractors; Compactification of Superstring Theory; Supergravity.

## Further Reading


Acharya BS, Denef F, and Valandro R (2005) Statistics of M theory vacua. *Journal of High Energy Physics* 0506: 056.

Ashok S and Douglas MR (2004) Counting flux vacua. *Journal of High Energy Physics* 0401: 060.

Blumenhagen R, Gmeiner F, Honecker G, Lust D, and Weigand T (2005) The statistics of supersymmetric D-brane models. *Nuclear Physics B* 713: 83.

Bousso R and Polchinski J (2000) Quantization of four-form fluxes and dynamical neutralization of the cosmological constant. *Journal of High Energy Physics* 06: 06.

Cox DA and Katz S (1999) *Mirror Symmetry and Algebraic Geometry*. Providence, RI: American Mathematical Society.

Denef F and Douglas MR (2004) Distributions of flux vacua. *Journal of High Energy Physics* 0405: 072.

Denef F and Douglas MR (2005) Distributions of nonsupersymmetric flux vacua. *Journal of High Energy Physics* 0503: 061.

Douglas MR (2003) The statistics of string/M theory vacua. *Journal of High Energy Physics* 5: 046.

Douglas MR, Shiffman B, and Zelditch S (2004) Critical points and supersymmetric vacua I. *Communications in Mathematical Physics* 252(1–3): 325–358; II: Asymptotics (to appear).

Douglas MR, Shiffman B, and Zelditch S (2005) Critical points and supersymmetric vacua, III: String/M models. Submitted to *Communications in Mathematical Physics*. http://arxiv./org/abs/math-ph/0506015.

Edelman A and Kostlan E (1995) How many zeros of a random polynomial are real? *Bull. Amer. Math. Soc. (N.S.)* 32: 1–37.

Ferrara S, Kallosh R, and Strominger A (1995) $N=2$ extremal black holes, *Physical Review D* 52: 5412.

Giddings SB, Kachru S, and Polchinski J (2002) Hierarchies from fluxes in string compactifications. *Physical Review D (3)* 66(10): 106006.

Gross M, Huybrechts D, and Joyce D (2003) *Calabi–Yau Manifolds and Related Geometries*. New York: Springer.

Moore GW (2004) Les Houches lectures on strings and arithmetic, arXiv:hep-th/0401049.

Shiffman B and Zelditch S (2003) Equilibrium distribution of zeros of random polynomials. *Int. Math. Res. Not.* 1: 25–49.

Zelditch S (2001) From random polynomials to symplectic geometry. In: *XIIIth International Congress of Mathematical Physics*, pp. 367–376. Beijing: Higher Education Press.

Zelditch S (2005) Random complex geometry and vacua, or: How to count universes in string/M theory, 2005 preprint.

# Random Dynamical Systems

**V Araújo**, Universidade do Porto, Porto, Portugal



## Introduction

The concept of random dynamical system is a comparatively recent development combining ideas and methods from the well-developed areas of probability theory and dynamical systems.

Let us consider a mathematical model of some physical process given by the iterates $T_0^k = T_0 \circ \overset{k}{\cdots} \circ T_0$, $k \geq 1$, of a smooth transformation $T_0 : M \circlearrowleft$ of a manifold into itself. A realization of the process with initial condition $x_0$ is modeled by the sequence $(T_0^k(x_0))_{k\geq 1}$, the orbit of $x_0$.

Due to our inaccurate knowledge of the particular physical system or due to computational or theoretical limitations (e.g., lack of sufficient computational power, inefficient algorithms, or insufficiently developed mathematical or physical theory), the mathematical models never correspond exactly to the phenomenon they are meant to model. Moreover, when considering practical systems, we cannot avoid either external noise or measurement or inaccuracy errors, so every realistic mathematical model should allow for small errors along orbits not to disturb the long-term behavior too much. To be able to cope with unavoidable uncertainty about the "correct" parameter values, observed initial states and even the specific mathematical formulation involved, let randomness be embedded within the model to begin with.

This article presents the most basic classes of models, defines the general concept, and presents some developments and examples of applications.

## Dynamics with Noise

To model random perturbations of a transformation $T_0$, we may consider a transition from the image $T_0(x)$ to some point according to a given probability law, obtaining a Markov chain, or, if $T_0$ depends on a parameter $p$, we may choose $p$ at random at each iteration, which also can be seen as a Markov chain but whose transitions are strongly correlated.

### Random Noise

Given $T_0 : M \circlearrowleft$ and a family $\{p(\cdot \mid x) : x \in M\}$ of probability measures on $M$ such that the support of $p(\cdot \mid x)$ is close to $T_0(x)$, the random orbits are sequences $(x_k)_{k\geq 1}$ where each $x_{k+1}$ is a random variable with law $p(\cdot \mid x_k)$. This is a Markov chain with state space $M$ and transition probabilities $\{p(\cdot \mid x)\}_{x\in M}$. To extend the concept of invariant measure of a transformation to this setting, a probability measure $\mu$ is said to be "stationary" if $\mu(A) = \int p(A \mid x)\, d\mu(x)$ for every measurable (Borel) subset $A$. This can be conveniently translated by saying that the skew-product measure $\mu \times p^{\mathbb{N}}$ on $M \times M^{\mathbb{N}}$ given by

$$\begin{aligned} &d(\mu \times p^{\mathbb{N}})(x_0, x_1, \ldots, x_n, \ldots) \\ &\quad = d\mu(x_0) p(dx_1 \mid x_0) \cdots p(dx_{n+1} \mid x_n) \cdots \end{aligned}$$

is invariant by the shift map $\mathcal{S} : M \times M^{\mathbb{N}} \circlearrowleft$ on the space of orbits. Hence, we may use the ergodic theorem and get that time averages of all continuous observables $\varphi : M \to \mathbb{R}$, that is, writing $\underline{x} = (x_k)_{k\geq 0}$ and

$$\begin{aligned} \tilde{\varphi}(\underline{x}) &= \lim_{n\to+\infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi(x_k) \\ &= \lim_{n\to+\infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi(\pi_0(\mathcal{S}^k(\underline{x}))) \end{aligned}$$

exist for $\mu \times p^{\mathbb{N}}$-almost all sequences $\underline{x}$, where $\pi_0 : M \times M^{\mathbb{N}} \to M$ is the natural projection on the first coordinate. It is well known that stationary measures always exist if the transition probabilities $p(\cdot \mid x)$ depend continuously on $x$.

A function $\varphi : M \to \mathbb{R}$ is invariant if $\varphi(x) = \int \varphi(z) p(dz \mid x)$ for $\mu$-almost every $x$. We then say that $\mu$ is ergodic if every invariant function is constant $\mu$-almost everywhere. Using the ergodic theorem again, if $\mu$ is ergodic, then $\tilde{\varphi} = \int \varphi\, d\mu$, $\mu$-almost everywhere.

Stationary measures are the building blocks for more sophisticated analysis involving, for example, asymptotic sojourn times, Lyapunov exponents, decay of correlations, entropy and/or dimensions, exit/entrance times from/to subsets of $M$, to name just a few frequent notions of dynamical and probabilistic/statistical nature.

**Example 1** (Random jumps). Given $\epsilon > 0$ and $T_0 : M \to M$, let us define

$$p^{\epsilon}(A \mid x) = \frac{m(A \cap B(T_0(x), \epsilon))}{m(B(T_0(x), \epsilon))}$$

where $m$ denotes some choice of Riemannian volume form on $M$. Then $p^{\epsilon}; (\cdot \mid x)$ is the normalized volume restricted to the $\epsilon$-neighborhood of $T_0(x)$.

This defines a family of transition probabilities allowing the points to "jump" from $T_0(x)$ to any point in the $\epsilon$-neighborhood of $T_0(x)$ following a uniform distribution law.

### Random Maps

Alternatively, we may choose maps $T_1, T_2, \ldots, T_k$ independently at random near $T_0$ according to a probability law $\nu$ on the space $T(M)$ of maps, whose support is close to $T_0$ in some topology, and consider sequences $x_k = T_k \circ \cdots \circ T_1(x_0)$ obtained through random iteration, $k \geq 1$, $x_0 \in M$.

This is again a Markov chain whose transition probabilities are given for any $x \in M$ by

$$p(A \mid x) = \nu(\{T \in T(M)\colon T(x) \in A\})$$

so this model may be reduced to the first one. However, in the random-maps setting, we may associate, with each random orbit, a sequence of maps which are iterated, enabling us to use "robust properties" of the transformation $T_0$ (i.e., properties which are known to hold for $T_0$ and for every nearby map $T$) to derive properties of the random orbits.

Under some regularity conditions on the map $x \mapsto p(A \mid x)$ for every Borel subset $A$, it is possible to represent random noise by random maps on suitably chosen spaces of transformations. In fact, the transition probability measures obtained in the random-maps setting exhibit strong spatial correlation: $p(\cdot \mid x)$ is close to $p(\cdot \mid y)$ as $x$ is near $y$.

If we have a parametrized family $T\colon \mathcal{U} \times M \to M$ of maps, we can specify the law $\nu$ by giving a probability $\theta$ on $\mathcal{U}$. Then with every sequence $T_1, \ldots, T_k, \ldots$ of maps of the given family, we associate a sequence $\omega_1, \ldots, \omega_k, \ldots$ of parameters in $\mathcal{U}$ since

$$T_k \circ \cdots \circ T_1 = T_{\omega_k} \circ \cdots \circ T_{\omega_1} = T^k_{\omega_1, \ldots, \omega_k}$$

for all $k \geq 1$, where we write $T_\omega(x) = T(\omega, x)$. In this setting, the shift map $\mathcal{S}$ becomes a skew-product transformation

$$\mathcal{S} : M \times \mathcal{U}^{\mathbb{N}} \circlearrowleft \quad (x, \underline{\omega}) \mapsto (T_{\omega_1}(x), \sigma(\underline{\omega}))$$

to which many of the standard methods of dynamical systems and ergodic theory can be applied, yielding stronger results that can be interpreted in random terms.

**Example 2** (Parametric noise). Let $T : P \times M \to M$ be a smooth map where $P$, $M$ are finite-dimensional Riemannian manifolds. We fix $p_0 \in P$, denote by $m$ some choice of Riemannian volume form on $P$, set $T_w(x) = T(w, x)$, and for every $\epsilon > 0$ write $\theta_\epsilon = (m(B(p_0, \epsilon))^{-1} \cdot (m \mid B(p_0, \epsilon))$, the normalized restriction of $m$ to the $\epsilon$-neighborhood of $p_0$. Then $(T_w)_{w \in P}$, together with $\theta_\epsilon$, defines a random perturbation of $T_{p_0}$, for every small enough $\epsilon > 0$.

**Example 3** (Global additive perturbations). Let $M$ be a homogeneous space, that is, a compact connected Lie group admitting an invariant Riemannian metric. Fixing a neighborhood $\mathcal{U}$ of the identity $e \in M$, we can define a map $T : \mathcal{U} \times M \to M$, $(u, x) \mapsto L_u(T_0(x))$, where $L_u(x) = u \cdot x$ is the left translation associated with $u \in M$. The invariance of the metric means that left (and also right) translations are isometries, hence fixing $u \in \mathcal{U}$ and taking any $(x, v) \in TM$, we get

$$\begin{aligned} \|DT_u(x) \cdot v\| &= \|DL_u(T_0(x))(DT_0(x) \cdot v)\| \\ &= \|DT_0(x) \cdot v\| \end{aligned}$$

In the particular case of $M = \mathbb{T}^d$, the $d$-dimensional torus, we have $T_u(x) = T_0(x) + u$, and this simplest case suggests the name "additive random perturbations" for random perturbations defined using families of maps of this type.

For the probability measure on $\mathcal{U}$, we may take $\theta_\epsilon$, any probability measure supported in the $\epsilon$-neighborhood of $e$ and absolutely continuous with respect to the Riemannian metric on $M$, for any $\epsilon > 0$ small enough.

**Example 4** (Local additive perturbations). If $M = \mathbb{R}^d$ and $U_0$ is a bounded open subset of $M$ strictly invariant under a diffeomorphism $T_0$, that is, closure $(T_0(U_0)) \subset U_0$, then we can define an isometric random perturbation setting:

(i) $V = T_0(U_0)$ (so that closure $(V) =$ closure $(T_0(U_0)) \subset U_0$);
(ii) $G \simeq \mathbb{R}^d$ the group of translations of $\mathbb{R}^d$; and
(iii) $\mathcal{V}$ a small enough neighborhood of 0 in $G$.

Then for $v \in \mathcal{V}$ and $x \in V$, we set $T_v(x) = x + v$, with the standard notation for vector addition, and clearly $T_v$ is an isometry. For $\theta_\epsilon$, we may take any probability measure on the $\epsilon$-neighborhood of 0, supported in $\mathcal{V}$ and absolutely continuous with respect to the volume in $\mathbb{R}^d$, for every small enough $\epsilon > 0$.

### Random Perturbations of Flows

In the continuous-time case, the basic model to start with is an ordinary differential equation $dX_t = f(t, X_t)dt$, where $f\colon [0, +\infty) \to \mathcal{X}(M)$ and $\mathcal{X}(M)$ is the family of vector fields in $M$. We embed randomness in the differential equation

basically through "diffusion," the perturbation is given by white noise or Brownian motion "added" to the ordinary solution.

In this setting, assuming for simplicity that $M = \mathbb{R}^n$, the random orbits are solutions of stochastic differential equations

$$\begin{aligned} \mathrm{d}X_t &= f(t, X_t)\mathrm{d}t + \epsilon \cdot \sigma(t, X_t)\mathrm{d}W_t, \\ & 0 \le t \le T,\ X_0 = Z \end{aligned}$$

where $Z$ is a random variable, $\epsilon, T > 0$ and both $f : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : [0, T] \times \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^k, \mathbb{R}^n)$ are measurable functions. The space of linear maps $\mathbb{R}^k \to \mathbb{R}^n$ is written on $\mathcal{L}(\mathbb{R}^k, \mathbb{R}^n)$ and $W_t$ is the white-noise process on $\mathbb{R}^k$. The solution of this equation is a stochastic process:

$$X : \mathbb{R} \times \Omega \to M \quad (t, \omega) \mapsto X_t(\omega)$$

for some (abstract) probability space $\Omega$, given by

$$X_t = Z + \int_0^T f(s, X_s)\mathrm{d}s + \int_0^T \epsilon \cdot \sigma(s, X_s)\mathrm{d}W_s$$

where the last term is a stochastic integral in the sense of Itô. Under reasonable conditions on $f$ and $\sigma$, there exists a unique solution with continuous paths, that is,

$$[0, +\infty) \ni t \mapsto X_t(\omega)$$

is continuous for almost all $\omega \in \Omega$ (in general these paths are nowhere differentiable).

Setting $Z = \delta_{x_0}$, the probability measure concentrated on the point $x_0$, the initial point of the path is $x_0$ with probability 1. We write $X_t(\omega)x_0$ for paths of this type. Hence, $x \mapsto X_t(\omega)x$ defines a map $X_t(\omega) : M\circlearrowleft$ which can be shown to be a homeomorphism and even diffeomorphisms under suitable conditions on $f$ and $\sigma$. These maps satisfy a cocycle property

$$\begin{aligned} X_0(\omega) &= \mathrm{Id}_M \quad \text{(identity map of } M\text{)} \\ X_{t+s}(\omega) &= X_t(\theta(s)(\omega)) \circ X_s(\omega) \end{aligned}$$

for $s, t \ge 0$ and $\omega \in \Omega$, for a family of measure-preserving transformations $\theta(s) : (\Omega, \mathbb{P})\circlearrowleft$ on a suitably chosen probability space $(\Omega, \mathbb{P})$. This enables us to write the solution of this kind of equations also as a skew product.

### The Abstract Framework

The illustrative particular cases presented can all be written in skew-product form as follows.

Let $(\Omega, \mathbb{P})$ be a given probability space, which will be the model for the noise, and let $\mathbb{T}$ be time, which usually means $\mathbb{Z}_+, \mathbb{Z}$ (discrete, resp. invertible system) or $\mathbb{R}_+, \mathbb{R}$ (continuous, resp. invertible system). A random dynamical system is a skew product

$$\mathcal{S}_t : \Omega \times M\circlearrowleft, (\omega, x) \mapsto (\theta(t)(\omega), \varphi(t, \omega)(x))$$

for all $t \in \mathbb{T}$, where $\theta : \mathbb{T} \times \Omega \to \Omega$ is a family of measure-preserving maps $\theta(t) : (\Omega, \mathbb{P})\circlearrowleft$ and $\varphi : \mathbb{T} \times \Omega \times M \to M$ is a family of maps $\varphi(t, \omega) : M\circlearrowleft$ satisfying the cocycle property: for $s, t \in \mathbb{T}$, $\omega \in \Omega$,

$$\begin{aligned} \varphi(0, \omega) &= \mathrm{Id}_M \\ \varphi(t + s, \omega) &= \varphi(t, \theta(s)(\omega)) \circ \varphi(s, \omega) \end{aligned}$$

In this general setting an invariant measure for the random dynamical system is any probability measure $\mu$ on $\Omega \times M$ which is $\mathcal{S}_t$-invariant for all $t \in \mathbb{T}$ and whose marginal is $\mathbb{P}$, that is, $\mu(\mathcal{S}_t^{-1}(U)) = \mu(U)$ and $\mu(\pi_\Omega^{-1}(U)) = \mathbb{P}(U)$ for every measurable $U \subset \Omega \times M$, respectively, with $\pi_\Omega : \Omega \times M \to \Omega$ the natural projection.

**Example 5** In the setting of the previous examples of random perturbations of maps, the product measure $\eta = \mathbb{P} \times \mu$ on $\Omega \times M$, with $\Omega = \mathcal{U}^{\mathbb{N}}$, $\mathbb{P} = \theta_\epsilon^{\mathbb{N}}$ and $\mu$ any stationary measure, is clearly invariant. However, not all invariant measures are product measures of this type.

Naturally an invariant measure is ergodic if every $\mathcal{S}_t$-invariant function is $\mu$-almost everywhere constant. That is, if $\psi : \Omega \times M \to \mathbb{R}$ satisfies $\psi \circ \mathcal{S}_t = \psi$ $\mu$-almost everywhere for every $t \in \mathbb{T}$, then $\psi$ is $\mu$-almost everywhere constant.

## Applications

The well-established applications of both probability or stochastic differential equations (solution of boundary value problems, optimal stopping, stochastic control etc.) and dynamical systems (all kinds of models of physical, economic or biological phenomena, solutions of differential equations, control systems etc.) will not be presented here. Instead, this section focuses on topics where the subject sheds new light on these areas.

### Products of Random Matrices and the Multiplicative Ergodic Theorem

The following celebrated result on products of random matrices has far-reaching applications on dynamical systems theory.

Let $(X_n)_{n\ge0}$ be a sequence of independent and identically distributed random variables on the probability space $(\Omega, P)$ with values in $\mathcal{L}(\mathbb{R}^k, \mathbb{R}^k)$ such that $E(\log^+ \|X_1\|) < +\infty$, where $\log^+ x = \max\{0, \log x\}$ and $\|\cdot\|$ is a given norm on

$\mathcal{L}(\mathbb{R}^k, \mathbb{R}^k)$. Writing $\varphi_n(\omega) = X_n(\omega) \circ \cdots \circ X_1(\omega)$ for all $n \geq 1$ and $\omega \in \Omega$ we obtain a cocycle. If we set

$$B = \left\{ (\omega, y) \in \Omega \times \mathbb{R}^k : \lim_{n \to +\infty} \frac{1}{n} \log \|\varphi_n(\omega) y\| \right.$$
$$\left. \text{exists and is finite or is } -\infty \right\},$$
$$\Omega' = \{\omega \in \Omega : (\omega, y) \in B \text{ for all } y \in \mathbb{R}^k\}$$

then $\Omega'$ contains a subset $\Omega''$ of full probability and there exist random variables (which might take the value $-\infty$) $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$ with the following properties.

1. Let $I = \{k+1 = i_1 > i_2 > \cdots > i_{l+1} = 1\}$ be any $(l+1)$-tuple of integers and then we define

   $$\Omega_I = \{\omega \in \Omega'' : \lambda_i(\omega) = \lambda_j(\omega), i_h > i, j \geq i_{h+1},$$
   $$\text{and } \lambda_{i_h}(\omega) > \lambda_{i_{h+1}}(\omega) \text{ for all } 1 < h < l\}$$

   the set of elements where the sequence $\lambda_i$ jumps exactly at the indexes in $I$. Then for $\omega \in \Omega_I, 1 < h \leq l$,

   $$\Sigma_{I,h}(\omega) = \left\{ y \in \mathbb{R}^k : \lim_{n \to +\infty} \frac{1}{n} \log \|\varphi_n(\omega)\| \leq \lambda_{i_h}(\omega) \right\}$$

   is a vector subspace with dimension $i_{h-1} - 1$.
2. Setting $\Sigma_{I,k+1}(\omega) = \{0\}$, then

   $$\lim_{n \to +\infty} \frac{1}{n} \log \|\varphi_n(\omega)\| = \lambda_{i_h}(\omega)$$

   for every $y \in \Sigma_{I,h}(\omega) \backslash \Sigma_{I,h+1}(\omega)$.
3. For all $\omega \in \Omega''$ there exists the matrix

   $$A(\omega) = \lim_{n \to +\infty} [(\varphi_n(\omega))^* \varphi_n(\omega)]^{1/2n}$$

   whose eigenvalues form the set $\{e^{\lambda_i} : i = 1, \ldots, k\}$.

The values of $\lambda_i$ are the random Lyapunov characteristics and the corresponding subspaces are analogous to random eigenspaces. If the sequence $(X_n)_{n \geq 0}$ is ergodic, then the Lyapunov characteristics become nonrandom constants, but the Lyapunov subspaces are still random.

We can easily deduce the multiplicative ergodic theorem for measure-preserving differentiable maps $(T_0, \mu)$ on manifolds $M$ from this result. For simplicity, we assume that $M \subset \mathbb{R}^k$ and set $p(A \mid x) = \delta_{T_0(x)}(A) = 1$ if $T_0(x) \in A$ and 0 otherwise. Then the measure $\mu \times p^{\mathbb{N}}$ on $M \times M^{\mathbb{N}}$ is $\sigma$-invariant (as defined earlier) and we have that $\pi_0 \circ \sigma = T_0 \circ \pi_0$, where $\pi_0 : M^{\mathbb{N}} \to M$ is the projection on the first coordinate, and also $(\pi_0)_*(\mu \times p^{\mathbb{N}}) = \mu$. Then, setting for $n \geq 1$

$$X : M \to \mathcal{L}(\mathbb{R}^k, \mathbb{R}^k) \quad \text{and} \quad X_n = X \circ \pi_0 \circ \sigma^n$$
$$x \mapsto DT_0(x)$$

we obtain a stationary sequence to which we can apply the previous result, obtaining the existence of Lyapunov exponents and of Lyapunov subspaces on a full measure subset for any $C^1$ measure-preserving dynamical system.

By a standard extension of the previous setup, we obtain a random version of the multiplicative ergodic theorem. We take a family of skew-product maps $\mathcal{S}_t : \Omega \times M \circlearrowleft$ as in the section "The abstract framework" with an invariant probability measure $\mu$ and such that $\varphi(t, \omega) : M \circlearrowleft$ is (for simplicity) a local diffeomorphism. We then consider the stationary family

$$X_t : \Omega \to \mathcal{L}(TM),\ \omega \mapsto D\varphi(t, \omega) : TM \circlearrowleft \quad t \in \mathbb{T}$$

where $D\varphi(t, \omega)$ is the tangent map to $\varphi(t, \omega)$. This is a cocycle since for all $t, s \in \mathbb{T}$, $\omega \in \Omega$ we have

$$X(s+t, \omega) = X(s, \theta(t)\omega) \circ X(t, \omega)$$

If we assume that

$$\sup_{0 \leq t \leq 1} \sup_{x \in M} \left( \log^+ \|D\varphi(t, \omega)(x)\| \right) \in L^1(\Omega, \mathbb{P})$$

where $\|\cdot\|$ denotes the norm on the corresponding space of linear maps given by the induced norm (from the Riemannian metric) on the appropriate tangent spaces, then we obtain a sequence of random variables (which might take the value $-\infty$) $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$, with $k$ being the dimension of $M$, such that

$$\lim_{t \to +\infty} \frac{1}{t} \log \|X_t(\omega, x) y\| = \lambda_i(\omega, x)$$

for every $y \in E_i \omega, x) = \Sigma_i(\omega, x) \setminus \Sigma_{i+1}(\omega, x)$ and $i = 1, \ldots, k+1$, where $(\Sigma_i(\omega, x))_i$ is a sequence of vector subspaces in $T_x M$ as before, measurable with respect to $(\omega, x)$. In this setting, the subspaces $E_i(\omega, x)$ and the Lyapunov exponents are invariant, that is, for all $t \in \mathbb{T}$ and $\mu$-almost every $(\omega, x) \in \Omega \times M$, we have

$$\lambda_i(\mathcal{S}_t(\omega, x)) = \lambda_i(\omega, x) \quad \text{and} \quad E_i(\mathcal{S}_t(\omega, x)) = E_i(\omega, x)$$

The dependence of Lyapunov exponents on the map $T_0$ has been a fruitful and central research program in dynamical systems for decades extending to the present day. The random multiplicative ergodic theorem sets the stage for the study of the stability of Lyapunov exponents under random perturbations.

### Stochastic Stability of Physical Measures

The development of the theory of dynamical systems has shown that models involving expressions as simple as quadratic polynomials (as the logistic family or Hénon attractor), or autonomous ordinary

differential equations with a hyperbolic singularity of saddle type, as the Lorenz flow, exhibit sensitive dependence on initial conditions, a common feature of chaotic dynamics: small initial differences are rapidly augmented as time passes, causing two trajectories originally coming from practically indistinguishable points to behave in a completely different manner after a short while. Long-term predictions based on such models are unfeasible, since it is not possible to both specify initial conditions with arbitrary accuracy and numerically calculate with arbitrary precision.

**Physical measures** Inspired by an analogous situation of unpredictability faced in the field of statistical mechanics/thermodynamics, researchers focused on the statistics of the data provided by the time averages of some observable (a continuous function on the manifold) of the system. Time averages are guaranteed to exist for a positive-volume subset of initial states (also called an observable subset) on the mathematical model if the transformation, or the flow associated with the ordinary differential equation, admits a smooth invariant measure (a density) or a physical measure.

Indeed, if $\mu_0$ is an ergodic invariant measure for the transformation $T_0$, then the ergodic theorem ensures that for every $\mu$-integrable function $\varphi : M \to \mathbb{R}$ and for $\mu$-almost every point $x$ in the manifold $M$, the time average $\tilde{\varphi}(x) = \lim_{n\to+\infty} n^{-1} \sum_{j=0}^{n-1} \varphi(T_0^j(x))$ exists and equals the space average $\int \varphi \, \mathrm{d}\mu_0$. A physical measure $\mu$ is an invariant probability measure for which it is required that time averages of every continuous function $\varphi$ exist for a positive Lebesgue measure (volume) subset of the space and be equal to the space average $\mu(\varphi)$.

We note that if $\mu$ is a density, that is, absolutely continuous with respect to the volume measure, then the ergodic theorem ensures that $\mu$ is physical. However, not every physical measure is absolutely continuous. To see why in a simple example, we consider a singularity $p$ of a vector field which is an attracting fixed point (a sink), then the Dirac mass $\delta_p$ concentrated on $p$ is a physical probability measure, since every orbit in the basin of attraction of $p$ will have asymptotic time averages for any continuous observable $\varphi$ given by $\varphi(p) = \delta_p(\varphi)$.

Physical measures need not be unique or even exist in general but, when they do exist, it is desirable that the set of points whose asymptotic time averages are described by physical measures (such a set is called the basin of the physical measures) be of full Lebesgue measure – only an exceptional set of points with zero volume would not have a well-defined asymptotic behavior. This is yet far from being proved for most dynamical systems, in spite of much recent progress in this direction.

There are robust examples of systems admitting several physical measures whose basins together are of full Lebesgue measure, where "robust" means that there are whole open sets of maps of a manifold in the $C^2$ topology exhibiting these features. For typical parametrized families of one-dimensional unimodal maps (maps of the circle or of the interval with a unique critical point), it is known that the above scenario holds true for Lebesgue almost every parameter. It is known that there are systems admitting no physical measure, but the only known cases are not robust, that is, there are systems arbitrarily close which admit physical measures.

It is hoped that conclusions drawn from models admitting physical measures to be effectively observable in the physical processes being modeled. In order to lend more weight to this expectation, researchers demand stability properties from such invariant measures.

**Stochastic stability** There are two main issues concerning a mathematical model, both from theoretical and practical standpoints. The first one is to describe the asymptotic behavior of most orbits, that is, to understand what happens to orbits when time tends to infinity. The second and equally important one is to ascertain whether the asymptotic behavior is stable under small changes of the system, that is, whether the limiting behavior is still essentially the same after small changes to the law of evolution. In fact, since models are always simplifications of the real system (we cannot ever take into account the whole state of the universe in any model), the lack of stability considerably weakens the conclusions drawn from such models, because some properties might be specific to it and not in any way resembling the real system.

Random dynamical systems come into play in this setting when we need to check whether a given model is stable under small random changes to the law of evolution.

In more precise terms, we suppose that there is a dynamical system (a transformation or a flow) admitting a physical measure $\mu_0$ and we take any random dynamical system obtained from this one through the introduction of small random perturbations on the dynamics, as in Examples 1–4 or in the section on "Random perturbations of flows," with the noise level $\epsilon > 0$ close to zero.

In this setting if, for any choice $\mu_\epsilon$ of invariant measure for the random dynamical system for all $\epsilon > 0$ small enough, the set of accumulation points of

the family $(\mu_\epsilon)_{\epsilon>0}$, when $\epsilon$ tends to 0 – also known as zero-noise limits – is formed by physical measures or, more generally, by convex linear combinations of physical measures, then the original unperturbed dynamical system is stochastically stable.

This intuitively means that the asymptotic behavior measured through time averages of continuous observables for the random system is close to the behavior of the unperturbed system.

Recent progress in one-dimensional dynamics has shown that, for typical families $(f_t)_{t\in(0,1)}$ of maps of the circle or of the interval having a unique critical point, a full Lebesgue measure subset $T$ of the set of parameters is such that, for $t \in T$, the dynamics of $f_t$ admits a unique stochastically stable (under additive noise type random perturbations) physical measure $\mu_t$ whose basin has full measure in the ambient space (either the circle or the interval). Therefore, models involving one-dimensional unimodal maps typically are stochastically stable.

In many settings (e.g., low-dimensional dynamical systems), Lyapunov exponents can be given by time averages of continuous functions – for example, the time average of $\log\|DT_0\|$ gives the biggest exponent. In this case, stochastic stability directly implies stability of the Lyapunov exponents under small random perturbations of the dynamics.

**Example 6** (Stochastically stable examples). Let $T_0 : \mathbb{S}^1 \circlearrowleft$ be a map such that $\lambda$, the Lebesgue (length) measure on the circle, is $T_0$-invariant and ergodic. Then $\lambda$ is physical.

We consider the parametrized family $T_t : \mathbb{S}^1 \times \mathbb{S}^1 \to \mathbb{S}^1$, $(t, x) \mapsto x + t$ and a family of probability measures $\theta_\epsilon = (\lambda(-\epsilon, \epsilon))^{-1} \cdot (\lambda \mid (-\epsilon, \epsilon))$ given by the normalized restriction of $\lambda$ to the $\epsilon$-neighborhood of 0, where we regard $\mathbb{S}^1$ as the Lie group $\mathbb{R}/\mathbb{Z}$ and use additive notation for the group operation. Since $\lambda$ is $T_t$-invariant for every $t \in \mathbb{S}^1, \lambda$ is also an invariant measure for the measure-preserving random system

$$\mathcal{S} : (\mathbb{S}^1 \times \Omega^{\mathbb{N}}, \lambda \times \theta_\epsilon^{\mathbb{N}})\circlearrowleft$$

for every $\epsilon > 0$, where $\Omega = (\mathbb{S}^1)^{\mathbb{N}}$. Hence, $(T_0, \lambda)$ is stochastically stable under additive noise perturbations.

Concrete examples can be irrational rotations, $T_0(x) = x + \alpha$ with $\alpha \in \mathbb{R}\backslash\mathbb{Q}$, or expanding maps of the circle, $T_0(x) = b \cdot x$ for some $b \in \mathbb{N}, n \geq 2$. Analogous examples exist in higher-dimensional tori.

**Example 7** (Stochastic stability depends on the type of noise). In spite of the straightforward method for obtaining stochastic stability in Example 6, for example, an expanding circle map $T_0(x) = 2 \cdot x$, we can choose a continuous family of probability measures $\theta_\epsilon$ such that the same map $T_0$ is not stochastically stable.

It is well known that $\lambda$ is the unique absolutely continuous invariant measure for $T_0$ and also the unique physical measure. Given $\epsilon > 0$ small, let us define transition probability measures as follows:

$$p_\epsilon(\cdot \mid z) = \frac{\lambda \mid [\phi_\epsilon(z) - \epsilon, \phi_\epsilon(z) + \epsilon]}{\lambda([\phi_\epsilon(z) - \epsilon, \phi_\epsilon(z) + \epsilon])}$$

where $\phi_\epsilon \mid (-\epsilon, \epsilon) \equiv 0$, $\phi_\epsilon \mid [\mathbb{S}^1 \setminus (-2\epsilon, 2\epsilon)] \equiv T_0$, and over $(-2\epsilon, -\epsilon] \cup [\epsilon, 2\epsilon)$, we can define $\phi_\epsilon$ by interpolation in order that it be smooth.

In this setting, every random orbit starting at $(-\epsilon, \epsilon)$ never leaves this neighborhood in the future. Moreover, it is easy to see that every random orbit eventually enters $(-\epsilon, \epsilon)$. Hence, every invariant probability measure $\mu_\epsilon$ for this Markov chain model is supported in $[-\epsilon, \epsilon]$. Thus, letting $\epsilon \to 0$, we see that the only zero-noise limit is $\delta_0$, the Dirac mass concentrated at 0, which is not a physical measure for $T_0$.

This construction can be achieved in a random-maps setting, but only in the $C^0$ topology – it is not possible to realize this Markov chain by random maps that are $C^1$ close to $T_0$ for $\epsilon$ near 0.

### Characterization of Measures Satisfying the Entropy Formula

Significant effort has been put in recent years in extending important results from dynamical systems to the random setting. Among many examples are: the local conjugacy between the dynamics near a hyperbolic fixed point and the action of the derivative of the map on the tangent space, the stable/unstable manifold theorems for hyperbolic invariant sets and the notions and properties of metric and topological entropy, dimensions and equilibrium states for potentials on random (or fuzzy) sets.

The characterization of measures satisfying the entropy formula is one important result whose extension to the setting of iteration of independent and identically distributed random maps has recently had interesting new consequences back into nonrandom dynamical systems.

**Metric entropy for random perturbations** Given a probability measure $\mu$ and a partition $\xi$ of $M$, except perhaps for a subset of $\mu$-null measure, the entropy of $\mu$ with respect to $\xi$ is defined to be

$$H_\mu(\xi) = -\sum_{R\in\xi} \mu(R) \log \mu(R)$$

where the convention that $0 \log 0 = 0$ has been used. Given another finite partition $\zeta$, we write $\xi \vee \zeta$ to indicate the partition obtained through intersection of every element of $\xi$ with every element of $\zeta$, and analogously for any finite number of partitions. If $\mu$ is also a stationary measure for a random-maps model (see the section "Random maps"), then for any finite measurable partition $\xi$ of $M$,

$$h_\mu(\xi) = \inf_{n \geq 1} \frac{1}{n} \int H_\mu \left( \bigvee_{i=0}^{n-1} \left( T^i_{\underline{\omega}} \right)^{-1} (\xi) \right) \mathrm{d}p^{\mathbb{N}}(\underline{\omega})$$

is finite and is called the entropy of the random dynamical system with respect to $\xi$ and to $\mu$.

We define $h_\mu = \sup_\xi h_\mu(\xi)$ as the metric entropy of the random dynamical system, where the supremo is taken over all $\mu$-measurable partitions. An important point here is the following notion: setting $\mathcal{A}$ the Borel $\sigma$-algebra of $M$, we say that a finite partition $\xi$ of $M$ is a random generating partition for $\mathcal{A}$ if

$$\bigvee_{i=0}^{+\infty} (T^i_{\underline{\omega}})^{-1}(\xi) = \mathcal{A}$$

(except $\mu$-null sets) for $p^{\mathbb{N}}$-almost all $\omega \in \Omega = \mathcal{U}^{\mathbb{N}}$. Then a classical result from ergodic theory ensures that we can calculate the entropy using only a random generating partition $\xi$, that is, $h_\mu = h_\mu(\xi)$.

**The entropy formula** There exists a general relation ensuring that the entropy of a measure-preserving differentiable transformation $(T_0, \mu)$ on a compact Riemannian manifold is bounded from above by the sum of the positive Lyapunov exponents of $T_0$

$$h_\mu(T_0) \leq \int \sum_{\lambda_i(x) > 0} \lambda_i(x) \, \mathrm{d}\mu(x)$$

The equality (entropy formula) was first shown to hold for diffeomorphisms preserving a measure equivalent to the Riemannian volume, and then the measures satisfying the entropy formula were characterized: for $C^2$ diffeomorphisms the equality holds if and only if the disintegration of $\mu$ along the unstable manifolds is formed by measures absolutely continuous with respect to the Riemannian volume restricted to those submanifolds. The unstable manifolds are the submanifolds of $M$ everywhere tangent to the Lyapunov subspaces corresponding to all positive Lyapunov exponents, analogous to "integrating the distribution of Lyapunov subspaces corresponding to positive exponents" – this particular point is a main subject of smooth ergodic theory for nonuniformly hyperbolic dynamics.

Both the inequality and the characterization of stationary measures satisfying the entropy formula were extended to random iterations of independent and identically distributed $C^2$ maps (noninjective and admitting critical points), and the inequality reads

$$h_\mu \leq \iint \sum_{\lambda_i(x,\omega) > 0} \lambda_i(x, \omega) \, \mathrm{d}\mu(x) \, \mathrm{d}p^{\mathbb{N}}(\omega)$$

where the functions $\lambda_i$ are the random variables provided by the random multiplicative ergodic theorem.

## Construction of Physical Measures as Zero-Noise Limits

The characterization of measures which satisfy the entropy formula enables us to construct physical measures as zero-noise limits of random invariant measures in some settings, outlined in the following, obtaining in the process that the physical measures so constructed are also stochastically stable.

The physical measures obtained in this manner arguably are natural measures for the system, since they are both stable under (certain types of) random perturbations and describe the asymptotic behavior of the system for a positive-volume subset of initial conditions. This is a significant contribution to the state-of-the-art of present knowledge on dynamics from the perspective of random dynamical systems.

**Hyperbolic measures and the entropy formula** The main idea is that an ergodic invariant measure $\mu$ for a diffeomorphism $T_0$ which satisfies the entropy formula and whose Lyapunov exponents are everywhere nonzero (known as hyperbolic measure) necessarily is a physical measure for $T_0$. This follows from standard arguments of smooth nonuniformly hyperbolic ergodic theory.

Indeed $\mu$ satisfies the entropy formula if and only if $\mu$ disintegrates into densities along the unstable submanifolds of $T_0$. The unstable manifolds $W^{\mathrm{u}}(x)$ are tangent to the subspace corresponding to every positive Lyapunov exponent at $\mu$-almost every point $x$, they are an invariant family, that is, $T_0(W^{\mathrm{u}}(x)) = W^{\mathrm{u}}(x)$ for $\mu$-almost every $x$, and distances on them are uniformly contracted under iteration by $T_0^{-1}$.

If the exponents along the complementary directions are nonzero, then they must be negative and smooth ergodic theory ensures that there exist stable manifolds, which are submanifolds $W^{\mathrm{s}}(x)$ of

$M$ everywhere tangent to the subspace of negative Lyapunov exponents at $\mu$-almost every point $x$, form a $T_0$-invariant family ($T_0(W^s(x)) = W^s(x)$, $\mu$-almost everywhere), and distances on them are uniformly contracted under iteration by $T_0$.

We still need to understand that time averages are constant along both stable and unstable manifolds, and that the families of stable and unstable manifolds are absolutely continuous, in order to realize how a hyperbolic measure is a physical measure.

Given $y \in W^s(x)$, the time averages of $x$ and $y$ coincide for continuous observables simply because dist $(T_0^n(x), T_0^n(y)) \to 0$ when $n \to +\infty$. For unstable manifolds, the same holds when considering time averages for $T_0^{-1}$. Since forward and backward time averages are equal $\mu$-almost everywhere, the set of points having asymptotic time averages given by $\mu$ has positive Lebesgue measure if the set

$$B = \bigcup \{ W^s(y) \colon y \in W^u(x) \cap \mathrm{supp}(\mu) \}$$

has positive volume in $M$, for some $x$ whose time averages are well defined.

Now, stable and unstable manifolds are transverse everywhere where they are defined, but they are only defined $\mu$-almost everywhere and depend measurably on the base point, so we cannot use transversality arguments from differential topology, in spite of $W^u(x) \cap \mathrm{supp}(\mu)$ having positive volume in $W^u(x)$ by the existence of a smooth disintegration of $\mu$ along the unstable manifolds. However, it is known for smooth ($C^2$) transformations that the families of stable and unstable manifolds are absolutely continuous, meaning that projections along leaves preserve sets of zero volume. This is precisely what is needed for measure-theoretic arguments to show that $B$ has positive volume.

**Zero-noise limits satisfying the entropy formula** Using the extension of the characterization of measures satisfying the entropy formula for the random-maps setting, we can build random dynamical systems, which are small random perturbations of a map $T_0$, having invariant measures $\mu_\epsilon$ satisfying the entropy formula for all sufficiently small $\epsilon > 0$. Indeed, it is enough to construct small random perturbations of $T_0$ having absolutely continuous invariant probability measures $\mu_\epsilon$ for all small enough $\epsilon > 0$.

In order to obtain such random dynamical systems, we choose families of maps $T : \mathcal{U} \times M \to M$ and of probability measures $(\theta_\epsilon)_{\epsilon>0}$ as in Examples 3 and 4, where we assume that $o \in \mathcal{U}$, so that $T_0$ belongs to the family. Letting $T_x(u) = T(u, x)$ for all $(u, x) \in \mathcal{U} \times M$, we then have that $T_x(\theta_\epsilon)$ is absolutely continuous. This means that sets of perturbations of positive $\theta_\epsilon$-measure send points of $M$ onto positive-volume subsets of $M$. Such a perturbation can be constructed for every continuous map of any manifold.

In this setting, any invariant probability measure for the associated skew-product map $\mathcal{S} : \Omega \times M \circlearrowleft$ of the form $\theta_\epsilon^{\mathbb{N}} \times \mu_\epsilon$ is such that $\mu_\epsilon$ is absolutely continuous with respect to volume on $M$. Then the entropy formula holds:

$$h_{\mu_\epsilon} = \iint \sum_{\lambda_i(x,\omega)>0} \lambda_i(x, \omega) \, \mathrm{d}\mu_\epsilon(x) \, \mathrm{d}\theta_\epsilon^{\mathbb{N}}(\omega)$$

Having this and knowing the characterization of measures satisfying the entropy formula, it is natural to look for conditions under which we can guarantee that the above inequality extends to any zero-noise limit $\mu_0$ of $\mu_\epsilon$ when $\epsilon \to 0$. In this case, $\mu_0$ satisfies the entropy formula for $T_0$.

If, in addition, we are able to show that $\mu_0$ is a hyperbolic measure, then we obtain a physical measure for $T_0$ which is stochastically stable by construction.

These ideas can be carried out completely for hyperbolic diffeomorphisms, that is, maps admitting a continuous invariant splitting of the tangent space into two sub-bundles $E \oplus F$ defined everywhere with bounded angles, whose Lyapunov exponents are negative along $E$ and positive along $F$. Recently, maps satisfying weaker conditions were shown to admit stochastically stable physical measures following the same ideas.

These ideas also have applications to the construction and stochastic stability of physical measure for strange attractors and for all mathematical models involving ordinary differential equations or iterations of maps.

*See also:* Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Homeomorphisms and Diffeomorphisms of the Circle; Lyapunov Exponents and Strange Attractors; Nonequilibrium Statistical Mechanics (Stationary): Overview; Random Walks in Random Environments; Stochastic Differential Equations.

## Further Reading

Arnold L (1998) *Random Dynamical Systems*. Berlin: Springer.
Billingsley P (1965) *Ergodic Theory and Information*. New York: Wiley.
Billingsley P (1985) *Probability and Measure*, 3rd edn. New York: Wiley.
Bonatti C, Díaz L, and Viana M (2004) *Dynamics Beyond Hyperbolicity: A Global Geometric and Probabilistic Perspective*. Berlin: Springer.

Bonatti C, Díaz L, and Viana M (2005) *Dynamics Beyond Uniform Hyperbolicity. A Global Geometric and Probabilistic Perspective*, Encyclopaedia of Mathematical Sciences, 102 Mathematical Physics III. Berlin: Springer.
Doob J (1953) *Stochastic Processes*. New York: Wiley.
Fathi A, Herman M-R, and Yoccoz J-C (1983) A proof of Pesin's stable manifold theorem. In: Palis J (ed.) *Geometric Dynamics (Rio de Janeiro, 1981)*, Lecture Notes in Mathematics, vol. 1007, 177–215. Berlin: Springer.
Kifer Y (1986) *Ergodic Theory of Random Perturbations*. Boston: Birkhäuser.
Kifer Y (1988) *Random Perturbations of Dynamical Systems*. Boston: Birkhäuser.
Kunita H (1990) *Stochastic Flows and Stochastic Differential Equations*. Cambridge: Cambridge University Press.
Ledrappier F and Young L-S (1998) Entropy formula for random transformations. *Probability Theory and Related Fields* 80(2): 217–240.
Liu P-D and Qian M (1995) *Smooth Ergodic Theory of Random Dynamical Systems*, Lecture Notes in Mathematics, vol. 1606. Berlin: Springer.
Øskendal B (1992) *Stochastic Differential Equations. Universitext*, 3rd edn. Berlin: Springer.
Viana M (2000) What's new on Lorenz strange attractor. *Mathematical Intelligencer* 22(3): 6–19.
Walters P (1982) *An Introduction to Ergodic Theory*. Berlin: Springer.

# Random Matrix Theory in Physics

**T Guhr**, Lunds Universitet, Lund, Sweden



## Introduction

We wish to study energy correlations of quantum spectra. Suppose the spectrum of a quantum system has been measured or calculated. All levels in the total spectrum having the same quantum numbers form one particular *subspectrum*. Its energy levels are at positions $x_n$, $n = 1, 2, \ldots, N$, say. We assume that $N$, the number of levels in this subspectrum, is large. With a proper *smoothing* procedure, we obtain the level density $R_1(x)$, that is, the probability density of finding a level at the energy $x$. As indicated in the top part of Figure 1, the level density $R_1(x)$ increases with $x$ for most physics systems. In the present context, however, we are not so interested in the level density. We want to measure the spectral correlations independently of it. Hence, we have to remove the level density from the subspectrum. This is referred to as *unfolding*. We introduce a new dimensionless energy scale $\xi$ such that $\mathrm{d}\xi = R_1(x)\,\mathrm{d}x$. By construction, the resulting subspectrum in $\xi$ has level density unity, as shown schematically in the bottom part of Figure 1. It is always understood that the energy correlations are analyzed in the *unfolded subspectra*.

Surprisingly, a remarkable *universality* is found in the spectral correlations of a large class of systems, including nuclei, atoms, molecules, quantum chaotic and disordered systems, and even quantum chromodynamics on the lattice. Consider the *nearest-neighbor spacing distribution* $p(s)$. It is the probability density of finding two adjacent levels in the distance $s$. If the positions of the levels are uncorrelated, the nearest-neighbor spacing distribution can be shown to follow the *Poisson law*

$$p^{(\mathrm{P})}(s) = \exp(-s) \qquad [1]$$

While this is occasionally found, many more systems show a rather different nearest-neighbor spacing distribution, the *Wigner surmise*

$$p^{(\mathrm{W})}(s) = \frac{\pi}{2} s \exp\left(-\frac{\pi}{4} s^2\right) \qquad [2]$$

As shown in Figure 2, the Wigner surmise excludes degeneracies, $p^{(\mathrm{W})}(0) = 0$, the levels repel each other. This is only possible if they are correlated. Thus, the Poisson law and the Wigner surmise reflect the absence or the presence of energy correlations, respectively.

Now, the question arises: if these correlation patterns are so frequently found in physics, is there some simple, phenomenological model? – Yes, random matrix theory (RMT) is precisely this. To describe the absence of correlations, we choose, in view of what has been said above, a diagonal Hamiltonian

$$H = \mathrm{diag}(x_1, \ldots, x_N) \qquad [3]$$

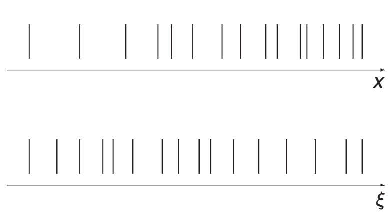



**Figure 1** Original (top) and unfolded (bottom) spectrum.

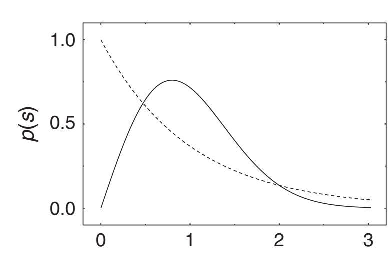



**Figure 2** Wigner surmise (solid) and Poisson law (dashed).

whose elements, the eigenvalues $x_n$, are uncorrelated random numbers. To model the presence of correlations, we insert off-diagonal matrix elements,

$$H = \begin{bmatrix} H_{11} & \cdots & H_{1N} \\ \vdots & & \vdots \\ H_{N1} & \cdots & H_{NN} \end{bmatrix} \quad [4]$$

We require that H is real symmetric, $H^{\mathrm{T}} = H$. The independent elements $H_{nm}$ are random numbers. The random matrix $H$ is diagonalized to obtain the energy levels $x_n, n = 1, 2, \ldots, N$. Indeed, a numerical simulation shows that these two models yield, after unfolding, the Poisson law and the Wigner surmise for large $N$, that is, the absence or presence of correlations. This is the most important insight into the phenomenology of RMT.

In this article, we set up RMT in a more formal way; we discuss analytical calculations of correlation functions, demonstrate how this relates to supersymmetry and stochastic field theory and show the connection to chaos, and we briefly sketch the numerous applications in many-body physics, in disordered and mesoscopic systems, in models for interacting fermions, and in quantum chromodynamics. We also mention applications in other fields, even beyond physics.

## Random Matrix Theory

### Classical Gaussian Ensembles

For now, we consider a system whose energy levels are correlated. The $N \times N$ matrix $H$ modeling it has no fixed zeros but random entries everywhere. There are three possible *symmetry classes* of random matrices in standard Schrödinger quantum mechanics. They are labeled by the Dyson index $\beta$. If the system is not time-reversal invariant, $H$ has to be Hermitian and the random entries $H_{nm}$ are complex ($\beta = 2$). If time-reversal invariance holds, two possibilities must be distinguished: if either the system is rotational symmetric, or it has integer spin and rotational symmetry is broken, the Hamilton matrix $H$ can be chosen to be real symmetric ($\beta = 1$). This is the case in eqn [4]. If, on the other hand, the system has half-integer spin and rotational symmetry is broken, $H$ is self-dual ($\beta = 4$) and the random entries $H_{nm}$ are $2 \times 2$ quaternionic. The Dyson index $\beta$ is the dimension of the number field over which $H$ is constructed.

As we are interested in the eigenvalue correlations, we diagonalize the random matrix, $H = U^{-1}xU$. Here, $x = \mathrm{diag}(x_1, \ldots, x_N)$ is the diagonal matrix of the $N$ eigenvalues. For $\beta = 4$, every eigenvalue is doubly degenerate. This is *Kramers' degeneracy*. The diagonalizing matrix $U$ is in the orthogonal group $\mathrm{O}(N)$ for $\beta = 1$, in the unitary group $\mathrm{U}(N)$ for $\beta = 2$ and in the unitary–symplectic group $\mathrm{USp}(2N)$ for $\beta = 4$. Accordingly, the three symmetry classes are referred to as *orthogonal, unitary*, and *symplectic*.

We have not yet chosen the probability densities for the random entries $H_{nm}$. To keep our assumptions about the system at a minimum, we treat all entries on equal footing. This is achieved by *rotational invariance* of the probability density $P_N^{(\beta)}(H)$, not to be confused with the rotational symmetry employed above to define the symmetry classes. No basis for the matrices is preferred in any way if we construct $P_N^{(\beta)}(H)$ from matrix invariants, that is, from traces and determinants, such that it depends only on the eigenvalues, $P_N^{(\beta)}(H) = P_N^{(\beta)}(x)$. A particularly convenient choice is the Gaussian

$$P_N^{(\beta)}(H) = C_N^{(\beta)} \exp\left(-\frac{\beta}{4v^2} \operatorname{tr} H^2\right) \quad [5]$$

where the constant $v$ sets the energy scale and the constant $C_N^{(\beta)}$ ensures normalization. The three symmetry classes together with the probability densities [5] define the *Gaussian ensembles*: the Gaussian orthogonal (GOE), unitary (GUE) and symplectic (GSE) ensemble for $\beta = 1, 2, 4$.

The phenomenology of the three Gaussian ensembles differs considerably. The higher $\beta$, the stronger the level repulsion between the eigenvalues $x_n$. Numerical simulation quickly shows that the nearest-neighbor spacing distribution behaves like $p^{(\beta)}(s) \sim s^\beta$ for small spacings $s$. This also becomes obvious by working out the differential probability $P_N^{(\beta)}(H)\mathrm{d}[H]$ of the random matrices $H$ in eigenvalue–angle coordinates $x$ and $U$. Here, $\mathrm{d}[H]$ is the invariant measure or volume element in the matrix space. When writing $\mathrm{d}[\cdot]$, we always mean the product of all differentials of independent variables for the quantity in the square brackets. Up to constants, we have

$$\mathrm{d}[H] = |\Delta_N(x)|^\beta\, \mathrm{d}[x]\, \mathrm{d}\mu(U) \quad [6]$$

where $\mathrm{d}\mu(U)$ is, apart from certain phase contributions, the invariant or Haar measure on $\mathrm{O}(N)$, $\mathrm{U}(N)$, or $\mathrm{USp}(2N)$, respectively. The Jacobian of the transformation is the modulus of the Vandermonde determinant

$$\Delta_N(x) = \prod_{n<m} (x_n - x_m) \quad [7]$$

raised to the power $\beta$. Thus, the differential probability $P_N^{(\beta)}(H)\,\mathrm{d}[H]$ vanishes whenever any two eigenvalues $x_n$ degenerate. This is the level

repulsion. It immediately explains the behavior of the nearest-neighbor spacing distribution for small spacings.

Additional symmetry constraints lead to new random matrix ensembles relevant in physics, the *Andreev* and the *chiral Gaussian ensembles*. If one refers to the classical Gaussian ensembles, one usually means the three ensembles introduced above.

## Correlation Functions

The probability density to find $k$ energy levels at positions $x_1, \ldots, x_k$ is the $k$-level correlation function $R_k^{(\beta)}(x_1, \ldots, x_k)$. We find it by integrating out $N-k$ levels in the $N$-level differential probability $P_N^{(\beta)}(H)\,\mathrm{d}[H]$. We also have to average over the bases, that is, over the diagonalizing matrices $U$. Due to rotational invariance, this simply yields the group volume. Thus, we have

$$R_k^{(\beta)}(x_1, \ldots, x_k) = \frac{N!}{(N-k)!}\int_{-\infty}^{+\infty} \mathrm{d}x_{k+1} \cdots \int_{-\infty}^{+\infty} \mathrm{d}x_N |\Delta_N(x)|^{\beta} P_N^{(\beta)}(x) \quad [8]$$

Once more, we used rotational invariance which implies that $P_N^{(\beta)}(x)$ is invariant under permutation of the levels $x_n$. Since the same then also holds for the correlation functions [8], it is convenient to normalize them to the combinatorial factor in front of the integrals. A constant ensuring this has been absorbed into $P_N^{(\beta)}(x)$.

Remarkably, the integrals in eqn [8] can be done in closed form. The GUE case ($\beta=2$) is mathematically the simplest, and one finds the determinant structure

$$R_k^{(2)}(x_1, \ldots, x_k) = \det[K_N^{(2)}(x_p, x_q)]_{p,q=1,\ldots,k} \quad [9]$$

All entries of the determinant can be expressed in terms of the kernel $K_N^{(2)}(x_p, x_q)$, which depends on two energy arguments $(x_p, x_q)$. Analogous but more complicated formulae are valid for the GOE ($\beta=1$) and the GSE ($\beta=4$), involving quaternion determinants and integrals and derivatives of the kernel.

As argued in the Introduction, we are interested in the energy correlations on the unfolded energy scale. The level density is formally the one-level correlation function. For the three Gaussian ensembles it is, to leading order in the level number $N$, the *Wigner semicircle*

$$R_1^{(\beta)}(x_1) = \frac{1}{2\pi v^2}\sqrt{4\,Nv^2 - x_1^2} \quad [10]$$

for $|x_1| \leq 2\sqrt{N}v$ and zero for $|x_1| > 2\sqrt{N}v$. None of the common systems in physics has such a level density. When unfolding, we also want to take the limit of infinitely many levels $N \to \infty$ to remove cutoff effects due to the finite dimension of the random matrices. It suffices to stay in the center of the semicircle where the mean level spacing is $D = 1/R_1^{(\beta)}(0) = \pi v/\sqrt{N}$. We introduce the dimensionless energies $\xi_p = x_p/D, p = 1, \ldots, k$, which have to be held fixed when taking the limit $N \to \infty$. The *unfolded correlation functions* are given by

$$X_k^{(\beta)}(\xi_1, \ldots, \xi_k) = \lim_{N\to\infty} D^k R_k^{(\beta)}(D\xi_1, \ldots, D\xi_k) \quad [11]$$

As we are dealing with probability densities, the Jacobians $\mathrm{d}x_p/\mathrm{d}\xi_p$ enter the reformulation in the new energy variables. This explains the factor $D^k$. Unfolding makes the correlation functions *translation invariant*; they depend only on the differences $\xi_p - \xi_q$. The unfolded correlation functions can be written in a rather compact form. For the GUE ($\beta=2$), they read

$$X_k^{(2)}(\xi_1, \ldots, \xi_k) = \det\left[\frac{\sin \pi(\xi_p - \xi_q)}{\pi(\xi_p - \xi_q)}\right]_{p,q=1,\ldots,k} \quad [12]$$

There are similar, but more complicated, formulae for the GOE ($\beta=1$) and the GSE ($\beta=4$). By construction, one has $X_1^{(\beta)}(\xi_1) = 1$.

It is useful to formulate the case where correlations are absent, that is, the *Poisson case*, accordingly. The level density $R_1^{(\mathrm{P})}(x_1)$ is simply $N$ times the (smooth) probability density chosen for the entries in the diagonal matrix [4]. Lack of correlations means that the $k$-level correlation function only involves one-level correlations,

$$R_k^{(\mathrm{P})}(x_1, \ldots, x_k) = \frac{N!}{(N-k)!N^k}\prod_{p=1}^{k} R_1^{(\mathrm{P})}(x_p) \quad [13]$$

The combinatorial factor is important, since we always normalize to $N!/(N-k)!$. Hence, one finds

$$X_k^{(\mathrm{P})}(\xi_1, \ldots, \xi_k) = 1 \quad [14]$$

for all unfolded correlation functions.

## Statistical Observables

The unfolded correlation functions yield all statistical observables. The two-level correlation function $X_2(r)$ with $r = \xi_1 - \xi_2$ is of particular interest in applications. If we do not write the superscript ($\beta$) or (P), we mean either of the functions. For the Gaussian ensembles, $X_2^{(\beta)}(r)$ is shown in Figure 3. One often writes $X_2(r) = 1 - Y_2(r)$. The two-level *cluster function* $Y_2(r)$ nicely measures the deviation from the uncorrelated Poisson case, where one has $X_2^{(\mathrm{P})}(r) = 1$ and $Y_2^{(\mathrm{P})}(r) = 0$.

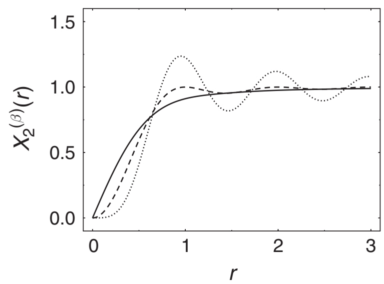



**Figure 3** Two-level correlation function $X_2^{(\beta)}(r)$ for GOE (solid), GUE (dashed) and GSE (dotted).

By construction, the average level number in an interval of length $L$ in the unfolded spectrum is $L$. The level *number variance* $\Sigma^2(L)$ is shown to be an average over the two-level cluster function,

$$\Sigma^2(L) = L - 2\int_0^L (L-r)Y_2(r)\mathrm{d}r \qquad [15]$$

We find $L \pm \sqrt{\Sigma^2(L)}$ levels in an interval of length $L$. In the uncorrelated Poisson case, one has $\Sigma^{2(\mathrm{P})}(L) = L$. This is just *Poisson's error law*. For the Gaussian ensembles $\Sigma^{2(\beta)}(L)$ behaves logarithmically for large $L$. The spectrum is said to be more *rigid* than in the Poisson case. As Figure 4 shows, the level number variance probes longer distances in the spectrum, in contrast to the nearest-neighbor spacing distribution.

Many more observables, also sensitive to higher order, $k > 2$ correlations, have been defined. In practice, however, one is often restricted to analyzing two-level correlations. An exception is, to some extent, the nearest-neighbor spacing distribution $p(s)$. It is the two-level correlation function with the *additional* requirement that the two levels in question are adjacent, that is, that there are no levels between them. Thus, *all* correlation functions are needed if one wishes to calculate the exact nearest-neighbor spacing distribution $p^{(\beta)}(s)$ for the Gaussian ensembles. These considerations explain that we have $X_2^{(\beta)}(s) \simeq p^{(\beta)}(s)$ for small $s$. But while $X_2^{(\beta)}(s)$ saturates for large $s$, $p^{(\beta)}(s)$ quickly goes to zero in a Gaussian fashion. Thus, although the nearest-neighbor spacing distribution mathematically involves all correlations, it makes in practice only a meaningful statement about the two-level correlations. Luckily, $p^{(\beta)}(s)$ differs only very slightly from the heuristic Wigner surmise [2] (corresponding to $\beta = 1$), respectively from its extensions (corresponding to $\beta = 2$ and $\beta = 4$).

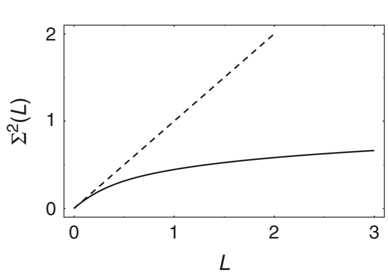



**Figure 4** Level number variance $\Sigma^2(L)$ for GOE (solid) and Poisson case (dashed).

## Ergodicity and Universality

We constructed the correlation functions as averages over an ensemble of random matrices. But this is not how we proceeded in the data analysis sketched in the Introduction. There, we started from *one single* spectrum with very many levels and obtained the statistical observable just by sampling and, if necessary, smoothing. Do these two averages, the *ensemble average* and the *spectral average*, yield the same? Indeed, one can show that the answer is affirmative, if the level number $N$ goes to infinity. This is referred to as *ergodicity* in RMT.

Moreover, as already briefly indicated in the Introduction, very many systems from different areas of physics are well described by RMT. This seems to be at odds with the Gaussian assumption [5]. There is hardly any system whose Hamilton matrix elements follow a Gaussian probability density. The solution for this puzzle lies in the unfolding. Indeed, it has been shown that almost all functional forms of the probability density $P_N^{(\beta)}(H)$ yield the same unfolded correlation functions, if no new scale comparable to the mean level spacing is present in $P_N^{(\beta)}(H)$. This is the mathematical side of the empirically found *universality*.

Ergodicity and universality are of crucial importance for the applicability of RMT in data analysis.

## Wave Functions

By modeling the Hamiltonian of a system with a random matrix $H$, we do not only make an assumption about the statistics of the energies, but also about those of the wave functions. Because of the eigenvalue equation $Hu_n = x_n u_n, n = 1, \ldots, N$, the wave function belonging to the eigenenergy $x_n$ is modeled by the eigenvector $u_n$. The columns of the diagonalizing matrix $U = [u_1\, u_2 \cdots u_N]$ are these eigenvectors. The probability density of the components $u_{nm}$ of the eigenvector $u_n$ can be calculated rather easily. For large $N$ it approaches a Gaussian. This is equivalent to the *Porter–Thomas distribution*. While wave functions are often not accessible in an experiment, one can measure transition amplitudes and widths, giving information about the matrix elements of a transition operator and a

projection of the wave functions onto a certain state in Hilbert space. If the latter are represented by a fixed matrix $A$ or a fixed vector $a$, respectively, one can calculate the RMT prediction for the probability densities of the matrix elements $u_n^\dagger A u_m$ or the widths $a^\dagger u_n$ from the probability density of the eigenvectors.

### Scattering Systems

It is important that RMT can be used as a powerful tool in scattering theory, because the major part of the experimental information about quantum systems comes from scattering experiments. Consider an example from compound nucleus scattering. In an accelerator, a proton is shot on a nucleus, with which it forms a compound nucleus. This then decays by emitting a neutron. More generally, the ingoing *channel* $\nu$ (the proton in our example) connects to the *interaction region* (the nucleus), which also connects to an outgoing channel $\mu$ (the neutron). There are $\Lambda$ channels with channel wave functions which are labeled $\nu = 1, \ldots, \Lambda$. The interaction region is described by an $N \times N$ Hamiltonian matrix $H$ whose eigenvalues $x_n$ are bound-state energies labeled $n = 1, \ldots, N$. The dimension $N$ is a cutoff which has to be taken to infinity at the end of a calculation. The $\Lambda \times \Lambda$ *scattering matrix* $S$ contains the information about how the ingoing channels are transformed into the outgoing channels. The scattering matrix $S$ is unitary. Under certain and often justified assumptions, a scattering matrix element can be cast into the form

$$S_{\nu\mu} = \delta_{\nu\mu} - \mathrm{i}2\pi W_\nu^\dagger G^{-1} W_\mu \qquad [16]$$

The couplings $W_{n\nu}$ between the bound states $n$ and the channels $\nu$ are collected in the $N \times \Lambda$ matrix $W$, $W_\nu$ is its $\nu$th column. The propagator $G^{-1}$ is the inverse of

$$G = z1_N - H + \mathrm{i}\pi \sum_{\nu \text{ open}} W_\nu W_\nu^\dagger \qquad [17]$$

Here, $z$ is the scattering energy and the summation is only over channels which are open, that is, accessible. Formula [16] has a clear intuitive interpretation. The scattering region is entered through channel $\nu$, the bound states of $H$ become *resonances* in the scattering process according to eqn [17], the interaction region is left through channel $\mu$. This formulation applies in many areas of physics. All observables such as transmission coefficients, cross sections, and others can be calculated from the scattering matrix $S$.

We have not made any statistical assumptions yet. Often, one can understand generic features of a scattering system by assuming that the Hamiltonian $H$ is a random matrix, taken from one of the three classical ensembles. This is one RMT approach used in scattering theory.

Another RMT approach is based on the scattering matrix itself, $S$ is modeled by a $\Lambda \times \Lambda$ unitary random matrix. Taking into account additional symmetries, one arrives at the *three circular ensembles*, circular orthogonal (COE), unitary (CUE) and symplectic (CSE). They correspond to the three classical Gaussian ensembles and are also labeled with the Dyson index $\beta = 1, 2, 4$. The eigenphases of the random scattering matrix correspond to the eigenvalues of the random Hamiltonian matrix. The unfolded correlation functions of the circular ensembles are identical to those of the Gaussian ensembles.

### Supersymmetry

Apart from the symmetries, random matrices contain nothing but random numbers. Thus, a certain type of redundancy is present in RMT. Remarkably, this redundancy can be removed, without losing any piece of information by using *supersymmetry*, that is, by a reformulation of the random matrix model involving commuting and anticommuting variables. For the sake of simplicity, we sketch the main ideas for the GUE, but they apply to the GOE and the GSE accordingly.

One defines the $k$-level correlation functions by using the resolvent of the Schrödinger equation,

$$\widehat{R}_k^{(2)}(x_1, \ldots, x_k) = \frac{1}{\pi^k} \int P_N^{(2)}(H) \prod_{p=1}^{k} \mathrm{tr} \frac{1}{x_p^\pm - H} \mathrm{d}[H] \qquad [18]$$

The energies carry an imaginary increment $x_p^\pm = x_p \pm \mathrm{i}\varepsilon$ and the limit $\varepsilon \to 0$ has to be taken at the end of the calculation. The $k$-level correlation functions $R_k^{(2)}(x_1, \ldots, x_k)$ as defined in eqn [8] can always be obtained from the functions [18] by constructing a linear combination of the $\widehat{R}_k^{(2)}(x_1, \ldots, x_k)$ in which the signs of the imaginary increments are chosen such that only the imaginary parts of the traces contribute. Some trivial $\delta$-distributions have to be removed. The $k$-level correlation functions [18] can be written as the $k$-fold derivative

$$\widehat{R}_k^{(2)}(x_1, \ldots, x_k) = \frac{1}{(2\pi)^k} \frac{\partial^k}{\prod_{p=1}^{k} \partial J_p} Z_k^{(2)}(x + J) \bigg|_{J=0} \qquad [19]$$

of the generating function

$$Z_k^{(2)}(x+J) = \int P_N^{(2)}(H) \prod_{p=1}^{k} \frac{\det\left(x_p^{\pm} + J_p - H\right)}{\det\left(x_p^{\pm} - J_p - H\right)} \mathrm{d}[H] \qquad [20]$$

which depends on the energies and $k$ new source variables $J_p, p = 1, \ldots, k$, ordered in $2k \times 2k$ diagonal matrices

$$\begin{aligned} x &= \mathrm{diag}(x_1, x_1, \ldots, x_k, x_k) \\ J &= \mathrm{diag}(+J_1, -J_1, \ldots, +J_k, -J_k) \end{aligned} \qquad [21]$$

We notice the normalization $Z_k^{(2)}(x) = 1$ at $J = 0$. The generating function [20] is an integral over an ordinary $N \times N$ matrix $H$. It can be exactly rewritten as an integral over a $2k \times 2k$ *supermatrix* $\sigma$ containing commuting and anticommuting variables,

$$Z_k^{(2)}(x+J) = \int Q_k^{(2)}(\sigma) \mathrm{sdet}^{-N}(x^{\pm} + J - \sigma) \mathrm{d}[\sigma] \qquad [22]$$

The integrals over the commuting variables are of the ordinary *Riemann–Stiltjes* type, while those over the anticommuting variables are *Berezin* integrals. The Gaussian probability density [5] is mapped onto its counterpart in superspace

$$Q_k^{(2)}(\sigma) = c_k^{(2)} \exp\left(-\frac{1}{2v^2} \mathrm{str}\, \sigma^2\right) \qquad [23]$$

where $c_k^{(2)}$ is a normalization constant. The supertrace str and the superdeterminant sdet generalize the corresponding invariants for ordinary matrices. The total number of integrations in eqn [22] is drastically reduced as compared to eqn [20]. Importantly, it is independent of the level number $N$ which now only appears as the negative power of the superdeterminant in eqn [22], that is, as an explicit parameter. This most convenient feature makes it possible to take the limit of infinitely many levels by means of a saddle point approximation to the generating function.

Loosely speaking, the supersymmetric formulation can be viewed as an irreducible representation of RMT which yields a clearer insight into the mathematical structures. The same is true for applications in scattering theory and in models for crossover transitions to be discussed below. This explains why supersymmetry is so often used in RMT calculations.

It should be emphasized that the rôle of supersymmetry in RMT is quite different from the one in high-energy physics, where the commuting and anticommuting variables represent physical particles, bosons and fermions, respectively. This is not so in the RMT context. The commuting and anticommuting variables have no direct physics interpretation; they appear simply as helpful mathematical devices to cast the RMT model into an often much more convenient form.

### Crossover Transitions

The RMT models discussed up to now describe four extreme situations, the absence of correlations in the Poisson case and the presence of correlations as in the three fully rotational invariant models GOE, GUE, and GSE. A real physics system, however, is often between these extreme situations. The corresponding RMT models can vary considerably, depending on the specific situation. Nevertheless, those models in which the random matrices for two extreme situations are simply added with some weight are useful in so many applications that they acquired a rather generic standing. One writes

$$H(\alpha) = H^{(0)} + \alpha H^{(\beta)} \qquad [24]$$

where $H^{(0)}$ is a random matrix drawn from an ensemble with a completely arbitrary probability density $P_N^{(0)}(H^{(0)})$. The case of a fixed matrix is included, because one may choose a product of $\delta$-distributions for the probability density. The matrix $H^{(\beta)}$ is random and drawn from the classical Gaussian ensembles with probability density $P_N^{(\beta)}(H^{(\beta)})$ for $\beta = 1, 2, 4$. One requires that the group diagonalizing $H^{(0)}$ is a subgroup of the one diagonalizing $H^{(\beta)}$. The model [24] describes a *crossover transition*. The weight $\alpha$ is referred to as *transition parameter*. It is useful to choose the spectral support of $H^{(0)}$ and $H^{(\beta)}$ equal. One can then view $\alpha$ as the root-mean-square matrix element of $H^{(\beta)}$. At $\alpha = 0$, one has the arbitrary ensemble. The Gaussian ensembles are formally recovered in the limit $\alpha \to \infty$, to be taken in a proper way such that the energies remain finite.

We are always interested in the unfolded correlation functions. Thus, $\alpha$ has to be measured in units of the mean level spacing $D$ such that $\lambda = \alpha / D$ is the physically relevant transition parameter. It means that, depending on the numerical value of $D$, even a small effect on the original energy scale can have sizeable impact on the spectral statistics. This is referred to as *statistical enhancement*. The nearest-neighbor spacing distribution is already very close to $p^{(\beta)}(s)$ for the Gaussian ensembles if $\lambda$ is larger than 0.5 or so. In the long-range observables such as the level number variance $\Sigma^2(L)$, the deviation from the Gaussian ensemble statistics becomes visible at interval lengths $L$ comparable to $\lambda$.

Crossover transitions can be interpreted as *diffusion processes*. With the fictitious time $t = \alpha^2/2$, the probability density $P_N(x,t)$ of the eigenvalues $x$ of the total Hamilton matrix $H = H(t) = H(\alpha)$ satisfies the diffusion equation

$$\Delta_x P_N(x,t) = \frac{4}{\beta}\frac{\partial}{\partial t} P_N(x,t) \qquad [25]$$

where the probability density for the arbitrary ensemble is the initial condition $P_N(x,0) = P_N^{(0)}(x)$. The Laplacian

$$\Delta_x = \sum_{n=1}^{N} \frac{\partial^2}{\partial x_n^2} + \sum_{n<m} \frac{\beta}{x_n - x_m}\left(\frac{\partial}{\partial x_n} - \frac{\partial}{\partial x_m}\right) \qquad [26]$$

lives in the curved space of the eigenvalues $x$. This diffusion process is *Dyson's Brownian motion* in slightly simplified form. It has a rather general meaning for *harmonic analysis on symmetric spaces*, connecting to the spherical functions of Gelfand and Harish-Chandra, Itzykson–Zuber integrals, and to *Calogero–Sutherland models of interacting particles*. All this generalizes to superspace. In the supersymmetric version of Dyson's Brownian motion the generating function of the correlation functions is propagated,

$$\Delta_s Z_k(s,t) = \frac{4}{\beta}\frac{\partial}{\partial t} Z_k(s,t) \qquad [27]$$

where the initial condition $Z_k(s,0) = Z_k^{(0)}(s)$ is the generating function of the correlation functions for the arbitrary ensemble. Here, $s$ denotes the eigenvalues of some supermatrices, not to be confused with the spacing between adjacent levels. Since the Laplacian $\Delta_s$ lives in this curved eigenvalue space, this diffusion process establishes an intimate connection to harmonic analysis on superspaces. Advantageously, the diffusion [27] is the same on the original and on the unfolded energy scales.

## Fields of Application

### Many-Body Systems

Numerous studies apply RMT to nuclear physics which is also the field of its origin. If the total number of nucleons, that is, protons and neutrons, is not too small, nuclei show *single-particle* and *collective* motion. Roughly speaking, the former is decoherent out-of-phase motion of the nucleons confined in the nucleus, while the latter is coherent in-phase motion of all nucleons or of large groups of them such that any additional individual motion of the nucleons becomes largely irrelevant. It has been shown empirically that the single-particle excitations lead to GOE statistics, while collective excitations produce different statistics, often of the Poisson type. Mixed statistics as described by crossover transitions are then of particular interest to investigate the character of excitations. For example, one applies the model [24] with $H^{(0)}$ drawn from a Poisson ensemble and $H^{(\beta)}$ from a GOE. Another application of crossover transitions is breaking of *time-reversal invariance* in nuclei. Here, $H^{(0)}$ is from a GOE and $H^{(\beta)}$ from a GUE. Indeed, a fit of spectral data to this model yields an upper bound for the *time-reversal invariance* violating root-mean-square matrix element in nuclei. Yet another application is breaking of symmetries such as parity or isospin. In the case of two quantum numbers, positive and negative parity, say, one chooses $H^{(0)} = \mathrm{diag}(H^{(+)}, H^{(-)})$ block-diagonal with $H^{(+)}$ and $H^{(-)}$ drawn from two uncorrelated GOE and $H^{(\beta)}$ from a third uncorrelated GOE which breaks the block structure. Again, root-mean-square matrix elements for symmetry breaking have been derived from the data.

Nuclear excitation spectra are extracted from scattering experiments. An analysis as described above is only possible if the resonances are isolated. Often, this is not the case and the resonance widths are comparable to or even much larger than the mean level spacing, making it impossible to obtain the excitation energies directly from the cross sections. One then analyzes the latter and their fluctuations as measured and applies the concepts sketched above for scattering systems. This approach has also been successful for crossover transitions.

Due to the complexity of the nuclear many-body problem, one has to use effective or phenomenological interactions when calculating spectra. Hence, one often studies whether the statistical features found in the experimental data are also present in the calculated spectra which result from the various models for nuclei.

Other many-body systems, such as complex atoms and molecules, have also been studied with RMT concepts, but the main focus has always been on nuclei.

### Quantum Chaos

Originally, RMT was intended for modeling systems with many degrees of freedom such as nuclei. Surprisingly, RMT proved useful for systems with few degrees of freedom as well. Most of these studies aim at establishing a link between RMT and *classical chaos*. Consider as an example the classical motion of a point-like particle in a rectangle billiard. Ideal reflection at the boundaries and absence of friction are assumed, implying that the particle is reflected infinitely many times. A second billiard is built by taking a rectangle and replacing one corner with a quarter circle as shown in **Figure 5**. The motion of the particle in this *Sinai*

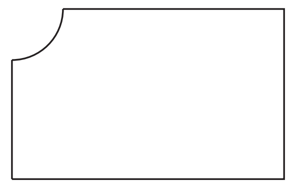


**Figure 5** The Sinai billiard.

*billiard* is very different from the one in the rectangle. The quarter circle acts like a convex mirror which spreads out the rays of light upon reflection. This effect accumulates, because the vast majority of the possible trajectories hit the quarter circle infinitely many times under different angles. This makes the motion in the Sinai billiard *classically chaotic*, while the one in the rectangle is *classically regular*. The rectangle is *separable* and *integrable*, while this feature is destroyed in the Sinai billiard. One now quantizes these billiard systems, calculates the spectra, and analyzes their statistics. Up to certain scales, the rectangle (for irrational squared ratio of the side lengths) shows Poisson behavior, the Sinai billiard yields GOE statistics.

A wealth of such empirical studies led to the Bohigas–Giannoni–Schmit conjecture. We state it here not in its original, but in a frequently used form: *spectra of systems whose classical analogues are fully chaotic show correlation properties as modeled by the Gaussian ensembles*. The Berry–Tabor conjecture is complementary: *spectra of systems whose classical analogs are fully regular show correlation properties which are often those of the Poisson type*. As far as concrete physics applications are concerned, these conjectures are well-posed. From a strict mathematical viewpoint, they have to be supplemented with certain conditions to exclude exceptions such as Artin's billiard. Due to the defnition of this system on the hyperbolic plane, its quantum version shows Poisson-like statistics, although the classical dynamics is chaotic. Up to now, no general and mathematically rigorous proofs could be given. However, *semiclassical* reasoning involving periodic orbit theory and, in particular, the *Gutzwiller trace formula*, yields at least a heuristic understanding.

Quantum chaos has been studied in numerous systems. An especially prominent example is the Hydrogen atom put in a strong magnetic field, which breaks the integrability and drives the correlations towards the GOE limit.

### Disordered and Mesoscopic Systems

An electron moving in a probe, a piece of wire, say, is scattered many times at impurities in the material. This renders the motion diffusive. In a statistical model, one writes the Hamilton operator as a sum of the kinetic part, that is, the Laplacian, and a white-noise disorder potential $V(\boldsymbol{r})$ with second moment

$$\langle V(\boldsymbol{r})V(\boldsymbol{r}')\rangle = c_V\delta^{(d)}(\boldsymbol{r}-\boldsymbol{r}') \qquad [28]$$

Here, $\boldsymbol{r}$ is the position vector in $d$ dimensions. The constant $c_V$ determines the mean free time between two scattering processes in relation to the density of states. It is assumed that phase coherence is present such that quantum effects are still significant. This defines the *mesoscopic regime*. The average over the disorder potential can be done with supersymmetry. In fact, this is the context in which supersymmetric techniques in statistical physics were developed, before they were applied to RMT models. In the case of weak disorder, the resulting field theory in superspace for two-level correlations acquires the form

$$\int d\mu(Q) f(Q)\exp(-S(Q)) \qquad [29]$$

where $f(Q)$ projects out the observable under consideration and where $S(Q)$ is the *effective Lagrangian*

$$S(Q) = -\int \mathrm{str}\Big(\mathcal{D}(\nabla Q(\boldsymbol{r}))^2 + i2rMQ(\boldsymbol{r})\Big)d^d r \qquad [30]$$

This is the *supersymmetric nonlinear $\sigma$ model*. It is used to study level correlations, but also to obtain information about the conductance and conductance fluctuations when the probe is coupled to external leads. The supermatrix field $Q(\boldsymbol{r})$ is the remainder of the disorder average, its matrix dimension is four or eight, depending on the symmetry class. This field is a Goldstone mode. It does not directly represent a particle as often the case in high-energy physics. The matrix $Q(\boldsymbol{r})$ lives in a coset space of certain supergroups. A tensor $M$ appears in the calculation, and $r$ is the energy difference on the unfolded scale, not to be confused with the position vector $\boldsymbol{r}$.

The first term in the effective Lagrangian involving a gradient squared is the *kinetic term*, it stems from the Laplacian in the Hamiltonian. The constant $\mathcal{D}$ is the classical diffusion constant for the motion of the electron through the probe. The second term is the ergodic term. In the limit of zero dimensions, $d \to 0$, the kinetic term vanishes and the remaining ergodic term yields precisely the unfolded two-level correlations of the Gaussian ensembles. Thus, RMT can be viewed as the zero-dimensional limit of field theory for disordered systems. For $d > 0$, there is a competition between the two terms. The diffusion constant $\mathcal{D}$ and the system size determine an energy scale, the *Thouless*

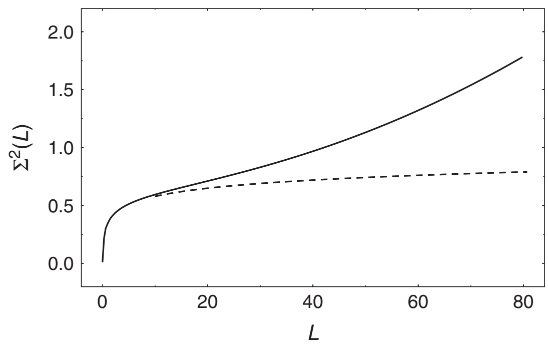



**Figure 6** Level number variance $\Sigma^2(L)$. In this example, the Thouless energy is $E_c \approx 10$ on the unfolded scale. The Gaussian ensemble behavior is dashed.

*energy* $E_c$, within which the spectral statistics is of the Gaussian ensemble type and beyond which it approaches the Poisson limit. In Figure 6, this is schematically shown for the level number variance $\Sigma^2(L)$, which bends from Gaussian ensemble to Poisson behavior when $L > E_c$. This relates to the crossover transitions in RMT. Gaussian ensemble statistics means that the electron states extend over the probe, while Poisson statistics implies their spatial localization. Hence, the Thouless energy is directly the dimensionless conductance.

A large number of issues in disordered and mesoscopic systems have been studied with the supersymmetric nonlinear $\sigma$ model. Most results have been derived for *quasi-one-dimensional* systems. Through a proper discretization, a link is established to models involving chains of random matrices. As the conductance can be formulated in terms of the scattering matrix, the experience with RMT for scattering systems can be applied and indeed leads to numerous new results.

### Quantum Chromodynamics

Quarks interact by exchanging gluons. In quantum chromodynamics, the gluons are described by gauge fields. Relativistic quantum mechanics has to be used. Analytical calculations are only possible after some drastic assumptions and one must resort to *lattice gauge theory*, that is, to demanding numerics, to study the full problem.

The massless Dirac operator has *chiral symmetry*, implying that all nonzero eigenvalues come in pairs $(-\lambda_n, +\lambda_n)$ symmetrically around zero. In chiral RMT, the Dirac operator is replaced with block off-diagonal matrices

$$W = \begin{bmatrix} 0 & W_b \\ W_b^\dagger & 0 \end{bmatrix} \qquad [31]$$

where $W_b$ is a random matrix without further symmetries. By construction, $W$ has chiral symmetry. The assumption underlying chiral RMT is that the gauge fields effectively randomize the motion of the quark. Indeed, this simple schematic model correctly reproduces low-energy sum rules and spectral statistics of lattice gauge calculations. Near the center of the spectrum, there is a direct connection to the partition function of quantum chromodynamics. Furthermore, a similarity to disordered systems exists and an analog of the Thouless energy could be found.

### Other Fields

Of the wealth of further investigations, we can mention but a few. RMT is in general useful for wave phenomena of all kinds, including classical ones. This has been shown for *elastomechanical* and *electromagnetic* resonances.

An important field of application is *quantum gravity* and matrix model aspects of *string theory*. We decided not to go into this, because the reason for the emergence of RMT concepts there is very different from everything else discussed above.

RMT is also successful beyond physics. Not surprisingly, it always received interest in mathematical statistics, but, as already said, it also relates to harmonic analysis. A connection to number theory exists as well. The high-lying zeros of the Riemann $\zeta$ function follow the GUE predictions over certain interval lengths. Unfortunately, a deeper understanding is still lacking.

As the interest in statistical concepts grows, RMT keeps finding new applications. Recently, one even started using RMT for risk management in finance.

*See also:* Arithmetic Quantum Chaos; Chaos and Attractors; Determinantal Random Fields; Free Probability Theory; Growth Processes in Random Matrix Theory; Hyperbolic Billiards; Integrable Systems in Random Matrix Theory; Integrable Systems: Overview; Number Theory in Physics; Ordinary Special Functions; Quantum Chromodynamics; Quantum Mechanical Scattering Theory; Random Partitions; Random Walks in Random Environments; Semi-Classical Spectra and Closed Orbits; Supermanifolds; Supersymmetry Methods in Random Matrix Theory; Symmetry Classes in Random Matrix Theory.

## Further Reading

Beenakker CWJ (1997) Random-matrix theory of quantum Transport. *Reviews of Modern Physics* 69: 731–808.

Bohigas O (1984) Chaotic motion and random matrix theory. In: Dehesa JS, Gomez JMG, and Polls A (eds.) *Mathematical and Computational Methods in Physics*, Lecture Notes in Physics, vol. 209. Berlin: Springer.

Brody TA, Flores J, French JB, Mello PA, Pandey A, and Wong SSM (1981) Random-matrix physics: spectrum and strength fluctuations. *Reviews of Modern Physics* 53: 385–479.

Efetov K (1997) *Supersymmetry in Disorder and Chaos*. Cambridge: Cambridge University Press.

Forrester PJ, Snaith NC, and Verbaarschot JJM (eds.) (2003) Special issue: random matrix theory. *Journal of Physics A* 36.

Guhr T, Müller-Groeling A, and Weidenmüller HA (1998) Random-matrix theories in quantum physics: common concepts. *Physics Reports* 299: 189–428.

Haake F (2001) *Quantum Signatures of Chaos*, 2nd edn. Berlin: Springer.

Mehta ML (2004) *Random Matrices*, 3rd edn. New York: Academic Press.

Stöckmann HJ (1999) *Quantum Chaos: An Introduction*. Cambridge: Cambridge University Press.

Verbaarschot JJM and Wettig T (2000) Random matrix theory and chiral symmetry in QCD. *Annual Review of Nuclear and Particle Science* 50: 343–410.


# Random Partitions

**A Okounkov**, Princeton University, Princeton, NJ, USA



## Partitions

A partition of $n$ is a monotone sequence of nonnegative integers,

$$\lambda = (\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq 0)$$

with sum $n$. The number $n$ is also denoted by $|\lambda|$ and is called the size of $n$. The number of nonzero terms in $\lambda$ is called the length of $\lambda$ and often denoted by $\ell(\lambda)$. It is convenient to make the sequence $\lambda$ infinite by adding a string of zeros at the end.

A geometric object associated to partition is its diagram. The diagram of $\lambda = (4, 2, 2, 1)$ is shown in Figure 1. A larger diagram, flipped and rotated by 135°, can be seen in Figure 2. Flipping the diagram introduces an involution on the set of partitions of $n$ known as transposition. The transposed partition is denoted by $\lambda'$.

Partitions serve as natural combinatorial labels for many basic objects in mathematics and physics. For example, partitions of $n$ index both conjugacy classes and irreducible representations of the symmetric group $S(n)$. Partitions $\lambda$ with $\ell(\lambda) \leq n$ index irreducible polynomial representations of the general linear group $GL(n)$. More generally, the highest weight of a rational representation of $GL(n)$ can be naturally viewed as two partitions of total length $\leq n$.

For an even more basic example, partitions $\lambda$ with $\lambda_1 \leq m$ and $\ell(\lambda) \leq n$ are the same as upright lattice paths making $n$ steps up and $m$ steps to the right (just follow the boundary of $\lambda$). In particular, there are $\binom{n+m}{n}$ of such. By a variation on this theme, partitions label the standard basis of fermionic Fock space (Miwa *et al.* 2000). They also label a standard basis of the bosonic Fock space.

In most instances, partitions naturally occur together with some weight function. For example, the dimension, $\dim \lambda$, of an irreducible representation of $S(n)$, or some power of it, is what always appears in harmonic analysis on $S(n)$. By a theorem of Burnside,

$$\mathfrak{M}_{\text{Planch}}(\lambda) = \frac{(\dim \lambda)^2}{n!} \qquad [1]$$

is a probability measure on the set of partitions of $n$; it is known as the Plancherel measure. Besides harmonic analysis, there are many other contexts in which it appears, for example, by a theorem of Schensted (see Sagan (2001) and Stanley (1999)), the distribution of the first part $\lambda_1$ of a Plancherel random partition $\lambda$ is the same as the distribution of the longest increasing subsequence in a uniformly random permutation of $\{1, 2, \ldots, n\}$.

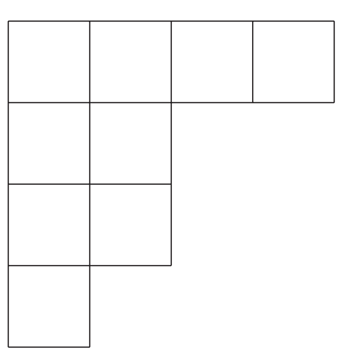


Figure 1 Diagram of a partition.

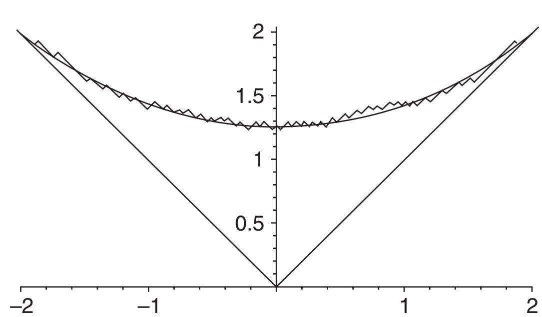



Figure 2 A Plancherel-random partition of 1000 and the limit shape.

Partitions of $n$ being just a finite set, one is often interested in letting $n \to \infty$. Even if the original problem was not of a probabilistic origin, one can still often benefit from adopting a probabilistic viewpoint because of the intuition and techniques that it brings. This is best illustrated by concrete examples, which is what we now turn to. These examples are not meant to be a panorama of random partitions. This is an old and still rapidly growing field and a simple list of all major contributions will take more space than is allowed. The books Kerov (2003), Pitman (n.d.) Sagan (2001), and Stanley (1999) offer much more information on the topics discussed below.

## Plancherel Measure

### Dimension of a Diagram

There are several formulas and interpretations for the number $\dim\lambda$ in [1]; see Sagan (2001) and Stanley (1999). The one that often appears in the context of growth processes is the following: $\dim\lambda$ is the number of ways to grow the diagram $\lambda$ from the empty diagram $\emptyset$ by adding a square at a time. That is, $\dim\lambda$ is number of chains of the form

$$\emptyset = \lambda^{(0)} \subset \lambda^{(1)} \subset \cdots \subset \lambda^{(n-1)} \subset \lambda^{(n)} = \lambda$$

where $|\lambda^{(k)}| = k$ and $\mu \subset \lambda$ means inclusion of diagrams.

From the classical formula

$$\dim\lambda = \frac{|\lambda|!}{\prod(\lambda_i + k - i)!} \prod_{i \leq j \leq k} (\lambda_i - \lambda_j + j - i) \qquad [2]$$

where $k$ is any number such that $\lambda_{k+1} = 0$, one sees that the Plancherel measure is a discrete analog of the eigenvalue density

$$\mathrm{e}^{-(1/2)\sum x_i^2} \prod_{i<j} (x_i - x_j)^2$$

of a GUE random matrix (Mehta 1991). Indeed, the first factor in [2], which looks like a multinomial coefficient, is the analog of the Gaussian weight. Kerov (2003) and Johansson were among the first to recognize the analogy between Plancherel measure and GUE. One comes across many partition sums that are discrete analogs of random matrix integrals.

The most compact formula for $\dim\lambda$ is the hook formula

$$\frac{\dim\lambda}{|\lambda|!} = \prod_{\approx \in \lambda} h(\approx)^{-1} \qquad [3]$$

Here the product is over all squares $\approx$ in the diagram of $\lambda$ and

$$h(\approx) = 1 + a(\approx) + l(\approx)$$

where $a(\approx)$ and $l(\approx)$ is the number of squares to the right of the square $\approx$ and below it, respectively. (These are known as arm-length and leg-length.)

### Limit Shape and Edge Scaling

When the diagram of $\lambda$ is very large, the logarithm of the hook product approximates a double integral. The analysis of the corresponding integral plays the central role, (see Kerov (2003), chapter 3) in the proof of the following law of large numbers for the Plancherel measure.

Take the diagram of $\lambda$, flip and rotate it as in **Figure 1** and rescale by a factor of $\sqrt{n}$ so that it has unit area. In this way one obtains a measure on continuous and, in fact, Lipschitz functions. By a result of Logan and Shepp and, independently, Vershik and Kerov these measures converge as $n \to \infty$ to the $\delta$-measure on a single function $\Omega(x)$. This limit shape for the Plancherel measure, is also plotted in **Figure 2**. Explicitly,

$$\Omega(x) = \begin{cases} \frac{2}{\pi}\left(x \arcsin(x/2) + \sqrt{4 - x^2}\right), & |x| \leq 2 \\ |x|, & |x| > 2 \end{cases}$$

This is an analog of Wigner's semicircle law (Mehta 1991) for spectra of random matrices. The Gaussian correction to the limit shape was also found by Kerov (2003).

The limit shape result can be refined to show that $\lambda_1/\sqrt{n} \to 2$ in probability. Together with Schensted's theorem, this answers the question posed by Ulam about the longest increasing subsequence in a random permutation. Further progress came in the work of Baik, Deift, and Johansson (see Deift (2000)), who conjectured (and proved for $i = 1$ and 2) that as $n \to \infty$ the joint distribution

$$\frac{\lambda_i - 2\sqrt{n}}{n^{1/6}}, \quad i = 1, 2, \ldots$$

becomes exactly the same as the distribution of largest eigenvalues of a GUE random matrix. In particular, the longest increasing subsequence, suitably scaled, is distributed exactly like the largest eigenvalue. The distribution of the latter is known as the Tracy–Widom distribution; it is given in terms of a particular solution of the Painlevé II equation. For more information about the proof of the full conjecture, see Aldous and Diaconis (1999), Deift (2000), and Okounkov (2002).

### Correlation Functions

One way to prove the full BDJ conjecture is to use the following exact formula first obtained in a more general setting by Borodin and Olshanski (see Olshanski (2003), and Okounkov (2002) for further generalizations). Look at the downsteps of the zig-zag curve in Figure 2. The $x$-coordinates of their midpoints are the numbers

$$\mathfrak{S}(\lambda) = \left\{\lambda_i - \mathrm{i} + \tfrac{1}{2}\right\} \subset \mathbb{Z} + \tfrac{1}{2} \qquad [4]$$

The map $\lambda \mapsto \mathfrak{S}(\lambda)$ makes a random partition a random subset of $\mathbb{Z} + \frac{1}{2}$, that is, a random point field on a lattice. These random points should be treated like eigenvalues of a random matrix. In particular, it is natural to consider their correlations, that is, the probability that $X \subset \mathfrak{S}(\lambda)$ for some fixed $X \subset \mathbb{Z} + \frac{1}{2}$.

Many formulas work better if we replace the Plancherel measures $\mathfrak{M}_{\mathrm{Planch},n}$ on partitions of a fixed number $n$ by their Poisson average,

$$\mathfrak{M}_{\xi} = \mathrm{e}^{-\xi} \sum_{n \geq 0} \frac{\xi^n}{n!} \mathfrak{M}_{\mathrm{Planch},n}$$

Here $\xi > 0$ is a parameter. It equals the expected size of $\lambda$. For any finite set $X$, we have

$$\mathrm{Prob}_{\xi}(X \subset \mathfrak{S}(\lambda)) = \det\left[\mathbb{K}_{\mathrm{Bessel}}(x_i, x_j; \xi)\right]_{x_i, x_j \in X} \qquad [5]$$

where $\mathbb{K}_{\mathrm{Bessel}}$ is the discrete Bessel kernel given by

$$\mathbb{K}_{\mathrm{Bessel}}(x, y; \xi) = \sqrt{\xi}\, \frac{J_{x-1/2}(2\sqrt{\xi}) J_{y+1/2}(2\sqrt{\xi}) - J_{x+1/2}(2\sqrt{\xi}) J_{y-1/2}(2\sqrt{\xi})}{x - y}$$

Note that only Bessel function of integral order enter this formula.

For large argument $\xi$, $J_n(2\sqrt{\xi})$ has sine asymptotics if $n \ll 2\sqrt{\xi}$ and Airy function asymptotics if $n \approx 2\sqrt{\xi}$. Consequently, one gets the random matrix behavior near the edge of the limit shape and discrete sine kernel asymptotics of correlations in the bulk of the limit shape.

### Permutation Enumeration

A basic combinatorial problem is to count permutations $\sigma_1, \ldots, \sigma_p \in S(n)$ of given cycle types $\mu^{(1)}, \ldots, \mu^{(p)}$ such that

$$\sigma_1 \cdots \sigma_p = 1 \qquad [6]$$

A geometric interpretation of this problem is to count covers of the sphere $S^2 = \mathbb{CP}^1$ branched over $p$ given points with monodromy $\mu^{(1)}, \ldots, \mu^{(p)}$. Elementary character theory of $S(n)$ gives (Jones 1998)

$$\#\{\sigma_i \in C_{\mu^{(i)}}, \prod \sigma_i = 1\} = \left\langle \prod f_{\mu^{(i)}} \right\rangle_{\mathrm{Planch}} \qquad [7]$$

where $C_\mu$ is the conjugacy class with cycle type $\mu$ and

$$f_\mu(\lambda) = |C_\mu| \frac{\chi^\lambda_\mu}{\dim \lambda}$$

is the central character of the irreducible representation $\lambda$. Here $\chi^\lambda_\mu$ is the character of any $\sigma \in C_\mu$ in the representation $\lambda$.

Let $\mu$ be of the form $(\bar{\mu}, 1, 1, \ldots)$ with $\bar{\mu}$ fixed. By a result of Kerov and Olshanski, $\binom{n}{|\bar{\mu}|}^{-1} f_\mu(\lambda)$, is a polynomial in $\lambda$ of degree $|\bar{\mu}|$. See [11] for the simplest example $\bar{\mu} = (2)$, that is, for the central character of a transposition. We thus recognize in [7] a discrete analog of the GUE expectation of a polynomial in traces of a random matrix. This analogy becomes even clearer in the Gromov–Witten theory of $\mathbb{CP}^1$, which can be viewed as taking into account contributions of certain degenerate covers, see Okounkov (2002).

There is a generalization, due to Burnside, of [7] to counting branched covers of surfaces of any genus $g$; see Jones (1998). The only modification required is that a representation $\lambda$ is now counted with the weight $(\dim \lambda)^{2-2g}$. For example, covers of the torus correspond to a uniform measure on partitions. In particular, the probability that two random permutation from $S(n)$ commute is $p(n)/n!$, where $p(n)$ is the number of partitions of $n$.

## Generalizations of Plancherel Measure

### Schur Functions and Cauchy Identity

Schur functions $s_\lambda(x_1, \ldots, x_n)$, where $\lambda$ is a partition with at most $n$ parts, form a distinguished linear basis of the algebra of symmetric polynomials in $x_1, \ldots, x_n$. Various definitions and many remarkable properties of these function are discussed in, for example, Sagan (2001) and Stanley (1999). One of them is that $s_\lambda(x)$ is the trace of a matrix with eigenvalues $\{x_i\}$ in an irreducible $\mathrm{GL}(n)$ module with highest weight $\lambda$. The following stability of $s_\lambda$,

$$s_\lambda(x_1, \ldots, x_n, 0) = s_\lambda(x_1, \ldots, x_n), \quad \ell(\lambda) \leq n$$

allows one to define Schur functions in infinitely many variables. The formulas

$$p_\mu = \sum_\lambda \chi^\lambda_\mu s_\lambda, \qquad s_\lambda = \sum_\mu \frac{\chi^\lambda_\mu}{\mathfrak{z}(\mu)} p_\mu$$

where

$$\mathfrak{z}(\mu) = \frac{|\mu|!}{|C_\mu|} = |\mathrm{Aut}(\mu)| \prod \mu_i$$

establish the transition between the basis of Schur function and the basis of power sum functions

$$p_\mu = \prod p_{\mu_i}, \qquad p_k = \sum_i x_i^k$$

In particular, the dimension function $\dim \lambda$ is the following specialization of the Schur function:

$$\frac{\dim \lambda}{|\lambda|!} = s_\lambda \big|_{p_1=1,\, p_2=p_3=\cdots=0}$$

We will discuss other important specializations of Schur functions later.

A typical situation in which a random matrix integral can be reduced to a sum over partition is when one uses the Cauchy identity

$$\begin{aligned} \frac{1}{\prod(1-x_i y_i)} &= \exp\left(\sum \frac{p_k(x)p_k(y)}{k}\right) \\ &= \sum_\lambda s_\lambda(x)s_\lambda(y) \end{aligned} \qquad [8]$$

to expand the integrand in Schur function and integrate term by term using, for example, the orthogonality of characters or the identity

$$\int_{\mathrm{U}(n)} s_\lambda(AgBg^{-1})\, \mathrm{d}g = \frac{1}{\dim_n \lambda} s_\lambda(A)s_\lambda(B) \qquad [9]$$

Here $s_\lambda(A)$ denotes the Schur function in eigenvalues of a matrix $A$, $\mathrm{d}g$ is the normalized Haar measure on the unitary group $\mathrm{U}(n)$, and

$$\dim_n \lambda = s_\lambda(\underbrace{1,\ldots,1}_{n \text{ times}})$$

is the dimension of irreducible $\mathrm{GL}(n)$ module $V^\lambda$ with highest weight $\lambda$. The meaning of [9] is that normalized characters are algebra homomorphisms from the center of the group algebra of $\mathrm{U}(n)$ to numbers. This method of converting a random matrix problem to a random partition problem is known as character expansion (see, e.g., Kazakov (2001)).

Inspired by the Cauchy identity, one can generalize Plancherel measure to

$$\mathfrak{M}_{\mathrm{Schur}} = \prod(1-x_i y_i)s_\lambda(x)s_\lambda(y)$$

where $x$ and $y$, or, equivalently, $p_k(x)$ and $p_k(y)$, are viewed as parameters. This is known as the Schur measure. If $p_1(x)=p_2(y)=\sqrt{\xi}$ and all other $p_k$'s vanish, we get $\mathfrak{M}_{\mathrm{Schur}} = \mathfrak{M}_\xi$. Many properties of the Plancherel measure can be generalized to Schur measure, in particular, exact formulas for correlation functions, description of the limit shape, etc. (Okounkov 2002).

### Dimension Functions

We already met the function $\dim_n \lambda$. There is a useful formula

$$\dim_n \lambda = \prod_{\square\in\lambda} \frac{n+c(\square)}{h(\square)} \qquad [10]$$

where $c((i,j)) = j - i$ is the content of the square $\square$ in $i$th row and $j$th column. From [10] it is clear that $\dim_n$ makes sense for arbitrary complex values of $n$. The corresponding specializations of the Schur measure

$$x = \underbrace{\sqrt{\xi},\ldots,\sqrt{\xi}}_{z \text{ times}}, \qquad y = \underbrace{\sqrt{\xi},\ldots,\sqrt{\xi}}_{z' \text{ times}}$$

where $\xi, z, z'$ are parameters, are related to the so-called Z-measures and their theory is much-developed (Olshanski 2003). As $z, z', \xi^{-1} \to \infty$ in such a way that $zz'\xi \to \xi_0$, we get $\mathfrak{M}_{\xi_0}$ in the limit.

The enumerative problems discussed in the section "Permutation enumeration" have analogs for the unitary groups $\mathrm{U}(n)$ and, suitably interpreted, the answers are the same with the dimension $\dim_n \lambda$ replacing $\dim \lambda$. For example, instead of counting the solutions to [6], one may be interested in the volume of the set of $p$-tuples of unitary matrices with given eigenvalues that multiply to 1. Geometrically, such data arise as the monodromy of a flat unitary connection over $S^2 \setminus \{p \text{ points}\}$, which is a $\mathrm{U}(n)$ analog of a branched cover. The analog of Burnside's formula is Witten's formula for the volumes of moduli spaces of flat connections on a genus $g$ surface with given holonomy around $p$ punctures, (see, e.g., Witten (1991) and Woodward (2004)). It involves summing normalized characters over all representations $V^\lambda$, not necessarily polynomial, with the weight $(\dim V^\lambda)^{2-2g}$. If additionally weighted by a Gaussian of the form $\exp(-A(f_2(\lambda) + (n/2)|\lambda|))$, where

$$\begin{aligned} f_2(\lambda) &= \frac{1}{2}\sum_i \left[\left(\lambda_i - i + \tfrac{1}{2}\right)^2 - \left(-i + \tfrac{1}{2}\right)^2\right] \\ &= \sum_{\square\in\lambda} c(\square) \end{aligned} \qquad [11]$$

this becomes Migdal's formula for the partition function of the 2D Yang–Mills theory, the positive constant $A$ being the area of the surface (see, e.g., Witten (1991) and Woodward (2004)).

A further generalization naturally arising in the theory of quantum groups is the quantum dimension

$$\begin{aligned} \dim_{n,q} \lambda &= s_\lambda(q^{1-n}, q^{3-n}, \ldots, q^{n-3}, q^{n-1}) \\ &= \prod_{\square\in\lambda} \frac{q^{n+c(\square)} - q^{-n-c(\square)}}{q^{h(\square)} - q^{-h(\square)}} \end{aligned}$$

where $q$ is a parameter (it is more common to use $\dim_{n,q^{1/2}}$ instead). Obviously, $\dim_{n,q} \to \dim_n$ as $q \to 1$. The function $\dim_{n,q}$ is an important building block of, for example, quantum invariants of knots and 3-folds, and various related objects (see, e.g., Bakalov and Kirillov (2001)). The Verlinde formula (Bakalov and Kirillov 2001) can be viewed as an analog of Burnside's formula with weight $\dim_{n,q}$. When $q$ is a root of unity the summation over $\lambda$ is naturally truncated to a finite sum.

The next level of generalization is obtained by deforming Schur function to Jack and, more generally, Macdonald symmetric functions (Macdonald 1995). In particular, the Jack polynomial analog of the Plancherel measure is

$$\mathfrak{M}_{\text{Jack}}(\lambda) = \prod_{\square \in \lambda} \frac{n!(t_1 t_2)^n}{((a(\square)+1)t_1 + l(\square)t_2)(a(\square)t_1 + (l(\square)+1)t_2)}$$

where $t_1, t_2$ are parameters, and $a(\square)$ and $l(\square)$ denote, as above, the arm- and leg-length of a square $\square$. This measure depends only on the ratio $t_2/t_1$ which is the usual parameter of Jack polynomials. To continue the analogy with random matrices, this should be viewed as a general $\beta$ analog of the Plancherel measure.

The measure $\mathfrak{M}_{\text{Jack}}$ naturally arises in Atiyah–Bott localization computations on the Hilbert scheme of $n$ points in $\mathbb{C}^2$. By definition, this Hilbert scheme parametrizes ideals $I \subset \mathbb{C}[x,y]$ of codimension $n$ as linear spaces. The torus $(\mathbb{C}^*)^2$ acts on it by rescaling $x$ and $y$ and the fixed points of this action are

$$I_\lambda = \text{Span of } \{x^{j-1} y^{i-1}\}_{(i,j) \notin \lambda}$$

where $\lambda$ is a partition of $n$. The weight of this fixed point in the Atiyah–Bott formula is proportional to $\mathfrak{M}_{\text{Jack}}(\lambda)$, the parameters $t_1$ and $-t_2$ being the standard torus weights. Corresponding formulas in $K$-theory involve a Macdonald polynomial analog of $\dim \lambda$.

Nekrasov defines the partition functions of $\mathcal{N}=2$ supersymmetric gauge theories by formally applying the Atiyah–Bott localization formula to (noncompact) instanton moduli spaces. The resulting expression is a sum over partitions with a weight which is a generalization of $\mathfrak{M}_{\text{Jack}}$. In this way, random partitions enter gauge theory. What is more, statistical properties of these random partitions are reflected in the dynamics of gauge theories. For example, the limit shape turns out to be precisely the Seiberg–Witten curve (see Nekrasov and Okounkov (2003), Okounkov (2002), and also Nakajima and Yoshioka (2003)).

## Harmonic Functions on Young Graph

### Definitions

Partitions form a natural directed graph $\mathbb{Y}$, known as Young graph, in which there is an edge from $\mu$ to $\lambda$ if $\lambda$ is obtained from $\mu$ by adding a square. We will denote this by $\mu \nearrow \lambda$. Let $\kappa$ be a non-negative function (called multiplicity) on edges of $\mathbb{Y}$. A function $\phi$ on the vertices of $\mathbb{Y}$ is harmonic if it satisfies

$$\phi(\mu) = \sum_{\lambda \nwarrow \mu} \kappa(\mu, \lambda) \phi(\lambda) \qquad [12]$$

for any $\mu$. For given edge multiplicities $\kappa$, non-negative harmonic functions normalized by $\phi(\emptyset) = 1$ form a convex compact (with respect to pointwise convergence) set, which we will denote by $\boldsymbol{H}(\kappa)$. The extreme points of $\boldsymbol{H}(\kappa)$ are the indecomposable or ergodic harmonic functions. They are the most important ones. One defines

$$\dim_\kappa \lambda/\mu = \sum_{\mu=\nu_0 \nearrow \nu_1 \nearrow \cdots \nearrow \nu_{|\lambda|-|\mu|}=\lambda} \prod \kappa(\nu_i, \nu_{i+1})$$

and $\dim_\kappa \lambda = \dim_\kappa \lambda/\emptyset$. For example, if $\kappa \equiv 1$ then $\dim_\kappa \lambda = \dim \lambda$. Any function $\phi \in \boldsymbol{H}(\kappa)$ defines a probability measure on partitions of fixed size $n, n = 0, 1, 2, \ldots$, by

$$\mathfrak{M}_{\phi,n}(\lambda) = \phi(\lambda) \dim_\kappa \lambda, \quad |\lambda| = n \qquad [13]$$

The mean value property [12] implies a certain coherence of these measures for different values of $n$, which, in general, does not hold for measures like $\mathfrak{M}_{\text{Schur}}$. Two multiplicity functions $\kappa$ and $\kappa'$ are gauge equivalent if

$$\kappa'(\mu, \lambda) = f(\mu) \kappa(\mu, \lambda) f(\lambda)^{-1}$$

for some function $f$. In this case, $\boldsymbol{H}(\kappa)$ and $\boldsymbol{H}(\kappa')$ are naturally isomorphic and the measures $\mathfrak{M}_\phi$ are the same.

### First Example: Thoma Theorem

Let $F$ be a central function on the infinite symmetric group $S(\infty) = \bigcup_n S(n)$, normalized by $F(1) = 1$. Restricted to $S(n)$, $F$ is a linear combination of irreducible characters

$$F|_{S(n)} = \sum_{|\lambda|=n} \phi(\lambda) \chi^\lambda$$

The branching rule $\chi^\lambda|_{S(n-1)} = \sum_{\mu \nearrow \lambda} \chi^\mu$ implies that the Fourier coefficients $\phi$ are harmonic with respect to $\kappa \equiv 1$. They are non-negative if and only if $F$ is a positive-definite function on $S(\infty)$, which means that the matrix $(F(g_i g_j^{-1}))$ is non-negative definite for any $\{g_i\} \subset S(\infty)$. The description of all indecomposable positive-definite central functions on $S(\infty)$ was first

obtained by Thoma (see Kerov (2003, 1998) and Olshanski (2003)). Rephrased in our language, it says that the functions

$$\phi(\lambda) = s_\lambda|_{p_1=1, p_k=\sum \alpha_i^k + (-1)^{k+1}\sum \beta_i^k, k>1}$$

are the extreme points of $\boldsymbol{H}(1)$. Here $\alpha_i$ and $\beta_i$ are parameters satisfying

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq 0, \qquad \beta_1 \geq \beta_2 \geq \cdots \geq 0$$
$$\sum \alpha_i + \beta_i \leq 1$$

This set is known as the Thoma simplex. The origin $\alpha_i = \beta_i = 0$ corresponds to the Plancherel measure.

A general positive-definite central functions on $S(\infty)$ defines a measure on the Thoma simplex. This measure can be interpreted as a point process on the real line, for example, by placing particles at positions $\{\alpha_i\}$ and $\{-\beta_i\}$. Interesting central functions lead to interesting processes (see Olshanski (2003)).

### Second Example: Kingman Theorem

Let $\Pi$ be a partition of the naturals $\mathbb{N}$ into disjoint subsets. For any $n = 1, 2, \ldots, \Pi$ defines the induced partition $\Pi_n$ of $\{1, \ldots, n\}$ and hence a partition $\lambda(\Pi_n)$ of the number $n$. A measure $\mathcal{M}$ on partitions $\Pi$ is called exchangeable if

$$\mathcal{M}(\Pi_n) = \phi(\lambda(\Pi_n)), \quad n = 1, 2, \ldots$$

for some function $\phi$ on $\mathbb{Y}$. This implies that $\phi$ is harmonic for

$$\kappa_{\mathrm{K}}(\mu, \lambda) = \rho_k$$

where $\mu = 1^{\rho_1} 2^{\rho_2} \cdots$ and $\lambda = 1^{\rho_1} 2^{\rho_2} \cdots k^{\rho_k - 1} (k+1)^{\rho_{k+1}+1} \cdots$. The description of all exchangeable measures $\mathcal{M}$ was first obtained by Kingman. In our language, it says that the extreme points of $\boldsymbol{H}(\kappa_{\mathrm{K}})$ are

$$\phi(\lambda) = m_\lambda|_{p_1=1, p_k=\sum \alpha_i^k, k>1}$$

where $m_\lambda$ is the monomial symmetric function (sum of all monomials with exponents $\lambda$) and $\alpha_i$ are parameters as before. The corresponding measure $\mathcal{M}_\alpha$ can be described as follows. Let $X_i$ be a sequence of independent, identically distributed random variables such that $\{\alpha_i\}$ are the measures of atoms of their distribution. This defines a random partition $\Pi$ of $\mathbb{N}$ by putting $i$ and $j$ in the same block of $\Pi$ if and only if $X_i = X_j$. A general exchangeable measure $\mathcal{M}$ is then a convex linear combination of $\mathcal{M}_\alpha$, which can be viewed as making the common distribution of $X_i$ also random. See Pitman (n.d.) for a lot more about Kingman's theorem.

The multiplicities $\kappa_{\mathrm{K}}$ are gauge equivalent to multiplicities

$$\kappa_{\mathrm{VP}}(\mu, \lambda) = k\rho_k \qquad [14]$$

which arise in the study of probability measures on virtual permutations $\mathfrak{S}$ (Olshanski 2003). By definition,

$$\mathfrak{S} = \varprojlim S(n)$$

with respect to the maps $S(n) \to S(n-1)$ that delete $n$ from the disjoint cycle decomposition of a permutation $\sigma \in S(n)$. For $n \geq 5$, this is the unique map that commutes with the right and left action of $S(n-1)$. Thus, $\mathfrak{S}$ has a natural $S(\infty) \times S(\infty)$ action; however, it is not a group. A measure $\mathcal{M}$ on $\mathfrak{S}$ is central if it is invariant under the action of the diagonal subgroup in $S(\infty) \times S(\infty)$. Let the push-forward of $\mathcal{M}$ to $S(n)$ give mass $\phi(\lambda)$ to a permutation with cycle type $\lambda$. It is then easy to see that $\phi$ is harmonic with respect to [14]. Thus, Kingman's theorem gives a description of ergodic central measures on $\mathfrak{S}$. For example, $\alpha_i = 0$ corresponds to the $\delta$-measure at the identity.

### Ergodic Method

A unified approach to this type of problems was proposed and developed by Vershik and Kerov. It is based on the following ergodic theorem. Let $\phi$ be an ergodic harmonic function. Then

$$\phi(\mu) = \lim \frac{\dim_\kappa \lambda/\mu}{\dim_\kappa \lambda}, \quad |\lambda| \to \infty \qquad [15]$$

for almost all $\lambda$ with respect to the measure [13] (Kerov 2003). This is similar to approximating a Gibbs measure in infinite volume by a sequence of finite-volume Gibbs measures with appropriate boundary conditions. The ratio on the RHS of [15] is known as the Martin kernel. Its asymptotics as $|\lambda| \to \infty$ plays the essential role.

Let us call a sequence $\{\lambda(n)\}$ of partitions of $n$ regular if the limit in [15] exists for all $\mu$. For $\kappa \equiv 1$, Vershik and Kerov proved that $\{\lambda(n)\}$ is regular if and only if the following limits exist:

$$\frac{\lambda(n)_i}{n} \to \alpha_i, \qquad \frac{\lambda(n)'_i}{n} \to \beta_i \qquad [16]$$

that is, if the rows and columns of $\lambda(n)$, scaled by $n$, have a limit. In this case, the limit in [15] is the harmonic function with Thoma parameters $\alpha_i$ and $\beta_i$. This simultaneously proves Thoma classification and gives a law of large numbers for the corresponding measures [13]. It also gives a transparent geometric interpretation of Thoma parameters. Note that the behavior [16] is very different from the formation of a smooth limit shape that we saw earlier. For a common generalization of this result and Kingman's theorem see Kerov (1998).

*See also:* Determinantal Random Fields; Growth Processes in Random Matrix Theory; Integrable Systems in Random Matrix Theory; Random Matrix Theory in Physics; Symmetry Classes in Random Matrix Theory.

## Further Reading


Aldous D and Diaconis P (1999) Longest increasing subsequences: from patience sorting to the Baik–Deift–Johansson theorem. *Bulletin of the American Mathematical Society* 36(4): 413–432.

Bakalov B and Kirillov A Jr. (2001) Lectures on Tensor Categories and Modular Functors. *University Lecture Series,* vol. 21. American Mathematical Society.

Deift P (2000) Integrable systems and combinatorial theory. *Notices of the American Mathematical Society* 47(6): 631–640.

Jones G (1998) *Characters and Surfaces: A Survey*, The Atlas of Finite Groups: Ten Years on (Birmingham, 1995), London Mathematical Society Lecture Note Series, vol. 249, pp. 90–118. Cambridge: Cambridge University Press.

Kazakov V (2001) Solvable Matrix Models, *Random Matrices and Their Applications,* vol. 40. MSRI Publications; Cambridge: Cambridge University Press.

Kerov S (2003) *Asymptotic Representation Theory of the Symmetric Group and its Applications in Analysis*. American Mathematical Society.

Kerov S, Okounkov A, and Olshanski G (1998) The boundary of the Young graph with Jack edge multiplicities. *International Mathematics Research Notices* 4: 173–199.

Macdonald IG (1995) *Symmetric Functions and Hall Polynomials*. Oxford: Clarendon.

Mehta ML (1991) *Random Matrices*, 2nd edn. Boston, MA: Academic Press.

Miwa T, Jimbo M, and Date E (2000) *Solitons. Differential Equations, Symmetries and Infinite-Dimensional Algebras*. Cambridge: Cambridge University Press.

Nakajima H and Yoshioka K (2003) Lectures on instanton counting, math.AG/0311058.

Nekrasov N and Okounkov A (2003) Seiberg–Witten theory and random partitions, hep-th/0306238.

Okounkov A (2002) Symmetric Functions and Random Partitions, *Symmetric Functions 2001: Surveys of Developments and Perspectives*, pp. 223–252, NATO Sci. Ser. II Math. Phys. Chem., 74. (math.CO/0309074). Dordrecht: Kluwer Academic.

Olshanski G (2003) An introduction to harmonic analysis on the infinite symmetric group, math.RT/0311369.

Okounkov A (2002) The uses of random partitions, math-ph/0309015.

Pitman J (n.d.) *Combinatorial Stochastic Processes*, Lecture Notes from St. Four Course, available from www.stat.berekeley.edu.

Sagan B (2001) The Symmetric Group. Representations, Combinatorial Algorithms, and Symmetric Functions. *Graduate Texts in Mathematics*, 2nd edn., vol. 203, New York: Springer.

Stanley R (1999) *Enumerative Combinatorics, II*. Cambridge: Cambridge University Press.

Witten E (1991) On quantum gauge theories in two dimensions. *Communications in Mathematical Physics* 141(1): 153–209.

Woodward C (2004) Localization for the norm-square of the moment map and the two-dimensional Yang–Mills integral, math.SG/0404413.


# Random Walks in Random Environments

**L V Bogachev**, University of Leeds, Leeds, UK



## Introduction

Random walks provide a simple conventional model to describe various transport processes, for example, propagation of heat or diffusion of matter through a medium (for a general reference see, e.g., Hughes (1995)). However, in many practical cases, the medium where the system evolves is highly irregular, due to factors such as defects, impurities, fluctuations, etc. It is natural to model such irregularities as "random environment," treating the observable sample as a statistical realization of an ensemble, obtained by choosing the local characteristics of the motion (e.g., transport coefficients and driving fields) at random, according to a certain probability distribution.

In the random walks context, such models are referred to as "random walks in random environments" (RWRE). This is a relatively new chapter in applied probability and physics of disordered systems initiated in the 1970s. Early interest in RWRE models was motivated by some problems in biology, crystallography, and metal physics, but later applications have spread through numerous areas (see review papers by Alexander *et al.* (1981), Bouchaud and Georges (1990), and a comprehensive monograph by Hughes (1996)). After 30 years of extensive work, RWRE remain a very active area of research, which has been a rich source of hard and challenging questions and has already led to many surprising discoveries, such as subdiffusive behavior, trapping effects, localization, etc. It is fair to say that the RWRE paradigm has become firmly established in physics of random media, and its models, ideas, methods, results, and general effects have become an indispensable part of the standard tool kit of a mathematical physicist.

One of the central problems in random media theory is to establish conditions ensuring homogenization, whereby a given stochastic system evolving in a random medium can be adequately described, on some spatial–temporal scale, using a suitable effective system in a homogeneous (nonrandom) medium. In particular, such systems would exhibit classical diffusive behavior with effective drift and diffusion coefficient. Such an approximation, called "effective medium approximation" (EMA), may be expected to

be successful for systems exposed to a relatively small disorder of the environment. However, in certain circumstances, EMA may fail due to atypical environment configurations ("large deviations") leading to various anomalous effects. For instance, with small but positive probability a realization of the environment may create "traps" that would hold the particle for an anomalously long time, resulting in the subdiffusive behavior, with the mean square displacement growing slower than linearly in time.

RWRE models have been studied by various nonrigorous methods including Monte Carlo simulations, series expansions, and the renormalization group techniques (see more details in the above references), but only a few models have been analyzed rigorously, especially in dimensions greater than one. The situation is much more satisfactory in the one-dimensional case, where the mathematical theory has matured and the RWRE dynamics has been understood fairly well.

The goal of this article is to give a brief introduction to the beautiful area of RWRE. The principal model to be discussed is a random walk with nearest-neighbor jumps in independent and identically distributed (i.i.d.) random environment in one dimension, although we shall also comment on some generalizations. The focus is on rigorous results; however, heuristics will be used freely to motivate the ideas and explain the approaches and proofs. In a few cases, sketches of the proofs have been included, which should help appreciate the flavor of the results and methods.

## Ordinary Random Walks: A Reminder

To put our exposition in perspective, let us give a brief account of a few basic concepts and facts for ordinary random walks, that is, evolving in a nonrandom environment (see further details in Hughes (1995)). In such models, space is modeled using a suitable graph, for example, a $d$-dimensional integer lattice $\mathbb{Z}^d$, while time may be discrete or continuous. The latter distinction is not essential, and in this article we will mostly focus on the discrete-time case. The random mechanism of spatial motion is then determined by the given transition probabilities (probabilities of jumps) at each site of the graph. In the lattice case, it is usually assumed that the walk is translation invariant, so that at each step distribution of jumps is the same, with no regard to the current location of the walk.

In one dimension ($d=1$), the simple (nearest-neighbor) random walk may move one step to right or to the left at a time, with some probabilities $p$ and $q=1-p$, respectively. An important assumption is that only the current location of the walk determines the random motion mechanism, whereas the past history is not relevant. In terms of probability theory, such a process is referred to as "Markov chain." Thus, assuming that the walk starts at the origin, its position after $n$ steps can be represented as the sum of consecutive displacements, $X_n = Z_1 + \cdots + Z_n$, where $Z_i$ are independent random variables with the same distribution $P\{Z_i = 1\} = p$, $P\{Z_i = -1\} = q$.

The strong law of large numbers (LLN) states that almost surely (i.e., with probability 1)

$$\lim_{n\to\infty} \frac{X_n}{n} = EZ_1 = p - q, \quad P\text{-a.s.} \qquad [1]$$

where $E$ denotes expectation (mean value) with respect to $P$. This result shows that the random walk moves with the asymptotic average velocity close to $p-q$. It follows that if $p-q \neq 0$, then the process $X_n$, with probability 1, will ultimately drift to infinity (more precisely, $+\infty$ if $p-q>0$ and $-\infty$ if $p-q<0$). In particular, in this case, the random walk may return to the origin (and in fact visit any site on $\mathbb{Z}$) only finitely many times. Such behavior is called "transient." However, in the symmetric case (i.e., $p=q=0.5$) the average velocity vanishes, so the above argument fails. In this case, the walk behavior appears to be more complicated, as it makes increasingly large excursions both to the right and to the left, so that $\overline{\lim}_{n\to\infty} X_n = +\infty$, $\underline{\lim}_{n\to\infty} X_n = -\infty$ ($P$-a.s.). This implies that a symmetric random walk in one dimension is "recurrent," in that it visits the origin (and indeed any site on $\mathbb{Z}$) infinitely often. Moreover, it can be shown to be "null-recurrent," which means that the expected time to return to the origin is infinite. That is to say, return to the origin is guaranteed, but it takes very long until this happens.

Fluctuations of the random walk can be characterized further via the central limit theorem (CLT), which amounts to saying that the probability distribution of $X_n$ is asymptotically normal, with mean $n(p-q)$ and variance $4npq$:

$$\lim_{n\to\infty} P\left\{\frac{X_n - n(p-q)}{\sqrt{4npq}} \le x\right\} = \Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-y^2/2}\,\mathrm{d}y \qquad [2]$$

These results can be extended to more general walks in one dimension, and also to higher dimensions. For instance, the criterion of recurrence for a general one-dimensional random walk is that it is unbiased, $E(X_1 - X_0) = 0$. In the two-dimensional case, in addition one needs $E|X_1 - X_0|^2 < \infty$. In higher dimensions, any random walk (which does not reduce to lower dimension) is transient.

### Random Environments and Random Walks

The definition of an RWRE involves two ingredients: (1) the environment, which is randomly chosen but remains fixed throughout the time evolution, and (2) the random walk, whose transition probabilities are determined by the environment. The set of environments (sample space) is denoted by $\Omega=\{\omega\}$, and we use $\mathbb{P}$ to denote the probability distribution on this space. For each $\omega\in\Omega$, we define the random walk in the environment $\omega$ as the (time-homogeneous) Markov chain $\{X_t,\ t=0,1,2,\ldots\}$ on $\mathbb{Z}^d$ with certain (random) transition probabilities

$$p(x,y,\omega)=P^\omega\{X_1=y|X_0=x\} \qquad [3]$$

The probability measure $P^\omega$ that determines the distribution of the random walk in a given environment $\omega$ is referred to as the "quenched" law. We often use a subindex to indicate the initial position of the walk, so that, for example, $P_x^\omega\{X_0=x\}=1$.

By averaging the quenched probability $P_x^\omega$ further, with respect to the environment distribution, we obtain the "annealed" measure $\boldsymbol{P}_x=\mathbb{P}\times P_x^\omega$, which determines the probability law of the RWRE:

$$\boldsymbol{P}_x(A)=\int_\Omega P_x^\omega(A)\ \mathbb{P}(\mathrm{d}\omega)=\mathbb{E}P_x^\omega(A) \qquad [4]$$

Expectation with respect to the annealed measure $\boldsymbol{P_x}$ will be denoted by $\boldsymbol{E_x}$.

Equation [4] implies that if some property $A$ of the RWRE holds almost surely with respect to the quenched law $P_x^\omega$ for almost all environments (i.e., for all $\omega\in\Omega'$ such that $\mathbb{P}(\Omega')=1$), then this property is also true with probability 1 under the annealed law $\boldsymbol{P}_x$.

Note that the random walk $X_n$ is a Markov chain only conditionally on the fixed environment (i.e., with respect to $P_x^\omega$), but the Markov property fails under the annealed measure $\boldsymbol{P}_x$. This is because the past history cannot be neglected, as it tells what information about the medium must be taken into account when averaging with respect to environment. That is to say, the walk learns more about the environment by taking more steps. (This idea motivates the method of "environment viewed from the particle," see related section below.)

The simplest model is the nearest-neighbor one-dimensional walk, with transition probabilities

$$p(x,y,\omega)=\begin{cases} p_x & \text{if } y=x+1\\ q_x & \text{if } y=x-1\\ 0 & \text{otherwise}\end{cases}$$

where $p_x$ and $q_x=1-p_x$ $(x\in\mathbb{Z})$ are random variables on the probability space $(\Omega,\mathbb{P})$. That is to say, given the environment $\omega\in\Omega$, the random walk currently at point $x\in\mathbb{Z}$ will make a one-unit step to the right, with probability $p_x$, or to the left, with probability $q_x$. Here the environment is determined by the sequence of random variables $\{p_x\}$. For most of the article, we assume that the random probabilities $\{p_x, x\in\mathbb{Z}\}$ are i.i.d., which is referred to as "i.i.d. environment." Some extensions to more general environments will be mentioned briefly in the section "Some generalizations and variations." The study of RWRE is simplified under the following natural condition called "(uniform) ellipticity:"

$$0<\delta\le p_x\le 1-\delta<1,\quad x\in\mathbb{Z},\ \mathbb{P}\text{-a.s.} \qquad [5]$$

which will be frequently assumed in the sequel.

## Transience and Recurrence

In this section, we discuss a criterion for the RWRE to be transient or recurrent. The following theorem is due to Solomon (1975).

**Theorem 1** *Set $\rho_x:=q_x/p_x$, $x\in\mathbb{Z}$, and $\eta:=\mathbb{E}\ln\rho_0$.*

(i) *If $\eta\neq 0$ then $X_t$ is transient ($\boldsymbol{P}_0$-a.s.); moreover, if $\eta<0$ then $\lim_{t\to 0}X_t=+\infty$, while if $\eta>0$ then $\lim_{t\to 0}X_t=-\infty$ ($\boldsymbol{P}_0$-a.s.).*

(ii) *If $\eta=0$ then $X_t$ is recurrent ($\boldsymbol{P}_0$-a.s.); moreover,*

$$\overline{\lim_{t\to\infty}}\, X_t=+\infty,\quad \underline{\lim_{t\to\infty}}\, X_t=-\infty,\quad \boldsymbol{P}_0\text{-a.s.}$$

Let us sketch the proof. Consider the hitting times $T_x:=\min\{t\ge 0: X_t=x\}$ and denote by $f_{xy}$ the quenched first-passage probability from $x$ to $y$:

$$f_{xy}:=P_x^\omega\{1\le T_y<\infty\}$$

Starting from 0, the first step of the walk may be either to the right or to the left, hence by the Markov property the return probability $f_{00}$ can be decomposed as

$$f_{00}=p_0f_{10}+q_0f_{-1,0} \qquad [6]$$

To evaluate $f_{10}$, for $n\ge 1$ set

$$u_x\equiv u_x^{(n)}:=P_x^\omega\{T_0<T_n\},\quad 0\le x\le n$$

which is the probability to reach 0 prior to $n$, starting from $x$. Clearly,

$$f_{10}=\lim_{n\to\infty}u_1^{(n)} \qquad [7]$$

Decomposition with respect to the first step yields the difference equation

$$u_x=p_xu_{x+1}+q_xu_{x-1},\quad 0<x<n \qquad [8]$$

with the boundary conditions

$$u_0=1,\qquad u_n=0 \qquad [9]$$

Using $p_x + q_x = 1$, eqn [8] can be rewritten as

$$u_{x+1} - u_x = \rho_x(u_x - u_{x-1})$$

whence by iterations

$$u_{x+1} - u_x = (u_1 - u_0)\prod_{j=1}^{x} \rho_j \qquad [10]$$

Summing over $x$ and using the boundary conditions [9] we obtain

$$1 - u_1 = \left(\sum_{x=0}^{n-1}\prod_{j=1}^{x} \rho_j\right)^{-1} \qquad [11]$$

(if $x = 0$, the product over $j$ is interpreted as 1). In view of eqn [7] it follows that $f_{10} = 1$ if and only if the right-hand side of eqn [11] tends to 0, that is,

$$\sum_{x=1}^{\infty} \exp(Y_x) = \infty, \quad Y_x := \sum_{j=1}^{x} \ln \rho_j \qquad [12]$$

Note that the random variables $\ln \rho_j$ are i.i.d., hence by the strong LLN

$$\lim_{x\to\infty} \frac{Y_x}{x} = \mathbb{E}\ln \rho_0 \equiv \eta, \quad \mathbb{P}\text{-a.s.}$$

That is, the general term of the series [12] for large $x$ behaves like $\exp(x\eta)$; hence, for $\eta > 0$ the condition [12] holds true (and so $f_{10} = 1$), whereas for $\eta < 0$ it fails (and so $f_{10} < 1$).

By interchanging the roles of $p_x$ and $q_x$, we also have $f_{-1,0} < 1$ if $\eta > 0$ and $f_{-1,0} = 1$ if $\eta < 0$. From eqn [6], it then follows that in both cases $f_{00} < 1$, that is, the random walk is transient.

In the critical case, $\eta = 0$, by a general result from probability theory, $Y_x \geq 0$ for infinitely many $x$ ($\mathbb{P}$-a.s.), and so the series in eqn [12] diverges. Hence, $f_{10} = 1$ and, similarly, $f_{-1,0} = 1$, so by eqn [6] $f_{00} = 1$, that is, the random walk is recurrent.

It may be surprising that the critical parameter appears in the form $\eta = \mathbb{E}\ln \rho_0$, as it is probably more natural to expect, by analogy with the ordinary random walk, that the RWRE criterion would be based on the mean drift, $\mathbb{E}(p_0 - q_0)$. In the next section, we will see that the sign of $d$ may be misleading.

A canonical model of RWRE is specified by the assumption that the random variables $p_x$ take only two values, $\beta$ and $1 - \beta$, with probabilities

$$\mathbb{P}\{p_x = \beta\} = \alpha, \quad \mathbb{P}\{p_x = 1 - \beta\} = 1 - \alpha \qquad [13]$$

where $0 < \alpha < 1$, $0 < \beta < 1$. Here $\eta = (2\alpha - 1)\times \ln(1 + (1 - 2\beta)/\beta)$, and it is easy to see that, for example, $\eta < 0$ if $\alpha < 1/2$, $\beta < 1/2$ or $\alpha > 1/2$, $\beta > 1/2$. The recurrent region where $\eta = 0$ splits into two lines, $\beta = 1/2$ and $\alpha = 1/2$. Note that the first case is degenerate and amounts to the ordinary symmetric random walk, while the second one (except where $\beta = 1/2$) corresponds to Sinai's problem (see the section "Sinai's localization"). A "phase diagram" for this model, showing various limiting regimes as a function of the parameters $\alpha, \beta$, is presented in **Figure 1**.

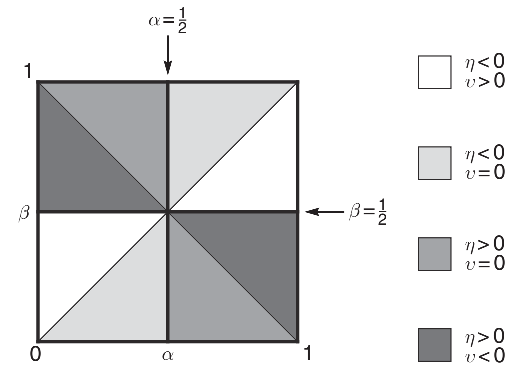



**Figure 1** Phase diagram for the canonical model, eqn [13]. In the regions where $\eta < 0$ or $\eta > 0$, the RWRE is transient to $+\infty$ or $-\infty$, respectively. The recurrent case, $\eta = 0$, arises when $\alpha = 1/2$ or $\beta = 1/2$. The asymptotic velocity $v := \lim_{t\to 0} X_t/t$ is given by eqn [14]. Adapted from Hughes BD (1996) *Random Walks and Random Environments. Volume 2: Random Environments*, Ch. 6, p. 391. Oxford: Clarendon, by permission of Oxford University Press.

## Asymptotic Velocity

In the transient case the walk escapes to infinity, and it is reasonable to ask at what speed. For a nonrandom environment, $p_x \equiv p$, the answer is given by the LLN, eqn [1]. For the simple RWRE, the asymptotic velocity was obtained by Solomon (1975). Note that by Jensen's inequality, $(\mathbb{E}\rho_0)^{-1} \leq \mathbb{E}\rho_0^{-1}$.

**Theorem 2** *The limit* $v := \lim_{t\to\infty} X_t/t$ *exists* ($\boldsymbol{P}_0$*-a.s.*) *and is given by*

$$v = \begin{cases} \dfrac{1 - \mathbb{E}\rho_0}{1 + \mathbb{E}\rho_0} & \text{if } \mathbb{E}\rho_0 < 1 \\ -\dfrac{1 - \mathbb{E}\rho_0^{-1}}{1 + \mathbb{E}\rho_0^{-1}} & \text{if } \mathbb{E}\rho_0^{-1} < 1 \\ 0 & \text{otherwise} \end{cases} \qquad [14]$$

Thus, the RWRE has a well-defined nonzero asymptotic velocity except when $(\mathbb{E}\rho_0)^{-1} \leq 1 \leq \mathbb{E}\rho_0^{-1}$. For instance, in the canonical example eqn [13] (see **Figure 1**), the criterion $\mathbb{E}\rho_0 < 1$ for the velocity $v$ to be positive amounts to the condition that both $(1 - \alpha)/\alpha$ and $(1 - \beta)/\beta$ lie on the same side of point 1.

The key idea of the proof is to analyze the hitting times $T_n$ first, deducing results for the walk $X_t$ later. More specifically, set $\tau_i = T_i - T_{i-1}$, which is the time to hit $i$ after hitting $i-1$ (providing that $i > X_0$). If $X_0 = 0$ and $n \geq 1$, then $T_n = \tau_1 + \cdots + \tau_n$. Note that in fixed environment $\omega$ the random variables $\{\tau_i\}$ are independent, since the quenched random walk "forgets" its past. Although there is no independence with respect to the annealed probability measure $\boldsymbol{P}_0$, one can show that, due to the i.i.d. property of the environment, the sequence $\{\tau_i\}$ is ergodic and therefore satisfies the LLN:

$$\frac{T_n}{n} = \frac{\tau_1 + \cdots + \tau_n}{n} \to \boldsymbol{E}_0\tau_1, \quad \boldsymbol{P}_0\text{-a.s.}$$

In turn, this implies

$$\frac{X_t}{t} \to \frac{1}{\boldsymbol{E}_0\tau_1}, \quad \boldsymbol{P}_0\text{-a.s.} \qquad [15]$$

(the clue is to note that $X_{T_n} = n$).

To compute the mean value $\boldsymbol{E}_0\tau_1$, observe that

$$\tau_1 = \mathbf{1}_{\{X_1=1\}} + \mathbf{1}_{\{X_1=-1\}}(1 + \tau_0' + \tau_1') \qquad [16]$$

where $\mathbf{1}_A$ is the indicator of event $A$ and $\tau_0', \tau_1'$ are, respectively, the times to get from $-1$ to 0 and then from 0 to 1. Taking expectations in a fixed environment $\omega$, we obtain

$$E_0^\omega\tau_1 = p_0 + q_0(1 + E_0^\omega\tau_0' + E_0^\omega\tau_1) \qquad [17]$$

and so

$$E_0^\omega\tau_1 = 1 + \rho_0 + \rho_0 E_0^\omega\tau_0' \qquad [18]$$

Note that $E_0^\omega\tau_0'$ is a function of $\{p_x, x<0\}$ and hence is independent of $\rho_0 = q_0/p_0$. Averaging eqn [18] over the environment and using $\boldsymbol{E}_0\tau_0' = \boldsymbol{E}_0\tau_1$ yields

$$\boldsymbol{E}_0\tau_1 = \begin{cases} \dfrac{1 + \mathbb{E}\rho_0}{1 - \mathbb{E}\rho_0} & \text{if } \mathbb{E}\rho_0 < 1 \\ \infty & \text{if } \mathbb{E}\rho_0 \geq 1 \end{cases} \qquad [19]$$

and by eqn [15] "half" of eqn [14] follows. The other half, in terms of $\mathbb{E}\rho_0^{-1}$, can be obtained by interchanging the roles of $p_x$ and $q_x$, whereby $\rho_0$ is replaced with $\rho_0^{-1}$.

Let us make a few remarks concerning Theorems 1 and 2. First of all, note that by Jensen's inequality $\mathbb{E}\ln\rho_0 \leq \ln\mathbb{E}\rho_0$, with a strict inequality whenever $\rho_0$ is nondegenerate. Therefore, it may be possible that, with $P_0$-probability 1, $X_t \to \infty$ but $X_t/t \to 0$ (see **Figure 1**). This is quite unusual as compared to the ordinary random walk (see the subsection "Ordinary random walks: a reminder"), and indicates some kind of slowdown in the transient case.

Furthermore, by Jensen's inequality

$$\mathbb{E}\rho_0 = \mathbb{E}p_0^{-1} - 1 \geq (\mathbb{E}p_0)^{-1} - 1$$

so eqn [14] implies that if $\mathbb{E}\rho_0 < 1$, then

$$0 < v \leq 2\,\mathbb{E}p_0 - 1 = \mathbb{E}\,(p_0 - q_0)$$

and the inequality is strict if $p_0$ is genuinely random (i.e., does not reduce to a constant). Hence, the asymptotic velocity $v$ is less than the mean drift $\mathbb{E}(p_0 - q_0)$, which is yet another evidence of slowdown. What is even more surprising is that it is possible to have $\mathbb{E}(p_0 - q_0) > 0$ but $\eta = \mathbb{E}\ln\rho_0 > 0$, so that $\mathbb{P}_0$-a.s. $X_t \to -\infty$ (although with velocity $v = 0$). Indeed, following Sznitman (2004) suppose that

$$\mathbb{P}\{p_0 = \beta\} = \alpha, \qquad \mathbb{P}\{p_0 = \gamma\} = 1 - \alpha$$

with $\alpha > 1/2$. Then $\mathbb{E}p_0 \geq \alpha\beta > 1/2$ if $1 > \beta > 1/2\alpha$, hence $\mathbb{E}(p_0 - q_0) = 2\,\mathbb{E}p_0 - 1 > 0$. On the other hand,

$$\mathbb{E}\ln\rho_0 = \alpha\ln\frac{1-\beta}{\beta} + (1-\alpha)\ln\frac{1-\gamma}{\gamma} > 0$$

if $\gamma$ is sufficiently small.

## Critical Exponent, Excursions, and Traps

Extending the previous analysis of the hitting times, one can obtain useful information about the limit distribution of $T_n$ (and hence $X_t$). To appreciate this, note that from the recursion eqn [16] it follows

$$\tau_1^s = \mathbf{1}_{\{X_1=1\}} + \mathbf{1}_{\{X_1=-1\}}(1 + \tau_0' + \tau_1')^s$$

and, similarly to [17],

$$E_0^\omega\tau_1^s = p_0 + q_0 E_0^\omega(1 + \tau_0' + \tau_1')^s$$

Taking here expectation $\mathbb{E}$, one can deduce that $\boldsymbol{E}_0\tau_1^s < \infty$ if and only if $\mathbb{E}\rho_0^s < 1$. Therefore, it is natural to expect that the root $\kappa$ of the equation

$$\mathbb{E}\rho_0^\kappa = 1 \qquad [20]$$

plays the role of a critical exponent responsible for the growth rate (and hence, for the type of the limit distribution) of the sum $T_n = \tau_1 + \cdots + \tau_n$. In particular, by analogy with sums of i.i.d. random variables one can expect that if $\kappa > 2$, then $T_n$ is asymptotically normal, with the standard scaling $\sqrt{n}$, while for $\kappa < 2$ the limit law of $T_n$ is stable (with index $\kappa$) under scaling $\approx n^{1/\kappa}$.

Alternatively, eqn [20] can be obtained from consideration of excursions of the random walk. Let $T_{11}^L$ be the left-excursion time from site 1, that is the time to return to 1 after moving to the left at the first step. If $\eta = \mathbb{E}\ln\rho_0 < 0$, then $T_{11}^L < \infty$ ($\boldsymbol{P}_0$-a.s.). Fixing an environment $\omega$, let $w_1 = E_1^\omega T_{11}^L$ be the

quenched mean duration of the excursion $T_{11}^L$ and observe that $w_1 = 1 + E_0^\omega \tau_1$, where $\tau_1$ is the time to get back to 1 after stepping to 0.

As a matter of fact, this representation and eqn [19] imply that the annealed mean duration of the left excursion, $\boldsymbol{E}_0 T_{11}^L$, is given by

$$\mathbb{E}w_1 = \begin{cases} \dfrac{2}{1 - \mathbb{E}\rho_0} & \text{if } \mathbb{E}\rho_0 < 1 \\ \infty & \text{if } \mathbb{E}\rho_0 \geq 1 \end{cases} \qquad [21]$$

Note that in the latter case (and bearing in mind $\eta < 0$), the random walk starting from 1 will eventually drift to $+\infty$, thus making only a finite number of visits to 0, but the expected number of such visits is infinite.

In fact, our goal here is to characterize the distribution of $w_1$ under the law $\mathbb{P}$. To this end, observe that the excursion $T_{11}^L$ involves at least two steps (the first and the last ones) and, possibly, several left excursions from 0, each with mean time $w_0 = E_0^\omega T_{00}^L$. Therefore,

$$w_1 = 2 + \sum_{j=1}^{\infty} q_0^j p_0 (j w_0) = 2 + \rho_0 w_0 \qquad [22]$$

By the translation invariance of the environment, the random variables $w_1$ and $w_0$ have the same distribution. Furthermore, similarly to recursion [22], we have $w_0 = 2 + \rho_{-1} w_{-1}$. This implies that $w_0$ is a function of $p_x$ with $x \leq -1$ only, and hence $w_0$ and $\rho_0$ are independent random variables. Introducing the Laplace transform $\phi(s) = \mathbb{E}\exp(-s w_1)$ and conditioning on $\rho_0$, from eqn [22] we get the equation

$$\phi(s) = \mathrm{e}^{-2s} \mathbb{E}\phi(s\rho_0) \qquad [23]$$

Suppose that

$$1 - \phi(s) \sim a s^\kappa, \quad s \to 0$$

then eqn [23] amounts to

$$1 - as^\kappa + \cdots = (1 - 2s + \cdots)(1 - as^\kappa\, \mathbb{E}\rho_0^\kappa + \cdots)$$

Expanding the product on the right, one can see that a solution with $\kappa = 1$ is possible only if $\mathbb{E}\rho_0 < 1$, in which case

$$a = \mathbb{E}w_1 = \frac{2}{1 - \mathbb{E}\rho_0}$$

We have already obtained this result in eqn [21].

The case $\kappa < 1$ is possible if $\mathbb{E}\rho_0^\kappa = 1$, which is exactly eqn [20]. Returning to $w_1$, one expects a slow decay of the distribution tail,

$$\mathbb{P}\{w_1 > t\} \sim bt^{-1/\kappa}, \quad t \to \infty$$

In particular, in this case the annealed mean duration of the left excursion appears to be infinite.

Although the above considerations point to the critical parameter $\kappa$, eqn [20], which may be expected to determine the slowdown scale, they provide little explanation of a mechanism of the slowdown phenomenon. Heuristically, it is natural to attribute the slowdown effects to the presence of "traps" in the environment, which may be thought of as regions that are easy to enter but hard to leave. In the one-dimensional case, such a trap would occur, for example, between two long series of successive sites where the probabilities $p_x$ are fairly large (on the left) and small (on the right).

Remarkably, traps can be characterized quantitatively with regard to the properties of the random environment, by linking them to certain large-deviation effects (see Sznitman (2002, 2004)). The key role in this analysis is played by the function $F(u) := \ln \mathbb{E}\rho_0^u$, $u \in \mathbb{R}$. Suppose that $\eta = \mathbb{E}\ln\rho_0 < 0$ (so that by Theorem 1 the RWRE tends to $+\infty$, $\boldsymbol{P}_0$-a.s.) and also that $\mathbb{E}\rho_0 > 1$ and $\mathbb{E}\rho_0^{-1} > 1$ (so that by Theorem 2, $v = 0$). The latter means that $F(1) > 0$ and $F(-1) > 0$, and since $F$ is a smooth strictly convex function and $F(0) = 0$, it follows that there is the second root $0 < \kappa < 1$, so that $F(\kappa) = 0$, that is, $\mathbb{E}\rho_0^\kappa = 1$ (cf. eqn [20]).

Let us estimate the probability to have a trap in $U = [-L, L]$ where the RWRE will spend anomalously long time. Using eqn [11], observe that

$$P_1^\omega\{T_0 < T_{L+1}\} \geq 1 - \exp\{-LS_L\}$$

where $S_L := L^{-1}\sum_{x=1}^{L} \ln\rho_x \to \eta < 0$ as $L \to \infty$. However, due to large deviations $S_L$ may exceed level $\epsilon > 0$ with probability

$$\mathbb{P}\{S_L > \epsilon\} \sim \exp\{-LI(\epsilon)\}, \quad L \to \infty$$

where $I(x) := \sup_u \{ux - F(u)\}$ is the Legendre transform of $F$. We can optimize this estimate by assuming that $\epsilon L \geq \ln n$ and minimizing the ratio $I(\epsilon)/\epsilon$. Note that $F(u)$ can be expressed via the inverse Legendre transform, $F(u) = \sup_x\{xu - I(x)\}$, and it is easy to see that if $\kappa := \min_{\epsilon > 0} I(\epsilon)/\epsilon$, then $F(\kappa) = 0$, so $\kappa$ is the second (positive) root of $F$.

The "left" probability $P_{-1}^\omega\{T_0 < T_{-L-1}\}$ is estimated in a similar fashion, and one can deduce that for some constants $K > 0, c > 0$, and any $\kappa' > \kappa$, for large $n$

$$\mathbb{P}\left\{P_0^\omega\left\{\max_{k \leq n} |X_k| \leq K\ln n\right\} \geq c\right\} \geq n^{-\kappa'}$$

That is to say, this is a bound on the probability to see a trap centered at 0, of size $\approx \ln n$, which will retain the RWRE for at least time $n$. It can be shown that, typically, there will be many such traps both in $[-n^{\kappa'}, 0]$ and $[0, n^{\kappa'}]$, which will essentially

prevent the RWRE from moving at distance $n^{\kappa'}$ from the origin before time $n$. In particular, it follows that $\lim_{n\to\infty} X_n/n^{\kappa'} = 0$ for any $\kappa' > \kappa$, so recalling that $0 < \kappa < 1$, we have indeed a sublinear growth of $X_n$. This result is more informative as compared to Theorem 2 (the case $v = 0$), and it clarifies the role of traps (see more details in Sznitman (2004)). The nontrivial behavior of the RWRE on the precise growth scale, $n^{\kappa}$, is characterized in the next section.

## Limit Distributions

Considerations in the previous section suggest that the exponent $\kappa$, defined as the solution of eqn [20], characterizes environments in terms of duration of left excursions. These heuristic arguments are confirmed by a limit theorem by Kesten *et al.* (1975), which specifies the slowdown scale. We state here the most striking part of their result. Denote $\ln^+ u := \max\{\ln u, 0\}$; by an arithmetic distribution one means a probability law on $\mathbb{R}$ concentrated on the set of points of the form $0, \pm c, \pm 2c, \ldots$.

**Theorem 3** *Assume that* $-\infty \le \eta = \mathbb{E} \ln \rho_0 < 0$ *and the distribution of* $\ln \rho_0$ *is nonarithmetic (excluding a possible atom at* $-\infty$*). Suppose that the root* $\kappa$ *of* eqn [20] *is such that* $0 < \kappa < 1$ *and* $\mathbb{E}\rho_0^{\kappa} \ln^+ \rho_0 < \infty$. *Then*

$$\lim_{n\to\infty} \boldsymbol{P}_0\{n^{-1/\kappa} T_n \le t\} = L_\kappa(t)$$

$$\lim_{t\to\infty} \boldsymbol{P}_0\{t^{-\kappa} X_t \le x\} = 1 - L_\kappa(x^{-1/\kappa})$$

*where* $L_\kappa(\cdot)$ *is the distribution function of a stable law with index* $\kappa$*, concentrated on* $[0, \infty)$.

General information on stable laws can be found in many probability books; we only mention here that the Laplace transform of a stable distribution on $[0, \infty)$ with index $\kappa$ has the form $\phi(s) = \exp\{-Cs^{\kappa}\}$.

Kesten *et al.* (1975) also consider the case $\kappa \ge 1$. Note that for $\kappa > 1$, we have $\mathbb{E}_{\rho_0} < (\mathbb{E}_{\rho_0^{\kappa}})^{1/\kappa} = 1$, so $v > 0$ by eqn [14]. For example, if $\kappa > 2$ then, as expected (see the previous section), there exists a nonrandom $\sigma^2 > 0$ such that

$$\lim_{n\to\infty} \boldsymbol{P}_0\left\{\frac{T_n - n/v}{\sigma\sqrt{n}} \le t\right\} = \Phi(t)$$

$$\lim_{t\to\infty} \boldsymbol{P}_0\left\{\frac{X_t - tv}{v^{3/2}\sigma\sqrt{t}} \le x\right\} = \Phi(x)$$

Let us describe an elegant idea of the proof based on a suitable renewal structure. (1) Let $U_i^n$ $(i \le n)$ be the number of left excursions starting from $i$ up to time $T_n$, and note that $T_n = n + 2\sum_i U_i^n$. Since the walk is transient to $+\infty$, the sum $\sum_{i\le 0} U_i^n$ is finite ($\boldsymbol{P}_0$-a.s.) and so does not affect the limit. (2) Observe that if the environment $\omega$ is fixed then the conditional distribution of $U_j^n$, given $U_{j+1}^n, \ldots, U_n^n = 0$, is the same as the distribution of the sum of $1 + U_{j+1}^n$ i.i.d. random variables $V_1, V_2, \ldots,$ each with geometric distribution $P_0^{\omega}\{V_i = k\} = p_j q_j^k$ $(k = 0, 1, 2, \ldots)$. Therefore, the sum $\sum_{i=1}^n U_i^n$ (read from right to left) can be represented as $\sum_{t=0}^{n-1} Z_t$, where $Z_0 = 0, Z_1, Z_2, \ldots$ is a branching process (in random environment $\{p_j\}$) with one immigrant at each step and the geometric offspring distribution with parameter $p_j$ for each particle present at time $j$. (3) Consider the successive "regeneration" times $\tau_k^*$, at which the process $Z_t$ vanishes. The partial sums $W_k := \sum_{\tau_k^* \le t < \tau_{k+1}^*} Z_t$ form an i.i.d. sequence, and the proof amounts to showing that the sum of $W_k$ has a stable limit of index $\kappa$. (4) Finally, the distribution of $W_0$ can be approximated using $M_0 := \sum_{t=1}^{\infty} \prod_{j=0}^{n-1} \rho_j$ (cf. eqn [11]), which is the quenched mean number of total progeny of the immigrant at time $t = 0$. Using Kesten's renewal theorem, it can be checked that $\mathbb{P}\{M_0 > x\} \sim Kx^{-\kappa}$ as $x \to \infty$, so $M_0$ is in the domain of attraction of a stable law with index $\kappa$, and the result follows.

Let us emphasize the significance of the regeneration times $\tau_i^*$. Returning to the original random walk, one can see that these are times at which the RWRE hits a new "record" on its way to $+\infty$, never to backtrack again. The same idea plays a crucial role in the analysis of the RWRE in higher dimensions (see the subsections "Zero–one laws and LLNs" and "Kalikow's condition and Sznitman's condition (T′)").

Finally, note that the condition $-\infty \le \eta < 0$ allows $\mathbb{P}\{p_0 = 1\} > 0$, so the distribution of $\rho_0$ may have an atom at 0 (and hence $\ln \rho_0$ at $-\infty$). In view of eqn [20], no atom is possible at $+\infty$. The restriction for the distribution of $\ln \rho_0$ to be nonarithmetic is important. This will be illustrated in the section "Diode model," where we discuss the model of random diodes.

## Sinai's Localization

The results discussed in the previous section indicate that the less transient the RWRE is (i.e., the critical exponent decreasing to zero), the slower it moves. Sinai (1982) proved a remarkable theorem showing that for the recurrent RWRE (i.e., with $\eta = \mathbb{E} \ln \rho_0 = 0$), the slowdown effect is exhibited in a striking way.

**Theorem 4** *Suppose that the environment $\{p_x\}$ is i.i.d. and elliptic, eqn [5], and assume that $\mathbb{E}\ln\rho_0 = 0$, with $\mathbb{P}\{\rho_0 = 1\} < 1$. Denote $\sigma^2 := \mathbb{E}\ln^2 \rho_0, 0 < \sigma^2 < \infty$. Then there exists a function $W_n = W_n(\omega)$ of the random environment such that for any $\varepsilon > 0$*

$$\lim_{n\to\infty} \boldsymbol{P}_0\left\{\left|\frac{\sigma^2 X_n}{\ln^2 n} - W_n\right| > \varepsilon\right\} = 0 \qquad [24]$$

*Moreover, $W_n$ has a limit distribution:*

$$\lim_{n\to\infty} \mathbb{P}\{W_n \le x\} = G(x) \qquad [25]$$

*and thus also the distribution of $\sigma^2 X_n / \ln^2 n$ under $\boldsymbol{P}_0$ converges to the same distribution $G(x)$.*

Sinai's theorem shows that in the recurrent case, the RWRE considered on the spatial scale $\ln^2 n$ becomes localized near some random point (depending on the environment only). This phenomenon, frequently referred to as "Sinai's localization," indicates an extremely strong slowdown of the motion as compared with the ordinary diffusive behavior.

Following Révész (1990), let us explain heuristically why $X_n$ is measured on the scale $\ln^2 n$. Rewrite eqn [11] as

$$P_1^\omega\{T_n < T_0\} = \left(1 + \sum_{x=1}^{n-1} \exp(Y_x)\right)^{-1} \qquad [26]$$

where $Y_x$ is defined in eqn [12]. By the CLT, the typical size of $|Y_x|$ for large $x$ is of order of $\sqrt{x}$, and so eqn [26] yields

$$P_1^\omega\{T_n < T_0\} \approx \exp\{-\sqrt{n}\}$$

This suggests that the walk started at site 1 will make about $\exp\{\sqrt{n}\}$ visits to the origin before reaching level $n$. Therefore, the first passage to site $n$ takes at least time $\approx \exp\{\sqrt{n}\}$. In other words, one may expect that a typical displacement after $n$ steps will be of order of $\ln^2 n$ (cf. eqn [24]). This argument also indicates, in the spirit of the trapping mechanism of slowdown discussed at the end of the section "Critical exponent, excursions, and traps," that there is typically a trap of size $\approx \ln^2 n$, which retains the RWRE until time $n$.

It has been shown (independently by H Kesten and A O Golosov) that the limit in [25] coincides with the distribution of a certain functional of the standard Brownian motion, with the density function

$$G'(x) = \frac{2}{\pi}\sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\left\{-\frac{(2k+1)^2\pi^2}{8}|x|\right\}$$

## Environment Viewed from the Particle

This important technique, dating back to Kozlov and Molchanov (1984), has proved to be quite efficient in the study of random motions in random media. The basic idea is to focus on the evolution of the environment viewed from the current position of the walk.

Let $\theta$ be the shift operator acting on the space of environments $\Omega = \{\omega\}$ as follows:

$$\omega = \{p_x\} \overset{\theta}{\mapsto} \bar{\omega} = \{p_{x-1}\}$$

Consider the process

$$\omega_n := \theta^{X_n}\omega, \qquad \omega_0 = \omega$$

which describes the state of the environment from the point of view of an observer moving along with the random walk $X_n$. One can show that $\omega_n$ is a Markov chain (with respect to both $P_0^\omega$ and $\boldsymbol{P}_0$), with the transition kernel

$$T(\omega, \mathrm{d}\omega') = p_0\, \delta_{\theta\omega}(\mathrm{d}\omega') + q_0\, \delta_{\theta^{-1}\omega}(\mathrm{d}\omega') \qquad [27]$$

and the respective initial law $\delta_\omega$ or $\mathbb{P}$ (here $\delta_\omega$ is the Dirac measure, i.e., unit mass at $\omega$).

This fact as it stands may not seem to be of any practical use, since the state space of this Markov chain is very complex. However, the great advantage is that one can find an explicit invariant probability $\mathbb{Q}$ for the kernel $T$ (i.e., such that $\mathbb{Q}T = \mathbb{Q}$), which is absolutely continuous with respect to $\mathbb{P}$.

More specifically, assume that $\mathbb{E}\rho_0 < 1$ and set $\mathbb{Q} = f(\omega)\mathbb{P}$, where (cf. eqn [14])

$$f = v\,(1+\rho_0)\sum_{x=0}^{\infty}\prod_{j=1}^{x}\rho_j \qquad v = \frac{1-\mathbb{E}\rho_0}{1+\mathbb{E}\rho_0} \qquad [28]$$

Using independence of $\{\rho_x\}$, we note

$$\int_\Omega \mathbb{Q}(\mathrm{d}\omega) = \mathbb{E}f = (1-\mathbb{E}\rho_0)\sum_{x=0}^{\infty}(\mathbb{E}\rho_0)^x = 1$$

hence $\mathbb{Q}$ is a probability measure on $\Omega$. Furthermore, for any bounded measurable function $g$ on $\Omega$ we have

$$\begin{aligned}\mathbb{Q}Tg &= \int_\Omega Tg(\omega)\mathbb{Q}(\mathrm{d}\omega) = \mathbb{E}f\,Tg \\ &= \mathbb{E}\{f[p_0\,(g\circ\theta) + q_0\,(g\circ\theta^{-1})]\} \\ &= \mathbb{E}\{g[(p_0 f)\circ\theta^{-1} + (q_0 f)\circ\theta]\}\end{aligned} \qquad [29]$$

By eqn [28],

$$(p_0 f)\circ\theta^{-1} = vp_{-1}(1+\rho_{-1})\sum_{x=0}^{\infty}\prod_{j=1}^{x}\rho_{j-1}$$

$$= v\left(1+\rho_0\sum_{x=0}^{\infty}\prod_{j=1}^{x}\rho_j\right) = v + \frac{\rho_0}{1+\rho_0} f$$

and similarly

$$(q_0 f)\circ\theta = -v + \frac{1}{1+\rho_0} f$$

So from eqn [29] we obtain

$$\mathbb{Q}Tg = \mathbb{E}(gf) = \int_\Omega g(\omega)\ \mathbb{Q}(\mathrm{d}\omega) = \mathbb{Q}g$$

which proves the invariance of $\mathbb{Q}$.

To illustrate the environment method, let us sketch the proof of Solomon's result on the asymptotic velocity (see Theorem 2). Set $d(x,\omega) := E_x^\omega(X_1 - X_0) = p_x - q_x$. Noting that $d(x,\omega) = d(0,\theta^x\omega)$, define

$$D_n := \sum_{i=1}^{n} d(X_{i-1},\omega) = \sum_{i=1}^{n} d(0,\theta^{X_{i-1}}\omega)$$

Due to the Markov property, the process $M_n := X_n - D_n$ is a martingale with respect to the natural filtration $\mathcal{F}_n = \sigma\{X_1,\ldots,X_n\}$ and the law $P_0^\omega$,

$$E_0^\omega[M_{n+1}\,|\,\mathcal{F}_n] = M_n, \quad P_0^\omega\text{-a.s.}$$

and it has bounded jumps, $|M_n - M_{n-1}| \le 2$. By general results, this implies $M_n/n \to 0$ ($P_0^\omega$-a.s.).

On the other hand, by Birkhoff's ergodic theorem

$$\lim_{n\to\infty}\frac{D_n}{n} = \int_\Omega d(0,\omega)\ \mathbb{Q}(\mathrm{d}\omega), \quad \boldsymbol{P}_0\text{-a.s.}$$

The last integral is easily evaluated to yield

$$\mathbb{E}(p_0 - q_0)f = v\mathbb{E}\sum_{x=0}^{\infty}\prod_{j=1}^{x}\rho_j(1-\rho_0)$$

$$= v(1-\mathbb{E}\rho_0)\sum_{x=0}^{\infty}(\mathbb{E}\rho_0)^x = v$$

and the first part of the formula [14] follows.

The case $\mathbb{E}\rho_0 \ge 1$ can be handled using a comparison argument (Sznitman 2004). Observe that if $p_x \le \tilde{p}_x$ for all $x$ then for the corresponding random walks we have $X_t \le \tilde{X}_t$ ($P_0^\omega$-a.s.). We now define a suitable dominating random medium by setting (for $\gamma > 0$)

$$\tilde{p}_x := \frac{p_x}{1+\gamma} + \frac{\gamma}{1+\gamma} \ge p_x$$

Then $\mathbb{E}\,\tilde{\rho}_0 = \mathbb{E}q_0/(p_0+\gamma) < 1$ if $\gamma$ is large enough, so by the first part of the theorem, $P_0^\omega$-a.s.,

$$\overline{\lim_{n\to\infty}}\frac{X_n}{n} \le \lim_{n\to\infty}\frac{\tilde{X}_n}{n} = \frac{1-\mathbb{E}\tilde{\rho}_0}{1+\mathbb{E}\tilde{\rho}_0} \qquad [30]$$

Note that $\mathbb{E}\tilde{\rho}_0$ is a continuous function of $\gamma$ with values in $[0,\mathbb{E}\rho_0] \ni 1$, so there exists $\gamma^*$ such that $\mathbb{E}\tilde{\rho}_0$ attains the value 1. Passing to the limit in eqn [30] as $\gamma \uparrow \gamma^*$, we obtain $\overline{\lim}_{n\to\infty}X_n/n \le 0$ ($P_0^\omega$-a.s.). Similarly, we get the reverse inequality, which proves the second part of the theorem.

A more prominent advantage of the environment method is that it naturally leads to statements of CLT type. A key step is to find a function $H(x,t,\omega) = x - vt + h(x,\omega)$ (called "harmonic coordinate") such that the process $H(X_n,n,\omega)$ is a martingale. To this end, by the Markov property it suffices to have

$$E_{X_n}^\omega H(X_{n+1},n+1,\omega) = H(X_n,n,\omega), \quad P_0^\omega\text{-a.s.}$$

For $\Delta(x,\omega) := h(x+1,\omega) - h(x,\omega)$ this condition leads to the equation

$$\Delta(x,\omega) = \rho_x\Delta(x-1,\omega) + v - 1 + (1+v)\rho_x$$

If $\mathbb{E}\rho_0 < 1$ (so that $v > 0$), there exists a bounded solution

$$\Delta(x,\omega) = v - 1 + 2v\sum_{k=0}^{\infty}\prod_{i=0}^{k}\rho_{x-i}$$

and we note that $\Delta(x,\omega) = \Delta(0,\theta^x\omega)$ is a stationary sequence with mean $\mathbb{E}\Delta(x,\omega) = 0$. Finally, setting $h(0,\omega) = 0$ we find

$$h(x,\omega) = \begin{cases} \sum_{k=0}^{x-1}\Delta(k,\omega), & x > 0 \\ -\sum_{k=1}^{-x}\Delta(-k,\omega), & x < 0 \end{cases}$$

As a result, we have the representation

$$X_n - nv = H(X_n,n,\omega) + h(X_n,\omega) \qquad [31]$$

For a fixed $\omega$, one can apply a suitable CLT for martingale differences to the martingale term in eqn [31], while using that $X_n \sim nv$ ($\boldsymbol{P}_0$-a.s.), the second term in eqn [31] is approximated by the sum $\sum_{k=0}^{nv}\Delta(k,\omega)$, which can be handled via a CLT for stationary sequences. This way, we arrive at the following result.

**Theorem 5** *Suppose that the environment is elliptic, eqn [5], and such that $\mathbb{E}\rho_0^{2+\varepsilon} < 1$ for some $\varepsilon > 0$ (which implies that $\mathbb{E}\rho_0 < 1$ and hence $v > 0$). Then there exists a nonrandom $\sigma^2 > 0$ such that*

$$\lim_{n\to\infty}\boldsymbol{P}_0\left\{\frac{X_n - nv}{\sqrt{n\sigma^2}} \le x\right\} = \Phi(x)$$

Note that this theorem is parallel to the result by Kesten *et al.* (1975) on asymptotic normality when $\kappa > 2$ (see the section "Limit distributions"). The moment assumptions in Theorem 5 are more restrictive, but they can be relaxed. On the other hand, Theorem 5 does not impose the nonarithmetic condition on the distribution of the environment (cf. Theorem 3). More importantly, the environment method proves to be quite efficient in more general situations, including non-i.i.d. environments and higher dimensions (at least in some cases, e.g., for random bonds RWRE and balanced RWRE discussed subsequently).

## Diode Model

In the preceding sections (except in the section "Limit distributions," where however we were limited to a nonarithmetic case), we assumed that $0 < p_x < 1$ and therefore excluded the situation where there are sites through which motion is permitted in one direction only. Allowing for such a possibility leads to the "diode model" (Solomon 1975). Specifically, suppose that

$$\mathbb{P}\{p_x = \beta\} = \alpha, \qquad \mathbb{P}\{p_x = 1\} = 1 - \alpha \qquad [32]$$

with $0 < \alpha < 1$, $0 < \beta < 1$, so that with probability $\alpha$ a point $x \in \mathbb{Z}$ is a usual two-way site and with probability $1 - \alpha$ it is a repelling barrier ("diode"), through which passage is only possible from left to right. This is an interesting example of statistically inhomogeneous medium, where the particle motion is strongly irreversible due to the presence of special semipenetrable nodes. The principal mathematical advantage of such a model is that the random walk can be decomposed into independent excursions from one diode to the next.

Due to diodes, the RWRE will eventually drift to $+\infty$. If $\beta > 1/2$, then on average it moves faster than in a nonrandom environment with $p_x \equiv \beta$. The situation where $\beta \le 1/2$ is potentially more interesting, as then there is a competition between the local drift of the walk to the left (in ordinary sites) and the presence of repelling diodes on its way. Note that $\mathbb{E}\rho_0 = \alpha\rho$, where $\rho := (1-\beta)/\beta$, so the condition $\mathbb{E}\rho_0 < 1$ amounts to $\beta > \alpha/(1+\alpha)$. In this case (which includes $\beta > 1/2$), formula [14] for the asymptotic velocity applies.

As explained in the section "Critical exponent, excursions, and traps," the quenched mean duration $w$ of the left excursion has Laplace transform given by eqn [23], which now reads

$$\phi(s) = \mathrm{e}^{-2s}\{1 - \alpha + \alpha\,\phi(s\rho)\}$$

This equation is easily solved by iterations:

$$\phi(s) = (1-\alpha)\sum_{k=0}^{\infty} \alpha^k \mathrm{e}^{-st_k}$$
$$t_k := 2\sum_{j=0}^{k} \rho^j \qquad [33]$$

hence the distribution of $w$ is given by

$$\mathbb{P}\{w = t_k\} = (1-\alpha)\alpha^k, \quad k = 0, 1, \ldots$$

This result has a transparent probabilistic meaning. In fact, the factor $(1-\alpha)\alpha^k$ is the probability that the nearest diode on the left of the starting point occurs at distance $k+1$, whereas $t_k$ is the corresponding mean excursion time. Note that formula [33] for $t_k$ easily follows from the recursion $t_k = 2 + \rho t_{k-1}$ (cf. eqn [22]) with the boundary condition $t_0 = 2$.

A self-similar hierarchy of timescales [33] indicates that the process will exhibit temporal oscillations. Indeed, for $\alpha\rho > 1$ the average waiting time until passing through a valley of ordinary sites of length $k$ is asymptotically proportional to $t_k \sim 2\rho^k$, so one may expect the annealed mean displacement $\boldsymbol{E}_0 X_n$ to have a local minimum at $n \approx t_k$. Passing to logarithms, we note that $\ln t_{k+1} - \ln t_k \sim \ln\rho$, which suggests the occurrence of persistent oscillations on the logarithmic timescale, with period $\ln\rho$ (see **Figure 2**). This was confirmed by Bernasconi and Schneider (1985) who showed that for $\alpha\rho > 1$

$$\boldsymbol{E}_0 X_n \sim n^{\kappa} F(\ln n), \quad n \to \infty \qquad [34]$$

where $\kappa = -\ln\alpha/\ln\rho < 1$ is the solution of eqn [20] and the function $F$ is periodic with period $\ln\rho$ (see **Figure 2**).

In contrast, for $\alpha\rho = 1$ one has

$$\boldsymbol{E}_0 X_n \sim \frac{n\ln\rho}{2\ln n}, \quad n \to \infty$$

and there are no oscillations of the above kind.

These results illuminate the earlier analysis of the diode model by Solomon (1975), which in the main has revealed the following. If $\alpha\rho = 1$, then $X_n$ satisfies the strong LLN:

$$\lim_{n\to\infty} \frac{X_n}{n/\ln n} = \frac{\ln\rho}{2}, \quad \boldsymbol{P}_0\text{-a.s.}$$

while in the case $\alpha\rho > 1$ the asymptotic behavior of $X_n$ is quite complicated and unusual: if $n_i \to \infty$ is a sequence of integers such that $\{\ln n_i\} \to \gamma$ (here $\{a\} = a - [a]$ denotes the fractional part of $a$), then the distribution of $n_i^{-\kappa} X_{n_i}$ under $\boldsymbol{P}_0$ converges to a nondegenerate distribution which depends on $\gamma$. Thus, the very existence of the limiting distribution

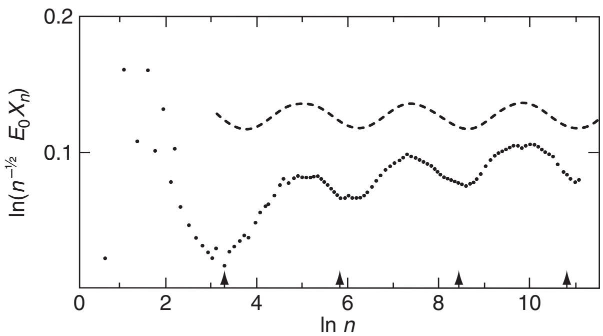



**Figure 2** Temporal oscillations for the diode model, eqn [32]. Here $\alpha=0.3$ and $\rho=1/0.09$, so that $\alpha\rho>1$ and $\kappa=1/2$. The dots represent an average of Monte Carlo simulations over 10 000 samples of the environment with a random walk of 200 000 steps in each realization. The broken curve refers to the exact asymptotic solution [34]. The arrows indicate the simulated locations of the minima $t_k$, the asymptotic spacing of which is predicted to be $\ln\rho\approx 241$. Reproduced from Bernasconi J and Schneider WR (1982). Diffusion on a one-dimensional lattice with random asymmetric transition rates. *Journal of Physics A: Mathematical and General* 15: L729–L734, by permission of IOP Publishing Ltd.

of $X_n$ and the limit itself heavily depend on the subsequence $n_i$ chosen to approach infinity.

This should be compared with a more "regular" result Theorem 3. Note that almost all the conditions of this theorem are satisfied in the diode model, except that here the distribution of $\ln\rho_0$ is arithmetic (recall that the value $\ln\rho_0=-\infty$ is permissible), so it is the discreteness of the environment distribution that does not provide enough "mixing" and hence leads to such peculiar features of the asymptotics.

## Some Generalizations and Variations

Most of the results discussed above in the simplest context of RWRE with nearest-neighbor jumps in an i.i.d. random environment have been extended to some other cases. One natural generalization is to relax the i.i.d. assumption, for example, by considering stationary ergodic environments (see details in Zeitouni (2004)). In this context, one relies on an ergodic theorem instead of the usual strong LLN. For instance, this way one readily obtains an extension of Solomon's criterion of transience versus recurrence (see Theorem 1). Other examples include an LLN (along with a formula for the asymptotic velocity, cf. Theorem 2), a CLT and stable laws for the asymptotic distribution of $X_n$ (cf. Theorem 3), and Sinai's localization result for the recurrent RWRE (cf. Theorem 4). Usually, however, ergodic theorems cannot be applied directly (like, e.g., to $X_n$, as the sequence $X_n-X_{n-1}$ is not stationary). In this case, one rather uses the hitting times which possess the desired stationarity (cf. the sections "Asymptotic velocity" and "Critical exponent, excursions, and traps"). In some situations, in addition to stationarity, one needs suitable mixing conditions in order to ensure enough decoupling (e.g., in Sinai's problem). The method of environment viewed from the particle (discussed earlier) is also suited very well to dealing with stationarity.

In the remainder of this section, we describe some other generalizations including RWRE with bounded jumps, RWRE where randomness is attached to bonds rather than sites, and continuous-time (symmetric) RWRE driven by the randomized master equation.

### RWRE with Bounded Jumps

The previous discussion was restricted to the case of RWRE with nearest-neighbor jumps. A natural extension is RWRE with bounded jumps. Let $L$, $R$ be fixed natural numbers, and suppose that from each site $x\in\mathbb{Z}$ jumps are only possible to the sites $x+i$, $i=-L,\ldots,R$, with (random) probabilities

$$p_x(i)\geq 0,\quad \sum_{i=-L}^{R} p_x(i)=1 \qquad [35]$$

We assume that the random vectors $p_x(\cdot)$ determining the environment are i.i.d. for different $x\in\mathbb{Z}$ (although many results can be extended to the stationary ergodic case).

The study of asymptotic properties of such a model is essentially more complex, as it involves products of certain random matrices and hence must use extensively the theory of Lyapunov exponents (see details and further references in Brémont (2004)). Lyapunov exponents, being natural analogs of logarithms of eigenvalues, characterize the asymptotic action of the product of random matrices along (random) principal directions, as described by Oseledec's multiplicative ergodic theorem. In most situations, however, the Lyapunov spectrum can

only be accessed implicitly, which makes the analysis rather hard.

To explain how random matrices arise here, let us first consider a particular case $R=1, L\geq 1$. Assume that $p_x(-L), p_x(1)\geq \delta>0$ for all $x\in\mathbb{Z}$ (ellipticity condition, cf. eqn [5]), and consider the hitting probabilities $u_n := P_n^{\omega}\{T_0<\infty\}$, where $T_0:=\min\{t\geq 0: X_t\leq 0\}$ (cf. the section "Transience and recurrence"). By decomposing with respect to the first step, for $n\geq 1$ we obtain the difference equation

$$u_n = p_n(1)u_{n+1} + \sum_{i=0}^{L} p_n(-i)u_{n-i} \qquad [36]$$

with the boundary conditions $u_0=\cdots=u_{-L+1}=1$. Using that $1=p_n(1)+\sum_{i=0}^{L} p_n(-i)$, we can rewrite eqn [36] as

$$p_n(1)(u_n-u_{n+1}) = \sum_{i=1}^{L} p_n(-i)(u_{n-i}-u_n)$$

or, equivalently,

$$v_n = \sum_{i=1}^{L} b_n(i)v_{n-i} \qquad [37]$$

where $v_i := u_i - u_{i+1}$ and

$$b_n(i) := \frac{p_n(-i)+\cdots+p_n(-L)}{p_n(1)} \qquad [38]$$

Recursion [37] can be written in a matrix form, $V_n = M_n V_{n-1}$, where $V_n := (v_n,\ldots,v_{n-L+1})^{\top}$,

$$M_n := \begin{pmatrix} b_n(1) & \cdots & \cdots & b_n(L) \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix} \qquad [39]$$

and by iterations we get (cf. eqn [10])

$$V_n = M_n\cdots M_1 V_0, \qquad V_0 = (1-u_1, 0,\ldots,0)^{\top}$$

Note that $M_n$ depends only on the transition probability vector $p_n(\cdot)$, and hence $M_n\cdots M_1$ is the product of i.i.d. random (non-negative) matrices. By Furstenberg–Kesten's theorem, the limiting behavior of such a product, as $n\to\infty$, is controlled by the largest Lyapunov exponent

$$\gamma_1 := \lim_{n\to\infty} n^{-1}\ln\|M_n\ldots M_1\| \qquad [40]$$

(by Kingman's subadditive ergodic theorem, the limit exists $\mathbb{P}$-a.s. and is nonrandom). It follows that, $P_0$-a.s., the RWRE $X_n$ is transient if and only if $\gamma_1\neq 0$, and moreover, $\lim_{n\to\infty}X_n=+\infty\,(-\infty)$ when $\gamma_1<0\,(>0)$, whereas $\underline{\lim}_{n\to\infty}X_n=-\infty$, $\overline{\lim}_{n\to\infty}X_n=+\infty$ when $\gamma_1=0$.

For orientation, note that if $p_n(i)=p(i)$ are nonrandom constants, then $\gamma_1=\ln\lambda_1$, where $\lambda_1>0$ is the largest eigenvalue of $M_0$, and so $\gamma_1<0$ if and only if $\lambda_1<1$. The latter means that the characteristic polynomial $\varphi(\lambda):=\det(M_0-\lambda I)$ satisfies the condition $(-1)^L\varphi(1)>0$. To evaluate $\det(M_0-I)$, replace the first column by the sum of all columns and expand to get $\varphi(1)=(-1)^{L-1}(b_1+\cdots+b_L)$. Substituting expressions [38] it is easy to see that the above condition amounts to $p(1)-\sum_{i=1}^{L} ip\times(-i)>0$, that is, the mean drift of the random walk is positive and hence $X_n\to+\infty$ a.s.

In the general case, $L\geq 1, R\geq 1$, similar considerations lead to the following matrices of order $d:=L+R-1$ (cf. eqn [39]):

$$M_n = \begin{pmatrix} a_n(R-1) & \cdots & a_n(1) & b_n(1) & \cdots & b_n(L) \\ 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 & 0 \end{pmatrix}$$

where $b_n(i)$ are given by eqn [38] and

$$a_n(i) := -\frac{p_n(i)+\cdots+p_n(R)}{p_n(R)}$$

Suppose that the ellipticity condition is satisfied in the form $p_n(i)\geq\delta>0$, $i\neq 0$, $-L\leq i\leq R$, and let $\gamma_1\geq\gamma_2\geq\cdots\geq\gamma_d$ be the (nonrandom) Lyapunov exponents of $\{M_n\}$. The largest exponent $\gamma_1$ is again given by eqn [40], while other exponents are determined recursively from the equalities

$$\gamma_1+\cdots+\gamma_k = \lim_{n\to\infty} n^{-1}\ln\|\wedge^k(M_n\cdots M_1)\|$$

$(1\leq k\leq d)$. Here $\wedge$ denotes the external (antisymmetric) product: $x\wedge y=-y\wedge x$ $(x,y\in\mathbb{R}^d)$, and $\wedge^k M$ acts on the external product space $\wedge^k\mathbb{R}^d$, generated by the canonical basis $\{e_{i_1}\wedge\cdots\wedge e_{i_k}, 1\leq i_1<\cdots<i_k\leq d\}$, as follows:

$$\wedge^k M(x_1\wedge\cdots\wedge x_k) := M(x_1)\wedge\cdots\wedge M(x_k)$$

One can show that all exponents except $\gamma_R$ are sign-definite: $\gamma_{R-1}>0>\gamma_{R+1}$. Moreover, it is the sign of $\gamma_R$ that determines whether the RWRE is transient or recurrent, the dichotomy being the same as in the case $R=1$ above (with $\gamma_1$ replaced by $\gamma_R$). Let us also mention that an LLN and CLT can be proved here (see Brémont (2004)).

In conclusion, let us point out an alternative approach due to Bolthausen and Goldsheid (2000)

who studied a more general RWRE on a strip $\mathbb{Z}\times\{0,1,\dots,m-1\}$. The link between these two models is given by the representation $X_n = mY_n + Z_n$, where $m := \max\{L,R\}$, $Y_n\in\mathbb{Z}$, $Z_n\in\{0,\dots,m-1\}$. Random matrices arising here are constructed indirectly using an auxiliary stationary sequence. Even though these matrices are nonindependent, thanks to their positivity the criterion of transience can be given in terms of the sign of the largest Lyapunov exponent, which is usually much easier to deal with. An additional attractive feature of this approach is that the condition $p_x(R)>0$ ($\mathbb{P}$-a.s.), which was essential for the previous technique, can be replaced with a more natural condition $\mathbb{P}\{p_x(R)>0\}>0$.

### Random Bonds RWRE

Instead of having random probabilities of jumps at each site, one could assign random weights to bonds between the sites. For instance, the transition probabilities $p_x = p(x,x+1,\omega)$ can be defined by

$$p_x = \frac{c_{x,x+1}}{c_{x-1,x}+c_{x,x+1}} \qquad [41]$$

where $c_{x,x+1}>0$ are i.i.d. random variables on the environment space $\Omega$.

The difference between the two models may not seem very prominent, but the behavior of the walk in the modified model [41] appears to be quite different. Indeed, working as in the section "Transience and recurrence," we note that

$$\rho_x = \frac{q_x}{p_x} = \frac{c_{x-1,x}}{c_{x,x+1}}$$

hence, exploiting formulas [11] and [41], we obtain, $\mathbb{P}$-a.s.,

$$\frac{1}{1-u_1} = \sum_{x=0}^{n-1}\frac{c_{01}}{c_{x,x+1}} \sim c_{01}\, n\,\mathbb{E}c_{01}^{-1} \to \infty \qquad [42]$$

since $\mathbb{E}c_{01}^{-1}>0$. Therefore, $f_{00}=1$, that is, the random walk is recurrent ($\boldsymbol{P}_0$-a.s.).

The method of environment viewed from the particle can also be applied here (see Sznitman (2004)). Similarly to the section "Environment viewed from the particle," we define a new probability measure $\mathbb{Q}=f(\omega)\,\mathbb{P}$ using the density

$$f(\omega) = Z^{-1}\big(c_{-1,0}(\omega)+c_{01}(\omega)\big)$$

where $Z=2\mathbb{E}c_{01}$ is the normalizing constant (we assume that $\mathbb{E}c_{01}<\infty$). One can check that $\mathbb{Q}$ is invariant with respect to the transition kernel eqn [41], and by similar arguments as in that section, we obtain that $\lim_{n\to\infty}X_n/n$ exists ($P_0^\omega$-a.s.) and is given by

$$\int_\Omega d(0,\omega)\;\mathbb{Q}(\mathrm{d}\omega) = Z^{-1}\mathbb{E}\big[c_{01}-c_{-1,0}\big] = 0$$

so the asymptotic velocity vanishes.

Furthermore, under suitable technical conditions on the environment (e.g., $c_{01}$ being bounded away from 0 and $\infty$, cf. eqn [5]), one can prove the following CLT:

$$\lim_{n\to\infty}\boldsymbol{P}_0\left\{\frac{X_n}{\sqrt{n\sigma^2}}\le x\right\} = \Phi(x) \qquad [43]$$

where $\sigma^2=(\mathbb{E}c_{01}\cdot\mathbb{E}c_{01}^{-1})^{-1}$. Note that $\sigma^2\le 1$ (with a strict inequality if $c_{01}$ is not reduced to a constant), which indicates some slowdown in the spatial spread of the random bonds RWRE, as compared to the ordinary symmetric random walk.

Thus, there is a dramatic distinction between the random bonds RWRE, which is recurrent and diffusive, and the random sites RWRE, with a much more complex asymptotics including both transient and recurrent scenarios, slowdown effects, and subdiffusive behavior. This can be explained heuristically by noting that the random bonds RWRE is reversible, that is, $m(x)p(x,y)=m(y)\times p(y,x)$ for all $x,y\in\mathbb{Z}$, with $m(x):=c_{x-1,x}+c_{x,x+1}$ (this property also easily extends to multidimensional versions). Hence, it appears impossible to create extended traps which would retain the particle for a very long time. Instead, the mechanism of the diffusive slowdown in a reversible case is associated with the natural variability of the environment resulting in the occasional occurrence of isolated "screening" bonds with an anomalously small weight $c_{x,x+1}$.

Let us point out that the RWRE determined by eqn [41] can be interpreted in terms of the random conductivity model (see Hughes (1996)). Suppose that each random variable $c_{x,x+1}$ attached to the bond $(x,x+1)$ has the meaning of the conductance of this bond (the reciprocal, $c_{x,x+1}^{-1}$, being its resistance). If a voltage drop $V$ is applied across the system of $N$ successive bonds, say from 0 to $N$, then the same current $I$ flows in each of the conductors and by Ohm's law we have $I=c_{x,x+1}\;V_{x,x+1}$, where $V_{x,x+1}$ is the voltage drop across the corresponding bond. Hence

$$V=\sum_{x=0}^{N}V_{x,x+1} = I\sum_{x=0}^{N}c_{x,x+1}^{-1}$$

which amounts to saying that the total resistance of the system of consecutive elements is given by the sum of the individual resistances. The effective

conductivity of the finite system, $\bar{c}_N$, is defined as the average conductance per bond, so that

$$\bar{c}_N^{-1} = \frac{1}{N}\sum_{x=0}^{N} c_{x,x+1}^{-1}$$

and by the strong LLN, $\bar{c}_N^{-1} \to \mathbb{E}c_{01}^{-1}$ as $N \to \infty$ ($\mathbb{P}$-a.s.). Therefore, the effective conductivity of the infinite system is given by $\bar{c} = (\mathbb{E}c_{01}^{-1})^{-1}$, and we note that $\bar{c} < \mathbb{E}c_{01}$ if the random medium is nondegenerate.

Returning to the random bonds RWRE, eqn [41], it is easy to see that a site $j$ is recurrent if and only if the conductance $c_{j,\infty}$ between $x$ and $\infty$ equals zero. Using again Ohm's law, we have (cf. eqn [42])

$$c_{j,+\infty}^{-1} = \sum_{x=j}^{\infty} c_{x,x+1}^{-1} = 0, \quad \mathbb{P}\text{-a.s.}$$

and we recover the result about recurrence.

### Continuous-Time RWRE

As in the discrete-time case, a random walk on $\mathbb{Z}$ with continuous time is a homogeneous Markov chain $X_t$, $t \in [0, \infty)$, with state space $\mathbb{Z}$ and nearest-neighbor (or at least bounded) jumps. The term "Markov" as usual refers to the "lack of memory" property, which amounts to saying that from the entire history of the process development up to a given time, only the current position of the walk is important for the future evolution while all other information is irrelevant.

Since there is no smallest time unit as in the discrete-time case, it is convenient to describe transitions of $X_t$ in terms of transition rates characterizing the likelihood of various jumps during a very short time. More precisely, if $p_{xy}(t) := P\{X_t = y \mid X_0 = x\}$ are the transition probabilities over time $t$, then for $h \to 0$

$$\begin{aligned} p_{xy}(h) &= c_{xy}h + o(h) \quad (x \neq y) \\ p_{xx}(h) &= 1 - h\sum_{y\neq x} c_{xy} + o(h) \end{aligned} \qquad [44]$$

Equations for the functions $p_{xy}(t)$ can then be derived by adapting the method of decomposition commonly used for discrete-time Markov chains (cf. the section "Transience and recurrence"). Here it is more convenient to decompose with respect to the "last" step, that is, by considering all possible transitions during a small increment of time at the end of the time interval $[0, t+h]$. Using Markov property and eqn [44] we can write

$$\begin{aligned} p_{0x}(t+h) &= h\sum_{y\neq x} p_{0y}(t)\, c_{yx} \\ &\quad + p_{0x}(t)\left(1 - h\sum_{y\neq x} c_{xy}\right) + o(h) \end{aligned}$$

which in the limit $h \to 0$ yields the master equation (or Chapman–Kolmogorov's forward equation)

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} p_{0x}(t) &= \sum_{y\neq x} \{c_{yx}p_{0y}(t) - c_{xy}p_{0x}(t)\} \\ p_{0x}(0) &= \delta_0(x) \end{aligned} \qquad [45]$$

where $\delta_0(x)$ is the Kronecker symbol.

Continuous-time RWRE are therefore naturally described via the randomized master equation, that is, with random transition rates. The canonical example, originally motivated by Dyson's study of the chain of harmonic oscillators with random couplings, is a symmetric nearest-neighbor RWRE, where the random transition rates $c_{xy}$ are nonzero only for $y = x \pm 1$ and satisfy the condition $c_{x,x+1} = c_{x+1,x}$, otherwise being i.i.d. (see Alexander *et al.* (1981)). In this case, the problem [45] can be formally solved using the Laplace transform, leading to the equations

$$s + G_0^+ + G_0^- = [\hat{p}_0(s)]^{-1} \qquad [46]$$

$$s + G_x^- + G_x^+ = 0 \quad (x \neq 0) \qquad [47]$$

where $G_x^-, G_x^+$ are defined as

$$G_x^{\pm} := c_{x,x\pm1}\,\frac{\hat{p}_{0x}(s) - \hat{p}_{0,x\pm1}(s)}{\hat{p}_{0x}(s)} \qquad [48]$$

and $\hat{p}_{0x}(s) := \int_0^\infty p_{0x}(t)\, \mathrm{e}^{-st}\mathrm{d}t$. From eqns [47] and [48] one obtains the recursion

$$\begin{aligned} G_x^{\pm} &= \left(\frac{1}{c_{x,x\pm1}} + \frac{1}{s + G_{x\pm1}^{+}}\right)^{-1} \\ x &= 0, \pm1, \pm2, \ldots \end{aligned} \qquad [49]$$

The quantities $G_0^{\pm}$ are therefore expressed as infinite continued fractions depending on $s$ and the random variables $c_{x,x\pm1}$, $c_{x,x\pm2}, \ldots$. The function $\hat{p}_{00}(s)$ can then be found from eqn [46].

In its generality, the problem is far too hard, and we shall only comment on how one can evaluate the annealed mean

$$\mathbb{E}\hat{p}_{00}(s) = \mathbb{E}\big(s + G_0^+ + G_0^-\big)^{-1}$$

According to eqn [49], the random variables $G_0^+, G_0^-$ are determined by the same algebraic formula, but involve the rate coefficients from different sides of site $x$, and hence are i.i.d. Furthermore, eqn [49] implies that the random variables $G_0^+, G_1^+$ have the same distribution and, moreover, $G_1^+$ and $c_{01}$ are independent. Therefore, eqn [49] may be used as an integral equation for the unknown density function of $G_0^+$. It can be proved that the suitable solution exists and is unique, and

although an explicit solution is not available, one can obtain the asymptotics of small values of $s$, thereby rendering information about the behavior of $p_{00}(t)$ for large $t$. More specifically, one can show that if $c_* := (\mathbb{E}c_{01}^{-1})^{-1} > 0$, then

$$\mathbb{E}\hat{p}_{00}(s) \sim (4c_* s)^{-1/2}, \quad s \to 0$$

and so by a Tauberian theorem

$$\mathbb{E}p_{00}(t) \sim (4\pi c_* t)^{-1/2}, \quad t \to \infty \qquad [50]$$

Note that asymptotics [50] appears to be the same as for an ordinary symmetric random walk with constant transition rates $c_{x,x+1} = c_{x+1,x} = c_*$, suggesting that the latter provides an EMA for the RWRE considered above.

This is further confirmed by the asymptotic calculation of the annealed mean square displacement, $\boldsymbol{E}_0 X_t^2 \sim 2c_* t$ as $t \to \infty$ (Alexander *et al.* 1981). Moreover, Kawazu and Kesten (1984) proved that $X_t$ is asymptotically normal:

$$\lim_{t\to\infty} \boldsymbol{P}_0 \left\{ \frac{X_t}{\sqrt{2c_* t}} \le x \right\} = \Phi(x) \qquad [51]$$

Therefore, if $c_* > 0$, then the RWRE has the same diffusive behavior as the corresponding ordered system, with a well-defined diffusion constant $D = c_*$.

In the case where $c_* = 0$ (i.e., $\mathbb{E}c_{01}^{-1} = \infty$), one may expect that the RWRE exhibits subdiffusive behavior. For example, if the density function of the transition rates is modeled by

$$f_a(u) = (1-\alpha)\, u^{-\alpha} \mathbf{1}_{\{0<u<1\}} \quad (0 < \alpha < 1)$$

then, as shown by Alexander *et al.* (1981),

$$\mathbb{E}p_{00}(t) \sim C_\alpha t^{-(1-\alpha)/(2-\alpha)}$$
$$\boldsymbol{E}_0 X_t^2 \sim C'_\alpha t^{2(1-\alpha)/(2-\alpha)}$$

In fact, Kawazu and Kesten (1984) proved that in this case $t^{-\alpha/(1+\alpha)} X_t$ has a (non-Gaussian) limit distribution as $t \to \infty$.

To conclude the discussion of the continuous-time case, let us point out that some useful information about recurrence of $X_t$ can be obtained by considering an imbedded (discrete-time) random walk $\tilde{X}_n$, defined as the position of $X_t$ after $n$ jumps. Note that continuous-time Markov chains admit an alternative description of their evolution in terms of sojourn times and the distribution of transitions at a jump. Namely, if the environment $\omega$ is fixed, then the random sojourn time of $X_t$ in each state $x$ is exponentially distributed with mean $1/c_x$, where $c_x := \sum_{y \ne x} c_{xy}$, while the distribution of transitions from $x$ is given by the probabilities $p_{xy} = c_{xy}/c_x$.

For the symmetric nearest-neighbor RWRE considered above, the transition probabilities of the imbedded random walk are given by

$$p_x := p_{x,x+1} = \frac{c_{x,x+1}}{c_{x-1,x} + c_{x,x+1}}$$
$$q_x := p_{x,x-1} = 1 - p_x$$

and we recognize here the transition law of a random walk in the random bonds environment considered in the previous subsection (cf. eqn [41]). Recurrence and zero asymptotic velocity established there are consistent with the results discussed in the present section (e.g., note that the CLT for both $X_n$, eqn [43], and $X_t$, eqn [51], does not involve any centering). Let us point out, however, that a "naive" discretization of time using the mean sojourn time appears to be incorrect, as this would lead to the scaling $t = n\delta_1$ with $\delta_1 := \mathbb{E}(c_{-1,0} + c_{01})^{-1}$, while from comparing the limit theorems in these two cases, one can conclude that the true value of the effective discretization step is given by $\delta_* := (2c_*)^{-1} = (1/2)\mathbb{E}c_{01}^{-1}$. In fact, by the arithmetic–harmonic mean inequality we have $\delta_* > \delta_1$, which is a manifestation of the RWRE's diffusive slowdown.

## RWRE in Higher Dimensions

Multidimensional RWRE with nearest-neighbor jumps are defined in a similar fashion: from site $x \in \mathbb{Z}^d$ the random walk can jump to one of the $2d$ adjacent sites $x + e \in \mathbb{Z}^d$ (such that $|e| = 1$), with probabilities $p_x(e) \ge 0$, $\sum_{|e|=1} p_x(e) = 1$, where the random vectors $p_x(\cdot)$ are assumed to be i.i.d. for different $x \in \mathbb{Z}^d$. As usual, we will also impose the condition of uniform ellipticity:

$$p_x(e) \ge \delta > 0, \quad \mathbb{P}\text{-a.s.}$$
$$|e| = 1, \quad x \in \mathbb{Z}^d \qquad [52]$$

In contrast to the one-dimensional case, theory of RWRE in higher dimensions is far from maturity. Possible asymptotic behaviors of the RWRE for $d \ge 2$ are not understood well enough, and many basic questions remain open. For instance, no definitive classification of the RWRE is available regarding transience and recurrence. Similarly, LLN and CLT have been proved only for a limited number of specific models, while no general sharp results have been obtained. On a more positive note, there has been considerable progress in recent years in the so-called ballistic case, where powerful techniques have been developed (see Sznitman (2002, 2004) and Zeitouni (2003, 2004)). Unfortunately, not much is

known for nonballistic RWRE, apart from special cases of balanced RWRE in $d \geq 2$ (Lawler 1982), small isotropic perturbations of ordinary symmetric random walks in $d \geq 3$ (Bricmont and Kupiainen 1991), and some examples based on combining components of ordinary random walks and RWRE in $d \geq 7$ (Bolthausen *et al.* 2003). In particular, there are no examples of subdiffusive behavior in any dimension $d \geq 2$, and in fact it is largely believed that a CLT is always true in any uniformly elliptic, i.i.d. random environment in dimensions $d \geq 3$, with somewhat less certainty about $d = 2$. A heuristic explanation for such a striking difference with the case $d = 1$ is that due to a less restricted topology of space in higher dimensions, it is much harder to force the random walk to visit traps, and hence the slowdown is not so pronounced.

In what follows, we give a brief account of some of the known results and methods in this fast-developing area (for further information and specific references, see an extensive review by Zeitouni (2004)).

### Zero–One Laws and LLNs

A natural first step in a multidimensional context is to explore the behavior of the random walk $X_n$ as projected on various one-dimensional straight lines. Let us fix a test unit vector $\ell \in \mathbb{R}^d$, and consider the process $Z_n^\ell := X_n \cdot \ell$. Then for the events $A_{\pm\ell} := \{\lim_{n\to\infty} Z_n^\ell = \pm\infty\}$ one can show that

$$\boldsymbol{P}_0(A_\ell \cup A_{-\ell}) \in \{0, 1\} \qquad [53]$$

That is to say, for each $\ell$ the probability that the random walk escapes to infinity in the direction $\ell$ is either 0 or 1.

Let us sketch the proof. We say that $\tau$ is "record time" if $|Z_t^\ell| > |Z_k^\ell|$ for all $k < t$, and "regeneration time" if in addition $|Z_\tau^\ell| \leq |Z_n^\ell|$ for all $n \geq \tau$. Note that by the ellipticity condition [52], $\overline{\lim}_{n\to\infty} |Z_n^\ell| = \infty$ ($\boldsymbol{P}_0$-a.s.), hence there is an infinite sequence of record times $0 = \tau_0 < \tau_1 < \tau_2 < \cdots$. If $\boldsymbol{P}_0(A_\ell \cup A_{-\ell}) > 0$, we can pick a subsequence of record times $\tau_i'$, each of which has a positive $\boldsymbol{P}_0$-probability to be a regeneration time (because otherwise $|Z_n^\ell|$ would persistently backtrack towards the origin and the event $A_\ell \cup A_{-\ell}$ could not occur). Since the trials for different record times are independent, it follows that a regeneration time $\tau^*$ occurs $\boldsymbol{P}_0$-a.s. Repeating this argument, we conclude that there exists an infinite sequence of regeneration times $\tau_i^*$, which implies that $|Z_n^\ell| \to \infty$ ($\boldsymbol{P}_0$-a.s.), that is, $\boldsymbol{P}(A_\ell \cup A_{-\ell}) = 1$.

Regeneration structure introduced by the sequence $\{\tau_i^*\}$ plays a key role in further analysis of the RWRE and is particularly useful for proving an LLN and a CLT, due to the fact that pieces of the random walk between consecutive regeneration times (and fragments of the random environment involved thereby) are independent and identically distributed (at least starting from $\tau_1^*$). In this vein, one can prove a "directional" version of the LLN, stating that for each $\ell$ there exist deterministic $v_\ell, v_{-\ell}$ (possibly zero) such that

$$\lim_{n\to\infty} \frac{Z_n^\ell}{n} = v_\ell \mathbf{1}_{A_\ell} + v_{-\ell} \mathbf{1}_{A_{-\ell}}, \quad \boldsymbol{P}_0\text{-a.s.} \qquad [54]$$

Note that if $\boldsymbol{P}_0(A_\ell) \in \{0, 1\}$, then eqn [54] in conjunction with eqn [53] would readily imply

$$\lim_{n\to\infty} \frac{Z_n^\ell}{n} = v_\ell, \quad \boldsymbol{P}_0\text{-a.s.} \qquad [55]$$

Moreover, if $\boldsymbol{P}_0(A_\ell) \in \{0, 1\}$ for any $\ell$, then there exists a deterministic $v$ (possibly zero) such that

$$\lim_{n\to\infty} \frac{X_n}{n} = v, \quad \boldsymbol{P}_0\text{-a.s.} \qquad [56]$$

Therefore, it is natural to ask if a zero–one law [53] can be enhanced to that for the individual probabilities $\boldsymbol{P}_0(A_\ell)$. It is known that the answer is affirmative for i.i.d. environments in $d = 2$, where indeed $\boldsymbol{P}(A_\ell) \in \{0, 1\}$ for any $\ell$, with counterexamples in certain stationary ergodic (but not uniformly elliptic) environments. However, in the case $d \geq 3$ this is an open problem.

### Kalikow's Condition and Sznitman's Condition (T′)

An RWRE is called "ballistic" (ballistic in direction $\ell$) if $v \neq 0$ ($v_\ell \neq 0$), see eqns [55] and [56]. In this section, we describe conditions on the random environment which ensure that the RWRE is ballistic.

Let $U$ be a connected strict subset of $\mathbb{Z}^d$ containing the origin. For $x \in U$, denote by

$$g(x, \omega) := E_0^\omega \sum_{n=0}^{T_U} \mathbf{1}_{\{X_n = x\}}$$

the quenched mean number of visits to $x$ prior to the exit time $T_U := \min\{n \geq 0 : X_n \notin U\}$. Consider an auxiliary Markov chain $\widehat{X}_n$, which starts from 0, makes nearest-neighbor jumps while in $U$, with (nonrandom) probabilities

$$\widehat{p}_x(e) = \frac{\mathbb{E}[g(x, \omega) p_x(e)]}{\mathbb{E}[g(x, \omega)]}, \quad x \in U \qquad [57]$$

and is absorbed as soon as it first leaves $U$. Note that the expectations in eqn [57] are finite; indeed, if $\alpha_x$ is the probability to return to $x$ before leaving $U$,

then, by the Markov property, the mean number of returns is given by

$$\sum_{k=1}^{\infty} k\alpha_x^k(1-\alpha_x) = \frac{\alpha_x}{1-\alpha_x} < \infty$$

since, due to ellipticity, $\alpha_x < 1$.

An important property, highlighting the usefulness of $\widehat{X}_n$, is that if $\widehat{X}_n$ leaves $U$ with probability 1, then the same is true for the original RWRE $X_n$ (under the annealed law $\boldsymbol{P}_0$), and moreover, the exit points $\widehat{X}_{\hat{T}_U}$ and $X_{T_U}$ have the same distribution laws.

Let $\ell \in \mathbb{R}^d, |\ell| = 1$. One says that Kalikow's condition with respect to $\ell$ holds if the local drift of $\widehat{X}_n$ in the direction $\ell$ is uniformly bounded away from zero:

$$\inf_U \inf_{x\in U} \sum_{|e|=1} (e\cdot\ell)\,\widehat{p}_x(e) > 0 \qquad [58]$$

A sufficient condition for [58] is, for example, that for some $\kappa > 0$

$$\mathbb{E}\big[(d(0,\omega)\cdot\ell)_+\big] \geq \kappa\mathbb{E}\big[(d(0,\omega)\cdot\ell)_-\big]$$

where $d(0,\omega) = E_0^\omega X_1$ and $u_\pm := \max\{\pm u, 0\}$.

A natural implication of Kalikow's condition [58] is that $\boldsymbol{P}_0(A_\ell) = 1$ and $v_\ell > 0$ (see eqn [55]). Moreover, noting that eqn [58] also holds for all $\ell'$ in a vicinity of $\ell$ and applying the above result with $d$ noncollinear vectors from that vicinity, we conclude that under Kalikow's condition there exists a deterministic $v \neq 0$ such that $X_n/n \to v$ as $n \to \infty$ ($\boldsymbol{P}_0$-a.s.). Furthermore, it can be proved that $(X_n - nv)/\sqrt{n}$ converges in law to a Gaussian distribution (see Sznitman (2004)).

It is not hard to check that in dimension $d = 1$ Kalikow's condition is equivalent to $v \neq 0$ and therefore characterizes completely all ballistic walks. For $d \geq 2$, the situation is less clear; for instance, it is not known if there exist RWRE with $\boldsymbol{P}(A_\ell) > 0$ and $v_\ell = 0$ (of course, such RWRE cannot satisfy Kalikow's condition).

Sznitman (2004) has proposed a more complicated transience condition (T′) involving certain regeneration times $\tau_i^*$ similar to those described in the previous subsection. An RWRE is said to satisfy Sznitman's condition (T′) relative to direction $\ell$ if $\boldsymbol{P}_0(A_\ell) = 1$ and for some $c > 0$ and all $0 < \gamma < 1$

$$\boldsymbol{E}_0 \exp\left(c \sup_{n\leq\tau_1^*} |X_n|^\gamma\right) < \infty \qquad [59]$$

This condition provides a powerful control over $\tau_1^*$ for $d \geq 2$ and in particular ensures that $\tau_1^*$ has finite moments of any order. This is in sharp contrast with the one-dimensional case, and should be viewed as a reflection of much weaker traps in dimensions $d \geq 2$.

Condition [59] can also be reformulated in terms of the exit distribution of the RWRE from infinite thick slabs "orthonormal" to directions $\ell'$ sufficiently close to $\ell$. As it stands, the latter reformulation is difficult to check, but Sznitman (2004) has developed a remarkable "effective" criterion reducing the job to a similar condition in finite boxes, which is much more tractable and can be checked in a number of cases.

In fact, condition (T′) follows from Kalikow's condition, but not the other way around. In the one-dimensional case, condition (T′) (applied to $\ell = 1$ and $\ell = -1$) proves to be equivalent to the transient behavior of the RWRE, which, as we have seen in Theorem 2, may happen with $v = 0$, that is, in a nonballistic scenario. The situation in $d \geq 2$ is quite different, as condition (T′) implies that the RWRE is ballistic in the direction $\ell$ (with $v_\ell > 0$) and satisfies a CLT (under $\boldsymbol{P}_0$). It is not known whether the ballistic behavior for $d \geq 2$ is completely characterized by condition (T′), although this is expected to be true.

### Balanced RWRE

In this section we discuss a particular case of nonballistic RWRE, for which LLN and CLT can be proved. Following Lawler (1982), we say that an RWRE is "balanced" if $p_x(e) = p_x(-e)$ for all $x \in \mathbb{Z}^d, |e| = 1$ ($\mathbb{P}$-a.s.). In this case, the local drift vanishes, $d(x,\omega) = 0$, hence the coordinate processes $X_n^i$ $(i = 1, \ldots, d)$ are martingales with respect to the natural filtration $\mathcal{F}_n = \sigma\{X_0, \ldots, X_n\}$. The quenched covariance matrix of the increments $\Delta X_n^i := X_{n+1}^i - X_n^i$ $(i = 1, \ldots, d)$ is given by

$$E_0^\omega\big[\Delta X_n^i \Delta X_n^j | \mathcal{F}_n\big] = 2\delta_{ij} p_{X_n}(e_i) \qquad [60]$$

Since the right-hand side of eqn [60] is uniformly bounded, it follows that $X_n/n \to 0$ ($\boldsymbol{P}_0$-a.s.). Further, it can be proved that there exist deterministic positive constants $a_1, \ldots, a_d$ such that for $i = 1, \ldots, d$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} p_{X_k}(e_i) = \frac{a_i}{2}, \quad \boldsymbol{P}_0\text{-a.s.} \qquad [61]$$

Once this is proved, a multidimensional CLT for martingale differences yields that $X_n/\sqrt{n}$ converges in law to a Gaussian distribution with zero mean and the covariances $b_{ij} = \delta_{ij} a_i$.

The proof of [61] employs the method of environment viewed from the particle. Namely, define a Markov chain $\omega_n := \theta^{X_n}\omega$ with the transition kernel

$$T(\omega, \mathrm{d}\omega') = \sum_{i=1}^{d} [p_0(e_i)\delta_{\theta\omega}(\mathrm{d}\omega') + p_0(-e_i)\delta_{\theta^{-1}\omega}(\mathrm{d}\omega')]$$

(cf. eqn [27]). The next step is to find a probability measure $\mathbb{Q}$ on $\Omega$ invariant under $T$ and absolutely continuous with respect to $\mathbb{P}$. Unlike the one-dimensional case, however, an explicit form of $\mathbb{Q}$ is not available, and $\mathbb{Q}$ is constructed indirectly as the limit of invariant measures of certain periodic modifications of the RWRE. Birkhoff's ergodic theorem then yields, $\boldsymbol{P}_0$-a.s.,

$$\frac{1}{n}\sum_{k=0}^{n-1} p_{X_k}(e_i,\omega) = \frac{1}{n}\sum_{k=0}^{n-1} p_0(e_i,\omega_k) \to \int_\Omega p_0(e_i,\omega)\ \mathbb{Q}(\mathrm{d}\omega) \geq \delta$$

by the ellipticity condition [52], and eqn [61] follows.

With regard to transience, balanced RWREs admit a complete and simple classification. Namely, it has been proved (see Zeitouni (2004)) that any balanced RWRE is transient for $d \geq 3$ and recurrent for $d = 2$ ($\boldsymbol{P}_0$-a.s.). It is interesting to note, however, that these answers may be false for certain balanced random walks in a fixed environment ($\mathbb{P}$-probability of such environments being zero, of course). Indeed, examples can be constructed of balanced random walks in $\mathbb{Z}^2$ and in $\mathbb{Z}^d$ with $d \geq 3$, which are transient and recurrent, respectively (Zeitouni 2004).

### RWRE Based on Modification of Ordinary Random Walks

A number of partial results are known for RWRE constructed on the basis of ordinary random walks via certain randomization of the environment. A natural model is obtained by a small perturbation of a simple symmetric random walk. To be more precise, suppose that: (1) $|p_x(e) - 1/2d| < \varepsilon$ for all $x \in \mathbb{Z}^d$ and any $|e| = 1$, where $\varepsilon > 0$ is small enough; (2) $\mathbb{E}p_x(e) = 1/2d$; (3) vectors $p_x(\cdot)$ are i.i.d. for different $x \in \mathbb{Z}^d$; and (4) the distribution of the vector $p_x(\cdot)$ is isotropic, that is, invariant with respect to permutations of its coordinates. Then for $d \geq 3$ Bricmont and Kupiainen (1991) have proved an LLN (with zero asymptotic velocity) and a quenched CLT (with nondegenerate covariance matrix). The proof is based on the renormalization group method, which involves decimation in time combined with a suitable spatial–temporal scaling. This transformation replaces an RWRE by another RWRE with weaker randomness, and it can be shown that iterations converge to a Gaussian fixed point.

Another class of examples is also built using small perturbations of simple symmetric random walks, but is anisotropic and exhibits ballistic behavior, providing that the annealed local drift in some direction is strong enough (see Sznitman (2004)). More precisely, suppose that $d \geq 3$ and $\eta \in (0,1)$. Then there exists $\varepsilon_0 = \varepsilon_0(d,\eta) > 0$ such that if $|p_x(e) - 1/2d| < \varepsilon$ ($x \in \mathbb{Z}^d, |e| = 1$) with $0 < \varepsilon < \varepsilon_0$, and for some $e_0$ one has $\mathbb{E}[d(x,\omega)\cdot e_0] \geq \varepsilon^{2.5-\eta}$ ($d = 3$) or $\geq \varepsilon^{3-\eta}$ ($d \geq 4$), then Sznitman's condition (T′) is satisfied with respect to $e_0$ and therefore the RWRE is ballistic in the direction $e_0$ (cf. the subsection "Kalikow's condition and Sznitman's condition (T′)").

Examples of a different type are constructed in dimensions $d \geq 6$ by letting the first $d_1 \geq 5$ coordinates of the RWRE $X_n$ behave according to an ordinary random walk, while the remaining $d_2 = d - d_1$ coordinates are exposed to a random environment (see Bolthausen *et al.* (2003)). One can show that there exists a deterministic $v$ (possibly zero) such that $X_n/n \to v$ ($\boldsymbol{P}_0$-a.s.). Moreover, if $d_1 \geq 13$, then $(X_n - nv)/\sqrt{n}$ satisfies both quenched and annealed CLT. Incidentally, such models can be used to demonstrate the surprising features of the multidimensional RWRE. For instance, for $d \geq 7$ one can construct an RWRE $X_n$ such that the annealed local drift does not vanish, $\mathbb{E}d(x,\omega) \neq 0$, but the asymptotic velocity is zero, $X_n/n \to 0$ ($\boldsymbol{P}_0$ - a.s.), and furthermore, if $d \geq 15$, then in this example $X_n/\sqrt{n}$ satisfies a quenched CLT. (In fact, one can construct such RWRE as small perturbations of a simple symmetric walk.) On the other hand, there exist examples (in high enough dimensions) where the walk is ballistic with a velocity which has an opposite direction to the annealed drift $\mathbb{E}d(x,\omega) \neq 0$. These striking examples provide "experimental" evidence of many unusual properties of the multidimensional RWRE, which, no doubt, will be discovered in the years to come.

*See also:* Averaging Methods; Growth Processes in Random Matrix Theory; Lagrangian Dispersion (Passive Scalar); Random Dynamical Systems; Random Matrix Theory in Physics; Stochastic Differential Equations; Stochastic Loewner Evolutions.

## Further Reading

Alexander S, Bernasconi J, Schneider WR, and Orbach R (1981) Excitation dynamics in random one-dimensional systems. *Reviews of Modern Physics* 53: 175–198.

Bernasconi J and Schneider WR (1985) Random walks in one-dimensional random media. *Helvetica Physica Acta* 58: 597–621.

Bolthausen E and Goldsheid I (2000) Recurrence and transience of random walks in random environments on a strip. *Communications in Mathematical Physics* 214: 429–447.

Bolthausen E, Sznitman A-S, and Zeitouni O (2003) Cut points and diffusive random walks in random environments. *Annales de l'Institut Henri Poincaré. Probabilités et Statistiques* 39: 527–555.

Bouchaud J-P and Georges A (1990) Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications. *Physical Reports* 195: 127–293.

Brémont J (2004) Random walks in random medium on $\mathbb{Z}$ and Lyapunov spectrum. *Annales de l'Institut Henri Poincaré. Probabilités et Statistiques* 40: 309–336.

Bricmont J and Kupiainen A (1991) Random walks in asymmetric random environments. *Communications in Mathematical Physics* 142: 345–420.

Hughes BD (1995) *Random Walks and Random Environments. Volume 1: Random Walks*. Oxford: Clarendon.

Hughes BD (1996) *Random Walks and Random Environments. Volume 2: Random Environments*. Oxford: Clarendon.

Kawazu K and Kesten H (1984) On birth and death processes in symmetric random environment. *Journal of Statistical Physics* 37: 561–576.

Kesten H, Kozlov MV, and Spitzer F (1975) A limit law for random walk in a random environment. *Compositio Mathematica* 30: 145–168.

Kozlov SM and Molchanov SA (1984) On conditions for applicability of the central limit theorem to random walks on a lattice. *Soviet Mathematics Doklady* 30: 410–413.

Lawler GF (1982) Weak convergence of a random walk in a random environment. *Communications in Mathematical Physics* 87: 81–87.

Molchanov SA (1994) Lectures on random media. In: Bernard P (ed.) *Lectures on Probability Theory*, Ecole d'Eté de Probabilités de Saint-Flour XXII-1992, Lecture Notes in Mathematics, vol. 1581, pp. 242–411. Berlin: Springer.

Révész P (1990) *Random Walk in Random and Non-Random Environments*. Singapore: World Scientific.

Sinai YaG (1982) The limiting behavior of a one-dimensional random walk in a random medium. *Theory of Probability and Its Applications* 27: 256–268.

Solomon F (1975) Random walks in a random environment. *The Annals of Probability* 3: 1–31.

Sznitman A-S (2002) Lectures on random motions in random media. In: Bolthausen E and Sznitman A-S. *Ten Lectures on Random Media*, DMV Seminar, vol. 32. Basel: Birkhäuser.

Sznitman A-S (2004) Topics in random walks in random environment. In: Lawler GF (ed.) *School and Conference on Probability Theory* (Trieste, 2002), ICTP Lecture Notes Series, vol. XVII, pp. 203–266 (Available at http://www.ictp.trieste.it/~pub_off/lectures/vol17.html).

Zeitouni O (2003) Random walks in random environments. In: Tatsien Li (ed.) *Proceedings of the International Congress of Mathematicians* (Beijing, 2002), vol. III, pp. 117–127. Beijing: Higher Education Press.

Zeitouni O (2004) Random walks in random environment. In: Picard J (ed.) *Lectures on Probability Theory and Statistics*, Ecole d'Eté de Probabilités de Saint-Flour XXXI-2001, Lecture Notes in Mathematics, vol. 1837, pp. 189–312. New York: Springer.


# Recursion Operators in Classical Mechanics


**F Magri**, Università di Milano Bicocca, Milan, Italy
**M Pedroni**, Università di Bergamo, Dalmine (BG), Italy




## Introduction

One of the tasks of classical mechanics has always been to identify those Hamiltonian systems which, by their peculiar properties, are considered solvable. The integrable systems of Liouville and the separable systems of Jacobi can serve as representative examples here. The bi-Hamiltonian geometry, a branch of Poisson geometry dealing with a special kind of deformation of Poisson bracket, suggests two further classes of Hamiltonian systems – the bi-Hamiltonian systems and the cyclic systems of Levi-Civita. The purpose of this article is to investigate the second class of systems mentioned above, and to explain why they are relevant for classical mechanics. (*see* Bi-Hamiltonian Methods in Soliton Theory and Multi-Hamiltonian Systems for further details).

To define a cyclic system of Levi-Civita, one must consider a symplectic manifold $(S, \omega)$ endowed with a tensor field of type $(1, 1)$, seen as an endomorphism $N : TS \to TS$ that obeys two conditions. The first condition is that the vector-valued 2-form

$$T_N(X, Y) = [NX, NY] - N[NX, Y] - N[X, NY] + N^2[X, Y]$$

(called the Nijenhuis torsion of $N$) vanishes identically. In this case $N$ is termed a "recursion operator." The second condition is that

$$\omega'(X, Y) = \omega(NX, Y)$$

is a closed 2-form. The manifolds where these conditions are fulfilled are called $\omega N$ manifolds. On these manifolds, each Hamiltonian vector field $X_h$ is embedded into the distribution

$$D_h = \langle X_h, NX_h, N^2X_h, \ldots \rangle$$

which is the minimal invariant distribution containing $X_h$. This can be called the Levi-Civita distribution generated by $X_h$. Experience has shown that $D_h$ is seldom integrable. The cyclic systems of Levi-Civita are, by definition, the generators of the integrable Levi-Civita distributions. Even though this notion is new in classical mechanics, many interesting classical systems display this property.

The aim of this article is to show that the cyclic systems of Levi-Civita are closely related to

separable systems of Jacobi. To this end, the article is organized in four sections, of which the first three clarify the above-mentioned concepts. In the section "$\omega N$ manifolds," the idea of $\omega N$ manifolds is explained from the viewpoint of bi-Hamiltonian geometry. The section "Cotangent bundles" shows that cotangent bundles provide a large class of $\omega N$ manifolds, proving that such manifolds are not rare. Next, two basic examples of cyclic systems of Levi-Civita are presented. Finally, the relation between cyclic systems of Levi-Civita and separable systems of Jacobi is explained briefly.

## $\omega N$ Manifolds

Let us consider a symplectic manifold $(S, \omega)$ with its Hamiltonian vector fields $X_h$ defined by

$$\omega(X_h, \cdot) = -\mathrm{d}h$$

and with the Poisson bracket

$$\{f, g\} = \omega(X_f, X_g)$$

Both the Hamiltonian vector fields and the functions on $S$ form a Lie algebra, and these algebras are homomorphic, since

$$[X_f, X_g] = X_{\{f,g\}}$$

The bi-Hamiltonian geometry is the study of the deformations of the Lie algebras which preserve the above morphism.

We start from the deformations of the Poisson algebra of functions, by replacing the bracket $\{f, g\}$ with the linear pencil

$$\{f, g\}_\epsilon = \{f, g\} + \epsilon\{f, g\}', \quad \epsilon \in \mathbb{R}$$

The problem is to find $\{f, g\}'$ in such a way that the linear pencil satisfies the Jacobi identity for any value of the parameter $\epsilon$. To solve this problem it is convenient to represent the bracket $\{f, g\}'$ in the form

$$\{f, g\}' = \omega'(X_f, X_g)$$

(which is analogous to the standard representation of the Poisson bracket of $S$) and then to notice that there exists a unique (1, 1) tensor field $N: TS \to TS$ such that

$$\omega'(X_f, X_g) = \omega(NX_f, X_g)$$

Due to the skew-symmetry of $\omega'$, the tensor field $N$ must satisfy the condition

$$\omega(NX_f, X_g) = \omega(X_f, NX_g)$$

To the first order in $\epsilon$, the Jacobi identity on $\{f, g\}_\epsilon$ gives

$$\{\{f, g\}, h\}' + \{\{f, g\}', h\} + \text{cyclic permutations} = 0$$

This condition entails a constraint on $\omega'$. One can readily check that $\omega'$ must be a closed 2-form:

$$\mathrm{d}\omega' = 0$$

In turn, this constraint imposes a condition on $N$. The translation of the closure of $\omega'$ on $N$ is

$$[NX_f, X_g] + [X_f, NX_g] - N[X_f, X_g] = X_{\{f,g\}'}$$

To the second order in $\epsilon$, the Jacobi identity on $\{f, g\}_\epsilon$ gives

$$\{\{f, g\}', h\}' + \text{cyclic permutations} = 0$$

entailing the condition

$$[NX_f, NX_g] = NX_{\{f,g\}'}$$

on $N$. Thus, the Jacobi identity is satisfied at any order in $\epsilon$ if and only if $N$ is torsion free and $\omega'$ is a closed 2-form. Hence, according to the definition given in the "Introduction," the manifold $S$ is an $\omega N$ manifold.

It may be of interest to notice that the bracket

$$[X, Y]_N = [NX, Y] + [X, NY] - N[X, Y]$$

is a new (deformed) commutator on vector fields, since the torsion of $N$ vanishes. The same is also true for

$$[X, Y]_\epsilon = [X, Y] + \epsilon[X, Y]_N$$

since the torsion of $(\mathrm{Id} + \epsilon N)$ vanishes too. Therefore, one can write

$$[X_f, X_g]_\epsilon = X_{\{f,g\}_\epsilon}$$

This formula shows that this process of deformation is rigid. For each change of the Poisson bracket, there is a deformation of the commutator of vector fields such that the basic correspondence between functions and Hamiltonian vector fields, established by the symplectic form $\omega$, remains a Lie algebra morphism.

The same phenomenon can be observed in connection with the definition of Hamiltonian vector field. If one introduces the pencil of 2-forms

$$\omega_\epsilon = \omega + \epsilon\omega'$$

and the pencil of derivations

$$\mathrm{d}_\epsilon = \mathrm{d} + \epsilon \mathrm{d}_N$$

where $\mathrm{d}_N$ is the derivation of type d and degree 1 canonically associated with $N$ according to the

theory of graded derivations of Frölicher and Nijenhuis, one can prove that

$$d_\epsilon^2 = 0, \qquad d_\epsilon \omega_\epsilon = 0$$

and that

$$\omega_\epsilon(X_h, \cdot) = -d_\epsilon h$$

This means that, on an $\omega N$ manifold, the symplectic form $\omega$ and the de Rham differential d are deformed in such a way that the basic relation between functions and Hamiltonian vector fields established by $\omega$ holds true.

## Cotangent Bundles

Cotangent bundles are a source of examples of $\omega N$ manifolds. The construction begins on the base manifold $Q$. For any $(1,1)$ tensor field $L: TQ \to TQ$ with vanishing Nijenhuis torsion, one constructs the deformed Liouville 1-form

$$\theta' = \sum_{i=1}^{n} y_i L^*(dx^i)$$

and its exterior derivative

$$\omega' = d\theta'$$

It can be proved that $\omega'$ satisfies the conditions explained in the previous section, and conclude that $T^*Q$, endowed with the pencil of 2-form $\omega_\epsilon = \omega +_\epsilon \omega'$, is an $\omega N$ manifold.

A subclass of these structures merits attention. It is related to the polynomials

$$s(\lambda) = \lambda^n - \left(s_1\lambda^{n-1} + s_2\lambda^{n-2} + \cdots + s_n\right)$$

the coefficients of which are functions on $Q$ satisfying the condition

$$ds_1 \wedge ds_2 \wedge \cdots \wedge ds_n \neq 0$$

(almost) everywhere on $Q$. Moreover, it is convenient to assume that the roots $(\lambda_1, \lambda_2, \ldots, \lambda_n)$ of $s(\lambda)$ are distinct and real, so that they are functionally independent and can be used as coordinates on $Q$. Therefore, the choice of $s(\lambda)$ is equivalent to fix a special system of coordinates on $Q$, as it happens in $\mathbb{R}^3$ when one introduces the elliptical coordinates as the roots of the polynomial

$$s(\lambda) = (\lambda - a)(\lambda - b)(\lambda - c) \times \left(1 + \frac{x^2}{\lambda - a} + \frac{y^2}{\lambda - b} + \frac{z^2}{\lambda - c}\right)$$

The peculiarity of this situation is that there exists a unique recursion operator $L: TQ \to TQ$ whose characteristic polynomial is $s(\lambda)$. Thus, the choice of $s(\lambda)$ also determines an $\omega N$ structure on $T^*Q$ according to the previous prescription. The conclusion is that there is a relation between pencils of Poisson brackets on $T^*Q$ and coordinate systems on $Q$. This relation is the clue to understand the geometry of separable systems of Jacobi.

## Cyclic Systems of Levi-Civita

The systems of coupled harmonic oscillators are the first example of cyclic systems of Levi-Civita. Let us consider, for simplicity, a system formed by only two particles, with masses $m_1$ and $m_2$, moving on a line under the action of an internal elastic force. The Lagrangian of the system is

$$L = \tfrac{1}{2}\left(m_1\dot{x}_1^2 + m_2\dot{x}_2^2\right) - \tfrac{1}{2}k(x_1 - x_2)^2$$

and the equations of motion are

$$M\ddot{x} + Kx = 0, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

where

$$M = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}, \qquad K = \begin{pmatrix} k & -k \\ -k & k \end{pmatrix}$$

Under a change of coordinates, the entries of the matrices $M$ and $K$ obey the transformation law of the components of a second-order covariant tensor. Therefore, the entries of the matrix $L = M^{-1}K$ are the components of a tensor field of type $(1,1)$ on $\mathbb{R}^2$. The defining equations of the associated endomorphism $L: T\mathbb{R}^2 \to \mathbb{R}^2$ are

$$L^*(dx_1) = \omega_1^2(dx_2 - dx_1)$$
$$L^*(dx_2) = \omega_2^2(dx_1 - dx_2)$$

if $\omega_1^2 = k/m_1$ and $\omega_2^2 = k/m_2$, and these equations clearly show that $L$ is torsion free. The same argument holds for any system of coupled harmonic oscillators. Therefore, the cotangent bundle associated with any system of coupled harmonic oscillators is an $\omega N$ manifold.

To compute the tensor field $N$ in our example, one has to follow the prescription, passing from

$$\theta' = \left(\omega_1^2 y_1 - \omega_2^2 y_2\right)(dx_2 - dx_1)$$

to

$$\omega' = \left(\omega_1^2\, dy_1 - \omega_2^2\, dy_2\right) \wedge (dx_2 - dx_1)$$

and to

$$N\left(\frac{\partial}{\partial x_1}\right) = \omega_1^2 \frac{\partial}{\partial x_1} - \omega_2^2 \frac{\partial}{\partial x_2}$$
$$N\left(\frac{\partial}{\partial x_2}\right) = -\omega_1^2 \frac{\partial}{\partial x_1} + \omega_2^2 \frac{\partial}{\partial x_2}$$
$$N\left(\frac{\partial}{\partial y_1}\right) = \omega_1^2 \left(\frac{\partial}{\partial y_1} - \frac{\partial}{\partial y_2}\right)$$
$$N\left(\frac{\partial}{\partial y_2}\right) = \omega_2^2 \left(-\frac{\partial}{\partial y_1} + \frac{\partial}{\partial y_2}\right)$$

The Levi-Civita distribution $D_h$ is therefore spanned by the vector fields

$$X_h = k\left[\frac{y_1}{\omega_1^2}\frac{\partial}{\partial x_1} + \frac{y_2}{\omega_2^2}\frac{\partial}{\partial x_2} + (x_2 - x_1)\left(\frac{\partial}{\partial y_1} - \frac{\partial}{\partial y_2}\right)\right]$$
$$NX_h = k\left[\left(y_1 - \frac{\omega_1^2}{\omega_2^2}y_2\right)\frac{\partial}{\partial x_1} + \left(y_2 - \frac{\omega_2^2}{\omega_1^2}y_1\right)\frac{\partial}{\partial x_2} + (\omega_1^2 + \omega_2^2)(x_2 - x_1)\left(\frac{\partial}{\partial y_1} - \frac{\partial}{\partial y_2}\right)\right]$$

related to the Hamiltonian

$$h = \frac{y_1^2}{2m_1} + \frac{y_2^2}{2m_2} + \frac{1}{2}k(x_1 - x_2)^2$$

of the system of coupled oscillators. Since $[X_h, NX_h] = 0$, the distribution is integrable; therefore, the system is a cyclic system of Levi-Civita. This property holds for any system of coupled harmonic oscillators. It will be apparent at the end of this article that this result is due to the eigenvectors of $L$ defining the separation coordinates of the coupled oscillators.

The second and final example of cyclic systems of Levi-Civita is the Neumann system, that is, the anisotropic harmonic oscillator on the sphere $S^2$, whose Lagrangian is

$$L = \tfrac{1}{2}m\left(\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2\right) - \tfrac{1}{2}\left(a_1x_1^2 + a_2x_2^2 + a_3x_3^2\right)$$

with the constraint

$$x_1^2 + x_2^2 + x_3^2 = 1$$

This constraint can be avoided by using the first two Cartesian coordinates $(x_1, x_2)$ as local coordinates on $S^2$. The Hamiltonian of the system can then be written in the form

$$h = \tfrac{1}{2}\left(1 + x_1^2\right)y_1^2 - x_1x_2y_1y_2 + \tfrac{1}{2}\left(1 + x_2^2\right)y_2^2 + \tfrac{1}{2}(a_1 - a_3)x_1^2 + \tfrac{1}{2}(a_2 - a_3)x_2^2$$

where, for simplicity, $m = 1$. Formally one is back in $\mathbb{R}^2$ as in the previous example, but the nonlinearity of the equations of motion hinders us to readily see the appropriate recursion operator $L : T\mathbb{R}^2 \to T\mathbb{R}^2$ to be used to construct the $\omega N$ structure on $T^*\mathbb{R}^2$. Let us however recall that according to Neumann, the system is separable in elliptical spherical (also called spheroconical) coordinates, defined as the roots of the restriction to $S^2$ of the polynomial

$$s(\lambda) = (\lambda - a)(\lambda - b)(\lambda - c)\left(\frac{x_1^2}{\lambda - a} + \frac{x_2^2}{\lambda - b} + \frac{x_3^2}{\lambda - c}\right) = \lambda^2 - (s_1\lambda + s_2)$$

Let us, therefore, use this polynomial to construct the unique recursion operator $L$ having $s(\lambda)$ as its characteristic polynomial. It is given by

$$L^*(\mathrm{d}s_1) = \mathrm{d}s_2 + s_1\,\mathrm{d}s_1$$
$$L^*(\mathrm{d}s_2) = s_2\,\mathrm{d}s_1$$

or, after a brief computation, by

$$L^*(\mathrm{d}x_1) = a_1\,\mathrm{d}x_1 - x_1\,\mathrm{d}\left[\tfrac{1}{2}(a_1 - a_3)x_1^2 + \tfrac{1}{2}(a_2 - a_3)x_2^2\right]$$
$$L^*(\mathrm{d}x_2) = a_2\,\mathrm{d}x_2 - x_2\,\mathrm{d}\left[\tfrac{1}{2}(a_1 - a_3)x_1^2 + \tfrac{1}{2}(a_2 - a_3)x_2^2\right]$$

The situation stays the same as in the previous example. Accordingly, the recursion operator $N$ on $T^*\mathbb{R}^2$ is now given by

$$N^*\,\mathrm{d}x_1 = a_1\,\mathrm{d}x_1 - x_1\,\mathrm{d}f$$
$$N^*\,\mathrm{d}x_2 = a_2\,\mathrm{d}x_2 - x_2\,\mathrm{d}f$$
$$N^*\,\mathrm{d}y_1 = a_1\,\mathrm{d}y_1 - (a_1 - a_3)x_1\,\mathrm{d}g + y_1\,\mathrm{d}f$$
$$N^*\,\mathrm{d}y_2 = a_2\,\mathrm{d}y_2 - (a_2 - a_3)x_2\,\mathrm{d}g + y_2\,\mathrm{d}f$$

where the shorthand notations

$$f = \tfrac{1}{2}(a_1 - a_3)x_1^2 + \tfrac{1}{2}(a_2 - a_3)x_2^2$$
$$g = x_1y_1 + x_2y_2$$

have been used. The derivation $\mathrm{d}_N$, associated with $N$, is accordingly defined by

$$\mathrm{d}_Nx_1 = N^*\,\mathrm{d}x_1 = \left[a_1 + (a_3 - a_1)x_1^2\right]\mathrm{d}x_1 + (a_3 - a_2)x_1x_2\,\mathrm{d}x_2$$
$$\mathrm{d}_Nx_2 = N^*\,\mathrm{d}x_2 = (a_3 - a_1)x_1x_2\,\mathrm{d}x_1 + \left[a_2 + (a_3 - a_2)x_2^2\right]\mathrm{d}x_2$$
$$\mathrm{d}_Ny_1 = N^*\,\mathrm{d}y_1 = \left[(a_3 - a_1)x_1y_2 - (a_3 - a_2)x_2y_1\right]\mathrm{d}x_2 + \left[a_1 + (a_3 - a_1)x_1^2\right]\mathrm{d}y_1 + (a_3 - a_1)x_1x_2\,\mathrm{d}y_2$$
$$\mathrm{d}_Ny_2 = N^*\,\mathrm{d}y_2 = \left[(a_3 - a_2)x_2y_1 - (a_3 - a_1)x_1y_2\right]\mathrm{d}x_1 + (a_3 - a_2)x_1x_2\,\mathrm{d}y_1 + \left[a_2 + (a_3 - a_2)x_2^2\right]\mathrm{d}y_2$$

on the coordinate functions. Recalling that $\mathrm{d}_N$ anticommutes with d, one can then easily check the condition

$$\mathrm{d}\mathrm{d}_Nh = \mathrm{d}s_1 \wedge \mathrm{d}h$$

where $s_1$ is the first coefficient of the polynomial defining the elliptical spherical coordinates, and $h$ is the Hamiltonian of the Neumann system. By the Frobenius theorem, this equation alone entails the integrability of the distribution $D_h$, without the need of computing $X_h, NX_h$, and their commutator $[X_h, NX_h]$. Thus, it can be concluded that the Neumann system too is a cyclic system of Levi-Civita, and that the recursion operator $N$, generating the distribution $D_h$, is closely related to the polynomial defining the separation coordinates of the Neumann system.

## Separable System of Jacobi

In 1838, Jacobi noticed that the Hamilton–Jacobi equation

$$h\left(x_1, x_2, \ldots, x_n, \frac{\partial W}{\partial x_1}, \ldots, \frac{\partial W}{\partial x_n}\right) = e$$

of many Hamiltonian systems splits owing to an appropriate choice of coordinates in a set of ordinary differential equations. On account of this property, these systems have been called separable. In 1904, Levi-Civita gave a first partial characterization of separable Hamiltonians by means of his separability conditions. In a letter addressed to Stäckel, he proved that $h$ is separable in a preassigned system of canonical coordinates if and only if the conditions

$$\frac{\partial^2 h}{\partial x_j \partial x_k}\frac{\partial h}{\partial y_j}\frac{\partial h}{\partial y_k} - \frac{\partial^2 h}{\partial x_j \partial y_k}\frac{\partial h}{\partial y_j}\frac{\partial h}{\partial x_k} - \frac{\partial^2 h}{\partial y_j \partial x_k}\frac{\partial h}{\partial x_j}\frac{\partial h}{\partial y_k} + \frac{\partial^2 h}{\partial y_j \partial y_k}\frac{\partial h}{\partial x_j}\frac{\partial h}{\partial x_k} = 0$$

are satisfied by $h$. One must notice the nontensorial character of these conditions; they hold only in a specific coordinate system, and if the coordinates are changed, it is not possible to reconstruct the form of the separability conditions in the new coordinates. The nontensorial character is the major drawback of the separability conditions of Levi-Civita, making them practically useless in the search of separation coordinates.

The contact between the theory of separable system of Jacobi and the theory of cyclic systems of Levi-Civita rests on two occurrences. The first is the form of the integrability conditions of the distribution $D_h$ generated by any vector field $X_h$ on an $\omega N$ manifold. Exploiting the Frobenius integrability conditions and the properties of the differential operator $d_N$ associated with the recursion operator $N$, it can be proved that $D_h$ is integrable if and only if the 2-form $dd_N h$ vanishes on $D_h$:

$$dd_N h = 0 \quad \text{on} \quad D_h$$

Suppose now that the dimension of $D_h$ is maximal, that is, equal to $n = (1/2)\dim S$. Then $D_h$ is spanned by the $n$ vector fields $(X_h, NX_h, \ldots, N^{n-1}X_h)$, and the vanishing condition of $dd_N h$ on $D_h$ turns out to be equivalent to

$$dd_N h(N^j X_h, N^k X_h) = 0$$

for any value of $j$ and $k$ from 0 to $n-1$. Thus, the number of separability conditions of $h$ and the number of integrability conditions of $D_h$ are equal. This circumstance strongly suggests that the two sets of conditions are related. The nontensorial character of the Levi-Civita conditions, compared with the tensorial character of the integrability conditions of $D_h$, further suggests that the former should be the evaluation of the latter in a specific system of coordinates. These coordinates are the "normal coordinates" of an $\omega N$ manifold, that will be introduced in the following.

Assume that the minimal polynomial of $N$ has real and distinct roots $(l_1, \ldots, l_n)$. In this case, the $\omega N$ manifold is said to be semisimple. A two-dimensional eigenspace is associated with each root $l_k$. Let us consider the distribution $E_k$ spanned by all the eigenvectors of $N$, except those associated with $l_k$. Since $N$ is torsion free, each distribution $E_k$ is integrable. Let us fix the attention on one of these distributions. It turns out that its leaves are symplectic submanifolds of codimension 2. So they are the level surfaces of a pair of (local) functions which are not in involution. By collecting together the pairs of functions associated with the $n$ distributions $(E_1, \ldots, E_n)$, one obtains, at the end, a coordinate system $(\lambda_1, \mu_1, \lambda_2, \mu_2, \ldots, \lambda_n, \mu_n)$ on $S$. Moreover, these functions can be chosen in such a way to form a system of canonical coordinates. The final result is that, on a semisimple $\omega N$ manifold, one can construct a coordinate system such that

$$\omega = \sum_{j=1}^{n} d\mu_j \wedge d\lambda_j$$

and

$$N^*(d\lambda_j) = l_j d\lambda_j$$
$$N^*(d\mu_j) = l_j d\mu_j$$

These coordinates are called the normal coordinates (or sometimes, the Darboux–Nijenhuis coordinates) of the $\omega N$ manifold. One can prove that the separability conditions of Levi-Civita are the integrability

conditions of $D_h$, written in normal coordinates. This result allows us to claim that the cyclic systems of Levi-Civita on semisimple $\omega N$ manifolds are all separable.

The reverse is also true. As has already been shown in the example of the Neumann system, a given separable system of Jacobi can be associated with a recursion operator $N$ in such a way that its phase space (with the possible exclusion of a singular locus) becomes an $\omega N$ manifold, and the Hamiltonian vector field $X_h$ becomes a cyclic system of Levi-Civita. A new interpretation of the process of separation of variables follows from this result. Indeed, to find separation coordinates for a given system on a symplectic manifold $S$ is equivalent to deforming the Poisson bracket of $S$ into a pencil

$$\{f,g\}_\epsilon = \{f,g\} + \epsilon\{f,g\}'$$

in such a way that the recursion operator $N$ defining the pencil $\{f,g\}_\epsilon$ generates, with $X_h$, an integrable distribution $D_h$. Therefore, classical mechanics is deeply entangled with the theory of recursion operators, even if the insistence on the use of separation coordinates has hidden this factor for a long time.

*See also:* Bi-Hamiltonian Methods in Soliton Theory; Classical *r*-Matrices, Lie Bialgebras, and Poisson Lie Groups; Integrable Systems and Algebraic Geometry; Integrable Systems and Recursion Operators on Symplectic and Jacobi Manifolds; Integrable Systems: Overview; Multi-Hamiltonian Systems; Separation of Variables for Differential Equations; Solitons and Kac–Moody Lie Algebras.

## Further Reading

Dubrovin BA, Krichever IM, and Novikov SP (2001) Integrable systems I. In: Arnol'd VI (ed.) *Encyclopaedia of Mathematical Sciences. Dynamical Systems IV*, pp. 177–332. Berlin: Springer.
Jacobi CGJ (1996) *Vorlesungen ber analytische Mechanik*, Deutsche Mathematiker Vereinigung, Freiburg. Braunschweig: Friedrich Vieweg and Sohn.
Ivan K, Michor PW, and Slovák J (1993) *Natural Operations in Differential Geometry*. Berlin: Springer.
Kalnins EG (1986) *Separation of Variables for Riemannian Spaces of Constant Curvature*. New York: Wiley.
Krasilshchik IS and Kersten PHM (2000) *Symmetries and Recursion Operators for Classical and Supersymmetric Differential Equations*. Dordrecht: Kluwer.
Magri F, Falqui G, and Pedroni M (2003) The method of Poisson pairs in the theory of nonlinear PDEs. In: Conte R, Magri F, Musette M, Satsuma J, and Winternitz P (eds.) *Direct and Inverse Methods in Nonlinear Evolution Equations*, Lecture Notes in Physics, vol. 632, pp. 85–136. Berlin: Springer.
Miller W (1977) *Symmetry and Separation of Variables*. Reading, MA–London–Amsterdam: Addison-Wesley.
Olver PJ (1993) *Applications of Lie Groups to Differential Equations*, 2nd edn. New York: Springer.
Pars LA (1965) *A Treatise on Analytical Dynamics*. London: Heinemann.
Vaisman I (1994) *Lectures on the Geometry of Poisson Manifolds*. Basel: Birkhäuser.
Vilasi G (2001) *Hamiltonian Dynamics*. River Edge, NJ: World Scientific.
Yano K and Ishihara S (1973) *Tangent and Cotangent Bundles: Differential Geometry*. New York: Dekker.

# Reflection Positivity and Phase Transitions

**Y Kondratiev**, Universität Bielefeld, Bielefeld, Germany
**Y Kozitsky**, Uniwersytet Marii Curie-Sklodowskiej, Lublin, Poland



## Phase Transitions in Lattice Systems

### Introduction

Phase transitions are among the main objects of equilibrium statistical mechanics, both classical and quantum. There exist several approaches to the description of these phenomena. Their common point is that the macroscopic behavior of a statistical mechanical model can be different at the same values of the model parameters. This corresponds to the multiplicity of equilibrium phases, each of which has its own properties. In the mathematical formulation, models are defined by interaction potentials and equilibrium phases appear as states – positive linear functionals on algebras of observables. In the classical case the states are defined by means of the probability measures which satisfy equilibrium conditions, formulated in terms of the interaction potentials. Such measures are called Gibbs measures and the corresponding states are called Gibbs states. The observables are then integrable functions. In the quantum case the states mostly are introduced by means of the Kubo–Martin–Schwinger condition – a quantum analog of the equilibrium conditions used for classical models. The quantum observables constitute noncommutative von Neumann algebras.

Infinite systems of particles studied in statistical mechanics fall into two main groups. These are continuous systems and lattice systems. In the latter case, particles are attached to the points of various crystalline lattices. In view of the specifics of our subject, in this article we will deal with lattice systems only.

One of the main problems of the mathematical theory of phase transitions is to prove that the Gibbs states of a given model can be multiple, that is, that this model undergoes a phase transition. To solve this problem one has to elaborate corresponding mathematical tools. Typically, at high temperatures (equivalently, for weak interactions), a model, which undergoes a phase transition, has only one Gibbs state. This state inherits all the symmetries possessed by the interaction potentials. At low temperatures this model has multiple Gibbs states, which may lose the symmetries. In this case the phase transition is accompanied by a symmetry breaking. Among the symmetries important in the theory of lattice systems, there is the invariance with respect to the lattice translations. If the Gibbs state of a translation invariant lattice model is unique, it ought to be ergodic with respect to the group of lattice translations. This means in particular that the spacial correlations in this state decay to zero at long distances. Therefore, the lack of the latter property may indicate a phase transition. In a number of lattice models, phase transitions can be established by means of their special property – reflection positivity. The most important consequence of reflection positivity are chessboard (another name checkerboard) estimates, being extended versions of Hölder's inequalities. The proof of a phase transition is then performed either by means of a combination of such estimates and contour methods, or by means of infrared estimates obtained from the chessboard estimates.

In this article we show how to prove phase transitions by means of the infrared estimates for some simple reflection positive models, both classical and quantum. The details on the reflection positivity method in all its versions may be found in the literature listed at the end of the article. There we also provide short bibliographic comments.

#### Nonergodicity and Infrared Estimates

The following heuristic arguments should give an idea how to establish the nonergodicity of a Gibbs state by means of infrared estimates. Let us consider a classical ferromagnetic translation-invariant model. (Of course, we assume that it possesses Gibbs states, which for models with unbounded spins is a nontrivial property. A particular case of this model is described in more detail in the subsection "Gaussian domination.") This model describes the system of interacting $N$-dimensional spins $x_\ell \in \mathbb{R}^N$, indexed by the elements $\ell \in \mathbb{Z}^d$ of the $d$-dimensional simple cubic lattice. The interaction is pairwise, attractive, nearest-neighbor, and invariant with respect to the rotations in $\mathbb{R}^N$. Consider a translation-invariant Gibbs state of this model, which always exists. Let $K(\ell, \ell'), \ell, \ell' \in \mathbb{Z}^d$, be the expectation of the scalar product $(x_\ell, x_{\ell'})$ of spins in this state. Then $K(\ell, \ell')$ is also translation invariant and hence may be written as

$$K(\ell, \ell') = \frac{1}{(2\pi)^d} \int_{(-\pi,\pi]^d} \widehat{K}(p) e^{i(p, \ell - \ell')} dp, \quad i = \sqrt{-1} \qquad [1]$$

where the generalized function $\widehat{K}$ is defined by the Fourier series

$$\widehat{K}(p) = \sum_{\ell' \in \mathbb{Z}^d} K(\ell, \ell') e^{-i(p, \ell - \ell')}, \quad p \in (-\pi, \pi]^d \qquad [2]$$

As the model is ferromagnetic, $K(\ell, \ell') \geq 0$. The Gibbs state is nonergodic if $K(\ell, \ell')$ does not tend to zero as $|\ell - \ell'| \to \infty$. In this case $\widehat{K}$ should be singular at $p = 0$. Set

$$\widehat{K}(p) = (2\pi)^d \lambda \delta(p) + g(p) \qquad [3]$$

where $\delta(p)$ is the Dirac $\delta$-function and $g(p)$ is regular at $p = 0$. Then the Gibbs state is nonergodic if $\lambda \neq 0$. Suppose we know that $g(p) \geq 0$ and that the following two estimates hold. The first one is

$$g(p) \leq \gamma / J |p|^2, \quad p \neq 0 \qquad [4]$$

where $\gamma > 0$ is a constant and $J > 0$ is the interaction intensity multiplied by the inverse temperature $\beta$. This is the infrared estimate. The second estimate is

$$K(\ell, \ell) \geq \varkappa > 0 \qquad [5]$$

where $\varkappa$ is independent of $J$. By these estimates and [1], [2], we get

$$\lambda \geq \varkappa - \frac{\gamma}{(2\pi)^d J} \int_{(-\pi,\pi]^d} \frac{dp}{|p|^2} \qquad [6]$$

For $d \geq 3$, the latter integral exists; hence, $\lambda > 0$ for $J$ large enough, which means that the state we consider is nonergodic.

The quantum case is more involved. The infrared bounds are obtained not for functions like $\widehat{K}(p)$ but for the so-called Duhamel two-point functions. Then one has to prove a number of additional statements, which finally lead to the proof of the result desired. In the section on reflection positivity in quantum systems we indicate how to do this for a simple quantum spin model.

### Reflection Positivity and Phase Transitions in Classical Systems

We begin by studying reflection positive (RP) functionals. Gibbs states of RP models are such functionals.

### Reflection Positive Functionals

Let $\Lambda$ be a finite set of indices consisting of an even number $|\Lambda|$ of elements, which label real variables $x_\ell, \ell \in \Lambda$. For $\Lambda' \subseteq \Lambda$, we write $x_{\Lambda'} = (x_\ell)_{\ell \in \Lambda'} \in \mathbb{R}^{|\Lambda'|}$. Suppose we are given a bijection $\rho : \Lambda \to \Lambda, \rho \circ \rho = \mathrm{id}$, such that the set $\Lambda$ falls into two disjoint parts $\Lambda_\pm$ with the property $\rho : \Lambda_+ \to \Lambda_-$. Therefore, $|\Lambda_+| = |\Lambda_-|$, and the map $\rho$ may be regarded as a reflection. For $x_\Lambda \in \mathbb{R}^{|\Lambda|}$, we set $\rho(x_\Lambda) = (x_{\rho(\ell)})_{\ell \in \Lambda}$. Now let $\mathcal{A}$ be an algebra of functions $A : \mathbb{R}^{|\Lambda|} \to \mathbb{R}$. Then we define the map $\vartheta : \mathcal{A} \to \mathcal{A}$ by setting

$$\vartheta(A)(x_\Lambda) = A(\rho(x_\Lambda)) \qquad [7]$$

Clearly, for all $A, B \in \mathcal{A}$ and $\xi, \eta \in \mathbb{R}$,

$$\begin{aligned} \vartheta(\xi A + \eta B) &= \xi\vartheta(A) + \eta\vartheta(B) \\ \vartheta(A \cdot B) &= \vartheta(A) \cdot \vartheta(B) \end{aligned} \qquad [8]$$

By $\mathcal{A}^+$ (respectively, $\mathcal{A}^-$), we denote the subalgebra of $\mathcal{A}$ consisting of functions dependent on $x_{\Lambda_+}$ (respectively, $x_{\Lambda_-}$). Then $\vartheta(\mathcal{A}^+) = \mathcal{A}^-$ and $\vartheta \circ \vartheta = \mathrm{id}$.

**Definition 1** A linear functional $\phi : \mathcal{A} \to \mathbb{R}$ is called RP with respect to the maps $\rho$ and $\vartheta$, if

$$\forall A \in \mathcal{A}^+ : \ \phi[A\vartheta(A)] \geq 0 \qquad [9]$$

**Example 2** Let $\chi$ be a Borel measure on the real line (not necessarily positive), with respect to which all real polynomials are integrable. Let also $\mathcal{A}$ be the algebra of all real-valued polynomials on $\mathbb{R}^{|\Lambda|}$, $|\Lambda|$ being even. Finally, let $\rho$ and $\vartheta$ be any of the maps with the properties described above. Then the functional

$$\begin{aligned} \phi(A) &= \int_{\mathbb{R}^{|\Lambda|}} A(x_\Lambda)\, \mathrm{d}\chi_\Lambda(x_\Lambda) \\ \mathrm{d}\chi_\Lambda(x_\Lambda) &= \prod_{\ell \in \Lambda} \mathrm{d}\chi(x_\ell) \end{aligned} \qquad [10]$$

is RP. Indeed, let $F : \mathbb{R}^{|\Lambda|/2} \to \mathbb{R}$ be such that $A(x_\Lambda) = F(x_{\Lambda_+})$. Then

$$\begin{aligned} \phi[A\vartheta(A)] &= \int F(x_{\Lambda_+}) \prod_{\ell \in \Lambda_+} \mathrm{d}\chi(x_\ell) \cdot \int F(x_{\Lambda_-}) \prod_{\ell \in \Lambda_-} \mathrm{d}\chi(x_\ell) \\ &= \left[ \int F(x_{\Lambda_+}) \prod_{\ell \in \Lambda_+} \mathrm{d}\chi(x_\ell) \right]^2 \geq 0 \end{aligned}$$

In the above example the multiplicative structure of the measure $\chi_\Lambda$ is crucial. It results in the positivity of $\phi$ with respect to all reflections. If one has just one such reflection, the measure which defines $\phi$ may be decomposable onto two measures only. Let $\Lambda, \mathcal{A}, \rho$, and $\vartheta$ be as above. Consider a Borel measure $\nu$ on $\mathbb{R}^{|\Lambda|/2}$ such that every real-valued polynomial on $\mathbb{R}^{|\Lambda|/2}$ is $\nu$-integrable.

**Proposition 3** *The functional*

$$\phi(A) = \int_{\mathbb{R}^{|\Lambda|}} A(x_\Lambda)\, \mathrm{d}\nu(x_{\Lambda_+})\, \mathrm{d}\nu(x_{\Lambda_-}) \qquad [11]$$

*is RP.*

In both these examples the states are symmetric, that is,

$$\phi[A\vartheta(B)] = \phi[B\vartheta(A)], \quad \text{for all } A, B \in \mathcal{A}^+ \qquad [12]$$

In the sequel we shall suppose that all RP functionals possess this property. Therefore, RP functionals obey a Cauchy–Schwarz type inequality.

**Lemma 4** *If $\phi$ is RP, then for any $A, B \in \mathcal{A}^+$,*

$$\{\phi[A\vartheta(B)]\}^2 \leq \phi[A\vartheta(A)] \cdot \phi[B\vartheta(B)] \qquad [13]$$

**Proof** For $\xi \in \mathbb{R}$, by [8] we have

$$\begin{aligned} &\phi[(A + \xi B)\vartheta(A + \xi B)] \\ &\quad = \phi[(A + \xi B)(\vartheta(A) + \xi\vartheta(B))] \geq 0 \end{aligned}$$

Since $\phi$ is linear, the latter can be written as a 3-nomial, whose positivity for all $\xi \in \mathbb{R}$ is equivalent to [13]. □

Now let an RP functional $\phi$ be such that for

$$A, B, C_1, \ldots, C_m, D_1, \ldots, D_m \in \mathcal{A}^+$$

there exists

$$\phi\left[\exp\left(A + \vartheta(B) + \sum_{i=1}^m C_i\vartheta(D_i)\right)\right]$$

and that the series

$$\begin{aligned} &\sum_{n_1, \ldots, n_m = 0}^{\infty} \frac{1}{n_1! \cdots n_m!} \cdot \phi\{[C_1\vartheta(C_1)]^{n_1} \cdots [C_m\vartheta(C_m)]^{n_m} \\ &\quad \times \exp[A + \vartheta(B)]\} \end{aligned} \qquad [14]$$

as well as the one with all $C_i$s replaced by $D_i$s converge absolutely.

**Lemma 5** *Let the functional $\phi$ and the functions $A, B, C_i, D_i, i = 1, \ldots, m$, be as above. Then*

$$\begin{aligned} &\left\{\phi\left[\exp\left(A + \vartheta(B) + \sum_{i=1}^m C_i\vartheta(D_i)\right)\right]\right\}^2 \\ &\quad \leq \phi\left[\exp\left(A + \vartheta(A) + \sum_{i=1}^m C_i\vartheta(C_i)\right)\right] \\ &\quad \times \phi\left[\exp\left(B + \vartheta(B) + \sum_{i=1}^m D_i\vartheta(D_i)\right)\right] \end{aligned} \qquad [15]$$

**Proof** By the above assumptions

$$\phi\left[\exp\left(A+\vartheta(B)+\sum_{i=1}^{m}C_i\vartheta(D_i)\right)\right]$$
$$=\phi\left[F\vartheta(G)\exp\left(\sum_{i=1}^{m}C_i\vartheta(D_i)\right)\right]$$
$$=\sum_{n_1,\dots,n_m=0}^{\infty}\frac{1}{n_1!\cdots n_m!}\cdot\phi[F\vartheta(G)[C_1\vartheta(D_1)]^{n_1}\cdots\times[C_m\vartheta(D_m)]^{n_m}] \qquad [16]$$

where $F=\mathrm{e}^A, G=\mathrm{e}^B$. Then by [13] and the Cauchy–Schwarz inequality for sums we get

$$\begin{aligned}\text{RHS}[16] &\le \sum_{n_1,\dots,n_m=0}^{\infty}\left\{\frac{1}{n_1!\cdots n_m!}\cdot\phi[F\vartheta(F)[C_1\vartheta(C_1)]^{n_1}\cdots[C_m\vartheta(C_m)]^{n_m}]\right\}^{1/2}\\ &\quad\times\left\{\frac{1}{n_1!\cdots n_m!}\cdot\phi[G\vartheta(G)[D_1\vartheta(D_1)]^{n_1}\cdots[D_m\vartheta(D_m)]^{n_m}]\right\}^{1/2}\\ &\le\left\{\sum_{n_1,\dots,n_m=0}^{\infty}\frac{1}{n_1!\cdots n_m!}\cdot\phi[F\vartheta(F)[C_1\vartheta(C_1)]^{n_1}\cdots[C_m\vartheta(C_m)]^{n_m}]\right\}^{1/2}\\ &\quad\times\left\{\sum_{n_1,\dots,n_m=0}^{\infty}\frac{1}{n_1!\cdots n_m!}\cdot\phi[G\vartheta(G)[D_1\vartheta(D_1)]^{n_1}\cdots[D_m\vartheta(D_m)]^{n_m}]\right\}^{1/2}\\ &=\left\{\phi\left[\exp\left(A+\vartheta(A)+\sum_{i=1}^{m}C_i\vartheta(C_i)\right)\right]\right\}^{1/2}\\ &\quad\times\left\{\phi\left[\exp\left(B+\vartheta(B)+\sum_{i=1}^{m}D_i\vartheta(D_i)\right)\right]\right\}^{1/2}\end{aligned}$$

which yields [15]. □

### Main Estimate

Let $\Delta$ be a finite set and $\Delta'$ be its nonempty subset. Let also $\mu$ and $\nu$ be finite Borel measures on $\mathbb{R}^{N|\Delta|}, N\in\mathbb{N}$. For vectors $b,c\in\mathbb{R}^N$, by $(b,c)$ and $|b|,|c|$ we denote their scalar product $\sum_{k=1}^{N}b^{(k)}c^{(k)}$ and the corresponding norms, respectively. By $x_\Delta$ we denote $(x_\ell)_{\ell\in\Delta}, x_\ell\in\mathbb{R}^N$; hence, $x_\Delta\in\mathbb{R}^{N|\Delta|}$.

**Lemma 6** *Let the sets $\Delta,\Delta'$ and the measures $\mu,\nu$ be as above. Then for every $(a_\ell)_{\ell\in\Delta'}\in\mathbb{R}^{N|\Delta'|}$ and $J\ge 0$,*

$$\left[\int_{\mathbb{R}^{2N|\Delta|}}\exp\left(-\frac{J}{2}\sum_{\ell\in\Delta'}|x_\ell-y_\ell-a_\ell|^2\right)\mathrm{d}\mu(x_\Delta)\mathrm{d}\nu(y_\Delta)\right]^2$$
$$\le\int_{\mathbb{R}^{2N|\Delta|}}\exp\left(-\frac{J}{2}\sum_{\ell\in\Delta'}|x_\ell-y_\ell|^2\right)\mathrm{d}\mu(x_\Delta)\,\mathrm{d}\mu(y_\Delta)$$
$$\times\int_{\mathbb{R}^{2N|\Delta|}}\exp\left(-\frac{J}{2}\sum_{\ell\in\Delta'}|x_\ell-y_\ell|^2\right)\mathrm{d}\nu(x_\Delta)\,\mathrm{d}\nu(y_\Delta) \qquad [17]$$

**Proof** Take two copies of $\Delta$ and denote them by $\Lambda_\pm$. Furthermore, by $\Lambda'_\pm\subset\Lambda_\pm$ we denote the subsets consisting of the elements of $\Delta'\subset\Delta$. For an $\ell\in\Lambda_+$, by $\rho(\ell)$ we denote its counterpart in $\Lambda_-$. Then $\rho$ is a reflection and $\rho(\Lambda'_+)=\Lambda'_-$. Let $\Lambda=\Lambda_+\cup\Lambda_-, \Lambda'=\Lambda'_+\cup\Lambda'_-$, and $\mathcal{A}$ be the algebra of all polynomials of $(x_{\Lambda'},y_{\Lambda'})\in\mathbb{R}^{2N|\Lambda'|}$. Note that $x_{\Lambda'}$ may be regarded as the pair $(x_{\Lambda'_+},x_{\Lambda'_-})$. Let $\mathcal{A}^+$ (respectively, $\mathcal{A}^-$) be the subalgebra of $\mathcal{A}$ consisting of the polynomials which depend on $x_{\Lambda'_+},y_{\Lambda'_+}$ (respectively, $x_{\Lambda'_-},y_{\Lambda'_-}$) only. Introduce the measures

$$\mathrm{d}\tilde{\mu}(x_\Delta)=\exp\left(-\frac{J}{2}\sum_{\ell\in\Delta'}|x_\ell|^2\right)\mathrm{d}\mu(x_\Delta)$$
$$\mathrm{d}\tilde{\nu}(x_\Delta)=\exp\left(-\frac{J}{2}\sum_{\ell\in\Delta'}|x|_\ell^2\right)\mathrm{d}\nu(x_\Delta)$$

and define the following functional on $\mathcal{A}$:

$$\phi(F)=\int_{\mathbb{R}^{2N|\Lambda|}}F(x_{\Lambda'},y_{\Lambda'})\,\mathrm{d}\tilde{\mu}(x_{\Lambda_+})\times\mathrm{d}\tilde{\nu}(y_{\Lambda_+})\,\mathrm{d}\tilde{\mu}(x_{\Lambda_-})\,\mathrm{d}\tilde{\nu}(y_{\Lambda_-}) \qquad [18]$$

It has the same structure as the one described by Proposition 3, hence is RP with respect to the map $\vartheta$ defined by the reflection $\rho$. Set

$$\Upsilon_\mu=\int_{\mathbb{R}^{N|\Delta|}}\mathrm{d}\tilde{\mu}(x_\Delta),\qquad\Upsilon_\nu=\int_{\mathbb{R}^{N|\Delta|}}\mathrm{d}\tilde{\nu}(y_\Delta) \qquad [19]$$

and

$$\begin{aligned}&A\equiv 0,\qquad B=-J\sum_{\ell\in\Lambda'_+}\left[\frac{1}{2}|a_\ell|^2+(a_\ell,y_\ell)\right]\\ &C_\ell^{(k)}=\sqrt{J}x_\ell^{(k)},\qquad D_\ell^{(k)}=\sqrt{J}\left(y_\ell^{(k)}+a_\ell^{(k)}\right)\\ &\ell\in\Lambda'_+,\quad k=1,\dots,N\end{aligned} \qquad [20]$$

Then the left-hand side of [17] is

$$\text{LHS }[17]=\frac{1}{(\Upsilon_\mu\Upsilon_\nu)^2}\left|\phi\left[\exp\left(A+\vartheta(B)+\sum_{\ell\in\Lambda'_+}\sum_{k=1}^{N}C_\ell^{(k)}\vartheta\left(D_\ell^{(k)}\right)\right)\right]\right|^2 \qquad [21]$$

with $\phi$ given by [18]. Applying [15] and taking into account [19], we arrive at

$$\begin{aligned}\text{LHS }[17]&\le\frac{1}{(\Upsilon_\mu\Upsilon_\nu)^2}\int_{\mathbb{R}^{2N|\Lambda|}}\exp\left(J\sum_{\ell\in\Lambda'_+}x_\ell x_{\rho(\ell)}\right)\\ &\quad\times\mathrm{d}\tilde{\mu}(x_{\Lambda_+})\,\mathrm{d}\tilde{\mu}(x_{\Lambda_-})\,\mathrm{d}\tilde{\nu}(y_{\Lambda_+})\,\mathrm{d}\tilde{\nu}(y_{\Lambda_-})\\ &\quad\times\int_{\mathbb{R}^{2N|\Lambda|}}\exp\left(J\sum_{\ell\in\Lambda'_+}y_\ell y_{\rho(\ell)}\right)\\ &\quad\times\mathrm{d}\tilde{\mu}(x_{\Lambda_+})\,\mathrm{d}\tilde{\mu}(x_{\Lambda_-})\,\mathrm{d}\tilde{\nu}(y_{\Lambda_+})\,\mathrm{d}\tilde{\nu}(y_{\Lambda_-})=\text{RHS }[17]\end{aligned}$$

which completes the proof. □

### Gaussian Domination

Let $\Lambda$ be a finite set, $|\Lambda|$ even, and $E$ be a set of unordered pairs of elements of $\Lambda$, such that the graph $(\Lambda, E)$ is connected. If $e \in E$ connects given $\ell, \ell' \in \Lambda$, we write $e = \langle \ell, \ell' \rangle$. We suppose that $E$ contains no loops $\langle \ell, \ell \rangle$. With each $\ell \in \Lambda$ we associate a random $N$-component vector $x_\ell$, called spin. The joint probability distribution of the spins $(x_\ell)_{\ell \in \Lambda}$ is defined by means of the local Gibbs measure

$$d\mu_\Lambda(x_\Lambda) = \frac{1}{Z_\Lambda} \exp\left( -\frac{J}{2} \sum_{\langle \ell, \ell' \rangle \in E} |x_\ell - x_{\ell'}|^2 \right) d\chi_\Lambda(x_\Lambda), \quad x_\Lambda \in \mathbb{R}^{N|\Lambda|} \qquad [22]$$

Here the measure

$$d\chi_\Lambda(x_\Lambda) = \prod_{\ell \in \Lambda} d\chi(x_\ell) \qquad [23]$$

describes the system if the interaction intensity $J$ equals zero. In general, $J \geq 0$, that is, the model [22], [23] is ferromagnetic. The single-spin measure $\chi$ is a probability measure on $\mathbb{R}^N$ and

$$Z_\Lambda = \int_{\mathbb{R}^{N|\Lambda|}} \exp\left( -\frac{J}{2} \sum_{\langle \ell, \ell' \rangle \in E} |x_\ell - x_{\ell'}|^2 \right) d\chi_\Lambda(x_\Lambda) \qquad [24]$$

is the partition function. Set

$$Z_\Lambda(h) = \int_{\mathbb{R}^{N|\Lambda|}} \exp\left( -\frac{J}{2} \sum_{\langle \ell, \ell' \rangle \in E} |x_\ell - x_{\ell'} - h_{\ell\ell'}|^2 \right) \times d\chi_\Lambda(x_\Lambda) \qquad [25]$$

where $h_{\ell\ell'} = h_{\ell'\ell} \in \mathbb{R}^N$, $\langle \ell, \ell' \rangle \in E$.

**Definition 7** The model [22]–[23] admits Gaussian domination if for all $h = (h_{\ell\ell'})_{\langle \ell, \ell' \rangle \in E}$,

$$Z_\Lambda(h) \leq Z_\Lambda(0) \qquad [26]$$

We prove that our model admits Gaussian domination if the graph satisfies the following:

**Assumption 8** The set of edges $E$ can be decomposed

$$E = \bigcup_{n=1}^{m} E_n, \qquad E_n \bigcap E_{n'} = \emptyset, \quad \text{if } n \neq n' \qquad [27]$$

in such a way that for every $n = 1, \ldots, m$, the graph $(\Lambda, E \backslash E_n)$ is disconnected and falls into two connected components, $(\Lambda_+^{(n)}, E_+^{(n)})$ and $(\Lambda_-^{(n)}, E_-^{(n)})$, which are isomorphic. This means that there exists a bijection $\rho_n : \Lambda \to \Lambda, \rho_n \circ \rho_n = \text{id}$, such that $\rho_n(\Lambda_+^{(n)}) = \Lambda_-^{(n)}$ and $\langle \rho_n(\ell), \rho_n(\ell') \rangle \in E_-^{(n)}$ whenever $\langle \ell, \ell' \rangle \in E_+^{(n)}$. Finally, we assume that if $\langle \ell, \ell' \rangle \in E_n$ and $\ell \in \Lambda_+^{(n)}$, then $\rho_n(\ell) = \ell'$.

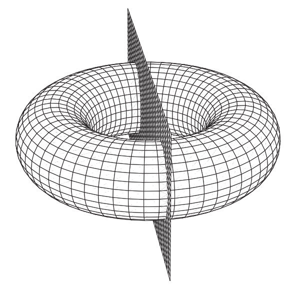


**Figure 1** The torus.

By this assumption if $\langle \ell, \ell' \rangle \in E_n$, then no other elements of $E_n$ can be of the form $\langle \ell, \ell'' \rangle$ or $\langle \ell'', \ell' \rangle$. The basic example here is the torus which one obtains from a rectangular box $\Lambda \subset \mathbb{Z}^d$, $|\Lambda|$ even, by imposing periodic conditions on its boundaries. The set of edges is $E = \{\langle \ell, \ell' \rangle \,|\, |\ell - \ell'|_\Lambda = 1\}$, where $|\ell - \ell'|_\Lambda$ is the periodic distance on $\Lambda$ (see the next subsection). Then every plane which contains the center of the torus and its axis cuts it out along a family of edges onto two subgraphs with the property desired (see Figure 1).

**Theorem 9** *The model* [22]–[23] *defined on the graph obeying* Assumption 8 *admits Gaussian domination.*

**Proof** For $\sigma = \pm 1, h = (h_{\ell\ell'})_{\langle \ell, \ell' \rangle \in E}$, and $n = 1, \ldots, m$, we define the map

$$(T_n^\sigma h)_{\ell\ell'} = \begin{cases} h_{\ell\ell'}, & \text{if } \langle \ell, \ell' \rangle \in E_\sigma^{(n)} \\ h_{\rho_n(\ell)\rho_n(\ell')}, & \text{if } \langle \ell, \ell' \rangle \in E_{-\sigma}^{(n)} \\ 0, & \text{if } \langle \ell, \ell' \rangle \in E_n \end{cases} \qquad [28]$$

According to Assumption 8

$$Z_\Lambda(h) = \int_{\mathbb{R}^{N|\Lambda|}} \exp\left( -\frac{J}{2} \sum_{\langle \ell, \ell' \rangle \in E_1} |x_\ell - x_{\ell'} - h_{\ell\ell'}|^2 \right) \times d\nu^+_{\Lambda_+^{(1)}}\left(x_{\Lambda_+^{(1)}}\right) d\nu^-_{\Lambda_-^{(1)}}\left(x_{\Lambda_-^{(1)}}\right) \qquad [29]$$

where

$$d\nu^\sigma_{\Lambda_\sigma^{(1)}}\left(x_{\Lambda_\sigma^{(1)}}\right) = \exp\left( -\frac{J}{2} \sum_{\langle \ell, \ell' \rangle \in E_\sigma^{(1)}} |x_\ell - x_{\ell'} - h_{\ell\ell'}|^2 \right) d\chi_{\Lambda_\sigma^{(1)}}\left(x_{\Lambda_\sigma^{(1)}}\right), \quad \sigma = \pm 1$$

Set

$$\mathrm{d}\nu^-_{\Lambda^{(1)}_+}(x_{\Lambda^{(1)}_+}) = \exp\left(-\frac{J}{2}\sum_{\langle \ell,\ell'\rangle\in E^{(1)}_+} |x_\ell - x_{\ell'} - h_{\rho(\ell)\rho(\ell')}|^2\right) \times \mathrm{d}\chi_{\Lambda^{(1)}_+}(x_{\Lambda^{(1)}_+})$$

$$\mathrm{d}\nu^+_{\Lambda^{(1)}_-}\left(x_{\Lambda^{(1)}_-}\right) = \exp\left(-\frac{J}{2}\sum_{\langle \ell,\ell'\rangle\in E^{(1)}_+} |x_{\rho(\ell)} - x_{\rho(\ell')} - h_{\ell\ell'}|^2\right) \times \mathrm{d}\chi_{\Lambda^{(1)}_-}(x_{\Lambda^{(1)}_-})$$

Then we apply here Lemma 6, with $\Lambda'_+ = \{\ell \in \Lambda^{(1)}_+ | \langle \ell,\ell'\rangle \in E_1\}$, and obtain

$$\begin{aligned}[Z_\Lambda(h)]^2 \le & \int_{\mathbb{R}^{N|\Lambda|}} \exp\left(-\frac{J}{2}\sum_{\langle \ell,\ell'\rangle\in E_1} |x_\ell - x_{\ell'}|^2\right) \\ & \times \mathrm{d}\nu^+_{\Lambda^{(1)}_+}\left(x_{\Lambda^{(1)}_+}\right)\mathrm{d}\nu^+_{\Lambda^{(1)}_-}\left(x_{\Lambda^{(1)}_-}\right) \\ & \times \int_{\mathbb{R}^{N|\Lambda|}} \exp\left(-\frac{J}{2}\sum_{\langle \ell,\ell'\rangle\in E_1} |x_\ell - x_{\ell'}|^2\right) \\ & \times \mathrm{d}\nu^-_{\Lambda^{(1)}_-}\left(x_{\Lambda^{(1)}_-}\right)\mathrm{d}\nu^-_{\Lambda^{(1)}_+}\left(x_{\Lambda^{(1)}_+}\right) \\ = & Z_\Lambda(T_1^+h)Z_\Lambda(T_1^-h)\end{aligned}$$

Next we estimate both $Z_\Lambda(T_1^\pm h)$ employing $E_2$ and $T_2^\sigma$. Repeating this procedure due times we finally get

$$[Z_\Lambda(h)]^{2^m} \le \prod_{\sigma_1,\ldots,\sigma_m=\pm 1} Z_\Lambda(T_m^{\sigma_m}\cdots T_1^{\sigma_1}h) = [Z_\Lambda(0)]^{2^m} \quad [30]$$

Note that $T_m^{\sigma_m}\cdots T_1^{\sigma_1}h = 0$ for any $h\in\mathbb{R}^{N|E|}$ and any sequence $\sigma_1,\ldots,\sigma_m = \pm 1$, which follows from [27] and [28]. □

As might be clear from the proof given above, the local Gibbs state

$$\phi_\Lambda(A) = \int_{\mathbb{R}^{N|\Lambda|}} A(x_\Lambda)\,\mathrm{d}\mu_\Lambda(x_\Lambda) \quad [31]$$

defined by means of the measure [22], is RP with respect to all reflections $\rho_n, n = 1,\ldots,m$. Indeed, the functional defined by the product measure

$$\mathrm{d}\tilde{\chi}_\Lambda(x_\Lambda) \stackrel{\text{def}}{=} \exp\left(-\frac{J}{2}\sum_{\ell\in\Lambda}|x_\ell|^2\right)\mathrm{d}\chi_\Lambda(x_\Lambda) \quad [32]$$

is RP (see Example 2). The Gibbs measure [22] can be written as

$$\mathrm{d}\mu_\Lambda(x_\Lambda) = \frac{1}{Z_\Lambda(0)}\exp\left(\sum_{n=1}^m\sum_{k=1}^N\sum_{\ell\in\Lambda'_{+,n}} C_\ell^{(k)}\vartheta_n\left(C_\ell^{(k)}\right)\right) \times \mathrm{d}\tilde{\chi}_\Lambda(x_\Lambda) \quad [33]$$

where $C_\ell^{(k)}, k = 1,\ldots,N$, are the same as in [20] and $\Lambda'_{+,n} \stackrel{\text{def}}{=} \{\ell\in\Lambda_+^{(n)} | \langle\ell,\ell'\rangle\in E_n\}$. Then the reflection positivity of the Gibbs state [31] can be obtained along the line of arguments used for proving Lemma 6. It appears that this is the only possible way to construct an RP functional from another RP functional.

Repeated application of the estimate [15] also yields

$$\phi_\Lambda\left(\prod_{\ell\in\Lambda}F_\ell(x_\ell)\right) \le \prod_{\ell\in\Lambda}\left[\phi_\Lambda\left(\prod_{\ell'\in\Lambda}F_\ell(x_{\ell'})\right)\right]^{1/|\Lambda|} \quad [34]$$

which holds for any family of functions $\{F_\ell : \mathbb{R}^N \to [0,+\infty)\}_{\ell\in\Lambda}$, for which the above expressions make sense. The estimate [34] is a chessboard estimate, which is a very important element of the theory of phase transitions in RP models. The estimate [26] may be obtained from [34].

### Infrared Bound

Let us show now how to derive the infrared estimates from the Gaussian domination [26]. Consider the system of $N$-dimensional spins indexed by the elements of $\mathbb{Z}^d$ with the nearest-neighbor ferromagnetic interaction and the single-spin measure $\chi$. To construct the periodic local Gibbs measure of this system, we take the box

$$\Lambda = (-L, L]^d \bigcap \mathbb{Z}^d, \quad L\in\mathbb{N} \quad [35]$$

and impose periodic conditions on its boundaries. This defines the periodic distance

$$\begin{aligned}|\ell-\ell'|_\Lambda &= \left[\sum_{j=1}^d |\ell_j - \ell'_j|_L^2\right]^{1/2}, \quad \ell,\ell'\in\Lambda \\ |\ell_j-\ell'_j|_L &= \min\{|\ell_j-\ell'_j|;\ L - |\ell_j - \ell'_j|\}\end{aligned} \quad [36]$$

and hence the set of edges $E$, being unordered pairs $\langle\ell,\ell'\rangle$ such that $|\ell-\ell'|_\Lambda = 1$. Thus, we have the graph $(\Lambda, E)$ and the measure [22]. This is the periodic local Gibbs measure of our model. By [31] it defines the periodic local Gibbs state $\phi_\Lambda$. We have included the inverse temperature $\beta$ into $J$

and assumed that the single-spin measure $\chi$ is rotation invariant. Let us introduce the Fourier transformation

$$\hat{x}(p) = \frac{1}{\sqrt{|\Lambda|}} \sum_{\ell \in \Lambda} x_\ell e^{i(\ell, p)}$$
$$x_\ell = \frac{1}{\sqrt{|\Lambda|}} \sum_{p \in \Lambda_*} \hat{x}(p) e^{-i(\ell, p)} \quad [37]$$

$$\Lambda_* = \left\{ p = (p_1, \ldots, p_d) | \, p_j = -\pi + \frac{\pi}{L}\kappa_j, \right.$$
$$\left. \kappa_j = 1, \ldots, 2L, \; j = 1, \ldots, d \right\} \quad [38]$$

Then we can set

$$\widehat{K}_\Lambda^{(k)}(p) = \phi_\Lambda \left[ \hat{x}^{(k)}(p) \hat{x}^{(k)}(-p) \right]$$
$$\widehat{K}_\Lambda(p) = \sum_{k=1}^{N} \widehat{K}_\Lambda^{(k)}(p) \quad [39]$$

Thereby, cf. [1], [2],

$$K_\Lambda(\ell, \ell') \stackrel{\text{def}}{=} \phi_\Lambda[(x_\ell, x_{\ell'})] = \frac{1}{|\Lambda|} \sum_{p \in \Lambda_*} \widehat{K}_\Lambda(p) e^{i(p, \ell - \ell')} \quad [40]$$

By construction, for any $\ell_0 \in \Lambda$,

$$K_\Lambda(\ell, \ell') = K_\Lambda(\ell + \ell_0, \ell' + \ell_0) \quad [41]$$

where addition is componentwise modulo $2L$. This means that $K_\Lambda(\ell, \ell')$ is invariant with respect to the translations on the corresponding torus. One can show that $K_\Lambda(\ell, \ell')$ converges, as $L \to +\infty$, to $K(\ell, \ell')$ discussed in the Introduction. The corresponding Gibbs state of the whole model is called the periodic Gibbs state. By construction, it is translation invariant. Set

$$\mathcal{E}(p) = \sum_{j=1}^{d} [1 - \cos p_j], \quad p \in (-\pi, \pi]^d \quad [42]$$

**Theorem 10** *For all* $p \in \Lambda_* \setminus \{0\}$,

$$\widehat{K}_\Lambda(p) \le \frac{N}{2J\mathcal{E}(p)} \quad [43]$$

**Proof** Consider the function $f(\xi) = Z_\Lambda(\xi h), \xi \in \mathbb{R}$, where $Z_\Lambda(h)$ is defined by [25]. By Theorem 9 it has a maximum at $\xi = 0$; hence,

$$f''(0) \le 0 \quad [44]$$

Obviously, $f''(0)$ depends on $h = (h_{\ell\ell'})_{\langle \ell, \ell' \rangle \in E}$, $h_{\ell\ell'} \in \mathbb{R}^N$. Let us choose $h$ such that only the first components $h_{\ell\ell'}^{(1)}$ are nonzero. Then [44] holds if

$$J \sum_{\langle \ell_1, \ell_1' \rangle \in E} \sum_{\langle \ell_2, \ell_2' \rangle \in E} \phi_\Lambda \left[ \left( x_{\ell_1}^{(1)} - x_{\ell_1'}^{(1)} \right) \left( x_{\ell_2}^{(1)} - x_{\ell_2'}^{(1)} \right) \right] h_{\ell_1 \ell_1'}^{(1)} h_{\ell_2 \ell_2'}^{(1)}$$
$$\le \sum_{\langle \ell, \ell' \rangle \in E} \left[ h_{\ell\ell'}^{(1)} \right]^2 \quad [45]$$

This means that the eigenvalues of the matrix of the real quadratic form (with respect to $h$) defined by the left-hand side of [45] do not exceed one. The same ought to be true for the extension of this form to the complex case. Let us show that the complex eigenvectors $h_{\ell\ell'}^{(1)}(p)$ of this matrix and the corresponding eigenvalues $\lambda(p)$ are

$$h_{\ell\ell'}^{(1)}(p) = \left( e^{i(p, \ell)} - e^{i(p, \ell')} \right) / \sqrt{|\Lambda|}$$
$$\lambda(p) = 2J\mathcal{E}(p) \widehat{K}_\Lambda^{(1)}(p) \qquad p \in \Lambda_* \quad [46]$$

For $j = 1, \ldots, d$, let $\theta_j \in \mathbb{Z}^d$ be the unit vector with the $j$th component equal to 1. Then for $\langle \ell, \ell' \rangle \in E$, there exists $\theta_j$ such that $\ell - \ell' = \pm \theta_j$. Since the edge $\langle \ell, \ell' \rangle$ is an unordered set, let us fix $\ell' = \ell + \theta_j$. Thereby,

$$\frac{1}{|\Lambda|^{1/2}} \sum_{\langle \ell, \ell' \rangle \in E} \left( x_\ell^{(1)} - x_{\ell'}^{(1)} \right) \left( e^{i(p, \ell)} - e^{i(p, \ell')} \right)$$
$$= \frac{2}{|\Lambda|^{1/2}} \sum_{\ell \in \Lambda} \sum_{j=1}^{d} \left[ x_\ell^{(1)} e^{i(p, \ell)} - x_\ell^{(1)} e^{i(p, \ell)} \cos(p, \theta_j) \right]$$
$$= 2\hat{x}^{(1)}(p) \mathcal{E}(p)$$

In view of [41], one has

$$\phi_\Lambda[\hat{x}^{(1)}(p) \hat{x}^{(1)}(p')] = \delta_{0, p+p'} \widehat{K}_\Lambda^{(1)}(p)$$

Then employing the latter two facts and [37], we get

$$J \sum_{\langle \ell_2, \ell_2' \rangle \in E} \phi_\Lambda \left[ \left( x_{\ell_1}^{(1)} - x_{\ell_1'}^{(1)} \right) \left( x_{\ell_2}^{(1)} - x_{\ell_2'}^{(1)} \right) h_{\ell_2 \ell_2'}(p) \right.$$
$$= 2J\mathcal{E}(p) \phi_\Lambda \left[ \left( x_{\ell_1}^{(1)} - x_{\ell_1'}^{(1)} \right) \hat{x}^{(1)}(p) \right]$$
$$= 2J\mathcal{E}(p) \cdot \frac{1}{|\Lambda|^{1/2}} \sum_{p' \in \Lambda_*} \phi_\Lambda \left[ \hat{x}^{(1)}(p') \hat{x}^{(1)}(p) \right]$$
$$\times \left( e^{-i(p', \ell_1)} - e^{-i(p', \ell_1')} \right)$$
$$= 2J\mathcal{E}(p) \widehat{K}_\Lambda^{(1)}(p) h_{\ell_1 \ell_1'}(p)$$

which proves [46]. Then by [45] $\widehat{K}_\Lambda^{(1)}(p) \le 1/2J\mathcal{E}(p)$, for $p \ne 0$. The same holds for $\widehat{K}_\Lambda^{(k)}(p), k = 2, \ldots, N$, which by [39] yields [43]. □

The result just proved and the convergence of $K_\Lambda(\ell, \ell') \to K(\ell, \ell')$, as $L \to +\infty$, imply the infrared bound [4]. It turns out that the estimate [43]

may be used directly to prove the phase transition. Consider

$$P_\Lambda \stackrel{\text{def}}{=} \frac{1}{|\Lambda|^2} \sum_{\ell_1, \ell_2 \in \Lambda} \phi_\Lambda[(x_{\ell_1}, x_{\ell_2})] = \phi_\Lambda\left(\left|\frac{1}{|\Lambda|}\sum_{\ell \in \Lambda} x_\ell\right|^2\right) \geq 0 \quad [47]$$

where $\Lambda$ is the box [35]. By [40] and [41], we have

$$P_\Lambda = \frac{1}{|\Lambda|}\widehat{K}_\Lambda(0) \quad [48]$$

One can show that if $P \stackrel{\text{def}}{=} \lim_{L \to +\infty} P_\Lambda$ is positive, then there exist multiple Gibbs states. By [40], [41], and [48], we get that for any $\ell \in \Lambda$,

$$K_\Lambda(\ell, \ell) = P_\Lambda + \frac{1}{|\Lambda|}\sum_{p \in \Lambda_* \setminus \{0\}} \widehat{K}(p) \quad [49]$$

Suppose that, cf. [5],

$$K_\Lambda(\ell, \ell) \geq \varkappa > 0 \quad [50]$$

with $\varkappa$ independent of $\Lambda$ and $J$. Employing in [49] this estimate and [43], and passing to the limit $L \to +\infty$, we get

$$P \geq \varkappa - \mathcal{I}(d)N/2J \quad [51]$$

where

$$\mathcal{I}(d) \stackrel{\text{def}}{=} \frac{1}{(2\pi)^d}\int_{(-\pi,\pi]^d} \frac{\mathrm{d}p}{\mathcal{E}(p)} \quad [52]$$

which is finite for $d \geq 3$. Thereby, we have proved the following:

**Theorem 11** *For the spin model* [22], [23], *there exist multiple Gibbs states, and hence multiple phases, if* $d \geq 3$ *and* $J > \mathcal{I}(d)N/2\varkappa$.

Finally, let us pay some attention to the estimate [50], which is closely related with the properties of the single-spin measure $\chi$ (note that $\chi$ played no role in obtaining [26] and [43]). If it is the uniform measure on the unit sphere $S_{N-1} \subset \mathbb{R}^N$, then $K_\Lambda(\ell, \ell) = 1$ and [50] is trivial. In general, one has to employ some technique to obtain such an estimate.

## Reflection Positivity and Phase Transitions in Quantum Systems

As in the classical case, the way of proving the phase transition for appropriate models leads from an estimate like [17] to Gaussian domination and then to the infrared bound. However, here this way is much more complicated, so in the frames of this article we can only sketch its main elements basing on the original paper by Dyson *et al.* (1978), where the interested reader can find the details. As above, we start by studying reflection positive functionals.

### Reflection Positivity in Nonabelian Case

Again we consider a finite set $\Lambda$, $|\Lambda|$ being even. For every $\ell \in \Lambda$, let a complex Hilbert space $\mathcal{H}_\ell$ be given. This is the single-spin physical Hilbert space for our quantum system. We suppose that all $\mathcal{H}_\ell$, $\ell \in \Lambda$, are the copies of a certain finite-dimensional space $\mathcal{H}$. The physical Hilbert space $\mathcal{H}_\Lambda$ corresponding to $\Lambda$ is the tensor product of $\mathcal{H}_\ell$, $\ell \in \Lambda$. Let $\mathcal{A}_\Lambda$ be the algebra of all linear operators defined on $\mathcal{H}_\Lambda$. This is the algebra of observables in our case; it is noncommutative (nonabelian) and contains the unit element $I$ – the identity operator. As above, $\Lambda$ splits into two subsets $\Lambda_\pm$, which are the mirror images of each other, that is, we are given a reflection $\rho : \Lambda \to \Lambda$, such that $\rho(\Lambda_+) = \Lambda_-$. This allows us to introduce the corresponding subalgebras $\mathcal{A}_\Lambda^\pm$ by setting the elements of $\mathcal{A}_\Lambda^+$ to be of the form $A \otimes I$, where $A : \mathcal{H}_{\Lambda_+} \to \mathcal{H}_{\Lambda_+}$ is a linear operator and $I$ is the identity operator on $\mathcal{H}_{\Lambda_-}$. Respectively, the elements of $\mathcal{A}_\Lambda^-$ are to be of the form $I \otimes A$. Then we define the map $\vartheta : \mathcal{A}_\Lambda^+ \to \mathcal{A}_\Lambda^-$ as

$$\vartheta(A \otimes I) = I \otimes \bar{A} \quad [53]$$

where $A \mapsto \bar{A}$ is complex (not Hermitian) conjugation; it may be realized as transposing and taking Hermitian conjugation. For $A_1, \ldots, A_n \in \mathcal{A}$, one has $\bar{A}_1 \cdots \bar{A}_n = \overline{A_1, \cdots A_n}$. We also suppose that $\vartheta$ possesses the properties [8]. A linear functional $\phi : \mathcal{A}_\Lambda \to \mathbb{R}$ is called RP (with respect to the pair $\rho, \vartheta$) if it has the property [9].

**Definition 12** A functional $\phi$ is called generalized reflection positive (GRP) if for any $A_1, \ldots, A_n \in \mathcal{A}_\Lambda^+$,

$$\phi[A_1\vartheta(A_1) \cdots A_n\vartheta(A_n)] \geq 0 \quad [54]$$

In principle, this notion differs from the reflection positivity only in the nonabelian case. However, if the algebras $\mathcal{A}_\Lambda^\pm$ commute (they do commute in our case), a functional $\phi$ is RP if and only if it is GRP.

**Example 13** Let

$$\phi(A) = \operatorname{trace}(A), \quad A \in \mathcal{A}_\Lambda \quad [55]$$

Since the space $\mathcal{H}_\Lambda$ is finite dimensional, this $\phi$ is well defined. It is GRP. Indeed, as the algebras $\mathcal{A}_\Lambda^\pm$ commute, we have

$$\begin{aligned}
&\phi[A_1 \otimes I \cdot \vartheta(A_1 \otimes I) \cdots A_n \otimes I \cdot \vartheta(A_n \otimes I)] \\
&\quad = \phi[A_1 \otimes I \cdots A_n \otimes I \cdot \vartheta(A_1 \otimes I) \cdots \vartheta(A_n \otimes I)] \\
&\quad = \phi[A_1 \otimes I \cdots A_n \otimes I \cdot \vartheta(A_1 \otimes I \cdots A_n \otimes I)] \\
&\quad = \operatorname{trace}[A_1 \cdots A_n] \cdot \operatorname{trace}[\bar{A}_1 \cdots \bar{A}_n] \\
&\quad = |\operatorname{trace}[A_1 \cdots A_n]|^2 \geq 0
\end{aligned}$$

The Cauchy–Schwarz inequality [13] obviously holds also in the quantum case. By means of this inequality and the Trotter product formula

$$\exp(A+B) = \lim_{n\to+\infty} [\exp(A/n)\exp(B/n)]^n \quad [56]$$

one can prove that every RP functional obeys an estimate like [17]. Thereby, we have the following analog of Lemma 6:

**Lemma 14** *Let $A, B, C_1, \ldots, C_n \in \mathcal{A}_\Lambda^+$ be any self-adjoint operators possessing real matrix representation and $a_1, \ldots, a_m$ be any real numbers. Then*

$$\left[\mathrm{trace}\left\{\exp\left(A+\vartheta(B)-\sum_{n=1}^{m}[C_n-\vartheta(C_n)-a_n]^2\right)\right\}\right]^2 \leq \mathrm{trace}\left\{\exp\left(A+\vartheta(A)-\sum_{n=1}^{m}[C_n-\vartheta(C_n)]^2\right)\right\} \times \mathrm{trace}\left\{\exp\left(B+\vartheta(B)-\sum_{n=1}^{m}[C_n-\vartheta(C_n)]^2\right)\right\} \quad [57]$$

### Gaussian Domination and Phase Transitions

To proceed further we need a concrete model with finite-dimensional physical Hilbert spaces. As every quantum model, it is defined by its Hamiltonian. Let $\Lambda \subset \mathbb{Z}^d$ be the box [35] and $(\Lambda, E)$ be the same graph as in the subsection "Infrared bound." The periodic Hamiltonian of our model is

$$H_\Lambda = \sum_{\ell\in\Lambda} Q_\ell + \frac{1}{2}\sum_{\langle\ell,\ell'\rangle\in E} |S_\ell - S_{\ell'}|^2 \quad [58]$$

where at each $\ell \in \Lambda$ we have the copies $Q_\ell$, $S_\ell^{(1)}, \ldots, S_\ell^{(N)}$ of $N+1$ basic operators, acting in the Hilbert space $\mathcal{H}_\ell$, and

$$|S_\ell - S_{\ell'}|^2 = \sum_{k=1}^{N}\left(S_\ell^{(k)} - S_{\ell'}^{(k)}\right)^2$$

The only condition we impose so far is that all these operators can simultaneously be chosen as real matrices. For $h = (h_{\ell\ell'})_{\langle\ell,\ell'\rangle\in E} \in \mathbb{R}^{N|E|}$, we set

$$Z_\Lambda(h) = \mathrm{trace}\left\{\exp\left(-\beta\sum_{\ell\in\Lambda}Q_\ell - \frac{\beta}{2}\sum_{\langle\ell,\ell'\rangle\in E}|S_\ell - S_{\ell'} - h_{\ell\ell'}|^2\right)\right\} \quad [59]$$

where $\beta > 0$ is the inverse temperature.

**Theorem 15** *For the model [58] and any $h = (h_{\ell\ell'})_{\langle\ell,\ell'\rangle\in E} \in \mathbb{R}^{N|E|}$,*

$$Z_\Lambda(h) \leq Z_\Lambda(0) \quad [60]$$

The proof is performed by means of Lemma 14.

The periodic local Gibbs state of the model [58] at the inverse temperature $\beta$, analogous to the state [31], is

$$\phi_\Lambda(A) = \mathrm{trace}\{A\exp(-\beta H_\Lambda)\}/Z_\Lambda(0), \quad A \in \mathcal{A}_\Lambda \quad [61]$$

As in the classical case, one can define the parameter [47]. However, now the fact that $\lim_{L\to+\infty} P_\Lambda > 0$ does not yet imply the phase transition. One has to prove a more general fact

$$\lim_{L'\to+\infty}\left\{\lim_{L\to+\infty}\phi_\Lambda\left(\left|\frac{1}{|\Lambda'|}\sum_{\ell\in\Lambda'}S_\ell\right|^2\right)\right\} > 0 \quad [62]$$

where $\Lambda'$ is the box [35] of side $2L'$. Furthermore, in the quantum case the Gaussian domination [60] does not lead directly to the estimate [43], which yields [51]. Instead, one can get a bound like [43] but for the Duhamel two-point function (DTF). Given $A, B \in \mathcal{A}_\Lambda$, their DTF is

$$(A,B) = \int_0^1 \phi_\Lambda(A\mathrm{e}^{-\xi\beta H_\Lambda}B\mathrm{e}^{\xi\beta H_\Lambda})\mathrm{d}\xi \quad [63]$$

By means of [56] one can show that

$$(A,B) = \frac{1}{Z_\Lambda(0)} \times \left\{\frac{\partial^2}{\partial\xi\partial\eta}\mathrm{trace}[\exp(\xi A + \eta B - \beta H_\Lambda)]\right\}_{\xi=\eta=0} \quad [64]$$

Let $\hat{S}(p) = (\hat{S}^{(1)}(p), \ldots, \hat{S}^{(N)}(p)), p \in \Lambda_*$, be the Fourier image of $S_\ell$, defined by [37], [38]. Then

$$\left(\hat{S}(p), \hat{S}(-p)\right) = \sum_{k=1}^{N}\left(\hat{S}^{(k)}(p), \hat{S}^{(k)}(-p)\right)$$

**Theorem 16** *For all $p \in \Lambda_*\backslash\{0\}$, it follows that*

$$\left(\hat{S}(p), \hat{S}(-p)\right) \leq \frac{N}{2\beta\mathcal{E}(p)} \quad [65]$$

To prove this statement one has to use the Gaussian bound [60] exactly as in the case of Theorem 10. The second derivative with respect to $\xi$ gives the corresponding DTF (see [64]).

Now let us indicate how the infrared bound [65] leads to the phase transition. To this end we use the simplest quantum spin model with the Hamiltonian [58], for which $Q_\ell = 0, N = 2$, and $S_\ell^{(k)}, k = 1, 2$, being the copies of the Pauli matrices

$$S^{(1)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad S^{(2)} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Then

$$K_\Lambda^{(k)}(\ell,\ell) = \phi_\Lambda\left(S_\ell^{(k)}\cdot S_\ell^{(k)}\right) = 1 \quad \text{for all } \ell \in \Lambda, \quad k = 1, 2 \quad [66]$$

which gives the bound $\varkappa$ (see [50]). For $A, B \in \mathcal{A}_\Lambda$, by $[A, B]$ we denote the commutator $AB - BA$. Set

$$\Sigma_\Lambda^{(k)}(p) = \phi_\Lambda\left(\left[\hat{S}^{(k)}(p), \left[H_\Lambda, \hat{S}^{(k)}(-p)\right]\right]\right)$$
$$k = 1, 2 \qquad [67]$$

The phase transition in the model we consider can be established by means of the following statement (see Dyson 1978, Theorem 5.1).

**Proposition 17** *Suppose there exist* $\Sigma^{(k)}(p), k = 1, 2$, $p \in (-\pi, \pi]^d$ *such that, for all* $L \in \mathbb{N}$,

$$\Sigma_\Lambda^{(k)}(p) \le \Sigma^{(k)}(p), \quad k = 1, 2, \ p \in \Lambda_* \qquad [68]$$

*Then the model undergoes a phase transition at a certain finite* $\beta$ *if* $d \ge 3$ *and*

$$\frac{1}{(2\pi)^d}\int_{(-\pi,\pi]^d}\left[\frac{\Sigma^{(k)}(p)}{8\mathcal{E}(p)}\right]^{1/2} \mathrm{d}p < 1 \qquad [69]$$

*for a certain, and hence for both,* $k = 1, 2$.

Thus to prove the phase transition we have to estimate $\Sigma_\Lambda^{(k)}(p), k = 1, 2$. By means of the Cauchy–Schwarz inequality, the estimate [69] may be transformed into the following:

$$\frac{1}{(2\pi)^d}\int_{(-\pi,\pi]^d}\left[\Sigma^{(1)}(p) + \Sigma^{(2)}(p)\right]\mathrm{d}p < 16/\mathcal{I}(d)$$

where $\mathcal{I}(d)$ is the same as in [52]. The integral on the left-hand side can be estimated from above by $8\sqrt{d(d+1)}$; hence, the latter inequality holds if

$$\mathcal{I}(d)\sqrt{d(d+1)} < 2$$

which holds for all $d \ge 3$. In particular, $\mathcal{I}(3) \approx 0.505$.

## Bibliographic Notes

As the original sources on the RP method in the theory of phase transitions we mention the papers Fröhlich *et al.* (1976) (classical case), Dyson *et al.* (1978) (quantum case), and Fröhlich and Lieb (1978) (both cases). In a unified way and with many examples, this method is described in Fröhlich *et al.* (1978, 1980). A detailed analysis of the method, especially in its applications to classical models with unbounded spins, was given in Shlosman (1986). The techniques based on the chessboard and contour estimates are described in Fröhlich and Lieb (1978) and Shlosman (1986). As was mentioned above, the quantum case is much more complicated; it gets even more complicated if one deals with quantum models employing infinite-dimensional physical Hilbert spaces and unbounded operators, such as quantum crystals. The adaptation of the RP method to such models was made in Driessler *et al.* (1979), Pastur and Khoruzhenko (1987), Barbulyak and Kondratiev (1992), and Kondratiev (1994). In the latter two papers a general version of the quantum crystal was studied in the framework of the Euclidean approach, based on functional integrals (see Albeverio *et al.* (2002)). In this approach the quantum crystal is represented as a lattice spin model with unbounded infinite-dimensional spins. Like in the case of classical models with unbounded spins, here establishing the estimate [5] becomes a highly nontrivial task. In particular cases, for example, for $\phi^4$-models, one applies special tools like the Bogoliubov inequalities (see Driessler *et al.* (1979) and Pastur and Khoruzhenko (1987)). In the general case quasiclassical asymptotics allow us to get the lower bound [5] (see Barbulyak and Kondratiev (1992) and Kondratiev (1994)). There is one more technique based on reflection positivity (see Lieb (1989)). It employs reflections in spin spaces, whereas the properties of the index sets (lattices) play no role. This technique proved to be useful in the theory of strongly correlated electron systems, see Tian (2004). Finally, we mention the books of Georgii (1988), Prum (1986), and Sinai (1982) where different aspects of the RP method are described. In Georgii (1988), one can also find extended bibliographical and historical comments on this subject.

*See also:* Phase Transition Dynamics; Phase Transitions in Continuous Systems; Quantum Spin Systems; Renormalization: Statistical Mechanics and Condensed Matter.

## Further Reading


Albeverio S, Kondratiev YG, Kozitsky Y, and Röckner M (2002) Euclidean Gibbs states of quantum lattice systems. *Reviews in Mathematical Physics* 14: 1–67.

Barbulyak VS and Kondratiev YG (1992) The quasiclassical limit for the Schrödinger operator and phase transitions in quantum statistical physics. *Functional Analysis and its Applications* 26(2): 61–64.

Driessler W, Landau L, and Fernando-Perez J (1979) Estimates of critical lengths and critical temperatures for classical and quantum lattice systems. *Journal of Statistical Physics* 20: 123–162.

Dyson FJ, Lieb EH, and Simon B (1978) Phase transitions in quantum spin systems with isotropic and nonisotropic interactions. *Journal of Statistical Physics* 18: 335–383.

Fröhlich J, Israel R, Lieb EH, and Simon B (1978) Phase transitions and reflection positivity, I. General theory and long range lattice models. *Communications in Mathematical Physics* 62: 1–34.

Fröhlich J, Israel R, Lieb EH, and Simon B (1980) Phase transitions and reflection positivity, II. Lattice systems with short-range and Coulomb interactions. *Journal of Statistical Physics* 22: 297–347.

Fröhlich J and Lieb EH (1978) Phase transitions in anisotropic lattice spin systems. *Communications in Mathematical Physics* 60: 233–267.

Fröhlich J, Simon B, and Spenser T (1976) Infrared bounds, phase transitions and continuous symmetry breaking. *Communications in Mathematical Physics* 50: 79–85.

Georgii H-O (1988) *Gibbs Measures and Phase Transitions*. Studies in Mathematics, vol.9. Berlin: Walter de Gruyter.

Kondratiev JG (1994) Phase transitions in quantum models of ferroelectrics. In: *Stochastic Processes, Physics and Geometry*, vol. II, pp. 465–475. Singapure: World Scientific.

Lieb EH (1989) Two Theorems on the Hubbard Model. *Physical Review Letters* 62: 1201–1204.

Pastur LA and Khoruzhenko BA (1987) Phase transitions in quantum models of rotators and ferroelectrics. *Theoretical and Mathematical Physics* 73: 111–124.

Prum P (1986) *Processus sur un Réseau et Mesures de Gibbs*. Applications. Techniques Stochastiques. Paris: Masson.

Prum B and Fort J-C (1991) *Stochastic Processes on a Lattice and Gibbs Measures* (translated from the French by Bertram Eugene Schwarzbach and revised by the authors). Mathematical Physics Studies, vol.11. Dordrecht: Kluwer Academic Publishers Group.

Shlosman SB (1986) The method of reflection positivity in the mathematical theory of first-order phase transitions. *Russian Mathematical Surveys* 41: 83–134.

Sinai Ya (1982) *Theory of Phase Transitions. Rigorous Results*. Oxford: Pergamon.

Tian G-S (2004) Lieb's spin-reflection-positivity method and its applications to strongly correlated electron systems. *Journal of Statistical Physics* 116: 629–680.

# Regularization for Dynamical $\zeta$-Functions

**V Baladi**, Institut Mathématique de Jussieu, Paris, France



## Introduction

If $A$ is a finite, say $N \times N$, matrix with complex coefficients, the following easy equality gives an expression for the polynomial $\prod_{k=1}^{N}(1 - z\lambda_k) = \det(\mathrm{Id} - zA)$:

$$\det(\mathrm{Id} - zA) = \exp\left(-\sum_{n=1}^{\infty} \frac{z^n}{n} \mathrm{tr}\, A^n\right) \quad [1]$$

(here, Id denotes the identity matrix and tr is the trace of a matrix). Even in this trivial finite-dimensional case, the $z$-radius of convergence of the logarithm of the right-hand side only gives information about the spectral radius (the modulus of the largest eigenvalue) of $A$. The zeros of the left-hand side (i.e., the inverses $z = 1/\lambda_k$ of the nonzero eigenvalues of $A$) can only be located after extending holomorphically the right-hand side. The purpose of this article is to discuss some dynamical situations in which $A$ is replaced by a linear bounded operator $\mathcal{L}$, acting on an infinite-dimensional space, and for which a dynamical determinant (or dynamical $\zeta$-function), constructed from periodic orbits, takes the part of the right-hand side. In the examples presented, $\mathcal{L}$ will be a transfer operator associated to a weighted discrete-time dynamical system: given a transformation $f: M \to M$ on a compact manifold $M$ and a function $g: M \to \mathbb{C}$, we set

$$\mathcal{L}\varphi = g \cdot \varphi \circ f^{-1} \quad [2]$$

(If $f$ is not inversible, it is understood, e.g., that $f$ has at most finitely many inverse branches, and that the right-hand side of [2] is the sum over these inverse branches, see the next section.) We let $\mathcal{L}$ act on a Banach space of functions or distributions $\varphi$ on $M$. For suitable $g$ (in particular $g = |\det Tf^{-1}|$ when this Jacobian makes sense), the spectrum of $\mathcal{L}$ is related to the fine statistical properties of the dynamics $f$: existence and uniqueness of equilibrium states (related to the maximal eigenvector of $\mathcal{L}$), decay of correlations (related to the spectral gap), limit laws, entropies, etc: see, for example, Baladi (1998) or Cvitanović *et al.* (2005). The operator $\mathcal{L}$ is not always trace-class, indeed, it sometimes is not compact on any reasonable space. Even worse, its essential spectral radius may coincide with its spectral radius. (Recall that the essential spectral radius of a bounded linear operator $\mathcal{L}$ acting on a Banach space is the infimum of those $\rho > 0$, such that the spectrum of $\mathcal{L}$ outside of the disk of radius $\rho$ is a finite set of eigenvalues of finite algebraic multiplicity.) However, various techniques allow us to prove that a suitable dynamically defined replacement for the right-hand side of [1] extends holomorphically to a disk in which its zeros describe at least part of the spectrum of $\mathcal{L}$. Some of these techniques have a "regularization" flavor, and we shall concentrate on them.

In the following section, we present the simplest case: analytic expanding or hyperbolic dynamics, for which no regularization is necessary and the Grothendieck–Fredholm theory can be applied. Next, we consider analytic situations where finitely many neutral periodic orbits introduce branch cuts in the dynamical determinant, and see how to "regularize" them. Finally, we discuss a

kneading operator regularization approach, inspired by the work of Milnor and Thurston, and applicable to dynamical systems with finite smoothness.

Despite the terminology, none of the regularization techniques discussed below match the following "$\zeta$-regularization" formula:

$$\prod_{k=1}^{\infty} a_k = \exp\left(-\frac{\mathrm{d}}{\mathrm{d}s}\sum_{k=1}^{\infty} a_k^{-s}|_{s=0}\right) \qquad [3]$$

(For information about the above $\zeta$-regularization and its applications to physics, we refer, e.g., to Elizalde 1995. See also Voros (1987) and Fried (1986) for more geometrical approaches and further references, e.g., to the work of Ray and Singer.)

We do not cover all aspects of dynamical $\zeta$-functions here. For more information and references, we refer to our survey Baladi (1998), to the more recent surveys by Pollicott (2001) and Ruelle (2002), and also to the exhaustive account by Cvitanović *et al.* (2005), which contains a rich array of physical applications.

## The Grothendieck–Fredholm Case

Let $M$ be a real analytic compact manifold (e.g., the circle or the $d$-torus), and let $f: M \to M$ be real analytic and $g: M \to \mathbb{C}$ be analytic.

First suppose that $f$ is uniformly expanding, that is, there is $\lambda > 1$ so that $\|Tf(v)\| \geq \lambda\|v\|$. (For example, $f(z) = z^2$ on the unit circle, or a small analytic perturbation thereof.) Consider

$$\mathcal{L}_{f,g}\varphi(x) = \sum_{y:\, f(y)=x} g(y)\varphi(y) \qquad [4]$$

(For example, with $g(y) = 1/|\det Tf(y)|$ or $1/|\det Tf(y)|^s$.) Ruelle (1976) proved that an operator $\mathcal{L}_0$, which is essentially the same as $\mathcal{L}_{f,g}$ (the difference, if any, arises from the use of Markov partitions, especially in higher dimensions), acting on a Banach space of holomorphic and bounded functions, is not only compact, but is in fact a nuclear operator in the sense of Grothendieck. In particular, the traces of all its powers are well defined, and the Grothendieck–Fredholm (Gohberg *et al.* 2000) determinant

$$d_0(z) = \exp\left(-\sum_{n=1}^{\infty}\frac{z^n}{n}\operatorname{tr}\mathcal{L}_0^n\right) \qquad [5]$$

extends to an entire function of finite order, the zeros of which are exactly the inverses of the nonzero eigenvalues of $\mathcal{L}_0$. (The order of the zero coincides with the algebraic multiplicity of the eigenvalue.) Ruelle also proved that the traces can be written as sums over periodic orbits:

$$\operatorname{tr}\mathcal{L}_0^n = \sum_{x:\, f^n(x)=x}^{*} \frac{\prod_{k=0}^{n-1} g(f^k x)}{|\det(\mathrm{Id} - Tf_x^{-n})|}$$

where $\sum^*$ means that the fixed points of $f^n$ lying in the intersection of two or more elements of the Markov partition must be counted two or more times. (Note that if $f^n(x) = x$, then this closed orbit gives a natural inverse branch for $f^{-n}$.) Taking into account the periodic orbits on the boundaries of the Markov partition, Ruelle expresses the following "dynamical determinant":

$$d_{f,g}(z) = \exp\left[-\sum_{n=1}^{\infty}\frac{z^n}{n}\sum_{x:\, f^n(x)=x}\frac{\prod_{k=0}^{n-1} g(f^k x)}{|\det(\mathrm{Id} - Tf_x^{-n})|}\right] \qquad [6]$$

as an alternated product of determinants $d_0(z)$ as in [5].

The expression [6] is sometimes also called a "dynamical $\zeta$-function," but we prefer to reserve this terminology for the following power series:

$$\zeta_{f,g}(z) = \exp\left[+\sum_{n=1}^{\infty}\frac{z^n}{n}\sum_{x:\, f^n(x)=x}\prod_{k=0}^{n-1} g(f^k x)\right] \qquad [7]$$

It is not difficult to write $\zeta_{f,g}(z)$ as (Baladi 1998) an alternated product of determinants $d_{f,g_i}$, for $i = 0, \ldots, d$, and appropriate weights $g_i$.

In fact, the results just described hold in more generality, for example, for piecewise bijective and analytic interval maps. Such maps, $f$, appear naturally, for example, when considering Schottky subgroups of $PSL(2,\mathbb{Z})$. We mention the recent work of Guillopé–Lin–Zworski (2004), who let the transfer operator associated to such $f$ and weights $g_s(y) = 1/|f'(y)|^s$ act (as trace-class operators) on suitable Hilbert spaces of holomorphic functions. This allows them to obtain precise estimates for the number of zeros of $s \mapsto d_{f,g_s}$ [1] in the complex plane: these zeros are the resonances (in the sense of the spectrum of the Laplacian).

Note that the nuclearity properties extend also to the Gauss map $f(x) = \{1/x\}$, which has infinitely many inverse branches, if the weight $g$ has summability properties over the branches (e.g., $g_s(y) = |1/f'(y)|^s$, where $s$ is a complex parameter, with $\Re s > 1/2$). The dynamical determinant $d_{f,g_s}(z)$ for the transfer operator of the Gauss map is related to the Selberg $\zeta$-function (see e.g., Chang and Mayer (2001) and references therein).

Next, assume that $M$ and $g$ are as before, but $f$ is a uniformly hyperbolic real analytic diffeomorphism.

For example, $M$ is the 2-torus and $f$ is a small real analytic perturbation of the linear automorphism

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

More generally, we may assume that $f$ is a real analytic Anosov diffeomorphism, that is, there are $C \geq 1$ and $\lambda > 1$ such that the tangent bundle decomposes as $TM = E^u \oplus E^s$, where the dynamical bundles $E^u$ and $E^s$ are $Tf$-invariant, with $\|Tf^n|_{E^s}\| \leq C\lambda^{-n}$ and $\|Tf^{-n}|_{E^u}\| \leq C\lambda^{-n}$ for all $n \in \mathbb{Z}_+$. In general, the smoothness of $x \mapsto E^u(x)$ and $E^s(x)$ is only Hölder. Under the very strong additional assumption that $E^u(x)$ and $E^s(x)$ are real analytic, Ruelle (1976) (see also Fried (1986)) showed that the power series $d_{f,g}(z)$ can again be written as a finite alternated product (this product being again an artifact of the Markov partition) of entire functions of finite order. For this, he constructed auxiliary transfer operators associated to the expanding (and analytic!) quotiented dynamics acting on holomorphic functions on disks. The analyticity assumption on the dynamical bundles was later lifted by Rugh (1996) (see also Fried (1995)), who let their transfer operators act on Banach topological tensor products of spaces of holomorphic functions on a disk with the dual of such a space. In all these cases, the transfer operator is a nuclear operator in the sense of Grothendieck and no regularization is needed. (More recent work of Kitaev (1999), when applied to this analytic setting, shows that the "meromorphic" function $d_{f,g}(z)$ in fact does not have poles.)

## Regularization and Intermittency

Consider the interval $M = [0, 1]$, and $f$ defined on $M$ by $f(x) = f_1(x) = x/(1-x)$ on $[0, 1/2]$, and $f(x) = f_2(x) = (1-x)/x$ on $[1/2, 1]$. (This is the Farey map, which appears naturally when considering continued fractions.) Each of the two branches is an analytic bijection onto [0, 1]. The second branch is expanding, but the first one, $f_1$, has a (parabolic) neutral fixed point at $x = 0$ (the expansion is $f(x) = x + x^2 + x^3 + \cdots$). Let $g = g_s$ be an analytic weight of the form $g(y) = 1/|f'(y)|^s$ for $\Re s \geq 1/2$. We are interested in the spectrum of the operator $\mathcal{L}_{f,g}$ associated with the pair $(f, g)$ by [4]. Clearly, the expression [6] is not a good candidate for an analog of the Fredholm determinant of $\mathcal{L}_{f,g}$. Rugh (1996) introduced a Banach space $\mathcal{B}$ of functions in a complex neighborhood of $M$, having a controlled singularity at 0, and such that the spectral radius of $\mathcal{L}_{f,g}$ on $\mathcal{B}$ is equal to 1, and such that the following regularized determinant

$$d_{f,g}(z) = \exp\left[-\sum_{n=1}^{\infty}\frac{z^n}{n}\sum_{x\in(0,1]:\, f^n(x)=x}\frac{\prod_{k=0}^{n-1} g_s(f^k x)}{1 - Tf_x^{-n}}\right] \quad [8]$$

is a holomorphic function in the cut complex plane $\{z \in \mathbb{C} \mid z \notin [1, \infty)\}$. Furthermore, its zeros $z$ in this cut plane are in bijection with the spectrum of $\mathcal{L}_{f,g}|_{\mathcal{B}}$ outside of the unit interval $[0, 1]$, and this spectrum consists of eigenvalues $1/z$ of finite multiplicities. Finally, these eigenvalues can only accumulate at 0 or 1, although each point in the unit interval belongs to the spectrum of $\mathcal{L}_{f,g}$. In particular, the essential spectral radius of $\mathcal{L}_{f,g}$ on $\mathcal{B}$ coincides with its spectral radius.

Let us define the Banach space $\mathcal{B}$ and explain the key ideas in the proof of the above result (Rugh's claim is in fact more general than the statement above and applies to a class of maps $f$ with neutral fixed points). The starting point is the decomposition

$$\mathcal{L}_{f,g} = \mathcal{L}_1 + \mathcal{L}_2$$

where $\mathcal{L}_i\varphi = \varphi \circ f_i^{-1} \cdot |(f_i^{-1})'|^s$. The operator $\mathcal{L}_2$ is of the type discussed in the previous section, and it is nuclear when acting, for example, on bounded holomorphic functions in a complex neighborhood of $M$. Since $f_1$ is not expanding (because of the parabolic fixed point at 0), other ideas must be used to handle the operator $\mathcal{L}_1$. The change of coordinates (this idea goes back to Fatou) $w = 1/x$ replaces the weak contraction $f_1^{-1}$ by the translation $w \mapsto w + 1$ in a suitable domain containing a half-plane $\Re w > w_0$. In order to take into account the weight $g_s$, it is convenient to use the change of variables $\Psi(w) = \varphi(1/w) \cdot w^{-2s}$. Indeed, in the new coordinates the operator $\mathcal{L}_1$ reads as

$$\mathcal{M}_1\Psi(w) = \Psi(w+1)$$

The next step consists in letting $\mathcal{M}_1$ act on the Banach space $\mathcal{B}_w$ of Laplace transforms of $L^1(\mathbb{R}^+, \text{Lebesgue})$, that is, functions

$$\Psi(w) = \int_0^{\infty} \mathrm{e}^{-(w-w_0)t}\psi(t)\,\mathrm{d}t$$

with the induced norm $\|\Psi\|_{\mathcal{B}_w} = \int |\psi(t)|\,\mathrm{d}t$. Since $\mathcal{M}_1$ maps $\psi$ to $\mathrm{e}^{-t}\psi(t)$, it is not difficult to see that the spectrum of $\mathcal{M}_1$ on $\mathcal{B}_w$ (and thus of $\mathcal{L}_1$ on the pullback $\mathcal{B}$ of $\mathcal{B}_w$ by $\Phi$, which consists of functions in a complex neighborhood of [0,1], holomorphic in a sector at 0, and with a possible, but controlled, singularity at 0) is the closed unit interval. One can check that $\mathcal{L}_2$ is nuclear on $\mathcal{B}$. Composing a bounded operator with a

nuclear operator gives a nuclear operator. If $1/z \notin [0,1]$, the resolvent $(1 - z\mathcal{L}_1)^{-1}$ is a bounded operator, and therefore, for such $z$, the operator

$$\mathcal{P}(z) := z\mathcal{L}_2(1 - z\mathcal{L}_1)^{-1} \quad [9]$$

is nuclear on $\mathcal{B}$. We view $\mathcal{P}(z)$ as a "regularized" version of $\mathcal{L}_{f,g} = \mathcal{L}_1 + \mathcal{L}_2$. Now, since

$$\begin{aligned}(1 - z\mathcal{L}_{f,g})^{-1} &= (1 - z(\mathcal{L}_1 + \mathcal{L}_2))^{-1} \\ &= (1 - z\mathcal{L}_1)^{-1}\left(1 - z\mathcal{L}_2(1 - z\mathcal{L}_1)^{-1}\right)^{-1}\end{aligned}$$

it is not surprising that one can prove (Rugh 1996) that the Fredholm determinant

$$u \mapsto \det\left(1 - \mathcal{L}_2(u - \mathcal{L}_1)^{-1}\right)$$

(which is holomorphic in $u \notin [0,1]$) has as its zero set $\mathrm{sp}(\mathcal{L}_{f,g}|_{\mathcal{B}}) \setminus [0,1]$, and that this set consists in isolated eigenvalues of finite multiplicity (equal to the order of the corresponding zero) for $\mathcal{L}_{f,g}$. Formally,

$$(1 - z\mathcal{L}_1)^{-1} = \sum_{k=0}^{\infty} z^k \mathcal{L}_1^k \quad [10]$$

so that the regularization we just described can be viewed as mirroring an induction (or renormalization) procedure, where the dynamics $f$ is replaced by the first-return map to the "chaotic" part of the phase space [0, 1/2]. (For the Farey map, the induced map is just the Gauss map.) The formal equality [10] is also behind the fact that (Rugh 1996)

$$\mathrm{tr}\,\mathcal{P}(z)^n = \sum_{x \neq 0:\, f^n(x) = x} \frac{\prod_{k=0}^{n-1} g_s(f^k x)}{1 - Tf_x^{-n}}$$

An extension of this theory to the two-dimensional setting has been obtained by Baladi, Pujals, and Sambarino.

## Regularization and Kneading Determinants

Up to now we have only discussed analytic dynamical systems, for which hyperbolicity (or uniform expansion) guaranteed that the transfer operator (or a regularized version thereof) was compact, even nuclear, on a natural Banach space. When considering hyperbolic invertible (or expanding noninvertible) maps $f$, and weights $g$ with "finite smoothness," say $C^r$ for some finite $r > 1$, the transfer operator defined by [2] or [4] is usually not compact on any infinite-dimensional space. However, one can often prove a "Lasota–Yorke" type inequality (see e.g., Baladi (1998)) which ensures that the essential spectral radius $\rho_{\mathrm{ess}}(\mathcal{L}_{f,g})$, defined in the "Introduction," is strictly smaller than the spectral radius. Then, the goal is to prove that the dynamical determinant [6] defines a holomorphic function in the disk of radius $1/\rho_{\mathrm{ess}}$, and that its zeros in this disk are exactly the inverses of the eigenvalues of $\mathcal{L}_{f,g}$. For uniformly expanding $C^r$ maps $f$ on compact manifolds, and $C^r$ weights, denoting by $\lambda > 1$ the expansion coefficient as in the section "The Grothendieck–Fredholm case," this goal was essentially attained by Ruelle (1990). For $\mathcal{L}_{f,g}$ acting on the Banach space of $C^r$ functions on $M$, Ruelle proved $\rho_{\mathrm{ess}}(\mathcal{L}_{f,g}) \leq \lambda^{-r}$ and was able to extend $d_{f,g}(z)$ (and interpret its zeros) in the disk of radius $\lambda^r$.

For $C^r$ Anosov diffeomorphisms $f$, and $C^r$ weights $g$, Pollicott, Ruelle, Haydn, and others obtained important results using the symbolic dynamics description (for which the maximal smoothness which can be used is $r \leq 1$, because of the metric-space model). Later, Kitaev (1999) was able to show that $d_{f,g}(z)$ extends to a holomorphic function in the disk of radius $\lambda^{-r/2}$, but did not give any spectral interpretation of the zeros of $d_{f,g}(z)$. More recently, Liverani (2005) was able to give such an interpretation, in a smaller disk however.

All the works mentioned in the previous paragraph are based on some approximation scheme (Taylor expansion style). In the early 1990s, a new approach, with a regularization flavor, was launched (see e.g., Baladi and Ruelle (1996)), initially for piecewise monotone interval maps. We present it next.

Consider a finite set of local homeomorphisms $\psi_\omega : U_\omega \to \psi_\omega(U_\omega)$, where each $U_\omega$ is a bounded open interval of $\mathbb{R}$, and of associated weight functions $g_\omega$ which are continuous, of bounded variation, and have support inside $U_\omega$. For example, the $\psi_\omega$ can be the inverse branches of a single piecewise monotone interval map $f$, and $g_\omega$ can be $g \circ \psi_\omega$ for a single $g$. (No contraction assumption is required on the $\psi_\omega$: their graph can even coincide with the diagonal on a segment.) The transfer operator is now

$$\mathcal{M}\varphi = \sum_\omega g_\omega \cdot (\varphi \circ \psi_\omega)$$

Ruelle obtained an estimate, noted $\widehat{R}$, for the essential spectral radius of $\mathcal{M}$ acting on the Banach space BV of functions of bounded variation. The main result of Baladi and Ruelle (1996) links the eigenvalues of $\mathcal{M} : \mathrm{BV} \to \mathrm{BV}$ outside of the disk of radius $\widehat{R}$, with the zeros of the following "sharp determinant":

$$\det{}^{\#}(\mathrm{Id} - z\mathcal{M}) = \exp\left(-\sum_{n=1}^{\infty} \frac{z^n}{n} \mathrm{tr}^{\#} \mathcal{M}^n\right) \quad [11]$$

where (with the understanding that $y/|y| = 0$ if $y = 0$)

$$\mathrm{tr}^{\#}\mathcal{M} = \sum_\omega \int \frac{1}{2} \frac{\psi_\omega(x) - x}{|\psi_\omega(x) - x|}\, \mathrm{d}g_\omega(x)$$

If the $\psi_\omega$ are strict contractions which form the set of inverse branches of a piecewise monotone interval map $f$, and $g_\omega = g \circ \psi_\omega$, then integration by parts together with the key property that

$$\mathrm{d}\frac{x}{2|x|} = \delta, \text{the Dirac delta at the origin of } \mathbb{R}$$

show that $\det^\#(\mathrm{Id} - z\mathcal{M}) = 1/\zeta_{f,g}(z)$ (recall [7]). If one assumes instead only that the graph of each admissible composition $\psi^n_{\boldsymbol{\omega}}$ of $n$ successive $\psi_\omega$'s (with $n \geq 1$) intersects the diagonal transversally, then

$$\det{}^\#(\mathrm{Id} - z\mathcal{M}) = \exp\left[-\sum_{n=1}^{\infty}\frac{z^n}{n}\sum_{\text{admissible } \psi^n_{\boldsymbol{\omega}}}\sum_{x:\psi^n_{\boldsymbol{\omega}}(x)=x} L\left(x, \psi^n_{\boldsymbol{\omega}}\right) \times \prod_{k=0}^{n-1} g_{\omega_k}\left(\psi^k_{\boldsymbol{\omega}}(x)\right)\right] \qquad [12]$$

where $L(x,\psi) \in \{-1,1\}$ is the Lefschetz number of a transversal fixed point $x = \psi(x)$ (if $\psi$ is $C^1$ this is just $\mathrm{sgn}\,(1 - \psi'(x))$). Therefore, we call the sharp determinant $\det^\#(\mathrm{Id} - z\mathcal{M})$ a Ruelle–Lefschetz (dynamical) determinant. For a class of "unimodal" interval maps $f$ and constant weight $g = 1$, the expression [12] with Lefschetz numbers, coming from the additional transversality assumption, gives that $\det^\#(\mathrm{Id} - z\mathcal{M})$ is just $1/\zeta^-(z)$, where the "negative $\zeta$-function"

$$\zeta^-(z) = \exp\left[+\sum_{n=1}^{\infty}\frac{z^n}{n}(2\#\mathrm{Fix}^-(f^n) - 1)\right] \qquad [13]$$

is defined by counting (twice) the sets

$$\mathrm{Fix}^-(f^n) = \{x | f^n(x) = x,\ f \text{ strictly decreasing in a neighborhood of } x\}$$

of "negative fixed points." This negative $\zeta$-function was studied by Milnor and Thurston, who proved the remarkable identity

$$(\zeta^-(z))^{-1} = \det(1 + \widehat{D}(z))$$

where $\widehat{D}(z)$ is a $1 \times 1$ "matrix," which is just a power series in $z$ with coefficients in $\{-1, 0, +1\}$, given by the signed itinerary of the image of the turning point (the so-called "kneading" data).

Returning now to the general setup $\psi_\omega$, $g_\omega$, the crucial step in the proof of the spectral interpretation of the zeros of this Ruelle–Lefschetz determinant consists in establishing the following continuous version of the Milnor–Thurston identity:

$$\det{}^\#(\mathrm{Id} - z\mathcal{M}) = \det{}^*(\mathrm{Id} + \widehat{\mathcal{D}}(z)) \qquad [14]$$

where the "kneading operator" $\widehat{\mathcal{D}}(z)$ replaces (formally) the finite kneading matrix of Milnor and Thurston. In a suitable $z$-disk, one proves that this operator $\widehat{\mathcal{D}}(z)$ is a Hilbert–Schmidt operator on an $L^2$ space (its kernel is bounded and compactly supported), thus allowing the use of regularized determinants of order 2 (see e.g., Gohberg *et al.* (2000)). By definition, $\det^*(\mathrm{Id} + \widehat{\mathcal{D}}(z))$ is the product of this regularized determinant with the exponential of the average of the kernel of $\widehat{\mathcal{D}}(z)$ along the diagonal, which is well defined. Another kneading operator, $\mathcal{D}(z)$, is essential. If $1/z$ is not in the spectrum of $\mathcal{M}$ (on BV), then $\mathcal{D}(z)$ is also Hilbert–Schmidt, and one can show $\det^*(\mathrm{Id} + \widehat{\mathcal{D}}(z)) = \det^*(\mathrm{Id} + \mathcal{D}(z))^{-1}$. The initial definitions of $\widehat{\mathcal{D}}(z)$ and $\mathcal{D}(z)$ were technical and we shall not give them here. However, a more conceptual definition of the $\mathcal{D}(z)$ was later implemented:

$$\mathcal{D}(z) = \mathcal{N}(\mathrm{Id} - z\mathcal{M})^{-1}\mathcal{S} \qquad [15]$$

where $\mathcal{N}$ is an auxiliary transfer operator and $\mathcal{S}$ is the convolution

$$\mathcal{S}\varphi(x) = \int \frac{1}{2}\frac{x-y}{|x-y|}\varphi(y)\,\mathrm{d}\mu$$

where $\mu$ is an auxiliary non-negative finite measure. From [15], it becomes clear that the kneading operator is a regularized (through the convolution $\mathcal{S}$) object which describes the inverse spectrum of the transfer operator: the resolvent $(\mathrm{Id} - z\mathcal{M})^{-1}$ in [15] means that poles can only appear if $1/z$ is an eigenvalue. Since $\det^*(\mathrm{Id} + \widehat{\mathcal{D}}(z)) = \det^*(\mathrm{Id} + \mathcal{D}(z))^{-1}$, this can be translated into a statement for zeros of $\det^*(\mathrm{Id} + \widehat{\mathcal{D}}(z))$. The Milnor–Thurston identity [14] then implies that any zero of $\det^\#(\mathrm{Id} - z\mathcal{M})$ is an inverse eigenvalue of $\mathcal{M}$.

The one-dimensional kneading regularization we just presented is well understood. The higher-dimensional theory is not as developed yet. Let $U_\omega$ be now finitely many bounded open subsets of $\mathbb{R}^d$, $\psi_\omega : U_\omega \to \psi_\omega(U_\omega)$ be local $C^r$ homeomorphisms or diffeomorphisms, while $g_\omega : U_\omega \to \mathbb{C}$ are compactly supported $C^r$ functions, for $r \geq 1$.

In 1995, A Kitaev wrote a two-page sketch proving a higher-dimensional Milnor–Thurston formula, under an additional transversality assumption. This assumption guarantees that the set of fixed points of each fixed period $m$ is finite, so that the Ruelle–Lefschetz determinant $\det^\#(\mathrm{Id} - z\mathcal{M})$ can be defined through [12]. Inspired by Kitaev's unpublished note, Baillif (2004) proved the following Milnor–Thurston formula:

$$\det{}^\#(\mathrm{Id} - z\mathcal{M}) = \prod_{k=0}^{d-1} \det{}^\flat(\mathrm{Id} + \mathcal{D}_k(z))^{(-1)^{k+1}} \qquad [16]$$

Here, the $\mathcal{D}_k(z)$ are kernel operators acting on $(k+1)$-forms, constructed with the resolvent $(\mathrm{Id} - z\mathcal{M}_k)^{-1}$, together with a convolution operator $\mathcal{S}_k$, mapping

$(k+1)$-forms to $k$-forms and which satisfies the homotopy equation $\mathrm{d}\mathcal{S}+\mathcal{S}\mathrm{d}=1$. The kernel $\sigma_k(x,y)$ of $\mathcal{S}_k$ has singularities of the form $(x-y)/\|x-y\|^d$. The transversality assumption allows Baillif to interpret the determinant obtained by integrating the kernels along the diagonal as a flat determinant in the sense of Atiyah and Bott, whence the notation $\det^\flat$ in the right-hand side of [16].

Baillif (2004) did not give a spectral interpretation of zeros or poles of the sharp determinant [16], but he noticed that for $|z|$ very small, suitably high iterates of the $\mathcal{D}_k(z)$ are trace-class on $L^2(\mathbb{R}^d)$, showing that the corresponding regularized determinant has a nonzero radius of convergence under weak assumptions. The spectral interpretation of the sharp determinant [12] in arbitrary dimension, but under additional assumptions, was subsequently carried out by Baillif and the author of the present article, giving a new proof of some of the results in Ruelle (1990).

*See also:* Chaos and Attractors; Dynamical Systems and Thermodynamics; Ergodic Theory; Hyperbolic Dynamical Systems; Number Theory in Physics; Quantum Ergodicity and Mixing of Eigenfunctions; Quillen Determinant; Semi-Classical Spectra and Closed Orbits; Spectral Theory for Linear Operators.

## Further Reading


Baillif M (2004) Kneading operators, sharp determinants, and weighted Lefschetz zeta functions in higher dimensions. *Duke Mathematical Journal* 124: 145–175.

Baladi V (1998) *Periodic Orbits and Dynamical Spectra, Ergodic Theory Dynam. Systems*, vol. 18, pp. 255–292 (with an addendum by Dolgopyat D and Pollicott M, pp. 293–301.)

Baladi V and Ruelle D (1996) Sharp determinants. *Inventiones Mathematicae* 123: 553–574.

Chang CH and Mayer DH (2001) An extension of the thermodynamic formalism approach to Selberg's zeta function for general modular groups. In: Fiedler B (ed.) *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems*, pp. 523–562. Springer: Berlin.

Cvitanović P, Artuso R, Mainieri R, Tanner G, and Vattay G (2005) *Chaos: Classical and Quantum*, ChaosBook.org. Copenhagen: Niels Bohr Institute.

Elizalde E (1995) Ten Physical Applications of Spectral Zeta Functions, *Lecture Notes in Physics*, New Series m:35. Springer: Berlin.

Fried D (1986) The zeta functions of Ruelle and Selberg. I. *Ann. Sci. École Norm. Sup* 19: 491–517.

Fried D (1986) Analytic torsion and closed geodesics on hyperbolic manifolds. *Inventiones Mathematicae* 84: 523–540.

Fried D (1995) Meromorphic zeta functions for analytic flows. *Communications in Mathematical Physics* 174: 161–190.

Gohberg I, Goldberg S, and Krupnik N (2000) *Traces and Determinants of Linear Operators*. Basel: Birkhäuser.

Guillopé L, Lin K, and Zworski M (2004) The Selberg zeta function for convex co-compact Schottky groups. *Communications in Mathematical Physics* 245: 149–176.

Kitaev AY (1999) Fredholm determinants for hyperbolic diffeomorphisms of finite smoothness. *Nonlinearity* 12: 141–179.

Liverani C (2005) Fredholm determinants, Anosov maps and Ruelle resonances. *Discrete and Continuous Dynamical Systems* 13: 1203–1215.

Pollicott M (2001) Dynamical zeta functions. In: Katok A, de la Llave R, Pesin Y, and Weiss H (eds.) *Smooth Ergodic Theory and Its Applications (Seattle, WA, 1999)*, Proc. Sympos. Pure Math., vol. 69, pp. 409–427. Providence, RI: American Mathematical Society.

Ruelle D (1976) Zeta functions for expanding maps and Anosov flows. *Inventiones Mathematicae* 34: 231–242.

Ruelle D (1990) *An Extension of the Theory of Fredholm Determinants, Inst. Hautes Études Sci. Publ. Math.* 175–193.

Ruelle D (2002) *Dynamical Zeta Functions and Transfer Operators, Notices American Mathematical Society*: 887–895.

Rugh HH (1996) *Generalized Fredholm Determinants and Selberg Zeta Functions for Axiom A Dynamical Systems, Ergodic Theory Dynam. Systems*. 805–819.

Rugh HH (1999) Intermittency and regularized Fredholm determinants. *Inventiones Mathematicae* 135: 1–25.

Voros A (1987) Spectral functions, special functions and the Selberg zeta function. *Communications in Mathematical Physics* 110: 439–465.


# Relativistic Wave Equations Including Higher Spin Fields

**R Illge and V Wünsch**, Friedrich-Schiller-Universität Jena, Jena, Germany



## Introduction

The description of phenomena at high energies requires the investigation of relativistic wave equations, that is, equations which are invariant under Lorentz transformations. Our discussion will be given classically (i.e., nonquantum). A classification of the wave equations may be based on the spin of the particles (or physical fields), which was discovered for the electron by Goudsmith and Uhlenbeck in 1925. For the greater part of physics, the three spin numbers $s=0, 1/2$, and 1 are sufficient; the respective equations named after their discoverers Klein–Gordon, Dirac, and Proca for massive fields and D'Alembert, Weyl, and Maxwell for massless fields, respectively (see the following section).

In their original form, these equations look rather different. However, their translation into spinor form shows that the wave equations for bosons and fermions

have the same structure, if $s > 0$. Therefore, most of the equations dealt with in this article are formulated for spinor fields. (Strictly speaking, the exclusive use of 2-spinors restricts the relativistic invariance to the proper Lorentz group $SO^+(1,3)$. However, all the results presented here can be "translated back" into tensor or bispinor form, respectively (Illge 1993).) Relativistic wave equations for free fields with arbitrary spin $s > 0$ in Minkowski spacetime are discussed in the section "Higher spin in Minkowski spacetime"; they were first given by Dirac (1936).

In the subsequent section, we explain how the field theory can be extended to curved spacetimes. If a Lagrangian is known, then there exists a well-known mathematical procedure ("Lagrange formalism") to obtain the field equations, the energy–momentum tensor, etc. All field equations for "low" spin $s \leq 1$ arise from an action principle. Consequently, they can be extended to curved spacetime by simply replacing the flat metric and connection with their curved versions.

If $s > 1$, then the wave equations do not follow from a variation principle without supplementary conditions. Nevertheless, one can try to generalize the equations of the section "Higher spin in Minkowski spacetime" to curved spacetime by the "principle of minimal coupling," too. However, the arising equations are not satisfactory, since there is an algebraic consistency condition in curved space if $s > 1$ (Buchdahl 1962), and another for charged fields in the presence of electromagnetism if $s > 1/2$ (Fierz and Pauli 1939).

There have been numerous attempts to avoid these inconsistencies. As a rule, the alternative theories require an extended spacetime structure or additional new fields or they give up some important principle. An extensive literature is devoted to just this problem – unfortunately, a survey article or book is missing.

Finally, we present a possibility to describe fields with arbitrary spin $s > 0$ within the framework of Einstein's general relativity without any auxiliary fields and subsidiary conditions in a uniform manner. The approach is based on irreducible representations of type $D(s,0)$ and $D(s-1/2,1/2)$ instead of $D(s/2,s/2)$ in the Fierz theory for bosons and $D(s/2+1/4,s/2-1/4)$ in the Rarita–Schwinger theory for fermions. It was first pointed out by Buchdahl (1982) that this type of field equations can be generalized to a curved spacetime if the mass is positive. After a short time Wünsch (1985) simplified them to their final form:

$$\begin{aligned} \nabla^A_{P'}\varphi_{AB\ldots E} + m_1\chi_{B\ldots EP'} &= 0 \\ \nabla^{P'}_{(A}\chi_{B\ldots E)P'} - m_2\varphi_{AB\ldots E} &= 0 \end{aligned} \qquad [1]$$

This system contains the well-known wave equations for low spin $s = 1/2$ and $s = 1$ as special cases. By iteration we obtain second-order wave equations of normal hyperbolic type. Further, Cauchy's initial-value problem is well posed and a Lagrangian is known. For zero mass, we state the wave equations

$$\nabla^A_{(A'}\Theta_{|A|B'\ldots E')} = 0 \qquad [2]$$

which are just the curved versions of the equations for the potential of a massless field. They are consistent in curved spacetime, too, and the Cauchy problem is well posed (Illge 1988).

Last but not least, let us mention the esthetic aspect. Equations [1] and [2] satisfy Dirac's demand: "Physical laws should have mathematical beauty."

In the following, we assume that the spacetime and all the spinor and tensor fields are of class $C^\infty$. All considerations are purely local. We will call a symmetric ("irreducible") spinor to be of type $(n,k)$ if and only if it has $n$ unprimed and $k$ primed indices (irrespective of their position). Moreover, we use the notations and conventions of Penrose and Rindler (1984), especially for the curvature spinors $\Psi_{ABCD}$ and $\Phi_{ABA'B'}$.

## Wave Equations for Low Spin in Minkowski Spacetime

The spin (or intrinsic angular momentum) of a particle is found to be quantized. Its projection on any fixed direction is an integer or half-integer multiple of Planck's constant $\hbar$; the only possible values are

$$-s\hbar,\ (-s+1)\hbar,\ \ldots,\ (s-1)\hbar,\ s\hbar$$

The spin quantum number $s$ so defined can have one of the values $s = 0, 1/2, 1, 3/2, 2, \ldots$ and is a characteristic for all elementary particles along with their mass $m$ and electric charge $e$. The particles with integer $s$ are called "bosons," those with half-integer $s$ "fermions." The three numbers $s = 0, 1/2$, and 1 are referred to as "low" spin; they are sufficient for the greater part of physics.

The principle of first quantization associates a type of field and a field equation to each type of elementary particles. Massive particles, with rest mass $m > 0$, and massless particles, with rest mass $m = 0$, are to be distinguished. Accordingly, we obtain six linear wave equations for $s \leq 1$, which read as follows in units such that $c = \hbar = 1$ (see Table 1):

For the sake of simplicity, we consider only free fields in Table 1; no source terms or interaction terms appear here. The associated "free" Lagrangians are given in Table 2.

Since the electromagnetic field tensor $F_{ab}$ satisfies the first part of Maxwell's equations $\partial_{[c}F_{ab]} = 0$, it follows

**Table 1** Relativistic wave equations for low spin $s = 0, 1/2$, and 1

| *Spin, mass* | *Wave equation* | *Associated particles* |
|---|---|---|
| $s = 0,\ m > 0$ | Klein–Gordon eqn. $(\approx + m^2)u = 0$ | Scalar mesons $\pi, \eta, K, \ldots$ |
| $s = 0,\ m = 0$ | D'Alembert eqn. $\approx u = 0$ | – |
| $s = 1/2,\ m > 0$ | Dirac eqn. $\partial^A_{A'}\varphi_A + \frac{im}{\sqrt{2}}\chi_{A'} = 0$ $\partial^{A'}_A\chi_{A'} - \frac{im}{\sqrt{2}}\varphi_A = 0$ | Leptons $e, \mu, \tau$ Baryons $p, n, \Lambda, \Xi, \Sigma, \ldots$ |
| $s = 1/2,\ m = 0$ | Weyl eqn. $\partial^A_{A'}\nu_A = 0$ | Massless(?) neutrinos $\nu_e, \nu_\mu, \nu_\tau$ |
| $s = 1,\ m > 0$ | Proca eqn. $H_{ab} = \partial_a U_b - \partial_b U_a$ $\partial^c H_{ca} + m^2 U_a = 0$ | Vector mesons $\rho, \omega, \psi, \Phi, \ldots$ |
| $s = 1,\ m = 0$ | Maxwell eqn. $\partial_{[a}F_{bc]} = 0$ $\partial_a F^{ab} = 0$ | Photon $\gamma$ |

**Table 2** The Lagrangian densities for free (i.e., noninteracting) fields with low spin

| *Field* | *Lagrangian density* |
|---|---|
| Scalar field | $\mathfrak{L} = \frac{1}{2}\{(\partial^a u)(\partial_a u) - m^2 u^2\}$ |
| Dirac field | $\mathfrak{L} = \frac{i}{\sqrt{2}}(\bar{\chi}_A\partial^{AA'}\chi_{A'} + \bar{\varphi}^{B'}\partial_{BB'}\varphi^B - \varphi^B\partial_{BB'}\bar{\varphi}^{B'} - \chi_{A'}\partial^{AA'}\bar{\chi}_A) + m(\bar{\chi}_A\varphi^A + \bar{\varphi}^{A'}\chi_{A'})$ |
| Weyl field | $\mathfrak{L} = \frac{i}{\sqrt{2}}(\bar{\nu}_{A'}\partial^{AA'}\nu_A - \nu_A\partial^{AA'}\bar{\nu}_{A'})$ |
| Proca field | $\mathfrak{L} = \frac{1}{4}H_{ab}H^{ab} - H^{ab}\partial_{[a}U_{b]} + \frac{m^2}{2}U_a U^a$ |
| Maxwell field | $\mathfrak{L} = -\frac{1}{4}F_{ab}F^{ab} = -(\partial_{[a}A_{b]})(\partial^{[a}A^{b]})$ |

that a vector field $A_a$ exists such that $F_{ab} = \partial_a A_b - \partial_b A_a$. This vector field is called the "electromagnetic 4-potential." It is not uniquely determined by the field $F_{ab}$; the freedom in $A_a$ is $A_a \to A_a + \partial_a\tau$ where $\tau = \tau(x)$ is a real-valued function. This gauge transformation of $A_a$ can be used, for example, to obtain the Lorentz gauge condition $\partial^a A_a = 0$.

The wave equations listed in Table 1 look rather different, but this formal disadvantage can be overcome. To begin with, we remark that fermions require spinors for their description. The Dirac and Weyl equations are not describable by linear equations for tensor fields. On the other hand, bosons can be described by spinors as well. All tensor equations can be "translated" into spinor form using the mixed spinor–tensor $\sigma^a_{AA'}$. We will demonstrate this procedure for the Proca field in some detail.

The (possibly complex) skew-symmetric tensor $H_{ab}$ and the vector $U_a$ have the spinor equivalents

$$H_{ab}\sigma^a_{AA'}\sigma^b_{BB'} = \varphi_{AB}\varepsilon_{A'B'} + \xi_{A'B'}\varepsilon_{AB}$$
$$U_a\sigma^a_{AA'} = \chi_{AA'}$$

where $\varphi$ and $\xi$ are both symmetric spinors: $\varphi_{AB} = \varphi_{(AB)}$, $\xi_{A'B'} = \xi_{(A'B')}$. After a straightforward calculation the Proca equation yields

$$\partial^{C'}_{(A}\chi_{B)C'} + \varphi_{AB} = 0, \qquad \partial^{C'}_{(A'}\chi_{B')C} + \xi_{A'B'} = 0$$
$$\partial^C_{A'}\varphi_{CA} + \partial^{C'}_A\xi_{C'A'} + m^2\chi_{AA'} = 0$$

Further, from the equation $\partial_{[c}H_{ab]} = 0$, we obtain $\partial^{C'}_A\xi_{A'C'} = \partial^C_{A'}\varphi_{AC}$; thus, the first and second summand in the third equation are equal. Consequently, we find the following spinor form of the Proca equations:

$$\partial^C_{A'}\varphi_{CA} + \frac{m^2}{2}\chi_{AA'} = 0, \qquad \partial^{C'}_{(A}\chi_{B)C'} + \varphi_{AB} = 0$$
$$\partial^{C'}_A\xi_{C'A'} + \frac{m^2}{2}\chi_{AA'} = 0, \qquad \partial^{C'}_{(A'}\chi_{B')C} + \xi_{A'B'} = 0 \qquad [3]$$

If the tensor fields $H$ and $U$ are real, then we have $\xi_{A'B'} = \bar{\varphi}_{A'B'}$, $\chi_{AA'} = \bar{\chi}_{AA'}$, and the second pair of equations is just the complex conjugate of the first.

Now it is readily seen that the Dirac and Proca equations have the same structure. They are coupled first-order systems of differential equations for pairs of spinor fields. The only decisive difference is that the spinors have one index if $s = 1/2$ and two indices if $s = 1$.

We obtain a similar result for Maxwell fields. The real tensor $F_{ab}$ has the spinor equivalent

$$F_{ab}\sigma^a_{AA'}\sigma^b_{BB'} = \varphi_{AB}\varepsilon_{A'B'} + \bar{\varphi}_{A'B'}\varepsilon_{AB}$$

with a symmetric spinor $\varphi_{AB}$. The spinor form of Maxwell's equations is (Penrose and Rindler 1984)

$$\partial^A_{A'}\varphi_{AB} = 0 \qquad [4]$$

and has the same structure as the Weyl equation.

Here we found an example for the power and utility of spinor techniques since they allow the formulation of the wave equations for bosons and fermions in a uniform manner. Only the cases $m > 0$ and $m = 0$ are to be distinguished. Moreover, the above results suggest the way for generalizing the wave equations to higher spin. Therefore, we can already end the discussion of the fields with low spin and take them as special cases of those with arbitrary spin.

## Higher Spin in Minkowski Spacetime

### Massive Fields

Relativistic wave equations for particles with arbitrary spin were first considered by Dirac (1936). His equations read

$$\partial^A_{P'}\varphi_{AB\ldots DQ'\ldots T'} + m_1\chi_{B\ldots DP'Q'\ldots T'} = 0$$
$$\partial^{P'}_A\chi_{B\ldots DP'Q'\ldots T'} - m_2\varphi_{AB\ldots DQ'\ldots T'} = 0 \qquad [5]$$

where the spinors $\varphi$ and $\chi$ are of type $(n,k)$ and $(n-1,k+1)$, respectively (corresponding to irreducible representations of the restricted Lorentz group $SO^{+}(1,3)$). The constants $m_1$ and $m_2$ are mass parameters ($m^2 = -2m_1m_2$) and the spin $s$ is one half of the total number of indices of each spinor, $s = (1/2)(n+k)$. As in the preceding section, we assume that electromagnetism and other interactions are absent. We should mention that equations for higher spin were not motivated by observations or empirical facts in that period of time, because only a few elementary particles were known (proton, neutron, electron, positron, and photon), and all of them have low spin (see **Table 1**). Since that time, particles with $s > 1$ were found in nature, for example, resonances in scattering experiments.

The system [5] allows a uniform description of free fields with arbitrary spin $s > 0$, including Dirac and Proca fields, as we know from the preceding section. (Remark: The symmetrization in eqns [3] can be omitted since the vector field $U$ is divergence-free as a consequence of the second Proca equation.) Various other field equations proposed subsequently can be comprehended as its special cases (Corson 1953). Examples are the Rarita–Schwinger equations for fermions: if they are written in terms of 2-spinors, then one obtains just the system [5] where the spinor $\varphi$ is of type $(s+1/2,\ s-1/2)$ and the spinor $\chi$ is of type $(s-1/2,\ s+1/2)$.

If we apply $\partial_E^{P'}$ to the first of the equations in [5] and use the second, we obtain

$$(\Box + m^2)\varphi_{AB\ldots DQ'\ldots T'} = 0 \qquad [6a]$$

since the second derivatives commute in flat spacetimes. Similarly,

$$(\Box + m^2)\chi_{B\ldots DP'Q'\ldots T'} = 0 \qquad [6b]$$

so both fields $\varphi$ and $\chi$ satisfy a Klein–Gordon type equation. Moreover, eqns [5] imply that each of $\varphi$ and $\chi$ is divergence-free

$$\partial^{AQ'}\varphi_{AB\ldots DQ'\ldots T'} = 0 = \partial^{BP'}\chi_{B\ldots DP'Q'\ldots T'} \qquad [7]$$

if they have at least one index of each kind.

In a sense, this procedure can be reversed. Let a symmetric spinor field $\varphi$ be given that satisfies [6a] and [7]. (Remark: A significant example is the Fierz system

$$(\Box + m^2)U_{ab\ldots d} = 0, \qquad \partial^a U_{ab\ldots d} = 0$$

for a symmetric, tracefree tensor field $U$, since the spinor equivalent of $U$ is of type $(k,k)$.)

Define

$$\chi_{B\ldots DP'Q'\ldots T'} := \partial^A_{P'}\varphi_{AB\ldots DQ'\ldots T'}$$

Then $\chi$ is symmetric in all its indices since $\varphi$ is divergence-free. Further, we obtain

$$\begin{aligned}\partial_E^{P'}\chi_{B\ldots DP'Q'\ldots T'} &= \partial_E^{P'}\partial^A_{P'}\varphi_{AB\ldots DQ'\ldots T'} \\ &\equiv -\frac{1}{2}\Box\varphi_{EB\ldots DQ'\ldots T'} \\ &= \frac{m^2}{2}\varphi_{EB\ldots DQ'\ldots T'}\end{aligned}$$

since $\varphi$ satisfies the Klein–Gordon equation [6a]. Consequently, the pair $(\varphi,\chi)$ satisfies a system [5]. Obviously, this procedure can be continued: define

$$\eta_{C\ldots DO'P'Q'\ldots T'} := \partial^B_{O'}\chi_{B\ldots DP'Q'\ldots T'}$$

etc. We obtain a sequence of spinors of type $(0,2s)$, $(1,2s-1), \ldots, (2s,0)$ each of which is obtainable from its immediate neighbors by a differentiation contracted on one index. Together, these spinors form an invariant exact set (Penrose and Rindler 1984).

The just given arguments show that there is an ambiguity in the system [5]. The spin $s$ fixes only the total number of indices of $\varphi$ and $\chi$. However, their partition into primed and unprimed ones is not *a priori* fixed. Therefore, we can choose a "convenient" partition for the respective needs.

### Massless Fields

If $m=0$, then the Dirac system [5] is decoupled. Therefore, we have to state a single equation for a single field. Let $\varphi$ be a spinor field of type $(n,0)$. The massless free-field equation for spin $(1/2)n$ is then taken to be

$$\partial^A_{A'}\varphi_{AB\ldots E} = 0 \qquad [8]$$

More precisely, the solutions of [8] represent left-handed massless particles with helicity $-(1/2)n\hbar$, whereas the solutions of the complex-conjugate form of this equation are right-handed particles (helicity $+(1/2)n\hbar$). Recall that the Weyl equation ($n=1$) and the source-free Maxwell equation ($n=2$) have this form. (Remark: The Bianchi identity in Einstein spaces also falls in this category, with the Weyl spinor $\Psi_{ABCD}$ taking the place of $\varphi\ldots$. Moreover, we may think of [8] with $n=4$ as the gauge-invariant equation for the weak vacuum gravitational field.)

The massless field equation [8] can be solved using methods of twistor geometry. Moreover, there is an explicit integral formula for representing massless free fields in terms of arbitrarily chosen null data on a light cone (Penrose and Rindler 1984, 1986, Ward and Wells 1990). We do not discuss eqns [8] in detail since they are generally inconsistent in curved spacetimes if $n > 2$ (see the next

section). We only indicate that each solution of [8] satisfies the second-order wave equation

$$\Box\varphi_{AB\ldots E}=0$$

Maxwell's equations imply the existence of an electromagnetic potential (cf. section "Wave equations for low spin in Minkowski spacetime"). This concept can be generalized to higher spin. A "potential" for a spinor field $\varphi_{AB\ldots E}$ of type $(n,0)$ is a spinor field $\Theta_{AB'\ldots E'}$ of type $(1,n-1)$ such that

$$\partial^{A}_{(A'}\Theta_{|A|B'\ldots E')}=0 \qquad [9]$$

and

$$\varphi_{AB\ldots E}=\partial^{B'}_{(B}\cdots\partial^{E'}_{E}\Theta_{A)B'\ldots E'} \qquad [10]$$

One can check in a straightforward manner that a spinor field $\varphi$ that is given by [9] and [10] satisfies the massless equation [8]. If $n>1$, there is a gauge freedom in these potentials; it turns out to be

$$\Theta_{AB'\ldots E'}\to\Theta_{AB'\ldots E'}+\partial_{A(B'}\omega_{C'\ldots E')}$$

for any spinor field $\omega$ of type $(0,n-2)$. Furthermore, the general massless field $\varphi$ can locally be expressed in this way (Penrose and Rindler 1986).

## Wave Equations in Curved Spacetimes, Consistency Conditions

First of all we emphasize that Hamilton's principle of stationary action is extremely important in field theories (see, e.g., Schmutzer (1968)). Assume that the Lagrangian $\mathfrak{L}$ contains at most first derivatives of a field $\psi_\Sigma:\mathfrak{L}=\mathfrak{L}(\psi_\Sigma(x),\partial_a\psi_\Sigma(x))$. "Special relativity" states that $\mathfrak{L}$ is invariant under Lorentz transformations. The Euler–Lagrange equations with respect to variation of $\psi_\Sigma$ read

$$\frac{\partial\mathfrak{L}}{\partial\psi_\Sigma}-\partial_a\frac{\partial\mathfrak{L}}{\partial(\partial_a\psi_\Sigma)}=0 \qquad [11]$$

and these are the field equations that $\psi_\Sigma$ is required to satisfy.

In "general relativity," the Lagrangian $\mathfrak{L}$ has to be generally covariant. So we have $\mathfrak{L}=\mathfrak{L}(\psi_\Sigma(x),\nabla_a\psi_\Sigma(x))$ and the Euler–Lagrange equations

$$\frac{\partial\mathfrak{L}}{\partial\psi_\Sigma}-\nabla_a\frac{\partial\mathfrak{L}}{\partial(\nabla_a\psi_\Sigma)}=0 \qquad [12]$$

emerge. If we assume that the Lagrangian $\mathfrak{L}$ does not contain the curvature tensors and their derivatives explicitly and compare [11] and [12], then it is easily seen how the wave equations in curved spacetime can be obtained: by simply replacing the flat metric and connection with their curved versions. This procedure is called the "principle of minimal coupling."

All equations for low spin in Minkowski spacetime are the Euler–Lagrange equations of a variation principle (see Table 2). Consequently, they can be extended to curved spacetime by simply using the principle of minimal coupling. The arising equations are perfectly acceptable. No complications arise, and so we do not repeat them in this section.

If $s>1$, then neither the massive nor the massless wave equations follow from a variation principle without supplementary conditions. Nevertheless, we can try to generalize the equations of the previous section to a curved spacetime by formally replacing the flat metric and connection with their curved versions, too. However, serious problems arise:

Let us first consider massless fields of helicity $-(1/2)n\hbar$. The principle of minimal coupling yields

$$\nabla^{A}_{A'}\varphi_{AB\ldots E}=0 \qquad [13]$$

If we apply $\nabla^{A'}_{F}$ to this equation, we obtain

$$\nabla^{A'}_{F}\nabla^{A}_{A'}\varphi_{AB\ldots E}=0$$

Since the covariant derivatives do not commute with each other, the term on the left-hand side is not completely symmetric in the unprimed indices. Therefore, this equation can be decomposed into two nontrivial irreducible parts if $n>1$: symmetrization yields the covariant D'Alembert equation

$$\nabla^a\nabla_a\varphi_{B\ldots EF}=0$$

as required, while antisymmetrization yields by use of the spinor Ricci identities

$$(n-2)\Psi^{KLM}{}_{(C}\varphi_{D\ldots E)KLM}=0 \qquad [14]$$

where $\Psi_{ABCD}$ is the Weyl spinor. If $n>2$ and the spacetime is not conformally flat, then this algebraic consistency condition effectively renders eqn [13] useless as physical field equations.

If $m>0$, the situation is not better. In somewhat similar way, we obtain the algebraic consistency conditions

$$\begin{aligned}&(n-2)\Psi^{KLM}{}_{(C}\varphi_{D\ldots E)KLMQ'P'\ldots T'}\\&\quad+k\Phi^{KLX'}{}_{(Q'}\varphi_{|KLC\ldots E|P'\ldots T')X'}=0\quad(n>1)\\&(k-1)\bar{\Psi}^{X'Y'Z'}{}_{(S'}\chi_{|B\ldots DX'Y'Z'|T'\ldots U')}\\&\quad+(n-1)\Phi_{(B}{}^{KX'Y'}\chi_{C\ldots D)KX'Y'S'T'\ldots U'}=0\quad(k>0)\end{aligned} \qquad [15]$$

if the spinor field $\varphi$ is of type $(n,k)$ (Buchdahl 1962).

We remark that similar consistency conditions occur if we have no gravitation, but an interaction with an electromagnetic field. Then the partial

derivative is to be replaced by $D_a = \partial_a - ieA_a$ and we obtain consistency conditions like [14] and [15], where the curvature spinors are to be replaced by the electromagnetic spinor (Fierz and Pauli 1939).

So far one is left with the problem: "Find the 'correct' laws for arbitrary spin, that means field equations which coincide with the well-known approved ones for low spin and which remain consistent even for higher spin when electromagnetism and/or gravitation is coupled!"

An extensive literature is devoted to just this problem. Let us briefly sketch some means by which the authors tried to solve it:

- derivation of the desired field equations from a variation principle where the original spinor fields are supplemented by auxiliary fields;
- extension of the four-dimensional spacetime geometry to a richer one: higher number of dimensions, complexification, addition of torsion, nonmetrical connection, ...;
- replacement of the algebra of spinors by some richer algebra;
- disclaim of the principle of minimal coupling; and
- supergravity theories.

Some of these attempts are able to solve the problem, at least partially. But, as a rule, they pay a price of new difficulties. In the next section, we offer "good" equations for arbitrary $s > 0$ within the conventional framework of the minimal coupling principle and of a curved spacetime background.

## Wave Equations for Arbitrary Spin without Consistency Conditions

### Massive Fields

The ansatz which leads to the desired result is surprisingly simple. We avoid the ambiguity in the Dirac system [5] that has been discussed earlier as well as any consistency condition if we state the wave equations

$$\begin{aligned} \nabla^A_{P'}\varphi_{AB\ldots E} + m_1\chi_{B\ldots EP'} &= 0 \\ \nabla^{P'}_{(A}\chi_{B\ldots E)P'} - m_2\varphi_{AB\ldots E} &= 0 \end{aligned} \qquad [16]$$

This system was first proposed by Wünsch (1985); it is equivalent to a pair of equations given by Buchdahl (1982) which contains the Weyl spinor explicitly. As before, $\varphi$ and $\chi$ are symmetric spinor fields, $\varphi$ has $n$ unprimed indices (and no one else!) and the constants $m_1, m_2$ are mass parameters ($m^2 = -2m_1m_2$). We assume $m_1 \neq 0$ in this section. Obviously, the Dirac and Proca equations are special cases of [16], choose $n=1$ and $n=2$, respectively. (Remark: An electromagnetic field can be included in [16] by $\nabla_a \to D_a = \nabla_a - ieA_a$, and the equations remain consistent (Illge 1993).)

First of all, we remark that eqns [16] are the Euler–Lagrange equations of an action principle. The existence of a Lagrangian is plausible since the number of equations and the number of degrees of freedom are equal. We do not state the Lagrangian, the energy–momentum tensor, and the current vector in this article and refer the reader to Illge (1993).

If $n > 1$, we can apply $\nabla^{BP'}$ to the first equation of [16] and obtain using the spinor Ricci identities:

$$\begin{aligned} \nabla^{BP'}\chi_{BC\ldots EP'} &= -\frac{1}{m_1}\nabla^{BP'}\nabla^A_{P'}\varphi_{ABC\ldots E} \\ &= -\frac{n-2}{m_1}\Psi^{KLM}{}_{(C}\varphi_{D\ldots E)KLM} \end{aligned} \qquad [17]$$

Hence the divergence of $\chi$ vanishes if $n=2$ or if the spacetime is conformally flat. These are exactly the cases where the symmetrization in the second equation of [16] can be omitted.

Now we are going to derive the second-order equations for $\varphi$ and $\chi$. Substituting

$$\chi_{BC\ldots EP'} = -\frac{1}{m_1}\nabla^A_{P'}\varphi_{AB\ldots E} \qquad [18]$$

into the second equation of [16], we obtain, after a bit of algebra,

$$\begin{aligned} &\nabla^a\nabla_a\varphi_{AB\ldots E} - 2(n-1)\Psi^{KL}{}_{(AB}\varphi_{C\ldots E)KL} \\ &\quad + \left(\frac{n+2}{12}R + m^2\right)\varphi_{AB\ldots E} = 0 \end{aligned} \qquad [19]$$

This is a linear second-order equation of normal hyperbolic type for the spinor field $\varphi$. It can be used to solve Cauchy's problem for the system [16].

Similarly, we get a second-order equation for $\chi$:

$$\begin{aligned} &\nabla^a\nabla_a\chi_{B\ldots EP'} - 2(n-1)\Phi_{(B}{}^K{}_{P'}{}^{W'}\chi_{C\ldots E)KW'} \\ &\quad + \left(\frac{R}{4} + m^2\right)\chi_{B\ldots EP'} \\ &= 2\frac{n-1}{n}\nabla_{(BP'}\nabla^{KW'}\chi_{C\ldots E)KW'} \end{aligned} \qquad [20]$$

Seemingly this is not an equation of hyperbolic type if $n > 1$. However, the second derivatives of $\chi$ on the right-hand side of [20] can be eliminated using [17]. Therefore, if the spinor field $\varphi$ is already known by solving [19], then [20] is an equation of Klein–Gordon type, too. However, it is generally inhomogeneous if $n > 2$. A wave equation that contains the spinor field $\chi$ alone exists only if $n=1, n=2$, or the spacetime is conformally flat.

Now we are going to discuss the "Cauchy problem" for the wave equations [16] (for details see Wünsch (1985)). Let a spacelike hypersurface $S$ be given and let $n^a$ denote the future-directed unit normal vector on $S$ and $\nabla_n = n^a \nabla_a$. The local Cauchy problem is to find a solution $(\varphi, \chi)$ of [16] with given Cauchy data $\varphi^0, \chi^0$ on $S$.

In general, the initial data $\varphi^0$ and $\chi^0$ cannot be prescribed arbitrarily. Suppose that a solution $(\varphi, \chi)$ of [16] does exist. Then the differential equations have to be satisfied on $S$, too. Thus, we obtain

$$(\nabla_n \varphi_{AB\ldots E})_{|S} = 2n_A^{A'} \left(\tilde{\nabla}^F_{A'} \varphi_{B\ldots EF} + m_1 \chi_{B\ldots EA'}\right)_{|S} \quad [21]$$

where the differential operator $\tilde{\nabla}_{AA'} = \nabla_{AA'} - n_{AA'} \nabla_n$ is just the tangential part of $\nabla_{AA'}$ with respect to $S$. Therefore, the right-hand side of [21] is completely determined by the initial data. Now the symmetry of the solution $\varphi_{AB\ldots E}$ implies the symmetry of $\nabla_n \varphi_{AB\ldots E}$. Consequently, the right-hand side of [21] has to be symmetric with respect to the unprimed indices and so we obtain the following constraints for the initial data if $\varphi$ has at least two indices:

$$n^{BA'} \left(\tilde{\nabla}^F_{A'} \varphi^0{}_{B\ldots EF} + m_1 \chi^0{}_{B\ldots EA'}\right)_{|S} = 0 \quad [22]$$

Now we can state:

**Theorem 1** *If the Cauchy data $\varphi^0$ and $\chi^0$ satisfy the constraints [22], then the Cauchy problem has a unique solution in a neighborhood of $S$.*

For each differential equation of hyperbolic type we can ask the question whether the wave propagation is "sharp," that is, free of tails. If this property is valid we say that the equation satisfies "Huygens' principle" (for an exact definition, see, e.g., Wünsch (1994)). Using invariant Taylor expansions of the parallel propagator and of the Riesz kernels in normal coordinates we can prove (Wünsch 1985):

**Theorem 2** *The massive wave equations [16] for spin $s > 0$ satisfy Huygens' principle if and only if the spacetime is of constant curvature and $R = -(6m^2/s)$.*

### Massless Fields

In the preceding section, we have seen that the premise $m_1 \neq 0$ is decisive for the consistency of [16] if $s > 1$. This fact agrees with the result of the previous section, that eqn [13] is inconsistent if $s > 1$ and the spacetime is not conformally flat. On the other hand, $m_2 = 0$ is possible. Therefore we state the wave equations

$$\nabla^A_{(A'} \Theta_{|A|B'\ldots E')} = 0 \quad [23]$$

for a spinor field $\Theta$ of type $(1, n-1)$. This is just eqn [9] for the potential of a massless field. We will show that [23] is a satisfactory equation in a generally curved spacetime (Illge 1988). Unfortunately, no Lagrangian has been found if $n > 1$.

To begin with, we remark that there is a gauge freedom in curved spacetimes, too, since the solution $\Theta$ of [23] cannot be uniquely determined if $n > 1$. We use this freedom to prescribe the divergence of $\Theta$. So let an arbitrary spinor field $\omega$ of type $(0, n-2)$ be given. We consider eqns [23] and

$$\nabla^{AB'} \Theta_{AB'C'\ldots E'} = \omega_{C'\ldots E'}$$

or, together,

$$\nabla^A_{A'} \Theta_{AB'\ldots E'} = -\frac{n-1}{n} \varepsilon_{A'(B'} \omega_{C'\ldots E')} \quad [24]$$

If we apply $\nabla^{A'}_B$ to this equation, we obtain using the spinor Ricci identities

$$\nabla^a \nabla_a \Theta_{BB'\ldots E'} - 2(n-1) \Phi_B{}^K{}_{(B'}{}^{W'} \Theta_{|K|C'\ldots E')W'} + \frac{R}{4} \Theta_{BB'\ldots E'} = \frac{2(n-1)}{n} \nabla_{B(B'} \omega_{C'\ldots E')} \quad [25]$$

This is a linear second-order equation of normal hyperbolic type for the spinor field $\Theta$ (cf. [20]).

Now let us discuss some particular cases. If $n = 1$, then [23] is just the Weyl equation itself. Therefore, the equations for the field and its potential are identical and there is no gauge freedom. If $n = 2$, then the spinor field $\Theta_{AA'}$ is a (complex) vector field and eqn [23] yields

$$\nabla^A_{(A'} \Theta_{|A|B')} = 0$$

The gauge field $\omega$ is just a scalar function, especially we can choose $\omega = 0$ (Lorentz gauge). As in eqn [10] we define the field spinor as

$$\varphi_{AB} = \nabla^{B'}_{(A} \Theta_{B)B'}$$

Since we have the identity

$$\nabla^A_{B'} \nabla^{A'}_{(B} \eta_{A)A'} = \nabla^{A'}_B \nabla^A_{(B'} \eta_{|A|A')}$$

for arbitrary spinor fields $\eta_{AA'}$ (which must not have additional free indices!), the spinor field $\varphi_{AB}$ satisfies the massless free-field equation

$$\nabla^A_{B'} \varphi_{AB} = 0$$

If $n > 2$, we can define a field $\varphi_{AB\ldots E}$ via the relation [10], too, replacing the partial with the covariant derivatives. But the field equation for $\varphi_{AB\ldots E}$ becomes more complicated than [13]. This

fact is not surprising, since eqn [23] is a consistent one, whereas [13] is inconsistent.

We continue with some remarks on "conformal rescalings of the metric." The equations for massless fields have to be invariant with respect to such transformations. Therefore, the "curved space" scalar wave equation is

$$\left(\Box + \frac{R}{6}\right)\varphi = 0 \qquad [26]$$

Further, the equations

$$\nabla^{A}_{(A'}\eta_{|AB\ldots E|B'\ldots F')} = 0 \qquad [27]$$

for any spinor field $\eta$ of type $(n, k)$ are conformally invariant (Penrose and Rindler 1984). Especially, eqns [23] for the massless potential and [13] for the massless field have this property.

We mention a further special case of [27]. If $\eta$ is of type $(k+1, k)$, then these equations are consistent, too (Frauendiener and Sparling 1999). The Cauchy problem is well posed and a Lagrangian is known. Unfortunately, the solutions do not satisfy a wave equation of second order if $k > 0$.

We conclude with the discussion of the Cauchy problem for eqn [24]. As in the preceding section, let a spacelike hypersurface $S$ and initial data $\Theta^0$ on $S$ be given. We can state:

**Theorem 3** *If a symmetric spinor field $\omega$ of type $(0, n-2)$ is given, then there exists a neighborhood of $S$ in which eqn [24] has one and only one solution satisfying $\Theta_{|S} = \Theta^0$.*

The proof is given in Illge (1988). We emphasize that there are no constraints on the Cauchy data for the massless equation [24].

In contrast to massive fields we are far away from an answer to the question whether Huygens principle is valid for the massless equations. A particular result is Wünsch (1994):

**Theorem 4** *Huygen's principle for the conformally invariant scalar wave equation [26], the Weyl, and the Maxwell equations is valid only for conformally flat and plane wave metrics within the classes of centrally symmetric, recurrent, $(2,2)$-decomposable, Petrov type N, III or D spacetimes as well as those with $\nabla_{[a}R_{b]c} = 0$.*

*See also:* Clifford Algebras and Their Representations; Dirac Fields in Gravitation and Nonabelian Gauge Theory; Euclidean Field Theory; Evolution Equations: Linear and Nonlinear; Spinors and Spin Coefficients; Standard Model of Particle Physics; Twistors.

## Further Reading


Buchdahl HA (1962) On the compatibility of relativistic wave equations in Riemann spaces. *Nuovo Cimento* 25: 486–496.

Buchdahl HA (1982) On the compatibility of relativistic wave equations in Riemann spaces II. *Journal of Physics A* 15: 1–5.

Corson EM (1953) *Introduction to Tensors, Spinors, and Relativistic Wave-Equations*. London and Glasgow: Blackie and Son Ltd.

Dirac PAM (1936) Relativistic wave equations. *Proceedings of the Royal Society London Series A* 155: 447–459.

Fierz M and Pauli W (1939) On relativistic wave equations for particles of arbitrary spin in an electromagnetic field. *Proceedings of the Royal Society London Series A* 173: 211–232.

Frauendiener J and Sparling GAJ (1999) On a class of consistent higher spin equations on curved manifolds. *Journal of Geometry and Physics* 30: 54–101.

Greiner W (1997) *Relativistic Quantum Mechanics – Wave Equations*, 2nd edn. Berlin: Springer.

Illge R (1988) On potentials for several classes of spinor and tensor fields in curved spacetimes. *General Relativity and Gravitation* 20: 551–564.

Illge R (1993) Massive fields of arbitrary spin in curved spacetimes. *Communications in Mathematical Physics* 158: 433–457.

Penrose R and Rindler W (1984) *Spinors and Space-Time, Two-Spinor Calculus and Relativistic Fields,* vol. 1. Cambridge: Cambridge University Press.

Penrose R and Rindler W (1986) *Spinors and Space-Time, Spinor and Twistor Methods in Space-Time Geometry,* vol. 2. Cambridge: Cambridge University Press.

Schmutzer E (1968) *Relativistische Physik*. Leipzig: Teubner-Verlag.

Ward RS and Wells RO (1990) *Twistor Geometry and Field Theory*. Cambridge: Cambridge University Press.

Wünsch V (1985) Cauchy's problem and Huygens' principle for relativistic higher spin wave equations in an arbitrarily curved space-time. *General Relativity and Gravitation* 17: 15–38.

Wünsch V (1994) Moments and Huygens' principle for conformally invariant field equations in curved space-times. *Annales de l'Institute Henri Poincaré – Physique théorique* 60: 433–455.

# Renormalization: General Theory

**J C Collins**, Penn State University,
University Park, PA, USA



## Introduction

Quantum field theories (QFTs) provide a natural framework for quantum theories that obey the principles of special relativity. Among their most striking features are ultraviolet (UV) divergences, which at first sight invalidate the existence of the theories. The divergences arise from Fourier modes of very high wave number, and hence from the structure of the theories at very short distances. In the very restricted class of theories called "renormalizable," the divergences may be removed by a singular redefinition of the parameters of the theory. This is the process of renormalization that defines a QFT as a nontrivial limit of a theory with a UV cutoff.

A very important QFT is the standard model, an accurate and successful theory for all the known interactions except gravity. Calculations using renormalization and related methods are vital to the theory's success.

The basic idea of renormalization predates QFT. Suppose we treat an observed electron as a combination of a bare electron of mass $m_0$ and the associated classical electromagnetic field down to a radius $a$. The observed mass of the electron is its bare mass plus the energy in the field (divided by $c^2$). The field energy is substantial, for example, 0.7 MeV when $a = 10^{-15}$ m, and it diverges when $a \to 0$. The observed mass, 0.5 MeV, is the sum of the large (or infinite) field contribution compensated by a negative and large (or infinite) bare mass. This calculation needs replacing by a more correct version for short distances, of course, but it remains a good motivation.

In this article, we review the theory of renormalization in its classic form, as applied to weak-coupling perturbation theory, or Feynman graphs. It is this method, rather than the Wilsonian approach (*see* Exact Renormalization Group), that is typically used in practice for perturbative calculations in the standard model, especially its QCD part.

Much of the emphasis is on weak-coupling perturbation theory, where there are well-known algorithmic rules for performing calculations and renormalization. Applications (*see* Quantum Chromodynamics for some important nontrivial examples) involve further related results, such as the operator product expansion, factorization theorems, and the renormalization group (RG), to go far beyond simple fixed-order perturbation theory. The construction of fully rigorous mathematical treatments for the exact theory is a topic of future research.

## Formulation of QFT

A QFT is specified by its Lagrangian density. A simple example is $\phi^4$ theory:

$$\mathcal{L} \stackrel{?}{=} \frac{(\partial\phi)^2}{2} - \frac{m^2\phi^2}{2} - \frac{\lambda\phi^4}{4!} \qquad [1]$$

where $\phi(x) = \phi(t, \boldsymbol{x})$ is a single component Hermitian field. The Lagrangian density and the resulting equation of motion, $\partial^2\phi + m^2\phi + (1/6)\lambda\phi^3 = 0$, are local; they involve only products of fields at the same spacetime point. Such locality is characteristic of relativistic theories, where otherwise it is difficult or impossible to preserve causality, but it is also the source of the UV divergences. The question mark over the equality symbol in eqn [1] is a reminder that renormalization of UV divergences will force us to modify the equation.

The Feynman rules for perturbation theory are given by a free propagator $\mathrm{i}/(p^2 - m^2 + \mathrm{i}0)$ and an interaction vertex $-\mathrm{i}\lambda$. Although we will usually work in four spacetime dimensions, it is useful also to consider the theory in a general spacetime dimensionality $n$, where the coupling has energy dimension $[\lambda] = E^{4-n}$. We use "natural units," that is, with $\hbar = c = 1$. The "i0" in the propagator $\mathrm{i}/(p^2 - m^2 + \mathrm{i}0)$ symbolizes the location of the pole relative to the integration contour; it is often written as $\mathrm{i}\epsilon$.

The primary targets of calculations are the vacuum expectation values of time-ordered products of $\phi$; in QFT these are called the Green functions of the theory. From these can be reconstructed the scattering matrix, scattering cross sections, and other measurable quantities.

## One-Loop Calculations

Low-order graphs for the connected and amputated four-point Green function are shown in Figure 1. Each one-loop graph has the form

$$-\mathrm{i}\lambda^2 I(p^2) \stackrel{?}{=} \frac{\lambda^2}{2}\int \frac{\mathrm{d}^4k}{(2\pi)^4} \frac{1}{(k^2 - m^2 + \mathrm{i}0)[(p-k)^2 - m^2 + \mathrm{i}0]} \qquad [2]$$

where $p$ is a combination of external momenta. There is a divergence from where the loop

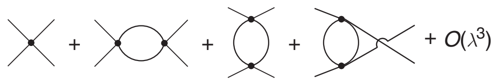



**Figure 1** One-loop approximation to connected and amputated four-point function, before renormalization.

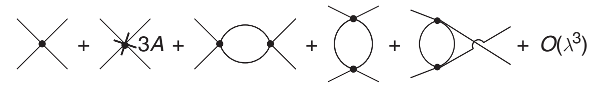



**Figure 2** One-loop approximation to renormalized connected and amputated four-point function, with counter-term.

momentum $k$ goes to infinity. We define the degree of divergence, $\Delta$, by counting powers of $k$ at large $k$, to get $\Delta = 0$. In an $n$-dimensional spacetime we would have $\Delta = n - 4$. The integral is divergent whenever $\Delta \geq 0$. Comparing the dimensions of the one-loop and tree graphs shows that $\Delta$ equals the negative of the energy dimension of the coupling $\lambda$. Thus, the dimensionlessness of $\lambda$ at the physical spacetime dimension is equivalent to the integral being just divergent.

The infinity in the integral implies that the theory *in its naive formulation* is not defined. With the aid of RG methods, it has been shown that the problem is with the complete theory, not just perturbation theory.

The divergence only arises because we use a continuum spacetime. So suppose that we formulate the theory initially on a lattice of spacing $a$ (in space or spacetime). Our loop graph is now

$$-\mathrm{i}\lambda^2 I(p; m, a) = \frac{-\lambda^2}{32\pi^4} \int \mathrm{d}^4 k\, S(k, m; a)\, S(p - k, m; a) \qquad [3]$$

where the free propagator $S(k, m; a)$ approaches the usual value $\mathrm{i}/(k^2 - m^2 + \mathrm{i}0)$ when $k$ is much smaller than $1/a$, and it falls off more rapidly for large $k$. The basic observation that propels the renormalization program is that the divergence as $a \to 0$ is independent of $p$. This is most easily seen by differentiating once with respect to $p$, after which the integral is convergent when $a = 0$, because the differentiated integral has degree of divergence –1.

Thus we can cancel the divergence in eqn [2] by replacing the coupling in the first term in Figure 1, by the so-called bare coupling

$$\lambda_0 = \lambda + 3A(a)\lambda^2 + O(\lambda^3) \qquad [4]$$

Here $A(a)$ is chosen so that the renormalized value of our one-loop graph,

$$-\mathrm{i}\lambda^2 I_R(p^2, m^2) = -\mathrm{i}\lambda^2 \lim_{a \to 0} [I(p; m, a) + A(a)] \qquad [5]$$

exists, at $a = 0$, with $A(a)$ in fact being real valued. The factor 3 multiplying $A(a)$ in eqn [4] is because there are three one-loop graphs, with equal divergent parts. The replacement for the coupling is made in the tree graph in Figure 1, but not yet at the vertices of the other graphs, because at the moment we are only doing a calculation accurate to order $\lambda^2$; the appropriate expansion parameter of the theory is the finite renormalized coupling $\lambda$, held fixed as $a \to 0$. We call the extra term in eqn [5] a counter-term. The diagrams for the correct renormalized calculation are represented in Figure 2, which has a counter-term graph compared with Figure 1.

In the physics terminology, used here, the cutting-off of the divergence by using a modified theory is called a regularization. This contrasts with the mathematics literature, where "regularized integral" usually means the same as a physicist's "renormalized integral."

There is always freedom to add a finite term to a counter-term. When we discuss the RG, we will see that this corresponds to a reorganization of the perturbation expansion and provides a powerful tool for improving perturbatively based calculations, especially in QCD. Contrary to the impression given in some parts of the literature, it is not necessary that a renormalized mass equal a corresponding physical particle mass, with similar statements for coupling and field renormalization. While such a prescription is common and natural in a simple theory like QED, it is by no means required and certainly may not always be best. If nothing else, the correspondence between fields and stable particles may be poor or nonexistent (as in QCD).

One classic possibility is to subtract the value of the graph at $p = 0$, a prescription associated with Bogoliubov, Parasiuk, and Hepp (BPH), which leads to

$$-\mathrm{i}\lambda^2 I_{\mathrm{R,\,BPH}}(p^2) = \frac{-\mathrm{i}\lambda^2}{32\pi^2} \int_0^1 \mathrm{d}x \ln\left[1 - p^2 x(1 - x)/m^2\right] \qquad [6]$$

In obtaining this from [2], we used a standard Feynman parameter formula,

$$\frac{1}{AB} = \int_0^1 \mathrm{d}x \frac{1}{[Ax + B(1 - x)]^2} \qquad [7]$$

to combine the propagator denominators, after which the integral over the momentum variable $k$ is elementary. We then obtain the renormalized one-loop (four-point and amputated) Green function

$$-\mathrm{i}\lambda - \mathrm{i}\lambda^2 [I_{\mathrm{R}}(s) + I_{\mathrm{R}}(t) + I_{\mathrm{R}}(u)] + O(\lambda^3) \qquad [8]$$

where $s$, $t$, and $u$ are the three standard Mandelstam invariants for the Green function. (For a $2 \to 2$

scattering process, or a corresponding off-shell Green function, in which particles of momenta $p_1$ and $p_2$ scatter to particles of momenta $p'_1$ and $p'_2$, the Mandelstam variables are defined as $s = (p_1 + p_2)^2$, $t = (p_1 - p'_1)^2$, and $u = (p_1 - p'_2)^2$.)

In the general case, with a nonzero degree of divergence, the divergent part of an integral is a polynomial in $p$ and $m$ of degree $D$, where $D$ is the smallest positive integer less than or equal to $\Delta$. In a higher spacetime dimension, this implies that renormalization of the original, momentum-independent, interaction vertex is not sufficient to cancel the divergences. We would need higher derivative terms, and this is evidence that the theory is not renormalizable in higher than 4 spacetime dimensions. Even so, the terms needed would be local, because of the polynomiality in $p$.

## Complete Formulation of Renormalization Program

The full renormalization program motivated by example calculations is:

- the theory is regulated to cut off the divergences;
- the numerical value of each coefficient in $\mathcal{L}$ is allowed to depend on the regulator parameter (e.g., $a$); and
- these dependences are adjusted so that finite results for Green functions are obtained after removal of the regulator.

In $\phi^4$ theory, we therefore replace $\mathcal{L}$ by

$$\mathcal{L} = \frac{Z}{2}(\partial\phi)^2 - \frac{Zm_0^2}{2}\phi^2 - \frac{Z^2\lambda_0}{4!}\phi^4 \qquad [9]$$

with the bare parameters, $Z$, $m_0$ and $\lambda_0$, having a regulator dependence such that Green functions of $\phi$ are finite at $a = 0$.

The slightly odd labeling of the coefficients in eqn [9] arises because observables like cross sections are invariant under a redefinition of the field by a factor. In terms of the bare field $\phi_0 \stackrel{\text{def}}{=} \sqrt{Z}\phi$, we have

$$\mathcal{L} = \frac{1}{2}(\partial\phi_0)^2 - \frac{m_0^2}{2}\phi_0^2 - \frac{\lambda_0}{4!}\phi_0^4 \qquad [10]$$

The unit coefficient of $(1/2)(\partial\phi_0)^2$ implies that $\phi_0$ has canonical commutation relations (in the regulated theory). This provides a natural standard for the normalization of the bare mass $m_0$ and the bare coupling $\lambda_0$.

All terms in $\mathcal{L}$ have coefficients with dimension zero or larger. This is commonly characterized by saying that the terms $\mathcal{L}$ "have dimension 4 or less," which refers to the products of field operators and derivatives in each term. A generalization of the power-counting analysis shows that if we start with a theory whose $\mathcal{L}$ only has terms of dimension 4 or less, then no terms of higher dimension are needed as counter-terms, at least not in perturbation theory. This is a very powerful restriction on self-contained QFTs, and was critical in the discovery of the standard model.

Sometimes it is found that the description of some piece of physics appears to need higher-dimension operators, as was the case originally with weak-interaction physics. The lack of renormalizability of such theories indicates that they cannot be complete, and an upper bound on the scale of their applicability can be computed, for example, a few hundred GeV for the four-fermion theory of weak interactions. Eventually, this theory was superseded by the renormalizable Weinberg–Salam theory of weak interactions, now a part of the standard model, to which the four-fermion theory provides a low-energy approximation for charged current weak interactions.

Certain operators of allowed dimensions are missing in eqn [9]: the unit operator, and $\phi$ and $\phi^3$. Symmetry under the transformation $\phi \to -\phi$ implies that Green functions with an odd number of fields vanish, so that no $\phi$ and $\phi^3$ counter-terms are needed. Divergences with the unit operator do appear, but not for ordinary Green functions. In gravitational physics, the coefficient of the unit operator gives renormalization of the cosmological constant.

To implement renormalized perturbation theory, we partition $\mathcal{L}$ (nonuniquely) as

$$\mathcal{L} = \mathcal{L}_{\text{free}} + \mathcal{L}_{\text{basic interaction}} + \mathcal{L}_{\text{counter-term}} \qquad [11]$$

where the free, the basic interaction, and the counter-term Lagrangians are

$$\mathcal{L}_{\text{free}} = \frac{1}{2}(\partial\phi)^2 - \frac{m^2}{2}\phi^2 \qquad [12]$$

$$\mathcal{L}_{\text{basic interaction}} = -\frac{\lambda}{4!}\phi^4 \qquad [13]$$

$$\mathcal{L}_{\text{counter-term}} = \frac{Z-1}{2}(\partial\phi)^2 - \frac{(Zm_0^2 - m^2)}{2}\phi^2 - \frac{(Z^2\lambda_0 - \lambda)}{4!}\phi^4 \qquad [14]$$

The renormalized coupling and mass, $\lambda$ and $m$, are to be fixed and finite when the UV regulator is removed. Both the basic interaction and the counter-terms are treated as interactions. First we compute "basic graphs" for Green functions using only the basic

interaction. The counter-terms are expanded in powers of $\lambda$, and then all graphs involving counter-term vertices at the chosen order in $\lambda$ are added to the calculation. The counter-terms are arranged to cancel all the divergences, so that the UV regulator can be removed, with $m$ and $\lambda$ held fixed. The counter-terms cancel the parts of the basic Feynman graphs associated with large loop momenta. An algorithmic specification of the otherwise arbitrary finite parts of the counter-terms is called a renormalization prescription or a renormalization scheme. Thus, it gives a definite relation between the renormalized and bare parameters, and hence a definite specification of the partitioning of $\mathcal{L}$ into its three parts.

It has been proved that this procedure works to all orders in $\lambda$, with corresponding results for other theories. Even in the absence of fully rigorous nonperturbative proofs, it appears clear that the results extend beyond perturbation theory, at least in asymptotically free theories like QCD: see the discussion on Wilsonian RG (*see* Exact Renormalization Group).

## Dimensional Regularization and Minimal Subtraction

The final result for renormalized graphs does not depend on the particular regularization procedure. A particularly convenient procedure, especially in QCD, is dimensional regularization, where divergences are removed by going to a low spacetime dimension $n$. To make a useful regularization method, $n$ is treated as a continuous variable, $n = 4 - 2\epsilon$.

Great advantages of the method are that it preserves Poincaré invariance and many other symmetries (including the gauge symmetry of QCD), and that Feynman graph calculations are minimally more complicated than for finite graphs at $n = 4$, particularly when all the lines are massless, as in many QCD calculations.

Although there is no such object as a genuine vector space of finite noninteger dimension, it is possible to construct an operation that behaves as if it were an integration over such a space. The operation was proved unique by Wilson, and explicit constructions have been made, so that consistency is assured at the level of all Feynman graphs. Whether a satisfactory definition beyond perturbation theory exists remains to be determined.

It is convenient to arrange that the renormalized coupling is dimensionless in the regulated theory. This is done by changing the normalization of $\lambda$ with the aid of an extra parameter, the unit of mass $\mu$:

$$\lambda_0 = \mu^{2\epsilon}(\lambda + \text{counter-terms}) \qquad [15]$$

with $\lambda$ and $\mu$ being held fixed when $\epsilon \longrightarrow 0$. (Thus, the basic interaction in eqn [13] is changed to $-\lambda\mu^{2\epsilon}\phi^4/4!$.) Then for the one-loop graph of eqn [2], dimensionally regularized Feynman parameter methods give

$$-\mathrm{i}\lambda^2 I(p; m, \epsilon) = \frac{\mathrm{i}\lambda^2}{32\pi^2}(4\pi)^\epsilon \Gamma(\epsilon) \times \int_0^1 \mathrm{d}x \left[\frac{m^2 - p^2 x(1-x) - \mathrm{i}0}{\mu^2}\right]^{-\epsilon} \qquad [16]$$

A natural renormalization procedure is to subtract the pole at $\epsilon = 0$, but it is convenient to accompany this with other factors to remove some universally occurring finite terms. So $\overline{\mathrm{MS}}$ renormalization ("modified minimal subtraction") is defined by using the counter-term

$$-\mathrm{i}A(\epsilon)\lambda^2 = -\mathrm{i}\frac{\lambda^2 S_\epsilon}{32\pi^2\epsilon} \qquad [17]$$

where $S_\epsilon \stackrel{\text{def}}{=} (4\pi\, \mathrm{e}^{-\gamma_{\mathrm{E}}})^\epsilon$, with $\gamma_{\mathrm{E}} = 0.5772\ldots$ being the Euler constant. This gives a renormalized integral (at $\epsilon = 0$)

$$-\frac{\mathrm{i}\lambda^2}{32\pi^2}\int_0^1 \mathrm{d}x \ln\left[\frac{m^2 - p^2 x(1-x)}{\mu^2}\right] \qquad [18]$$

which can be evaluated easily. A particularly simple result is obtained at $m = 0$:

$$\frac{\mathrm{i}\lambda^2}{32\pi^2}\left[-\ln\frac{-p^2}{\mu^2} + 2\right] \qquad [19]$$

This formula symptomizes important and very useful algorithmic simplifications in the higher-order massless calculations common in QCD.

The $\overline{\mathrm{MS}}$ scheme amounts to a *de facto* standard for QCD. At higher orders a factor of ${S_\epsilon}^L$ is used in the counter-terms, with $L$ being the number of loops.

## Coordinate Space

Quantum fields are written as if they are functions of $x$, but they are in fact distributions or generalized functions, with quantum-mechanical operator values. This indicates that using products of fields is dangerous and in need of careful definition. The relation with ordinary distribution theory is simplest in the coordinate-space version of Feynman graphs. Indeed in the 1950s, Bogoliubov and Shirkov formulated renormalization as a problem of defining products of the singular numeric-valued distributions in coordinate-space Feynman graphs; theirs was perhaps the best treatment of renormalization in that era.

For example, the coordinate-space version of eqn [5] is

$$-\lambda^2 \lim_{a\to 0} \int \mathrm{d}^4x\,\mathrm{d}^4y\, f(x,y) \times \left[\tfrac{1}{2}\tilde{S}(x-y;m,a)^2 + \mathrm{i}A(a)\delta^{(4)}(x-y)\right] \quad [20]$$

where $x$ and $y$ are the coordinates for the interaction vertices, $f(x,y)$ is the product of external-line free propagators, and $\tilde{S}(x-y;m,a)$ is the coordinate-space free propagator, which at $a=0$ has a singularity

$$\frac{1}{4\pi^2[-(x-y)^2 + \mathrm{i}0]} \quad [21]$$

as $(x-y)^2 \to 0$. We see in eqn [20] a version of the Hadamard finite part of a divergent integral, and renormalization theory generalizes this to particular kinds of arbitrarily high-dimension integrals. The physical realization and justification of the use of the finite-part procedure is in terms of renormalization of parameters in the Lagrangian; this also gives the procedure a significance that goes beyond the integrals themselves and involves the full nonperturbative formulation of QFT.

## General Counter-Term Formulation

We have written $\mathcal{L}$ as a basic Lagrangian density plus counter-terms, and have seen in an example how to cancel divergences at one-loop order. In this section, we will see how the procedure works to all orders. The central mathematical tool is Bogoliubov's $R$-operation. Here the counter-terms are expanded as a sum of terms, one for each basic one-particle irreducible (1PI) graph with a non-negative degree of divergence. To each basic graph for a Green function is added a set of counter-term graphs associated with divergences for subgraphs. The central theorem of renormalization is that this procedure does in fact remove all the UV divergences, with the form of the counter-terms being determined by the simple computation of the degree of divergence for 1PI graphs.

To see the essential difficulty to be solved, consider a two-loop graph like the first one in Figure 3. Its divergence is not a polynomial in external momenta, and is therefore not canceled by an allowed counter-term. This is shown by differentiation with respect to external momenta, which does not produce a finite result because of the divergent one-loop subgraph. But for consistency of the theory, the one-loop counter-terms already computed must be themselves put into loop graphs. Among others, this gives the second graph of Figure 3, where the cross denotes that a counter-term contribution is used. The contribution used here is actually 2/3 of the total one-loop counter-term, for reasons of symmetry factors that are not fully evident at first sight. The remainder of the one-loop coupling renormalization cancels a subdivergence in another two-loop graph. It is readily shown that the divergence of the sum of the first two graphs in Figure 3 is momentum independent, and thus can be canceled by a vertex counter-term.

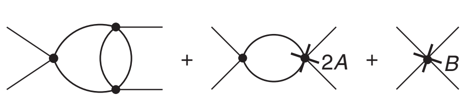



**Figure 3** A two-loop graph and its counter-terms. The label $B$ indicates that it is the two-loop overall counter-term for this graph.

This method is fully general, and is formalized in the Bogoliubov $R$-operation, which gives a recursive specification of the renormalized value $R(G)$ of a graph $G$:

$$R(G) \stackrel{\text{def}}{=} G + \sum_{\{\gamma_1,\ldots,\gamma_n\}} G|_{\gamma_i\to C(\gamma_i)} \quad [22]$$

The sum is over all sets of nonintersecting 1PI subgraphs of $G$, and the notation $G|_{\gamma_i\to C(\gamma_i)}$ denotes $G$ with all the subgraphs $\gamma_i$ replaced by associated counter-terms $C(\gamma_i)$. The counter-term $C(\gamma)$ of a 1PI graph $\gamma$ has the form

$$C(\gamma) \stackrel{\text{def}}{=} -T(\gamma + \text{counter-terms for subdivergences}) \quad [23]$$

Here $T$ is an operation that extracts the divergent part of its argument and whose precise definition gives the renormalization scheme. For example, in minimal subtraction we define

$$T(\Gamma) = \text{pole part at } \epsilon = 0 \text{ of } \Gamma \quad [24]$$

We formalize the term inside parentheses in eqn [23] as

$$\bar{R}(\gamma) \stackrel{\text{def}}{=} \gamma + \text{counterterms for subdivergences} = \gamma + {\sum_{\{\gamma_1,\ldots,\gamma_n\}}}' G|_{\gamma_i\to C(\gamma_i)} \quad [25]$$

where the prime on the $\sum'$ denotes that we sum over all sets of nonintersecting 1PI subgraphs except for the case that there is a single $\gamma_i$ equal to the whole graph (i.e., the term with $n=1$ and $\gamma_1=\gamma$ is omitted).

Note that, for the $\overline{\text{MS}}$ scheme, we define the $T$ operation to be applied to a factor of constant dimension obtained by taking the appropriate power of $\mu^\epsilon$ outside of the pole-part operation. Moreover, it is not a strict pole-part operation; instead each

pole is to be multiplied by ${S_\epsilon}^L$, where $L$ is the number of loops, and $S_\epsilon$ is defined after eqn [17].

Equations [22]–[25] give a recursive construction of the renormalization of an arbitrary graph. The recursion starts on one-loop graphs, since they have no subdivergences, that is, $C(\gamma) = -T(\gamma)$ for a one-loop 1PI graph.

Each counter-term $C(\gamma)$ is implemented as a contribution to the counter-term Lagrangian. The Feynman rules ensure that once $C(\gamma)$ has been computed, it appears as a vertex in bigger graphs in such a way as to give exactly the counter-terms for subdivergences used in the $R$-operation. It has been proved that the $R$-operation does in fact give finite results for Feynman graphs, and that basic power counting in exactly the same fashion as at one-loop determines the relevant operators.

In early treatments of renormalization, a problem was caused by graphs like **Figure 4**. This graph has three divergent subgraphs which overlap, rather than being nested. Within the $R$-operation approach, such cases are no harder to deal with than merely nested divergences.

The recursive specification of $R$-operation can be converted to a nonrecursive formulation by the forest formula of Zavyalov and Stepanov, later rediscovered by Zimmerman. It is normally the recursive formulation that is suited to all-orders proofs.

Whether these results, proved to all orders of perturbation theory, genuinely extend to the complete theory is not so easy to answer, certainly in a realistic four-dimensional QFT. One illuminating case is of a nonrelativistic quantum mechanics model with a delta-function potential in a two-dimensional space. Renormalization can be applied just as in field theory, but the model can also be treated exactly, and it has been shown that the results agree with perturbation theory.

Perturbation series in relativistic QFTs can at best be expected to be asymptotic, not convergent. So instead of a radius of convergence, we should talk about a region of applicability of a weak-coupling expansion. In a direct calculation of counter-terms, etc., the radius of applicability shrinks to zero as the regulator is removed. However, we can deduce the expansion for a renormalized quantity, whose expansion is expected to have a nonzero range of applicability. We can therefore appeal to the uniqueness of power series expansions to allow the

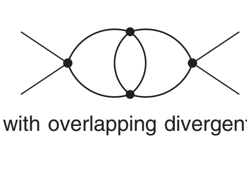


**Figure 4** Graph with overlapping divergent subgraphs.

calculation, at intermediate stages, to use bare quantities that are divergent as the regulator is removed.

## Renormalizability, Non-Renormalizability, and Super-Renormalizability

The basic power-counting method shows that if a theory with conventional fields (at $n = 4$) has only operators of dimension 4 or less in its $\mathcal{L}$, then the necessary counter-term operators are also of dimension 4 or less. So if we start with a Lagrangian with all possible such operators, given the field content, then the theory is renormalizable. This is not the whole story, as we will see in the discussion of gauge theories.

If we start with a Lagrangian containing operators of dimension higher than 4, then renormalization requires operators of ever higher dimension as counter-terms when one goes to higher orders in perturbation theory. Therefore, such a theory is said to be perturbatively non-renormalizable. Some very powerful methods of cancelation or some nonperturbative effects are needed to evade this result.

In the case of dimension-4 interactions, there is only a finite set of operators given the set of basic fields, but divergences occur at arbitrarily high orders in perturbation theory. If, instead, all the operators have at most dimension 3, then only a finite number of graphs need counter-terms. Such theories are called super-renormalizable. The divergent graphs also occur as subgraphs inside bigger graphs, of course. There is only one such theory in a four-dimensional spacetime: $\phi^3$ theory, which suffers from an energy density that is unbounded from below, so it is not physical. In lower spacetime dimension, where the requirements on operator dimension are different, there are many more known super-renormalizable theories, some with a very rigorous proof of existence.

All the above characterizations rely primarily on perturbative analysis, so they are subject to being not quite accurate in an exact theory, but they form a guide to the relevant issues.

## Renormalization and Symmetries: Gauge Theories

In most physical applications, we are interested in QFTs whose Lagrangian is restricted to obey certain symmetry requirements. Are these symmetries preserved by renormalization? That is, is the Lagrangian with all necessary counter-terms still invariant under the symmetry?

We first discuss nonchiral symmetries; these are symmetries in which the left-handed and right-handed parts of Dirac fields transform identically.

For Poincaré invariance and simple global internal symmetries, it is simplest to use a regulator, like dimensional regularization, which respects the symmetries. Then it is easily shown that the symmetries are preserved under renormalization. This holds even if the internal symmetries are spontaneously broken (as happens with a "wrong-sign mass term," e.g., negative $m^2$ in eqn [1]).

The case of local gauge symmetries is harder. But their preservation is more important, because gauge theories contain vector fields which, without a gauge symmetry, generally give unphysical features to the theory. For perturbation theory, BRST quantization is usually used, in which, instead of gauge symmetry, there is a BRST supersymmetry. This is manifested at the Green function level by Slavnov–Taylor identities that are more complicated, in general, than the Ward identities for simple global symmetries and for abelian local symmetries.

Dimensional regularization preserves these symmetries and the Slavnov–Taylor identities. Moreover, the $R$-operation still produces finite results with local counter-terms, but cancelations and relations occur between divergences for different graphs in order to preserve the symmetry. A simple example is QED, which has an abelian U(1) gauge symmetry, and whose gauge-invariant Lagrangian is

$$\mathcal{L} = -\frac{1}{4}\left(\partial_\mu A_\nu^{(0)} - \partial_\nu A_\mu^{(0)}\right)^2 + \bar{\psi}_0\left(\mathrm{i}\gamma^\mu\partial_\mu - e_0 A_\mu^{(0)} - m_0\right)\psi_0 \qquad [26]$$

At the level of individual divergent 1PI graphs, we get counter-terms proportional to ${A_\mu}^2$ and to $({A_\mu}^2)^2$, operators not present in the gauge-invariant Lagrangian. The Ward identities and Slavnov–Taylor identities show that these counter-terms cancel when they are summed over all graphs at a given order of renormalized perturbation theory. Moreover, the renormalization of coupling and the gauge field are inverse, so that $e_0 A_\mu^{(0)}$ equals the corresponding object with renormalized quantities, $\mu^\epsilon e A_\mu$. Naturally, sums of contributions to a counter-term in $\mathcal{L}$ can only be quantified with use of a regulator.

In nonabelian theories, the gauge-invariance properties are not just the absence of certain terms in $\mathcal{L}$ but quantitative relations between the coefficients of terms with different numbers of fields. Even so, the argument with Slavnov–Taylor identities generalizes appropriately and proves renormalizability of QCD, for example. But note that the relation concerning the product of the coupling and the gauge field does not generally hold; the form of the gauge transformation is itself renormalized, in a certain sense.

## Anomalies

Chiral symmetries, as in the weak-interaction part of the gauge symmetry of the standard model, are much harder to deal with. Chiral symmetries are ones for which the left-handed and right-handed components of Dirac field transform independently under different components of the symmetry group, local or global as the case may be. Occasionally, some or other of the left-handed or right-handed components may not even be present.

In general, chiral symmetries are not preserved by regularization, at least not without some other pathology. At best one can adjust the finite parts of counter-terms such that in the limit of the removal of the regulator, the Ward or Slavnov–Taylor identities hold. But in general, this cannot be done consistently, and the theory is said to suffer from an anomaly. In the case of chiral gauge theories, the presence of an anomaly prevents the (candidate) theory from being valid. A dramatic and nontrivial result (Adler–Bardeen theorem and some nontrivial generalizations) is that if chiral anomalies cancel at the one-loop level, then they cancel at all orders.

Similar results, but more difficult ones, hold for supersymmetries.

The anomaly cancelation conditions in the standard model lead to constraints that relate the lepton content to the quark content in each generation. For example, given the existence of the $b$ quark, and the $\tau$ and $\nu_\tau$ leptons (of masses around 4.5 GeV, 1.8 GeV, and zero respectively), it was strongly predicted on the grounds of anomaly cancelation that there must be a $t$ quark partner of the $b$ to complete the third generation of quark doublets. This prediction was much later vindicated by the discovery of the much heavier top quark with $m_t \simeq 175$ GeV.

## Renormalization Schemes

A precise definition of the counter-terms entails a specification of the renormalization prescription (or scheme), so that the finite parts of the counter-terms are determined. This apparently induces extra arbitrariness in the results. However, in the $\phi^4$ Lagrangian (for example), there are really only two independent parameters. (A scaling of the field does not affect any observables, so we do not count $Z$ as a parameter here.) Thus, at fixed regulator parameter $a$ or $\epsilon$, renormalization actually just gives a reparametrization of a two-parameter collection of theories. A renormalization prescription gives the

change of variables between bare and renormalized parameters, a rather singular transformation when the regulator is removed. If we have two different prescriptions, we can deduce a transformation between the renormalized parameters in the two schemes. The renormalized mass and coupling $m_1$ and $\lambda_1$ in one scheme can be obtained as functions of their values $m_2$ and $\lambda_2$ in the other scheme, with the bare parameters, and hence the physics, being the same in both schemes. Since these are renormalized parameters, the removal of the regulator leaves the transformation well behaved.

Generalization to all renormalizable theories is immediate.

## Renormalization Group and Applications and Generalizations

One part of the choice of renormalization scheme is that of a scale parameter such as the unit of mass $\mu$ of the $\overline{\text{MS}}$ scheme. The physical predictions of the theory are invariant if a change of $\mu$ is accompanied by a suitable change of the renormalized parameters, now considered as $\mu$-dependent parameters $\lambda(\mu)$ and $m(\mu)$. These are called the effective, or running, coupling and mass. The transformation of the parametrization of the theory is called an RG transformation.

The bare coupling and mass $\lambda_0$ and $m_0$ are RG invariant, and this can be used to obtain equations for the RG evolution of the effective parameters from the perturbatively computed counter-terms. For example, in $\phi^4$ theory, we have (in the renormalized theory after removal of the regulator)

$$\frac{d\lambda}{d\ln\mu^2} = \beta(\lambda) \qquad [27]$$

with $\beta(\lambda) = 3\lambda^2/(16\pi^2) + O(\lambda^3)$. As exemplified in eqns. [18] and [19], Feynman diagrams depend logarithmically on $\mu$. By choosing $\mu$ to be comparable to the physical external momentum scale, we remove possible large logarithms in this and higher orders. Thus, provided that the effective coupling at this scale is weak, we get an effective perturbation expansion.

This is a basic technique for exploiting perturbation theory in QCD, for the strong interactions, where the interactions are not automatically weak. In this theory the RG $\beta$ function is negative so that the coupling decreases to zero as $\mu \to \infty$; this is the asymptotic freedom of QCD.

A closely related method is that associated with the Callan–Symanzik equation, which is a formulation of a Ward identity for anomalously broken scale invariance. However, RG methods are the actually used ones, normally, even if sometimes an RG equation is incorrectly labeled as a Callan–Symanzik equation.

The elementary use of the RG is not sufficient for most interesting processes, which involve a set of widely different scales. Then more powerful theorems come into play. Typical are the factorization theorems of QCD (*see* Quantum Chromodynamics). These express differential cross sections for certain important reactions as a product of quantities that involve a single scale:

$$d\sigma = C(Q, \mu, \lambda(\mu)) \otimes f(m, \mu, \lambda(\mu)) + \text{small correction} \qquad [28]$$

The product is typically a matrix or a convolution product. The factors obey nontrivial RG equations, and these enable different values of $\mu$ to be used in the different factors. Predictions arise because some factors and the kernels of the RG equation are perturbatively calculable, with a weak effective coupling. Other factors, such as $f$ in eqn [28], are not perturbative. These are quantities with names like "parton distribution functions," and they are universal between many different processes. Thus, the nonperturbative functions can be measured in a limited set of reactions and used to predict cross sections for many other reactions with the aid of calculations of the perturbative factors.

Ultimately, this whole area depends on physical phenomena associated with renormalization.

## Concluding Remarks

The actual ability to remove the divergences in certain QFTs to produce consistent, finite, and nontrivial theories is a quite dramatic result. Moreover, associated with the integrals that give the divergences is behavior of the kind that is analyzed with RG methods and generalizations. So the properties of QFTs associated with renormalization get tightly coupled to many interesting consequences of the theories, most notably in QCD.

QFTs are actually very abstruse and difficult theories; only certain aspects currently lend themselves to practical calculations. So the reader should not assume that all aspects of their rigorous mathematical treatment are perfect. Experience, both within the theories and in their comparison with experiment, indicates, nevertheless, that we have a good approximation to the truth.

When one examines the mathematics associated with the *R*-operation and its generalizations with factorization theorems, there are clearly present some interesting mathematical structures that are not yet formulated in their most general terms. Some

indications of this can be seen in the work by Connes and Kreimer (*see* Hopf Algebra Structure of Renormalizable Quantum Field Theory), where it is seen that renormalization is associated with a Hopf algebra structure for Feynman graphs.

With such a deep subject, it is not surprising that it lends itself to other approaches, notably the Connes–Kreimer one and the Wilsonian one (*see* Exact Renormalization Group). Readers new to the subject should not be surprised if it is difficult to get a fully unified view of these different approaches.

## Notes on Bibliography

Reliable textbooks on quantum field theory are Sterman (1993) and Weinberg (1995). A clear account of the foundations of perturbative QCD methods is given by Sterman (1996).

A pedagogical account of renormalization and related subjects may be found in Collins (1984). The best account of renormalization theory before the 1970s is given by Bogoliubov and Shirkov (1959); the viewpoint is very modern, including a coordinate-space distribution-theoretic view. A full account of the Wilsonian method as applied to renormalization is given by Polchinski (1984).

Manuel and Tarrach (1994) give an excellent account of renormalization for a theory with a non-relativistic delta-function potential in 2 space dimensions, which provides a fully tractable model.

Tkachov (1994) reviews a systematic application of distribution theoretic methods to asymptotic problems in QFT. Finally, Weinzierl (1999) provides a construction of dimensional regularization with the aid of *K*-theory using an underlying vector space of the physical integer dimension. Other constructions, referred to in this paper, follow Wilson and use an infinite-dimensional underlying space.

## Acknowledgments

The author would like to thank Professor K Goeke for hospitality and support at the Ruhr-Universität-Bochum. This work was also supported by the US DOE.

*See also:* Anomalies; BRST Quantization; Effective Field Theories; Electroweak Theory; Euclidean Field Theory; Exact Renormalization Group; High $T_c$ Superconductor Theory; Holomorphic Dynamics; Hopf Algebra Structure of Renormalizable Quantum Field Theory; Lattice Gauge Theory; Operator Product Expansion in Quantum Field Theory; Perturbation Theory and its Techniques; Perturbative Renormalization Theory and BRST; Quantum Chromodynamics; Quantum Field Theory: A Brief Introduction; Singularities of the Ricci Flow; Standard Model of Particle Physics; Supergravity.

## Further Reading

Bogoliubov NN and Shirkov DV (1959) *Introduction to the Theory of Quantized Fields*. New York: Wiley-Interscience.

Collins JC (1984) *Renormalization*. Cambridge: Cambridge University Press.

Kraus E (1998) Renormalization of the electroweak standard model to all orders. *Annals of Physics* 262: 155–259.

Manuel C and Tarrach R (1994) Perturbative renormalization in quantum mechanics. *Physics Letters B* 328: 113–118.

Polchinski J (1984) Renormalization and effective Lagrangians. *Nuclear Physics B* 231: 269–295.

Sterman G (1993) *An Introduction to Quantum Field Theory*. Cambridge: Cambridge University Press.

Sterman G (1996) Partons, factorization and resummation. In: Soper DE (ed.) *QCD and Beyond*, pp. 327–406. (hep-ph/9606312). Singapore: World Scientific.

Tkachov FV (1994) Theory of asymptotic operation. A summary of basic principles. *Soviet Journal of Particles and Nuclei* 25: 649.

Weinberg S (1995) *The Quantum Theory of Fields, Vol. I. Foundations*. Cambridge: Cambridge University Press.

Weinzierl S (1999) Equivariant dimensional regularization, hep-ph/9903380.

Zavyalov OI (1990) *Renormalized Quantum Field Theory*. Dordrecht: Kluwer.

# Renormalization: Statistical Mechanics and Condensed Matter

**M Salmhofer**, Universität Leipzig, Leipzig, Germany



## Renormalization Group and Condensed Matter

Statistical mechanical systems at critical points exhibit scaling laws of order parameters, susceptibilities, and other observables. The exponents of these laws are universal, that is, independent of most details of the system. For example, the liquid–gas transition for real gases has the same exponents as the magnetization transition in the three-dimensional Ising model.

The renormalization group (RG) was developed by Kadanoff, Wilson, and Wegner, to understand these critical phenomena (Domb and Green 1976). The central idea is that the system becomes scale invariant at the critical point, which makes it natural to average over degrees of freedom on increasing length scales successively in the

calculation of the partition function. This leads to a map between effective interactions associated to different length scales. Thus, the focus shifts from the analysis of a single interaction to that of a flow on a space of interactions. This space is in general much larger than the original formulation of the model would suggest: the description of long-distance or low-energy properties may be in terms of variables that were not even present in the original formulation of the system. Phenomenologically, this corresponds to the emergence of collective degrees of freedom.

Condensed matter theory is itself already an effective theory, and its "microscopic" formulation gets inputs from the underlying theories, which determine in particular the statistics of the particles and their interactions at the scale of atomic energies. At much lower-energy scales, which are relevant for low-temperature phenomena in condensed matter, collective excitations of different, sometimes exotic, statistics may emerge, but the starting point is given naturally in terms of fermionic and bosonic particles. For this reason, the discussion given below will be split in these two cases.

A major difference between high-energy and condensed matter systems is that the latter have a well-defined Hamiltonian which can be used to define the finite-volume ensembles of quantum statistical mechanics and which determines the time evolution, as well as various analyticity properties.

The relevant spatial dimensions in condensed matter are $d \leq 3$, but some results in higher dimensions relevant for the development of the method will also be discussed below. The cases $d = 1$ and $d = 2$ have always been of mathematical interest but in recent years have become important for the theory of new materials.

Some interesting topics cannot be covered here due to space restrictions, notably the application of renormalization methods to membrane theory (see Wiese (2001)) and renormalization methods for operators (see Bach *et al.* (1998)).

## The Renormalization Group

In this section we briefly describe the setup of two important versions of the RG, namely the block spin RG and the RG based on scale decompositions of singular covariances.

### Block spin RG

Let $\Lambda$ be a finite lattice, for example, a finite subset of $\mathbb{Z}^d$. For the following, it is convenient to take $\Lambda$ to be a cube of side-length $L^K$ for $L > 1$ and some large $K$. Let $T$ be a set and $\Phi_\Lambda = \{\phi : \Lambda \to T\}$ be the set of spin configurations. Common examples for the target space $T$ are $T = \{-1, 1\}$ for the Ising model, $T = S^{N-1}$ for the O($N$) model, and $T = \mathbb{R}^n$ for unbounded spins. Let $S_\Lambda : \Phi_\Lambda \to \mathbb{R}, \phi \mapsto S_\Lambda(\phi)$ be an interaction and

$$Z(\Lambda, S_\Lambda) = \int \prod_{x \in \Lambda} \mathrm{d}\phi(x) \mathrm{e}^{-S_\Lambda(\phi)} \qquad [1]$$

In the unbounded case, $S_\Lambda$ is assumed to grow sufficiently fast for $|\phi| \to \infty$, so that $Z$ exists; for the case of a finite set $T$, the integral is replaced by a sum. Denote the corresponding Boltzmann factor by $\rho(\Lambda, S_\Lambda)$,

$$\rho(\Lambda, S_\Lambda)(\phi) = \frac{1}{Z(\Lambda, S_\Lambda)} \mathrm{e}^{-S_\Lambda(\phi)} \qquad [2]$$

The block spin transformation consists of an integration step and a rescaling step. Divide the lattice into cubic blocks of side-length $L$ and define a new lattice $\Lambda'$ by associating one lattice site of the new lattice to each $L$-block of the old lattice. For any $\phi' : \Lambda' \to T$, let

$$\rho'(\phi') = \int \prod_{x \in \Lambda} \mathrm{d}\phi(x) P(\phi', \phi) \mathrm{e}^{-S_\Lambda(\phi)} \qquad [3]$$

where $P(\phi', \phi) \geq 0$ and $\int \prod_{x' \in \Lambda'} \mathrm{d}\phi'(x') P(\phi', \phi) = 1$ for all $\phi$, so that $\rho'$ remains a probability distribution. Since $\rho'$ is positive, one defines

$$S'_{\Lambda'}(\phi') = -\log \rho'(\phi') \qquad [4]$$

By construction, the partition function is invariant: $Z(\Lambda', S'_{\Lambda'}) = Z(\Lambda, S_\Lambda)$. The new lattice $\Lambda'$ has spacing $L$; now rescale to make it a unit lattice. This completes the RG step in finite volume.

In an algorithmic sense, the "blocking rule" $P(\phi', \phi)$ can be viewed as a transition probability of a configuration $\phi$ to a configuration $\phi'$. $P$ may be deterministic, that is, simply fix $\phi'$ as a function of $\phi$. From the intuition of averaging over local fluctuations, $\phi'$ is often taken to be some average of $\phi(x)$ at $x$ in a block around $x'$, hence the name.

Obviously, the thus defined RG transformation often cannot be iterated arbitrarily, since in every application, the number of points of the lattice shrinks by a factor $L^d$, so that after $K$ iterations, a lattice with only a single point is left over. It is necessary to take the infinite-volume limit $L \to \infty$ to obtain a map that operates from a space to itself. However, [4] can become problematic in that limit: Gibbs measures $\rho$ can map to measures $\rho'$ whose large-deviation properties differ from those of Gibbs measures. The discussion of this problem and its solution is reviewed in Bricmont and Kupiainen (2001). The problem can be

solved in different ways, relaxing conditions on Gibbs measures or, in the Ising model, changing the description from the spins to the contours. The crucial point is that the difficulties arise only because [4] is applied globally, that is, to every $\phi'$. The set of bad $\phi'$ has very small probability.

Block spin methods have been used in mathematical construction of quantum field theories, for example, in the work of Gawedzki and Kupiainen (1985) and Balaban (1988) (see the subsection "Field theory and statistical mechanics"). The above-mentioned problem was avoided there by not taking a logarithm in the so-called large-field region (which has very small probability).

### Scale Decomposition RG

The generating functionals of quantum field theory and quantum statistical mechanics can be cast into the form

$$Z(C,V,\phi)=\int \mathrm{d}\mu_C(\phi')\,\mathrm{e}^{-V(\phi'+\phi)} \qquad [5]$$

Here $\mathrm{d}\mu_C$ denotes the Gaussian measure with covariance $C$, and $V$ is the two-body interaction between the particles. The field variables are real or complex for bosons and Grassmann-valued for fermions. Differentiating $\log Z$ with respect to the external field $\phi$ generates the connected amputated correlation functions. The covariance determines the free propagation of particles; the interaction their collisions.

In most cases, such functional integrals are *a priori* ill-defined, even if $V$ is small (and bounded from below) because the covariance $C$ is singular. That is, the integral kernel $C(X,X')$ of the operator $C$ either diverges as $|x-x'|\to 0$ (ultraviolet (UV) problem) or $C(X,X')$ has a slow decay as $|x-x'|\to\infty$ (infrared (IR) problem). In our notational convention, $X$ may, in addition to the configuration variable $x$, also contain discrete indices of the fields, such as a spin or color index. The dependence of $C$ on $x$ and $x'$ is assumed to be of the form $x-x'$. A typical example is the massless Gaussian field in $d$ dimensions, where $C$ is the inverse Fourier transform of $\hat{C}(k)=1/k^2$, $k\in\mathbb{R}^d$, which has both a UV and an IR problem, or its lattice analog,

$$\hat{D}(k)=\left(\frac{2}{a^2}\sum_{i=1}^{d}(1-\cos(ak_i)\right)^{-1}$$

with $a$ the lattice constant, which has only an IR problem. A typical interaction is of the type

$$V(\phi)=\int \mathrm{d}X\,\mathrm{d}Y\bar{\phi}(X)\phi(X)v(X,Y)\bar{\phi}(Y)\phi(Y) \qquad [6]$$

Again, we assume that the potential $v$ depends on $x$ and $y$ only via $x-y$, so that translation invariance holds. In both UV and IR cases, naive perturbation theory fails even as a formal power series. That is, writing $V=\lambda V_0$, with a coupling constant $\lambda$ which is treated as a formal expansion parameter, the singularity of $C$ leads to termwise divergences in the series. The theory is called perturbatively renormalizable if all divergences can be removed by posing counter-terms of certain types, which are fixed by physically sensible renormalization conditions. Identifying the UV renormalizable theories was a breakthrough in high-energy physics. The IR renormalization problem is different, and in some respects harder, because there is almost no freedom to put counter-terms: the microscopic model is given from the start. This will be discussed in more detail below for an example.

A much more ambitious, and largely open, project is to do this renormalization nonperturbatively, that is, to treat $\lambda$ as a real (typically, small) parameter. Some results will be discussed below.

The RG is set up by a scale decomposition $C=\sum_j C_j$. In the example of the massless Gaussian field, one would take each $\hat{C}_j$ to be a $C^\infty$ function supported in the region $\{k\in\mathbb{R}^d : M^j\le k^2\le M^{j+1}\}$, where $M>1$ is a fixed constant, and the summation over $j$ runs over $\mathbb{Z}$.

The scale decomposition of $C$ leads to a representation of [5] by an iteration of Gaussian convolution integrals with covariances $C_j$, hence a sequence of effective interactions $V_j$, defined recursively by

$$\mathrm{e}^{-V_j(\phi)}=\int \mathrm{d}\mu_{C_{j+1}}(\phi')\,\mathrm{e}^{-V_{j+1}(\phi'+\phi)}, \quad V_0=V \qquad [7]$$

For a singular covariance, the scale decomposition is an infinite sum. A formal object like [5] is now regularized by starting with a finite sum, that is, imposing a UV and IR cutoff, which is mathematically well defined, and then taking limits of the thus defined objects. Again, in condensed matter applications, imposing an IR cutoff is an operation that needs to be justified, for example, by showing that taking the limit as the cutoff is removed commutes with the infinite-volume limit.

Note that the RG map, which is the iteration $V_j\mapsto V_{j-1}$, goes to lower and lower $j$, corresponding to longer and longer length scales. The convention that the iteration starts at some fixed $j$, for example, $j=0$, is appropriate for IR problems. In UV problems, the iteration would start at some large $J_{\mathrm{UV}}$, which defines a UV cutoff and is taken to infinity, to remove the cutoff, at the end.

A variant using a continuous scale decomposition, $C=\int \mathrm{d}s\dot{C}_s$, originally due to Wegner and Houghton, became very popular after Polchinski (1984) used it

to give a short argument for perturbative renormalizability. Polchinski's equation, the analog of the recursion [7], reads

$$\frac{\partial V}{\partial s} = -\frac{1}{2} e^{V} \Delta_{\dot{C}_s} e^{-V} = \frac{1}{2} \Delta_{\dot{C}_s} V - \frac{1}{2} \left( \frac{\delta V}{\delta \phi}, \dot{C}_s \frac{\delta V}{\delta \phi} \right) \quad [8]$$

Here

$$\Delta_C = \left( \frac{\delta}{\delta \phi}, C \frac{\delta}{\delta \phi} \right)$$

denotes the Laplacian in field space associated to the covariance $C$. Polchinski's argument has been developed into a mathematical tool that applies to many models. For an introduction to perturbative renormalization using this method, see Salmhofer (1998). Equations of the type [8] have also been very useful beyond perturbation theory: much work has been done based on the beautiful representation of Mayer expansions found in Brydges and Kennedy (1987) using RG equations.

### Mathematical Structure and Difficulties

The RG flow is thus, depending on the implementation, either a sequence or a continuous flow of interactions. Setting up this flow in mathematical terms is not easy and indeed part of the mathematical RG analysis is to find a suitable space of interactions that is left invariant by the successive convolutions, and then to control the RG iteration. A serious problem is the proliferation of interactions: already a single application of the RG transformation [7] maps a simple interaction, such as [6], to a nonlocal functional of the fields,

$$V_j(\phi) = \sum_{m \geq 0} \int dX_1 \cdots dX_m \times v_m^{(j)}(X_1, \ldots, X_m)\, \phi(X_1) \cdots \phi(X_m) \quad [9]$$

Already for perturbative renormalization, one needs to extract local terms, calculate their flow more explicitly, and control the power counting of the remainder. The convergence of the series is not an issue in formal perturbation theory because in every finite order $r$ in $\lambda$, the sum over $m$ is finite.

For nonperturbative renormalization, however, the problem is much more serious. For bosonic systems, the expansion in powers of the fields in [9] is divergent, and one needs a split into small-field and large-field regions and cluster expansions to obtain a well-defined sequence of effective actions (Gawedzki and Kupiainen 1985, Feldman *et al.* 1987, Rivassean 1993). That is, the local parts are extracted and treated explicitly only in the small-field region, and this is combined with estimates on the rareness of large-field regions using cluster expansions. For fermions, the expansion in powers of the fields can be proved to converge for regular, summable covariances, which leads to substantial technical simplifications.

The spatial proliferation of interactions is absent only in certain one-dimensional and in specially constructed higher-dimensional models, the so-called "hierarchical models." In these models, the search for an RG fixed point is still a nonlinear fixed-point problem, whose treatment leads to interesting mathematical results.

This article will be restricted to the mathematical use of the RG both in perturbative and nonperturbative quantum field theory of condensed matter systems. Many nonrigorous but very interesting applications have also come out of this method, showing that it also works well in practice, but they will not be reviewed here. Before discussing condensed matter systems, the pioneering works done on the mathematical RG, which were largely motivated by high-energy physics, will be reviewed briefly, as they laid the foundation of much of the technique used later in the condensed matter case.

### Field Theory and Statistical Mechanics

Because of the close connection between quantum field theory and statistical mechanics given by formulas of the Feynman–Kac type, a significant amount of work on the mathematical RG focused on models of classical statistical mechanics in connection with field theories and gauge theories. Here we mention some of the pioneering results in that field.

The scale decomposition method was developed in a mathematical form and applied to perturbative UV renormalization of scalar field theories, as well as nonperturbative analysis of some models, by Gallavotti and Nicolò (Gallavotti 1985).

Infrared $\phi^4$ theory in four dimensions was constructed using block spin methods (Gawedzki and Kupiainen 1985) and scale decomposition RG (Feldman *et al.* 1987). An essential feature of the $\phi_4^4$ model is its IR asymptotic freedom, meaning that the local part of the effective quartic interaction tends to zero in the IR limit.

Block spin methods were used by Balaban (1988) to construct gauge theories in three and four dimensions. For gauge theories, the block spin RG has the major advantage that it allows to define a gauge-invariant RG flow. The scale decomposition violates gauge invariance, which creates substantial technical problems (Rivasseau 1993).

## Condensed Matter: Fermions

Starting with the seminal work of Feldman and Trubowitz (1990, 1991) and Benfatto and Gallavotti (1995), this field has become one of the most successful applications of the mathematical RG. We use this example to discuss the scale decomposition method in a bit more detail.

We shall mainly focus on models in $d \geq 2$ dimensions (the case $d = 1$ is described in detail in Benfatto and Gallavotti (1995)). The system is put into a finite (very large) box $\Lambda$ of side-length $L$. For simplicity we take periodic boundary conditions. The Hilbert space for spin-1/2 electrons is the fermionic Fock space $\mathcal{F} = \bigoplus_{n\geq 0} \bigwedge^n L^2(\Lambda, \mathbb{C}^2)$. The grand canonical ensemble in finite volume is given by the density operator $\rho = Z^{-1} e^{-\beta(H-\mu N)}$, with the Hamiltonian $H$ and the number operator $N$, in the usual second quantized form. The parameter $\beta = T^{-1}$ is the inverse temperature and the chemical potential $\mu$ is an auxiliary parameter used to fix the average particle number.

The grand canonical trace defining the ensemble can be rewritten in functional-integral form. It takes the form [5], but now $d\mu_C$ stands for a Grassmann Gaussian "measure," which is really only a linear functional (for definitions, see, e.g., Salmhofer (1998, chapter 4 and appendix B)). A two-body interaction corresponds to a quartic interaction polynomial $V$, as in [6]. The covariance is (in the infinite-volume limit $L \to \infty$)

$$C(\tau, x) = \frac{1}{\beta} \sum_{\omega \in \mathbb{M}_F} \int \frac{dk}{(2\pi)^d} e^{i(k\cdot x - \omega\tau)} \hat{C}(\omega, k)$$
$$\hat{C}(\omega, k) = \frac{1}{i\omega - e(k)} \qquad [10]$$

where $\tau \in (0, \beta]$ is a Euclidian time variable and $k$ is the spatial momentum. The summation over $\omega$ runs over the set of fermionic Matsubara frequencies $\mathbb{M}_F = \pi T(2\mathbb{Z} + 1)$. The function $e(k) = \varepsilon(k) - \mu$, where $\varepsilon(k)$ is the band function given by the single-particle term in the Hamiltonian. For a lattice system, $k \in \mathcal{B}_d$, the momentum space torus (e.g., for the lattice $\mathbb{Z}^d, \mathcal{B}_d = \mathbb{R}^d/2\pi\mathbb{Z}^d$); for a continuous system, $k \in \mathbb{R}^d$, hence there is a spatial UV problem. Electrons in a crystal have a natural spatial UV cutoff (see Salmhofer (1998, chapter 4) for a discussion) so we assume in the following that there is either a UV cutoff or that the system is on a lattice. A nonperturbative definition of the functional integral involves a limit from discrete times (by the Trotter product formula); see, for example, Salmhofer (1998) or Feldman *et al.* (2003, 2004).

### Perturbative Renormalization

**Renormalization of the Fermi surface at zero temperature** In the limit $T \to 0$, the Matsubara frequency $\omega$ becomes a real variable, hence the propagator has a singularity at $\omega = 0$ and $k \in S$, where $S = \{k : e(k) = 0\}$, a codimension-1 subset of $\mathcal{B}_d$, is the Fermi surface. The existence of a Fermi surface which does not degenerate to a point is a characteristic feature of systems showing metallic behavior.

The singularity implies that $\hat{C} \notin L^p(\mathbb{R} \times \mathcal{B}_d)$ for any $p \geq 2$. Because terms of the type

$$\int d\omega \int dk F(\omega, k)\hat{C}(\omega, k) \times \prod_{i=1}^{p-1} \Big(T_i(\omega, k)\hat{C}(\omega, k)\Big) \qquad [11]$$

appear for all $p \geq 1$ in the formal perturbation expansion, with functions $T_i$ and $F$ that do not vanish on the singularity set of $C$, the perturbation expansion for observables is termwise divergent. The deeper reason for these problems is that the interaction shifts the Fermi surface so that the true propagator has a singularity of the form $G(\omega, k) = (i\omega - e(k) - \sigma(\omega, k))^{-1}$. If the self-energy $\sigma$ is a sufficiently regular function, $G$ has the same integrability properties as $C$, but the singularity of $G$ is on the set $\tilde{S} = \{k : e(k) + \sigma(0, k) = 0\}$ (the singularity in $\omega$ remains at $\omega = 0$).

Let $1 = \sum_{j\leq 0} \chi_j(\omega, k)$ be a $C^\infty$ partition of unity such that

$$\text{for } j < 0 \quad \text{supp } \chi_j \subset \{(\omega, k) : \epsilon_0 M^{j-2} \leq |i\omega - e(k)| \leq \epsilon_0 M^j\} \qquad [12]$$

where $M > 1$ and $\epsilon_0$ is a fixed constant (an energy scale determined by the global properties of the function $e$; see Salmhofer (1998, chapter 4)). The corresponding covariances $\hat{C}_j = \hat{C}_{\chi_j}$ have the properties that for $j < 0, \|\hat{C}_j\|_1 \leq \text{const.} M^j$ and $\|\hat{C}_j\|_\infty \leq \text{const.} M^{-j}$. Using these bounds and expanding $v_m^{(j)} = \sum_{r\geq 1} v_{m,r}^{(j)} \lambda^r$, one can derive estimates for the coefficient functions $v_{m,r}^{(j)}$.

Of course, the scale decomposition by itself does not solve the problem of the moving singularity. It only allows us to pinpoint the problematic terms in the expansion. To construct the self-energy $\sigma$, as well as all higher Green functions, a two-step method is used (Feldman and Trubowitz 1990, 1991, Feldman *et al.* 1996, 2000). First, a counterterm function $K$ which modifies $e$ is introduced, so that all two-point insertions $T_i$ get subtracted on the Fermi surface, hence replaced by $\tilde{T}_i(\omega, k) = T_i(\omega, k) - T_i(0, k')$, with $k'$ obtained from $k$ by a

projection to the Fermi surface (Feldman and Trubouitz 1990, 1991). Consequently, the $\tilde{T}_i$ vanish linearly on the Fermi surface, so that the integral over $k$ in [11] converges. The effect of the counter-term function $K$ can be described less technically: it fixes the Fermi surface to be $S$, the zero set of $e$. Thus, $K$ forces $S$ to be the Fermi surface of the interacting system. To achieve this, $K$ must be chosen a function of $e$, $k$, and $\lambda V$. In contrast to the situation for covariances with point singularities, the function $K$ will, for a nontrivial Fermi surface, be very different from the original $e$. It can, however, be constructed to all orders in perturbation theory for a large class of Fermi surfaces. More precisely, one can prove: if $e \in C^2(\mathcal{B}_d, \mathbb{R})$, $\hat{v} \in C^2(\mathcal{B}_d, \mathbb{R})$, and the Fermi surface $S$ contains no points $k$ with $\nabla e(k) = 0$ and no flat sides, then $K = \sum_r \lambda^r K_r$ exists as a formal power series in $\lambda$ and the map $e \mapsto e + K$ is locally injective on this set of $e$'s (Feldman *et al.* 1996, 2000). With this counter-term, the order-$r$ $m$-point functions on scale $j$ satisfy the bounds

$$\left\| \hat{v}_{m,r}^{(j)} \right\|_\infty \leq w_{m,r} M^{(4-m)j/2} |j|^r$$

and

$$\left\| \hat{v}_{m,r}^{(j)} \right\|_1 \leq \tilde{w}_{m,r} \qquad [13]$$

with constants $w_{m,r}$ and $\tilde{w}_{m,r}$. Here $\hat{v}_{m,r}^{(j)}$ is the Fourier transform of $v_{m,r}^{(j)}$ (see [9], with the momentum conservation delta function from translation invariance removed.

Equation [13] implies that in the RG sense, the two-point function is relevant, the four-point function is marginal, and all higher $m$-point functions are irrelevant.

In one dimension, the Fermi "surface" reduces to two points which are related by a symmetry, so the counter-term function $K$ is just a constant, that is, an adjustment of the chemical potential $\mu$, which is justified because $\mu$ is only an auxiliary parameter used to fix the average value of the particle number. The counter-term function is a constant also in higher dimensions in the special case $e(k) = k^2 - \mu$: there, rotational symmetry implies that $K$ can be chosen independent of $k$ (if $v$ is also rotationally symmetric). However, in the generic case of non-spherical Fermi surfaces, $K$ depends nontrivially on $k$, and an inversion problem arises: adding the counter-term changes the model. To obtain the Green functions of a model with a given dispersion relation and interaction $(E, V)$, one needs to show that given $E$ in a suitable set, the equation

$$e(k) + K(\lambda, e, V)(k) = E(k) \qquad [14]$$

has a unique solution. If this is done, the procedure for renormalization is as follows. For a model given by dispersion relation and interaction $(E, V)$, solve [14], then add and subtract $e$ in the kinetic term. This automatically puts $K = E - e$ as a counter-term, and the expansion is now set up automatically with the right counter-term. The function $K$ describes the shift from the Fermi surface of the free system (the zero set of $E$) to that of the interacting system (the zero set of $e$). Proving that $K$ is sufficiently regular and solving [14] is nontrivial. Uniqueness of the solution follows from the above stated properties of $K$ as a function of $e$. Existence was shown for a class of Fermi surfaces with strictly positive curvature in Feldman *et al.* (1996, 2000), to every order in perturbation theory. This implies a bijective relation between the Fermi surfaces of the free and the interacting model.

**Positive temperature and the zero-limit temperature**
One advantage of the functional-integral approach is that the setup at positive temperatures is identical to that at zero temperature, save for the discreteness of the set $\mathbb{M}_\mathrm{F}$ at $T > 0$. Because $0 \notin \mathbb{M}_\mathrm{F}$, the temperature effectively provides an IR cutoff, so that all term-by-term divergences are regularized in a natural way. However, renormalization is still necessary because the temperature is a physical parameter and unrenormalized expansions give disastrous bounds for the behavior of observables as functions of the temperature. Renormalization carries over essentially unchanged (the counter-term function is constructed slightly differently).

Because $|\omega| \geq \pi/\beta$ for all $\omega \in \mathbb{M}_\mathrm{F}$, [12] implies $\mathrm{supp}\ \chi_j = \emptyset$ for $j < -J_\beta$, where

$$J_\beta = \log_M \frac{\beta \epsilon_0}{\pi} \qquad [15]$$

Thus, the scale decomposition is now a finite sum over $0 \geq j \geq -J_\beta$. This restriction is inessential for the problem of renormalizing the Fermi surface, but it puts a cutoff on the marginal growth of the four-point function: [15] and [13] imply that

$$\|\hat{v}_{m,r}^{(j)}\|_1 \leq \tilde{w}_{m,r} \left( \log \frac{\beta \epsilon_0}{\pi} \right)^r \qquad [16]$$

If one can show that $\tilde{w}_{m,r} \leq AB^r$ with constants $A$ and $B$, this implies that perturbation theory converges for $|\lambda| \log(\beta\epsilon_0/\pi) < B^{-1}$. Such a bound has been shown using constructive methods (Disertori and Rivasseau 2000, Feldman *et al.* 2003, 2004) (see below). The logarithm of $\beta$ is due to the Cooper instability (see Feldman and Trubowitz (1990, 1991) and Salmhofer (1998, section 4.5)).

The application of renormalization at positive temperature also led to the solution of a longstanding puzzle in solid-state physics, namely the (seeming) discontinuity of the results of perturbation theory as a function of the temperature claimed in the early literature. When renormalization is done correctly, there is no discontinuity in the temperature.

### Nonperturbative Renormalization for Fermions

It is a remarkable feature of fermionic field theories that for a covariance for which $\|\hat{C}\|_1$ and $\|C\|_1$ are both finite, the effective action defined in [7] exists and is analytic in the fields and in the original interaction $V$, thanks to determinant bounds. For a $V$ as in [6], with $\lambda v$ weak and of short range, the skeleton functions (where all relevant $m$-point functions are projected back to their initial values in the RG iteration) satisfy

$$\|\hat{v}_m^{(j)}\|_\infty \leq \text{const.} \|\hat{C}_j\|_1^{-(m/2)+1} \|C_j\|_1^{-1} \qquad [17]$$

For the many-electron covariance [10], with a positively curved $C^d$ Fermi surface and with the scale decomposition [12], $\|\hat{C}_j\|_1$ is of order $M^j$ and $\|C_j\|_1$ is of order $M^{-j(d+1)/2}$. The right-hand side of [17] then contains $M^{(d+3-m)j/2}$, which agrees (up to logarithms) with the perturbative power-counting bounds [13] only for $d=1$. In dimension $d=2$, the method has been refined by dividing the Fermi surface into angular sectors. The corresponding sectorized propagators have a better decay bound $\|C_j\|_1$, but the trade-off is sector sums at every vertex. Momentum conservation restricts these sector sums sufficiently in two dimensions to allow for good power-counting bounds. This has allowed for the construction of an interesting class of interacting fermionic models.

The major results obtained with the RG method are as follows.

Luttinger liquid behavior at zero temperature was proved for one-dimensional models with a repulsive interaction (Benfatto and Gallavotti 1995).

Fermi liquid behavior in the region where $|\lambda| \log(\beta\epsilon_0) \ll 1$ was proved for the two-dimensional model with $e(k)=k^2-1$, a local potential $V$, and a UV cutoff both on $k$ and the Matsubara frequencies $\omega$ in Disertori and Rivasseau (2000).

A two-dimensional model with a band function $e(k)$ that is nonsymmetric under $k \to -k$ and a general short-range interaction was proved to be a Fermi liquid at zero temperature (Feldman *et al.* 2003, 2004). Due to the asymmetry under $k \to -k$, the Cooper instability can be proved to be absent. In Feldman *et al.* (2003, 2004), a counter-term function as in Feldman *et al.* (1996, 2000) was used. The nonperturbative proof of the corresponding inversion theorem remains open.

In $d=3$, the proof of Fermi liquid behavior remains an open problem, despite some partial results.

## Condensed Matter: Bosons

Recent advances in quantum optice, in particular the trapping of ultracold atoms, have led to the experimental realization of Bose–Einstein condensation (BEC), which caused a surge of theoretical and mathematical works. For bosons, the definition of the ensembles is similar to, but more involved than in, the fermionic case. On a formal level, the functional-integral representation is analogous to fermions, except that the fields are not Grassmann fields but complex fields, and the covariance is given by a sum as in [10], but now the summation over $\omega$ runs over the bosonic Matsubara frequencies $\mathbb{M}_B = 2\pi T\mathbb{Z}$. The existence of even the free partition function in finite volume restricts the chemical potential (for free particles, $\mu < \inf_k \varepsilon(k)$ must hold). Note that $C$ is complex and Gaussian measures with complex covariances exist in infinite dimensions only under rather restricted conditions, which are not satisfied by [10]. This is inessential for perturbative studies, where everything can be reduced to finite-dimensional integrals involving the covariance, but a nonperturbative definition of functional integrals for such systems requires again a carefully regularized (e.g., discrete-time) definition of the functional integral.

### Bose–Einstein Condensation

The problem was treated to all orders in perturbation theory at positive particle density $\rho > 0$ by Benfatto (Benfatto and Gallavotti 1995). The initial interaction is again quartic, $\varepsilon(k)=k^2$, and one considers the problem at zero temperature, in the limit $\mu \to 0^-$, which is the limit in which BEC occurs for free particles. The interaction is expected to change the value of $\mu$, given the density, so a chemical potential term is included in the action, to give the interaction

$$V(\phi) = \int \mathrm{d}\tau \mathrm{d}x \mathrm{d}y |\phi(\tau,x)|^2 v(x-y) |\phi(\tau,y)|^2 + \nu \int \mathrm{d}\tau \mathrm{d}x |\phi(\tau,x)|^2 \qquad [18]$$

After writing $\phi(\tau,x)=\xi+\varphi(\tau,x)$, where $\xi$ is independent of $\tau$ and $x$, the density condition becomes $\rho=|\xi|^2$. $\nu$ now needs to be chosen such that the free energy has a minimum at $\xi=\sqrt{\rho}$. This can be reformulated in terms of the self-energy of the boson.

Benfatto uses the RG to prove that the propagator of the interacting system no longer has the singularity structure $(i\omega - k^2)^{-1}$ but instead $(\omega^2 + c^2k^2)^{-1}$, where $c$ is a constant. This requires a nontrivial analysis of Ward identities in the RG flow.

BEC has been proved in the Gross–Pitaevskii limit (Lieb *et al.* 2002). In the present formulation, this limit corresponds to an infinite-volume limit $L \to \infty$ where the density $\rho$ is taken to zero as an inverse power of $L$. A nonperturbative proof of BEC at fixed positive particle density remains an open problem.

### Superconductivity

Superconductivity (SC) occurs in fermionic systems, but it happens at energy scales where the relevant excitations have bosonic character: the Cooper pairs are bosons. In the RG framework, they arise naturally when the fermionic RG flow discussed above is stopped before it leaves the weak-coupling region and the dominant Cooper pairing term is rewritten by a Hubbard–Stratonovich transformation. The fermions can then be integrated over, resulting in the typical Mexican hat potential of an O(2) nonlinear sigma model. Effectively, one now has to deal with a problem similar to the one for BEC, but the action is considerably more complicated.

### The Nonlinear Sigma Models

The prototypical model, into whose universality class both examples mentioned above fall, is that of O($N$) nonlinear sigma models: both BEC and SC can be reformulated as spontaneous symmetry breaking (SSB) in the O(2) model in dimensions $d \geq 3$. For $d = 2$, long-range order is possible only at zero temperature because only then does the time direction truly represent a third dimension, preventing the Mermin–Wagner theorem from applying.

SSB has been proved for lattice O($N$) models by reflection positivity and Gaussian domination methods (Fröhlich *et al.* 1976). The elegance and simplicity of this method is unsurpassed, but only very special actions satisfy reflection positivity, so that the method cannot be used for the effective actions obtained in condensed matter models. Results in the direction of proving SSB in O($N$) models for $d \geq 3$ by RG methods, which apply to much more general actions, have been obtained by Balaban (1995).

*See also:* Bose–Einstein Condensates; Fermionic Systems; High $T_c$ Superconductor Theory; Holomorphic Dynamics; Operator Product Expansion in Quantum Field Theory; Perturbative Renormalization Theory and BRST; Phase Transition Dynamics; Reflection Positivity and Phase Transitions.

## Further Reading

Bach V, Fröhlich J, and Sigal IM (1998) Renormalization group analysis of spectral problems in quantum field theory. *Advances in Mathematics* 137: 205–298.

Balaban T (1988) Convergent renormalization expansions for lattice gauge theories. *Communications in Mathematical Physics* 119: 243–285.

Balaban T (1995) A low-temperature expansion for classical $N$-vector models I. A renormalization group flow. *Communications in Mathematical Physics* 167: 103–154.

Benfatto G and Gallavotti G (1995) *Renormalization Group*. Princeton: Princeton University Press.

Bricmont J and Kupiainen A (2001) Renormalizing the renormalization group pathologies. *Physics Reports* 348: 5–31.

Brydges DC and Kennedy T (1987) Mayer expansions and the Hamilton–Jacobi equation. *Journal of Statistical Physics* 48: 19–49.

Disertori M and Rivasseau V (2000) Interacting Fermi liquid in two dimensions at finite temperature I. Convergent contributions. *Communications in Mathematical Physics* 215: 251–290.

Domb C and Green M (eds.) (1976) *Phase Transitions and Critical Phenomena*, vol. 6. London: Academic Press.

Feldman J, Magnen J, Rivasseau V, and Sénéor R (1987) Construction of infrared $\phi_4^4$ by a phase space expansion. *Communications in Mathematical Physics* 109: 437.

Feldman J and Trubowitz E (1990) Perturbation theory for many-fermion systems. *Helvetica Physica Acta* 63: 157.

Feldman J and Trubowitz E (1991) The flow of an electron-phonon system to the superconducting state. *Helvetica Physica Acta* 64: 213.

Feldman J, Salmhofer M, and Trubowitz E (1996) Perturbation theory around non-nested Fermi surfaces I. Keeping the Fermi surface fixed. *Journal of Statistical Physics* 84: 1209–1336.

Feldman J, Salmhofer M, and Trubowitz E (2000) An inversion theorem in Fermi surface theory. *Communications on Pure and Applied Mathematics* 53: 1350–1384.

Feldman J, Knörrer H, and Trubowitz E (2003) A class of Fermi liquids. *Reviews in Mathematical Physics* 15: 949–1169.

Feldman J, Knörrer H, and Trubowitz E (2004) *Communications in Mathematical Physics* 247: 1–319.

Fröhlich J, Simon B, and Spencer T (1976) Infrared bounds, phase transitions, and continuous symmetry breaking. *Communications in Mathematical Physics* 50: 79.

Gallavotti G (1985) Renormalization theory and ultraviolet stability via renormalization group methods. *Reviews of Modern Physics* 57: 471–569.

Gawedzki K and Kupiainen A (1985) Massless lattice $\phi_4^4$ theory: Rigorous control of a renormalizable asymptotically free model. *Communications in Mathematical Physics* 99: 197–252.

Lieb E, Seiringer R, Solovej JP, and Yngvason J (2002) The ground state of the Bose gas. In: *Current Developments in Mathematics, 2001*, pp. 131–178. Cambridge: International Press.

Polchinski J (1984) Renormalization and effective Lagrangians. *Nuclear Physics B* 231: 269.

Rivasseau V (1993) *From Perturbative to Constructive Renormalization*. Princeton, NJ: Princeton University Press.

Salmhofer M (1998) *Renormalization: An Introduction*, Springer Texts and Monographs in Physics. Heidelberg: Springer.

Wiese KJ (2001) Polymerized membranes, a review. In: Domb C and Lebowitz J (eds.) *Phase Transitions and Critical Phenomena*, vol. 19. Academic Press.

# Resonances

**N Burq**, Université Paris-Sud, Orsay, France



## Introduction

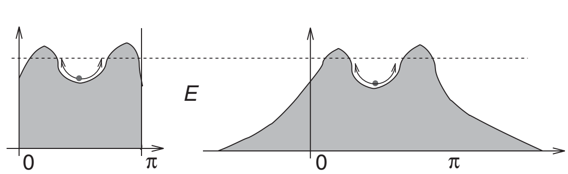
In quantum mechanics and wave propagation, eigenvalues (and eigenfunctions) appear naturally as they describe the behavior of a quantum system (or the vibration of a structure). There are however some cases where these simple notions do not suffice and one has to appeal to the more subtle notion of resonances. For example, if the vibration of a drum is well understood in terms of eigenvalues (the audible frequencies) and eigenfunctions (the corresponding vibrating modes), the notion of resonances is necessary to understand the propagation of waves in the exterior of a bounded obstacle. Another example (taken from Zworski (2002)) which allows us to understand both the similarities of resonances with eigenvalues and their differences is the following: consider the motion of a classical particle submitted to a force field deriving from the potential $V_1(x)$ on a bounded interval as shown in **Figure 1a**. If the classical momentum is denoted by $\xi$, then the classical energy is given by

$$E = |\xi|^2 + V_1(x)$$

and the classical motion is given by the relations of Hamiltonian mechanics:

$$\dot{x} = \frac{\partial E}{\partial \xi} = 2\xi, \qquad \dot{\xi} = -\frac{\partial E}{\partial x} = -V'(x)$$

Since energy is conserved, if the initial energy is smaller than the top of the barrier, then the classical particle bounces forever in the well. Now we can consider the same example with the potential $V_2(x)$ on $\mathbb{R}$ as shown in **Figure 1b**. Of course, if the particle is initially inside the well (with the same energy as before), the classical motion remains the same.




(a) (b)

**Figure 1a, b** A particle trapped in a well.

On the quantum mechanics point of view, both systems are described by the Hamiltonians

$$H_i = -h^2 \frac{d^2}{dx^2} + V_i(x)$$

acting on $L^2([-1,1])$ (with boundary conditions) and $L^2(\mathbb{R})$, respectively. In the first case, $H_1$ has a discrete spectrum, $\lambda_{j,h} \in \mathbb{R}$ with eigenfunctions $e_{j,h}(x), j \in \mathbb{N}$, and the time evolution of the system is given by

$$e^{itH_1} u = \sum_j e^{it\lambda_{j,h}} u_{j,h} \times e_{j,h} \qquad [1]$$

where $u_{j,h} \times e_{j,h}$ is the orthogonal projection of $u$ on the eigenspace $\mathbb{C}e_{j,h}$. In the second case, $H_2$ has no square integrable eigenfunction, and no simple description as [1] can consequently hold. However as $h \to 0$, the correspondence principle tells us that quantum mechanics should get close to classical mechanics. Since for both quantum problems the classical limit is the same (at least for initial states confined in the well with energy $E$), we expect that for the second potential there should exist a quantum state corresponding to the classical one. In fact, this is indeed the case and one can show that there exist resonant states $e_{j,h}$ associated to resonances $E_{j,h}$ which are solution of the equation

$$H_2 e_{j,h} - E_{j,h} e_{j,h}, \quad E_{j,h} \sim E$$

are not square integrable, but still have moderate growth at infinity and are confined in the interior of the well (see sections "Definition" and "Location of resonances"). On the other hand, the first quantum system is confined, whereas the second one is not and we know that even for initial states confined in the well, tunneling effect allows the quantum particle to escape to infinity. This fact should be described by the theory as a main difference between eigenvalues and resonances. This is indeed the case as the resonances $E_{j,h}$ are not real (contrarily to eigenvalues of self-adjoint operators) but have a nonvanishing imaginary part (see section "Resonance-free regions")

$$\operatorname{Im} E_{j,h} \sim e^{-C/h}$$

If we assume that a similar description as [1] still holds for the second system, at least locally in space (see section "Resonances and time asymptotics"), then, for time $t >> e^{C/h}$, the factor $e^{itE_h}$ becomes very small (the quantum particle has left the well due to tunneling effect).

There have been several studies on resonances and scattering theory and the presentation here cannot be complete. For a more in-depth presentation, one can

consult the books by Lax and Phillips (1989) and Hislop and Sigal (1987), or the reviews on resonances by Vodev (2001) and Zworski (1994) for example.

## Definition

There are different (equivalent) definitions of resonances. The most elegant is certainly the Helffer and Sjöstrand (1986) definition (see also the presentation of complex scaling by Combes *et al.* (1984) and the very general "black box" framework by Sjöstrand and Zworski (1991)). However, it requires a few prerequisites and we preferred to stick to the more elementary (but less general) resolvent point of view. The starting point for this definition of resonances is the fact that the eigenvalues of a (self-adjoint) operator $P$ are the points where $P$ is not injective. The more general resonances will be the points where the operator is not invertible (on suitable spaces).

More precisely, consider a perturbation of the Laplace operator on $\mathbb{R}^n$, $P_0(h) = -h^2\Delta$ in the following sense: let $\Theta \subset \mathbb{R}^d$ be a (possibly empty) smooth obstacle whose complementary, $\Omega = \Theta^c$, is connected. Consider a classical self-adjoint operator defined on $L^2(\Omega)$:

$$P_h u = (-h^2\Delta + V(x))u \qquad [2]$$

with boundary conditions (Dirichlet)

$$u\,|_{\partial\Theta} = 0 \qquad [3]$$

(Neumann boundary conditions could be used too). This setting contains both the Schrödinger operator ($P_h = -h^2\Delta + V(x) - E$ on $\Omega = \mathbb{R}^n$) and the Helmoltz equation with Dirichlet conditions, in the exterior of an obstacle (waves at large frequencies: $P = -\Delta - \tau^2$; in this case, define $h = \tau^{-1}$ and $P_h = -h^2\Delta$), which we shall define as acoustical scattering.

We assume that $P$ is a perturbation of $P_0$, that is, $V \to 0, |x| \to +\infty$ sufficiently fast (see Sjöstrand and Zworski (1991) for the very general black box assumptions). For example, this perturbation assumption is fulfilled if $V$ has compact support. Then the resolvent $P_h(z) = (P_h - z)^{-1}$ is well defined for $\mathrm{Im}\, z \neq 0$ as a bounded operator from $L^2(\Omega)$ to

$$H^2(\Omega) \cap H_0^1(\Omega)$$

(because the operator $P_h$ is self-adjoint). However, it is not bounded for $z > 0$ on $L^2(\Omega)$ because the essential spectrum of $P_h$ is precisely the semiaxis $z > 0$, but it admits a meromorphic continuation from $\mathbf{Im}\, z > 0$ toward the lower half-plane:

$$R_h(z) : L^2(\Omega)_{\mathrm{comp}} \to L^2(\Omega)_{\mathrm{loc}}$$

The poles of this resolvent $R_h$ are by definition the semiclassical resonances, $\mathrm{Res}_{\mathrm{sc}}(P_h)$.

**Remark 1** In the case of acoustical scattering ($P = -\Delta - \tau^2, \tau = h^{-1}$), the introduction of the additional parameter $z$ is pointless and one works directly with the parameter $\tau = h^{-1}\sqrt{z}$. In that case the resolvent $R(\tau)(-\Delta - \tau^2)^{-1}$ is well defined for $\mathbf{Im}\, \tau < 0$, the essential spectrum is precisely the axis $\tau \in \mathbb{R}$ and the resolvent admits a meromorphic continuation from $\mathbf{Im}\, z < 0$ toward the upper half-plane (with possibly a cut at 0):

$$R(\tau) : L^2(\Omega)_{\mathrm{comp}} \to L^2(\Omega)_{\mathrm{loc}}$$

The acoustic resonances are by definition the poles of this meromorphic continuation. They are related to semiclassical resonances by the relation

$$\mathrm{Res}_{\mathrm{sc}} = h\sqrt{\mathrm{Res}_{\mathrm{ac}}}$$

It can also be shown that if $z$ is a resonance, there exists an associated resonant state $e_z$ such that

$$(P_h - z)e_z = 0$$

the function $e_z$ satisfies Sommerfeld radiation conditions (in polar coordinates $(r, \theta) \in [0, +\infty) \times \mathbb{S}^{n-1}$)

$$|h\partial_r e - i\sqrt{z}e| \leq C|\mathrm{e}^{\mathrm{i}\sqrt{z}r}|/r^{1+n/2}$$

and the function

$$\frac{e_z}{1 + r^{(1/2)+\epsilon}}\ \mathrm{e}^{\mathrm{i}\sqrt{z}r}$$

is square integrable.

## Resonance-Free Regions

The very first result about resonance-free regions is based on Rellich uniqueness theorem (uniqueness for solutions of elliptic second-order equations) and says that there are no real resonances (except possibly 0). The more precise determination of resonance-free regions (originally in acoustical scattering) has been a subject of study from the 1960s and it has motivated a large range of works from the multiplier methods of Morawetz (1975) to the general propagation of singularity theorem of Melrose and Sjöstrand (1978). To state the main result in this direction, we need the notion of nontrapping perturbation.

**Definition 1** A generalized bicharacteristic at energy $E(x(s), \xi(s))$ is an integral curve of the Hamiltonian field

$$H_p = \frac{\partial p}{\partial \xi}\frac{\partial}{\partial x} - \frac{\partial p}{\partial x}\frac{\partial}{\partial \xi}$$

of the principal symbol $p(x, \xi) = |\xi|^2 + V(x)$ of the operator $P$, included in the characteristic set $p(x, \xi) = E$ and which, when hitting the boundary of the obstacle, reflects according to the laws of geometric optics (see (Melrose and Sjöstrand 1978)).

The operator $P$ (or by extension the obstacle in the case of acoustic scattering) is said to be nontrapping at energy $E$ if all generalized bicharacteristics go to the infinity:

$$\lim_{s\to\pm\infty} |x(s)| = +\infty$$

The operator $P$ (or by extension the obstacle in the case of acoustic scattering) is said to be nontrapping near energy $E$ if $P$ is nontrapping at energy $E'$ for $E'$ in a neighborhood of $E$.

The following result was obtained in different generalities by Morawetz (1975), Melrose and Sjöstrand (1978), and others.

**Theorem 1** *Assume that the operator $P$ is nontrapping near energy $E$. Then for any $N > 0$ there exist $h_0 > 0$ such that for $0 < h < h_0$ there are no resonances in the set*

$$\{z; |\mathrm{Im}\, z| \leq -Nh\log(h)\}$$

In the case of analytic geometries (and coefficients), this result (see Bardos *et al.* 1987) can be improved to

**Theorem 2** *Assume that the operator $P$ is non trapping. Then there exist $\epsilon > 0, N_0 > 0$ and $h_0 > 0$ such that for $0 < h < h_0$ there are no resonances in the set*

$$\{z; |\mathrm{Im}\, z| \leq N_0 h^{1-(1/3)}\} \cap \{|z - E| \leq \epsilon\}$$

**Remark 2** In the case of acoustical scattering, with the new definition of resonances, $\tau = h^{-1}\sqrt{z}$, the resonance-free zones have respectively the forms

$$\{z; |\mathrm{Im}\, z| \leq -N\log(|z|), |z| >> 1\}$$

$$\{z; |\mathrm{Im}\, z| \leq N_0|z|^{1/3}, |z| >> 1\}$$

In the case of trapping perturbations, the first result was obtained by Burq (1998).

**Theorem 3** *There exist $C > 0$ and $h_0 > 0$ such that for $0 < h < h_0$ there are no resonances in the set*

$$\{z; |\mathrm{Im}\, z| \leq N_0\, \mathrm{e}^{-C/h}\} \cap \{|z - E| \leq \epsilon\}$$

## Resonances and Time Asymptotics

The relationship between eigenfunctions/eigenvalues and time asymptotics is straightforward. This is no longer the case for resonances. For nontrapping problems however, this question has been studied in the late 1960s by Lax and Phillips (1989) and Vainberg (1968). In particular, this approach was decisive to study the local energy decay in acoustical scattering. As a consequence of Theorem 1, we have

**Theorem 4** *If the acoustical problem is nontrapping, then there exist $C, \alpha > 0$ such that for any solution of the wave equation*

$$\Box u = 0,\quad u|_{t=0} = u_0,\ \partial_t u|_{t=0} = u_1,\ u|_{\Gamma_D} = 0,\ \frac{\partial u}{\partial_n}|_{\Gamma_N} = 0$$

*with compactly supported initial data $(u_0, u_1)$ (in a fixed compact), one has*

$$E_{\mathrm{loc}}(u) = \int_{\Omega\cap\{|x|\leq C} |\nabla u|^2 + |\partial_t u|^2 \leq \begin{cases} C\mathrm{e}^{-\alpha t} & \text{if the space dimension is even} \\ \dfrac{C}{t^d} & \text{if the space dimension is odd} \end{cases} \quad [4]$$

Trapping perturbations were investigated more recently. In that case, the local energy decays, but the rate cannot be uniform. The first trapping example in acoustic scattering was studied by Ikawa (1983): the obstacle is the union of a finite number (and at least two) convex bodies. In that case, one has

**Theorem 5** *For any $\epsilon > 0$ there exists $C > 0$ such that for any initial data supported in a fixed compact set*

$$E_{\mathrm{loc}}(u)(t) \leq C\mathrm{e}^{-\alpha t}\|(u_0, u_1)\|^2_{D((1-\Delta)^{(1+\epsilon)/2})}$$

*where $D((1-\Delta)^{(1+\epsilon)/2})$ is the domain of the operator $(1-\Delta)^{(1+\epsilon)/2}$. Remark that the norm in $D((1-\Delta)^{1/2})$ is the natural energy and consequently the estimate above exhibits a loss of $\epsilon$ derivatives. For strongly trapping perturbations, the results are worse. They are consequences of Theorem 3.*

**Theorem 6** *For any $k$ there exists $C_k > 0$ such that for any initial data supported in a fixed compact set*

$$E_{\mathrm{loc}}(u)(t) \leq \frac{C_k}{\log(t)^{2k}}\|(u_0, u_1)\|^2_{D((1-\Delta)^{(1+k)/2})}$$

One can also obtain real asymptotic expansions in terms of resonances (see the work by Tang and Zworski (2000)).

**Theorem 7** *Let $\chi \in C_c^\infty(\mathbb{R}^n)$ and $\psi \in C_c^\infty((0,\infty))$ and let chsupp $\psi = [a, b]$. There exists $0 < \delta < c(h) < 2\delta$ such that for every $M > M_0$ there exists $L = L(M)$, and we have*

$$\chi\mathrm{e}^{-\mathrm{i}t\mu(P)/h}\chi\psi(P) = \sum_{z\in\Omega(h)\cap\mathrm{Res}(P)} \chi\mathrm{Res}(\mathrm{e}^{-\mathrm{i}t\mu(\bullet)/h} \times R(\bullet, h), z)\chi\psi(P) + \mathcal{O}_{\mathcal{H}\to\mathcal{H}}(h^\infty),\quad \text{for } t > h^{-L} \quad [5]$$

$$\Omega(h) = (a - c(h), b + c(h)) - \mathrm{i}[0, h^M)$$

*where* Res($f(\bullet), z$) *denotes the residue of a meromorphic family of operators*, $f$, *at* $z$.

The function $c(h)$ depends on the distribution of resonances: roughly speaking we cannot "cut" through a dense cloud of resonances. Even in the very well understood case of the modular surface there is, currently at least, a need for some nonexplicit grouping of terms. The same ideas can be applied to acoustic scattering.

## Trace Formulas

Trace formulas provide a description of the classical/quantum correspondence: one side is given by the trace of a certain function of the operator $f(P_h)$, whereas the other side is described in terms of classical objects (closed orbits of the classical flow). In the case of discrete eigenvalues, the question is relatively simple and can be solved by using the spectral theorem. In the case of continuous spectrum, the problem is much more subtle (self-adjoint operators with continuous spectrum behave in some ways as non-normal operators). It has been studied by Lax and Phillips (1989), Bardos *et al.* (1982), and Melrose (1982). More recently, Sjöstrand (1997) introduced a local notion of trace formulas.

Let $W \subset \Omega$ be an open precompact subsets of $e^{i[-2\theta_0, 0]}]0, +\infty[$. Assume that the intersections $I$ and $J$ of $W$ and $\Omega$ with the real axis are intervals and that $\Omega$ is simply connected.

**Theorem 8** *Let* $f(z, h)$ *be a family of holomorphic functions on* $z \in \Omega$ *such that* $|f|_{\Omega\setminus W}| \leq 1$. *Let* $\chi \in C_0^\infty(\mathbb{R})$ *equal to 1 on a neighborhood of* $\bar{I}$. *Then*

$$\begin{aligned}&\mathrm{Trace}\big((\chi f)(P_h) - (\chi f)(-h^2\Delta)\big)\\ &\quad = \sum_{\lambda \text{ a resonance of } P_h \cap \Omega} f(\lambda, h) + \mathcal{O}(h^{-n})\end{aligned}$$

The use of this result with a clever choice of functions $f$ allows Sjöstrand to show that an analytic singularity of the function $E \mapsto \mathrm{Vol}(\{x; V(x) \geq E\})$ (observe that if $V$ is bounded, this function vanishes for large $E$ and consequently it has analytic singularities) gives a lower bound for $\Omega$ a neighborhood of $E$

$$\sharp \mathrm{Res}(P_h) \cap \Omega \geq ch^{-n}$$

which coincides with the upper bound (see Zworski (2002) and the references given there).

## Location of Resonances

In some particular cases, one can expect to have a precise description of the location of resonances. This is the case in Ikawa's example in acoustic scattering where the obstacle is the union of two disjoint convex bodies. In this case, the line minimizing the distance, $d$, between the bodies is trapped. However, this trapped trajectory is isolated and of hyperbolic type (unstable). Ikawa (1983) and Gérard (1988) have obtained:

**Theorem 9** *There exist geometric positive constants* $k_p \to +\infty$ *as* $p \to +\infty$ *such that all resonances located above the line* $\mathrm{Im}\, z \geq -C$ (*C arbitrary large but fixed*) *have an asymptotic expansion*

$$\lambda \sim \lambda_{j,p} + \sum_l a_{l,p} \lambda_{j,p}^{-l/2} + \mathcal{O}(\lambda_{j,p}^{-\infty}), \quad j \to +\infty$$

*where the approximate resonances*

$$\lambda_{j,p} = j\frac{\pi}{d} - \mathrm{i}k_p$$

*are located on horizontal lines.*

Another example is when the obstacle is convex. This example is nontrapping and Sjöstrand and Zworski (1999) are able to prove that the resonances in any region $\mathrm{Im}\, z \geq N|z|^{1/3}$ ($N$ arbitrary large) are asymptotically distributed near cubic curves

$$\mathcal{C}_j = \{z \in \mathbb{C}; \mathrm{Im}\, z = -c_j |z|^{1/3}\}$$

Finally, the last main example where one can give a precise asymptotic for resonances is when there exists a stable (elliptic) periodic trajectory for the Hamiltonian flow. In that case it had been known from the 1960s (see the works by Babič (1968)) that one can construct quasimodes, that is, compactly supported approximate solutions of the eigenfunctions equation:

$$(P_h - E_h)e_j = \mathcal{O}(h^\infty)$$

It is only recently that Tang and Zworski (1998) and Stefanov (1999) proved that these quasimodes constructions imply the existence of resonances asymptotic to $E_h, h \to 0$.

*See also:* $h$-Pseudodifferential Operators and Applications; Semi-Classical Spectra and Closed Orbits.

## Further Reading


Babič VM (1968) Eigenfunctions which are concentrated in the neighborhood of a closed geodesic. *Zapiski Nauchnykh Seminanov Leningradkogo Otdeleniya Matematicheskogo Instituta Imeni V.A. Steklova* 9: 15–63.

Bardos C, Guillot JC, and Ralston J (1982) La relation de Poisson pour l'équation des ondes dans un ouvert non borné. Application à la théorie de la diffusion. *Communications in Partial Differential Equations* 7: 905–958.

Bardos C, Lebeau G, and Rauch J (1987) Scattering frequencies and Gevrey 3 singularities. *Inventiones Mathematicae* 90: 77–114.

Burq N (1998) Décroissance de l'énergie locale de l'équation des ondes pour le problème extérieur et absence de résonance au voisinage du réel. *Acta Mathematica* 180: 1–29.

Combes J-M, Duclos P, and Seiler R (1984) On the shape resonance. In: *Resonances – Models and Phenomena* (*Bielefeld, 1984*), Lecture Notes in Physics, vol. 211, pp. 64–77. Berlin: Springer.

Gérard C (1988) Asymptotique des pôles de la matrice de scattering pour deux obstacles strictement convexes. *Supplément au Bulletin de la Société Mathématique de France* 116: 146 pp.

Helffer B and Sjöstrand J (1986) Resonances en limite semi-classique. *Mémoire de la S.M.F* 114(24–25): 228 pp.

Hislop PD and Sigal IM (1987) Shape resonances in quantum mechanics. In: *Differential Equations and Mathematical Physics* (*Birmingham, Ala., 1986*), Lecture Notes in Mathematics, vol. 1285, pp. 180–196. Berlin: Springer.

Ikawa M (1983) On the poles of the scattering matrix for two convex obstacles. *Journal of Mathematics of the Kyoto University* 23: 127–194.

Lax PD and Phillips RS (1989) *Scattering Theory*, Pure and Applied Mathematics, 2nd edn., vol. 26. Boston: Academic Press.

Melrose RB (1982) Scattering theory and the trace of the wave group. *Journal of Functional Analysis* 45: 429–440.

Melrose RB and Sjöstrand J (1978) Singularities of boundary value problems. I. *Communications in Pure and Applied Mathematics* 31: 593–617.

Melrose RB and Sjöstrand J (1982) Singularities of boundary value problems. II. *Communications in Pure and Applied Mathematics* 35: 129–168.

Morawetz CS (1975) Decay for solutions of the exterior problem for the wave equation. *Communication in Pure Applied Mathematics* 28: 229–264.

Sjöstrand J (1997) A trace formula and review of some estimates for resonances. In: Rodino L (ed.) *Microlocal Analysis and Spectral Theory* (*Lucca, 1996*), NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., vol. 490, pp. 377–437. Dordrecht: Kluwer Academic.

Sjöstrand J and Zworski M (1991) Complex scaling and the distribution of scattering poles. *Journal of the American Mathematical Society* 4(4): 729–769.

Sjöstrand J and Zworski M (1999) Asymptotic distribution of resonances for convex obstacles. *Acta Mathematica* 183(2): 191–253.

Stefanov P (1999) Quasimodes and resonances: sharp lower bounds. *Duke Mathematical Journal* 99(1): 75–92.

Tang S-H and Zworski M (1998) From quasimodes to resonances. *Mathematical Research Letters* 5(3): 261–272.

Tang SH and Zworski M (2000) Resonance expansions of scattered waves. *Communication in Pure and Applied Mathematics* 53(10): 1305–1334.

Vainberg BR (1968) On the analytical properties of the resolvent for a certain class of operator pencils. *Mathematics of the USSR Sbornik* 6(2): 241–273.

Vodev G (2001) *Resonances in euclidean scattering. Cubo Matematica Educacional* 3: 317–360. http://www.math.sciences.univ-nantes.fr.

Zworski M (1994) Counting scattering poles. In: Ikawa M (ed.) *Spectral and Scattering Theory (Sanda, 1992)*, Lecture Notes in Pure and Applied Mathematics, vol. 161, pp. 301–331. New York: Dekker.

Zworski M (2002) Quantum resonances and partial differential equations. In: Li Ta Tsien (ed.) *Proceedings of the International Congress of Mathematicians* (*Beijing, 2002*), vol. III, pp. 243–252. Beijing: Higher Education Press.

**Ricci Flow** *see* Singularities of the Ricci Flow

# Riemann Surfaces

**K Hulek**, Universität Hannover, Hannover, Germany



## Introduction

Riemann surfaces were first studied as the natural domain of definition of (multivalued) holomorphic or meromorphic functions. They were the starting point for the development of the theory of real and complex manifolds (see Weyl (1997)). Nowadays, Riemann surfaces are simply defined as one-dimensional complex manifolds (see the next section). Compact Riemann surfaces can be embedded into projective spaces and are thus, by virtue of Chow's theorem, algebraic curves. By uniformization theory, the universal cover of a connected Riemann surface is either the unit disk, the complex plane, or the Riemann sphere (see the section "Uniformization").

This article discusses the basic theory of compact Riemann surfaces, such as their topology, their periods, and the definition of the Jacobian variety. Studying the zeros and poles of meromorphic functions leads to the notion of divisors and linear systems. In modern language this can be rephrased in terms of line bundles, resp. locally free sheaves (see the section "Divisors, linear systems, and line bundles"). One of the fundamental results is the Riemann–Roch theorem which expresses the difference between the dimension of a linear system and that of its adjoint system in terms of the degree of the linear system and the genus of the curve. This theorem has been vastly generalized and is truly one of the cornerstones of algebraic geometry. A formulation of this result and a discussion of some of its applications are also discussed.

A study of the subsets of the Jacobians parametrizing linear systems of given degree and dimension leads to Brill–Noether theory, which is discussed in the section "Brill–Noether theory." This is followed by a brief introduction to the theory of equations and syzygies of canonical curves.

Moduli spaces play a central role in the theory of complex variables and in algebraic geometry. Arguably, the most important of these is the moduli space of curves of genus $g$. This and related moduli problems are treated in the section "Moduli of compact Riemann surfaces." In particular, the space of stable maps is closely related to quantum cohomology. Finally, we present a brief discussion of the Verlinde formula and conformal blocks.

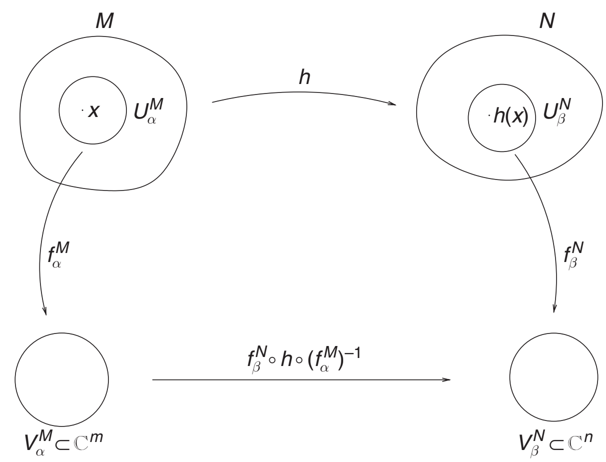



**Figure 2** Holomorphic map between manifolds.

## Basic Definitions

Riemann surfaces are one-dimensional complex manifolds. An $n$-dimensional complex manifold $M$ is a topological Hausdorff space (i.e., for any two points $x \neq y$ on $M$, there are disjoint open neighborhoods containing $x$ and $y$), which has a countable basis for its topology, together with a complex atlas $\mathcal{A}$. The latter is an open covering $(U_\alpha)_{\alpha \in \mathcal{A}}$ together with homeomorphisms $f_\alpha : U_\alpha \to V_\alpha \subset \mathbb{C}^n$, where the $U_\alpha$ are open subsets of $M$ and the $V_\alpha$ are open sets in $\mathbb{C}^n$. The main requirement is that these charts are holomorphically compatible, that is, for $U_\alpha \cap U_\beta \neq \emptyset$, the map shown in Figure 1,

$$f_\beta \circ f_\alpha^{-1}|_{f_\alpha(U_\alpha \cap U_\beta)} : f_\alpha(U_\alpha \cap U_\beta) \to f_\beta(U_\alpha \cap U_\beta) \subset \mathbb{C}^n$$

is biholomorphic. A map $h : M \to N$ between two complex manifolds is holomorphic if it is so with respect to the local charts. This means the following:

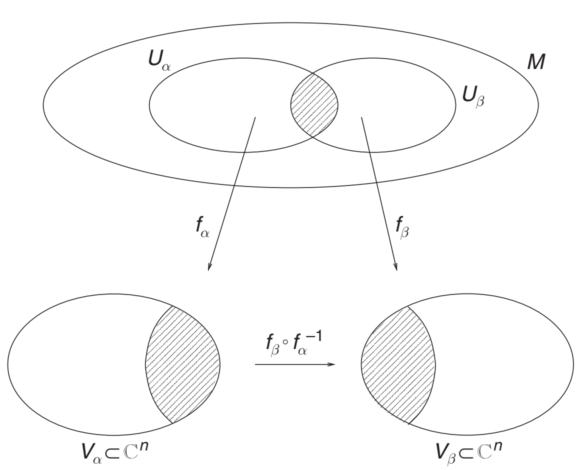



**Figure 1** Charts of a complex manifold.

for each point $x \in M$, there are charts $f_\alpha^M : U_\alpha^M \to V_\alpha^M \subset \mathbb{C}^n$ near $x$ and $f_\beta^N : U_\beta^N \to V_\beta^N \subset \mathbb{C}^m$ near $h(x)$ with $h(U_\alpha^M) \subset U_\beta^N$ such that the map shown in Figure 2

$$f_\beta^N \circ h \circ (f_\alpha^M)^{-1} : V_\alpha^M \to V_\beta^N \subset \mathbb{C}^m$$

is holomorphic (one checks easily that this does not depend on the choice of the charts).

A Riemann surface is a one-dimensional complex manifold. Trivial examples are given by open sets in $\mathbb{C}$ (where one chart suffices). Another example is the Riemann sphere $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, which can be covered by the two charts given by $z \neq \infty$ and $z \neq 0$. Both of these charts are homeomorphic to $\mathbb{C}$ with the transition function given by $z \mapsto 1/z$. Historically, Riemann surfaces were viewed as (branched) coverings of $\mathbb{C}$ or of the sphere, where they appear as the natural domain of definition of multivalued holomorphic or meromorphic functions.

## Uniformization

If $M$ is a Riemann surface, then its universal covering $\tilde{M}$ is again a Riemann surface. The connected and simply connected Riemann surfaces can be fully classified. Let

$$\mathbb{E} = \{z \in \mathbb{C}; |z| < 1\}$$

be the unit disk and $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ the Riemann sphere. The latter can be identified with the complex projective line $\mathbb{P}^1_\mathbb{C}$.

**Theorem 1** (Generalized Riemann mapping theorem). *Every connected and simply connected Riemann surface is biholomorphically equivalent*

*to the unit disk $\mathbb{E}$, the complex plane $\mathbb{C}$, or the Riemann sphere $\hat{\mathbb{C}}$.*

This theorem was proved rigorously by Koebe and Poincaré at the beginning of the twentieth century.

## Compact Riemann Surfaces

The topological structure of a compact Riemann surface $C$ is determined by its genus $g$ (Figure 3). Topologically, a Riemann surface of genus $g$ is a sphere with $g$ handles or, equivalently, a torus with $g$ holes.

Analytically, the genus can be characterized as the maximal number of linearly independent holomorphic forms on $C$ (see also the section "The Riemann–Roch theorem and applications").

There exists a very close link with algebraic geometry: every compact Riemann surface $C$ can be embedded into some projective space $\mathbb{P}^n_{\mathbb{C}}$ (in fact already into $\mathbb{P}^3_{\mathbb{C}}$). By Chow's theorem, $C$ is then a (projective) algebraic variety, that is, it can be described by finitely many homogeneous equations. It should be noted that such a phenomenon is special to complex dimension 1. The crucial point is that one can always construct a nonconstant meromorphic function on a Riemann surface (e.g., by Dirichlet's principle). Given such a function, it is not difficult to find a projective embedding of a compact Riemann surface $C$. On the other hand, it is easy to construct a compact two-dimensional torus $T=\mathbb{C}^2/L$ for some suitably chosen lattice $L$, which cannot be embedded into any projective space $\mathbb{P}^n_{\mathbb{C}}$.

The dichotomy Riemann surface/algebraic curve arises from different points of view: analysts think of a real two-dimensional surface with a Riemannian metric which, via isothermal coordinates, defines a holomorphic structure, whereas algebraic geometers think of a complex one-dimensional object.

In this article, the expressions compact Riemann surface and (projective) algebraic curve are both used interchangeably. The choice depends on which expression is more commonly used in the part of the theory which is discussed in the relevant section.

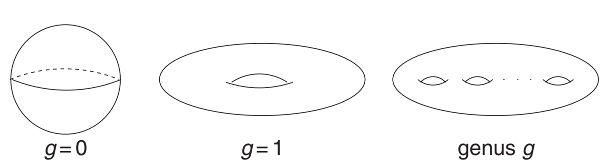



**Figure 3** Genus of Riemann surfaces.

## Periods and the Jacobian

On a compact Riemann surface $C$ of genus $g$, there exist $2g$ homologically independent paths, that is, $H_1(C,\mathbb{Z})\cong\mathbb{Z}^{2g}$.

Let $\gamma_1,\ldots,\gamma_{2g}$ be a basis of $H_1(C,\mathbb{Z})$ and let $\omega_1,\ldots,\omega_g$ be a basis of the space of holomorphic 1-forms on $C$. Integrating these forms over the paths $\gamma_1,\ldots,\gamma_{2g}$ defines the period matrix

$$\Omega=\begin{pmatrix}\int_{\gamma_1}\omega_1 & \cdots & \int_{\gamma_{2g}}\omega_1\\ \vdots & & \vdots\\ \int_{\gamma_1}\omega_g & \cdots & \int_{\gamma_{2g}}\omega_g\end{pmatrix}$$

If $Q=(\gamma_i,\gamma_j)$ is the intersection matrix of the paths $\gamma_1,\ldots,\gamma_{2g}$, then $\Omega$ satisfies the Riemann bilinear relations

$$\Omega Q\Omega^t=0,\quad \sqrt{-1}\,\Omega Q\overline{\Omega}^t>0 \qquad [1]$$

where the latter condition means positive definite. One can choose (see Figure 4) $\gamma_1,\ldots,\gamma_{2g}$ such that

$$Q=J=\begin{pmatrix}0 & \mathbf{1}_g\\ -\mathbf{1}_g & 0\end{pmatrix}$$

where $\mathbf{1}_g$ is the $g\times g$ unit matrix. Moreover, $\omega_1,\ldots,\omega_g$ can be chosen such that

$$\Omega=\begin{pmatrix}1 & \cdots & 0 & \tau_{11} & \cdots & \tau_{1g}\\ \vdots & \ddots & \vdots & \vdots & & \vdots\\ 0 & \cdots & 1 & \tau_{g1} & \cdots & \tau_{gg}\end{pmatrix}$$

Let

$$\Omega_0=(\tau_{ij})_{1\le i,j\le g}$$

Then the Riemann bilinear relations [1] become

$$\Omega_0=\Omega_0^t,\quad \operatorname{Im}\Omega_0^t>0$$

that is, $\Omega_0$ is an element of the Siegel upper half-space

$$\mathbb{H}_g=\{\tau\in\operatorname{Mat}(g\times g,\mathbb{C});\tau=\tau^t,\operatorname{Im}\tau>0\}$$

The matrix $\Omega_0$ is defined by the Riemann surface $C$ only up to the action of the symplectic group

$$\operatorname{Sp}(2g,\mathbb{Z})=\{M\in\operatorname{Mat}(2g\times 2g,\mathbb{Z});MJM^t=J\}$$

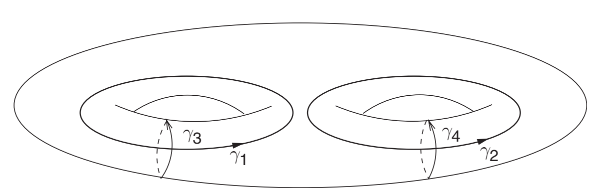



**Figure 4** Homology of a compact Riemann surface.

which acts on the Siegel space $\mathbb{H}_g$ by

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} : \tau \mapsto (A\tau + B)(C\tau + D)^{-1}$$

Here $A, \ldots, D$ are $g \times g$ blocks.

The rows of the matrix $\Omega$ define a rank-$2g$ lattice $L_\Omega$ in $\mathbb{C}^g$ and the Jacobian of $C$ is the torus

$$J(C) = \mathbb{C}^g / L_\Omega$$

More intrinsically, one can define $J(C)$ as follows. Let $H^0(C, \omega_C)$ be the space of holomorphic differential forms on $C$. Then, integration over cycles defines a monomorphism

$$H_1(C, \mathbb{Z}) \to H^0(C, \omega_C)^*$$
$$\gamma \mapsto \int_\gamma$$

and

$$J(C) = H^0(C, \omega_C)^* / H_1(C, \mathbb{Z})$$

For a fixed base point $P_0 \in C$, the Abel–Jacobi map is defined by

$$u : C \to J(C)$$
$$P \mapsto \left( \int_{P_0}^{P} \omega_1, \ldots, \int_{P_0}^{P} \omega_g \right)$$

Here, the integration is taken over some path from $P_0$ to $P$. Obviously, the integral depends on the choice of this path, but since $J(C)$ was obtained by dividing out the periods given by integrating over a basis of $H_1(C, \mathbb{Z})$, the map is well defined.

Let $C^d$ be the $d$th Cartesian product of $C$, that is, the set of all ordered $d$-tuples $(P_1, \ldots, P_d)$. Then, $u$ defines a map

$$u^d : C^d \to J(C)$$
$$(P_1, \ldots, P_d) \mapsto u(P_1) + \cdots + u(P_d)$$

where $+$ is the usual addition on the torus $J(C)$. If $d = g - 1$, then

$$\Theta = \mathrm{Im}(u^{g-1}) \subset J(C)$$

is a hypersurface (i.e., has codimension 1 in $J(C)$) and is called a theta divisor. A different choice of the base point $P_0$ results in a translation of the theta divisor. Using the theta divisor, one can show that $J(C)$ is an abelian variety, that is, $J(C)$ can be embedded into some projective space $\mathbb{P}^n_{\mathbb{C}}$. The pair $(J(C), \Theta)$ is a principally polarized abelian variety and Torelli's theorem states that $C$ can be reconstructed from its Jacobian $J(C)$ and the theta divisor $\Theta$.

## Divisors, Linear Systems, and Line Bundles

A divisor $D$ on $C$ is a formal sum

$$D = n_1 P_1 + \cdots + n_k P_k, \quad P_i \in C, \; n_i \in \mathbb{Z}$$

The degree of $D$ is defined as

$$\deg D = n_1 + \cdots + n_k$$

and $D$ is called "effective" if all $n_i \geq 0$. Every meromorphic function $f \neq 0$ defines a divisor

$$(f) = f_0 - f_\infty$$

where $f_0$ are the zeros of $f$ and $f_\infty$ the poles (each counted with multiplicity). Divisors of the form $(f)$ are called principal divisors and the degree of any principal divisor is 0 (see the next section). Two divisors $D_1$ and $D_2$ are called linearly equivalent ($D_1 \sim D_2$) if their difference is a principal divisor, that is,

$$D_1 - D_2 = (f)$$

for some meromorphic function $f \neq 0$. This defines an equivalence relation on the group $\mathrm{Div}(C)$ of all divisors on $C$. Since principal divisors have degree 0, the notion of degree also makes sense for classes of linearly equivalent divisors. We define the divisor class group of $C$ by

$$\mathrm{Cl}(C) = \mathrm{Div}(C)/\sim$$

The degree map defines an exact sequence

$$0 \to \mathrm{Cl}^0(C) \to \mathrm{Cl}(C) \overset{\deg}{\to} \mathbb{Z} \to 0$$

where $\mathrm{Cl}^0(C)$ is the subgroup of $\mathrm{Cl}(C)$ of divisor classes of degree 0.

Let $C_d$ be the set of unordered $d$-tuples of points on $C$, that is,

$$C_d = C^d / S_d$$

where the symmetric group $S_d$ acts on the Cartesian product $C^d$ by permutation. This is again a smooth projective variety and the Abel–Jacobi map $u^d : C^d \to J(C)$ clearly factors through a map

$$u_d : C_d \to J(C)$$

The fibers of this map are of particular interest.

**Theorem 2** (Abel). *Two effective divisors $D_1$ and $D_2$ on $C$ of the same degree $d$ are linearly equivalent if and only if $u_d(D_1) = u_d(D_2)$.*

One normally denotes the inverse image of $u_d(D)$ by

$$|D| = u_d^{-1}(u_d(D)) = \{D'; D' \geq 0, D' \sim D\}$$

Note that the latter description also makes sense if $D$ itself is not necessarily effective. One calls $|D|$ the

complete linear system defined by the divisor $D$. If $\deg D < 0$, then automatically $|D| = \emptyset$, but the converse is not necessarily true. Let $\mathcal{M}_C$ be the field of meromorphic (or equivalently rational) functions on $C$. Then, one defines

$$L(D) = \{f \in \mathcal{M}_C; (f) \geq -D\}$$

This is a $\mathbb{C}$-vector space and it is not difficult to see that $L(D)$ has finite dimension. To every function $0 \neq f \in L(D)$, one can associate the effective divisor

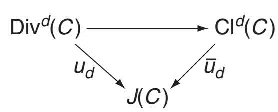
$$D_f = (f) + D \geq 0$$

Clearly, $D_f \sim D$ and every effective divisor with this property arises in this way. This gives a bijection

$$\mathbb{P}(L(D)) = |D|$$

showing that the complete linear system $|D|$ has the structure of a projective space. A linear system is a projective subspace of some complete linear system $|D|$.

Clearly, the map $u_d : C_d \to J(C)$ can be extended to the set $\mathrm{Div}^d(C)$ of degree $d$ divisors and Abel's theorem then states that this map factors through $\mathrm{Cl}^d(C)$, that is, that we have a commutative diagram



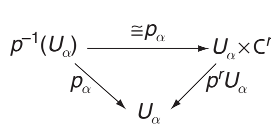

where $\overline{u}_d$ is injective.

**Theorem 3** (Jacobi's Inversion Theorem). *The map $u_d$ is surjective and hence induces an isomorphism*

$$\overline{u}_d : \mathrm{Cl}^d(C) \cong J(C)$$

It should be noted that the definition of the maps $u_d$ depends on the choice of a base point $P_0 \in C$. Hence, the maps $\overline{u}_d$ are not canonical, with the exception of the isomorphism $\overline{u}_0 : \mathrm{Cl}^0(C) \cong J(C)$ where the choice of $P_0$ drops out.

The concepts of divisors and linear systems can be rephrased in the language of line bundles. A (holomorphic) vector bundle on a complex manifold $M$ is a complex manifold $E$ together with a projection $p : E \to M$ which is a locally trivial $\mathbb{C}^r$-bundle. This means that an open covering $(U_\alpha)_{\alpha \in A}$ of $M$ and local trivializations




exist, such that the transition maps

$$\varphi_\beta \circ \varphi_\alpha^{-1}|_{(U_\alpha \cap U_\beta) \times \mathbb{C}^r} : \\ (U_\alpha \cap U_\beta) \times \mathbb{C}^r \to (U_\alpha \cap U_\beta) \times \mathbb{C}^r$$

are fiberwise linear isomorphisms. If $M$ is connected, then $r$ is constant and is called the rank of the vector bundle. A line bundle is simply a rank-1 vector bundle.

Alternatively, one can view vector bundles as locally free $\mathcal{O}_M$-modules, where $\mathcal{O}_M$ denotes the structure sheaf of holomorphic (or in the algebro-geometric setting regular) functions on $M$. An $\mathcal{O}_M$-module $\mathcal{E}$ is called locally free of rank $r$, if an open covering $(U_\alpha)_{\alpha \in A}$ of $M$ exists such that $\mathcal{E}|_{U_\alpha} \cong \mathcal{O}_{U_\alpha}^{\oplus r}$. The transition functions of a locally free sheaf can be used to define a vector bundle and vice versa, and hence the concepts of vector bundles and locally free sheaves can be used interchangeably. The open coverings $U_\alpha$ can be viewed either in the complex topology, or, if $M$ is an algebraic variety, in the Zariski topology, thus leading to either holomorphic vector bundles (locally free sheaves in the $\mathbb{C}$-topology) or algebraic vector bundles (locally free sheaves in the Zariski topology). Clearly, every algebraic vector bundle defines a holomorphic vector bundle. Conversely, on a projective variety $M$, Serre's GAGA theorem (géométrie algébriques et géométrie analytique), a vast generalization of Chow's theorem, states that there exists a bijection between the equivalence classes of algebraic and holomorphic vector bundles (locally free sheaves).

The Picard group $\mathrm{Pic}\, M$ is the set of all isomorphism classes of line bundles on $M$. The tensor product defines a group structure on $\mathrm{Pic}\, M$ where the neutral element is the trivial line bundle $\mathcal{O}_\mathcal{M}$ and the inverse of a line bundle $\mathcal{L}$ is its dual bundle $\mathcal{L}^*$, which is also denoted by $\mathcal{L}^{-1}$. For this reason, locally free sheaves of rank 1 are also called invertible sheaves.

We now return to the case of a compact Riemann surface (algebraic curve) $C$. The concept of line bundles and divisors can be translated into each other. If $D = \sum n_i P_i$ is a divisor on $C$ and $U$ an open set, then we denote by $D_U$ the restriction of $D$ to $U$, that is, the divisor consisting of all points $P_i \in U$ with multiplicity $n_i$. One then defines a locally free sheaf (line bundle) $\mathcal{L}(D)$ by

$$\mathcal{L}(D)(U) = \{f \in \mathcal{M}_C(U); (f) \geq -D_U\}$$

To see that this is locally free, it is enough to consider for each point $P_i$ a neighborhood $U_i$ on which a holomorphic function $t_i$ exists, which vanishes only at $P_i$ and there of order 1 (i.e., it is a local parameter near the point $P_i$). Then,

$$\mathcal{L}(D)(U_i) = t_i^{-n_i}\mathcal{O}_{U_i} \cong \mathcal{O}_{U_i}$$

This correspondence defines a map

$$\begin{aligned} \mathrm{Div}\, C &\to \mathrm{Pic}\, C \\ D &\mapsto \mathcal{L}(D) \end{aligned}$$

It is not hard to show that:

1. every line bundle $\mathcal{L} \in \operatorname{Pic} C$ is of the form $\mathcal{L} = \mathcal{L}(D)$ for some divisor $D$ on the curve C;
2. $D_1 \sim D_2 \Longleftrightarrow \mathcal{L}(D_1) \cong \mathcal{L}(D_2)$;
3. $\mathcal{L}(D_1) \otimes \mathcal{L}(D_2) \cong \mathcal{L}(D_1 + D_2)$; and
4. $\mathcal{L}(-D) \cong \mathcal{L}(D)^{-1}$.

Hence, there is an isomorphism of abelian groups

$$\mathrm{Cl}(C) \cong \operatorname{Pic} C$$

This correspondence allows to define the degree of a line bundle $\mathcal{L}$. In the complex analytic setting this can also be interpreted as follows. Let $\mathcal{O}_C^*$ be the sheaf of nowhere-vanishing functions. Using cocycles, one easily identifies

$$H^1(C, \mathcal{O}_C^*) \cong \operatorname{Pic} C$$

and the exponential sequence

$$0 \to \mathbb{Z} \to \mathcal{O}_C \overset{\exp}{\to} \mathcal{O}_C^* \to 0$$

induces an exact sequence

$$\begin{aligned} 0 &\to H^1(C, \mathbb{Z}) \to H^1(C, \mathcal{O}_C) \\ &\to H^1(C, \mathcal{O}_C^*) = \operatorname{Pic} C \to H^2(C, \mathbb{Z}) \end{aligned}$$

The last map in this exact sequence associates to each line bundle $\mathcal{L}$ its first Chern class $c_1(\mathcal{L}) \in H^2(C, \mathbb{Z}) \cong \mathbb{Z}$, which can be identified with the degree of $\mathcal{L}$. Hence, the subgroup $\operatorname{Pic}^0 C$ of degree 0 line bundles on $C$ is isomorphic to

$$\operatorname{Pic}^0 C \cong H^1(C, \mathcal{O}_C)/H^1(C, \mathbb{Z})$$

Altogether there are identifications

$$\operatorname{Pic}^0 C \cong \mathrm{Cl}^0 C \cong J(C)$$

## The Riemann–Roch Theorem and Applications

For every divisor $D$ on a compact Riemann surface $C$, the discussion of the preceding section shows that there is an identification of finite-dimensional vector spaces

$$L(D) = H^0(C, \mathcal{L}(D))$$

where $H^0(C, \mathcal{L}(D))$ is the space of global sections of the line bundle $\mathcal{L}(D)$. One defines

$$l(D) = \dim_{\mathbb{C}} L(D)$$

It is a crucial question in the theory of compact Riemann surfaces to study the dimension $l(D)$ as $D$ varies.

The canonical bundle $\omega_C$ of $C$ is defined as the dual of the tangent bundle of $C$. Its global sections are holomorphic 1-forms. Every divisor $K_C$ on $C$ with $\omega_C = \mathcal{L}(K_C)$ is called (a) canonical divisor. The canonical divisors are the divisors of the meromorphic 1-forms on C, whereas the effective canonical divisors correspond to the divisors of holomorphic 1-forms (here, we simply write a 1-form locally as $f(z)\,dz$ and define a divisor by taking the zeros, resp. poles of $f(z)$). By abuse of notation, we also denote the divisor class corresponding to canonical divisors by $K_C$. There is a natural identification

$$\mathbb{P}(H^0(C, \omega_C)) = |K_C|$$

For a divisor $D$, the index of speciality is defined by

$$i(D) = l(K_C - D) = \dim_{\mathbb{C}} L(K_C - D)$$

The linear system $|K_C - D|$ is called the adjoint system of $|D|$. A crucial role is played by the

**Theorem 4** (Riemann–Roch). *For any divisor D on a compact Riemann surface C of genus g, the equality*

$$l(D) - i(D) = \deg D + 1 - g \qquad [2]$$

*holds.*

This can also be written in terms of line bundles. If $\mathcal{L}$ is any line bundle, then we denote the dimension of the space of global sections by

$$h^0(\mathcal{L}) = \dim_{\mathbb{C}} H^0(C, \mathcal{L})$$

Then, the Riemann–Roch theorem can be written as

$$h^0(\mathcal{L}) - h^0(\omega_C \otimes \mathcal{L}^{-1}) = \deg \mathcal{L} + 1 - g \qquad [3]$$

This can be written yet again in a different way, if we use sheaf cohomology. By Serre duality, there is an isomorphism of cohomology groups

$$H^1(C, \mathcal{L}) \cong H^0(C, \omega_C \otimes \mathcal{L}^{-1})^*$$

and hence if we set

$$h^1(\mathcal{L}) = \dim_{\mathbb{C}} H^1(C, \mathcal{L})$$

then [3] reads

$$h^0(\mathcal{L}) - h^1(\mathcal{L}) = \deg \mathcal{L} + 1 - g \qquad [4]$$

Whereas [2] is the classical formulation of the Riemann–Roch theorem, formula [4] is the formulation which is more suitable for generalizations. From this point of view, the classical Riemann–Roch theorem is a combination of the cohomological formulation [4] together with Serre duality.

The Riemann–Roch theorem has been vastly generalized. This was first achieved by Hirzebruch who proved what is nowadays called the Hirzebruch–Riemann–Roch theorem for vector bundles on projective manifolds. A further generalization is due to Grothendieck, who proved a "relative" version involving maps between varieties. Nowadays, theorems like the Hirzebruch–Riemann–Roch theorem can be

viewed as special cases of the Atiyah–Singer index theorem for elliptic operators. The latter also contains the Gauss–Bonnet theorem from differential geometry as a special case. Moreover, Serre duality holds in much greater generality, namely for coherent sheaves on projective varieties.

Applying the Riemann–Roch theorem [3] to the zero divisor $D = 0$, resp. the trivial line bundle $\mathcal{O}_C$, one obtains

$$h^0(\omega_C) = g \qquad [5]$$

that is, the number of independent global holomorphic 1-forms equals the genus of the curve $C$. Similarly, for $D = K_C$, resp. $\mathcal{L} = \omega_C$, we find from [3] and [5] that

$$\deg K_C = 2g - 2$$

These relations show, how the Riemann–Roch theorem links analytic, resp. algebraic, invariants with the topology of the curve $C$.

Finally, if $\deg D > 2g - 2$, then $\deg(K_C - D) < 0$ and hence $i(D) = l(K_C - D) = 0$ and [2] becomes

$$l(D) = \deg D + 1 - g \quad \text{if } \deg D > 2g - 2$$

which is Riemann's original version of the theorem.

Classically, linear series arose in the study of projective embeddings of algebraic curves. For a nonzero effective divisor

$$D = \sum_{i=1}^{k} n_i P_i, \quad n_i > 0$$

the support of $D$ is defined by

$$\operatorname{supp}(D) = \{P_1, \ldots, P_k\}$$

A complete linear system $|D|$ is called base point free, if no point $P$ exists which is in the support of every divisor $D' \in |D|$. This is the same as saying that for every $P \in C$ a section $s \in H^0(C, \mathcal{L}(D))$ exists which does not vanish at $P$. Let $|D|$ be base point free and let $s_0, \ldots, s_n \in H^0(C, \mathcal{L}(D))$ be a basis of the space of sections. Then, one obtains a map

$$\varphi_{|D|} : C \to \mathbb{P}(H^0(C, \mathcal{L}(D))) = \mathbb{P}^n$$
$$P \mapsto (s_0(P) : \ldots : s_n(P))$$

The divisors $D' \in |D|$ are then exactly the pullbacks of the hyperplanes $H$ of $\mathbb{P}^n$ under the map $\varphi_{|D|}$. Note that the map $\varphi_{|D|}$ as defined here depends on the choice of the basis $s_0, \ldots, s_n$, but any two such choices only differ by an automorphism of $\mathbb{P}^n$. We say that $|D|$, resp. the associated line bundle $\mathcal{L} = \mathcal{L}(D)$, is very ample if $\varphi_{|D|}$ defines an embedding. Using the Riemann–Roch theorem, it is not difficult to prove:

**Proposition 1** *Let $D$ be a divisor of degree $d$ on the curve $C$. Then*

(i) *$|D|$ is base point free if $d \geq 2g$ and*
(ii) *$|D|$ is very ample if $d \geq 2g + 1$.*

If the genus $g(C) \geq 2$, then one can prove that $|K_C|$ is base point free and consider the canonical map

$$\varphi_{|K_C|} : C \to \mathbb{P}^{g-1}$$

A curve $C$ is called hyperelliptic if there exists a surjective map $f : C \to \mathbb{P}^1$ which is a covering of degree 2. In genus 2 every curve is hyperelliptic, whereas for genus $g \geq 3$ hyperelliptic curves are special. The connection with the canonical map is given by

**Theorem 5** (Clifford). *Let $C$ be a curve of genus $g \geq 2$. Then the canonical map is an embedding if and only if $C$ is not hyperelliptic.*

We end this section by stating Hurwitz's theorem: Let $f : C \to D$ be a surjective holomorphic map between compact Riemann surfaces (if $f$ is not constant then it is automatically surjective). Then, near a point $P \in C$ the map $f$ is given in local analytic coordinates by $f(t) = t^{n_P}$ and we call $f$ "ramified" of order $n_P$ if $n_P > 1$. The ramification divisor of $f$ is defined as

$$R = \sum_{P \in C} (n_P - 1)P$$

Note that this is a finite sum. If we define

$$f^*(Q) = \sum_{P \in f^{-1}(Q)} n_P P$$

then one can show that

$$\deg f = \deg f^*(Q) = \sum_{P \in f^{-1}(Q)} n_P$$

is independent of the point $Q$. This number is called the degree of the map $f$. (This should not be confused with the degree $\deg(f)$ of the principal divisor $(f)$ defined by $f$.) In fact, applying the above equality to the map $f : C \to \mathbb{P}^1$ associated to a nonconstant meromorphic function $f$ shows that the degree of the principal divisor $(f)$ is zero, since

$$\deg(f) = \deg f^*(0) - \deg f^*(\infty) = 0$$

**Theorem 6** (Hurwitz). *Let $f : C \to D$ be a surjective holomorphic map between compact Riemann surfaces of genus $g(C)$ and $g(D)$, respectively. Then,*

$$2g(C) - 2 = \deg f \cdot (2g(D) - 2) + \deg R$$

*where $R$ is the ramification divisor.*

## Brill–Noether Theory

In this section, we state the main results of Brill–Noether theory. For a divisor $D$ on a curve $C$ we denote by

$$r(D) = l(D) - 1$$

the projective dimension of the complete linear system $|D|$. The principal objects of Brill–Noether theory are the sets $W_d^r \subset \mathrm{Cl}^d(C) = \mathrm{Pic}^d(C)$ given by

$$W_d^r(C) = \{D; \deg D = d, r(D) \geq r\}$$

These sets are subvarieties of $\mathrm{Cl}^d(C) = \mathrm{Pic}^d(C)$.

We denote by $g_d^r$ a linear system (not necessarily complete) of degree $d$ and projective dimension $r$. Closely related to the varieties $W_d^r$ are the sets

$$G_d^r(C) = \{\delta; \delta \text{ is a } g_d^r \text{ on } C\}$$

These sets also have a natural structure as a projective variety. Clearly, there are maps $G_d^r(C) \to W_d^r(C)$.

If $g = g(C)$ is the genus of the curve $C$, then the Brill–Noether number is defined as

$$\rho(g, r, d) = g - (r + 1)(g - d + r)$$

Its significance is that it is the expected dimension of the varieties $G_d^r(C)$. The two basic results of Brill–Noether theory are:

**Theorem 7** (Existence Theorem). *Let $C$ be a curve of genus $g$. Let $d, r$ be integers such that $d \geq 1$, $r \geq 0$, and $\rho(g, r, d) \geq 0$. Then $G_d^r(C)$ and hence $W_d^r(C)$ are nonempty and every component of $G_d^r(C)$ has dimension at least $\rho$. If $r \geq d - g$, then the same is true for $W_d^r(C)$.*

**Theorem 8** (Connectedness Theorem). *Let $C$ be a curve of genus $g$ and $d, r$ integers such that $d \geq 1$, $r \geq 0$, and $\rho(g, r, d) \geq 1$. Then $G_d^r(C)$ and hence also $W_d^r(C)$ are connected.*

The above theorems hold for all curves $C$. There are other theorems which only hold for general curves (where general means outside a countable union of proper subvarieties in the moduli space, see the section "Moduli of compact Riemann surfaces").

**Theorem 9** (Dimension Theorem). *Let $C$ be a general curve of genus $g$ and $d \geq 1$, $r \geq 0$ integers. If $\rho(g, r, d) < 0$, then $G_d^r(C) = \emptyset$. If $\rho \geq 0$, then every component of $G_d^r(C)$ has dimension $\rho$.*

**Theorem 10** (Smoothness Theorem). *Let $C$ be a general curve of genus $g$ and $d \geq 1$, $r \geq 0$. Then, $G_d^r(C)$ is smooth of dimension $\rho$. If $\rho \geq 1$, then $G_d^r(C)$ and hence $W_d^r(C)$ are irreducible.*

Brill–Noether theory started with a paper of Brill and Noether in 1873. It was, however, only from the 1970s onwards that the main theorems could be proved rigorously, due to the work of Griffiths, Harris, Kleiman, Mumford, and many others. For an extensive treatment of the theory, as well as a list of references, the reader is referred to Arbarello *et al.* (1985).

## Green's Conjecture

In recent years, much progress was achieved in understanding the equations of canonical curves. If the curve $C$ is not hyperelliptic, then the canonical map $\varphi_{|K_C|} : C \to \mathbb{P}^{g-1}$ defines an embedding. We shall, in this case, identify $C$ with its image in $\mathbb{P}^{g-1}$ and call this a canonical curve. The Clifford index (for a precise definition see Lazarsfeld (1989)) is a first measure of how special a curve $C$ is with respect to the canonical map. Hyperelliptic curves, where the canonical map fails to be an embedding, have, by definition, Clifford index 0. The two next special cases are plane quintic curves (they have a $g_5^2$) and trigonal curves. A curve $C$ is called trigonal, if there is a $3:1$ map $C \to \mathbb{P}^1$, in which case $C$ has a $g_3^1$. More generally, the gonality of a curve $C$ is the minimal degree of a surjective map $C \to \mathbb{P}^1$. Plane quintics and trigonal curves are precisely the curves which have Clifford index 1.

**Theorem 11** (Enriques–Babbage). *If $C \subset \mathbb{P}^{g-1}$ is a canonical curve, then $C$ is either defined by quadratic equations, or it is trigonal or isomorphic to a plane quintic curve (i.e., it has Clifford index 1).*

One can now ask more refined questions about the equations defining canonical curves and the relations (syzygies) among these equations. This leads to looking at the minimal free resolution of a canonical curve $C$, which is of the form

$$0 \leftarrow \mathcal{I}_C \leftarrow \oplus_j \mathcal{O}_{\mathbb{P}^{g-1}}(-j)^{\beta_{0j}} \leftarrow \cdots \leftarrow \oplus_j \mathcal{O}_{\mathbb{P}^{g-1}}(-j)^{\beta_{kj}} \leftarrow 0$$

Here, $\mathcal{I}_C$ is the ideal sheaf of $C$ and $\mathcal{O}_{\mathbb{P}^{g-1}}(n)$ is the $n$th power of the dual of the Hopf bundle (or tautological sub-bundle) on $\mathbb{P}^{g-1}$ if $n \geq 0$, resp. the $|n|$th power of the Hopf bundle if $n < 0$. The $\beta_{ij}(C)$ are called the Betti numbers of $C$. The Green conjecture predicts a link between the nonvanishing of certain Betti numbers and geometric properties of the canonical curve, such as the existence of multisecants. Recently, C Voisin and M Teixidor have proved the Green conjecture for general curves of given gonality (see Beauville (2003)).

## Moduli of Compact Riemann Surfaces

As a set, the moduli space of compact Riemann surfaces of genus $g$ is defined as

$$\mathcal{M}_g = \{C; C \text{ is a compact Riemann surface of genus } g\}/\cong$$

For genus $g = 0$, the only Riemann surface is the Riemann sphere $\hat{\mathbb{C}} = \mathbb{P}^1$ and hence $\mathcal{M}_0$ consists of one point only. Every Riemann surface of genus 1 is a torus

$$E = \mathbb{C}/L$$

for some lattice $L$, which can be written in the form

$$L_\tau = \mathbb{Z}\tau + \mathbb{Z}, \quad \operatorname{Im}\tau > 0$$

Two elliptic curves $E_\tau = \mathbb{C}/L_\tau$ and $E_{\tau'} = \mathbb{C}/L_{\tau'}$ are isomorphic if and only if a matrix

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{Z})$$

exists with

$$\tau' = \frac{a\tau + b}{c\tau + d}$$

This proves that

$$\mathcal{M}_1 = \mathbb{H}_1/\mathrm{SL}(2, \mathbb{Z})$$

and this construction also shows that $\mathcal{M}_1$ can itself be given the structure of a Riemann surface. Using the $j$-function, one obtains that

$$\mathcal{M}_1 \cong \mathbb{C}$$

The situation is considerably more complicated for genus $g \geq 2$. The space of infinitesimal deformations of a curve $C$ is given by $H^1(C, T_C)$ where $T_C$ is the tangent bundle. By Serre duality

$$H^1(C, T_C) \cong H^0(C, \omega_C^{\otimes 2})^*$$

and by Riemann's theorem it then follows that

$$\dim H^1(C, T_C) = \dim H^0(C, \omega_C^{\otimes 2}) = 3g - 3$$

This shows that a curve of genus $g$ depends on $3g - 3$ parameters or moduli, a dimension count which was first performed by Riemann.

In genus 2 every curve has the hyperelliptic involution, and for a general curve of genus 2 this is the only automorphism. In genus $g \geq 3$ the general curve has no automorphisms, but some curves do. The order of the automorphism group is bounded by $84(g - 1)$. The existence of automorphism for some curves means that $\mathcal{M}_g$ is not a manifold, but has singularities. The singularities are, however, fairly mild. Locally, $\mathcal{M}_g$ always looks like $\mathbb{C}^{3g-3}/G$ near the origin, where $G$ is a finite group acting linearly on $\mathbb{C}^{3g-3}$. One expresses this by saying that $\mathcal{M}_g$ has only finite quotient singularities. A space with this property is also sometimes referred to as a $V$-manifold or an orbifold. Moreover, $\mathcal{M}_g$ is a quasiprojective variety, that is, a Zariski-open subset of a projective variety. As the above parameter count implies, the dimension of $\mathcal{M}_g$ is $3g - 3$. At this point it can also be clarified what is meant by a general curve in the context of Brill–Noether theory: a property is said to hold for the general curve in Brill–Noether theory if it holds outside a countable number of proper subvarieties of $\mathcal{M}_g$.

It is often useful to work with projective, rather than quasiprojective, varieties. This means that one wants to compactify $\mathcal{M}_g$ to a projective variety $\overline{\mathcal{M}}_g$, preferably in such a way that the points one adds still correspond to geometric objects. The crucial concept in this context is that of a stable curve. A stable curve of genus $g$ is a one-dimensional projective variety with the following properties:

1. $C$ is connected (but not necessarily irreducible),
2. $C$ has at most nodal singularities (i.e., two local analytic branches meet transversally),
3. the arithmetic genus $p_a(C) = h^1(C, \mathcal{O}_C) = g$, and
4. the automorphism group $\mathrm{Aut}(C)$ of $C$ is finite.

The last of these conditions is equivalent to the following: if a component of $C$ is an elliptic curve, then this must either meet another component or have a node, and if a component is a rational curve, then this component must either have at least two nodes or one node and intersect another component, or it is smooth and has at least three points of intersection with other components.

It should be noted that, in contrast to the previous illustrations, Figure 5 is drawn from the complex point of view, that is, the curves appear as one-dimensional objects.

The concept of stable curves leads to what is generally known as the Deligne–Mumford compactification of $\mathcal{M}_g$:

$$\overline{\mathcal{M}}_g = \{C; C \text{ is a stable curve of genus } g\}/\cong$$

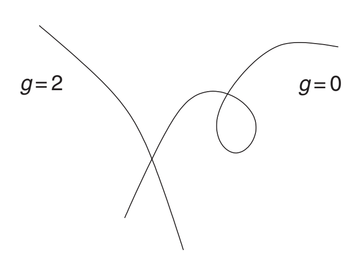



**Figure 5** An example of a stable curve of genus 3.

**Theorem 12** (Deligne–Mumford, Knudsen). *$\overline{\mathcal{M}}_g$ is an irreducible, projective variety of dimension $3g-3$ with only finite quotient singularities.*

The spaces $\overline{\mathcal{M}}_g$ have been studied intensively over the last 30 years. From the point of view of classification, an important question is to determine the Kodaira dimension of these spaces.

**Theorem 13** (Harris–Mumford, Eisenbud–Harris). *The moduli spaces $\overline{\mathcal{M}}_g$ are of general type for $g > 23$.*

On the other hand, it is known that $\overline{\mathcal{M}}_g$ is rational for $g \leq 6$, unirational for $g \leq 14$, and has negative Kodaira dimension for $g \leq 16$.

A further topic is to understand the cohomology of $\overline{\mathcal{M}}_g$, resp. the Chow ring, and to compute the intersection theory on $\overline{\mathcal{M}}_g$. For these topics we refer the reader to Vakil (2003).

Closely related is the moduli problem of stable $n$-pointed curves. A stable $n$-pointed curve (**Figure 6**) is an $(n+1)$-tuple $(C, x_1, \ldots, x_n)$, where $C$ is a connected nodal curve and $x_1, \ldots, x_n$ are smooth points of $C$ with the stability condition that the automorphism group of $(C, x_1, \ldots, x_n)$ is finite. These curves can be parametrized by a coarse moduli space $\overline{\mathcal{M}}_{g,n}$. These spaces share many properties of the spaces $\overline{\mathcal{M}}_g$: they are irreducible, projective varieties with finite quotient singularities and of dimension $3g-3+n$.

A further development, which has become very important in recent years, is that of moduli spaces of stable maps. These were introduced by Kontsevich in the context of quantum cohomology. To define stable maps, one first fixes a projective variety $X$ and then considers $(n+2)$-tuples $(C, x_1, \ldots, x_n, f)$ where $(C, x_1, \ldots, x_n)$ is an $n$-pointed curve of genus $g$ and $f: C \to X$ a map. The stability condition is, that this object allows only finitely many automorphisms $\varphi: C \to C$, fixing the marked points $x_1, \ldots, x_n$, such that $f \circ \varphi = f$. In order to obtain meaningful moduli spaces, one also fixes a class $\gamma \in H_2(X, \mathbb{Z})$. One then asks for a space parametrizing all stable $(n+2)$-tuples $(C, x_1, \ldots, x_n, f)$ with the additional property that $f_*[C] = \gamma$. This construction is best treated in the language of stacks, and one can show that this moduli problem gives rise to a proper Deligne–Mumford stack $\overline{\mathcal{M}}_{g,n}(X, \gamma)$. In general, this stack is very complicated, it need not be connected, can be very singular, and may have several components of different dimensions. Its expected dimension is

$$\begin{aligned}\exp.\dim \overline{\mathcal{M}}_{g,n}(X,\gamma) \\ = (\dim X - 3)(1-g) + n + \int_\gamma c_1(T_X)\end{aligned}$$

Quantum cohomology can now be rephrased as intersection theory on the stack $\overline{\mathcal{M}}_{g,n}(X, \gamma)$. In general, these stacks do not have the expected dimension. For this reason, Behrend and Fantechi (1997) have constructed a virtual fundamental class of the right dimension, which is the correct tool for the intersection theory which gives the algebro-geometric definition of quantum cohomology. In addition to this, there is also a symplectic formulation. It was shown by B Siebert that both approaches coincide.

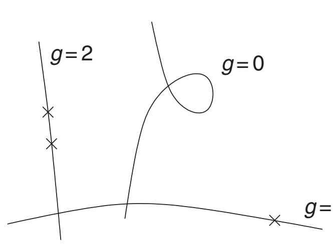



**Figure 6** An example of marked stable curve.

## Verlinde Formula and Conformal Blocks

The study of vector bundles (locally free sheaves) on a compact Riemann surface is an area of research in its own right. For a rank-$r$ bundle $\mathcal{E}$, the slope of $\mathcal{E}$ is defined by

$$\mu(\mathcal{E}) = \frac{\deg \mathcal{E}}{r}$$

where the degree of $\mathcal{E}$ is defined as the degree of the line bundle $\bigwedge^r \mathcal{E} = \det \mathcal{E}$. The bundle $\mathcal{E}$ is called stable, resp. semistable, if

$$\mu(\mathcal{F}) < \mu(\mathcal{E}), \quad \text{resp.} \quad \mu(\mathcal{F}) \leq \mu(\mathcal{E})$$

for every proper sub-bundle $\{0\} \subsetneq \mathcal{F} \subsetneq \mathcal{E}$. Let $C$ be a compact Riemann surface of genus $g \geq 2$ and let $\mathrm{SU}_C(r)$ be the moduli space of semistable rank-$r$ vector bundles with trivial determinant $\det \mathcal{E} = \mathcal{O}_C$. This is a projective variety of dimension $(r^2-1)(g-1)$. It contains a smooth open set, whose points correspond to the isomorphism classes of stable vector bundles. The complement of this set is in general the singular locus of $\mathrm{SU}_C(r)$ and its points correspond to direct sums of line bundles of degree 0. These are the so-called graded objects of the semistable, but not stable, bundles. By a theorem of Narasimhan and Seshadri, the points of $\mathrm{SU}_C(r)$ are also in one-to-one correspondence with the isomorphism classes of representations $\pi_1(C) \to \mathrm{SU}(r)$.

Let $L \in \mathrm{Pic}^{g-1}(C)$ be any line bundle of degree $g-1$ on C. Then, the set

$$\Theta_L = \left\{\mathcal{E} \in \mathrm{SU}_C(r); \dim H^0(C, \mathcal{E} \otimes L) > 0\right\}$$

is a Cartier divisor on $SU_C(r)$ and thus defines a line bundle $\mathcal{L}$ on $SU_C(r)$. This is a natural generalization of the construction of the classical theta divisor. The line bundle $\mathcal{L}$ generates the Picard group of the moduli space $SU_C(r)$.

**Theorem 14** (Verlinde Formula). *If C has genus g and k is a positive integer, then*

$$\dim H^0(SU_C(r), \mathcal{L}^k) = \left(\frac{r}{r+k}\right)^g \sum_{\substack{S \sqcup T = \{1,\ldots,r+k\} \\ |S|=r}} \prod_{\substack{s\in S \\ t \in T}} \left| \sin \pi \frac{s-t}{r+k} \right|^{g-1}$$

This formula was first found by Verlinde in the context of conformal field theory. Due to this relationship, the spaces $H^0(SU_C(r), \mathcal{L}^k)$ are also called conformal blocks. These spaces can also be defined for principal bundles. Rigorous proofs for the general case of the Verlinde formula are due to Beauville–Laszlo and Faltings. For a survey, see Beauville (1995).

*See also:* Characteristic Classes; Cohomology Theories; Index Theorems; Mirror Symmetry: a Geometric Survey; Moduli Spaces: An Introduction; Polygonal Billiards; Several Complex Variables: Basic Geometric Theory; Several Complex Variables: Compact Manifolds; Topological Gravity, Two-Dimensional.

## Further Reading


Arbarello E, Cornalba M, Griffiths Ph, and Harris J (1985) *Geometry of Algebraic Curves*, vol. I. New York: Springer.

Beauville A (1995) Vector bundles on curves and generalized theta functions: recent results and open problems. In: Boutet de Montel A and Morchenko V (eds.) *Current Topics in Complex Algebraic Geometry*, (Berkeley, CA, 1992/93), Math. Sci. Res. Inst. Publ., vol. 28, pp. 17–33. Cambridge: Cambridge University Press.

Beauville A (2003) La conjecture de green générique [d'après C. Voisin]. Exposé 924 du Séminaire Bourbaki.

Behrend K and Fantechi B (1997) The intrinsic normal cone. *Inventiones Mathematicae* 128: 45–88.

Faber C and Looijenga E (1999) Remarks on moduli of curves. In: Faber C and Looijenga E (eds.) *Moduli of Curves and Abelian Varieties*, Aspects Math. E33, pp. 23–45. Braunschweig: Vieweg.

Farkas H and Kra I (1992) *Riemann Surfaces*, 2nd edn. New York: Springer.

Forster O (1991) *Lectures on Riemann Surfaces* (translated from the 1977 German Original by Bruce Gilligan), Reprint of the 1981 English Translation. New York: Springer.

Griffiths Ph and Harris J (1994) *Principles of Algebraic Geometry*, Reprint of the 1978 Edition, Wiley Classics Library. New York: Wiley.

Hartshorne R (1977) *Algebraic Geometry*. Heidelberg: Springer.

Jost J (1997) *Compact Riemann Surfaces. An Introduction to Contemporary Mathematics* (translated from the German Manuscript by Simha RR). Berlin: Springer.

Kirwan F (1992) *Complex Algebraic Curves*. Cambridge: Cambridge University Press.

Lazarsfeld R (1989) A sampling of vector bundle techniques in the study of linear series. In: Cornalba M, Gomez-Mont X, and Verjovsk (eds.) *Lectures on Riemann Surfaces*, pp. 500–559. Teaneck, NJ: World Scientific.

Miranda R (1995) *Algebraic Curves and Riemann Surfaces*. Providence: American Mathematical Society.

Mumford D (1995) *Algebraic Geometry. I. Complex Projective Varieties*, Reprint of the 1976 Edition. Berlin: Springer.

Vakil R (2003) The moduli space of curves and its tautological ring. *Notices of the American Mathematical Society* 50(6): 647–658.

Weyl H (1997) *Die Idee der Riemannschen Fläche* (Reprint of the 1913 German Original, With Essays by Reinhold Remmert, Michael Schneider, Stefan Hildebrandt, Klaus Hulek and Samuel Patterson. Edited and with a Preface and a Biography of Weyl by Remmert). Stuttgart: Teubner.


# Riemann–Hilbert Methods in Integrable Systems

**D Shepelsky**, Institute for Low Temperature Physics and Engineering, Kharkov, Ukraine



## Introduction

The Riemann–Hilbert (RH) method in mathematical physics and analysis consists in reducing a particular problem to the problem of reconstruction of an analytic, scalar- or matrix-valued function in the complex plane from a prescribed jump across a given curve. More precisely, let an oriented contour $\Sigma$ be given in the complex $\lambda$-plane. The contour $\Sigma$ may have points of self-intersections, and it may consist of several connected components; typical contours appearing in applications to integrable systems are shown in Figure 1.

The orientation of an arc in $\Sigma$ defines the $+$ and the $-$ side of $\Sigma$. Suppose in addition that we are given a map $v : \Sigma \to GL(N, \mathbb{C})$ with $v, v^{-1} \in L^\infty(\Sigma)$. The (normalized) RH problem determined by the pair $(\Sigma, v)$ consists in finding an $N \times N$

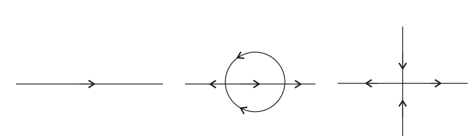


**Figure 1** Typical contours for RH problems.

matrix-valued function $m(\lambda)$ with the following properties:

$$m(\lambda) \text{ is analytic in } \mathbb{C}\backslash\Sigma \quad [1a]$$

$$m_+(\lambda) = m_-(\lambda)v(\lambda) \text{ for } \lambda \in \Sigma \text{ where } m_+(\lambda)(m_-(\lambda)) \text{ is the limit of } m \text{ from the } +(-) \text{ side of } \Sigma \quad [1b]$$

$$m(\lambda) \to I \text{ (identity matrix) as } \lambda \to \infty \quad [1c]$$

The precise sense in which the limit at $\infty$ and the boundary values $m_\pm$ are attained are technical matter that should be specified for each given RH problem $(\Sigma, v)$.

Concerning the name RH problem we note that in literature (particularly, in the theory of boundary values of analytic functions), the problem of reconstructing a function from its jump across a curve is often called the Hilbert boundary-value problem. The closely related problem of analytic matrix factorization (given $\Sigma$ and $v$, find $G(\lambda)$ analytic and nondegenerate in $\mathbb{C}\backslash\Sigma$ such that $G_+G_- = v$ on $\Sigma$) is sometimes called the Riemann problem. The name "RH problem" is also attributed to the reconstruction of a Fuchsian system with given poles and a given monodromy group.

In applications, the jump matrix $v$ also depends on certain parameters, in which the original problem at hand is naturally formulated (e.g., $v = v(\lambda; x, t)$ in applications to the integrable nonlinear differential equations in dimension $1 + 1$, with $x$ being the space variable and $t$ the time variable), and the main concern is the behavior of the solution of the RH problem, $m(\lambda; x, t)$, as a function of $x$ and $t$. Particular interest is in the behavior of $m(\lambda; x, t)$ as $x$ and $t$ become large.

In the scalar case, $N = 1$, rewriting the original multiplicative jump condition in the additive form

$$\log m_+(\lambda) = \log m_+(\lambda) + \log v(\lambda)$$

and using the Cauchy–Plemelj–Sokhotskii formula give an explicit integral representation for the solution

$$m(\lambda) = \exp\left\{\frac{1}{2\pi \mathrm{i}}\int_\Sigma \frac{\log v(\mu)}{\mu - \lambda}\mathrm{d}\mu\right\} \quad [2]$$

(in the case of nonzero index, $\Delta \log v|_\Sigma \neq 0$, formula [2] admits a suitable modification).

A generic (nonabelian) matrix RH problem cannot be solved explicitly in terms of contour integrals; however, it can always be reduced to a system of linear singular-integral equations, thus linearizing an originally nonlinear system.

The main benefit of reducing an originally nonlinear problem to the analytic factorization of a given matrix function arises in asymptotic analysis. Typically, the dependence of the jump matrix on the external parameters (say, $x$ and $t$) is oscillatory. In analogy of asymptotic evaluation of oscillatory contour integrals via the classical method of steepest descent, in the asymptotic evaluation of the solution $m(\lambda; x, t)$ of the matrix RH problem as $x, t \to \infty$, the nonlinear steepest-descent method examines the analytic structure of the jump matrix $v(\lambda; x, t)$ in order to deform the contour $\Sigma$ to contours where the oscillatory factors become exponentially small as $x, t \to \infty$, and hence the original RH problem reduces to a collection of local RH problems associated with the relevant points of stationary phase. Although the method has (in the matrix case) noncommutative and nonlinear elements, the final result of the analysis is as efficient as the asymptotic evaluation of the oscillatory integrals.

## Dressing Method

The RH method allows describing the solution of a differential system independently of the theory of differential equations. The solution might be explicit, that is, given in terms of elementary or elliptic or abelian functions and contour integrals of such functions. In general (transcendental) case, the solution can be represented in terms of the solution of certain linear singular integral equations.

In the modern theory of integrable systems, a system of nonlinear differential equations is often called integrable if it can be represented as a compatibility condition of an auxiliary overdetermined linear system of differential equations called a Lax pair of the given nonlinear system (actually it might involve more than two linear equations). In order that the compatibility condition represents a nontrivial nonlinear system of equations, the Lax pair is required to depend rationally on an auxiliary parameter (called a spectral parameter). The RH problem formulated in the complex plane of the spectral parameter allows, given a particular solution of the compatibility equations, to construct directly new solutions of the compatibility system by "dressing" the initial one.

For example, let $D(x, \lambda), x \in \mathbb{R}^n, \lambda \in \mathbb{C}$ be an $N \times N$ diagonal, polynomial in $\lambda$ with smooth coefficients, function such that $a_j := \partial D/\partial x_j$ are polynomials in $\lambda$ of degree $d_j$. Then $\Psi_0 := \exp D(x, \lambda)$ solves the system of linear equations $\partial\Psi_0/\partial x_j = a_j\Psi_0$, whose compatibility conditions $\partial^2\Psi_0/\partial x_j\partial x_k = \partial^2\Psi_0/\partial x_k\partial x_j$ are trivially satisfied. Given a contour $\Sigma$ and a smooth function $v$, consider the matrix RH problem [1]

with the jump matrix $\tilde{v}(\lambda; x) := \exp D(x, \lambda) v(\lambda) \exp -D(x, \lambda)$. Let $m(\lambda; x)$ be the solution of this RH problem. Then $(D_j m)_+ = (D_j m)_- \tilde{v}$, where $D_j f := \partial f / \partial x_j + [a_j, f]$ with $[a, b] := ab - ba$. The Liouville theorem implies that $(D_j m) m^{-1}$ is an entire function which is $o(\lambda^{d_j})$ as $\lambda \to \infty$. Setting $\Psi(x, \lambda) := m(\lambda; x) \exp D(x, \lambda)$ gives the system of linear equations

$$\frac{\partial \Psi}{\partial x_j} = a_j + \sum_{k<d_j} \lambda^k q_{jk}(x) \equiv R_j(x, \lambda)\Psi \qquad [3]$$

the compatibility conditions for which are

$$\frac{\partial R_k}{\partial x_j} - \frac{\partial R_j}{\partial x_k} = [R_j, R_k] \qquad [4]$$

Equating coefficients of various powers of $\lambda$ in [4] gives a (generally) nonlinear system of partial differential equations for the coefficient matrices $q_{jk}$. Thus, given $D(x, \lambda)$, the RH problem, if it is solvable, maps the pair $(\Sigma, v)$ to solutions of [4].

Specializing to $n = 2$ with variables $(x, t) \in \mathbb{R}^2$, the overdetermined system of linear equations and the corresponding compatibility conditions are

$$\Psi_x = U\Psi, \qquad \Psi_t = V\Psi \qquad [5]$$

and

$$U_t - V_x + [U, V] = 0 \qquad [6]$$

respectively. Conditions [6] are sometimes called the zero-curvature conditions.

Equations [5] and [6] with $U$ and $V$ depending rationally on the spectral parameter $\lambda$ represents the integrable nonlinear systems in $1 + 1$ dimension. A typical example of such a system is the (defocusing) nonlinear Schrödinger (NLS) equation

$$\mathrm{i}q_t + q_{xx} - 2|q|^2 q = 0 \qquad [7]$$

Starting from the RH problem with the $2 \times 2$ jump matrix

$$v(\lambda; x, t) = \mathrm{e}^{\mathrm{i}\theta\sigma_3/2} v(\lambda) \mathrm{e}^{-\mathrm{i}\theta\sigma_3/2} \qquad [8]$$

where $\theta(\lambda; x, t) = -t\lambda^2 + x\lambda$, $\sigma_3 = \mathrm{diag}\{1, -1\}$, and $v(\lambda)$ satisfies the involution $\sigma_3 v^*(\lambda)\sigma_3 = v(\bar{\lambda})$, expanding out the limit of the solution of the RH problem as $\lambda \to \infty$

$$m(\lambda; x, t) = I + \frac{m_1(x, t)}{\lambda} + o\left(\frac{1}{\lambda}\right) \qquad [9]$$

and arguing as above gives [5], with

$$U = \frac{\mathrm{i}\lambda\sigma_3}{2} + \begin{pmatrix} 0 & q \\ \bar{q} & 0 \end{pmatrix} \qquad [10]$$

and $q = -\mathrm{i}(m_1)_{12}$, whereas the compatibility condition [6] reduces to [7].

The relation between the RH problem and the differential equations [5] is local in $x$ and $t$; it is based only on the unique solvability of the RH problem, the Liouville theorem, and the explicit dependence of the jump matrix in $x$ and $t$. The uniqueness of the solution of an RH problem is basically provided by the Liouville theorem: the ratio $m^{(1)}(m^{(2)})^{-1}$ of any two solutions is analytic in $\mathbb{C}\backslash\Sigma$ and continuous across $\Sigma$ and is therefore identically equal to $I$ by the normalization condition [1c].

On the other hand, there are no completely general effective criteria for the solvability. Nevertheless, many RH problems seen in applications to integrable systems satisfy the following sufficient condition: if $\Sigma$ is symmetric with respect to $\mathbb{R}$ and contains $\mathbb{R}$, and if, in addition, $v^*(\lambda) = v(\bar{\lambda})$ for $\lambda \in \Sigma\backslash\mathbb{R}$ and $\mathrm{Re}\, v(\lambda) > 0$ for $\lambda \in \mathbb{R}$, then the RH problem is solvable.

For nonlinear equations supporting solitons, the RH problem appears naturally in a more general setting, as a meromorphic factorization problem, where $m$ in [1] is sought to be a (piecewise) meromorphic function, with additionally prescribed poles and respective residue conditions. Alternatively, in the Riemann factorization problem $G_+ G_- = v$, one assumes that $G$ degenerates at some given points $\lambda_1, \ldots, \lambda_n \in \Omega^+$ and $\mu_1, \ldots, \mu_n \in \Omega^-$, where $\mathbb{C} = \Omega^+ \cup \Omega^- \cup \Sigma$, and prescribes two sets of subspaces, $\mathrm{Im}\, G|_{\lambda=\lambda_j}$ and $\mathrm{Ker}\, G|_{\lambda=\mu_j}$. In the case $v \equiv I$, the solution of the factorization problem with zeros (meromorphic RH problem) is purely algebraic, and gives formulas describing multisoliton solutions. In the general case, $v \not\equiv I$, the meromorphic RH problem can be algebraically converted to a holomorphic RH problem, by subsequently removing the poles with the help of the Blaschke–Potapov factors.

Alternatively, a meromorphic RH problem can be converted to a holomorphic one by adding to $\Sigma$ an additional contour $\Sigma_{\mathrm{aux}}$ enclosing all the poles, interpolating the constants involved in the residue conditions inside the region surrounded by $\Sigma_{\mathrm{aux}}$, and defining a new jump matrix on $\Sigma_{\mathrm{aux}}$ using the interpolant and the Blaschke–Potapov factors.

RH problems formulated on the complex plane $\mathbb{C}$ correspond typically to solutions of relevant nonlinear problems decaying at infinity. For other types of boundary conditions (e.g., nonzero constants or periodic or quasiperiodic boundary conditions), the corresponding RH problem is naturally formulated on a Riemann surface. For example, the RH problem associated with finite density conditions $q(x, t) \to \rho \mathrm{e}^{\mathrm{i}\phi_\pm}$ as $x \to \pm\infty$ for the NLS equation [7] is naturally formulated on the two-sheet Riemann surface of the function $k(\lambda) = \sqrt{\lambda^2 - 4\rho^2}$ with

the contour $\Sigma$ consisting of the points $(\lambda, \varepsilon)$, where $|\lambda| \geq 2\rho$ and $\varepsilon = \pm 1$ marks the surface sheet.

## Inverse-Scattering Transform

The inverse-scattering transform method for solving initial-value problems for integrable nonlinear equations written as the compatibility conditions [6] for linear equations [5] consists in the following: starting from the given initial data, solve the direct problem, that is, determine appropriate eigenfunctions (solutions of the differential $x$-equation in the Lax pair [5]) having well-controlled analytic properties as functions of the auxiliary (spectral) parameter $\lambda$ and the associated spectral functions of $\lambda$; then, by virtue of the $t$-equations in the Lax pair [5], the associated functions evolve in a simple, explicit way. Finally, using the explicit evolution of the spectral functions, solve the inverse problem of finding the associated coefficients in the $x$-equation, which, by [5], evolve according to the given nonlinear equation and thus solve the Cauchy problem for this equation. The last step in this procedure, the inverse-scattering problem, can be effectively solved by reformulating it as an RH problem, which in turn can be related to a system of singular integral equations. The classical Gelfand–Levitan–Marchenko integral equation of the inverse-scattering problem is the Fourier transform of some special cases of these singular integral equations.

To fix ideas, consider the initial-value problem for the NLS equation [7], where the data $q(x, t=0) = q_0(x)$ have sufficient smooth and decay as $|x| \to \infty$. For each $\lambda \in \mathbb{C}\backslash\mathbb{R}$, one constructs solutions $\Psi(x, \lambda)$ of $\Psi_x = U\Psi$ with $U$ given by [10], having the properties

$$m(x, \lambda) := \Psi(x, \lambda) \exp\left(\frac{-\mathrm{i}x\lambda\sigma_3}{2}\right) \to I \quad \text{as } x \to -\infty$$

and $m(x, \lambda)$ is bounded as $x \to \infty$. For each fixed $x$, the $2 \times 2$ matrix function $m(x, \lambda)$ solves the RH problem in $\lambda$, where $\Sigma = \mathbb{R}$ and the jump matrix is

$$v = v(\lambda; x) = \begin{pmatrix} 1 - |r(\lambda)|^2 & r(\lambda)\,\mathrm{e}^{\mathrm{i}\lambda x} \\ -\bar{r}(\lambda)\,\mathrm{e}^{-\mathrm{i}\lambda x} & 1 \end{pmatrix} \qquad [11]$$

Here $r(\lambda)$ is the reflection coefficient of $q_0(x)$.

The direct scattering map $\mathcal{R}$ is described by mapping $q \mapsto r$,

$$q \mapsto m(x, \lambda) = m(x, \lambda; q) \mapsto v(\lambda; x) \mapsto r = \mathcal{R}(q)$$

By virtue of the $t$-equations in [5], if $q(t) = q(x, t)$ solves the NLS equation, then $r(t) = \mathcal{R}(q(\,\cdot\,, t))$ evolves as $r(t) = r(t, \lambda) = \mathrm{e}^{-\mathrm{i}t\lambda^2} r_0(\lambda)$, where $r_0 = \mathcal{R}(q_0)$. Given $r$, the inverse-scattering map $\mathcal{R}^{-1}$ is obtained by solving the normalized RH problem (RHP) with the jump matrix [11] and evaluating its solution $m(x, \lambda)$ as $\lambda \to \infty$ [9]:

$$\begin{aligned} r \mapsto v \mapsto \mathrm{RHP} &\mapsto m(x, \lambda) \\ &= m(x, \lambda; r) \mapsto m_1(x) \mapsto q(x) \\ &= -\mathrm{i}(m_1(x))_{12} \end{aligned}$$

and thus

$$q(x, t) = \mathcal{R}^{-1}\left(\mathrm{e}^{\mathrm{i}x(\cdot) - \mathrm{i}t(\cdot)^2} r(\cdot)\right) \qquad [12]$$

The mathematical rigor to this scheme is provided by the general theory of analytic matrix factorization making use of the relation between the factorization problem and certain singular integral equations; this relation can be established with the help of the Cauchy operators

$$Ch(\lambda) = \int_\Sigma \frac{h(\mu)}{\mu - \lambda} \frac{\mathrm{d}\mu}{2\pi\mathrm{i}}, \quad \lambda \in \mathbb{C}\backslash\Sigma$$

and

$$C^{\pm}h(\lambda) = \lim_{\substack{\lambda' \to \lambda \\ \lambda' \in (\pm)-\text{side of } \Sigma}} (Ch)(\lambda')$$

For a very general class of contours, the Cauchy operators $C^{\pm} : L^p \to L^p, 1 < p < \infty$, are bounded, $C^+ - C^- = I$, and $C^+ + C^- = -H$, where

$$Hh(\lambda) := \lim_{\varepsilon \to 0} \int_{\substack{\Sigma \\ |\mu - \lambda| > \varepsilon}} \frac{h(\mu)}{\lambda - \mu} \frac{\mathrm{d}\mu}{\pi\mathrm{i}}$$

is the Hilbert transform.

The map $\mathcal{R}$ is often considered as a nonlinear Fourier-type map; this point of view is supported by the fact that $\mathcal{R}$ is a bijection between the corresponding Schwartz spaces of functions. Making use of the $L^p$ or Hölder theory of the Cauchy operators and the related factorization problems, it is possible to analyze the action of $\mathcal{R}$ and $\mathcal{R}^{-1}$ in various functional spaces. This also requires making more precise the definition of the RH problem: for fixed $1 < p < \infty$, given $\Sigma$ and $v$ such that $v, v^{-1} \in L^\infty(\Sigma \to GL(N, \mathbb{C}))$, we say that $m_\pm$ solves an RH $L^p$-problem if $m_\pm \in I + \partial C(L^p)$ and $m_+(\lambda) = m_-(\lambda)v(\lambda)$ for $\lambda \in \Sigma$. Here a pair of $L^p(\Sigma)$-functions $f_\pm \in \partial C(L^p)$ if there exists a unique function $h \in L^p(\Sigma)$ such that $f_\pm(\lambda) = (C^\pm h)(\lambda)$. Then $f(\lambda) = Ch(\lambda)$, $\lambda \in \mathbb{C}\backslash\Sigma$, is called the extension of $f_\pm$ off $\Sigma$.

Given a factorization of $v = (v^-)^{-1}v^+ = (I - w^-)^{-1}(I + w^+)$ on $\Sigma$ with $v^\pm, (v^\pm)^{-1} \in L^p$, the basic associated singular integral operator is defined by

$$C_w h := C^+(hw^-) + C^-(hw^+)$$

If the operator $I - C_w$ is invertible on $L^p(\Sigma)$, with $\mu \in I + L^p(\Sigma)$, solving $(I - C_w)m = I$, then $m(\lambda) = I + (C(\mu(w_+ + w^-)))(\lambda)$ is the unique solution of the

RH problem $(\Sigma, v)$. Although the operator $C_w$ need not be compact, in many cases it is Fredholm with zero index. Then the existence of $(I - C_w)^{-1}$ is equivalent to the solvability of the RH problem $(\Sigma, v)$, and the normalized RH problem ($m \to I$ as $\lambda \to \infty$) has a unique solution if and only if the corresponding homogeneous RH problem (with $m \to 0$ as $\lambda \to \infty$) has only the trivial solution (vanishing lemma).

The most complete theory for RH problem relative to simple contours is the theory when $v$ is in an inverse, closed, decomposing Banach algebra $\mathcal{A}$, that is, the algebra of continuous functions with the Hilbert transform bounded in it such that if $f \in \mathcal{A}$, then $f^{-1} \in \mathcal{A}$. For contours with self-intersections, the RH factorization theory is formulated in terms of a pair of decomposing algebras: choosing the orientation of the contour in such a way that it divides the $\lambda$-plane into two disjoint regions, $\Omega^+$ and $\Omega^-$, and each arc of $\Sigma$ forms part of the positively oriented boundary of $\Omega^+$, the functions in the $+(-)$ algebra are continuous up to the boundary in each connected component of $\Omega^+$ $(\Omega^-)$.

The choice of functional spaces in the RH problem should be based on the integrable system at hand. For example, an integrable flow connected to the scattering problem for $\Psi_x = U\Psi$, with $U$ defined by [10], has in general the form $e^{it\lambda^p\sigma_3}v(\lambda)e^{-it\lambda^p\sigma_3}$ (Ablowitz–Kaup–Newell–Segur (AKNS) hierarchy) in the scattering space (for the NLS equation, $p=2$), so that appropriate spaces are $L^2((1+x^2)\,dx) \cap H^{p-1}$ for $q(\cdot, t)$ and $L^2((1+|\lambda|^{2p-2})|d\lambda|) \cap H^1$ as the scattering space. Deift and Zhou showed that in this case the scattering map $\mathcal{R}$ and the inverse-scattering map $\mathcal{R}^{-1}$ indeed involve no "loss" of smoothness or decay.

A generalization of the inverse-scattering transform method to the initial boundary-value problems for integrable nonlinear equations (on the half-line or on a finite interval with respect to the space variable $x$) can be also developed on the basis of the RH problem formalism. It this case, the construction of the corresponding RH problem involves simultaneous spectral analysis of the both linear equations in the Lax pair [5]. The boundary values generate an additional set of spectral functions, which generally makes the construction of the associated RH problem more complicated than in the case of the corresponding initial-value problem (particularly, the contour is to be enhanced by adding the part coming from the spectral analysis of the $t$-equation); however, this RH problem again depends explicitly on $x$ and $t$, which makes it possible to develop relevant techniques (such as the nonlinear steepest-descent method for the asymptotic analysis) in the same spirit as in the case of initial-value problems.

An RH problem may be viewed as a special case in a more general setting of problems of reconstructing an analytic function from the known structure of its singularities. The departure from analyticity of a function $m$ of the complex variable $\lambda$ can be described in terms of the "d-bar" derivative, $\partial m/\partial\bar{\lambda}$. If $\partial m/\partial\bar{\lambda}$ can be linearly related to $m$ itself, then the use of the extension of Cauchy's formula

$$m(\lambda) = \frac{1}{2\pi i}\int_D d\mu \wedge d\bar{\mu}\frac{1}{\mu-\lambda}\frac{\partial m}{\partial\bar{\mu}} + \frac{1}{2\pi i}\int_{\partial D} d\mu\frac{m(\mu)}{\mu-\lambda}$$

leads to a linear integral equation for $m$. This is the case for some multidimensional $(2+1)$ nonlinear integrable equations. For example, for the Kadomtsev–Petviashvili-I equation (the two-dimensional generalization of the Korteweg–de Vries equation) $(q_t + 6qq_x + q_{xxx})_x = 3q_{yy}$, the appropriate eigenfunctions are still sectionally meromorphic, but their jumps across a contour are connected nonlocally to $m$ on the contour, which leads to nonlocal RH problem of the type

$$m_+(\lambda) = m_-(\lambda) + \int_\Sigma d\mu m_-(\mu)f(\mu,\lambda), \quad \lambda \in \Sigma$$

with given $f(\mu,\lambda)$ (analogue of scattering data). Contrarily, the eigenfunctions for the Kadomtsev–Petviashvili-II equation $(q_t + 6qq_x + q_{xxx})_x = -3q_{yy}$ are nowhere analytic, with $\partial m/\partial\bar{\lambda}$ related to $m$ by

$$\frac{\partial m}{\partial\bar{\lambda}}(\lambda) = F(\operatorname{Re}\lambda, \operatorname{Im}\lambda)m(-\bar{\lambda}), \quad \lambda \in \mathbb{C}$$

## Nonlinear Steepest-Descent Method

The nonlinear steepest-descent method is based on a direct asymptotic analysis of the relevant RH problem; it is general and algorithmic in the sense that it does not require *a priori* information (anzatz) about the form of the solution of the asymptotic problem. However, the noncommutativity of the matrix setting requires developing rather sophisticated technical ideas, which, in particular, enable an explicit solution of the associated local RH problems.

To fix ideas, let us again consider the NLS equation. The dependence of the jump matrix $v(\lambda; x, t)$ on $x$ and $t$ is oscillatory; it is the same as in the integral

$$q(x,t) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{i(x\lambda - t\lambda^2)}\hat{q}_0(\lambda)d\lambda \qquad [13]$$

which solves the initial-value problem for the linearized version of [7]:

$$iq_t + q_{xx} = 0, \quad q(x,0) = q_0(x) \qquad [14]$$

(here $\hat{q}_0(\lambda)$ is the Fourier transform of the initial data $q_0$). The main contribution to [13] as $|x|$ and $t$ tend to $\infty$ comes from the point of stationary phase of $e^{i(x\lambda - t\lambda^2)}$, that is, the point $\lambda = \lambda_0 = x/2t$, for which

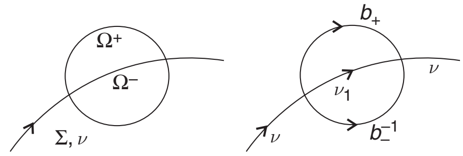
$$\frac{d}{d\lambda}(x\lambda - t\lambda^2) = 0$$

If $\hat{q}_0(\lambda)$ is analytic in a strip $|\mathrm{Im}\,\lambda| < \varepsilon$, then one can use Cauchy's theorem to deform [13] to an integral on a contour $\Sigma_\varepsilon$ such that $|e^{i(x\lambda - t\lambda^2)}|$ decreases rapidly on $\Sigma_\varepsilon$ away from $\lambda = \lambda_0$. Hence, as $t \to \infty$, the problem localizes to a neighborhood of $\lambda = \lambda_0$; this constitutes the standard method of steepest descent.

In the spirit of the oscillatory contour integral case, the nonlinear steepest-descent method for an oscillatory RH problem introduced by Deift and Zhou consists in the following: deform the contour and (rationally) approximate the jump matrix in order to obtain an RH problem with a jump matrix that decays to the identity away from stationary phase points; then, rescaling the problem near the stationary phase points, obtain a (local) RH problem with a piecewise constant jump matrix, which can be solved in closed form, usually in terms of certain special functions.

The contour deformation means the following. Suppose that the jump matrix of an RH problem $(\Sigma, v)$ has a factorization $v = b_-^{-1} v_1 b_+$ between two points on $\Sigma$, where $b_+(b_-)$ has holomorphic and nondegenerating continuation to the part $\Omega^+(\Omega^-)$ of a disk $\Omega$ supported by these points, see Figure 2a. Then the contour $\Sigma$ may be deformed to the contour $\Sigma' = \Sigma \cup \partial\Omega$, and the jump matrices across $\Sigma'$ may be defined as indicated in Figure 2b. If $m$ solves the RH problem $(\Sigma, v)$, then $m'$ defined by $m' = m b_\pm^{-1}$ in $\Omega^\pm$ and $m' = m$ outside $\Omega$ solves the deformed RH problem associated with $\Sigma'$.

The appropriate factorization of $v$ given by [8] and the contour deformation are to be chosen in accordance with signature table; for the NLS equation, it is given in Figure 3. The key step is to move algebraically the factors $e^{\pm i\theta}$ in $v(\lambda; x, t)$ into regions of the complex plane, where they are exponentially decreasing as $t \to \infty$. The jump matrix admits two algebraic factorizations:




(a) (b)

**Figure 2** Deformation of an RH problem.

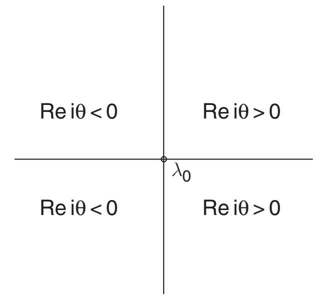



**Figure 3** Signature table.

$$\begin{aligned} v &= \begin{pmatrix} 1 - |r|^2 & r e^{i\theta} \\ -\bar{r} e^{-i\theta} & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & r e^{i\theta} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\bar{r} e^{-i\theta} & 1 \end{pmatrix} \quad (\lambda > \lambda_0) \\ &= \begin{pmatrix} 1 & 0 \\ -\dfrac{\bar{r} e^{-i\theta}}{1 - |r|^2} & 1 \end{pmatrix} \begin{pmatrix} 1 - |r|^2 & 0 \\ 0 & \dfrac{1}{1 - |r|^2} \end{pmatrix} \\ &\quad \times \begin{pmatrix} 1 & \dfrac{r e^{i\theta}}{1 - |r|^2} \\ 0 & 1 \end{pmatrix} \quad (\lambda < \lambda_0) \end{aligned}$$

The diagonal factors $(1 - |r|^2)^{\pm 1}$ can be removed by conjugating $v$ by $\delta_\pm^{\sigma_3}$, where $\delta(\lambda)$ solves the scalar, normalized RH problem on $\mathbb{R}: \delta_+ = \delta_-(1 - |r|^2)$ for $\lambda < \lambda_0$ and $\delta_+ = \delta_-$ for $\lambda > \lambda_0$; the solution of the latter can be written in a closed form:

$$\delta(\lambda) = \exp\left\{ \frac{1}{2\pi i} \int_{-\infty}^{\lambda_0} \frac{\log(1 - |r(\mu)|^2)}{\mu - \lambda} \, d\mu \right\}$$

Then $\tilde{m} := m \delta^{-\sigma_3}$ solves the RH problem across $\Sigma = \mathbb{R}$, with the jump matrix

$$\begin{aligned} \tilde{v} &= \begin{pmatrix} 1 & r\delta^2 e^{i\theta} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\bar{r}\delta^{-2} e^{-i\theta} & 1 \end{pmatrix} \quad (\lambda > \lambda_0) \\ &= \begin{pmatrix} 1 & 0 \\ -\dfrac{\bar{r}\delta_-^{-2} e^{-i\theta}}{1 - |r|^2} & 1 \end{pmatrix} \begin{pmatrix} 1 & \dfrac{r\delta_+^2 e^{i\theta}}{1 - |r|^2} \\ 0 & 1 \end{pmatrix} \quad (\lambda < \lambda_0) \end{aligned}$$

Replacing $r, \bar{r}$, etc., by appropriate rational approximations $[r], [\bar{r}]$, matching at $\lambda = \lambda_0$,

$$\tilde{m}_+ \begin{pmatrix} 1 & 0 \\ -[\bar{r}]\delta^{-2} e^{-i\theta} & 1 \end{pmatrix}$$

can be continued to the sector above $\mathbb{R}_+ + \lambda_0$ and

$$\tilde{m}_- \begin{pmatrix} 1 & [r]\delta^2 e^{i\theta} \\ 0 & 1 \end{pmatrix}$$

can be continued to the sector below $\mathbb{R}_+ + \lambda_0$, where the factors $e^{\pm i\theta}$ are exponentially decreasing. Doing the same for the appropriate factors on $\mathbb{R}_- + \lambda_0$, we obtain an RH problem on a cross, say, $(\lambda_0 + e^{i\pi/4}\mathbb{R}) \cup (\lambda_0 + e^{-i\pi/4}\mathbb{R})$. As $t \to \infty$, the RH problem then localizes at $\lambda_0$.

Performing an appropriate scaling, a straightforward computation shows that, as $t \to \infty$, the problem reduces to an RH problem with the jump matrix that does not depend on $\lambda$ (it is determined by $r(\lambda_0)$), which make it possible to solve this problem explicitly (in terms of the parabolic cylinder functions, in the case of the NLS equation). Using explicit asymptotics for these functions and controlling the error terms, it is possible to obtain the uniform (for all $x \in \mathbb{R}$) asymptotics for the solution of the initial-value problem for the NLS equation with $q_0 \in L^2((1+x^2)\,dx) \cap H^1$ of the form

$$q(x,t) = t^{-1/2}\alpha(\lambda_0)\exp(ix^2/(4t) - i\nu(\lambda_0)\log 2t) + O(t^{-(1/2+\kappa)})$$

for any fixed $0 < \kappa < 1/4$, where $\alpha$ and $\nu$ are given in terms of $r = \mathcal{R}(q_0)$:

$$\nu(\lambda) = -\frac{1}{2\pi}\log(1 - |r(\lambda)|^2)$$

$$|\alpha(\lambda)|^2 = \frac{\nu(\lambda)}{2}$$

and

$$\arg\alpha(\lambda) = \frac{1}{\pi}\int_{-\infty}^{\lambda}\log(\lambda-\mu)\,d(\log(1-|r(\mu)|^2)) + \frac{\pi}{4} + \arg\Gamma(i\nu(\lambda)) + \arg r(\lambda)$$

The method can be used to obtain asymptotic expansions to all orders. Also, for nonlinear equations supporting solitons, the soliton part of the asymptotics can be incorporated via the dressing method.

Further applications include long-time asymptotics for near-integrable systems, such as the perturbed NLS equation $iq_t + q_{xx} - 2|q|^2q - \varepsilon|q|^l q = 0$ for $l > 2$ and $\varepsilon > 0$, and the small-dispersion limits of integrable equations (e.g., for the Korteweg–de Vries equation $q_t - 6qq_x + \varepsilon^2 q_{xxx} = 0$ with small dispersion $\varepsilon \searrow 0$).

The RH formalism makes possible a comprehensive global asymptotic analysis of the Painlevé transcendents (which, due to their increasing role in the modern mathematical physics, should be considered as new nonlinear special functions), including explicit connection formulas, as $x$ approaches relevant critical points along different directions in the complex plane.

The development of the RH method in the theory of integrable systems caused emerging new analytic and algebraic ideas for other branches of mathematics and theoretical physics. The recent examples are the study of the asymptotics in the theory of orthogonal polynomials and random matrices and in combinatories (random permutations).

*See also:* Boundary-Value Problems for Integrable Equations; $\bar{\delta}$ Approach to Integrable Systems; Integrable Systems and Algebraic Geometry; Integrable Systems and the Inverse Scattering Method; Integrable Systems: Overview; Nonlinear Schrödinger Equations; Painlevé Equations; Twistor Theory: Some Applications [in Integrable Systems, Complex Geometry and String Theory]; Riemann–Hilbert Problem.

## Further Reading


Ablowitz MJ and Clarkson PA (1991) Solitons, Nonlinear Evolution Equations and Inverse Scatting, *London Math. Soc.*, Lecture Notes Series, vol. 149. Cambridge: Cambridge University Press.

Beals R, Deift PA, and Tomei C (1988) Direct and Inverse Scattering on the Line. *Mathematical Surveys and Monographs 28*. Providence, RI: American Mathematical Society.

Belokolos ED, Bobenko AI, Enol'skii VZ, and Its AR (1994) Algebro-Geometric Approach to Nonlinear Integrable Equations. *Springer Series in Nonlinear Dynamics*. Berlin: Springer.

Deift PA (1999) Orthogonal Polynomials and Random Matrices: A Riemann–Hilbert Approach. *Courant Lecture Notes in Mathematics*, vol. 3. New York: CIMS.

Deift PA and Zhou X (2003) Long-time asymptotics for solutions of the NLS equation with initial data in a weighted Sobolev space. *Communications on Pure and Applied Mathematics* 56(8): 1029–1077.

Deift PA, Its AR, and Zhou X (1993) Long-time asymptotics for integrable nonlinear wave equations. In: Fokas AS and Zakharov VE (eds.) *Important Developments in Soliton Theory*, pp. 181–204. Berlin: Springer.

Faddeev LD and Takhtajan LA (1987) *Hamiltonian Methods in the Theory of Solitons*. Berlin: Springer.

Fokas AS (2000) On the integrability of linear and nonlinear partial differential equations. *Journal of Mathematical Physics* 41: 4188–4237.

Its AR (2003) The Riemann–Hilbert problem and integrable systems. *Notices of the AMS* 50(11): 1389–1400.

Novikov SP, Manakov SV, Pitaevskii LP, and Zakharov VE (1984) *Theory of Solitons. The Inverse Scattering Method*. New York: Consultants Bureau.

Zhou X (1989) The Riemann–Hilbert problem and inverse scattering. *Journal on Mathematical Analysis. Society for Industrial and Applied Mathematics (SIAM)* 20: 966–986.

# Riemann–Hilbert Problem

**V P Kostov**, Université de Nice Sophia Antipolis, Nice, France



## Regular and Fuchsian Linear Systems on the Riemann Sphere

Consider a system of ordinary linear differential equations with time belonging to the Riemann sphere $CP^1 = \mathbf{C} \cup \infty$:

$$dX/dt = A(t)X \qquad [1]$$

The $n \times n$ matrix $A$ is meromorphic on $CP^1$, with poles at $a_1, \ldots, a_{p+1}$; the dependent variables $X$ form an $n \times n$ matrix. One can assume that $\infty$ is not among the poles $a_j$ and it is not a pole of the 1-form $A(t)dt$ (this can be achieved by a fractionally-linear transformation of $t$).

P Deligne has introduced a terminology of meromorphic connections and sections which is often preferred in modern literature to the one of meromorphic linear systems and their solutions, and there is a one-to-one correspondence between the two languages.

**Definition 1** System [1] is regular at the pole $a_j$ if its solutions have a moderate (or polynomial) growth rate there, that is, for every sector $S$ centered at $a_j$ and not containing other poles of the system and for every solution $X$ restricted to $S$ there exists $N_j \in \mathbf{R}$ such that $\|X(t - a_j)\| = O(|t - a_j|^{N_j})$ for all $t \in S$. System [1] is regular if it is regular at all poles $a_j$. System [1] is Fuchsian if its poles are logarithmic (i.e., of first order). Every Fuchsian system is regular.

**Remark 2** The opening of the sector $S$ might be $> 2\pi$. Restricting to a sector is necessary because the solutions are, in general, ramified at the poles $a_j$ and by turning around the poles much faster than approaching them one can obtain any growth rate.

A Fuchsian system can be presented in the form

$$dX/dt = \left(\sum_{j=1}^{p+1} A_j/(t - a_j)\right)X, \quad A_j \in \mathrm{gl}(n, \mathbf{C}) \qquad [2]$$

The sum of its *matrices-residua* $A_j$ is 0, that is,

$$A_1 + \cdots + A_{p+1} = 0 \qquad [3]$$

(recall that $\infty$ is not a pole of the system).

**Remark 3** The linear equation (with meromorphic coefficients) $\sum_{j=0}^{n} a_j(t)x^{(j)} = 0$ is Fuchsian if $a_j$ has poles of order only $\leq n - j$. A linear equation is Fuchsian if and only if it is regular. The best-studied Fuchsian equations are the hypergeometric one and its generalizations and the Jordan–Pochhammer equation.

The linear change of the dependent variables

$$X \mapsto W(t)X \qquad [4]$$

(where $W$ is meromorphic on $CP^1$) makes system [2] undergo the gauge transformation

$$A \to -W^{-1}(dW/dt) + W^{-1}AW \qquad [5]$$

(Most often one requires $W$ to be holomorphic and holomorphically invertible for $t \neq a_j, j = 1, \ldots, p+1$, so that no new singular points appear in the system.) This transformation preserves regularity but not necessarily being Fuchsian. The only invariant under the group of linear transformations [4] is the monodromy group of the system.

**Definition 4** Set $\Sigma = CP^1 \backslash \{a_1, \ldots, a_{p+1}\}$. Fix a base point $a_0 \in \Sigma$ and a matrix $B \in \mathrm{GL}(n, \mathbf{C})$. Consider a closed contour $\gamma$ with base point $a_0$ and bypassing the poles of the system. The monodromy operator of system [1] defined by this contour is the linear operator $M$ acting on the solution space of the system which maps the solution $X$ with $X|_{t=a_0} = B$ into the value of its analytic continuation along $\gamma$. Notation: $X \overset{\gamma}{\mapsto} XM$. The monodromy operator depends only on the class of homotopy equivalence of $\gamma$.

The monodromy group is the subgroup of $\mathrm{GL}(n, \mathbf{C})$ generated by all monodromy operators. It is defined only up to conjugacy due to the freedom to choose $a_0$ and $B$.

**Definition 5** Define the product (concatenation) $\gamma_1\gamma_2$ of two paths $\gamma_1, \gamma_2$ in $\Sigma$ (where the end of $\gamma_1$ coincides with the beginning of $\gamma_2$) as the path obtained by running $\gamma_1$ first and $\gamma_2$ next.

**Remark 6** The monodromy group is an antirepresentation of the fundamental group $\pi_1(\Sigma)$ into $\mathrm{GL}(n, \mathbf{C})$ because one has

$$X \overset{\gamma_1}{\mapsto} XM_1 \overset{\gamma_2}{\mapsto} XM_2M_1 \qquad [6]$$

that is, the concatenation $\gamma_1\gamma_2$ of the two contours defines the monodromy operator $M_2M_1$. In the text, the monodromy group is referred to as to a representation, not an antirepresentation.

One usually chooses a standard set of generators of $\pi_1(\Sigma)$ (see Figure 1) defined by contours $\gamma_j, j = 1, \ldots, p+1$, where $\gamma_j$ consists of a segment

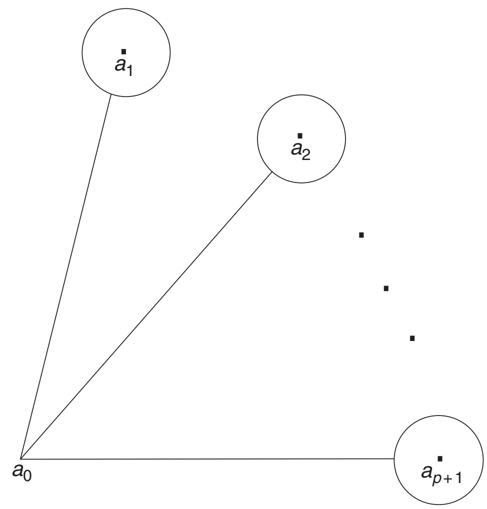



**Figure 1** The standard set of generators.

$[a_0, a'_j]$ ($a'_j$ being a point close to $a_j$), of a small circumference run counterclockwise (centered at $a_j$, passing through $a'_j$ and containing inside no pole of the system other than $a_j$), and of the segment $[a'_j, a_0]$. Thus, $\gamma_j$ is freely homotopic to a small loop circumventing counterclockwise $a_j$ (and no other pole $a_i$). The indices of the poles are chosen such that the indices of the contours increase from 1 to $p+1$ when one turns around $a_0$ clockwise.

For the standard choice of the contours the generators $M_j$ satisfy the relation

$$M_1 \ldots M_{p+1} = I \qquad [7]$$

Indeed, the concatenation of contours $\gamma_{p+1} \ldots \gamma_1$ is homotopy equivalent to 0 and equality [7] results from Remark 6.

**Remarks 7**

(i) If the matrix-residuum $A_j$ of a Fuchsian system has no eigenvalues differing by a nonzero integer, then the monodromy operator $M_j$ defined as above is conjugate to $\exp(2\pi i A_j)$. It is always true that the eigenvalues $\sigma_{k,j}$ of $M_j$ equal $\exp(2\pi i \lambda_{k,j})$, where $\lambda_{k,j}$ are the eigenvalues of $A_j$.

(ii) If the generators $M_j$ of the monodromy group are defined after a standard set of contours $\gamma_j$, then they are conjugate to the corresponding operators $L_j$ of local monodromy, that is, when the poles $a_j$ are circumvented counterclockwise along small loops. The operators $L_j$ of a regular system can be computed (up to conjugacy) algorithmically – one first makes the system Fuchsian at $a_j$ by means of a change [4] and then carries out the computation. Thus, $M_j = Q_j^{-1} L_j Q_j$ for some $Q_j \in GL(n, \mathbf{C})$ and the difficulty when computing the monodromy group of system [1] consists in computing the matrices $Q_j$ which is a transcendental problem.

(iii) As will be noted in Theorem 9, every component of every solution to a regular linear system is a function of the class of Nilsson, that is, representable as a convergent (on sectors) series $\sum_{k \in \mathbf{N},\, 1 \le i \le n,\, 0 \le \nu \le n-1} a_{i,k,\nu} t^{\alpha_i + k} \ln^\nu t$, $\alpha_i \in \mathbf{C}$, $a_{i,k,\nu} \in \mathbf{C}$.

**Example 8** The Fuchsian system $dX/dt = (A/t)X$, $A \in \mathrm{gl}(n, \mathbf{C})$, has two poles – at 0 and at $\infty$, with matrices-residua $A$ and $-A$. Any solution is of the form $X = \exp(A \ln t)G$, $G \in \mathrm{GL}(n, \mathbf{C})$. To compute the local monodromy around 0, change the argument of $t$ by $2\pi i$. This results in $\ln t \mapsto \ln t + 2\pi i$ and $X \mapsto XG^{-1}\exp(2\pi i A)G$, that is the monodromy operator at 0 equals $G^{-1}\exp(2\pi i A)G$ (and in the same way the one at $\infty$ equals $G^{-1}\exp(-2\pi i A)G$).

## Formulation and History of the Problem

The Riemann–Hilbert problem (or Hilbert's twenty-first problem) is formulated as follows:

*Prove that for any set of points $a_1, \ldots, a_{p+1} \in \mathbf{C}P^1$ and for any set of matrices $M_1, \ldots, M_p \in \mathrm{GL}(n, \mathbf{C})$ there exists a Fuchsian linear system with poles at and only at $a_1, \ldots, a_{p+1}$ for which the corresponding monodromy operators are $M_1, \ldots, M_p$, $M_{p+1} = (M_1 \ldots M_p)^{-1}$.*

Historically, the Riemann–Hilbert problem was first stated for Fuchsian equations, not for systems – Riemann mentions in a note at the end of the 1850s the problem how to reconstruct a Fuchsian equation from its monodromy representation and Hilbert includes it in 1900 as the twenty-first problem on his list in a formulation mentioning equations and not systems. However, the number of parameters necessary to parametrize a Fuchsian equation is, in general, smaller than the one necessary to parametrize a monodromy group generated by $p$ matrices. Therefore, one has to allow the presence of additional apparent singularities in the equation, that is, singularities the monodromy around which is trivial.

It had been believed for a long time that the Riemann–Hilbert problem has a positive solution for any $n \in \mathbf{N}$, after J. Plemelj in 1908 gave a proof with a gap. In his proof, Plemelj tries to reduce the Riemann–Hilbert problem to the so-called homogeneous Hilbert boundary-value problem of the theory of singular integral equations. It follows from the correct part of the proof that if one of

the monodromy operators of system [1] is diagonalizable, then system [1] is equivalent to a Fuchsian one; this is due to Yu S Il'yashenko. (In particular, if one allows just one additional apparent singularity, then the Riemann–Hilbert problem is positively solvable. The author has shown that the result still holds if one of the monodromy operators has one Jordan block of size 2 and $n-2$ Jordan blocks of size 1. The result is sharp – it would be false if one allows one Jordan block of size $\geq 3$ or two blocks of size 2.) It also follows that any finitely generated subgroup of $GL(n, \mathbf{C})$ is the monodromy group of a regular system with prescribed poles which is Fuchsian at all the poles with the possible exception of one (where the system is regular) which can be chosen among them at random.

After the publication of Plemelj's result, the interest shifted basically towards the question how to construct a Fuchsian system given the monodromy operators $M_j$. At the end of the 1920s IA Lappo-Danilevskii expressed the solutions to a Fuchsian system as series of the monodromy operators. These series are convergent for monodromy operators close to the identity matrix and for such operators one can express the residua $A_j$ of the Fuchsian system as convergent series of the monodromy operators.

In 1956 BL Krylov proved that the Riemann–Hilbert problem is solvable for $n=p=2$ by constructing a Fuchsian system after its monodromy group. In 1983 NP Erugin did the same in the case $n=2, p=3$, and established a connection between the Riemann–Hilbert problem and Painlevé's equations.

In 1957 H Röhrl reformulated the problem in terms of fibre bundles. His approach is more geometric; however, it does not require the system realizing a given monodromy group to be Fuchsian, but only regular.

In 1978 W Dekkers considered the particular case $n=2$ of the Riemann–Hilbert problem, and gave a positive answer to it. The gap in Plemelj's proof was detected in the 1980s by AT Kohn and YuS Il'yashenko.

It was proved by AA Bolibrukh in 1989 that, for $n \geq 3$, the problem has a negative answer. For $n=3$, the answer is negative precisely for those couples (monodromy group, set of poles) for which each monodromy operator $M_1, \ldots, M_{p+1}$ is conjugate to a Jordan block of size 3, the monodromy group is reducible, with an invariant subspace or factor-space of dimension 2, the monodromy sub- or factor-representation corresponding to it is irreducible and cannot be realized by a Fuchsian system having all its matrices-residua conjugate to Jordan blocks of size 2. In Bolibrukh's work, the last condition is formulated in a different (but equivalent) way using the notion of Fuchsian weight.

## The New Setting of the Problem

After the negative answer to the Riemann–Hilbert problem for $n \geq 3$, it is reasonable to reformulate it as follows:

*Find necessary and/or sufficient conditions for the choice of the monodromy operators $M_1, \ldots, M_p$ and the points $a_1, \ldots, a_{p+1}$ so that there should exist a Fuchsian system with poles at and only at the given points and whose monodromy operators $M_j$ should be the given ones.*

In the new setting of the Riemann–Hilbert problem, the answer is positive if the monodromy group is irreducible (for any positions of the poles $a_j$). This has been first proved by Bolibrukh for $n=3$ and then independently by the author and by him for any $n$.

Bolibrukh found many examples of couples (reducible monodromy group, poles) for which the answer to the Riemann–Hilbert problem is negative. For $n=3$, the negative answer is due to possible "bad position" of the poles and a small shift from this position while keeping the same monodromy group leads to a couple for which the answer is positive. For $n \geq 4$, there are couples where the negative answer is due to arithmetic properties of the eigenvalues of the matrices-residua and the corresponding monodromy groups are not realizable by Fuchsian systems for any position of the poles. During the last years of his life, Bolibrukh studied upper-triangular monodromy representations and found other examples with negative answer to the Riemann–Hilbert problem.

Bolibrukh also found some sufficient conditions for the positive resolvability of the Riemann–Hilbert problem in the case of a reducible monodromy group. For example, suppose that the monodromy group is a semidirect sum:

$$M_j = \begin{pmatrix} M_j^1 & * \\ 0 & M_j^2 \end{pmatrix}$$

where the matrices $M_j^i$ (of size $l_i \times l_i, i=1,2$) define the representations $\chi_i$. Suppose that the representation $\chi_2$ is realizable by a Fuchsian system, that the representation $\chi_1$ is irreducible, and that one of the matrices $M_j$ is block-diagonal, with left upper block of size $s \times s$, where $s \leq l_1$. Then for any choice of the poles $a_j$ the monodromy group can be realized by some Fuchsian system.

Bolibrukh also gave an estimation upon the number $m$ of additional apparent singularities in a Fuchsian equation which are sufficient to realize a given irreducible monodromy group. It follows from his result that

$$m \leq \frac{n(n-1)(p-1)}{2} + 1 - n$$

One can ask the question what the codimension of the subset in the space (monodromy group, poles) is which provides the negative answer to the Riemann–Hilbert problem in its initial setting. The (author's) answer for $p \geq 3$ is $2p(n-1)$, and for $n \geq 7$ this codimension is attained only at couples (monodromy group, poles) for which every monodromy operator $M_j$ is conjugate to a Jordan block of size $n$, the group has an invariant subspace or factor-space of dimension $n-1$, the corresponding sub- or factor-representation is irreducible and cannot be realized by a Fuchsian system in which all matrices-residua are conjugate to Jordan blocks of size $n-1$. For $n \leq 6$ there are examples where the same codimension is attained (but cannot be decreased) on other couples as well.

## Levelt's Result and Bolibrukh's Method

In 1961, AHM Levelt described the form of the solution to a regular system at its pole. His result is in the core of Bolibrukh's method for solving the Riemann–Hilbert problem.

**Theorem 9** *In the neighborhood of a pole, the solution to a regular linear system is representable in the form*

$$X = U_j(t-a_j)(t-a_j)^{D_j}(t-a_j)^{E_j}G_j \qquad [8]$$

*where the matrix* $U_j$ *is holomorphic in a neighborhood of* 0, $D_j = \mathrm{diag}(\varphi_{1,j}, \ldots, \varphi_{n,j}), \varphi_{n,j} \in \mathbf{Z}$, $\det G_j \neq 0$. *The matrix* $E_j$ *is in upper-triangular form and the real parts of its eigenvalues belong to* $[0,1)$ (*by definition*, $(t-a_j)^{E_j} = \mathrm{e}^{E_j \ln(t-a_j)}$). *The numbers* $\varphi_{k,j}$ *satisfy the condition* [10] *formulated below. They are valuations in the eigenspaces of the monodromy operator* $M_j$ (*i.e., in the maximal subspaces invariant for* $M_j$ *on which it acts as an operator with a single eigenvalue*).

*A regular system is Fuchsian at* $a_j$ *if and only if*

$$\det U_j(0) \neq 0 \qquad [9]$$

*The condition on* $\varphi_{k,j}$ *can be formulated as follows: let* $E_j$ *have one and the same eigenvalue in the rows with indices* $s_1 < s_2 < \cdots < s_q$. *Then one has*

$$\varphi_{s_1,j} \geq \varphi_{s_2,j} \geq \cdots \geq \varphi_{s_q,j} \qquad [10]$$

**Remark 10** Denote by $\beta_{k,j}$ the diagonal entries (i.e., the eigenvalues) of the matrix $E_j$. Then the sums $\beta_{k,j} + \varphi_{k,j}$ are the eigenvalues of the matrix-residuum $A_j$ at $a_j$.

In proving that the Riemann–Hilbert problem is positively solved in the case of an irreducible monodromy group, Bolibrukh (or the author) uses the correct part of Plemelj's proof – namely, that the given monodromy group can be realized by a regular system which is Fuchsian at all poles but one. After this, a suitable change [4] is sought which makes the system Fuchsian at the last pole. The criterium to be Fuchsian is provided by the above theorem; one checks how the matrices $D_j$, that is, the exponents $\varphi_{k,j}$ and the matrices $U_j$ change as a result of the transformation [4]. This is easier (one has only to multiply to the left by $W(t)$) than to see how the matrix $A(t)$ of system [1] changes because one has conjugation in rule [5]. This idea is also due to Bolibrukh.

When Bolibrukh obtains the negative answer to the Riemann–Hilbert problem in some case of reducible monodromy group, he often uses the following two propositions:

**Proposition 11** *The sum* $\sum \beta_{k,j} + \varphi_{k,j}$ *relative to a subspace of the solution space invariant for all monodromy operators is a non-positive integer.*

In particular, the sum of all exponents $\beta_{k,j} + \varphi_{k,j}$ is a non-positive integer which is 0 if and only if the system is Fuchsian.

**Proposition 12** *If some component of some column of some matrix solution to a regular system is identically equal to* 0, *then the monodromy group of the system is reducible.*

A reducible monodromy group can be conjugated to a block upper-triangular form, with the diagonal blocks defining irreducible representations. Thus, the Riemann–Hilbert problem for reducible monodromy groups makes necessary the answer to the question "given the set of poles $a_j$, for which sets of exponents $\varphi_{k,j}$ can a given irreducible monodromy group be realized by such a Fuchsian system?" For $n \geq 2$, an irreducible monodromy group can be *a priori* realized by infinitely many Fuchsian systems, with different sets of exponents $\varphi_{k,j}$. Consider the case when these exponents are fixed for $j \neq 1$; suppose that $a_1 = 0$. The author has shown that then infinitely many of the *a priori* possible choices of the exponents $\varphi_{k,1}$ cannot be realized by Fuchsian systems if and only if the given monodromy group is realized by a Fuchsian system which is obtained from another one via the change of time $t \mapsto t^k/(b_k t^k + b_{k-1}t^{k-1} + \cdots + b_0)$, $b_i \in \mathbf{C}, b_0 \neq 0, k \in \mathbf{N}^*, k > 1$. This change increases the number of poles.

## Further Developments – The Deligne–Simpson Problem

The Riemann–Hilbert problem can be generalized for irregular systems as follows. One asks whether for given poles $a_j$ there exists a linear system of ordinary differential equations on the Riemann sphere with these and only these poles which is Fuchsian at the regular singular points, which has prescribed formal normal forms, formal monodromies and Stokes multipliers at the irregular singular points, and which has a prescribed global monodromy.

The Riemann–Hilbert problem has been considered in some papers (of H Esnault, E Vieweg, and C Hertling) in the context of algebraic curves of higher genus instead of $CP^1$.

The study of the so-called Riemann–Hilbert correspondence between the category of holonomic $\mathcal{D}$-modules and the one of perverse sheaves with constructible cohomology has been initiated in the works of J Bernstein in the algebraic aspect and of M Sato, T Kawai, and M Kashiwara in the analytic one. This has been done in the case of a variety of arbitrary dimension (not necessarily $CP^1$), with codimension one pole divisor. Perversity has been defined by P Deligne, M Goresky, and R MacPherson. Regularity has been defined by M Kashiwara in the analytic aspect and by Z Mebkhout in the geometric one. Important contributions in the domain are due to Ph Maisonobe, M Merle, N Nitsure, C Sabbah, and the list is far from being exhaustive. The Riemann–Hilbert correspondence plays an important role in other trends of mathematics as well.

The Deligne–Simpson problem is formulated like this: *Give necessary and sufficient conditions upon the choice of the conjugacy classes* $c_j \subset \mathrm{gl}(n, \mathbf{C})$ *or* $C_j \subset \mathrm{GL}(n, \mathbf{C})$ *so that there should exist an irreducible* (*i.e., without proper invariant subspace*) $(p+1)$*-tuple of matrices* $A_j \in c_j$ *satisfying* [3] *or of matrices* $M_j$ *satisfying* [7].

The problem was stated in the 1980s by P Deligne for matrices $M_j$ and in the 1990s by the author for matrices $A_j$. C Simpson was the first to obtain results towards its resolution in the case of matrices $M_j$. The problem admits the following geometric interpretation in the case of matrices $M_j$: *For which* $(p+1)$*-tuples of local monodromies does there exist an irreducible global monodromy with such local monodromies?*

For generic eigenvalues the problem has found a complete solution in the author's papers in the form of a criterium upon the Jordan normal forms defined by the conjugacy classes. The author has treated the case of nilpotent matrices $A_j$ and the one of unipotent matrices $M_j$ as well. For matrices $A_j$, the problem has been completely solved (for any eigenvalues) by W Crawley-Boevey. The case of matrices $A_j$ with $p = 2$ has been treated by O Gleizer using results of A Klyachko. The case when the matrices $M_j$ are unitary is considered in papers of S Agnihotri, P Belkale, I Biswas, C Teleman, and C Woodward. Several cases of finite groups have been considered by M Dettweiler, S Reiter, K Strambach, J Thompson, and H Völklein. The important rigid case has been studied by NM Katz. Y Haraoka has considered the problem in the context of linear systems in Okubo's normal form. One can find details in an author's survey on the Deligne–Simpson problem (Kostov, 2004).

*See also:* Affine Quantum Groups; Bicrossproduct Hopf Algebras and Non-Commutative Spacetime; Einstein Equations: Exact Solutions; Holonomic Quantum Fields; Integrable Systems: Overview; Isomonodromic Deformations; Leray–Schauder Theory and Mapping Degree; Painlevé Equations; Riemann–Hilbert Methods in Integrable Systems; Twistors; WDVV Equations and Frobenius Manifolds.

## Further Reading

Anosov DV and Bolibruch AA (1994) *The Riemann–Hilbert Problem*, A Publication from the Moscow Institute of Mathematics, Aspects of Mathematics, Vieweg.

Arnol'd VI and Ilyashenko YuS (1988) Ordinary differential equations. In: *Dynamical Systems I*, Encyclopedia of Mathematical Sciences, t. 1. Berlin: Springer.

Beukers F and Heckman G (1989) Monodromy for the hypergeometric function ${}_nF_{n-1}$. *Inventiones Mathematicae* 95: 325–354.

Bolibrukh AA (1990) The Riemann–Hilbert problem. *Russian Mathematical Surveys* 45(2): 1–49.

Bolibrukh AA (1992) Sufficient conditions for the positive solvability of the Riemann–Hilbert problem. *Mathematical Notes* 51(1–2): 110–117.

Crawley-Boevey W (2003) On matrices in prescribed conjugacy classes with no common invariant subspace and sum zero. *Duke Mathematical Journal* 118(2): 339–352.

Dekkers W (1979) The matrix of a connection having regular singularities on a vector bundle of rank 2 on $\boldsymbol{P}^1(C)$. *Lecture Notes in Mathematics* 712: 33–43.

Deligne P (1970) *Equations différentielles à points singuliers réguliers*, Lecture Notes in Mathematics, vol. 163, pp. 133. Berlin: Springer.

Dettweiler M and Reiter S (1999) On rigid tuples in linear groups of odd dimension. *Journal of Algebra* 222(2): 550–560.

Esnault H and Viehweg E (1999) Semistable bundles on curves and irreducible representations of the fundamental group. *Algebraic geometry: Hirzebruch 70 (Warsaw, 1998), Contemporary Mathematics AMS, Providence, RI* 241: 129–138.

Esnault H and Hertling C (2001) Semistable bundles on curves and reducible representations of the fundamental group. *International Journal of Mathematics* 12(7): 847–855.

Haraoka Y (1994) Finite monodromy of Pochhammer equation. *Annales de l'Institut Fourier* 44(3): 767–810.

Katz NM (1995) Rigid Local systems. *Annnals of Mathematics, Studies Series, Study*, vol. 139. Princeton: Princeton University Press.

Kohn A and Treibich (1983) *Un résultat de Plemelj, Mathematics and Physics (Paris 1979/1982)*. In: Progr. Math., vol. 37, pp. 307–312. Boston: Birkhäuser.

Kostov VP (1992) Fuchsian linear systems on $\mathbf{C}P^1$ and the Riemann–Hilbert problem. *Comptes Rendus de l'Académie des Sciences à Paris*, 143–148.

Kostov VP (1999) The Deligne–Simpson problem. *C.R. Acad. Sci. Paris, t. 329 Série I*, 657–662.

Kostov VP (2004) The Deligne–Simpson problem – a survey. *Journal of Algebra* 281: 83–108.

Levelt AHM (1961) Hypergeometric functions. *Indagationes Mathematicae* 23: 361–401.

Maisonobe Ph and Narváez-Macarro L (eds.) (2004) *Eléments de la théorie des systèmes différentiels géométriques*. Cours du C.I.M.P.A. Ecole d'été de Séville. Séminaires et Congrès 8, xx + 430 pages.

Völklein H (1998) Rigid generators of classical groups. *Mathematische Annalen* 311(3): 421–438.

Wasow WR (1976) *Asymptotic Expansions for Ordinary Differential Equations*. New York: Huntington.

# Riemannian Holonomy Groups and Exceptional Holonomy

**D D Joyce**, University of Oxford, Oxford, UK



## Riemannian Holonomy Groups

Let $(M, g)$ be a Riemannian $n$-manifold. The holonomy group $\mathrm{Hol}(g)$ is a Lie subgroup of $\mathrm{O}(n)$, a global invariant of $g$ which measures the constant tensors $S$ on $M$ preserved by the Levi-Civita connection $\nabla$ of $g$. The most well-known examples of metrics with special holonomy are Kähler metrics, with $\mathrm{Hol}(g) \subseteq U(m) \subset \mathrm{O}(2m)$. A Kähler manifold $(M, g)$ also carries a complex structure $J$ and Kähler 2-form $\omega$ with $\nabla J = \nabla\omega = 0$.

The classification of Riemannian holonomy groups gives a list of interesting special Riemannian geometries such as Calabi–Yau manifolds and the exceptional holonomy groups $G_2$ and Spin(7), all of which are important in physics. These geometries have many features in common with Kähler geometry, and are characterized by the existence of constant exterior forms.

### General Properties of Holonomy Groups

Let $M$ be a connected manifold of dimension $n$ and $g$ a Riemannian metric on $M$, with Levi-Civita connection $\nabla$, regarded as a connection on the tangent bundle $TM$ of $M$. Suppose $\gamma : [0,1] \to M$ is a smooth path, with $\gamma(0) = x$ and $\gamma(1) = y$. Let $s$ be a smooth section of $\gamma^*(TM)$, so that $s : [0,1] \to TM$ with $s(t) \in T_{\gamma(t)}M$ for each $t \in [0,1]$. Then we say that $s$ is parallel if $\nabla_{\dot\gamma(t)} s(t) = 0$ for all $t \in [0,1]$, where $\dot\gamma(t)$ is

$$\frac{\mathrm{d}}{\mathrm{d}t}\gamma(t) \in T_{\gamma(t)}M$$

For each $v \in T_xM$, there is a unique parallel section $s$ of $\gamma^*(TM)$ with $s(0) = v$. Define a map $P_\gamma : T_xM \to T_yM$ by $P_\gamma(v) = s(1)$. Then $P_\gamma$ is well defined and linear, and is called the parallel transport map along $\gamma$. This easily generalizes to continuous, piecewise-smooth paths $\gamma$. As $\nabla g = 0$, we see that $P_\gamma : T_xM \to T_yM$ is orthogonal with respect to the metric $g$ on $T_xM$ and $T_yM$.

**Definition 1** Fix a point $x \in M$. $\gamma$ is said to be loop based at $x$ if $\gamma : [0,1] \to M$ is a continuous, piecewise-smooth path with $\gamma(0) = \gamma(1) = x$. If $\gamma$ is a loop based at $x$, then the parallel transport map $P_\gamma$ lies in $\mathrm{O}(T_xM)$, the group of orthogonal linear transformations of $T_xM$. Define the (Riemannian) holonomy group $\mathrm{Hol}_x(g)$ of $g$ based at $x$ to be

$$\mathrm{Hol}_x(g) = \{P_\gamma : \gamma \text{ is a loop based at } x\} \subseteq \mathrm{O}(T_xM) \qquad [1]$$

Here are some elementary properties of $\mathrm{Hol}_x(g)$. The only difficult part is showing that $\mathrm{Hol}_x(g)$ is a (closed) Lie subgroup.

**Theorem 2** $\mathrm{Hol}_x(g)$ *is a Lie subgroup of* $\mathrm{O}(T_xM)$, *which is closed and connected if* $M$ *is simply connected, but need not be closed or connected otherwise. Let* $x, y \in M$, *and suppose* $\gamma : [0,1] \to M$ *is a continuous, piecewise-smooth path with* $\gamma(0) = x$ *and* $\gamma(1) = y$, *so that* $P_\gamma : T_xM \to T_yM$. *Then*

$$P_\gamma \, \mathrm{Hol}_x(g) P_\gamma^{-1} = \mathrm{Hol}_y(g) \qquad [2]$$

By choosing an orthonormal basis for $T_xM$ we can identify $\mathrm{O}(T_xM)$ with the Lie group $\mathrm{O}(n)$, and so identify $\mathrm{Hol}_x(g)$ with a Lie subgroup of $\mathrm{O}(n)$. Changing the basis changes the subgroups by conjugation by an element of $\mathrm{O}(n)$. Thus, $\mathrm{Hol}_x(g)$ may be regarded as a Lie subgroup of $\mathrm{O}(n)$ defined up to conjugation. Equation [2] shows that in this sense, $\mathrm{Hol}_x(g)$ is independent of the base point $x$. Therefore, we omit the subscript $x$ and write $\mathrm{Hol}(g)$ for the holonomy group of $g$, regarded as a subgroup of $\mathrm{O}(n)$ defined up to conjugation.

It is significant that $\mathrm{Hol}(g)$ is a global invariant of $g$, that is, it does not vary from point to point like local invariants of $g$ such as the curvature. Generic metrics $g$ on $M$ have $\mathrm{Hol}(g) = \mathrm{SO}(n)$ if $M$ is orientable, and $\mathrm{Hol}(g) = \mathrm{O}(n)$ otherwise. But some special metrics $g$ can have $\mathrm{Hol}(g)$ a proper

subgroup of SO($n$) or O($n$). Then $M$ carries some extra geometric structures compatible with $g$.

Broadly, the smaller Hol($g$) is as a subgroup of O($n$), the more special $g$ is, and the more extra geometric structures there are. Therefore, understanding and classifying the possible holonomy groups gives a family of interesting special Riemannian geometries, such as Kähler geometry. All of these special geometries have cropped up in physics.

Define the holonomy algebra $\mathfrak{hol}(g)$ to be the Lie algebra of Hol($g$), regarded as a Lie subalgebra of $\mathfrak{o}(n)$, defined up to the adjoint action of O($n$). Define $\mathfrak{hol}_x(g)$ to be the Lie algebra of $\mathrm{Hol}_x(g)$, as a Lie subalgebra of $\mathfrak{o}(T_xM) \cong \Lambda^2 T_x^* M$. The holonomy algebra $\mathfrak{hol}(g)$ is intimately connected with the Riemann curvature tensor $R_{abcd} = g_{ae}R^e{}_{bcd}$ of $g$.

**Theorem 3** *The Riemann curvature tensor $R_{abcd}$ lies in $S^2\mathfrak{hol}_x(g)$ at $x$, where $\mathfrak{hol}_x(g)$ is regarded as a subspace of $\Lambda^2 T_x^* M$. It also satisfies the first and second Bianchi identities*

$$R_{abcd} + R_{adbc} + R_{acdb} = 0 \qquad [3]$$

$$\nabla_e R_{abcd} + \nabla_c R_{abde} + \nabla_d R_{abec} = 0 \qquad [4]$$

A related result is the Ambrose–Singer holonomy theorem, which, roughly speaking, says that $\mathfrak{hol}_x(g)$ may be reconstructed from $R_{abcd}|_y$ for all $y \in M$, moved to $x$ by parallel transport.

If $(M, g)$ and $(N, h)$ are Riemannian manifolds, the product $M \times N$ carries a product metric $g \times h$. It is easy to show that $\mathrm{Hol}(g \times h) = \mathrm{Hol}(g) \times \mathrm{Hol}(h)$. A Riemannian manifold $(M, g)$ is called reducible if every point has an open neighborhood isometric to a Riemannian product and irreducible otherwise.

**Theorem 4** *Let $(M, g)$ be Riemannian $n$-manifold. Then the natural representation of* Hol($g$) *on $\mathbb{R}^n$ is reducible if and only if $g$ is reducible.*

There is a class of Riemannian manifolds called the "Riemannian symmetric spaces" which are important in the theory of Riemannian holonomy groups. A Riemannian symmetric space is a special kind of Riemannian manifold with a transitive isometry group. The theory of symmetric spaces was worked out by Élie Cartan in the 1920s, who classified them completely, using his own classification of Lie groups and their representations.

A Riemannian metric $g$ is called "locally symmetric" if $\nabla_e R_{abcd} \equiv 0$, and "nonsymmetric" otherwise. Every locally symmetric metric is locally isometric to a Riemannian symmetric space. The relevance of symmetric spaces to holonomy groups is that many possible holonomy groups are the holonomy group of a Riemannian symmetric space, but are not realized by any nonsymmetric metric. Therefore, by restricting attention to nonsymmetric metrics, one considerably reduces the number of possible Riemannian holonomy groups.

A tensor $S$ on $M$ is constant if $\nabla S = 0$. An important property of Hol($g$) is that it determines the constant tensors on $M$.

**Theorem 5** *Let $(M, g)$ be a Riemannian manifold, with Levi-Civita connection $\nabla$. Fix $x \in M$, so that* $\mathrm{Hol}_x(g)$ *acts on $T_xM$, and so on the tensor powers $\bigotimes^k T_xM \otimes \bigotimes^l T_x^*M$. Suppose $S \in C^\infty(\bigotimes^k TM \otimes \bigotimes^l T^*M)$ is a constant tensor. Then $S|_x$ is fixed by the action of* $\mathrm{Hol}_x(g)$. *Conversely, if $S|_x \in \bigotimes^k T_xM \otimes \bigotimes^l T_x^*M$ is fixed by* $\mathrm{Hol}_x(g)$, *it extends to a unique constant tensor $S \in C^\infty(\bigotimes^k TM \otimes \bigotimes^l T^*M)$.*

The main idea in the proof is that if $S$ is a constant tensor and $\gamma : [0,1] \to M$ is a path from $x$ to $y$, then $P_\gamma(S|_x) = S|_y$, that is, "constant tensors are invariant under parallel transport." In particular, they are invariant under parallel transport around closed loops based at $x$, and so under elements of $\mathrm{Hol}_x(g)$.

### Berger's Classification of Holonomy Groups

Berger classified Riemannian holonomy groups in 1955.

**Theorem 6** *Let $M$ be a simply connected, $n$-dimensional manifold, and $g$ an irreducible, nonsymmetric Riemannian metric on $M$. Then*

(i) $\mathrm{Hol}(g) = \mathrm{SO}(n)$,
(ii) $n = 2m$ *and* $\mathrm{Hol}(g) = \mathrm{SU}(m)$ *or* $\mathrm{U}(m)$,
(iii) $n = 4m$ *and* $\mathrm{Hol}(g) = \mathrm{Sp}(m)$ *or* $\mathrm{Sp}(m)\mathrm{Sp}(1)$,
(iv) $n = 7$ *and* $\mathrm{Hol}(g) = G_2$, *or*
(v) $n = 8$ *and* $\mathrm{Hol}(g) = \mathrm{Spin}(7)$.

To simplify the classification, Berger makes three assumptions: $M$ is simply connected, $g$ is irreducible, and $g$ is nonsymmetric. We can make $M$ simply connected by passing to the "universal cover." The holonomy group of a reducible metric is a product of holonomy groups of irreducible metrics, and the holonomy groups of locally symmetric metrics follow from Cartan's classification of Riemannian symmetric spaces. Thus, these three assumptions can easily be removed.

Here is a sketch of Berger's proof of Theorem 6. As $M$ is simply connected, Theorem 2 shows Hol($g$) is a closed, connected Lie subgroup of SO($n$), and since $g$ is irreducible, Theorem 4 shows the representation of Hol($g$) on $\mathbb{R}^n$ is irreducible. So, suppose that $H$ is a closed, connected subgroup of

SO($n$) acting irreducibly on $\mathbb{R}^n$, with Lie algebra $\mathfrak{h}$. The classification of all such $H$ follows from the classification of Lie groups (and is of considerable complexity). Berger's method was to take the list of all such groups $H$, and to apply two tests to each possibility to find out if it could be a holonomy group. The only groups $H$ which passed both tests are those in the theorem.

Berger's tests are algebraic and involve the curvature tensor. Suppose that $R_{abcd}$ is the Riemann curvature of a metric $g$ with $\mathrm{Hol}(g) = H$. Then Theorem 3 gives $R_{abcd} \in S^2\mathfrak{h}$, and the first Bianchi identity [3] applies. But if $\mathfrak{h}$ has large codimension in $\mathfrak{o}(n)$, then the vector space $\mathfrak{R}^H$ of elements of $S^2\mathfrak{h}$ satisfying [3] will be small, or even zero. However, the "Ambrose–Singer holonomy theorem" shows that $\mathfrak{R}^H$ must be big enough to generate $\mathfrak{h}$. For many of the candidate groups $H$, this does not hold, and so $H$ cannot be a holonomy group. This is the first test.

Now $\nabla_e R_{abcd}$ lies in $(\mathbb{R}^n)^* \otimes \mathfrak{R}^H$, and also satisfies the second Bianchi identity, eqn [4]. Frequently, these imply that $\nabla R = 0$, so that $g$ is locally symmetric. Therefore, we may exclude such $H$, and this is Berger's second test.

Berger's proof does not show that the groups on his list actually occur as Riemannian holonomy groups – only that no others do. It is now known, though this took another thirty years to find out, that all possibilities in Theorem 6 do occur.

### The Groups on Berger's List

Here are some brief remarks about each group on Berger's list.

(i) SO($n$) is the holonomy group of generic Riemannian metrics.

(ii) Riemannian metrics $g$ with $\mathrm{Hol}(g) \subseteq \mathrm{U}(m)$ are called "Kähler metrics." Kähler metrics are a natural class of metrics on complex manifolds, and generic Kähler metrics on a given complex manifold have holonomy U($m$).

Metrics $g$ with $\mathrm{Hol}(g) = \mathrm{SU}(m)$ are called Calabi–Yau metrics. Since SU($m$) is a subgroup of U($m$), all Calabi–Yau metrics are Kähler. If $g$ is Kähler and $M$ is simply connected, then $\mathrm{Hol}(g) \subseteq \mathrm{SU}(m)$ if and only if $g$ is Ricci-flat. Thus, Calabi–Yau metrics are locally more or less the same as Ricci-flat Kähler metrics.

If $(M, J)$ is a compact complex manifold with trivial canonical bundle admitting Kähler metrics, then Yau's solution of the Calabi conjecture gives a unique Ricci-flat Kähler metric in each canonical class. This gives a way to construct many examples of Calabi–Yau manifolds, and explains why these have been named after them.

(iii) Metrics $g$ with $\mathrm{Hol}(g) = \mathrm{Sp}(m)$ are called "hyper-Kähler." As $\mathrm{Sp}(m) \subseteq \mathrm{SU}(2m) \subset \mathrm{U}(2m)$, hyper-Kähler metrics are Ricci-flat and Kähler.

Metrics $g$ with holonomy group Sp($m$)Sp(1) for $m \geq 2$ are called "quaternionic Kähler." (Note that quaternionic Kähler metrics are not in fact Kähler.) They are Einstein, but not Ricci-flat.

(iv), (v) $G_2$ and Spin(7) are the exceptional cases, so they are called the "exceptional holonomy groups." Metrics with these holonomy groups are Ricci-flat.

The groups can be understood in terms of the four division algebras: the real numbers $\mathbb{R}$, the complex numbers $\mathbb{C}$, the quaternions $\mathbb{H}$, and the octonions or Cayley numbers $\mathbb{O}$.

- SO($n$) is a group of automorphisms of $\mathbb{R}^n$.
- U($m$) and SU($m$) are groups of automorphisms of $\mathbb{C}^m$.
- Sp($m$) and Sp($m$) Sp(1) are automorphism groups of $\mathbb{H}^m$.
- $G_2$ is the automorphism group of $\mathrm{Im}\,\mathbb{O} \cong \mathbb{R}^7$. Spin(7) is a group of automorphisms of $\mathbb{O} \cong \mathbb{R}^8$, preserving part of the structure on $\mathbb{O}$.

## The Exceptional Holonomy Groups

For some time after Berger's classification, the exceptional holonomy groups remained a mystery. In 1987, Bryant used the theory of exterior differential systems to show that locally there exist many metrics with these holonomy groups, and gave some explicit, incomplete examples. Then in 1989, Bryant and Salamon found explicit, complete metrics with holonomy $G_2$ and Spin(7) on noncompact manifolds. In 1994–95, the author constructed the first examples of metrics with holonomy $G_2$ and Spin(7) on compact manifolds. For more information on exceptional holonomy, see Joyce (2000, 2002).

### The Holonomy Group $G_2$

Let $(x_1, \ldots, x_7)$ be coordinates on $\mathbb{R}^7$. Write $\mathrm{d}\boldsymbol{x}_{ij\ldots l}$ for the exterior form $\mathrm{d}x_i \wedge \mathrm{d}x_j \wedge \cdots \wedge \mathrm{d}x_l$ on $\mathbb{R}^7$. Define a metric $g_0$, a 3-form $\varphi_0$, and a 4-form $*\varphi_0$ on $\mathbb{R}^7$ by

$$\begin{aligned} g_0 &= \mathrm{d}x_1^2 + \cdots + \mathrm{d}x_7^2 \\ \varphi_0 &= \mathrm{d}\boldsymbol{x}_{123} + \mathrm{d}\boldsymbol{x}_{145} + \mathrm{d}\boldsymbol{x}_{167} + \mathrm{d}\boldsymbol{x}_{246} \\ &\quad - \mathrm{d}\boldsymbol{x}_{257} - \mathrm{d}\boldsymbol{x}_{347} - \mathrm{d}\boldsymbol{x}_{356} \\ *\varphi_0 &= \mathrm{d}\boldsymbol{x}_{4567} + \mathrm{d}\boldsymbol{x}_{2367} + \mathrm{d}\boldsymbol{x}_{2345} + \mathrm{d}\boldsymbol{x}_{1357} \\ &\quad - \mathrm{d}\boldsymbol{x}_{1346} - \mathrm{d}\boldsymbol{x}_{1256} - \mathrm{d}\boldsymbol{x}_{1247} \end{aligned} \qquad [5]$$

The subgroup of $\mathrm{GL}(7, \mathbb{R})$ preserving $\varphi_0$ is the exceptional Lie group $G_2$. It also preserves $g_0, *\varphi_0$,

and the orientation on $\mathbb{R}^7$. It is a compact, semisimple, 14-dimensional Lie group, a subgroup of SO(7).

A $G_2$-structure on a 7-manifold $M$ is a principal sub-bundle of the frame bundle of $M$, with structure group $G_2$. Each $G_2$-structure gives rise to a 3-form $\varphi$ and a metric $g$ on $M$, such that every tangent space of $M$ admits an isomorphism with $\mathbb{R}^7$ identifying $\varphi$ and $g$ with $\varphi_0$ and $g_0$, respectively. By an abuse of notation, $(\varphi, g)$ can be referred to as a $G_2$-structure.

**Proposition 7** *Let $M$ be a 7-manifold and $(\varphi, g)$ a $G_2$-structure on $M$. Then the following are equivalent*:

(i) $\mathrm{Hol}(g) \subseteq G_2$, *and $\varphi$ is the induced 3-form*;
(ii) $\nabla\varphi = 0$ *on $M$, where $\nabla$ is the Levi-Civita connection of $g$; and*
(iii) $d\varphi = d(*\varphi) = 0$ *on $M$.*

The equations $d\varphi = d(*\varphi) = 0$ look like linear partial differential equations on $\varphi$. However, it is better to consider them as nonlinear, for the following reason. The 3-form $\varphi$ determines the metric $g$, and $g$ gives the Hodge star $*$ on $M$. So $*\varphi$ is a nonlinear function of $\varphi$, and $d(*\varphi) = 0$ a nonlinear equation. Thus, constructing and studying $G_2$-manifolds come down to studying solutions of nonlinear elliptic partial differential equations.

Note that $\mathrm{Hol}(g) \subseteq G_2$ if and only if $\nabla\varphi = 0$ follows from Theorem 5. We call $\nabla\varphi$ the "torsion" of the $G_2$-structure $(\varphi, g)$, and when $\nabla\varphi = 0$ the $G_2$-structure is "torsion-free." A triple $(M, \varphi, g)$ is called a $G_2$-manifold if $M$ is a 7-manifold and $(\varphi, g)$ a torsion-free $G_2$-structure on $M$. If $g$ has holonomy $\mathrm{Hol}(g) \subseteq G_2$, then $g$ is Ricci-flat.

**Theorem 8** *Let $M$ be a compact 7-manifold, and suppose that $(\varphi, g)$ is a torsion-free $G_2$-structure on $M$. Then $\mathrm{Hol}(g) = G_2$ if and only if $\pi_1(M)$ is finite. In this case, the moduli space of metrics with holonomy $G_2$ on $M$, up to diffeomorphisms isotopic to the identity, is a smooth manifold of dimension $b^3(M)$.*

### The Holonomy Group Spin(7)

Let $\mathbb{R}^8$ have coordinates $(x_1, \ldots, x_8)$. Define a 4-form $\Omega_0$ on $\mathbb{R}^8$ by

$$\begin{aligned}\Omega_0 = {}& dx_{1234} + dx_{1256} + dx_{1278} + dx_{1357} - dx_{1368} \\ & - dx_{1458} - dx_{1467} - dx_{2358} - dx_{2367} - dx_{2457} \\ & + dx_{2468} + dx_{3456} + dx_{3478} + dx_{5678}\end{aligned} \qquad [6]$$

The subgroup of $\mathrm{GL}(8, \mathbb{R})$ preserving $\Omega_0$ is the holonomy group Spin(7). It also preserves the orientation on $\mathbb{R}^8$ and the Euclidean metric $g_0 = dx_1^2 + \cdots + dx_8^2$. It is a compact, semisimple, 21-dimensional Lie group, a subgroup of SO(8).

A Spin(7)-structure on an 8-manifold $M$ gives rise to a 4-form $\Omega$ and a metric $g$ on $M$, such that each tangent space of $M$ admits an isomorphism with $\mathbb{R}^8$ identifying $\Omega$ and $g$ with $\Omega_0$ and $g_0$, respectively. By an abuse of notation, the pair $(\Omega, g)$ is referred to as a Spin(7)-structure.

**Proposition 9** *Let $M$ be an 8-manifold and $(\Omega, g)$ a* Spin(7)*-structure on $M$. Then the following are equivalent*:

(i) $\mathrm{Hol}(g) \subseteq \mathrm{Spin}(7)$ *and $\Omega$ is the induced 4-form*;
(ii) $\nabla\Omega = 0$ *on $M$, where $\nabla$ is the Levi-Civita connection of $g$; and*
(iii) $d\Omega = 0$ *on $M$.*

We call $\nabla\Omega$ the torsion of the Spin(7)-structure $(\Omega, g)$, and $(\Omega, g)$ torsion free if $\nabla\Omega = 0$. A triple $(M, \Omega, g)$ is called a Spin(7)-manifold if $M$ is an 8-manifold and $(\Omega, g)$ a torsion-free Spin(7)-structure on $M$. If $g$ has holonomy $\mathrm{Hol}(g) \subseteq \mathrm{Spin}(7)$, then $g$ is Ricci-flat.

Here is a result on compact 8-manifolds with holonomy Spin(7).

**Theorem 10** *Let $(M, \Omega, g)$ be a compact* Spin(7)*-manifold. Then, $\mathrm{Hol}(g) = \mathrm{Spin}(7)$ if and only if $M$ is simply connected, and $b^3(M) + b^4_+(M) = b^2(M) + 2b^4_-(M) + 25$. In this case, the moduli space of metrics with holonomy* Spin(7) *on $M$, up to diffeomorphisms isotopic to the identity, is a smooth manifold of dimension $1 + b^4_-(M)$.*

The inclusions between the holonomy groups $\mathrm{SU}(m)$, $G_2$, Spin(7) are

$$\begin{array}{ccccc} \mathrm{SU}(2) & \longrightarrow & \mathrm{SU}(3) & \longrightarrow & G_2 \\ \downarrow & & \downarrow & & \downarrow \\ \mathrm{SU}(2)\times\mathrm{SU}(2) & \longrightarrow & \mathrm{SU}(4) & \longrightarrow & \mathrm{Spin}(7) \end{array} \qquad [7]$$

The meaning of the above equation is illustrated by using the inclusion $\mathrm{SU}(3) \hookrightarrow G_2$. As SU(3) acts on $\mathbb{C}^3$, it also acts on $\mathbb{R} \oplus \mathbb{C}^3 \cong \mathbb{R}^7$, taking the SU(3)-action on $\mathbb{R}$ to be trivial. Thus, we embed SU(3) as a subgroup of $\mathrm{GL}(7, \mathbb{R})$. It turns out that SU(3) is contained in the subgroup $G_2$ of $\mathrm{GL}(7, \mathbb{R})$ defined in the section "The holonomy group $G_2$."

### Constructing Compact $G_2$- and Spin(7)-Manifolds

The author's method of constructing compact 7-manifolds with holonomy $G_2$ is based on the

Kummer construction for Calabi–Yau metrics on the $K3$ surface and may be divided into four steps.

*Step* 1. Let $T^7$ be the 7-torus and $(\varphi_0, g_0)$ a flat $G_2$-structure on $T^7$. Choose a finite group $\Gamma$ of isometries of $T^7$ preserving $(\varphi_0, g_0)$. Then the quotient $T^7/\Gamma$ is a singular, compact 7-manifold, an orbifold.

*Step* 2. For certain special groups $\Gamma$, there is a method to resolve the singularities of $T^7/\Gamma$ in a natural way, using complex geometry. We get a nonsingular, compact 7-manifold $M$, together with a map $\pi : M \to T^7/\Gamma$, the resolving map.

*Step* 3. On $M$, we explicitly write down a one-parameter family of $G_2$-structures $(\varphi_t, g_t)$ depending on $t \in (0, \epsilon)$. They are not torsion free, but have small torsion when $t$ is small. As $t \to 0$, the $G_2$-structure $(\varphi_t, g_t)$ converges to the singular $G_2$-structure $\pi^*(\varphi_0, g_0)$.

*Step* 4. We prove using analysis that for sufficiently small $t$, the $G_2$-structure $(\varphi_t, g_t)$ on $M$, with small torsion, can be deformed to a $G_2$-structure $(\varphi_t, \tilde{g}_t)$, with zero torsion. Finally, it is shown that $\tilde{g}_t$ is a metric with holonomy $G_2$ on the compact 7-manifold $M$.

We explain the first two steps in greater detail. For Step 1, an example of a suitable group $\Gamma$ is given here.

**Example 11** Let $(x_1, \ldots, x_7)$ be coordinates on $T^7 = \mathbb{R}^7/\mathbb{Z}^7$, where $x_i \in \mathbb{R}/\mathbb{Z}$. Let $(\varphi_0, g_0)$ be the flat $G_2$-structure on $T^7$ defined by [5]. Let $\alpha, \beta$, and $\gamma$ be the involutions of $T^7$ defined by

$$\alpha : (x_1, \ldots, x_7) \mapsto (x_1, x_2, x_3, -x_4, -x_5, -x_6, -x_7) \qquad [8]$$

$$\beta : (x_1, \ldots, x_7) \mapsto (x_1, -x_2, -x_3, x_4, x_5, \tfrac{1}{2} - x_6, -x_7) \qquad [9]$$

$$\gamma : (x_1, \ldots, x_7) \mapsto (-x_1, x_2, -x_3, x_4, \tfrac{1}{2} - x_5, x_6, \tfrac{1}{2} - x_7) \qquad [10]$$

By inspection, $\alpha$, $\beta$, and $\gamma$ preserve $(\varphi_0, g_0)$, because of the careful choice of exactly which signs to change. Also, $\alpha^2 = \beta^2 = \gamma^2 = 1$, and $\alpha$, $\beta$, and $\gamma$ commute. Thus, they generate a group $\Gamma = \langle \alpha, \beta, \gamma \rangle \cong \mathbb{Z}_2^3$ of isometries of $T^7$ preserving the flat $G_2$-structure $(\varphi_0, g_0)$.

Having chosen a lattice $\Lambda$ and finite group $\Gamma$, the quotient $T^7/\Gamma$ is an orbifold, a singular manifold with only quotient singularities. The singularities of $T^7/\Gamma$ come from the fixed points of nonidentity elements of $\Gamma$. We now describe the singularities in the example.

**Lemma 12** *In Example 11, $\beta\gamma$, $\gamma\alpha$, $\alpha\beta$, and $\alpha\beta\gamma$ have no fixed points on $T^7$. The fixed points of $\alpha, \beta, \gamma$ are each 16 copies of $T^3$. The singular set $S$ of $T^7/\Gamma$ is a disjoint union of 12 copies of $T^3$, 4 copies from each of $\alpha, \beta, \gamma$. Each component of $S$ is a singularity modeled on that of $T^3 \times \mathbb{C}^2/\{\pm 1\}$.*

The most important consideration in choosing $\Gamma$ is that we should be able to resolve the singularities of $T^7/\Gamma$ within holonomy $G_2$, in Step 2. We have no idea how to resolve general orbifold singularities of $G_2$-manifolds. However, after fifty years of hard work we understand well how to resolve orbifold singularities of Calabi–Yau manifolds, with holonomy SU($m$). This is done by a combination of algebraic geometry, which produces the underlying complex manifold by a crepant resolution, and Calabi–Yau analysis, which produces the Ricci-flat Kähler metric on this complex manifold.

Now the holonomy groups SU(2) and SU(3) are subgroups of $G_2$, as in [7]. Our tactic in Step 2 is to ensure that all of the singular set $S$ of $T^7/\Gamma$ can locally be resolved with holonomy SU(2) or SU(3), and then use Calabi–Yau geometry to do this. In particular, suppose each connected component of $S$ is isomorphic to either

1. $T^3 \times \mathbb{C}^2/G$, for $G$ a finite subgroup of SU(2); or
2. $\mathcal{S}^1 \times \mathbb{C}^3/G$, for $G$ a finite subgroup of SU(3) acting freely on $\mathbb{C}^3 \setminus \{0\}$.

One can use complex algebraic geometry to find a crepant resolution $X$ of $\mathbb{C}^2/G$ or $Y$ of $\mathbb{C}^3/G$. Then $T^3 \times X$ or $\mathcal{S}^1 \times Y$ gives a local model for how to resolve the corresponding component of $S$ in $T^7/\Gamma$. Thus we construct a nonsingular, compact 7-manifold $M$ by using the patches $T^3 \times X$ or $\mathcal{S}^1 \times Y$ to repair the singularities of $T^7/\Gamma$. In the case of Example 11, this means gluing 12 copies of $T^3 \times X$ into $T^7/\Gamma$, where $X$ is the blow-up of $\mathbb{C}^2/\{\pm 1\}$ at its singular point.

By considering different groups $\Gamma$ acting on $T^7$, and also by finding topologically distinct resolutions $M_1, \ldots, M_k$ of the same orbifold $T^7/\Gamma$, we can construct many compact Riemannian 7-manifolds with holonomy $G_2$. A good number of examples are given in Joyce (2000, chapter 12). **Figure 1** displays the 252 different sets of Betti numbers of compact, simply connected 7-manifolds with holonomy $G_2$ constructed there together with 5 more sets from Kovalev. It seems likely to the author that the Betti numbers given in **Figure 1** are only a small proportion of

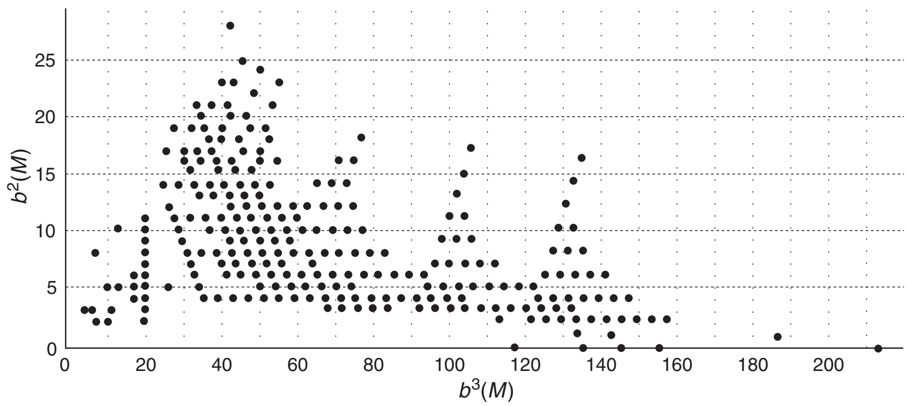



**Figure 1** Betti numbers $(b^2, b^3)$ of compact $G_2$-manifolds. (From Joyce (2000) and Kovalev (2003).)

the Betti numbers of all compact 7-manifolds with holonomy $G_2$.

A different construction of compact 7-manifolds with holonomy $G_2$ was given by Kovalev (2003), involving gluing together asymptotically cylindrical Calabi–Yau 3-folds. Compact 8-manifolds with holonomy Spin(7) were constructed by the author using two different methods: first, by resolving singularities of torus orbifolds $T^8/\Gamma$ in a similar way to the $G_2$ case (though the details are different and more difficult), and second, by resolving $Y/\langle\sigma\rangle$ for $Y$ a Calabi–Yau 4-orbifold with singularities of a special kind, and $\sigma$ an antiholomorphic isometric involution of $Y$. Details can be found in Joyce (2000).

*See also:* Calibrated Geometry and Special Lagrangian Submanifolds.

## Further Reading


Bryant RL (1987) Metrics with exceptional holonomy. *Annals of Mathematics* 126: 525–576.

Bryant RL and Salamon SM (1989) On the construction of some complete metrics with exceptional holonomy. *Duke Mathematical Journal* 58: 829–850.

Gross M, Huybrechts D, and Joyce D (2003) Calabi–Yau *Manifolds and Related Geometries*, Universitext Series. Berlin: Springer.

Joyce DD (2002) Constructing compact manifolds with exceptional holonomy, math.DG/0203158, 2002 (a survey paper).

Joyce DD (2004) Constructing compact manifolds with exceptional holonomy. In: Douglas M, Gauntlett J, and Gross M (eds.) *Strings and Geometry*, Clay Mathematics Proceedings 3. pp. 177–191. Providence, RI: American Mathematical Society. math.DG/0203158.

Kovalev AG (2003) Twisted connected sums and special Riemannian holonomy. *Journal für die Reine und Angewandte Mathematik* 565: 125–160, math.DG/0012189.

# S

## Saddle Point Problems

**M Schechter**, University of California at Irvine, Irvine, CA, USA



### Introduction

Many problems arising in science and engineering call for the solving of the Euler equations of functionals, that is, equations of the form

$$G'(u) = 0 \qquad [1]$$

where $G(u)$ is a $C^1$-functional (usually representing the energy) arising from the given data. As an illustration, the equation

$$-\Delta u(x) = f(x, u(x))$$

is the Euler equation of the functional

$$G(u) = \frac{1}{2}\|\nabla u\|^2 - \int F(x, u(x))\,\mathrm{d}x$$

on an appropriate space, where

$$F(x,t) = \int_0^t f(x,s)\,\mathrm{d}s \qquad [2]$$

and the norm is that of $L^2$. The solving of the Euler equations is tantamount to finding critical points of the corresponding functional. The classical approach was to look for maxima or minima. If one is looking for a minimum, it is not sufficient to know that the functional is bounded from below, as is easily checked. However, one can show that there is a sequence satisfying

$$G(u_k) \to a, \qquad G'(u_k) \to 0 \qquad [3]$$

for $a = \inf G$. If the sequence has a convergent subsequence, this will produce a minimum.

However, when extrema do not exist, there is no clear way of obtaining critical points. In particular, this happens when the functional is not bounded from either above or below. Until recently, there was no organized procedure for producing critical points which are not extrema. We shall describe an approach which is very useful in such cases.

To illustrate the technique, we consider the problem of finding a solution of

$$-u''(x) + u(x) = f(x, u(x)) \qquad [4]$$

$x \in I = [0, 2\pi]$, under the conditions

$$u(0) = u(2\pi), \qquad u'(0) = u'(2\pi) \qquad [5]$$

We assume that the function $f(x,t)$ is continuous in $I \times \mathbb{R}$ and is periodic in $x$ with period $2\pi$. The approach begins by asking the question, "does there exist a differentiable function $G$ from a space $H$ to $\mathbb{R}$ such that [4], [5] are equivalent to [1]?" It is hoped that one can mimic the methods of calculus to find critical points and thus solve [1].

Actually, we are asking the following: does there exist a mapping $G$ from a space $H$ to $\mathbb{R}$ such that $G$ has a critical point $u$ satisfying $G'(u) = -u'' + u - f(x, u(x))$?

In order to solve the problem one has to

1. find $G(u)$ such that
$$(G'(u), v)_H = (u, v)_H - (f(\cdot, u), v) \qquad [6]$$
holds for each $u, v \in H$,
2. show that there is a function $u(x)$ such that $G'(u) = 0$,
3. show that $u''$ exists in $I$,
4. show that [1] implies [4].

We used the notation

$$(u, v) = \int_0^{2\pi} u(x)v(x)\,\mathrm{d}x$$

In order to carry out the procedure, we assume that for each $R > 0$ there is a constant $C_R$ such that

$$|f(x,t)| \le C_R, \quad x \in I, \ t \in \mathbb{R}, \ |t| \le R \qquad [7]$$

This assumption is used to carry out step (1). We define

$$G(u) = \frac{1}{2}\|u\|_H^2 - \int_0^{2\pi} F(x, u(x))\,\mathrm{d}x \qquad [8]$$

where $F(x,t)$ is given by [2] and we take $H$ to be the completion of $C^1(I)$ with respect to the norm

$$\|u\|_H = (\|u'\|^2 + \|u\|^2)^{1/2} \qquad [9]$$

where $\|u\|^2 = (u, u)$. We have

**Theorem 1** *If $f(x,t)$ satisfies [7], then $G(u)$ given by [8] is continuously differentiable and satisfies [6].*

Once we have reduced the problem to solving [1], we can search for critical points. The easiest type to locate are "saddle points" which are local minima in some directions and local maxima in all others. For instance, we obtain theorems such as

**Theorem 2** *Assume that*

$$\begin{aligned} &|f(x,t)| \le C(|t|+1), \quad x \in I,\ t \in \mathbb{R} \\ &2F(x,t)/t^2 \to \beta(x) \quad \text{a.e. as } |t| \to \infty \end{aligned} \qquad [10]$$

*with $\beta(x)$ satisfying*

$$\begin{aligned} &1+n^2 \le \beta(x) \le 1+(n+1)^2 \\ &1+n^2 \not\equiv \beta(x) \not\equiv 1+(n+1)^2 \end{aligned} \qquad [11]$$

*and $n$ an integer $\ge 0$. If $G(u)$ is given by [8], then there is a $u_0 \in H$ such that*

$$G'(u_0) = 0 \qquad [12]$$

*In particular, $u_0$ is a solution of [4] and [5] in the usual sense.*

In proving this theorem, we shall make use of

**Theorem 3** *Let $M, N$ be closed subspaces of a Hilbert space $E$ such that $M = N^\perp$. Assume that at least one of these subspaces is finite dimensional. Let $G$ be a continuously differentiable functional on $E$ satisfying*

$$m_0 = \sup_{v\in N} \inf_{w \in M} G(v+w) \ne -\infty \qquad [13]$$

*and*

$$m_1 = \inf_{w\in M} \sup_{v \in N} G(v+w) \ne \infty \qquad [14]$$

*Then there is a sequence $\{u_k\} \subset E$ such that*

$$G(u_k) \to c, \quad m_0 \le c \le m_1, \qquad G'(u_k) \to 0 \qquad [15]$$

Theorem 3 allows us to obtain solutions if we can find subspaces of $H$ such that [13] and [14] hold. We use it to give the proof of Theorem 2.

**Proof.** Note that

$$\|u\|_H^2 = \sum (1+k^2)|\alpha_k|^2, \quad u \in H \qquad [16]$$

where the $\alpha_k$ are given by

$$\alpha_k = (u, \bar{\varphi}_k), \quad k = 0, \pm 1, \pm 2, \ldots \qquad [17]$$

and

$$\varphi_k(x) = \frac{1}{\sqrt{2\pi}} e^{ikx}, \quad k = 0, \pm 1, \pm 2, \ldots \qquad [18]$$

Let

$$N = \{u \in H : \alpha_k = 0 \text{ for } |k| > n\}$$

Thus,

$$\begin{aligned} \|u\|_H^2 &= \sum_{|k|\le n} (1+k^2)|\alpha_k|^2 \\ &\le (1+n^2)\|u\|^2, \quad u \in N \end{aligned} \qquad [19]$$

Let

$$M = \{u \in H : \alpha_k = 0 \text{ for } |k| \le n\}$$

In this case,

$$\begin{aligned} \|u\|_H^2 &= \sum_{|k|\ge n+1} (1+k^2)|\alpha_k|^2 \\ &\ge (1+(n+1)^2)\|u\|^2, \quad u \in M \end{aligned} \qquad [20]$$

Note that $M, N$ are closed subspaces of $H$ and that $M = N^\perp$. Note also that $N$ is finite dimensional. If we consider the functional [8], it is not difficult to show that [11] implies

$$\inf_M G > -\infty; \qquad \sup_N G < \infty \qquad [21]$$

We are now in a position to apply Theorem 3. This produces a saddle point satisfying [1]. □

## Minimax

Theorem 3 is very useful when extrema do not exist, but it is not always applicable. One is then forced to search for other ways of obtaining critical points. Again, one is faced with the fact that there is no systematic method of finding them. A useful idea is to try to find sets that separate the functional. By this we mean the following:

**Definition 1** Two sets $A, B$ separate the functional $G(u)$ if

$$a_0 := \sup_A G \le b_0 := \inf_B G \qquad [22]$$

We would like to find sets $A$ and $B$ such that [22] will imply

$$\exists u : G(u) \ge b_0, \quad G'(u) = 0 \qquad [23]$$

This is too much to expect since even semiboundedness does not imply the existence of an extremum. Consequently, we weaken our requirements and look for sets $A, B$ such that [22] implies

$$G(u_k) \to a, \qquad G'(u_k) \to 0 \qquad [24]$$

with $a \ge b_0$. This leads to

**Definition 2** We shall say that the set $A$ links the set $B$ if [22] implies [24] with $a \ge b_0$ for every $C^1$ functional $G(u)$.

Of course, [24] is a far cry from [23], but if, for example, the sequence [24] has a convergent subsequence, then [24] implies [23]. Whether or not [24] implies [23] is a property of the functional $G(u)$. We state this as

**Definition 3** We say that $G(u)$ satisfies the Palais–Smale (PS) condition if [24] always implies [23].

The usual way of verifying this is to show that every sequence satisfying [24] has a convergent subsequence (there are other ways).

All of this leads to

**Theorem 4** *If G satisfies the PS condition and is separated by a pair of linking sets, then it has a critical point satisfying* [23].

This theorem cannot be applied until one knows if there are linking sets and functionals that satisfy the PS condition. Fortunately, they exist. Examples and sufficient conditions for $A$ to link $B$ are found in the literature. Obviously, the weaker the conditions, the more pairs will qualify. To date, the conditions described in the next section allow all known examples.

## The Details

Let $E$ be a Banach space, and let $\Phi$ be the set of all continuous maps $\Gamma = \Gamma(t)$ from $E \times [0,1]$ to $E$ such that

1. $\Gamma(0) = I$, the identity map;
2. for each $t \in [0,1), \Gamma(t)$ is a homeomorphism of $E$ onto $E$ and $\Gamma^{-1}(t) \in C(E \times [0,1), E)$;
3. $\Gamma(1)E$ is a single point in $E$ and $\Gamma(t)A$ converges uniformly to $\Gamma(1)E$ as $t \to 1$ for each bounded set $A \subset E$; and
4. for each $t_0 \in [0,1)$ and each bounded set $A \subset E$,

$$\sup_{0 \le t \le t_0, u \in A} \{\|\Gamma(t)u\| + \|\Gamma^{-1}(t)u\|\} < \infty \qquad [25]$$

We have the following

**Theorem 5** *A sufficient condition for A to link B is*

(i) $A \cap B = \phi$ and
(ii) *for each* $\Gamma \in \Phi$ *there is a* $t \in (0,1]$ such that

$$\Gamma(t)A \cap B \neq \phi$$

**Theorem 6** *Let G be a* $C^1$*-functional on E, and let A, B be subsets of E such that A, B satisfy* [22] *and the hypotheses of* Theorem 5. *Assume that*

$$a := \inf_{\Gamma \in \Phi} \sup_{0 \le s \le 1, u \in A} G(\Gamma(s)u) \qquad [26]$$

*is finite. Let* $\psi(t)$ *be a positive, locally Lipschitz continuous function on* $[0,\infty)$ *such that*

$$\int_0^\infty \psi(r)\,\mathrm{d}r = \infty \qquad [27]$$

*Then there is a sequence* $\{u_k\} \subset E$ *such that*

$$G(u_k) \to a, \qquad G'(u_k)/\psi(\|u_k\|) \to 0 \qquad [28]$$

*If* $a = b_0$*, then we can also require that*

$$\mathrm{d}(u_k, B) \to 0 \qquad [29]$$

**Corollary 1** *Under the hypotheses of* Theorem 6 *there is a sequence* $\{u_k\} \subset E$ *such that*

$$G(u_k) \to a, \qquad (1 + \|u_k\|)G'(u_k) \to 0 \qquad [30]$$

**Proof.** We merely take $\psi(u) = 1/(1 + \|u\|)$ in Theorem 6. □

A useful criterion for finding linking subsets is

**Theorem 7** *Let F be a continuous map from a Banach space E to* $\mathbb{R}^n$*, and let* $Q \subset E$ *be such that* $F_0 = F|_Q$ *is a homeomorphism of Q onto the closure of a bounded open subset* $\Omega$ *of* $\mathbb{R}^n$*. If* $p \in \Omega$*, then* $F_0^{-1}(\partial\Omega)$ *links* $F^{-1}(p)$.

## Some Examples

The following are examples of sets that link.

**Example 1** Let $M, N$ be closed subspaces such that $E = M \oplus N$ (with one finite dimensional). Let

$$B_R = \{u \in E : \|u\| < R\}$$

and take $A = \partial B_R \cap N, B = M$. Then $A$ links $B$. To see this, we identify $N$ with some $\mathbb{R}^n$ and take $\Omega = B_R \cap N, Q = \bar{\Omega}$. For $u \in E$, we write

$$u = v + w, \quad v \in N, \ w \in M \qquad [31]$$

and take $F$ to be the projection

$$Fu = v$$

Since $F|_Q = I$ and $M = F^{-1}(0)$, we see from Theorem 7 that $A$ links $B$.

**Example 2** We take $M, N$ as in Example 1. Let $w_0 \neq 0$ be an element of $M$, and take

$$\begin{aligned} A &= \{v \in N : \|v\| \le R\} \\ &\quad \cup \{sw_0 + v : v \in N, s \ge 0, \|sw_0 + v\| = R\} \\ B &= \partial B_\delta \cap M, \quad 0 < \delta < R. \end{aligned}$$

Then $A$ links $B$. Again we identify $N$ with some $\mathbb{R}^n$, and we may assume $\|w_0\| = 1$. Let

$$Q = \{sw_0 + v : v \in N, s \ge 0, \|sw_0 + v\| \le R\}$$

Then $A = \partial Q$ in $\mathbb{R}^{n+1}$. If $u$ is given by [31], we define

$$Fu = v + \|w\|w_0$$

Then $F|_Q = I$ and $B = F^{-1}(\delta w_0)$. We can now apply Theorem 7 to conclude that $A$ links $B$.

**Example 3** Take $M, N$ as before and let $v_0 \neq 0$ be an element of $N$. We write $N = \{v_0\} \oplus N'$. We take

$$A = \{v' \in N' : \|v'\| \le R\}$$
$$\cup \{sv_0 + v' : v' \in N', s \ge 0, \|sv_0 + v'\| = R\}$$
$$B = \{w \in M : \|w\| \ge \delta\}$$
$$\cup \{sv_0 + w : w \in M, s \ge 0, \|sv_0 + w\| = \delta\}$$

where $0 < \delta < R$. Then $A$ links $B$. To see this, we let

$$Q = \{sv_0 + v' : v' \in N', s \ge 0, \|sv_0 + v'\| \le R\}$$

and reason as before. For simplicity, we assume that $\|v_0\| = 1$, $E$ is a Hilbert space and that the splitting $E = N' \oplus \{v_0\} \oplus M$ is orthogonal. If

$$u = v' + w + sv_0, \quad v' \in N',\ w \in M,\ s \in \mathbb{R} \qquad [32]$$

we define

$$F(u) = v' + \left(s + \delta - \sqrt{\delta^2 - \|w\|^2}\right)v_0, \quad \|w\| \le \delta$$
$$= v' + (s + \delta)v_0, \quad \|w\| > \delta$$

Note that $F|_Q = I$ while $F^{-1}(\delta v_0)$ is precisely the set $B$. Hence we can conclude via Theorem 7 that $A$ links $B$.

**Example 4** This is the same as Example 3 with $A$ replaced by $A = \partial B_R \cap N$. The proof is the same with $Q$ replaced by $Q = \bar{B}_R \cap N$.

**Example 5** Let $M, N$ be as in Example 1. Take $A = \partial B_\delta \cap N$, and let $v_0$ be any element in $\partial B_1 \cap N$. Take $B$ to be the set of all $u$ of the form

$$u = w + sv_0, \quad w \in M$$

satisfying any of the following:

(i) $\|w\| \le R, s = 0$,
(ii) $\|w\| \le R, s = 2R_0$, and
(iii) $\|w\| = R, 0 \le s \le 2R_0$

where $0 < \delta < \min(R, R_0)$. Then $A$ links $B$. To see this, take $N = \{v_0\} \oplus N'$. Then any $u \in E$ can be written in the form [32]. Define

$$F(u) = v' + \left(R_0 - \max\left\{\frac{R_0}{R}\|w\|, |s - R_0|\right\}\right)v_0$$

and $Q = \bar{B}_\delta \cap N$. Again we may identify $N$ with some $\mathbb{R}^n$. Then $F \in C(E, N)$ and $F|_Q = I$. Moreover, $A = F^{-1}(0)$. Hence, $A$ links $B$ by Theorem 7.

**Example 6** Let $M, N$ be as in Example 1. Let $v_0$ be in $\partial B_1 \cap N$ and write $N = \{v_0\} \oplus N'$. Let $A = \partial B_\delta \cap N$, $Q = \bar{B}_\delta \cap N$, and

$$B = \{w \in M : \|w\| \le R\}$$
$$\cup \{w + sv_0 : w \in M, s \ge 0, \|w + sv_0\| = R\}$$

where $0 < \delta < R$. Then $A$ links $B$. To see this, write $u = w + v' + sv_0$, $w \in M$, $v' \in N'$, $s \in \mathbb{R}$ and take

$$F(u) = (cR - \max\{c\|w + sv_0\|, |cR - s|\})v_0 + v'$$

where $c = \delta/(R - \delta)$. Then $F$ is the identity operator on $Q$, and $F^{-1}(0) = B$. Apply Theorem 7.

## Some Applications

Many elliptic semilinear problems can be described in the following way. Let $\Omega$ be a domain in $\mathbb{R}^n$, and let $A$ be a self-adjoint operator on $L^2(\Omega)$. We assume that $A \ge \lambda_0 > 0$ and that

$$C_0^\infty(\Omega) \subset D := D(A^{1/2}) \subset H^{m,2}(\Omega) \qquad [33]$$

for some $m > 0$, where $C_0^\infty(\Omega)$ denotes the set of test functions in $\Omega$ (i.e., infinitely differentiable functions with compact supports in $\Omega$), and $H^{m,2}(\Omega)$ denotes the Sobolev space. If $m$ is an integer, the norm in $H^{m,2}(\Omega)$ is given by

$$\|u\|_{m,2} := \left(\sum_{|\mu| \le m} \|D^\mu u\|^2\right)^{1/2} \qquad [34]$$

Here $D^\mu$ represents the generic derivative of order $|\mu|$ and the norm on the right-hand side of [34] is that of $L^2(\Omega)$. We shall not assume that $m$ is an integer.

Let $q$ be any number satisfying

$$2 \le q \le 2n/(n - 2m), \quad 2m < n$$
$$2 \le q < \infty, \qquad n \le 2m$$

and let $f(x, t)$ be a continuous function on $\Omega \times \mathbb{R}$. We make the following assumptions.

**Assumption A** The function $f(x, t)$ satisfies

$$|f(x, t)| \le V_0(x)^q |t|^{q-1} + V_0(x) W_0(x) \qquad [35]$$

and

$$f(x, t)/V_0(x)^q = o(|t|^{q-1}) \text{ as } |t| \to \infty \qquad [36]$$

where $V_0(x) > 0$ is a function in $L^q(\Omega)$ such that

$$\|V_0 u\|_q \le C\|u\|_D, \quad u \in D \qquad [37]$$

and $W_0$ is a function in $L^{q'}(\Omega)$. Here

$$\|u\|_q := \left(\int_\Omega |u(x)|^q \, dx\right)^{1/q} \quad [38]$$

$$\|u\|_D := \|A^{1/2}u\| \quad [39]$$

and $q' = q/(q-1)$. With the norm [39], $D$ becomes a Hilbert space. Define $G$ and $F$ by [8] and [2]. It follows that $G$ is a continuously differentiable functional on the whole of $D$.

We assume further that

$$H(x,t) = 2F(x,t) - tf(x,t) \geq -W_1(x) \in L^1(\Omega), \quad x \in \Omega, t \in \mathbb{R} \quad [40]$$

and

$$H(x,t) \to \infty \quad \text{a.e. as } |t| \to \infty \quad [41]$$

Moreover, we assume that there are functions $V(x), W(x) \in L^2(\Omega)$ such that multiplication by $V(x)$ is a compact operator from $D$ to $L^2(\Omega)$ and

$$F(x,t) \leq C(V(x)^2|t|^2 + V(x)W(x)|t|) \quad [42]$$

We wish to obtain a solution of

$$Au = f(x,u), \quad u \in D \quad [43]$$

By a solution of [43] we shall mean a function $u \in D$ such that

$$(u,v)_D = (f(\cdot,u),v), \quad v \in D \quad [44]$$

If $f(x,u)$ is in $L^2(\Omega)$, then a solution of [44] is in $D(A)$ and solves [43] in the classical sense. Otherwise we call it a weak or semistrong solution. We have

**Theorem 8** *Let $A$ be a self-adjoint operator in $L^2(\Omega)$ such that $A \geq \lambda_0 > 0$ and [33] holds for some $m > 0$. Assume that $\lambda_0$ is an eigenvalue of $A$ with eigenfunction $\varphi_0$. Assume also*

$$2F(x,t) \leq \lambda_0 t^2, \quad |t| \leq \delta \text{ for some } \delta > 0 \quad [45]$$

*and*

$$2F(x,t) \geq \lambda_0 t^2 - W_0(x), \quad t > 0, x \in \Omega \quad [46]$$

*where $W_0 \in L^1(\Omega)$. Assume that $f(x,t)$ satisfies [35], [36], [40], [41], and [42]. Then [43] has a solution $u \neq 0$.*

**Proof.** Under the hypotheses of the theorem, it is known that the following alternative holds: either

(i) there is an infinite number of $y(x) \in D(A)\backslash\{0\}$ such that

$$Ay = f(x,y) = \lambda_0 y \quad [47]$$

or

(ii) for each $\rho > 0$ sufficiently small, there is an $\varepsilon > 0$ such that

$$G(u) \geq \varepsilon, \|u\|_D = \rho \quad [48]$$

We may assume that option (ii) holds, for otherwise we are done. By [46] we have

$$G(R\varphi_0) \leq R^2(\|\varphi_0\|_D^2 - \lambda_0\|\varphi_0\|^2) + \int_\Omega W_0(x)\, dx = \int_\Omega W_0(x)\, dx$$

By Theorem 6, there is a sequence satisfying [28]. Taking $\psi(r) = 1/(r+1)$, we conclude that there is a sequence $\{u_k\} \subset D$ such that

$$G(u_k) \to c, \quad m_0 \leq c \leq m_1, \quad (1 + \|u_k\|_D)G'(u_k) \to 0 \quad [49]$$

In particular, we have

$$\|u_k\|_D^2 - 2\int_\Omega F(x,u_k)\, dx \to c \quad [50]$$

and

$$\|u_k\|_D^2 - (f(\cdot,x_k),u_k) \to 0 \quad [51]$$

Consequently,

$$\int_\Omega H(x,u_k)\, dx \to -c \quad [52]$$

These imply

$$\int_\Omega H(x,u_k)\, dx \leq K \quad [53]$$

If $\rho_k = \|u_k\|_D \to \infty$, let $\tilde{u}_k = u_k/\rho_k$. Then $\|\tilde{u}_k\|_D = 1$. Consequently, there is a renamed subsequence such that $\tilde{u}_k \to \tilde{u}$ weakly in $D$, strongly in $L^2(\Omega)$, and a.e. in $\Omega$. We have from [42]

$$1 \leq (m_1 + \delta)/\rho_k^2 + 2C\int_\Omega \{V(x)^2\tilde{u}_k^2 + V(x)W(x)|\tilde{u}_k|\rho_k^{-1}\}\, dx$$

Consequently,

$$1 \leq 2C\int_\Omega V(x)^2\tilde{u}^2\, dx \quad [54]$$

This shows that $\tilde{u} \not\equiv 0$. Let $\Omega_0$ be the subset of $\Omega$ on which $\tilde{u} \neq 0$. Then

$$|u_k(x)| = \rho_k|\tilde{u}_k(x)| \to \infty, \quad x \in \Omega_0 \quad [55]$$

If $\Omega_1 = \Omega \backslash \Omega_0$, then we have

$$\int_{\Omega} H(x, u_k)\,\mathrm{d}x = \int_{\Omega_0} + \int_{\Omega_1} \geq \int_{\Omega_0} H(x, u_k)\,\mathrm{d}x - \int_{\Omega_1} W_1(x)\,\mathrm{d}x \to \infty \qquad [56]$$

This contradicts [53], and we see that $\rho_k = \|u_k\|_D$ is bounded. Once we know that the $\rho_k$ are bounded, we can apply well-known theorems to obtain the desired conclusion. □

**Remark 1** It should be noted that the crucial element in the proof of Theorem 8 was [51]. If we had been dealing with an ordinary Palais–Smale sequence, we could only conclude that

$$\|u_k\|_D^2 - (f(\cdot, u_k), u_k) = o(\rho_k)$$

which would imply only

$$\int_{\Omega} H(x, u_k)\,\mathrm{d}x = o(\rho_k)$$

This would not contradict [56], and the argument would not go through.

As another application, we wish to solve

$$-x''(t) = \nabla_x V(t, x(t)) \qquad [57]$$

where

$$x(t) = (x_1(t), \ldots, x_n(t)) \qquad [58]$$

is a map from $I = [0, 2\pi]$ to $\mathbb{R}^n$ such that each component $x_j(t)$ is a periodic function in $H^1$ with period $2\pi$, and the function

$$V(t, x) = V(t, x_1, \ldots, x_n)$$

is continuous from $\mathbb{R}^{n+1}$ to $\mathbb{R}$ with a gradient

$$\nabla_x V(t, x) = (\partial V/\partial x_1, \ldots, \partial V/\partial x_n) \in C(\mathbb{R}^{n+1}, \mathbb{R}^n) \qquad [59]$$

For each $x \in \mathbb{R}^n$, the function $V(t, x)$ is periodic in $t$ with period $2\pi$. We shall study this problem under the following assumptions:

1. $$0 \leq V(t, x) \leq C(|x|^2 + 1)$$
$$t \in I, x \in \mathbb{R}^n$$
2. There are constants $m > 0, \alpha \leq 3m^2/2\pi^2$ such that
$$V(t, x) \leq \alpha, \quad |x| \leq m, t \in I, x \in \mathbb{R}^n$$
3. There are constants $\beta > 1/2$ and $C$ such that
$$V(t, x) \geq \beta |x|^2$$
when
$$|x| > C, t \in I, x \in \mathbb{R}^n$$
4. The function given by
$$H(t, x) = 2V(t, x) - \nabla_x V(t, x) \cdot x \qquad [60]$$
satisfies
$$H(t, x) \leq W(t) \in L^1(I), \quad |x| \geq C \qquad [61]$$
$t \in I$, $x \in \mathbb{R}^n$, and
$$H(t, x) \to -\infty \quad \text{as } |x| \to \infty \qquad [62]$$

We have

**Theorem 9** *Under the above hypotheses, the system [57] has a nonconstant solution.*

**Proof.** Let $X$ be the set of vector functions $x(t)$ described above. It is a Hilbert space with norm satisfying

$$\|x\|_X^2 = \sum_{j=1}^{n} \|x_j\|_{H^1}^2$$

We also write

$$\|x\|^2 = \sum_{j=1}^{n} \|x_j\|^2$$

where $\|\cdot\|$ is the $L^2(I)$ norm. Let

$$N = \{x(t) \in X : x_j(t) \equiv \text{constant}, 1 \leq j \leq n\}$$

and $M = N^{\perp}$. The dimension of $N$ is $n$, and $X = M \oplus N$. The following is easily proved.

**Lemma 1** *If $x \in M$, then*

$$\|x\|_\infty^2 \leq \frac{\pi}{6} \|x'\|^2$$

*and*

$$\|x\| \leq \|x'\|$$

We define

$$G(x) = \|x'\|^2 - 2\int_I V(t, x(t))\,\mathrm{d}t, \quad x \in X \qquad [63]$$

For each $x \in X$ write $x = v + w$, where $v \in N$, $w \in M$. For convenience, we shall use the following equivalent norm for $X$:

$$\|x\|_X^2 = \|w'\|^2 + \|v\|^2$$

If $x \in M$ and

$$\|x'\|^2 = \rho^2 = \frac{6}{\pi} m^2$$

then Lemma 1 implies that $\|x\|_\infty \leq m$, and we have by Hypothesis 2 that $V(t, x) \leq \alpha$.

Hence,

$$G(x) \geq \|x'\|^2 - 2\int_{|x|<m} \alpha \, dt \geq \rho^2 - 2\alpha(2\pi) \geq 0 \qquad [64]$$

Note that Hypothesis 3 is equivalent to

$$V(t,x) \geq \beta|x|^2 - C, \quad t \in I, \ x \in \mathbb{R}^n \qquad [65]$$

for some constant $C$. Next, let

$$y(t) = v + sw_0$$

where $v \in N$, $s \geq 0$, and

$$w_0 = (\sin t, 0, \ldots, 0)$$

Then $w_0 \in M$, and

$$\|w_0\|^2 = \|w_0'\|^2 = \pi$$

Note that

$$\|y\|^2 = \|v\|^2 + s^2\pi = 2\pi|v|^2 + \pi s^2$$

Consequently,

$$\begin{aligned} G(y) &= s^2\|w_0'\|^2 - 2\int_I V(t, y(t))\, dt \\ &\leq \pi s^2 - 2\beta\int_I |y(t)|^2 \, dt + 2\pi C \\ &= \pi s^2 - 2\beta(\|v\|^2 + \pi s^2) + 2\pi C \\ &\leq (1-2\beta)\pi s^2 - 4\beta\pi|v|^2 + 2\pi C \\ &\to -\infty \text{ as } s^2 + |v|^2 \to \infty \end{aligned}$$

We also note that Hypothesis 1 implies

$$G(v) \leq 0, \quad v \in N \qquad [66]$$

Take

$$\begin{aligned} A &= \{v \in N : \|v\| \leq R\} \\ &\quad \cup \{sw_0 + v : v \in N, s \geq 0, \|sw_0 + v\|_X = R\} \\ B &= \partial B_\rho \cap M, 0 < \rho = 6m^2/\pi < R \end{aligned}$$

where

$$B_\sigma = \{x \in X : \|x\|_X < \sigma\}$$

By Example 2, $A$ links $B$. Moreover, if $R$ is sufficiently large,

$$\sup_A G = 0 \leq \inf_B G \qquad [67]$$

Hence, we may conclude that there is a sequence $\{x^{(k)}\} \subset X$ such that

$$G(x^{(k)}) \to c \geq 0, \quad (1 + \|x^{(k)}\|_X)\, G'(x^{(k)}) \to 0$$

Hence,

$$G(x^{(k)}) = \|[x^{(k)}]'\|^2 - 2\int_I V(t, x^{(k)}(t))\, dt \to c \geq 0 \qquad [68]$$

$$(G'(x^{(k)}), z)/2 = ([x^{(k)}]', z') - \int_I \nabla_x V(t, x^{(k)}) \cdot z(t)\, dt \to 0, \quad z \in X \qquad [69]$$

and

$$(G'(x^{(k)}), x^{(k)})/2 = \|[x^{(k)}]'\|^2 - \int_I \nabla_x V(t, x^{(k)}) \cdot x^{(k)}\, dt \to 0 \qquad [70]$$

If

$$\rho_k = \|x^{(k)}\|_X \leq C$$

then there is a renamed subsequence such that $x^{(k)}$ converges to a limit $x \in X$ weakly in $X$ and uniformly on $I$. From [69] we see that

$$(G'(x), z)/2 = (x', z') - \int_I \nabla_x V(t, x(t)) \cdot z(t)\, dt = 0, \quad z \in X$$

from which we conclude easily that $x$ is a solution of [57]. From [68], we see that

$$G(x) \geq c \geq 0$$

showing that $x(t)$ is not a constant. For if $c > 0$ and $x \in N$, then

$$G(x) = -2\int_I V(t, x(t))\, dt \leq 0$$

If $c = 0$, we see that $x \in B$ by Theorem 6. Hence, $x \in M$. If

$$\rho_k = \|x^{(k)}\|_X \to \infty$$

let $\tilde{x}^{(k)} = x^{(k)}/\rho_k$. Then, $\|\tilde{x}^{(k)}\|_X = 1$. Let $\tilde{x}^{(k)} = \tilde{w}^{(k)} + \tilde{v}^{(k)}$, where $\tilde{w}^{(k)} \in M$ and $\tilde{v}^{(k)} \in N$. There is a renamed subsequence such that $\tilde{x}^{(k)}$ converges uniformly in $I$ to a limit $\tilde{x}$ and $\|[\tilde{x}^{(k)}]'\| \to r$ and $\|\tilde{x}^{(k)}\| \to \tau$, where $r^2 + \tau^2 = 1$. From [68] and [70], we obtain

$$\|[\tilde{x}^{(k)}]'\|^2 - 2\int_I V(t, x^{(k)}(t))\, dt/\rho_k^2 \to 0$$

and

$$\|[\tilde{x}^{(k)}]'\|^2 - \int_I \nabla_x V(t, x^{(k)}) \cdot x^{(k)}\, dt/\rho_k^2 \to 0$$

Thus,

$$2\int_I V(t, x^{(k)}(t))\, dt/\rho_k^2 \to r^2 \qquad [71]$$

and

$$\int_I \nabla_x V(t, x^{(k)}) \cdot x^{(k)}\, dt/\rho_k^2 \to r^2 \qquad [72]$$

Hence,

$$\int_I H(t, x^{(k)}(t))\, dt/\rho_k^2 \to 0 \qquad [73]$$

By Hypothesis 3, the left-hand side of [71] is

$$\geq 2\beta \|\tilde{x}^{(k)}\|^2 - 4\pi C/\rho_k^2$$

Thus,

$$r^2 \geq 2\beta\tau^2 = 2\beta(1 - r^2)$$

showing that $r > 0$. Hence, $\tilde{x}(t) \not\equiv 0$. Let $\Omega_0 \subset I$ be the set on which $[\tilde{x}(t)] \neq 0$. The measure of $\Omega_0$ is positive. Thus, $|x^{(k)}(t)| \to \infty$ as $k \to \infty$ for $t \in \Omega_0$. Hence,

$$\int_I H(t, x^{(k)}(t))\, dt \leq \int_{\Omega_0} H(t, x^{(k)}(t))\, dt + \int_{I\setminus\Omega_0} W(t)\, dt \to -\infty$$

contrary to Hypothesis 4. Thus, the $\rho_k$ are bounded, and the proof is complete. □

## Superlinear Problems

Consider the problem

$$-\Delta u = f(x, u), \quad x \in \Omega;\ u = 0 \text{ on } \partial\Omega \qquad [74]$$

where $\Omega \subset \mathbb{R}^n$ is a bounded domain whose boundary is a smooth manifold, and $f(x, t)$ is a continuous function on $\bar{\Omega} \times \mathbb{R}$. This semilinear Dirichlet problem has been studied by many authors. It is called "sublinear" if there is a constant $C$ such that

$$|f(x, t)| \leq C(|t| + 1), \quad x \in \Omega, t \in \mathbb{R}$$

Otherwise, it is called "superlinear". Assume

($a_1$) There are constants $c_1, c_2 \geq 0$ such that

$$|f(x, t)| \leq c_1 + c_2|t|^s$$

where $0 \leq s < (n + 2)/(n - 2)$ if $n > 2$.

($a_2$) $f(x, t) = o(|t|)$ as $t \to 0$.

($a_3$) Either

$$F(x, t)/t^2 \to \infty \text{ as } t \to \infty$$

or

$$F(x, t)/t^2 \to \infty \text{ as } t \to -\infty.$$

We have

**Theorem 10** *Under hypotheses* ($a_1$)–($a_3$) *the boundary-value problem*

$$-\Delta u = \beta f(x, u),\ x \in \Omega; \quad u = 0 \text{ on } \partial\Omega \qquad [75]$$

*has a nontrivial solution for almost every positive* $\beta$.

Unfortunately, this theorem does not give any information for any specific $\beta$. It still leaves open the problem of solving [74]. For this purpose, we add the assumption

($a_4$) There are constants $\mu > 2, r \geq 0$ such that

$$\mu F(x, t) - tf(x, t) \leq C(t^2 + 1), \quad |t| \geq r \qquad [76]$$

We have

**Theorem 11** *Under hypotheses* ($a_1$)–($a_4$) *problem* [74] *has a nontrivial solution.*

We also have

**Theorem 12** *If we replace hypothesis* ($a_4$) *with*

($a_4'$) *The function* $-H(x, t)$ *is convex in* $t$,

*then the problem* [74] *has at least one nontrivial solution.*

## Weak Linking

It is not clear if it is possible for $A$ to link $B$ if neither is contained in a finite-dimensional manifold. For instance, if $E = M \oplus N$, where $M, N$ are closed infinite-dimensional subspaces of $E$ and $B_R$ is the ball centered at the origin of radius $R$ in $E$, it is unknown if the set $A = M \cap \partial B_R$ links $B = N$. (If either $M$ or $N$ is finite dimensional, then $A$ does link $B$.) Unfortunately, this is the situation which arises in some important applications including Hamiltonian systems, the wave equation and elliptic systems, to name a few.

We now consider linking when both $M$ and $N$ are infinite dimensional and $G'$ has some additional continuity property. A property that is very useful is that of weak-to-weak continuity:

$$u_k \to u \text{ weakly in } E \implies G'(u_k) \to G'(u) \text{ weakly} \qquad [77]$$

We make the following definition:

**Definition 3** A subset $A$ of a Banach space $E$ links a subset $B$ of $E$ "weakly" if for every $G \in C^1(E, \mathbb{R})$ satisfying [77] and

$$a_0 := \sup_A G \leq b_0 := \inf_B G \qquad [78]$$

there is a sequence $\{u_k\} \subset E$ and a constant $c$ such that

$$b_0 \leq c < \infty \qquad [79]$$

and

$$G(u_k) \to c, \qquad G'(u_k) \to 0 \qquad [80]$$

We have the following counterpart of Theorem 7.

**Theorem 13** *Let $E$ be a separable Hilbert space, and let $G$ be a continuous functional on $E$ with a continuous derivative satisfying [77]. Let $N$ be a closed subspace of $E$, and let $Q$ be a bounded open subset of $N$ containing the point $p$. Let $F$ be a continuous map of $E$ onto $N$ such that*

(i) $F|_Q = I$, *and*
(ii) *For each finite-dimensional subspace $S \neq \{0\}$ of $E$ containing $p$, there is a finite-dimensional subspace $S_0 \neq \{0\}$ of $N$ containing $p$ such that*

$$v \in \bar{Q} \cap S_0, \qquad w \in S \Longrightarrow F(v+w) \in S_0 \qquad [81]$$

*Set $A = \partial Q, B = F^{-1}(p)$. If*

$$a_1 = \sup_{\bar{Q}} G < \infty \qquad [82]$$

*and [22] holds, then there is a sequence $\{u_k\} \subset E$ such that [24] holds with $a \le a_1$.*

Theorem 13 states that if $Q, F, p$ satisfy the hypotheses of that theorem, then $A = \partial Q$ links $B = F^{-1}(p)$ weakly. It follows from this theorem that all sets $A, B$ known to link when one of the subspaces $M, N$ is finite dimensional will link weakly even when $M, N$ are both infinite dimensional.

Now we give some applications of Theorem 13 to semilinear boundary-value problems. Let $\Omega$ be a domain in $\mathbb{R}^n$ and let $A$ be a self–adjoint operator in $L^2(\Omega)$ having 0 in its resolvent set (thus, there is an interval $(a, b)$ in its resolvent set satisfying $a < 0 < b$). Let $f(x,t)$ be a continuous function on $\Omega \times \mathbb{R}$ such that

$$|f(x,t)| \le V(x)^2|t| + W(x)V(x) \qquad [83]$$

$x \in \Omega$, $t \in \mathbb{R}$, and

$$f(x,t)/t \to \alpha_\pm(x) \quad \text{as } t \to \pm\infty \qquad [84]$$

where $V, W \in L^2(\Omega)$, and multiplication by $V(x) > 0$ is a compact operator from $D = D(|A|^{1/2})$ to $L^2(\Omega)$. Let

$$M = \int_b^\infty \mathrm{d}E(\lambda)D, \qquad N = \int_{-\infty}^a \mathrm{d}E(\lambda)D$$

where $\{E(\lambda)\}$ is the spectral measure of $A$. Then $M, N$ are invariant subspaces for $A$ and $D = M \oplus N$. If

$$\alpha(u,v) = \int_\Omega (\alpha_+ u^+ - \alpha_- u^-)v\,\mathrm{d}x \qquad [85]$$

$\alpha(u) = \alpha(u,u)$, then we assume that

$$\alpha(v) \ge (Av, v), \quad v \in N \qquad [86]$$

$$(Aw, w) \ge \alpha(w), \quad w \in M \qquad [87]$$

We also assume that the only solution of

$$Au = \alpha_+ u^+ - \alpha_- u^- \qquad [88]$$

is $u \equiv 0$, where $u^\pm = \max\{\pm u, 0\}$. We have

**Theorem 14** *Under the above hypotheses there is at least one solution of*

$$Au = f(x,u), \quad u \in D(A) \qquad [89]$$

Next, we consider an application concerning radially symmetric solutions for the problem

$$u_{tt} - \Delta u = f(t,x,u), \quad t \in \mathbb{R},\ x \in B_R \qquad [90]$$

$$u(t,x) = 0, \quad t \in \mathbb{R},\ x \in \partial B_R \qquad [91]$$

$$u(t+T,x) = u(t,x), \quad t \in \mathbb{R},\ x \in B_R \qquad [92]$$

where $B_R = \{x \in \mathbb{R}^n : |x| < R\}$. We assume that the ratio $R/T$ is rational. Let

$$8R/T = a/b \qquad [93]$$

where $a, b$ are relatively prime positive integers. It can be shown that

$$n \not\equiv 3 \ (\mathrm{mod}(4,a)) \qquad [94]$$

implies that the linear problem corresponding to [90]–[92] has no essential spectrum. If

$$n \equiv 3 \ (\mathrm{mod}(4,a)) \qquad [95]$$

then the essential spectrum of the linear operator consists of precisely one point

$$\lambda_0 = -(n-3)(n-1)/4R^2 \qquad [96]$$

Consider the case

$$f(t,r,s) = \mu s + p(t,r,s) \qquad [97]$$

where $\mu$ is a point in the resolvent set, $r = |x|$, and

$$|p(t,r,s)| \le C(|s|^\theta + 1), \quad s \in \mathbb{R} \qquad [98]$$

for some number $\theta < 1$. We then have

**Theorem 15** *If [94] holds, then [90]–[92] have a weak rotationally invariant solution. If [95] holds and $\lambda_0 < \mu$, assume in addition that $p(t,r,s)$ is nondecreasing in $s$. If $\mu < \lambda_0$, assume that $p(t,r,s)$ is nonincreasing in $s$. Then [90]–[92] have a weak rotationally invariant solution.*

*See also:* Combinatorics: Overview; Homoclinic Phenomena; Ljusternik–Schnirelman Theory; Minimax Principle in the Calculus of Variations.

## Further Reading

Ambrosetti A and Prodi G (1993) A primer of nonlinear analysis. *Cambridge Studies in Advanced Mathematics* 34. Cambridge: Cambridge University Press.
Chang KC (1993) *Infinite Dimensional Morse Theory and Multiple Solution Problems*. Boston: Birkhäuser.
Ekeland I and Temam R (1976) *Convex Analysis and Variational Problems*. Amsterdam: North-Holland.
Ghoussoub N (1993) *Duality and Perturbation Methods in Critical Point Theory*. Cambridge: Cambridge University Press.
Mawhin J and Willem M (1989) *Critical Point Theory and Hamiltonian Systems*. Berlin: Springer.
Rabinowitz PR (1986) *Minimax Methods in Critical Point Theory with Applications to Differential Equations*, Conf. Board of Math. Sci. Reg. Conf. Ser. in Math. No. 65. Providence, RI: American Mathematical Society.
Schechter M (1999) *Linking Methods in Critical Point Theory*. Boston: Birkhäuser.
Schechter M (1986) *Spectra of Partial Differential Operators*, 2nd edn. Amsterdam: North-Holland.
Struwe M (1996) *Variational Methods*. Berlin: Springer.
Willem M (1996) *Minimax Theorems*. Boston: Birkhäuser.

# Scattering in Relativistic Quantum Field Theory: Fundamental Concepts and Tools

**D Buchholz**, Universität Göttingen, Göttingen, Germany
**S J Summers**, University of Florida, Gainesville, FL, USA



## Physical Motivation and Mathematical Setting

The primary connection of relativistic quantum field theory to experimental physics is through scattering theory, that is, the theory of the collision of elementary (or compound) particles. It is therefore a central topic in quantum field theory and has attracted the attention of leading mathematical physicists. Although a great deal of progress has been made in the mathematically rigorous understanding of the subject, there are important matters which are still unclear, some of which will be indicated below.

In the paradigmatic scattering experiment, several particles, which are initially sufficiently distant from each other that the idealization that they are not mutually interacting is physically reasonable, approach each other and interact (collide) in a region of microscopic extent. The products of this collision then fly apart until they are sufficiently well separated that the approximation of noninteraction is again reasonable. The initial and final states of the objects in the scattering experiment are therefore to be modeled by states of noninteracting, that is, free, fields, which are mathematically represented on Fock space. Typically, what is measured in such experiments is the probability distribution (cross section) for the transitions from a specified state of the incoming particles to a specified state of the outgoing particles.

It should be mentioned that until the late 1950s, the scattering theory of relativistic quantum particles relied upon ideas from nonrelativistic quantum-mechanical scattering theory (interaction representation, adiabatic limit, etc.), which were invalid in the relativistic context. Only with the advent of axiomatic quantum field theory did it become possible to properly formulate the concepts and mathematical techniques which will be outlined here.

Scattering theory can be rigorously formulated either in the context of quantum fields satisfying the Wightman axioms (Streater and Wightman 1964) or in terms of local algebras satisfying the Haag–Kastler–Araki axioms (Haag 1992). In brief, the relation between these two settings may be described as follows: in the Wightman setting, the theory is formulated in terms of operator-valued distributions $\phi$ on Minkowski space, the quantum fields, which act on the physical state space. These fields, integrated with test functions $f$ having support in a given region $\mathcal{O}$ of spacetime (only four-dimensional Minkowski space $\mathbb{R}^4$ will be treated here), $\phi(f) = \int d^4x\, f(x)\phi(x)$, form under the operations of addition, multiplication, and Hermitian conjugation a polynomial $^*$-algebra $\mathcal{P}(\mathcal{O})$ of unbounded operators. In the Haag–Kastler–Araki setting, one proceeds from these algebras to algebras $\mathcal{A}(\mathcal{O})$ of bounded operators which, roughly speaking, are formed by the bounded functions $A$ of the operators $\phi(f)$. This step requires some mathematical care, but these subtleties will not be discussed here. As the statements and proofs of the results in these two frameworks differ only in technical details, the theory is presented here in the more convenient setting of algebras of bounded operators ($C^*$-algebras).

Central to the theory is the notion of a particle, which, in fact, is a quite complex concept, the full nature of which is not completely understood, cf.

below. In order to maintain the focus on the essential points, we consider in the subsequent sections primarily a single massive particle of integer spin $s$, that is, a boson. In standard scattering theory based upon Wigner's characterization, this particle is simply identified with an irreducible unitary representation $U_1$ of the identity component $\mathcal{P}_+^\uparrow$ of the Poincaré group with spin $s$ and mass $m > 0$. The Hilbert space $\mathcal{H}_1$ upon which $U_1(\mathcal{P}_+^\uparrow)$ acts is called the one-particle space and determines the possible states of a single particle, alone in the universe. Assuming that configurations of several such particles do not interact, one can proceed by a standard construction to a Fock space describing freely propagating multiple particle states,

$$\mathcal{H}_\mathrm{F} = \bigoplus_{n\in\mathbb{N}_0} \mathcal{H}_n$$

where $\mathcal{H}_0 = \mathbb{C}$ and $\mathcal{H}_n$ is the $n$-fold symmetrized direct product of $\mathcal{H}_1$ with itself. This space is spanned by vectors $\Phi_1 \otimes \cdots \otimes \Phi_n$, where $\otimes$ denotes the symmetrized tensor product, representing an $n$-particle state wherein the $k$th particle is in the state $\Phi_k \in \mathcal{H}_1$, $k = 1, \ldots, n$. The representation $U_1(\mathcal{P}_+^\uparrow)$ induces a unitary representation $U_\mathrm{F}(\mathcal{P}_+^\uparrow)$ on $\mathcal{H}_\mathrm{F}$ by

$$U_\mathrm{F}(\lambda)(\Phi_1 \otimes \cdots \otimes \Phi_n) \doteq U_1(\lambda)\Phi_1 \otimes \cdots \otimes U_1(\lambda)\Phi_n \quad [1]$$

In interacting theories, the states in the corresponding physical Hilbert space $\mathcal{H}$ do not have such an *a priori* interpretation in physical terms, however. It is the primary goal of scattering theory to identify in $\mathcal{H}$ those vectors which describe, at asymptotic times, incoming, respectively, outgoing, configurations of freely moving particles. Mathematically, this amounts to the construction of certain specific isometries (generalized Møller operators), $\Omega^\mathrm{in}$ and $\Omega^\mathrm{out}$, mapping $\mathcal{H}_\mathrm{F}$ onto subspaces $\mathcal{H}^\mathrm{in} \subset \mathcal{H}$ and $\mathcal{H}^\mathrm{out} \subset \mathcal{H}$, respectively, and intertwining the unitary actions of the Poincaré group on $\mathcal{H}_\mathrm{F}$ and $\mathcal{H}$. The resulting vectors

$$(\Phi_1 \otimes \cdots \otimes \Phi_n)^\mathrm{in/out} \doteq \Omega^\mathrm{in/out}(\Phi_1 \otimes \cdots \otimes \Phi_n) \in \mathcal{H} \quad [2]$$

are interpreted as incoming and outgoing particle configurations in scattering processes wherein the $k$th particle is in the state $\Phi_k \in \mathcal{H}_1$.

If, in a theory, the equality $\mathcal{H}^\mathrm{in} = \mathcal{H}^\mathrm{out}$ holds, then every incoming scattering state evolves, after the collision processes at finite times, into an outgoing scattering state. It is then physically meaningful to define on this space of states the scattering matrix, setting $S = \Omega^\mathrm{in}\Omega^{\mathrm{out}*}$. Physical data such as collision cross sections can be derived from $S$ and the corresponding transition amplitudes $\langle(\Phi_1 \otimes \cdots \otimes \Phi_m)^\mathrm{in}, (\Phi'_1 \otimes \cdots \otimes \Phi'_n)^\mathrm{out}\rangle$, respectively, by a standard procedure. It should be noted, however, that neither the above physically mandatory equality of state spaces nor the more stringent requirement that every state has an interpretation in terms of incoming and outgoing scattering states, that is, $\mathcal{H} = \mathcal{H}^\mathrm{in} = \mathcal{H}^\mathrm{out}$ (asymptotic completeness), has been fully established in any interacting relativistic field theoretic model so far. This intriguing problem will be touched upon in the last section of this article.

Before going into details, let us state the few physically motivated postulates entering into the analysis. As discussed, the point of departure is a family of algebras $\mathcal{A}(\mathcal{O})$, more precisely a net, associated with the open subregions $\mathcal{O}$ of Minkowski space and acting on $\mathcal{H}$. Restricting attention to the case of bosons, we may assume that this net is local in the sense that if $\mathcal{O}_1$ is spacelike separated from $\mathcal{O}_2$, then all elements of $\mathcal{A}(\mathcal{O}_1)$ commute with all elements of $\mathcal{A}(\mathcal{O}_2)$. (In the presence of fermions, these algebras contain also fermionic operators which anticommute.) This is the mathematical expression of the principle of Einstein causality. The unitary representation $U$ of $\mathcal{P}_+^\uparrow$ acting on $\mathcal{H}$ is assumed to satisfy the relativistic spectrum condition (positivity of energy in all Lorentz frames) and, in the sense of equality of sets, $U(\lambda)\mathcal{A}(\mathcal{O})U(\lambda)^{-1} = \mathcal{A}(\lambda\mathcal{O})$ for all $\lambda \in \mathcal{P}_+^\uparrow$ and regions $\mathcal{O}$, where $\lambda\mathcal{O}$ denotes the Poincaré transformed region. It is also assumed that the subspace of $U(\mathcal{P}_+^\uparrow)$-invariant vectors is spanned by a single unit vector $\Omega$, representing the vacuum, which has the Reeh–Schlieder property, that is, each set of vectors $\mathcal{A}(\mathcal{O})\Omega$ is dense in $\mathcal{H}$. These standing assumptions will subsequently be amended by further conditions concerning the particle content of the theory.

## Haag–Ruelle Theory

Haag and Ruelle were the first to establish the existence of scattering states within this general framework (Jost 1965); further substantial improvements are due to Araki and Hepp (Araki 1999). In all of these investigations, the arguments were given for quantum field theories with associated particles (in the Wigner sense) which have strictly positive mass $m > 0$ and for which $m$ is an isolated eigenvalue of the mass operator (upper and lower mass gap). Moreover, it was assumed that states of a single particle can be created from the vacuum by local operations. In physical terms, these assumptions allow only for theories with short-range interactions and particles carrying strictly localizable charges.

In view of these limitations, Haag–Ruelle theory has been developed in a number of different directions. By now, the scattering theory of massive particles is under complete control, including also

particles carrying nonlocalizable (gauge or topological) charges and particles having exotic statistics (anyons, plektons) which can appear in theories in low spacetime dimensions. Due to constraints of space, these results must go without further mention; we refer the interested reader to the articles Buchholz and Fredenhagen (1982) and Fredenhagen *et al.* (1996). Theories of massless particles and of particles carrying charges of electric or magnetic type (infraparticles) will be discussed in subsequent sections.

We outline here a recent generalization of Haag–Ruelle scattering theory presented in Dybalski (2005), which covers massive particles with localizable charges without relying on any further constraints on the mass spectrum. In particular, the scattering of electrically neutral, stable particles fulfilling a sharp dispersion law in the presence of massless particles is included (e.g., neutral atoms in their ground states). Mathematically, this assumption can be expressed by the requirement that there exists a subspace $\mathcal{H}_1 \subset \mathcal{H}$ such that the restriction of $U(\mathcal{P}_+^\uparrow)$ to $\mathcal{H}_1$ is a representation of mass $m > 0$. We denote by $P_1$ the projection in $\mathcal{H}$ onto $\mathcal{H}_1$.

To establish notation, let $\mathcal{O}$ be a bounded spacetime region and let $A \in \mathcal{A}(\mathcal{O})$ be any operator such that $P_1 A\Omega \neq 0$. The existence of such localized (in brief, local) operators amounts to the assumption that the particle carries a localizable charge. That the particle is stable, that is, completely decouples from the underlying continuum states, can be cast into a condition first stated by Herbst: for all sufficiently small $\mu > 0$

$$\|E_\mu(1 - P_1)A\Omega\| \leq c\mu^\eta \qquad [3]$$

for some constants $c, \eta > 0$, where $E_\mu$ is the projection onto the spectral subspace of the mass operator corresponding to spectrum in the interval $(m - \mu, m + \mu)$. In the case originally considered by Haag and Ruelle, where $m$ is isolated from the rest of the mass spectrum, this condition is certainly satisfied.

Setting $A(x) \doteq U(x)AU(x)^{-1}$, where $U(x)$ is the unitary implementing the spacetime translation $x = (x_0, \boldsymbol{x})$ (the velocity of light and Planck's constant are set equal to 1 in what follows), one puts, for $t \neq 0$,

$$A_t(f) = \int \mathrm{d}^4x\, g_t(x_0) f_{x_o}(\boldsymbol{x}) A(x) \qquad [4]$$

Here $x_0 \mapsto g_t(x_0) \doteq g((x_0 - t)/|t|^\kappa)/|t|^\kappa$ induces a time averaging about $t$, $g$ being any test function which integrates to 1 and whose Fourier transform has compact support, and $1/(1+\eta) < \kappa < 1$ with $\eta$ as above. The Fourier transform of $f_{x_0}$ is given by $\widetilde{f_{x_0}}(\boldsymbol{p}) \doteq \widetilde{f}(\boldsymbol{p})\, \mathrm{e}^{-\mathrm{i}x_0\omega(\boldsymbol{p})}$, where $f$ is some test function on $\mathbb{R}^3$ with $\widetilde{f}(\boldsymbol{p})$ having compact support, and $\omega(\boldsymbol{p}) = (\boldsymbol{p}^2 + m^2)^{1/2}$. Note that $(x_0, \boldsymbol{x}) \mapsto f_{x_0}(\boldsymbol{x})$ is a solution of the Klein–Gordon equation of mass $m$.

With these assumptions, it follows by a straightforward application of the harmonic analysis of unitary groups that in the sense of strong convergence $A_t(f)\Omega \to P_1 A(f)\Omega$ and $A_t(f)^*\Omega \to 0$ as $t \to \pm\infty$, where $A(f) = \int \mathrm{d}^3x\, f(\boldsymbol{x}) A(0, \boldsymbol{x})$. Hence, the operators $A_t(f)$ may be thought of as creation operators and their adjoints as annihilation operators. These operators are the basic ingredients in the construction of scattering states. Choosing local operators $A_k$ as above and test functions $f^{(k)}$ with disjoint compact supports in momentum space, $k = 1, \ldots, n$, the scattering states are obtained as limits of the Haag–Ruelle approximants

$$A_{1t}(f^{(1)}) \cdots A_{nt}(f^{(n)})\Omega \qquad [5]$$

Roughly speaking, the operators $A_{kt}(f^{(k)})$ are localized in spacelike separated regions at asymptotic times $t$, due to the support properties of the Fourier transforms of the functions $f^{(k)}$. Hence they commute asymptotically because of locality and, by the clustering properties of the vacuum state, the above vector becomes a product state of single-particle states. In order to prove convergence, one proceeds, in analogy to Cook's method in quantum-mechanical scattering theory, to the time derivatives,

$$\begin{aligned} &\partial_t A_{1t}(f^{(1)}) \cdots A_{nt}(f^{(n)})\Omega \\ &\quad = \sum_{k \neq l} A_{1t}(f^{(1)}) \cdots [\partial_t A_{kt}(f^{(k)}), A_{lt}(f^{(l)})] \cdots A_{nt}(f^{(n)})\Omega \\ &\qquad + \sum_k A_{1t}(f^{(1)}) \cdots \overset{k}{\vee} \cdots A_{nt}(f^{(n)}) \partial_t A_{kt}(f^{(k)})\Omega \end{aligned} \qquad [6]$$

where $\overset{k}{\vee}$ denotes omission of $A_{kt}(f^{(k)})$. Employing techniques of Araki and Hepp, one can prove that the terms in the first summation on the right-hand side (RHS) of [6], involving commutators, decay rapidly in norm as $t$ approaches infinity because of locality, as indicated above. By applying condition [3] and the fact that the vectors $\partial_t A_{kt}(f^{(k)})\Omega$ do not have a component in the single-particle space $\mathcal{H}_1$, the terms in the second summation on the RHS of [6] can be shown to decay in norm like $|t|^{-\kappa(1+\eta)}$. Thus, the norm of the vector [6] is integrable in $t$, implying the existence of the strong limits

$$\begin{aligned} &\left(P_1 A_1(f^{(1)})\Omega \otimes \cdots \otimes P_1 A_n(f^{(n)})\Omega\right)^{\mathrm{in/out}} \\ &\quad \doteq \lim_{t \to \mp\infty} A_{1t}(f^{(1)}) \cdots A_{nt}(f^{(n)})\Omega \end{aligned} \qquad [7]$$

As indicated by the notation, these limits depend only on the single-particle vectors $P_1 A_k(f^{(k)})\Omega \in \mathcal{H}_1$, $k = 1, \ldots, n$, but not on the specific choice of operators and test functions. In order to establish their Fock structure, one employs results on clustering properties of vacuum correlation functions in theories without strictly positive minimal mass. Using this, one can compute inner products of arbitrary asymptotic states and verify that the maps

$$\left(P_1 A_1(f^{(1)})\Omega \otimes \cdots \otimes P_1 A_n(f^{(n)})\Omega\right) \mapsto \left(P_1 A_1(f^{(1)})\Omega \otimes \cdots \otimes P_1 A_n(f^{(n)})\Omega\right)^{\text{in/out}} \quad [8]$$

extend by linearity to isomorphisms $\Omega^{\text{in/out}}$ from the Fock space $\mathcal{H}_\text{F}$ onto the subspaces $\mathcal{H}^{\text{in/out}} \subset \mathcal{H}$ generated by the collision states. Moreover, the asymptotic states transform under the Poincaré transformations $U(\mathcal{P}_+^\uparrow)$ as

$$U(\lambda)\left(P_1 A_1(f^{(1)})\Omega \otimes \cdots \otimes P_1 A_n(f^{(n)})\Omega\right)^{\text{in/out}} = \left(U_1(\lambda) P_1 A_1(f^{(1)})\Omega \otimes \cdots \otimes \times\, U_1(\lambda) P_1 A_n(f^{(n)})\Omega\right)^{\text{in/out}} \quad [9]$$

Thus, the isomorphisms $\Omega^{\text{in/out}}$ intertwine the action of the Poincaré group on $\mathcal{H}_\text{F}$ and $\mathcal{H}^{\text{in/out}}$. We summarize these results, which are vital for the physical interpretation of the underlying theory, in the following theorem.

**Theorem 1** *Consider a theory of a particle of mass $m > 0$ which satisfies the standing assumptions and the stability condition [3]. Then there exist canonical isometries $\Omega^{\text{in/out}}$, mapping the Fock space $\mathcal{H}_\text{F}$ based on the single-particle space $\mathcal{H}_1$ onto subspaces $\mathcal{H}^{\text{in/out}} \subset \mathcal{H}$ of incoming and outgoing scattering states. Moreover, these isometries intertwine the action of the Poincaré transformations on the respective spaces.*

Since the scattering states have been identified with Fock space, asymptotic creation and annihilation operators act on $\mathcal{H}^{\text{in/out}}$ in a natural manner. This point will be explained in the following section.

## LSZ Formalism

Prior to the results of Haag and Ruelle, an axiomatic approach to scattering theory was developed by Lehmann, Symanzik, and Zimmermann (LSZ), based on time-ordered vacuum expectation values of quantum fields. The relative advantage of their approach with respect to Haag–Ruelle theory is that useful reduction formulas for the $S$-matrix greatly facilitate computations, in particular in perturbation theory. Moreover, these formulas are the starting point of general studies of the momentum space analyticity properties of the $S$-matrix (dispersion relations), as outlined in Dispersion Relations (cf. also Iagolnitzer (1993)). Within the present general setting, the LSZ method was established by Hepp.

For simplicity of discussion, we consider again a single particle type of mass $m > 0$ and integer spin $s$, subject to condition [3]. According to the results of the preceding section, one then can consistently define asymptotic creation operators on the scattering states, setting

$$A(f)^{\text{in/out}}\left(P_1 A_1(f^{(1)})\Omega \otimes \cdots \otimes P_1 A_n(f^{(n)})\Omega\right)^{\text{in/out}} \doteq \lim_{t\to\mp\infty} A_t(f)\left(P_1 A_1(f^{(1)})\Omega \otimes \cdots \otimes P_1 A_n(f^{(n)})\Omega\right)^{\text{in/out}} = \left(P_1 A(f)\Omega \otimes P_1 A_1(f^{(1)})\Omega \otimes \cdots \otimes P_1 A_n(f^{(n)})\Omega\right)^{\text{in/out}} \quad [10]$$

Similarly, one obtains the corresponding asymptotic annihilation operators,

$$A(f)^{\text{in/out}*}\left(P_1 A_1(f^{(1)})\Omega \otimes \cdots \otimes P_1 A_n(f^{(n)})\Omega\right)^{\text{in/out}} = \lim_{t\to\mp\infty} A_t(f)^*\left(P_1 A_1(f^{(1)})\Omega \otimes \cdots \otimes P_1 A_n(f^{(n)})\Omega\right)^{\text{in/out}} = 0 \quad [11]$$

where the latter equality holds if the Fourier transforms of the functions $f, f^{(1)}, \ldots, f^{(n)}$, have disjoint supports. We mention as an aside that, by replacing the time-averaging function $g$ in the definition of $A_t(f)$ by a delta function, the above formulas still hold. But the convergence is then to be understood in the weak Hilbert space topology. In this form, the above relations were anticipated by LSZ (asymptotic condition).

It is straightforward to proceed from these relations to reduction formulas. Let $B$ be any local operator. Then one has, in the sense of matrix elements between outgoing and incoming scattering states,

$$BA(f)^{\text{in}} - A(f)^{\text{out}}B = \lim_{t\to\infty}(BA(f_{-t}) - A(f_t)B) = \lim_{t\to\infty}\left(\int \mathrm{d}^4x f_{-t}(x)BA(x) - \int \mathrm{d}^4x f_t(x)A(x)B\right) \quad [12]$$

$f_t(x) \doteq g_t(x_0) f(x_0)(\text{vec}(x))$. Because of the (essential) support properties of the functions $f_{\pm t}$, the contributions to the latter integrals arise, for asymptotic $t$, from spacetime points $x$ where the localization

regions of $A(x)$ and $B$ have a negative timelike (first term), respectively, positive timelike (second term) distance. One may therefore proceed from the products of these operators to the time-ordered products $T(BA(x))$, where $T(BA(x)) = A(x)B$ if the localization region of $A(x)$ lies in the future of that of $B$, and $T(BA(x)) = BA(x)$ if it lies in the past. It is noteworthy that a precise definition of the time ordering for finite $x$ is irrelevant in the present context – any reasonable interpolation between the above relations will do. Similarly, one can define time-ordered products for an arbitrary number of local operators. The preceding limit can then be recast into

$$\lim_{t\to\infty} \int d^4x (f_{-t}(x) - f_t(x)) T(BA(x)) \qquad [13]$$

The latter expression has a particularly simple form in momentum space. Proceeding to the Fourier transforms of $f_{\pm t}$ and noticing that, in the limit of large $t$,

$$\left(\widetilde{f_{-t}}(p) - \widetilde{f_t}(p)\right)/(p_0 - \omega(\boldsymbol{p})) \longrightarrow -2\pi i \widetilde{f}(\boldsymbol{p})\, \delta(p_0 - \omega(\boldsymbol{p})) \qquad [14]$$

one gets

$$\begin{aligned} BA(f)^{\text{in}} - A(f)^{\text{out}}B &= -2\pi i \int d^3p \widetilde{f}(\boldsymbol{p})(p_0 - \omega(\boldsymbol{p})) \\ &\quad \times T(B\widetilde{A}(-p))\Big|_{p_0 = \omega(\boldsymbol{p})} \end{aligned} \qquad [15]$$

Here $T(B\widetilde{A}(p))$ denotes the Fourier transform of $T(BA(x))$, and it can be shown that the restriction of $(p_0 - \omega(\boldsymbol{p}))T(B\widetilde{A}(-p))$ to the manifold $\{p \in \mathbb{R}^4 : p_0 = \omega(\boldsymbol{p})\}$ (the "mass shell") is meaningful in the sense of distributions on $\mathbb{R}^3$. By the same token, one obtains

$$\begin{aligned} &A(f)^{\text{out}*}B - BA(f)^{\text{in}*} \\ &= -2\pi i \int d^3p \overline{\widetilde{f}(\boldsymbol{p})}\, (p_0 - \omega(\boldsymbol{p})) T(\widetilde{A^*}(p)B)\Big|_{p_0 = \omega(\boldsymbol{p})} \end{aligned} \qquad [16]$$

Similar relations, involving an arbitrary number of asymptotic creation and annihilation operators, can be established by analogous considerations. Taking matrix elements of these relations in the vacuum state and recalling the action of the asymptotic creation and annihilation operators on scattering states, one arrives at the following result, which is central in all applications of scattering theory.

**Theorem 2** *Consider the theory of a particle of mass $m > 0$ subject to the conditions stated in the preceding sections and let $f^{(1)}, \ldots, f^{(n)}$ be any family of test functions whose Fourier transforms have compact and nonoverlapping supports. Then*

$$\begin{aligned} &\Big\langle \left(P_1A_1(f^{(1)})\Omega \otimes \cdots \otimes P_1A_k(f^{(k)})\Omega\right)^{\text{out}}, \\ &\qquad \left(P_1A_{k+1}(f^{(k+1)})\Omega \otimes \cdots \otimes P_1A_n(f^{(n)})\Omega\right)^{\text{in}} \Big\rangle \\ &= (-2\pi i)^n \int \cdots \int d^3p_1 \cdots d^3p_n \overline{\widetilde{f^{(1)}}(\boldsymbol{p}_1)} \cdots \\ &\quad \times \overline{\widetilde{f^{(k)}}(\boldsymbol{p}_k)} \widetilde{f^{(k+1)}}(\boldsymbol{p}_{k+1}) \cdots \widetilde{f^{(n)}}(\boldsymbol{p}_n) \\ &\quad \times \prod_{i=1}^{n} (p_{i_0} - \omega(\boldsymbol{p}_i)) \Big\langle \Omega, T\Big(\widetilde{A_1^*}(p_1) \cdots \\ &\quad \times \widetilde{A_k^*}(p_k) \widetilde{A_{k+1}}(-p_{k+1}) \cdots \\ &\quad \times \widetilde{A_n}(-p_n)\Omega\Big)\Big\rangle\Big|_{p_{j_0} = \omega(\boldsymbol{p}_j)}^{j=1,\ldots,n} \end{aligned} \qquad [17]$$

*in an obvious notation.*

Thus, the kernels of the scattering amplitudes in momentum space are obtained by restricting the (by the factor $\prod_{i=1}^{n}(p_{i_0} - \omega(\boldsymbol{p}_i))$) amputated Fourier transforms of the vacuum expectation values of the time-ordered products to the positive and negative mass shells, respectively. These are the famous LSZ reduction formulas, which provide a convenient link between the time-ordered (Green's) functions of a theory and its asymptotic particle interpretation.

## Asymptotic Particle Counters

The preceding construction of scattering states applies to a significant class of theories; but even if one restricts attention to the case of massive particles, it does not cover all situations of physical interest. For an essential input in the construction is the existence of local operators interpolating between the vacuum and the single-particle states. There may be no such operators at one's disposal, however, either because the particle in question carries a nonlocalizable charge, or because the given family of operators is too small. The latter case appears, for example, in gauge theories, where in general only the observables are fixed by the principle of local gauge invariance, and the physical particle content as well as the corresponding interpolating operators are not known from the outset. As observables create from the vacuum only neutral states, the above construction of scattering states then fails if charged particles are present. Nevertheless, thinking in physical terms, one would expect that the observables contain all relevant information in order to determine the features of scattering states, in particular their collision cross section. That this is indeed the case was first shown by Araki and Haag (Araki 1999).

In scattering experiments, the measured data are provided by detectors (e.g., particle counters) and

coincidence arrangements of detectors. Essential features of detectors are their lack of response in the vacuum state and their macroscopic localization. Hence, within the present mathematical setting, a general detector is represented by a positive operator $C$ on the physical Hilbert space $\mathcal{H}$ such that $C\Omega = 0$. Because of the Reeh–Schlieder theorem, these conditions cannot be satisfied by local operators. However, they can be fulfilled by "almost-local" operators. Examples of such operators are easy to produce, putting $C = L^*L$ with

$$L = \int d^4x\, f(x)\, A(x) \qquad [18]$$

where $A$ is any local operator and $f$ any test function whose Fourier transform has compact support in the complement of the closed forward light cone (and hence in the complement of the energy momentum spectrum of the theory). In view of the properties of $f$ and the invariance of $\Omega$ under translations, it follows that $C = L^*L$ annihilates the vacuum and can be approximated with arbitrary precision by local operators. The algebra generated by these operators $C$ will be denoted by $\mathcal{C}$.

When preparing a scattering experiment, the first thing one must do with a detector is to calibrate it, that is, test its response to sources of single-particle states. Within the mathematical setting, this amounts to computing the matrix elements of $C$ in states $\Phi \in \mathcal{H}_1$:

$$\langle \Phi, C\Phi \rangle = \int\int d^3p\, d^3q\, \overline{\Phi(\boldsymbol{p})}\, \Phi(\boldsymbol{q})\, \langle \boldsymbol{p}|C|\boldsymbol{q}\rangle \qquad [19]$$

Here $\boldsymbol{p} \mapsto \Phi(\boldsymbol{p})$ is the momentum space wave function of $\Phi$, $\langle \cdot|C|\cdot\rangle$ is the kernel of $C$ in the single-particle space $\mathcal{H}_1$, and we have omitted (summations over) indices labeling internal degrees of freedom of the particle, if any. The relevant information about $C$ is encoded in its kernel. As a matter of fact, one only needs to know its restriction to the diagonal, $\boldsymbol{p} \mapsto \langle \boldsymbol{p}|C|\boldsymbol{p}\rangle$. It is called the sensitivity function of $C$ and can be shown to be regular under quite general circumstances (Araki 1999, Buchholz and Fredenhagen 1982).

Given a state $\Psi \in \mathcal{H}$ for which the expectation value $\langle \Psi, C(x))\Psi\rangle$ differs significantly from 0, one concludes that this state deviates from the vacuum in a region about $x$. For finite $x$, this does not mean, however, that $\Psi$ has a particle interpretation at $x$. For that spacetime point may, for example, be just the location of a collision center. Yet, if one proceeds to asymptotic times, one expects, in view of the spreading of wave packets, that the probability of finding two or more particles in the same spacetime region is dominated by the single-particle contributions. It is this physical insight which justifies the expectation that the detectors $C(x)$ become particle counters at asymptotic times. Accordingly, one considers for asymptotic $t$ the operators

$$C_t(h) \doteq \int d^3x\, h(\boldsymbol{x}/t) C(t, \boldsymbol{x}) \qquad [20]$$

where $h$ is any test function on $\mathbb{R}^3$. The role of the integral is to sum up all single-particle contributions with velocities in the support of $h$ in order to compensate for the decreasing probability of finding such particles at asymptotic times $t$ about the localization center of the detector. That these ideas are consistent was demonstrated by Araki and Haag, who established the following result (Araki 1999).

**Theorem 3** *Consider, as before, the theory of a massive particle. Let $C^{(1)}, \ldots, C^{(n)} \in \mathcal{C}$ be any family of detector operators and let $h^{(1)}, \ldots, h^{(n)}$ be any family of test functions on $\mathbb{R}^3$. Then, for any state $\Psi^{\text{out}} \in \mathcal{H}^{\text{out}}$ of finite energy,*

$$\lim_{t\to\infty} \left\langle \Psi^{\text{out}}, C_t^{(1)}(h^{(1)}) \cdots C_t^{(n)}(h^{(n)}) \Psi^{\text{out}} \right\rangle$$
$$= \int \cdots \int d^3p_1 \cdots d^3p_n \langle \Psi^{\text{out}}, \rho^{\text{out}}(\boldsymbol{p}_1) \cdots \rho^{\text{out}}(\boldsymbol{p}_n)\Psi^{\text{out}}\rangle$$
$$\times \prod_{k=1}^{n} h(\boldsymbol{p}_k/\omega(\boldsymbol{p}_k))\langle \boldsymbol{p}_k|C^{(k)}|\boldsymbol{p}_k\rangle \qquad [21]$$

*where $\rho^{\text{out}}(\boldsymbol{p})$ is the momentum space density (the product of creation and annihilation operators) of outgoing particles of momentum $\boldsymbol{p}$, and (summations over) possible indices labeling internal degrees of freedom of the particle are omitted. An analogous relation holds for incoming scattering states at negative asymptotic times.*

This result shows, first of all, that the scattering states have indeed the desired interpretation with regard to the observables, as anticipated in the preceding sections. Since the assertion holds for all scattering states of finite energy, one may replace in the above theorem the outgoing scattering states by any state of finite energy, if the theory is asymptotically complete, that is, $\mathcal{H} = \mathcal{H}^{\text{in}} = \mathcal{H}^{\text{out}}$. Then choosing, in particular, any incoming scattering state and making use of the arbitrariness of the test functions $h^{(k)}$ as well as the knowledge of the sensitivity functions of the detector operators, one can compute the probability distributions of outgoing particle momenta in this state, and thereby the corresponding collision cross sections.

The question of how to construct certain specific incoming scattering states by using only local observables was not settled by Araki and Haag,

however. A general method to that effect was outlined in Buchholz *et al.* (1991). As a matter of fact, for that method only the knowledge of states in the subspace of neutral states is required. Yet in this approach one would need for the computation of, say, elastic collision cross sections of charged particles the vacuum correlation functions involving at least eight local observables. This practical disadvantage of increased computational complexity of the method is offset by the conceptual advantage of making no appeal to quantities which are *a priori* nonobservable.

## Massless Particles and Huygens' Principle

The preceding general methods of scattering theory apply only to massive particles. Yet taking advantage of the salient fact that massless particles always move with the speed of light, Buchholz succeeded in establishing a scattering theory also for such particles (Haag 1992). Moreover, his arguments lead to a quantum version of Huygens' principle.

As in the case of massive particles, one assumes that there is a subspace $\mathcal{H}_1 \subset \mathcal{H}$ corresponding to a representation of $U(\mathcal{P}_+^\uparrow)$ of mass $m=0$ and, for simplicity, integer helicity; moreover, there must exist local operators interpolating between the vacuum and the single-particle states. These assumptions cover, in particular, the important examples of the photon and of Goldstone particles. Picking any suitable local operator $A$ interpolating between $\Omega$ and some vector in $\mathcal{H}_1$, one sets, in analogy to [4],

$$A_t \doteq \int \mathrm{d}^4x\, g_t(x_0) \times (-1/2\pi)\varepsilon(x_0)\,\delta(x_0^2 - \boldsymbol{x}^2)\partial_0 A(x) \qquad [22]$$

Here $g_t(x_0) \doteq (1/|\ln t|)\, g((x_0 - t)/|\ln t|)$ with $g$ as in [4], and the solution of the Klein–Gordon equation in [4] has been replaced by the fundamental solution of the wave equation; furthermore, $\partial_0 A(x)$ denotes the derivative of $A(x)$ with respect to $x_0$. Then, once again, the strong limit of $A_t\Omega$ as $t \to \pm\infty$ is $P_1 A\Omega$, with $P_1$ the projection onto $\mathcal{H}_1$.

In order to establish the convergence of $A_t$ as in the LSZ approach, one now uses the fact that these operators are, at asymptotic times $t$, localized in the complement of some forward, respectively, backward, light cone. Because of locality, they therefore commute with all operators which are localized in the interior of the respective cones. More specifically, let $\mathcal{O} \subset \mathbb{R}^4$ be the localization region of $A$ and let $\mathcal{O}_\pm \subset \mathbb{R}^4$ be the two regions having a positive, respectively, negative, timelike distance from all points in $\mathcal{O}$. Then, for any operator $B$ which is compactly localized in $\mathcal{O}_\pm$, respectively, one obtains $\lim_{t\to\pm\infty} A_t B\Omega = \lim_{t\to\pm\infty} BA_t\Omega = BP_1A\Omega$. This relation establishes the existence of the limits

$$A^{\mathrm{in/out}} = \lim_{t\to\mp\infty} A_t \qquad [23]$$

on the (by the Reeh–Schlieder property) dense sets of vectors $\{B\Omega : B \in \mathcal{A}(\mathcal{O}_\mp)\} \subset \mathcal{H}$. It requires some more detailed analysis to prove that the limits have all of the properties of a (smeared) free massless field, whose translates $x \mapsto A^{\mathrm{in/out}}(x)$ satisfy the wave equation and have c-number commutation relations. From these free fields, one can then proceed to asymptotic creation and annihilation operators and construct asymptotic Fock spaces $\mathcal{H}^{\mathrm{in/out}} \subset \mathcal{H}$ of massless particles and a corresponding scattering matrix as in the massive case. The details of this construction can be found in the original article, cf. Haag (1992).

It also follows from these arguments that the asymptotic fields $A^{\mathrm{in/out}}$ of massless particles emanating from a region $\mathcal{O}$, that is, for which the underlying interpolating operators $A$ are localized in $\mathcal{O}$, commute with all operators localized in $\mathcal{O}_\mp$, respectively. This result may be understood as an expression of Huygens' principle. More precisely, denoting by $\mathcal{A}^{\mathrm{in/out}}(\mathcal{O})$ the algebras of bounded operators generated by the asymptotic fields $A^{\mathrm{in/out}}$, respectively, one arrives at the following quantum version of Huygens' principle.

**Theorem 4** *Consider a theory of massless particles as described above and let $\mathcal{A}^{\mathrm{in/out}}(\mathcal{O})$ be the algebras generated by massless asymptotic fields $A^{\mathrm{in/out}}$ with $A \in \mathcal{A}(\mathcal{O})$. Then*

$$\mathcal{A}^{\mathrm{in}}(\mathcal{O}) \subset \mathcal{A}(\mathcal{O}_-)' \quad \text{and} \quad \mathcal{A}^{\mathrm{out}}(\mathcal{O}) \subset \mathcal{A}(\mathcal{O}_+)' \qquad [24]$$

*Here the prime denotes the set of bounded operators commuting with all elements of the respective algebras (i.e., their commutants).*

## Beyond Wigner's Concept of Particle

There is by now ample evidence that Wigner's concept of particle is too narrow in order to cover all particle-like structures appearing in quantum field theory. Examples are the partons which show up in nonabelian gauge theories at very small spacetime scales as constituents of hadrons, but which do not appear at large scales due to the confining forces. Their mathematical description

requires a quite different treatment, which cannot be discussed here. But even at large scales, Wigner's concept does not cover all stable particle-like systems, the most prominent examples being particles carrying an abelian gauge charge, such as the electron and the proton, which are inevitably accompanied by infinite clouds of ("on-shell") massless particles.

The latter problem was discussed first by Schroer, who coined the term "infraparticle" for such systems. Later, Buchholz showed in full generality that, as a consequence of Gauss' law, pure states with an abelian gauge charge can neither have a sharp mass nor carry a unitary representation of the Lorentz group, thereby uncovering the simple origin of results found by explicit computations, notably in quantum electrodynamics (Steinmann 2000). Thus, one is faced with the question of an appropriate mathematical characterization of infraparticles which generalizes the concept of particle invented by Wigner. Some significant steps in this direction were taken by Fröhlich, Morchio, and Strocchi, who based a definition of infraparticles on a detailed spectral analysis of the energy–momentum operators. For an account of these developments and further references, cf. Haag (1992).

We outline here an approach, originated by Buchholz, which covers all stable particle-like structures appearing in quantum field theory at asymptotic times. It is based on Dirac's idea of improper particle states with sharp energy and momentum. In the standard (rigged Hilbert space) approach to giving mathematical meaning to these quantities, one regards them as vector-valued distributions, whereby one tacitly assumes that the improper states can coherently be superimposed so as to yield normalizable states. This assumption is valid in the case of Wigner particles but fails in the case of infraparticles. A more adequate method of converting the improper states into normalizable ones is based on the idea of acting on them with suitable localizing operators. In the case of quantum mechanics, one could take as a localizing operator any sufficiently rapidly decreasing function of the position operator. It would map the improper "plane-wave states" of sharp momentum into finitely localized states which thereby become normalizable. In quantum mechanics, these two approaches can be shown to be mathematically equivalent. The situation is different, however, in quantum field theory.

In quantum field theory, the appropriate localizing operators $L$ are of the form [18]. They constitute a (nonclosed) left ideal $\mathcal{L}$ in the C*-algebra $\mathcal{A}$ generated by all local operators. Improper particle states of sharp energy–momentum $p$ can then be defined as linear maps $|\cdot\rangle_p : \mathcal{L} \to \mathcal{H}$ satisfying

$$U(x)|L\rangle_p = e^{ipx}|L(x)\rangle_p, \quad L \in \mathcal{L} \qquad [25]$$

It is instructive to (formally) replace $L$ here by the identity operator, making it clear that this relation indeed defines improper states of sharp energy–momentum.

In theories of massive particles, one can always find localizing operators $L \in \mathcal{L}$ such that their images $|L\rangle_p \in \mathcal{H}$ are states with a sharp mass. This is the situation covered in Wigner's approach. In theories with long-range forces there are, in general, no such operators, however, since the process of localization inevitably leads to the production of low-energy massless particles. Yet improper states of sharp momentum still exist in this situation, thereby leading to a meaningful generalization of Wigner's particle concept.

That this characterization of particles covers all situations of physical interest can be justified in the general setting of relativistic quantum field theory as follows. Picking $g_t$ as in [4] and any vector $\Psi \in \mathcal{H}$ with finite energy, one can show that the functionals $\rho_t, t \in \mathbb{R}$, given by

$$\rho_t(L^*L) \doteq \int d^4x \, g_t(x_0) \, \langle \Psi, (L^*L)(x)\Psi \rangle, \quad L \in \mathcal{L} \qquad [26]$$

are well defined and form an equicontinuous family with respect to a certain natural locally convex topology on the algebra $\mathcal{C} = \mathcal{L}^*\mathcal{L}$. This family of functionals therefore has, as $t \to \pm\infty$, weak-$*$ limit points, denoted by $\sigma$. The functionals $\sigma$ are positive on $\mathcal{C}$ but not normalizable. (Technically speaking, they are weights on the underlying algebra $\mathcal{A}$.) Any such $\sigma$ induces a positive-semidefinite scalar product on the left ideal $\mathcal{L}$ given by

$$\langle L_1 | L_2 \rangle \doteq \sigma(L_1^* L_2), \quad L_1, L_2 \in \mathcal{L} \qquad [27]$$

After quotienting out elements of zero norm and taking the completion, one obtains a Hilbert space and a linear map $L \mapsto |L\rangle$ from $\mathcal{L}$ into that space. Moreover, the spacetime translations act on this space by a unitary representation satisfying the relativistic spectrum condition.

It is instructive to compute these functionals and maps in theories of massive particles. Making use of relation [21] one obtains, with a slight change of notation,

$$\langle L_1 | L_2 \rangle = \int d\mu(p) \, \langle p \, | L_1^* L_2 | p \, \rangle \qquad [28]$$

where $\mu$ is a measure giving the probability density of finding at asymptotic times in state $\Psi$ a particle of energy–momentum $p$. Once again, possible summations over different particle types and internal degrees of freedom have been omitted here. Thus,

setting $|L\rangle_p \doteq L\,|p\rangle$, one concludes that the map $L \mapsto |L\rangle$ can be decomposed into a direct integral of improper particle states of sharp energy–momentum, $|\cdot\rangle = \int_\oplus \mathrm{d}\mu(p)^{1/2}|\cdot\rangle_p$. It is crucial that this result can also be established without any *a priori* input about the nature of the particle content of the theory, thereby providing evidence of the universal nature of the concept of improper particle states of sharp momentum, as outlined here.

**Theorem 5** *Consider a relativistic quantum field theory satisfying the standing assumptions. Then the maps $L \mapsto |L\rangle$ defined above can be decomposed into improper particle states of sharp energy–momentum $p$,*

$$|\cdot\rangle = \int_\oplus \mathrm{d}\mu(p)^{1/2}|\cdot\rangle_p \qquad [29]$$

*where $\mu$ is some measure depending on the state $\Psi$ and the respective time limit taken.*

It is noteworthy that whenever the space of improper particle states corresponding to fixed energy–momentum $p$ is finite dimensional (finite particle multiplets), then in the corresponding Hilbert space there exists a continuous unitary representation of the little group of $p$. This implies that improper momentum eigenstates of mass $m = (p^2)^{1/2} > 0$ carry definite (half)integer spin, in accordance with Wigner's classification. However, if $m = 0$, the helicity need not be quantized, in contrast to Wigner's results.

Though a general scattering theory based on improper particle states has not yet been developed, some progress has been made in Buchholz *et al.* (1991). There it is outlined how inclusive collision cross sections of scattering states, where an undetermined number of low-energy massless particles remains unobserved, can be defined in the presence of long-range forces, in spite of the fact that a meaningful scattering matrix may not exist.

## Asymptotic Completeness

Whereas the description of the asymptotic particle features of any relativistic quantum field theory can be based on an arsenal of powerful methods, the question of when such a theory has a complete particle interpretation remains open to date. Even in concrete models there exist only partial results, cf. Iagolnitzer (1993) for a comprehensive review of the current state of the art. This situation is in striking contrast to the case of quantum mechanics, where the problem of asymptotic completeness has been completely settled.

One may trace the difficulties in quantum field theory back to the possible formation of superselection sectors (Haag 1992) and the resulting complex particle structures, which cannot appear in quantum-mechanical systems with a finite number of degrees of freedom. Thus, the first step in establishing a complete particle interpretation in a quantum field theory has to be the determination of its full particle content. Here the methods outlined in the preceding section provide a systematic tool. From the resulting data, one must then reconstruct the full physical Hilbert space of the theory comprising all superselection sectors. For theories in which only massive particles appear, such a construction has been established in Buchholz and Fredenhagen (1982), and it has been shown that the resulting Hilbert space contains all scattering states. The question of completeness can then be recast into the familiar problem of the unitarity of the scattering matrix. It is believed that phase space (nuclearity) properties of the theory are of relevance here (Haag 1992).

However, in theories with long-range forces, where a meaningful scattering matrix may not exist, this strategy is bound to fail. Nonetheless, as in most high-energy scattering experiments, only some very specific aspects of the particle interpretation are really tested – one may think of other meaningful formulations of completeness. The interpretation of most scattering experiments relies on the existence of conservation laws, such as those for energy and momentum. If a state has a complete particle interpretation, it ought to be possible to fully recover its energy, say, from its asymptotic particle content, that is, there should be no contributions to its total energy which do not manifest themselves asymptotically in the form of particles. Now the mean energy–momentum of a state $\Psi \in \mathcal{H}$ is given by $\langle \Psi, P\Psi\rangle, P$ being the energy–momentum operators, and the mean energy–momentum contained in its asymptotic particle content is $\int \mathrm{d}\mu(p)p$, where $\mu$ is the measure appearing in the decomposition [29]. Hence, in case of a complete particle interpretation, the following should hold:

$$\langle \Psi, P\Psi\rangle = \int \mathrm{d}\mu(p)p \qquad [30]$$

Similar relations should also hold for other conserved quantities which can be attributed to particles, such as charge, spin, etc. It seems that such a weak condition of asymptotic completeness suffices for a consistent interpretation of most scattering experiments. One may conjecture that relation [30] and its generalizations hold in all theories admitting a local stress–energy tensor and local currents corresponding to the charges.

*See also:* Algebraic Approach to Quantum Field Theory; Axiomatic Quantum Field Theory; Dispersion Relations; Perturbation Theory and its Techniques; Quantum Chromodynamics; Quantum Field Theory in Curved

Spacetime; Quantum Mechanical Scattering Theory; Scattering, Asymptotic Completeness and Bound States; Scattering in Relativistic Quantum Field Theory: The Analytic Program.

## Further Reading


Araki H (1999) *Mathematical Theory of Quantum Fields*. Oxford: Oxford University Press.

Buchholz D and Fredenhagen K (1982) Locality and the structure of particle states. *Communications in Mathematical Physics* 84: 1–54.

Buchholz D, Porrmann M, and Stein U (1991) Dirac versus Wigner: towards a universal particle concept in quantum field theory. *Physics Letters B* 267: 377–381.

Dybalski W (2005) Haag–Ruelle scattering theory in presence of massless particles. *Letters in Mathematical Physics* 72: 27–38.

Fredenhagen K, Gaberdiel MR, and Ruger SM (1996) Scattering states of plektons (particles with braid group statistics) in $(2+1)$ dinemsional quantum field theory. *Communications in Mathematical Physics* 175: 319–336.

Haag R (1992) *Local Quantum Physics*. Berlin: Springer.

Iagolnitzer D (1993) *Scattering in Quantum Field Theories*. Princeton, NJ: Princeton University Press.

Jost R (1965) *General Theory of Quantized Fields*. Providence, RI: American Mathematical Society.

Steinmann O (2000) *Perturbative Quantum Electrodynamics and Axiomatic Field Theory*. Berlin: Springer.

Streater RF and Wightman AS (1964) *PCT, Spin and Statistics, and All That*. Reading, MA: Benjamin/Cummings.


# Scattering in Relativistic Quantum Field Theory: The Analytic Program

**J Bros**, CEA/DSM/SPhT, CEA Saclay, Gif-sur-Yvette, France



## Introduction to the Analytic Structures of Quantum Field Theory

The importance of complex variables and of the concept of analyticity in theoretical physics finds one of its best illustrations in the analytic structure of relativistic quantum field theory (QFT). The latter have been investigated from several viewpoints in the last 50 years, according to the successive progress in QFT.

In the two main axiomatic frameworks of QFT, namely the one based on Wightman axioms (for a short presentation, *see* Dispersion Relations and also Axiomatic Quantum Field Theory) and the Haag, Kastler, and Araki theory of "local observables" (*see* Algebraic Approach to Quantum Field Theory), there are general justifications of analyticity properties for relevant "$N$-point structure functions" both in complexified spacetime variables and in complexified energy–momentum variables.

In the Wightman framework, relativistic quantum fields are operator-valued distributions $\Phi_j(x)$ on four-dimensional Minkowski spacetime that transform covariantly under a unitary representation of the Poincaré group in the Hilbert space of states. The basic quantities of QFT are (tempered) distributions on $\boldsymbol{R}^{4N}$ of the form $<\Psi, \Phi(x_1)\cdots\Phi(x_N)\Psi'>$, which depend on pairs of states $\Psi, \Psi'$, belonging to the Hilbert space of the QFT considered: they can be called $N$-point structure functions of the field $\Phi$ "in $x$-space," namely in Minkowski spacetime (here, for brevity, we assume that the system is defined in terms of a single quantum field). In parallel, it is important to consider the Fourier transform $\tilde{\Phi}(p) = \int e^{ip\cdot x}\Phi(x)\,dx$ of the field in the Minkowskian energy–momentum space ($p\cdot x \doteq p_0x_0 - \boldsymbol{p}\cdot\boldsymbol{x}$ denoting the Minkowskian scalar product). The corresponding quantities $<\Psi, \tilde{\Phi}(p_1)\cdots\tilde{\Phi}(p_N)\Psi'>$, can then be called $N$-point structure functions of the field $\Phi$ "in $p$-space," namely in energy–momentum space.

In the algebraic QFT framework, each basic local observable $B$ affiliated to a certain bounded region of spacetime $\mathcal{O}$ generates a Haag–Kastler–Araki quantum field $B(x)$ by the action of the translations of spacetime, namely $B(x) \doteq U(x)BU(x)^{-1}$. Here $U(x)$ denotes the unitary representation of the group of spacetime translations in the Hilbert space of states: $B(x)$ is affiliated to the translated region $\mathcal{O}(x) = \{y; y - x \in \mathcal{O}\}$. Then again one can consider $N$-point structure functions of the theory of the form $<\Psi, B(x_1)\cdots B(x_N)\Psi'>$ and $<\Psi, \tilde{B}(p_1)\cdots\tilde{B}(p_N)\Psi'>$.

To summarize the situation as it occurs in both cases, one can say the following:

1. A certain postulate of relativistic causality implies the analyticity of structure functions of a certain class, often called "Green functions," in the complex energy–momentum variables $k_j = p_j + iq_j$, in particular for purely imaginary energies.
2. "Stability properties" of the states $\Psi, \Psi'$ such as a "bounded energy content" of these states imply

the analyticity of the previous structure functions in the complex spacetime variables, in particular for purely imaginary times.

In both cases, analyticity is obtained as a basic property of the Fourier–Laplace transformation in several variables. Let $V^+$ denote the forward cone of the Minkowskian space ($V^+ \doteq -V^- \doteq \{x;\, x^2 \doteq x \cdot x > 0,\, x_0 > 0\}$) and let

$$\tilde{f}(p + \mathrm{i}q) = \int_{V_a^+} \mathrm{e}^{\mathrm{i}(p+iq)\cdot x}\, f(x)\mathrm{d}x \qquad [1]$$

$$g(x + \mathrm{i}y) = (2\pi)^{-4} \int_{V_P^+} \mathrm{e}^{-\mathrm{i}p(x+iy)}\tilde{g}(p)\mathrm{d}p \qquad [2]$$

be the associated reciprocal Fourier formulas, applied, respectively, to functions $f(x)$ with support contained in the translated forward cone $V_a^+ = -a + V^+$, $a \in V^+$ (or in its closure), and to functions $\tilde{g}(p)$ with support contained in the translated forward cone $V_P^+ = -P + V^+$, $P \in V^+$ of energy–momentum space (or in its closure). Then in view of the convergence properties of the previous integrals, one easily checks that $\tilde{f}(k)$ is holomorphic with possible exponential increase in the imaginary directions controlled by the bound $\mathrm{e}^{q\cdot a}$ in the tube domain $\mathcal{T}^+ = \boldsymbol{R}^4 + \mathrm{i}V^+$; similarly, $g(z)$ is holomorphic with an increase controlled by the exponential bound $\mathrm{e}^{-y\cdot P}$ in the tube domain $\mathcal{T}^- = \boldsymbol{R}^4 + \mathrm{i}V^-$.

On the one hand, for each $N$ the structure functions $<\Psi, \tilde{\Phi}(p_1)\cdots\tilde{\Phi}(p_N)\Psi'>$ (or $<\Psi, \tilde{B}(p_1)\cdots\tilde{B}(p_N)\Psi'>$) have conical support properties of the previous type in the variables $p_j$, as a consequence of the relativistic shape of the energy–momentum spectrum. In both axiomatic frameworks, in fact, one postulates that there is a state of zero energy–momentum $\Omega$, called the vacuum, and that the energy–momentum spectrum $\Sigma$, namely the joint spectrum of the generators $P_\mu$ of the Lie algebra of the group $U(x)$, is contained in the closure of $V^+$: this is the so-called spectral condition. A more refined assumption introduced for the requirements in particle physics is that $\Sigma$ contains discrete parts localized on sheets of (mass-shell) hyperboloids inside $V^+$. These support properties in $p$-space imply that the corresponding inverse Fourier transforms $<\Psi, \Phi(x_1)\cdots\Phi(x_N)\Psi'>$ are boundary values of holomorphic functions in appropriate tube domains of the complex space variables $(z_1, \ldots, z_n)$.

On the other hand, in order to exhibit structure functions with conical support properties in $x$-space, one needs to build appropriate algebraic combinations of functions $<\Psi, \Phi(x_{j_1})\cdots\Phi(x_{j_N})\Psi'>$ with permuted arguments in order to take the benefit of the causality postulate, which is always formulated in terms of the commutator of two field operators. There are two versions of this postulate. In the Wightman framework, causality is expressed by the condition of local commutativity or microcausality,

$$[\Phi(x_1), \Phi(x_2)] = 0 \text{ for } (x_1 - x_2)^2 < 0 \qquad [3]$$

In the algebraic QFT framework, causality is expressed by a similar property in terms of any field $B(x)$ generated by a local observable $B \doteq B(0)$ affiliated to a region of spacetime enclosed in a given "double cone" $\mathcal{O}_b = V_b^+ \cap (-V_b^+)$. The corresponding expression of causality is

$$[B(x_1), B(x_2)] = 0 \quad \text{for } (x_1 - x_2) \notin (V_a^+ \cup (-V_a^+)) \qquad [4]$$

for all $a$ such that $a > 2b$.

So, we see that basically, causality and spectral condition generate analyticity respectively in complexified $p$-space and $x$-space. However, the situation is more intricate, since for each $N$ there are always several holomorphic branches (two in the case $N = 2$) in the variables $(z_1, \ldots, z_n)$ and also in the variables $(k_1, \ldots, k_n)$: each of these two sets is obtained essentially by permutations of the $N$ vector variables. The important point is that these various branches can be seen to "communicate together," thanks to the existence of "coincidence regions" of their boundary values on the reals. Here again the roles played by causality and stability are symmetric (but inverted): while causality produces coincidence regions for the holomorphic functions in complex spacetime, spectral conditions produce coincidence regions for the holomorphic functions in complex energy–momentum space.

In view of a basic theorem of several complex variable analysis, called the edge-of-the-wedge theorem (see below in (4)), the two sets of communicating holomorphic branches actually define by mutual analytic continuation two holomorphic function $H_N^{\Psi,\Psi'}(k_1, \ldots, k_N)$ and $\mathcal{W}_N^{\Psi,\Psi'}(z_1, \ldots, z_N)$ in respective domains $D_N^{\Psi,\Psi'}$ and $\Delta_N^{\Psi,\Psi'}$. However, these two primitive domains are not natural holomorphy domains (a phenomenon which is particular to complex geometry in several variables). The problem of finding their holomorphy envelopes, namely the smallest domains $\hat{D}_N^{\Psi,\Psi'}$ and $\hat{\Delta}_N^{\Psi,\Psi'}$ in which any functions holomorphic in the primitive domains can be analytically continued, is the idealistic purpose of what has been called the analytic program of axiomatic QFT. So, we see that there is an analytic program in $x$-space and there is an analytic program in $p$-space. In practice, except for the case $N = 2$, where the complete answer is known, only a partial knowledge of the holomorphy envelopes has been obtained.

The analytic program in $p$-space, which is the only one to be described in the rest of this article, was often considered as physically more interesting, in view of the fact that it aims to establish analyticity properties of the scattering kernels on the complex mass shell. As a matter of fact, an important part of it concerns the derivation of the analyticity domains of dispersion relations for two-particle scattering amplitudes. This part is important from the historical viewpoint as well as from conceptual, physical, and pedagogical viewpoints (the reader may find it useful to first check the article Dispersion Relations, which illustrates how a structure function of the form $H_2^{\Psi,\Psi'}(k_1, k_2)$ can be used for that purpose with a suitable choice of the states $\Psi$ and $\Psi'$). In the general development of the analytic program (in $x$-space as well as in $p$-space), it is recommended to consider the infinite set of structure functions $H_N \doteq H_N^{\Omega,\Omega}(k_1, \ldots, k_N)$ and $\mathcal{W}_N \doteq \mathcal{W}_N^{\Omega,\Omega}(z_1, \ldots, z_N)$ where $\Omega$ is the privileged vacuum state of the theory, in view of the fact that each of these sets characterizes entirely the field theory considered.

Before shifting to the analytic program in $p$-space, we would like to mention various points of interest of the analytic program in $x$-space:

1. Various results of this program have been extensively used for proving fundamental properties of QFT, such as the PCT-invariance theorem, the spin–statistics connection, etc. A good part of these can be found in the books by Streater and Wightman (1980) and by Jost (1965).
2. The functions $H_N$ and $\mathcal{W}_N$ are holomorphic in their respective $p$-space and $x$-space "Euclidean subspaces." To make this clear, let us assume that a Lorentz frame has been chosen once for all; the linear subspace of complex spacetime (resp. energy–momentum) vectors of the form $z = (\mathrm{i}y_0, \boldsymbol{x})$ (resp. $k = (\mathrm{i}q_0, \boldsymbol{p})$) is called the "Euclidean subspace" of the corresponding complex Minkowskian space, in view of the fact that the quadratic form $z^2 \doteq z \cdot z = -(y_0^2 + \boldsymbol{x}^2)$ (resp. $k^2 \doteq k \cdot k = -(q_0^2 + \boldsymbol{p}^2)$) has a definite (negative) sign on that subspace. Then it has been established that (for each $N$) the restrictions of $H_N$ and $\mathcal{W}_N$ to the corresponding $N$-vector Euclidean subspaces are the Fourier transforms of each other. This fact participates in the foundation of the Euclidean formulation of QFT or "QFT at imaginary times"; the latter has provided many important results in QFT, in particular for the rigorous study of field models (initiated by Glimm and Jaffe in the 1970s).
3. A more recent extension of QFT called thermal QFT (TQFT), which aims to study the behavior of quantum fields in a thermal bath, can be described in terms of a modified analytic program. In the latter, the spectral condition is replaced by the so-called KMS condition, which prescribes $x$-space analyticity properties of a particular type for the structure functions $\mathcal{W}_N$: it requires analyticity together with periodicity conditions with respect to imaginary times, the period being the inverse of the temperature (*see* Thermal Quantum Field Theory). The usual analytic structure for the theories with vacuum and spectral conditions is recovered in the zero-temperature limit.
4. In more recent investigations concerning quantum fields on (holomorphic) curved spacetimes, analyticity properties of the structure functions similar to those of thermal QFT can be established. This is the case in particular with de Sitter spacetime, for which a notion of "temperature of geometrical origin" is most simply exhibited.

In this article, an account of the general analytic program of axiomatic QFT in complex energy–momentum space will be presented; it will describe some of the methods which have been used for establishing analyticity properties of the $N$-point structure functions of QFT and corresponding properties of the $(n \to n')$-particle collision processes, for all $n, n'$ such that $n \geq 2, n' \geq 2, n + n' = N$. (For a more detailed study, in particular concerning the microlocal methods, see the book by Iagolnitzer (1992)).

Concerning the important case $N = 4$, this article gives complements to the results described in the article Dispersion Relations. In fact, the program allows one to justify other important analytic structures of the four-point functions and of two-particle scattering functions. They concern

- the field-theoretical basis of analyticity in the complexified variable of angular momentum, first introduced and developed in potential theory (Regge 1959);
- the Bethe–Salpeter (BS-) type structure (based on the additional postulate of asymptotic completeness), which is a relativistic field-theoretical generalization of the Lippmann–Schwinger structure of nonrelativistic scattering theory (for Schrödinger equations with Yukawa-type potentials).

The latter allows one to introduce the concept of composite particle in the field-theoretical framework (including bound states and unstable particles or "resonances") and also the concept of "Regge particle," thanks to complex angular momentum analysis.

## Various Aspects of the General Analytic Program of QFT in Complex Energy–Momentum Space

### The *N*-Point Structure Functions of QFT

It is proved in the Wightman QFT axiomatic framework that any QFT is completely characterized by the (infinite) sequence of its "*N*-point functions" or "vacuum expectation values" (also called "Wightman functions")

$$W_N(x_1, \ldots, x_N) \doteq < \Omega, \Phi(x_1) \cdots \Phi(x_N)\Omega >$$

which are tempered distributions on $\boldsymbol{R}^{4N}$ satisfying a set of general properties that can be split up into linear and nonlinear conditions. (This is known as the Wightman reconstruction theorem).

**Linear conditions** Each individual *N*-point function satisfies three sets of linear conditions which result, respectively, from:

1. *Poincaré invariance*: typically, for every Poincaré transformation $g$ of Minkowski spacetime
$$W_N(x_1, \ldots, x_N) = W_N(gx_1, \ldots, gx_N)$$
in particular, the $W_N$ are invariant under spacetime translations and therefore defined on the quotient subspace $\boldsymbol{R}^{4(N-1)} \doteq \boldsymbol{R}^{4N}/\boldsymbol{R}^4$ of the differences $x_j - x_k$.
2. *Microcausality*: support conditions on commutator functions of the following form:
$$C^{(j,j+1)}(x_1,\ldots,x_n) \doteq W_N(x_1,\ldots,x_j,x_{j+1},\ldots,x_N) - W_N(x_1,\ldots,x_{j+1}, x_j,\ldots,x_N) = 0$$
in the region of $\boldsymbol{R}^{4N}$ defined by $(x_j - x_{j+1})^2 < 0$.
3. *Spectral condition*: support conditions on the Fourier transform $\tilde{W}_N(p_1, \ldots, p_N) = \delta(p_1 + \cdots + p_N) \times \hat{w}_N(p_1, \ldots, p_{N-1})$ of $W_N$, which assert that $\hat{w}_N(p_1, \ldots, p_{N-1}) = 0$ if either one of the following conditions is fulfilled: $p_1 + \cdots + p_j \notin \Sigma$, for $j = 1, \ldots, N-1$.

For each $N$, one can then construct a set of distributions $R_N^{(\alpha)}(x_1, \ldots, x_N)$, called "generalized retarded functions" (Araki, Ruelle, Steinmann, 1960 (see Iagolnitzer (1992, ref. [EGS])) which are appropriate linear combinations of multiple commutator functions built from $W_N$ and multiplied by products of Heaviside step-functions $\theta(x_{j,0} - x_{k,0})$ of the differences of time coordinates. Each of these distributions $R_N^{(\alpha)}(x_1, \ldots, x_N)$ has its support contained in a convex salient cone $C_\alpha$. This construction can be seen as a generalization of the decomposition [23] of the commutator $C_{\Psi,\Psi'}$ in the article Dispersion Relations. Then in view of the Laplace-transform theorem in several variables, the Fourier transform $\tilde{R}_N^{(\alpha)}(p_1, \ldots, p_N) = \delta(p_1 + \cdots + p_N) \times \tilde{r}_N^{(\alpha)}([p]_N)$ is such that $\tilde{r}_N^{(\alpha)}([p]_N)$ is the boundary value of a holomorphic function $\tilde{r}_N^{(\alpha),(c)}([k]_N)$ defined in a tube $\mathcal{T}_\alpha = \boldsymbol{R}^{4(N-1)} + \mathrm{i}\tilde{C}_\alpha$. Here $[k]_N = [p]_N + \mathrm{i}[q]_N$ belongs to a $4(N-1)$-dimensional complex linear space $M_N^{(c)}$: this is the set of complex vectors $[k]_N \doteq (k_1, \ldots, k_N)$ such that $k_1 + \cdots + k_N = 0$. $\tilde{C}_\alpha$ is the dual cone of $C_\alpha$ in the real $(4(N-1)$-dimensional) $[q]_N$-space. Geometrically, each cone $\tilde{C}_\alpha$ is defined in terms of a certain "cell" of $[q]_N$-space which is defined by prescribing consistent conditions of the form $\varepsilon_J q_J \in V^+$ with $q_J = \sum_{j\in J} q_j$ and $\varepsilon_J = \pm 1$ for all proper subsets $J$ of the set $\{1, 2, \ldots, N\}$. This is the expression of the microcausality postulate (summarized in [3] or [4]) in complex energy–momentum space. Concerning the difference between the two formulations [3] and [4], one can see that there is no geometrical difference concerning the analyticity domains, but differences for the type of increase of the structure functions in their tube domains: in the case of [3], they are bounded by powers of the energy–momenta, while in the case of [4] they may have an exponential increase governed by factors of the type $\mathrm{e}^{q\cdot a}$.

For each $N$, the linear space generated by all the distributions $\tilde{r}_N^{(\alpha)}([p]_N)$ is constrained by a set of linear relations (called Steinmann relations) which result from algebraic expressions of discontinuities of the following type, called (generalized) "absorptive parts,"

$$\tilde{r}_N^{(\alpha)}([p]_N) - \tilde{r}_N^{(\alpha')}([p]_N) = < \Omega, [\tilde{\mathrm{R}}_{J_1}^{(\alpha_1)}([p]_{(J_1)}), \tilde{\mathrm{R}}_{J_2}^{(\alpha_2)}([p]_{(J_2)})]\Omega > \quad [5]$$

for all pairs of adjacent cells $(\alpha, \alpha')_{(J_1,J_2)}$ in the following sense: $\alpha$ and $\alpha'$ only differ by changing the value of $\varepsilon_{J_1} = -\varepsilon_{J_2}, (J_1, J_2)$ denoting any given partition of the set $\{1, 2, \ldots, N\}$. In [5], the symbols $\tilde{\mathrm{R}}_{J_i}^{(\alpha_i)}$ denote generalized retarded operators of lower order and the argument $[p]_{(J)}$ stands for the set of independent 4-momenta $\{p_j; j \in J\}$. Formula [5] may be seen as an $N$-point generalization of formula [26] of Dispersion Relations for the case when the state $\Psi = \Psi'$ is replaced by $\Omega$.

Then by applying to [5] the same argument based on spectral condition as in the exploitation of eqn [26] in Dispersion Relations, one concludes that the two distributions $\tilde{r}_N^{(\alpha)}$ and $\tilde{r}_N^{(\alpha')}$ coincide on an open set $\mathcal{R}_{\alpha,\alpha'}$ of the form $p_{J_1}^2 = p_{J_2}^2 < M_{J_1}^2$, where $p_{J_1} \doteq \sum_{j\in J_1} p_j = -p_{J_2}$. It then follows from the general "oblique edge-of-the-wedge theorem" (Epstein, 1960; see below) that the two corresponding holomorphic functions $\tilde{r}_N^{(\alpha),(c)}([k]_N)$ and $\tilde{r}_N^{(\alpha'),(c)}([k]_N)$

have a common analytic continuation in the union of their tubes together with a certain complex "connecting set," bordered by $\mathcal{R}_{\alpha,\alpha'}$. Since this argument applies to all pairs $(\alpha, \alpha')_{(J_1,J_2)}$, the following important property holds (see Iagolnitzer (1992, refs. [B2], [EGS])):

**Theorem 1**

(i) *All the holomorphic functions* $\tilde{r}_N^{(\alpha),(c)}([k]_N)$ *admit a common analytic continuation* $H_N([k]_N)$, *called the N-point structure function (or Green function) of the given quantum field in complex energy–momentum space. It is holomorphic in a "primitive domain"* $D_N$ *of* $M_N^{(c)}$, *which is the union of all tubes* $\mathcal{T}_\alpha$ *together with complex "connecting sets" bordered by all the coincidence regions* $\mathcal{R}_{\alpha,\alpha'}$ *defined previously.*

(ii) *For each N the complex domain* $D_N$ *contains the whole Euclidean subspace* $\mathcal{E}_N$ *of* $M_N^{(c)}$, *which is the set of all complex vectors* $[k]_N = (k_1, \ldots, k_N)$ *such that* $k_j = (k_{j,0}, \boldsymbol{k}_j); k_{j,0} = iq_{j,0}, \boldsymbol{k}_j = \boldsymbol{p}_J$ *for* $j = 1, 2, \ldots, N$. *(This Euclidean subspace depends on the choice of a given Lorentz frame in Minkowski spacetime.)*

**Positivity Conditions** The Hilbert space framework which underlies the axioms of QFT implies (an infinite set of) positivity inequalities on the $N$-point structure functions of the fields. As a typical example related to the previous formula [5] when $|J_1| = |J_2| = N/2$ (for $N$ even), one can mention the positive-definiteness property of the absorptive parts for appropriate pairs of adjacent cells $(\alpha_1, \alpha_2 = -\alpha_1)_{(J_1,J_2)}$, which simply expresses the positivity of the following Hilbertian squared norm:

$$\left| \int \bar{f}([p]_{(J_2)}) f([p]_{(J_1)}) [\tilde{r}_N^{(\alpha)}([p]_N) - \tilde{r}_N^{(\alpha')}([p]_N)] \mathrm{d}[p]_{(J_1)} \mathrm{d}[p]_{(J_2)} \right| = \left| \int f([p]_{(J_1)}) \tilde{\mathrm{R}}_J^{(\alpha_1)}([p]_{(J_1)}) \Omega > \mathrm{d}[p]_{(J_1)} \right|^2 \geq 0 \quad [6]$$

### Scattering Kernels of General $(n \to n')$-Particle Collisions and General Reduction Formulas

The presentation of $(2 \to 2)$-particle scattering kernels in the article Dispersion Relations can be generalized to arbitrary $(n \to n')$-particle collision processes, involving $n$ incoming massive particles $(n \geq 2)$ and $n'$ outgoing massive particles $(n' \geq 2)$. The big "scattering matrix" or "$S$-matrix" in the Hilbert space of states is the collection of all partial scattering matrices $S_{n,n'}$ or of the equivalent kernels $S_{n,n'}(p_{n,\mathrm{in}}; p_{n',\mathrm{out}})$, defined by a straightforward generalization of formula [20] of the quoted article:

$$\mathrm{S}_{n,n'}(\hat{f}_{n,\mathrm{in}}, \hat{g}_{n',\mathrm{out}}) = \int_{\mathcal{M}_{n,n'}} \hat{f}_{n,\mathrm{in}}(p_{n,\mathrm{in}}) \overline{\hat{g}_{n',\mathrm{out}}}(p_{n',\mathrm{out}}) \times S_{n,n'}(p_{n,\mathrm{in}}; p_{n',\mathrm{out}}) \mu_m^n(p_{n,\mathrm{in}}) \mu_m^{n'}(p_{n',\mathrm{out}}) \quad [7]$$

Here we have considered for simplicity the case of collisions involving a single type of particle with mass $m$. In the arguments of the wave packets, the kernel, and the measures $(\mu_m^n, \mu_m^{n'})$, $p_{n,\mathrm{in}}$ and $p_{n',\mathrm{out}}$, respectively, denote the sets of incoming and outgoing 4-momenta $(p_1, \ldots, p_n)$ and $(p'_1, \ldots, p'_{n'})$ which all belong to the physical mass shell $H_m^+ = \{p; p \in V^+, p^2 = m^2\}$. By supplementing these mass-shell constraints with the relativistic law of conservation of total energy–momentum $p_1 + \cdots + p_n = p'_1 + \cdots + p'_{n'}$, one obtains the definition of the mass-shell manifold $\mathcal{M}_{n,n'}$ of $(n \to n')$-particle collision processes.

We shall reserve the name of scattering kernel (or scattering amplitude), denoted by $T_{n,n'}(p_{n,\mathrm{in}}; p_{n',\mathrm{out}})$, to the so-called "connected component" of the $S$-matrix kernel $S_{n,n'}(p_{n,\mathrm{in}}; p_{n',\mathrm{out}})$. By analogy with the definition of $T$ in terms of $S$ for the two-particle collision processes (*see* Dispersion Relations) $T_{n,n'}$ is defined by a recursive algorithm, which amounts to subtract from $S_{n,n'}$ all the components of the $(n \to n')$-collision processes that are decomposable into independent collision processes involving smaller number of particles, according to all admissible partitions of the numbers $n$ and $n'$.

For any given $N$, let us consider all the "affiliated" scattering kernels $T_{n,n'}$ such that $n + n' = N$ and whose corresponding collision processes, also called "channels," are deduced from one another by the relevant exchange of incoming particles and outgoing antiparticles (e.g., $\Pi_1 + \Pi_2 + \Pi_3 \to \Pi_4 + \Pi_5 + \Pi_6$, $\Pi_1 + \Pi_2 \to \overline{\Pi_3} + \Pi_4 + \Pi_5 + \Pi_6$, and $\Pi_1 + \Pi_3 \to \overline{\Pi_2} + \Pi_4 + \Pi_5 + \Pi_6$). There exist general reduction formulas according to which all these scattering kernels are restrictions to the mass-shell manifold $\mathcal{M}_{(N)}$ of appropriate boundary values of the (so-called) "amputated $N$-point function" $\hat{H}_N(k_1, \ldots, k_N) \doteq (k_1^2 - m^2) \cdots (k_N^2 - m^2) \times H_N(k_1, \ldots, k_N)$. More precisely, these reduction formulas can be written as follows:

$$T_{n,n'}(-p_{n,\mathrm{in}}; p_{n',\mathrm{out}})_{|\mathcal{M}_{(N)}^{(\alpha)}} = \hat{H}_N^{(\alpha)}(p_1, \ldots, p_N)_{|\mathcal{M}_{(N)}^{(\alpha)}} \quad [8]$$

In the latter, $\hat{H}_N^{(\alpha)}$ denotes a certain boundary value of $\hat{H}_N$ on the reals: it is equal to a generalized retarded function $\tilde{r}_N^{(\alpha)}([p]_N)$ which depends in a specific way on a region of the mass shell, called $\mathcal{M}_{(N)}^{(\alpha)}$, in which

the $(n \to n')$-channel is considered. The important thing to be noted in [8] is the sign convention which attributes the notation $-p_j$ to the momentum of any incoming particle and therefore implies that $p_j$ belongs to the negative sheet of hyperboloid $H_m^- \doteq -H_m^+$. This is the price to pay for expressing symmetrically the energy–momentum conservation law as $p_1 + p_2 + \cdots + p_N = 0$ (according to the QFT formalism), but it also displays, as a nice feature, the fact that all the affiliated scattering kernels $T_{n,n'}$ such that $n + n' = N$ are located on the various connected components of the mass shell $\mathcal{M}_{(N)}(p_j \in H_m;\, j = 1, 2, \ldots, N)$: the choice of the sheet $H_m^-$ or $H_m^+$ of $H_m$ is exactly linked to the incoming or outgoing character of the particle considered.

**Remark 1** The reduction formulas are more usually expressed in terms of the Fourier transforms of the (connected parts of the) $N$-point amputated chronological functions $\tau_N([p]_N)$ (*see* Scattering in Relativistic Quantum Field Theory: Fundamental Concepts and Tools). As a matter of fact, the latter coincide with the boundary values $\tilde{r}_N^{(\alpha)}([p]_N)$ of $H_N$ in the corresponding relevant regions $\mathcal{M}_{(N)}^{(\alpha)}$.

**Remark 2** Coming back to the case of two-particle scattering amplitudes (i.e., $n = n' = 2, N = 4$), one can see that the general study presented here implies the consideration of the four-point function $H_4(k_1, k_2, k_3, k_4)$, which is a holomorphic function of three independent complex 4-momenta (since $k_1 + k_2 + k_3 + k_4 = 0$). In that case, the domain $D_4$ contains 32 tubes $\mathcal{T}_\alpha$ which are specified by triplets of conditions such as $q_1 \in V^+$, $q_2 \in V^+$, $q_3 \in V^+$, or $-q_1 \in V^+$, $q_1 + q_2 \in V^+$, $q_1 + q_3 \in V^+$, and those obtained by permutations of the subscripts $(1, 2, 3, 4)$ and also by a global substitution of the cone $V^-$ to $V^+$.

**Remark 3** The logical path from the postulates of QFT to the analyticity properties of two-particle scattering amplitudes that has been followed in the article Dispersion Relations can be seen as a partial exploitation of the general analyticity properties of the four-point function: one was specially interested there in the analyticity properties of $H_4$ in a single 4-momentum $k_1 = -k_3$ (at fixed real values of $p_2 = -p_4$). The "partial reduction formula" [27] of Dispersion Relations corresponds to the restriction of eqn [8] (for $N = 4$) to the linear submanifold $(p_1 = -p_3,\, p_2 = -p_4)$. It may also be worthwhile to stress the fact that, in spite of the exponential bounds on $H_4$ implied by the postulates of algebraic QFT, it has been possible to prove that the scattering function is still bounded by a power of $s$ in its cut-plane (or crossing) domain; the dispersion relations with two subtractions are still justified in that case (Epstein, Glaser, Martin, 1969 (see Martin (1969, preprint))).

### Off-Shell Character of $D_N$: Nontriviality of the Analytic Structure of the Scattering Kernels

One can now see that for each value of $N (N \geq 4)$ the situation created by complex geometry in the space $\mathbf{C}^{4(N-1)}$ of $[k]_N$ is a mere generalization of the one described in a simple situation in the article Dispersion Relations.

1. There exists a fundamental $(3N - 4)$-dimensional complex submanifold, namely the complex mass shell $\mathcal{M}_{(N)}^{(c)}$ defined by the equations $k_j^2 = m^2$; $j = 1, \ldots, N$, which connects together the various real mass-shell components $\mathcal{M}_{(N)}^{(\alpha)}$ interpreted as the various physical regions of a set of affiliated $(n \to n')$-collision processes. The problem of proving the "analyticity of $(n \to n')$-scattering functions" thus amounts to constructing such holomorphic functions on the complex manifold $\mathcal{M}_{(N)}^{(c)}$, whose boundary values on the various real regions $\mathcal{M}_{(N)}^{(\alpha)}$ would reproduce the relevant scattering kernels $T_{n,n'}(-p_{n,\text{in}}; p_{n',\text{out}})$.
2. All the tubes $\mathcal{T}_\alpha$ which generate the primitive domain $D_N$ are off-shell domains, namely their intersections with $\mathcal{M}_{(N)}^{(c)}$ are empty. This simply comes from the fact that the conditions $q_j \in V^\pm$ (included in their definition) and $k_j^2 = m^2 > 0$ are incompatible. One can also check that adding the coincidence regions $\mathcal{R}_{\alpha,\alpha'}$ between adjacent tubes does not improve the situation. However, one can state as a relevant scope the following program.
3. *Linear program* (so-called because it only relies on the linear conditions presented in the section "$N$-point structure functions of QFT"): find parts of the holomorphy envelope of $D_N$ (possibly improved by the exploitation of the Steinmann relations) whose intersections with the complex mass shell $\mathcal{M}_{(N)}^{(c)}$ are nonempty. In the best case, show that such intersections can exist which connect two different regions $\mathcal{M}_{(N)}^{(\alpha)}$ together, which means "proving the crossing property between these two regions."
4. We shall see in the following that, except for the case $N = 4$, the results of this linear program have been rather disappointing as far as reaching the complex mass shell is concerned; however, other interesting analytic structures also coming from positivity conditions and from the additional postulate of asymptotic completeness have been investigated under the general name of

nonlinear program. The "synergy" created by the combination of these two programs remains, to a large extent, to be explored.

### Results of Analytic Completion in the "Linear Program"

We can only outline here some of the geometrical methods which allow one to compute parts of the holomorphy envelopes of the domains $D_N$. One important method, which may be used after applying suitable conformal mappings, reduces to the following basic theorem.

**The tube theorem** *The holomorphy envelope of a "tube domain" of the form* $T_B = \boldsymbol{R}^n + \mathrm{i}B$, *where B is an arbitrary domain in* $\boldsymbol{R}^n$ *called the basis of the tube, is the convex tube* $T_{\hat{B}} = \boldsymbol{R}^n + \mathrm{i}\hat{B}$, *where* $\hat{B}$ *is the convex hull of B.*

The *opposite or oblique edge-of-the-wedge theorem* (Epstein 1960 (see Streater and Wightman (1980, ch. 2, ref. 18))) is a refined local version of the tube theorem, in which the basis $B$ is of the form $B = C_1 \cup C_2$, where $C_1, C_2$ are two disjoint (opposite or nonopposite) cones with apex at the origin and where $T_B$ is replaced by a pair of "local tubes" $(T^{(\mathrm{loc})}_{C_1}, T^{(\mathrm{loc})}_{C_2})$. Here the adjective "local" means that the real parts of the variables are confined in a given open set $\mathcal{U}$ (which can be arbitrarily small). The connectedness of $T_B$ is now replaced by the consideration of any pair of functions $(f_1, f_2)$ holomorphic in these local tubes whose boundary values on their common real set $\mathcal{U}$ coincide. The result is that $f_1$ and $f_2$ admit a common analytic continuation $f$ in a local tube $T^{(\mathrm{loc})}_{C}$, where $C$ is the convex hull of $C_1 \cup C_2$. In the case of opposite cones $(C_1 = -C_2)$, $f$ is then analytic in the real set $\mathcal{U}$, while in the general oblique case $f$ is only analytic in a complex connecting set bordered by $\mathcal{U}$ (namely a set which connects $T^{(\mathrm{loc})}_{C_1}$ and $T^{(\mathrm{loc})}_{C_2}$). There exists an extended version of the edge-of-the-wedge theorem in which the boundary values of $f_1$ and $f_2$ are only defined as distributions.

For simplicity, we shall just give a very rough classification of the type of results obtained. We shall distinguish:

- analyticity domains in the space of several (possibly all) variables: they can be of global type or of microlocal type, namely restricted to complex neighborhoods of real points;
- analyticity domains in special families of one-dimensional complex manifolds; and
- combinations of one-dimensional results which generate domains in several variables by a refined use of the tube theorem, called the *Malgrange–Zerner "flat tube theorem,"* or "flat edge-of-the-wedge theorem." In the latter, the local tubes $T^{(\mathrm{loc})}_{C_1}$ and $T^{(\mathrm{loc})}_{C_2}$ of $f_1$ and $f_2$ reduce to one-variable domains of the upper half-plane in separate variables $z_1 = x_1 + \mathrm{i}y_1, z_2 = x_2 + \mathrm{i}y_2$ but with a common range of real parts $(x_1, x_2) \in \mathcal{U}$. The data $f_1(z_1, x_2)$ and $f_2(x_1, z_2)$ have coinciding boundary values $(f_1(x_1, x_2) = f_2(x_1, x_2))$ in the limit $(y_1 \to 0, y_2 \to 0)$. The result is again the existence of a common analytic continuation to $f_1$ and $f_2$, which is a function of two complex variables $f(z_1, z_2)$ in the intersection of the quadrant $(y_1 > 0,\ y_2 > 0)$ with a complex neighborhood of $\mathcal{U}$. (Note that this result of complex analysis still holds when the real boundary values of the holomorphic functions have singularities, namely are only defined in the sense of distributions).

**Global analyticity properties** The following property (discovered by Streater for three-point functions) looks like an extension of the tube theorem. The holomorphy envelope of the union of two tubes $\mathcal{T}_\alpha, \mathcal{T}_{\alpha'}$ corresponding to adjacent pairs of cells $(\alpha, \alpha')_{(J_1, J_2)}$ together with a complex connecting set bordered by $\mathcal{R}_{\alpha,\alpha'} = \{[p]_N; p^2_{J_1} < m^2_{J_1}\}$ is the convex hull $\mathcal{T}_{\alpha,\alpha'}$ of the union of these tubes minus the following analytic hypersurface $\sigma_{J_1}$ which can be called "a cut": $\sigma_{J_1} = \{[k]_N\colon k^2_{J_1} = m^2_{J_1} + \rho, \rho \geq 0\}$. The interest of this result (although it remains by itself an off-shell result) is that it can generate larger cut-domains by additional analytic completions, which may have intersections with the complex mass shell (see below for the case $N = 4$).

**Microlocal analyticity properties** In the case of the four-point function $\hat{H}_4$, it is possible to consider opposite cut-domains of the previous type, for which $\sigma_{J_1} = \sigma_{\{1,2\}}$ is the energy-cut of the channel $(1, 2 \to 3, 4)$, and for which the spectral conditions prescribe an "edge-of-the-wedge situation" in the neighborhood of the corresponding mass-shell component $\mathcal{M}_{(1,2\to 3,4)}$. The result is that $H_4$ is proved to be holomorphic in a full complex cut-neighborhood of $\mathcal{M}_{(1,2\to 3,4)}$ in the ambient complex energy–momentum space. The intersection of this local domain with the complex mass shell $\mathcal{M}^{(c)}_{(4)}$ is of course a full complex cut-neighborhood of $\mathcal{M}_{(1,2\to 3,4)}$ in $\mathcal{M}^{(c)}_{(4)}$, and this proves that the corresponding scattering amplitude is the boundary value of an analytic scattering function defined as the restriction $\hat{F}(s,t) \doteq \hat{H}_{4|\mathcal{M}^{(c)}_{(4)}}$ of $\hat{H}_4$: it is holomorphic in a domain of complex $(s,t)$ space deprived from the $s$–cut.

In the general case $N > 4$, the results are less spectacular, although a more sophisticated microlocal method involving a "generalized edge-of-the-wedge

theorem" has been applied. This method, which was one of the three methods at the origin of the chapter of mathematics called microlocal analysis (the other two being Hörmander's "analytic wave-front" method and Sato's "microfunctions" method) is based on a local version of the Fourier–Laplace transformation called the FBI transformation (see, e.g., the book on "hypo-analytic structures" by Treves (1992) and in the present context the article "Causality and local analyticity" by Bros and Iagolnitzer (1973) (see Iagolnitzer (1992, ref. [BI1]))).

A first positive result (obtained at first by Hepp in 1965) is the fact that the various real boundary values of $\hat{H}_N$ admit well-defined restrictions as tempered distributions on the corresponding (real) mass shell $\mathcal{M}_{(N)}$; this result is in fact crucial for the rigorous proof of general reduction formulas. However, (according to Bros, Epstein, Glaser, 1972 (see Iagolnitzer (1992, ref. [BEG2])) the local existence of an analytic scattering function in $\mathcal{M}^{(c)}_{(N)}$ is not ensured at all points of the mass shell, but only in certain regions. A rather favourable situation still occurs for $(2 \to 3)$-particle collision amplitudes (i.e., for $N = 5$), but in the general case there are large regions of the mass shell where it is only possible to prove (at least in this linear program) that the amplitude is a sum of a limited number of boundary values of analytic functions, defined in local domains of $\mathcal{M}^{(c)}_{(N)}$ (see in this connection, Iagolnitzer (1992)).

**Analyticity at fixed total energy in momentum transfer variables** A remarkably simple situation had already been exploited before the general analysis of $H_N$ leading to Theorem 1 was carried out. It is the section of the domain of the $N$-point function in the space of the "initial relative 4-momentum" $k = (k_1 - k_2)/2$ of the $s$-channel with initial 4-momenta $(k_1, k_2)$, when the total energy–momentum $P = -(k_1 + k_2)$ with $P^2 = s$ is kept fixed and real. The remaining 4-momenta $p_3, \ldots, p_N$ such that $p_3 + \cdots + p_N = P$ are also kept fixed and real. Consider the case when $P$ is (positive) timelike and such that $s \geq 4m^2$. Then it can be seen that one obtains analyticity of (a certain "1-vector restriction" of) $H_N$ with respect to the vector variable $k$ in the union of the two opposite tubes $\mathcal{T}^+ = \boldsymbol{R}^4 + \mathrm{i}V^+$, $\mathcal{T}^- = \boldsymbol{R}^4 + \mathrm{i}V^-$. Moreover, an edge-of-the-wedge situation holds in view of the spectral coincidence region of the form $k_1^2 = (-P/2 + k)^2 < M_1^2$, $k_2^2 = (-P/2 - k)^2 < M_2^2$. The corresponding holomorphy envelope is given by a Jost–Lehmann–Dyson domain (*see* Dispersion Relations), whose section by the complex mass shell $k_1^2 = k_2^2 = m^2$ turns out to give a "spherical tube domain" of the form $\{k; k = p + iq; k.P = 0, k^2 = -s/4 + m^2; |q^2| < b^2\}$. The $(2 \to N-2)$-particle scattering kernel is therefore the boundary value of a scattering function holomorphic in the previous spherical domain of complex $k$-space. In the special case of the two-particle scattering amplitude $F(s,t)$, one checks that the previous domain yields for each $s, s \geq 4m^2$, an ellipse of analyticity for $\hat{F}(s,t)$ in the $t$-plane with foci at $t = 0$ and $u = 4m^2 - s - t = 0$; this ellipse is called the Lehmann ellipse. (We have considered for simplicity the case of a single type of particle with mass $m$ and two-particle threshold at $2m$.) In fact, the squared momentum transfer $t$ is equal to $(k - k')^2$, if $k' = (k_3 - k_4)/2$ denotes the "final relative momentum" of the $s$-channel, which was here taken to be fixed and real. Moreover, by a similar argument the corresponding absorptive part, namely the discontinuity across the $s$-cut of the scattering amplitude, can be shown to be holomorphic in a larger ellipse with the same foci called the large Lehmann ellipse.

It is interesting to compare the previous result with the one that one obtains when the fixed vector $P$ is chosen to be spacelike, namely when $s$ has a negative, namely "unphysical" value with respect to the distinguished channel $(1, 2 \to 3, 4)$. For that case, the exploitation of the primitive domain $D_4$ shows that for all negative (unphysical) values $\zeta_i = k_i^2 < 0$; $i = 1, 2, 3, 4$, of the squared mass variables, the function $\hat{H}_4$ is holomorphic in a cut-plane of the variable $t$, where the cuts are the $t$-cut ($t = 4m^2 + \rho$, $\rho \geq 0$) and the $u$-cut ($u = 4m^2 - s - t = 4m^2 + \rho'$, $\rho' \geq 0$). This cut-plane has of course to be compared with the off- shell cut-plane domain $\Delta_\zeta$ at the basis of the proof of dispersion relations (*see* Dispersion Relations). Here, however, the choice of the squared momentum transfer $t$ as the variable of analyticity allows one to shift to another interpretation in terms of the concept of angular momentum.

**Analyticity in the complex angular momentum variable** In all the situations previously considered for the case $N = 4$, one can see that at fixed real values of the squared energy $s$ and of the squared masses $\zeta = \{\zeta_i;\ i = 1, 2, 3, 4\}$, the complex initial and final relative 4-momenta $k$ and $k'$ have directions which vary on the complexified sphere $S^{(c)}$. Moreover, the corresponding restriction of $\hat{H}_4$ to that sphere turns out to be always well defined and analytic on the real part of that sphere: it therefore defines a kernel on the sphere, which, in view of Poincaré invariance, is invariant under the rotations and therefore admits a convergent expansion in Legendre polynomials. Let us call $h_\ell(s;\zeta)$ the corresponding sequence of Legendre coefficients.

In the first case considered above, this sequence coincides (all $\zeta_i$ being equal to $m^2$) with what the physicists call the set of partial waves $f_\ell(s)$ of the scattering amplitude. The analyticity of $\hat{H}_4$ on a complex spherical tube of $S^{(c)}$, namely of $\hat{F}(s,t)$ in the Lehmann ellipse, is then equivalent to a certain exponential decrease property with respect to $\ell$ of the sequence of partial waves.

In the second case, where $s$ and the $\zeta_i$ are negative, it can be seen that the sphere $S$ describes 4-momentum configurations which all belong to a certain Euclidean subspace $\mathcal{E}_4$ of $M_4^{(c)}$. But this situation is much more favourable from the viewpoint of analyticity, since $\hat{H}_4$ can be seen to be holomorphic on the full complex submanifold $S^{(c)} \times S^{(c)}$ minus two sets $\sigma_t$ and $\sigma_u$ which correspond to the $t$- and $u$-cuts of the complex $t$-plane. Then this larger analyticity property turns out to be equivalent to the fact that the sequence $h_\ell(s;\zeta)$ admits an interpolation $\tilde{H}(\lambda;s;\zeta)$ holomorphic in a certain half-plane of the form $\mathrm{Re}\,\lambda > \ell_0$ such that for all integers $\ell > \ell_0$ one has: $\tilde{H}(\ell;s;\zeta) = h_\ell(s;\zeta)$. The value of $\ell_0$ is linked to the power bound at large momenta that must be satisfied by $\hat{H}_4$ as a consequence of the temperateness property included in the Wightman axiomatic framework (Bros and Viano 2000).

Of course, this nice analytic structure in a complex angular momentum variable could extend to the set of physical partial waves $f_\ell(s)$ if one could establish the analytic continuation of $\hat{F}(s,t)$ in a cut-plane of $t$ containing the Lehmann ellipses, but this seems out of the possibilities at least of the linear program.

## The "Nonlinear Program" and Its Two Main Aspects

**The extension of the analyticity domains by positivity and the derivation of bounds by unitarity** Positivity conditions of the form [6] have been extensively applied to the case $N = 4$ (namely for subsets $J$ with two elements). The main result (Martin 1969) consists in the possibility of differentiating the forward dispersion relations with respect to $t$ and, as a consequence, to enlarge the analyticity domain in $t$ at fixed $s$: the Lehmann ellipse, whose size shrinks to zero when $s$ tends to infinity, can then be replaced by an ellipse (i.e., the Martin ellipse) whose maximal point $t = t_{\max} > 0$ is fixed when $s$ goes to infinity. This justifies the extension of dispersion relations in $s$ to positive values of $t$; then in a second step the use of unitarity relations for the partial waves allows one to obtain Froissart-type bounds on the scattering amplitudes (see Martin (1969)).

**Asymptotic completeness and BS-type structural analysis** The BS equations have been at first introduced as identities of formal series in the perturbative approach of QFT, and the idea of considering such identities as exact equations having a conceptual content in the general axiomatic framework of QFT has been introduced and developed by Symanzik in 1960. However, it took a long time before its integration in the analytic program of QFT (Bros 1970 (see Iagolnitzer (1992, ref. [B1]))). These developments belong to the nonlinear program since they rely on quadratic integral equations between the various $N$-point functions, which express the postulate of asymptotic completeness via the use of appropriate reduction formulas.

For brevity, the general set of BS-type equations for the $N$-point functions with $N > 4$ will not be presented. The simplest BS-type equation, which concerns the four-point function, can be written as follows:

$$\hat{H}_4(K;k,k') = B(K;k,k') + (\hat{H}_4 \circ_s B)(K;k,k') \quad [9]$$

where

$$(\hat{H}_4 \circ_s B)(K;k,k') = \int_\Gamma \hat{H}_4(K;k,k'')B(K;k'',k')G\left(\frac{K}{2}+k''\right) \times G\left(\frac{K}{2}-k''\right) \mathrm{d}_4k'' \quad [10]$$

In the latter, the $s$-channel is privileged, with $s = K^2, K = -(k_1 + k_2)$; $\hat{H}_4$ is seen as a $K$-dependent kernel ($k$ and $k'$ are the initial and final relative 4-momenta already defined), and the new object $B$ to be studied is also a $K$-dependent kernel. The function $G(k)$ is holomorphic in $k^2$ in a cut-plane except for a pole at $k^2 = m^2$ which plays a crucial role. (It is essentially the "propagator" or two-point function of the field theory considered). Apart from pathologies due to the Fredholm alternative, the correspondence between $\hat{H}_4$ and $B$ is one-to-one, but the peculiarity concerns the integration cycle $\Gamma$ of [10]: it is a complex cycle of real dimension 4, which coincides with the Euclidean space of the vector variable $k''$ when all the 4-momenta are Euclidean, and can always be distorted inside the analyticity domain of $\hat{H}_4$ together with the external variables. The exploitation of the Fredholm equation in complex space with "floating integration cycles" then implies that $B$ is holomorphic at least in the primitive domain of $\hat{H}_4$.

An important geometrical aspect of the integration on the cycle $\Gamma$ in [10] is the fact that this cycle is "pinched" between the pair of poles of the functions $G$ when $K^2$ tends to its threshold value ($s = 4m^2$).

The type of mathematical concept encountered here is closely related to those used in the study of analyticity properties and Landau singularities of the Feynman amplitudes in the perturbative approach of QFT (in this connection, see the books by Hwa and Teplitz (1966) and by F Pham (2005) and references therein).

The first basic result is that it is equivalent for $\hat{H}_4$ to satisfy an asymptotic completeness equation in the pure two-particle region $4m^2 < s < 9m^2$ and for $B$ to satisfy the following property called two-particle irreducibility: $B$ satisfies dispersion relations in $s$ such that the $s$-cut begins at the three-particle threshold: $s = 9m^2$.

The consequence of this extended analyticity property of $B$ is that it generates the following type of analyticity properties for $\hat{H}_4$:

1. *The existence of a two-sheeted analytic structure* for $\hat{H}_4$ over a domain of the $s$-plane containing the interval $4m^2 \leq s < 9m^2$, with a square-root-type branch point at the threshold $s = 4m^2$.
2. *Composite particles*. There exists a Fredholm-type expression

$$\hat{H}_4(K; k, k') = \frac{N(K; k, k')}{D(K^2)} \qquad [11]$$

where $N$ and $D$ are expressed in terms of $B$ via Fredholm determinants, which shows that in its second sheet $\hat{H}_4$ may have poles in $s = K^2$, generated by the zeros of $D$. These poles are interpreted as resonances or unstable particles. The generation of real poles in the first sheet (i.e., bound states) is also possible under special spectral assumptions of QFT.
3. *Complex angular momentum diagonalization of BS-type equations* (Bros and Viano 2000, 2003). The operation $\circ_s$ in the BS-type equation [9] contains not only an integration over squared-mass variables, but also a convolution product on the sphere $S$; the latter is transformed into a product by the Legendre expansion of four-point functions described previously in the subsection "Analyticity in the complex angular momentum variable." As a result, there is a partially diagonalized transform of eqn [9] in terms of the functions $\tilde{H}(\lambda;\, s; \zeta)$ and $\tilde{B}(\lambda;\, s; \zeta)$, which allows one to write a Fredholm formula similar to [11], namely

$$\tilde{H}(\lambda;\; s; \zeta) = \frac{\tilde{N}(\lambda;\; s; \zeta)}{\tilde{D}(\lambda; s)} \qquad [12]$$

Then under suitable increase assumptions on $B$, there may exist a half-plane of the form $\mathrm{Re}\,\lambda > \ell_1$ (with $\ell_1 < \ell_0$) such that $\tilde{H}(\lambda;\, s; \zeta)$ admits poles in the joint variables $\lambda$ and $s$, corresponding to the concept of Regge particle: the composite particles introduced in (2) might then be integrated in the Regge particle, although they manifest themselves physically only for integral values $\ell$ of $\lambda$ with the corresponding spin interpretation. Of course, this scenario is by no means proven to hold in the general analytic program of QFT, but we have seen that the relevant "embryonary structures" are conceptually built-in, so that the phenomenon might hopefully be produced in a definite quantum field model.
4. *Byproducts of BS-type structural analysis for* $N = 5$ *and* $N = 6$. Relativistic exact structural equations for $(3 \to 3)$-particle collision amplitudes, which generalize the Faddeev structural equations of nonrelativistic potential theory, have been shown to be valid in the energy region of "elastic" collisions (i.e., with total energy bounded by $4m$); relevant Landau singularities of tree diagrams and triangular diagrams have been exhibited as a by-product in this low-energy region (Bros, and also Combescure, Dunlop in two-dimensional field models, 1981 (see Iagolnitzer (1992, refs. [B3], [B4], [CD]))). Moreover, crossing domains on the complex mass shell for $(2 \to 3)$-particle collision amplitudes have been obtained (Bros 1986 (see Iagolnitzer (1992, ref. [B1]))) by conjointly using $(N = 5)$ BS-type equations together with analytic completion properties (*see*, e.g., the "Crossing lemma" in Dispersion Relations).

*See also:* Algebraic Approach to Quantum Field Theory; Axiomatic Quantum Field Theory; Dispersion Relations; Scattering, Asymptotic Completeness and Bound States; Scattering in Relativistic Quantum Field Theory: Fundamental Concepts and Tools; Thermal Quantum Field Theory.

## Further Reading


Bros J and Viano GA (2000) Complex angular momentum in general quantum field theory. *Annales Henri Poincaré* 1: 101–172.

Bros J and Viano GA (2003) Complex angular momentum diagonalization of the Bethe–Salpeter structure in general quantum field theory. *Annales Henri Poincaré* 4: 85–126.

Haag R (1992) *Local Quantum Physics*. Berlin: Springer.

Hwa RC and Teplitz VL (1966) *Homology and Feynman integrals*. New York: Benjamin.

Iagolnitzer D (1992) *Scattering in Quantum Field Theories: The Axiomatic and Constructive Approaches*, Princeton Series in Physics. Princeton: Princeton University Press.

Jost R (1965) *The General Theory of Quantized Fields*. Providence: American Mathematical Society.

Martin A (1969) *Scattering Theory: Unitarity, Analyticity and Crossing*, Lecture Notes in Physics. Berlin: Springer.
Pham F (2005) *Intégrales singulières*. Paris: EDP Sciences/CNRS Éditions.
Streater RF and Wightman AS (1964, 1980) *PCT, Spin and Statistics, and all that*. Princeton: Princeton University Press.
Treves F (1992) *Hypoanalytic Structures*. Princeton: Princeton University Press.

# Scattering, Asymptotic Completeness and Bound States

**D Iagolnitzer**, CEA/DSM/SPhT, CEA/Saclay, Gif-sur-Yvette, France
**J Magnen**, Ecole Polytechnique, France



## Introduction

Relativistic quantum field theory (QFT) has been mainly developed since the 1950s in the perturbative framework. Quantities of interest then appear as infinite sums of Feynman integrals, corresponding to infinite series expansions with respect to couplings. This approach has led to basic successes for practical purposes, but suffered due to crucial defects from conceptual and mathematical viewpoints. First, individual terms were *a priori* infinite: this was solved by perturbative renormalization. However, even so, the series remain divergent. Two rigorous approaches have been developed since the 1960s. The axiomatic approach aims to establish a general framework independent of any particular model (Lagrangian interaction) and to analyze general properties that can be derived in that framework from basic principles. The "constructive" approach aims to rigorously establish the existence of nontrivial QFT models (theories) and to directly analyze their properties. Some of the fundamental bases are described in this encyclopedia in the articles by J Bros, D Buchholz and J Summers, and by G Gallavotti, respectively. This article aims to a deeper study of particle analysis and scattering of theories. In contrast to the articles by Buchholz and Summers and G Gallavotti, it is restricted to massive theories, a rather strong restriction, but for the latter goes much beyond in particle analysis.

From a purely physical viewpoint, results remain limited: the models rigorously defined so far are weakly coupled models in spacetime dimensions 2 or 3, results on bound states depend on specific kinematical factors in these dimensions, proofs of asymptotic completeness (AC) are not yet complete, .... On the positive side, we might say that the analysis and results are of interest from both conceptual and physical viewpoints; on the other hand, these works have also largely been related and have contributed to important, purely mathematical developments, for example, in the domain of analytic functions of several complex variables, microlocal analysis, ....

The general framework of QFT based on Wightman axioms is introduced in the next section. Massive theories are characterized in that framework by a condition on the mass spectrum. Haag–Ruelle asymptotic theory then allows one to define, in the Hilbert space $\mathcal{H}$ of states, two subspaces $\mathcal{H}_{\text{in}}$ and $\mathcal{H}_{\text{out}}$ corresponding to states that are asymptotically tangent, before and after interactions, respectively, to free-particle states. The AC condition $\mathcal{H} = \mathcal{H}_{\text{in}} = \mathcal{H}_{\text{out}}$ introduces a further important implicit particle content in the theory. Collision amplitudes or scattering functions are then well defined in the space of on-mass-shell initial and final energy–momenta (satisfying energy–momentum conservation). The LSZ "reduction formulas" give their link with chronological functions of the fields.

Basic properties of scattering amplitudes that follow from the Wightman axioms are then outlined. In particular, these axioms allow one to define the "$N$-point functions," which are analytic in a domain of complex energy–momentum space containing the Euclidean region (imaginary energy components), and from which chronological and scattering functions can be recovered. Other results at that stage include the on-shell physical sheet analyticity properties of four-point functions, as also general asymptotic causality and local analyticity properties for $N \geq 4$.

Next, we describe results derived from AC and regularity conditions on analyticity and asymptotic causality in terms of particles. In particular, the analysis of the links between analyticity properties of irreducible kernels (satisfying Bethe–Salpeter type equations) and AC in low-energy regions are included, following ideas of K Symanzik.

The final three sections are devoted to the analysis of models.

Models of QFT have been rigorously defined in Euclidean spacetime, through cluster and, more generally, phase-space expansions which are shown

to be convergent at small coupling (and replace the nonconvergent expansions, of perturbative QFT). Examples of such models are the super-renormalizable massive $\varphi^4$ models in dimensions 2 or 3 (in the 1970s) and the "just renormalizable" massive (fermionic) Gross–Neveu model – in dimension 2 – in the 1980s. The $N$-point functions of these models can be shown to have exponential fall-off in Euclidean spacetime. By the usual Fourier–Laplace transform theorem, one obtains in turn analyticity properties in corresponding regions away from the Euclidean energy–momentum space.

On the other hand, *à la* Osterwalder–Schrader properties can be established in Euclidean spacetime. By analytic continuation from imaginary to real times, it is in turn shown that a corresponding nontrivial theory satisfying the Wightman axioms is recovered on the Minkowskian side. This analysis is omitted here. However, no information is obtained in that way on the mass spectrum, AC, energy–momentum space analyticity, .... Such results can be obtained through the use of irreducible kernels. This was initiated by T Spencer in the 1970s and then developed along the same line (Spencer and Zirilli, Dimock and Eckmann, Koch, Combescure, and Dunlop). We outline here the more general approach of the present authors. In the latter, irreducible kernels are directly defined through "higher-order" cluster expansions which are again convergent at sufficiently small coupling. They are shown to satisfy exponential fall-off in Euclidean spacetime with rates better than those of the $N$-point functions, and hence corresponding analyticity in larger regions around (and away from) the Euclidean energy–momentum space. Results will then be established by analytic continuation, from the Euclidean up to the Minkowskian energy–momentum space, of structure equations that express the $N$-point functions in terms of irreducible kernels. These structure equations are infinite series expansions, with again convergence properties at small coupling. In the cases $N=2$ and $N=4$ (even theories), the re-summation of these structure equations give, respectively, the Lippmann–Schwinger and Bethe–Salpeter (BS) integral equations (up to some regularization).

The one-particle irreducible (1PI) two-point kernel $G_1$ is analytic up to $s=(2m)^2-\varepsilon$, where $\varepsilon$ is small at small coupling ($s$ is the squared center of mass energy of the channel). A simple argument then allows one to show analyticity of the actual two-point function in the same region up to a pole at $k^2=m_{\text{ph}}^2$: this shows the existence of a first basic physical mass $m_{\text{ph}}$ (close at small coupling to the bare mass $m$). In a free theory (zero coupling) with one mass $m$, there is only one corresponding particle. At small coupling, the existence of other (stable) particles is not *a priori* expected; nevertheless, we will see that such particles (two-particle bound states) will occur in some models in view of kinematical threshold effects.

The 2PI four-point kernel $G_2$ is shown to be analytic up to $s=(4m)^2-\varepsilon$ in an even theory. On the other hand, it satisfies a (regularized) BS equation. In a way analogous to the section "AC and analyticity," starting here from the analyticity of $G_2$, the actual four-point function $F$ is in turn analytic or meromorphic in that region up to the cut at $s\geq 4m^2$, and the discontinuity formula associated with AC in the low-energy region is obtained.

For some models (depending on the signs of some couplings), it will be shown that $F$ has a pole in the physical sheet, below the two-particle threshold (at a distance from it which tends to zero as the coupling itself tends to zero). This pole then corresponds to a further stable particle.

More generally, and up to some technical problems, the structure equations should allow one to derive various discontinuity formulas of $N$-point functions including those associated with AC in increasingly higher-energy regions. Asymptotic causality in terms of particles and related analyticity properties (Landau singularities...) should also follow. However, in this approach, results should be obtained only for very small couplings as the energy region considered increases.

**Note**: Notations used are different in the next two sections on the one hand, and the final three sections on the other. These notations follow the use of, respectively, axiomatic and constructive field theory; for instance, $x$ and $p$ are real on the Minkowskian side in the next two sections whereas they are real on the Euclidean side in the last three sections. The mass $m$ in the next two sections is a physical mass, whereas it is a bare mass in the last three sections (where a physical mass is noted $m_{\text{ph}}$).

## The General Framework of Massive Field Theories

We denote by $x=(x_0,\boldsymbol{x})$ a (real) point in Minkowski spacetime with respective time and space components $x_0$ and $\boldsymbol{x}$ (in a given Lorentz frame); $x^2=x_0^2-\boldsymbol{x}^2$. Besides the usual spacetime dimension $d=4$, possible values 2 or 3 will also be considered. In all that follows, the unit system is such that the velocity $c$ of light is equal to 1. Energy–momentum variables, dual (by Fourier transformation) to time and space

variables, respectively, are denoted by $p=(p_0,\boldsymbol{p})$; $p^2=p_0^2-\boldsymbol{p}^2$.

We describe below the Wightman axiomatic framework, though alternative ones such as "local quantum physics" based on the Araki–Haag–Kastler axioms may be used similarly for present purposes. For simplicity, unless otherwise stated, we consider a theory with only one basic (neutral, scalar) field $A$; A is defined on spacetime as an operator-valued distribution: for each test function $f, A(f)$ (formally $\int A(x)f(x)\mathrm{d}x$) is an operator in a Hilbert space $\mathcal{H}$ of states. A physical state is represented by a (normalized) vector in $\mathcal{H}$ modulo scalar multiples. It has to be physically understood as "sub specie aeternitatis" (i.e., "with all its evolution," the Heisenberg picture of quantum mechanics being always adopted). It is assumed that there exists in $\mathcal{H}$ a representation of the Poincaré group (semidirect product of pure Lorentz transformations and spacetime translations).

The Wightman axioms include:

1. local commutativity: $A(x)$ and $A(y)$ commute if $x-y$ is spacelike: $(x-y)^2<0$.
2. the spectral condition ($=$ positivity of the energy in relativistic form): the spectrum of the energy–momentum operators (infinitesimal generators of spacetime translations) is contained in the cone $V_+(p^2\geq 0, p_0\geq 0)$. In a massive theory, the spectrum is more precisely assumed to be contained in the union of the origin (that will correspond to the vacuum vector introduced next), of one or more discrete mass-shell hyperboloids $H_+(m_i)(p^2=m_i^2, p_0>0)$ with strictly positive masses $m_i$, and of a continuum. For simplicity, and unless otherwise stated, we consider in this section a theory with only one mass $m$ and a continuum starting at $2m$ (but this will not be so in a theory with "two-particle bound states"). This condition introduces a first (partial) particle content of the theory. In models, physical masses will not be introduced at the outset but will have to be determined.
3. existence in $\mathcal{H}$ of a vacuum vector $\Omega$, which is the only invariant vector under Poincaré transformations up to scalar multiples; it is moreover assumed that the vector space generated by the action of field operators on the vacuum is dense in $\mathcal{H}$.
4. Poincaré covariance of the theory.

Subspaces $\mathcal{H}_{\text{in}}$ and $\mathcal{H}_{\text{out}}$ of $\mathcal{H}$ can be defined by limiting procedures. To that purpose, one considers test functions $f_{j,t}(x)$ with Fourier transforms of the form $\tilde{f}_j(p)\mathrm{e}^{\mathrm{i}(p_0-[\boldsymbol{p}^2+m^2]^{1/2})t}$, where the functions $\tilde{f}_j$ have their supports in a neighborhood of the mass-shell $H_+(m)$. It can then be shown that vectors of the form $\Psi_t=A(f_{1,t})A(f_{2,t})\cdots A(f_{n,t})\Omega$ converge to limits in $\mathcal{H}$ when $t\to\pm\infty$, respectively, and that these limits depend only on the mass-shell restrictions of the test functions $\tilde{f}_{j|H_+(m)}$.

$\mathcal{H}_{\text{in}}$ and $\mathcal{H}_{\text{out}}$ are interpreted physically as subspaces of states that are "asymptotically tangent" before, respectively, after the interactions, to free-particle states with particles of mass $m$. They are in fact both isomorphic to the free-particle Fock space $\mathcal{F}$, namely the direct sum of $n$-particle spaces of "wave functions" depending on $n$ on mass-shell energy–momenta $p_1,p_2,\ldots,p_n$.

AC is the assertion that $\mathcal{H}=\mathcal{H}_{\text{in}}=\mathcal{H}_{\text{out}}$, that is, that each state in $\mathcal{H}$ is asymptotically tangent to a free-particle state, with particles of mass $m$, both before and after interactions (the two free-particle states are different if there are interactions). This condition cannot be expected to always hold in the general framework introduced above, even if we restrict our attention to "physically reasonable" theories in which states of $\mathcal{H}$ are asymptotically tangent to free-particle states before and after interactions: the absence of other stable particles with different masses is not guaranteed. For instance, even if $A$ is "neutral," the action of field operators on the vacuum might generate pairs of "charged" particles with opposite charges, whatever "charge" one might imagine. Individual charged particles cannot occur in the neutral space $\mathcal{H}$ and their mass thus does not appear in the spectral condition. Hence, such states of pairs of charged particles will not belong to $\mathcal{H}_{\text{in}}$ or $\mathcal{H}_{\text{out}}$ although they belong to $\mathcal{H}$. However, if the set of charged particles is known, it can be shown that the above framework might be enlarged by defining charged fields, in such a way that AC might still be valid in the enlarged framework (see the article of Buchholz and Summers). For simplicity, we restrict below our attention to the simplest theories in which AC holds in the way stated above.

If AC holds, it is shown that there exists a linear operator $S$ from $\mathcal{H}$ to $\mathcal{H}$, called "collision operator" or "$S$-matrix," that relates the "initial" and "final" free-particle states to which a state in $\mathcal{H}$ is tangent before and after interactions, respectively; if AC does not hold, $S$ can also be defined as in operator in $\mathcal{F}$. Collision amplitudes or scattering functions are the energy–momentum kernels of $S$ for given numbers $m$ and $n$ of initial and final particles. As easily seen, they are well-defined distributions on the space of all initial and final on-shell energy–momenta. For convenience, we will denote by $p_k$ the physical energy–momentum of a final particle with index $k(p_k\in H_+(m))$, and by $-p_k$ the physical energy–momentum of an initial particle $(-p_k\in H_+(m))$.

### Wightman Functions, Chronological Functions, and LSZ Reduction Formulas

The $N$-point Wightman "functions" $W_N$ are defined as the vacuum expectation values (VEVs) of the products of $N$ field operators, namely:

$$W_N(x_1, x_2, \ldots, x_N) \\ = < \Omega, A(x_1)A(x_2)\cdots A(x_N)\Omega >$$

The chronological functions $T_N$ are the VEVs of the chronological products of the fields $A(x_1), \ldots, A(x_N)$: in the latter, fields are ordered according to decreasing values of the time components of the points $x_k$. $T_N$ is essentially well defined due to local commutativity with, however, problems not treated here at coinciding points.

$\tilde{T}_N(p_1, \ldots, p_N)$ will denote the Fourier transform of $T_N$. In view of the invariance of the theory under spacetime translations, functions above are invariant under global spacetime translation of all points $x_k$ together. Hence, their Fourier transforms contain an energy–momentum conservation (e.m.c.) delta function $\delta(p_1 + p_2 + \cdots + p_N)$. Connected $N$-point functions are defined by induction (over $N$) via a formula expressing each (nonconnected) function as the sum of the corresponding connected function and of products of connected functions depending on subsets of points. In contrast to nonconnected functions, the analysis shows that connected functions in energy–momentum space do not contain in general e.m.c. delta functions involving subsets of energy–momenta.

It can be shown that the two-point function $\tilde{T}_2(p_1, p_2) = \delta(p_1 + p_2)\tilde{T}_2(p_1)$ has a pole of the form $1/(p_1^2 - m^2)$ and that $\tilde{T}_N$ has similar poles for each energy–momentum variable $p_k$ on the mass-shell. The connected, amputated chronological function $\tilde{T}_N^{\mathrm{amp,c}}$ is defined by multiplying $(\tilde{T}_N)_{\mathrm{connected}} = \tilde{T}_N^{\mathrm{c}}$ (for $N \geq 2$) by the product of all factors $p_k^2 - m^2$ that cancel these poles. It is then shown that it can be restricted as a distribution to the mass-shell of any physical process with $m$ initial and $n$ final particles, with $m + n = N$, and that this restriction coincides with the collision amplitude of the process. A process is here characterized by fixing the initial and final indices.

The analyticity properties of interest (described below) will apply to the connected functions after factoring out their global e.m.c. delta functions.

### The Analytic *N*-point Functions

The Wightman axioms (without so far AC) yield general analyticity, as also asymptotic causality, properties that we now describe. The analysis is essentially based on the interplay of support properties in $x$-space arising from local commutativity and the definition of chronological operators, and support properties in $p$-space due to the spectral condition. Support properties in $x$-space apply to cell and more general "paracell" functions which are VEVs of adequate combinations of products of "partial" chronological operators. It is shown that each such function has support in $x$-space in a closed cone $C_S$ (with apex at the origin). Moreover, for cell functions, the cone $C_S$ is convex and salient. Hence, in view of the usual Laplace transform theorem, the cell function in $p$-space (after Fourier transformation) is the boundary value of a function analytic in complex space in the tube $\mathrm{Re}\, p$ arbitrary, $\mathrm{Im}\, p$ in the open dual cone $\tilde{C}_S$ of $C_S$. It is also shown that, near any real point $P = (P_1, \ldots, P_N)$, the chronological function in $p$-space coincides with one or more cell functions.

Together with support properties in $p$-space arising from the spectral condition and the use of coincidence relations between some cell functions (in adequate real regions in $p$-space), one then shows the existence, for each $N$, of a well-defined, unique analytic function $F_N$, called the "analytic $N$-point function," whose domain of analyticity, the "primitive domain of analyticity," in complex $p$-space contains all the tubes $\mathcal{T}_S$ associated with the cell functions. It also contains in particular a complex neighborhood of the Euclidean energy–momentum space which consists of energy momenta $P_k$ with real $\boldsymbol{P}_k$ and imaginary energies $(P_k)_0$. Moreover, the chronological function $\tilde{T}_N^{\mathrm{amp,c}}$ is the boundary value of $F_N$ at all real points $P$, from imaginary directions which include those of the convex envelope of the cones $\tilde{C}_S$ associated with cell functions that coincide locally with $\tilde{T}_N^{\mathrm{amp,c}}$.

However, the primitive domain has an empty intersection with the complex mass-shell, and thus gives no result on analyticity properties of collision amplitudes on the (real or complex) mass-shell. For $N = 4$, it has been possible to largely extend the primitive domain (which is not a "natural domain of holomorphy") by computing (parts of) its holomorphy envelope, which now has a nonempty intersection with the complex mass shell. It is shown in turn that the four-point function $F_4$ can be restricted to the complex mass-shell in a one-sheeted domain, called the "physical sheet," that admits each (real) physical region on its boundary (there is here one physical region for each choice of the two initial and the two final indices, the corresponding physical regions being disconnected from each other). In each physical region, the collision amplitude is the boundary value of the mass-shell restriction of $F_4$, from the corresponding half-space of "$+\mathrm{i}\varepsilon$" directions $\mathrm{Im}\ s > 0$, where $s$ is the (squared) energy of the process.

The analyticity domain on the complex mass-shell contains paths of analytic continuation between the various physical regions (“crossing property”) and admits cuts $s_{ij}$ real $\geq (2m)^2$ covering the various physical regions. From these analyticity properties in the physical sheet, one can also derive “dispersion relations” (*see* Dispersion Relations).

#### Asymptotic causality and analyticity properties for $N \geq 4$

No similar result has been achieved at $N > 4$, and as a matter of fact, no similar result is expected if the AC condition is not assumed. The best results achieved so far are decompositions of the collision amplitude, in various parts of its physical region, as a sum of boundary values of functions analytic in domains of the complex mass-shell. In contrast to the case $N = 4$, the sum reduces to one term only in a certain subset of the physical region. Near other points, the $N$-point analytic function cannot be restricted locally to the complex mass-shell, though it can be decomposed as a sum of terms which, individually, are locally analytic in a larger domain that intersects the complex mass-shell.

These analyticity properties for $N \geq 4$ are a direct consequence of (and equivalent to) an asymptotic causality property that we now outline. Let $f_{k,\tau}(p)$ be, for each index $k$, a test function of the form

$$f_{k,\tau}(p) = \mathrm{e}^{\mathrm{i}p.\tau u_k}\, \mathrm{e}^{-\gamma\tau|p_k - P_k|^2}$$

where each $u_k$ is a point in spacetime, $P_k$ is a given on-shell energy–momentum, and $\tau$ will be a spacetime dilatation parameter ($\gamma > 0$). It is well localized in $p$-space around the point $P_k$ and its Fourier transform is well localized in $x$-space around the point $\tau u_k$ up to an exponential fall-off of width $\sqrt{\gamma\tau}$ which is small compared to $\tau$ as $\tau \to \infty$.

We now consider the action of the (connected, amputated) chronological function on such test functions. A configuration $u = (u_1, \ldots, u_N)$ will be called “noncausal” at $P = (P_1, \ldots, P_N)$ if this action decays exponentially as $\tau \to \infty$. In mathematical terms, $u$ is then outside the “essential support” or “microsupport” at $P$. The asymptotic causality property established, has roughly the following content: the only possible causal configurations $u$ at $P$ are those for which energy–momentum can be transferred from the initial to the final points in future cones. Moreover, at least two initial “extremal” points must coincide, as also two extremal final points. The simplest example is the case $N = 4$; if, for example, indices 1,2 are initial and $3, 4$ final, then the only *a priori* possible causal situations are such that $u_3 = u_4$ is in the future cone of $u_1 = u_2$ (in this particular case Lorentz invariance implies that $u_3 - u_1$ must be proportional to $P_3 + P_4$). In more general cases, the possible causal configurations $u$ depend on $P$.

## AC and Analyticity

### Asymptotic Causality in Terms of Particles and Landau Singularities

As a matter of fact, a better causality property “in terms of particles” – which is the best possible one – is expected for “physically reasonable” theories if the (stable) particles of the theory are known. (By physically reasonable, we mean the absence of “à la Martin” pathologies such as the occurrence of an infinite number of unstable particles with arbitrary long lifetime). That property expresses the idea that the only causal configurations $u$ at $P$ are those for which the energy–momentum can be transferred from the initial to the final points via intermediate stable particles in accordance with classical laws: there should exist a classical connected multiple scattering diagram in spacetime joining the initial and final points $u_k$, with physical on-shell energy–momenta for each intermediate particle and energy–momentum conservation at each (pointwise) interaction vertex.

This property, if it holds, yields in turn (and is equivalent to) improved analyticity of the analytic $N$-point function near real physical regions: the (on-shell) collision amplitude is the boundary value of a unique analytic function in its physical region, at least away from some “exceptional points.” The boundary value (namely the collision amplitude) is moreover analytic outside Landau surfaces $L_+(\Gamma)$ of connected multiple scattering graphs $\Gamma$; and along these surfaces (which are in general smooth codimension-1 surfaces), it is in general obtained from well-specified “$+\mathrm{i}\varepsilon$” directions (that depend in general on the real point $P$ of $L_+$).

Exceptional points are those that lie at the intersection of two (or several) surfaces $L_+(\Gamma_1)$, $L_+(\Gamma_2) \ldots$, with opposite causal directions, and hence having no $+\mathrm{i}\varepsilon$ directions in common (in the on-shell framework). Such points do not occur at $N = 4$ for two-body processes, in which case the surfaces $L_+$ are the $n$-particle thresholds $s = (nm)^2$, with $n \geq 2, s = (p_1 + p_2)^2$. They do occur more generally: in a $3 \to 3$ process, 1,2,3 initial, 4,5,6 final, this is the case of all points $P$ such that $-P_1 = P_4$, $-P_2 = P_5$, $-P_3 = P_6$ which all belong to the Landau surfaces of the two graphs $\Gamma_1, \Gamma_2$, with only one internal line joining two interaction

vertices: in the case of $\Gamma_1$, (resp., $\Gamma_2$), the first vertex involves the external particles 1, 2, 4 (resp., 1, 3, 5), while the second one involves 3, 5, 6 (resp., 2, 4, 6). If moreover $P_1, P_2, P_3$ lie in a common plane, previous points $P$ also lie on surfaces $L_+$ of "triangle" graphs with again opposite causal directions at $P$. The fact that $+i\epsilon$ directions are opposite can equally be checked for the corresponding Feynman integrals of perturbative field theory.

**Remark** The above points are no longer exceptional in spacetime dimension 2. In fact, all surfaces $L_+$ mentioned then coincide with the (on-shell) codimension-1 surface $-p_1 = p_4, -p_2 = p_5, -p_3 = p_6$, with two opposite causal directions. The previous asymptotic causality property, together with a further "causal factorization" property for causal configurations, then yields along that surface an actual factorization of the three-body (nonconnected) $S$-matrix into a product of two-body scattering functions modulo an analytic background. The latter vanishes outside the surface, hence is identically zero, for some special two-dimensional models.

In the absence of the AC condition, one clearly sees why the above causality in terms of particles cannot be established: as we have seen, there is *a priori* no control on the stable particles of the theory and on their masses, and pathologies such as those mentioned above cannot be excluded. Hopefully, the first problem should be solved if AC is assumed, and the second one should be removed by adequate regularity assumptions. This is the purpose of the so-called axiomatic nonlinear program, in which one also wishes to examine further problems, for example, analytic continuation into unphysical sheets, with the occurrence of possible unstable particle poles and other singularities, nature of singularities, possible multiparticle dispersion relations, . . . , to cite only a few. Results so far remain limited but provide a first insight into such problems.

### The Nonlinear Axiomatic Program

Results described below are based on discontinuity formulas arising from – and essentially equivalent in adequate energy regions to – AC, together with some regularity conditions. They can be established either with or without the introduction of adequate "irreducible" kernels. The methods rely on some general preliminary results on Fredholm theory in complex space (and with complex parameters). Irreducible kernels are defined through integral (Fredholm type) equations, first in the Euclidean region (imaginary energies) and then by local distortions of integration contours allowing one to reach the Minkowskian region. From discontinuity formulas and algebraic arguments, these irreducible kernels are shown to have analyticity (or meromorphy) properties associated with the physical idea of irreducibility (see examples below).

Results obtained so far with or without irreducible kernels are comparable in the simplest cases. However, the method based on irreducible kernels gives more refined results and seems best adapted to "extricate" the analytic structure of $N$-point functions for $N > 4$.

### $N = 4$, Two-Body Processes in the Low-Energy Region

By even theory, we mean theories in which $N$-point function vanishes identically for $N$ odd.

Standard results on two-body processes with initial (resp., final) energy–momenta $p_1, p_2$ (resp., $p'_1, p'_2$) in the low-energy region $(2m)^2 \leq s < (3m)^2$ ($s = (p_1 + p_2)^2 = (p'_1 + p'_2)^2$) are based on the "off-shell unitarity equation"

$$F_+ - F_- = F_+ \star F_- \qquad [1]$$

where $F_+(p_1, p_2; p'_1, p'_2)$ and $F_-(p_1, p_2; p'_1, p'_2)$ denote, respectively, the $+i\varepsilon$ and $-i\varepsilon$ boundary values of the four-point function $F_4$ from above or below the cut $s \geq (2m)^2$ in the physical sheet, and $\star$ denotes on-shell convolution over two intermediate energy–momenta. This relation is a direct consequence of AC for $s$ less than $(3m)^2$, or less than $(4m)^2$ in an even theory. When the four external energy–momentum vectors $p_1$, $p_2$, $p'_1$, $p'_2$ are put on the mass shell (on both sides of that relation), one recovers the usual elastic unitarity relation for the collision amplitude $T_+$ and its complex conjugate $T_-$:

$$T_+ - T_- = T_+ \star T_-$$

In the exploitation of these relations outlined below, a regularity condition is moreover needed, for example, the continuity of $F_+$ in the low-energy region.

By considering the unitarity equation as a Fredholm equation for $T_+$ at fixed $s$ (in the complex mass shell), one obtains the following result: $T_+$ can be analytically continued as a meromorphic function of $s$ through the cut (in the low-energy region) in a two-sheeted ($d$ even) or multisheeted ($d$ odd) domain around the two-particle threshold. Possible poles in the second sheet (generated by Fredholm theory) will correspond physically to unstable particles. The singularity at the two-particle threshold is of the square-root type in $s$ for $d$ even, or in

$1/\log s$ for $d$ odd. The difference between the two cases is due to the power $(d-1)/2$ of $s$, integer or half-integer, in the kinematical factor arising from on-shell convolution. This result can also be extended to the off-shell function $F_4$ by applying a further argument of analytic continuation making use of the off-shell unitarity equation.

Restricting now our attention to an even theory (for simplicity), a similar result also follows from the introduction of a two 2PI BS type kernel $G$ satisfying (and here defined from $F$ through) a regularized BS equation of the form

$$F = G + F \circ_M G \qquad [2]$$

where $\circ_M$ denotes convolution over two intermediate energy–momenta with two-point functions on the internal lines and a regularization factor in order to avoid convergence problems at infinity ($G$ then depends on the choice of this factor but its properties and the subsequent analysis do not). Alternatively, one may also introduce a kernel satisfying a renormalized BS equation, but this is not useful for present purposes.

Starting from the above discontinuity formula [1], one shows in turn that $G$ is indeed "2PI" in the analytic sense:

$$G_+ = G_- \qquad [3]$$

in the low-energy region. More precisely, $G$ is analytic or meromorphic (with poles that may arise from Fredholm theory) in a domain that includes the two-particle threshold $s=(2m)^2$, in contrast to $F$ itself.

The proof of [3] is based on the relation independent of $M$ (and thus leaving the $M$ dependence implicit).

$$\circ_+ - \circ_- = \star \qquad [4]$$

(which is a nontrivial adaptation of the decomposition of a mass-shell delta function as a sum of plus and minus i$\varepsilon$ poles). A simple algebraic argument then shows essentially the equivalence between the discontinuity formulas [1] and [3].

In turn, assuming that $G$ has no poles, this analyticity allows one to recover the two-sheetedness ($d$ even) or multisheetedness ($d$ odd, singularity in $1/\log$) of $F$, in view of the BS type equation.

### $N=6$, 3–3 Process in the Low-Energy Region (Even Theory)

The result, in the neighborhood of the 3–3 physical region, is here a "structure equation" expressing the 3–3 function $F$ in the low-energy region as a sum of "à la Feynman contributions" associated with graphs with one internal line and with triangle graphs, with two-point functions on internal lines and four-point functions at each vertex, plus a remainder $R$. The latter is shown to be a boundary value from $+\mathrm{i}\varepsilon$ directions Im $s$ positive, where $s=(p_1+p_2+p_3)^2, p_1, p_2, p_3$ denoting the energy–momentum vectors of the initial particles. Further regularity conditions are needed to recover its local physical region analyticity. The various explicit contributions that we have just mentioned yield the actual physical region Landau singularities expected in the low-energy 3–3 physical region.

A more refined result, in the approach based on irreducible kernels outlined below, applies in a larger region and then includes further à la Feynman contributions associated with 2-loop and 3-loop diagrams (the latter do not contribute to "effective" singularities in the neighborhood of the physical region).

The first result can be established from discontinuity formulas for the three-point function around two-particle thresholds, arising from AC, and "microsupport" analysis of all terms involved. In the approach based on irreducible kernels, it is useful to introduce in particular a 3PI kernel $G_3$ that, in contrast to the 3–3 function, will be analytic or meromorphic in a domain including the three-particle threshold. To that purpose, an adequate set of integral equations is introduced and the three-particle irreducibility of $G_3$ in "the analytic sense" is then established. In turn it provides the complete structure equation mentioned above.

### More General Analysis

There are so far only preliminary steps in more general situations, in view of (difficult) technical problems involved and the need of *ad hoc* regularity assumption at each stage. As already mentioned, the approach based on irreducible kernels seems best adapted. The analysis should clearly involve more general irreducible kernels with various irreducibility properties with respect to various channels (and not only with respect to the basic channel considered such as the 3–3 channel in the case above). From a heuristic viewpoint, one may first consider to that purpose adequate formal expansions into (infinite) sums of "à la Feynman contributions" adapted to the energy regions under investigation. These à la Feynman contributions will involve adequate irreducible kernels in the graphical sense at each vertex, and the above expansions correspond formally to the best possible regroupings of Feynman integrals with respect to the energy region considered. From such expansions, one might

determine adequate sets of integral equations allowing one, together with regularity assumptions, to carry out an analysis similar to above.

## The Models

A Euclidean field-theoretical model can be defined by a probability measure $d\mu(\varphi)$ on the space of tempered distributions $\varphi$ in Euclidean spacetime, whose moments verify the Osterwalder–Schrader (or similar) axioms. The moments of $d\mu$ are, for each $N$, the Euclidean (Schwinger) $N$-point functions:

$$S(x_1, \ldots, x_N) = \int \varphi(x_1) \cdots \varphi(x_N)\, d\mu(\varphi) \qquad [5]$$

In what follows, the measure $d\mu$ will be a perturbed Gaussian measure which, for the massive $\varphi^4$ model with a volume cutoff $\Lambda$ and an ultraviolet cutoff $\rho$, is given in $d$ dimensions by

$$d\mu_{\Lambda,\rho} = e^{-\lambda(\rho)\int_\Lambda \varphi^4(z)\,dz + a(\rho)\int_\Lambda \varphi^2(z)dz}\, d\nu_\rho(\varphi)/Z_{\Lambda,\rho} \qquad [6]$$

where $Z_{\Lambda,\rho}$ is the normalization factor and where $d\nu_\rho(\varphi)$ is the Gaussian measure of mean zero ($\int \varphi\, d\nu = 0$) and covariance

$$C(x-y;\rho) = \int d^d p\, e^{ip(x-y)}\, e^{-p^2/\rho^2}/(\zeta(\rho)p^2 + m^2)$$

where by convention $m$ is called the bare mass.

For $d = 2$ or 3 one can show that, for $\lambda(\rho) = \lambda$ small enough (depending on $m$) and $\zeta(\rho) = 1$, there exists a function $a(\rho)$ ($a(\rho) = O(\lambda)$ as $\lambda \to 0$) such that, for any set of $N$ distinct points, the function $S(x_1, \ldots, x_N) = \lim_{\Lambda,\rho\to\infty} S_{\Lambda,\rho}(x_1, \ldots, x_N)$ exists, is not Gaussian (hence does not correspond to a trivial, free theory), and satisfies the Osterwalder–Schrader axioms. The connected part $S(x_1, \ldots, x_N)_{\text{connected}}$ has the following perturbative series:

$$\lim_{\Lambda,\rho\to\infty} \sum_n \frac{(-1)^n}{n!} \int \varphi(x_1) \ldots \varphi(x_N) \times \left[\int_\Lambda [\lambda\varphi^4 - a(\rho)\varphi^2](z)\, dz\right]^n d\nu_{\Lambda,\rho}(\varphi)\big|_{\text{connected}} \qquad [7]$$

which is the (divergent) sum of the connected renormalized (Euclidean) Feynman graphs.

The study of the perturbative series leads to the distinction of:

1. the super-renormalizable theories, where it is possible to take $\lambda(\rho), \zeta(\rho)$ not depending on $\rho$. In dimension 2, all the models where $\lambda\varphi^4$ is replaced by

$$c_{2p}\varphi^{2p} + c_{2p-1}\varphi^{2p-1} + \cdots + c_5\varphi^5 + \lambda\varphi^4 + c_3\varphi^3 \qquad [8]$$

   also exist provided that $c_{2p} > 0$ is small enough depending on $m$ and on the other coefficient $c$'s and $\lambda$, and
2. the just renormalizable theories where $\lambda(\rho)$ (and possibly $\zeta(\rho)$) depend in general on $\rho$. In models mentioned below $\lambda(\rho) \to 0$ as $\rho \to \infty$; this characterizes "asymptotic freedom."

The proof of the existence of the $N$-point functions makes use of Taylor type expansions with remainder. The first orders are used to compute $\lambda(\rho), \zeta(\rho), a(\rho)$. The idea is to consider the functional integral [5] – at $\Lambda, \rho$ finite – as an integral over roughly $\Lambda\rho^d$ "degrees of freedom" which are weakly coupled. This corresponds to a decomposition of the phase space (with cutoff both in $x$-space (the box $\Lambda$) and in $p$-space (roughly $|p| < \rho$)). The coupling between different regions in $x$-space comes from the propagators $C_\rho$; the coupling between different frequencies in $p$-space comes from the $\varphi^4$ term (the interaction vertex). The expansion is then, for each degree of freedom, a finite expansion in the coupling between this degree and the others so that, even if the expansion is perturbative up to the order $\Lambda\rho^d$, the bound on each term is qualitatively the one on a product of $\Lambda\rho^d$ finite order-independent expansions, the order of which can be fixed uniformly in $\rho$ (and depending only on $\lambda$). To achieve this program, the propagator linking two points of distance of order $L$ must have a decrease of order $e^{-L^{-1}|x-y|}$, that is, have momentum larger than $L^{-1}$, so that one must localize both in $x$-space and $p$-space ; for example, the smallest cells of phase space correspond to fields $\varphi$ localized in $x$, $p$-spaces, the $x$-boxes being of side $\rho^{-1}$ and the $p$-localization consisting of values such that roughly $(\rho/2) \le |p| \le \rho$. More generally, a generic cell (of index i) corresponds to fields $\varphi$ at point $x$ and momentum $p$, with $x$ in a box of side $2^i\rho^{-1}$ and $2^{-i-1}\rho < |p| < 2^{-i}\rho$.

These expansions are mimicking the à la Wilson renormalization group. For just renormalizable theories (where $\lambda(\rho)$ depends on $\rho$), one is led to introduce the effective coupling constant $\lambda(2^{-i}\rho)$ whose perturbative expansion is the value at momentum zero of the sum of all the (connected, amputated) four-point functions containing only propagators of momentum (roughly) bigger than $2^{-i}\rho$ (plus $\lambda(\rho)$ which in fact tends to zero as $\rho \to \infty$).

Then by small coupling we mean a theory where $\lambda(2^{-i}\rho)/\zeta(2^{-i}\rho)^2$ is small for all $i$.

By convention we write $\lambda_{\text{ren}}, \zeta_{\text{ren}}, a_{\text{ren}}$ for the effective parameters of the theory at zero momentum.

The expansion obtained expresses $S_{\text{connected}}$ as a sum of terms each of them being associated to a

given set of phase-space cells which are "connected" together by "links" that are either propagators or vertices. Each term decreases exponentially with the difference $i_{\max} - i_{\min}$ of the upper and lower indices of the phase-space cells involved. Moreover, each set must contain the cells associated to the fields $\varphi(x_1)\cdots\varphi(x_N)$ whose indices are fixed by the order of magnitude of the distances between the points. On the other hand, the difference between the theory of cutoff $\rho$ and the one of cutoff $2\rho$ are terms containing at least one cell of momentum of order $\rho$; these terms are thus small like $\mathrm{cst}(x_1,\ldots,x_N)\mathrm{e}^{-(\mathrm{cst})\rho}$, so that the limit as $\rho\to\infty$ exists.

So far, the "construction" of models is possible only at small coupling, apart from special cases. The $\varphi^4$ theory in dimension 4 is just renormalizable (from the perturbative viewpoint) but the above condition of small coupling cannot be achieved (and it is generally believed that this model cannot be defined as a nontrivial theory). A just renormalizable model has been shown to exist, namely the Gross–Neveu model which is a fermionic theory in dimension 2. The elementary particle physics models are just renormalizable but their construction has not been completed so far (in particular in view of the confinement problem). *See* Constructive Quantum Field Theory for details.

To state the result in a form convenient for our purposes here, we introduce a splitting of the covariance in two parts:

$$C(x-y;\rho) = C_M(x-y;\rho) + C_{>M}(x-y;\rho), \quad M > m$$
$$\tilde{C}_M(p;\rho) = (\mathrm{e}^{-p^2/\rho^2}/p^2 + m^2) - (\mathrm{e}^{-p^2/\rho^2}/p^2 + M^2)$$

so that $C_M(x-y)$ behaves like $C$ at large distances but has an ultraviolet cutoff of size $M$, and $|C_{>M}(x-y)| \le \mathrm{e}^{-M|x-y|}$ decreases exponentially depending on the (technical) choice of $M$. Let $\mathrm{d}\nu_M(\varphi)$ be the Gaussian measure of covariance $C_M$.

One divides also $\Lambda$ in unit cubes and obtains for the connected $N$-point function an expansion as a sum over connected trees; a tree $T$ is composed of lines $\ell$ and vertices $v$; each line joins two vertices or one of the external points $x_1,\ldots,x_N$ and a vertex; moreover, there are no loops.

To each line $\ell$ is associated a propagator $C_M(z_\ell, z'_\ell) = C_M(\ell)$.

To each vertex $v$ are associated:

1. two subsets $I_v, I'_v$ of $\{\ell\}$,
2. a connected set $X_v$ of unit cubes such that all the $z_\ell, \ell\in I_v$ and all the $z'_\ell, \ell \in I'_v$ are contained in $X_v$; $|X_v|$ is the volume of $X_v$, and
3. a kernel $K_{X_v}(\{z,z'\}_v;\varphi)$

Finally, the external points are by convention $z_\ell$ points; then:

$$\begin{aligned}&S_{\Lambda,\rho}(x_1,\ldots,x_N)_{\text{connected}}\\&\quad= \int \mathrm{d}\nu_M(\varphi)\sum_T \frac{1}{|T|!}\sum_{\substack{\{X_v\}\\ \text{nonoverlapping}}}\\&\qquad\times\int\Bigg[\prod_{\substack{\ell\\ z_\ell \text{ not external}}}\mathrm{d}z_\ell\Bigg]\Bigg[\prod_{\ell\in T}\mathrm{d}z'_\ell C_M(\ell)\Bigg]\\&\qquad\times\prod_{v\in T}K_{X_v}(\{z,z'\}_v;\varphi)\end{aligned} \qquad [9]$$

where for coupling small enough:

$$\int \mathrm{d}\nu_M(\varphi)\prod_{v\in T}|K_{X_v}(\{z,z'\}_v;\varphi)| \le \prod_{v\in T}\mathrm{e}^{-M(1-\epsilon)|X_v|} \qquad [10]$$

The $X$'s are $2\times 2$ nonoverlapping; however, it will suffice to sum over all $X$'s (without restriction) to get a bound showing the convergence of the expansion as $\Lambda\to\infty$. In this formula the $K(.,\varphi)$'s are still coupled by the measure $\mathrm{d}\nu_M(\varphi)$; all the nonperturbativity is hidden in the $K$'s (in particular the contribution of momentum bigger than $M$).

As a consequence of [9] and if $a(\rho,\lambda)$ has been chosen such that $a_{\text{ren}} = 0$, for $M$ large enough and at small coupling (depending on $M$, $m$):

$$\begin{aligned}&|S(x,y)_{\text{connected}}|\\&\quad\le |C_M(x-y)| + \int \mathrm{d}z'_1\mathrm{d}z'_2|C_M(x-z'_1)\\&\qquad\times \mathrm{e}^{-M(1-\epsilon)|z'_1-z'_2|}C_M(z'_2-y)| + \cdots\\&\quad\le (\mathrm{cst})\mathrm{e}^{-m(1-\epsilon)|x-y|}\end{aligned} \qquad [11]$$

More generally, the connected $N$-point function satisfies

$$|S(x_1,\ldots,x_N)_{\text{connected}}| \le \mathrm{cst}\ \mathrm{e}^{-m(1-\epsilon)d(x_1,\ldots,x_N)} \qquad [12]$$

where $d(x_1,\ldots,x_N)$ is the length of the smallest tree joining $x_1,\ldots,x_N$, with possibly intermediate points.

## The Irreducible Kernels

### The 1PI Kernel and a Lippmann–Schwinger Equation

To then show that a theory – if the perturbation series heuristically shows it – contains only one particle of mass smaller than $2m(1-\epsilon)$, it is necessary to expand further the coupling between the $K$'s in [9]. Each perturbative step relatively to this coupling will generate a sum of terms such that in each one there is a "new" propagator $C_M$ between two $K$'s.

The fact that in [9] the $X$'s are nonoverlapping has the consequence that an expansion where for each pair of $K_X$ the number of propagators $C_M$

remains bounded (say by $n+1$) is convergent (for small enough couplings depending on $m$, $n$); this is because, for a given $X$, the others must be farther and farther as their number increases, and in view of the exponential decrease (in $x$-space) of $C_M$.

We then consider the expansion where we have further expanded the two-point function $S(x,y)$ such that each term can be decomposed in the channel $x \to y$ in $C_M$ propagators and 1PI contributions (in the sense that any line cutting such a 1PI contribution (and outside the $X's$) cuts at least two propagators); that means that these 1PI contributions are no longer coupled by the $d\nu_M(\varphi)$ measure. They are made of propagators and of $K_X$ which still have nonoverlapping restrictions; the latter are straightforwardly expanded using a kind of (convergent) Mayer expansion; the result is finally a Lippmann–Schwinger type equation:

$$S(x,y)_{\text{connected}} = C_M(x-y) + \int dz_1\, dz_2\, C_M(x-z_1) \times G_1(z_1,z_2)\, C_M(z_2-y) + \cdots \qquad [13]$$

or

$$S(x,y)_{\text{connected}} = \left[ C_M \sum_{p\geq 0} [G_1 C_M]^p \right](x,y)$$

which is equivalent to

$$S_{\text{connected}} = C_M + C_M G_1 C_M + C_M G_1 S_{\text{connected}} \qquad [14]$$

where $G_1$ is a 1PI kernel that satisfies the bound

$$|G_1(t,u)| \leq \lambda_{\text{ren}} e^{-2m(1-\epsilon)|t-u|} \qquad [15]$$

In Fourier transform, eqn [14] becomes

$$F(p) = \tilde{C}_M(p) + \tilde{C}_M(p)\tilde{G}_1(p)\tilde{C}_M(p) + \tilde{C}_M(p) G_1(p) F(p) \qquad [16]$$

Denoting by $\delta(p+q)\, F(p,q)$ the Fourier transform of $S(x,y)_{\text{connected}}$, we can then compute $F(p)$:

$$F(p) = \frac{(p^2+m^2)[\tilde{C}_M + \tilde{C}_M \tilde{G}_1](p)}{(p^2+m^2) - (p^2+m^2)\tilde{C}_M \tilde{G}_1(p)} \qquad [17]$$

where $(p^2+m^2)\tilde{C}_M(p) \to (1-m^2/M^2)$ as $p\to 0$ and $|\tilde{G}_1(p)| \leq \lambda_{\text{ren}}\, \text{cst}(m)$ so that (as expected) $F$ has no pole in the Euclidean region at small coupling; but, as will be seen in the next section, it has a pole outside the Euclidean region.

### The 2PI Kernel and a BS Equation

From the previous discussion, it is clear that one can extract from [9] as many propagators as we want between kernels $K_X$. If one considers a splitting of the external points in incoming $x_1,\ldots,x_p$ and outgoing $x_{p+1},\ldots,x_N$ points, this defines a channel. One then obtains $n$PI kernels (in the given channel). In the same way as above, one obtains a relevant structure equation; this equation makes sense only if the kernels $K_X$ have a decrease corresponding to $n$-particle irreducibility; to that purpose we take $M > nm$. The expansion converges for couplings small enough depending on $m$ and $n$.

In the case $n=2$ this gives a kind of BS equation (the Lippmann–Schwinger equation corresponding to the case $n=1$); if we restrict, for simplicity, the analysis to even theories one is led to jump directly to the case $n=3$:

$$S(x_1,x_2;x_3,x_4)_{\text{connected}} = \int dz_1\, dt_1\, dz_2\, dt_2 (\circ_M)(x_1,x_2;z_1,t_1) \times G_2(z_1,t_1;z_2,t_2)(\circ_M)(z_2,t_2;x_3,x_4) + \cdots \qquad [18]$$

$$S = \circ_M \sum_{p\geq 1} [G_2 \circ_M]^p$$

or

$$S = \circ_M G_2 \circ_M + \circ_M G_2 S \qquad [19]$$

where

$$(\circ_M)(x_1,x_2;x_3,x_4) = S(x_1,x_3)S(x_2,x_4) + S(x_1,x_4)S(x_2,x_3)$$

and where

$$|G_2(t_1,t_2;u_1,u_2)| \leq \lambda_{\text{ren}} \exp\{-4m(1-\epsilon)\max_{i,j}(|t_i-u_j|)\} \qquad [20]$$

Equation [19] once amputated, and after Fourier transformation, is eqn [2].

### More General Irreducible Kernels and Structure Equations

Irreducible kernels with various degrees of irreducibility in various channels can be defined in a similar way. Corresponding expansions of $N$-point functions follow, in terms of integrals involving these kernels and two-point functions. These kernels are again convergent at small coupling ($\to 0$ as their irreducibility $\to \infty$) as well as the corresponding structure equations (which generalize eqn [18]).

## Analyticity, AC, and Bound States

As explained in the introduction, we now proceed by analytic continuation away from the Euclidean region in complex energy–momentum space.

First, it is easily seen that the two-point function is analytic in the region $s < (2m)^2 - \epsilon$ apart from a pole at $s = m_{\text{ph}}^2$ which defines the physical mass $m_{\text{ph}}$ ($m_{\text{ph}}^2$ is the zero in $p^2$ of the denominator in formula [17]). In view of the bounds of the previous two sections, $m_{\text{ph}}$ is close to the "bare" mass $m$.

The 2PI kernel, for even theories, is shown, again by Laplace transform theorem, to be analytic and bounded in domains around and away from the Euclidean region up to $s = (4m)^2 - \epsilon$, and is of the order of $\lambda_{\text{ren}}$.

As we have seen in the section "AC and analyticity," the analyticity of $G_2$ entails the analytic structure of $F$ (two-sheeted or multisheeted at the threshold). On the other hand, further poles of $F$ can be generated by the BS integral equation [2] in the physical or unphysical sheets. If a pole in the physical sheet occurs at $s < (2m_{\text{ph}})^2$ real, it will correspond to a new particle in the theory, namely a two-particle bound state.

### AC in the Low-Energy Region

The analysis of possible bound states, which will be presented in the following, will show that there might be at most one two-particle bound state of mass $m_{\text{B}} < 2m_{\text{ph}}$ which tends to $2m_{\text{ph}}$ as the couplings tends to zero.

On the other hand, for even theories, in view of the analyticity properties of the two-point function and of the 2PI kernel $G_2$, equation [1] holds in the region $(2m_{\text{ph}}^2) < s < (4m_{\text{ph}})^2 - \epsilon$, where $*$ is on-shell convolution with particles of mass $m_{\text{ph}}$.

If there is no two-particle bound state, this characterizes the AC of the theory for $s < (4m_{\text{ph}})^2 - \epsilon$.

If there is a bound state of mass $m_{\text{B}}$, AC is established only in the region $s < (3m_{\text{ph}})^2 - \epsilon$.

For non-even theories, the analysis is similar but requires the introduction of new irreducible kernels in view of the fact that the non-evenness opens new channels. AC in all cases can be established, for small couplings, up to $s < (3m_{\text{ph}})^2 - \epsilon$.

### Analysis of Possible Two-Particle Bound States for Even Theories at Small Coupling

It can be checked that such poles of $F$, if there are, either lie far away in the unphysical sheet(s) or are close to the two-particle threshold ($s = (2m_{\text{ph}})^2$). This is due to the convergence, at small coupling, of the Neumann series $F = G_2 + G_2 \circ_M G_2 + \cdots$. Individual terms $G_2 \circ_M \cdots \circ_M G_2$ are, in fact, defined away from the Euclidean region by analytic continuation in a two-sheeted ($d$ even) or multisheeted ($d$ odd) domain around the threshold: to that purpose locally distorted integration contours (initially the Euclidean region) are introduced as in the section "AC and analyticity," so as to avoid the pole singularities of the two-point functions involved in $\circ_M$, the threshold singularities being due to the pinching of this contour between the two poles as $s \to (2m_{\text{ph}})^2$. If a fixed neighborhood of the threshold is excluded, one does obtain uniform bounds of the form (cst $\lambda_{\text{ren}})^q$ (for a term with $q$ factors $G_2$) in any bounded domain, which ensures the convergence of the Neumann series.

It remains to study the neighborhood of the threshold. To that purpose, the following method is convenient. One shows that the convolution operator $\circ_M$ can be written in the form

$$\circ_M = g(s) * + \nabla \qquad [21]$$

where $*$ is, as in the section "AC and analyticity," on-shell convolution for $s > (2m_{\text{ph}})^2$ or is obtained by analytic continuation for complex value of $s$ around the threshold; $g(s) = 1/2$ for $d$ even and, if $d$ is odd, $g(s) = (\text{i}/2\pi)\log\sigma$, where $\sigma = 4m_{\text{ph}}^2 - s$. In view of this definition of $g(s)$, the operator $\nabla$ is regular: it is an analytic one-sheeted operation around the threshold (this is equivalent to [4]), and it has no pole singularities. This property of $\nabla$ can be established by geometric methods or by an explicit evaluation.

It is then useful to introduce a new kernel $U$ linked to $G_2$ by the integral equation

$$U = G_2 + U \nabla G_2 \qquad [22]$$

In view of the regularity and bounds of $\nabla$ and $G_2$, one sees (e.g., by a series expansion) that $U$, like $G_2$, is analytic in a neighborhood of the threshold and behaves in the same way at small $\lambda_{\text{ren}}$.

By a simple algebraic argument $F$ and $U$ are related by the integral equations

$$F = U + g(s)U * F = U + g(s)F * U \qquad [23]$$

**Two-dimensional models** We start the analysis with the case $d = 2$. The mass shell is trivial in this case; let $f$ be the restriction of $F$ to the mass shell; it depends only on $s = (p_3 + p_4)^2$ due to the mass shell and e.m.c. constraints (as also Lorentz invariance). On the mass shell, the operation $*$ becomes a mere multiplication and the integral equation [23] becomes

$$f(s) = u(s) + \frac{1}{a(s)} f(s)u(s) \qquad [24]$$

where $u$ is the mass shell restriction of $U$ and the factor $a(s)$ arising from $*$ is of the form $a(s) = \text{cst}\ s^{1/2}\sigma^{1/2}, \sigma = (2m_{\text{ph}})^2 - s$, which gives

$$f(s) = \frac{a(s)u(s)}{a(s) - u(s)} \qquad [25]$$

In turn one obtains

$$F = U + \frac{U_| \, {}_|U}{a(s) - u(s)} \qquad [26]$$

where $U_|$ (resp., ${}_|U$) is $U$ with $p_3, p_4$ (resp., $p_1, p_2$) restricted to the mass shell. Equation [26] completely characterizes the local structure of $F$ in view of the local analyticity of $U$.

The analysis of the possible poles follows from the fact that $U$ is equal to $G_2$ up to higher order in $\lambda_{\rm ren}$; on the other hand, $G_2$ is equal to a first known term plus higher-order corrections in $\lambda_{\rm ren}$ (if we expand in $\lambda_{\rm ren}$ the expression for $G_2$ obtained in the previous section), so that the leading contribution of $u(s)$ is known and the results follow.

For a theory (see [8]) containing a $\lambda_{\rm ren}\varphi^4$ term there is exactly one pole, which corresponds to the zero of $a(s) - u(s)$, lying in the region $(2m_{\rm ph})^2 - \epsilon < s < (2m_{\rm ph})^2$. This pole is either in the physical sheet for $\lambda_{\rm ren} < 0$ or in the second sheet if $\lambda_{\rm ren} > 0$. In the case $\lambda_{\rm ren} < 0$, this pole corresponds to a two-particle bound state of physical mass $m_{\rm B}$ which tends to $2m_{\rm ph}$ as $\lambda_{\rm ren} \to 0$.

In a model without $\varphi^4$ term ($\lambda_{\rm ren} = 0$) the lowest-order contribution to $G_2$, hence to $U$, is in general of the order of the square of the leading coupling, in which case there is always one bound state.

The treatment of the fermionic Gross–Neveu model, which involves spin and color indices, is analogous, with minor modifications. Equations now involve, in the two-particle region, $4 \times 4$ matrices; poles of $F$ are now the zeros of $\det(a(s)I - m(s)u(s))$, where $m(s)$ is the $4 \times 4$ matrix obtained from $2 \times 2$ residue matrices (whose leading matrix elements are explicitly computable). The detailed analysis, which requires the consideration of different channels (various color and spin indices) is omitted.

**Three-dimensional models** The results are similar: $F$ is decomposed as $F' + F''$, where $F'$ is the $\ell = 0$ "partial wave component" of $F$, namely $F' = (1/2\pi) \int F \, d\theta$, where $\theta$ is the "scattering angle" of the channel; its complement $F''$ is shown to be locally bounded in view of a further factor $\sigma$. The analysis is then analogous to the case $d = 2$ with $a(s)$ now behaving like $\mathrm{cst}/\log\sigma$ as $\sigma \to 0$. There is, *a priori*, either no pole, or one pole in the physical sheet at $s = m_{\rm B}^2 < (2m_{\rm ph})^2$ with $m_{\rm B} = 2m_{\rm ph} + O(e^{-\mathrm{cst}/\lambda_{\rm ren}})$, depending again on the signs of the couplings. For the existing even models such as the $\varphi^4$ model, there is no pole, hence no two-particle bound state.

**Four-dimensional models** The existence of the $\varphi^4$ model in dimension 4 is doubtful. If a four-dimensional model were defined, and if the 2PI kernel $G_2$ of a massive channel could be defined and shown to satisfy analyticity properties analogous to above, there would be no two-particle bound state at small coupling. In fact, the kinematical factor $\sigma^{(d-3)/2}$ (for $d$ even) generated by the mass shell convolution is no longer equal to $\sigma^{-1/2}$ as in the $d = 2$ case but now to $\sigma^{1/2}$. As a consequence, the Neumann series giving $F$ in terms of $G_2$ is convergent also in the neighborhood of the two-particle threshold.

**Non-even theories** The analysis for the non-even theories follows similar lines. As already mentioned, the analysis requires the introduction of new irreducible kernels. For the models $\lambda\varphi^4 + c_3\varphi^3$, which do exist at small couplings in dimensions 2 and 3, there will be either exactly one or no two-particle bound state, depending on the respective values of $\lambda, c_3$.

### Structure Equations and AC in Higher-Energy Regions

The structure equations of the previous section provide, after analytical continuation away from the Euclidean region, a rigorous version of the analysis presented at the end of the section "AC and analyticity." The irreducible kernels can here be defined in a direct way following the previous section, together with their analyticity properties. One has then to derive the discontinuity formulas that in turn characterize AC. This program has been carried out in the $3 \to 3$ particle region, and partly in the general case. It seems possible to complete general proofs up to some technical (difficult) problems. As already mentioned, in this approach, the coupling should be taken smaller and smaller as the energy region considered increases.

*See also:* Axiomatic Quantum Field Theory; Constructive Quantum Field Theory; Dispersion Relations; Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Perturbation Theory and its Techniques; Quantum Chromodynamics; Scattering in Relativistic Quantum Field Theory: Fundamental Concepts and Tools; Scattering in Relativistic Quantum Field Theory: the Analytic Program; Schrödinger operators.

## Further Reading

Bros J (1984) $r$-particle irreducible kernels, asymptotic completeness and analyticity properties of several particle collision amplitudes. *Physica A* 124: 145.

Bros J, Epstein H, and Glaser V (1965) A proof of the crossing property for two-particle amplitudes in general quantum field theory. *Communications in Mathematical Physics* 1: 240.

Bros J, Epstein H, and Glaser V (1972) Local analyticity properties of the $n$-particle scattering amplitude. *Helvetica Physica Acta* 43: 149.

Epstein H, Glaser V, and Iagolnitzer D (1981) Some analyticity properties arising from asymptotic completeness in quantum field theory. *Communications in Mathematical Physics* 80: 99.

Glimm J and Jaffe A (1981, 1987) *Quantum Physics: A Functionnal Integral Point of View*. Heidelberg: Springer.

Iagolnitzer D (1992) *Scattering in Quantum Field Theories: The Axiomatic and Constructive Approaches*, Princeton Series in Physics. Princeton University Press.
Iagolnitzer D and Magnen J (1987a) Asymptotic completeness and multiparticle structure in field theories. *Communications in Mathematical Physics* 110: 51.
Iagolnitzer D and Magnen J (1987b) Asymptotic completeness and multiparticle structure in field theories. II. Theories with renormalization: the Gross–Neveu model. *Communications in Mathematical Physics* 111: 89.
Martin A (1970) *Scattering Theory: Unitarity, Analyticity and Crossing*. Heidelberg: Springer.
Rivasseau V (1991) *From Perturbative to Constructive Renormalization*. Princeton: Princeton University Press.

# Schrödinger Operators

**V Bach**, Johannes Gutenberg-Universität, Mainz, Germany



Schrödinger operators are linear partial differential operators of the form

$$H_V = -\Delta + V(x) \qquad [1]$$

acting on a suitable dense domain $\mathrm{dom}(H_V) \subseteq L^2(\Omega)$ in the Hilbert space of square-integrable functions on a spatial domain $\Omega \subseteq \mathbb{R}^d$, where $d \in \mathbb{N}$. Here, $H_0 = -\Delta = -\sum_{\nu=1}^{d} \partial^2/\partial x_\nu^2$ is (minus) the Laplacian on $\Omega$, and the potential $V : \Omega \to \mathbb{R}$ acts as a multiplication operator, $[V\psi](x) := V(x)\psi(x)$.

## Historical Origin and Relation to Theoretical Physics

In 1926, Schrödinger formulated quantum theory as wave mechanics and proved later that it is equivalent to Heisenberg's matrix mechanics. He proposed that the state of a physical system at time $t \in \mathbb{R}$ is given by a normalized wave function $\psi_t \in L^2(\Omega)$ whose dynamics is determined by a linear Cauchy problem: $\psi_0$ is the state at time $t = 0$, and for $t > 0$, it evolves according to

$$\mathrm{i}\frac{\partial \psi_t}{\partial t} = H\,\psi_t \qquad [2]$$

the Schrödinger equation. More generally, $\psi_0$ is a normalized element of a Hilbert space $\mathcal{H}$, and the Hamiltonian $H_V$ is a self-adjoint operator, that is, $\mathrm{dom}(H_V) = \mathrm{dom}(H_V^*) \subseteq \mathcal{H}$ and $H_V = H_V^*$ on $\mathrm{dom}(H_V)$. Formally, eqn [2] is solved by the evolution operator or propagator $\exp(-\mathrm{i}tH_V)$ in the form $\psi_t = \exp(-\mathrm{i}tH_V)\psi_0$. The self-adjointness of $H_V$ insures the existence and unitarity of the propagator $\exp(-\mathrm{i}tH_V)$, for all $t \in \mathbb{R}$, so $\|\psi_t\| = \|\psi_0\| = 1$. For physics, this unitarity is crucial, because $\|\psi_t\|^2$ is interpreted as the total probability of the system to be at time $t$ in some state in $\mathcal{H}$. The general validity of eqn [2] as the fundamental dynamical law of all physical theories, including, for example, nonrelativistic and (special) relativistic quantum mechanics, quantum field theory, and string theory, deserves appreciation.

If the physical system under consideration is a nonrelativistic point particle of mass $m > 0$ in a potential $\widetilde{V} : \mathbb{R}^d \to \mathbb{R}$, then, according to the principles of classical (Newtonian) mechanics, its state is determined by its momentum $p \in \mathbb{R}^d$ and its position $x \in \mathbb{R}^d$, its kinetic energy is $(1/2m)p^2$, its potential energy is $\widetilde{V}(x)$, and the dynamics is given by the Hamiltonian flow generated by the Hamiltonian function $H_{\mathrm{class}}(p, x) = (1/2m)p^2 + \widetilde{V}(x)$. Schrödinger derived the Hamiltonian (operator) $H = -(\hbar^2/2m)\Delta + \widetilde{V}(x)$ in [2] from the replacement of the momentum $p \in \mathbb{R}^d$ by the momentum operator $-\mathrm{i}\hbar\nabla_x$. This prescription is called quantization and is further discussed in the section "Quantization and semiclassical limit." The Schrödinger operator $H_V$ in [1] is then obtained after an additional unitary rescaling, $\psi(x) \mapsto \mu^{d/2}\psi(\mu x)$, by $\mu := \hbar(2m)^{-1/2}$, and a redefinition $V(x) := \widetilde{V}(x/\mu)$ of the potential.

For more details, we refer the reader to Schrödinger (1926) and Messiah (1962).

## Self-Adjointness

Led by the requirement of unitarity of the propagator, the domain $\mathrm{dom}(H_V)$ in [1] is usually chosen such that $H_V$ is self-adjoint, which, in turn, is most often established by means of the Kato–Rellich perturbation theory, briefly described below. If $V \equiv 0$, then $H_0$ equals the Laplacian $-\Delta$, which is a positive self-adjoint operator, provided $\mathrm{dom}(H_0) = W^2_{\mathrm{b.c.}}(\Omega)$ is the second Sobolev space with suitable conditions on the boundary $\partial\Omega$ of $\Omega$. Typical examples are $\mathrm{dom}(H_0) = W^2(\mathbb{R}^d)$, for $\Omega = \mathbb{R}^d$, and $W^2_{\mathrm{Dir}}(\Omega)$ and $W^2_{\mathrm{Neu}}(\Omega)$ with Dirichlet or Neumann boundary conditions on $\partial\Omega$, respectively, in case that $\Omega$ is a bounded, open domain in

$\mathbb{R}^d$ with smooth boundary $\partial\Omega$. Starting from this situation, $V$ is required to be relatively $H_0$-bounded, that is, that $M(V,r) := V(-\Delta + r\mathbb{1})^{-1}$ defines (extends to) a bounded operator on $L^2(\Omega)$, for any $r > 0$. If $\lim_{r\to\infty} \|M(V,r)\| < 1$, then $H_V$ is self-adjoint on $\mathrm{dom}(H_0)$ and semibounded, that is, the infimum $\inf \sigma(H_V)$ of its spectrum $\sigma(H_V)$ is finite; in other words, $H_V \geq c\mathbb{1}$, for some $c \in \mathbb{R}$, as a quadratic form. (The semiboundedness corresponds to quasidissipativity, as a generator of the semigroup $\exp(-\beta H_V)$.)

A fairly large class of potentials fulfilling these requirements is defined by

$$\lim_{\alpha\searrow 0}\left\{\sup_{x\in\Omega}\int_{|x-y|\leq\alpha}|x-y|^{4-d}V(y)^2\,\mathrm{d}^d y\right\} = 0 \qquad [3]$$

for $d \neq 4$, and with $|x-y|^{4-d}$ replaced by $(\ln|x-y|)^{-1}$, for $d = 4$. For $d \leq 3$, [3] is equivalent to the uniform local square integrability of $V$, that is, $\sup_{x\in\Omega}\int_{|x-y|\leq 1} V(y)^2\,\mathrm{d}^d y < \infty$. Note that [3] allows for local singularities of $V$, provided they are not too severe; in this respect, quantum mechanics is more general than classical mechanics. Equation [3] is a sufficient condition for $H_V = -\Delta + V$ to be self-adjoint on $\mathrm{dom}(-\Delta)$ because $\lim_{r\to\infty}\|M(V,r)\| = 0$. Moreover, as eqn [3] only misses some borderline cases, it is also almost necessary for the self-adjointness of $H_V$. By means of Kato's inequality, the conditions on $V$, especially on its positive part $V_+ := \max\{V, 0\}$, can be further relaxed. Also, if one realizes $H_V$ as the Friedrichs extension of a semi-bounded quadratic form, the conditions to impose on $V$ are milder. One possibly loses, however, control over the operator domain $\mathrm{dom}(H_V)$, and typically $\mathrm{dom}(-\Delta)$ is only a core for $H_V$.

For further details on self-adjointness, we refer the reader to Reed and Simon (1980a, b), Kato (1976), and Cycon *et al.* (1987).

## Spectral Analysis

The self-adjointness of $H_V$ establishes a functional calculus, generalizing the notion of diagonalizability of finite-dimensional self-adjoint matrices: there exists a unitary transformation $W : L^2(\Omega) \to L^2(\sigma(H_V), \mathrm{d}\mu)$ such that $H_V$ acts on elements $\varphi$ of $L^2(\sigma(H_V), \mathrm{d}\mu_{H_V})$ as a multiplication operator, $[H_V\varphi](\omega) = \omega\varphi(\omega)$. The spectral measure $\mu_{H_V}$ decomposes into an absolutely continuous (ac) part $\mu_{H_V,\mathrm{ac}}$, a pure point (pp) part $\mu_{H_V,\mathrm{pp}}$, and a singular continuous (sc) part $\mu_{H_V,\mathrm{sc}}$, mutual disjointly supported on the ac spectrum $\sigma_{\mathrm{ac}}(H_V)$, the pp spectrum $\sigma_{\mathrm{pp}}(H_V)$, and the sc spectrum $\sigma_{\mathrm{sc}}(H_V) \subseteq \mathbb{R}$, respectively, whose union is the spectrum $\sigma(H_V)$ of $H_V$. There is an additional decomposition of the spectrum of $H_V$ into the discrete spectrum $\sigma_{\mathrm{disc}}(H_V)$, which consists of all isolated eigenvalues of $H_V$ of finite multiplicity, and its complement $\sigma_{\mathrm{ess}}(H_V) = \mathbb{R}\setminus\sigma_{\mathrm{disc}}(H_V)$, the essential spectrum of $H_V$, as its residual spectrum is void. One of the main goals of the spectral analysis is to determine the spectral measure for a given potential $V$ as precisely as possible.

In many applications, $\Omega = \mathbb{R}^d$ and the potential $V$ in $H_V$ is not only relatively $H_0$-bounded, but even relatively $H_0$-compact, that is, $M(V,1)$ is compact. In this case, $\lim_{r\to\infty}\|M(V,r)\| = 0$, insuring self-adjointness on $\mathrm{dom}(H_0)$ and semiboundedness of $H_V$. Moreover, a theorem of Weyl implies that its essential spectrum agrees with the one of $H_0$, that is, with the positive half-axis $\mathbb{R}_0^+$, and the discrete spectrum is contained in the negative half-axis $\mathbb{R}^-$. If, furthermore, $(H_0+1)^{-1}\,[x\cdot\nabla V(x)](H_0+1)^{-1}$ is compact, then the essential spectrum on the positive half-axis is purely absolutely continuous, $\sigma_{\mathrm{ess}}(H_V)\cap\mathbb{R}^+ = \sigma_{\mathrm{ac}}(H_V)\cap\mathbb{R}^+$, and hence $\sigma_{\mathrm{disc}}(H_V) \subseteq \sigma_{\mathrm{pp}}(H_V) \subseteq \sigma_{\mathrm{disc}}(H_V)\cup\{0\}$; the singular continuous spectrum is void.

We remark that the absence of singular continuous spectrum is not understood. Indeed, it is possible to explicitly construct potentials $V$ such that $H(V)$ has singular continuous spectrum. In terms of the Baire category, singular continuous spectrum is even typical. The appearance of singular continuous spectrum can, perhaps, be easier understood in terms of the dynamical properties of $\exp[-itH_V]$, rather than the spectral analysis of its generator $H_V$: Singular continuous spectrum occurs when initially localized states are not bound states, but move out to infinity very slowly.

The reader is referred to Simon (2000), Reed and Simon (1980a, b) and Cycon *et al.* (1987) for further detail.

## Properties of Eigenfunctions

Let us assume $\Omega = \mathbb{R}^d$, that $V \leq 0$ is nonpositive, fulfills [3], and that $\lim_{|x|\to\infty} V(x) = 0$. From the statements in the last section we conclude that $H_V = -\Delta + V(x)$ is semibounded, that the essential spectrum is the positive half-axis and that all eigenvalues are negative and of finite multiplicity, possibly accumulating only at 0. We collect some properties of the eigenfunctions $\psi_j \in L^2(\mathbb{R}^d)$ with corresponding eigenvalue $e_j < 0$, that is, $H_V\psi_j = e_j\psi_j$. The smallest eigenvalue $e_0 := \inf\sigma(H_V)$ (coinciding with the bottom of the spectrum) is simple, and the corresponding eigenfunction $\psi_0(x) > 0$ is strictly positive a.e. Elliptic regularity implies that at a given point $x \in \mathbb{R}^d$, the eigenfunction $\psi_j$ is almost $2 - d/2$ degrees more regular than $V$. For example,

if $V \in C^k[B_{2\epsilon}(x)]$, for some $\epsilon > 0$, then $\psi_j \in C^{k+\ell}[B_\varepsilon(x)]$, for all $\ell < 2 - d/2$. Agmon estimates (originally obtained by S'nol and also known in mathematical physics as Combes–Thomas argument) furthermore show that, for unbounded $\Omega$, the eigenfunction $\psi_j$ decays exponentially: $|\psi_j(x)| \leq C_\alpha e^{-\alpha|x|}$, for any $0 < \alpha < e_j$.

For more details, see Reed and Simon (1978, 1980a, b) and Cycon *et al.* (1987).

## One Dimension and Sturm–Liouville Theory

For $d = 1$, the stationary Schrödinger equation reduces to a second-order ordinary differential equation known as a Sturm–Liouville problem,

$$-\psi''(x) + V(x)\psi(x) = E\psi(x) \qquad [4]$$

on $L^2([a,b])$, with $V \in L^1([a,b])$ and independent boundary conditions at $-\infty \leq a < b \leq \infty$, say. Equation [4] admits an almost explicit solution by means of the Prüfer transformation defined by $\varphi(x) := \arctan[\psi(x)/\psi'(x)]$ and $R(x) := \ln\left(\sqrt{\psi(x)^2 + \psi'(x)^2}\right)$. The key point about the Prüfer transformation is that it effectively reduces the second-order differential equation [4] into a (nonlinear) first-order equation for $\varphi$,

$$\varphi'(x) = (E - V(x))\sin^2[\varphi(x)] + \cos^2[\varphi(x)] \qquad [5]$$

Note that [5] does not involve $R$ and that the boundary conditions on $\psi$ and $\psi'$ at $a$ and $b$ can be easily expressed in terms of $\varphi(a)$ and $\varphi(b)$. Moreover, having determined $\varphi$ on $[a,b]$ from [5], the function $R$ is immediately obtained by integrating $R'(x) = [1 + V(x) - E]\sin[\varphi(x)]\cos[\varphi(x)]$. In case of a bounded interval, $-\infty < a < b < \infty$, or a confining potential, $\lim_{x\to\pm\infty} V(x) = \infty$, it is not difficult to derive from [5] the following basic facts: the spectrum of $H(V)$ consists only of simple eigenvalues $E_0 < E_1 < E_2 < \cdots$ with $\lim_{n\to\infty} E_n = \infty$. Moreover, the corresponding eigenfunction $\psi_n \neq 0$, $n \in \mathbb{N}_0$, with $H(V)\psi_n = E_n\psi_n$, has precisely $n$ zeros, and Sturm's oscillation theorem holds.

See Amrein *et al.* (2005) for more details.

## Quantization and Semiclassical Limit

The quantization procedure postulated by Schrödinger is the replacement of the classical momentum $p \in \mathbb{R}^d$ by the quantum-mechanical momentum operator $-i\hbar\nabla_x$. It is known (and, in fact, easy to see, cf. Messiah (1962)) that the classical Hamiltonian equation of motions is invariant under symplectic transformations, but Schrödinger's quantization procedure does not commute with symplectic changes of the classical variables. The question of the geometrically sound definition of quantization, with a general $d$-dimensional manifold replacing the spatial domain $\Omega$, has attracted many mathematicians and has led to the mathematical fields of geometric quantization and deformation quantization.

It is remarkable, however, that Schrödinger himself discovered already in his early paper the fact that classical dynamics derives as the scaling limit $\hbar \to 0$ from quantum mechanics. The systematic study of the convergence of wave functions and of operators and their spectral properties is known as semiclassical analysis, which is nowadays considered to be part of microlocal analysis. We illustrate the type of results one obtains by the following example on $\Omega = \mathbb{R}^d$.

Let $F \in C_0^\infty(\mathbb{R};\mathbb{R})$ be a smooth characteristic function, compactly supported in an interval $I \subset \mathbb{R}^-$ away from the essential spectrum of the semiclassical Schrödinger operator $H_\hbar = -\hbar^2\Delta + V$ with a smooth potential $V \in C_0^\infty(\mathbb{R}^d)$ of compact support. We define the operator $F[H_\hbar]$ by functional calculus (note that $I \subset \sigma_d(H_V)$ and $F[H_\hbar]$ is of trace class).

Let, furthermore, $A_\hbar = \sum_{|\alpha|\leq M} a_\alpha(x)\partial_x^\alpha$ be a differential operator representing an observable. Then $\mathrm{tr}\{A_\hbar F[H_\hbar]\}$, which exists because the eigenfunctions of $H_\hbar$ are smooth and decay exponentially, is, up to normalization, interpreted to be the expectation of the observable $A_\hbar$ in the state represented by the spectral projection of $H_\hbar$ in $I$, approximated by $F[H_\hbar]$.

Semiclassical analysis then yields an asymptotic expansion of the form

$$\mathrm{tr}\{A_\hbar F[H_\hbar]\} = \hbar^{-d}(c_0 + c_1\hbar + \cdots + c_n\hbar^n + o(\hbar^n))$$

for arbitrarily large integers $n \in \mathbb{N}$. The leading-order coefficient $c_0$ is determined by Bohr's correspondence principle,

$$\begin{aligned}\mathrm{tr}\{A_\hbar\, F[H_\hbar]\} &= \int_{\mathbb{R}^{2d}} a[x,p]F[p^2 + V(x)]\frac{\mathrm{d}p\,\mathrm{d}y}{(2\pi\hbar)^d} \\ &\quad + o\left((2\pi\hbar)^{-d}\right)\end{aligned} \qquad [6]$$

Semiclassical analysis thus provides the mathematical link between quantum and classical mechanics. The proof of [6] usually involves pseudodifferential and/or Fourier integral operators, depending on the method. Advanced topics in semiclassical analysis studied more recently are the construction of quasimodes, that is, wave functions $\psi_{E,\hbar,n}$ which solve the eigenvalue problem $(H_\hbar - E)\psi_{E,\hbar,n} = O(\hbar^n)$ up to errors of order $\hbar^n$, for arbitrarily large $n \in \mathbb{N}$, and the relation between semiclassical asymptotics

and the KAM (Kolmogorov–Arnold–Moser) theory from classical mechanics.

For more details, see Dimassi and Sjöstrand (1999), and Robert (1987). *See also* Stability Theory and KAM, KAM Theory and Celestial Mechanics in this encyclopedia.

## Lieb–Thirring Inequalities

Lieb–Thirring inequalities are estimates on eigenvalue sums of $H_{-V} = -\Delta - V(x)$, where $V \geq 0$ is assumed to be non-negative (note that we changed the sign of $V$) and vanishing at $\infty$; the most important examples for these sums are the number of eigenvalues below a given $-E \leq 0$ and the sum of its negative eigenvalues, counting multiplicities. More generally, denoting by $[\lambda]_+ := \max\{\lambda, 0\}$ the positive part of $\lambda \in \mathbb{R}$, Lieb–Thirring inequalities are estimates on $\mathrm{tr}\{[-E - H_{-V}]_+^\gamma\}$, for $\gamma \geq 0$. The number of eigenvalues below $-E$ is then obtained in the limit $\gamma \to 0$, and the sum of the negative eigenvalues corresponds to $E = 0$ and $\gamma = 1$. We henceforth assume $E = 0$, for simplicity. A guess inspired by [6] with $F[\lambda] := [-\lambda]_+^\gamma$, $A = 1$, and $\hbar = 1$ then is that $\mathrm{tr}\{[-H_{-V}]_+^\gamma\}$ is approximately given by

$$\int_{\mathbb{R}^{2d}} [V(x) - p^2]_+^\gamma \frac{\mathrm{d}^d x\, \mathrm{d}^d p}{(2\pi)^d} = C_{\mathrm{SC}}(\gamma, d) \int_{\mathbb{R}^d} V(x)^{(d/2)+\gamma}\, \mathrm{d}^d x \qquad [7]$$

for a suitable constant $C_{\mathrm{SC}}(\gamma, d) > 0$ depending only on $\gamma$ and $d$ (but not on $V$). While this guess is wrong, it is nevertheless a useful guiding principle. Namely, in a rather large range of $\gamma$ and $d$, there exist constants $C_{\mathrm{LT}}(\gamma, d) > 0$ such that

$$\mathrm{tr}\{[-H_{-V}]_+^\gamma\} \leq C_{\mathrm{LT}}(\gamma, d) \int_{\mathbb{R}^d} V(x)^{(d/2)+\gamma}\, \mathrm{d}^d x \qquad [8]$$

for all $V \geq 0$, for which the right-hand side is finite (with the understanding that this finiteness also insure that $[-H_{-V}]_+^\gamma$ is trace class, in the first place).

Of course, $C_{\mathrm{LT}}(\gamma, d) \geq C_{\mathrm{SC}}(\gamma, d)$, by [6]. The Lieb–Thirring conjecture, which is still open today, says that the best possible choice of $C_{\mathrm{LT}}(1, 3)$ equals $C_{\mathrm{SC}}(1, 3)$ in the physically most relevant case $\gamma = 1$ and $d = 3$. It is known that $C_{\mathrm{LT}}(\gamma, d) > C_{\mathrm{SC}}(\gamma, d)$, for $\gamma < 1$ or $d < 3$.

Lieb–Thirring estimates have been derived for various modifications of the original model, depending on the application. One of these are pseudorelativistic Hamiltonians of the form $H = T(p) - V$, where $T(p) = \sqrt{p^2 + m^2}$, with $m \geq 0$, another one includes an external magnetic field, for example, $H = (p - A)^2 - V$ (see the next and the last section).

The reader is referred to Thirring (1997), Reed and Simon (1978), and Simon (1979) for further details.

## Magnetic Schrödinger Operators

Magnetic Schrödinger operators are Hamiltonians of the form

$$H_{\mathrm{mc}}(\boldsymbol{A}, V) = (\boldsymbol{p} - \boldsymbol{A}(\boldsymbol{x}))^2 - V(x) \quad \text{on } L^2(\mathbb{R}^3) \qquad [9]$$

or

$$H_{\mathrm{Pauli}}(\boldsymbol{A}, V) = [\boldsymbol{\sigma} \cdot (\boldsymbol{p} - \boldsymbol{A}(\boldsymbol{x}))]^2 - V(\boldsymbol{x}) \quad \text{on } L^2(\mathbb{R}^3) \otimes \mathbb{C}^2 \qquad [10]$$

where $V$ is the (electrostatic) potential; as before, $\boldsymbol{A} : \mathbb{R}^3 \to \mathbb{R}^3$ is the vector potential of the magnetic field $B = \boldsymbol{\nabla} \wedge \boldsymbol{A}$, and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ are the Pauli matrices. $H_{\mathrm{mc}}(\boldsymbol{A}, V)$ and $H_{\mathrm{pauli}}(\boldsymbol{A}, V)$ generate the dynamics of a particle moving in an external electromagnetic field of spin $s = 0$ and spin $s = 1/2$, respectively. The operator $H_{\mathrm{Pauli}}(\boldsymbol{A}, V)$ is usually called Pauli Hamiltonian, and we refer to $H_{\mathrm{mc}}(\boldsymbol{A}, V)$ as the magnetic Hamiltonian. To keep the exposition simple, we assume henceforth that $A_\nu$ and $\partial_\mu A_\nu$ are uniformly bounded, which suffices to prove the self-adjointness of both Hamiltonians.

At a first glance, the magnetic and the Pauli Hamiltonians may seem to differ only marginally, but in fact, some of their spectral properties are fundamentally different.

1. The magnetic Hamiltonian fulfills the diamagnetic inequality, $|\mathrm{e}^{-\beta H_{\mathrm{mc}}(\boldsymbol{A}, V)}(x, y)| \leq \mathrm{e}^{-\beta H_{\mathrm{mc}}(0, V)}(x, y)$, for almost all $x, y \in \mathbb{R}^3$, where $m(x, y)$ denotes the integral kernel of an operator $m$. As a consequence, $\inf \sigma[H_{\mathrm{mc}}(\boldsymbol{A}, V)] \leq \inf \sigma[H_{\mathrm{mc}}(0, V)] = \inf \sigma[H(V)]$, and the quadratic form of the magnetic Hamiltonian is semibounded, for all choices of $\boldsymbol{A}$, provided $H(V)$ is.
2. If $\inf \sigma[H_{\mathrm{mc}}(\boldsymbol{A}, V)]$ is an eigenvalue, the diamagnetic inequality reflects the fact that the corresponding eigenvector is not positive or of constant phase. The determination of the nodal set of eigenfunctions is a difficult task on its own.
3. For $V = 0$, the diamagnetic inequality and the minimax principle imply that $\boldsymbol{p} - \boldsymbol{A}$ has no zero eigenvalue.
4. The diamagnetic inequality fails to hold for the Pauli Hamiltonian. On the contrary, if $\boldsymbol{A}$ is carefully adjusted in $H_{\mathrm{mc}}(\boldsymbol{A}, -Z|x|^{-1})$, and $Z$ is sufficiently large, then the corresponding

quadratic form may assume arbitrarily small values (even if the corresponding field energy is added).

5. For many choices of $\boldsymbol{A}$, the (Dirac) operator $\boldsymbol{\sigma}\cdot(\boldsymbol{p}-\boldsymbol{A})$ has a nontrivial kernel.

From (1)–(4) it is clear that the proof of stability of matter (see the next section) in presence of a magnetic field is more difficult than in absence of it. This can be illustrated by the fact that magnetic Lieb–Thirring inequalities, being the natural analog of eqn [8], are more involved to derive than the original estimate [8]. The currently best bound is of the form

$$\begin{aligned}\mathrm{tr}\{[-H_{-V}]_+^{\gamma}\} &\\ \leq C_{\mathrm{mLT}}\int_{\mathbb{R}^d}\Big\{&[V(x)]_+^{5/2}+|B(x)|\,[V(x)]_+^{3/2}\\ &+\Big(|B(x)|+L_c(x)^{-2}\Big)L_c(x)^{-1}[V(x)]_+\Big\}\,\mathrm{d}^dx\end{aligned}\qquad [11]$$

for some universal $C_{\mathrm{mLT}}<\infty$, where $L_c(x)$ is a local length scale associated with $B$. It is nonlocal in $x$ and somewhat reminiscent of a maximal function.

We further remark that if restricted to two dimensions, $d=2$, both the magnetic and the Pauli Hamiltonians play an important role in the theory of the (integer) quantum Hall effect.

For more details, see Simon (1979), Cycon *et al.* (1987), Rauch and Simon (1997), and Erdös and Solovej (2004). See also the article Quantum Hall Effect in this encyclopedia.

## *N*-Body Schrödinger Operators

The origin of quantum mechanics is atomic ($K=1$ below) or molecular ($K\geq 2$) physics. If we regard the nuclei of the molecule as fixed point charges $\underline{Z}:=(Z_1,\ldots,Z_K)>0$ at respective positions $\underline{R}:=(R_1,\ldots,R_K)\in\mathbb{R}^3$, then the Hamiltonian (in convenient units) of this molecule with $N\in\mathbb{N}$ electrons is the following Schrödinger operator:

$$\begin{aligned}H_N(\underline{Z},\underline{R}) &=\sum_{n=1}^{N}\left\{-\Delta_n-\sum_{k=1}^{K}\frac{Z_k}{|x_n-R_k|}\right\}\\ &\quad+\sum_{1\leq m<n\leq N}\frac{1}{|x_m-x_n|}\end{aligned}\qquad [12]$$

defined on $\mathcal{H}^{(N)}:=\bigwedge_{n=1}^{N}L^2[\mathbb{R}^3\times\mathbb{Z}_2]\subseteq L^2[(\mathbb{R}^3\times\mathbb{Z}_2)^N]$, the space of totally antisymmetric, square-integrable wave functions in $N$ space–spin variables $(x_1,\sigma_1),\ldots,(x_N,\sigma_N)\in\mathbb{R}^3\times\mathbb{Z}_2$. The antisymmetry of the wave function accounts for the fact that electrons are fermions and is of crucial importance. Note that the number $N$ of electrons is possibly very large. It is clear that we cannot expect to carry out the spectral analysis of this Schrödinger operator directly, but rather only suitable approximations.

In spite of the fact that $H_N(\underline{Z},\underline{R})$ was one of the basic operators of quantum mechanics from its very beginning in the late 1920s, $H_N(\underline{Z},\underline{R})$ was, strictly speaking, not known to be self-adjoint before Kato developed the perturbation theory (described in the section "Self-adjointness") some 20 years later, which then also yielded the semiboundedness of $H_N(\underline{Z},\underline{R})$. So, the ground-state energy $E_N(\underline{Z},\underline{R}):=\inf\sigma[H_N(\underline{Z},\underline{R})]>-\infty$ is finite. From the HVZ (Hunziker–van Winter–Zishlin) theorem follows that $\inf\sigma_{\mathrm{ess}}[H_N(\underline{Z},\underline{R})]=E_{N-1}(\underline{Z},\underline{R})$, which particularly implies that $E_N(\underline{Z},\underline{R})$ is monotonically decreasing in $N$ and negative (because $E_1(\underline{Z},\underline{R})<0$).

It is known that $E_N(\underline{Z},\underline{R})=E_{N+1}(\underline{Z},\underline{R})$ and that $H_N(\underline{Z},\underline{R})$ has no eigenvalue, for $N\geq 2Z_{\mathrm{tot}}+1$, where $Z_{\mathrm{tot}}:=\sum_{k=1}^{K}Z_k$ is the total nuclear charge of the atom. On the other hand, it is known that $E_N(\underline{Z},\underline{R})$ is an eigenvalue, provided $N<Z_{\mathrm{tot}}$. Thus, defining $N_{\mathrm{crit}}$ to be the smallest number such that $E_N(\underline{Z},\underline{R})$ is not an eigenvalue, for all $N\geq N_{\mathrm{crit}}$, that is, $N_{\mathrm{crit}}$ is the maximal number of electrons the molecule can bind, we have that $Z_{\mathrm{tot}}\leq N_{\mathrm{crit}}\leq 2Z_{\mathrm{tot}}+1$. In increasing precision, asymptotic neutrality, $N_{\mathrm{crit}}=Z_{\mathrm{tot}}+R(Z_{\mathrm{tot}})$, with $R(Z_{\mathrm{tot}})=o(Z_{\mathrm{tot}})$ and $R(Z)=o(Z^{5/7})$, was shown for atoms and for molecules, respectively. The ionization conjecture states that $N_{\mathrm{crit}}\leq Z_{\mathrm{tot}}+\mathrm{C}$, for some universal constant $C$. It is still open for the full model represented by $H_N(\underline{Z},\underline{R})$, but has been proved in the Hartree–Fock approximation. It has been proved in the Hartree–Fock approximation by Solovej.

The semiboundedness of $H_N(\underline{Z},\underline{R})$, for fixed $\underline{Z},\underline{R}$, and $N$, alone does not rule out a physical collapse of the matter described by $H_N(\underline{Z},\underline{R})$, but the stronger property of stability of matter does. It holds if there exists a constant $C$, possibly depending on $\underline{Z}$, such that

$$E_N(\underline{Z},\underline{R})+\sum_{1\leq k<\ell\leq K}\frac{Z_k\,Z_\ell}{|R_k-R_\ell|}\geq -\mathrm{C}(N+K)\qquad [13]$$

that is, if the ground-state energy plus the repulsive electrostatic energy of the nuclei is bounded below by a constant times the total number $N+K$ of particles in the system. Equation [13] was shown to hold for $H_N(\underline{Z},\underline{R})$.

In connection with stability of matter, Thomas–Fermi theory and the question of the limit of large nuclear charge came into the focus of research. For simplicity, we restrict ourselves to atoms, $K=1$, that is, there is one nucleus of charge $Z:=Z_1$ at the origin, $R_1=0$, and we consider $E(Z):=\min_{N\in\mathbb{N}}E_N(Z,0)$ (which amounts to fixing $N:=N_{\mathrm{crit}}$). An asymptotic expansion for $E(Z)$ of increasing

precision in $Z$ was obtained by ever-finer estimates; presently, one knows that

$$E(Z) = E_{\mathrm{TF}} Z^{7/3} + \tfrac{1}{4} Z^2 + C_{\mathrm{DS}} Z^{5/3} + o(Z^{5/3}) \qquad [14]$$

where the leading contribution $E_{\mathrm{TF}} Z^{7/3}$ is the Thomas–Fermi energy, $(1/4)Z^2$ is the Scott correction, and $C_{\mathrm{DS}} Z^{5/3}$ is the Dirac–Schwinger term. The computation of this last term requires semiclassical analysis sketched in the section "Quantization and semiclassical limit."

For more details, see Cycon *et al.* (1987), Rauch and Simon (1997), Thirring (1997), and Solovej (2003). See also the article Stability of Matter in this encyclopedia.

## Scattering Theory

The study of the properties of the propagator $\exp(-\mathrm{i}tH)$ of a self-adjoint operator $H = H^*$, as $t \to \infty$, is the concern of scattering theory. To obtain a well-defined mathematical object in this limit, it is necessary to compose $\exp(-\mathrm{i}tH)$ with the inverse of some explicitly accessible comparison dynamics before passing to the limit $t \to \infty$. If $V$ is a short-range potential, that is, $V$ is relatively $H_0$-compact and $|V(x)| \le C|x|^{-\nu}$, for some $\nu > 1$ and $C < \infty$, then the comparison dynamics appropriate for $H_V$ is generated by $H_0$: the wave operators $\Omega^{\pm}$ are defined as the strong limits

$$\Omega^{\pm} := \lim_{t \to \pm\infty} \mathrm{e}^{\mp \mathrm{i}tH_V} \mathrm{e}^{\pm \mathrm{i}tH_0} \qquad [15]$$

A general technique in scattering theory to prove the existence of such limits is Cook's argument, which formally amounts to an application of the fundamental theorem of calculus. For example, for the existence of $\Omega^+$, one writes

$$\begin{aligned} \Omega^+ - \mathbb{1} &= \int_0^\infty \mathrm{d}t \left\{ \frac{\mathrm{d}}{\mathrm{d}t} \left( \mathrm{e}^{-\mathrm{i}tH_V} \mathrm{e}^{\mathrm{i}tH_0} \right) \right\} \\ &= -\mathrm{i} \int_0^\infty \mathrm{d}t \, \{ \mathrm{e}^{-\mathrm{i}tH_V} \, V \, \mathrm{e}^{\mathrm{i}tH_0} \} \end{aligned} \qquad [16]$$

and additionally proves the absolute integrability of $t \mapsto \mathrm{e}^{-\mathrm{i}tH_V} V \mathrm{e}^{\mathrm{i}tH_0} \varphi$, for $\varphi$ in a dense subset of $\mathcal{H}$, like $\mathrm{dom}(H_0) = \mathrm{dom}(H_V)$.

Research in scattering theory in the past two decades or so was focused around the question of asymptotic completeness, which is a mathematically precise formulation

$$\mathrm{Ran}\Omega^+ = \mathrm{Ran}\Omega^- = \mathcal{H}_{\mathrm{pp}}^{\perp}(H_V) \qquad [17]$$

of the physical expectation that the states in $\mathcal{H}$ are either bound states (eigenvectors) of $H_V$ or scattering states (states in the range of $\Omega^{\pm}$) of $H_V$. The intertwining property $H_V \Omega^{\pm} = \Omega^{\pm} H_0$ (which easily follows from [15]) implies that the restriction of $H_V$ to $\mathrm{Ran}\Omega^{\pm}$ is unitarily equivalent to $H_0$, hence $\mathrm{Ran}\Omega^{\pm} \subseteq \mathcal{H}_{\mathrm{ac}}(H_V) \subseteq \mathcal{H}_{\mathrm{pp}}^{\perp}(H_V)$. The difficult part of the proof of asymptotic completeness is to show that $\mathcal{H}_{\mathrm{pp}}^{\perp}(H_V) \subseteq \mathrm{Ran}\Omega^{\pm}$.

Much effort has been spent to prove asymptotic completeness for $N$-body Schrödinger operators on $\mathcal{H}^{(N)} := \bigotimes_{n=1}^N L^2(\mathbb{R}^3)$ of the form

$$\begin{aligned} H_N(V) &= \sum_{n=1}^N \frac{-\Delta_n}{2\, m_n} + V(x) \\ &\text{with } V(x) := \sum_{1 \le m < n \le N} V_{mn}(x_m - x_n) \end{aligned} \qquad [18]$$

where each pair potential $V_{mn}$ obeys $|\partial_y^{\alpha} V_{mn}(y)| \le C(1 + |y|)^{-\mu - |\alpha|}$, with $\alpha \in \mathbb{N}_0^d$ being a multi-index. If $\mu > 1$ for all $m \ne n$ then $V$ is called a short-range potential. Conversely, if $0 < \mu \le 1$ then $V$ is a long-range potential. Note that even though each $V_{mn}$ decays at infinity, $|x|^2 = x_1^2 + x_2^2 + \cdots + x_n^2 \to \infty$ alone does not imply that $V(x) \to \infty$. In fact, physical intuition tells us that for a cluster $\mathcal{C}$ of $N$ particles, whose dynamics is generated by $H_N(V)$, several scenarios for the long-time asymptotic behavior of the evolution are possible:

1. The $N$ particles stay together in their cluster $\mathcal{C}$ whose center of mass moves in space at constant velocity.
2. The cluster breaks up into two (or even more) subclusters, $\mathcal{C}_1$ and $\mathcal{C}_2$, of $N_1$ and $N_2 = N - N_1$ particles, respectively, whose centers of mass drift apart from each other at constant velocities (in the short-range case). For each subcluster $\mathcal{C}_1$ and $\mathcal{C}_2$, both scenarios may appear again, after waiting sufficiently longer.
3. In the limit $t \to \infty$, possibly after going through (1) and (2) several times, the initial cluster $\mathcal{C}$ is broken up into $1 \le K \le N$ subclusters $\mathcal{C}_1, \ldots, \mathcal{C}_K$, whose centers of mass drift apart from each other at constant velocities according to a free and independent dynamics of their centers of mass.

In some sense, asymptotic completeness says that nothing else than (1)–(3) can possibly happen. (Strictly speaking, asymptotic completeness is a statement about the limit $t \to \infty$ and only involves (3) – the actual behavior of $\exp[-itH_V]$ at intermediate times in terms of (1)–(3) is beyond the reach of current mathematics.) It is a key insight of scattering theory that the asymptotics of the time evolution in the sense of (3) is completely

characterized by the asymptotic velocity defined by the strong limit

$$P^+ := \lim_{t\to\infty}\left(\mathrm{e}^{-\mathrm{i}tH_N(V)}\frac{x}{t}\mathrm{e}^{\mathrm{i}tH_N(V)}\right) \qquad [19]$$

It is a nontrivial fact that $P^+$ exists, commutes with $H_N(V)$, and that bound states are precisely the states with zero asymptotic velocity, while states with nonzero asymptotic velocity are scattering states in $\mathrm{Ran}\Omega^{\pm}$. This then implies asymptotic completeness for short-range potentials. The proof of this dichotomy builds essentially upon positive commutator or Mourre estimates. Given an interval $J$ localized (in energy) away from any eigenvalue of any possible subcluster configuration $\mathcal{C}_1, \ldots, \mathcal{C}_K$ (called thresholds), the Mourre estimate asserts the existence of a positive constant $M > 0$ and a compact operator $R \in \mathcal{B}(\mathcal{H}^{(N)})$ such that

$$\mathbb{1}_{\mathcal{J}}\,\mathrm{i}[H_N(V), A]\,\mathbb{1}_{\mathcal{J}} \geq M\mathbb{1}_{\mathcal{J}} - R \qquad [20]$$

as a quadratic form, for some suitable operator $A$. This operator $A$ is often chosen to be the dilation generator $A = (1/2)\{p\cdot x + x\cdot p\}$ or a variant thereof.

Again, the proof of asymptotic completeness for long-range potentials is still more difficult and has been carried out only for $\mu > \sqrt{3} - 1$. The additional problem is the comparison dynamics of the relative motion of the clusters $\mathcal{C}_1$ and $\mathcal{C}_2$ in (2), which is not the free one; the clusters rather influence each other even at large distances.

For more details, see Reed and Simon (1980c) and Derezinski and Gérard (1997). See also the articles Scattering in Relativistic Quantum Field Theory: Fundamental Concepts and Tools, Scattering, Asymptotic Completeness and Bound States in this encyclopedia.

## Random Schrödinger Operators

Schrödinger operators $H(V_\omega)$ on $L^2(\mathbb{R}^d)$ or $\ell^2(\mathbb{Z}^d)$ with a random potential $V_\omega$ are called random Schrödinger operators. (If $H(V_\omega)$ acts on $\ell^2(\mathbb{Z}^d)$, then the (continuum) Laplacian $-\Delta$ is replaced by the discrete Laplacian on $\mathbb{Z}^d$ defined by $[-\Delta_{\mathrm{disc}} f](x) = \sum_{\nu=1}^{d}\{2f(x) - f(x - e_\nu) - f(x + e_\nu)\}$.) More precisely, given a probability space $(\Omega, P, \mu)$ and a random variable $\Omega \ni \omega \mapsto V_\omega$, the family $\{H(V_\omega)\}_{\omega\in\Omega}$ defines an operator-valued random variable that we refer to as a random Schrödinger operator. Random quantum systems are physically relevant as models for amorphous materials, and for solids in very heterogenous external fields or coupled to quantized fields. Suitable ergodicity assumptions on $\omega \to V_\omega$ ensure that the domain of $H_\omega$ and even many spectral properties (in particular, the spectrum $\sigma(H(V_\omega)) \subseteq \mathbb{R}$ itself) are independent of $\omega$ $P$-almost surely. For example, assuming an independent, identical distribution (i.i.d.) of $V_\omega$ in the discrete case on $\mathbb{Z}^d$, one arrives at the Anderson model, which has been most thoroughly studied. Its counterpart for continuum models is a Poisson-distributed $V_\omega$. A model which also has ergodic properties, although deterministic, is the Hofstadter or the Mathieu problem. Most research has been focused on localization, that is, spatial decay properties of the resolvent $\{H(\lambda V_\omega) - E\}^{-1}(x, y)$ of $H(\lambda V_\omega)$, as $|x - y| \to \infty$, and particularly the question of presence or absence of exponential decay (localization), as this is an important indicator for the transport properties of the material under consideration. Exponential localization of eigenstates has been established for $d = 1$ or strong disorder or sufficiently high energies $E \gg 1$. Localization is also intimately related to bounds on moments of the form $\|x^{\mu/2}\psi_t\| \leq C_\mu t^\beta$. The study of the asymptotic distribution of eigenvalues close to the lowest threshold leads to the so-called Lifshitz tails.

The reader is referred to Figotin and Pastur (1992), Cycon *et al.* (1987), and Stollmann (2001).

## (Pseudo)relativistic Schrödinger Operators

Schrödinger operators of the form $H(V) = p^2 + V(x)$ do not observe the invariance principles of (special) relativity, as their derivation is based in classical (Newtonian) mechanics. The free Dirac operator $D := \boldsymbol{\alpha}\cdot\boldsymbol{p} + m\beta$ (here, $\alpha_\nu$ and $\beta$ are self-adjoint $4\times 4$ matrices) possesses the desired relativistic invariance, but it is not semibounded, and the definition of an interacting Dirac operator is notoriously difficult (and unsolved). The replacement of the kinetic energy $(1/2m)p^2$ by the Klein–Gordon operator $\sqrt{p^2 + m^2}$ is a step towards relativistic invariance, which, at the same time, yields a positive operator. This replacement may also be viewed as the restriction of the free Dirac operator to its positive-energy subspace. The virtue of this replacement is that it immediately allows for the study of interacting $N$-particle operators,

$$H_N^{\mathrm{rel}}(\underline{Z}, \underline{R}) = \sum_{n=1}^{N}\left\{\sqrt{-\Delta_n + m^2} - \sum_{k=1}^{K}\frac{Z_k}{|x_n - R_k|}\right\} + \sum_{1\leq \ell < n \leq N}\frac{1}{|x_\ell - x_n|} \qquad [21]$$

much like in [12]. Since $\sqrt{p^2 + m^2} \sim |p|$, as $p \to \infty$, the pseudorelativistic kinetic energy $\sqrt{p^2 + m^2}$ can

balance only less severe local singularities of the potential $V$ than the nonrelativistic kinetic energy $(1/2m)p^2$. Indeed, already the quadratic form $\sqrt{p^2+m^2} - g|x|^{-1}$ on $C_0^\infty(\mathbb{R}^3)$ associated to a hydrogen-like atom is unbounded from below if $g > 2/\pi$. Hence, the stability of matter becomes a more subtle property of pseudorelativistic matter. The relaxation of the restriction onto the positive subspace of the free Dirac operator also got into the focus of research.

For more details, we refer the reader to Thirring (1997).

*See also:* Deformation Quantization; Elliptic Differential Equations: Linear Theory; *h*-Pseudodifferential Operators and Applications; Localization for Quasiperiodic Potentials; Nonlinear Schrödinger Equations; Normal Forms and Semiclassical Approximation; *N*-Particle Quantum Scattering; Quantum Hall Effect; Quantum Mechanical Scattering Theory; Scattering, Asymptotic Completeness and Bound States; Stability of Matter; Stationary Phase Approximation.

## Further Reading


Amrein W, Hinz A, and Pearson D (2005) *Sturm–Liouville Theory – Past and Present*. Boston: Birkhäuser.

Cycon H, Froese R, Kirsch W, and Simon B (1987) *Schrödinger Operators*, 1st edn. Berlin: Springer.

Derezinski J and Gérard C (1997) *Scattering Theory of Classical and Quantum N-Particle Systems*, Text and Monographs in Physics. Berlin: Springer-Verlag.

Dimassi M and Sjöstrand J (1999) *Spectral Asymptotics in the Semi-Classical Limit*. London Mathematical Society Lecture Notes Series, vol. 268. Cambridge: Cambridge University Press.

Erdös L and Solovej JP (2004) Uniform Lieb–Thirring inequality for the three-dimensional Pauli operator with a strong non-homogeneous magnetic field. *Annales Henri Poincaré* 5: 671–741.

Figotin A and Pastur L (1992) *Spectra of Random and Almost-Periodic Operators*. Grundlehren der Mathematischen Wissenschaften, vol. 297. Berlin: Springer-Verlag.

Kato T (1976) *Perturbation Theory of Linear Operators*, 2 edn., Grundlehren der mathematischen Wissenschaften, vol. 132. Berlin: Springer-Verlag.

Messiah A (1962) *Quantum Mechanics*, 1st edn., vol. 2. Amsterdam: North-Holland.

Rauch J and Simon B (eds.) (1997) *Quasiclassical Methods*. IMA Volumes in Mathematics and Its Applications, vol. 95. Berlin: Springer-Verlag.

Reed M and Simon B (1978) *Methods of Modern Mathematical Physics IV. Analysis of Operators*, 1st edn., vol. 4. San Diego: Academic Press.

Reed M and Simon B (1980a) *Methods of Modern Mathematical Physics: I. Functional Analysis*, 2nd edn., vol. 1. San Diego: Academic Press.

Reed M and Simon B (1980b) *Methods of Modern Mathematical Physics: II. Fourier Analysis and Self-Adjointness*, 2nd edn., vol. 2. San Diego: Academic Press.

Reed M and Simon B (1980c) *Methods of Modern Mathematical Physics: III. Scattering Theory*, 2nd edn., vol. 3. San Diego: Academic Press.

Robert D (1987) *Autour de l'Approximation Semi-Classique*, 1st edn. Boston: Birkhäuser.

Schrödinger E (1926) Quantisierung als Eigenwertproblem. *Annalen der Physik* 79: 489.

Simon B (1979) *Functional Integration and Quantum Physics*, Pure and Applied Mathematics. New York: Academic Press.

Simon B (2000) Schrödinger operators in the twentieth century. *Journal of Mathematical Physics* 41: 3523–3555.

Solovej JP (2003) The ionization conjecture in Hartree–Fock theory. *Annals of Mathematics* 158: 509–576.

Stollmann P (2001) *Caught by Disorder*. Progress in Mathematical Physics, vol. 20. Boston: Birkhäuser.

Thirring W (ed.) (1997) *The Stability of Matter: From Atoms to Stars – Selecta of Elliott H. Lieb*, 2 edn. Berlin: Springer-Verlag.


# Schwarz-Type Topological Quantum Field Theory


**R K Kaul and T R Govindarajan**, The Institute of Mathematical Sciences, Chennai, India

**P Ramadevi**, Indian Institute of Technology Bombay, Mumbai, India




## Introduction

Topological quantum field theories (TQFTs) provide powerful tools to probe topology of manifolds, specifically in low dimensions. This is achieved by incorporating very large gauge symmetries in the theory which lead to gauge-invariant sectors with only topological degrees of freedom. These theories are of two kinds: (1) Schwarz type and (2) Witten type.

In a Witten-type topological field theory, action is a BRST exact form, so is the stress energy tensor $T_{\mu\nu}$ so that their functional averages are zero (Witten 1988). The BRST charge is associated with a certain shift symmetry. The topological observables form cohomological classes and semiclassical approximation turns out to be exact. In four dimensions, such theories involving Yang–Mills gauge fields provide a field-theoretic representation for Donaldson invariants.

On the other hand, Schwarz-type TQFTs are described by local action functionals which are not total derivatives but are explicitly independent of metric (Schwarz 1978, 1979, 1987, Witten 1989).

The examples of such theories are topological Chern–Simons (CS) theories and BF theories.

Metric independence of the action $S$ of a Schwarz-type gauge theory implies that stress–energy tensor is zero:

$$\frac{\delta S}{\delta g_{\mu\nu}} \equiv T_{\mu\nu} = 0$$

More generally, in the gauge-fixed version of such theories, stress–energy can be BRST exact, where BRST charge corresponds to gauge fixing in contrast to Witten-type theories where corresponding BRST charge corresponds to a combination of shift symmetry and gauge symmetry. There are no local propagating degrees of freedom; the only degrees of freedom are topological. Expectation values of metric-independent operators $W$ are also independent of the metric:

$$\frac{\delta\langle W\rangle}{\delta g_{\mu\nu}} = 0$$

Three-dimensional CS theories are of particular interest, for these provide a framework for the study of knots and links in any 3-manifold. Pioneering indications of the fact that topological invariants can be found in such a setting came in very early when A S Schwarz demonstrated that a particular topological invariant, Ray–Singer analytic torsion (which is equivalent to combinatorial Reidemeister–Franz torsion) can be interpreted in terms of the partition function of a quantum gauge field theory (Schwarz 1978, 1979). In particular, in the weak-coupling limit of CS theory of gauge group $\mathcal{G}$ on a manifold $\mathcal{M}$, contribution from each topologically distinct flat connection (characterized by the equivalence classes of homomorphisms: $\pi_1(\mathcal{M}) \rightarrow \mathcal{G}$) to the partition function is given by metric-independent Ray–Singer torsion of the flat connection up to a phase. This phase factor is also a topological invariant of framed 3-manifold $\mathcal{M}$ (Witten 1989). It was Schwarz who first discussed CS theory as a topological field theory and also conjectured that the well-known Jones polynomial may be related to it (Schwarz 1987). In his famous paper Witten (1989) not only demonstrated this connection, but also set up a general field-theoretic framework to study the topological properties of knots and links in any arbitrary 3-manifold. In addition, this framework provides a method of obtaining some new manifold invariants. As discussed by A Achúcaro and P K Townsend, CS theory also describes gravity in three-dimensional spacetime (Carlip 2003).

BF theories in three dimensions provide another framework for field-theoretic description of topological properties of knots and links. These theories with bilinear action in fields can also be defined in higher dimensions. In particular in $D=4$, BF theory, besides describing two-dimensional generalizations of knots and links, also provides a field-theoretic interpretation of Donaldson invariants. This provides a connection of these theories with Witten-type TQFTs of Yang–Mills gauge fields. We shall not discuss BF theories in the following and refer to the article BF Theories in this Encyclopedia.

Witten (1995) has also formulated CS theories in three complex dimensions described in terms of holomorphic 1-forms. Such a theory on Calabi–Yau spaces can be interpreted as a string theory in terms of a Witten-type topological field theory of a sigma model coupled to gravity. General topological sigma models in Batalin–Vilkovisky formalism have been constructed by Alexandrov *et al.* (1997). This is a Schwarz-type theory. However, in its gauge-fixed version, it can also be interpreted as a Witten-type theory. This construction provides a general formulation from which numerous topological field theories emerge. In particular, the Witten $A$ and $B$ models and also multidimensional CS theories are special cases of this construction.

In the following, we shall survey three-dimensional CS theory as a description of knots/links, indicate how manifold invariants can be constructed from invariants for framed links, and also discuss its application to three-dimensional gravity.

## Three-Dimensional CS Theory with Gauge Group U(1)

The simplest Schwarz-type topological field theory is the U(1) CS theory described by the action:

$$S = -\frac{1}{8\pi}\int_{\mathcal{M}} A\,\mathrm{d}A \qquad [1]$$

where $A$ is a connection 1-form $A = A_\mu \mathrm{d}x^\mu$ and $\mathcal{M}$ is the 3-manifold, which we shall take to be $S^3$ for the discussion below. The action has no dependence on the metric. Besides being the U(1) gauge invariant, it is also general coordinate invariant.

In quantum CS field theory, we are interested in the functional averages of gauge-invariant and metric-independent functionals $W[A]$:

$$\langle W[A]\rangle = \frac{1}{\mathcal{Z}}\int [\mathcal{D}A]\, W[A] \exp\{\mathrm{i}kS\}$$
$$\mathcal{Z} = \int [\mathcal{D}A] \exp\{\mathrm{i}kS\} \qquad [2]$$

This theory captures some of the simple, but interesting, topological properties of knots and links in three dimensions. For a knot $K$, we associate a knot

operator $\oint_K A$ which is gauge invariant and also does not depend on the metric of the 3-manifold. Then for a link made of two knots $K_1$ and $K_2$, we have the loop correlation function $\langle \oint_{K_1} A \oint_{K_2} A \rangle$, which can be evaluated in terms of two-point correlator $\langle A_\mu(x) A_\nu(y) \rangle$ in $R^3$ (with flat metric). This correlator in Lorentz gauge ($\partial_\mu A^\mu = 0$) is:

$$\langle A_\mu(x) A_\nu(y) \rangle = \frac{\mathrm{i}}{k} \epsilon_{\mu\nu\rho} \frac{(x-y)^\rho}{|x-y|^3}$$

so that for two distinct knots $K_1$ and $K_2$

$$\left\langle \oint_{K_1} A \oint_{K_2} A \right\rangle = \frac{4\pi \mathrm{i}}{k} \mathcal{L}(K_1, K_2) \qquad [3]$$

where

$$\mathcal{L}(K_1, K_2) = \frac{1}{4\pi} \oint_{K_1} \oint_{K_2} \mathrm{d}x^\mu \mathrm{d}y^\nu \epsilon_{\mu\nu\rho} \frac{(x-y)^\rho}{|x-y|^3}$$

This integral is the well-known topological invariant called "Gauss linking number" of two distinct closed curves. It is an integer measuring the number of times one knot $K_1$ goes through the other knot $K_2$. Linking number does not depend on the location, size, or shape of the knots. In electrodynamics, it has the physical interpretation of work done to move a monopole around a knot while electric current runs through the other knot.

Abelian CS theory also provides a field-theoretic representation for another topological quantity called "self-linking number," also known as "framing number," of the knot. It is related to the functional average of $\langle \oint_K A \oint_K A \rangle$ where two loop integrals are over the same knot. Coincidence singularity is avoided by a topological loop-splitting regularization. For a knot $K$ given by $x^\mu(s)$ parametrized along the length of the knot by $s$, we associate another closed curve $K_\mathrm{f}$ given by $y^\mu(s) = x^\mu(s) + \epsilon\, n^\mu(s)$, where $\epsilon$ is a small parameter and $n^\mu(s)$ is a principal normal to the curve at $s$. The coincidence limit is then obtained at the end by taking the limit $\epsilon \to 0$. Such a limiting procedure is called framing and knot $K_\mathrm{f}$ is the "frame" of knot $K$. Linking number of the knot $K$ and its frame $K_\mathrm{f}$ is the self-linking number of the knot:

$$S\mathcal{L}(K, n^\mu) = \frac{1}{4\pi} \oint \oint \mathrm{d}x^\mu\, \mathrm{d}y^\nu \frac{\epsilon_{\mu\nu\rho}(x-y)^\rho}{|x-y|^3}$$

Hence coincidence two loop correlator is

$$\left\langle \oint_K A \oint_K A \right\rangle = \frac{4\pi \mathrm{i}}{k} S\mathcal{L}(K, n^\mu) \qquad [4]$$

Notice that the self-linking number of a knot is independent of the regularization parameter $\epsilon$, but does depend on the topological character of the normal vector field $n^\mu(s)$. It is also related to two geometric quantities called "twist" $T(K)$ and "writhe" $w(K)$ through a theorem due to Calugareanu:

$$S\mathcal{L}(K) = T(K) + \omega(K) \qquad [5]$$

where

$$T(K) = \frac{1}{2\pi} \oint_K \mathrm{d}s\, \epsilon_{\mu\nu\rho} \frac{\mathrm{d}x^\mu}{\mathrm{d}s} n^\nu \frac{\mathrm{d}x^\rho}{\mathrm{d}s}$$
$$\omega(K) = \frac{1}{4\pi} \oint_K \mathrm{d}s \oint_K \mathrm{d}t\, \epsilon_{\mu\nu\rho} \frac{\mathrm{d}e^\mu}{\mathrm{d}s} \frac{\mathrm{d}e^\nu}{\mathrm{d}t} e^\rho$$

Here

$$e^\mu(s,t) = \frac{y^\mu(t) - y^\mu(s)}{|y(t) - y(s)|}$$

is a unit map from $K \otimes K \longrightarrow S^2$ and $n^\mu(s)$ is a normal unit vector field. $T(K)$ and $\omega(K)$ are not in general integers and represent the amount of twist and coiling of the knot. These are not topological invariants but their sum, self-linking number, is indeed always an integer and a topological invariant. This result has found interesting applications in the studies of the action of enzymes on circular DNA.

## Nonabelian CS Theories

Nonabelian CS theories provide far more information about the topological properties of the manifolds as well as knots and links.

Nonabelian CS theory in a 3-manifold $\mathcal{M}$ (which as in last section is taken to be $S^3$) is described by the action functional

$$S = \frac{1}{4\pi} \int_{\mathcal{M}} \mathrm{tr}\left(A \wedge \mathrm{d}A + \tfrac{2}{3} A \wedge A \wedge A\right) \qquad [6]$$

where $A$ is a gauge field 1-form which takes its value in the Lie algebra $\mathcal{LG}$ of a compact semisimple Lie group $\mathcal{G}$. For example, we may take this group to be SU($N$) and $A = A^a T^a$, where $T^a$ is the fundamental $N$-dimensional representation with $\mathrm{tr}\, T^a T^b = -1/2\delta^{ab}$. Under homotopically nontrivial gauge transformations this action is not invariant, but changes by an amount $2\pi n$ where integers $n$ are the winding numbers characterizing the gauge transformations which fall in homotopic classes given by $\Pi_3(\mathcal{G}) = \mathcal{Z}$ for a compact semisimple group $\mathcal{G}$. However, for quantum theory what is relevant is $\exp[\mathrm{i}kS]$ which is invariant even under homotopically nontrivial gauge transformations provided the coupling $k$ takes integer values. This quantized nature of the coupling was pointed out by Deser *et al.* (1982a, b) (and also they were first to introduce the nonabelian CS term as a gauge-invariant topological

mass term in gauge theories). So for integer $k$, the quantum field theory we discuss here is gauge invariant.

The topological operators are Wilson loop operators for an oriented knot $K$:

$$W_R[K] = \mathrm{tr}\, P \exp \oint_K A_R \quad [7]$$

where $A_R = A^a T^a_R$ with $T^a_R$ as the representation matrices of a finite-dimensional representation $R$ of the $\mathcal{LG}$. $P$ stands for the path ordering of the exponential. The observable Wilson link operator for a link $L = \bigcup_1^n K_i$, carrying representations $R_i$ on the respective component knots, is

$$W_{R_1 R_2 \cdots R_n}[L] = \prod_1^n W_{R_i}[K_i] \quad [8]$$

Expectation values of these operators are:

$$V_{R_1, R_2 \cdots R_n}[L] = \frac{\int [\mathcal{D}A]\, W_{R_1 \cdots R_n}[L] e^{ikS}}{\int [\mathcal{D}A] e^{ikS}} \quad [9]$$

The measure $[\mathcal{D}A]$ has to be metric independent. These expectation values depend not only on the isotopy of the link $L$ but also on the set of the representations $\{R_i\}$. These can be evaluated in principle nonperturbatively. For example, when $\mathcal{LG} = \mathrm{su}(N)$ and each of the component knot of the links carries the fundamental $N$-dimensional representation, the Wilson link expectation values satisfy a recursion relation involving three link diagrams which are identical except for one crossing where they differ as over crossing ($L_+$), under crossing ($L_-$), and no crossing ($L_0$) as shown in the Figure 1.

The expectation values of these links are related as (Witten 1989):

$$q^{N/2} V_N[L_+] - q^{-N/2} V_N[L_-] = \left(q^{1/2} - q^{-1/2}\right) V_N[L_0] \quad [10]$$

where

$$q = \exp\left(\frac{2\pi i}{k+N}\right)$$

This is precisely the well-known skein relation for the HOMFLY polynomial. The famous Jones one-variable polynomial (whose two-variable generalization is the HOMFLY polynomial) corresponds to the case of spin-1/2 representation of SU(2) CS theory: $V_2[L] =$ Jones polynomial $[L]$, up to an overall normalization. These skein relations are sufficient to recursively find all the expectation values of links with only fundamental representation on the components. To obtain invariants for any other representation, more general methods have to be developed. A complete and explicit solution of the CS field theory is thus obtained. One such method has been reviewed in Kaul (1999). The method makes use of the following important statement:

**Proposition:** *CS theory on a 3-manifold $\mathcal{M}$ with boundary $\Sigma$ is described by a WZNW (Wess–Zumino–Novikov–Witten) conformal field theory (CFT) on the boundary* (Figure 2).

Using the same identification, functional average for Wilson lines ending at $n$ points on the boundary $\Sigma$ is obtained from WZNW field theory on the boundary with $n$ punctures carrying representations $R_i$ (Figure 3):

We can represent CS functional integral as a vector (Witten 1989) in the Hilbert space $\mathcal{H}$ associated with the $n$-point vacuum expectation values of primary fields in WZNW conformal field theory on the boundary $\Sigma$. Next, to obtain a complete and explicit nonperturbative solution of the CS theory, the theory of knots and links and their connection to braids is invoked.

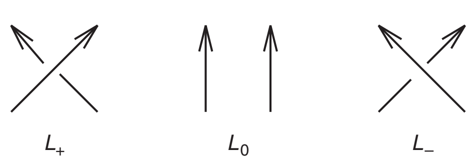



**Figure 1** Skein related links.

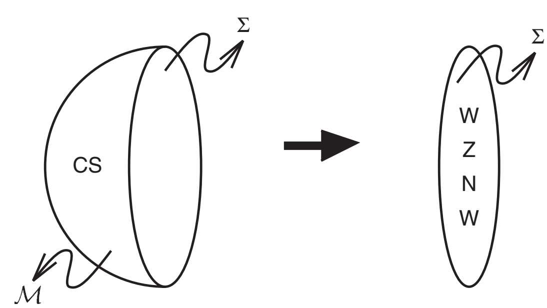



**Figure 2** Relation of CS to CFT.

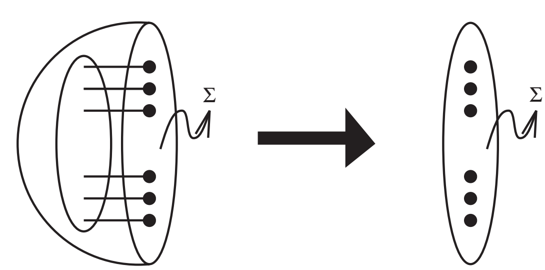



**Figure 3** CS functional integrals with Wilson lines and CFT on punctured boundary.

### Knots/Links and Braids

Braids have an intimate connection with knots and links which can be summarized as follows:

1. An $n$-braid is a collection of nonintersecting strands connecting $n$ points on a horizontal rod to $n$ points on another horizontal rod below strictly excluding any backward traversing of the strands. A general braid can be written as a word in terms of elementary braid generators.
2. We associate representations $R_i$ of the group with the strands as their colors. We also put an orientation on each strand. When all the representations are identical and also all strands are unoriented, we get ordinary braids, otherwise we get colored oriented braids.
3. The colored oriented braids form a groupoid where product of the different braids is obtained by joining them with both colors and orientations matching on the joined strands. Unoriented monochromatic braids form a group.
4. A knot/link can be formed from a given braid by a process called platting. We connect adjacent strands namely the $(2i+1)$th strand to $2i$th strand carrying the same color and opposite orientations in both the rods of an even-strand braid (Figure 4a).

   There is a theorem due to Birman which states that all colored oriented knots/links can be obtained through platting. This construction is not unique.
5. There is another construction associated with braids which relates them to knots and links. We obtain a closure of a braid by connecting the ends of the first, second, third, ... strands from above to those of the respective first, second, third, ... strands from below as shown in the Figure 4b. There is theorem due to Alexander which states that any knot or link can be obtained as a closure of a braid, though again not uniquely.

### Link Invariants

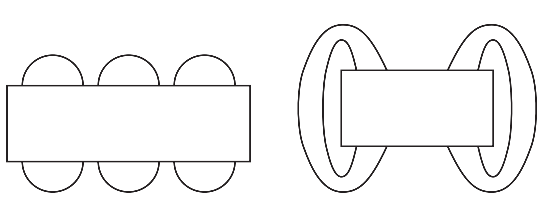
This connection of braids to knots and links can be used to construct link invariants, say in $S^3$. To do so, two nonintersecting 3-balls are removed from the 3-manifold $S^3$ to obtain a manifold with two $S^2$ boundaries. Then we arrange $2n$ Wilson lines of, say SU($N$) CS theory, as a $2n$-strand oriented braid carrying representations $R_i$ in this manifold. The CS functional integral over this manifold is a state in the tensor product of the Hilbert spaces $\mathcal{H}_1 \otimes \mathcal{H}_2$ associated with conformal field theory on the two boundaries. These boundaries have $2n$ punctures carrying the set of representations $\{R_i\}$ and $\{R'_i\}$, respectively, the two sets being permutations of each other. This state can be expanded in terms of some convenient basis given by the conformal blocks for the $2n$-point correlation functions of $\mathrm{SU}(N)_k$ WZNW conformal field theory. The duality of these correlation functions represents the transformation between different bases for the Hilbert space. Their monodromy properties allow us to write down representations of the braid generators. Since an arbitrary braid is just a word in terms of these generators, this construction provides us a matrix representation $\mathcal{B}(\{R_i\},\{R'_j\})$ for the colored oriented braid in the manifold with two $S^2$ boundaries. Then we plat this braid by gluing two balls $B_1$ and $B_2$ with Wilson lines as shown in Figure 5.

Each of the two caps again represents a state $|\psi(\{R_j\})\rangle$ in the Hilbert space associated with the conformal field theory on punctured boundary ($S^2$). Platting of the braid then simply is the matrix element of braid representation $\mathcal{B}(\{R_i\},\{R'_j\})$ with respect to these states $|\psi(\{R_i\})\rangle$ and $|\psi(\{R'_j\})\rangle$ corresponding to two caps $B_1, B_2$. Thus, for a link in $S^3$ the invariant is given by the following theorem:

**Theorem** *The vacuum expectation value of Wilson loop operator of a link L constructed from platting of a colored oriented 2n braid with representation* $\mathcal{B}(\{R_i\},\{R'_j\})$ *is given by* (Kaul 1999):

$$V[L] = \langle \psi(\{R_i\}) | \mathcal{B}(\{R_i\},\{R'_j\}) | \psi(\{R'_j\}) \rangle \qquad [11]$$

This theorem can be used to calculate the invariant for any arbitrary link. For an unknot $U$



(a) (b)

**Figure 4** (a) Platting and (b) closure of braids.

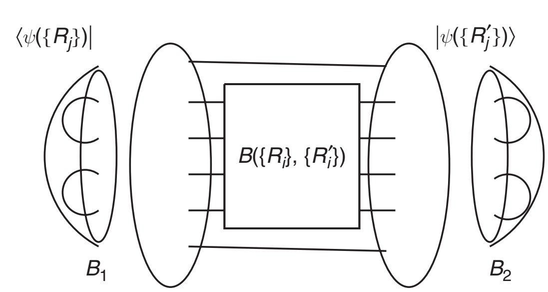



**Figure 5** Construction of the link invariant.

carrying an $N$-dimensional representation in an SU($N$) CS theory, the knot invariant is:

$$V_N[U] = [N], \quad \text{where } [N] = \frac{q^{N/2} - q^{-N/2}}{q^{1/2} - q^{-1/2}}$$

Wilson link expectation values calculated this way depend on the regularization, that is, the definition of framing used in defining coincident loop correlators. One such regularization usually used is the standard framing, where the frame for every knot is so chosen that its self-linking number is zero.

The procedure outlined here has been used for explicit computations of knot/link invariants. This has led to answers to several questions of knot theory. One such question relates to distinguishing chirality of knots (Kaul 1999). In this context, newer invariants constructed with arbitrary representations living on the knots are more powerful than the older polynomial invariants. For example, invariants with spin-3/2 representation in an SU(2) CS theory are sensitive to chirality of many knots which otherwise is not detected by Jones, HOMFLY, and Kauffman polynomials. However, invariants obtained from CS theories do not distinguish all chiral knots. There is a class of links known as "mutants" which are not distinguished by CS link invariants (Kaul 1999). A mutant link is obtained by removing a portion of weaving pattern in a link and then gluing it back after rotating it about any one of three orthogonal axes by an amount $\pi$.

The CS invariants of knots and links can also be used to construct special 3-manifold invariants. Hence, CS theory provides an important tool to study these.

## Manifold Invariants from CS Theory

Different 3-manifolds can be constructed through a procedure called "surgery of framed knots and links" in $S^3$ (Lickorish–Wallace theorem). This construction is not unique. That is, there are many framed knots and links which give the same manifold. However, rules of this equivalence are known: these are called "Kirby moves."

Classification of 3-manifolds would involve finding a method of associating a quantity with the manifold obtained by surgery on the corresponding framed knot/link on $S^3$. If the Kirby moves on the framed knot/link leave this quantity unchanged, then it is a 3-manifold invariant. Knot/link invariants of nonabelian CS theories provide a method of finding such 3-manifold invariants. Equivalently, this procedure gives an algebraic meaning to the surgery construction of 3-manifolds. Details of this method for generating manifold invariants are given in Kaul (1999) and Kaul and Ramadevi (2001).

### Surgery of Framed Knots/Links and Kirby Moves

As discussed earlier, frame of a knot $K$ is an associated closed curve $K_f$ going along the length of the knot wrapping around it certain number of times. Self-linking number (also called framing number) is equal to the linking number of the knot with its frame. There are several ways of fixing this framing. The "standard" framing is one in which the frame number of the knot, that is, the linking number of the knot and its frame is zero. On the other hand, "vertical" framing is obtained by choosing the frame vertically above the knot projected on to a plane. In such a frame, the framing number of a knot is the same as its crossing number. In constructing the 3-manifold invariants from CS theories, we need vertical framing. The framing number may be denoted by writing the integer by the side of knot. We denote a framed $r$-component link by $[L, f]$ where framing $f = (n(1), n(2), \ldots, n(r))$ is a set of integers denoting the framing number of component knots $K_1, K_2, \ldots, K_r$ in the link $L$.

According to the Lickorish–Wallace theorem, surgery over links with vertical framing in $S^3$ yields all the 3-manifolds. This surgery is performed in the following way.

Take a framed $r$-component link $[L, f]$ in $S^3$. Thicken the component knots $K_1, K_2, \ldots, K_r$ such that the solid tubes $N_1, N_2, \ldots, N_r$ so obtained are nonintersecting. Then the compliment $S^3 - (N_1 + N_2 + \cdots + N_r)$ will have $r$ toral boundaries. On the $i$th toral boundary, we imagine an appropriate curve winding $n(i)$ times around the meridian and once along the longitude. Perform a modular transformation so that this curve bounds a disk. This construction is done with each of the toral boundaries. The tubes $N_1, N_2, \ldots, N_r$ are then glued back in to the respective gaps. This surgery thus yields a new 3-manifold. This construction is not unique. The rules of equivalence for surgery on framed knots/links in $S^3$ are two independent Kirby moves.

**Kirby move I** Take an arbitrary $r$-component framed link $[L, f]$ in $S^3$ and consider a curve $C$ with framing number +1 going around the unlinked strands of $L$ as in Figure 6a. We refer to this $(r + 1)$-component link as $H[X]$, where $X$ represents a weaving pattern of the strands. Kirby move I consists of twisting the disk enclosed by $C$ in the clockwise direction from below by an amount $2\pi$. This twisting thereby introduces new crossings

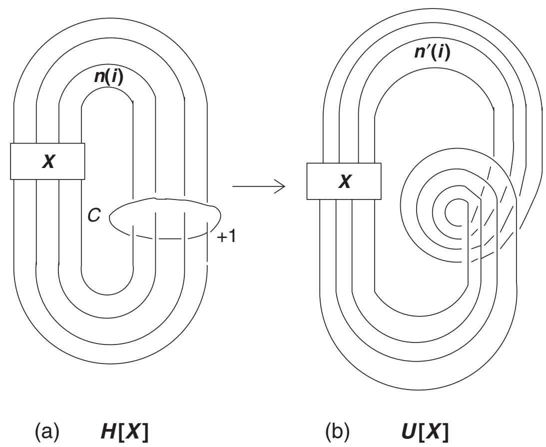



**Figure 6** Kirby move I.

between the curve $C$ and the strands enclosed by it. Then the curve $C$ is removed giving us a new $r$-component link $U[X]$ of Figure 6b. Framing numbers $n'(i)$ of the component knots in link $U[X]$ are related to the framing number $n(i)$ of framed link $[L, f]$ as $n'(i) = n(i) - (\mathcal{L}(K_i, C))^2$, where $\mathcal{L}(K_i, C)$ is the linking number of knot $K_i$ and closed curve $C$. The surgery of the framed links in Figures 6a and 6b will give the same 3-manifold.

Inverse Kirby move I involves removal of a curve $C$ with framing number –1 (instead of +1) after making one complete anticlockwise twist from below on the disk enclosed by $C$. In the process the unlinked strands get twisted in the anticlockwise direction leading to changed framing numbers $n'(i) = n(i) + (\mathcal{L}(K_i, C))^2$ of the component knots $K_i$.

**Kirby move II** This move consists of removing a disjoint unknot $C$ with framing –1 from framed link $[L, f]$ without changing the rest of the link as in Figure 7. Surgery of the two links in Figure 7 will give the same manifold.

Inverse Kirby move II involves removal of a disjoint unknot with framing +1 (instead of –1) from a framed link.

### 3-Manifold Invariants

Now a 3-manifold invariant can be constructed by an appropriate combination of the invariants of framed links in such a way that this algebraic expression is unchanged under the Kirby moves. We

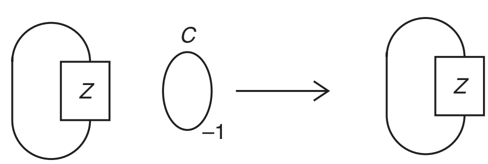



**Figure 7** Kirby move II.

need for this purpose invariants for links in $S^3$ with vertical framing.

Let $\mathcal{M}$ be the manifold obtained from surgery of an $r$-component framed link $[L, f]$ in $S^3$. Then a manifold invariant $\hat{F}^{(\mathcal{G})}[\mathcal{M}]$ is given as a linear combination of the framed link invariants $V^{(\mathcal{G})}_{R_1,\ldots,R_r}[L, f]$, with representations $R_1, R_2, \ldots, R_r$ living on component knots, obtained from CS theory based on a compact semisimple group $\mathcal{G}$:

$$\hat{F}^{(\mathcal{G})}[\mathcal{M}] = \alpha^{-\sigma[L,f]} \sum_{R_1,\ldots R_r} \left( \prod_{i=1}^{r} \mu_{R_i} \right) \times V^{(\mathcal{G})}_{R_1,R_2,\ldots,R_r}[L, f] \qquad [12]$$

Here $\sigma[L, f]$ is the signature of the linking matrix and $\mu_{R_i} = S_{0R_i}$, $\alpha = \mathrm{e}^{\mathrm{i}\pi c/4}$, where $c$ is the central charge of the associated WZNW conformal field theory and $S_{0R_i}$ denotes the matrix element of the modular matrix $S$. General $S$-matrix elements for any compact group are given by

$$S_{R_1R_2} = (-\mathrm{i})^{(d-r)/2} |L_\omega/L|^{-1/2} (k + C_v)^{-1/2} \times \sum_{\omega \in W} \epsilon(\omega) \exp\left( \frac{-2\pi\mathrm{i}}{k + C_v} (\omega(\Lambda_{R_1} + \rho), \Lambda_{R_2} + \rho) \right)$$

where $W$ denotes the Weyl group and its elements $\omega$ are words constructed using the generator $s_{\alpha_i}$ – that is, $\omega = \prod_i s_{\alpha_i}$ and $\epsilon(\omega) = (-1)^{\ell(\omega)}$ with $\ell(\omega)$ as length of the word. Here $\Lambda_{R_i}$'s denotes the highest weights of the representations $R_i$'s and $\rho$ is the Weyl vector. The action of the Weyl generator $s_\alpha$ on a weight $\Lambda_R$ is

$$s_\alpha(\Lambda_R) = \Lambda_R - 2\alpha \frac{(\Lambda_R, \alpha)}{(\alpha, \alpha)}$$

and $|L_\omega/L|$ is the ratio of weight and coroot lattices (equal to the determinant of the Cartan matrix for simply laced algebras). Also $C_v$ is quadratic Casimir invariant for the adjoint representation.

It is important to stress that the expression $\hat{F}^{(\mathcal{G})}[\mathcal{M}]$ is unchanged under both Kirby moves I and II (for detailed proof, see Kaul (1999) and Kaul and Ramadevi (2001)). Notice that for every compact gauge group, we have a new 3-manifold invariant.

**Few examples of 3-manifolds** Table 1 lists the algebraic expressions of this invariant calculated explicitly from the formula in eqn [12] for a few 3-manifolds. All these examples can be constructed by surgery on an unknot $U(f)$ with different frame numbers $f$.

In Table 1 $\mathcal{L}[p, q]$ stands for Lens spaces of the type $(p, q)$ and $C_R$ is the quadratic Casimir invariant

**Table 1** Invariants for some simple manifolds

| $U(f)$ | $\mathcal{M}$ | $\hat{F}^{(\mathcal{G})}[\mathcal{M}]$ |
|---|---|---|
| $U(0)$ | $S^2 \times S^1$ | $1/S_{00}$ |
| $U(\pm 1)$ | $S^3$ | $1$ |
| $U(+2)$ | $RP^3$ | $\alpha^{-1} \sum_R \frac{S_{0R}\, q^{2C_R}\, S_{0R}}{S_{00}}$ |
| $U(+p)$ | $\mathcal{L}[p,1]$ | $\alpha^{-1} \sum_R \frac{S_{0R}\, q^{pC_R}\, S_{0R}}{S_{00}}$ |

for representation $R$ of the Lie algebra of the gauge group $\mathcal{G}$.

Partition function of a CS theory on $\mathcal{M}$ is also an invariant characterizing the 3-manifold. This has been calculated for several manifolds by different methods. Invariant $\hat{F}^{(\mathcal{G})}[\mathcal{M}]$ listed above for various manifolds is related to the CS partition function $Z^{(\mathcal{G})}[\mathcal{M}]$: $\hat{F}^{(\mathcal{G})}[\mathcal{M}] = S_{00}^{-1} Z^{(\mathcal{G})}[\mathcal{M}]$. So the method of constructing 3-manifold invariants above can also be used to calculate the partition function of CS theories.

## 3D Gravity and CS Theory

Three-dimensional CS theory also provides a description of gravity. The 3D gravity including cosmological constant has been first discussed by Deser and Jackiw (1984). The action with cosmological constant $\Lambda = \pm 1/\ell^2$ is:

$$S = \frac{1}{16\pi G} \int_{\mathcal{M}} \mathrm{d}^3x \sqrt{-g}(R - 2\Lambda) \qquad [13]$$

G is the Newton's constant, $g_{\mu\nu}$ is the metric on the 3-manifold $\mathcal{M}$, and R is scalar curvature. Solutions of Einstein equations of motion have a constant positive (negative) curvature if $\Lambda$ is positive (negative). It is also well known that there are no dynamical degrees of freedom for gravity in dimensions $D \leq 3$; it is indeed described by topological field theories. The gravity action above can be rewritten as a CS gauge theory in first-order formulation (Carlip 2003). For triads $e^a_\mu$ and spin connection $\omega^a_\mu$ of Euclidean gravity, we define 1-forms $e = e^a_\mu T^a\, \mathrm{d}x^\mu, \omega = \omega^a_\mu T^a \mathrm{d}x^\mu$, which have values in the Lie algebra of SU(2) whose generators are $T^a = \mathrm{i}\sigma^a/2$ with $\sigma^a$ as three Pauli matrices. In terms of these we define two gauge field 1-forms $A$ and $\bar{A}$ as:

$$A = \left(\frac{\mathrm{i}e}{\ell} + \omega\right), \qquad \bar{A} = \left(\frac{\mathrm{i}e}{\ell} - \omega\right)$$

Then the Euclidean gravity action can be written in terms of two CS actions, $S_{\mathrm{CS}}[A]$ and $S_{\mathrm{CS}}[\bar{A}]$, as

$$S = kS_{\mathrm{CS}}[A] - kS_{\mathrm{CS}}[\bar{A}] \qquad [14]$$

where the coupling constant $k = \ell/(4G)$ for negative cosmological constant $\Lambda = -1/\ell^2$. The gauge group for this theory is $\mathrm{SL}(2, C)$. Infinitesimal diffeomorphisms are described by field-dependent gauge transformations. The corresponding gauge group for Minkowski gravity with negative cosmological constant $\Lambda$ is $\mathrm{SO}(2,R) \otimes \mathrm{SO}(2,R)$. For positive $\Lambda$, one gets $\mathrm{SO}(3,1)$ and $\mathrm{SO}(4)$ for Minkowski and Euclidean metrics, respectively. For $\Lambda = 0$, we have $\mathrm{ISO}(2,1)$ ($\mathrm{ISO}(3)$) as the gauge group for Minkowski (Euclidean) gravity. Hence, the sign of cosmological constant determines the gauge group of the CS theory.

Identification of 3D gravity with CS theory can be used with some advantage to find the partition function for a black hole in 3D gravity with negative cosmological constant. This in turn yields an expression for entropy of the black hole.

### BTZ Black Hole and Its Partition Function

Only for negative $\Lambda$ we have a black hole solution of the Einstein's equations. This solution, known as the BTZ black hole (Carlip 2003), in Euclidean gravity is given by the metric

$$\mathrm{d}s^2_{\mathrm{E}} = \left(-M + \frac{r^2}{l^2} - \frac{J^2}{4r^2}\right)\mathrm{d}\tau^2 + \left(-M + \frac{r^2}{l^2} - \frac{J^2}{4r^2}\right)^{-1} \mathrm{d}r^2 + r^2\left(\mathrm{d}\phi - \frac{J}{2r}\mathrm{d}\tau\right)^2$$

It is specified by two parameters $M$ and $J$ (the mass and angular momentum). By a coordinate transformation, this metric can be rewritten as $\mathrm{d}s^2_{\mathrm{E}} = (l^2/z^2)(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2)$, with $z > 0$. This is the 3D upper-half hyperbolic space and can be rewritten using spherical polar coordinates as

$$\mathrm{d}s^2_{\mathrm{E}} = \frac{l^2}{R^2 \sin^2\chi}\left(\mathrm{d}R^2 + R^2\mathrm{d}\theta^2 + R^2 \sin^2\theta \mathrm{d}\chi^2\right)$$

We have the identifications $(R, \theta, \chi) \sim (R \exp\{2\pi r_+/l\}, \theta + \{2\pi r_-/l\}, \chi)$ where $r_+$ and $r_-$ are the outer and inner horizon radii, respectively. It is clear from this identification that topologically the metric corresponds to a solid torus. Functional integral over this manifold represents a state in the Hilbert space specified by the mass and angular momentum. It is the microcanonical ensemble partition function and its logarithm is the entropy of the black hole.

To evaluate this partition function, the connection 1-form is kept at a constant value on the toroidal boundary through a gauge transformation. We

define local coordinates on the torus boundary $z = x + \tau y$ such that $\int_a \mathrm{d}z = 1$, $\int_b \mathrm{d}z = \tau$, where $a$ ($b$) stands for the contractible (noncontractible) cycle of solid torus and $\tau = \tau_1 + \mathrm{i}\tau_2$ is the modular parameter of the boundary torus. Then connection describing the black hole is

$$A = \left( \frac{-\mathrm{i}\pi \tilde{u}}{\tau_2} \mathrm{d}\bar{z} + \frac{\mathrm{i}\pi u}{\tau_2} \mathrm{d}z \right) T^3 \qquad [15]$$

where $u$ and $\tilde{u}$ are canonically conjugate with commutation relation: $[\tilde{u}, u] = (2/\pi)\tau_2(k+2)^{-1}$. These are related to black hole parameters through holonomies of gauge field $A$ around the $a$- and $b$-cycles (for a classical black hole solution $\Theta = 2\pi$):

$$u = -\frac{\mathrm{i}}{2\pi}\left( -\mathrm{i}\Theta\tau + \frac{2\pi(r_+ + \mathrm{i}|r_-|)}{l} \right)$$

$$\tilde{u} = -\frac{\mathrm{i}}{2\pi}\left( -\mathrm{i}\Theta\bar{\tau} + \frac{2\pi(r_+ + \mathrm{i}|r_-|)}{l} \right)$$

For a fixed value of connection, namely $u$, the functional integral is described by a state $\psi_0$ with no Wilson line in the bulk. The states with Wilson line carrying spin $j/2$ are given by Labastida and Ramallo:

$$\psi_j(u, \tau) = \exp\left\{ \frac{\pi k}{4\tau_2} u^2 \right\} \chi_j(u, \tau)$$

where the Weyl–Kac characters for affine su(2)

$$\chi_j(u, \tau) = \frac{\Theta_{j+1}^{(k+2)}(u, \tau) - \Theta_{-j-1}^{(k+2)}(u, \tau)}{\Theta_1^2(u, \tau) - \Theta_{-1}^2(u, \tau)}$$

and $\Theta$ functions are defined by

$$\Theta_\mu^k(u, \tau) = \sum_{n \in \mathcal{Z}} \exp\left\{ 2\pi \mathrm{i} k \left[ \left( n + \frac{\mu}{2k} \right)^2 \tau + \left( n + \frac{\mu}{2k} \right) u \right] \right\}$$

Given the collection of states $\psi_j$, we write the partition function by choosing an appropriate ensemble for fixed mass and angular momentum. This black hole partition function is:

$$Z_{\mathrm{BH}} = \int \mathrm{d}\mu(\tau, \bar{\tau}) \left| \sum_{j=0}^{k} (\psi_j(0, \tau))^* \psi_j(u, \tau) \right|^2$$

where modular invariant measure is $\mathrm{d}\mu(\tau, \bar{\tau}) = \mathrm{d}\tau\, \mathrm{d}\bar{\tau}/\tau_2^2$. This integral can be worked out for large black hole mass and zero angular momentum in saddle-point approximation. The computation yields (Govindarajan *et al.* 2001):

$$Z_{\mathrm{BH}} = \frac{l^2}{r_+^2} \sqrt{\frac{8 r_+ G}{\pi l^2}} \exp\left( \frac{2\pi r_+}{4G} \right) + \cdots \qquad [16]$$

This gives not only the leading Bekenstein–Hawking behavior of the black hole entropy $S$ but also a subleading logarithmic term:

$$S = \ln Z_{\mathrm{BH}} = \frac{2\pi r_+}{4G} - \frac{3}{2} \ln \frac{2\pi r_+}{4G} + \cdots$$

This is an interesting application of CS theory to 3D gravity. In fact, three-dimensional CS theory also has applications in the study of black holes in four-dimensional gravity: the boundary degrees of freedom of a black hole in 4D are also described by an SU(2) CS theory. This allows a calculation of the degrees of freedom of, for example, Schwarzschild black hole. For large area black holes, this in turn results in an expression for the entropy which, besides a Bekenstein–Hawking area term, has a logarithmic area correction with same coefficient –3/2 as above. This suggests a universal, dimension-independent, nature of the these logarithmic corrections.

*See also:* BF Theories; The Jones Polynomial; Knot Theory and Physics; Large-*N* and Topological Strings; Quantum 3-Manifold Invariants; Topological Quantum Field Theory: Overview.

## Further Reading

Alexandrov M, Konstsevich M, Schwarz A, and Zaboronsky O (1997) The geometry of the master equations and topological quantum field theory. *International Journal of Modern Physics A* 12: 1405–1430.

Atiyah M (1989) *The Geometry and Physics of Knots*. Cambridge: Cambridge University Press.

Carlip S (2003) *Quantum Gravity in 2+1 Dimensions*, Cambridge Monographs on Mathematical Physics, Cambridge: Cambridge University Press.

Deser S and Jackiw R (1984) Three-dimensional cosmological gravity: dynamics of constant curvature. *Annals of Physics* 153: 405–416.

Deser S, Jackiw R, and Templeton S (1982a) Three-dimensional massive gauge theory. *Physical Review Letters* 48: 975–978.

Deser S, Jackiw R, and Templeton S (1982b) Topological massive gauge theories. *Annals of Physics NY* 140: 372–411.

Govindarajan TR, Kaul RK, and Suneeta V (2001) Logarithmic correction to the Bekenstein–Hawking entropy of the BTZ black hole. *Classical and Quantum Gravity* 18: 2877–2886.

Kaul RK (1999) Chern–Simons theory, knot invariants, vertex models and three-manifold invariants. In: Kaul RK, Maharana J, Mukhi S, and Kalyana Rama S (eds.) *Frontiers of Field Theory, Quantum Gravity and Strings*, Horizons in World Physics, vol. 227, pp. 45–63. New York: NOVA Science Publishers.

Kaul RK and Ramadevi P (2001) Three-manifold invariants from Chern–Simons field theory with arbitrary semi-simple gauge groups. *Communications in Mathematical Physics* 217: 295–314.
Schwarz AS (1978) The partition function of degenerate quadratic functional and Ray–Singer invariants. *Letters in Mathematical Physics* 2: 247–252.
Schwarz AS (1979) The partition function of a degenerate functional. *Communications in Mathematical Physics* 67: 1–16.
Schwarz AS (1987) *New Topological Invariants in the Theory of Quantized Fields.* Abstracts in the Proceedings of International Topological Conference, Baku, Part II.
Witten E (1988) Topological quantum field theory. *Communications in Mathematical Physics* 117: 353–386.
Witten E (1989) Quantum field theory and the Jones polynomial. *Communications in Mathematical Physics* 121: 351–399.
Witten E (1995) Chern–Simons gauge theory as a string theory. *Progress in Mathematics* 133: 637–678.

# Seiberg–Witten Theory

**Siye Wu**, University of Colorado, Boulder, CO, USA



## Introduction

Gauge theory is the cornerstone of the standard model of elementary particles. The original motivation for studying supersymmetric gauge theories was phenomenological (such as the hierarchy problem). They display a large number of interesting phenomena and become the models for the dynamics of strongly coupled field theories. They also offer valuable insights to nonsupersymmetric models. In $N=1$ gauge theory, the low-energy effective superpotential is holomorphic both in the superfields and in the coupling constants. This powerful holomorphy principle, together with symmetry and various limits, often determines the effective superpotential completely. Such theories often have quantum moduli spaces where the classical singularities are smoothed out, continuous interpolation between Higgs and confinement phases, massless composite mesons and baryons, and dual theories weakly coupled at low energy. For $N=2$ pure gauge theory, the low-energy effective theory is an abelian gauge theory in which both the kinetic term and the coupling constant are determined by a holomorphic prepotential. The electric–magnetic duality is in the ambiguity of the low-energy description. Much physical information, such as the coupling constant, the Kähler metric on the quantum moduli, the monodromy around the singularities, can be incorporated in a family of elliptic curves. This low-energy exact solution is also useful to topological field theory that can be obtained from the $N=2$ theory by twisting. Much of the above was the work of Seiberg and Witten in the mid-1990s. In this article, we review some of the fascinating aspects of $N=1$ and $N=2$ supersymmetric gauge theories.

## $N=1$ Gauge Theory and Seiberg Dualities

### $N=1$ Yang–Mills Theory and QCD

Let $G$ be a compact Lie group and let $P$ be a principal $G$-bundle over the Minkowski space $\mathbb{R}^{3,1}$. In pure gauge theory, the dynamical variable is a connection $A$ in $P$; two connections are equivalent if they are related by a gauge transformation. Let $F \in \Omega^2(\mathbb{R}^{3,1}, \mathrm{ad}\,P)$ be the curvature of $A$. It decomposes into the self-dual and anti-self-dual parts, that is, $F = F^+ + F^-$, where $F^{\pm} = (1/2)(F \mp \sqrt{-1} * F)$. With a suitably normalized nondegenerate bilinear form $\langle \cdot,\cdot \rangle$ on the Lie algebra $\mathfrak{g}$, the classical action is

$$S_{\mathrm{YM}}[A] = \int_{\mathbb{R}^{3,1}} -\frac{1}{2g^2}\langle F \wedge *F\rangle + \frac{\theta}{16\pi^2}\langle F \wedge F\rangle$$
$$= \int_{\mathbb{R}^{3,1}} -\frac{\tau}{8\pi}\langle F^+ \wedge F^+\rangle - \frac{\bar{\tau}}{8\pi}\langle F^- \wedge F^-\rangle$$

Here $g > 0$ is the coupling constant and $\theta \in \mathbb{R}$, the $\theta$ angle, and

$$\tau = \frac{\theta}{2\pi} + \frac{4\pi\sqrt{-1}}{g^2}$$

is a complex number in the upper-half plane that incorporates both. Classically, the theory is conformally invariant and the dynamics is independent of the $\theta$-term. At the quantum level, $\theta(\mathrm{mod}2\pi)$ appears in the path integral and parametrizes inequivalent vacua. The coupling constant runs as energy $\mu$ varies, satisfying the renormalization group equation

$$\mu\frac{\mathrm{d}g}{\mathrm{d}\mu} = -\frac{b_0}{(4\pi)^2}g^3 + o(g^5)$$

where the right-hand side is called the $\beta$-function $\beta(g)$. This introduces, when $b_0 \neq 0$, a mass scale $\Lambda$ given by

$$(\Lambda/\mu)^{b_0} = \mathrm{e}^{-8\pi^2/g(\mu)^2}$$

up to one-loop. Consequently, the classical scale invariance is lost. It is convenient to redefine $\Lambda$ as a complex quantity such that

$$(\Lambda/\mu)^{b_0} = \mathrm{e}^{2\pi\sqrt{-1}\tau(\mu)}$$

For pure gauge theory, $b_0 = (11/3)\check{h}$, where $\check{h}$ is the dual Coxeter number of $\mathfrak{g}$. At high energy ($\mu \to \infty$), the coupling becomes weak ($g \to 0$); this is known as asymptotic freedom. On the contrary, the interaction becomes strong at low energy. It is believed that the theory exhibits confinement and has a mass gap.

QCD, or quantum chromodynamics, is gauge theory coupled to matter fields. Suppose the boson $\phi$ and the fermion $\psi$ are in the (complex) representations $R_b$ and $R_f$ of $G$, respectively. That is, $\phi \in \Gamma(P \times_G R_b)$, or $\phi$ is a section of the bundle $P \times_G R_b$, and $\psi \in \Gamma(S \otimes (P \times_G R_f))$, where $S$ is the spinor bundle over $\mathbb{R}^{3,1}$. The classical action is

$$S_{\mathrm{QCD}}[A,\phi,\psi] = S_{\mathrm{YM}}[A] + \frac{1}{g^2}\int \mathrm{d}^4x \frac{1}{2}|\nabla\phi|^2 + \sqrt{-1}(\psi, \nabla\!\!\!/\,\psi) + \cdots$$

where $\nabla$ is the covariant derivative, $\nabla\!\!\!/$ is the Dirac operator coupled to $A$, and we have omitted possible mass and potential terms. The quantum theory depends sensitively on the representations $R_b$ and $R_f$. In the $\beta$-function, we have

$$b_0 = \tfrac{11}{3}\check{h} - \tfrac{1}{6}\nu(R_b) - \tfrac{2}{3}\nu(R_{\mathrm{f}})$$

where $\nu(R)$ is the Dynkin index of a representation $R$. If $b_0 < 0$, the theory is free in the infrared but strongly interacting in the ultraviolet. If $b_0 > 0$, the converse is true; in particular, the theory exhibits asymptotic freedom. If $b_0 = 0$, the situation depends on the sign of the two or higher-loop contributions.

Pure $N = 1$ supersymmetric gauge theory is one on the superspace $\mathbb{R}^{3,1|(2,2)}$ with a constraint that the curvature vanishes in the odd directions. The dynamical variables are in the superfield strength $W$, a $1|(1,0)$-form valued in ad $P$. In components, the theory is gauge field coupled to a Majorana or Weyl fermion in the adjoint representation. Let $S^{\pm}$ be spinor bundles of positive (negative) chiralities, respectively, and let $\lambda$ be a section of $S^+ \otimes \mathrm{ad}P$. The action, written both in superspace and in ordinary spacetime, is

$$S_{\mathrm{SYM}}^{N=1}[A,\lambda] = \frac{1}{4\pi}\mathrm{Im}\left(\int \mathrm{d}^4x\, \mathrm{d}^2\theta\, \tau\langle W, W\rangle\right) = S_{\mathrm{YM}}[A] + \frac{1}{g^2}\int \mathrm{d}^4x\sqrt{-1}\langle\bar{\lambda}, \nabla\!\!\!/^{+}\lambda\rangle$$

Since $b_0 = 3\check{h}$, the theory is asymptotically free but strongly coupled at low energy. Classically, the theory has a $\mathrm{U}(1)_R$ chiral symmetry. However, due to anomaly, only the subgroup $\mathbb{Z}_{2\check{h}}$ survives at the quantum level. Instanton effect yields gaugino condensation $\langle\lambda\lambda\rangle \sim \Lambda^3$. The symmetry is thus further broken to $\mathbb{Z}_2$, resulting $\check{h}$ inequivalent vacua.

The $N = 1$ QCD has additional chiral superfields $\Phi$ in a representation $R$, including the bosons $\phi \in \Gamma(P \times_G R)$ and the fermions $\psi \in \Gamma(S^+ \otimes (P \times_G R))$. In the absence of superpotential, the action is

$$S_{\mathrm{SQCD}}^{N=1}[A,\lambda,\phi,\psi] = S_{\mathrm{SYM}}^{N=1}[A,\lambda] + \frac{1}{g^2}\int \mathrm{d}^4x\, \mathrm{d}^2\theta\, d^2\bar{\theta}\, \tfrac{1}{2}|\Phi|^2$$

In components, the second term is

$$\frac{1}{g^2}\int \mathrm{d}^4x\left(\tfrac{1}{2}|\nabla\phi|^2 + \sqrt{-1}(\psi, \nabla\!\!\!/^{+}\psi) - \tfrac{1}{2}\phi^*|D|^2 + \cdots\right)$$

where $D: R \to \mathfrak{g}^*$ is the moment map of the Hamiltonian $G$-action on $R$, and we have omitted other terms containing fermionic fields. The moduli space of classical vacua is the symplectic quotient $D^{-1}(0)/G = R//G$. It is the same as the Kähler quotient $R^{\mathrm{s}}/G^{\mathbb{C}}$, where the stable subset $R^{\mathrm{s}} = \{\phi \in R | G^{\mathbb{C}} \cdot \phi \cap D^{-1}(0) \neq \emptyset\}$ is open and dense in $R$. Again, the quantum theory depends on the representation $R$. Since $b_0 = 3\check{h} - (1/2)\nu(R)$, the theory is asymptotically free, infrared free, scale invariant (to one-loop) when $\nu(R) < 6\check{h}, \nu(R) > 6\check{h}, \nu(R) = 6\check{h}$, respectively. The moduli space may be lifted by a superpotential or modified by other quantum effects.

### $\mathrm{SU(N_c)}$ Theories at Low Energy

We now consider $N = 1$ QCD with $G = \mathrm{SU}(N_c)$; $N_c$ is the number of colors. The matter field consists of $N_f$ copies of quarks $Q^i (1 \le i \le N_f)$ in the fundamental representation of $\mathrm{SU}(N_c)$ and $N_f$ copies of antiquarks $Q'_{i'} (1 \le i' \le N_f)$ in the conjugate representation. Using the isomorphism of $\mathfrak{su}(N_c)$ with its dual, the moment map is

$$D(Q, Q') = \text{traceless part of } \sqrt{-1}(QQ^{\dagger} - Q'Q'^{\dagger})$$

So $(Q, Q') \in D^{-1}(0)$ if and only if $QQ^{\dagger} - Q'^{\dagger} = cI_{N_c}$ for some $c \in \mathbb{R}$. If $N_f < N_c$, then $c = 0$ and

$$Q, Q' \sim \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_{N_f} \end{pmatrix}$$

for some $a_k \geq 0$. Generically, these $a_k > 0$ and the gauge group $SU(N_c)$ is broken to $SU(N_c - N_f)$. If $N_f \geq N_c$, then

$$Q \sim \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_{N_c} \end{pmatrix}, \qquad Q' \sim \begin{pmatrix} a'_1 & & \\ & \ddots & \\ & & a'_{N_c} \end{pmatrix}$$

where $a_k, a'_k \geq 0$ satisfy $a_k^2 - a'^2_k = c$ for some $c \in \mathbb{R}$. The gauge group is completely broken. The low-energy superfields are the mesons $M^i_{i'} = Q^i Q'_{i'}$ and, if $N_f \geq N_c$, the baryons

$$B_{i_{N_c+1}\cdots i_{N_f}} = \frac{1}{N_c!}\epsilon_{i_1\cdots i_{N_f}} Q^{i_1}\cdots Q^{i_{N_c}}$$
$$B'^{i'_{N_c+1}\cdots i'_{N_f}} = \frac{1}{N_c!}\epsilon^{i'_1\cdots i'_{N_f}} Q'_{i'_1}\cdots Q'_{i'_{N_c}}$$

When $N_f < N_c$, Affleck *et al.* (1984) found a dynamically generated superpotential

$$W_{\text{eff}}(\hat{M}) = (N_c - N_f)\left(\frac{\Lambda^{3N_c - N_f}}{\det M}\right)^{1/(N_c - N_f)}$$

generated by instanton effect when $N_f = N_c - 1$ and by gaugino condensation in the unbroken $SU(N_c - N_f)$ theory when $N_f < N_c - 1$. It is also the unique superpotential (up to a multiplicative constant) that is consistent with the global and supersymmetry. The potential pushes the vacuum to infinity. Therefore, contrary to the classical picture, theories with $N_f < N_c$ do not have a vacuum at the quantum level.

When $N_f \geq 3N_c$, the theory is not strongly interacting at low energy, and perturbation methods are reliable. (When $N_f = 3N_c$, the two-loop contribution to the $\beta$-function is negative.) We now look at the range $N_c \leq N_f < 3N_c$. The cases $N_f = N_c, N_c + 1$ and $N_c + 2 \leq N_f < 3N_c$ were studied in Seiberg (1994) and Seiberg (1995), respectively.

When $N_f = N_c$, the classical moduli space is $\det M = BB'$. The quantum theory at low energy consists of the fields $M$, $B$, $B'$ satisfying the constraint $\det M - BB' = \Lambda^{2N_c}$. The quantum moduli space is smooth everywhere, and there are no additional massless particles. So the gluons are heavy throughout the moduli space. This is due to confinement near the origin, where the interaction is strong, and due to the Higgs mechanism far out in the flat direction, where the classical picture is a good approximation. We see a smooth transition between these two effects.

When $N_f = N_c + 1$, there is a dynamically generated superpotential

$$W_{\text{eff}} = \frac{1}{\Lambda^{2N_c - 1}}(B'MB - \det M)$$

The stationary points of $W_{\text{eff}}$ are at $BB' - \wedge^{N_c} M = 0$, $BM = 0, MB' = 0$; these are precisely the constraints that the classical configuration satisfies. However, the moduli space is interpreted differently: it is embedded into a larger space, and the constraints are satisfied only at stationary points. At the singularity $\langle M \rangle = 0$, the whole global symmetry group is unbroken, and $B, B'$ are the new massless fields resolving the singularity. So we have a continuous transition between confinement (without chiral symmetry breaking) and the Higgs mechanism in the semiclassical regime.

When $N_c + 2 \leq N_f \leq (3/2)N_c$, the original theory, called the electric theory, is still strongly coupled in the infrared. Seiberg (1995) proposed that there is a dual, magnetic theory, which is infrared free. The two theories are different classically, but are equivalent at the quantum level. The dual theory is an $N = 1 SU(\tilde{N}_c)$ gauge theory with $\tilde{N}_c = N_f - N_c$, coupled to dual quarks $\tilde{Q}_i, \tilde{Q}'^{i'}$, where $1 \leq i; i' \leq N_f$ are flavor indices. In addition, the mesons $M^i_{i'}$ become fundamental fields. They are not coupled to the $SU(\tilde{N}_c)$ gauge field but interact with the dual quarks through the superpotential

$$W = \mu^{-1} M^i_{i'} \tilde{Q}_i \tilde{Q}'^{i'}$$

The two theories have the same global symmetry and the same gauge-invariant operators. The dual quarks are fundamental in the magnetic theory but are solitonic excitations in the electric theory. At high energy, the electric theory is asymptotically free, while the magnetic theory is strongly coupled. At low energies, the converse is true. Therefore, reliable perturbative calculations can be performed by choosing an appropriate weakly coupled theory.

When $(3/2)N_c < N_f < 3N_c$, the theory has a nontrivial infrared fixed point. This is because up to two-loop,

$$\beta(g) = -\frac{g^3}{16\pi^2}(3N_c - N_f) + \frac{g^5}{128\pi^4}\left(2N_c N_f - 3N_c^2 - \frac{N_f}{N_c}\right) + o(g^7)$$

There is a solution $g_* > 0$ to $\beta(g) = 0$. We have $\beta(g) < 0$ when $0 < g < g_*, \beta(g) > 0$ when $g > g_*$. In the infrared limit, the coupling constant flows to $g = g_*$, where we have a nontrivial, interacting superconformal theory in four dimensions. The conformal dimension becomes anomalous and is equal to 3/2 of the charge of the chiral $U(1)_R$; for example, that of the meson $\mu^{-1}M$ is $3(N_f - N_c)/N_f > 1$ in this range.

### Other Classical Gauge Groups

We now consider $N=1$ supersymmetric gauge theory and QCD with gauge groups $Sp(N_c)$ and $SO(N_c)$. The $Sp(N_c)$ theories, studied by Intriligator and Pouliot (1995), are the simplest examples of the $N=1$ theories. We take $2N_f$ chiral superfields $Q_i (i=1,\ldots,2N_f)$ in the fundamental representation $\mathbb{C}^{2N_c} \cong \mathbb{H}^{N_c}$ of $Sp(N_c)$. The number of copies must be even so that the quantum theory is free from global gauge anomaly. The gauge-invariant quantities are the mesons $M_{ij}=Q_i^a Q_j^b \omega_{ab}$, where $\omega$ is the symplectic form on $\mathbb{C}^{2N_c}$, subject to a constraint $\epsilon^{1,\ldots,2N_c+2} M_{1,2} \cdots M_{2N_c+1,2N_c+2}=0$. Using the decomposition $\mathfrak{u}(2N_c)=\mathfrak{sp}(N_c) \oplus \sqrt{-1}\{\mathbb{H}\text{-self-adjoint matrices}\}$, the moment map $D(Q)$ is the projection of $\sqrt{-1}QQ^\dagger$ on $\mathfrak{sp}(N_c)$. So $D(Q)=0$ implies

$$Q \sim \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_{\min\{N_c,N_f\}} \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

where $a_k \geq 0$. At a generic point of the classical moduli space, the gauge group is broken to $Sp(N_c - N_f)$ if $N_c > N_f$; it is completely broken if $N_c \leq N_f$.

Since $b_0 = 3(N_c+1) - N_f$, the quantum theory is infrared free if $N_f \geq 3(N_c+1)$. (When $b_0=0$, the two-loop $\beta$-function is negative.) When $N_f \leq N_c$, there is a dynamically generated superpotential

$$W_{\text{eff}} = (N_c + 1 - N_f) \times \left( \frac{2^{N_c-1} \Lambda^{3(N_c+1)-N_f}}{\text{Pf}\, M} \right)^{1/(N_c+1-N_f)}$$

pushing the vacuum to infinity.

When $N_f = N_c$, the classical moduli space $\text{Pf}M=0$ has singularities. The quantum moduli space is $\text{Pf}\, M = 2^{N_c-1}\Lambda^{2(N_c+1)}$. The singularity is smoothed out and there are no light fields other than the mesons $M$. When $N_f = N_c+1$, all components of $M$ become dynamical in the low-energy theory, and there is a superpotential

$$W_{\text{eff}} = -\frac{\text{Pf}\, M}{2^{N_c-1}\Lambda^{2N_c+1}}$$

At the most singular point $\langle M \rangle = 0$, the global symmetry is unbroken, and all the light fields in $M$ become massless. In both cases, there is a transition between confinement and Higgs mechanism.

When $N_c + 3 \leq N_f \leq (3/2)(N_c+1)$, there is a dual, magnetic theory which is free in the infrared. The dual theory has $2N_f$ quarks $\tilde{Q}^i$ in the fundamental representation of $Sp(\tilde{N}_c)$, where $\tilde{N}_c = N_f - N_c - 2$. In addition, the mesons $M_{ij}$ become elementary and couple to $\tilde{Q}$ through a superpotential $W=(2\mu)^{-1} M_{ij} \tilde{Q}^{ia} \tilde{Q}^{jb} \tilde{\omega}_{ab}$, where $\tilde{\omega}$ is the symplectic form on $\mathbb{C}^{2\tilde{N}_c}$. When $(3/2)(N_c+1) < N_f < 3(N_c+1)$, the theory flows to an interacting superconformal field theory in the infrared.

Theories with the $SO(N_c)$ gauge group were studied by Seiberg (1995) and by Intriligator and Seiberg (1995). Since the fundamental representation is real, there is no constraint on the number $N_f$ of quarks $Q^i (1 \leq i \leq N_f)$. The gauge invariants are the mesons $M^{ij}=Q_a^i Q_b^j \delta^{ab}$ and, if $N_f \geq N_c$, the baryons $B_{i_{N_c+1}\cdots i_{N_f}} = \epsilon_{i_1 \cdots i_{N_f}} Q^{i_1} \cdots Q^{i_{N_c}}/N_c!$ They satisfy rank $M \leq N_c$ and $BB = \wedge^{N_c} M$. Using the decomposition $\mathfrak{u}(N_c) = \mathfrak{so}(N_c) \oplus \sqrt{-1}\{\mathbb{R}\text{-self-adjoint matrices}\}$, the moment map $D(Q)$ is the projection $\sqrt{-1}QQ^\dagger$ on $\mathfrak{so}(N_c)$. If $D(Q)=0$, then up to gauge and global symmetries, $Q$ is of the form

$$Q \sim \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_r \end{pmatrix}$$

where $a_1,\ldots,a_r > 0$ if $r = \text{rank}\, Q \leq N_c$ and $a_1,\ldots,a_{N_c-1} > 0$ and $a_{N_c} \neq 0$ if $r = N_c$. Generically, the gauge group is broken to $SO(N_c - N_f)$ if $N_c \geq N_f + 2$ and is totally broken if $N_c < N_f + 2$.

We have $b_0 = 3(N_c-2) - N_f$ if $N_c \geq 5$. For $N_c = 4$, the group is $(SU(2) \times SU(2))/\mathbb{Z}_2$ and $b_0 = 6 - N_f$ for each $SU(2)$ factor. If $N_c = 3$, the group is $SU(2)/\mathbb{Z}_2 b_0 = 6 - 2N_f$. The theory is asymptotically free if $N_f > 3(N_c-2)$ and infrared free if $N_f \geq 3(N_c-2)$.

When $N_f \leq N_c - 5$, there is a dynamically generated superpotential

$$W_{\text{eff}} = \frac{1}{2}(N_c - 2 - N_f) \times \left( \frac{16\Lambda^{3(N_c-2)-N_f}}{\det M} \right)^{1/(N_c-2-N_f)}$$

lifting the classical vacuum degeneracy. The coefficient is fixed by mass deformation and by matching the SU(4) theory when $N_c = 6$.

When $N_f = N_c - 4$, the unbroken gauge group is $SO(4) = (SU(2) \times SU(2))/\mathbb{Z}_2$ on the generic point of the moduli space. The superpotential of the original theory is

$$W_{\text{eff}} = 2(\epsilon_+ + \epsilon_-) \left( \frac{\Lambda^{2(N_c-1)}}{\det M} \right)^{1/2}$$

where the choices $\epsilon_+, \epsilon_- = \pm 1$ correspond to the fact that each of the SU(2) theory has two vacua. There are two physically inequivalent branches: $\epsilon_+ = \epsilon_-$ and $\epsilon_+ = -\epsilon_-$. For $\epsilon_+ = \epsilon_-$, the superpotential pushes the vacuum to infinity. For $\epsilon_+ = -\epsilon_-, W_{\text{eff}} = 0$. In the quantum theory, the singularity is smoothed out and

all the massless fermions are in $M$, even at the origin of the moduli space. Hence the quarks are confined.

When $N_f = N_c - 3$, the unbroken gauge group is SO(3) and the theory has two branches with

$$W_{\text{eff}} = 4(1+\epsilon)\frac{\Lambda^{2N_c-3}}{\det M}$$

where $\epsilon = \pm 1$. For $\epsilon = 1$, the quantum theory has no vacuum. For $\epsilon = -1$, $W_{\text{eff}} = 0$, but there are additional light fields $\tilde{Q}_i$ coupling to $M$ via the superpotential $W \sim (2\mu)^{-1} M^{ij} \tilde{Q}_i \tilde{Q}_j$ near $M = 0$.

When $N_f = N_c - 2$, the low-energy theory is related to the $N = 2$ gauge theory and will be addressed in the subsection "Seiberg–Witten's low-energy solution."

When $N_f \geq N_c - 1$, we define a dual, magnetic theory whose gauge group is $\mathrm{SO}(\tilde{N}_c)$, where $\tilde{N}_c = N_f - N_c + 4$. There are $N_f$ dual quarks $\tilde{Q}_i (1 \leq i \leq N_f)$ in the fundamental representation. This theory is infrared free if $N_f \leq (3/2)(N_c - 2)$. In the effective theory, the mesons $M^{ij}$ become fundamental and couple with the dual quarks through a superpotential $W = (2\mu)^{-1} M^{ij} \tilde{Q}_i \tilde{Q}_j$ if $N_f \geq N_c$; there is an additional term $\det M / 64\Lambda^{2N_c-5}$ if $N_f = N_c - 1$. When $(3/2)(N_c - 2) < N_f < 3(N_c - 2)$, the theory flows to an interacting superconformal field theory in the infrared.

## $N = 2$ Gauge Theory and Seiberg–Witten Duality

### $N = 2$ Yang–Mills Theory

Pure $N = 2$ supersymmetric gauge theory is a special case of $N = 1$ QCD when $R = \mathfrak{g}^{\mathbb{C}}$ is the (complexified) adjoint representation of $G$. The moment map is $D(\phi) = (1/2\sqrt{-1})[\phi, \bar{\phi}] \in \mathfrak{g} \cong \mathfrak{g}^*(\phi \in \mathfrak{g})$. Since the fermionic fields $\lambda$ and $\psi$ are sections of the same bundle, there is a second set of supersymmetry transformations by interchanging the roles of $\lambda$ and $\psi$. This makes the theory $N = 2$ supersymmetric. The classical action is

$$\begin{aligned} S^{N=2}_{\text{SYM}}[A, \lambda, \psi, \phi] = S_{\text{YM}}[A] & \\ + \frac{1}{g^2} \int \mathrm{d}^4x \sqrt{-1} (\langle \bar{\lambda}, \nabla\!\!\!/ \lambda \rangle & \\ + \langle \bar{\psi}, \nabla\!\!\!/ \psi \rangle) + \frac{1}{2} |\nabla\phi|^2 & \\ + \sqrt{-1} (\langle \bar{\phi}, [\lambda, \psi] \rangle + \langle \phi, [\bar{\lambda}, \bar{\psi}] \rangle) & \\ - \frac{1}{8} |[\phi, \bar{\phi}]|^2 & \end{aligned}$$

The energy reaches the minimum when $\phi$ takes a constant value $\phi \in \mathfrak{g}^{\mathbb{C}}$ that can be conjugated by $G$ to the Cartan subalgebra $\mathfrak{t}^{\mathbb{C}}$. ($\mathfrak{t}$ is the Lie algebra of the maximal torus $T$.) The classical moduli space is $\mathfrak{g}^{\mathbb{C}}/G^{\mathbb{C}} = \mathfrak{t}^{\mathbb{C}}/W$, where $W$ is the Weyl group. At a generic $\phi \in \mathfrak{t}^{\mathbb{C}}$, the gauge group is broken to $T$ by the Higgs mechanism. Classically, the massless degrees of freedom are excitations of $\phi$ and components of the gauge field in $\mathfrak{t}$. So the low-energy physics can be described by these massless fields. However, the moduli space is singular when $\phi$ is on the walls of the Weyl chambers. At these values, the unbroken gauge group is larger and there are extra massless fields that resolve the singularities.

Since $b_0 = 2\check{h} > 0$, the quantum theory is asymptotically free but strongly interacting at low energy. It can be shown that $N = 1$ supersymmetry already forbids a dynamically generated superpotential on $\mathfrak{t}^{\mathbb{C}}/W$. Therefore, the vacuum degeneracy is not lifted and the quantum moduli space is still a continuum. However, there are corrections to the part of classical moduli space where strong interactions occur. The quantum theory has a dynamically generated mass scale $\Lambda$. We pick the renormalization scale $\mu$ to be $|\phi|$, the typical energy scale where spontaneous symmetry breaking occurs. Far away from the origin, that is, when $|\phi| \gg |\Lambda|$, the theory is weakly interacting and the classical description of the moduli space is a good approximation. However, when $|\phi|$ is comparable to $|\Lambda|$, the classical language and perturbation methods fail due to strong interaction. At $\phi = 0$, the full gauge symmetry is restored classically. But since the theory becomes strongly interacting at low energy, it cannot be the low-energy solution of the original theory.

The classical $\mathrm{U}(1)_R$ symmetry extends to $\mathrm{U}(2)_R$, mixing $\lambda$ and $\psi$. The $\mathrm{U}(1)_R$ subgroup in $\mathrm{U}(2)_R$ is anomalous except for a subgroup $\mathbb{Z}_{4\check{h}}$. So we have a global $\mathrm{SU}(2)_R \times_{\mathbb{Z}_2} \mathbb{Z}_{4\check{h}}$ symmetry at the quantum level. This is consistent with a continuous moduli space of vacua, if the group $\mathrm{SU}(2)_R$ is to act nontrivially. Also, the space is not a single orbit of the global symmetry group. The generator of $\mathbb{Z}_{4\check{h}}$ acts on $\mathfrak{t}^{\mathbb{C}}$ by a phase $\mathrm{e}^{-\pi\sqrt{-1}/\check{h}}$. The group $\mathbb{Z}_{4\check{h}}$ is spontaneously broken to the subgroup which acts trivially on $\mathfrak{t}^{\mathbb{C}}/W$.

We study the general form of low-energy effective Lagrangian that is consistent with $N = 2$ supersymmetry. We assume that the quantum effect does not modify the topology of the moduli space $\mathfrak{t}^{\mathbb{C}}/W$, though it may alter the singularity and its nature. Suppose $U$ is the quantum moduli. At a generic point in $U$, the residual gauge group is $T$. In the $N = 1$ language, the theory is a supersymmetric gauged sigma model with target space $U$. It contains $N = 1$ vector multiplets $W^I$ and chiral multiplets $\Phi^I$, where $1 \leq I \leq r$, $r = \dim T$ being the rank of $G$. $N = 1$ supersymmetry requires that $U$ is Kähler, with

possible singularities where the effective theory breaks down. $N=2$ supersymmetry requires further that $U$ is special Kähler, that is, there is a flat, torsion-free connection $\nabla$ on $TU$ such that the Kähler form $\omega$ is parallel and such that $d_\nabla J=0$, where the complex structure $J$ is viewed as a 1-form valued in $TU$. See, for example, Freed (1999). Locally, there is a holomorphic prepotential $\mathcal{F}$ and special coordinates $\{z^I\}$. Let $\tilde{z}_I=\partial\mathcal{F}/\partial z^I$ be the dual coordinates and let $\tau_{IJ}=\partial^2\mathcal{F}/\partial z^I\partial z^J=\partial\tilde{z}_I/\partial z^J$. Then $K=\mathrm{Im}(\tilde{z}_I\bar{z}^I)$ is a Kähler potential and $\omega=(\sqrt{-1}/2)$ $\mathrm{Im}(\tau_{IJ})\mathrm{d}z^I\wedge\mathrm{d}\bar{z}^J$ is the Kähler form. The effective action is

$$S_{\mathrm{eff}}^{N=2}[W,\Phi]=\frac{1}{4\pi}\mathrm{Im}\left(\int \mathrm{d}^4x\,\mathrm{d}^2\theta\,\tfrac{1}{2}\tau_{IJ}(\Phi)(W^I,W^J)+\int \mathrm{d}^4x\,\mathrm{d}^2\theta\mathrm{d}^2\bar{\theta}K(\Phi)\right)$$

Note that both the coupling constants $\tau_{IJ}$ and the metric $\mathrm{Im}\tau_{IJ}$ on $U$ are determined by a holomorphic function $\mathcal{F}$, which is the hallmark of $N=2$ supersymmetry.

In the bare theory with abelian gauge group $T$, the action is given by choosing $\mathcal{F}_0(\Phi)=(1/2)\tau_{IJ}\langle\Phi^I,\Phi^J\rangle$, where the $\tau_{IJ}$ (and hence the metric $\mathrm{Im}\tau_{IJ}$) are constants. Due to one-loop and instanton effects, $\mathcal{F}$ is no longer quadratic in the effective theory. Since $\tau$ varies on $U$, it cannot be holomorphic (except at a few singular points), single valued, and having a positive-definite imaginary part. The solution to this apparent contradiction is that each set of special coordinates and the expression of $\mathcal{F}$ is valid only in part of $U$. Solving the $N=2$ gauge theory at low energy means understanding the singularity of $U$ in the strong coupling regime and obtaining the explicit form of $\mathcal{F}$ or $\tau_{IJ}$ in various regions of the moduli space.

### Seiberg–Witten's Low-Energy Solution

We consider $N=2$ gauge theory with $G=\mathrm{SU}(2)$. The Cartan subalgebra is $\mathfrak{t}\cong\mathbb{C}$; each $a\in\mathbb{C}$ determines an element $\phi=(1/2)\left(\begin{smallmatrix}a & 0\\ 0 & -a\end{smallmatrix}\right)$ in $\mathfrak{t}$. The Weyl group $W\cong\mathbb{Z}_2$ acts on $\mathbb{C}$ by $a\mapsto\pm a$. The moduli space of classical vacua is the $u$-plane $\mathbb{C}/\mathbb{Z}_2$ parametrized by $u=\mathrm{tr}\,\phi^2=(1/2)a^2$. When $u\neq 0$, the gauge group is broken to U(1). The generator of $\mathbb{Z}_{4\check{h}}=\mathbb{Z}_8\subset\mathrm{U}(1)_R$ acts as $a\mapsto\sqrt{-1}a, u\mapsto -u$. The $\mathbb{Z}_8$ symmetry is broken to $\mathbb{Z}_4$; the quotient $\mathbb{Z}_2=\mathbb{Z}_8/\mathbb{Z}_4$ acts on the $u$-plane by $u\mapsto\pm u$.

Abelian gauge theory and $N=4$ supersymmetric gauge theory exhibit exact electric–magnetic duality in the sense that the quantum theories are identical if the coupling constant $\tau$ undergoes an $\mathrm{SL}(2,\mathbb{Z})$ transformation. Seiberg and Witten (1994a, b) proposed that this is so for the low-energy effective theory of the $N=2$ gauge theory. An $\mathrm{SL}(2,\mathbb{Z})$ transformation maps one description of the low-energy theory to another, exchanging electricity and magnetism. It is however not an exact duality of the full SU(2) theory. Rather, duality is in the ambiguity of the choice of the low-energy description. More precisely, $\tau$ is a section of a flat $\mathrm{SL}(2,\mathbb{Z})$ bundle over $U$. Thus, $\tau$ is multivalued and exists as a function in local charts only. So we must use different Lagrangians in different regions of the $u$-plane. Around the singularities where $\tau$ is not defined, nontrivial monodromy can appear.

Away from infinity, the electric theory is strongly interacting but the magnetic theory is infrared free. The dual field is $\tilde{a}=\mathrm{d}\mathcal{F}(a)/\mathrm{d}a$, and $\tau_{\mathrm{eff}}(u)=\mathrm{d}\tilde{a}/\mathrm{d}a$. The group $\mathrm{SL}(2,\mathbb{Z})$ is generated by

$$P=\begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix},\qquad S=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}$$
$$T=\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$$

To see its action on $\left(\begin{smallmatrix}\tilde{a}\\ a\end{smallmatrix}\right)$, we use the central extension of the $N=2$ super-Poincaré algebra. In the classical theory, the central charge is $Z=(n_{\mathrm{e}}+\tau n_{\mathrm{m}})a$ from the boundary terms at infinity. As the electric–magnetic duality transformation $S$ interchanges $n_{\mathrm{e}}$ and $n_{\mathrm{m}}$, we have for any $\gamma\in\mathrm{SL}(2,\mathbb{Z})$, $\gamma:(n_{\mathrm{m}},n_{\mathrm{e}})\mapsto(n_{\mathrm{m}},n_{\mathrm{e}})\gamma^{-1}$. When $n_m=0$, the classical formula $Z=n_{\mathrm{e}}a$ is valid. Invariance of $Z$ under $\mathrm{SL}(2,\mathbb{Z})$ requires that $Z=n_{\mathrm{m}}\tilde{a}+n_{\mathrm{e}}a$ at the quantum level and that $\mathrm{SL}(2,\mathbb{Z})$ acts on $\left(\begin{smallmatrix}\tilde{a}\\ a\end{smallmatrix}\right)$ homogeneously as a column vector.

When $u=(1/2)a^2$ is large, perturbation is reliable. The classical and one-loop results are $a(u)\sim\sqrt{2u}$, $\tilde{a}\sim(\sqrt{-1}/\pi)a\log a^2$. As $u$ goes around infinity, the fields transforms as $a\mapsto -a, \tilde{a}\mapsto -\tilde{a}+2a$. The monodromy is $M_\infty=PT^{-2}$. The mass $M$ of a monopole state is bounded by $M^2=P^\mu P_\mu\geq|Z|^2$, which is precisely the Bogomol'nyi bound. Now as a consequence of the $N=2$ supersymmetry, it receives no quantum corrections as long as supersymmetry is not broken at the quantum level. The states that saturate the bound are the BPS states. The BPS spectrum at $u\in U$ is a subset of $H_1(E_u,\mathbb{Z})\cong\mathbb{Z}^2$ containing the pairs $(n_{\mathrm{m}},n_{\mathrm{e}})$ realized by the dyon charges. Near infinity, the condition is that either $n_{\mathrm{e}}=\pm1, n_{\mathrm{m}}=0$ (for $W_\pm$ particles) or $n_{\mathrm{m}}=\pm1$ (for monopoles or dyons). This spectrum is invariant under the monodromy $M_\infty$.

The nontrivial holonomy at infinity implies the existence of at least one singularity at a finite value $u=u_0$, where extra particles become massless. Seiberg and Witten (1994a, b) propose that these

particles are collective excitations in the perturbative regime. Suppose along a path connecting $u_0$ and some base point near infinity, a monopole of charges $(\pm 1, n_e) = (0, 1)(T^{\mp n_e} S^{\pm 1})^{-1}$ becomes massless at $u_0$. Then by the renormalization group analysis and duality, the monodromy at $u_0$ is $M_{u_0} = (T^{\mp n_e} S^{\pm 1}) T^2 (T^{\mp n_e} S^{\pm 1})^{-1}$. It turns out that there are two singularities $u = \pm \Lambda^2$ with monodromies $M_{\Lambda^2} = ST^2 S^{-1}$ and $M_{-\Lambda^2} = (TS)T^2(TS)^{-1}$. The particles that become massless at $\pm\Lambda^2$ are of charges $(n_m, n_e) = (1, 0)$ and $(1, -1)$, respectively. The only BPS states in the strong coupling regime are those which become massless at the singularities; the others decay as $u$ deforms towards strong interaction.

The monodromies $M_{\pm\Lambda^2}, M_\infty$ (or any two of them) generate the subgroup $\Gamma(2)$. The family of elliptic curves with these monodromies can be identified with $y^2 = (x - \Lambda^2)(x + \Lambda^2)(x - u)$ called the Seiberg–Witten curve. The singularities are at $u = \pm\Lambda^2$ and $u = \infty$, where the curve degenerates. Let

$$\lambda = \frac{\sqrt{2}}{2\pi} \frac{y\, dx}{x^2 - \Lambda^4}$$

be the Seiberg–Witten differential (of second kind on the total space $E$). Then in a suitable basis $(\alpha, \beta)$ of $H_1(E_u/U, \mathbb{Z})$, we have $a = \int_\alpha \lambda, \tilde{a} = \int_\beta \lambda$. At a singularity, if $\nu = n_m \beta + n_e \alpha$ is a vanishing cycle, then the dyon of charges $(n_m, n_e)$ becomes massless. This is because its central charge is $Z = n_m \tilde{a} + n_e a = \int_\nu \lambda$. The monodromy at a singularity where $\nu$ is a vanishing cycle is given by the Picard–Lefshetz formula $M: \gamma \mapsto \gamma - 2(\gamma \cdot \nu)\nu$. At $u = \pm\Lambda^2$, the vanishing cycles are $\beta$ and $\beta - \alpha$, respectively.

We return to the $N = 1$ $SO(N_c)$ gauge theory with $N_f = N_c - 2$. At a generic point in the moduli space, the gauge group is broken to SO(2), which is abelian. Much of the above discussion applies to this case. By $N = 1$ supersymmetry, the effective coupling $\tau_{\text{eff}}$ is holomorphic in $M$ but is not single valued. In fact, $\tau_{\text{eff}}$ depends on $u = \det M$, which is invariant under the (anomaly free) $SU(N_f)$ symmetry. For large $u$, we have $e^{2\pi\sqrt{-1}\tau_{\text{eff}}} = \Lambda^{4N_c - 8}/u^2$ and the monodromy around infinity is $M_\infty = PT^{-2}$. On the other hand, a large expectation value of $M$ of rank $N_c - 3$ breaks the gauge group to SO(3) and the theory is the $N = 2$ theory discussed earlier. Using these facts, Intriligator and Seiberg (1995) identified the family of elliptic curves as $y^2 = x(x - 16\Lambda^{2N_c - 4})(x - u)$. There are two singularities with inequivalent physics. At $u = 0$, the monodromy is $ST^2 S^{-1}$. A pair of monopoles $\tilde{Q}^\pm$ becomes massless. They couple with $M$ through the superpotential $W \sim (2\mu)^{-1} M^{ij} \tilde{Q}_i \tilde{Q}_j$. At $u = 16\Lambda^{2N_c - 4}$, the monodromy is $(T^2 S)T^2(T^2 S)^{-1}$. A pair of dyons $E^\pm$ of charges $\pm 1$ become massless. The effective action is $W_{\text{eff}} \sim (u - 16\Lambda^{2N_c - 4})E^+ E^-$.

Topological gauge theory is a twisted version of $N = 2$ Yang–Mills theory in which the observables at high energy are the Donaldson invariants. The work of Seiberg and Witten (1994a, b) yields new insight to it and has a tremendous impact on the geometry of 4-manifolds. See Witten (1994) for the initial steps.

After the work of Seiberg and Witten (1994a, b), there has been much progress on theories with other gauge groups. If the gauge group is a compact Lie group of rank $r$, the $u$-plane is replaced by $\mathfrak{t}^{\mathbb{C}}/W$; the singularities are modified by quantum effects. The duality group is $Sp(2r, \mathbb{Z})$ or its subgroup of finite index, acting on the coupling matrix $\tau = (\tau_{IJ})$ by fractional linear transformations. For example, for $G = SU(N_c)$, the moduli space is parametrized by gauge invariants $u_2, \ldots, u_{N_c}$ defined by $\det(xI - \phi) = x^{N_c} - \sum_{i=2}^{N_c} u_i x^{N_c - i} = P_{N_c}(x, u_i)$. Classically, the singular locus is a simple singularity of type $A_{N_c - 1}$. At the quantum level, the singularity consists of two copies of such locus shifted by $\pm\Lambda^n$ in the $u_n$ direction. The monodromies correspond to a family of hyperelliptic curves $y^2 = P_{N_c}(x, u_i)^2 - \Lambda^{2N_c}$ of genus $N_c - 1$. The Seiberg–Witten differential is

$$\lambda = \frac{\sqrt{2}}{\pi\sqrt{-1}} \frac{\partial P_{N_c}(x, u_i)}{\partial x} \frac{x\, dx}{y} + \partial(\cdots)$$

The $N_c - 1$ independent eigenvalues $a^i$ of $\phi$ and their duals $\tilde{a}_i = \partial\mathcal{F}/\partial a^i$ are the periods of $\lambda$ along the $2N_c - 2$ homology cycles in the curve. For more details, the reader is referred to Klemm *et al.* (1995) and Argyes and Faraggi (1995).

## $N = 2$ QCD

$N = 2$ supersymmetric QCD is $N = 2$ Yang–Mills theory coupled to $N = 2$ matter. The latter consists of $N = 1$ superfields $Q$ that form a quarternionic representation $R$ of the gauge group $G$. The space $R$ has a $G$-invariant hyper-Kähler structure. The hyper-Kähler moment map $\mu_{\mathbb{H}}: R \to \mathfrak{g}^* \otimes \text{Im}\,\mathbb{H}$ consists of a real moment map $\mu_{\mathbb{R}}: R \to \mathfrak{g}^*$ for the Kähler structure and a complex moment map $\mu_{\mathbb{C}}: R \to (\mathfrak{g}^*)^{\mathbb{C}}$ for the holomorphic symplectic structure. As an $N = 1$ theory, the matter superfields are valued in $R \times \mathfrak{g}^{\mathbb{C}}$ with a $D$-term $D(Q, \Phi) = \mu_{\mathbb{R}}(Q) + (1/2\sqrt{-1})[\Phi, \bar{\Phi}]$ and a superpotential $W(Q, \Phi) = \sqrt{2}\langle \mu_{\mathbb{C}}(Q), \Phi \rangle + m(Q)$, where the mass term $m$ is a $G$-invariant quadratic form on $R$. The classical moduli space of vacua has two branches. On the Coulomb branch where $Q = 0$ and $\Phi \neq 0$, the unbroken gauge group is abelian and the

photons are massless. If $Q \neq 0$ exists in the flat directions, the gauge group is broken according to the value of $Q$; these are the Higgs branches. If $m=0$, the moduli space of classical vacua is the hyper-Kähler quotient $\mu_{\mathbb{H}}^{-1}(0)/G$. The branches of two types touch at the origin, where the full gauge group is restored, and at other subvarieties in $R$. The global symmetry is the subgroup of $U(R)$ that commutes with the $G$-action on $R$ and preserves $m$; it contains $U(2)_R$.

Quantum mechanically, such a theory is free from local gauge anomalies. Consistency under large gauge transformations puts a torsion condition on $R$, such as $\nu(R) = 0(\mathrm{mod}\, 2)$. Since $b_0 = 2\check{h} - (1/2)\nu(R)$, the theory is asymptotically free if $\nu(R) < 4\check{h}$. If $\nu(R) = 4\check{h}$, the quantum theory is scale invariant up to one-loop (and hence to all loops), and is expected to be so nonperturbatively. If $\nu(R) > 4\check{h}$, the quantum theory may not be defined but it can be the low-energy solution of another asymptotically free theory. Due to the axial anomaly, the $U(2)_R$ global symmetry reduces to the subgroup $SU(2)_R \times_{\mathbb{Z}_2} \mathbb{Z}_{4\check{h}-\nu(R)}$. The metric on the Coulomb branch can be corrected by quantum effects, but those on the Higgs branches do not change because of the uniqueness of the hyper-Kähler metric. In the quantum theory, the Higgs branches still touch the Coulomb branch, but the photons of the Coulomb branch are the only massless gauge bosons at the point where they meet.

When $G = SU(N_c)$ we take $N_f$ quarks $Q^i (i = 1, \ldots, N_f)$ in the fundamental representation and $N_f$ antiquarks $\tilde{Q}_i (i = 1, \ldots, N_f)$ in the complex-conjugate representation. The moment map is the same as in $N=1$ QCD whereas the superpotential is $W = \sqrt{2}\tilde{Q}_i \Phi Q^i + \sum_i m_i \tilde{Q}_i Q^i$. Consider the case $G = SU(2)$ as in Seiberg and Witten (1994b). Since $b_0 = 4 - N_f$, the asymptotically free theories have $N_f \leq 3$ whereas the $N_f = 4$ theory is scale invariant. As the representations on $Q^i$ and $\tilde{Q}_i$ are isomorphic, the classical global symmetry is $O(2N_f) \times U(2)_R$ when all $m_i = 0$. The appearance of the even number of fundamental representations is necessary for the consistency of the theory at the quantum level. The $U(1)_R$ symmetry is anomalous if $N_f \neq 4$. When $N_f > 0$, $SO(2N_f)$ is anomaly free, whereas $O(2N_f)/SO(2N_f) = \mathbb{Z}_2$ is anomalous. The anomaly free subgroup of $\mathbb{Z}_2 \times U(1)_R$ is $\mathbb{Z}_{4(4-N_f)}$. Its $\mathbb{Z}_2$ subgroup acts in the same way as $\mathbb{Z}_2 \subset Z(SO(2N_f))$. A nonzero expectation value of $u = \mathrm{tr}\,\phi^2$ further breaks the symmetry to $\mathbb{Z}_4$. The quotient group that acts effectively on the $u$-plane (the Coulomb branch) is $\mathbb{Z}_{4-N_f}$ if $N_f > 0$ and $\mathbb{Z}_2$ if $N_f = 0$. When $N_f = 4$, the $U(1)_R$ symmetry is anomaly free but $\mathbb{Z}_2 = O(8)/SO(8)$ is still anomalous.

The $N_f = 0$ theory is the $N=2$ pure gauge theory. In order to compare it to the $N_f > 0$ theories, we multiply $n_e$ by 2 so that it has integer values on $Q^i$ and $\tilde{Q}_i$, and divide $a$ by 2 to preserve the formula $Z = n_m \tilde{a} + n_e a$. The monodromies around the singularities become $M_{\Lambda^2} = STS^{-1}$, $M_{-\Lambda^2} = (T^2 S)T(T^2 S)^{-1}$, $M_\infty = PT^{-4}$. They generate the subgroup $\Gamma_0(4)$ of $SL(2,\mathbb{Z})$. The coupling constant is

$$\tau = \frac{\theta}{\pi} + \frac{8\pi\sqrt{-1}}{g^2}$$

The Seiberg–Witten curve is $y^2 = x^3 - ux^2 + (1/4)\Lambda_0^4 x$, related to the earlier one $y^2 = (x-u)(x^2 - \Lambda_0^4)$ by an isogeny. Here and below, $\Lambda_{N_f}$ is the dynamically generated scale.

For $N_f > 0$, we consider the case with zero bare masses. The simplest BPS-saturated states are the elementary quarks with mass $\sqrt{2}|a|$, which form the vector representation of $SO(2N_f)$. In addition, the quarks have fermion zero modes in the monopole background. When $n_m = 1$, each $SU(2)$ doublet of quarks has one zero mode. With $N_f$ hypermultiplet, there are $2N_f$ zero modes in the vector representation of $SO(2N_f)$. Upon quantization, the quantum states are in the spinor representation. So the flavor symmetry is really $Spin(2N_f)$. The spectrum may also include states with $n_m > 1$. For $N_f = 2, 3, 4$, the center $Z(Spin(2N_f))$ are $\mathbb{Z}_2 \times \mathbb{Z}_2$, $\mathbb{Z}_4$, $\mathbb{Z}_2 \times \mathbb{Z}_2$, whose generators act on states of charges $(n_m, n_e)$ by $((-1)^{n_e+n_m}, (-1)^{n_e})$, $\sqrt{-1}^{n_m+2n_e}$, $((-1)^{n_m}, (-1)^{n_e})$, respectively.

Suppose at a singularity on the $u$-plane, the low-energy theory is QED with $k$ hypermultiplets. Let $m_i$ be the bare mass and $S_i$, the $U(1)$ charge of the $i$th hypermultiplet. With the expectation value of $\phi$, the actual masses are $|\sqrt{2}a + m_i| (1 \leq i \leq k)$. As the states form a small representation of the $N=2$ algebra, the central charge is modified as $Z = n_m \tilde{a} + n_e a + S \cdot m/\sqrt{2}$, where $m = (m_1, \ldots, m_k)$ and $S = (S_1, \ldots, S_k)$. Under a duality transformation $M \in SL(2,\mathbb{Z})$, the column vector $(m/\sqrt{2}, \tilde{a}, a)$ is multiplied by a matrix of the form $\hat{M} = \begin{pmatrix} I_k & 0 \\ * & M \end{pmatrix}$. (For example, if $M = T$, $\hat{M}$ can be derived by one-loop analysis.) So the row vector $W = (S, n_m, n_e)$ transforms as $W \mapsto W\hat{M}^{-1}$. The transformation on $(n_m, n_e)$ is not homogeneous when there are hypermultiplets. This phenomenon persists even when all the bare masses $m_i$ are zero.

When $N_f = 1$, the global symmetry of the $u$-plane is $\mathbb{Z}_3$. There are three singularities related by this symmetry, where monopoles with charges $(n_m, n_e) = (1,0), (1,1)$, and $(1,2)$ become massless. The low-energy theory at each singularity is QED with a single light hypermultiplet. Besides the photon, no other flat directions exist. This is consistent with the absence of Higgs branch in the original theory. The monodromies at the singularities are $STS^{-1}$,

$(TS)T(TS)^{-1}, (T^2S)T(T^2S)^{-1}$, respectively, and the corresponding Seiberg–Witten family of curves is $y^2 = x^2(x-u) - (1/64)\Lambda_1^6$. The Seiberg–Witten differential is

$$\lambda = -\frac{\sqrt{2}}{4\pi}\frac{y\,\mathrm{d}x}{x^2}$$

When $N_f = 2$, there are two singularities related by the global symmetry $\mathbb{Z}_2$ of the $u$-plane. The massless states at one singularity have $(n_\mathrm{m}, n_\mathrm{e}) = (1,0)$ and form a spinor representation of SO(4) while those at the other have $(n_\mathrm{m}, n_\mathrm{e}) = (1,1)$ and form the other spinor representation. The low-energy theory at each singularity is QED with two light hypermultiplets. There are additional flat directions along which $\mathrm{SO}(4) \times \mathrm{SU}(2)_R$ is broken. They form the two Higgs branches that touch the $u$-plane at the two singularities rather than at the origin. The metric and pattern of symmetry breaking are the same as classically. The monodromies are $ST^2S^{-1}, (TS)T^2(TS)^{-1}$. The Seiberg–Witten curve is $y^2 = (x^2 - u) - (1/64)\Lambda_2^4)(x-u)$ and the differential is

$$\lambda = -\frac{\sqrt{2}}{4\pi}\frac{y\,\mathrm{d}x}{x^2 - \Lambda_2^4/64}$$

When $N_f = 3$, the $u$-plane has no global symmetry. There are two singularities. At one of them, a single monopole bound state with $(n_\mathrm{m}, n_\mathrm{e}) = (2,1)$ becomes massless and there are no other light particles. At the other singularity, the massless states have $(n_\mathrm{m}, n_\mathrm{e}) = (1,0)$ and form a (four-dimensional) spinor representation of SO(6) with a definite chirality. Thus, the low-energy theory is QED with four light hypermultiplets. Along the flat directions, the $\mathrm{SO}(6) \times \mathrm{SU}(2)_R$ symmetry is further broken. This corresponds to a single Higgs branch touching the $u$-plane at the singularity. Again, the metric on the Higgs branch is not modified by quantum effects. The monodromies at the two singularities are $(ST^2S)T(ST^2S)^{-1}$ and $ST^4S^{-1}$, respectively. The Seiberg–Witten curve is $y^2 = x^2(x-u) - (1/64)\Lambda_3^2(x-u)^2$ and the differential is

$$\lambda = \frac{\sqrt{2}}{\pi\Lambda_3}\log\left(y + \sqrt{-1}\frac{\Lambda_3}{8}\left(x - u - \frac{32}{\Lambda_3^2}x^2\right)\right)\mathrm{d}x$$

When $N_f = 4$, the theory is characterized by classical coupling constant $\tau$, and there are no corrections to $a = (1/2)\sqrt{2u}, \tilde{a} = \tau a$. There is only one singularity at $u = 0$, where the monodromy is $P$. Seiberg and Witten (1994b) postulate that the full quantum theory is $\mathrm{SL}(2,\mathbb{Z})$ invariant, just like the $N = 4$ pure gauge theory. The elementary hypermultiplet has $(n_\mathrm{m}, n_\mathrm{e}) = (0,1)$ and form the vector representation $v$ of SO(8). Fermion zero modes give rise to hypermultiplets with $(n_\mathrm{m}, n_\mathrm{e}) = (1,0), (1,1)$ that transform under the spinor representations $s$, $c$ of Spin(8). $\mathrm{SL}(2,\mathbb{Z})$ acts on the spectrum via a homomorphism onto the outer-automorphism group $S_3$ of Spin(8), which then permutes $v$, $s$, and $c$. So duality is mixed in an interesting way with the SO(8) triality. In $v$, $s$, and $c$, the center $\mathbb{Z}_2 \times \mathbb{Z}_2$ acts as $((-1)^{n_\mathrm{m}}, (-1)^{n_\mathrm{e}}) = (1,-1), (-1,1), (-1,-1)$, respectively. The full $\mathrm{SL}(2,\mathbb{Z})$ invariance predicts the existence of multimonopole bound states: for every pair of relatively prime integers $(p, q)$, there are eight states with $(n_\mathrm{m}, n_\mathrm{e}) = (p,q)$ that form a representation of Spin(8) on which the center acts as $((-1)^p, (-1)^q)$.

Solutions when the bare masses are nonzero are also obtained by Seiberg and Witten (1994b). The masses can be deformed to relate theories with different values of $N_f$. $N = 2$ QCD with a general classical gauge group has also been studied. By adding to these theories a mass term $m\,\mathrm{tr}\,\Phi^2$ that explicitly breaks the supersymmetry to $N = 1$, the dualities of Seiberg can be recovered. For $\mathrm{SU}(N_c), \mathrm{SO}(N_c)$ and $\mathrm{Sp}(2N_c)$ gauge groups, see Hanany and Oz (1995), Argyes *et al.* (1996), Argyes *et al.* (1997) and references therein.

*See also:* Anomalies; Brane Construction of Gauge Theories; Donaldson–Witten Theory; Duality in Topological Quantum Field Theory; Effective Field Theories; Electric–Magnetic Duality; Floer Homology; Gauge Theories from Strings; Gauge Theory: Mathematical Applications; Nonperturbative and Topological Aspects of Gauge Theory; Quantum Chromodynamics; Topological Quantum Field Theory: Overview; Supersymmetric Particle Models.

## Further Reading


Affleck I, Dine M, and Seiberg N (1984) Dynamical supersymmetric breaking in supersymmetric QCD. *Nuclear Physics B* 241: 493–534.

Argyes PC and Faraggi AE (1995) Vacuum structure and spectrum of $N = 2$ supersymmetric SU($n$) gauge theory. *Physical Review Letters* 74: 3931–3934.

Argyes PC, Plesser MR, and Seiberg N (1996) The moduli space of vacua of $N = 2$ SUSY QCD and duality in $N = 1$ SUSY QCD. *Nuclear Physics B* 471: 159–194.

Argyes PC, Plesser MR, and Shapere AD (1997) $N = 2$ moduli spaces and $N = 1$ dualities for SO($n_c$) and USp($2n_c$) superQCD. *Nuclear Physics B* 483: 172–186.

Freed DS (1999) Special Kähler manifolds. *Communications in Mathematical Physics* 203: 31–52.

Hanany A and Oz Y (1995) On the quantum moduli space of vacua of $N = 2$ supersymmetric SU($N_c$) gauge theories. *Nuclear Physics B* 452: 283–312.

Klemm A, Lerche W, Yankielowicz S, and Theisen S (1995) Simple singularities and $N = 2$ supersymmetric Yang–Mills theory. *Physics Letters B* 344: 169–175.

Intriligator K and Pouliot P (1995) Exact superpotential, quantum vacua and duality in supersymmetric SP($N_c$) gauge theories. *Physics Letters B* 353: 471–476.
Intriligator K and Seiberg N (1995) Duality, monopoles, dyons, confinement and oblique confinement in supersymmetric SO($N_c$) gauge theories. *Nuclear Physics B* 444: 125–160.
Seiberg N (1994) Exact results on the space of vacua of four-dimensional SUSY gauge theories. *Physical Review D* 49: 6857–6863.
Seiberg N (1995) Electric–magnetic duality in supersymmetric non-Abelian gauge theories. *Nuclear Physics B* 435: 129–146.
Seiberg N and Witten E (1994a) Electric–magnetic duality, monopole condensation, and confinement in $N=2$ supersymmetric Yang–Mills theory. *Nuclear Physics B* 426: 19–52.
Seiberg N and Witten E (1994b) Monopoles, duality and chiral symmetry breaking in $N=2$ supersymmetric QCD. *Nuclear Physics B* 431: 484–550.
Witten E (1994) Monopoles and four-manifolds. *Mathematical Research Letters* 1: 769–796.

**Semiclassical Approximation** *see* Stationary Phase Approximation; Normal Forms and Semiclassical Approximation

# Semiclassical Spectra and Closed Orbits

**Y Colin de Verdière**, Université de Grenoble 1, Saint-Martin d'Hères, France



## Introduction

The purpose of this article is to describe the so-called "semiclassical trace formula" (SCTF) relating the "spectrum" of a semiclassical Hamiltonian to the "periods of closed orbits" of its classical limit. SCTF formula expresses the asymptotic behavior as $\hbar \to 0$ ($\hbar = h/2\pi$) of the regularized density of states as a sum of oscillatory contributions associated to the closed orbits of the classical limit.

We will mainly present the case of the Schrödinger operator on a Riemannian manifold which contains the purely Riemannian case.

We start with a section about the history of the subject. We then give a statement of the results and a heuristic proof using Feynman integrals. This proof can be transformed into a mathematical proof which we will not give here. After that we describe some applications of the SCTF.

## About the History

SCTF has several origins: on one side, Selberg trace formula (1956) is an exact summation formula concerning the case of locally symmetric spaces; this formula was interpreted by H Huber as a formula relating eigenvalues of the Laplace operator and lengths of closed geodesics (also called the "lengths spectrum") on a closed surface of curvature –1.

On the other side, around 1970, two groups of physicists developed independently asymptotic trace formulas:

- M Gutzwiller for the Schrödinger operator, using the quasiclassical approximation of the Green function (the "van Vleck's formula"); it is interesting to note that the word "trace formula" is not written, but Gutzwiller instead speaks of a new "quantization method" (the old one being "Einstein–Brillouin–Keller (EBK)" or "Bohr–Sommerfeld rules").
- R Balian and C Bloch, for the eigenfrequencies of a cavity, use what they call a "multiple reflection expansion." They asked about a possible application to Kac's problem.

At the same time, under the influence of Mark Kac's famous paper "Can one hear the shape of a drum?," mathematicians became quite interested in inverse spectral problems, mainly using heat kernel expansions (for the state of the art around 1970, see Berger *et al.* (1971)).

The SCTF was put into its final mathematical form for the Laplace operator on closed manifolds by three groups of people around 1973–75:

- Y Colin de Verdière in his thesis was using the short-time expansion of the Schrödinger kernel and an approximate Feynman path integral. He proved that the spectrum of the Laplace operator determines generically the lengths of closed geodesics.
- J Chazarain derived the qualitative form of the trace for the wave kernel using Fourier integral operators.

- Using the full power of the symbolic calculus of Fourier integral operators, H Duistermaat and V Guillemin were able to compute the main term of the singularity from the Poincaré map of the closed orbit. Their paper became a canonical reference on the subject.

  After that, people were able to extend SCTF to:

- general semiclassical Hamiltonians (Helffer–Robert, Guillemin–Uribe, Meinrenken),
- manifolds with boundary (Guillemin–Melrose),
- surfaces with conical singularities and polygonal billiards (Hillairet), and
- several commuting operators (Charbonnel–Popov).

Recently, some researchers have remarked about the nonprincipal terms in the singularities expansion which come from the semiclassical Birkhoff normal form (Zelditch, Guillemin).

## Selberg Trace Formula

We consider a compact hyperbolic surface $X$. "Hyperbolic" means that the Riemannian metric is locally $(\mathrm{d}x^2+\mathrm{d}y^2)/y^2$ or is of constant curvature –1. Such a surface is the quotient $X=\boldsymbol{H}/\Gamma$ where $\Gamma$ is a discrete co-compact subgroup of the group of isometries of the Poincaré half-plane $\boldsymbol{H}$. Closed geodesics of $X$ are in bijective correspondence with nontrivial conjugacy classes of $\Gamma$. More precisely, the set of loops $C(S^1,X)$ splits into connected components associated to conjugacy classes and each component of nontrivial loops contains exactly one periodic geodesic.

**Theorem 1** (Selberg trace formula). *If $\rho$ is a real-valued function on $\mathbb{R}$ whose Fourier transform $\hat{\rho}$ is compactly supported and $\lambda_j=1/4+\mu_j^2$ is the spectrum of the Laplace operator on $X$, we have*:

$$\sum_{j=1}^{\infty}\rho(\mu\pm\mu_j)=\frac{A}{2\pi}\int_{\mathbb{R}}\rho(\mu+s)s\tanh\pi s\,\mathrm{d}s + \sum_{\gamma\in\mathcal{P}}\sum_{n=1}^{\infty}\frac{l_\gamma}{2\pi\sinh(nl_\gamma/2)} \times \mathrm{Re}(\hat{\rho}(nl_\gamma)\mathrm{e}^{\mathrm{i}n\mu l_\gamma})$$

*where $A$ is the area of $X$, $\mathcal{P}$ the set of primitive conjugacy classes of $\Gamma$ and, for $\gamma\in\mathcal{P}$, $l_\gamma$ is the length of the unique closed geodesic associated to $\gamma$.*

A nice recent presentation of the Selberg trace formula can be found in Marklof (2003).

## Semiclassical Schrödinger Operators on Riemannian Manifolds

If $(X, g)$ is a (possibly noncompact) Riemannian manifold and $V: X\to\mathbb{R}$ a smooth function which satisfies $\liminf_{x\to\infty}V(x)=E_\infty>-\infty$, the differential operator $\hat{H}=(1/2)\hbar^2\Delta+V$ is semibounded from below and admits self-adjoint extensions. For all those extensions, the spectrum is discrete in the interval $]-\infty,E_\infty[$ and eigenfunctions $\hat{H}\varphi_j=E_j\varphi_j$ are localized in the domain $V\le E_j$. If $X$ is compact and $V=0$, we recover the case of the Laplace operator.

We will denote this part of the spectrum by

$$\inf V<E_1(\hbar)<E_2(\hbar)\le\cdots\le E_j(\hbar)\le\cdots<E_\infty$$

For the Laplace operator, we have $E_j=\hbar^2\lambda_j$, where $\lambda_1\le\lambda_2\le\cdots\le\lambda_j\le\cdots$ is the spectrum of the Laplace operator.

The SCTF can also be derived the same way for Schrödinger operators with magnetic field. One can even extend it to Hamiltonian systems which are not obtained by Legendre transform from a regular Lagrangian. In this case, Morse indices have to be replaced by the more general Maslov indices.

## Classical Dynamics

### Newton Flows

Euler–Lagrange equations for the Lagrangian $\mathcal{L}(x,v):=(1/2)\|v\|_g^2-V(x)$ admit a Hamiltonian formulation on $T^\star X$ whose energy is given by $H=(1/2)\|\xi\|_g^2+V(x)$. We will denote by $X_H$ the Hamiltonian vector field

$$X_H:=\sum_j\frac{\partial H}{\partial\xi_j}\partial_{x_j}-\frac{\partial H}{\partial x_j}\partial_{\xi_j}$$

Preservation of $H$ by the dynamics shows immediately that the Hamiltonian flow $\Phi_t$ restricted to $H<E_\infty$ is complete.

The Hamiltonian $H$ is the "classical limit" of $\hat{H}$; in more technical terms, $H$ is the semiclassical principal symbol of $\hat{H}$.

If $V=0$, $H=(1/2)g^{ij}\xi_i\xi_j$ and the flow is the geodesic flow.

### Periodic Orbits

**Definition 1** *A* periodic orbit $(\gamma,T)$ (also denoted p.o.) of the Hamiltonian $H$ consists of an orbit $\gamma$ of $X_H$ which is homeomorphic to a circle and a nonzero real number $T$ so that $\Phi_T(z)=z$ for all $z\in\gamma$. We will denote by $T_0(\gamma)>0$ (the primitive period) the smallest $T>0$ for which $\Phi_T(z)=z$.

If $(T, E)$ are given, $W_{T,E}$ is the set of $z$'s so that $H(z) = E$ and $\Phi_T(z) = z$.

- The (linear) Poincaré map $\Pi_\gamma$ of a p.o. $(\gamma, T)$ with $H(\gamma) = E$: we restrict the flow to $S_E := \{H = E\}$ and take a hypersurface $\Sigma$ inside $S_E$ transversal to $\gamma$ at the point $z_0$. The associated return map $P$ is a local diffeomorphism fixing $z_0$. Its linearization $\Pi_\gamma := P'(z_0)$ is the linear Poincaré map, an inversible (symplectic) endomorphism of the tangent space $T_{z_0}\Sigma$.
- The Morse index $\iota(\gamma)$: p.o. $(\gamma, T)$ is a critical point of the action integral $\int_0^T \mathcal{L}(\gamma(s), \dot\gamma(s))\,ds$ on the manifold $C^\infty(\mathbb{R}/T\mathbb{Z}, X)$. It always has a finite Morse index (Milnor 1967) which is denoted by $\iota(\gamma)$. For general Hamiltonian systems, the Morse index is replaced by the Conley–Zehnder index.
- The nullity index $\nu(\gamma)$ is the dimension of the space of infinitesimal deformations of the p.o. $\gamma$ by p.o. of the same energy and period. We always have $\nu(\gamma) \geq 1$ and $\nu(\gamma) = 1 + \dim\ker(\mathrm{Id} - \Pi_\gamma)$.

**Example 1** (Geodesic flows)

- Riemannian manifold with sectional curvature $< 0$: in this case, we have for all periodic geodesics $\iota(\gamma) = 0, \nu(\gamma) = 1$.
- Generic metrics: for a generic metric on a closed manifold, we have $\nu(\gamma) = 1$ for all periodic geodesics.
- For flat tori of dimension $d$: we have $\iota(\gamma) = 0$ and $\nu(\gamma) = d$.
- For sphere of dimension 2 with constant curvature: if $\gamma_n$ is the $n$th iterate of the great circle, we have $\iota(\gamma_n) = 2|n|$ and $\nu(\gamma_n) = 3$.

It is a beautiful result of J-P Serre that any pair of points on a closed Riemannian manifold are endpoints of infinitely many distinct geodesics. Counting geometrically distinct periodic geodesics is much harder especially for simple manifolds like the spheres. It is now known that every closed Riemannian manifold admits infinitely many geometrically distinct periodic geodesics (at least, in some cases, for generic metrics, (Berger 2000 chap. V). There exists significant knowledge concerning more general Hamiltonian systems as well.

### Nondegeneracy

There are several possible nondegeneracy assumptions. They can be formulated "à la Morse–Bott" (critical point of action integrals) or purely symplectically.

**Definition 2** Two submanifolds $Y$ and $Z$ of $X$ intersect cleanly iff $Y \cap Z$ is a manifold whose tangent space is the intersection of the tangent spaces of $Y$ and $Z$.

Fixed points of a smooth map are clean if the graph of the map intersects the diagonal cleanly.

**Definition 3** We will denote by (ND) the following property of the p.o. $(\gamma_0, T_0)$: the fixed points of the associated (nonlinear) Poincaré map $P$ are clean.

The set $W_{T,E}$ is ND if all p.o.'s inside are ND. $W_{T,E}$ is then a manifold of dimension $\nu(\gamma)$.

**Example 2**

- *Generic case*: $\nu = 1$; (ND) is equivalent to "1 is not an eigenvalue of the linear Poincaré map." In this case, we can deform the p.o. smoothly by moving the energy. This family of p.o.'s is called a cylinder of p.o.'s. The period $T(E)$ is then a smooth function of $E$.
- *Completely integrable systems*: $\nu = d$; (ND) is then a consequence of the so-called "isoenergetic KAM condition": assuming the Hamiltonian is expressed as $H(I_1, \ldots, I_d)$ using action-angle coordinates, this condition is that the mapping $I \to [\nabla H(I)]$ from the energy surface $H = E$ into the projective space is a local diffeomorphism. This condition implies that Diophantine invariant tori are not destructed by a small perturbation of the Hamiltonian.
- *Maximally degenerated systems*: it is the case where all orbits are periodic ($\nu = 2d - 1$). For example, the two-body problem with Newtonian potential and the geodesic flows on compact rank-1 symmetric spaces.

### Canonical Measures and Symplectic Reduction

Under the hypothesis (ND), the manifold $W_{T,E}$ admits a canonical measure $\mu_c$, invariant by $\Phi_t$. In the case $\nu = 1$, this measure is given by $|dt|/\sqrt{\det(\mathrm{Id} - \Pi)}$.

By using a Poincaré section, it is enough to understand the following fact: if $A$ is a symplectic linear map, the space $\ker(\mathrm{Id} - A)$ admits a canonical Lebesgue measure.

We start with the following construction: let $L_1$ and $L_2$ be two Lagrangian subspaces of a symplectic space $E$ and $\omega_j, j = 1, 2$, be half-densities on $L_j$, denoted by $\omega_j \in \Omega^{1/2}(L_j)$. If $W = L_1 \cap L_2$, we have the following canonical isomorphisms: $\Omega^{1/2}(L_j) = \Omega^{1/2}(W) \otimes \Omega^{1/2}(L_j/W)$. So $\Omega^{1/2}(L_1) \otimes \Omega^{1/2}(L_2) = \Omega^{1/2}(L_1/W) \otimes \Omega^{1/2}(L_2/W) \otimes \Omega^1(W)$. $M_j = L_j/W$ are two Lagrangian subspaces of the reduced space $W^o/W$ whose intersection is 0. Hence, by using the Liouville measure on it, we get $\Omega^{1/2}(M_1) \otimes \Omega^{1/2}(M_2) = \mathbb{C}$. Hence, we get a density $\omega_1 \star \omega_2$ on $W$. It turns out that the previous calculation is one of the main algebraic pieces of the symbolic calculus of

Fourier integral operators and the density $\omega_1 \star \omega_2$ arises in stationary-phase computations.

The graph of a symplectic map is equipped with a half-density by pullback of the Liouville half-density. So we can apply the previous construction to the intersection of the graph of $A$ and the graph of the identity map.

### Actions

**Definition 4** If $(\gamma, T)$ is a p.o., we define the following quantity which is called action of $\gamma$:

$$A(\gamma) = \int_\gamma \xi \, dx$$

In the (ND) case, $A(\gamma)$ is constant on each connected component of $W_{T,E}$.

In the generic case and if $T'(E) \neq 0$ (cylinder of p.o.), p.o.'s of the cylinder are also parametrized by $T$ (i.e., we note by $\gamma_E$ the p.o. of the cylinder of energy $E$ and $\gamma_T$ the p.o. of period $T$). If $a(E) = A(\gamma_E)$ and $b(T) = -\int_0^T \mathcal{L}(\gamma_T(s), \dot{\gamma}_T(s)) ds$, $a(E)$ and $b(T)$ are Legendre transforms of each other.

## Playing with Spectral Densities

We will define the "regularized spectral densities." The general idea is as follows: we want to study an $\hbar$-dependent sequence of numbers $E_j(\hbar)$ (a spectrum) in some interval $[a, b]$. We introduce a non negative function $\rho \in \mathcal{S}(\mathbb{R})$ which satisfies $\int \rho(t) dt = 1$, and also $D_{\rho,\varepsilon,\hbar}(E) = \sum \rho_\varepsilon(E - E_j)$, where $\rho_\varepsilon(E) = \varepsilon^{-1}\rho(E/\varepsilon)$. It gives the analysis of the spectrum at the scale $\varepsilon$. Of course, we will adapt the scaling $\varepsilon$ to the small parameter $\hbar$. If the scaling is of the size of the mean spacing of the spectrum, we will get a very precise resolution of the spectrum.

The general philosophy is:

- If $\hbar$ is the semiclassical parameter of a semiclassical Hamiltonian, the mean spacing of the eigenvalues is of order $\hbar^d$ (Weyl's law). The trace formula gives the asymptotic behavior of $D_{\rho,\varepsilon,\hbar}(E)$ for $\varepsilon \sim \hbar$ (and hence $\varepsilon >> \overline{\Delta E}$ except if $d = 1$). This behavior is not "universal" and thus contains a significant amount information of (in our case, on periodic trajectories).
- Better resolution of the spectrum needs the use of the long-time behavior of the classical dynamics and is conjecturally universal. It means that eigenvalues seen at very small scale behave like eigenvalues of an ensemble of random matrices, the most common one being the Wigner Gaussian orthogonal ensemble (GOE) and Gaussian unitary ensemble (GUE).

We fix some interval $[a, b]$ with $b < E_\infty$.

We define $D(E) := \sum_{a \leq E_j \leq b} \delta(E_j)$ as the sum of Dirac measures at the points $E_j$ and its $\hbar$-Fourier transform as

$$Z(t) = \text{trace}'(e^{-it\hat{H}/\hbar}) := \sum{}' \exp(-itE_j/\hbar) \quad [1]$$

where $\sum'$ is the sum over $E_j \in [a, b]$.

The Duistermaat–Guillemin trick relates the previous behavior to asymptotics of the regularized density of eigenvalues. Let us give a function $\rho \in \mathcal{S}(\mathbb{R})$ so that $\hat{\rho}(t) = \int e^{-itE} \rho(E) dE$ is compactly supported and

$$\hat{\rho}(t) = 1 + O(t^\infty), \quad t \to 0 \quad [2]$$

(all moments of $\rho$ vanish). We introduce, for $E \in [a, b]$, $D_\rho(E) := \sum_j' \frac{1}{\hbar} \rho(E - E_j/\hbar)$. $D_\rho(E)$ is independent modulo $O(\hbar^\infty)$ of $a, b$. We have

$$D_\rho(E) = \frac{1}{2\pi\hbar} \int \hat{\rho}(t) Z(t) \, dt$$

The idea is now to start from a semiclassical approximation of $U(t) = e^{-it\hat{H}/\hbar}$ and to insert it into eqn [1]. We need only a uniform approximation of $U(t)$ for $t \in \text{Support}(\hat{\rho})$. From the asymptotic expansion of $Z(t)$, we will deduce the asymptotic expansion of $D_\rho$, the regularized eigenvalue density.

## The Smoothed Density of States

The following statement expressing the smoothed density of eigenvalues is the main result of the subject. Under the (ND) assumption, it gives the existence of an asymptotic expansion for $D_\rho(E)$:

**Theorem 2** *If $E$ is not a critical value of $H$ and the* (ND) *condition is satisfied for all p.o.'s of energy $E \in [a, b]$ and period inside the support of $\hat{\rho}$,*

$$D_\rho(E) = D_{\text{Weyl}}(E) + \sum D_{W(T,E)} + O(\hbar^\infty) \quad [3]$$

*where*:

(i)
$$D_{\text{Weyl}}(E) = (2\pi\hbar)^{-d} \left( \sum_{j=0}^{\infty} a_j(E) \hbar^j \right)$$

*with* $a_0(E) = \int_{H=E} dL/dH$

(ii) *The sum is over all the manifolds $W_{T,E}$ so that* $T \in \text{Support}(\hat{\rho})$.

(iii)
$$D_{W(T,E)} = \frac{\varepsilon}{(2\pi i\hbar)^{(\nu(\gamma)+1)/2}} e^{-i\iota(\gamma)\pi/2} \times e^{iA(\gamma)/\hbar} \sum_{j \geq 0} b_j(E) \hbar^j$$

*with*

$$b_0(E) = \hat{\rho}(T)\int_{W_{T,E}} d\mu_c$$
$$\varepsilon = \begin{cases} 1 & \text{if } T'(E) > 0 \\ i & \text{if } T'(E) < 0 \end{cases}$$

*If* $\nu(\gamma) = 1$, *we get* $b_0 = \hat{\rho}(T_\gamma)T_0|\det(\mathrm{Id} - \Pi_\gamma)|^{-1/2}$.

## The Weyl Expansion

If Support($\hat{\rho}$) is contained in $[-T_{\min}, T_{\min}]$, where $T_{\min}$ is the smallest period of a p.o. $\gamma$ with $H(\gamma) = E$, and, if $E$ is not a critical value of $H$, formula [3] reduces to

$$D_\rho(E) \sim (2\pi\hbar)^{-d}\left(\sum_{j=0}^{\infty} a_j(E)\hbar^j\right)$$

From the previous formula, it is possible to deduce the following estimates:

**Theorem 3** *If a, b are not critical values of H*:

$$\#\{j | a \le E_j(\hbar) \le b\} = (2\pi\hbar)^{-d}\text{volume}(a \le H \le b)(1 + O(\hbar))$$

This remainder estimate is optimal and was first shown in rather great generality by Hörmander (1968).

## Derivation from the Feynman Integral

### The Feynman Integral

R Feynman (Feynman and Hibbs 1965) found a geometric representation of the propagator, that is, the kernel $p(t,x,y)$ of the unitary group $\exp(-it\hat{H}/\hbar)$ using an integral (FPI := Feynman path integral) on the manifold $\Omega_{t,x,y} := \{\gamma : [0,t] \to X | \gamma(0) = x, \gamma(t) = y\}$ of paths from $x$ to $y$ in the time $t$; if $\mathcal{L}(\gamma, \dot{\gamma})$ is the Lagrangian, we have, for $t > 0$:

$$p(t,x,y) = \int_{\Omega_{t,x,y}} \exp\left(\frac{i}{\hbar}\int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s))\, ds\right)|d\gamma|$$

where $|d\gamma|$ is a "Riemannian measure" on the manifold $\Omega_{t,x,y}$ with the natural Riemannian structure.

There is no justification FPI as a useful mathematical tool. Nevertheless, FPI gives good heuristics and right formulas.

### The Trace and Loop Manifolds

Let us try a formal calculation of the partition function and its semiclassical limit. We get

$$Z(t) = \int_X |dx| \int_{\Omega_{x,x,t}} \exp\left(\frac{i}{\hbar}\int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s))ds\right)|d\gamma|$$

If we denote by $\Omega_t$ the manifold of paths $\gamma : \mathbb{R}/t\mathbb{Z} \to X$, (loops) and we apply Fubini (sic !), we get

$$Z(t) = \int_{\Omega_t} \exp\left(\frac{i}{\hbar}\int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s))\, ds\right)|d\gamma|$$

### The Semiclassical Limit

We want to apply stationary phase in order to get the asymptotic expansion of $Z(t)$; critical points of $J_t : \Omega_t \to \mathbb{R}$ are the p.o.'s of the Euler–Lagrange flow and hence of the Hamiltonian flow of period $t$. We require the ND assumption (Morse–Bott), the Morse index, and the determinant of the Hessian:

1. The ND assumption is the original Morse–Bott one in Morse theory: we have smooth manifolds of critical points and the Hessian is transversally ND.
2. The Morse index is the Morse index of the action functional on periodic loops: $L(\gamma) := \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s))ds$.
3. The Hessian is associated to a periodic Sturm–Liouville operator for which many regularizations have already been proposed.

In this manner, we get a sum of contributions given by the components $W_{j,t}$ of $W_t$:

$$Z_j(t) = (i\hbar)^{-\nu_j/2} e^{(i/\hbar)L(\gamma)} c_j(\hbar)$$

with $c_j(\hbar) \sim \sum_{l=0}^{\infty} c_{j,l}\hbar^l$ and

$$c_{j,0} = \frac{e^{-i\mu(\pi/2)}}{|\delta|^{1/2}}$$

where $\mu$ is the Morse index and $\delta$ is a regularized determinant.

## The Integrable Case

As observed by Berry–Tabor, the trace formula in this case comes from Poisson summation formula using action-angle coordinates. Asymptotic of the eigenvalues to any order can then be given in the so-called quantum integrable case by Bohr–Sommerfeld rules.

## The Maximally Degenerated Case

Let us assume that $(X, g)$ is a compact Riemannian manifold for which all geodesics have the same smallest period $T_0 = 2\pi$. Then we have the following clustering property:

**Theorem 4** *There exists some constant C and some integer $\alpha$ so that*

(i) *the spectrum of $\Delta$ is contained in the union of the intervals*

$$I_k = \left[\left(k + \frac{\alpha}{4}\right)^2 - C, \left(k + \frac{\alpha}{4}\right)^2 + C\right], \quad k = 1, 2, \ldots$$

(ii) $N(k) = \#\text{Spectrum}(\Delta) \cap I_k$ *is a polynomial function of k for k large enough.*

The property (ii) is consequence of the trace formula.

## Applications to the Inverse Spectral Problem

We will now restrict ourselves to the case of the Laplace operator on a compact Riemannian manifold $(X, g)$. The main result is as follows:

**Theorem 5** (*Colin de Verdière*). *If X is given, there exists a generic subset $\mathcal{G}_X$, in the sense of Baire category, of the set of smooth Riemannian metrics on X, so that, if $g \in \mathcal{G}_X$, the length spectrum of $(X, g)$ can be recovered from the Laplace spectrum. The set $\mathcal{G}_X$ contains all metrics with $<0$ sectional curvature and (conjecturally) all metrics with $<0$ sectional curvature.*

We can take for $\mathcal{G}_X$ the set of metrics for which all periodic geodesics are nondegenerate and the length spectrum is simple.

Some cancelations may occur between the asymptotic expansions of two ND periodic trajectories with the same actions if the Morse indices differ by 2 mod 4.

## The Case with Boundary

If $(X, g)$ is a smooth compact manifold with boundary, one introduces the broken geodesic flow by extending the trajectories by reflection on the boundary. SCTFs have been extended to that case by Guillemin and Melrose. Periodic geodesics which are transversal to the boundary contribute to the density of states in the same way as for periodic manifolds. Periodic geodesics inside the boundary are in general accumulation of periodic geodesics near the boundary: their contributions is therefore very complicated analytically.

## Bifurcations

Let us denote by $C_H \subset \mathbb{R}^2_{T,E}$, the set of pairs $(T,E)$ for which $W_{T,E}$ is not empty. The previous results apply to the "smooth" part of the set $C_H$. Among other interesting points are points $(0,E)$ with critical value $E$ of $H$ (Brummelhuis–Paul–Uribe) and points corresponding to bifurcation of p.o. when moving the energy.

Detailed studies of some of these points have been done, for example, the results of suitable applications of the theory of singularities of functions of finitely many variables, their deformations (catastrophe theory), and applications to stationary-phase method, and a significant body of knowledge on these subjects now exists.

## SCTF and Eigenvalue Statistics

One of the main open mathematical problems is: "can one really use appropriate forms of the SCTF as quantization rules and use it in order to derive eigenvalues statistics?"

This problem is related to the fine-scale study of the eigenvalue spacings ($\varepsilon << \hbar$). It is one of the important unsolved problems of the so-called "quantum chaos." Many people think that progress in this field will allow us to solve the Bohigas–Giannoni–Schmit conjecture: "if the geodesic flow is hyperbolic, eigenvalue distribution follows random matrix asymptotics."

*See also:* Billiards in Bounded Convex Domains; *h*-Pseudodifferential Operators and Applications; Quantum Ergodicity and Mixing of Eigenfunctions; Random Matrix Theory in Physics; Regularization for Dynamical Zeta Functions; Resonances.

## Further Reading


Balian R and Bloch C (1970) Distribution of eigenfrequencies for the wave equation in a finite domain I. *Annals of Physics* 60: 401.

Balian R and Bloch C (1971) Distribution of eigenfrequencies for the wave equation in a finite domain II. *Annals of Physics* 64: 271.

Balian R and Bloch C (1972) Distribution of eigenfrequencies for the wave equation in a finite domain III. *Annals of Physics* 69: 76.

Berger M (2000) *Riemannian Geometry During the Second Half of the Twentieth Century*. University Lecture Series, vol. 17. Providence, RI: American Mathematical Society.

Berger M, Gauduchon P, and Mazet E (1971) *Le spectre d'une variété riemannienne compacte*. Berlin–Heidelberg–New York: Springer LNM.

Colin de Verdière Y (1973) Spectre du Laplacien et longueurs des géodésiques périodiques II. *Compositio Mathematica* 27: 159–184.

Duistermaat J (1976) On the Morse index in variational calculus. *Advances in Mathematics* 21: 173–195.

Duistermaat J and Guillemin V (1975) The spectrum of positive elliptic operators and periodic geodesics. *Inventiones Mathematicae* 29: 39–79.

Feynman R and Hibbs A (1965) *Quantum Mechanics and Path Integrals*. New-York: McGraw-Hill.
Gutzwiller M (1971) Periodic orbits and classical quantization conditions. *Journal of Mathematical Physics* 12: 343–358.
Gutzwiller M (1990) *Chaos in Classical and Quantum Mechanics*. Berlin–Heidelberg–New York: Springer.
Hörmander L (1968) The spectral function of an elliptic operator. *Acta Mathematica* 121: 193–218.
Marklof J (2003) Selberg's trace formula: an introduction. Lectures given at the International School "Quantum Chaos on Hypebolic Manifolds," Schloss Reisensburg, Gunsburg, Germany, 4–11 October 2003. To appear in Springer LNP, Berlin–Heidelberg, New York. http://fr.arxiv.org
Milnor J (1967) *Morse Theory*. Annals of Mathematics Studies no. 51. Princeton, NJ: Princeton University Press.
Selberg A (1956) Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series. *Journal of the Indian Mathematical Society* 20: 47–87.

# Semilinear Wave Equations

**P D'Ancona**, Università di Roma "La Sapienza," Rome, Italy



## Introduction

A semilinear wave equation is an equation of the form

$$\Box u = F(u, u'), \quad u : \Omega \subseteq \mathbb{R} \times \mathbb{R}^n \to \mathbb{R} \qquad [1]$$

where $F : \mathbb{R}^{n+2} \to \mathbb{R}$ is a smooth function, the d'Alembert operator $\Box$ is defined as

$$\Box = D_t^2 - D_{x_1}^2 \cdots - D_{x_n}^2, \quad D_t = \frac{\partial}{\partial t}, D_{x_j} = \frac{\partial}{\partial x_j} \qquad [2]$$

and $u'$ denotes the vector of all first-order derivatives of $u$:

$$u' = (D_t u, D_{x_1} u, \ldots, D_{x_n} u) \equiv (u_t, u_{x_1}, \ldots, u_{x_n})$$

Sometimes the term "semilinear" is used in a more restrictive sense and refers to the special class of equations

$$\Box u = f(u) \qquad [3]$$

The very particular case $f(u) = -mu, m > 0$, corresponds to the Klein–Gordon equation, used to model relativistic particles. True nonlinear terms of the form $f(u) = -mu - u^3, m \geq 0$ (meson equation), or $f(u) = -\sin u$ (sine-Gordon equation) have been proposed as models of self-interacting fields with a local interaction. Notice that for the physical applications it is natural to consider complex-valued functions $u(t, x)$; in the general case of eqn [1], this actually means that we are considering a $2 \times 2$ system in $\Re u$ and $\Im u$. However, the natural physical requirement of gauge invariance restricts the possible nonlinearities to the functions satisfying the condition

$$f(e^{i\theta} u) = f(u) e^{i\theta}, \quad \forall \theta \in \mathbb{R} \qquad [4]$$

Thus, in particular $f(0) = 0$ and we see that $f$ must be of the form $f(u) = g(|u|^2)u$ for some $g$. Since the gauge-invariant wave equation

$$\Box u = g(|u|^2)u \qquad [5]$$

has essentially the same properties as the real-valued equation [3], it is not too restrictive to study only real-valued functions as we shall mostly do in the following.

The more general equations of the form [1], involving the derivatives of $u$, are encountered in several physical theories, including the nonlinear $\sigma$-models and general relativity.

However, beyond the concrete physical applications, eqn [1] is important since it is a simplified but relevant model of much more general equations and systems of mathematical physics; despite its simple structure, the semilinear wave equation presents already all the main difficulties and phenomena of nonlinear wave interaction, and it represents an ideal laboratory for such problems.

In this article we plan to give a concise but, as far as possible, comprehensive review of the main research directions concerning eqn [1], and in particular we shall focus on the global existence of both large and small nonlinear waves, and the problem of local existence for low-regularity solutions. A large part of the theory extends to nonlinear perturbations of the form $\Box u = F(u, u', u'')$ and to the fully nonlinear case; we have no space here to give an account of these developments and we must refer the reader to the books and papers cited in the "Further reading" section.

## Classical Results

Equations [1] and [3] are hyperbolic with respect to the variable $t$. This is a precise way of stating that the "correct" problem for it is an initial-value problem (IVP) with data at some fixed time, or

more generally on some spacelike surface: this means that we assign two functions $u_0(x), u_1(x)$, called the "initial data," and we look for a function $u(t,x)$ satisfying the IVP:

$$\Box u = F(u,u'), \quad u(0,x) = u_0(x), \quad u_t(0,x) = u_1(x) \qquad [6]$$

This setting is in agreement with the physical picture of an evolution problem: the data represent the complete state of a system at a fixed time, and they uniquely determine the evolution of the system, which is described by the differential equation.

This rough statement of the problem is sufficient when working with smooth functions, as in the classical approach. By purely classical methods, that is, energy inequalities and nonlinear estimates, it is not difficult to prove the following local existence result, where $H^k = H^k(\mathbb{R}^n)$ denotes the Sobolev space of functions with $k$ derivatives in $L^2(\mathbb{R}^n)$:

**Theorem 1** *Assume $F$ is $C^\infty$. Let $(u_0,u_1) \in H^k \times H^{k-1}$ for some $k > 1 + n/2$. Then there exists a time $T = T(\|u_0\|_{H^k} + \|u_1\|_{H^{k-1}}) > 0$ such that problem [6] has a unique solution belonging to $(u,u_t) \in C([-T,T]; H^k) \times C([-T,T]; H^{k-1})$.*

*If $F = F(u)$ depends only on $u$, the result holds for all $k > n/2$.*

**Proof** We decided to include a sketchy but complete proof of this result since it shows the basic approach to nonlinear wave equations: many results of the theory, even some of the most delicate ones, are obtained by suitable variations of the contraction method, and are similar in spirit to this classical theorem.

Assume for a moment that the equation is linear so that $F = F(t,x)$ is a given smooth function of $(t,x)$. For the linear equation $\Box u = F$, we can construct a solution $u$ using explicit formulas. Moreover, $u$ satisfies the energy inequality

$$E_k(t) \le E_k(0) + \int_0^t \|F(s,\cdot)\|_{H^{k-1}}\, ds \qquad [7]$$

where the energy $E_k(t)$ is defined as

$$E_k(t) = \|u(t,\cdot)\|_{H^k} + \|u_t(t,\cdot)\|_{H^{k-1}} \qquad [8]$$

Now we introduce the space $X_T = C([-T,T]; H^k) \cap C^1([-T,T]; H^{k-1})$, the space $Y_T = C([-T,T]; H^{k-1})$, the mapping $\Phi : F \to u$ that takes the function $F(t,x)$ into the solution of $\Box u = F$ (with fixed data $u_0, u_1$), and the mapping $\Psi(u) = F(u,u')$ which is the original right-hand side of the equation.

The energy inequality tells us that $\Phi$ is bounded from $Y_T$ to $X_T$. Actually, for $M$ large enough with respect to $E_k(0)$ (the $H^k$ norm of the data), $\Phi$ takes any ball $B_Y(0,N)$ of $Y_T$ into the ball $B_X(0, M + NT)$ of $X_T$. Moreover, if we apply [7] to the difference of two equations $\Box u = F$ and $\Box v = G$, we also see that $\Phi$ is Lipschitz continuous from $Y_T$ to $X_T$, with a Lipschitz constant $CT$.

On the other hand, $\Psi(u) = F(u,u')$ takes $X_T$ to $Y_T$, provided $k > 1 + n/2$; we can even say that it is Lipschitz continuous from $B_X(0,M)$ to $B_Y(0, C(M))$ for some function $C(M)$, with a Lipschitz constant $C_1(M)$ also depending on $M$. This follows easily from Moser type estimates like

$$\|F(u,u')\|_{H^{k-1}} \le \phi(\|u\|_{L^\infty})\|u\|_{H^k}, \quad k > \frac{n}{2} + 1$$

or

$$\|F(u)\|_{H^k} \le \phi(\|u\|_{L^\infty})\|u\|_{H^k}, \quad k > \frac{n}{2}$$

Now it is easy to conclude: the composition $\Phi \circ \Psi$ maps $X_T$ into itself, and actually is a contraction of $B_X(0,M)$ into itself provided $M$ is large enough with respect to the data, and $T$ is small enough with respect to $M$. The unique fixed point is the required solution. $\Box$

The wave operator has an additional important property called the finite speed of propagation, which can be stated as follows: given the IVP

$$\Box u = 0, \qquad u(0,x) = u_0(x), \qquad u_t(0,x) = u_1(x)$$

if we modify the data "outside" a ball $B(x_0,R) \subset \mathbb{R}^n$, the values of the solution inside the cone

$$K(x_0,R) = \{(t,x) : t \ge 0,\ |x - x_0| < R - t\}$$

do not change. Notice that $K(x_0,R)$ is the cone with basis $B(x_0,R)$ and tip $(R,x_0)$; the slope of its mantle represents the speed of propagation of the signals, which for the wave operator $\Box$ is equal to 1. The property extends without modification to the semilinear problem [6], at least for the smooth solutions given by Theorem 1. Actually, it is not difficult to modify the proof of the theorem to work on cones instead of bands $[-T,T] \times \mathbb{R}^n$; in other words, given a ball $B = B(x_0,R)$, we can assign two data $u_0 \in H^k(B)$, $u_1 \in H^{k-1}(B) (k > n/2 + 1)$ and prove the existence of a local solution on the cone $K(x_0,R)$ for some time interval $t \in [0,T]$.

In general, the finite speed of propagation allows us to localize in space most of the results and the estimates; as a rule of thumb, we expect that what is true on a band $[0,T] \times \mathbb{R}^n$ should also be true on any truncated cone $K(x_0,R) \cap \{0 \le t \le T\}$.

## Symmetries

The linear wave equation can be written as the Euler–Lagrange equation of a suitable Lagrangian. This is still true for the semilinear perturbations of the form

$$\square u + f(u) = 0 \qquad [9]$$

Indeed, denoting with $F(s) = \int_0^s f(\sigma)\,d\sigma$ the primitive of $f$, the Lagrangian of [9] is

$$\mathcal{L}(u) = \iint \left[ -\frac{1}{2}|u_t|^2 + \frac{1}{2}|\nabla_x u|^2 + F(u) \right] dt\,dx \qquad [10]$$

The functional $\mathcal{L}$ is not positive definite; hence, the variational approach gives only weak results. However, this point of view allows us to apply Noether's principle: any invariance of the functional is related to a conservation law of the equation. These conserved quantities can also be obtained by taking the product of the equation by a suitable multiplier, although this method is far from obvious in many cases. We describe here this circle of ideas briefly.

The functional $\mathcal{L}$ is invariant under the Poincaré group, generated by time and space translations and the Lorentz transformations ($\lambda > 1, c \neq 0$):

$$t \mapsto \frac{\lambda t - x_j/c}{\sqrt{\lambda^2 - 1}}, \qquad x_j \mapsto \frac{\lambda x_j - ct}{\sqrt{\lambda^2 - 1}} \qquad [11]$$

The infinitesimal generators of the translations are simply the partial derivatives $D_t$ and $D_{x_j}$. The Lorentz transformations can be decomposed as a rotation followed by a boost, and indeed a corresponding complete set of infinitesimal generators are the operators

$$\Gamma_{jk} = x_j D_k - x_k D_j, \qquad \Gamma_j = x_j D_t + t D_j \qquad [12]$$

All the operators in the Poincaré group commute with $\square$ exactly.

The conservation law related to time translations (time derivative) is the fundamental "conservation of energy"

$$E(t) = \int \left[ \frac{1}{2}u_t^2 + \frac{1}{2}|\nabla_x u|^2 + F(u) \right] dx = E(0) \qquad [13]$$

while spatial translations (spatial derivatives) lead to the conservation of momenta

$$\int u_t u_{x_j}\, dx = \text{const.}, \quad j = 1, \ldots, n$$

On the other hand, infinitesimal rotations and boost [12] are connected to the conservation of angular momenta

$$\int [x_k D_j u - x_j D_k u] \cdot D_t u\, dx = \text{const.}, \quad j,k = 1, \ldots, n \qquad [14]$$

and

$$\int [x_k e(u) + D_k u\, D_t u]\, dx = \text{const.}, \quad k = 1, \ldots, n \qquad [15]$$

where

$$e(u) = \tfrac{1}{2}u_t^2 + \tfrac{1}{2}|\nabla_x u|^2 + F(u) \qquad [16]$$

is the energy density.

The Poincaré group does not exhaust the invariance properties of the free wave equation. Among the other transformations which commute or almost commute with $\square$, we mention the spacetime dilations and inversions (which together with translations and Lorentz transformations generate the larger conformal group), the scaling $u \mapsto \lambda u$, the spatial dilations, and, in the complex-valued case, the gauge transformation $u \mapsto e^{i\theta} u$. In this way several useful conservation laws can be obtained, including the conformal energy identities of K Morawetz.

## Strichartz Estimates

Energy estimates are very useful tools but they have some major shortcomings. The main one is clearly the large number of derivatives necessary to estimate the nonlinear term. This is why the modern theory of semilinear wave equations relies mainly on different tools, which go under the umbrella name of Strichartz estimates and express the decay properties of solutions when measured in $L^p$ or related norms. In this section we summarize these estimates in their most general form, and try to give a feeling of the techniques involved.

Consider the following IVP for a homogeneous linear wave equation:

$$\square u = 0, \qquad u(0,x) = 0, \qquad u_t(0,x) = f(x) \qquad [17]$$

The conservation of energy states that

$$\|u_t(t,\cdot)\|_{L^2}^2 + \|\nabla_x u(t,\cdot)\|_{L^2}^2 \equiv \|f\|_{L^2}^2 \qquad [18]$$

for all times $t$. Thus, we see that $L^2$-type norms of the solution do not decay. The interesting fact is that if we measure the solution $u$ in a different $L^p$-norm, $p > 2$, the norm decays as $t \to \infty$, and the decay is fastest for the $L^\infty$-norm.

To appreciate the dispersive phenomena at their best, let us assume that the Fourier transform of the data is localized in an annulus of order 1:

$$\operatorname{supp} \hat{f}(\xi) \subset \{1/2 \leq |\xi| \leq 2\} \qquad [19]$$

Then the corresponding solution $u(t,x)$ has the same property, and we see that

$$\|u\|_{L^2} = \|\hat{u}\|_{L^2} \le 2\||\xi|\hat{u}\|_{L^2} \equiv 2\|\nabla u\|_{L^2} \le 4\|u\|_{L^2}$$

We condense the last line in the shorthand notation

$$\|u\|_{L^2} \simeq \|\nabla u\|_{L^2}$$

We shall also write

$$\|v\|_X \lesssim \|w\|_Y \iff \|v\|_X \le C\|w\|_Y \quad \text{for some } C$$

We can now rewrite the conservation of energy [20] in a very simple form; for localized data (and hence a localized solution) as in [19], we have

$$\|u(t,\cdot)\|_{L^2} \lesssim \|f\|_{L^2} \qquad [20]$$

The basic $L^\infty$-estimate for a solution of [17] with localized data as in [19] is simply

$$\|u(t,\cdot)\|_{L^\infty} \lesssim t^{-(n-1)/2}\|f\|_{L^1} \qquad [21]$$

This estimate is well known since the 1960s; it can be proved easily by several techniques, notably by the stationary-phase method. Property [21] measures the fact that as time increases, the total energy of the solution remains constant but spreads over a region of increasing volume, due to the propagation of waves. If we interpolate between [20] and [21], we obtain the full set of dispersive estimates

$$\|u(t,\cdot)\|_{L^q} \lesssim t^{-(n-1)(1/2-1/q)}\|f\|_{L^p}$$
$$\frac{1}{q}+\frac{1}{p}=1, \quad 2 \le q \le \infty \qquad [22]$$

Recall that we are working with localized solutions on the annulus $|\xi| \sim 1$; it is easy to extend the above estimates to general solutions by a rescaling argument, exploiting the fact that, if $u(t,x)$ is a solution of the homogeneous wave equation, $u(\lambda t, \lambda x)$ is also a solution for any constant $\lambda$. Indeed, if $\hat{f}$ (and hence $\hat{u}$) is supported in the annulus $2^{j-1} \le |\xi| \le 2^{j+1}$, $j \in \mathbb{Z}$, by rescaling [21], we obtain

$$\|u(t,\cdot)\|_{L^\infty} \lesssim t^{-(n-1)/2}2^{j(n-1)/2}\|f\|_{L^1} \qquad [23]$$

If $f$ is any smooth function, not localized in frequency, we can still write it as a series

$$f = \sum_{j\in\mathbb{Z}} f_j$$

where $\operatorname{supp} \hat{f}_j \subset \{2^{j-1} \le |\xi| \le 2^{j+1}\}$. The quantity

$$\|f\|_{\dot{B}^s_{1,1}} = \sum_{j\in\mathbb{Z}} 2^{js}\|f_j\|_{L^1}$$

is by definition the $\dot{B}^s_{1,1}$ Besov norm of $f$. Thus, summing the estimates [23] over $j$, we conclude that a general solution of [17] satisfies the dispersive estimate

$$\|u(t,\cdot)|_{L^\infty} \lesssim t^{-(n-1)/2}\|f\|_{\dot{B}^{(n-1)/2}_{1,1}} \qquad [24]$$

The Strichartz estimates can be obtained as a consequence of the above dispersive estimates, plus some subtle functional analytic arguments. In the general form we give here, they were proved by J Ginibre and G Velo, and in the most difficult endpoint cases by Keel and T Tao. The solution of the homogeneous problem [17] studied above can be written as

$$u(t,x) = \frac{\sin(t|D|)}{|D|}f, \quad |D| \equiv \mathcal{F}^{-1}|\xi|\mathcal{F}$$

(here $\mathcal{F}$ denotes the Fourier transform). On the other hand, the solution of the complete nonhomogeneous problem

$$\Box u = F(t,x), \quad u(0,x) = u_0, \quad u_t(0,x) = u_1(x) \qquad [25]$$

can be written by Duhamel's formula as

$$u(t,x) = \frac{\partial}{\partial t}\frac{\sin(t|D|)}{|D|}u_0 + \frac{\sin(t|D|)}{|D|}u_1 + \int_0^t \frac{\sin((t-s)|D|)}{|D|}f\,ds$$

and we see that the above estimates [22] apply to all the operators appearing here. If we consider problem [25] and we assume that the data $F(t,x), u_0, u_1$ are localized in frequency so that $\hat{F}(t,\xi), \hat{u}_0, \hat{u}_1$ have support in the annulus $|\xi| \sim 1$, the Strichartz estimate takes the following form:

$$\|u\|_{L^p_I L^q} \lesssim \|u_0\|_{L^2} + \|u_1\|_{L^2} + \|F\|_{L^{\tilde{p}'}_I L^{\tilde{q}'}} \qquad [26]$$

Here the dimension is $n \ge 2$; $L^p_I L^q$ denotes the space with norm

$$\|u\|_{L^p_I L^q} = \left(\int_I \|u(t,\cdot)\|^p_{L^q(\mathbb{R}^n)}\,dt\right)^{1/p}, \quad I = [0,T]$$
$$\text{or } I = \mathbb{R}$$

the indices $p, q$ satisfy the conditions

$$\frac{1}{p}+\frac{1}{q}\frac{n-1}{2} \le \frac{1}{2}\frac{n-1}{2},$$
$$p \in [2,\infty], \ (n,p,q) \ne (3,2,\infty) \qquad [27]$$

while $\tilde{p}, \tilde{q}$ satisfy an identical condition (and $p'$ denotes the conjugate index to $p$). The constant in inequality [26] is uniform with respect to the interval $I$.

To get the most general form of the estimates, some additional function space trickery is required. As before, a simple rescaling argument extends estimate [26] to the case of data $F, u_0, u_1$, whose spatial Fourier transforms are localized in the annulus $2^{j-1} \leq |\xi| \leq 2^{j+1}$; we obtain

$$2^{j(1/p+n/q)}\|u\|_{L_I^p L^q} \lesssim 2^{jn/2}\|u_0\|_{L^2} + 2^{j(n/2-1)}\|u_1\|_{L^2} + 2^{j(1/\tilde{p}'+n/\tilde{q}'-2)}\|F\|_{L_I^{\tilde{p}'} L^{\tilde{q}'}}$$

Finally, if the data are arbitrary, we may decompose them as series of localized functions, and summing the corresponding estimates we obtain the general Strichartz estimates for the wave equation [25]: for all $(p, q)$ and $(\tilde{p}, \tilde{q})$ as in [27],

$$\|u\|_{L_I^p \dot{B}^{1/p+n/q}_{q,2}} \lesssim \|u_0\|_{\dot{H}^{n/2}} + \|u_1\|_{\dot{H}^{n/2-1}} + \|F\|_{L_I^{\tilde{p}'} \dot{B}^{1/\tilde{p}'+n/\tilde{q}'-2}_{\tilde{q}',2}} \quad [28]$$

Here, given a decomposition $f = \sum_{j\in\mathbb{Z}} f_j$, the homogeneous Besov and Sobolev norms are defined, respectively, by the identities (obvious modification for $r=\infty$):

$$\|f\|^r_{\dot{B}^s_{q,r}} = \sum_{j\in\mathbb{Z}} 2^{jsr}\|f_j\|^r_{L^q},$$
$$\|u\|_{\dot{H}^s} = \||\xi|^s\hat{u}\|_{L^2} \simeq \|u\|_{\dot{B}^s_{2,2}}$$

It is easy to convert the estimates [28] into a form that uses only the more traditional norms

$$\|f\|_{\dot{H}^s_q} \equiv \||D|^s f\|_{L^q}, \qquad |D|^\sigma \equiv \mathcal{F}^{-1}|\xi|^\sigma\mathcal{F}$$

since by the Besov–Sobolev embedding we have

$$\dot{B}^s_{q,2} \subseteq \dot{H}^s_q \text{ for } 2 \leq q < \infty,$$
$$\dot{B}^s_{q,2} \supseteq \dot{H}^s_q \text{ for } 1 < q \leq 2$$

Notice that if we apply to the equation and the data the operator $|D|^\sigma = \mathcal{F}^{-1}|\xi|^\sigma\mathcal{F}$, which commutes with $\square$, the Strichartz estimate [28] can be rewritten in an apparently more general form:

$$\|u\|_{L_I^p \dot{B}^{1/p+n/q+\sigma}_{q,2}} \lesssim \|u_0\|_{\dot{H}^{n/2+\sigma}} + \|u_1\|_{\dot{H}^{n/2-1+\sigma}} + \|F\|_{L_I^{\tilde{p}'} \dot{B}^{1/\tilde{p}'+n/\tilde{q}'-2+\sigma}_{\tilde{q}',2}} \quad [29]$$

In particular, it is possible to choose the indices in such a way that no derivatives appear on $u$ and $F$: this choice gives

$$\|u\|_{L^p(\mathbb{R}^{n+1})} \lesssim \|u_0\|_{\dot{H}^{1/2}} + \|u_1\|_{\dot{H}^{-1/2}} + \|F\|_{L^{p'}(\mathbb{R}^{n+1})}$$
$$p = \frac{2(n+1)}{n-1}$$

which is the estimate originally proved by Strichartz.

## Global Large Waves

As for ordinary differential equations (ODEs), the local solutions constructed in Theorem 1 can be extended to a maximal time interval $[0, T^*]$, and a natural question arises: are these maximal solutions global, that is, is $T^* = \infty$?

For generic nonlinearities and large data, the answer is negative; in a dramatic way, in general the norm $\|u(t, \cdot)\|_{L^\infty}$ is unbounded as $t\uparrow T^* < \infty$. The reason for this is simple: using the finite speed of propagation, we can localize the equation and work on a cone; then if we take constant functions as initial data, the solution inside the cone does not depend on $x$, and the equation restricted to the cone effectively reduces to an ODE:

$$\square u = f(u) \iff y''(t) = f(y), \quad y(t) \equiv u(t,x) \quad [30]$$

By this remark it is elementary to construct solutions of the IVP [6] that blow up in a finite time.

This construction does not apply if the equation has some positive conserved quantity. Indeed, consider a general gauge-invariant equation

$$\square u + g(|u|^2)u = 0, \quad u(0,x) = u_0(x), \quad u_t(0,x) = u_1(x) \quad [31]$$

for some smooth function $g(s)$. Writing $G(s) = \int_0^s g(\sigma)\,d\sigma$, multiplying the equation by $\overline{u}_t$, and integrating over $\mathbb{R}^n$, it is easy to check that the nonlinear energy

$$E(t) = \int \left[|u_t|^2 + |\nabla_x u|^2 + G(|u|^2)\right] dx \equiv E(0) \quad [32]$$

is constant in time, provided the solution $u$ is smooth enough. When $G(s)$ has no definite sign, we can proceed as above and construct solutions that blow up in finite time; this is usually called the "focusing" case. However, if we assume that $G(s) \geq 0$ ("defocusing" case), the energy $E(t)$ is non-negative. The corresponding ODE, which is $y'' + g(y^2)y = 0$, has only global solutions, and one may guess that also the solutions of [31] can be extended to global ones.

This innocent-looking guess turns out to be one of the most difficult problems of the theory of nonlinear waves, and is actually largely unsolved at present.

The only general result for eqns [31] is Segal's theorem, stating that the IVP has always a global weak solution:

**Theorem 2** *Let $g(s)$ be a $C^1$ non-negative function on $[0, +\infty)$, write $G(s) = \int_0^s g(\sigma)\,d\sigma$ and assume that for some constant C*

$$sg(s^2) \leq CG(s^2), \qquad \lim_{s\to+\infty} G(s) = +\infty \quad [33]$$

*Then for any* $(u_0, u_1) \in H^1 \times L^2$ *such that* $G(|u_0|^2) \in L^1$, *the IVP* [31] *has a global solution* $u(t, x)$ *in the sense of distributions, such that* $u' \in L^\infty(\mathbb{R}, L^2(\mathbb{R}^n))$ *and* $F(u) \in L^\infty(\mathbb{R}, L^1(\mathbb{R}^n))$.

The proof (see Shatah and Struwe (1998)) is delicate but elementary in spirit: by truncating the nonlinear term, we can approximate the problem at hand with a sequence of problems with global solution; then the conservation law [32] yields some extra compactness, which allows us to extract a subsequence converging to a solution of the original equation.

Thus we see that, despite its generality, this result does not shed much light on the difficulties of the problem. Indeed, the weak solution obtained might not be unique, nor smooth, and in these questions the real obstruction to solving [31] is hidden.

Notice that in the one-dimensional case $n = 1$ the solution is always unique and smooth when the data are smooth, since in this case $E(t)$ controls the $L^\infty$-norm of $u$. For higher dimensions $n \geq 2$, something more can be proved if we assume that the nonlinear term has a polynomial growth:

$$sg(s^2) = |s|^{p-1}s \quad \text{for } s \text{ large, } p > 1 \qquad [34]$$

In particular, the defocusing wave equation with a power nonlinearity

$$\Box u + |u|^{p-1}u = 0 \qquad [35]$$

has been studied extensively. Notice that when $p$ is close to 1, the term $|u|^{p-1}u$ becomes singular near 0; this introduces additional difficulties in the problem; for this reason, it is better to consider a smooth term as in [34].

We can summarize the best-known results concerning [31] under [34] as follows. Let $p_0(n)$ be the number

$$p_0(1) = p_0(2) = \infty$$
$$p_0(n) = 1 + \frac{4}{n-2} \quad \text{for } n \geq 3$$

Then

- in the subcritical case $1 \leq p < p_0(n)$, for any data $(u_0, u_1) \in H^1 \times L^2$, there exists a unique solution $u \in C(\mathbb{R}; H^1)$ such that $u' \in C(\mathbb{R}; L^2)$;
- the same result holds in the critical case $p = p_0(n)$ for $n \geq 3$; and
- when $3 \leq n \leq 7, 1 \leq p \leq p_0(n)$, the solution is smoother if the data are smoother.

These results have been achieved in the course of more than 30 years through the works of several authors (it is indispensable to mention at least the names of K Jörgens, I Segal, W Strauss, W von Wahl, P Brenner, H Pecher, J Ginibre, G Velo, R Glassey and the more recent contributions of J Shatah, M Struwe, L Kapitanski, M Grillakis, omitting many others). Actually modern proofs are remarkably simple, and are based again on a variation of the fixed-point argument. Roughly speaking, the linear equation $\Box u + g(|v|^2)v = 0$ defines a mapping $v \mapsto u$; the Strichartz estimates localized on a cone imply that this mapping is Lipschitz continuous in suitable spaces, the Lipschitz constant being estimated by the nonlinear energy of the solution restricted to the cone. In order to show that this mapping is actually a contraction, it is sufficient to prove that the localized energy tends to zero near the tip of the cone, that is, it cannot concentrate at a point. Once this is known, it is easy to continue the solution beyond any maximal time of existence and prove the global existence and uniqueness of the solution.

In the supercritical case $p > p_0(n)$, very little is known at present; there is some indication that the problem is much more unstable than in the subcritical case (Kumlin, Brenner, Lebeau), and there is some numerical evidence in the same direction.

## Global Small Waves

It was noted already in the 1960s (Segal, Strauss) that the equation in dimension $n \geq 2$

$$\Box u = f(u), \quad u(0,x) = \varepsilon u_0(x), \quad u_t(0,x) = \varepsilon u_1(x)$$
$$f(u) = O(|u|^\gamma) \quad \text{for } u \sim 0$$

with small data can be considered as a perturbation of the free wave equation and admits global solutions. The phenomenon may be regarded as follows: the wave operator tends to spread waves and reduce their size (see [21]); the nonlinear term tends to concentrate the peaks and make them higher, but at the same time it makes small waves smaller. If the rate of dispersion is fast enough, the initial data are small enough, and the power of the nonlinear term is high enough, the peaks have no time to concentrate, and the solution quickly flattens out to 0. Notice that in dimension 1 there is no dispersion, and this kind of mechanism does not occur.

It was, however, F John who initiated the modern study of this question by giving the complete picture in dimension 3: for the IVP

$$\Box u = |u|^\gamma, \quad u(0,x) = \varepsilon u_0(x)$$
$$u_t(0,x) = \varepsilon u_1(x), \quad n = 3$$

he proved that, for fixed $u_0, u_1 \in C_0^\infty$,

- if $\gamma > 1 + \sqrt{2}$ and $\varepsilon$ is small enough, the solution is global and
- if $1 < \gamma < 1 + \sqrt{2}$ and the data are not identically zero, the solutions blow up in a finite time for all $\varepsilon$ (i.e., the $L^\infty$-norm is unbounded).

Later Schaeffer proved that blow-up occurs also at the critical value $\gamma = 1 + \sqrt{2}$.

W Strauss guessed the correct critical value for all dimensions – $\gamma_0(n)$ is the positive root of the algebraic equation

$$\left(\frac{n-1}{2}\gamma - \frac{n+1}{2}\right)\gamma = 1$$

and conjectured that the same picture as in dimension 3 is valid for all dimensions $n \geq 2$.

Soon Sideris proved that, for $1 < \gamma < \gamma_0(n)$ and the quite general and small data, one always has blow-up. Also it was proved by Klainerman, Shatah, Christodoulou, and others that the positive part of the conjecture was true for $\gamma > \gamma_0(n)$, with a small gap near the critical value. The gap was closed by Georgiev, Lindblad, Sogge, who proved global existence for all $\gamma > \gamma_0(n)$. We also mention that the solution at the critical value $\gamma = \gamma_0(n)$ always seems to blow up; this is settled for low dimension (Schaeffer, Yordanov, Zhang and others), but the question is still not completely clear for large dimensions.

This problem has spurred a great deal of creativity, eventually leading to very fruitful results: the different approaches have proved useful in a variety of problems, sometimes quite different from the original semilinear equation. We mention a few:

- The weighted estimates of F John are estimates of the solution in spacetime $L^p$ norms with weights of the form $(1 + |t| + |x|)^\alpha (1 + ||t| - |x||)^\beta$. An extension of this method was also used in the final complete proof of the conjecture.
- The vector field approach of S Klainerman. If we regard energy estimates as norms generated by the plain derivatives, it is natural to extend them to more general norms generated by vector fields commuting, or quasicommuting, with the wave operator. The conservation of energy expressed in these generalized norms has a built-in decay that allows us to prove global existence of small waves. This circle of ideas led very far, and we might even regard Christodoulou and Klainerman's proof of the stability of Minkowski space for the Einstein equation as an extreme consequence of this approach.
- The normal forms of J Shatah. The idea is to apply a nonlinear (and nonlocal) transformation to the equation in order to increase the power $\gamma$. This method is effective for a variety of equations, including the semilinear wave, Klein–Gordon, and Schrödinger equations.
- The conformal transform method of D Christodoulou. The Penrose transform takes the wave operator on $\mathbb{R}^{1+n}$ to the wave operator on a bounded subset of $\mathbb{R} \times \mathbb{S}^n$, the so-called Einstein diamond (here $\mathbb{S}^n$ is the $n$-dimensional sphere). Thanks to the fact that a problem of global existence is converted into a problem of local existence, the proof reduces to showing that the lifespan of the local solution becomes large enough to cover the whole diamond when $\varepsilon$ decreases.

A similar theory has been developed for the more general semilinear equation

$$\Box u = F(u, u'), \quad F(u, u') = O(|u, u'|^\gamma) \text{ for } u \sim 0$$

but the results are less complete. The general picture is similar: for $\gamma \geq 2$ when $n \geq 4$, and for $\gamma \geq 3$ when $n = 3$, one has global small solutions, while for $\gamma$ close to 1 one in general has blow-up.

A very interesting phenomenon in this context was discovered by S Klainerman: some nonlinearities with a special structure, called "null structure," behave better than the others. This structure is clearly related to the wave operator, and in the end it can be precisely explained in terms of interaction of waves in phase space. We illustrate these ideas in the most interesting special case. Consider the equation in three dimensions

$$\Box u = F(D_t u, D_x u), \quad F = O(|u'|^\gamma), \ n = 3$$

In the "cubic" case $\gamma = 3$, one has global existence for all data small enough. On the other hand, in the "quadratic" case $\gamma = 2$, it is possible to construct examples where the solution blows up in a finite time no matter how small the data. Now, assume that the nonlinear term has the following structure:

$$F(u') = aQ_0(u') + \sum_{0 \leq j < k \leq 3} c_{jk} Q_{jk}(u') + O(|u'|^3) \quad [36]$$

which is called a "null structure". Here $a, c_{jk}$ are constants, and the quadratic forms $Q$ are the following:

$$Q_0(u') = |D_t u|^2 - |D_{x_1 u}|^2 - |D_{x_2 u}|^2 - |D_{x_3 u}|^2 \quad [37]$$

$$Q_{0j}(u') = D_t u \cdot D_{x_j} u - D_{x_j} u \cdot D_t u, \quad j = 1, 2, 3 \quad [38]$$

$$Q_{jk}(u') = D_{x_j} u \cdot D_{x_k} u - D_{x_k} u \cdot D_{x_j} u \quad j, k = 1, 2, 3, \ j < k \quad [39]$$

Then the problem has a global solution for all small enough data. The extensions and applications of this idea are very wide (see the "Further reading" section for further information). Another situation where the null structure plays an important role is discussed in the next section.

## Low Regularity

Theorem 1, although optimal in the classical framework, is not satisfactory for a few reasons. From a physicist's point of view, requiring $n/2+1$ derivatives of the data is not meaningful, since the measurable quantities involve only low-order derivatives, the most important one being the energy, that is, the $H^1$-norm of the solution. Moreover, the wave equation has a rich set of conserved quantities, symmetries and decay properties which may be useful to prove stronger results, and in particular the global existence. However, many of these structures appear only at a low-regularity level ($H^1$ or even $L^p$); in order to exploit them it is essential to work with low-regularity solutions.

As an example, if we were able to prove Theorem 1 for $k=1$, then we could deduce that the local solutions can be extended to global ones in all cases when the $H^1$-norm is conserved. For instance, this would allow us to solve globally the equations of the form

$$\Box u + G'(|u|^2)u = 0, \quad G(s) \geq 0$$

The problem of the lowest value of $s$ such that a unique local solution exists in $H^s$ is quite difficult, and still not completely solved. In order to state the results we precise the definition of solution as follows: the IVP is said to be locally well posed in $H^s$, if, for all $(u_0, u_1)$ in a bounded set $B$ of $H^s \times H^{s-1}$, there exist a $T>0$, a Banach space $X_T$ (depending on $B$) continuously embedded in $C([0,T]; H^s)$, and a unique solution $u \in X_T$, such that the map $(u_0, u_1) \mapsto u$ is continuous from $B$ to $X_T$.

For the wave equation with a power nonlinearity

$$\Box u = |u|^p \qquad [40]$$

or more generally

$$\Box u = F(u), \quad F(u) = 0$$
$$|F(u) - F(v)| \leq C|u-v|(|u|^{p-1} + |v|^{p-1}) \qquad [41]$$

the picture is almost complete. Indeed, by using the scaling

$$t \mapsto \lambda t, \qquad x \mapsto \lambda x \qquad [42]$$

and the Lorentz transformation

$$t \mapsto \frac{\lambda t - x_1}{\sqrt{\lambda^2 - 1}}, \qquad x \mapsto \frac{t - \lambda x_1}{\sqrt{\lambda^2 - 1}} \qquad [43]$$

it is possible to show by explicit constructions that

- the equation is not locally well posed for $p(n/2 - s) \leq (n/2 + 2 - s)$ (scaling) and
- the equation is not locally well posed for $p(n/4 + 1/4 - s) \leq n/4 + 5/4 - s$ (Lorentz).

On the positive side, local well-posedness has been almost fully proved in the complementary region of indices, with the exception of a tiny spot near the endpoint $s=0$, $p=(n+5)/(n+1)$ where the problem is still open (and the conjecture is that the equation is ill posed for indices in that region). These results are due to several authors, among the others we cite C Kenig, G Ponce, L Vega, H Lindblad, C Sogge, L Kapitanski, and T Tao.

When the nonlinearity depends also on the first-order derivatives of $u$, the situation becomes more complex. In the general case, the best result available is still the local existence theorem (Theorem 1); the only possible refinement is the use of fractional Sobolev spaces $H^s$, but in general local solvability only holds for $s > n/2+1$. If we assume that $F = F(u')$ is a quadratic form in the first-order derivatives, a clever use of Strichartz estimates allows us to prove local solvability down to $s > n/2 + 1/2$ for $n \geq 3$ and $s > 7/4$ for $n=2$ (Ponce and Sideris).

However, exactly as in the case of the small nonlinear waves examined in the previous section, if the nonlinear term has a null structure the result can be improved. Indeed, when $F(u')$ is a combination of the forms [37]–[39], then local solvability and uniqueness can be proved for all $s > n/2$, as in the case of a nonlinear term of the type $F(u)$. This result is due to Klainerman, Machedon, and Selberg. Again, the proof is based on a variation of the contraction method; the additional ingredient here is the use of suitable function spaces, which are the counterpart for the wave equation of the spaces used by Bourgain in the study of the nonlinear Schrödinger equation. The norm of these spaces is defined as follows:

$$\|u\|_{H^{s,\theta}} \equiv \|\langle\xi\rangle^s \langle|\tau| - |\xi|\rangle^\theta \widetilde{u}(\tau,\xi)\|_{L^2(\mathbb{R}^{n+1})}$$

where $\langle\xi\rangle = (1+|\xi|^2)^{1/2}$ and $\widetilde{u}$ is the spacetime Fourier transform of $u(t,x)$. The wave operator can be regarded as a spacetime Fourier multiplier of the form $\tau^2 - |\xi|^2 = (|\tau| - |\xi|)(|\tau| + |\xi|)$, and we see that "inverting" the operator $\Box$ has a regularizing effect in the scale of $H^{s,\theta}$ spaces, since it decreases both

$s$ and $\theta$ by one unit. Substantiating this formal argument and complementing it with suitable estimates for the nonlinear term requires some hard work, which is contained in the theory of bilinear estimates developed by Klainerman and his school.

*See also:* Evolution Equations: Linear and Nonlinear; Symmetric Hyperbolic Systems and Shock Waves; Wave Equations and Diffraction.

## Further Reading


Choquet-Bruhat Y (1988) Global existence for nonlinear $\sigma$-models, Rend. Sem. Mat. Univ. Pol. Torino, Fascicolo speciale "Hyperbolic equations," 65–86.

D'Ancona P, Georgiev V, and Kubo H (2001) Weighted decay estimates for the wave equation. *Journal of Differential Equations* 177: 146–208.

Georgiev V, Lindblad H, and Sogge CD (1997) Weighted Strichartz estimates and global existence for semilinear wave equations. *American Journal of Mathematics* 119: 1291–1319.

Ginibre J and Velo G (1982) The Cauchy problem for the o($N$), cp($N-1$), and $G_C(N,p)$ models. *Annals of Physics* 142: 393–415.

Ginibre J and Velo G (1995) Generalized Strichartz inequalities for the wave equation. *Journal of Functional Analysis* 133: 50–68.

Hörmander L (1997) *Lectures on Nonlinear Hyperbolic Equations*. Berlin: Springer.

John F (1979) Blow-up of solutions of nonlinear wave equations in three space dimensions. *Manuscripta Mathematica* 28(1–3): 235–268.

Keel M and Tao T (1998) Endpoint Strichartz estimates. *American Journal of Mathematics* 120: 955–980.

Klainerman S (1986) The null condition and global existence to nonlinear wave equations. *Lecture in Applied Mathematics* 23: 293–326.

Klainerman S and Selberg S (2002) Bilinear estimates and applications to nonlinear wave equations. *Communications in Contemporary Mathematics* 4: 223–295.

Lindblad H and Sogge C (1995) Existence and scattering with minimal regularity for semilinear wave equations. *Journal of Functional Analysis* 130: 357–426.

Schiff LI (1951) Nonlinear meson theory of nuclear forces I. *Physical Review* 84: 1–9.

Segal IE (1963) The global Cauchy problem for a relativistic scalar field with power interaction. *Bulletin de la Société Mathématique de France* 91: 129–135.

Shatah J (1988) Weak solutions and the development of singularities in the SU(2) $\sigma$-model. *Communications on Pure and Applied Mathematics* 41: 459–469.

Shatah J and Struwe M (1993) Regularity results for nonlinear wave equations. *Annals of Mathematics* 138: 503–518.

Shatah J and Struwe M (1998) *Geometric Wave Equations*. Courant Lecture Notes in Mathematics, vol. 2. New York: Courant Institute, New York University.

Strauss W (1989) *Nonlinear Wave Equations*. CBMS Lecture Notes, vol. 75. Providence, RI: American Mathematical Society.


# Separation of Variables for Differential Equations


**S Rauch-Wojciechowski**, Linköping University, Linköping, Sweden

**K Marciniak**, Linköping University, Norrköping, Sweden




## Introduction

The method of separation of variables (SoV) is a way of finding particular and general solutions of certain types of partial differential equations (PDEs). Its main idea is to consider the additive ansatz $u(x)=\sum_{\mathrm{i}} w_i(x^i,\alpha)$ or the multiplicative ansatz $u(x)=\prod_{\mathrm{i}} u_i(x^i,\alpha)$ for a solution of a PDE that allows for reducing this PDE to a set of (uncoupled) ordinary differential equations (ODEs) for the unknown functions $w_i(x^i,\alpha)$ or $u_i(x^i,\alpha)$ of one variable $x^i$, where $x=(x^1,\ldots,x^n)$. Locally, the additive ansatz is, through the change of variables $u(x)=\exp(\sum_{\mathrm{i}} w_i(x^i,\alpha))$, equivalent to the multiplicative ansatz.

Many well-known equations of mathematical physics such as the heat equation, the wave equation, the Schrödinger equation, and the Hamilton–Jacobi equation are solved by separating variables in suitably chosen systems of coordinates.

## Fourier Method

The SoV method can be attributed to Fourier (1945), who solved the heat equation

$$\partial_t u = \partial_{xx} u \qquad [1]$$

for distribution of temperature $u(x,t)$ in a one-dimensional metal rod (of length $L$) by looking first for special solutions of the product type $u(x,t)=X(x)T(t)$. This ansatz, substituted to [1], reduces it to two ODEs: $\partial_t T=-k^2 T$ and $\partial_{xx}X=-k^2X$ that can be solved by quadratures:

$$T_k(t)=A\mathrm{e}^{-k^2 t}, \qquad X_k(x)=B\cos(kx)+C\sin(kx)$$

Due to linearity of [1], any formal linear combination $u(x,t)=\sum_k c_k X_k(x)T_k(t)$ is again a solution of the heat equation and can be used for solving an initial boundary-value problem (IBVP). For instance,

in the case of the IBVP on the interval $0 \le x \le L$ and with zero boundary conditions

$$\begin{aligned} &\partial_t u = \partial_{xx} u, && 0 < t, 0 < x < L \\ &u(0,t) = u(L,t) = 0, && 0 < t \\ &u(x,0) = f(x), && 0 < x < L \end{aligned}$$

only a countable set of values for the separation constant $k$ is admissible: $k_n = (n\pi/L)$, $n = 1, 2, \ldots$. Then the general solution has the form of the Fourier series

$$u(x,t) = \sum_{n=1}^{\infty} c_n \exp\left(-k_n^2 t\right) \sin(k_n x)$$

where the coefficients $c_n$ are given by the integrals

$$c_n = \frac{2}{L} \int_0^L f(x) \sin(k_n x)\, \mathrm{d}x$$

The sequence of functions $\sin(k_n x)$ is complete on the interval $[0, L]$. That means that any regular (continuous and differentiable) initial data function $f(x)$ such that $f(0) = f(L) = 0$ can be uniquely expressed as an infinite convergent sum of the orthogonal set of functions $\sin(k_n x)$. The study of mathematical properties of the Fourier expansion gave rise to the classical theory of Fourier series and Fourier integrals.

## Separability of PDEs in General Setting

A general setting for an additive separability of a single, usually nonlinear, PDE has been developed by Levi-Civita (1904) and by Kalnins and Miller (1980) (see also Miller (1983)). Let

$$\begin{aligned} &H(x^1, \ldots, x^n; u, u_i, u_{ij}, u_{ijk}, \ldots) = E \\ &\quad 1 \le i, j, k \le n \end{aligned} \qquad [2]$$

be a finite-order PDE for an unknown function $u(x)$, where $u_i(x) = \partial_{x^i} u$, $u_{ij} = \partial_{x^j} \partial_{x^i} u$, etc., and $E$ is a constant. A separable solution $u(x) = \sum_i W_i(x^i)$ satisfies the simpler equation

$$E = H(x; u, u_i, u_{ii}, \ldots) \equiv H[x, u] \qquad [3]$$

where all mixed derivatives $u_{ij}$, etc., disappear. If a separable solution is admissible by eqn [2], then the function $H(x; u, u_i, u_{ii}, \ldots)$ has to satisfy a set of integrability conditions following from the total derivatives of [3]. Let

$$\begin{aligned} D_i &= \partial_{x^i} + u_{i,1} \partial_u + u_{i,2} \partial_{u_{i,1}} + \cdots + u_{i,m_i+1} \partial_{u_{i,m_i}} \\ &\equiv \widetilde{D}_i + u_{i,m_i+1} \partial_{u_{i,m_i}} \end{aligned}$$

(where $u_{i,1} = u_i, u_{i,j+1} = \partial_{x^i} u_{i,j}$, etc., and $m_i$ is the largest number $l$ such that $\partial_{u_{i,l}} H \ne 0$) denote the operator of total derivative with respect to (w.r.t.) $x^i$; then, $D_i H[x, u] = 0$ or

$$u_{i,m_i+1} = -\frac{\widetilde{D}_i H}{H_{u_i,m_i}}$$

where $H_{u_i,m_i} = \partial_{u_i,m_i} H$. The integrability conditions $D_j u_{i,m_i+1} = 0$, $j \ne i$, give rise to a large set of differential conditions to be satisfied by $H[x, u]$:

$$\begin{aligned} &H_{u_i,m_i} H_{u_j,m_j} \left(\widetilde{D}_i \widetilde{D}_j H\right) + H_{u_i,m_i u_j,m_j} \left(\widetilde{D}_i H\right)\left(\widetilde{D}_j H\right) \\ &\quad = H_{u_j,m_j} \left(\widetilde{D}_i H\right)\left(\widetilde{D}_j H_{u_i,m_i}\right) \\ &\qquad + H_{u_i,m_i} \left(\widetilde{D}_j H\right)\left(\widetilde{D}_i H_{u_j,m_j}\right) \end{aligned} \qquad [4]$$

In general, the conditions [4] are restrictions for both $H$ and the form of a particular separable solution $u(x)$. If [4] is satisfied identically w.r.t. all $u$, $u_{k,l}$, we say that the corresponding coordinate system $x^i$ is a regular separable coordinate system; then the PDE [3] admits a $(\sum_i m_i + 1)$-parameter family of separable solutions. Most cases considered in literature are regular; since then the separable solution is usually sufficiently general for solving various IBVPs.

A given PDE, however, usually does not satisfy [4]; since these equations are not of tensorial type, the natural question arises if there exists a suitable change of coordinates $y(x)$ such that the transformed PDE satisfies [4]. Such separation coordinates may or may not exist; it is usually very difficult to decide.

Here and in what follows, we speak about separability of a single (scalar) PDE. The theory of separability of systems of PDEs is still not developed fully, although it is of relevance in the theory of Maxwell equations and of the Dirac equation.

We present here the most classical part of SoV theory: orthogonal separability of the Hamilton–Jacobi equation for geodesic motions on Riemannian manifolds.

## Configurational Separation of Hamilton–Jacobi Equation on Riemannian Manifolds

Around 1842, C G J Jacobi invented the method of generating function for solving the canonical Hamilton equations

$$\begin{aligned} &\dot{x} = \frac{\partial H(x,y)}{\partial y}, && \dot{y} = -\frac{\partial H(x,y)}{\partial x} \\ &x = (x^1, \ldots, x^n) && y = (y^1, \ldots, y^n) \end{aligned} \qquad [5]$$

where $H(x, y)$ is a Hamiltonian and dot denotes the time derivative (Landau and Lifshitz 1976). In this

method, one looks for a generating function $W(x, \alpha)$ of a canonical transformation

$$y = \frac{\partial W(x, \alpha)}{\partial x}, \qquad \beta = \frac{\partial W(x, \alpha)}{\partial \alpha}$$

that transforms Hamiltonian equations [5] into simple equations for the new variables $\beta \in R^n$, $\alpha \in R^n$. Since the transformation is canonical, the transformed equations are again Hamiltonian with the new Hamiltonian $\widetilde{H}(\beta, \alpha) = H(x(\beta, \alpha), y(\beta, \alpha))$. If we choose this transformation so that $\widetilde{H}(\beta, \alpha) = \alpha_1$, then the transformed Hamilton equations become

$$\dot{\beta} = \frac{\partial \widetilde{H}(\beta, \alpha)}{\partial \alpha} = (1, 0, \ldots, 0)$$
$$\dot{\alpha} = -\frac{\partial \widetilde{H}(\beta, \alpha)}{\partial \beta} = 0$$

so that $\beta(t) = (t + \beta_{10}, \beta_{20}, \ldots, \beta_{n0})$, $\alpha(t) = (\alpha_{10}, \ldots, \alpha_{n0}) = \text{const.}$ and the solution $x(t), y(t)$ of the Hamilton equations [5] is then given implicitly by the equations

$$\beta(t) = \frac{\partial W(x(t), \alpha)}{\partial \alpha}, \qquad y(t) = \frac{\partial W(x(t), \alpha)}{\partial x}$$

Since

$$y = \frac{\partial W(x, \alpha)}{\partial x}$$

the generating function $W(x, \alpha)$ has to satisfy (identically w.r.t. $(x, \alpha)$) the first-order nonlinear PDE

$$H\left(x, \frac{\partial W(x, \alpha)}{\partial x}\right) = \alpha_1 \qquad [6]$$

This equation is called the Hamilton–Jacobi equation for the generating function $W(x, \alpha)$. It is solved when its complete integral $W(x, \alpha)$, complete means that

$$\det\left(\frac{\partial^2 W(x, \alpha)}{\partial x^i \partial \alpha_j}\right) \neq 0$$

depending on $n$ independent constants $\alpha$ is known. In general, it is very difficult to find solutions of [6]. The most important method is the method of separation of variables when one looks for a solution in the form $W(x, \alpha) = \sum_{k=1}^{n} W_k(x^k, \alpha)$ which is a sum of $n$ functions $W_k(x^k, \alpha)$, each depending on a single variable $x^k$ and, possibly, all constants $a$. If the Hamilton–Jacobi equation [6] admits such a solution, then integrating this equation is reduced to integrating $n$ (uncoupled) first-order ODEs for functions $W_k(x^k, \alpha)$. The constants $\alpha_k$ acquire then the meaning of integration constants.

A separable solution $W(x, \alpha)$ of [6] exists whenever the Hamiltonian $H(x, y)$ satisfies (identically) the integrability conditions [4] which in this case acquire the (nonlinear) form

$$\begin{aligned} L_{ij}(H) &\equiv \partial_i H \partial_j H \partial^i \partial^j H + \partial^i H \partial^j H \partial_i \partial_j H \\ &\quad - \partial_i H \partial^j H \partial^i \partial_j H - \partial^i H \partial_j H \partial_i \partial^j H \\ &= 0 \quad \text{for all } i, j = 1, \ldots, n \end{aligned} \qquad [7]$$

($\partial_i = \partial/\partial x^i, \partial^i = \partial/\partial y_i$) found by Levi-Civita (1904).

In classical mechanics the most important Hamiltonians are natural ones:

$$H(x, y) = \frac{1}{2}\sum_{i,j} g^{ij}(x) y_i y_j + V(x) \equiv G + V \qquad [8]$$

They are defined on the cotangent bundle $T^*Q$ of a configurational Riemannian manifold $Q$ with the metric tensor $g$. The function $G$ is the geodesic Hamiltonian associated with the metric tensor $g$. For such natural Hamiltonians, the Levi-Civita condition $L_{ij}(G + V) = 0$ splits into the condition $L_{ij}(G) = 0$ and a condition for the potential $V(x)$. The condition $L_{ij}(G) = 0$, depending solely on the kinetic energy term, is thus a necessary condition for coordinates $x^i$ on $Q$ to be separation coordinates for [8].

In the fundamental case of orthogonal separation (i.e., when $g^{ij} = 0$ for $i \neq j$), the Levi-Civita conditions $L_{ij}(G + V) = 0$ read

$$\begin{aligned} &\partial_i \partial_j g^{kk} - \left(\partial_i \ln g^{jj}\right)\partial_j g^{kk} \\ &\quad - \left(\partial_j \ln g^{ii}\right)\partial_i g^{kk} = 0, \quad i \neq j \end{aligned} \qquad [9]$$

$$\begin{aligned} &\partial_i \partial_j V - \left(\partial_i \ln g^{jj}\right)\partial_j V \\ &\quad - \left(\partial_j \ln g^{ii}\right)\partial_i V = 0, \quad i \neq j \end{aligned} \qquad [10]$$

The main questions arising here are

1. What is the algebraic form of orthogonally separable Riemannian metrics?
2. What is the form of separable coordinates on Riemannian manifolds?

The first question is answered by the Stäckel theorem (Stäckel 1891) that provides an algebraic characterization of orthogonal separability of a natural Hamiltonian $H = G + V$.

**Theorem 1** *The Hamilton–Jacobi equation for the natural Hamiltonian*

$$H = G + V = \frac{1}{2}\sum_i g^{ii}(x) y_i^2 + V(x)$$

*is separable in the (orthogonal) coordinates x if and only if*

(i) *There exists a matrix* $\Phi = [\varphi_{ij}(x^i)]$, $\det(\Phi) \neq 0$ *(so that the row i depends only on* $x^i$*) such that* $[g^{11}, \ldots, g^{nn}]$ *is the first row of the inverse matrix* $\Psi = \Phi^{-1}$.

(ii) *The potential V has the form* $V(x) = \sum_i g^{ii} f_i(x^i)$, *where each* $f_i(x^i)$ *is a function of one variable* $x^i$ *only.*

Such matrix $\Phi$ is called a Stäckel matrix.

**Proof** If

$$[g^{11}, \ldots, g^{nn}] \begin{bmatrix} \varphi_{11}(x^1) & \cdots & \varphi_{1n}(x^1) \\ \vdots & \ddots & \vdots \\ \varphi_{n1}(x^n) & \cdots & \varphi_{nn}(x^n) \end{bmatrix} = [1, 0, \ldots, 0] \qquad [11]$$

then the Hamilton–Jacobi equation for $H$ can be written as

$$\begin{aligned} \frac{1}{2}\sum_i g^{ii}\left(\frac{\partial W}{\partial x^i}\right)^2 + \sum_i g^{ii} f_i(x^i) = \alpha_1 \\ = \alpha_1 \sum_i g^{ii}\varphi_{i1}(x^i) + \alpha_2 \sum_i g^{ii}\varphi_{i2}(x^i) \\ + \cdots + \alpha_n \sum_i g^{ii}\varphi_{in}(x^i) \end{aligned} \qquad [12]$$

This equation admits an additively separable solution $W = \sum_i W_i(x^i)$, where the functions $W_i$ satisfy $n$ ODEs (separation equations):

$$\begin{aligned} \frac{1}{2}\left(\frac{\partial W_i}{\partial x^i}\right)^2 + f_i(x^i) \\ = \alpha_1\varphi_{i1}(x^i) + \alpha_2\varphi_{i2}(x^i) + \cdots + \alpha_n\varphi_{in}(x^i) \\ i = 1, \ldots, n \end{aligned} \qquad [13]$$

By differentiating [13] w.r.t. $\alpha_j$, we get

$$\varphi_{ij}(x^i) = \frac{\partial W_i}{\partial x^i}\frac{\partial^2 W_i}{\partial x^i \partial \alpha_j}$$

and thus

$$\det[\varphi_{ij}(x^i)] = \frac{\partial W_1}{\partial x^1} \cdots \frac{\partial W_n}{\partial x^n} \det\left(\frac{\partial^2 W}{\partial x^i \partial \alpha_j}\right) \neq 0$$

so that $W = \sum_i W_i(x^i)$ is indeed a complete integral of the Hamilton–Jacobi equation [12]. Conversely, if $W = \sum_i W_i(x^i)$ is a complete integral of the Hamilton–Jacobi equation [12], then by differentiating it w.r.t. $\alpha_j$ we get for $j = 1$

$$\sum_i g^{ii}\frac{\partial W_i}{\partial x^i}\frac{\partial^2 W_i}{\partial x^i \partial \alpha_j} = 1$$

and

$$\sum_i g^{ii}\frac{\partial W_i}{\partial x^i}\frac{\partial^2 W_i}{\partial x^i \partial \alpha_j} = 0$$

(for $j = 2, \ldots, n$), that is, the condition [11] for the Stäckel matrix

$$\Phi = \left[\frac{\partial W_i}{\partial x^i}\frac{\partial^2 W_i}{\partial x^i \partial \alpha_j}\right]$$

Further, we see that

$$\begin{aligned} V &= \alpha_1 - \frac{1}{2}\sum_i g^{ii}(\partial_{x^i} W_i)^2 \\ &= \frac{1}{2}\sum_i g^{ii}\left[\alpha_1\varphi_{i1}(x^i) - \frac{1}{2}(\partial_{x^i} W_i)^2\right] \\ &= \sum_i g^{ii} f_i(x^i) \end{aligned} \qquad \square$$

**Remark 2** The Stäckel characterization of orthogonal separability is equivalent to Levi-Civita conditions [9] and [10]. It is in fact a solution of these conditions.

**Remark 3** With every Stäckel matrix, one can relate a family of $n$ quadratic in momenta Hamiltonians defined by $n$ rows of the inverse Stäckel matrix $\Psi = \Phi^{-1} = [\psi_{kr}]$:

$$H_k = \frac{1}{2}\sum_{r=1}^{n} \psi_{kr} y_r^2, \quad k = 1, \ldots, n \qquad [14]$$

(so that $H_1 = G$). These Hamiltonians are linearly and functionally independent; they Poisson-commute (so that they form a Liouville integrable system) and are all diagonal so that they have common eigenvectors.

These properties are the main ingredients of an intrinsic (coordinate-independent) characterization of separable geodesic Hamiltonians $G$ in terms of involutive Killing tensors that is due to works of Eisenhart (1934), Kalnins and Miller (1980), and Benenti (1997).

**Theorem 4** *A necessary and sufficient condition for the existence of an orthogonal additive separable coordinate system x for the Hamilton–Jacobi equation of the geodesic Hamiltonian* $H_1 = G$ *on an n-dimensional (pseudo)-Riemannian manifold is that there exist n quadratic forms* $H_r = \sum_{i,j}^{n} h_r^{ij}(x) y_i y_j$ *such that*

(i) *They all Poisson-commute:* $\{H_r, H_s\} = 0, 1 \leq r, s \leq n$.

(ii) *The set* $\{H_r\}_{r=1}^{n}$ *is linearly independent.*

(iii) *There is a basis* $\{\omega_{(j)}\}_{j=1}^{n}$ *of n simultaneous eigenforms for all* $H_r$.

*If conditions(i)–(iii) are satisfied then there exist functions $g_j(x)$ such that $\omega_{(j)} = g_j \mathrm{d}x^j$, $j = 1, \ldots, n$.*

This theorem has been further simplified by Benenti (1997), who has shown that for separability it is sufficient that $g^{ij}$ admits a single Killing 2-tensor with simple eigenvalues and normal eigenvectors. He has also explained the role of ignorable coordinates.

These results are key ingredients of an answer to the question (2). Eisenhart (1934), starting from the fact that every separable geodesic Hamiltonian $H = G$ admits $n$ quadratic (w.r.t. momenta $y_i$) integrals of motion, derived a set of nonlinear PDEs characterizing separable Riemannian metrics. He has solved these equations for spaces of constant curvature. This solution is the basis of the Kalnins and Miller's (1986) diagrammatic classification of all orthogonal separation coordinates on $\boldsymbol{R}^n$ and the sphere $\boldsymbol{S}^n$. Separable coordinates on the Minkowski space $\boldsymbol{M}^n$ have not been classified yet.

Since the work of Robertson (1927) and Eisenhart (1934), it is known that in $\boldsymbol{R}^n, \boldsymbol{S}^n$ and, in general, in the space with diagonal Ricci tensor, the (additive) separability of Hamilton–Jacobi equation for the natural Hamiltonian $H = G + V$ is equivalent to multiplicative separability of the stationary Schrödinger equation with the same potential $V$:

$$(\Delta + V(x))\Theta(x) = E\Theta(x) \qquad [15]$$

where

$$\Delta = \sum_{i,j=1}^{n} \frac{1}{\sqrt{\det(g)}} \partial_i \left( \sqrt{\det(g)} g^{ij} \partial_j \right)$$

is the Laplace–Beltrami operator. Usually, multiplicative separated solutions $\Theta(x) = \prod_{i=1}^{n} \Theta_i(x)$ is considered but the change of the dependent variable $u = \ln \Theta$ transforms it into an additive separable solution. If we restrict our considerations to orthogonal separation coordinates ($g^{ij} = 0$ for $i \neq j$), eqn [15] becomes

$$\sum_{i=1}^{n} \left( g^{ii} (u_{ii} + u_i^2) + \frac{1}{\sqrt{\det(g)}} \partial_i \times \left( \sqrt{\det(g)} g^{ii} \right) u_i \right) + V(x) = E$$

where $u_i = \partial_i u$, $u_{ii} = \partial_i \partial_i u$. The integrability conditions [4] for regular separation lead to the Levi-Civita condition [9] on the components $g^{ii}$ of the metric tensor, upon comparison of the coefficients at $u_i^2$. The coefficients at $u_{ii}$ yield the Robertson condition

$$\partial_i \partial_j \ln \left( \sqrt{\det(g)} g^{ii} \right) = 0, \quad i \neq j$$

and the constant terms in [4] give the Levi-Civita equation [10] meaning that $V(x) = \sum_{i=1}^{n} g^{ii} f_i(x^i)$. Eisenhart has shown that the Robertson condition is equivalent to the requirement that the Ricci tensor is diagonal: $R_{ij} = 0$, $i \neq j$ in variables $x$ so that the Robertson condition is satisfied automatically in the Euclidean space, in spaces of constant curvature and in Einstein spaces. Thus every orthogonal coordinate system permitting multiplicative separation of the Schrödinger equation corresponds to the Stäckel form.

### Jacobi Problem of Separability

In order to apply the separability theory to physical Hamiltonians $H = (1/2)p^2 + V(q)$, $p = (p_1, \ldots, p_n)$, $q = (q^1, \ldots, q^n)$, it is essential to solve the following problem: "given a potential $V(q)$, decide if there exists a point transformation $x(q)$ to some curvilinear coordinates $x$ such that the Hamilton–Jacobi equation associated with $H$ is separable in coordinates $x$, and if such transformation exists, determine it and solve the obtained Hamilton–Jacobi equation."

This problem has been raised by Jacobi (1884) in connection with the problem of finding geodesic motions on a 3-axial ellipsoid. For solving this problem Jacobi introduced his "remarkable change of coordinates" to the generalized elliptic coordinates $x(q)$ defined through zeros of the rational function

$$1 + \sum_{i=1}^{n} \frac{(q^i)^2}{(z - \lambda_i)} \equiv \frac{\prod_j (z - x^j)}{\prod_i (z - \lambda_i)} \qquad [16]$$

where the constants $\lambda_i > 0$ are all different. From the graph of the left-hand side of [16], it is easy to see that there are exactly $n$ simple, real zeros. For given values of elliptic coordinates $x^j$, the values of $(q^i)^2$ are uniquely determined as residues at $\lambda_i$ while Cartesian coordinates $q^i$ are determined uniquely only in each $n$-tant of $\boldsymbol{R}^n$.

The Jacobi elliptic coordinates play a pivotal role in orthogonal separability on $\boldsymbol{R}^n$ and $\boldsymbol{S}^n$ since they are the mother of all other separation coordinates that can be obtained through proper and improper degenerations of $\lambda_i$'s. By using these coordinates Jacobi solved not only the geodesic motions on the ellipsoid but also the motion on the ellipsoid under the action of harmonic potential $V(q) = (1/2)q^2$. He has also found separation coordinates for a system of three interacting particles on the line known today as the Calogero system. In general, however, Jacobi considered the problem of finding separation coordinates for a given potential $V(q)$ to be very difficult. In *Vörlesungen über Dynamik*, ch. 26, he writes: "The main difficulty in integrating a given

differential equation lies in introducing convenient variables, which there is no rule for finding. Therefore, we must travel the reverse path and after finding some notable substitution, look for problems to which it can be successfully applied". This statement had a profound influence on further development of SoV theory that concentrated on characterizing separable Hamiltonians (as expressed in terms of separation coordinates) and on describing and classifying separation coordinates.

The original problem of Jacobi of finding separation variables for a given natural Hamiltonian has been taken up by Rauch-Wojciechowski (1986), who found a characterization of separable potentials $V(q)$ in terms of Cartesian coordinates $q_i$. Its invariant geometric form has been given by Benenti. A complete criterion of separability that allows for an effective testing and calculation of separation coordinates (if they exist) for $V(q)$ has been solved by Waksjö and Rauch-Wojciechowski (2003). This criterion is directly applicable to the problem of finding SoV for the Schrödinger equation.

### Criterion of Separability for $n=2$

The criterion of separability for $n=2$ can be read from the Bertrand–Darboux theorem.

**Theorem 5** (Bertrand–Darboux). *For the Hamiltonian*:

$$H=\tfrac{1}{2}(p_1^2+p_2^2)+V(q_1,\,q_2)$$

*the following statements are equivalent*:

(i) *H has a functionally independent integral of motion $\{H,K\}=0$ of the form*

$$\begin{aligned}K=&\left(aq_2^2+bq_2+c\right)p_1^2+\left(aq_1^2+\widetilde{b}q_1+\widetilde{c}\right)p_2^2\\&+\left(-2aq_1q_2-bq_1-\widetilde{b}q_2+d\right)p_1p_2\\&+k(q_1,q_2)\end{aligned}$$

(ii) *The potential $V(q_1,q_2)$ satisfies the following linear second-order PDE with quadratic coefficients*

$$\begin{aligned}0=&\,2\left(aq_2^2-aq_1^2+bq_2-\widetilde{b}q_1+c-\widetilde{c}\right)\partial_1\partial_2V\\&+(-2aq_1q_2-bq_1-\widetilde{b}q_2+d)\left(\partial_2^2V-\partial_1^2V\right)\\&+(6aq_2+3b)\partial_1V-(6aq_1+3\widetilde{b})\partial_2V\end{aligned}\qquad[17]$$

*where $a,b,\widetilde{b},c,\widetilde{c},d$ are some constants, $\partial_1=\partial_{q_1},\partial_2=\partial_{q_2}$.*

(iii) *The Hamilton–Jacobi equation for H is separable in one of the four orthogonal coordinate systems in the plane: elliptic, parabolic, polar, or Cartesian.*

**Remark 6** If the potential $V(q_1,q_2)$ is separable, then it admits an integral of motion $K$ that is quadratic w.r.t. momenta and $V$ satisfies (identically w.r.t. $q_1,q_2$) eqn [17] for certain values of the undetermined constants $a,b,\widetilde{b},c,\widetilde{c},d$. Since coefficients at linearly independent expressions of $q_1,q_2$ have to be equal to zero, the parameters $a,b,\widetilde{b},c,\widetilde{c},d$ have to satisfy a set of linear, algebraic, homogeneous equations. If there is a nonzero solution for $a,b,\widetilde{b},c,\widetilde{c},d$, then there exists an integral of motion $K$ and separation coordinates can be determined as characteristic variables for equation [17].

**Example 7** Separable cases of the Henon–Heiles potential

$$V=\tfrac{1}{2}\left(\omega_1q_1^2+\omega_2q_2^2\right)+\alpha q_1^2q_2-\tfrac{1}{3}\beta q_2^3$$

By substituting this form of $V$ into [17], we get two sets of admissible solutions for parameters $\alpha$, $\beta$, $\omega_1$, $\omega_2$: (i) $\beta=-\alpha$, $\omega_1=\omega_2$ with $V$ separable in rotated (by $\pi/4$) Cartesian coordinates; (ii) $\beta=-6\alpha$, $\omega_1$, $\omega_2$-arbitrary with $V$ separable in the shifted parabolic coordinates. In case (ii) eqn [17] becomes

$$\begin{aligned}2&\left(q_2-\frac{1}{4\alpha}(4\omega_1-\omega_2)\right)\partial_1\partial_2V\\&+q_1\left(\partial_2^2V-\partial_1^2V\right)+3\partial_1V=0\end{aligned}$$

and in its characteristic coordinates defined as $q_1=\sqrt{\xi\eta}$, $q_2=(1/2)(\xi-\eta)+(1/4\alpha)(4\omega_1-\omega_2)$ it takes the form $(\xi-\eta)\partial_\xi\partial_\eta V+\partial_\xi V+\partial_\eta V=0$ solved by $V(\xi,\eta)=(\xi+\eta)^2[f(\xi)+g(\eta)]$ which is separable in the parabolic coordinates.

### Effective Criterion of Separability for Arbitrary Dimension

For $n>2$, a similar theorem characterizing separability in generalized elliptic coordinates has been formulated by Rauch-Wojciechowski (1986).

**Theorem 8** (Elliptic Bertrand–Darboux). *For a natural Hamiltonian $H=(1/2)p^2+V(q)$, the following statements are equivalent*:

(i) *H has n global, functionally independent and involutive integrals of motion $\{H,K_i\}=0$, $\{K_i,K_j\}=0$, $i,j=1,\ldots,n$, having the form*

$$\begin{aligned}K_i&=\sum_{r=1,r\neq i}^{n}(\lambda_i-\lambda_r)^{-1}l_{ir}^2+p_i^2+k_i(q)\\l_{ir}&=q_ip_r-q_rp_i\end{aligned}\qquad[18]$$

(ii) *The potential V satisfies the following system of linear second-order PDEs*

$$(\lambda_i - \lambda_j)\partial_i\partial_j V - \Im_{ij}(2 + \Re)V = 0 \quad i,j = 1,\ldots,n,\ i \neq j \qquad [19]$$

$$\lambda_i\partial_i\Im_{jk}V + \lambda_j\partial_j\Im_{ki}V + \lambda_k\partial_k\Im_{ij}V = 0 \quad \text{all } i,j,k \text{ different} \qquad [20]$$

*where* $\Im_{ij} = q_i\partial_j - q_j\partial_i$, $\Re = \sum_{r=1}^{n} q_r\partial_r$.

(iii) *The Hamilton–Jacobi equation for H is separable in the generalized elliptic coordinates [16] with parameters* $\lambda_i$.

**Remark 9** Equations [19]–[20] follow from the compatibility conditions that mixed derivatives of $k_i(q)$ calculated from the conditions $\{H, K_r\} = 0$, are equal. This leads to an overdetermined system [19]–[20] of PDEs for $V(q)$. Equations [19]–[20] are not linearly independent but we keep both sets [19]–[20] in the formulation of this theorem because eqns [19] give rise to the basic Bertrand–Darboux equations [21] used in the criterion of separability while eqns [20] give rise to cyclic Bertrand–Darboux equations [22] used for testing the level of spherical symmetry in the potential.

For testing elliptic separability of any given potential $V(q)$, it is necessary to introduce into eqns [19] and [20] the freedom of choice of the Euclidean reference frame (as described by the Euclidean transformation $\widetilde{q} = A^t(q - b)$, $A \in \mathrm{SO}(n)$, $b \in \boldsymbol{R}^n$). By substituting it into [19]–[20], omitting tildes and summing over one of the indices, we obtain new equations

$$\begin{aligned} 0 = \sum_{k=1}^{n}\big(&(\alpha q_i q_k + \beta_i q_k + \beta_k q_i + \gamma_{ik})\partial_k\partial_j V \\ &- (\alpha q_j q_k + \beta_j q_k + \beta_k q_j + \gamma_{jk})\partial_k\partial_i V \\ &+ 3\big((\alpha q_i + \beta_i)\partial_j V - (\alpha q_j + \beta_j)\partial_i V\big)\big) \\ &i,j = 1,\ldots,n,\ i \neq j \end{aligned} \qquad [21]$$

$$\begin{aligned} 0 = \sum_{l=1}^{n}\big(&(\gamma_{il}q_j - \gamma_{jl}q_i)\partial_k\partial_l V + (\gamma_{jl}q_k - \gamma_{kl}q_j)\partial_i\partial_l V \\ &+ (\gamma_{kl}q_i - \gamma_{il}q_k)\partial_j\partial_l V\big) \end{aligned} \qquad [22]$$

with the new coefficients $\alpha$, $\beta_i$, $\gamma_{ij}$ that are unconstrained despite that the orthogonal matrix $A$ satisfies the quadratic algebraic constraint $AA^t = I$.

Theorem 8 provides the following test of elliptic separability for a potential $V(q)$ given in Cartesian coordinates.

1. Insert $V(q)$ into the Bertrand–Darboux equations [21]. This gives a system of linear, homogeneous, algebraic equations for the unknown parameters $\alpha$, $\beta_i$, $\gamma_{ij}$. If $\alpha = 0$, then $V(q)$ is not separable in elliptic coordinates.
2. If $\alpha \neq 0$, set $b = -\alpha^{-1}\beta$, $S = bb^t - \alpha^{-1}\gamma$ and diagonalize $S$: $S = A\,\mathrm{diag}(\lambda_1,\ldots,\lambda_n)A^t$. If some eigenvalues $\lambda_i$ coincide, then $V(q)$ is not separable in elliptic coordinates. Otherwise $V(q)$ is separable in the elliptic coordinates $x = (x^1,\ldots,x^n)$ given by

$$1 + \sum_{i=1}^{n}\frac{(\widetilde{q}^i)^2}{(z - \lambda_i)} \equiv \frac{\prod_j (z - x^j)}{\prod_i (z - \lambda_i)}$$

(compare with [16]), where $q = A\widetilde{q} + b$, with $b$ and $A$ found as above.

If $\alpha = 0$, $\beta \neq 0$, then there exists a similar algorithm for separability in generalized parabolic coordinates and for $\alpha = 0$, $\beta = 0$, $\gamma \neq tI$, we have separability in Cartesian coordinates if all $\lambda_i$ are different. For giving an idea of what happens when degenerations occur, consider the case $\alpha = 0$, $\beta = 0$. Then the Bertrand–Darboux equations [21] are Euclidean equivalent to the canonical form $(\lambda_i - \lambda_j)\ \partial_i\partial_j V = 0$ and if all $\lambda_i$ are different, then equations $\partial_i\partial_j V = 0$ imply that $V(q)$ is a sum of functions of one variable only: $V(q) = \sum_{i=1}^{n} V_i(q^i)$.

The main problem is to handle all possible degenerations when certain $\lambda$'s coincide. Let $\lambda_1 = \cdots = \lambda_j < \lambda_{j+1} < \cdots < \lambda_n$, where $1 < j < n$. Then $V(q) = V_j(q^1,\ldots,q^j) + V_{j+1}(q^{j+1}) + \cdots + V_n(q^n)$ which means that variables $q^{j+1},\ldots,q^n$ separate off while the potential $V_j(q^1,\ldots,q^j)$ has to be tested again on $\boldsymbol{R}^j$ with the use of eqns [21]. Degenerations for $\alpha \neq 0$ or $\beta \neq 0$ are more complicated and the cyclic Bertrand–Darboux equations [22] have to be used. They unfold the level of spherical degeneracy of spheres and embedded subspheres. A complete analysis of all possible degenerations is technical. It requires considering of all possible degenerations of the sequences $\lambda_1 < \cdots < \lambda_n$ and of the related equations [21]–[22] for the potential $V(q)$. It has been proved by Waksjö and Rauch-Wojciechowski (2003) that there is a one-to-one correspondence between all possible sets of PDEs [21]–[22] characterizing separable potentials and all possible types of Riemannian metrics (in the Kalnins and Miller (1986) classification of all separable coordinates on $\boldsymbol{R}^n$ and $\boldsymbol{S}^n$) so that no completely separable case is missed. The most important is that after maximally $n$ steps separation coordinates are always determined (if they exist) by a sequential use of the Bertrand–Darboux and cyclic Bertrand–Darboux equations [21]–[22].

## Separation of Eigenvalues Problems

Eigenvalues problems (in a given domain $D$) of the form

$$\Delta w(q) + \lambda\rho(q)w(q) = 0, \quad \rho > 0 \qquad [23]$$

(where $\Delta$ is the Laplace operator) arise when separating the wave equation $\rho(q)u_{tt} = \Delta u$ and the diffusion equation $\rho(q)u_t = \Delta u$ (Courant and Hilbert 1989). The multiplicative ansatz $u(q,t) = w(q)g(t)$ yields eqn [23] together with $\ddot{g} = \lambda g$ or $\dot{g} = \lambda g$. The problem [23] is also used for solving the inhomogeneous equation $\Delta u = f$ with the zero boundary condition $u|_{\partial D} = 0$. In general, the properties of the eigenvalues $\lambda_i$ and of the corresponding eigenfunctions $w_i$ of the problem [23] depend on the regularity requirements for $w_i$ and on the boundary conditions at $\partial D$.

For the zero boundary conditions $w(q)|_{\partial D} = 0$, one seeks a nontrivial ($w \neq 0$) solution having in the region $D$ continuous first- and second-order derivatives. General theorems (Courant and Hilbert 1989) state that for such problems there exists a growing sequence $\{\lambda_i\}_{i=1}^n$ of positive eigenvalues $\lambda_i$ such that $\lambda_i \to \infty$ as $i$ increases, and that there is a related sequence of normalized eigenfunctions $\sqrt{\rho}w_1$, $\sqrt{\rho}w_2, \ldots$ that form a complete weighted-orthogonal (in the sense that $\int_D \rho w_i w_j = \delta_{ij}$, $i, j = 1, 2, \ldots$) system of functions so that every regular initial function $\nu(q)$ with $\nu(q)|_{\partial D} = 0$ may be expanded in terms of the eigenfunctions $w_m$ in an absolutely and uniformly convergent series $\nu(q) = \sum_{m=1}^{\infty} c_m w_m(q)$ with $c_m = \int_D \rho\nu w_m$. This makes it possible to express a solution of the IBVP for the wave or for the diffusion equation with zero boundary conditions:

$$\begin{aligned} \rho(q)u_{tt} = \Delta u \text{ respectively } \rho(q)u_t = \Delta u \\ u(q,t)|_{\partial D} = 0, \quad u(q, t = 0) = \nu(q) \end{aligned} \qquad [24]$$

as a convergent infinite series $u(q,t) = \sum_{m=1}^{\infty} c_m w_m(q) g_m(t)$, where $g_m(t)$ satisfy $\ddot{g} = \lambda g$ respectively $\dot{g} = \lambda g$. Further determination of properties of the eigenfunctions $w_n$ is possible only in special domains $D$ when the problem [23] can be reduced to one-dimensional eigenvalue problems by separating variables in some suitable coordinates.

**Example 10** Consider the spherical domain $r^2 = x^2 + y^2 + z^2 \leq 1$. Equation [23] with $\rho = 1$ attains in the spherical coordinates $(r, \varphi, \theta)$ the form

$$\Delta w + \lambda w \equiv \frac{1}{r^2 \sin\theta}\left(\partial_r(r^2 \sin\theta\, \partial_r w) + \partial_\varphi\left(\frac{1}{\sin\theta}\partial_\varphi w\right) + \partial_\theta(\sin\theta\, \partial_\theta w)\right) + \lambda w = 0$$

The ansatz $w = f(r)Y(\theta, \varphi)$ gives the separated equation

$$\frac{(r^2 f')' + \lambda r^2 f}{f} = -\frac{1}{Y\sin\theta}\left(\partial_\varphi\left(\frac{1}{\sin\theta}\partial_\varphi Y\right) + \partial_\theta(\sin\theta\, \partial_\theta Y)\right)$$

so that its both sides must be equal to a constant $\kappa$. Continuity of $Y$ implies that it has to be periodic in $\varphi$ (with period $2\pi$) and regular at $\theta = 0$, $\theta = \pi$. It can only be satisfied for $\kappa = n(n+1)$. The left-hand side of the above equation yields then $(r^2 f')' - n(n+1)f + \lambda r^2 f = 0$. Solutions that are regular at $r = 0$ are the Bessel functions $(1/\sqrt{r})J_{n+(1/2)}(\sqrt{\lambda}r)$. The equation for spherical harmonics

$$\frac{1}{Y\sin\theta}\left(\partial_\varphi\left(\frac{1}{\sin\theta}\partial_\varphi Y\right) + \partial_\theta(\sin\theta\, \partial_\theta Y)\right) + n(n+1)Y = 0$$

can be further multiplicatively separated by assuming $Y = \Theta(\theta)\Phi(\varphi)$. The function $P(z = \cos\theta) = \Theta(\theta)$ satisfies then the Legendre equation

$$\left((1 - z^2)P'(z)\right)' + \left(n(n+1) - \frac{\sigma}{1 - z^2}\right)P(z) = 0$$

$P(z)$ is regular at $z = \pm 1$ only when $\sigma = k^2$, $k = 0, 1, 2, \ldots$. The function $\Phi(\varphi)$ satisfies then $\Phi'' = -k^2\Phi$ with solutions $\Phi_k(\varphi) = a_k \cos(k\varphi) + b_k \sin(k\varphi)$. The full solution of the eigenvalue problem $\Delta w + \lambda w = 0$ has the form of an infinite series

$$w_m(r, \varphi, \theta) = \sum_{n=0}^{\infty} \frac{1}{\sqrt{r}} J_{n+(1/2)}\left(\sqrt{\lambda_{m,n}}r\right)\Big[a_{n,0}P(\cos\theta) + \sum_{k=1}^{n}\left(a_{n,k}\cos(k\theta) + b_{n,k}\sin(k\theta)\right) \times P_{n,k}(\cos\theta)\Big]$$

where the constants $\lambda_{m,n}$, $m = 1, 2, \ldots$, are determined by the transcendental equation $J_{n+(1/2)}(\sqrt{\lambda}) = 0$ that follows from the boundary condition $u(q,t)|_{\partial D} = 0$.

Almost all BVPs that can be reduced to one-dimensional eigenvalue problems may be considered as a special or limiting case of the Lame problem where the boundary $\partial D$ is given by pieces of confocal quadrics corresponding to some separation coordinates. If $D = \{q(x) \in \boldsymbol{R}^3 : x_i^0 \leq x_i \leq x_i^1,\ i = 1, 2, 3\}$ is a domain defined by parametrizing $q$ with the elliptic coordinates $x_i$ given by [16], then the eigenvalue problem $\Delta w + \lambda w = 0$ splits into three one-dimensional equations of the form

$$\varphi(s)Y''(s) + \tfrac{1}{2}\varphi'(s)Y'(s) + (\lambda s + \mu)Y(s) = 0$$

where $\varphi(s) = 4(s - e_1)(s - e_2)(s - e_3)$ and $e_i$ are parameters of the elliptic coordinates. This is the Lame equation; its solutions define new transcendental functions that depend on the choice of the constants $\lambda, \mu$.

The approach presented here extends to diverse modifications such as vibrations with forcing term $\Delta w(q) + \lambda w(q) = f(q)$, vibrations of a nonhomogeneousmedium $\Delta w(q) + \lambda \rho(q) w(q) = 0$, the stationary Schrödinger equation $\Delta w(q) + V(q) w(q) = \lambda w(q)$ whenever the functions $\rho(q)$, $f(q)$, $V(q)$ are compatible with the separation coordinates.

Separation equations for the second-order BVP are the source of one-dimensional eigenvalue problems of the Sturm–Liouville type

$$(p(s)u')' - q(s)u + \lambda \rho(s) u = 0$$

with singularities that may occur at the endpoints of the fundamental domain. Majority of orthogonal polynomials and special functions appearing in mathematical physics are solutions of Sturm–Liouville problems.

In the complex domain the study of singularities of Laurent series solutions of the same equations led to development of theory of linear ODEs with singular points of the Fuchs class and the Böcher class.

## Constructive Approach to Separability of Liouville Integrable Systems

In the constructive approach to separability, one considers simultaneously all Hamilton–Jacobi equations following from a set of $n$, functionally independent, commuting integrals $H_1(x, y), \ldots, H_n(x, y)$, $\{H_i, H_j\} = 0$, that define a Liouville integrable system (Sklyanin 1995).

One starts with the separation equations, a set of $n$ decoupled ODEs for the functions $W_i(x_i, \alpha)$ depending on one variable $x_i$ and parametric $\alpha \in \boldsymbol{R}^n$:

$$f_i\left(x_i, y_i = \frac{\partial W_i(x_i, \alpha)}{\partial x_i}; \alpha\right) = 0 \qquad [25]$$

Assume that the dependence on $\alpha_i$ is essential (i.e., that $\det(\partial f_i / \partial \alpha_j) \neq 0$) so that we can resolve eqns [25] w.r.t. $\alpha_i$ so that $\alpha_i = H_i(x, y)$ for some functions $H_i$. If the functions $W_i(x_i, \alpha)$ solve [25] identically w.r.t. $x$ and $\alpha$, then the function $W(x, \alpha) = \sum_{i=1}^{n} W_i(x_i, \alpha)$ is simultaneously an additively separable solution of eqns [25] and of the equations

$$\alpha_i \overset{x,\alpha}{\equiv} H_i\left(x, y = \frac{\partial W(x, \alpha)}{\partial x}\right), \quad i = 1, \ldots, n \qquad [26]$$

since solving [25] w.r.t. $\alpha$ is a purely algebraic operation. We can treat eqns [26] as a set of simultaneously separable (in the canonical variables $(x, y)$) Hamilton–Jacobi equations related to the Hamiltonians $H_i$. Assume now that

$$\det\left(\frac{\partial^2 W}{\partial x_i \, \partial \alpha_j}\right) = \det\left(\frac{\partial^2 W_i}{\partial x_i \, \partial \alpha_j}\right) \neq 0$$

i.e. that $W$ is a complete integral for [26]. Then the Hamiltonians $H_i(x, y) = \alpha_i$ Poisson-commute since $\alpha_i$ can be treated as new canonical variables obtained by the canonical transformation $(x, y) \rightarrow (\beta, \alpha)$ given by

$$y = \frac{\partial W(x, \alpha)}{\partial x}, \qquad \beta = \frac{\partial W(x, \alpha)}{\partial \alpha}$$

Thus, any solvable w.r.t. $\alpha$ set of separation relations [25] defines a Liouville integrable system.

If we perform a canonical transformation from $(x, y)$ to new variables $(q, p)$, then the new set of commuting Hamiltonians $\widetilde{H}_i(q, p) = H_i(x(q, p), y(q, p))$ is also called separable.

The main problem for any given set of commuting Hamiltonians $\widetilde{H}_i(q, p)$ is to decide if there exists a canonical transformation $(q, p) \rightarrow (x, y)$ to the separation variables $(x, y)$ so that the related Hamilton–Jacobi equations [26] are simultaneously separable. An answer to this problem is known for integrable Hamiltonians solvable through the spectral curve method (Sklyanin 1995) and for the whole class of natural Hamiltonians discussed earlier.

This approach brings new, wider perspective to the classical separability mechanism stated in the Stäckel theorem. It contains majority of all known separable Hamiltonian systems. For example, if we specify the separation relations [25] to be affine in $\alpha_i$,

$$\sum_{k=1}^{n} f_{ik}(x_i, y_i) \alpha_k = g_i(x_i, y_i), \quad i = 1, \ldots, n \qquad [27]$$

then [27] are called generalized Stäckel separability conditions. To recover the explicit form of Hamiltonians $H_k = \alpha_k$, it is enough to solve relations [27] w.r.t. $\alpha_k$. It has been proved that the Stäckel Hamiltonians in [27] constitute a quasi-bi-Hamiltonian chain. If we specify further relations [27] by assuming that functions $f_{ik}$ do not depend on $y_i$ and functions $g_i$ are quadratic in $y_i$, then we obtain the classical Stäckel separability conditions (see Theorem 1)

$$\sum_{k=1}^{n} f_{ik}(x_i) \alpha_k = \frac{1}{2} g_i(x_i) y_i^2 + h_i(x_i) \qquad [28]$$

that can be solved for $\alpha_k$ yielding

$$\alpha_k(x,y) = \frac{1}{2}\sum_{i=1}^{n}(\Phi^{-1})_{ik}\left(y_i^2 + \frac{h_i(x_i)}{g_i(x_i)}\right)$$

that is, the Stäckel Hamiltonians [14] with the Stäckel matrix $\Phi = [\varphi_{ik}]$, where $\varphi_{ik} = f_{ik}(x_i)/g_i(x_i)$. By specifying [28] further, we obtain separation relations

$$x_i^{n-1}\alpha_1 + x_i^{n-2}\alpha_2 + \cdots + \alpha_n = \tfrac{1}{2}g(x_i)y_i^2 + h(x_i)$$

which give the so-called Benenti systems associated with conformal Killing tensors and cofactor pair systems.

Relations [27], with $g_i(x_i, y_i)$ depending exponentially on momenta $y$, contain several well-known systems such as periodic Toda lattice, the KdV dressing chain, and the Ruijsenaar–Schneider system. Relations with $g_i$ cubic in momenta $y$ yield stationary flows of Boussinesq hierarchy and integrable systems on the loop algebra sl(3).

*See also:* Boundary-Value Problems for Integrable Equations; Calogero–Moser–Sutherland Systems of Nonrelativistic and Relativistic Type; Elliptic Differential Equations: Linear Theory; Evolution Equations: Linear and Nonlinear; Integrable Systems: Overview; Multi-Hamiltonian Systems; Ordinary Special Functions; Recursion Operators in Classical Mechanics; Toda Lattices.

## Further Reading


Benenti S (1997) Intrinsic characterization of the variable separation in the Hamilton–Jacobi equation. *Journal of Mathematical Physics* 38(12): 6578–6602.

Courant R and Hilbert D (1989) *Methods of Mathematical Physics. vol. II. Partial Differential Equations*, *Wiley Classics Library*. A Wiley-Interscience Publication. New York: Wiley.

Eisenhart LP (1934) Separable systems of Stäckel. *Annals of Mathematics* 35(2): 284–305.

Fourier J (1945) *The Analytical Theory of Heat*. New York: G. E. Stechert and Co.

Jacobi CGJ (1884) *Vörlesungen über Dynamik. Herausgegeben A. Clebsch.*, ch. 26. Berlin: Verlag von G. Reimer.

Kalnins EG and Miller W Jr. (1980) Killing tensors and variable separation for Hamilton–Jacobi and Helmholtz equations. *SIAM Journal on Mathematical Analysis* 11(6): 1011–1026.

Kalnins EG and Miller W Jr. (1986) Separation of variables on $n$-dimensional Riemannian manifolds. I. The $n$-sphere $S^n$ and Euclidean $n$-space $R^n$. *Journal of Mathematical Physics* 27(7): 1721–1736.

Landau LD and Lifshitz EM (1976) *Course of Theoretical Physics. vol. 1. Mechanics*, Third Edition. Translated from the Russian by JB Sykes and JS Bell. Oxford–New York–Toronto, Ont: Pergamon Press.

Levi-Civita T (1904) Sulla integrazione della equazione di Hamilton-Jacobi per separazione di variabili. *Mathematische Annalen* 59: 383–397.

Miller W Jr. (1983) The technique of variable separation for partial differential equations. In: Wolf BK (ed.) *Nonlinear Phenomena* (Oaxtepec, 1982), Lecture Notes in Physics, vol. 189, pp. 184–208. Berlin: Springer.

Sklyanin EK (1995) Separation of variables. New trends. *Progress in Theoretical Physics* 118: 35–60.

Stäckel P (1897) Über die Integration der Hamilton'schen Differentialgleichung mittelst Separation der Variabeln. *Math. Ann.* 49(1): 145–147.

Waksjö C and Rauch-Wojciechowski S (2003) How to find separation coordinates for the Hamilton–Jacobi equation: a criterion of separability for natural Hamiltonian systems. *Mathematical Physics, Analysis and Geometry* 6(4): 301–348.

Wojciechowski S (1986) Review of the recent results on integrability of natural Hamiltonian systems. In: Winternitz P (ed.) *Systèmes dynamiques non linéaires: intégrabilité et comportement qualitatif*, Sém. Math. Sup, vol. 102, pp. 294–327. Montreal, QC: Presses Univ. Montréal.


# Separatrix Splitting

**D Treschev**, Moscow State University, Moscow, Russia



Separatrices are asymptotic manifolds in dynamical systems. However, this term is applied usually in the case of a small dimension of the phase space, where these manifolds are hypersurfaces. In the context of separatrix splitting manifolds asymptotic to hyperbolic tori are usually considered, where tori of dimension 0 and 1 are called equilibrium positions and periodic trajectories, respectively. A separatrix can be stable (asymptotic as $t \to +\infty$) and unstable (asymptotic as $t \to -\infty$).

In this article we consider the case of systems with finite-dimensional phase space. Basically we deal with nonautonomous Hamiltonian systems $2\pi$-periodic in time. However, it is useful to keep in mind the fact that the cases of autonomous Hamiltonian systems and symplectic maps are dynamically the same. Some results for non-Hamiltonian perturbations will also be presented. Hamiltonian systems with one-and-a-half or two degrees of freedom as well as area-preserving two-dimensional maps are especially important for us because the results on the separatrix splitting in this case are more clear and complete. Dynamics in such systems is essentially the same. Below we call these systems two dimensional.

We assume that all systems are at least $C^\infty$-smooth.

## Poincaré Integral

Consider a Liouville integrable Hamiltonian system. Then any separatrix either goes to infinity or joins two hyperbolic tori. From a dynamical point of view, the latter case is more interesting. If these tori are different, the situation is called heteroclinic, otherwise homoclinic. Poincaré was the first to notice that after a generic perturbation stable and unstable separatrices become different submanifolds of the phase space. This phenomenon is called the separatrix splitting.

Poincaré (1987) considered perturbations of separatrices homoclinic to a periodic solution in a Hamiltonian system with one-and-a-half degrees of freedom. In this case the system has the form

$$\dot{x} = \frac{\partial H}{\partial y}, \qquad \dot{y} = -\frac{\partial H}{\partial x}, \qquad (x, y) \in D \subset \boldsymbol{R}^2 \qquad [1]$$

where $D$ is an open domain and

$$H(x, y, t, \varepsilon) = H_0(x, y) + \varepsilon H_1(x, y, t) + O(\varepsilon^2) \qquad [2]$$

We assume that $H$ is $2\pi$-periodic in $t$ and $\varepsilon$ is a small parameter. Let $(x_0, y_0) \in D$ be an equilibrium position for the unperturbed $(\varepsilon = 0)$ system: $\operatorname{grad} H_0(x_0, y_0) = 0$. Without loss of generality, $(x_0, y_0) = 0$. In the extended phase space $D \times \boldsymbol{T}$ ($\boldsymbol{T} = \{t \bmod 2\pi\}$ is a one-dimensional torus) instead of the equilibrium we have a $2\pi$-periodic solution $0 \times \boldsymbol{T}$. Suppose that the equilibrium (and therefore, the periodic solution) is hyperbolic and the corresponding stable and unstable separatrices $\Lambda^{s,u}$ are doubled: $\Lambda^s = \Lambda^u = \Lambda$. Let $\gamma(t)$ be a natural parametrization of $\Lambda$, that is, $(x(t), y(t)) = \gamma(t)$ is a solution of eqns [1]. In the extended phase space, we have the asymptotic surface

$$(\gamma(t + \tau), t), \quad t \in \boldsymbol{T}, \ \tau \in \boldsymbol{R}$$

For small values of $\varepsilon$, the perturbed system has a hyperbolic periodic solution $(\sigma_\varepsilon(t), t), \sigma_\varepsilon(t) = O(\varepsilon) \in D$ and the separatrices

$$(\gamma_\varepsilon^{s,u}(t, \tau), t), \quad \gamma_0^{s,u}(t, \tau) = \gamma(t + \tau)$$

Since the addition to the Hamiltonian of a function, depending only on $t$ and $\varepsilon$, does not change the dynamics, without loss of generality we can assume that $H_1(0, 0, t) \equiv 0$. Hence the Poincaré integral

$$\mathcal{P}(\tau) = \int_{-\infty}^{+\infty} H_1(\gamma(t + \tau), t)\, \mathrm{d}t$$

converges. The function $\mathcal{P}$ carries all information on the separatrix splitting in the first approximation in $\varepsilon$.

Periodicity of $H_1$ in $t$ implies $2\pi$-periodicity of $\mathcal{P}(\tau)$. There is also the following obvious identity:

$$\frac{\mathrm{d}\mathcal{P}(\tau)}{\mathrm{d}\tau} = \int_{-\infty}^{+\infty} \{H_0, H_1\}(\gamma(t + \tau), t)\, \mathrm{d}t$$

where $\{,\}$ is the Poisson bracket.

## Melnikov Integral

Melnikov (1963) considered general (not necessarily Hamiltonian) $2\pi$-periodic in $t$ perturbations:

$$\dot{x} = \frac{\partial H_0}{\partial y} + \varepsilon v_1(x, y, t) + O(\varepsilon^2)$$
$$\dot{y} = -\frac{\partial H_0}{\partial x} + \varepsilon v_2(x, y, t) + O(\varepsilon^2)$$

In this case, information on the separatrix splitting in the first approximation is contained in the Melnikov integral

$$\mathcal{M}(\tau) = \int_{-\infty}^{+\infty} vH_0(\gamma(t + \tau), t)\, \mathrm{d}t$$

where $vH_0 = v_1 \partial H_0/\partial x + v_2 \partial H_0/\partial y$.

Note that if the vector field $v$ is Hamiltonian and $H_1$ is the corresponding Hamiltonian function, we have: $vH_0 = -\{H_0, H_1\}$. Hence in Hamiltonian systems we have: $\mathcal{M}(\tau) = -\mathrm{d}\mathcal{P}(\tau)/\mathrm{d}\tau$.

A multidimensional version of the Melnikov integral is presented in Wiggins (1988).

## Geometric Meaning of $\mathcal{M}(\tau)$

Let $\Gamma_T$ be a compact piece of the unperturbed separatrix

$$\Gamma_T = \{(x, y) \in D\colon (x, y) = \gamma(t),\ |t| \le T\}$$

Then for any $T > 0$ there exists a neighborhood $U$ of $\Gamma_T$ and symplectic coordinates (time–energy coordinates) $\tau$, $h$ on $U$ such that the section of the perturbed separatrices $\Lambda_\varepsilon^{s,u}$ by the plane $\{t = 0\}$ is as follows:

$$\Lambda_\varepsilon^{s,u}|_{t=0} = \{(\tau, h) : h = h_\varepsilon^{s,u}(\tau)\}$$

where

1. $h_\varepsilon^u(\tau) = O(\varepsilon^2)$,
2. $h_\varepsilon^s(\tau) = -\varepsilon\mathcal{M}(\tau) + O(\varepsilon^2)$.

   Moreover, let $g_\varepsilon^t : D \to D$ be the phase flow of the perturbed system. The map $g_\varepsilon^{2\pi}$ is called the Poincaré map. The following statement holds.
3. For any two points $z_0, z_1 \in U$ such that $z_1 = g_\varepsilon^{2\pi}(z_0)$, let $(\tau_0, h_0)$ and $(\tau_1, h_1)$ be their time–energy coordinates. Then

$$\tau_1 = \tau_0 + 2\pi + O(\varepsilon), \qquad h_1 = h_0 + O(\varepsilon)$$

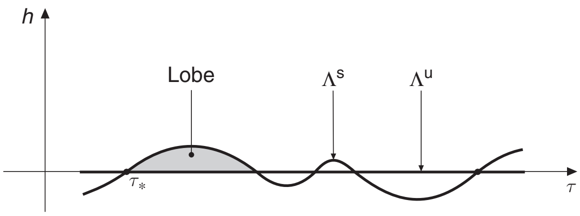



**Figure 1** Perturbed separatrices in time–energy coordinates.

Existence of such coordinates has several corollaries.

- If $\mathcal{P}$ is not identically constant, the separatrices split and this splitting is of the first order in $\varepsilon$.
- Let $\tau_*$ be a simple zero of $\mathcal{M}$. Then the perturbed separatrices intersect transversally at a point $z_*(\varepsilon)$ with time–energy coordinates $(\tau_* + O(\varepsilon), O(\varepsilon^2), t = 0)$. Such a point $z_*(\varepsilon)$ is called a transversal homoclinic point. It generates a doubly asymptotic solution in the perturbed system.
- Consider a lobe domain $\mathcal{L}(\tau_*, \varepsilon)$ bounded by two segments of separatrices on the section $\{t = 0\}$ (see Figure 1). Let another "corner point" of the lobe $\mathcal{L}(\tau_*, \varepsilon)$ correspond to the simple zero $\tau_*'$ of $\mathcal{M}$. Then the symplectic area of $\mathcal{L}(\tau_*, \varepsilon)$ equals

$$\mathcal{A}_{\mathcal{L}}(\tau_*, \varepsilon) = -\varepsilon \int_{\tau_*}^{\tau_*'} \mathcal{M}(\tau)\, d\tau + O(\varepsilon^2)$$

## A Standard Example

Consider as an example a pendulum with periodically oscillating suspension point. The Hamiltonian of the system can be presented in the form

$$H(x, y, t, \varepsilon) = \tfrac{1}{2}y^2 + \Omega^2 \cos x + \varepsilon\theta(t)\cos x \qquad [3]$$

where $\Omega$ is the "internal" frequency of the pendulum. The function $\theta$ is $2\pi$-periodic in time. Hence the frequency of the suspension point oscillation equals 1. In this case, the unperturbed homoclinic solution $\gamma(t)$ can be computed explicitly. In particular,

$$\cos(x(t)) = 1 - 2\cosh^{-2}(\Omega t)$$

Therefore, $\mathcal{P}(\tau) = \int_{-\infty}^{+\infty} \theta(t)(\cos(x(t+\tau)) - 1)\, dt$. For example, if $\theta(t) = \cos t$, we have

$$\mathcal{P}(\tau) = -\frac{2\pi \cos\tau}{\Omega^2 \sinh(\pi/2\Omega)}$$

In this case, different lobes have the same area

$$\mathcal{A}_{\mathcal{L}} = \frac{4\varepsilon\pi}{\Omega^2 \sinh(\pi/2\Omega)} + O(\varepsilon^2)$$

## Multidimensional Case

Multidimensional generalization of the Poincaré–Melnikov construction is strongly connected to the concept of a (partially) hyperbolic torus. Let $(M, \omega, H)$ be a Hamiltonian system on the $2m$-dimensional symplectic manifold $(M, \omega)$.

An invariant $n$-torus $N \subset M\, (0 \le n < m)$ is called hyperbolic if there exist coordinates $x, y, z$ on $M$ in a neighborhood of $N$ such that

1. $y = (y_1, \ldots, y_n)$, $x = (x_1, \ldots, x_n) \bmod 2\pi$, $z = (z^s, z^u)$, $z^{s,u} = (z_1^{s,u}, \ldots, z_l^{s,u})$, $l + n = m$;
2. $\omega = dy \wedge dx + dz^u \wedge dz^s$;
3. $N = \{(x, y, z) : y = 0,\ z = 0\}$; and
4. $H = \langle \nu, y\rangle + (1/2)\langle Ay, y\rangle + \langle z^u, \Omega(x) z^s\rangle + O_3(y, z)$,

where $\nu \in \boldsymbol{R}^n$ is a constant vector, $A$ is a constant $n \times n$ matrix, $\Omega$ is an $l \times l$ matrix such that $\Omega(x) + \Omega^T(x)$ is positive definite for any $x \bmod 2\pi$, the symbol $O_3$ denotes terms of order not less than 3, and $\langle a, b\rangle = \sum a_j b_j$.

If $\det A \ne 0$, the torus is called nondegenerate. If $\nu$ is Diophantine, that is, for some $\alpha, \beta > 0$ and any $0 \ne k \in \boldsymbol{Z}^n$

$$|\langle \nu, k\rangle| \ge \alpha |k|^{-\beta}$$

the torus $N$ is called Diophantine. The coordinates $(x, y, z)$ are called canonical for $N$.

Now suppose that the Hamiltonian $H$ depends smoothly on the parameter $\varepsilon$:

$$H = H_0 + \varepsilon H_1 + O(\varepsilon^2)$$

and for $\varepsilon = 0$ the system is Liouville integrable with the commuting first integrals $F_1, \ldots, F_m$:

$$\{F_j, F_k\} = 0, \quad 1 \le j, k \le m$$

Let $M_0 = \{F_1 = \cdots = F_m = 0\} \subset M$ be their zero common level and let $N \subset M$ be an $n$-dimensional nondegenerate Diophantine hyperbolic torus. The torus $N$ generates the invariant Lagrangian asymptotic manifolds $\Lambda^{s,u} \subset M$. Suppose that the separatrices are doubled, that is, there is a Lagrangian manifold $\Lambda \subset \Lambda^s \cap \Lambda^u$.

Consider the perturbed Hamiltonian $H = H_0 + \varepsilon H_1 + O(\varepsilon^2)$. The torus $N$ as well as the asymptotic manifolds $\Lambda^{s,u}$ survive the perturbation. Let $N_\varepsilon$ be the corresponding hyperbolic torus in the perturbed system and $\Lambda_\varepsilon^{s,u}$ its asymptotic manifolds: $N_\varepsilon$ and $\Lambda_\varepsilon^{s,u}$ depend smoothly on $\varepsilon$ and $N_0 = N$, $\Lambda_0^{s,u} = \Lambda^{s,u}$.

Let the function $\chi(x)$ satisfy the equation

$$\langle \nu, \partial\chi(x)/\partial x\rangle + H_1(x, 0, 0) = \frac{1}{(2\pi)^n}\int_{T^n} H_1(x, 0, 0)\, dx$$

This equation has a smooth solution unique up to an additive constant.

Consider a solution of the unperturbed Hamiltonian equations $\gamma(t) \subset \Lambda$. Let $I_j^\gamma, I_{j,l}^\gamma, 1 \le j, l \le m$ be the following quantities (Treschev 1994):

$$I_j^\gamma = \lim_{T \to +\infty} \Big( - \int_{-T}^{T} \{F_j, H_1\}(\gamma(t))\, dt + \{F_j, \chi\}(\gamma(-T)) - \{F_j, \chi\}(\gamma(T)) \Big)$$

$$I_{j,l}^\gamma = \lim_{T \to +\infty} \Big( - \int_{-T}^{T} \{F_j, \{F_l, H_1\}\}(\gamma(t))\, dt + \{F_j\{F_l, \chi\}\}(\gamma(-T)) - \{F_j\{F_l, \chi\}\}(\gamma(T)) \Big)$$

The numbers $I_j^\gamma, I_{j,l}^\gamma$ play the role of the first and second derivatives of the Poincaré integral at some point.

If any of the quantities $I_j^\gamma, I_{j,l}^\gamma$ does not vanish, the asymptotic manifolds $\Lambda^{s,u}$ split. Moreover, suppose that $I_1^\gamma = \cdots = I_m^\gamma = 0$ and the rank of the matrix $(I_{j,l}^\gamma)$ equals $m-1$. Then for small values of $\varepsilon$, the manifolds $\Lambda^s$ and $\Lambda^u$ intersect transversally on the energy level at points of the solution $\gamma_\varepsilon(t)$, where $\gamma_\varepsilon \to \gamma$ as $\varepsilon \to 0$.

## Poincaré Integral in Multidimensional Case

Suppose that the Hamiltonian from the previous section equals

$$H(x, y, u, v, t, \varepsilon) = H_0(y, u, v) + \varepsilon H_1(x, y, u, v, t) + O(\varepsilon^2)$$

Here $x = (x_1, \ldots, x_n) \bmod 2\pi$, $y = (y_1, \ldots, y_n) \in \boldsymbol{R}^n$, and $(u, v) \in \boldsymbol{R}^2$. The symplectic structure $\omega = dy \wedge dx + dv \wedge du$.

We assume that in the unperturbed integrable system the variables separate:

$$H_0(y, u, v) = F(y) + f(u, v)$$

and the system with one degree of freedom and Hamiltonian $f$ has a hyperbolic equilibrium $(u, v) = 0$ with a homoclinic solution $\gamma(t)$. Any torus

$$N(y^0) = \{(x, y, u, v, t) \colon y = y^0,\ u = v = 0\}$$

is a hyperbolic torus of the unperturbed system with frequency vector

$$\begin{pmatrix} \nu(y^0) \\ 1 \end{pmatrix}, \quad \nu(y) = \partial F / \partial y$$

Suppose that $N = N(0)$ is Diophantine and nondegenerate. Then in the perturbed system there is smooth in $\varepsilon$ hyperbolic torus $N_\varepsilon$, $N_0 = N$. Consider the Poincaré function

$$\mathcal{P}(\xi, \tau) = \int_{-\infty}^{+\infty} \Big( H_1(\xi + \nu(t+\tau), 0, \gamma(t+\tau), t) - H_1(\xi + \nu(t+\tau), 0, 0, 0, t) \Big) dt$$

Obviously, $\mathcal{P}(\xi, \tau)$ is $2\pi$-periodic in $\xi$ and $\tau$.

If $\mathcal{P}$ is not identically constant, asymptotic surfaces of $N_\varepsilon$ split in the first approximation in $\varepsilon$. Nondegenerate critical points of $\mathcal{P}$ correspond to transversal homoclinic solutions of the perturbed system.

Other results on the splitting of multidimensional asymptotic manifolds are presented in Arnol'd *et al.* (1988) and Lochak *et al.* (2003).

## Exponentially Small Separatrix Splitting

If in the unperturbed (integrable) system there are no asymptotic manifolds, they can appear after a perturbation. Consider, for example, perturbation of a real-analytic Liouville integrable system near a simple resonance:

$$\dot{x} = \frac{\partial H}{\partial y}, \quad \dot{y} = \frac{\partial H}{\partial x}, \quad x \in \boldsymbol{T}^m, \quad y \in D \subset \boldsymbol{R}^m$$

$$H(x, y, t, \varepsilon) = H_0(y) + \varepsilon H_1(x, y, t, \varepsilon)$$

As usual, we assume $2\pi$-periodicity in $t$. A simple resonance corresponds to a value of the action variable $y = y^0$ such that the frequency vector

$$\hat{\nu} = \begin{pmatrix} \nu^0 \\ 1 \end{pmatrix}, \quad \nu^0 = \frac{\partial H_0}{\partial y}(y^0) \in \boldsymbol{R}^m$$

(here 1 is the frequency, corresponding to the time variable) admits only one resonance. More precisely, there exists a nonzero $\hat{k} \in \boldsymbol{Z}^{m+1}$, satisfying $\langle \hat{k}, \hat{\nu} \rangle = 0$ and any $k \in \boldsymbol{Z}^{m+1}$ such that $\langle k, \hat{\nu} \rangle = 0$ is collinear with $\hat{k}$.

Without loss of generality, we can assume that $y^0 = 0$ and $\nu^0 = \binom{0}{\tilde{\nu}}$, $\tilde{\nu} \in \boldsymbol{R}^{m-1}$. Then the vector $\bar{\nu} = \binom{\tilde{\nu}}{1} \in \boldsymbol{R}^m$ is nonresonant.

In a $\sqrt{\varepsilon}$-neighborhood of the resonance we have a system with fast variables $X = (x_2, \ldots, x_m, t) \bmod 2\pi$ and slow variables $Y = (x_1, \varepsilon^{-1/2} y_1, \ldots, \varepsilon^{-1/2} y_m)$ variables:

$$\dot{Y} = O(\sqrt{\varepsilon}), \qquad \dot{X} = \bar{\nu} + O(\sqrt{\varepsilon}) \qquad [4]$$

If the frequency vector $\bar{\nu}$ is Diophantine, by using the Neishtadt averaging procedure, we can reduce the dependence of the right-hand sides of eqns [4] on

the fast variables to exponentially small in $\varepsilon$ terms. This means that there exist new symplectic variables

$$P = Y + O(\sqrt{\varepsilon}), \quad Q = X + O(\sqrt{\varepsilon})$$

(new time coincides with the old one) such that system [4] takes the form

$$\dot{P} = \sqrt{\varepsilon} F(P, \sqrt{\varepsilon}) + O\big(\exp(-a\varepsilon^{-b})\big)$$
$$\dot{Q} = \overline{\nu} + \sqrt{\varepsilon} G(P, \sqrt{\varepsilon}) + O\big(\exp(-a\varepsilon^{-b})\big)$$

with positive constants $a, b$.

If we neglect the exponentially small reminders, the system turns out to be integrable. Generically, it has a family of hyperbolic $m$-tori of the form $\{(P, Q)\colon P = \text{const.}\}$ with doubled asymptotic manifolds. However, the terms $O(\exp(-a\varepsilon^{-b}))$ generically cannot be removed completely. They produce an exponentially small splitting of the asymptotic manifolds. This splitting implies nonintegrability, chaotic behavior, Arnol'd diffusion, and other dynamical effects.

It is important to note that exponentially small splitting appears only in the analytic case. In smooth systems the splitting is much stronger.

Unfortunately, at present there are no quantitative methods for studying such splittings except obvious upper estimates and the case of two-dimensional systems.

## Exponentially Small Splitting in Two-Dimensional Systems

The main results on exponentially small separatrix splitting were obtained by Lasutkin and his students (Gelfreich and others). Another effective approach was proposed by Treschev. There are no general theorems in this situation; however, many examples were studied. We discuss the splitting in the pendulum with rapidly oscillating suspension point. The Hamiltonian of the system has the form

$$H = \tfrac{1}{2} y^2 + (1 + 2b\cos(t/\varepsilon))\cos x$$

(cf. [3]). For any value of $\varepsilon$ the circle $\{(x, y, t)\colon x = y = 0\}$ is a periodic trajectory. For small $\varepsilon > 0$ the trajectory is hyperbolic.

Poincaré integral can be formally written in this system. It predicts the area of lobes $16\pi b\varepsilon^{-1} \mathrm{e}^{-\pi(2\varepsilon)^{-1}}$. However, there is no reason to expect that this asymptotics of the splitting is correct. Indeed, its value is exponentially small in $\varepsilon$, while the error of the Poincaré–Melnikov method is in general quadratic in the perturbation. To obtain correct asymptotics of the separatrix splitting, one has to study singularities of the solutions with respect to complex time. Area of lobes in this system equals (Treschev 1997)

$$\mathcal{A}_{\mathcal{L}} = 4bf(b, \varepsilon)\varepsilon^{-1} \mathrm{e}^{-\pi(2\varepsilon)^{-1}}$$

Here $f(b, \varepsilon), \varepsilon \geq 0$ is a smooth function. The function $f(b, 0)$ is even and entire. It can be computed numerically as a solution of a problem which does not contain $\varepsilon$. The value $f(0, 0) = 4\pi$ corresponds to the Poincaré integral, but the function $f(b, 0)$ is not constant. It is possible to prove that $f$ can be expanded in a power series in $\varepsilon$. Apparently, this series diverges for any $b \neq 0$.

## Separatrix Splitting and Dynamics

1. Separatrix splitting can be regarded as an obstacle to the integrability of the perturbed system. However, this statement needs some comments. Doubled asymptotic surfaces in an integrable Hamiltonian system can have self-intersections. In the case of equilibrium, such intersections can even be transversal. In the literature, there is no general result saying that separatrix splitting implies nonintegrability. Some particular cases (studied by Kozlov, Ziglin, Bolotin, and others) are presented in Arnol'd *et al.* (1989). For example, in the two-dimensional case, this is seen to be true.
2. Conceptual reason for the nonintegrability, discussed in the previous item, is a complicated dynamics near the splitted separatrices. In many situations, it is possible to find in this domain a Smale horseshoe. This implies positive topological entropy, existence of nontrivial hyperbolic sets, symbolic dynamics, etc.
3. Consider a near-integrable area-preserving two-dimensional map. In the perturbed system in the vicinity of the splitted separatrices of a hyperbolic fixed point $z_\varepsilon$ the so-called stochastic layer is formed. Here we mean the domain bounded by invariant curves, closest to the separatrices. An important quantity, describing the rate of chaos, is the area of the stochastic layer $\mathcal{A}_{\mathrm{SL}}(\varepsilon)$. It turns out (Treschev 1998b) that $\mathcal{A}_{\mathrm{SL}}(\varepsilon)$ is connected with the area of the largest lobe $\mathcal{A}_{\mathcal{L}}(\varepsilon)$ by the simple formula

$$c_1 \mathcal{A}_{\mathrm{SL}}(\varepsilon) < \frac{\mathcal{A}_{\mathcal{L}}(\varepsilon)\log(\mathcal{A}_{\mathcal{L}}(\varepsilon))}{\log^2 \mu} < c_2 \mathcal{A}_{\mathrm{SL}}(\varepsilon)$$

   with some constants $c_1, c_2 > 0$, where $\mu$ is the largest multiplier (Lyapunov exponent) of the fixed point $z_0$.
4. Let $\hat{z}$ be a hyperbolic fixed point of an area-preserving two-dimensional map. The point $\hat{z}$

divides the corresponding separatrices $\Lambda^{s,u}$ in 4 branches $\Lambda^s_{1,2}$ and $\Lambda^u_{1,2}$. Suppose that the pair of branches $\Lambda^s_j$ and $\Lambda^u_l$ satisfies the following conditions:

- $\Lambda^s_j$ and $\Lambda^u_l$ lie in a compact invariant domain;
- $\Lambda^s_j$ and $\Lambda^u_l$ do not coincide and intersect at a homoclinic point.

Then the closures $\overline{\Lambda}^s_j, \overline{\Lambda}^u_l$ are compact invariant sets. Very little is known about these sets. For example, it is not known if their measure is positive. However, by using the Poincaré recurrence theorem, it is possible to prove (Treschev 1998a) that $\overline{\Lambda}^s_j = \overline{\Lambda}^u_l$.

*See also:* Averaging Methods; Billiards in Bounded Convex Domains; Hamiltonian Systems: Obstructions to Integrability; Hamiltonian Systems: Stability and Instability Theory.

## Further Reading

Arnol'd V, Kozlov V, and Neishtadt A (1988) Mathematical aspects of classical and celestial mechanics. In: *Encyclopaedia of Mathematical Sciences*, vol. 3. Berlin: Springer.

Lochak P, Marco J-P, and Sauzin D (2003) On the splitting of invariant manifolds in multidimensional near-integrable Hamiltonian systems. *Memoirs of the American Mathematical Society* 163(775): viii+145.

Melnikov V (1963) On the stability of the center for time-periodic perturbations. *Trudy Moskovskogo Metern. Obschestva* 12: 3–52 (Russian). (English transl.: Trans. *Moscow Math. Soc.* 1963, pp. 1–56 (1965).)

Poincaré H (1987) *Les méthodes nouvelles de la Mécanique Céleste*, vols. 1–3. Paris: Gauthier–Villars, (Original publication: 1892, 1893, 1899). New Printing: Librairie Scientifique et Technique Albert Blanchard 9, Rue Medecin 75006, Paris.

Treschev D (1994) Hyperbolic tori and asymptotic surfaces in Hamiltonian systems. *Russian Journal of Mathematical Physics* 2(1): 93–110.

Treschev D (1997) Separatrix splitting for a pendulum with rapidly oscillating suspension point. *Russian Journal of Mathematical Physics* 5(1): 63–98.

Treschev D (1998a) Closures of asymptotic curves in a two-dimensional symplectic map. *J. Dynam. Control Systems* 4(3): 305–314.

Treschev D (1998b) Width of stochastic layers in near-integrable two-dimensional symplectic maps. *Physica D* 116(1–2): 21–43.

Wiggins S (1988) *Global Bifurcations and Chaos. Analytical Methods. Applied Mathematical Sciences,* vol. 73, 494pp New York: Springer.

# Several Complex Variables: Basic Geometric Theory

**A Huckleberry**, Ruhr-Universität Bochum, Bochum, Germany
**T Peternell**, Universität Bayreuth, Bayreuth, Germany



## Introduction

The rubric "several complex variables" is attached to a wide area of mathematics which involves the study of holomorphic phenomena in dimensions higher than one. In this area there are viewpoints, methods and results which range from those on the analytic side, where analytic techniques of partial differential equations (PDEs) are involved, to those of algebraic geometry which pertain to varieties defined over finite fields. Here we outline selected basic methods which are aimed at understanding global geometric phenomena. Detailed presentations of most results discussed here can be found in the basic texts (Demailly, Grauert and Fritzsche 2001, Griffiths and Harris 1978, Grauert *et al.* 1994, Grauert and Remmert 1979, 1984).

## Domains in $\mathbb{C}^n$

Complex analysis begins with the study of holomorphic functions on domains $D$ in $\mathbb{C}^n$. These are smooth complex-valued functions $f$ which satisfy

$$\bar{\partial} f := \sum \frac{\partial f}{\partial \bar{z}_i} \mathrm{d}\bar{z}_i = 0$$

Some results from the one-dimensional theory extend to the case where $n > 1$. However, even at the early stages of development, one sees that there are many new phenomena in the higher-dimensional setting.

### Extending Results from the One-Dimensional Theory

For local results one may restrict considerations to functions $f$ which are holomorphic in a neighborhood of $0 \in \mathbb{C}^n$. The restriction of $f$ to, for example, any complex line through 0 is holomorphic, and therefore the maximum principle can be immediately transferred to the higher-dimensional setting.

The zero-set $V(f)$ of a nonconstant holomorphic function is one-codimensional over the complex numbers (two-codimensional over the reals). Thus the identity principle must be formulated in a different way from its one-dimensional version. For example, under the usual connectivity assumptions, if $f$ vanishes on a set $E$ with Hausdorff dimension bigger than $2n-2$, then it vanishes identically. Here

is another useful version: if $M$ is a real submanifold such that the real tangent space $T_zM$ generates the full complex tangent space at one of its points, that is, $T_zM + iT_zM = T_z\mathbb{C}^n$, and $f|M \equiv 0$, then $f \equiv 0$.

In the one-dimensional theory, after choosing appropriate holomorphic coordinates, $f(z) = z^k$ for some $k$. This local normal form implies that nonconstant holomorphic functions are open mappings. Positive results in the mapping theory of several complex variables are discussed below. The simple example $F: \mathbb{C}^2 \to \mathbb{C}^2, (z, w) \to (z, zw)$, shows that the open mapping theorem cannot be transferred without further assumptions.

The local normal-form theorem in several complex variables is called the "Weierstraß preparation theorem." It states that after appropriate normalization of the coordinates, $f$ is locally the product of a nonvanishing holomorphic function with a "polynomial"

$$P(z, z') = z^k + a_{k-1}(z')z^{k-1} + \cdots + a_0(z')$$

where $z$ is a single complex variable, $z'$ denotes the remaining $n-1$ variables, and the coefficients are holomorphic in $z'$. This is a strong inductive device for the local theory.

If $D$ is a product $D = D_1 \times \cdots \times D_n$ of relatively compact domains in the complex plane $\mathbb{C}$, then repeated integration transfers the one-variable Cauchy integral formula from the $D_i$ to $D$. The resulting integral is over the product $\mathrm{bd}(D_1) \times \cdots \times \mathrm{bd}(D_n)$ of the boundaries which is topologically a small set in $\mathrm{bd}(D)$. Complex analytically it is, however, large in the sense of the above identity principle.

It follows from, for example, the $n$-variable Cauchy integral formula that holomorphic functions agree with their convergent power series developments. As in the one-variable theory, the appropriate topology on the space $\mathcal{O}(D)$ of holomorphic functions on $D$ is that of uniform convergence on compact subsets. In this way $\mathcal{O}(D)$ is equipped with the topology of a Fréchet space.

### First Theorems on Analytic Continuation

Analytic continuation is a fundamental phenomenon in complex geometry. One type of continuation theorem which is known in the one-variable theory is of the following type: If $E$ is a small closed set in $D$ and $f \in \mathcal{O}(D\backslash E)$ is a holomorphic function which satisfies some growth condition near $E$, then it extends holomorphically to $D$. The notion "small" can be discussed in terms of measure, but it is more appropriate to discuss it in complex analytic terms.

An analytic subset $A$ of $D$ is locally the common zero set $\{a \in D; f_1(a) = \cdots = f_m(a) = 0\}$ of finitely many holomorphic functions. A function $g$ on $A$ is said to be holomorphic if at each $a \in A$ it is the restriction of a holomorphic function on some neighborhood of $a$ in $D$. There is an appropriate notion of an irreducible component of $A$. If $A$ is irreducible, it contains a dense open set $A_{\mathrm{reg}}$, which is a connected $k$-dimensional complex manifold, that is, at each of its points $a$ there are functions $f_1, \ldots, f_k$ which define a map $F := (f_1, \ldots, f_k)$, which is a holomorphic diffeomorphism of $A_{reg}$ onto an open set in $\mathbb{C}^k$. The boundary $A_{\mathrm{sing}}$ is the set of singular points of $A$, which is a lower-dimensional analytic set. The dimension of an analytic set is the maximum of the dimensions of its irreducible components.

Here are typical examples of theorems on continuing holomorphic functions across small analytic sets $E$. If codim $E \geq 2$, then every function which is holomorphic on $D\backslash E$ extends to a holomorphic function on $D$. The same is true of meromorphic functions, that is, functions which are locally defined as quotients $m = f/g$ of holomorphic functions. If $f$ is holomorphic on $D$, then $g := 1/f$ is holomorphic outside the analytic set $E := V(f)$. Thus $g$ cannot be holomorphically continued across this one-codimensional set. However, Riemann's Hebbarkeitssatz is valid in several complex variables: if $f$ is locally bounded outside an analytic subset $E$ of any positive codimension, then it extends holomorphically to $D$.

With a bit of care, continuation results of this type can be proved for (reduced) complex spaces. These are defined as paracompact Hausdorff spaces which possess charts $(U_\alpha, \varphi_\alpha)$, where the local homeomorphism $\varphi_\alpha$ identifies the open set $U_\alpha$ with a closed analytic subset $A_\alpha$ of a domain $D_\alpha$ in some $\mathbb{C}^{n_\alpha}$. As indicated above, a continuous function on $A_\alpha$ is holomorphic if at each point it can be holomorphically extended to some neighborhood of that point in $D_\alpha$. Finally, just as in the case of manifolds, the compatibility between charts is guaranteed by requiring that coordinate change $\varphi_{\alpha\beta}: U_{\alpha\beta} \to U_{\beta\alpha}$ is biholomorphic, that is, it is a homeomorphism so that it and its inverse are given by holomorphic functions as $F = (f_1, \ldots, f_m)$. The discussion of irreducible components, sets of singularities, and dimension for complex spaces goes exactly in the same way as that above for analytic sets.

If $E$ is everywhere at least two-codimensional, then the above result on continuation of meromorphic functions holds in complete generality. The Hebbarkeitssatz requires the additional condition that the complex space is normal. In many situations this causes no problem at all, because, in general, there is a canonically defined associated normal

complex space $\tilde{X}$ and a proper, surjective, finite-fibered holomorphic map $\tilde{X} \to X$ which is biholomorphic outside a nowhere-dense proper analytic subset. Difficulties can be overcome by simply lifting functions to this normalization and applying the Hebbarkeitssatz.

Continuation theorems of Hartogs-type reflect the fact that complex analysis in dimensions larger than one is really quite different from the one-variable version. The following is such a theorem. Let $(z, w)$ be the standard coordinates in $\mathbb{C}^2$ and think of the $z$-axis as a parameter space for geometric figures in the $w$-plane. For example, let $D_z := \{(z, w) : |w| < 1\}$ be a disk and $A_z := \{(z, w) : 1 - \varepsilon < |w| < 1\}$ be an annulus. An example of a Hartogs figure $H$ in $\mathbb{C}^2$ is the union of the family of disks $D_z$ for $|z| < 1 - \delta$ with the family $A_z$ of annuli for $1 - \delta \leq |z| < 1$. One should visualize the moving disks which suddenly change to moving annuli. One speaks of filling in the Hartogs figure to obtain the polydisk $\hat{H} := \{(z, w); |z|, |w| < 1\}$. Hartogs' continuation theorem states that a function which is holomorphic on $H$ extends holomorphically to $\hat{H}$.

### Cartan–Thullen Theorem

One of the major developments in complex analysis in several variables was the realization that certain convexity concepts lie behind the strong continuation properties. At the analytic level one such is defined as follows by the full algebra of holomorphic functions $\mathcal{O}(D)$. If $K$ is a compact subset of $D$, then its holomorphic convex hull $\hat{K}$ is defined as the intersection of the sets $P(f) := \{p \in D : |f(p)| \leq |f|_K\}$ as $f$ runs through $\mathcal{O}(D)$. One says that $D$ is holomorphically convex if $\hat{K}$ is compact for every compact subset $K$ of $D$.

The theorem of H. Cartan and Thullen relates this concept to analytic continuation phenomena as follows. A domain $D$ is said to be a domain of holomorphy if, given a divergent sequence $\{z_n\} \subset D$, there exists $f \in \mathcal{O}(D)$ which is unbounded along it. In other words, the phenomenon of being able to extend all holomorphic functions on $D$ to a truly larger domain $\hat{D}$ does not occur. The Cartan–Thullen theorem states that $D$ is a domain of holomorphy if and only if it is holomorphically convex. In the next paragraph the relation between this type of convexity and a certain complex geometric convexity of the boundary bd$(D)$ will be indicated.

### Levi Theorem and the Levi Problem

Consider a smooth (local) real hypersurface $\Sigma$ containing $0 \in \mathbb{C}^n$ with $n > 1$. It is the zero-set $\{\rho = 0\}$ in some neighborhood $U$ of 0 of a smooth function with $\mathrm{d}\rho \neq 0$ on $U$. This is viewed as a piece of a boundary of a domain $D$, where $U \cap D = \{\rho < 0\}$. The real tangent space $T_0\Sigma = \mathrm{Ker}(\mathrm{d}\rho(0))$ contains a unique maximal (one-codimensional) complex subspace $T_0^{\mathbb{C}}\Sigma = \mathrm{Ker}(\partial\rho(0)) = H$. The signature of the restriction of the complex Hessian (or Levi form) $\mathrm{i}\partial\bar{\partial}\rho$ to $H$ is a biholomorphic invariant of $\Sigma$. In this notation the Hessian is a real alternating 2-form which is compatible with the complex structure, and its signature is defined to be the signature of the associated symmetric form.

If the restriction of this Levi form to the complex tangent space has a negative eigenvalue, that is, if the boundary bd$(D)$ has a certain degree of concavity, then there is a map $F : \Delta \to U$ of the unit disk $\Delta$ which is biholomorphic onto its image with $F(0) = 0$ and otherwise $F(\mathrm{cl}(\Delta)) \subset D$. The reader can imagine pushing the image of this map into the domain to obtain a family of disks which are in the domain, and pushing it in the outward pointing direction to obtain annuli which are also in the domain. Making this precise, one builds a (higher-dimensional) Hartogs figure $H$ at the base point 0 so that $\hat{H}$ is an open neighborhood of 0. In particular this proves the theorem of E. E. Levi: *every function holomorphic on $U \cap D$ extends to a neighborhood of* 0. This can be globally formulated as follows:

**Theorem** *If $D$ is a domain of holomorphy with smooth boundary in $\mathbb{C}^n$, then* bd$(D)$ *is Levi-pseudoconvex.*

Here the terminology Levi-pseudoconvex is used to denote the condition that the restriction of the Levi form to the complex tangent space of every boundary point is positive semidefinite.

One of the guiding problems of complex analysis in higher dimensions is the Levi problem. This is the converse statement to that of the Levi's theorem:

**Levi Problem** *Is a domain $D$ with smooth Levi-pseudoconvex boundary in a complex manifold necessarily a domain of holomorphy?*

Stated in this form it is not true, but for domains in $\mathbb{C}^n$ it is true. As will be sketched below, under stronger assumptions on the Levi form it is almost true. However, there are still interesting open problems in complex analysis which are related to the Levi problem.

### Bounded Domains and Their Automorphisms

The unit disk in the complex plane is particularly important, because, with the exceptions of projective

space $\mathbb{P}_1(\mathbb{C})$, the complex plane $\mathbb{C}$, the punctured plane $\mathbb{C}\backslash\{0\}$, and compact complex tori, it is the universal cover of every (connected) one-dimensional complex manifold.

In higher dimensions it should first be underlined that, without some further condition, there is no best bounded domain in $\mathbb{C}^n$. For example, two randomly chosen small perturbations of the unit ball $\mathbb{B}_2 := \{(z, w); |z|^2 + |w|^2 < 1\}$, with, for example, real analytic boundary, are not biholomorphically equivalent.

On the other hand, the following theorem of H. Cartan shows that bounded domains $D$ are good candidates for covering spaces:

**Theorem** *Equipped with the compact open topology, the group* Aut($D$) *of holomorphic automorphisms of $D$ is a Lie group acting properly on $D$.*

The notion of a proper group action of a topological group on a topological space is fundamental and should be underlined. It means that if $\{x_n\}$ is a convergent sequence in the space where the group is acting, then a sequence of group elements $\{g_n\}$, with the property that $\{g_n(x_n)\}$ is convergent, itself possesses a convergent subsequence. As a consequence, isotropy groups are compact and orbits are closed.

In the context of bounded domains $D$ this implies that if $\Gamma$ is a discrete subgroup of Aut($D$), then $X = D/\Gamma$ carries a natural structure of a complex space. If in addition $\Gamma$ is acting freely, something that, with minor modifications, can be arranged, then $X$ is a complex manifold.

Many nontrivial compact complex manifolds arise as quotients $D/\Gamma$ of bounded domains. Even very concrete quotients, for example, where $D = \mathbb{B}_2$, are extremely interesting. Conversely, if Aut($D$) contains a discrete subgroup $\Gamma$ so that $D/\Gamma$ is compact, then $D$ is probably very special. For example, it is known to be holomorphically convex!

Any compact quotient $X = D/\Gamma$ of a bounded domain is projective algebraic in the sense that it can be realized as a complex (algebraic) submanifold of some complex projective space. In fact the embedding can be given by quite special $\Gamma$-invariant holomorphic tensors on $D$, and this in turn implies that $X$ is of general type (see below). For further details, in particular on Cartan's theorem on the automorphism group of a bounded domain, the reader is referred to Narasimhan (1971).

## Stein Manifolds

The founding fathers of the first phase of "modern complex analysis" (Cartan, Oka, and Thullen) realized that domains of holomorphy form the basic class of spaces where it would be possible to solve the important problems of the subject concerning the existence of holomorphic or meromorphic functions with reasonably prescribed properties. In fact, Oka formulated a principle which more or less states that if a complex analytic problem which is well formulated on a domain of holomorphy has a continuous solution, then it should have a holomorphic solution. Given the flexibility of continuous functions and the rigidity of holomorphic functions, this would seem impossible but in fact is true!

Beginning in the late 1930s, Stein worked on problems related to this Oka principle, in particular on those related to what we would now call the algebraic topological aspects of the subject, and he was led to formulate conditions on a general complex manifold $X$ which should hold if problems of the above type are to be solved. First, his axiom of holomorphic convexity was simply that, given a divergent sequence $\{x_n\}$ in $X$, there should be a function $f \in \mathcal{O}(X)$ such that $\{f(x_n)\}$ is unbounded. Secondly, holomorphic functions should separate points in the sense that, given distinct points $x_1, x_2 \in X$, there exists $f \in \mathcal{O}(X)$ with $f(x_1) \neq f(x_2)$. Finally, globally defined holomorphic functions should give local coordinates. Assuming that $X$ is $n$-dimensional, this means that, given a point $x \in X$, there exist $f_1, \ldots, f_n \in \mathcal{O}(X)$ such that $\mathrm{d}f_1(x) \wedge \cdots \wedge \mathrm{d}f_n(x) \neq 0$.

Assuming Stein's axioms, Cartan and Serre then produced a powerful theory in the context of sheaf cohomology which proved certain vanishing theorems that led to the desired existence theorems. This theory and typical applications are sketched below. Before going into this, we would like to mention that Grauert's version of the Cartan–Serre theory requires only very weak versions of Stein's axioms: (1) The connected component containing $K$ of the holomorphic convex hull $\hat{K}$ of every compact set should be compact. (2) Given $x \in X$, there are functions $f_1, \ldots, f_m \in \mathcal{O}(X)$ so that $x$ is an isolated point in the fiber of the map $F := (f_1, \ldots, f_m) : X \to \mathbb{C}^m$. Of course the results also hold for complex spaces.

Holomorphically convex domains in $\mathbb{C}^n$ are Stein manifolds, and since closed complex manifolds of Stein manifolds are Stein, it follows that any complex submanifold of $\mathbb{C}^n$ is Stein. In particular, affine varieties are Stein spaces. Remmert's theorem states the converse: an $n$-dimensional Stein manifold can be embedded as a closed complex submanifold of $\mathbb{C}^{2n+1}$. A nontrivial result of Behnke and Stein implies that every noncompact Riemann surface is also Stein.

### Basic Formalism

The following first Cousin problem is typical of those which can be solved by Stein theory. Let $X$ be a complex manifold which is covered by open sets $U_i$. Suppose that on each such set a meromorphic function $m_i$ is given so that on the overlap $U_{ij} := U_i \cap U_j$ the difference $m_{ij} = m_j - m_i =: f_{ij}$ is holomorphic. This means that the distribution of polar parts of these functions is well defined. The question is whether or not there exists a globally defined meromorphic function $m \in \mathcal{M}(X)$ with these prescribed polar parts, that is, with $m - m_i =: f_i \in \mathcal{O}(U_i)$.

If one applies the Oka principle, this problem can be easily solved. For this one can assume that the covering is locally finite and take $\gamma_i$ to be a partition of unity subordinate to the cover. Using standard shrinking and cut-off arguments, one extends the $f_{ij}$ to the full space $X$ as smooth functions so that the alternating cocycle relations $f_{ij} + f_{jk} + f_{ki} = 0$ and $f_{ij} = -f_{ji}$ still hold. Then $f_j := \sum \gamma_k f_{jk}$ is a smooth function on $U_j$ which satisfies $f_j - f_i = f_{ij}$ on the overlap $U_{ij}$. It follows that $f := m_i + f_i = m_j + f_j$ is a globally well-defined "smooth" function with the prescribed polar parts. The Oka principle would then imply that there is a globally defined meromorphic function with the same property.

The basic sheaf cohomological formalism for Stein theory can be seen in the above argument. Suppose that instead of applying extension and cut-off techniques from the smooth category, we could answer positively the question "given the holomorphic functions $\{f_{ij}\}$ on the $U_{ij}$, do there exist holomorphic functions $\{f_i\}$ on the $U_i$ such that $f_j - f_i = f_{ij}$ on the $U_{ij}$?" Then we would immediately have the desired globally defined meromorphic function $m := m_i + f_i$. This question is exactly the question of whether or not the Cech cohomology class of the alternating cocycle $\{f_{ij}\}$ vanishes.

Let us quickly summarize the language of Cech cohomology. A presheaf of abelian groups is a mapping $U \to S(U)$ which associates to every open subset of $X$ an abelian group. Typical examples are $U \to \mathcal{O}(U)$, $U \to C^\infty(U)$, $U \to H^*(U, \mathbb{Z}), \ldots$ . The last example which associates to $U$ its topological cohomology does not localize well in terms of following the basic axioms for a sheaf: (1) Given a covering $\{U_i\}$ of an open subset $U$ of $X$ and elements $s_i \in S(U_i)$ with $s_j - s_i = 0$ on $U_{ij}$, there exists $s \in S(U)$ with $s|U_i = s_i$. (2) If $s, t \in S(U)$ are such that $s|U_i = t|U_i$ for all $i$, then $s = t$. For this we have assumed that the restriction mappings have been built into the definition of a presheaf.

Associated to a sheaf $S$ on $X$ and a covering $\mathcal{U} = \{U_i\}$ is the space of alternating $q$-cocycles $\mathcal{C}^q(U, S)$, which is the set of alternating maps $\xi$ from the set of $(q+1)$-fold indices of the form $(i_0, \ldots, i_q) \mapsto s_{i_0,\ldots,i_q} \in S(U_{i_0,\ldots,i_q})$. Here $U_{i_0,\ldots,i_q} := U_{i_0} \cap \cdots \cap U_{i_q}$. The boundary mapping $\delta : \mathcal{C}^q \to \mathcal{C}^{q+1}$ is defined by $\delta(\xi)_{i_0,\ldots i_{q+1}} = \sum_k (-1)^k s_{i_0,\ldots,i_{k-1},i_{k+1},\ldots,i_{q+1}}$. It follows that $\delta^2 = 0$, and $H^*(\mathcal{U}, S)$ is defined to be the cohomology of the associated complex.

In any consideration it is necessary to refine coverings, shrink, etc., and therefore one goes to the limit $H^*(X, \mathcal{S})$ over all refinements of the coverings. The script notation $\mathcal{S}$ is used to denote that we have then localized the sheaf to the germ level. Due to a theorem of Leray one can, however, always take a suitable covering so that $H^q(\mathcal{U}, \mathcal{S}) = H^q(X, \mathcal{S})$ for all $q$, where now $\mathcal{S}(U)$ satisfies the above axioms.

One of the important facts in this cohomology theory is that a short exact sequence of sheaves $0 \to \mathcal{S}' \to \mathcal{S} \to \mathcal{S}'' \to 0$ yields a long exact sequence

$$\begin{aligned} 0 &\to H^0(X, \mathcal{S}') \to H^0(X, \mathcal{S}) \to H^0(X, \mathcal{S}'') \\ &\to H^1(X, \mathcal{S}') \to H^1(X, \mathcal{S}) \to H^1(X, \mathcal{S}'') \to \cdots \end{aligned}$$

in cohomology.

A fundamental theorem of Stein theory, Theorem B, states that for the basic analytic sheaves $\mathcal{S}$ of complex analysis, the so-called coherent sheaves, all cohomology spaces $H^q(X, \mathcal{S})$ vanish for all $q \geq 1$. In the above example of the first Cousin problem the desired vanishing is that of $H^1(\mathcal{U}, \mathcal{O})$.

### Coherent Sheaves

Numerous important sheaves in complex analysis are associated to vector bundles on complex manifolds. A holomorphic vector bundle $\pi : E \to X$ over a complex manifold is a holomorphic surjective maximal rank fibration. Every fiber $E_x := \pi^{-1}(x)$ is a complex vector space, and the vector space structure is defined holomorphically over $X$. For example, addition is a holomorphic map $E \times_X E \to E$. Such bundles are locally trivial, that is, there is a covering $\{U_i\}$ of the base such that $\pi^{-1}(U_i)$ is isomorphic to $U_i \times \mathbb{C}^r$ and on the overlap the gluing maps in the fibers are holomorphic maps $\varphi_{ij} : U_{ij} \to \mathrm{GL}_n(\mathbb{C})$. The number $r$ is called the rank of the bundle. Holomorphic bundles of rank 1 are referred to as holomorphic line bundles. Of course all of these definitions make sense in other categories, for example, topological, smooth, real analytic, etc.

A holomorphic section of $E$ over an open set $U$ is a holomorphic map $s : U \to E$ with $\pi \circ s = \mathrm{Id}_U$. The space of these sections is denoted by $\mathcal{E}(U)$, and the map $U \to \mathcal{E}(U)$ defines a sheaf which is locally just $\mathcal{O}_X^r$. It is therefore called a locally free sheaf of

$\mathcal{O}$-modules. Conversely, by taking bases of a locally free sheaf $\mathcal{S}$ on the open sets where it is isomorphic to a direct sum $\mathcal{O}^r$, one builds an associated holomorphic vector bundle $E$ so that $\mathcal{E} = \mathcal{S}$.

It is not possible to restrict our attention to these locally free sheaves or equivalently to holomorphic vector bundles. One important reason is that images of holomorphic vector bundle maps are not necessarily vector bundles. A related reason is that the sheaf of ideals of holomorphic functions which vanish on a given analytic set $A$ is not always a vector bundle. This is caused by the presence of singularities in $A$. There are many other reasons, but these should suffice for this sketch.

The sheaves $\mathcal{S}$ that arise naturally in complex analysis are almost vector bundles. If $X$ is the base complex manifold or complex space under consideration, then $\mathcal{S}$ will come from a vector bundle on some big open subset $X_0$ whose boundary is an analytic set $X_1$, and then on the irreducible components of $X_1$ it will come from vector bundles on such big open sets, etc. These sheaves are called coherent analytic sheaves of $\mathcal{O}_X$-modules. The correct algebraic definition is that locally there exists an exact sequence

$$0 \to \mathcal{O}^{p_d} \to \cdots \to \mathcal{O}^{p_1} \to \mathcal{O}^{p_0} \to \mathcal{S} \to 0 \qquad [1]$$

of sheaves of $\mathcal{O}$-modules. This implies in particular that, although $\mathcal{S}$ might not be locally free, it is locally finitely generated, and the relations among the generators are also finitely generated.

## Selected Theorems

The following efficiently formulated fundamental theorem contains a great deal of information about Stein manifolds.

**Theorem B** *A complex space $X$ is Stein if and only if for every coherent sheaf $\mathcal{S}$ of $\mathcal{O}_X$-modules $H^q(X, \mathcal{S}) = 0$ for all $q \geq 1$.*

Since $\mathcal{S}$ is a sheaf, it follows that $H^0(X, \mathcal{S}) = \mathcal{S}(X)$. This is referred to as the space of sections of $\mathcal{S}$ over $X$. As a result of Theorem B, we are able to construct sections with prescribed properties. Let us give two concrete applications (there are many more!).

**Example** Let $A$ be a closed analytic subset of a Stein space $X$, and let $\mathcal{I}$ denote the subsheaf of $\mathcal{O}_X$ which consists of those functions which vanish on $A$. Note that this must be defined for every open subset $U$ of $X$. Then we have the short exact sequence $0 \to \mathcal{I} \to \mathcal{O}_X \to \mathcal{O}_X/\mathcal{I} \to 0$. The restriction of $\mathcal{O}_X/\mathcal{I}$ to $A$ is called the (reduced) structure sheaf $\mathcal{O}_A$ of $A$. In other words, for $U$ open in $A$ the space $\mathcal{O}_A(U)$ should be regarded as the space of holomorphic functions on $U$.

Now, $\mathcal{I}$ is a coherent sheaf on $X$ and therefore by Theorem B the cohomology group $H^1(X, \mathcal{I})$ vanishes. Consequently, the associated long exact sequence in cohomology implies that the restriction mapping $\mathcal{O}_X(X) \to \mathcal{O}_A(A)$ is surjective. This special case of Theorem A means that every (global!) holomorphic function on $A$ is the restriction of a holomorphic function on $X$. ◇

**Example** Let us consider the multiplicative (second) Cousin problem. In this case meromorphic functions $m_i$ are given on the open subsets $U_i$ of a covering $\mathcal{U}$ with the property that $m_i = f_{ij} m_j$, where $f_{ij}$ is holomorphic and nowhere vanishing on the overlap $U_{ij}$. This is a distribution $D$ of the zero and polar parts of meromorphic functions, which in complex geometry is called a divisor, and the interesting question is whether or not there exists a globally defined meromorphic function which has $D$ as its divisor.

Now we note that $\mathrm{GL}_1(\mathbb{C}) = \mathbb{C}^*$ and thus $f_{ij} : U_{ij} \to \mathbb{C}^*$ defines a line bundle $L$ on $X$ and we regard it as an element of the space $H^1(X, \mathcal{O}^*)$ of equivalence classes of line bundles on $X$. Here $\mathcal{O}^*$ is the sheaf of nowhere-vanishing holomorphic functions on $X$. It is not even a sheaf of $\mathcal{O}$-modules; therefore coherence is not discussed in this case.

The long exact sequence in cohomology associated to the short exact sequence $0 \to \mathbb{Z} \longrightarrow \mathcal{O} \xrightarrow{\exp} \mathcal{O}^* \to 1$ yields an element $c_1(L) \in H^2(X, \mathbb{Z})$, which is a purely topological invariant. It is called the Chern class of $L$, and one knows that $L$ is topologically trivial if and only if $c_1(L) = 0$.

Coming back to the Cousin II problem, using the same argument as in the Cousin I problem, we can solve it if and only if we can find nowhere-vanishing functions $f_i \in \mathcal{O}^*(U_i)$ with $f_i = f_{ij} f_j$. This is equivalent to finding a nowhere-vanishing section of $L$. But a line bundle has a nowhere-vanishing section if and only if it is isomorphic to the trivial bundle. In other words, the Cousin II problem can be solved for a given divisor $D$ if and only if the associated line bundle $L(D)$ is trivial in $H^1(X, \mathcal{O}^*)$. For this, a necessary condition is that the Chern class $c_1(L(D))$ vanishes. But if $X$ is Stein, this is also sufficient, because the vanishing of $H^1(X, \mathcal{O})$ together with the long exact sequence in cohomology shows that $H^1(X, \mathcal{O}^*) \xrightarrow{c_1} H^2(X, \mathbb{Z})$ is injective.

Hence, in this case we have the following precise formulation of the Oka principle: "A given divisor $D$ on a Stein manifold is the divisor of a globally defined meromorphic function if and only if the associated line bundle is topologically trivial." ◇

A slightly refined statement from that above is the fact that on a Stein manifold the space of topological line bundles is the same as the space of holomorphic line bundles. In the case of (higher rank) vector bundles this is a deep and important theorem of Grauert. It can be formulated as follows.

**Grauert's Oka principle** On a Stein space the map $F:\mathrm{Vect}_{\mathrm{holo}}(X) \to \mathrm{Vect}_{\mathrm{top}}(X)$ from the space of holomorphic vector bundles to the space of topological vector bundles which forgets the complex structure is bijective.

In closing this section, a few words concerning the proofs of the major theorems, for example, Theorem B, should be mentioned. In all cases one must solve something like an additive Cousin problem and one first does this on special relatively compact subsets. For this step there are at least two different ways to proceed. One is to delicately piece together solutions which are known to exist on very special polyhedral-type domains or build up from lower-dimensional pieces of such.

Another method is to solve certain systems of PDEs on relatively compact domains where control at the boundary is given by the positivity of the Levi-form. An example of how such PDEs occur can already be seen at the level of the above Cousin I problem. At the point where we have solved it topologically, that is, the holomorphic cocycle $\{f_{ij}\}$ is a coboundary $f_{ij} = f_j - f_i$ of smooth functions, we observe that since $\bar{\partial} f_{ij} = 0$, it follows that $\alpha = \bar{\partial} f_i$ is a globally defined $(0,1)$-form. It is $\bar{\partial}$-closed, that is, the compatibility condition for solving the system $\bar{\partial} u = \alpha$ is fulfilled. If this system can be solved, then we use the solution $u$ to adjust the topological solutions of the Cousin problem by replacing $f_i$ by $f_i - u$. We still have $f_{ij} = f_j - f_i$, but now the $f_i$ are holomorphic on $U_i$.

To obtain the global solution to a Cousin-type problem, one exhausts the Stein space by the special relatively compact subsets $U_n$ where, by one method or another, we have solved the problem with solutions $s_n$. One would like to say that the $s_n$ converge to a global solution $s$. However, there is no way to *a priori* guarantee this without making some sort of estimates. One main way of handling this problem is to adjust the solutions as $n \to \infty$ by an approximation procedure. For this one needs to know that holomorphic objects, for example, functions on $U_n$, can be approximated on $U_n$ by objects of the same type which are defined on the bigger set $U_{n+1}$. This Runge-type theorem, which is a nontrivial ingredient in the whole theory, requires the introduction of an appropriate Fréchet structure on the spaces of sections of a coherent sheaf. This is in itself a point that needs some attention.

## Montel's Theorem and Fredholm Mappings

If $U$ is an open subset of a complex space $X$, then $\mathcal{O}(U)$ has the Fréchet topology of convergence on compact subsets $K$ defined by the seminorms $|\cdot|_K$. Using resolutions of type (1) above, one shows that the space of sections $\mathcal{S}(U)$ of every coherent sheaf $\mathcal{S}$ also possesses a canonical Fréchet topology. This is then extended to the spaces $\mathcal{C}^q(\mathcal{U}, \mathcal{S})$, and consequently one is able to equip the cohomology spaces $H^q(X, \mathcal{S})$ with (often non-Hausdorff) quotient topology.

Elements of such cohomology groups can be regarded as obstructions to solving complex analytic problems. One often expects such obstructions, and is satisfied whenever it can be shown if there are only finitely many, that is, a finiteness theorem of the type $\dim H^q(X, \mathcal{S}) < \infty$ is desirable. Here we sketch two finiteness theorems which hold in seemingly different contexts, but their proofs are based on one principle: use the compactness guaranteed by Montel's theorem as the necessary input for the Fredholm theorem in the context of Fréchet spaces.

Recall that a continuous linear map $T: E \to F$ between topological vector spaces is said to be compact if there is an open neighborhood $U$ of $0 \in E$ such that $T(U)$ is relatively compact in $F$. If $Y$ is a relatively compact open subset of a complex space $X$, then Montel's theorem states that the restriction map $r^X_Y: \mathcal{O}(X) \to \mathcal{O}(Y)$ is compact. This can be extended to coherent sheaves, and using the Fredholm theorem for certain natural restriction and boundary maps, one proves the following fundamental fact.

**Lemma 1** *If the restriction map* $r^X_Y: H^q(X, \mathcal{S}) \to H^q(Y, \mathcal{S})$ *is surjective, then* $H^q(Y, \mathcal{S})$ *is finite dimensional.*

Since the methods for the proof are basic in complex analysis, we outline it here. Take a covering $\tilde{\mathcal{U}}$ of $X$ such that $H^q(\mathcal{U}, \mathcal{S}) = H^q(X, \mathcal{S})$. Then intersect its elements with $Y$ to obtain a covering $\tilde{\mathcal{U}}$ of $Y$. Finally, refine that covering with refinement mapping $\tau$ to a covering $\mathcal{V}$ of $Y$ such that $H^q(\mathcal{V}, \mathcal{S}) = H^q(X, \mathcal{S})$ and so that $U_i$ contains $V_{\tau(i)}$ as a relatively compact subset for all $i$. Let $Z^q(\mathcal{U}, \mathcal{S})$ denote the kernel of the boundary map $\delta$ for the covering $\mathcal{U}$, and consider the map $Z^q(\mathcal{U}, \mathcal{S}) \oplus \mathcal{C}^{q-1}(\mathcal{V}, \mathcal{S}) \to \mathcal{C}^q(\mathcal{V}, \mathcal{S})$ which is the direct sum $\tau \oplus \delta$ of the restriction and boundary maps. By assumption it is surjective. Since $\delta$ is the difference of this map and the compact map $\tau$, L Schwartz's version of the Fredholm theorem for Fréchet spaces implies that its image is of finite

codimension, that is, $H^q(Y, \mathcal{S}) = H^q(\mathcal{V}, \mathcal{S})$ is finite dimensional.

Applying this Lemma in the case of compact spaces where $X = Y$, one has the following theorem of Cartan and Serre:

**Theorem** *If $X$ is a compact complex space and $\mathcal{S}$ is a coherent sheaf on $X$, then* $\dim H^q(X, \mathcal{S}) < \infty$ *for all $q$.*

Grauert made use of this technique in solving the Levi problem for a strongly pseudoconvex relatively compact domain $D$ with smooth boundary in a complex manifold $X$. Here strongly pseudoconvex means that the restriction of the Levi form to the complex tangent space of every boundary point is positive definite. To do this he sequentially made "bumps" at boundary points to obtain a finite sequence of domains $D = D_0 \subset D_1 \subset \cdots \subset D_m$ in such a way that the restriction mappings at the level of $q$th cohomology, $q \geq 1$, are all surjective and such that at the last step $D$ is relatively compact in $D_m$. Applying the above Lemma, $\dim H^q(D, \mathcal{S}) < \infty$. Using another bumping procedure, it then follows that $D$ is holomorphically convex and, in fact, that $D$ is almost Stein.

This last statement means that one can guarantee that $\mathcal{O}(D)$ separates points outside of some compact subset which could contain compact subvarietes on which the global holomorphic functions are constant. In this situation one can apply Remmert's reduction theorem which implies that there is a canonically defined proper surjective holomorphic map $\pi : D \to Z$ to a Stein space which is biholomorphic outside of finitely many fibers. One says that, in order to obtain the Stein space $Z$, finitely many compact analytic subsets must be blown down to points.

The above mentioned reduction theorem is a general result which applies to any holomorphically convex complex space $X$. For this one observes that if $X$ is holomorphically convex, then for $x \in X$ the level set $L(x) := \{y \in X; f(y) = f(x) \text{ for all } f \in \mathcal{O}(X)\}$ is a compact analytic subset of $X$. One then defines an equivalence relation: $x \sim y$ if and only if the connected component of $L(x)$ containing $x$ and that of $L(y)$ which contains $y$ are the same. One then equips $X/\sim$ with the quotient topology and proves that the canonical quotient $\pi : X \to X/\sim =: Z$ is proper. Finally, for $U$ open in $Z$ one defines $\mathcal{O}_Z(U) = \mathcal{O}_X(\pi^{-1}(U))$ and proves that, equipped with this structure, $Z$ is a Stein space. This Remmert reduction is universal with respect to holomorphic maps to holomorphically separable complex spaces, that is, if $\varphi : X \to Y$ and $\mathcal{O}_Y(Y)$ separates the points of $Y$, then there exists a uniquely defined holomorphic map $\tau_\varphi : Z \to Y$ so that $\tau_\varphi \circ \pi = \varphi$. It should be noted that, even if the original space $X$ is a complex manifold, the associated Stein space $Z$ may be singular. This reflects the fact that it is difficult to avoid singularities in complex geometry.

## Mapping Theory

Above we have attempted to make it clear that holomorphic maps play a central role in complex geometry. It is even important to regard a holomorphic function as a map. Here we outline the basic background necessary for dealing with maps and then state three basic theorems which involve proper holomorphic mappings.

### Basic Facts

A holomorphic map $F : X \to Y$ between (reduced) complex spaces is a continuous map which can be represented locally as a holomorphic map between analytic subsets of the spaces in which $X$ and $Y$ are locally embedded. In other words, $F$ is the restriction of a map $F = (f_1, \ldots, f_m)$ which is defined by holomorphic functions.

If $X$ is irreducible and $X$ and $Y$ are one-dimensional, then a nonconstant holomorphic map $F : X \to Y$ is an open mapping. This statement is far from being true in the higher-dimensional setting. The reader need only consider the example $F : \mathbb{C}^2 \to \mathbb{C}^2, (z, w) \to (zw, z)$.

Despite the fact that holomorphic maps can be quite complicated, they have properties that in certain respects render them tenable. Let us sketch these in the case where $X$ is irreducible. First, one notes that every fiber $F^{-1}(y)$ is a closed analytic subset of $X$. One defines $\text{rank}_x F$ to be the codimension at $x$ of the fiber $F^{-1}(F(x))$ at $x$. Then $\text{rank}\, F := \max\{\text{rank}_x F; x \in X\}$. It then can be shown that $\{x \in X; \text{rank}_x F \leq k\}$ is a closed analytic subset of $X$ for every $k$. Applying this for $k = \text{rank}\, F - 1$ we see that, outside a proper closed analytic subset, $F$ has constant maximal rank.

If $F : X \to Y$ has constant rank $k$ in a neighborhood of some point $x \in X$, then one can choose neighborhoods $U$ of $x$ in $X$ and $V$ of $F(x)$ in $Y$ so that $F|U$ maps $U$ onto a closed analytic subset of $Y$. By restricting $F$ to the sets where it has lower rank and applying this local-image theorem, it follows that the local images of the set where $F$ has lower rank are at least two dimensions smaller than those of top rank. Conversely, the fiber dimension $d_F(x) := \dim_x F^{-1}(F(x))$ is semicontinuous in the sense that $d_F(x) \geq d_F(z)$ for all $z$ near $x$. Finally, we note that if $Y$ is $m$-dimensional, then $F : X \to Y$ is an open map if and only if it is of constant rank $m$.

### Proper Mappings

By definition a mapping $F: X \to Y$ between topological spaces is proper if and only if the inverse image $F^{-1}(K)$ of an arbitrary compact subset in $Y$ is compact in $X$. This is a more delicate condition than meets the eye. For example, if $F: X \to Y$ is a proper map and one removes one point from some fiber, then it is normally no longer proper! On the other hand, the restriction of a proper map to a closed subset is still proper.

Remmert's "Proper mapping theorem" is the first basic theorem on proper holomorphic maps:

**Theorem** *The image of a proper holomorphic map $F: X \to Y$ is a closed analytic subset of $Y$.*

Given another basic theorem of complex analysis, the reader can imagine how this might be proved. This is the continuation theorem for analytic sets due to Remmert and Stein:

> *If $X$ is a complex space and $Y$ is a closed analytic subset with $\dim_y Y \leq k$ for all $y \in Y$ and $Z$ is a closed analytic subset of the complement $X \backslash Y$ with $\dim_z Z \geq k+1$ at all $z \in Z$, then the topological closure $\mathrm{cl}(Z)$ of $Z$ in $X$ is a closed analytic subset of $X$ with $E = \mathrm{cl}(Z) \backslash Z = \mathrm{cl}(Z) \cap Y$ a proper analytic subset of $\mathrm{cl}(Z)$.*

Similar results hold for more general complex analytic objects. For example, closed positive currents with (locally) finite volume can be continued across any proper analytic subset (Skoda 1982). A sketch of the proof of the proper mapping theorem (for $X$ irreducible) goes as follows. From the assumption that $F$ is proper, the image $F(X)$ is closed. If $F$ has constant rank $k$, then, by the local result stated above, its image is everywhere locally a $k$-dimensional analytic set. Since the image is closed, the desired result follows. If rank $F = k$ and $E := \{x \in X;\ \mathrm{rank}_x F < k\} \neq \emptyset$, then by induction $F(E)$ is a closed analytic subset of dimension at most $k-2$. Let $A := F^{-1}(E)$ and apply the previous discussion for constant rank maps to $F|(X \backslash A): \ X \backslash A \to Y \backslash E$. The image is a closed $k$-dimensional analytic subset of $Y \backslash E$ and its Remmert–Stein extension is the full image $F(X)$.

In this framework the Stein factorization theorem is an important tool. Here $F: X \to Y$ is again a proper holomorphic map which we may now assume to be surjective. Analogous to the construction of the reduction of a holomorphically convex space, one says that two points in $X$ are equivalent if they are in the same connected component of an $F$-fiber. This is indeed an equivalence relation, and the quotient $Z := X/\sim$ is a complex space equipped with the direct image sheaf. Thus one decomposes $F$ into two maps $X \to Z \to Y$, where $X \to Z$ is a canonically associated surjective map with connected fiber, and $Z \to Y$ is a finite map.

This geometric proper mapping theorem is a preview of one of the deepest results in complex analysis: Grauert's direct image theorem. This concerns the images of sheaves, not just the images of points. For this, given a sheaf $\mathcal{S}$ on $X$ one defines the $q$th direct image sheaf on $Y$ as the sheaf associated to the presheaf which attaches to an open set $U$ in $Y$ the cohomology space $H^q(F^{-1}(U), \mathcal{S})$. Grauert's "Bildgarbensatz" states the following: "If $F: X \to Y$ is a proper holomorphic map, then all direct image sheaves of any coherent sheaf on $X$ are coherent on $Y$."

## Complex Analysis and Algebraic Geometry

The interplay between these subjects has motivated research and produced deep results on both sides. Here we indicate just a few results of the type which show that objects which are *a priori* of an analytic nature are in fact algebraic geometric.

### Projective Varieties

Let us begin with the algebraic geometric side of the picture where we consider algebraic subvarieties $X$ of projective space $\mathbb{P}_n(\mathbb{C})$. If $[z_0 : z_1 : \cdots : z_n]$ are homogeneous coordinates of $\mathbb{P}_n$, such a variety is the simultaneous zero-set, $X := V(P_1, \ldots, P_m)$, of finitely many (holomorphic) homogeneous polynomials $P_i = P_i(z_0, \ldots, z_m)$. Chow's theorem states that in this context there are no further analytic phenomena:

**Theorem** *Closed complex analytic subsets of projective space $\mathbb{P}_n(\mathbb{C})$ are algebraic subvarieties.*

This observation has numerous consequences. For example, if $F: X \to Y$ is a holomorphic map between algebraic varieties, then, by applying Chow's theorem to its graph, it follows that $F$ is algebraic.

Chow's theorem can be proved via an application of the Remmert–Stein theorem in a very simple situation. For this, let $\pi: \mathbb{C}^{n+1} \backslash \{0\} \to \mathbb{P}_n(\mathbb{C})$ be the standard projection, and let $Z := \pi^{-1}(X)$. Since $Z$ is positive dimensional, by the Remmert–Stein theorem it can be extended to an analytic subset of $\mathbb{C}^{n+1}$. The resulting subvariety $K(X)$ (the cone over $X$) is invariant by the $\mathbb{C}^*$-action which is defined by $v \to \lambda v$ for $\lambda \in \mathbb{C}^*$. If $f$ is a holomorphic function on $\mathbb{C}^{n+1}$ which vanishes on $K(X)$, then we develop it in homogeneous polynomials $f = \sum P_d$ and note that $\lambda(f)(z) = f(\lambda z) = \sum \lambda^d P_d$ also vanishes for all $\lambda$. Hence, all $P_d$ vanish identically and therefore the ideal of holomorphic functions which vanish on $K(X)$

is generated by the homogeneous polynomials which vanish on $K(X)$ and consequently finitely many of these define $X$ as a subvariety of $\mathbb{P}_n(\mathbb{C})$.

Complements of subvarieties in projective varieties occur in numerous applications and are important objects in complex geometry. Even complements $\mathbb{P}_n \backslash Y$ of subvarieties $Y$ in the full projective space are not well understood. If $Y$ is the intersection of a compact projective variety $X$ with a projective hyperplane, that is, $Y$ is a hyperplane section, then $X \backslash Y$ is affine. If $Y$ is $q$-codimensional in $X$, then $X \backslash Y$ possesses a certain degree of Levi convexity and general theorems of Andreotti and Grauert (1962) on the finiteness and vanishing of cohomology indeed apply. However, not nearly as much is understood in this case as in the case of a hyperplane section.

### Kodaira Embedding Theorem

Given that analytic subvarieties of projective space are algebraic, one would like to understand whether a given compact complex manifold or complex space can be realized as such a subvariety. Kodaira's theorem is a prototype of such an embedding theorem. Most often one formulates projective embedding theorems in the language of bundles.

For this, observe that if $L \to X$ is a holomorphic line bundle over a compact complex manifold, then its space $\Gamma(X, L)$ of holomorphic sections is a finite-dimensional vector space $V$. The zero-set of a section $s \in V$ is a one-codimensional subvariety of $X$. Let us restrict our attention to bundles which are generated by their sections which for line bundles simply means that for every $x \in X$ there is some section $s \in V$ with $s(x) \neq 0$. It then follows that for every $x \in X$ the space $H_x := \{x \in X; s(x) = 0\}$ is a one-codimensional vector subspace of $V$. Thus $L$ defines a holomorphic map $\varphi_L : X \to \mathbb{P}(V^*), x \mapsto H_x$. Note that we must go to the projective space $\mathbb{P}(V^*)$, because a linear function defining such an $H_x$ is only unique up to a complex multiple.

Projective embedding theorems state that under certain conditions on $L$ the map $\varphi_L$ is a holomorphic embedding, that is, it is injective and is everywhere of maximal rank in the analytic sense that its differential has maximal rank. Here we outline a complex analytic approach of Grauert for proving embedding theorems. It makes strong use of the complex geometry of bundle spaces.

Let $L \to X$ be a holomorphic line bundle over a compact complex manifold. A Hermitian bundle metric is a smoothly varying metric $h$ in the fibers of $L$. This defines a norm function $v \mapsto |v|^2 := h(v, v)$ on the bundle space $L$. One says that $L$ is positive if the tubular neighborhood $T := \{v \in L; |v|^3 < 1\}$ is strongly pseudoconcave, that is, when regarded from outside $T$, its boundary is strongly pseudoconvex.

To prove an embedding theorem, one must produce sections with prescribed properties. Sections of powers $L^k$ are closely related to holomorphic functions on the dual bundle space $L^*$. This is due to the fact that if $\pi : L \to X$ is the bundle projection, $\pi^{-1}(U_\alpha) \cong U_\alpha \times \mathbb{C}$ is a local trivialization, and $z_\alpha$ is a fiber coordinate, then a holomorphic function $f$ on $L^*$ has a Taylor series development

$$f(v) = \sum s_\alpha(n)(\pi(v)) z_\alpha^n(v)$$

The function $f$ is well defined on $L$. Hence, the transformation law for the $z_\alpha^n$ must be canceled out by a transformation law for the coefficient functions $s_\alpha(n)$. This implies that the $s_\alpha(n)$ are sections of $L^n$. Hence, proving the existence of sections in the powers of $L$ with prescribed properties amounts to the same thing as proving the existence of holomorphic funtions on $L^*$ with analogous properties.

The positivity assumption on $L$ is equivalent to assuming that the tubular neighborhoods of the zero-section in $L^*$ defined by the norm function associated to the dual metric are strongly pseudoconvex. The solution to the Levi problem, which was sketched above, then shows that $L^*$ is holomorphically convex, and its Remmert reduction is achieved by simply blowing down its zero-section. In other words, $L^*$ is essentially a Stein manifold, and using Stein theory, it is possible to produce enough holomorphic functions on $L$ to show that some power $L^k$ defines a holomorphic embedding $\varphi_{L^k} : X \to \mathbb{P}(\Gamma(X, L^k)^*)$. Bundles with this property are said to be ample, and thus we have outlined the following fact: "a line bundle which is Grauert-positive is ample."

It should be underlined that we defined the Chern class of $L$ as the image in $H^2(X, \mathbb{Z})$ of its equivalence class in $H^1(X, \mathcal{O}^*)$, that is, in this formulation the Chern class is a Cech cohomology class. It is, however, often more useful to consider it as a deRham class where it lies in the (1, 1)-part of $H^2_{deR}(X, \mathbb{C})$. If $h$ is a bundle metric as above, then the Levi form of the norm function is a representative $-c_1(L, h)$ of the Chern class of $L^*$. Thus $c_1(L, h)$ is an integral (1, 1)-form which represents $c_1(L)$. It is called the Chern form of $L$ associated to the metric $h$. The following is Kodaira's formulation of his embedding theorem:

**Theorem** *A line bundle $L$ is ample if and only if it possesses a metric $h$ so that $c_1(L, h)$ is positive definite.*

Kodaira's proof of this fact follows from his vanishing theorem (*see* Several Complex Variables: Compact Manifolds) in the same way the example of Theorem A was derived from Theorem B in the

first example in the subsection "Selected theorems." That an ample bundle is positive follows immediately from the fact that if $\varphi_{L^k}$ is an embedding, then its pullback of the (positive) hyperplane bundle on projective space agrees with $L^k$.

Finally, one asks the question "under what natural conditions can one construct a bundle $L$ which is positive?" The following is an example of an answer which is related to geometric quantization.

Suppose that $X$ is a compact complex manifold equipped with a symplectic structure $\omega$, that is, $\omega$ is a $d$-closed, nondegenerate 2-form. One says that $\omega$ is Kählerian if it is compatible with the complex structure $J$ in the sense that $\omega(Jv, Jw) = \omega(v, w)$ and $\omega(Jv, v) > 0$ for every $v$ and $w$ in every tangent space of $X$. Note that if $L$ is a positive line bundle, then it possesses a Hermitian metric $h$ such that $\omega = c_1(L, h)$ is a Kählerian structure on $X$.

It should be underlined that there are Kähler manifolds without positive bundles, for example, every compact complex torus $T = \mathbb{C}^n/\Gamma$ possesses the Kählerian structure which comes from the standard linear structure on $\mathbb{C}^n$. However, for $n > 1$ most such tori are not projective algebraic and therefore do not have positive bundles.

If, on the other hand, the Kählerian structure is integral, a condition that is automatic for the Chern form $c_1(L, h)$ of a bundle, then there is indeed a line bundle $L \to X$ equipped with a Hermitian metric $h$ such that $c_1(L, h) = \omega$. The condition of integrality can be formulated in terms of the integrals of $\omega$ over homology classes being integral or that its deRham class is in the image of the deRham isomorphism from the Cech cohomology $H^2(X, \mathbb{Z}) \otimes \mathbb{C}$ to $H^2_{\text{de R}}(X, \mathbb{C})$. Coupling this with the embedding theorem for positive bundles, we have the following theorem of Kodaira:

**Theorem** *If $(X, \omega)$ is Kählerian and $\omega$ is integral, then $X$ is projective algebraic.*

This result has been refined in the following important way (a conjecture of Grauert and Riemenschneider proved with different methods by Siu (1984) and by Demailly (1985)): *the same result holds if $\omega$ is only assumed to be semipositive and positive in at least one point.*

For Grauert's proof of the Kodaira embedding theorem and a number of other important and beautiful results, we recommend the original paper (Grauert 1962).

### Quotients of Bounded Domains

Let $D$ be a bounded domain in $\mathbb{C}^n$ and $\Gamma$ be a discrete subgroup of $\text{Aut}(D)$ which is acting freely on $D$ with a compact quotient $X := D/\Gamma$. For $\gamma \in \Gamma$ let $J(\gamma, z)$ be the determinant of the Jacobian $d\gamma/dz$ and, given a holomorphic function $f$, consider (at least formally) the Poincaré series $\sum f(\gamma(z))J(\gamma, z)^k$ of weight $k$. If $f$ is bounded and $k \geq 2$, then this series converges to a holomorphic function $P(f)$ on $D$ which satisfies the transformation rule $P(f)\,(\gamma(z)) = J(\gamma, z)^{-k}P(f)(z)$.

Now the differential volume form $\Omega := dz_1 \wedge \cdots \wedge dz_n$ transforms in the opposite way (for $k = 1$). Therefore $s(f) = P(f)(\Omega)^k$ is a $\Gamma$-invariant section of the $k$th power of the determinant bundle $K := \Lambda^n T^*D$ of the holomorphic cotangent bundle of $D$. In other words, $s(f) \in \Gamma(X, K^k)$. Since the choice of $f$ may be varied to show that there are sufficiently many sections to separate points and to guarantee the maximal rank condition, it follows that the canonical bundle $K$ of $X$ is ample. Compact complex manifolds with ample canonical bundle are examples of manifolds which are said to be of general type (*see* Several Complex Variables: Compact Manifolds). Thus, this construction with Poincaré series proves the following: "Every compact quotient $D/\Gamma$ of a bounded domain is of general type and is in particular projective algebraic."

*See also:* Gauge Theoretic Invariants of 4-Manifolds; Moduli Spaces: An Introduction; Riemann Surfaces; Several Complex Variables: Compact Manifolds; Twistor Theory: Some Applications [in Integrable Systems, Complex Geometry and String Theory].

## Further Reading


Andreotti A and Grauert H (1962) Théorèmes de finitude pour la cohomologie des espaces complexes. *Bulletin de la Société Mathématique de France* 90: 193–259.

Demailly J-P (1985) Champs magnfique et inégalitiés de Morse pour lat $d''$-cohomologie. *Annales de l'Institut de Fourier* 35: 189–229.

Demailly J-P, Complex analytic and algebraic geometry, http://www-fourier.ujf-grenoble.fr/demailly.

Grauert H (1962) Uber Modifikationen und exzeptionelle analytische Mengen. *Mathematische Annalen* 129: 331–368.

Grauert H and Fritzsche K (2001) *From Holomorphic Functions to Complex Manifolds*. Heidelberg: Springer.

Griffiths PhA and Harris J (1978) *Principles of Algebraic Geometry*. New York: Wiley.

Grauert H and Remmert R (1979) *Theory of Stein Spaces*. Heidelberg: Springer.

Grauert H and Remmert R (1984) *Coherent Analytic Sheaves*. Heidelberg: Springer.

Grauert H, Peternell Th, and Remmert R (1994) Several Complex Variables VII. *Encyclopedia of Mathematical Science*, vol. 74. Heidelberg: Springer.

Narasimhan R (1971) Several Complex Variables. *Chicago Lectures in Mathematics*. Chicago, IL: University of Chicago Press.

Siu YT (1984) A vanishing theorem for semi-positive line bundles over non-Kähler manifolds. *Journal of Differntial Geometry* 19: 431–452.

Skoda H (1982) Prolongement des courants positifs fermés de masse finie. *Inventiones Mathematicae* 66: 361–376.

# Several Complex Variables: Compact Manifolds

**A Huckleberry**, Ruhr-Universität Bochum, Bochum, Germany
**T Peternell**, Universität Bayreuth, Bayreuth, Germany



## Introduction

The aim of this article is to give an overview of the classification theory of compact complex manifolds. Very roughly, compact manifolds can be divided into three disjoint classes:

- Projective manifolds, that is, manifolds which can be embedded into some projective space, or manifolds birational to those, usually called Moishezon manifolds. These manifolds are treated by algebraic geometric methods, but very often transcendental methods are also indispensable.
- Compact (nonalgebraic) Kähler manifolds, that is, manifolds carrying a positive closed $(1,1)$-form, or manifolds bimeromorphic to those. This class is treated mainly by transcendental methods from complex analysis and complex differential geometry. However, some algebraic methods are also of use here.
- General compact manifolds which are not bimeromorphic to Kähler manifolds. For two reasons we will essentially ignore this class in our survey. First, because of the lack of methods, not much is known, for example, there is still no complete classification of compact complex surfaces, and it is still unknown whether or not the 6-sphere carries a complex structure. And second, for the purpose of this encyclopedia, this class seems to be less important.

The main problems of classification theory can be described as follows.

- Birational classification: describe all projective (Kähler) manifolds up to birational (bimeromorphic) equivalence; find good models in every equivalence class. This includes the study of invariants.
- Biholomorphic classification: classify all projective (Kähler) manifolds with some nice property, for example, curvature, many symmetries, etc.
- Topological classification and moduli: study all complex structures on a given topological manifold – including the study of topological invariants of complex manifolds; describe complex structures up to deformations and describe moduli spaces.
- Symmetries: describe group actions and invariants – this is deeply related with the moduli problem.

In this article we will assume familiarity with basic notions and methods from several complex variables and/or algebraic geometry. In particular we refer to Several Complex Variables: Basic Geometric Theory in this encyclopedia.

We first note some standard notation used in this article. If $X$ is a complex manifold of dimension $n$, then $T_X$ will denote its holomorphic tangent bundle and $\Omega_X^p$ the sheaf of holomorphic $p$-forms, that is, the sheaf of sections of the bundle $\bigwedge^p T_X^*$. The bundle $\bigwedge^n T_X^*$ is usually denoted by $K_X$, the canonical bundle of $X$ and its sheaf of sections is the dualizing sheaf $\omega_X$, but frequently we will not distinguish between vector bundles and their sheaves of sections. An effective (Cartier) divisor on a normal space $X$ is a finite linear combination $\sum n_i Y_i$, where $n_i > 0$ and $Y_i \subset X$ are irreducible reduced subvarieties of codimension, which are locally given by one equation. If $L$ is a line bundle, then instead of $L^{\otimes m}$ we often write $mL$. If $X$ is a compact variety and $E$ a vector bundle or coherent sheaf, then the dimension of the finite-dimensional vector space $H^q(X,E)$ will be denoted by $h^q(X,E)$.

## Birational Classification

Two compact manifolds $X$ and $Y$ are bimeromorphically equivalent, if there exist nowhere dense analytic subsets $A \subset X$ and $B \subset Y$ and a biholomorphic map $X \backslash A \to Y \backslash B$ such that the closure of the graph is an analytic set in $X \times Y$. In case $X$ and $Y$ are algebraic, one rather says that $X$ and $Y$ are birationally equivalent. This induces an isomorphism between the function fields of $X$ and $Y$. If $X$ and $Y$ are projective or Moishezon(see below), then conversely an isomorphism of their function fields induces a birational equivalence between $X$ and $Y$. Important examples are blow-ups of submanifolds; locally they can be described as follows. Suppose that locally $X$ is an open set $U \subset \mathbb{C}^n$ with coordinates $z_1, \ldots, z_n$ and that $A \subset X$ is given by $z_1 = \cdots = z_m = 0$. Then the blow-up $\hat{X} \to X$ is the submanifold $\hat{X} \subset U \times \mathbb{P}_{n-m-1}$ given by the equations

$$y_j t_i - y_i t_j = 0$$

where $t_j$ are homogeneous coordinates in $\mathbb{P}_{n-m-1}$.

The Chow lemma says that any birational – even rational – maps can be dominated by a sequence of blow-ups with smooth centers. Recently other factorizations ("weak factorization," using blow-ups and blow-downs) have been established.

A projective manifold is a compact manifold which is a submanifold of some projective space $\mathbb{P}_N$. Of course, a projective manifold can be embedded into projective spaces in many ways. According to Chow's theorem (*see* Several Complex Variables: Basic Geometric Theory), $X \subset \mathbb{P}_N$ is automatically given by polynomial equations and is therefore an algebraic variety. This is part of Serre's GAGA principle which roughly says that all global analytic objects on a projective manifold, for example, vector bundles or coherent sheaves and their cohomology are automatically algebraic. A compact manifold which is bimeromorphically equivalent to a projective manifold is called a Moishezon manifold. These arise naturally, for example, as quotient of group actions, compactifications, etc.

The most important birational invariant of compact manifolds is certainly the Kodaira dimension $\kappa(X)$. It is defined in three steps:

- $\kappa(X) = -\infty$ iff $h^0(mK_X) = 0$ for all $m \geq 1$.
- $\kappa(X) = 0$ iff $h^0(mK_X) \leq 1$ for all $m$, and $h^0(mK_X) = 1$ for some $m$.
- In all other cases we can consider the meromorphic map $f_m : X \to \mathbb{P}_{N(m)}$ associated to $H^0(mK_X)$ for all those $m$ for which $h^0(mK_X) \geq 2$. Let $V_m$ denote the (closure of the) image of $f_m$. Then $\kappa(X)$ is defined to be the maximal possible $\dim V_m$.

Recall that $f_m$ is defined by $[s_0 : \cdots : s_N]$ for a given base $s_i$ of $H^0(mK_X)$, cf. Several Complex Variables: Basic Geometric Theory.

In the same way one defines the Kodaira (or Iitaka) dimension $\kappa(L)$ of a holomorphic line bundle $L$ (instead of $L = K_X$).

We are now going to describe geometrically the different birational equivalence classes and how to single out nice models in each class. Using methods in characteristic $p$, Miyaoka and Mori proved the following theorem:

**Theorem 1** *Let X be a projective manifold and suppose that through a general point $x \in X$ there is a curve C such that $K_X \cdot C < 0$. Then X is uniruled, that is, there is a family of rational curves covering X.*

A rational curve is simply the image of nonconstant map $f : \mathbb{P}_1 \to X$. It is a simple matter to prove that uniruled manifolds have $\kappa(X) = -\infty$, but the converse is an important open problem. A step towards this conjecture has recently been made by Boucksom *et al.* (2004) if $K_X$ is not pseudoeffective, that is, $K_X$ "cannot be approximated by effective divisors," then $X$ is uniruled. Here one also finds a discussion of the case when $K_X$ is pseudoeffective.

Mori theory is central in birational geometry. To state the main results in this theory, we recall the notion of ampleness: a line bundle $L$ is ample if $L$ carries a metric of positive curvature. Alternatively some tensor power of $L$ has enough global section to separate points and tangents and there gives an embedding into some projective space; *see* Several Complex Variables: Basic Geometric Theory for more details. The notion of nefness, which is in a certain sense the degenerate version of ampleness, plays a central role in Mori theory: a line bundle or divisor $L$ is nef if

$$L \cdot C = \deg(L|C) \geq 0$$

for all curves $C \subset X$. Examples are those $L$ carrying a metric of semipositive curvature, but the converse is not true. However, if $L$ is nef, there exists for all positive $\epsilon > 0$ a metric $h_\epsilon$ with curvature $\Theta_\epsilon > -\epsilon\omega$, where $\omega$ is a fixed positive form. In this context singular metrics on $L$ are also important. Locally they are given by $e^{-\varphi}$ with a locally integrable weight function $\varphi$ and they still have a curvature current $\Theta$. If $L$ has a singular metric with $\Theta$ bounded from below as current by a Kähler form, then $L$ is big, that is, $\kappa(L) = \dim X$, the birational version of ampleness. If one simply has $\Theta \geq 0$ as current, then $L$ is pseudoeffective (and vice versa). All these positivity notions only depend on the Chern class $c_1(L)$ of $L$ and therefore one considers the ample cone

$$K_{\text{amp}} \subset (H^{1,1}(X) \cap H^2(X, \mathbb{Z})) \otimes \mathbb{R}$$

and the cone of curves

$$\overline{NE}(X) \subset (H^{n-1,n-1}(X) \cap H^{2n-2}(X, \mathbb{Z})) \otimes \mathbb{R}$$

The ample cone is by definition the closed cone of nef divisors, the interior being the ample classes, while the cone of curves is the closed cone generated by the fundamental classes of irreducible curves.

A basic result says that these cones are dual to each other. The structure of $\overline{NE}(X)$ in the part where $K_X$ is negative is very nice; one has the following cone theorem:

**Theorem 2** $\overline{NE}(X)$ *is locally finite polyhedral in the half-space $\{K_X < 0\}$; the (geometrically) extremal rays contain classes of rational curves.*

A ray $R = \mathbb{R}_+[a]$ is said to be extremal in a closed cone $K$ if the following holds: given $b, c \in K$ with $b + c \in R$, then $b, c \in R$. Given such an extremal ray $R \subset \overline{NE}(X)$, one can find an ample line bundle $H$ and a rational number $t$ such that $K_X + tH$ is nef and $K_X + tH \cdot R = 0$. Using the Kawamata–Viehweg vanishing theorem, a generalization of Kodaira's vanishing theorem, which is one of the technical corner stones of the theory, one proves the so-called

**Base point free theorem** *Some multiple of $K_X + tH$ is spanned by global sections and therefore defines a holomorphic map $f: X \to Y$ to some normal projective variety Y contracting exactly those curves whose classes belong to R.*

These maps are called "contractions of extremal rays" or "Mori contractions." In dimension 2 they are classical: either $X = \mathbb{P}_2$ and $f$ is the constant map, or $f$ is a $\mathbb{P}_1$-bundle or $f$ is birational and the contraction of a $\mathbb{P}_1$ with normal bundle $\mathcal{O}(-1)$, that is, $f$ contracts a $(-1)$-curve. In particular $Y$ is again smooth. In the first two cases $X$ has a very precise structure, but in the third birational case one proceeds by asking whether or not $K_Y$ is nef. If it is not nef, we start again by choosing the contraction of an extremal ray; if $K_Y$ is nef, then a fundamental result says that a multiple of $K_Y$ is spanned. The class of manifolds with this property will be discussed later.

The situation in higher dimensions is much more complicated. For example, $Y$ need no longer be smooth. However the singularities which appear are rather special.

**Definition 1** A normal variety $X$ is said to have only terminal singularities if first some multiple of the canonical (Weil) divisor $K_X$ is a Cartier divisor, that is, a line bundle (one says that $X$ is $\mathbb{Q}$-Gorenstein) and second if for some (hence for every) resolution of singularities $\pi: X \to X$ the following holds:

$$K_{\hat{X}} = \pi^*(K_X) + \sum a_i E_i$$

where the $E_i$ run over the irreducible $\pi$-exceptional divisors and the $a_i$ are strictly positive.

A brief remark concerning Weil divisors is in order: a Weil divisor is a finite linear combination $\sum a_i Y_i$ with $Y_i$ irreducible of codimension 1, but $Y_i$ is not necessarily locally defined by one equation. Recall that if each $Y_i$ is given locally by one equation, then the Weil divisor is Cartier. On a smooth variety these notions coincide.

One important consequence is that $\kappa(X) = \kappa(\hat{X})$ in case of terminal singularities, which is completely false for arbitrary singularities. Also notice that terminal singularities are rational: $R^q\pi_*(\mathcal{O}_{\hat{X}}) = 0$ for $q \geq 1$. Terminal singularities occur in codimension at least 3. Thus they are not present on surfaces. In dimension 3 terminal singularities are well understood. The main point in this context is that for a birational Mori contraction the image $Y$ often has terminal singularities.

Now the scheme of Mori theory is the following. Start with a projective manifold $X$. If $K_X$ is nef, we stop; this class is discussed later. If $K_X$ is not nef, then perform a Mori contraction $f: X \to Y$. There are two cases:

- If $\dim Y < \dim X$, then the general fiber $F$ is a manifold with ample $-K_F$, that is, a Fano manifold (discussed in the next section). Here we stop and observe that $\kappa(X) = -\infty$. Of course one can still investigate $Y$ and try to say more on the structure of the fibration $f$.
- If $\dim Y = \dim X$, then $Y$ has terminal singularities – unless $f$ is a small contraction which means that no divisors are contracted. Thus if $f$ is not small, we may attempt to proceed by substituting $X$ by $Y$.

As a result one must develop the entire theory for varieties with terminal singularities. The big problem arises from small contractions $f$. In that case $K_Y$ cannot be $\mathbb{Q}$-Cartier and the machinery stops. So new methods are required. At this stage, other aspects of the theory lead one to attempt a certain surgery procedure which should improve the situation and allow one to continue as above. The expected surgery $Y \rightharpoonup Y'$, which takes place in codimension at least 2, is a "flip." The idea is that we should substitute a small set, namely the exceptional set of a small contraction, by some other small set (on which the canonical bundle will be positive) to improve the situation. Of course $Y'$ should possess only terminal singularities. The existence of flips is very deep and has been proved by S Mori in dimension 3. Moreover, there cannot be an infinite sequence of flips, at least in dimension at most 4.

In summary, by performing contractions and flips one constructs from $X$ a birational model $X'$ with terminal singularities such that either

- $K_{X'}$ is nef in which case we call $X'$ a minimal model for $X$, or
- $X'$ admits a Fano fibration $f': X' \to Y'$ (discussed below), in which case $\kappa(X) = \kappa(X') = -\infty$.

Up to now, Mori theory (via the work of Kawamata, Kollár, Mori, Reid, Shokurov, and others) works well in dimension 3 (and possibly in the near future in dimension 4) but in higher dimensions there are big problems with the existence of flips. Of course there might be completely different and possibly less precise ways to construct a minimal model. One way is to consider the canonical ring $R$ of a manifold of general type:

$$R = \sum H^0(mK_X)$$

If $R$ is finitely generated as $\mathbb{C}$-algebra, then $\mathrm{Proj}(R)$ would be at least a canonical model which

has slightly more complicated singularities than a minimal model. However, it is known that this "finite generatedness problem" is equivalent to the existence of minimal models. On the other hand, if $X$ is of general type with $K_X$ nef (hence essentially ample) or more generally when some positive multiple $mK_X$ is generated by global sections, then $R$ is finitely generated.

We now must discuss the case of a nef canonical bundle. The behavior is predicted by the

**Abundance conjecture.** *If $X$ has only terminal singularities and $K_X$ is nef, then some multiple $mK_X$ is spanned.*

Up to now this conjecture is known only in dimension 3 (Kawamata, Kollár, Miyaoka). In higher dimensions it is even unknown if there is a single section in some multiple $mK_X$. If $mK_X$ is spanned, one considers the Stein factorization $f: X \to Y$ of the associated map, which is called the Iitaka fibration (if not birational) and we have $\dim Y = \kappa(X)$ by definition. The general fiber $F$ is a variety with $K_F \equiv 0$, a class discussed in the next section. If $f$ is birational, then $Y$ will be slightly singular (so-called canonical singularities) and $K_Y$ will be ample. Essentially we are in the case of negative Ricci curvature.

Everything that was outlined above holds for projective manifolds. In the Kähler case one would expect the same picture, but the methods completely fail, and new, analytic methods must be found. Only very few results are known in this context.

We come back to the case of a Fano fibration $f: X \to Y$. By definition the anticanoical bundle $-K_X$ is relatively ample so that the general fiber is a Fano variety. In this case there are no constraints on $Y$.

To see how much of the geometry of $X$ is dictated by the rational curves, one considers the so-called rational quotient of $X$. Here we identify two very general points on $X$ if they can be joined by a chain of rational curves. In that way we obtain the rational quotient

$$f: X \rightharpoonup Y$$

This map is merely meromorphic, but has the remarkable property of being "almost holomorphic," that is, the set of indeterminacies does not project onto $Y$. In other words, one has nice compact fibers not meeting the indeterminacy set. If $Y$ is just a point, then all points of $X$ can be joined by chains of rational curves and $X$ is called rationally connected. This notion is clearly birationally invariant.

A deep theorem of Graber–Harris–Starr states that, given a Fano fibration (or a fibration with rationally connected fibers) $f: X \to Y$, then $X$ is rationally connected if and only if $Y$ is.

Manifolds $X_n$ which are birational to $\mathbb{P}_n$ are called rational. If there merely exists a surjective ("dominant") rational map $\mathbb{P}_n \rightharpoonup X$, then $X$ is said to be unirational. Of course rational (resp. unirational) manifolds are rationally connected, but to decide whether a given manifold is rational/unirational is often a very deep problem. Therefore, rational connectedness is often viewed as a practical substitute for (uni)rationality.

Often it is very important to compute the Kodaira dimension of fiber spaces. Let us fix a holomorphic surjective map $f: X \to Y$ between projective manifolds and we suppose $f$ has connected fibers. Then the so-called conjecture $C_{nm}$ states that

$$\kappa(X) \geq \kappa(F) + \kappa(Y)$$

where $F$ is the general fiber of $f$. This conjecture is known in many cases, for example, when the general fiber is of general type, but it is wide open in general. It is deeply related to the existence of minimal models (Kawamata).

## Biholomorphic Classification

In this section we discuss manifolds $X$ with

- ample anticanonical bundles $-K_X$ (Fano manifolds),
- trivial canonical bundles, and
- ample canonical bundles $K_X$.

Due to the solution of the Calabi conjecture by Yau and Aubin, these classes are characterized by a Kähler metric of positive (resp. zero, resp. negative) Ricci curvature. In principle, in view of the results of Mori theory, one should rather consider varieties with terminal singularities, but we ignore this aspect completely. Philosophically, up to birational equivalence all manifolds are via fibrations somehow composed of those classes via fibrations, possibly also up to étale coverings.

Examples of Fano manifolds are hypersurfaces of degree at most $n+1$ in $\mathbb{P}_{n+1}$, Grassmannians, or more generally homogenenous varieties $G/P$ with $G$ semisimple and $P$ a parabolic subgroup. Fano manifolds are simply connected. This can be seen either by classical differential geometric methods using a Kähler metric of positive curvature or via the fundamental

**Theorem 3** *Fano manifolds are rationally connected.*

The only known proof of this fact uses, as in the uniruled criterion mentioned above, characteristic $p$ methods. By just using complex methods it is not

known how to construct a single rational curve (of course, in concrete examples the rational curves are seen immediately). One still has to observe that rationally connected manifolds are simply connected, which is not so surprising, since rational curves lift to the universal cover.

At least in principle, Fano manifolds can be classified:

**Theorem 4** *There are only finitely many families of Fano manifolds in every dimension.*

A family (of Fano manifolds) is a submersion $\pi: \mathcal{X} \to S$ (with $S$ irreducible) such that all fibers are Fano manifolds. The essential step is to bound $(-K_X)^n$. An actual classification has been carried out only in dimension up to 3; in dimension 2 one finds $\mathbb{P}_2, \mathbb{P}_1 \times \mathbb{P}_1$ and the so-called del Pezzo surfaces ($\mathbb{P}_2$ blown up in at most eight points in general position). In dimension 3 there are already 17 families of Fano 3-folds with $b_2 = 1$ and 88 families with $b_2 \geq 2$.

An extremely hard question is to decide whether a given Fano manifold is rational or unirational. Even in dimension 3 this is not completely decided.

The next class to be discussed are the manifolds with trivial canonical class $K_X$. This means that there is a holomorphic $n$-form without zeros ($n = \dim X$). Important examples are tori and hypersurface in $\mathbb{P}_{n+1}$ of degree $n+2$. Simply connected manifolds with trivial canonical bundles are further divided into irreducible Calabi–Yau manifolds and irreducible symplectic manifolds. The first class is defined by requiring that there are no holomorphic $p$-forms for $p < \dim X$ whereas the second is characterized by the existence of a holomorphic 2-form of everywhere maximal rank. A completely different characterization is by holonomy: an irreducible Calabi–Yau manifold has SU-holonomy whereas irreducible symplectic manifolds have Sp-holonomy (with respect to a suitable Kähler metric).

The splitting theorem of Beauville–Bogomolov–Kobayashi says

**Theorem 5** *Let $X$ be a projective (or compact Kähler) manifold with trivial canonical bundle. Then there exists a finite unbranched cover $\tilde{X} \to X$ such that*

$$X = A \times \Pi X_i \times \Pi Y_j$$

*with $A$ a torus, $X_i$ irreducible Calabi–Yau, and $Y_j$ irreducible symplectic.*

The key to the proof of this theorem is the existence of a Ricci-flat Kähler metric on $X$, a Kähler–Einstein metric with zero Ricci curvature. Actually one has a stronger result: instead of assuming $K_X$ to be trivial, just assume that $c_1(X) = 0$ in $H^2(X, \mathbb{R})$. Then there exists a finite unramified cover $X \to X$ such that $K_{\tilde{X}}$ is trivial. In view of Mori theory, normal projective varieties $X$ with at most terminal singularities and $K_X \equiv 0$ (i.e., $K_X \cdot C = 0$ for all curves) should also be investigated. It is expected that similar structure theorems hold; in particular $\pi_1(X)$ should be finite. The main difficulty is that there are no differential methods available; on the other hand an algebraic proof even for the splitting theorem in the smooth case is unknown.

Calabi–Yau manifolds play an important role in string theory and mirror symmetry (*see* Mirror Symmetry: A Geometric Survey). Here we mention two basic problems. The first is the problem of boundedness:

*Are there only finitely many families of Calabi–Yau manifolds in any dimension?*

This problem is wide open; in particular one might ask:

*Is the Hodge number $h^{1,2}$ bounded for Calabi–Yau 3-folds?*

The other problem asks for the existence of rational curves. In all known examples there are rational curves, but a general existence proof is not known. The case where $b_2(X) = 1$ seems to be particularly difficult. If $b_2(X) \geq 2$, then in may cases one can hope to find a fibration or a birational map, at least for 3-folds. Given such a map, the existence of rational curves is simple. For example, if $D \subset X$ is an irreducible hypersurface which is not nef, choose $H$ ample and consider the *a priori* positive real number $p$ such that $D + pH$ is on the boundary of the ample cone. Then actually $p$ is rational and a suitable multiple $m(D + pH)$ is spanned and defines a contraction on $X$. This comes from "logarithmic Mori theory."

The above splitting theorem exhibits a torus factor and all holomorphic 1-forms on $X$ come from this torus. This principle generalizes: given any projective or compact Kähler manifold $X$, there exists a "universal object," the Albanese torus

$$\mathrm{Alb}(X) = H^0(\Omega^1_X)^* / H_1(X, \mathbb{Z})$$

(which is algebraic if $X$ is) together with a holomorphic map

$$\alpha : X \to \mathrm{Alb}(X)$$

the Albanese map. This Albanese map is given by integrating 1-forms and is often far from being surjective. The important property is now that, given a holomorphic 1-form $\omega$ on $X$, there exists a holomorphic 1-form $\eta$ on the Albanese torus such that $\omega = \alpha^*(\eta)$. The universal property reads as

follows: every map $X \to T$ to a torus factors via an affine map $\mathrm{Alb}(X) \to T$.

There is a nonabelian analog, the so-called Shafarevich map, but at the moment this map is only known to be meromorphic. It is an important tool to study the fundamental group $\pi_1(X)$. We refer to Campana (1996) and Kollár (1995).

In the following, Chern classes of holomorphic vector bundles will be important. Let $X$ be a compact complex manifold and $E$ a holomorphic vector bundle on $X$. The $j$th Chern class of $E$ is an element

$$c_j(E) \in H^{2j}(X, \mathbb{Q}) \cap H^{j,j}(X)$$

It can be defined, for example, by putting a Hermitian metric on $E$, computing the curvature of the canonical connection compatible with both the metric and the holomorphic structure and then by applying certain linear operators coming from symmetric functions such as determinant and trace. Actually Chern classes can be attached to every complex topological vector bundle on a topological manifold; then $c_j(E)$ will simply live in $H^{2j}(X, \mathbb{R})$. There is also a purely algebraic construction by Grothendieck. We refer, for example, to Fulton (1984) as well as for a discussion of the elementary functorial properties of Chern classes. Here we just recall that for a rank-$r$ vector bundle $E$ the first Chern class

$$c_1(E) = c_1\left(\bigwedge E\right)$$

where the Chern class of the line bundle $\bigwedge^r E$ as given in Several Complex Variables: Basic Geometric Theory actually lives in $H^2(X, \mathbb{Z})$.

Finally we discuss manifolds with ample canonical class $K_X$. Here moduli question often plays a central role. Moduli spaces of surfaces with fixed $c_1^2$ and $c_2$ are very intensively studied (by Catanese, Ciliberto, and others). Here, without going into details, we will concentrate on the very interesting topic of Kähler–Einstein metrics.

A Kähler metric $\omega$ is said to be Kähler–Einstein, if its Ricci curvature $\mathrm{Ric}(\omega)$ is proportional to $\omega$. The proportionality factor $\lambda$ can be taken to be $-1, 0, 1$. In case $K_X$ is ample or trivial, Kähler–Einstein metrics always exist by Yau and Aubin (cases $\lambda = -1$, resp. $\lambda = 0$). However if $X$ is Fano, there are obstructions, and a Kähler–Einstein metric does not always exist. An important consequence of the existence of a Kähler–Einstein metric on a manifold $X_n$ with ample canonical class is the Miyaoka–Yau inequality:

$$c_1^2 \omega^{n-2} \leq \frac{2n+1}{n} \omega^{n-2}$$

In case of equality, $X$ is covered by the $n$-dimensional unit ball.

The same inequality holds in case $K_X = 0$, and as a consequence the Chern class $c_2(X)$ is in some sense semipositive. If $c_2(X) = 0$, then some finite unramified cover of $X$ is a torus.

There is an interesting relation to stability. Recall that a vector bundle $E$ on a compact Kähler manifold $X_n$ is semistable with respect to a given Kähler form $\omega$, if for all proper coherent subsheaves $\mathcal{F} \subset E$ of rank-$r$ the following inequality holds:

$$\frac{c_1(\mathcal{F}) \cdot \omega^{n-1}}{r} \leq \frac{c_1(E) \cdot \omega^{n-1}}{n}$$

In case of strict inequality, $E$ is said to be stable.

The basic observation is now that the tangent bundle of a manifold with a Kähler–Einstein metric is semistable (with respect to the Kähler–Einstein metric). It is expected that Fano manifolds with $b_2 = 1$ have (semi?-)stable tangent bundles, although in certain situations they do not admit a Kähler–Einstein metric.

Again the first two Chern classes of a semistable vector bundle fulfill an inequality:

$$c_1^2(E) \cdot \omega^{n-2} \leq \frac{2r}{r-1} c_2(E) \cdot \omega^{n-2}$$

Equally important, semistable bundles with fixed numerical data form moduli spaces, this being the origin of the stability notion (Mumford). In this context, the notion of an Hermite–Einstein bundle is also important. Given a holomorphic vector bundle $E$ with a Hermitian metric $h$, there is a unique connection $F_h$ on $E$ compatible both with $h$ and the complex structure. $F_h$ is a (1,1)-form with values in $\mathrm{End}(E)$. Now suppose $(X, \omega)$ is Kähler and let $\lambda F_h$ be the contraction of $F_h$ with $\omega$. Then $(E, h)$ is said to Hermite–Einstein on $(X, \omega)$, if

$$\lambda F_h = \gamma \mathrm{id}$$

with some constant $\gamma$ and id: $E \to E$ the identity. Notice that $(X, \omega)$ is Kähler–Einstein if $(T_X, h)$ is Hermite–Einstein over $(X, \omega)$ with $h$ the Kähler metric with Kähler form $\omega$. It is not so difficult to see that Hermite–Einstein bundles are semistable (with respect to the underlying Kähler form) and actually are directs sum of stable Hermite–Einstein bundles. Conversely, a very deep theorem of Uhlenbeck–Yau says that *every stable vector bundle on a compact Kähler manifold is Hermite–Einstein.* This is known as the Kobayashi–Hitchin correspondence; see Lübke and Teleman (1995).

## Topology, Invariants and Cohomology

Besides the Kodaira dimension there are other important invariants of compact complex manifolds. Of course there are topological invariants such as the Betti number $b_i(X) = \dim H^i(X, \mathbb{R})$ or the fundamental group $\pi_1(X)$. The fundamental group has been studied intensively in the last decade. A central question asks which groups can occur as fundamental groups of compact Kähler manifolds; another problem is the so-called Shafarevitch conjecture which says that the universal cover of a compact Kähler manifold should be holomorphically convex. We refer to Campana (1996) and Kollár (1995).

The plurigenera,

$$P_m(X) = \dim h^0(mK_X)$$

are also extremely important. Here, Siu recently proved that $P_m(X)$ is constant in families of projective manifolds. Other important invariants are $h^0(X, (\Omega^1_X)^{\otimes m})$. For example, it is conjectured that if

$$h^0(X, (\Omega^1_X)^{\otimes m}) = 0$$

for all positive $m$, then $X$ is rationally connected. Tensor powers of the cotangent bundle somehow capture more of the structure of $X$ than the Kodaria dimension but they are more difficult to treat. The relevance of the dimensions

$$h^0(X, \Omega^p_X)$$

of holomorphic forms is easier to understand. More generally one has the Hodge numbers

$$h^{p,q}(X) = \dim H^q(X, \Omega^p_X)$$

For compact Kähler manifolds, the Hodge decomposition states

$$H^r(X, \mathbb{C}) = \bigoplus_{p+q=r} H^{p,q}(X)$$

Furthermore, Hodge duality,

$$H^{p,q}(X) = H^{q,p}(X)$$

holds. These results form a cornerstone for the geometry of compact Kähler manifolds and the starting point of Hodge theory. Hodge theory is, for example, extremely important in the study of families of manifolds and moduli.

Concerning the topology of projective (Kähler) manifolds, the following two questions are very basic.

- Which invariants are topological (or diffeomorphic) invariants?
- What are the projective or Kähler structures on a given compact topological manifold?

Concerning the first, Hodge decomposition implies that the irregularity $h^0(\Omega^1_X)$ is actually a topological invariant. However it is unknown whether the number of holomorphic 2-forms is a topological invariant of Kähler 3-folds. Both questions have been intensively studied in dimension 2. However, in higher dimensions almost nothing is known. For example, it is not known whether there is projective manifold of general type of even dimension which is homeomorphic to a quadric, that is, a hypersurface of degree 2 in projective space.

Other important tools in the study of projective/Kähler manifolds are listed below.

- Cohomological methods: Riemann–Roch theorem and holomorphic Morse inequalities; vanishing theorems (Kodaira, Kawamata–Viehweg, etc.); Serre duality. References: Demailly (2000), Demailly and Lazarsfeld, Fulton (1984), Grauert *et al.* (1994), Lazarsfeld (2004).
- $L^2$ methods: extension theorems, singular metrics, multiplier ideals, etc. Reference: Demailly and Lazarsfeld (2001), Lazarsfeld (2004).
- Theory of currents. Reference: Demailly 2000.
- Cycle space and Douady space, resp. Chow scheme and Hilbert scheme. Reference: Fulton 1984, Grauert 1994, Kollár 1996.

We restrict our remarks on just one of these topics, vanishing theorems. The classical Kodaira–Nakano vanishing theorem says that if $X$ is a compact manifold of dimension $n$ with a positive (ample) line bundle $L$, then

$$H^q(X, L \otimes \Omega^p) = 0$$

for $p + q > n$. This is usually proved via harmonic theory, that is, by representing the cohomology space by harmonic $(p, q)$-forms with values in $L$ and by computing integrals of these forms. For many purposes, for example, for Mori theory, it is important to generalize this to a line bundle which have some positivity properties but which are not ample. This works only for $p = n$, however this is the most important part of the Kodaira–Nakano vanishing. The Kawamata–Viehweg vanishing theorem in its most basic version says that given a nef and big line bundle $L$, then Kodaira vanishing still holds:

$$H^q(X, L \otimes K_X) = 0$$

for $q \geq 1$. But actually it is not necessary to assume $L$ nef, in fact the following is true. Let

$$D = \sum a_i D_i$$

be an effective $\mathbb{Q}$-divisor, that is, all $a_i$ are positive rational numbers. Let $\langle a_i \rangle$ be the fractional part of $a_i$ and suppose that the $\mathbb{Q}$-divisor $\sum \langle a_i \rangle D_i$ has normal crossings. Let $\lceil a_i \rceil$ be the roundup of $a_i$ and put $L = \sum \lceil a_i \rceil D_i$. If $D$ is big and nef, then

$$H^q(X, L \otimes K_X) = 0$$

for $q \geq 1$. Of course $L$ itself need not be nef! This generalization is technically very important and yields substantial freedom for birational manipulations. We refer to Kawamata *et al.* (1987) and Lazarsfeld (2004). Even this is not the end of the story: the Kawamata–Viehweg theorem is embedded in the broader context of the Nadel vanishing theorem where multiplier ideal sheaves come into the play. See Demailly and Lazarsfeld and Lazarsfeld (2004).

## Homogeneous Manifolds

In this section we consider vector fields and holomorphic group actions on compact (Kähler) manifolds. Our main reference is Huckleberry (1990) with further literature given there.

We denote by $\mathrm{Aut}(X)$ the group of holomorphic automorphisms of the compact manifold $X$ (well known to be a complex Lie group), and by $G := \mathrm{Aut}^0(X)$ the connected component containing the identity. The tangent space at any point of $\mathrm{Aut}^0(X)$ can naturally be identified with $H^0(X, T_X)$, the (finite-dimensional) space of holomorphic vector fields on $X$. In fact, by integration, a vector field determines a one-parameter group of automorphisms.

One says that $X$ is homogeneous if $G$ acts transitively on $X$. Therefore, one can write

$$X = G/H$$

where $H$ is the isotropy subgroup of any point $x_0 \in X$, that is, the subgroup of automorphisms fixed $x_0$. Conversely one can take a complex Lie group $G$ and a closed subgroup $H$ and form the quotient $G/H$ which is again a complex manifold and in fact homogeneous (of course not necessarily compact).

Going back to a compact manifold $X$, the condition to be homogenenous can be rephrased by saying that the tangent bundle is generated by global sections, that is, if $x \in X$ and $e \in T_{X,x}$, then there exists $v \in H^0(X, T_X)$ such that $v(x) = e$. The easiest case is when $T_X$ is trivial. If $X$ is Kähler, this is exactly the case when $X$ is torus, $X = \mathbb{C}^n/\Gamma$ with $\Gamma \simeq \mathbb{Z}^{2n}$ a lattice, but without the Kähler assumption there are many more examples (the so-called parallelizable manifolds).

More generally, let us consider the case that the compact Kähler manifold $X$ admits a vector field $v$ without zeros, but $X$ is not required to be homogeneous. Then a theorem of Lieberman says that there is a finite unramified cover $f: X \to X$ and a splitting

$$\tilde{X} \simeq F \times T$$

with $T$ a torus, such that $f^*(v)$ is the pullback of a vector field on $T$. On the other hand, if $v$ has a zero, then a classical theorem of Rosenlicht says that $X$ is covered by rational curves, that is, $X$ is uniruled. In particular $\kappa(X) = -\infty$. Notice also that a manifold of general type can never carry a vector field, in other words, the automorphism group is discrete, even finite.

Coming back to compact homogeneous Kähler manifolds, the first thing to study is the Albanese map. The Borel–Remmert theorem says that

$$X \simeq T \times Q$$

where $T$ is the Albanese torus. This is proved using a maximal compact subgroup $K \subset G$ and by some averaging process over $K$. Moreover, $Q$ is a rational homogeneous manifold. The structure of $Q$ is more precisely the following. One can write $Q = S/P$ with $S$ a semisimple Lie group and $P \subset S$ parabolic, which means that $P$ contains a maximal connected solvable subgroup (the so-called Borel subgroup). The main ingredients of the proof are the Tits fibration, the Levi–Malcev decomposition of a Lie group into its radical and a semisimple group, and the Borel fixed point theorem:

**Theorem 6** *Let $G \subset \mathrm{GL}_n(\mathbb{C})$ be a connected solvable subgroup and $X \subset \mathbb{P}_{n-1}$ be a $G$-stable subvariety. Then $G$ has a fixed point on $X$.*

In the homogenenous Kähler case, the rationality of $Q$ is seen by exhibiting an open subset in $Q$ which is algebraically isomorphic to $\mathbb{C}^n$.

Now things come down to classify all rational homogenenous manifold $S/P$ which is of course classical. Notice that all rational homogeneous manifolds are Fano. One knows that a rational homogeneous manifold with Betti number $b_2 \geq 2$ can be fibered over another rational homogenenous manifold with fibers rational homogeneous – this is actually a fiber bundle. The case that $b_2 = 1$ can be rephrased by saying that $P$ is maximal parabolic. This fiber bundle might not be trivial as shown by the projectivized tangent bundle $\mathbb{P}(T_{\mathbb{P}_n})$.

Compact Hermitian symmetric spaces form a particularly interesting subclass of homogeneous Kähler manifolds. A manifold equipped with a

Hermitian metric is called Hermitian symmetric, if for every $x \in X$ there exists an involutive holomorphic isometry fixing $x$. Mok has shown the remarkable fact that the simply connected compact Hermitian symmetric spaces are exactly those simply connected compact manifolds carrying a Kähler metric with semipositive holomorphic bisectional curvature. The only manifold having a metric with positive holomorphic bisectional curvature is $\mathbb{P}_n$ (Siu-Yau, Mori).

*See also:* Classical Groups and Homogeneous Spaces; Einstein Manifolds; Mirror Symmetry: A Geometric Survey; Moduli Spaces: An Introduction; Riemann Surfaces; Several Complex Variables: Basic Geometric Theory; Topological Sigma Models; Twistor Theory: Some Applications [in Integrable Systems, Complex Geometry and String Theory].

## Further Reading


Beltrametti M and Sommese AJ (1995) *The Adjunction Theory of Complex Projective Varieties*. Berlin: de Gruyter.

Boucksom S, Demailly JP, Paun M, and Peternell T (2004) The pseudo-effective cone of a compact Kähler manifold and varieties of negative Kodaira dimension math.AG/0405285

Campana F (1996) Kodaira dimension and fundamental group of compact Kähler manifolds. In: Andreatta M and Peternell T (eds.) *Higher Dimensional Complex Varieties*, pp. 89–162. Berlin: de Gruyter.

Demailly JP (2000) Complex analytic and algebraic geometry, http://www-fourier.ujf-grenoble.fr/ demailly.

Demailly JP and Lazarsfeld R (eds.) Vanishing Theorems and Effective Results in Algebraic Geometry, ICTP Lecture Notes, vol. 6, Trieste.

Fulton W (1984) *Intersection Theory*. Heidelberg: Springer.

Griffiths Ph and Harris J (1978) *Priniciples of Algebraic Geometry*. New York: Wiley.

Grauert H, Peternell Th, and Remmert R (1994) Several Complex Variables VII. *Encyclopedia of Mathematical Sciences*, vol. 74. Heidelberg: Springer.

Huckleberry A (1990) Actions of groups of holomorphic transformations. In: Barth W and Narasimhan R (eds.) *Several Complex Variables VI*, Encyclopedia of Mathematical Science, vol. 69, pp. 143–196. Berlin: Springer.

Kawamata Y, Matsuda K, and Matsuki K (1987) Introduction to the minimal model problem. *Advance Studies in Pure Mathematics* 10: 283–360.

Kollár J (1995) *Shafarevitch Maps and Automorphic Forms*. Princeton University Press.

Kollár J (1996) Rational Curves on Algebraic Varieties. *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 32. Heidelberg: Springer.

Lazarsfeld R (2004) Positivity in Algebraic Geometry I, II. *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 48 and 49. Heidelberg: Springer.

Lübke M and Teleman A (1995) *The Kobayashi–Hitchin Correspondence*. Singapore: World Scientific.

Matsuki K (2002) *Introduction to the Mori Program*, Universitext. Heidelberg: Springer.

Mori S (1987) Classification of higher-dimensional varieties. *Proceedings of Symposia in Pure Mathematics* 46: 269–331.

Parshin AN and Shafarevich IR (eds.) (1999) *Algebraic Geometry V – Fano Varieties*, vol. 47, *Encyclopedia of Mathematical Sciences*. Heidelberg: Springer.

Siu YT (1987) Lectures on Hermitian–Einstein Metrics for Stable Bundles and Kähler–Einstein Metrics. *DMV Seminar*, vol. 8. Basel: Birkhäuser.

Ueno K (1975) Classification Theory of Compact Complex spaces. *Lecture Notes in Math.*, vol. 439. Heidelberg: Springer.

Viehweg E (1995) Quasi-Projective Moduli for Polarozed Manifolds. *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 30. Heidelberg: Springer.


# Shock Wave Refinement of the Friedman–Robertson–Walker Metric


**B Temple**, University of California at Davis, Davis, CA, USA

**J A Smoller**, University of Michigan, Ann Arbor, MI, USA




## Introduction

In the standard model of cosmology, the expanding universe of galaxies is described by a Friedman–Robertson–Walker (FRW) metric, which in spherical coordinates has a line element given by (Blau and Guth 1987, Weinberg 1972)

$$ds^2 = -dt^2 + R^2(t)\left\{\frac{dr^2}{1-kr^2} + r^2[d\theta^2 + \sin^2\theta\, d\phi^2]\right\} \quad [1]$$

In this model, which accounts for things on the largest length scale, the universe is approximated by a space of uniform density and pressure at each fixed time, and the expansion rate is determined by the cosmological scale factor $R(t)$ that evolves according to the Einstein equations. Astronomical observations show that the galaxies are uniform on a scale of about one billion light years, and the expansion is critical – that is, $k=0$ in [1] – and so, according to [1], on the largest scale, the universe is infinite flat Euclidian space $R^3$ at each fixed time. Matching the Hubble constant to its observed values, and invoking the Einstein equations, the FRW model implies that the entire infinite universe $R^3$ emerged all at once from a singularity ($R=0$), some 14 billion years ago, and this event is referred to as the big bang.

In this article, which summarizes the work of the authors in Smoller and Temple (1995, 2003), we describe a two-parameter family of exact solutions of the Einstein equations that refine the FRW metric by a spherical shock wave cutoff. In these exact solutions, the expanding FRW metric is reduced to a region of finite extent and finite total mass at each fixed time, and this FRW region is bounded by an entropy-satisfying shock wave that emerges from the origin (the center of the explosion), at the instant of the big bang, $t=0$. The shock wave, which marks the leading edge of the FRW expansion, propagates outward into a larger ambient spacetime from time $t=0$ onward. Thus, in this refinement of the FRW metric, the big bang that set the galaxies in motion is an explosion of finite mass that looks more like a classical shock wave explosion than does the big bang of the standard model. (The fact that the entire infinite space $R^3$ emerges at the instant of the big bang, is, loosely speaking, a consequence of the Copernican principle, the principle that the Earth is not in a special place in the universe on the largest scale of things. With a shock wave present, the Copernican principle is violated, in the sense that the Earth then has a special position relative to the shock wave. But, of course, in these shock wave refinements of the FRW metric, there is a spacetime on the other side of the shock wave, beyond the galaxies, and so the scale of uniformity of the FRW metric, the scale on which the density of the galaxies is uniform, is no longer the largest length scale.)

In order to construct a mathematically simple family of shock wave refinements of the FRW metric that meet the Einstein equations exactly, we assume $k=0$ (critical expansion), and we restrict to the case that the sound speed in the fluid on the FRW side of the shock wave is constant. That is, we assume an FRW equation of state $p=\sigma\rho$, where $\sigma$, the square of the sound speed $\sqrt{\partial p/\partial\rho}$, is constant, $0<\sigma\leq c^2$. At $\sigma=c^2/3$, this catches the important equation of state $p=(c^2/3)\rho$ which is correct at the earliest stage of big bang physics (Weinberg 1972). Also, as $\sigma$ ranges from 0 to $c^2$, we obtain qualitatively correct approximations to general equations of state. Taking $c=1$ (we use the convention that $c=1$, and Newton's constant $\mathcal{G}=1$ when convenient), the family of solutions is then determined by two parameters, $0<\sigma\leq 1$ and $r_*\geq 0$. The second parameter, $r_*$, is the FRW radial coordinate $r$ of the shock in the limit $t\to 0$, the instant of the big bang. (Since, when $k=0$, the FRW metric is invariant under the rescaling $r\to\alpha r$ and $R\to\alpha^{-1}R$, we fix the radial coordinate $r$ by fixing the scale factor $\alpha$ with the condition that $R(t_0)=1$ for some time $t_0$, say present time.) The FRW radial coordinate $r$ is singular with respect to radial arclength $\bar{r}=rR$ at the big bang $R=0$, so setting $r_*>0$ does not place the shock wave away from the origin at time $t=0$. The distance from the FRW center to the shock wave tends to zero in the limit $t\to 0$ even when $r_*>0$. In the limit $r_*\to\infty$, we recover from the family of solutions the usual (infinite) FRW metric with equation of state $p=\sigma\rho$ – that is, we recover the standard FRW metric in the limit that the shock wave is infinitely far out. In this sense our family of exact solutions of the Einstein equations considered here represents a two-parameter refinement of the standard FRW metric.

The exact solutions for the case $r_*=0$ were first constructed in Smoller and Temple (1995) (see also the notes by Smoller and Temple (1999)), and are qualitatively different from the solutions when $r_*>0$, which were constructed later in Smoller and Temple (2003). The difference is that, when $r_*=0$, the shock wave lies closer than one Hubble length from the center of the FRW spacetime throughout its motion (Smoller and Temple 2000), but when $r_*>0$, the shock wave emerges at the big bang at a distance beyond one Hubble length. (The Hubble length depends on time, and tends to zero as $t\to 0$.) We show in Smoller and Temple (2003) that one Hubble length, equal to $c/H$, where $H=\dot{R}/R$, is a critical length scale in a $k=0$ FRW metric because the total mass inside one Hubble length has a Schwarzschild radius equal exactly to one Hubble length. (Since $c/H$ is a good estimate for the age of the universe, it follows that the Hubble length $c/H$ is approximately the distance of light travel starting at the big bang up until the present time. In this sense, the Hubble length is a rough estimate for the distance to the further most objects visible in the universe.) That is, one Hubble length marks precisely the distance at which the Schwarzschild radius $\bar{r}_s\equiv 2M$ of the mass $M$ inside a radial shock wave at distance $\bar{r}$ from the FRW center, crosses from inside ($\bar{r}_s<\bar{r}$) to outside ($\bar{r}_s>\bar{r}$) the shock wave. If the shock wave is at a distance closer than one Hubble length from the FRW center, then $2M<\bar{r}$ and we say that the solution lies outside the black hole, but if the shock wave is at a distance greater than one Hubble length, then $2M>\bar{r}$ at the shock, and we say that the solution lies "inside" the black hole. Since $M$ increases like $\bar{r}^3$, it follows that $2M<\bar{r}$ for $\bar{r}$ sufficiently small, and $2M>\bar{r}$ for $\bar{r}$ sufficiently large, so there must be a critical radius at which $2M=\bar{r}$, and we show in what follows (see also Smoller and Temple (2003)) that when $k=0$, this critical radius is exactly the Hubble length. When the parameter $r_*=0$, the family of solutions for $0<\sigma\leq 1$ starts at the big bang, and evolves thereafter

"outside" the black hole, satisfying $2M/\bar{r} < 1$ everywhere from $t = 0$ onward. But, when $r_* > 0$, the shock wave is further out than one Hubble length at the instant of the big bang, and the solution begins with $2M/\bar{r} > 1$ at the shock wave. From this time onward, the spacetime expands until eventually the Hubble length catches up to the shock wave at $2M/\bar{r} = 1$, and then passes the shock wave, making $2M/\bar{r} < 1$ thereafter. Thus, when $r_* > 0$, the whole spacetime begins inside the black hole (with $2M/\bar{r} > 1$ for sufficiently large $\bar{r}$), but eventually evolves to a solution outside the black hole. The time when $\bar{r} = 2M$ actually marks the event horizon of a white hole (the time reversal of a black hole) in the ambient spacetime beyond the shock wave. We show that, when $r_* > 0$, the time when the Hubble length catches up to the shock wave comes after the time when the shock wave comes into view at the FRW center, and when $2M = \bar{r}$ (assuming $t$ is so large that we can neglect the pressure from this time onward), the whole solution emerges from the white hole as a finite ball of mass expanding into empty space, satisfying $2M/\bar{r} < 1$ everywhere thereafter. In fact, when $r_* > 0$, the zero pressure Oppenheimer–Snyder solution outside the black hole gives the large-time asymptotics of the solution (Oppenheimer and Snyder 1939, Smoller and Temple 1988, 2004 and the comments after Theorems 6–8 below).

The exact solutions in the case $r_* = 0$ give a general-relativistic version of an explosion into a static, singular, isothermal sphere of gas, qualitatively similar to the corresponding classical explosion outside the black hole (Smoller and Temple 1995). The main difference physically between the cases $r_* > 0$ and $r_* = 0$ is that, when $r_* > 0$ (the case when the shock wave emerges from the big bang at a distance beyond one Hubble length), a large region of uniform expansion is created behind the shock wave at the instant of the big bang. Thus, when $r_* > 0$, lightlike information about the shock wave propagates inward from the wave, rather than outward from the center, as is the case when $r_* = 0$ and the shock lies inside one Hubble length. (One can imagine that when $r_* > 0$, the shock wave can get out through a great deal of matter early on when everything is dense and compressed, and still not violate the speed of light bound. Thus, when $r_* > 0$, the shock wave "thermalizes," or more accurately "makes uniform," a large region at the center, early on in the explosion.) It follows that, when $r_* > 0$, an observer positioned in the FRW spacetime inside the shock wave will see exactly what the standard model of cosmology predicts, up until the time when the shock wave comes into view in the far field. In this sense, the case $r_* > 0$ gives a black hole cosmology that refines the standard FRW model of cosmology to the case of finite mass. One of the surprising differences between the case $r_* = 0$ and the case $r_* > 0$ is that, when $r_* > 0$, the important equation of state $p = \rho/3$ comes out of the analysis as special at the big bang. When $r_* > 0$, the shock wave emerges at the instant of the big bang at a finite nonzero speed (the speed of light) only for the special value $\sigma = 1/3$. In this case, the equation of state on both sides of the shock wave tends to the correct relation $p = \rho/3$ as $t \to 0$, and the shock wave decelerates to subluminous speed for all positive times thereafter (see Smoller and Temple (2003) and Theorem 8 below).

In all cases $0 < \sigma \leq 1$, $r_* \geq 0$, the spacetime metric that lies beyond the shock wave is taken to be a metric of Tolmann–Oppenheimer–Volkoff (TOV) form (Oppenheimar and Volkoff 1939):

$$ds^2 = -B(\bar{r})d\bar{t}^2 + A^{-1}(\bar{r})d\bar{r}^2 + \bar{r}^2[d\theta^2 + \sin^2\theta\, d\phi^2] \quad [2]$$

The metric [2] is in standard Schwarzschild coordinates (diagonal with radial coordinate equal to the area of the spheres of symmetry), and the metric components depend only on the radial coordinate $\bar{r}$. Barred coordinates are used to distinguish TOV coordinates from unbarred FRW coordinates for shock matching. The mass function $M(\bar{r})$ enters as a metric component through the relation

$$A = 1 - \frac{2M(\bar{r})}{\bar{r}} \quad [3]$$

The TOV metric [2] has a very different character depending on whether $A > 0$ or $A < 0$; that is, depending on whether the solution lies outside the black hole or inside the black hole. In the case $A > 0$, $\bar{r}$ is a spacelike coordinate, and the TOV metric describes a static fluid sphere in general relativity. (When $A > 0$, for example, the metric [2] is the starting point for the stability limits of Buchdahl and Chandresekhar for stars (Weinberg 1972, Smoller and Temple 1997, 1998).) When $A < 0$, $\bar{r}$ is the timelike coordinate, and [2] is a dynamical metric that evolves in time. The exact shock wave solutions are obtained by taking $\bar{r} = R(t)r$ to match the spheres of symmetry, and then matching the metrics [1] and [2] at an interface $\bar{r} = \bar{r}(t)$ across which the metrics are Lipschitz continuous. This can be done in general. In order for the interface to be a physically meaningful shock surface, we use the result in Theorem 4 below (see Smoller and Temple (1994)) that a single additional conservation constraint is sufficient to rule out $\delta$-function sources at the shock (the Einstein equations $G = \kappa T$ are second order in the metric, and

so $\delta$-function sources will in general be present at a Lipschitz continuous matching of metrics), and guarantee that the matched metric solves the Einstein equations in the weak sense. The Lipschitz matching of the metrics, together with the conservation constraint, leads to a system of ordinary differential equations (ODEs) that determine the shock position, together with the TOV density and pressure at the shock. Since the TOV metric depends only on $\bar{r}$, the equations thus determine the TOV spacetime beyond the shock wave. To obtain a physically meaningful outgoing shock wave, we impose the constriant $\bar{p} \leq \bar{\rho}$ to ensure that the equation of state on the TOV side of the shock is physically reasonable, and as the entropy condition we impose the condition that the shock be compressive. For an outgoing shock wave, this is the condition $\rho > \bar{\rho}$, $p > \bar{p}$, that the pressure and density be larger on the side of the shock that receives the mass flux – the FRW side when the shock wave is propagating away from the FRW center. This condition breaks the time-reversal symmetry of the equations, and is sufficient to rule out rarefaction shocks in classical gas dynamics (Smoller 1983, Smoller and Temple 2003). The ODEs, together with the equation-of-state bound and the conservation and entropy constraints, determine a unique solution of the ODEs for every $0 < \sigma \leq 1$ and $\bar{r}_* \geq 0$, and this provides the two-parameter family of solutions discussed here (Smoller and Temple 1995, 2003). The Lipschitz matching of the metrics implies that the total mass $M$ is continuous across the interface, and so when $r_* > 0$, the total mass of the entire solution, inside and outside the shock wave, is finite at each time $t > 0$, and both the FRW and TOV spacetimes emerge at the big bang. The total mass $M$ on the FRW side of the shock has the meaning of total mass inside the radius $\bar{r}$ at fixed time, but on the TOV side of the shock, $M$ does not evolve according to equations that give it the interpretation as a total mass because the metric is inside the black hole. Nevertheless, after the spacetime emerges from the black hole, the total mass takes on its usual meaning outside the black hole, and time asymptotically the big bang ends with an expansion of finite total mass in the usual sense. Thus, when $r_* > 0$, our shock wave refinement of the FRW metric leads to a big bang of finite total mass.

A final comment is in order regarding our overall philosophy. The family of exact shock wave solutions described here are rough models in the sense that the equation of state on the FRW side satisfies the condition $\sigma = \text{const.}$, and the equation of state on the TOV side is determined by the equations, and therefore cannot be imposed. Nevertheless, the bounds on the equations of state imply that the equations of state are qualitatively reasonable, and we expect that this family of solutions will capture the gross dynamics of solutions when more general equations of state are imposed. For more general equations of state, other waves, such as rarefaction waves and entropy waves, would need to be present to meet the conservation constraint, and thereby mediate the transition across the shock wave. Such transitional waves would be very difficult to model in an exact solution. But, the fact that we can find global solutions that meet our physical bounds, and that are qualitatively the same for all values of $\sigma \in (0,1]$ and all initial shock positions, strongly suggests that such a shock wave would be the dominant wave in a large class of problems.

In the next section, the FRW solution is derived for the case $\sigma = \text{const.}$, and the Hubble length is discussed as a critical length scale. Subsequently, the general theorems in Smoller and Temple (1994) for matching gravitational metrics across shock waves are employed. This is followed by a discussion of the construction of the family of solutions in the case $r_* = 0$. Finally, the case $r_* > 0$ is discussed. (Details can be found in Smoller and Temple (1995, 2003, 2004).)

## The FRW Metric

According to Einstein's theory of general relativity, all properties of the gravitational field are determined by a Lorentzian spacetime metric tensor $g$, whose line element in a given coordinate system $x = (x^0, \ldots, x^3)$ is given by

$$ds^2 = g_{ij} dx^i dx^j \tag{4}$$

(We use the Einstein summation convention, whereby repeated up–down indices are assumed summed from 0 to 3.) The components $g_{ij}$ of the gravitational metric $g$ satisfy the Einstein equations

$$G^{ij} = \kappa T^{ij}, \quad T^{ij} = (\rho c^2 + p) w^i w^j + p g^{ij} \tag{5}$$

where we assume that the stress-energy tensor $T$ corresponds to that of a perfect fluid. Here $G$ is the Einstein curvature tensor,

$$\kappa = \frac{8\pi\mathcal{G}}{c^4} \tag{6}$$

is the coupling constant, $\mathcal{G}$ is Newton's gravitational constant, $c$ is the speed of light, $\rho c^2$ is the energy density, $p$ is the pressure, and $\boldsymbol{w} = (w^0, \ldots, w^3)$ are the components of the 4-velocity of the fluid (cf. Weinberg 1972), and again we use the convention that $c = 1$ and $\mathcal{G} = 1$ when convenient.

Putting the metric ansatz [1] into the Einstein equations [5] gives the equations for the FRW metric (Weinberg 1972),

$$H^2 = \left(\frac{\dot{R}}{R}\right)^2 = \frac{\kappa}{3}\rho - \frac{k}{R^2} \qquad [7]$$

and

$$\dot{\rho} = -3(p+\rho)H \qquad [8]$$

The unknown quantities $R$, $\rho$, and $p$ are assumed to be functions of the FRW coordinate time $t$ alone, and the "dot" denotes differentiation with respect to $t$.

To verify that the Hubble length $\bar{r}_{\mathrm{crit}} = 1/H$ is the limit for FRW–TOV shock matching outside a black hole, write the FRW metric [1] in standard Schwarzschild coordinates $x = (\bar{r}, \bar{t})$, where the metric takes the form

$$\mathrm{d}s^2 = -B(\bar{r}, \bar{t})\mathrm{d}\bar{t}^2 + A(\bar{r}, \bar{t})^{-1}\mathrm{d}\bar{r}^2 + \bar{r}^2\mathrm{d}\Omega^2 \qquad [9]$$

and the mass function $M(\bar{r}, \bar{t})$ is defined through the relation

$$A = 1 - \frac{2M}{\bar{r}} \qquad [10]$$

It is well known that a general spherically symmetric metric can be transformed to the form [9] by coordinate transformation (see Weinberg (1972) and Groah and Temple (2004)). Substituting $\bar{r} = Rr$ into [1] and diagonalizing the resulting metric, we obtain (see Smoller and Temple (2004) for details)

$$\mathrm{d}s^2 = -\frac{1}{\psi^2}\left\{\frac{1-kr^2}{1-kr^2-H^2\bar{r}^2}\right\}\mathrm{d}\bar{t}^2 + \left\{\frac{1}{1-kr^2-H^2\bar{r}^2}\right\}\mathrm{d}\bar{r}^2 + \bar{r}^2\mathrm{d}\Omega^2 \qquad [11]$$

where $\psi$ is an integrating factor that solves the equation

$$\frac{\partial}{\partial \bar{r}}\left(\psi\frac{1-kr^2-H^2\bar{r}^2}{1-kr^2}\right) - \frac{\partial}{\partial t}\left(\psi\frac{H\bar{r}}{1-kr^2}\right) = 0 \qquad [12]$$

and the time coordinate $\bar{t} = \bar{t}(t, \bar{r})$ is defined by the exact differential

$$\mathrm{d}\bar{t} = \left(\psi\frac{1-kr^2-H^2\bar{r}^2}{1-kr^2}\right)\mathrm{d}t + \left(\psi\frac{H\bar{r}}{1-kr^2}\right)\mathrm{d}\bar{r} \qquad [13]$$

Now using [10] in [7], it follows that

$$M(t, \bar{r}) = \frac{\kappa}{2}\int_0^{\bar{r}} \rho(t)s^2\mathrm{d}s = \frac{1}{3}\frac{\kappa}{2}\rho\bar{r}^3 \qquad [14]$$

Since in the FRW metric, $\bar{r} = Rr$ measures arclength along radial geodesics at fixed time, we see from [14] that $M(t, \bar{r})$ has the physical interpretation as the total mass inside radius $\bar{r}$ at time $t$ in the FRW metric. Restricting to the case of critical expansion $k = 0$, we see from [7], [14], and [13] that $\bar{r} = H^{-1}$ is equivalent to $2M/\bar{r} = 1$, and so at fixed time $t$, the following equivalences are valid:

$$\bar{r} = H^{-1} \quad \text{iff} \quad \frac{2M}{\bar{r}} = 1 \quad \text{iff} \quad A = 0 \qquad [15]$$

We conclude that $\bar{r} = H^{-1}$ is the critical length scale for the FRW metric at fixed time $t$ in the sense that $A = 1 - 2M/\bar{r}$ changes sign at $\bar{r} = H^{-1}$, and so the universe lies inside a black hole beyond $\bar{r} = H^{-1}$, as claimed above. Now, we proved in Smoller and Temple (1998) that the standard TOV metric outside the black hole cannot be continued into $A = 0$ except in the very special case $\rho = 0$. (It takes an infinite pressure to hold up a static configuration at the event horizon of a black hole.) Thus, shock matching beyond one Hubble length requires a metric of a different character, and for this purpose, we introduce the TOV metric inside the black hole – a metric of TOV form, with $A < 0$, whose fluid is comoving with the timelike radial coordinate $\bar{r}$ (Smoller and Temple 2004).

The Hubble length $\bar{r}_{\mathrm{crit}} = c/H$ is also the critical distance at which the outward expansion of the FRW metric exactly cancels the inward advance of a radial light ray impinging on an observer positioned at the origin of a $k = 0$ FRW metric. Indeed, by [1], a light ray traveling radially inward toward the center of an FRW coordinate system satisfies the condition

$$c^2\,\mathrm{d}t^2 = R^2\,\mathrm{d}r^2 \qquad [16]$$

so that

$$\frac{\mathrm{d}\bar{r}}{\mathrm{d}t} = \dot{R}r + R\dot{r} = H\bar{r} - c = H\left(\bar{r} - \frac{c}{H}\right) > 0 \qquad [17]$$

if and only if

$$\bar{r} > \frac{c}{H}$$

Thus, the arclength distance from the origin to an inward moving light ray at fixed time $t$ in a $k = 0$ FRW metric will actually increase as long as the light ray lies beyond the Hubble length. An inward moving light ray will, however, eventually cross the Hubble length and reach the origin in finite proper time, due to the increase in the Hubble length with time.

We now calculate the infinite redshift limit in terms of the Hubble length. It is well known that light emitted at $(t_e, r_e)$ at wavelength $\lambda_e$ in an FRW spacetime will be observed at $(t_0, r_0)$ at wavelength $\lambda_0$ if

$$\frac{R_0}{R_e} = \frac{\lambda_0}{\lambda_e}$$

Moreover, the redshift factor $z$ is defined by

$$z = \frac{\lambda_0}{\lambda_e} - 1$$

Thus, infinite redshifting occurs in the limit $R_e \to 0$, where $R = 0, t = 0$ is the big bang. Consider now a light ray emitted at the instant of the big bang, and observed at the FRW origin at present time $t = t_0$. Let $r_\infty$ denote the FRW coordinate at time $t \to 0$ of the furthest objects that can be observed at the FRW origin before time $t = t_0$. Then $r_\infty$ marks the position of objects at time $t = 0$ whose radiation would be observed as infinitly redshifted (assuming no scattering). Note then that a shock wave emanating from $\bar{r} = 0$ at the instant of the big bang, will be observed at the FRW origin before present time $t = t_0$ only if its position $r$ at the instant of the big bang satisfies the condition $r < r_\infty$. To estimate $r_\infty$, note first that from [16] it follows that an incoming radial light ray in an FRW metric follows a lightlike trajectory $r = r(t)$ if

$$r - r_e = -\int_{t_e}^{t} \frac{d\tau}{R(\tau)}$$

and thus

$$r_\infty = \int_0^{t_0} \frac{d\tau}{R(\tau)} \qquad [18]$$

Using this, the following theorem can be proved (Smoller and Temple 2004).

**Theorem 1** *If the pressure p satisfies the bounds*

$$0 \le p \le \tfrac{1}{3}\rho \qquad [19]$$

*then, for any equation of state, the age of the universe $t_0$ and the infinite red shift limit $r_\infty$ are bounded in terms of the Hubble length by*

$$\frac{1}{2H_0} \le t_0 \le \frac{2}{3H_0} \qquad [20]$$

$$\frac{1}{H_0} \le r_\infty \le \frac{2}{H_0} \qquad [21]$$

(We have assumed in Theorem 1 that $R = 0$ when $t = 0$ and $R = 1$ when $t = t_0, H = H_0$.)

The next theorem gives closed-form solutions of the FRW equations [7], [8] in the case when $\sigma = \text{const.}$ As a special case, we recover the bounds in [20] and [21] from the cases $\sigma = 0$ and 1/3.

**Theorem 2** *Assume $k = 0$ and the equation of state*

$$p = \sigma\rho \qquad [22]$$

*where $\sigma$ is taken to be constant,*

$$0 \le \sigma \le 1$$

*then (assuming an expanding universe $\dot{R} > 0$), the solution of system [7], [8] satisfying $R = 0$ at $t = 0$ and $R = 1$ at $t = t_0$ is given by*

$$\rho = \frac{4}{3\kappa(1+\sigma)^2}\frac{1}{t^2} \qquad [23]$$

$$R = \left(\frac{t}{t_0}\right)^{2/[3(1+\sigma)]} \qquad [24]$$

$$\frac{H}{H_0} = \frac{t_0}{t} \qquad [25]$$

*Moreover, the age of the universe $t_0$ and the infinite red shift limit $r_\infty$ are given exactly in terms of the Hubble length by*

$$t_0 = \frac{2}{3(1+\sigma)}\frac{1}{H_0} \qquad [26]$$

$$r_\infty = \frac{2}{1+3\sigma}\frac{1}{H_0} \qquad [27]$$

From [27] we conclude that a shock wave will be observed at the FRW origin before present time $t = t_0$ only if its position $r$ at the instant of the big bang satisfies the condition

$$r < \frac{2}{1+3\sigma}\frac{1}{H_0}$$

Note that $r_\infty$ ranges from one-half to two Hubble lengths as $\sigma$ ranges from 1 to 0, taking the intermediate value of one Hubble length at $\sigma = 1/3$ (cf. [21]).

Note that using [23] and [24] in [14], it follows that

$$\begin{aligned} M &= \frac{\kappa}{2}\int_0^{\bar{r}} \rho(t)s^2 ds \\ &= \frac{2\bar{r}^3}{9(1+\sigma)^2 t_0^{2/(1+\sigma)}} t^{-2\sigma/(1+\sigma)} \end{aligned} \qquad [28]$$

so $\dot{M} < 0$ if $\sigma > 0$. It follows that if $p = \sigma\rho$, $\sigma = \text{const.} > 0$, then the total mass inside radius $r = \text{const.}$ decreases in time.

## The General Theory of Shock Matching

The matching of the FRW and TOV metrics in the next two sections is based on the following theorems that were derived in Smoller and Temple (1994) (Theorems 3 and 4 apply to non-lightlike shock surfaces. The lightlike case was discussed by Scott (2002).)

**Theorem 3** *Let $\Sigma$ denote a smooth, three-dimensional shock surface in spacetime with spacelike*

*normal vector* $\boldsymbol{n}$ *relative to the spacetime metric* $g$*; let* $K$ *denote the second fundamental form on* $\Sigma$*; and let* $G$ *denote the Einstein curvature tensor. Assume that the components* $g_{ij}$ *of the gravitational metric* $g$ *are smooth on either side of* $\Sigma$ *(continuous up to the boundary on either side separately), and Lipschitz continuous across* $\Sigma$ *in some fixed coordinate system. Then the following statements are equivalent*:

(*i*) $[K]=0$ *at each point of* $\Sigma$.
(*ii*) *The curvature tensors* $R^i_{jkl}$ *and* $G_{ij}$*, viewed as second-order operators on the metric components* $g_{ij}$*, produce no* $\delta$*-function sources on* $\Sigma$.
(*iii*) *For each point* $P\in\Sigma$*, there exists a* $C^{1,1}$ *coordinate transformation defined in a neighborhood of* $P$*, such that, in the new coordinates (which can be taken to be the Gaussian normal coordinates for the surface), the metric components are* $C^{1,1}$ *functions of these coordinates.*
(*iv*) *For each* $P\in\Sigma$*, there exists a coordinate frame that is locally Lorentzian at* $P$*, and can be reached within the class of* $C^{1,1}$ *coordinate transformations.*

*Moreover, if any one of these equivalencies hold, then the Rankine–Hugoniot jump conditions,* $[G]^{\sigma}_{i}n_{\sigma}=0$ *(which express the weak form of conservation of energy and momentum across* $\Sigma$ *when* $G=\kappa T$*), hold at each point on* $\Sigma$.

Here $[f]$ denotes the jump in the quantity $f$ across $\Sigma$ (this being determined by the metric separately on each side of $\Sigma$ because $g_{ij}$ is only Lipschitz continuous across $\Sigma$), and by $C^{1,1}$ we mean that the first derivatives are Lipschitz continuous.

In the case of spherical symmetry, the following stronger result holds. In this case, the jump conditions $[G^{ij}]n_i=0$, which express the weak form of conservation across a shock surface, are implied by a single condition $[G^{ij}]n_in_j=0$, so long as the shock is non-null, and the areas of the spheres of symmetry match smoothly at the shock and change monotonically as the shock evolves. Note that, in general, assuming that the angular variables are identified across the shock, we expect conservation to entail two conditions, one for the time and one for the radial components. The fact that the smooth matching of the spheres of symmetry reduces conservation to one condition can be interpreted as an instance of the general principle that directions of smoothness in the metric imply directions of conservation of the sources.

**Theorem 4** *Assume that* $g$ *and* $\bar{g}$ *are two spherically symmetric metrics that match Lipschitz continuously across a three-dimensional shock interface* $\Sigma$ *to form the matched metric* $g\cup\bar{g}$*. That is, assume that* $g$ *and* $\bar{g}$ *are Lorentzian metrics given by*

$$\mathrm{d}s^2=-a(t,r)\mathrm{d}t^2+b(t,r)\mathrm{d}r^2+c(t,r)\mathrm{d}\Omega^2 \qquad [29]$$

*and*

$$\mathrm{d}\bar{s}^2=-\bar{a}(\bar{t},\bar{r})\mathrm{d}\bar{t}^2+\bar{b}(\bar{t},\bar{r})\mathrm{d}\bar{r}^2+\bar{c}(\bar{t},\bar{r})\mathrm{d}\Omega^2 \qquad [30]$$

*and that there exists a smooth coordinate transformation* $\Psi:(t,r)\to(\bar{t},\bar{r})$*, defined in a neighborhood of a shock surface* $\Sigma$ *given by* $r=r(t)$*, such that the metrics agree on* $\Sigma$*. (We implicitly assume that* $\theta$ *and* $\varphi$ *are continuous across the surface.) Assume that*

$$c(t,r)=\bar{c}(\Psi(t,r)) \qquad [31]$$

*in an open neighborhood of the shock surface* $\Sigma$*, so that, in particular, the areas of the 2-spheres of symmetry in the barred and unbarred metrics agree on the shock surface. Assume also that the shock surface* $r=r(t)$ *in unbarred coordinates is mapped to the surface* $\bar{r}=\bar{r}(\bar{t})$ *by* $(\bar{t},\bar{r}(\bar{t}))=\Psi(t,r(t))$*. Assume, finally, that the normal* $\boldsymbol{n}$ *to* $\Sigma$ *is non-null, and that*

$$\boldsymbol{n}(c)\neq 0 \qquad [32]$$

*where* $\boldsymbol{n}(c)$ *denotes the derivative of the function* $c$ *in the direction of the vector* $\boldsymbol{n}$*. Then the following are equivalent to the statement that the components of the metric* $g\cup\bar{g}$ *in any Gaussian normal coordinate system are* $C^{1,1}$ *functions of these coordinates across the surface* $\Sigma$:

$$[G^i_j]n_i=0 \qquad [33]$$

$$[G^{ij}]n_in_j=0 \qquad [34]$$

$$[K]=0 \qquad [35]$$

*Here again,* $[f]=\bar{f}-f$ *denotes the jump in the quantity* $f$ *across* $\Sigma$*, and* $K$ *is the second fundamental form on the shock surface.*

We assume in Theorem 4 that the areas of the 2-spheres of symmetry change monotonically in the direction normal to the surface. For example, if $c=r^2$, then $\partial c/\partial t=0$, so the assumption $\boldsymbol{n}(c)\neq 0$ is valid except when $\boldsymbol{n}=\partial/\partial t$, in which case the rays of the shock surface would be spacelike. Thus, the shock speed would be faster than the speed of light if our assumption $\boldsymbol{n}(c)\neq 0$ failed in the case $c=r^2$.

## FRW–TOV Shock Matching Outside the Black Hole – The Case $r_*=0$

To construct the family of shock wave solutions for parameter values $0<\sigma\leq 1$ and $r_*=0$, we match the exact solution [23]–[25] of the FRW metric [1] to the TOV metric [2] outside the black hole,

assuming $A > 0$. In this case, we can bypass the problem of deriving and solving the ODEs for the shock surface and constraints discussed above, by actually deriving the exact solution of the Einstein equations of TOV form that meets these equations. This exact solution represents the general-relativistic version of a static, singular isothermal sphere – singular because it has an inverse square density profile, and isothermal because the relationship between the density and pressure is $\bar{p} = \bar{\sigma}\bar{\rho}$, $\bar{\sigma} = \text{const}$.

Assuming the stress tensor for a perfect fluid, and assuming that the density and pressure depend only on $\bar{r}$, the Einstein equations for the TOV metric [2] outside the black hole (i.e., when $A = 1 - 2M/\bar{r} > 0$) are equivalent to the Oppenheimer–Volkoff system

$$\frac{dM}{d\bar{r}} = 4\pi\bar{r}^2\bar{\rho} \qquad [36]$$

$$-\bar{r}^2\frac{d}{d\bar{r}}\bar{p} = \mathcal{G}M\bar{\rho}\left\{1 + \frac{\bar{p}}{\bar{\rho}}\right\} \times \left\{1 + \frac{4\pi\bar{r}^3\bar{p}}{M}\right\}\left\{1 - \frac{2\mathcal{G}M}{\bar{r}}\right\}^{-1} \qquad [37]$$

Integrating [36], we obtain the usual interpretation of $M$ as the total mass inside radius $\bar{r}$,

$$M(\bar{r}) = \int_0^{\bar{r}} 4\pi\xi^2\bar{\rho}(\xi)d\xi \qquad [38]$$

The metric component $B \equiv B(\bar{r})$ is determined from $\bar{\rho}$ and $M$ through the equation

$$\frac{B'(\bar{r})}{B} = -2\frac{\bar{p}'(\bar{r})}{\bar{p} + \bar{\rho}} \qquad [39]$$

Assuming

$$\bar{p} = \bar{\sigma}\bar{\rho}, \qquad \bar{\rho}(\bar{r}) = \frac{\gamma}{\bar{r}^2} \qquad [40]$$

for some constants $\bar{\sigma}$ and $\gamma$, and substituting into [3], we obtain

$$M(\bar{r}) = 4\pi\gamma\bar{r} \qquad [41]$$

Putting [40] and [41] into [37] and simplifying yields the identity

$$\gamma = \frac{1}{2\pi\mathcal{G}}\left(\frac{\bar{\sigma}}{1 + 6\bar{\sigma} + \bar{\sigma}^2}\right) \qquad [42]$$

From [38] we obtain

$$A = 1 - 8\pi\mathcal{G}\gamma < 1 \qquad [43]$$

Applying [39] leads to

$$B = B_0\left(\frac{\bar{\rho}}{\bar{\rho}_0}\right)^{-2\bar{\sigma}/(1+\bar{\sigma})} = B_0\left(\frac{\bar{r}}{\bar{r}_0}\right)^{4\bar{\sigma}/(1+\bar{\sigma})} \qquad [44]$$

By rescaling the time coordinate, we can take $B_0 = 1$ at $\bar{r}_0 = 1$, in which case [44] reduces to

$$B = \bar{r}^{4\bar{\sigma}/(1+\bar{\sigma})} \qquad [45]$$

We conclude that when [42] holds, [40]–[43] and [44] provide an exact solution of the Einstein field equations of TOV type, for each $0 \le \bar{\sigma} \le 1$. (In this case, an exact solution of TOV type was first found by Tolman (1939), and rediscovered in the case $\bar{\sigma} = 1/3$ by Misner and Zapolsky (cf. Weinberg (1972 p. 320)).) By [43], these solutions are defined outside the black hole, since $2M/\bar{r} < 1$. When $\bar{\sigma} = 1/3$, [42] yields $\gamma = 3/56\pi\mathcal{G}$ (cf. Weinberg (1972, equation (11.4.13))).

To match the FRW exact solution [23]–[25] with equation of state $p = \sigma\rho$ to the TOV exact solution [40]–[45] with equation of state $\bar{p} = \bar{\sigma}\bar{\rho}$ across a shock interface, we first set $\bar{r} = Rr$ to match the spheres of symmetry, and then match the timelike and spacelike components of the corresponding metrics in standard Schwarzschild coordinates. The matching of the $d\bar{r}^2$ coefficient $A^{-1}$ yields the conservation of mass condition that implicitly gives the shock surface $\bar{r} = \bar{r}(t)$,

$$M(\bar{r}) = \frac{4\pi}{3}\rho(t)\bar{r}^3 \qquad [46]$$

Using this together with [41] gives the following two relations that hold at the shock surface:

$$\bar{r} = \sqrt{\frac{3\gamma}{\rho(t)}}$$

$$\rho = \frac{3}{4\pi}\frac{M}{\bar{r}(t)^3} = \frac{3\gamma}{\bar{r}(t)^2} = 3\bar{\rho} \qquad [47]$$

Matching the coefficient $B$ of $d\bar{t}^2$ on the shock surface determines the integrating factor $\psi$ in a neighborhood of the shock surface by assigning initial conditions for [44]. Finally, the conservation constraint $[T_{ij}]n_i n_j = 0$ leads to the single condition

$$0 = (1 - A)(\rho + \bar{p})(p + \bar{\rho})^2 + \left(1 - \frac{1}{A}\right)(\bar{\rho} + \bar{p})(\rho + p)^2 + (p - \bar{p})(\rho - \bar{\rho})^2 \qquad [48]$$

which upon using $p = \sigma\rho$ and $\bar{p} = \bar{\sigma}\bar{\rho}$ is satisfied assuming the condition

$$\bar{\sigma} = \tfrac{1}{2}\sqrt{9\sigma^2 + 54\sigma + 49} - \tfrac{3}{2}\sigma - \tfrac{7}{2} \equiv H(\sigma) \qquad [49]$$

Alternatively, we can solve for $\sigma$ in [49] and write this relation as

$$\sigma = \frac{\bar{\sigma}(\bar{\sigma} + 7)}{3(1 - \bar{\sigma})} \qquad [50]$$

This guarantees that conservation holds across the shock surface, and so it follows from Theorem 4 that all of the equivalencies in Theorem 3 hold across the shock surface. Note that $H(0) = 0$, and to leading order $\bar{\sigma} = (3\sigma/7) + O(\sigma^2)$ as $\sigma \to 0$. Within the physical region $0 \le \sigma, \bar{\sigma} \le 1, H'(\sigma) > 0, \bar{\sigma} < \sigma$, and $H(1/3) = \sqrt{17} - 4 \approx 0.1231, H(1) = \sqrt{112}/2 - 5 \approx 0.2915$.

Using the exact formulas for the FRW metric in [23]–[25], and setting $R_0 = 1$ at $\rho = \rho_0, t = t_0$, we obtain the following exact formulas for the shock position:

$$\bar{r}(t) = \alpha t \tag{51}$$

$$r(t) = \bar{r}(t)R(t)^{-1} = \beta t^{(1+3\sigma)/(3+3\sigma)} \tag{52}$$

where

$$\alpha = 3(1+\sigma)\sqrt{\frac{\bar{\sigma}}{1+6\bar{\sigma}+\bar{\sigma}^2}}$$
$$\beta = \alpha^{(1+3\sigma)/(3+3\sigma)}\left(\frac{3\gamma}{\rho_0}\right)^{1/(3+3\sigma)} \tag{53}$$

It follows from [41] that $A > 0$, and from [52] that $r_* = \lim_{t\to 0} r(t) = 0$. The entropy condition that the shock wave be compressive follows from the fact that $\bar{\sigma} = H(\sigma) < \sigma$. Thus, we conclude that for each $0 < \sigma \le 1, r_* = 0$, the solutions constructed in [40]–[53] define a one-parameter family of shock wave solutions that evolve everywhere outside the black hole, which implies that the distance from the shock wave to the FRW center is less than one Hubble length for all $t > 0$.

Using [51] and [52], one can determine the shock speed, and check when the Lax characteristic condition (Smoller 1983) holds at the shock. The result is the following theorem. (Note that even when the shock speed is larger than $c$, only the wave, and not the sound speeds or any other physical motion, exceeds the speed of light. See Scott (2002) for the case when the shock speed is equal to the speed of light.) The reader is referred to Smoller and Temple (1995) for details.

**Theorem 5** *There exist values* $0 < \sigma_1 < \sigma_2 < 1$, $(\sigma_1 \approx 0.458, \sigma_2 = \sqrt{5}/3 \approx 0.745)$, *such that, for* $0 < \sigma \le 1$, *the Lax characteristic condition holds at the shock if and only if* $0 < \sigma < \sigma_1$; *and the shock speed is less than the speed of light if and only if* $0 < \sigma < \sigma_2$.

The explicit solution in the case $r_* = 0$ can be interpreted as a general-relativistic version of a shock wave explosion into a static, singular, isothermal sphere, known in the Newtonian case as a simple model for star formation (Smoller and Temple 2000). As the scenario goes, a star begins as a diffuse cloud of gas. The cloud slowly contracts under its own gravitational force by radiating energy out through the gas cloud as gravitational potential energy is converted into kinetic energy. This contraction continues until the gas cloud reaches the point where the mean free path for transmission of light is small enough that light is scattered, instead of being transmitted, through the cloud. The scattering of light within the gas cloud has the effect of equalizing the temperature within the cloud, and at this point the gas begins to drift toward the most compact configuration of the density that balances the pressure when the equation of state is isothermal. This configuration is a static, singular, isothermal sphere, the general-relativistic version of which is the exact TOV solution beyond the shock wave when $r_* = 0$. This solution in the Newtonian case is also inverse square in the density and pressure, and so the density tends to infinity at the center of the sphere. Eventually, the high densities at the center ingnite thermonuclear reactions. The result is a shock wave explosion emanating from the center of the sphere, and this signifies the birth of the star. The exact solutions when $r_* = 0$ represent a general-relativistic version of such a shock wave explosion.

## Shock Wave Solutions Inside the Black Hole – The Case $r_* > 0$

When the shock wave is beyond one Hubble length from the FRW center, we obtain a family of shock wave solutions for each $0 < \sigma \le 1$ and $r_* > 0$ by shock matching the FRW metric [1] to a TOV metric of form [2] under the assumption that

$$A(\bar{r}) = 1 - \frac{2M(\bar{r})}{\bar{r}} \equiv 1 - N(\bar{r}) < 0 \tag{54}$$

In this case, $\bar{r}$ is the timelike variable. Assuming that the stress tensor $T$ is taken to be that of a perfect fluid comoving with the TOV metric, the Einstein equations $G = \kappa T$, inside the black hole, take the form (see Smoller and Temple (2004) for details)

$$\bar{p}' = \frac{\bar{p}+\bar{\rho}}{2}\frac{N'}{N-1} \tag{55}$$

$$N' = -\left\{\frac{N}{\bar{r}} + \kappa\bar{p}\bar{r}\right\} \tag{56}$$

$$\frac{B'}{B} = -\frac{1}{N-1}\left\{\frac{N}{\bar{r}} + \kappa\bar{\rho}\right\} \tag{57}$$

The system [55]–[57] defines the simplest class of gravitational metrics that contain matter, evolve inside the black hole, and such that the mass function $M(\bar{r}) < \infty$ at each fixed time $\bar{r}$. System [55]–[57] for $A < 0$ differs substantially from the TOV equations for $A > 0$ because, for example, the energy density $T^{00}$ is equated with the timelike component $G^{rr}$ when $A < 0$, but with $G^{tt}$ when $A > 0$. In particular, this implies that, inside the black hole, the mass function $M(\bar{r})$ does not have the interpretation as a total mass inside the radius $\bar{r}$ as it does outside the black hole.

Equations [56], [57] do not have the same character as [54], [55] and the relation $\bar{p} = \bar{\sigma}\bar{\rho}$ with $\bar{\sigma} = \text{const.}$ is inconsistent with [56], [57] together with the conservation constraint and the FRW assumption $p = \sigma\rho$ for shock matching. Thus, instead of looking for an exact solution of [56], [57] ahead of time, as in the case $r_* = 0$, we assume the FRW solution [23]–[25], and derive the ODEs that describe the TOV metrics that match this FRW metric Lipschitz-continuously across a shock surface, and then impose the conservation, entropy, and equation of state constraints at the end. Matching a given $k = 0$ FRW metric to a TOV metric inside the black hole across a shock interface leads to the system of ODEs, (see Smoller and Temple (2004) for details),

$$\frac{\mathrm{d}u}{\mathrm{d}N} = -\left\{\frac{(1+u)}{2(1+3u)N}\right\} \times \left\{\frac{(3u-1)(\sigma-u)N + 6u(1+u)}{(\sigma-u)N + (1+u)}\right\} \qquad [58]$$

$$\frac{\mathrm{d}\bar{r}}{\mathrm{d}N} = -\frac{1}{1+3u}\frac{\bar{r}}{N} \qquad [59]$$

with conservation constraint

$$v = \frac{-\sigma(1+u) + (\sigma-u)N}{(1+u) + (\sigma-u)N} \qquad [60]$$

where

$$u = \frac{\bar{p}}{\rho}, \quad v = \frac{\bar{\rho}}{\rho}, \quad \sigma = \frac{p}{\rho} \qquad [61]$$

Here $\rho$ and $p$ denote the (known) FRW density and pressure, and all variables are evaluated at the shock. Solutions of [58]–[60] determine the (unknown) TOV metrics that match the given FRW metric Lipschitz-continuously across a shock interface, such that conservation of energy and momentum hold across the shock, and such that there are no $\delta$-function sources at the shock (Israel 1966, Smoller and Temple 1997). Note that the dependence of [58]–[60] on the FRW metric is only through the variable $\sigma$, and so the advantage of taking $\sigma = \text{const.}$ is that the whole solution is determined by the inhomogeneous scalar equation [58] when $\sigma = \text{const.}$ We take as the entropy constraint the condition that

$$0 < \bar{p} < p, \qquad 0 < \bar{\rho} < \rho \qquad [62]$$

and to insure a physically reasonable solution, we impose the equation of state constriant on the TOV side of the shock (this is equivalent to the dominant energy condition (Blau and Guth 1987))

$$0 < \bar{p} < \bar{\rho} \qquad [63]$$

Condition [62] implies that outgoing shock waves are compressive. Inequalities [62] and [63] are both implied by the single condition (Smoller and Temple 2004),

$$\frac{1}{N} < \left(\frac{1-u}{1+u}\right)\left(\frac{\sigma-u}{\sigma+u}\right) \qquad [64]$$

Since $\sigma$ is constant, eqn [58] uncouples from [59], and thus solutions of system [58]–[60] are determined by the scalar nonautonomous equation [58]. Making the change of variable $S = 1/N$, which transforms the "big bang" $N \to \infty$ over to a rest point at $S \to 0$, we obtain

$$\frac{\mathrm{d}u}{\mathrm{d}S} = \left\{\frac{(1+u)}{2(1+3u)S}\right\} \times \left\{\frac{(3u-1)(\sigma-u) + 6u(1+u)S}{(\sigma-u) + (1+u)S}\right\} \qquad [65]$$

Note that the conditions $N > 1$ and $0 < \bar{p} < p$ restrict the domain of [65] to the region $0 < u < \sigma < 1, 0 < S < 1$. The next theorem gives the existence of solutions for $0 < \sigma \leq 1, r_* > 0$, inside the black hole (Smoller and Temple 2003).

**Theorem 6** *For every $\sigma$, $0 < \sigma < 1$, there exists a unique solution $u_\sigma(S)$ of [65], such that [64] holds on the solution for all $S, 0 < S < 1$, and on this solution, $0 < u_\sigma(S) < \bar{u}$, $\lim_{S\to 0} u_\sigma(S) = \bar{u}$, where*

$$\bar{u} = \mathrm{Min}\{1/3, \sigma\} \qquad [66]$$

*and*

$$\lim_{S\to 1} \bar{p} = 0 = \lim_{S\to 1} \bar{\rho} \qquad [67]$$

*For each of these solutions $u_\sigma(S)$, the shock position is determined by the solution of [59], which in turn is determined uniquely by an initial condition which can be taken to be the FRW radial position of the shock wave at the instant of the big bang,*

$$r_* = \lim_{S\to 0} r(S) > 0 \qquad [68]$$

Concerning the shock speed, we have

**Theorem 7** *Let $0 < \sigma < 1$. Then the shock wave is everywhere subluminous, that is, the shock speed $s_\sigma(S) \equiv s(u_\sigma(S)) < 1$ for all $0 < S \leq 1$, if and only if $\sigma \leq 1/3$.*

Concerning the shock speed near the big bang $S=0$, the following is true:

**Theorem 8** *The shock speed at the big bang $S=0$ is given by*

$$\lim_{S\to 0} s_\sigma(S) = 0, \quad \sigma < 1/3 \qquad [69]$$

$$\lim_{S\to 0} s_\sigma(S) = \infty, \quad \sigma > 1/3 \qquad [70]$$

$$\lim_{S\to 0} s_\sigma(S) = 1, \quad \sigma = 1/3 \qquad [71]$$

Theorem 8 shows that the equation of state $p=\rho/3$ plays a special role in the analysis when $r_* > 0$, and only for this equation of state does the shock wave emerge at the big bang at a finite nonzero speed, the speed of light. Moreover, [66] implies that in this case, the correct relation $\bar{p}/\bar{\rho}=\bar{\sigma}$ is also achieved in the limit $S \to 0$. The result [67] implies that (neglecting the pressure $p$ at this time onward), the solution continues to a $k=0$ Oppenheimer–Snyder solution outside the black hole for $S > 1$.

It follows that the shock wave will first become visible at the FRW center $\bar{r}=0$ at the moment $t=t_0, (R(t_0)=1)$, when the Hubble length $H_0^{-1}=H^{-1}(t_0)$ satisfies

$$\frac{1}{H_0} = \frac{1+3\sigma}{2} r_* \qquad [72]$$

where $r_*$ is the FRW position of the shock at the instant of the big bang. At this time, the number of Hubble lengths $\sqrt{N_0}$ from the FRW center to the shock wave at time $t=t_0$ can be estimated by

$$1 \leq \frac{2}{1+3\sigma} \leq \sqrt{N_0} \leq \frac{2}{1+3\sigma} e^{\sqrt{3\sigma}((1+3\sigma)/(1+\sigma))}$$

Thus, in particular, the shock wave will still lie beyond the Hubble length $1/H_0$ at the FRW time $t_0$ when it first becomes visible. Furthermore, the time $t_{\text{crit}} > t_0$ at which the shock wave will emerge from the white hole given that $t_0$ is the first instant at which the shock becomes visible at the FRW center, can be estimated by

$$\frac{2}{1+3\sigma} e^{\sigma/4} \leq \frac{t_{\text{crit}}}{t_0} \leq \frac{2}{1+3\sigma} e^{2\sqrt{3\sigma}/(1+\sigma)} \qquad [73]$$

for $0 < \sigma \leq 1/3$, and by the better estimate

$$e^{\sqrt{6}/4} \leq \frac{t_{\text{crit}}}{t_0} \leq e^{3/2} \qquad [74]$$

in the case $\sigma=1/3$. Inequalities [73], [74] imply, for example, that at the Oppenheimer–Snyder limit $\sigma=0$,

$$\sqrt{N_0} = 2, \qquad \frac{t_{\text{crit}}}{t_0} = 2$$

and in the limit $\sigma=1/3$,

$$1.8 \leq \frac{t_{\text{crit}}}{t_0} \leq 4.5, \qquad 1 < \sqrt{N_0} \leq 4.5$$

We can conclude that at the moment $t_0$ when the shock wave first becomes visible at the FRW center, the shock wave must lie within 4.5 Hubble lengths of the FRW center. Throughout the expansion up until this time, the expanding universe must lie entirely within a white hole – the universe will eventually emerge from this white hole, but not until some later time $t_{\text{crit}}$, where $t_{\text{crit}}$ does not exceed $4.5t_0$.

## Conclusion

We believe that the existence of a wave at the leading edge of the expansion of the galaxies is the most likely possibility. The alternatives are that either the universe of expanding galaxies goes on out to infinity, or else the universe is not simply connected. Although the first possibility has been believed for most of the history of cosmology based on the Friedmann universe, we find this implausible and arbitrary in light of the shock wave refinements of the FRW metric discussed here. The second possibility, that the universe is not simply connected, has received considerable attention recently (Klarreich 2003). However, since we have not seen, and cannot create, any non-simply-connected 3-spaces on any other length scale, and since there is no observational evidence to support this, we view this as less likely than the existence of a wave at the leading edge of the expansion of the galaxies, left over from the big bang. Recent analysis of the microwave background radiation data shows a cutoff in the angular frequencies consistent with a length scale of around one Hubble length (Andy Abrecht, private communication). This certainly makes one wonder whether this cutoff is evidence of a wave at this length scale, especially given the consistency of this possibility with the case $r_* > 0$ of the family of exact solutions discussed here.

## Acknowledgments

The work of JS was supported in part by NSF Applied Mathematics Grant Number DMS-010-3998, and that of BT by NSF Applied Mathematics Grant Number DMS-010-2493.

*See also:* Black Hole Mechanics; Cosmology: Mathematical Aspects; Newtonian Limit of General Relativity; Symmetric Hyperbolic Systems and Shock Waves.

## Further Reading

Blau SK and Guth AH (1987) Inflationary cosmology. In: Hawking SW and Israel W (eds.) *Three Hundred Years of Gravitation*, pp. 524–603. Cambridge: Cambridge University Press.
Groah J, Smoller J, and Temple B (2003) Solving the Einstein equations by Lipschitz continuous metrics: shockwaves in general relativity. In: Friedlander S and Serre D (eds.) *Handbook of Mathematical Fluid Dynamics*, vol. 2, pp. 501–597. Amsterdam: North Holland.
Groah J and Temple B (2004) Shock-Wave Solutions of the Einstein Equations: Existence and Consistency by a Locally Inertial Glimm Scheme. *Memoirs of the AMS*, 84 pages, vol. 172, no. B13.
Israel W (1966) Singular hypersurfaces and thin shells in general relativity. *IL Nuovo Cimento* XLIV B(1): 1–14.
Klarreich E (2003) The shape of space. *Science News* 164–165.
Oppenheimer JR and Snyder JR (1939) On continued gravitational contraction. *Physical Review* 56: 455–459.
Scott M (2002) *General Relativistic Shock Waves Propagating at the Speed of Light*. Ph.D. thesis, UC-Davis.
Smoller J (1983) *Shock Waves and Reaction Diffusion Equations*. New York, Berlin: Springer.
Smoller J and Temple B (1994) Shock-wave solutions of the Einstein equations: the Oppenheimer–Snyder model of gravitational collapse extended to the case of non-zero pressure. *Archives for Rational and Mechanical Analysis* 128: 249–297.
Smoller J and Temple B (1995) Astrophysical shock-wave solutions of the Einstein equations. *Physical Review D* 51(6): 2733–2743.
Smoller J and Temple B (1997) Solutions of the Oppenheimer–Volkoff equations inside 9/8'ths of the Schwarzschild radius. *Communications in Mathematical Physics* 184: 597–617.
Smoller J and Temple B (1998) On the Oppenheimer–Volkov equations in general relativity. *Archives for Rational and Mechanical Analysis* 142: 177–191.
Smoller J and Temple B (2000) Cosmology with a shock wave. *Communications in Mathematical Physics* 210: 275–308.
Smoller J and Temple B (2003) Shock wave cosmology inside a black hole. *Proceedings of the National Academy of Sciences of the United States of America* 100(20): 11216–11218.
Smoller J and Temple B (2004) Cosmology, black holes, and shock waves beyond the Hubble length. *Methods and Applications of Analysis* 11(1): 77–132.
Tolman R (1939) Static solutions of Einstein's field equations for spheres of fluid. *Physical Review* 55: 364–374.
Weinberg S (1972) *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. New York: Wiley.

**Shock Waves** *see* Symmetric Hyperbolic Systems and Shock Waves

# Short-Range Spin Glasses: The Metastate Approach

**C M Newman**, New York University, New York, NY, USA
**D L Stein**, University of Arizona, Tucson, AZ, USA



## Introduction

The nature of the low-temperature spin glass phase in short-range models remains one of the central problems in the statistical mechanics of disordered systems (Binder and Young 1986, Chowdhury 1986, Mézard *et al.* 1987, Stein 1989, Fischer and Hertz 1991, Dotsenko 2001, Newman and Stein 2003). While many of the basic questions remain unanswered, analytical and rigorous work over the past decade have greatly streamlined the number of possible scenarios for pure state structure and organization at low temperatures, and have clarified the thermodynamic behavior of these systems.

The unifying concept behind this work is that of the "metastate." It arose independently in two different constructions (Aizenman and Wehr 1990, Newman and Stein 1996b), which were later shown to be equivalent (Newman and Stein 1998a). The metastate is a probability measure on the space of all thermodynamic states. Its usefulness arises in situations where multiple "competing" pure states may be present. In such situations it may be difficult to construct individual states in a measurable and canonical way; the metastate avoids this difficulty by focusing instead on the statistical properties of the states.

An important aspect of the metastate approach is that it relates, by its very construction (Newman and Stein 1996b), the observed behavior of a system in large but finite volumes with its thermodynamic properties. It therefore serves as a (possibly indispensable) tool for analyzing and understanding both the infinite-volume and finite-volume properties of a system, particularly in cases where a straightforward interpolation between the two may be incorrect, or their relation otherwise difficult to analyze.

We will focus on the Edwards–Anderson (EA) Ising spin glass model (Edwards and Anderson

1975), although most of our discussion is relevant to a much larger class of realistic models. The EA model is described by the Hamiltonian

$$\mathcal{H}_{\mathcal{J}} = -\sum_{\langle x,y\rangle} J_{xy}\sigma_x\sigma_y \qquad [1]$$

where $\mathcal{J}$ denotes a particular realization of all of the couplings $J_{xy}$ and the brackets indicate that the sum is over nearest-neighbor pairs only, with $x, y \in \mathbf{Z}^d$. We will take Ising spins $\sigma_x = \pm 1$; although this will affect the details of our discussion, it is unimportant for our main conclusions. The couplings $J_{xy}$ are quenched, independent, identically distributed random variables whose common distribution $\mu$ is symmetric about zero.

## States and Metastates

We are interested in both finite-volume and infinite-volume Gibbs states. For the cube of length scale $L$, $\Lambda_L = \{-L, -L+1, \ldots, L\}^d$, we define $\mathcal{H}_{\mathcal{J},L}$ to be the restriction of the EA Hamiltonian to $\Lambda_L$ with a specified boundary condition such as free, fixed, or periodic. Then the finite-volume Gibbs distribution $\rho_{\mathcal{J}}^{(L)} = \rho_{\mathcal{J},\beta}^{(L)}$ on $\Lambda_L$ (at inverse temperature $\beta = 1/T$) is

$$\rho_{\mathcal{J},\beta}^{(L)}(\sigma) = Z_L^{-1}\exp\{-\beta\mathcal{H}_{\mathcal{J},L}(\sigma)\} \qquad [2]$$

where the partition function $Z_L(\beta)$ is such that the sum of $\rho_{\mathcal{J},\beta}^{(L)}$ over all $\sigma$ yields 1. (In this and all succeeding definitions, the dependence on spatial dimension $d$ will be suppressed.)

Thermodynamic states are described by infinite-volume Gibbs measures. At fixed inverse temperature $\beta$ and coupling realization $\mathcal{J}$, a thermodynamic state $\rho_{\mathcal{J},\beta}$ is the limit, as $L \to \infty$, of some sequence of such finite-volume measures (each with a specified boundary condition, which may remain the same or may change with $L$). A thermodynamic state $\rho_{\mathcal{J},\beta}$ can also be characterized intrinsically through the Dobrushin–Lanford–Ruelle (DLR) equations (see, e.g., Georgii 1988): for any $\Lambda_L$, the conditional distribution of $\rho_{\mathcal{J},\beta}$ (conditioned on the sigma-field generated by $\{\sigma_x : x \in \mathbf{Z}^d \backslash \Lambda_L\}$ is $\rho_{\mathcal{J},\beta}^{(L),\overline{\sigma}}$, where $\overline{\sigma}$ is given by the conditioned values of $\sigma_x$ for $x$ on the boundary of $\Lambda_L$.

Consider now the set $\mathcal{G} = \mathcal{G}(\mathcal{J},\beta)$ of all thermodynamic states at a fixed $(\mathcal{J},\beta)$. The set of extremal, or pure, Gibbs states is defined by

$$\text{ex}\,\mathcal{G} = \mathcal{G} \setminus \{a\rho_1 + (1-a)\rho_2 : \\ a \in (0,1);\ \rho_1, \rho_2 \in \mathcal{G};\ \rho_1 \neq \rho_2\} \qquad [3]$$

and the number of pure states $\mathcal{N}(\mathcal{J},\beta)$ at $(\mathcal{J},\beta)$ is the cardinality $|\text{ex}\,\mathcal{G}|$ of ex $\mathcal{G}$. It is not hard to show that, in any $d$ and for a.e. $\mathcal{J}$, the following two statements are true: (1) $\mathcal{N} = 1$ at sufficiently low $\beta > 0$; (2) at any fixed $\beta$, $\mathcal{N}$ is constant a.s. with respect to the $\mathcal{J}$'s. (The last assertion follows from the measurability and translation invariance of $\mathcal{N}$, and the translation ergodicity of the disorder distribution of $\mathcal{J}$.)

A pure state $\rho_\alpha$ (where $\alpha$ is a pure-state index) can also be intrinsically characterized by a "clustering property"; for two-point correlation functions, this reads

$$\langle\sigma_x\sigma_y\rangle_{\rho_\alpha} - \langle\sigma_x\rangle_{\rho_\alpha}\langle\sigma_y\rangle_{\rho_\alpha} \to 0 \qquad [4]$$

as $|x - y| \to \infty$. A simple observation (Newman and Stein 1992), with important consequences for spin glasses, is that if many pure states exist, a sequence of $\rho_{\mathcal{J},\beta}^{(L)}$'s, with boundary conditions and $L$'s chosen independently of $\mathcal{J}$, will generally not have a (single) limit. We call this phenomenon "chaotic size dependence" (CSD).

We will be interested in the properties of ex $\mathcal{G}$ at low temperatures. If the spin-flip symmetry present in the EA Hamiltonian equation [1] is spontaneously broken above some dimension $d_0$ and below some temperature $T_c(d)$, there will be at least a pair of pure states such that their even-spin correlations are identical and their odd-spin correlations have the opposite sign. Assuming that such broken spin-flip symmetry indeed exists for $d > d_0$ and $T < T_c(d)$, the question of whether there exists more than one such pair (of spin-flip related extremal infinite-volume Gibbs distributions) is a central unresolved issue for the EA and related models. If many such pairs should exist, we can ask about the structure of their relations with one another, and how this structure would manifest itself in large but finite volumes. To do this, we use an approach, introduced by Newman and Stein (1996b), to study inhomogeneous and other systems with many competing pure states. This approach, based on an analogy with chaotic dynamical systems, requires the construction of a new thermodynamic quantity which is called the "metastate" – a probability measure $\kappa_{\mathcal{J}}$ on the thermodynamic states. The metastate allows an understanding of CSD by analyzing the way in which $\rho_{\mathcal{J},\beta}^{(L)}$ "samples" from its various possible limits as $L \to \infty$.

The analogy with chaotic dynamical systems can be understood as follows. In dynamical systems, the chaotic motion along a deterministic orbit is analyzed in terms of some appropriately selected probability measure, invariant under the dynamics. Time along the orbit is replaced, in our context, by $L$ and the phase space of the dynamical system is replaced by the space of Gibbs states.

Newman and Stein (1996b) considered a "microcanonical ensemble" (as always, at fixed $\beta$, which will hereafter be suppressed for ease of notation) $\kappa_N$ in which each of the finite-volume Gibbs states

$\rho_{\mathcal{J}}^{(L_1)}, \rho_{\mathcal{J}}^{(L_2)}, \ldots, \rho_{\mathcal{J}}^{(L_N)}$ has weight $N^{-1}$. The ensemble $\kappa_N$ converges to a metastate $\kappa_{\mathcal{J}}$ as $N \to \infty$, in the following sense: for every (nice) function $g$ on states (e.g., a function of finitely many correlations),

$$\lim_{N\to\infty} N^{-1} \sum_{\ell=1}^{N} g(\rho^{(L_\ell)}) = \int g(\Gamma)\, \mathrm{d}\kappa_{\mathcal{J}}(\Gamma) \qquad [5]$$

The information contained in $\kappa_{\mathcal{J}}$ effectively specifies the fraction of cube sizes $L_\ell$ which the system spends in different (possibly mixed) thermodynamic states $\Gamma$ as $\ell \to \infty$.

A different, but in the end equivalent, approach based on $\mathcal{J}$-randomness is due to Aizenman and Wehr (1990). Here one considers the random pair $(\mathcal{J}, \rho_{\mathcal{J}}^{(L)})$, defined on the underlying probability space of $\mathcal{J}$, and takes the limit $\kappa^\dagger$ (with conditional distribution $\kappa_{\mathcal{J}}^\dagger$, given $\mathcal{J}$), via finite-dimensional distributions along some subsequence. The details are omitted here, and the reader is referred to the work by Aizenman and Wehr (1990) and Newman and Stein (1998a). We note, however, the important result that a "deterministic" subsequence of volumes can be found on which [5] is valid and also $(\mathcal{J}, \rho_{\mathcal{J}}^{(L)})$ converges, with $\kappa_{\mathcal{J}}^\dagger = \kappa_{\mathcal{J}}$ (Newman and Stein 1998a).

In what follows we use the term "metastate" as shorthand for the $\kappa_{\mathcal{J}}$ constructed using periodic boundary conditions on a sequence of volumes chosen independently of the couplings, and along which $\kappa_{\mathcal{J}} = \kappa_{\mathcal{J}}^\dagger$. We choose periodic boundary conditions for specificity; the results and claims discussed are expected to be independent of the boundary conditions used, as long as they are chosen independently of the couplings.

## Low-Temperature Structure of the EA Model

There have been several scenarios proposed for the spin-glass phase of the Edwards–Anderson model at sufficiently low temperature and high dimension. These remain speculative, because it has not even been proved that a phase transition from the high-temperature phase exists at positive temperature in any finite dimension.

As noted earlier, at sufficiently high temperature in any dimension (and at all nonzero temperatures in one and presumably two dimensions, although the latter assertion has not been proved), there is a unique Gibbs state. It is conceivable that this remains the case in all dimensions and at all nonzero temperatures, in which case the metastate $\kappa_{\mathcal{J}}$ is, for a.e. $\mathcal{J}$, supported on a single, pure Gibbs state $\rho_{\mathcal{J}}$. (It is important to note, however, that in principle such a trivial metastate could occur even if $\mathcal{N} > 1$; indeed, just such a situation of "weak uniqueness" (van Enter and Fröhlich 1985, Campanino *et al.* 1987) happens in very long range spin glasses at high temperatures (Fröhlich and Zegarlinski 1987, Gandolfi *et al.* 1993).)

A phase transition has been proved to exist (Aizenman *et al.* 1987) in the Sherrington–Kirkpatrick (SK) model (Sherrington and Kirkpatrick 1975), which is the infinite-range version of the EA model. Numerical (Ogielski 1985, Ogielski and Morgenstern 1985, Binder and Young 1986, Kawashima and Young 1996) and some analytical (Fisher and Singh 1990, Thill and Hilhorst 1996) work has led to a general consensus that above some dimension (typically around three or four) there does exist a positive-temperature phase transition below which spin-flip symmetry is broken, that is, in which pure states come in pairs, as discussed below eqn [4]. Because much of the literature has focused on this possibility, we assume it in what follows, and the metastate approach turns out to be highly useful in restricting the scenarios that can occur. The simplest such scenario is a two-state picture in which, below the transition temperature $T_c$, there exists a single pair of global flip-related pure states $\rho_{\mathcal{J}}^{\alpha}$ and $\rho_{\mathcal{J}}^{-\alpha}$. In this case, there is no CSD for periodic boundary conditions and the metastate can be written as

$$\kappa_{\mathcal{J}} = \delta_{\frac{1}{2}\rho_{\mathcal{J}}^{\alpha} + \frac{1}{2}\rho_{\mathcal{J}}^{-\alpha}} \qquad [6]$$

That is, the metastate is supported on a single (mixed) thermodynamic state.

The two-state scenario that has received the most attention in the literature is the "droplet/scaling" picture (McMillan 1984, Fisher and Huse 1986, 1988, Bray and Moore 1985). In this picture a low-energy excitation above the ground state in $\Lambda_L$ is a droplet whose surface area scales as $l^{d_s}$, with $l \sim O(L)$ and $d_s < d$, and whose surface energy scales as $l^\theta$, with $\theta > 0$ (in dimensions where $T_c > 0$). More recently, an alternative picture has arisen (Krzakala and Martin 2000, Palassini and Young 2000) in which the low-energy excitations differ from those of droplet/scaling, in that their energies scale as $l^{\theta'}$, with $\theta' = 0$.

The low-temperature picture that has perhaps generated the most attention in the literature is the replica symmetry breaking (RSB) scenario (Binder and Young 1986, Marinari *et al.* 1994, 1997, Franz *et al.* 1998, Marinari *et al.* 2000, Marinari and Parisi 2000, 2001, Dotsenko 2001), which assumes a rather complicated pure-state structure, inspired by Parisi's solution of the SK model (Parisi 1979, 1983, Mézard *et al.* 1984, 1987). This is a many-state picture ($\mathcal{N} = \infty$ for a.e.

$\mathcal{J}$) in which the ordering is described in terms of the "overlaps" between states. There has been some ambiguity in how to describe such a picture for short-range models; the prevailing, or standard, view. Consider any reasonably constructed thermodynamic state $\rho_{\mathcal{J}}$ (see Newman and Stein (1998a) for more details) – e.g., the "average" over the metastate $\kappa_{\mathcal{J}}$

$$\rho_{\mathcal{J}} = \int \Gamma \, d\kappa_{\mathcal{J}}(\Gamma) \qquad [7]$$

Now choose $\sigma$ and $\sigma'$ from the product distribution $\rho_{\mathcal{J}}(\sigma)\rho_{\mathcal{J}}(\sigma')$. The overlap $Q$ is defined as

$$Q = \lim_{L\to\infty} |\Lambda_L|^{-1} \sum_{x\in\Lambda_L} \sigma_x \sigma'_x \qquad [8]$$

and $P_{\mathcal{J}}(q)$ is defined to be its probability distribution.

In the standard RSB picture, $\rho_{\mathcal{J}}$ is a mixture of infinitely many pure states, each with a specific $\mathcal{J}$-dependent weight $W$:

$$\rho_{\mathcal{J}}(\sigma) = \sum_{\alpha} W^{\alpha}_{\mathcal{J}} \rho^{\alpha}_{\mathcal{J}}(\sigma) \qquad [9]$$

If $\sigma$ is drawn from $\rho^{\alpha}_{\mathcal{J}}$ and $\sigma'$ from $\rho^{\beta}_{\mathcal{J}}$, then the expression in eqn [8] equals its thermal mean,

$$q^{\alpha\beta}_{\mathcal{J}} = \lim_{L\to\infty} |\Lambda_L|^{-1} \sum_{x\in\Lambda_L} \langle\sigma_x\rangle_{\alpha} \langle\sigma_x\rangle_{\beta} \qquad [10]$$

and hence $P_{\mathcal{J}}$ is given by

$$P_{\mathcal{J}}(q) = \sum_{\alpha,\beta} W^{\alpha}_{\mathcal{J}} W^{\beta}_{\mathcal{J}} \delta(q - q^{\alpha\beta}_{\mathcal{J}}) \qquad [11]$$

The "self-overlap," or EA order parameter, is given by $q_{\mathrm{EA}} = q^{\alpha\alpha}_{\mathcal{J}}$ and (at fixed $T$) is thought to be independent of both $\alpha$ and $\mathcal{J}$ (with probability 1).

According to the standard RSB scenario, the $W^{\alpha}_{\mathcal{J}}$'s and $q^{\alpha\beta}_{\mathcal{J}}$'s are non-self-averaging (i.e., $\mathcal{J}$-dependent) quantities, except for $\alpha = \beta$ or its global flip, where $q^{\alpha\beta}_{\mathcal{J}} = \pm q_{\mathrm{EA}}$. The average $P_s(q)$ of $P_{\mathcal{J}}(q)$ over the disorder distribution of $\mathcal{J}$ is predicted to be a mixture of two delta-function components at $\pm q_{\mathrm{EA}}$ and a continuous part between them. However, it was proved by Newman and Stein (1996c) that this scenario cannot occur, because of the translation invariance of $P_{\mathcal{J}}(q)$ and the translation ergodicity of the disorder distribution. Nevertheless, the metastate approach suggests an alternative, nonstandard, RSB scenario, which is described next.

The idea behind the nonstandard RSB picture (referred to by us as the nonstandard SK picture in earlier papers) is to produce the finite-volume behavior of the SK model to the maximum extent possible. We therefore assume in this picture that in each $\Lambda_L$, the finite-volume Gibbs state $\rho^{(L)}_{\mathcal{J}}$ is well approximated deep in the interior by a mixed thermodynamic state $\Gamma^{(L)}$, decomposable into many pure states $\rho_{\alpha_L}$ (explicit dependence on $\mathcal{J}$ is suppressed for ease of notation). More precisely, each $\Gamma$ in $\kappa_{\mathcal{J}}$ satisfies

$$\Gamma = \sum_{\alpha_\Gamma} W^{\alpha_\Gamma}_{\Gamma} \rho_{\alpha_\Gamma} \qquad [12]$$

and is presumed to have a nontrivial overlap distribution for $\sigma, \sigma'$ from $\Gamma(\sigma)\Gamma(\sigma')$:

$$P_{\Gamma}(q) = \sum_{\alpha_\Gamma,\beta_\Gamma} W^{\alpha_\Gamma}_{\Gamma} W^{\beta_\Gamma}_{\Gamma} \delta(q - q_{\alpha_\Gamma\beta_\Gamma}) \qquad [13]$$

as did $\rho_{\mathcal{J}}$ in the standard RSB picture.

Because $\kappa_{\mathcal{J}}$, like its counterpart $\rho_{\mathcal{J}}$ in the standard SK picture, is translation covariant, the resulting ensemble of overlap distributions $P_{\Gamma}$ is independent of $\mathcal{J}$. Because of the CSD present in this scenario, the overlap distribution for $\rho^{(L)}_{\mathcal{J}}$ varies with $L$, no matter how large $L$ becomes. So, instead of averaging the overlap distribution over $\mathcal{J}$, the averaging must now be done over the states $\Gamma$ within the metastate $\kappa_{\mathcal{J}}$, all at fixed $\mathcal{J}$:

$$P_{\mathrm{ns}}(q) = \int P_{\Gamma}(q) \kappa_{\mathcal{J}}(\Gamma) d\Gamma \qquad [14]$$

The $P_{\mathrm{ns}}(q)$ is the same for a.e. $\mathcal{J}$, and has a form analogous to the $P_{\mathrm{s}}(q)$ in the standard RSB picture.

However, the nonstandard RSB scenario seems rather unlikely to occur in any natural setting, because of the following result:

**Theorem** Newman and Stein 1998b). (*Consider two metastates constructed along (the same) deterministic sequence of $\Lambda_L$'s, using two different sequences of flip-related, coupling-independent boundary conditions (such as periodic and antiperiodic). Then with probability one, these two metastates are the same.*

The proof is given by Newman and Stein (1998b), but the essential idea can be easily described here. As discussed earlier, $\kappa_{\mathcal{J}} = \kappa^{\dagger}_{\mathcal{J}}$; but $\kappa^{\dagger}_{\mathcal{J}}$ is constructed by a limit of finite-dimensional distributions, which means averaging over other couplings including the ones near the system boundary, and hence gives the same metastate for two flip-related boundary conditions.

This invariance with respect to different sequences of periodic and antiperiodic boundary conditions means essentially that the frequency of appearance of various thermodynamic states $\Gamma^{(L)}$ in finite volumes $\Lambda_L$ is independent of the choice of boundary conditions. Moreover, this same

invariance property holds among any two sequences of fixed boundary conditions (and the fixed boundary condition of choice may even be allowed to vary arbitrarily along any single sequence of volumes)! It follows that, with respect to changes of boundary conditions, the metastate is extraordinarily robust.

This should rule out all but the simplest overlap structures, and in particular the nonstandard RSB and related pictures (for a full discussion, see Newman and Stein 1998b). It is therefore natural to ask whether the property of metastate invariance allows any many-state picture.

There is one such picture, namely the "chaotic pairs" picture, which is fully consistent with metastate invariance (our belief is that it is the only many-state picture that fits naturally and easily into results obtained about the metastate.)

Here the periodic boundary condition metastate is supported on infinitely many pairs of pure states, but instead of eqn [12] one has

$$\Gamma = (1/2)\rho_{\alpha_\Gamma} + (1/2)\rho_{-\alpha_\Gamma} \qquad [15]$$

with overlap

$$P_\Gamma = (1/2)\delta(q - q_{\mathrm{EA}}) + (1/2)\delta(q + q_{\mathrm{EA}}) \qquad [16]$$

So there is CSD in the states but not in the overlaps, which have the same form as a two-state picture in every volume. The difference is that, while in the latter case, one has the "same" pair of states in every volume, in chaotic pairs the pure-state pair varies chaotically as volume changes. If the chaotic pairs picture is to be consistent with metastate invariance in a natural way, then the number of pure-state pairs should be "uncountable." This allows for a "uniform" distribution (within the metastate) over all of the pure states, and invariance of the metastate with respect to boundary conditions could follow naturally.

## Open Questions

We have discussed how the metastate approach to the EA spin glass has narrowed considerably the set of possible scenarios for low-temperature ordering in any finite dimension, should broken spin-flip symmetry occur. The remaining possibilities are either a two-state scenario, such as droplet/scaling, or the chaotic-pairs picture if there exist many pure states at some $(\beta, d)$. Both have simple overlap structures. The metastate approach appears to rule out more complicated scenarios such as RSB, in which the approximate pure-state decomposition in a typical large, finite volume is a nontrivial mixture of many pure-state pairs.

Of course, this does not answer the question of which, if either, of the remaining pictures actually does occur in real spin glasses. In this section we list a number of open questions relevant to the above discussion.

**Open Question 1** Determine whether a phase transition occurs in any finite dimension greater than one. If it does, find the lower critical dimension.

Existence of a phase transition does not necessarily imply two or more pure states below $T_c$. It could happen, for example, that in some dimension there exists a single pure state at all nonzero temperatures, with two-point spin correlations decaying exponentially above $T_c$ and more slowly (e.g., as a power law) below $T_c$. This leads to:

**Open Question 2** If there does exist a phase transition above some lower critical dimension, determine whether the low-temperature spin-glass phase exhibits broken spin-flip symmetry.

If broken symmetry does occur in some dimension, then of course an obvious open question is to determine the number of pure-state pairs, and hence the nature of ordering at low temperature. A (possibly) easier question (but still very difficult), and one which does not rely on knowing whether a phase transition occurs, is to determine the zero-temperature – i.e., ground state – properties of spin glasses as a function of dimensionality. A ground state is an infinite-volume spin configuration whose energy (governed by eqn [1]) cannot be lowered by flipping any finite subset of spins. That is, all ground state spin configurations must satisfy the constraint

$$\sum_{\langle x,y\rangle \in \mathcal{C}} J_{xy}\sigma_x\sigma_y \geq 0 \qquad [17]$$

along any closed loop $\mathcal{C}$ in the dual lattice.

**Open Question 3** How many ground state pairs is the $T = 0$ periodic boundary condition metastate supported on, as a function of $d$?

The answer is known to be one for 1D, and a partial result (Newman and Stein 2000, 2001a) points towards the answer being one for 2D as well. There are no rigorous, or even heuristic (except based on underlying "ansätze") arguments in higher dimension.

An interesting – but unrealistic – spin-glass model in which the ground state structure can be exactly solved (although not yet completely rigorously) was proposed by the authors (Newman and Stein 1994, 1996a) (see also Banavar 1994). This "highly disordered" spin glass is one in which the coupling magnitudes scale nonlinearly with the volume (and so are no longer distributed independently of the volume, although they remain independent and identically distributed for each volume). The model

displays a transition in ground state multiplicity: below eight dimensions, it has only a single pair of ground states, while above eight it has uncountably many such pairs. The mechanism behind the transition arises from a mapping to invasion percolation and minimal spanning trees (Lenormand and Bories 1980, Chandler *et al*. 1982, Wilkinson and Willemsen 1983): the number of ground state pairs can be shown to equal $2^{\mathcal{N}}$, where $\mathcal{N} = \mathcal{N}(d)$ is the number of distinct global components in the "minimal spanning forest." The zero-temperature free boundary condition metastate above eight dimensions is supported on a uniform distribution (in a natural sense) on uncountably many ground state pairs.

Interestingly, the high-dimensional ground state multiplicity in this model can be shown to be unaffected by the presence of frustration, although frustration still plays an interesting role: it leads to the appearance of chaotic size dependence when free boundary conditions are used.

Returning to the more difficult problem of ground state multiplicity in the EA model, we note as a final remark that there could, in principle, exist ground state pairs that are not in the support of metastates generated through the use of coupling-independent boundary conditions. If such states exist, they may be of some interest mathematically, but are not expected to play any significant physical role. A discussion of these putative "invisible states" is given by Newman and Stein (2003).

**Open Question 4** If there exists broken spin-flip symmetry at a range of positive temperatures in some dimensions, then what is the number of pure-state pairs as a function of $(\beta, d)$?

Again, the answer to this is not known above one dimension; indeed, the prerequisite existence of spontaneously broken spin-flip symmetry has not been proved in any dimension. A speculative paper by the authors (Newman and Stein 2001b), using a variant of the highly disordered model, suggests that there is at most one pair of pure states in the EA model below eight dimensions; but no rigorous arguments are known at this time.

## Acknowledgments

This work was supported in part by NSF Grants DMS-01-02587 and DMS-01-02541.

*See also:* Glassy Disordered Systems: Dynamical Evolution; Mean Field Spin Glasses and Neural Networks; Spin Glasses.

## Further Reading

Aizenman M, Lebowitz JL, and Ruelle D (1987) Some rigorous results on the Sherrington–Kirkpatrick spin glass model. *Communications in Mathematical Physics* 112: 3–20.

Aizenman M and Wehr J (1990) Rounding effects of quenched randomness on first-order phase transitions. *Communications in Mathematical Physics* 130: 489–528.

Banavar JR, Cieplak M, and Maritan A (1994) Optimal paths and domain walls in the strong disorder limit. *Physical Review Letters* 72: 2320–2323.

Binder K and Young AP (1986) Spin glasses: experimental facts, theoretical concepts, and open questions. *Review of Modern Physics* 58: 801–976.

Bray AJ and Moore MA (1985) Critical behavior of the three-dimensional Ising spin glass. *Physical Review B* 31: 631–633.

Campanino M, Olivieri E, and van Enter ACD (1987) One dimensional spin glasses with potential decay $1/r^{1+\epsilon}$. Absence of phase transitions and cluster properties. *Communications in Mathematical Physics* 108: 241–255.

Chandler R, Koplick J, Lerman K, and Willemsen JF (1982) Capillary displacement and percolation in porous media. *Journal of Fluid Mechanics* 119: 249–267.

Chowdhury D (1986) *Spin Glasses and Other Frustrated Systems*. New York: Wiley.

Dotsenko V (2001) *Introduction to the Replica Theory of Disordered Statistical Systems*. Cambridge: Cambridge University Press.

Edwards S and Anderson PW (1975) Theory of spin glasses. *Journal of Physics F* 5: 965–974.

Fischer KH and Hertz JA (1991) *Spin Glasses*. Cambridge: Cambridge University Press.

Fisher DS and Huse DA (1986) Ordered phase of short-range Ising spin-glasses. *Physical Review Letters* 56: 1601–1604.

Fisher DS and Huse DA (1988) Equilibrium behavior of the spin-glass ordered phase. *Physical Review B* 38: 386–411.

Fisher ME and Singh RRP (1990) Critical points, large-dimensionality expansions and Ising spin glasses. In: Grimmett G and Welsh DJA (eds.) *Disorder in Physical Systems*, pp. 87–111. Oxford: Clarendon Press.

Franz S, Mézard M, Parisi G, and Peliti L (1998) Measuring equilibrium properties in aging systems. *Physical Review Letters* 81: 1758–1761.

Fröhlich J and Zegarlinski B (1987) The high-temperature phase of long-range spin glasses. *Communications in Mathematical Physics* 110: 121–155.

Gandolfi A, Newman CM, and Stein DL (1993) Exotic states in long-range spin glasses. *Communications in Mathematical Physics* 157: 371–387.

Georgii HO (1988) *Gibbs Measures and Phase Transitions*. Berlin: de Gruyter.

Kawashima N and Young AP (1996) Phase transition in the three-dimensional $\pm J$ Ising spin glass. *Physical Review B* 53: R484–R487.

Krzakala F and Martin OC (2000) Spin and link overlaps in three-dimensional spin glasses. *Physical Review Letters* 85: 3013–3016.

Lenormand R and Bories S (1980) Description d'un mécanisme de connexion de liaison destiné à l'étude du drainage avec piègeage en milieu poreux. *Comptes Rendus de l'Académie des Sciences* 291: 279–282.

Marinari E and Parisi G (2000) Effects of changing the boundary conditions on the ground state of Ising spin glasses. *Physical Review B* 62: 11677–11685.

Marinari E and Parisi G (2001) Effects of a bulk perturbation on the ground state of 3*D* Ising spin glasses. *Physical Review Letters* 86: 3887–3890.
Marinari E, Parisi G, Ricci-Tersenghi F, Ruiz-Lorenzo JJ, and Zuliani F (2000) Replica symmetry breaking in spin glasses: Theoretical foundations and numerical evidences. *Journal of Statistical Physics* 98: 973–1047.
Marinari E, Parisi G, and Ritort F (1994) On the 3*D* Ising spin glass. *Journal of Physics A* 27: 2687–2708.
Marinari E, Parisi G, and Ruiz-Lorenzo J (1997) Numerical simulations of spin glass systems. In: Young AP (ed.) *Spin Glasses and Random Fields*, pp. 59–98. Singapore: World Scientific.
McMillan WL (1984) Scaling theory of Ising spin glasses. *Journal of Physics C* 17: 3179–3187.
Mézard M, Parisi G, Sourlas N, Toulouse G, and Virasoro M (1984) Nature of spin-glass phase. *Physical Review Letters* 52: 1156–1159.
Mézard M, Parisi G, and Virasoro MA (eds.) (1987) *Spin Glass Theory and Beyond*. Singapore: World Scientific.
Newman CM and Stein DL (1992) Multiple states and thermodynamic limits in short-ranged Ising spin glass models. *Physical Review B* 46: 973–982.
Newman CM and Stein DL (1994) Spin-glass model with dimension-dependent ground state multiplicity. *Physical Review Letters* 72: 2286–2289.
Newman CM and Stein DL (1996a) Ground state structure in a highly disordered spin glass model. *Journal of Statistical Physics* 82: 1113–1132.
Newman CM and Stein DL (1996b) Non-mean-field behavior of realistic spin glasses. *Physical Review Letters* 76: 515–518.
Newman CM and Stein DL (1996c) Spatial inhomogeneity and thermodynamic chaos. *Physical Review Letters* 76: 4821–4824.
Newman CM and Stein DL (1998a) Simplicity of state and overlap structure in finite volume realistic spin glasses. *Physical Review E* 57: 1356–1366.
Newman CM and Stein DL (1998b) Thermodynamic chaos and the structure of short-range spin glasses. In: Bovier A and Picco P (eds.) *Mathematics of Spin Glasses and Neural Networks*, pp. 243–287. Boston: Birkhauser.
Newman CM and Stein DL (2000) Nature of ground state incongruence in two-dimensional spin glasses. *Physical Review Letters* 84: 3966–3969.
Newman CM and Stein DL (2001a) Are there incongruent ground states in 2*D* Edwards–Anderson spin glasses? *Communications in Mathematical Physics* 224: 205–218.
Newman CM and Stein DL (2001b) Realistic spin glasses below eight dimensions: a highly disordered view. *Physical Review E* 63: 16101-1–16101-9.
Newman CM and Stein DL (2003) Topical review: Ordering and broken symmetry in short-ranged spin glasses. *Journal of Physics: Condensed Matter* 15: R1319–R1364.
Ogielski AT (1985) Dynamics of three-dimensional spin glasses in thermal equilibrium. *Physical Review B* 32: 7384–7398.
Ogielski AT and Morgenstern I (1985) Critical behavior of the three-dimensional Ising spin-glass model. *Physical Review Letters* 54: 928–931.
Palassini M and Young AP (2000) Nature of the spin glass state. *Physical Review Letters* 85: 3017–3020.
Parisi G (1979) Infinite number of order parameters for spin-glasses. *Physical Review Letters* 43: 1754–1756.
Parisi G (1983) Order parameter for spin-glasses. *Physical Review Letters* 50: 1946–1948.
Sherrington D and Kirkpatrick S (1975) Solvable model of a spin glass. *Physical Review Letters* 35: 1792–1796.
Stein DL (1989) Disordered systems: Mostly spin glasses. In: Stein DL (ed.) *Lectures in the Sciences of Complexity*, pp. 301–355. New York: Addison–Wesley.
Thill MJ and Hilhorst HJ (1996) Theory of the critical state of low-dimensional spin glass. *Journal of Physics I* 6: 67–95.
van Enter ACD and Fröhlich J (1985) Absence of symmetry breaking for *N*-vector spin glass models in two dimensions. *Communications in Mathematical Physics* 98: 425–432.
Wilkinson D and Willemsen JF (1983) Invasion percolation: A new form of percolation theory. *Journal of Physics A* 16: 3365–3376.

# Sine-Gordon Equation

**S N M Ruijsenaars**, Centre for Mathematics and Computer Science, Amsterdam, The Netherlands



## Introduction

The sine-Gordon equation

$$\left(\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial t^2}\right)\phi = \sin \phi \qquad [1]$$

may be viewed as a prototype for a nonlinear integrable field theory. It is manifestly invariant under spacetime translations and Lorentz boosts,

$$\begin{aligned}(x,t) &\mapsto (x-\alpha, t-\beta)\\ (x,t) &\mapsto (x\cosh\theta - t\sinh\theta, t\cosh\theta - x\sinh\theta)\end{aligned} \qquad [2]$$

It shares this relativistic invariance property with the linear Klein–Gordon equation, which is obtained upon replacing $\sin\phi$ by $\phi$. (The name sine-Gordon equation is derived from this relation, and was introduced by Kruskal.) The sine-Gordon equation can also be defined and studied in the form

$$\frac{\partial^2}{\partial u\,\partial v}\tilde{\phi} = \sin\tilde{\phi}, \quad \tilde{\phi}(u,v) = \phi(t,x) \qquad [3]$$

where

$$u = (x+t)/2, \quad v = (x-t)/2 \qquad [4]$$

are the so-called light-cone variables.

There are two interpretations of the field $\phi(t,x)$ that are quite different, both from a physical and from a mathematical viewpoint. The first one consists in viewing it as a real-valued function, so

that [1] is simply a nonlinear PDE in two variables. In the second version, one views $\phi(t,x)$ as an operator-valued distribution on a Hilbert space. (Thus, one should smear $\phi(t,x)$ with a test function $f(t,x)$ in Schwartz space to obtain a genuine operator on the Hilbert space.) In spite of their different character, the classical and quantum field theory versions have several striking features in common, including the presence of an infinite number of conservation laws and the occurrence of solitonic excitations.

The classical sine-Gordon equation has been used as a model for various wave phenomena, including the propagation of dislocations in crystals, phase differences across Josephson junctions, torsion waves in strings and pendula, and waves along lipid membranes. It was already studied in the nineteenth century in connection with the theory of pseudospherical surfaces. The quantum version is used as a simple model for solid-state excitations.

The designation "sine-Gordon" is also used for various equations that generalize [1] or bear resemblance to it. These include the so-called homogeneous and symmetric space sine-Gordon models, discrete and supersymmetric versions, and generalizations to higher-dimensional spacetimes (i.e., in [1] the spatial derivative is replaced by the Laplace operator in several variables). In this contribution we focus on [1], however.

Our main goal is to discuss the integrability and solitonic properties, both at the classical and at the quantum level. First, we sketch the inverse-scattering transform (IST) solution to the Cauchy problem for [1]. Following Faddeev and Takhtajan, we emphasize the interpretation of the IST as an action-angle transformation for an infinite-dimensional Hamiltonian system. Next, the particle-like solutions are surveyed by using a description in terms of variables that may be viewed as relativistic action-angle coordinates. This is followed by a section on the quantum field theory version, paying special attention to the factorized scattering that is the quantum analog of the solitonic classical scattering. Finally, we sketch the intimate relation between the $N$-particle subspaces of the classical and quantum sine-Gordon field theory and certain integrable relativistic systems of $N$ point particles on the line.

## The Classical Version: An Integrable Hamiltonian System

In order to tie in the hyperbolic evolution equation [1] with the notion of infinite-dimensional integrable system, it is necessary to restrict attention to initial data $\phi(0,x)=\phi(x)$ and $\partial_t\phi(0,x)=\pi(x)$ with special properties. First of all, the energy functional

$$H=\int_{-\infty}^{\infty}\left(\frac{1}{2}\pi(x)^2+\frac{1}{2}\partial_x\phi(x)^2+(1-\cos\phi(x))\right)\mathrm{d}x \qquad [5]$$

and symplectic form

$$\omega=\int_{-\infty}^{\infty}\mathrm{d}\pi(x)\wedge\mathrm{d}\phi(x)\,\mathrm{d}x \qquad [6]$$

should be well defined on the phase space of initial data. Indeed, in that case [1] amounts to the Hamilton equation associated to [5] via [6].

Second, there exists a sequence of functionals

$$I_{2l+1}(\phi,\pi),\quad l\in\mathbb{Z} \qquad [7]$$

that formally Poisson-commute with $H$ and among themselves.

In particular, $H$ equals $2(I_1+I_{-1})$, whereas $2(I_1-I_{-1})$ equals the momentum functional

$$P=-\int_{-\infty}^{\infty}\pi(x)\partial_x\phi(x)\,\mathrm{d}x \qquad [8]$$

The functional $I_{2l+1}$ contains $x$-derivatives of order up to $|2l+1|$, so one needs to require that the functions $\partial_x\phi(x)$ and $\pi(x)$ be smooth and that all of their derivatives have sufficient decrease for $x\to\pm\infty$.

A natural choice guaranteeing the latter requirements is

$$\partial_x\phi(x),\pi(x)\in S_{\mathbb{R}}(\mathbb{R}) \qquad [9]$$

where $S_{\mathbb{R}}(\mathbb{R})$ denotes the Schwartz space of real-valued functions on the line. To render the integral over $1-\cos\phi(x)$ (and similar integrals occurring for the sequence [7]) finite, one also needs to require

$$\phi(x)\to 2\pi k_{\pm},\quad x\to\pm\infty,\ k_{\pm}\in\mathbb{Z} \qquad [10]$$

On this phase space $\Omega$ of initial data, the Cauchy problem for the evolution equation [1] is not only well posed, but can be solved in explicit form by using the IST. More generally, the Hamiltonians $I_{2l+1}$ give rise to evolution equations that are simultaneously solved via the IST, yielding an infinite sequence of commuting Hamiltonian flows on $\Omega$.

Before sketching the overall picture resulting from the IST, it should be mentioned at this point that [1] admits explicit solutions of interest that do not belong to $\Omega$. First, there is a class of algebro-geometric solutions that have no limits as $x\to\pm\infty$. These solutions can be obtained via finite-gap integration methods, yielding formulas involving

the Riemann theta functions associated to compact Riemann surfaces. Second, there are the tachyon solutions. They arise from the particle-like solutions that do belong to $\Omega$ by the transformation

$$\phi(t,x) \to \phi(x,t) + \pi \qquad [11]$$

(Observe that the equation of motion [1] is invariant under [11], whereas due to the finite-energy requirement [10] this is not true for solutions evolving in $\Omega$.)

The IST via which the above Cauchy problem can be solved starts from an auxiliary system of two linear ordinary differential equations involving $\phi(0,x)$ and $\partial_t\phi(0,x)$. It is beyond our scope to describe the system in detail. The results derived from it, however, are to a large extent the same as those obtained via a simpler auxiliary linear operator that is associated to the light-cone Cauchy problem. The latter operator is of the Ablowitz–Kaup–Newell–Segur (AKNS) form. That is, the linear operator is an ordinary differential operator of Dirac type given by

$$L = \begin{pmatrix} \mathrm{i}\frac{\mathrm{d}}{\mathrm{d}x} & -\mathrm{i}q \\ \mathrm{i}r & -\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}x} \end{pmatrix} \qquad [12]$$

where the external potentials $r(u)$ and $q(u)$ depend on the evolution equation at hand. For the light-cone sine-Gordon equation [3], one needs to choose

$$r = -q = (\partial_u\tilde{\phi})(u,0)/2 \qquad [13]$$

In both settings, the associated spectral features are invariant under the sine-Gordon evolution and all of the evolutions generated by the Hamiltonians $I_{2l+1}$, yielding the so-called isospectral flows. More specifically, if the initial data give rise to bound-state solutions of the linear problem (square-integrable wave functions), then the corresponding eigenvalues are time independent. Furthermore, due to the decay requirements on the potential in the linear system, there exist scattering solutions with plane-wave asymptotics for all initial data in $\Omega$. A suitable normalization leads to the so-called Jost solutions $\Psi(x,\lambda)$. (Here $\lambda$ is the spectral parameter, which varies over the real line for scattering solutions.) Their $x \to \pm\infty$ asymptotics is encoded in transition coefficients $a(\lambda)$ and $b(\lambda)$, with $a(\lambda)$ and $|b(\lambda)|$ being time independent, whereas $\arg b(\lambda)$ has a linear dependence on time when the potential evolves according to the sine-Gordon equation. The bound states correspond to special $\lambda$-values $\lambda_1, \ldots, \lambda_N$ with positive imaginary part (namely the zeros of the coefficient $a(\lambda)$, which is analytic in the upper-half $\lambda$-plane); their normalization coefficients $\nu_1, \ldots, \nu_N$ have an essentially linear time evolution, just as $b(\lambda)$.

The crux of the IST is now that the potentials can be reconstructed from the spectral data

$$\{b(\lambda), \lambda_1, \ldots, \lambda_N, \nu_1, \ldots, \nu_N\} \qquad [14]$$

by solving a linear integral equation of Gelfand–Levitan–Marchenko (GLM) type. (Alternatively, Riemann–Hilbert problem techniques can be used.) Hence, the nonlinear Cauchy problem can be replaced by the far simpler linear problems of determining the spectral data [14] of a linear operator (the direct problem) and then solving the linear GLM equation for the time-evolved scattering data (the inverse problem).

From the Hamiltonian perspective, the IST may be reinterpreted as a transformation to action-angle variables. The action variables are defined in terms of $|b(\lambda)|$ and $\lambda_1, \ldots, \lambda_N$. They are time independent under the sine-Gordon and higher Hamiltonian flows. The angle variables are $\arg b(\lambda)$ and suitable functions of the normalization coefficients. They depend linearly on the evolution times of the flows. The Hamiltonians can be explicitly expressed in action variables.

Next, we point out that there is a large subspace of Cauchy data $(\phi(x), \pi(x))$ that do not give rise to bound states in the auxiliary linear problem. The associated solutions are the so-called radiation solutions: they decrease to 0 for large times. These solutions can be obtained from the inverse transform involving the GLM equation by only taking $b(\lambda)$ into account.

The other extreme is to choose $b(\lambda) = 0$ and arbitrary bound states and normalization coefficients in the GLM equation. This special case of vanishing reflection leads to the particle-like solutions that are studied in the next section. For general Cauchy data, one has both $b(\lambda) \neq 0$ and a finite number of bound states. These so-called mixed solutions have a radiation component (encoded in $b(\lambda)$) which decays for asymptotic times, whereas the bound states show up for $t \to \pm\infty$ as isolated solitons, antisolitons, and breathers.

## Classical Solitons, Antisolitons, and Breathers

Just as for other classical soliton equations, the case of reflectionless data can be handled in complete detail, since the GLM equation reduces to an $N \times N$ system of linear equations. The case $N = 1$ yields the 1-soliton and 1-antisoliton solutions. Resting at the origin, these one-particle solutions are given by

$$\pm 4 \arctan(\mathrm{e}^{-x}) \qquad [15]$$

and have energy 8 (cf. [5]). (We normalize all solutions by requiring

$$\lim_{x\to\infty}\phi(t,x)=0 \qquad [16]$$

Note that one can add arbitrary multiples of $2\pi$ without changing the energy $H$ [5].) A spatial translation and Lorentz boost then yields the general solutions

$$\phi_{\pm}(t,x) = \pm 4\arctan(\exp(q - x\cosh\theta + t\sinh\theta)) \qquad [17]$$

with energy $8\cosh\theta$ and momentum $8\sinh\theta$ (cf. [8]).

Defining the topological charge of a solution (with normalization [16]) by

$$Q=\frac{1}{2\pi}\lim_{x\to-\infty}\phi(t,x) \qquad [18]$$

the different charges $Q=1$ and $Q=-1$ of the soliton and antisoliton reflect a signature associated to the special value of the spectral parameter on the imaginary axis for which a bound state in the linear problem occurs. More generally, for bound-state eigenvalues on the imaginary axis these signatures must be specified in the IST setting, a point glossed over in the previous section.

Bound states in the linear problem can also arise from $\lambda$-values off the imaginary axis, which come in pairs $\mathrm{i}a\pm b$, with $a,b>0$. Such pairs give rise to solutions containing breathers, which can be viewed as bound states of a soliton and an antisoliton. The one-breather solution breathing at the origin is given by

$$4\arctan\left(\cot\eta\frac{\sin(t\sin\eta)}{\cosh(x\cos\eta)}\right), \quad \eta\in(0,\pi/2) \qquad [19]$$

and has energy $16\cos\eta$. A spacetime translation and Lorentz boost then yields the general solution

$$\phi_b(t,x) = 4\arctan\left(\cot\eta\frac{\sin[-\gamma/2+\sin\eta(t\cosh\theta-x\sinh\theta)]}{\cosh[y/2-\cos\eta(x\cosh\theta-t\sinh\theta)]}\right) \qquad [20]$$

which has energy $16\cosh\theta\cos\eta$ and momentum $16\sinh\theta\cos\eta$. It may be obtained by analytic continuation from the solution describing a collision between a soliton with velocity $\tanh\theta_1$ and an antisoliton with velocity $\tanh\theta_2$, taking $\theta_2<\theta_1$. The latter is given by

$$\phi_{+-}(t,x) = 4\arctan\left(\coth((\theta_1-\theta_2)/2)\frac{\sinh((\mu_1-\mu_2)/2)}{\cosh((\mu_1+\mu_2)/2)}\right) \quad \theta_2<\theta_1 \qquad [21]$$

where

$$\mu_j = q_j - x\cosh\theta_j + t\sinh\theta_j, \quad q_j,\theta_j\in\mathbb{R} \qquad [22]$$

and $\phi_b$ results from $\phi_{+-}$ by substituting

$$\theta_{\substack{1\\2}}\to\theta\pm\mathrm{i}\eta, \qquad q_{\substack{1\\2}}\to(y\mp\mathrm{i}\gamma)/2 \qquad [23]$$

(For the case $\theta_1<\theta_2$, one needs an extra minus sign on the right-hand side of [21].)

There is yet another possibility for an eigenvalue on the imaginary axis we have not mentioned thus far: it may have an arbitrary multiplicity, giving rise to the so-called multipole solutions. This is illustrated by the breather solution $\phi_b$: when one sets $\gamma=-2q_0\eta$ and lets $\eta$ tend to 0, one obtains a solution

$$\phi_{\mathrm{sep}}(t,x) = 4\arctan\left(\frac{q_0+t\cosh\theta-x\sinh\theta}{\cosh[y/2-x\cosh\theta+t\sinh\theta]}\right) \qquad [24]$$

From a physical viewpoint, the soliton and antisoliton have just enough energy to prevent a bound state from being formed. Notice that in this case the distance between soliton and antisoliton diverges logarithmically in $|t|$ as $t\to\pm\infty$, whereas for $\phi_{+-}$ one obtains linear increase.

The 2-soliton and 2-antisoliton solutions can also be obtained by analytic continuation of $\phi_{+-}$. They read

$$\phi_{\pm\pm} = \mp 4\arctan\left(\coth((\theta_1-\theta_2)/2) \times\frac{\cosh((\mu_1-\mu_2)/2)}{\sinh((\mu_1+\mu_2)/2)}\right), \quad \theta_2<\theta_1 \qquad [25]$$

where $\mu_j$ is given by [22]. Thus, they arise by taking $q_2\to q_2+\mathrm{i}\pi$ and $q_1\to q_1+\mathrm{i}\pi$ in [21], resp. The equal-signature eigenvalues corresponding to these two solutions cannot collide and move off the imaginary axis; physically speaking, equal-charge particles repel each other. The energy and momentum of the solutions [25] and [21] are given by $8\cosh\theta_1+8\cosh\theta_2$ and $8\sinh\theta_1+8\sinh\theta_2$, respectively.

Up to scale factors, the above variables $\theta_1,\theta_2$ and $\theta,\eta$ are the action variables resulting from the IST, whereas $q_1,q_2$ and $y,\gamma$ are the canonically conjugated angle variables. Accordingly, the time and space translation flows (generated by $H$ [5] and $P$ [8], resp.) shift the angles linearly in the evolution parameters $t$ and $x$.

We conclude this section with a description of the $N$-soliton solution and its large time asymptotics. It can be expressed in terms of the $N\times N$ matrix

$$\mathcal{L}_{jk} = \exp(\mu_j)\frac{\prod_{l\neq j}|\coth((\theta_j-\theta_l)/2)|}{\cosh((\theta_j-\theta_k)/2)} \qquad [26]$$

where $\mu_j$ is given by [22] with

$$q_1, \ldots, q_N \in \mathbb{R}, \qquad \theta_N < \cdots < \theta_1 \qquad [27]$$

Specifically, one has

$$\begin{aligned}\phi_{N+}(t,x) &= 4\,\mathrm{tr}\arctan(\mathcal{L}) \\ &= -2\mathrm{i}\,\ln(|\mathbf{1}_N + \mathrm{i}\mathcal{L}|/|\mathbf{1}_N - \mathrm{i}\mathcal{L}|) \\ &= -2\mathrm{i}\,\ln\left(\left(1 + \sum_{l=1}^{N} \mathrm{i}^l S_l(\mathcal{L})\right)/\mathrm{c.c.}\right)\end{aligned} \qquad [28]$$

where $S_l$ is the $l$th symmetric function of $\mathcal{L}$. Using Cauchy's identity, one obtains the explicit formula

$$S_l = \sum_{\substack{I\subset\{1,\ldots,N\}\\|I|=l}} \exp\left(\sum_{j\in I}\mu_j\right) \prod_{\substack{j\in I\\k\notin I}} |\coth((\theta_j - \theta_k)/2)| \qquad [29]$$

In order to specify the $t \to \pm\infty$ asymptotics of $\phi_{N+}$, we introduce the 1-soliton solutions

$$\phi_j^{\pm}(t,x) = 4\,\arctan(\exp(\mu_j \mp \Delta_j/2)) \qquad [30]$$

where

$$\Delta_j = \left(\sum_{k<j} - \sum_{k>j}\right)\delta(\theta_j - \theta_k) \qquad [31]$$

$$\delta(\theta) = \ln\left(\coth(\theta/2)^2\right) \qquad [32]$$

Then, one has

$$\sup_{x\in\mathbb{R}}\left|\phi_{N+}(t,x) - \sum_{j=1}^{N}\phi_j^{\pm}(t,x)\right| = O(\exp(-|t|r)) \qquad t \to \pm\infty \qquad [33]$$

where the decay rate is given by

$$r = \min_{j\neq k}(\cosh(\theta_j)|\tanh\theta_j - \tanh\theta_k|) \qquad [34]$$

Thus, the soliton profile with velocity $\tanh\theta_j$ incurs a shift $\Delta_j/\cosh\theta_j$ as a result of the collision. The factor $1/\cosh\theta_j$ may be viewed as a Lorentz contraction factor.

## The Quantum Version: A Soliton Quantum Field Theory

From a perturbation-theoretic viewpoint, the quantum sine-Gordon Hamiltonian is given by

$$H = \int_{-\infty}^{\infty} : \left(\frac{1}{2}(\partial_t\phi)^2 + \frac{1}{2}(\partial_x\phi)^2 + \frac{m^2}{\beta^2}(1 - \cos\beta\phi)\right) : \mathrm{d}x, \quad m, \beta > 0 \qquad [35]$$

Here, $\phi(0,x)$ is a neutral Klein–Gordon field with mass $m$ and the double dots denote a suitable ordering prescription. The associated equation of motion

$$\phi_{xx} - \phi_{tt} = \frac{m^2}{\beta}\sin\beta\phi \qquad [36]$$

is equivalent to [1] on the classical level, but not on the quantum level. (If $\phi(t,x)$ is a classical solution to [36], then $\beta\phi(t/m, x/m)$ solves [1].) This difference is due to the extremely singular character of interacting relativistic quantum field theory, a context in which "solving" the field theory has slowly acquired a meaning that is vastly different from the classical notion. Indeed, one can at best hope to verify [36] in the sense of expectation values in suitable quantum states, and this is precisely what has been achieved within the form-factor program sketched later on.

From the perspective of functional analysis, the existence of a well-defined Wightman field theory with all of the features mentioned below is wide open. More precisely, beginning with pioneering work by Fröhlich some 30 years ago, various authors have contributed to a mathematically rigorous construction of a sine-Gordon quantum field theory version, but to date it seems not feasible to verify that the resulting Wightman field theory has any of the explicit features we are going to sketch. (For example, not even the free character of the field theory for $\beta^2 = 4\pi$ has been established; cf. below.)

That said, we proceed to sketch some highlights of the impressive, but partly heuristic lore that has been assembled in a great many theoretical physics papers. A key result we begin with is the equivalence to a field theory that looks very different at face value. This is the massive Thirring model, formally given by the Hamiltonian

$$H_{\mathrm{T}} = \int_{-\infty}^{\infty} : \left(\Psi^*(-\mathrm{i}\gamma^5\partial_x + \gamma^0 M)\Psi + \frac{g}{2}(J_0^2 - J_1^2)\right) : \mathrm{d}x \qquad M \in (0,\infty),\ g \in \mathbb{R} \qquad [37]$$

Here, $\Psi(0,x)$ is the charged Dirac field with mass $M$ and the double dots stand for normal ordering. For the $\gamma$-algebra, one may choose

$$\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \gamma^1 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$
$$\gamma^5 = \gamma^0\gamma^1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad [38]$$

and $J_\mu$ is the Dirac current,

$$J_0 = \Psi^*\Psi, \qquad J_1 = \Psi^*\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\Psi \qquad [39]$$

The equivalence argument (due to Coleman) consists in showing that the quantities

$$-\frac{\beta}{2\pi}\epsilon_{\mu\nu}\partial^{\nu}\phi, \qquad \frac{m^2}{\beta^2} : \cos\beta\phi : \qquad [40]$$

in the sine-Gordon theory have the same vacuum expectation values (in perturbation theory) as the massive Thirring quantities

$$: J_{\mu} :, \qquad -M : \Psi^{*}\gamma^{0}\Psi : \qquad [41]$$

resp., provided the parameters are related by

$$\frac{4\pi}{\beta^2} = 1 + \frac{g}{\pi} \qquad [42]$$

This yields an equivalence between the charge-0 sector of the massive Thirring model and the sector of the sine-Gordon theory obtained by the action of the fields [40] on the vacuum vector. But the charged sectors of the Thirring model can also be viewed as new sectors in the sine-Gordon theory, obtained by a solitonic field construction (first performed by Mandelstam).

In this picture, the fermions and antifermions in the massive Thirring model correspond to new excitations in the sine-Gordon theory, the quantum solitons and antisolitons. The latter are viewed as coherent states of the sine-Gordon "mesons" in the vacuum sector, the rest masses being related by

$$M = \frac{8m}{\beta^2}\left(1 - \frac{\beta^2}{8\pi}\right) \qquad [43]$$

in the semiclassical limit $\beta^2 \to 0$.

Even at the formal level involved in the correspondence, the theories are not believed to exist for $\beta^2 > 8\pi$ and $g < -\pi/2$, since there is positivity breakdown for this range of couplings. The free Dirac case $g = 0$ corresponds to $\beta^2 = 4\pi$. In particular, there is no interaction between the sine-Gordon solitons and antisolitons for this $\beta$-value. In the range $\beta^2 \in (4\pi, 8\pi)$ there is interaction, but bound soliton–antisoliton pairs (quantum breathers, alias sine-Gordon mesons) do not occur.

By contrast, for $\beta^2 < 4\pi$ there exist breathers with rest masses

$$m_n = 2M\sin(n+1)\alpha, \quad \alpha \equiv m/2M,$$
$$n+1 = 1, 2, \ldots, L < \pi/2\alpha \qquad [44]$$

Thus, the "particle spectrum" consists of solitons and antisolitons with mass $M$ and mesons $C_1, \ldots, C_L$ with masses $m_1, \ldots, m_L$ given by [44]. The latter formula was first established by semiclassical quantization of the classical breathers (Dashen–Hasslacher–Neveu), and ever since is usually called the DHN formula. Notice that for $\alpha$ near zero $m_1$ and $m$ are nearly equal, and that for $\beta^2 \geq 4\pi$ there are no longer any sine-Gordon mesons present in the theory.

*A priori*, the existence of infinitely many classical conserved Hamiltonians does not even formally imply the same feature for the quantum field theory, as anomalies may occur. For the sine-Gordon and massive Thirring cases, anomalies have been shown to be absent, however. This entails not only that the number of solitons, antisolitons, and breathers in a scattering process is conserved, but also that the set of incoming rapidities equals the set of outgoing rapidities.

The latter stability features and the DHN formula [44] are corroborated by the $S$-matrix, which is known in complete detail. The two-body amplitudes involving solitons and antisolitons can be written in terms of the function

$$u(z) = \exp\left(\mathrm{i}\int_0^{\infty}\frac{\mathrm{d}x}{x}\frac{\sinh(\alpha - \pi/2)x}{\sinh\alpha x\cosh\pi x/2}\sin 2xz\right) \qquad [45]$$

They are given by

$$(u_{++}, t_{+-}, r_{+-}, u_{--})(\theta) = u(\theta/2)\left(1, \frac{\sinh(\pi\theta/2\alpha)}{\sinh(\pi(\mathrm{i}\pi - \theta)/2\alpha)}, \frac{\mathrm{i}\sin(\pi^2/2\alpha)}{\sinh(\pi(\mathrm{i}\pi - \theta)/2\alpha)}, 1\right) \qquad [46]$$

where $\theta$ denotes the rapidity difference. (Due to fermion statistics, one gets only one amplitude for a soliton or antisoliton pair. But a soliton and an antisoliton have opposite charge, so they can be distinguished. In that case, therefore, the notion of reflection and transmission coefficients makes sense.)

The $S$-matrix involving an arbitrary number of solitons, antisolitons, and their bound states is also explicitly known. The amplitudes involving no breathers are readily described in terms of the above two-body amplitudes. Indeed, the $S$-matrix factorizes as a sum of products of the amplitudes [46], yielding a picture of particles scattering independently in pairs, just as at the classical level. The factorization can be performed irrespective of the temporal ordering assumed for the pair scattering processes, since the four functions occurring inside the parentheses of [46] satisfy the Yang–Baxter equations.

Roughly speaking, the $S$-matrix for processes involving breathers can be calculated by analytic continuation from the soliton–antisoliton $S$-matrix. The details are however quite substantial. We only add that scattering amplitudes involving solely breathers can be expressed using only hyperbolic functions.

Since the 1980s, a lot of information has also been gathered concerning matrix elements of suitable sine-Gordon field quantities between special quantum states (form factors). Unfortunately, the correlation functions involve infinite sums of form factors that are quite difficult to control analytically. Hence, it is not known whether the correlation functions associated with the form factors give rise to a Wightman field theory with the usual axiomatic properties.

## The Relation to Relativistic Calogero–Moser Systems

The behavior of the special classical solutions discussed earlier is very similar to that of classical point particles. Furthermore, the picture of classical solitons, antisolitons, and their bound states scattering independently in pairs is essentially preserved on the quantum level, just as one would expect for the quantization of an integrable particle system.

Next, we note that from the quantum viewpoint there is no physical distinction between wave functions and point particles, whereas a classical wave is a physical entity that is clearly very different from a point particle. Even so, it is a natural question whether there exist classical Hamiltonian systems of $N$ point particles on the line whose physical characteristics (charges, bound states, scattering, etc.) are the same as those of the particle-like sine-Gordon solutions. If so, a second question is equally obvious: does the quantum version of the $N$-particle systems still have the same features as that of the quantum sine-Gordon excitations?

As we now sketch, the first question has been answered in the affirmative, whereas the second one has not been completely answered yet. However, all of the information on the pertinent quantum $N$-particle systems collected thus far points to an affirmative answer. The systems at issue are relativistic versions of the well-known nonrelativistic Calogero–Moser $N$-particle systems.

To begin with the classical two-particle system, its Hamiltonian is given by

$$H = (\cosh p_1 + \cosh p_2)\coth((x_1 - x_2)/2) \qquad [47]$$

on the phase space

$$\Omega = \{(x, p) \in \mathbb{R}^4 | x_2 < x_1\}, \quad \omega = \mathrm{d}x \wedge \mathrm{d}p \qquad [48]$$

Taking $x_2 \to x_2 + \mathrm{i}\pi$ yields the particle–antiparticle Hamiltonian

$$\tilde{H} = (\cosh p_1 + \cosh p_2)|\tanh((x_1 - x_2)/2)| \qquad [49]$$

on the phase space

$$\tilde{\Omega} = \{(x, p) \in \mathbb{R}^4\}, \quad \omega = \mathrm{d}x \wedge \mathrm{d}p \qquad [50]$$

The two-antiparticle Hamiltonian is again given by [47] and [48]. The interaction potential in [47] is repulsive, whereas it is attractive in [49]. Hence, any initial point in $\Omega$ gives rise to a scattering state, whereas points in $\tilde{\Omega}$ yield scattering states if and only if the reduced Hamiltonian

$$\tilde{H}_{\mathrm{r}} = \cosh p|\tanh(x/2)|, \quad p = (p_1 - p_2)/2$$
$$x = x_1 - x_2 \qquad [51]$$

satisfies $\tilde{H}_{\mathrm{r}} > 1$. More specifically, in both cases the distance $|x_1(t) - x_2(t)|$ increases linearly as $t \to \pm\infty$, the scattering (position shift) being encoded by the same function [32] as for the sine-Gordon solitons. The phase-space points on the separatrix $\{\tilde{H}_{\mathrm{r}} = 1\}$ have the same temporal asymptotics as the multipole solution [24], whereas the bound-state oscillations for $\tilde{H}_{\mathrm{r}} < 1$ match those of the breathers [20].

More generally, the Hamiltonian for $N_+$ particles and $N_-$ antiparticles is given by the function

$$\sum_{j=1}^{N_+} \cosh(p_j^+) \prod_{\substack{k=1 \\ k \neq j}}^{N_+} |\coth((x_j^+ - x_k^+)/2)|$$
$$\times \prod_{l=1}^{N_-} |\tanh((x_j^+ - x_l^-)/2)| + \sum_{l=1}^{N_-} \cosh(p_l^-)$$
$$\times \prod_{\substack{m=1 \\ m \neq l}}^{N_-} |\coth((x_l^- - x_m^-)/2)|$$
$$\times \prod_{j=1}^{N_+} |\tanh((x_l^- - x_j^+)/2)| \qquad [52]$$

on the phase space

$$\Omega_{N_+,N_-} = \Big\{(x^+, p^+) \in \mathbb{R}^{2N_+}, (x^-, p^-) \in \mathbb{R}^{2N_-} | x_{N_+}^+ < \cdots < x_1^+, x_{N_-}^- < \cdots < x_1^- \Big\} \qquad [53]$$

$$\omega_{N_+,N_-} = \mathrm{d}x^+ \wedge \mathrm{d}p^+ + \mathrm{d}x^- \wedge \mathrm{d}p^- \qquad [54]$$

This defining Hamiltonian can be supplemented by $(N_+ + N_- - 1)$ independent Hamiltonians that pairwise commute. The action-angle map of this integrable system can be used to relate the scattering and bound-state behavior to that of the sine-Gordon solutions from an earlier section, yielding an exact correspondence. Indeed, the variables we used to describe the particle-like sine-Gordon solutions amount to the action-angle variables associated to [52]. Moreover, the matrix $\mathcal{L}$ [26] with $t = x = 0$

equals the Lax matrix for the $N$-particle system, which is the manifestation of a remarkable self-duality property of the equal-charge case. There is an equally close relation between the general particle-like solutions and the general systems encoded in [52].

As a matter of fact, the connection can be further strengthened by introducing spacetime trajectories for the solitons, antisolitons, and breathers, which are defined in terms of the evolution of an initial point in $\Omega_{N_+,N_-}$ under the time translation generator [52] and the space translation generator, obtained from [52] by the replacement $\cosh \to \sinh$. These point particle and antiparticle trajectories make it possible to follow the motion of the solitons, antisolitons, and breathers during the temporal interval in which the nonlinear interaction takes place, whereas for large times the trajectories are located at the (then) clearly discernible positions of the individual solitons, antisolitons, and breathers.

Before sketching the soliton-particle correspondence at the quantum level, we add a remark on the finite-gap solutions of the classical sine-Gordon equation, already mentioned in the paragraph containing [11]. These solutions may be viewed as generalizations of the particle-like solutions discussed earlier, and they can also be obtained via relativistic $N$-particle Calogero–Moser systems. The pertinent systems are generalizations of the hyperbolic systems just described to the elliptic level.

Turning now to the quantum level, we begin by mentioning that the Poisson-commuting Hamiltonians admit a quantization in terms of commuting analytic difference operators. This involves a special ordering choice of the $p$-dependent and $x$-dependent factors in the classical Hamiltonians, which is required to preserve commutativity. The resulting quantum two-body problem can be explicitly solved in terms of a generalization of the Gauss hypergeometric function. For the case of equal charges, the scattering is encoded in the sine-Gordon amplitudes $u_{\pm\pm}(\theta)$ (cf. [45] and [46]). For the unequal-charge case, one should distinguish an even and odd channel. The scattering on these channels is encoded in the sine-Gordon amplitudes $t_{+-}(\theta) \pm r_{+-}(\theta)$. Moreover, the bound-state spectrum agrees with the DHN formula [44] and the bound-state wave functions are given by hyperbolic functions.

As a consequence of these results, the physics encoded in the two-body subspace of the sine-Gordon quantum field theory is indistinguishable from that of the corresponding two-body relativistic Calogero–Moser systems. To extend this equivalence to the arbitrary-$N$ case, one needs first of all sufficiently explicit solutions to the $N$-body Schrödinger equation. To date, this has only been achieved for the case of $N$ equal charges and the special couplings for which the reflection amplitude $r_{+-}$ vanishes. The asymptotics of the pertinent solutions is factorized in terms of $u_{\pm\pm}(\theta)$, in agreement with the sine-Gordon picture.

*See also:* Bäcklund Transformations; Boundary-Value Problems for Integrable Equations; Calogero–Moser–Sutherland Systems of Nonrelativistic and Relativistic Type; Infinite-dimensional Hamiltonian Systems; Integrability and Quantum Field Theory; Integrable Systems and Discrete Geometry; Integrable Systems and Inverse Scattering Method; Integrable Systems: Overview; Ljusternik–Schnirelman Theory; Solitons and Other Extended Field Configurations; Solitons and Kac–Moody Lie Algebras; Symmetries and Conservation Laws; Two-Dimensional Models; Yang–Baxter Equations.

## Further Reading


Ablowitz MJ, Kaup DJ, Newell AC, and Segur H (1974) The inverse scattering transform – Fourier analysis for nonlinear problems. *Studies in Applied Mathematics* 53: 249–315.

Coleman S (1977) Classical lumps and their quantum descendants. In: Zichichi A (ed.) *New Phenomena in Subnuclear Physics, Proceedings Erice 1975*, pp. 297–421. New York: Plenum.

Faddeev LD and Takhtajan LA (1987) *Hamiltonian Methods in the Theory of Solitons*. Berlin: Springer.

Flaschka HF and Newell AC (1975) Integrable systems of nonlinear evolution equations. In: Moser J (ed.) *Dynamical Systems, Theory and Applications*, Lecture Notes in Physics, vol. 38, pp. 355–440. Berlin: Springer.

Karowski M (1979) Exact $S$-matrices and form factors in 1+1 dimensional field theoretic models with soliton behaviour. *Physics Reports* 49: 229–237.

Ruijsenaars SNM (2001) Sine-Gordon solitons vs. Calogero–Moser particles. In: Pakuliak S and von Gehlen G (eds.) *Proceedings of the Kiev NATO Advanced Study Institute Integrable Structures of Exactly Solvable Two-Dimensional Models of Quantum Field Theory*, NATO Science Series, vol. 35, pp. 273–292. Dordrecht: Kluwer.

Scott AC, Chu FYF, and McLaughlin DW (1973) The soliton: a new concept in applied science. *Proceedings of the Institute of Electrical and Electronics Engineers* 61: 1443–1483.

Smirnov FA (1992) Form Factors in Completely Integrable Models of Quantum Field Theory. *Advanced Series in Mathematical Physics*, vol. 14. Singapore: World Scientific.

Thacker HB (1981) Exact integrability in quantum field theory and statistical systems. *Reviews of Modern Physics* 53: 253–285.

Zamolodchikov AB and Zamolodchikov AlB (1979) Factorized $S$-matrices in two dimensions as the exact solutions of certain relativistic quantum field theory models. *Annals of Physics (NY)* 120: 253–291.

# Singularities of the Ricci Flow

**M Anderson**, State University of New York at Stony Brook, Stony Brook, NY, USA



## Introduction

Fix a closed $n$-dimensional manifold $M$, and let $\mathbb{M}$ be the space of Riemannian metrics on $M$. As in the reasoning leading to the Einstein equations in general relativity, there is basically a unique simple and natural vector field on the space $\mathbb{M}$. Namely, the tangent space $T_g\mathbb{M}$ consists of symmetric bilinear forms; besides multiples of the metric itself, the Ricci curvature $\mathrm{Ric}_g$ of $g$ is the only symmetric form that depends on at most the second derivatives of the metric, and is invariant under coordinate changes, that is, a (0, 2)-tensor formed from the metric. Thus, consider

$$X_g = \mu \mathrm{Ric}_g + \lambda g$$

where $\mu, \lambda$ are scalars. Setting $\mu = -2$, the corresponding equation for the flow of $X$ is

$$\frac{\mathrm{d}}{\mathrm{d}t} g(t) = -2\mathrm{Ric}_{g(t)} + \lambda g(t) \qquad [1]$$

The Ricci flow, introduced by Hamilton (1982), is obtained by setting $\lambda = 0$:

$$\frac{\mathrm{d}}{\mathrm{d}t} g(t) = -2\mathrm{Ric}_{g(t)} \qquad [2]$$

Rescaling the metric and time variable $t$ transforms [2] into [1], with $\lambda = \lambda(t)$. For example, rescaling the Ricci flow [2] so that the volume of $(M, g(t))$ is preserved leads to the flow equation [1] with $\lambda = 2 \oint R$, twice the mean value of the scalar curvature $R$.

The Ricci flow [2] bears some relation with the metric part of the $\beta$-function or renormalization group (RG) flow equation

$$\frac{\mathrm{d}}{\mathrm{d}t} g(t) = \beta(g(t))$$

for the two-dimensional sigma model of maps $\Sigma^2 \to M$. The $\beta$-function is a vector field on $\mathbb{M}$, invariant under diffeomorphisms, which has an expansion of the form

$$-\beta(g) = \mathrm{Ric}_g + \varepsilon \mathrm{Riem}^2 + \cdots$$

where $\mathrm{Riem}^2$ is quadratic in the Riemann curvature tensor. The Ricci flow corresponds to the one-loop term or semiclassical limit in the RG flow (cf. D'Hoker (1999) and Friedan (1985)).

Recently, G Perelman (2002, 2003a, b) has developed new insights into the geometry of the Ricci flow which has led to a solution of long-standing mathematical conjectures on the structure of 3-manifolds, namely the Thurston geometrization conjecture (Thurston 1982), and hence the Poincaré conjecture.

## Basic Properties of the Ricci Flow

In charts where the coordinate functions are locally defined harmonic functions in the metric $g(t)$, [2] takes the form

$$\frac{\mathrm{d}}{\mathrm{d}t} g_{ij} = \Delta g_{ij} + Q_{ij}(g, \partial g)$$

where $\Delta$ is the Laplace operator on functions with respect to the metric $g = g(t)$ and $Q$ is a lower-order term quadratic in $g$ and its first-order partial derivatives. This is a nonlinear heat-type equation for $g_{ij}$ and leads to the existence and uniqueness of solutions to the Ricci flow on some time interval starting at any smooth initial metric. This is the reason for the minus sign in [2]; a plus sign gives a backwards heat-type equation, which has no solutions in general.

The flow [2] gives a natural method to try to construct canonical metrics on the manifold $M$. Stationary points of the flow [2] are Ricci-flat metrics, while stationary points of the flow [1] are (Riemannian) Einstein metrics, where $\mathrm{Ric}_g = (R/n)g$, with $R$ the scalar curvature of $g$. One of Hamilton's motivations for studying the Ricci flow were results on an analogous question for nonlinear sigma models. Consider maps $f$ between Riemannian manifolds $M$, $N$ with Lagrangian given by the Dirichlet energy. Eells–Sampson studied the heat equation for this action and proved that when the target $N$ has nonpositive curvature, the flow exists for all time and converges to a stationary point of the action, that is, a harmonic map $f_\infty : M \to N$. The idea is to see if an analogous program can be developed on the space of metrics $\mathbb{M}$.

There are a number of well-known obstructions to the existence of Einstein metrics on manifolds, in particular, in dimensions 3 and 4. Thus, the Ricci flow will not exist for all time on a general manifold. Hence, it must develop singularities. A fundamental issue is to try to relate the structure of the singularities of the flow with the topology of the underlying manifold $M$.

A few simple qualitative features of the Ricci flow [2] are as follows: if $\mathrm{Ric}(x, t) > 0$, then the flow contracts the metric $g(t)$ near $x$, to the future, while

if $\text{Ric}(x,t) < 0$, then the flow expands $g(t)$ near $x$. At a general point, there will be directions of positive and negative Ricci curvature, along which the metric locally contracts or expands. The flow preserves product structures of metrics, and preserves the isometry group of the initial metric.

The form of [2] shows that the Ricci flow continues as long as Ricci curvature remains bounded. On a bounded time interval where $\text{Ric}_{g(t)}$ is bounded, the metrics $g(t)$ are quasi-isometric, that is, they have bounded distortion compared with the initial metric $g(0)$. Thus, one needs to consider evolution equations for the curvature, induced by the flow for the metric. The simplest of these is the evolution equation for the scalar curvature $R$:

$$\frac{\mathrm{d}}{\mathrm{d}t}R = \Delta R + 2|\text{Ric}|^2 \qquad [3]$$

Evaluating [3] at a point realizing the minimum $R_{\min}$ of $R$ on $M$ shows that $R_{\min}$ is monotone nondecreasing along the flow. In particular, the Ricci flow preserves positive scalar curvature. Moreover, if $R_{\min}(0) > 0$, then

$$t \leq \frac{n}{2R_{\min}(0)} \qquad [4]$$

Thus, the Ricci flow exists only up to a maximal time $T \leq n/2R_{\min}(0)$ when $R_{\min}(0) > 0$. In contrast, in regions where the Ricci curvature stays negative definite, the flow exists for infinite time.

The evolution of the Ricci curvature has the same general form as [3]:

$$\frac{\mathrm{d}}{\mathrm{d}t}R_{ij} = \Delta R_{ij} + \widetilde{Q}_{ij} \qquad [5]$$

The expression for $\widetilde{Q}$ is much more complicated than the Ricci curvature term in [3] but involves only quadratic expressions in the curvature. However, $\widetilde{Q}$ involves the full Riemann curvature tensor Riem of $g$, and not just the Ricci curvature (as [3] involves Ricci and not just scalar curvature). An important feature of dimension 3 is that the full Riemann curvature Riem is determined algebraically by the Ricci curvature. So the Ricci flow has a much better chance of "working" in dimension 3. For example, an analysis of $\widetilde{Q}$ shows that the Ricci flow preserves positive Ricci curvature in dimension 3; if $\text{Ric}_{g(0)} > 0$, then $\text{Ric}_{g(t)} > 0$, for $t > 0$. This is not the case in higher dimensions. On the other hand, in any dimension $>2$, the Ricci flow does not preserve negative Ricci curvature, or even a general lower bound $\text{Ric} \geq -\lambda$, for $\lambda > 0$. For the remainder of the article, we usually assume then that $\dim M = 3$.

The first basic result on the Ricci flow is the following, due to Hamilton (1982).

*Space-form theorem*. If $g(0)$ is a metric of positive Ricci curvature on a 3-manifold $M$, then the volume normalized Ricci flow exists for all time, and converges to the round metric on $S^3/\Gamma$, where $\Gamma$ is a finite subgroup of SO(4), acting freely on $S^3$.

Thus, the Ricci flow "geometrizes" 3-manifolds of positive Ricci curvature. There are two further important structural results on the Ricci flow.

*Curvature pinching estimate* (Hamilton 1982, Ivey 1993). For $g(t)$ a solution to the Ricci flow on a closed 3-manifold $M$, there is a nonincreasing function $\phi : (-\infty, \infty) \to \mathbb{R}$, tending to 0 at $\infty$, and a constant $C$, depending only on $g(0)$, such that,

$$\text{Riem}(x,t) \geq -C - \phi(R(x,t)) \cdot |R(x,t)| \qquad [6]$$

This estimate does not imply a lower bound on $\text{Riem}(x,t)$ uniform in time. However, when combined with the fact that the scalar curvature $R(x,t)$ is uniformly bounded below (cf. [3]), it implies that $|\text{Riem}|(x,t) \gg 1$ only where $R(x,t) \gg 1$. To control the size of $|\text{Riem}|$, it thus suffices to obtain just an upper bound on $R$. This is remarkable, since the scalar curvature is a much weaker invariant of the metric than the full curvature. Moreover, at points where the curvature is sufficiently large, [6] shows that $\text{Riem}(x,t)/R(x,t) \geq -\delta$, for $\delta$ small. Thus, if one scales the metric to make $R(x,t) = 1$, then $\text{Riem}(x,t) \geq -\delta$. In such a scale, the metric then has almost non-negative curvature near $(x,t)$.

*Harnack estimate* (Hamilton 1982). Let $(N, g(t))$ be a solution to the Ricci flow with bounded and non-negative curvature $\text{Riem} \geq 0$, and suppose $g(t)$ is a complete Riemannian metric on $N$. Then for $0 < t_1 \leq t_2$,

$$R(x_2,t_2) \geq \frac{t_1}{t_2}\exp\left(-\frac{d_{t_1}^2(x_1,x_2)}{2(t_2-t_1)}\right)R(x_1,t_1) \qquad [7]$$

where $d_{t_1}$ is the distance function on $(M, g_{t_1})$. This allows one to control the geometry of the solution at different spacetime points, given control at an initial point.

## Singularity Formation

The deeper analysis of the Ricci flow is concerned with the singularities that arise in finite time. Equation [3] shows that the Ricci flow will not exist for arbitrarily long time in general. In the case of initial metrics with positive Ricci curvature, this is resolved by rescaling the Ricci flow to constant volume. However, the general situation is necessarily much more complicated. For example, any manifold which is a connected sum of $S^3/\Gamma$ or $S^2 \times S^1$ factors has metrics of positive scalar curvature. For obvious

topological reasons, the volume normalized Ricci flow then cannot converge nicely to a round metric; even the renormalized flow must develop singularities.

The usual method to understand the structure of singularities, particularly in geometric PDEs, is to rescale or renormalize the solution on a sequence converging to the singularity to make the solution bounded, and try to pass to a limit of the renormalization. Such a limit solution models the singularity formation, and one hopes (or expects) that the singularity models have special features making them much simpler than an arbitrary solution of the flow.

A singularity forms for the Ricci flow only where the curvature becomes unbounded. Suppose then that $\lambda_i^2 = |\mathrm{Riem}|(x_i, t_i) \to \infty$, on a sequence of points $x_i \in M$, and times $t_i \to T < \infty$. Consider the rescaled or blow-up metrics and times

$$\bar{g}_i(\bar{t}_i) = \lambda_i^2 \phi_i^* g(t), \quad \bar{t}_i = \lambda_i^2 (t - t_i) \qquad [8]$$

where $\phi_i$ are diffeomorphisms giving local dilations of the manifold near $x_i$ by the factor $\lambda_i$.

The flow $\bar{g}_i$ is also a solution of the Ricci flow, and has bounded curvature at $(x_i, 0)$. For suitable choices of $x_i$ and $t_i$, the curvature will be bounded near $x_i$, and for nearby times to the past, $\bar{t}_i \le 0$; for example, one might choose points $(x_i, t_i)$ where the curvature is maximal on $(M, g(t)), t \le t_i$.

The rescaling [8] expands all distances by the factor $\lambda_i$, and time by the factor $\lambda_i^2$. Thus, in effect one is studying very small regions, of spatial size on the order of $r_i = \lambda_i^{-1}$ about $(x_i, t_i)$, and "using a microscope" to examine the small-scale features in this region on a scale of size about 1.

A limit solution of the Ricci flow, defined at least locally in space and time, will exist provided that the local volumes of the rescalings are bounded below (Gromov compactness). In terms of the original unscaled flow, this requires that the metric $g(t)$ should not be locally collapsed on the scale of its curvature, that is,

$$\mathrm{vol}\, B_{x_i}(r_i, t_i) \ge \nu r_i^n \qquad [9]$$

for some fixed but arbitrary $\nu > 0$. A maximal connected limit $(N, \bar{g}(\bar{t}), x)$ containing the base point $x = \lim x_i$, is then called a "singularity model." Observe that the topology of the limit $N$ may well be distinct from the original manifold $M$, most of which may have been blown off to infinity in the rescaling.

To see the potential usefulness of this process, suppose one does have local noncollapse on the scale of the curvature, and that base points of maximal curvature in space and time $t \le t_i$ have been chosen. At least in a subsequence, one then obtains a limit solution to the Ricci flow $(N, \bar{g}(\bar{t}), x)$, based at $x$, defined at least for times $(-\infty, 0]$, with $\bar{g}(\bar{t})$ a complete Riemannian metric on $N$. Such solutions are called ancient solutions of the Ricci flow. The estimate [6] shows that the limit has non-negative curvature in dimension 3, and so [7] holds on $N$. Thus, the limit is indeed quite special. The topology of complete manifolds $N$ of non-negative curvature is completely understood in dimension 3. If $N$ is noncompact, then $N$ is diffeomorphic to $\mathbb{R}^3, S^2 \times \mathbb{R}$, or a quotient of these spaces. If $N$ is compact, then a slightly stronger form of the space-form theorem implies $N$ is diffeomorphic to $S^3/\Gamma, S^2 \times S^1$, or $S^2 \times_{\mathbb{Z}_2} S^1$.

The study of the formation of singularities in the Ricci flow was initiated by Hamilton (1995). Recently, Perelman has obtained an essentially complete understanding of the singularity behavior of the Ricci flow, at least in dimension 3.

## Perelman's Work

### Noncollapse

Consider the Einstein–Hilbert action

$$\mathcal{R}(g) = \int_M R(g)\, \mathrm{d}V_g \qquad [10]$$

as a functional on $\mathbb{M}$. Critical points of $\mathcal{R}$ are Ricci-flat metrics. It is natural and tempting to try to relate the Ricci flow with the gradient flow of $\mathcal{R}$ (with respect to a natural $L^2$ metric on the space $\mathbb{M}$). However, it has long been recognized that this cannot be done directly. In fact, the gradient flow of $\mathcal{R}$ does not even exist, since it implies a backwards heat-type equation for the scalar curvature $R$ (similar to [3] but with a minus sign before $\Delta$).

Consider however the following functional extending $\mathcal{R}$:

$$\mathcal{F}(g, f) = \int_M (R + |\nabla f|^2) \mathrm{e}^{-f}\, \mathrm{d}V_g \qquad [11]$$

as a functional on the larger space $\mathbb{M} \times C^\infty(M, \mathbb{R})$, or equivalently a family of functionals on $\mathbb{M}$, parametrized by $C^\infty(M, \mathbb{R})$. The functional [11] also arises in string theory as the low-energy effective action; the scalar field $f$ is called the dilaton. Fix any smooth measure $\mathrm{d}m$ on $M$ and define the Perelman coupling by requiring that $(g, f)$ satisfy

$$\mathrm{e}^{-f}\, \mathrm{d}V_g = \mathrm{d}m \qquad [12]$$

The resulting functional

$$\mathcal{F}^m(g, f) = \int_M (R + |\nabla f|^2)\, \mathrm{d}m \qquad [13]$$

becomes a functional on $\mathbb{M}$. (This coupling does not appear to have been considered in string theory.) The $L^2$ gradient flow of $\mathcal{F}^m$ is given simply by

$$\frac{\mathrm{d}\widetilde{g}}{\mathrm{d}t} = -2(\mathrm{Ric}_{\widetilde{g}} + \widetilde{D}^2 f) \qquad [14]$$

where $\widetilde{D}^2 f$ is the Hessian of $f$ with respect to $\widetilde{g}$. The evolution equation [14] for $\widetilde{g}$ is just the Ricci flow [2] modified by an infinitesimal diffeomorphism: $\widetilde{D}^2 f = (\mathrm{d}/\mathrm{d}t)(\phi_t^* \widetilde{g})$, where $(\mathrm{d}/\mathrm{d}t)\phi_t = \widetilde{\nabla} f$. Thus, the gradient flow of $\mathcal{F}^m$ is the Ricci flow, up to diffeomorphisms. The evolution equation for the scalar field $f$,

$$f_t = -\widetilde{\Delta} f - \widetilde{R} \qquad [15]$$

is a backward heat equation (balancing the forward evolution of the volume form of $\widetilde{g}(t)$). Thus, this flow will not exist for general $f$, going forward in $t$. However, one of the basic points of view is to let the (pure) Ricci flow [2] flow for a time $t_0 > 0$. At $t_0$, one may then take an arbitrary $f = f(t_0)$ and flow this $f$ backward in time $(\tau = t_0 - t)$ to obtain an initial value $f(0)$ for $f$. The choice of $f(t_0)$ determines, together with the choice of volume form of $g(0)$), (or $g(t_0)$), the measure $\mathrm{d}m$ and so the choice of $\mathcal{F}^m$. The process of passing from $\mathcal{F}$ to $\mathcal{F}^m$ corresponds to a reduction of the symmetry group of all diffeomorphisms $\mathcal{D}$ of $\mathcal{F}$ to the group $\mathcal{D}_0$ of volume-preserving diffeomorphisms; the quotient space $\mathcal{D}/\mathcal{D}_0$ has been decoupled into a space $C^\infty(M, \mathbb{R})$ of parameters.

The functionals $\mathcal{F}^m$ are not scale invariant. To achieve scale invariance, Perelman includes an explicit insertion of the scale parameter, related to time, by setting

$$\mathcal{W}(g, f, \tau) = \int \Big(\tau(|\nabla f|^2 + R) + f - n\Big) \times (4\pi\tau)^{-n/2}\, \mathrm{e}^{-f}\, \mathrm{d}V \qquad [16]$$

with coupling so that $\mathrm{d}m = (4\pi\tau)^{-n/2}\, \mathrm{e}^{-f}\, \mathrm{d}V$ is fixed. The entropy functional $\mathcal{W}$ is invariant under simultaneous rescaling of $\tau$ and $g$, and $\tau_t = -1$. Again, the gradient flow of $\mathcal{W}$ is the Ricci flow modulo diffeomorphisms and rescalings and the stationary points of the gradient flow are the gradient Ricci solitons,

$$\mathrm{Ric}_g + D^2 f - \frac{1}{2\tau} g = 0$$

for which the metrics evolve by diffeomorphisms and rescalings. Gradient solitons arise naturally as singularity models, due to the rescalings and diffeomorphisms in the blow-up procedure [8]. An important example is the cigar soliton on $\mathbb{R}^2 \times \mathbb{R}$, (or $\mathbb{R}^2 \times S^1$),

$$g = (1 + r^2)^{-1} g_{\mathrm{Eucl}} + \mathrm{d}s^2 \qquad [17]$$

Perelman then uses the scalar field $f$ to probe the geometry of $g(t)$. For instance, the collapse or noncollapse of the metric $g(t)$ near a point $x \in M$ can be detected from the size of $\mathcal{W}(g(t))$ by choosing $\mathrm{e}^{-f}$ to be an approximation to a delta function centered at $(x, t)$. The more collapsed $g(t)$ is near $x$, the more negative the value of $\mathcal{W}(g(t))$. The collapse of the metric $g(t)$ on any scale in finite time is then ruled out by combining this with the fact that the entropy functional $\mathcal{W}$ is increasing along the Ricci flow.

Much more detailed information can be obtained by studying the path integral associated to the evolution equation [15] for $f$, given by

$$\mathcal{L}(\gamma) = \int_\gamma \sqrt{\tau}[|\dot{\gamma}(\tau)|^2 + R(\gamma(\tau))]\, \mathrm{d}\tau$$

where $R$ and $|\dot{\gamma}(\tau)|$ are computed with respect to the evolving metrics $g(\tau)$. In particular, the study of the geodesics and the associated variational theory of the length functional $\mathcal{L}$ are important in understanding the geometry of the Ricci flow near the singularities.

## Singularity Models

A major accomplishment of Perelman is essentially a classification of all complete singularity models $(N, g(t))$ that arise in finite time. In the simple case where $N$ is compact, then as noted above, $N$ is diffeomorphic to $S^3/\Gamma$, $S^2 \times S^1$, or $S^2 \times_{\mathbb{Z}_2} S^1$.

In the much more important case where $N$ is complete and noncompact, Perelman proves that the geometry of $N$ near infinity is that of a union of $\varepsilon$-necks. Thus, at time 0, and at points $x$ with $r(x) = \mathrm{dist}(x, x_0) \gg 1$, for a fixed base point $x_0$, a region of radius $\varepsilon^{-1}$ about $x$, in the scale where $R(x) = 1$, is $\varepsilon$-close to such a region in the standard round product metric on $S^2 \times \mathbb{R}$; $\varepsilon$ may be made arbitrarily small by choosing $r(x)$ sufficiently large. For example, this shows that the cigar soliton [17] cannot arise as a singularity model. Moreover, this structure also holds on a time interval on the order of $\varepsilon^{-1}$ to the past, so that on such regions the solution is close to the (backwards) evolving Ricci flow on $S^2 \times \mathbb{R}$.

Perelman shows that this structural result for the singularity models themselves also holds for the solution $g(t)$ very near any singularity time $T$. Thus, at any base point $(x, t)$ where the curvature is sufficiently large, the rescaling as in [8] of the spacetime by the curvature is smoothly close, on large compact domains, to corresponding large domains in a complete singularity model. The "ideal" complete singularity models do actually describe the geometry and topology near any singularity. Consequently, one has a detailed understanding of the small-scale geometry and topology in a neighborhood of every point where the curvature is large on $(M, g(t))$, for $t$ near $T$.

The main consequence of this analysis is the existence of canonical, almost round 2-spheres $S^2$ in any region of $(M, g(t))$ where the curvature is sufficiently large; the radius of the $S^2$'s is on the order of the curvature radius. One then disconnects the manifold $M$ into pieces, by cutting $M$ along a judicious choice of such 2-spheres, and gluing in round 3-balls in a natural way. This surgery process allows one to excise out the regions of $(M, g(t))$ where the Ricci flow is almost singular, and thus leads to a naturally defined Ricci flow with surgery, valid for all times $t \in [0, \infty)$.

The surgery process disconnects the original connected 3-manifold $M$ into a collection of disjoint (connected) 3-manifolds $M_i$, with the Ricci flow running on each. However, topologically, there is a canonical relation between $M$ and the components $M_i$; $M$ is the connected sum of $\{M_i\}$. An analysis of the long-time behavior of the volume-normalized Ricci flow confirms the expectation that the flow approaches a fixed point, that is, an Einstein metric, or collapses along 3-manifolds admitting an $S^1$ fibration. This then leads to the proof of Thurston's geometrization conjecture for 3-manifolds and consequently the proof of the Poincaré conjecture. It gives a full classification of all closed 3-manifolds, much like the classification of surfaces given by the classical uniformization theorem.

*See also:* Einstein Manifolds; Evolution Equations: Linear and Nonlinear; Minimal Submanifolds; Renormalization: General Theory; Topological Sigma Models.

## Further Reading


Anderson M (1997) Scalar curvature and geometrization conjectures for 3-manifolds. In: Grove K and Petersen P (eds.) *Comparison Geometry*, MSRI Publications, vol. 30, pp. 49–82. Cambridge: Cambridge University Press.

Cao HT, Chow B, Cheng SC, and Yau ST (eds.) (2003) *Collected Papers on Ricci Flow*. Boston: International Press.

D'Hoker E (1999) String theory. In: Deligne P *et al.* (eds.) *Quantum Fields and Strings: A Course for Mathematicians*, vol. 2, Providence, RI: American Mathematical Society.

Friedan D (1985) Nonlinear models in $2 + \varepsilon$ dimensions. *Annals of Physics* 163: 318–419.

Hamilton R (1982) Three manifolds of positive Ricci curvature. *Journal of Differential Geometry* 17: 255–306.

Hamilton R (1993) The Harnack estimate for the Ricci flow. *Journal of Differential Geometry* 37: 225–243.

Hamilton R (1995) Formation of singularities in the Ricci flow. In: Yau ST (ed.) *Surveys in Differential Geometry* vol. 2, pp. 7–136. Boston, MA: International Press.

Ivey T (1993) Ricci solitons on compact three-manifolds. *Differential Geometry and Its Applications* 3: 301–307.

Perelman G (2002) *The entropy formula for the Ricci flow and its geometric applications*, math.DG/0211159.

Perelman G (2003a) *Ricci flow with surgery on three-manifolds*, math.DG/0303109.

Perelman G (2003b) *Finite extinction time for the solutions to the Ricci flow on certain three-manifolds*, math.DG/0307245.

Thurston W (1982) Three dimensional manifolds, Kleinian groups and hyperbolic geometry. *Bulletin of American Mathematical Society* 6: 357–381.


# Singularity and Bifurcation Theory

**J-P Françoise and C Piquet**, Université P.-M. Curie, Paris VI, Paris, France



## Introduction

Dynamical systems first developed from the geometry of Newton's equations (see Goodstein and Goodstein (1997)) and the question of the stability of the solar system motivated further researches inspired by celestial mechanics (cf. Siegel and Moser (1971)). Then dynamical systems developed intensively from stability theory (Lyapunov's theory) to generic properties (based on functional analysis techniques,) hyperbolic structures (Anosov's flows, Smale axiom A) and to perturbation theory (Pugh's closing lemma, KAM theorem). There are many links with ergodic theory dating back to Birkhoff's ergodic theorem (motivated by Boltzmann–Gibbs contributions to thermodynamics). These aspects have been developed in several articles of the encyclopedia (*see* Generic Properties of

Dynamical Systems; Ergodic Theory; Hyperbolic Dynamical Systems). This article develops another aspect of dynamical systems, namely bifurcation theory. In contrast, the mathematics involved relates more to local analytic geometry in the broad sense and provides local models like normal forms, uses blow-up techniques and asymptotic developments. This contains the singularity theory of functions (related to singularities of gradient flows). A recent development of the whole subject deals with bifurcation theory of fast-slow systems.

## Singularity Theory of Functions

A singular point of a gradient dynamics

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \operatorname{grad} V(x)$$

is a critical point of the function $V$. Assume that the function $V: U \to R$ is defined and infinitely differentiable on an open set $U$. Let $x_0 \in U$ be a critical point of $V$.

**Definition 1** The critical point $x_0$ is said to be of Morse type if the Hessian of $V$ at $x_0 : D_x^2 V(x_0)$ is of maximal rank $n$. The corank of a singular point $x_0$ is the corank of the matrix $D_x^2 V(x_0)$.

Denote by O the local ring of germs of $C^\infty$ functions at point $x_0$.

**Definition 2** The Jacobian ideal of the function $V$ at $x_0$, denoted as Jac(V), is the ideal generated in the ring $O$ by the partial derivatives of $V: \partial V/\partial x_i, i = 1, \ldots, n$, considered as elements of the local ring $O$.

The singularity (or the singular point) is isolated if

$$\dim_R O/\mathrm{Jac}(V) < \infty$$

In that case, the Milnor number is defined as the dimension

$$\mu = \dim_R O/\mathrm{Jac}(V)$$

Local models of singularities at a point are simple expressions that germs of functions singular at this point have in local coordinates.

R Thom proposed to focus more particularly on the singularities whose Milnor number is less than or equal to 4 and whose corank is less than or equal to 2.

The list of local models $V_\lambda(x)$ of functions whose singularities at 0 display a Milnor number less than or equal to 4 and a corank less than or equal to 2 is the following:

$V_\lambda(x) = \frac{1}{3}x^3 + \lambda_1 x$, the *fold*,
$V_\lambda(x) = \frac{1}{4}x^4 + \frac{1}{2}\lambda_1 x^2 + \lambda_2 x$, the *cusp*,
$V_\lambda(x) = \frac{1}{5}x^5 + \frac{1}{3}\lambda_1 x^3 + \frac{1}{2}\lambda_2 x^2 + \lambda_3 x$, the *swallow tail*,
$V_\lambda(x) = \frac{1}{6}x^6 + \frac{1}{4}\lambda_1 x^4 + \frac{1}{3}\lambda_2 x^3 + \frac{1}{2}\lambda_3 x^2 + \lambda_4 x$, the *butterfly*,
$V_\lambda(x) = x^3 - 3xy^2 + \lambda_1(x^2 + y^2) + \lambda_2 x + \lambda_3 y$, the *elliptic umbilic*,
$V_\lambda(x) = x^3 + y^3 + \lambda_1 xy + \lambda_2 x + \lambda_3 y$, the *hyperbolic umbilic*, and
$V_\lambda(x) = y^4 + x^2 y + \lambda_1 x^2 + \lambda_2 y^2 + \lambda_3 x + \lambda_4 y$, the *parabolic umbilic*.

Consider more particularly the first four cases. The "state equation" defines the critical points of $V_\lambda$:

$$\frac{\partial V_\lambda}{\partial x} = 0$$

which contains the subset of the stable equilibrium points of the associated gradient dynamics. The nature of these equilibrium states changes at points contained in the set defined by the equation

$$\frac{\partial^2 V_\lambda}{\partial x^2} = 0$$

The projection of this set on the space of parameters contains the set of values of the parameters for which the equilibrium position is susceptible to change of topological type (in other terms to undergo a bifurcation). This set is called the catastrophe set (see Figure 1).

Consider now the case of umbilics where there are two state equations:

$$\frac{\partial V}{\partial x} = \frac{\partial V}{\partial y} = 0$$

The catastrophe set $S$ is determined by one further equation:

$$\text{Hess } V = \frac{\partial^2 V}{\partial x^2}\frac{\partial^2 V}{\partial y^2} - \left(\frac{\partial^2 V}{\partial x \partial y}\right)^2 = 0$$

In both cases of hyperbolic and elliptic umbilics, the set $S$ is a singular surface. For the last case of the parabolic umbilic, the set $S$ is of dimension 3 and again it is only possible to represent it by a family of its sections by a variable hyperplane (see Figure 2).

All possible deformations (in the space of functions) of a function with an isolated singularity can be induced by a single $\mu$-dimensional family of deformations named the "universal deformation." In general, the "codimension" of a bifurcation is the minimal number of parameters needed to display all possible phase diagrams of all possible unfoldings. Several deep mathematical techniques, like the Malgrange division theorem and preparation theorem, allowed J Mather to prove the theorem (local, then global) of existence of the universal unfolding.

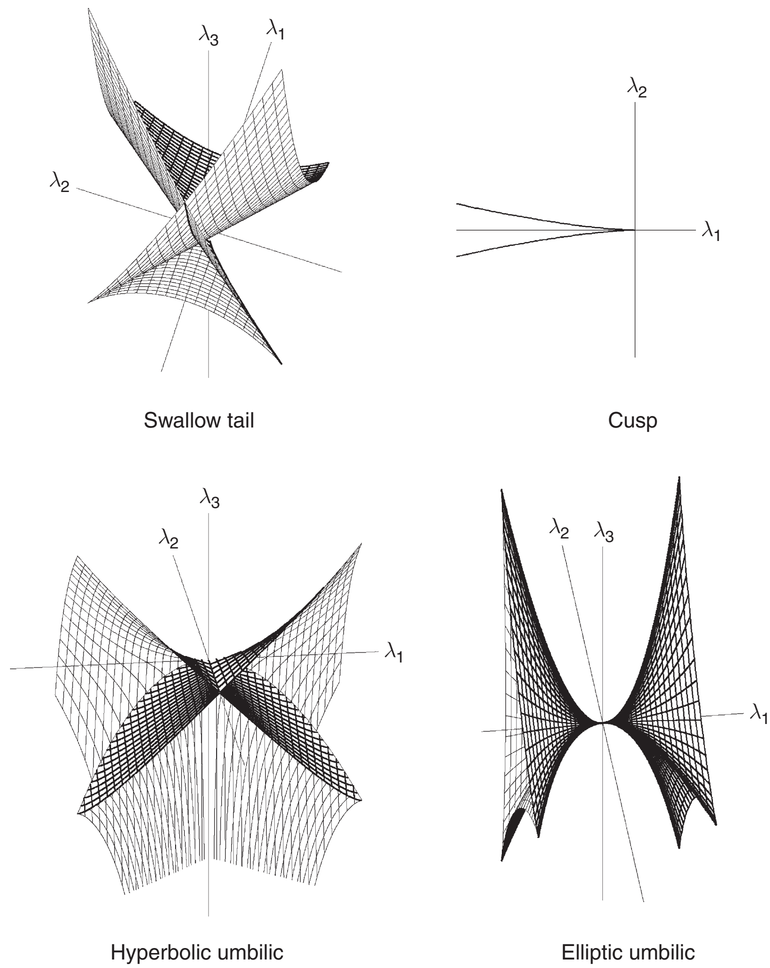



**Figure 1** Examples of catastrophe sets. Adapted with permission from Françoise J-P (2005) *Oscillations en Biologie: Analyse Qualitative et Modèles* (Mathématiques et Applications, vol. 46). Heidelberg: Springer.

The theory of unfoldings of singularities can be used, for instance, to provide asymptotic expression of stationary phase integrals when critical points of the phase are not of Morse type. This relates to monodromy, Bernstein polynomials, Milnor fibration near a singular point, and simultaneous local models of forms and functions (cf. Malgrange (1974)) and *see* Feynman Path Integrals).

## Singularity Theory of Vector Fields

### Transcritical Bifurcation

The transcritical bifurcation is the standard mechanism for changes in stability. The local model is given by

$$\dot{x} = rx - x^2$$

For $r < 0$, there is an unstable fixed point at $x^* = r$ and a stable fixed point at $x^* = 0$. As $r$ increases, the unstable and the stable fixed points coalesce when $r = 0$ and when $r > 0$, they exchange their stability. A simplified model of the essential physics of a laser is due to Haken (1983). It is given by

$$\dot{n} = GnN - kn$$

were $n$ is the number of photons in the laser field, $N$ is the number of excited atoms, and the gain term comes from the process of stimulated emission which occurs at a rate proportional to the product $n.N$. Furthermore, the number of excited atoms drops down by the emission of photons $N = N_0 - \alpha n$. Then we obtain

$$\dot{n} = (GN_0 - k)n - \alpha G n^2$$

This model displays a transcritical bifurcation, which explains in elementary terms the laser threshold.

### Pitchfork Bifurcation

The local model for supercritical pitchfork bifurcation is

$$\dot{x} = rx - x^3$$

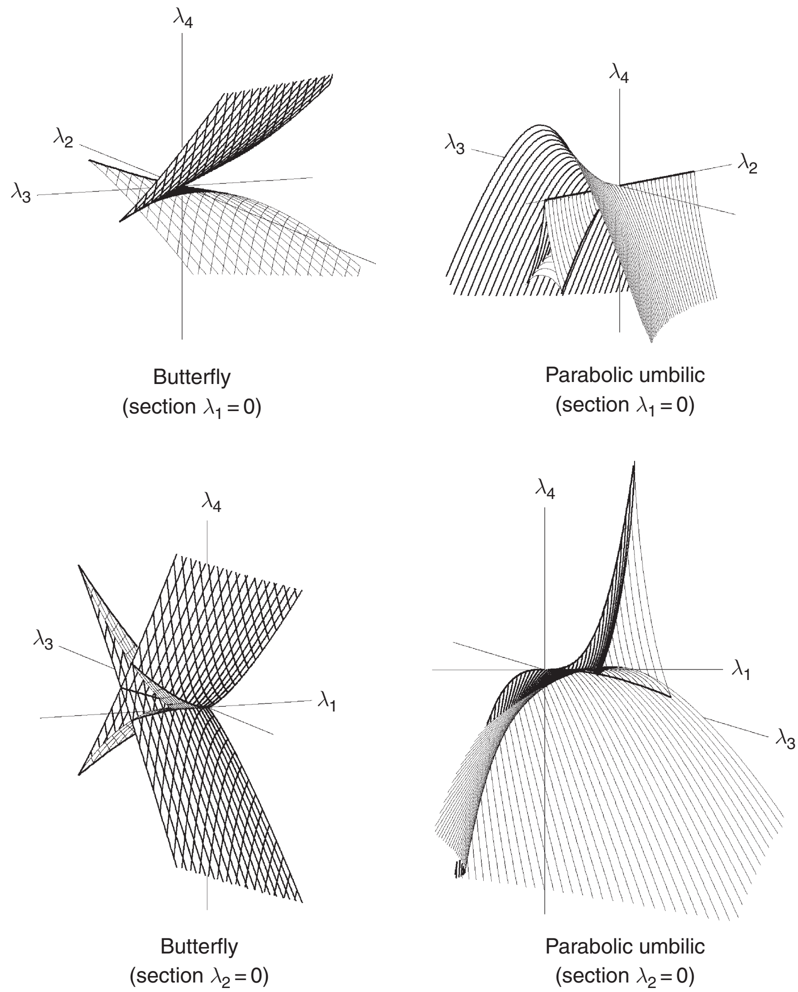



**Figure 2** Sections of catastrophe sets. Adapted with permission from Françoise J-P (2005) *Oscillations en Biologie: Analyse Qualitative et Modèles* (Mathématiques et Applications, vol. 46). Heidelberg: Springer.

When the parameter $r < 0$, it displays one stable equilibrium position. As $r$ increases, this equilibrium bifurcates (for $r > 0$) into two stable equilibria and an unstable equilibrium. Its drawing suggests "the pitchfork." In case of subcritical pitchfork bifurcation

$$\dot{x} = rx + x^3$$

there is a single stable state for $r < 0$ that bifurcates into two stable states and one unstable as $r > 0$.

### Normal Forms

Local analysis of vector fields proceeds with local models called normal forms. A local vector field near a singular point (zero) is seen as a derivation of the local ring of functions which preserves the unique maximal ideal (of the functions which vanish at the singular point). It yields a linear operator of the finite-dimensional vector spaces of truncated Taylor expansions of functions. This leads to decomposition of the vector fields into semisimple and nilpotent parts (at the level of formal series). A normal form is a formal coordinate system in which the semisimple part is linear. If the vector field preserves a structure (like volume form or symplectic form) the change of coordinates which brings it to its normal form is also (volume-preserving, symplectic). The simplicity of the normal form depends on the number of allowed resonances for the eigenvalues of the first-order jet of the vector field at the singular point. The best-known example is the Birkhoff normal form of Hamiltonian vector fields that we recall now, but we should also mention the Sternberg normal form of volume-preserving vector fields.

Local analysis of a Hamiltonian vector field under symplectic changes of coordinates is the same as the local analysis of functions (namely its associated Hamiltonian). Birkhoff normal form deals with the

case of a Hamiltonian that is a perturbation at the origin:

$$H_0(p) = \sum_{j=1}^{m} \lambda_j p_j$$

$$p_j = x_j^2 + y_j^2, \quad j = 1, \ldots, m$$

where the symplectic form is

$$\omega = \sum_{j=1}^{m} \mathrm{d}x_j \wedge \mathrm{d}y_j$$

If the eigenvalues $\lambda_j$ are assumed to be independent over the integers (no resonances), then there is a formal system of symplectic coordinates $\hat{p}_j, \hat{q}_j$, $j = 1, \ldots, m$, called action-angle variables, in which the Hamiltonian only depends of the action variables $\hat{p}_j$. Such a coordinate system is generically divergent because, under generic assumptions on the 3-jet of the Hamiltonian, the system displays isolated periodic orbits in any neighborhood of the origin (see Moser, Vey, Francoise). Normal forms are normally used in applications (e.g., Nekhoroshev theorem, Hopf bifurcation theorem) in their truncated versions. Birkhoff normal form was conjectured (A Weinstein) to enter in the asymptotic expansion of the fundamental solution of the wave equation on a Riemannian manifold near elliptic geodesics. This conjecture was recently proved by V Guillemin.

### Stability Theory of Hamiltonian Systems, Nekhoroshev Theorem, Arnol'd Diffusion

The generic divergence of the Birkhoff normal form does not allow one to conclude about the stability of the elliptic singular point. In the case where it is convergent, the motion is trapped inside invariant tori (conservation of the actions). The KAM theorem (see Gallavotti (1983)) provides the existence of many invariant tori but, except in low dimensions, this does not exclude the existence of trajectories that would escape to infinity. Arnol'd indeed provided a mechanism and examples of such situations (this is now called Arnol'd diffusion) (*see* Introductory Articles: Classical Mechanics). This diffusion process needs some time, which is estimated below by a theorem of Nekhoroshev.

Consider the Hamiltonian

$$H_\epsilon(p, q) = h(p) + \epsilon f(p, q)$$

where $h(p)$ is strictly convex, analytic, anisochronous on the closure $\overline{U}$ of an open bounded region $U$ of $R^m$ and the perturbation $f(p, q)$ is analytic on $\overline{U} \times R^m$. Nekhoroshev's theorem tells that there are positive constants $a, b, d, g, \tau$ such that for any initial data $p_0, q_0$, the actions $p$ do not change by more than $a\epsilon^g$ before a time bounded below by $\tau e^{b/\epsilon^d}$ (see Gallavotti (1983)).

### Bifurcations of Periodic Orbits

Consider a one-parameter family of vector fields $X_\lambda$ of class $C^k, k \geq 3$,

$$\dot{x} = F(x, \lambda)$$

Assume that $X_\lambda(0) = 0$ and that the linear part of the vector field at 0 has two complex-conjugated eigenvalues $\mu(\lambda)$ and $\overline{\mu(\lambda)}$ such that $\mathrm{Re}(\mu(\lambda)) > 0$ for $\lambda > 0, \mathrm{Re}(\mu(0)) = 0$ and $(\mathrm{Re}(\mu(\lambda)))/d\lambda|_{\lambda=0} \neq 0$. Then, for $\lambda > 0$ but small enough, the vector field $X_\lambda$ has a periodic orbit $\gamma_\lambda$ which tends to 0 as $\lambda$ tends to 0.

This bifurcation of codimension 1 is named Hopf bifurcation and it occurs in many models.

When several oscillators (conservative or dissipative) are weakly coupled, they may display frequency locking (existence of an attractive periodic orbit) phase locking, and synchronization. The fact that we always see the same face of the Moon from the Earth can be explained by a synchronization of the rotation of the Moon onto itself with its rotation around the Earth. Synchronization also plays a fundamental role in living organisms (e.g., heart, population dynamics: see D Attenborough's movie "The Trials of Life"). It is sometimes possible to be convinced of synchronization via computer experiments, but the main theoretical approach is due to Malkin. *See* Bifurcations of Periodic Orbits, where a full mathematical proof is included.

### Homoclinic Bifurcation, Newhouse's Phenomenon

Homoclinic bifurcation occurs in the family $X_\lambda$ at the bifurcation value of the parameter $\lambda = 0$ if $X_0$ displays a singular orbit which tends to 0 both for $t \to +\infty$ and for $t \to -\infty$. In dimension 2, if $\lambda$ is slightly deformed around 0, one periodic orbit may appear (or disappear). For planar systems, the Bogdanov–Takens bifurcation is the codimension-2 bifurcation, which mixes the homoclinic and the Hopf bifurcations. In dimension 3, more complicated phase diagrams may occur (such as in the Shilnikov bifurcation) with the appearance of infinitely many periodic orbits or homoclinic loops (in a stable way: Newhouse phenomenon). This eventually gives rise to strange attractors (the Roessler attractor).

### The Poincaré Center-Focus Problem, Local Hilbert's 16th Problem, Abel Equations, Algebraic Moments

Hopf bifurcation theory for two-dimensional systems deals with the first case of a general situation

often referred to as degeneracies of Hopf bifurcations or alternatively Hopf–Takens bifurcations. Consider more generally a planar vector field, tangent at the origin to a linear focus:

$$\dot{x} = y + \mu x + f(x, y)$$
$$\dot{y} = -x + \mu y + g(x, y)$$

The Poincaré center-focus problem asks for necessary and sufficient conditions on the perturbation terms so that all orbits are periodic in a neighborhood of the origin. This problem is still pending in the case, for instance, where $f$ and $g$ are homogeneous of degrees 4 and 5. It was solved a long time ago for degrees 2 and 3. Part (b) of Hilbert's 16th Problem asks for finding a bound in terms of the degrees of polynomial perturbations for the number of limit cycles (isolated periodic orbits) in the neighborhood of the origin. In the case of homogeneous perturbations, a Cherkas transformation allows the reduction of both problems to the so-called one-dimensional periodic Abel equations:

$$\mathrm{d}y/\mathrm{d}x = p(x)y^2 + q(x)y^3$$

where $p$ and $q$ are trigonometric polynomials in $x$. A perturbative approach was developed for several years and yields a theory of algebraic moments related to Livsic's generalized problem of moments.

## Fast–Slow Systems

Fast–slow systems

$$\epsilon\,\dot{x} = f(x, y), \qquad \dot{y} = g(x, y)$$

are characterized by the existence of two timescales. Variables $x$ are called fast variables and $y$ are called slow variables. Different approximation techniques can be used (averaging method, multiscale approach (*see* Multiscale Approaches)). The behavior of solutions is approximated as follows (when the scale $\epsilon$ is small). The orbit jumps to an attractor of the fast dynamics. This attractor may eventually lose its stability and/or bifurcate as time evolves. Then the orbit jumps to another attractor of the fast dynamics. Once again, this attractor may evolve/bifurcate/disappear, depending on the slow variables $y$. This explains why bifurcation theory enters in the process in a crucial way, and it has to be adapted to this special context where some new phenomena may occur (e.g., singular Hopf bifurcation theory, Canards, etc.). Fundamental tools to be used in this context are Takens theorem, Fenichel central manifold theorem, blowing-up (Dumortier–Roussarie).

Excitability is also an important feature which occurs in some fast–slow systems. Consider initial data in a neighborhood of an excitable attractive point. For some initial data, the orbit goes very quickly to the attractor. For some others instead (usually below some threshold), the orbit undergoes a long incursion in the phase diagram before turning back to the attractive point.

Singular Hopf bifurcation, hysteresis, and excitability can, for instance, occur in the electrodissolution and passivation of iron in sulfuric acid (see Alligood *et al.* (1997)).

Sometimes, the orbit leaves the neighborhood of a first attractor to jump to a second one and then this second one disappears and the orbit jumps back to the initial attractor as the slow variables have undergone a cycle. This is called a hysteresis cycle. In case one of the attractors is a point while the other is an attractive periodic orbit, it may lead to bursting oscillations. These oscillations are characterized by the periodic succession of silent phases (attractor of the fast dynamics) and active (pulsatile) phases (periodic attractor of the fast dynamics). They are ubiquitous in physiology, where they were first discovered and can be also observed in physics (laser beams) and in population dynamics.

### Example

The Hindmarsh–Rose model displays bursting oscillations:

$$\epsilon\dot{x} = y - x^3 + 3x^2 + I - z$$
$$\epsilon\dot{y} = 1 - 5x^2 - y$$
$$\dot{z} = s(x - x_1) - z$$

The fast dynamics is two dimensional. For some values of the parameters, it displays an attractive node, a saddle and a repulsive focus. Under the slow variation of $z$, the fast dynamics displays a saddle–node bifurcation, a Hopf bifurcation from which emerges a stable limit cycle which disappears into a homoclinic bifurcation. The fast–slow system undergoes a hysteresis loop which yields to bursting oscillations.

## Conclusions

Over the past three decades, mathematical techniques gathered under the names of singularity theory and bifurcation theory of dynamical systems have offered a powerful means to explore nonlinear phenomena in diverse settings. These include mechanical vibrations, lasers, superconducting circuits, and chemical oscillators. Many such instances are further developed in this encyclopedia.

*See also:* Bifurcation Theory; Bifurcations of Periodic Orbits; Chaos and Attractors; Entropy and Quantitative Transversality; Ergodic Theory; Feynman Path Integrals; Generic Properties of Dynamical Systems; Gravitational Lensing; Homoclinic Phenomena; Hyperbolic Dynamical Systems; Multiscale Approaches; Optical Caustics; Poisson Reduction; Stationary Phase Approximation; Symmetry and Symmetry Breaking in Dynamical Systems; Symmetry and Symplectic Reduction; Synchronization of Chaos; Weakly Coupled Oscillators.

## Further Reading

Alligood KT, Sauer TD, and Yorke JA (1997) *Chaos, An Introduction to Dynamical Systems, Textbooks in Mathematical Sciences*. New York: Springer.

Alpay D and Vinikov V (eds.) (2001) *Operator Theory, System Theory and Related Topics, The Mosche Livsic Anniversary Volume, Operator Theory, Advances and Applications* vol. 123. Birkhauser.

Briskin M, Francoise JP, and Yomdin (2001) *Generalized Moments, Cener-Focus Conditions and Compositions of Polynomials*. Operator Theory, Advances and Applications 123 ( in honor of M Livsic, 80th birthday).

Diener M (1994) The canard unchained, or how fast–slow dynamical systems bifurcate? *The Mathematical Intelligencer* 6: 38–49.

Francoise JP and Guillemin V (1991) On the period spectrum of a symplectic map. *Journal of Functional Analysis* 100: 317–358.

Gallavotti G (1983) *The Elements of Mechanics*. New York: Springer.

Goodstein DL and Goodstein JR (1997) *Feynmann's lost lecture*. London: Vintage.

Guckenheimer J (2004) Bifurcations of relaxation oscillations. In: Ilyashenko Y, Rousseau C, and Sabidussi G (eds.) *Normal Forms, Bifurcations and Finiteness Problems in Differential Equations. Séminaire de mathématiques supérieures de Montréal*, Nato Sciences Series, II. Mathematics, vol. 137, pp. 295–316. Kluwer.

Haken H (1983) *Synergetics*, 3rd edn. Berlin: Springer.

Keener J and Sneyd J (1998) Mathematical Physiology. *Interdisciplinary Applied Mathematics*, vol. 8. New York: Springer.

Malgrange B (1974) Intégrales asymptotiques et monodromie. *Annales de l'ENS* 7: 405–430.

May R-M (1976) Simple mathematical models with very complicated dynamics. *Nature* 261: 459–467.

Nekhoroshev V (1977) An exponential estimate of the time of stability of nearly integrable Hamiltonian systems. *Russian Mathematical Surveys* 32(6): 1–65.

Palis J and de Melo W (1982) *Geometric Theory of Dynamical Systems, An Introduction*. New York: Springer.

Perko L (2000) *Differential Equations and Dynamical Systems*, 3rd edn, Text in Applied Mathematics, vol. 7. New York: Springer.

Siegel C-L and Moser J (1971) *Lectures on Celestial Mechanics*, Die Grundleheren der mathematischen Wissenschaften, vol. 187. Berlin: Springer.

Smale S (1998) Mathematical problems for the next century. *The Mathematical Intelligencer* 20: 7–15.

Smale S. Dynamics retrospective: great problems, attempts that failed. *Physica D* 51: 267–273.

**Sobolev Spaces** *see* Inequalities in Sobolev Spaces

# Solitons and Kac–Moody Lie Algebras

**E Date**, Osaka University, Osaka, Japan



## Introduction

Solitons and Kac–Moody Lie algebras were born at almost the same time in the 1960s, although they did not have a connection at first. They both have roots in the history of mathematics. From the 1970s on, they became intersection points for many (previously known and new) results.

The notion of solitons has many facets and it is difficult to give a mathematically precise definition; closely related to solitons is the notion of "completely integrable systems." The latter is usually used in a much broader sense.

The terminology "soliton" was originally used for a particular phenomenon in shallow water waves. Now, in its broadest sense, it is used to represent an area of research relating to this particular phenomenon in direct or indirect ways. From the viewpoint of solitons, particular solutions of differential equations are of special interest. Although particular solutions have been studied for a long time, interest in them was overshadowed by the method of functional analysis in the 1950s. In the late nineteenth century, in parallel with the theory of algebraic functions, several studies undertook the solution of mechanical problems by elliptic or hyperelliptic integrals. Subsequently, however, there was a drop in activity in this area of work.

Originally it was hoped that this kind of phenomenon could be used for practical applications. No mention of practical application of solitons will be made in this article.

First we list several topics which constitute the main body of the notion of solitons in the early stages; we will then explain relations with Kac–Moody Lie algebras.

## Birth of Solitons

The name "soliton" itself was coined by Martin D Kruskal around 1965. It was originally employed for the solitary wave solution Korteweg–de Vries (KdV) equation

$$u_t - \frac{1}{4}(6uu_x + u_{xxx}) = 0, \quad u = u(x,t) \qquad [1]$$

The coefficients here are not important. We can change them arbitrarily. The unknown function $u$, or rather $-u$, represents the height of the wave.

The solitary wave solution in question is given by

$$u(x,t) = -2c\ \mathrm{sech}^2\left(\sqrt{c}(x - ct - d)\right) \qquad [2]$$

This is a traveling-wave solution with the height of the wave proportional to the speed. This is one feature of the nonlinearity of this differential equation.

A reason for this nomenclature comes from the particle-like property of solitary wave observed via numerical computations. That is, if we have two solitons [2] with different speeds, with the faster one on the left and the slower one on the right, then after some time they collide and their shapes are distorted. After a long enough time, they are separated and recover their original shapes, the only difference being in the change of the phase shift $d$ in [2].

Solitary waves in shallow water (like a canal) were first observed by Scott Russell in Scotland in the middle of the nineteenth century. Differential equations which possess solitary waves in shallow water as solutions were sought after Scott Russell's report. Boussinesq derived one (now called the Boussinesq equation, which contains second partial derivatives with respect to time) from the Euler equation of water wave; then in 1895 Korteweg and his student de Vries derived the KdV equation. They also showed that the KdV equation possesses solutions expressible in terms of elliptic functions.

In the 1960s Kruskal and Zabusky carried out numerical computations for the Fermi–Pasta–Ulam problem; they also came across the KdV equation and found the aforementioned phenomenon.

## Inverse-Scattering Method

Kruskal and his co-workers further pursued the origin of the particle-like property of solitons and proposed the so-called inverse-scattering method.

The inverse problem of scattering theory of the one-dimensional Schrödinger operator

$$L = -\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^2 + u(x)$$

was studied by Gelfand–Dikii, Marchenko, and Krein in the 1950s, motivated by scattering theory in quantum mechanics.

It gives a one-to-one correspondence between rapidly decreasing potentials $u(x)$ and scattering data which consist of discrete eigenvalues $-\eta_j^2$ and normalization $c_j, j = 1, \ldots, n$, of the eigenfunctions corresponding to them and the reflection coefficient $r(\xi)$. The reflection coefficient represents the ratio of reflection of the unit plane wave $\mathrm{e}^{\mathrm{i}\xi x}$ by the potential field. The scattering data $\{r(\xi), \eta_j, c_i, j = 1, \ldots, n\}$ are a mathematical idealization of observable data in quantum scattering. The procedure of reconstructing a potential from given scattering data is called the inverse problem. The heart of this procedure is solving an integral equation (the Gelfand–Dikii–Marchenko equation). In the reflectionless case ($r(\xi) = 0$), this integral equation reduces to a system of linear algebraic equations.

Kruskal and co-workers found that the scattering data of these operators with solutions of [1] as potentials depend very simply on $t$:

$$\begin{aligned} \eta_j(t) &= \eta_j(0), \qquad c_j(t) = c_j(0)\,\mathrm{e}^{2\mathrm{i}\eta^3 t} \\ r(\xi,t) &= r(\xi,0)\,\mathrm{e}^{2\mathrm{i}\xi^3 t} \end{aligned} \qquad [3]$$

It was realized at the same time that soliton solutions correspond to a reflectionless potential ($r(\xi) = 0$) with only one discrete eigenvalue, while reflectionless potentials correspond to a nonlinear "superposition" of soliton solutions (called multisoliton solutions) and describe the interaction of solitons.

As was pointed out by Zakharov and others, the inverse-scattering method has an intimate relation with the Riemann–Hilbert problem.

## Lax Representation

Looking at this invariance of the spectrum, Lax reformulated the KdV equation [1] as an evolution equation for the one-dimensional Schrödinger operator:

$$\begin{aligned} \frac{\mathrm{d}L}{\mathrm{d}t} &= [A, L], \quad L = \left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^2 + u \\ A &= \left(\frac{\partial}{\partial x}\right)^3 + \frac{3}{4}\left(u\frac{\partial}{\partial x} + \frac{\partial}{\partial x}u\right) \end{aligned} \qquad [4]$$

Here we have changed the sign of the operator for later convenience. This form of representation together with the inverse-scattering method gave a framework for finding nonlinear differential (difference) equations that have solutions with properties similar to solitons (soliton equations).

Among such are the sine-Gordon equation

$$u_{tt} - u_{xx} = \sin u$$

the nonlinear Schrödinger equation

$$\mathrm{i}u_t + u_{xx} + |u|^2 u = 0$$

the modified KdV equation

$$u_t - \frac{1}{6}\left(6u^2 u_x + u_{xxx}\right) = 0$$

the Toda lattice equation

$$\begin{aligned}\frac{\mathrm{d}Q_n}{\mathrm{d}t} &= P_n \\ \frac{\mathrm{d}P_n}{\mathrm{d}t} &= -\exp(Q_n - Q_{n+1}) + \exp(Q_{n-1} - Q_n)\end{aligned} \qquad [5]$$

and so on. The first three are obtained by replacing $L$ by a $2 \times 2$ matrix differential operator of first order. For eqn [5], the linear operator corresponding to $L$ in the case of the KdV equation is a difference operator of order 2 and has a connection with the theory of orthogonal polynomials in one variable as well as with the theory of moment problems.

Later it was remarked that the differential operator $A$ in eqn [4] is nothing but the differential operator part of the fractional power of $L: A = (L^{3/2})_+$. By replacing $A$ in [4] by $(L^{(2n+1)/2})_+$ we obtain higher ($n$th) KdV equations.

## Basic Representations of Affine Lie Algebras

In the 1960s Kac and Moody introduced independently a class of infinite-dimensional Lie algebras which are in many respects close to finite-dimensional semisimple Lie algebras. Each of them is constructed for a given generalized Cartan matrix (GCM),

$$\begin{aligned}C = \left(a_{ij}\right), \quad a_{ii} = 2,\ a_{ij} \leq 0 \text{ for } i \neq j \\ \text{and if } a_{ij} = 0 \text{ then } a_{ji} = 0\end{aligned} \qquad [6]$$

There is a special class of Kac–Moody Lie algebras that are now called affine Lie algebras. They correspond to positive-semidefinite GCM and are realized as central extensions of loop algebras (current algebras)

$$\mathbf{C}[\lambda, \lambda^{-1}] \otimes \mathfrak{g}$$

of finite-dimensional semisimple Lie algebras $\mathfrak{g}$. They have many applications in physics, in particular as current algebras. The Sugawara construction in current algebra plays an essential role in conformal field theory. Note that finite-dimensional semisimple Lie algebras correspond to positive-definite GCMs.

In the late 1970s, there was interest in constructing representations of these algebras after the general theory of representations was constructed. Among them was the work of Lepowsky–Wilson, who constructed basic representations of the affine Lie algebra $A_1^{(1)} (= \widehat{\mathfrak{sl}}_2)$ using differential operators of infinite order in infinitely many variables. These operators were called vertex operators by Garland, in view of the resemblance to objects in string theory. Character formulas for these new Lie algebras were intensively studied and many combinatorial identities were (re)derived.

## Geometric Interpretation

How do Kac–Moody Lie algebras enter into this picture?

In the early stages of the history of solitons Kac–Moody Lie algebras appeared rather artificially. Some authors tried to understand solitons from geometric viewpoints. A typical example is the sine-Gordon equation. This equation appears as the Gauß–Codazzi equation in the theory of embeddings of two-dimensional surfaces of constant negative curvature into three-dimensional Euclidean space, while the Gauß–Weingarten equation is the linear equation that appears in the Lax representation of the sine-Gordon equation. Another approach of a geometric nature, involving the prolongation structure, was the direction initiated by Wahlquist–Estabrook. In this approach, the Lie algebra appeared in a natural way, although the nature of such Lie algebras was not so clear. This direction of research is close in spirit to the method of Cartan for treating partial differential equations.

Several authors considered generalizations of the Toda lattice equation. Bogoyavlenskii and others observed that the original Toda lattice equation [5] is related to the Cartan matrix of the affine Lie algebra of type $A$. Viewed in this way, it was straightforward to generalize the Toda lattice equation to Cartan matrices of another type of affine Lie algebras and also to ordinary Cartan matrices. These were typical appearances of Kac–Moody Lie algebras in the theory of solitons; they were used to produce soliton equations. The climax of this is the work of Drinfel'd–Sokolov.

It needed some time to understand another role of affine Lie algebras in the theory of solitons.

## Bäcklund Transformation

In the theory of two-dimensional surfaces of constant negative curvature, a method of obtaining another surface of constant negative curvature from the given one with some parameter was known by the work of Bäcklund. If we apply this to the trivial solutions $u = 0$ of the sine-Gordon equation, we

obtain a one-soliton solution of the sine-Gordon equation. From this fact, the transformation of solutions of soliton equations to other solutions is called a Bäcklund transformation. The original Darboux transformation is a special case of a Bäcklund transformation.

## Hamiltonian Formalism

Another discovery of Gardner–Greene–Kruskal–Miura was the Hamiltonian structure of the KdV equation. In the process of showing the existence of infinitely many conservation laws, they used the so-called Miura transformation, which relates the KdV and the modified KdV equation. Faddeev–Zakharov showed that the transformation to scattering data is a canonical transformation, and conserved quantities are obtained from the expansion of the reflection coefficients.

Gelfand–Dikii studied Hamiltonian structures of the KdV equation using the formal variational calculus they initiated.

M Adler was the first to try to study the KdV equation by using the orbit method known for finite-dimensional Lie algebras. It was known by the works of Kostant and Kirillov or even earlier by Lie that the co-adjoint orbits of Lie algebras admit symplectic structures (the Kostant–Kirillov bracket). Adler considered the algebra of pseudodifferential operators in one variable. This acquires the structure of Lie algebra by the commutation relation. This algebra admits a natural triangular decomposition by order. He showed that the KdV equation can be viewed as a Hamiltonian system in the co-adjoint orbit of the one-dimensional Schrödinger operator with the Kostant–Kirillov bracket. By introducing the notion of residue of pseudodifferential operators he rederived conserved quantities. The work of Drinfeld–Sokolov can be regarded as a thorough generalization of this direction. Hamiltonian structures of the KdV equation and other soliton equations are now understood in this way.

The method is also applicable to finite-dimensional Lie algebras. Symes, Kostant, and others treated the finite Toda lattice in this way.

The motion of tops, including that of Kovalevskaya, was also studied in this way.

## Hirota's Method

There was another approach to soliton equations, quite different from the above. This was the method initiated by Hirota. He placed stress on the form of multisoliton solutions of the KdV equation, the sine-Gordon equation, and so on. He made a dependent-variable transformation of the KdV equation [1],

$$u = 2\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)\log f$$

This form naturally arises when we reconstruct the potential of the one-dimensional Schrödinger operator from the scattering data by solving the Gelfand–Dikii–Marchenko integral equation. In this new dependent variable, eqn [1] takes the following form:

$$\left(D_x^4 - 4D_x D_t\right) f(x,t)\cdot f(x,t) = 0$$

where the operator $D_x$ is defined by

$$D_x(f\cdot g) = \frac{\mathrm{d}}{\mathrm{d}x'} f(x+x')\, g(x-x')|_{x'=0} \qquad [7]$$

This operator is called Hirota's bilinear differential operator. In such transformed form, he tried to solve the resulting equation in a perturbative way,

$$\begin{aligned} f = 1 &+ \sum_{j=1}^{n} \exp(2p_j x + 2p^3 t + q_j) \\ &+ \sum_{1\le j<k\le n} c_{ij}\exp(2(p_j+p_k)x \\ &+ 2(p_j^3+p_k^3)t + q_j + q_k) + \cdots \end{aligned} \qquad [8]$$

It is rather miraculous that in the soliton equation case we can truncate such a perturbative procedure at a finite point. The number of steps corresponds to the number of solitons.

Most of the soliton equations are rewritten in bilinear form with such bilinear differentiation after a suitable dependent-variable transformation. (Some equations need several new dependent variables.) Once we have a differential equation in Hirota's bilinear differential form, it always has two-soliton solutions.

Up to 1980, keywords characterizing solitons were; inverse-scattering method, Bäcklund transformation, multisolitons, Hirota's method, quasi-periodic solutions, etc. No explicit mention was made of representation theory.

## Hierarchy of Soliton Equations

As was stated above, soliton equations viewed as Hamiltonian systems have infinitely many conservation laws. This implies that we can introduce infinitely many independent time variables consistently. From this viewpoint, it is natural to consider the KdV equation and its higher-order analogs simultaneously. They have many properties in common. For example, the $t$-dependence of the scattering data of the higher

KdV equation is given by replacing $\xi^3$ by $\xi^{2n+1}$ and $\eta_j^3$ by $\eta_j^{2n+1}$ in eqn [3]. The totality of soliton equations organized in this way is called a hierarchy of soliton equations; in the KdV case, it is called the KdV hierarchy. This notion of hierarchy was introduced by M Sato. He tried to understand the nature of the bilinear method of Hirota. First, he counted the number of Hirota bilinear operators of given degree for hierarchies of soliton equations. For the number of bilinear equations, M Sato and Y Sato made extensive computations and made many conjectures that involve eumeration of partitions.

## Kadomtsev–Petviashvili Hierarchy

Although it was included in a family of soliton equations slightly later, the Kadomtsev–Petviashvili (KP) equation is a soliton equation in three independent variables, which first appeared in plasma physics:

$$\tfrac{3}{4}u_{yy} - \left(u_t - \tfrac{1}{4}(6uu_x + u_{xxx})\right) = 0 \qquad [9]$$

For this equation we have to replace the Lax representation by

$$\left[\left(\frac{\partial}{\partial x}\right)^2 + u - \frac{\partial}{\partial y}, \left(\frac{\partial}{\partial x}\right)^3 + \frac{3}{2}u\frac{\partial}{\partial x} + v - \frac{\partial}{\partial t}\right] = 0 \qquad [10]$$

This form of representation was introduced by Zakharov–Shabat. Sometimes it is referred to as the zero-curvature representation or the Zakharov–Shabat representation. The KP equation is universal in the sense that it contains the KdV equation [1] and the Boussinesq equation as special cases. If $u$ does not depend on $y$, resp. $t$, this gives the KdV, resp. the Boussinesq equation.

## Work of Sato

Sato stressed the importance of the study of the KP equation. He first introduced the KP hierarchy. Instead of the one-dimensional Schrödinger operator in the KdV case consider a pseudo- (micro) differential operator of first order,

$$\begin{aligned} L &= \partial + u_2(x)\partial^{-1} + u_3(x)\partial^{-3} + \cdots \\ \partial &= \frac{\partial}{\partial x_1}, \quad x = (x_1, x_2, x_3, \ldots) \end{aligned} \qquad [11]$$

Setting $B_n = (L^n)_+$, the KP hierarchy is defined by the Zakharov–Shabat representation

$$\left[\frac{\partial}{\partial x_m} - B_m, \frac{\partial}{\partial x_n} - B_n\right] = 0, \quad m, n = 2, 3, \ldots$$

If we assume that $L^2$ is a differential operator, we have the KdV hierarchy and the constraint that $L^3$ is a differential operator gives the Boussinesq hierarchy. This process is called reduction.

Sato found that character polynomials (Schur functions) solve the KP hierarchy and, based on this observation, he created the theory of the infinite-dimensional (universal) Grassmann manifold and showed that the Hirota bilinear equations are nothing but the Plücker relations for this Grassmann manifold.

Sato also gave an (infinite-dimensional) determinantal formula for Hirota's dependent variable and called the latter the $\tau$-function. Using this $\tau$-function, the wave function (the eigenfunction corresponding to the KP hierarchy) is expressed as

$$\begin{aligned} w(x,k) &= \exp\left(\sum_{n=1}^{\infty} x_n k^n\right)\frac{\tau(x - \epsilon(k^{-1}))}{\tau(x)} \\ \epsilon(k) &= \left(k, \frac{k^2}{2}, \frac{k^3}{3}, \ldots\right) \\ Lw &= kw \end{aligned} \qquad [12]$$

where $L$ is given by eqn [11].

## Affine Lie Algebras as Infinitesimal Transformation Groups for Soliton Equations

Date–Jimbo–Kashiwara–Miwa found another relation among soliton equations and affine Lie algebras. After noticing some similarity between the formula in the paper by Lepowsky–Wilson on the Rogers–Ramanujan identity using the vertex operators for $A_1^{(1)}$ and the formula in the computation of numbers of bilinear operators in Sato's paper, they applied the vertex operator for $A_1^{(1)}$,

$$X(p) = \exp\left(\sum_{j=1}^{\infty} 2x_{2j-1}p^{2j-1}\right) \times \exp\left(-\sum_{j=1} \frac{2}{jp^{2j-1}}\frac{\partial}{\partial x_{2j-1}}\right)$$

to 1 (which is the simplest $\tau$-function for the KdV hierarchy), where $p$ is a parameter. They found that the result is the $\tau$-function corresponding to the one-soliton solution of the KP hierarchy. They also found that successive application of $X(p)$'s to 1 produced all multisoliton $\tau$-functions. Therefore, applications of vertex operators are precisely

Bäcklund transformations. This implies that the affine Lie algebra $A_1^{(1)}$ is the infinitesimal transformation group for solutions of the KdV hierarchy.

After this discovery, it was realized that the totality of $\tau$-functions of the KdV hierarchy is the group orbit of the highest weight vector (=1) of the basic representation of $A_1^{(1)}$.

The vertex operators for the KP hierarchy were also found:

$$X(p,q) = \exp\left(\sum_{j=1}^{\infty} x_j(p^j + q^j)\right) \times \exp\left(-\sum_{j=1}\left(\frac{1}{jp^j} + \frac{1}{jq^j}\right)\frac{\partial}{\partial x_j}\right)$$

If we put $q = -p$, the vertex operator for $A_1^{(1)}$ ([12]) is recovered.

Viewed in this way the Lie algebra corresponding to the KP hierarchy is $\mathfrak{gl}_\infty(=A_\infty)$. And an embedding of $A_1^{(1)}$ into $A_\infty$ was also found. Subsequently, the method using free fermions (Clifford algebras) was established. Frenkel–Kac had already used free fermions to construct basic representations. In this approach, the $\tau$-functions are defined as vacuum expectation values. Based on this connection with affine Lie algebras, many conjectures of Sato on the number of bilinear equations are (re)proved by using specialized characters of affine Lie algebras.

The use of free fermions was exploited by Ishibashi–Matsuo–Ooguri to relate soliton equations with conformal field theory on Riemann surfaces. This aspect was further studied by Tsuchiya–Ueno–Yamada using $D$-modules.

Once such a viewpoint was established, it was easy to construct soliton equations corresponding to other affine Lie algebras. Hierarchies similar to the KP hierarchies (the simplest equation contains three variables) were also found, which correspond to Lie algebras like $\mathfrak{go}_\infty, \mathfrak{sp}_\infty$ (the BKP hierarchy, the CKP hierarchy, and so on).

Summarizing these developments, we can say that affine Lie algebras, or slightly larger ones like $\mathfrak{gl}_\infty$, appear naturally as infinitesimal transformation groups for soliton equations and the solution spaces are the (completed) group orbits of highest weight vector $\tau$-functions of level-1 representations. The Hirota bilinear equations are the equations describing these orbits (analogs of Plücker relations).

Soon afterwards, the notion of $\tau$-functions was introduced in the study of Painlevé equations by Okamoto, revealing Hamiltonian structures in Painlevé equations.

## The Method of Drinfeld–Sokolov

The KdV or the KP hierarchies are related to scalar linear differential operators. A parallel treatment using matrix differential operators is also possible. In fact, the nonlinear Schrödinger equation, modified KdV equation, the sine-Gordon equation, etc., are treated in this way.

Drinfel'd and Sokolov gave a general framework along these lines. The first step is to choose the starting (matrix-valued) linear differential operator of order one. For that they use the language of Lie algebras.

Let us start with a matrix realization of a Lie algebra (for an affine Lie algebra, the elements are Laurent polynomials in one variable). Consider a linear differential operator of the following form:

$$L = \frac{\mathrm{d}}{\mathrm{d}x} + q(x) + \Lambda$$

where $q(x)$ is an element of the Borel subalgebra and $\Lambda$ is a sum of positive Chevalley generators in the case of affine Lie algebras. By using gauge transformations (adjoint group), they consider several normal forms. One normal form is obtained by choosing a node of the corresponding Dynkin diagram. The resulting matrix system is equivalent to the one obtained by scalar Lax representation (or a slight generalization of it). In this way, the generalized KdV equations for affine Lie algebras are obtained. Another normal form is to make $q$ $\mathfrak{h}$-valued. Soliton equations obtained in this way are called the modified KdV equations. This is a generalization of the Miura transformation. They also comment on the construction of partially modified soliton equations, which correspond to taking various parabolic subalgebras. The Hamiltonian formalism is also treated from their viewpoint.

In summary, in their approach affine algebras are used to construct soliton equations, or one can say that they consider the space of initial values of soliton equations.

They also discuss two-dimensional Toda lattices in their setting and show that modified equations in their sense are symmetries of the two-dimensional Toda lattices.

## Common Features of the Roles of Affine Lie Algebras in Solitons

In $\tau$-function approach as well as in the method of Drinfeld–Sokolov, the existence of triangular decomposition of Lie algebras was essential. In the former case, it was basic when considering highest-weight representations and, for the latter, it was used for the setup.

## Special Solutions of Soliton Equations (Multisoliton and Rational Solutions)

One of the characteristic features of soliton equations is that they allow rich special solutions. Multisoliton solutions were the starting point of the whole story. They directly relate to vertex operators of affine Lie algebras.

Rational solutions (in terms of $\tau$-function polynomial solutions) can be viewed as degenerations of multisoliton solutions. Motions of poles (or zeros) of the solutions are interesting. Airault–McKean–Moser studied the motion of poles of rational solutions of the KdV equation and found that they are identical to the motion of particles on a line (Calogero–Moser–Sutherland system). This viewpoint has now been generalized by Veselov and others.

Another discovery of Sato was that polynomial $\tau$-functions of the KP hierarchy are precisely Schur functions (character polynomials).

In accordance with the process of reduction, polynomial $\tau$-functions of the KdV hierarchy are Schur functions of special type.

## Quasiperiodic Solutions of Soliton Equations

As mentioned above, the KdV equation admits solutions expressible in terms of elliptic functions. Dubrovin–Novikov and Its–Matveev, almost at the same time, studied solutions of the KdV equation with periodic initial condition.

To the Sturm–Liouville (i.e., one-dimensional Schrödinger) operator with periodic potential

$$L = \left(\frac{\partial}{\partial x}\right)^2 + u(x), \quad u(x+l) = u(x)$$

there corresponds the discriminant, which is an entire function of the spectral parameter. Its zeros represent the periodic and antiperiodic spectrum $\lambda_j$ of the operator:

$$Lf_j(x) = \lambda_j f_j(x), \quad f_j(x+l) = \pm f_j(x)$$

It turns out that, except for a finite number of zeros, other zeros are double. Such a potential is called a finite-zone potential. These zones correspond to the spectrum of the operator in the $L^2$-sense. To a finite-zone potential $u(x)$ there corresponds a hyperelliptic curve

$$\mu^2 = \prod_{j=0}^{2n} (\lambda - \lambda_j)$$

with simple zeros $\lambda_j$ of the discriminant as zeros of polynomials defining the curve. If we consider the Dirichlet boundary value problem for the operator $L$,

$$Lf = \mu, f$$
$$f(s, \mu) = 0 = f(s+l, \mu)$$

the eigenvalues are discrete and each eigenvalue $\mu_j$ is located in a zone:

$$\lambda_{2j-1} \le \mu_j(s) \le \lambda_{2j}$$

So, for the double zeros ($\lambda_{2j-1} = \lambda_{2j}$), the corresponding Dirichlet eigenvalue $\mu_j(s)$ does not depend on $s$.

Dubrovin–Novikov also showed that a finite-zone potential is a stationary solution of the higher-order KdV equation (the order being equal to the number of nontrivial zones) and the $n$-zonal potentials form a finite-dimensional integrable system. In other words, the linear operators $L$, $A_n$ defining the $n$th order KdV equations commute,

$$[L, A_n] = 0$$

In passing, it was later found that such a pair of commuting linear differential operators was first studied by Burchnall–Chaundy in the 1920s. H F Baker remarked on the corresponding simultaneous eigenfunctions by relating them to multiplicative functions on algebraic curves.

## The Work of Krichever

Krichever reversed the above argument, utilizing the properties of corresponding eigenfunctions as a function of the spectral parameter. In this approach, we start with a compact Riemann surface $C$ (= nonsingular algebraic curve) of genus $g$. Here we apply his method to the KP hierarchy. Take a point $P_0$ on $C$ together with the inverse of a local parameter $k^{-1}$. Also take a general divisor $\delta$ on C of degree $g$. Consider a function $\psi(x, P), x = (x_1, x_2, \ldots)$, with the following properties:

1. $\psi$ is meromorphic on $C \backslash P_0$ with the pole divisor $\delta$, and
2. near $P_0, \psi$ behaves like

$$\psi(x, P) = \exp\left(\sum_{j=1}^{\infty} x_j k^j\right)\left(1 + \mathcal{O}(k^{-1})\right)$$

Such a $\psi$ exists uniquely and can be constructed using the theory of abelian integrals and the Jacobi problems on algebraic curves. Such a function was called the Baker–Akhiezer function, since Akhiezer constructed it by using abelian integrals and Jacobi's

problem in his study of moment problems (orthogonal polynomials).

It was later realized that Schur had much earlier considered such functions in the study of ordinary differential equations.

It is easy to show that such a function satisfies the following linear differential equations:

$$\frac{\partial}{\partial x_n}\psi = \left(\left(\frac{\partial}{\partial x_1}\right)^n + \sum_{j=0}^{n-1} u_j(x)\left(\frac{\partial}{\partial x_1}\right)^j\right)\psi, \quad n = 2, 3, \ldots$$

In this way, we obtain a solution of the KP hierarchy.

If there exists a rational function $f(P)$ on $C$ with poles only at $P_0$ with singular part $k^n, \psi$ can be factorized as

$$\psi(x, P) = \exp f(P)\psi'(x', P)$$

where $x'$ indicates the set of variables other than $x_n$. Consequently, we have

$$\frac{\partial}{\partial x_n}\psi(x, P) = f(P)\psi(x, P)$$

In this way, for a hyperelliptic curve $C$ and a branch point of it, viewed as the double cover of $\mathbf{C}P^1$, we recover the case of the KdV hierarchy.

Multisolitons correspond to rational algebraic curves with ordinary double points, while rational solutions correspond to further degeneration.

The study of quasiperiodic solutions of soliton equations revealed an intimate relationship with the theory of algebraic curves. One particular outcome was the characterization of Jacobian varieties among abelian varieties. This was originally posed by Schottky and subsequently reformulated by S P Novikov using soliton equations (Schottky problem, Novikov conjecture). This problem was solved through studies by Shiota, Mulase, and Arbarello–De Concini.

Another aspect was finding commutative subalgebras in the ring of linear differential operators. This problem is related to the theory of stable vector bundles on algebraic curves.

## Similarity Solutions of Soliton Equations

Ablowitz and Segur have shown that the Painlevé transcendent of the second kind solves the KdV equation as a similarity solution. This was the starting point of the study of similarity solutions of soliton equations.

Flaschka and Newell tried to construct the theory of multisimilarity solutions. As a by-product, they discussed modulation of the KdV equation by using the averaging method of Whitham. This opens the way to study the quasiclassical limit of soliton equations. This aspect was further studied by Dubrovin and others in connection with topological field theory.

Quite recently, Noumi and Yamada gave a generalization of the Painlevé equation in many variables by using the idea of similarity solutions of soliton equations. In the work of Noumi–Yamada, the affine Weyl group and $\tau$-functions play an essential role in constructing generalizations of the Painlevé equation. The shift or the unit of difference corresponds to imaginary null roots of affine Lie algebras. The idea is further applied to elliptic Painlevé equations.

## Integrable Many-Body Problems

As mentioned in relation with the rational solutions of soliton equations, the theory of integrable many-body problems has an intimate relationship with the theory of solitons. Recently, Veselov and his co-workers introduced the notion of Baker–Akhiezer functions of many variables. This concerns a commutative subring of differential operators in many variables. The structure of vector bundles on algebraic varieties of higher dimensions is quite different from that of algebraic curves. For this reason, a naïve generalization of soliton equations to higher dimensions is not possible. Veselov and others have set up a class of functions which they call multidimensional Baker–Akhiezer functions. They are defined by giving a finite set of vectors in a Euclidean space. The first problem is the existence. For the existence of the multidimensional Baker–Akhiezer function the set must satisfy several constraints. This is quite different from the case of solitons. Root systems satisfy these constraints and the corresponding Baker–Akhiezer function becomes the common eigenfunction of linear differential operators appearing in the Calogero–Sutherland–Moser model corresponding to root systems.

## Ball–Box Systems

Satsuma–Takahashi found a soliton-like phenomenon in cellular automata. It took much time for a mathematical explanation of this. Now it is understand that these systems are obtained by a limiting procedure from soliton equations. Sometimes this is called ultra-discretization. The system thus obtained can also be obtained from the theory of crystal bases of affine Lie algebras. They are now called ball–box systems.

## Other Topics

A quantized version of the inverse-scattering method was initiated by Faddeev and his co-workers, which makes a connection with two-dimensional solvable lattice models and produced the notion of quantum groups. Through the Bethe ansatz, another relation of two-dimensional lattice models and ball–box systems has been discussed.

*See also:* Affine Quantum Groups; Bäcklund Transformations; Bi-Hamiltonian Methods in Soliton Theory; Coherent States; Current Algebra; Integrable Systems and Algebraic Geometry; Integrable Systems: Overview; Multi-Hamiltonian Systems; Painlevé Equations; Partial Differential Equations: Some Examples; *q*-Special Functions; Recursion Operators in Classical Mechanics; Sine-Gordon Equation; Toda Lattices.

## Further Reading

Cherednik I (1996) *Basic Methods of Soliton Theory*. Advanced Series in Mathematical Physics, vol. 25. Singapore, New Jersey, London and Hong Kong: World Scientific.

Date E, Kashiwara M, Jimbo M, and Miwa T (1983) Transformation groups for soliton equations. In: Jimbo M and Miwa T (eds.) *Proceedings of RIMS Symposium on Non-Linear Integrable Systems – Classical Theory and Quantum theory*, pp. 39–119. Singapore: World Scientific.

Drinfel'd VG and Sokolov VV (1985) Lie algebras and equations of Korteweg–de Vries type. *Journal of Soviet Mathematics* 30: 1975–2036.

Gardner CS, Greene JM, Kruskal MD, and Miura RM (1967) Methods for solving the Korteweg–de Vries equation. *Physical Review Letter* 19: 1095–1097.

Kac VG (1990) *Infinite Dimensional Lie Algebras*, 3rd edition. Cambridge: Cambridge University Press.

Manin YuI (1979) Algebraic aspects of nonlinear differential equations. *Journal of Soviet Mathematics* 11: 1–122.

Miwa T, Jimbo M, and Date E (2000) *Solitons*, (Translated by Reid, M). Cambridge: Cambridge University Press.

Noumi M (2002) Affine Weyl group approach to Painlevé equations. In: Tatsien LI (ed.) *International Congress of Mathematicians (2002, Beijing)*, Proceedings of the International Congress of Mathematicians, August 20–28, 2002, Beijing, pp. 497–509. Beijing: Higher Education Press.

Novikov SP, Manakov SV, Pitaevskii LP, and Zakharov VE (1984) *Theory of Solitons. The Inverse Scattering Method*. New York and London: Consultants Bureau.

Sato M and Sato Y (1983) Soliton equations as dynamical systems on infinite-dimensional Grassmann manifold. In: Fujita H, Lax PD, and Strang G (eds.) *Nonlinear Partial Differential Equations in Applied Science (Tokyo, 1982)*, North-Holland Math. Stud., vol. 81, pp. 259–271. Amsterdam: North-Holland.

# Solitons and Other Extended Field Configurations

**R S Ward**, University of Durham, Durham, UK



## Introduction

A soliton is a localized lump (or string or wall, etc.) of energy, which can move without distortion, dispersion, or dissipation, and which is stable under perturbations (and collisions with other solitons). The word was coined by Zabusky and Kruskal in 1965 to describe a solitary wave with particle-like properties (as in electron, proton, etc.). Solitons are relevant to numerous areas of physics – condensed matter, cosmology, fluids/plasmas, biophysics (e.g., DNA), nuclear physics, high-energy physics, etc. Mathematically, they are modeled as solutions of appropriate partial differential equations.

Systems which admit solitons may be classified according to the mechanism by which stability is ensured. Such mechanisms include complete integrability, nontrivial topology plus dynamical balancing, and *Q*-balls/breathers.

Sometimes the term "soliton" is used in a restricted sense, to refer to stable localized lumps which have purely elastic interactions: solitons which collide without any radiation being emitted. This is possible only in very special systems, namely, those that are completely integrable. For these systems, soliton stability (and the elasticity of collisions) arises from a number of characteristic properties, including a precise balance between dispersion and nonlinearity, solvability by the inverse scattering transform from linear data, infinitely many conserved quantities, a Lax formulation (associated linear problem), and Bäcklund transformations. Examples of such integrable soliton systems are the sine-Gordon, Korteweg–deVries, and nonlinear Schrödinger equations.

The category of topological solitons is the most varied, and includes such examples as kinks, vortices, monopoles, skyrmions, and instantons. The requirement of dynamical balancing for these can be understood in terms of Derrick's theorem, which provides necessary conditions for a classical field theory to admit static localized solutions. The

Derrick argument involves studying what happens to the energy of a field when one changes the scale of space. If one has a scalar field (or multiplet of scalar fields) $\phi$, and/or a gauge field $F_{\mu\nu}$, then the static energy $E$ is the sum of terms such as

$$E_0 = \int V(\phi)\,\mathrm{d}^n x, \qquad E_d = \int T_d(D_j\phi)\,\mathrm{d}^n x,$$
$$E_F = \int F_{jk}F_{jk}\,\mathrm{d}^n x$$

where each integral is over ($n$-dimensional) space $\boldsymbol{R}^n$, $D_j\phi$ denotes the covariant spatial derivative of $\phi$, and $T_d(\xi_j)$ is a real-valued polynomial of degree $d$. In particular, for example, we could have $T_2(D_j\phi) = (D_j\phi)(D_j\phi)$, the standard gradient term. Under the dilation $x^j \mapsto \lambda x^j$, these functionals transform as

$$E_0 \mapsto \lambda^{-n}E_0, \qquad E_d \mapsto \lambda^{d-n}E_d, \qquad E_F \mapsto \lambda^{4-n}E_F$$

In order to have a static solution (critical point of the static energy functional), one needs to have a zero exponent on $\lambda$, and/or a balance between positive and negative exponents. A negative exponent indicates a compressing force (tending to implode a localized lump), whereas a positive exponent indicates an expanding force; so to have a static lump solution, these two forces have to balance each other. For $n=1$, a system involving only a scalar field, with terms of the form $E_0$ and $E_2$, can admit static solitons (e.g., kinks); the scaling argument implies a virial theorem, which in this case says that $E_0 = E_2$. For $n=2$, one can have a scalar system with only $E_2$, since in this case the relevant exponent is zero (e.g., the two-dimensional sigma model). Another $n=2$ example is that of vortices in the abelian Higgs model, where the energy contains terms $E_0, E_2$, and $E_F$. For $n=3$, interesting systems have $E_2$ together with either $E_4$ (e.g., skyrmions) or $E_F$ (e.g., monopoles). An $E_0$ term is optional in these cases; its presence affects, in particular, the long-range properties of the solitons. For $n=4$, one can have instantons in a pure gauge theory (term $E_F$ only).

It should be noted that if there are no restrictions on the fields $\phi$ and $A_j$ (such as those arising, e.g., from nontrivial topology), then there is a more obvious mode of instability, which will inevitably be present: $\phi \mapsto \mu\phi$ and/or $A_j \mapsto \mu A_j$, where $0 \le \mu \le 1$. In other words, the fields can simply be scaled away altogether, so that the height of the soliton (and its energy) go smoothly to zero. This can be prevented by nontrivial topology.

Another way of preventing solitons from shrinking is to allow the field to have some "internal" time dependence, so that it is stationary rather than static. For example, one could allow the complex scalar field $\phi$ to have the form $\phi = \psi \exp(\mathrm{i}\omega t)$, where $\psi$ is independent of time $t$. This leads to something like a centrifugal force, which can have a stabilizing effect in the absence of Skyrme or magnetic terms. The corresponding solitons are $Q$-balls.

## Kinks and Breathers

The simplest topological solitons are kinks, in systems involving a real-valued scalar field $\phi(x)$ in one spatial dimension. The dynamics is governed by the Lagrangian density

$$\mathcal{L} = \tfrac{1}{2}\left[(\phi_t)^2 - (\phi_x)^2 - W(\phi)^2\right]$$

where $W(\phi)$ is a (fixed) smooth function. The system can admit kinks if $W(\phi)$ has at least two zeros, for example, $W(A) = W(B) = 0$ with $W(\phi) > 0$ for $A < \phi < B$. Two well-known systems are: sine-Gordon (where $W(\phi) = 2\sin(\phi/2), A = 0$, and $B = 2\pi$) and $\phi^4$ (where $W(\phi) = 1 - \phi^2, A = -1$, and $B = 1$). The corresponding field equations are the Euler–Lagrange equations for $\mathcal{L}$; for example, the sine-Gordon equation is

$$\phi_{tt} - \phi_{xx} + \sin\phi = 0 \qquad [1]$$

Configurations satisfying the boundary conditions $\phi \to A$ as $x \to -\infty$ and $\phi \to B$ as $x \to \infty$ are called kinks (and the corresponding ones with $x = \infty$ and $x = -\infty$ interchanged are antikinks). For kink (or antikink) configurations, there is a lower bound, called the Bogomol'nyi bound, on the static energy $E[\phi]$; for kink boundary conditions, we have

$$\begin{aligned} E[\phi] &= \frac{1}{2}\int_{-\infty}^{\infty}\left[(\phi_x)^2 + W(\phi)^2\right]\mathrm{d}x \\ &= \frac{1}{2}\int_{-\infty}^{\infty}[\phi_x - W(\phi)]^2\,\mathrm{d}x + \int_{-\infty}^{\infty} W(\phi)\phi_x\,\mathrm{d}x \\ &\ge \int_A^B W(\phi)\,\mathrm{d}\phi \end{aligned}$$

with equality if and only if the Bogomol'nyi equation

$$\frac{\mathrm{d}\phi}{\mathrm{d}x} = W(\phi) \qquad [2]$$

is satisfied. A static solution of the Bogomol'nyi equation is a kink solution – it is a static minimum of the energy functional in the kink sector. For example, for the sine-Gordon system, we get $E[\phi] \ge 8$, with equality for the sine-Gordon kink

$$\phi(x) = 4\tan^{-1}\exp(x - x_0)$$

while for the $\phi^4$ system, we get $E[\phi] \ge 4/3$, with equality for the phi-four kink

$$\phi(x) = \tanh(x - x_0)$$

These kinks are stable topological solitons; the nontrivial topology corresponds to the fact that the boundary value of $\phi(t,x)$ at $x=\infty$ is different from that at $x=-\infty$. With trivial boundary conditions (say $\phi \to A$ as $x \to \pm\infty$), stable static solitons are unlikely to exist, but solitons with periodic time dependence (which in this context are called breathers) may exist. For example, the sine-Gordon equation and the nonlinear Schrödinger equation, both, admit breathers – but these owe their existence to complete integrability. By contrast, the $\phi^4$ system (which is not integrable) does not admit breathers; a collision between a $\phi^4$ kink and an antikink (with suitable impact speed) produces a long-lived state which looks like a breather, but eventually decays into radiation.

In lattice systems, however, breathers are more generic. In a one-dimensional lattice system, the continuous space $\boldsymbol{R}$ is replaced by the lattice $\boldsymbol{Z}$, so $\phi(t,x)$ is replaced by $\phi_n(t)$, where $n \in \boldsymbol{Z}$. The Lagrangian is

$$L = \frac{1}{2}\sum_n \left[(\dot{\phi}_n)^2 - h^{-2}(\phi_{n+1} - \phi_n)^2 - W(\phi_n)\right]$$

where $h$ is a positive parameter, corresponding to the dimensionless ratio between the lattice spacing and the size of a kink. The continuum limit is $h \to 0$. This system admits kink solutions as in the continuum case; and for $h$ large enough, it admits breathers as well, but these disappear as $h$ becomes small.

Interpreted in three dimensions, the kink becomes a domain wall separating two regions in which the order parameter $\phi$ takes distinct values; this has applications in such diverse areas as cosmology and condensed matter physics.

## Sigma Models and Skyrmions

In a sigma model or Skyrme system, the field is a map $\phi$ from spacetime to a Riemannian manifold $M$; generally, $M$ is taken to be a Lie group or a symmetric space. The energy density of a static field can be constructed as follows (the Lorentz-invariant extension of this gives a relativistic Lagrangian for fields on spacetime). Let $\phi^a$ be local coordinates on the $m$-dimensional manifold $M$, let $h_{ab}$ denote the metric of $M$, and let $x^j$ denote the spatial coordinates on space $\boldsymbol{R}^n$. An $m \times m$ matrix $D$ is defined by

$$D_a{}^b = (\partial_j\phi^c)h_{ac}(\partial_j\phi^b)$$

where $\partial_j$ denotes derivatives with respect to the $x^j$. Then the invariants $\mathcal{E}_2 = \mathrm{tr}(D) = |\partial_j\phi^a|^2$ and $\mathcal{E}_4 = (1/2)[(\mathrm{tr}\, D)^2 - \mathrm{tr}(D^2)]$ can be terms in the energy density, as well as a zeroth-order term $\mathcal{E}_0 = V(\phi^a)$ not involving derivatives of $\phi$. A term of the form $\mathcal{E}_4$ is called a Skyrme term.

The boundary condition on field configurations is that $\phi$ tends to some constant value $\phi_0 \in M$ as $|x| \to \infty$ in $\boldsymbol{R}^n$. From the topological point of view, this compactifies $\boldsymbol{R}^n$ to $S^n$. In other words, $\phi$ extends to a map from $S^n$ to $M$; and such maps are classified topologically by the homotopy group $\pi_n(M)$. For topological solitons to exist, this group has to be nontrivial.

In one spatial dimension ($n=1$) with $M=S^1$ (say), the expression $\mathcal{E}_4$ is identically zero, and we just have kink-type systems such as sine-Gordon. The simplest two-dimensional example ($n=2$) is the $O(3)$ sigma model, which has $M=S^2$ with its standard metric. In this system, the field is often expressed as a unit 3-vector field $\boldsymbol{\phi} = (\phi^1, \phi^2, \phi^3)$, with $\mathcal{E}_2 = (\partial_j\boldsymbol{\phi}) \cdot (\partial_j\boldsymbol{\phi})$. Here the configurations are classified topologically by their degree (or winding number, or topological charge) $N \in \pi_2(S^2) \cong \boldsymbol{Z}$, which equals

$$N = \frac{1}{4\pi}\int \boldsymbol{\phi} \cdot \partial_1\boldsymbol{\phi} \times \partial_2\boldsymbol{\phi}\, \mathrm{d}x^1\, \mathrm{d}x^2$$

Instead of $\boldsymbol{\phi}$, it is often convenient to use a single complex-valued function $W$ related to $\boldsymbol{\phi}$ by the stereographic projection $W = (\phi^1 + \mathrm{i}\phi^2)/(1-\phi^3)$. In terms of $W$, the formula for the degree $N$ is

$$N = \frac{\mathrm{i}}{2\pi}\int \frac{W_1\overline{W}_2 - W_2\overline{W}_1}{(1+|W|^2)^2}\, \mathrm{d}x^1\, \mathrm{d}x^2$$

and the static energy is (with $z = x^1 + \mathrm{i}x^2$)

$$\begin{aligned} E &= \int \mathcal{E}_2\, \mathrm{d}^2x \\ &= 8\int \frac{|W_z|^2 + |W_{\bar z}|^2}{(1+|W|^2)^2}\, \mathrm{d}^2x \\ &= 16\int \frac{|W_{\bar z}|^2}{(1+|W|^2)^2}\, \mathrm{d}^2x + 8\int \frac{|W_z|^2 - |W_{\bar z}|^2}{(1+|W|^2)^2}\, \mathrm{d}^2x \\ &= 16\int \frac{|W_{\bar z}|^2}{(1+|W|^2)^2}\, \mathrm{d}^2x + 8\pi N \end{aligned}$$

From this, one sees that $E$ satisfies the Bogomol'nyi bound $E \geq 8\pi N$, and that minimal-energy solutions correspond to solutions of the Cauchy–Riemann equations $W_{\bar z} = 0$. To have finite energy, $W(z)$ has to be a rational function, and so solutions with winding number $N$ correspond to rational meromorphic functions $W(z)$, of degree $|N|$. (If $N < 0$, then $W$ is a rational function of $\bar z$.) The energy is scale invariant (conformally invariant), and consequently these solutions are not solitons – they are not quite stable, since their size is not fixed. Adding terms $\mathcal{E}_4$ and $\mathcal{E}_0$

to the energy density fixes the soliton size, and the resulting two-dimensional Skyrme systems admit true topological solitons.

The three-dimensional case ($n=3$), with $M$ being a simple Lie group, is the original Skyrme model of nuclear physics. If $M=\mathrm{SU}(2)$, then the integer $N \in \pi_3(\mathrm{SU}(2)) \cong \boldsymbol{Z}$ is interpreted as the baryon number. The (quantum) excitations of the $\phi$-field correspond to the pions, whereas the (semiclassical) solitons correspond to the nucleons. This model emerges as an effective theory of quantum chromodynamics (QCD), in the limit where the number of colors is large. If we express the field as a function $U(x^j)$ taking values in a Lie group, then $L_j = U^{-1}\partial_j U$ takes values in the corresponding Lie algebra, and $\mathcal{E}_2$ and $\mathcal{E}_4$ take the form

$$\begin{aligned}\mathcal{E}_2 &= -\tfrac{1}{2}\,\mathrm{tr}(L_j L_j)\\ \mathcal{E}_4 &= -\tfrac{1}{16}\,\mathrm{tr}\big([L_j, L_k][L_j, L_k]\big)\end{aligned}$$

The static energy density in the basic Skyrme system is the sum of these two terms. The static energy satisfies a Bogomol'nyi bound $E \geq 12\pi^2|N|$, and it is believed that stable solitons (skyrmions) exist for each value of $N$. Classical skyrmions have been investigated numerically; for values of $N$ up to $\sim 25$, they turn out to resemble polyhedral shells. Comparison with nucleon phenomenology requires semiclassical quantization, and this leads to results which are at least qualitatively correct.

A variant of the Skyrme model is the Skyrme–Faddeev system, which has $n=3$ and $M=S^2$; the solitons in this case resemble loops which can be linked or knotted, and which are classified by their Hopf number $N \in \pi_3(S^2)$. In this case, the energy satisfies a lower bound of the form $E \geq cN^{3/4}$. Numerical experiments indicate that for each $N$, there is a minimal-energy solution with Hopf number $N$, and with energy close to this topological lower bound.

## Abelian Higgs Vortices

Vortices live in two spatial dimensions; viewed in three dimensions, they are string-like. Two of their applications are as cosmic strings and as magnetic flux tubes in superconductors. They occur as static topological solitons in the the abelian Higgs model (or Ginzburg–Landau model), and involve a magnetic field $B=\partial_1 A_2 - \partial_2 A_1$, coupled to a complex scalar field $\phi$, on the plane $\boldsymbol{R}^2$. The energy density is

$$\mathcal{E} = \tfrac{1}{2}(D_j\phi)(\overline{D_j\phi}) + \tfrac{1}{2}B^2 + \tfrac{1}{8}\lambda(1-|\phi|^2)^2 \qquad [3]$$

where $D_j\phi := \partial_j\phi - \mathrm{i}A_j\phi$, and where $\lambda$ is a positive constant. The boundary conditions are

$$D_j\phi = 0, \qquad B = 0, \qquad |\phi| = 1 \qquad [4]$$

as $r \to \infty$. If we consider a very large circle $C$ on $\boldsymbol{R}^2$, so that [4] holds on C, then $\phi|_C$ is a map from the circle $C$ to the circle of unit radius in the complex plane, and therefore it has an integer winding number $N$. Thus configurations are labeled by this vortex number $N$.

Note that if $\mathcal{E}$ vanishes, then $B=0$ and $|\phi|=1$: the gauge symmetry is spontaneously broken, and the photon "acquires a mass": this is a standard example of spontaneous symmetry breaking.

The total magnetic flux $\int B\,\mathrm{d}^2x$ equals $2\pi N$; a proof of this is as follows. Let $\theta$ be the usual polar coordinate around C. Because $|\phi|=1$ on C, we can write $\phi = \exp[\mathrm{i}f(\theta)]$ for some function $f$; this $f$ need not be single-valued, but must satisfy $f(2\pi) - f(0) = 2\pi N$ with $N$ being an integer (in order that $\phi$ be single-valued). In fact, this defines the winding number. Now since $D_j\phi = \partial_j\phi - \mathrm{i}A_j\phi = 0$ on C, we have

$$A_j = -\mathrm{i}\phi^{-1}\partial_j\phi = \partial_j f$$

on $C$. So, using Stokes' theorem, we get

$$\begin{aligned}\int_{\boldsymbol{R}^2} B\,\mathrm{d}^2x &= \int_C A_j\,\mathrm{d}x^j\\ &= \int_0^{2\pi} \frac{\mathrm{d}f}{\mathrm{d}\theta}\,\mathrm{d}\theta\\ &= 2\pi N\end{aligned}$$

If $\lambda=1$, then the total energy $E=\int \mathcal{E}\,\mathrm{d}^2x$ satisfies the Bogomol'nyi bound $E \geq \pi N; E=\pi N$ if and only if a set of partial differential equations (the Bogomol'nyi equations) are satisfied. Since like charges repel, the magnetic force between vortices is repulsive. However, there is also a force from the Higgs field, and this is attractive. The balance between the two forces is determined by $\lambda$: if $\lambda > 1$, the vortices repel each other; whereas if $\lambda < 1$, the vortices attract. In the critical case $\lambda = 1$, the force between vortices is exactly balanced, and there exist static multivortex solutions. In fact, one has the following: given $N$ points in the plane, there exists an $N$-vortex solution of the Bogomol'nyi equations (and hence of the full field equations) with $\phi$ vanishing at the chosen points (and nowhere else). All static solutions are of this form. These solutions cannot, however, be written down explicitly in terms of elementary functions (except of course for $N=0$).

## Monopoles

The abelian Higgs model does not admit three-dimensional solitons, but a nonabelian generalization does – such nonabelian Higgs solitons are called magnetic monopoles. The field content, in the simplest version, is as follows. First, there is a gauge (Yang–Mills) field $F_{\mu\nu}$, with gauge potential $A_\mu$, and with the gauge group being a simple Lie group $G$. Second, there is a Higgs scalar field $\phi$, transforming under the adjoint representation of $G$ (thus $\phi$ takes values in the Lie algebra of $G$). For simplicity, $G$ is taken to be SU(2) in what follows. So we may write $A_\mu = \mathrm{i}A_\mu^a \sigma_a, F_{\mu\nu} = \mathrm{i}F_{\mu\nu}^a \sigma_a$, and $\phi = \mathrm{i}\phi^a \sigma_a$, where $\sigma_a$ are the Pauli matrices. The energy of static ($\partial_0 \phi = 0 = \partial_0 A_j$), purely magnetic ($A_0 = 0$) configurations is

$$E = \int \left[\tfrac{1}{2} B_j^a B_j^a + \tfrac{1}{2}(D_j\phi)^a (D_j\phi)^a + \tfrac{1}{4}\lambda(1 - \phi^a\phi^a)^2\right] \mathrm{d}^3 x$$

where $B_j^a = (1/2)\epsilon_{jkl} F_{kl}$ is the magnetic field. The boundary conditions are $B_j^a \to 0$ and $\phi^a\phi^a \to 1$ as $r \to \infty$; so $\phi$ restricted to a large spatial 2-sphere becomes a map from $S^2$ to the unit 2-sphere in the Lie algebra su(2), and as such it has a degree $N \in \mathbf{Z}$. An analytic expression for $N$ is

$$\int B_j^a (D_j\phi)^a \, \mathrm{d}^3 x = 2\pi N \qquad [5]$$

At long range, the field resembles an isolated magnetic pole (a Dirac magnetic monopole), with magnetic charge $2\pi N$. Asymptotically, the SU(2) gauge symmetry is spontaneously broken to U(1), which is interpreted as the electromagnetic gauge group.

In 1974, it was observed that this system admits a smooth, finite-energy, stable, spherically symmetric $N = 1$ solution – this is the 't Hooft–Polyakov monopole. There is a Bogomol'nyi lower bound on the energy $E$: from $0 \le (B + D\phi)^2 = B^2 + (D\phi)^2 + 2B \cdot D\phi$, we get

$$E \ge 2\pi N + \int \tfrac{1}{4}\lambda(1 - \phi^a\phi^a)^2 \, \mathrm{d}^3 x \qquad [6]$$

where [5] has been used. The inequality [6] is saturated if and only if the Prasad–Sommerfield limit $\lambda = 0$ is used, and the Bogomol'nyi equations

$$(D_j\phi)^a = -B_j^a \qquad [7]$$

hold. The corresponding solitons are called Bogomol'nyi–Prasad–Sommerfield (BPS) monopoles.

The Bogomol'nyi equations [7], together with the boundary conditions described above, form a completely integrable elliptic system of partial differential equations. For any positive integer $N$, the space of BPS monopoles of charge $N$, with gauge freedom factored out, is parametrized by a $(4N - 1)$-dimensional manifold $\mathcal{M}_N$. This is the moduli space of $N$ monopoles. Roughly speaking, each monopole has a position in space (three parameters) plus a phase (one parameter), making a total of $4|N|$ parameters; an overall phase can be removed by a gauge transformation, leaving $(4|N| - 1)$ parameters. In fact, it is often useful to retain the overall phase, and to work with the corresponding $4|N|$-dimensional manifold $\widetilde{\mathcal{M}}_N$. This manifold has a natural metric, which corresponds to the expression for the kinetic energy of the system. A point in $\widetilde{\mathcal{M}}_N$ represents an $N$-monopole configuration, and the slow-motion dynamics of $N$ monopoles corresponds to geodesics on $\widetilde{\mathcal{M}}_N$; this is the geodesic approximation of monopole dynamics.

The $N = 1$ monopole is spherically symmetric, and the corresponding fields take a simple form; for example, the Higgs field of a 1-monopole located at $r = 0$ is

$$\phi^a = \left[\frac{\coth(2r)}{r} - \frac{1}{2r^2}\right] x^a$$

For $N > 1$, the expressions tend to be less explicit; but monopole solutions can nevertheless be characterized in a fairly complete way. The Bogomol'nyi equations [7] are a dimensional reduction of the self-dual Yang–Mills equations in $\mathbf{R}^4$, and BPS monopoles correspond to holomorphic vector bundles over a certain two-dimensional complex manifold ("mini-twistor space"). This leads to various other characterizations of monopole solutions, for example, in terms of certain curves ("spectral curves") on mini-twistor space, and in terms of solutions of a set of ordinary differential equations called the Nahm equations. Having all these descriptions enables one to deduce much about the monopole moduli space, and to characterize many monopole solutions. In particular, there are explicit solutions of the Nahm equations involving elliptic functions, which correspond to monopoles with certain discrete symmetries, such as a 3-monopole with tetrahedral symmetry, and a 4-monopole with the appearance and symmetries of a cube.

## Yang–Mills Instantons

Consider gauge fields in four-dimensional Euclidean space $\mathbf{R}^4$, with gauge group $G$. For simplicity, in what follows, $G$ is taken to be SU(2); one can extend much of the structure to more general groups, for example, the simple Lie groups. Let $A_\mu$ and $F_{\mu\nu}$

denote the gauge potential and gauge field. The Yang–Mills action is

$$S = -\frac{1}{4}\int \mathrm{tr}\big(F_{\mu\nu}F_{\mu\nu}\big)\,\mathrm{d}^4x \qquad [8]$$

where we assume a boundary condition, at infinity in $\boldsymbol{R}^4$, such that this integral converges. The Euler–Lagrange equations which describe critical points of the functional $S$ are the Yang–Mills equations

$$D_\mu F_{\mu\nu} = 0 \qquad [9]$$

Finite-action Yang–Mills fields are called instantons. The Euclidean action [8] is used in the path-integral approach to quantum gauge field theory; therefore, instantons are crucial in understanding the path integral.

The dual of the field tensor $F_{\mu\nu}$ is

$$*F_{\mu\nu} = \tfrac{1}{2}\varepsilon_{\mu\nu\alpha\beta}F_{\alpha\beta}$$

The gauge field is self-dual if $*F_{\mu\nu} = F_{\mu\nu}$, and anti-self-dual if $*F_{\mu\nu} = -F_{\mu\nu}$. In view of the Bianchi identity $D_\mu * F_{\mu\nu} = 0$, any self-dual or anti-self-dual gauge field is automatically a solution of the Yang–Mills equations [9]. This fact also follows from the discussion below, where we see that self-dual instantons give local minima of the action.

The Yang–Mills action (and Yang–Mills equations) are conformally invariant; any finite-action solution of the Yang–Mills equations on $R^4$ extends smoothly to the conformal compactification $S^4$. Gauge fields on $S^4$, with gauge group SU(2), are classified topologically by an integer $N$, namely, the second Chern number

$$N = c_2 = -\frac{1}{8\pi^2}\int \mathrm{tr}\big(F_{\mu\nu} * F_{\mu\nu}\big)\,\mathrm{d}^4x \qquad [10]$$

From [8] and [10] a topological lower bound on the action is given as follows:

$$\begin{aligned} 0 &\leq -\int \mathrm{tr}\big(F_{\mu\nu} - *F_{\mu\nu}\big)\big(F_{\mu\nu} - *F_{\mu\nu}\big)\,\mathrm{d}^4x \\ &= 8S - 16\pi^2 N \end{aligned}$$

and so $S \geq 2\pi^2 N$, with equality if and only if the field is self-dual. If $N < 0$, we get $S \geq 2\pi^2|N|$, with equality if and only if $F$ is anti-self-dual. So the self-dual (or anti-self-dual) fields minimize the action in each topological class.

For the remainder of this section, we restrict to self-dual instantons with instanton number $N > 0$. The space (moduli space) of such instantons, with gauge equivalence factored out, is an $(8N-3)$-dimensional real manifold. In principle, all these gauge fields can be constructed using algebraic-geometry (twistor) methods: instantons correspond to holomorphic vector bundles over complex projective 3-space (twistor space). One large class of solutions which can be written out explicitly is as follows: for $N = 1$ and $N = 2$ it gives all instantons, while for $N \geq 3$ it gives a $(5N+4)$-dimensional subfamily of the full $(8N-3)$-dimensional solution space. The gauge potentials in this class have the form

$$A_\mu = \mathrm{i}\sigma_{\mu\nu}\partial_\nu \log \phi \qquad [11]$$

where the $\sigma_{\mu\nu}$ are constant matrices (antisymmetric in $\mu\nu$) defined in terms of the Pauli matrices $\sigma_a$ by

$$\begin{aligned} \sigma_{10} &= \sigma_{23} = \tfrac{1}{2}\sigma_1 \\ \sigma_{20} &= \sigma_{31} = \tfrac{1}{2}\sigma_2 \\ \sigma_{30} &= \sigma_{12} = \tfrac{1}{2}\sigma_3 \end{aligned}$$

The real-valued function $\phi = \phi(x^\mu)$ is a solution of the four-dimensional Laplace equation given by

$$\phi(x^\mu) = \sum_{k=0}^{N} \frac{\lambda_k}{(x^\mu - x_k^\mu)(x^\mu - x_k^\mu)}$$

where the $x_k^\mu$ are $N+1$ distinct points in $\boldsymbol{R}^4$, and the $\lambda_k$ are $N+1$ positive constants: a total of $5N+5$ parameters. It is clear from [11] that the overall scale of $\phi$ is irrelevant, leaving a $(5N+4)$-parameter family. For $N = 1$ and $N = 2$, symmetries reduce the parameter count further, to 5 and 13, respectively. Although $\phi$ has poles at the points $x = x_k$, the gauge potentials are smooth (possibly after a gauge transformation).

Finally, it is worth noting that (as one might expect) there is a gravitational analog of the gauge-theoretic structures described here. In other words, one has self-dual gravitational instantons – these are four-dimensional Riemannian spaces for which the conformal-curvature tensor (the Weyl tensor) is self-dual, and the Ricci tensor satisfies Einstein's equations $R_{\mu\nu} = \Lambda g_{\mu\nu}$. As before, such spaces can be constructed using a twistor-geometrical correspondence.

## Q-Balls

A $Q$-ball (or nontopological soliton) is a soliton which has a periodic time dependence in a degree of freedom which corresponds to a global symmetry. The simplest class of $Q$-ball systems involves a complex scalar field $\phi$, with an invariance under the constant phase transformation $\phi \mapsto \mathrm{e}^{\mathrm{i}\theta}\phi$; the $Q$-balls are soliton solutions of the form

$$\phi(t, \boldsymbol{x}) = \mathrm{e}^{\mathrm{i}\omega t}\psi(\boldsymbol{x}) \qquad [12]$$

where $\psi(\boldsymbol{x})$ is a complex scalar field depending only on the spatial variables $\boldsymbol{x}$. The best-known case is the 1-soliton solution

$$\phi(t,x) = a\sqrt{2}\exp(\mathrm{i}a^2 t)\mathrm{sech}(ax)$$

of the nonlinear Schrödinger equation $\mathrm{i}\phi_t + \phi_{xx} + \phi|\phi|^2 = 0$.

More generally, consider a system (in $n$ spatial dimensions) with Lagrangian

$$\mathcal{L} = \tfrac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - U(|\phi|)$$

where $\phi(x^\mu)$ is a complex-valued field. Associated with the global phase symmetry is the conserved Noether charge $Q = \int \mathrm{Im}(\bar{\phi}\phi_t)\,\mathrm{d}^n x$. Minimizing the energy of a configuration subject to $Q$ being fixed implies that $\phi$ has the form [12]. Without loss of generality, we may take $\omega \geq 0$. Note that $Q = \omega I$, where $I = \int |\psi|^2\,\mathrm{d}^n x$. The energy of a configuration of the form [12] is $E = E_q + E_k + E_p$, where

$$E_q = \frac{1}{2}\int |\partial_j \psi|^2\,\mathrm{d}^n x$$

$$E_k = \tfrac{1}{2} I\omega^2 = \tfrac{1}{2} Q^2/I$$

$$E_p = \int U(|\psi|)\,\mathrm{d}^n x$$

Let us take $U(0) = 0 = U'(0)$, with the field satisfying the boundary condition $\psi \to 0$ as $r \to \infty$.

A stationary $Q$-lump is a critical point of the energy functional $E[\psi]$, subject to $Q$ having some fixed value. The usual (Derrick) scaling argument shows that any stationary $Q$-lump must satisfy

$$(2-n)E_q - nE_p + nE_k = 0 \qquad [13]$$

For simplicity, in what follows, let us take $n \geq 3$. Define $m > 0$ by $U''(0) = m^2$; then, near spatial infinity, the Euler–Lagrange equations give $\nabla^2\psi - (m^2 - \omega^2)\psi = 0$. So, in order to satisfy the boundary condition $\psi \to 0$ as $r \to \infty$, we need $\omega < m$.

It is clear from [13] that if $U \geq (1/2)m^2|\psi|^2$ everywhere, then there can be no solution. So $K = \min[2U(|\psi|)/|\psi|^2]$ has to satisfy $K < m^2$. Also, we have

$$E_p = \int U \geq \tfrac{1}{2}KI = (K/\omega^2)E_k > (K/\omega^2)E_p \qquad [14]$$

where the final inequality comes from [13]. As a consequence, we see that $\omega^2$ is restricted to the range

$$K < \omega^2 < m^2 \qquad [15]$$

An example which has been studied in some detail is $\mathrm{U}(f) = f^2[1 + (1-f^2)^2]$; here $m^2 = 4$ and $K = 2$, so the range of frequency for $Q$-balls in this system is $\sqrt{2} < \omega < 2$. The dynamics of $Q$-balls in systems such as these turns out to be quite complicated.

*See also:* Abelian Higgs Vortices; Homoclinic Phenomena; Integrable Systems: Overview; Instantons: Topological Aspects; Noncommutative Geometry from Strings; Sine-Gordon Equation; Topological Defects and Their Homotopy Classification.

## Further Reading

Atiyah MF and Hitchin NJ (1988) *The Geometry and Dynamics of Magnetic Monopoles*. Princeton: Princeton University Press.

Coleman S (1988) *Aspects of Symmetry*. Cambridge: Cambridge University Press.

Drazin PG and Johnson RS (1989) *Solitons: An Introduction*. Cambridge: Cambridge University Press.

Goddard P and Mansfield P (1986) Topological structures in field theories. *Reports on Progress in Physics* 49: 725–781.

Jaffe A and Taubes C (1980) *Vortices and Monopoles*. Boston: Birkhäuser.

Lee TD and Pang Y (1992) Nontopological solitons. *Physics Reports* 221: 251–350.

Makhankov VG, Rubakov YP, and Sanyuk VI (1993) *The Skyrme Model: Fundamentals, Methods, Applications*. Berlin: Springer.

Manton NS and Sutcliffe PM (2004) *Topological Solitons*. Cambridge: Cambridge University Press.

Rajaraman R (1982) *Soliton and Instantons*. New York: North-Holland.

Rebbi C and Soliani G (1984) *Solitons and Particles*. Singapore: World Scientific.

Vilenkin A and Shellard EPS (1994) *Cosmic Strings and Other Cosmological Defects*. Cambridge: Cambridge University Press.

Ward RS and Wells RO Jr. (1990) *Twistor Geometry and Field Theory*. Cambridge: Cambridge University Press.

Zakrzewski WJ (1989) *Low Dimensional Sigma Models*. Bristol: IOP.

# Source Coding in Quantum Information Theory

**N Datta**, University of Cambridge, Cambridge, UK
**T C Dorlas**, Dublin Institute for Advanced Studies, Dublin, Republic of Ireland



## Introduction

Two key issues of classical and quantum information theory are storage and transmission of information. An information source produces some outputs (or signals) more frequently than others. Due to this redundancy, one can reduce the amount of space needed for its storage without compromising on its content. This data compression is done by a suitable encoding of the output of the source. In contrast, in the transmission of information through a channel, it is often advantageous to add redundancy to a message, in order to combat the effects of noise. This is done in the form of error-correcting codes. The amount of redundancy which needs to be added to the original message depends on how much noise is present in the channel (see, e.g., Nielson and Chuang (2000)). Hence, redundancy plays complementary roles in data compression and transmission of data through a noisy channel. In this review we focus only on data compression in quantum information theory.

In classical information theory, Shannon showed that there is a natural limit to the amount of compression that can be achieved. It is given by the Shannon entropy. The analogous concept in quantum information theory is the von Neumann entropy. Here, we review some of the main results of quantum data compression and the significance of the von Neumann entropy in this context.

The review is structured as follows. We first give a brief introduction to the Shannon entropy and classical data compression. This is followed by a discussion of quantum entropy and the idea behind quantum source coding. We elaborate on data compression schemes for three different classes of quantum sources, namely memoryless sources, ergodic sources, and sources modeled by Gibbs states of quantum spin systems. In the bulk of the review, we concentrate on source-dependent, fixed-length coding schemes. We conclude with a brief discussion of universal and variable-length coding. We would like to point out that this review article is by no means complete. Due to a restriction on its length, we had to leave out various important aspects and developments of quantum source coding.

## Classical Data Compression

### Entropy and Source Coding

A simple model of a classical information source consists of a sequence of discrete random variables $X_1, X_2, \ldots, X_n$, whose values represent the output of the source. Each random variable $X_i, 1 \le i \le n$, takes values $x_i$ from a finite set, the source alphabet $\mathcal{X}$. Hence, $\underline{X}^{(n)} := (X_1, \ldots, X_n)$ takes values $\underline{x}^{(n)} := (x_1, \ldots, x_n) \in \mathcal{X}^n$. We recall the definition of entropy (or information content) of a source.

If the discrete random variables $X_1, \ldots, X_n$ which take values from a finite alphabet $\mathcal{X}$ have joint probabilities

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = p_n(x_1, \ldots, x_n)$$

then the Shannon entropy of this source is defined by

$$\begin{aligned} H(X_1, \ldots, X_n) \\ = -\sum_{x_1 \in \mathcal{X}} \cdots \sum_{x_n \in \mathcal{X}} p_n(x_1, \ldots, x_n) \\ \times \log p_n(x_1, \ldots, x_n) \end{aligned} \qquad [1]$$

Here and in the following, the logarithm is taken to the base 2. This is because the fundamental unit of classical information is a "bit," which takes two values 0 and 1. Notice that $H(X_1, \ldots, X_n)$ in fact only depends on the (joint) probability mass function (p.m.f.) $p_n$ and can also be denoted as $H(p_n)$.

There are several other concepts of entropy, for example, relative entropy, conditional entropy, and mutual information. See, for example, Cover and Thomas (1991) and Nielson and Chuang (2000). It is easy to see that

1. $0 \le H(X_1, \ldots, X_n) \le n \log |\mathcal{X}|$, where $|\mathcal{X}|$ denotes the number of letters in the alphabet $\mathcal{X}$. Two other important properties are as follows:
2. $H(X_1, \ldots, X_n)$ is jointly concave in $X_1, \ldots, X_n$ and
3. $H(X_1, \ldots, X_n) \le H(X_1, \ldots, X_m) + H(X_{m+1}, \ldots, X_n)$ for $m < n$.

The latter property is called subadditivity.

In the next section, analogous quantities are introduced for quantum information and the corresponding properties are stated.

Suppose that the random variables $X_1, X_2, \ldots, X_n$ are independent and identically distributed (i.i.d.). Then the entropy of each random variable modeling the source is the same and can be denoted by $H(X)$. From the point of view of classical information theory, the Shannon entropy has an important operational definition. It quantifies the minimal

physical resources needed to store data from a classical information source and provides a limit to which data can be compressed reliably (i.e., in a manner in which the original data can be recovered later with a low probability of error). Shannon showed that the original data can be reliably obtained from the compressed version only if the rate of compression is greater than the Shannon entropy. This result is formulated in Shannon's noiseless channel coding theorem (Shannon 1918, Cover and Thomas 1991, Nielson and Chuang 2000) given later.

### The Asymptotic Equipartition Property

The main idea behind Shannon's noiseless channel coding theorem is to divide the possible values $x_1, x_2, \dots, x_n$ of random variables $X_1, \dots, X_n$ into two classes – one consisting of sequences which have a high probability of occurrence, known as "typical sequences," and the other consisting of sequences which occur rarely, known as "atypical sequences." The idea is that there are far fewer typical sequences than the total number of possible sequences, but they occur with high probability. The existence of typical sequences follows from the so-called "asymptotic equipartition property":

**Theorem 1** (AEP). *If $X_1, X_2, X_3, \dots$ are i.i.d. random variables with p.m.f. $p(x)$, then*

$$-\frac{1}{n}\log p_n(X_1, \dots, X_n) \xrightarrow{\mathbb{P}} H(X) \qquad [2]$$

*where $H(X)$ is the Shannon entropy for a single variable, and $p_n(X_1, \dots, X_n)$ denotes the random variable taking values $p_n(x_1, \dots, x_n) = \prod_{i=1}^{n} p(x_i)$ with probabilities $p_n(x_1, \dots, x_n)$.*

This theorem has been generalized to the case of sequences of dependent variables $(X_n)_{n\in\mathbb{Z}}$ which are ergodic for the shift transformation defined below. It is easiest to formulate this for an information stream which extends from $-\infty$ to $+\infty$:

**Definition** A sequence $(X_n)_{n\in\mathbb{Z}}$ is called "stationary" if for any $n_1 < n_2$ and any $x_{n_1}, \dots, x_{n_2} \in \mathcal{X}$,

$$\mathbb{P}(X_{n_1} = x_{n_1}, \dots, X_{n_2} = x_{n_2}) = \mathbb{P}(X_{n_1+1} = x_{n_1}, \dots, X_{n_2+1} = x_{n_2})$$

We define the shift transformation $\tau$ by

$$\tau\big((x_n)_{n\in\mathbb{Z}}\big) = (x'_n)_{n\in\mathbb{Z}}, \quad x'_n = x_{n-1} \qquad [3]$$

Then $(X_n)_{n\in\mathbb{Z}}$ is called "ergodic" if it is stationary and if every subset $A \subset \mathcal{X}^{\mathbb{Z}}$ such that $\tau(A) = A$ has probability 0 or 1, that is, $\mathbb{P}((X_n)_{n\in\mathbb{Z}} \in A) = 0$ or 1.

It is known that $(X_n)_{n\in\mathbb{Z}}$ is ergodic if and only if its probability distribution is extremal in the set of invariant probability measures. The generalization of Theorem 1 (McMillan 1953, Breiman 1957) now reads:

**Theorem 2** (Shannon–McMillan–Breiman theorem). *Suppose that the sequence $(X_n)_{n\in\mathbb{Z}}$ is ergodic. Then*

$$\lim_{n\to\infty}\left\{-\frac{1}{n}\log p_n(X_1, \dots, X_n)\right\} = h_{KS} \quad \text{with probability } 1 \qquad [4]$$

*where $h_{KS}$ is the Kolmogorov–Sinai entropy defined by*

$$h_{\mathrm{KS}} = \lim_{n\to\infty}\frac{1}{n}H(X_1, \dots, X_n) = \inf_n \frac{1}{n}H(X_1, \dots, X_n) \qquad [5]$$

**Remark.** It follows from the subadditivity property (3) above that the sequence $(1/n)H(p_n)$ is decreasing, and it is obviously bounded below by 0.

We now define the set of typical sequences (or more precisely, $\epsilon$-typical sequences) as follows:

**Definition** Let $X_1, \dots, X_n$ be i.i.d. random variables with p.m.f. $p(x)$. Given $\epsilon > 0$, $\epsilon$-typical set $T_\epsilon^{(n)}$ is the set of sequences $(x_1 \dots x_n)$ for which

$$2^{-n(H(X)+\epsilon)} \le p(x_1 \dots x_n) \le 2^{-n(H(X)-\epsilon)} \qquad [6]$$

In the case of an ergodic sequence, $H(X)$ is replaced by $h_{\mathrm{KS}}$ in [6].

Let $|T_\epsilon^{(n)}|$ denote the total number of typical sequences and $\mathbb{P}\{T_\epsilon^{(n)}\}$ denote the probability of the typical set. Then the following is an easy consequence of Theorem 1.

**Theorem 3** (Theorem of typical sequences). *For any $\delta > 0 \,\exists\, n_0(\delta) > 0$ such that $\forall n \ge n_0(\delta)$ the following hold:*

(i) $\mathbb{P}\{T_\epsilon^{(n)}\} > 1 - \delta$ *and*
(ii) $(1-\delta)2^{n(H(X)-\epsilon)} \le |T_\epsilon^{(n)}| \le 2^{n(H(X)+\epsilon)}$

### Shannon's Noiseless Channel Coding Theorem

Shannon's noiseless channel coding theorem is a simple application of the theorem of typical sequences and says that the optimal rate at which one can reliably compress data from an i.i.d. classical information source is given by the Shannon entropy $H(X)$ of the source.

A "compression scheme" $C^n$ of rate $R$ maps possible sequences $\underline{x} = (x_1, \dots, x_n)$ to a binary string of length $\lceil nR \rceil$: $C^n : \underline{x} \mapsto \underline{y} = (y_1, \dots, y_{\lceil nR \rceil})$, where $x_i \in \mathcal{X}$; $|\mathcal{X}| = d$ and $y_i \in \{0,1\}\,\forall\, 1 \le i \le \lceil nR \rceil$. The corresponding decompression scheme takes the $\lceil nR \rceil$

compressed bits and maps them back to a string of $n$ letters from the alphabet $\mathcal{X}$: $D^n : \underline{y} \in \{0,1\}^{\lceil nR \rceil} \mapsto \underline{x}' = (x'_1, \ldots, x'_n)$. A compression–decompression scheme is said to be "reliable" if the probability that $\underline{x}' \neq \underline{x}$ tends to 0 as $n \to \infty$. Shannon's noiseless channel coding theorem (Shannon 1918, Cover and Thomas 1991) now states

**Theorem 4** (Shannon). *Suppose that $\{X_i\}$ is an i.i.d. information source, with $X_i \sim p(x)$ and Shannon entropy $H(X)$. If $R > H(X)$ then there exists a reliable compression scheme of rate $R$ for the source. Conversely, any compression scheme with rate $R < H(X)$ is not reliable.*

**Proof** (sketch). Suppose $R > H(X)$. Choose $\epsilon > 0$ such that $H(X) + \epsilon < R$. Consider the set $T_\epsilon^{(n)}$ of typical sequences. The method of compression is then to examine the output of the source, to see if it belongs to $T_\epsilon^{(n)}$. If the output is a typical sequence, then we compress the data by simply storing an index for the particular sequence using $\lceil nR \rceil$ bits in the obvious way. If the input string is not typical, then we compress the string to some fixed $\lceil nR \rceil$ bit string, for example, $(00\ldots000)$. In this case, data compression effectively fails, but, in spite of this, the compression–decompression scheme succeeds with probability tending to 1 as $n \to \infty$, since by Theorem 3 the probability of atypical sequences can be made small by choosing $n$ large enough.

If $R < H(X)$, then any compression scheme of rate $R$ is not reliable. This also follows from Theorem 3 by the following argument. Let $\mathcal{S}(n)$ be a collection of sequences $\underline{x}^{(n)}$ of size $|\mathcal{S}(n)| \leq 2^{\lceil nR \rceil}$. Then the subset of atypical sequences in $\mathcal{S}(n)$ is highly improbable, whereas the corresponding subset of typical sequences has probability bounded by $2^{nR} 2^{-nH(X)} \to 0$ as $n \to \infty$. □

## Quantum Data Compression

### Quantum Sources and Entropy

In quantum information processing systems, information is stored in quantum states of physical systems. The most general description of a quantum state is provided by a density matrix.

A "density matrix" $\rho$ is a positive semidefinite operator on a Hilbert space $\mathcal{H}$, with $\mathrm{tr}\rho = 1$, and the expected value of an operator $A$ on $\mathcal{H}$ is given by

$$\phi(A) = \mathrm{tr}\,(\rho A) \qquad [7]$$

The functional $\phi$ on $\mathcal{M} = \mathcal{B}(\mathcal{H})$, the algebra of linear operators on $\mathcal{H}$, is positive (i.e., $\phi(A) \geq 0$, if $A \geq 0$) and maps the identity $\boldsymbol{I} \in \mathcal{M}$ to 1. Such a functional is also called a state. Conversely, given such a state on a finite-dimensional algebra $\mathcal{M}$, there exists a unique density matrix $\rho_\phi$ such that [7] holds, so the concepts can be used interchangeably. (This is not true in the infinite-dimensional case.)

The quantum analog of the Shannon entropy is called the von Neumann entropy. For any quantum state $\phi$ (or equivalently $\rho_\phi$), it is defined by

$$S(\phi) \equiv S(\rho_\phi) := -\mathrm{tr}\big(\rho_\phi \log \rho_\phi\big) \qquad [8]$$

Here we use log to denote $\log_2$ and define 0 log $0 \equiv 0$, as for the Shannon entropy. Let the density matrix $\rho_\phi$ have a spectral decomposition

$$\rho_\phi = \sum_{i=1}^{d} \lambda_i |\psi_i\rangle\langle\psi_i| \qquad [9]$$

Here $\{|\psi_i\rangle\}$ is the set of eigenvectors of $\rho_\phi$. They form an orthonormal basis of the Hilbert space $\mathcal{H}$. By the fact that $\rho_\phi$ is positive definite and has trace 1, the eigenvalues $\lambda_i$ of $\rho_\phi$ determine a probability distribution. When expressed in terms of the $\lambda_i$, the von Neumann entropy of $\rho$ reduces to the Shannon entropy corresponding to this probability distribution (henceforth, the subscript $\phi$ of $\rho_\phi$ will be omitted): $S(\rho) = H(\underline{\lambda})$, where $\underline{\lambda} = \{\lambda_1, \ldots, \lambda_d\}$.

The von Neumann entropy has properties analogous to $H(X_1, \ldots, X_n)$, in particular (Ohya and Petz 1993, Nielson and Chuang 2000)

1. $0 \leq S(\phi) \leq \log(\dim(\mathcal{H}))$;
2. $S(\phi)$ is concave in $\phi$; and
3. if $\phi$ is a state on $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ then $S(\phi) \leq S(\phi_1) + S(\phi_2)$ if $\phi_1$ and $\phi_2$ are the restrictions of $\phi$ to $\mathcal{H}_1 \otimes \boldsymbol{I}$ and $\boldsymbol{I} \otimes \mathcal{H}_2$ respectively.

A "quantum information source" in general is defined by a sequence of density matrices $\rho^{(n)}$ on Hilbert spaces $\mathcal{H}_n$ of increasing dimensions $N_n$ given by a decomposition

$$\rho^{(n)} = \sum_k p_k^{(n)} |\Psi_k^{(n)}\rangle\langle\Psi_k^{(n)}| \qquad [10]$$

where the states $|\Psi_k^{(n)}\rangle$ are interpreted as the signal states, and the numbers $p_k^{(n)} \geq 0$ with $\sum_k p_k^{(n)} = 1$, as their probabilities of occurrence. The vectors $|\Psi_k^{(n)}\rangle \in \mathcal{H}_n$ need not be mutually orthogonal.

### Compression–Decompression Scheme and Fidelity

To compress data from such a source one encodes each signal state $|\Psi_k^{(n)}\rangle$ by a state $\widetilde{\rho}_k^{(n)} \in \mathcal{B}(\widetilde{\mathcal{H}}_n)$ where $\dim \widetilde{\mathcal{H}}_n = d_c(n) < N_n$. Thus, a compression scheme is a map $\mathcal{C}^{(n)} : |\Psi_k^{(n)}\rangle\langle\Psi_k^{(n)}| \mapsto \widetilde{\rho}_k^{(n)} \in \mathcal{B}(\widetilde{\mathcal{H}}_n)$. The state $\widetilde{\rho}_k^{(n)}$ is referred to as the compressed state. A corresponding decompression scheme is a map $\mathcal{D}^{(n)}: \mathcal{B}(\widetilde{\mathcal{H}}_n) \mapsto \mathcal{B}(\mathcal{H}_n)$. Both $\mathcal{C}^{(n)}$ and $\mathcal{D}^{(n)}$ must be

completely positive maps. In particular, this implies that $\mathcal{D}^{(n)}$ must be of the form

$$\mathcal{D}^{(n)}(\rho) = \sum_i D_i \rho D_i^* \qquad [11]$$

for linear operators $D_i : \widetilde{\mathcal{H}}_n \mapsto \mathcal{H}_n$ such that $\sum_i D_i^* D_i = \boldsymbol{I}$ (see Nielson and Chuang 2000). Obviously, in order to achieve the maximum possible compression of Hilbert space dimensions per signal state, the goal must be to make the dimension $d_c(n)$ as small as possible, subject to the condition that the information carried in the signal states can be retrieved with high accuracy upon decompression.

The "rate of compression" is defined as

$$R_n := \frac{\log(\dim \widetilde{\mathcal{H}}_n)}{\log(\dim \mathcal{H}_n)} = \frac{\log d_c(n)}{\log N_n}$$

It is natural to consider the original Hilbert space $\mathcal{H}_n$ to be the $n$-qubit space. In this case $N_n = 2^n$ and hence $\log N_n = n$. As in the case of classical data compression, we are interested in finding the optimal limiting rate of data compression, which in this case is given by

$$R_\infty := \lim_{n \to \infty} \frac{\log d_c(n)}{n} \qquad [12]$$

Unlike classical signals, quantum signal states are not completely distinguishable. This is because they are, in general, not mutually orthogonal. As a result, perfectly reconstructing a quantum signal state from its compressed version is often an impossible task and therefore too stringent a requirement for the reliability of a compression–decompression scheme. Instead, a reasonable requirement is that a state can be reconstructed from the compressed version which is nearly indistinguishable from the original signal state. A measure of indistinguishability useful for this purpose is the average fidelity defined as follows:

$$F_n := \sum_k p_k^{(n)} \langle \Psi_k^{(n)} | \mathcal{D}^{(n)}(\widetilde{\rho}_k^{(n)}) | \Psi_k^{(n)} \rangle \qquad [13]$$

This fidelity satisfies $0 \le F_n \le 1$ and $F_n = 1$ if and only if $\mathcal{D}^{(n)}(\widetilde{\rho}_k^{(n)}) = |\Psi_k^{(n)}\rangle\langle\Psi_k^{(n)}|$ for all $k$. A compression–decompression scheme is said to be reliable if $F_n \to 1$ as $n \to \infty$.

The key idea behind data compression is the fact that some signal states have a higher probability of occurrence than others (these states playing a role analogous to the typical sequences of classical information theory). These signal states span a subspace of the original Hilbert space of the source and is referred to as the typical subspace.

### Schumacher's Theorem for Memoryless Quantum Sources

The notion of a typical subspace was first introduced in the context of quantum information theory by Schumacher (1995) in his seminal paper. He considered the simplest class of quantum information sources, namely quantum memoryless or i.i.d sources. For such a source the density matrix $\rho^{(n)}$, defined through [10], acts on a tensor product Hilbert space $\mathcal{H}_n = \mathcal{H}^{\otimes n}$ and is itself given by a tensor product

$$\rho^{(n)} = \pi^{\otimes n} \qquad [14]$$

Here $\mathcal{H}$ is a fixed Hilbert space (representing an elementary quantum subsystem) and $\pi$ is a density matrix acting on $\mathcal{H}$; for example, $\mathcal{H}$ can be a single qubit Hilbert space, in which case $\dim \mathcal{H} = 2, \mathcal{H}_n$ is the Hilbert space of $n$ qubits and $\pi$ is the density matrix of a single qubit. If the spectral decomposition of $\pi$ is given by

$$\pi = \sum_{i=1}^{\dim \mathcal{H}} q_i |\phi_i\rangle\langle\phi_i| \qquad [15]$$

then the eigenvalues and eigenvectors of $\rho^{(n)}$ are given by

$$\lambda_{\underline{k}}^{(n)} = q_{k_1} q_{k_2} \ldots q_{k_n} \qquad [16]$$

and

$$|\psi_{\underline{k}}^{(n)}\rangle = |\phi_{k_1}\rangle \otimes |\phi_{k_2}\rangle \otimes \cdots \otimes |\phi_{k_n}\rangle \qquad [17]$$

Thus, we can write the spectral decomposition of the density matrix $\rho^{(n)}$ of an i.i.d. source as

$$\rho^{(n)} = \sum_{\underline{k}} \lambda_{\underline{k}}^{(n)} |\psi_{\underline{k}}^{(n)}\rangle\langle\psi_{\underline{k}}^{(n)}| \qquad [18]$$

where the sum is over all possible sequences $\underline{k} = (k_1 \ldots k_n)$, with each $k_i$ taking $(\dim \mathcal{H})$ values. Hence, we see that the eigenvalues $\rho^{(n)}$ are labeled by a classical sequence of indices $\underline{k} = k_1 \ldots k_n$.

The von Neumann entropy of such a source is given by

$$S(\rho^{(n)}) \equiv S(\pi^{\otimes n}) = nS(\pi) = nH(X) \qquad [19]$$

where $X$ is the classical random variable with probability distribution $\{q_i\}$.

Let $T_\epsilon^{(n)}$ be the classical typical subset of indices $(k_1 \ldots k_n)$ for which

$$\left| -\frac{1}{n} \log(q_{k_1} \ldots q_{k_n}) - S(\pi) \right| \le \epsilon \qquad [20]$$

as in the theorem of typical sequences. Defining $\mathcal{T}_\epsilon^{(n)}$ as the space spanned by the eigenvectors $|\psi_{\underline{k}}^{(n)}\rangle$

with $\underline{k} \in T_\epsilon^{(n)}$ then immediately yields the quantum analog of the theorem of typical sequences – Theorem 4 given below. We refer to $\mathcal{T}_\epsilon^{(n)}$ as the typical subspace (or more precisely, the $\epsilon$–typical subspace).

**Theorem 4** (Typical subspace theorem). *Fix* $\epsilon > 0$. *Then for any* $\delta > 0 \, \exists \, n_0(\delta) > 0$ *such that* $\forall n \geq n_0(\delta)$ *and* $\rho^{(n)} = \pi^{\otimes n}$, *the following are true*:

(i) $\mathrm{Tr}(P_\epsilon^{(n)} \rho^{(n)}) > 1 - \delta$ *and*
(ii) $(1-\delta)2^{n(S(\pi)-\epsilon)} \leq \dim(\mathcal{T}_\epsilon^{(n)}) \leq 2^{n(S(\pi)+\epsilon)}$, *where* $P_\epsilon^{(n)}$ *is the orthogonal projection onto the subspace* $\mathcal{T}_\epsilon^{(n)}$.

Note that $\mathrm{tr}(P_\epsilon^{(n)} \rho^{(n)})$ gives the probability of the typical subspace. As $\mathrm{tr}(P_\epsilon^{(n)} \rho^{(n)})$ approaches unity for $n$ sufficiently large, $\mathcal{T}_\epsilon^{(n)}$ carries almost all the weight of $\rho^{(n)}$. Let $\mathcal{T}_\epsilon^{(n)\perp}$ denote the orthocomplement of the typical subspace, that is, for any pair of vectors $|\psi\rangle \in \mathcal{T}_\epsilon^{(n)}$ and $|\phi\rangle \in \mathcal{T}_\epsilon^{(n)\perp}$, $\langle\phi|\psi\rangle = 0$. It follows from the above theorem that the probability of a signal state belonging to $\mathcal{T}_\epsilon^{(n)\perp}$ can be made arbitrarily small for $n$ sufficiently large.

Let $P_\epsilon^{(n)}$ denote the orthogonal projection onto the typical subspace $\mathcal{T}_\epsilon^{(n)}$. The encoding (compression) of the signal states $|\Psi_k^{(n)}\rangle$ of [10], is done in the following manner. $\mathcal{C}^{(n)} : |\Psi_k^{(n)}\rangle \langle\Psi_k^{(n)}| \mapsto \tilde{\rho}_k^{(n)}$, where

$$\tilde{\rho}_k^{(n)} := \alpha_k^2 |\tilde{\Psi}_k^{(n)}\rangle\langle\tilde{\Psi}_k^{(n)}| + \beta_k^2 |\Phi_0\rangle\langle\Phi_0| \qquad [21]$$

Here

$$|\tilde{\Psi}_k^{(n)}\rangle := \frac{P_\epsilon^{(n)}|\Psi_k^{(n)}\rangle}{\|P_\epsilon^{(n)}|\Psi_k^{(n)}\rangle\|}$$
$$\alpha_k := \|P_\epsilon^{(n)}|\Psi_k^{(n)}\rangle\|, \quad \beta_k = \|(I - P_\epsilon^{(n)})|\Psi_k^{(n)}\rangle\| \qquad [22]$$

and $|\Phi_0\rangle$ is any fixed state in $\mathcal{T}_\epsilon^{(n)}$.

Obviously $\tilde{\rho}_k^{(n)} \in \mathcal{B}(\mathcal{T}_\epsilon^{(n)})$, and hence the typical subspace $\mathcal{T}_\epsilon^{(n)}$ plays the role of the compressed space. The decompression $\mathcal{D}^{(n)}(\tilde{\rho}_k^{(n)})$ is defined as the extension of $\tilde{\rho}_k^{(n)}$ on $\mathcal{T}_\epsilon^{(n)}$ to $\mathcal{H}_n$:

$$\mathcal{D}^{(n)}\left(\tilde{\rho}_{\underline{k}}^{(n)}\right) = \tilde{\rho}_{\underline{k}}^{(n)} \oplus 0$$

The fidelity of this compression–decompression scheme satisfies

$$\begin{aligned} F_n &= \sum_k p_k^{(n)} \langle\Psi_k^{(n)}|\tilde{\rho}_k^{(n)}|\Psi_k^{(n)}\rangle \\ &= \sum_k p_k^{(n)} \left[\alpha_k^2 |\langle\Psi_k^{(n)}|\tilde{\Psi}_k^{(n)}\rangle|^2 + \beta_k^2 |\langle\Psi_k^{(n)}|\Phi_0\rangle|^2\right] \\ &\geq \sum_k p_k^{(n)} \alpha_k^2 |\langle\Psi_k^{(n)}|\tilde{\Psi}_k^{(n)}\rangle|^2 = \sum_k p_k^{(n)} \alpha_k^4 \\ &\geq \sum_k p_k^{(n)} (2\alpha_k^2 - 1) = 2A_n - 1 \end{aligned} \qquad [23]$$

where $A_n = \mathrm{tr}(P_\epsilon^{(n)} \rho_n)$.

Using the typical subspace theorem, Schumacher (1995) proved the following analog of Shannon's noiseless channel coding theorem for memoryless quantum information sources:

**Theorem 5** (Schumacher's quantum coding theorem). *Let* $\{\rho_n, \mathcal{H}_n\}$ *be an i.i.d. quantum source*: $\rho_n = \pi^{\otimes n}$ *and* $\mathcal{H}_n = \mathcal{H}^{\otimes n}$. *If* $R > S(\pi)$, *then there exists a reliable compression scheme of rate R. If* $R < S(\pi)$, *then any compression scheme of rate R is not reliable.*

**Proof**

(i) $R > S(\pi)$. Choose $\epsilon > 0$ such that $R > S(\pi) + \epsilon$. For a given $\delta > 0$, choose the typical subspace as above and choose $n$ large enough so that (i) and (ii) in the typical subspace theorem hold. In particular, $A_n = \mathrm{tr}(P_\epsilon^{(n)}\rho_n) > 1 - \delta$. Thus, the fidelity tends to 1 as $n \to \infty$.

(ii) *Suppose* $R < S(\pi)$. Let the compression map be $\mathcal{C}^{(n)}$. We may assume that $\widetilde{\mathcal{H}}_n$ is a subspace of $\mathcal{H}_n$ with $\dim \widetilde{\mathcal{H}}_n = 2^{nR}$. We denote the projection onto $\widetilde{\mathcal{H}}_n$ as $\widetilde{P}_n$ and let $\tilde{\rho}_k^{(n)} = \mathcal{C}^{(n)}(|\Psi_k^{(n)}\rangle\langle\Psi_k^{(n)}|)$. Since $\tilde{\rho}_k^{(n)}$ is concentrated on $\widetilde{\mathcal{H}}_n$, we have $\tilde{\rho}_k^{(n)} \leq \widetilde{P}_n$ and hence $\mathcal{D}^{(n)}(\tilde{\rho}_k^{(n)}) \leq \mathcal{D}^{(n)}(\widetilde{P}_n)$, for any decompression map $\mathcal{D}^{(n)}$. Inserting into the definition of the fidelity, we then have

$$\begin{aligned} F &\leq \sum_k p_k^{(n)} \langle\Psi_k^{(n)}|\mathcal{D}^{(n)}(\widetilde{P}_n)|\Psi_k^{(n)}\rangle = \mathrm{tr}\left(\rho^{(n)}\mathcal{D}^{(n)}(\widetilde{P}_n)\right) \\ &\leq \sum_{\underline{k}\in T_\epsilon^{(n)}} \lambda_{\underline{k}}^{(n)} \langle\psi_{\underline{k}}^{(n)}|\mathcal{D}^{(n)}(\widetilde{P}_n)|\psi_{\underline{k}}^{(n)}\rangle + \sum_{\underline{k}\notin T_\epsilon^{(n)}} \lambda_{\underline{k}}^{(n)} \end{aligned} \qquad [24]$$

By the typical subspace theorem, the latter sum tends to 0 as $n \to \infty$, and in the sum over $\underline{k} \in T_\epsilon^{(n)}$ we have $\lambda_{\underline{k}}^{(n)} \leq 2^{-n(S(\pi)-\epsilon)}$. The first sum can therefore be bounded as follows:

$$\begin{aligned} &\sum_{\underline{k}\in T_\epsilon^{(n)}} \lambda_{\underline{k}}^{(n)} \langle\psi_{\underline{k}}^{(n)}|\mathcal{D}^{(n)}(\widetilde{P}_n)|\psi_{\underline{k}}^{(n)}\rangle \\ &\quad \leq 2^{-n(S(\pi)-\epsilon)} \sum_{\underline{k}} \langle\psi_{\underline{k}}^{(n)}|\mathcal{D}^{(n)}(\widetilde{P}_n)|\psi_{\underline{k}}^{(n)}\rangle \\ &\quad = 2^{-n(S(\pi)-\epsilon)} \mathrm{tr}\left(\mathcal{D}^{(n)}(\widetilde{P}_n)\right) \\ &\quad = 2^{-n(S(\pi)-\epsilon)} \mathrm{tr}\left(\sum_i D_i \widetilde{P}_n D_i^*\right) \\ &\quad = 2^{-n(S(\pi)-\epsilon)} 2^{nR} \end{aligned} \qquad [25]$$

by the cyclic property of the trace and the fact that $\sum_i D_i^* D_i = I$ and $\dim \widetilde{\mathcal{H}}_n = 2^{nR}$. □

Even for a quantum source with memory, reliable data compression is achieved by looking for a typical subspace $\mathcal{T}_\epsilon^{(n)}$ of the Hilbert space $\mathcal{H}_n$ for a given $\epsilon > 0$. In the following subsections, we discuss two different classes of such sources for which one

can find typical subspaces $\mathcal{T}_\epsilon^{(n)}$ such that the fidelity $F_n$ tends to 1 as $n \to \infty$.

## Ergodic Quantum Sources

A quantum generalization of classical ergodic sources is defined as follows. First consider the analog of an infinite sequence of random variables which is a state on the infinite tensor product of a finite-dimensional $*$-algebra $\mathcal{M}$. The latter is given by the norm closure of the increasing sequence of finite tensor products

$$\mathcal{M}_\infty = \overline{\bigcup_n \otimes_{k=-n}^n \mathcal{M}} \qquad [26]$$

A translation-invariant state $\phi_\infty$ on $\mathcal{M}_\infty$ is said to be ergodic if it cannot be decomposed as a (nontrivial) convex combination of other translation-invariant states. The analog of the Kolmogorov–Sinai entropy [5] for an ergodic state $\phi_\infty$ is called the mean entropy and is given by

$$S_M(\phi_\infty) = \lim_{n\to\infty} \frac{1}{n} S(\phi_n) = \inf_{n\in\mathbb{N}} \frac{1}{n} S(\phi_n) \qquad [27]$$

where $\phi_n$ is the restriction of $\phi_\infty$ to $\mathcal{M}_n := \mathcal{M}^{\otimes n}$. Following Hiai and Petz (1991), we define the following quantity for any state $\phi$ on an arbitrary finite-dimensional $*$-algebra $\mathcal{M}$ and a given $\delta > 0$:

$$\beta_\delta(\phi) = \inf\{\log \mathrm{tr}(q) : q \in \mathcal{M}, q^* = q, \\ q^2 = q, \phi(q) \geq 1 - \delta\} \qquad [28]$$

We also define a state $\phi_\infty$ on $\mathcal{M}_\infty$ to be completely ergodic if it is ergodic under transformations on $\mathcal{M}_\infty$, induced by $l$-fold shifts on $\mathbb{Z}$, for arbitrary $l \in \mathbb{N}$. The following theorem is due to Hiai and Petz (1991), who proved it in a slightly more general setting:

**Theorem 6** (Hiai and Petz). *Suppose that $\phi_\infty$ is a completely ergodic state on $\mathcal{M}_\infty$ and $d := \dim \mathcal{M} < \infty$, and set $\phi_n = \phi_\infty \restriction_{\mathcal{M}_n}$. Then, for any $\delta > 0$, the following hold*:

(i) $$\limsup_{n\to\infty} \frac{1}{n} \beta_\delta(\phi_n) \leq S_M(\phi_\infty) \qquad [29]$$

(ii) $$\liminf_{n\to\infty} \frac{1}{n} \beta_\delta(\phi_n) \geq S_M(\phi_\infty) - \delta \log d \qquad [30]$$

**Proof of (i)** Choose $r > S_M(\phi_\infty)$ and let $\epsilon < r - S_M(\phi_\infty)$ and $h = r - \epsilon$. By the definition of $S_M(\phi_\infty)$, there exists $l \in \mathbb{N}$ such that $S(\phi_l) < lh$. Let $\{|e_i\rangle\}_{i=1}^{ld}$ be an orthonormal set of eigenvectors of $\rho_{\phi_l}$, with corresponding eigenvalues $\lambda_i$, that is, let

$$\rho_{\phi_l} = \sum_{i=1}^{ld} \lambda_i p_i \qquad [31]$$

where $p_i = |e_i\rangle\langle e_i|$ is the projection onto $|e_i\rangle$, be the spectral decomposition for $\rho_{\phi_l}$. Denote the spectrum $\mathcal{X} = \{\lambda_i\}_{i=1}^{ld}$. For $n \in \mathbb{N}$, introduce the probability measures $\nu_n$ on $\mathcal{X}^n$ by

$$\nu_n(A) = \phi_{nl}(q_A) \qquad [32]$$

where, for any $A \subset \mathcal{X}^n$, the projection $q_A$ is defined by

$$q_A = \sum_{(\lambda_{i_1},\ldots,\lambda_{i_n})\in A} p_{i_1} \otimes \ldots \otimes p_{i_n} \qquad [33]$$

Similarly, we define $\nu_\infty$ on $\mathcal{X}^{\mathbb{Z}}$. The sequence of random variables $(X_n)_{n\in\mathbb{Z}}$ with distribution $\nu_\infty$ is then ergodic since $\phi_\infty$ is completely ergodic (and hence $l$-ergodic).

By the Shannon–McMillan–Breiman theorem (Theorem 2),

$$-\frac{1}{n} \log \nu_n(\{(x_1, \ldots, x_n)\}) \to h_{\mathrm{KS}} \qquad [34]$$

almost surely w.r.t. $\nu_\infty$, where $h_{\mathrm{KS}}$ is the Kolmogorov–Sinai entropy. The latter is given by $h_{\mathrm{KS}} = \lim_{n\to\infty} (1/n)H_n = \inf_{n\in\mathbb{N}} (1/n)H_n$, where

$$H_n = -\sum_{(x_1,\ldots,x_n)\in\mathcal{X}^n} \nu_n(\{(x_1, \ldots, x_n)\}) \\ \times \log \nu_n(\{(x_1, \ldots, x_n)\}) \qquad [35]$$

Notice in particular that

$$h_{\mathrm{KS}} \leq H_1 = S(\phi_l) < lh \qquad [36]$$

If let $T_\epsilon^{(n)}$ be the (typical) subset of $\mathcal{X}^n$ such that

$$-\frac{1}{n} \log \nu_n(\{(x_1, \ldots, x_n)\}) \in (h_{\mathrm{KS}} - \epsilon, h_{\mathrm{KS}} + \epsilon) \qquad [37]$$

for $(x_1, \ldots, x_n) \in T_\epsilon^{(n)}$ then we have $\nu_\infty(T_\epsilon^{(n)}) \geq 1 - \delta$ for $n$ large enough. Moreover, since $\nu_n(\{(x_1, \ldots, x_n)\}) \geq \mathrm{e}^{-n(h_{\mathrm{KS}}+\epsilon)}$ for all $(x_1, \ldots, x_n) \in T_\epsilon^{(n)}$, and the total measure is 1,

$$|T_\epsilon^{(n)}| \leq \mathrm{e}^{n(h_{\mathrm{KS}}+\epsilon)} \leq \mathrm{e}^{n(lh+\epsilon)} \qquad [38]$$

It follows that $\mathrm{tr}(q_{T_\epsilon^{(n)}}) \leq \mathrm{e}^{n(lh+\epsilon)}$ whereas $\phi_{nl}(q_{T_\epsilon^{(n)}}) = \nu_n(T_\epsilon^{(n)}) \geq 1 - \delta$ and we conclude that

$$\frac{1}{nl} \beta_\delta(\phi_{nl}) \leq \frac{n(lh + \epsilon)}{nl} < r \qquad [39]$$

from which [29] follows upon taking $n \to \infty$, since $r > S_M(\phi_\infty)$ was arbitrary. (Notice that $\beta_\delta(\phi_n)$ is decreasing in $n$ since $\mathcal{M}_n \subset \mathcal{M}_{n+1}$.) □

**Proof of (ii)** Given $\epsilon, \delta > 0$ and $n \in \mathbb{N}$, choose a projection $q_n$ with $\phi_n(q_n) \geq 1 - \delta$ and $\log \mathrm{tr}(q_n) < \beta_\delta(\phi_n) + \epsilon$. Since $S_M(\phi_\infty) = \inf (1/n) S(\phi_n)$ we have

$S_M(\phi_\infty) \leq (1/n)S(\phi_n)$. We now use the following lemma:

**Lemma 7** *If $\phi$ is a state on a finite-dimensional $*$-algebra $\mathcal{M}$, and $q \in \mathcal{M}$ is a projection, then*

$$S(\phi) \leq H(p) + \phi(q)\log \mathrm{tr}(q) + (1-\phi(q))\log \mathrm{tr}(1-q) \qquad [40]$$

*where $H(p) = -p\log p - (1-p)\log(1-p)$ (the binary entropy) with $p = \phi(q)$.*

**Proof** First notice that if $[\rho_\phi, q] = 0$ then the result [40] follows from the simple inequality:

$$-\sum_{i=1}^{m} \tilde{\lambda}_i \log \tilde{\lambda}_i \leq \log m \quad \text{if} \quad \sum_{i=1}^{m} \tilde{\lambda}_i = 1 \qquad [41]$$

Indeed, diagonalizing $\rho_\phi$, the eigenvalues $\lambda_i$ divide into two subsets with corresponding eigenvectors belonging to the range of $q$, respectively, its complement. Considering the first set, we have, if $m = \dim(\mathrm{Ran}(q))$, and taking $\tilde{\lambda}_i = \lambda_i/(\sum_{i=1}^m \lambda_i)$ in [41],

$$-\sum_{i=1}^{m} \lambda_i \log \lambda_i \leq -\left(\sum_{i=1}^{m} \lambda_i\right) \log\left(\frac{1}{m}\sum_{i=1}^{m} \lambda_i\right) = -\mathrm{tr}(q\rho_\phi)\left[\log \mathrm{tr}(q\rho_\phi) - \log \mathrm{tr}(q)\right]$$

Adding the analogous inequality for the part of the spectrum corresponding to $1-q$, we obtain [40].

In the general case, that is, if $[\rho_\phi, q] \neq 0$, define the unitary $u = 2q - 1$ and the state

$$\phi'(x) = \tfrac{1}{2}[\phi(x) + \phi(uxu)] \qquad [42]$$

Then $[\rho_{\phi'}, q] = 0$ and by concavity of $S(\phi)$ and the result for the previous case

$$H(X) + \phi(q) \log \mathrm{tr}(q) + (1-\phi(q)) \log \mathrm{tr}(1-q) \geq S(\phi') \geq S(\phi) \qquad [43]$$

since $\phi'(q) = \phi(q)$. □

Continuing with the proof of (ii), we conclude that

$$S(\phi_n) \leq H(p) + \phi_n(q_n)\log \mathrm{tr}(q_n) + (1-\phi(q_n))\log \mathrm{tr}(1-q_n) \leq 1 + \beta_\delta(\phi_n) + \epsilon + \delta n \log d$$

Dividing by $n$ and taking the limit we obtain (30). □

It follows from this theorem that we can define a typical subspace in the same way as in Schumacher's theorem. Indeed, given $\delta > 0$ and $\epsilon > 0$, we have that for $n$ large enough, there exists a subspace $\mathcal{T}_\epsilon^{(n)}$ equal to the range of a projection $q_n$ such that $\phi_n(q_n) > 1 - \delta$ and $\mathrm{e}^{n(S_M(\phi_\infty) - \delta\log d - \epsilon)} < \dim(\mathcal{T}_\epsilon^{(n)}) = \mathrm{tr}(q_n) < \mathrm{e}^{n(S_M(\phi_\infty)+\epsilon)}$. The proof of the quantum analog of the Shannon–McMillan theorem is then similar to that of Schumacher's theorem (Petz and Mosonyi 2001, Bjelaković *et al.* 2004):

**Theorem 8** *Let $\phi_\infty$ be a completely ergodic stationary state on the infinite tensor product algebra $\mathcal{M}_\infty$. If $R > S_M(\phi_\infty)$, then for any decomposition of the form*

$$\rho^{(n)} = \sum p_k^{(n)} |\Psi_k^{(n)}\rangle\langle\Psi_k^{(n)}| \qquad [44]$$

*there exists a reliable quantum code of rate $R$. Conversely, if $R < S_M(\phi_\infty)$ then any quantum compression–decompression scheme of rate $R$ is not reliable.*

**Remarks** Theorem 6 also holds for higher-dimensional information streams, with essentially the same proof. (The existence of the mean entropy is more complicated in that case.) The condition of complete ergodicity in this theorem is unnecessary. Indeed, Bjelaković *et al.* (2004) showed that the result remains valid (also in more than one dimensions) if the state $\phi_\infty$ of the source is simply ergodic. They achieved this by decomposing a general ergodic state into a finite number of $l$-ergodic states, and then applying the above strategy to each. It should also be mentioned that a weaker version of Theorem 6 was proved by King and Lesniewski (1998). They considered the entropy of an associated classical source, but did not show that this classical entropy can be optimized to approximate the von Neumann entropy. This had in fact already been proved by Hiai and Petz (1991). The relevance of the latter work for quantum information theory was finally pointed out by Mosonyi and Petz (2001).

## Source Coding for Quantum Spin Systems

In this section we consider a class of quantum sources modeled by Gibbs states of a finite strongly interacting quantum spin system in $\Lambda \subset \mathbb{Z}^d$ with $d \geq 2$. Due to the interaction between spins, the density matrix of the source is not given by a tensor product of the density matrices of the individual spins and hence the quantum information source is non-i.i.d. We consider the density matrix to be written in the standard Gibbsian form:

$$\rho^{\omega,\Lambda} = \frac{\mathrm{e}^{-\beta H_\Lambda^\omega}}{\Xi^{\omega,\Lambda}} \qquad [45]$$

where $\beta > 0$ is the inverse temperature. Here $\omega$ denotes the boundary condition, that is, the configuration of the spins in $\Lambda^c = \mathbb{Z}^d \setminus \Lambda$, and $H_\Lambda^\omega$ is the Hamiltonian acting on the spin system in $\Lambda$ under this boundary condition. (see Datta and Suhov (2002)

for precise definitions of these quantities). The denominator on the right-hand side of [45] is the partition function.

Note that any faithful density matrix can be written in the form [45] for some self-adjoint operator $H_\Lambda^\omega$ with discrete spectrum, such that $e^{-\beta H_\Lambda^\omega}$ is trace class. However, we consider $H_\Lambda^\omega$ to be a small quantum perturbation of a classical Hamiltonian and require it to satisfy certain hypotheses (see Datta and Suhov (2002)). In particular, we assume that $H_\Lambda = H_{0\Lambda} + \lambda V_\Lambda$, where (1) $H_{0\Lambda}$ is a classical, finite-range, translation-invariant Hamiltonian with a finite number of periodic ground states, and the excitations of these ground states have an energy proportional to the size of their boundaries (Peierls condition); (2) $\lambda V_\Lambda$ is a translation-invariant, exponentially decaying, quantum perturbation, $\lambda$ being the perturbation parameter. These hypotheses ensure that the quantum Pirogov–Sinai theory of phase transitions in lattice systems (see, e.g., Datta *et al.* (1996)) applies.

The power of quantum Pirogov–Sinai theory is such that, in proving reliable data compression for such sources, we do not need to invoke the concept of ergodicity.

Using the concavity of the von Neumann entropy $S(\rho^{\omega,\Lambda})$, one can prove that the von Neumann entropy rate (or mean entropy) of the source

$$h := \lim_{\Lambda \nearrow \mathbb{Z}^d} \frac{S(\rho^{\omega,\Lambda})}{|\Lambda|}$$

exists. For a general van Hove sequence, this follows from the strong subadditivity of the von Neumann entropy (see, e.g., Ohya and Petz (1993)).

Let $\rho^{\omega,\Lambda}$ have a spectral decomposition

$$\rho^{\omega,\Lambda} = \sum_j \lambda_j |\psi_j\rangle\langle\psi_j|$$

where the eigenvalues $\lambda_j, 1 \le j \le 2^{|\Lambda|}$, and the corresponding eigenstates $|\psi_j\rangle$, depend on $\omega$ and $\Lambda$. Let $\mathcal{P}^{\omega,\Lambda}$ denote the probability distribution $\{\lambda_j\}$ and consider a random variable $K^{\omega,\Lambda}$ which takes a value $\lambda_j$ with probability $\lambda_j$:

$$K^{\omega,\Lambda}(\psi_j) = \lambda_j; \quad \mathcal{P}^{\omega,\Lambda}(K^{\omega,\Lambda} = \lambda_j) = \lambda_j$$

The data compression limit is related to asymptotical properties of the random variables $K^{\omega,\Lambda}$ as $\Lambda \nearrow \mathbb{Z}^d$. As in the case of i.i.d. sources, we prove the reliability of data compression by first proving the existence of a typical subspace. The latter follows from Theorem 9 below. The proof of this crucial theorem relies on results of quantum Pirogov–Sinai theory (Datta *et al.* 1996).

**Theorem 9** *Under the above assumptions, for $\beta$ large and $\lambda$ small enough, for all $\epsilon > 0$*

$$\lim_{\Lambda \nearrow \mathbb{Z}^d} \mathcal{P}^{\omega,\Lambda}\left(\left|\frac{-1}{|\Lambda|}\log K^{\omega,\Lambda} - h\right| \le \epsilon\right) = \lim_{\Lambda \nearrow \mathbb{Z}^d} \sum_j \lambda_j \chi_{\{|-|\Lambda|^{-1}\log\lambda_j - h| \le \epsilon\}} = 1 \qquad [46]$$

*where $\chi_{\{\ldots\}}$ denotes an indicator function.*

Theorem 9 is essentially a law of large numbers for random variables ($-\log K^{\omega,\Lambda}$). The statement of the theorem can be alternatively expressed as follows. For any $\epsilon > 0$,

$$\lim_{\Lambda \nearrow \mathbb{Z}^d} \mathcal{P}^{\omega,\Lambda}\left(2^{-|\Lambda|(h+\epsilon)} \le K^{\omega,\Lambda} \le 2^{-|\Lambda|(h-\epsilon)}\right) = 1 \qquad [47]$$

Thus, we can define a typical subspace $\mathcal{T}_\epsilon^{\omega,\Lambda}$ by

$$\mathcal{T}_\epsilon^{\omega,\Lambda} := \text{span}\,\{|\psi_j\rangle : 2^{-|\Lambda|(h+\epsilon)} \le \lambda_j \le 2^{-|\Lambda|(h-\epsilon)}\} \qquad [48]$$

It clearly satisfies the analogs of (i) and (ii) of the typical subspace theorem, which implies as before that a compression scheme of rate $R$ is reliable if and only if $R > h$.

## Universal and Variable Length Data Compression

Thus far we discussed source-dependent data compression for various classes of quantum sources. In each case data compression relied on the identification of the typical subspace of the source, which in turn required a knowledge of its density matrix. In classical information theory, there exists a generalization of the theorem of typical sequences due to Csiszár and Körner (1981) where the typical set is universal, in that it is typical for every possible probability distribution with a given entropy. This result was used by Jozsa *et al.* (1998) to construct a universal compression scheme for quantum i.i.d sources with a given von Neumann entropy $S$ using a counting argument for symmetric subspaces. This was generalized to ergodic sources by Kaltchenko and Yang (2003) along the lines of Theorem 6. Hayashi and Matsumoto (2002) supplemented the work of Jozsa *et al.* (1998) with an estimation of the eigenvalues of the source (using the measurement smearing technique) to show that a reliable compression scheme exists for any quantum i.i.d source, independent of the value of its von Neumann entropy $S$, the limiting rate of compression being given by $S$. If one admits variable length coding, the Lempel–Ziv algorithm gives a completely universal compression scheme, independent of the value of the entropy, in the classical case (Cover and Thomas 1991). This algorithm was generalized to the quantum case for i.i.d sources by Jozsa and Presnell (2003), and to

sources modeled by Gibbs states of free bosons or fermions on a lattice by Johnson and Suhov (2002).

Another important question is the efficiency of the various coding schemes. The above-mentioned schemes for quantum i.i.d. sources are not efficient, in the sense that they have no polynomial time implementation. Recently, it was shown by Bennett *et al.* (2004) that an efficient, universal compression scheme for i.i.d sources can be constructed by employing quantum state tomography.

## Acknowledgment

The authors would like to thank Y M Suhov for helpful discussions.

*See also:* Capacity for Quantum Information; Channels in Quantum Information Theory; Positive Maps on $C^*$-Algebras.

## Further Reading


Bennett CH, Harrow AW, and Lloyd S (2004) Universal quantum data compression via gentle tomography, quant-ph/0403078.

Bjelaković I, Krüger T, Siegmund-Schultze R, and Szkoła A (2004) The Shannon–McMillan theorem for ergodic quantum lattice systems. *Inventiones Mathematicae* 155: 203–222.

Breiman L (1957) The individual ergodic theorem of information theory. *Annals of Mathematical Statistics* 28: 809–811.

Breiman L (1960) The individual ergodic theorem of information theory – correction note. *Annals of Mathematical Statistics* 31: 809–810.

Cover TM and Thomas JA (1991) *Elements of Information Theory*. New York: Wiley.

Csiszár I and Körner J (1981) *Information Theory. Coding Theorems for Discrete Memoryless Systems*. Budapest: Akadémiai Kiadó.

Datta N, Fernández R, and Fröhlich J (1996) Low-temperature phase diagrams of quantum lattice systems. I. Stability for quantum perturbations of classical systems with finitely-many ground states. *Journal of Statistical Physics* 84: 455–534.

Datta N and Suhov Y (2002) Data compression limit for an information source of interacting qubits. *Quantum Information Processing* 1(4): 257–281.

Hayashi M and Matsumoto K (2002) Quantum universal variable-length source coding. *Physical Review A* 66: 022311.

Hiai F and Petz D (1991) The proper formula for the relative entropy and its asymptotics in quantum probability. *Communications in Mathematical Physics* 143: 257–281.

Johnson O and Suhov YM (2002) The von Neumann entropy and information rate for integrable quantum Gibbs ensembles. *Quantum Computers and Computing* 3/1: 3–24.

Jozsa R, Horodecki M, Horodecki P, and Horodecki R (1998) Universal quantum information compression. *Physical Review Letters* 81: 1714–1717.

Jozsa R and Presnell S (2003) Universal quantum information compression and degrees of prior knowledge. *Proceedings of Royal Society of London Series A* 459: 3061–3077.

Kaltchenko A and Yang E-H (2003) Universal compression of ergodic quantum sources. *Quantum Information & Computation* 3: 359–375.

King C and Lesniewski A (1998) Quantum sources and a quantum coding theorem. *Journal of Mathematical Physics* 39: 88–101.

McMillan B (1953) The basic theorems of information theory. *Annals of Mathematical Statistics* 24: 196–219.

Nielson MA and Chuang IL (2000) *Quantum Computation and Quantum Information*. Cambridge: Cambridge University Press.

Ohya M and Petz D (1993) *Quantum Entropy and Its Use*. Heidelberg: Springer.

Petz D and Mosonyi M (2001) Stationary quantum source coding. *Journal of Mathematical Physics* 42: 4857–4864.

Schumacher B (1995) Quantum coding. *Physical Review A* 51: 2738–2747.

Shannon CE (1918) A mathematical theory of communication. *Bell System Technical Journal* 27: 379–423, 623–656.


# Spacetime Topology, Causal Structure and Singularities

**R Penrose**, University of Oxford, Oxford, UK



## The Value of Topological Reasoning in General Relativity

Solving the equations of Einstein's general relativity (*see* General Relativity: Overview) can be an exceedingly complicated business; it is commonly found necessary to resort to numerical solutions involving very complex computer codes (*see* Computational Methods in General Relativity: The Theory). The essential content of the basic equations of the theory itself is, however, something that can be phrased in simple geometrical terms, using only basic concepts of differential geometry (*see* General Relativity: Overview). By virtue of this, it is sometimes the case, in general relativity, that geometrical arguments of various kinds – including purely topological ones (i.e., arguments depending only upon the properties of continuity or smoothness) – can be used to great effect to obtain results that are not readily accessible by standard procedures of differential equation theory or by direct numerical calculation.

One particularly significant family of situations where this kind of argument has a key role to play is in the important issue of the singularities that arise in many solutions of the Einstein equations, in which spacetime curvatures may be expected to diverge to infinity. These are exemplified, particularly, by two important classes of solutions of the

Einstein field equations in which singularities arise. In the first instance, we have cosmological models, which tend to exhibit the presence of an initial singularity referred to as the "Big Bang," as was first noted in the standard Friedmann models (which are solutions of the Einstein equations with simple matter sources; *see* Cosmology: Mathematical Aspects). Secondly, we find a final singularity (for local observers) at the endpoint of gravitational collapse to a black hole (where in the relevant region, outside the collapsing matter, Einstein's vacuum equations are normally taken to hold). In either case, there are canonical exact models, in which considerable symmetry is assumed, and where the models indeed become singular at places where the spacetime curvature diverges to infinity. For many years (prior to 1965), there had been much debate as to whether these singularities were an inevitable feature of the general physical situation under consideration, or whether the presence of singularities might be an artifact of the assumed high symmetry. The use of topological-type arguments has established that, in general terms, the occurrence of a singularity is not merely an artifact of symmetry, and cannot generally be removed by the introduction of small (finite) perturbations.

Let us first consider the standard picture, put forward in 1939 by Oppenheimer and Snyder (OS), of the gravitational collapse of an over-massive star to a black hole; see Figure 1 (and *see* Stationary Black Holes). This assumes exact spherical symmetry. The region external to the matter is described by the well-known Schwarzschild solution of the Einstein vacuum equations, appropriately extended to inside the "Schwarzschild radius" $r = 2mG/c^2$ ($G$ being Newton's gravitational constant and $c$, the speed of light, and where $m$ is the total mass of the collapsing material; from now, for convenience, we choose units so that $G = c = 1$). In Figure 1, this internal extension is conveniently expressed using Eddington–Finkelstein coordinates $(r, v, \theta, \phi)$ (see Eddington (1924) and Finkelstein (1958)), where $v = t + r + 2m \log(r - 2m)$, the metric form being

$$ds^2 = (1 - 2m/r)dv^2 - 2dvdr - r^2(d\theta^2 + \sin^2\theta d\phi^2)$$

(The signature convention $+---$ is being adopted here; *see* General Relativity: Overview.) We find that, in this model, there is a singularity (at $r = 0$) at the future endpoint of each world line of collapsing matter. Moreover, no future-timelike line starting inside the horizon can avoid reaching the singularity when we try to extend it, as a timelike curve,

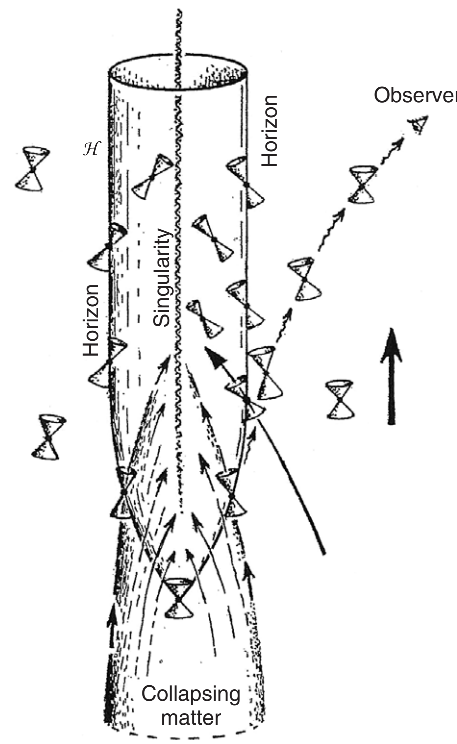



**Figure 1** Spacetime diagram of collapse to a black hole. (One spatial dimension is suppressed.) Matter collapses inwards, through the 3-surface that becomes the (absolute) event horizon. No matter or information can escape the hole once it has been formed. The null cones are tangent to the horizon and allow matter or signals to pass inwards but not outwards. An external observer cannot see inside the hole, but only the matter – vastly dimmed and redshifted – just before it enters the hole. (Reproduced with permission from Penrose R. (2004) *The Road to Reality: a Complete Guide to the Laws of the Universe*. London: Jonathan Cape.)

indefinitely into the future, where the "horizon" is the three-dimensional region obtained by rotating, over the $(\theta, \phi)$ 2-sphere, the null (lightlike) line which is $r = 2m$ outside the matter region and which is the extension of this line, as a null line, into the past until it meets the axis. It is easy to see that any observer's world line within this horizon is indeed trapped in this sense.

The question naturally arises: how representative is this model? Here, the singularity occurs at the center ($r = 0$), the place where all the matter is directed, and where it all reaches without rebounding. So it may be regarded as unsurprising that the density becomes infinite there. Now, let us suppose that the collapsing material is not exactly spherically symmetrical. Even if it is only slightly (though finitely) perturbed away from this symmetrical situation, having slight (but finite) transverse motions, the collapsing matter is now not all directed exactly towards the center, as it is in the OS model. One might imagine that the singularity

could now be avoided, the different portions of matter just “missing” each other and then being finally flung out again, after some complicated motions, where the density and spacetime curvatures might well become large but presumably still finite. To follow such an irregular collapse in full detail would present a very difficult task, and one would have to carry it out by numerical means. As yet, despite enormous advances in computational technique, a fully effective simulation of such a “generic” collapse is still not in hand. In any case, it is hard to make a convincing case as to whether or not a singularity arises, because as soon as metric or curvature quantities begin to diverge, the computation becomes fundamentally unreliable and simply “gives up.” So we cannot really tell whether the failure is due to some genuine divergence or whether it is an artifact. It is thus fortunate that other mathematical techniques are available. Indeed, by use of a differential–topological–causal argument, we find that such perturbations do not help, at least so long as they are small enough not to alter the general character of the collapse, which we find has an “unstoppable” character, so long as a certain criterion is satisfied its early stages.

## Trapped Surfaces

But how are we to characterize the collapse as “unstoppable,” where no symmetries are to be assumed, and the simple picture illustrated in Figure 1 cannot be appealed to? A convenient characterization is the presence of what is called a “trapped surface.” This notion generalizes a key feature of the $0 < r < 2m$ region inside the horizon of the vacuum (Eddington–Finkelstein) picture of Figure 1. To understand what this feature is, consider fixing a point $s$ in the vacuum region of the $(v, r)$-plane of Figure 1. We must, of course, bear in mind that, because this plane is to be “rotated” about the central vertical axis ($r = 0$) by letting $\theta$ and $\phi$ vary as coordinates on a 2-sphere $S^2$, the point $s$ actually describes a closed 2-surface $\mathtt{S}$ (coordinatized by $\theta$ and $\phi$) with topology $S^2$ (so $\mathtt{S}$ is intrinsically an ordinary 2-sphere). We shall be concerned with the region $I^+(\mathtt{S})$, which is the (chronological) “future” of $\mathtt{S}$, that is, the locus of points $q$ for which a timelike curve exists having a future endpoint at $q$ and a past endpoint on $\mathtt{S}$. We shall also be interested, particularly, in the boundary $\partial I^+(\mathtt{S})$ of $I^+(\mathtt{S})$. This boundary is described, in Figure 1, by the pair of null curves $v = \text{const.}$ and $2r + 4m \log(r - 2m) = \text{const.}$, proceeding into the future from $s$ (and rotated in $\theta$ and $\phi$). The region $I^+(\mathtt{S})$ itself is represented by that part of Figure 1 which lies between these null curves.

We observe that, in this symmetrical case ($s$ being chosen in the vacuum region), a characterization of $s$ as being “trapped,” in the sense that it lies in a region that is within the horizon, is that the future tangents to these null curves both point “inwards,” in the sense of decreasing $r$. Since $r$ is the metric radius of the $S^2$ of rotation, so that the element of surface area of this sphere is proportional to $r^2$, it follows that the surface area of the boundary $\partial I^+(\mathtt{S})$ reduces, on both branches, as we move away from $\mathtt{S}$ into the future. The three-dimensional region $\partial I^+(\mathtt{S})$ consists of two null surfaces joined along $\mathtt{S}$, in the sense that their Lorentzian normals are null 4-vectors. For each fixed value of $\theta$ and $\phi$, this normal is a tangent to one or other of the two null curves of Figure 1, starting at $s$. For a trapped $s$, these normals point in the direction of decreasing $r$, and it follows that the divergence of these normals is negative (so $\rho > 0$ in what follows below).

In the general case, it is this property of negativity of the divergence, at $\mathtt{S}$, of both sets of Lorentzian normals (i.e., of null tangents to $\partial I^+(\mathtt{S})$), that characterizes $\mathtt{S}$ as a trapped surface, where in the general case we must also prescribe $\mathtt{S}$ to be compact and spacelike. But now there are to be no assumptions of symmetry whatever. Such a characterization is stable against small, but finite, perturbations of the location of $\mathtt{S}$, within the spacetime manifold $\mathtt{M}$, and also against small, but finite, perturbations of $\mathtt{M}$ itself.

We can think of a trapped surface in more direct physical/geometrical terms. Imagine a flash of light emitted all over some spacelike compact spherical surface such as $\mathtt{S}$, but now in ordinary flat spacetime, where for simplicity we suppose that $\mathtt{S}$ is situated in some spacelike (flat) 3-hypersurface $\mathtt{H}$, of constant time $t = 0$. There will be one component to the flash proceeding outwards and another proceeding inwards. Provided that $\mathtt{S}$ is convex, the outgoing flash will represent an initial increase of the surface area at every point of $\mathtt{S}$ and the ingoing flash, an initial decrease. In four-dimensional spacetime terms, we express this as positivity of the divergence of the outward null normal and the negativity of the divergence of the inward one. The characteristic feature of a trapped surface is that whereas the ingoing flash will still have an initially reducing surface area, the “outgoing” flash now has the curious property that its surface area is also initially decreasing, this holding at every point of $\mathtt{S}$.

Locally, this is not particularly strange. For a surface wiggling in and out, we are quite likely to find portions of ingoing flash with increasing area,

and portions of outgoing flash with decreasing area. An extreme case in Minkowski spacetime has S as the intersection of two past light cones. All the null normals to S point along the generators of these past cones, and therefore all converge into the future. Such a surface S (indeed spacelike) looks "trapped" everywhere locally, but fails to count as trapped, not being compact. Since there is nothing causally extreme about Minkowski space, it is appropriate not to count such surfaces as "trapped." What *is* the peculiar about a trapped surface is that both ingoing and outgoing flashes are initially decreasing in area, over the entire compact S. (N. B. Hawking and Ellis (1973) adopt a slightly different terminology; the term "trapped," used here, refers to their "closed trapped.")

## The Null Raychaudhuri Equation

What do we deduce from the existence of a trapped surface? A glance at **Figure 1** gives us some indication of the trouble. As we trace $\partial I^{+}(S)$ into the future, we find that its cross-sectional area continues to decrease, until becoming zero at the central singularity. This last feature need not reflect closely what happens in more general cases, with no spherical symmetry. But the reduction in surface area is a general property. This is the first point to appreciate in a theorem (Penrose 1965, 1968, Hawking and Ellis 1973) which indicates the profoundly disturbing physical implications of the existence of a trapped surface in physically realistic gravitational collapse, according to Einstein's general relativity. The surface-area reduction arises from a result known as "Raychaudhuri's equation," in the case of null rays – where we refer to this as the "Sachs" equations. We come to this next.

Although many different notations are used to express the needed quantities, we can here conveniently employ the spin-coefficient formalism, as described elsewhere in this Encyclopedia (*see* Spinors and Spin Coefficients).

Suppose that we have a congruence (smooth three-parameter family) of rays (null geodesics) in four-dimensional spacetime. Let $\ell^a$ be a real future-null vector, tangent to a null geodesic $\gamma$ of the congruence, and let $m^b$ be complex-null, also defined along $\gamma$, where its real and imginary parts are unit vectors spanning a 2-surface element orthogonal to $\ell^a$ at each point of $\gamma$, so we have

$$\ell_a \ell^a = 0,\ \ell_a m^a = 0,$$
$$m_a m^a = 0,\ \bar{m}_a m^a = -1,$$
$$\ell^a = \bar{\ell}^a$$

where it is assumed that each of $\ell^a, m^a$ is parallel-propagated along $\gamma$:

$$\ell^a \nabla_a \ell^b = 0, \qquad \ell^a \nabla_a m^b = 0$$

($\nabla_a$ denoting covariant derivative). The spin-coefficient quantities

$$\rho = m^a \bar{m}^b \nabla_a \ell_b \qquad \text{and} \qquad \sigma = m^a m^b \nabla_a \ell_b$$

are of importance. Here, the real part of $\rho$ measures the convergence of the congruence and the imaginary part defines its rotation; $\sigma$ measures its shear, where the argument of $\sigma$ defines the direction (perpendicular to $\gamma$) of the axis of shear, and whose strength is defined by $|\sigma|$ (see Penrose and Rindler (1986) for a graphic description of these quantities). Defining propagation derivative along $\gamma$ by

$$\mathrm{D} = \ell^a \nabla_a$$

we can write the Sachs equations as

$$\mathrm{D}\rho = \rho^2 + \bar{\sigma}\sigma + \Phi$$
$$\mathrm{D}\sigma = 2\rho\sigma + \Psi$$

where $\Phi = -(1/2)R_{ab}\ell^a\ell^b$ and $\Psi = C_{abcd}\ell^a m^b \ell^c m^d$, conventions for the Ricci tensor $R_{ab}$ and the Weyl tensor $C_{abcd}$ being those of General Relativity: Overview (and of Penrose and Rindler (1984)). We note that it is the real Ricci component $\Phi$ which governs the propagation of the divergence and the complex Weyl component $\Psi$ which governs the propagation of shear, though there are some non-linear terms. The quantity $\Phi$ is normally taken non-negative, since it measures the energy flux across $\gamma$ (with, in fact $\Phi = 4\pi G T_{ab}\ell^a\ell^b$, where $T_{ab}$ is the energy tensor). The condition that $\Phi \geq 0$ at all points of spacetime and for all null directions $\ell^a$, is called the "weak energy condition." (Again there is a minor discrepancy with Hawking and Ellis (1973) who adopt a somewhat stronger "weak energy condition," which is the above but where $\ell^a$ is also allowed to be future-timelike. Unfortunately, with this terminology, their "weak energy condition" is not strictly weaker than their "strong energy condition.")

It will now be assumed that $\rho$ is real:

$$\rho = \bar{\rho}$$

which is always the case for propagation along the generators of a null hypersurface. The weak energy condition then has an important implication for us. We find that if $A$ is an element of 2-surface area within the plane spanned by the real and imaginary parts of $m^a$, then (this area element being propagated by D along the lines $\gamma$)

$$\mathrm{D}A^{1/2} = -\rho A^{1/2}$$

As a consequence, assuming $\Phi \geq 0$,

$$\mathrm{D}^2 A^{1/2} = -(\bar{\sigma}\sigma + \Phi)A^{1/2} \leq 0$$

This tells us that once the divergence $(-\rho)$ becomes negative, then the area element must reduce to zero sometime in the future along $\gamma$, assuming that $\gamma$ is future-null-complete in the sense that it extends to indefinitely large values of an affine parameter $u$ defined along it, where an affine parameter associated with the parallel-propagated $\ell^a$ satisfies

$$\ell^a \nabla_a u = 1$$

Such a place where the cross-sectional area pinches down to zero is a singularity of the congruence or null hypersurface, referred to as a "caustic." (There are also terminological confusions arising from different authors defining the term "caustic" in slightly different ways. The terminology used here is slightly discrepant from that of Arnol'd (1992) (Chapter 3).)

From this property, it follows that if we have a trapped surface $\mathtt{S}$, then every generator of $\partial I^+(\mathtt{S})$, if extended indefinitely into the future, must eventually encounter a caustic. This, so far, tells us nothing about actual singularities in the spacetime $\mathtt{M}$; even Minkowski space contains many null hypersurfaces with multitudes of caustic points. However, caustics do tell us something significant about sets like $\partial I^+(\mathtt{S})$, which are the boundaries of future sets, and we come to this shortly.

## Causality Properties

First, consider the basic causal relations. If $a$ an $b$ are two points of $\mathtt{M}$, then if there is a nontrivial future-timelike curve in $\mathtt{M}$ from $a$ to $b$ we say that $a$ "chronologically" precedes $b$ and write

$$a \ll b$$

(so it would be possible for some observer's world line to encounter first $a$ and then $b$). If there is a future-null curve in $\mathtt{M}$ from $a$ to $b$ (trivial or otherwise), we say that $a$ "causally" precedes $b$ and write

$$a \prec b$$

(so it would be possible for a signal to get from $a$ to $b$). We have the following elementary properties (see Penrose (1972)):

$$\begin{gathered} a \prec a \\ \text{if } a \ll b \text{ then } a \prec b \\ \text{if } a \ll b \text{ and } b \ll c \text{ then } a \ll c \\ \text{if } a \ll b \text{ and } b \prec c \text{ then } a \ll c \\ \text{if } a \prec b \text{ and } b \ll c \text{ then } a \ll c \\ \text{if } a \prec b \text{ and } b \prec c \text{ then } a \prec c \end{gathered}$$

We generalize the definition of $I^+(\mathtt{S})$, above, to an arbitrary subset $\mathtt{Q}$ in $\mathtt{M}$, obtaining the *chronological future* $I^+(\mathtt{Q})$ and *past* $I^-(\mathtt{Q})$ of $\mathtt{Q}$ in $\mathtt{M}$ by

$$\begin{aligned} I^+(\mathtt{Q}) &= \{q | p \ll q \text{ for some } p \in \mathtt{Q}\} \\ I^-(\mathtt{Q}) &= \{q | q \ll p \text{ for some } p \in \mathtt{Q}\} \end{aligned}$$

The notation $\{q|$ some property of $q\}$ denotes the set of $q$'s with the stated property and the *causal future* $J^+(\mathtt{Q})$ and past $J^-(\mathtt{Q})$ of $\mathtt{Q}$ in $\mathtt{M}$ by

$$\begin{aligned} J^+(\mathtt{Q}) &= \{q | p \prec q \text{ for some } p \in \mathtt{Q}\} \\ J^-(\mathtt{Q}) &= \{q | q \prec p \text{ for some } p \in \mathtt{Q}\} \end{aligned}$$

The $I^{\pm}(\mathtt{Q})$ are always open sets, but the $J^{\pm}(\mathtt{Q})$ are not always closed (though they are for any closed set $\mathtt{Q}$ in Minkowski space). Thus, the sets $I^{\pm}(\mathtt{Q})$ have a more uniform character than the $J^{\pm}(\mathtt{Q})$, and it is simpler to concentrate, here, on the $I^{\pm}(\mathtt{Q})$ sets.

The boundary $\partial I^+(\mathtt{Q})$ of $I^+(\mathtt{Q})$ has an elegant characterization:

$$\partial I^+(\mathtt{Q}) = \{q | I^+(q) \subseteq \partial I^+(\mathtt{Q}), \text{but } q \notin I^+(\mathtt{Q})\}$$

and the corresponding statement holds for $\partial I^-(\mathtt{Q})$. Boundaries of futures also have a relatively simple structure, as is exhibited in the following result (for which there is also a version with past and future interchanged):

**Lemma** *Let $\mathtt{Q} \subseteq \mathtt{M}$ be closed, and $p \in \partial I^+(\mathtt{Q}) - \mathtt{Q}$, then there exists a null geodesic on $\partial I^+(\mathtt{Q})$ with future endpoint at $p$ and which either extends along $\partial I^+(\mathtt{Q})$ indefinitely into the past, or until it reaches a point of $\mathtt{Q}$. It can only extend into the future along $\partial I^+(\mathtt{Q})$ if $p$ is not a caustic point of $\partial I^+(\mathtt{Q})$.*

Beyond a caustic point, the null geodesic would enter into the interior of $I^+(\mathtt{Q})$, but this also happens (more commonly) when crossing another region of null hypersurface on $\partial I^+(\mathtt{Q})$.

We wish to apply this to $\partial I^+(\mathtt{S})$, for a trapped surface $\mathtt{S}$, but we first need a further assumption that $\mathtt{S}$ lies in the interior of the (future) domain of dependence $D^+(\mathtt{H})$ of some spacelike hypersurface $\mathtt{H}$. This region is defined as the totality of points $q$ for which every timelike curve with future endpoint $q$ can be extended into the past until it meets $\mathtt{H}$. One can consider domains of dependence for regions $\mathtt{H}$ other than smooth spacelike surfaces, but it is usual to assume, more generally, that $\mathtt{H}$ is a closed achronal set, where "achronal" means that $\mathtt{H}$ contains no pair of points $a$, $b$ for which $a \ll b$. We find that every point $q$ in the interior int$D^+(\mathtt{H})$ of $D^+(\mathtt{H})$ has the further property that all null curves into the past from $q$ will also eventually meet $\mathtt{H}$ if extended sufficiently. The physical significance of $D^+(\mathtt{H})$ is that, for fields with locally Lorentz-invariant and deterministic evolution equations, the (appropriate) initial data on $\mathtt{H}$ will fix the fields throughout $D^+(\mathtt{H})$ (and also

throughout the similarly defined past domain of dependence $D^{-}(\mathtt{H})$). We find that points in the future Cauchy horizon $H^{+}(\mathtt{H})$, which is the future boundary of $D^{+}(\mathtt{H})$ defined by

$$H^{+}(\mathtt{H}) = D^{+}(\mathtt{H}) - I^{-}(D^{+}(\mathtt{H})),$$

has properties similar to the boundary of a past set, in accordance with the above lemma, and also for the past Cauchy horizon $H^{-}(\mathtt{H})$, defined correspondingly.

## Singularity Theorems and Related Questions

Now, applying our lemma to $\partial I^{+}(\mathtt{S})$, for a trapped surface $\mathtt{S} \subseteq \mathrm{int}D^{+}(\mathtt{H})$, we find that every one of its points lies on a null-geodesic segment $\gamma$ on $\partial I^{+}(\mathtt{S})$, with past endpoint on $\mathtt{S}$ (for if $\gamma$ did not terminate at $\mathtt{S}$ it would have to reach $\mathtt{H}$, which is impossible). Assuming future-null completeness and weak energy ($\Phi \geq 0$), we conclude that if extended far enough into the future, the family of such null geodesics $\gamma$ must encounter a caustic, and therefore they must leave $\partial I^{+}(\mathtt{S})$ and enter $I^{+}(\mathtt{S})$. We finally conclude that $\partial I^{+}(\mathtt{S})$ must be a compact topological 3-manifold. Using basic theorems, we construct an everywhere timelike vector field in $\mathrm{int}D^{+}(\mathtt{H})$ which provides a (1–1) continuous map from the compact $\partial I^{+}(\mathtt{S})$ to $\mathtt{H}$, yielding a contradiction if $\mathtt{H}$ is noncompact, thereby establishing the following (Penrose 1965, 1968):

**Theorem** *The requirement that there be a trapped surface which, together with its closed future, lies in the interior of the domain of dependence of a noncompact spacelike hypersurface, is incompatible with future null completeness and the weak energy condition.*

We notice that this “singularity theorem” gives no indication of the nature of the failure of future null completeness in a spatially open spacetime subject to weak positivity of energy and containing a trapped surface. The natural assumption is that in an actual physical situation of such gravitational collapse, the failure of completeness would arise at places where curvatures mount to such extreme values that classical general relativity breaks down, and must be replaced by the appropriate “quantum geometry” (*see* Quantum Geometry and its Applications, etc.). Hawking (1965) showed how this theorem (in time-reversed form) could also be applied on a cosmological scale to provide a strong argument that the Big-Bang singularity of the standard cosmologies is correspondingly stable. He subsequently introduced techniques from “Morse theory” which could be applied to timelike rather than just null geodesics and, using arguments applied to Cauchy horizons, was able to remove assumptions concerning domains of dependence (e.g., Hawking (1967)). A later theorem (Hawking and Penrose 1970) encompassed most of the earlier ones and had, as one of its implications, that virtually all spatially closed universe models, satisfying a reasonable energy condition and without closed timelike curves, would have to be singular, in this sense of “incompleteness,” but again the topological-type arguments used give little indication of the nature or location of the singularities.

Another issue that is not addressed by these arguments is whether the singularities arising from gravitational collapse are inevitably “hidden,” as in **Figure 1**, by the presence of a horizon – a conjecture referred to as “cosmic censorship” (see Penrose (1969, 1998)). Without this assumption, one cannot deduce that gravitational collapse, in which a trapped surface forms, will lead to a black hole, or to the alternative which would be a “naked singularity.” There are many results in the literature having a bearing on this issue, but it still remains open.

A related issue is that of strong cosmic censorship which has to do with the question of whether singularities might be observable to local observers. Roughly speaking, a naked singularity would be one which is “timelike,” whereas the singularities in black holes might in general be expected to be spacelike (or future-null), and in the Big Bang, spacelike (or past-null). There are ways of characterizing these distinctions purely causally, in terms of past sets or future sets (sets $\mathtt{Q}$ for which $\mathtt{Q} = I^{-}(\mathtt{Q})$ or $\mathtt{Q} = I^{+}(\mathtt{Q})$); see Penrose (1998). If (strong) cosmic censorship is valid, so there are no timelike singularities, the remaining singularities would be cleanly divided into past-type and future-type. In the observed universe, there appears to be a vast difference between the structure of the two, which is intimately connected with the second law of thermodynamics, there appearing to be an enormous constraint on the Weyl curvature (*see* General Relativity: Overview) in the initial singularities but not in the final ones.

Despite the likelihood of singularities arising in their time evolution, it is possible to set up initial data for the Einstein vacuum equations for a wide variety of complicated spatial topologies (*see* Einstein Equations: Initial Value Formulation). On the observational side, however, there seems to be little evidence for anything other than Euclidean spatial topology in our actual universe (which includes black holes). Speculation on the nature of spacetime at the tiniest scales, however, where quantum gravity might be relevant, often involves non-Euclidean topology, however. It may be noted that an early theorem of Geroch established that the constraints of classical Lorentzian geometry do not permit the spatial topology to change without violations of causality (closed timelike curves).

*See also:* Asymptotic Structure and Conformal Infinity; Boundaries for Spacetimes; Computational Methods in General Relativity: The Theory; Cosmology: Mathematical Aspects; Critical Phenomena in Gravitational Collapse; Einstein Equations: Exact Solutions; Einstein Equations: Initial Value Formulation; General Relativity: Overview; Geometric Analysis and General Relativity; Lorentzian Geometry; Quantum Cosmology; Quantum Geometry and its Applications; Spinors and Spin Coefficients; Stationary Black Holes.

## Further Reading


Arno'ld, Beem JK, and Ehrlich PE (1996) *Global Lorentzian Geometry*, (2nd edn). New York: Marcel Dekker.

Eddington AS (1924) A comparison of Whitehead's and Einstein's formulas. *Nature* 113: 192.

Finkelstein D (1958) Past–future asymmetry of the gravitational field of a point particle. *Physical Review* 110: 965–967.

Hawking SW (1965) *Physical Review Letters* 15: 689.

Hawking SW (1967) The occurrence of singularities in cosmology III. Causality and singularities. *Proceedings of the Royal Society (London) A* 300: 187–.

Hawking SW and Ellis GFR (1973) *The Large-Scale Structure of Space-Time:* Cambridge: Cambridge University Press.

Hawking SW and Penrose R (1970) The singularities of gravitational collapse and cosmology. *Proceedings of the Royal Society (London) A* 314: 529–548.

Oppenheimer JR and Snyder H (1939) On continued gravitational contraction. *Physical Review* 56: 455–459.

Penrose R (1965) Gravitational collapse and space-time singularities. *Physics Review Letters* 14: 57–59.

Penrose R (1972) *Techniques of Differential Topology in Relativity*. CBMS Regional Conference Series in Applied Mathematics, no.7. Philadelphia: SIAM.

Penrose R and Rindler W (1984) *Spinors and Space-Time, Vol. 1: Two-Spinor Calculus and Relativistic Fields*. Cambridge: Cambridge University Press.

Penrose R and Rindler W (1984) *Spinors and Space-Time, Vol. 2: Spinor and Twistor Methods in Space-Time Geometry*. Cambridge: Cambridge University Press.

Penrose R (1968) Structure of space-time. In: DeWitt CM and Wheeler JA (eds.) *Battelle Rencontres, 1967 Lectures in Mathematics and Physics*. New York: Benjamin.

Penrose R (1969) Gravitational collapse: the role of general relativity. *Rivista del Nuovo Cimento Serie I, Numero Speciale* 1: 252–276.

Penrose R (1998) The question of cosmic censorship. In: Wald RM (ed.) *Black Holes and Relativistic Stars*, (reprinted in (1999) Journal of Astrophysics and Astronomy 20, 233–248.) Chicago, IL: University of Chicago Press.


**Special Lagrangian Submanifolds** *see* Calibrated Geometry and Special Lagrangian Submanifolds

# Spectral Sequences

**P Selick**, University of Toronto, Toronto, ON, Canada



## Introduction

Spectral sequences are a tool for collecting and distilling the information contained in an infinite number of long exact sequences. Their most common use is the calculation of homology by filtering the object under study and using a spectral sequence to pass from knowledge of the homology of the filtration quotients to that of the object itself. This article will discuss the construction of spectral sequences and the notion of convergence including conditions sufficient to guarantee convergence. Some sample applications of spectral sequences are given.

A differential on an abelian group $G$ is a self-map $d: G \to G$ such that $d^2 = 0$. A morphism of differential groups is a map $f: G \to G'$ such that $d'f = fd$. The condition $d^2 = 0$ guarantees that $\operatorname{Im} d \subset \operatorname{Ker} d$, so to the differential group $(G, d)$ we can associate its homology, $H(G, d) := \operatorname{Ker} d / \operatorname{Im} d$. Often $G$ has extra structure and we require $d$ to satisfy some compatibility condition in order that $H(G, d)$ should also have this structure. For example, a differential graded Lie algebra $(L, d)$ requires a differential $d$ *which satisfies the condition* $d[x, y] = [dx, y] + (-1)^{|x|}[x, dy]$. *While, for simplicity, throughout this article we will always assume that* $G$ is an abelian group, the concepts are readily extended to the case where $G$ is an object of some abelian category and generalizations to nonabelian situations have also been studied.

An important example of extra structure is the case where $G = \bigoplus_{n=-\infty}^{\infty} G_n$ is a graded abelian group. The appropriate compatibility condition for a differential graded group is that $d$ should be homogeneous of degree –1. That is, $d(G_n) \subset G_{n-1}$. In many contexts it is more natural to use superscripts and regard $d$ as having degree +1; the two concepts are equivalent via the reindexing convention $G^n := G_{-n}$. Another important example is that

where $G$ forms a graded algebra, meaning that it has a multiplication $G_n \otimes G_k \to G_{n+k}$. To form a differential graded algebra, in addition to having degree –1, $d$ is required to satisfy the Leibniz rule $d(xy) = d(x)y + -(1)^{|x|}xd(y)$ (where $|x|$ denotes the degree of $x$) familiar from the differentiation of differential forms.

In many cases, $G$ itself is not the main object of interest, but is a relatively large and complicated object, $G = G(X)$, formed by applying some functor $G$ to the object $X$ being studied. For example, $X$ might be some manifold and $G$ could be the set of all differential forms on $X$ with the exterior derivative as $d$. The presumption is that $H(G(X))$ carries the information we want about $X$ in a much simpler form than the whole of $G(X)$.

A *spectral sequence* (Leray 1946) is defined simply as a sequence $((E^r, d^r))_{r=n_0, n_0+1, \ldots,}$ of differential abelian groups such that $E^{r+1} = H(E^r, d^r)$. By reindexing, we could always arrange that $n_0 = 1$, but sometimes it is more natural to begin with some other integer. If all terms $(E^r, d^r)$ of the spectral sequence have the appropriate additional structure, we might refer, for example, to a spectral sequence of Lie algebras. If there exists $N$ such that $E^r = E^N$ for all $r \geq N$ (equivalently $d^r = 0$ for all $r \geq N$), the spectral sequence is said to "collapse" at $E^N$.

The definition of spectral sequence is so broad that we can say almost nothing of interest about them without putting on some additional conditions. We will begin by considering the most common type of spectral sequence, historically the one that formed the motivating example: the spectral sequence of a filtered chain complex.

## Filtered Objects

To study a complicated object $X$, it often helps to filter $X$ and study it one filtration at a time. A filtration $\mathcal{F}_X$ of a group $X$ is a nested collection of subgroups

$$\mathcal{F}_X := \ldots F_nX \subset F_{n+1}X \subset \cdots \subset X \quad -\infty < n < \infty$$

A morphism $f : \mathcal{F}_X \to \mathcal{F}_Y$ of filtered groups is a homomorphism $f : X \to Y$ such that $f(F_n(X)) \subset F_n(Y)$. The groups $F_nX/F_{n-1}X$ are called the "filtration quotients" and their direct sum $\mathrm{Gr}(\mathcal{F}_X) := \bigoplus_n F_nX/F_{n-1}X$ is called the *associated graded group* of the filtered group $\mathcal{F}_X$. In cases where $X$ has additional structure, we might define special types of filtrations satisfying some compatibility conditions so that $\mathrm{Gr}(\mathcal{F}_X)$ inherits the additional structure. For example, an *algebra filtration* of an algebra $X$ is defined as one for which $(F_nX)(F_kX) \subset F_{n+k}X$.

Since our plan is to study $X$ by computing $\mathrm{Gr}(\mathcal{F}_X)$, the first question we need to consider is what conditions we need to place on our filtration so that $\mathrm{Gr}(\mathcal{F}_X)$ retains enough information to recover $X$. Our experience from the "5-lemma" suggests that the appropriate way to phrase the requirement is to ask for conditions on the filtrations which are sufficient to conclude that $f : X \to Y$ is an isomorphism whenever $f : \mathcal{F}_X \to \mathcal{F}_Y$ is a morphism of filtered groups for which the induced $\mathrm{Gr}(f) : \mathrm{Gr}(X) \to \mathrm{Gr}(Y)$ is an isomorphism.

It is clear that $\mathrm{Gr}\mathcal{F}_X$ can tell us nothing about $X - (\cup X_n)$ so we require that $X = \cup X_n$. Similarly we need that $\cap X_n = 0$. However, the latter condition is insufficient as can be seen from the following example.

**Example 1** Let $X := \bigoplus_{k=1}^{\infty} \mathbb{Z}$ and $Y := \prod_{k=1}^{\infty} \mathbb{Z}$. Set

$$F_nX := \begin{cases} X & \text{if } n \geq 0 \\ \bigoplus_{k=-n}^{\infty} \mathbb{Z} & \text{if } n < 0 \end{cases}$$

$$F_nY := \begin{cases} Y & \text{if } n \geq 0 \\ \prod_{k=-n}^{\infty} \mathbb{Z} & \text{if } n < 0 \end{cases}$$

and let $f : X \to Y$ be the inclusion. Then $\mathrm{Gr}(f)$ is an isomorphism but $f$ is not.

To phrase the appropriate condition we need the concept of algebraic limits. Given a sequence of objects $\{X_n\}_{n \in \mathbb{Z}}$ and morphisms $f_n : X_n \to X_{n+1}$ in some category, the "direct limit" or "colimit" of the sequence, written $\varinjlim_n F_nX$, is an object $X$ together with morphisms $g_n : X_n \to X$ satisfying $g_{n+1} \circ f_n = g_n$, having the universal property that given any object $X'$ together with maps $g'_n : X_n \to X'$ satisfying $g'_{n+1} \circ f_n = g'_n$, there exists a unique morphism $h : X \to X'$ such that $g'_n = h \circ g_n$ for all $n$. By the usual categorical argument the object $X$, if it exists, is unique up to isomorphism. The dual concept, "inverse limit" or simply "limit" of the sequence, written $\varprojlim_n F_nX$, is obtained by reversing the directions of the morphisms. For intuition, we note that these notions share, with the notion of limits of sequences in calculus, the properties that changing the terms $X_n$ only for $n < N$ does not affect $\varinjlim_n F_nX$, and if the sequence stabilizes at $N$ (i.e., the morphisms $f_n$ are isomorphisms for all $n \geq N$), then $\varinjlim_n F_nX \cong X_N$. Similarly $\varprojlim_n F_nX$ depends only upon behavior of the sequence as $n \to -\infty$. Limits over partially ordered sets other than $\mathbb{Z}$ can also be taken but we shall not need them in this article. Although limits need not exist in general, in the category of abelian groups, both the direct and inverse limit exist for any sequence and are given explicitly by the following constructions. $\varinjlim_n F_nX = \bigoplus X_n/\sim$ where, letting $i_k : X_k \to \bigoplus X_n$ be the

canonical inclusion, the equivalence relation is generated by $i_n(x) \sim i_{n+1}f(x)$ for $x \in X_n$. $\varprojlim_n F_n X = \{(x_n) \in \Pi X_n \mid f_n(x_n) = x_{n+1}\, \forall n\}$.

The condition needed is that our filtrations should be *bicomplete*, defined as follows. $\mathcal{F}_X$ is called "cocomplete" if the canonical map $X \to \varinjlim_n F_n X$ is an isomorphism and $\mathcal{F}_X$ is called "complete" if $X \to \varprojlim_n X/F_n X$ is an isomorphism. $\mathcal{F}_X$ is called bicomplete if it is both complete and cocomplete. Note that $\mathcal{F}_X$ cocomplete is equivalent to $\cup F_n X = X$ but $\mathcal{F}_X$ complete is stronger than $\cap F_n X = 0$.

**Theorem 1** (Comparison theorem). *Let $\mathcal{F}_X$ be bicomplete and let $\mathcal{F}_Y$ be cocomplete with $\cap F_n Y = 0$. Suppose that $f : \mathcal{F}_X \to \mathcal{F}_Y$ is a morphism such that* $\mathrm{Gr}(f) : \mathrm{Gr}(X) \to \mathrm{Gr}(Y)$ *is an isomorphism. Then $f : X \to Y$ is an isomorphism.*

## Filtered Chain Complexes

A *chain complex* $(C, d)$ of abelian groups consists of abelian groups $C_n$ for $n \in \mathbb{Z}$ together with homomorphisms $d_n : C_n \to C_{n-1}$ such that $d_n \circ d_{n+1} = 0$ for all $n$. To the chain complex $(C, d)$ we can associate the differential (abelian) group $(C_*, d) := \bigoplus_{n=-\infty}^{\infty} C_n$ with $d|_{C_n}$ induced by $d_n$. We often write simply $C$ if the differential is understood. The dual notion in which $d$ has degree +1 is called a *cochain complex* and the concepts are equivalent through our convention $C^n := C_{-n}$.

**Theorem 2** (Homology commutes with direct limits). $H(\varinjlim_n C_n) = \varinjlim_n H(C_n)$.

As we shall see later, failure of homology to commute with inverse limits is a source of great complication in working with spectral sequences.

Let $\mathcal{F}_C$ be a filtered chain complex. In many applications, our goal is to compute $H_*(C)$ from a knowledge of $H_*(F_n C/F_{n-1}C)$ for all $n$. The overall plan, which is not guaranteed to be successful in general, would be:

1. use the given filtration on $C$ to define a filtration on $H_*(C)$,
2. use our knowledge of $H_*(\mathrm{Gr}\, C)$ to compute $\mathrm{Gr}\, H_*(C)$,
3. reconstruct $H_*(C)$ from $\mathrm{Gr}\, H_*(C)$.

To begin, set $F_n(H_*C) := \mathrm{Im}(s_n)_*$, where $s_n : F_n(C) \to C$ is the inclusion (chain) map from the filtration. The spectral sequence which we will define for this situation can be regarded as a method of keeping track of the information contained in the infinite collection of long exact homology sequences coming from the short exact sequences $0 \to F_{n-1}C \to F_n C \to F_n C/F_{n-1}C \to 0$. When working with a long exact sequence, knowledge of two of every three terms gives a handle on computing the remaining terms but does not, in general, completely determine those terms, which explains intuitively why we have some reason to hope that a spectral sequence might be useful and also why it is not guaranteed to solve our problem.

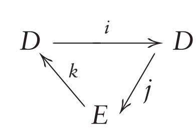
Before proceeding with our motivating example, we digress to discuss spectral sequences formed from *exact couples*.

## Exact Couples

In this section, we will define exact couples, show how to associate a spectral sequence to an exact couple, and discuss some properties of spectral sequences coming from exact couples. As we shall see, a filtered chain complex gives rise to an exact couple and we will examine this spectral sequence in greater detail.

Exact couples were invented by Massey and many books use them as a convenient method of constructing spectral sequences. Other books bypass discussion of exact couples and define the spectral sequence coming from a filtered chain complex directly.

**Definition 1** An "exact couple" consists of a triangle




containing abelian groups $D$, $E$, and together with homomorphisms $i$, $j$, $k$ such that the diagram is exact at each vertex.

In the following, to avoid conflicting notation considering the many superscripts and subscripts which will be needed, we use the convention that an $n$-fold composition will be written $f^{\circ n}$ rather than the usual $f^n$.

Given an exact couple, set $d := jk : E \to E$. By exactness, $kj = 0$, so $d^{\circ 2} = jkjk = 0$ and therefore $(E, d)$ forms a differential group. To the exact couple we can associate another exact couple, called its derived couple, as follows. Set $D' := \mathrm{Im}\, i \subset D$ and $E' := H(E, d)$. Define $i' := i|_D$ and let $j' : D' \to E'$ be given by $j'(iy) := \overline{j(y)}$, where $\overline{x}$ denotes the equivalence class of $x$. The map $k' : E' \to D'$ is defined by $k'(\overline{z}) := kz$. One checks that the maps $j'$ and $k'$ are well defined and that $(D', E', i', j', k')$ forms an exact couple. Therefore, from our original exact couple, we can inductively form a sequence of exact couples $(D^r, E^r, i^r, j^r, k^r)_{r=1}^{\infty}$ with $D^1 := D, E^1 := E, D^r := (D^{r-1})'$

and $E^r := (E^{r-1})'$. This gives a spectral sequence $(E^r, d^r)_{r=1}^{\infty}$ with $d^r = j^r k^r$.

To the filtered chain complex $\mathcal{F}_C$, we can associate an exact couple as follows. Set $D := \bigoplus_{p,q} D_{p,q}$ where $D_{p,q} = H_{p+q}(F_p C)$ and $E := \bigoplus_{p,q} E_{p,q}$ where $E_{p,q} = H_{p+q}(F_p C/F_{p-1}C)$. The long exact homology sequences coming from the sequences $0 \to F_{p-1}C \xrightarrow{a} F_p C \xrightarrow{b} F_p C/F_{p-1}C \to 0$ give rise, for each $p$ and $q$, to maps $a_* : D_{p-1,q+1} \to D_{p,q}, b_* : D_{p,q} \to E_{p,q}$, and $\partial : E_{p,q} \to D_{p-1,q}$. Define $i : D \to D$ to be the map whose restriction to $D_{p-1,q+1}$ is the composition of $a_*$ with the canonical inclusion $D_{p,q} \to D$. Similarly, define $j : D \to E$ and $k : E \to D$ to be the maps whose restrictions to each summand are the compositions of $b_*$ and $\partial$ with the inclusions. The indexing scheme for the bigradations is motivated by the fact that in many applications it causes all of the nonzero terms to appear in the first quadrant, so it is the most common choice, although one sometimes sees other conventions.

There is actually a second exact couple we could associate to $\mathcal{F}_C$, which yields the same spectral sequence: use the same $E$ as above but replace $D$ by $\bigoplus D_{p,q}$ with $D_{p,q} = H_{p+q+1}(C/F_p C)$, and define $i$, $j$, and $k$ in a manner similar to that above.

When dealing with cohomology rather than homology, the usual starting point would be a system of inclusions of cochain complexes $\cdots F^{n+1}C \subset F^n C \subset F^{n-1} \subset \cdots \subset C$. This can be reduced to the previous case by replacing the cochain complex $C$ by a chain complex $C_*$ using the convention $C_p := C^{-p}$ and filtering the result by $F_n C_* := F^{-n}C$. The usual practice, equivalent to the above followed by a rotation of 180°, is to leave the original indices and instead reverse the arrows in the exact couple. In this case, it is customary to write $D_r^{p,q}$ and $E_r^{p,q}$ for the terms in the exact couple and spectral sequence.

In applications, it is often the case that $E^1$ is known and that our goal includes computing $D^1$. The example of the filtered chain complex with the assumption that we know $H_*(F_p C/F_{p-1}C)$ for all $p$ is fairly typical.

Since each $D^r$ is contained in $D^{r-1}$ and each $E^r$ is a subquotient of $E^{r-1}$, the terms of these exact couples get smaller as we progress. To get properties of the spectral sequence, we need to examine this process and, in particular, analyze that which remains in the spectral sequence as we let $r$ go to infinity.

For $x \in E$, if $dx = 0$ then $\bar{x}$ belongs to $E^2$ and so $d^2(\bar{x})$ is defined. In the following, we shall usually simplify the notation by writing simply $x$ in place of $\bar{x}$ and writing $d^r x = 0$ to mean "$d^r x$ is defined and equals 0."

If $dx = 0, \ldots, d^{r-1}x = 0$, then $x$ represents an element of $E^r$ and $d^r x$ is defined. Set $Z^r := \{x \in E \mid d^m x = 0 \, \forall m \leq r\}$. Then $E^{r+1} \cong Z^r/\sim$ where $x \sim y$ if there exists $z \in E$ such that for some $t \leq r$ we have $d^m z = 0$ for $m < t$ (thus $d^t z$ is defined) and $d^t z = x - y$. With this as motivation, we set $Z^\infty := \cap_r Z^r = \{x \in E \mid d^m x = 0 \ \forall m\}$ (known as the "infinite cycles") and define $E^\infty := Z^\infty/\sim$ where $x \sim y$ if there exists $z \in E$ such that for some $t$ we have $d^m z = 0$ for $m < t$ and $d^t z = x - y$.

Notice that $D^{r+1} = \mathrm{Im} i^{\circ r} \cong D/\mathrm{Ker}\, i^{\circ r}$. There is no analog of this statement for $r = \infty$. Instead we have separate concepts so we set $D^\infty := D/\cup_r \mathrm{Ker}\, i^{\circ r}$ and ${}^\infty D := \cap_r \mathrm{Im} i^{\circ r}$. The analog of the $r$th-derived exact couple when $r = \infty$ is the following exact sequence.

**Theorem 3** *There are maps induced by $i$, $j$, and $k$ producing an exact sequence*

$$0 \to D^\infty \xrightarrow{i^\infty} D^\infty \xrightarrow{j^\infty} E^\infty \xrightarrow{{}^\infty k} {}^\infty D \xrightarrow{{}^\infty i} {}^\infty D$$

The fact that we were able to add the 0 term to the left of this sequence but not the right can be traced to the fact that $\varinjlim$ preserves exactness but $\varprojlim$ does not.

In our motivating example, the terms of the initial exact couple came with a bigrading $D = \bigoplus D_{p,q}$ and $E = \bigoplus E_{p,q}$ and writing $|f|$ for the bidegree of a morphism $f$ we had: $|i| = (1, -1); |j| = (0, 0); |k| = (-1, 0); d = (-1, 0)$. It follows that $|i^r| = (1, -1); |j^r| = (-r+1, r-1); |k^r| = (-1, 0); |d^r| = (-r, r-1)$ which is considered the standard bigrading for a bigraded exact couple. Similarly, the standard bigrading for a bigraded spectral sequence is one such that $|d^r| = (-r, r-1)$.

We observed earlier that terms of an exact couple and its corresponding spectral sequence get smaller as $r \to \infty$ as each is a subquotient of its predecessor. Note that the bigrading is such that this applies to each pair of coordinates individually (e.g., $E_{p,q}^{r+1}$ is a subquotient of $E_{p,q}^r$) and so in particular if the $p$, $q$-position ever becomes 0 that position remains 0 forevermore.

## Convergence of Graded Spectral Sequences

As noted earlier, the definition of spectral sequence is so broad that we need to put some conditions on our spectral sequences to make them useful as a computational tool. From now on, we will restrict attention to spectral sequences arising from exact couples in which $D = \bigoplus D_p$ and $E = \bigoplus E_p$ are graded with $i|_{D_p} \subset D_{p+1}, j|_{D_p} \subset E_p$, and $k|_{E_p} \subset D_{p-1}$. All the spectral sequences which have been studied

to date satisfy this condition and in fact most also have a second gradation as in the case of our motivating example. To see how to proceed, we examine that case more closely.

For a filtered chain complex $\mathcal{F}_C$ with structure maps $s_p : F_pC \to C$ we defined $F_p(H_*(C)) = \operatorname{Im} s_{p*}$. If $x = i^{\circ(r-1)}y$ belongs to

$$D^r_{p,q} = \operatorname{Im} i^{\circ(r-1)} : H_{p+q}(F_{p-r+1}C) \to H_{p+q}(F_pC)$$

then $(s_p)_*x = (s_p)_* i^{\circ(r-1)}y = (s_{p+1})_* i^{\circ r}y = (s_{p+1})_* ix$. Therefore, we have a commutative diagram

$$\begin{array}{ccc} D^r_{p,q} & \longrightarrow & F_p(H_{p+q}(C)) \\ \downarrow i & & \downarrow \\ D^{r+1}_{p+1,q-1} & \longrightarrow & F_{p+1}(H_{p+q}(C)) \end{array}$$

yielding a map

$$D^{r+1}_{p+1,q-1}/D^r_{p,q} \to F_{p+1}\big(H_{p+q}(C)\big)/F_p\big(H_{p+q}(C)\big) = \operatorname{Gr}_{p+1}(H_{p+q})C$$

Letting $r$ go to infinity, we get an induced map $\phi : D^\infty / i^\infty(D^\infty) \to \operatorname{Gr}(H(C))$.

**Theorem 4** *If $\mathcal{F}_H(C)$ is cocomplete then*

(i) $D^\infty = F_n(H(C))$;
(ii) $\phi : D^\infty / i^\infty(D^\infty) \to \operatorname{Gr}(H(C))$ *is an isomorphism*;
(iii) *There is an exact sequence* $0 \to \operatorname{Gr}(H(C)) \xrightarrow{i^\infty} E^\infty \xrightarrow{\infty k} {}^\infty D \xrightarrow{\infty i} {}^\infty D$.

We say that the spectral sequence $(E^r)$ "abuts" to $\mathcal{F}_L$ if there is an isomorphism $\operatorname{Gr}L \to E^\infty$. Here we mean an isomorphism of graded abelian groups, which makes sense since under our assumptions $E^r$ inherits a grading from $E^1$ for each $r$. If in addition the filtration on $L$ is cocomplete, we say that $(E^r)$ "weakly converges" to $\mathcal{F}_L$ and if it is bicomplete we say that $(E^r)$ "converges" (or *strongly converges*) to $\mathcal{F}_L$. The notation $(E^r) \Rightarrow \mathcal{F}_L$ (or simply $(E^r) \Rightarrow L$ when the filtration on $L$ is either understood or unimportant) is often used in connection with convergence but there is no universal agreement as to which of the three concepts (abuts, weakly converges, or converges) it refers to! In this article, we will also use the expression $(E^r)$ "quasiconverges" to $\mathcal{F}_L$ to mean that the spectral sequence weakly converges to $\mathcal{F}_L$ with $\cap_n F_n L = 0$. (Note: the terminology quasiconverges is nonstandard although the concept has appeared in the literature, sometimes under the name *converges*.)

While it would be overstating things to claim that convergence of the spectral sequence shows that $E^\infty$ determines $H(C)$, it is clear that convergence is what we need in order to expect that $E^\infty$ contains enough information to possibly reconstruct $H(C)$. The sense in which this is true is stated more precisely in the following theorem.

**Theorem 5** (Spectral sequence comparison theorem). *Let $f = (f^r) : (E^r) \to \tilde{E}^r$ be a morphism of spectral sequences.*

(i) *If $f : E^N \to \tilde{E}^N$ is an isomorphism for some $N$, then $f^r$ is an isomorphism for all $r \geq N$ (including $r = \infty$).*
(ii) *Suppose in addition that $(E^r)$ converges to $\mathcal{F}_X$ and $(\tilde{E}^r)$ quasiconverges to $\mathcal{F}_{\tilde{X}}$. Let $\phi : \mathcal{F}_X \to \mathcal{F}_{\tilde{X}}$ be a morphism of filtered abelian groups which is compatible with $f$. (i.e., there exist isomorphisms $\eta : \operatorname{Gr} X \cong E^\infty$ and $\tilde{\eta} : \operatorname{Gr} \tilde{X} \cong \tilde{E}^\infty$ such that $f^\infty \circ \eta = \tilde{\eta} \circ \operatorname{Gr}(f)$). Then $f : X \to \tilde{X}$ is an isomorphism.*

Within the constraints provided by Theorem 5, a spectral sequence might have many limits. A typical calculation of some group $Y$ by means of spectral sequences might proceed as an application of Theorem 5 along the lines of the following plan.

1. Subgroups $F_nY$ forming a filtration of $Y$ are defined, although usually not computable at this point. The subgroups are chosen in a manner that seems natural bearing in mind that to be useful it will be necessary to show convergence properties.
2. Directly or by means of an exact couple, a spectral sequence is defined in a manner that seems to be related to the filtration.
3. Some early term of the spectral sequence (usually $E^1$ or $E^2$) is calculated explicitly and the differentials $d^r$ are calculated successively resulting in a computation of $E^\infty$.
4. With the aid of the knowledge of $E^\infty$, a conjecture $Y = G$ is formulated for some $G$.
5. A suitable filtration on G and a map of filtrations $\mathcal{F}_G \to \mathcal{F}_Y$ or $\mathcal{F}_Y \to \mathcal{F}_G$ are defined.
6. The spectral sequence arising from $\mathcal{F}_G$ is demonstrated to converge to $G$.
7. The original spectral sequence is demonstrated to converge to $Y$ and Theorem 5 is applied.

The hardest steps are usually (3) and (7). For step (3), in most cases the calculations require knowledge which cannot be obtained from the spectral sequence itself, although the spectral sequence machinery plays its role in distilling the information and pointing the way to exactly what needs to be calculated. Steps (4)–(6) are frequently very easy, and often not stated explicitly, with "by construction of $G$" being the most common justification of (6). We now discuss the types of considerations involved in step (7).

Convergence of a spectral sequence to a desired $L$ can be difficult to verify in general partly because

the conditions are stated in terms of some filtration (usually understood only in a theoretical sense) on an initially unknown $L$ rather than in terms of properties of the spectral sequence itself or an exact couple from which it arose. Theorems 2 and 4(ii) give us the following extremely important special case in which we can conclude convergence to $H(C)$ of the spectral sequence for $\mathcal{F}_C$ based on conditions that are often easily checked.

**Theorem 6** *If $\mathcal{F}_C$ is a filtered chain complex such that $\mathcal{F}_C$ is cocomplete and there exists M such that $H(F_nC)=0$ for $n<M$, then the spectral sequence for $\mathcal{F}_C$ converges to $H(C)$.*

Although the second hypothesis, which implies that ${}^{\infty}D=0$, is very strong it handles the large numbers of commonly used filtrations which are 0 in negative degrees.

Under the conditions of Theorem 6, inserting the bigradings into Theorem 4 gives a short exact sequence $0\to D^{\infty}_{p-1,q+1}\to D^{\infty}_{p,q}\to E^{\infty}_{p,q}\to 0$ with $D^{\infty}_{p,q}\cong F_p(H_{p+q}(X))$; equivalently

$$F_k(H_n(C))/F_{k-1}(H_n(C))\cong E^{\infty}_{k,n-k}$$

Thus, the only $E^{\infty}$-terms relevant to the computation to $H_n(C)$ are those on the diagonal $p+q=n$. In the important case of a first quadrant spectral sequence ($E^r_{p,q}=0$ if $p<0$ or $q<0$), the number of nonzero terms on any diagonal is finite so the $E^{\infty}$-terms on the diagonal $p+q=n$ give a finite composition series for each $H_n(C)$.

Here is an elementary example of an application of a spectral sequence.

**Example 2** Let $S_*(\,)$ denote the singular chain complex, let $H_*(\,):=H_*(S_*(\,))$ denote singular homology, and let $H^{\text{cell}}_*(\,)$ denote cellular homology. Let $X$ be a CW-complex with $n$-skeleton $X^{(n)}$. The inclusions $S_*(X^{(n)})\to S_*(X)$ yield a filtration on $S_*(X)$. In the associated spectral sequence,

$$E^1_{p,q}=H_{p+q}\left(X^{(p)}/X^{(p-1)}\right)\cong\begin{cases}\text{free abelian group on the } p\text{-cells of } X & \text{if } q=0\\ 0 & \text{if } q\neq 0\end{cases}$$

The differential

$$d^1_{p,0}\,:\,H_p\left(X^{(p)}/X^{(p-1)}\right)\to H_{p-1}\left(X^{(p-1)}/X^{(p-2)}\right)$$

is the definition of the differential in cellular homology. Therefore,

$$E^2_{p,q}=\begin{cases}H^{\text{cell}}(X) & \text{if } q=0\\ 0 & \text{if } q\neq 0\end{cases}$$

Looking at the bidegrees, the domain or range of $d^2_{p,q}$ is zero for each $p$ and $q$ so $d^2=0$, and similarly $d^r=0$ for all $r>2$. Therefore, the spectral sequence collapses with $E^2=E^{\infty}$. The spectral sequence converges to $H_*(X)$ so the terms on the diagonal $p+q=n$ form a composition series for $H_n(X)$. Since the $(n,0)$ term is the only nonzero term on this diagonal, $H_n(X)\cong H^{\text{cell}}_n(X)$. That is, "cellular homology equals singular homology."

Returning to the general situation, set $L_{\infty}:=\varinjlim_n D_n$ and $L_{-\infty}:=\varprojlim_n D_n$. Filter $L_{\infty}$ by $F_nL_{\infty}:=\operatorname{Im}(D_n\to L_{\infty})$ and filter $L_{-\infty}$ by $F_nL_{-\infty}:=\operatorname{Ker}(L_{-\infty}\to D_n)$. It follows from the definitions that $F_nL_{\infty}=D^{\infty}_n$ and so $D^{\infty}_n/i^{\infty}(D^{\infty}_{n-1})=\operatorname{Gr}_nL_{\infty}$. At the other end, the canonical map $L_{-\infty}\to D_n$ lifts to ${}^{\infty}D_n$ yielding an injection $L_{-\infty}/F_nL_{-\infty}\to {}^{\infty}D_n$. Therefore, for each $n$ there is an injection $\operatorname{Gr}_nL_{-\infty}\to K_n$ where $K_n=\operatorname{Ker}({}^{\infty}D_{n-1}\to {}^{\infty}D_n)$. In general, the map $L_{-\infty}\to {}^{\infty}D_n$ need not be surjective (an element could be in the image of $i^{\circ r}$ for each finite $r$ without being part of a consistent infinite sequence), although it is surjective in the special case when ${}^{\infty}D_s\to {}^{\infty}D_{s+1}$ is surjective for each $s$. In the latter case we get $\operatorname{Gr}L_{-\infty}\cong K$. As we will see in the next section, the exact sequence of Theorem 3 extends to the right (Theorem 8) giving $\varprojlim^1_r Z^r=0$ as a sufficient condition that ${}^{\infty}D_s\to {}^{\infty}D_{s+1}$ be surjective for each $s$, where $\varprojlim^1$ is described in that section and $(Z^r)$ refers to the system of inclusions $\cdots\subset Z^{r+1}\subset Z^r\subset Z^{r-1}\subset\cdots$. Thus, $\varprojlim^1_r Z^r=0$ is a sufficient condition for $\operatorname{Gr}L_{-\infty}\cong K$.

Taking into account the short exact sequence $0\to D^{\infty}/i^{\infty}(D^{\infty})\to E^{\infty}\to K\to 0$ coming from Theorem 3, the preceding discussion yields two obvious candidates for a suitable $\mathcal{F}_L$: $\mathcal{F}_{L_{\infty}}$ or $\mathcal{F}_{L_{-\infty}}$. In theory there are other possibilities, but in practice one of these two cases usually occurs. We examine them individually and see what additional conditions are required for convergence.

*Case* I: *Conditions for convergence to $\mathcal{F}_{L_{\infty}}$* It is easily checked from the definitions that $\varinjlim_n D^{\infty}_n=\varinjlim_n D_n$ so $\mathcal{F}_{L_{\infty}}$ is always cocomplete. Therefore, besides $\operatorname{Gr}L_{\infty}\cong E^{\infty}$ (equivalently, $K=0$), it is required to verify that $\mathcal{F}_{L_{\infty}}$ is complete. As we will see in the next section, the completeness condition can be restated as $\cap D_n=0$ and $\varprojlim^1_n D_n=0$. According to the preceding discussion, under the assumption that $L_{-\infty}=\cap D^n=0$, which we need anyway as part of the requirement that $\mathcal{F}_{L_{\infty}}$ be complete, $\varprojlim^1_r Z_rX=0$ is sufficient to show $K=0$.

*Case* II: *Conditions for convergence to $\mathcal{F}_{L_{-\infty}}$* Any inverse limit is complete in its canonical filtration, so $\mathcal{F}_{L_{-\infty}}$ is always complete and the issues are whether $\operatorname{Gr}L_{-\infty}\cong E^{\infty}$ and whether $\mathcal{F}_{L_{-\infty}}$ is cocomplete. $\mathcal{F}_{L_{-\infty}}$ is cocomplete if and only if every element of

$L_{-\infty}$ lies in $\mathrm{Ker}(L_{-\infty} \to D_n)$ for some $n$, for which a sufficient condition is that $L_\infty = 0$ or equivalently $E^\infty \cong K$. Therefore, if the reason for the isomorphism $\mathrm{Gr}\, L_{-\infty} \cong E^\infty$ is that the maps $E^\infty \twoheadrightarrow K$ and $\mathrm{Gr}\, L_{-\infty} \rightarrowtail K$ are isomorphisms, then the rest of the convergence conditions are automatic. In particular, to deduce convergence to $\mathcal{F}_{L_{-\infty}}$ it suffices to know that $L_\infty = 0$ and $\varprojlim_r{}^1 Z_r = 0$.

## Derived Functors

The left and right derived functors $L_nT, R^nT$ of a functor $T$ provide a measure of the amount by which the functor deviates from preserving exactness.

The category $\mathcal{I}nv$ of inverse systems indexed over $\mathbb{Z}$ (i.e., the category whose objects are diagrams of abelian groups $\cdots \to A_{n-1} \to A_n \to A_{n+1} \to \cdots$) forms an abelian category in which a sequence of morphisms $A' \to A \to A''$ is exact if and only if the sequence $A_n{}' \to A_n \to A_n{}''$ of abelian groups is exact for each $n$. The functor of interest to us is $\varprojlim : \mathcal{I}nv \to \mathcal{AB}$ where $\mathcal{AB}$ denotes the category of abelian groups.

Let $T : \underline{\underline{A}} \to \underline{\underline{B}}$ be an additive functor between abelian categories. Suppose that $X$ in $\mathrm{Obj}\, \underline{\underline{A}}$ has an injective resolution $I_X$. The definition of additive functor implies that $T$ takes zero morphisms to zero morphisms, so $TI_X$ forms a cochain complex in $\underline{\underline{B}}$. The *right derived functors* of $T$ are defined by $(R^nT)(X) := H^n(TI_X)$. The result is independent of the choice of injective resolution (assuming one exists) and satisfies:

1. If $T$ is "left exact" (meaning that $T$ preserves monomorphisms), then $R^0T(X) = T(X)$;
2. If $T$ preserves exactness, then $(R^nT)(X) = 0$ for $n > 0$.

**Theorem 7** *Let $0 \to X' \to X \to X'' \to 0$ be a short exact sequence in $\underline{\underline{A}}$. Suppose $T$ is left exact and that all the objects have injective resolutions. Then there is a (long) exact sequence*

$$0 \to T(X') \to T(X) \to T(X'') \to (R^1T)(X') \to \cdots$$
$$\to (R^{n-1}T)(X'') \to (R^nT)(X') \to (R^nT)(X) \to$$
$$(R^nT)(X'') \to \cdots$$

Similarly, the *left derived functors* of $T$ are defined by using projective resolutions and have similar properties with respect to the obvious duality.

The functor $\varprojlim_n$ is left exact and in the category $\mathcal{I}nv$ every object has an injective resolution. Therefore $\varprojlim_n{}^q$ is defined and $\varprojlim_n{}^0 X_n = \varprojlim_n X_n$, where $\varprojlim_n{}^q$ denotes the derived functor $R^q(\varprojlim_n)$. It turns out that $\varprojlim_n{}^q$ is 0 for $q > 1$, but we are particularly interested in $\varprojlim_n{}^1$.

Let $(X_n)$ be an inverse system with structure maps $i_{n-1} : X_{n-1} \to X_n$. An explicit construction for $\varprojlim_n{}^1 X_n$ is as follows. Define $\phi : \prod_n X_n \to \prod_n X_n$ by letting $\phi(x_n)$ be the sequence whose $n$th component is $(x_n - i_{n-1}x_{n-1})$. Then $\varprojlim_n{}^1 X_n \cong \mathrm{Coker}\,\phi$. Observe that $\mathrm{Ker}\,\phi \cong \varprojlim_n X_n$ according to the explicit formula for $\varprojlim_n X_n$ given earlier.

Recall that we defined ${}^\infty D = \cap_r \mathrm{Im}\, i^{\circ r} \cong \varprojlim_r D^r$. The exact sequence of Theorem 3 can be extended to give:

**Theorem 8** *There is an exact sequence*

$$0 \to D^\infty \xrightarrow{i} D^\infty \xrightarrow{j} E^\infty \xrightarrow{k} {}^\infty D \xrightarrow{i} {}^\infty D$$
$$\xrightarrow{j} \varprojlim_r{}^1 Z^r \xrightarrow{k} \varprojlim_r{}^1 D^r \xrightarrow{i} \varprojlim_r{}^1 D^r \to 0$$

It is clear from the explicit construction that if the system $(X_n)$ stabilizes with $X_n = G$ for all sufficiently small $n$, then $\varprojlim_n X = G$ and $\varprojlim_n{}^1 X = 0$. If the spectral sequence collapses at any stage then the system $(Z^r)$ stabilizes at that point, and so for a spectral sequence which collapses, the condition $\varprojlim_r{}^1 Z^r = 0$, which arose in the discussion of convergence in the previous section, is automatic.

Let $\mathcal{F}_X$ be a filtered abelian group. Applying Theorem 7 to the short exact sequence $0 \to F_nX \to X \to X/F_nX \to 0$ of inverse systems gives an exact sequence

$$0 \to \varprojlim_n F_nX \to \varprojlim_n X \to \varprojlim_n X/F_n$$
$$\to \varprojlim_n{}^1 F_nX \to \varprojlim_n{}^1 X$$

Since $\varprojlim_n X = X$ and $\varprojlim_n{}^1 X = 0$, we get

**Theorem 9** *$\mathcal{F}_X$ is complete if and only if $\varprojlim_n F_nX = 0$ and $\varprojlim_n{}^1 F_nX = 0$.*

When working with $\varprojlim_n{}^1$ the following sufficient condition for its vanishing, known as the Mittag–Leffler condition, is often useful.

**Theorem 10** *Suppose $A$ is an inverse system in which for each $n$ there exists $k(n) \le n$ such that $\mathrm{Im}(A_i \to A_n)$ equals $\mathrm{Im}(A_{k(n)} \to A_n)$ for all $i \le k(n)$. Then $\varprojlim_n{}^1 A = 0$.*

Of course, this will not be (directly) useful in establishing $\varprojlim_n{}^1 F_nX = 0$ since the structure maps in that system are all monomorphisms.

## Some Examples of Standard Spectral Sequences and Their Use

To this point we have considered the general theory of spectral sequences. The properties of the spectral sequences arising in many specific situations have

been well studied. Usually the spectral sequence would be defined either directly, through an exact couple, or by giving some filtration on a chain complex. This defines the $E^1$-term. Typically, a theorem would then be proved giving some formula for the resulting $E^2$-term. In many cases, conditions under which the spectral sequence converges may also be well known.

In this section, we shall take a brief look at the Serre spectral sequence, Atiyah–Hirzebruch spectral sequence, spectral sequence of a double complex, Grothendieck spectral sequence, change of ring spectral sequence, and Eilenberg–Moore spectral sequence, and carry out a few sample calculations.

### Serre Spectral Sequence

Let $F \to X \xrightarrow{\pi} B$ be a fiber bundle (or more generally a fibration) in which the base $B$ is a CW-complex. Define a filtration on the total space by $F_nX := \pi^{-1}B^{(n)}$. This yields a filtration on $H_*(X)$ by setting $F_nH_*(X) := \mathrm{Im}(H_*(F_nX) \to H_*(X))$. The spectral sequence coming from the exact couple in which $D^1_{p,q} := H_{p+q}(F_pX)$ and $E^1_{p,q} := H_{p+q}(F_pX, F_{p-1}X)$ is called the "Serre spectral sequence" of the fibration. Theorems from topology guarantee that this filtration is cocomplete and that $E^1_{p,q} = 0$ if either $p < 0$ or $q < 0$. Therefore, the Serre spectral sequence is always a first quadrant spectral sequence converging to $H_*(X)$.

**Theorem 11** (Serre). *In the Serre spectral sequence of the fibration $F \to E \to B$ there is an isomorphism $E^2_{p,q} \cong H_p(B; {}^tH_q(F))$.*

Here ${}^tH_*(F)$ denotes a "twisted" or "local" coefficient system in which the differential is modified to take into account the action, coming from the fibration, of the fundamental groupoid of the base $B$ on the fiber $F$. In the special case where $B$ is simply connected and $\mathrm{Tor}(H_*(B), H_*(F)) = 0$, the "universal coefficient theorem" says that the $E^2$-term reduces to $E^2_{p,q} \cong H_p(B) \otimes H_q(F)$.

The Serre spectral sequence for cohomology, $E_2^{p,q} \cong H^p(B; {}^tH^q(F)) \Rightarrow H^{p+q}(X)$, has the advantage that it is a spectral sequence of algebras which greatly simplifies calculation of the differentials $d_r$ which are restricted by the requirement that they satisfy the Leibniz rule with respect to the cup product on $H^*(B)$ and $H^*(F)$, and which also allows the computation of the cup product on $H^*(X)$. Since it is a first quadrant spectral sequence, convergence is not an issue.

Frequently in applications of the Serre spectral sequence, instead of using the spectral sequence to calculate $H_*(X)$ from knowledge of $H_*(F)$ and $H_*(B)$ it is instead $H_*(X)$ and one of the other two homologies which is known, and one is working backwards from the spectral sequence to find the homology of the third space.

**Example 3** The universal $S^1$-bundle is the bundle $S^1 \to S^\infty \to \mathbb{C}P^\infty$ where $S^\infty$ is contractible. We will calculate $H^*(\mathbb{C}P^\infty)$ from the Serre spectral sequence of this bundle, taking $H^*(S^1)$ and $H^*(S^\infty)$ as known. We also take as known that $\mathbb{C}P^\infty$ is path connected, so $H^0(\mathbb{C}P^\infty) \cong \mathbb{Z}$.

$$E_2^{p,q} \cong H^p(\mathbb{C}P^\infty) \otimes H^q(S^1) \cong \begin{cases} H^p(\mathbb{C}P^\infty) & \text{if } q = 0 \text{ or } 1 \\ 0 & \text{otherwise} \end{cases}$$

$E_\infty$-terms on the diagonal $p + q = n$ form a composition series for $H^n(S^\infty)$ which is zero for $n \neq 0$. Therefore $E_\infty^{p,q} = 0$ unless $p = 0$ and $q = 0$, with $E_\infty^{0,0} \cong \mathbb{Z}$. Because all nonzero terms lie in the first quadrant, the bidegrees of the differentials show that $d_r(E_2^{1,0}) = 0$ for all $r \geq 2$, so $0 = E_\infty^{1,0} = E_2^{1,0} = H^1(\mathbb{C}P^\infty)$. Since $E_2^{1,q} \cong E_2^{1,0} \otimes E_2^{0,q}$, it follows that $E_2^{1,q} = 0$ for all $q$. Taking into the account the known zero terms, the bidegrees of the differentials show that $E_3^{0,1} \cong \mathrm{Ker}(d_2 : E_2^{0,1} \to E_2^{2,0})$ and $E_\infty^{0,1} = E_3^{0,1}$. Similarly, $E_\infty^{2,0} = E_3^{2,0} \cong \mathrm{Coker}(d_2 : E_2^{0,1} \to E_2^{2,0})$. Therefore, the vanishing of these $E_\infty$-terms shows that $d_2 : E_2^{0,1} \cong E_2^{2,0}$ and in particular $H^2(\mathbb{C}P^\infty) \cong E_2^{0,1} = H^1(S^1) \cong \mathbb{Z}$. It follows that $E_2^{2,q} \cong \mathbb{Z} \otimes E_2^{0,q} \cong E_2^{0,q}$ for all $q$. With the aid of the fact that we showed $E_2^{1,1} = 0$, we can repeat the argument used to show $E_2^{1,q} = 0$ for all $q$ to conclude that $E_2^{3,q} = 0$ for all $q$. Repeating the procedure, we inductively find that $E_2^{p,q} \cong E_2^{p-2,q}$ for all $p > 0$ and all $q$ and in particular

$$H^n(\mathbb{C}P^\infty) \cong \begin{cases} \mathbb{Z} & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases}$$

The cup products in $H^*(\mathbb{C}P^\infty)$ can also be determined by taking advantage of the fact that the spectral sequence is a spectral sequence of algebras. Let $a \in E_2^{2,0} \cong \mathbb{Z}$ be a generator and set $x := d_2 a$. By the preceding calculation, $d_2$ is an isomorphism so $x$ is a generator of $H^2(\mathbb{C}P^\infty)$. Therefore, $x \otimes a$ is a generator of $E_2^{2,2}$ and the isomorphism $d_2$ gives that $d_2(x \otimes a)$ is a generator of $H^4(\mathbb{C}P^\infty)$. However, $d_2(x \otimes a) = d_2(x \otimes 1)(1 \otimes a) = 0 \otimes 1 + (-1)^2(x \otimes 1)d_2 a = x^2 \otimes 1$ and thus, $x^2$ is a generator of $H^4(\mathbb{C}P^\infty)$. Inductively, it follows that $x^n$ is a generator of $H^{2n}(\mathbb{C}P^\infty)$ for all $n$ and so $H^*(\mathbb{C}P^\infty) \cong \mathbb{Z}[x]$.

When working backwards from the Serre or other first quadrant spectral sequences in which $E^2_{p,q} \cong E^2_{p,0} \otimes E^2_{0,q}$ the following analog of the comparison theorem (Theorem 5) is often useful.

**Theorem 12** (Zeeman comparison theorem). *Let $E$ and $E'$ be first quadrant spectral sequences such that $E^2_{p,q} = E^2_{p,0} \otimes E^2_{0,q}$ and $E'^2_{p,q} = E'^2_{p,0} \otimes E'^2_{0,q}$. Let $f : E \to E'$ be a homomorphism of spectral sequences such that $f^2_{p,q} = f^2_{p,0} \otimes f^2_{0,q}$. Suppose that $f^\infty_{p,q} : E^\infty_{p,q} \to E'^\infty_{p,q}$ is an isomorphism for all $p$ and $q$. Then the following are equivalent*:

(i) $f^2_{p,0} : E^2_{p,0} \to E'^2_{p,0}$ *is an isomorphism for* $p \le n-1$;
(ii) $f^2_{0,q} : E^2_{0,q} \to E'^2_{0,q}$ *is an isomorphism for* $q \le n$.

There is a version of the Serre spectral sequence for generalized homology theories coming from the exact couple obtained by applying the generalized homology theory to the Serre filtration of $X$.

**Theorem 13** (Serre spectral sequence for generalized homology). *Let $F \to X \to B$ be a fibration and let $Y$ be an* (*unreduced*) *homology theory satisfying the Milnor wedge axiom. Then there is a* (*right half-plane*) *spectral sequence with $E^2_{p,q} \cong H_p(B; {}^tY_q(F))$ converging to $Y_{p+q}(X)$.*

Cocompleteness of the filtration follows from the properties of generalized homology theories satisfying the wedge axiom (Milnor 1962), and the rest of the convergence conditions are trivial since the filtration is 0 in negative degrees. Here, unlike the Serre spectral sequence for ordinary homology, the existence of terms in the fourth quadrant opens the possibility for composition series of infinite length, although in the case where $B$ is a finite-dimensional complex all the nonzero terms of the spectral sequence will live in the strip between $p=0$ and $p=\dim B$ and so the filtrations will be finite.

The special case of the fibration $* \to X \to X$ yields what is known as the "Atiyah–Hirzebruch spectral sequence".

**Theorem 14** (Atiyah–Hirzebruch spectral sequence). *Let $X$ be a CW-complex and let $Y$ be an* (*unreduced*) *homology theory satisfying the Milnor wedge axiom. Then there is a* (*right half-plane*) *spectral sequence with $E^2_{p,q} \cong H_p(X; Y_q(*))$ converging to $Y_{p+q}(X)$.*

In the cohomology Serre spectral sequence for generalized cohomology (including the cohomology Atiyah–Hirzebruch spectral sequence), convergence of the spectral sequence to $Y^*(X)$ is not guaranteed. Convergence to $\varprojlim_n Y^*(F_nX)$, should that occur, would be of the type discussed in case II in the section "Convergence of graded spectral sequences". Since $X_n = \emptyset$ for $n < 0$, the system defining $L_\infty$ stabilizes to 0. Therefore, $L_\infty = 0$ and, by the discussion in that section, $\varprojlim_r{}^1 Z_rX = 0$ becomes a sufficient condition for convergence to $\varprojlim_n Y^*(F_nX)$. However since the real object of study is usually $Y^*(X)$, the spectral sequence is most useful when one is also able to show $\varprojlim_n{}^1 Y^*(F_nX) = 0$ in which case the Milnor exact sequence (Milnor 1962)

$$\begin{aligned} 0 \to \varprojlim_n{}^1 Y^*(F_nX) \to Y^*(X) \\ \to \varprojlim_n Y^*(F_nX) \to 0 \end{aligned}$$

gives $Y^*(X) \cong \varprojlim_n Y^*(F_nX)$.

If $Y^*(\,)$ has cup products then the spectral sequence has the extra structure of a spectral sequence of $Y^*(*)$-algebras. In the case where $B$ is finite dimensional, all convergence problems disappear since the spectral sequence lives in a strip and the filtrations are finite.

**Example 4** Let $K^*(\,)$ be complex $K$-theory. Since $K^*(*) \cong \mathbb{Z}[z, z^{-1}]$ with $|z| = 2$, in the Atiyah–Hirzebruch spectral sequence for $K^*(\mathbb{C}P^n)$ we have

$$E_2^{p,q} = \begin{cases} \mathbb{Z} & \text{if } q \text{ is even and } p \text{ is even with } 0 \le p \le 2n \\ 0 & \text{otherwise} \end{cases}$$

Because $\mathbb{C}P^n$ is a finite complex, the spectral sequence converges to $K^*(\mathbb{C}P^n)$. Since all the nonzero terms have even total degree and all the differentials have total degree $+1$, the spectral sequence collapses at $E_2$ and we conclude that $K^q(\mathbb{C}P^n) = 0$ if $q$ is odd and that it has a composition series consisting of $(n+1)$ copies of $\mathbb{Z}$ when $q$ is even. Since $\mathbb{Z}$ is a free abelian group, this uniquely identifies the group structure of $K^{\text{even}}(\mathbb{C}P^n)$ as $\mathbb{Z}^{n+1}$. To find the ring structure we can make use of the fact that this is a spectral sequence of $K^*(*)$-algebras. The result is $K^*(\mathbb{C}P^n) \cong K^*(*)[x]/(x^{n+1})$, where $|x| = 2$.

In the Atiyah–Hirzebruch spectral sequence for $K^*(\mathbb{C}P^\infty)$ again all the terms have even total degree so the spectral sequence collapses at $E_2$. We noted earlier that collapse of the spectral sequence implies that $\varprojlim_r{}^1 Z_rX = 0$ and so the spectral sequence convergences to $\varprojlim_n K^*(\mathbb{C}P^n)$, where we used $F_{2n}\mathbb{C}P^\infty = \mathbb{C}P^n$. Since our preceding calculation shows that $K^*(\mathbb{C}P^n) \to K^*(\mathbb{C}P^{n-1})$ is onto, Mittag–Leffler (Theorem 10) implies that $\varprojlim_n{}^1 K^*(\mathbb{C}P^n) = 0$. Therefore, the spectral sequence converges to $K^*(\mathbb{C}P^\infty)$ and we find that $K^*(\mathbb{C}P^\infty) \cong \varprojlim_n K^*(\mathbb{C}P^n)$, which is isomorphic to the power series ring $K^*(*)[[x]]$, where $|x| = 2$.

In topology one might be interested in the Atiyah–Hirzebruch spectral sequence in the case where $X$ is a spectrum rather than a space (a spectrum being a generalization in which cells in negative degrees are allowed including the possibility that the dimensions

of the cells are not bounded below). In such cases, the spectral sequence is no longer constrained to lie in the right half-plane and convergence criteria are not well understood for either the homology or cohomology version.

### Spectral Sequence of a Double Complex

A double complex is a chain complex of chain complexes. That is, it is a bigraded abelian group $C_{p,q}$ together with two differentials $d' : C_{p,q} \to C_{p-1,q}$ and $d'' : C_{p,q} \to C_{p,q-1}$ satisfying $d' \circ d' = 0, d'' \circ d'' = 0$, and $d'd'' = d''d'$. Given a double complex C its total complex Tot C is defined by $(\text{Tot}\, C)_n := \bigoplus_{p+q=n} C_{p,q}$ with differential defined by $d|_{C_{p,q}} := d' + (-1)^p d'' : C_{p,q} \to C_{p-1,q} \oplus C_{p,q-1} \subset \text{Tot}_{n-1} C$.

There are two natural filtrations, $\mathcal{F}'_{\text{TotC}}$ and $\mathcal{F}''_{\text{Tot C}}$, on Tot C given by

$$\left(F'_p(\text{TotC})\right)_n = \bigoplus_{\substack{s+t=n \\ s \leq p}} C_{s,t}$$

$$\left(F''_p(\text{TotC})\right)_n = \bigoplus_{\substack{s+t=n \\ t \leq p}} C_{s,t}$$

yielding two spectral sequences abutting to $H_*(\text{TotC})$. In the first $E'^2_{p,q} = H_p(H_q(C_{*,*}))$ and in the other $E''^2_{p,q} = H_q(H_p(C_{*,*}))$. Convergence of these spectral sequences is not guaranteed, although the first will always converge if there exists $N$ such that $C_{p,q} = 0$ for $p < N$ and the second will converge if there exists $N$ such that $C_{p,q} = 0$ for $q < N$. From the double complex $C$ one could instead form the product total complex $(\text{Tot}^\pi C)_n := \prod_{p+q=n} C_{p,q}$ and proceed in a similar manner to construct the same spectral sequences with different convergence problems. In the important special case of a first quadrant double complex both spectral sequences converge and information is often obtained by playing one off against the other.

**Example 5** Let $M$ and $N$ be $R$-modules. Let $\text{Tor}'^R_*(M,N)$ and $\text{Tor}''^R_*(M,N)$ be the derived functors of $(_) \otimes N$ and $M \otimes (_)$, respectively. Let $P_*$ and $Q_*$ be projective resolutions of $M$ and $N$ respectively. Define a first quadrant double complex by $C_{p,q} := P_p \otimes Q_q$. Since $P_p$ is projective,

$$H_q(C_{p,*}) = P_p \otimes H_q(C_{p,*}) = \begin{cases} 0 & \text{if } q \neq 0 \\ N & \text{if } q = 0 \end{cases}$$

and so in the first spectral sequence of the double complex,

$$E'^2_{p,q} = \begin{cases} 0 & \text{if } q \neq 0 \\ \text{Tor}'^R_p(M,N) & \text{if } q = 0 \end{cases}$$

Therefore, the spectral sequence collapses to give $H_n(\text{Tot}\ C) \cong \text{Tor}'^R_n(M,N)$. Similarly, the second spectral sequence shows that $H_n(\text{Tot}\, C) \cong \text{Tor}''^R_n(M,N)$. Thus, $\text{Tor}^R_*(M,N)$ can be computed equally well from a projective resolution of either variable.

The technique of using a double complex in which one spectral sequence yields the homology the total complex to which both converge can be used to prove.

**Theorem 15** (Grothendieck spectral sequence). *Let $\underline{\underline{C}} \xrightarrow{F} \underline{\underline{B}} \xrightarrow{G} \underline{\underline{A}}$ be a composition of additive functors, where $\underline{\underline{C}}$, $\underline{\underline{B}}$, and $\underline{\underline{A}}$ are abelian categories. Assume that all objects in $\underline{\underline{C}}$ and $\underline{\underline{B}}$ have projective resolutions. Suppose that F takes projectives to projectives. Then for all objects C of $\underline{\underline{C}}$ there exists a (first quadrant) spectral sequence with $E^2_{p,q} = (L_p G)((L_q F)(C))$ converging to $(L_{p+q}(GF))(C)$.*

Naturally, there is a corresponding version for right derived functors.

An application of the Grothendieck spectral sequence is the following "change of rings spectral sequence." Let $f : R \to S$ be a ring homomorphism, let $M$ be a right $S$-module and let $N$ be a left $R$-module. Let $F(A) = S \otimes_R A$ and $G(B) = M \otimes_S B$, and note that $GF(A) = M \otimes_R A$. Applying the Grothendieck spectral sequence to the composition (left $R$-modules $\xrightarrow{F}$ left $S$-modules $\xrightarrow{G}$ abelian groups) yields a convergent spectral sequence $E^2_{p,q} \cong \text{Tor}^S_p(M, \text{Tor}^R_q(S,N)) \Rightarrow \text{Tor}^R_{p+q}(M,N)$.

### Eilenberg–Moore Spectral Sequence

For a topological group $G$, Milnor showed how to construct a universal $G$-bundle $G \to EG \to BG$ in which $EG$ is the infinite join $G^{*\infty}$ with diagonal $G$-action. There is a natural filtration $F_n BG := G^{*(n+1)}/G$ on $BG$ and therefore an induced filtration on the base of any principal $G$-bundle. This filtration yields a spectral sequence including as a special case a tool for calculating $H_*(BG)$ from knowledge of $H_*(G)$.

**Theorem 16** *Let $G \to X \to B$ be a principal $G$-bundle and let $H_*(\,)$ denote homology with coefficients in a field. Then there is a first quadrant spectral sequence with $E^2_{p,q} = \text{Tor}^{H_*(G)}_{pq}(H_*(X), H_*(*))$ converging to $H_{p+q}(BG)$.*

Here the group structure makes $H_*(G)$ into an algebra and $\text{Tor}^A_{pq}(M,N)$ denotes degree $q$ of the graded object formed as the $p$th-derived functor of the tensor product of the graded modules $M$ and $N$ over the graded ring $A$.

There is also a version (Eilenberg and Moore 1962) which, like the Serre spectral sequence, is suitable for computing $H^*(G)$ from $H^*(BG)$.

**Theorem 17** *Let*

$$\begin{array}{ccc} W & \longrightarrow & Y \\ \downarrow & & \downarrow \pi \\ X & \xrightarrow{f} & B \end{array}$$

*be a pullback square in which $\pi$ is a fibration and $X$ and $B$ are simply connected. Suppose that $H^*(X), H^*(Y)$, and $H^*(B)$ are flat $R$-modules of finite type, where $H^*(\,)$ denotes cohomology with coefficients in the Noetherian ring $R$. Then there is a (second quadrant) spectral sequence with $E_2^{p,q} \cong \mathrm{Tor}_{pq}^{H^*(B)}(H^*(X), H^*(Y))$ converging to $H^{p+q}(W)$.*

The cohomological version of the Eilenberg–Moore spectral sequence, stated above, contains the more familiar Tor for modules over an algebra. For the homological version, one must dualize these notions appropriately to define the cotensor product of comodules over a coalgebra, and its derived functors Cotor.

Provided the action of the fundamental group of $B$ is sufficiently nice there are extensions of the Eilenberg–Moore spectral sequence to the case where $B$ is not simply connected, although they do not always converge, and extensions to generalized (co)homology theories have also been studied.

*See also:* Cohomology Theories; Derived Categories; K-Theory; Spectral Theory for Linear Operators.

## Further Reading


Adams JF (1974) *Stable Homotopy and Generalized Homology*. Chicago: Chicago University Press.

Atiyah M and Hirzebruch F (1969) Vector bundles and homogeneous spaces. *Proceedings of Symposia in Pure Mathematics* 3: 7–38.

Boardman JM (1999) Conditionally convergent spectral sequences. *Contemporary Mathematics* 239: 49–84.

Bousfield AK and Kan DM (1972) *Homotopy Limits, Completions and Localization*. Lecture Notes in Math, vol. 304. Berlin: Springer.

Cartan H and Eilenberg S (1956) *Homological Algebra*. Princeton: Princeton University Press.

Eilenberg S and Moore J (1962) Homology and fibrations I. Coalgebras cotensor product and its derived functors. *Commentarii Mathematici Helvetici* 40: 199–236.

Grothendieck A (1957) Sur quelques points d'algèbre homologique. *Tôhoku Mathematical Journal* 9: 119–221.

Hilton PJ and Stammbach U (1970) *A Course in Homological Algebra*. Graduate Texts in Mathematics, vol. 4. Berlin: Springer.

Koszul J-L (1947) Sur l'opérateurs de derivation dans un anneau. *Comptes Rendus de l'Académie des Sciences de Paris* 225: 217–219.

Leray J (1946) L'anneau d'une representation; Propriétés d'homologie de la projections d'un espace fibré sur sa base; Sur l'anneau d'homologie de l'espace homogène, quotient d'un groupe clos par un sous-groupe abélien, connexe, maximum. *C.R. Acad. Sci.* 222: 1366–1368; 1419–1422 and 223: 412–415.

Massey W (1952) Exact couples in algebraic topology I, II. *Annals of Mathematics* 56: 363–396.

Massey W (1953) Exact couples in algebraic topology III, IV, V. *Annals of Mathematics* 57: 248–286.

McCleary J (2001) *User's Guide to Spectral Sequences*. Cambridge Studies in Advanced Mathematics, vol. 58. Cambridge: Cambridge University Press.

Milnor J (1956) Constructions of universal bundles, II. *Annals of Mathematics* 63: 430–436.

Milnor J (1962) On axiomatic homology theory. *Pacific Journal of Mathematics* 12: 337–341.

Selick P (1997) *Introduction to Homotopy Theory*. Fields Institute Monographs, vol. 9. Providence, RI: American Mathematical Society.

Serre J-P (1951) Homologie singulière des espaces fibrés. *Annals of Mathematics* 54: 425–505.

Smith L (1969) *Lectures on the Eilenberg–Moore Spectral Sequence*. Lecture Notes in Math, vol. 134. Berlin: Springer.

Zeeman EC (1958) A proof of the comparison theorem for spectral sequences. *Proceedings of the Cambridge Philosophical Society* 54: 57–62.


# Spectral Theory of Linear Operators

**M Schechter**, University of California at Irvine, Irvine, CA, USA



## Introduction

We begin with the study of linear operators on normed vector spaces (for definitions, see, e.g., Schechter (2002) or the appendix at the end of this article). If the scalars are complex numbers, we shall call the space complex. If the scalars are real, we shall call it real.

Let $X, Y$ be normed vector spaces. A mapping $A$ which assigns to each element $x$ of a set $D(A) \subset X$ a unique element $y \in Y$ is called an operator (or transformation). The set $D(A)$ on which $A$ acts is called the domain of $A$. The operator $A$ is called linear if

1. $D(A)$ is a subspace of $X$, and
2. $A(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 A x_1 + \alpha_2 A x_2$

for all scalars $\alpha_1, \alpha_2$ and all elements $x_1, x_2 \in D(A)$.

To begin, we shall only consider operators $A$ with $D(A) = X$.

An operator $A$ is called bounded if there is a constant $M$ such that

$$\|Ax\| \le M\|x\|, \quad x \in X \qquad [1]$$

The norm of such an operator is defined by

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} \qquad [2]$$

It is the smallest $M$ which works in [1]. An operator $A$ is called continuous at a point $x \in X$ if $x_n \to x$ in $X$ implies $Ax_n \to Ax$ in $Y$. A bounded linear operator is continuous at each point. For if $x_n \to x$ in $X$, then

$$\|Ax_n - Ax\| \le \|A\| \cdot \|x_n - x\| \to 0$$

We also have

**Theorem 1** *If a linear operator $A$ is continuous at one point $x_0 \in X$, then it is bounded, and hence continuous at every point.*

We let $B(X, Y)$ be the set of bounded linear operators from $X$ to $Y$. Under the norm [2], one easily checks that $B(X, Y)$ is a normed vector space.

### The Adjoint Operator

An assignment $F$ of a number to each element $x$ of a vector space is called a functional and denoted by $F(x)$. If it satisfies

$$F(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 F(x_1) + \alpha_2 F(x_2) \qquad [3]$$

for $\alpha_1, \alpha_2$ scalars, it is called linear. It is called bounded if

$$|F(x)| \le M\|x\|, \quad x \in X \qquad [4]$$

If $F$ is a bounded linear functional on a normed vector space $X$, the norm of $F$ is defined by

$$\|F\| = \sup_{x \in X,\, x \neq 0} \frac{|F(x)|}{\|x\|} \qquad [5]$$

It is equal to the smallest number $M$ satisfying [4].

For any normed vector space $X$, let $X'$ denote the set of bounded linear functionals on $X$. If $f, g \in X'$, we say that $f = g$ if

$$f(x) = g(x) \text{ for all } x \in X$$

The "zero" functional is the one assigning zero to all $x \in X$. We define $h = f + g$ by

$$h(x) = f(x) + g(x), \quad x \in X$$

and $g = \alpha f$ by

$$g(x) = \alpha f(x), \quad x \in X$$

Under these definitions, $X'$ becomes a vector space. The expression

$$\|f\| = \sup_{x \neq 0} \frac{|f(x)|}{\|x\|}, \quad f \in X' \qquad [6]$$

is easily seen to be a norm. Thus, $X'$ is a normed vector space. It is therefore natural to ask when $X'$ will be complete. A rather surprising answer is given by

**Theorem 2** *$X'$ is a Banach space whether or not $X$ is.*

(For the definition of a Banach space, see, e.g., Schechter (2002) or the appendix at the end of this article.)

Suppose $X, Y$ are normed vector spaces and $A \in B(X, Y)$. For each $y' \in Y'$, the expression $y'(Ax)$ assigns a scalar to each $x \in X$. Thus, it is a functional $F(x)$. Clearly $F$ is linear. It is also bounded since

$$|F(x)| = |y'(Ax)| \le \|y'\| \cdot \|Ax\| \le \|y'\| \cdot \|A\| \cdot \|x\|$$

Thus, there is an $x' \in X'$ such that

$$y'(Ax) = x'(x), \quad x \in X \qquad [7]$$

This functional $x'$ is unique. Thus, to each $y' \in Y'$ we have assigned a unique $x' \in X'$. We designate this assignment by $A'$ and note that it is a linear operator from $Y'$ to $X'$. Thus, [7] can be written in the form

$$y'(Ax) = A'y'(x) \qquad [8]$$

The operator $A'$ is called the adjoint (or conjugate) of $A$. We note

**Theorem 3** *$A' \in B(Y', X')$, and $\|A'\| = \|A\|$.*

The adjoint has the following easily verified properties:

$$(A + B)' = A' + B' \qquad [9]$$

$$(\alpha A)' = \alpha A' \qquad [10]$$

$$(AB)' = B'A' \qquad [11]$$

Why should we consider adjoints? One reason is as follows. Many problems in mathematics and its applications can be put in the form: given normed vector spaces $X, Y$ and an operator $A \in B(X, Y)$, one wishes to solve

$$Ax = y \qquad [12]$$

The set of all $y$ for which one can solve [12] is called the "range" of $A$ and is denoted by $R(A)$. The set of all $x$ for which $Ax = 0$ is called the "null space" of $A$ and is denoted by $N(A)$. Since $A$ is linear, it is easily checked that $N(A)$ and $R(A)$ are subspaces of $X$ and $Y$,

respectively (for definitions, see, e.g., Schechter (2002) or the appendix at the end of this article). The dimension of $N(A)$ is denoted by $\alpha(A)$.

If $y \in R(A)$, there is an $x \in X$ satisfying [12]. For any $y' \in Y'$ we have

$$y'(Ax) = y'(y)$$

Taking adjoints we get

$$A'y'(x) = y'(y)$$

If $y' \in N(A')$, this gives $y'(y) = 0$. Thus, a necessary condition that $y \in R(A)$ is that $y'(y) = 0$ for all $y' \in N(A')$. Obviously, it would be of great interest to know when this condition is also sufficient.

## The Spectrum and Resolvent Sets

From this point henceforth we shall assume that $X = Y$. We can then speak of the identity operator $I$ defined by

$$Ix = x, \quad x \in X$$

For a scalar $\lambda$, the operator $\lambda I$ is given by

$$\lambda Ix = \lambda x, \quad x \in X$$

We shall denote the operator $\lambda I$ by $\lambda$.

We shall denote the space $B(X, X)$ by $B(X)$. For any operator $A \in B(X)$, a scalar $\lambda$ for which $\alpha(A - \lambda) \neq 0$ is called an eigenvalue of $A$. Any element $x \neq 0$ of $X$ such that $(A - \lambda)x = 0$ is called an eigenvector (or eigenelement). The points $\lambda$ for which $(A - \lambda)$ has a bounded inverse in $B(X)$ comprise the resolvent set $\rho(A)$ of $A$ (for definitions, see, e.g., Schechter (2002) or the appendix at the end of this article). If $X$ is a Banach space, it is the set of those $\lambda$ such that $\alpha(A - \lambda) = 0$ and $R(A - \lambda) = X$. The spectrum $\sigma(A)$ of $A$ consists of all scalars not in $\rho(A)$. The set of eigenvalues of $A$ is sometimes called the point spectrum of $A$ and is denoted by $P\sigma(A)$.

We note that

**Theorem 4** *For $A$ in $B(X)$, $\sigma(A') = \sigma(A)$.*

We are now going to examine the sets $\rho(A)$ and $\sigma(A)$ for arbitrary $A \in B(X)$.

**Theorem 5** *$\rho(A)$ is an open set and hence $\sigma(A)$ is a closed set.*

Does every operator $A \in B(X)$ have points in its resolvent set? Yes. In fact, we have

**Theorem 6** *For $A$ in $B(X)$, set*

$$r_\sigma(A) = \inf_n \|A^n\|^{1/n} \qquad [13]$$

*Then $\rho(A)$ contains all scalars $\lambda$ such that $|\lambda| > r_\sigma(A)$.*

Let $p(t)$ be a polynomial of the form

$$p(t) = \sum_0^n a_k t^k$$

Then for any operator $A \in B(X)$, we define the operator

$$p(A) = \sum_0^n a_k A^k$$

where we take $A^0 = I$. We have

**Theorem 7** *If $\lambda \in \sigma(A)$, then $p(\lambda) \in \sigma(p(A))$ for any polynomial $p(t)$.*

Proof Since $\lambda$ is a root of $p(t) - p(\lambda)$, we have

$$p(t) - p(\lambda) = (t - \lambda)q(t)$$

where $q(t)$ is a polynomial with real coefficients. Hence,

$$p(A) - p(\lambda) = (A - \lambda)q(A) = q(A)(A - \lambda) \qquad [14]$$

Now, if $p(\lambda)$ is in $\rho(p(A))$, then [14] shows that $\alpha(A - \lambda) = 0$ and $R(A - \lambda) = X$. This means that $\lambda \in \rho(A)$, and the theorem is proved. □

A symbolic way of writing Theorem 7 is

$$p(\sigma(A)) \subset \sigma(p(A)) \qquad [15]$$

Note that, in general, there may be points in $\sigma(p(A))$ which may not be of the form $p(\lambda)$ for some $\lambda \in \sigma(A)$. As an example, consider the operator on $\mathbb{R}^2$ given by

$$A(\alpha_1, \alpha_2) = (-\alpha_2, \alpha_1)$$

$A$ has no spectrum; $A - \lambda$ is invertible for all real $\lambda$. However, $A^2$ has –1 as an eigenvalue. What is the reason for this? It is simply that our scalars are real. Consequently, imaginary numbers cannot be considered as eigenvalues. We shall see later that in order to obtain a more complete theory, we shall have to consider complex Banach spaces. Another question is whether every operator $A \in B(X)$ has points in its spectrum. For complex Banach spaces, the answer is yes.

## The Spectral Mapping Theorem

Suppose we want to solve an equation of the form

$$p(A)x = y, \quad x, \ y \in X \qquad [16]$$

where $p(t)$ is a polynomial and $A \in B(X)$. If 0 is not in the spectrum of $p(A)$, then $p(A)$ has an inverse in $B(X)$ and, hence, [16] can be solved for all $y \in X$. So a natural question to ask is: what is the spectrum of $p(A)$? By Theorem 7 we see that it contains $p(\sigma(A))$,

but by the remark at the end of the preceding section it can contain other points. If it were true that

$$p(\sigma(A)) = \sigma(p(A)) \qquad [17]$$

then we could say that [16] can be solved uniquely for all $y \in X$ if and only if $p(\lambda) \neq 0$ for all $\lambda \in \sigma(A)$.

For a complex Banach space we have

**Theorem 8** *If X is a complex Banach space, then* $\mu \in \sigma(p(A))$ *if and only if* $\mu = p(\lambda)$ *for some* $\lambda \in \sigma(A)$, *that is, if [17] holds.*

Proof We have proved it in one direction already (Theorem 7). To prove it in the other, let $\gamma_1, \ldots, \gamma_n$ be the (complex) roots of $p(t) - \mu$. For a complex Banach space they are all scalars. Thus,

$$p(A) - \mu = c(A - \gamma_1) \cdots (A - \gamma_n), \quad c \neq 0$$

Now suppose that all of the $\gamma_j$ are in $\rho(A)$. Then each $A - \gamma_j$ has an inverse in $B(X)$. Hence, the same is true for $p(A) - \mu$. In other words, $\mu \in \rho(p(A))$. Thus, if $\mu \in \sigma(p(A))$, then at least one of the $\gamma_j$ must be in $\sigma(A)$, say $\gamma_k$. Hence, $\mu = p(\gamma_k)$, where $\gamma_k \in \sigma(A)$. This completes the proof. □

Theorem 8 is called the "spectral mapping theorem" for polynomials. As mentioned before, it has the useful consequence:

**Corollary 1** *If X is a complex Banach space, then eqn [16] has a unique solution for every y in X if and only if* $p(\lambda) \neq 0$ *for all* $\lambda \in \sigma(A)$.

## Operational Calculus

Other things can be done in a complex Banach space that cannot be done in a real Banach space. For instance, we can get a formula for $p(A)^{-1}$ when it exists. To obtain this formula, we first note

**Theorem 9** *If X is a complex Banach space, then* $(z - A)^{-1}$ *is a complex analytic function of z for* $z \in \rho(A)$.

By this, we mean that in a neighborhood of each $z_0 \in \rho(A)$, the operator $(z - A)^{-1}$ can be expanded in a "Taylor series," which converges in norm to $(z - A)^{-1}$, just like analytic functions of a complex variable.

Now, by Theorem 6, $\rho(A)$ contains the set $|z| > \|A\|$. We can expand $(z - A)^{-1}$ in powers of $z^{-1}$ on this set. In fact, we have

**Lemma 1** *If* $|z| > \limsup \|A^n\|^{1/n}$, *then*

$$(z - A)^{-1} = \sum_1^\infty z^{-n} A^{n-1} \qquad [18]$$

*where the convergence is in the norm of* $B(X)$.

Let $C$ be any circle with center at the origin and radius greater than, say, $\|A\|$. Then, by Lemma 1,

$$\oint_C z^n (z - A)^{-1} dz = \sum_{k=1}^\infty A^{k-1} \oint_C z^{n-k} dz = 2\pi i A^n \qquad [19]$$

or

$$A^n = \frac{1}{2\pi i} \oint_C z^n (z - A)^{-1} dz \qquad [20]$$

where the line integral is taken in the right direction.

Note that the line integrals are defined in the same way as is done in the theory of functions of a complex variable. The existence of the integrals and their independence of path (so long as the integrands remain analytic) are proved in the same way. Since $(z - A)^{-1}$ is analytic on $\rho(A)$, we have

**Theorem 10** *Let C be any closed curve containing* $\sigma(A)$ *in its interior. Then [20] holds.*

As a direct consequence of this, we have

**Theorem 11** $r_\sigma(A) = \max_{\lambda \in \sigma(A)} |\lambda|$ *and* $\|A^n\|^{1/n} \to r_\sigma(A)$ *as* $n \to \infty$.

We can now put Lemma 1 in the following form:

**Theorem 12** *If* $|z| > r_\sigma(A)$, *then [18] holds with convergence in* $B(X)$.

Now let $b$ be any number greater than $r_\sigma(A)$, and let $f(z)$ be a complex-valued function that is analytic in $|z| < b$. Thus,

$$f(z) = \sum_0^\infty a_k z^k, \quad |z| < b \qquad [21]$$

We can define $f(A)$ as follows: the operators

$$\sum_0^n a_k A^k$$

converge in norm, since

$$\sum_0^\infty |a_k| \cdot \|A^k\| < \infty$$

This last statement follows from the fact that if $c$ is any number satisfying $r_\sigma(A) < c < b$, then

$$\|A^k\|^{1/k} \leq c$$

for $k$ sufficiently large, and the series

$$\sum_0^\infty |a_k| c^k$$

is convergent. We define $f(A)$ to be

$$\sum_{0}^{\infty} a_k A^k \qquad [22]$$

By Theorem 10, this gives

$$\begin{aligned} f(A) &= \frac{1}{2\pi i} \sum_{0}^{\infty} a_k \oint_C z^k (z - A)^{-1} \, dz \\ &= \frac{1}{2\pi i} \oint_C \sum_{0}^{\infty} a_k z^k (z - A)^{-1} \, dz \\ &= \frac{1}{2\pi i} \oint_C f(z)(z - A)^{-1} \, dz \end{aligned} \qquad [23]$$

where $C$ is any circle about the origin with radius greater than $r_\sigma(A)$ and less than $b$.

We can now give the formula that we promised. Suppose $f(z)$ does not vanish for $|z| < b$. Set $g(z) = 1/f(z)$. Then $g(z)$ is analytic in $|z| < b$, and hence $g(A)$ is defined. Moreover,

$$\begin{aligned} f(A)g(A) &= \frac{1}{2\pi i} \oint_C f(z)g(z)(z - A)^{-1} \, dz \\ &= \frac{1}{2\pi i} \oint_C (z - A)^{-1} \, dz = I \end{aligned}$$

Since $f(A)$ and $g(A)$ clearly commute, we see that $f(A)^{-1}$ exists and equals $g(A)$. Hence,

$$f(A)^{-1} = \frac{1}{2\pi i} \oint_C \frac{1}{f(z)} (z - A)^{-1} \, dz \qquad [24]$$

In particular, if

$$g(z) = 1/f(z) = \sum_{0}^{\infty} c_k z^k, \quad |z| < b$$

then

$$f(A)^{-1} = \sum_{0}^{\infty} c_k A^k \qquad [25]$$

Now, suppose $f(z)$ is analytic in an open set $\Omega$ containing $\sigma(A)$, but not analytic in a disk of radius greater than $r_\sigma(A)$. In this case, we cannot say that the series [22] converges in norm to an operator in $B(X)$. However, we can still define $f(A)$ in the following way: there exists an open set $\omega$ whose closure $\bar{\omega} \subset \Omega$ and whose boundary $\partial\omega$ consists of a finite number of simple closed curves that do not intersect, and such that $\sigma(A) \subset \omega$. (That such a set always exists is left as an exercise; see, e.g., Schechter (2002).) We now define $f(A)$ by

$$f(A) = \frac{1}{2\pi i} \oint_{\partial\omega} f(z)(z - A)^{-1} \, dz \qquad [26]$$

where the line integrals are to be taken in the proper directions. It is easily checked that $f(A) \in B(X)$ and is independent of the choice of the set $\omega$. By [23], this definition agrees with the one given above for the case when $\Omega$ contains a disk of radius greater than $r_\sigma(A)$. Note that if $\Omega$ is not connected, $f(z)$ need not be the same function on different components of $\Omega$.

Now suppose $f(z)$ does not vanish on $\sigma(A)$. Then we can choose $\omega$ so that $f(z)$ does not vanish on $\bar{\omega}$ (this is also an exercise). Thus, $g(z) = 1/f(z)$ is analytic on an open set containing $\bar{\omega}$ so that $g(A)$ is defined. Since $f(z)g(z) = 1$, one would expect that $f(A)g(A) = g(A)f(A) = I$, in which case, it would follow that $f(A)^{-1}$ exists and is equal to $g(A)$. This follows from

**Lemma 2** *If $f(z)$ and $g(z)$ are analytic in an open set $\Omega$ containing $\sigma(A)$ and*

$$h(z) = f(z)g(z)$$

*then $h(A) = f(A)g(A)$.*

Therefore, it follows that we have

**Theorem 13** *If $A$ is in $B(X)$ and $f(z)$ is a function analytic in an open set $\Omega$ containing $\sigma(A)$ such that $f(z) \neq 0$ on $\sigma(A)$, then $f(A)^{-1}$ exists and is given by*

$$f(A)^{-1} = \frac{1}{2\pi i} \oint_{\partial\omega} \frac{1}{f(z)} (z - A)^{-1} \, dz$$

*where $\omega$ is any open set such that*

(i) $\sigma(A) \subset \omega$, $\bar{\omega} \subset \Omega$,
(ii) *$\partial\omega$ consists of a finite number of simple closed curves, and*
(iii) *$f(z) \neq 0$ on $\bar{\omega}$.*

Now that we have defined $f(A)$ for functions analytic in a neighborhood of $\sigma(A)$, we can show that the spectral mapping theorem holds for such functions as well (see Theorem 8). We have

**Theorem 14** *If $f(z)$ is analytic in a neighborhood of $\sigma(A)$, then*

$$\sigma(f(A)) = f(\sigma(A)) \qquad [27]$$

*that is, $\mu \in \sigma(f(A))$ if and only if $\mu = f(\lambda)$ for some $\lambda \in \sigma(A)$.*

## Complexification

What we have just done is valid for complex Banach spaces. Suppose, however, we are dealing with a real Banach space. What can be said then?

Let $X$ be a real Banach space. Consider the set $Z$ of all ordered pairs $\langle x, y\rangle$ of elements of $X$. We set

$$\langle x_1, y_1\rangle + \langle x_2, y_2\rangle = \langle x_1 + x_2,\ y_1 + y_2\rangle$$
$$(\alpha + \mathrm{i}\beta)\langle x, y\rangle = \langle(\alpha x - \beta y), (\beta x + \alpha y)\rangle$$
$$\alpha, \beta \in \mathbb{R}$$

With these definitions, one checks easily that $Z$ is a complex vector space. The set of elements of $Z$ of the form $\langle x, 0\rangle$ can be identified with $X$. We would like to introduce a norm on $Z$ that would make $Z$ into a Banach space and satisfy

$$\|\langle x, 0\rangle\| = \|x\|, \quad x \in X$$

An obvious suggestion is

$$(\|x\|^2 + \|y\|^2)^{1/2}$$

However, it is soon discovered that this is not a norm on $Z$ (why?). We have to be more careful. One that works is given by

$$\|\langle x, y\rangle\| = \max_{\alpha^2+\beta^2=1} (\|\alpha x - \beta y\|^2 + \|\beta x + \alpha y\|^2)^{1/2}$$

With this norm, $Z$ becomes a complex Banach space having the desired properties.

Now let $A$ be an operator in $B(X)$. We define an operator $\hat{A}$ in $B(Z)$ by

$$\hat{A}\langle x, y\rangle = \langle Ax, Ay\rangle$$

Then

$$\begin{aligned}
&\|\hat{A}\langle x, y\rangle\| \\
&\quad = \max_{\alpha^2+\beta^2=1} (\|\alpha Ax - \beta Ay\|^2 + \|\beta Ax + \alpha Ay\|^2)^{1/2} \\
&\quad = \max_{\alpha^2+\beta^2=1} (\|A(\alpha x - \beta y)\|^2 + \|A(\beta x + \alpha y)\|^2)^{1/2} \\
&\quad \le \|A\| \cdot \|\langle x, y\rangle\|
\end{aligned}$$

Thus,

$$\|\hat{A}\| \le \|A\|$$

But,

$$\|\hat{A}\| \ge \sup_{x \neq 0} \frac{\|\langle Ax, 0\rangle\|}{\|\langle x, 0\rangle\|} = \|A\|$$

Hence,

$$\|\hat{A}\| = \|A\|$$

If $\lambda$ is real, then

$$(\hat{A} - \lambda)\langle x, y\rangle = \langle(A - \lambda)x, (A - \lambda)y\rangle$$

This shows that $\lambda \in \rho(\hat{A})$ if and only if $\lambda \in \rho(A)$. Similarly, if $p(t)$ is a polynomial with real coefficients, then

$$p(\hat{A})\langle x, y\rangle = \langle p(A)x, p(A)y\rangle$$

showing that $p(\hat{A})$ has an inverse in $B(Z)$ if and only if $p(A)$ has an inverse in $B(X)$. Hence, we have

**Theorem 15** *Equation [16] has a unique solution for each $y$ in $X$ if and only if $p(\lambda) \neq 0$ for all $\lambda \in \sigma(\hat{A})$.*

In the example given earlier, the operator $\hat{A}$ has eigenvalues i and –i. Hence, –1 is in the spectrum of $\hat{A}^2$ and also in that of $A^2$. Thus, the equation

$$(A^2 + 1)x = y$$

cannot be solved uniquely for all $y$.

## Compact Operators

Let $X$, $Y$ be normed vector spaces. A linear operator $K$ from $X$ to $Y$ is called compact (or completely continuous) if $D(K) = X$ and for every sequence $\{x_n\} \subset X$ such that $\|x_n\| \le C$, the sequence $\{Kx_n\}$ has a subsequence which converges in $Y$. The set of all compact operators from $X$ to $Y$ is denoted by $K(X, Y)$.

A compact operator is bounded. Otherwise, there would be a sequence $\{x_n\}$ such that $\|x_n\| \le C$, while $\|Kx_n\| \to \infty$. Then $\{Kx_n\}$ could not have a convergent subsequence. The sum of two compact operators is compact, and the same is true of the product of a scalar and a compact operator. Hence, $K(X, Y)$ is a subspace of $B(X, Y)$.

If $A \in B(X, Y)$ and $K \in K(Y, Z)$, then $KA \in K(X, Z)$. Similarly, if $L \in K(X, Y)$ and $B \in B(Y, Z)$, then $BL \in K(X, Z)$.

Suppose $K \in B(X, Y)$, and there is a sequence $\{F_n\}$ of compact operators such that

$$\|K - F_n\| \longrightarrow 0 \text{ as } n \longrightarrow \infty \qquad [28]$$

We claim that if $Y$ is a Banach space, then $K$ is compact.

**Theorem 16** *Let $X$ be a normed vector space and $Y$ a Banach space. If $L$ is in $B(X, Y)$ and there is a sequence $\{K_n\} \subset K(X, Y)$ such that*

$$\|L - K_n\| \longrightarrow 0 \text{ as } n \longrightarrow 0$$

*then $L$ is in $K(X, Y)$.*

**Theorem 17** *Let $X$ be a Banach space and let $K$ be an operator in $K(X)$. Set $A = I - K$. Then, $R(A)$ is closed in $X$ and $\dim N(A) = \dim N(A')$ is finite.*

*In particular, either* $R(A) = X$ *and* $N(A) = \{0\}$, *or* $R(A) \neq X$ *and* $N(A) \neq \{0\}$.

The last statement of Theorem 17 is known as the "Fredholm alternative."

Let $X, Y$ be Banach spaces. An operator $A \in B(X, Y)$ is said to be a Fredholm operator from $X$ to $Y$ if

1. $\alpha(A) = \dim N(A)$ is finite,
2. $R(A)$ is closed in $Y$, and
3. $\beta(A) = \dim N(A')$ is finite.

The set of Fredholm operators from $X$ to $Y$ is denoted by $\Phi(X, Y)$. If $X = Y$ and $K \in K(X)$, then, clearly, $I - K$ is a Fredholm operator. The index of a Fredholm operator is defined as

$$i(A) = \alpha(A) - \beta(A) \qquad [29]$$

For $K \in K(X)$, we have shown that $i(I - K) = 0$ (Theorem 17).

**Theorem 18** *Let* $X$, $Y$ *be normed vector spaces, and assume that* $K$ *is in* $K(X, Y)$. *Then* $K'$ *is in* $K(Y', X')$.

Let $X$ be a Banach space, and suppose $K \in K(X)$. If $\lambda$ is a nonzero scalar, then

$$\lambda I - K = \lambda(I - \lambda^{-1}K) \in \Phi(X) \qquad [30]$$

For an arbitrary operator $A \in B(X)$, the set of all scalars $\lambda$ for which $\lambda I - A \in \Phi(X)$ is called the $\Phi$-set of $A$ and is denoted by $\Phi_A$. Thus, [30] gives

**Theorem 19** *If* $X$ *is a Banach space and* $K$ *is in* $K(X)$, *then* $\Phi_K$ *contains all scalars* $\lambda \neq 0$.

**Theorem 20** *Under the hypothesis of Theorem 19*, $\alpha(K - \lambda) = 0$ *except for, at most, a denumerable set* $S$ *of values of* $\lambda$. *The set* $S$ *depends on* $K$ *and has 0 as its only possible limit point. Moreover, if* $\lambda \neq 0$ *and* $\lambda \notin S$, *then* $\alpha(K - \lambda) = 0, R(K - \lambda) = X$ *and* $K - \lambda$ *has an inverse in* $B(X)$.

## Unbounded Operators

In many applications, one runs into unbounded operators instead of bounded ones. This is particularly true in the case of differential equations. For instance, consider the operator $\mathrm{d}/\mathrm{d}t$ on $C[0, 1]$ with domain consisting of continuously differentiable functions. It is clearly unbounded. In fact, the sequence $x_n(t) = t^n$ satisfies $\|x_n\| = 1, \|\mathrm{d}x_n/\mathrm{d}t\| = n \to \infty$ as $n \to \infty$. It would, therefore, be useful if some of the results that we have stated for bounded operators would also hold for unbounded ones. We shall see that, indeed, many of them do. Unless otherwise specified, $X, Y, Z$, and $W$ will denote Banach spaces in this article.

Let $X$, $Y$ be normed vector spaces, and let $A$ be a linear operator from $X$ to $Y$. We now officially lift our restriction that $D(A) = X$. However, if $A \in B(X, Y)$, it is still to be assumed that $D(A) = X$.

The operator $A$ is called closed if whenever $\{x_n\} \subset D(A)$ is a sequence satisfying

$$x_n \longrightarrow x \text{ in } X, \qquad Ax_n \longrightarrow y \text{ in } Y \qquad [31]$$

then $x \in D(A)$ and $Ax = y$. Clearly, all operators in $B(X, Y)$ are closed.

To define $A'$ for an unbounded operator, we follow the definition for bounded operators, and exercise a bit of care. We want

$$A'y'(x) = y'(Ax), \quad x \in D(A) \qquad [32]$$

Thus, we say that $y' \in D(A')$ if there is an $x' \in X'$ such that

$$x'(x) = y'(Ax), \quad x \in D(A) \qquad [33]$$

Then we define $A'y'$ to be $x'$. In order that this definition make sense, we need $x'$ to be unique, that is, that $x'(x) = 0$ for all $x \in D(A)$ should imply that $x' = 0$. This is true if and only if $D(A)$ is dense in $X$. To summarize, we can define $A'$ for any linear operator from $X$ to $Y$ provided $D(A)$ is dense in $X$. We take $D(A')$ to be the set of those $y' \in Y'$ for which there is an $x' \in X'$ satisfying [33]. This $x'$ is unique, and we set $A'y' = x'$. Note that if

$$|y'(Ax)| \leq C\|x\|, \quad x \in D(A)$$

then a simple application of the Hahn–Banach theorem (see e.g., Schechter (2002) or the appendix) shows that $y' \in D(A')$.

We define unbounded Fredholm operators in the following way: let $X$, $Y$ be Banach spaces. Then the set $\Phi(X, Y)$ of Fredholm operators from $X$ to $Y$ consists of linear operators from $X$ to $Y$ such that

1. $D(A)$ is dense in $X$,
2. $A$ is closed,
3. $\alpha(A) = \dim N(A) < \infty$,
4. $R(A)$ is closed in $Y$, and
5. $\beta(A) = \dim N(A') < \infty$.

## The Essential Spectrum

Let $A$ be a linear operator on a normed vector space $X$. We say that $\lambda \in \rho(A)$ if $R(A - \lambda)$ is dense in $X$ and there is a $T \in B(X)$ such that

$$\begin{aligned} T(A - \lambda) &= I \text{ on } D(A) \\ (A - \lambda)T &= I \text{ on } R(A - \lambda) \end{aligned} \qquad [34]$$

Otherwise, $\lambda \in \sigma(A)$. As before, $\rho(A)$ and $\sigma(A)$ are called the resolvent set and spectrum of $A$, respectively. To show the relationship of this definition to the one given before, we note the following.

**Lemma 3** *If $X$ is a Banach space and $A$ is closed, then $\lambda \in \rho(A)$ if and only if*

$$\alpha(A - \lambda) = 0, \qquad R(A - \lambda) = X \qquad [35]$$

Throughout the remainder of this section, we shall assume that $X$ is a Banach space, and that $A$ is a densely defined, closed linear operator on $X$. We ask the following question: what points of $\sigma(A)$ can be removed from the spectrum by the addition of a compact operator to $A$? The answer to this question is closely related to the set $\Phi_A$. We define this to be the set of all scalars $\lambda$ such that $A - \lambda \in \Phi(X)$. We have

**Theorem 21** *The set $\Phi_A$ is open, and $i(A - \lambda)$ is constant on each of its components.*

We also have

**Theorem 22** *$\Phi_{A+K} = \Phi_A$ for all $K$ which are $A$-compact, and $i(A + K - \lambda) = i(A - \lambda)$ for all $\lambda \in \Phi_A$.*

Set

$$\sigma_e(A) = \bigcap_{K \in K(X)} \sigma(A + K)$$

We call $\sigma_e(A)$ the essential spectrum of $A$ (there are other definitions). It consists of those points of $\sigma(A)$ which cannot be removed from the spectrum by the addition of a compact operator to $A$. We now characterize $\sigma_e(A)$.

**Theorem 23** *$\lambda \notin \sigma_e(A)$ if and only if $\lambda \in \Phi_A$ and $i(A - \lambda) = 0$.*

## Normal Operators

A sequence of elements $\{\varphi_n\}$ in a Hilbert space is called orthonormal if

$$(\varphi_m, \varphi_n) = \begin{cases} 0, & m \neq n \\ 1, & m = n \end{cases} \qquad [36]$$

(for definitions, see, e.g., Schechter (2002) or the appendix at the end of this article).

Let $\{\varphi_n\}$ be an orthonormal sequence (finite or infinite) in a Hilbert space $H$. Let $\{\lambda_k\}$ be a sequence (of the same length) of scalars satisfying

$$|\lambda_k| \leq C$$

Then for each element $f \in H$, the series

$$\sum \lambda_k (f, \varphi_k) \varphi_k$$

converges in $H$. Define the operator $A$ on $H$ by

$$Af = \sum \lambda_k (f, \varphi_k) \varphi_k \qquad [37]$$

Clearly, $A$ is a linear operator. It is also bounded, since

$$\|Af\|^2 = \sum |\lambda_k|^2 |(f, \varphi_k)|^2 \leq C^2 \|f\|^2 \qquad [38]$$

by Bessel's inequality

$$\sum_1^\infty (f, \varphi_k)^2 \leq \|f\|^2 \qquad [39]$$

For convenience, let us assume that each $\lambda_k \neq 0$ (just remove those $\varphi_k$ corresponding to the $\lambda_k$ that vanish). In this case, $N(A)$ consists of precisely those $f \in H$ which are orthogonal to all of the $\varphi_k$. Clearly, such $f$ are in $N(A)$. Conversely, if $f \in N(A)$, then

$$0 = (Af, \varphi_k) = \lambda_k (f, \varphi_k)$$

Hence, $(f, \varphi_k) = 0$ for each $k$. Moreover, each $\lambda_k$ is an eigenvalue of $A$ with $\varphi_k$ the corresponding eigenvector. This follows immediately from [37]. Since $\sigma(A)$ is closed, it also contains the limit points of the $\lambda_k$.

Next, we shall see that if $\lambda \neq 0$ is not a limit point of the $\lambda_k$, then $\lambda \in \rho(A)$. To show this, we solve

$$(\lambda - A)u = f \qquad [40]$$

for any $f \in H$. Any solution of [40] satisfies

$$\lambda u = f + Au = f + \sum \lambda_k (u, \varphi_k) \varphi_k \qquad [41]$$

Hence,

$$\lambda (u, \varphi_k) = (f, \varphi_k) + \lambda_k (u, \varphi_k)$$

or

$$(u, \varphi_k) = \frac{(f, \varphi_k)}{\lambda - \lambda_k} \qquad [42]$$

Substituting back in [41], we obtain

$$\lambda u = f + \sum \frac{\lambda_k (f, \varphi_k) \varphi_k}{\lambda - \lambda_k} \qquad [43]$$

Since $\lambda$ is not a limit point of the $\lambda_k$, there is a $\delta > 0$ such that

$$|\lambda - \lambda_k| \geq \delta, \quad k = 1, 2, \ldots$$

Hence, the series in [43] converges for each $f \in H$. It is an easy exercise to verify that [43] is indeed a solution of [40]. To see that $(\lambda - A)^{-1}$ is bounded, note that

$$|\lambda| \cdot \|u\| \leq \|f\| + C\|f\|/\delta \qquad [44]$$

(cf. [38]). Thus, we have proved

**Lemma 4** *If the operator A is given by [37], then* $\sigma(A)$ *consists of the points* $\lambda_k$*, their limit points and possibly* 0. $N(A)$ *consists of those u which are orthogonal to all of the* $\varphi_k$*. For* $\lambda \in \rho(A)$*, the solution of [40] is given by [43].*

We see from all this that the operator [37] has many useful properties. Therefore, it would be desirable to determine conditions under which operators are guaranteed to be of that form. For this purpose, we note another property of $A$. It is expressed in terms of the Hilbert space adjoint of $A$.

Let $H_1$ and $H_2$ be Hilbert spaces, and let $A$ be an operator in $B(H_1, H_2)$. For fixed $y \in H_2$, the expression $Fx = (Ax, y)$ is a bounded linear functional on $H_1$. By the Riesz representation theorem (see, e.g., Schechter (2002) or the appendix at the end of this article), there is a $z \in H_1$ such that $Fx = (x, z)$ for all $x \in H_1$. Set $z = A^*y$. Then $A^*$ is a linear operator from $H_2$ to $H_1$ satisfying

$$(Ax, y) = (x, A^*y) \qquad [45]$$

$A^*$ is called the Hilbert space adjoint of $A$. Note the difference between $A^*$ and the operator $A'$ defined for a Banach space. As in the case of the operator $A'$, we note that $A^*$ is bounded and

$$\|A^*\| = \|A\| \qquad [46]$$

Returning to the operator $A$, we remove the assumption that each $\lambda_k \neq 0$ and note that

$$(Au, v) = \sum \lambda_k (u, \varphi_k)(\varphi_k, v) = \left(u, \sum \bar{\lambda}_k (v, \varphi_k)\varphi_k\right)$$

showing that

$$A^*v = \sum \bar{\lambda}_k (v, \varphi_k)\varphi_k \qquad [47]$$

(If $H$ is a complex Hilbert space, then the complex conjugates $\bar{\lambda}_k$ of the $\lambda_k$ are required. If $H$ is a real Hilbert space, then the $\lambda_k$ are real, and it does not matter.) Now, by Lemma 4, we see that each $\bar{\lambda}_k$ is an eigenvalue of $A^*$ with $\varphi_k$ a corresponding eigenvector. Note also that

$$\|A^*f\|^2 = \sum |\lambda_k|^2 |(f, \varphi_k)|^2 \qquad [48]$$

showing that

$$\|A^*f\| = \|Af\|, \quad f \in H \qquad [49]$$

An operator satisfying [49] is called normal. An important characterization is given by

**Theorem 24** *An operator is normal and compact if and only if it is of the form [37] with* $\{\varphi_k\}$ *an orthonormal set and* $\lambda_k \to 0$ *as* $k \to \infty$.

We also have

**Lemma 5** *If A is normal, then*

$$\|(A^* - \bar{\lambda})u\| = \|(A - \lambda)u\|, \quad u \in H \qquad [50]$$

**Corollary 2** *If A is normal and* $A\varphi = \lambda\varphi$*, then* $A^*\varphi = \bar{\lambda}\varphi$.

**Lemma 6** *If A is normal and compact, then it has an eigenvalue* $\lambda$ *such that* $|\lambda| = \|A\|$.

We also have

**Corollary 3** *If A is a normal compact operator, then there is an orthonormal sequence* $\{\varphi_k\}$ *of eigenvectors of A such that every element u in H can be written in the form*

$$u = h + \sum (u, \varphi_k)\varphi_k \qquad [51]$$

*where* $h \in N(A)$.

## Hyponormal Operators

An operator $A$ in $B(H)$ is called hyponormal if

$$\|A^*u\| \leq \|Au\|, \quad u \in H \qquad [52]$$

or, equivalently, if

$$([AA^* - A^*A]u, u) \leq 0, \quad u \in H \qquad [53]$$

Of course, a normal operator is hyponormal. An operator $A \in B(H)$ is called seminormal if either $A$ or $A^*$ is hyponormal. We have

**Theorem 25** *If A is seminormal, then*

$$r_\sigma(A) = \|A\| \qquad [54]$$

We have earlier defined the essential spectrum of an operator $A$ to be

$$\sigma_e(A) = \bigcap_{K \in K(H)} \sigma(A + K) \qquad [55]$$

It was shown that $\lambda \notin \sigma_e(A)$ if and only if $\lambda \in \Phi_A$ and $i(A - \lambda) = 0$ (Theorem 23). Let us show that we can be more specific in the case of seminormal operators.

**Theorem 26** *If A is a seminormal operator, then* $\lambda \in \sigma(A) \backslash \sigma_e(A)$ *if and only if* $\lambda$ *is an isolated eigenvalue with* $r(A - \lambda) = \lim_{n\to\infty} \alpha[(A - \lambda)^n] < \infty$.

**Lemma 7** *If A is hyponormal, then so is* $B = A - \lambda$ *for any complex* $\lambda$.

**Lemma 8** *If B is hyponormal with 0 an isolated point of* $\sigma(B)$ *and either* $\alpha(B)$ *or* $\beta(B)$ *is finite, then* $B \in \Phi(H)$ *and* $i(B) = 0$.

There is a simple consequence of Lemma 8.

**Corollary 4** *If A is seminormal and $\lambda$ is an isolated point of $\sigma(A)$, then $\lambda$ is an eigenvalue of A.*

We also have the following:

**Theorem 27** *Let A be a seminormal operator such that $\sigma(A)$ has no nonzero limit points. Then A is compact and normal. Thus, it is of the form [37] with the $\{\varphi_k\}$ orthonormal and $\lambda_k \to 0$.*

**Corollary 5** *If A is seminormal and compact, then it is normal.*

## Spectral Resolution

We saw in the section "Operational calculus" that, in a Banach space $X$, we can define $f(A)$ for any $A \in B(X)$ provided $f(z)$ is a function analytic in a neighborhood of $\sigma(A)$. In this section, we shall show that we can do better in the case of self-adjoint operators.

A linear operator $A$ on a Hilbert space $X$ is called self-adjoint if it has the property that $x \in D(A)$ and $Ax = f$ if and only if

$$(x, Ay) = (f, y), \quad y \in D(A)$$

In particular, it satisfies

$$(Ax, y) = (x, Ay), \quad x, y \in D(A)$$

A bounded self-adjoint operator is normal.

To get an idea, let $A$ be a compact, self-adjoint operator on $H$. Then by Theorem 24,

$$Au = \sum \lambda_k (u, \varphi_k)\varphi_k \qquad [56]$$

where $\{\varphi_k\}$ is an orthonormal sequence of eigenvectors and the $\lambda_k$ are the corresponding eigenvalues of $A$. Now let $p(t)$ be a polynomial with real coefficients having no constant term

$$p(t) = \sum_{1}^{m} a_k t^k \qquad [57]$$

Then $p(A)$ is compact and self-adjoint. Let $\mu \neq 0$ be a point in $\sigma(p(A))$. Then $\mu = p(\lambda)$ for some $\lambda \in \sigma(A)$ (Theorem 8). Now $\lambda \neq 0$ (otherwise we would have $\mu = p(0) = 0$). Hence, it is an eigenvalue of $A$ (see the section "The spectrum and resolvent sets"). If $\varphi$ is a corresponding eigenvector, then

$$\begin{aligned}[p(A) - \mu]\varphi &= \sum a_k A^k \varphi - \mu\varphi \\ &= \sum a_k \lambda^k \varphi - \mu\varphi \\ &= [p(\lambda) - \mu]\varphi = 0\end{aligned}$$

Thus $\mu$ is an eigenvalue of $p(A)$ and $\varphi$ is a corresponding eigenvector. This shows that

$$p(A)u = \sum p(\lambda_k)(u, \varphi_k)\varphi_k \qquad [58]$$

Now, the right-hand side of [58] makes sense if $p(t)$ is any function bounded on $\sigma(A)$ (see the section "Normal operators"). Therefore it seems plausible to define $p(A)$ by means of [58]. Of course, for such a definition to be useful, one would need certain relationships to hold. In particular, one would want $f(t)g(t) = h(t)$ to imply $f(A)g(A) = h(A)$. We shall discuss this a bit later.

If $A$ is not compact, we cannot, in general, obtain an expansion in the form [56]. However, we can obtain something similar. In fact, we have

**Theorem 28** *Let A be a self-adjoint operator in $B(H)$. Set*

$$m = \inf_{\|u\|=1} (Au, u), \qquad M = \sup_{\|u\|=1} (Au, u)$$

*Then there is a family $\{E(\lambda)\}$ of orthogonal projection operators on H depending on a real parameter $\lambda$ and such that:*

(i) $E(\lambda_1) \le E(\lambda_2)$ for $\lambda_1 \le \lambda_2$;
(ii) $E(\lambda)u \to E(\lambda_0)u$ as $\lambda_0 < \lambda \to \lambda_0, u \in H$;
(iii) $E(\lambda) = 0$ for $\lambda < m, E(\lambda) = I$ for $\lambda \ge M$;
(iv) $AE(\lambda) = E(\lambda)A$; *and*
(v) *if $a < m, b \ge M$ and $p(t)$ is any polynomial, then*

$$p(A) = \int_a^b p(\lambda)\, \mathrm{d}E(\lambda) \qquad [59]$$

*This means the following. Let $a = \lambda_0 < \lambda_1 < \cdots < \lambda_n = b$ be any partition of $[a, b]$, and let $\lambda'_k$ be any number satisfying $\lambda_{k-1} \le \lambda'_k \le \lambda_k$. Then*

$$\sum_{1}^{n} p(\lambda'_k)[E(\lambda_k) - E(\lambda_{k-1})] \to p(A) \qquad [60]$$

*in $B(H)$ as $\eta = \max(\lambda_k - \lambda_{k-1}) \to 0$.*

**Theorem 29** *Let A be a self-adjoint operator on H. Then there is a family $\{E(\lambda)\}$ of orthogonal projection operators on H satisfying* (i) *and* (ii) *of Theorem 28 and*

(i) $$E(\lambda) \to \begin{cases} 0 & \text{as } \lambda \to -\infty \\ I & \text{as } \lambda \to +\infty \end{cases}$$

(ii) $$E(\lambda)A \subset AE(\lambda)$$

(iii) $$p(A) = \int_{-\infty}^{\infty} p(\lambda)\, \mathrm{d}E(\lambda)$$

*for any polynomial $p(t)$.*

These theorems are known as the spectral theorems for self-adjoint operators.

## Appendix

Here we include some background material related to the text.

Consider a collection $C$ of elements or "vectors" with the following properties:

1. They can be added. If $f$ and $g$ are in $C$, so is $f+g$.
2. $f+(g+h)=(f+g)+h, \quad f,g,h \in C$.
3. There is an element $0 \in C$ such that $h+0=h$ for all $h \in C$.
4. For each $h \in C$ there is an element $-h \in C$ such that $h+(-h)=0$.
5. $g+h=h+g, \quad g,h \in C$.
6. For each real number $\alpha, \alpha h \in C$.
7. $\alpha(g+h)=\alpha g+\alpha h$.
8. $(\alpha+\beta)h=\alpha h+\beta h$.
9. $\alpha(\beta h)=(\alpha\beta)h$.
10. To each $h \in C$ there corresponds a real number $\|h\|$ with the following properties:
11. $\|\alpha h\|=|\alpha|\|h\|$.
12. $\|h\|=0$ if, and only if, $h=0$.
13. $\|g+h\| \leq \|g\|+\|h\|$.
14. If $\{h_n\}$ is a sequence of elements of $C$ such that $\|h_n-h_m\| \to 0$ as $m,n \to \infty$, then there is an element $h \in C$ such that $\|h_n-h\| \to 0$ as $n \to \infty$.

A collection of objects which satisfies statements (1)–(9) and the additional statement

15. $1h=h$

is called a vector space or linear space.

A set of objects satisfying statements (1)–(13) is called a normed vector space, and the number $\|h\|$ is called the norm of $h$. Although statement (15) is not implied by statements (1)–(9), it is implied by statements (1)–(13). A sequence satisfying

$$\|h_n-h_m\| \to 0 \quad \text{as } m,n \to \infty$$

is called a Cauchy sequence. Property (14) states that every Cauchy sequence converges in norm to a limit (i.e., satisfies $\|h_n-h\| \to 0$ as $n \to \infty$). Property (14) is called completeness, and a normed vector space satisfying it is called a complete normed vector space or a Banach space.

We shall write

$$h_n \to h \quad \text{as } n \to \infty$$

when we mean

$$\|h_n-h\| \to 0 \quad \text{as } n \to \infty$$

A subset $U$ of a vector space $V$ is called a subspace of $V$ if $\alpha_1 x_1+\alpha_2 x_2$ is in $U$ whenever $x_1, x_2$ are in $U$ and $\alpha_1, \alpha_2$ are scalars.

A subset $U$ of a normed vector space $X$ is called closed if for every sequence $\{x_n\}$ of elements in $U$ having a limit in $X$, the limit is actually in $U$.

Consider a vector space $X$ having a mapping $(f,g)$ from pairs of its elements to the reals such that

1. $(\alpha f,g)=\alpha(f,g)$
2. $(f+g,h)=(f,h)+(g,h)$
3. $(f,g)=(g,f)$
4. $(f,f)>0$ unless $f=0$.

Then

$$(f,g)^2 \leq (f,f)(g,g), \quad f,g \in X \qquad [61]$$

An expression $(f,g)$ that assigns a real number to each pair of elements of a vector space and satisfies the aforementioned properties is called a scalar (or inner) product.

If a vector space $X$ has a scalar product $(f,g)$, then it is a normed vector space with norm $\|f\|=(f,f)^{1/2}$. A vector space which has a scalar product and is complete with respect to the induced norm is called a Hilbert space. Every Hilbert space is a Banach space, but the converse is not true. Inequality [61] is known as the Cauchy–Schwarz inequality. $\mathbb{R}^n$ is a Hilbert space.

Let $H$ be a Hilbert space and let $(x,y)$ denote its scalar product. If we fix $y$, then the expression $(x,y)$ assigns to each $x \in H$ a number. An assignment $F$ of a number to each element $x$ of a vector space is called a functional and denoted by $F(x)$. The scalar product is not the first functional we have encountered. In any normed vector space, the norm is also a functional. The functional $F(x)=(x,y)$ satisfies

$$F(\alpha_1 x_1+\alpha_2 x_2)=\alpha_1 F(x_1)+\alpha_2 F(x_2) \qquad [62]$$

for $\alpha_1, \alpha_2$ scalars. A functional satisfying [62] is called linear. Another property is

$$|F(x)| \leq M\|x\|, \quad x \in H \qquad [63]$$

which follows immediately from Schwarz's inequality (cf. [61]). A functional satisfying [63] is called bounded. The norm of such a functional is defined to be

$$\|F\| = \sup_{x \in H, x \neq 0} \frac{|F(x)|}{\|x\|}$$

Thus for $y$ fixed, $F(x)=(x,y)$ is a bounded linear functional in the Hilbert space $H$. We have

**Theorem 30** *For every bounded linear functional F on a Hilbert space H there is a unique element y ∈ H such that*

$$F(x) = (x, y) \quad \text{for all } x \in H \qquad [64]$$

*Moreover,*

$$\|y\| = \sup_{x\in H, x\neq 0} \frac{|F(x)|}{\|x\|} = \|F\| \qquad [65]$$

Theorem 30 is known as the "Riesz representation theorem."

For any normed vector space $X$, let $X'$ denote the set of bounded linear functionals on $X$. If $f, g \in X'$, we say that $f = g$ if

$$f(x) = g(x) \quad \text{for all } x \in X$$

The "zero" functional is the one assigning zero to all $x \in X$. We define $h = f + g$ by

$$h(x) = f(x) + g(x), \quad x \in X$$

and $g = \alpha f$ by

$$g(x) = \alpha f(x), \quad x \in X$$

Under these definitions, $X'$ becomes a vector space. We have been employing the expression

$$\|f\| = \sup_{x\neq 0} \frac{|f(x)|}{\|x\|}, \quad f \in X' \qquad [66]$$

This is easily seen to be a norm. Thus $X'$ is a normed vector space.

We also have

**Theorem 31** *Let M be a subspace of a normed vector space X, and suppose that f(x) is a bounded linear functional on M. Set*

$$\|f\| = \sup_{x\in M, x\neq 0} \frac{|f(x)|}{\|x\|}$$

*Then there is a bounded linear functional F(x) on the whole of X such that*

$$F(x) = f(x), \quad x \in M \qquad [67]$$

*and*

$$\|F\| = \sup_{x\in X, x\neq 0} \frac{|F(x)|}{\|x\|} = \|f\| = \sup_{x\in M, x\neq 0} \frac{|f(x)|}{\|x\|} \qquad [68]$$

Theorem 31 is known as the "Hahn–Banach theorem."

If $A$ is a linear operator from $X$ to $Y$, with $R(A) = Y$ and $N(A) = \{0\}$ (i.e., consists only of the vector 0), we can assign to each $y \in Y$ the unique solution of

$$Ax = y$$

This assignment is an operator from $Y$ to $X$ and is usually denoted by $A^{-1}$ and called the inverse operator of $A$. It is linear because of the linearity of $A$. One can ask: "when is $A^{-1}$ continuous?" or, equivalent by, "when is it bounded?" A very important answer to this question is given by

**Theorem 32** *If X, Y are Banach spaces and A is a closed linear operator from X to Y with* $R(A) = Y, N(A) = \{0\}$, *then* $A^{-1} \in B(Y, X)$.

This theorem is sometimes referred to as the "bounded inverse theorem."

If $A$ is self-adjoint and

$$(A - \lambda)x = 0, \qquad (A - \mu)y = 0$$

with $\lambda \neq \mu$, then

$$(x, y) = 0$$

If $A$ has a compact inverse, its eigenvalues cannot have limit points. If $A^{-1}$ is compact, then the eigenelements corresponding to the same eigenvalue form a finite-dimensional subspace.

*See also:* Ljusternik–Schnirelman Theory; Quantum Mechanical Scattering Theory; Regularization for Dynamical Zeta Functions; Spectral Sequences; Stochastic Resonance.

## Further Reading

Bachman G and Narici L (1966) *Functional Analysis*. New York: Academic Press.
Banach S (1955) *Théorie des Opérations Linéaires*. New York: Chelsea.
Berberian SK (1974) *Lectures in Functional Analysis and Operator Theory*. New York: Springer.
Brown A and Pearcy C (1977) *Introduction to Operator Theory*. New York: Springer.
Day MM (1958) *Normed Linear Spaces*. Berlin: Springer.
Dunford N and Schwartz IT (1958, 1963) *Linear Operators, I, II*. New York: Wiley.
Edwards RE (1965) *Functional Analysis*. New York: Holt.
Epstein B (1970) *Linear Functional Analysis*. Philadelphia: W. B. Saunders.
Gohberg IC and Krein MS (1960) The basic propositions on defect numbers, root numbers and indices of linear operators. *Amer. Math. Soc. Transl*, Ser. 2, 13, 185–264.
Goldberg S (1966) *Unbounded Linear Operators*. New York: McGraw-Hill.
Halmos PR (1951) *Introduction to Hilbert Space*. New York: Chelsea.
Hille E and Phillips R (1957) *Functional Analysis and Sem-Groups*. Providence: American Mathematical Society.
Kato T (1966, 1976) *Perturbation Theory for Linear Operators*. Berlin: Springer.

Müller V (2003) *Spectral Theory of Linear Operators – and Spactral System in Banach Algebras*. Basel: Birkhäuser Verlag.
Reed M and Simon B (1972) *Methods of Modern Mathematical Physics, I*. Academic Press.
Riesz F and St.-Nagy B (1955) *Functional Analysis*. New York: Ungar.
Schechter M (2002) *Principles of Functional Analysis*. Providence: American Mathematical Society.
Stone MH (1932) *Linear Transformations in Hilbert Space*. Providence: American Mathematical Society.
Taylor AE (1958) *Introduction to Functional Analysis*. New York: Wiley.
Weidmann J (1980) *Linear Operators in Hilbert Space*. New York: Springer.
Yosida K (1965, 1971) *Functional Analysis*. Berlin: Springer.

# Spin Foams

**A Perez**, Penn State University, University Park, PA, USA



## Introduction

In loop quantum gravity (LQG) (*see* Loop Quantum Gravity) – a background independent formulation of quantum gravity – the full quantum dynamics is governed by the following (constraint) operator equations or quantum Einstein equations:

Gauss Law

$$\widehat{G}_i(A,E)|\Psi>:=\widehat{D_aE_i^a}|\Psi>=0$$

Vector constraint

$$\widehat{V}_a(A,E)|\Psi>:=\widehat{E_i^aF_{ab}^i(A)}|\Psi>=0$$

Scalar constraint

$$\widehat{S}(A,E)\Big|\Psi>:=\left[\widehat{\sqrt{\det E}^{-1}E_i^aE_j^bF_{ab}^{ij}(A)+\cdots}\right]\Big| \Psi>=0 \qquad [1]$$

where $A_a^i$ is an SU(2) connection ($i=1,2,3$, $a=1,2,3$), $E_i^a$ is its conjugate momentum (the triad field), $F^{ij}_{ab}(A)$ is the curvature of $A_a^i$, and $D_a$ is the covariant derivative (*see* Canonical General Relativity). The hat means that the classical phase-space functions are promoted to operators in a kinematical Hilbert space $\mathcal{H}_{\rm kin}$; the solutions are in the so-called physical Hilbert space $\mathcal{H}_{\rm phys}$. The goal of the spin foam approach is to construct a mathematically well-defined notion of path integral for LQG as a device for computing the solutions of the previous equations.

The space of solution of the Gauss and vector constraints [1] is well understood in LQG (*see* Loop Quantum Gravity), and often also called kinematical Hilbert space $\mathcal{H}_{\rm kin}$. The solutions of the scalar constraint can be characterized by the definition of the generalized projection operator $P$ from the kinematical Hilbert space $\mathcal{H}_{\rm kin}$ into the kernel of the scalar constraint $\mathcal{H}_{\rm phys}$. Formally, one can write $P$ as

$$P=\text{“}\prod_{x\in}\delta(\widehat{S}(x))\text{”} = \int D[N]\exp\left[i\int_\Sigma N(x)\widehat{S(x)}\right] \qquad [2]$$

A formal argument shows that $P$ can also be defined in a manifestly covariant manner as a regularization of the formal path integral of general relativity. In first-order variables, it becomes

$$P=\int D[e]\ D[A]\ \mu[A,e]\exp[iS_{\rm GR}(e,A)] \qquad [3]$$

where $e$ is the tetrad field, $A$ is the spacetime connection, and $\mu[A,e]$ denotes the appropriate measure.

In both cases, $P$ characterizes the space of solutions of quantum Einstein equations as for any arbitrary state $|\phi>\in\mathcal{H}_{\rm kin}$ then $P|\phi>$ is a (formal) solution of [1]. Moreover, the matrix elements of $P$ define the physical inner product $(<,>_{\rm p})$ providing the vector space of solutions of [1] with the Hilbert space structure that defines $\mathcal{H}_{\rm phys}$. Explicitly,

$$<s,s'>_{\rm p}\ :=\ <Ps,s'>$$

for $s,s'\in\mathcal{H}_{\rm kin}$.

When these matrix elements are computed in the spin network basis (see **Figure 1**) (*see* Loop Quantum Gravity), they can be expressed as a sum over amplitudes of “spin network histories”: spin foams (**Figure 2**). The latter are naturally given by foam-like combinatorial structures whose basic elements carry quantum numbers of geometry (*see* Loop Quantum Gravity). A spin foam history, from the state $|s>$ to the state $|s'>$, is denoted by a pair $(F_{s\to s'},\{j\})$, where $F_{s\to s'}$ is the 2-complex with boundary given by the graphs of the spin network states $|s'>$ and $|s>$, respectively, and $\{j\}$ is the set of spin quantum numbers labeling its edges (denoted $e\in F_{s\to s'}$) and faces (denoted $f\in F_{s\to s'}$). Vertices are denoted

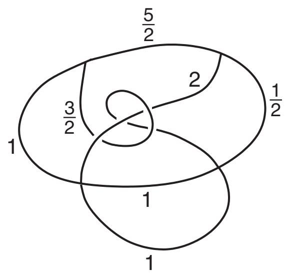



**Figure 1** A spin network state is given by a graph embedded in space whose links and nodes are labeled by unitary irreducible representations of SU(2). These states form a complete basis of the kinematical Hilbert space of LQG where the operator equations [1] are defined.

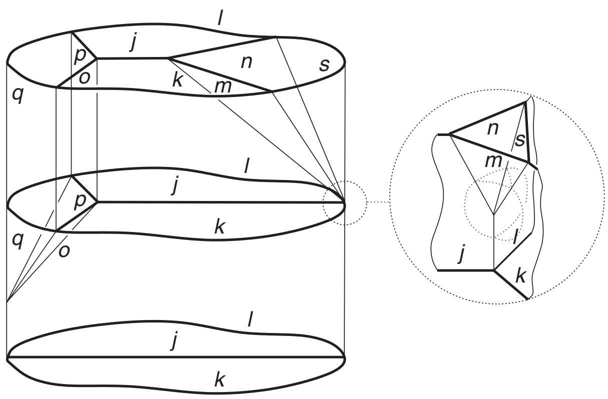



**Figure 2** A spin foam as the "colored" 2-complex representing the transition between three different spin network states. A transition vertex is magnified on the right.

$v \in F_{s \to s'}$. The physical inner product can be expressed as a sum over spin foam amplitudes

$$
\begin{aligned}
\langle s', s\rangle_{\mathrm{p}} &= \langle Ps', s\rangle \\
&= \sum_{F_{s\to s'}} N(F_{s\to s'}) \sum_{\{j\}} \prod_{f\in F_{s\to s'}} A_f(j_f) \\
&\quad \times \prod_{e\in F_{s\to s'}} A_e(j_e) \prod_{v\in F_{s\to s'}} A_v(j_v)
\end{aligned} \qquad [4]
$$

where $N(F_{s\to s'})$ is a (possible) normalization factor, and $A_f(j_f), A_e(j_e)$, and $A_v(j_v)$ are the 2-cell or face amplitude, the edge or 1-cell amplitude, and the 0-cell or vertex amplitude, respectively. These local amplitudes depend on the spin quantum numbers labeling neighboring cells in $F_{s\to s'}$ (e.g., the vertex amplitude of the vertex magnified in Figure 2 is $A_v(j,k,l,m,n,s)$).

The underlying discreteness discovered in LQG is crucial: in the spin foam representation, the functional integral for gravity is replaced by a sum over amplitudes of combinatorial objects given by foam-like configurations (spin foams) as in [4]. A spin foam represents a possible history of the gravitational field and can be interpreted as a set of transitions through different quantum states of space. Boundary data in the path integral are given by the polymer-like excitations (spin network states, Figure 1) representing 3-geometry states in LQG.

## Spin Foams in 3D Quantum Gravity

Now we introduce the concept of spin foams in a more explicit way in the context of the quantization of three-dimensional (3D) Riemannian gravity. Later in this section we will present the definition of $P$ from the canonical and covariant viewpoint formally stated in the introduction by eqns [2] and [3], respectively.

### The Classical Theory

Riemannian gravity in 3D is a theory with no local degrees of freedom, that is, a topological theory (*see* Topological Quantum Field Theory: Overview). Its action (in the first-order formalism) is given by

$$S(e,\omega) = \int_M \mathrm{tr}(e \wedge F(\omega)) \qquad [5]$$

where $M = \Sigma \times \mathbb{R}$ (for $\Sigma$ an arbitrary Riemann surface), $\omega$ is an SU(2) connection, and the triad $e$ is an su(2)-valued 1-form. The gauge symmetries of the action are the local SU(2) gauge transformations

$$\delta e = [e, \alpha], \qquad \delta\omega = d_\omega \alpha \qquad [6]$$

where $\alpha$ is an su(2)-valued 0-form, and the "topological" gauge transformation

$$\delta e = d_\omega \eta, \qquad \delta\omega = 0 \qquad [7]$$

where $d_\omega$ denotes the covariant exterior derivative and $\eta$ is an su(2)-valued 0-form. The first invariance is manifest from the form of the action, while the second is a consequence of the Bianchi identity, $d_\omega F(\omega) = 0$. The gauge symmetries are so large that all the solutions to the equations of motion are locally pure gauge. The theory has only global or topological degrees of freedom.

Upon the standard $2+1$ decomposition (*see* Canonical General Relativity), the phase space in these variables is parametrized by the pullback to $\Sigma$ of $\omega$ and $e$. In local coordinates, one can express them in terms of the two 2D connection $A_a^i$ and the triad field $E_j^b = \epsilon^{bc} e_c^k \eta_{jk}$, where $a = 1, 2$ are space coordinate indices and $i, j = 1, 2, 3$ are su(2) indices. The symplectic structure is defined by

$$\{A_a^i(x), E_j^b(y)\} = \delta_a^{\ b} \delta^i_{\ j} \delta^{(2)}(x,y) \qquad [8]$$

Local symmetries of the theory are generated by the first-class constraints

$$D_b E_j^b = 0, \qquad F_{ab}^i(A) = 0 \qquad [9]$$

which are referred to as the Gauss law and the curvature constraint, respectively – the quantization of these is the analog of [1] in 4D. This simple theory has been quantized in various ways in the literature; here we will use it to introduce the spin foam quantization.

### Kinematical Hilbert Space

In analogy with the 4D case, one follows Dirac's procedure finding first a representation of the basic variables in an auxiliary or kinematical Hilbert space $\mathcal{H}_{\text{kin}}$. The basic states are functionals of the connection depending on the parallel transport along paths $\gamma \subset \Sigma$: the so-called holonomy. Given a connection $A_a^i(x)$ and a path $\gamma$, one defines the holonomy $h_\gamma[A]$ as the path-ordered exponential

$$h_\gamma[A] = P \exp \int_\gamma A \qquad [10]$$

The kinematical Hilbert space, $\mathcal{H}_{\text{kin}}$, corresponds to the Ashtekar–Lewandowski (AL) representation of the algebra of functions of holonomies or generalized connections. This algebra is in fact a $C^*$-algebra and is denoted Cyl (*see* Loop Quantum Gravity). Functionals of the connection act in the AL representation simply by multiplication. For example, the holonomy operator acts as follows:

$$\widehat{h_\gamma[A]}\Psi[A] = h_\gamma[A]\Psi[A] \qquad [11]$$

As in 4D, an orthonormal basis of $\mathcal{H}_{\text{kin}}$ is defined by the spin network states. Each spin network is labeled by a graph $\gamma \subset \Sigma$, a set of spins $\{j_\ell\}$ labeling links $\ell \in \gamma$, and a set of intertwiners $\{\iota_n\}$ labeling nodes $n \in \gamma$ (**Figure 3**), namely:

$$s_{\gamma,\{j_\ell\},\{\iota_n\}}[A] = \bigotimes_{n\in\gamma} \iota_n \bigotimes_{\ell\in\gamma} \overset{j_\ell}{\prod}(h_\ell[A]) \qquad [12]$$

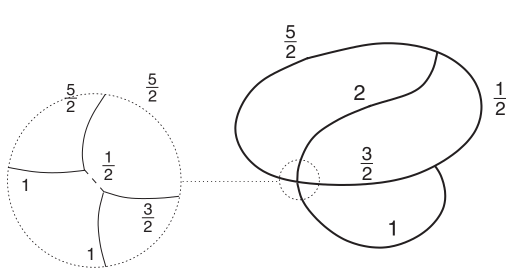



**Figure 3** A spin network state in $2+1$ LQG. The decomposition of a 4-valent node in terms of basic 3-valent intertwiners is shown.

where $\overset{j}{\Pi}$ is the unitary irreducible representation matrix of spin $j$ (for a precise definition, *see* Loop Quantum Gravity). For simplicity, we will often denote spin network states $|s>$ omitting the graph and spin labels.

### Spin Foams from the Hamiltonian Formulation

The physical Hilbert space, $\mathcal{H}_{\text{phys}}$, is defined by those "states" that are annihilated by the constraints. By construction, spin-network states solve the Gauss constraint – $\widehat{D_a E_i^a}|s>=0$ – as they are manifestly SU(2) gauge invariant (*see* Loop Quantum Gravity). To complete the quantization, one needs to characterize the space of solutions of the quantum curvature constraints $(\widehat{F}_{ab}^i)$, and to provide it with the physical inner product. The existence of $\mathcal{H}_{\text{phys}}$ is granted by the following:

**Theorem 1** *There exists a normalized positive linear form P over* Cyl, *that is,* $P(\psi^*\psi) \geq 0$ *for* $\psi \in$ Cyl *and* $P(1)=1$, *yielding (through the GNS construction* (*see* Algebraic Approach to Quantum Field Theory)) *the physical Hilbert space* $\mathcal{H}_{\text{phys}}$ *and the physical representation* $\pi_p$ *of* Cyl.

The state $P$ contains a very large Gelfand ideal (set of zero norm states) $J := \{\alpha \in \text{Cyl s.t. } P(\alpha^*\alpha)=0\}$. In fact, the physical Hilbert space $\mathcal{H}_{\text{phys}} := \overline{\text{Cyl}/J}$ corresponds to the quantization of finitely many degrees of freedom. This is expected in 3D gravity as the theory does not have local excitations (no "gravitons") (*see* Topological Quantum Field Theory: Overview). The representation $\pi_p$ of Cyl solves the curvature constraint in the sense that for any functional $f_\gamma[A] \in$ Cyl defined on the subalgebra of functionals defined on contractible graphs $\gamma \in \Sigma$, one has that

$$\pi_p[f_\gamma]\Psi = f_\gamma[0]\Psi \qquad [13]$$

This equation expresses the fact that "$\widehat{F}=0$" in $\mathcal{H}_{\text{phys}}$ (for flat connections, parallel transport is trivial around a contractible region). For $s, s' \in \mathcal{H}_{\text{kin}}$, the physical inner product is given by

$$<s, s'>_{\text{p}} := P(s^*s) \qquad [14]$$

where the $*$-operation and the product are defined in Cyl.

The previous equation admits a "sum over histories" representation. We shall introduce the concept of the spin foam representation as an explicit construction of the positive linear form $P$ which, as in [2], is formally given by

$$\begin{aligned} P &= \int D[N] \exp\left(\mathrm{i} \int_\Sigma \text{tr}[N\widehat{F}(A)]\right) \\ &= \prod_{x\in\Sigma} \delta[\widehat{F(A)}] \end{aligned} \qquad [15]$$

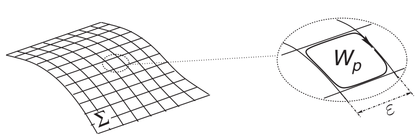



**Figure 4** Cellular decomposition of the space manifold Σ (a square lattice in this example), and the infinitesimal plaquette holonomy $W_p[A]$.

where $N(x) \in \mathrm{su}(2)$. One can make the previous formal expression a rigorous definition if one introduces a regularization. Given a partition of Σ in terms of 2D plaquettes of coordinate area $\epsilon^2$, one has that

$$\int_\Sigma \mathrm{tr}[NF(A)] = \lim_{\epsilon \to 0} \sum_{p^i} \epsilon^2 \mathrm{tr}[N_{p^i} F_{p^i}] \qquad [16]$$

where $N_{p^i}$ and $F_{p^i}$ are values of $N^i$ and $\epsilon^{ab} F^i_{ab}[A]$ at some interior point of the plaquette $p^i$ and $\epsilon^{ab}$ is the Levi-Civita tensor. Similarly, the holonomy $W_{p^i}[A]$ around the boundary of the plaquette $p^i$ (see Figure 4) is given by

$$W_{p^i}[A] = \mathbb{1} + \epsilon^2 F_{p^i}(A) + \mathcal{O}(\epsilon^2) \qquad [17]$$

where $F_{p^i} = \tau_j \epsilon^{ab} F^j_{ab}(x_{p^i})$ ($\tau_j$ are the generators of su(2) in the fundamental representation). The previous two equations lead to the following definition: given $s \in \mathrm{Cyl}$ (think of spin network state based on a graph $\gamma$), the linear form $P(s)$ is defined as

$$P(s) := \lim_{\epsilon \to 0} \left\langle \Omega \prod_{p^i} \int \mathrm{d}N_{p^i} \exp(\mathrm{i}\ \mathrm{tr}[N_{p^i} W_{p^i}]), s \right\rangle \qquad [18]$$

where $<,>$ is the inner product in the AL representation and $|\Omega>$ is the "vacuum" ($1 \in \mathrm{Cyl}$) in the AL representation. The partition is chosen so that the links of the underlying graph $\gamma$ border the plaquettes. One can easily perform the integration over the $N_{p^i}$ using the identity (Peter–Weyl theorem)

$$\int \mathrm{d}N \exp(\mathrm{i}\ \mathrm{tr}[NW]) = \sum_j (2j+1) \mathrm{tr}\left[\overset{j}{\Pi}(W)\right] \qquad [19]$$

Using the previous equation

$$P(s) := \lim_{\epsilon \to 0} \prod_{p^i} \sum_{j(p^i)} (2j(p^i) + 1) <\Omega\ \mathrm{tr}\left[\overset{j(p^i)}{\Pi}(W_{p^i})\right], s> \qquad [20]$$

where $j(p^i)$ is the spin labeling element of the sum [19] associated to the $i$th plaquette. Since the $\mathrm{tr}[\Pi^j(W)]$ commute, the ordering of plaquette operators in the previous product does not matter. It can be shown that the limit $\epsilon \to 0$ exists and one can give a closed expression of $P(s)$.

Now in the AL representation (see eqn [11]), each $\mathrm{tr}[\Pi^{j(p^i)}(W_{p^i})]$ acts by creating a closed loop in the $j_{p^i}$ representation at the boundary of the corresponding plaquette (Figures 5 and 6).

One can introduce a (nonphysical) time parameter that works simply as a coordinate providing the means of organizing the series of actions of plaquette loop operators in [20]; that is, one assumes that each of the loop actions occurs at different "times." We have introduced an auxiliary time slicing (arbitrary parametrization). If one inserts the AL partition of unity

$$\mathbb{1} = \sum_{\gamma \in \Sigma} \sum_{\{j\}_\gamma} |\gamma, \{j\} >< \gamma, \{j\}| \qquad [21]$$

where the sum is over the complete basis of spin network states $\{|\gamma, \{j\}>\}$ – based on all graphs $\gamma \in \Sigma$ and with all possible spin labeling – between each time

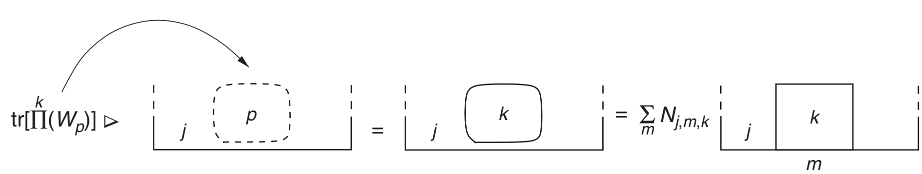


**Figure 5** Graphical notation representing the action of one plaquette holonomy on a spin network state. On the right is the result written in terms of the spin network basis. The amplitude $N_{j,m,k}$ can be expressed in terms of Clebsch–Gordan coefficients.

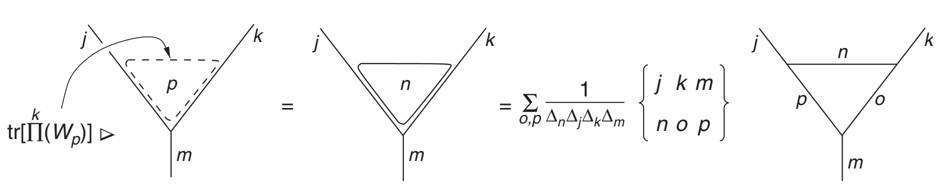


**Figure 6** Graphical notation representing the action of one plaquette holonomy on a spin network vertex. The object in brackets ({}) is a 6$j$-symbol and $\Delta_j := 2j + 1$.

slice, one arrives at a sum over spin network histories representation of $P(s)$. More precisely, $P(s)$ can be expressed as a sum over amplitudes corresponding to a series of transitions that can be viewed as the "time evolution" between the "initial" spin network $s$ and the "final" "vacuum state" $\Omega$. The physical inner product between spin networks $s$ and $s'$ is defined as

$$<s, s'>_{\mathrm{p}} := P(s^* s')$$

and can be expressed as a sum over amplitudes corresponding to transitions interpolating between the "initial" spin network $s'$ and the "final" spin network $s$ (e.g., Figures 7 and 8).

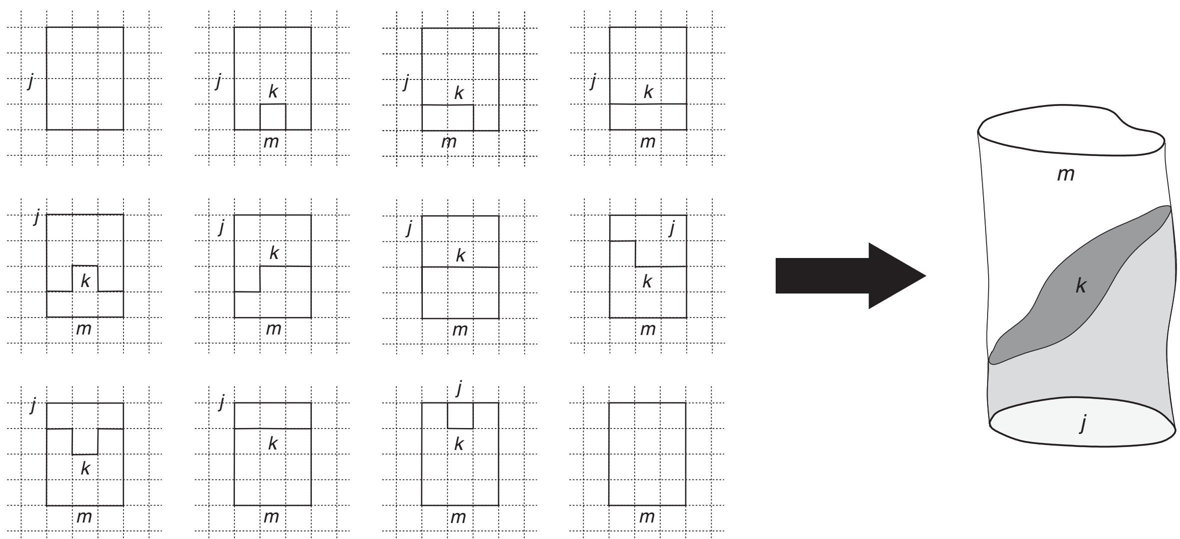



**Figure 7** A set of discrete transitions in the loop-to-loop physical inner product obtained by a series of transitions as in Figure 5. On the right, the continuous spin foam representation in the limit $\epsilon \to 0$.

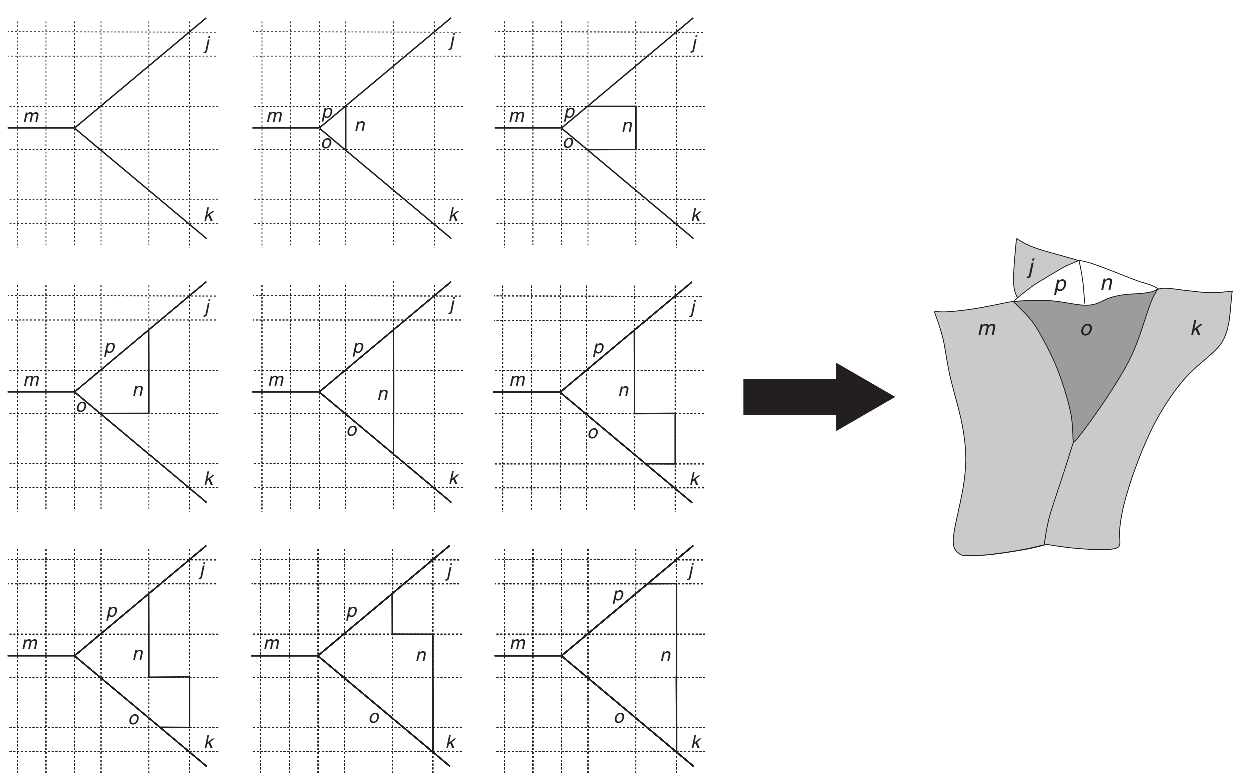
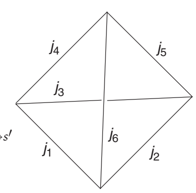



**Figure 8** A set of discrete transitions representing one of the contributing histories at a fixed value of the regulator. On the right, the continuous spin foam representation when the regulator is removed.

Spin network nodes evolve into edges while spin network links evolve into 2D faces. Edges inherit the intertwiners associated to the nodes and faces inherit the spins associated to links. Therefore, the series of transitions can be represented by a 2-complex whose 1-cells are labeled by intertwiners and whose 2-cells are labeled by spins. The places where the action of the plaquette loop operators create new links (Figures 6 and 8) define 0-cells or vertices. These foam-like structures are the so-called spin foams. The spin foam amplitudes are purely combinatorial and can be explicitly computed from the simple action of the loop operator in the AL representation (*see* Loop Quantum Gravity). A particularly simple case arises when the spin network states $s$ and $s'$ have only 3-valent nodes. Explicitly,

$$<s, s'>_{\mathrm{p}} := P(s^* s')$$

$$= \sum_{\{j\}} \prod_{f \in F_{s \to s'}} (2j_f + 1)^{\frac{\nu_f}{2}} \prod_{v \in F_{s \to s'}} \qquad [22]$$




where the notation is that of [4], and $\nu_f = 0$ if $f \cap s \neq 0 \wedge f \cap s' \neq 0, \nu_f = 1$ if $f \cap s \neq 0 \vee f \cap s' \neq 0$, and $\nu_f = 2$ if $f \cap s = 0 \wedge f \cap s' = 0$. The tetrahedral diagram denotes a 6*j*-symbol: the amplitude obtained by means of the natural contraction of the four intertwiners corresponding to the 1-cells converging at a vertex. More generally, for arbitrary spin networks, the vertex amplitude corresponds to 3*nj*-symbols, and $<s, s'>_{\mathrm{p}}$ takes the general form [4].

### Spin Foams from the Covariant Path Integral

In this section we re-derive the spin foam representation of the physical scalar product of $2+1$ (Riemannian) quantum gravity directly as a regularization of the covariant path integral. The formal path integral for 3D gravity can be written as

$$P = \int D[e] D[A] \exp\left[ \mathrm{i} \int_M \mathrm{tr}[e \wedge F(A)] \right] \qquad [23]$$

Assume $M = \Sigma \times I$, where $I \subset \mathbb{R}$ is a closed (time) interval (for simplicity, we ignore boundary terms).

In order to give a meaning to the formal expression above, one replaces the 3D manifold (with boundary) $M$ with an arbitrary cellular decomposition $\Delta$. One also needs the notion of the associated dual 2-complex of $\Delta$ denoted by $\Delta^*$. The dual 2-complex $\Delta^*$ is a combinatorial object defined by a set of vertices $v \in \Delta^*$ (dual to 3-cells in $\Delta$), edges $e \in \Delta^*$ (dual to 2-cells in $\Delta$), and faces $f \in \Delta^*$

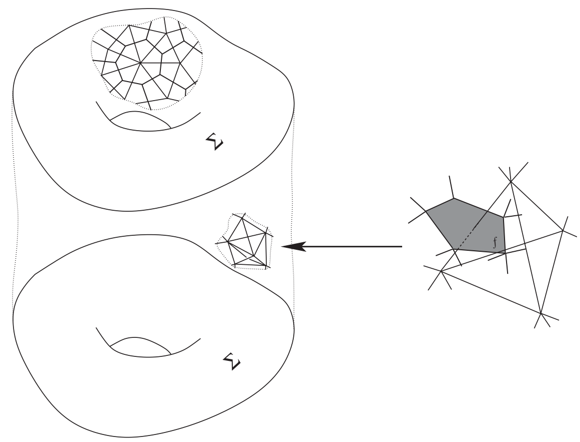



**Figure 9** The cellular decomposition of $M = \Sigma \times I$ ($\Sigma = T^2$ in this example). The illustration shows part of the induced graph on the boundary and the detail of a tetrahedron in $\Delta$ and a face $f \in \Delta^*$ in the bulk.

(dual to 1-cells in $\Delta$). The intersection of the dual 2-complex $\Delta^*$ with the boundaries defines two graphs $\gamma_1, \gamma_2 \in \Sigma$ (see Figure 9). For simplicity, we ignore the boundaries until the end of this section. The fields $e$ and $A$ are discretized as follows. The su(2)-valued 1-form field $e$ is represented by the assignment of $e_f \in \mathrm{su}(2)$ to each 1-cell in $\Delta$. We use the fact that faces in $\Delta^*$ are in one-to-one correspondence with 1-cells in $\Delta$ and label $e_f$ with a face subindex (Figure 9). The connection field $A$ is represented by the assignment of group elements $g_e \in \mathrm{SU}(2)$ to each edge in $e \in \Delta^*$ (see Figure 10).

With all this, [23] becomes the regularized version $P_\Delta$ defined as

$$P_\Delta = \int \prod_{f \in \Delta^*} \mathrm{d}e_f \prod_{e \in \Delta^*} \mathrm{d}g_e \exp\left[\mathrm{i}\ \mathrm{tr}\left[e_f W_f\right]\right] \qquad [24]$$

where $\mathrm{d}e_f$ is the regular Lebesgue measure on $\mathbb{R}^3$, $\mathrm{d}g_e$ is the Haar measure on SU(2), and $W_f$ denotes the holonomy around (spacetime) faces, that is, $W_f = g_e^1 \cdots g_e^N$ for $N$ being the number of edges bounding the corresponding face (see Figure 10). The discretization procedure is reminiscent of the one used in standard lattice gauge theory (*see* Lattice

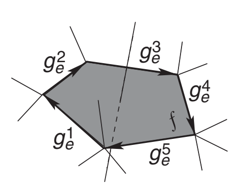



**Figure 10** A (2-cell) face $f \in \Delta^*$ in a cellular decomposition of the spacetime manifold $M$ and the corresponding dual 1-cell. The connection field is discretized by the assignment of the parallel transport group elements $g_e^i \in \mathrm{SU}(2)$ to edges $e \in \Delta^*$ ($i = 1, \ldots, 5$ in the face shown here).

Gauge Theory). The previous definition can be motivated by an analysis equivalent to the one presented in [16].

Integrating over $e_f$, and using [19], one obtains

$$P_\Delta = \sum_{\{j\}} \int \prod_{e\in\Delta^*} \mathrm{d}g_e \prod_{f\in\Delta^*} (2j_f+1) \times \mathrm{tr}\left[\overset{j_f}{\Pi}(g_e^1 \ldots g_e^N)\right] \qquad [25]$$

Now it remains to integrate over the lattice connection $\{g_e\}$. If an edge $e \in \Delta^*$ bounds $n$ faces $f \in \Delta^*$ there will be $n$ traces of the form $\mathrm{tr}[\Pi^{j_f}(\cdots g_e \cdots)]$ in [25] containing $g_e$ in the argument. In order to integrate over $g_e$ we can use the following identity:

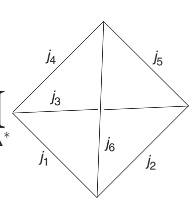
$$I^n_{\mathrm{inv}} := \int \mathrm{d}g \overset{j_1}{\Pi}(g) \otimes \overset{j_2}{\Pi}(g) \otimes \cdots \otimes \overset{j_n}{\Pi}(g) = \sum_\iota C^\iota_{j_1 j_2 \cdots j_n} C^{*\iota}_{j_1 j_2 \cdots j_n} \qquad [26]$$

where $I^n_{\mathrm{inv}}$ is the projector from the tensor product of irreducible representations $\mathcal{H}_{j_1\cdots j_n} = j_1 \otimes j_2 \otimes \cdots \otimes j_n$ onto the invariant component $\mathcal{H}^0_{j_1\cdots j_n} = \mathrm{Inv}[j_1 \otimes j_2 \otimes \cdots \otimes j_n]$. On the right-hand side, we have chosen an orthonormal basis of invariant vectors (intertwiners) in $\mathcal{H}_{j_1\cdots j_n}$ to express the projector. Notice that the assignment of intertwiners to edges is a consequence of the integration over the connection. Using [26] one can write $P_\Delta$ in the general spin foam representation form [4]

$$P_\Delta = \sum_{\{f\}} \prod_{f\in\Delta^*} (2j_f+1) \prod_{v\in\Delta^*} A_v(j_v) \qquad [27]$$

where $A_v(\iota_v, j_v)$ is given by the appropriate trace of the intertwiners corresponding to the edges bounded by the vertex. As in the previous section, this amplitude is given in general by an SU(2) $3Nj$-symbol. When $\Delta$ is a simplicial complex, all the edges in $\Delta^*$ are 3-valent and vertices are 4-valent. Consequently, the vertex amplitude is given by the contraction of the corresponding four 3-valent intertwiners, that is, a $6j$ symbol. In that case, the path integral takes the (Ponzano–Regge) form

$$P_\Delta = \sum_{\{j\}} \prod_{f\in\Delta^*} (2j_f+1) \prod_{v\in\Delta^*} \qquad [28]$$




The labeling of faces that intersect the boundary naturally induces a labeling of the edges of the graphs $\gamma_1$ and $\gamma_2$ induced by the discretization. Thus, the boundary states are given by spin network states on $\gamma_1$ and $\gamma_2$, respectively. A careful analysis of the boundary contribution shows that only the face amplitude is modified to $(\Delta_{j_\ell})^{\nu_f/2}$, and that the spin foam amplitudes are as in eqn [22].

A crucial property of the path integral in 3D gravity (and of the transition amplitudes in general) is that it does not depend on the discretization $\Delta$ – this is due to the absence of local degrees of freedom in 3D gravity and not expected to hold in 4D. Given two different cellular decompositions $\Delta$ and $\Delta'$, one has

$$\tau^{-n_0} P_\Delta = \tau^{-n'_0} P_{\Delta'} \qquad [29]$$

where $n_0$ is the number of 0-simplexes in $\Delta$, and $\tau = \sum_j (2j+1)^2$. As $\tau$ is given by a divergent sum, the discretization independence statement is formal. Moreover, the sum over spins in [28] is typically divergent. Divergences occur due to infinite gauge-volume factors in the path integral corresponding to the topological gauge freedom [7]. Freidel and Louapre have shown how these divergences can be avoided by gauge-fixing unphysical degrees of freedom in [24]. In the case of 3D gravity with positive cosmological constant, the state sum generalizes to the Turaev–Viro invariant (*see* Topological Quantum Field Theory: Overview) defined in terms of the quantum group $\mathrm{SU}_q(2)$ with $q^n = 1$ where the representations are finitely many and thus $\tau < \infty$. Equation [29] is a rigorous statement in that case. No such infrared divergences appear in the canonical treatment of the previous section.

## Spin Foams in 4D

### Spin Foam from the Canonical Formulation

There is no rigorous construction of the physical inner product of LQG in 4D. The spin foam representation as a device for its definition has been introduced formally by Rovelli. In 4D LQG, difficulties in understanding dynamics are centered around the quantum scalar constraint $\widehat{S} = \widehat{\sqrt{\det E}^{-1} E^a_i E^b_j F^{ij}_{ab}(A)} + \cdots$ (see [1]) – the vector constraint $\widehat{V}_a(A, E)$ is solved in a simple manner (*see* Loop Quantum Gravity). The physical inner product formally becomes

$$\langle Ps, s' \rangle_{\mathrm{diff}} = \prod_x \delta[\widehat{S}(x)] = \int D[N] < \exp\left[\mathrm{i}\int_\Sigma N(x)\widehat{S}(x)\right] s, s' >_{\mathrm{diff}} = \int D[N] \sum_{n=0}^{\infty} \frac{i^n}{n!} < \left[\int_\Sigma N(x)\widehat{S}(x)\right]^n s, s' >_{\mathrm{diff}} \qquad [30]$$

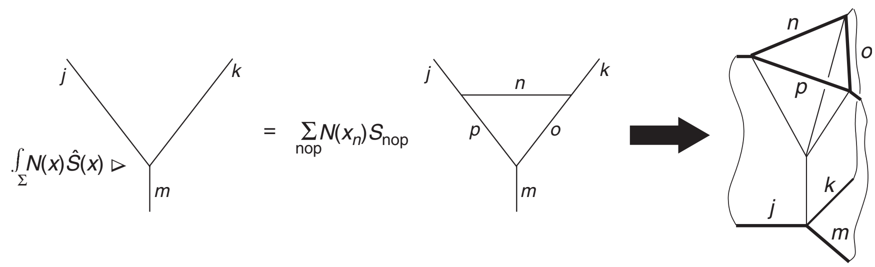



**Figure 11** The action of the scalar constraint and its spin foam representation. $N(x_n)$ is the value of $N$ at the node and $S_{nop}$ are the matrix elements of $\widehat{S}$.

where $< , >_{\text{diff}}$ denotes the inner product in the Hilbert space of solutions of the vector constraint, and the exponential has been expanded in powers in the last expression on the right-hand side.

From early on, it was realized that smooth loop states are naturally annihilated by $\widehat{S}$ (independently of any quantization ambiguity). Consequently, $\widehat{S}$ acts only on spin network nodes. Generically, it does so by creating new links and nodes modifying the underlying graph of the spin network states (**Figure 11**).

Therefore, each term in the sum [30] represents a series of transitions – given by the local action of $\widehat{S}$ at spin network nodes – through different spin network states interpolating the boundary states $s$ and $s'$, respectively. The action of $\widehat{S}$ can be visualized as an "interaction vertex" in the "time" evolution of the node (**Figure 11**). As in the explicit 3D case, eqn [30] can be expressed as sum over "histories" of spin networks pictured as a system of branching surfaces described by a 2-complex whose elements inherit the representation labels on the intermediate states. The value of the "transition" amplitudes is controlled by the matrix elements of $\widehat{S}$. Therefore, although the qualitative picture is independent of quantization ambiguities, transition amplitudes are sensitive to them.

Before even considering the issue of convergence of [30], the problem with this definition is evident: every single term in the sum is a divergent integral! Therefore, this way of presenting spin foams has to be considered as formal until a well-defined regularization of [2] is provided. That is the goal of the spin foam approach.

Instead of dealing with an infinite number of constraints Thiemann recently proposed to impose one single master constraint defined as

$$M = \int_\Sigma \mathrm{d}x^3 \, \frac{S^2(x) - q^{ab} V_a(x) V_b(x)}{\sqrt{\det q(x)}} \qquad [31]$$

Using techniques developed by Thiemann, this constraint can indeed be promoted to a quantum operator acting on $\mathcal{H}_{\text{kin}}$. The physical inner product is given by

$$<s, s'>_{\text{p}} := \lim_{T\to\infty} <s, \int_{-T}^{T} \mathrm{d}t \; \mathrm{e}^{\mathrm{i}t\widehat{M}} s'> \qquad [32]$$

A spin foam representation of the previous expression could now be achieved by the standard skeletonization that leads to the path-integral representation in quantum mechanics. In this context, one splits the $t$-parameter in discrete steps and writes

$$\mathrm{e}^{\mathrm{i}t\widehat{M}} = \lim_{N\to\infty} [\mathrm{e}^{\mathrm{i}t\widehat{M}/N}]^N = \lim_{N\to\infty} [1 + \mathrm{i}t\widehat{M}/N]^N \qquad [33]$$

The spin foam representation follows from the fact that the action of the basic operator $1 + \mathrm{i}t\widehat{M}/N$ on a spin network can be written as a linear combination of new spin networks whose graphs and labels have been modified by the creation of new nodes (in a way qualitatively analogous to the local action shown in **Figure 11**). An explicit derivation of the physical inner product of 4D LQG along these lines is under current investigation.

### Spin Foams from the Covariant Formulation

In 4D, the spin foam representation of the dynamics of LQG has been investigated more intensively in the covariant formulation. This has led to a series of constructions which are referred to as spin foam models. These treatments are related more closely to the construction based on the covariant path-integral approach of the last section. Here we illustrate the formulation which has captured much interest in the literature: the Barrett–Crane (BC) model.

**Spin foam models for gravity as constrained quantum BF theory** The BC model is one of the most extensively studied spin foam models for quantum gravity. To introduce the main ideas involved, we concentrate on the definition of the model in the Riemannian sector. The BC model can be formally

viewed as a spin foam quantization of SO(4) Plebanski's formulation of general relativity. Plebanski's Riemannian action depends on an SO(4) connection $A$, a Lie-algebra-valued 2-form $B$, and Lagrange multiplier fields $\lambda$ and $\mu$. Writing explicitly the Lie algebra indices, the action is given by

$$S[B, A, \lambda, \mu] = \int \left[ B^{IJ} \wedge F_{IJ}(A) + \lambda_{IJKL}\, B^{IJ} \wedge B^{KL} + \mu \epsilon^{IJKL} \lambda_{IJKL} \right] \quad [34]$$

where $\mu$ is a 4-form and $\lambda_{IJKL} = -\lambda_{JIKL} = -\lambda_{IJLK} = \lambda_{KLIJ}$ is a tensor in the internal space. Variation with respect to $\mu$ imposes the constraint $\epsilon^{IJKL}\lambda_{IJKL} = 0$ on $\lambda_{IJKL}$. The Lagrange multiplier tensor $\lambda_{IJKL}$ has then 20 independent components. Variation with respect to $\lambda$ imposes 20 algebraic equations on the 36 components of $B$. The (nondegenerate) solutions to the equations obtained by varying the multipliers $\lambda$ and $\mu$ are

$$B^{IJ} = \pm \epsilon^{IJKL} e_K \wedge e_L$$

and

$$B^{IJ} = \pm e^I \wedge e^J \quad [35]$$

in terms of the 16 remaining degrees of freedom of the tetrad field $e^I_a$. If one substitutes the first solution into the original action, one obtains Palatini's formulation of general relativity; therefore, on shell (and on the right sector), the action is that of classical gravity.

The key idea in the definition of the model is that the path integral for the theory corresponding to the action $S[B, A, 0, 0]$, namely

$$P_{\text{topo}} = \int D[B]D[A] \exp\left[ \mathrm{i} \int \left[ B^{IJ} \wedge F_{IJ}(A) \right] \right] \quad [36]$$

can be given a meaning as a spin foam sum, [4], in terms of a simple generalization of the construction of the previous section. In fact, $S[B, A, 0, 0]$ corresponds to a simple theory known as BF theory that is formally very similar to 3D gravity (*see* BF Theories). The result is independent of the chosen discretization because BF theory does not have local degrees of freedom (just as 3D gravity).

The BC model aims at providing a definition of the path integral of gravity pursuing a well-posed definition of the formal expression

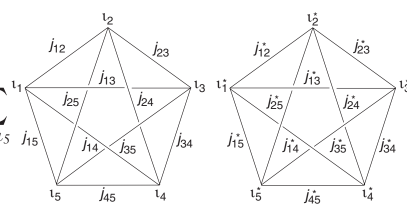
$$P_{\text{GR}} = \int D[B]D[A] \delta\left[ B \to \epsilon^{IJKL} e_K \wedge e_L \right] \times \exp\left[ \mathrm{i} \int \left[ B^{IJ} \wedge F_{IJ}(A) \right] \right] \quad [37]$$

where $D[B]D[A]\delta(B \to \epsilon^{IJKL} e_K \wedge e_L)$ means that one must restrict the sum in [36] to those configurations of the topological theory satisfying the constraints $B = *(e \wedge e)$ for some tetrad $e$. The remarkable fact is that this restriction can be implemented in a systematic way directly on the spin foam configurations that define $P_{\text{topo}}$.

In $P_{\text{topo}}$ spin foams are labeled with spins corresponding to the unitary irreducible representations of SO(4) (given by two spin quantum numbers $(j_{\text{R}}, j_{\text{L}})$). Essentially, the factor "$\delta(B \to \epsilon^{IJKL} e_K \wedge e_L)$" restricts the set of spin foam quantum numbers to the so-called simple representations (for which $j_{\text{R}} = j_{\text{L}} = j$). This is the "quantum" version of the solution to the constraints [35]. There are various versions of this model. The simplest definition of the transition amplitudes in the BC model is given by

$$P(s^* s) = \sum_{\{j\}} \prod_{f \in F_{s \to s'}} (2j_f + 1)^{\nu_f} \prod_{v \in F_{s \to s'}}$$

$$\sum_{\iota_1 \cdots \iota_5} \quad [38]$$




where we use the notation of [22], the graphs denote $15j$-symbols, and $\iota_i$ are half-integers labeling SU(2) normalized 4-intertwiners. No rigorous connection with the Hilbert space picture of LQG has yet been established. The self-dual version of Plebanski's action leads, through a similar construction, to Reisenberger's model.

The simplest amplitude in the BC model corresponds to a single 4-simplex, which can be viewed as the simplest triangulation of the 4D spacetime given by the interior of a 3-sphere (the corresponding 2-complex is shown in Figure 12). States of the 4-simplex are labeled by ten spins $j$ (labeling the ten edges of the boundary spin network, see Figure 12) which can be shown to be related to the area in

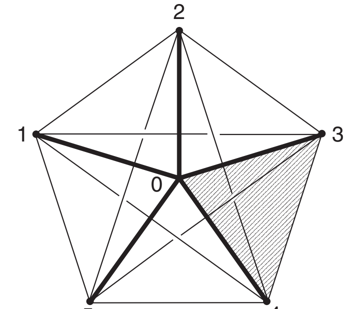



**Figure 12** The dual of a 4-simplex.

Planck units of the ten triangular faces that form the 4-simplex. A first indication of the connection of the model with gravity was that the large-$j$ asymptotics appeared to be dominated by the exponential of the Regge action (the action derived by Regge as a discretization of general relativity). This estimate was done using the stationary-phase approximation to the integral that gives the amplitude of a 4-simplex in the BC model. However, more detailed calculations showed that the amplitude is dominated by configurations corresponding to degenerate 4-simplexes. This seems to invalidate a simple connection to general relativity and is one of the main puzzles in the model.

### Spin Foams as Feynman Diagrams

The main problem with the models of the previous section is that they are defined on a discretization $\Delta$ of $M$ and that – contrary to what happens with a topological theory, for example, 3D gravity (eqn [29]) – the amplitudes depend on the discretization $\Delta$. Various possibilities to eliminate this regulator have been discussed in the literature but no explicit results are yet known in 4D. An interesting proposal is a discretization-independent definition of spin foam models achieved by the introduction of an auxiliary field theory living on an abstract group manifold – $\mathrm{Spin}(4)^4$ and $\mathrm{SL}(2,\mathbb{C})^4$ for Riemannian and Lorentzian gravity, respectively. The action of the auxiliary group field theory (GFT) takes the form

$$S[\phi] = \int_{G^4} \phi^2 + \frac{\lambda}{5!}\int_{G^{10}} M^{(5)}[\phi] \qquad [39]$$

where $M^{(5)}[\phi]$ is a fifth-order monomial, and $G$ is the corresponding group. In the simplest model, $M^{(5)}[\phi] = \phi(g_1, g_2, g_3, g_4)\phi(g_4, g_5, g_6, g_7)\times \phi(g_7, g_3, g_8, g_9)\phi(g_9, g_6, g_2, g_{10})\phi(g_{10}, g_8, g_5, g_1)$. The field $\phi$ is required to be invariant under the (simultaneous) right action of the group on its four arguments in addition to other symmetries (not described here for simplicity). The perturbative expansion in $\lambda$ of the GFT Euclidean path integral is given by

$$P = \int D[\phi]\mathrm{e}^{-S[\phi]} = \sum_{F_N} \frac{\lambda^N}{\mathrm{sym}[F_N]} A[F_N] \qquad [40]$$

where $A[F_N]$ corresponds to a sum of Feynman-diagram amplitudes for diagrams with $N$ interaction vertices, and $\mathrm{sym}[F_N]$ denotes the standard symmetry factor. A remarkable property of this expansion is that $A[F_N]$ can be expressed as a sum over spin foam amplitudes, that is, 2-complexes labeled by unitary irreducible representations of $G$. Moreover, for very simple interaction $M^{(5)}[\phi]$, the spin foam amplitudes are in one-to-one correspondence to those found in the models of the previous section (e.g., the BC model). This duality is regarded as a way of providing a fully combinatorial definition of quantum gravity where no reference to any discretization or even a manifold structure is made. Transition amplitudes between spin network states correspond to $n$-point functions of the field theory. These models have been inspired by generalizations of matrix models applied to BF theory.

Divergent transition amplitudes can arise by the contribution of "loop" diagrams as in standard quantum field theory. In spin foams, diagrams corresponding to 2D bubbles are potentially divergent because spin labels can be arbitrarily high leading to unbounded sums in [4]. Such divergences do not occur in certain field theories dual (in the sense above) to the BC model. However, little is known about the convergence of the series in $\lambda$ and the physical meaning of this constant. Nevertheless, Freidel and Louapre have shown that the series can be re-summed in certain models dual to lower-dimensional theories.

### Causal Spin Foams

Let us conclude by presenting a fundamentally different construction leading to spin foams. Using the kinematical setting of LQG with the assumption of the existence of a microlocal (in the sense of Planck scale) causal structure, Markopoulou and Smolin define a general class of (causal) spin foam models for gravity. The elementary transition amplitude $A_{s_I \to s_{I+1}}$ from an initial spin network $s_I$ to another spin network $s_{I+1}$ is defined by a set of simple combinatorial rules based on a definition of causal propagation of the information at nodes. The rules and amplitudes have to satisfy certain causal restrictions (motivated by the standard concepts in classical Lorentzian physics). These rules generate surface-like excitations of the same kind one encounters in the previous formulations. Spin foams $F^N_{s_i \to s_f}$ are labeled by the number of times, $N$, these elementary transitions take place. Transition amplitudes are defined as

$$\langle s_i, s_f\rangle = \sum_N A(F^N_{s_i \to s_f}) \qquad [41]$$

which is of the generic form [4]. The models are not related to any continuum action. The only guiding principles in the construction are the restrictions imposed by causality, and the requirement of the existence of a nontrivial critical behavior that reproduces general relativity at large scales. Some indirect evidence of a possible nontrivial continuum limit has been obtained in certain versions of these models in $1 + 1$ dimensions.

*See also:* Algebraic Approach to Quantum Field Theory; BF Theories; Canonical General Relativity; Chern–Simons Models: Rigorous Results; Lattice Gauge Theory; Loop Quantum Gravity; Quantum Dynamics in Loop Quantum Gravity; Quantum Geometry and its Applications; Topological Quantum Field Theory: Overview.

## Further Reading

Ashtekar A (1991) *Lectures on Nonperturbative Canonical Gravity*. Singapore: World Scientific.

Ashtekar A and Lewandowski J (2004) Background independent quantum gravity: a status report.

Baez J (1998) Spin foams. *Classical and Quantum Gravity* 15: 1827–1858.

Baez J (2000) An introduction to spin foam models of quantum gravity and BF theory. *Lecture Notes in Physics* 543: 25–94.

Baez J and Muniain JP (1995) *Gauge fields, Knots and Gravity*. Singapore: World Scientific.

Oriti D (2001) Spacetime geometry from algebra: spin foam models for nonperturbative quantum gravity. *Reports on Progress in Physics* 64: 1489–1544.

Perez A (2003) Spin foam models for quantum gravity. *Classical and Quantum Gravity* 20: R43.

Rovelli C *Quantum Gravity*. Cambridge: Cambridge University Press (to appear).

Thiemann T *Modern Canonical Quantum General Relativity*. Cambridge: Cambridge University Press (to appear).

# Spin Glasses

**F Guerra**, Università di Roma "La Sapienza", Rome, Italy



## Introduction

From a physical point of view, spin glasses, as dilute magnetic alloys, are very interesting systems. They are characterized by such features as exhibiting a new magnetic phase, where magnetic moments are frozen into disordered equilibrium orientations, without any long-range order. See, for example, Young (1987) for general reviews, and also Stein (1989) for a very readable account about the physical properties of spin glasses. The experimental laboratory study of spin glasses is a very difficult subject, because of their peculiar properties. In particular, the existence of very slowly relaxing modes, with consequent memory effects, makes it difficult to realize the very basic physical concept of a system at thermodynamical equilibrium, at a given temperature.

From a theoretical point of view some models have been proposed, which try to capture the essential physical features of spin glasses, in the frame of very simple assumptions.

The basic model has been proposed by Edwards and Anderson (1975) many years ago. It is a simple extension of the well-known nearest-neighbor Ising model. On a large region $\Lambda$ of the unit lattice in $d$ dimensions, we associate an Ising spin $\sigma(n)$ to each lattice site $n$, and then we introduce a lattice Hamiltonian

$$H_\Lambda(\sigma, J) = -\sum_{(n,n')} J(n,n')\sigma(n)\sigma(n') \qquad [1]$$

Here, the sum runs over all couples of nearest-neighbor sites in $\Lambda$, and $J$ are quenched random couplings, assumed for simplicity to be independent identically distributed random variables, with centered unit Gaussian distribution. The quenched character of the $J$ means that they do not contribute to thermodynamic equilibrium, but act as a kind of random external noise on the coupling of the $\sigma$ variables. In the expression of the Hamiltonian, we have indicated with $\sigma$ the set of all $\sigma(n)$, and with $J$ the set of all $J(n,n')$. The region $\Lambda$ must be taken very large, by letting it invade all lattice in the limit. The physical motivation for this choice is that for real spin glasses the interaction between the spins dissolved in the matrix of the alloy oscillates in sign according to distance. This effect is taken into account in the model through the random character of the couplings between spins.

Even though very drastic simplifications have been introduced in the formulation of this model, as compared to the extremely complicated nature of physical spin glasses, nevertheless a rigorous study of all properties emerging from the static and dynamic behavior of a thermodynamic system of this kind is far from being complete. In particular, with reference to static equilibrium properties, it is not yet possible to reach a completely substantiated description of the phases emerging in the low-temperature region. Even physical intuition gives completely different guesses for different people.

In the same way as a mean-field version can be associated to the ordinary Ising model, so it is possible for the disordered model described by [1]. Now we consider a number of sites $i = 1, 2, \ldots, N$, and let each spin $\sigma(i)$ at site $i$ interact with all other spins, with the intervention of a quenched noise $J_{ij}$. The precise form of the Hamiltonian will be given in the following.

This is the mean-field model for spin glasses, introduced by Sherrington and Kirkpatrick (1975).

It is a celebrated model. Numerous articles have been devoted to its study during the years, appearing in the theoretical physics literature.

The relevance of the model stems surely from the fact that it is intended to represent some important features of the physical spin glass systems, of great interest for their peculiar properties, at least at the level of the mean-field approximation.

But another important source of interest is connected with the fact that disordered systems, of the Sherrington–Kirkpatrick type, and their generalizations, seem to play a very important role for theoretical and practical assessments about hard optimization problems, as it is shown, for example, by Mézard *et al.* (2002).

It is interesting to remark that the original paper was entitled "Solvable model of a spin-glass," while a previous draft, as told by David Sherrington, contained the even stronger designation "Exactly solvable." However, it turned out that the very natural solution devised by the authors is valid only at high temperatures, or for large external magnetic fields. At low temperatures, the proposed solution exhibits a nonphysical drawback given by a negative entropy, as properly recognized by the authors in their very first paper.

It took some years to find an acceptable solution. This was done by Giorgio Parisi in a series of papers, marking a radical departure from the previous methods. In fact, a very intense method of "spontaneous replica symmetry breaking" was developed. As a consequence, the physical content of the theory was encoded in a functional order parameter of new type, and a remarkable structure emerged for the pure states of the theory, a kind of hierarchical, ultrametric organization. These very interesting developments, due to Parisi, and his coworkers, are explained in a brilliant way in the classical book by Mézard *et al.* (1987). Part of this structure will be recalled in the following.

It is important to remark that the Parisi solution is presented in the form of an ingenious and clever "ansatz." Until few years ago, it was not known whether this ansatz would give the true solution for the model, in the so-called thermodynamic limit, when the size of the system becomes infinite, or it would be only a very good approximation for the true solution.

The general structures offered by the Parisi solution, and their possible generalizations for similar models, exhibit an extremely rich and interesting mathematical content. Very appropriately, Talagrand (2003) has used a strongly suggestive sentence in the title to his recent book: "Spin glasses: a challenge for mathematicians."

As a matter of fact, how to face this challenge is a very difficult problem. Here we would like to recall the main features of a very powerful method, yet extremely simple in its very essence, based on a comparison and interpolation argument on sets of Gaussian random variables.

The method found its first simple application in Guerra (2001), where it was shown that the Sherrington–Kirkpatrick replica symmetric approximate solution was a rigorous lower bound for the quenched free energy of the system, uniformly in the size. Then, it was possible to reach a long-awaited result (Guerra and Toninelli 2002): the convergence of the free energy density in the thermodynamic limit, by an intermediate step where the quenched free energy was shown to be subadditive in the size of the system.

Moreover, still by interpolation on families of Gaussian random variables, the first mentioned result was extended to give a rigorous proof that the expression given by the Parisi ansatz is also a lower bound for the quenched free energy of the system, uniformly in the size (Guerra 2003). The method gives not only the bound, but also the explicit form of the correction in a complex form. As a recent and very important result, along the task of facing the challenge, Michel Talagrand has been able to dominate these correction terms, showing that they vanish in the thermodynamic limit. This milestone achievement was first announced in a short note, containing only a synthetic sketch of the proof, and then presented with all details in a long paper (Talagrand 2006).

The interpolation method is also at the basis of the far-reaching generalized variational principle proved by Aizenman *et al.* (2003).

In our presentation, we will try to be as self-contained as possible. We will give all definitions, explain the basic structure of the interpolation method, and show how some of the results are obtained. We will concentrate mostly on questions connected with the free energy, its properties of subadditivity, the existence of the infinite-volume limit, and the replica bounds.

For the sake of comparison, and in order to provide a kind of warm-up, we will recall also some features of the standard elementary mean-field model of ferromagnetism, the so-called Curie–Weiss model. We will concentrate also here on the free energy, and systematically exploit elementary comparison and interpolation arguments. This will show the strict analogy between the treatment of the ferromagnetic model and the developments in the mean-field spin glass case. Basic roles will be played in the two cases, but with different expressions, by positivity and convexity properties.

Then, we will consider the problem of connecting results for the mean-field case to the short-range case. An intermediate position is occupied by the so-called diluted models. They can be studied through a generalization of the methods exploited in the mean-field case, as shown, for example, in De Sanctis (2005).

The organization of the paper is as follows. We first introduce the ferromagnetic model and discuss behavior and properties of the free energy in the thermodynamic limit, by emphasizing, in this very elementary case, the comparison and interpolation methods that will be also exploited, in a different context, in the spin glass case.

The basic features of the mean-field spin glass models are discussed next, by introducing all necessary definitions. This is followed by the introduction, for generic Gaussian interactions, of some important formulas, concerning the derivation with respect to the strength of the interaction, and the Gaussian comparison and interpolation method.

We then give simple applications to the mean-field spin glass model, in particular to the existence of the infinite-volume limit of the quenched free energy (Guerra and Toninelli 2002), and to the proof of general variational bounds, by following the useful strategy developed in Aizenman *et al.* (2003).

The main features of the Parisi representation are recalled briefly, and the main theorem concerning the free energy is stated. This is followed by a brief mention of results for diluted models.

We also attack the problem of connecting the results for the mean-field case to the more realistic short-range models.

Finally we provide conclusions and outlook for future foreseen developments.

Our treatment will be as simple as possible, by relying on the basic structural properties, and by describing methods of presumably very long lasting power. The emphasis given to the mean-field case reflects the status of research. After some years from now this review would perhaps be written according to completely different patterns.

## A Warm-up. The Mean-field Ferromagnetic Model: Structure and Results

The mean-field ferromagnetic model is among the simplest models of statistical mechanics. However, it contains very interesting features, in particular a phase transition, characterized by spontaneous magnetization, at low temperatures. We refer to standard textbooks for a full treatment and a complete appreciation of the model in the frame of the theory of ferromagnetism. Here we first consider some properties of the free energy, easily obtained through comparison methods.

The generic configuration of the mean-field ferromagnetic model is defined through Ising spin variables $\sigma_i = \pm 1$, attached to each site $i = 1, 2, \ldots, N$.

The Hamiltonian of the model, in some external field of strength $h$, is given by the mean-field expression

$$H_N(\sigma, h) = -\frac{1}{N}\sum_{(i,j)} \sigma_i\sigma_j - h\sum_i \sigma_i \qquad [2]$$

Here, the first sum extends to all $N(N-1)/2$ site couples, and the second to all sites.

For a given inverse temperature $\beta$, let us now introduce the partition function $Z_N(\beta, h)$ and the free energy per site $f_N(\beta, h)$, according to the well-known definitions

$$Z_N(\beta, h) = \sum_{\sigma_1 \ldots \sigma_N} \exp(-\beta H_N(\sigma, h)) \qquad [3]$$

$$-\beta f_N(\beta, h) = N^{-1} E \log Z_N(\beta, h) \qquad [4]$$

It is also convenient to define the average spin magnetization

$$m = \frac{1}{N}\sum_i \sigma_i \qquad [5]$$

Then, it is immediately seen that the Hamiltonian in [2] can be equivalently written as

$$H_N(\sigma, h) = -\frac{1}{2}Nm^2 - h\sum_i \sigma_i \qquad [6]$$

where an unessential constant term has been neglected. In fact, we have

$$\sum_{(i,j)} \sigma_i\sigma_j = \frac{1}{2}\sum_{i,j;i\neq j} \sigma_i\sigma_j = \frac{1}{2}N^2m^2 - \frac{1}{2}N \qquad [7]$$

where the sum over all couples has been equivalently written as one half the sum over all $i$, $j$ with $i \neq j$, and the diagonal terms with $i = j$ have been added and subtracted out. Notice that they give a constant because $\sigma_i^2 = 1$.

Therefore, the partition function in [3] can be equivalently substituted by the expression

$$Z_N(\beta, h) = \sum_{\sigma_1 \ldots \sigma_N} \exp\left(\frac{1}{2}\beta N m^2\right) \exp\left(\beta h \sum_i \sigma_i\right) \qquad [8]$$

which will be our starting point.

Our interest will be in the $\lim_{N\to\infty} N^{-1} \log Z_N(\beta, h)$. To this purpose, let us establish the important subadditivity property, holding for the splitting of the

large-$N$ system in two smaller systems with $N_1$ and $N_2$ sites, respectively, with $N = N_1 + N_2$,

$$\log Z_N(\beta, h) \leq \log Z_{N_1}(\beta, h) + \log Z_{N_2}(\beta, h) \qquad [9]$$

The proof is very simple. Let us denote, in the most natural way, by $\sigma_1, \ldots, \sigma_{N_1}$ the spin variables for the first subsystem, and by $\sigma_{N_1+1}, \ldots, \sigma_N$ the $N_2$ spin variables of the second subsystem. Introduce also the subsystem magnetizations $m_1$ and $m_2$, by adapting the definition [5] to the smaller systems, in such a way that

$$Nm = N_1 m_1 + N_2 m_2 \qquad [10]$$

Therefore, we see that the large system magnetization $m$ is the linear convex combination of the smaller system ones, according to the obvious

$$m = \frac{N_1}{N} m_1 + \frac{N_2}{N} m_2 \qquad [11]$$

Since the mapping $m \to m^2$ is convex, we also have the general bound, holding for all values of the $\sigma$ variables

$$m^2 \leq \frac{N_1}{N} m_1^2 + \frac{N_2}{N} m_2^2 \qquad [12]$$

Then, it is enough to substitute the inequality in the definition [8] of $Z_N(\beta, h)$, and recognize that we achieve factorization with respect to the two subsystems, and therefore the inequality $Z_N \leq Z_{N_1} Z_{N_2}$. So we have established [9]. From subadditivity, the existence of the limit follows by standard arguments. In fact, we have

$$\lim_{N\to\infty} N^{-1} \log Z_N(\beta, h) = \inf_N N^{-1} \log Z_N(\beta, h) \qquad [13]$$

Now we will calculate explicitly this limit, by introducing an order parameter $M$, a trial function, and an appropriate variational scheme. In order to get a lower bound, we start from the elementary inequality $m^2 \geq 2mM - M^2$, holding for any value of $m$ and $M$. By inserting the inequality in the definition [8] we arrive at a factorization of the sum over $\sigma$'s. The sum can be explicitly calculated, and we arrive immediately to the lower bound, uniform in the size of the system,

$$N^{-1} \log Z_N(\beta, h) \geq \log 2 + \log\cosh \beta(h + M) - \tfrac{1}{2}\beta M^2 \qquad [14]$$

holding for any value of the trial order parameter $M$. Clearly, it is convenient to take the supremum over $M$. Then, we establish the optimal uniform lower bound

$$N^{-1} \log Z_N(\beta, h) \geq \sup_M \left(\log 2 + \log\cosh \beta(h + M) - \tfrac{1}{2}\beta M^2\right) \qquad [15]$$

It is simple to realize that the supremum coincides with the limit as $N \to \infty$. To this purpose we follow the following simple procedure. Let us consider all possible values of the variable $m$. There are $N + 1$ of them, corresponding to any number $K$ of possible spin flips, starting from a given $\sigma$ configuration, $K = 0, 1, \ldots, N$. Let us consider the trivial decomposition of the identity, holding for any $m$,

$$1 = \sum_M \delta_{mM} \qquad [16]$$

where $M$ in the sum runs over the $N + 1$ possible values of $m$, and $\delta$ is Kroneker delta, being equal to 1 if $M = N$, and zero otherwise. Let us now insert [16] in the definition [8] of the partition function inside the sum over $\sigma$'s, and invert the two sums. Because of the forcing $m = M$ given by the $\delta$, we can write $m^2 = 2mM - M^2$ inside the sum. Then if we neglect the $\delta$, by using the trivial $\delta \leq 1$, we have an upper bound, where the sum over $\sigma$'s can be explicitly performed as before. Then it is enough to take the upper bound with respect to $M$, and consider that there are $N + 1$ terms in the now trivial sum over $M$, in order to arrive at the upper bound

$$N^{-1} \log Z_N(\beta, h) \leq \sup_M \left(\log 2 + \log\cosh \beta(h + M) - \tfrac{1}{2}\beta M^2\right) + N^{-1} \log(N + 1) \qquad [17]$$

Therefore, by going to the limit as $N \to \infty$, we can collect all our results in the form of the following theorem giving the full characterization of the thermodynamic limit of the free energy.

**Theorem 1** *For the mean-field ferromagnetic model we have*

$$\lim_{N\to\infty} N^{-1} \log Z_N(\beta, h) = \inf_N N^{-1} \log Z_N(\beta, h) \qquad [18]$$

$$= \sup_M \left(\log 2 + \log\cosh \beta(h + M) - \tfrac{1}{2}\beta M^2\right) \qquad [19]$$

This ends our discussion about the free energy in the ferromagnetic model.

Other properties of the model can be easily established. Introduce the Boltzmann–Gibbs state

$$\omega_N(A) = Z_N^{-1} \sum_{\sigma_1 \ldots \sigma_N} A \exp\left(\frac{1}{2}\beta N m^2\right) \exp\left(\beta h \sum_i \sigma_i\right) \qquad [20]$$

where $A$ is any function of $\sigma_1 \ldots \sigma_N$.

The observable $m(\sigma)$ becomes self-averaging under $\omega_N$, in the infinite-volume limit, in the sense that

$$\lim_{N\to\infty} \omega_N((m - M(\beta, h))^2) = 0 \qquad [21]$$

This property of $m$ is the deep reason for the success of the strategy exploited earlier for the convergence of the free energy. Easy consequences are the following. In the infinite-volume limit, for $h \neq 0$, the Boltzmann–Gibbs state becomes a factor state

$$\lim_{N\to\infty} \omega_N(\sigma_1 \ldots \sigma_s) = M(\beta, h)^s \qquad [22]$$

A phase transition appears in the form of spontaneous magnetization. In fact, while for $h = 0$ and $\beta \leq 1$ we have $M(\beta, h) = 0$, on the other hand, for $\beta > 1$, we have the discontinuity

$$\lim_{h\to 0^+} M(\beta, h) = -\lim_{h\to 0^-} M(\beta, h) \equiv M(\beta) > 0 \qquad [23]$$

Fluctuations can also be easily controlled. In fact, one proves that the rescaled random variable $\sqrt{N}(m - M(\beta, h))$ tends in distribution, under $\omega_N$, to a centered Gaussian with variance given by the susceptibility

$$\chi(\beta, h) \equiv \frac{\partial}{\partial h} M(\beta, h) \equiv \frac{\beta(1 - M^2)}{1 - \beta(1 - M^2)} \qquad [24]$$

Notice that the variance becomes infinite only at the critical point $h = 0, \beta = 1$, where $M = 0$.

Now we are ready to attack the much more difficult spin glass model. But it will be surprising to see that, by following a simple extension of the methods described here, we will arrive at similar results.

## Basic Definitions for the Mean-Field Spin Glass Model

As in the ferromagnetic case, the generic configuration of the mean-field spin glass model is defined through Ising spin variables $\sigma_i = \pm 1$, attached to each site $i = 1, 2, \ldots, N$.

But now there is an external quenched disorder given by the $N(N-1)/2$ independent and identical distributed random variables $J_{ij}$, defined for each pair of sites. For the sake of simplicity, we assume each $J_{ij}$ to be a centered unit Gaussian with averages $E(J_{ij}) = 0, E(J_{ij}^2) = 1$. By quenched disorder we mean that the $J$ have a kind of stochastic external influence on the system, without contributing to the thermal equilibrium.

Now the Hamiltonian of the model, in some external field of strength $h$, is given by the mean-field expression

$$H_N(\sigma, h, J) = -\frac{1}{\sqrt{N}} \sum_{(i,j)} J_{ij}\sigma_i\sigma_j - h \sum_i \sigma_i \qquad [25]$$

Here, the first sum extends to all site pairs, and the second to all sites. Notice the $\sqrt{N}$, necessary to ensure a good thermodynamic behavior to the free energy.

For a given inverse temperature $\beta$, let us now introduce the disorder-dependent partition function $Z_N(\beta, h, J)$ and the quenched average of the free energy per site $f_N(\beta, h)$, according to the definitions

$$Z_N(\beta, h, J) = \sum_{\sigma_1 \ldots \sigma_N} \exp(-\beta H_N(\sigma, h, J)) \qquad [26]$$

$$-\beta f_N(\beta, h) = N^{-1} E \log Z_N(\beta, h, J) \qquad [27]$$

Notice that in [27] the average $E$ with respect to the external noise is made "after" the log is taken. This procedure is called quenched averaging. It represents the physical idea that the external noise does not contribute to the thermal equilibrium. Only the $\sigma$'s are thermalized.

For the sake of simplicity, it is also convenient to write the partition function in the following equivalent form. First of all let us introduce a family of centered Gaussian random variables $\mathcal{K}(\sigma)$, indexed by the configurations $\sigma$, and characterized by the covariances

$$E(\mathcal{K}(\sigma)\mathcal{K}(\sigma')) = q^2(\sigma, \sigma') \qquad [28]$$

where $q(\sigma, \sigma')$ are the overlaps between two generic configurations, defined by

$$q(\sigma, \sigma') = N^{-1} \sum_i \sigma_i \sigma'_i \qquad [29]$$

with the obvious bounds $-1 \leq q(\sigma, \sigma') \leq 1$, and the normalization $q(\sigma, \sigma) = 1$. Then, starting from the definition [25], it is immediately seen that the partition function in [26] can also be written, by neglecting unessential constant terms, in the form

$$Z_N(\beta, h, J) = \sum_{\sigma_1 \ldots \sigma_N} \exp\left(\beta \sqrt{\frac{N}{2}} \mathcal{K}(\sigma)\right) \exp\left(\beta h \sum_i \sigma_i\right) \qquad [30]$$

which will be the starting point of our treatment.

## Basic Formulas of Derivation and Interpolation

We work in the following general setting. Let $U_i$ be a family of centered Gaussian random variables, $i = 1, \ldots, K$, with covariance matrix given by $E(U_i U_j) \equiv S_{ij}$. We treat the index $i$ now as configuration space for some statistical mechanics system, with partition function $Z$ and quenched free energy given by

$$E \log \sum_i w_i \exp(\sqrt{t} U_i) \equiv E \log Z \qquad [31]$$

where $w_i \geq 0$ are generic weights, and $t$ is a parameter ruling the strength of the interaction.

It would be hard to underestimate the relevance of the following derivation formula

$$\frac{\mathrm{d}}{\mathrm{d}t} E \log \sum_i w_i \exp(\sqrt{t} U_i) = \frac{1}{2} E\left( Z^{-1} \sum_i w_i \exp(\sqrt{t} U_i\right) S_{ii} - \frac{1}{2} E\left( Z^{-2} \sum_i \sum_j w_i w_j \exp(\sqrt{t} U_i) \right. \left. \times \exp(\sqrt{t} U_j) S_{ij} \right) \quad [32]$$

The proof is straightforward. First we perform directly the $t$-derivative. Then, we notice that the random variables appear in expressions of the form $E(U_i F)$, where $F$ are functions of the $U$'s. These can be easily handled through the following integration by parts formula for generic Gaussian random variables, strongly reminiscent of the Wick theorem in quantum field theory,

$$E(U_i F) = \sum_j S_{ij} E\left( \frac{\partial}{\partial U_j} F \right) \quad [33]$$

Therefore, we see that always two derivatives are involved. The two terms in [32] come from the action of the $U_j$ derivatives, the first acting on the Boltzmann factor, and giving rise to a Kronecker $\delta_{ij}$, the second acting on $Z^{-1}$, and giving rise to the minus sign and the duplication of variables.

The derivation formula can be expressed in a more compact form by introducing replicas and suitable averages. In fact, let us introduce the state $\omega$ acting on functions $F$ of $i$ as follows

$$\omega(F(i)) = Z^{-1} \sum_i w_i \exp(\sqrt{t} U_i) F(i) \quad [34]$$

together with the associated product state $\Omega$ acting on replicated configuration spaces $i_1, i_2, \ldots, i_s$. By performing also a global $E$ average, finally we define the averages

$$\langle F \rangle_t \equiv E\Omega(F) \quad [35]$$

where the subscript is introduced in order to recall the $t$ dependence of these averages.

Then, eqn [32] can be written in a more compact form

$$\frac{\mathrm{d}}{\mathrm{d}t} E \log \sum_i w_i \exp(\sqrt{t} U_i) = \tfrac{1}{2} \langle S_{i_1 i_1} \rangle - \tfrac{1}{2} \langle S_{i_1 i_2} \rangle \quad [36]$$

Our basic comparison argument will be based on the following very simple theorem.

**Theorem 2** *Let $U_i$ and $\hat{U}_i$, for $i = 1, \ldots, K$, be independent families of centered Gaussian random variables, whose covariances satisfy the inequalities for generic configurations*

$$E(U_i U_j) \equiv S_{ij} \geq E(\hat{U}_i \hat{U}_j) \equiv \hat{S}_{ij} \quad [37]$$

*and the equalities along the diagonal*

$$E(U_i U_i) \equiv S_{ii} = E(\hat{U}_i \hat{U}_i) \equiv \hat{S}_{ii} \quad [38]$$

*then for the quenched averages we have the inequality in the opposite sense*

$$E \log \sum_i w_i \exp(U_i) \leq E \log \sum_i w_i \exp(\hat{U}_i) \quad [39]$$

*where the $w_i \geq 0$ are the same in the two expressions.*

Considerations of this kind are present in the mathematical literature, as mentioned, for example, in Talagrand (2003).

The proof is extremely simple and amounts to a straightforward calculation. In fact, let us consider the interpolating expression

$$E \log \sum_i w_i \exp(\sqrt{t} U_i + \sqrt{1-t} \hat{U}_i) \quad [40]$$

where $0 \leq t \leq 1$. Clearly, the two expressions under comparison correspond to the values $t = 0$ and $t = 1$, respectively. By taking the derivative with respect to $t$, with the help of the previous derivation formula, we arrive at the evaluation of the $t$ derivative in the form

$$\frac{\mathrm{d}}{\mathrm{d}t} E \log \sum_i w_i \exp(\sqrt{t} U_i + \sqrt{1-t} \hat{U}_i) = \frac{1}{2} E\left( Z^{-1} \sum_i w_i \exp(\sqrt{t} U_i)(S_{ii} - \hat{S}_{ii} \right) - \frac{1}{2} E\left( Z^{-2} \sum_i \sum_j w_i w_j \exp(\sqrt{t} U_i) \right. \left. \times \exp(\sqrt{t} U_j)(S_{ij} - \hat{S}_{ij} \right) \quad [41]$$

From the conditions assumed for the covariances, we immediately see that the interpolating function is nonincreasing in $t$, and the theorem follows.

The derivation formula and the comparison theorem are not restricted to the Gaussian case. Generalizations in many directions are possible. For the diluted spin glass models and optimization problems we refer, for example, to Franz and Leone (2003), and to De Sanctis (2005), and references therein.

## Thermodynamic Limit and the Variational Bounds

We give here some striking applications of the basic comparison theorem. Guerra and Toninelli (2002) have given a very simple proof of a long-awaited result, about the convergence of the free energy per site in the thermodynamic limit. Let us show the argument. Let us consider a system of size $N$ and two smaller systems of sizes $N_1$ and $N_2$ respectively, with $N = N_1 + N_2$, as before in the ferromagnetic case. Let us now compare

$$E \log Z_N(\beta, h, J) = E \log \sum_{\sigma_1 \ldots \sigma_N} \exp\left(\beta \sqrt{\frac{N}{2}} \mathcal{K}(\sigma)\right) \times \exp\left(\beta h \sum_i \sigma_i\right) \qquad [42]$$

with

$$E \log \sum_{\sigma_1 \ldots \sigma_N} \exp\left(\beta \sqrt{\frac{N_1}{2}} \mathcal{K}^{(1)}(\sigma^{(1)})\right) \times \exp\left(\beta \sqrt{\frac{N_2}{2}} \mathcal{K}^{(2)}(\sigma^{(2)})\right) \exp\left(\beta h \sum_i \sigma_i\right) \equiv E \log Z_{N_1}(\beta, h, J) + E \log Z_{N_2}(\beta, h, J) \qquad [43]$$

where $\sigma^{(1)}$ stands for $\sigma_i, i = 1, \ldots, N_1$, and $\sigma^{(2)}$ for $\sigma_i, i = N_1 + 1, \ldots, N$. Covariances for $\mathcal{K}^{(1)}$ and $\mathcal{K}^{(2)}$ are expressed as in [28], but now the overlaps are substituted with the partial overlaps of the first and second block, $q_1$ and $q_2$, respectively. It is very simple to apply the comparison theorem. All one has to do is to observe that the obvious

$$Nq = N_1 q_1 + N_2 q_2 \qquad [44]$$

analogous to [10], implies, as in [12],

$$q^2 \le \frac{N_1}{N} q_1^2 + \frac{N_2}{N} q_2^2 \qquad [45]$$

Therefore, the comparison gives the superaddivity property, to be compared with [9],

$$E \log Z_N(\beta, h, J) \ge E \log Z_{N_1}(\beta, h, J) + E \log Z_{N_2}(\beta, h, J) \qquad [46]$$

From the superaddivity property the existence of the limit follows in the form

$$\lim_{N\to\infty} N^{-1} E \log Z_N(\beta, h, J) = \sup_N N^{-1} E \log Z_N(\beta, h, J) \qquad [47]$$

to be compared with [13].

The second application is in the form of the Aizenman–Sims–Starr generalized variational principle. Here, we will need to introduce some auxiliary system. The denumerable configuration space is given by the values of $\alpha = 1, 2, \ldots$. We introduce also weights $w_\alpha \ge 0$ for the $\alpha$ system, and suitably defined overlaps between two generic configurations $p(\alpha, \alpha')$, with $p(\alpha, \alpha) = 1$.

A family of centered Gaussian random variables $\hat{\mathcal{K}}(\alpha)$, now indexed by the configurations $\alpha$, will be defined by the covariances

$$E(\hat{\mathcal{K}}(\alpha)\hat{\mathcal{K}}(\alpha')) = p^2(\alpha, \alpha') \qquad [48]$$

We will also need a family of centered Gaussian random variables $\eta_i(\alpha)$, indexed by the sites $i$ of our original system and the configurations $\alpha$ of the auxiliary system, so that

$$E(\eta_i(\alpha)\eta_{i'}(\alpha')) = \delta_{ii'} p(\alpha, \alpha') \qquad [49]$$

Both the probability measure $w_\alpha$, and the overlaps $p(\alpha, \alpha')$ could depend on some additional external quenched noise, which does not appear explicitly in our notation.

In the following, we will denote by $E$ averages with respect to all random variables involved.

In order to start the comparison argument, we will consider first the case where the two $\sigma$ and $\alpha$ systems are not coupled, so as to appear factorized in the form

$$E \log \sum_{\sigma_1 \ldots \sigma_N} \sum_\alpha w_\alpha \exp\left(\beta \sqrt{\frac{N}{2}} \mathcal{K}(\sigma)\right) \times \exp\left(\beta \sqrt{\frac{N}{2}} \hat{\mathcal{K}}(\alpha)\right) \exp\left(\beta h \sum_i \sigma_i\right) \equiv E \log Z_N(\beta, h, J) + E \log \sum_\alpha w_\alpha \times \exp\left(\beta \sqrt{\frac{N}{2}} \hat{\mathcal{K}}(\alpha)\right) \qquad [50]$$

In the second case, the $\mathcal{K}$ fields are suppressed and the coupling between the two systems will be taken in a very simple form, by allowing the $\eta$ field to act as an external field on the $\sigma$ system. In this way the $\sigma$'s appear as factorized, and the sums can be explicitly performed. The chosen form for the second term in the comparison is

$$E \log \sum_{\sigma_1 \ldots \sigma_N} \sum_\alpha w_\alpha \exp\left(\beta \sum_i \eta_i(\alpha)\sigma_i\right) \exp\left(\beta h \sum_i \sigma_i\right) \equiv N \log 2 + E \log \sum_\alpha w_\alpha (c_1 c_2 \ldots c_N) \qquad [51]$$

where we have defined

$$c_i = \cosh\beta(h + \eta_i(\alpha)) \qquad [52]$$

as arising from the sums over $\sigma$'s.

Now we apply the comparison theorem. In the first case, the covariances involve the sums of squares of overlaps

$$\tfrac{1}{2}(q^2(\sigma,\sigma') + p^2(\alpha,\alpha')) \qquad [53]$$

In the second case, a very simple calculation shows that the covariances involve the overlap products

$$q(\sigma,\sigma')p(\alpha,\alpha'\prime) \qquad [54]$$

Therefore, the comparison is very easy and, by collecting all expressions, we end up with the useful estimate, as in Aizenman *et al.* (2003), holding for any auxiliary system as defined before,

$$\begin{aligned} &N^{-1}E\log Z_N(\beta,h,J) \\ &\quad\le \log 2 + N^{-1}\left(E\log\sum_\alpha w_\alpha(c_1c_2\cdots c_N)\right. \\ &\qquad \left. - E\log\sum_\alpha w_\alpha \exp\left(\beta\sqrt{\frac{N}{2}}\hat{\mathcal{K}}(\alpha)\right)\right) \end{aligned} \qquad [55]$$

## The Parisi Representation for the Free Energy

We refer to the original papers, reprinted in the extensive review given in Mézard *et al.* (2002), for the general motivations, and the derivation of the broken replica ansatz, in the frame of the ingenious replica trick. Here, we limit ourselves to a synthetic description of its general structure, independently from the replica trick.

First of all, let us introduce the convex space $\mathcal{X}$ of the functional order parameters $x$, as nondecreasing functions of the auxiliary variable $q$, both $x$ and $q$ taking values on the interval $[0,1]$, that is,

$$\mathcal{X} \ni x : [0,1] \ni q \to x(q) \in [0,1] \qquad [56]$$

Notice that we call $x$ the function, and $x(q)$ its values. We introduce a metric on $\mathcal{X}$ through the $L^1([0,1], \mathrm{d}q)$-norm, where $\mathrm{d}q$ is the Lebesgue measure.

For our purposes, we will consider the case of piecewise constant functional order parameters, characterized by an integer $K$, and two sequences $q_0, q_1, \ldots, q_K, m_1, m_2, \ldots, m_K$ of numbers satisfying

$$\begin{aligned} 0 = q_0 \le q_1 \le \cdots \le q_{K-1} \le q_K = 1 \\ 0 \le m_1 \le m_2 \le \cdots \le m_K \le 1 \end{aligned} \qquad [57]$$

such that

$$\begin{aligned} &x(q) = m_1 \quad \text{for } 0 = q_0 \le q < q_1 \\ &x(q) = m_2 \quad \text{for } q_1 \le q < q_2 \\ &\vdots \\ &x(q) = m_K \quad \text{for } q_{K-1} \le q \le q_K \end{aligned} \qquad [58]$$

In the following, we will find it convenient to define also $m_0 \equiv 0$, and $m_{K+1} \equiv 1$. The replica symmetric case of Sherrington and Kirkpatrick corresponds to

$$K = 2, \quad q_1 = \bar{q}, \quad m_1 = 0, \quad m_2 = 1 \qquad [59]$$

Let us now introduce the function $f$, with values $f(q,y;x,\beta)$, of the variables $q \in [0,1], y \in R$, depending also on the functional order parameter $x$, and on the inverse temperature $\beta$, defined as the solution of the nonlinear antiparabolic equation

$$\begin{aligned} &(\partial_q f)(q,y) + \tfrac{1}{2}(\partial_y^2 f)(q,y) \\ &\qquad + \tfrac{1}{2}x(q)(\partial_y f)^2(q,y) = 0 \end{aligned} \qquad [60]$$

with final condition

$$f(1,y) = \log\cosh(\beta y) \qquad [61]$$

Here, we have stressed only the dependence of $f$ on $q$ and $y$.

It is very simple to integrate eqn [60] when $x$ is piecewise constant. In fact, consider $x(q) = m_a$, for $q_{a-1} \le q \le q_a$, firstly with $m_a > 0$. Then, it is immediately seen that the correct solution of eqn [60] in this interval, with the right final boundary condition at $q = q_a$, is given by

$$\begin{aligned} &f(q,y) \\ &= \frac{1}{m_a}\log\int \exp(m_a f(q_a, y + z\sqrt{q_a - q}))\,\mathrm{d}\mu(z) \end{aligned} \qquad [62]$$

where $\mathrm{d}\mu(z)$ is the centered unit Gaussian measure on the real line. On the other hand, if $m_a = 0$, then [60] loses the nonlinear part and the solution is given by

$$f(q,y) = \int f(q_a, y + z\sqrt{q_a - q})\,\mathrm{d}\mu(z) \qquad [63]$$

which can be seen also as deriving from [62] in the limit $m_a \to 0$. Starting from the last interval $K$, and using [62] iteratively on each interval, we easily get the solution of [60], [61], in the case of piecewise order parameter $x$, as in [58], through a chain of interconnected Gaussian integrations.

Now, we introduce the following important definitions. The trial auxiliary function, associated to a given mean-field spin glass system, as described

earlier, depending on the functional order parameter $x$, is defined as

$$\log 2 + f(0, h; x, \beta) - \frac{\beta^2}{2}\int_0^1 q\, x(q)\, \mathrm{d}q \qquad [64]$$

Notice that in this expression the function $f$ appears evaluated at $q = 0$, and $y = h$, where $h$ is the value of the external magnetic field. This trial expression shoul be considered as the analog of that appearing in [14] for the ferromagnetic case.

The Parisi spontaneously broken replica symmetry expression for the free energy is given by the definition

$$-\beta f_P(\beta, h) \equiv \inf_x \left(\log 2 + f(0, h; x, \beta) - \frac{\beta^2}{2}\int_0^1 q\, x(q)\, \mathrm{d}q\right) \qquad [65]$$

where the infimum is taken with respect to all functional order parameters $x$. Notice that the infimum appears here, as compared to the supremum in the ferromagnetic case.

By exploiting a kind of generalized comparison argument, involving a suitably defined interpolation function, Guerra (2003) has established the following important result.

**Theorem 3** *For all values of the inverse temperature* $\beta$, *and the external magnetic field* $h$, *and for any functional order parameter* $x$, *the following bound holds*:

$$N^{-1} E \log Z_N(\beta, h, J) \le \log 2 + f(0, h; x, \beta) - \frac{\beta^2}{2}\int_0^1 q\, x(q)\, \mathrm{d}q$$

*uniformly in* $N$. *Consequently, we have also*

$$N^{-1} E \log Z_N(\beta, h, J) \le \inf_x \left(\log 2 + f(0, h; x, \beta) - \frac{\beta^2}{2}\int_0^1 q\, x(q)\, \mathrm{d}q\right)$$

*uniformly in* $N$.

However, this result can also be understood in the framework of the generalized variational principle established by Aizenman–Sims–Starr as described earlier.

In fact, one can easily show that there exist $\alpha$ systems such that

$$N^{-1} E \log \sum_\alpha w_\alpha c_1 c_2 \ldots c_N \equiv f(0, h; x, \beta) \qquad [66]$$

$$N^{-1} E \log \sum_\alpha w_\alpha \exp\left(\beta \sqrt{\frac{N}{2}} \hat{\mathcal{K}}(\alpha)\right) \equiv \frac{\beta^2}{2}\int_0^1 q\, x(q)\, \mathrm{d}q \qquad [67]$$

uniformly in $N$. This result stems from earlier work of Derrida, Ruelle, Neveu, Bolthausen, Sznitman, Aizenman, Contucci, Talagrand, Bovier, and others, and in a sense is implicit in the treatment given in Mézard *et al.* (1987). It can be reached in a very simple way. Let us sketch the argument.

First of all, let us consider the Poisson point process $y_1 \ge y_2 \ge y_3 \ldots$, uniquely characterized by the following conditions. For any interval $A$, introduce the occupation numbers $N(A)$, defined by

$$N(A) = \sum_\alpha \chi(y_\alpha \in A) \qquad [68]$$

where $\chi(\,) = 1$, if the random variable $y_\alpha$ belongs to the interval $A$, and $\chi(\,) = 0$, otherwise. We assume that $N(A)$ and $N(B)$ are independent if the intervals $A$ and $B$ are disjoint, and moreover that for each $A$, the random variable $N(A)$ has a Poisson distribution with parameter

$$\mu(A) = \int_a^b \exp(-y)\, \mathrm{d}y \qquad [69]$$

if $A$ is the interval $(a, b)$, that is,

$$P(N(A) = k) = \exp(-\mu(A))\mu(A)^k / k! \qquad [70]$$

We will exploit $-y_\alpha$ as energy levels for a statistical mechanics system with configurations indexed by $\alpha$. For a parameter $0 < m < 1$, playing the role of inverse temperature, we can introduce the partition function

$$v = \sum_\alpha \exp\left(\frac{y_\alpha}{m}\right) \qquad [71]$$

For $m$ in the given interval it turns out that $v$ is a very well defined random variable, with the sum over $\alpha$ extending to infinity. In fact, there is a strong inbuilt smooth cutoff in the very definition of the stochastic energy levels.

From the general properties of Poisson point processes, it is very well known that the following basic invariance property holds. Introduce a random variable $b$, independent of $y$, subject to the condition $E(\exp b) = 1$, and let $b_\alpha$ be independent copies. Then, the randomly biased point process $y'_\alpha = y_\alpha + b_\alpha$, $\alpha = 1, 2, \ldots$, is equivalent to the original one in distribution. An immediate consequence is the following. Let $f$ be a random variable, independent of $y$, such that $E(\exp f) < \infty$, and let $f_\alpha$ be independent copies. Then, the two random variables

$$\sum_\alpha \exp\left(\frac{y_\alpha}{m}\right) \exp(f_\alpha) \qquad [72]$$

$$\sum_\alpha \exp\left(\frac{y_\alpha}{m}\right) E(\exp(mf))^{1/m} \qquad [73]$$

have the same distribution. In particular, they can be freely substituted under averages.

The auxiliary system which gives rise to the Parisi representation according to [66] and [67], for a piecewise constant order parameter, is expressed in the following way. Now $\alpha$ will be a multi-index $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_K)$, where each $\alpha_a$ runs on $1, 2, 3, \ldots$. Define the Poisson point process $y_{\alpha_1}$, then, independently, for each value of $\alpha_1$ processes $y_{\alpha_1\alpha_2}$, and so on up to $y_{\alpha_1\alpha_2\ldots\alpha_K}$. Notice that in the cascade of independent processes $y_{\alpha_1}, y_{\alpha_1\alpha_2}, \ldots, y_{\alpha_1\alpha_2\ldots\alpha_K}$, the last index refers to the numbering of the various points of the process, while the first indices denote independent copies labeled by the corresponding $\alpha$'s.

The weights $w_\alpha$ have to be chosen according to the definition

$$w_\alpha = \exp\frac{y_{\alpha_1}}{m_1}\exp\frac{y_{\alpha_1\alpha_2}}{m_2}\ldots\exp\frac{y_{\alpha_1\alpha_2\ldots\alpha_K}}{m_K} \qquad [74]$$

The cavity fields $\eta$ and $\mathcal{K}$ have the following expression in terms of independent unit Gaussian random variables $J^i_{\alpha_1}, J^i_{\alpha_1\alpha_2}, \ldots, J^i_{\alpha_1\alpha_2\ldots\alpha_K}, J'_{\alpha_1}, J'_{\alpha_1\alpha_2}, \ldots, J'_{\alpha_1\alpha_2\ldots\alpha_K}$,

$$\eta_i(\alpha) = \sqrt{q_1 - q_0}\,J^i_{\alpha_1} + \sqrt{q_2 - q_1}\,J^i_{\alpha_1\alpha_2} + \cdots + \sqrt{q_K - q_{K-1}}\,J^i_{\alpha_1\alpha_2\ldots\alpha_K} \qquad [75]$$

$$\mathcal{K}(\alpha) = \sqrt{q_1^2 - q_0^2}\,J'_{\alpha_1} + \sqrt{q_2^2 - q_1^2}\,J'_{\alpha_1\alpha_2} + \cdots + \sqrt{q_K^2 - q_{K-1}^2}\,J'_{\alpha_1\alpha_2\ldots\alpha_K} \qquad [76]$$

It is immediate to verify that $E(\eta_i(\alpha)\eta_{i'}(\alpha'))$ is zero if $i \neq i'$, while

$$E(\eta_i(\alpha)\eta_i(\alpha')) = \begin{cases} 0 & \text{if } \alpha_1 \neq \alpha'_1 \\ q_1 & \text{if } \alpha_1 = \alpha'_1, \alpha_2 \neq \alpha'_2 \\ q_2 & \text{if } \alpha_1 = \alpha'_1, \alpha_2 = \alpha'_2, \alpha_3 \neq \alpha'_3, \\ \vdots & \\ 1 & \text{if } \alpha_1 = \alpha'_1, \alpha_2 = \alpha'_2, \ldots, \\ & \alpha_K = \alpha'_K \end{cases} \qquad [77]$$

Similarly, we have

$$E(\mathcal{K}(\alpha)\mathcal{K}(\alpha')) = \begin{cases} 0 & \text{if } \alpha_1 \neq \alpha'_1 \\ q_1^2 & \text{if } \alpha_1 = \alpha'_1, \alpha_2 \neq \alpha'_2 \\ q_2^2 & \text{if } \alpha_1 = \alpha'_1, \alpha_2 = \alpha'_2, \alpha_3 \neq \alpha'_3, \\ \vdots & \\ 1 & \text{if } \alpha_1 = \alpha'_1, \alpha_2 = \alpha'_2, \ldots, \\ & \alpha_K = \alpha'_K \end{cases} \qquad [78]$$

This ends the definition of the $\alpha$ system, associated to a given piecewise constant order parameter.

Now, it is simple to verify that [66] and [67] hold. Let us consider, for example, [66]. With the $\alpha$ system chosen as before, the repeated application of the stochastic equivalence of [72] and [73] will give rise to a sequence of interchained Gaussian integrations exactly equivalent to those arising from the expression for $f$, as solution of the eqn [60]. For [73], there are equivalent considerations.

Therefore, we see that the estimate in Theorem 3 is also a consequence of the generalized variational principle.

Up to this point we have seen how to obtain upper bounds. The problem arises whether, as in the ferromagnetic case, we can also get lower bounds, so as to shrink the thermodynamic limit to the value given by the $\inf_x$ in Theorem 3. After a short announcement, Talagrand (2005) has firmly established the complete proof of the control of the lower bound. We refer to the original paper for the complete details of this remarkable achievement. About the methods, here we only recall that in Guerra (2003) we have given also the corrections to the bounds appearing in Theorem 3, albeit in a quite complicated form. Talagrand has been able to establish that these corrections do in fact vanish in the thermodynamic limit.

In conclusion, we can establish the following extension of Theorem 1 to spin glasses.

**Theorem 4** *For the mean-field spin glass model we have*

$$\lim_{N\to\infty} N^{-1} E \log Z_N(\beta, h, J) = \sup_N N^{-1} E \log Z_N(\beta, h, J) \qquad [79]$$

$$= \inf_x \left( \log 2 + f(0, h; x, \beta) - \frac{\beta^2}{2}\int_0^1 q x(q)\, dq \right) \qquad [80]$$

## Diluted Models

Diluted models, in a sense, play a role intermediate between the mean-field case and the short-range case. In fact, while in the mean-field model each site is interacting with all other sites, on the other hand, in the diluted model, each site is interacting with only a fixed number of other sites. However, while for the short-range models there is a definition of distance among sites, relevant for the interaction, no such definition appears in the diluted models, where all sites are in any case equivalent. From this point of view, the diluted models are structurally similar to the mean-field models, and most of the

techniques and results explained before can be extended to them.

Let us define a typical diluted model. The quenched noise is described as follows. Let $K$ be a Poisson random variable with parameter $\alpha N$, where $N$ is the number of sites, and $\alpha$ is a parameter entering the theory, together with the temperature. We consider also a sequence of independent centered random variables $J_1, J_2, \ldots$, and a sequence of discrete independent random variables $i_1, j_1, i_2, j_2, \ldots$, uniformly distributed over the set of sites $1, 2, \ldots, N$. Then we assume as Hamiltonian

$$H_N(\sigma) = -\sum_{k=0}^{K} J_k \sigma_{i_k} \sigma_{j_k} \qquad [81]$$

Only the variables $\sigma$ contribute to thermodynamic equilibrium. All noise coming from $K, J_k, i_k, j_k$ is considered quenched, and it is not explicitly indicated in our notation for $H$.

The role played by Gaussian integration by parts in the Sherrington–Kirckpatrick model, here is assumed by the following elementary derivation formula, holding for Poisson distributions,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} P(K = k, t\alpha N) &\equiv \frac{\mathrm{d}}{\mathrm{d}t} \exp(-t\alpha N)(t\alpha N)^k / k! \\ &= \alpha N (P(K = k-1, t\alpha N) \\ &\quad - P(K = k, t\alpha N)) \end{aligned} \qquad [82]$$

Then, all machinery of interpolation can be easily extended to the diluted models, as firstly recognized by Franz and Leone in (2003).

In this way, the superaddivity property, the thermodynamic limit, and the generalized variational principle can be easily established. We refer to Franz and Leone (2003), and De Sanctis (2005), for a complete treatment.

There is an important open problem here. While in the fully connected case, the Poisson probability cascades provide the right auxiliary $\alpha$ systems to be exploited in the variational principle, on the other hand in the diluted case more complicated probability cascades have been proposed, as shown, for example, in Franz and Leone (2003), and in Panchenko and Talagrand (2004). On the other hand, in De Sanctis (2005), the very interesting proposal has been made that also in the case of diluted models the Poisson probability cascades play a very important role. Of course, here the auxiliary system interacts with the original system differently, and involves a multi-overlap structure as explained in De Sanctis (2005). In this way a kind of very deep universality is emerging. Poisson probability cascades are a kind of universal class of auxiliary systems. The different models require different cavity fields ruling the interaction between the original system and the auxiliary system. But further work will be necessary in order to clarify this very important issue. For results about diluted models in the high-temperature region, we refer to Guerra and Toninelli (2004).

## Short-Range Model and Its Connections with the Mean-Field Version

The investigations of the connections between the short-range version of the model and its mean-field version are at the beginning. Here, we limit ourselves to a synthetic description of what should be done, and to a short presentation of the results obtained so far.

First of all, according to the conventional wisdom, the mean-field version should be a kind of limit of the short-range model on a lattice in dimension $d$, when $d \to \infty$, with a proper rescaling of the strength of the Hamiltonian, of the form $d^{-1/2}$. Results of this kind are very well known in the ferromagnetic case, but the present technology of interpolation does not seem sufficient to assure a proof in the spin glass case. So, this very basic result is still missing. In analogy with the ferromagnetic case, it would be necessary to arrive at the notion of a critical dimension, beyond which the features of the mean-field case still hold, for example, in the expression of the critical exponents and in the ultrametric hierarchical structure of the pure phases, or at least for the overlap distributions. For physical dimensions less than the critical one, the short-range model would need corrections with respect to its mean-field version. Therefore, this is a completely open problem.

Moreover, always according to the conventional wisdom, the mean-field version should be a kind of limit of the short-range models, in finite fixed dimensions, as the range of the interaction goes to infinity, with proper rescaling. Important work of Franz and Toninelli shows that this is effectively the case, if a properly defined Kac limit is performed. Here, interpolation methods are effective, and we refer to Franz and Toninelli (2004), and references quoted there, for full details.

Due to the lack of efficient analytical methods, it is clear that numerical simulations play a very important role in the study of the physical properties emerging from short-range spin glass models. In particular, we refer to Marinari *et al.* (2000) for a detailed account of the evidence, coming from theoretical considerations and extensive computer simulations, that some of the more relevant features of the spontaneous replica breaking scheme of the mean field are also present in

short-range models in three dimensions. Different views are expressed, for example, in Newman and Stein (1998), where it is argued that the phase-space structure of short-range spin glass models is much simpler than that foreseen by the Parisi spontaneous replica symmetry mechanism.

Such very different views, both apparently strongly supported by reasonable theoretical considerations and powerful numerical simulations, are a natural consequence of the extraordinary difficulty of the problem.

It is clear that extensive additional work will be necessary before the clarification of the physical features exhibited by the realistic short-range spin glass models.

## Conclusion and Outlook for Future Developments

As we have seen, in these last few years, there has been an impressive progress in the understanding of the mathematical structure of spin glass models, mainly due to the systematic exploration of comparison and interpolation methods. However, many important problems are still open. The most important one is to establish rigorously the full hierarchical ultrametric organization of the overlap distributions, as appears in Parisi theory, and to fully understand the decomposition in pure states of the glassy phase, at low temperatures.

Moreover, it would be important to extend these methods to other important disordered models as, for example, neural networks. Here the difficulty is that the positivity arguments, so essential in comparison methods, do not seem to emerge naturally inside the structure of the theory.

Finally, the problem of connecting properties of the short-range model, with those arising in the mean-field case, is still almost completely open.

## Acknowledgment

We gratefully acknowledge useful conversations with Michael Aizenman, Pierluigi Contucci, Giorgio Parisi, and Michel Talagrand. The strategy explained in this report grew out from a systematic exploration of comparison and interpolation methods, developed in collaboration with Fabio Lucio Toninelli, and Luca De Sanctis.

This work was supported in part by MIUR (Italian Ministry of Instruction, University and Research), and by INFN (Italian National Institute for Nuclear Physics).

*See also:* Glassy Disordered Systems: Dynamical Evolution; Large Deviations in Equilibrium Statistical Mechanics; Mean Field Spin Glasses and Neural Networks; Short-Range Spin Glasses: The Metastate Approach; Statistical Mechanics and Combinatorial Problems.

## Further Reading

Aizenman M, Sims R, and Starr S (2003) Extended variational principle for the Sherrington–Kirkpatrick spin-glass model. *Physical Review B* 68: 214403.

De Sanctis L (2005) *Structural Approachs to Spin Glasses and Optimization Problems*. Ph.D. thesis, Department of Mathematics, Princeton University.

Edwards SF and Anderson PW (1975) Theory of spin glasses. *Journal of Physics F: Metal Physics* 5: 965–974.

Franz S and Leone M (2003) Replica bounds for optimization problems and diluted spin systems. *Journal of Statistical Physics* 111: 535–564.

Franz S and Toninelli FL (2004) The Kac limit for finite-range spin glasses. *Physical Review Letters* 92: 030602.

Guerra F (2001) Sum rules for the free energy in the mean field spin glass model. *Fields Institute Communications* 30: 161.

Guerra F (2003) Broken replica symmetry bounds in the mean field spin glass model. *Communications in Mathematical Physics* 233: 1–12.

Guerra F and Toninelli FL (2002) The thermodynamic limit in mean field spin glass models. *Communications in Mathematical Physics* 230: 71–79.

Guerra F and Toninelli FL (2004) The high temperature region of the Viana–Bray diluted spin glass model. *Journal of Statistical Physics* 115: 531–555.

Marinari E, Parisi G, Ricci-Tersenghi F, Ruiz-Lorenzo JJ, and Zuliani F (2000) Replica symmetry breaking in short range spin glasses: A review of the theoretical foundations and of the numerical evidence. *Journal of Statistical Physics* 98: 973–1074.

Mézard M, Parisi G, and Virasoro MA (1987) *Spin Glass Theory and Beyond*. Singapore: World Scientific.

Mézard M, Parisi G, and Zecchina R (2002) Analytic and algorithmic solution of random satisfiability problems. *Science* 297: 812.

Newman CM and Stein DL (1998) Simplicity of state and overlap structure in finite-volume realistic spin glasses. *Physical Review E* 57: 1356–1366.

Panchenko D and Talagrand M (2004) Bounds for diluted mean-field spin glass models. *Probability Theory Related Fields* 130: 319–336.

Sherrington D and Kirkpatrick S (1975) Solvable model of a spin-glass. *Physical Review Letters* 35: 1792–1796.

Stein DL (1989) Disordered systems: Mostly spin glasses. In: Stein DL (ed.) *Lectures in the Sciences of Complexity*. New York: Addison-Wesley.

Talagrand M (2003) *Spin Glasses: A Challenge for Mathematicians. Mean Field Models and Cavity Method*. Berlin: Springer.

Talagrand M (2006) The Parisi formula. *Annals of Mathematics* 163: 221–263.

Young P (ed.) (1987) *Spin Glasses and Random Fields*. Singapore: World Scientific.

# Spinors and Spin Coefficients

**K P Tod**, University of Oxford, Oxford, UK



## Introduction

Spinors were invented by the mathematician E Cartan (see, e.g., Cartan (1981)) in the early years of the last century in the course of his study of rotation groups. The physicist Pauli reinvented what Cartan would have called the spinors of SU(2), which is the double cover of the rotation group SO(3), in order to explain the spectroscopy of alkali atoms and the anomalous Zeeman effect. For this, he needed an essential two-valuedness of the electron, an internal quantum number to contribute to the angular momentum, which is now called spin. Now the wave function becomes a two-component column vector. It is worth noting that, despite the name, Pauli resisted the picture of an electron as a spinning "thing" on the grounds that, as a representation of SU(2) which was not a representation of SO(3), it should have no classical kinematic model, which a spinning object would have.

According to the review article of van der Waerden (1960), the term "spinor" is due to Ehrenfest in 1929, and was introduced in the flurry of interest after the next important step in the evolution of spinors in the physics literature, which was the introduction of a relativistic equation for the electron by Dirac (1928).

Dirac sought a linear, first-order but Lorentz-invariant equation for the electron which was to be the square root of the linear, Lorentz-invariant but second-order Klein–Gordon equation. He assumed the equation for the wave function $\psi$ would take the form

$$L\psi := (\mathrm{i}\gamma^a p_a + mcI)\psi = 0 \qquad [1]$$

where $p_a = -\mathrm{i}\hbar\partial/\partial x^a$ for $a = 0, 1, 2, 3$, but where $\gamma^a$ are complex square matrices, of a size to be determined, and $I$ is the corresponding identity matrix. Differentiating [1] again, one obtains the Klein–Gordon equation for $\psi$ provided these matrices satisfy the equation

$$\gamma^a\gamma^b + \gamma^b\gamma^a = 2\eta^{ab}I \qquad [2]$$

where $\eta^{ab}$ is the Minkowski metric, diag(1, −1, −1, −1).

Assuming the $\gamma^a$ have been found, the usual substitution $p \to p - \mathrm{i}eA$, for a particle in a magnetic field with vector potential $A$, leads to the correct magnetic moment for the electron, so that this equation does describe an electron with spin in the form made familiar by Pauli.

To decide on the size of the matrices $\gamma^a$ and therefore the dimension of the space of $\psi$'s, one notices, with the aid of [2], that the following are a basis for the algebra generated by the $\gamma^a$:

$$1, \gamma^a, \gamma^{[a}\gamma^{b]}, \gamma^{[a}\gamma^b\gamma^{c]}, \gamma^{[a}\gamma^b\gamma^c\gamma^{d]} \qquad [3]$$

There are 16 elements in this basis, assuming that there are no extra identities among them, so that we might hope to find a representation as $4 \times 4$ matrices. This can be done, and Dirac gave explicit formulas in terms of Pauli matrices. The space of Dirac spinors is now a complex four-dimensional vector space, which turns out to split as the sum of a complex two-dimensional vector space $S$, which is referred to as a spin space, and its complex conjugate $\bar{S}$ (the relationship between a complex vector space and its complex conjugate is described in the text below and eqn [9]). Under proper, orthochronous Lorentz transformations, $S$ transforms into itself by $SL(2, \mathbb{C})$ transformation, but space and time reflections relate $S$ to $\bar{S}$. The fact that there are two spin spaces $S$ and $\bar{S}$ in dimension 4 is the basis of chirality: an electron is represented by a Dirac spinor, which is a pair of spinors, one in each of $S$ and $\bar{S}$, which are related under space reflection; a particle represented just by a spinor in $S$ cannot be invariant under space reflection.

The Clifford algebra (see Clifford Algebras and their Representations) associated with a vector space $V$ with metric $g$ is defined as the algebra generated by elements $v$, $w$ of $V$ with the multiplication $\cup$ satisfying

$$v \cup w + w \cup v = 2g(v, w) \qquad [4]$$

The matrices $\gamma^a$ define a representation of the Clifford algebra by associating a covector $v_a$ with a matrix $v = v_a\gamma^a$, since [2] is then equivalent to [4].

This part of the process works in any dimension $n$ and signature $s$. For odd $n$, as, for example, with Pauli spinors, the $\gamma^a$ are square matrices of size $2^N \times 2^N$, where $N = (n-1)/2$, and there is a single spin space of dimension $2^N$. For even $n$, as with the original Dirac spinors, the $\gamma^a$ are square matrices of size $2^N \times 2^N$, where $N = n/2$, but there are two spin spaces each of dimension $2^{N-1}$. Reality properties of the spin spaces and the existence of other structures on them depend in an intricate way on $n$ and $s$ (Penrose and Rindler 1984, 1986, Benn and Tucker 1987).

The dimension of the space of spinors rises rapidly with $n$, which is one reason why historically spinors have been most useful in spaces of dimensions 3 and 4, where the spin space has dimension 2. In a space of dimension 11, a case considered in supergravity, the spin space already has dimension 32.

## Spinors in General Relativity: Spinor Algebra

In this section, we start again with a different emphasis. Conventions follow Penrose and Rindler (1984, 1986). To introduce spinors as a calculus in a four-dimensional, Lorentzian spacetime $\mathcal{M}$, one can begin by choosing an orthonormal tetrad of vectors $(\boldsymbol{e}_0, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$ at a point $p$. The following conventions are used:

$$g(\boldsymbol{e}_a, \boldsymbol{e}_b) = \eta_{ab} = \mathrm{diag}(1, -1, -1, -1)$$

Any vector $\boldsymbol{v}$ in the tangent space $V = T_p\mathcal{M}$ at $p$ has components $v^a$ in this basis, which we arrange as a matrix and label in two ways:

$$\Psi(\boldsymbol{v}) = \frac{1}{\sqrt{2}}\begin{pmatrix} v^0 + v^3 & v^1 + \mathrm{i}v^2 \\ v^1 - \mathrm{i}v^2 & v^0 - v^3 \end{pmatrix} = \begin{pmatrix} v^{00'} & v^{01'} \\ v^{10'} & v^{11'} \end{pmatrix} \quad [5]$$

The reason for the factor $1/\sqrt{2}$ will be seen below, as will the rationale for the second form of the matrix. Note that $\Psi(\boldsymbol{v})$ is Hermitian and that

$$2 \det \Psi(\boldsymbol{v}) = g(\boldsymbol{v}, \boldsymbol{v}) = \eta_{ab} v^a v^b \quad [6]$$

Clearly, there is a one-to-one correspondence between elements of $V$ and Hermitian $2 \times 2$ matrices. Further, if $t$ is any matrix in $\mathrm{SL}(2, \mathbb{C})$,then the transformation

$$\Psi(\boldsymbol{v}) \to t\Psi(\boldsymbol{v})t^\dagger \quad [7]$$

where $t^\dagger$ is the Hermitian conjugate of $t$, is linear in $\boldsymbol{v}$, and preserves both Hermiticity and the norm of $\boldsymbol{v}$. Thus, it must represent a Lorentz transformation. It is straightforward to check that it is a proper, orthochronous Lorentz transformation and that all such transformations arise in this way (recall that "proper" means transformations of determinant 1 so that orientation is preserved, and "orthochronous" means that future-pointing timelike or null vectors are taken to future-pointing timelike or null vectors, so that time orientation is preserved; the proper, orthochronous Lorentz group is equivalently the identity-connected component of the Lorentz group). Since both $t$ and $-t$ give the same Lorentz transformation, this provides an explicit demonstration of the (2 – 1)-homomorphism of SL(2,$\mathbb{C}$) with the proper, orthochronous Lorentz group $\mathrm{O}_+^\uparrow(1,3)$.

If the vector $\boldsymbol{v}$ in [5] is null, then the matrix has vanishing determinant, or, equivalently, it has rank 1, and so it can be written as the outer product of a two-component column vector $\alpha = (\alpha^0, \alpha^1)^{\mathrm{T}}$ and its Hermitian conjugate:

$$\Psi(\boldsymbol{v}) = \alpha\alpha^\dagger \quad [8]$$

Furthermore, under [7], $\alpha$ transforms as

$$\alpha \to t\alpha \quad [9]$$

The two-complex-dimensional space to which $\alpha$ belongs is the spin space $S$ at $p$, already met in the previous section, and it follows from [8], since null vectors span $V$, that the tensor product $S \otimes \bar{S}$ of $S$ with its complex conjugate vector space $\bar{S}$ is the complexification of $V$. Complex conjugation gives an antilinear map from $S$ to $\bar{S}$. (One associates the complex-conjugate vector space $\bar{V}$ to any given complex vector space $V$ as follows: scalar multiplication for $V$ can be considered as a function $\phi : \mathbb{C} \times V \to V$ given by $\phi(z, v) = zv$, while vector addition is a map $\psi : V \times V \to V$ given by $\psi(u, v) = u + v$. Define another complex vector space by taking the same vectors and the same $\psi$ but with scalar multiplication $\bar{\phi}$, where $\bar{\phi}(z, v) = \phi(\bar{z}, v)$. This is the complex-conjugate vector space $\bar{V}$. Given a choice of basis, we think of $V$ as, say, $n$-component column vectors of complex numbers, and then $\bar{V}$ is the corresponding complex-conjugate columns.)

Conventionally, $S$ is the space of unprimed spinors and $\bar{S}$ the space of primed spinors, and one also has the two duals $S'$ and $\bar{S}'$ which are associated in the corresponding way to the dual $V'$ of $V$. Analogously to the situation with vectors and covectors, index conventions for spinors are as follows:

$$\alpha^A \in S, \quad \beta^{A'} \in \bar{S}, \quad \gamma_A \in S', \quad \delta_{A'} \in \bar{S}'$$

where $A = 0, 1, A' = 0', 1'$.

Spinor algebra mirrors tensor algebra: a spinor $\phi^{A_1 \ldots A_p A'_1 \ldots A'_q}{}_{B_1 \ldots B_r B'_1 \ldots B'_s}$ is an element of the tensor product of $p$ copies of $S$, $q$ copies of $\bar{S}$, $r$ copies of $S'$, and $s$ copies of $\bar{S}'$. The second way of writing the matrix in [5] enables the identification of a vector with a matrix to be conventionally written as

$$v^a = v^{AA'} \quad [10]$$

and then extended to any tensor $T^{a \ldots b}{}_{c \ldots d}$ by replacing each vector index, say $b$, with a pair $BB'$ of spinor indices. In particular, from [8], it follows that any real null vector $n^a$ can be written in the form

$$n^a = \nu^A \bar{\nu}^{A'}$$

for some spinor $\nu^A$.

One must pay attention to the order of spinor indices of a given type, primed or unprimed, but by convention may permute primed and unprimed indices. A spinor with an equal number $n$ of primed and unprimed indices corresponds to a tensor of valence $n$, and the tensor is real if the spinor satisfies a suitable Hermiticity relation.

Spinors may have various symmetries among their indices, much as tensors have. However, since $S$ is two dimensional, there is only a one-dimensional space of 2-forms on $S$. This has two consequences: no spinor can be antisymmetric over more than two indices; and if we make a choice of canonical 2-form, all spinors can be written in terms of symmetric spinors and the canonical 2-form. This is a decomposition of spinors into irreducibles for $SL(2,\mathbb{C})$.

One makes a choice of 2-form $\epsilon_{AB}$ according to

$$\epsilon_{AB} = -\epsilon_{BA}, \quad \epsilon_{01} = 1$$

There is an inverse $\epsilon^{AB}$ defined by

$$\epsilon^{AC}\epsilon_{BC} = \delta^A_B \qquad [11]$$

where $\delta^A_B$ is the Kronecker delta. The complex conjugate of $\epsilon_{AB}$ is conventionally written without an overbar as $\epsilon_{A'B'}$, and analogously $\epsilon^{A'B'}$ is the complex conjugate of $\epsilon^{AB}$.

Because of the antisymmetry of $\epsilon_{AB}$, order of indices is crucial in equations such as [11]. The 2-form $\epsilon_{AB}$ has a role akin to that of a metric as it provides an identification of $S$ and its dual, according to

$$\alpha^A \to \alpha_B = \alpha^A \epsilon_{AB}$$

$$\beta_B \to \beta^A = \epsilon^{AB}\beta_B$$

with corresponding formulas for primed spinors. Note that, because of the antisymmetry of $\epsilon_{AB}$, necessarily $\alpha_A\alpha^A = 0$ for any $\alpha^A$.

With conventions made so far, it can be checked that

$$g_{ab}v^a v^b = \epsilon_{AB}\epsilon_{A'B'}v^{AA'}v^{BB'} \qquad [12]$$

for any vector $v^a$, where $g_{ab}$ is the spacetime metric at $p$, so that

$$g_{ab} = \epsilon_{AB}\epsilon_{A'B'}$$

It is the desire to have this formula without constants that necessitates the choice of the factor $1/\sqrt{2}$ in [5].

One final piece of spinor algebra that we note is the following: given a symmetric spinor $\phi_{A_1\ldots A_n}$ there is a factorization

$$\phi_{A_1\ldots A_n} = \alpha^{(1)}_{(A_1} \cdots \alpha^{(n)}_{A_n)} \qquad [13]$$

where the round brackets indicate symmetrization over the indices $A_1, \ldots, A_n$, and the $n$ spinors $\alpha^{(1)}_{A_1}, \ldots, \alpha^{(n)}_{A_n}$, which are determined only up to reordering and rescaling, are known as the principal spinors of $\phi$. To prove this, note that the principal spinors can be identified with the solutions $\zeta^A$ of the equation

$$\phi_{A_1\ldots A_n}\zeta^{A_1}\cdots\zeta^{A_n} = 0$$

and there are $n$ of these, counting multiplicities, by the "fundamental theorem of algebra."

## Spinors in General Relativity: Spinor Calculus

We now want to define spinor fields on the spacetime $\mathcal{M}$ as sections of a spinor bundle $\mathcal{S}$ whose fiber at each point is $S$ and such that the tensor product $\mathcal{S}\otimes\bar{\mathcal{S}}$ is the complexified tangent bundle. The existence of such an $\mathcal{S}$ imposes global restrictions on $\mathcal{M}$: $\mathcal{M}$ must be orientable and time orientable, and a certain characteristic class, the second Stiefel–Whitney class, must vanish (for an explanation of these terms see, e.g., Penrose and Rindler (1984, 1986)). Assuming that $\mathcal{M}$ satisfies these conditions, spinor fields can be defined. It is convenient to retain the algebraic formulas from the previous section (e.g., [10] or [12]) but with indices now regarded as abstract (a note on the abstract index convention appears in Twistors).

By an argument analogous to that for the fundamental theorem of Riemannian geometry, there is a unique covariant derivative that satisfies the Leibniz condition, coincides with the Levi-Civita derivative on tensors and the gradient on scalars, and annihilates $\epsilon_{AB}$ and $\epsilon_{A'B'}$. Following the conventions of the previous section, the spinor covariant derivative will be denoted as $\nabla_{AA'}$. The commutator of derivatives can be written in terms of irreducible parts (for $SL(2,\mathbb{C})$) according to the formula

$$\nabla_{AA'}\nabla_{BB'} - \nabla_{BB'}\nabla_{AA'} = \epsilon_{A'B'}\Delta_{AB} + \epsilon_{AB}\Delta_{A'B'}$$

where $\Delta_{AB} = \nabla_{C'(A}\nabla^{C'}_{B)}$. The definition of the Riemann curvature tensor is in terms of the Ricci identity

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)v^c = R_{abd}{}^c v^d$$

and then this translates into two Ricci identities for a spinor field:

$$\Delta_{AB}\alpha_C = \chi_{ABCD}\alpha^D$$

$$\Delta_{A'B'}\alpha_C = \Phi_{A'B'CD}\alpha^D$$

The curvature spinors $\chi_{ABCD}$ and $\Phi_{A'B'CD}$ are related to the curvature tensor. The Ricci spinor $\Phi_{A'B'AB}$ is Hermitian and symmetric on both index pairs and is a multiple of the trace-free part of the Ricci tensor:

$$\Phi_{A'B'AB} = -\tfrac{1}{2}\left(R_{ab} - \tfrac{1}{4}Rg_{ab}\right)$$

The spinor $\chi_{ABCD}$ is symmetric on the first and last pairs of indices and decomposes into irreducibles according to

$$\chi_{ABCD} = \Psi_{ABCD} - 2\Lambda\epsilon_{D(A}\epsilon_{B)C}$$

where $\Lambda = R/24$ in terms of the Ricci scalar or scalar curvature $R$, while $\Psi_{ABCD}$, which is totally symmetric and is known as the Weyl spinor, is related to the Weyl tensor $C_{abcd}$ by the equation

$$C_{abcd} = \Psi_{ABCD}\epsilon_{A'B'}\epsilon_{C'D'} + \bar{\Psi}_{A'B'C'D'}\epsilon_{AB}\epsilon_{CD}$$

Thus, the ten real components of the Weyl tensor are coded into the five complex components of the Weyl spinor.

Following the last remark in the previous section, the Weyl spinor has four principal spinors, each of which defines a null direction, the principal null directions of the Weyl tensor. There is a classification of Weyl tensors, the Petrov–Pirani–Penrose classification, based on coincidences among the principal null directions (Penrose and Rindler 1984, 1986).

As a final exercise in spinor calculus, we recall the zero-rest-mass equations (see Twistors). In flat spacetime, these are the equations

$$\nabla^{A'A}\phi_{AB\ldots C} = 0$$

on a totally symmetric spinor field $\phi_{AB\ldots C}$. The field is said to have spin $s$ if it has $2s$ indices, and the cases $s = 1/2$, 1, or 2, respectively, are the Weyl neutrino equation, the Maxwell equation, and the linearized Einstein equation. In flat spacetime, these hyperbolic equations are well understood and solvable in a variety of ways. In curved spacetime, however, if $s \geq 3/2$, then there are curvature obstructions to the existence of solutions, known as Buchdahl conditions. This can be seen at once by differentiating again, say by $\nabla^{B}_{A'}$, and using the spinor Ricci identity. After a little algebra, one finds

$$\Psi^{ABC}{}_{(D}\phi_{E\ldots F)ABC} = 0$$

so that, whenever the field has three or more indices, there are algebraic constraints on its components in terms of the Weyl spinor.

## The Spin-Coefficient Formalism

The spin-coefficient formalism of Newman and Penrose is a formalism for spinor calculus in space-times (see, e.g., Penrose and Rindler (1984, 1986) and Stewart (1990)). It finds application in any calculation dealing with curvature tensors, including solving the Einstein equations. The formalism exploits the compression of terminology which the introduction of complex quantities permits.

The formalism starts with a choice of spinor dyad, a basis of spinor fields $(o^A, \iota^A)$ normalized so that $o_A\iota^A = 1$. From the dyad, one constructs a null tetrad, which is a basis of vector fields, according to the scheme

$$\ell^a = o^A\bar{o}^{A'},\ n^a = \iota^A\bar{\iota}^{A'},\ m^a = o^A\bar{\iota}^{A'},\ \bar{m}^a = \iota^A\bar{o}^{A'}$$

Given the normalization of the spinor dyad, each of the vectors in the null tetrad is null (hence the name) and all inner products are zero, except for

$$\ell^a n_a = 1 = -m^a\bar{m}_a$$

It follows that the metric can be written in the basis as

$$g_{ab} = 2\ell_{(a}n_{b)} - 2m_{(a}\bar{m}_{b)}$$

The components of the covariant derivative in the null tetrad are given separate names according to the following scheme:

$$\ell^a\nabla_a = D,\ n^a\nabla_a = \Delta,\ m^a\nabla_a = \delta,\ \bar{m}^a\nabla_a = \bar{\delta}$$

and the spin coefficients are the 12 components of the covariant derivative of the basis. Each is labeled with a Greek letter according to the following scheme:

$$\begin{aligned} Do^A &= \epsilon o^A - \kappa\iota^A, & \Delta o^A &= \gamma o^A - \tau\iota^A \\ \delta o^A &= \beta o^A - \sigma\iota^A, & \bar{\delta}o^A &= \alpha o^A - \rho\iota^A \\ D\iota^A &= \pi o^A - \epsilon\iota^A, & \Delta\iota^A &= \nu o^A - \gamma\iota^A \\ \delta\iota^A &= \mu o^A - \beta\iota^A, & \bar{\delta}\iota^A &= \lambda o^A - \alpha\iota^A \end{aligned} \quad [14]$$

The spin coefficients code the 24 real Ricci rotation coefficients into 12 complex quantities. Some of the spin coefficients have direct geometrical interpretation. For example, the vanishing of $\kappa$ is the condition for the integral curves of $\ell^a$ to be geodesic, while, if $\sigma$ is also zero, this congruence of geodesics is shear free. The same role is played by $\nu$ and $\lambda$ for the $n^a$-congruence. The real and imaginary parts of $\rho$ are, respectively (minus), the expansion and the twist of the congruence of integral curves of $\ell^a$.

In practice, it is often simpler to calculate the spin coefficients from the commutators of the basis vectors, now regarded as directional derivatives, as follows:

$$\begin{aligned}\Delta D - D\Delta &= (\gamma+\bar{\gamma})D + (\epsilon+\bar{\epsilon})\Delta - (\bar{\tau}+\pi)\delta - (\tau+\bar{\pi})\bar{\delta}\\ \delta D - D\delta &= (\bar{\alpha}+\beta-\bar{\pi})D + \kappa\Delta - (\bar{\rho}+\epsilon-\bar{\epsilon})\delta - \sigma\bar{\delta}\\ \delta\Delta - \Delta\delta &= -\bar{\nu}D + (\tau-\bar{\alpha}-\beta)\Delta + (\mu-\gamma+\bar{\gamma})\delta + \bar{\lambda}\bar{\delta}\\ \bar{\delta}\delta - \delta\bar{\delta} &= (\bar{\mu}-\mu)D + (\bar{\rho}-\rho)\Delta + (\alpha-\bar{\beta})\delta - (\bar{\alpha}-\beta)\bar{\delta}\end{aligned} \qquad [15]$$

The commutator of second derivatives applied to the spinor dyad expresses the components of the curvature tensor in terms of the derivatives of the spin coefficients. Before presenting these, we adopt a convention for labeling the components of curvature. The components of the Weyl spinor are given as follows:

$$\begin{aligned}\Psi_0 &= \Psi_{ABCD}o^A o^B o^C o^D\\ \Psi_1 &= \Psi_{ABCD}o^A o^B o^C \iota^D\\ \Psi_2 &= \Psi_{ABCD}o^A o^B \iota^C \iota^D\\ \Psi_3 &= \Psi_{ABCD}o^A \iota^B \iota^C \iota^D\\ \Psi_4 &= \Psi_{ABCD}\iota^A \iota^B \iota^C \iota^D\end{aligned} \qquad [16]$$

so that these five complex scalars encode the ten real components of the Weyl tensor. For the Ricci spinor, set

$$\begin{aligned}\Phi_{00} &= \Phi_{ABA'B'}o^A o^B \bar{o}^{A'} \bar{o}^{B'}, & \Phi_{01} &= \Phi_{ABA'B'}o^A o^B \bar{o}^{A'} \bar{\iota}^{B'}\\ \Phi_{02} &= \Phi_{ABA'B'}o^A o^B \bar{\iota}^{A'} \bar{\iota}^{B'}, & \Phi_{11} &= \Phi_{ABA'B'}o^A \iota^B \bar{o}^{A'} \bar{\iota}^{B'}\\ \Phi_{12} &= \Phi_{ABA'B'}o^A \iota^B \bar{\iota}^{A'} \bar{\iota}^{B'}, & \Phi_{22} &= \Phi_{ABA'B'}\iota^A \iota^B \bar{\iota}^{A'} \bar{\iota}^{B'}\end{aligned}$$

together with $\Phi_{10} = \overline{\Phi_{01}}, \Phi_{20} = \overline{\Phi_{02}}$, and $\Phi_{21} = \overline{\Phi_{12}}$. The nine components of the trace-free Ricci tensor are encoded in these scalars of which three are real and three complex. The Ricci scalar, as before, is replaced by the real scalar $\Lambda = R/24$.

Now the commutators of covariant derivatives on the spinor dyad lead to the following system:

$$\begin{aligned}D\rho - \bar{\delta}\kappa &= \rho^2 + \sigma\bar{\sigma} + (\epsilon+\bar{\epsilon})\rho - \bar{\kappa}\tau\\ &\quad - (3\alpha + \bar{\beta} - \pi)\kappa + \Phi_{00}\\ D\sigma - \delta\kappa &= (\rho+\bar{\rho}+3\epsilon-\bar{\epsilon})\sigma\\ &\quad - (\tau - \bar{\pi} + \bar{\alpha} + 3\beta)\kappa + \Psi_0\\ D\tau - \Delta\kappa &= (\tau+\bar{\pi})\rho + (\bar{\tau}+\pi)\sigma + (\epsilon-\bar{\epsilon})\tau\\ &\quad - (3\gamma+\bar{\gamma})\kappa + \Psi_1 + \Phi_{01}\\ D\alpha - \bar{\delta}\epsilon &= (\rho+\bar{\epsilon}-2\epsilon)\alpha + \beta\bar{\sigma} - \bar{\beta}\epsilon - \kappa\lambda - \bar{\kappa}\gamma\\ &\quad + (\epsilon+\rho)\pi + \Phi_{10}\\ D\beta - \delta\epsilon &= (\alpha+\pi)\sigma + (\bar{\rho}-\bar{\epsilon})\beta - (\mu+\gamma)\kappa\\ &\quad - (\bar{\alpha}-\bar{\pi})\epsilon + \Psi_1\\ D\gamma - \Delta\epsilon &= (\tau+\bar{\pi})\alpha + (\bar{\tau}+\pi)\beta - (\epsilon+\bar{\epsilon})\gamma - (\gamma+\bar{\gamma})\epsilon\\ &\quad + \tau\kappa - \nu\kappa + \Psi_2 - \Lambda + \Phi_{11}\\ D\lambda - \bar{\delta}\pi &= (\rho - 3\epsilon + \bar{\epsilon})\lambda + \bar{\sigma}\mu + (\pi+\alpha-\bar{\beta})\pi\\ &\quad - \nu\bar{\kappa} + \Phi_{20}\\ D\mu - \delta\pi &= (\bar{\rho}-\epsilon-\bar{\epsilon})\mu + \sigma\lambda + (\bar{\pi}-\bar{\alpha}+\beta)\pi\\ &\quad - \nu\kappa + \Psi_2 + 2\Lambda\\ D\nu - \Delta\pi &= (\pi+\bar{\tau})\mu + (\bar{\pi}+\tau)\lambda + (\gamma-\bar{\gamma})\pi\\ &\quad - (3\epsilon+\bar{\epsilon})\nu + \Psi_3 + \Phi_{21}\\ \Delta\lambda - \bar{\delta}\nu &= -(\mu+\bar{\mu}+3\gamma-\bar{\gamma})\lambda\\ &\quad + (3\alpha+\bar{\beta}+\pi-\bar{\tau})\nu - \Psi_4\\ \delta\rho - \bar{\delta}\sigma &= (\bar{\alpha}+\beta)\rho - (3\alpha-\bar{\beta})\sigma + (\rho-\bar{\rho})\tau\\ &\quad + (\mu-\bar{\mu})\kappa - \Psi_1 + \Phi_{01}\\ \delta\alpha - \bar{\delta}\beta &= \mu\rho - \lambda\sigma + \alpha\bar{\alpha} + \beta\bar{\beta} - 2\alpha\beta + (\rho-\bar{\rho})\gamma\\ &\quad + (\mu-\bar{\mu})\epsilon - \Psi_2 + \Lambda + \Phi_{11}\\ \delta\lambda - \bar{\delta}\mu &= (\rho-\bar{\rho})\nu + (\mu-\bar{\mu})\pi + (\alpha+\bar{\beta})\mu\\ &\quad + (\bar{\alpha}-3\beta)\lambda - \Psi_3 + \Phi_{21}\\ \Delta\mu - \delta\nu &= -(\mu+\gamma+\bar{\gamma})\mu - \lambda\bar{\lambda} + \bar{\nu}\pi\\ &\quad + (\bar{\alpha}+3\beta-\tau)\nu - \Phi_{22}\\ \Delta\beta - \delta\gamma &= (\bar{\alpha}+\beta-\tau)\gamma - \mu\tau + \sigma\nu + \epsilon\bar{\nu}\\ &\quad + (\gamma-\bar{\gamma}-\mu)\beta - \alpha\bar{\lambda} - \Phi_{12}\\ \Delta\sigma - \delta\tau &= -(\mu-3\gamma+\bar{\gamma})\sigma - \bar{\lambda}\rho - (\tau+\beta-\bar{\alpha})\tau\\ &\quad + \kappa\bar{\nu} - \Phi_{02}\\ \Delta\rho - \bar{\delta}\tau &= (\gamma+\bar{\gamma}-\bar{\mu})\rho - \sigma\lambda + (\bar{\beta}-\alpha-\bar{\tau})\tau\\ &\quad + \nu\kappa - \Psi_2 - 2\Lambda\\ \Delta\alpha - \bar{\delta}\gamma &= (\rho+\epsilon)\nu - (\tau+\beta)\lambda + (\bar{\gamma}-\bar{\mu})\alpha\\ &\quad + (\bar{\beta}-\bar{\tau})\gamma - \Psi_3\end{aligned} \qquad [17]$$

Finally, it is possible to write out the Bianchi identities in this formalism. For simplicity, and with a view to an application, we do this below only for vacuum, so that the Ricci tensor is zero:

$$\begin{aligned}D\Psi_1 - \bar{\delta}\Psi_0 &= (\pi - 4\alpha)\Psi_0 + 2(2\rho+\epsilon)\Psi_1 - 3\kappa\Psi_2\\ \Delta\Psi_0 - \delta\Psi_1 &= (4\gamma-\mu)\Psi_0 - 2(2\tau+\beta)\Psi_1 + 3\sigma\Psi_2\\ D\Psi_2 - \bar{\delta}\Psi_1 &= -\lambda\Psi_0 + 2(\pi-\alpha)\Psi_1 + 3\rho\Psi_2 - 2\kappa\Psi_3\\ \Delta\Psi_1 - \delta\Psi_2 &= \nu\Psi_0 + 2(\gamma-\mu)\Psi_1 - 3\tau\Psi_2 + 2\sigma\Psi_3\\ D\Psi_3 - \bar{\delta}\Psi_2 &= -2\lambda\Psi_1 + 3\pi\Psi_2 + 2(\rho-\epsilon)\Psi_3 - \kappa\Psi_4\\ \Delta\Psi_2 - \delta\Psi_3 &= 2\nu\Psi_1 - 3\mu\Psi_2 + 2(\beta-\tau)\Psi_3 + \sigma\Psi_4\\ D\Psi_4 - \bar{\delta}\Psi_3 &= -3\lambda\Psi_2 + 2(\alpha+2\pi)\Psi_3 + (\rho-4\epsilon)\Psi_4\\ \Delta\Psi_3 - \delta\Psi_4 &= 3\nu\Psi_2 - 2(\gamma+2\mu)\Psi_3 + (4\beta-\tau)\Psi_4\end{aligned} \qquad [18]$$

The whole system is then loosely described as the spin-coefficient equations.

As a simple application, we shall prove the Goldberg–Sachs theorem: for vacuum spacetimes, a

spinor field $o^A$ is geodesic and shear free iff it is a repeated principal spinor of the Weyl spinor.

In the spin-coefficient formalism, $o^A$ is geodesic and shear-free iff $\kappa$ and $\sigma$ vanish, and, from [16], is a repeated principal spinor of the Weyl spinor provided $\Psi_0 = \Psi_1 = 0$. It will be repeated three times if also $\Psi_2 = 0$ and four times if $\Psi_3 = 0$, but one must have $\Psi_k \neq 0$ for some $k$ if the spacetime is not to be flat.

Suppose that $o^A$ is a (twice) repeated principal spinor of the Weyl spinor, then at once from the first two expressions in [18] both $\kappa$ and $\sigma$ vanish. If it is repeated three times, one gets the same result from the third and fourth expressions in [18], while if $o^A$ is repeated four times then the fifth and sixth expressions of [18] should be used.

For the converse, suppose that $\kappa = \sigma = 0$. Then, by the first equation in [14], $o^A$ can be rescaled to ensure that $\epsilon = 0$ and a spinor field $\iota^A$ can be chosen which is normalized against $o^A$ and parallelly propagated along $\ell^a$, so that, by the fifth equation in [14], $\pi = 0$. From the second expression in [17], one can see at once that $\Psi_0 = 0$, so that the first two equations in [18] simplify to give expressions for $D\Psi_1$ and $\delta\Psi_1$. By commuting $D$ and $\delta$ on $\Psi_1$ and using the second expression of [15] with the relevant parts of [17], it can be concluded that $\Psi_1 = 0$, as required.

Another application which is easy to describe is the solution of the type-D vacuum equations. A type-D solution is one for which the Weyl spinor has two (linearly independent) repeated principal spinors. If these are taken as the normalized dyad, then from [16] only $\Psi_2$ is nonzero among the $\Psi_k$. By the Goldberg–Sachs theorem, both spinors are geodesic and shear free, so that the spin coefficients $\sigma, \kappa, \lambda$, and $\nu$ all vanish. With these conditions, the spin-coefficient equations simplify to the point that careful choices of coordinates and the remaining freedom in the dyad enable the equations to be solved explicitly. One obtains metrics that depend only on a few parameters. Analogous methods reduce the Einstein equations to simpler systems for the other vacuum algebraically special metrics, that is, the other vacuum metrics for which the Weyl spinor does not have four distinct principal null directions (Mason 1998).

The spin-coefficient formalism has also been extensively used in the study of asymptotically flat spacetimes and gravitational radiation (Penrose and Rindler 1984, 1986, Stewart 1990).

## The Positive-Mass Theorem

A very important application of spinor calculus in recent years was the proof by Witten (1981) of the positive-mass (or positive-energy) theorem. The proof was motivated by ideas from supergravity and gave rise to an increased interest in spinors in general relativity.

The positive-mass theorem is the following assertion: *given an asymptotically flat spacetime $\mathcal{M}$ with a spacelike hypersurface $\Sigma$, which is topologically $\mathbb{R}^3$ and in which the dominant energy condition holds, the total (or Arnowitt–Deser–Misner (ADM)) momentum is timelike and future-pointing*. (The dominant-energy condition is the requirement that $T_{ab}U^aV^b$ is non-negative for every pair of future-pointing timelike or null vectors $U^a$ and $V^b$.)

We follow the notation of Penrose and Rindler (1984, 1986), where the proof begins by considering the 2-form $\Xi$ defined in terms of a spinor field $\lambda^A$ on $\Sigma$ by

$$\Xi = -\mathrm{i}\bar{\lambda}_{B'}\nabla_a\lambda_B \mathrm{d}x^a \wedge \mathrm{d}x^b$$

If $\lambda^a$ tends to a constant spinor at spatial infinity on $\Sigma$, then

$$\frac{1}{4\pi G}\oint_S \Xi \to p_a\lambda^A\bar{\lambda}^{A'} \qquad [19]$$

as the spacelike spherical surface $S$ tends to spatial infinity, where $p_a$ is the ADM momentum. Suppose $\Sigma$ has unit normal $t^a$, intrinsic metric $h_{ab} = g_{ab} - t_at_b$ and the dual-volume 3-form is $\mathrm{d}\Sigma^a = t^a\mathrm{d}\Sigma$. Then Stokes' theorem states that

$$\oint_S \Xi = \int_\Sigma \mathrm{d}\Xi$$

We calculate

$$\mathrm{d}\Xi = \alpha + \beta$$

where

$$\alpha = 4\pi G T_{ab}\ell^a \mathrm{d}\Sigma^b$$
$$\beta = -\mathrm{i}\epsilon_{ab}{}^{cd}\nabla_c\lambda^B\nabla_d\bar{\lambda}^{B'}\mathrm{d}\Sigma^a$$

where $\ell^a = \lambda^a\bar{\lambda}^{A'}$ and we have used the Einstein field equations to replace curvature terms in $\alpha$ by the energy–momentum tensor $T_{ab}$. Provided the matter satisfies the dominant-energy condition, $\alpha$ is everywhere a positive multiple of the volume form on $\Sigma$ and its integral is positive (it can vanish only in vacuum). To make the integral of $\beta$ positive, $\lambda^A$ is required to satisfy

$$D_{AA'}\lambda^A = 0 \qquad [20]$$

where $D_a = h_a^b\nabla_b$, which is the projection of the four-dimensional covariant derivative rather than the intrinsic covariant derivative of $\Sigma$. Equation [20] is the Sen–Witten equation; it is elliptic and reduces to

the Dirac equation on a maximal surface; furthermore, given an asymptotically constant value for $\lambda^A$ on an asymptotically flat 3-surface $\Sigma$ with the topology of $\mathbb{R}^3$, it has a unique solution. Equation [20] removes part of the derivative of $\lambda^A$ from $\beta$ to leave

$$\beta = -h^{ab} D_a \lambda_C D_b \bar{\lambda}_{C'} \mathrm{d}\Sigma^c$$

Now $h_{ab}$ is negative definite and $\Sigma$ has timelike normal so that $\beta$ is a positive multiple of the volume form on $\Sigma$ (unless $\lambda^A$ is covariantly constant, a case which is dealt with separately). Thus, the integral of $\mathrm{d}\Xi$ is non-negative and therefore, by [19], so is the inner product of the ADM momentum $p_a$ with any null vector constructed from asymptotically constant spinors. Furthermore, this inner product is strictly positive, except in a vacuum spacetime admitting a constant spinor. Such spacetimes can be found explicitly and cannot be asymptotically flat, so that the ADM momentum is always timelike and future pointing, and vanishes only in flat spacetime.

The basic positive-energy theorem outlined above can be extended in several directions:

- to prove that the total momentum at future null infinity is also timelike and future pointing;
- to deal with surfaces $\Sigma$ which have inner boundaries, for example, at black holes;
- to prove inequalities between charge and mass; and
- to deal with spacetimes which are asymptotically anti-de Sitter rather than flat.

## Further Applications of Spinors

Supersymmetry is a symmetry in quantum field theory relating bosons and fermions. In the language of spinors, bosons are represented by fields with an even number of spinor indices and fermions by fields with an odd number of indices. Thus, the gauge transformations of supersymmetry are generated by spinors with a single index.

Supergravity is supersymmetry in the case that one of the fields is the graviton. A supergravity theory is labeled by an integer $N$ for the number of independent supersymmetries and much of the numerology of these theories follows from properties of spinors. $N=1$ supergravity contains a graviton and a spin-3/2 field coupled together, and the presence of the supersymmetry allows the Buchdahl condition to be evaded. Supergravity theory with one supersymmetry in 11 spacetime dimensions depends on one spinor, which, in 11 dimensions, has 32 components. This is as many components as eight Dirac spinors in a four-dimensional spacetime, and, by a process of dimensional reduction, $N=1$ supergravity in 11 dimensions is related to $N=8$ supergravity in four dimensions. For reasons related to the Buchdahl conditions, 8 is the largest $N$ that is considered in four dimensions.

In superstring theory and in some supergravity theories, one often wishes to consider spaces with "residual supersymmetry," by which is meant that there is a spinor field satisfying a condition of covariant constancy in some connection (Candelas *et al.* 1985). The existence of such constant spinors, as a result of spinor Ricci identities analogous to those given above, typically imposes strong restrictions on the curvature. Riemannian manifolds admitting constant spinors for the Levi-Civita connection are Ricci-flat (Hitchin 1974); Lorentzian ones can often be found in terms of a few functions. Manifolds of special holomorphy, which are of interest in superstring theory, can usually be characterized as admitting special spinors (Wang 1989).

*See also:* Clifford Algebras and Their Representations; Dirac Operator and Dirac Field; Einstein Equations: Exact Solutions; Einstein's Equations with Matter; General Relativity: Overview; Geometric Flows and the Penrose Inequality; Index Theorems; Relativistic Wave Equations Including Higher Spin Fields; Spacetime Topology, Causal Structure and Singularities; Supergravity; Twistor Theory: Some Applications [in Integrable Systems, Complex Geometry and String Theory]; Twistors.

## Further Reading

Benn IM and Tucker RW (1987) *An Introduction to Spinors and Geometry with Applications in Physics.* Bristol: Adam Hilger.

Budinich P and Trautman A (1988) *The Spinorial Chessboard.* Berlin: Springer.

Candelas P, Horowitz G, Strominger A, and Witten E (1985) Vacuum configurations for superstrings. *Nuclear Physics B* 258: 46–74.

Cartan E (1981) *The Theory of Spinors.* New York: Dover.

Chevalley CC (1954) *The Algebraic Theory of Spinors.* New York: Columbia University Press.

Dirac PAM (1928) The quantum theory of the electron. *Proceedings of the Royal Society of London* A 117: 610–624.

Harvey FR (1990) *Spinors and Calibrations.* Boston: Academic Press.

Hitchin NJ (1974) Harmonic spinors. *Advances in Mathematics.* 14: 1–55.

Mason LJ (1998) The asymptotic structure of algebraically special spacetimes. *Classical Quantum Gravity* 15: 1019–1030.

Penrose R and Rindler W (1984, 1986) *Spinors and Space–Time.* vol. 1 and 2. Cambridge: Cambridge University Press.

Stewart J (1990) *Advanced General Relativity.* Cambridge: Cambridge University Press.

van der Waerden BL (1960) Exclusion principle and spin. In: Fierz M and Weisskopf VF (eds.) *Theoretical Physics in the Twentieth Century: A Memorial Volume to Wolfgang Pauli*, pp. 199–244. New York: Interscience.

Wang MY (1989) Parallel spinors and parallel forms. *Annals of Global Analysis and Geometry* 7: 59–68.

Witten E (1981) A new proof of the positive energy theorem. *Communications in Mathematical Physics* 80: 381–402.

# Stability of Flows

**S Friedlander**, University of Illinois-Chicago, Chicago, IL, USA



## Introduction

This article gives a brief discussion of a topic with an enormous literature, namely the stability/instability of fluid flows. Following the seminal observations and experiments of Reynolds in 1883, the issue of stability of a fluid flow became one of the central problems in fluid dynamics: stable flows are robust under inevitable disturbances in the environment, while unstable flows may break up, sometimes rapidly. These possibilities were demonstrated in a relatively simple experiment where flow in a pipe is examined at increasing speeds. As a dimensionless parameter (now known as the Reynolds number) increases, the flow completely changes its nature from a stable flow to a completely different regime that is irregular in space and time. Reynolds called this "turbulence" and observed that the transition from the simple flow to the chaotic flow was caused by the phenomenon of instability.

Even though the topic has been the subject of intense study over more than a century, Reynolds experiment is still not fully explained by current theory. Although there is no rigorous proof of stability of the simple flow (known as Poiseuille flow in a circular pipe), analytical and numerical investigations of the equations suggest theoretical stability for all Reynolds numbers. However, experiments show instability for sufficiently large Reynolds numbers. A plausible explanation for this phenomenon is the instability of such flows with respect to small but finite disturbances combined with their stability to infinitesimal disturbances.

The issue of fluid stability, in contexts much more complex than the fundamental experiment of Reynolds, arises in a multitude of branches of science, including engineeering, physics, astrophysics, oceanography, and meteorology. It is far beyond the scope of this short article to even touch upon most of the extensive literature. In the bibliography we list just a few of the substantive books where classical results can be found (Chandrasekhar 1961, Drazin and Reid 1981, Gershuni and Zhukovitiskii 1976, Joseph 1976, Lin 1967, Swinney and Gollub 1985). Recent extensive bibliographies on mathematical aspects of fluid instability are given in several articles in the *Handbook of Mathematical Fluid Dynamics* (Friedlander and Serre 2003) and the compendium of articles on hydrodynamics and nonlinear instabilities in Godreche and Maneville (1998).

## The Equations of Motion

The Navier–Stokes equations for the motion of an incompressible, constant density, viscous fluid are

$$\frac{\partial \boldsymbol{q}}{\partial t} + (\boldsymbol{q} \cdot \nabla)\, \boldsymbol{q} = -\frac{1}{\rho}\, \nabla P + \nu\, \nabla^2 \boldsymbol{q} \qquad [1a]$$

$$\operatorname{div} \boldsymbol{q} = 0 \qquad [1b]$$

where $\boldsymbol{q}(\boldsymbol{x}, t)$ denotes the velocity vector, $P(\boldsymbol{x}, t)$ the pressure, and the constants $\rho$ and $\nu$ are the density and kinematic viscosity, respectively. This system is considered in three (or sometimes two) spatial dimensions with a specified initial velocity field

$$\boldsymbol{q}(\boldsymbol{x}, 0) = \boldsymbol{q}_0(\boldsymbol{x}) \qquad [1c]$$

and physically appropriate boundary conditions: for example, zero velocity on a rigid boundary, or periodicity conditions for flow on a torus. This nonlinear system of partial differential equations (PDEs) has proved to be remarkably challenging, and in three dimensions the fundamental issues of existence and uniqueness of physically reasonable solutions are still open problems.

It is often useful to consider the Navier–Stokes equations in nondimensional form by scaling the velocity and length by some intrinsic scale in the problem, for example, in Reynolds' experiment by the mean speed $U$ and the diameter of the pipe $d$. This leads to the nondimensional equations

$$\frac{\partial \boldsymbol{q}}{\partial t} + (\boldsymbol{q} \cdot \nabla)\, \boldsymbol{q} = -\nabla P + \frac{1}{R}\, \nabla^2 \boldsymbol{q} \qquad [2a]$$

$$\operatorname{div} \boldsymbol{q} = 0 \qquad [2b]$$

where the Reynolds number $R$ is

$$R = Ud/\nu \qquad [3]$$

In many situations, the size of $R$ has a crucial influence on stability. Roughly speaking, when $R$ is small the flow is very sluggish and likely to be stable. However, the effects of viscosity are actually very complicated and not only is viscosity able to smooth and stabilize fluid motions, sometimes it actually also destroys and destabilizes flows.

The Euler equations, which predate the Navier–Stokes equations by many decades, neglect the effects of viscosity and are obtained from [1a] by setting the viscosity parameter $\nu$ to zero. Since this

removes the highest-derivative term from the equations, the nature of the Euler equations is fundamentally different from that of the Navier–Stokes equations and the limit of vanishing viscosity (or infinite Reynolds number) is a very singular limit. Since all real fluids are at least very weakly viscous, it could be argued that only the the Navier–Stokes equations are physically relevant. However, many important physical phenomena, such as turbulence, involve flows at very high Reynolds numbers ($10^4$ or higher). Hence, an understanding of turbulence is likely to involve the asymptotics of the Navier–Stokes equations as $R \to \infty$. The first step towards the construction of such asymptotics is the study of inviscid fluids governed by the Euler equations:

$$\frac{\partial \boldsymbol{q}}{\partial t} + (\boldsymbol{q} \cdot \nabla)\, \boldsymbol{q} = -\nabla P \qquad [4a]$$

$$\operatorname{div} \boldsymbol{q} = 0 \qquad [4b]$$

Stability issues for the Euler equations are in many respects distinct from those of the Navier–Stokes equations and in this article we will briefly touch upon stability results for both systems.

## Comments on Some "Classical" Instabilities

To illustrate the complexity of the structure of instabilities that can arise in the Navier–Stokes equations, we mention one classical example, namely the centrifugal instabilities called Taylor–Couette instabilities. Consider a fluid between two concentric cylinders rotating with different angular velocities. If the inner cylinder rotates sufficiently faster than the outer one, the centrifugal force is stronger on inside particles than outside particles and a disturbance which exchanges the radial position of particles is enhanced, that is, the configuration is unstable. As the angular velocity of the inner cylinder is increased above a certain critical rate, the instability is manifested in a series of small toroidal (Taylor) vortices that fill the space between the cylinders. There follows a hierarchy of successive instabilities: azimuthal traveling waves, twisting regimes, and quasiperiodic regimes until chaotic solutions appear. Such a sequence of bifurcations is a scenario for a transition to turbulence postulated by Ruelle–Takens. Details concerning bifurcation theory and fluid behavior can be found in the book of Chossat and Iooss (1994).

We note that phenomena of successive bifurcations connected with loss of stability, such as regimes of Taylor–Couette instabilities, occur at moderately large Reynolds numbers. Fully developed turbulence is a phenomenon associated with very high Reynolds numbers. These are parameter regimes basically inaccessible in current numerical investigations of the Navier–Stokes equations and turbulent models. The Euler equations lie at the limit as $R \to \infty$. It is an interesting observation that results at the limit of infinite Reynolds number are sometimes also applicable and consistent with experiments for flows with only moderate Reynolds number.

There is a huge diversity of forces that couple with fluid motion to produce instability. We will merely mention a few of these which an interested reader could pursue in consultation with texts listed in the "Further reading" section and references therein.

1. The so-called Bénard problem of convective instability concerns a horizontal layer of fluid between parallel plates and subject to a temperature gradient. The governing equations are the Navier–Stokes equation for a nonconstant density fluid and the heat equation. In this problem, the critical parameter governing the onset of instability is called the Rayleigh number. The patterns that can develop as a result of instability are strongly influenced by the boundary conditions in the horizontal coordinates. With lattice type conditions, bifurcating solutions include rolls, rectangles, and hexagons. Convection rolls are themselves subject to secondary instabilities that may break the translation symmetry and deform the rolls into meandering shapes. Further refinements of convective instabilities include doubly diffusive convection, where the density depends on concentration as well as temperature. Competition between stabilizing diffusivity and destabilizing diffusivity can lead to the so-called "salt-finger" instabilities.
2. Of considerable interest in astrophysics and plasma physics are the instabilities that occur in electrically conducting fluilds. Here the fluid equations are coupled with Maxwell's equations. Much work has been done on the topic of magnetohydrodynamical (MHD) stability, which was developed to address various important physical issues such as thermonuclear fusion, stellar and planetary interiors, and dynamo theory. For example, dynamo theory addresses the issue of how a magnetic field can be generated and sustained by the motion of an electrically conducting fluid. In the simplest scenario, the fluid motion is assumed to be a given divergence-free vector field and the study of

the instabilities that may occur in the evolution of the magnetic field is called the kinematic dynamo problem. This gives rise to interesting problems in dynamical systems and actually is closely analogous to the topic of vorticity generation in the three-dimensional (3D) fluid equations in the absence of MHD effects.

In the next section we discuss certain mathematical results that have been rigorously proved for particular problems in the stability of fluid flows. We restrict our attention to the "basic" equations, that is, [2a] and [2b], [4a] and [4b], observing that even in rather simple configurations there are still more open problems than precise rigorous results.

## The Navier–Stokes Equations: Mathematical Definitions of Stability/Instability

Instability occurs when there is some disturbance of the internal or external forces acting on the fluid and, loosely speaking, the question of stability or instability considers whether there exist disturbances that grow with time. There are many mathematical definitions of stability of a solution to a PDE. Most of these definitions are closely related but they may not be equivalent. Because of the distinctly different nature of the Navier–Stokes equations for a viscous fluid and the Euler equations for an inviscid fluid, we will adopt somewhat different precise definitions of stability for the two systems of PDEs. Both definitions are related to the concept known as Lyapunov stability. A steady state described by a velocity field $\boldsymbol{U}_0(\boldsymbol{x})$ is called Lyapunov stable if every state $\boldsymbol{q}(\boldsymbol{x},t)$ "close" to $\boldsymbol{U}_0(\boldsymbol{x})$ at $t=0$ stays close for all $t>0$. In mathematical terms, "closeness" is defined by considering metrics in a normed space $X$. While in finite-dimensional systems the choice of norm is not significant because all Banach norms are equivalent, in infinite-dimensional systems, such as a fluid configuration, this choice is crucial. The point was emphasized by Yudovich (1989) and it is a version of the definition of stability given in this book that we will adopt in connection with the parabolic Navier–Stokes equations.

### Definitions for a General Nonlinear Evolution Equation

Consider an evolution equation for $\boldsymbol{u}(\boldsymbol{x},t)$ whose phase space is a Banach space $X$:

$$\frac{\partial \boldsymbol{u}}{\partial t} = L\boldsymbol{u} + N(\boldsymbol{u},\boldsymbol{u})$$

We assume that if the initial value $\boldsymbol{u}(\boldsymbol{x},0) \in X$ is given, the future evolution $\boldsymbol{u}(\boldsymbol{x},t), t>0$, of the equation is uniquely defined (at least for sufficiently small initial data). Without loss of generality, we can assume that zero is a steady state.

We define a version of Lyapunov (nonlinear) stability and its converse instability.

**Definition 1** Let $(X, Z)$ be a pair of Banach spaces. The zero steady state is called $(X, Z)$ nonlinearly stable if, no matter how small $\epsilon > 0$, there exists $\delta > 0$ so that $\boldsymbol{u}(\boldsymbol{x},0) \in X$ and

$$\|\boldsymbol{u}(\boldsymbol{x},0)\|_Z < \delta$$

imply the following two assertions:

(i) there exists a global in time solution such that $\boldsymbol{u}(\boldsymbol{x},t) \in (([0,\infty); X)$;
(ii) $\|\boldsymbol{u}(\boldsymbol{x},t)\|_Z < \epsilon$ for a.e. $t \in [0,\infty)$.

The zero state is called nonlinearly unstable if either of the above assertions is violated. We note that under this strong definition of stability, loss of existence of a solution is a particular case of instability. The concept of existence that we will invoke in considering the Navier–Stokes equations is the existence of "mild" solutions introduced by Kato and Fujita (1962). Local-in-time existence of mild solutions is known in $X = L^q$ for $q \geq n$, where $n$ denotes the space dimension. ($L^q$ denotes the usual Lebesque space).

We now state two theorems for the Navier–Stokes equations [2a] and [2b]. The theorems are valid in any space dimension $n$ and in finite or infinite domains. Of course, the most physically relevant cases are $n=3$ or 2. Both theorems relate properties of the spectrum of the linearized Navier–Stokes equations to stability or instability of the full nonlinear system. Let $\boldsymbol{U}_0(\boldsymbol{x})$, $P_0(\boldsymbol{x})$ be a steady state flow:

$$(\boldsymbol{U}_0 \cdot \nabla)\boldsymbol{U}_0 = -\nabla P_0 + \frac{1}{R}\nabla^2 \boldsymbol{U}_0 + \frac{1}{R}\boldsymbol{F} \qquad [5a]$$

$$\nabla \cdot \boldsymbol{U}_0 = 0 \qquad [5b]$$

where $\boldsymbol{U}_0 \in C^\infty$ vanishes on the boundary of the domain $D$ and $\boldsymbol{F}$ is a suitable external force. We write [2a] and [2b] in perturbation form as

$$\boldsymbol{q}(\boldsymbol{x},t) = \boldsymbol{U}_0(\boldsymbol{x}) + \boldsymbol{u}(\boldsymbol{x},t) \qquad [6]$$

where

$$\frac{\partial \boldsymbol{u}}{\partial t} = L_{\mathrm{NS}}\boldsymbol{u} + N(\boldsymbol{u},\boldsymbol{u}) \qquad [7a]$$

$$\nabla \cdot \boldsymbol{u} = 0 \qquad [7b]$$

with

$$L_{NS}\boldsymbol{u} \equiv -(\boldsymbol{U}_0 \cdot \nabla)\boldsymbol{u} - (\boldsymbol{u} \cdot \nabla)\boldsymbol{U}_0 + \frac{1}{R}\nabla^2\boldsymbol{u} - \nabla P_1 \qquad [8]$$

$$N(\boldsymbol{u}, \boldsymbol{u}) \equiv -(\boldsymbol{u} \cdot \nabla)\boldsymbol{u} - \nabla P_2 \qquad [9]$$

Here $P_1$ and $P_2$ are, respectively, the portions of the pressure required to ensure that $L_{NS}u$ and $N(\boldsymbol{u}, \boldsymbol{u})$ remain divergence free. The operators $L_{NS}$ and $N$ act on the space of divergence-free vector-valued functions in the closure of the Sobolev space $W^{s,p}$ that vanish on the boundary of $D$.

We note that the spectrum of the elliptic linear operator $L_{NS}$ with appropriate boundary conditions in a bounded domain is purely discrete: that is, it consists of a countable number of eigenvalues of finite multiplicity with the sole limit point being at infinity.

**Theorem 2** (Nonlinear instability). *Let $1 < p < \infty$ be arbitrary. Suppose that the operator $L_{NS}$ over $L^p$ has spectrum in the right half of the complex plane. Then the flow $\mathbf{U}_0(\boldsymbol{x})$ is $(L^q, L^p)$ nonlinearly unstable for any $q > \max(p, n)$.*

**Theorem 3** (Asymptotic Lyapunov stability). *Let $q > n$ be arbitrary. Assume that the operator $L_{NS}$ over $L^q$ has spectrum confined to the left half of the complex plane. Then the flow $\mathbf{U}_0(\boldsymbol{x})$ is $(L^q, L^q)$ nonlinearly stable.*

A recent proof of these theorems is given in Friedlander *et al.* (2006) using a bootstrap type argument. In Theorem 2, the space $L^q, q > n$, is used as an auxiliary space inwhich the norm of the nonlinear term is controlled, while the final instability result is proved in $L^p$ for $p \in (1, \infty)$. We note that this includes the most physically relevant case of instability in the $L^2$ energy norm. An earlier proof of the theorems under the restriction $p \geq n$ was given by Yudovich (1989).

To apply Theorem 2 or 3 to conclude nonlinear instability or stability of a given flow $\boldsymbol{U}_0$, it is necessary to have information concerning the spectrum of the linear operator $L_{NS}$. Obtaining such information has been the goal of much of the literature concerning fluid stability (see the bibliography and the references therein). However, except in the case of some relatively simple flows, the eigenvalues of $L_{NS}$ have not yet been calculated explicitly. Perhaps the example that is the most tractable is plane parallel shear flows. Here the eigenvalue problem is governed by an ordinary differential equation (ODE) known as the Orr–Sommerfeld equation, which has been the subject of extensive analytical and numerical investigations. Consider the parallel flow $\boldsymbol{U}_0 = (U(z), 0, 0)$ in the strip $-1 \leq z \leq 1$. For disturbances of the form

$$\phi(z)\, e^{i(k_1 x + k_2 y)}\, e^{\lambda t} \qquad [10]$$

the eigenvalue $\lambda$ is determined by the following equation with $k^2 = k_1^2 + k_2^2$:

$$\left(U - i\frac{\lambda}{k}\right)\left[\frac{d^2}{dz^2} - k^2\right]\phi - U''\phi = \frac{1}{ikR}\left[\frac{d^2}{dz^2} - k^2\right]\phi \qquad [11]$$

with boundary conditions $\phi = 0$ at $z = \pm 1$. We note that the discreteness of the spectrum is preserved if periodicity conditions are imposed in the $(x, y)$ plane.

The complexity of the spectral problem [11] is apparent even for the simple case $U(z) = 1 - z^2$ (known as plane Poiseuille flow). Unstable eigenvalues exist but only in certain regions of $(k, R)$ parameter space. There is a critical Reynolds number, $R_c = 5772$, below which $\text{Re}\,\lambda < 0$ for all wave numbers $k$. For $R > R_c$, instability occurs in a band of wave numbers and the thickness of this band shrinks to zero as $R \to \infty$ (i.e., the inviscid limit). Hence, Poiseuille flow with $R < R_c$ can be considered as an example where the stability Theorem 3 can be applied, that is, the flow is nonlinearly stable to infinitesimal disturbances. However, extremely careful experiments are needed to obtain agreement with the theoretical value of $R_c = 5772$. Rather it is more usual in an experiment with $R \sim 2000$ that the flow exhibits instability in the form of streamwise streaks that appear near the walls. These structures do not look like traveling waves of the form given by expression [10], rather they are finite-amplitude effects of nonmodal growth. Such linear growth of disturbances, along with energy growth and pseudospectra have recently been investigated extensively.

An example where Theorem 3, proving nonlinear instability, can be applied is the so-called Kolmogorov flow. This is also a shear flow with the spectral problem for the linearized operator given by eqn [11]. In this example, the profile is oscillatory in $z$ with $U(z) = \sin mz$. In an elegant paper, Meshalkin and Sinai (1961) used continued fractions to prove the existence of a real unstable positive eigenvalue. It is interesting, and in some sense surprising, that the particular case of sinusoidal profiles leads to a nonconstant-coefficient eigenvalue problem, where it is possible to construct in explicit form the transcendental characteristic equation that relates the eigenvalues $\lambda$ and the wave numbers. Usually,

this can be done only for constant-coefficient equations. In the case $U(z) = \sin mz$, a Fourier series representation for the eigenfunctions leads to a tridiagonal infinite matrix for the algebraic system satisfied by the Fourier coefficients. This is amenable to examination using continued fractions. Analysis of the characteristic equation shows that there exist real eigenvalues $\lambda > 0$ provided $R$ is larger than some critical value for each wave number $k$ with $k^2 < m^2$.

## The Euler Equation: Linear and Nonlinear Stability/Instability

We conclude this brief article with some discussion of instabilities in the inviscid Euler equations whose existence is likely to be important as a "trigger" for the development of instabilities in high-Reynolds-number viscous flows. As we mentioned, the Euler equations are very different from the Navier–Stokes equations in their mathematical structure. The Euler equations are degenerate and nonelliptic. As such, the spectrum of the linearized operator $L_{\mathrm{E}}$ is not amenable to standard spectral theory of elliptic operators. For example, unlike the Navier–Stokes operator, the spectrum of $L_{\mathrm{E}}$ is not purely discrete even in bounded domains. To define $L_{\mathrm{E}}$ we consider a steady Euler flow $\{\boldsymbol{U}_0(\boldsymbol{x}), P_0(\boldsymbol{x})\}$, where

$$\boldsymbol{U}_0 \cdot \nabla \boldsymbol{U}_0 = -\nabla P_0 \qquad [12a]$$

$$\nabla \cdot \boldsymbol{U}_0 = 0 \qquad [12b]$$

We assume that $\boldsymbol{U}_0 \in C^\infty$. For the Euler equations, appropriate boundary conditions include zero normal component of $\boldsymbol{U}_0$ on a rigid boundary, or periodicity conditions (i.e., flow on a torus) or suitable decay at infinity in an unbounded domain. The theorems that we will be describing have been proved mainly in the cases of the second and third conditions stated above. There are many classes of vector fields $\boldsymbol{U}_0(\boldsymbol{x})$, in two and three dimensions, that satisfy [12a] and [12b]. We write [4a] and [4b] in perturbation form as

$$\boldsymbol{q}(\boldsymbol{x}, t) = \boldsymbol{U}_0(\boldsymbol{x}) + \boldsymbol{u}(\boldsymbol{x}, t) \qquad [13]$$

with

$$\frac{\partial \boldsymbol{u}}{\partial t} = L_{\mathrm{E}}\, \boldsymbol{u} + N(\boldsymbol{u}, \boldsymbol{u}) \qquad [14a]$$

$$\nabla \cdot \boldsymbol{u} = 0 \qquad [14b]$$

Here

$$L_{\mathrm{E}}\, \boldsymbol{u} \equiv -(\boldsymbol{U}_0 \cdot \nabla)\, \boldsymbol{u} - (\boldsymbol{u} \cdot \nabla) \boldsymbol{U}_0 - \nabla P_1 \qquad [15]$$

$$N(\boldsymbol{u}, \boldsymbol{u}) \equiv -(\boldsymbol{u} \cdot \nabla)\, \boldsymbol{u} - \nabla P_2 \qquad [16]$$

Linear (spectral) instability of a steady Euler flow $\boldsymbol{U}_0(\boldsymbol{x})$ concerns the structure of the spectrum of $L_{\mathrm{E}}$. Assuming $\boldsymbol{U}_0 \in \mathrm{C}^\infty(T^n)$, the linear equation

$$\frac{\partial \boldsymbol{u}}{\partial t} = L_{\mathrm{E}}\, \boldsymbol{u}, \qquad \nabla \cdot \boldsymbol{u} = 0 \qquad [17]$$

defines a strongly continuous group in every Sobolev space $W^{s,p}$ with generator $L_{\mathrm{E}}$. We denote this group by $\exp\{L_{\mathrm{E}}\, t\}$. For the issue of spectral instability of the Euler equation it proves useful to study not only the spectrum of $L_{\mathrm{E}}$ but also the spectrum of the evolution operator $\exp\{L_{\mathrm{E}}\, t\}$. This permits the development of an explicit formula for the growth rate of a small perturbation due to the essential (or continuous) spectrum. It was proved by Vishik (1996) that a quantity $\Lambda$, refered to as a "fluid Lyapunov exponent" gives the maximum growth rate of the essential spectrum of $\exp\{L_{\mathrm{E}}\, t\}$. This quantity is obtained by computing the exponential growth rate of a certain vector that satisfies a specific system of ODEs over the trajectories of the flow $\boldsymbol{U}_0(\boldsymbol{x})$. This proves to be an effective mechanism for detecting instabilities in the essential spectrum which result due to high-spatial-frequency perturbations. For example, for this reason any flow $\boldsymbol{U}_0(\boldsymbol{x})$ with a hyperbolic fixed point is linearly unstable with growth in the sense of the $L^2$-norm. In two dimensions, $\Lambda$ is equal to the maximal classical Lyapunov exponent (i.e., the exponential growth of a tangent vector over the ODE $\dot{\boldsymbol{x}} = \boldsymbol{U}_0(\boldsymbol{x})$). In three dimensions, the existence of a nonzero classical Lyapunov exponent implies that $\Lambda > 0$. However, in three dimensions there are also examples where the classical Lyapunov exponent is zero and yet $\Lambda > 0$. We note that the delicate issue of the unstable essential spectrum is strongly dependent on the function space for the perturbations and that $\Lambda$, for a given $\boldsymbol{U}_0$, will vary with this function space. More details and examples of instabilities in the essential spectrum can be found in references in the bibliography.

In contrast with instabilities in the essential spectrum, the existence of discrete unstable eigenvalues is independent of the norm in which growth is measured. From this point of view, such instabilities can be considered as "strong." However, for most flows $\boldsymbol{U}_0(\boldsymbol{x})$ we do not know the existence of such unstable eigenvalues. For fully 3D flows there are no examples, to our knowledge, where such unstable eigenvalues have been proved to exist for flows with standard metrics. The case that has received the most attention in the literature is the "relatively simple" case of plane parallel shear flow. The eigenvalue problem is governed by the Rayleigh

equation (which is the inviscid version of the Orr–Sommerfeld equation [11]):

$$\left(U - \frac{\mathrm{i}\lambda}{k}\right)\left[\frac{\mathrm{d}^2}{\mathrm{d}z^2} - k^2\right]\phi - U''\phi = 0$$
$$\phi = 0 \quad \text{at} \quad z = \pm 1 \qquad [18]$$

The celebrated Rayleigh stability criterion says that a sufficient condition for the eigenvalues $\lambda$ to be pure imaginary is the absence of an inflection point in the shear profile $U(z)$. It is more difficult to prove the converse; however, there have been several recent results that show that oscillating profiles indeed produce unstable eigenvalues. For example, if $U(z) = \sin mz$ the continued fraction proof of Meshalkin and Sinai can be adapted to exhibit the full unstable spectrum for [18]. We note the "fluid Lyapunov exponent" $\Lambda$ is zero for all shear flows; thus the only way the unstable spectrum can be nonempty for shear flows is via discrete unstable eigenvalues.

As we have discussed, it is possible to show that many classes of steady Euler flows are linearly unstable, either due to a nonempty unstable essential spectrum (i.e., cases where $\Lambda > 0$) or due to unstable eigenvalues or possibly for both reasons. It is natural to ask what this means about the stability/instability of the full nonlinear Euler equations [14]–[16]. The issue of nonlinear stability is complex and there are several natural precise definitions of nonlinear stability and its converse instability.

One definition is to consider nonlinear stability in the energy norm $L^2$ and the enstrophy norm $H^1$, which are natural function spaces to measure growth of disturbances but are not "correct" spaces for the Euler equations in terms of proven properties of existence and uniqueness of solutions to the nonlinear equation. Falling under this definition is the most frequently employed method to prove nonlinear stability, which is an elegant technique developed by Arnol'd (cf. Arnol'd and Khesin (1998) and references therein). This is based on the existence of the so-called energy-Casimirs. The vorticity curl $q$ is transported by the motion of the fluid so that at time $t$ it is obtained from the vorticity at time $t = 0$ by a volume-preserving diffeomorphism. In the terminology of Arnol'd, the velocity fields obtained in this manner at any two times are called isovortical. For a given field $\boldsymbol{U}_0(\boldsymbol{x})$, the class of isovortical fields is an infinite-dimensional manifold $M$, which is the orbit of the group of volume-preserving diffeomorphisms in the space of divergence-free vector fields. The steady flows are exactly the critical points of the energy functional $E$ restricted to $M$. If a critical point is a strict local maximum or minimum of $E$, then the steady flow is nonlinearly stable in the space $J_1$ of divergence-free vectors $\boldsymbol{u}(\boldsymbol{x}, t)$ (satisfying the boundary conditions) that have finite norm,

$$\|\boldsymbol{u}\|_{J_1} \equiv \|\boldsymbol{u}\|_{L^2} + \|\text{curl } \boldsymbol{u}\|_{L^2} \qquad [19]$$

This theory can be applied, for example, to show that any shear flow with no inflection points in the profile $U(z)$ is nonlinearly unstable in the function space $J_1$, that is, the classical Rayleigh criterion implies not only spectral stability but also nonlinear stability.

We note that Arnol'd's stability method cannot be applied to the Euler equations in three dimensions because the second variation of the energy defined on the tangent space to $M$ is never definite at a critical point $\boldsymbol{U}_0(\boldsymbol{x})$. This result is suggestive, but does not prove, that most Euler flows in three dimensions are nonlinearly unstable in the Arnol'd sense. To quote Arnol'd, in the context of the Euler equations "there appear to be an infinitely great number of unstable configurations."

In recent years, there have been a number of results concerning nonlinear instability for the Euler equation. Most of these results prove nonlinear instability under certain assumptions on the structure of the spectrum of the linearized Euler operator. To date, none of the approaches prove the definitive result that in general linear instability implies nonlinear instability. As we have remarked, this is a much more delicate issue for Euler than for Navier–Stokes because of the existence, for a generic Euler flow, of a nonempty essential unstable spectrum. To give a flavor of the mathematical treatment of nonlinear instability for the Euler equations, we present one recent result and refer the interested reader to articles listed in the "Further reading" section for further results and discussions.

In the context of Euler equations in two dimensions, we adopt the following definition of Lyapunov stability.

**Definition 4** An equilibrium solution $\boldsymbol{U}_0(\boldsymbol{x})$ is called Lyapunov stable if for every $\varepsilon > 0$ there exists $\delta > 0$ so that for any divergence-free vector $\boldsymbol{u}(\boldsymbol{x}, 0) \in W^{1+s,p}$, $s > 2/p$, such that $\|\boldsymbol{u}(\boldsymbol{x}, 0)\|_{L^2} < \delta$ the unique solution $\boldsymbol{u}(\boldsymbol{x}, t)$ to [14]–[16] satisfies

$$\|\boldsymbol{u}(\boldsymbol{x}, t)\|_{L^2} < \varepsilon \quad \text{for} \quad t \in [0, \infty)$$

We note that we require the initial value $\boldsymbol{u}(\boldsymbol{x}, 0)$ to be in the Sobolev space $W^{1+s,p}$, $s > p/2$, since it is known that the 2D Euler equations are globally in time well posed in this function space.

**Definition 5** Any steady flow $U_0(x)$ for which the conditions of Definition 4 are violated is called nonlinearly unstable in $L^2$.

Observe that the open issues (in three dimensions) of nonuniqueness or nonexistence of solutions to [14]–[16] would, under Definition 5, be scenarios for instability.

**Theorem 6** (Nonlinear instability for 2D Euler flows). *Let $U_0(x) \in C^\infty(T^2)$ be satisfy [12]. Let $\Lambda$ be the maximal Lyapunov exponent to the ODE $\dot{x} = U_0(x)$. Assume that there exists an eigenvalue $\lambda$ in the $L^2$ spectrum of the linear operator $L_E$ given by [15] with $\mathrm{Re}\,\lambda > \Lambda$. Then in the sense of Definition 5, $U_0(x)$ is Lyapunov unstable with respect to growth in the $L^2$-norm.*

The proof of this result is given in Vishik and Friedlander (2003) and uses a so-called "bootstrap" argument whose origins can be found in references in that article. We remark that the above result gives nonlinear instability with respect to growth of the energy of a perturbation which seems to be a physically reasonable measure of instability.

In order to apply Theorem 6 to a specific 2D flow it is necessary to know that the linear operator $L_E$ has an eigenvalue with $\mathrm{Re}\,\lambda > \Lambda$. As we have discussed, such knowledge is lacking for a generic flow $U_0(x)$. Once again, we turn to shear flows. As we noted $\Lambda = 0$ for shear flows, any shear profile for which unstable eigenvalues have been proved to exist provides an example of nonlinear instability with respect to growth in the energy.

We conclude with the observation that it is tempting to speculate that, given the complexity of flows in three dimensions, most, if not all, such inviscid flows are nonlinearly unstable. It is clear from the concept of the fluid Lyapunov exponent that stretching in a flow is associated with instabilities and there are more mechanisms for stretching in three, as opposed to two, dimensions. However, to date there are virtually no mathematical results for the nonlinear stability problem for fully 3D flows and many challenging issues remain entirely open.

## Acknowledgments

The author is very grateful to IHES and ENS-Cachan for their hospitality during the writing of this paper. She thanks Misha Vishik for much helpful advice.

The work is partially supported by NSF grant DMS-0202767.

*See also:* Compressible Flows: Mathematical Theory; Incompressible Euler Equations: Mathematical Theory; Magnetohydrodynamics; Newtonian Fluids and Thermohydraulics; Non-Newtonian Fluids; Topological Knot Theory and Macroscopic Physics.

## Further Reading

Arnol'd VI and Khesin B (1998) *Topological Methods in Hydrodynamics*. New York: Springer.

Chandrasekhar S (1961) *Hydrodynamic and Hydromagnetic Stability*. Oxford: Oxford University Press.

Chossat P and Iooss G (1994) *The Couette–Taylor Problem*. Berlin: Springer.

Drazin PG and Reid WH (1981) *Hydrodynamic Stability*. Cambridge: Cambridge University Press.

Friedlander S, Pavlovic N, and Shvydkoy R (2006) Nonlinear instability for the Navier–Stokes equations (to appear in *Communications in Mathematical Physics*).

Friedlander S and Serre D (eds.) (2003) *Handbook of Mathematical Fluid Dynamics*, vol. 2. Amsterdam: Elsevier.

Friedlander S and Yudovich VI (1999) Instabilities in fluid motion. *Notices of the American Mathematical Society* 46: 1358–1367.

Gershuni GZ and Zhukovitiskii EM (1976) *Convective Instability of Incompressible Fluids*. Jerusalem: Keter Publishing House.

Godreche C and Manneville P (eds.) (1998) *Hydrodynamics and Nonlinear Instabilities*. Cambridge: Cambridge University Press.

Joseph DD (1976) *Stability of Fluid Motions*, 2 vols. Berlin: Springer.

Kato T and Fujita H (1962) On the nonstationary Navier–Stokes system. *Rend. Sem. Mat. Univ. Padova* 32: 243–260.

Lin CC (1967) *The Theory of Hydrodynamic Stability*. Cambridge: Cambridge University Press.

Meshalkin LD and Sinai IaG (1961) Investigation of stability for a system of equations describing the plane movement of an incompressible viscous liquid. *App. Math. Mech.* 25: 1700–1705.

Swinney H and Gollub L (eds.) (1985) *Hydrodynamic Instabilities and Transition to Turbulence*. New York: Springer.

Vishik MM (1996) Spectrum of small oscillations of an ideal fluid and Lyapunov exponents. *Journal de Mathématiques Pures et Appliquées* 75: 531–557.

Vishik M and Friedlander S (2003) Nonlinear instability in 2 dimensional ideal fluids: the case of a dominant eigenvalue. *Communications in Mathematical Physics* 243: 261–273.

Yudovich VI (1989, US translation) *Linearization Method in Hydrodynamical Stability Theory*, Transl. Math. Monog. vol. 74. Providence: American Mathematical Society.

# Stability of Matter

**J P Solovej**, University of Copenhagen, Copenhagen, Denmark



## Introduction

The theorem on stability of matter is one of the most celebrated results in mathematical physics. It is one of the rare cases where a result of such great importance to our understanding of the world around us appeared first in a completely rigorous formulation.

Issues of stability are, of course, extremely important in physics. One of the major triumphs of the theory of quantum mechanics is the explanation it gives of the stability of the hydrogen atom (and the complete description of its spectrum). Quantum mechanics or, more precisely, the uncertainty principle explains not only the stability of tiny microscopic objects, but also the stability of gigantic stellar objects such as white dwarfs. Chandrasekhar's famous theory on the stability of white dwarfs required, however, not only the usual uncertainty principle, but also the Pauli exclusion principle for the fermionic electrons.

Whereas both the stability of atoms and the stability of white dwarfs were early triumphs of quantum mechanics, it, surprisingly, took nearly 40 years before the question of stability of everyday macroscopic objects was even raised (Fisher and Ruelle 1966). The rigorous answer to the question came shortly thereafter in what came to be known as the "theorem on stability of matter" proved first by Dyson and Lenard (1967).

Both the stability of hydrogen and the stability of white dwarfs simply mean that the total energy of the system cannot be arbitrarily negative. If there were no such lower bound to the energy, one would have a system from which it would be possible, in principle, to extract an infinite amount of energy. One often refers to this kind of stability as *stability of the first kind*.

Stability of matter is somewhat different. Stability of the first kind for atoms generalizes, as noted later, to objects of macroscopic size. The question arises as to how the lowest possible energy depends on the size or, more precisely, on the (macroscopic) number of particles in the object. Stability of matter in its precise mathematical formulation is the requirement that the lowest possible energy depends at most linearly on the number of particles. Put differently, the lowest possible energy calculated per particle cannot be arbitrarily negative as the number of particles increases. This is often referred to as "stability of the second kind." If stability of the second kind does not hold, one would be able to extract an arbitrarily large amount of energy by adding a single atomic particle to a sufficiently large macroscopic object.

A perhaps more intuitive notion of stability is related to the volume occupied by a macroscopic object. More precisely, the volume of the object, when its total energy is close to the lowest possible energy, grows at least linearly in the number of particles. This volume dependence is a fairly simple consequence of stability of matter as formulated above.

The first mention of stability of the second kind for a charged system is perhaps by Onsager (1939), who studied a system of charged classical particles with a hard core and proved the stability of the second kind. The proof of stability of matter by Dyson and Lenard, which does not rely on any hard-core assumption, but rather on the properties of fermionic quantum particles, used results from Onsager's paper.

The real relevance of the notion of stability of the second kind was first realized by Fisher and Ruelle (1966) in an attempt to understand the thermodynamic properties of matter and to give meaning to thermodynamic quantities such as the energy density (energy per volume). Stability of matter is a necessary ingredient in explaining the existence of thermodynamics, that is, that the energy per volume has a well-defined limit as the volume and number of particles tend to infinity, with the ratio (i.e., the density of particles) kept fixed. The existence of this limit is, however, not just a simple consequence of stability of matter. The existence of the thermodynamic limit for ordinary charged matter was proved rigorously by Lieb and Lebowitz (1972) using the result on stability of matter as an input.

After the original proof of stability of matter by Dyson and Lenard, several other proofs were given (see, e.g., reviews by Lieb (1976, 1990, 2004) for detailed references). Lieb and Thirring (1975) in particular presented an elegant and simple proof relying on an uncertainty principle for fermions. As explained in a later section, the best mathematical formulation of the usual uncertainty principle is in terms of a Sobolev inequality. The method of Lieb and Thirring is related to a Sobolev type inequality for antisymmetric functions. The Lieb–Thirring inequality is discussed later. The proof by Dyson

and Lenard gave a very poor bound on the lowest possible energy per particle. The proof by Lieb and Thirring gave a much more realistic bound on this quantity (see below). Two proofs of stability of matter will be sketched here. Both proofs rely on the Lieb–Thirring inequality. The first proof described is mathematically simple to explain, whereas the second proof (Lieb–Thirring) is based on the Thomas–Fermi theory. It is mathematically somewhat more involved but, from a physical point of view, more intuitive.

As in the case of white dwarfs, stability of matter relies on the fermionic property of electrons. Dyson (1967) proved that the stability of the second kind fails if we ignore the Pauli exclusion principle. In physics textbooks, the importance of the Pauli exclusion principle for the stability of white dwarfs is often emphasized. Its importance for the stability of everything around us is usually ignored.

As mentioned above the result on stability of matter appeared from the beginning as a completely rigorously proved theorem. In contrast, the stability of white dwarfs was only derived rigorously by Lieb and Thirring (1984) and Lieb and Yau (1987) over 50 years after the original work of Chandrasekhar.

The original formulation of stability of matter, which is given in the next section, dealt with charged matter consisting of electrons and nuclei interacting only through electrostatic interactions and being described by nonrelativistic quantum mechanics. Over the years, many generalizations of stability of matter have been derived in order to include relativistic effects and electromagnetic interactions. Some of these generalizations will be discussed in this article. A complete understanding of stability of matter in quantum electrodynamics (QED) does not exist as yet, which is intimately related to the fact that this theory still awaits a mathematically satisfactory formulation.

## The Formulation of Stability of Matter

Consider $K$ nuclei with nuclear charges $z_1, \ldots, z_K > 0$ at positions $r_1, \ldots, r_K \in \mathbb{R}^3$, and $N$ electrons with charges –1 (this amounts to a choice of units) at positions $x_1, \ldots, x_N \in \mathbb{R}^3$. In order to discuss stability, it turns out that one can consider the nuclei as fixed in space, whereas the electrons are dynamic. More precisely, this means that the kinetic energy of the nuclei is ignored. It is important to realize that if stability holds for static nuclei, it also holds for dynamic nuclei. This is simply because the kinetic energy is positive, so that the effect of ignoring it is to lower the total energy.

Since we consider only electrostatic interactions, the quantum Hamiltonian describing this system is

$$H_N = \sum_{i=1}^{N} T_i - \sum_{k=1}^{K}\sum_{i=1}^{N} \frac{z_k}{|x_i - r_k|} + \sum_{1\le i<j\le N} \frac{1}{|x_i - x_j|} + \sum_{1\le k<\ell\le K} \frac{z_k z_\ell}{|r_k - r_\ell|} \quad [1]$$

The kinetic energy operator $T_i$ is (half) the Laplacian in the variable $x_i$, i.e., $T_i = -(1/2)\Delta_i$. Atomic units are used, where not only the electron charge is –1, but the mass of the electron is also 1 and $\hbar = 1$. The unit of energy is then 2 Ry.

The Hamiltonian $H_N$ depends on the parameters $\boldsymbol{z} = (z_1, \ldots, z_K)$ and $\boldsymbol{r} = (r_1, \ldots, r_K)$. It acts on the Hilbert space of fermionic, that is, antisymmetric wave functions. More precisely, the fermionic Hilbert space is

$$\mathcal{H}_N^{\mathrm{F}} = \bigwedge^{N} L^2(\mathbb{R}^3; \mathbb{C}^2)$$

Here the target space is $\mathbb{C}^2$, in order to describe spin-1/2 particles. One can, of course, also consider the Hamiltonian $H_N$ on the full Hilbert space,

$$\mathcal{H}_N = \bigotimes^{N} L^2(\mathbb{R}^3; \mathbb{C}^2) = L^2(\mathbb{R}^{3N}; \mathbb{C}^{2^N})$$

of which $\mathcal{H}_N^{\mathrm{F}}$ is a subspace.

The quantity of interest is the ground-state energy

$$\begin{aligned} E^{\mathrm{F}}(\boldsymbol{z}, N, K) &= \inf_{\boldsymbol{r}} \inf \operatorname{spec}_{\mathcal{H}_N^{\mathrm{F}}} H_N \\ &= \inf_{\boldsymbol{r}} \inf \Big\{ (\Psi, H_N \Psi) \mid \Psi \\ &\qquad \in \mathcal{H}_N^{\mathrm{F}} \cap C^\infty(\mathbb{R}^{3N}; \mathbb{C}^{2^N}), \|\Psi\| = 1 \Big\} \end{aligned} \quad [2]$$

and likewise for the ground-state energy $E(\boldsymbol{z}, N, K)$ on the full space $\mathcal{H}_N$. Clearly, $E^{\mathrm{F}}(\boldsymbol{z}, N, K) \ge E(\boldsymbol{z}, N, K)$. It turns out that the energy $E(\boldsymbol{z}, N, K)$ is the same as one would get by restricting to symmetric functions instead of antisymmetric ones. Therefore, the energy $E(\boldsymbol{z}, N, K)$ is often referred to as the lowest possible energy for bosonic particles.

The Hamiltonian $H_N$ is an unbounded operator and we must discuss its domain to be able to talk about its spectrum. Also, it should be self-adjoint. It turns out that these questions are intimately related to stability. The operator $H_N$ is well defined on smooth (i.e., $C^\infty$) functions. Thus, the last definition of $E^{\mathrm{F}}(\boldsymbol{z}, N, K)$ in [2] is meaningful. If this ground-state energy is finite (i.e., not $-\infty$), then the Hamiltonian has an extension, the Friedrichs' extension, to a

self-adjoint operator with the property that the second equality in [2] holds.

In the definition of $E^{\mathrm{F}}$, we have minimized over all the positions $\boldsymbol{r}$ of the nuclei. Even though the nuclear dynamics is not considered, one is still interested in finding the lowest possible energy independent of where they are located.

**Theorem 1** (Stability of the first kind). *For all* $N$, $K$, *and* $z$, *we have*

$$E(z, N, K) > -\infty$$

**Theorem 2** (Stability of matter). *There exists a constant* $C_{|z|} > 0$ *depending only on* $|z| = \max\{z_1, \ldots, z_k\}$ *such that*

$$E^{\mathrm{F}}(z, N, K) \geq -C_{|z|}(N + K)$$

The constant $C_{|z|}$ bounds the binding energy per particle. In the case of hydrogen atoms, when $|z| = 1$, Dyson and Lenard arrived at a bound with $C_1 \approx 10^{14}$ Ry. Lieb and Thirring arrive at $C_1 \approx 5 = 10$ Ry. Since the binding energy of a single hydrogen atom is 1 Ry, it is easy to see that one must have $C_1 \geq 1/4$. Over the years, there have been some improvements on the estimated value of this constant in the theory of stability of matter.

That the Pauli exclusion principle, that is, the fermionic character of the electrons, is necessary for stability of matter is a consequence of the next theorem.

**Theorem 3** (the $N^{5/3}$ law for bosons). *If* $N = K$ *and* $z_1 = \cdots = z_K = z > 0$, *then there exist constants* $C_\pm > 0$ *depending on* $z$ *such that*

$$-C_- N^{5/3} < E(z, N, N) < -C_+ N^{5/3}$$

It is the superlinear (exponent 5/3) behavior in $N$ of the upper bound that violates stability of matter. This upper bound was proved by Lieb (1979) by a fairly simple variational argument. The lower bound above, which shows that the exponent 5/3 is optimal, was proved by Dyson and Lenard (1968) in their original paper on stability of matter.

This theorem leaves open the possibility that the stability of matter could be recovered by introducing finite nuclear masses. That this, indeed, is not the case was proved by Dyson (1967) by a complicated variational argument based on the Bogolubov pair theory for superfluid helium. We now add the kinetic energy $\sum_{k=1}^{K} -(1/2)\Delta_{r_k}$ of the nuclei (assuming, for simplicity, that they have the same mass as the electrons) to the Hamiltonian $H_N$ and consider the case where $z_1 = z_2 = \cdots = z_K = 1$. We denote the ground-state energy over the space $L^2(\mathbb{R}^{3(N+K)})$ (ignoring spin) by $\widetilde{E}(N,K)$. Then, Dyson proved that

$$\min_{N+K=M} \widetilde{E}(N, K) \leq -CM^{7/5}$$

for some constant $C > 0$. It was later shown by Conlon *et al.* (1988) that the exponent 7/5 is indeed optimal. Dyson (1967) made a conjecture for the precise asymptotic behavior of this energy. This conjecture, which was proved by Lieb and Solovej (2005) and Solovej (2004), is given in the next theorem.

**Theorem 4** (Dyson's 7/5-law for the charged Bose gas).

$$\lim_{M\to\infty} \min_{N+K=M} \frac{\widetilde{E}(N, K)}{M^{7/5}} = \inf\left\{\frac{1}{2}\int |\nabla\phi|^2 - J\int \phi^{5/2} \,|\, \phi \geq 0, \int \phi^2 = 1\right\} \quad [3]$$

*where*

$$J = \left(\frac{4}{\pi}\right)^{3/4} \frac{\Gamma(1/2)\Gamma(3/4)}{5\Gamma(5/4)}$$

## Generalizations of Stability of Matter

Over the years, generalizations of stability of matter including relativistic effects and interactions with the electromagnetic field have been attempted. Since the relativistic Dirac operator is not bounded below, we cannot simply replace the standard nonrelativistic kinetic energy operator $T_j = -(1/2)\Delta_j$ by the free Dirac operator.

Relativistic effects have been included by considering the (pseudo) relativistic kinetic energy

$$T_j^{\mathrm{Rel}} = \sqrt{-c^2\Delta_j + c^4} - c^2$$

In the units used in this article, the physical value of the speed of light $c$ is approximately 137 or, more precisely, the reciprocal of the fine-structure constant $\alpha$.

For this relativistic kinetic energy, Lieb and Yau (1988) proved that stability of matter holds in the sense formulated in Theorem 2 if $\alpha(=c^{-1})$ is small enough and $\max_j\{z_j\}\alpha \leq 2/\pi$. It is known here that the value $2/\pi$ is the best possible, since it is so for the one-atom case. The one-atom case had been studied by Herbst. The corresponding case of a one-electron molecule was studied by Lieb and Daubechies. Less optimal results on the stability of matter with relativistic kinetic energy had been

obtained prior to the work of Lieb and Yau by Conlon and later by Fefferman and de la Llave. References to these works can be found in the work of Lieb and Yau (1988).

The relativistic kinetic energy $T_j^{\text{Rel}}$ agrees with the free Dirac operator on the positive spectral subspace of the free Dirac operator (i.e., a subspace of $L^2(\mathbb{R}^3;\mathbb{C}^4)$). Therefore, the stability of matter follows if $T_j$ is replaced by the free Dirac operator and if one restricts to the Hilbert space obtained as in [2] but with $L^2(\mathbb{R}^3;\mathbb{C}^2)$ replaced by the positive spectral subspace of the free Dirac operator. This formulation is often referred to as the "no-pair" model. In the usual Dirac picture, the negative spectral subspace, the Dirac sea, is occupied. As long as one ignores pair creation, only the positive spectral subspace is available.

Magnetic fields may be included by considering the "magnetic kinetic energy"

$$T_j^{\text{Mag}} = \tfrac{1}{2}\left(-\mathrm{i}\nabla_j - c^{-1}\boldsymbol{A}(x_j)\right)^2$$

It turns out that the stability of matter theorem (Theorem 2) holds for all magnetic vector potentials $\boldsymbol{A}:\mathbb{R}^3 \to \mathbb{R}^3$ with a constant $C_{|z|}$ independent of $\boldsymbol{A}$. This is, therefore, also the case if we consider the magnetic field (or rather the vector potential) as a dynamic variable and add the (positive) field energy

$$\mathcal{U} = \frac{1}{8\pi}\int_{\mathbb{R}^3} |\nabla \times \boldsymbol{A}(x)|^2\,\mathrm{d}x \qquad [4]$$

to the Hamiltonian. The resulting Hamiltonian describes a charged spinless particle interacting with a classical electromagnetic field.

A more complicated situation is described by the "magnetic Pauli kinetic energy"

$$T_j^{\text{Pauli}} = \tfrac{1}{2}((-\mathrm{i}\nabla_j - c^{-1}\boldsymbol{A}(x_j))\cdot\boldsymbol{\sigma}_j)^2$$

where the coupling of the spin to the magnetic field is included through the vector of $2\times 2$ Pauli matrices acting on the spin components of particle $j$, that is, $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$, with

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

For the Pauli kinetic energy, stability of matter will not hold independently of the magnetic field (or even for a fixed unbounded magnetic field) unless the field energy $\mathcal{U}$ in [4] is included in the Hamiltonian. If the field energy is included, stability of matter holds independently of the magnetic field, that is, even if one minimizes over the dynamic variable $\boldsymbol{A}$, if $\alpha(=c^{-1})$ and $\max_j\{z_j\}\alpha^2$ are small enough. This was proved by Fefferman (1997) and by Lieb *et al.* (1995). The latter result includes the physical value of $\alpha$. The fact that a bound on $\alpha$ is needed had been proved by Loss and Yau. Stability for a one-electron atom had been proved in this model by Fröhlich, Lieb, and Loss. The many-electron atom and the one-electron molecule had been studied by Lieb and Loss. Most relevant references may be found in the work of Lieb *et al.* (1995).

The possibility of quantizing the magnetic field has also been studied. In this case, one must introduce an ultraviolet cutoff in the momentum modes of the vector potential. Stability of matter in the resulting model of (ultraviolet cutoff) QED coupled to non-relativistic matter was proved by Fefferman *et al.* improving results of Bugliaro, Fröhlich, and Graf.

Finally, one may include both relativistic effects and electromagnetic interactions. Let us first discuss the case of classical electromagnetic fields. If instead of the Pauli kinetic energy one uses the Dirac operator with a magnetic vector potential then there would be no lower bound on the energy. But, as previously described, one can study a no-pair formulation of relativistic particles coupled to electromagnetic fields. The question arises which subspace of $L^2(\mathbb{R}^3;\mathbb{C}^4)$ one should restrict to (i.e., which subspace is filled and which one is available). There are two obvious choices. Either one should, as before, restrict to the positive spectral subspace of the free Dirac operator or one should restrict to the positive spectral subspace of the magnetic Dirac operator. It is proved by Lieb *et al.* (1997) that the former choice leads to instability, whereas stability of matter holds for the latter choice under some conditions on $\alpha$ and $\max_j\{z_j\}$. Stability requires that the field energy $\mathcal{U}$ is included in the Hamiltonian. It then holds independently of the magnetic field.

This final stability result also holds if the magnetic field is quantized with an ultraviolet cutoff as proved by Lieb and Loss (2002).

The no-pair model even with the ultraviolet cutoff quantized field is not fully relativistically invariant. As mentioned above, there is still no mathematical formulation of QED, a fully relativistically invariant model for quantum particles interacting with electromagnetic fields.

## The Proof of Stability of the First Kind

The proof of stability of the first kind will now be sketched for charged quantum systems.

As mentioned in the introduction, stability of the first kind is a consequence of the uncertainty principle. Contrary to what is often stated in physics textbooks, stability does not follow from the Heisenberg formulation of the uncertainty principle.

A mathematically more flexible formulation is provided by the classical Sobolev inequality, which states that for all square-integrable functions $\psi \in L^2(\mathbb{R}^3)$, one has

$$\int |\nabla\psi|^2 \geq C_S \left(\int |\psi|^6\right)^{1/3} \qquad [5]$$

for $C_S > 0$. It follows from this inequality that for any attractive potential $V$, there is a lower bound on the energy expectation

$$\begin{aligned}&\left(\psi, \left(-\frac{1}{2}\Delta - V\right)\psi\right)\\ &\quad = \frac{1}{2}\int |\nabla\psi|^2 - \int V|\psi|^2 \geq \frac{1}{2} C_S \left(\int \psi^6\right)^{1/3}\\ &\quad - \left(\int V^{5/2} \int |\psi|^2\right)^{2/5} \left(\int |\psi|^6\right)^{1/5}\\ &\quad \geq -C \int V^{5/2} \int |\psi|^2\end{aligned}$$

for some $C > 0$. Thus, the lowest possible energy of one particle moving in the potential $V$ is bounded below by $-C\int V^{5/2}$. For $N$ (noninteracting) particles, the lower bound is $-CN\int V^{5/2}$. This holds whether or not the particles have spin. If, more generally, the potential can be written as $V = U + W, U, W \geq 0$, where $\int U^{5/2} < \infty$ and $W$ is bounded $W \leq \|W\|_\infty$, then the energy of $N$ noninteracting particles moving in the potential $V$ is bounded below by

$$-NC \int U^{5/2} - N\|W\|_\infty \qquad [6]$$

For the Hamiltonian $H_N$ from [1], one can get a lower bound on the energy $E(z, N, K)$ by ignoring all the positive potential terms, that is, the last two sums in [1]. The remaining Hamiltonian describes $N$ independent particles moving in the potential

$$-V = -\sum_{k=1}^{K} \frac{z_k}{|x - r_k|} = -\sum_{k=1}^{K} (U_k + W_k)$$

where $U_k$ is the restriction of $z_k/|x - r_k|$ to the set $|x - r_k| < R$ for some $R > 0$ and $W_k$ is the restriction to the complementary set. Using [6], one can easily see that the energy expectation is bounded below by

$$\begin{aligned}&- C\,N\,K^{5/2} \max_k\{z_k\}^{5/2} R^{1/2} - N\,K \max_k\{z_k\} R^{-1}\\ &\quad = -C'\,N\,K^2 \max_k\{z_k\}^2\end{aligned}$$

where we have made the optimal choice for $R \sim (K \max_k \{z_k\})^{-1}$.

This finite lower bound on the energy proves the stability of the first kind, but it clearly does not have the form required for the stability of the second kind.

## The Proof of Stability of Matter

The proof of stability of the first kind presented in the previous section must be improved in two ways in order to conclude the stability of matter.

For fermions, it turns out that the lower bound in [6] can be improved in such a way that there is no factor $N$ in the first term. This is the content of the bound of Lieb and Thirring discussed in the introduction.

**Theorem 5** (Lieb–Thirring inequality 1975). *The sum of all the negative eigenvalues of the operator* $-(1/2)\Delta - V(x)$ *is bounded below by*

$$-L_{\mathrm{LT}} \int V^{5/2}$$

*for some constant* $L_{\mathrm{LT}} > 0$

For $N$ noninteracting fermions moving in the potential $V$, the lowest possible energy is given by the sum of the $N$ lowest eigenvalues of the operator in the above theorem. Thus, the theorem gives a lower bound on this energy independently of $N$.

The second point where the argument from the previous section has to be improved is the control of the electrostatic energy. In the above discussion, all repulsive terms have simply been ignored. For stability of matter, a much more delicate bound is needed. Many versions of such bounds have been given going back to the work of Onsager (1939). Here, a result of Baxter (1980) will be used.

**Theorem 6** (Baxter's correlation estimate). *For all positions* $x_i, \ldots, x_N, r_1, \ldots, r_K \in \mathbb{R}^3$ *and all charges* $z_1, \ldots, z_K > 0$, *we have the pointwise inequality*

$$\begin{aligned}&-\sum_{k=1}^{K}\sum_{i=1}^{N} \frac{z_k}{|x_i - r_k|} + \sum_{1 \leq i < j \leq N} \frac{1}{|x_i - x_j|}\\ &\quad + \sum_{1 \leq k < \ell \leq K} \frac{z_k z_\ell}{|r_k - r_\ell|} \geq -\sum_{i=1}^{N} V(x_i)\end{aligned}$$

*where* $V(x) = (1 + 2\max_k\{z_k\}) \max_k \{|x - r_k|^{-1}\}$.

This theorem simply states that, for a lower bound, one can replace the full electrostatic Coulomb energy by the energy of independent electrons moving in the potential where they always see only the closest nuclei (with a modified charge). Baxter (1980) used probabilistic techniques to prove the inequality. An improved version of the inequality was given by Lieb and Yau (1988), with an analytic proof.

Similarly to the argument in the previous section, one can write $V(x) = U(x) + W(x)$, where $U$ is the restriction of $V$ to the set where $\min_k \{|x - r_k|\} < R$ for some $R > 0$ and $W$ is the restriction to the complementary set. It then follows from Baxter's correlation estimate and the Lieb–Thirring inequality that the lowest eigenvalue of the Hamiltonian $H_N$ on the fermionic Hilbert space $\mathcal{H}_N^{\mathrm{F}}$ is bounded below by

$$\begin{aligned} &-L_{\mathrm{LT}} \int U^{5/2} - N(1 + 2\max_k\{z_k\})R^{-1} \\ &\geq -C(1 + 2\max_k\{z_k\})^{5/2} K R^{1/2} \\ &\quad - N(1 + 2\max_k\{z_k\})R^{-1} \\ &= -C'(1 + 2\max_k\{z_k\})^2 (N + K) \end{aligned}$$

where $R \sim (1 + 2\max_k\{z_k\})^{-1}$. This lower bound is linear in the total particle number $N + K$, as required by stability of matter.

## From Thomas–Fermi Theory to Stability of Matter

In this final section, the proof of stability of matter by Lieb and Thirring (1975), where they use the Thomas–Fermi theory, is discussed briefly. First note that there is a dual formulation of the Lieb–Thirring inequality theorem (Theorem 5), which makes the connection to the Sobolev inequality [5] much more transparent.

**Theorem 7** (Lieb–Thirring inequality as a kinetic energy bound). *For any normalized antisymmetric (fermionic) wave function $\Psi \in \mathcal{H}_N^{\mathrm{F}}$ we have with $C_{\mathrm{LT}} = \frac{3}{5}(\frac{2}{5}L_{\mathrm{LT}}^{-1})^{2/3}$ the following lower bound on the kinetic energy*:

$$\begin{aligned} &\sum_{i=1}^{N} \frac{1}{2} \int_{\mathbb{R}^{3N}} \|\nabla_i \Psi(x_1, \ldots, x_N)\|^2 \, \mathrm{d}x_1 \cdots \mathrm{d}x_N \\ &\quad \geq C_{\mathrm{LT}} \int_{\mathbb{R}^3} \rho(x)^{5/3} \, \mathrm{d}x \end{aligned}$$

*where $\|\cdot\|$ is the norm in spin space $(\mathbb{C}^{2^N})$ and the one-electron density is given by*

$$\rho(x) = N \int_{\mathbb{R}^{3(N-1)}} \|\Psi(x, x_2, \ldots, x_N)\|^2 \, \mathrm{d}x_2 \cdots \mathrm{d}x_N$$

This estimate follows immediately from Theorem 5, which implies that

$$\sum_{i=1}^{N} \frac{1}{2} \int \|\nabla_i \Psi\|^2 - \int \rho V \geq -L_{\mathrm{LT}} \int V^{5/2}$$

To arrive at Theorem 7, simply choose $V = ((2/5(L_{\mathrm{LT}}^{-1}\rho)^{2/3}$.

One should compare the Lieb–Thirring kinetic energy bound with the expression $(3/10)(3\pi^2)^{2/3}\rho^{5/3}$ for the (thermodynamic) energy density of a free Fermi gas. One of the yet unproven conjectures is that the Lieb–Thirring bound holds with $C_{\mathrm{LT}}$ replaced by the free Fermi constant $(3/10)(3\pi^2)^{2/3}$.

The idea in the Lieb–Thirring proof of stability of matter is to bound the energy below by an expression depending only on the one-electron density. Theorem 7 achieves this for the kinetic energy. What is missing is a lower bound on the electrostatic Coulomb energy depending only on the density. One can show (see Lieb (1976) or Lieb and Thirring (1975)) that, except for an error of the form "$-\mathrm{const} \times N$," the total energy expectation $(\Psi, H_N \Psi)$ may be bounded below by

$$\begin{aligned} &C_{\mathrm{LT}} \int \rho^{5/3} - \sum_{k=1}^{K} \int \rho(x) \frac{z_k}{|x - r_k|} \, \mathrm{d}x \\ &\quad + \frac{1}{2} \iint \frac{\rho(x)\rho(y)}{|x - y|} \, \mathrm{d}x \, \mathrm{d}y + \sum_{1 \leq k < \ell \leq K} \frac{z_k z_\ell}{|r_k - r_\ell|} \end{aligned} \qquad [7]$$

Here, as before, $\rho$ is the one-electron density of the $N$-body wave function $\Psi$. The expression [7] is the famous Thomas–Fermi energy functional. It has been studied rigorously by Lieb and Simon (1977). The Thomas–Fermi energy is the infimum of the expression (7) over all $\rho$ with $\int \rho = N$. One of the important results about the Thomas–Fermi energy is Teller's no-binding theorem (Lieb and Simon 1977). It states that in Thomas–Fermi theory atoms do not bind to form molecules. This means that the Thomas–Fermi energy is greater than the sum of the individual atomic energies (these energies in turn depend only on the nuclear charges).

The above Thomas–Fermi lower bound on the energy expectation $(\Psi, H_N \Psi)$ together with the no-binding theorem implies stability of matter.

The generalizations to stability of matter discussed earlier are proved in a way similar to the proof presented in the previous section.

*See also*: *h*-Pseudodifferential Operators and Applications; Quantum Statistical Mechanics: Overview; Schrödinger Operators.

## Further Reading

Baxter JR (1980) Inequalities for potentials of particle systems. *Illinois Journal of Mathematics* 24: 645–652.

Conlon JG, Lieb EH, and Yau H-T (1988) The $N^{7/5}$ law for charged bosons. *Communications in Mathematical Physics* 116: 417–448.

Dyson FJ (1967) Ground state energy of a finite system of charged particles. *Journal of Mathematical Physics* 8: 1538–1545.

Dyson FJ and Lenard A (1967) Stability of matter. I. *Journal of Mathematical Physics* 8: 423–434.

Dyson FJ and Lenard A (1968) Stability of matter. II. *Journal of Mathematical Physics* 9: 698–711.

Fefferman CL (1997) Stability of matter with magnetic fields. *CRM Proceedings and Lecture Notes* 12: 119–133.

Fefferman C, Fröhlich J, and Graf GM (1997) Stability of ultraviolet cutoff quantum electrodynamics with non-relativistic matter. *Communications in Mathematical Physics* 190: 309–330.

Fisher M and Ruelle D (1966) The stability of many-particle systems. *Journal of Mathematical Physics* 7: 260–270.

Lieb EH and Lebowitz JL (1972) The constitution of matter: existence of thermodynamics for systems composed of electrons and nuclei. *Advances in Mathematics* 9: 316–398.

Lieb EH and Thirring WE (1975) Bound for the kinetic energy of fermions which proves the stability of matter. *Physical Review Letters* 35: 687–689.

Lieb EH (1976) The stability of matter. *Reviews of Modern Physics* 48: 553–569.

Lieb EH and Simon B (1977) Thomas–Fermi theory of atoms, molecules and solids. *Advances in Mathematics* 23: 22–116.

Lieb EH (1979) The $N^{5/3}$ law for bosons. *Physics Letters A* 70: 71–73.

Lieb EH and Thirring WE (1984) Gravitational collapse in quantum mechanics with relativistic kinetic energy. *Annals of Physics, NY* 155: 494–512.

Lieb EH and Yau H-T (1987) The Chandrasekhar theory of stellar collapse as the limit of quantum mechanics. *Communications in Mathematical Physics* 112: 147–174.

Lieb EH and Yau H-T (1988) The stability and instability of relativistic matter. *Communications in Mathematical Physics* 118: 177–213.

Lieb EH (1990) The stability of matter: from atoms to stars. 1989 Gibbs Lecture. *Bulletin of the American Mathematical Society* 22: 1–49.

Lieb EH, Loss M, and Solovej JP (1995) Stability of matter in magnetic fields. *Physical Review Letters* 75: 985–989.

Lieb EH, Siedentop H, and Solovej JP (1997) Stability and instability of relativistic electrons in classical electromagnetic fields. *Journal of Statistical Physics* 89: 37–59.

Lieb EH and Loss M (2002) Stability of a model of relativistic quantum electrodynamics. *Communications in Mathematical Physics* 228: 561–588.

Lieb EH (2004) The stability of matter and quantum electrodynamics, Proceedings of the Heisenberg Symposium, Munich, Dec. 2001. In: Buschhorn G and Wess J (eds.) *Fundamental Physics – Heisenberg and Beyond*, pp. 53–68. Berlin: Springer (arXiv math-ph/0209034).

Lieb EH and Solovej JP Ground state energy of the two-component charged Bose gas. *Communications in Mathematical Physics* (to appear).

Onsager L (1939) Electrostatic interaction of molecules. *Journal of Physical Chemistry* 43: 189–196.

Solovej JP (2004) Upper bounds to the ground state energies of the one- and two-component charged Bose gases. *Preprint*.


# Stability of Minkowski Space

**S Klainerman**, Princeton University, Princeton, NJ, USA



## Introduction

The Minkowski space, which is the simplest solution of the Einstein field equations in vacuum, that is, in the absence of matter, plays a fundamental role in modern physics as it provides the natural mathematical background of the special theory of relativity. It is most reasonable to ask whether it is stable under small perturbations. In other words, can arbitrary small perturbations of flat initial conditions lead to developments which are radically different, in the large, from the flat Minkowski space? It turns out to be a highly nontrivial problem as the Einstein equations are of a quasilinear hyperbolic character. Typical systems of this type, in three space dimensions, do form singularities in finite time even for small disturbances of their trivial initial data. To avoid finite-time singularities, we must require that sufficiently small perturbations of Minkowski space are geodesically complete. This, however, is not enough; one should also insist that the corresponding spacetimes become flat along all possible directions, that is, globally asymptotically flat. This is measured by the decay of the curvature tensor to zero. The precise rate of decay is also of interest. One expects that various null-frame components of the curvature tensor decay at different rates along outgoing null hypersurfaces; this goes under the name of "peeling estimates." It turns out in fact that we cannot prove geodesic completeness without establishing at the same time sufficiently fast rates of decay to flatness corresponding to at least some peeling.

The problem of stability of Minkowski space is intimately related to that of describing the asymptotic properties of the gravitational field at large distances from an isolated, weakly radiating physical system. Precise laws of gravitational radiation can be deduced from the assumption that the spacetime $(\boldsymbol{M}, \boldsymbol{g})$ under consideration can be conformally compactified by adding a boundary $\mathcal{S}$, called skry, to $\boldsymbol{M}$ so that an appropriate conformal rescaling of $\boldsymbol{g}$ can be extended smoothly to the new manifold $(\hat{\boldsymbol{M}}, \hat{\boldsymbol{g}})$ with boundary. In reality, the compactified spacetime cannot be smooth at the particular point $i^0$ corresponding to spacelike infinity. A spacetime

$(\boldsymbol{M}, \boldsymbol{g})$ is called asymptotically simple (AS) if its conformal completion is smooth everywhere except $i^0$ and every null geodesic intersects $\mathcal{S}$ at precisely two endpoints. The AS assumption allows one to derive precise decay asymptotic for various curvature components of $(\boldsymbol{M}, \boldsymbol{g})$ along null geodesics which are referred to as strong peeling. The obvious questions raised by this procedure are: do there exist nontrivial AS spacetimes and, if so, do they contain a sufficiently large class of radiating spacetimes including those which appear in all relevant applications?

Clearly, the two problems mentioned above are related but not equivalent. Asymptotically simple spacetimes verify strong peeling, in particular they are globally asymptotically flat, that is, their curvature tensor tends to zero along all geodesics. Yet, it is perfectly possible that arbitrarily small perturbations of the Minkowski space are geodesically complete and globally asymptotically flat without being asymptotically simple.

The first global stability result of the Minkowski metric was proved by Christodoulou and Klainerman (1993). Their result proves sufficiently strong peeling estimates to allow one to derive the most important properties of gravitational radiation, such as the Bondi mass-law formula, but not as strong as those consistent with asymptotic simplicity. A companion result was proved by Klainerman and Nicolò (2003). Recently, Rodnianski and Lindblad (submitted) have obtained a surprising global stability of Minkowski result for the Einstein vacuum equations in the Lorentz gauge, which provides considerable weaker peeling than Christodoulou and Klainerman (1993) and Klainerman and Nicolò (1999) but is much easier to prove.

The goal of this article is to describe various results obtained since the early 1980s concerning both aspects of the problem of stability of Minkowski mentioned above.

## Initial Data Formulation

The proper mathematical context for the stability of Minkowski is that provided by the initial-value problem for vacuum solutions to the Einstein field equations, that is, Ricci flat spacetimes $(\boldsymbol{M}, \boldsymbol{g})$, $\boldsymbol{R}_{\mu\nu} = 0$. We recall the following simple definitions:

**Definition 1** An initial data set is a triplet $(\Sigma, g, k)$ consisting of a three-dimensional complete Riemannian manifold $(\Sigma, g)$ and a 2-covariant symmetric tensor $k$ on $\Sigma$ satisfying the constraint equations:

$$\nabla^j k_{ij} - \nabla_i \, tr_g \, k = 0, \qquad R - |k|^2 + (\mathrm{tr}\, k)^2 = 0$$

where $\nabla$ is the covariant derivative, $R$ the scalar curvature of $(\Sigma, g)$. An initial data set is said to be maximal if $\mathrm{tr}_g \, k = 0$. This is a gauge condition which can be imposed without loss of generality. For simplicity we shall assume, throughout this article, that all initial data sets we consider are maximal.

**Definition 2** An initial data set is said to be flat, or trivial, if it corresponds to a complete spacelike hypersurface in Minkowski space with its induced metric and second fundamental form. An initial data set is said to be asymptotically flat if there exists a system of coordinates $(x^1, x^2, x^3)$ defined in a neighborhood of infinity on $\Sigma$, with $r = \sqrt{(x^1)^2 + (x^2)^2 + (x_3)^2}$, relative to which the metric $g$ approaches the Euclidian metric and $k$ approaches zero as $r \to \infty$. We assume, for simplicity, that $\Sigma$ has only one end. A neighborhood of infinity means the complement of a sufficiently large compact set on $\Sigma$.

**Remark 1** Because of the constraint equations, the asymptotic behavior cannot be arbitrarily prescribed. A precise definition of asymptotic flatness has to involve the ADM mass of $(\Sigma, g)$. Taking the mass into account, we write

$$g_{ij} = \left(1 + \frac{2M}{r}\right)\delta_{ij} + o(r^{-1})$$

According to the positive-mass theorem, $M \geq 0$ and $M = 0$ implies that the initial data set is flat.

**Definition 3** We say that an initial data set is strongly asymptotically flat if, for some $\delta 1/2$, relative to the coordinate system mentioned above,

$$g_{ij} - \left(1 + \frac{2M}{r}\right)\delta_{ij} = O(r^{-1-\delta}), \quad k_{ij} = O(r^{-2-\delta})$$

as $r \to \infty$

Moreover, every derivative of $g - (1 + 2M/r)\delta$ and $k$ improves the asymptotics by one.

**Definition 4** A Cauchy development of an initial data set $(\Sigma, g, k)$ is a spacetime manifold $(\boldsymbol{M}, \boldsymbol{g})$ satisfying the Einstein equations together with an embedding $i: \Sigma \longrightarrow \boldsymbol{M}$ such that $i_*(g), i_*(k)$ are the first and second fundamental forms of $i(\Sigma)$ in $\boldsymbol{M}$. A development is required to be also globally hyperbolic (which means that $i(\Sigma)$ is a Cauchy hypersurface, i.e., each causal curve in $\boldsymbol{M}$ intersects $i(\Sigma)$ at precisely one point) in order to assure the unique dependence of solutions on the data. A future development of $(\Sigma, g, k)$ consists of a globally hyperbolic manifold $(\boldsymbol{M}, \boldsymbol{g})$ with boundary, satisfying the Einstein equations, and an embedding $i$ as before which identifies $\Sigma$ to the boundary of $\boldsymbol{M}$.

The most primitive question asked about the initial-value problem, solved in a satisfactory way, for very large classes of evolution equations, is that of local existence and uniqueness of solutions. For the Einstein equations, this type of result was first established by Bruhat (1952) with the help of wave coordinates which allowed her to cast the Einstein equations in the form of a system of nonlinear wave equations to which one can apply the standard theory of symmetric hyperbolic systems. A stronger result, due to Hughes *et al.* (1976), states the following:

**Theorem 1** *Let* $(\Sigma, g, k)$ *be an initial data set for the Einstein vacuum equations. Assume that* $\Sigma$ *can be covered by a locally finite system of coordinate charts* $U_\alpha$ *related to each other by* $\mathbf{C}^1$ *diffeomorphisms, such that* $(g,k) \in H^s_{\mathrm{loc}}(U_\alpha) \times H^{s-1}_{\mathrm{loc}}(U_\alpha)$ *with* $s > 5/2$. *Then there exists a unique (up to an isometry) globally hyperbolic, Hausdorff, development (***M***, g) for which* $\Sigma$ *is a Cauchy hypersurface.*

In Theorem 1, the uniqueness up to an isometry requires additional regularity, $s > (5/2) + 1$, on the data. One has uniqueness, however, without additional regularity for the reduced Einstein equations system in wave coordinates.

**Remark 2** In the case of nonlinear systems of differential equations, the local existence and uniqueness result leads, through a straightforward extension argument, to a global result. The formulation of the same type of result for the Einstein equations is a little more subtle; it was done by Bruhat and Geroch.

**Theorem 2** (Bruhat–Geroch). *For each smooth initial data set, there exists a unique maximal future development.*

Thus, any construction, obtained by an evolutionary approach from a specific initial data set, must be necessarily contained in its maximal development. This may be said to solve the problem of global existence and uniqueness in general relativity. This is of course misleading, for equations defined in a fixed background global is a solution which exists for all time. In general relativity, however, we have no such background as the spacetime itself is the unknown. The connection with the classical meaning of a global solution requires a special discussion concerning the proper time of timelike geodesics; all further questions may be said to concern the qualitative properties of the maximal development. The central issue is that of existence and character of singularities. First, we can define a regular maximal development as one which is complete in the sense that all future timelike and null geodesics can be indefinitely extended relative to their proper time (or affine parameter in the case of null geodesics). If the initial data set is sufficiently far off from the trivial one, the corresponding future development may not be regular. This is the content of the following well-known theorem of Penrose (1979).

**Theorem 3** *If the manifold support of an initial data set is noncompact and contains a closed trapped surface, the corresponding maximal development is incomplete.*

## Stability of Minkowski Space

At the opposite end of Penrose's trapped-surface condition, the problem of stability of Minkowski space concerns the development of asymptotically flat initial data sets which are sufficiently close to the trivial one. Although it may be reasonable to expect the existence of a sufficiently small neighborhood of the trivial initial data set, in an appropriate topology, such that all corresponding developments are geodesically complete and globally asymptotically flat, such a result was by no means preordained. First, all known explicit asymptotically flat solutions of the Einstein vacuum equations, that is, the Kerr family, are singular. The attempts to construct nonexplicit, dynamic, solutions based on the conformal compactification method, due to Penrose (1962), were obstructed by the irregular behavior of initial data sets at $i^0$. (The problem is that the singularity at $i^0$ could propagate and thus destroy the expected smoothness of scry. This problem has been recently solved by constructing initial data sets which are precisely stationary at spacelike infinity.) Finally, the attempts, using partial differential equation hyperbolic methods, to extend the classical local result of Bruhat ran into the usual difficulties of establishing global in time existence to solutions of quasilinear hyperbolic systems. Indeed, as mentioned above, the wave coordinate gauge allows one to express the Einstein vacuum equations in the form of a system of nonlinear wave equations which does not satisfy Klainerman's null condition (the null condition (Klainerman 1983, 1986) identifies an important class of quasilinear systems of wave equations in four spacetime dimensions for which one can prove global in time existence of small solutions) and thus was sought to lead to formation of singularities. (The conjectured singular behavior of wave coordinates was sought, however, to reflect only the instability of the specific choice of gauge condition and not a true singularity of the equations.) According to Bruhat (personal communication),

Einstein himself had reasons to believe that the Minkowski space may not be stable. The problem of stability of the Minkowski space was first settled by Christodoulou and Klainerman (1990).

**Theorem 4** (Global stability of Minkowski). *Any asymptotically flat initial data set which is sufficiently close to the trivial one has a complete maximal future development.*

A related result (Theorem 5) proved recently by Klainerman and Nicolò (2003a), solves the problem of radiation for arbitrary asymptotically flat initial data sets: a proof the result below can also be derived, indirectly, from Christodoulou and Klainerman (1993). The proof of Klainerman and Nicolò (2003a) avoids, however, a great deal of the technical complications of this proof.

**Theorem 5** *For any, suitably defined, asymptotically flat initial data set* $(\Sigma, g, k)$ *with maximal future development* $(\mathbf{M}, \mathbf{g})$, *one can find a suitable domain* $\Omega_0 \subset \Sigma$ *with compact closure in* $\Sigma$ *such that the boundary* $\mathcal{D}_0^+$ *of its domain of influence* $\mathcal{C}^+(\Omega_0)$, *or causal future of* $\Omega$, *in* $\mathbf{M}$ *has complete null geodesic generators with respect to the corresponding affine parameters.*

Both the results of Christodoulou–Klainerman and Klainerman–Nicolò prove in fact a lot more than stated above. They provide a wealth of information concerning the behavior of null hypersurfaces as well as the rate at which various components of the Riemann curvature tensor approach zero along timelike and null geodesics. Here are more precise versions for Theorems 4 and 5.

**Theorem 4** (Expanded version). *Assume that* $(\Sigma, g, k)$ *is maximal and strong asymptotically flat,* $g - (1 + 2M/r)\delta = 0(r^{-3/2})$, $k = 0(r^{-5/2})$ *plus an appropriate global smallness assumption. We can construct complete spacetime* $(M, g)$ *together with a maximal foliation* $\Sigma_t$ *given by the level hypersurfaces of a time function t and null foliation* $C_u$, *given by the level hypersurfaces of an outgoing optical function u such that relative to an adapted null frame* $e_4 = L$, $e_3 = \underline{L}$, *and* $(e_a)_{a=1,2}$ *we have, along the null hypersurfaces* $C_u$ *the weak peeling decay,*

$$\begin{aligned} \alpha_{ab} &= R(L, e_a, L, e_b) = O(r^{-7/2}) \\ 2\beta_a &= R(L, \underline{L}, L, e_a) = O(r^{-7/2}) \\ 4\rho &= R(L, \underline{L}, L, \underline{L}) = O(r^{-3}) \\ 4\sigma &= {}^{*}R(L, \underline{L}, L, \underline{L}) = O(r^{-3}) \\ 2\underline{\beta}_a &= R(L, \underline{L}, L, e_a) = O(r^{-2}) \\ \underline{\alpha}_{ab} &= R(\underline{L}, e_a, \underline{L}, e_b) = O(r^{-1}) \end{aligned} \quad [1]$$

*as* $r \to \infty$ *with* $4\pi r^2 = \mathrm{Area}(S_{t,u} = \Sigma_t \cap C_u)$. *Also,* $\rho - \bar{\rho}, \sigma = O(r^{-7/2})$, *with* $\bar{\rho}$ *the average of* $\rho$ *over the compact 2-surfaces* $S_{t,u} = \Sigma_t \cap C_u$.

Three points are noteworthy. (1) The outgoing optical solution refers to the solution of the Eikonal equation $\mathbf{g}^{\alpha\beta}\partial_a u \partial_\beta u = 0$ whose level hypersurfaces $C_u$ intersect $\Sigma_t$ in expanding wave fronts for increasing $t$; (2) the generators $L$ and $\underline{L}$ are given by: $L = -\mathbf{g}^{\alpha\beta}\partial_\beta u \partial_\alpha$, the null geodesic generator of $C_u$; $\underline{L}$ is then the null conjugate of $L$, perpendicular to $S_{t,u} = C_u \cap \Sigma_t$; and (3) $e_a$ is an orthonormal frame on $S_{t,u}$.

**Theorem 5** (Expanded version). *For any asymptotically flat initial data sets* $(\Sigma, g, k)$, *verifying the same asymptotically flat conditions as in Theorem 4 one can find a suitable domain* $\Omega_0 \subset \Sigma$ *with compact closure in* $\Sigma$ *such that its future domain of influence* $\mathcal{C}^+(\Omega_0)$ *can be foliated by two null foliations; one outgoing* $C(u)$ *whose leaves are complete towards the future and the second one* $\underline{C}(\underline{u})$ *which is incoming. Let* $S(u, \underline{u}) = C(u) \cap \underline{C}(\underline{u})$ *denote the compact 2-surfaces of intersection between the outgoing and incoming null hypersurfaces, whose area is denoted by* $4\pi r^2$, *and consider an adapted null frame (that is, L is a the geodesic null generator of* $C(u)$, $\underline{L}$ *its null conjugate perpendicular to* $S(u, \underline{u})$, *and* $e_a$ *an orthonormal frame on* $S(u, \underline{u})$) $L, \underline{L}, (e_a)_{a=1,2}$ *at every point along an outgoing null cone* $C(u)$. *Then, denoting by* $\alpha, \beta, \rho, \sigma, \underline{\beta}, \underline{\alpha}$ *the null components of the curvature tensor, as in Theorem 5, we have, along* $C(u)$ *as* $r \to \infty$,

$$\begin{aligned} &\alpha, \beta, \rho - \bar{\rho}, \sigma = O(r^{-7/2}), \quad \underline{\beta} = O(r^{-2}), \\ &\underline{\alpha} = O(r^{-1}) \end{aligned} \quad [2]$$

Observe that the rates of decay in [1] and [2] are the same. This will be referred to as weak peeling to distinguish from the rates of decay compatible with asymptotic simplicity, that is,

$$\begin{aligned} &\alpha = O(r^{-5}), \quad \beta = O(r^{-4}) \\ &\rho, \sigma = O(r^{-3}), \quad \underline{\beta} = O(r^{-2}), \quad \underline{\alpha} = O(r^{-1}) \end{aligned} \quad [3]$$

to which we shall refer as strong peeling. We shall discuss more about these in the next section, following a review, of a recent result of Lindblad–Rodnianski.

Even the expanded forms of Theorems 4 and 5 stated here do not exhaust, all the information provided by global stability results in Christodoulou and Klainerman (1993) and Klainerman and Nicolò (2003a). Of particular interest are the main asymptotic conclusions which can be derived with the help of these information, the most

important being the Bondi mass-law formula which calculates the gravitational energy radiated at null infinity.

The simplest gauge condition in which the hyperbolic character of the Einstein field equations are easiest to exhibit is the wave coordinate condition; that is, one solves the Einstein vacuum equations relative to a special system of coordinates $x^\alpha$ which satisfy the equation $\Box \mathbf{g} x^\alpha = 0$. Then, denoting by $h_{\alpha\beta} = \mathbf{g}_{\alpha\beta} - \mathbf{m}_{\alpha\beta}$ with $\mathbf{m}$ the standard Minkowski metric, we obtain the following system of quasilinear wave equations in $h$,

$$g^{\mu\nu}\partial_\mu\partial_\nu h = N(h, \partial h) \qquad [4]$$

with $N(h, \partial h)$ a nonlinear term, quadratic in $\partial h$, which can be exhibited explicitly. This form of the Einstein field equations, called the wave coordinates reduced Einstein equations, is precisely the one which allowed Bruhat (1952) to prove the first local existence result. Later, she also pointed out that the first nontrivial iterate of [4] behaves like $t^{-1}\log t$ rather than $t^{-1}$ as expected from the decay properties of solutions to $\Box h = 0$ in Minkowski space. This seems to indicate that the wave coordinates may not be suitable to study the long-time behavior of solutions to the Einstein field equations. This negative conclusion is also consistent with the fact that the eqns [4] do not verify Klainerman's null condition. (Klainerman's null condition (Klainerman 1983) is an algebraic condition on systems of nonlinear wave equations in $(1+3)$ dimensions, similar to [4], which allows one to extend all local solutions, corresponding to small initial data, for all time. Moreover, these solutions decay at the rate of $t^{-1}$ as $t \to \infty$ consistent to the decay of free waves.) Lindblad and Rodnianski (2003) were able to isolate a new condition, which they call the weak null condition, verified by the wave coordinates reduced Einstein eqns [4], for which one can prove a small data global existence result consistent with the weaker decay rates suggested by the linear asymptotic analysis of Bruhat. Although the new result provides far weaker peeling information than [1], it is much simpler to prove than both Theorems 4 and 5. Moreover, the result seems to apply to a broader class of initial data than in Theorems 4 and 5. It remains an intriguing open problem whether the result of Lindblad–Rodnianski can be used as a stepping stone towards the more complete results of Theorems 4 and 5; that it is once a complete solution, with limited peeling, is known to exist whether one can improve, using the more precise techniques employed in Theorems 4 and 5 minus an important part of their technical complications, the weak peeling properties of [1].

## Strong Peeling

The weak peeling properties [1] derived in Theorems 4 and 5 are consistent, from a scaling point of view, with the SAF condition. To derive strong peeling, see [3], one needs stronger asymptotic conditions. Recently, Corvino–Schoen and Chruściel and Delay (2002) have proved the existence of a large class of asymptotically flat initial data sets $(\Sigma, g, k)$ which are precisely stationary (here $g_{\text{kerr}}, k_{\text{kerr}}$ are the initial data of the a Kerr solution in standard coordinates) $g = g_{\text{kerr}}, k = k_{\text{kerr}}$ outside a sufficiently large compact set. Moreover, they have proved the existence of sufficiently small solutions in this class which satisfy the requirements needed in Friedrich's conformal compactification method (see Friedrich (2002) and the references within) to produce asymptotically simple spacetimes, that is, spacetimes satisfying Penrose's regular compactification condition (Penrose 1962). Simultaneously, Klainerman and Nicolò (1999) were able to refine the methods used in the proof of Theorem 5 to prove the following:

**Theorem 6** *Assume that the initial data set* $(\Sigma, g, k)$ *of Theorem 5 satisfies the stronger assumption,*

$$g - g_S = O(r^{-(3/2+\gamma)}), \qquad k = (r^{-(5/2+\gamma)}) \qquad [5]$$

*for some* $\gamma > 3/2$. *Here*

$$g_S = \left(1 - 2\frac{M}{r}\right)^{-1} dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)$$

*denotes the restriction of the Schwarzschild to* $t = 0$ *in standard polar coordinates. Then, in addition to the results reported in Theorem 5, we have the strong peeling estimates,*

$$\alpha = O(r^{-5}), \qquad \beta = O(r^{-4})$$

*as* $r \to \infty$ *along the outgoing null leaves* $C(u)$. *Moreover, the same conclusions hold true if [5] is replaced by*

$$g - g_{\text{kerr}} = O(r^{-(3/2+\gamma)}), \qquad k - k_{\text{kerr}} = (r^{-(5/2+\gamma)}) \qquad [6]$$

*for some* $\gamma > 5/2$.

The first part of the theorem was proved in Klainerman and Nicolò (2003b). The second part is work in progress by Klainerman and Nicolò. The existence of initial conditions of the type required in Theorem 6 was established in the works of Corvino (2000) and Chruściel and Delay (2002).

## Open Problems

**Problem 1** *Extend results of Theorems 5 and 6 to the whole domain of dependence, for small data sets.*

The results of Theorems 5 and 6 give a satisfactory description of gravitational radiation of general classes of asymptotically flat initial data sets outside the domain of dependence of a sufficiently large compact set. It would be desirable to extend these results to the whole domain of dependence of initial data sets which satisfy an additional global smallness assumption similar to that of Theorem 4.

**Problem 2** *Is strong peeling (and implicitly asymptotic simplicity) consistent with physically relevant data? If not, is weak peeling a good substitute?*

Damour and Christodoulou (2000) have given conclusive evidence that under no-incoming-radiation condition the future null infinity cannot be smooth. In fact, $\beta = O(r^{-4} \log r)$ as $r \to \infty$.

**Problem 3** *Can one weaken the AF conditions to include, for example, initial data sets with infinite ADM angular momentum?*

It is reasonable to expect a global stability of Minkowski result for small initial data sets which verify, for arbitrarily small $\epsilon$,

$$g - \left(1 + 2\frac{M}{r}\right)\delta = 0(r^{-1-c}), \qquad k = 0(r^{-2-c})$$

One expects in this case that the top null components $\alpha$ and $\beta$ decay only like $O(r^{-3})$ as $r \to \infty$ along the null hypersurfaces $C(u)$. It seems that the methods of Lindblad–Rodnianski can treat this case but can only give decay estimates for $\alpha, \beta$ of the form $O(r^{-3+c})$.

**Problem 4** *Is the Kerr solution in the exterior of the black hole stable?*

The problem remains wide open.

*See also:* Asymptotic Structure and Conformal Infinity; Classical Groups and Homogeneous Spaces; Critical Phenomena in Gravitational Collapse; Einstein Equations: Exact Solutions; Geometric Analysis and General Relativity; Supergravity.

## Further Reading

Bruhat YC (1952) Theoreme d'existence pour certain systemes d'equations aux derivee partielles nonlineaires. *Acta Mathematica* 88: 141–225.

Bruhat YCh and Geroch RP Global aspects of the Cauchy problem in general relativity. *Communications in Mathematical Physics* 14: 329–335.

Christodoulou D (2000) *The Global Initial Value Problem in General Relativity*, Lecture Given at the Ninth Marcel Grossman Meeting. 2–8 July, Rome.

Christodoulou D and Klainerman S (1990) Asymptotic properties of linear field equations in Minkowski space. *Communications in Pure and Applied Mathematics* XLIII: 137–199.

Christodoulou D and Klainerman S (1993) The global non-linear stability of the Minkowski space. *Princeton Mathematical Series.*

Chruściel PT and Delay E (2002) Existence of non-trivial, vacuum, asymptotically simple spacetimes. *Classical and Quantum Gravity* 19: L71–L79.

Corvino J (2000) Scalar curvature deformation and a gluing construction for the Einstein constraint equations. *Communications in Mathematical Physics* 214: 137–189.

Friedrich H (2002) Conformal Einstein evolution. In: Frauendiener J and Friedrich H (eds.) *The Conformal Structure of Space-Time*, Lecture Notes in Physics. Springer.

Hughes T, Kato T, and Marsden J (1976) Well-posed quasilinear second-order hypebolic systems. *Archives for Rational and Mechanical Analysis* 63: 273–294.

Hawking SW and Ellis GFR (1973) The large scale structure of spacetime. *Cambridge Monographs on Mathematical Physics.*

Klainerman S (1983) Long-Time Behavior of Solutions to Nonlinear Wave Equations Proc. ICM Warszawa.

Klainerman S (1986) The null condition and global existence to non-linear wave equations. *Lectures in Applied Mathematics* 23: 293–326.

Klainerman S and Nicolò F (1999) On local and global aspects of the Cauchy problem in general relativity. *Classical and Quantum Gravity* 16: R73–R157.

Klainerman S and Nicolò F (2003a) The evolution problem in general relativity. *Progress in Mathematical Physics*, vol. 25. Boston: Birkhaüser.

Klainerman S and Nicolò F (2003b) Peeling properties of asymptotically flat solutions to the Einstein vacuum equations. *Classical and Quantum Gravity.*

Lindblad H and Rodnianski I (2003) The weak null condition for Eisnstein's equations. *Comptes Rendus Hebdomodaires des Seances de l'Academic des Sciences, Paris* 336(11): 901–906.

Lindblad H, and Rodnianski I Global existence for the Einstein vacuum equations in wave coordinates. *Annals of Mathematics* (submitted).

Penrose R (1965) Zero rest-mass fields including gravitation: asymptotic behaviour. *Proceedings of the Royal Society of London. Series* A 284: 159–203.

Penrose R (1979) Singularities and time asymmetry in general relativity – an Einstein centenary survey. In: Hawking S and Israel W (eds.) Cambridge: Cambridge University Press.

Penrose R (1962) Zero rest mass fields including gravitation: asymptotic behavior. *Proceedings of the Royal Society of London. Series A* 270: 159–203.

Wald R (1984) *General Relativity.* University of Chicago Press.

# Stability Problems in Celestial Mechanics

**A Celletti**, Università di Roma “Tor Vergata,” Rome, Italy



## Introduction

The long-term stability of planets and satellites might be desumed by the regular dynamics that we constantly observe. However, the ultimate fate of the solar system is an intriguing question, which has puzzled scientists since antiquity. In the past centuries, the common belief of a regular motion of the main planets was strengthened by the discovery of a simple law, due to J D Titius and J E Bode (eighteenth century), which provides a recipe to compute the approximate distances of the planets from the Sun. Adopting astronomical units as a measure of the distance, the Titius–Bode law can be stated as

$$d_n = 0.4 + 0.3 \times 2^n \text{ AU} \qquad [1]$$

where the index $n$ must be selected as provided in Table 1, which compares the distances computed according to [1] with the observed values. Titius and Bode already noticed that it was necessary to skip one unit in $n$ from Mars to Jupiter; indeed, the quantity $d_3 = 2.8$ AU might correspond to an average distance of some minor bodies of the asteroid belt, which had been discovered since the beginning of the nineteenth century. The studies of the $N$-body problem, namely the dynamics of $N$ mutually attracting bodies (according to Newton’s law), inspired several mathematical and physical theories: from the development of perturbation methods to the discovery of chaotic systems, as attested by the masterly work of H Poincaré (1892). In particular, perturbation theory had relevant applications in celestial mechanics; for example, it led to the prediction of the existence of Neptune in the nineteenth century by J C Adams and U Leverrier and later to the discovery of Pluto by C Tombaugh, as a result of unexplained perturbations on Uranus and Neptune, respectively. Modern advances in perturbation theories have been provided by the Kolgomorov–Arnol’d–Moser (KAM) and Nekhoroshev theorems, which find broad applications in celestial mechanics insofar as simple model problems are concerned.

The stability of the solar system can also be approached through numerical investigations, which allow one to predict the motion of the celestial bodies using more realistic models. The results of the numerical integrations undermine in some cases the apparent regularity of the solar system: in the following sections, we shall review many examples of regular and chaotic motions in different contexts of celestial mechanics, from the $N$-body problem to the rotational dynamics.

**Table 1** Tititus–Bode law and observed data

| | *Index n* (*of* [1]) | *Distance computed from* [1] | *Observed distance* (AU) |
|---|---|---|---|
| Mercury | $-\infty$ | 0.4 | 0.39 |
| Venus | 0 | 0.7 | 0.72 |
| Earth | 1 | 1 | 1 |
| Mars | 2 | 1.6 | 1.52 |
| Jupiter | 4 | 5.2 | 5.2 |
| Saturn | 5 | 10 | 9.54 |
| Uranus | 6 | 19.6 | 19.19 |

## The Restricted Three-Body Problem

Let $P_1, \ldots, P_N$ be $N$ bodies with masses $m_1, \ldots, m_N$, which interact through Newton’s law. Let $u^{(i)} \in \boldsymbol{R}^3, i = 1, 2, \ldots, N$, denote the position of the bodies in an inertial reference frame. Normalizing the gravitational constant to 1, the equations of motion of the $N$-body problem have the form

$$\frac{\mathrm{d}^2 u^{(i)}}{\mathrm{d}t^2} = -\sum_{j=1, j\neq i}^{N} \frac{m_j (u^{(i)} - u^{(j)})}{|u^{(i)} - u^{(j)}|^3}, \quad i = 1, \ldots, N \qquad [2]$$

In the case $N = 2$, one reduces to the two-body problem, which can be explicitly solved by means of Kepler’s laws as follows. Consider, for example, the Earth–Sun case: for negative values of the energy, the trajectory of the Earth is an ellipse with one focus coinciding with the barycenter, which can practically be identified with the Sun; the Earth–Sun radius vector describes equal areas in equal times; the cube of the semimajor axis is proportional to the square of the period of revolution.

Consider now an extension to the study of three bodies such that in the Keplerian approximation $P_2$ and $P_3$ move around $P_1$ and such that the semimajor axis of $P_2$ is greater than that of $P_3$ (an example is obtained identifying $P_1$ with the Sun, $P_2$ with the Jupiter, and $P_3$ with an asteroid of the main belt). The three-body problem is described by [2] setting $N = 3$; a special case is given by the restricted three-body problem, which describes the evolution of a “zero-mass” body under the gravitational attraction exerted by an assigned two-body system. Setting $N = 3$ and $m_3 = 0$ in [2], the

equations governing the restricted three-body problem are given by

$$\frac{d^2u^{(1)}}{dt^2} = -\frac{m_2(u^{(1)} - u^{(2)})}{|u^{(1)} - u^{(2)}|^3}$$
$$\frac{d^2u^{(2)}}{dt^2} = -\frac{m_1(u^{(2)} - u^{(1)})}{|u^{(2)} - u^{(1)}|^3}$$
$$\frac{d^2u^{(3)}}{dt^2} = -\frac{m_1(u^{(3)} - u^{(1)})}{|u^{(3)} - u^{(1)}|^3} - \frac{m_2(u^{(3)} - u^{(2)})}{|u^{(3)} - u^{(2)}|^3}$$

The first two equations concern the motion of the primaries $P_1$ and $P_2$ and they correspond to a Keplerian two-body problem, whose solution can be inserted in the equation for $u^{(3)}$, which becomes a periodically forced second-order equation. The restricted three-body problem can be conveniently described in terms of suitable action-angle coordinates, known as Delaunay variables. The present discussion is restricted to the planar case, namely we assume that the motion of the three bodies takes place on the same plane. The corresponding Delaunay variables, say $(L, G, \ell, \gamma) \in \boldsymbol{R}^2 \times \boldsymbol{T}^2$, are defined as follows (Szebehely 1967). Let $a$ and $e$ be, respectively, the semimajor axis and the eccentricity of the osculating orbit of $P_3$ and let $\mu = 1/m_1^{2/3}$; then Delaunay's action variables are given by

$$L = \mu\sqrt{m_1 a}, \qquad G = L\sqrt{1-e^2}$$

Next, introduce the angle variables: we denote by $\lambda$ and $\varphi$ the longitudes of Jupiter and of the asteroid; let $\gamma$ be the argument of perihelion, namely the angle formed by the periapsis direction with a preassigned reference line, and let $u$ denote the eccentric anomaly, which can be defined through

$$\tan\frac{\varphi-\gamma}{2} = \sqrt{\frac{1+e}{1-e}}\tan\frac{u}{2} \qquad [3]$$

Let $\ell$ be the mean anomaly, which is related to the eccentric anomaly by means of Kepler's equation

$$\ell = u - e\sin u \qquad [4]$$

Delaunay's angle variables are represented by the mean anomaly $\ell$ and by the argument of perihelion $\gamma$. For completeness, it should be remarked that the distance $r$ between the minor body $P_3$ and the primary $P_1$ is related to the longitude and to the eccentric anomaly by means of the relations

$$r = \frac{a(1-e^2)}{1+e\cos(\varphi-\gamma)} = a(1-e\cos u) \qquad [5]$$

In a reference frame centered at one of the primaries, say $P_1$, let $H = H(L, G, \ell, \gamma, \lambda)$ denote the Hamiltonian function describing the planar problem; notice that $H(L, G, \ell, \gamma, \lambda)$ has two degrees of freedom and an explicit time dependence through the longitude $\lambda$ of $P_2$. If the primaries are assumed to move in circular orbits around their common center of mass, the Hamiltonian function reduces to two degrees of freedom, where a new variable $g$ is introduced as the difference between the argument of perihelion $\gamma$ and the longitude $\lambda$ of the primary. Normalizing the units of measure so that the distance between the primaries and the sum of their masses is unity, the Hamiltonian function $H$ describing the circular, planar, restricted three-body problem is given by

$$H(L, G, \ell, g) = -\frac{1}{2L^2} - G + \varepsilon F(L, G, \ell, g) \qquad [6]$$

where $\varepsilon = \mu m_2$. The perturbing function takes the form

$$F = r\cos(f+g) - \frac{1}{\sqrt{1+r^2-2r\cos(f+g)}}$$

where $f = \varphi - \gamma$ represents the true anomaly, namely the angle formed by the instantaneous orbital radius with the periapsis line. Notice that the quantities $r$ and $f$ are functions of the Delaunay variables through the relations [3]–[5]. As a consequence, one can expand the perturbing function in the form (Delaunay 1860)

$$F(L, G, \ell, g) = \sum_{j,k\geq 0} F_{jk}(\ell, g)e^j a^k$$

where $F_{jk}$ are cosine terms with arguments given by a linear combination of the variables $\ell$ and $g$. For example, the first few terms of the series development are given by the following expression:

$$\begin{aligned} F(L, G, \ell, g) = & -1 - \frac{L^4}{4} - \frac{9}{64}L^8 + \frac{L^4 e}{2}\cos\ell \\ & - \left(\frac{3}{8}L^6 + \frac{15}{64}L^{10}\right)\cos(\ell+g) \\ & + \frac{9}{4}L^4 e\cos(\ell+2g) \\ & - \left(\frac{3}{4}L^4 + \frac{5}{16}L^8\right)\cos(2\ell+2g) \\ & - \frac{3}{4}L^4 e\cos(3\ell+2g) \\ & - \left(\frac{5}{8}L^6 + \frac{35}{128}L^{10}\right)\cos(3\ell+3g) \\ & - \frac{35}{64}L^8\cos(4\ell+4g) \\ & - \frac{63}{128}L^{10}\cos(5\ell+5g) + \cdots \end{aligned} \qquad [7]$$

where the eccentricity is a function of the actions through $e = \sqrt{1 - G^2/L^2}$. We remark that the Hamiltonian [6] is nearly integrable with perturbing parameter $\varepsilon$; indeed, for $\varepsilon = 0$ one recovers the two-body problem describing the interaction between $P_1$ and $P_3$, which can be explicitly solved according to Kepler's laws.

### KAM Stability

Classical perturbation theory, as developed by Laplace, Lagrange, Delaunay, Poincaré, etc., does not allow investigation of the stability of the $N$-body problem, since the series defining the solution are generally divergent. In order to justify this statement, let us start by rewriting the unperturbed Hamiltonian in [6] as

$$h(L, G) = -\frac{1}{2L^2} - G \qquad [8]$$

so that [6] becomes $H(L, G, \ell, g) = h(L, G) + \varepsilon F(L, G, \ell, g)$. In order to remove the perturbation to the second order in the perturbing parameter, one looks for a change of variables $(L, G, \ell, g) \to (L', G', \ell', g')$ close to the identity, that is,

$$\begin{aligned} L &= L' + \varepsilon \frac{\partial \Phi}{\partial \ell}(L', G', \ell, g) \\ G &= G' + \varepsilon \frac{\partial \Phi}{\partial g}(L', G', \ell, g) \\ \ell' &= \ell + \varepsilon \frac{\partial \Phi}{\partial L'}(L', G', \ell, g) \\ g' &= g + \varepsilon \frac{\partial \Phi}{\partial G'}(L', G', \ell, g) \end{aligned}$$

where $\Phi(L', G', \ell, g)$ is the generating function of the transformation. Let

$$\frac{\partial h}{\partial L}(L, G) = \frac{1}{L^3} \equiv \omega(L)$$

In order to perform a first-order perturbation theory, we look for a generating function $\Phi(L', G', \ell, g)$, such that the transformed Hamiltonian is integrable up to $O(\varepsilon^2)$, namely

$$\begin{aligned} &h\left(L' + \varepsilon \frac{\partial \Phi}{\partial \ell}(L', G', \ell, g), G' + \varepsilon \frac{\partial \Phi}{\partial g}(L', G', \ell, g)\right) \\ &\quad + \varepsilon F\left(L' + \varepsilon \frac{\partial \Phi}{\partial \ell}(L', G', \ell, g), G' \right. \\ &\quad \left. + \varepsilon \frac{\partial \Phi}{\partial g}(L', G', \ell, g), \ell, g\right) \\ &= h_1(L', G') + \varepsilon\left[\omega(L') \frac{\partial \Phi}{\partial \ell}(L', G', \ell', g') \right. \\ &\quad \left. - \frac{\partial \Phi}{\partial g}(L', G', \ell', g') + F(L', G', \ell', g')\right] + O(\varepsilon^2) \end{aligned}$$

where $h_1(L', G')$ is the new unperturbed Hamiltonian. If we denote by $F_0(L', G')$ the average of the perturbing function over the angle variables, the new unperturbed Hamiltonian takes the form

$$h_1(L', G') = h(L', G') + \varepsilon F_0(L', G')$$

Expanding $F$ in Fourier series as $F(L, G, \ell, g) = \sum_{n, m \in \mathbf{Z}} F_{nm}(L, G) e^{i(n\ell + mg)}$, the generating function is given by the following expression:

$$\Phi(L', G', \ell, g) = -i \sum_{n, m \in \mathbf{Z} \setminus \{0\}} \frac{F_{nm}(L', G')}{\omega(L')n - m} e^{i(n\ell + mg)}$$

The occurrence of small divisors of the form

$$\frac{1}{\omega(L)n - m}, \quad n, m \in \mathbf{Z}$$

might prevent the convergence of the series defining the generating function. In particular, we remark that zero divisors occur whenever $\omega(L) = m/n$. This situation, which is called an $m:n$ orbit–orbit resonance, implies that during a given interval of time the body $P_3$ makes $m$ revolutions, whereas $P_2$ makes exactly $n$ orbits about $P_1$.

The control of the occurrence of the small divisors was obtained through a theorem by A N Kolmogorov, who made a major breakthrough in the study of nearly integrable systems. He proved, under general assumptions, that some regions of the phase space are almost filled by maximal invariant tori. The theorem provides a constructive algorithm to give estimates on the perturbing parameter, ensuring the existence of some invariant surfaces. Kolmogorov's theorem was later extended by V I Arnol'd and J Moser, giving rise to the so-called KAM theory. More precisely, the KAM theorem can be stated as follows (see, e.g., Arnol'd *et al.* (1997)): consider a real-analytic, nearly integrable Hamiltonian function and fix a rationally independent frequency vector $\omega$; if the unperturbed Hamiltonian is not degenerate and if the frequency satisfies a strong nonresonance assumption (called the diophantine condition), for sufficiently small values of the perturbing parameter, there exists an invariant torus on which a quasiperiodic motion with frequency $\omega$ takes place. A preliminary investigation of the stability of the $N$-body problem by means of KAM theory (Arnol'd *et al.* 1997) leads to the existence of large regions filled by quasiperiodic motions, provided the masses of the planets are sufficiently small. Arnol'd's version of KAM theorem has been applied by J Laskar and P Robutel to the spatial three-body planetary problem (the planetary problem concerns the study of the

dynamics of two bodies with comparable masses, moving in the gravitational field of a larger primary) and the existence of quasiperiodic motions has been proved for values of the ratio of semimajor axis less than 0.8 and for inclinations up to $\sim 1°$.

Concrete estimates on the strength of the perturbation were given by M Hénon: in the context of the three-body problem, the application of the original version of Arnol'd's theorem allows one to prove the existence of invariant tori for values of the perturbing parameter (representing the Jupiter–Sun mass ratio) $\leq 10^{-333}$ while the implementation of Moser's theorem provides an estimate of $10^{-50}$. We remark that the astronomical value of the Jupiter–Sun mass ratio amounts to $\sim 10^{-3}$, showing a relevant discrepancy between KAM results and physical measurements. More recently, KAM estimates have been refined and adapted to the study of significant problems of celestial mechanics (Celletti and Chierchia 1995). Strong improvements have been obtained combining accurate estimates with a computer-assisted implementation, where the computer is used to perform long computations concerning the development of the perturbing series and the check of KAM estimates. The numerical errors are controlled through the implementation of a suitable technique, known as interval arithmetic. In the framework of the planar, circular, restricted three-body problem, the stability of some asteroids has been proved by A Celletti and L Chierchia for realistic values of the perturbing parameter (e.g., for $\varepsilon = 10^{-3}$). A suitable approximation of the disturbing function (namely, a finite truncation of the series development as in [7]) has been considered. The result relies on an implementation of a computer-assisted isoenergetic KAM theorem and on the following remark: in the four-dimensional phase space, on a fixed energy level the invariant two-dimensional surfaces separate the phase space, providing the stability of the actions for all motions trapped between any two invariant tori. Since the action variables are related to the semimajor axis and to the eccentricity of the orbit, one obtains that the elliptic elements remain close to their initial values.

A computer-assisted KAM theorem has been applied by A Giorgilli and U Locatelli to the planetary (Jupiter–Saturn) problem. Using a suitable secular approximation, it can be shown that this model admits two invariant tori, which bound the orbits corresponding to the initial data of Jupiter and Saturn.

#### Nekhoroshev Stability

A different approach in order to study the stability of nearly integrable systems is provided by Nekhoroshev's theorem (see, e.g., Arnol'd *et al.* (1997)), which guarantees, under smallness requirements, the stability of the motions on an open set of initial conditions for exponentially long times. Consider a Hamiltonian function of the form

$$H(y, x) = h(y) + \varepsilon f(y, x), \quad (y, x) \in B \times \boldsymbol{T}^n \qquad [9]$$

where $B$ is an open subset of $\boldsymbol{R}^n$. We assume that $h$ and $f$ are analytic functions and that the integrable Hamiltonian $h$ satisfies a geometric condition, called steepness. We remark that functions such as $h(L, G)$ in [8] satisfy the steepness condition. For sufficiently small values of $\varepsilon$, Nekhoroshev's theorem states that any motion $(y(t), x(t))$ satisfying Hamilton's equations associated with [9] is bounded for a finite (but exponentially long) time, that is,

$$\|y(t) - y(0)\| \leq y_0 \varepsilon^a, \quad \text{for } |t| \leq t_0 \mathrm{e}^{(\varepsilon_0/\varepsilon)^b}$$

where $y_0, t_0, \varepsilon_0, a$, and $b$ are suitable positive constants.

Nekhoroshev's theorem can be conveniently applied to the three-body problem, where it provides a confinement of the action variables, representing the semimajor axis and the eccentricity of the osculating orbit. Interesting applications of Nekhoroshev's theorem concern the investigation of the triangular Lagrangian points in the spatial, restricted three-body problem. (The Lagrangian points are five equilibrium positions of the planar, restricted three-body problem in a synodic reference frame, which rotates with the angular velocity of the primaries. Two of such positions are called triangular, since the configuration of the three bodies is an equilateral triangle in the orbital plane.) Effective estimates were developed by A Giorgilli and C Skokos, showing the existence of a stability region around the Lagrangian point $L_4$, large enough to include some known asteroids. In the same framework, the exponential stability was proven by G Benettin, F Fassó, and M Guzzo for all values of the mass-ratio parameter, except for a few values of the reduced mass $\mu$ up to $\mu \simeq 0.038$.

#### Numerical Results

The study of the stability of the $N$-body problem can be investigated by performing numerical integrations of the equations of motion. The dynamics of the outer planets of the solar system (from Jupiter to Pluto) has been explored by Sussman and Wisdom (1992) using a dedicated computer, the Digital Orrery. The integration of the equations of motion was performed over 845 million years; the results provided evidence of the stability of the major planets and a chaotic behavior of Pluto. An

alternative approach, based on an average of the equations of motion over fast angles, was adopted by Laskar (1995), where the perturbing function of the spatial problem was expanded up to the second order in the masses and up to the fifth powers of the eccentricity and the inclination. The dynamics of all planets (excluding Pluto) was investigated by means of frequency analysis over a time span ranging from −15 Gyr to +10 Gyr. The numerical integrations provided evidence of the regularity of the external planets (from Jupiter to Neptune), a moderate chaotic behavior of Venus and the Earth, and a marked chaotic dynamics of Mercury and Mars. The computations show that the inner solar system is chaotic, with a Lyapunov time of $\sim 5$ Myr, thus preventing any prediction of the evolution over 100 Myr.

## The Spin–Orbit Problem

The dynamics of the bodies of the solar system results from a combination of a revolutionary motion around a primary body and a rotation about an internal axis. A simple mathematical model describing the spin–orbit interaction can be introduced as follows. Let S be a triaxial ellipsoidal satellite, which moves about a central planet P. We denote by $T_{rev}$ and $T_{rot}$ the periods of revolution and rotation. A $p:q$ spin–orbit resonance occurs if

$$\frac{T_{rev}}{T_{rot}} = \frac{p}{q}, \quad \text{for } p, q \in \mathbf{N}, \ q \neq 0$$

Whenever $p = q = 1$, the satellite always points the same face to the host planet. Most of the evolved satellites or planets are trapped in a 1 : 1 resonance, with the only exception of Mercury, which is observed in a nearly 3 : 2 resonance. In order to introduce a simple mathematical model which describes the spin–orbit interaction, we assume that:

1. the satellite moves on a Keplerian orbit around the planet (with semimajor axis $a$ and eccentricity $e$);
2. the spin axis is perpendicular to the orbit plane;
3. the spin axis coincides with the shortest physical axis; and
4. dissipative effects as well as perturbations due to other planets or satellites are neglected.

We denote by $A < B < C$ the principal moments of inertia of the satellite and by $r$ and $f$, respectively, the instantaneous orbital radius and the true anomaly of the Keplerian orbit. Let $x$ be the angle between the longest axis of the ellipsoid and a preassigned reference line. From standard Euler's equations for rigid body, the equation of motion in normalized units (i.e., assuming that the period of revolution is $2\pi$) takes the form

$$\ddot{x} + \frac{\varepsilon}{r^3} \sin(2x - 2f) = 0 \qquad [10]$$

where $\varepsilon \equiv \frac{3}{2}(B - A)/C$. This equation is integrable whenever $A = B$ or in the case of zero orbital eccentricity. Due to the assumption of Keplerian motion, both $r$ and $f$ are known functions of the time. Therefore, we can expand [10] in Fourier series as

$$\ddot{x} + \varepsilon \sum_{m \neq 0,\, m=-\infty}^{\infty} W\left(\frac{m}{2}, e\right) \sin(2x - mt) = 0 \qquad [11]$$

where the coefficients $W(m/2, e)$ decay as $W(m/2, e) \propto e^{|m-2|}$. A further simplification of the model is obtained as follows. According to (4), we neglected the dissipative forces and perturbations due to other bodies. The most important contribution is due to the nonrigidity of the satellite, provoking a tidal torque caused by the internal friction. The size of the dissipative effects is significantly small compared to the gravitational terms. Therefore, we decide to retain in [11] only those terms which are of the same order or larger than the average effect of the tidal torque. The following equation results:

$$\ddot{x} + \varepsilon \sum_{m \neq 0,\, m=N_1}^{N_2} \tilde{W}\left(\frac{m}{2}, e\right) \sin(2x - mt) = 0 \qquad [12]$$

where $N_1$ and $N_2$ are suitable integers, which depend on the physical and orbital parameters of the satellite, while $\tilde{W}(m/2, e)$ are suitable truncations of the coefficients $W(m/2, e)$. We remark that eqn [12] can be derived from Hamilton's equations associated with a one-dimensional, time-dependent, nearly integrable Hamiltonian function with perturbing parameter $\varepsilon$ and a trigonometric disturbing function.

### Analytical Results

The phase space associated with [12] admits a Poincaré map showing a pendulum-like structure: the periodic orbits are surrounded by librational curves and the chaotic separatrix divides the librational regime from the region where rotational motions can take place. The three-dimensional phase space is separated by KAM rotational tori into invariant regions, providing a strong stability property for all motions confined between any pair of KAM rotational tori. Let us denote by $\mathcal{P}(p/q)$ a periodic orbit associated with the $p:q$ resonance; in the context of the model associated with [12], the

stability of the periodic orbit $\mathcal{P}(p/q)$ is obtained by showing the existence of two invariant tori $\mathcal{T}(\omega_1)$ and $\mathcal{T}(\omega_2)$ with $\omega_1 < p/q < \omega_2$. A refined computer-assisted KAM theorem has been implemented (Celletti 1990) with the aim of proving the existence of trapping invariant surfaces. Realistic estimates, in agreement with the physical values of the parameters (namely, the equatorial oblateness $\varepsilon$ and the eccentricity $e$), have been obtained in several examples of spin–orbit commensurabilities, like the 1 : 1 Moon–Earth interaction or the 3 : 2 Mercury–Sun resonance.

Concerning Nekhoroshev-type estimates, the classical D'Alembert problem has been studied by Biasco and Chierchia (2002). In particular, an equatorially symmetric oblate planet moving on a Keplerian orbit around a primary body has been investigated; the model does not assume any further constraint on the spin axis. Although the Hamiltonian describing this model is properly degenerate, it is shown that Nekhoroshev-like results apply to the D'Alembert problem in the proximity of a 1 : 1 resonance.

### Numerical Results

The model introduced in [10]–[12] often represents an unrealistic simplification of the spin–orbit dynamics. In particular, assumption (1) implies that secular perturbations of the orbital parameters are neglected, whereas the hypothesis (2) corresponds to disregarding the spin–orbit obliquity, namely the angle formed by the rotational axis with the normal to the orbital plane. Due to the presence of an equatorial bulge, the gravitational attraction of the other bodies of the solar system induces a torque, resulting in a precessional motion. It is also important to take into account the changes of the obliquity angle, whose variations might affect the climatic behavior.

A realistic model for the precession and the variation of the obliquity has been presented by Laskar (1995). The numerical simulations and the frequency-map analysis show that the Earth's obliquity is actually stable, although a large chaotic region is found in the interval between 60° and 90°. Since the present obliquity of the Earth amounts to ~23.3°, the Earth is outside the dangerous region. An interesting simulation was performed to evaluate the role played by the Moon. Without the Moon, the extent of the chaotic region would greatly increase, eventually preventing the birth of an evoluted life. Among the inner planets, Mars' obliquity shows larger chaotic extent, which drives to variations from 0° to 60° in a few million years. On the contrary, the external planets do not show significant chaotic regions and their obliquities are essentially stable.

*See also:* Averaging Methods; Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Gravitational *N*-Body Problem (Classical); Hamiltonian Systems: Stability and Instability Theory; KAM Theory and Celestial Mechanics; Multiscale Approaches; Stability Theory and KAM.

## Further Reading

Arnol'd VI, Kozlov VV, and Neishtadt AI (1997) *Mathematical Aspects of Classical and Celestial Mechanics*. Berlin: Springer.

Biasco L and Chierchia L (2002) Nekhoroshev stability for the D'Alembert problem of celestial mechanics. *Atti Accademia Nazionale Lincei Classe Scienze Fisiche Matematiche Naturali Rendiconti Lincei (9) Matematica Applicata* 13(2): 85–89.

Celletti A (1990) Analysis of resonances in the spin–orbit problem in celestial mechanics: the synchronous resonance (Part I). *Journal of Applied Mathematics and Physics (ZAMP)* 41: 174–204.

Celletti A and Chierchia L (1995) *A Constructive Theory of Lagrangian Tori and Computer-Assisted Applications*. Dynamics Reported, New Series, vol. 4, pp. 60–129. Berlin: Springer.

Delaunay C (1860) Théorie du Mouvement de la Lune, Mémoires de l'Académie des Sciences 1, Tome XXVIII, Paris.

Laskar J (1995) Large scale chaos and marginal stability in the solar system. In: *XIth ICMP Meeting (Paris, July 1994)*, pp. 75–120. Cambridge, MA: International Press.

Poincarè H (1892) *Les Méthodes Nouvelles de la Méchanique Céleste*. Paris: Gauthier-Villars.

Sussman GJ and Wisdom J (1992) Chaotic evolution of the solar system. *Science* 241: 56–62.

Szebehely V (1967) *Theory of Orbits*. New York: Academic Press.

# Stability Theory and KAM

**G Gentile**, Università degli Studi "Roma Tre," Rome, Italy



## Introduction

A Hamiltonian system is a dynamical system whose equations of motions can be written in terms of a scalar function, called the Hamiltonian of the system: if one uses coordinates $(\boldsymbol{p}, \boldsymbol{q})$ in a domain (phase space) $\mathcal{D} \subset \boldsymbol{R}^{2N}$, where $N$ is the number of independent variables one needs to identify a configuration of the system (degrees of freedom), there is a function $\mathcal{H}(\boldsymbol{p}, \boldsymbol{q})$ such that $\dot{\boldsymbol{p}} = -\partial\mathcal{H}/\partial\boldsymbol{q}$ and $\dot{\boldsymbol{q}} = \partial\mathcal{H}/\partial\boldsymbol{p}$. An integrable (Hamiltonian) system is a Hamiltonian system which, in suitable coordinates $(\boldsymbol{A}, \boldsymbol{\alpha}) \in \mathcal{A} \times \boldsymbol{T}^N$, where $\mathcal{A}$ is an open subset of $\boldsymbol{R}^N$ and $\boldsymbol{T} = \boldsymbol{R}/2\pi\boldsymbol{Z}$ is the standard torus, can be described by a Hamiltonian $\mathcal{H}_0(\boldsymbol{A})$, that is, depending only on $\boldsymbol{A}$. The coordinates $(\boldsymbol{A}, \boldsymbol{\alpha})$ are called action-angle variables. In such a case the dynamics is trivial: any initial condition $(\boldsymbol{A}_0, \boldsymbol{\alpha}_0)$ evolves in such a way that the action variables are constants of motion (i.e., $\boldsymbol{A}(t) = \boldsymbol{A}_0$ for all $t \in \boldsymbol{R}$), while the angles grow linearly in time as $\boldsymbol{\alpha}(t) = \boldsymbol{\alpha}_0 + \boldsymbol{\omega}t$, where $\boldsymbol{\omega} = \boldsymbol{\omega}(\boldsymbol{A}_0) \equiv \partial_{\boldsymbol{A}}\mathcal{H}_0(\boldsymbol{A}_0)$ is called the rotation (or frequency) vector. An integrable system can be thought of as a collection of decoupled (i.e., independent) rotators: the entire phase space $\mathcal{A} \times \boldsymbol{T}^N$ is foliated into invariant tori and all motions are quasiperiodic. Integrable systems are stable, in the sense that nearby initial conditions separate at most linearly in time (in particular, the actions do not separate at all): mathematically, this is expressed by the fact that all the Lyapunov exponents are nonpositive.

An example of an integrable system is any one-dimensional conservative mechanical system, in any region of phase space in which motions are bounded. By increasing the number of degrees of freedom, exhibiting nontrivial integrable systems can become a difficult task. The problem of studying the effects of even small Hamiltonian perturbations on integrable systems and of understanding if the latter remain stable, in the aforementioned sense, was considered by Poincaré to be the fundamental problem of dynamics. For a long time, it was commonly thought that all motions could be reduced to superpositions of periodic motions, hence to quasiperiodic motions, but at the end of nineteenth century it was realized by Boltzmann and Poincaré that such a picture was too naive, and that in reality more complicated motions were possible. As a consequence of this, it became a widespread belief that, even when starting from an integrable system, the introduction of an arbitrarily small perturbation would break integrability.

This belief was strengthened by the work of Poincaré (1898), who showed that the series describing the solution in a perturbation theory approach are in general divergent. The source of divergence in perturbation series is the presence of small divisors, that is, of denominators of the kind of $\boldsymbol{\omega} \cdot \boldsymbol{\nu}$, where $\boldsymbol{\omega}$ is the rotation vector that should characterize the invariant torus (if existent) and $\boldsymbol{\nu}$ is any integer vector. Despite this, however, perturbation series (known as Lindstedt series) continued to be extensively used by astronomers in problems of celestial mechanics, such as the study of planetary motions, for the simple reason that they provided predictions in good agreement with the observations. But the feeling that the underlying mathematical tools were unsatisfactory persisted.

In fact, the well-known Fermi–Pasta–Ulam numerical experiment, in 1955, was originally conceived in the spirit of confirming that integrability would in general be easily lost. Consider a chain with $N$ harmonic oscillators, with, say, periodic boundary conditions, coupled with cubic and quartic two-body potentials, so that the Hamiltonian is

$$\begin{gathered} \mathcal{H}(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{N} \frac{1}{2} p_i^2 + W(q_{i+1} - q_i) \\ W(x) = \frac{1}{2}x^2 + \frac{\alpha}{3}x^3 + \frac{\beta}{4}x^4 \end{gathered} \qquad [1]$$

for $\alpha, \beta$ real parameters and $(\boldsymbol{p}, \boldsymbol{q}) \in \boldsymbol{R}^N \times \boldsymbol{R}^N$. One can introduce new variables such that the Hamiltonian, for $\alpha = \beta = 0$, can be written as

$$\mathcal{H}_0(\boldsymbol{A}) = \sum_{i=1}^{N} \frac{1}{2}\left(P_k^2 + \omega_k Q_k^2\right) = \boldsymbol{\omega} \cdot \boldsymbol{A} \qquad [2]$$

for a suitable rotation vector $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_N) \in \boldsymbol{R}^N$ (an explicit computation gives $\omega_k = 2\sin(k\pi/N)$).

Consider an initial condition in which all the energy is confined to a few modes, that is, $A_k \neq 0$ at $t = 0$ only for a few values of $k$. For $\alpha = \beta = 0$, the system is integrable, so that $A_k(t) = 0$ for all $t \in \boldsymbol{R}$ and for all $k$ such that $A_k(0) = 0$. If the system ceases to be integrable when the perturbation is switched on, the energy is likely to start to be shared among the various modes, and after a long enough time has

elapsed, an equidistribution of the energy among all modes (thermalization) might be expected. At least this behavior was expected by Fermi, Pasta, and Ulam, but it was not what they found numerically: on the contrary, all the energy seemed to remain associated with the modes close to the few initially excited ones.

At about the same time, Kolmogorov (1954) published a breakthrough paper going exactly in the opposite direction: if one perturbs an integrable system, under some mild conditions on the integrable part, most of the tori are preserved, although slightly deformed. A more precise statement is the following.

**Theorem 1** *Let an N-degree-of-freedom Hamiltonian system be described by an analytic Hamiltonian of the form*

$$\mathcal{H}(\boldsymbol{A},\boldsymbol{\alpha}) = \mathcal{H}_0(\boldsymbol{A}) + \varepsilon f(\boldsymbol{A},\boldsymbol{\alpha}) \qquad [3]$$

*with $\varepsilon$ a real parameter (perturbation parameter), $f$ a $2\pi$-periodic function of each angle variable (potential or perturbation), and $\mathcal{H}_0(\boldsymbol{A})$ satisfying the nondegeneracy condition $\det \partial_{\boldsymbol{A}}^2 \mathcal{H}_0(\boldsymbol{A}) \neq 0$ (anisochrony condition). If $\boldsymbol{\omega} = \boldsymbol{\omega}(\boldsymbol{A}) \equiv \partial_{\boldsymbol{A}} \mathcal{H}_0(\boldsymbol{A})$ is fixed to satisfy the Diophantine condition*

$$|\boldsymbol{\omega}\cdot\boldsymbol{\nu}| > \frac{C_0}{|\boldsymbol{\nu}|^{\tau}} \quad \forall \boldsymbol{\nu} \in \boldsymbol{Z}^N \setminus \boldsymbol{0} \qquad [4]$$

*for some constants $C_0 > 0$ and $\tau > N-1$ (here $|\boldsymbol{\nu}| = |\nu_1| + \cdots + |\nu_N|$ and $\cdot$ denotes the standard inner product: $\boldsymbol{\omega}\cdot\boldsymbol{\nu} = \omega_1\nu_1 + \cdots + \omega_N\nu_N$), then there is an invariant torus with rotation vector $\boldsymbol{\omega}$ for $\varepsilon$ small enough, say for $\varepsilon$ smaller than some value $\varepsilon_0$ depending on $C_0$ and $\tau$ (and on the function $f$).*

By saying that there is an invariant torus with rotation vector $\boldsymbol{\omega}$, one means that there is an invariant surface in phase space on which, in suitable coordinates, the dynamics is the same as in the unperturbed case, and the conjugation (i.e., the change of variables which leads to such coordinates) is analytic in the angle variables and in the perturbation parameter. One also says that the torus of an integrable system ($\varepsilon = 0$) is preserved (or even persists) under a small perturbation.

Note that, *a posteriori*, this proves convergence of the perturbation series: however, a direct check of convergence was performed only recently by Eliasson (1996). Kolmogorov's proof was based on a completely different idea, that is, by performing iteratively a sequence of canonical transformations (which are changes of coordinates preserving the Hamiltonian structure of the equations of motion) such that at each step the size of the perturbation is reduced. Of course, on the basis of Poincaré's result, this iterative procedure cannot work for all initial conditions (e.g., when $\boldsymbol{\omega}$ does not satisfy [4]). The key point in Kolmogorov's scheme is to fix the rotation vector $\boldsymbol{\omega}$ of the torus one is looking for, in such a way that the small divisors are controlled through the Diophantine condition [4] and the exponentially fast convergence of the algorithm.

New proofs and extensions of Kolmogorov's theorem were given later by Arnol'd (1962) and by Moser (1962); hence, the acronym KAM to denote such a theorem. Arnol'd gave a more detailed (and slightly different) proof compared to the original one by Kolmogorov, and applied the result to the planar three-body problem, thus showing that physical applications of the theorem were possible. Moser, on the other hand, proposed a modified method using a technique introduced by Nash (which approximates smooth functions with analytical ones) to deal with the case of systems with finite smoothness.

For fixed small enough $\varepsilon$, the surviving invariant tori cover a large portion of the phase space, called the Kolmogorov set; the relative measure of the region of phase space which is not filled by such tori tends to zero at least as $\sqrt{\varepsilon}$ for $\varepsilon \to 0$. A system described by a Hamiltonian like [3] is then called a quasi-integrable Hamiltonian system.

The excluded region of phase space corresponds to the unperturbed tori which are destroyed by the perturbation: the rotation vectors of such tori are close to a resonance, that is, to a value $\boldsymbol{\omega}$ such that $\boldsymbol{\omega}\cdot\boldsymbol{\nu} = 0$ for some integer vector $\boldsymbol{\nu}$, and these are exactly the vectors which do not satisfy the Diophantine condition [4] for any value $C_0$. A subset of phase space of this kind is called a resonance region.

At first sight, this would seem to provide an explanation for the results found by Fermi, Pasta, and Ulam, but this is not quite the case. First, the threshold value $\varepsilon_0$ depends on $N$, and goes to zero very fast as $N \to \infty$ (in general as $N!^{-\alpha}$ for some $\alpha > 0$); however, the results of the numerical experiments apparently were insensitive to the number $N$ of oscillators. Second, the KAM theorem deals with maximal tori, that is, tori characterized by rotation vectors which have as many components as the number of degrees of freedom, while the rotation vectors of the numerical quasiperiodic solutions seem to involve just a small number of components.

Finally, as an extra problem, the validity of the nondegeneracy condition for the unperturbed

Hamiltonian is violated, because the unperturbed Hamiltonian is linear in the action variables (one says that the Hamiltonian is isochronous). Recently, Rink (2001), by continuing the work by Nishida, showed that in the Fermi–Pasta–Ulam problem it is possible to perform a canonical change of coordinates such that in the new variables the Hamiltonian becomes anisochronous: one uses part of the perturbation to remove isochrony. But the other two obstacles remain.

## Lower-Dimensional Tori

A natural question is what happens to the invariant tori corresponding to rotation vectors which are not rationally independent, that is, vectors satisfying $n$ resonance conditions, such as $\boldsymbol{\omega}\cdot\boldsymbol{\nu}_i=0$ for $n$ independent vectors $\boldsymbol{\nu}_1,\ldots,\boldsymbol{\nu}_n$, with $1\le n\le N-2$ (the case $n=N-1$ corresponds to periodic orbits and is comparatively easy); for instance, one can take $\boldsymbol{\omega}=(\omega_1,\ldots,\omega_n,0,\ldots,0)$ and, by a suitable linear change of coordinates, one can always make the reduction to a case of this kind. In particular, one can ask if a result analogous to the KAM theorem holds for these tori. Such a problem for the model [3] has not been studied very widely in the literature. What has usually been considered is a system of $n$ rotators coupled with a system with $s=N-n$ degrees of freedom near an equilibrium point: then one calls normal coordinates the coordinates describing the latter, and the role of the parameter $\varepsilon$ is played by the size of the normal coordinates (if their initial conditions are chosen near the equilibrium point). In the absence of perturbation (i.e., for $\varepsilon=0$), one has either hyperbolic or elliptic or, more generally, mixed tori, according to the nature of the equilibrium points: one refers to these tori as lower-dimensional tori, as they represent $n$-dimensional invariant surfaces in a system with $N$ degrees of freedom. Then one can study the preservation of such tori.

One can prove that, in such a case, at least if certain generic conditions are satisfied, in suitable coordinates, $n$ angles rotate with frequencies $\omega_1,\ldots,\omega_n$, respectively, while the remaining $N-n$ angles have to be fixed close to some values corresponding to the extremal points of the function obtained by averaging the potential over the rotating angles.

The case of hyperbolic tori is easier, as in the case of elliptic tori one has to exclude some values of $\varepsilon$ to avoid some further resonance conditions between the rotation vector $\boldsymbol{\omega}$ and the normal frequencies $\lambda_k$ (i.e., the eigenvalues of the linearized system corresponding to the normal coordinates), known as the first and second Mel'nikov conditions:

$$\begin{aligned}&|\boldsymbol{\omega}\cdot\boldsymbol{\nu}\pm\lambda_k|>\frac{C_0}{|\boldsymbol{\nu}|^\tau}\quad\forall\boldsymbol{\nu}\in\mathbf{Z}^N\setminus 0,\ \forall 1\le k\le s\\&|\boldsymbol{\omega}\cdot\boldsymbol{\nu}\pm\lambda_k\pm\lambda_{k'}|>\frac{C_0}{|\boldsymbol{\nu}|^\tau}\quad\forall\boldsymbol{\nu}\in\mathbf{Z}^N\setminus 0\\&\quad\forall 1\le k,k'\le s\end{aligned}\qquad[5]$$

Such conditions appear, with the values of the normal frequencies slightly modified by terms depending on $\varepsilon$, at each iterative step, and at the end only for values of $\varepsilon$ belonging to some Cantor set one can have elliptic lower-dimensional tori.

The second Mel'nikov conditions are not really necessary, and in fact they can be relaxed as Bourgain (1994) has shown; this is an important fact, as it allows degenerate normal frequencies, which were forbidden in the previous works by Kuksin (1987), Eliasson (1988), and Pöschel (1989).

Similar results also apply in the case of lower-dimensional tori for the model [3], which represents sort of a degenerate situation, as the normal frequencies vanish for $\varepsilon=0$. Again, one has to use part of the perturbation to remove the complete degeneracy of normal frequencies.

## Quasiperiodic Solutions in Partial Differential Equations

For explaining the Fermi–Pasta–Ulam experiment, one has to deal with systems with arbitrarily many degrees of freedom. Hence, it is natural to investigate systems which have *ab initio* infinitely many degrees of freedom, such as the nonlinear wave equation, $u_{tt}-u_{xx}+V(x)u=\varphi(u)$, the nonlinear Schrödinger equation, $\mathrm{i}u_t-u_{xx}+V(x)u=\varphi(u)$, the nonlinear Korteweg–de Vries equation $u_t+u_{xxx}-6u_xu=\varphi(u)$, and other systems of nonlinear partial differential equations (PDEs); the continuum limit of the Fermi–Pasta–Ulam model gives indeed a nonlinear Korteweg–de Vries equation, as shown by Zabuski and Kruskal (1965). Here $(t,x)\in\mathbf{R}\times[0,\pi]^d$, if $d$ is the space dimension, and either periodic $(u(0,t)=u(\pi,t))$ or Dirichlet $(u(0,t)=u(\pi,t)=0)$ boundary conditions can be considered; $\varphi(u)$ is a function analytic in $u$ and starting from orders strictly higher than one, while $V(x)$ is an analytic function of $x$, depending on extra parameters $\xi_1,\ldots,\xi_n$. Such a function is introduced essentially for technical reasons, as we shall see that the eigenvalues $\lambda_k$ of the Sturm–Liouville operator $-\partial_x^2+V(x)$ must satisfy some Diophantine conditions. If we set $V(x)=\mu\in\mathbf{R}$ in the nonlinear wave equation, we obtain the Klein–Gordon equation, which, in the particular case $\mu=0$,

reduces to the string equation. Again, the role of the perturbation parameter is played by the size of the solution itself.

Small-amplitude periodic and quasiperiodic solutions for PDE systems have been extensively studied, among others, by Kuksin, Wayne, Craig, Pöschel, and Bourgain. Results for such systems read as follows. Consider for concreteness the one-dimensional nonlinear wave equation with Dirichlet boundary conditions and with $\varphi(u) = u^3 + O(u^5)$. When the nonlinear function $\varphi(u)$ is absent, any solution of the linear wave equation $u_{tt} - u_{xx} + V(x)u = 0$ is a superposition of either finitely or infinitely many periodic solutions with frequencies $\lambda_k$ determined by the function $V(x)$. Let $u_0(\boldsymbol{\omega}t, x)$ be a quasiperiodic solution of the linear wave equation with rotation vector $\boldsymbol{\omega} \in \boldsymbol{R}^n$, where $\omega_k = \lambda_{m_k}$, for some $n$-tuple $\{m_1, \ldots, m_n\}$. Then for $\varepsilon$ small enough there exists a subset $\Xi_\varepsilon$ of the space of parameters with large Lebesgue measure (more precisely, with complementary Lebesgue measure which tends to zero when $\varepsilon \to 0$) such that for all $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n) \in \Xi_\varepsilon$ there is a solution $u_\varepsilon(t, x)$ of the nonlinear wave equation and a rotation vector $\boldsymbol{\omega}_\varepsilon$ satisfying the conditions

$$\begin{aligned} \left| u_\varepsilon(t, x) - \sqrt{\varepsilon} u_0(\boldsymbol{\omega}_\varepsilon t, x) \right| &\leq C\varepsilon \\ |\boldsymbol{\omega}_\varepsilon - \boldsymbol{\omega}| &< C\varepsilon \end{aligned} \qquad [6]$$

for some positive constant $C$.

The case $n = 1$ (periodic solutions) is not as easy as the finite-dimensional case, because there are infinitely many normal frequencies, so that there are small divisor problems which for finite-dimensional systems appear only for $n \geq 2$.

For the nonlinear wave equation and the Schrödinger equation, if $n \geq 1$, one can take $V(x) = \mu$, but one needs $\mu \neq 0$; for $n > 1$, one can take $V(x) = \mu$, as one can perform a preliminary transformation leading to an equation in which a function depending on parameters naturally appears, as shown by Kuksin and Pöschel (1996). For $n = 1$, the case $\mu = 0$ has been very recently solved by Gentile *et al.* (2005).

Statements for more general situations can also be obtained, while extensions to space dimensions $d \geq 2$ are not trivial and have been obtained only recently by Bourgain (1998). The above result also holds if the number of components of the rotation vector is less than the number of parameters: one uses such parameters because one needs to impose some Diophantine conditions such as [5], now for all the frequencies $\lambda_k = \omega_k, k \notin \{m_1, \ldots, m_n\}$. Again, the second Mel'nikov conditions were shown by Bourgain to be unnecessary, and this is an essential ingredient for the higher-dimensional case.

Even if systems of the type considered above have been widely studied, they remain significantly different from a discrete system such as the chain of oscillators [1] for $N$ large enough (also in the limit $N \to \infty$), so that the results which have been found for PDE systems do not really provide an explanation for the numerical findings.

Also in the case of lower-dimensional tori for finite-dimensional systems the main problem is that, even if such tori exist, it is not clear what relevance they can have for the dynamics (a case in which hyperbolic tori play a role is considered later). An important feature of maximal tori is that they fill most of the phase space, a property which certainly does not hold for lower-dimensional tori, which lie outside the Kolmogorov set.

In the Fermi–Pasta–Ulam experiment, one considers initial conditions close to lower-dimensional tori; hence, an interesting problem is to study their stability, that is, how fast the trajectories starting from such initial conditions drift away.

## Arnol'd Diffusion and Nekhoroshev's Theorem

Consider again the maximal tori. For $N = 2$, the preservation of most of the invariant tori prevents the possibility of diffusion in phase space: the tori represent two-dimensional surfaces in a three-dimensional space (as dynamics occur on the level surfaces of the energy in a four-dimensional space), so that, if an initial condition is trapped in a gap between two tori, the corresponding trajectory remains confined forever between them. The situation is quite different for $N \geq 3$: in such a case, the tori do not represent a topological obstruction to diffusion any more.

That mechanisms of diffusion are really possible was shown by Arnol'd (1963). Because of the perturbation, lower-dimensional hyperbolic tori appear inside the resonance regions, with their stable and unstable manifolds (whiskers). It is possible that these manifolds of the same torus intersect with a nonvanishing angle (homoclinic angle); as a consequence, the angles between the stable and unstable manifolds of nearby tori (heteroclinic angles) can also be different from zero, and one can find a set of hyperbolic lower-dimensional tori such that the unstable manifold of each of them intersects the stable manifold of the torus next to it: one says that such tori form a transition chain of heteroclinic connections. Then there can be trajectories moving along such connections, producing at the end a drift of order 1 (in $\varepsilon$) in the action variables. Such a phenomenon is referred to as Arnol'd diffusion.

Of course, diffusing trajectories should be located in the region of phase space where there are no invariant tori (hence, a very small region when $\varepsilon$ is small), but an important consequence is that, unlike what happens in the unperturbed case, not all motions are stable: in particular, the action variables can change by a large amount over long times.

Providing interesting examples of Hamiltonian systems in which Arnol'd diffusion can occur is not so easy: in fact, for the diffusion to really occur, one needs a lower bound on the homoclinic angles, and to evaluate these angles can be difficult. For instance, Arnold's (1963) original example, which describes a system near a resonance region, is a two-parameter system given by

$$\frac{1}{2}\left(A_1^2 + A_2^2\right) + A_3 + \mu(\cos\alpha_1 - 1) + \varepsilon\mu(\cos\alpha_1 - 1)(\sin\alpha_2 + \cos\alpha_3) \qquad [7]$$

and the angles can be proved to be bounded from below only by assuming that the perturbation parameter $\varepsilon$ is exponentially small with respect to the other parameter $\mu$, which in turn implies a situation not really convincing from a physical point of view. More generally, for all the examples which are discussed in literature, the relation with physics (as the d'Alembert problem on the possibility for a planet to change the inclination of the precession cone) is not obvious.

So the question naturally arises as to how fast can such a mechanism of diffusion be, and how relevant is it for practical purposes. A first answer is provided by a theorem of Nekhoroshev (1977), which states the following result.

**Theorem 2** *Suppose we have an N-degree-of-freedom quasi-integrable Hamiltonian system, where the unperturbed Hamiltonian satisfies some condition such as convexity (or a weaker one, known as steepness, which is rather involved, to state in a concise way); for concreteness consider a function $\mathcal{H}_0(\boldsymbol{A})$ in [2] which is quadratic in $\boldsymbol{A}$. Then there are two positive constants a and b such that for times t up to $O(\exp(\varepsilon^{-b}))$ the variations of the action variables cannot be larger than $O(\varepsilon^a)$.*

The constants $a$ and $b$ depend on $N$, and they tend to zero when $N \to \infty$; Lochak and Neĭshtadt (1992) and Pöschel (1993) found estimates $a = b = 1/2N$, which are probably in general optimal. Nekhoroshev's theorem is usually stated in the form above, but it provides more information than that explicitly written: the trajectories, when trapped into a resonance region, drift away and come close to some invariant torus, and then they behave like quasiperiodic motions, up to very small corrections, for a long time, until they enter some other resonance region, and so on. Of course, for initial conditions on some invariant torus, KAM theorem applies, but the new result concerns initial conditions which do not belong to any tori.

Nekhoroshev's theorem gives a lower bound for the diffusion time, that is, the time required for a drift of order 1 to occur in the action variables. But, of course, an upper bound would also be desirable. The diffusion times are related to the amplitude of the homoclinic angles, which are very small (and difficult to estimate as stated before). The strongest results in this direction have been obtained with variational methods, for instance, by Bessi, Bernard, Berti, and Bolle: at best, for the diffusion time, one finds an estimate $O(\mu^{-1}\log\mu^{-1})$, if $\mu$ is the amplitude of the homoclinic angles (which in turn are exponentially small in some power of $\epsilon$, as one can expect as a consequence of Nekhoroshev's theorem).

Then one can imagine that the results of the Fermi–Pasta–Ulam experiment can also be interpreted in the light of Nekhoroshev's theorem. The solutions one finds numerically certainly do not correspond to maximal tori, but one could expect that they could be solutions which appear to be quasiperiodic for long but finite times (e.g., moving near some lower-dimensional torus determined by the initial conditions), and that if one really insists on observing the time evolution for a very long time, then deviations from quasiperiodic behavior could be detected. This is an appealing interpretation, and the most recent numerical results make it plausible: Galgani and Giorgilli (2003) have found numerically that the energy, even if initially confined to the lower modes, tend to be shared among all the other modes, and higher the modes the longer is the time needed for the energy to flow to them. Of course, this does not settle the problem, as there is still the issue of the large number of degrees of freedom; furthermore, for large $N$ the spacing between the frequencies is small, and they become almost degenerate. Hence, the problem still has to be considered as open.

## Stability versus Chaos

The main problem in applying the KAM theorem seems to be related to the small value of the threshold $\epsilon_0$ which is required. In general, when the size of the perturbation parameter is very large, the region of phase space filled with invariant tori decreases (or even disappears), and chaotic motions appear. By the latter, one generally means motions which are highly sensitive to the initial conditions: a small variation of the initial conditions produces a catastrophic variation in the corresponding trajectories (this is due to the appearance of strictly positive Lyapunov exponents).

A natural question is then how such a result as the KAM theorem is meaningful in physical situations: in other words, for which systems the KAM theorem can really apply.

One of the main motivations to study such a problem was to explain astronomical observations and to study the stability of the solar system. In order to apply the KAM theorem to the solar system, one has to interpret the gravitational forces between the planets as perturbations of a collection of several decoupled two-body systems (each planet with the Sun). One can write the masses of the planets as $\varepsilon m_i$, and $\varepsilon$ plays the role of the perturbation parameter. The corresponding Hamiltonian (after suitable reductions and scalings) is

$$\sum_{i=1}^{N} \frac{p_i^2}{2\mu_i} - \sum_{i=1}^{N} \frac{m_i m_0}{|q_i|} + \varepsilon \sum_{1 \le i < j \le N} \frac{p_i \cdot p_j}{m_0} + \varepsilon \sum_{1 \le i < j \le N} \frac{m_i m_j}{|q_i - q_j|} \qquad [8]$$

where $i = 0$ corresponds to the Sun, while $i = 1, \ldots, N$ correspond to the planets (hence $N = 9$), $m_0$ is the mass of the Sun, and $\varepsilon\mu_i$ are the reduced masses ($\mu_i^{-1} = m_i^{-1} + \varepsilon m_0^{-1}$); here $(q_i, p_i) \in \boldsymbol{R}^3 \times \boldsymbol{R}^3, i = 0, \ldots, N$, the inner product in $p_i \cdot p_j$ is in $\boldsymbol{R}^3$, and the norm $|\cdot|$ is the Euclidean one.

A first difficulty is that the solar system is a properly degenerate system; that is, the unperturbed Hamiltonian does not depend on all the action variables. But such a degeneracy can be removed by performing a canonical change of coordinates which produces a new Hamiltonian in which the integrable part contains new terms of order $\varepsilon$ depending on all action variables and is nondegenerate, while the perturbation becomes of order $\varepsilon^2$: the angle variables corresponding to the actions not originally appearing in the unperturbed Hamiltonian are called the slow variables, while the others are called the fast variables.

However, a naive implementation of the KAM theorem, in general, even for simplified but still realistic systems, would provide a preposterously small value of the threshold $\varepsilon_0$. The problem could be just a computational one: in principle, a very refined estimate of the threshold could give a better value, so that it is very difficult to decide analytically if the real values of the planetary masses allow the solar system to fall inside the regime of applicability of the KAM theorem. Results in this direction have been obtained, but only for special situations: for instance, by considering the restricted planar circular three-body problem (which provides a simplified description of the system "Sun + Jupiter + asteroid"), Celletti and Chierchia (1997) found analytical bounds on the perturbation parameters comparable with the physical values. Of course, this is not at all conclusive for the general situation in which all planets (with their satellites and the asteroids) are considered together; in particular, it does not shed light on the problem of the stability of the entire solar system.

On the contrary, extensive numerical simulations performed by Laskar (starting from 1989) seem to suggest that the solar system is unstable. Deflections from the current orbits could be produced to such an extent that collisions between planets could not be avoided: Mercury could collide with Venus and be ejected from the solar system. An important issue is to consider the times over which such phenomena can occur. Laskar's numerical simulations show that such times are less than the estimated age of the solar system, and that one can make accurate predictions for the planetary motions only for a finite amount of time (~100 Myr). Furthermore, the assumed partial instability of the solar system has also been used by Laskar (2004) to explain some observed phenomena such as the evolution of the obliquity (which is the angle between equator and orbital plane) of some planets. Of course, these simulations have been carried out with several approximations, as that of averaging over the fast variables, which allows one to use a large integration step in the numerical integration of the equations of motion for the resulting system. This is the so-called secular system introduced by Lagrange: instead of the fast motion of the planets, one describes the slow deformations of the planetary orbits (imagining the planets as regions of mass spread along their orbits).

*See also:* Averaging Methods; Bifurcation Theory; Billiards in Bounded Convex Domains; Diagrammatic Techniques in Perturbation Theory; Dynamical Systems and Thermodynamics; Gravitational *N*-Body Problem (Classical); Hamiltonian Systems: Stability and Instability Theory; Hamilton–Jacobi Equations and Dynamical Systems: Variational Aspects; Integrable Systems and Discrete Geometry; KAM Theory and Celestial Mechanics; Localization for Quasiperiodic Potentials; Stability Problems in Celestial Mechanics; Synchronization of Chaos; Weakly Coupled Oscillators.

## Further Reading

Arnol'd VI (1963) Proof of a theorem of A. N. Kolmogorov on the preservation of conditionally periodic motions under a small perturbation of the Hamiltonian. *Russian Mathematical Surveys* 18: 85–192.

Arnol'd VI (1964) Instability of dynamical systems with many degrees of freedom. *Soviet Mathematics Doklady* 5: 581–585.

Arnol'd VI, Kozlov VV, and Neĭshtadt AI (1988) *Dynamical Systems* III. Encyclopedia of Mathematical Sciences, vol. 3. Berlin: Springer.

Bourgain J (1994) Construction of quasi-periodic solutions for Hamiltonian perturbations of linear equations and applications to nonlinear PDE. *International Mathematics Research Notices* 1994(11), 475–497.

Bourgain J (1998) Quasi-periodic solutions of Hamiltonian perturbations of 2D linear Schrödinger equations. *Annals of Mathematics* 148: 363–439.

Bourgain J (2005) *Green's Function Estimates for Lattice Schrödinger Operators and Applications*. Princeton: Princeton University Press.

Celletti A and Chierchia L (1997) On the stability of realistic three-body problems. *Communications in Mathematical Physics* 186: 413–449.

Eliasson LH (1996) Absolutely convergent series expansions for quasi periodic motions. *Mathematical Physics Electronic Journal* 2, paper 4 (electronic). Preprint 1988.

Eliasson LH (1988) Perturbations of stable invariant tori for Hamiltonian systems. *Annali della Scuola Normale Superiore di Pisa* 15: 115–147.

Ford J (1992) The Fermi–Pasta–Ulam problem: paradox turns discovery. *Physics Reports* 213: 271–310.z

Galgani L and Giorgilli A (2003) Recent results on the Fermi–Pasta–Ulam problem. *Rossiĭskaya Akademiya Nauk. Sankt-Peterburgskoe Otdelenie. MatematicheskiĭInstitut im. V. A. Steklova. Zapiski Nauchnykh Seminarov (POMI)* 300: 145–154.

Gallavotti G (1986) Quasi-integrable mechanical systems. In: *Phénomènes critiques, systèmes aléatoires, théories de jauge* (Les Houches, 1984), pp. 539–624. Amsterdam: North-Holland.

Gallavotti G, Bonetto F, and Gentile G (2004) *Aspects of Ergodic, Qualitative and Statistical Theory of Motion*. Berlin: Springer.

Gentile G, Mastropietro V, and Procesi M (2005) Periodic solutions for completely resonant nonlinear wave equations with Dirichlet boundary conditions. *Communications in Mathematical Physics* 257: 319–362.

Kolmogorov AN (1954) On conservation of conditionally periodic motions for a small change in Hamilton's function. *Doklady Akademii Nauk SSSR* 98: 527–530 (Russian).

Kuksin SB (1987) Hamiltonian perturbations of infinite-dimensional linear systems with imaginary spectrum. *Functional Analysis and its Applications* 21: 192–205.

Kuksin SB (1993) *Nearly Integrable Infinite-Dimensional Hamiltonian Systems*, Lecture Notes in Mathematics, vol. 1556. Berlin: Springer.

Kuksin SB and Pöschel J (1996) Invariant Cantor manifolds of quasi-periodic oscillations for a nonlinear Schrödinger equation. *Annals of Mathematics* 143: 149–179.

Laskar J (2004) Chaos in the solar system. In: Iagolnitzer D, Rivasseau V, and Zinn-Justin J (eds.) *Proceedings of the International Conference on Theoretical Physics TH2002*, (Paris, 2002). Basel: Birkhäuser.

Lochak P and Neĭshtadt AI (1992) Estimates of stability time for nearly integrable systems with a quasiconvex Hamiltonian. *Chaos* 2: 495–499.

Moser J (1962) On invariant curves of area-preserving mappings of an annulus. *Nachrichten der Akademie der Wissenschaften in Göttingen* 1962: 1–20.

Moser J (1973) *Stable and Random Motions in Dynamical Systems*, Annals of Mathematical Studies. Princeton: Princeton University Press.

NekhoroševNN (1977) An exponential estimate of the time of stability of nearly integrable Hamiltonian systems. *Russian Mathematical Surveys* 32: 1–65.

Pöschel J (1989) On elliptic lower-dimensional tori in Hamiltonian systems. *Mathematische Zeitscrift* 202: 559–608.

Pöschel J (1993) Nekhoroshev estimates for quasi-convex Hamiltonian systems. *Mathematische Zeitscrift* 213: 187–216.

Rink B (2001) Symmetry and resonance in periodic FPU chains. *Communications in Mathematical Physics* 218: 665–685.

Zabusky NJ and Kruskal MD (1965) Interaction of "solitons" in a collisionless plasma and the recurrence of initial states. *Physical Review Letters* 15: 240–243.


# Standard Model of Particle Physics

**G Altarelli**, CERN, Geneva, Switzerland



## Introduction

The standard model (SM) is a consistent, finite, and – within the limitations of our present technical ability – computable theory of fundamental microscopic interactions that successfully explains most of the known phenomena in elementary particle physics. The SM describes strong, electromagnetic, and weak interactions. All microscopic phenomena observed to date can be attributed to one or the other of these interactions. For example, the forces that hold together the protons and the neutrons in the atomic nuclei are due to strong interactions; the binding of electrons to nuclei in atoms or of atoms in molecules is caused by electromagnetism; and the energy production in the Sun and the other stars occurs through nuclear reactions induced by weak interactions. In principle, gravitational forces should also be included in the list of fundamental interactions but their impact on fundamental particle processes at accessible energies is totally negligible.

The structure of the SM is a generalization of that of quantum electrodynamics (QED), in the sense that it is a renormalizable field theory based on a local symmetry (i.e., separately valid at each spacetime point $x$) that extends the gauge invariance of electrodynamics to a larger set of

conserved currents and charges. There are eight strong charges, called "color" charges and four electroweak charges (which, in particular, include the electric charge). The commutators of these charges form the $SU(3) \otimes SU(2) \otimes U(1)$ algebra. In QED, the interaction between two matter particles with electric charges (e.g., two electrons) is mediated by the exchange of one (or more) photons emitted by one electron and reabsorbed by the second. In the SM the matter fields, all of spin 1/2, are the quarks, the constituents of protons, neutrons, and all hadrons, endowed with both color and electroweak charges, and the leptons (the electron $e^-$, the muon $\mu^-$, the tauon $\tau^-$, plus the three associated neutrinos $\nu_e$, $\nu_\mu$, and $\nu_\tau$) with no color but with electroweak charges. The matter fermions come in three generations or families with identical quantum numbers but different masses. The pattern is as follows:

$$\begin{bmatrix} u & u & u & \nu_e \\ d & d & d & e \end{bmatrix}, \quad \begin{bmatrix} c & c & c & \nu_\mu \\ s & s & s & \mu \end{bmatrix}, \quad \begin{bmatrix} t & t & t & \nu_\tau \\ b & b & b & \tau \end{bmatrix} \qquad [1]$$

Each family contains a weakly charged doublet of quarks, in three color replicas, and a colorless weakly charged doublet with a neutrino and a charged lepton. At present, there is no explanation for this triple repetition of fermion families. The force carriers, of spin 1, are the photon $\gamma$, the weak interaction gauge bosons $W^+$, $W^-$, and $Z_0$ and the eight gluons $g$ that mediate the strong interactions. The photon and the gluons have zero masses as a consequence of the exact conservation of the corresponding symmetry generators, the electric charge and the eight color charges. The weak bosons $W^+$, $W^-$, and $Z_0$ have large masses ($m_W \sim 80.4$ GeV, $m_Z = 91.2$ GeV), signaling that the corresponding symmetries are badly broken. In the SM, the spontaneous breaking of the electroweak gauge symmetry is induced by the Higgs mechanism, which predicts the presence of one (or more) spin 0 particles in the physical spectrum, the Higgs boson(s), not yet experimentally observed. A tremendous experimental effort is underway or planned to reveal the Higgs sector as the last crucial missing link in the SM verification.

## Quantum Chromodynamics

The statement that quantum chromodynamics (QCD) is a renormalizable gauge theory based on the group SU(3) with color triplet quark matter fields fixes the QCD Lagrangian density to be

$$\mathcal{L} = -\frac{1}{4}\sum_{A=1}^{8} F^{A\mu\nu}F^{A}_{\mu\nu} + \sum_{j=1}^{n_f} \bar{q}_j(\mathrm{i}\not{D} - m_j)q_j \qquad [2]$$

Here $q_j$ are the quark fields (of $n_f$ different flavors) with mass $m_j$; $\not{D} = D_\mu\gamma^\mu$, where $\gamma^\mu$ are the Dirac matrices and $D_\mu$ is the covariant derivative

$$D_\mu = \partial_\mu - \mathrm{i}e_s \sum_A t^A g^A_\mu \qquad [3]$$

$e_s$ is the gauge coupling (in analogy with QED,

$$\alpha_s = \frac{e_s^2}{4\pi} \qquad [4]$$

here and throughout this article natural units, $\hbar = c = 1$, are used); $g^A_\mu$, $A = 1, \ldots, 8$, are the gluon fields, and $t^A$ are the SU(3) group generators in the triplet representation of quarks (i.e., $t_A$ are $3 \times 3$ matrices acting on $q$); the generators obey the commutation relations $[t^A, t^B] = \mathrm{i}C_{ABC}t^C$, where $C_{ABC}$ are the complete antisymmetric structure constants of SU(3) (the normalization of $C_{ABC}$ and of $e_s$ is specified by $tr[t^A t^B] = 1/2\delta^{AB}$);

$$F^A_{\mu\nu} = \partial_\mu g^A_\nu - \partial_\nu g^A_\mu - e_s C_{ABC} g^B_\mu g^C_\nu \qquad [5]$$

The physical vertices in QCD include the gluon–quark–antiquark vertex, analogous to the QED photon–fermion–antifermion coupling, but also the three-gluon and four-gluon vertices, of order $e_s$ and $e_s^2$, respectively, which have no analog in an abelian theory like QED. In QED, the photon (a neutral particle) is coupled to all electrically charged particles. In QCD, the gluons are colored, hence self-coupled. This is reflected in the fact that in QED $F_{\mu\nu}$ is linear in the gauge field, so that the term $F^2_{\mu\nu}$ in the Lagrangian is a pure kinetic term, while in QCD $F^A_{\mu\nu}$ is quadratic in the gauge field, so that in $F^{A2}_{\mu\nu}$ we find cubic and quartic vertices beyond the kinetic term.

The QCD Lagrangian in eqn [2] has a simple structure but a very rich dynamical content, including the observed complex spectroscopy with a large number of hadrons. The most prominent properties of QCD are asymptotic freedom and confinement. In field theory, the effective coupling of a given interaction vertex is modified by the interaction. As a result, the measured intensity of the force depends on the transferred (four)momentum squared, $Q^2$, among the participants. In QCD, the relevant coupling parameter that appears in physical processes is $\alpha_s$ (see eqn [4]). Asymptotic freedom means that the effective coupling becomes a function of

$Q^2$: $\alpha_s(Q^2)$ decreases for increasing $Q^2$ and vanishes asymptotically. Thus, the QCD interaction becomes very weak in processes with large $Q^2$, called hard processes or deep inelastic processes (i.e., with a final-state distribution of momenta and a particle content very different from that in the initial state). One can prove that in four spacetime dimensions all gauge theories based on a noncommuting group of symmetry are asymptotically free, and conversely. The effective coupling decreases very slowly at large momenta with the inverse logarithm of $Q^2$: $\alpha_s(Q^2) = 1/b \log Q^2/\Lambda^2$, where $b$ is a known constant and $\Lambda$ is an energy of the order of a few hundred MeV. Since in quantum mechanics large momenta imply short wavelengths, the result is that at short distances the potential between two color charges is similar to the Coulomb potential, that is, proportional to $\alpha_s(r)/r$, with an effective color charge which is small at short distances. On the contrary the interaction strength becomes large at large distances or small transferred momenta, of order $Q \lesssim \Lambda$. In fact, the observed hadrons are tightly bound composite states of quarks, with compensating color charges so that they are overall neutral in color. The property of confinement is the impossibility of separating color charges, like individual quarks and gluons. This is because in QCD the interaction potential between color charges increases, at long distances, linearly in $r$. When we try to separate the quark and the antiquark that form a color-neutral meson the interaction energy grows until pairs of quarks and antiquarks are created from the vacuum and new neutral mesons are coalesced instead of free quarks. For example, consider the process $e^+e^- \to q\bar{q}$ at large center-of-mass energies. The final-state quark and antiquark have large energies, so they separate in opposite directions very fast. But the color-confinement forces create new pairs in between them. Two back-to-back jets of colorless hadrons are observed with a number of slow pions that make the exact separation of the two jets impossible. In some cases, a third well-separated jet of hadrons is also observed: these events correspond to the radiation of an energetic gluon from the parent quark–antiquark pair.

## Electroweak Interactions

We split the electroweak Lagrangian into two parts by separating the Higgs boson couplings:

$$\mathcal{L} = \mathcal{L}_{\text{symm}} + \mathcal{L}_{\text{Higgs}} \qquad [6]$$

We start by specifying $\mathcal{L}_{\text{symm}}$, which involves only gauge bosons and fermions (a sum over all flavors of quarks and leptons, generally indicated by $\psi$ is understood):

$$\mathcal{L}_{\text{symm}} = -\frac{1}{4}\sum_{A=1}^{3} F^A_{\mu\nu}F^{A\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + \bar{\psi}_{\text{L}} \text{i}\gamma^\mu D_\mu \psi_{\text{L}} + \bar{\psi}_{\text{R}} \text{i}\gamma^\mu D_\mu \psi_{\text{R}} \qquad [7]$$

This is the Yang–Mills Lagrangian for the gauge group SU(2) $\otimes$ U(1) with fermion matter fields. Here

$$\begin{aligned} B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \\ F^A_{\mu\nu} &= \partial_\mu W^A_\nu - \partial_\nu W^A_\mu - g\epsilon_{ABC} W^B_\mu W^C_\nu \end{aligned} \qquad [8]$$

are the gauge antisymmetric tensors constructed out of the gauge field $B_\mu$ associated with U(1), and $W^A_\mu$ corresponding to the three SU(2) generators; $\epsilon_{ABC}$ are the group structure constants (see eqn [11]), which, for SU(2), coincide with the totally antisymmetric Levi-Civita tensor (recall the familiar angular-momentum commutators).

The fermion fields are described through their left- and right-hand components:

$$\psi_{\text{L,R}} = [(1 \mp \gamma_5)/2]\psi, \qquad \bar{\psi}_{\text{L,R}} = \bar{\psi}[(1 \pm \gamma_5)/2] \qquad [9]$$

Note that, as given in eqn [9],

$$\begin{aligned} \bar{\psi}_{\text{L}} &= \psi^\dagger_{\text{L}}\gamma_0 = \psi^\dagger[(1-\gamma_5)/2]\gamma_0 \\ &= \bar{\psi}[\gamma_0(1-\gamma_5)/2]\gamma_0 = \bar{\psi}[(1+\gamma_5)/2] \end{aligned}$$

The matrices $P_\pm = (1 \pm \gamma_5)/2$ are projectors. They satisfy the relations $P_\pm P_\pm = P_\pm, P_\pm P_\mp = 0$, $P_+ + P_- = 1$.

The standard electroweak theory is a chiral theory, in the sense that $\psi_{\text{L}}$ and $\psi_{\text{R}}$ behave differently under the gauge group. In particular, all $\psi_{\text{R}}$ are singlets and all $\psi_{\text{L}}$ are doublets in the minimal SM (MSM). Thus, mass terms for fermions (of the form $\bar{\psi}_{\text{L}}\psi_{\text{R}}$ + h.c.) are forbidden in the symmetric limit. Fermion masses are introduced, together with $W^\pm$ and $Z$ masses, by the mechanism of symmetry breaking. The covariant derivatives $D_\mu\psi_{\text{L,R}}$ are explicitly given by

$$D_\mu\psi_{\text{L,R}} = \left[\partial_\mu + \text{i}g\sum_{A=1}^{3} t^A_{\text{L,R}} W^A_\mu + \text{i}g'\frac{1}{2}Y_{\text{L,R}}B_\mu\right]\psi_{\text{L,R}} \qquad [10]$$

where $t^A_{\text{L,R}}$ and $1/2Y_{\text{L,R}}$ are the SU(2) and U(1) generators, respectively, in the reducible representations $\psi_{\text{L,R}}$. The commutation relations of the SU(2) generators are given by

$$[t^A_{\text{L}}, t^B_{\text{L}}] = \text{i}\epsilon_{ABC}t^C_{\text{L}} \quad \text{and} \quad [t^A_{\text{R}}, t^B_{\text{R}}] = \text{i}\epsilon_{ABC}t^C_{\text{R}} \qquad [11]$$

We use the normalization $\text{tr}[t^At^B] = 1/2\delta^{AB}$ in the fundamental representation of SU(2). The electric

charge generator $Q$ (in units of $e$, the positron charge) is given by

$$Q = t_{\mathrm{L}}^3 + 1/2\, Y_{\mathrm{L}} = t_{\mathrm{R}}^3 + 1/2\, Y_{\mathrm{R}} \qquad [12]$$

All fermion couplings to the gauge bosons can be derived directly from eqns [7] and [10]. The charged-current (CC) couplings are the simplest. From

$$\begin{aligned} g\left(t^1 W_\mu^1 + t^2 W_\mu^2\right) &= g\left\{\left[\left(t^1 + \mathrm{i}t^2\right)/\sqrt{2}\right]\right. \\ &\quad \times \left.\left[\left(W_\mu^1 - \mathrm{i}W_\mu^2\right)/\sqrt{2}\right] + \text{h.c.}\right\} \\ &= g\left\{\left[\left(t^+ W_\mu^-\right)/\sqrt{2}\right] + \text{h.c.}\right\} \end{aligned} \qquad [13]$$

where $t^\pm = t^1 \pm \mathrm{i}t^2$ and $W^\pm = (W^1 \pm \mathrm{i}W^2)/\sqrt{2}$, we obtain the vertex

$$\begin{aligned} V_{\bar\psi\psi W} &= g\bar\psi\gamma_\mu\left[\left(t_{\mathrm{L}}^+/\sqrt{2}\right)(1-\gamma_5)/2 + \left(t_{\mathrm{R}}^+/\sqrt{2}\right)\right. \\ &\quad \times \left.(1+\gamma_5)/2\right]\psi W_\mu^- + \text{h.c.} \end{aligned} \qquad [14]$$

In the neutral-current (NC) sector, the photon $A_\mu$ and the mediator $Z_\mu$ of the weak NC are orthogonal and normalized linear combinations of $B_\mu$ and $W_\mu^3$:

$$\begin{aligned} A_\mu &= \cos\theta_{\mathrm{W}} B_\mu + \sin\theta_{\mathrm{W}} W_\mu^3 \\ Z_\mu &= -\sin\theta_{\mathrm{W}} B_\mu + \cos\theta_{\mathrm{W}} W_\mu^3 \end{aligned} \qquad [15]$$

Equations [15] define the weak mixing angle $\theta_{\mathrm{W}}$. The photon is characterized by equal couplings to left and right fermions with a strength equal to the electric charge. Recalling eqn [12] for the charge matrix $Q$, we immediately obtain

$$g\sin\theta_{\mathrm{W}} = g'\cos\theta_{\mathrm{W}} = e \qquad [16]$$

or, equivalently,

$$\tan\theta_{\mathrm{W}} = g'/g \qquad [17]$$

Once $\theta_{\mathrm{W}}$ has been fixed by the photon couplings, it is a simple matter of algebra to derive the $Z$ couplings, with the result

$$\begin{aligned} \Gamma_{\bar\psi\psi Z} &= g/(2\cos\theta_{\mathrm{W}})\bar\psi\gamma_\mu\left[t_{\mathrm{L}}^3(1-\gamma_5) + t_{\mathrm{R}}^3(1+\gamma_5)\right. \\ &\quad \left.-2Q\sin^2\theta_{\mathrm{W}}\right]\psi Z^\mu \end{aligned} \qquad [18]$$

where $\Gamma_{\bar\psi\psi Z}$ is a notation for the vertex. In the MSM, $t_{\mathrm{R}}^3 = 0$ and $t_{\mathrm{L}}^3 = \pm 1/2$. Note that the CC and NC weak couplings do not conserve P (parity) and C (charge conjugation).

In order to derive the effective four-fermion interactions that are equivalent, at low energies, to the CC and NC couplings given in eqns [14] and [18], we anticipate that large masses, as experimentally observed, are provided for $W^\pm$ and $Z$ by $\mathcal{L}_{\mathrm{Higgs}}$. For left–left CC couplings, when the momentum transfer squared can be neglected with respect to $m_{\mathrm{W}}^2$ in the propagator of Born diagrams with single $W$ exchange, from eqn [14], we can write

$$\begin{aligned} \mathcal{L}_{\mathrm{eff}}^{\mathrm{CC}} &\simeq \left(g^2/8m_{\mathrm{W}}^2\right)\left[\bar\psi\gamma_\mu(1-\gamma_5)t_{\mathrm{L}}^+\psi\right] \\ &\quad \times \left[\bar\psi\gamma^\mu(1-\gamma_5)t_{\mathrm{L}}^-\psi\right] \end{aligned} \qquad [19]$$

By specializing further in the case of doublet fields such as $\nu_e - e^-$ or $\nu_\mu - \mu^-$, we obtain the tree-level relation of $g$ with the Fermi coupling constant $G_{\mathrm{F}}$ measured from $\mu$ decay ($G_{\mathrm{F}} = 1.16639(2) \times 10^{-5}\,\mathrm{GeV}^{-2}$):

$$G_{\mathrm{F}}/\sqrt{2} = g^2/8m_{\mathrm{W}}^2 \qquad [20]$$

By recalling that $g\sin\theta_{\mathrm{W}} = e$, we can also cast this relation in the form

$$m_{\mathrm{W}} = \mu_{\mathrm{Born}}/\sin\theta_{\mathrm{W}} \qquad [21]$$

with

$$\mu_{\mathrm{Born}} = \left(\pi\alpha/\sqrt{2}G_{\mathrm{F}}\right)^{1/2} \simeq 37.2802\,\mathrm{GeV} \qquad [22]$$

where $\alpha$ is the fine-structure constant of QED ($\alpha \equiv e^2/4\pi = 1/137.036$).

In the same way, for neutral currents we obtain, in Born approximation, from eqn [18], the effective four-fermion interaction given by

$$\mathcal{L}_{\mathrm{eff}}^{\mathrm{NC}} \simeq \sqrt{2}G_{\mathrm{F}}\rho_0\bar\psi\gamma_\mu[\ldots]\psi\bar\psi\gamma^\mu[\ldots]\psi \qquad [23]$$

where

$$[\ldots] \equiv t_{\mathrm{L}}^3(1-\gamma_5) + t_{\mathrm{R}}^3(1+\gamma_5) - 2Q\sin^2\theta_{\mathrm{W}} \qquad [24]$$

and

$$\rho_0 = m_{\mathrm{W}}^2/m_{\mathrm{Z}}^2\cos^2\theta_{\mathrm{W}} \qquad [25]$$

All couplings given in this section are obtained at tree level and are modified in higher orders of perturbation theory. In particular, the relations between $m_{\mathrm{W}}$ and $\sin\theta_{\mathrm{W}}$ (eqns [21] and [22]) and the observed values of $\rho$ ($\rho = \rho_0$ at tree level) in different NC processes are altered by computable small electroweak radiative corrections.

The gauge-boson self-interactions can be derived from the $F_{\mu\nu}$ term in $\mathcal{L}_{\mathrm{symm}}$, by using eqn [15] and $W^\pm = (W^1 \pm \mathrm{i}W^2)/\sqrt{2}$. For the three-gauge-boson vertex $W^+W^-V$ with $V = Z, \gamma$, we obtain

$$\begin{aligned} \Gamma_{W^-W^+V} &= \mathrm{i}g_{W^-W^+V}\left[g_{\mu\nu}(q-p)_\lambda + g_{\mu\lambda}(p-r)_\nu\right. \\ &\quad \left.+ g_{\nu\lambda}(r-q)_\mu\right] \end{aligned} \qquad [26]$$

with

$$\begin{aligned} g_{W^-W^+\gamma} &= g\sin\theta_{\mathrm{W}} = e \quad \text{and} \\ g_{W^-W^+Z} &= g\cos\theta_{\mathrm{W}} \end{aligned} \qquad [27]$$

This form of the triple gauge vertex is very special: in general, there could be departures from the above SM expression, even restricting us to $SU(2) \otimes U(1)$ gauge symmetric and C and P invariant couplings. In fact, some small corrections are already induced by the radiative corrections. The SM form of the triple gauge vertex has been experimentally confirmed by measuring the cross section $e^+e^- \rightarrow W^+W^-$ at LEP.

We now turn to the Higgs sector of the electroweak Lagrangian. The Higgs Lagrangian is specified by the gauge principle and the requirement of renormalizability to be

$$\mathcal{L}_{\text{Higgs}} = \left(D_\mu \phi\right)^\dagger \left(D^\mu \phi\right) - V\left(\phi^\dagger \phi\right) - \bar{\psi}_{\text{L}} \Gamma \psi_{\text{R}} \phi - \bar{\psi}_{\text{R}} \Gamma^\dagger \psi_{\text{L}} \phi^\dagger \qquad [28]$$

where $\phi$ is a column vector including all Higgs scalar fields; it transforms as a reducible representation of the gauge group. The quantities $\Gamma$ (which include all coupling constants) are matrices that make the Yukawa couplings invariant under the Lorentz and gauge groups. The potential $V(\phi^\dagger\phi)$, symmetric under $SU(2) \otimes U(1)$, contains, at most, quartic terms in $\phi$ so that the theory is renormalizable:

$$V\left(\phi^\dagger \phi\right) = -\tfrac{1}{2}\mu^2 \phi^\dagger \phi + \tfrac{1}{4}\lambda\left(\phi^\dagger \phi\right)^2 \qquad [29]$$

Spontaneous symmetry breaking is induced if the minimum of $V$, which is the classical analog of the quantum-mechanical vacuum state (both are the states of minimum energy) is obtained for nonvanishing $\phi$ values. This occurs because we have taken $\mu^2$ and $\lambda$ positive in $V$ (note the "wrong" sign of the mass term). Precisely, we denote the vacuum expectation value (VEV) of $\phi$, that is, the position of the minimum, by $v$:

$$\langle 0|\phi(x)|0\rangle = v \neq 0 \qquad [30]$$

The fermion mass matrix is obtained from the Yukawa couplings by replacing $\phi(x)$ by $v$:

$$M = \bar{\psi}_{\text{L}} \mathcal{M} \psi_{\text{R}} + \bar{\psi}_{\text{R}} \mathcal{M}^\dagger \psi_{\text{L}} \qquad [31]$$

with

$$\mathcal{M} = \Gamma \cdot v \qquad [32]$$

In the SM, where all left fermions, $\psi_{\text{L}}$, are doublets and all right fermions, $\psi_{\text{R}}$, are singlets, only Higgs doublets can contribute to fermion masses. There are enough free couplings in $\Gamma$, so that one single complex Higgs doublet is indeed sufficient to generate the most general fermion mass matrix. It is important to observe that by a suitable change of basis we can always make the matrix $\mathcal{M}$ Hermitian, $\gamma_5$-free and diagonal. In fact, we can make separate unitary transformations on $\psi_{\text{L}}$ and $\psi_{\text{R}}$ according to

$$\psi'_{\text{L}} = U\psi_{\text{L}}, \qquad \psi'_{\text{R}} = V\psi_{\text{R}} \qquad [33]$$

and consequently

$$\mathcal{M} \rightarrow \mathcal{M}' = U^\dagger \mathcal{M} V \qquad [34]$$

This transformation does not alter the general structure of the fermion couplings in $\mathcal{L}_{\text{symm}}$.

If only one Higgs doublet is present, the change of basis that makes $\mathcal{M}$ diagonal will at the same time diagonalize also the fermion–Higgs Yukawa couplings. Thus, in this case, no flavor-changing neutral Higgs exchanges are present. This is not true, in general, when there are several Higgs doublets. But one Higgs doublet for each electric charge sector, that is, one doublet coupled only to $u$-type quarks, one doublet to $d$-type quarks, one doublet to charged leptons would also be satisfactory, because the mass matrices of fermions with different charges are diagonalized separately. In fact, at the moment, the simplest model with only one Higgs doublet seems adequate for describing all observed phenomena.

Weak charged currents are the only tree-level interactions in the SM that change flavor: by emission of a $W$, a $u$-type quark is turned into a $d$-type quark, or a $\nu_l$ neutrino is turned into an $l^-$ charged lepton (all fermions are left-handed). If we start from a $u$-type quark that is a mass eigenstate, emission of a $W$ turns it into a $d$-type quark state $d'$ (the weak isospin partner of $u$) that in general is not a mass eigenstate. In general, the mass eigenstates and the weak eigenstates do not coincide and a unitary transformation connects the two sets:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V \begin{pmatrix} d \\ s \\ b \end{pmatrix} \qquad [35]$$

or, in shorthand, $D' = VD$, where $V$ is the Cabibbo–Kobayashi–Maskawa (CKM) matrix. Thus, in terms of mass eigenstates the charged weak current of quarks is of the form

$$J^+_\mu \propto \bar{u}\gamma_\mu(1 - \gamma_5)VD \qquad [36]$$

Since $V$ is unitary (i.e., $VV^\dagger = V^\dagger V = 1$) and commutes with $T^2$, $T_3$, and Q (because all $d$-type quarks have the same isospin and charge) the neutral current couplings are diagonal both in the primed and unprimed basis (if the $Z$ $d$-type quark current is abbreviated as $\bar{D}'\Gamma D'$ then by changing basis we get $\bar{D}V^\dagger\Gamma VD$ and $V$ and $\Gamma$ commute because, as seen from eqn [24], $\Gamma$ is made of Dirac matrices and $T_3$ and Q generator matrices). It follows that $\bar{D}'\Gamma D' = \bar{D}\Gamma D$. This is the Glashow–Iliopoulos–Maiani (GIM)

mechanism that ensures natural flavor conservation of the neutral current couplings at the tree level. For three generations of quarks, the CKM matrix depends on four physical parameters: three mixing angles and one phase. This phase is the unique source of CP violation in the SM.

We now consider the gauge-boson masses and their couplings to the Higgs. These effects are induced by the $(D_\mu\phi)^\dagger(D^\mu\phi)$ term in $\mathcal{L}_{\text{Higgs}}$ (eqn [28]), where

$$D_\mu\phi = \left[\partial_\mu + \mathrm{i}g\sum_{A=1}^{3} t^A W_\mu^A + \mathrm{i}g'(Y/2)B_\mu\right]\phi \qquad [37]$$

Here $t^A$ and $1/2Y$ are the $\mathrm{SU}(2)\otimes\mathrm{U}(1)$ generators in the reducible representation spanned by $\phi$. Not only doublets but all non-singlet Higgs representations can contribute to gauge-boson masses. The condition that the photon remains massless is equivalent to the condition that the vacuum is electrically neutral:

$$Q|v\rangle = \left(t^3 + \tfrac{1}{2}Y\right)|v\rangle = 0 \qquad [38]$$

The charged $W$ mass is given by the quadratic terms in the $W$ field arising from $\mathcal{L}_{\text{Higgs}}$, when $\phi(x)$ is replaced by $v$. We obtain

$$m_{\mathrm{W}}^2 W_\mu^+ W^{-\mu} = g^2\left|\left(t^+ v/\sqrt{2}\right)\right|^2 W_\mu^+ W^{-\mu} \qquad [39]$$

whilst for the $Z$ mass we get (recalling eqn [15])

$$\tfrac{1}{2}m_{\mathrm{Z}}^2 Z_\mu Z^\mu = \left|\left[g\cos\theta_{\mathrm{W}} t^3 - g'\sin\theta_{\mathrm{W}}(Y/2)\right]v\right|^2 Z_\mu Z^\mu \qquad [40]$$

where the factor of 1/2 on the left-hand side is the correct normalization for the definition of the mass of a neutral field. For Higgs doublets

$$\phi = \begin{pmatrix}\phi^+\\ \phi^0\end{pmatrix}, \qquad v = \begin{pmatrix}0\\ v\end{pmatrix} \qquad [41]$$

we obtain

$$m_{\mathrm{W}}^2 = 1/2g^2v^2, \qquad m_{\mathrm{Z}}^2 = 1/2g^2v^2/\cos^2\theta_{\mathrm{W}} \qquad [42]$$

Note that by using eqn [20] we obtain

$$v = 2^{-3/4}G_{\mathrm{F}}^{-1/2} = 174.1\,\mathrm{GeV} \qquad [43]$$

It is also evident that for Higgs doublets

$$\rho_0 = m_{\mathrm{W}}^2/m_{\mathrm{Z}}^2\cos^2\theta_{\mathrm{W}} = 1 \qquad [44]$$

This relation is typical of one or more Higgs doublets and would be spoiled by the existence of, for example, Higgs triplets. This result is valid at the tree level and is modified by calculable small electroweak radiative corrections. The $\rho_0$ parameter has been measured from the intensity of NC interactions (recall eqn [25]) and confirmed to be close to unity at a few per milli level.

In MSM only one Higgs doublet is present. Then the fermion–Higgs couplings are in proportion to the fermion masses. In fact, from the Yukawa couplings $g_{\phi\bar{f}f}(\bar{f}_{\mathrm{L}}\phi f_{\mathrm{R}} + \mathrm{h.c.})$, the mass $m_f$ is obtained by replacing $\phi$ by $v$, so that $m_f = g_{\phi\bar{f}f}v$. In MSM, three out of the four Hermitian fields are removed from the physical spectrum by the Higgs mechanism and become the longitudinal modes of $W^+, W^-$, and $Z$ which acquire a mass. The fourth neutral Higgs is physical and should be found. If more doublets are present, two more charged and two more neutral Higgs scalars should be around for each additional doublet.

The couplings of the physical Higgs $H$ to the gauge bosons can be simply obtained from $\mathcal{L}_{\text{Higgs}}$, by the replacement

$$\phi(x) = \begin{pmatrix}\phi^+(x)\\ \phi^0(x)\end{pmatrix} \rightarrow \begin{pmatrix}0\\ v + (H/\sqrt{2})\end{pmatrix} \qquad [45]$$

(so that $(D_\mu\phi)^\dagger(D^\mu\phi) = 1/2(\partial_\mu H)^2 + \cdots$), with the result

$$\begin{aligned}\mathcal{L}[H, W, Z] &= g^2\left(v/\sqrt{2}\right)W_\mu^+W^{-\mu}H + \left(g^2/4\right)W_\mu^+W^{-\mu}H^2\\ &\quad + \left[\left(g^2vZ_\mu Z^\mu\right)/\left(2\sqrt{2}\cos^2\theta_{\mathrm{W}}\right)\right]H\\ &\quad + \left[g^2/\left(8\cos^2\theta_{\mathrm{W}}\right)\right]Z_\mu Z^\mu H^2\end{aligned}$$

In MSM, the Higgs mass $m_{\mathrm{H}}^2 \sim \lambda v^2$ is of order of the weak scale $v$ but cannot be predicted because the value of $\lambda$ is not fixed. The dominant decay mode of the Higgs is in the $b\bar{b}$ channel below the $WW$ threshold, while the $W^+W^-$ channel is dominant for sufficiently large $m_{\mathrm{H}}$. The width is small below the $WW$ threshold, not exceeding a few MeV, but increases steeply beyond the threshold, reaching the asymptotic value of $\Gamma \sim 1/2m_{\mathrm{H}}^3$ at large $m_{\mathrm{H}}$, where all energies and masses are in TeV.

A central role in the experimental verification of the standard electroweak theory has been played by CERN, the European Laboratory for Particle Physics, located near Geneva, between France and Switzerland. The indirect effects of the $Z_0$, that is, the occurrence of weak processes induced by the neutral current, were first observed in 1974 at CERN by the Collaboration Gargamelle (the name of the bubble chamber used in the experiment). Later, in 1982, the $W^\pm$ and the $Z_0$ were, for the first time, directly produced and observed in proton–antiproton collisions by the UA1 and UA2 collaborations and then further studied with the same technique both at CERN and subsequently at the Tevatron of Fermilab near Chicago. Starting from 1989 LEP, the large $e^+e^-$ collider was functioning at CERN till 2000. In the LEP circular ring of circumference $\sim$27 km, electrons and

positrons were accelerated in opposite directions to an equal energy in the range between 45 and 103 GeV. The beams were made to cross and collide in correspondence of four experimental areas where the ALEPH, DELPHI, L3, and OPAL detectors were located to study the final states produced in the collisions. In its first phase, called LEP1, from 1989 to 1995 the LEP operation had been completely dedicated to a precise study of the $Z_0$ properties, mass, lifetime, and decay modes in order to accurately test the predictions of the SM. The main lessons of the precision tests of the standard electroweak theory can be summarized as follows. It has been checked that the couplings of quarks and leptons to the weak gauge bosons $W^{\pm}$ and $Z$ are indeed precisely those prescribed by the gauge symmetry. The accuracy of a few tenths of 1% for these tests implies that, not only the tree level, but also the structure of quantum corrections has been verified. Then, since the end of 1995, the energy of LEP was increased and the phase of LEP2 was started. The total energy was gradually increased up to 206 GeV. The main physics goals of LEP2 were the search for the Higgs and for possible new particles, the precise measurement of $m_W$ and the experimental study of the triple gauge vertices $WW\gamma$ and $WWZ_0$. The Higgs particle of the SM could in principle be produced at LEP2 in the reaction $e+e^- \to Z_0 H$, which proceeds by $Z_0$ exchange. The nonobservation of the Higgs particle at LEP2 has allowed to establish a lower limit on its mass: $m_{\rm H} \gtrsim 114$ GeV. Indirect indications on the Higgs mass were also obtained from the precision tests of the SM, as the radiative effects depend logarithmically on $m_{\rm H}$. The indication is that the Higgs mass cannot be too heavy if the SM is valid: $m_{\rm H} \lesssim 219$ GeV at 95% c.l. In 2001, LEP was dismantled and, in its tunnel, a new double ring of superconducting magnets is being installed. The new accelerator, the LHC (Large Hadron Collider), will be a proton–proton collider of total center-of-mass energy 14 TeV. Two large experiments ATLAS and CMS will continue to search for the Higgs starting in the year 2007. The sensitivity of LHC experiments to the SM Higgs will go up to masses $m_{\rm H}$ of $\sim$1 TeV.

*See also:* Effective Field Theories; Electric–Magnetic Duality; Electroweak Theory; General Relativity: Experimental Tests; Noncommutative Geometry and the Standard Model; Perturbative Renormalization Theory and BRST; Quantum Chromodynamics; Quantum Electrodynamics and its Precision Tests; Quantum Field Theory: a Brief Introduction; Relativistic Wave Equations Including Higher Spin Fields; Renormalization: General Theory; Supersymmetric Particle Models.

## Further Reading

Altarelli G (2000) *The Standard Electroweak Theory and Beyond*, Proceedings of the Summer School on Phenomenology of Gauge Interactions. Zuoz, Switzerland hep-ph/0011078.

Altarelli G (2001) *A QCD Primer*, Proceedings of the 2001 European School of High Energy Physics. Beatenberg, Switzerland, hep-ph/0204179.

Close F, Marten, and Sutton MC (2002) *The Particle Odyssey*. Oxford: Oxford University Press.

Fraser G (ed.) (1998) *The Particle Century*. London: The Institute of Physics.

Martin BR and Shaw G (1997) *Particle Physics*, 2nd edn. Chichester: Wiley.

Particle Data Group (2004) Review of particle physics. *Physics Letters B* 592: 1.

Perkins DH (2000) *Introduction to High Energy Physics*. Reading: Addison-Wesley.

# Stationary Black Holes

**R Beig**, Universität Wien, Vienna, Austria
**P T Chruściel**, Université de Tours, Tours, France



## Introduction

This article treats a specific class of stationary solutions to the Einstein field equations which read

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu} \qquad [1]$$

Here $R_{\mu\nu}$ and $R = g^{\mu\nu} R_{\mu\nu}$ are, respectively, the Ricci tensor and the Ricci scalar of the spacetime metric $g_{\mu\nu}$, $G$ the Newton constant, and $c$ the speed of light. The tensor $T_{\mu\nu}$ is the stress–energy tensor of matter. Spacetimes, or regions thereof, where $T_{\mu\nu} = 0$ are called vacuum.

Stationary solutions are of interest for a variety of reasons. As models for compact objects at rest, or in steady rotation, they play a key role in astrophysics. They are easier to study than nonstationary systems because stationary solutions are governed by elliptic rather than hyperbolic equations. Finally, like in any field theory, one expects that large classes of dynamical solutions approach ("settle down to") a stationary state in the final stages of their evolution.

The simplest stationary solutions describing compact isolated objects are the spherically symmetric

ones. In the vacuum region, these are all given by the Schwarzschild family. A theorem of Birkhoff shows that in the vacuum region any spherically symmetric metric, even without assuming stationarity, belongs to the family of Schwarzschild metrics, parametrized by a positive mass parameter $m$. Thus, regardless of possible motions of the matter, as long as they remain spherically symmetric, the exterior metric is the Schwarzschild one for some constant $m$. This has the following consequence for stellar dynamics: imagine following the collapse of a cloud of pressureless fluid ("dust"). Within Newtonian gravity, this dust cloud will, after finite time, contract to a point at which the density and the gravitational potential diverge. However, this result cannot be trusted as a sensible physical prediction because, even if one supposes that Newtonian gravity is still valid at very high densities, a matter model based on noninteracting point particles is certainly not. Consider, next, the same situation in the Einstein theory of gravity: here a new question arises, related to the form of the Schwarzschild metric outside of the spherically symmetric body:

$$
\begin{aligned}
g &= -V^2\,\mathrm{d}t^2 + V^{-2}\,\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2,\\
V^2 &= 1 - \frac{2Gm}{rc^2},\\
t &\in \mathbb{R}, \quad r \in \left(\frac{2Gm}{c^2}, \infty\right)
\end{aligned}
\qquad [2]
$$

Here $\mathrm{d}\Omega^2$ is the line element of the standard 2-sphere. Since the metric [2] seems to be singular as $r = 2m$ is approached (from now on, we use units in which $G = c = 1$), there arises the need to understand what happens at the surface of the star when the radius $r = 2m$ is reached. One thus faces the need of a careful study of the geometry of the metric [2] when $r = 2m$ is approached, and crossed.

The first key feature of the metric [2] is its stationarity, of course, with Killing vector field $X$ given by $X = \partial_t$. A Killing field, by definition, is a vector field the local flow of which generates isometries. A spacetime (the term spacetime denotes a smooth, paracompact, connected, orientable, and time-orientable Lorentzian manifold) is called stationary if there exists a Killing vector field $X$ which approaches $\partial_t$ in the asymptotically flat region (where $r$ goes to $\infty$; see below for precise definitions) and generates a one-parameter group of isometries. A spacetime is called static if it is stationary and if the stationary Killing vector $X$ is hypersurface orthogonal, that is, $X^\flat \wedge \mathrm{d}X^\flat = 0$, where $X^\flat = X_\mu\,\mathrm{d}x^\mu = g_{\mu\nu}X^\nu\,\mathrm{d}x^\mu$. A spacetime is called axisymmetric if there exists a Killing vector field $Y$, which generates a one-parameter group of isometries and which behaves like a rotation in the asymptotically flat region, with all orbits $2\pi$-periodic. In asymptotically flat spacetimes, this implies that there exists an axis of symmetry, that is, a set on which the Killing vector vanishes. Killing vector fields which are a nontrivial linear combination of a time translation and of a rotation in the asymptotically flat region are called stationary rotating, or helical.

There exists a technique, due independently to Kruskal and Szekeres, of attaching together two regions $r > 2m$ and two regions $r < 2m$ of the Schwarzschild metric, as in Figure 1, to obtain a manifold with a metric which is smooth at $r = 2m$. In the extended spacetime, the hypersurface $\{r = 2m\}$ is a null hypersurface $\mathscr{E}$, the Schwarzschild event horizon. The stationary Killing vector $X = \partial_t$ extends to a Killing vector in the extended spacetime which becomes tangent to and null on $\mathscr{E}$. The global properties of the Kruskal–Szekeres extension of the exterior Schwarzschild spacetime make this spacetime a natural model for a nonrotating black hole. It is worth noting here that the exterior Schwarzschild spacetime [2] admits an infinite number of nonisometric vacuum extensions, even in the class of maximal, analytic, simply connected ones. The Kruskal–Szekeres extension is singled out by the properties that it is maximal, vacuum, analytic, simply connected, with all maximally extended geodesics either complete, or with the area $r$ of the orbits of the isometry groups tending to zero along them.

We can now come back to the problem of the contracting dust cloud according to the Einstein theory. For simplicity, we take the density of the dust to be uniform – the so-called Oppenheimer–Snyder solution. It then turns out that, in the course of collapse, the surface of the dust will eventually cross the Schwarzschild radius, leaving behind a Schwarzschild black hole. If one follows the dust cloud further, a singularity will eventually form, but will not be visible from the "outside region" where $r > 2m$. For a collapsing body of the mass of the Sun, say, one has $2m = 3\,\mathrm{km}$. Thus, standard phenomenological matter models such as that for dust can still be trusted, so that the previous objection to the Newtonian scenario does not apply.

There is a rotating generalization of the Schwarzschild metric, namely the two-parameter family of exterior Kerr metrics, which in Boyer–Lindquist coordinates takes the form

$$
\begin{aligned}
g = &-\frac{\Delta - a^2\sin^2\theta}{\Sigma}\,\mathrm{d}t^2 - \frac{2a\sin^2\theta(r^2 + a^2 - \Delta)}{\Sigma}\,\mathrm{d}t\,\mathrm{d}\varphi\\
&+ \frac{(r^2 + a^2)^2 - \Delta a^2\sin^2\theta}{\Sigma}\sin^2\theta\,\mathrm{d}\varphi^2\\
&+ \frac{\Sigma}{\Delta}\,\mathrm{d}r^2 + \Sigma\,\mathrm{d}\theta^2
\end{aligned}
\qquad [3]
$$

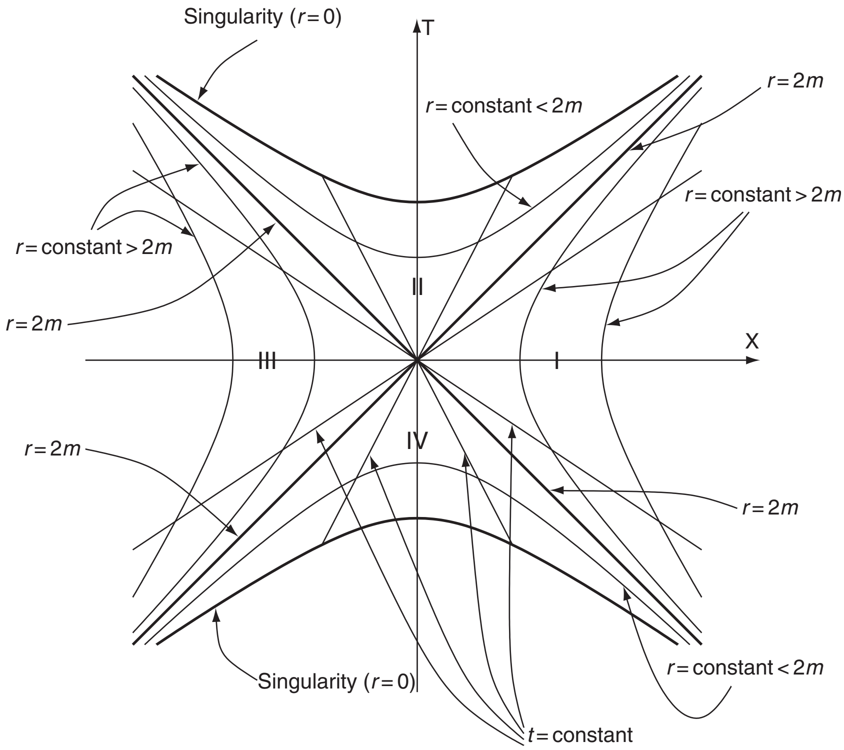



**Figure 1** The Kruskal–Szekeres extension of the Schwarzschild solution. (Adapted with permission from Nicolas J-P (2002) Dirac fields on asymptotically flat space-times. *Dissertationes Mathematicae* 408: 1–85.)

with $0 \leq a < m$. Here $\Sigma = r^2 + a^2 \cos^2\theta, \Delta = r^2 + a^2 - 2mr$ and $r_+ < r < \infty$ where $r_+ = m + (m^2 - a^2)^{1/2}$. When $a = 0$, the Kerr metric reduces to the Schwarzschild metric. The Kerr metric is again a vacuum solution, and it is stationary with $X = \partial_t$ the asymptotic time translation, as well as axisymmetric with $Y = \partial_\varphi$ the generator of rotations. Similarly to the Schwarzschild case, it turns out that the metric can be smoothly extended across $r = r_+$, with $\{r = r_+\}$ being a smooth null hypersurface $\mathscr{E}$ in the extension. The null generator $K$ of $\mathscr{E}$ is the limit of the stationary-rotating Killing field $X + \omega Y$, where $\omega = a/(2mr_+)$. On the other hand, the Killing vector $X$ is timelike only outside the hypersurface $\{r = m + (m^2 - a^2 \cos^2\theta)^{1/2}\}$, on which $X$ becomes null. In the region between $r_+$ and $r = m + (m^2 - a^2\cos^2\theta)^{1/2}$, which is called the ergoregion, $X$ is spacelike. It is also spacelike on and tangent to $\mathscr{E}$, except where the axis of rotation meets $\mathscr{E}$, where $X$ is null. Based on the above properties, the Kerr family provides natural models for rotating black holes.

Unfortunately, as opposed to the spherically symmetric case, there are no known explicit collapsing solutions with rotating matter, in particular no known solutions having the Kerr metric as final state.

The aim of the theory outlined below is to understand the general geometrical features of stationary black holes, and to give a classification of models satisfying the field equations.

## Model-Independent Concepts

Some of the notions used informally in the introductory section will now be made more precise. The mathematical notion of black hole is meant to capture the idea of a region of spacetime which cannot be seen by "outside observers." Thus, at the outset, one assumes that there exists a family of physically preferred observers in the spacetime under consideration. When considering isolated physical systems, it is natural to define the "exterior observers" as observers which are "very far" away from the system under consideration. The standard way of making this mathematically precise is by using conformal completions, discussed in more detail in the article about asymptotic structure in this encyclopedia: a pair $(\tilde{\mathscr{M}}, \tilde{g})$ is called a conformal completion at infinity, or simply conformal completion, of $(\mathscr{M}, g)$ if $\tilde{\mathscr{M}}$ is a manifold with boundary such that:

1. $\mathscr{M}$ is the interior of $\tilde{\mathscr{M}}$;
2. there exists a function $\Omega$, with the property that the metric $\tilde{g}$, defined as $\Omega^2 g$ on $\mathscr{M}$, extends by continuity to the boundary of $\tilde{\mathscr{M}}$, with the

extended metric remaining of Lorentzian signature; and

3. $\Omega$ is positive on $\mathscr{M}$, differentiable on $\tilde{\mathscr{M}}$, vanishes on the boundary

$$\mathscr{I} := \tilde{\mathscr{M}} \setminus \mathscr{M}$$

with $d\Omega$ nowhere vanishing on $\mathscr{I}$.

The boundary $\mathscr{I}$ of $\tilde{\mathscr{M}}$ is called Scri, a phonic shortcut for "script I." The idea here is the following: forcing $\Omega$ to vanish on $\mathscr{I}$ ensures that $\mathscr{I}$ lies infinitely far away from any physical object – a mathematical way of capturing the notion "very far away." The condition that $d\Omega$ does not vanish is a convenient technical condition which ensures that $\mathscr{I}$ is a smooth three-dimensional hypersurface, instead of some, say, one- or two-dimensional object, or of a set with singularities here and there. Thus, $\mathscr{I}$ is an idealized description of a family of observers at infinity.

To distinguish between various points of $\mathscr{I}$, one sets

$$\mathscr{I}^+ = \{\text{points in } \mathscr{I} \text{ which are to the future of the physical spacetime}\}$$

$$\mathscr{I}^- = \{\text{points in } \mathscr{I} \text{ which are to the past of the physical spacetime}\}$$

(Recall that a point $q$ is to the future, respectively to the past, of $p$ if there exists a future directed, respectively past directed, causal curve from $p$ to $q$. Causal curves are curves $\gamma$ such that their tangent vector $\dot{\gamma}$ is causal everywhere, $g(\dot{\gamma}, \dot{\gamma}) \leq 0$.) One then defines the black hole region $\mathscr{B}$ as

$$\mathscr{B} := \{\text{points in } \mathscr{M} \text{ which are not in the past of } \mathscr{I}^+\} \qquad [4]$$

By definition, points in the black hole region cannot thus send information to $\mathscr{I}^+$; equivalently, observers on $\mathscr{I}^+$ cannot see points in $\mathscr{B}$. The white-hole region $\mathscr{W}$ is defined by changing the time orientation in [4]. A key notion related to the concept of a black hole is that of future ($\mathscr{E}^+$) and past ($\mathscr{E}^-$) event horizons,

$$\mathscr{E}^+ := \partial\mathscr{B}, \qquad \mathscr{E}^- := \partial\mathscr{W} \qquad [5]$$

Under mild assumptions, event horizons in stationary spacetimes with matter satisfying the null-energy condition,

$$T_{\mu\nu}\ell^\mu\ell^\nu \geq 0 \quad \text{for all null vectors } \ell^\mu \qquad [6]$$

are smooth null hypersurfaces, analytic if the metric is analytic.

In order to develop a reasonable theory, one also needs a regularity condition for the interior of spacetime. This has to be a condition which does not exclude singularities (otherwise the Schwarzschild and Kerr black holes would be excluded), but which nevertheless guarantees a well-behaved exterior region. One such condition, assumed in all the results described below, is the existence in $\mathscr{M}$ of an asymptotically flat spacelike hypersurface $\mathscr{S}$ with compact interior. Further, either $\mathscr{S}$ has no boundary or the boundary of $\mathscr{S}$ lies on $\overline{\mathscr{E}^+ \cup \mathscr{E}^-}$. To make things precise, for any spacelike hypersurface let $g_{ij}$ be the induced metric, and let $K_{ij}$ denote its extrinsic curvature. A spacelike hypersurface $\mathscr{S}_{\text{ext}}$ diffeomorphic to $\mathbb{R}^3$ minus a ball will be called asymptotically flat if the fields $(g_{ij}, K_{ij})$ satisfy the fall-off conditions

$$|g_{ij} - \delta_{ij}| + r|\partial_\ell g_{ij}| + \cdots + r^k|\partial_{\ell_1\cdots\ell_k} g_{ij}| + r|K_{ij}| + \cdots + r^k|\partial_{\ell_1\cdots\ell_{k-1}} K_{ij}| \leq Cr^{-1} \qquad [7]$$

for some constants $C, k \geq 1$. A hypersurface $\mathscr{S}$ (with or without boundary) will be said to be asymptotically flat with compact interior if $\mathscr{S}$ is of the form $\mathscr{S}_{\text{int}} \cup \mathscr{S}_{\text{ext}}$, with $\mathscr{S}_{\text{int}}$ compact and $\mathscr{S}_{\text{ext}}$ asymptotically flat.

There exists a canonical way of constructing a conformal completion with good global properties for stationary spacetimes which are asymptotically flat in the sense of [7], and which are vacuum sufficiently far out in the asymptotic region. This conformal completion is referred to as the standard completion and will be assumed from now on.

Returning to the event horizon $\mathscr{E} = \mathscr{E}^+ \cup \mathscr{E}^-$, it is not very difficult to show that every Killing vector field $X$ is necessarily tangent to $\mathscr{E}$. Since the latter set is a null Lipschitz hypersurface, it follows that $X$ is either null or spacelike on $\mathscr{E}$. This leads to a preferred class of event horizons, called Killing horizons. By definition, a Killing horizon associated with a Killing vector $K$ is a null hypersurface which coincides with a connected component of the set

$$\mathcal{H}(K) := \{p \in \mathscr{M}\colon\ g(K,K)(p) = 0,\ K(p) \neq 0\} \qquad [8]$$

A simple example is provided by the "boost Killing vector field" $K = z\partial_t + t\partial_z$ in Minkowski spacetime: $\mathcal{H}(K)$ has four connected components,

$$\mathcal{H}_{\epsilon\delta} := \{t = \epsilon z, \delta t > 0\}, \quad \epsilon, \delta \in \{\pm 1\}$$

The closure $\bar{\mathcal{H}}$ of $\mathcal{H}$ is the set $\{|t| = |z|\}$, which is not a manifold, because of the crossing of the null hyperplanes $\{t = \pm z\}$ at $t = z = 0$. Horizons of this type are referred to as bifurcate Killing horizons, with the set $\{K(p) = 0\}$ being called the bifurcation surface of $\mathcal{H}(K)$. The bifurcate horizon structure in the Kruszkal–Szekeres–Schwarzschild spacetime can be clearly seen in **Figures 1** and **2**.

The Vishveshwara–Carter lemma shows that if a Killing vector $K$ is hypersurface orthogonal, $K^\flat \wedge dK^\flat = 0$, then the set $\mathcal{H}(K)$ defined in [8] is a union of smooth null hypersurfaces, with $K$ being tangent to the null geodesics threading $\mathcal{H}$ ("$\mathcal{H}$ is generated by $K$"), and so is indeed a Killing horizon. It has been shown by Carter that the same conclusion can be reached if the hypothesis of hypersurface orthogonality is replaced by that of existence of two linearly independent Killing vector fields.

In stationary-axisymmetric spacetimes, a Killing vector $K$ tangent to the generators of a Killing horizon $\mathcal{H}$ can be normalized so that $K = X + \omega Y$, where $X$ is the Killing vector field which asymptotes to a time translation in the asymptotic region, and $Y$ is the Killing vector field which generates rotations in the asymptotic region. The constant $\omega$ is called the angular velocity of the Killing horizon $\mathcal{H}$.

On a Killing horizon $\mathcal{H}(K)$, one necessarily has

$$\nabla^\mu (K^\nu K_\nu) = -2\kappa K^\mu \qquad [9]$$

Assuming the so-called dominant-energy condition on $T_{\mu\nu}$ (*see* Positive Energy Theorem and Other Inequalities in GR), it can be shown that $\kappa$ is constant (recall that Killing horizons are always connected in the terminology used in this article); it is called the surface gravity of $\mathcal{H}$. A Killing horizon is called degenerate when $\kappa = 0$, and nondegenerate otherwise; by an abuse of terminology, one similarly talks of degenerate black holes, etc. In Kerr spacetimes we have $\kappa = 0$ if and only if $m = a$. A fundamental theorem of Boyer shows that degenerate horizons are closed. This implies that a horizon $\mathcal{H}(K)$ such that $K$ has zeros in $\bar{\mathcal{H}}$ is nondegenerate, and is of bifurcate type, as described above. Further, a nondegenerate Killing horizon with complete geodesic generators always contains zeros of $K$ in its closure. However, it is not true that existence of a nondegenerate horizon implies that of zeros of $K$: take the Killing vector field $z\partial_t + t\partial_z$ in Minkowski spacetime from which the 2-plane $\{z = t = 0\}$ has been removed. The universal cover of that last spacetime provides a spacetime in which one cannot restore the points which have been artificially removed, without violating the manifold property.

The domain of outer communications (DOC) of a black hole spacetime is defined as

$$\langle\langle \mathscr{M} \rangle\rangle := \mathscr{M} \setminus \{\mathscr{B} \cup \mathscr{W}\} \qquad [10]$$

Thus, $\langle\langle \mathscr{M} \rangle\rangle$ is the region lying outside of the white-hole region and outside of the black hole region; it is the region which can both be seen by the outside observers and influenced by those.

The subset of $\langle\langle \mathscr{M} \rangle\rangle$ where $X$ is spacelike is called the ergoregion. In the Schwarzschild spacetime, $\omega = 0$ and the ergoregion is empty, but neither of these is true in Kerr with $a \neq 0$.

A very convenient method for visualizing the global structure of spacetimes is provided by the Carter–Penrose diagrams. An example of such a diagram is given in Figure 2.

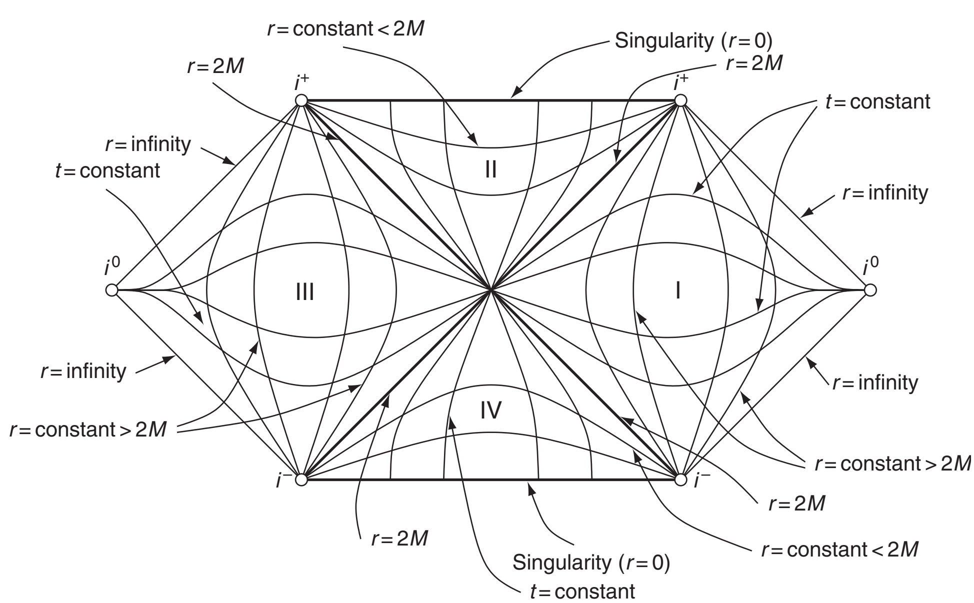



**Figure 2** The Carter–Penrose diagram for the Kruskal–Szekeres spacetime. There are actually two asymptotically flat regions, with corresponding $\mathscr{I}^\pm$ and $\mathscr{E}^\pm$ defined with respect to the second region, but not indicated on this diagram. Each point in this diagram represents a two-dimensional sphere, and coordinates are chosen so that light cones have slopes ±1. Regions are numbered as in Figure 1. (Adapted with permission from Nicolas J-P (2002) Dirac fields on asymptotically flat space-times. *Dissertationes Mathematicae* 408: 1–85.)

A corollary of the topological censorship theorem of Friedman, Schleich, and Witt is that DOCs of regular black hole spacetimes satisfying the dominant-energy condition are simply connected. This implies that connected components of event horizons in stationary spacetimes have $\mathbb{R} \times S^2$ topology.

The discussion of the concepts associated with stationary-black hole spacetimes can be concluded by summarizing the properties of the Schwarzschild and Kerr geometries: the extended Kerr spacetime with $m > a$ is a black hole spacetime with the hypersurface $\{r = r_+\}$ forming a nondegenerate, bifurcate Killing horizon generated by the vector field $X + \omega Y$ and surface gravity given by

$$\kappa = \frac{(m^2 - a^2)^{1/2}}{2m[m + (m^2 - a^2)^{1/2}]}$$

In the case $a = 0$, where the angular velocity $\omega$ vanishes, $X$ is hypersurface orthogonal and becomes the generator of $\mathcal{H}$. The bifurcation surface in this case is the totally geodesic 2-sphere, along which the four regions in **Figure 1** are joined.

## Classification of Stationary Solutions ("No-Hair Theorems")

We confine attention to the "outside region" of black holes, the DOC. (Except for the degenerate case discussed later, the "inside"(black hole) region is not stationary, so that this restriction already follows from the requirement of stationarity.) For reasons of space, we only consider vacuum solutions; there exists a similar theory for electro-vacuum black holes. (There is a somewhat less developed theory for black hole spacetimes in the presence of nonabelian gauge fields.) In connection with a collapse scenario, the vacuum condition begs the question: collapse of what? The answer is twofold: first, there are large classes of solutions of Einstein equations describing pure gravitational waves. It is believed that sufficiently strong such solutions will form black holes. (Whether or not they will do that is related to the cosmic censorship conjecture, *see* Spacetime Toplogy, Casual Structure and Singularities.) Consider, next, a dynamical situation in which matter is initially present. The conditions imposed in this section correspond then to a final state in which matter has either been radiated away to infinity, or has been swallowed by the black hole (as in the spherically symmetric Oppenheimer–Snyder collapse described above).

Based on the facts below, it is expected that the DOCs of appropriately regular, stationary, vacuum black holes are isometrically diffeomorphic to those of Kerr black holes:

1. *The rigidity theorem* (Hawking). Event horizons in regular, nondegenerate, stationary, analytic vacuum black holes are either Killing horizons for $X$, or there exists a second Killing vector in $\langle\langle \mathscr{M} \rangle\rangle$.
2. *The Killing horizons theorem* (Sudarsky–Wald). Nondegenerate stationary vacuum black holes such that the event horizon is the union of Killing horizons of $X$ are static.
3. The Schwarzschild black holes exhaust the family of static regular vacuum black holes (Israel, Bunting – Masood-ul-Alam, Chruściel).
4. The Kerr black holes satisfying

$$m^2 > a^2 \qquad [11]$$

exhaust the family of nondegenerate, stationary-axisymmetric, vacuum, connected black holes. Here $m$ is the total Arnowitt–Deser–Misner (ADM) mass, while the product $am$ is the total ADM angular momentum. (Of course, these quantities generalize the constants $a$ and $m$ appearing in the Kerr metric.) The framework for the proof has been set up by Carter, and the statement above is due to Robinson.

The above results are collectively known under the name of no-hair theorems, and they have not provided the final answer to the problem so far. There are no *a priori* reasons known for the analyticity hypothesis in the rigidity theorem. Further, degenerate horizons have been completely understood in the static case only.

Yet another key open question is that of the existence of nonconnected regular stationary-axisymmetric vacuum black holes. The following result is due to Weinstein: let $\partial\mathscr{S}_a$, $a = 1, \ldots, N$, be the connected components of $\partial\mathscr{S}$. Let $X^\flat = g_{\mu\nu} X^\mu \, dx^\nu$, where $X^\mu$ is the Killing vector field which asymptotically approaches the unit normal to $\mathscr{S}_{\text{ext}}$. Similarly, set $Y^\flat = g_{\mu\nu} Y^\mu dx^\nu$, $Y^\mu$ being the Killing vector field associated with rotations. On each $\partial\mathscr{S}_a$, there exists a constant $\omega_a$ such that the vector $X + \omega_a Y$ is tangent to the generators of the Killing horizon intersecting $\partial\mathscr{S}_a$. The constant $\omega_a$ is called the angular velocity of the associated Killing horizon. Define

$$m_a = -\frac{1}{8\pi} \int_{\partial\mathscr{S}_a} *dX^\flat \qquad [12]$$

$$L_a = -\frac{1}{4\pi} \int_{\partial\mathscr{S}_a} *dY^\flat \qquad [13]$$

Such integrals are called Komar integrals. One usually thinks of $L_a$ as the angular momentum of each connected component of the black hole. Set

$$\mu_a = m_a - 2\omega_a L_a \tag{14}$$

Weinstein shows that one necessarily has $\mu_a > 0$. The problem at hand can be reduced to a harmonic-map equation, also known as the Ernst equation, involving a singular map from $\mathbb{R}^3$ with Euclidean metric $\delta$ to the two-dimensional hyperbolic space. Let $r_a > 0$, $a = 1, \ldots, N-1$, be the distance in $\mathbb{R}^3$ along the axis between neighboring black holes as measured with respect to the (unphysical) metric $\delta$. Weinstein proved that for nondegenerate regular black holes the inequality [11] holds, and that the metric on $\langle\langle \mathscr{M} \rangle\rangle$ is determined up to isometry by the $3N-1$ parameters

$$(\mu_1, \ldots, \mu_N, L_1, \ldots, L_N, r_1, \ldots, r_{N-1}) \tag{15}$$

just described, with $r_a, \mu_a > 0$. These results by Weinstein contain the no-hair theorem of Carter and Robinson as a special case. Weinstein also shows that, for every $N \geq 2$ and for every set of parameters [15] with $\mu_a, r_a > 0$, there exists a solution of the problem at hand. It is known that for some sets of parameters [15] the solutions will have "strut singularities" between some pairs of neighboring black holes, but the existence of the "struts" for all sets of parameters as above is not known, and is one of the main open problems in our understanding of stationary-axisymmetric electro-vacuum black holes. The existence and uniqueness results of Weinstein remain valid when strut singularities are allowed in the metric at the outset, although such solutions do not fall into the category of regular black holes discussed here.

## Acknowledgments

This work was partially supported by a Polish Research Committee grant 2 P03B 073 24 and by the Erwin Schroedinger Institute, Vienna.

*See also:* Asymptotic Structure and Conformal Infinity; Black Hole Mechanics; Critical Phenomena in Gravitational Collapse; Einstein Equations: Exact Solutions; Einstein Equations: Initial Value Formulation; Geometric Flows and the Penrose Inequality; Spacetime Topology, Causal Structure and Singularities.

## Further Reading

Carter B (1973) Black hole equilibrium states. In: de Witt C and de Witt B (eds.) *Black Holes*, Proceedings of the Les Houches Summer School. New York: Gordon and Breach.

Chruściel PT (2002) Black holes. In: Friedrich H and Frauendiener J (eds.) *Proceedings of the Tübingen Workshop on the Conformal Structure of Space-Times*, Springer Lecture Notes in Physics, vol. 604, pp. 61–102 (gr-qc/0201053). Springer: Heidelberg.

Hawking SW and Ellis GFR (1973) *The Large Scale Structure of Space–Time*. Cambridge: Cambridge University Press.

Heusler M (1996) *Black Hole Uniqueness Theorems*. Cambridge: Cambridge University Press.

Heusler M (1998) Stationary black holes: uniqueness and beyond. *Living Reviews* Relativity 1 (http://www.livingreviews.org/lrr-1988-6).

Nicolas J-P (2002) Dirac fields on asymptotically flat space-times. *Dissertationes Mathematicae* 408: 1–85.

O'Neill B (1995) *The Geometry of Kerr Black Holes*. Wellesley, MA: A.K. Peters.

Volkov MS and Gal'tsov DV (1999) Gravitating non-abelian solitons and black holes with Yang–Mills fields. *Physics Reports* 319: 1–83 (hep-th/9810070).

Wald RM (1984) *General Relativity*. Chicago: University of Chicago Press.

# Stationary Phase Approximation

**J J Duistermaat**, Universiteit Utrecht, Utrecht, The Netherlands



## Introduction

An oscillatory integral is an integral of the form

$$I(\omega) = \int e^{i\omega\varphi(\theta)} a(\theta) d\theta \tag{1}$$

Here the integration is over a smooth $k$-dimensional manifold $\Theta$ which is provided with a smooth density $d\theta$. The real variable $\omega$ plays the role of a frequency variable, whereas the real-valued smooth function $\varphi$ on $\Theta$ is called the phase function. The amplitude function $a$ is assumed to be a compactly supported complex (vector-) valued smooth function on $\Theta$. The topic of this article is the asymptotic behavior of the oscillatory integral $I(\omega)$ as the frequency $\omega$ tends to infinity.

When the manifold $\Theta$ is not compact and the amplitude function is not compactly supported, then a smooth cutoff function may be used to write the integral as the sum of an integral with a compactly supported amplitude and one with an amplitude which is equal to zero in a large compact subset of $\Theta$. The

latter integral can be studied if suitable assumptions are made about the asymptotic behavior of the phase function and the amplitude at infinity, but this is not the subject of this article. The use of the exponential function with purely imaginary argument instead of the sine and the cosine is just a matter of convenience.

The first observation about oscillatory integrals in the next section is the principle of stationary phase, which states that the contributions to the integral which are not rapidly decreasing as $\omega \to \infty$ only come from the stationary points of $\varphi$, the points $\theta \in \Theta$ where the total derivative $\mathrm{d}\varphi(\theta)$ of $\varphi$ is equal to zero. This principle is closely related to the observation that a superposition of waves is maximal at points where the waves are in phase, an observation which goes back to Huygens (1690).

Assume that $\theta_0$ is a nondegenerate stationary point of $\varphi$. That is, $\mathrm{d}\varphi(\theta_0) = 0$ and the Hessian $\mathrm{D}^2\varphi(\theta_0)$ of $\varphi$ at $\theta_0$ is nondegenerate. Then $\theta_0$ is an isolated stationary point of $\varphi$, and the contribution to $I(\omega)$ of a neighborhood of $\theta_0$ has an asymptotic expansion of the form

$$I(\omega) \sim \mathrm{e}^{\mathrm{i}\omega\varphi(\theta_0)} \sum_{r=0}^{\infty} c_r\, \omega^{-k/2-r}, \quad r \to \infty$$

Here the leading coefficient $c_0$ is the product of $a(\theta_0)$ with a nonzero constant which only depends on $\mathrm{D}^2\varphi(\theta_0)$ and the density $\mathrm{d}\theta$ at $\theta_0$. For increasing $r$ the coefficients $c_r$ depend on the derivatives of $\varphi$ and $a$ at $\theta_0$ of increasing order (see the section "The method of stationary phase").

Usually, even if all the objects are analytic in a neighborhood of $\theta_0$, the asymptotic power series does not converge. However, there are exceptional cases where the stationary phase approximation is exact. Assume, for instance, that $\Theta$ is a compact manifold provided with a symplectic form $\sigma, \varphi$ is the Hamiltonian function of a Hamiltonian circle action on $\Theta$ with isolated fixed points, and $a(\theta)\,\mathrm{d}\theta = \sigma^k/k!$. Then the stationary points of $\varphi$ are the fixed points of the circle action, each stationary point of $\varphi$ is nondegenerate and $I(\omega)$ is equal to the sum over the finitely many stationary points of only the leading terms of the asymptotic expansions at the stationary points. This Duistermaat–Heckman formula is a consequence of a more general localization formula in equivariant cohomology (see the section "Exact stationary phase").

For the purpose of applications, but also in the analysis of oscillatory integrals, it is worthwhile to allow complex-valued phase functions, but with a local minimum for the imaginary part at the stationary point $\theta_0$ of the real parts. That is, the real part of the exponent $\mathrm{i}\omega\varphi(\theta)$ has a local maximum at $\theta_0$. An extreme case occurs when $\varphi(\theta) = \mathrm{i}\psi(\theta)$ for a real-valued function $\psi$ which has a nondegenerate local minimum at $\theta_0$, in which case the integrand is a sharply peaking Gaussian density at $\theta_0$. When $\varphi$ and $a$ are analytic near $\theta_0$, then the method of steepest descent consists of deforming the path of integration in the complex domain in such a way that the integrand becomes such a sharply peaking Gaussian density. During the deformation, the integral does not change because of Cauchy's integral theorem.

An important extension of the theory occurs if the real-valued phase function and the amplitude are allowed to depend smoothly on additional parameters $x$, which vary in an $n$-dimensional smooth manifold $M$. The amplitude is also allowed to depend on $\omega$, with an asymptotic expansion of the form

$$a(x,\theta,\omega) \sim \sum_{r=0}^{\infty} a_r(x,\theta)\omega^{m+(k/2)-r} \quad \text{as } \omega \to \infty \qquad [2]$$

The expansion is supposed to be locally uniformly in $(x,\theta)$ and to allow termwise differentiations of any order with respect to the variables $(x,\theta)$. Then the integral

$$I(x,\omega) = \int \mathrm{e}^{\mathrm{i}\omega\varphi(x,\theta)} a(x,\theta,\omega)\,\mathrm{d}\theta$$

is called an oscillatory integral of order $m$. Here the function $x \mapsto I(x,\omega)$ is viewed as a continuous superposition of the $\theta$-dependent family of oscillatory functions $x \mapsto \mathrm{e}^{\mathrm{i}\omega\varphi(x,\theta)}a(x,\theta)$.

The example which formed the point of departure of Airy (1838) is that $\mathrm{e}^{\mathrm{i}\omega\varphi(x,\theta)}a(x,\theta,\omega)$ is the wave which arrives at the points $x$ in spacetime which is sent out by a point $\theta$ on a reflecting mirror. That is, at $x$ one collects (= integrates over $\Theta$) all the waves sent out by the various points $\theta$ of the mirror $\Theta$. The main point of the theory, however, is that in great generality the solutions of linear partial differential equations, such as classical wave equations or quantum mechanical Schrödinger equations, can be represented, as functions of $x$, as oscillatory integrals. This construction has led to decisive progress in the general theory of linear partial differential equations with smoothly varying coefficients.

According to the principle of stationary phase, the main asymptotic contributions to the integral come from the points $\theta$ such that $\partial\varphi(x,\theta)/\partial\theta = 0$. The phase function $\varphi \in \mathrm{C}^\infty(M \times \Theta)$ is called nondegenerate if the $(n+k) \times k$-matrix

$$\frac{\partial^2\varphi(x,\theta)}{\partial(x,\theta)\partial\theta} \text{ has rank } k \text{ when } \frac{\partial\varphi(x,\theta)}{\partial\theta} = 0 \qquad [3]$$

This is the natural condition to ensure that the set

$$S_\varphi := \left\{ (x, \theta) \in M \times \Theta \middle| \frac{\partial \varphi(x, \theta)}{\partial \theta} = 0 \right\}$$

is a smooth $n$-dimensional submanifold of $M \times \Theta$. The condition [3], moreover, implies that the mapping

$$\iota_\varphi : S_\varphi \ni (x, \theta) \mapsto \left( x, \frac{\partial \varphi(x, \theta)}{\partial x} \right) \in \mathrm{T}^* M$$

is a smooth immersion from $S_\varphi$ into the cotangent bundle $\mathrm{T}^*M$ of $M$. Note that $\xi = \partial\varphi(x,\theta)/\partial x$ is coordinate invariantly defined as a linear form on the tangent space $\mathrm{T}_xM$ of $M$ at the point $x$. That is, $\xi \in (\mathrm{T}_xM)^*$ = the dual space of $\mathrm{T}_xM$, and $(\mathrm{T}_xM)^*$ is the fiber of $\mathrm{T}^*M$ over $x$. In classical mechanics, $\mathrm{T}^*M$ is the phase space of the position space $M$, and a linear form $\xi$ on $\mathrm{T}_xM$ is called a momentum vector at the position $x$. If $\sigma$ denotes the canonical symplectic form on $\mathrm{T}^*M$, then $\iota_\varphi^*\sigma = 0$. The immersion $\iota_\varphi$ locally embeds $S_\varphi$ onto a smooth $n$-dimensional submanifold $\Lambda_\varphi$ of $M$, which is a Lagrangian manifold in $\mathrm{T}^*M$, which by definition means that $\iota_\varphi^*\sigma = 0$.

Oscillatory integrals with very different phase functions and amplitudes can define the same $\omega$-dependent functions on $M$. The theory of Hörmander (1971, section 3.1) says that the germs of the Lagrangian manifolds $\Lambda_\varphi$ and $\Lambda_\psi$ are the same if and only if $\varphi$ and $\psi$ define the same class of oscillatory integrals. Moreover, every Lagrangian submanifold $\Lambda$ of $\mathrm{T}^*M$ is locally of the form $\Lambda_\varphi$ for some nondegenerate phase function $\varphi$. In this way, the mapping $\varphi \mapsto \Lambda_\varphi$ defines a bijection between the set of equivalence classes of germs of nondegenerate phase functions and the set of germs of Lagrangian submanifolds of $\mathrm{T}^*M$. Let $\Lambda$ be an immersed Lagrangian submanifold of $\mathrm{T}^*M$. A global oscillatory integral of order $m$ on $M$, defined by $\Lambda$, is a locally finite sum $u(x,\omega)$ of oscillatory integrals of order $m$ with nondegenerate phase functions $\varphi$ such that $\Lambda_\varphi \subset \Lambda$. The leading terms of the amplitudes correspond to a section $s$ of a canonically defined complex line bundle $\lambda$ over $\Lambda$, which is called the principal symbol of $u$ (see the section "The principal symbol on the Lagrangian manifold").

If $P$ is a linear partial differential operator, such as the wave operators, in which the coefficients may depend in a smooth way on $x$ and in a polynomial way on $\omega$, then the condition that $Pu$ is asymptotically small implies that $p = 0$ on $\Lambda$, in which $p$ is a smooth function on $\mathrm{T}^*M$, called the principal symbol of $P$. Because $\Lambda$ is a Lagrangian manifold, the equation $p = 0$ implies that $\Lambda$ is invariant under the flow of the Hamiltonian system with Hamilton function equal to $p$. Furthermore, the principal symbol $s$ of $u$ satisfies a homogeneous first-order ordinary differential equation along the solution curves of the Hamiltonian system. Conversely, these properties can be used to construct global oscillatory integrals $u$ which asymptotically satisfy $Pu = 0$ and have prescribed initial values. This theory, due to Maslov (1972), may be viewed as a far reaching generalization of the WKB method.

Let $\pi : \mathrm{T}^*M \to M : (x, \xi) \mapsto x$ denote the canonical projection from $\mathrm{T}^*M$ onto $M$. The projections into $M$ of the solution curves in a Lagrangian submanifold $\Lambda$ of $\mathrm{T}^*M$, of a Hamiltonian system which leaves $\Lambda$ invariant, are the ray bundles of geometrical optics. If $\Lambda$ is not transversal to the fiber of $\mathrm{T}^*M$ at $(x, \xi)$, then the ray bundle exhibits a caustic at the point $x \in M$, and the oscillatory integral is asymptotically of larger order than $\omega^m$ near $x$. Applying the theory of unfoldings of singularities to the phase function, one can determine the structurally stable caustics and obtain normal forms of the oscillatory integrals in the structurally stable cases (see the section "Caustics").

If we also integrate over the frequency variable $\omega$, then we obtain the Fourier integral distributions $u$ of Hörmander (1971, sections 1.2 and 3.2). In this case the corresponding Lagrangian manifold is conic in the sense that if $(x, \xi) \in \Lambda$, then $(x, \tau\xi) \in \Lambda$ for every $\tau > 0$. The wave front set of $u$, which is the microlocal singular locus of the distribution $u$, is contained in $\Lambda$, with equality if the principal symbol of $u$ is not equal to zero at the corresponding stationary points of the phase function. Fourier integral operators are defined as the linear operators acting on distributions, of which the distribution kernels are Fourier integral distributions. Under a suitable transversality condition for the Lagrangian manifolds of the distribution kernels, the composition of two Fourier integral operators is again a Fourier integral operator, and the principal symbol of the composition is a product of the principal symbols. The proof is an application of the method of stationary phase. Fourier integral operators are a very powerful tool in the analysis of linear partial differential operators with smoothly varying coefficients (see Hörmander (1985)).

## The Principle of Stationary Phase

The principle of stationary phase says that if the phase function $\varphi$ has no stationary points in the support of the amplitude function $a$, then the

oscillatory integral [1] is rapidly decreasing, in the sense that for every $N$ we have $I(\omega) = O(\omega^{-N})$ as $\omega \to \infty$. For the proof, one introduces a vector field $v$ on $\Theta$ such that $v\varphi = 1$ on a neighborhood of the support of $a$. Then $e^{i\omega\varphi} = (i\omega)^{-1} v(e^{i\omega\varphi})$, and an integration by parts in [1] yields that

$$I(\omega) = \frac{1}{i\omega} \int e^{i\omega\varphi(\theta)} ({}^{t}va)(\theta)\, d\theta$$

where ${}^{t}v$ denotes the transposed of the linear partial differential operator $v$. Iterating this, the rapid decrease of $I(\omega)$ follows.

Using cutoff functions, $I(\omega)$ is, modulo a rapidly decreasing function, equal to an oscillatory integral with phase function $\varphi$ and an amplitude which has support in an arbitrarily small neighborhood of the set of stationary points of $\varphi$. In this sense, the contributions to the integral which are not rapidly decreasing come only from the stationary points of $\varphi$.

## The Method of Stationary Phase

Assume that $\theta_0$ is a nondegenerate stationary point of $\varphi$. Then $\theta_0$ is an isolated stationary point of $\varphi$. Using local coordinates near $\theta_0$, the contribution to [1] from the neighborhood of $\theta_0$ can be written as an oscillatory integral with $\Theta = \boldsymbol{R}^k$ and a pase function $\varphi$ which has a nondegenerate stationary point at 0. Write $Q = D^2\varphi(0)$. According to the Morse lemma, there is smooth substitution of variables $\theta = T(y)$ such that $T(0) = 0, DT(0) = I$, and $\varphi(T(\mu)) = \varphi(0) + \langle Qy, y\rangle/2$ for all $y$ in a neighborhood of 0 in $\boldsymbol{R}^k$. Applying this substitution of variables to [1] we obtain

$$I(\omega) = e^{i\omega\varphi(0)} \int_{\boldsymbol{R}^k} e^{i\omega\langle Qy,y\rangle/2} b(y)\, dy$$

where $b$ is a compactly supported smooth function on $\boldsymbol{R}^k$ with $b(0) = a(0)$. Now the Fourier transform of the function $y \mapsto e^{i\omega\langle Qy,y\rangle/2}$ is equal to the function

$$\eta \mapsto \left(\det\left(\frac{\omega}{2\pi i} Q\right)\right)^{-1/2} e^{i\omega^{-1}\langle Q^{-1}\eta,\eta\rangle/2} \qquad [4]$$

Both in the definition of the square root of the determinant and in the proof one uses the analytic continuation to the domain of complex-valued symmetric bilinear forms $Q$ for which the imaginary part of $Q$ is positive definite. For purely imaginary $Q$ we have the familiar formula for the Fourier transform of a Gaussian density (see Hörmander (1990, theorem 7.6.1)). The Taylor expansion of the exponential factor in [4] then yields that

$$\begin{aligned}\int_{\boldsymbol{R}^k} & e^{i\omega\langle Qy,y\rangle/2} b(y)\, dy \\ & \sim \left(\det\left(\frac{\omega}{2\pi i} Q\right)\right)^{-1/2} \sum_{r=0}^{\infty} (-2i\omega)^{-r} \frac{1}{r!} \\ & \times \left\langle Q^{-1} \frac{\partial}{\partial y}, \frac{\partial}{\partial y} \right\rangle^{r} b(y)\bigg|_{y=0}\end{aligned}$$

as $\omega \to \infty$ (see Hörmander (1990, lemma 7.7.3)).

It is important for the applications that, if the phase function and amplitude depend smoothly on parameters, all the constructions can be made to depend smoothly on the parameters.

## Exact Stationary Phase

Suppose that we have given an action of a Lie group $G$ on the manifold $\Theta$. Let $\mathfrak{g}$ denote the Lie algebra of $G$. For any $g \in G$ and $X \in \mathfrak{g}$ the corresponding diffeomorphism of $\Theta$ and vector field on $\Theta$ is denoted by $g_\Theta$ and $X_\Theta$, respectively. If $\Omega(\Theta)$ denotes the algebra of smooth differential forms on $\Theta$, then we consider the algebra $S\mathfrak{g}^* \otimes \Omega(\Theta)$ of all $\Omega(\Theta)$-valued polynomials on $\mathfrak{g}$, where $S\mathfrak{g}^*$ denotes the algebra of all polynomial functions on $\mathfrak{g}$. On $S\mathfrak{g}^* \otimes \Omega(\Theta)$ we have the action of $g \in G$ which sends $\alpha$ to $X \mapsto g_\Theta^*(\alpha(\mathrm{Ad}\, g\, X))$. Let $A = (S\mathfrak{g}^* \otimes \Omega(\Theta))^G$ denote the subalgebra of all $G$-invariant elements of $S\mathfrak{g}^* \otimes \Omega(\Theta)$. The equivariant exterior derivative $D$ is defined by

$$(D\alpha)(X) = d(\alpha(X)) - i_{X_\Theta}(\alpha(X))$$

If $\alpha$ is homogeneous as a differential form of degree $p$ and homogeneous as a polynomial on $\mathfrak{g}$ of degree $q$, then $r = p + 2q$ is called the total degree of $\alpha$. Let $A^r$ denote the space of sums of such $\alpha \in A$ of total degree $r$. Then $D_r = D\colon A^r \to A^{r+1}$ and $D_r \circ D_{r-1} = 0$. The space $H_G^r(\Theta) := \ker D_r / \mathrm{Im}\, D_{r-1}$ is called the equivariant cohomology in degree $r$, in the model of Cartan (1950).

Assume that $\Theta$ is compact and oriented, and that the action of $G$ preserves the orientation. If $\alpha \in A$, then we denote by $\alpha(X)^{[k]}$ the volume part of the differential form $\alpha(X)$, and

$$\left(\int \alpha\right)(X) := \int_\Theta \alpha(X)^{[k]}, \quad X \in \mathfrak{g}$$

defines an Ad $G$-invariant function $\int\alpha$ on $\mathfrak{g}$. Now $\alpha = D\beta$ implies that $\alpha(X)^{[k]}$ id equal to the exterior derivative of $\beta(X)^{[k-1]}$, and therefore $\int\alpha = 0$, in view of Stokes' theorem. It follows that integration over $\Theta$ yields a linear mapping $\int$ from $H_G(\Theta)$ to $(S\mathfrak{g}^*)^{\mathrm{Ad}\, G}$, which is called integration in equivariant cohomology.

Now assume that also the Lie group $G$ is compact, and let $X \in \mathfrak{g}$. Then the zero-set $Z_X$ of $X_\Theta$ in $\Theta$ has finitely many connected components $F$, each of which is a smooth and compact submanifold of $\Theta$. In general, the $F$'s can have different dimensions. The linearization L$X$ of the vector field $X_\Theta$ along $F$ acts linearly on the normal bundle N$F$ of $F$. If $\Omega$ is the curvature form of N$F$, then

$$\varepsilon(X) := \det_C\left[\frac{\mathrm{i}}{2\pi}(\mathrm{L}X - \Omega)\right]$$

is called the equivariant Euler form of N$F$. $\varepsilon(X)$ is an invertible element in the algebra $\Omega^{\mathrm{even}}(F)$. The localization formula of Berline–Vergne (1982) and Atiyah–Bott (1984) now says that if $D\alpha = 0$ then

$$\left(\int\alpha\right)(X) = \sum_F \int_F \left(\mathrm{i}_F^*\alpha(X)/\varepsilon(X)\right)^{[\dim F]}$$

Assume that $\sigma$ is a symplectic form on $\Theta$, which implies that $k = 2l$ is even. Furthermore, assume that the infinitesimal action of $\mathfrak{g}$ on $\Theta$ is Hamiltonian, which means that there exists a $G$-equivariant smooth mapping $\mu : \Theta \to \mathfrak{g}^*$, called the momentum mapping, such that $\mathrm{i}_{X_\Theta}\sigma = -\mathrm{d}(\mu(X))$ for every $X \in \mathfrak{g}$. Here $\mu$ is viewed as an element of $(\mathfrak{g}^* \otimes \Omega^0(\Theta))^G \subset A$. Then $\widehat{\sigma}(X) := \sigma - \mu(X)$ defines an element $\widehat{\sigma} \in A$ such that $D\widehat{\sigma} = 0$. In turn, this implies that the form

$$\beta(X) := \mathrm{e}^{-\mathrm{i}\omega\widehat{\sigma}(X)} = \mathrm{e}^{\mathrm{i}\omega\mu(X)} \sum_{r=0}^{l} (-\mathrm{i}\omega\sigma)^r / r!$$

is equivariantly closed, and the localization formula of equivariant cohomology applied to this case yields the Duistermaat–Heckman (1982, 1983) formula. Because $\beta(X)^{[k]} = \mathrm{e}^{\mathrm{i}\omega\mu(X)}(-\mathrm{i}\omega\sigma)^l/l!$, its integral over $\Theta$ is an oscillatory integral with phase function $\mu(X)$. The stationary points of $\mu(X)$ are the zeros of $X_\Theta$ and the stationary points of $\mu(X)$ are nondegenerate if and only if the zeros of $X_\Theta$ are isolated. It follows that in this case the oscillatory integral is equal to the leading term in the stationary-phase approximation.

## The Principal Symbol on the Lagrangian Manifold

Let $u(x, \omega)$ be a global oscillatory integral of order $m$ defined by $\Lambda$, and let $(x_0, \xi_0) \in \Lambda$. One way to define the principal symbol of $u$ at $(x_0, \xi_0) \in \Lambda$ is to test $u$ with an oscillatory function of the form $\mathrm{e}^{-\mathrm{i}\omega\psi(x)}b(x)$, in which $\mathrm{d}\psi(x_0) = \xi_0$, the support of $b$ is contained in a small neighborhood of $x_0$, and $b(x_0) = 1$. If $u$ is locally represented by the phase function $\varphi$ and amplitude $a$, and $(x_0, \xi_0) = \iota_\varphi(x_0, \theta_0)$, then the phase function $\varphi(x, \theta) - \psi(x)$ in the oscillatory integral

$$\langle u, \mathrm{e}^{-\mathrm{i}\psi} b\rangle = \int_M \int_\Theta \mathrm{e}^{\mathrm{i}(\varphi(x,\theta) - \psi(x))} a(x, \theta) b(x)\, \mathrm{d}\theta\, \mathrm{d}x$$

has a stationary point at $(x_0, \theta_0)$, which means that $\Lambda$ and $\mathrm{d}\psi$ intersect at $(x_0, \xi_0)$. Here the 1-form $\mathrm{d}\psi$ on $M$, which is a section of $\pi : \mathrm{T}^*M \to M$, is viewed as a submanifold of $\mathrm{T}^*M$. Locally the Lagrangian submanifolds of $\mathrm{T}^*M$ which are transversal to the fibers of $\pi : \mathrm{T}^*M \to M$ are precisely the manifolds of the form $\mathrm{d}\psi$. The stationary point of $\varphi - \psi$ is nondegenerate if and only if $L := \mathrm{T}_{(x_0,\xi_0)}\Lambda$ and $L_\psi := \mathrm{T}_{(x_0,\xi_0)}(\mathrm{d}\psi)$ are transversal. In this case, the method of stationary phase can be applied in order to obtain an asymptotic expansion in terms of powers of $\omega$. The coefficient of the leading term of order $\omega^m$ depends only on the Lagrangian plane $L_\psi$, which is transversal to both $L$ and the tangent space of the fiber of $\mathrm{T}^*M$, and not on the other data of $\psi$ and $b$. If $\mathcal{L}$ denotes the set of all Lagrangian planes in $\mathrm{T}_{(x_0,\xi_0)}(\mathrm{T}^*M)$ which are transversal to both $L$ and the fiber, then the complex-valued functions on $\mathcal{L}$ which arise in this way form a one-dimensional complex vector space $L_{(x_0,\xi_0)}$. The $L_{(x_0,\xi_0)}$ for $(x_0, \xi_0) \in \Lambda$ form a complex line bundle $\lambda$ over $\Lambda$ which is canonically isomorphic to the tensor product of the line bundle of half-densities and the Maslov line bundle, a line bundle with structure group $\mathbf{Z}/4\mathbf{Z}$ (see Duistermaat (1974, section 1.2)). In this way, the principal symbol $s$ of $u$ can be viewed as a section of the line bundle $\lambda$ over $\Lambda$.

## Caustics

Let $(x_0, \xi_0)$ be a point in the Lagrangian submanifold $\Lambda$ of $\mathrm{T}^*M$. The restriction to $\Lambda$ of the projection $\pi : \mathrm{T}^*M \to M$ is a diffeomorphism from an open neighborhood of $(x_0, \xi_0)$ in $\Lambda$ onto an open neighborhood of $x_0$ in $M$, if and only if $\Lambda$ is transversal to the fiber of $\mathrm{T}^*M$ at $(x_0, \xi_0)$. If $\Lambda = \Lambda_\varphi$ for a nondegenerate phase function $\varphi, (x_0, \xi_0) \in S_\varphi$ and $(x_0, \xi_0) = \iota_\varphi(x_0, \theta_0)$, then this condition is in turn equivalent to the condition that $\theta_0$ is a nondegenerate stationary point of $\theta \mapsto \varphi(x_0, \theta)$. An application of the method of stationary phase shows that in this case the oscillatory integral is equal to a progressing wave of the form $\mathrm{e}^{\mathrm{i}\omega\psi(x)}b(x, \omega)$. Here $\psi(x) = \varphi(x, \theta(x))$, where $\theta(x)$ is the stationary point of $\theta \mapsto \varphi(x, \theta)$, and $b(x, \omega)$ has an asymptotic expansion as in [2] with $k = 0$.

If $\theta_0$ is a degenerate stationary point of $\theta \mapsto \varphi(x_0, \theta)$ and $a_0(x_0, \theta_0) \neq 0$, then the oscillatory integral is not of order $O(\omega^m)$. That is, it is of larger order than at points where we have a nondegenerate

stationary point. For this reason, the points $(x_0, \xi_0)$ at which $\Lambda$ is not transversal to the fibers of $\pi : \mathrm{T}^*M \to M$ are called the caustic points of $\Lambda$. Their projections $x_0 \in M$ form the caustic set in $M$.

In the theory of unfoldings of singularities, the germs of the families of functions $x \mapsto (\theta \mapsto \varphi(x, \theta))$ and $y \mapsto (\mu \mapsto \psi(y, \mu))$ are called equivalent if there exists a germ of a diffeomorphism of the form $H : (x, \theta) \mapsto (y(x), \mu(x, \theta))$ and a smooth function $\chi(x)$ such that $\psi(y(x), \mu(x, \theta)) = \varphi(x, \theta) + \chi(x)$. If $J(y, \omega)$ is an oscillatory integral with phase function $\psi$, integration variable $\mu$ and parameter $y$, then the substitution of variables $\mu = \mu(x, \theta)$ in the integral, followed by the substitution of variables $y = y(x)$ in the parameters, yields that $J(y, \omega) = \mathrm{e}^{\mathrm{i}\omega\chi(x)} I(x, \omega)$, in which $I(x, \omega)$ is an oscillatory integral with phase function $\varphi$ and an amplitude function of the same order as the amplitude function of $J$. The germ $\varphi$ is called stable if every nearby germ $\psi$ is equivalent to $\varphi$. The Morse lemma with parameters implies that this is the case if $\theta \mapsto \varphi(x_0, \theta)$ has a nondegenerate stationary point at $\theta_0$. However, the theory of unfoldings of singularities of Thom and Mather shows that there are many stable germs with degenerate critical points. Moreover, in dimension $n \leq 5$ the generic germ is stable, and is equivalent to a germ in a finite list of normal forms.

The simplest example of a normal form with degenerate critical points is $\varphi(x, \theta) = \theta^3 + x_1\theta$. Here we have taken $k = 1$, but still allowed an arbitrary dimension $n \geq 1$ of $M$. In this normal form, the stationary points correspond to $3\theta^2 + x_1 = 0$, which is a manifold which over the $x$-space folds over at $x_1 = 0$. The stationary point is degenerate if and only if $6\theta = 0$, hence $x_1 = 0$, which means that $x_1 = 0$ is the caustic set. If the amplitude is equal to 1, then the oscillatory integral is equal to $\omega^{-1/3}\,\mathrm{Ai}(\omega^{2/3} x_1)$, in which $\mathrm{Ai}(z)$ denotes the Airy function. If the amplitude is nonzero at a degenerate critical point, then the oscillatory integral near the corresponding caustic point is asymptotically of the same order as $\omega^{-1/3}\,\mathrm{Ai}(\omega^{2/3} x_1)$, which implies that the oscillatory integral is a factor $\omega^{1/6}$ larger at these caustic points than at the points away from the caustic set. In Airy (1838), where the Airy function was introduced, Airy considered light in a neighborhood of a caustic as an oscillatory integral. Then, under suitable genericity conditions, he brought the phase function into the normal form $\theta^3 + x_1\theta$. Even for stable normal forms in low dimensions, the interference patterns near the caustic points can be very intricate (see, e.g., Berry *et al.* (1979)). A survey of the application of the theory of unfoldings to caustics in oscillatory integrals can be found in Duistermaat (1974).

*See also:* Equivariant Cohomology and the Cartan Model; Feynman Path Integrals; Functional Integration in Quantum Physics; Hamiltonian Group Actions; *h*-Pseudodifferential Operators and Applications; Multiscale Approaches; Normal Forms and Semiclassical Approximation; Optical Caustics; Path Integrals in Noncommutative Geometry; Perturbation Theory and its Techniques; Schrödinger Operators; Singularity and Bifurcation Theory; Wave Equations and Diffraction.

## Further Reading


Airy GB (1838) On the intensity of light in a neighborhood of a caustic. *Transactions of the Cambridge Philosophical Society* 6: 379–403.

Atiyah MF and Bott R (1984) The moment map and equivariant cohomology. *Topology* 23: 1–28.

Berline N and Vergne M (1982) Classes caractéristiques équivariante. Formules de localisation en cohomologie équivariante. *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences, Paris* 295: 539–541.

Berry MV, Nye JF, and Wright FJ (1979) The elliptic umbilic diffraction catastrophe. *Philosophical Transactions of the Royal Society of London* A291: 453–484.

Cartan H (1950) Notion d'algèbre différentielle; applications ou opère un groupe de Lie, and: La transgression dans un groupe de Lie et dans un fibré principal. In: *Colloque de Topologie*, pp. 15–27, 57–71 Bruxelles: C.B.R.M.

Duistermaat JJ (1974) Oscillatory integrals, Lagrange immersions and unfoldings of singularities. *Communications in Pure and Applied Mathematics* 27: 207–281.

Duistermaat JJ and Heckman GJ (1982) On the variation in the cohomology of the symplectic form of the reduced phase space. *Inventiones Mathematical* 69: 259–268.

Duistermaat JJ and Heckman GJ (1983) On the variation in the cohomology of the symplectic form of the reduced phase space. *Inventiones Mathematical* 72: 153–158.

Hörmander L (1971) Fourier integral operators I. *Acta Mathematica* 127: 79–183.

Hörmander L (1983, 1990) *The Analysis of Linear Partial Differential Operators I*. Berlin: Springer.

Hörmander L (1985) *The Analysis of Linear Partial Differential Operators IV*. Berlin: Springer.

Huygens C (1690) *Traité de la Lumière*, Leyden: Van der Aa. (English translation: Treatise on Light, 1690, reprint by Dover Publications, New York, 1962.)

Maslov VP (1965) *Perturbation Theory and Asymptotic Methods* (In Russian), Moscow: Moskov. Gos. University. (French translation by Lascoux J and Sénéor R (1972) Paris: Dunod.)

# Statistical Mechanics and Combinatorial Problems

**R Zecchina**, International Centre for Theoretical Physics (ICTP), Trieste, Italy



## Introduction

Equilibrium statistical mechanics and combinatorial optimization – which is viewed here as a branch of discrete mathematics and theoretical computer science – have common roots. Phase transition are mathematical phenomena which are not limited to physical systems but are typical of many combinatorial problems, one famous example being the percolation transition in random graphs. Similarly, the understanding of relevant physical problems, such as three-dimensional lattice statistics or two-dimensional quantum statistical mechanics problems, is strictly related to the question of purely combinatorial origin of solving counting problems over nonplanar lattices. Most of the tools and concepts which have allowed to solve problems in one field have a natural counterpart in the other. While the possibility of solving exactly physical models is always connected to the presence of some algebraic properties which guarantee integrability, in the combinatorial approach the emphasis is more on algorithms that can be applied to problem instances in which the symmetries behind intergrability might be absent. Also at the level of out-of-equilibrium phenomena, there exists a deep connection between physics and combinatorics: just like physical processes, local algorithms have to deal with an exponentially large set of possible configurations and their out-of-equilibrium analysis constitutes a theory of how problems are actually solved.

Computational complexity theory deals with classifying problems in terms of the computational resources, typically time, required for their solution. What can be measured (or computed) is the time that a particular algorithm uses to solve the problem. This time in turn depends on the implementation of the algorithm as well as on the computer the program is running on. The theory of computational complexity provides us with a notion of complexity that is largely independent of implementational details and the computer at hand. This is not surprising, since it is related to a highly nontrivial question, that is: what do we mean by saying that a combinatorial problem is solvable?

Problems which can be solved in polynomial time are considered to be tractable and compose the so-called polynomial (P) class. The harder problems are grouped in a larger class called NP, where NP stands for "nondeterministic polynomial time." These problems are such that a potential solution can be checked rapidly in polynomial time, while finding a solution may require exponential time in the worst case. In turn, the hardest problems in NP belong to a subclass called NP-complete: an efficient algorithm for solving one NP-complete problem could be easily modified to effectively solve any problem in NP. By now, a huge number of NP-complete problems has been identified, and the lack of such an algorithm corroborates the widespread conjecture $P \neq NP$, that is, that no such algorithm exists. However, NP-complete problems are not always hard: when their resolution complexity is measured with respect to some underlying probability distribution of problem instances, NP-complete problems are often easy to solve on average. To deepen the understanding of the average-case complexity (and of the huge variability of running times observed in numerical experiments), computer scientists, mathematicians, and physicists have focused their attention on the study of random instances of hard combinatorial problems, seeking for a link between the onset of exponential-time complexity and some intrinsic (i.e., algorithm independent) properties of the randomized NP-complete problems. These types of questions have merged combinatorial optimization with statistical physics of disordered systems.

Computational complexity theory can also be formulated for counting problems: similarly to optimization problems, equivalence classes can be defined which separate polynomially solvable counting problems with the hard ones – the so-called #P and #P-complete problems. Complexity theory for counting problems makes the connections with statistical mechanics even more direct in that counting solutions is nothing but a computation of a partition function.

Two simple theorems by Jerrum and Sinclair (1989) (which can be easily extended to many combinatorial problems) can help in clarifying these connections.

The first theorem tells us that any randomized algorithm (e.g., Monte Carlo) for approximating the partition function of a generic spin glass model – the so-called spin glass problem – could be used to solve all the other NP combinatorial problems. The second theorem tells us that an algorithm for evaluating exactly the partition function of the ferromagnetic Ising model over a general graph would again solve any other problem in the class #P, which, as mentioned above, is the generalization of

the NP class to counting problems and obviously contains the class NP as a particular case.

Let us consider the following sightly simplified definition of the Ising and the spin glass problems.

**Problem instance** A symmetric matrix $J_{ij}$ with entries in $\{-1, 0, 1\}$ and an inverse temperature $\beta$.

**Output** The partition function $Z = \sum_{\{\sigma_i\}} 2^{-\beta H(\boldsymbol{\sigma})}$, where $H(\boldsymbol{\sigma}) = -\sum_{i,j} J_{ij}\sigma_i\sigma_j$ with $\sigma_i = \pm 1$.

Moreover, let us define the fully polynomial randomized approximation scheme (FPRAS) for counting and decision problems. A FPRAS for a function *f* from problem instances to real numbers is a probabilistic algorithm that in polynomial time in the problem size *n* and in the relative error $\epsilon \in [0, 1]$, outputs with high probability a number which approximates $f(n)$ within a ratio $1 + \epsilon$. Given the above definitions, the theorems can be stated as follows:

**Theorem 1** *There can be no FPRAS for the spin glass problem unless* P = NP, *that is*, *all problems in* NP *turn out to be solvable in polynomial time*.

**Theorem 2** *The Ising problem is* #P-*complete even when the matrix* $J_{ij}$ *is non-negative*, *that is*, *an algorithm which outputs in polynomial time the exact Ising partition function for an arbitrary graph could be used to solve any other counting problem in* #P.

The above theorems hold for arbitrary graphs, in particular for those graph or lattice realizations which are particularly hard to analyze, the so-called worst cases. There exist no similar proofs of computational hardness for more restricted and realistic structures, such as, for instance, three-dimensional regular lattices for the Ising problem or finite connectivity random graphs for spin glasses.

As a final introductory remark, it is worth mentioning that the connections between worst-case complexity and the average case one is the building block of modern cryptography and communication theory. On the one hand, the so-called RSA cryptosystem is based on factoring large integers, a problem which is believed to be hard on average while it is not known to be so in the worst case. On the other hand, alternative cryptographic systems have been proposed which rely on a worst-case/average-case equivalence (see, e.g., the theorem of Ajtai (1996) concerning some hidden vector problems in high-dimensional lattices.)

As far as communication theory is concerned, average-case complexity is indeed crucial: while Shannon's theorem (1948) provides a very general result stating that many optimal codes do exist (in fact, random codes are optimal), the decoding problem is in general NP-complete and therefore potentially intractable. However, since the choice of the coding scheme is part of the design, what matters are the average-case behavior of the decoding algorithm (and its large deviations) and very efficient codes which can solve on average the decoding problem close to Shannon's bounds are known.

In what follows, we will limit the discussion to two basic examples of combinatorial and counting problems which are representative and central to both computer science and statistical physics.

## Constraint Satisfaction Problems

Combinatorial problems are usually written as constraint satisfaction problems (CSPs): *n* discrete variables are given which have to satisfy *m* constraints, all at the same time. Each constraint can take different forms depending on the problem under study: famous examples are the *K*-satisfiability (*K*-SAT) problem in which constraints are an "OR" function of *K* variables in the ensemble (or their negations) and the graph *Q*-coloring problem in which constraints simply enforce the condition that the endpoints of the edges in the graph must not have the same color (among the *Q* possible ones). Quite in general a generic CSP can be written as the problem of finding a zero-energy ground state of an appropriate energy function and its analysis amounts at performing a zero-temperature statistical physics study. Hard combinatorial problems are those which correspond to frustrated physical model systems.

Given an instance of a CSP, one wants to know whether there exists a solution, that is, an assignment of the variables which satisfies all the constraints (e.g., a proper coloring). When it exists, the instance is called SAT, and one wants to find a solution. Most of the interesting CSPs are NP-complete: in the worst case, the number of operations needed to decide whether an instance is SAT or not is expected to grow exponentially with the number of variables. But recent years have seen an upsurge of interest in the theory of typical-case complexity, where one tries to identify random ensembles of CSPs which are hard to solve, and the reason for this difficulty. As already mentioned, random ensembles of CSPs are also of great theoretical and practical importance in communication theory, since some of the best modern error-correcting codes (the so-called low-density parity check codes) are based on such constructions.

### Satisfiability and Spin Glass Models

The archetypical example of CSP is satisfiability (SAT). This is a core problem in computational complexity: it is the first one to have been shown NP-complete, and since then thousands of problems have been shown to be computationally equivalent to it. Yet it is not so easy to find difficult instances. The main ensemble which has been used for this goal is the random $K$-SAT ensemble (for $K > 2$, K-SAT is NP-complete).

The SAT problem is defined as follows. Given a vector of $\{0,1\}$ Boolean variables $\boldsymbol{x} = \{x_i\}_{i \in I}$, where $I = \{1, \ldots, n\}$, consider a SAT formula defined by

$$\mathcal{F}(\boldsymbol{x}) = \bigwedge_{a \in A} C_a(\boldsymbol{x})$$

where $A$ is an arbitrary finite set (disjoint with $I$) labeling the clauses $C_a$; $C_a(\boldsymbol{x}) = \bigvee_{i \in I(a)} J_{a,i}(x_i)$; any literal $J_{a,i}(x_i)$ is either $x_i$ or $\sim x_i$ ("not" $x_i$); and finally, $I(a) \subset I$ for every $a \in A$. Similarly to $I(a)$, we can define the set $A(i) \subset A$ as $A(i) = \{a : i \in I(a)\}$, that is, the set of clauses containing variable $x_i$ or its negation.

Given a formula $\mathcal{F}$, the problem of finding a variable assignment $s$ such that $\mathcal{F}(s) = 1$, if it exists, can also be written as a spin glass problem as follows: if we consider a set of $n$ Ising spins, $\sigma_i \in \{\pm 1\}$ in place of the Boolean variables ($\sigma_i = -1, 1 \leftrightarrow x_i = 0, 1$) we may write the energy function associated to each clause as follows:

$$E_a = \prod_{r=1}^{K} \frac{(1 + J_{a,i_r}\sigma_{i_r})}{2}$$

where $J_{a,i} = -1$ (resp. $J_{a,i} = 1$) if $x_i$ (resp. $\tilde{x}_i$) appears in clause $a$. The total energy of a configuration $E = \sum_{a=1}^{|A|} E_a$ is nothing but a $K$-spin spin glass model.

Random $K$-SAT is a version of SAT in which each clause is taken to involve exactly $K$ distinct variables, randomly chosen and negated with uniform distribution. Its energy function corresponds to a spin glass system over a finite connectivity (diluted) random graph.

In recent years random $K$-SAT has attracted much interest in computer science and in statistical physics. The interesting limit is the thermodynamic limit $n \to \infty$, $m = |A| \to \infty$ at fixed clause density $\alpha = m/n$.

Its most striking feature is certainly its sharp threshold. It is strongly believed that there exists a phase transition for this problem: numerical and heuristic analytical arguments are in support of the so-called satisfiability threshold conjecture:

**Conjecture** There exists $\alpha_c(K)$ such that with high probability:

- if $\alpha < \alpha_c(K)$, a random instance is satisfiable;
- if $\alpha > \alpha_c(K)$, a random instance is unsatisfiable.

Although this conjecture remains unproven, the existence of a nonuniform sharp threshold has been established by Friedgut (1997). A lot of effort has been devoted to understanding this phase transition. This is interesting both from physics and the computer science points of view, because the random instances with $\alpha$ close to $\alpha_c$ are the hardest to solve. There exist rigorous results that give bounds for the threshold $\alpha_c(K)$: using these bounds, it was shown that $\alpha_c(K)$ scales as $2^K \ln(2)$ when $K \to \infty$.

On the statistical physics side, the cavity method (which is the generalization to disordered systems characterized by ergodicity breaking of the iterative method used to solve exactly physical models on the Bethe lattice), is a powerful tool which is claimed to be able to compute the exact value of the threshold, giving for instance $\alpha_c(3) \simeq 4.2667\ldots$ It is a nonrigorous method but the self-consistency of its results have been checked by a "stability analysis," and it has also led to the development of a new family of algorithms – the so-called "survey propagation" – which can solve efficiently very large instances at clause densities which are very close to the threshold (for technical details see Mézard and Zecchina (2002) and Braunstein *et al.* (2005) and references therein).

The main hypothesis on which the cavity analysis of random $K$-SAT relies is the existence, in a region of clause density $[\alpha_d, \alpha_c]$ close to the threshold, of an intermediate phase called the "hard-SAT" phase; see **Figure 1**. In this phase the set $\mathcal{S}$ of solutions (a subset of the vertices in an $n$-dimensional hypercube) is supposed to split into many disconnected clusters $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots$. If one considers two solutions $X, Y$ in the same cluster $\mathcal{S}_j$, it is

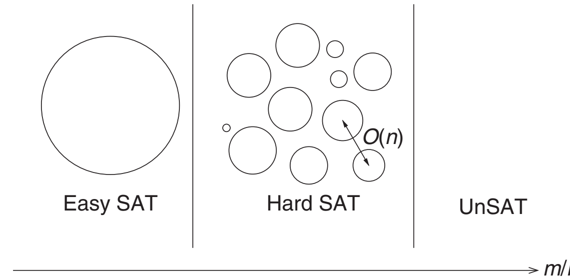



**Figure 1** A pictorial representation of the clustering transition in random $K$-SAT.

possible to walk from $X$ to $Y$ (staying in $\mathcal{S}$) by flipping at each step a finite numbers of variables. If, on the other hand, $X$ and $Y$ are in different clusters, in order to walk from $X$ to $Y$ (staying in $\mathcal{S}$), at least one step will involve an extensive number (i.e., $\propto n$) of flips. This clustered phase is held responsible for entrapping many local-search algorithms into non-optimal metastable states. This phenomenon is not exclusive to random $K$-SAT. It is also predicted to appear in many other hard SAT and optimization problems such as "coloring," and corresponds to the so called "one-step replica symmetry breaking" (1RSB) phase in the language of statistical physics. It is also a crucial limiting feature for decoding algorithms in some error correcting codes.

The only CSP for which the existence of the clustering phase has been established rigorously is the polynomial problem of solving random linear equation in GF (Motwani and Raghavan 2000). For random $K$-SAT, rigorous probabilistic bounds can be used to prove the existence of the clustering phenomenon, for large enough $K$, in some region of $\alpha$ included in the interval $[\alpha_d(K), \alpha_c(K)]$ predicted by the statistical physics analysis.

In the analysis of CSP like $K$-SAT, two main questions are in order. The first is of algorithmic nature and asks for an algorithm which decides whether for a given CSP instance all the constraints can be simultaneously satisfied or not. The second question is more theoretical and deals with large random instances, for which one wants to know the structure of the solution space and predict the typical behavior of classes of algorithms.

#### Message-Passing Algorithms from Statistical Physics

The algorithmic contributions of statistical mechanics to combinatorial optimization are numerous and important (a representative example being the celebrated "simulated annealing algorithm"). For the sake of brevity, here we limit the discussion to the so-called "message-passing algorithms" which are also of great interest in coding theory.

The statistical analysis of the cavity equations allows to study the average properties of ensemble of problems and it is totally equivalent to the replica method in which the average over the ensemble is the first step in any calculation. The survey propagation (SP) equations are a formulation of the cavity equations which is valid for each specific instance and is able to provide information about the statistical behavior of the individual variables in the stable and metastable states of a given energy density (i.e., given fraction of violated constraints).

The single-sample SP equations are nicely described in terms of the factor graph representation used in information theory to characterize error-correcting codes. In the factor graph, the $N$ variables $i, j, k, \ldots$ are represented by circular "variable nodes," whereas the $M$ clauses $a, b, c, \ldots$ are represented by square "function nodes." For random $K$-SAT, the function nodes have connectivity $K$, while the variable nodes have an average Poisson connectivity $K\alpha$.

The iterative SP equations are examples of message-passing procedures. In message-passing algorithms such as the so-called "belief propagation (BP) algorithm" used in error-correcting codes and statistical inference problems, the unknowns which are self-consistently evaluated by iteration are the marginals over the solution space of the variables characterizing the combinatorial problem (the probability space is the set of all solutions sampled with uniform measure). According to the physical interpretation, the quantities that are evaluated by SP are the probability distributions of local fields over the set of clusters. That is, while BP performs a "white" average over solutions, SP takes care of cluster-to-cluster fluctuations, telling us which is the probability of picking up a cluster at random and finding a given variable biased in a certain direction (or unfrozen if it is paramagnetic in the cluster). SP computes quantities which are probabilities over different pure states: the order parameter which is evaluated as fixed point of the SP equations is a probability measure in a space of functions, or for finite $n$, the full list of probability densities describing the cluster-to-cluster fluctuations of the variables.

In both SP and BP one assumes knowledge of the marginals of all variables in the temporary absence of one of them and then writes the marginal probability induced on this "cavity" variable in absence of another third variable interacting with it (i.e., the so-called Bethe lattice approximation for the problem). These relations define a closed set of equations for such cavity marginals that can be solved iteratively (this fact is known as message-passing technique). The equations become exact if the cavity variables acting as inputs are uncorrelated. They are conjectured to be an asymptotically exact approximation over random locally tree–like structures such as, for instance, the random $K$-SAT factor graph. Both BP and SP can be derived in a variational framework.

### Complexity of Counting Problems

In order to describe the nature of computational complexity of counting in physical models, it is enough to consider the classical Ising problem. The

computation of the Ising partition function or, more in general, of the weighted matching polynomial, is the root problem of lattice statistics.

For planar graphs like, for example, two-dimensional regular lattices, counting problems can often be solved by a variety of different methods, for example, transfer matrices and Pfaffians, which require a number of operations which are polynomial in the number of vertices.

The complexity of the counting problems changes if one considers nonplanar graphs, that is, graphs with a nontrivial topological genus. In discrete mathematics, such problems are classified as #P-complete, meaning that the existence of an exact polynomial algorithm for the evaluation of the generating functions would imply the polynomial solvability of many known counting combinatorial problems, the most famous one being the evaluation of the permanent of 0–1 matrices. In statistical mechanics and mathematical chemistry, the interest in nonplanar lattices is obviously related to their $D > 2$ character: the three-dimensional cubic lattice is nothing but a nonplanar graph of topological genus $g = 1 + N/4$, where $N$ is the number of sites.

The planar two-dimensional Ising model was solved in 1944 by Onsager using the algebraic transfer matrix method. Successively, alternative exact solutions have been proposed which resorted to simple combinatorial and geometrical reasoning. As is well known, the underlying idea of the combinatorial methods consists in recasting the sum over spin configurations of the Boltzmann weights as a sum over closed curves (loops) weighted by the activity of their bonds. Double counting is avoided by a proper cancellation mechanism which takes care of the different intrinsic topologies of loops which give rise to the same contribution in the partition function. Such an approach has been developed first by Kac and Ward (1952) and provides a direct way of taking the field theoretic continuum limit. In $D > 2$, the generalization of the above method encounters enormous difficulties due to the variety of intrinsic topologies of surfaces immersed in $D > 2$ lattices.

Another combinatorial method proposed in the 1960s by Kasteleyn is the so-called Pfaffian method. It consists in writing the weighted sum over loops as a dimer covering or prefect matching generating function. Once the relationship between loop counting and dimer coverings (or perfect matchings) over a suitably decorated and properly oriented lattice is established, the Pfaffian method turns out to be a simple technique for the derivation of exact solutions or for the definition of polynomial algorithms over planar lattices which are applicable also to the two-dimensional Ising spin glass.

The generalization of the Pfaffian construction to the nonplanar case must deal with the ambiguity of orienting the homology cycles of the graph. Such a problem can be formally solved in full generality for any orientable lattice and leads to an expression of the Ising partition function or the dimer coverings generating function given as a sum over all possible inequivalent orientations of the lattice (or its embedding surface): for a graph of genus $g$, the homology basis is composed of $2g$ cycles and, therefore, there are $2^{2g}$ inequivalent orientations. It is only for graphs of logarithmic genus that the generalized Pfaffian formalism provides a polynomial algorithm.

Counting perfect matchings can be thought of as the problem of evaluating the permanent of 0–1 matrices over properly constructed bipartite graphs, which is among the oldest and most famous #P-complete problems.

The Pfaffian formalism when applied to the permanent problem leads to a simple general result, that is, it provides a general formula for writing the permanent of a matrix in terms of a number of determinants which is exponential in the genus of the underlying graph.

*See also:* Combinatorics: Overview; Determinantal Random Fields; Dimer Problems; Phase Transitions in Continuous Systems; Spin Glasses; Two-Dimensional Ising Model.

## Further Reading


Achlioptas D, Naor A, and Peres Y (2005) Rigorous location of phase transitions in hard optimization problems. *Nature* 435: 759–764.

Ajtai N (1996) Generating hard instances of lattice problems. *Electronic Colloquium on Computational Complexity (ECCC)* 7: 3.

Braunstein A, Mézard M, and Zecchina R (2005) Survey Propagation: an Algorithm for Satisfiability. *Random Structures and Algorithms* 27: 201–226.

Cocco S and Monasson R (2004) Heuristic average-case analysis of the backtrack resolution of random 3-satisfiability instances. *Theoretical Computer Science A* 320: 345.

Distler J (1992) A note on the 3D Ising model as a string theory. *Nuclear Physics B* 388: 648.

Dubois O, Monasson R, Selman B, and Zecchina R (eds.) (2001) *NP-hardness and Phase transitions* (Special Issue), *Theoretical Computer Science,* vol. 265 (1–2). Elsevier.

Friedgut E (1999) Sharp threshold of graph properties, and the KSat problem. *Journal of American Mathematical Society* 12: 1017–1054.

Jerrum M and Sinclair A (1989) Approximating the permanent. *SIAM Journal on Computing* 18: 1149.

Lovász L and Plummer MD (1986) Matching Theory. *North-Holland Mathematics Studies 121, Annals of Discrete Mathematics (29).* New York: North-Holland.

Mézard M, Mora T, and Zecchina R (2005) Clustering of solutions in the random satisfiability problem. *Physics Review Letters* 94: 197205.

Mézard M and Parisi G (2001) The Bethe lattice spin glass revisited. *European Physical Journal B* 20: 217.
Mézard M, Parisi G, and Virasoro MA (1987) *Sping Glass Theory and Beyond*. Singapore: World Scientific.
Mézard M, Parisi G, and Zecchina R (2002) Analytic and algorithmic solution of random satisfiability problems. *Science* 297: 812–815.
Mézard M and Zecchina R (2002) Random K-satisfiability: from an analytic solution to a new efficient algorithm. *Physical Review E* 66: 056126.
Monasson R, Zecchina R, Kirkpatrick S, Selman B, and Troyansky L (1999) Determining computational complexity from characteristic "phase transitions". *Nature* 400: 133.
Motwani R and Raghavan P (2000) *Randomized Algorithms*. Cambridge: Cambridge University Press.
Nishimori H (2001) *Statistical Physics of Spin Glasses and Information Processing*. Oxford: Oxford University Press.
Papadimitriou CH (1994) *Computational Complexity*. Addison-Wesley.
Regge T and Zecchina R (2000) Combinatorial and topological approach to the 3D Ising model. *Journal of Physics A: Mathematical and General* 33: 741.
Richardson T and Urbanke R (2001) An introduction to the analysis of iterative coding systems. In: Marcus B and Rosenthal J (eds.) *Codes, Systems, and Graphical Models*. New York: Springer.

# Statistical Mechanics of Interfaces

**S Miracle-Solé**, Centre de Physique Théorique, CNRS, Marseille, France



## Introduction

When a fluid is in contact with another fluid, or with a gas, a portion of the total free energy of the system is proportional to the area of the surface of contact, and to a coefficient, the surface tension, which is specific for each pair of substances. Equilibrium will accordingly be obtained when the free energy of the surfaces in contact is a minimum.

Suppose that we have a drop of some fluid, $b$, over a flat substrate, $w$, while both are exposed to air, $a$. We have then three different surfaces of contact, and the total free energy of the system consists of three parts, associated to these three surfaces. A drop of fluid $b$ will exist provided its own two surface tensions exceed the surface tension between the substrate $w$ and the air, that is, provided that

$$\tau^{wb} + \tau^{ba} > \tau^{wa}$$

If equality is attained, then a film of fluid $b$ is formed, a situation which is known as perfect, or complete wetting (see Figure 1).

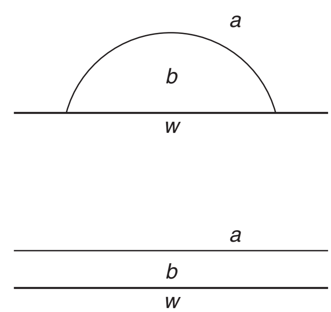



**Figure 1** Partial and complete wetting.

When one of the substances involved is anisotropic, such as a crystal, the contribution to the total free energy of each element of area depends on its orientation. The minimum surface free energy for a given volume then determines the ideal form of the crystal in equilibrium.

It is only in recent times that equilibrium crystals have been produced in the laboratory, first, in negative crystals (vapor bubbles) of organic substances. Most crystals grow under nonequilibrium conditions and it is a subsequent relaxation of the macroscopic crystal that restores the equilibrium.

An interesting phenomenon that can be observed on these crystals is the roughening transition, characterized by the disappearance of the facets of a given orientation, when the temperature attains a certain particular value. The best observations have been made on helium crystals, in equilibrium with superfluid helium, since the transport of matter and heat is then extremely fast. Crystals grow to sizes of 1–5 mm and relaxation times vary from milliseconds to minutes. Roughening transitions for three different types of facets have been observed (see, e.g., Wolf *et al.* (1983)).

These are some classical examples among a variety of interesting phenomena connected with the behavior of the interface between two phases in a physical system. The study of the nature and properties of the interfaces, at least for some simple systems in statistical mechanics, is also an interesting subject of mathematical physics. Some aspects of this study will be discussed in the present article.

We assume that the interatomic forces can be modeled by a lattice gas, and consider, as a simple example, the ferromagnetic Ising model. In a typical two-phase equilibrium state, there is a dense

component, which can be interpreted as a solid or liquid phase, and a dilute phase, which can be interpreted as the vapor phase. Considering certain particular cases of such situations, we first introduce a precise definition of the surface tension and then proceed on the mathematical analysis of some preliminary properties of the corresponding interfaces. The next topic concerns the wetting properties of the system, and the final section is devoted to the associated equilibrium crystal.

## Pure Phases and Surface Tension

The Ising model is defined on the cubic lattice $\mathcal{L} = \boldsymbol{Z}^3$, with configuration space $\Omega = \{-1, 1\}^{\mathcal{L}}$. If $\sigma \in \Omega$, the value $\sigma(i) = -1$ or 1 is the spin at the site $i = (i_1, i_2, i_3) \in \mathcal{L}$, and corresponds to an empty or an occupied site in the lattice gas version of the model. The system is first considered in a finite box $\Lambda \subset \mathcal{L}$, with fixed values of the spins outside.

In order to simplify the exposition, we shall mainly consider the three-dimensional Ising model, though some of the results to be discussed hold in any dimension $d \geq 2$. We shall also, sometimes, refer to the two-dimensional model, it being understood that the definitions have been adapted in the obvious way. We assume that the box $\Lambda$ is a parallelepiped, centered at the origin of $\mathcal{L}$, of sides $L_1, L_2, L_3$, parallel to the axes.

A configuration of spins on $\Lambda(\sigma(i), i \in \Lambda)$, denoted $\sigma_\Lambda$, has an energy defined by the Hamiltonian

$$H_\Lambda(\sigma_\Lambda|\bar{\sigma}) = -J \sum_{\langle i,j\rangle \cap \Lambda \neq \emptyset} \sigma(i)\sigma(j) \qquad [1]$$

where $J$ is a positive constant (ferromagnetic or attractive interaction). The sum runs over all nearest-neighbor pairs $\langle i,j\rangle \subset \mathcal{L}$, such that at least one of the sites belongs to $\Lambda$, and one takes $\sigma(i) = \bar{\sigma}(i)$ when $i \notin \Lambda$, the configuration $\bar{\sigma} \in \Omega$ being the given boundary condition. The probability of the configuration $\sigma_\Lambda$, at the inverse temperature $\beta = 1/kT$, is given by the Gibbs measure

$$\mu_\Lambda(\sigma_\Lambda|\bar{\sigma}) = Z^{\bar{\sigma}}(\Lambda)^{-1} \exp(-\beta H_\Lambda(\sigma_\Lambda|\bar{\sigma})) \qquad [2]$$

where $Z^{\bar{\sigma}}(\Lambda)$ is the partition function

$$Z^{\bar{\sigma}}(\Lambda) = \sum_{\sigma_\Lambda} \exp(-\beta H_\Lambda(\sigma_\Lambda|\bar{\sigma})) \qquad [3]$$

Local properties at equilibrium can be described by the correlation functions between the spins on finite sets of sites,

$$\mu_\Lambda^{\bar{\sigma}}(\sigma(A)) = \sum_{\sigma_\Lambda} \mu_\Lambda(\sigma_\Lambda|\bar{\sigma}) \prod_{i \in A} \sigma(i) \qquad [4]$$

The measures [2] determine (by the Dobrushin–Lanford–Ruelle equations) the set of Gibbs states of the infinite system, as measures on the set $\Omega$ of all configurations. If a Gibbs state happens to be equal to $\lim \mu_\Lambda(\cdot|\bar{\sigma})$, when $L_1, L_2, L_3 \to \infty$, under a fixed boundary condition $\bar{\sigma}$, we shall call it the Gibbs state associated to the boundary condition $\bar{\sigma}$. One also says that this state exists in the thermodynamic limit. Then, equivalently, the correlation functions [4] converge to the corresponding expectation values in this state.

This model presents, at low temperatures (i.e., for $\beta > \beta_c$, where $\beta_c$ is the critical inverse temperature), two different thermodynamic pure phases, a dense and a dilute phase in the lattice gas language (called here the positive and the negative phase). This means two extremal translation-invariant Gibbs states, $\mu^+$ and $\mu^-$, obtained as the Gibbs states associated with the boundary conditions $\bar{\sigma}$, respectively equal to the ground configurations $\bar{\sigma}(i) = 1$ and $\bar{\sigma}(i) = -1$, for all $i \in \mathcal{L}$. The spontaneous magnetization

$$m^*(\beta) = \mu^+(\sigma(i)) = -\mu^-(\sigma(i)) \qquad [5]$$

is then strictly positive. On the other hand, if $\beta \leq \beta_c$, then the Gibbs state is unique and $m^* = 0$.

Each configuration inside $\Lambda$ can be described in a geometric way by specifying the set of Peierls contours which indicate the boundaries between the regions of spin 1 and the regions of spin –1. Unit-square faces are placed midway between the pairs of nearest-neighbor sites $i$ and $j$, perpendicularly to these bonds, whenever $\sigma(i)\sigma(j) = -1$. The connected components of this set of faces are the Peierls contours. Under the boundary conditions (+) and (–), the contours form a set of closed surfaces. They describe the defects of the considered configuration with respect to the ground states of the system (the constant configurations 1 and –1), and are a basic tool for the investigation of the model at low temperatures.

In order to study the interface between the two pure phases, one needs to construct a state describing the coexistence of these phases. This can be done by means of a new boundary condition. Let $\boldsymbol{n} = (n_1, n_2, n_3)$ be a unit vector in $\boldsymbol{R}^3$, such that $n_3 > 0$, and introduce the mixed boundary condition $(\pm, \boldsymbol{n})$, for which

$$\bar{\sigma}(i) = \begin{cases} 1 & \text{if } i \cdot \boldsymbol{n} \geq 0 \\ -1 & \text{if } i \cdot \boldsymbol{n} < 0 \end{cases} \qquad [6]$$

This boundary condition forces the system to produce a defect going transversally through the box $\Lambda$, a large Peierls contour that can be

interpreted as the microscopic interface (also called a domain wall). The other defects that appear above and below the interface can be described by closed contours inside the pure phases.

The free energy per unit area due to the presence of the interface is the surface tension. It is defined by

$$\tau(\boldsymbol{n}) = \lim_{L_1,L_2\to\infty} \lim_{L_3\to\infty} -\frac{n_3}{\beta L_1 L_2} \ln \frac{Z^{\pm,\boldsymbol{n}}(\Lambda)}{Z^{+}(\Lambda)} \qquad [7]$$

In this expression the volume contributions proportional to the free energy of the coexisting phases, as well as the boundary effects, cancel, and only the contributions to the free energy due to the interface are left. The existence of such a quantity indicates that the macroscopic interface, separating the regions occupied by the pure phases in a large volume $\Lambda$, has a microscopic thickness and can therefore be regarded as a surface in a thermodynamic approach.

**Theorem 1** *The interfacial free energy per unit area*, $\tau(\boldsymbol{n})$, *exists*, *is bounded*, *and its extension by positive homogeneity*, $f(\boldsymbol{x}) = |\boldsymbol{x}|\,\tau(\boldsymbol{x}/|\boldsymbol{x}|)$, *is a convex function on* $\boldsymbol{R}^3$. *Moreover*, $\tau(\boldsymbol{n})$ *is strictly positive for* $\beta > \beta_c$, *and vanishes if* $\beta \le \beta_c$.

The existence of $\tau(\boldsymbol{n})$ and also the last statement were proved by Lebowitz and Pfister (1981), in the particular case $\boldsymbol{n} = (0,0,1)$, with the help of correlation inequalities. A complete proof of the theorem was given later with similar arguments. The convexity of $f$ is equivalent to the fact that the surface tension $\tau$ satisfies a thermodynamic stability condition known as the pyramidal inequality (see Messager *et al.* (1992)).

## Gibbs States and Interfaces

In this section we consider the $(\pm, \boldsymbol{n}_0)$ boundary condition, also simply denoted $(\pm)$, associated to the vertical direction $\boldsymbol{n}_0 = (0,0,1)$,

$$\bar{\sigma}(i) = 1 \text{ if } i_3 \ge 0, \quad \bar{\sigma}(i) = -1 \text{ if } i_3 < 0 \qquad [8]$$

The corresponding surface tension is $\tau^{\pm} = \tau(\boldsymbol{n}_0)$. We shall first recall some classical results which concern the Gibbs states and interfaces at low temperatures.

According to the geometrical description of the configurations introduced in the last section, we observe that

$$Z^{\pm,\boldsymbol{n}}(\Lambda)/Z^{+}(\Lambda) = \sum_{\lambda} \exp(-2\beta J|\lambda| - U_\Lambda(\lambda)) \qquad [9]$$

where the sum runs over all microscopic interfaces $\lambda$ compatible with the boundary condition and $|\lambda|$ is the number of faces of $\lambda$ (inside $\Lambda$). The term $U_\Lambda(\lambda)$ equals $-\ln Z^{+}(\lambda,\Lambda)/Z^{+}(\Lambda)$, the sum in the partition function $Z^{+}(\lambda,\Lambda)$ being extended to all configurations whose associated contours do not intersect $\lambda$. Each term in sum [9] gives a weight proportional to the probability of the corresponding microscopic interface.

At low (positive) temperatures, we expect the microscopic interface corresponding to this boundary condition, which at zero temperature coincides with the plane $i_3 = -1/2$, to be modified by small deformations. Each microscopic interface $\lambda$ can then be described by its defects, with respect to the interface at $\beta = \infty$. To this end, one introduces some objects, called walls, which form the boundaries between the horizontal plane portions of the microscopic interface, also called the ceilings of the interface.

More precisely, one says that a face of $\lambda$ is a ceiling face if it is horizontal and such that the vertical line passing through its center does not have other intersections with $\lambda$. Otherwise, one says that it is a wall face. The set of wall faces splits into maximal connected components. The set of walls, associated to $\lambda$, is the set of these components, each component being identified by its geometric form and its projection on the plane $i_3 = -1/2$. Every wall $\omega$, with projection $\pi(\omega)$, increases the energy of the interface by a quantity $2J\|\omega\|$, where $\|\omega\| = |\omega| - |\pi(\omega)|$, and two walls are compatible if their projections do not intersect. In this way, the microscopic interfaces may be interpreted as a "gas of walls" on the two-dimensional lattice.

Dobrushin, who developed the above analysis, also proved the dilute character of this "gas" at low temperatures. This implies that the microscopic interface is essentially flat, or rigid. One can understand this fact by noticing first that the probability of a wall is less than $\exp(-2\beta J\|\omega\|)$ and, second, that in order to create a ceiling in $\lambda$, which is not in the plane $i_3 = -1/2$, one needs to surround it by a wall, that one has to grow when the ceiling is made over a larger area.

Using correlation inequalities one proves that the Gibbs state $\mu^{\pm}$, associated to the $(\pm)$ boundary conditions, always exists, and that it is invariant under horizontal translations of the lattice, that is, $\mu^{\pm}(\sigma(A+a)) = \mu^{\pm}(\sigma(A))$ for all $a = (a_1, a_2, 0)$. It is also an extremal Gibbs state. Let $m(z)$ be the magnetization $\mu^{\pm}((\sigma(z))$ at the site $z = (0,0,z)$. The function $m(z)$ is monotone increasing and satisfies the symmetry property $m(-z) = -m(z+1)$. Some consequences of Dobrushin's work are the following properties.

**Theorem 2** *If the temperature is low enough, that is, if $\beta J \geq c_1$, where $c_1$ is a given constant, then*

$$m^{\pm}(0) \text{ is strictly positive} \qquad [10]$$

$$m^{\pm}(z) \to m^*, \text{ when } z \to \infty, \text{ exponentially fast} \qquad [11]$$

Equation [10] is just another way of saying that the interface is rigid and that the state $\mu^{\pm}$ is non-translation invariant (in the vertical direction). Then, the correlation functions $\mu^{\pm}(\sigma(A))$ describe the local properties, or local structure, of the macroscopic interface. In particular, the function $m(z)$ represents the magnetization profile. Then statement [11], together with the symmetry property, tells us that the thickness of this interface is finite, with respect to the unit lattice spacing.

The statistics of interfaces has been rewritten in terms of a gas of walls and this system may further be studied by cluster expansion techniques. There is an interaction between the walls, coming from the term $U_\Lambda(\lambda)$ in eqn [9], but a convenient mathematical description of this interaction can be obtained by applying the standard low-temperature cluster expansion, in terms of contours, to the regions above and below the interface.

This method was introduced by Gallavotti in his study (mentioned below) of the two-dimensional Ising model. It has been applied by Bricmont and co-workers to examine the interface structure in the present case. As a consequence, it follows that the surface tension, more exactly $\beta\tau^{\pm}(\beta)$, and also the correlation functions, are analytic functions at low temperatures. They can be obtained as explicit convergent series in the variable $\zeta = e^{-2\beta J}$.

The same analysis applied to the two-dimensional model shows a very different behavior at low temperatures. In this case, the microscopic interface $\lambda$ is a polygonal line and the walls belong to the one-dimensional lattice. One can then increase the size of a ceiling without modifying the walls attached to it.

Indeed, Gallavotti turned this observation into a proof that the Gibbs state $\mu^{\pm}$ is now translation invariant. The line $\lambda$ undergoes large fluctuations of order $\sqrt{L_1}$, and disappears from any finite region of the lattice, in the thermodynamic limit. In particular, we have then $\mu^{\pm} = (1/2)(\mu^+ + \mu^-)$, a result that extends to all boundary conditions $(\pm, \boldsymbol{n})$.

Using these results Bricmont and co-workers also studied the local structure of the interface at low temperatures and showed that its intrinsic thickness is finite. To study the global fluctuations, one can compute the magnetization profile by introducing, before taking the thermodynamic limit, a change of scale: $\mu_\Lambda^{\pm}(\sigma(zL_1^{\delta}))$, with $\delta = 1/2$ or near to this value. This is an exact computation that has been done by Abraham and Reed.

Let us come back to the three-dimensional Ising model where we know that the interface orthogonal to a lattice axis is rigid at low temperatures.

**Question 1** At higher temperatures, but before reaching the critical temperature, do the fluctuations of this interface become unbounded, in the thermodynamic limit, so that the corresponding Gibbs state is translation invariant?

One says then that the interface is rough, and it is believed that, effectively, the interface becomes rough when the temperature is raised, undergoing a roughening transition at an inverse temperature $\beta_R > \beta_c$.

It is known that $\beta_R \leq \beta_c^{d=2}$, the critical inverse temperature of the two-dimensional Ising model, since van Beijeren proved, using correlation inequalities, that above this value, the state $\mu^{\pm}$ is not translation invariant. Recalling that the rigid interface may be viewed as a two-dimensional system, the system of walls, a representation that would become inappropriate for a rough interface, one might think that the phase transition of the two-dimensional Ising model is relevant for the roughening transition, and that $\beta_R$ is somewhere near $\beta_c^{d=2}$. Indeed, approximate methods, used by Weeks and co-workers give some evidence for the existence of such a $\beta_R$ and suggest a value slightly smaller than $\beta_c^{d=2}$, as shown in **Table 1**. To this day, however, there appears to be no proof of the fact that $\beta_R > \beta_c$, that is, that the roughening transition for the three-dimensional Ising model really occurs.

At present one is able to study the roughening transition rigorously only for some simplified models with a restricted set of admissible microscopic interfaces. Moreover, the closed contours, describing the defects above and below $\lambda$, are neglected, so that these two regions have the constant configurations 1 or –1, and one has $U_\Lambda(\lambda) = 0$ in eqn [9].

The best known of these models is the classic SOS (solid-on-solid) model in which the interfaces $\lambda$ have the property of being cut only once by all vertical lines of the lattice. This means that $\lambda$ is the graph of a function that can equivalently be used to define the possible configurations of $\lambda$. If $\lambda$ contains the horizontal face with center $(i_1, i_2, i_3 - 1/2)$, then

**Table 1** Some temperature values

| | | |
|---|---|---|
| $d = 3$ | $\beta_c J \sim 0.22$ | approximate critical temperature |
| $d = 3$ | $\beta_R J \sim 0.41$ | conjectured roughening temperature |
| $d = 2$ | $\beta_c J = 0.44$ | exact critical temperature |

the value at $(i_1, i_2)$ of the associated function is $\phi(i_1, i_2) = i_3$.

The proof that the SOS model with the boundary condition (±) has a roughening transition is a highly nontrivial result due to Fröhlich and Spencer. When $\beta$ is small enough, the fluctuations of $\lambda$ are of order $\sqrt{\ln L}$ (in a cubic box of side $L$).

Moreover, other interface models, with additional conditions on the allowed microscopic interfaces, are exactly solvable. The BCSOS (body-centered SOS) model, introduced by van Beijeren, belongs to this class. It is, in fact, the first model for which the existence of a roughening transition has been proved. More recently, also the TISOS (triangular Ising SOS) model, introduced by Blöte and Hilhorst and further studied by Nienhuis and co-workers, has been considered in this context.

The interested reader can find more information and references, concerning the subject of this section, in the review article by Abraham (1986).

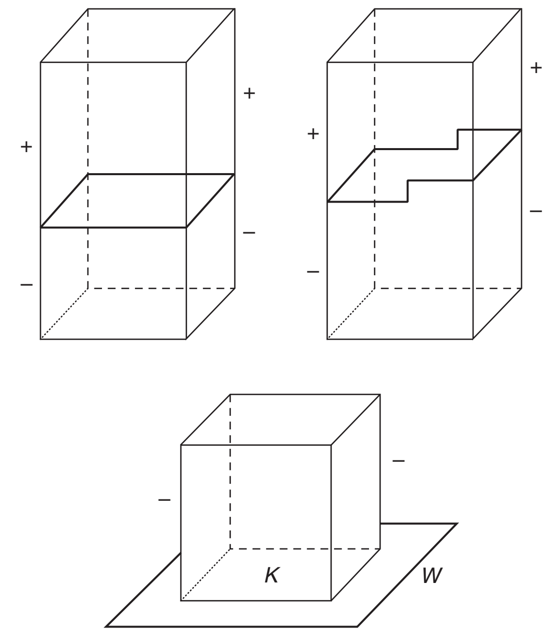



**Figure 2** Boundary conditions for the cubic lattice. Above, the box $\Lambda$ with the (±) and (step) boundary conditions. Below, the box $\Lambda'$ and the wall $W$ with the ($w$ −) boundary conditions.

## Wetting Phenomena

Next we consider the Ising model over a plane horizontal substrate (also called a wall) and study the difference of surface tensions which governs the wetting properties of this system.

We first describe the approach developed by Fröhlich and Pfister (1987) and briefly report some results of their study. We consider the model on the semi-infinite lattice

$$\mathcal{L}' = \{i \in \mathbf{Z}^3 : i_3 \geq 0\} \quad [12]$$

A magnetic field, $K \geq 0$, is added on the boundary sites, $i_3 = 0$, which describes the interaction with the substrate, supposed to occupy the complementary region $\mathcal{L} \setminus \mathcal{L}'$.

We constrain the model in the finite box $\Lambda' = \Lambda \cap \mathcal{L}'$, with $\Lambda$ as above, and impose the value of the spins outside. The Hamiltonian becomes

$$H^w_{\Lambda'}(\sigma_{\Lambda'}|\bar{\sigma}) = -J \sum_{\langle i,j\rangle \cap \Lambda' \neq \emptyset} \sigma(i)\sigma(j) - K \sum_{i \in \Lambda', i_3 = 0} \sigma(i) \quad [13]$$

Here $\sigma_{\Lambda'}$ represents the configuration inside $\Lambda'$, the pairs $\langle i,j \rangle$ are contained in $\mathcal{L}'$, and $\sigma(i) = \bar{\sigma}(i)$ when $i \notin \Lambda'$, the configuration $\bar{\sigma}$ being the given boundary condition (see **Figure 2**). The corresponding partition function is denoted by $Z^{w\bar{\sigma}}(\Lambda')$.

Since there are two pure phases in the model, we must consider two surface free energies, or surfaces tensions, $\tau^{w+}$ and $\tau^{w-}$, between the wall and the positive or negative phase present in the bulk. They are defined through the choice of the boundary condition, $\bar{\sigma}(i) = 1$ or $\bar{\sigma}(i) = -1$, for all $i \in \mathcal{L}'$. Let us consider first the case of the (−) boundary condition.

The surface free energy contribution per unit area due to the presence of the wall, when we have the negative phase in the bulk, is

$$\tau^{w-}(\beta, K) = \lim_{L_1, L_2 \to \infty} \lim_{L_3 \to \infty} -\frac{1}{\beta L_1 L_2} \ln \frac{Z^{w-}(\Lambda')}{Z^-(\Lambda)^{1/2}} \quad [14]$$

The division by $Z^-(\Lambda)^{1/2}$ allows us to subtract from the total free energy, $\ln Z^{w-}(\Lambda')$, the bulk term and all boundary terms which are not related to the presence of the wall. The existence of limit [14] follows from correlation inequalities, and we have $\tau^{w-} \geq 0$.

One can prove, as well, the existence of the Gibbs state $\mu^{w-}$ of the semi-infinite system, associated to the (−) boundary condition. This state is the limit of the finite volume Gibbs measures $\mu_{\Lambda'}(\sigma_{\Lambda'}|(-))$ defined by the Hamiltonian [13]. It describes the local equilibrium properties of the system near the wall, when deep inside the bulk the system is in the negative phase. Similar definitions give the surface tension $\tau^{w+}$ and the Gibbs state $\mu^{w+}$, corresponding to the boundary condition $\bar{\sigma}(i) = 1$, for all $i \in \Lambda'$.

We remark that the states $\mu^{w+}$ and $\mu^{w-}$ are invariant by translations parallel to the plane $i_3 = 0$, and introduce the magnetizations, $m^{w-}(z) = \mu^{w-}(\sigma(z))$, where $z$ denotes the site $(0, 0, z)$, $m^{w-} = m^{w-}(0)$, and

similarly $m^{w+}(z)$ and $m^{w+}$. Their connection with the surface free energies is given by the formula

$$\tau^{w-}(\beta,K) - \tau^{w+}(\beta,K) = \int_0^K (m^{w+}(\beta,s) - m^{w-}(\beta,s))\,ds \qquad [15]$$

We mention in the following theorem some results of Fröhlich and Pfister's study. Here $\tau^{\pm}$ is, as before, the usual surface tension between the two pure phases of the system, for a horizontal interface.

**Theorem 3** *With the above definitions, we have*

$$\tau^{w-}(\beta,K) - \tau^{w+}(\beta,K) \le \tau^{\pm}(\beta) \qquad [16]$$

$$m^{w+}(\beta,K) - m^{w-}(\beta,K) \ge 0 \qquad [17]$$

*and the difference in [17] is a monotone decreasing function of the parameter K. Moreover, if $m^{w+} = m^{w-}$, then the Gibbs states $\mu^{w+}$ and $\mu^{w-}$ coincide.*

The proof is a subtle application of correlation inequalities. Since, from Theorem 3, the integrand in eqn [15] is a positive and decreasing function, the difference $\Delta\tau = \tau^{w-} - \tau^{w+}$ is a monotone increasing and concave (and hence continuous) function of the parameter $K$. On the other hand, one can prove that $\Delta\tau = \tau^{\pm}$, if $K \ge J$. This justifies the following definition:

$$K_w(\beta) = \min\{K\colon \Delta\tau(\beta,K) = \tau^{+-}(\beta)\} \qquad [18]$$

In the thermodynamic description of wetting, the partial-wetting regime is characterized by the strict inequality in [16]. Equivalently, by $K < K_w(\beta)$. We must have then $m^{w+} \neq m^{w-}$, because of eqn [15]. This shows that, in the case of partial wetting, $\mu^{w+}$ and $\mu^{w-}$ are different Gibbs states.

The complete-wetting regime is characterized by the equality in [16], that is, by $K \ge K_w(\beta)$. Then, we have $m^{w+} = m^{w-}$, and taking into account the last statement in Theorem 3, also $\mu^{w+} = \mu^{w-}$. This last result implies that there is only one Gibbs state. Thus, complete wetting corresponds to unicity of the Gibbs state.

In this case, we also have $\lim m^{w-}(z) = m^*$, when $z \to \infty$, because this is always true for $m^{w+}(z)$. This indicates that we are in the positive phase of the system although we have used the (–) boundary condition, so that the bulk negative phase cannot reach the wall anymore. The film of positive phase, which wets the wall completely, has an infinite thickness with respect to the unit lattice spacing, in the thermodynamic limit.

When $\beta = \infty$, only a few particular ground configurations contribute to the partition functions, such as the configuration $\sigma(i) = -1$ for the partition function $Z^{w-}$, etc., and we obtain $\Delta\tau = 2K$ and $\tau^{\pm} = 2J$. For nonzero but low temperatures, the small perturbations of these ground states have to be considered, a problem that can be treated by the method of cluster expansions. In fact, the corresponding defects can be described by closed contours as in the case of pure phases.

**Theorem 4** *For $K < J$, the functions $\beta\tau^{w-}(\beta,K)$ and $\beta\tau^{w+}(\beta,K)$ are analytic at low temperatures, that is, provided that $\beta(J-K) \ge c_2$, where $c_2$ is a given constant. Moreover, $m^{w+}(z)$ and $m^{w-}(z)$ tend, respectively, to $m^*$ and to $-m^*$, when $z \to \infty$, exponentially fast.*

The last statement in Theorem 4 tells us that the wall affects only a layer of finite thickness (with respect to the lattice spacing). From a macroscopic point of view, the negative phase reaches the wall, and we are in the partial-wetting regime. Indeed, a strict inequality holds in [16].

Thus, for $K < J$ there is always partial wetting at low temperatures. Then the following question arises:

**Question 2** Is there a situation of complete wetting at higher temperatures? It is understood here that $K$ takes a fixed value, characteristic of the substrate, such that $0 < K < J$.

This is known to be the case in dimension $d = 2$, where the exact value of $K_w(\beta)$ can be obtained from Abraham's solution of the model:

$$\cosh 2\beta K_w = \cosh 2\beta J - e^{-2\beta J}\sinh 2\beta J$$

Then complete wetting occurs for $\beta$ in the interval $\beta_c < \beta \le \beta_w(K)$, where $\beta_c$ is the critical inverse temperature and $\beta_w(K)$ is the solution of $K_w(\beta) = K$. The case $d = 2$ has been reviewed in Abraham (1986).

To our knowledge, the above question remains an open problem for the Ising model in dimension $d = 3$. The problem has, however, been solved for the simpler case of a SOS interface model. In this case, a nice and rather brief proof of the following result has been given by Chalker (1982): one has $m^{w+} = m^{w-}$, and hence complete wetting, if

$$2\beta(J-K) < -\ln(1 - e^{-8\beta J})$$

It is very plausible that a similar statement is valid for the semi-infinite Ising model and, also that Chalker's method could play a role for extending the proof to this case, provided an additional assumption is made. Namely, that $\beta$ is sufficiently large, and hence $J - K$ small enough, in order to insure the convergence of the cluster expansions and to be able to use them.

## Equilibrium Crystals

The shape of an equilibrium crystal is obtained, according to thermodynamics, by minimizing the surface free energy between the crystal and the medium, for a fixed volume of the crystal phase. Given the orientation-dependent surface tension $\tau(\boldsymbol{n})$, the solution to this variational problem, known under the name of Wulff construction, is the following set:

$$\mathcal{W} = \{\boldsymbol{x} \in \boldsymbol{R}^3 : \boldsymbol{x} \cdot \boldsymbol{n} \leq \tau(\boldsymbol{n}) \text{ for all } \boldsymbol{n}\} \qquad [19]$$

Notice that the problem is scale invariant, so that if we solve it for a given volume of the crystal, we get the solution for other volumes by an appropriate scaling. We notice also that the symmetry $\tau(\boldsymbol{n}) = \tau(-\boldsymbol{n})$ is not required for the validity of formula [19]. In the present case, $\tau(\boldsymbol{n})$ is obviously a symmetric function, but nonsymmetric situations are also physically interesting and appear, for instance, in the case of a drop on a wall discussed in the last section.

The surface tension in the Ising model between the positive and negative phases has been defined in eqn [7]. In the two-dimensional case, this function $\tau(\boldsymbol{n})$ has (as shown by Abraham) an exact expression in terms of some Onsager's function. It follows (as explained in Miracle-Sole (1999)) that the Wulff shape $\mathcal{W}$, in the plane $(x_1, x_2)$, is given by

$$\cosh \beta x_1 + \cosh \beta x_2 \leq \cosh^2 2\beta J / \sinh 2\beta J$$

This shape reduces to the empty set for $\beta \leq \beta_c$, since the critical $\beta_c$ satisfies $\sinh 2J\beta_c = 1$. For $\beta > \beta_c$, it is a strictly convex set with smooth boundary.

In the three-dimensional case, only certain interface models can be exactly solved (see the section "Gibbs states and interfaces"). Consider the Ising model at zero temperature. The ground configurations have only one defect, the microscopic interface $\lambda$, imposed by the boundary condition $(\pm, \boldsymbol{n})$. Then, from eqn [9], we may write

$$\tau(\boldsymbol{n}) = \lim_{L_1, L_2 \to \infty} \frac{n_3}{L_1 L_2} (E_\Lambda(\boldsymbol{n}) - \beta^{-1} N_\Lambda(\boldsymbol{n})) \qquad [20]$$

where $E_\Lambda = 2J|\lambda|$ is the energy (all $\lambda$ have the same minimal area) and $N_\Lambda$ the number of ground states. Every such $\lambda$ has the property of being cut only once by all straight lines orthogonal to the diagonal plane $i_1 + i_2 + i_3 = 0$, provided that $n_k > 0$, for $k = 1, 2, 3$. Each $\lambda$ can then be described by an integer function defined on a triangular plane lattice, the projection of the cubic lattice $\mathcal{L}$ on the diagonal plane. The model defined by this set of admissible microscopic interfaces is precisely the TISOS model. A similar definition can be given for the BCSOS model that describes the ground configurations on the body-centered cubic lattice.

From a macroscopic point of view, the roughness or the rigidity of an interface should be apparent when considering the shape of the equilibrium crystal associated with the system. A typical equilibrium crystal at low temperatures has smooth plane facets linked by rounded edges and corners. The area of a particular facet decreases as the temperature is raised and the facet finally disappears at a temperature characteristic of its orientation. It can be argued that the disappearance of the facet corresponds to the roughening transition of the interface whose orientation is the same as that of the considered facet.

The exactly solvable interface models mentioned above, for which the function $\tau(\boldsymbol{n})$ has been computed, are interesting examples of this behavior, and provide a valuable information on several aspects of the roughening transition. This subject has been reviewed by Abraham (1986), van Beijeren and Nolden (1987), and Kotecky (1989).

For example, we show in **Figure 3** the shape predicted by the TISOS model (one-eighth of the shape because of the condition $n_k > 0$). In this model, the interfaces orthogonal to the three coordinate axes are rigid at low temperatures.

For the three-dimensional Ising model at positive temperatures, the description of the microscopic interface, for any orientation $\boldsymbol{n}$, appears as a very difficult problem. It has been possible, however, to analyze the interfaces which are very near to the particular orientations $\boldsymbol{n}_0$, discussed in the

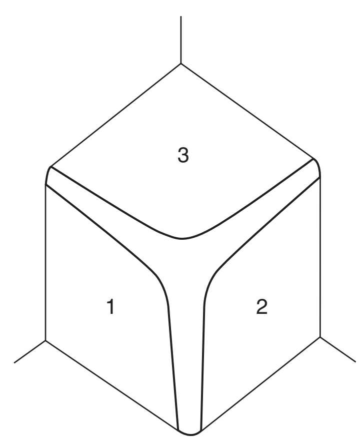



**Figure 3** Cubic equilibrium crystal shown in a projection parallel to the (1,1,1) direction. The three regions (1, 2, and 3) indicate the facets and the remaining area represents a curved part of the crystal surface.

section "Gibbs states and interfaces." This analysis allows us to determine the shape of the facets in a rigorous way.

We first observe that the appearance of a facet in the equilibrium crystal shape is related, according to the Wulff construction, to the existence of a discontinuity in the derivative of the surface tension with respect to the orientation. More precisely, assume that the surface tension satisfies the convexity condition of Theorem 1, and let this function $\tau(\boldsymbol{n}) = \tau(\theta, \phi)$ be expressed in terms of the spherical coordinates of $\boldsymbol{n}$, the vector $\boldsymbol{n}_0$ being taken as the $x_3$-axis. A facet orthogonal to $\boldsymbol{n}_0$ appears in the Wulff shape if and only if the derivative $\partial\tau(\theta, \phi)/\partial\theta$ is discontinuous at the point $\theta = 0$, for all $\phi$. The facet $\mathcal{F} \subset \partial\mathcal{W}$ consists of the points $\boldsymbol{x} \in \boldsymbol{R}^3$ belonging to the plane $x_3 = \tau(\boldsymbol{n}_0)$ and such that, for all $\phi$ between 0 and $2\pi$,

$$x_1 \cos\phi + x_2 \sin\phi \leq \partial\tau(\theta, \phi)/\partial\theta|_{\theta=0^+} \qquad [21]$$

The step free energy is expected to play an important role in the facet formation. It is defined as the free energy associated with the introduction of a step of height 1 on the interface, and can be regarded as an order parameter for the roughening transition. Let $\Lambda$ be a parallelepiped as in the section "Pure phases and surface tension," and introduce the (step, $\boldsymbol{m}$) boundary conditions (see Figure 2), associated to the unit vectors $\boldsymbol{m} = (\cos\phi, \sin\phi) \in \boldsymbol{R}^2$, by

$$\bar{\sigma}(i) = \begin{cases} 1 & \text{if } i > 0 \text{ or if } i_3 = 0 \text{ and} \\ & \quad i_1 m_1 + i_2 m_2 \geq 0 \\ -1 & \text{otherwise} \end{cases} \qquad [22]$$

Then, the step free energy per unit length for a step orthogonal to $\boldsymbol{m}$ (with $m_2 > 0$) on the horizontal interface, is

$$\tau^{\text{step}}(\phi) = \lim_{L_1\to\infty} \lim_{L_2\to\infty} \lim_{L_3\to\infty} -\frac{\cos\phi}{\beta L_1} \ln \frac{Z^{\text{step},\boldsymbol{m}}(\Lambda)}{Z^{\pm,\boldsymbol{n}_0}(\Lambda)} \qquad [23]$$

A first result concerning this point was obtained by Bricmont and co-workers, by proving a correlation inequality which establish $\tau^{\text{step}}(0)$ as a lower bound to the one-sided derivative $\partial\tau(\theta, 0)/\partial\theta$ at $\theta = 0^+$ (the inequality extends also to $\phi \neq 0$). Thus, when $\tau^{\text{step}} > 0$, a facet is expected.

Using the perturbation theory of the horizontal interface, it is possible to also study the microscopic interfaces associated with the (step, $\boldsymbol{m}$) boundary conditions. When considering these configurations, the step may be viewed as an additional defect on the rigid interface described in the section "Pure phases and surface tension." It is, in fact, a long wall going from one side to the other side of the box $\Lambda$. The step structure at low temperatures can then be analyzed with the help of a new cluster expansion. As a consequence of this analysis, we have the following theorem.

**Theorem 5** *If the temperature is low enough, that is, if $\beta J \geq c_3$, where $c_3$ is a given constant, then the step free energy, $\tau^{\text{step}}(\phi)$, exists, is strictly positive, and extends by positive homogeneity to a strictly convex function. Moreover, $\beta\tau^{\text{step}}(\phi)$ is an analytic function of $\zeta = e^{-2J\beta}$, for which an explicit convergent series expansion can be found.*

Using the above results on the step structure, similar methods allow us to evaluate the increment in surface tension of an interface tilted by a very small angle $\theta$ with respect to the rigid horizontal interface. This increment can be expressed in terms of the step free energy, and one obtains the following relation.

**Theorem 6** *For $\beta J \geq c_3$, we have*

$$\partial\tau(\theta, \phi)/\partial\theta|_{\theta=0^+} = \tau^{\text{step}}(\phi) \qquad [24]$$

This relation, together with eqn [21], implies that one obtains the shape of the facet by means of the two-dimensional Wulff construction applied to the step free energy. The reader will find a detailed discussion on these points, as well as the proofs of Theorems 5 and 6, in Miracle-Sole (1995).

From the properties of $\tau^{\text{step}}$ stated in Theorem 5, it follows that the Wulff equilibrium crystal presents well-defined boundary lines, smooth and without straight segments, between a rounded part of the crystal surface and the facets parallel to the three main lattice planes.

It is expected, but not proved, that at a higher temperature, but before reaching the critical temperature, the facets associated with the Ising model undergo a roughening transition. It is then natural to believe that the equality [24] is true for any $\beta$ larger than $\beta_R$, allowing us to determine the facet shape from eqns [21] and [24], and that for $\beta \leq \beta_R$, both sides in this equality vanish, and thus, the disappearance of the facet is involved. However, the condition that the temperature is low enough is needed in the proofs of Theorems 5 and 6.

*See also:* Dimer Problems; Phase Transitions in Continuous Systems; Phase Transition Dynamics; Two-Dimensional Ising Model; Wulff Droplets.

## Further Reading

Abraham DB (1986) Surface structures and phase transitions. In: Domb C and Lebowitz JL (eds.) *Critical Phenomena*, vol. 10, pp. 1–74. London: Academic Press.
Chalker JT (1982) The pinning of an interface by a planar defect. *Journal of Physics A: Mathematical and General* 15: L481–L485.
Fröhlich J and Pfister CE (1987) Semi-infinite Ising model. I. *Communications in Mathematical Physics* 109: 493–523.
Fröhlich J and Pfister CE (1987) Semi-infinite Ising model. II. *Communication in Mathematical Physics* 112: 51–74.
Gallavotti G (1999) *Statistical Mechanics: A Short Treatise*. Berlin: Springer.
Kotecky R (1989) Statistical mechanics of interfaces and equilibrium crystal shapes. In: Simon B, Truman A, and Davies IM (eds.) *IX International Congress of Mathematical Physics*, pp. 148–163. Bristol: Adam Hilger.
Lebowitz JL and Pfister CE (1981) Surface tension and phase coexistence. *Physical Review Letters* 46: 1031–1033.
Messager A, Miracle-Sole S, and Ruiz J (1992) Convexity properties of the surface tension and equilibrium crystals. *Journal of Statistical Physics* 67: 449–470.
Miracle-Sole S (1995) Surface tension, step free energy and facets in the equilibrium crystal shape. *Journal of Statistical Physics* 79: 183–214.
Miracle-Sole S (1999) Facet shapes in a Wulff crystal. In: Miracle-Sole S, Ruiz J, and Zagrebnov V (eds.) *Mathematical Results in Statistical Mechanics*, pp. 83–101. Singapore: World Scientific.
van Beijeren H and Nolden I (1987) The roughening transition. In: Schommers W and von Blackenhagen P (eds.) *Topics in Current Physics*, vol. 43. pp. 259–300. Berlin: Springer.
Widom B (1991) Interfacial phenomena. In: Hansen JP, Levesque D, and Zinn-Justin J (eds.) *Liquid, Freezing and the Glass Transition*, pp. 506–546. Amsterdam: North-Holland.
Wolf PE, Balibar S, and Gallet F (1983) Experimental observations of a third roughening transition in hcp $^4$He crystals. *Physical Review Letters* 51: 1366–1369.

# Stochastic Differential Equations

**F Russo**, Université Paris 13, Villetaneuse, France



## Introduction

Stochastic differential equations (SDEs) appear today as a modeling tool in several sciences as telecommunications, economics, finance, biology, and quantum field theory.

An SDE is essentially a classical differential equation which is perturbed by a random noise. When nothing else is specified, SDE means in fact ordinary SDE; in that case it corresponds to the perturbation of an ordinary differential equation. Stochastic partial differential equations (SPDEs) are obtained as random perturbation of partial differential equations (PDEs).

One of the most important difference between deterministic and stochastic ordinary differential equations is described by the so-called Peano type phenomenon. A classical differential equation with continuous and linear growth coefficients admits global existence but not uniqueness as classical calculus text books illustrate studying equations of the type

$$\frac{\mathrm{d}X}{\mathrm{d}t}(t) = \sqrt{X(t)}, \quad X(0) = 0$$

However, if one perturbs the right member of the equality with an additive Gaussian white noise $(\xi_t)$ (even with very small intensity), then the problem becomes well stated. A similar phenomenon happens with linear PDEs of evolution type perturbed with a spacetime white noise.

SDEs constitute a vast subject and account for an incredible amount of relevant contributions. We try to orientate the reader about the main axes trying to indicate references to the different subfields. We will prefer to refer to monographs when available, instead of articles.

## Motivation and Preliminaries

In the whole article $T$ will be a strictly positive real number. Let us consider continuous functions $b:\mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$, $a:\mathbb{R}_+ \times \mathbb{R}^{d\times m} \to \mathbb{R}^d$ and $x_0 \in \mathbb{R}^d$. We consider a differential problem of the following type:

$$\begin{aligned} \frac{\mathrm{d}X_t}{\mathrm{d}t} &= b(t, X_t) \\ X_0 &= x_0 \end{aligned} \qquad [1]$$

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space. Suppose that previous equation is perturbed by a random noise $(\xi_t)_{t\geq 0}$. Because of modeling reasons it could be reasonable to suppose $(\xi_t)_{t\geq 0}$ satisfying the following properties.

1. It is a family of independent random variables (r.v.'s)
2. $(\xi_t)_{t\geq 0}$ is "stationary", that is, for any positive integer $n$, positive reals $h, t_0, t_1, \ldots, t_n$, the law of $(\xi_{t_0+h}, \ldots, \xi_{t_n+h})$ does not depend on $h$.

More precisely we perturb eqn [1] as follows:

$$\frac{dX_t}{dt} = b(t, X_t) + a(t, X_t)\xi_t \qquad X_0 = x_0 \tag{2}$$

We suppose for a moment that $d = m = 1$. In reality no reasonable real-valued process $(\xi_t)_{t\geq 0}$ fulfilling previous assumptions exists. In particular, if process $(\xi_t)$ exists (resp. $(\xi_t)$ exists and each $\xi_t$ is a square-integrable r.v.), then the process cannot have continuous paths (resp. it cannot be measurable with respect to $\Omega \times \mathbb{R}_+$). However, suppose that such a process exists; we set $B_t = \int_0^t \xi_s ds$. In that case, properties (1) and (2) can be translated into the following on $(B_t)$.

(P1) It has independent increments, which means that for any $t_0, \dots, t_n, h \geq 0, B_{t_1+h} - B_{t_0+h}, \dots, B_{t_n+h} - B_{t_{n-1}+h}$ are independent r.v.'s.
(P2) It has stationary increments, which means that for any $t_0, \dots, t_n, h \geq 0$, the law of $(B_{t_1+h} - B_{t_0+h}, \dots, B_{t_n+h} - B_{t_{n-1}+h})$ does not depend on $h$.

On the other hand, it is natural to require that

(C1) $B_0 = 0$ a.s.,
(C2) it is a continuous process, that is, it has continuous paths a.s.

Equation [2] should be rewritten in some integral form

$$X_t = X_0 + \int_0^t b(s, X_s)ds + \int_0^t a(s, X_s)dB_s \tag{3}$$

Clearly the paths of process $(B_t)$ cannot be differentiable, so one has to give meaning to integral $\int_0^t a(s, X_s)dB_s$. This will be intended in the "Itô" sense, see considerations below.

An important result of probability theory says that a stochastic process $(B_t)$ fulfilling properties P1, P2 and C1, C2 is essentially a "Brownian motion". More precisely, there are real constants $b, \sigma$ such that $B_t = bt + \sigma W_t$, where $(W_t)$ is a classical Brownian motion defined below.

**Definition 1**

(i) A (continuous) stochastic process $(W_t)$ is called classical "Brownian motion" if $W_0 = 0$ a.s., it has independent increments and the law of $W_t - W_s$ is a Gaussian $N(0, t - s)$ *r.v.*
(ii) A $m$-dimensional Brownian motion is a vector $(W^1, \dots, W^m)$ of independent classical Brownian motions.

Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration fulfilling the usual conditions, see (Karatzas and Shreve (1991, section 1.1). There one can find basic concepts of the theory of stochastic processes as the concept of adapted, progressively measurable process. An adapted process is also said to be nonanticipating towards the filtration $(\mathcal{F}_t)$ which represents the state of the information at each time $t$. A process $(X_t)$ is said to be adapted if for any $t, X_t$ is $\mathcal{F}_t$-measurable. The notion of progressively measurable process is a slight refinement of the notion of adapted process.

**Definition 2**

(i) A (continuous) $(\mathcal{F}_t)$ adapted process $(W_t)$ is called (classical) $(\mathcal{F}_t)$-Brownian motion if $W_0 = 0$, if for any $s < t$ $W_t - W_s$ is an $N(0, t - s)$ distributed r.v. which is independent of $\mathcal{F}_s$.
(ii) An $(\mathcal{F}_t)$-$m$-dimensional Brownian motion is a vector $(W^1, \dots, W^m)$ of $(\mathcal{F}_t)$-classical independent Brownian motions.

From now on, we will consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $(\mathcal{F}_t)_{t\geq 0}$ fulfilling the usual conditions. From now on all the considered filtrations will have that property.

Let $W = (W_t)_{t\geq 0}$ be an $(\mathcal{F}_t)_{t\geq 0}$-$m$-dimensional classical Brownian motion. In Karatzas and Shreve (1991, chapter 3) and Revuz and Yor (1999, chapter 4), one introduces the notion of stochastic Itô integral announced before. Let $Y = (Y^1, \dots, Y^m)$ be a progressively measurable $m$-dimensional process such that $\int_0^T \|Y_s\|^2 ds < \infty$, then the Itô integral $\int_0^T Y_s dW_s$ is well defined. In particular the indefinite integral $\int_0^{\cdot} Y_s dW_s$ is an $(\mathcal{F}_t)$-progressively measurable continuous process. If $Y$ is an $\mathbb{R}^{d\times m}$ matrix-valued process, the integral $\int_0^t Y_s dW_s$ is componentwise defined and it will be a vector in $\mathbb{R}^d$. The analogous of differential calculus in the framework of stochastic processes is Itô calculus, see again Karatzas and Shreve (1991, chapter 3) and Revuz and Yor (1999, chapter 4). Important tools are the concept of quadratic variation $[X]$ of a stochastic process when it exists. For instance, the quadratic variation $[W]_t$ of a classical Brownian motion equals $t$. If $M_t = \int_0^t Y_s dW_s$, then $[M]_t = \int_0^t \|Y_s\|^2 ds$. One celebrated theorem of P Lévy states the following: if $(M_t)$ defines a continuous $(\mathcal{F}_t)$-local martingale such that $[M_t] \equiv t$, then $M$ is an $(\mathcal{F}_t)$-classical Brownian motion. That theorem is called the "Lévy characterization theorem of Brownian motion." Itô formula constitutes the natural generalization of fundamental theorem of differential calculus to the stochastic calculus. Another significant tool is Girsanov theorem; it states essentially the following: suppose that the following so-called "Novikov condition" is verified:

$$\mathbb{E}\left(\exp\left(\frac{1}{2}\int_0^T \|Y_t\|^2\right)dt\right) < \infty$$

Then the process $\tilde{W}_t = W_t + \int_0^t Y_s \mathrm{d}s, t \in [0,T]$ is again an $m$-dimensional $(\mathcal{F}_t)$-classical Brownian motion under a new probability measure $Q$ on $(\Omega, \mathcal{F}_T)$ defined by

$$\mathrm{d}Q = \mathrm{d}P \exp\left(\int_0^t Y_s \mathrm{d}W_s - \frac{1}{2}\|Y_s\|^2 \mathrm{d}s\right)$$

Let $\xi$ be an $\mathcal{F}_0$-measurable r.v., for instance, $\xi \equiv x \in \mathbb{R}^d$. We are interested in the SDE

$$\begin{aligned} \mathrm{d}X_t &= a(t, X_t)\,\mathrm{d}W_t + b(t, X_t)\,\mathrm{d}t \\ X_0 &= \xi \end{aligned} \quad [4]$$

**Definition 3** A progressively measurable process $(X_t)_{t\in[0,T]}$ is said to be solution of [4] if a.s.

$$X_t = Z + \int_0^t a(t, X_t)\,\mathrm{d}W_t + \int_0^t b(t, X_t)\,\mathrm{d}t \quad \forall t \in [0,T] \quad [5]$$

provided that the right-hand side member makes sense. In particular, such a solution is continuous. The function a (resp. b) is called the diffusion (drift) coefficient of the SDE. $a$ and $b$ may sometimes be allowed to be random; however, this dependence has to be progressively measurable. Clearly, we can define the notion of solution $(X_t)_{t\geq 0}$ on the whole positive real axis.

We remark that those equations are called Itô SDEs. A solution of previous equation is named diffusion process.

## The Lipschitz Case

The most natural framework for studying the existence and uniqueness for SDEs appears when the coefficients are Lipschitz.

A function $\gamma : [0,T] \times \mathbb{R}^m \longrightarrow \mathbb{R}^d$ is said to have "polynomial growth" (with respect to $x$ uniformly in $t$), if for some $n$ there is a constant $C > 0$ with

$$\sup_{t\in[0,T]} \|\gamma(t,x)\| \leq C(1 + \|x\|^n) \quad [6]$$

The same function is said to have "linear growth" if [6] holds with $n = 1$. A function $\gamma : \mathbb{R}_+ \times \mathbb{R}^m \to \mathbb{R}^d$ is said to be "locally Lipschitz" (with respect to $x$ uniformly in $t$), if for every $t \in [0,T], K > 0$, $\gamma_{|[0,T]\times[-K,K]}$ is Lipschitz (with respect to $x$ uniformly with respect to $t$).

Let $a : \mathbb{R}_+ \times \mathbb{R}^{d\times m} \longrightarrow \mathbb{R}^d, b : \mathbb{R}_+ \times \mathbb{R}^d \longrightarrow \mathbb{R}^d$, be Borel functions, $\xi$ an $\mathbb{R}^d$-valued r.v. $\mathcal{F}_0$-measurable and $(W_t)_{t\geq 0}$ be a $m$-dimensional $(\mathcal{F}_t)$-Brownian motion.

Classical fixed-point theorems allow to establish the following classical result.

**Theorem 1** *We suppose a and b locally Lipschitz with linear growth. Let $\xi$ be a square-integrable r.v. that is $\mathcal{F}_0$-measurable. Then [4] has a unique solution X. Moreover,*

$$\mathbb{E}\left(\sup_{t\leq T} |X_t|^2\right) < \infty$$

**Remark 1**

(i) Equation [4] can be settled similarly by putting initial condition $x$ at some time $s$. In that case the problem is again well stated. If $\xi \equiv x$ is a deterministic point of $\mathbb{R}^d$, then we will often denote by $X^{s,x}$ the solution of that problem.

(ii) If the coefficients are only locally Lipschitz, the equation may be solved until a stopping time. If $d = 1$, it is possible to state necessary and sufficient conditions for nonexplosion (Feller test).

(iii) The theorem above admits several generalizations. For instance, the Brownian motion can be replaced by general semimartingales, (possibly with jumps as Lévy processes).

An important role of diffusion processes is the fact that they provide probabilistic representation to PDEs of parabolic (and even elliptic) type. We will only mention here the parabolic framework.

We denote $A(t,x) = a(t,x)a(t,x)^*$, where $*$ means transposition for matrices. $(t,x) \to A(t,x) = (A_{ij}(t,x))$ is a $d \times d$ matrix-valued function. Let us consider also continuous functions $k : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d, g : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ with polynomial growth or non-negative.

Given a solution of [4], we can associate its generator $(L_t, t \in [0,T])$ setting

$$L_t f(x) = \frac{1}{2}\sum_{i,j=1}^d A_{ij}(t,x)\partial_{ij}^2 f(x) + b(t,x) \cdot \nabla f(x)$$

Feynman–Kac theorem is stated below and it provides probabilistic representation of an associated parabolic linear PDEs.

**Theorem 2** *Suppose there is a function $v : [0,T[ \times \mathbb{R}^d \to \mathbb{R}^d$ continuous with polynomial growth of class $C^{1,2}([0,T] \times \mathbb{R}^d)$ satisfying the following Cauchy problem:*

$$\begin{aligned} (\partial_t v + L_t)v - kv &= g \\ v(T,x) &= f(x) \end{aligned} \quad [7]$$

*Then*

$$v(s,x) = E\bigg(f(X_T)\exp\left(-\int_s^T k(\theta, X_\theta)\,\mathrm{d}\theta\right) - \int_s^T g(t, X_t)\exp\left\{-\int_s^t k(\theta, X_\theta)\,\mathrm{d}\theta\right\}\mathrm{d}t\bigg)$$

*for* $(s,x) \in [0,T] \times \mathbb{R}^d$, *where* $X = X^{s,x}$. *In particular, such a solution is unique.*

**Remark 2**

(i) In order to obtain "classical solutions" of the above Cauchy problem, one needs some conditions. It is the case, for instance, when the following ellipticity condition holds on $A$:

$$\exists c > 0,\ \forall (t,x) \in [0,T] \times \mathbb{R}^n,\ \forall (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n$$

$$\sum_{i,j}^{d} A_{ij}(t,x)\xi_i\xi_j \geq c \sum_{i=1}^{d} |\xi_i|^2 \qquad [8]$$

In the degenerate case, it is possible to deal with viscosity solutions, in the sense of P L Lions. This theorem establishes an important link between deterministic PDEs and SDEs.

(ii) A natural generalization of Feynman–Kac theorem comes from the system of forward–backward SDEs in the sense of Pardoux and Peng.

(iii) Other types of probabilistic representation do appear in stochastic control theory through the so-called verification theorems, see for instance, Fleming and Soner (1993) and Yong and Zhou (1999). In that case, the (nonlinear) Hamilton–Jacobi–Bellmann deterministic equation is represented by a controlled SDE.

(iv) Another bridge between nonlinear PDEs and diffusions can be provided in the framework of interacting particle systems with chaos propagation, see Graham *et al.* (1996) for a survey on those problems. Among the most significant nonlinear PDEs investigated probabilistically, we quote the case of porous media equations. For instance, for a positive integer $m$, a solution to

$$\partial_t u = \tfrac{1}{2}\partial^2_{xx}(u^{2m+1}) \qquad [9]$$

can be represented by a (nonlinear) diffusion of the type, see Benachour *et al.* (1996),

$$\begin{aligned} \mathrm{d}X_t &= u^m(s, X_s)\,\mathrm{d}W_t \\ u(t,\cdot) &= \text{law density of } X_t \end{aligned} \qquad [10]$$

## Different Notions of Solutions

Let $a$ and $b$ as at the beginning of the previous section. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, a filtration $(\mathcal{F}_t)_{t\geq 0}$ fulfilling the usual conditions, an $(\mathcal{F}_t)_{t\geq 0}$-classical Brownian motion $(W_t)_{t\geq 0}$. Let $\xi$ be an $\mathcal{F}_0$-measurable r.v. In the section "Motivation and preliminaries," we defined the notion of solution of the following equation:

$$\begin{aligned} \mathrm{d}X_t &= b(t, X_t)\,\mathrm{d}t + a(t, X_t)\,\mathrm{d}W_t \\ X_0 &= \xi \end{aligned} \qquad [11]$$

This equation will be denoted by $E(a,b)$ (without initial condition). However, as we will see, the general concept of solution of an SDE is more sophisticated and subtle than in the deterministic case. We distinguish several variants of existence and uniqueness.

**Definition 4** (Strong existence). We will say that equation $E(a,b)$ admits strong existence if the following holds. Given any probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a filtration $(\mathcal{F}_t)_{t\geq 0}$, an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion $(W_t)_{t\geq 0}$, an $\mathcal{F}_0$-measurable and square-integrable r.v. $\xi$, there is a process $(X_t)_{t\geq 0}$ solution to $E(a,b)$ with $X_0 = \xi$ a.s.

**Definition 5** (Pathwise uniqueness). We will say that equation $E(a,b)$ admits pathwise uniqueness if the following property is fulfilled. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, a filtration $(\mathcal{F}_t)_{t\geq 0}$, an $(\mathcal{F}_t)_{t\geq 0}$ Brownian motion $(W_t)_{t\geq 0}$. If two processes $X, \tilde{X}$ are two solutions such that $X_0 = \tilde{X}_0$ a.s., then $X$ and $\tilde{X}$ coincide.

**Definition 6** (Existence in law or weak existence). Let $\nu$ be a probability law on $\mathbb{R}^d$. We will say that $E(a,b;\nu)$ admits weak existence if there is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a filtration $(\mathcal{F}_t)_{t\geq 0}$, an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion $(W_t)_{t\geq 0}$, and a process $(X_t)_{t\geq 0}$ solution of $E(a,b)$ with $\nu$ being the law of $X_0$.

We say that $E(a,b)$ admits weak existence if $E(a,b;\nu)$ admits weak existence for every $\nu$.

**Definition 7** (Uniqueness in law). Let $\nu$ be a probability law on $\mathbb{R}^d$. We say that $E(a,b;\nu)$ has a unique solution in law if the following holds. We consider an arbitrary probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a filtration $(\mathcal{F}_t)_{t\geq 0}$ on it; we consider also another probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ equipped with another filtration $(\tilde{\mathcal{F}}_t)_{t\geq 0}$; we consider an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion $(W_t)_{t\geq 0}$, and an $(\tilde{\mathcal{F}}_t)_{t\geq 0}$-Brownian motion $(\tilde{W}_t)_{t\geq 0}$; we suppose having a process $(X_t)_{t\geq 0}$ (resp. a process $(\tilde{X}_t)_{t\geq 0}$) solution of $E(a,b)$ on the first (resp. on the second) probability space such that both the law of $X_0$ and $\tilde{X}_0$ are identical to $\nu$. Then $X$ and $\tilde{X}$ must have the same law as r.v. with values in $E = C(\mathbb{R}_+)$ (or $C[0,T]$).

We say that $E(a,b)$ has a unique solution in law if $E(a,b;\nu)$ has a unique solution in law for every $\nu$.

There are important theorems which establish bridges among the preceding notions. One of the most celebrated is the following.

**Proposition 1** (Yamada–Watanabe). *Consider the equation* $E(a,b)$.

(i) *Pathwise uniqueness implies uniqueness in law.*

(ii) *Weak existence and pathwise uniqueness imply strong existence.*

*A version can be stated for $E(a,b;\nu)$ where $\nu$ is a fixed probability law.*

**Remark 3**

(i) If $a$ and $b$ are locally Lipschitz with linear growth, Theorem [1] implies that $E(a,b)$ admits strong existence and pathwise uniqueness.
(ii) If $a$ and $b$ are only locally Lipschitz, then pathwise uniqueness is fulfilled.

### Existence and Uniqueness in Law

A way to create weak solutions of $E(1,b)$ when $(t,x)\to b(t,x)$ is Borel with linear growth is the Girsanov theorem. Suppose $d=1$ for simplicity. Let us consider an $(\mathcal{F}_t)$-classical Brownian motion $(X_t)$. We set

$$W_t = X_t - \int_0^t b(s, X_s)\mathrm{d}s$$

Under some suitable probability $Q,(W_t)$ is an $(\mathcal{F}_t)$-classical Brownian motion. Therefore, $(X_t)$ provides a solution to $E(1,b;\delta_0)$.

We continue with an example where $E(a,b)$ does not admit pathwise uniqueness, even though it admits uniqueness in law.

**Example 1** We consider the stochastic equation

$$X_t = \int_0^t \mathrm{sign}(X_s)\mathrm{d}W_s \qquad [12]$$

with

$$\mathrm{sign}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}$$

It corresponds to $E(a,b;\delta_0)$ with $b=0$ and $a(x)=\mathrm{sign}(x)$.

If $(W_t)_{t\geq 0}$ is an $(\mathcal{F}_t)$-classical Brownian motion, then $(X_t)_{t\geq 0}$ is $(\mathcal{F}_t)_{t\geq 0}$-continuous local martingale vanishing at zero such that $[X]_t \equiv t$. According to Lévy characterization theorem stated earlier, $X$ is an $(\mathcal{F}_t)_{t\geq 0}$-classical Brownian motion. This shows in particular that $E(a,b;\delta_0)$ admits uniqueness in law. In the sequel, we will show that $E(a,b;\delta_0)$ also admits weak existence.

Let now $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, an $(\mathcal{F}_t)_{t\geq 0}$-classical Brownian motion with respect to a filtration and $(X_t)_{t\geq 0}$ such that [12] is verified. Then $\tilde{X}_t = -X_t$ can also be shown to be a solution. Therefore, $E(a,b;\delta_0)$ does not admit pathwise uniqueness.

We continue stating a result true in the multi-dimensional case.

**Proposition 3** (Stroock–Varadhan). *Let $\nu$ be a probability on $\mathbb{R}^d$ such that*

$$\int_{\mathbb{R}} \|x\|^{2m}\nu(\mathrm{d}x) < +\infty \qquad [13]$$

*for a certain $m>1$. We suppose that $a,b$ are continuous with linear growth. Then $E(a,b;\nu)$ admits weak existence.*

From now on, a function $\gamma:[0,T]\times\mathbb{R}^m\to\mathbb{R}^d$ will be said Hölder-continuous if it is Hölder-continuous in the space variable $x\in\mathbb{R}^m$ uniformly with respect to the time variable $t\in[0,T]$.

Stroock and Varadhan (1979) also provide the following result, which is an easy consequence of their theorem 7.2.1.

**Proposition 4** *We suppose $a,b$ both Hölder-continuous, bounded such that condition; [8] is fulfilled. Then SDE $E(a,b;\nu)$ admits weak uniqueness.*

**Remark 4**

(i) The Hölder condition and [8] in Proposition 4 may be relaxed and replaced with the solvability of a Cauchy problem of a parabolic PDE with suitable terminal value.
(ii) In the case $d=1$, if $a,b$ are bounded and just Borel with [8] for $x$ on each compact, then $E(a,b;\nu)$ admits weak existence and uniqueness in law. See Stroock and Varadhan (1979, exercises 7.3.2 and 7.3.3).
(iii) If $d=2$, the same holds as at previous point provided that moreover $a$ does not depend on time.

We proceed with some more specifically unidimensional material stating some results from K J Engelbert and W Schmidt, who furnished necessary and sufficient conditions for weak existence and uniqueness in law of SDEs.

For a Borel function $\sigma:\mathbb{R}\to\mathbb{R}$, we first define

$$Z(\sigma) = \{x\in\mathbb{R} | \sigma(x)=0\}$$

then we define the set $I(\sigma)$ as the set of real numbers $x$ such that

$$\int_{x-\varepsilon}^{x+\varepsilon} \frac{\mathrm{d}y}{\sigma^2(y)} = \infty, \quad \forall\varepsilon>0$$

**Proposition 5** (*Engelbert–Schmidt criterion*). *Suppose that $a:\mathbb{R}\to\mathbb{R}$, that is, does not depend on time and we consider the equation without drift $E(a,0)$.*

(i) *$E(a,0)$ admits weak existence (without explosion) if and only if*

$$I(a) \subset Z(a) \qquad [14]$$

(ii) $E(a, 0)$ *admits weak existence and uniqueness in law if and only if*

$$I(a) = Z(a) \quad [15]$$

**Remark 5**

(i) If $a$ is continuous then, [14] is always verified. Indeed, if $a(x) \neq 0$, there is $\varepsilon > 0$ such that

$$|a(y)| > 0, \quad \forall y \in [x - \varepsilon, x + \varepsilon]$$

Therefore, $x$ cannot belong to $I(a)$.

(ii) Equation [14] is verified also for some discontinuous functions as, for instance, $a(x) = \mathrm{sign}(x)$. This confirms what was affirmed previously, that is, the weak existence (and uniqueness in law) for $E(a, 0)$.

(iii) If $a(x) = 1_{\{0\}}(x)$, [14] is not verified.

(iv) If $a(x) = |x|^\alpha, \alpha \geq 1/2$, then

$$Z(a) = I(a) = \{0\}$$

So there is at most one solution in law for $E(a, 0)$.

(v) The proof is technical and makes use of Lévy characterisation theorem of Brownian motion.

### Results on Pathwise Uniqueness

**Proposition 6** (Yamada–Watanabe). *Let $a, b : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ and consider again $E(a, b)$. Suppose $b$ globally Lipschitz and $h : \mathbb{R}_+ \to \mathbb{R}_+$ strictly increasing continuous such that*

(i) $h(0) = 0$;

(ii) $\int_0^\varepsilon (1/h^2)(y)\mathrm{d}y = \infty, \quad \forall \varepsilon > 0$; and

(iii) $|a(t, x) - a(t, y)| \leq h(x - y)$.

*Then pathwise uniqueness is verified.*

**Remark 6**

(i) In Proposition 6, one typical choice is $h(u) = u^\alpha, \alpha > 1/2$.

(ii) Pathwise uniqueness for $E(a, b)$ holds therefore if $b$ is globally Lipschitz and $a$ is Hölder-continuous with parameter equal to 1/2.

**Corollary 1** *Suppose that the assumptions of Proposition 6 are verified and $a, b$ continuous with linear growth. Then $E(a, b; \nu)$ admits strong existence and pathwise uniqueness, whenever $\nu$ verifies condition [13].*

**Proof** It follows from Propositions 6 and 3 together with Proposition 1 (ii). □

**Remark 7** Suppose $d = 1$. Pathwise uniqueness for $E(a, b)$ also holds under the following assumptions.

(i) $a$, $b$ are bounded, $a$ is time independent and $a \geq \text{const.} > 0$, $h$ as in Proposition 6. This result has an analogous form in the case of spacetime white noise driven SPDEs of parabolic type, as proved by Bally, Gyongy, and Pardoux in 1994.

(ii) a independent on time, $b$ bounded and $a \geq \text{const.} > 0$; moreover, $|a(x) - a(y)|^2 \leq |f(y) - f(x)|$ and $f$ is increasing and bounded.

For illustration we provide some significant examples.

**Example 2**

$$X_t = \int_0^t |X_s|^\alpha \, \mathrm{d}W_s, \quad t \geq 0 \quad [16]$$

We set $a(x) = |x|^\alpha, 0 < \alpha < 1$. This is equation $E(a, 0)$ with $a(x) = |x|^\alpha$. According to Engelbert–Schmidt notations, we have $Z(a) = \{0\}$. Moreover

(i) If $\alpha \geq 1/2$, then $I(a) = \{0\}$.

(ii) If $\alpha < 1/2$ then $I(a) = \emptyset$.

Therefore, according to Proposition 5, $E(a, 0)$ admits weak existence. On the other hand, if $\alpha \geq 1/2$,

$$|x^\alpha - y^\alpha| \leq h(|x - y|) \quad [17]$$

where $h(z) = z^\alpha$. According to Proposition 6, [16] admits pathwise uniqueness and by Corollary [1], also strong existence. The unique solution is $X \equiv 0$.

If $\alpha < 1/2, X \equiv 0$ is always a solution. This is not the only one; even uniqueness in law is not true.

**Example 3** Let $a(x) = \sqrt{|x|}, b$ Lipschitz. Then $E(a, b)$ admits strong existence and pathwise uniqueness. In fact, $a$ is Hölder-continuous with parameter 1/2 and the second item of Remark 6 applies; so pathwise uniqueness holds. Strong existence is a consequence of Propositions 3 and 1 (ii).

An interesting particular case is provided by the following equation. Let $x_0, \sigma, \delta \geq 0, k \in \mathbb{R}$. The following equation admits strong existence and pathwise uniqueness.

$$Z_t = x_0 + \sigma \int_0^t \sqrt{|Z_s|} \, \mathrm{d}W_s + \int_0^t (\delta - kX_s) \, \mathrm{d}s \qquad t \in [0, T] \quad [18]$$

Equation [18] is widely used in mathematical finance and it constitutes the model of Cox–Ingersoll–Ross: the solution of the mentioned equation represents the short interest rate.

Consider now the particular case where $k = 0$, $\sigma = 2$. According to some comparison theorem for SDEs, the solution $Z$ is always non-negative and

therefore the absolute value may be omitted. The equation becomes

$$Z_t = x_0 + 2\int_0^t \sqrt{Z_s}\,\mathrm{d}W_s + \delta t \qquad [19]$$

**Definition 8** The unique solution $Z$ to

$$Z_t = x_0 + 2\int_0^t \sqrt{Z_s}\mathrm{d}W_s + \delta t \qquad [20]$$

is called "square $\delta$-dimensional Bessel process" starting at $x_0$; it is denoted by $\mathrm{BESQ}^\delta(x_0)$; for fine properties of this process, see Revuz and Yor (1999, ch. IX.3).

Since $Z \geq 0$, we call $\delta$-dimensional Bessel process starting from $x_0$ the process $X = \sqrt{Z}$. It is denoted by $\mathrm{BES}^\delta(x_0)$.

**Remark 8** Let $d \geq 1$. Let $W = (W^1, \ldots, W^d)$ be a classical $d$-dimensional Brownian motion. We set $X_t = \| W_t \|$. $(X_t)_{t\geq 0}$ is a $d$-dimensional Bessel process.

**Remark 9** If $\delta > 1$, it is possible to see that

$$X_t = W_t + \frac{\delta - 1}{2}\int_0^t \frac{\mathrm{d}s}{X_s}$$

### The Case with Distributional Drift

Pioneering work about diffusions with generalized drift was presented by N I Portenko, but in the framework of semimartingale processes. Recently, some work was done characterizing solutions in the class of the so-called Dirichlet processes, with some motivations in random irregular environment.

A useful transformation in the theory of SDE is the so-called "Zvonkin transformation." Let $(W_t)$ be an $(\mathcal{F}_t)$-classical Brownian motion. Let $a$ (resp. $b$) : $\mathbb{R} \to \mathbb{R}$ (resp. $C^1$) be locally bounded. We suppose moreover $a > 0$. We fix $x_0 \in \mathbb{R}$. Let $(X_t)_{t\geq 0}$ be a solution of

$$X_t = x_0 + \int_0^t b(X_s)\,\mathrm{d}s + \int_0^t a(X_s)\,\mathrm{d}W_s \qquad [21]$$

We set

$$\Sigma(x) = \int_0^x \frac{2b}{a^2}(y)\,\mathrm{d}y$$

and we define $h : \mathbb{R} \to \mathbb{R}$ such that

$$h(0) = 0, \qquad h' = \mathrm{e}^{-\Sigma}$$

$h$ is strictly increasing. We set $\tilde{a}(x) = (ah')(h^{-1}(x))$, where $h^{-1}$ is the inverse of $h$. We set $Y_t = h(X_t)$. Without entering into details, the classical Itô formula allows to show that $(Y_t)$ defines a solution of

$$\mathrm{d}Y_t = \tilde{a}(Y_t)\,\mathrm{d}W_t \qquad Y_0 = h(x_0) \qquad [22]$$

Now, eqn [22] fulfills the requirements of the Engelbert–Schmidt criterion so that it admits weak existence and uniqueness in law. Consequently, unless explosion, one can easily establish the same well-posedness for [21].

Zvonkin transformation also allows to prove strong existence and pathwise uniqueness results for [21]; for instance, when

- $a$ has linear growth, and
- $$y \to \int_0^y \frac{b(s)}{a^2(s)}\mathrm{d}s$$

  is a bounded function.

In fact, problem [22] satisfies pathwise uniqueness and strong existence since the coefficients are Lipschitz with linear growth. Therefore, one can deduce the same for [21].

Veretennikov generalized Zvonkin transformation to the $d$-dimensional case in some cases which include the case $a = 1$ and $b$ bounded Borel.

Zvonkin's procedure suggests also to consider a formal equation of the type

$$\mathrm{d}X_t = \mathrm{d}W_t + \gamma'(X_t)\,\mathrm{d}t \qquad [23]$$

where $\gamma$ is only a continuous function and so $b = \gamma'$ is a Schwartz distribution; $\gamma$ could be, for instance, the realization of an independent Brownian motion of $W$. Therefore, eqn [23] is motivated by the study of irregular random media. When $\sigma = 1, b = \gamma'$, SDE [22], $h' = \mathrm{e}^{-2\gamma}$ still makes sense.

Using the Engelbert–Schmidt criterion, one can see that problem [22] still admits weak existence and uniqueness in the sense of distribution laws. If $Y$ is a solution of [22], $X = h^{-1}(Y)$ provides a natural candidate solution for [21]. R F Bass, Z-Q Chen and F Flandoli, F Russo, and J Wolf investigated generalized SDEs as [23]: in particular, they made previous reasoning rigorous, respectively, in the case of strong and weak solutions, see Flandoli *et al.* (2003).

## Connected Topics

We aim here at giving some basic references about topics which are closely connected to SDEs.

### Stochastic Partial Differential Equations (SPDEs)

If a SDE is a random perturbation of an ordinary differential equation, an SPDE is a random

perturbation of a PDE. Several studies were performed in the parabolic (evolution equation) and hyperbolic case (wave equation). Most of the work was done in the case of a fixed underlying probability spaces. We only quote two basic monographies which should be consulted at first before getting into the subject: the one of Walsh (1986) and the one of Da Prato and Zabczyk (1992).

However, it was possible to establish some results about weak existence and uniqueness in law for SPDEs. One possible tool was a generalization of Girsanov theorem to the case of Gaussian spacetime white noise. Weak existence for the stochastic quantization equation was proved with the help of infinite-dimensional Dirichlet forms by S Albeverio and M Röckner.

We also indicate a beautiful recent monography by Da Prato (2004) which pays particular attention to Kolmogorov equations with infinitely many variables.

### Numerical Approximations

Relevant work was done in numerical approximation of solutions to SDEs and related approximations of solutions to linear parabolic equations via Feynman–Kac probabilistic representation, see Theorem 2). It seems that the stochastic simulations (of improved Monte Carlo type and related topics) for solving deterministic problems are efficient when the space dimension is greater than 4.

### Malliavin Calculus

Malliavin calculus is a wide topic (*see* Malliavin Calculus). Relevant applications of it concern stochastic (ordinary and partial) differential equations. We only quote a monography of Nualart (1995) on those applications. Two main objects were studied.

- Given a solution of an SDE, $(X_t)$, sufficient conditions so that $X_t, t > 0$, has a (smooth) density $p(t, \cdot)$. Small-time asymptotics of this density, when $t \to 0$, and small-drift perturbation were performed, refining Freidlin–Ventsell large-deviation estimates.
- Coming back to SDE [11], one can conceive to consider coefficients $a$, $b$ nonadapted with respect to the underlying filtration $(\mathcal{F}_t)$. On the other hand, the initial condition $\xi$ may be anticipating, that is, not $\mathcal{F}_0$-measurable. In that case, the Itô integral $\int_0^t a(s, X_s) \mathrm{d}W_s$ is not defined. A replacement tool is the so-called "Skorohod integral."

### Rough Paths Approach

A very successful and significant research field is the rough path theory. In the case of dimension $d = 1$, Doss–Sussmann method allows to transform the solution of an SDE into the solution of an ordinary (random) differential equation. In particular, that solution can be seen as depending (pathwise) continuously from the driving Brownian motion $(W_t)$ with respect to the usual topology of $C([0, T])$. Unless exceptions, this continuity does not hold in case of general dimension $d > 1$. Rough paths theory, introduced by T Lyons, allows to recover somehow this lack of continuity and establishes a true pathwise stochastic integration.

### SDEs Driven by Non-semimartingales

At the moment, there is a very intense activity towards SDEs driven by processes which are not semimartingales. In this perspective, we list SDEs driven by fractional Brownian motion with the help of rough paths theory, using fractional and Young type integrals and involving finite cubic variation processes. Among the contributors in that area we quote L Coutin, R Coviello, M Errami, M Gubinelli, Z Qian, F Russo, P Vallois, and M Zähle.

*See also:* Fractal Dimensions in Dynamics; Image Processing: Mathematics; Interacting Stochastic Particle Systems; Lagrangian Dispersion (Passive Scalar); Malliavin Calculus; Path Integrals in Noncommutative Geometry; Quantum Dynamical Semigroups; Quantum Fields with Indefinite Metric: Non-Trivial Models; Random Dynamical Systems; Random Walks in Random Environments; Stochastic Hydrodynamics; Stochastic Resonance.

## Further Reading


Bally V, Gyongy I, and Pardoux E (1994) White noise driven parabolic SPDEs with measurable drift. *Journal of Functional Analysis* 120: 484–510.

Benachour S, Chassaing P, Roynette B, and Vallois P (1996) Processus associés l' équation des milieux poreux. *Annali Scuola Normale Superiore di Pisa, Classe di Scienze* 23(4): 793–832.

Bouleau N and Lépingle D (1994) *Numerical Methods for Stochastic Processes.* New York: Wiley.

Da Prato G (2004) *Kolmogorov Equations for Stochastic PDEs.* Basel: Birkäuser.

Da Prato G and Zabczyk J (1992) *Stochastic Equations in Infinite Dimensions*, Encyclopedia of Mathematics and its Applications, vol. 44. Cambridge: Cambridge University Press.

Flandoli F, Russo F, and Wolf J (2003) Some stochastic differential equations with distributional drift. Part I. General calculus. *Osaka Journal of Mathematics* 40(2): 493–542.

Fleming WH and Soner M (1993) *Controlled Markov Processes and Viscosity Solutions.* New York: Springer.

Graham C, Kurtz Th G, Méléard S, Protter Ph E, Pulvirenti M, and Talay D (1996) Probabilistic models for nonlinear partial

differential equations. In: Talay and Tubaro L (eds.) *Lectures given at the 1st Session and Summer School held in Montecatini Terme*, 22–30 May 1995, Lecture notes in Mathematics, vol. 1627. Springer-Verlag. Centro Internazionale Matematico Estivo (C.I.M.E.), Florence, 1996.
Karatzas I and Shreve SE (1991) *Brownian Motion and Stochastic Calculus*, 2nd edn. New York: Springer.
Kloeden PE and Platen E (1992) *Numerical Solutions of Stochastic Differential Equations*. Berlin: Springer.
Lamberton D and Lapeyre B (1997) *Introduction au Calcul Stochastique et Applications la Finance*. Paris: Collection Ellipses.
Ma J and Yong J (1999) *Forward–Backward Stochastic Differential Equations and Their Applications*, Lecture Notes in Mathematics, vol. 1702. Berlin: Springer.
Nualart D (1995) *The Malliavin Calculus and Related Topics*. New York: Springer.
Øksendal B (2003) *Stochastic Differential Equations. An Introduction with Applications*. Sixth edition, Universitext. Berlin: Springer-Verlag.
Pardoux E (1998) *Backward stochastic differential equations and viscosity solutions of systems of semilinear parabolic and elliptic PDEs of second order*. Stochastic analysis and related topics, VI (Geilo, 1996), 79–127, Progr. Probab., 42, Boston: Birkäuser, 1998.
Protter P (1992) *Stochastic Integration and Differential Equations. A new approach*. Berlin: Springer.
Lyons T and Qian Z (2002) *Controlled Systems and Rough Paths*. Oxford: Oxford University Press.
Revuz D and Yor M (1999) *Continuous Martingales and Brownian Motion*, Third edition. Berlin: Springer-Verlag.
Stroock D and Varadhan SRS (1979) *Multidimensional Diffusion Processes*. Berlin–New York: Springer.
Walsh JB (1986) An introduction to stochastic partial differential equations. In: *Ecole d'été de probabilités de Saint–Flour, XIV – 1984*, pp. 265–439. Lecture Notes in Mathematics. vol. 118, Springer.
Yong J and Zhou XY (1999) *Stochastic Controls: Hamiltonian Systems and HJB Equations*. New York: Springer.

# Stochastic Hydrodynamics

**B Ferrario**, Università di Pavia, Pavia, Italy



## Introduction

Mathematical models in hydrodynamics are introduced to describe the motion of fluids. The basic equations for Newtonian incompressible fluids are the Euler and the Navier–Stokes equations, for inviscid and viscous fluids, respectively. For a given set of body forces acting on the fluid, these nonlinear partial differential equations (PDEs) model the evolution in time of the velocity and pressure at each point of the fluid, given the initial velocity and suitable boundary conditions (*see* Partial Differential Equations: Some Examples). The equations of hydrodynamics offer challenging mathematical problems, like proving the existence and uniqueness of solutions, determining their regularity, their asymptotic behavior for large time, and their stability. To gain some insight into the behavior of fluids, stochastic analysis is introduced into hydrodynamics. In fact, there are various attempts to describe turbulent regime (*see* Turbulence Theories). But, analyzing individual solutions that determine the flow at any time, for a given initial condition, is a desperate task, since the dynamics in a turbulent regime is chaotic and highly unstable. This is a particular chaotic motion with some characteristic statistical properties (see Monin and Yaglom (1987)). The aim of a statistical description of turbulent flow is to single out some relevant collective properties of the flow that, hopefully, make it possible to grasp the salient features of the dynamics. In this sense, stochastic hydrodynamics is germane to the kinetic gas theory. In the next section we shall review a typical topic of stochastic hydrodynamics, the evolution of probability measures. Results on stationary probability measures will be given in the subsequent sections.

Another characteristic of turbulent flows is the lack of space regularity of the velocity field. We shall introduce in the section "The stochastic Navier–Stokes equations" a stochastic model of turbulence, which exhibits lack of regularity of the solutions.

The Euler equations are a singular limit of the Navier–Stokes equations, since they are first order, instead of second-order PDEs. It is little surprise if they involve different mathematical techniques. A full section will be devoted to a discussion of Euler equations and another to the Navier–Stokes equations. Statistics of an inviscid flow, when approximated by vortex motion, will be described in the final section.

## Statistical Solutions

Let $u(t, x)$ be the fluid velocity at time $t$ and point $x \in D \subseteq \mathbb{R}^d$; since the initial velocity is always affected by experimental errors, it is reasonable to assign a measure $\nu$ determining the probability that the initial velocity belongs to a Borel set $\Gamma$ of the space $\mathbb{H}$ of all admissible velocity fields $u = u(x)$.

A spatial statistical solution is a family of probability measures $\mu(t, \cdot), t \geq 0$, each supported

on the set $\mathbb{H}$ such that, given any Borel set $\Gamma$ in $\mathbb{H}$, we have

$$\text{Prob}\{u(t,x) \in \Gamma\} = \mu(t,\Gamma), \quad \forall t > 0 \qquad [1]$$

with the initial condition $\mu(0,\Gamma) = \nu(\Gamma)$. The construction and analysis of statistical solutions $\mu(t,\cdot)$ is one of the crucial mathematical problems in stochastic hydrodynamics (see, e.g., Vishik and Fursikov (1988)).

Hopf gave the first mathematical formulation of the problem of describing turbulent flows by statistical solutions. The first result on the existence of statistical solutions is by Foias in 1973. Hopf (1952) presented an equation in variational derivatives satisfied by the characteristic functional $\chi(t,\phi)$ of the family of measures $\mu(t,\cdot)$ associated with the Navier–Stokes equations. The characteristic functional $\chi(t,\phi)$ is the Fourier transform of the measure $\mu(t,\cdot)$:

$$\chi(t,\phi) = \int_{\mathbb{H}} e^{i\langle\phi,u\rangle}\mu(t,du) \qquad [2]$$

defined for any smooth test function $\phi$.

We now derive the evolution equation for $\chi(t,\phi)$, by assuming that the dynamics takes place in the phase space $\mathbb{H}$ and follows the nonlinear equation

$$\frac{du}{dt} = F(u) \qquad [3]$$

If $u^{\nu}(t)$ is the solution started from $\nu$ at time $t=0$, then its probability distribution is represented by the time-evolved measure $\mu(t,\cdot)$. Therefore, we have that

$$\int_{\mathbb{H}} e^{i\langle\phi,u\rangle}\mu(t,du) = \int_{\mathbb{H}} e^{i\langle\phi,u^{\nu}(t)\rangle}\mu(0,d\nu) \qquad [4]$$

Differentiating in time, we obtain

$$\begin{aligned}\frac{d}{dt}\chi(t,\phi) &= \int_{\mathbb{H}} e^{i\langle\phi,u^{\nu}(t)\rangle} i\langle\phi,F(u^{\nu}(t))\rangle\mu(0,d\nu)\\ &= i\int_{\mathbb{H}} e^{i\langle\phi,\nu\rangle}\langle\phi,F(\nu)\rangle\mu(t,d\nu)\end{aligned} \qquad [5]$$

The last integral is uniquely determined by $\chi$, since the measure $\mu(t,\cdot)$ is uniquely determined by $\chi(t,\phi)$. We denote by $\Phi\chi(t,\phi)$ the last integral in [5]. The evolution equation thus obtained for the characteristic functional $\chi$ is

$$\frac{d}{dt}\chi(t,\phi) = i\Phi\chi(t,\phi), \quad \forall\phi \qquad [6]$$

This is called the Hopf equation associated with the dynamical system [3].

Another way to analyze the evolution of measures is through the moments; instead of the measure $\mu(t,\cdot)$ describing the spatial statistical solution, we deal with the moments of $\mu(t,\cdot)$ of any order. For a nonlinear dynamics [3], the moments equations are an infinite chain of coupled equations, the so-called Friedman–Keller equations.

A prominent role among statistical solutions is played by stationary solutions. They contain all the statistical information in the case of equilibrium in time. We have that the characteristic functional of an invariant measure is constant in time. Therefore,

$$\frac{d}{dt}\chi(t,\phi) = 0$$

Bearing in mind equation [5], this is equivalent to say that the signed measure $\langle\phi,F(\nu)\rangle\mu(t,d\nu)$ vanishes, for any test function $\phi$ and time $t$. Setting $t=0$, we obtain that an invariant measure $\nu$ in the space $\mathbb{H}$ satisfies the Liouville equation

$$\int_{\mathbb{H}} \langle\phi(\nu),F(\nu)\rangle\, d\nu(\nu) = 0 \qquad [7]$$

for appropriate test functions $\phi$. This equation is also called the relation of infinitesimal invariance and the measure $\nu$ is said to be infinitesimally invariant.

The stationary measures are natural candidates to describe the statistical asymptotic behavior of the system when $t\to\infty$. Notice that, in a chaotic system two motions that are arbitrarily close to one another at $t=0$ can evolve in completely different ways. So, to describe satisfactorily the dynamics we take average over a big number of experiments. This is the so-called ensemble average. These averages are assumed to be with respect to an invariant measure $\mu$. The invariant measures must exist and either they are unique or at most one has physical meaning and enters in the functional integral defining the ensemble average. According to the ergodic principle (an assumption not yet proved in hydrodynamics), ensemble averages replace long-time averages: for every initial velocity field $\nu$, except for a set of initial values negligible in some sense, the time average of an observable $\psi$ tends, as time goes to infinity, to the ensemble average

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T \psi(u^{\nu}(t))\, dt = \int_{\mathbb{H}} \psi\, d\mu \qquad [8]$$

However, it is extremely difficult to prove the existence of stationary probability measures for the Navier–Stokes equations solving directly equation [7]. The situation is formally the same as in equilibrium statistical mechanics, where the Liouville equation is in fact solved, leading to the Boltzmann–Gibbs distribution. However, the results in statistical hydrodynamics are far from being satisfactory.

Recent studies to prove the existence of invariant measures for the Navier–Stokes equations are based on stochastic models (see the section "The stochastic Navier–Stokes equations"). On the other hand, for the Euler equations it is possible to construct formally invariant measures, by means of invariant quantities of the classical motion (see the next section).

Finally, we point out that there are techniques using invariant measures to show some results for the time evolution (e.g., the motion exists for almost all initial values with respect to an invariant measure).

## The Euler Equations

We start recalling some basic facts on Euler equations (*see* Incompressible Euler Equations: Mathematical Theory).

The motion of an inviscid, incompressible, and homogeneous fluid is described by the Euler equations, which in Eulerian coordinates read as

$$\begin{aligned} &\frac{\partial u}{\partial t} + (u \cdot \nabla)u + \nabla p = f \\ &\nabla \cdot u = 0 \qquad\qquad \text{in } D \\ &u \cdot n = 0 \quad \text{on } \partial D \end{aligned} \qquad [9]$$

where, at time $t \geq 0$ and position $x \in D, u = u(t,x)$ is the vector velocity, $p = p(t,x)$ the hydrodynamic pressure. The units have been chosen so that the mass density $\rho = 1$. $\nabla$ denotes the nabla vector operator so

$$u \cdot \nabla = \sum_{j=1}^{d} u_j \frac{\partial}{\partial x_j}$$

$$\nabla \cdot u = \sum_{j=1}^{d} \frac{\partial u_j}{\partial x_j}$$

$$\nabla p = \left(\frac{\partial p}{\partial x_1}, \ldots, \frac{\partial p}{\partial x_d}\right)$$

Finally, $f$ denotes the external force. If the spatial domain $D$ has a boundary $\partial D$, then the velocity is assumed to be tangent to the boundary ($n$ denotes the exterior normal vector to the boundary). Some initial condition $u_0$ at time $t = 0$ is assigned.

When $f = 0$, there are invariant quantities for system [9]. In the literature, there are many works suggesting a Gaussian stationary statistics (see, e.g., the paper by Kraichnan (1980)). We consider invariants that are quadratic in the velocity so as to construct (formally) invariant measures of Gibbs type: the *energy*

$$\mathcal{E}(u) := \frac{1}{2}\int_D |u|^2 \, \mathrm{d}x$$

and, only in the two-dimensional case ($d = 2$), the *enstrophy*

$$\mathcal{S}(u) := \frac{1}{2}\int_D |\text{curl } u|^2 \, \mathrm{d}x$$

(with $\text{curl}\, u = \nabla^{\perp} \cdot u \equiv \partial u_2/\partial x_1 - \partial u_1/\partial x_2$ for $d = 2$).

It is natural to look for velocity fields in the following function spaces: the space $H^0$ of finite kinetic energy and the space $H^1$ of finite enstrophy. Clearly, the admissible fields should also obey the boundary conditions and divergence-free condition. If $P$ is the projection operator onto the space of divergence-free vectors, and $B$ is the bilinear form $B(u,v) := P[(u \cdot \nabla)v]$, the Euler equations can be given the structure of an evolution,

$$\frac{\mathrm{d}u}{\mathrm{d}t} = -B(u,u) \qquad [10]$$

obtained by applying the projection operator $P$ to the first equation in [9]. The pressure disappears and can be regarded as a Lagrange multiplier associated with the divergence-free constraint ($\nabla \cdot u = 0$); it can be fully recovered once the velocity field is known. The dynamics is considered in the phase space of divergence-free velocity vectors $\mathbb{H}$ (a large space containing $H^0$ and $H^1$), which is an infinite-dimensional functional space. More precisely, identifying $H^0$ with its dual $(H^0)'$, we introduce the Gelfand's triplet

$$H^1 \subset H^0 \subset (H^1)' = H^{-1}$$

The space $H^\alpha$, with $\alpha = 1, 2, \ldots,$ are the usual Sobolev spaces but with the additional divergence-free and boundary conditions. For $\alpha > 0$ noninteger, the spaces $H^\alpha$ are defined by interpolation, whereas those with $\alpha < 0$ by duality. As usual, regularity in space is related to the spaces $H^\alpha$ with higher exponent $\alpha$. We have that $\mathbb{H} = \cup_{\alpha \in \mathbb{R}} H^\alpha$.

Invariance of $\mathcal{E}$ and $\mathcal{S}$ can be proved resorting to eqn [9] and assuming that $u$ is a smooth vector field. For instance,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(u(t)) &= \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{2}\int_D |u|^2 \, \mathrm{d}x = \int_D u \cdot \frac{\partial u}{\partial t} \, \mathrm{d}x \\ &= -\int_D u \cdot [(u \cdot \nabla)u] \, \mathrm{d}x - \int_D u \cdot \nabla p \, \mathrm{d}x \end{aligned}$$

By integrating by parts and bearing in mind the divergence-free condition and the boundary condition, we conclude that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}(u) = 0$$

In the same way, the invariance of $\mathcal{S}$ can be proved.

As a consequence, the following Gibbs measures which are defined on the space $\mathbb{H}$

$$\mu_{\mathcal{E}}(\mathrm{d}u) = \frac{1}{Z_{\mathcal{E}}}\,\mathrm{e}^{-\mathcal{E}(u)}\,\mathrm{d}u \qquad \mu_{\mathcal{S}}(\mathrm{d}u) = \frac{1}{Z_{\mathcal{S}}}\,\mathrm{e}^{-\mathcal{S}(u)}\,\mathrm{d}u \qquad [11]$$

are heuristically invariant in time. In [11], $Z_{\cdot}$ are the partition functions, that is, they are normalization constants needed to guarantee that $\mu_{\mathcal{E}}$ and $\mu_{\mathcal{S}}$ are genuine probability measures (e.g., $Z_{\mathcal{E}} = \int_{\mathbb{H}} \mathrm{e}^{-\mathcal{E}(u)}\mathrm{d}u$).

Actually, these measures $\mu$ solve the Liouville equation

$$\int_{\mathbb{H}} \langle \phi(u), B(u,u)\rangle \mathrm{d}\mu(u) = 0 \qquad [12]$$

for any test function $\phi$, cylindrical, infinitely differentiable, bounded, and with bounded derivatives.

On the other hand, the (global and not only infinitesimal) invariance means that if there exists a global flow in time which is well defined in a phase space of full measure $\mu$, then the measure $\mu$ is invariant under this dynamics. The measures $\mu_{\mathcal{E}}$ and $\mu_{\mathcal{S}}$ are centered Gaussian measures whose support is in a space larger than $H^0$, as can be proved by standard methods in the theory of Gaussian measures on infinite-dimensional spaces. By the very definition, $\mu_{\mathcal{E}}$ is a cylindrical measure in $H^0$ and $\mu_{\mathcal{S}}$ is cylindrical in $H^1$. Then the support of $\mu_{\mathcal{E}}$ is any Hilbert space $\tilde{\mathbb{H}}$ such that $H^0 \subset \tilde{\mathbb{H}}$ is a Hilbert–Schmidt embedding, and the support of $\mu_{\mathcal{S}}$ is any space $\tilde{\mathbb{H}}$ such that $H^1 \subset \tilde{\mathbb{H}}$ is a Hilbert–Schmidt embedding. When the spatial dimension $d$ is 2, $\mathrm{supp}(\mu_{\mathcal{E}}) = \cap_{\alpha<-1} H^{\alpha}$ and $\mathrm{supp}(\mu_{\mathcal{S}}) = \cap_{\alpha<0} H^{\alpha}$. When $d$ is 3, $\mathrm{supp}(\mu_{\mathcal{E}}) = \cap_{\alpha<-3/2} H^{\alpha}$.

Moreover, $\mu_{\mathcal{E}}(H^0) = \mu_{\mathcal{S}}(H^0) = 0$, that is, the space of finite energy $H^0$ is negligible with respect to these measures. Let us show this property for the "enstrophy measure" $\mu_{\mathcal{S}}$ when $d = 2$. Let $\{e_j\}_{j=1}^{\infty}$ be a complete orthonormal system in $H^0$. Hence, for $u = \sum_j u_j e_j$, we have $\|u\|^2_{H^0} = \sum_j |u_j|^2$ and $\|u\|^2_{H^1} = \sum_j \lambda_j |u_j|^2$ (with $0 < \lambda_1 \leq \lambda_2 \leq \cdots$ and $\lambda_j \sim j$ as $j \to \infty$). Keeping in mind its definition, the measure $\mu_{\mathcal{S}}$ can be considered as a measure on the space of the sequences $\{u_j\}_j$ and written as an infinite product of one-dimensional centered Gaussian measures

$$\mu_{\mathcal{S}}(\mathrm{d}u) = \otimes_j \frac{1}{\sqrt{2\pi\lambda_j^{-1}}}\,\mathrm{e}^{-(\lambda_j/2)|u_j|^2}\,\mathrm{d}u_j \qquad [13]$$

The energy is

$$\mathcal{E}(u) = \frac{1}{2}\sum_j |u_j|^2$$

and the renormalized energy is

$$:\mathcal{E}:(u) = \frac{1}{2}\sum_j \left( |u_j|^2 - \int |u_j|^2 \mu_{\mathcal{S}}(\mathrm{d}u) \right)$$

Since, as can be easily shown $\int (:\mathcal{E}:(u))^2 \mu_{\mathcal{S}}(\mathrm{d}u) < \infty$, $:\mathcal{E}:(u)$ is finite for $\mu_{\mathcal{S}}$-almost every $u$. On the contrary, since $\sum_j \int |u_j|^2 \mu_{\mathcal{S}}(\mathrm{d}u) = \sum_j \lambda_j^{-1} = +\infty$, $\mathcal{E}(u)$ is infinite for $\mu_{\mathcal{S}}$-almost every $u$.

We also note in passing that, for any $\gamma > 0$ and $\beta > -\gamma$

$$\int_{\mathbb{H}} \mathrm{e}^{-\beta:\mathcal{E}:(u)} \frac{\mathrm{e}^{-\gamma\mathcal{S}(u)}}{Z}\,\mathrm{d}u < \infty$$

so that

$$\mu_{\mathcal{S}}^{(\beta),(\gamma)}(\mathrm{d}u) = \frac{\mathrm{e}^{-\beta:\mathcal{E}(u):-\gamma\mathcal{S}(u)}}{\int \mathrm{e}^{-\beta:\mathcal{E}(u):-\gamma\mathcal{S}(u)}\,\mathrm{d}u}\,\mathrm{d}u \qquad [14]$$

is a probability measure, which is infinitesimally invariant for the Euler flow.

Since the space of finite-energy velocity is negligible with respect to these measures, it is necessary to replace the classical solutions having finite energy with generalized solutions. This is not an easy task in the three-dimensional case, whereas some results have been proved for the two-dimensional problem, where the following existence result holds. Let us analyze the quadratic term $B(u,u) = -P[(u\cdot\nabla)u].(u\cdot\nabla)u$ can be rewritten as $\nabla(u\otimes u)$, taking in account the divergence-free condition. Trivially, we have that $\nabla(u\otimes u) = \nabla(u\otimes u - :u\otimes u:)$, where $:u\otimes u: = \int u\otimes u;\ \mu_{\mathcal{S}}(\mathrm{d}u)$. We consider the quadratic expression $(u\otimes u - :u\otimes u:)$. This is integrable with respect to the measure $\mu_{\mathcal{S}}$ in the sense that

$$\int \|u\otimes u - :u\otimes u:\|^2_{H^{-\varepsilon}}\,\mu_{\mathcal{S}}(\mathrm{d}u) < \infty \qquad [15]$$

for any $\varepsilon > 0$. We remark that this property is similar to the integrability of the renormalized energy, which is a quadratic expression as well. This implies that the $H^{-1-\varepsilon}$-norm of $\nabla(u\otimes u)$ is integrable with respect to the measure $\mu_{\mathcal{S}}$. Therefore, $B(u,u)$ is defined for $\mu_{\mathcal{S}}$-a.e. $u$.

Now, let us replace eqn [10] with a system of infinite equations for all the components $u_j$ with respect to the orthonormal basis $\{e_j\}_j$, obtained by taking the scalar product with $e_j$ of both sides of eqn [10]:

$$\frac{\mathrm{d}u_j}{\mathrm{d}t} = B_j(u,u), \quad j = 1,2,\ldots \qquad [16]$$

Each component $B_j(u,u)$ is defined for $\mu_S$-a.e. $u$. These estimates lead to define a weak solution (see Albeverio and Cruzeiro (1990)):

**Theorem 1** *Let $d=2$. There exists a flow $U(t,\omega)$ defined on a probability space $(\Omega,\mathcal{F},P)$ with values in $H^{-\varepsilon-1}$ for any $\varepsilon>0$, $U(\cdot,\omega)\in C(\mathbb{R},H^{-\varepsilon-1})$ P-a.e. $\omega$, such that for each component $U_j$ we have*

$$\begin{aligned} &U_j(t,\omega)\\ &\quad = U_j(0,\omega)+\int_0^t B_j(U(s,\omega),U(s,\omega))\,\mathrm{d}s,\\ &\qquad P-a.e.\omega, \qquad \forall t\in\mathbb{R} \end{aligned}$$

*Moreover, the measure $\mu_S$ is invariant under this flow.*

We point out that uniqueness is an open problem also for $d=2$. But already in the classical analysis of the Euler equations in a bounded domain, uniqueness for initial velocity of finite energy is not known. Working with the measure $\mu_{\mathcal{E}}$ is even worse, especially when $d=3$, because its support is a larger space within which more irregular velocity vectors live. The more irregular the spaces where the flow lives, the more difficult is to handle the nonlinear term $B(u,u)$.

On the other hand, for $d=1$, the mathematical analysis is much easier. For instance, it can be proved (see Robert (2003)) that the one-dimensional inviscid Burgers equation on the line

$$\frac{\partial u}{\partial t}+\frac{\partial}{\partial x}\left(\frac{1}{2}u^2\right)=0 \qquad [17]$$

has intrinsic invariant statistical solution, given by a class of Lévy's processes with negative jumps.

## The Stochastic Navier–Stokes Equations

The Navier–Stokes equations describe advection with velocity $u$ and diffusion with kinematic viscosity $\nu>0$ (*see* Viscous Incompressible Fluids: Mathematical Theory)

$$\left.\begin{aligned} &\frac{\partial u}{\partial t}-\nu\Delta u+(u\cdot\nabla)u+\nabla p=f\\ &\nabla\cdot u=0\\ &u=0 \quad \text{on}\,\partial D \end{aligned}\right. \quad \text{in}\, D \qquad [18]$$

where $\Delta$ is the Laplace operator. Nonslip boundary conditions are assumed. Although the Euler equations [9] are formally obtained from [18] by setting $\nu=0$, the presence of the second-order operator $-\nu\Delta$ makes the analysis needed to prove the existence, uniqueness, and regularity of solutions easier than for the Euler equations. However, at variance with the Euler equations, the Navier–Stokes equations do not possess invariants, since the viscosity dissipates energy. Hence, it is difficult to find explicit expressions of invariant measures for the deterministic Navier–Stokes equations, except the trivial invariant measures concentrated on a stationary solution. However, as soon as a stochastic force is introduced in these equations, it is possible to have nontrivial invariant measures. It is impossible to review here the wide literature concerning the stochastic Navier–Stokes equations and we confine ourselves to make some remarks. Most results are concerned with proving the existence and/or uniqueness of an invariant measure $\mu$, without giving an explicit representation, apart some attempts like Gallavotti (2002), where a formal representation of stationary distributions is given in terms of functional integrals. Some properties of the not explicit invariant measures are given like, for instance, estimates of moments, exponential convergence of the statistical solution for large time.

Stochastic forces can enter in the Navier–Stokes equations in different ways. We can consider randomness in the forcing term, so that the force $f$ in [18] has a deterministic component which represents its mean varying slowly and a stochastic one, which accounts small fluctuations around the mean and varying very rapidly. Alternatively, since the molecules are not rigidly connected to one another in the fluid, they are subjected to fluctuations. A complete description of fluctuations relating the microscopic and macroscopic motion is not achieved at present. However, we shall introduce some models for which rigorous mathematical results can be proved.

The first part of this section concerns the Navier–Stokes equations with noise $n$:

$$\begin{aligned} &\frac{\partial u}{\partial t}-\nu\Delta u+(u\cdot\nabla)u+\nabla p=n\\ &\nabla\cdot u=0 \end{aligned} \qquad [19]$$

for which invariant measures exist, one of which can be ergodic provided that the noise is suitably chosen. In the second part, a Navier–Stokes-type stochastic system is described, which has irregular solutions, as expected in turbulence.

Let us introduce the stochastic Navier–Stokes equations with time white noise. The first equation in [19] is an Itô equation:

$$\partial_t u+[-\nu\Delta u+(u\cdot\nabla)u+\nabla p]=\partial_t w \qquad [20]$$

Here $w=w_{(1)},\ldots,w_{(d)}$ is a Brownian motion, that is, its time derivative $n=\partial w/\partial t$ is a Gaussian

stochastic field with zero mean and correlation function given by

$$\mathbb{E}[n_{(j)}(t,x)n_{(k)}(t',x')] = \delta_{jk}q(x-x')\delta(t-t') \qquad [21]$$

for $j,k=1,\ldots,d$.

We shall use the differential form for the Itô equation [20] always understood in the integral form

$$u(t)-u(0)+\int_0^t[-\nu\Delta u(s)+(u(s)\cdot\nabla)u(s) + \nabla p(s)]\mathrm{d}s = w(t) \qquad [22]$$

Modeling perturbations by a white noise process represents the first step to understand how a random perturbation acts in the mathematical equations, rather than a good physical or numerical model. The first results are in a paper by Bensoussan and Temam (1973).

Obviously, the regularity of the solutions depends on the spatial covariance $q$ of the noise.

Let us consider the following cases.

- $q=\delta$: the noise is white also in space.

An invariant measure is known explicitly. Indeed, assume periodic boundary conditions on the square ($d=2$) or the cube ($d=3$) $D$, which makes the spatial domain a torus. In this case, the Euler and Navier–Stokes equations are set in the same functional spaces. The generator of the stochastic Navier–Stokes equations [20] corresponds to the sum of the generator of the Euler equations [9] and of the stochastic Stokes equations

$$\begin{aligned}\partial_t u &= [\nu\Delta u - \nabla p] + \partial_t w \\ \nabla\cdot u &= 0\end{aligned} \qquad [23]$$

Since the first equation in [23] is linear in the unknown velocity $u$, the Stokes system has a unique invariant measure which is a centered Gaussian measure. In particular, when the noise is a space-time white noise and $d=2$, this is the invariant measure [14] of the enstrophy:

$$\mu_S^{(0),(2\nu)}(\mathrm{d}u) = \frac{1}{Z}\mathrm{e}^{-2\nu S(u)}\mathrm{d}u$$

On a bidimensional torus, it is proved that this measure is not only infinitesimally invariant, but also globally invariant for a unique flow [20] defined for $\mu_S^{(0),(2\nu)}$-a.e. initial velocity. We recall that initial velocities of finite energy are negligible with respect to the measure $\mu_S^{(0),(2\nu)}$.

- $q$ more regular than above, that is, the noise is colored in space.

As soon as the forcing term is more regular in space, the Navier–Stokes system has a solution of finite energy. These are solutions close to those of the deterministic equation. Techniques similar to those used to prove the existence and/or uniqueness of solutions for the deterministic equations work also in the stochastic case with an additive noise (or even a multiplicative noise) to get weak or strong solutions. Global existence in the space $H^0$ is proved for $d=2,3$ and uniqueness only for $d=2$, as is the case for the deterministic Navier–Stokes equations.

The interesting feature is that by adding a noise which acts on all the components with respect to a Hilbert basis (or at least on many components), the stochastic Navier–Stokes system has a unique invariant measure, which is ergodic. This is proved for the spatial dimension $d=2$. By means of the Krylov–Bogoliubov's method, existence of at least an invariant measure is proved by compactness of a family of averaged measures; the limit measures are stationary measures. But, when many modes are perturbed by a noise, there is a mixing effect on the dynamics, avoiding existence of many stationary measures. For the spatial dimension $d=2$, the best result in this context is in Hairer and Mattingly (2004), where the noise acts on very few modes. For the spatial dimension $d=3$, the result in Da Prato and Debussche (2003) shows the existence of an invariant measure; even if there is no uniqueness of the solutions (as in the deterministic case), by a selection principle, they construct a transition semigroup, which has a unique invariant measure, ergodic and strongly mixing.

Mathematical proofs are given for very different noises. (The reader is urged to consult, among the others, the papers by E, Mattingly and Sinai; Flandoli and Maslowski; Mikulevicius and Rozovskii; Vishik and Fursikov. The latter authors study also statistical solutions in two and three dimensions. For a kick noise $n=\sum_k\delta(t-k)q_k(x)$ in equations [19], there are results for $d=2$ by Bricmont, Kupiainen and Lefevere; Kuksin and Shirikyan.)

We conclude that, as far as invariant measures and their ergodicity are concerned, the stochastic Navier–Stokes equations have richer results than the deterministic Navier–Stokes equations. It is appealing to investigate the limit as the intensity of the noise goes to zero, so as to recover the deterministic equation. Now, think of equation [19] with a noise $\varepsilon n$, for $n$ fixed and $\varepsilon\to 0$. Due to the sensitive dependence on initial conditions, even a small noise may have important effects on the dynamics. A conjecture by Kolmogorov is that the unique invariant measure $\mu_\varepsilon$ tends, when $\varepsilon\to 0$, to a specific measure, the so-called Kolmogorov measure, which

would enter into the ergodic principle. This is a difficult problem, not yet solved.

We also mention the analysis of the inviscid limit. Kuksin (2004) showed that the solution $u_\nu$ of the two-dimensional stochastic Navier–Stokes equations

$$\frac{\partial u}{\partial t} - \nu\Delta u + (u \cdot \nabla)u + \nabla p = \sqrt{\nu}n, \quad 0 < \nu \leq 1 \quad [24]$$

on the torus converges in distribution to a stationary solution of the Euler equations. Here $n$ is a random force white in time and smooth in space. More precisely, for each subsequence $u_{\nu_j}$,

$$\lim_{\nu_j \to 0} \lim_{T \to \infty} u_{\nu_j}(T + t) = U(t) \quad [25]$$

and almost every trajectory of the nontrivial limit process $U$ solves the Euler equations [9] without the forcing term. Moreover, the process $U$ keeps memory of some features of the noise force $n$, since the mean values of the enstrophy and of the energy of $U$ depend on the noise $n$.

We now present the second part on stochastic models for viscous fluids. In his 1884 paper, Reynolds introduced the decomposition of turbulent flow into mean and fluctuating flows. The equations obtained are difficult to study. We shall show now a tractable model for a one-dimensional problem ($d = 1$) with a suitable model of fluctuations. Decompose the velocity field into the sum of a mean flow $\overline{u}$ and a fluctuation $\delta$

$$u = \overline{u} + \delta$$

The fluctuation is assumed to be highly irregular; it is reasonable to model it by a stochastic process. If we choose

$$\delta = b\frac{\mathrm{d}w}{\mathrm{d}t}$$

where $b$ is a given velocity field and $\mathrm{d}w/\mathrm{d}t$ is white noise, then the motion of the fluid is governed by a stochastic equation of Itô type. Indeed, the Navier–Stokes equations are balance equations of linear momentum:

$$\frac{Du}{Dt} = \nu\Delta u - \nabla p \quad [26]$$

where $Du/Dt$ is the material time derivative along the trajectory of a particle which is at time $t$ in position $x(t)$ moving with velocity $u$ (so $u(x(t)) = (\mathrm{d}x/\mathrm{d}t)(t)$):

$$\frac{Du}{Dt} = \frac{\mathrm{d}}{\mathrm{d}t}u(t, x(t)) = \frac{\partial u}{\partial t} + (u \cdot \nabla)u \quad [27]$$

According to the mathematical model for the fluctuation, we have

$$\mathrm{d}x(t) = \overline{u}(t, x(t))\mathrm{d}t + b(x(t))\mathrm{d}w(t) \quad [28]$$

Therefore, $D\overline{u}$ is computed by means of Itô's formula

$$D\overline{u}(t, x(t)) = \frac{\partial \overline{u}}{\partial t}\mathrm{d}t + \sum_{k=1}^{d} \frac{\partial \overline{u}}{\partial x_k}\mathrm{d}x_k(t) + \frac{1}{2}\sum_{k,s=1}^{d} \frac{\partial^2 \overline{u}}{\partial x_k \partial x_s} b_k b_s \mathrm{d}t \quad [29]$$

This leads to the stochastic Navier–Stokes-type equations (we neglect the overline symbol)

$$\begin{aligned} &\mathrm{d}_t u + [-\nu\Delta u + (u \cdot \nabla)u + \nabla p + \tfrac{1}{2}Qu]\,\mathrm{d}t \\ &\quad = -(b \cdot \nabla)u\,\mathrm{d}w(t) \quad [30] \\ &\nabla \cdot u = 0 \end{aligned}$$

where $Q$ is the second-order differential operator given by the last term in [29].

Rigorous mathematical results for the above equations have been proved for the one-dimensional case, that is, the Burgers equations on the line. Given an initial velocity of finite energy $u_0 \in H^0$, there exists a unique solution $u \in C([0, T]; H^0) \cap L^2(0, T; H^1)$ (P-a.s.). But it can be shown that for a more regular initial velocity there is no higher regularity of the solution of eqn [30], if $b \neq 0$. This means that these stochastic Burgers equations cannot have too regular nontrivial solutions, as expected in turbulent motion.

## Statistics of Vortices and Bidimensional Turbulence

Onsager (1949) proposed to investigate bidimensional turbulent flows, extending in a rigorous way to hydrodynamics the statistical mechanics approach of Boltzmann. If we are interested in flows of finite energy, the results of the section "The Euler equations" provide no answer to the problem. Another way to proceed is by approximating the Euler equations in a suitable way. Actually, in a two-dimensional turbulent flow, there appears a large-scale organization leading to coherent structures. These are hydrodynamical vortices, whose dynamics is governed by the Euler equations. Onsager suggested to approximate the continuous Euler equations by a great (but finite) number of point vortices. This leads to a finite-dimensional Hamiltonian system, to which the methods of statistical mechanics can be successfully applied. Of course, the crucial point is to pass to the limit, to

recover the continuous system. But there are many different ways to approximate a continuous vorticity by a cloud of point vortices and different approximations may lead to very different statistical equilibrium states.

We present here the approach presented in Lions (1997). To get an idea of a completely different approximation, see, for example, Robert (2003).

Let $D$ be a bounded open smooth simply connected subset of $\mathbb{R}^2$. Then there exists a function $\psi$ (the stream function) such that $u = \nabla^{\perp}\psi$ and $\psi|_{\partial D} = 0$. Given the velocity $u$, we recover the stream function by means of the vorticity $\omega = \operatorname{curl} u = -\Delta\psi$, so $\psi(x) = \int_D g(x,y)\omega(y)\,dy$ (here $g$ is the Green's function of the Laplacian $-\Delta$ and $x$, $y$ are points in $D$). The Euler equations can be written as

$$\frac{\partial\omega}{\partial t} + u \cdot \nabla\omega = 0 \qquad \omega = \operatorname{curl} u \quad [31]$$

Consider now a solution given by vorticity concentrated in a finite number $N$ of points:

$$\omega = \sum_{i=1}^{N} \lambda_i \delta_{x_i(t)} \quad [32]$$

Here the vortex intensities $\lambda_i$ are real values and $x_i(t)$ are distinct points in $D$ for $i = 1, \ldots, N$.

According to the Euler equations, these points evolve as follows (see also Marchioro and Pulvirenti (1994)):

$$\frac{d}{dt}x_j(t) = \nabla^{\perp}_{x_j} \sum_{l=1, l\neq j}^{N} \lambda_l\, g(x_j(t), x_l(t)) + \lambda_j \nabla^{\perp}_{x_j} \tilde{g}(x_j), \quad j = 1, \ldots, N \quad [33]$$

where $\tilde{g}$ is related to the Green's function $g$. This is a Hamiltonian system in $D^N$. Hereafter, we shall suppose that the vortex intensities are the same ($\lambda_i = \lambda\, \forall i$), so that the Hamiltonian is

$$\mathcal{H}(x_1, \ldots, x_N) = \frac{1}{2} \sum_{l,j=1, l\neq j}^{N} g(x_j, x_l) + \sum_{j=1}^{N} \tilde{g}(x_j) \quad [34]$$

By means of $\mathcal{H}$, we define the canonical Gibbs measure

$$\mu^N(dx_1\, dx_2 \cdots dx_N) = \frac{1}{Z(N)}\, e^{-\tilde{\beta}\lambda\mathcal{H}(x_1,\ldots,x_N)} dx_1\, dx_2 \cdots dx_N \quad [35]$$

where $Z(N)$ is the partition function. If $Z(N) < \infty$, then $\mu^N$ is a well-defined probability measure on $D^N$ and, by construction, it is an invariant measure for system [33]. We can prove that $Z(N)$ is finite for $\tilde{\beta}\lambda \in (-8\pi/N, 4\pi)$, so that it is natural to choose as a scaling $\tilde{\beta}\lambda N = \beta$. Hence,

$$\mu^N(dx_1\, dx_2 \cdots dx_N) = \frac{1}{Z(N)}\, e^{-(\beta/N)\mathcal{H}}\, dx_1\, dx_2 \cdots dx_N \quad [36]$$

is considered for $-8\pi < \beta \leq 0$, or $\beta > 0$ with $N > \beta/4\pi$.

Bearing in mind the Onsager approach to approximate the turbulent Euler motion by means of point vortices, we are interested in the limit as $N$ goes to $+\infty$, for $\beta$ fixed in $(-8\pi, +\infty)$. It turns out that, when the number of point vortices becomes very large, their statistical behavior corresponds to a very large number of independent particles moving in a mean force field that they create.

More precisely, consider $\lambda = 1/N, \tilde{\beta} = \beta$. The empirical measure

$$\frac{1}{N}\sum_{i=1}^{N} \delta_{x_i(t)}$$

describing the vorticity, weakly converges to a probability density $\rho$ and each correlation function

$$\rho_j^N(x_1, \cdots, x_j) = \int_D dx_{j+1} \cdots \int_D dx_N \frac{1}{Z(N)}\, e^{-(\beta/N)\mathcal{H}} \quad \text{for } j = 1, \ldots, N-1 \quad [37]$$

weakly converges to $\otimes_{i=1}^{j}\rho = \prod_{i=1}^{j} \rho(x_i)$.

The equation satisfied by $\rho$, also called the mean-field equation, is

$$\rho(x) = \frac{e^{-\beta U(x)}}{\int_D e^{-\beta U(y)}\, dy}, \quad \text{with } U(x) = \int_D g(x,y)\rho(y)\, dy \quad [38]$$

The relation between $U$ and $\rho$ can also be written as $-\Delta U = \rho$ in $D$, $U = 0$ on $\partial D$. We point out that $u = \nabla^{\perp} U$ is a stationary solution of the Euler equations. Indeed, $\omega = -\Delta U = \rho$ and $\rho$ is a function of $U$, let us say $\rho = F(U)$. This gives that $\nabla\omega = \nabla U F'(U)$ and thus the term $u \cdot \nabla\omega$ in the Euler equation [31] vanishes.

It can be proved that there exists a solution of the mean-field equation when $\beta \geq 0$ or when $\beta < 0$ and $D$ is simply connected. Uniqueness is known in some cases, for instance, when $D$ is a bounded open smooth simply connected domain and the velocity is assumed tangent to the boundary.

There are numerical evidences of this approximation approach (see references in Lions (1997) referring to the periodic case). They show that for

large time and large Reynolds number (viscosity $\nu$ close to 0), the vorticity of the solution of the Navier–Stokes equations appears in a simple and organized structure. This stays intact until the viscous dissipation damps it. The important observation is that the organized structure is described quite precisely by the solution of the mean-field equation for some specific $\beta$.

Actually, to say that a fluid is inviscid is an approximation (which may be justified in many contexts), since every fluid displays some kind of viscosity. But turbulence is a phenomenon occurring at very small viscosity. In this sense, the above result provides a description of stationary regime in an ideal fluid, which is a good approximation of some numerical simulations of real fluids. Besides this good agreement with numerical simulations, there is no proof on how to deduce the mean-field equation from the Euler equations (e.g., which parameter $\beta$ has to be chosen in eqn [38]?).

**Remark** The extension of this analysis to three-dimensional flows involves vortex filaments, instead of point vortices. There are attempts to describe interacting vortex filaments as proposed by Chorin. Idealizations of behavior of vortices are introduced to have a tractable mathematical model. The reader is referred to Lions (1997) for a description of nearly parallel vortex filaments and to Flandoli and Bessaih (2003) for more realistic filaments which fold.

*See also:* Cauchy Problem for Burgers-Type Equations; Hamiltonian Fluid Dynamics; Incompressible Euler Equations: Mathematical Theory; Malliavin Calculus; Non-Newtonian Fluids; Partial Differential Equations: Some Examples; Stochastic Differential Equations; Turbulence Theories; Viscous Incompressible Fluids: Mathematical Theory; Vortex Dynamics.

## Further Reading

Albeverio S and Cruzeiro AB (1990) Global flows with invariant (Gibbs) measures for Euler and Navier–Stokes two dimensional fluids. *Communications in Mathematical Physics* 129: 431–444.

Bensoussan A and Temam R (1973) Équations stochastiques du type Navier–Stokes. *Journal of Functional Analysis* 13: 195–222.

Bricmont J, Kupiainen A, and Lefevere R (2001) Ergodicity of the 2D Navier–Stokes equations with random forcing. *Communications in Mathematical Physics* 224(1): 65–81.

Da Prato G and Debussche A (2003) Ergodicity for the 3D stochastic Navier–Stokes equations. *Journal de Mathématiques Pures et Appliquées* 82(8): 877–947.

E Weinan, Mattingly JC, and Sinai Y (2001) Gibbsian dynamics and ergodicity for the stochastically forced Navier–Stokes equation. *Communications in Mathematical Physics* 224(1): 83–106.

Flandoli F and Bessaih H (2003) A mean field result for 3D vortex filaments. In: Davies IM, Jacob N, Truman A, Hassan O, Morgan K, and Weatherill NP (eds.) *Probabilistic Methods in Fluids*, 22–34. River Edge, NJ: World Scientific.

Flandoli F and Maslowski B (1995) Ergodicity of the 2-D Navier–Stokes equation under random perturbations. *Communications in Mathematical Physics* 172(1): 119–141.

Frisch U (1995) *Turbulence. The legacy of A. N. Kolmogorov*. Cambridge: Cambridge University Press.

Gallavotti G (2002) *Foundations of Fluid Dynamics*. Berlin: Springer.

Hairer M and Mattingly JC (2004) Ergodic properties of highly degenerate 2D stochastic Navier–Stokes equations (English. English, French summary). *Comptes Rendus Mathématique. Académie des Sciences. Paris* 339(12): 879–888 (see also the paper "Ergodicity of the 2D Navier–Stokes equations with degenerate stochastic forcing" to appear on *Annals of Mathematics*).

Hopf E (1952) Statistical hydromechanics and functional calculus. *Journal of Rational Mechanics and Analysis* 1: 87–123.

Kraichnan RH and Montgomery D (1980) Two–dimensional turbulence. *Reports on Progress in Physics* 43(5): 547–619.

Kuksin SB (2004) The Eulerian limit for 2D statistical hydrodynamics. *Journal of Statistical Physics* 115(1/2): 469–492.

Kuksin S and Shirikyan A (2000) Stochastic dissipative PDEs and Gibbs measures. *Communications in Mathematical Physics* 213(2): 291–330.

Lions P-L (1997) *On Euler Equations and Statistical Physics*, Pubblicazione della Scuola Normale Superiore. Pisa: Cattedra Galileiana.

Marchioro C and Pulvirenti M (1994) Mathematical Theory of Incompressible Nonviscous Fluids. *Applied Mathematical Science*, vol. 96. New York: Springer.

Mikulevicius R and Rozovskii BL (2004) Stochastic Navier–Stokes equations for turbulent flows. *SIAM Journal on Mathematical Analysis* 35(5): 1250–1310.

Monin AS and Yaglom AM (1987) *Statistical Fluid Mechanics: Mechanics of Turbulence*.Cambridge, MA: MIT Press.

Onsager L (1949) Statistical hydrodynamics. *Nuovo Cimento* 6(suppl. 2): 279–287.

Robert R (2003) Statistical hydrodynamics (Onsager revisited). In: Friedlander S and Serre D (eds.) *Handbook of Mathematical Fluid Dynamics*, vol. II, pp. 1–54. Amsterdam: North-Holland.

Vishik MJ and Fursikov AV (1988) *Mathematical Problems of Statistical Hydromechanics*. Dordrecht: Kluwer Academic Publishers.

# Stochastic Loewner Evolutions

**G F Lawler**, Cornell University, Ithaca, NY, USA



## Introduction

The stochastic Loewner evolution or Schramm–Loewner evolution (SLE) is a family of random curves that appear as scaling limits of curves or cluster boundaries of discrete statistical mechanical models in two dimensions at criticality. The stochastic Loewner evolution was introduced by Oded Schramm as a candidate for the limit of loop-erased random walk and the boundary of percolation clusters, and it is now believed that SLE curves appear in most planar critical systems whose scaling limit satisfies conformal invariance. The curves are defined by solving a Loewner differential equation with a random input.

## Definition

There are three major one-parameter families of SLE curves – chordal, radial, and whole-plane – which correspond to curves connecting two boundary points in a domain, a boundary point and an interior point in a domain, and two points in $\mathbb{C}$, respectively. The parameter is usually denoted $\kappa > 0$. The starting point for defining SLE is to write down the assumptions that one expects from a scaling limit, assuming that the limit is conformally invariant.

In the chordal case, we assume that there is a family of probability measures $\{\mu_D(z,w)\}$, indexed by simply connected proper domains $D \subset \mathbb{C}$ and distinct boundary points $z, w \in \partial D$, supported on continuous curves $\gamma:[0,t_\gamma] \to \bar{D}$ with $\gamma(0)=z$, $\gamma(t_\gamma)=w$, which satisfies the following:

- *Conformal invariance*. If $f: D \to D'$ is a conformal transformation, then the image of $\mu_D(z,w)$ under $f$ is the same as $\mu_{D'}(f(z), f(w))$, up to a time change.
- *Conformal Markov property for* $\mu_D(z,w)$. Suppose $\gamma[0,t]$ is known, and let $g_t$ be a conformal transformation of the slit domain $D \backslash \gamma[0,t]$ onto $D$ with $g_t(\gamma(t)) = z, g_t(w) = w$ (see Figure 1). Then the conditional distribution on $g_t \circ \gamma[t, t_\gamma]$, given $\gamma[0,t]$, is the same, up to a change of parametrization, as the original distribution. (Implicit in this is the assumption that $\gamma(t)$ is on the boundary of $D \backslash \gamma(0,t]$, which will be true, e.g., if $\gamma$ is non-self-intersecting and $\gamma(0,t_\gamma) \subset D$.)

Using the Riemann mapping theorem, one can see that such a family $\{\mu_D(z,w)\}$ is determined (up to reparametrization) by $\mu_{\mathbb{H}}(0,\infty)$, where $\mathbb{H} = \{x + iy : y > 0\}$ denotes the upper half-plane. Suppose $\gamma:[0,\infty) \to \mathbb{C}$ is a simple (i.e., no self-intersections) curve with $\gamma(0)=0$, $\gamma(0,\infty) \subset \mathbb{H}$, and $\sup_t \mathrm{Im}[\gamma(t)] = \infty$. Let $H_t = \mathbb{H} \backslash \gamma[0,t]$. There is a unique conformal transformation $g_t: H_t \to \mathbb{H}$ whose expansion at infinity is

$$g_t(z) = z + \frac{b(t)}{z} + O(|z|^{-2}), \quad z \to \infty$$

(see Figure 2). The coefficient $b(t)$, which is sometimes called the half-plane capacity of $\gamma[0,t]$ and denoted hcap$[\gamma[0,t]]$, is continuous, strictly increasing, and tending to $\infty$. In fact,

$$b(t) = \lim_{y \to \infty} y E[\mathrm{Im}[X_\tau] \mid X_0 = \mathrm{i}y]$$

where $X_s$ denotes a complex Brownian motion and $\tau = \tau_{\gamma[0,t]}$ is the first time $s$ such that $X_s \in \mathbb{R} \cup \gamma[0,t]$. By reparametrizing $\gamma, b(t) = 2t$. With this parametrization, the maps $g_t$ satisfy the Loewner differential equation

$$\dot{g}_t(z) = \frac{2}{g_t(z) - U_t}, \quad g_0(z) = z$$

where $U:[0,\infty) \to \mathbb{R}$ is a continuous function with $U_0 = 0$. In fact, $U_t = g_t(\gamma(t))$. Schramm observed that the measure $\mu_{\mathbb{H}}(0,\infty)$, at least if it were supported on simple curves and the curves were parametrized using half-plane capacity, would produce a random $U_t$. If the assumptions above on $\{\mu_D(z,w)\}$ are translated into assumptions on the "driving function" $U_t$, one shows readily that $U_t$ must be a driftless Brownian motion, that is, $U_t = \sqrt{\kappa} B_t$, for a standard one-dimensional Brownian motion $B_t$.

Chordal SLE$_\kappa$ (in $\mathbb{H}$ connecting 0 and $\infty$) is defined to be the random collection of conformal maps $g_t$ obtained by solving the initial-value problem

$$\dot{g}_t(z) = \frac{2}{g_t(z) - \sqrt{\kappa} B_t}, \quad g_0(z) = z \qquad [1]$$

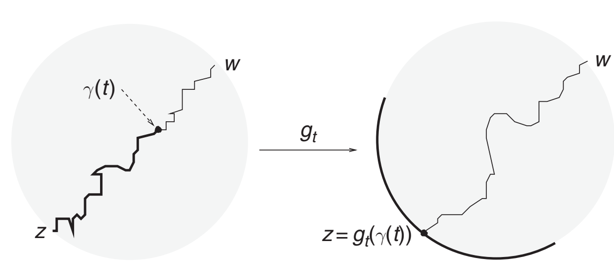



**Figure 1** The map $g_t$ from $D \backslash \gamma[0,t]$ onto $D$.

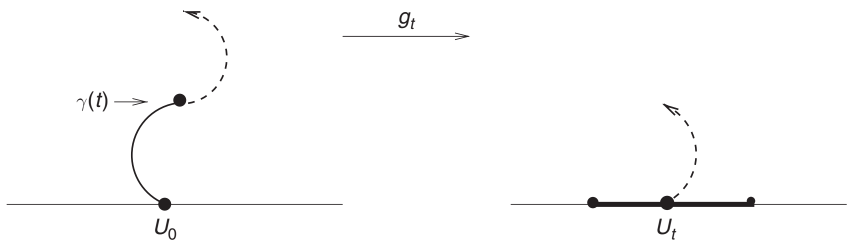



**Figure 2** The map $g_t$ from $\mathbb{H}\backslash\gamma[0,t]$ onto $\mathbb{H}$.

where $B_t$ is a standard one-dimensional Brownian motion. Equation [1] is often given in terms of the inverse $f_t = g_t^{-1}$:

$$\dot{f}_t(z) = -f_t'(z)\frac{2}{z - \sqrt{\kappa}\,B_t}$$

This equation describes a random evolution of conformal maps $f_t$ from $\mathbb{H}$ into subdomains of $\mathbb{H}$. For each $z \in \bar{\mathbb{H}}$, the solution of [1] is defined up to a time $T_z \in [0, \infty]$ with $T_z > 0$ for $z \neq 0$. For fixed $t, g_t$ is the unique conformal transformation of $H_t := \{z \in \mathbb{H} : T_z > t\}$ onto $\mathbb{H}$ with expansion

$$g_t(z) = z + \frac{2t}{z} + \cdots, \quad z \to \infty$$

The chordal $\mathrm{SLE}_\kappa$ path is the random curve $\gamma : [0, \infty) \to \mathbb{H}$ such that for each $t, H_t$ is the unbounded component of $\mathbb{H}\backslash\gamma[0, t]$. It is not immediate from the definition that such a curve $\gamma$ exists, but its existence has been proved. If $G_t = g_{t/\kappa}$, then we can write eqn [1] as

$$\dot{G}_t(z) = \frac{a}{G_t(z) + W_t} \qquad [2]$$

where $a = 2/\kappa$ and $W_t := -\sqrt{\kappa}B_{t/\kappa}$ is a standard Brownian motion. Then $Z_t^z := G_t(z) + W_t$ satisfies the Bessel stochastic differential equation

$$\mathrm{d}Z_t^z = \frac{a}{Z_t^z}\,\mathrm{d}t + \mathrm{d}W_t, \quad Z_0^z = z \qquad [3]$$

This equation is valid up to time $\kappa T_z$, which is the first time that $Z_t^z = 0$.

Although chordal $\mathrm{SLE}_\kappa$ is defined with a particular parametrization, one generally thinks of it as a measure on curves modulo reparametrization. The scaling properties of Brownian motion imply that this measure is invariant under dilations of $\mathbb{H}$. If $D$ is a simply connected domain and $z$, $w$ are distinct boundary points of $D$, chordal $\mathrm{SLE}_\kappa$ in $D$ connecting $z$ and $w$ is defined to be the conformal image of $\mathrm{SLE}_\kappa$ in $\mathbb{H}$ from 0 to $\infty$ under a conformal transformation of $\mathbb{H}$ onto $D$ taking 0 to $z$ and $\infty$ to $w$. There is a one-parameter family of such transformations, but the scale invariance of $\mathrm{SLE}_\kappa$ in $\mathbb{H}$ shows that the image measure is independent of the choice of transformation.

The geometric and fractal properties of the curve $\gamma$ vary greatly as the parameter $\kappa$ changes:

- if $\kappa \leq 4$, $\gamma$ is a simple curve;
- if $4 < \kappa < 8$, $\gamma$ has self-intersections, but is not space filling; and
- if $\kappa \geq 8$, $\gamma$ is a space filling curve.

To see this, one notes that the conformal Markov property implies that there can be double points with positive probability if and only if $T_x < \infty$ occurs with positive probability for $x > 0$. In addition, the curve is space filling if and only if $T_z < \infty$ for all $z$ and $T_w \neq T_z$ for $w \neq z$. The problem is then reduced to a problem about the Bessel equation [3] for which the following holds:

- if $a \geq 1/2$ and $z \neq 0$, the probability that $T_z < \infty$ is zero. If $a < 1/2$, this probability equals 1.
- if $1/4 < a < 1/2$, and $w, z$ are distinct points in $\mathbb{H}$, then there is a positive probability that $T_w = T_z$.
- if $0 < a \leq 1/4$, then with probability 1, $T_w \neq T_z$ for all $w \neq z$.

This kind of argument is typical when studying SLE – geometric properties of the curve are established by analyzing a stochastic differential equation. The Hausdorff dimension of the path $\gamma$ is given by

$$\dim[\gamma[0, \infty)] = \min\left\{1 + \frac{\kappa}{8}, 2\right\}$$

The radial Loewner equation describes the evolution of a curve from the boundary of the unit disk $\mathbb{D} = \{z : |z| < 1\}$ to the origin. Suppose $\gamma : [0, \infty) \to \bar{\mathbb{D}}$ is a simple curve with $\gamma(0) = 1$, $\gamma(0, \infty) \subset \mathbb{D}\backslash\{0\}$, and $\gamma(t) \to 0$ as $t \to \infty$. Let $g_t$ be the unique conformal transformation of $\mathbb{D}\backslash\gamma[0, t]$ onto $\mathbb{D}$ such that $g_t(0) = 0, g_t'(0) > 0$. One can check that $g_t'(0)$ is continuous and strictly increasing in $t$, and hence we can parametrize $\gamma$ in such a way that $g_t'(0) = \mathrm{e}^t$. Using this reparametrization, there is a continuous

$U_t : [0, \infty) \to \mathbb{R}$ with $U_0 = 0$ such that $g_t$ satisfies the radial Loewner equation

$$\dot{g}_t(z) = g_t(z)\frac{e^{iU_t} + g_t(z)}{e^{iU_t} - g_t(z)}, \quad g_0(z) = z$$

If $z \neq 0$, then we can define $h_t(z) = -i \log g_t(z)$ locally near $z$, and this equation becomes

$$\dot{h}_t(z) = \cot\left(\frac{h_t(z) - U_t}{2}\right)$$

Radial $SLE_\kappa$ (connecting 1 and 0 in $\mathbb{D}$) is obtained by setting $U_t = \sqrt{\kappa}B_t$. If $D$ is a simply connected domain, $z \in D, w \in \partial D$, then radial $SLE_\kappa$ in $D$ connecting $w$ and $z$ is obtained by conformal transformation using the unique transformation $f$ of $\mathbb{D}$ onto $D$ with $f(0) = z, f(1) = w$. Again, we think of this as being defined modulo time change. If $a = 2/\kappa$ and $v_t = h_{at/2}$, then

$$\dot{v}_t(z) = \frac{a}{2}\cot\left(\frac{v_t(z) + W_t}{2}\right) \qquad [4]$$

where $W_t := -\sqrt{\kappa}B_{t/\kappa}$ is a standard Brownian motion. If $L_t^z = v_t(z) + W_t$, then we get

$$dL_t^z = \frac{a}{2}\cot\left(\frac{L_t^z}{2}\right)dt + dW_t$$

Radial and chordal SLE are closely related. In fact, if $\gamma$ is a chordal SLE path in $\mathbb{H}$ from 0 to $\infty$, $\tilde{\gamma}$ is a radial SLE path in $\mathbb{D}$ from 1 to 0, and $\eta = -i \log \tilde{\gamma}$, then for small $t$ the distribution of $\eta$ is absolutely continuous to the distribution of a (random time change of) $\gamma$. Showing this involves understanding the behavior of the Loewner equation under conformal transformations. Suppose $\gamma, \tilde{\gamma}$ have been parametrized as in [2] and [4] with $a = 2/\kappa$. Let $g_t^*$ be the conformal transformation of $\mathbb{H} \setminus \eta[0, t]$ onto $\mathbb{H}$ such that

$$g_t^*(z) = z + \frac{a^*(t)}{z} + \cdots, \quad z \to \infty$$

and let $U_t^*$ be the Loewner driving function such that

$$\dot{g}_t^*(z) = \frac{\dot{a}^*(t)}{g_t^*(z) - U_t^*}$$

Here $a^*(t) = \mathrm{hcap}[\eta[0, t]]$. If we consider a time change $\sigma$ such that $a^*(\sigma(t)) = at$ and let $U_t = \tilde{U}_{\sigma(t)}$ be the time-changed driving function, Itô's formula can be used to show that

$$dU_t = \tfrac{1}{2}(1 - 3a)F_t\,dt + d\tilde{W}_t \qquad [5]$$

where the $F_t$ in the drift term depends on $\gamma[0, t]$ and is independent of $a$, and $\tilde{W}$ is a standard Brownian motion. Girsanov's theorem implies that Brownian motions with the same variance but different drifts have absolutely continuous distributions. In particular, qualitative properties such as existence of double points or Hausdorff dimension of paths are the same for radial and chordal SLE. $U_t$ is a driftless Brownian motion if $a = 1/3$, $\kappa = 6$.

Whole-plane $SLE_\kappa$ from 0 to $\infty$ is a path $\gamma : (-\infty, \infty) \to \mathbb{C}$ with $\gamma(-\infty) = 0$, $\gamma(\infty) = \infty$, such that given $\gamma(-\infty, t]$, the distribution of $\gamma(t, \infty)$ is that of radial $SLE_\kappa$ from boundary point $\gamma(t)$ to interior point $\infty$ in the domain $\mathbb{C} \setminus \gamma[-\infty, t]$. One can define whole-plane $SLE_\kappa$ connecting two distinct points in $\mathbb{C}$ by conformal transformation.

## Locality and Restriction

There are two special values of $\kappa : \kappa = 6, a = 1/3$ that satisfies the "locality" property and $\kappa = 8/3$, $a = 3/4$ that satisfies the "restriction" property. Suppose $\gamma$ is a chordal $SLE_\kappa$ curve from 0 to $\infty$ in $\mathbb{H}$ parametrized as in [2]. Suppose $\Phi : \mathcal{N} \to H$ is a conformal map taking a neighborhood $\mathcal{N}$ of 0 in $\mathbb{H}$ to $\Phi(\mathcal{N})$ and that locally maps $\mathbb{R}$ into $\mathbb{R}$. Let $\tilde{\gamma}(t) = \Phi \circ \gamma(t)$, which is defined for sufficiently small $t$. Let $g_t^*$ be the conformal transformation of $\mathbb{H} \setminus \tilde{\gamma}[0, t]$ onto $\mathbb{H}$ with

$$g_t^*(z) = z + \frac{a^*(t)}{z} + \cdots$$

and let $\tilde{U}_t$ be the driving function such that

$$\dot{g}_t^*(z) = \frac{\dot{a}^*(t)}{g_t^*(z) - \tilde{U}_t}$$

Here $a^*(t) = \mathrm{hcap}[\tilde{\gamma}[0, t]]$. If we change time, $\gamma_t^* = \tilde{\gamma}_{\sigma(t)}$, so that $a^*(\sigma(t)) = at$, then an application of Itô's formula shows that $U_t^* := \tilde{U}_{\sigma(t)}$ satisfies

$$dU_t^* = \frac{1}{2}(3a - 1)\frac{\Phi''_{\sigma(t)}(W_{\sigma(t)})}{\Phi'_{\sigma(t)}(W_{\sigma(t)})^2}dt + d\tilde{W}_t$$

Here $\tilde{W}_t$ is a standard Brownian motion, $\Phi_t = g_t^* \circ \Phi \circ g_t^{-1}$, and $g_t$ is the conformal map associated to $\gamma$. In particular, if $a = 1/3$, $\kappa = 6$, $U_t^*$ is a standard Brownian motion; hence, $\tilde{\gamma}^*$ has the distribution of $SLE_6$. The locality property for $SLE_6$ can be stated as "the conformal image of $SLE_6$ is (a time change of) $SLE_6$." Intuitively, the $SLE_6$ path in a restricted domain does not feel the boundary of the domain until it reaches it. Radial $SLE_6$ satisfies a similar locality property. Moreover, [5] can be used to show that the image of chordal $SLE_6$ under the exponential map is the same (for small time $t$) as radial $SLE_6$. The locality property explains why

$SLE_6$ is a natural candidate for the boundary of percolation clusters.

If $\kappa \leq 4, SLE_\kappa$ paths are simple, that is, with no self-intersections. Suppose $A \subset \bar{\mathbb{H}} \backslash \{0\}$ is a compact set such that $\mathbb{H} \backslash A$ is simply connected. Let $\gamma$ denote a chordal $SLE_\kappa$ in $\mathbb{H}$ connecting 0 and $\infty$ and let $E_A$ be the event $E_A = \{\gamma(0, \infty) \cap A = \emptyset\}$. Let $\Phi_A : \mathbb{H} \backslash A \to \mathbb{H}$ be the unique conformal transformation with $\Phi_A(0) = 0$, $\Phi_A(\infty) = \infty$, $\Phi'_A(\infty) = 1$. On the event $E_A$, we can define $\tilde{\gamma}(t) = \Phi_A \circ \gamma(t)$. Chordal $SLE_\kappa$ is said to satisfy the restriction property if the conditional distribution of $\tilde{\gamma}$ given $E_A$ is the same as (a time change of) $\gamma$. The only $\kappa \leq 4$ that satisfies this property is $\kappa = 8/3$. The proof of this fact also establishes the formula: if $\gamma$ is a chordal $SLE_{8/3}$ curve in $\mathbb{H}$ from 0 to $\infty$, then

$$\boldsymbol{P}\{\gamma(0,\infty) \cap A = \emptyset\} = \Phi'_A(0)^{5/8} \qquad [6]$$

There is a similar formula for radial $SLE_{8/3}$, which establishes a radial restriction property. Suppose $A \subset \bar{\mathbb{D}} \backslash \{0, 1\}$ is a compact set such that $\mathbb{D} \backslash A$ is simply connected. Let $\Psi_A$ be the unique conformal transformation of $\mathbb{D} \backslash A$ onto $\mathbb{D}$ with $\Psi_A(0)$, $\Psi'_A(0) > 0$. Then, if $\gamma$ is a radial $SLE_{8/3}$ curve from 1 to 0 in $\mathbb{D}$, then

$$\boldsymbol{P}\{\gamma(0,\infty) \cap A = \emptyset\} = \Psi'_A(0)^{5/48} |\Psi'_A(1)|^{5/8}$$

The restriction property makes $SLE_{8/3}$ the candidate for the scaling limit of self-avoiding walks.

## Relation to Conformal Field Theory

The Schramm–Loewner evolution is one of the tools used to rigorously prove predictions made using powerful, yet nonrigorous, arguments of conformal field theory. In conformal field theory, there is a parameter $c$, called the central charge, which classifies theories. To each $c \leq 1$, there corresponds a $\kappa \leq 4$ and a "dual" $\kappa' = 16/\kappa \geq 4$:

$$c = c_\kappa = \frac{(8 - 3\kappa)(\kappa - 6)}{2\kappa}$$

In particular, $\kappa = 8/3$, $\kappa' = 6$ corresponds to central charge zero. It is expected, and has been proved in a number of cases, that $SLE_\kappa$ or $SLE_{\kappa'}$ curves will appear in scaling limits of systems with central charge $c_\kappa$. These systems can also be parametrized by the boundary scaling exponent or conformal weight

$$\alpha = \alpha_\kappa = \frac{6 - \kappa}{2\kappa}$$

For $\kappa = 8/3$, $\alpha = 5/8$ which is the exponent in [6].

In studying the relationship between SLE and conformal field theories, two other probabilistic objects, restriction measures and the (Brownian) loop soup, arise. An $\mathbb{H}$-hull (connecting 0 and $\infty$) is an unbounded, connected, closed set $K \subset \bar{\mathbb{H}}$ with $K \cap \mathbb{R} = \{0\}$ and such that $\mathbb{H} \backslash K$ consists of two connected components, one whose boundary includes the positive reals and the other whose boundary includes the negative reals. A (chordal) restriction measure on hulls $K$ is a probability measure with the property that for any $A$ as in [6], the distribution of $\Phi_A \circ K$ given $\{K \cap A = \emptyset\}$ is the same as the original measure. The (Brownian) loop measure is a measure on unrooted loops derived from Brownian bridges. It is the scaling limit of the measure on random walk loops that gives each unrooted simple random walk loop of length $2n$ measure $4^{-2n}$. The loop measure in a bounded domain is obtained by restricting to loops that stay in that domain. We can consider this as a measure on "hulls" by filling in the bounded holes (so that the complement of the hull is connected). By doing this we get a family of infinite measures on hulls, indexed by domains $D$, and this family satisfies conformal invariance and the restriction property. The loop soup with parameter $\lambda$ is a Poissonian realization from this measure with parameter $\lambda$.

The set of all restriction measures is parametrized by $\alpha \geq 5/8$; the $\alpha$-restriction measure has the property that

$$\boldsymbol{P}\{K \cap A \neq \emptyset\} = \Phi'_A(0)^\alpha$$

For $\alpha = 5/8$, $K$ is given by the path of $SLE_{8/3}$. For integer $\alpha$, the hull $K$ can be constructed by taking $\alpha$-independent Brownian excursions in $\mathbb{H}$ (Brownian motions starting at 0 conditioned to stay in $\mathbb{H}$ for all times), and letting $K$ be the hull obtained by taking the union of the paths and filling in the bounded holes. If $\kappa \leq 8/3, c_\kappa \leq 0$, then the restriction measure with exponent $\alpha_\kappa \geq 5/8$ can also be constructed as follows: take a chordal $SLE_\kappa$ path and an independent realization of the loop soup with intensity $\lambda_\kappa = -c_\kappa$; add to the SLE path all the loops in the soup that intersect the $SLE_\kappa$ curve; and then fill in all the bounded hulls. The limiting case $\alpha = 5/8$, $\lambda = 0$ gives the only measure supported on simple curves that is also a restriction measure, $SLE_{8/3}$.

For $8/3 < \kappa \leq 4, 0 < c_\kappa \leq 1$, it is conjectured, and proved for small $c_\kappa$, that $SLE_\kappa$ curves can be found by taking a loop soup with parameter $\lambda = c_\kappa$ and looking at connected curves in the fractal set given by the complement of the union of all the hulls generated by the loops.

## Examples

The scaling limit of simple random walk, Brownian motion, is known to be conformally invariant. A two-dimensional Brownian bridge or loop is a Brownian motion, $B_t, 0 \leq t \leq 1$, conditioned so that $B_0 = B_1$. The frontier or outer boundary of the Brownian motion is the boundary of the unbounded component of the complement. Benoit Mandelbrot first observed numerically that the outer boundary of Brownian motion had fractal dimension $\sim 4/3$. Gregory Lawler, Oded Schramm, and Wendelin Werner used SLE to prove that the boundary has Hausdorff dimension 4/3. In fact, the outer boundary can be considered as an $SLE_{8/3}$ loop.

$SLE_6$ and $SLE_{8/3}$ arise in the scaling limit of critical percolation on the triangular lattice. Suppose that each vertex in the upper half-plane triangular lattice is colored black or white each with a probability 1/2. Suppose the real line gives a boundary condition of black on the negative real line and white on the positive real line. Then if we represent the vertices in the lattice as hexagons as in the figure, a curve is formed which is the boundary between the black and white clusters. This curve is called the "percolation exploration process." Stanislav Smirnov proved that the scaling limit of this curve is conformally invariant, and from this it can be concluded that the curve is chordal $SLE_6$. In particular, the Hausdorff dimension is 7/4 and the scaling limit has double points. In the scaling limit, the "outer boundary" of this curve has Hausdorff dimension 4/3 and its dimension is absolutely continuous with respect to that of $SLE_{8/3}$. While this result is expected for other critical percolation model, such as bond percolation in $\mathbb{Z}^2$ with critical probability 1/2, it has only been proved for the triangular lattice. Percolation has central charge 0 and the "locality" property can be seen in the lattice model. The outer boundary of the curve has the same distribution as the outer boundary of a Brownian motion that is reflected at angle $\pi/3$ off the real line. Locally, the outer boundary of percolation, the outer boundary of complex Brownian motion, and $SLE_{8/3}$ all look the same, and it is expected that this will also be true for the scaling limit of self-avoiding walks.

There are three models derived in some way from simple random walk that have been proved to have scaling limits of $SLE_\kappa$. The loop-erased random walk (LERW) in a finite subset $V$ of $\mathbb{Z}^2$ connecting two distinct points is obtained by taking a simple random walk from one point to the other and erasing loops chronologically. The LERW is closely related to uniform spanning trees; in fact, if one chooses a spanning tree of $V$ from the uniform distribution on all spanning trees, then the distribution of the unique path connecting the two points is exactly that of the LERW (see Figure 3). Another description of the LERW is as the Laplacian random walk: the LERW from $z$ to $w$ in $V$ chooses a new step weighted by the value of the function that is harmonic on the complement of $w$ and the path up to that point with boundary values 0 on the path and 1 on $w$. The LERW in the discrete upper half-plane can be obtained by erasing loops from a simple random walk excursion. The LERW and the uniform spanning tree are systems with central charge $c = -2$. It has been proved that the scaling limit of the LERW is $SLE_2$; hence, the paths have Hausdorff dimension 5/8.

There is another path associated to spanning trees given by the one-to-one correspondence between spanning trees and Hamiltonian walks on a corresponding directed (Manhattan) lattice on the dual graph (see Figure 4). If the spanning trees, or equivalently the Hamiltonian walks, are chosen using the uniform distribution, then the scaling limit of this walk is the space-filling curve $SLE_8$. Note that 2 and 8 are the dual values of $\kappa$ associated to $c = -2$.

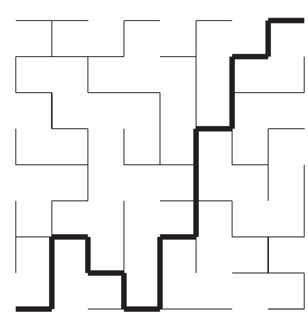


**Figure 3** A spanning tree and the path between two vertices. If the tree has the uniform distribution, the path has the distribution of the LERW.

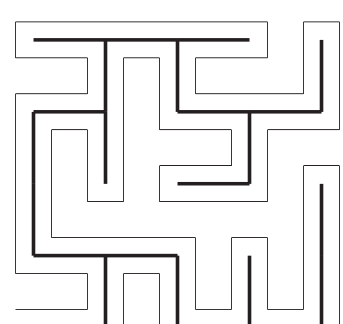


**Figure 4** A spanning tree and the corresponding Hamiltonian walk.

Another discrete process derived from simple random walk, the harmonic explorer, has a scaling limit of $SLE_4$. There is a particular property of $SLE_\kappa$ that leads to the definition of this discrete process. Consider a chordal $SLE_\kappa$ curve, let $z \in \mathbb{H}$, and let $Z_t^z$ be as in [3] with $a = 2/\kappa$. Itô's formula shows that $\Theta_t := \arg(Z_t^z)$ satisfies

$$d\Theta_t = \left(\frac{1}{2} - a\right) \frac{\sin(2\Theta_t)}{|Z_t^z|^2}\, dt - \frac{\sin\Theta_t}{|Z_t^z|}\, dW_t$$

In particular, $\Theta_t$ is a martingale if and only if $a = 1/2$, $\kappa = 4$. The probability that a complex Brownian motion starting at $z \in \mathbb{H}$ first hits $\mathbb{R}$ on the negative half-line can be shown to be $\arg(z)$. If $\kappa \le 4$, then we can see that $\Theta_\infty$ equals 0 or $\pi$, depending on whether $z$ is on the right or left side of the path $\gamma(0, \infty)$. For the martingale case $\kappa = 4$, $\Theta_t$ represents the probability that $z$ is on the left side of $\gamma(0, \infty)$, given $\gamma(0, t]$. The harmonic explorer is a process on the hexagonal lattice defined to have this property. In a way similar to the percolation process, the walk is defined as the boundary between black and white hexagons on the triangular lattice. However, when an unexplored hexagon is reached in the harmonic explorer, it is colored black with probability $q$, where $q$ is the probability that a simple random walk on the triangular lattice starting at that hexagon (considered as a vertex in the triangular lattice) hits a black hexagon before hitting a white hexagon. It is not difficult to show that this process has the property that for $z$ away from the curve, the "probability of $z$ ending on the left given the curve of $n$ steps" is a martingale.

There are many other models for which $SLE_\kappa$ curves are expected in the limit, but it has not been established. The most difficult part is to show the existence of a limit that is conformally invariant. One example is the self-avoiding walk (SAW). It is an open problem to establish that there exists a scaling limit of the uniform measure on SAWs and to establish conformal invariance of the limit. However, the nature of the discrete model is such that if the limit exists, it must satisfy the restriction property. Hence, under the assumption of conformal invariance, the only possible limit is $SLE_{8/3}$. Numerical simulations strongly support the conjecture that $SLE_{8/3}$ is the limit of SAWs, and this gives strong evidence for the conformal invariance conjecture for SAWs. Critical exponents for SAWs (as well as critical exponents for many other models) can be predicted nonrigorously from rigorous scaling exponents for the corresponding SLE paths.

## Generalizations

One of the reasons that the theory of SLE is nice for simply connected domains is that a simply connected domain with an arc connected to the boundary of the domain removed is again simply connected. For nonsimply connected domains, it is more difficult to describe because the conformal type of the slit domain changes as time evolves. In the case of a curve crossing an annulus, this can be done with an added parameter referring to the conformal type of the annulus (two annuli of the form $\{z : r_j < |z| < s_j\}$ are conformally equivalent if and only if $r_1/s_1 = r_2/s_2$). It is not immediately obvious what the correct definition of SLE should be in general domains and, more generally, on Riemann surfaces. One possibility for $\kappa \le 4$ is to consider a configurational (equilibrium statistical mechanics) view of SLE. Consider a family of measures $\{\mu_D(z, w)\}$, where $D$ ranges over domains and $z$, $w$ are distinct boundary points at $\partial D$ is locally analytic, supported on simple curves from $z$ to $w$ (modulo time change). Let $\mu_D^\#(z, w) = \mu_D(z, w)/|\mu_D(z, w)|$ be the corresponding probability measures, which may be defined even if $\partial D$ is not smooth at $z, w$. Then the following axioms should hold:

- *Conformal invariance*. If $f : D \to D'$ is a conformal transformation, $f \circ \mu_D^\#(z, w) = \mu_{D'}^\#(f(z), f(w))$.
- *Conformal Markov property*.
- *Perturbation of domains*. Suppose $D_1 \subset D$ and $\partial D_1, \partial D$ agree near $z$, $w$. Then $\mu_{D_1}(z, w)$ should be absolutely continuous with respect to $\mu_D(z, w)$. Let $Y$ denote the Radon–Nikodym derivative of $\mu_{D_1}(z, w)$ with respect to $\mu_D(z, w)$. Then

$$Y(\gamma) = 1\{\gamma(0, t_\gamma) \subset D_1\}\ F_c(D; \gamma, D\backslash D_1)$$

where $F_c$ is to be determined. In the case where $D, D_1$ are simply connected, $F_c(D; \gamma, D\backslash D_1) = J(\gamma, D, D_1)^{-c}$, where $J(\gamma, D, D_1)$ denotes the probability that there is a loop in the Brownian loop soup in $D$ that intersects both $\gamma$ and $D\backslash D_1$. (There is no problem defining this quantity in nonsimply connected domains, but it is not clear that it is the right quantity.) Here $c = c_\kappa$. The restriction property tells us that $F_0 \equiv 1$.

- *Conformal covariance*. If $f$ is as above, $\partial D$, $\partial D'$ are smooth near $z, w$ and $f(z), f(w)$, respectively, then

$$f \circ \mu_D(z, w) = |f'(z)|^\alpha |f'(w)|^\alpha \mu_{D'}(f(z), f(w))$$

Here $\alpha = \alpha_\kappa$ is the boundary scaling exponent.

*See also:* Boundary Conformal Field Theory; Percolation Theory; Random Walks in Random Environments.

## Further Reading

Kager W and Nienhuis B (2004) A guide to stochastic Loewner evolution and its applications. *Journal of Statistical Physics* 115: 1149–1229.
Lawler G (2004) Conformally invariant processes in the plane. In: *School and Conference on Probability Theory*, ICTP Lecture Notes, vol. 17, pp. 305–351. Trieste, Italy: ICTP.
Lawler G (2005) *Conformally Invariant Processes in the Plane.* Providence, RI: American Mathematical Society.
Lawler G and Werner W (2004) The Brownian loop soup. *Probability Theory Related Fields* 128: 565–588.
Lawler G, Schramm O, and Werner W (2004a) Conformal restriction, the chordal case. *The Journal of American Mathematical Society* 16: 917–955.
Lawler G, Schramm O, and Werner W (2004b) Conformal invariance of planar loop-erased random walks and uniform spanning trees. *Annals of Probability* 32: 939–995.
Schramm O (2000) Scaling limits of loop-erased random walks and uniform spanning trees. *Israel Journal of Mathematics* 118: 221–288.
Smirnov S (2001) Critical percolation in the plane: conformal invariance, Cardy's formula, scaling limits. *Comptes rendus de l'Academie des sciences – series i – mathematics* 333: 239–244.
Werner W (2003) SLEs as boundaries of clusters of Brownian loops. *Comptes rendus de l'Academie des sciences – series i – mathematics* 337: 481–486.
Werner W (2004) Random planar curves and Schramm–Loewner evolutions. In: *Ecole d'Eté de Probabilités de Saint-Flour XXXII – 2002*, Lecture Notes in Mathematics, vol. 1840, pp. 113–195. Berlin: Springer.

# Stochastic Resonance

**S Herrmann**, Université Henri Poincaré, Nancy 1 Vandoeuvre-lès-Nancy, France
**P Imkeller**, Humboldt Universität zu Berlin, Berlin, Germany



## Introduction

The concept of stochastic resonance was introduced by physicists. It originated in a toy model designed for a qualitative description of periodicity phenomena in the recurrences of glacial eras in Earth's history. It spread its popularity over numerous areas of natural sciences: neuronal response to periodic stimuli, variations of magnetization in a ferromagnetic system, voltage variations in the simple Schmitt trigger electronic circuit or in more complicated devices, behavior of lasers in optical bi-stability, etc. The interest in this ubiquitous phenomenon is enhanced by signal analysis: an optimal dose of noise in some system can essentially boost signal transduction. Noise in this context does not enter the system as an impurity perturbing its performance, but on the contrary as a catalyst triggering amplified stochastic response to weak periodic signals.

## The Climate Paradigm

The phenomenon of stochastic resonance was first discovered in an elementary climate model serving in an explanation of major transitions in paleoclimatic time series confining glacial cycles. Data collected for instance from ice or deep sea cores allow one to deduce estimates of the average temperature on Earth over the last 700 000 years. They exhibit periodic switching between ice and warm ages with fast spontaneous transitions. The average periodicity of the glaciation time series obtained is $\sim 10^5$ years. In order to explain temperature variations, Benzi *et al.* (1981) introduced random perturbations into an energy balance model of the Budyko–Sellers type. This model describes the evolution of the seasonal and global average temperature $X$ caused by defects in the balance between incoming and outgoing radiation

$$c\frac{\mathrm{d}X(t)}{\mathrm{d}t} = E_{\mathrm{in}} - E_{\mathrm{out}}$$

where $c$ is the active thermal inertia of the system. The incoming energy is modeled as proportional to the "solar constant" $Q$:

$$E_{\mathrm{in}} = Q\left(1 + A\cos\frac{2\pi t}{T}\right), \quad \text{with } T \approx 92\,000 \text{ years}$$

and $A \approx 0.1\%$ of $Q$. This exceedingly small variation of the solar constant is caused by a modulation of the orbital eccentricity of the Earth's trajectory (Figure 1). The outgoing radiation $E_{\mathrm{out}}$ is composed

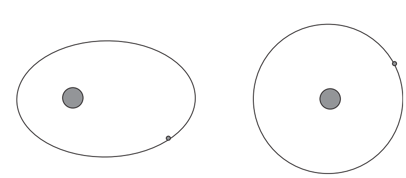


**Figure 1** Modulation of the orbital eccentricity.

of two essential parts. The first part $a(X)E_{\text{in}}$ is dominated by the albedo $a(X)$ representing the proportion of energy reflected back to space. It is a decreasing function of temperature, due to the higher rate of reflection from a brighter Earth at low temperatures implying a bigger volume of ice. The second part of the outgoing radiation comes from the fact that the Earth radiates energy like a black body, and is given by the Boltzmann law $\gamma X^4$, where $\gamma$ is the Stefan constant. Describing the balance of energy terms as a slowly and weakly time-varying gradient of a potential $U$, the balance model can be expressed by

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = -\frac{\partial U}{\partial x}\left(\frac{t}{T}, X(t)\right)$$

where the time period 1 is blown up to (large) $T$ by time scaling. The roles of deep and shallow wells switch periodically (Figure 2). Since the variation of the solar constant is extremely small, we can assume that the height of the barrier between the two wells is lower-bounded by a positive constant. The system then admits three steady states two of which are stable and separated by roughly 10 K. As the solar constant, they fluctuate slowly and very weakly. Therefore, this deterministic system cannot account for climate changes with temperature variations of $\sim$10 K. They can only be explained by allowing transitions between the two steady states which become possible by adding noise to the system. In general, short timescale phenomena such as annual fluctuations in solar radiation are modeled by Gaussian white noise of intensity $\varepsilon$ and lead to equations of the type

$$\mathrm{d}X_t^\varepsilon = -\frac{\partial U}{\partial x}\left(\frac{t}{T}, X_t^\varepsilon\right)\mathrm{d}t + \sqrt{\varepsilon}\,\mathrm{d}W_t \qquad [1]$$

which are generic for studying stochastic resonance in numerous physical and biological models. Generally, the input of noise amplifies a weak periodic signal by creating trajectories fluctuating randomly periodically between meta-stable states. An optimal tuning of noise intensity to period length ("stochastic resonance") significantly enhances the response of the random system to weak perturbations with long periods.

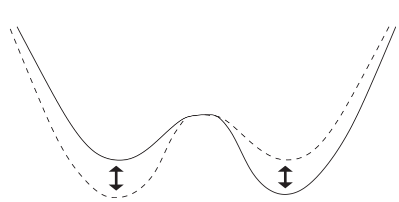


**Figure 2** Deep and shallow wells switching periodically.

## Strongly Damped Brownian Particle

It is useful to roughly compare solutions of stochastic differential equations and motions of Brownian particles in double-well landscapes (Figure 3) in order to understand properties of their trajectories (see Schweitzer 2003, Mazo 2002). As in the previous section, let us concentrate on a one-dimensional setting, remarking that we shall give a treatment that easily generalizes to the finite-dimensional setting. Due to Newton's law, the motion of a particle is governed by the impact of all forces acting on it. Let us denote $F$ the sum of these forces, $m$ the mass, $x$ the space coordinate, and $v$ the velocity of the particle. Then

$$m\dot{v} = F$$

Let us first assume the potential to be switched off. In their pioneering work at the turn of the twentieth century, Marian v. Smoluchowski and Paul Langevin introduced stochastic concepts to describe the Brownian particle motion by claiming that at time $t$

$$F(t) = -\gamma_0 v(t) + \sqrt{2k_{\mathrm{B}}T\gamma_0}\,\dot{W}_t$$

The first term results from friction $\gamma_0$ and is velocity dependent. An additional stochastic force represents random interactions between Brownian particles and their simple molecular random environment. The white noise $\dot{W}$ (formal derivative of the Wiener process) plays the crucial role. The diffusion coefficient (standard deviation of the random impact) is composed of Boltzmann's constant $k_{\mathrm{B}}$, friction, and environmental temperature $T$. It satisfies the condition of the fluctuation–dissipation theorem expressing the balance of energy loss due to friction and energy gain resulting from noise. The equation of motion becomes

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = v(t)$$

$$\mathrm{d}v(t) = -\frac{\gamma_0}{m}v(t)\,\mathrm{d}t + \frac{\sqrt{2k_{\mathrm{B}}T\gamma_0}}{m}\,\mathrm{d}W_t$$

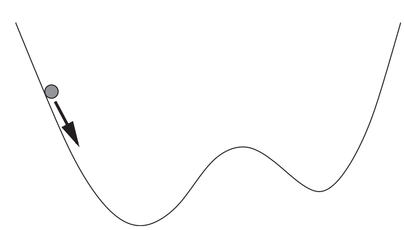


**Figure 3** Brownian particle in a double-well landscape.

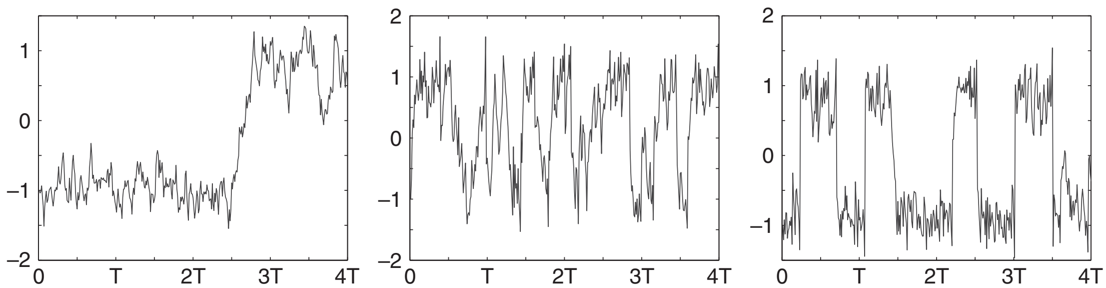


**Figure 4** Resonance pictures for diffusions.

In the stationary regime, the stationary Ornstein–Uhlenbeck process provides its solution

$$v(t) = v(0)\,\mathrm{e}^{-(\gamma_0/m)t} + \frac{\sqrt{2k_\mathrm{B}T\gamma_0}}{m}\int_0^t \mathrm{e}^{-(\gamma_0/m)(t-s)}\,\mathrm{d}W_s$$

The ratio $\beta := \gamma_0/m$ determines the dynamic behavior. Let us focus on the over-damped situation with large friction and very small mass. Then for $t >> 1/\beta = \tau$ (relaxation time), the first term in the expression for velocity can be neglected, while the stochastic integral represents a Gaussian process. By integrating, we obtain in the over-damped limit ($\beta \to \infty$) that $v$ and thus $x$ is Gaussian with almost constant mean

$$m(t) = x(0) + \frac{1-\mathrm{e}^{-\beta t}}{\beta}\,v(0) \approx x(0)$$

and covariance close to the covariance of white noise see Nelson (1967):

$$K(s,t) = \frac{2k_\mathrm{B}T}{\gamma_0}\min(s,t) + \frac{k_\mathrm{B}T}{\gamma_0\beta}(-2 + 2\mathrm{e}^{-\beta t} + 2\mathrm{e}^{-\beta s} - \mathrm{e}^{-\beta|t-s|} - \mathrm{e}^{-\beta(t+s)})$$
$$\approx \frac{2k_\mathrm{B}T}{\gamma_0}\min(s,t)$$

Hence, the time-dependent change of the velocity of the Brownian particle can be neglected, the velocity rapidly thermalizes ($\dot{v} \approx 0$), while the spactial coordinate remains far from equilibrium. In the so-called adiabatic transformation, the evolution of the particle's position is thus given by the transformed Langevin equation

$$\mathrm{d}x(t) = \frac{\sqrt{2k_\mathrm{B}T}}{\gamma_0}\,\mathrm{d}W_t$$

Let us next suppose that we have a Brownian particle in an external field of force (see Figure 3), generating a potential $U(t,x)$. This leads to the Langevin equation

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = v(t)$$
$$m\,\mathrm{d}v(t) = -\gamma_0\,v(t)\mathrm{d}t - \frac{\partial U}{\partial x}(t,x(t)) + \sqrt{2k_\mathrm{B}T\gamma_0}\,\mathrm{d}W_t$$

In the over-damped limit, after relaxation time, the adiabatic elimination of the fast variables (Gardiner 2004) leads to an equation similar to the one encountered in the previous section:

$$\mathrm{d}x(t) = -\frac{1}{\gamma_0}\frac{\partial U}{\partial x}(t,x(t)) + \frac{\sqrt{2k_\mathrm{B}T}}{\gamma_0}\,\mathrm{d}W_t$$

In the particular case of some double-well potential $x \to U(t,x)$ with slow periodic variation, the following patterns of behavior of the solution trajectories will be experienced. If temperature is high, noise has a predominant influence on the motion, and the particle often crosses the barrier separating the two wells during one period. The behavior of the particle does not seem to be periodic but rather chaotic. If temperature is small, the particle stays for a very long time in the starting well, fluctuating weakly around the equilibrium position. It has too low energy to follow the periodic variation of the potential. So in this case too, the trajectories do not look periodic. Between these two extreme situations, there exists a regime of noise intensities for which the energy transmitted by the noise is sufficient to cross the barrier almost twice per period. The parameters are then near to the resonance point and the motion exhibits periodic switching (Figure 4).

## Transition Criteria and Quasideterministic Motion

Studying stochastic resonance accordingly means looking for the range of regimes for which periodic behavior is enhanced and eventually optimal. The optimal relation between period $T$ and noise

intensity $\varepsilon$ emerges in the small noise limit. To explain this, let us focus on the basic indicator for periodic transitions – the time the Brownian particle needs to exit from the starting well, say the left one. In the "frozen" case, that is, if the time variation of the potential term is eliminated just by freezing it at some time $s$, the asymptotics of the exit time is derived from the classical large deviation theory of randomly perturbed dynamical systems (see Freidlin and Wentzell 1998). Let us assume that $U$ is locally Lipschitz. We denote by $D_l$ (resp. $D_r$) the domain corresponding to the left (resp. right) well and $\chi$ their common boundary. The law of the first exit time $\tau^{\varepsilon}_{D_l} = \inf\{t \geq 0, X^{\varepsilon}_t \notin D_l\}$ is described by some particular functional related to large deviation. For $t > 0$, we introduce the "action functional" on the space of continuous functions $\mathcal{C}([0,t])$ on $[0,t]$ by

$$S^s_t(\varphi) = \begin{cases} \frac{1}{2}\int_0^t \left(\dot{\varphi}_u + \frac{\partial U}{\partial x}(s, \varphi_u)\right)^2 \mathrm{d}u, & \text{if } \varphi \text{ is abs. continuous} \\ +\infty & \text{otherwise} \end{cases}$$

which is non-negative and vanishes on the set of solutions of the ordinary differential equation $\dot{x} = -(\partial U/\partial x)(s,x)$. Let $x$ and $y \in \mathbb{R}$. In relation with the action functional, we define the quasipotential

$$V_s(x,y) = \inf\{S^s_t(\varphi) : \varphi \in \mathcal{C}([0,t]), \varphi_0 = x, \varphi_t = y, t \geq 0\}$$

It represents the minimal work the diffusion starting in $x$ has to do in order to reach $y$. To switch wells, the Brownian particle starting in the left well's bottom $x_l$ has to overcome the barrier. So we let

$$\overline{V}_s = \inf_{y \in \chi} V_s(x_l, y)$$

This minimal work needed to exit from the left well can be computed explicitly, and is seen to equal to twice its depth. The asymptotic behavior of the exit time is expressed by

$$\lim_{\varepsilon \to 0} \varepsilon \ln \mathbb{E}[\tau^{\varepsilon}_{D_l}] = \overline{V}_s$$

and

$$\lim_{\varepsilon \to 0} \mathbb{P}_x\left(\mathrm{e}^{(\overline{V}_s - \delta)/\varepsilon} < \tau^{\varepsilon}_{D_l} < \mathrm{e}^{(\overline{V}_s + \delta/\varepsilon)}\right) = 1 \quad \text{for any } \delta > 0$$

The prefactor for the exponential rate, derived by Freidlin and Wentzell (1998), was first given by Eyring and Kramers and then by Bovier *et al.* (2004).

Let us now assume that the left well is the deeper one at time $s$. If the Brownian particle has enough time to cross the barrier, that is, if $T > \mathrm{e}^{\overline{V}_s/\epsilon}$, then whatever the starting point is, Freidlin (2000) proved that it should stay near $x_l$ in the following sense:

$$\Lambda(t \in [0,1] : |X^{\varepsilon}_{tT} - x_l| > \delta) \to 0$$

in probability as $\varepsilon \to 0$. Here $\Lambda$ denotes Lebesgue measure on $\mathbb{R}$. If $T < \mathrm{e}^{\overline{V}_s/\epsilon}$, the time left is not long enough for crossings: the particle stays in the starting well, near the stable equilibrium point:

$$\Lambda(t \in [0,1] : |X^{\varepsilon}_{tT} - (x_l 1_{\{x \in D_l\}} + x_r 1_{\{x \in D_r\}})| > \delta) \to 0$$

This observation is at the basis of Freidlin's law of quasideterministic periodic motion discussed in the subsequent section. The lesson it teaches is this: to observe switching of the position to the energetically most favorable well, $T$ should be larger than some critical level $\mathrm{e}^{\lambda/\epsilon}$. Measuring time in exponential scales by $\mu$ through the equation $T^{\varepsilon} = \mathrm{e}^{\mu/\varepsilon}$, the condition becomes $\mu > \lambda$.

## Stochastic Resonance for Landscapes, Frozen on Half-Periods

This particular case has analytical advantages, since it allows one to employ classical techniques of semigroup and operator theory. The situation is the following: let $U$ be a double-well potential with minima $x_l = -1$ and $x_r = 1$ and a saddle point at the origin. We assume that $U(x) \to \infty$ as $|x| \to \infty$ and $U(-1) = -V/2 = -\overline{V}_l/2$, $U(1) = -v/2 = -\overline{V}_r/2$, $U(0) = 0$, and $0 < v < V$. We define the 1-periodic potential by $U(t,x) = U(t + 1/2, -x)$. Hence on each half-period the corresponding diffusion is time homogeneous. The critical level $\lambda$ is then easily defined by $\lambda = v$, that is, twice the depth of the shallow well. By letting

$$\phi(t) = \begin{cases} -1 & \text{for } t \in [k, k + \frac{1}{2}) \\ 1 & \text{for } t \in [k + \frac{1}{2}, k + 1), \quad k = 0, 1, 2, \ldots \end{cases}$$

the periodic function which describes the location of the global minimum of the potential, we get in the small noise limit

$$\Lambda(t \in [0,1] : |X^{\varepsilon}_{tT} - \phi(t)| > \delta) \to 0$$

in probability as $\varepsilon \to 0$. This result expresses Freidlin's law of quasideterministic motion: for large periods, the trajectories of the particle approach a periodic deterministic function. But the sense in which this notion measures periodicity does not take into account that for large periods short excursions to the wrong well may occur in an erratic way without counting much for Lebesgue measure of time. In fact, if the period is too large, that is, $\mu > V$, the time available in one period permits the exit of not only the shallow well but also that of the deep well. So, whatever the starting position of

the particle is, the number of observed transitions in one half period becomes very large. Indeed the first time $\xi$ the particle starting in $x_l$ hits again $x_l$ after visiting the position $x_r$ satisfies

$$\mathbb{E}(\xi) = e^{v/\varepsilon} + e^{V/\varepsilon} < T^\varepsilon = e^{\mu/\varepsilon}$$

The motion of the particle appears more chaotical than periodic: noise intensity is too large compared to period length. We avoid this range of chaotic spontaneous transitions by defining the resonance interval $I_R = [v, V]$, as the range of admissible energy parameters $\mu$ for randomly periodic behavior. In this regime, the trajectories possess periodicity properties. In these terms the resonance point describes the tuning rate $\mu_R \in I_R$ for which the stochastic response to weak external periodic forcing is optimal. To make sense, this point has to refer to some measure of quality for periodicity of random trajectories. In the huge physics literature concerning resonance, two families of criteria can be distinguished. The first one is based on invariant measures and spectral properties of the infinitesimal generator associated with the diffusion $X^\varepsilon$. Now, $X^\varepsilon$ is not Markovian and consequently does not admit invariant measures. But by taking into account deterministic motion of time in the interval of periodicity and considering the process $Z_t = (t \bmod(T^\varepsilon), X_t)$, we obtain a Markov process with an invariant measure $\nu_t(x)\mathrm{d}x$. In other words, the law of $X_t \sim \nu_t(x)\mathrm{d}x$ and the law of $X_{t+T} \sim \nu_{t+T}(x)\mathrm{d}x$, under this measure, are the same for all $t \geq 0$. Let us present the most important ones:

- the spectral power amplification (SPA) which plays an eminent role in the physics literature describes the energy carried by the spectral component of the averaged trajectories of $X^\varepsilon$ corresponding to the period:

$$M_{\mathrm{SPA}}(\varepsilon, T) = \left| \int_0^1 \mathbb{E}_\nu[X^\varepsilon_{sT}] e^{2\pi i s}\, \mathrm{d}s \right|^2$$

- the SPA-to-noise ratio, giving the ratio of the amplitude of the response and the noise intensity, which is also related to the signal-to-noise ratio:

$$M_{\mathrm{SPN}}(\varepsilon, T) = M_{\mathrm{SPA}}(\varepsilon, T)/\varepsilon^2$$

- the total energy of the averaged trajectories

$$M_{\mathrm{EN}}(\varepsilon, T) = \int_0^1 (\mathbb{E}_\nu[X_{sT}])^2\, \mathrm{d}s$$

The second family of criteria is more probabilistic. It refers to quality measures based on transition times between the domains of attraction of the local minima, residence times distributions measuring the time spent in one well between two transitions, or interspike times. This family is certainly less popular in the physics community.

However, measures related to invariant measures may suffer from robustness deficiency (Imkeller and Pavlyukevich 2002). To explain what we mean by robustness, let us introduce a model reduction first discussed by McNamara and Wiesenfeld (1989). Instead of studying the diffusion $X^\varepsilon$ in the double-well landscape, they introduce a two-state Markov chain $Y^\varepsilon$ (**Figure 5**) the dynamics of which just takes account of the domain of attraction the diffusion is in, and therefore with state space $\{-1, 1\}$. A reasonable choice of the infinitesimal generator should retain the dynamics of the diffusion's transitions characterized by Kramers' rate. We may take

$$Q(t) = \begin{pmatrix} -\varphi & \varphi \\ \psi & -\psi \end{pmatrix}, \quad 0 \leq t \leq \frac{T}{2}$$
$$Q(t) = \begin{pmatrix} -\psi & \psi \\ \varphi & -\varphi \end{pmatrix}, \quad \frac{T}{2} \leq t < T$$

periodically continued on $\mathbb{R}_+$. Here, $\varphi = p e^{-V/\varepsilon}$ and $\psi = q e^{-v/\varepsilon}$. The prefactors of subexponential order are beyond the scope of large deviation theory. They are related to the curvature of the potential in the

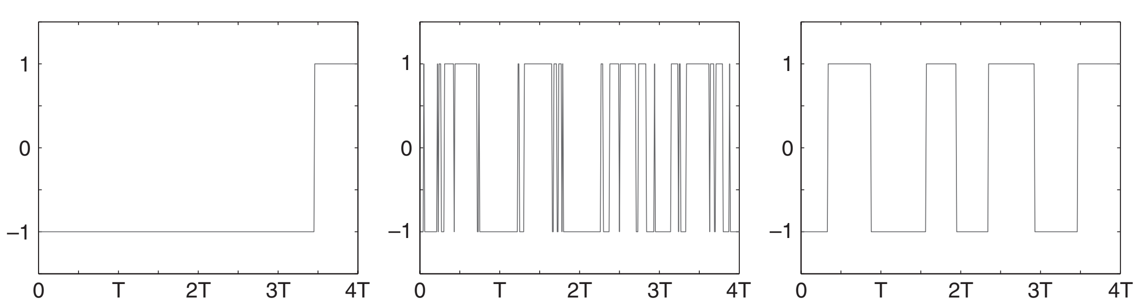



**Figure 5** Resonance pictures for Markov chain.

minima and the saddle point of the landscape and given by

$$p = \frac{1}{2\pi}\sqrt{U''(-1)|U''(0)|}$$
$$q = \frac{1}{2\pi}\sqrt{U''(1)|U''(0)|}$$

On the intervals $[kT/2,(k+1)T/2[, k \geq 0,$ the Markov chain $Y^\varepsilon$ is time-homogeneous and its transition probabilities can be expressed in terms of $\varphi$ and $\psi$. For instance, the probability with which the chain jumps from state –1 to state +1 in the time window $[t, t+h]$ equals $\varphi h + o(h)$, if this time interval is contained in $[kT/2,(k+1)T/2[$ for some even $k$. The stationary measure of the Markov chain denoted by $\nu$ can be explicitly calculated, and so can the classical quality measures based on the spectral notions. For instance, the spectral power amplification coefficient equals

$$M_{\text{SPA}}(\varepsilon, T) = \left| \int_0^1 \mathbb{E}_\nu[Y^\varepsilon_{st}]e^{2\pi is}\,ds \right|^2 = \frac{4}{\pi^2}\frac{T^2(\varphi-\psi)^2}{(\varphi+\psi)^2T^2+\pi^2}$$

This simple expression admits asymptotically a unique maximum which exhibits the resonance point:

$$T^\varepsilon_{\text{opt}} = \frac{\pi}{\sqrt{2pq}}\sqrt{\frac{v}{V-v}}e^{(V+v)/2\varepsilon}\left\{1+\mathcal{O}(e^{-(V-v)/\varepsilon})\right\}$$

The optimal period is then exponentially large – as was suggested by large deviation theory – and the growth rate is the sum of the two wells' depths. The simple Markov chain model is popular since the usual physical quantities are easy computable and since it is believed to mimic the dynamics of a Brownian particle in the corresponding double-well landscape. However, the models are not as similar as expected (Freidlin 2003). Indeed, in a reasonably large time window around the resonance point for $Y^\varepsilon$, the tuning picture of the spectral power amplification for the diffusion is different. Under weak regularity conditions on the potential, it exhibits strict monotonicity in the window. Hence, optimal tuning points for diffusion and Markov chain differ essentially. In other words, the SPA tuning behavior of the diffusion is not robust for passage to the reduced model. This strange deficiency is difficult to explain. The main reason of this subtle effect appears to be that the diffusive nature of the Brownian particle is neglected in the reduced model. In order to point out this feature, we may compute the SPA coefficient of $g(X^\varepsilon)$, where $g$ is a particular function designed to cut out the small fluctuations of the diffusion in the neighborhood of the bottoms of the wells, by identifying all states there. So $g(x) = -1$ (resp. 1) in some neighborhood of –1 (resp. 1) and otherwise $g$ is the identity. This results in

$$\widetilde{M}_{\text{SPA}}(\varepsilon, T) = \left| \int_0^1 \mathbb{E}_\nu g(X^\varepsilon_{sT})e^{2\pi is}\,ds \right|^2$$

In the small noise limit this quality function admits a local maximum close to the resonance point of the reduced model: the growth rate of $T^\varepsilon_{\text{opt}}$ is also given by the sum of the wells' depths. So the lack of robustness seems to be due to the small fluctuations of the particle in the wells' bottoms. In any case, this clearly calls for other quality measures to be used to transfer properties of the reduced model to the original one. Our discussion indicates that due to their emphasis on the pure transition dynamics, the second family of quality measures should be used. For these notions there is no need to restrict to landscapes frozen in time-independent potential states on half-period intervals.

## Stochastic Resonance for Continuously Varying Landscapes

From now on the potential $U(t, x)$ is supposed to be continuously varying in $(t, x)$. For simplicity, its local minima are assumed to be located at $\pm 1$, and its only saddle point at 0, independently of time. So the only meta-stable states on the whole time axis are $\pm 1$. Let us denote by $\Delta_-(t)$ (resp. $\Delta_+(t)$) the depth of the left (resp. right) well at time $t$. Together with $U$, these functions are continuous and 1-periodic. Assume that they are strictly monotonous between their global extrema. Let us now come back to the motion of a Brownian particle in this landscape. The exit time law by Eyring–Kramers–Freidlin entails that trajectories get close to the global minimum, if the period is large enough. Stated as before in exponential rates $T = e^{\mu/\varepsilon}$, with $\mu \geq \max_{i=\pm}\sup_{t\geq 0} 2\Delta_\pm(t)$, that is, $\mu$ exceeds the maximal work needed to cross the barrier, the particle often switches between the two wells and should stay close to the deepest position in the landscape. This position being described by the function $\phi(t) = 21_{\{\Delta_+(t)>\Delta_-(t)\}} - 1$, we get in the small noise limit

$$\Lambda(t \in [0,1] : |X^\varepsilon_{tT} - \phi(t)| > \delta) \to 0$$

in probability. But on these long timescales, many short excursions to the wrong well are observed, and trajectories look chaotical instead of periodic. So we

have to look at smaller periods even at the cost that the particle may not stay close to the global minimum. Let us study the transition dynamics. Assume that the starting point is –1 corresponding to the bottom of the deep well. If the depth of the well is always larger than $\mu = \varepsilon \log T^\varepsilon$, the particle has too little time during one period to climb the barrier, and should stay in the starting well. If, on the contrary, the minimal work to leave the starting well, given by $2\Delta_-(s)$, becomes smaller than $\mu$ at some time $s$, then the transition can and will happen. More formally, for $\mu \in [\inf_{t\geq 0} 2\Delta_-(t), \sup_{t\geq 0} 2\Delta_-(t)]$, we define (**Figure 6**).

$$a_\mu^-(s) = \inf\{t \geq s : 2\Delta_-(t) \leq \mu\}$$

The first transition time from –1 to 1 denoted $\tau_+$ has the following asymptotic behavior as $\varepsilon \to 0: \tau_+/T^\varepsilon \to a_\mu^-(0)$. At the second transition the particle returns to the starting well. If $a_\mu^+$ is defined analogously with respect to the depth function $\Delta_+$, this transition will occur near the deterministic time $a_\mu^+(a_\mu^-(s))T^\varepsilon$. In order to observe periodicity, and to exclude chaoticity from all parts of its trajectories, the particle has to stay for some time in the other well before returning. This will happen under the assumption $2\Delta_+(a_\mu(0)) > \mu$, that is, the right well is the deep one at transition time. In fact, we can define the resonance interval $I_R$ (**Figure 7**), as the set of all scales $\mu$ for which trajectories exhibit periodicity in the small noise limit, by

$$I_R = \left[\max_{i=\pm}\inf_{t\geq 0} 2\Delta_i(t), \inf_{t\geq 0}\max_{i=\pm} 2\Delta_i(t)\right]$$

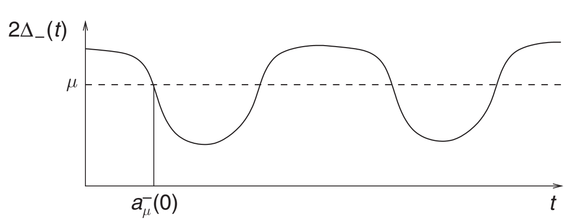



**Figure 6** Definition of $a_\mu^-$.

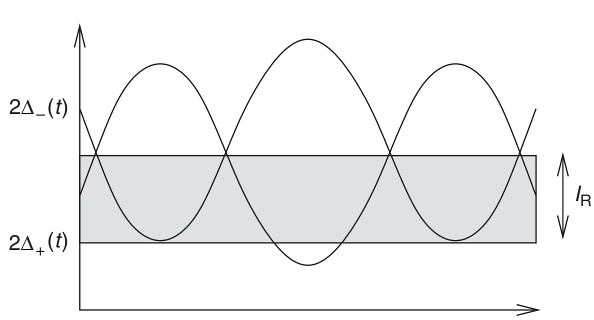



**Figure 7** Resonance interval.

On this interval they get close to deterministic periodic ones. Again, periodicity is quantified by a quality measure, to be maximized in order to obtain resonance as the best possible response to periodic forcing. One interesting measure is based on the probability that random transitions happen in some small time window around a deterministic time, in the small noise limit (Herrmann and Inkeller 2005). Formally, for $h > 0$, the measure gives

$$M_h(\varepsilon, T) = \min_{i=\pm} \mathbb{P}_i(\tau_\mp/T^\varepsilon \in [a_\mu^i - h, a_\mu^i + h])$$

where $\mathbb{P}_i$ is the law of the diffusion starting in $i$. In the small noise limit, this quality measure tends to 1, and optimal tuning can be related to the exponential rate at which this happens. This is due to the following large deviations principle:

$$\lim_{\varepsilon\to 0} \varepsilon \log(1 - M_h(\varepsilon, T)) = \max_{i=\pm}\{\mu - 2\Delta_i(a_\mu^i - h)\}$$

for $\mu \in I_R$, with uniform convergence on each compact subset of $I_R$. The result is established using classical large-deviation techniques applied to locally time homogeneous approximations of the diffusion. Maximizing the transition probability in the time window position means minimizing the default rate obtained by the large deviations principle. This can be easily achieved. In fact, if the window length $2h$ is small, then $\mu - 2\Delta_i(a_\mu^i - h) \approx 2h\Delta_i'(a_\mu^i)$, since $2\Delta_i(a_\mu^i) = \mu$ by definition. The value $\Delta_i'(a_\mu^i)$ is negative, so we have to find the position where its absolute value is maximal. In this position the depth of the starting well has the most rapid drop under the level $\mu$, characterizing the link between the noise intensity and the period. So the transition time is best concentrated around it.

It is clear that a good candidate for the resonance point is given by the eventually existing limit of the global minimizer $\mu_R(h)$ as the window length $h$ tends to 0. This limit is therefore called the resonance point of the diffusion with time-periodic landscape $U$. Let us note that for sinusoidal depth functions

$$\Delta_-(t) = \frac{V+v}{4} + \frac{V-v}{4}\cos(2\pi t)$$

and

$$\Delta_+(t) = \Delta_-(t+\pi)$$

the optimal tuning is given by $T^\varepsilon = \exp \mu_R/\varepsilon$ with $\mu_R = (v+V)/2$. This optimal rate is equivalent to the optimal rate given by the SPA coefficient of the reduced dynamics' Markov chain in the preceding section.

The big advantage of the quality measure $M_h$ is its robustness. Indeed, consider the reduced model

consisting of a two-state Markov chain with infinitesimal generator

$$Q(t) = \begin{pmatrix} -\varphi(t) & \varphi(t) \\ \psi(t) & -\psi(t) \end{pmatrix}$$

where $\varphi(t) = \exp -2\Delta_-(t/T)/\varepsilon$ and $\psi(t) = \exp -2\Delta_+(t/T)/\varepsilon$. The law of transition times of this Markov chain is readily computed from Laplace transforms. Normalized by $T^\varepsilon$ it converges to $a^i_\mu$. This calculation even reveals a rigorous underlying pattern for the second- and higher-order transition times interpreting the interspike distributions of the physics literature. The dynamics of diffusion and Markov chain are similar. Resonance points provided by $M_b$ for the diffusion and its analog for the Markov chain agree.

## Related Notions: Synchronization

In the preceding sections, we interpreted stochastic resonance as optimal response of a randomly perturbed dynamical system to weak periodic forcing, in the spirit of the physics literature (see Gammaitoni *et al.* (1998)). Our crucial assumption concerned the barrier heights a Brownian particle has to overcome in the potential landscape of the dynamical system: it is uniformly lower bounded in time. Measures for the quality of tuning were based on essentially two concepts: one concerning spectral criteria, with the spectral power amplification as most prominent member, the other one concerning the pure transitions dynamics between the domains of attraction of the local minima. A number of different criteria can be used to create an optimal tuning between the intensity of the noise perturbation and the large period of the dynamical system. The relations have to be of an exponential type $T = \exp \mu/\varepsilon$, since the Brownian particle needs exponentially long times to cross the barrier separating the wells according to the Eyring–Kramers–Freidlin transition law. Our barrier height assumption seems natural in many situations, but can fail in others. If it becomes small periodically, and eventually scales with the noise-intensity parameter, the Brownian particle does not need to wait an exponentially long time to climb it. So periodicity obtains for essentially smaller timescales. In this setting, the slowness of periodic forcing may also be assumed to be essentially subexponential in the noise intensity.

If it is fast enough to allow for substantial changes before large deviation effects can take over, we are in the situation of Berglund and Gentz (2002). They in fact consider the case in which the barrier between the wells becomes low twice per period, to the effect of modulating periodically a bifurcation parameter: at time zero the right-hand well becomes almost flat, and at the same time the bottom of the well and the saddle approach each other; half a period later, a spatially symmetric scenario is encountered. In this situation, there is a threshold value for the noise intensity under which transitions become unlikely. Above this threshold, the trajectories typically contain two transitions per period. Results are formulated in terms of concentration properties for random trajectories. The intuitive picture is this: with overwhelming probability, sample paths will be concentrated in spacetime sets scaling with the small parameters of the problem. In higher dimensions, these sets may be given by adiabatic or center manifolds of the deterministic system, which allow model reduction of higher-dimensional systems to lower-dimensional ones. Asymptotic results hold for any choice of the small parameters in a whole parameter region. A passage to the small noise limit as for optimal tuning in the preceding sections is not needed.

Related problems studied by Berglund and Gentz in the multidimensional case concern the noise-induced passage through periodic orbits, where unexpected phenomena arise. Here, as opposed to the classical Freidlin–Wentzell theory, the distribution of first-exit points depends nontrivially on the noise intensity. Again aiming at results valid for small but nonvanishing parameters in subexponential scale ranges, they investigate the density of first-passage times in a large regime of parameter values, and obtain insight into the transition from the stochastic resonance regime into the synchronization regime.

*See also:* Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Magnetic Resonance Imaging; Spectral Theory for Linear Operators; Stochastic Differential Equations.

## Further Reading

Benzi R, Sutera A, and Vulpiani A (1981) The mechanism of stochastic resonance. *Journal of Physics A* 14: L453–L457.

Berglund N and Gentz B (2002) A sample-paths approach to noise-induced synchronization: stochastic resonance in a double-well potential. *Annals of Applied Probability* 12(4): 1419–1470.

Bovier A, Eckhoff M, Gayrard V, and Klein M (2004) Metastability in reversible diffusion processes. I. Sharp asymptotics for capacities and exit times. *Journal of the European Mathematical Society* 6(4): 399–424.

Freidlin MI (2000) Quasi-deterministic approximation, metastability and stochastic resonance. *Physica D* 137(3–4): 333–352.

Freidlin M (2003) Noise sensitivity of stochastic resonance and other problems related to large deviations. In: Sri Namachchivaya and Lin YK (eds.) *IUTAM Symposium on Nonlinear Stochastic Dynamics*, Solid Mech. Appl, vol. 110, pp. 43–55. Dordrecht: Kluwer Academic.
Freidlin MI and Wentzell AD (1998) *Random Perturbations of Dynamical Systems*, 2nd edn. New York: Springer.
Gammaitoni L, Hänggi P, Jung P, and Marchesoni F (1998) Stochastic resonance. *Reviews of Modern Physics* 70(1): 223–287.
Gardiner CW (2004) *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Springer Series in Synergetics, 3rd edn., vol. 13. Berlin: Springer.
Herrmann S and Imkeller P (2005) The exit problem for diffusions with time-periodic drift and stochastic resonance. *Annals of Applied Probability* 15(1A): 39–68.
Imkeller P and Pavlyukevich I (2002) Model reduction and stochastic resonance. *Stochastics and Dynamics* 2(4): 463–506.
Mazo RM (2002) *Brownian Motion: Fluctuations, Dynamics, and Applications*. International Series of Monographs on Physics, vol. 112. New York: Oxford University Press.
McNamara B and Wiesenfeld K (1989) Theory of stochastic resonance. *Physical Review A* 39(9): 4854–4869.
Nelson E (1967) *Dynamical Theories of Brownian Motion*. Princeton, NJ: Princeton University Press.
Schweitzer F (2003) *Brownian Agents and Active Particles*, Springer Series in Synergetics, (Collective dynamics in the natural and social sciences, With a foreword by J. Doyne Farmer). Berlin: Springer.

**Strange Attractors** *see* Lyapunov Exponents and Strange Attractors

# String Field Theory

**L Rastelli**, Princeton University, Princeton, NJ, USA



## Introduction

String field theory (SFT) is the second-quantized approach to string theory. In the usual, first-quantized, formulation of string perturbation theory, one postulates a recipe for the string $S$-matrix in terms of a sum over two-dimensional (2D) world sheets embedded in spacetime. Very schematically,

$$\langle\langle V_1(k_1)\ldots V_n(k_n)\rangle\rangle = \sum_{\text{topologies}} g_s^{-\chi}\int [d\mu_\alpha]\,\langle V_1(k_1)\ldots V_n(k_n)\rangle_{\{\mu_\alpha\}} \qquad [1]$$

Here the left-hand side stands for the $S$-matrix of the physical string states $\{V_a(k_a)\}$. The symbol $\langle\ldots\rangle_{\{\mu_\alpha\}}$ denotes a correlation function on the 2D world sheet, which is a punctured Riemann surface of Euler number $\chi$ and given moduli $\{\mu_\alpha\}$. In SFT, one aims to recover this standard prescription from the Feynman rules of a second-quantized spacetime action $S[\Phi]$. The string field $\Phi$, the fundamental dynamical variable, can be thought of as an infinite-dimensional array of spacetime fields $\{\phi^i(x^\mu)\}$, one field for each basis state in the Fock space of the first-quantized string.

The most straightforward way to construct $S[\Phi]$ uses the unitary light-cone gauge. Light-cone SFT is an almost immediate transcription of Mandelstam's light-cone diagrams in a second-quantized language. While often useful as a bookkeeping device, light-cone SFT seems unlikely to represent a real improvement over the first-quantized approach. By contrast, from our experience in ordinary quantum field theory, we should expect Poincaré-covariant SFTs to give important insights into the issues of vacuum selection, background independence and the nonperturbative definition of string theory.

Covariant SFT actions are well established for the open (Witten 1986), closed (Zwiebach 1993) and open/closed (Zwiebach 1998) bosonic string. These theories are based on the BRST formalism, where the world sheet variables include the $bc$ ghosts introduced in gauge-fixing the world sheet metric to the conformal gauge $g_{ab}\sim\delta_{ab}$. (An alternative approach (Hata *et al.*), based on covariantizing light-cone SFT, will not be described in this article.) Much less is presently known for the superstring: classical actions have been established for the Neveu-Schwarz sector of the open superstring (Berkovits 2001) and for the heterotic string (Berkovits *et al.* 2004).

During the first period of intense activity in SFT (1985–1992), the covariant bosonic actions were constructed and shown to pass the basic test of reproducing the $S$-matrix [1] to each order in the perturbative expansion. The more recent revival of the subject (since 1999) was triggered by the realization that SFT contains nonperturbative information as well: D-branes emerge as solitonic solutions of the classical equations of motion in

open SFT (OSFT). We can hope that the nonperturbative string dualities will also be understood in the framework of SFT, once covariant SFTs for the superstring are better developed.

In this article, we review the basic formalism of covariant SFT, using for illustration purposes the simplest model – cubic bosonic OSFT. We then briefly sketch the generalization to bosonic SFTs that include closed strings. Finally, we turn to the subjects of classical solutions in OSFT and the physics of the open-string tachyon.

## Open Bosonic SFT

The standard formulation of string theory starts with the choice of an on-shell spacetime background where strings propagate. In the bosonic string, the closed string background is described by a conformal field theory of central charge 26 (the "matter" CFT). The total world sheet CFT is the direct sum of this matter CFT and of the universal ghost CFT, of central charge –26. To describe open strings, we must further specify boundary conditions for the string endpoints. The open-string background is encoded in a boundary CFT (BCFT), a CFT defined in the upper-half plane, with conformal boundary conditions on the real axis (see Boundary Conformal Field Theory in this encyclopedia). In modern language, the choice of BCFT corresponds to specifying a D-brane state.

In classical OSFT, we fix the closed-string background (the bulk CFT) and consider varying the D-brane configuration (the boundary conditions). To lowest order in $g_s$, we can neglect the backreaction of the D-brane on the closed-string fields, since this is a quantum effect from the open-string viewpoint. Let us prepare the ground by recalling the standard $\sigma$-model philosophy. To describe off-shell open-string configurations, we should allow for general (not necessarily conformal) boundary conditions. We can imagine to proceed as follows:

1. We choose an initial open-string background, a reference BCFT that we shall call $\text{BCFT}_0$. For example, a D$p$ brane in flat 26 dimensions (Neumann boundary conditions on $p+1$ coordinates, Dirichlet on $25-p$ coordinates).
2. We then write a basis of boundary perturbations around this background. Taking, for example, $\text{BCFT}_0$ to be a D25 brane in flat space, the world sheet action $S_{\text{WS}}$ takes the schematic form

$$S_{\text{WS}} = \frac{1}{2\pi\alpha'}\int_{\text{UHP}} \partial X_\mu \bar{\partial} X^\mu + \int_{\boldsymbol{R}} \tilde{T}(x^\mu) + \tilde{A}_\nu(x^\mu)\partial X^\nu + \tilde{B}_\nu(x^\mu)\partial^2 X^\nu + \cdots \qquad [2]$$

Here to the standard free bulk action (integrated over the upper-half complex plane UHP) we have added perturbation localized on the real axis $\boldsymbol{R}$. Notice that the basis of perturbations depends on the chosen $\text{BCFT}_0$.
3. We interpret the coefficients $\{\tilde{\phi}^i(x^\mu)\}$ of the perturbations as spacetime fields. (The tilde on $\tilde{\phi}^i(x)$ serves as a reminder that these fields are not quite the same as the fields $\phi^i(x)$ that will appear in the OSFT action). We are after a spacetime action $S[\{\tilde{\phi}^i\}]$ such that solutions of its classical equations of motion correspond to conformal boundary conditions:

$$\frac{\delta S}{\delta \tilde{\phi}^i} = 0(\text{spacetime}) \leftrightarrow \beta_i[\{\tilde{\phi}^j\}] = 0 \text{ (world sheet)} \qquad [3]$$

We recognize in [2] the familiar open-string tachyon $\tilde{T}(x)$ and gauge field $\tilde{A}_\mu(x)$, which are the lowest modes in an infinite tower of fields. Relevant perturbations on the world sheet (with conformal dimension $h<1$) correspond to tachyonic fields in spacetime ($m^2<0$), whereas marginal world sheet perturbations ($h=1$) give massless spacetime fields. To achieve a complete description, we must include all the higher massive open-string modes as well, which correspond to nonrenormalizable boundary perturbations ($h>1$). In the traditional $\sigma$-model approach, this appears like a daunting task. The formalism of OSFT will automatically circumvent this difficulty.

### The Open-String Field

In covariant SFT the reparametrization ghosts play a crucial role. The ghost CFT consists of the Grassmann odd fields $b(z), c(z), \bar{b}(\bar{z}), \bar{c}(\bar{z})$, of dimensions $(2,0), (-1,0), (0,2), (0,-1)$, respectively. The boundary conditions on the real axis are $b=\bar{b}, c=\bar{c}$. The state space $\mathcal{H}_{\text{BCFT}_0}$ of the full matter + ghost BCFT can be broken up into subspaces of definite ghost number,

$$\mathcal{H}_{\text{BCFT}_0} = \bigoplus_{G=-\infty}^{\infty} \mathcal{H}^{(G)}_{\text{BCFT}_0} \qquad [4]$$

We use conventions where the SL(2, $\boldsymbol{R}$) vacuum $|0\rangle$ carries zero ghost number, $G(|0\rangle)=0$, while $G(c)=+1$ and $G(b)=-1$. As is familiar from the first-quantized treatment, physical open-string states are identified with $G=+1$ cohomology classes of the BRST operator,

$$Q|V_{\text{phys}}\rangle = 0, \quad |V_{\text{phys}}\rangle \sim |V_{\text{phys}}\rangle + Q|\Lambda\rangle \qquad [5]$$
$$G(|V_{\text{phys}}\rangle) = +1$$

where the nilpotent BRST operator $Q$ has the standard expression

$$Q = \frac{1}{2\pi i}\oint (c\,T_{\text{matter}} + :bc\partial c:) \qquad [6]$$

Though not *a priori* obvious, it turns out that the simplest form of the OSFT action is achieved by taking as the fundamental off-shell variable an arbitrary $G = +1$ element of the first-quantized Fock space,

$$|\Phi\rangle \in \mathcal{H}^{(1)}_{\text{BCFT}_0} \qquad [7]$$

By the usual state–operator correspondence of CFT, we can also represent $|\Phi\rangle$ as a local (boundary) vertex operator acting on the vacuum,

$$|\Phi\rangle = \Phi(0)|0\rangle \qquad [8]$$

The open-string field $|\Phi\rangle$ is really an infinite-dimensional array of spacetime fields. We can make this transparent by expanding it as

$$|\Phi\rangle = \sum_i \int d^{p+1}k|\Phi_i(k)\rangle\,\phi^i(k_\mu) \qquad [9]$$

where $\{|\Phi_i(k)\rangle\}$ is some convenient basis of $\mathcal{H}^{(1)}_{\text{BCFT}_0}$ that diagonalizes the momentum $k_\mu$. The fields $\phi^i$ are *a priori* complex. This is remedied by imposing a suitable reality condition on the string field, which will be stated momentarily. Notice that there are many more elements in $\{|\Phi_i(k)\rangle\}$ than in the physical subspace (the cohomology classes of $Q$). Some of the extra fields will turn out to be nondynamical and could be integrated out, but at the price of making the OSFT action look much more complicated.

It is often useful to think of the string field in terms of its Schrödinger representation, that is, as a functional on the configuration space of open strings. Consider the unit half-disk in the upper-half plane, $D_H \equiv \{|z| \leq 1, \Im z \geq 0\}$, with the vertex operator $\Phi(0)$ inserted at the origin. Impose $\text{BCFT}_0$ open string boundary conditions for the fields $X(z,\bar{z})$ on the real axis (here $X(z,\bar{z})$ is a short-hand notation for all matter and ghost fields), and boundary conditions $X(\sigma) = X_b(\sigma)$ on the curved boundary of $D_H, z = \exp(i\sigma), 0 \leq \sigma \leq \pi$. The path integral over $X(z,\bar{z})$ in the interior of the half-disk assigns a complex number to any given $X_b(\sigma)$, so we obtain a functional $\Phi[X_b(\sigma)]$. This is the Schrödinger wave function of the state $\Phi(0)|0\rangle$. Thus, we can think of open-string functionals $\Phi[X_b(\sigma)]$ as the fundamental variables of OSFT. This is as it should be: the first-quantized wave functions are promoted to dynamical fields in the second-quantized theory. Finally, let us quote the reality condition for the string field, which takes a compact form in the Schrödinger representation:

$$\begin{aligned}\Phi[X^\mu(\sigma), b(\sigma), c(\sigma)]\\ = \Phi^*[X^\mu(\pi-\sigma), b(\pi-\sigma), c(\pi-\sigma)]\end{aligned} \qquad [10]$$

where the superscript $*$ denotes complex conjugation.

### The Classical Action

With all the ingredients in place, it is immediate to write the quadratic part of the OSFT action. The linearized equations of motion must reproduce the physical-state condition [5]. This suggests

$$S \sim \langle\Phi|Q|\Phi\rangle \qquad [11]$$

Here $\langle|\rangle$ is the usual BPZ inner product of $\text{BCFT}_0$, which is defined in terms of a two-point correlator on the disk, as we review below. The ghost anomaly implies that on the disk we must have $G_{\text{tot}} = +3$, which happily is the case in [11]. Moreover, since the inner product is nondegenerate, variation of [11] gives

$$Q|\Phi\rangle = 0 \qquad [12]$$

as desired. The equivalence relation $|V_{\text{phys}}\rangle \sim |V_{\text{phys}}\rangle + Q|\Lambda\rangle$ is interpreted in the second-quantized language as the spacetime gauge invariance

$$\delta_\Lambda|\Phi\rangle = Q|\Lambda\rangle, \quad |\Lambda\rangle \in \mathcal{H}^{(0)}_{\text{BCFT}_0} \qquad [13]$$

valid for the general off-shell field. This equation is a very compact generalization of the linearized gauge invariance for the massless gauge field. Indeed, focusing on the level-zero components, $|\Phi\rangle \sim A_\mu(x)(c\partial X^\mu)(0)|0\rangle$ and $|\Lambda\rangle \sim \lambda(x)|0\rangle$, we find $\delta A_\mu(x) = \partial_\mu\lambda(x)$. It is then plausible to guess that the nonlinear gauge invariance should take the form

$$\delta_\Lambda|\Phi\rangle = Q|\Lambda\rangle + |\Phi\rangle * |\Lambda\rangle - |\Lambda\rangle * |\Phi\rangle \qquad [14]$$

where $*$ is some suitable product operation that conserves ghost number

$$* : \mathcal{H}^{(n)}_{\text{BCFT}_0} \otimes \mathcal{H}^{(m)}_{\text{BCFT}_0} \to \mathcal{H}^{(n+m)}_{\text{BCFT}_0} \qquad [15]$$

Based on a formal analogy with 3D nonabelian Chern–Simons theory, Witten proposed the cubic action

$$S = -\frac{1}{g_o^2}\left(\frac{1}{2}\langle\Phi|Q|\Phi\rangle + \frac{1}{3}\langle\Phi|\Phi * \Phi\rangle\right) \qquad [16]$$

The string field $|\Phi\rangle$ is analogous to the Chern–Simons gauge potential $A = A_i dx^i$, the $*$ product to the $\wedge$ product of differential forms, $Q$ to the exterior derivative $d$, and the ghost number $G$ to the degree

of the form. The analogy also suggests a number of algebraic identities:

$$
\begin{aligned}
Q^2 &= 0 \\
\langle QA|B\rangle &= -(-1)^{G(A)}\langle A|QB\rangle \\
Q(A*B) &= (QA)*B + (-1)^{G(A)}A*(QB) \\
\langle A|B\rangle &= (-1)^{G(A)G(B)}\langle B|A\rangle \\
\langle A|B*C\rangle &= \langle B|C*A\rangle \\
A*(B*C) &= (A*B)*C
\end{aligned} \qquad [17]
$$

Note in particular the associativity of the $*$-product. It is straightforward to check that this algebraic structure implies the gauge invariance of the cubic action under [14]. A $*$-product satisfying all required formal properties can indeed be defined. The most intuitive presentation is in the functional language. Given an open-string curve $X(\sigma), 0 \le \sigma \le \pi$, we single out the string mid-point $\sigma = \pi/2$ and define the left and right "half-string" curves

$$
\begin{aligned}
X_{\mathrm{L}}(\sigma) &\equiv X(\sigma) && \text{for } 0 \le \sigma \le \frac{\pi}{2} \\
X_{\mathrm{R}}(\sigma) &\equiv X(\pi - \sigma) && \text{for } \frac{\pi}{2} \le \sigma \le \pi
\end{aligned} \qquad [18]
$$

A functional $\Phi[X(\sigma)]$ can, of course, be regarded as a functional of the two half-strings, $\Phi[X] \to \Phi[X_{\mathrm{L}}, X_{\mathrm{R}}]$. We define

$$
(\Phi_1 * \Phi_2)[X_{\mathrm{L}}, X_{\mathrm{R}}] \equiv \int [\mathrm{d}Y]\Phi_1[X_{\mathrm{L}}, Y]\Phi_2[Y, X_{\mathrm{R}}] \qquad [19]
$$

where $\int [\mathrm{d}Y]$ is meant as the functional integral over the space of half-strings $Y(\sigma)$, with $Y(\pi/2) = X_{\mathrm{L}}(\pi/2) = X_{\mathrm{R}}(\pi/2)$. Figure 1a shows two open strings interacting (to form a single open string) if and only if the right half of the first string precisely overlaps with the left half of the second string. Associativity is transparent (Figure 1b).

We can now translate this formal construction in the precise CFT language. Very generally, an $n$-point vertex of open strings can be defined by specifying an $n$-punctured disk, that is, a disk with marked points on the boundary (punctures) and a choice of local coordinates around each puncture. Local coordinates are essential since we are dealing with off-shell open-string states. The BPZ inner product (two-point vertex) is given by

$$
\begin{aligned}
\langle \Phi_1|\Phi_2\rangle &\equiv \langle I \circ \Phi_1(0)\ \Phi_2(0)\rangle_{\mathrm{UPH}} \\
I(z) &= -\frac{1}{z}
\end{aligned} \qquad [20]
$$

The symbol $f \circ \Phi(0)$, where $f$ is a complex map, means the conformal transform of $\Phi(0)$ by $f$. For example, if $\Phi$ is a dimension-$d$ primary field, then $f \circ \Phi(0) = f'(0)^d \Phi(f(0))$. If $\Phi$ is nonprimary, the transformation rule will be more complicated and involve extra terms with higher derivatives of $f$. By performing the $\mathrm{SL}(2, \mathbf{C})$ transformation

$$
w = h(z) \equiv \frac{1 + \mathrm{i}z}{1 - \mathrm{i}z} \qquad [21]
$$

we can represent the two-point vertex as a correlator on the unit disk $D = \{|w| \le 1\}$,

$$
\begin{aligned}
\langle \Phi_1|\Phi_2\rangle &= \langle f_1 \circ \Phi_1(0),\ f_2 \circ \Phi_2(0)\rangle_D \\
f_1(z_1) &= -h(z_1), \quad f_2(z_2) = h(z_2)
\end{aligned} \qquad [22]
$$

The vertex operators are inserted as $w = -1$ and $w = +1$ on $D$ (see Figure 2a) and correspond to the two open strings at (Euclidean) world sheet time $\tau = -\infty$ (we take $z = \exp(\mathrm{i}\sigma + \tau)$). The left half of $D$ is the world sheet of the first open string; the right half of $D$ is the world sheet of the second string. The two strings meet at $\tau = 0$ on the imaginary $w$ axis. The three-point Witten vertex is given by

$$
\begin{aligned}
&\langle \Phi_1, \Phi_2, \Phi_3\rangle \\
&\quad \equiv \langle g_1 \circ \Phi_1(0) g_2 \circ \Phi_2(0) g_3 \circ \Phi_3(0)\rangle_D
\end{aligned} \qquad [23]
$$

where

$$
\begin{aligned}
g_1(z_1) &= \mathrm{e}^{2\pi\mathrm{i}/3}\left(\frac{1 + \mathrm{i}z_1}{1 - \mathrm{i}z_1}\right)^{2/3} \\
g_2(z_2) &= \left(\frac{1 + \mathrm{i}z_2}{1 - \mathrm{i}z_2}\right)^{2/3} \\
g_3(z_3) &= \mathrm{e}^{-2\pi\mathrm{i}/3}\left(\frac{1 + \mathrm{i}z_3}{1 - \mathrm{i}z_3}\right)^{2/3}
\end{aligned} \qquad [24]
$$

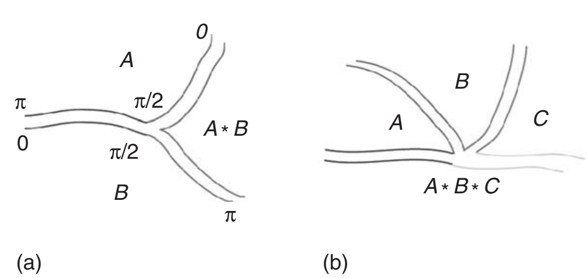



**Figure 1** Midpoint overlaps of open strings.

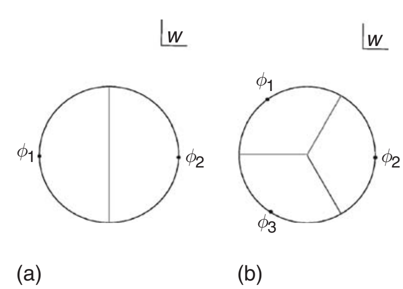



**Figure 2** Representation of the quadratic and cubic vertices as 2- and 3-punctured unit disks.

The 3-punctured disk is depicted in **Figure 2b**, and describes the symmetric mid-point overlap of the three strings at $\tau = 0$. Finally, the relation between the three-point vertex and the $*$-product is

$$\langle \Phi_1 | \Phi_2 * \Phi_3 \rangle \equiv \langle \Phi_1, \Phi_2, \Phi_3 \rangle \quad [25]$$

Knowledge of the right-hand side (RHS) in [25] for all $\Phi$ allows to reconstruct the $*$-product. All formal properties [17] are easily shown to hold in the CFT language. This completes the definition of the OSFT action.

Evaluation of the classical action is completely algorithmic and can be carried out for arbitrary massive states, with no fear of divergences, since in all required correlators the operators are inserted well apart from each other.

#### Quantization

Quantization is defined by the path integral over the second-quantized string field. The first step is to deal with the gauge invariance [14] of the classical action. The gauge symmetry is reducible: not all gauge parameters $\Lambda^{(0)}$ (the superscript labels ghost number) lead to a gauge transformation. This is clear at the linearized level; indeed, if $\Lambda^{(0)} = Q\Lambda^{(-1)}$, then $\delta_{\Lambda^{(0)}}\Phi^{(1)} = Q^2\Lambda^{(0)} = 0$. Thus, the set $\{\Lambda^{(0)}\}$ gives a redundant parametrization of the gauge group. Characterizing this redundancy is somewhat subtle, since fields of the form $\Lambda^{(-1)} = Q\Lambda^{(-2)}$ do not really lead to a redundancy in $\Lambda^{(0)}$, and so on, *ad infinitum*. It is clear that we need to introduce an infinite tower of (second-quantized) ghosts for ghosts.

The Batalin–Vilkovisky formalism is a powerful way to handle the problem. The basic object is the master action $S(\phi^s, \phi^*_s)$, which is a function of the "fields" $\phi^s$ and of the "antifields" $\phi^*_s$. Each field is paired with a corresponding antifield of opposite Grassmanality. ("Grassmanality" is defined to be even or odd: a Grassmann even (odd) field is a commuting (anticommuting) field). The master action must obey the boundary condition of reducing to the classical action when the antifields are set to zero. (Note that in general the set of fields $\phi^s$ will be larger than the set of fields $\phi^i$ that appear in the classical action). Independence of the $S$-matrix on the gauge-fixing procedure is equivalent to the BV master equation

$$\tfrac{1}{2}\{S, S\} = -\hbar \Delta S \quad [26]$$

The antibracket $\{,\}$ and the $\Delta$ operator are defined as

$$\{A, B\} \equiv \frac{\partial_r A}{\partial \phi^s}\frac{\partial_l B}{\partial \phi^*_s} - \frac{\partial_r A}{\partial \phi^*_s}\frac{\partial_l B}{\partial \phi^s}$$
$$\Delta \equiv \frac{\partial_r}{\partial \phi^s}\frac{\partial_l}{\partial \phi^*_s} \quad [27]$$

where $\partial_l$ and $\partial_r$ are derivatives from the left and from the right. It is often convenient to expand $S$ in powers of $\hbar$, $S = S_0 + \hbar S_1 + \hbar^2 S_2 + \cdots$, with

$$\{S_0, S_0\} = 0$$
$$\{S_0, S_1\} + \{S_0, S_1\} = -2\hbar\Delta S_0, \ldots \quad [28]$$

With these definitions in place, we shall simply describe the answer, which is extremely elegant. In OSFT the full set of fields and antifields is packaged in a single string field $|\Phi\rangle$ of unrestricted ghost number. If we write

$$|\Phi\rangle = |\Phi_-\rangle + |\Phi_+\rangle$$
$$\text{with } G(\Phi_-) \leq 1 \text{ and } G(\Phi_+) \geq 2 \quad [29]$$

all the fields are contained in $|\Phi_-\rangle$ and all the antifields in $|\Phi_+\rangle$. To make the pairing explicit, we pick a basis $\{|\Phi_s\rangle\}$ of $\mathcal{H}_{\mathrm{BCFT}_0}$, and define a conjugate basis $\{|\Phi^C_s\rangle\}$ by

$$\langle \Phi^C_r | \Phi_s \rangle = \delta_{rs} \quad [30]$$

Clearly, $G(\Phi^C_s) + G(\Phi_s) = 3$. Then

$$|\Phi_-\rangle = \sum_{G(\Phi_s)\leq 1} |\Phi_s\rangle \phi^s, \quad |\Phi_+\rangle = \sum_{G(\Phi_s)\leq 1} |\Phi^C_s\rangle \phi^*_s \quad [31]$$

Basis states $|\Phi_s\rangle$ with even (odd) ghost number $G(\Phi_s)$ are defined to be Grassmann even (odd). The full string field $|\Phi\rangle$ is declared to be Grassmann odd. It follows that $\phi^s$ is Grassmann even (odd) for $G(\Phi_s)$ odd (even), and that the corresponding antifield $\phi^*_s$ has the opposite Grassmanality of $\phi^s$, as it must be. With this understanding of $|\Phi\rangle$, the classical master action $S_0$ is identical in form to the Witten action [16]! The boundary condition is satisfied; indeed, setting $|\Phi_+\rangle = 0$, the ghost number anomaly implies that only the terms with $G = +1$ survive. The equation $\{S_0, S_0\} = 0$ follows from straightforward manipulations using the algebraic identities [17]. On the other hand, the issue of whether $\Delta S_0 = 0$, or whether instead quantum corrections are needed to satisfy full BV master equation, is more subtle and has never been fully resolved. The $\Delta$ operator receives singular contributions from the same region of moduli space responsible for the appearance of closed-string poles, which are discussed below. (See Thorn (1989) for a classic statement of this issue). It seems possible to choose a basis in $\mathcal{H}_{\mathrm{BCFT}_0}$ such that there are no quantum corrections to $S_0$ (Erler and Gross 2004). In the following we shall derive the Feynman rules implied by $S_0$ alone.

### SFT Diagrams and Minimal Area Metrics

Imposing the Siegel gauge condition $b_0\Phi = 0$, one finds the gauge-fixed action

$$S_{\rm gf} = -\frac{1}{g_o^2}\left(\frac{1}{2}\langle\Phi|c_0L_0|\Phi\rangle + \frac{1}{3}\langle\Phi|\Phi * \Phi\rangle + \langle\beta|\, b_0|\Phi\rangle\right) \qquad [32]$$

where $\beta$ is a Lagrangian multiplier. The propagator reads

$$\frac{b_0}{L_0} = b_0\int_0^\infty {\rm d}T\, {\rm e}^{-TL_0} \qquad [33]$$

Since $L_0$ is the first-quantized open-string Hamiltonian, ${\rm e}^{-TL_0}$ is the operator that evolves the open-string wave functions $\Psi[X(\sigma)]$ by Euclidean world sheet time $T$. It can be visualized as a flat rectangular strip of "horizontal" width $\pi$ and "vertical" height $T$. Each propagator comes with an antighost insertion

$$b_0 = \int_0^\pi b(\sigma) \qquad [34]$$

integrated on a horizontal trajectory.

The only elementary interaction vertex is the mid-point three-string overlap, visualized in Figure 3. We are instructed to draw all possible diagrams with given external legs (represented as semi-infinite strips), and to integrate over all Schwinger parameters $T_i \in [0,\infty)$ associated with the internal propagators. The claim is that this prescription reproduce precisely the first-quantized result [1]. This follows if we can show that (1) the OSFT Feynman rules give a unique cover of the moduli space of open Riemann surfaces; (2) the integration measure agrees with the measure $[{\rm d}\mu_\alpha]$ in [1]. The latter property holds because the antighost insertion [34] is precisely the one prescribed by the Polyakov formalism for integrating over the moduli $T_i$. To show point (1), we introduce the concept of minimal-area metrics, which has proved very fruitful. (Here and below, our discussion of minimal-area metrics will summarize ideas developed mainly by Zwiebach.) Quite generally, the Feynman rules of an SFT provide us with a cell decomposition of the appropriate moduli space of Riemann surfaces, a way to construct surfaces in terms of vertices and propagators. Given a Riemann surface (for fixed values of its complex moduli), the SFT must associate with it one and only one string diagram. The diagram has more structure than the Riemann surface: it defines a metric on it. In all known covariant SFTs, this is the metric of minimal area obeying suitable length conditions. Consider the following:

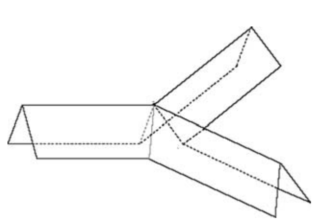


**Figure 3** The cubic vertex represented as the mid-point gluing of three strips.

**Minimal-area problem for open SFT** Let $\mathcal{R}_o$ be a Riemann surface with at least one boundary component and possibly punctures on the boundary. Find the (conformal) metric of minimal area on $\mathcal{R}_o$ such that all nontrivial Jordan open curves have length greater than or equal to $\pi$. (A curve is said to be nontrivial if it cannot be continuously shrunk to a point without crossing a puncture.)

An OSFT diagram (for fixed values of its $T_i$), defines a Riemann surface $\mathcal{R}_o$ endowed with a metric solving this minimal-area problem. This is the metric implicit in its picture: flat everywhere except at the conical singularities of defect angle $(n-2)\pi$ when $n$ propagators meet symmetrically. (For $n = 3$, these are the elementary cubic vertices; for $n > 3$, they are effective vertices, obtained when propagators joining cubic vertices collapse to zero length.) It is not difficult to see both that the length conditions are obeyed, and that the metric cannot be made smaller without violating a length condition. Conversely, any surface $\mathcal{R}_o$ endowed with a minimal-area metric, corresponds to an OSFT diagram. The idea is that the minimal-area metric must have open geodesics ("horizontal trajectories") of length $\pi$ foliating the surface. The geodesics intersect on a set of measure zero – the "critical graph" where the propagators are glued. Bands of open geodesics of infinite height are the external legs of the diagram, while bands of finite height are the internal propagators.

The single cover of moduli space is then ensured by an existence and uniqueness theorem for metrics solving the minimal-area problem for OSFT. These metrics are seen to arise from Jenkins–Strebel quadratic differentials. Existence shows that the Feynman rules of OSFT generate each Riemann surface $\mathcal{R}_o$ at least once. Uniqueness shows that there is no overcounting: since different diagrams correspond to different metrics (by inspection of their picture), no Riemann surface can be generated twice.

### Closed Strings in OSFT

As is familiar, the open-string $S$-matrix contains poles due to the exchange of on-shell open and closed strings. The closed-string poles are present in nonplanar loop amplitudes. We have seen that OSFT reproduces the standard $S$-matrix. Factorization over the open-string poles is manifest, it corresponds to propagator lengths $T_i$ going to infinity. Surprisingly, the closed-string poles are also correctly reproduced, despite the fact that OSFT treats only the open strings as fundamental dynamical variables. In some sense, closed strings must be considered as derived objects in OSFT. Factorizing the amplitudes over the closed-string poles, one finds that on-shell closed-string states can be represented, at least formally, as certain singular open-string fields with $G = +2$, closely related to the (formal) identity string field. The picture is that of a folded open string, whose left and right halves precisely overlap, with an extra closed-string vertex operator inserted at the mid-point. The corresponding open/closed vertex is given by

$$\langle \Psi_{\text{phys}} | \Phi \rangle_{\text{OC}} \equiv \langle \Psi_{\text{phys}}(0) \mathcal{I} \circ \Phi(0) \rangle_D \qquad \mathcal{I} = \left( \frac{1 + iz}{1 - iz} \right)^2 \tag{35}$$

and describes the coupling to the open-string field of a nondynamical, on-shell closed string $|\Psi_{\text{phys}}\rangle$. It is possible to add this open/closed vertex to the OSFT action. Remarkably, the resulting Feynman rules give a single cover of the moduli space of Riemann surfaces with at least one boundary, with open and closed punctures. This is shown using the same minimal-area problem as above, but now allowing for surfaces with closed punctures as well.

We should finally mention that the structure of OSFT emerges frequently in topological string theory, in contexts where open/closed duality plays a central role. Two examples are the interpretation of Chern–Simons theory as the OSFT for the $A$-model on the conifold, and the intepretation of the Kontsevich matrix integral for topological gravity as the OSFT on FZZT branes in $(2, 1)$ minimal string theory.

## Closed Bosonic SFT

The generalization to covariant closed SFT is nontrivial, essentially because the requisite closed-string decomposition of moduli space is much more complicated.

The free theory parallels the open case, with a minor complication in the treatment of the CFT zero modes. The closed-string field is taken to live in a subspace of the matter + ghost state space, $|\Psi\rangle \in \tilde{\mathcal{H}}_{\text{CFT}_0}$, where the tilde means that we impose the subsidiary conditions

$$b_0^- |\Psi\rangle = L_0^- |\Psi\rangle = 0, \quad b_0^- \equiv b_0 - \bar{b}_0, \quad L_0^- \equiv L_0 - \bar{L}_0 \tag{36}$$

In the classical theory, the string field carries ghost number $G = +2$, since it is the off-shell extension of the familiar closed-string physical states, and the quadratic action reads

$$S \sim \langle \Psi, Q_c \Psi \rangle \tag{37}$$

Here $Q_c$ is the usual closed BRST operator. The inner product $\langle \, , \rangle$ is defined in terms of the BPZ inner product, with an extra insertion of $c_0^- \equiv c_0 - \bar{c}_0$,

$$\langle A, B \rangle \equiv \langle A | c_0^- | B \rangle \tag{38}$$

In [37] $G_{\text{top}} = +6$, as it should be. Without the extra ghost insertion and the subsidiary conditions [36] it would not be possible to write a quadratic action. The linearized equations of motion and gauge invariance,

$$Q_c |\Psi\rangle = 0, \quad |\Psi\rangle \sim |\Psi\rangle + Q_c |\Lambda\rangle, \quad |\Lambda\rangle \in \tilde{\mathcal{H}}^{(1)}_{\text{CFT}_0} \tag{39}$$

give the expected cohomological problem. The fact that the cohomology is computed in the semirelative complex, $b_0^- |\Psi\rangle = b_0^- |\Lambda\rangle = 0$, well known from the operator formalism of the first-quantized theory, is recovered naturally in the second-quantized treatment.

The interacting action is constructed iteratively, by demanding that the resulting Feynman rules give a (unique) cover of moduli space. This requires the introduction of infinitely many elementary string vertices $\mathcal{V}_{g,n}$, where $n$ is the number of closed-string punctures and $g$ the genus. This decomposition of moduli space is more intricate than the decomposition that arises in OSFT, but is in fact analogous to it, when characterized in terms of the following.

**Minimal-area problem for closed SFT** Let $\mathcal{R}_c$ be a closed Riemann surface, possibly with punctures. Find the (conformal) metric of minimal area on $\mathcal{R}$ such that all nontrivial Jordan closed curves have length greater than or equal to $2\pi$.

The minimal-area metric induces a foliation of $\mathcal{R}_c$ by closed geodesics of length $2\pi$. In the classical theory ($g = 0$), the minimal-area metrics arise from Jenkins–Strebel quadratic differentials (as in the open case), and geodesics intersect on a measure-zero set. For $g > 0$, however, there can be foliation bands of geodesics that cross. By staring at the foliation, we can break up the surface into vertices and propagators. In correspondence with each puncture, there is a band of

infinite height, a flat semi-infinite cylinder of circumference $2\pi$, which we identify as an external leg of the diagram. We mark a closed geodesic on each semi-infinite cylinder, at a distance $\pi$ from its boundary. Bands of finite height (internal bands not associated to punctures) correspond to propagators if their height is greater than $2\pi$, otherwise they are considered part of an elementary vertex. Along any internal cylinder of height greater than $2\pi$, we mark two closed geodesics, at a distance $\pi$ from the boundary of the cylinder. If we now cut open all the marked curves, the surface decomposes into a number of semi-infinite cylinders (external legs), finite cylinders (internal propagators) and surfaces with boundaries (elementary interactions). Each elementary interaction of genus $g$ and with $n$ boundaries is an element of $\mathcal{V}_{g,n}$. A crucial point of this construction is that we took care of leaving a "stub" of length $\pi$ attached to each boundary. Stubs ensure that sewing of surfaces preserves the length condition on the metric (no closed curve shorter than $2\pi$).

These geometric data can be translated into an iterative algebraic construction of the full quantum action $S[\Psi]$. The $\mathcal{V}_{g,n}$ satisfy geometric recursion relations whose algebraic counterpart is the quantum BV master equation for $S[\Psi]$. Remarkably, the singularities of the $\Delta$ operator encountered in OSFT are absent here, precisely because of the presence of the stubs. We refer to Zwiebach (1993) for a complete discussion of closed SFT.

#### Open/Closed SFT

There is also a covariant SFT that includes both open and closed strings as fundamental variables. The Feynman rules arise from the following problem.

**Minimal-area problem for open/closed SFT** Let $\mathcal{R}_{oc}$ be a Riemann surface, with or without boundaries, possibly with open and closed punctures. Find the (conformal) metric of minimal area on $\mathcal{R}_{oc}$ such that all nontrivial Jordan open curves have length greater than or equal to $l_o = \pi$, and all nontrivial Jordan closed curves have length greater than or equal to $l_c = 2\pi$.

The surface $\mathcal{R}_{oc}$ is decomposed in terms of elementary vertices $\mathcal{V}^{g,n}_{b,m}$ (of genus $g$, $b$ boundary components, $n$ closed-string punctures and $m$ open-string punctures) joined by open and closed propagators. Degenerations of the surface correspond always to propagators becoming of infinite length – factorization is manifest both in the open and in the closed channel.

The SFT described in the section "Closed strings in OSFT" (Witten OSFT augmented with the single open/closed vertex [35]) corresponds to taking $l_o = \pi$ and $l_c = 0$. Varying $l_c \in [0, 2\pi]$, we find a whole family of interpolating SFTs. This construction clarifies the special status of the Witten theory: moduli space is covered by a single cubic open overlap vertex, with no need to introduce dynamical closed strings, but at the price of a somewhat singular formulation.

### Classical Solutions in Open SFT

In the present formulation of SFT, a background (a classical solution of string theory) must be chosen from the outset. The very definition of the string field requires to specify a (B)CFT$_0$. Intuitively, the string field lives in the "tangent" to the "theory space" at a specific point – where "theory space" is some notion of a "space of 2D (boundary) quantum field theories," not necessarily conformal. In the early 1990s independence from the choice of background was demonstrated for infinitesimal deformations: the SFT actions written using neighboring (B)CFTs are indeed related by a field redefinition. In recent years, it has become apparent that at least the open-string field reaches out to open-string backgrounds a finite distance away – possibly covering the whole of theory space. (Classical solutions of closed SFT are beginning to be investigated at the time of this writing (2005)).

The OSFT action written using BCFT$_0$ data is just the full world volume action of the D-brane with BCFT$_0$ boundary conditions. Which classical solutions should we expect in this OSFT? In the bosonic string, D$p$ branes carry no conserved charge and are unstable. This instability is reflected in the presence of a mode with $m^2 = -1/\alpha'$, the open-string tachyon $T(x^\mu), \mu = 0, \ldots, p$. From this physical picture, Sen argued that:

1. the tachyon potential, obtained by eliminating the higher modes of the string field by their equations of motion, must admit a local minimum corresponding to the vacuum with no D-brane at all (henceforth, the tachyon vacuum, $T(x^\mu) = T_0$);
2. the value of the potential at $T_0$ (measured with respect to the BCFT$_0$ point $T = 0$) must be exactly equal to minus the tension of the brane with BCFT$_0$ boundary conditions;
3. there must be no perturbative open-string excitations around the tachyon vacuum; and
4. there must be space-dependent "lump" solutions corresponding to lower-dimensional branes. For example, a lump localized along one world volume direction, say $x^1$, such that $T(x^1) \to T_0$ as $x^1 \to \pm\infty$, is identified with a D$(p-1)$ brane.

Sen's conjectures have all been verified in OSFT. (See Sen (2004) and Taylor and Zwiebach (2003) for reviews). The deceptively simple-looking equations of motion (in Siegel gauge)

$$L_0|\Phi\rangle + b_0(|\Phi\rangle * |\Phi\rangle) = 0 \qquad [40]$$

are really an infinite system of coupled equations, and no analytic solutions are known. Turning on a vacuum expectation value (VEV) for the tachyon drives into condensation an infinite tower of modes. Fortunately, the approximation technique of "level truncation" is surprisingly effective. The string field is restricted to modes with an $L_0$ eigenvalue smaller than a prescribed maximal level $L$. For any finite $L$, the truncated OSFT contains a finite number of fields and numerical computations are possible. Numerical results for various classical solutions converge quite rapidly as the level $L$ is increased.

The most important solution is the string field $|\mathcal{T}\rangle$ that corresponds to the tachyon vacuum. A remarkable feature of $|\mathcal{T}\rangle$ is universality: it can be written as a linear combination of modes obtained by acting on the tachyon $c_1|0\rangle$ with ghost oscillators and matter Virasoro operators,

$$|\mathcal{T}\rangle = T_0\, c_1|0\rangle + u\, L^m_{-2} c_1|0\rangle + v\, c_{-1}|0\rangle + \cdots$$

This implies that the properties of $|\mathcal{T}\rangle$ are independent of any detail of $\mathrm{BCFT}_0$, since all computations involving $|\mathcal{T}\rangle$ can be reduced to purely combinatorial manipulations involving the ghosts and the Virasoro algebra. The numerical results strongly confirm Sen's conjectures, and indicate that the tachyon vacuum is located at a non-singular point in configuration space. Numerical solutions describing lower-dimensional branes and exactly marginal deformations are also available. For example, the full family of solutions interpolating between a D1 and a D0 brane at the self-dual radius has been found. There is increasing evidence that the open-string field provides a faithful map of the open-string landscape.

### Vacuum SFT: D-branes as Projectors

In the absence of a closed-form expression for $|\mathcal{T}\rangle$, we are led to guesswork. When expanded around $|\mathcal{T}\rangle$, the OSFT is still cubic, only with a different kinetic term $\mathcal{Q}$,

$$S = -\kappa_0 \left[\frac{1}{2}\langle\Phi|\mathcal{Q}|\Phi\rangle + \frac{1}{3}\langle\Phi|\Phi * \Phi\rangle\right] \qquad [41]$$

The operator $\mathcal{Q}$ must obey all the formal properties [17], must be universal (constructed from ghosts and matter Virasoro operators), and must have trivial cohomology at $G = +1$. Another constraint comes from requiring that [41] admits classical solutions in Siegel gauge. The choice

$$\begin{aligned}\mathcal{Q} &= \frac{1}{2\mathrm{i}}(c(i) - \bar{c}(i)) \\ &= c_0 - (c_2 + c_{-2}) + (c_4 + c_{-4}) - \cdots\end{aligned} \qquad [42]$$

satisfies all these requirements. The conjecture (Rastelli *et al.* 2001) is that, by a field redefinition, the kinetic term around the tachyon vacuum can be cast into this form. This "purely ghost" $\mathcal{Q}$ is somewhat singular (it acts at the delicate string mid-point), and presumably should be regarded as the leading term of a more complicated operator that includes matter pieces as well. The normalization constant $\kappa_0$ is formally infinite. Nevertheless, a regulator (e.g., level truncation) can be introduced, and physical observables are finite and independent of the regulator. The vacuum SFT ([41]–[42]) appears to capture the correct physics, at least at the classical level. Taking a matter/ghost factorized ansatz

$$|\Phi_{\mathrm{g}}\rangle \otimes |\Phi_{\mathrm{m}}\rangle \qquad [43]$$

and assuming that the ghost part is universal for all D-branes solutions, the equations of motion reduce to following equations for the matter part:

$$|\Phi_{\mathrm{m}}\rangle * |\Phi_{\mathrm{m}}\rangle = |\Phi_{\mathrm{m}}\rangle \qquad [44]$$

A solution $|\Phi_{\mathrm{m}}\rangle$ can be regarded as a projector acting in "half-string space." Recall that the $*$-product looks formally like a matrix multiplication [19]: the matrices are the string fields, whose "indices" run over the half-string curves. These projector equations have been exactly solved by many different techniques (see Rastelli (2004) for a review). In particular, there is a general BCFT construction that shows that one can obtain solutions corresponding to any D-brane configuration, including multiple branes – the rank of the projector is the number of branes. A rank-one projector corresponds to an open-string functional which is left/right split, $\Phi[X(\sigma)] = F_{\mathrm{L}}(X_{\mathrm{L}})F_{\mathrm{R}}(X_{\mathrm{R}})$. There is also clear analogy between these solutions and the soliton solutions of noncommutative field theory. The analogy can be made sharper using a formalism that rewrites the open-string $*$-product as the tensor product of infinitely many Moyal products. (See Bars (2002) and references therein).

It is unclear whether or not multiple-brane solutions (should) exist in the original OSFT – they are yet to be found in level truncation. Understanding this and other issues, like the precise role of closed strings in the quantum theory seems to require a precise characterization of the allowed

space of open-string functionals. In principle, the path integral over such functionals would define the theory at the full nonperturbative level. This remains a challenge for the future.

**Note Added in Proof** Very recently, M Schnabl, building on previous work on star algebra projectors and related surface states (Rastelli L (2004) and references therein) was able to find the exact solution for the universal tachyon condensate in OSFT. This breakthrough is likely to lead to rapid new developments in SFT.

*See also:* Boundary Conformal Field Theory; BRST Quantization; Chern–Simons Models: Rigorous Results; Fedosov Quantization; The Jones Polynomial; Large-*N* and Topological Strings; Large-*N* Dualities; Noncommutative Geometry from Strings; Noncommutative Tori, Yang–Mills, and String Theory; Operads; Superstring Theories; Topological Quantum Field Theory: Overview; Two-Dimensional Conformal Field Theory and Vertex Operator Algebras.

## Further Reading


Bars I (2002) *MSFT: Moyal Star Formulation of String Field Theory*. Proceedings of 3rd International Sakharov Conference on Physics, Moscow, Russia, 24–29 Jun, arXiv:hep-th/0211238.

Berkovits N (2001) Review of open superstring field theory, arXiv:hep-th/0105230.

Berkovits N, Okawa Y, and Zwiebach B (2004) WZW-like action for heterotic string field theory. *Journal of High Energy Physics* 0411: 038 (arXiv:hep-th/0409018).

Bordes J, Chan HM, Nellen L, and Tsou ST (1991) Half string oscillator approach to string field theory. *Nuclear Physics B* 351: 441.

Erler TG and Gross DJ (2004) Locality, causality, and an initial value formulation for open string field theory, arXiv:hep-th/0406199.

Ohmori K (2001) *A Review on Tachyon Condensation in Open String Field Theories*. Master's thesis, University of Tokyo (arXiv:hep-th/0102085).

Okawa Y (2002) Open string states and D-brane tension from vacuum string field theory. *Journal of High Energy Physics* 0207: 003 (arXiv:hep-th/0204012).

Rastelli L (2004) Open string fields and D-branes. *Fortschritte der Physics*. 52: 302.

Rastelli L, Sen A, and Zwiebach B *Vacuum String Field Theory*, Proceedings of the Strings 2001 Conference, TIFR, Mumbai, India, arXiv:hep-th/0106010.

Schnabl M (2005) Analytic solution for tachyon condensation in open string theory. arXiv:hep-th/0511286.

Sen A (2004) Tachyon dynamics in open string theory, arXiv:hep-th/0410103.

Shatashvili SL *On Field Theory of Open Strings, Tachyon Condensation and Closed Strings*, Proceedings of the Strings 2001 Conference, TIFR, Mumbai, India, arXiv:hep-th/0105076.

Siegel W (1988) Introduction To String Field Theory. *Advanced Series in Mathematical Physics* 8: 1–244 (arXiv:hep-th/0107094).

Taylor W and Zwiebach B (2001) D-branes, tachyons, and string field theory. In: Gubser SS and Lykken JD (eds.) *Boulder 2001, Strings, Branes and Extra Dimensions, TASI 2001 Lectures*, pp. 641–759. Singapore: World Scientific.

Thorn CB (1989) String field theory. *Physics Report* 175: 1.

Witten E (1986) Noncommutative geometry and string field theory. *Nuclear Physics B* 268: 253.

Zwiebach B (1993) Closed string field theory: quantum action and the B-V master equation. *Nuclear Physics B* 390: 33 (arXiv:hep-th/9206084).

Zwiebach B (1998) Oriented open–closed string theory revisited. *Annals of Physics* 267: 193 (arXiv:hep-th/9705241).


# String Theory: Phenomenology

**A M Uranga**, Consejo Superior de Investigaciones Cientificas, Madrid, Spain



## String Theory and Compactification

The string theory provides a setup in which gauge and gravitational interactions can be described in a unified framework consistently at the quantum level. As such, it provides a candidate theory in which to describe the standard model of particle physics (describing quarks and leptons and their strong and electroweak interactions) and gravity within the same quantum theory.

The string theory has a unique fundamental scale $M_s$, fixed by the string tension, often encoded in the parameter $\alpha'$ of dimension $(\text{length})^{-2}$. All other scales are derived from this one and are background dependent.

Most of the string theory phenomenological model building has centered on the critical superstrings, which are ten dimensional (10D) and involve spacetime (as well as world-sheet) supersymmetry. There are five such different 10D theories: type IIA, type IIB, type I, and the $E_8 \times E_8$ and SO(32) heterotic theories. The heterotic theories include nonabelian gauge fields and charged fermions in ten dimensions; hence, they constitute a promising setup to embed the standard model. On the other hand, the possibility of including D-branes

(which carry nonabelian gauge symmetries and charged matter) in compactifications of type II theories (and orientifolds thereof, like the type I theory itself) makes the latter reasonable alternative setups to embed the standard model as a brane world. The different 10D theories (as well as the 11D M-theory) are related by diverse dualities, also upon compactification. This suggests that they are just different limits of a unique underlying theory. For 4D models, this implies that the different classes of constructions are ultimately related by dualities, and that often a given model may be realized using different string theory constructions as starting points.

In order to recover 4D physics at low energies, compactification of the theory is required. In geometrical terms, the theory is required to propagate on a spacetime with geometry $M_4 \times X_6$, where $M_4$ is a 4D Minkowski space, and $X_6$ is a compact manifold. This description is valid in the regime of a large compactification volume, $\alpha'/R^2 \ll 1$ (where $R$ is the overall scale of the compact manifold), where $\alpha'$ string theory corrections are negligible. Other 4D string models may be constructed using abstract conformal field theories. They may often be regarded as extrapolations of geometric compactifications to the regime of sizes comparable with the string length, where string theory corrections are relevant and the classical geometric picture does not hold.

In the simplest situation of geometrical compactification, not including additional backgrounds beyond the metric, the requirement of 4D spacetime supersymmetry (useful for the stability of the model, as well as of phenomenological interest) implies that the space $X_6$ is endowed with an SU(3) holonomy metric. Existence of such metrics is guaranteed for Calabi–Yau spaces, namely Kähler manifolds with vanishing first Chern class.

There are a very large number of 4D supersymmetric string models that can be constructed using different starting string theories and different compactification manifolds. They lead to different 4D spectra, often including nonabelian gauge symmetries and charged chiral fermions (but only rarely resembling the actual standard model). In addition, for each given model, there exist, in general, a large number of massless 4D scalars, known as moduli, whose vacuum expectation values are not fixed. They parametrize different choices of the compactification data in a given topological sector (e.g., Kähler and complex structure moduli of the internal Calabi–Yau space). All physical parameters of the 4D theory vary continuously with the vacuum expectation values of these scalars.

All such models are on equal footing from the point of view of the theory. Hence, 4D string models suffer from a large arbitrariness. Although the breaking of supersymmetry clearly changes the picture qualitatively (e.g., flat directions associated to moduli are lifted by radiative corrections), it is difficult to evaluate this impact.

In this situation, most of the research in string theory phenomenology has centered on the study of generic properties of certain classes of compactifications, with the potential to lead to realistic structures (such as $N=1$ or no supersymmetry, nonabelian gauge symmetries with replicated sets of charged chiral fermions). Within each class, explicit models (as close as possible to the standard model) have also been constructed. Generic predictions or expectations for phenomenology can be obtained within each setup, but quantitative results, even for explicit models, are always functions of undetermined moduli vacuum expectation values. Tractable mechanisms for moduli stabilization are under active research, although only preliminary results are available presently.

The better-studied classes of models are compactifications of heterotic theories on Calabi–Yau spaces, and compactifications of type II theories (or orientifolds thereof) with D-branes. Other possibilities include the heterotic M-theory, the M-theory on $G_2$ holonomy varieties, the F-theory on Calabi–Yau 4-folds, etc. As already mentioned, different classes (or even explicit models) are often related by string duality.

## Heterotic String Phenomenology

A large class of phenomenologically interesting string vacua, which has been explored in depth, is provided by 4D compactifications of (any of the two) perturbative heterotic string theories. Compactification on large volume manifolds can be described in the supergravity approximation. As described by Candelas, Horowitz, Strominger, and Witten, the requirement of 4D $N=1$ supersymmetry requires the internal manifold to be of SU(3) holonomy, a condition which is satisfied by Calabi–Yau manifolds. In the presence of a curvature, the Bianchi identity for the Kalb–Ramond 2-form $B$ is modified, so that, in general, it reads

$$\mathrm{d}H = \mathrm{tr}\, R^2 - \frac{1}{30}\mathrm{tr}\, F^2 \qquad [1]$$

where $H$ is the field strength 3-form, $R$ is the Ricci 2-form, and $F$ is the field strength, in the adjoint representation, of the 10D gauge fields. Regarding the above equation in cohomology leads to a

consistency condition, forcing the background gauge bundle $V$ to be topologically nontrivial, with

$$c_2(V) = c_2(TX_6) \qquad [2]$$

where $c_2$ denotes the second Chern class, and $TX_6$ is the compactification tangent space.

The condition of supersymmetry implies that the gauge fields must be solutions of the Donaldson–Uhlenbeck–Yau equations. Existence of such a solution is guaranteed for holomorphic and stable gauge bundles. The simplest solution to these conditions is the so-called standard embedding, where the gauge connection is locally identical to the spin connection, but more general solutions exist and have been characterized for particular classes of Calabi–Yau manifolds (e.g., when they are elliptically fibered). The gauge background bundle $V$, with structure group $H$, breaks the 10D gauge symmetry $G$ to its commutant subgroup $G_{4\mathrm{D}}$. The latter corresponds to the 4D gauge symmetry. Moreover, the background bundle modifies the Kaluza–Klein reduction of the 10D charged fermions, leading to a nonzero number of replicated 4D chiral fermions. Decomposing the adjoint representation of $G$ (in which 10D fermions transform) with respect to $G_{4\mathrm{D}} \times H$,

$$\mathrm{Adj}\, G = \bigoplus_i (R_{G_{4\mathrm{D}},i}, R_{H,i}) \qquad [3]$$

the net number of 4D chiral fermions in the representation $R_{G_{4\mathrm{D}}}$ is given by the index of the Dirac operator coupled to $V$ in the representation $R_{H,i}$. Condition [1] implies proper cancellation of chiral anomalies in the resulting theory. A simple and well-studied class is provided by standard embedding compactifications of the $E_8 \times E_8$ heterotic string theory, whose unbroken 4D gauge group is $E_6 \times E_8$. The number of families (i.e., chiral multiplets in the representation 27 of $E_6$ ) and conjugate families (in the $\overline{27}$) are given by the Hodge numbers

$$n_{27} = h_{1,1}(X_6), \quad n_{\overline{27}} = h_{2,1}(X_6) \qquad [4]$$

More specifically, the harmonic representatives in each cohomology class represent the internal profile of the corresponding 4D fields. The net number of families is thus determined by the Euler characteristic $\chi(X_6)$

$$n_{\mathrm{fam}} = |h_{1,1} - h_{2,1}| = \tfrac{1}{2}|\chi(X_6)| \qquad [5]$$

Recently, much progress in heterotic model building has been achieved in nonstandard embedding compactifications by the detailed construction of holomorphic stable bundles and the computation of the diverse indexes. In particular, explicit models with just the minimal supersymmetric standard model spectrum have been constructed.

The above geometric approach has several limitations. On the technical side, the construction of explicit holomorphic and stable gauge bundles is nontrivial from the mathematical viewpoint. On the more fundamental side, it allows one to explore only the large volume limit of heterotic compactifications.

Further insight into the latter aspect can be obtained via constructions based on exactly solvable conformal field theories (CFTs), which describe the world-sheet string dynamics in compactifications, including all $\alpha'$ corrections, and, therefore, allowing one to enter the small volume regime. The simplest such compactifications are provided by toroidal orbifolds, which describe string propagation in quotients of toroidal compactifications by a discrete group $\Gamma$. From the world-sheet viewpoint, they are described by 2D free CFT, but which include sectors of closed strings with boundary conditions twisted by elements of $\Gamma$. The resulting 4D theory contains chiral fermions, arising from the untwisted and twisted sectors. In the former, the nonchiral spectrum of toroidal compactification suffers a projection onto the $\Gamma$-invariant states and leads to chirality. Twisted sectors are localized at the fixed points of the orbifold action, where the local supersymmetry is reduced, leading naturally to chiral fermions.

Many of these models can be regarded as limits of compactifications on Calabi–Yau spaces in the limit in which they become locally flat and develop conical singularities (and similarly, their gauge bundles become locally flat and with curvature localized near the singular points). Indeed, flat directions involving moduli fields in the twisted sector often exist, which correspond to geometric blow-ups of the singular point that resolve the conical singularities to yield a smooth Calabi–Yau.

The theories remain simple and solvable for any value of the untwisted moduli (namely moduli of the underlying toroidal compactification). This allows the discussion of their low-energy effective action including the explicit dependence on the untwisted moduli, while only partial results for the dependence on twisted moduli are known.

Other approaches, such as free fermion constructions or Gepner models, also provide exact descriptions of compactifications, although only at a point of the moduli space, deep inside the small volume regime.

Exact CFT constructions provide a small volume description of Calabi–Yau compactifications, at least for particular models. Moreover, their consistency conditions (modular invariance of the partition function) provide a stringy version of the large volume geometric condition implied by eqn [2]. The

constructions also show the existence of full-fledged string theory constructions with properties similar to geometric compactifications, but incorporating all $\alpha'$ corrections.

Within the general class of perturbative heterotic string models, a certain number of phenomenologically interesting statements are quite generic.

- The 4D Planck scale $M_{\rm P}$ and gauge couplings $g_{YM}$ (at the string scale) are related to the fundamental string scale by

$$M_{\rm s} = M_{\rm P} g_{YM} \qquad [6]$$

  This implies that the string scale is close to the 4D Planck scale. In this situation, supersymmetry can stabilize the electroweak scale against radiative corrections.
- 4D heterotic models contain certain U(1) symmetries, whose gauge bosons actually get Stuckelberg masses due to $B \wedge F$ couplings to components of the 2-form. Such U(1)'s would correspond to global symmetries, but are violated at tree level by $\alpha'$ nonperturbative effects, namely world-sheet instantons. Hence, no continuous global symmetries exist, even perturbatively, in these models. Proton decay might, however, be avoided by discrete global symmetries. In any event, even without such symmetries, the large fundamental scale suppresses the processes mediating proton decay. Thus, the proton lifetime is naturally larger than present experimental bounds.
- Gauge coupling constants for the different gauge factors in the standard model unify at the string scale. This agrees with extrapolation from their electroweak values, assuming the minimal supersymmetric standard model content between the electroweak and string scale, up to a mismatch of scales (by a factor of 20). The latter may be addressed in diverse ways, such as threshold corrections, intermediate scales, or in the heterotic M-theory.
- Yukawa couplings are, in principle, computable. Explicit computations have been carried out in standard embedding geometric compactifications (where they amount to the overlap integral of the internal profiles of the 4D fields, namely a topological intersection number), and in orbifold models. They are in general moduli dependent, so their quantitative analysis is involved. Qualitatively, however, interesting patterns, such as hierarchical structures, are possible, for example, in specific orbifold models.

Heterotic models have been studied beyond the perturbative regime. For instance, the construction of compactifications including nonperturbative objects, namely 5-branes, has been pursued; so has been the strong coupling limit of the $E_8 \times E_8$ heterotic, described by compactifications of the M-theory on an interval (the so-called heterotic M-theory or Horava–Witten theory). The strong coupling phenomena of the SO(32) heterotic theory can be addressed using dual type I (or other type II orientifold) constructions.

## D-Brane Phenomenology

A different setup for realistic string theory compactifications, within the so-called brane-world constructions, is provided by compactifications of type II string theories containing D-branes, or quotients thereof. A particularly relevant class of quotients involves quotienting out by world-sheet parity, accompanied by some $Z_2$ geometric action. The resulting theories are denoted type II orientifolds, and contain orientifold planes, subspaces fixed under the geometric action, corresponding to regions where the orientation of a string can flip. Type II compactifications with D-branes filling the noncompact dimensions must satisfy a set of consistency conditions, known as RR tadpole cancellation. This is the condition that, in the compact space, the charge of D-branes and orientifold planes under the different RR forms must cancel. For the $Z$-valued charges, the conditions read

$$\sum_a N_a Q_a + Q_{\rm Op} = 0 \qquad [7]$$

where $N_a$ denotes the multiplicity of D-branes with charge vector and $Q_a$ under the RR fields, $Q_{\rm Op}$ is the charge vector of the orientifold planes. Additional discrete conditions may be present if the relevant K-theory group (classifying D-brane charges in the corresponding background) contains torsion pieces.

The most familiar example of these constructions is provided by the type I string theory, which is an orientifold quotient of the type IIB theory by world-sheet parity (with no geometric action). The model can be regarded as containing one orientifold 9-plane and 32 D9 branes (all filling out 10D spacetime), such that their RR charges with respect to the (nondynamical) RR 10-form cancel.

Supersymmetric geometric compactifications of type II theories and orientifolds must correspond to compactification on Calabi–Yau spaces in order to have a preserved spinor. Models with D-branes filling the noncompact dimensions may be broadly classified into two classes: type IIB compactifications with $D(3 + 2p)$-branes, wrapped on holomorphic $2p$-cycles, and carrying holomorphic and stable

world-volume gauge bundles, and type IIA compactifications with D6 branes wrapped on special Lagrangian 3-cycles (in general, models with D4 and D8 branes are not allowed since Calabi–Yau spaces do not have nontrivial 1- or 5-cycles on which to wrap the branes). This classification is a large volume realization of the general classification of supersymmetric configurations of D-branes into two classes, denoted A and B.

### Intersecting Brane Worlds

Type IIA compactifications with A-branes correspond to compactifications of type IIA theory (or orientifolds thereof) with D6 branes wrapped on 3-cycles of the internal Calabi–Yau space. In these models, each stack of $N$ D6 branes generically leads to a U($N$) gauge factor. Chirality arises from open strings stretched between pairs of branes at the corresponding intersections. The chiral fermions from an open string stretched between branes $a$ and $b$ transform in the bifundamental representation $(\square_a, \overline{\square}_b)$ of the gauge factors $U(N_a) \times U(N_b)$ of the intersecting D6 brane stacks. In general, two 3-cycles in a 6D manifold intersect at points of the internal space. Hence, such fermions arise in several families, whose (net) number is given by the (net) number of intersections of the corresponding 3-cycles $\Pi_a, \Pi_b$, namely the topological invariant intersection number of their homology classes

$$I_{ab} = [\Pi_a] \cdot [\Pi_b] \qquad [8]$$

Simple modifications of the above rules arise in some sectors in the presence of orientifold planes (e.g., the reduction of the gauge symmetry from unitary to orthogonal or symplectic factors for branes on top of orientifold planes).

The RR tadpole cancellation conditions specify that the total homological charge carried by the D6 branes (and the orientifold 6-planes) cancel. They imply automatic cancellation of cubic nonabelian anomalies, and the cancellation of mixed U(1) anomalies by a Green–Schwarz mechanism mediated by 4D scalars from the RR closed-string sector.

Explicit models with SM spectrum have been constructed in orientifolds of toroidal compactifications in the nonsupersymmetric case, and in orbifolds thereof in supersymmetric cases. The generalization of the above construction beyond toroidal situations is, in principle, possible, but difficult, due to the mathematically challenging task of constructing special Lagrangian submanifolds for general Calabi–Yau manifolds.

Certain phenomenologically interesting quantities, such as gauge couplings and their threshold corrections, Yukawa couplings, and other diverse correlation functions have been computed in toroidal cases, where the corresponding correlators are computable exactly in $\alpha'$. Particularly interesting is the computation of Yukawa couplings, or, in general, of couplings involving only fields at intersections. These couplings arise from open-string world-sheet instantons, namely disks with boundaries on the D-branes corresponding to those intersections.

### Type IIB Orientifolds

Type IIB compactifications with B-type branes contain several familiar classes of 4D models, for instance, compactifications of type I string theory on smooth Calabi–Yau spaces (whose description may be carried out using the effective supergravity action, in close analogy with the heterotic compactifications). Compactifications of type I string theory on orbifolds can be regarded as a particular realization of this, easily described using exact CFTs (although from the viewpoint of the general description as B-branes, the appearance of lower-dimensional branes requires their mathematical description to involve coherent sheaves). Since open strings at orbifolds do not have twisted boundary conditions, chirality arises from the orbifold projection of the toroidally compactified theory on the spectrum.

Another example within this kind is provided by the so-called magnetized D-brane models. These correspond to toroidal compactifications of type I theory, with D9 branes carrying constant magnetic backgrounds for the internal components of the world-volume gauge fields. In this kind of model, although the closed-string sector is highly supersymmetric, the open-string spectrum has reduced supersymmetry, or no supersymmetry (if the bundle stability condition is relaxed). Chirality arises from the nontrivial index of the Dirac operator for open strings ending on D-branes with different world-volume magnetic fields. Explicit models have mainly centered on nonsupersymmetric models from orientifolds of $T^6$, and on supersymmetric models from orientifolds of the $T^6/(Z_2 \times Z_2)$ orbifold. In both contexts, models with semirealistic spectra have been obtained: concretely nonsupersymmetric models with just the standard model spectrum, or supersymmetric models with the minimal supersymmetric standard model spectrum, plus nonchiral matter. Further, properties of the gauge coupling constants and the computation of the Yukawa couplings have been studied as functions of undetermined moduli.

Finally, a second large class of models constructed using B-type branes are given by lower-dimensional D-branes, for example, D3 branes, located at singular points in the internal compactification space. Since the massless sector of open strings is determined only in terms of the local structure of the singularity, these models have been mostly studied in noncompact setups. Resulting spectra can be encoded in quiver diagrams, related to those in the mathematical literature on the McKay correspondence. Semirealistic three-family models have been constructed based on systems of D3 and D7 branes at the $C^3/Z_3$ orbifold singularity.

Type IIB orientifold compactifications are also intimately related to F-theory compactifications on Calabi–Yau 4-folds, which provide a nonperturbative completion for such models.

Mirror symmetry exchanges type IIB and IIA compactifications with B- and A-type branes. Hence, it provides a map between the above two kinds of compactifications. This shows that type IIB orientifold models lead to spectra with structure similar to that of intersecting-branes worlds, and that they share many of their general properties.

As a particular example, toroidal models of intersecting D6 branes are mapped under mirror symmetry to models of magnetized D9 branes. This mirror map has been exploited to construct the same theories from both starting points and to recover certain quantities, such as the $\alpha'$-exact Yukawa couplings in the IIA picture from a purely classical (no $\alpha'$ corrections) computation in the mirror IIB model. This is a particular application of the general proposal of homological mirror symmetry in compactifications with branes.

Type II orientifold compactifications with D-branes have also been explored beyond the geometric regime, using exact CFTs to describe the (analog of the) internal space, and crosscap and boundary states to describe (the analogs of) orientifold planes and D-branes. Formal developments in the construction of the latter in Gepner models have been successfully applied to obtain large classes of semirealistic 4D string models in this setup.

As compared with heterotic compactifications, the setup of D-brane models leads to several generic features:

- Since gauge sectors are localized on D-branes, and have a dilaton dependence different from gravitational interactions, the relation between the fundamental string scale and the 4D Planck scale and gauge coupling reads

$$M_{\mathrm{P}}^2 g_{YM}^2 = \frac{M_{\mathrm{s}}^{11-p} V_{\mathrm{T}}}{g_{\mathrm{s}}} \qquad [9]$$

  where $V_{\mathrm{T}}$ is a measure of the volume in the directions transverse to the brane, and $g_{\mathrm{s}}$ is the 10D string coupling. The above relation shows that it is possible to achieve large 4D Planck mass with a lower fundamental string scale by adjusting the transverse volume and the string coupling. This has been proposed by Antoniadis, Arkani-Hamed, Dimopoulos, and Dvali as an alternative to explain the Planck/weak hierarchy without supersymmetry.
- The compactifications contain several U(1) gauge symmetries. For some of the corresponding gauge bosons, the 4D effective theory contains Stuckelberg masses of order $M_{\mathrm{s}}$, due to $B \wedge F$ couplings to fields in the RR sector. These couplings make the U(1) gauge bosons massive; hence, they are absent from the low-energy physics. Nevertheless, the U(1)'s remain as global symmetries exact in $\alpha'$ and to all orders in the perturbation theory in $g_{\mathrm{s}}$. They are violated by D-brane instantons, which are nonperturbative in $g_{\mathrm{s}}$. In many realistic models, the baryon number is one such global symmetry, and it prevents proton decay, even if the string scale is not large.
- In general, each gauge factor in the standard model arises from a different brane stack, and their gauge couplings at the string scale are controlled by different moduli. This implies that, generically, it is not natural to have gauge coupling unification in D-brane models. Particular models may enjoy enhanced discrete global symmetries at special points in moduli space where unification is achieved, thus making unification appear more natural in such examples. Similar statements apply for constructions which realize complete or partial unification of gauge groups at large scales (like string models of grand unification or of Pati–Salam type).
- As already mentioned, important quantities such as Yukawa couplings are, in principle, computable, although quantitative expressions have been derived only in a few examples, mostly in toroidal compactifications or quotients thereof. The results are moduli dependent, making it difficult to derive model-independent patterns.

## M-Theory Phenomenology

Most of the phenomenological models from the M-theory have been constructed using the Horava–Witten theory (compactification of M-theory on $S^1/Z_2$) as starting point. This theory provides a description of the strong coupling regime of the $E_8 \times E_8$ heterotic theory, and many of its basic features are similar to those in the perturbative regime. In particular, the techniques used in model

building involve the construction of stable and holomorphic vector bundles and the computation of the relevant indexes to obtain the 4D gauge group and charge matter content. An important difference is that gauge interactions propagate only over the 10D boundaries of spacetime, while gravity propagates over the 11 dimensions. This makes the setup share some features of brane-world constructions, and, in particular, it allows one to lower the fundamental scale of the theory (the 11D Planck scales) to reconcile it with the traditional unification scale.

A different setup for M-theory phenomenology involves the compactification of the 11D theory on a 7-manifold of $G_2$ holonomy $X_7$, in order to lead to $N=1$ supersymmetry in four dimensions. Although a fundamental formulation of the M-theory is lacking, duality arguments and indirect evidence can be used to show that nonabelian gauge symmetries of the A–D–E classical groups arise if $X_7$ contains 3-cycles of codimension-4 singularities, locally of the form $C^2/\Gamma$, with $\Gamma$ an A–D–E Kleinian subgroup of SU(2). Similarly, it can be shown that chiral multiplets charged under these gauge symmetries arise if $X_7$ contains certain codimension-7 singularities. The local geometry of the latter has been explicitly described, and can be regarded as lying at the intersections of codimension-4 singularities.

The direct construction of such singular $G_2$ holonomy manifolds is very difficult, and there are no known topological conditions that guarantee existence of such a metric for a fixed topology. However, the existence of large classes of such models can be indirectly shown by using duality arguments. Namely, any type IIA models of intersecting D6 branes and O6 planes, preserving $N=1$ supersymmetry, lifts to an M-theory compactification on a singular $G_2$ holonomy manifold. In fact, the local structure of the codimension-4 and -7 singularities agrees in particular cases with the local structure of D6 branes on 3-cycles and D6 brane intersections.

## Further Topics

Some additional topics related to the phenomenology of the string theory, but not covered by the above model building description are discussed in the following.

### Effective Actions

The construction of effective actions for such classes of models has been carried out in general in supersymmetric compactifications, using the parametrization of the general 4D $N=1$ supergravity action in terms of the Kähler potential for the moduli and matter fields, the gauge kinetic functions, and the superpotential. The moduli action is quite universal, at least for geometric compactifications and for untwisted moduli in orbifold compactifications. For instance, the Kähler potential for the 4D dilaton multiplet $S$ and the modulus $T$ controlling the size of the internal manifold, in the large volume and weak coupling regime, reads

$$K = -\log(S+S^*) - 3\log(T+T^*) \qquad [10]$$

The corresponding expression including matter fields is more model dependent, but known within each particular class.

### Moduli Stabilization and Supersymmetry Breaking

Both issues are often related. Although moduli stabilization preserving supersymmetry is possible, it often occurs that the potential stabilizing moduli has its origin in mechanisms related to supersymmetry breaking.

The description of purely string theoretical mechanisms to break supersymmetry is difficult, and most approaches rely on field-theoretical mechanisms in the effective action. One of the better-studied mechanisms, mostly in the heterotic string setup (but also in type II compactifications), is gaugino condensation in a strongly coupled hidden sector, interacting with the standard model sector via gravitational (or perhaps additional gauge) interactions. Although explicit models with such hidden sectors and strong dynamics exist, they often result in runaway potentials for moduli. Racetrack scenarios where several condensates balance each other are possible but contrived.

A second mechanism to break supersymmetry, mostly explored in type IIB/F-theory compactifications, is the introduction of field-strength fluxes for $p$-form fields. Interestingly, such fluxes lead to nontrivial potentials depending on moduli, and generically breaking supersymmetry. The existence of several remnant flat directions in the leading $\alpha', g_s$ approximation, leaves unanswered the question of possible runaway moduli potentials in those directions. However, evidence for nonperturbative contributions stabilizing the remaining moduli at finite distance has been proposed. Preliminary results in the analysis of flux stabilized vacua have been obtained in simple examples of (still unrealistic) Calabi–Yau compactifications with small number of moduli.

Most explored mechanisms propose supersymmetry breaking below the Kaluza–Klein compactification

scale, and, therefore, can be described in the 4D effective theory. They can be nicely parametrized in terms of vacuum expectation values for the dilaton and geometric moduli of the compactification. This description allows for a computation of the soft terms using the expansion of the $N = 1$ supergravity formulas in components. Concrete patterns, such as the universality of squark masses, or the complex phases of diverse soft terms, can be explored using this approach.

Alternative mechanisms of breaking supersymmetry at higher scales, such as the introduction of antibranes or nonsupersymmetric compactifications, lead to generic difficulties with stability.

Related to the question of supersymmetry breaking is the question of the cosmological constant. Unfortunately, there is no manifest mechanism in the string theory that explains the smallness of the observed value of this scale. Given that many aspects of both quantum gravity in the string theory and realistic model building (with proper supersymmetry breaking and moduli stabilization) are still under progress, an open-minded point of view on this problem and the proposed solutions is kept.

### Cosmology

Although somewhat different from the traditional focus of string phenomenology, recent progress in observational cosmology has triggered much interest in string theory realizations of inflationary models (or alternatives such as pre-big bang scenarios). Most inflationary models have centered on using moduli as the inflaton field, due to their flat potentials. A simple setup in type II compactifications, known as brane inflation models, uses the modulus controlling a brane position as the inflaton field, which has a flat enough potential with a moderate fine-tuning. Such setups may lead to interesting additional features, such as a moderate but potentially observable density of cosmic strings created in the reheating process.

On the other hand, many interesting questions in string cosmology await further understanding of time-dependent backgrounds in the string theory.

## Retrospect

It is remarkable that the formal framework of the string theory admits tractable solutions with reasonable resemblance to the structure of the standard model. In particular, generic features such as nonabelian gauge symmetry and chirality, coupled to gravity, are generic in 4D compactifications. This is already a success. In addition, much progress has been made in the general description of the relevant mathematical tools, and physical mechanisms and ingredients involved in these vacua, as well as in the explicit construction of models with the standard model spectrum (or supersymmetric extensions of it). Yet, many questions remain open and much more work is needed in order to make contact with the physics observed in nature.

*See also:* Brane Worlds; Compactification of Superstring Theory; Cosmology: Mathematical Aspects; Superstring Theories.

## Further Reading

Acharya B and Witten E (2001) Chiral fermions from manifolds of G(2) holonomy, hep-th/0109152.

Aldazabal G, Ibáñez LE, Quevedo F, and Uranga AM (2000) D-branes at singularities: a bottom up approach to the string embedding of the standard model. *Journal of High Energy Physics* 0008: 002.

Angelantonj C and Sagnotti A (2002) Open strings. *Physics Reports* 371: 1–150.

Angelantonj C and Sagnotti A (2003) Open strings – erratum. *Physics Reports* 376: 339–405.

Antoniadis I, Arkani-Hamed N, Dimopoulos S, and Dvali GR (1998) New dimensions at a millimeter to a Fermi and superstrings at a TeV. *Physics Letters B* 436: 257–263.

Bachas C (1995) A way to break supersymmetry, hep-th/9503030.

Blumenhagen R, Cvetič M, Langacker P, and Shiu G (2005) Toward realistic intersecting D-brane models, hep-th/0502005.

Candelas P, Horowitz GT, Strominger A, and Witten E (1985) Vacuum configurations for superstrings. *Nuclear Physics B* 258: 46–74.

Donagi R, He Y-H, Ovrut BA, and Reinbacher R (2004) The spectra of heterotic standard model vacua, hep-th/0411156.

Green MB, Schwarz JH, and Witten E (1987) *Superstring Theory*. Cambridge Monographs On Mathematical Physics, vols. 1 and 2. Cambridge: Cambridge University Press.

Ibáñez LE (1987) The search for a standard model SU(3) × SU(2) × U(1) superstring: an introduction to orbifold constructions. *Seoul Sympos*. 1986, 46.

Polchinski J (1998) *String Theory*. vols. 1 and 2. Cambridge: Cambridge University Press.

Uranga AM (2003) Chiral four-dimensional string compactifications with intersecting D-branes. *Classical and Quantum Gravity* 20: S373–S394.

Witten E (1996) Strong coupling expansion of Calabi–Yau compactification. *Nuclear Physics B* 471: 135–158.

# String Topology: Homotopy and Geometric Perspectives

**R L Cohen**, Stanford University, Stanford, CA, USA



String topology is a new field of study involving the geometric and algebraic topology of spaces of loops and paths in manifolds. The subject was initiated in the important work of Chas and Sullivan (1999) who uncovered previously unknown algebraic structure in the homology and equivariant homology of loop spaces. While the structure is purely topological, it was motivated by formalisms in quantum field theory and string theory. Since that time this subject has attracted the attention of many mathematicians, but one of the main lines of research continues to be motivated by the attempt to understand the relation between this structure (and its generalizations) with topological and conformal field theories.

In order to describe some of the recent advances in this field, we begin with some notation. Throughout this article $M^n$ will denote a closed, $n$-dimensional, oriented manifold. $LM$ will denote the free loop space,

$$LM = \mathrm{Map}(S^1, M)$$

For $D_1, D_2 \subset M$ closed submanifolds, $\mathcal{P}_M(D_1, D_2)$ will denote the space of paths in $M$ that start at $D_1$ and end at $D_2$,

$$\mathcal{P}_M(D_1, D_2) = \{\gamma : [0,1] \to M, \gamma(0) \in D_1, \gamma(1) \in D_2\}$$

The paths and loops we consider will always be assumed to be piecewise smooth. Such spaces of paths and loops are well known to be infinite-dimensional manifolds, and roughly speaking, string topology is the study of the intersection theory in these manifolds.

Recall that for closed, oriented manifolds, there is an intersection pairing,

$$H_r(M) \times H_s(M) \to H_{r+s-n}(M)$$

which is defined to be Poincaré dual to the cup product,

$$H^{n-r}(M) \times H^{n-s}(M) \xrightarrow{\cup} H^{2n-r-s}(M)$$

The geometric significance of this pairing is that if the homology classes are represented by submanifolds, $P^r$ and $Q^s$ with transverse intersection, then the image of the intersection pairing is represented by the geometric intersection, $P \cap Q$.

The remarkable result of Chas and Sullivan says that even without Poincaré duality, there is an intersection type product

$$\mu : H_p(LM) \times H_q(LM) \to H_{p+q-n}(LM)$$

that is compatible with both the intersection product on $H_*(M)$ via the map $ev : LM \to M (\gamma \to \gamma(0))$, and with the Pontrjagin product in $H_*(\Omega M)$.

The construction of this pairing involves consideration of the diagram,

$$LM \xleftarrow{\gamma} \mathrm{Map}(8, M) \xrightarrow{e} LM \times LM \qquad [1]$$

Here $\mathrm{Map}(8, M)$ is the mapping space from the figure 8 to $M$, which can be viewed as the subspace of $LM \times LM$ consisting of those pairs of loops that agree at the basepoint. $\gamma : \mathrm{Map}(8, M) \to LM$ is the map on mapping spaces induced by the pinch map $S^1 \to S^1 \vee S^1$.

Chas and Sullivan constructed this pairing by studying intersections of chains in loop spaces. A more homotopy-theoretic viewpoint was taken by Cohen and Jones (2002) who viewed $e : \mathrm{Map}(8, M) \to LM \times LM$ as an embedding, and showed there is a tubular neighborhood homeomorphic to a normal given by the pullback bundle, $ev^*(TM)$, where $ev : LM \to M$ is the evaluation map mentioned above. They then constructed a Pontrjagin–Thom collapse map whose target is the Thom space of the normal bundle, $\tau_e : LM \times LM \to \mathrm{Map}(8, M)^{ev^*(TM)}$. Computing $\tau_e$ in homology and applying the Thom isomorphism defines an "umkehr map,"

$$e_! : H_*(LM \times LM) \to H_{*-n}(\mathrm{Map}(8, M))$$

The Chas–Sullivan loop product is defined to be the composition

$$\mu_* = \gamma_* \circ e_! : H_*(LM \times LM) \to H_{*-n}(\mathrm{Map}(8, M)) \to H_{*-n}(LM)$$

Notice that the umkehr map $e_!$ can be defined for a generalized homology theory $h_*$ whenever one has a Thom isomorphism of the tangent bundle, $TM$, which is to say a generalized homology theory $h_*$ for which the representing spectrum is a ring spectrum, and which supports an orientation of $M$.

By twisting the Pontrjagin–Thom construction by the virtual bundle $-TM$, one obtains a map of spectra,

$$\tau_e : LM^{-TM} \wedge LM^{-TM} \to \mathrm{Map}(8, M)^{ev^*(-TM)}$$

where $LM^{-TM}$ is the Thom spectrum of the pullback of the virtual bundle $ev^*(-TM)$. Now we can compose, to obtain a multiplication,

$$LM^{-TM} \wedge LM^{-TM} \xrightarrow{\tau_e} \mathrm{Map}(8, M)^{ev_0^*(-TM)} \xrightarrow{\gamma} LM^{-TM}$$

The following was proved by Cohen and Jones (2002).

**Theorem 1** *Let M be a closed manifold, then $LM^{-TM}$ is a ring spectrum. If M is orientable the ring structure on $LM^{-TM}$ induces the Chas–Sullivan loop product on $H_*(LM)$ by applying homology and the Thom isomorphism.*

The ring structure on the spectrum $LM^{-TM}$ was also observed by Dwyer and Miller using different methods.

Cohen and Godin (2004) generalized the loop product in the following way. Observe that the figure 8 is homotopy equivalent to the pair of pants surface $P$, which we think of as a genus 0 cobordism between two circles and one circle.

Furthermore, **Figure 1** is homotopic to the diagram of mapping spaces,

$$LM \xleftarrow{\rho_{\text{out}}} \text{Map}(P, M) \xrightarrow{\rho_{\text{in}}} (LM)^2$$

where $\rho_{\text{in}}$ and $\rho_{\text{out}}$ are restriction maps to the "incoming" and "outgoing" boundary components of the surface $P$. So the loop product can be viewed as a composition,

$$\begin{aligned}\mu &= \mu_P \\ &= (\rho_{\text{out}})_* \circ (\rho_{\text{in}})_! : (H_*(LM))^{\otimes 2} \to H_*(\text{Map}(P, M)) \\ &\quad \to H_*(LM)\end{aligned}$$

where using the figure 8 to replace the surface $P$ can be viewed as a technical device that allows one to define the umkehr map $(\rho_{in})_!$.

In general if one considers a surface of genus $g$, viewed as a cobordism from $p$ incoming circles to $q$ outgoing circles, $\Sigma_{g,p+q}$, one gets a similar diagram (**Figure 2**)

$$(LM)^q \xleftarrow{\rho_{\text{out}}} \text{Map}(\Sigma_{g,p+q}, M) \xrightarrow{\rho_{\text{in}}} (LM)^p$$

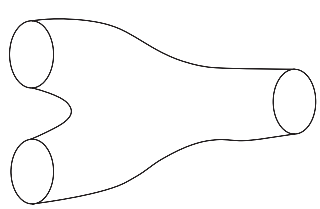


**Figure 1** Pair of pants *P*.

Cohen and Godin (2004) used the theory of "fat" or "ribbon" graphs to represent surfaces as developed by Harer (1985), Penner (1987), and Strebel (1984), in order to define Pontrjagin–Thom maps,

$$\tau_{\Sigma_{g,p+q}} : (LM)^p \to \text{Map}(\Sigma_{g,p+q}, M)^{\nu(\Sigma_{g,p+q})}$$

where $\nu(\Sigma_{g,p+q})$ is the appropriately defined normal bundle of $\rho_{\text{in}}$. By applying (perhaps generalized) homology and the Thom isomorphism, they defined the umkehr map,

$$(\rho_{\text{in}})_! : H_*((LM)^p) \to H_{*+\chi(\Sigma_{g,p+q})\cdot n}(\text{Map}(\Sigma_{g,p+q}, M))$$

where $\chi(\Sigma_{g,p+q}) = 2 - 2g - p - q$ is the Euler characteristic. Cohen and Godin then defined the string topology operation to be the composition,

$$\begin{aligned}\mu_{\Sigma_{g,p+q}} = \rho_{\text{out}} \circ (\rho_{\text{in}})_! : H_*((LM)^p) \to H_{*+\chi(\Sigma_{g,p+q})\cdot n} \\ \times (\text{Map}(\Sigma_{g,p+q}, M)) \to H_{*+\chi(\Sigma_{g,p+q})\cdot n}((LM)^q)\end{aligned}$$

They proved that these operations respect gluing of surfaces,

$$\mu_{\Sigma_1 \# \Sigma_2} = \mu_{\Sigma_2} \circ \mu_{\Sigma_1}$$

where $\Sigma_1 \# \Sigma_2$ is the glued surface as shown in **Figure 3**.

The coherence of these operations is summarized in the following theorem.

**Theorem 2** (Cohen and Godin 2004). *Let $h_*$ be any multiplicative generalized homology theory that supports an orientation of M. Then the assignment*

$$\Sigma_{g,p+q} \to \mu_{\Sigma_{g,p+q}} : h_*((LM)^p) \to h_*((LM)^q)$$

*is a positive boundary topological quantum field theory. "Positive boundary" refers to the fact that the number of outgoing boundary components, q, must be positive.*

A theory with open strings was initiated by Sullivan (2004) and developed further by A Ramirez (2005) and by Harrelson (2004). In this

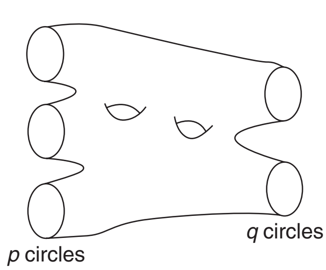



**Figure 2** $\Sigma_{g,p+q}$.

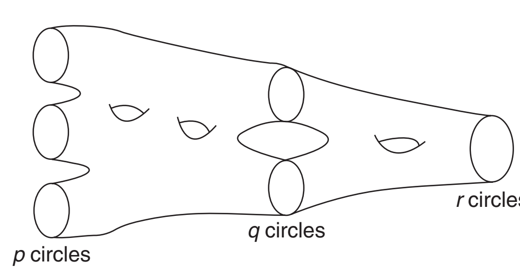



**Figure 3** $\Sigma_1 \# \Sigma_2$.

setting one has a collection of submanifolds, $D_i \subset M$, referred to as "D-branes." This theory studies intersections in the path spaces $P_M(D_i, D_j)$.

A theory with D-branes involves "open–closed cobordisms" which are cobordisms between compact one-dimensional manifolds whose boundary is partitioned into three parts:

1. Incoming circles and intervals.
2. Outgoing circles and intervals.
3. The rest is the "free boundary" which is itself a cobordism between the boundary of the incoming and boundary of the outgoing intervals. Each connected component of the "free boundary" is labeled by a D-brane (see **Figure 4**).

In a topological field theory with D-branes, one associates to each boundary circle a vector space $V_{S^1}$ (in our case $V_{S^1} = H_*(LM)$) and to an interval whose endpoints are labeled by $D_i, D_j$, one associates a vector space $V_{D_i,D_j}$ (in our case $V_{D_i,D_j} = H_*(P_M(D_i, D_j))$).

To an open–closed cobordism as above, one associates an operation from the tensor product of these vector spaces corresponding to the incoming boundaries to the tensor product of the vector spaces corresponding to the outgoing boundaries. Of course, these operations have to respect the relevant gluing of open–closed cobordisms.

By developing a theory of fat graphs that encode the open–closed boundary data, Ramirez was able to prove that there are string topology operations that form a positive boundary, topological quantum field theory with D-branes (Ramirez 2005).

We end these notes by a discussion of three applications of string topology to classifying spaces of groups.

**Example 1** Application to Poincaré duality groups – (Abbaspour *et al.* to appear). For $G$ any discrete group, one has that the loop space of the classifying space satisfies

$$LBG \simeq \coprod_{[g]} BC_g$$

where $[g]$ is the conjugacy class determined by $g \in G$, and $C_g < G$ is the centralizer of $g$.

When $BG$ is represented by a closed manifold, or more generally, when $G$ is a Poincaré duality group, the Chas–Sullivan loop product then defines pairings among the homologies of the centralizer subgroups. Abbaspour *et al.* describe this loop product entirely in terms of group homology, thus giving structure to the homology of Poincaré-duality groups that previously had not been known.

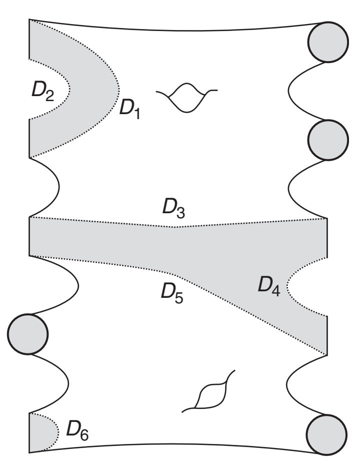



**Figure 4** Open–closed cobordism.

**Example 2** Applications to 3-manifolds. (Abbaspour 2005). Let $\iota : H_*M \to H_*(LM)$ be induced by inclusion of constant loops. This is a split injection of rings. Write $H_*(LM) = H_*(M) \oplus A_M$. We say $H_*(LM)$ has nontrivial extended loop products if the composition

$$A_M \otimes A_M \hookrightarrow H_*(LM) \otimes H_*(LM) \xrightarrow{\mu} H_*(LM)$$

is nontrivial.

Let $M$ be a closed, irreducible 3-manifold. In a remarkable piece of work, Abbaspour showed the relationship between having a trivial extended loop product and $M$ being "algebraically hyperbolic." This means that $M$ is a $K(\pi, 1)$ and its fundamental group has no rank-2 abelian subgroup. (If geometrization conjecture is true, this is equivalent to $M$ admitting a complete hyperbolic metric.)

**Example 3** The string topology of classifying spaces of compact Lie groups (Gruher (to appear) and of Gruher and Salvatore (to appear)). The goal of Gruher's work is to construct string topological invariants of $LBG \simeq EG \times_G G$, where $G$ acts on itself via conjugation. Ultimately, one would like to understand the relationship between this structure and the work of Freed (2003) on twisted equivariant $K$-theory, $K_G^\tau(G)$ and the Verlinde algebra.

The first observation in this program was to notice that the key ingredient in the forming of the Chas–Sullivan loop product is that the fibration $ev : LM \to M$ is a fiberwise monoid over a closed oriented manifold. The fiber is $\Omega M$, which has the usual Pontrjagin product.

The following was proved by Gruher and Salvatore:

**Lemma 3** *Let $G \to E \to M$ be a fiberwise monoid over a closed manifold $M$. Then $E^{-TM}$ is a ring spectrum.*

The following construction gives a large supply of examples of such fiberwise monoids over manifolds.

Let $G \to P \to M$ be a principal $G$ bundle over a closed manifold $M$. We can construct the corresponding adjoint bundle,

$$\mathrm{Ad}(P) = P \times_G G \to M$$

It is an easy observation that $G \to \mathrm{Ad}(P) \to M$ is a fiberwise monoid.

**Theorem 4** $\mathrm{Ad}(P)^{-TM}$ *is a ring spectrum. This ring structure is natural with respect to maps of principal G-bundles.*

Let $BG$ be classifying space of compact Lie groups. It is possible to construct a filtration of $BG$,

$$M_1 \hookrightarrow M_2 \hookrightarrow \cdots \hookrightarrow M_i \subset M_{i+1} \hookrightarrow \cdots \hookrightarrow BG$$

where the $M_i$'s are compact, closed manifolds. An example of this is filtering $BU(n)$ by Grassmannians.

Let $G \to P_i \to M_i$ be the restriction of $EG \to BG$. By the above theorem one obtains an inverse system of ring spectra

$$P_1^{-TM_1} \leftarrow P_2^{-TM_2} \leftarrow \cdots \leftarrow P_i^{-TM_i} \leftarrow P_{i+1}^{-TM_{i+1}} \leftarrow \cdots$$

**Theorem 5** *The homotopy type of this pro-ring-spectrum is a well-defined invariant of BG. It is referred to as the "string topology of BG."*

## Potential Application: Twisted K-theory and the Verlinde Algebra

Let $G$ be a connected, compact Lie group. Using the observation that the loop space of a classifying space is the classifying space of the loop group, $L(BG) \simeq B(LG)$, the string topology gives new structure on the classifying space of these loop groups. In particular, one has new structure on the $K$-theory of these classifying spaces. Now classical results of Atiyah and Segal suggest that $K$-theory of classifying spaces should be related to the representation theory of the group. In this case, the representation theory of loop groups has been widely studied and is very important in conformal field theory.

Understanding the precise relationship between the string topology of the classifying space and this representation theory is an interesting area of current research. To motivate this, first recall that the loop space, $LBG$, has a well-known description as

$$LBG \simeq EG \times_{\mathrm{Ad}} G$$

where the right-hand side refers to the homotopy orbit space of the conjugation (or adjoint) action of $G$ on itself. Thus, the homology $H_*(LBG)$ is the equivariant homology $H_*^G(G)$. Similarly, the $K$-theory $K^*(LBG)$ maps to the equivariant $K$-theory, $K_G^*(G)$. Now in recent work of Freed (2003) twisted equivariant $K$-homology, $K_G^\tau(G)$ was shown to be isomorphic to the Verlinde algebra. This algebra is a space of representations of the loop group, $LG$. The multiplication in this algebra is the "fusion product," coming from conformal field theory. One topic of current research is to understand the relationship between multiplicative structure coming from the string topology of $BG$, and this fusion product in the Verlinde algebra. More generally, the goal is to bring to bear the considerable calculational techniques of algebraic topology that are available in string topology, to understand the recently uncovered field theoretic structure of twisted $K$-theory (Freed 2003), and its applications to string theory.

## Acknowledgment

The author was partially supported by a grant from the NSF.

*See also:* Mathematical Knot Theory; Topological Defects and Their Homotopy Classification.

## Further Reading


Abbaspour H (2005) On string topology of three manifolds. *Topology* 44: 1059–1091.

Abbaspour H, Cohen RL, and Gruher K *String Topology of Poincaré Duality Groups* (in preparation).

Chas M and Sullivan D (1999) String topology. To appear in *Annals of Mathematics*. Preprint: math.GT/9911159.

Cohen RL and Godin V (2004) A polarized view of string topology. In: *Topology, Geometry, and Quantum Field Theory,* London Math. Soc. Lecture Notes, vol. 308, pp. 127–154.

Cohen RL and Jones JDS (2002) A homotopy theoretic realization of string topology. *Mathematisches Annalen* 324: 773–798.

Freed D, Hopkins M, and Teleman C Twisted $K$-theory and loop group representations. Preprint: math.AT/0312155.

Godin V (2004) *A Category of Bordered Fat Graphs and the Mapping Class Group of a Bordered Surface*. Ph.D. thesis, Stanford University.

Gruher K Ph.D. thesis, Stanford University (in preparation).

Gruher K and Salvatore P *String Topology of Classifying Spaces* (in preparation).

Harrelson E (2004) On the homology of open–closed string theory. Preprint: math.AT/0412249.

Harer JL (1985) Stability of the homology of the mapping class groups of orientable surfaces. *Annals of Mathematics* 121: 215–249.

Penner R (1987) The decorated Teichmuller space of punctured surfaces. *Communications in Mathematical Physics* 113: 299–339.

Ramirez A (2005) *Open–Closed String Topology*. Ph.D thesis, Stanford University.

Strebel K (1984) *Quadratic Differentials*. Berlin: Springer.

Sullivan D (2004) Open and closed string field theory interpreted in classical algebraic topology. In: Tillman U (ed.) *Topology, Geometry, and Quantum Field Theory*, London Math. Soc. Lecture Notes, vol. 308, pp. 344–357. Cambridge: Cambridge University Press.

# Superfluids

**D Einzel**, Bayerische Akademie der Wissenschaften, Garching, Germany



## Introduction

Superfluidity has been known to exist since the 1930s. This widespread phenomenon occurs in many-particle Bose and Fermi systems as different as liquid $^4$He, liquid $^3$He, atomic gases like Rb and Li, atomic nuclei, pulsars and last, but not least, in metals, where the itinerant electrons may become superfluid. This article is devoted to a unifying theoretical description of Bose and Fermi superfluidity. The mechanisms leading to superfluidity include Bose–Einstein condensation (BEC) and Bardeen, Cooper, and Schrieffer (BCS)–Leggett pairing correlations. We hope to be able to demonstrate why this fascinating phenomenon is – even roughly 80 years after its experimental discovery and its first theoretical explanation – still a subject of intensive research.

The phenomenon of superfluidity is closely connected with the apparent lack of any measurable flow resistance, which scales with the shear viscosity of the fluid. Its complete absence implies that the system is frictionless moving with zero viscosity. The observation of superfluidity is usually precluded by the solidification of most liquids as the temperature is lowered. Only systems with particularly light atoms (like the helium isotopes $^4$He and $^3$He) stay liquid down to the lowest temperatures. These systems are referred to as "quantum liquids," since their liquid state is caused by the quantum-mechanical zero-point motion of the atoms. It should be noted that the Helium isotopes belong to two different kinds of elementary particles which can be distinguished by their statistics: $^4$He is a spin-0 boson and $^3$He a spin-1/2 fermion.

In 1924, Satyendra Nath Bose and Albert Einstein proposed that below a characteristic degeneracy temperature $T_B$, a macroscopic number of bosons can condense into the state of lowest energy $\epsilon_k = 0$. In the 1930s, Fritz London and Heinz London showed that this so-called Bose–Einstein condensate can be described by a macroscopic quantum-mechanical wave function like the one for a single elementary particle, but with the probability density replaced by the density of the condensed particles. By the end of the 1930s, the experimental results of Allen, Kamerlingh–Onnes, Keesom, Kapitza, Miesener, Wolfke, and others accumulated the evidence that liquid $^4$He undergoes a second-order phase transition at $T_\lambda = 2.17\,K$ to a state referred to as a superfluid, since the liquid could flow without any sign of a flow resistance. This superfluid state was interpreted in terms of Bose condensation of the $^4$He atoms in the liquid (London 1938).

In Figure 1 the $P$–$T$ phase diagram of liquid $^4$He is shown with a normal liquid phase, a solid phase and the superfluid phase below the $\lambda$-line at about 2 K.

Fermions cannot condense in a way similar to the BEC, due to the Pauli exclusion principle. In 1957 Bardeen, Cooper, and Schrieffer came up with their ingenious proposal that the superfluidity of the electron system (usually referred to as superconductivity) comes about through the formation of fermion pairs (quasibosons) in $\boldsymbol{k}$-space in a spin-singlet state. In 1971, several superfluid phases of liquid $^3$He at a few mK were discovered by Lee, Osheroff, and Richardson at Cornell University. Experimental aspects connected with the spin degrees of freedom of the quantum liquid gave strong evidence for Cooper pairing of the $^3$He atoms in a spin-triplet state. In Figure 2 the zero-field $P$–$T$ phase diagram of liquid $^3$He is shown with a normal (Fermi) liquid phase, a solid phase and the superfluid A and B phases.

Immediately after this discovery, Anthony J Leggett applied the BCS ideas to liquid $^3$He and introduced a generalized scheme, that allowed for triplet-pairing correlations. His theory turned out to describe a large variety of experimental results accurately. A new and exciting development set in when Bose–Einstein condensates were discovered for the first time in dilute gases of alkali atoms in 1995 by Cornell and Wiemann *et al.* (Rb), Ketterle *et al.* (Na), and Hulet *et al.* (Li).

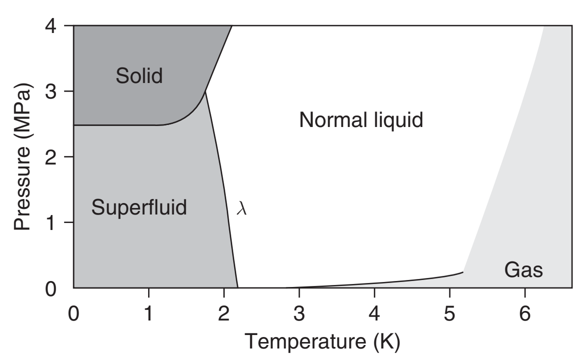



**Figure 1** The phase diagram of liquid $^4$He. Courtesy of Erkki Thuneberg.

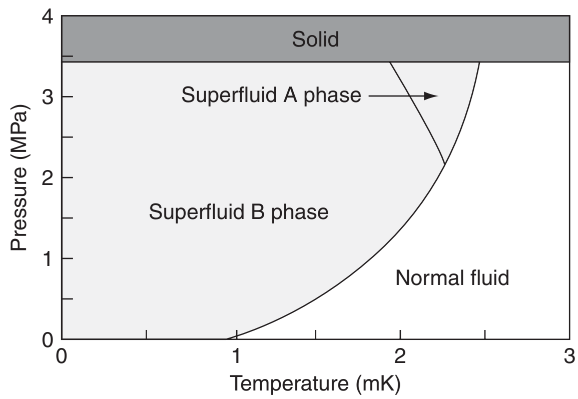



**Figure 2** The phase diagram of liquid $^3$He. Courtesy of Erkki Thuneberg.

## Boson and Fermion Degeneracy

In what follows, the energy dispersion of Bose and Fermi systems is denoted as $\epsilon_{\boldsymbol{k}}$ (free bosons/fermions would be represented by $\epsilon_{\boldsymbol{k}} = \hbar^2 \boldsymbol{k}^2/2m$). A large number of bosons can occupy Bose quantum states $|\boldsymbol{k}\rangle$, the average occupation is dictated by the Bose–Einstein distribution

$$n_{\boldsymbol{k}} = \frac{1}{e^{(\epsilon_{\boldsymbol{k}}-\mu)/k_B T} - 1} \qquad [1]$$

For Bose systems, the chemical potential is negative $\mu = -k_B T\alpha$ and $\alpha$ is fixed by the condition

$$n = \frac{1}{V}{\sum_{\boldsymbol{k}}}' n_{\boldsymbol{k}} = \frac{1}{\lambda_T^3} B_{\frac{3}{2}}(\alpha) \qquad [2]$$

where the prime indicates the summation over excited states $|\boldsymbol{k}| > 0$. In [2], $\lambda_T = h/\sqrt{2\pi m k_B T}$ denotes the thermal de Broglie wavelength which provides a criterion for the importance of quantum effects or degeneracy through $n\lambda_T^3 \geq O(1)$. The Bose integrals $B_\sigma(\alpha)$ originate from the conversion of the momentum sum into an energy integral and read for parabolic dispersion:

$$B_\sigma(\alpha) = \frac{1}{\Gamma(\sigma)} \int_0^\infty \frac{dy\, y^{\sigma-1}}{e^{y+\alpha} - 1} = \sum_{\nu=1}^{\infty} \frac{e^{-\nu\alpha}}{\nu^\sigma} \qquad [3]$$

with $B_\sigma(0) = \zeta(\sigma), \Gamma$ the Euler $\Gamma$-function and $\zeta$ denoting the Riemann $\zeta$-function. It is important to understand that in order to have a constant total density $n, B_{3/2}(\alpha)$ has to increase $\propto T^{-3/2}$ in the same way as $\lambda_T^3$. This is, however, impossible at all temperatures since the chemical potential of the Bose gas vanishes $(\alpha \to 0)$ at a finite temperature $T_B$ given by

$$T_B = \frac{2\pi\hbar^2}{mk_B}\left[\frac{n}{\zeta(3/2)}\right]^{2/3} \qquad [4]$$

for which $n\lambda_{T_B}^3 = B_{3/2}(0) = \zeta(3/2) = 2.612\ldots$.

In sharp contrast, fermions obey the Pauli exclusion principle, which states that only one fermion can occupy a quantum state $|\boldsymbol{k}, \sigma\rangle$ specified in addition by the spin projection $\sigma$. The average statistical occupation is given by the Fermi–Dirac distribution

$$f_{\boldsymbol{k}} = \frac{1}{e^{(\epsilon_{\boldsymbol{k}}-\mu)/k_B T} + 1} \qquad [5]$$

Figure 3 shows a comparison of Bose–Einstein and Fermi–Dirac momentum distributions $n_{\boldsymbol{k}}$ plotted vs. $\epsilon_{\boldsymbol{k}}$. The chemical potential is shown for fermions only, $\mu_F = k_B T\alpha$ is always positive and the total density can be expressed as

$$n = \frac{1}{V}\sum_{\boldsymbol{k},\sigma} f_{\boldsymbol{k}} = \frac{2}{\lambda_T^3} F_{\frac{3}{2}}(\alpha) \qquad [6]$$

where the factor of 2 originates from the spin degeneracy. For parabolic dispersion, the Fermi integral reads:

$$F_\sigma(\alpha) = \frac{1}{\Gamma(\sigma)} \int_0^\infty \frac{dy\, y^{\sigma-1}}{e^{y-\alpha} + 1} \overset{T\to 0}{=} \frac{(\mu/k_B T)^\sigma}{\Gamma(\sigma+1)} \qquad [7]$$

One recognizes that the degeneracy condition $n\lambda_T^{3/2} \geq 1$ corresponds to the limit $T \leq T_F = \mu(0)/k_B$, which is connected with the formation of a "Fermi sea," with $\mu(0) \equiv E_F$ the Fermi energy:

$$\mu \overset{T\to 0}{=} \frac{\hbar^2}{2m}(3\pi^2 n)^{\frac{2}{3}} = E_F \qquad [8]$$

To summarize, quantum behavior in Bose and Fermi system sets in below the degeneracy temperature $T^*$, defined through $n\lambda_{T^*}^3 = O(1)$. For bosons, $T^* = T_B$ is the temperature at which the chemical potential vanishes, whereas for fermions $T^* = T_F$ is the Fermi temperature.

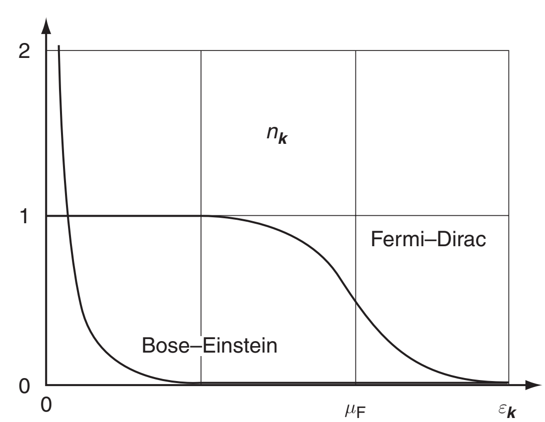



**Figure 3** The Fermi and Bose momentum distribution.

## London Quantum Hydrodynamics

For a general treatment of the quantum-mechanical origin of the equations describing Bose and Fermi superfluidity, it is convenient to introduce a parameter $\nu$ which describes single bosons ($\nu = 1$) or Fermion pairs ($\nu = 2$) of mass $M = \nu m$. The basic assumption (London 1938) is that the laws of quantum mechanics are applicable also to a macroscopic number of single ($\nu = 1$) or composite ($\nu = 2$) particles of density $\rho^s/\nu m$, the so-called condensate, which is represented by a macroscopic wave function $\psi(\boldsymbol{r}, t)$. $\psi$ has the property

$$\psi(\boldsymbol{r}, t)\psi^*(\boldsymbol{r}, t) = \frac{\rho^s(\boldsymbol{r}, t)}{\nu m}; \quad \nu = 1, 2$$

The dynamics of the condensate is governed by the Schrödinger equation

$$\mathrm{i}\frac{\hbar}{\nu}\frac{\partial\psi}{\partial t} = \left(-\frac{\hbar^2\nabla^2}{2\nu^2 m} + \mu\right)\psi \qquad [9]$$

in which $\mu$ represents the condensate's chemical potential. After performing a Madelung transformation (Madelung 1926):

$$\psi = a\mathrm{e}^{\mathrm{i}\varphi}, \quad a^2 = \frac{\rho^s}{\nu m}$$

one arrives at two coupled hydrodynamic equations, the first of which reads

$$\begin{gathered}\frac{\partial\rho^s}{\partial t} + \nabla\cdot\boldsymbol{j}^s_m = 0 \\ \boldsymbol{j}^s_m = \rho^s\boldsymbol{v}^s, \quad \boldsymbol{v}^s = \frac{\hbar}{\nu m}\nabla\varphi\end{gathered} \qquad [10]$$

Equation [10] can be interpreted as a continuity equation, which represents the conservation law for the condensate mass density $\rho^s$. The second equation

$$-\frac{\hbar}{\nu}\frac{\partial\varphi}{\partial t} = \frac{1}{2}m v^{s2} + \mu + \mathrm{O}(\hbar^2\nabla^2) \qquad [11]$$

assumes the form of the Hamilton–Jacobi equation for the action field of classical mechanics $\hbar\varphi$, if the quasiclassical limit (terms $\propto \mathrm{O}(\hbar^2\nabla^2) \to 0$) is taken.

From [10] and [11] a condensate acceleration equation can be derived, which resembles the Euler equation of classical hydrodynamics ($\mu = \mu_0 + \delta\mu$):

$$\frac{\partial\boldsymbol{v}^s}{\partial t} + (\boldsymbol{v}^s\cdot\nabla)\boldsymbol{v}^s = -\frac{1}{m}\nabla\delta\mu \qquad [12]$$

The physical nature of the driving force becomes evident after applying the Gibbs–Duhem relation $n\delta\mu = \delta P - \sigma_0\delta T$. Finally, the acceleration of the mass supercurrent $\boldsymbol{j}^s_m$ is of the form

$$\frac{\partial\boldsymbol{j}^s_m}{\partial t} = -\frac{\rho^s}{\rho}\nabla(\delta P - \sigma_0\delta T) \qquad [13]$$

It turns out that the London equations [10] and [13], in which $\rho^s$ is an unknown phenomenological parameter, explain many experimental observations such as persistent currents, U-tube oscillations, thermomechanical (e.g., fountain-) effects, beaker flow phenomena, and many others.

## Bose–Einstein Condensation (BEC)

In order to understand the macroscopic quantum state in case of Bose systems, we consider first the simple case of a Bose gas. Let us decompose the energy eigenstates $\epsilon_{\boldsymbol{k}}$ into those with $\epsilon_{\boldsymbol{k}} = \epsilon_0 = 0$ (condensate) and average occupation number

$$n_0 = \frac{N_0}{V} = \frac{1}{V}\frac{1}{\mathrm{e}^{\alpha} - 1} \qquad [14]$$

and those with $\epsilon_{\boldsymbol{k}} > 0$ (excited states) and average occupation number

$$n_{ex} = \frac{N_{ex}}{V} = \sum_{\boldsymbol{k}}{}' n_{\boldsymbol{k}} = n\frac{B_{3/2}(\alpha)}{B_{3/2}(0)}\left(\frac{T}{T_B}\right)^{3/2} \qquad [15]$$

with the total density $n = n_{ex} + n_0$. The consequence of the chemical potential vanishing at $T_B$ clearly is a macroscopic occupation of the ground state of the Bose gas:

$$N_0 \overset{\alpha\to 0}{=} \frac{1}{1 + \alpha + \cdots - 1} = \frac{1}{\alpha} \to \infty \qquad [16]$$

This phenomenon is referred to as BEC. Below $T_B, \alpha = 0$ and from [15] we see that

$$n_{ex} \overset{\text{free bosons}}{=} n\left(\frac{T}{T_B}\right)^{3/2}, \quad T < T_B \qquad [17]$$

The average occupation of the ground state is given by

$$n_0(T) = n - n_{ex}(T), \quad T < T_B \qquad [18]$$

It is important to understand that the number density of condensed particles $n_{ex}$ has nothing to do with the current response function $\rho^s$ (eqn [10]). A derivation of $\rho^s$ will be given in the section "Local response of condensates and excitation gases."

Let us now discuss the structure of the excitation spectrum, which will turn out to be crucial for the observability of superfluidity, in some more detail. Suppose that a macroscopic object of mass $M$ moves through the superfluid. Then one may ask the question, at what velocity does this motion cause the creation of an excitation of energy $E_{\boldsymbol{p}}$ and momentum $\boldsymbol{p}$. The condition can be formulated in terms of the velocity

difference $v_i - v_f$ as $E_p = M(v_i^2 - v_f^2)/2$ and $p = M(v_i - v_f)$. Eliminating $v_f$ yields $\epsilon_p = p \cdot v_i + O(M^{-1})$ so that condition for the creation of an excitation leads to the so-called Landau critical velocity

$$v_L = \min\left\{\frac{E_p}{|p|}\right\} > 0 \qquad [19]$$

It is immediately clear that for free bosons $v_L = 0$. This means that a free Bose gas can never be a superfluid, since drag forces on moving objects will start to act even at smallest velocities.

It turns out that interaction effects can drastically modify the nature of the elementary excitations. In 1947, Nikolai Bogoliubov showed (for the first time using the method of second quantization) that even in the limit of weak repulsive interactions the excitation spectrum is phonon-like $E_p = c|p|$, with $c$ the sound velocity. Lev Landau and Richard Feynman investigated the situation for superfluid $^4$He, where the interactions between the atoms are far from weak. Landau (1947) postulated the following form for the excitation spectrum, for which Feynman (1953) gave the microscopic justification. At low momenta, the spectrum is phonon-like and linear in $p$:

$$\lim_{p \to 0} E_p = E_p^{\text{phon}} = c|p| \qquad [20]$$

At higher momenta, the spectrum is reminiscent of that of crystal phonons in that $E_p$ passes though a maximum, and then, at a characteristic momentum $p_0$ approaches the next minimum, which, however, is located at a finite energy $\Delta$. Feynman called this part of the spectrum the "roton" (mass $m_r$) in an analogy with a "smoke ring," since it is connected with the forward motion of a particle accompanied by a ring of back-flowing other particles:

$$\lim_{|p| \to p_0} E_p = E_p^{\text{rot}} = \Delta + \frac{(|p| - p_0)^2}{2m_r} \qquad [21]$$

Figure 4 shows a sketch of the phonon–roton spectrum of superfluid $^4$He. Clearly, the Landau critical velocity for the phonon–roton spectrum is characterized by the roton minimum and is given by $v_L \approx \Delta/p_0$.

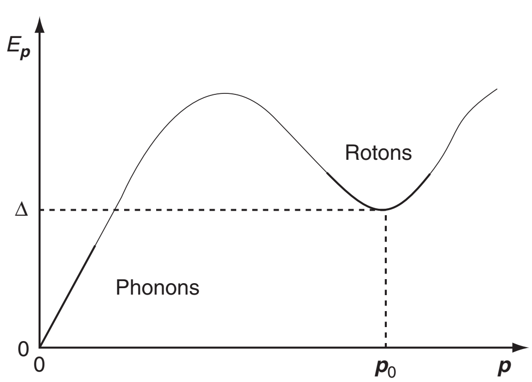



**Figure 4** The phonon–roton spectrum.

## BCS–Leggett Pair Condensation

The key assumptions of the weak-coupling mean-field BCS–Leggett pairing model can be summarized as follows: one first assumes that at sufficiently low temperatures it is energetically favorable that a temperature-dependent part of the fermions forms so-called Cooper pairs. This pair formation is caused by an attractive interaction in $k$-space near the Fermi surface:

$$\Gamma_{kp}^{(s)} < 0, \qquad |\xi_k|, |\xi_p| < \epsilon_c$$

Here $\xi_k = \epsilon_k - \mu$ measures the energy from the chemical potential. The index $s$ denotes the total spin of the pair. Classical superconductors have pairs in a relative singlet state $s = 0, m_s = 0$ whereas the superfluid phases of liquid $^3$He have pairs in a relative spin-triplet state $s = 1, m_s = 0, \pm 1$, with $m_s$ the magnetic quantum number. The amplitude of spontaneous pair formation is

$$g_{k\sigma_1\sigma_2} \equiv \langle \hat{c}_{-k\sigma_1} \hat{c}_{k\sigma_2} \rangle \neq 0, \qquad T \leq T_c \qquad [22]$$

with $k = k_1 - k_2$ the relative momentum of the pair. The attractive interaction that drives the Cooper-pair formation connects the pairing amplitude $g_{k\sigma_1\sigma_2}$ with a new energy scale, the so-called pair potential

$$\Delta_{k\sigma_1\sigma_2} = \sum_p \Gamma_{kp}^{(s)} g_{p\sigma_1\sigma_2} \qquad [23]$$

As a consequence of triplet pairing the spin part of the pair potential is "even" upon interchange of $\sigma_1$ and $\sigma_2$: $\Delta_{k\sigma_2\sigma_1} = \Delta_{k\sigma_1\sigma_2}$. Then the Pauli principle requires that $\Delta_{k\sigma_1\sigma_2}$ must be "odd" with respect to the interchange of $k_1$ and $k_2$ or, equivalently, $k \to -k$. The $k$-dependence can now be classified by an orbital quantum number $\ell$ with the special cases of $\ell = 1$ ($p$-wave) pairing, $\ell = 3$ ($f$-wave) pairing, etc. All superfluid phases of $^3$He are characterized by $p$-wave orbital symmetry.

The transition temperature $T_c$ from [23] reads

$$k_B T_c = \frac{2e^\gamma}{\pi} \epsilon_c e^{-1/(N_F \Gamma^{(s)})}$$

with $N_F = 3n/2E_F$ the density of states at the Fermi level and $\gamma = 0.577\ldots$ the Euler constant. The energies $\xi_k$ can trivially be divided into particle-like ($\xi_k > 0$) and hole-like ($\xi_k < 0$) terms. The presence of the pair potential $\Delta_k$ leads to a mixing of particle- and hole-like contributions to the energy, which

becomes a matrix in particle–hole, or Nambu space (Nambu 1960), and generates what is referred to as off-diagonal long-range order (ODLRO):

$$\underline{\xi}_{k} = \begin{pmatrix} \xi_k \mathbf{1} & \Delta_k \\ \Delta_k^\dagger & -\xi_k \mathbf{1} \end{pmatrix} \quad [24]$$

As usual, the diagonalization of $\underline{\xi}_k$ (Bogoliubov 1958) leads to the energy dispersion of the relevant thermal excitations of the superfluid state, the so-called Bogoliubov quasiparticles or "bogolons":

$$E_k = \sqrt{\xi_k^2 + \Delta_k^2}, \quad \Delta_k^2 = \Delta_k \cdot \Delta_k^\dagger \quad [25]$$

In Figure 5, the dispersion $E_p$ of Bogoliubov quasiparticles vs. $|\boldsymbol{p}|$ is shown. It turns out that the superfluid phases (A and B) of liquid $^3$He in zero magnetic field are characterized by unitary matrices $\Delta_k$, so that the scalar quantity $\Delta_k$ can be interpreted as the energy gap in the bogolon spectrum, which, in general, may be anisotropic in $\boldsymbol{k}$-space.

The energy gap $\Delta_k$ of the superfluid B-phase can be represented in the simple nodeless (pseudoisotropic) and BCS-like form (Balian and Werthamer, (BW), 1963):

$$\Delta_k = \Delta(T), \qquad \frac{\Delta(0)}{k_B T_c} = \frac{\pi}{e^\gamma} \quad [26]$$

Its spin structure is characterized by the presence of all three triplet components $m_s = 0, \pm 1$ and will be discussed further with respect to the magnetization response (see next section). The gap symmetry of $^3$He-A is uniaxial with respect to an axis $\hat{\ell}$ (Anderson and Morel 1960; Anderson and Brinkman 1973)

$$\Delta_k = \Delta_0(T)\sin\phi_k, \qquad \frac{\Delta_0(0)}{k_B T_c} = \frac{\pi e^{5/6}}{2e^\gamma} \quad [27]$$

where $\cos\phi_k = \boldsymbol{k}\cdot\hat{\ell}$, and characterized by two point nodes of $\Delta_k$ at the zeros ($\phi_k = 0, \pi$) on the Fermi surface. It has furthermore turned out that only the $m_s = \pm 1$ components of the spin triplet contribute to its spin dependence (equal spin pairing (ESP)).

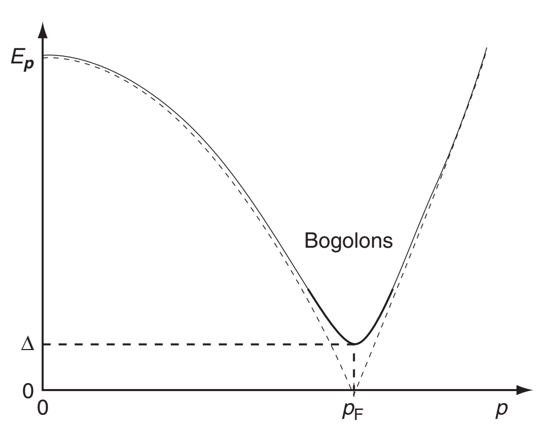



Figure 5 The bogolon energy dispersion.

## Local Response of Condensates and Excitation Gases

In the previous sections we have seen that the structure (energy dispersion, statistics, critical flow velocity) of the relevant thermal excitations is of crucial importance for the superfluidity. We can now aim at a generalized statistical description of bosonic (phonons, rotons) and fermionic (bogolons) excitation gases, by introducing a generalized momentum distribution

$$n_\theta\{E_k\} = \frac{1}{e^{E_k/k_B T} - \theta} \quad [28]$$

and its energy derivative

$$\varphi_{k,\theta} = -\frac{\partial n_\theta\{E_k\}}{\partial E_k} = \frac{1}{2k_B T[\cosh(E_k/k_B T) - \theta]} \quad [29]$$

Special cases are

$$\theta = \begin{cases} 1, & \text{Bose (phonons, rotons)} \\ -1, & \text{Fermi (bogolons)} \end{cases}$$

Introducing the spin $s = (1-\theta)/4$, the total momentum density response to the presence of a superfluid velocity

$$\boldsymbol{v}^s = \frac{\hbar\nabla\varphi}{(2s+1)m}, \quad s = 0, \frac{1}{2}$$

and a normal fluid velocity $\boldsymbol{v}^n$ can be written in the general form

$$\boldsymbol{j}_m = \frac{2s+1}{V}\sum_k \boldsymbol{p} n_\theta\{E_k + \delta E_k\} + \rho\boldsymbol{v}^s \quad [30]$$

After Taylor-expanding $n_\theta$ with respect to the small energy shifts $\delta E_k = \boldsymbol{p}\cdot(\boldsymbol{v}^s - \boldsymbol{v}^n)$, one may introduce the so-called normal fluid density tensor

$$\rho_{ij}^n = \frac{2s+1}{V}\sum_k \varphi_{k,\theta} p_i p_j \quad [31]$$

and the momentum density assumes the form

$$\boldsymbol{j}_m = \rho^s\cdot\boldsymbol{v}^s + \rho^n\cdot\boldsymbol{v}^n, \quad \rho^s = \rho\mathbf{1} - \rho^n \quad [32]$$

Equation [32] forms the central result of this essay because it represents the microscopic counterpart of the generalized London equation [10]. It is clearly seen how the phenomenon of superfluidity originates from $\rho^s > 0$ due to a qualitative change in the dispersion of the elementary excitations, which may in particular be characterized by a gap in

the excitation spectrum. Equation [32] is more general than [10] in that it introduces a two-fluid picture in which the mass supercurrent $\boldsymbol{j}_{\mathrm{m}}^{\mathrm{s}} = \rho^{\mathrm{s}}\boldsymbol{v}^{\mathrm{s}}$ (eqn [10]) is complemented by a normal (excitation) mass current $\boldsymbol{j}_{\mathrm{m}}^{\mathrm{n}} = \rho^{\mathrm{n}}\boldsymbol{v}^{\mathrm{n}}$ in the presence of a macroscopic velocity field $\boldsymbol{v}^{\mathrm{n}}$ of the excitation gas obeying arbitrary statistics. The temperature dependence of $\rho^{\mathrm{s}}(T)$ can now be computed via [31] and the result depends on the dispersion of the thermal excitation under consideration. **Figure 6** shows the temperature dependence of the normal and superfluid density of superfluid $^4$He. The normal fluid density of superfluid $^3$He is, in general, a tensor quantity

$$\rho_{ij}^{\mathrm{n}} = \begin{cases} \rho_{\|}^{\mathrm{n}}\hat{\ell}_i\hat{\ell}_j + \rho_{\perp}^{\mathrm{n}}(\delta_{ij} - \hat{\ell}_i\hat{\ell}_j), & {}^3\text{He-A} \\ \rho^{\mathrm{n}}\delta_{ij}, & {}^3\text{He-B} \end{cases} \qquad [33]$$

The short-range Fermi liquid interaction leads to a quasiparticle mass enhancement $m^*/m = 1 + F_1^{\mathrm{s}}/3$ characterized by the pressure-dependent dimensionless Landau parameter $F_1^{\mathrm{s}}$. In **Figure 7**, the normal fluid density ($\rho_{\|,\perp}^{\mathrm{n}}$ for $^3$He-A, $\rho^{\mathrm{n}}$ for $^3$He-B) is shown as a function of reduced temperature at a pressure of 27 bar, where $F_1^{\mathrm{s}} = 12.53$. The entropy density of an excitation system of arbitrary statistics below the transition can be written as

$$\sigma_0 = k_{\mathrm{B}}\frac{(2s+1)}{V}\sum_{\boldsymbol{k}} P_{\boldsymbol{k}\theta}$$
$$P_{\boldsymbol{k}\theta} = \theta(1+\theta n_\theta)\ln(1+\theta n_\theta) - n_\theta \ln n_\theta \qquad [34]$$

with $n_\theta = n_\theta\{E_{\boldsymbol{k}}\}$, from which one may derive the specific heat capacity

$$T\delta\sigma = \frac{2s+1}{V}\sum_{\boldsymbol{k}} E_{\boldsymbol{k}} n_\theta \left\{ \frac{E_{\boldsymbol{k}} + \frac{\partial E_{\boldsymbol{k}}}{\partial T}\delta T}{k_{\mathrm{B}}(T+\delta T)} \right\}$$
$$= c_V \delta T \qquad [35]$$

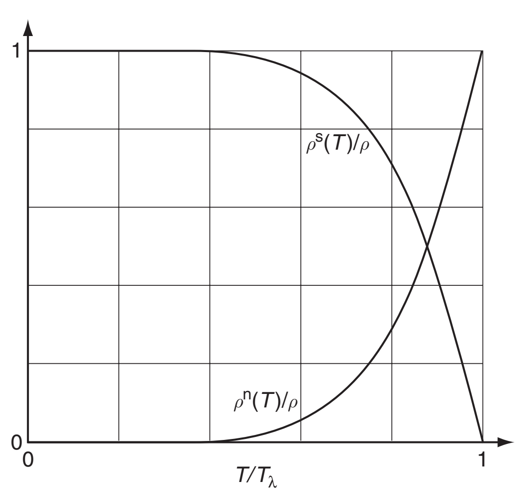



**Figure 6** The normal and superfluid density for He-II.

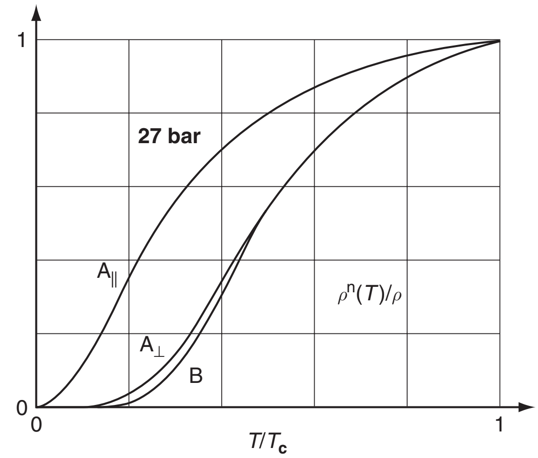



**Figure 7** The normal fluid density for $^3$He-A, B.

After a Taylor expansion of $n_\theta$ with respect to the small local temperature change $\delta T$, the result for $c_V(T)$ reads

$$c_V = \frac{2s+1}{V}\sum_{\boldsymbol{k}} \varphi_{\boldsymbol{k},\theta}\left(\frac{E_{\boldsymbol{k}}^2}{T} - E_{\boldsymbol{k}}\frac{\partial E_{\boldsymbol{k}}}{\partial T}\right) \qquad [36]$$

In **Figures 8 and 9** we show the cusp-like specific heat of a Bose gas as compared with the specific heat of $^3$He-A, B, which display discontinuities at $T_{\mathrm{c}}$.

Finally, the superfluid phases of $^3$He are characterized in addition by the spin degrees of freedom, reflected by the bogolon spin magnetization response to an external magnetic field $B$:

$$M^{\mathrm{n}} = \frac{\gamma\hbar}{2V}\sum_{\boldsymbol{k},\sigma} \sigma n_{-1}\{E_{\boldsymbol{k}} - \gamma\hbar\sigma B/2\} = \chi_0 B \qquad [37]$$

where $\gamma$ denotes the gyromagnetic ratio of the fermions. The bogolon spin susceptibility $\chi_0$ is obtained after a Taylor expansion of $n_{-1}$ with respect to $B$ as

$$\chi_0 = \left(\frac{\gamma\hbar}{2}\right)^2 \frac{1}{V}\sum_{\boldsymbol{k},\sigma}\varphi_{\boldsymbol{k},-1} \equiv \left(\frac{\gamma\hbar}{2}\right)^2 N_{\mathrm{F}} Y(T) \qquad [38]$$

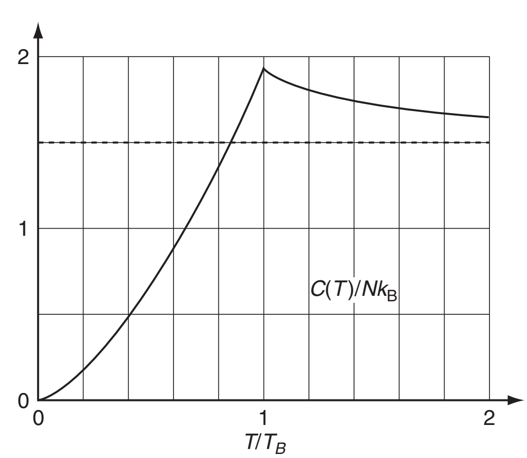



**Figure 8** The specific heat capacity of a Bose gas.

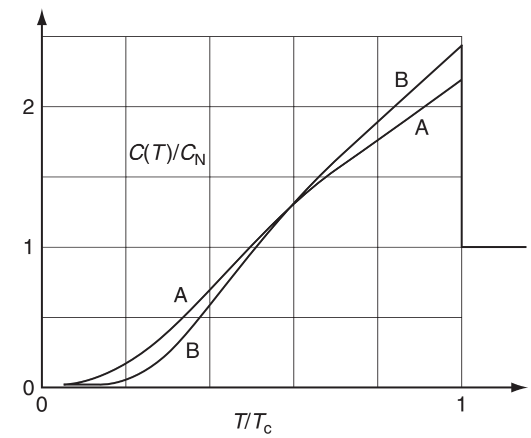



**Figure 9** The specific heat of $^3$He-A, B.

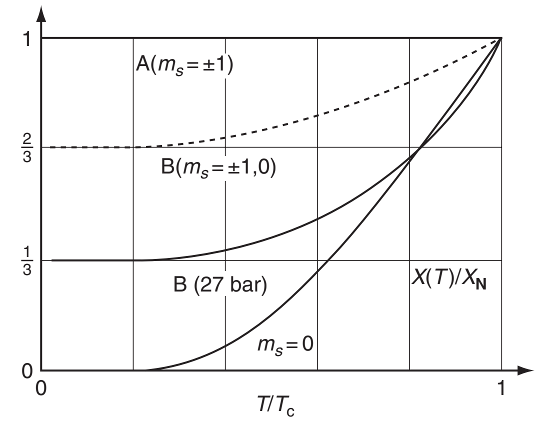



**Figure 10** The spin susceptibility of $^3$He-A, B.

Note that eqn [38] accounts only for the $m_s = 0$ (bogolon) contribution to the spin-triplet susceptibility, the temperature dependence of which is given by the so-called Yosida function $Y(T) = N_F^{-1} \sum_{k\sigma} \varphi_{k,-1}$. The total susceptibility reads

$$\chi_{\text{tot}} = \underbrace{\chi_0}_{\text{bogolons}} + \underbrace{\chi_1 + \chi_{-1}}_{\text{condensate}} \qquad [39]$$

with the condensate contributing through $\chi_{m_s=\pm 1}$ a fraction of 2/3 of the normal state Pauli susceptibility. In Figure 10, the reduced spin susceptibility $\chi/\chi_N$ of $^3$He-A, B is plotted vs. reduced temperature. While the constant susceptibility is characteristic of the ESP pairing state, the reduction of the B-phase susceptibility is due to the lack of the nonmagnetic $m_s = 0$ contribution to the spin triplet in the low-temperature limit. Exchange interaction effects, characterized by the dimensionless Landau parameter $F_0^a$, lead to a further reduction of the Balian-Werthamer (BW)-state susceptibility, which is shown for 27 bar, where $F_0^a = -0.755$. Note that the theoretical picture reflected in Figure 10, and also in Figures 6, 7, and 9, is in quantitative agreement with experimental observations.

In summary, superfluidity is a quantum-mechanical phenomenon seen on a macroscopic scale. It occurs below the degeneracy temperature $T^* \propto n^{2/3}/m$ of both Bose and Fermi many-particle systems (like liquid $^4$He and $^3$He) and is a property of a macroscopic number of particles, the condensate. The role of (weak or strong) interactions is manifested in the structure of the relevant elementary excitations, which always exist in addition to the condensate at finite temperatures and above certain critical velocities. These excitations form a gas, referred to as the normal fluid, since it gives rise to temperature-dependent thermodynamic and response functions and contributes to the entropy and the flow dissipation. Superfluidity is now well understood using various aspects of the concept of the macroscopic wave function. On the microscopic level, the mechanisms of BEC and BCS–Leggett pair formation have been successfully invoked to understand the fascinating properties of Bose and Fermi superfluids.

*See also:* Bose–Einstein Condensates; Bosons and Fermions in External Fields; High $T_c$ Superconductor Theory; Topological Knot Theory and Macroscopic Physics; Variational Techniques for Ginzburg–Landau Energies; Vortex Dynamics.

## Further Reading

Anderson PW and Brinkman WF (1973) *Physical Review Letters* 30: 1911.
Anderson PW and Morel P (1960) *Physical Review Letters* 5: 136.
Annett J (2004) *Superconductivity, Superfluids and Condensates*. Oxford: Oxford University Press.
Balian R and Werthamer NR (1963) *Physical Review* 131: 1553.
Bogoliubov NN (1947) *Journal of Physics USSR* 11: 23.
Bogoliubov NN (1958) *Soviet Physics JETP* 7: 794.
Dobbs ER (2000) *Helium Three*. Oxford: Oxford University Press.
Feynman RP (1953) *Physical Review* 91: 1301.
Keller WE (1969) *Helium-3 and Helium-4*. New York: Plenum.
Landau LD (1947) *Journal of Physics USSR* 11: 91.
London F (1938) *Physical Review* 54: 947.
Madelung E (1926) *Z. Physik* 40: 322.
Nambu Y (1960) *Physical Review* 117: 648.
Tilley DR and Tilley J (1990) *Superfluidity and Superconductivity*. Adam Hilger.
Tisza L (1938) *Journal de Physique et Radium* 1: 164.
Tsuneto T (1998) *Superconductivity and Superfluidity*. Cambridge: Cambridge University Press.
Vollhardt D and Wölfle P (1990) *The Superfluid Phases of Helium 3*. Taylor and Francis.

# Supergravity

**K S Stelle**, Imperial College, London, UK



## Introduction: Minimal $D=4$ Supergravity

The essential idea of supersymmetry is an extension of the relativistic structure group of spacetime, which in ordinary four-dimensional physics in the absence of gravity is the Poincaré group ISO(3, 1). In a minimal supersymmetric theory in flat $D=4$ spacetime, the minimal supersymmetry algebra (the "graded Poincaré algebra") adds spinorial generators $Q_\alpha$ to the Lorentz generators $M_{mn}$ and the translational generators (momenta) $P_m$, where $m=0,1,2,3$. The core relation is the "anticommutator" of two $Q_\alpha$:

$$\{Q_\alpha, \bar{Q}_\beta\} = -2\gamma^m_{\alpha\beta}P_m \qquad [1]$$

where $\bar{Q}=Q^\dagger\gamma^0$ and the $\gamma^m$ are the Dirac gamma matrices. In the minimal $D=4$ supersymmetry algebra, the spinor generator $Q_\alpha$ is taken to be Majorana: $Q=C(\bar{Q})^{\mathrm{T}}$, where $C$ is the charge-conjugation matrix and $A^{\mathrm{T}}$ denotes the transpose of the matrix $A$. The full supersymmetry algebra adjoins to the anticommutation relation [1] the usual commutation relations among the Lorentz generators and the commutators of the Lorentz generators with the momenta and the spinors $Q_\alpha$; the latter express respectively the vectorial and spinorial characters of $P_m$ and $Q_\alpha$:

$$\mathrm{i}[M_{mn}, M_{pq}] = \eta_{np}M_{mq} - \eta_{mp}M_{nq} \qquad [2]$$

$$\mathrm{i}[M_{mn}, P_q] = \eta_{nq}P_m - \eta_{mq}P_n \qquad [3]$$

$$\mathrm{i}[M_{mn}, Q_\alpha] = \tfrac{1}{2}(\gamma_{mn}Q)_\alpha \qquad [4]$$

where $\gamma_{mn}=(1/2)(\gamma_m\gamma_n-\gamma_n\gamma_m)$ and $\eta_{mn}=\mathrm{diag}(-1,1,1,1)$ is the Minkowski metric. The final relation in the supersymmetry algebra expresses the flatness of Minkowski space:

$$[P_m, P_n] = 0 \qquad [5]$$

This algebra has been considered as an extension of the symmetry algebra of particle physics since the work of Gol'fand and Likhtman in 1971, and especially since the linearly realized supersymmetric model of Wess and Zumino in 1974. That model contains a pair of $D=4$ scalar fields and a $D=4$ Majorana spinor, so the numbers of bosonic and fermionic degrees of freedom are equal; this is a fundamental characteristic of supersymmetric theories.

The work of Wess and Zumino led to an explosion of interest in supersymmetry, especially once it was realized that renormalizable supersymmetric models display a cancellation of some of the divergences that have plagued relativistic quantum field theory since its inception in the 1930s. In particular, in renormalizable flat-space field theory models, divergences quadratic in a high-momentum cutoff vanish as a result of cancellations between virtual bosonic and fermionic particles. This is a very attractive feature for control of the "hierarchy problem" in particle physics, especially for the instability inherent in having vastly different scales within the same theory, for example, the TeV scale of ordinary electroweak physics and the $10^{16}$ GeV scale where unification with the strong interactions might come in.

When one includes gravity, the stability problems of particle physics become much more severe. Einstein's theory of general relativity is itself non-renormalizable, that is, its ultraviolet divergences are of different forms from the terms present in the original "classical" action and there is no acceptable finite set of correction terms that can be added to it to remove this defect. Moreover, when otherwise tolerably behaved matter field theories that are renormalizable in a flat-spacetime context are coupled to general relativity, the gravitational couplings pollute the matter theories with non-renormalizable divergences. This is a key aspect of the great difficulty that has been encountered in interpreting gravity as a quantum theory.

Supersymmetry, with its divergence-canceling powers, was thus a very attractive option in the struggle to formulate a quantum theory of gravity, and the creation of a supergravity theory was thus a very high priority task. This was achieved in 1976 by Freedman, Ferrara, and Van Nieuwenhuizen using the technique of iterative Noether coupling to build up this nonlinear theory order-by-order in powers of the fermionic fields. The fermionic partner of the massless spin-2 "graviton" field is a massless fermionic spin-3/2 field that has come to be called the "gravitino."

A second 1976 paper by Deser and Zumino soon followed, emphasizing how supergravity manages to circumvent the well-known problems of coupling spins higher than 1 to gravity. A key point in achieving this result is the role played by the local version of the supersymmetry algebra [1]–[5]. As one can see from the translations occurring on the right-hand side of [1], when one replaces translation symmetry by local general coordinate invariance in a gravitational context, the supersymmetry transformations must themselves become local as well. Local symmetries allow for transformation parameters

that are local in the spacetime coordinates $x^m$, and in interacting theories they require coupling of the corresponding “gauge field” to a conserved current. In the case of supergravity, the gravitino field $\psi_{m\alpha}$ plays this gauge-field role, and its coupling to the conserved current of supersymmetry is the key to allowing a consistent coupling between the spin-2 graviton and the spin-3/2 gravitino.

## The Minimal Supergravity Action

The action for minimal supergravity in $D=4$ dimensions can be written, using the vierbein formalism where the metric is expressed as a quadratic expression in a nonsymmetric $4\times 4$ vierbein matrix $e_m^a, g_{mn}=e_m^a e_n^b \eta_{ab}$, as

$$I=\frac{1}{2\kappa^2}\int \mathrm{d}^4x\,\det(e)R(e,\omega(e)+K(\psi)) - \frac{\mathrm{i}}{2}\int \mathrm{d}^4x\epsilon^{mnpq}\bar{\psi}_m\gamma_5\gamma_n D_p(e,\omega(e)+K(\psi))\psi_q \quad [6]$$

where $\kappa=\sqrt{8\pi G}$ is the gravitational coupling constant,

$$\omega_m^{ab}(e)=\tfrac{1}{2}e^{na}\left(e^b_{n,m}-e^b_{m,n}\right)-\tfrac{1}{2}e^{nb}\left(e^a_{n,m}-e^a_{m,n}\right) + \tfrac{1}{2}e^{na}e^{rb}(e_{nc,r}-e_{rc,n})e^c_m \quad [7]$$

is the usual vierbein formalism spin connection (in which $e^b_{n,m}=\partial_m e^b_n$ and $e^{ma}$ is the matrix inverse of $e_{ma}$), and

$$K_m^{ab}(\psi)=\frac{\mathrm{i}\kappa^2}{4}(\bar{\psi}_m\gamma^a\psi^b+\bar{\psi}^a\gamma_m\psi^b-\bar{\psi}_m\gamma^b\psi^a) \quad [8]$$

is the fermionic contorsion, an additional part of the covariant derivative $D_m(e+K(\psi))$ appearing in the action [6]. (Indices $m,n$ are taken to be “world” indices while indices $a,b$ are “tangent space” indices; one can convert from one type to another using the vierbein $e_m^a$ and its inverse, e.g., $\psi_{a\alpha}= e_a^m\psi_{m\alpha}$.)

Keeping the terms in the action grouped as above using the nonstandard covariant derivative $e_m^{ab}+K_m^{ab}$ is what has been called “1.5 order formalism”: this greatly simplifies the writing and analysis of the supergravity action [6]. In the action [6], one has the Ricci scalar $R(e,\omega(e)+K(\psi))$ written in terms of this generalized torsional spin connection. One may of course expand out all the $\omega_m^{ab}+K_m^{ab}$ combinations and write the nonlinear fermionic terms separately. Doing this produces a quartic term

$$L_4=\frac{\kappa^2}{32}[\bar{\psi}^b\gamma^a\psi^c(\bar{\psi}_b\gamma_a\psi_c+2\bar{\psi}_a\gamma_b\psi_c) - 4(\bar{\psi}^a\gamma^b\psi^c)(\bar{\psi}_a\gamma_b\psi_c)]$$

showing the highly nonlinear nature of supergravity theory – when expanded out, the theory becomes much more cumbersome to study. The 1.5 order formalism trick is one of a large number of algebraic simplifications that had to be developed in order to master the technical aspects of supergravity. It also reveals a characteristic physical feature: this theory naturally involves a connection with torsion built from the fermionic fields.

In terms of the torsional covariant derivative $D_m\epsilon(x)=(\partial_m+(1/4)(\omega_m^{ab}(e)+K_m^{ab}(\psi))\gamma_{ab})\epsilon(x)$ of the infinitesimal supersymmetry parameter $\epsilon(x)$, the local supersymmetry transformations which leave the action [6] invariant (up to the integral of a total derivative) are

$$\delta e_m^a=\mathrm{i}\bar{\epsilon}\gamma_a\psi_m \quad [9]$$

$$\delta\psi_m=2\kappa^{-1}D_m\epsilon \quad [10]$$

The inhomogeneous part $2\kappa^{-1}\partial_m\epsilon$ in the gravitino transformation [10] demonstrates the gauge-field nature of the gravitino field. For a distribution of “supermatter” fields (e.g., Wess–Zumino model scalars and spinors), the integrated “charge” that one would get from a Gauss’s law surface integral at spatial infinity using the gravitino gauge field is the total supercharge $Q_\alpha$, which in turn plays the role of the supersymmetry generator in the original matter-sector supersymmetry algebra [1].

Both the gravitational field and the gravitino field are thus effectively gauge fields, albeit not of a standard Yang–Mills type. The local algebra is a deformation of the rigid supersymmetry algebra [1]–[5], generalizing the relation between general covariance and flat-space Poincaré symmetry. Some basic consequences of the flat-space algebra are preserved, however. An extremely important instance of this is energy positivity. As one can see by multiplying [1] by $\gamma^0$ and then contracting on the spinor index,

$$E=P^0=\frac{1}{2}\sum_\alpha\{Q_\alpha,Q_\alpha^\dagger\}$$

The right-hand side is manifestly non-negative provided the theory is quantized in a positive-metric Hilbert space. One can see this even more explicitly in a Majorana spinor basis, where $Q_\alpha^\dagger=Q_\alpha$. Accordingly, for flat-space supersymmetric theories, one obtains directly the result that energy is non-negative. This carries over to the local algebra of supergravity, where the total energy is obtained from a Gauss’s law integral over the sphere at spatial infinity.

In general relativity, an integrated energy can be defined with respect to an asymptotic timelike Killing vector at spatial infinity. Showing that this

energy is non-negative remained for decades a famously unsolved problem in gravitational physics; it was ultimately proven in Yau's positive-energy theorem. The algebraic structure of supergravity makes energy positivity much more transparent, however. Since pure general relativity can be obtained by setting the gravitino field to zero, this result is inherited by pure Einstein theory as a consequence of its being embeddable into supergravity. Energy positivity can thus be proved even at the classical level using ideas taken from supergravity, as was done by Witten and later streamlined by Nester, in an argument much simpler than Yau's proof. This argument writes the energy as an integral over a positive-semidefinite expression quadratic in a commuting spinor field which is analogous to the (anticommuting) spinor parameter of supergravity in the transformations [9] and [10].

## Auxiliary Fields and Superspace

Supergravity shares with flat-space supersymmetric theories a curious technical feature that gives a hint of a new underlying geometry. Standard counting of the gauge-invariant continuous degrees freedom of the graviton and the gravitino in momentum space yield the same result per momentum value: two bosonic degrees of freedom and two fermionic degrees of freedom. This accords with the general requirement in supersymmetric theories that the numbers of bosonic and fermionic degrees of freedom match. This count follows from the Einstein and spin-3/2 equations of motion, or "on-shell." If one compares the count of nongauge degrees of freedom without using the equations of motion (i.e., "off-shell"), one obtains an imbalance, however: six nongauge graviton versus 12 nongauge fermion fields. This is directly related to another puzzling feature of the supergravity realization of local supersymmetry: the local supersymmetry algebra closes onto a finite set of transformations only when the equations of motion are imposed.

As in flat-space supersymmetry, the cure for this problem is to add nondynamical "auxiliary" fields to the action. In the supergravity case, the imbalance in the off-shell bose–fermi field count indicates that an additional six bosonic fields are needed. In the minimal set of auxiliary fields, these organize into a vector $b_m$ and a scalar-pseudoscalar pair $M$, $N$; the additional terms in the action [6] are simply

$$\int d^4x \det(e)\left(-\tfrac{1}{3}M^2 - \tfrac{1}{3}N^2 + \tfrac{1}{3}b_m b^m\right)$$

while the local supersymmetry transformations are changed to include the auxiliary fields, e.g., the gravitino transformation becomes

$$\delta\psi_m = 2\kappa^{-1}D_m(\omega, K)\epsilon + \gamma_5\left(b_m - \tfrac{1}{3}\gamma_m\gamma^n b_n\right)\epsilon - \tfrac{1}{3}\gamma_m(M + \gamma_5 N)\epsilon)$$

while the auxiliary fields transform into expressions that vanish on-shell. Since the field equations for the auxiliary fields are algebraic in character and since for source-free supergravity they have the simple solution $b_m = M = N = 0$, one can directly regain the on-shell formalism by algebraically eliminating the auxiliary fields.

The inclusion of auxiliary fields is not an empty trick, however. The local supersymmetry transformations including the auxiliary fields form a closed set without the use of equations of motion ("off-shell closure"). This standardizes the form of the supersymmetry transformations so that they remain the same even when supermatter is coupled to supergravity instead of needing a case-by-case Noether construction as in the case without the auxiliary fields. In this way, a standard set of coupling rules can be drawn up, known as the "tensor calculus." This tensor calculus is of great importance as it allows for the construction of general models of supergravity coupled to supermatter (Wess–Zumino multiplets and super Yang–Mills multiplets consisting of spin-1 gauge fields and spin-1/2 "gaugino" fields). These general couplings form the basis for essentially all supersymmetric phenomenology, and in particular for the formulation of the Minimal Supersymmetric Standard Model. Since supersymmetry is not directly observed in low-energy physics, it must be spontaneously broken, like many other gauge symmetries. As it happens, the physically realistic mechanisms of supersymmetry breaking all originate from supergravity couplings derived using the tensor calculus.

Given the regular set of tensor calculus rules for coupling supergravity to supermatter, one is led to suspect that a geometrical structure lies in the background. This is indeed the case; the corresponding construction is known as "superspace."

The basic idea of superspace is a generalization of the coset space construction of Minkowski space as the coset space given by the Poincaré group divided by the Lorentz group: $\mathcal{M}_4(x^m) = \mathrm{ISO}(3,1)/\mathrm{SO}(3,1)$. For supersymmetric theories, one analogously constructs $\mathrm{Superspace}(x^m, \theta^\alpha) = \mathrm{Graded\ Poincaré}/\mathrm{SO}(3,1)$. The basic ideas of superspace were introduced by Akulov and Volkov in 1972, while the idea of expanding in "functions" on this space, thus yielding "superfield," was introduced by Salam and Strathdee

in 1974. This led to a formulation of the Wess–Zumino model in terms of a chiral superfield $\phi(x, \theta)$, which is subjected to a covariant superspace constraint.

In order to manage the formalism of superspace more efficiently, it is convenient to use a two-component spinor formalism corresponding to the Weyl basis for the Dirac gamma matrices, in which the Majorana spinor coordinate $\theta$ is represented as

$$\theta = \begin{pmatrix} \theta_\alpha \\ \bar{\theta}^{\dot{\alpha}} \end{pmatrix}$$

where two-component indices $\alpha, \dot{\alpha} = 1, 2$ are raised and lowered with the covariant two-index antisymmetric tensors $\epsilon^{\alpha\beta}, \epsilon^{\dot{\alpha}\dot{\beta}}$, which both take the numerical value $\mathrm{i}\sigma_2$. The flat-space fermionic covariant derivatives are then

$$\begin{aligned} D_\alpha &= \frac{\partial}{\partial\theta^\alpha} + \mathrm{i}\sigma^m_{\alpha\dot{\beta}}\bar{\theta}^{\dot{\beta}}\partial_m \\ \bar{D}_{\dot{\alpha}} &= -\frac{\partial}{\partial\bar{\theta}^{\dot{\alpha}}} + \mathrm{i}\bar{\theta}^{\beta}\sigma^m_{\beta\dot{\alpha}}\partial_m \end{aligned} \qquad [11]$$

where the $\sigma^m_{\alpha\dot{\beta}} = (\mathbb{1}, \sigma_i)$ for $m = (0, i)$ (where $\sigma_i$ are the Pauli matrices) are the Van der Waerden matrices which establish the mapping between vector indices and (chiral, antichiral) spinor index pairs. The Wess–Zumino multiplet is then described by a complex chiral superfield satisfying the constraint $\bar{D}_{\dot{\alpha}}\phi = 0$. Unlike the situation in Minkowski space, where the only Lorentz-covariant solution to a constraint that sets to zero the $\partial/\partial x^m$ derivatives is a constant, superspace has a reducible set of coordinates $(x^m, \theta^\alpha, \bar{\theta}^{\dot{\alpha}})$ and, as a result, requiring $\phi$ to be annihilated by $\bar{D}_{\dot{\alpha}}$ does not require the whole superfield to be a constant.

Since the fermionic coordinates of superspace $\theta^\alpha, \bar{\theta}^{\dot{\alpha}}$ are anticommuting (i.e., they are elements of a Grassman algebra), and since $\alpha, \dot{\alpha} = 1, 2$ have an index range of two, powers of them higher than the second order necessarily vanish. As a result, superfields like $\phi$ can be expanded into sets of component fields, each of which is an ordinary field in Minkowski space. In this way, a chiral superfield expands into $(A(x), B(x), \chi_\alpha(x), \bar{\chi}_{\dot{\alpha}}(x), F(x), G(x))$, where the fields $A$, $B$, $\chi$, and $\bar{\chi}$ are the physical fields of the Wess–Zumino model, while $F$ and $G$ are dimension-2 auxiliary fields. In this way, the auxiliary fields of supersymmetry naturally fit into a superspace formalism as higher components in a superfield expansion. It is in this sense that they point toward the superspace formulations of supersymmetric theories.

For supergravity, there are a number of different approaches to realizing the theory in superspace, and these correspond naturally to the various possible choices of auxiliary-field sets. With the minimal set, the supergravity multiplet is described by a superfield carrying a vector index $H_m(x, \theta, \bar{\theta})$; this superfield is called the prepotential of supergravity. Note the fact that since the divisor group in the coset-space construction of superspace is the Lorentz group, superfields may carry indices corresponding to any Lorentz representation. The component-field expansion of the $H_m$ superfield yields the physical $e^a_m, \psi_{m\alpha}, \bar{\psi}_{m\dot{\alpha}}$ and auxiliary fields $(b_m, M, N)$ together with a number of other components of dimension lower than those of the physical fields. This is not, however, all that surprising: even the physical fields $e^a_m, \psi_{m\alpha}, \bar{\psi}_{m\dot{\alpha}}$ contain components that are not directly related to the physical modes because we are dealing with a gauge theory. What occurs in superspace is a redundant expression of the supergravity multiplet with the presence of various component gauge fields.

The full expression of local supersymmetry in superspace can be given in a number of different formalisms. Suffice it here to indicate the transformation of the linearized theory expanded in small fluctuations about empty flat superspace. Converting the vector index of $H_m$ into a (chiral, antichiral) spinor index pair via $H_{\alpha,\dot{\beta}} = \sigma^m_{\alpha\dot{\beta}}H_m$, the linearized local symmetry transformation of the supergravity multiplet is

$$\delta H_{\alpha\dot{\beta}} = D_\alpha\bar{L}_{\dot{\beta}} - \bar{D}_{\dot{\beta}}L_\alpha \qquad [12]$$

where the transformation parameter superfield $L_\alpha$ carrying a spinor index is antichiral: $D_\alpha\bar{L}_{\dot{\beta}} = 0$ (while the conjugate parameter superfield $\bar{L}_\alpha$ is chiral). Expanding in component fields and comparing with the expansion of $H_m$, one sees that the chiral spinor superfield contains precisely the components needed to provide the standard gauge symmetries of $e^a_m$ and $\psi_{m\alpha}, \psi_{m\dot{\alpha}}$ and also to transform the other gauge components of $H_m$ as well. One can then make various gauge choices according to taste in a given context.

One frequently encountered superspace gauge choice sets to zero all the fields in $H_m$ except for the physical and auxiliary fields $(e^a_m, \psi_{m\alpha}, \bar{\psi}_{m\dot{\alpha}}, b_m, M, N)$. This is called a Wess–Zumino gauge following the analogy to a similar construction for super Maxwell theory (containing spins 1 and 1/2). Wess–Zumino gauge choices are not, however, supersymmetrically covariant. This shows up when one works out the supersymmetry algebra in such a gauge: the presence of auxiliary fields gives closure, as required, without use of the equations of motion, but the anticommutator of two supersymmetry

transformations when acting on a gauge field such as the Maxwell field or the vierbein gives a combination of the anticipated translation with an admixture of a gauge transformation with a field-dependent parameter.

The prepotential superfield of minimal supergravity can itself be fit into larger formalisms in superspace that are analogous to standard differential geometry, with supervielbeins, superspin connections and so forth. An unavoidable feature of these more seemingly geometric constructions, however, is their high degree of redundancy: superspace vielbeins and spin connections carrying Lorentz indices have many component fields in addition to those found in the prepotential. This redundancy is then cut down in turn by imposing superspace constraints on the geometrical superfields, for example, on the components of the torsion tensor in superspace.

## Extended Supergravities and Supergravities in Higher Dimensions

The possible graded extensions of the Poincaré algebra allow for more than one spinorial generator. Thus, one can have $N$ supersymmetry generators $Q^i_\alpha, \bar{Q}_{\dot\beta j}, i, j = 1, \ldots N$, with basic anticommutators (in Lorentz two-component notation)

$$\{Q^i_\alpha, \bar{Q}_{\dot\beta i}\} = 2\delta^i_j \sigma^m_{\alpha\dot\beta} P_m \qquad [13]$$

$$\{Q^i_\alpha, Q^j_\beta\} = 2\epsilon_{\alpha\beta} a^{\ell ij} Z_\ell \qquad [14]$$

$$\{\bar{Q}_{\dot\alpha i}, \bar{Q}_{\dot\beta j}\} = 2\epsilon_{\dot\alpha\dot\beta} \bar{a}^\ell_{ij} Z_\ell \qquad [15]$$

The right-hand sides of [14] and [15] allow for the possibility of nonvanishing commutators between supersymmetry generators of the same chirality. As one can see from the overall symmetry in pairs of indices $(\alpha i, \beta j)$, the coefficients $a^{\ell ij}$ must be antisymmetric in the $i, j$ indices, so such nonvanishing same-chirality anticommutators cannot occur for $N = 1$. The corresponding abelian generators $Z_\ell$ are called central charges since they must commute with all the other $(Q^i_\alpha, \bar{Q}_{\dot\beta],j}, P_m)$ elements of the algebra.

The $i, j$ indices may be endowed with a symmetry meaning as well, although this is not obligatory in every model. When the central charges are absent, $Z_\ell = 0$, one has U($N$) (or SU($N$)) as the maximal such external automorphism; the choice of index placement on $Q^i_\alpha$ and $\bar{Q}_{\dot\beta j}$ anticipates this. If such a symmetry is realized in a given model, the fact that the $Q^i_\alpha, \bar{Q}_{\dot\beta j}$ carry representations both for that symmetry and for the spacetime Poincaré symmetry demonstrates how supersymmetry evades the no-go theorem barring unified spacetime and internal symmetries. This theorem (the Coleman–Mandula theorem) can be evaded, since at the time it was written, graded Lie symmetry algebras were not yet considered. For nonzero central charges, the external automorphism algebra becomes a subalgebra of U($N$) determined by the requirement that invariant antisymmetric tensors $a^{\ell ij}$ exist.

The representations of the algebra [13]–[14] span an increasing range of spins as the number $N$ of $D = 4$ supersymmetries increases. For massive representations without central charges, the spins of the smallest supersymmetry representation extend from states of spin 0 (scalars) up to spin $N/2$; with central charges, the spin range can be shortened down to a minimum range of $N/4$. For massless representations, the range of helicities in a PCT (parity–change–time reversal) symmetric multiplet is from $-N/4$ to $N/4$. This spin range has an important implication for the maximal extension of supersymmetry that can be realized in an interacting supersymmetric field theory, because no interacting theories with a finite set of spins exist for spins $>2$. Accordingly, the maximal extension of supersymmetry is $N = 8$ for massless theories, and in order to have massive states with spins that do not exceed spin 2 in an $N = 8$ theory, the central charges have to be active for maximal multiplet shortening.

The $N = 8$ supergravity theory, found by Cremmer and Julia in 1978, is thus the largest possible supergravity in $D = 4$ dimensions. It contains the following "spin" range (allowing for a certain imprecision of expression: for massless fields one should really speak only of helicities)

$N = 8$ supergravity spins

| Spin | 2 | $\frac{3}{2}$ | 1 | $\frac{1}{2}$ | 0 |
|---|---|---|---|---|---|
| Multiplicity | 1 | 8 | 28 | 56 | 70 |

In order to realize the automorphism SU(8) symmetry, one has to consider the field strengths for the 28 spin-1 fields, separated into complex self-dual and anti-self-dual parts in their antisymmetric Lorentz indices. These complex field strengths can then be endowed with a complex 28-dimensional representation of SU(8). The 70 scalars, on the other hand, fit precisely into the four-index antisymmetric self-dual representation of SU(8), $\phi^{i_1 i_2 i_3 i_4} = 1/(4!)\epsilon^{i_1 i_2 i_3 i_4 j_1 j_2 j_3 j_4} \bar{\phi}_{j_1 j_2 j_3 j_4}$. It is the use of the eight-index epsilon tensor here that restricts the automorphism group to SU(8) instead of U(8).

The SU(8) automorphism symmetry of $N = 8$ supergravity theory is linearly realized. It plays an important role in another symmetry of this theory which is highly nonlinear. This theory has a

remarkable nonlinear $E_7$ symmetry. In fact, the 70 scalars form a nonlinear sigma model with the fields taking their values in the coset space $E_7/\mathrm{SU}(8)$ (of dimension $133 - 63 = 70$), where the SU(8) divisor is the linearly realized automorphism group discussed above.

The extended supergravities point to another aspect of supergravity theory: the existence of higher-dimensional supergravities, from which the extended theories in $D = 4$ spacetime can be derived by Kaluza–Klein dimensional reduction. If one considers a $D'$ dimensional massless theory in a spacetime where $d$ dimensions form a compact $d$-torus, then the theory can be viewed as a $D = D' - d$ dimensional theory in which the discrete Fourier modes arising from the periodicity requirements on the $d$-torus give rise to towers of equally spaced massive Kaluza–Klein states, plus a massless sector in $D' - d$ dimensions corresponding to the modes with no dependence on the $d$-torus coordinates.

Importantly, $N = 8$ supergravity in four-dimensional spacetime can be obtained in this way from a supergravity theory that exists in 11 spacetime dimensions. Upon dimensional reduction on a 7-torus to four dimensions, one obtains $N = 8, D = 4$ supergravity at the massless level, plus an infinite tower of massive $N = 8$ supermultiplets with central charges so that their spin range extends only up to spin 2. This $D = 11$ supergravity was in fact found before the $N = 8$ theory by Cremmer, Julia, and Scherk, with the details of the more complicated $N = 8, D = 4$ theory being worked out via the techniques of Kaluza–Klein dimensional reduction. The fields of the $D = 11$ theory include an exotic field type not encountered in $D = 4$ theories: the bosonic fields of the theory comprise the graviton $e_M^A$ plus a three-index antisymmetric tensor gauge field $C_{MNP}$. Counting the number of propagating modes of these fields for a given momentum value gives $44 + 84 = 128$ bosonic degrees of freedom. This precisely balances the 128 fermionic degrees of freedom coming from the $D = 11$ gravitino $\psi_{M\alpha}$.

## Supergravity Effective Theories, Strings and Branes

The hope for a cancellation of the ultraviolet divergences in a supersymmetric theory of gravity turned out to be ephemeral, although there is in fact a postponement of the divergence onset until a higher order in quantum field loops. There is agreement that the nonmaximal supergravities diverge at the three-loop order. For the $N = 8, D = 4$ theory, the situation remains unclear, but divergences are nonetheless expected to occur at some finite loop order.

This persistence of nonrenormalizability in $D = 4$ supergravity theories is no longer seen as a disaster, however, because these theories are now seen as effective theories for the massless modes arising from a deeper microscopic quantum theory. In addition, the theories that are most directly connected to this underlying quantum theory are, surprisingly, the maximal supergravities in spacetime dimensions 10 and 11. $D = 11$ supergravity can be dimensionally reduced on a 1-torus (i.e., a circle) to $D = 10$ where the massless sector yields type IIA supergravity theory. This theory is the effective theory for a consistent quantum theory of type IIA superstrings in $D = 10$. Theories of relativistic strings (i.e., one-dimensional extended objects) have strikingly different properties from theories of point particles. In particular, the spread-out nature of the interactions leads to a damping out of the quantum field theory divergences, while the underlying supersymmetry causes a cancellation of other infinities that could have arisen owing to the two-dimensional nature of the string world sheets. This gives, for the first time, a perturbatively well-defined quantum theory including gravity.

In addition to the type IIA theory, there are four other consistent superstring theories in $D = 10$, and these are in turn related to various $D = 10$ supergravity effective theories for the massless modes: type IIB, $E_8 \times E_8$ heterotic, SO(32) heterotic, and SO(32) type I. Remarkably, the maximal $D = 11$ supergravity enters into this picture as well, as a consequence of a pattern of duality symmetries that have been found among the superstring theories.

The dualities of string theory are directly related to the nonlinear symmetries of the dimensionally reduced supergravities in $D = 4$. The string quantum corrections do not respect the $E_7$ symmetry of the classical $N = 8$ theory, but they do respect a discrete subgroup of this symmetry in which the $E_7$ group elements are required to take integer values: $E_7(\mathbb{Z})$.

This quantum-level restriction to a discrete subgroup can be seen from another phenomenon characteristic of superstring theories: the existence of "electric" and "magnetic" brane solutions. The antisymmetric-tensor (or "form") fields of the higher-dimensional supergravities naturally give rise to solitonic solutions in which $p + 1$ dimensions form a flat Poincaré invariant subspace. This can be interpreted as the world volume of an infinite $p$-brane extended object. In the $D = 11$ supergravity theory, the branes that emerge in this way are a 2-brane and a 5-brane. The three-dimensional world volume of the 2-brane naturally couples to the

3-form field $C_{MNP}$, just as an ordinary Maxwell vector field couples to the one-dimensional world line of a point particle (or 0-brane). The 2-brane is thus naturally electrically charged with respect to the 3-form field; its charge can be obtained, in a direct generalization of the Maxwell case, from a Gauss' law integral of the field strength $H_{[4]} = dC_{[3]}$ over a 7-sphere at spatial infinity in the eight directions transverse to the brane worldvolume. The 5-brane, on the other hand, has a magnetic type charge; it is the 7-form dual to $H_{[4]}$ that is integrated to give its charge. In addition to these static infinite $p$-branes, the theory contains dynamical finite-extent branes as well, although for these one generally does not have explicit solutions.

As one reduces a higher-dimensional supergravity to lower and lower dimensions, there is a proliferation of solitonic brane solutions of varying dimensionality, and of both electric and magnetic charge types. In a quantum theory context, these electrically and magnetically charged branes pair up in ways that must satisfy a generalization of the Dirac quantization condition for $D = 4$ electric and magnetic point particles. This ends up requiring all the supergravity solitonic brane charges to lie on a charge lattice. It is the requirement that this discrete brane-charge lattice be respected that restricts the classical supergravity nonlinear symmetry groups to discrete duality subgroups.

The dualities relate brane solutions within a given theory and also between different string theories. They include transformations that invert the radii of compactifying tori, giving a large–small compactification scale duality. They also include transformations that invert the string coupling constant, thus interchanging strong and weak coupling. The type IIB theory, for example, is self-dual under strong–weak coupling duality. In the case of the type IIA theory, however, something remarkable happens. The strong coupling limit of this theory turns out to be related by duality, not to another string theory, but to the maximal $D = 11$ supergravity. The role of the Kaluza–Klein massive modes for the 11 to 10 reduction is played by an infinite tower of extremal charged black holes.

Thus, even $D = 11$ supergravity theory has a role to play in the effective theory of the underlying quantum dynamics. This underlying theory has been dubbed "M-theory." It is still only partially understood, but many of its most important properties are presaged by the remarkable nonlinear structure of the classical supergravities.

*See also:* Brane Construction of Gauge Theories; Brane Worlds; Branes and Black Hole Statistical Mechanics; Random Algebraic Geometry, Attractors and Flux Vacua; Renormalization: General Theory; Spinors and Spin Coefficients; Stability of Minkowski Space; Supermanifolds; Superstring Theories; Supersymmetric Particle Models; Symmetries and Conservation Laws; Symmetries in Quantum Field Theory: Algebraic Aspects.

## Further Reading

Buchbinder JL and Kuzenko SM (1998) *Ideas and Methods of Supersymmetry and Supergravity*. Bristol: IoP Publishing Ltd.
Stelle KS (1998) BPS branes in supergravity, Trieste 1987 School of High-Energy Physics and Cosmology, arXiv:hep-th/9803116.
Van Nieuwenhuizen P (1981) Supergravity. *Physics Reports* 68: 189–398.
Wess J and Bagger J (1983) *Supersymmetry and Supergravity*. Princeton: Princeton University Press.

# Supermanifolds

**F A Rogers**, King's College London, London, UK



## Introduction

A supermanifold is a generalization of a classical manifold to include coordinates that are in some sense anticommuting. Much of the motivation for the study of supermanifolds comes from supersymmetric physics, where it is useful to have a formalism which treats fermions and bosons in the same way. The underlying reason for the effectiveness of supermanifolds is that anticommuting coordinates allow the fermionic canonical anticommutation relations to be handled in a way analogous to the bosonic canonical commutation relations. Supersymmetric methods have proved immensely effective in fundamental physics; they also play a considerable role in geometrical index theory in mathematics. In this article we describe supermanifolds from two points of view – geometric and algebraic – and consider some of the standard features of manifold calculus, including integration since this is an area where the distinctive features of this generalized geometry are particularly apparent.

Thus, even $D = 11$ supergravity theory has a role to play in the effective theory of the underlying quantum dynamics. This underlying theory has been dubbed "M-theory." It is still only partially understood, but many of its most important properties are presaged by the remarkable nonlinear structure of the classical supergravities.

*See also:* Brane Construction of Gauge Theories; Brane Worlds; Branes and Black Hole Statistical Mechanics; Random Algebraic Geometry, Attractors and Flux Vacua; Renormalization: General Theory; Spinors and Spin Coefficients; Stability of Minkowski Space; Supermanifolds; Superstring Theories; Supersymmetric Particle Models; Symmetries and Conservation Laws; Symmetries in Quantum Field Theory: Algebraic Aspects.

## Further Reading

Buchbinder JL and Kuzenko SM (1998) *Ideas and Methods of Supersymmetry and Supergravity*. Bristol: IoP Publishing Ltd.
Stelle KS (1998) BPS branes in supergravity, Trieste 1987 School of High-Energy Physics and Cosmology, arXiv:hep-th/9803116.
Van Nieuwenhuizen P (1981) Supergravity. *Physics Reports* 68: 189–398.
Wess J and Bagger J (1983) *Supersymmetry and Supergravity*. Princeton: Princeton University Press.

# Supermanifolds

**F A Rogers**, King's College London, London, UK



## Introduction

A supermanifold is a generalization of a classical manifold to include coordinates that are in some sense anticommuting. Much of the motivation for the study of supermanifolds comes from supersymmetric physics, where it is useful to have a formalism which treats fermions and bosons in the same way. The underlying reason for the effectiveness of supermanifolds is that anticommuting coordinates allow the fermionic canonical anticommutation relations to be handled in a way analogous to the bosonic canonical commutation relations. Supersymmetric methods have proved immensely effective in fundamental physics; they also play a considerable role in geometrical index theory in mathematics. In this article we describe supermanifolds from two points of view – geometric and algebraic – and consider some of the standard features of manifold calculus, including integration since this is an area where the distinctive features of this generalized geometry are particularly apparent. One situation where supermanifolds are used in physics is in the superspace formulation of supergravity, where the physical fields are found in the component fields in the Taylor expansion of functions on the supermanifold in anticommuting variables. More fundamentally, the symmetry groups of supersymmetric theories have commuting and anticommuting generators, and are examples of super Lie groups, which are supermanifolds with a compatible group structure.

## Some Algebraic Preliminaries

The coordinates of a supermanifold have particular algebraic features which are best understood by introducing some of the basic concepts of superalgebra. (The word super here does not imply superiority, simply the extension of some classical concept to have odd as well as even, anticommuting as well as commuting, elements.) A "super vector space" is a vector space $V$ together with a direct sum decomposition

$$V = V_0 \oplus V_1 \qquad [1]$$

The subspaces $V_0$ and $V_1$ are referred to, respectively, as the even and odd parts of $V$. A general element $v$ of $V$ thus has the unique decomposition $v = v_0 + v_1$ with $v_0$ in $V_0$ and $v_1$ in $V_1$. We will normally consider homogeneous elements, that is, elements $v$ which are either even or odd, with parity denoted by $|v|$, so that $|v| = i$ if $v$ is in $V_i, i = 0, 1$. (Arithmetic of parity indices $i = 0, 1$ is always modulo 2.) A superalgebra is a super vector space whose elements can be multiplied together in such a way that the product of an even element with an even element and that of an odd element with an odd element are both even, while the product of an odd element with an even element is odd; more formally:

**Definition 1**

(i) A "superalgebra" is a super vector space $A = A_0 \oplus A_1$ which is also an algebra which satisfies $A_i A_j \subset A_{i+j}$.
(ii) The superalgebra is "supercommutative" if, for all homogeneous $a, b$ in $A, ab = (-1)^{(|a||b|)} ba$.

If the algebra is supercommutative then odd elements anticommute, and the square of an odd element is zero. The basic supercommutative superalgebra used is the real Grassmann algebra with generators $\mathbf{1}, \beta_1, \beta_2, \ldots$ and relations

$$\mathbf{1}\beta_i = \beta_i \mathbf{1} = \beta_i, \quad \beta_i \beta_j = -\beta_j \beta_i \qquad [2]$$

A typical element of this algebra is then

$$a = a_\emptyset \mathbf{1} + \sum_i a_i \beta_i + \sum_{i<j} a_{ij} \beta_i \beta_j \cdots \qquad [3]$$

This algebra, which is denoted $\mathbb{R}_S$, is a superalgebra with $\mathbb{R}_S := \mathbb{R}_{S,0} \oplus \mathbb{R}_{S,1}$, where $\mathbb{R}_{S,0}$ consists of linear combinations of products of even numbers of the anticommuting generators, while $\mathbb{R}_{S,1}$ is built similarly from odd products.

The Grassmann algebra $\mathbb{R}_S$ is used to build the $(m, n)$-dimensional superspace $\mathbb{R}_S^{m,n}$ in the following way:

**Definition 2.** An $(m, n)$-dimensional superspace is the space

$$\mathbb{R}_S^{m,n} = \underbrace{\mathbb{R}_{S0} \times \cdots \times \mathbb{R}_{S0}}_{m \text{ copies}} \times \underbrace{\mathbb{R}_{S1} \times \cdots \times \mathbb{R}_{S1}}_{n \text{ copies}} \qquad [4]$$

A typical element of $\mathbb{R}_S^{m,n}$ is written as $(x^1, \ldots, x^m; \xi^1, \ldots, \xi^n)$, where the convention is used that lower case Latin letters represent even objects and lower case Greek letters represent odd objects, while small capitals are used for objects of mixed or unspecified parity.

As will be described in more detail below, in the geometric approach supermanifolds are spaces locally modeled on $\mathbb{R}_S^{m,n}$. In order to define a supermanifold, we will need to define a topology on this space, and to have some notion of differentiation. Consider first multilinear functions of purely anticommuting variables. If there are $n$ such variables, $\xi^1, \ldots, \xi^n$, then a multilinear function $\textsc{f}$ can be expressed in the form

$$\textsc{f}(\xi^1, \ldots, \xi^n) = \textsc{f}_\emptyset + \sum_{i=1}^{n} \textsc{f}_i \xi^i + \sum_{1=i<j}^{n} \textsc{f}_{ij} \xi^i \xi^j + \cdots + \textsc{f}_{1\ldots n} \xi^1 \ldots \xi^n \qquad [5]$$

where the coefficients $\textsc{f}_\emptyset, \textsc{f}_i$ and so on are real numbers. Such functions will be known (anticipating the terminology for functions of both odd and even variables) as supersmooth. (A useful notation will be to write

$$\textsc{f}(\xi^1, \ldots, \xi^n) = \sum_\mu \textsc{f}_\mu \xi^\mu \qquad [6]$$

with $\mu$ a multi-index $\mu = \mu_1 \cdots \mu_k$ and $\xi^\mu = \xi^{\mu_1} \cdots \xi^{\mu_k} \mathbf{1}$. The set of multi-indices is restricted to those where $1 \le \mu_1 < \cdots < \mu_k \le n$.) More general supersmooth functions, with the coefficients $\textsc{f}_\emptyset, \ldots$ taking values in $\mathbb{C}, \mathbb{R}_S$, or some other algebra are also possible.

Differentiation of supersmooth functions of anticommuting variables is defined by linearity together with the rule

$$\frac{\partial(\xi^{\mu_1}\xi^{\mu_2}\ldots\xi^{\mu_r})}{\partial \xi^j} = \begin{cases} (-1)^{k-1} \xi^{\mu_1} \ldots \widehat{\xi^{\mu_k}} \ldots \xi^{\mu_r} & \text{if } j = \mu_k \\ 0 & \text{otherwise} \end{cases} \qquad [7]$$

where the caret $\widehat{}$ indicates an omitted factor.

In order to extend the notion of supersmoothness to functions on the more general superspace $\mathbb{R}_S^{m,n}$, we should strictly take note of the fact that an even Grassmann variable is not simply a real or complex variable, as explained in the appendix. Assuming this done, a supersmooth function on the general superspace $\mathbb{R}_S^{m,n}$ can then be defined as a function of the form

$$F(x^1, \ldots, x^m, \xi^1, \ldots, \xi^n) = \sum_\mu \textsc{f}_\mu(x^1, \ldots, x^m) \xi^\mu \qquad [8]$$

with each coefficient function $\textsc{f}_\mu$ a smooth function on $\mathbb{R}^m$.

The final preparatory idea needed is the topology on the superspace $\mathbb{R}_S^{m,n}$. It turns out that a coarse, non-Hausdorff topology leads to most of the supermanifolds used in physics. In order to define this topology, we introduce a mapping

$$\epsilon : \mathbb{R}_S \to \mathbb{R}$$

defined by

$$\epsilon\left(a_\emptyset \mathbf{1} + \sum_i a_i \beta_i + \sum_{i<j} a_{ij} \beta_i \beta_j \cdots\right) = a_\emptyset \qquad [9]$$

and the related mapping

$$\epsilon : \mathbb{R}_S^{m,n} \to \mathbb{R}^m$$

defined by

$$\epsilon((x^1,\ldots,x^m;\,\xi^1,\ldots,\xi^n)) = (\epsilon(x^1),\ldots,\epsilon(x^m)) \qquad [10]$$

These maps project out all the nilpotent Grassmann generators, leaving simply the real part. The topology involves the inverse of these projection maps: a subset $U$ of $\mathbb{R}_S^{m,n}$ is said to be open if and only if there exists an open set $V$ in $\mathbb{R}^m$ such that $U = \epsilon^{-1}(V)$. Thus, an open set is unlimited in the nilpotent directions.

In the sequel, where we consider integration, the superdeterminant of the matrix $M$ of an endomorphism of a super vector space $V$ will be useful. If $V$ is an $(m,n)$-dimensional super vector space (so that $V_0$ has dimension $m$ and $V_1$ dimension $n$), then $M$ will have the block diagonal form

$$\begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix}$$

where the entries of $M_{00}$ and $M_{11}$ are even, whereas those of $M_{10}$ and $M_{10}$ are odd. If $N = M^{-1}$ has block form

$$\begin{pmatrix} N_{00} & N_{01} \\ N_{10} & N_{11} \end{pmatrix}$$

then the superdeterminant of $M$ is defined by

$$\mathrm{S}\,\det M = \det M_{00} \det N_{11}$$

It can be shown that the superdeterminant obeys the product rule, unlike the obvious generalization of the determinant to the super case.

## The Geometric Approach to Supermanifolds

A manifold is a space locally modeled on the topological space $\mathbb{R}^m$, where $m$ is the dimension of the manifold. Thus, each point in a manifold has a neighborhood which is essentially a neighborhood in $\mathbb{R}^m$. The most geometrically intuitive approach to supermanifolds is to generalize this directly by modeling a space locally on an extension of $\mathbb{R}^m$ to include anticommuting variables; the most straightforward space with the required algebraic property is the superspace $\mathbb{R}_S^{m,n}$ built from a Grassmann algebra, leading to a supermanifold of dimension $(m,n)$. (The dimension of a supermanifold is a pair of integers, indicating the numbers of even and odd coordinates of each point.)

The formal definition of a supermanifold will now be given in a manner very closely analogous to that of a classical manifold.

**Definition 3.** Let $M$ be a set.

(i) An $(m,n)$ open chart on $M$ is a pair $(U,\phi)$ such that $U$ is a subset of $M$ and $\phi$ is an injective map of $U$ into $\mathbb{R}_S^{m,n}$, with the image $\phi(U)$ an open set in $\mathbb{R}_S^{m,n}$.

(ii) An $(m,n)$ atlas on $M$ is a collection $\{(U_\alpha,\phi_\alpha)\}$ of $(m,n)$ charts on $M$ such that the $U_\alpha$ cover $M$ and, whenever $U_\alpha \cap U_\beta$ is not empty, the change of coordinate function $\phi_\alpha \circ \phi_\beta^{-1}$ is supersmooth.

An $(m,n)$-dimensional supermanifold is a set $M$ together with a maximal $(m,n)$ atlas on $M$.

The space $M$ is given a topology by defining $U \subset M$ to be open if and only if, for each $\alpha$ such that $U \cap U_\alpha$ is not empty, the set $\phi_\alpha(U \cap U_\alpha)$ is an open subset of $\mathbb{R}_S^{m,n}$.

Examples of supermanifolds include $\mathbb{R}_S^{m,n}$ itself, and also supermanifolds constructed from the data of a vector bundle over a classical manifold in a manner which will now be described. If $N$ is a classical $m$-dimensional real manifold and $E$ is an $n$-dimensional vector bundle over $N$, then an $(m,n)$-dimensional supermanifold can be constructed in the following way: suppose that $\{(V_\alpha,\psi_\alpha)\}$ is an atlas of charts on $N$, so that each $V_\alpha$ is an open subset of $N$ and each $\psi_\alpha$ is an injective map of $V_\alpha$ onto an open subset of $\mathbb{R}^m$, with $\psi_\alpha \circ \psi_\beta^{-1}$ smooth. Suppose further that the $V_\alpha$ are also local trivialization neighborhoods of the bundle $E$ with transition functions $g_{\alpha\beta}: V_\alpha \cap V_\beta \to \mathrm{GL}(n)$. Then we build the supermanifold $M$ by patching together the sets $\epsilon^{-1}(\psi_\alpha(V_\alpha) \times \mathbb{R}_S^{0,n})$ in a consistent way. This leads to a supermanifold with coordinate change functions

$$\phi_\alpha \circ \phi_\beta^{-1}\left(x_\beta^1,\ldots x_\beta^m,\xi_\beta^1,\ldots,\xi_\beta^n\right) = \left(x_\alpha^1,\ldots x_\alpha^m,\xi_\alpha^1,\ldots,\xi_\alpha^n\right)$$

where

$$\begin{aligned}\left(x_\alpha^1,\ldots x_\alpha^m\right) &= \psi_\alpha \circ \psi_\beta^{-1}\left(x_\beta^1,\ldots x_\beta^m\right) \\ \xi_\alpha^j &= \sum_{k=1}^n g_{\alpha\beta jk}\left(x_\beta^1,\ldots,x_\beta^m\right)\xi_\beta^k\end{aligned} \qquad [11]$$

(Here again we refer to the appendix for the way in which functions of even Grassmann variables, as opposed simply to real numbers, are handled.) Particular examples of this construction are the tangent bundle over $N$ and bundles of spinors over $N$. It was actually shown by Batchelor that all real, supersmooth supermanifolds are of this form.

A similar definition may be made of a complex supermanifold using a complex Grassmann algebra, with the coordinate transition functions required to be superanalytic. In this case, supermanifolds which

are not related to vector bundles in the manner described above are possible, basically because partitions of unity do not exist in the analytic setting. An example is the twisted supertorus, which is built over the standard torus and has transition functions $(z, \zeta) \to (z + 1, \zeta)$ and $(z, \zeta) \to (z + a + \alpha\zeta, \zeta + \alpha)$, extending the standard torus with transition functions $z \to z + 1$, $z \to z + a$. (Here $a, \alpha$ are, respectively, even and odd constants.) This supermanifold is an example of a super Riemann surface; such surfaces play an important role in the quantization of the spinning string.

As with classical manifolds, a natural class of functions can be defined on a supermanifold: a function $f$ on an open subset $U$ of the supermanifold $M$ is said to be supersmooth if, for each $\alpha$ such that $U \cap U_\alpha$ is nonempty, the function $f \circ \phi_\alpha^{-1}$ is supersmooth on $\phi_\alpha(U \cap U_\alpha)$. In local coordinates supersmooth functions are such that $f(x^1, \ldots, x^m, \xi^1, \ldots, \xi^n) = \sum_\mu f_{\alpha\mu}(x^1, \ldots, x^m)\xi^\mu$ with each $f_{\alpha\mu}$ a smooth function.

## The Algebraic Approach to Supermanifolds

In the algebraic approach to supermanifolds, it is the algebra of functions, rather than the manifold itself, which is extended to include anticommuting elements. In this approach an $(m, n)$-dimensional supermanifold is defined to be a pair $(N, A)$, where $N$ is an $m$-dimensional classical manifold and $A$ is a sheaf of superalgebras over $N$ with various properties, described below. The statement that $A$ is a sheaf of algebras over $N$ means that corresponding to each open subset $U$ of $N$ there is an algebra $A(U)$; also, if $V \subset U$, there is a "restriction map" $\rho_{U,V}$ mapping $A(U)$ into $A(V)$, and the various restriction maps obey certain consistency conditions. A particular example of such a sheaf (with trivial odd part) is the sheaf $A_\emptyset$ of real-valued functions on $N$, with $A_\emptyset(U) = C^\infty(U)$, the set of real-valued smooth functions on $U$ and $\rho_{U,V}$ mapping a function in $C^\infty(U)$ to its restriction in $C^\infty(V)$. The defining property of the sheaf corresponding to an $(m, n)$-dimensional supermanifold is that there is a cover $\{U_\alpha\}$ of $N$ for which the algebras $A(U_\alpha)$ have the form $A(U_\alpha) \cong C^\infty(U_\alpha) \otimes \Lambda(\mathbb{R}^n)$, so that a typical element $f$ of $A(U_\alpha)$ may be expressed as $f = \sum_\mu f_\mu \xi^\mu$, where $f_\mu \in C^\infty(U_\alpha)$ and $\xi^1, \ldots, \xi^n$ are generators of $\Lambda(\mathbb{R}^n)$. The notation here is chosen to emphasize the close correspondence with the algebra of smooth functions described at the end of the previous section. This makes it clear that, despite an apparent difference, the two approaches lead to essentially equivalent supermanifolds.

The advantage of the algebraic approach is its mathematical elegance and economy – there is no need to introduce the auxiliary Grassmann algebra $\mathbb{R}_S$ in which coordinate functions take values – but from the point of view of physicists, the geometric point of view has two advantages: first, it is closer to the standard manifold picture and thus easier to grasp, and, second, it allows a wider class of supermanifolds, because Grassmann constants are allowed; for instance, the twisted supertorus described above cannot be included in the algebraic approach without either introducing an auxiliary algebra or moving to the more difficult concept of a family of supermanifolds.

While there have been various attempts to develop infinite-dimensional supermanifolds, most of the constructions have been developed for very specific purposes, such as path integration and functional integration methods for theories with fermions. Even the question of defining a basic infinite-dimensional superalgebra with the necessary analytic properties, such as a Hilbert–Banach superalgebra, requires sophisticated procedures, so that the development of a theory of infinite-dimensional supermanifolds becomes extremely technical.

## Calculus on Supermanifolds

Much of the calculus of functions on supermanifolds proceeds in simple analogy to that of classical manifolds, with addition sign factors occurring whenever two odd quantities are transposed. For instance, a vector field on $M$ may be described as a superderivation of the algebra of supersmooth functions on $M$, that is, a linear mapping of this space obeying the super Leibniz rule $X fg = Xf\, g + (-1)^{(|X||f|)} f\, Xg$. Standard examples of vector fields (defined locally) are coordinate derivatives $\partial/\partial x^i$ and $\partial/\partial \xi^j$, defined by $(\partial/\partial x^i)f = \partial_i(f \circ \phi)$ and $(\partial/\partial \xi^j)f = \partial_{j+m}(f \circ \phi)$ with $\phi$ the coordinate function corresponding to the coordinates $(x^1, \ldots, x^m; \xi^1, \ldots, \xi^n)$. Equipped with this concept of vector field, much of differential calculus on manifolds can be directly generalized to supermanifolds in a relatively straightforward way. However, in the case of integration the situation is quite different. The standard approach to integration of anticommuting variables is the Berezin integral, which is a formal, algebraic integral that is not an antiderivative and has no measure-theoretic features. There are various reasons why such an integral is used: for instance, even the simple function $\xi$ of a single anticommuting variable has no antiderivative, while the topology on $\mathbb{R}_S^{m,n}$ does not allow open sets which discriminate in

odd directions. Additionally, when changing variables on $\mathbb{R}_S^{m,n}$ it is the superdeterminant of the Jacobian matrix which must be used. In the purely odd sector, differentials thus transform the "wrong" way.

The Berezin integral of a function $f$ of $n$ anti-commuting variables is defined by

$$\int \mathrm{d}^n\xi \left( \sum_\mu f_\mu \xi^\mu \right) = f_{1\dots n} \qquad [12]$$

In other words, Berezin integration simply picks out the coefficient of the highest-order term, thus resembling differentiation more than integration in the classical sense. Nonetheless, the Berezin integral has very useful properties, in particular allowing direct analoges of Fourier transformations and integral kernel. Given that it is the algebra of functions, and the operators acting on these algebras, which is the key element in supergeometry, these are vital properties of the integral.

The transformation rule under change of variable is the inverse of that which one expects. For instance, in the case of a single variable, if one makes the transformation $\xi \to \phi = a\xi + \beta$ with $a$ and $\beta$ constants, a direct calculation shows that the integral is invariant provided that one sets $\mathrm{d}\xi = a\,\mathrm{d}\phi$.

Integration on $\mathbb{R}_S^{m,n}$ is essentially defined by combining classical integration for the even variables with Berezin integration for odd variables, giving

$$\int_{\epsilon^{-1}(V)} \mathrm{d}^m x\, \mathrm{d}^n\xi \left( \sum_\mu f_\mu(x^1, \dots, x^m)\xi^\mu \right) = \int_V \mathrm{d}^m x \big( f_{1\dots n}(x^1, \dots, x^m) \big) \qquad [13]$$

This also defines integration on supermanifolds, provided that we can find a rule for the change of variable. This, as indicated above, may be done by using the superdeterminant of the Jacobian matrix. Suppose that $(y, \phi)$ are a new set of coordinates on our supermanifold. Then an invariant definition of integral is obtained if we set

$$\mathrm{d}^m y\, \mathrm{d}^n\xi = \mathrm{Sdet} \begin{pmatrix} \dfrac{\partial y}{\partial x} & \dfrac{\partial y}{\partial \xi} \\ \dfrac{\partial \phi}{\partial x} & \dfrac{\partial \phi}{\partial \xi} \end{pmatrix} \mathrm{d}^m x\, \mathrm{d}^n \xi \qquad [14]$$

## Appendix

We now describe the device which allows functions of even Grassmann variables to be handled simply as functions of conventional variables. The necessary class of functions is captured by defining supersmooth functions on $\mathbb{R}_S^{m,0}$ as extensions by Taylor expansion from smooth functions on $\mathbb{R}^m$.

**Definition 4.** The function $F : \mathbb{R}_S^{m,0} \to \mathbb{R}_S$ is said to be supersmooth if there exists a smooth function $\tilde{F} : \mathbb{R}^m \to \mathbb{R}$, such that

$$\begin{aligned} F(x^1, \dots, x^m) &= \tilde{F}(\epsilon(x)) + \sum_{i=1}^m (x^i - \epsilon(x^i)\mathbf{1}) \frac{\partial \tilde{F}}{\partial x^i}(\epsilon(x)) \\ &\quad + \frac{1}{2} \sum_{i,j=1}^m (x^i - \epsilon(x^i)\mathbf{1}) \\ &\quad \times (x^j - \epsilon(x^j)\mathbf{1}) \frac{\partial^2 \tilde{F}}{\partial x^i \partial x^j}(\epsilon(x)) \dots \end{aligned} \qquad [15]$$

(Although this Taylor series will in general be infinite, it gives well-defined coefficients for each $\beta_\mu$ in the expansion [3], so that the value of $F$ is a well-defined element of $\mathbb{R}_S$.) A number of different classes of function can be obtained, by varying the space in which the function $\tilde{F}$ takes its value.

*See also:* Batalin–Vilkovisky Quantization; BRST Quantization; Graded Poisson Algebras; Path-Integrals in Non Commutative Geometry; Random Matrix Theory in Physics; Supergravity; Superstring Theories; Supersymmetric Particle Models; Supersymmetric Quantum Mechanics.

## Further Reading

Batchelor M (1979) The structure of supermanifolds. *Transactions of the American Mathematical Society* 253: 329–338.
Batchelor M (1980) Two approaches to supermanifolds. *Transactions of the American Mathematical Society* 258: 257–270.
Berezin FA (1987) *Introduction to Superanalysis*. Dordrecht: Reidel.
Berezin FA and Leĭtes DA (1976) Supermanifolds. *Soviet Maths Doklady* 16: 1218–1222.
Crane L and Rabin JM (1988) Super Riemann surfaces: uniformization and Teichmüller theory. *Communications in Mathematical Physics* 113: 601–623.
DeWitt BS (1992) *Supermanifolds*. Cambridge: Cambridge University Press.
Howe PS (1979) Super Weyl transformations in two dimensions. *Journal of Physics A* 12: 393–402.
Jadczyk A and Pilch K (1981) Superspace and supersymmetries. *Communications in Mathematical Physics* 78: 373–390.
Kostant B (1977) Graded manifolds, graded Lie theory and prequantization. In: *Differential Geometric Methods in Mathematical Physics*,. Lecture Notes in Mathematics, Springer.
Kupsch J and Smolyanov O (2000) Hilbert norms for graded algebras. *Proceedings of the American Mathematical Society* 128: 1647–1653.
Polchinski J (1998) *String Theory*, vol. II. Cambridge: Cambridge University Press.
Rogers A (2003) Supersymmetry and Brownian motion on supermanifolds. *Infinite Dimensional Analysis, Quantum Probability and Related Topics* 6(suppl. 1): 83–102.
Salam A and Strathdee J (1974) Super-gauge transformations. *Nuclear Physics B* 76: 477–482.

Voronov AA (1992) Geometric integration theory on supermanifolds. *Soviet Scientific Reviews C: Mathematical Physics Reviews* 9: 1–138.
Wess J and Bagger J (1983) *Supersymmetry and Supergravity*. Princeton: Princeton University Press.
West P (1990) *Introduction to Supersymmetry and Supergravity*. Singapore: World Scientific.

# Superstring Theories

**C Bachas and J Troost**, Ecole Normale Supérieure, Paris, France



## Introduction

String theory postulates that all elementary particles in nature correspond to different vibration states of an underlying relativistic string. In the quantum theory both the frequencies and the amplitudes of vibration are quantized, so that the quantum states of a string are discrete. They can be characterized by their mass, spin, and various gauge charges. One of these states has zero mass and spin equal to $2\hbar$, and can be identified with the messenger of gravitational interactions, the graviton. Thus, string theory is a candidate for a unified theory of all fundamental interactions, including quantum gravity.

In this article, we discuss the theory of superstrings as consistent theories of quantum gravity. The aim is to provide a quick (mostly lexicographic and bibliographic) entry to some of the salient features of the subject for a nonspecialist audience. Our treatment is thus neither complete nor comprehensive – there exist for this several excellent expert books, in particular by Green, *et al.* (1987) and by Polchinski (1998). An introductory textbook by Zwiebach (2004) is also highly recommended for beginners. Several other complementary reviews on various aspects of superstring theories are available on the internet (see the "Further reading" section); some more will be given as we proceed.

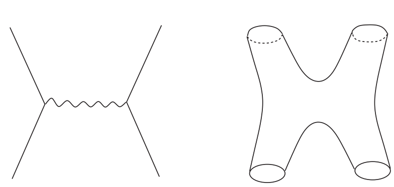


**Figure 1** A four-particle and a four-string interaction.

## The Five Superstring Theories

Theories of relativistic extended objects are tightly constrained by anomalies, that is, quantum violations of classical symmetries. These arise because the classical trajectory of an extended $p$-dimensional object (or "$p$-brane") is described by the embedding $X^\mu(\zeta^a)$, where $\zeta^{a=0,\dots,p}$ parametrize the brane world volume, and $X^{\mu=0,\dots,D-1}$ are coordinates of the target space. The quantum mechanics of a single $p$-brane is therefore a $(p+1)$-dimensional quantum field theory, and as such suffers *a priori* from ultraviolet divergences and anomalies. The case $p=1$ is special in that these problems can be exactly handled. The story for higher values of $p$ is much more complicated, as will become apparent later on.

The theory of ordinary loops in space is called closed bosonic string theory. The classical trajectory of a bosonic string extremizes the Nambu–Goto action (proportional to the invariant area of the world sheet)

$$S_{\mathrm{NG}} = -\frac{1}{2\pi\alpha'} \int \mathrm{d}^2\zeta \sqrt{-\det(G_{\mu\nu}\partial_a X^\mu \partial_b X^\nu)} \qquad [1]$$

where $G_{\mu\nu}(X)$ is the target-space metric, and $\alpha'$ is the Regge slope (which is inversely proportional to the string tension and has dimensions of length squared). In flat spacetime, and for a conformal choice of world-sheet parameters $\zeta^\pm = \zeta^0 \pm \zeta^1$, the equations of motion read:

$$\partial_+\partial_- X^\mu = 0 \quad \text{and} \quad \eta_{\mu\nu}\partial_\pm X^\mu \partial_\pm X^\nu = 0 \qquad [2]$$

with $\eta_{\mu\nu}$ the Minkowski metric. The $X^\mu$ are thus free two-dimensional fields, subject to quadratic phase-space constraints known as the Virasoro conditions. These can be solved consistently at the quantum level in the critical dimension $D=26$. Otherwise, the symmetries of eqns [2] are anomalous: either Lorentz invariance is broken, or there is a conformal anomaly leading to unitarity problems. (For $D<26$, unitary noncritical string theories in highly curved rather than in the originally flat background can be constructed.)

Even for $D=26$, bosonic string theory is, however, sick because its lowest-lying state is a tachyon,

The finiteness of string perturbation theory has been, strictly speaking, only established up to two loops – for a recent review see D'Hoker and Phong (2002). However, even though the technical problem is open and hard, the qualitative case for all-order finiteness is convincing. It can be illustrated with the torus diagram which makes a one-loop contribution to string amplitudes. The thin torus of **Figure 2** could be traced either by a short, light string propagating (virtually) for a long time, or by a long, heavy string propagating for a short period of time. In conventional field theory, these two virtual trajectories would have made distinct contributions to the amplitude, one in the infrared and the second in the ultraviolet region. In string theory, on the other hand, they are related by a modular transformation (that exchanges $\zeta^0$ with $\zeta^1$) and must not, therefore, be counted twice. A similar kind of argument shows that all potential divergences of string theory are infrared – they are therefore kinematical (i.e., occur for special values of the external momenta), or else they signal an instability of the vacuum and should cancel if one expands around a stable ground state.

The low-energy limit of the heterotic and type I string theories is $N=1$ supergravity plus super Yang–Mills. In addition to the $N=1$ graviton multiplet, the massless spectrum now also includes gauge bosons and their associated gauginos. The two-derivative effective action in the heterotic case reads:

$$S_{\text{het}} = \frac{1}{2\kappa^2}\int \mathrm{d}^{10}x\sqrt{-G}\,\mathrm{e}^{-2\Phi} \times \left[R + 4\partial_\mu\Phi\partial^\mu\Phi + \frac{\kappa^2}{g^2_{\text{YM}}}\mathrm{tr}(F_{\mu\nu}F^{\mu\nu}) - \frac{1}{2}\left|\mathrm{d}B_2 - \frac{\kappa^2}{g^2_{\text{YM}}}\omega_3^{\text{gauge}}\right|^2\right] + \text{fermions} \qquad [7]$$

where $\omega_3^{\text{gauge}} = \mathrm{tr}(AdA + (2/3)A^3)$ is the Chern–Simons gauge 3-form. Again, supersymmetry fixes completely the above action – the only freedom is in the choice of the gauge group and of the Yang–Mills coupling $g_{\text{YM}}$. Thus, up to redefinitions of the fields, the type I theory has necessarily the same low-energy limit.

The $D=10$ supergravity plus super Yang–Mills has a hexagon diagram that gives rise to gauge and gravitational anomalies, similar to the triangle anomaly in $D=4$. It turns out that for the two special groups $E_8 \times E_8$ and SO(32), the structure of these anomalies is such that they can be canceled by a combination of local counter-terms. One of them is of the form $\int B_2 \wedge X_8(F,R)$, where $X_8$ is an 8-form quartic in the curvature and/or Yang–Mills field strength. The other is already present in the lower line of expression [7], with the replacement $\omega_3^{\text{gauge}} \rightarrow \omega_3^{\text{gauge}} - \omega_3^{\text{Lorentz}}$, where the second Chern–Simons form is built out of the spin connection. Note that these modifications of the effective action involve terms with more than two derivatives, and are not required by supersymmetry at the classical level. The discovery by Green and Schwarz that string theory produces precisely these terms (from integrating out the massive string modes) was called the "first superstring revolution."

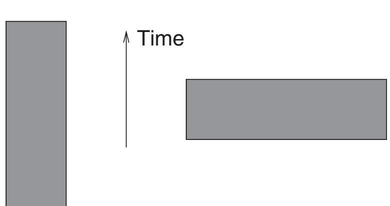



**Figure 2** The same torus diagram viewed in two different channels.

## D-Branes

A large window into the nonperturbative structure of string theory has been opened by the discovery of D(irichlet)-branes, and of strong/weak-coupling duality symmetries. A D$p$ brane is a solitonic $p$-dimensional excitation, defined indirectly by the property that open string endpoints can attach to its world volume (see **Figure 3**). Stable D$p$ branes exist in the type IIA and type IIB theories for $p$ even, respectively, odd, and in the type I theory for $p=1$ and 5. They are charged under the R–R $(p+1)$-form potential or, for $p > 4$, under its magnetic dual. Strictly speaking, only for $0 \leq p \leq 6$ do D-branes resemble regular solitons the word stands for "solitary waves"). The D7 branes are more like

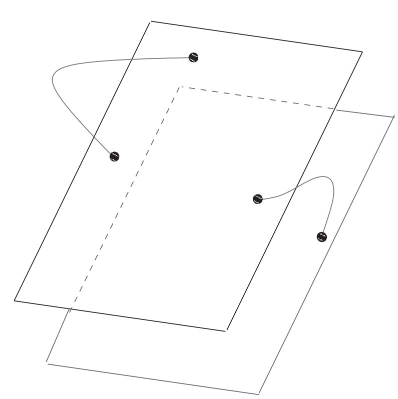


**Figure 3** D-branes and open strings.

cosmic strings, the D8 branes are domain walls, while the D9 branes are spacetime filling. Indeed, type I string theory can be thought as arising from type IIB through the introduction of an orientifold 9-plane (required for tadpole cancelation) and of 32 D9 branes.

The low-energy dynamics of a D$p$ brane is described by a supersymmetric abelian gauge theory, reduced from ten down to $p+1$ dimensions. The gauge field multiplet includes $9-p$ real scalars, plus gauginos in the spinor representation of the R-symmetry group SO$(9-p)$. These are precisely the massless states of an open string with endpoints moving freely on a hyperplane. The real scalar fields are Goldstone modes of the broken translation invariance, that is, they are the transverse coordinate fields $\vec{Y}(\xi^a)$ of the D-brane. The bosonic part of the low-energy effective action is the sum of a Dirac–Born–Infeld (DBI) and a Chern–Simons (CS) like term:

$$I_p = -T_p \int d^{p+1}\xi \, e^{-\Phi} \sqrt{-\det(\hat{G}_{ab} + \mathcal{F}_{ab})} - \rho_p \int \sum_n \hat{C}_n \wedge e^{\mathcal{F}} \qquad [8]$$

where $\mathcal{F}_{ab} = \hat{B}_{ab} + 2\pi\alpha' F_{ab}$, hats denote pullbacks on the brane of bulk tensor fields (e.g., $\hat{G}_{ab} = G_{\mu\nu}\partial_a Y^\mu \partial_b Y^\nu$), $F_{ab}$ is the field strength of the world-volume gauge field, and in the CS term one is instructed to keep the $(p+1)$-form of the expression under the integration sign. The constants $T_p$ and $\rho_p$ are the tension and charge density of the D-brane. As was the case for the effective supergravities, the above action receives curvature corrections that are higher order in the $\alpha'$ expansion. Note however that a class of higher-order terms have been already resummed in expression [8]. These involve arbitrary powers of $F_{ab}$, and are closely related more precisely $T$-dual, see later) to relativistic effects which can be important even in the weak-acceleration limit. When refereing to the D9 branes of the type I superstring, the action [8] includes the GS terms required to cancel the gauge anomaly.

The tension and charge density of a D$p$ brane can be extracted from its coupling to the (closed-string) graviton and R–R $(p+1)$-form, with the result:

$$T_p^2 = \rho_p^2 = \frac{\pi}{\kappa^2}(4\pi^2\alpha')^{3-p} \qquad [9]$$

The equality of tension and charge follows from unbroken supersymmetry, and is also known as a Bogomol'nyi–Prasad–Sommerfeld (BPS) condition. It implies that two or more identical D-branes exert no net static force on each other, because their R–R repulsion cancels exactly their gravitational attraction. A nontrivial check of the result [9] comes from the Dirac quantization condition (generalized to extended objects by Nepomechie and Teitelboim). Indeed, a D$p$ brane and a D$(6-p)$-brane are dual excitations, like electric and magnetic charges in four dimensions, so their couplings must obey

$$2\kappa^2 \rho_p \rho_{6-p} = 2\pi k \quad \text{where } k \in Z \qquad [10]$$

This ensures that the Dirac singularity of the long-range R–R fields of the branes does not lead to an observable Bohm–Aharonov phase. The couplings [9] obey this condition with $k=1$, so that D-branes carry the smallest allowed R–R charges in the theory.

A simple but important observation is that open strings living on a collection of $n$ identical D-branes have matrix-valued wave functions $\psi_{ij}$, where $i, j = 1, \ldots, n$ label the possible endpoints of the string. The low-energy dynamics of the branes is thus described by a nonabelian gauge theory, with group U$(n)$ if the open strings are oriented, and SO$(n)$ or Sp$(n)$ if they are not. We have already encountered such Chan–Paton factors in our discussion of the type I superstring. More generally, this simple property of D-branes has led to many insights on the geometric interpretation and engineering of gauge theories, which are reviewed in the articles Brane Construction of Gauge Theories and Gauge Theories from Strings. It has also placed on a firmer footing the idea of a brane world, according to which the fields and interactions of the standard model would be confined to a set of D-branes, while gravitons are free to propagate in the bulk (for reviews, *see* Brane Worlds and reference Lust (2004)). It has, finally, inspired the gauge/string theory or AdS/CFT correspondence (*see* Ads/CFT Correspondence and Aharony *et al.* (2000)) on which we will comment later.

## Dualities and M Theory

One other key role of D-branes has been to provide evidence for the various nonperturbative duality conjectures. Dual descriptions of the same physics arise also in conventional field theory. A prime example is the Montonen–Olive duality of four-dimensional, $N=4$ supersymmetric Yang–Mills, which is the low-energy theory describing the dynamics of a collection of D3 branes. The action

for the gauge field and six associated scalars $\Phi^I$ (all in the adjoint representations of the gauge group $G$) is

$$S_{N=4} = -\frac{1}{4g^2}\int d^4x\,\mathrm{tr} \times \left(F_{\mu\nu}F^{\mu\nu} + 2\sum_I D_\mu\Phi^I D^\mu\Phi^I + \sum_{I<J} 2[\Phi^I,\Phi^J]^2\right) - \frac{\theta}{32\pi^2}\int d^4x\,\mathrm{tr}\left(F_{\mu\nu}{}^*F^{\mu\nu}\right) + \text{fermionic terms} \qquad [11]$$

Consider for simplicity the case $G = SU(2)$. The scalar potential has flat directions along which the six $\Phi^I$ commute. By an SO(6) R-symmetry rotation, we can set all but one of them to zero, and let $<\mathrm{tr}(\Phi^1\Phi^1)> = v^2$ in the vacuum. In this "Coulomb phase" of the theory, a U(1) gauge multiplet stays massless, while the charged states become massive by the Higgs effect. The theory admits furthermore smooth magnetic-monopole and dyon solutions, and there is an elegant formula for their mass:

$$M = v|n_{el} + \tau n_{mg}|, \quad \text{where } \tau = \frac{\theta}{2\pi} + \frac{4\pi i}{g^2} \qquad [12]$$

and $n_{el}(n_{mg})$ denotes the quantized electric (magnetic) charge. This is a BPS formula that receives no quantum corrections. It exhibits the $SL(2,Z)$ covariance of the theory,

$$\tau \to \frac{a\tau + b}{c\tau + d}$$

and

$$(n_{el}, n_{mg}) \to (n_{el}, n_{mg})\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} \qquad [13]$$

Here $a, b, c, d$ are integers subject to the condition $ad - bc = 1$. Of special importance is the transformation $\tau \to -1/\tau$, which exchanges electric and magnetic charges and (at least for $\theta = 0$) the strong- with the weak-coupling regimes. For more details see the review by Harvey (1996).

The extension of these ideas to string theory can be illustrated with the strong/weak- coupling duality between the type I theory, and the $\mathrm{Spin}(32)/Z_2$ heterotic string. Both have the same massless spectrum and low-energy action, whose form is dictated entirely by supersymmetry. The only difference lies in the relations between the string and supergravity parameters. Eliminating the latter, one finds

$$\lambda_{het} = \frac{1}{2\lambda_I} \qquad \text{and} \qquad \alpha'_{het} = \sqrt{2}\lambda_I\alpha'_I \qquad [14]$$

It is thus tempting to conjecture that the strongly coupled type I theory has a dual description as a weakly coupled heterotic string. These are, indeed, the only known ultraviolet completions of the theory [7]. Furthermore, for $\lambda_I \gg 1$, the D1 brane of the type I theory becomes light, and could be plausibly identified with the heterotic string. This conjecture has been tested successfully by comparing various supersymmetry-protected quantities (such as the tensions of BPS excitations and special higher-derivative terms in the effective action), which can be calculated exactly either semiclassically, or at a given order in the perturbative expansion. Testing the duality for nonprotected quantities is a hard and important problem, which looks currently out of reach.

The other three string theories have also well-motivated dual descriptions at strong coupling $\lambda$. The type IIB theory is believed to have an $SL(2,Z)$ symmetry, similar to that of the $N = 4$ super Yang–Mills. (Note that $\lambda$ is a dynamical parameter, that changes with the vacuum expectation value of the dilaton $<\phi>$. Thus, dualities are discrete gauge symmetries of string theory.) The type IIA theory has a more surprising strong-coupling limit: it grows one extra dimension (of radius $R_{11} = 1/\lambda\sqrt{\alpha'}$), and can be approximated at low energy by the maximal 11-dimensional supergravity of Cremmer, Julia, and Scherk. The latter is a very economical theory – its massless bosonic fields are only the graviton and a 3-form potential $A_3$. The bosonic part of the action reads

$$S_{11D} = \frac{1}{2\kappa_{11}^2}\int d^{11}x\sqrt{-G}\left(R - \tfrac{1}{2}|F_4|^2\right) - \frac{1}{12\kappa_{11}^2}\int A_3 \wedge F_4 \wedge F_4 \qquad [15]$$

The electric and magnetic charges of the 3-form are a (fundamental?) membrane and a solitonic 5-brane. Standard Kaluza–Klein reduction on a circle maps $S_{11D}$ to the IIA supergravity action [6], where $G_{\mu\nu}$, $\phi$, and $C_1$ descend from the 11-dimensional graviton, and $B_2$ and $C_3$ from the 3-form $A_3$. Furthermore, all BPS excitations of the type IIA string theory have a counterpart in 11 dimensions, as summarized in **Table 1**. Finally, if one compactifies the eleventh dimension on an interval (rather than a circle), one finds the conjectured strong-coupling limit of the $E_8 \times E_8$ heterotic string.

The web of duality relations can be extended by compactifying further to $D \leq 9$ dimensions. Readers interested in more details should consult Polchinski (1998) or one of the many existing reviews of the subject (Townsend (1996), see also "Further Reading" section). In nine dimensions, in particular, the two type II theories, as well as the two heterotic superstrings, are pairwise $T$-dual. $T$-duality is a perturbative symmetry (thus firmly established, not

**Table 1** BPS excitations of type IIA string theory, and their counterparts in $\mathcal{M}$ theory compactified on a circle of radius $R_{11}$

| *Tension* | *Type IIA* | $\mathcal{M}$ on $S^1$ | *Tension* |
|---|---|---|---|
| $(\sqrt{\pi}/\kappa_{10})\,(2\pi\sqrt{\alpha'})^3$ | D0 brane | K–K excitation | $1/R_{11}$ |
| $T_F = (2\pi\alpha')^{-1}$ | String | Wrapped membrane | $2\pi R_{11}\,(2\pi^2/\kappa_{11}^2)^{1/3}$ |
| $(\sqrt{\pi}/\kappa_{10})(2\pi\sqrt{\alpha'})$ | D2 brane | Membrane | $T_2^M = (2\pi^2/\kappa_{11}^2)^{1/3}$ |
| $(\sqrt{\pi}/\kappa_{10})(2\pi\sqrt{\alpha'})^{-1}$ | D4 brane | Wrapped 5-brane | $R_{11}\,(2\pi^2/\kappa_{11}^2)^{2/3}$ |
| $(\pi/\kappa_{10}^2)(2\pi\alpha')$ | NS-5-brane | 5-brane | $(1/2\pi)\,(2\pi^2/\kappa_{11}^2)^{2/3}$ |
| $(\sqrt{\pi}/\kappa_{10})(2\pi\sqrt{\alpha'})^{-3}$ | D6 brane | K–K monopole | $2\pi^2 R_{11}^2/\kappa_{11}^2$ |

From Bachas CP (1997) Lectures on D-branes. In: Olive DI and West PC (eds.) *Duality and Supersymmetric Theories*, Proceedings, Easter School, Newton Institute, Euroconference, Cambridge, UK, April 7–18. With permission of Cambridge University Press.

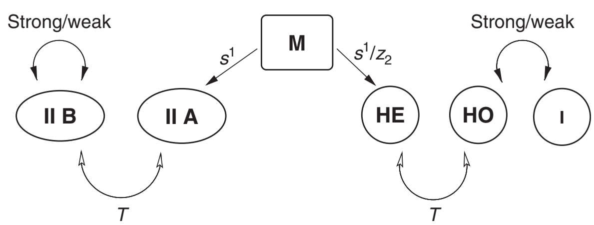



**Figure 4** Web of dualities in nine dimensions. From Bachas CP (1997) Lectures on D-branes. In: Olive DI and West PC (eds.) *Duality and Supersymmetric Theories*, Proceedings, Easter School, Newton Institute, Euroconference, Cambridge, UK, April 7–18. With permission of Cambridge University Press.

only conjectured) which exchanges momentum and winding modes. Putting together all the links one arrives at the fully connected web of Figure 4. This makes the point that all five consistent superstrings, and also 11-dimensional supergravity, are limits of a unique underlying structure called M theory. (For lack of a better definition, "M" is sometimes also used to denote the $D = 11$ supergravity plus supermembranes, as in Figure 4.) A background-independent definition of M theory has remained elusive. Attempts to define it as a matrix model of D0 branes, or by quantizing a fundamental membrane, proved interesting but incomplete. A difficulty stems from the fact that in a generic background, or in $D = 11$ Minkowski spacetime, there is only a dimensionful parameter fixing the scale at which the theory becomes strongly coupled.

## Other Developments and Outlook

We have not discussed in this brief review some important developments covered in other contributions to the encyclopedia. For the reader's convenience, and for completeness, we enumerate (some of) them giving the appropriate cross-references:

*Compactification*. To make contact with the standard model of particle physics, one has to compactify string theory on a six-dimensional manifold. There is an embarassment of riches, but no completely realistic vacuum and, more significantly, no guiding dynamical principle to help us decide (*see* Compactification of Superstring Theory). The controlled (and phenomenologically required) breaking of spacetime supersymmetry is also a problem.

*Conformal field theory and quantum geometry*. The algebraic tools of 2D conformal field theory, both bulk and boundary (*see* Two-Dimensional Conformal Field Theory and Vertex Operator Algebras), play an important role in string theory. They allow, in certain cases, a resummation of $\alpha'$ effects, thereby probing the regime where classical geometric notions do not apply.

*Microscopic models of black holes*. Charged extremal black holes can be modeled in string theory by BPS configurations of D-branes. This has led to the first microscopic derivation of the Bekenstein–Hawking entropy formula, a result expected from any consistent theory of quantum gravity. As with the tests of duality, the extension of these results to neutral black holes is a difficult open problem – *see* Branes and Black Hole Statistical Mechanics.

*AdS/CFT and holography*. A new type of (holographic) duality is the one that relates supersymmetric gauge theories in four dimensions to string theory in asymptotically anti-de Sitter spacetimes. The sharpest and best-tested version of this duality relates $N = 4$ super Yang–Mills to string theory in $AdS_5 \times S_5$. Solving the $\sigma$-model in this latter background is one of the keys to further progress in the subject (*see* AdS/CFT Correspondence).

*String phenomenology*. Finding an experimental confirmation of string theory is clearly one of the most pressing outstanding questions. There exist several interesting possibilities for this – cosmic strings, large extra dimensions, modifications of gravity, primordial cosmology (*see* String Theory: Phenomenology for a

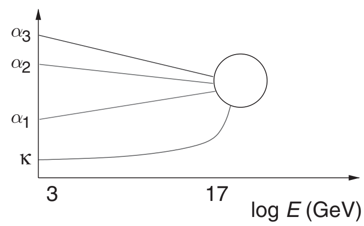



**Figure 5** The unification of couplings.

review). Here we point out the one supporting piece of experimental evidence: the unification of the gauge couplings of the (supersymmetric, minimal) standard model at a scale close to, but below the Planck scale, as illustrated in Figure 5. This is a generic "prediction" of string theory, especially in its heterotic version.

*See also:* AdS/CFT Correspondence; Boundary Conformal Field Theory; Brane Construction of Gauge Theories; Brane Worlds; Branes and Black Hole Statistical Mechanics; Compactification of Superstring Theory; Derived Categories; Electroweak Theory; Gauge Theories from Strings; Noncommutative Geometry from Strings; Supermanifolds; String Field Theory; String Theory: Phenomenology; Supergravity; Two-Dimensional Conformal Field Theory and Vertex Operator Algebras; Wheeler–DeWitt Theory.

## Further Reading

Aharony O, Gubser SS, Maldacena JM, Ooguri H, and Oz Y (2000) Large *N* field theories, string theory and gravity. *Physics Report* 323: 183 (arXiv:hep-th/9905111).
Angelantonj C and Sagnotti A (2002) Open strings. *Physics Report* 371: 1.
Angelantonj C and Sagnotti A (2003) Open strings – erratum. *Physics Report* 376: 339 (arXiv:hep-th/0204089).
Bachas CP (1997) Lectures on D-branes. In: Olive DI and West PC (eds.) *Duality and Supersymmetric Theories*, Proceedings, Easter School, Newton Institute, Euroconference, Cambridge, UK, April 7–18, arXiv:hep-th/9806199.
Berkovits N (2002) ICTP lectures on covariant quantization of the superstring, arXiv:hep-th/0209059.
D'Hoker E and Phong DH (2002) Lectures on two-loop superstrings, arXiv:hep-th/0211111.
Green MB, Schwarz JH, and Witten E (1987) *Superstring Theory*. Vol. 1: *Introduction*; Vol. 2: *Loop Amplitudes, Anomalies and Phenomenology*. Cambridge: Cambridge University Press.
Harvey JA (1996) Magnetic monopoles, duality, and supersymmetry. arXiv:hep-th/9603086.
Kiritsis E (2004) D-branes in standard model building, gravity and cosmology. *Fortschritte der Physik* 52: 200 (arXiv:hep-th/0310001).
Lust D (2004) Intersecting brane worlds: A path to the standard model? *Classical and Quantum Gravity* 21: S1399 (arXiv:hep-th/0401156).
Obers NA and Pioline B (1999) U-duality and M-theory. *Physics Report* 318: 113 (arXiv:hep-th/9809039).
Polchinski J (1998) *String Theory. Vol. 1: An Introduction to the Bosonic String;* Vol. 2: *Superstring Theory and Beyond*. Cambridge: Cambridge University Press.
Salam A and Sezgin E (1989) *Supergravities in Diverse Dimensions*. Amsterdam: North-Holland/World Scientific.
Sen A (1997) An introduction to non-perturbative string theory, arXiv:hep-th/9802051.
The following URL of the HEP-SPIRES database: http://www.slac.stanford.edu collects many popular string theory reviews.
Townsend PK (1996) Four lectures on M-theory, arXiv:hep-th/9612121.
Zwiebach B (2004) *A First Course in String Theory*. Cambridge: Cambridge University Press.

# Supersymmetric Particle Models

**S Pokorski**, Warsaw University, Warsaw, Poland



## Introduction

Supersymmetric quantum field theories (*see* Supergravity) are characterized by the existence of one ($N = 1$ supersymmetry) or several ($N > 1$ extended supersymmetry) conserved Noether-like charges $Q_A$ $A = 1, \ldots, N$, which establish symmetry links between particle states of different spin. Supersymmetry ensures equal numbers of bosonic and fermionic particle states. If it is exact, bosons and fermions related by supersymmetry transformations have equal masses. Moreover, supersymmetry imposes stringent relations between interactions which involve particles of different spin. This gives rise to a special ultraviolet behavior of supersymmetric theories. Their ultraviolet divergences are much softer than in nonsupersymmetric theories. In particular, $N = 4$ supersymmetric quantum field theories are finite and for any $N$ they are free from quadratic divergences plaguing ordinary theories with elementary scalars. $N > 4$ supersymmetric theories necessarily involve particles of spin higher than 1 and are not renormalizable. Supersymmetry promoted to a local symmetry includes gravity.

Only $N = 1$ supersymmetric theories allow for chiral fermions which are the fundamental objects in elementary particle interactions (*see* Standard Model of Particle Physics). This is because parity and

charge conjugation symmetries are violated in weak interactions. Therefore, $N > 1$ theories may not be of immediate phenomenological relevance. However, they may be useful for constructing supersymmetric theories in more than four dimensions (more than three spatial dimensions). Chiral (effective) theory in four dimensions can be then obtained after compactification of extra dimensions. For instance, $N=2$ theory in five dimensions $(x_\mu, y)$ compactified on a circle with reflection symmetry $y \to -y$ (orbifold compactification) gives chiral $N=1$ theory in four dimensions.

Absence of quadratic divergences in supersymmetric theories is the main argument supporting the belief that fundamental interactions of elementary particles at energies not higher that $\mathcal{O}(1\,\text{TeV})$ should be described by an (approximately) $N=1$ supersymmetric extension of the standard model (SM). Indeed, supersymmetric models elegantly solve the so-called hierarchy problem of the SM. At present, supersymmetry remains a theoretical hypothesis. No experimental evidence for it has been found yet (for experimental lower bounds on the masses of supersymmetric particles see Eidelman *et al.* (2004)). Supersymmetric models will be tested experimentally at the Large Linear Collider at CERN (Geneva), after the completion of its construction in 2007. Supergravity theories may be physically relevant as an intermediate step in constructing phenomenologically viable models from superstring theories.

The essence of the hierarchy problem of the standard model (SM) – the successful $\mathrm{SU}(3)_c \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ gauge theory of interactions of quarks and leptons at energies up to about 100 GeV – is the following. By itself, the SM does not explain the value of the Fermi scale $v$ of the electroweak $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry breaking ($v \sim G_{\mathrm{F}}^{-1/2}$ where $G_{\mathrm{F}}$ is the Fermi constant determined by the life time of the muon). Indeed, in the SM, the electroweak symmetry breaking is realized by an elementary Higgs field $H$ (an SU(2) doublet) with a potential

$$V = m^2 H^\dagger H + \frac{\lambda}{2}(H^\dagger H)^2 \qquad [1]$$

where $m$ and $\lambda$ are free parameters of the SM. When $m^2 < 0$ is chosen, the minimum of the potential occurs when

$$\langle H^\dagger H\rangle = -\frac{m^2}{\lambda} \equiv \frac{v^2}{2} \qquad [2]$$

that is, the Higgs doublet acquires $\mathrm{SU}(2) \times \mathrm{U}(1)_Y$ breaking vacuum expectation value $v$ which is just the Fermi scale. The masses of the intermediate vector bosons $W^\pm$ and $Z^0$ are proportional to $v$ and depend also on the gauge couplings. Within the SM understood as a theory with the momentum cut-off $\Lambda_{\mathrm{SM}}$, quantum corrections to the mass parameter $m^2$ in eqn [1] are quadratically divergent:

$$\delta m^2 = \frac{3}{64\pi^2}\left(3g_2^2 + g_1^2 + \lambda - 8y_t^2\right)\Lambda_{\mathrm{SM}}^2 + \cdots \qquad [3]$$

Here, $g_1, g_2$, and $y_t$ are the gauge couplings of the groups $\mathrm{U}(1)_Y, \mathrm{SU}(2)_L$, and the top-quark Yukawa coupling, respectively. This means that if, above the energy scale $\Lambda_{\mathrm{SM}}$, the SM is replaced by some more fundamental theory, in which there are particles of masses $M \gtrsim \Lambda_{\mathrm{SM}}$, the quantum corrections to $m^2$ are quadratically dependent on the new mass scale $M$. For $M \gg v$, this is very unnatural even if the original parameter $m^2$ remains a free parameter of this underlying theory and particularly difficult to accept if in the underlying theory $m^2$ is fixed by some more fundamental considerations. If the SM was the correct theory up to, for example, the mass scale suggested by the see-saw mechanism for the neutrino masses, $\Lambda_{\mathrm{SM}} \sim 10^{15}$ GeV

$$|\delta m^2| \sim 10^{28}\,\mathrm{GeV}^2 \sim 10^{24} v^2!$$

Clearly, this excludes the possibility of understanding the magnitude of the Fermi scale $v$ in any sensible way. Thus, for naturalness of the Higgs mechanism in the SM there should exist a new mass scale $M \gtrsim v$, say only one order of magnitude higher than $v$ and the theory describing the physics above that scale should be free of quadratic divergences. (Approximate) supersymmetry is at present the most elegant and theoretically most complete solution to the hierarchy problem of the SM.

## Supersymmetric Extensions of the SM

In supersymmetry, the gauge fields $A_\mu^a$ are promoted to vector superfields $\hat{V}^a = (A_\mu^a, \lambda^a, D^a)$, one for each gauge symmetry group generator, where $\lambda^a$'s are Weyl fermions (called gauginos) and $D^a$'s are nondynamical auxiliary fields. A renormalizable supersymmetric gauge theory is completely defined (see, e.g., Sohnius (1985) and Wess and Bagger (1992)) by specifying the gauge group, the set of chiral supermultiplets $\hat{\Phi}_i = (\phi_i, \psi_i, F_i)$ representing matter fields, and the superpotential – a holomorphic polynomial function of at most third order in the chiral superfields which determines Yukawa couplings of the fermions $\psi_i$ and scalars $\phi_i$. Auxiliary fields $D^a$ and $F_i$ can be eliminated via their (algebraic) equations of motion.

The so-called minimal supersymmetric SM (MSSM) encodes the main features of any supersymmetric extension of the SM. Its gauge group is

$SU(3) \times SU(2) \times U(1)$ – the same as in the SM – and the chiral superfields are associated to each of the SM quark and lepton fields. Thus, quarks and leptons get scalar spin zero superpartners, the squarks and sleptons, carrying the same quantum numbers as their corresponding fermions and the vector superfields provide spin 1/2 superpartners for the gauge fields – the gluinos, the winos, and the bino. The SM Higgs doublet with weak hypercharge $Y = 1/2$ becomes a scalar component of a chiral superfield $\hat{H}_u$ which contains in addition one doublet of Weyl fermions – the Higgsinos. The chiral anomaly cancelation condition requires that there be also a second Higgs chiral superfield $\hat{H}_d$ with $Y = -1/2$. Such a superfield is also required for giving masses to all flavors of quarks; because of the holomorphicity of the superpotential the same Higgs doublet cannot couple simultaneously to all quarks.

With the MSSM superfield content, the most general renormalizable superpotential consistent with the gauge symmetry has the form

$$W = Y_u \hat{U}^c \hat{Q} \hat{H}_u + Y_d \hat{D}^c \hat{Q} \hat{H}_d + Y_l \hat{E}^c \hat{L} \hat{H}_d + \mu \hat{H}_d \hat{H}_u + \lambda_1 \hat{D}^c \hat{Q} \hat{L} + \lambda_2 \hat{E}^c \hat{L} \hat{L} + \lambda_3 \hat{U}^c \hat{D}^c \hat{D}^c + \lambda_4 \hat{L} \hat{H}_u \quad [4]$$

(flavor indices are suppressed) where the superfield $\hat{Q}$ contains the SU(2) quark doublet $Q$ and its scalar superpartner $\tilde{Q}$ and similarly for the lepton doublet $\hat{L}$, quark singlets $\hat{U}, \hat{D}$, and lepton singlet $\hat{E}$ superfields. The three first terms in [4] give the SM-like Yukawa couplings of quarks and leptons to the Higgs fields together with Yukawa couplings of the corresponding superpartners. The fourth term has no SM analogy; it gives supersymmetric masses to the Higgs scalar and Higgsinos. The interactions in the second line do not conserve baryon and lepton numbers, respectively $B$ and $L$, and should be forbidden (or strongly suppressed) by some additional symmetry of the theory as they would lead to rapid proton decay. A discrete symmetry, called $R$-parity $R = (-1)^{2S+3(B-L)}$, where $S$ is the spin of the field, is an interesting possibility. $R$-parity acts differently on the different components of the superfields: it is even for all SM particles and odd for their superpartners. Its conservation implies that superpartners must appear in pairs in any interaction vertex. Thus, with $R$-parity imposed, the lightest supersymmetric particle is stable and it is an excellent candidate for the dark matter in the universe.

Supersymmetry cannot be an exact symmetry of nature because there do not exist elementary fermions and bosons degenerate in mass. The superpotential [4] does not break supersymmetry spontaneously but even if it did the elementary fermions and bosons would on average have equal masses (they would satisfy some mass sum rule) which is also contradicted by the experimental data. Therefore, in the MSSM, supersymmetry has to be broken explicitly but in such a way that the soft ultraviolet behavior remains intact. Remarkably, the supersymmetry breaking terms which can be added to the MSSM Lagrangian without reintroducing quadratic divergences make heavy just those fields which are opposite statistics superpartners of the SM gauge bosons and fermions. These so-called soft terms are:

$$\begin{aligned}\mathcal{L}_{\text{soft}} = &-\tfrac{1}{2}\tilde{G}\tilde{G}^a\tilde{G}^a - \tfrac{1}{2}\tilde{W}\tilde{W}^a\tilde{W}^a - \tfrac{1}{2}\tilde{B}\tilde{B}\tilde{B} \\ &- m_Q^2|\tilde{Q}|^2 - m_U^2|\tilde{U}^c|^2 - m_D^2|\tilde{D}^c|^2 \\ &- m_L^2|\tilde{L}|^2 - m_E^2|\tilde{E}^c|^2 - m_{H_u}^2|H^u|^2 \\ &- m_{H_d}^2|H^d|^2 - m_3^2(H^u H^d + \text{c.c.}) \\ &+ A_U \tilde{U}^c \tilde{Q} H_u + A_D \tilde{D}^c \tilde{Q} H_d + A_E \tilde{E}^c \tilde{L} H_d \qquad [5]\end{aligned}$$

and yield gaugino (gluino $\tilde{G}$, wino $\tilde{W}$, and bino $\tilde{B}$) and scalar mass terms as well as explicit trilinear couplings between scalars (scalar mass terms and $A$-terms are $3 \times 3$ matrices in the flavor space). As a result, supersymmetry is broken in the mass spectra but not in the dimensionless couplings.

The origin of the soft supersymmetry breaking remains an open issue. Terms [5] are most probably remnants of the spontaneous supersymmetry breaking in the so-called "hidden" sector – a hypothetical set of fields that do not interact directly with the MSSM fields. For example, in the popular scenario, they interact with the MSSM fields only gravitationally and spontaneous supersymmetry breaking in the hidden sector is communicated to the MSSM sector by gravitational interactions giving rise to terms [5]. Several other mechanisms of supersymmetry breaking transmission have also been proposed (gauge mediation, anomaly mediation, etc.).

The mass parameters and $A$-terms in [5] are free parameters of the low-energy supersymmetric theory and, combined with the interactions like $Q\tilde{Q}\tilde{G}$ originating from supersymmetric kinetic terms, may be a new, troublesome, source of flavor changing neutral currents and of $CP$ violation.

## Higgs Sector of the MSSM

The MSSM Higgs potential reads

$$V = m_1^2|H_d|^2 + m_2^2|H_u|^2 + m_3^2(H_u H_d + \text{c.c.}) + \frac{g_1^2 + g_2^2}{8}\left(|H_d|^2 - |H_u|^2\right)^2 \qquad [6]$$

Its quartic part is uniquely determined by the structure of the supersymmetric gauge theory. The parameters $m_1^2$, $m_2^2$, and $m_3^2$ are determined by

the soft supersymmetry breaking Higgs boson masses [5] and the $\mu$ parameter in [4]. The potential [6] is bounded from below for $m_1^2 + m_2^2 > 2m_3^4$, and for $m_1^2 m_2^2 - m_3^2 < 0$ it has the electroweak symmetry breaking minimum at $v_u = \langle H_u^0 \rangle \neq 0, v_u = \langle H_d^0 \rangle \neq 0$. The ratio $v_u/v_d \equiv \tan\beta$ is then phenomenologically a very important parameter.

Quantum corrections to the mass parameters in [6] are controlled by the mass scale $M_{\rm soft}$ of the supersymmetry breaking terms [5]; at the one-loop level instead of [3], one finds

$$\delta m_{1,2}^2 \sim \frac{1}{16\pi^2}\left(3g_2^2 + g_1^2 - 12y_{\rm b,t}^2\right) M_{\rm soft}^2 \ln\frac{\Lambda_{\rm NEW}^2}{M_{\rm soft}^2} \qquad [7]$$

where $y_{\rm b}$ and $y_{\rm t}$ are the bottom- and top-quark Yukawa couplings, respectively and $\Lambda_{\rm NEW}$ is the scale at which the soft supersymmetry breaking terms are generated by the putative supersymmetry breaking transmission mechanism. In gravity mediation scenarios, $\Lambda_{\rm NEW} \sim M_{\rm Pl}$. In gauge mediation scenarios, $\Lambda_{\rm NEW}$ is low but it is a new scale, introduced by hand.

In the softly broken supersymmetric models, the hierarchy problem is solved for $M_{\rm soft} \lesssim \mathcal{O}(10)v$. Moreover, eqn [7] shows that via quantum corrections the large top-quark Yukawa coupling $y_{\rm t}$ drives the mass parameter $m_2^2$ to a negative value, inducing the electroweak symmetry breaking. This means that in supersymmetric models the electroweak scale is calculable in terms of the known coupling constants and the (unknown) scales $M_{\rm soft}$ and cutoff scale $\Lambda_{\rm NEW}$ to the MSSM. If $M_{\rm soft} \lesssim \mathcal{O}(10)v$, the correct electroweak scale is obtained for $\Lambda_{\rm NEW} \sim M_{\rm GUT}$. This nicely fits with unification of the gauge couplings.

In supersymmetric models, the quartic couplings in the Higgs potential are restricted. This typically leads to a strong upper bound on the mass of the lightest Higgs particle. In the minimal model with the potential [6], at the tree level

$$M_{\rm Higgs} < M_Z \approx 91\,{\rm GeV} \qquad [8]$$

This bound is substantially modified by quantum corrections. They depend quadratically on the top-quark mass and logarithmically on the stop mass scale $M_{\tilde{t}} \sim M_{\rm soft}$:

$$M_{\rm Higgs}^2 < \lambda v^2 \qquad [9]$$

where $\lambda$ is given by

$$\lambda = \tfrac{1}{8}\left(g_2^2 + g_1^2\right)\cos^2 2\beta + \Delta\lambda$$

$$\text{with } \Delta\lambda = \frac{3g_2^2}{8\pi^2}\frac{m_t^4}{v^2 M_W^2}\ln\frac{M_{\tilde{t}}^2}{m_t^2} \qquad [10]$$

For $M_{\tilde{t}} < 1\,{\rm TeV}, M_{\rm Higgs} < 130\,{\rm GeV}$.

The minimal-model bound on the Higgs mass can be relaxed in models with extended Higgs sector. For instance, if an additional gauge group singlet chiral superfield couples to the Higgs doublets, the Higgs self-coupling $\lambda$ in [9] receives additional contributions. Explicit calculations show that in such and other models, with $M_{\rm soft} \lesssim 1\,{\rm TeV}$, the bound on the Higgs mass cannot be raised above $\sim$150 GeV if one wants to preserve perturbative gauge coupling unification.

## Supersymmetric Grand Unified Theories

There are two striking aspects of the matter spectrum in the SM. One is the chiral anomalies cancelation (Weinberg 1996–2000, Pokorski 2000), which is necessary for a unitary (and renormalizable) theory, and occurs thanks to certain conspiracy between quarks and leptons suggesting a deeper link between them. The second one is that the spectrum fits into simple representations of the SU(5) and SO(10) groups (Ross 1985). Indeed, each generation of the SM matter fills $\mathbf{5}^* + \mathbf{10} + \mathbf{1}$ (if the right-handed neutrino is included into the spectrum) representations of SU(5) and for SO(10), $\mathbf{16} = \mathbf{5}^* + \mathbf{10} + \mathbf{1}$. The assignment of fermions to the SU(5) or SO(10) representations fixes the normalization of the $U(1)_Y$ generator. Both facts suggest unification of strong and electroweak elementary forces in a grand unified theory with some bigger gauge symmetry group. Such unification implies that all the SM gauge forces become of equal strength at some unification scale. Their strength is measured by the running gauge couplings $\alpha_i = g_i^2/4\pi$, $i = 1, 2, 3$, of the three group factors $SU(3)_c \times SU(2)_L \times U(1)_Y$. The energy scale dependence of $\alpha_i$ is governed by the renormalization group equations. In the first nontrivial approximation, they read:

$$\frac{1}{\alpha_i(Q)} = \frac{1}{\alpha_i(M_Z)} - \frac{b_0^{(i)}}{2\pi}\ln\left(\frac{Q}{M_Z}\right) \qquad [11]$$

Here, $1/\alpha_i(M_Z) = (58.98 \pm 0.04, 29.57 \pm 0.03, 8.40 \pm 0.14)$ are the experimental values of the gauge couplings at the Fermi scale and $b_0^{(i)}$ are the coefficients which depend on the matter content of the theory. They are

$$b_0 = \left(\tfrac{1}{10} + \tfrac{4}{3}N_g, -\tfrac{43}{6} + \tfrac{4}{3}N_g, -11 + \tfrac{4}{3}N_g\right)$$

in the SM and

$$b_0 = \left(\tfrac{3}{5} + 2N_g, -5 + 2N_g, -9 + 2N_g\right)$$

in the MSSM, where $N_g$ is the number of fermion generations. In the SM, the running gauge couplings

approach each other at high scale of order $10^{13}$ GeV but never unify.

In the MSSM, with sparticle spectrum characterized by $M_{\text{soft}} \approx 1$ TeV and for the initial Fermi scale values given above, the three gauge couplings unify with high precision at the scale $M_{\text{GUT}} \sim 10^{16}$ GeV. Therefore, the MSSM can be embedded into supersymmetric grand unified theories with no hierarchy problem for the Fermi scale (it is stable with respect to radiative corrections generated by particles with masses $\sim M_{\text{GUT}}$) and no conflict with the measured values of the gauge couplings.

In the SM, the baryon number is (perturbatively) conserved since there are no renormalizable couplings violating this symmetry. Experimental search for proton decay, for example, $p \to e^+\pi^0$, $p \to K^+\nu$, is one of the most fundamental tests for particle physics. The present limit on the proton life time is $\tau_p > 10^{33}$ yr. In grand unified theories, baryon number conservation is violated by interactions mediated by the heavy gauge bosons corresponding to the enlarged gauge symmetry (e.g., SU(5)), spontaneously broken at $M_{\text{GUT}}$ to the SM gauge symmetry. Such interactions manifest themselves at low energy as additional, nonrenormalizable interactions added to the SM Lagrangian. Proton decay is then induced by the set of dimension-6 operators of the form

$$\mathrm{O}_i^{(6)} = \frac{c_i^{(6)}}{M_{(6)}^2} qqql \qquad [12]$$

where $q, l$ denote quarks and leptons, respectively. For $c_i^{(6)} \sim \alpha_{\text{GUT}} \approx 1/25$, the experimental limit on $\tau_p$ requires $M_{(6)} \gtrsim 10^{15}$ GeV, consistently with $M_{\text{GUT}} = 10^{16}$ GeV in supersymmetric GUTs. However, in supersymmetric GUTs, there is still another, genuinely supersymmetric, source of contributions to the proton decay amplitudes. These are the dimension-5 operators

$$\mathrm{O}_i^{(5)} = \frac{c_i^{(5)}}{M_{(5)}^2} qq\tilde{q}\tilde{l} \qquad [13]$$

where $\tilde{q}, \tilde{l}$ denote squarks and sleptons, respectively. Such operators originate from the exchange of the color triplet scalars present in the Higgs boson GUT multiplets, with $M_{(5)} \sim M_{\text{GUT}} \sim 10^{16}$ GeV, and $c^{(5)} \gtrsim 10^{-7}$ is given by the Yukawa couplings. Inserted into diagrams with gaugino exchanges they give rise to dimension-6 operators of the form [12]. One then gets $c^{(6)} = \alpha_{\text{GUT}} c^{(5)}$, $M_{(6)}^2 = M_{(5)} M_{\text{SUSY}}$. Given various uncertainties, for example, in the unknown squark, gaugino, and heavy Higgs boson mass spectrum, such contributions in supersymmetric GUT models predict the proton life time to be consistent with but close to the present experimental limits.

## Summary

Supersymmetry is distinct in several very important points from all other proposed solutions to the hierarchy problem. First of all, it provides a general theoretical framework which allows one to address many physical questions. Supersymmetric models, like the MSSM or its simple extensions, satisfy a very important criterion of "perturbative calculability." In particular, they are easily consistent with the precision electroweak data. The SM is their low-energy approximation in the sense of the Appelquist–Carazzone decoupling, so most of the successful structure of the SM is built into supersymmetric models. The quadratically divergent quantum corrections to the Higgs mass parameter (the origin of the hierarchy problem in the SM) are absent in any order of perturbation theory. Therefore, the cutoff to a supersymmetric theory can be as high as the Planck scale, and "small" scale of the electroweak breaking is still natural. Supersymmetry is not only consistent with grand unification of elementary forces but, in fact, makes it very successful. And, finally, supersymmetry is needed for string theory.

However, there are also some problems to be solved: the hierarchy problem of the electroweak scale is solved but the origin of the soft supersymmetry breaking scale $M_{\text{soft}}$ remains an open question: spontaneous supersymmetry breaking and its transmission to the visible sector is a difficult problem and a fully satisfactory mechanism which would yield $M_{\text{soft}}$ hierarchically smaller than the Planck (string) scale has not yet been found. On the phenomenological side, there are new potential sources of flavor-changing neutral current transitions and of CP violation, and baryon and lepton numbers are not automatically conserved by the renormalizable couplings. But even those problems can at least be discussed in a concrete quantitative way.

*See also:* Brane Construction of Gauge Theories; Perturbation Theory and its Techniques; Seiberg–Witten Theory; Standard Model of Particle Physics; Supergravity; Supermanifolds.

## Further Reading

Eidelman S *et al.* (The Particle Data Group) (2004) Review of particle physics. *Physics Letters B* 592: 1.

Kane GL (ed.) (1998a) *Perspectives in Supersymmetry*. Singapore: World Scientific.

Kane GL (ed.) (1998b) *Perspectives on Higgs Physics II*. Singapore: World Scientific.

Nilles H-P (1984) *Physics Reports C* 110: 1.

Pokorski S (2000) *Gauge Field Theories*, Cambridge Monographs on Mathematical Physics, 2nd edn. Cambridge: Cambridge University Press.
Ross GG (1985) *Grand Unified Theories*. Redwood City, CA: Addison-Wesley.
Sohnius M-F (1985) *Physics Reports C* 128: 39.
Weinberg S (1996–2000) *The Quantum Theory of Fields*, vols. I–III. Cambridge: Cambridge University Press.
Wess J and Bagger J (1992) *Supersymmetry and Supergravity*, Princeton Series in Physics, 2nd edn. Princeton: Princeton University Press.

# Supersymmetric Quantum Mechanics

**J-W van Holten**, NIKHEF, Amsterdam, The Netherlands



## Introduction

Supersymmetric quantum mechanics is a specific extension of quantum mechanics with fermionic degrees of freedom. In quantum field theory and many-body theory, a fermionic degree of freedom is one which is subject to Pauli's principle: any nondegenerate quantum state associated with a fermionic degree of freedom can be occupied at most once at any time. Similarly, in quantum mechanics, one associates a fermionic degree of freedom with an observable, the eigenvalue spectrum of which is restricted to the discrete set (0, 1).

The simplest example of a purely fermionic quantum system is the fermionic oscillator. It is represented by conjugate operators $(f, f^\dagger)$ such that

$$f^2 = 0, \quad f^{\dagger 2} = 0, \quad ff^\dagger + f^\dagger f = 1 \qquad [1]$$

with a Hamiltonian $H$ given by the bilinear expression

$$H_f = \varepsilon_f + \hbar\omega f^\dagger f \qquad [2]$$

The state space of this system is spanned by two independent state vectors $|0\rangle$ and $|1\rangle$, such that

$$\begin{aligned} f|0\rangle &= 0, \qquad f^\dagger|0\rangle = |1\rangle \\ f|1\rangle &= |0\rangle, \qquad f^\dagger|1\rangle = 0 \end{aligned} \qquad [3]$$

By construction, the states $|n_f\rangle$ are eigenstates of fermion number,

$$N_f = f^\dagger f, \qquad N_f^2 = N_f \qquad [4]$$

with eigenvalue $n_f = (0, 1)$; this implements the Pauli principle. The states have energy eigenvalues

$$E_{n_f} = \varepsilon_f + n_f\hbar\omega, \quad n_f = (0, 1) \qquad [5]$$

differing in energy by $\Delta E = \hbar\omega$. Physically, the system can be identified with a single fixed magnetic dipole in an external magnetic field, the only polarization states of the dipole being spin up or spin down.

In the Schrödinger representation of quantum mechanics (wave mechanics), fermionic degrees of freedom are represented by anticommuting Grassmann variables. These have no immediate classical analog, but can be used to construct quasiclassical observables like spin.

A supersymmetric quantum system is a system possessing both fermionic and bosonic degrees of freedom, characterized by a degeneracy between states with even and odd fermion number. In the Schrödinger representation, this is manifest in a symmetry transforming bosonic (Grassmann-even) into fermionic (Grassmann-odd) variables. The generators of the supersymmetry transformations square to the Hamiltonian of the system.

## The Supersymmetric Oscillator

An elementary example of a supersymmetric quantum system is the supersymmetric oscillator. It is a physical system combining a standard bosonic quantum oscillator with a fermionic oscillator of the same frequency. The ordinary harmonic oscillator is described by the pair of lowering and raising operators $(b, b^\dagger)$, with commutator

$$bb^\dagger - b^\dagger b = 1 \qquad [6]$$

and the Hamiltonian

$$H_b = \varepsilon_b + \hbar\omega b^\dagger b \qquad [7]$$

In this case, the eigenvalue spectrum of the occupation number

$$N_b = b^\dagger b \qquad [8]$$

consists of all non-negative integers $n_b = 0, 1, 2, \ldots$, with corresponding energy eigenvalues. To construct the supersymmetric oscillator, the harmonic oscillator is combined with a fermionic oscillator [2] of the same frequency:

$$H_s = \varepsilon_0 + \hbar\omega\left(b^\dagger b + f^\dagger f\right) \qquad [9]$$

where $\varepsilon_0 = \varepsilon_b + \varepsilon_f$. The ground state of this system is the state annihilated by both $b$ and $f$:

$$b|0,0\rangle = f|0,0\rangle = 0 \qquad [10]$$

The full set of energy eigenstates of the system is constructed by taking

$$|n_b, n_f\rangle = \frac{1}{\sqrt{n_b!}} b^{\dagger n_b} f^{\dagger n_f} |0,0\rangle$$
$$n_b = (0,1,2,\ldots), \qquad n_f = (0,1) \qquad [11]$$

with the energy eigenvalue spectrum

$$E(n_b, n_f) = \varepsilon_0 + n\hbar\omega, \quad n = n_b + n_f \qquad [12]$$

Clearly, there is a degeneracy in energy between the states $|n_b+1,0\rangle$ and $|n_b,1\rangle$, which have the same total occupation number $n$, but differ in the bosonic and fermionic occupation number by one unit. This is illustrated in Figure 1. Such pairs of states which are degenerate in energy can be transformed into each other by the operators

$$Q = \sqrt{2\hbar\omega}\, b^\dagger f, \qquad Q^\dagger = \sqrt{2\hbar\omega}\, f^\dagger b \qquad [13]$$

The explicit transformations are

$$|n_b+1,0\rangle = \frac{1}{\sqrt{2(n_b+1)\hbar\omega}} Q|n_b,1\rangle$$
$$|n_b,1\rangle = \frac{1}{\sqrt{2(n_b+1)\hbar\omega}} Q^\dagger|n_b+1,0\rangle \qquad [14]$$

The operations [14] are called supersymmetry transformations, and the operators $Q$ and $Q^\dagger$ are called supercharges.

As the zero point of energy is arbitrary in systems without gravitational interactions, it is customary to take $\varepsilon_0 = 0$, that is, $\varepsilon_f = -\varepsilon_b$; with the normalization [13], the Hamiltonian $H$ is then the symmetrized absolute square of the supercharges:

$$QQ^\dagger + Q^\dagger Q = 2H \qquad [15]$$

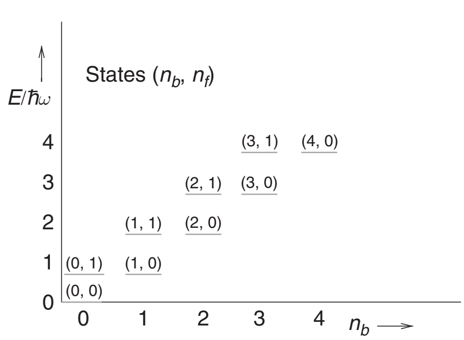



**Figure 1** Spectrum of states of the supersymmetric oscillator.

whilst

$$Q^2 = Q^{\dagger 2} = 0 \qquad [16]$$

The above relations suffice to guarantee that the supercharges $(Q, Q^\dagger)$ are conserved:

$$[Q,H] = [Q^\dagger, H] = 0 \qquad [17]$$

a result re-expressing the degeneracy between states with the same $n$ but different $n_b$ and $n_f$. The real form of the supercharges is

$$Q_1 = \frac{1}{2}(Q + Q^\dagger), \qquad Q_2 = \frac{1}{2\mathrm{i}}(Q - Q^\dagger) \qquad [18]$$

In this representation

$$H = Q_1^2 + Q_2^2 \qquad [19]$$

An important observation is that the ground state is the only state annihilated by both supersymmetry operators:

$$Q|0,0\rangle = 0, \qquad Q^\dagger|0,0\rangle = 0 \qquad [20]$$

Indeed, it is the only state with zero energy eigenvalue, and only such a state can be an invariant supersinglet; all other states have positive energy and they necessarily occur in supersymmetry pairs.

## Anticommuting Variables

Fermionic degrees of freedom can be described in a pseudoclassical formulation by anticommuting variables $\xi$ taking values in an infinite-dimensional Grassmann algebra:

$$\xi\xi' + \xi'\xi = 0 \qquad [21]$$

With an anticommuting variable $\xi$, we can associate a derivative operator $\partial/\partial\xi$, which is an element of another Grassmann algebra such that

$$\left\{\frac{\partial}{\partial\xi}, \xi\right\} = \frac{\partial}{\partial\xi}\xi + \xi\frac{\partial}{\partial\xi} = 1, \quad \frac{\partial^2}{\partial\xi^2} = 0 \qquad [22]$$

This extends the original Grassmann algebra to a Clifford algebra. Integration with respect to an anticommuting variable is defined in the same way:

$$\int \mathrm{d}\xi \cdot \xi = 1, \quad \int \mathrm{d}\xi \cdot 1 = 0 \qquad [23]$$

that is, integration is the same as differentiation for anticommuting variables. With these definitions, we can represent the fermionic raising and lowering operators in terms of anticommuting variables as

$$f^\dagger \to \xi, \qquad f \to \frac{\partial}{\partial\xi} \qquad [24]$$

and the states by

$$|0\rangle \rightarrow 1, \qquad |1\rangle \rightarrow \xi \qquad [25]$$

Then an arbitrary state takes the form of a linear superposition

$$|\Psi\rangle = \psi_0|0\rangle + \psi_1|1\rangle \;\rightarrow\; \Psi(\xi) = \psi_0 + \psi_1\xi \qquad [26]$$

and the standard positive-semidefinite inner product on the state space is represented on the wave functions by the double integral

$$\langle\Phi|\Psi\rangle = \int \mathrm{d}\xi\,\mathrm{d}\bar{\xi}\,\mathrm{e}^{\bar{\xi}\xi}\,\Phi^*(\bar{\xi})\Psi(\xi) = \phi_0^*\psi_0 + \phi_1^*\psi_1 \quad [27]$$

By construction, $f^\dagger = \xi$ and $f = \partial/\partial\xi$ are conjugates with respect to this inner product:

$$\int \mathrm{d}\xi\,\mathrm{d}\bar{\xi}\,\mathrm{e}^{\bar{\xi}\xi}\Phi^*(\bar{\xi})\xi\Psi(\xi) = \int \mathrm{d}\bar{\xi}\,\mathrm{e}^{\bar{\xi}\xi}\left(\frac{\partial\Phi}{\partial\xi}\right)^*(\bar{\xi})\Psi(\xi) \quad [28]$$

The real (self-conjugate) forms of the fermion operators are, therefore, defined by

$$\sigma_1 = \left(\xi + \frac{\partial}{\partial\xi}\right), \qquad \sigma_2 = \mathrm{i}\left(\xi - \frac{\partial}{\partial\xi}\right) \qquad [29]$$

which satisfy the Pauli–Dirac anticommutation relations

$$\sigma_i\sigma_j + \sigma_j\sigma_i = 2\delta_{ij} \qquad [30]$$

By taking the product, we obtain

$$\sigma_3 = -\mathrm{i}\sigma_1\sigma_2 = 1 - 2\xi\frac{\partial}{\partial\xi} = 1 - 2N_f$$
$$\Leftrightarrow\; N_f = \frac{1}{2}(1 - \sigma_3) \qquad [31]$$

Thus, we may think of the wave functions as two-component spinors, the components being labeled either by the eigenvalues of the spin operator $\sigma_3$, or equivalently by the fermion number $N_f$, which is a projection operator on the states with negative spin.

The action of the Hamiltonian on a wave function $\Psi(\xi)$ is represented by the integral

$$[H\Psi](\xi) = \int \mathrm{d}\xi'\,\mathrm{d}\bar{\xi}\,\mathrm{e}^{\bar{\xi}(\xi'-\xi)}H(\xi,\bar{\xi})\Psi(\xi') \qquad [32]$$

where $H(\xi,\bar{\xi})$ is the ordered symbol of the Hamiltonian:

$$H(\xi,\bar{\xi}) = \varepsilon_f + \hbar\omega\xi\bar{\xi} \qquad [33]$$

This expression is to be considered as the classical Hamiltonian of the system. In particular, the exponent of the action

$$S = \int_1^2 \mathrm{d}t\big(\mathrm{i}\hbar\bar{\xi}\dot{\xi} - H(\xi,\bar{\xi})\big)$$
$$= \hbar\int_1^2 \mathrm{d}t\big(\mathrm{i}\bar{\xi}\dot{\xi} + \omega\bar{\xi}\xi\big) + \varepsilon_f(t_2 - t_1) \qquad [34]$$

provides the integrand for the path-integral representation of the evolution operator in the quantum theory. The proof is not given here; the reader is referred to the literature. In passing, note that as the anticommuting variables $(\xi,\bar{\xi})$ are taken to be dimentionless, one actually should identify the momentum conjugate to $\xi$ with $\pi = -\mathrm{i}\hbar\bar{\xi}$; in the quantum theory, this is replaced by the operator $-\mathrm{i}\hbar\partial/\partial\xi$.

## Classical Supersymmetry

The classical action for the supersymmetric oscillator with bosonic amplitude $x$ and fermionic amplitude $\xi$ is

$$S = \int_1^2 \mathrm{d}t\left(\frac{1}{2}\dot{x}^2 - \frac{\omega}{2}x^2 + \mathrm{i}\bar{\xi}\dot{\xi} + \omega\bar{\xi}\xi\right) \qquad [35]$$

As inferred from the quantum theory, it is a combination of a linear harmonic oscillator and a fermionic oscillator of the same frequency. A factor $\sqrt{\hbar}$ is also absorbed in $\xi$ and $\bar{\xi}$; equivalently, we can use natural units in which $\hbar = 1$. In the following, we use this convention.

The action [35] is invariant under infinitesimal symmetry transformations

$$\delta x = -\mathrm{i}\big(\bar{\epsilon}\xi + \epsilon\bar{\xi}\big)$$
$$\delta\xi = (\dot{x} + \mathrm{i}\omega x)\epsilon, \qquad \delta\bar{\xi} = (\dot{x} - \mathrm{i}\omega x)\bar{\epsilon} \qquad [36]$$

with $(\bar{\epsilon},\epsilon)$ Grassmann-odd parameters. The Noether theorem then implies that there are conserved fermionic charges

$$Q = (p - \mathrm{i}\omega x)\xi, \qquad \bar{Q} = (p + \mathrm{i}\omega x)\bar{\xi} \qquad [37]$$

with the momentum defined by $p = \dot{x}$. The other conserved quantity is the energy, represented by the Hamiltonian

$$H = \frac{1}{2}\left(p^2 + \omega^2x^2\right) + \omega\xi\bar{\xi} \qquad [38]$$

The canonical phase-space formulation is obtained by defining brackets of two functions $(A, B)$ on the phase space $(x, p\,; \xi, \bar{\xi})$ by

$$\{A,B\} = \frac{\partial A}{\partial x}\frac{\partial B}{\partial p} - \frac{\partial A}{\partial p}\frac{\partial B}{\partial x}$$
$$+\,\mathrm{i}(-1)^A\left(\frac{\partial A}{\partial\xi}\frac{\partial B}{\partial\bar{\xi}} + \frac{\partial A}{\partial\bar{\xi}}\frac{\partial B}{\partial\xi}\right) \qquad [39]$$

where $(-1)^A$ is the Grassmann parity of $A$. In terms of these brackets, the time evolution and supersymmetry transformations take the form

$$\dot{A} = -\{H, A\}, \qquad \delta A = \mathrm{i}\big\{\bar{\epsilon}Q + \epsilon\bar{Q}, A\big\} \qquad [40]$$

Moreover, the charges $Q$ and $\bar{Q}$ satisfy the bracket algebra

$$\{Q, \bar{Q}\} = -2\mathrm{i}H, \qquad \{Q, H\} = \{\bar{Q}, H\} = 0 \quad [41]$$

Thus, the action [35] is the classical counterpart of the quantum theory [9]–[17] in the correspondence limit $\mathrm{i}\{A, B\} \to [A, B]_{\pm} = AB \pm BA$. For these theories, supersymmetry is rooted in the classical transformations [36].

## Supersymmetric Quantum Mechanics

The construction for the supersymmetric oscillator can be generalized to other dynamical systems in two ways. First, the nature of the interactions as represented by the potential can be modified. Second, the number of degrees of freedom can be varied. This section presents a generalization of the supersymmetric oscillator to anharmonic interactions, obtained by modification of the supercharges [37] with a general function $\Phi(x)$ as follows:

$$Q = (p - \mathrm{i}\Phi(x))\xi, \qquad \bar{Q} = (p + \mathrm{i}\Phi(x))\bar{\xi} \quad [42]$$

The brackets [39] imply the supersymmetry algebra [41] with the Hamiltonian

$$\begin{aligned} H &= \frac{\mathrm{i}}{2}\{Q, \bar{Q}\} \\ &= \frac{1}{2}p^2 + \frac{1}{2}\Phi^2(x) + \frac{1}{2}\Phi'(x)(\xi\bar{\xi} - \bar{\xi}\xi) \end{aligned} \quad [43]$$

In quantum mechanics, the supercharges become operators $Q$ and $Q^\dagger$ upon reinterpretation of $(x, p)$ as canonically conjugate operators, and the replacement $\xi \to f^\dagger$ and $\bar{\xi} \to f$; this procedure involves no ordering ambiguity. The Hamiltonian operator defined by the anticommutator of $Q$ and $Q^\dagger$ then takes the operator form associated with [43]. With the identification

$$A = \frac{1}{\sqrt{2}}(p - \mathrm{i}\Phi), \qquad A^\dagger = \frac{1}{\sqrt{2}}(p + \mathrm{i}\Phi) \quad [44]$$

and making use of the (anti)commutation relations

$$AA^\dagger - A^\dagger A = \Phi'(x), \qquad ff^\dagger + f^\dagger f = 1 \quad [45]$$

this Hamilton operator can be written in normal-ordered form as

$$H = \tfrac{1}{2}(QQ^\dagger + Q^\dagger Q) = A^\dagger A + \Phi'(x)f^\dagger f \quad [46]$$

It is positive-semidefinite by construction. All results for the supersymmetric oscillator are reproduced upon taking $\Phi(x) = \omega x$.

As the Hamiltonian commutes with the fermion number operator $N_f$, we can label all stationary states $|E, n_f\rangle$ by the energy $E$ and the fermion number $n_f = (0, 1)$. Moreover, all states of positive energy are degenerate with respect to fermion number, as they form pairs related by supersymmetry:

$$Q|E, 0\rangle = \sqrt{2E}\,|E, 1\rangle, \quad \bar{Q}|E, 1\rangle = \sqrt{2E}\,|E, 0\rangle \quad [47]$$

Only ground states with $E_0 = 0$ can occur as singlets under supersymmetry. The existence of such a ground state with fermion number $n_f$ amounts to the existence of a state $|0, n_f\rangle$ satisfying

$$A^\dagger f|0, n_f\rangle = Af^\dagger|0, n_f\rangle = 0 \quad [48]$$

The corresponding wave functions are of the form

$$\begin{aligned} |0, 0\rangle &\to \Psi_0(x, \xi) = \psi_-(x) \\ |0, 1\rangle &\to \Psi_1(x, \xi) = \psi_+(x)\xi \end{aligned} \quad [49]$$

where $\psi_\pm(x)$ are solutions of the equations

$$A\psi_- = 0, \qquad A^\dagger\psi_+ = 0 \quad [50]$$

These functions are formally given by the expressions

$$\psi_\pm(x) = C_\pm \mathrm{e}^{\pm\int_0^x \Phi(y)\,\mathrm{d}y} \quad [51]$$

For a zero-energy ground state to exist, one of these functions must be normalizable. For example, if $\Phi(x)$ is a polynomial of positive odd degree $2k - 1$, then, depending on the sign of the coefficient of $x^{2k-1}$, one of the exponents is bounded, approaching zero for $x \to \pm\infty$, and as a result becomes square integrable.

If no normalizable wave functions of the form [51] exist, the ground state cannot have zero energy ($E_0 > 0$) and all states necessarily belong to superdoublets.

## Spinning-Particle Mechanics

Minimal supersymmetric classical or quantum mechanics requires equal number of bosonic and fermionic coordinates in configuration space $(x_i, \xi_i)$, rather than equal number of bosonic and fermionic degrees of freedom in phase space. Specifically, minimal free supersymmetric particle mechanics in $n$ dimensions is described by the classical Lagrangian

$$L = \frac{1}{2}\dot{x}_i^2 + \frac{i}{2}\xi_i\dot{\xi}_i, \quad i = 1, \ldots, n \quad [52]$$

It is invariant modulo a total time derivative under infinitesimal supersymmetry transformations

$$\delta x_i = -\mathrm{i}\epsilon\xi_i, \qquad \delta\xi_i = \dot{x}_i\epsilon \quad [53]$$

The canonical phase-space formulation is phrased in terms of the free-particle momentum and Hamiltonian

$$p_i = \dot{x}_i, \qquad H = \tfrac{1}{2}p_i^2 \tag{54}$$

and the brackets

$$\{A, B\} = \frac{\partial A}{\partial x_i}\frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i}\frac{\partial B}{\partial x_i} + \mathrm{i}(-1)^A \frac{\partial A}{\partial \xi_i}\frac{\partial B}{\partial \xi_i} \tag{55}$$

The supersymmetry transformations are generated by the supercharge

$$Q = p_i \xi_i, \qquad \delta A = \mathrm{i}\epsilon\{Q, A\} \tag{56}$$

with the supersymmetry algebra

$$\mathrm{i}\{Q, Q\} = 2H, \qquad \{Q, H\} = 0 \tag{57}$$

An important quantity in these models is the bilinear (Grassmann-even) antisymmetric tensor

$$\sigma_{ij} = \mathrm{i}\xi_i\xi_j \tag{58}$$

For a free particle, it is a set of constants of motion forming a representation of so($n$), the Lie algebra of $n$-dimensional rotations:

$$\left\{\sigma_{ij}, \sigma_{kl}\right\} = \delta_{jk}\sigma_{il} - \delta_{jl}\sigma_{ik} - \delta_{ik}\sigma_{jl} + \delta_{il}\sigma_{jk} \tag{59}$$

Therefore, the physical interpretation of $\sigma_{ij}$ is that it represents the particle spin. For this reason, supersymmetric particle mechanics is often called spinning-particle mechanics.

Quantum mechanics of the spinning particle has the same algebraic structure, with $(x_i, p_i)$ the standard canonically conjugate operators, and the fermionic coordinates $\xi_i$ represented by the generators of a Clifford algebra; the irreducble representation in terms of Pauli–Dirac matrices of dimension $2^{[n/2]} \times 2^{[n/2]}$ is

$$\xi_i \to \frac{1}{\sqrt{2}}\gamma_i, \quad \gamma_i\gamma_j + \gamma_j\gamma_i = 2\delta_{ij} \tag{60}$$

It follows that the wave functions have $2^{[n/2]}$ components, describing different polarization states. Furthermore, in minimal supersymmetric quantum mechanics, the supersymmetry operator is represented by the Dirac operator:

$$Q \to \frac{1}{\sqrt{2}}\gamma \cdot p, \quad (\gamma \cdot p)^2 = p_i^2 = 2H \tag{61}$$

Hence, the stationary states of the system solve the Dirac equation

$$\gamma \cdot p\Psi = \sqrt{2E}\Psi \tag{62}$$

The models can, without difficulty, be extended to include interactions with external fields. As an example, we consider the coupling to a magnetic field described by a vector potential $A_i(x)$. An extension of the free-particle action [52], invariant under the same supersymmetry transformations [53], is

$$S = \int \mathrm{d}t\left(\frac{1}{2}\dot{x}_i^2 + \frac{\mathrm{i}}{2}\xi_i\dot{\xi}_i + qA_i(x)\dot{x}_i - \frac{\mathrm{i}q}{2}F_{ij}(x)\xi_i\xi_j\right) \tag{63}$$

where $F_{ij} = \nabla_i A_j - \nabla_j A_i$ is the field strength. The canonical momentum in this model is

$$p_i = \dot{x}_i + qA_i(x) \tag{64}$$

with the result that the canonical expressions for the Hamiltonian and supercharge become

$$H = \tfrac{1}{2}(p_i - qA_i(x))^2, \qquad Q = (p_i - qA_i(x))\xi_i \tag{65}$$

In the quantum theory, these constants of motion become the covariant Laplacian and Dirac operator in an external vector potential $A_i(x)$. Observe that supersymmetry requires the spin to couple to the magnetic field with gyromagnetic ratio $g = 2$. Explicitly, the equation of motion for $\xi$ can be transformed into an equation for the spin precession:

$$\dot{\xi}_i = qF_{ij}\xi_j \;\Rightarrow\; \dot{\sigma}_{ij} = q(F_{ik}\sigma_{kj} - \sigma_{ik}F_{kj}) \tag{66}$$

In three dimensions, this is equivalent to an equation in terms of axial vectors:

$$F_{ij} = \varepsilon_{ijk}B_k,\ \sigma_{ij} = \varepsilon_{ijk}s_k \;\Rightarrow\; \dot{\mathbf{s}} = -q\boldsymbol{B} \times \mathbf{s} \tag{67}$$

showing that the precession rate of $\mathbf{s}$ is given by twice the Larmor frequency.

## Extended Supersymmetry

It is possible to construct theories with more supersymmetries by associating with every bosonic coordinate several fermionic coordinates. An example is the supersymmetric oscillator and its generalizations considered earlier, which has equal number of bosonic and fermionic degrees of freedom in phase space, rather than equal number of bosonic and fermionic coordinates in configuration space. The classical phase space, spanned by variables $(x_i, p_i; \xi_i, \bar{\xi}_i)$ with $i = 1, \ldots, n$, then has double the number of fermionic variables compared to the minimal supersymmetric particle models. Such models can be constructed for systems with an $n$-dimensional bosonic configuration space. Their supercharges take the form

$$Q = (p_i - \mathrm{i}\Phi_i(\boldsymbol{x}))\xi_i, \quad \bar{Q} = (p_i + \mathrm{i}\Phi_i(\boldsymbol{x}))\bar{\xi}_i \\ \boldsymbol{x} = (x_1, \ldots, x_n) \tag{68}$$

whilst the Hamiltonian becomes

$$H = \tfrac{1}{2}p_i^2 + \tfrac{1}{2}\Phi_i^2(\boldsymbol{x}) + \tfrac{1}{4}(\nabla_j\Phi_i + \nabla_i\Phi_j)(\xi_i\bar{\xi}_j - \bar{\xi}_i\xi_j) \quad [69]$$

The supercharges are conserved if the curl of $\Phi_i(\boldsymbol{x})$ vanishes: $\nabla_i\Phi_j - \nabla_j\Phi_i = 0$. It follows that at least locally there exists a single function $W(\boldsymbol{x})$ such that

$$\Phi_i(\boldsymbol{x}) = \nabla_i W(\boldsymbol{x}) \quad [70]$$

$W(\boldsymbol{x})$ is called the superpotential. Defining the operators

$$A_i = p_i - \mathrm{i}\Phi_i(\boldsymbol{x}), \qquad A_i^\dagger = p_i + \mathrm{i}\Phi_i(\boldsymbol{x})$$
$$A_i A_j^\dagger - A_j^\dagger A_i = \nabla_i\Phi_j + \nabla_j\Phi_i \quad [71]$$

the supersymmetric quantum theory is defined by

$$Q = A_i f_i^\dagger, \qquad Q^\dagger = A_i^\dagger f_i$$
$$H = \tfrac{1}{2}(QQ^\dagger + Q^\dagger Q) \quad [72]$$

The Hamiltonian is the direct operator translation of the classical expression [69]; its normal-ordered form is

$$H = A_i^\dagger A_i + \tfrac{1}{2}(\nabla_i\Phi_j + \nabla_j\Phi_i) f_i^\dagger f_j \quad [73]$$

The total fermion number operator

$$N_f = f_i^\dagger f_i \quad [74]$$

(summed over $i$) satisfying the commutation relations

$$\left[N_f, f_j^\dagger\right]_- = f_j^\dagger, \qquad [N_f, f_j]_- = -f_j \quad [75]$$

commutes with the Hamiltonian. Hence, the stationary states can be labeled by the energy $E$ and the total fermion number $n_f = (0, \ldots, n)$. The energy spectrum being positive semidefinite, all positive-energy states occur in pairs of fermion number $(n_f, n_f + 1)$; zero-energy states exist only if the equations

$$A_i f_i^\dagger\,|0, n_f\rangle = A_i^\dagger f_i\,|0\,, n_f\rangle = 0 \quad [76]$$

admit a normalizable solution. In this context, the vanishing of the curl of $\Phi_i(\mathbf{x})$ is important, as it is a necessary condition for the formal solutions

$$\psi_\pm(\boldsymbol{x}) = C_\pm \exp\left(\pm\int_0^{\mathbf{x}} \Phi(\boldsymbol{y})\cdot \mathrm{d}\boldsymbol{y}\right) = C'_\pm\,\mathrm{e}^{\pm W(\mathbf{x})} \quad [77]$$

to be single-valued. If one of them is normalizable, there exists a zero-energy ground state with $n_f = 0$ or $n_f = n$, represented by a wave function:

$$|0,0\rangle \rightarrow \Psi_0(\boldsymbol{x},\boldsymbol{\xi}) = \psi_-(\boldsymbol{x})$$
$$|0,r\rangle \rightarrow \Psi_r(\boldsymbol{x},\boldsymbol{\xi}) = \psi_+(\boldsymbol{x})\xi_1\ldots\xi_n \quad [78]$$

Alternatively, we can represent the wave functions as spinors of dimension $2^n$, on which the fermion operators $f_i^\dagger$ and $f_i$ act as a $2^n$-dimensional matrix representation of the Clifford algebra with generators $\gamma_a, a = 1, \ldots, 2n$, defined by

$$\gamma_i = f_i + f_i^\dagger, \qquad \gamma_{i+r} = \mathrm{i}\left(f_i - f_i^\dagger\right) \quad [79]$$

These operators indeed satisfy the anticommutation rule

$$\gamma_a\gamma_b + \gamma_b\gamma_a = 2\delta_{ab} \quad [80]$$

Thus, the wave functions have $2^n$ components, as compared to the $2^{[n/2]}$ polarization states of the minimal models.

## The Witten Index

We have noted that for supersymmetric quantum systems, like the harmonic and anharmonic supersymmetric oscillator, states exist in pairs of different fermion number, degenerate in energy, except for possibly one or more zero-energy states which are superinvariant in the sense that

$$Q|0,n\rangle = \bar{Q}|0,n\rangle = 0 \;\Leftrightarrow\; H|0,n\rangle = 0 \quad [81]$$

In the Schrödinger representation, these states are characterized as zero modes of the Dirac operator:

$$\gamma\cdot D\,\Psi = 0 \quad [82]$$

where $D_i$ is an ordinary or field-dependent (e.g., covariant) derivative. Clearly, the existence of such states can, in some cases, be guaranteed if there is no state which can pair up with a given state to form a superdoublet. Witten developed a topological characterization of this condition, encoded in an index defined by

$$I = \mathrm{tr}(-1)^{N_f} = n_b(E=0) - n_f(E=0) \quad [83]$$

where $N_f$ is the fermion number operator, and $n_{b,f}(E=0)$ are the number of bosonic and fermionic zero-energy states. The trace is taken over the complete space of states, but as all nonzero energy states occur in pairs of a bosonic and a fermionic state, their contributions to the trace cancel, having opposite sign. Therefore, the trace is actually only over the zero-energy states, and counts the number of bosonic states with positive sign, and the number of fermionic states with negative sign. If the index vanishes, $I = 0$, then any zero-energy states necessarily exist in equal number of bosonic and fermionic states; under perturbations of the potential, these states can form pairs and change their energy to a positive value. However, if the index does not vanish, $I \neq 0$, then there are states which have no

partner of complementary fermion number; these states can never get a nonzero energy under changes in the parameters of the potential, as long as the changes respect supersymmetry. Such systems, therefore, necessarily possess exact zero-energy states which are invariant under all supersymmetries.

Deformations of the potential respecting supersymmetry are those obtained by changing the parameters in the superpotential. The usefulness of this concept is, therefore, that the index for models with complicated superpotentials can be computed by comparing them with models with simple superpotentials having similar topological properties.

Counting the number of states is not always a simple procedure, in particular when the spectrum includes continuum states. Therefore, in practice one often needs a regularization procedure, by taking the trace over the full state space of the exponentially damped quantity

$$I(\beta) = \mathrm{tr}(-1)^{N_f} \mathrm{e}^{-\beta H} \qquad [84]$$

and taking the limit $\beta \to 0$. The quantity [84] can be computed in terms of a path integral with periodic boundary conditions for the fermionic degrees of freedom.

Finally, as the wave function representation of supersymmetric quantum mechanics [82] links the Witten index to the space of zero modes of a Dirac operator, in particular cases it can be used to describe topological aspects of sigma models and gauge theories, and related mathematical quantities such as the Atiyah–Singer index.

More details and references to the original literature can be found in the reviews listed in the Further Reading section.

*See also:* Path-Integrals in Non Commutative Geometry; Supermanifolds.

## Further Reading

Cooper F, Khare A, and Sukhatme U (1995) Supersymmetry and quantum mechanics. *Physics Reports* 251: 267.

De Witt BS (1984, 1992) *Supermanifolds*. Cambridge: Cambridge University Press.

Shifman MA (1999) *ITEP Lectures on Particle Physics and Field Theory*, vol. 1, ch. 4. Singapore: World Scientific.

van Holten JW (1996) D = 1 Supergravity and spinning particles. In: Jancewicz B and Sobezyk J (eds.) *From Field Theory to Quantum Groups*, p. 173. Singapore: World Scientific.

# Supersymmetry Methods in Random Matrix Theory

**M R Zirnbauer**, Universität Köln, Köln, Germany



## Introduction

A prominent theme of modern condensed matter physics is electronic transport – in particular, the electrical conductivity – of disordered metallic systems at very low temperatures. From the Landau theory of weakly interacting Fermi liquids, one expects the essential aspects of the situation to be captured by the single-electron approximation. Mathematical models that have been proposed and studied in this context include random Schrödinger operators and band random matrices.

If the physical system has infinite size, two distinct possibilities exist: the quantum single-electron motion may either be bounded or unbounded. In the former case, the disordered electron system is an insulator, in the latter case, a metal with finite conductivity (if the electron motion is not critical but diffusive). Metallic behavior is expected for weakly disordered systems in three dimensions; insulating behavior sets in when the disorder strength is increased or the space dimension reduced.

The main theoretical tool used in the physics literature on the subject is the "supersymmetry method" pioneered by Wegner and Efetov (1979–83). Over the past 20 years, physicists have applied the method in many instances, and a rather complete picture of weakly disordered metals has emerged. Several excellent reviews of these developments are available in print.

From the perspective of mathematics, however, the method has not always been described correctly, and what is sorely lacking at present is an exposition of how to implement the method rigorously. (Unfortunately, the correct exposition by Schäfer and Wegner (1980) was largely ignored or forgotten by later authors.) In this article, an attempt is made to help remedy the situation, by giving a careful review of the Wegner–Efetov supersymmetry method for the case of Hermitian band random matrices.

## Gaussian Ensembles

Let $V$ be a unitary vector space of finite dimension. A Hermitian random matrix model on $V$ is defined

by some probability distribution on Herm($V$), the Hermitian linear operators on $V$. We may fix some orthonormal basis of $V$ and represent the elements $H$ of Herm($V$) by Hermitian square matrices.

Quite generally, probability distributions are characterized by their Fourier transform or characteristic function. In the present case this is

$$\Omega(K) = \left\langle \mathrm{e}^{\mathrm{itr} HK} \right\rangle$$

where the Fourier variable $K$ is some other linear operator on $V$, and $\langle \dots \rangle$ denotes the expectation value with respect to the probability distribution for $H$. Later, it will be important that, if $\Omega(K)$ is an analytic function of $K$, the matrix entries of $K$ need not be from $\mathbb{R}$ or $\mathbb{C}$ but can be taken from the even part of some exterior algebra.

The probability distributions to be considered in this article are Gaussian with zero mean, $\langle H \rangle = 0$. Their Fourier transform is also Gaussian:

$$\Omega(K) = \mathrm{e}^{-(1/2)J(K,K)}$$

with $J$ some quadratic form. We now describe $J$ for a large family of hierarchical models that includes the case of band random matrices.

Let $V$ be given a decomposition by orthogonal vector spaces:

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_{|\Lambda|}$$

We should imagine that every vector space $V_i$ corresponds to one site $i$ of some lattice $\Lambda$, and the total number of sites is $|\Lambda|$. For simplicity, we take all dimensions to be equal: $\dim V_1 = \cdots = \dim V_{|\Lambda|} = N$. Thus, the dimension of $V$ is $N|\Lambda|$. The integer $N$ is called the number of orbitals per site.

If $\Pi_i$ is the orthogonal projector on the linear subspace $V_i \subset V$, we take the bilinear form $J$ to be

$$J(K, K') = \sum_{i,j=1}^{|\Lambda|} J_{ij} \, \mathrm{tr}(\Pi_i K \, \Pi_j K')$$

where the coefficients $J_{ij}$ are real, symmetric, and positive. This choice of $J$ implies invariance under the group $\mathscr{U}$ of unitary transformations in each subspace:

$$\mathscr{U} = \mathrm{U}(V_1) \times \mathrm{U}(V_2) \times \cdots \times \mathrm{U}(V_{|\Lambda|})$$

Clearly, $\Omega(K) = \Omega(UKU^{-1})$ or, equivalently, the probability distribution for $H$ is invariant under conjugation $H \mapsto UHU^{-1}$, for $U \in \mathscr{U}$.

If $\{e_i^a\}_{a=1,\dots,N}$ is an orthonormal basis of $V_i$, we define linear operators $E_{ij}^{ab} : V_j \to V_i$ by $E_{ij}^{ab} e_j^b = e_i^a$. By evaluating $J(E_{ij}^{ab}, E_{j'i'}^{b'a'}) = J_{ij}\delta_{ii'}\delta_{jj'}\delta^{aa'}\delta^{bb'}$, one sees that the matrix entries of $H$ all are statistically independent.

By varying the lattice $\Lambda$, the number of orbitals $N$, and the variances $J_{ij}$, one obtains a large class of Hermitian random matrix models, two prominent subclasses of which are the following:

1. For $|\Lambda| = 1$, one gets the Gaussian Unitary Ensemble (GUE). Its symmetry group is $\mathscr{U} = \mathrm{U}(N)$, the largest one possible in dimension $N = \dim V$.
2. If $|i - j|$ denotes a distance function for $\Lambda$, and $f$ a rapidly decreasing positive function on $\mathbb{R}_+$ of width $W$, the choice $J_{ij} = f(|i - j|)$ with $N = 1$ gives an ensemble of band random matrices with bandwidth $W$ and symmetry group $\mathscr{U} = \mathrm{U}(1)^{|\Lambda|}$.

Beyond being real, symmetric, and positive, the variances $J_{ij}$ are required to have two extra properties in order for all of the following treatment to go through:

- They must be positive as a quadratic form. This is to guarantee the existence of an inverse, which we denote by $w_{ij} = (J^{-1})_{ij}$.
- The off-diagonal matrix entries of the inverse must be nonpositive: $w_{ij} \leq 0$ for $i \neq j$.

## Basic Tools

### Green's Functions

A major goal of random matrix theory is to understand the statistical behavior of the spectrum and the eigenstates of a random Hamiltonian $H$. Spectral and eigenstate information can be extracted from the Green's function, that is, from matrix elements of the operator $(z - H)^{-1}$ with complex parameter $z \in \mathbb{C} \backslash \mathbb{R}$. For the models at hand, the good objects to consider are averages of $\mathscr{U}$-invariant observables such as

$$G_i^{(1)}(z) = \left\langle \mathrm{tr}\, \Pi_i (z - H)^{-1} \right\rangle \quad [1]$$

$$G_{ij}^{(2)}(z_1, z_2) = \left\langle \mathrm{tr}\, \Pi_i (z_1 - H)^{-1} \Pi_j (z_2 - H)^{-1} \right\rangle \quad [2]$$

The discontinuity of $G_i^{(1)}(z)$ across the real $z$-axis yields the local density of states. In the limit of infinite volume ($|\Lambda| \to \infty$), the function $G_{ij}^{(2)}(z_1, z_2)$ for $z_1 = E + \mathrm{i}\varepsilon$, $z_2 = E - \mathrm{i}\varepsilon$, real energy $E$, and $\varepsilon > 0$ going to zero, gives information on transport, for example, the electrical conductivity by the Kubo–Greenwood formula.

Mathematically speaking, if $G_{ij}^{(2)}(E + \mathrm{i}\varepsilon, E - \mathrm{i}\varepsilon)$ is bounded (for infinite volume) in $\varepsilon$ and decreases algebraically with distance $|i - j|$ at $\varepsilon = 0+$, the

spectrum is absolutely continuous and the eigenstates are extended at energy $E$. On the other hand, a pure point spectrum and localized eigenstates are signaled by the behavior $G^{(2)}_{ij} \sim \varepsilon^{-1} e^{-\lambda |i-j|}$ with positive Lyapunov exponent $\lambda$.

#### Green's Functions from Determinants

For any pair of linear operators $A$, $B$ on a finite-dimensional vector space $V$, the following formula from basic linear algebra holds if $A$ has an inverse:

$$\frac{\mathrm{d}}{\mathrm{d}t} \det(A + tB)\Big|_{t=0} = \det(A)\,\mathrm{tr}(A^{-1}B)$$

Using it with $A = z - H$ and $z \in \mathbb{C}\backslash\mathbb{R}$, all Green's functions can be expressed in terms of determinants; for example,

$$G^{(2)}_{ij}(w,z) = \sum_{a,b=1}^{N} \frac{\partial^2}{\partial s \partial t} \left\langle \frac{\det(w-H)\det(z-H+tE^{ab}_{ij})}{\det(w-H-sE^{ba}_{ji})\det(z-H)} \right\rangle \Bigg|_{s=t=0}$$

It is clear that, given a formula of this kind, what one wants is a method to handle ensemble averages of ratios of determinants. This is what is reviewed in the sequel.

#### Determinants as Gaussian Integrals

Let the Hermitian scalar product of the unitary vector space $V$ be written as $\varphi_1, \varphi_2 \mapsto (\bar{\varphi}_1, \varphi_2)$, and denote the adjoint or Hermitian conjugate of a linear operator $A$ on $V$ by $A^*$. If $\mathfrak{Re}\, A := (1/2)(A + A^*) > 0$, the standard Lebesgue integral of the Gaussian function $\varphi \mapsto e^{-(\bar{\varphi}, A\varphi)}$ makes sense and gives

$$\int e^{-(\bar{\varphi}, A\varphi)} = \det A^{-1} \qquad [3]$$

where it is understood that we are integrating with the Lebesgue measure on (the normed vector space) $V$ normalized by $\int e^{-(\bar{\varphi},\varphi)} = 1$. The same integral with anticommuting $\psi$ instead of the (commuting) $\varphi \in V$ gives

$$\int e^{-(\bar{\psi}, A\psi)} = \det A \qquad [4]$$

This basic formula from the field theory of fermionic particles is a consequence of the integration over anticommuting variables actually being *differentiation*:

$$\int \mathrm{d}\bar{\psi}_1\, \mathrm{d}\psi_1 f(\bar{\psi}_1, \psi_1, \ldots) := \frac{\partial^2}{\partial \bar{\psi}_1\, \partial \psi_1} f(\bar{\psi}_1, \psi_1, \ldots)$$

## Fermionic Variant

The supersymmetry method of random matrix theory is a theme with many variations. The first variation to be described is the "fermionic" one. To optimize the notation, we now write $\mathrm{d}\mu_{N,J}(H)$ for the density of the Gaussian probability distribution of $H$:

$$\langle F(H) \rangle = \int F(H)\, \mathrm{d}\mu_{N,J}(H)$$

All determinants and traces appearing below will be taken over vector spaces that are clear from the context.

Let $z_1, \ldots, z_n$ be any set of $n$ complex numbers, put $z := \mathrm{diag}(z_1, \ldots, z_n)$ for later purposes, and consider

$$\Omega^{\mathrm{ferm}}_{n,N}(z, J) = \int \prod_{\alpha=1}^{n} \det(z_\alpha - H)\, \mathrm{d}\mu_{N,J}(H) \qquad [5]$$

The supersymmetry method expresses this average of a product of determinants in an alternative way, by integrating over a "dual" measure as follows.

Introducing an auxiliary unitary vector space $\mathbb{C}^n$, one associates with every site $i$ of the lattice $\Lambda$ an object $Q_i \in \mathrm{Herm}(\mathbb{C}^n)$, the space of Hermitian $n \times n$ matrices. If $\mathrm{d}Q_i$ for $i = 1, \ldots, |\Lambda|$ are Lebesgue measures on $\mathrm{Herm}(\mathbb{C}^n)$, one puts $DQ = \mathrm{const.} \times \prod_i \mathrm{d}Q_i$ and

$$\mathrm{d}\nu_{n,J}(Q) := e^{-(1/2)\Sigma_{i,j}(J^{-1})_{ij}\,\mathrm{tr}\, Q_i Q_j} DQ \qquad [6]$$

The multiplicative constant in $DQ$ is fixed by requiring the density to be normalized: $\int \mathrm{d}\nu_{n,J}(Q) = 1$. By completing the square, this Gaussian probability measure has the characteristic function

$$\int e^{\mathrm{i}\Sigma_j \mathrm{tr} Q_j K_j}\, \mathrm{d}\nu_{n,J}(Q) = e^{-(1/2)\Sigma_{ij} J_{ij}\,\mathrm{tr}\, K_i K_j}$$

where the Fourier variables $K_1, \ldots, K_{|\Lambda|}$ are $n \times n$ matrices with matrix entries taken from $\mathbb{C}$ or another commutative algebra.

The key relation of the fermionic variant of the supersymmetry method is that the expectation of the product of determinants [5] has another expression as

$$\Omega^{\mathrm{ferm}}_{n,N}(z, J) = \int \prod_{j=1}^{|\Lambda|} \det{}^N(z - \mathrm{i}Q_j)\, \mathrm{d}\nu_{n,J}(Q) \qquad [7]$$

($\mathrm{i} = \sqrt{-1}$). The strategy of the proof is quite simple: one writes the determinants in both expressions for $\Omega^{\mathrm{ferm}}_{n,N}$ as Gaussian integrals over $nN|\Lambda|$ complex fermionic variables $\psi_1, \ldots, \psi_n$ (each $\psi_\alpha$ is a vector in $V$ with anticommuting coefficients), using the basic

formula [4]. The integrals then encountered are essentially the Fourier transforms of the distributions $d\mu_{N,J}(H)$ resp., $d\nu_{n,J}(Q)$. The result is

$$\int e^{-\Sigma_\gamma z_\gamma(\bar\psi_\gamma,\psi_\gamma)}\, e^{-(1/2)\Sigma_{ij}J_{ij}\Sigma_{\alpha\beta}(\bar\psi_\alpha,\Pi_i\psi_\beta)(\bar\psi_\beta,\Pi_j\psi_\alpha)}$$

for both expressions of $\Omega^{\text{ferm}}_{n,N}$. In other words, although the probability distributions $d\mu_{N,J}(H)$ and $d\nu_{n,J}(Q)$ are distinct (they are defined on different spaces), their characteristic functions coincide when evaluated on the Fourier variables $K = \sum_\alpha \psi_\alpha(\bar\psi_\alpha, \bullet)$ for $H$ and $(K_i)_{\alpha\beta} = (\bar\psi_\alpha, \Pi_i\psi_\beta)$ for $Q_i$. This establishes the claimed equality of the expressions [5] and [7] for $\Omega^{\text{ferm}}_{n,N}(z,J)$.

What is the advantage of passing to the alternative expression by $d\nu_{n,J}(Q)$? The answer is that, while $H$ is made up of independent random variables, the new variables $Q_i$, called the Hubbard–Stratonovich field, are correlated: they interact through the "exchange" constants $w_{ij} = (J^{-1})_{ij}$. If that interaction creates enough collectivity, a kind of mean-field behavior results.

For the simple case of GUE ($|\Lambda| = 1, w_{11} = N/\lambda^2$) with $z_1 = \cdots = z_n = E$, one gets the relation

$$\langle \det^n(E-H)\rangle = \int \det^N(E-iQ)\, e^{-(N/2\lambda^2)\operatorname{tr} Q^2}\, dQ$$

the right-hand side of which is easily analyzed by the steepest descent method in the limit of large $N$.

For band random matrices in the so-called ergodic regime, the physical behavior turns out to be governed by the constant mode $Q_1 = \cdots = Q_{|\Lambda|}$ – a fact that can be used to establish GUE universality in that regime.

## Bosonic Variant

The bosonic variant of the present method, due to Wegner, computes averages of products of determinants placed in the denominator:

$$\Omega^{\text{bos}}_{n,N}(z,J) = \int \prod_{\alpha=1}^{n} \det^{-1}(z_\alpha - H)\, d\mu_{N,J}(H) \qquad [8]$$

where we now require $\mathfrak{Im} z_\alpha \neq 0$ for all $\alpha = 1,\ldots,n$. Complications relative to the fermionic case arise from the fact that the integrand in [8] has poles. If one replaces the anticommuting vectors $\psi_\alpha$ by commuting ones $\varphi_\alpha$, and then simply repeats the previous calculation in a naive manner, one arrives at

$$\Omega^{\text{bos}}_{n,N}(z,J) \overset{?}{=} \int \prod_{j=1}^{|\Lambda|} \det^{-N}(z - Q_j)\, d\nu_{n,J}(Q) \qquad [9]$$

where the integral is still over $Q_j \in \text{Herm}(\mathbb{C}^n)$. The calculation is correct, and relation [9] therefore holds true, provided that the parameters $z_1,\ldots,z_n$ all lie in the same half (upper or lower) of the complex plane. To obtain information on transport properties, however, one needs parameters in both the upper and lower halves; see the paragraph following [2]. The general case to be addressed below is $\mathfrak{Im} z_\alpha > 0$ for $\alpha = 1,\ldots,p$, and $\mathfrak{Im} z_\alpha < 0$ for $\alpha = p+1,\ldots,n$. Careful inspection of the steps leading to eqn [9] reveals a convergence problem for $0 < p < n$. In fact, [9] with $Q_j$ in $\text{Herm}(\mathbb{C}^n)$ turns out to be false in that range. Learning how to resolve this problem is the main step toward mathematical mastery of the method. Let us therefore give the details.

If $s_\alpha := \operatorname{sgn} \mathfrak{Im} z_\alpha$, the good (meaning convergent) Gaussian integral to consider is

$$\int e^{i\Sigma_\alpha s_\alpha(\bar\varphi_\alpha,(z_\alpha - H)\varphi_\alpha)} = \prod_{\alpha=1}^{n} \det^{-1}(-is_\alpha(z_\alpha - H))$$

To avoid carrying around trivial constants, we now assume $i^{(n-2p)N|\Lambda|} = 1$. Use of the characteristic function of the distribution for $H$ then gives

$$\Omega^{\text{bos}}_{n,N}(z,J) = \int e^{i\Sigma_\gamma s_\gamma z_\gamma(\bar\varphi_\gamma,\varphi_\gamma)} \times e^{-\frac{1}{2}\Sigma_{ij}J_{ij}\Sigma_{\alpha\beta}s_\alpha(\bar\varphi_\alpha,\Pi_i\varphi_\beta)s_\beta(\bar\varphi_\beta,\Pi_j\varphi_\alpha)} \qquad [10]$$

The difficulty of analyzing this expression stems from the "hyperbolic" nature (due to the indefiniteness of the signs $s_\alpha = \pm 1$) of the term quartic in the $\varphi_\alpha, \bar\varphi_\alpha$.

### Fyodorov's Method

The integrand for $\Omega^{\text{bos}}$ is naturally expressed in terms of $n \times n$ matrices $M_i$ with matrix elements $(M_i)_{\alpha\beta} = (\bar\varphi_\alpha, \Pi_i\varphi_\beta)$. These matrices lie in $\text{Herm}^+(\mathbb{C}^n)$, that is, they are non-negative as well as Hermitian. Fyodorov's idea was to introduce them as the new variables of integration. To do that step, recall the basic fact that, given two differentiable spaces $X$ and $Y$ and a smooth map $\psi: X \to Y$, a distribution $\mu$ on $X$ is pushed forward to a distribution $\psi(\mu)$ on $Y$ by $\psi(\mu)[f] := \mu[f \circ \psi]$, where $f$ is any test function on $Y$.

We apply this universal principle to the case at hand by identifying $X$ with $V^n$, and $Y$ with $(\text{Herm}^+(\mathbb{C}^n))^{|\Lambda|}$, and $\psi$ with the mapping that sends

$$(\varphi_1,\ldots,\varphi_n) \in X \quad \text{to} \quad (M_1,\ldots,M_{|\Lambda|}) \in Y$$

by $(M_i)_{\alpha\beta} = (\bar\varphi_\alpha, \Pi_i\varphi_\beta)$. On $X = V^n$ we are integrating with the product Lebesgue measure normalized by $\int e^{-\Sigma_\alpha(\bar\varphi_\alpha,\varphi_\alpha)} = 1$. We now want the push-forward of this flat measure (or distribution) by the mapping $\psi$. In general, the push-forward of a measure is not

guaranteed to have a density but may be singular (like a Dirac $\delta$-distribution). This is in fact what happens if $N < n$. The matrices $M_i$ then have less than the maximal rank, so they fail to be positive but possess zero eigenvalues, which implies that the flat measure on $X$ is pushed forward by $\psi$ into the boundary of $Y$. For $N \geq n$, on the other hand, the push-forward measure does have a density on $Y$; and that density is $\prod_{i=1}^{|\Lambda|}(\det M_i)^{N-n}\mathrm{d}M_i$, as is seen by transforming to the eigenvalue representation and comparing Jacobians. The $\mathrm{d}M_i$ are Lebesgue measures on $\mathrm{Herm}(\mathbb{C}^n)$, normalized by the condition

$$\int_{M_i>0} \mathrm{e}^{-\mathrm{tr}M_i}(\det M_i)^{N-n}\,\mathrm{d}M_i = \int \mathrm{e}^{-\Sigma_\alpha(\bar{\varphi}_\alpha,\Pi_i\varphi_\alpha)} = 1$$

Assembling the sign information for $\mathfrak{Im} z_\alpha$ in a diagonal matrix $s := \mathrm{diag}(s_1,\ldots,s_n)$, and pushing the integral over $X$ forward to an integral over $Y$ with measure $DM := \prod_i \mathrm{d}M_i$, we obtain Fyodorov's formula:

$$\Omega^{\mathrm{bos}}_{n,N}(z,J) = \int_Y \mathrm{e}^{-(1/2)\Sigma_{ij}J_{ij}\mathrm{tr}(sM_isM_j)} \times \mathrm{e}^{\Sigma_k \mathrm{tr}(\mathrm{i}szM_k+(N-n)\ln M_k)}DM \qquad [11]$$

This formula has a number of attractive features. One is ease of derivation, another is ready generalizability to the case of non-Gaussian distributions. The main disadvantage of the formula is that it does not apply to the case of band random matrices (because of the restriction $N \geq n$); nor does it combine nicely with the fermionic formula [7] to give a supersymmetric formalism, as one formula is built on $J_{ij}$ and the other on $w_{ij}$.

Note that [11] clearly displays the dependence on the signature of $\mathfrak{Im} z$: you cannot remove the $s_1,\ldots,s_n$ from the integrand without changing the domain of integration $Y = (\mathrm{Herm}^+(\mathbb{C}^n))^{|\Lambda|}$. This important feature is missing from the naive formula [9].

Setting $q = n - p$, let $\mathrm{U}(p,q)$ be the pseudounitary group of complex $n \times n$ matrices $T$ with inverse $T^{-1} = sT^*s$. Since $|\det T| = 1$ for $T \in \mathrm{U}(p,q)$, the integration domain $Y$ and density $DM = \prod_i \mathrm{d}M_i$ of Fyodorov's formula are invariant under $\mathrm{U}(p,q)$ transformations $M_i \mapsto TM_iT^*$, and so is actually the integrand in the limit where all parameters $z_1,\ldots,z_n$ become equal. Thus, the elements of $\mathrm{U}(p,q)$ are global symmetries in that limit. This observation holds the key to another method of transforming the expression [10].

### The Method of Schäfer and Wegner

To rescue the naive formula [9], what needs to be abandoned is the integration domain $\mathrm{Herm}(\mathbb{C}^n)$ for the matrices $Q_i$. The good domain to use was constructed by Schäfer and Wegner, but was largely forgotten in later physics work.

Writing $(M_k)_{\alpha\beta} = (\bar{\varphi}_\alpha, \Pi_k\varphi_\beta)$ as before, consider the function

$$F_M(Q) = \mathrm{e}^{(1/2)\Sigma_{ij}w_{ij}\mathrm{tr}(sQ_i+\mathrm{i}z)(sQ_j+\mathrm{i}z)-\Sigma_k \mathrm{tr}M_kQ_k} \qquad [12]$$

viewed as a holomorphic function of

$$Q = (Q_1,\ldots,Q_{|\Lambda|}) \in \mathrm{End}(\mathbb{C}^n)^{|\Lambda|}$$

If the Gaussian integral $\int F_M(Q)DQ$ with holomorphic density $DQ = \prod_i \mathrm{d}Q_i$ is formally carried out by completing the square, one gets the integrand of [10]. This is just what we want, as it would allow us to pass to a $Q$-matrix formulation akin to the one of the previous section. But how can that formal step be made rigorous? To that end, one needs to (1) construct a domain on which $|F_M(Q)|$ decreases rapidly so that the integral exists, and (2) justify completion of the square and shifting of variables.

To begin, take the absolute value of $F_M(Q)$. Putting $(1/2)(Q_j + Q_j^*) =: \mathfrak{Re}Q_j$ and $(1/2\mathrm{i})(Q_j - Q_j^*) =: \mathfrak{Im}Q_j$, we have $|F_M| = \mathrm{e}^{-(1/4)(f_1+f_2+f_3)}$ with

$$f_1(Q) = \sum_{ij} w_{ij}\mathrm{tr}(s\mathfrak{Im}Q_i + z)(s\mathfrak{Im}Q_j + z) + \mathrm{c.c.}$$

$$f_2(Q) = -2\sum_{ij} w_{ij}\mathrm{tr}(s\mathfrak{Re}Q_i)(s\mathfrak{Re}Q_j)$$

$$f_3(Q) = 4\sum_i \mathrm{tr}\left(M_i + s\mathfrak{Im}z\sum_j w_{ij}\right)\mathfrak{Re}Q_i$$

These expressions suggest making the following choice of integration domain for $Q_i (i = 1,\ldots,|\Lambda|)$. Pick some real constant $\lambda > 0$ and put

$$\mathfrak{Re}Q_i = \lambda T_iT_i^*, \qquad \mathfrak{Im}Q_i = P_i := \begin{pmatrix} P_i^+ & 0 \\ 0 & P_i^- \end{pmatrix}$$

with $T_i \in \mathrm{U}(p,q), P_i^+ \in \mathrm{Herm}(\mathbb{C}^p), P_i^- \in \mathrm{Herm}(\mathbb{C}^q)$. The set of matrices $Q_i$ so defined is referred to as the Schäfer–Wegner domain $X_\lambda^{p,q}$. The range of the field $Q = (Q_1,\ldots,Q_{|\Lambda|})$ is the direct product $\mathscr{X} := (X_\lambda^{p,q})^{|\Lambda|}$.

To show that this is a good choice of domain, we first of all show convergence of the integral $\int_{\mathscr{X}} F_M(Q)DQ$. The matrices $P_i$ commute with $s$, so

$$f_1(Q)|_{\mathscr{X}} = 2\mathfrak{Re}\sum_{ij} w_{ij}\mathrm{tr}(P_i + sz)(P_j + sz)$$

Since the coefficients $w_{ij}$ are positive as a quadratic form, this expression is convex (with a positive Hessian) in the Hermitian matrices $P_i$. Second, the function

$$f_2(Q)|_{\mathscr{X}} = -2\lambda^2\sum_{ij} w_{ij}\mathrm{tr}\left(T_iT_i^*\right)^{-1}T_jT_j^*$$

is bounded from below by the constant $-2\lambda^2 n\Sigma_i w_{ii}$. This holds true because $w_{ij}$ is negative for $i \neq j$, and because $T_i T_i^* > 0$ and the trace of a product of two positive Hermitian matrices is always positive. Third,

$$f_3(Q)|_{\mathscr{X}} = 4\lambda \sum_i \mathrm{tr}\left(M_i + s\Im\mathfrak{m} z \sum_j w_{ij}\right) T_i T_i^*$$

is positive, as $(\ldots)$ is positive Hermitian. As long as $s\Im\mathfrak{m} z > 0$, the function $f_3$ goes to infinity for all possible directions of taking the $T_i$ to infinity on $\mathrm{U}(p,q)$.

Thus, when the matrices $Q_i$ are taken to vary on the Schäfer–Wegner domain $X_\lambda^{p,q}$, the absolute value $|F_M| = \mathrm{e}^{-(1/4)(f_1+f_2+f_3)}$ decreases rapidly at infinity. This establishes the convergence of $\int_{\mathscr{X}} F_M(Q)DQ$.

Next, let us count dimensions. The mapping $T \mapsto TT^*$ for $T \in \mathrm{U}(p,q) =: G$ is invariant under right multiplication of $T$ by elements of the unitary subgroup $H := \mathrm{U}(p) \times \mathrm{U}(q)$ – it is called the "Cartan embedding" of $G/H$ into $G$. The real manifold $G/H$ has dimension $2pq$ and so does its image under the Cartan embedding. Augmenting this by the dimension of $\mathrm{Herm}(\mathbb{C}^p)$ and $\mathrm{Herm}(\mathbb{C}^q)$ (from $P_i$), one gets $\dim X_\lambda^{p,q} = 2pq + p^2 + q^2 = (p+q)^2 = n^2$, which is as it should be.

Finally, why can one shift variables and do the Gaussian integral over $Q$ (with translation-invariant $DQ$) by completing the square? This question is legitimate as the Schäfer–Wegner domain $X_\lambda^{p,q}$ lacks invariance under the required shift, which is $Q_i \mapsto Q_i - \mathrm{i}sz + \sum_j J_{ij} s M_j s$.

To complete the square in [12], introduce a parameter $t \in [0,1]$ and consider the family of shifts

$$Q_i \mapsto Q_i + t\left(-\mathrm{i}sz + \Sigma_j J_{ij} s M_j s\right)$$

For fixed $t$, this shift takes $\mathscr{X} = (X_\lambda^{p,q})^{|\Lambda|}$ into another domain, $\mathscr{X}(t)$. Inspection shows that the function [12] still decreases rapidly (uniformly in the $M_i$) on $\mathscr{X}(t)$, as long as $t < 1$. Without changing the integral, one can add pieces to $\mathscr{X}(t)$ (for $t < 1$) at infinity to arrange for the chain $\mathscr{X} - \mathscr{X}(t)$ to be a cycle. Because $\mathscr{X}(t)$ is homotopic to $\mathscr{X}(0) = \mathscr{X}$, this cycle is a boundary: there exists a manifold $\mathscr{Y}(t)$ of dimension $\dim \mathscr{X} + 1$ such that $\partial\mathscr{Y}(t) = \mathscr{X} - \mathscr{X}(t)$. Viewed as a holomorphic differential form of degree $(n^2|\Lambda|, 0)$ in the complex space $\mathrm{End}(\mathbb{C}^n)^{|\Lambda|}$, the integrand $\omega := F_M(Q)DQ$ is closed (i.e., $\mathrm{d}\omega = 0$). Therefore, by Stokes' theorem,

$$\int_{\mathscr{X}} \omega - \int_{\mathscr{X}(t)} \omega = \int_{\partial\mathscr{Y}(t)} \omega = \int_{\mathscr{Y}(t)} \mathrm{d}\omega = 0$$

which proves $\int_{\mathscr{X}(t)} F_M(Q)DQ = \int_{\mathscr{X}} F_M(Q)DQ$, independent of $t$. (This argument does not go through for the nonrigorous choice $sQ_i := T_i P_i T_i^{-1}$ usually made!)

In the limit $t \to 1$, one encounters the expression

$$\int_{\mathscr{X}(1)} F_M(Q)DQ = \int_{\mathscr{X}} \mathrm{d}\nu_{n,J}(\mathrm{i}sQ) \times \mathrm{e}^{-(1/2)\Sigma_{ij} J_{ij} \mathrm{tr}(sM_i sM_j) + \mathrm{i}\Sigma_k \mathrm{tr}(szM_k)}$$

with $\mathrm{d}\nu_{n,J}$ as in [6]. The normalization integral over $\mathscr{X}$ is defined by taking the Hermitian matrices $P_i$ to be the inner variables of integration. The outer integrals over the $T_i$ then demonstrably exist, and one can fix the (otherwise arbitrary) normalization of $DQ$ by setting $\int_{\mathscr{X}} \mathrm{d}\nu_{n,J}(\mathrm{i}sQ) = 1$. Making that choice, and comparing with [10], one has proved

$$\Omega_{n,N}^{\mathrm{bos}} = \int_{\varphi,\bar\varphi} \left( \int_{\mathscr{X}} F_{(M_i)_{\alpha\beta} = (\bar\varphi_\alpha, \Pi_i \varphi_\beta)}(Q)DQ \right)$$

The final step is to change the order of integration over the $Q$- and $\varphi$-variables, which is permitted since the $Q$-integral converges uniformly in $\varphi$. Doing the Gaussian $\varphi$-integral and shifting $Q_k \to Q_k - \mathrm{i}sz$, one arrives at the Schäfer–Wegner formula for $\Omega_{n,N}^{\mathrm{bos}}$:

$$\Omega_{n,N}^{\mathrm{bos}}(z, w^{-1}) = \int_{\mathscr{X}} \mathrm{e}^{(1/2)\Sigma_{ij} w_{ij} \mathrm{tr}(sQ_i sQ_j)} \times \mathrm{e}^{-N\Sigma_k \mathrm{tr}\ln(Q_k - \mathrm{i}sz)} DQ \qquad [13]$$

which is a rigorous version of the naive formula [9]. Compared to Fyodorov's formula, it has the disadvantage of not being manifestly invariant under global hyperbolic transformations $Q_i \mapsto TQ_iT^*$ (the integration domain $\mathscr{X}$ is not invariant). Its best feature is that it does apply to the case of band random matrices with one orbital per site ($N = 1$).

## Supersymmetric Variant

We are now in a position to tackle the problem of averaging ratios of determinants. For concreteness, we shall discuss the case where the number of determinants is two for both the numerator and the denominator, which is what is needed for the calculation of the function $G_{ij}^{(2)}(z_1, z_2)$ defined in eqn [2]. We will consider the case of relevance for the electrical conductivity: $z_1 = E + \mathrm{i}\epsilon, z_2 = E - \mathrm{i}\epsilon$, with $E \in \mathbb{R}$ and $\epsilon > 0$.

A $Q$-integral formula for $G_{ij}^{(2)}(z_1, z_2)$ can be derived by combining the fermionic method for

$$\left\langle \det(z_1 - H)\det\left(z_2 - H + t_2 E_{ij}^{ab}\right) \right\rangle$$

with the Schäfer–Wegner bosonic formalism for

$$\left\langle \det^{-1}\left(z_1 - H - t_1 E_{ji}^{ba}\right)\det^{-1}(z_2 - H)\right\rangle$$

and eventually differentiating with respect to $t_1, t_2$ at $t_1 = t_2 = 0$ and summing over $a, b$; see the subsection "Green's functions from determinants." All steps are formally the same as before, but with traces and determinants replaced by their supersymmetric analogs. Having given a great many technical details in the last two sections, we now just present the final formula along with the necessary definitions and some indication of what are the new elements involved in the proof.

Let each of $Q_{\mathrm{BB}}, Q_{\mathrm{FF}}, Q_{\mathrm{BF}}$, and $Q_{\mathrm{FB}}$ stand for a $2 \times 2$ matrix. If the first two matrices have commuting entries and the last two anticommuting ones, they combine to a $4 \times 4$ supermatrix:

$$Q = \begin{pmatrix} Q_{\mathrm{BB}} & Q_{\mathrm{BF}} \\ Q_{\mathrm{FB}} & Q_{\mathrm{FF}} \end{pmatrix}$$

Relevant operations on supermatrices are the supertrace,

$$\mathrm{Str}Q = \mathrm{tr}Q_{\mathrm{BB}} - \mathrm{tr}Q_{\mathrm{FF}}$$

and the superdeterminant,

$$\mathrm{Sdet}Q = \frac{\det(Q_{\mathrm{BB}})}{\det(Q_{\mathrm{FF}} - Q_{\mathrm{FB}}{Q_{\mathrm{BB}}}^{-1}Q_{\mathrm{BF}})}$$

These are related by the identity $\mathrm{Sdet} = \exp \circ \mathrm{Str} \circ \ln$ whenever the superdeterminant exists and is nonzero.

In the process of applying the method described earlier, a supermatrix $Q_i$ gets introduced at every site $i$ of the lattice $\Lambda$. The domain of integration for each of the matrix blocks $(Q_i)_{\mathrm{BB}} (i = 1, \ldots, |\Lambda|)$ is taken to be the Schäfer–Wegner domain $X_\lambda^{1,1}$ (with some choice of $\lambda > 0$); the integration domain for each of the $(Q_i)_{\mathrm{FF}}$ is the space of Hermitian $2 \times 2$ matrices, as before.

Let $E_{\mathrm{BB}}^{11}$ be the $4 \times 4$ (super)matrix with unit entry in the upper-left corner and zeros elsewhere; similarly, $E_{\mathrm{FF}}^{22}$ has unity in the lower-right corner and zeros elsewhere. Putting $s = \mathrm{diag}(1, -1, 1, 1)$ and $z = \mathrm{diag}(z_1, z_2, z_1, z_2)$, the supersymmetric $Q$-integral formula for the generating function of $G_{ij}^{(2)}$ – obtained by combining the Schäfer–Wegner bosonic method with the fermionic variant – is written as

$$\left\langle \frac{\det(z_1 - H)\det(z_2 - H + t_2 E_{ij}^{ab})}{\det(z_1 - H - t_1 E_{ji}^{ba})\det(z_2 - H)} \right\rangle$$
$$= \int DQ \mathrm{e}^{(1/2)\Sigma_{kl} w_{kl} \mathrm{Str}(sQ_k s Q_l)}$$
$$\times\, \mathrm{e}^{-\mathrm{Str}\ln\left(\Sigma_{r,c}(Q_r - \mathrm{i}sz)\otimes E_{rr}^{cc} + \mathrm{i}t_1 E_{\mathrm{BB}}^{11}\otimes E_{ji}^{ba} - \mathrm{i}t_2 E_{\mathrm{FF}}^{22}\otimes E_{ij}^{ab}\right)} \qquad [14]$$

where the second supertrace includes a sum over sites and orbitals, and on setting $t_1 = t_2 = 0$ becomes

$$\mathrm{e}^{-N\Sigma_r \mathrm{Str}\ln(Q_r - \mathrm{i}sz)} = \prod_r \mathrm{Sdet}^{-N}(Q_r - \mathrm{i}sz)$$

The superintegral "measure" $DQ = \prod_r DQ_r$ is the flat Berezin form, that is, the product of differentials for all the commuting matrix entries in $(Q_r)_{\mathrm{BB}}$ and $(Q_r)_{\mathrm{FF}}$, times the product of derivatives for all the anticommuting matrix entries in $(Q_r)_{\mathrm{BF}}$ and $(Q_r)_{\mathrm{FB}}$.

To prove the formula [14], two new tools are needed, a brief account of which is as follows.

### Gaussian Superintegrals

There exists a supersymmetric generalization of the Gaussian integration formulas given in the subsection "Determinants as Gaussian integrals": if $A, D(B, C)$ are linear operators or matrices with commuting (resp., anticommuting) entries, and $\mathfrak{Re}A > 0$, one has

$$\mathrm{Sdet}^{-1}\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \int \mathrm{e}^{-(\bar\varphi, A\varphi) - (\bar\varphi, B\psi) - (\bar\psi, C\varphi) - (\bar\psi, D\psi)}$$

Verification of this formula is straightforward. Using it, one writes the last factor in [14] as a Gaussian superintegral over four vectors: $\varphi_1, \varphi_2, \psi_1$, and $\psi_2$. The integrand then becomes Gaussian in the matrices $Q_r$.

### Shifting Variables

The next step in the proof is to do the "Gaussian" integral over the supermatrices $Q_r$. By definition, in a superintegral, one first carries out the Fermi integral, and afterwards the ordinary integrations. The Gaussian integral over the anticommuting parts $(Q_r)_{\mathrm{BF}}$ and $(Q_r)_{\mathrm{FB}}$ is readily done by completing the square and shifting variables using the fact that fermionic integration is differentiation:

$$\int \mathrm{d}\xi f(\xi - \xi') = \frac{\partial}{\partial \xi} f(\xi - \xi') = \int \mathrm{d}\xi f(\xi)$$

Similarly, the Gaussian integral over the Hermitian matrices $(Q_r)_{\mathrm{FF}}$ is done by completing the square and shifting. The integral over $(Q_r)_{\mathrm{BB}}$, however, is not Gaussian, as the domain is not $\mathbb{R}^n$ but the Schäfer–Wegner domain. Here, more advanced calculus is required: these integrations are done by using a supersymmetric change-of-variables theorem due to Berezin to make the necessary shifts by nilpotents. (There is not enough space to describe this here, so please consult Berezin's (1987) book.) Without difficulty, one finds the result to agree with the left-hand side of eqn [14], thereby establishing that formula.

## Approximations

All manipulations so far have been exact and, in fact, rigorous (or can be made so with little extra effort). Now we turn to a sequence of approximations that have been used by physicists to develop a quantitative understanding of weakly disordered quantum dots, wires, films, etc. While physically satisfactory, not all of these approximations are under full mathematical control. We will briefly comment on their validity as we go along.

### Saddle-Point Manifold

We continue to consider $G^{(2)}_{kl}(E+\mathrm{i}\varepsilon, E-\mathrm{i}\varepsilon)$ and focus on $E=0$ (the center of the energy band) for simplicity. By varying the exponent on the right-hand side of [14] and setting the variation to zero one obtains, for $t_1 = t_2 = 0$,

$$\sum_j w_{ij} s Q_j s - N Q_i^{-1} = 0$$

which is called the saddle-point equation.

Let us now assume translational invariance, $w_{ij} = f(|i-j|)$. Then, if $\lambda = \sqrt{N/\Sigma_j w_{ij}}$, the saddle-point equation has $i$-independent solutions of the form

$$Q_i = \lambda \begin{pmatrix} q_{\mathrm{BB}} & 0 \\ 0 & q_{\mathrm{FF}} \end{pmatrix}$$

where for $q_{\mathrm{FF}}$ there are three possibilities: two isolated points $q_{\mathrm{FF}} = \pm\mathbf{1}$ (unit matrix) coexist with a manifold

$$q_{\mathrm{FF}} = \begin{pmatrix} \cos\theta_1 & \sin\theta_1\, \mathrm{e}^{\mathrm{i}\phi_1} \\ \sin\theta_1\, \mathrm{e}^{-\mathrm{i}\phi_1} & -\cos\theta_1 \end{pmatrix} \qquad [15]$$

which is two-dimensional; whereas the solution space for $q_{\mathrm{BB}}$ consists of a single connected 2-manifold:

$$q_{\mathrm{BB}} = \begin{pmatrix} \cosh\theta_0 & \sinh\theta_0\, \mathrm{e}^{\mathrm{i}\phi_0} \\ \sinh\theta_0\, \mathrm{e}^{-\mathrm{i}\phi_0} & \cosh\theta_0 \end{pmatrix} \qquad [16]$$

The solutions $q_{\mathrm{FF}} = \pm\mathbf{1}$ are usually discarded in the physics literature. (The argument is that they break supersymmetry and therefore get suppressed by fermionic zero modes. For the simpler case of the one-point function [1] and in three space dimensions, such suppression has recently been proved by Disertori, Pinson, and Spencer.) Other solutions for $q_{\mathrm{BB}}$ are ruled out by the requirement $\mathfrak{Re}\, Q_i > 0$ for the Schäfer–Wegner domain.

The set of matrices [16] and [15] – the "saddle-point manifold" – is diffeomorphic to the product of a 2-hyperboloid $\mathrm{H}^2$ with a 2-sphere $\mathrm{S}^2$. Moving along that manifold $M := \mathrm{H}^2 \times \mathrm{S}^2$ leaves the $Q$-field integrand [14] unchanged (for $z_1 = z_2 = t_1 = t_2 = 0$).

One can actually anticipate the existence of such a manifold from the symmetries at hand. These are most transparent in the starting point of the formalism as given by the characteristic function $\langle \mathrm{e}^{-\mathrm{i}K_H} \rangle$ with

$$K_H = (\bar\varphi_1, H\varphi_1) - (\bar\varphi_2, H\varphi_2) + (\bar\psi_1, H\psi_1) + (\bar\psi_2, H\psi_2)$$

The signs of this quadratic expression are what is encoded in the signature matrix $s = \mathrm{diag}(1, -1, 1, 1)$ (recall that the first two entries are forced by $\mathfrak{Im}\, z_1 > 0$ and $\mathfrak{Im}\, z_2 < 0$). The Hermitian form $K_H$ is invariant under the product of two Lie groups: $\mathrm{U}(1,1)$ acting on the $\varphi$'s, and $\mathrm{U}(2)$ acting on the $\psi$'s. This invariance gets transferred by the formalism to the $Q$-side; the saddle-point manifold $M$ is in fact an "orbit" of the group action of $G := \mathrm{U}(1,1) \times \mathrm{U}(2)$ on the $Q$-field. In the language of physics, the degrees of freedom of $M$ correspond to the Goldstone bosons of a broken symmetry.

$K_{\mathrm{H}}$ also has some supersymmetries, mixing $\varphi$'s with $\psi$'s. At the infinitesimal level, these combine with the generators of $G$ to give a Lie superalgebra of symmetries $\mathfrak{g} := \mathfrak{u}(1,1|2)$. One therefore expects some kind of saddle-point supermanifold, say $\mathcal{M}$, on the $Q$-side.

$\mathcal{M}$ can be constructed by extending the above solution $q_0 := \mathrm{diag}(q_{\mathrm{BB}}, q_{\mathrm{FF}})$ of the dimensionless saddle-point equation $sqs = q^{-1}$ to the full $4 \times 4$ supermatrix space. Putting $q = q_0 + q_1$ with

$$q_1 = \begin{pmatrix} 0 & q_{\mathrm{BF}} \\ q_{\mathrm{FB}} & 0 \end{pmatrix}$$

and linearizing in $q_1$, one gets

$$s q_1 s = -q_0^{-1} q_1 q_0^{-1} \qquad [17]$$

The solution space of this linear equation for $q_1$ has dimension 4 for all $q_0 \in M$. Based on it, one expects four Goldstone fermions to emerge along with the four Goldstone bosons of $M$.

For the simple case under consideration, one can introduce local coordinates and push the analysis to nonlinear order, but things get quickly out of hand (when done in this way) for more challenging, higher-rank cases. Fortunately, there exists an alternative, coordinate-independent approach, as the mathematical object to be constructed is completely determined by symmetry!

### Riemannian Symmetric Superspace

The linear equation [17] associates with every point $x \in M$ a four-dimensional vector space of solutions,

$V_x$. As the point $x$ moves on $M$ the vector spaces $V_x$ turn and twist; thus, they form what is called a vector bundle $V$ over $M$. (The bundle at hand turns out to be nontrivial, i.e., there exists no global choice of coordinates for it.)

A section of $V$ is a smooth mapping $\nu : M \to V$ such that $\nu(x) \in V_x$ for all $x \in M$. The sections of $V$ are to be multiplied in the exterior sense, as they represent anticommuting degrees of freedom; hence the proper object to consider is the exterior bundle, $\wedge V$.

It is a beautiful fact that there exists a unique action of the Lie superalgebra $\mathfrak{g}$ on the sections of $\wedge V$ by first-order differential operators, or derivations for short. (Be advised however that this canonical $\mathfrak{g}$-action is not well known in physics or mathematics.)

The manifold $M$ is a symmetric space, that is, a Riemannian manifold with $G$-invariant geometry. Its metric tensor, $g$, uniquely extends to a second-rank tensor field (still denoted by $g$) which maps pairs of derivations of $\wedge V$ to sections of $\wedge V$, and is invariant with respect to the $\mathfrak{g}$-action. This collection of objects – the symmetric space $M$, the exterior bundle $\wedge V$ over it, the action of the Lie superalgebra $\mathfrak{g}$ on the sections of $\wedge V$, and the $\mathfrak{g}$-invariant second-rank tensor $g$ – form what the author calls a "Riemannian symmetric superspace," $\mathcal{M}$.

#### Nonlinear Sigma Model

According to the Landau–Ginzburg–Wilson (LGW) paradigm of the theory of phase transitions, the large-scale physics of a statistical mechanical system near criticality is expected to be controlled by an effective field theory for the long-wavelength excitations of the order parameter of the system.

Wegner is credited for the profound insight that the LGW paradigm applies to the random matrix situation at hand, with the role of the order parameter being taken by the matrix $Q$. He argued that transport observables (such as the electrical conductivity) are governed by slow spatial variations of the $Q$-field inside the saddle-point manifold. Efetov skilfully implemented this insight in a supersymmetric variant of Wegner's method.

While the direct construction of the effective continuum field theory by gradient expansion of [14] is not an entirely easy task, the outcome of the calculation is predetermined by symmetry. On general grounds, the effective field theory has to be a nonlinear sigma model for the Goldstone bosons and fermions of $\mathcal{M}$: if $\{\phi^A\}$ are local coordinates for the bundle $V$ with metric $g_{AB}(\phi) = g(\partial/\partial\phi^A, \partial/\partial\phi^B)$, the action functional is

$$S = \sigma \int \mathrm{d}^d x \partial_\mu \phi^A g_{AB}(\phi) \partial_\mu \phi^B$$

The coupling parameter $\sigma$ has the physical meaning of bare (i.e., unrenormalized) conductivity. In the present model $\sigma = NW^2 a^{2-d}$, where $W$ is essentially the width of the band random matrix in units of the lattice spacing $a$ (the short-distance cutoff of the continuum field theory). $S$ is the effective action in the limit $z_1 = z_2$. For a finite frequency $\omega = z_1 - z_2$, a symmetry-breaking term of the form $\mathrm{i}\omega\nu \int \mathrm{d}^d x f(\phi)$, where $\nu = N(\pi\lambda)^{-1} a^{-d}$ is the local density of states, has to be added to $S$.

By perturbative renormalization group analysis, that is, by integrating out the rapid field fluctuations, one finds for $d = 2$ that $\sigma$ decreases on increasing the cutoff $a$. This property is referred to as "asymptotic freedom" in field theory. On its basis one expects exponentially decaying correlations, and hence localization of all states, in two dimensions. However, a mathematical proof of this conjecture is not currently available.

In three dimensions and for a sufficiently large bare conductivity, the renormalization flow goes toward the metallic fixed point ($\sigma \to \infty$), where $G$-symmetry is broken spontaneously. A rigorous proof of this important conjecture (existence of disordered metals in three space dimensions) is not available either.

#### Zero-Mode Approximation

For a system in a box of linear size $L$, the cost of exciting fluctuations in the sigma model field is estimated as the Thouless energy $E_{\mathrm{Th}} = \sigma/\nu L^2$. In the limit of small frequency, $|\omega| \ll E_{\mathrm{Th}}$, the physical behavior is dominated by the constant modes $\phi^A(x) = \phi^A$ (independent of $x$). By computing the integral over these modes, Efetov found the energy-level correlations in the small-frequency limit to be those of the GUE.

*See also:* Random Matrix Theory in Physics; Symmetry Classes in Random Matrix Theory.

## Further Reading


Berezin FA (1987) *Introduction to Superanalysis*. Dordrecht: Reidel.

Disertori M, Pinson H, and Spencer T (2002) Density of states of random band matrices. *Communication in Mathematical Physics* 232: 83–124.

Efetov KB (1997) *Supersymmetry in Disorder and Chaos*. Cambridge: Cambridge University Press.

Fyodorov YV (2002) Negative moments of characteristic polynomials of random matrices: Ingham–Siegel integral as an

alternative to Hubbard–Stratonovich transformation. *Nuclear Physics B* 621: 643–674.

Mirlin AD (2000) Statistics of energy levels and eigen functions in disordered systems. *Physics Reports* 326: 260–382.

Schäfer L and Wegner F (1980) Disordered system with *n* orbitals per site: Lagrange formulation, hyperbolic symmetry, and Goldstone modes. *Zeitschrift für Physik B* 38: 113–126.

Wegner F (1979) The mobility edge problem: continuous symmetry and a conjecture. *Zeitschrift für Physics B* 35: 207–210.

Zirnbauer MR (1996) Riemannian symmetric superspaces and their origin in random matrix theory. *Journal of Mathematical Physics* 37: 4986–5018.

# Symmetric Hyperbolic Systems and Shock Waves

**S Kichenassamy**, Université de Reims Champagne-Ardenne, Reims, France



## Introduction

Many systems of partial differential equations arising in mathematical physics and differential geometry are quasilinear: the top-order derivatives enter only linearly. They may be cast in the form of first-order systems by introducing, if needed, derivatives of the unknowns as additional unknowns. For such systems, the theory of symmetric–hyperbolic (SH) systems provides a unified framework for proving the local existence of smooth solutions if the initial data are smooth. It is also convenient for constructing numerical schemes, and for studying shock waves. Despite what the name suggests, the impact of the theory of SH systems is not limited to hyperbolic problems, two examples being Tricomi's equation, and equations of Cauchy–Kowalewska type.

Application of the SH framework usually requires a preliminary reduction to SH form ("symmetrization").

After comparing briefly the theory of SH systems with other functional-analytic approaches, we collect basic definitions and notation. We then present two general rules, for symmetrizing conservation laws and strictly hyperbolic equations, respectively. We next turn to special features possessed by linear SH systems, and give a general procedure to prove existence, which covers both linear and nonlinear systems. We then summarize those results on shock waves, and on blow-up singularities, which are related to SH structure. Examples and applications are collected in the last section.

The advantages of SH theory are: a standardized procedure for constructing solutions; the availability of standard numerical schemes; a natural way to prove that the speed of propagation of support is finite. On the other hand, the symmetrization process is sometimes *ad hoc*, and does not respect the physical or geometric nature of the unknowns; to obviate this defect to some extent, we remark that symmetrizers may be viewed as introducing a new Riemannian metric on the space of unknowns. The search for a comprehensive criterion for identifying equations and boundary conditions compatible with SH structure is still the object of current research.

The most important fields of application of the theory today are general relativity and fluid dynamics, including magnetohydrodynamics.

### Context of SH Theory in Modern Terms

The basic reason why the theory works may be summarized as follows for the modern reader; the history of the subject is, however, more involved.

Let $H$ be a real Hilbert space. Consider a linear initial-value problem $\mathrm{d}u/\mathrm{d}t + Au = 0; u(0) = u_0 \in H$, where $A$ is unbounded, with domain $D(A)$. By Stone's theorem, one can solve it in a generalized sense, if the unbounded operator $A$ satisfies $A + A^* = 0$. This condition contains two ingredients: a symmetry condition on $A$, and a maximality condition on $D(A)$, which incorporate boundary conditions (von Neumann, Friedrichs). Semigroup theory (Hille and Yosida, Phillips, and many others) handles more general operators $A$: it is possible to solve this problem in the form $u(t) = S(t)u_0$ for $t > 0$, where $\{S(t)\}_{t\geq 0}$ is a continuous contraction semigroup, if and only if $(Au, u) \geq 0$, and equation $x + Ax = y$ has a solution for every $y$ in $H$ (this is a maximality condition on $D(A)$). One then says that $A$ is maximal monotone. For such operators, $A + A^* \geq 0$. SH systems are systems $Qu_t + Au = F$, satisfying two algebraic conditions ensuring formally that $A + A^*$ is bounded, and that $Q$ is symmetric and positive definite. This algebraic structure enables one to solve the problem directly, without explicit reference to semigroup theory. Precise definitions are given next.

We assume throughout that all coefficients, nonlinearities, and data are smooth unless otherwise specified.

### Definitions

Consider a quasilinear system

$$M_B^{A\alpha}(x,u)\partial_\alpha u^B = N^A(x,u) \qquad [1]$$

where $u=(u^A)_{A=1,\dots,m}$, $x=(x^\alpha)_{\alpha=0,\dots,n}$, and $\partial_\alpha = \partial/\partial x^\alpha$. The components of $u$ may be real or complex. We follow the summation convention on repeated indices in different positions; $x^0=t$ may be thought of as the evolution variable; we write $x=(t,\boldsymbol{x})$, with $\boldsymbol{x}=(x^1,\dots,x^n)$. Indices $A$, $B,\dots$ run from 1 to $m$, indices $j,k,\dots$ from 1 to $n$, and Greek indices from 0 to $n$. The complex conjugate of $u^A$ is written $\bar{u}^A$.

- Equation [1] is symmetrizable if there are functions $\sigma_{AB}(x,u)$ such that

$$M_{AB}^\alpha := \sigma_{AC}M_B^{C\alpha}$$

  satisfies the condition $M_{AB}^\alpha = \bar{M}_{BA}^\alpha$ for every $\alpha$.
- It is symmetric if it is symmetrizable with $\sigma_{AB}=\delta_{AB}$.
- It is symmetric-hyperbolic with respect to $k_\alpha$ if it is symmetric and if $k_\alpha M_{AB}^\alpha$ is positive definite: $k_\alpha M_{AB}^\alpha \xi^A\xi^B > 0$ for $\xi=(\xi^A)\neq 0$.

Thus, a symmetrizer $(\sigma_{AB})$ gives rise to a Riemannian metric $(k_\alpha\sigma_{AC}M_B^{C\alpha})$ on the space of unknowns, independent of any Riemannian structure on $x$-space. The system is SH with respect to $x^0$ if $k_\alpha=\delta_\alpha^0$.

The simplest class of SH systems is provided by real semilinear systems of the form

$$A^0(\boldsymbol{x})\partial_t u + A^i(\boldsymbol{x})\partial_i u = N(x,u) \qquad [2]$$

where the $A^\alpha$ are real symmetric matrices, $A^0$ is symmetric and positive definite, and $k_\alpha=\delta_{\alpha 0}$. Writing $A^0=P^2$, with $P$ symmetric and positive definite, one finds that $v=Pu$ solves a SH system with $A^0=I$ (identity matrix).

Conservation laws (with "reaction" or "source" term $N^A$) are usually defined as quasilinear systems of the form

$$\partial_\alpha f^{A\alpha}(x,u) = N^A(x,u) \qquad [3]$$

They are common in fluid dynamics and combustion. They are limiting cases of nonlinear diffusion equations of the typical form

$$\partial_\alpha f^{A\alpha}(x,u) = N^A(x,u) + \varepsilon\partial_j(B_B^{Ajk}\partial_k u^B) \qquad [4]$$

The determination of the form of the coefficients $B_B^{Ajk}$ is a nontrivial modeling issue; they may reflect varied physical processes such as heat conduction, viscosity, or bulk viscosity. They may depend on $x$, $u$, and the derivatives of $u$. The simplest case is $B_B^{Ajk}=D^{jk}\delta_B^A$ with $(D^{jk})$ diagonal. Some authors require the symmetry condition

$$\delta_{AC}B_B^{Cjk} = \delta_{BC}B_A^{Ckj} \qquad [5]$$

Equations in which $f^{A\alpha}=u^A\delta_0^\alpha$ are called reaction–diffusion equations; they arise in physical and biological problems in which chemical reactions and diffusion phenomena are combined, and in population dynamics.

A conservation law is symmetric if and only if $\partial f^{A\alpha}/\partial u^B$ is symmetric in $A$ and $B$, which means that there are, locally, functions $g^\alpha(x,u)$ such that $f^{A\alpha}=\partial g^\alpha/\partial u^A$.

A more fundamental derivation of conservation laws would take us beyond the scope of this survey.

## Symmetrization

Two general procedures for symmetrization are available: one for conservation laws, the other for semilinear strictly hyperbolic problems.

### Conservation Laws with a Convex Entropy

Consider, for simplicity, a conservation law of the form

$$\partial_t u^A + \partial_j f^{Aj}(u) = 0 \qquad [6]$$

We, therefore, assume that the $f^{A\alpha}=f^{A\alpha}(u)$ and $f^{A0}(u)=u^A$. We show that the following three statements are equivalent locally: (1) there is a strictly convex function $U(u)$ such that $\sigma_{AB}=\partial^2 U/\partial u^A\partial u^B$ is a symmetrizer; (2) eqn [6] implies a scalar relation of the form $\partial_\alpha U^\alpha=0$, with $U^0$ strictly convex; and (3) there is a change of unknowns $v_A=v_A(u)$ such that the system satisfied by $v=(v_A)$ is SH and $(\partial v_A/\partial u^B)$ is positive definite.

In fluid dynamics, $U^0$ may sometimes be related to specific entropy, and $U^j$ to entropy flux. For this reason, if (2) holds, one says that $U^0$ is an entropy for eqn [6], and that $(U^0,U^j)$ is an entropy pair. A system may have several entropies in this sense; this fact is sometimes useful in studying convergence properties of approximate solutions of eqn [6].

Let us now prove the equivalence of these properties.

Assume first (3): there are new unknowns $v_A=v_A(u)$ and functions $g^\alpha(v)$ such that $f^{A\alpha}=\partial g^\alpha/\partial v_A$. One finds that if eqn [6] holds,

$$\partial_\alpha U^\alpha = 0 \quad \text{where } U^\alpha = v_A\frac{\partial g^\alpha}{\partial v_A} - g^\alpha \qquad [7]$$

Furthermore, we have $f^{A0}=u^A$; therefore, eqn [7] gives: $U^0=v_Au^A-g^0$, so that $U^0$ is the Legendre

transform (familiar from mechanics) of $g^0$. It follows that $v_A = \partial U^0/\partial u^A$. Finally, $(\partial v_A/\partial u^B) = (\partial^2 U^0/\partial u^A \partial u^B)$ is positive definite, and $U^0$ is strictly convex.

We have proved that (3) implies (2). Next, assume (2): the entropy equality $U_t + \partial_j U^j = 0$ holds identically – and not just for the solution at hand. Using [6], we find

$$0 = \frac{\partial U}{\partial u^A}\partial_t u^A + \frac{\partial U^j}{\partial u^B}\partial_j u^B = \left[-\frac{\partial U}{\partial u^A}\frac{\partial f^{Aj}}{\partial u^B} + \frac{\partial U^j}{\partial u^B}\right]\partial_j u^B$$

Assumption (2), therefore, means that $U$ is strictly convex and satisfies

$$\frac{\partial U}{\partial u^A}\frac{\partial f^{Aj}}{\partial u^B} = \frac{\partial U^j}{\partial u^B} \qquad [8]$$

Now, letting $v_A = \partial U/\partial u^A$ and $g^j(v) = v_A f^{Aj} - U^j$, we find

$$\frac{\partial g^j}{\partial v_A} = f^{Aj} + \left[\frac{\partial U}{\partial u^A}\frac{\partial f^{Aj}}{\partial u^C} - \frac{\partial U^j}{\partial u^C}\right]\frac{\partial u^C}{\partial v_A} = f^{Aj} \qquad [9]$$

Let $\sigma_{AB} = \partial^2 U/\partial u^A \partial u^B$. Since $U$ is strictly convex, $(\sigma_{AB})$ is positive definite, and so is its inverse. We have now proved (3). Note that $u^A = \partial g^0/\partial v_A$, where $g^0(v) = u^A v_A - U(u)$ is the Legendre transform of $U$.

Next, using eqn [9], and the relations $\sigma_{AB} = \partial v_A/\partial u^B = \partial v_B/\partial u^A$, we find

$$\begin{aligned} 0 &= \sigma_{AB}[\partial_t u^B + \partial_j f^{Bj}] \\ &= \sigma_{AB}\partial_t u^B + \frac{\partial v_B}{\partial u^A}\frac{\partial^2 g^j}{\partial v^B \partial u^C}\partial_j u^C \\ &= \sigma_{AB}\partial_t u^B + \frac{\partial^2 g^j}{\partial u^A \partial u^C}\partial_j u^C \end{aligned}$$

which is SH; therefore, $\sigma_{AB}$ is a symmetrizer for eqn [6], and (1) is proved. Thus, (2) implies (1) and (3).

Finally, if (1) holds, $\sigma_{AC}\partial f^{Cj}/\partial u^B$ is symmetric in $A$ and $B$. It follows that

$$\frac{\partial}{\partial u^C}\left[\frac{\partial U}{\partial u^A}\frac{\partial f^{Aj}}{\partial u^B}\right] = \sigma_{AC}\frac{\partial f^{Aj}}{\partial u^B} + \frac{\partial U}{\partial u^A}\frac{\partial^2 f^{Aj}}{\partial u^B \partial u^C}$$

is symmetric in $B$ and $C$, so that there are, locally, functions $U^j$ such that eqn [8] holds. Therefore, $(U, U^j)$ is an entropy pair, and we see that (1) implies (2).

This completes the proof of the equivalence of (1), (2), and (3).

### Strictly Hyperbolic Equations

Consider the scalar equation $Pf = g(t, \boldsymbol{x})$, where $P$ is the linear operator

$$P = \partial_t^N - \sum_{j=0}^{N-1} p_{N-j}(t, \boldsymbol{x})\partial_t^j$$

of order $N$. Let $\Lambda = (1 - \Delta)^{1/2}$, where $\Delta$ is the Laplace operator on the space variables. Then $u = (u^A)$, where $u^A = \partial_t^{A-1}\Lambda^{N-A} f$ for $A = 1, \ldots, N$, solves a first-order pseudodifferential system of the form

$$u_t - Lu = G$$

If $P$ is strictly hyperbolic, the principal symbol $a_1(t, \boldsymbol{x}, \xi)$ of $L$ has a diagonal form with real eigenvalues $\lambda_j(t, \boldsymbol{x}, \xi)$, and there are projectors $p_j(t, \boldsymbol{x}, \xi)(p_j^2 = p_j)$ which commute with $a_1$, such that $1 = \sum_j p_j$, and $a_1 = \sum_j \lambda_j p_j$. Let $r_0 = \sum_j p_j^* p_j$, and $r_0(D)$ the corresponding operator. Equation

$$r_0(D)\partial_t u - r_0(D)Lu = r_0(D)G$$

is formally SH in the following sense: $r_0$ is positive definite and $r_0 a_1$ is Hermitian.

## Linear Problems

Consider a linear system

$$Lu := Q(t, \boldsymbol{x})\partial_t u + A^j(t, \boldsymbol{x})\partial_j u + B(t, \boldsymbol{x})u = f(t, \boldsymbol{x}) \qquad [10]$$

We assume that $Q$ and the $A^j$ are real and symmetric, $Q \geq c$ with $c$ positive, and all coefficients and their first-order derivatives are bounded.

### Energy Identity

Multiplying the equation by $u^{\mathrm{T}}$ (transpose of $u$), one derives the "energy identity"

$$\partial_t(u^{\mathrm{T}}Qu) + \partial_j(u^{\mathrm{T}}A^j u) + u^{\mathrm{T}}Cu = 2u^{\mathrm{T}}f(t, \boldsymbol{x}) \qquad [11]$$

where $C = 2B - \partial_t Q - \partial_j A^j$. $C$ is not necessarily positive. However, $v := u\exp(-\lambda t)$ satisfies a linear SH system for which $C$ is positive definite if $\lambda$ is large enough.

### Propagation of Support

A basic property of wave-like equations is finite speed of propagation of support: if the right-hand side vanishes, and if the solution at time 0 is localized in the ball of radius $r$, then the solution at time $t$ is localized in the ball of radius $r + ct$ for a suitable constant $c$.

This property also holds for SH systems. To see this, let us consider the set where a solution $u$

vanishes: if the initial condition vanishes for $|\boldsymbol{x}| \leq R$, we claim that $u$ at some later time vanishes for $|\boldsymbol{x}| \leq R - t/a$, for $a$ large enough.

Indeed, let us integrate the energy identity on a truncated cone $\Gamma := \{|\boldsymbol{x}| \leq a(t_0 - t)/t_0; 0 \leq t \leq t_1\}$ with $t_1 < t_0$. The boundary of $\Gamma$ consists of three parts: $\partial\Gamma = \Omega_0 \cup \Omega_1 \cup S$, where $\Omega_0$ and $\Omega_1$ represent the portions of the boundary on which $t = 0$ and $t_1$, respectively. The outer normal to $S$ is proportional to $(a, t_0 x^j/|\boldsymbol{x}|)$. Let $E(s)$ denote the integral of $u^{\mathrm{T}}Qu$ on $\Gamma \cap \{t = s\}$. Integrating eqn [11] by parts, we obtain

$$E(t_1) - E(0) + \int_S u^{\mathrm{T}}\Phi u \,\mathrm{d}s = \iint_\Gamma (2u^{\mathrm{T}}f - u^{\mathrm{T}}Cu)\mathrm{d}t\,\mathrm{d}\boldsymbol{x} \qquad [12]$$

where $\Phi$ is proportional to $aQ + t_0 \sum_j x^j A^j/|\boldsymbol{x}|$. Take $a$ so large that $\Phi$ is positive definite. The integral over $S$ is then non-negative. If $C$ is positive definite and $f \equiv 0$, so that $E(0) = 0$, we find that $E(t_1) \leq 0$. Since $Q$ is positive definite, this implies $u \equiv 0$ on $\Omega_1$, as claimed.

### A Numerical Scheme

System $Lu = f$ may be discretized, for example, by the Lax–Friedrichs method: let $h$ be the discretization step in space, and $k$ the time step; write $\tau_j u(t, \boldsymbol{x}) = u(t, x^1, \ldots, x^j + h, \ldots, x^n)$ (translation in the $j$ direction). One replaces $\partial_j u$ by the centered difference in the $j$ direction: $(\tau_j u - \tau_j^{-1}u)/2h$; and the time derivative by

$$[u(t+k, \boldsymbol{x}) - \frac{1}{2n}\sum_j (\tau_j u(t, \boldsymbol{x}) + \tau_j^{-1}u(t, \boldsymbol{x}))]/k \qquad [13]$$

For consistency of the scheme, we require $k/h = \lambda > 0$ to be fixed as $k$ and $h$ tend to zero; stability then holds if $\lambda$ is small.

## Nonlinear Problems and Singularities

We give a simple setup for proving the existence of smooth solutions to SH systems for small times. Such solutions may develop singularities. We limit ourselves to two types of singularities, on which SH structure provides some information: jump discontinuities and blow-up patterns. Caustic formation is not considered.

### Construction of a Smooth Solution

Consider a real SH system (eqn [1]). Recall that a function of $\boldsymbol{x}$ belongs to the Sobolev space $H^s$ if its derivatives of order $s$ or less are square-integrable. One constructs a solution defined for $t$ small, which is in $H^s$, $s > n/2 + 1$, as a function of $\boldsymbol{x}$, by the following procedure:

(1) Replace spatial derivatives by regularized operators, which should be bounded in Sobolev spaces; the regularized equation is an ODE in $H^s$; let $u_\varepsilon$ be its solution.
(2) Write the equation satisfied by derivatives of order $s$ of $u_\varepsilon$, and apply the energy identity to it.
(3) Find a positive $T$ such that the solution is bounded in $H^s$ for $|t| \leq T$, uniformly in $\varepsilon$; this implies a $C^1$ bound.
(4) Prove the convergence of the approximations in $L^2$.
(5) Prove the continuity in time of the $H^s$ norm; conclude that the $u_\varepsilon$ tend to a solution in $C(-T, T; H^s)$.

The result admits a local version, in which Sobolev spaces are replaced by Kato's "uniformly local" spaces. Uniqueness of the solution is proved along similar lines. We do not attempt to identify the infimum of the values of $s$ for which the Cauchy problem is well-posed.

### Jump Discontinuities: Shock Waves

A "shock wave" is a weak solution of a system of conservation laws admitting a jump discontinuity. By definition, weak solutions satisfy, for any smooth function $\phi_A(x)$ with compact support,

$$\iint \{f^{A\alpha}\partial_\alpha \phi_A + N^A \phi_A\}\,\mathrm{d}t\,\mathrm{d}\boldsymbol{x} = 0$$

The theory of shock waves is an attempt to understand solutions of conservation laws which are limits of solutions of diffusion equations; the hope is that the influence of second-derivative terms is appreciable only near shocks, and that, for given initial data, there is a unique weak solution of the conservation law which may be obtained as such a limit, if modeling has been done correctly. This problem may be difficult already for a single shock ("shock structure").

The theory of shock waves follows the one-dimensional theory closely. We therefore describe the main facts for a conservation law in one space dimension ($u = u(t, x)$):

$$\partial_t u + \partial_x f(u) = 0$$

If a shock travels at speed $c$, the weak formulation of the equations gives the Rankine–Hugoniot relation $c[u] = [f(u)]$, where square brackets denote jumps. There may be several weak solutions having the same initial condition. One restricts solutions by

making two further requirements: (1) the system admits an entropy pair $(U, F)$ with a convex entropy and (2) to be admissible, weak solutions must be limits of "viscous approximations"

$$\partial_t u + \partial_x f(u) = \varepsilon \partial_x^2 u$$

as $\varepsilon \to 0$. One then finds easily that the entropy equality $(\partial_t U + \partial_x F = 0)$ must be replaced, for such weak solutions, by the entropy condition: $\partial_t U + \partial_x F \leq 0$ in the weak sense. This condition admits a concrete interpretation if the gradient of each characteristic speed is never orthogonal to the corresponding right eigenvector ("genuine nonlinearity"); in that case, characteristics must impinge on the shock ("shock inequalities").

For the equations of gas dynamics with polytropic law $(pv^\gamma = \text{const.})$, there is a unique solution with initial condition $u = u_l$ for $x < 0, u = u_r$ for $x > 0$, where $u_l$ and $u_r$ are constant ("Riemann problem") which satisfies the entropy condition, provided $|u_l - u_r|$ is small. More generally, if the equation of state $p = p(v, s) > 0$ satisfies $\partial p/\partial v < 0$ and $\partial^2 p/\partial v^2 > 0$, the shock inequalities are equivalent to the fact that the entropy increases after the passage of a shock with $|u_l - u_r|$ small.

On the numerical side, one should mention: (1) the widely used idea of upstream differencing; (2) the Lax–Wendroff scheme, the complete analysis of which requires tools from soliton theory; and (3) the availability of general results for dissipative schemes for SH systems.

Recent trends include: (1) admissibility conditions when genuine nonlinearity does not hold and (2) other approximations of shock wave problems, most notably kinetic formulations.

Some of the ideas of shock wave theory have been applied to Hamilton–Jacobi equations and to motion by mean curvature, with applications to front propagation problems and "computer vision."

### Stronger Singularities: Blow-Up Patterns

The amplitude of a solution may also grow without bound. Examples include optical pulse propagation in Kerr media and singularities in general relativity. The phenomenon is common when reaction terms are allowed. As we now explain, this phenomenon is reducible to SH theory in many cases of interest.

Blow-up singularities are usually not governed by the characteristic speeds defined by the principal part, because top-order derivatives are balanced by lower-order terms. In many applications, a systematic process (Fuchsian reduction) enables one to identify the correct model near blow-up; as a result, one can write the solution as the sum of a singular part, known in closed form, and a regular part. If the singularity locus is represented by $t = 0$, the regular part solves a renormalized equation of the typical form

$$tMu + Au = t^\varepsilon N \qquad [14]$$

where $Mu = 0$ is SH. Under natural conditions, for any initial condition $u_0$ such that $Au_0 = 0$, there is a unique solution of eqn [14] defined for small $t$.

The upshot is an asymptotic representation of solutions which renders the same services as an exact solution, and is valid precisely where numerical computation breaks down.

Fuchsian reduction enables one in particular to study (1) the blow-up time; (2) how the singularity locus varies when Cauchy data, prescribed in the smooth region, are varied; and (3) expressions which remain finite at blow-up. It is the only known general procedure for constructing analytically singular spacetimes involving arbitrary functions, rather than arbitrary parameters, and is therefore relevant to the search for alternatives to the big bang.

## Examples and Applications

### Wave Equation with Variable Coefficients

Consider the equation

$$\partial_{tt} u + 2a^j(x)\partial_{jt} u - a^{jk}(x)\partial_{jk} u = f(t, \boldsymbol{x}, u, \nabla u)$$

with $(a^{jk})$ positive definite. Letting $v = (v_0, \ldots, v_{n+1}) := (u, \partial_j u, \partial_t u)$, we find the system

$$\begin{aligned} \partial_t v_0 &= v_{n+1} \\ \partial_t v_k - \partial_k v_{n+1} &= 0 \\ \partial_t v_{n+1} + 2a^k \partial_k v_{n+1} - a^{jk} \partial_k v_j &= f \end{aligned}$$

It is symmetrizable, using the quadratic form $\sigma_{AB} v^A v^B = v_0^2 + a^{jk} v_j v_k + v_{n+1}^2$.

One proves directly that, if $v_j = \partial_j v_0$ for $t = 0$, this relation remains true for all $t$.

### Maxwell's Equations

Maxwell's equations may be split into six evolution equations: $\partial_t \boldsymbol{E} - \operatorname{curl} \boldsymbol{B} + \boldsymbol{j} = 0$ and $\partial_t \boldsymbol{B} + \operatorname{curl} \boldsymbol{E} = 0$, and two "constraints" $\operatorname{div} \boldsymbol{E} - \rho = 0, \operatorname{div} \boldsymbol{B} = 0$. The system of evolution equations is already in symmetric form; the quadratic form $\sigma_{AB} u^A u^B$ is here $|\boldsymbol{E}|^2 + |\boldsymbol{B}|^2$.

### Compressible Fluids

Consider first the case of a polytropic gas:

$$\begin{aligned} \partial_t \boldsymbol{v} + (\boldsymbol{v}\cdot\nabla)\boldsymbol{v} + \rho^{-1}\nabla p = 0 \\ \partial_t \rho + \operatorname{div}(\rho \boldsymbol{v}) = 0 \end{aligned} \qquad [15]$$

with $p$ proportional to $\rho^\gamma$. Taking $(p, \boldsymbol{v})$ as unknowns, one readily finds the SH system

$$\frac{1}{\gamma p}\partial_t p + \frac{1}{\gamma p}(\boldsymbol{v}\cdot\nabla)p + \operatorname{div}\boldsymbol{v} = 0 \qquad [16]$$

$$\rho\partial_t \boldsymbol{v} + \nabla p + \rho(\boldsymbol{v}\cdot\nabla)\boldsymbol{v} = 0 \qquad [17]$$

Symmetrization for more general compressible fluids with dissipation, including bulk viscosity, so as to satisfy the additional condition [5] may be achieved if we take as thermodynamic variables $\rho$ and $T$, and assume pressure $p$ and internal energy $\varepsilon$ satisfy $\partial p/\partial\rho > 0$ and $\partial\varepsilon/\partial T > 0$, by taking as unknowns $(\rho, \rho\boldsymbol{v}, \rho(\varepsilon + |\boldsymbol{v}|^2/2))$. The specific entropy $s$ satisfies $\mathrm{d}\varepsilon = T\mathrm{d}s - p\mathrm{d}(1/\rho)$. If the viscosity and heat conduction coefficients are positive, one finds that $U = -\rho s$ is a convex entropy (in the sense of SH theory) on the set where $\rho > 0, T > 0$.

### Einstein's Equations

The computation of solutions of Einstein's equations over long times, in particular in the study of coalescence of binary stars, has recently led to unexplained difficulties in the standard Arnowitt–Deser–Misner (ADM) formulation of the initial-value problem in general relativity. One way to tackle these difficulties is to rewrite the field equations in SH form; we focus on this particular aspect of recent research.

Recall the problem: find a four-dimensional metric $g_{ab}$ with Lorentzian signature, such that $R_{ab} - \frac{1}{2}Rg_{ab} = \chi T_{ab}$, with $\nabla^a T_{ab} = 0$, combined with an equation of state if necessary. $R_{ab}$ is the Ricci tensor and $R = g^{ab}R_{ab}$ is the scalar curvature; they depend on derivatives of the metric up to order 2. In addition to the metric, $T_{ab}$ involves physical quantities such as fluid 4-velocity or an electromagnetic field. The conservation laws of classical mathematical physics are all contained in the relation $\nabla^a T_{ab} = 0$.

Now, the field equations cannot be solved for $\partial_t^2 g_{ab}$, and, as a consequence, the Taylor series of $g_{ab}$ with respect to time cannot be determined, even formally, from the values of $g_{ab}$ and $\partial_t g_{ab}$ for $t = 0$ (i.e., the Cauchy data). Furthermore, these data must satisfy four *constraint equations*. If the constraints are satisfied initially, they "propagate." But in numerical computation, these constraints are never exactly satisfied, and the computed solution may deviate considerably from the exact solution. Also, numerical computations depend heavily on the way Einstein's equations are formulated.

The simplest way to derive a SH system is to replace $R_{ab}$ by $R_{ab}^{(h)} = R_{ab} - \frac{1}{2}[g_{bc}\partial_a F^c + g_{ac}\partial_b F^c]$, where $F^c := g^{ab}\Gamma^c_{ab}$. It turns out that $R_{ab}^{(h)} = -\frac{1}{2}g^{cd}\partial_{cd}g_{ab} + H_{ab}(g, \partial g)$, where the expression of $H_{ab}$ is immaterial. Applying to each component of the metric the treatment of the first example above (wave equation with variable coefficients), one easily derives an SH system of 50 equations for 50 unknowns: the ten independent components of the metric, and their 40 first-order derivatives. Now, if the $\Gamma^c$ are initially zero (coordinates are "harmonic"), they remain so at later times.

Unfortunately, the harmonic coordinate condition does not seem to be stable in the large. More recent formulations start with one of the standard setups (ADM formalism, conformal equations, tetrad formalism, Newman–Penrose formalism) and proceed by adding combinations of the constraints to the equations, multiplied by parameters adjusted so as to ensure hyperbolicity or symmetric–hyperbolicity if needed. Another recent idea is to add a new unknown $\lambda$ which monitors the failure of the constraint equations; one adds to the equations a new relation of the form $\partial_t\lambda = \alpha C - \beta\lambda$, where $C = 0$ is equivalent to the constraints, and $\alpha$ and $\beta$ are parameters. One then adds coupling terms to make the extended system SH. It is expected that the set of constraints acts as an attractor.

Reported computations indicate that these methods have resulted in an improvement of the time over which numerical computations are valid.

### Tricomi's Equation

Let $\varphi(x, y)$ solve $(y\partial_x^2 - \partial_y^2)\varphi = 0$. Letting $u = \mathrm{e}^{-\lambda x}(\partial_x\varphi, \partial_y\varphi)$, one finds a symmetric system $Lu = 0$, with

$$L = \begin{pmatrix} y & 0 \\ 0 & 1 \end{pmatrix}(\partial_x + \lambda) - \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\partial_y$$

If

$$Z = \begin{pmatrix} 1 & y \\ 1 & 1 \end{pmatrix}$$

we find that $K = ZL = A^1\partial_x + A^2\partial_y + B$, where

$$B - \tfrac{1}{2}(\partial_x A^1 + \partial_y A^2) = \begin{pmatrix} \frac{1}{2} + \lambda y & \lambda y \\ \lambda y & \lambda \end{pmatrix}$$

is positive definite if $y$ is bounded, of arbitrary sign, and $\lambda$ is small.

### Cauchy–Kowalewska Systems

Consider a complex system

$$\partial_t u = A^j(z,t,u)\frac{\partial u}{\partial z^j} + B(z,t,u) \qquad [18]$$

where $u = (u^A), z = (z^1, \ldots, z^n)$. The coefficients are analytic in their arguments when $z$ and $t$ are close to the origin and $u$ is bounded by some constant $K$. The Cauchy–Kowalewska theorem ensures that, for any analytic initial condition near the origin, this system has a unique analytic solution near $z = 0$, even without any symmetry assumption on the $A^j$. This result is a consequence of SH theory (Garabedian).

Indeed, write $z^j = x^j + iy^j$, $\partial_{z^j} = (1/2)(\partial_{x^j} - i\partial_{y^j})$, and $\partial_{\bar{z}_j} = (1/2)(\partial_{x^j} + i\partial_{y^j})$. Recall that analytic functions of $z$ satisfy the Cauchy–Riemann equations $\partial_{\bar{z}_j} u = 0$.

Adding $(\bar{A}^j)^{\mathrm{T}}\partial_{\bar{z}_j}$ to [18], and using the definition of $\partial_{z^j}$ and $\partial_{\bar{z}_j}$, we find the symmetric system

$$u_t = \frac{1}{2}(A^j + (\bar{A}^j)^{\mathrm{T}})\partial_{x^j} u + \frac{1}{2i}(A^j - (\bar{A}^j)^{\mathrm{T}})\partial_{y^j} u + B \qquad [19]$$

Solving this system, we find a candidate $u$ for a solution of eqn [18]. To show that $u$ is analytic if the data are, we solve a second SH system for $w = w^{(j)} := \partial_{\bar{z}_j} u$. If the data are analytic, $w$ vanishes initially, and therefore remains zero for all $t$. Therefore, $u$ is indeed analytic.

*See also:* Computational Methods in General Relativity: The Theory; Einstein Equations: Initial Value Formulation; Evolution Equations: Linear and Nonlinear; Magnetohydrodynamics; Partial Differential Equations: Some Examples; Semilinear Wave Equations; Shock Wave Refinement of the Friedman–Robertson–Walker Metric.

## Further Reading

Dautray R and Lions J (1988) *Mathematical Analysis and Numerical Methods for Science and Technology*, vols. 1–6. (transl. from French). New York: Springer.
Friedrichs KO (1986) In: Morawetz CS (ed.) *Selecta*, two volumes: collection of reprints with comments. Boston: Birkhäuser.
Godlewski E and Raviart P-A (1996) *Numerical Approximation of Hyperbolic Systems of Conservation Laws*. Berlin: Springer.
Gustafsson B, Kreiss H-O, and Oliger J (1996) *Time-Dependent Problems and Difference Methods*. New York: Wiley-Interscience.
John F (1982) *Partial Differential Equations*, 4th edn. Berlin: Springer.
Kato T (1975a) Quasilinear equations of evolution with applications to partial differential equations. *Lecture Notes in Mathematics* 448: 25–70.
Kato T (1975b) The Cauchy problem for quasilinear symmetric hyperbolic systems. *Archive for Rational Mechanics and Analysis* 58: 181–205.
Kawashima S and Shizuta Y (1988) On the normal form of the symmetric hyperbolic–parabolic systems associated with the conservation laws. *Tôhoku Mathematical Journal* 40: 449–464.
Kichenassamy S (1996) *Nonlinear Wave Equations*. New York: Dekker.
Lax PD (1973) Hyperbolic Conservation Laws and the Mathematical Theory of Shock Waves. *Regional Conference Series in Applied Mathematics*. Philadelphia: SIAM.
Lax PD and Phillips RS (1989) *Scattering Theory*, (revised edition). Boston: Academic Press.
Majda A (1984) *Compressible Fluid Flow and Systems of Conservation Laws*. Berlin: Springer.
Serre D (1999/2000) *Systems of Conservation Laws, I & II*. Cambridge: Cambridge University Press.
Smoller J (1983) *Shock Waves and Reaction-Diffusion Equations*. Berlin: Springer.
Taylor M (1991) Pseudodifferential Operators and Nonlinear P.D.E., *Progress in Mathematics*, vol. 100. Boston: Birkhäuser.

# Symmetries and Conservation Laws

**L H Ryder**, University of Kent, Canterbury, UK



## Introduction: Spacetime Symmetries

Symmetries have played, and continue to play, an important role in fundamental physics, but the part they play is today seen as more complicated and many-sided than it was in the early days of particle physics, just after the Second World War. The area in which symmetries have had their most dramatic consequences is elementary particle physics, or high-energy physics, and the majority of this article is concerned with this subject. The article concludes with some observations about symmetries and conservation laws in general relativity.

In the early days, considerations of symmetry were almost limited to Lorentz transformations: we begin by reviewing this crucially important topic. Invariance of the laws of nature under *translations* in space and time are actually necessary for the existence of science itself; if experiments did not yield the same results today and tomorrow, and in Paris and Moscow and on the Moon, then in effect there would be no laws of nature. Almost as strong

a statement could be made about invariance under *rotations*; if space were not isotropic, experimental results would depend on which direction the apparatus was aligned in, and again any laws would be extremely hard to find. Turning to the question of *motion*, Newton and Galileo realized that the laws of dynamics are the same in all inertial frames in relative motion. In the Newton–Galileo scheme, the rule for relating the space and time coordinates of two frames of reference is (for relative motion along the common $x$-axis)

$$x' = x - vt, \qquad t' = t \qquad [1]$$

This *principle of relativity* was reaffirmed by Einstein, but with the crucial modification that the rules for relating coordinates in two frames are given by Lorentz transformations, so that [1] is replaced by

$$x' = \gamma(x - vt), \qquad t' = \gamma\left(t - \frac{vx}{c^2}\right) \qquad [2]$$

Time is absolute in [1] but relative in [2]. Einstein was of course motivated by the fact that Maxwell's equations are covariant under Lorentz transformations, but not under Newton–Galileo ones.

The above considerations reveal that the laws of nature should be covariant under ten types of transformation: three translations in space, one in time, three parameters (angles) for rotations and three velocities. These transformations together form a group, the inhomogeneous Lorentz, or Poincaré group. It is a nonabelian group whose ten generators correspond to 4-momentum, angular momentum, and Lorentz boosts. The seminal work on the significance of this group in fundamental physics is that of Wigner in 1939. Assuming that the states of fundamental quantum systems (particles, atoms, molecules) form the basis states for *representations* of this group, these entities are described by two quantities, mass and spin. Spin, moreover, which was already familiar from earlier investigations in quantum physics, was described by the rotation group (SU(2), which is homomorphic to SO(3)) only for states with timelike momentum. For photons, for example, with null momentum, spin is described by the (noncompact) Euclidean group in the plane, with the consequence that there are only two polarization states for this massless particle.

Noether's theorem provides the crucial link between symmetries and *conservation laws*, via the principle of least action. Noether showed that the invariance of the action under a continuous symmetry operation implied the existence of a conserved quantity. The conserved quantities corresponding to invariance under translation in space and time are momentum and energy; conservation of angular momentum follows from invariance under rotations and invariance under Lorentz transformations gives rise to conservation of motion of the center of mass.

## Gauge Theories: Electromagnetism and Yang–Mills Theories

A quantity whose conservation has been well known for a long time is electric charge. The question may then be asked: invariance under what symmetry gives rise to conservation of electric charge? A classical complex field has the Lagrangian density

$$L = (\partial_\mu \phi)(\partial^\mu \phi^*) - m^2 \phi^* \phi \qquad [3]$$

which is invariant under

$$\phi \to \exp(-\mathrm{i}Q\,\Lambda)\phi \qquad [4]$$

$\Lambda$ being the parameter for the transformation. Noether's theorem then yields conservation of $Q$, interpreted as electric charge. With $\Lambda$ a constant, as above, the Lagrangian possesses a "global" symmetry. This becomes a "local" symmetry when $\Lambda$ becomes space and time dependent, $\Lambda(\boldsymbol{r}, t)$ or $\Lambda(x^\mu)$. In that case, however, the Lagrangian [3] is no longer invariant under [4], because of the derivative terms. To preserve invariance an extra field $A_\mu$ must be introduced, so that [4] then becomes

$$\begin{aligned} \phi &\to \exp(-\mathrm{i}Q\,\Lambda(x^\mu))\phi \\ A_\mu &\to A_\mu + \frac{1}{Q}\partial_\mu \Lambda \end{aligned} \qquad [5]$$

and the Lagrangian acquires extra terms, involving $A_\mu$. The field $A_\mu$ is called a *gauge field* and is identified with the electromagnetic potential. The transformation [5] is called a *gauge transformation*, and since the phase factor $\exp(-\mathrm{i}Q\,\Lambda)$ may be regarded as a unitary $1 \times 1$ matrix, we have here a theory with U(1) gauge invariance, which describes electromagnetism and conservation of charge.

The notion of isospin had been introduced by Heisenberg in 1932. Isospin (then called isotopic spin) was a vector-like quantity conserved in strong (nuclear) interactions. Yang and Mills in 1954 made the pioneering suggestion that isospin conservation could also be recast as a gauge theory, by enlarging the U(1) group of electromagnetism to SU(2) (corresponding to rotations in "isospin space"), and at the same time treating the rotation angles as functions of spacetime. Then, eqn [4] will change: if

for example ψ is an isospinor field, then local isospin rotations are given by

$$\psi(x) \to \exp\left\{-i\frac{\tau}{2}\cdot\theta(x)\right\}\psi(x) = U(x)\psi(x) \quad [6]$$

where τ are the Pauli matrices: τ/2 are the generators of SU(2). The gauge field then has three components $A^i_\mu (i = 1, 2, 3)$ which may be written as a matrix

$$\boldsymbol{A}_\mu = A^i_\mu \frac{\tau^i}{2}$$

transforming as

$$\boldsymbol{A}_\mu \to \boldsymbol{A}'_\mu = U(x)\boldsymbol{A}_\mu\, U^{-1}(x) - \frac{i}{g}(\partial_\mu\, U(x))U^{-1}(x) \quad [7]$$

where $g$ is the coupling constant, analogous to electric charge. The problem with this idea was that the isospin gauge field, analogous to the photon in electrodynamics, should, like the photon, be massless and have polarization states ±1 (commonly, but inaccurately – see the work of Wigner (1939) – called spin 1); whereas the Yukawa particle, identified as the $\pi$ meson, was massive and had spin 0, so could not act as the isospin gauge field.

The Yang–Mills idea really came into its own with the standard model (SM) of particle physics. This (gauge) model has an invariance group SU(2) ⊗ U(1) ⊗ SU(3), the first two groups corresponding to electroweak interactions (a unification of weak interactions and electromagnetism) and the final SU(3) to quantum chromodynamics (QCD), the gauge theory describing quark interactions, which "glues" them together to make hadrons – protons, neutrons, pions, etc. This model is a dramatically successful one. The QCD sector of the theory requires essentially no further elaboration on the Yang–Mills idea than replacing the group SU(2) by SU(3). This is a straightforward matter of replacing the generators τ/2 of SU(2) with the eight generators (3 × 3 matrices) of SU(3). U($x$) then also becomes a 3 × 3 matrix. The three degrees of freedom are the three quark "colors," for which there is good experimental evidence, and the gluons, the quanta of the gauge fields, are indeed massless and have good experimental support. In the electroweak sector, however, the gauge fields, the ***W*** and ***Z*** bosons, were found with the predicted masses of 80.3 and 91.2 GeV respectively (the proton mass, for comparison, is 0.98 GeV). They are certainly not massless, as the straightforward Yang–Mills theory would require, and the explanation for this requires the introduction of the concept of spontaneous symmetry breaking.

## Spontaneous Symmetry Breaking

The general idea of spontaneous symmetry breaking is that the *vacuum* – the state of lowest energy – is not invariant under the symmetry in question. A simple and common illustration is a pencil balanced vertically on its tip on a horizontal plane. The pencil is in unstable equilibrium but the system has a symmetry under rotations in the plane about the axis coincident with the pencil. Eventually, the pencil will fall into its lowest-energy state (vacuum), lying on the table in some direction – and the rotational symmetry is then lost. In fact, under rotations the actual lowest-energy (vacuum) state will be changed into another such state. There is a *degenerate vacuum*.

A similar scenario may be constructed in a complex scalar field theory. Consider such a theory with a Lagrangian given by

$$L = (\partial_\mu\phi)(\partial^\mu\phi^*) - m^2\phi^*\phi - \lambda(\phi^*\phi)^2 \quad [8]$$

that is, with a potential energy function given by

$$V(\phi, \phi^*) = m^2\phi^*\phi + \lambda(\phi^*\phi)^2 \quad [9]$$

where $m$ is the mass of the field (quantum) and $\lambda$ is the coupling of its self-interaction. The ground state is obtained by minimizing $V$, hence $\partial V/\partial\phi = 0$, giving (assuming that $m^2 > 0$) a minimum at $\phi = \phi^* = 0$. If, however, $m^2 < 0$, there is a local maximum at $\phi = 0$ and a minimum at $|\phi|^2 = -m^2/2\lambda > 0$. In quantum theory language, the vacuum expectation value $< 0|\phi|0 >$ of the field is nonzero. Goldstone showed that this implied the presence of a massless scalar particle – a Goldstone boson. There was some interest in this result in particle physics, where the hypothesis of "partial conservation of the axial vector current" (PCAC) might result in a Goldstone boson that could be identified with the pion; although not massless, the pion is the lightest hadron, so "almost" massless.

Higgs analyzed what happens to the Goldstone model if electromagnetism is included. The Lagrangian [8] is invariant under the global transformation [4], but if this is made local, as in [5], a gauge field must be introduced and it is found that the massless Goldstone boson disappears and the massless gauge field (photon) becomes massive. Thus, spontaneous symmetry breaking of a gauge theory results in the appearance of a massive, rather than massless, gauge particle. (It is relevant to remark that a massless photon possesses two polarization states, but a massive one possesses three, so the number of spin-polarization states is preserved – the massless photon "eats" the Goldstone boson and becomes massive.) The Higgs model was generalized to the

case of a nonabelian symmetry group by Guralnik, Hagen, and Kibble and invoked by Weinberg in his 1971 model for the electroweak interaction in which the gauge quanta were massive.

Higgs' work was motivated by the theory of superconductivity, where the Meissner effect (expulsion of magnetic flux from a superconductor), when relativistic, implies that the effective mass of a photon in a superconductor is nonzero – this is, the "reason" that the flux does not penetrate. In the theory of Bardeen, Cooper, and Schrieffer (BCS), a superconductor is described by an effective scalar field, a composite of electron pairs (though paired in momentum space rather than coordinate space), and this provides a physical analogy with the model above. The SM of particle physics postulates a Higgs scalar field analogous to the BCS composite scalar field. If this field exists, Higgs particles should also exist, but they have not yet been found. This is an outstanding problem for the SM.

## Baryon and Lepton Numbers

The fact that the proton $p$ does not decay into positron plus photon, $e^+ + \gamma$, or muon plus photon, $\mu^+ + \gamma$, implies a conservation law of baryon number $B$ (the proton possessing $B = 1$ and the others $B = 0$). Furthermore, the stability of $\mu^-$ and $\tau^-$ against decay into $e^- + \gamma$ implies conservation of lepton numbers $L_e$, $L_\mu$, and $L_\tau$. These are regarded as global, not local, symmetries, so there are no associated gauge fields or interactions. Interestingly, however, these symmetries are not built into the SM, so are not guaranteed by it. More interestingly, these symmetries are actually destroyed in one attempt to go beyond the SM. This is the hypothesis that QCD may be unified with electroweak interactions to produce a "grand unified" theory (GUT). The simplest GUT is the one in which the SU(2) $\otimes$ U(1) $\otimes$ SU(3) symmetry is assumed to be a subgroup of the much tighter symmetry SU(5), and in that theory the proton is unstable:

$$p \to e^+ + \pi^0 \qquad [10]$$

The predicted lifetime is $10^{30\pm1}$ years, while a recent estimate of the lifetime for this decay mode is $> 5 \times 10^{32}$ years. It may be that GUTs do not exist in nature, but since the decay [10] violates conservation of the quantities $B$ and $L_e$, even entertaining the idea that the decay might take place begs the question, "are these conservation laws sacrosanct?"

Another recent development which leads to the same question is the subject of neutrino oscillations. A strong motivation for this is the solar neutrino problem; this is the problem that the number of electron neutrinos detected on Earth, originating in the Sun, is less than the number predicted, by a factor close to 3. The mismatch could be at least partly, and perhaps completely, explained if electron neutrinos "oscillated" into muon and/or tau neutrinos on their passage from the Sun to the Earth, since the reaction which detects the neutrinos on Earth is sensitive only to electron neutrinos, and not to the other species. But oscillation is only permitted if $L_e, L_\mu$, and $L_\tau$ are not separately conserved quantities. Oscillation can also only take place if the masses of the different neutrinos are different – the oscillation rate depends on $\Delta m^2$ – hence not all the neutrinos may be massless.

## Discrete Symmetries

Ever since parity violation was discovered in weak interactions (nuclear beta decay) by Wu in 1957, the whole subject of discrete symmetries has presented problems which are still not resolved. The symmetries in question are

$P$ (space inversion): $(x, y, z) \to (-x, -y, -z)$
$T$ (time reversal): $t \to -t$
$C$ (particle–antiparticle conjugation): particle $\leftrightarrow$ antiparticle

Are the laws of physics invariant under these operations? The Wu experiment revealed that weak interactions are not invariant under $P$, but what about other interactions and other operations? In this context, the $CPT$ theorem is highly important. According to this theorem (based on very general assumptions), all laws of nature must be invariant under the combined operation $CPT$, so that, for example, the fact that weak interactions are not invariant under $P$ means that they are not invariant under the product $CT$ either.

The violation of $P$ invariance in beta decay was soon related to the fact that the neutrino involved (the electron neutrino – or, to be precise, antineutrino) was massless. Spin-1/2 particles like the electron and neutrino obey the Dirac equation, which may be written out as a pair of coupled equations for left- and right-handed states. In the case $m = 0$, however, these equations decouple so it is possible to have a massless spin-1/2 particle which is either left-handed or right-handed. Any interaction involving this particle would automatically violate parity (which turns a left-handed state into a right-handed one). Experiments have verified that the neutrino is indeed left-handed. The SM incorporates this in the sense that the left-handed electron $e^-{}_L$ and the electron neutrino $\nu_e$ are assigned to a

weak isospin SU(2) doublet, while the right-handed electron $e^-_R$ transforms as a singlet. A similar pattern is repeated for the $\mu$ and $\tau$ particles and their neutrinos. The phenomenon of neutrino oscillations, on the other hand, does not allow all the neutrino states also to be purely left-handed (since they cannot be massless). This poses a potential problem for the SM.

For a few years after 1957 it was believed that beta decay violated $C$ as well as $P$, but conserved the product $CP$; and indeed that all weak interactions were $CP$ invariant. In 1964, however, it was found that there is a small element of $CP$ violation in $K^0$ decay. CP-violating effects are also expected in $B^0$ decays. The physical origin of $CP$ violation is still not understood, but its importance is that it implies $T$ violation, so that in (at least some) weak interactions, there is an "arrow of time" on the subnuclear scale. (Such an arrow of time is, of course, familiar in thermodynamics.) This is used in a cosmological context to explain baryon–antibaryon asymmetry in the Universe.

#### Baryon–Antibaryon Asymmetry

In the standard model of cosmology it is shown that applying the known laws of physics to the early Universe (the first few minutes) leads to the conclusion that at an age of 226 s nuclear fusion reactions took place resulting in a mixture of 74% protons and 26% $\alpha$ particles, so that, hundreds of thousands of years later, when galactic condensation took place, it would involve precisely this admixture of hydrogen and helium gases. Just this amount of helium has been found in the Sun, giving great confidence to the "big bang" model. Assuming that at extremely small times the baryon number of the Universe was zero, $B = 0$, and assuming also (a big assumption, but one nevertheless made by cosmologists) that the Universe is made of matter and not antimatter, we may then ask, why is this – where has the antimatter gone?

Surprisingly, this question was addressed as early as 1966 by Sakharov, who showed that, starting with an initial state with $B = 0$, it would be possible to reach a state with $B \neq 0$ as long as three conditions obtained: $B$ violating interactions, $CP$ and $C$ violating interactions, and lack of thermal equilibrium. GUTs and ordinary weak interactions already provide possibilities for the first two of these conditions. Breakdown of thermal equilibrium will be expected to occur as the Universe expands. When the particle density is high, reactions such as $p + \bar{p} \to \gamma + \gamma$ will ensure an equal population of baryons and antibaryons, even in the presence of $B$ violating interactions, but as the density increases and this reaction rate becomes less than the expansion rate, thermal equilibrium can no longer be maintained. Thus, GUTs offer an explanation of why there is no antimatter in the Universe. It might be thought that this sort of explanation is implausible, since the $B$-violating and $CP$-violating forces are so weak, but actually this is not a problem, since the ratio of baryon number to photon number in the Universe is of the order $N_B/N_\gamma \approx 10^{-9}$; so we may conjure up a scenario in which the $B$ and $CP$ violating forces give rise to a volume of space in which there are, say, $10^9$ antibaryons, $10^9 + 1$ baryons and approximately the same number of photons. Then, all the antibaryons become annihilated leaving one baryon and $10^9$ photons – as observed.

A recent development in the area of discrete symmetries has been the suggestion by Kostelecky and coworkers that there might exist spontaneous violation of $CPT$ and Lorentz symmetry.

### Topological Charges

Conserved quantities of a quite different type have received a lot of attention in recent decades. Their conservation is a consequence of nontrivial boundary conditions for the fields. A famous example is the sine-Gordon "kink." The sine-Gordon equation

$$\frac{\partial^2\phi}{\partial t^2} - \frac{\partial^2\phi}{\partial x^2} + \frac{1}{b^2}\sin(b\phi) = 0 \qquad [11]$$

describes a scalar field in one space and one time dimension. It is a nonlinear equation which possesses, among others, the interesting solution

$$f(\xi) = \frac{4}{b}\ \arctan\exp[\pm(\gamma/\sqrt{b})\xi]$$

where $\xi = x - vt$ and $\gamma = (1 - v^2)^{-1/2}$. This corresponds to a solitary wave which moves, preserving its shape and size – in distinction to usual waves, which spread out and dissipate. Waves of this type are called solitons, and solitons have in fact been observed moving along canals. In this case, they are solutions to the Korteveg de Vries equation. Equation [11] clearly possesses the constant solutions

$$\phi = \frac{2\pi n}{b}, \quad n = 0, \pm1, \pm2, \ldots$$

which, it may be shown, all have zero energy. We may then construct a solution of the above type, but with $n = 0$ as $x \to -\infty$ and $n = N$ as $x \to +\infty$. This so-called "kink" solution has finite energy and is not continuously deformable into a solution with $n = 0$ everywhere, since this would involve overcoming an

infinite energy barrier. The "kink number" may be characterized as a charge: defining the current

$$J^{\mu} = \frac{b}{2\pi}\varepsilon^{\mu\nu}\partial_{\nu}\phi$$

with $\varepsilon^{\mu\nu}$ the totally antisymmetric symbol, it is clear that this is identically conserved, $\partial_{\mu}J^{\mu} = 0$. This is a consequence of the definition of $\varepsilon^{\mu\nu}$; it is not a consequence of invariance of the sine-Gordon Lagrangian under a symmetry operation, so the current $J^{\mu}$ is not a Noether current. The associated conserved charge is

$$\begin{aligned} Q &= \int J^0\,\mathrm{d}x = \frac{b}{2\pi}\int \frac{\partial\phi}{\partial x}\,\mathrm{d}x \\ &= \frac{b}{2\pi}[\phi(\infty) - \phi(-\infty)] = N \end{aligned}$$

Models of the above type may be written down in a spacetime with more than two dimensions. In that case the above solution depends only on one coordinate, so represents an infinite planar "domain wall," on the two sides of which the field assumes different values. Such domain walls, as well as "cosmic strings," are considered as serious possibilities in cosmology.

Nonabelian gauge theories and the sigma model also provide a fertile ground for topological excitations – field configurations which for topological reasons do not decay. Gauge theories with spontaneous symmetry breaking have two-dimensional solutions corresponding to vortex lines and three-dimensional solutions corresponding to magnetic monopoles. In spacetime ($3+1$ dimensions), there is a solution to the gauge field equations, with no spontaneous symmetry breaking, corresponding to an "instanton," a finite-energy field configuration, localized in time as well as in space (hence the name). The gauge group here is SU(2), whose group space is $S^3$. Spacetime is "Euclideanized" into $R^4$, whose boundary is then $S^3$. Asymptotic field configurations may then be characterized by mappings of $S^3$ in field space into $S^3$ in parameter space, and since the third homotopy group of $S^3$ is nontrivial, $\pi_3(S^3) = Z$, these field configurations belong to different classes and are not deformable into each other. These define "degenerate vacua" of the gauge field equations. In quantum theory, tunneling between these vacua is allowed and 't Hooft has shown how this may give rise to deuteron decay $d \to e^+ + \bar{\nu}_{\mu}$. Other examples of topologically nontrivial configurations are so-called sphalerons, which may also contribute to baryon number violation in the early Universe, and skyrmions, constructs in the nonlinear sigma model which serve as a model for baryon number.

## Supersymmetry

Supersymmetry is a fermion–boson symmetry, postulating that multiplets of fundamental particles contain both fermions and bosons. Thus, for example, since electrons exist there should also be "selectrons" – "scalar" electrons, with spin 0. There should also be photinos, with spin 1/2, to take their place alongside photons, and so on. If supersymmetry were exact, these particles would have the same mass as their partners and would have all been found, but in fact none have yet been discovered, so presumably supersymmetry is a broken symmetry. The feature that makes supersymmetry attractive is that it holds some promise for solving divergence problems in quantum field theory, since the radiative corrections from fermion and boson loops are opposite in sign and may exactly cancel. Supersymmetric models can also help to solve the so-called hierarchy problem in quantum field theory.

If supersymmetry is made into a local symmetry, rather than simply a global one, extra fields must be introduced (as the photon field was introduced above), and it turns out that one of these is a spin-2 field, which may be identified with the *graviton*. Local supersymmetry thus becomes supergravity.

## General Relativity

Symmetries and conservation laws take on new aspects when general relativity is considered. Einstein's field equations relate the energy–momentum tensor of matter (and radiation) to the Ricci tensor of spacetime. The Ricci tensor has vanishing *covariant* divergence, which means that the energy–momentum tensor possesses the same property, but conservation of energy and momentum requires that it is the ordinary derivative, not the covariant one, of this tensor that should vanish. It might be expected that this problem could be alleviated by including the contribution of the gravitational field itself in energy–momentum tensor. This is quite reasonable, but then problems of interpretation arise, since at any one point in a general spacetime, a coordinate system might be found which is inertial (this is the force of the equivalence principle), corresponding to no gravitational field, and therefore no energy. The usual procedure is to introduce an energy–momentum "pseudotensor," and to conclude that energy in a gravitational field is not localizable.

The role of symmetries in general relativity is rather different from its role in particle physics, which is set in Minkowski spacetime. In a general spacetime there are no symmetries, but many examples of particular spacetimes with their own symmetries are now known. The symmetry operations involved are

*isometries*, with corresponding groups of motion (so that the isometry group of Minkowski space is the Poincaré group). These groups are an important subject of study in cosmology; for example, there is a classification of homogeneous cosmological models, labeled according to the Bianchi classification.

*See also:* Cotangent Bundle Reduction; Effective Field Theories; Electroweak Theory; General Relativity: Overview; Infinite-Dimensional Hamiltonian Systems; Noncommutative Geometry and the Standard Model; Quantum Field Theory: A Brief Introduction; Quasiperiodic Systems; Sine-Gordon Equation; Supergravity; Symmetries in Quantum Field Theory of Lower Spacetime dimensions; Symmetry and Symplectic Reduction; Symmetry Classes in Random Matrix Theory; Topological Defects and Their Homotopy Classification.

## Further Reading

Aitchison IJ and Hey AJ (1981) *Gauge Theories in Particle Physics*. Bristol: Adam Hilger.
Cheng T-P and Li L-F (1984) *Gauge Theory of Elementary Particle Physics*. Oxford: Clarendon Press.
Eguchi T, Gilkey PB, and Hanson AJ (1980) Gravitation, gauge theories and differential geometry. *Physics Reports* 66: 213.
Huang K (1998) *Quantum Field Theory: From Operators to Path Integrals*. New York: Wiley.
Kostelecky VA (ed.) (2004) *CPT and Lorentz Symmetry: Proceedings of the Third Meeting*. Singapore: World Scientific.
Landau LD and Lifshitz EM (1971) *The Classical Theory of Fields*. Oxford: Pergamon Press.
Manton N and Sutcliffe P (2004) *Topological Solitons*. Cambridge: Cambridge University Press.
Perkins DH (2000) *Introduction to High Energy Physics*, 4th edn. Cambridge: Cambridge University Press.
Review of Particle Properties (2002), *Physical Review D* 66: 01002, July 2002, Part 1.
Rubakov V (2002) *Classical Theory of Gauge Fields*. Princeton: Princeton University Press.
Ryder LH (1996) *Quantum Field Theory*, 2nd edn. Cambridge: Cambridge University Press.
Stephani H (2004) *Relativity: An Introduction to Special and General Relativity*, 3rd edn. Cambridge: Cambridge University Press.
Weinberg S (1983) *The First Three Minutes*. London: Fontana.
Wess J and Bagger (1983) *Supersymmetry and Supergravity*. Princeton: Princeton University Press.
Wigner E (1939) On unitary representations of the inhomogeneous Lorentz group. *Annals of Mathematics* 40: 149.

# Symmetries in Quantum Field Theory of Lower Spacetime Dimensions

**J Mund**, Universidade de São Paulo, São Paulo, Brazil
**K-H Rehren**, Universität Göttingen, Göttingen, Germany



## Symmetries in Quantum Field Theory

Symmetries have proved to be one of the most powerful concepts in quantum theory, and in quantum field theory in particular. From the beginnings of quantum mechanics, it is well known that the presence of a symmetry allows one to predict relations between different measurements, to classify spectra (energy or other), and to understand the Pauli exclusion principle, to name only a few applications. Much more remarkably, in modern relativistic quantum field theory, designed to describe the interactions of elementary particles, fundamental interactions have been found to be induced by the principle of local gauge invariance.

One distinguishes spacetime symmetries (Poincaré or conformal transformations), which change the position and orientation of the system in space and time, and internal symmetries, which preserve the localization, acting on certain internal degrees of freedom. The Coleman–Mandula (1967) theorem states that internal and spacetime symmetries cannot be mixed, in the sense that the generators of internal symmetries must be Lorentz scalars, hence the total group of symmetries factorizes into a direct product. Supersymmetries are an exception of this theorem because their generators do not form a Lie algebra, and they were in fact designed to circumvent the Coleman–Mandula theorem.

It is well known that the structure of symmetries of quantum systems in low-dimensional spacetime differs significantly from that in four-dimensional spacetime. ("Low" means in our context two or three, depending on the type of charge localization, c.f. below.) To name some examples:

- Two-dimensional quantum systems may have much higher symmetries than four-dimensional ones:
  - In two dimensions, there exist massive integrable models with infinitely many conservation laws and factorizable scattering matrices (*see* Integrability and Quantum Field Theory). These models exhibit solitonic superselection sectors, c.f. below.
  - The conformal group of two-dimensional spacetime is infinite dimensional, allowing for

the exact computation of correlation functions by the help of Ward identities (Belavin, Polyakov, and Zamolodchikov 1984). Only the finite-dimensional Möbius group, however, is also a symmetry of the vacuum state. Möbius covariance implies that the theory contains two subtheories of chiral fields defined on the light rays $t - x = \text{constant}$, resp. $t + x = \text{constant}$, and that these can be extended to fields defined on a circle, by adding a "point at infinity" to the light ray (Lüscher and Mack 1976). One arrives thus at one-dimensional chiral quantum field theories on a circle, which will play an important role in the discussion below.

- Continuous symmetries cannot be spontaneously broken in two dimensions. The latter is true not only for relativistic quantum field theory (Coleman 1973), but also in quantum statistical mechanics (Mermin and Wagner 1966) where it is responsible for the absence of ferromagnetism (*see* Symmetry Breaking in Field Theory). Spontaneous symmetry breakdown requires long-range order which is overcome by thermal fluctuations down to zero temperature, because these diverge logarithmically (in the thermodynamical limit) in two dimensions. This theorem thus illustrates how the spacetime dimension-dependent size of phase space has an effect on internal symmetries of quantum systems. A detailed mathematical analysis of the balance between phase space (thermal fluctuations) and long-range order (symmetry breakdown) has been given in a recent discussion of the Goldstone theorem (Buchholz, Doplicher, Longo and Roberts 1992).
- The Coleman–Mandula theorem, excluding a mixing between internal and spacetime symmetries (see above), is valid only in higher dimensions.

In more recent times, it has become apparent that low-dimensional quantum systems do not only admit more symmetries, but they may exhibit internal symmetries of an entirely new type, not describable by groups of transformations. In this article, we shall focus on the various ways in which the new symmetries can arise, and how they can be understood. In order to properly appreciate these issues, let us first recall some basic symmetry concepts in the conventional case.

In the traditional setting, symmetries arise in the form of groups of transformations of the quantum system which leave observable quantities (e.g., vacuum expectation values and correlation functions) invariant. The symmetries form a group of $*$-automorphisms of the algebra of fields:

$$\begin{aligned} \alpha_g(\phi_1\phi_2) &= \alpha_g(\phi_1)\alpha_g(\phi_2) \\ \left(\alpha_g(\phi)\right)^* &= \alpha_g(\phi^*) \\ \alpha_{g_1}\alpha_{g_2} &= \alpha_{g_1 g_2} \end{aligned} \qquad [1]$$

(typically given by linear transformations of field multiplets). In the strongest case, the automorphisms are implemented by unitary operators on the state space

$$U(g)\phi U(g)^* = \alpha_g(\phi) \qquad [2]$$

The implementers form a representation of the group of automorphisms,

$$U(g_1)U(g_2) = U(g_1 g_2) \qquad [3]$$

and there is an invariant vector state (a ground state, or the vacuum state in relativistic quantum field theory),

$$U(g)\Omega = \Omega \qquad [4]$$

However, depending on the dynamics of the quantum system, these relations cannot always be fully realized. One therefore considers several weaker or more general notions of symmetries relevant in four dimensions:

- *Spontaneously broken symmetries.* The transformations are given as automorphisms of an algebra, but which are not unitarily implemented in a given irreducible representation of the algebra. Invariant pure states do not exist.
- *Projective representations.* The symmetries are unitarily implemented, but the implementers fail to satisfy the group law [3]. They give rise to ray (projective) representations or representations of a covering group. In particular, an invariant state vector as in [4] cannot exist in an irreducible representation.
- *Infinitesimal symmetries.* Lie algebras of infinitesimal transformations, given as derivations of an algebra, which cannot be integrated to finite transformations. Derivations may or may not be implemented in a given representation of the algebra by commutators with self-adjoint generators.
- *Supersymmetry.* The infinitesimal transformations form a graded Lie algebra.
- Local gauge symmetries form an infinite-dimensional group which are, however, not realized as automorphisms of the quantum algebra. Quantization of classical gauge interactions usually proceeds by breaking the gauge invariance in some way and restoring it at a later stage.

## The Connection between Symmetry and Superselection Sectors

It is often convenient to describe a model in terms of localized fields which do not represent an observable (in the sense of quantum mechanics that an operator corresponds to some measurement prescription). For example, Fermi fields which violate the principle of causality because they anticommute with each other at spacelike distance rather than commute are not observables. Only fields which are quadratic in the Fermi fields (densities of charge, current, energy) are observables. This means that an internal symmetry is used in order to distinguish the observables as those operators which are invariant under the symmetry: in the example, the symmetry transformation multiplies each Fermi field by –1 (by the spin-statistics theorem, this transformation coincides with the univalence of the Lorentz group). We characterize this situation by writing

$$A(O) = F(O)^G \qquad [5]$$

where $A(O)$ and $F(O)$ stand for the algebras of observables and fields localized in some spacetime region $O$, respectively, $G$ is the internal symmetry group acting by automorphisms on each $F(O)$ without affecting the localization, and $F(O)^G \equiv \{a \in F(O), \alpha_g(a) = a$ for all $g \in G\}$ denotes the subalgebra of invariants. The internal symmetry group $G$ which distinguishes the observables according to [5] is usually called the "(global) gauge group."

If the gauge symmetry $G$ is unbroken in the vacuum state, then there is a well-known connection between symmetry and superselection rules (*see* Symmetries and Conservation Laws): namely, the observables act reducibly on the vacuum Hilbert space representation of $F$ because they commute with the unitary operators which implement the symmetry (or with their infinitesimal generators, usually called charges). As a consequence, the validity of the superposition principle is restricted because two eigenstates of different eigenvalues of the charges cannot exhibit interference. In other words, they belong to different superselection sectors. Wick, Wightman, and Wigner (1952) were the first to point out this relation. We therefore call this scenario the "WWW scenario" for brevity.

In the WWW scenario, the decomposition of the Hilbert space is determined by the central decomposition of the internal symmetry group (the eigenvalues of the Casimir operators). In this way, the superselection sectors are in one-to-one correspondence with the irreducible representations of the internal symmetry group.

Superselection sectors of two-dimensional models do not follow this scheme expected by the WWW scenario (see below). This was most strikingly demonstrated through the classification of the unitary highest-weight representations of the Virasoro algebra (Friedan, Qiu, and Shenker) which is nothing other than the classification of the superselection sectors of the observable algebra generated by the chiral stress–energy tensor, and through the determination of their fusion rules by Belavin, Polyakov, and Zamolodchikov (1984).

In two dimensions, one is therefore lacking a compelling *a priori* ansatz, like the WWW scenario, for describing the system in terms of auxiliary nonobservable charged fields. At this point, one may argue that from an operational point of view, a quantum field theory, and in particular its symmetries, should be understood entirely in terms of its observables. (This viewpoint is emphasized in the algebraic approach to QFT, *see* Algebraic Approach to Quantum Field Theory.) We shall therefore now ask the opposite question: suppose we are given an algebra $A$ of local observables (without knowledge of a field algebra and its gauge group). We define the superselection sectors intrinsically as (the unitary equivalence classes of) the positive-energy representations of $A$. Then the question is: do these sectors arise through a WWW scenario from some field algebra and a gauge symmetry, and if so, can the latter be reconstructed from the given observables alone?

The answer in four dimensions is positive, thanks to a deep result due to Doplicher and Roberts (1990). Let us sketch the line of reasoning leading to this result in some detail, because it shows how the connection between (global) gauge symmetry on the one hand and spacetime geometry on the other hand emerges through the principle of causality (locality) of relativistic quantum field theory, and because it makes apparent what is different in low-dimensional spacetime.

The analysis is based on the general structure theory of superselection sectors due to Doplicher, Haag, and Roberts (DHR, 1971). The latter starts with a selection criterion invoking the concept of a localized charge: a superselection sector which by measurements within the causal complement of some spacetime region $O$ cannot be distinguished from the vacuum sector. The heuristic idea is, of course, that the sector is obtained from the vacuum sector by placing some charge in the region $O$ (e.g., by the application of a localized charged field operator to the vacuum vector).

It has been shown (Buchholz and Fredenhagen 1982) that positive-energy representations of

massive theories always satisfy this selection criterion with a localization region $O$ of the form of a narrow cone extending in spacelike direction. (In massless theories with long-range interactions, such as QED, the situation is more complicated because the charge creates an electric field whose flux at infinity does not vanish (Gauss' law) and is not Lorentz invariant.) DHR assume that the localization region is even compact, and can be chosen arbitrarily within the unitary equivalence class of the representation.

Exploiting a strong version of locality (Haag duality) for the vacuum representation of the observables, DHR proceed to define an associative composition (or fusion) law for positive-energy representations. This law is commutative only up to unitary equivalence. The crucial point is that the unitary intertwiner establishing this equivalence (the statistics operator) can be chosen in a unique way provided any pair of spacelike disconnected localization regions can be continuously deformed into any other such pair.

This point marks the separation between high and low dimensions. In two dimensions, in each pair of spacelike disconnected regions, one region is to the left of the other, thus distinguishing the pair $(O_1, O_2)$ from $(O_2, O_1)$. Consequently, they cannot be deformed into each other, and there arise two statistics operators. The same holds in three dimensions when the localization regions are spacelike cones, and $O_1, O_2$ are taken within (the causal complement of) some larger spacelike cone. If the spacetime dimension is at least 4, or if in three dimensions the localization regions are compact, then the statistics operator is unique and, as a consequence, coincides with its inverse.

The (non-)uniqueness of the statistics operator has far-reaching consequences concerning our original question about the underlying gauge symmetry. Namely, the DHR analysis proceeds to show that the set of positive-energy representations equipped with the composition law, and the linear spaces of inertwiners between different representations, together form the mathematical structure of a $C^*$ tensor category. The statistics operators which are distinguished intertwiners give additional structure to this category: this structure is called a (permutation) symmetry if the statistics operators coincide with their inverse, and it is called a braiding otherwise. (It gives rise to a representation of the permutation group or the braid group, respectively.) In other words, the spacetime topology, through the intervention of the uniqueness of the statistics operator, causes the tensor category to be symmetric in high dimensions, and braided in low dimensions.

At a more elementary level, one may think of statistics operators as reflecting commutation relations between the searched-for charged fields. Making an ansatz for the commutation relations at spacelike separation, essentially the same topological argument as before implies, together with Poincaré invariance, that the coefficients appearing in this relation should form a representation of the permutation group, or of the braid group, respectively. The DHR approach, however, is entirely intrinsic, avoiding any *a priori* assumption of charged fields.

The duality theorem due to Doplicher and Roberts (1990) now states that every symmetric $C^*$ tensor category (with some further qualifications valid in the DHR setting) is isomorphic to the category of unitary representations of a compact group, in which the composition law is the tensor product and the (permutation) symmetry is the natural one. Moreover, the category uniquely determines the group, and by a crossed product construction (an action of the category on the algebra $A$) one reconstructs a field algebra $F$ such that [5] holds. If fermionic sectors are present, then there is some arbitrariness in the commutation relations among the corresponding fermionic fields, which can be exploited to produce the normal commutation relations (fermionic fields anticommute among each other, and bosonic fields commute with any field at spacelike separation). This fixes the field algebra $F$ up to unitary equivalence. The conclusion is that the WWW scenario is the most general in four dimensions (apart from the reservations due to long-range forces, see above).

## Generalized Symmetries in Low Dimensions

In view of the success of this program in four dimensions and the advantage of the WWW scenario for model building, the obvious challenge is to search for an analogous understanding of superselection sectors (charges) in low dimensions in terms of an algebra of charged fields and a gauge symmetry distinguishing the observables. This gauge symmetry cannot, in general, be a group for several reasons:

- As stated before, the tensor category of superselection sectors possesses only a braiding, rather than a (permutation) symmetry, hence the duality theorem fails.
- One can associate a (statistical) dimension $d_\pi$ to each superselection sector $[\pi]$ which is multiplicative under the composition law (fusion), and additive under direct sums. In a symmetric

category, the dimensions are necessarily positive integers. Indeed, in the WWW scenario, they coincide with the naive dimension of the associated representation of the gauge group. But in the low-dimensional models, the dimensions turn out to be nonintegers in general.

- Moore and Seiberg (1988) have axiomatized the superselection structure of chiral and two-dimensional conformal field theories in terms of a system of recoupling and braiding coefficients controlling the fusion of sectors and its noncommutativity. (In fact, this system is basically equivalent to the DHR category.) For models such as SU(2) current algebras at level $k$, these coefficients turn out to coincide with the recoupling and braiding coefficients one can associate with a quantum group deformation (Drinfel'd 1986) of SU(2) with deformation parameter $q = -\exp \mathrm{i}\pi/k$. Representations of quantum groups (quasitriangular Hopf algebras, *see* Hopf Algebras and $q$-Deformation Quantum Groups) have a tensor product defined in terms of a noncocommutative coproduct. Moreover, they possess a quantum dimension which is a $q$-deformation of an integer. The quantum dimensions precisely match the statistical dimensions of the superselection sectors. All this strongly suggests that quantum groups appear as generalized symmetries in two dimensions, at least in a large class of models.

A natural testing ground for the search for appropriate generalized symmetry concepts in low dimensions is the abundance of models in chiral and two-dimensional conformal QFT (*see* Two-Dimensional Models). As mentioned before, conformal symmetry in two dimensions has far-reaching consequences, especially the existence of chiral quantum fields which are defined on a one-dimensional light ray. As a null direction in the two-dimensional spacetime, this ray unites both the spacelike property of carrying a causal structure, and the timelike property that the generator of translations has positive spectrum (energy). These two features together with Möbius covariance are so powerful that they allow for the exact construction of large classes of models. The most elementary ones (minimal models) are completely described by the chiral stress–energy density field, that is, the local generator of the conformal symmetry. Other models also contain currents which are the local generators of internal symmetries. These models exhibit many nontrivial superselection structures, which illustrate the wide range of possible deviations from higher-dimensional QFT, and at the same time exhibit possible approaches to appropriate symmetry concepts in low dimensions.

Attempts to classify the possible algebraic structures of generalized internal symmetries in a model-independent setting start from the idea that the representation category of the internal symmetries of a given model should be equivalent to the tensor category of its superselection sectors. Several algebraic structures have been proposed as candidates, complying with this idea. They all assume specific modifications or deformations of eqns [1]–[5] above, highly constrained by self-consistency. Among these proposals are:

- quantum groups (see e.g., Fröhlich and Kerler 1993),
- weak quasiquantum groups (Mack and Schomerus 1992) and rational Hopf algebras (Fuchs *et al.* 1994),
- weak C* Hopf algebras (Rehren 1997, Böhm and Szlachányi 1996) or quantum groupoids (Nikshych and Vainerman 1998), and
- braided groups (Majid 1991).

In several cases, the respective "symmetry algebra" can be reconstructed from the tensor category of superselection sectors, and a field algebra with linear transformation behavior can be constructed which contains the observables as invariant elements as in [5]. However, the situation is unsatisfactory for various reasons. First, the class of QFT models for which these constructions have been performed is quite restricted (most constructions work only for rational models, i.e., models with a finite set of charges); second, the reconstructed symmetry algebra is not unique and finally, the constructed field algebras have features which diverge significantly from the WWW scenario. For example, it is not always warranted that the quantum symmetries are consistent with the *-structure, indispensable for Hilbert space positivity (a necessary prerequisite for the probability interpretation of quantum theory). Moreover, typically there are global gauge transformations which are implemented by localized field operators, thus exhibiting a mixing of local and global concepts. It also happens that this holds for elements in the center of the symmetry algebra, which implies that the field algebra is not local relative to its gauge invariant elements, that is, the charged fields do not commute with the gauge-invariant elements at spacelike separation. In other constructions, the field algebra is not associative, or there are no finite field multiplets.

Historically, the first candidate for a "symmetry algebra" compatible with braid group statistics has

been the structure of a quantum group, as mentioned above. However, in physically interesting models, the quantum group is not semisimple and thus has too many (namely, indecomposable) representations. Solutions to this problem have been:

1. A BRS approach in an indefinite-metric framework (Hadjiivanov *et al.* 1991),
2. "Truncation," that is, discarding the "unphysical" representations. Fröhlich and Kerler (1993) have done this consistently in a categorical framework. In fact, they have given a complete classification of the possible braided tensor categories generated by a single irreducible object with statistical dimension $d$ satisfying $1 < d < 2$, in terms of categories constructed from the "truncated" representations of $U_q(\mathrm{sl}_2)$. Truncation can also be performed by dividing the quantum group itself through the ideal which is annihilated by all "physical" representations, leading to a weak quasiquantum group (Mack and Schomerus 1992).
3. Relaxing the axioms, thus admitting the more general structures mentioned above.

All the above approaches assume a given generalized symmetry concept and show to what extent field algebras complying with it can be constructed. They thus concern nonobservable objects, and it is no contradiction if different symmetry concepts can be associated with the same observable data.

A more radical concept of global gauge symmetry, applicable to the low-dimensional case, has been developed by Longo and Rehren (1995). Its point of departure is the notion of a conditional expectation, which has the same abstract properties as a group average. In the WWW scenario, the Haar measure of the compact gauge group defines an average

$$\mu : F \ni \phi \mapsto \int \mathrm{d}\mu(g)\, \alpha_g(\phi) \in A \qquad [6]$$

which is a positive linear map respecting the localization, and the observables are invariant, $\mu(a) = a$. In fact, the observables are exactly the image of this map, that is, [5] is equivalently formulated, but without reference to the group transformations, as

$$A(O) = \mu(F(O)) \qquad [7]$$

Turning to the observables $A$ of a quantum field theory in low dimensions, one looks for a quantum field theory $F$, containing $A$ and equipped with a conditional expectation $\mu$ such that [7] holds, and which preserves the vaccum state. $F$ may not satisfy local commutativity, but it should be local relative to the observables in the sense mentioned before. In rational chiral CFT, such extensions can be classified (and indeed constructed) in terms of the superselection category of $A$, giving direct access to the decomposition of the vacuum Hilbert space of $F$ into superselection sectors of $A$. The advantage here is that no problems with Hilbert space structure can arise (because the approach is entirely in terms of operator algebras); a drawback is that in general $F$ is not unique, and nonvacuum representations of $F$ also have to be considered in order to generate all sectors of $A$.

The method can be used to classify and construct both nonlocal chiral extensions as candidates for sector-generating field algebras for a theory $A$ of chiral observables, and local two-dimensional quantum field theories containing two given chiral subtheories, that is, observable algebras of two-dimensional models (Kawahigashi and Longo 2004). The chiral sector structure of the latter models is described by a "modular invariant." In many cases, this means that their thermal partition functions are invariant under the group $\mathrm{PSL}(2, \mathbb{Z})$ of modular transformations of the temperature (see below).

At this point, another link between spacetime and internal symmetries may be noted. The modular theory of von Neumann algebras (*see* Tomita–Takesaki Modular Theory) associates a one-parameter group of automorphisms (called the "modular group") with a state and an algebra "in standard position." In quantum field theory, for the vacuum state and an algebra of observables localized in certain wedge regions of Minkowski spacetime, this group can be identified with a boost subgroup of the Lorentz group (Bisognano and Wichmann 1975). Similarly, in chiral CFT on the circle, the modular group associated with the observables in an interval and the vacuum coincides with a subgroup of the Möbius group. For nonlocal theories, there may be an obstruction, however. On the other hand, if a subalgebra is stable under the modular group of some algebra, then there is a conditional expectation from the larger algebra onto the smaller algebra. Combining these general theorems, the Möbius covariance of the inclusions $A(O) \subset F(O)$ implies the existence of a conditional expectation, that is, the above generalization of the average over the internal symmetry. Moreover, assuming a generalized notion of compactness ("finite index") for the generalized internal symmetry, the Bisognano–Wichmann property holds also for nonlocal theories (Longo and Rehren 2004).

Of course, there is also a WWW scenario in chiral theories, that is, one may restrict a local theory to its invariants under some group of internal gauge

symmetries ("orbifold models"). It then happens that the invariants not only have the expected superselection sectors in correspondence with the representations of the gauge group, but in addition "twisted" sectors appear which, together with the former, constitute a "quantum double" structure. The twisted sectors arise by restriction of solitonic sectors of the original theory, which are in one-to-one correspondence with the elements of the gauge group (Müger 2005). Solitonic sectors are localizable with respect to two different vacua, and do not admit an unrestricted composition law.

## Special Issues

A particularly simple situation is the case of anyons, that is, when all sectors have statistical dimension 1. Then the sectors form an abelian group $\hat{G}$ under fusion, and one can construct a WWW scenario with global gauge group $G$ the dual of $\hat{G}$. The ensuing quantum fields satisfy generalized commutation relations at spacelike separation, given by an abelian representation of the braid group, where the coefficients can be arbitrary complex phases (responsible for the name "anyons"). However, it is known that there can arise an obstruction, which enforces the "local" global gauge transformations (mentioned before) to be present. In this case, the gauge symmetry can also be described by a quasiquantum group. It is noteworthy that free anyon fields have been constructed in two-dimensional spacetime, while in three dimensions there can be no (cone-) localized massive anyon fields which are free in the sense that they generate only single-particle states from the vacuum (Mund 1998).

The charge structure of massive quantum field theories in two dimensions is very different both from that encountered in conformal quantum field theories, and from the charge structure in high dimensions. It has been observed long ago that, in contrast to four dimensions, the strong locality property (Haag duality) which is necessary to set up the DHR analysis of superselection sectors, fails for the algebra of invariants under an internal gauge group in two dimensions. This algebraic feature can be traced back to the fact that the causal complement of a point is disconnected in two dimensions, or, in physical terms, that "a charge cannot be transported around a detector" without passing through its region of causal dependence. Müger (1998) has shown that any algebra of observables which satisfies Haag duality, cannot possess any nontrivial DHR superselection sectors at all, and that the only sectors which can exist are solitonic sectors. This general result nicely complies with the experience with integrable models, as mentioned before.

There are also some results giving interesting insight, which can be obtained intrinsically in terms of the observables. One of them concerns "central" observables (generalized Casimir operators).

Casimir operators in the WWW scenario are functions of the generators of the internal symmetry which usually are integrals over densities belonging to the field algebra $F$ (Noether's theorem). Since they also commute with the generators, they can be approximated by local observables, and are therefore defined in each representation of the latter. By Schur's lemma, they are multiples of the identity in each irreducible sector. Since the eigenvalues of Casimir operators distinguish the representations of the gauge group, they also distinguish the sectors.

In chiral CFT extended to the circle (see above), one can find global "charge measuring operators" $C_i$, one for each sector $\pi_i$, in the center of the observable algebra (Fredenhagen *et al.* 1992) which have similar properties. They arise as a consequence of an algebraic obstruction to define the charged sectors on the circle, related to a nontrivial effect if a charge is "transported once around the circle," and form an operator representation of the fusion rules within the global algebra of observables. Under rather natural conditions clarified by Kawahigashi, Longo, and Müger (2001), the matrix of eigenvalues $\pi_j(C_i)$ is nondegenerate, that is, the generalized Casimir operators completely distinguish the superselection sectors. In this case, the superselection category is a modular category (*see* Braided and Modular Tensor Categories): the matrix with entries $d_{\pi_j}\pi_j(C_i)$ and the diagonal matrix with entries $\pi_j(U)$ (where $U$ is the Möbius rotation by $2\pi$) are multiples of the generators $S$ and $T$ of the "modular group" $PSL(2, \mathbb{Z})$, in a matrix representation labeled by the superselection sectors of the chiral observables. The physical significance of this matrix representation is that it relates thermal expectation values for different values of the temperature (Cardy 1986, Kac and Peterson 1984, Verlinde 1988)

These examples, together with the failure of the Coleman–Mandula theorem, may illustrate the intricate relations among spacetime geometry, covariance, and internal symmetry (charge structure) in low dimensions. In relativistic quantum field theory, the link is provided by the principle of locality, which "turns geometry into algebra."

*See also:* Algebraic Approach to Quantum Field Theory; Axiomatic Quantum Field Theory; Braided and Modular Tensor Categories; Hopf Algebras and $q$-Deformation

Quantum Groups; Integrability and Quantum Field Theory; Quantum Field Theory: A Brief Introduction; Quantum Fields with Topological Defects; Symmetries and Conservation Laws; Symmetries in Quantum Field Theory: Algebraic Aspects; Symmetry Breaking in Field Theory; Tomita–Takesaki Modular Theory; Two-Dimensional Conformal Field Theory and Vertex Operator Algebras; Two-Dimensional Models.

## Further Reading


Böhm G and Szlachányi K (1996) A coassociative C*-quantum group with non-integral dimensions. *Letters in Mathematical Physics* 35: 437–448.

Coleman S and Mandula J (1967) All possible symmetries of the *S*-matrix. *Physical Review* 159: 1251–1256.

Doplicher S and Roberts JE (1990) Why there is a field algebra with a compact gauge group describing the superselection structure in particle physics. *Communications in Mathematical Physics* 131: 51–107.

Fredenhagen K, Rehren K-H, and Schroer B (1992) Superselection sectors with braid group statistics and exchange algebras II: geometric aspects and conformal covariance. *Reviews in Mathematical Physics* SI1: 113–157.

Fröhlich J and Kerler T (1993) *Quantum Groups, Quantum Categories, and Quantum Field Theory*. Lecture Notes in Mathematics, vol. 1542. Berlin: Springer.

Fuchs J, Ganchev A, and Vecsernyés P (1994) On the quantum symmetry of rational field theories. *Theoretical Mathematical Physics* 98: 266–276.

Hadjiivanov LK, Paunov RR, and Todorov IT (1991) Quantum group extended chiral *p*-models. *Nuclear Physics B* 356: 387–438.

Longo R and Rehren K-H (1995) Nets of subfactors. *Rev. Mathematical Physics* 7: 567–597.

Mack G and Schomerus V (1992) Quasi Hopf quantum symmetry in quantum theory. *Nuclear Physics B* 370: 185–230.

Majid S (1991) Braided groups and algebraic quantum field theories. *Letters in Mathematical Physics* 22: 167–175.

Moore G and Seiberg N (1989) Classical and quantum conformal field theory. *Communications in Mathematical Physics* 123: 177–254.

Müger M (1998) Superselection structure of massive quantum field theories in $1+1$ dimensions. *Reviews in Mathematical Physics* 10: 1147–1170.

Mund J (1998) No-go theorem for “free” relativistic anyons in $d=2+1$. *Letters in Mathematical Physics* 43: 319–328.

Rehren K-H (1997) Weak C* Hopf symmetry. In: Doebner H-D and Dobrev VK (eds.) *Group Theoretical Methods in Physics*, pp. 62–69. (q-alg/9611007). Sofia: Heron Press.

Schomerus V (1995) Construction of field algebras with quantum symmetries from local observables. *Communications in Mathematical Physics* 169: 193–236.

Wick GC, Wightman AS, and Wigner EP (1952) The intrinsic parity of elementary particles. *Physical Review* 88: 101–105.


# Symmetries in Quantum Field Theory: Algebraic Aspects

**J E Roberts**, Università di Roma “Tor Vergata,” Rome, Italy



## Introduction

This article treats the most important results and concepts relating to symmetry and conservation laws in quantum field theory. It includes such results as Wigner’s theorem, Goldstone’s theorem, the Bisognano–Wichmann theorem, the quantum Noether theorem, and the theorem on the existence of gauge groups and a field net. It is written within the framework of algebraic quantum field theory, this being the simplest setting capable of expressing all these concepts and results.

Symmetries come in many guises. They are to a physical system what automorphisms are to a mathematical theory. In fact, when a physical system is described in mathematical terms, its symmetries correspond to the automorphisms of the mathematical structure and in particular form a group, its symmetry group. The reader should bear in mind this simple picture throughout its diverse variations. Readers unfamiliar with the mathematical terminology should consult the appendix.

## Elementary Quantum Mechanics

Before turning to quantum field theory, let us comment on symmetries in elementary quantum mechanics. These systems have the density matrices, that is, positive operators of trace 1, on an infinite-dimensional separable Hilbert space as states, the self-adjoint operators as observables. The expectation value of the bounded observable $A$ in the state determined by $\rho$ is given by $\mathrm{tr}\,\rho \mathrm{A}$. Having specified the mathematical structure, the notion of symmetry follows. With a suggestive notation, it is a pair of mappings $A \mapsto \alpha A, \rho \mapsto \rho\alpha^{-1}$ such that

$$\mathrm{tr}\,\rho\alpha^{-1}\alpha \mathrm{A} = \mathrm{tr}\,\rho \mathrm{A}$$

for all observables $A$ and states $\rho$.

If we take $\rho$ and $A$ to be the projections onto $\mathbb{C}\phi$ and $\mathbb{C}\psi$ for unit vectors $\phi$ and $\psi$, then the above condition corresponds to the conservation of

transition probabilities $|(\phi,\psi)|^2$. This formed the starting point for Wigner's analysis, who concluded:

**Theorem** *Every symmetry is of the form* $A \mapsto UAU^{-1}$ *and* $\rho \mapsto U\rho U^{-1}$, *where* $U$ *is a unitary or antiunitary operator.*

As could have been foreseen from the outset, this simple result in no way distinguishes one elementary quantum-mechanical system from another. A more useful notion of symmetry results if the Hamiltonian is reckoned as part of the information describing the system and, therefore, has to be left invariant by a symmetry. The operator $U$ above must therefore satisfy the condition $UHU^{-1} = H$ and it commutes with the Hamiltonian. As the Hamiltonian is the generator of time translations, $U$ is a constant of motion. This is the genesis of the relation between symmetries and conservation laws.

## Quantum Field Theories

The simplest types of quantum field theories can be described by von Neumann algebras $\mathfrak{A}(\mathcal{O})$ depending on double cones $\mathcal{O}$ and subject to

$$\mathcal{O}_1 \subset \mathcal{O}_2 \Rightarrow \mathfrak{A}(\mathcal{O}_1) \subset \mathfrak{A}(\mathcal{O}_2)$$

a structure referred to as the net of observables.

An alternative approach would be to use the Wightman formalism. This would need a discussion of pointlike fields and the domains of definition of unbounded operators, thus complicating a general exposition of symmetry.

Comparing this description of a quantum field theory with that of an elementary quantum-mechanical system, the net clearly substitutes observables but nothing has yet been said about states. Since the set of double cones is directed under inclusion, the union of the $\mathfrak{A}(\mathcal{O})$ is a $*$-algebra $\mathfrak{A}$ and a state of our system is a state on this algebra.

Most states are of no physical relevance. A characterization of the states of physical relevance, even say to elementary particle physics, is not known although some progress has been made.

The net structure is the hallmark of a field theory and allows us to distinguish two important classes of symmetries. An internal symmetry $\alpha$ satisfies the condition

$$\alpha(\mathfrak{A}(\mathcal{O})) = \mathfrak{A}(\mathcal{O})$$

for all double cones $\mathcal{O}$. By contrast, a spacetime symmetry is an automorphism $\alpha_L$ implementing a Poincaré transformation $L$ and hence satisfying the condition

$$\alpha_L(\mathfrak{A}(\mathcal{O})) = \mathfrak{A}(L\mathcal{O})$$

for every double cone $\mathcal{O}$. It is usually the case that internal symmetries commute with spacetime symmetries.

The state of prime relevance to elementary particle physics is the vacuum state $\omega_0$. The corresponding Gelfand–Naimark–Segal (GNS) representation $\pi_0$ is called the vacuum representation. Now the vacuum state of a quantum field theory is typically unique and as such invariant under a symmetry of the system $\omega_0\alpha^{-1} = \omega_0$.

## Spacetime Symmetries

Since the vacuum state is invariant, we have a unitary representation of the Poincaré group implementing the spacetime symmetries in the vacuum representation. To illustrate the role of representations up to a factor, we take instead the GNS representation of a pure state corresponding to a particle of half-integral spin. Here we need a unitary representation of the covering group of the Poincaré group, inhomogeneous $SL(2,\mathbb{C})$ to implement the symmetries. The situation for the subgroup of rotations is the same.

The most important property of these representations is positivity of the energy. More precisely, in a representation of relevance to elementary particle physics such as the vacuum representation, the generator $P^0$ of time translations is a positive operator $P^0 \geq 0$. Expressed in a frame-independent way, the spectrum of spacetime translations is contained in the closed forward light cone. It is one of the basic principles to be exploited in applying quantum field theory to elementary particle physics. Notice that the principle is no longer valid for an equilibrium state.

A similar situation arises in conformal field theory. Here the role of double cones in Minkowski space is played by intervals on the circle and that of the Poincaré group by the Möbius group on the circle $PSL(2,\mathbb{R})$. Again, the Möbius group cannot always be unitarily implemented and conformal invariance is defined via a continuous unitary representation of its covering group. Most importantly, there is an analog of positivity of the energy. The generator of rotations of the circle is a positive operator.

A remarkable aspect of spacetime symmetries was discovered by Bisognano and Wichmann in an application of modular theory in the field-theoretical context looking not at double cones but at wedges. A wedge $\mathcal{W}$ is a Poincaré transform of the standard wedge $x^1 > |x^0|$. They found that the modular automorphisms of $\mathfrak{A}(\mathcal{W})$ and the vacuum vector $\Omega_0$

have a geometric significance. For the standard wedge, they got the following result.

**Theorem** *If the net is derived from Wightman fields, the modular operator is* $e^{-2\pi K}$, *where K is the generator of boosts in the* 1*-direction and the modular conjugation is* $ZR\Theta$, *where* $\Theta$ *is the* TCP-*operator, R is the rotation through* $\pi$ *about the* 1*-axis, and Z is the unitary operator equal to* 1 *on the Bose subspace and* $-\mathrm{i}$ *on the Fermi subspace.*

The modular data for $\mathfrak{A}(\mathcal{O})$ and $\Omega_0$ also admit a geometric interpretation for the free massless scalar field.

These facts enhance our understanding of spacetime symmetries. The ideas have meanwhile been applied to curved spacetime to select a state with vacuum-like properties using the principle of the geometric action of the modular conjugation.

## Gauge Symmetry

Gauge symmetries do not fit into our scheme in that they act trivially on the observable algebra $\mathfrak{A}$. To exhibit a gauge symmetry we need a larger net $\mathfrak{F}$ called the field net. The gauge group will be the group of automorphisms of $\mathfrak{F}$ leaving the subnet $\mathfrak{A}$ pointwise fixed and $\mathfrak{A}$ the subnet of $\mathfrak{F}$ of fixed points under $G$. This has the merit of indicating the mathematical framework for gauge symmetry but otherwise begs important questions. *A priori* one does not know what properties $\mathfrak{F}$ should have nor how it should be constructed.

The right approach is to understand what intrinsic structure of $\mathfrak{A}$ governs the existence of a nontrivial gauge group. This brings us back to the states or representations relevant to elementary particle physics. A condition for selecting some of these relevant representations is that asymptotically they be like the vacuum in spacelike directions. More precisely, $\pi$ must be unitarily equivalent to the vacuum representation $\pi_0$ on the spacelike complement of every double cone.

The resulting theory of superselection sectors hinges on the property of Haag duality that, for each double cone $\mathcal{O}$,

$$\mathfrak{A}(\mathcal{O}) = \mathfrak{A}(\mathcal{O}')'$$

where $\mathcal{O}'$ denotes the spacelike complement of $\mathcal{O}$. It implies that every representation satisfying the selection criterion is unitarily equivalent to one of the form $\pi_0\rho$, where $\rho$ is an endomorphism of $\mathfrak{A}$ localized in some fixed but arbitrary double cone, that is, $\rho(A) = A$ if $A \in \mathfrak{A}(\mathcal{O}')$. The endomorphisms thus obtained are closed under composition and hence the objects of a full tensor subcategory $\mathcal{T}$ of the category of all endomorphisms and their intertwiners. There is a dimension function $d$ defined on the objects of $\mathcal{T}$, $d(\rho) = 1, 2, \ldots, \infty$. If $\mathcal{T}_f$ denotes the full subcategory whose objects have finite dimension, then the following result holds.

**Theorem** $\mathcal{T}_f$ *is equivalent to the tensor category of finite-dimensional continuous unitary representations of a canonical compact group G. There is a canonical field net* $\mathfrak{F}$ *with Bose–Fermi commutation relations extending* $\mathfrak{A}$ *such that G is the group of automorphisms of* $\mathfrak{F}$ *leaving* $\mathfrak{A}$ *pointwise fixed.*

The first step in the proof is to define and analyze the statistics of the representations in question. The statistics of an irreducible representation $\rho$ can be classified as being para-Bose or para-Fermi of order $d(\rho)$. The second step is to show that each $\rho$ of finite dimension has a well-defined conjugate up to equivalence. The third and most difficult step is showing that $\mathcal{T}_f$ can be embedded in the tensor category of Hilbert spaces.

## The Local Implementation of Symmetries

Gauge symmetry has its associated conservation laws in that the different sectors of the last section are labeled by conserved quantities such as baryon number, lepton number, or electric charge, generically called charges. The theory is built round the idea of creating charge and elements of the field net carry charges. But there should be a dual approach based on measuring charges. One would like to prove the existence of local conserved currents corresponding to these charges. This has not proved possible but there is a good substitute, described below, which can be regarded as a weak version of a quantum Noether theorem.

If $\mathcal{O}_1 \subset \mathcal{O}_2$ is a strict inclusion of double cones, then the theory is said to satisfy the split property if there is a type I factor $\mathcal{M}$ such that

$$\mathfrak{A}(\mathcal{O}_1) \subset \mathcal{M} \subset \mathfrak{A}(\mathcal{O}_2)$$

where a type I factor is a von Neumann algebra isomorphic to some $\mathcal{B}(\mathcal{H})$. In this case $\mathcal{M}$ can be chosen in a canonical fashion and there is an isomorphism $\psi$ called the universal localizing map of $\mathcal{B}(\mathcal{H})$ onto $\mathcal{M}$, where $\mathcal{H}$ is the underlying Hilbert space. We have $\psi(A) = A$ for $A \in \mathfrak{A}(\mathcal{O}_1)$.

**Theorem** *If U is an implementing representation of the internal symmetry group G,* $\psi(U)$ *will be a representation of G in* $\mathcal{M}$ *that continues to implement the symmetry on* $\mathfrak{A}(\mathcal{O}_1)$. *If G is a Lie group*

*then the infinitesimal generators in the representation are an analog of locally integrated current densities.*

## Spontaneously Broken Symmetry

The standard physical example of a spontaneously broken symmetry is magnetization. Despite the overall rotational symmetry, a magnet picks out a preferred direction as its direction of magnetization. The chosen state breaks the symmetry.

The phenomenon of spontaneously broken symmetry involves an interplay of symmetries and certain classes of states, vacuum states, ground states, or equilibrium states. If such an $\omega$ is induced by a vector cyclic and separating for a local algebra $\mathfrak{A}(\mathcal{O})$, then, as explained in the appendix, given $\mathcal{O}$, modular theory yields a canonical unitary representation $V$ of the internal symmetry group $G$:

$$gA = V_g A V_g^*, \quad A \in \mathfrak{A}(\mathcal{O})$$

The results concern the breaking of a one-parameter group $\lambda \mapsto \alpha_\lambda$ of symmetries. More precisely, one asks whether $\omega\delta = 0$ or not, where $\delta$ is the infinitesimal generator of $\lambda \mapsto \alpha_\lambda$,

$$\delta(F) = \lim_{\lambda \to 0} \lambda^{-1}(\alpha_\lambda(F) - F)$$

where norm convergence is understood and holds on a dense domain. $\delta$, the derivation, is an infinitesimal symmetry. Goldstone first showed that the spontaneous breaking of such symmetries requires the presence of massless bosons. The following result is taken from a more modern treatment. $\mathcal{O}_R$ here denotes the double cone whose base is the ball in $t = 0$ of radius $R$ centered on the origin and $\mathcal{D}$ the domain of $\delta$.

**Theorem** *Let $\delta$ be a derivation on a field net $\mathfrak{F}$ in $s > 1$ spatial dimensions such that for $F \in \mathfrak{F}(\mathcal{O}_R) \cap \mathcal{D}$*

$$|\omega_0 \delta F| \leq c_{R,\varepsilon}(\|F\Omega\| + \|F^*\Omega\|) + \varepsilon\|\delta F\|$$

(i) *If $\liminf_{R\to\infty} c_{R,\varepsilon} R^{-(s-1)/2} = 0$, then $\omega_0\delta = 0$.*
(ii) *If $\liminf_{R\to\infty} c_{R,\varepsilon} R^{-(s-1)/2} < \infty$, then $\omega_0\delta \neq 0$ is only possible if the spectrum of the translations coincides with the forward light cone $V_+$ and the boundary $\partial V_+/\{0\}$ has non-trivial spectral measure (i.e., there are massless particles in the theory).*
(iii) *If $c_{R,\varepsilon}$ is polynomially bounded in $R$, then $\omega_0\delta \neq 0$ is only possible if the spectrum of translations coincides with $V_+$ but there are not necessarily any massless particles.*

## Symmetries of the *S*-matrix

Scattering theory not only allows one to construct the multiparticle scattering states but also shows that internal symmetries and spacetime symmetries continue to act on these states and are therefore symmetries of the *S*-matrix. We can, however, ask what are all the symmetries of the *S*-matrix. An answer was provided by Coleman and Mandula, who showed that, when there is nontrivial scattering, there are no further symmetries of the *S*-matrix.

## Appendix

In an effort to make this article more self-contained, this appendix collects together a few simple pertinent concepts and results from the theory of operator algebras. A C*-algebra is a $*$-algebra $\mathcal{A}$ with a norm $\|\cdot\|$ making it into a Banach algebra and satisfying

$$\|A^*A\| = \|A\|^2$$

for every $A \in \mathcal{A}$. Any C*-algebra can be realized as a norm closed $*$-subalgebra of the C*-algebra $\mathcal{B}(\mathcal{H})$ of all bounded operators on a Hilbert space $\mathcal{H}$. A von Neumann algebra $\mathcal{R}$ is a C*-algebra that is the dual space of a Banach space. This Banach space $\mathcal{R}_*$, the predual of $\mathcal{R}$, is intrinsically defined. The topology on $\mathcal{R}$ determined by duality with $\mathcal{R}_*$ is called the $\sigma$-topology. $\mathcal{B}(\mathcal{H})$ is a von Neumann algebra and its predual is the set of trace class operators. Any von Neumann algebra can be realized as a $\sigma$-closed unital $*$-subalgebra of some $\mathcal{B}(\mathcal{H})$.

A state on a C*-algebra $\mathcal{A}$ is a positive linear functional $\omega$ of norm 1. If $\mathcal{A}$ has a unit $I$ the normalization condition can be expressed as $\omega(I) = 1$. Of fundamental importance is the relation between representations and states. A representation of $\mathcal{A}$ on a Hilbert space $\mathcal{H}$ is just a structure-preserving mapping or morphism of $\mathcal{A}$ into $\mathcal{B}(\mathcal{H})$. For simplicity, we suppose that $\mathcal{A}$ has a unit. Given a state $\omega$, there is an associated representation $\pi_\omega$ defined by a vector $\Omega$ such that $\pi_\omega(\mathcal{A})\Omega$ is dense in the Hilbert space in question, that is, it is a cyclic vector for the representation and

$$\omega(A) = (\Omega, \pi_\omega(A)\Omega), \quad A \in \mathcal{A}$$

that is, the cyclic vector implements the given state. This is referred to as the GNS construction. Given any two such representations, there is a unique unitary operator mapping the one cyclic vector onto the other and realizing the equivalence of the representations.

A state of a von Neumann algebra is said to be normal if it is continuous in the $\sigma$-topology. If $\omega$ is normal, then $\pi_\omega(\mathcal{R})$ is $\sigma$-closed.

An inclusion of unital von Neumann algebras has the split property if there is an intermediate type I factor, that is, if it has the form $\mathcal{R}_1 \subset \mathcal{B}(\mathcal{H}) \subset \mathcal{R}_2$.

The following elementary observation is often used in treating symmetries. If $\alpha$ is an automorphism of $\mathcal{A}$ with $\omega\alpha^{-1} = \omega$, there is a unique unitary operator leaving the cyclic vector $\Omega$ invariant and inducing $\alpha$ in the representation $\pi_\omega$. In other words, $U\Omega = \Omega$ and

$$U\pi_\omega(A)U^{-1} = \pi_\omega(\alpha A)$$

If we apply the above lemma to a group $G$ of symmetries leaving a state invariant, it yields a group $U(g)$ of unitaries satisfying the condition

$$U(gh) = U(g)U(h), \quad g, h \in G$$

since $U(g)$ is uniquely defined by the above conditions.

When there is no invariant state, the situation is more complicated. Suppose there is a group $G$ of symmetries and a representation $\pi$ of $\mathfrak{A}$ where each $g$ is unitarily implemented. Thus, there is a unitary $U(g)$ with

$$U(g)\pi(A)U(g)^{-1} = \pi(gA), \quad A \in \mathfrak{A}$$

All we can now conclude is that

$$U(gh) = Z(g, h)U(g)U(h)$$

where $Z(g, h)$ is a unitary in $\mathfrak{A}'$, the commutant of $\mathfrak{A}$, satisfying the 2–cocycle identity

$$Z(gh, k)Z(g, h) = Z(g, hk)^g Z(h, k)$$

where ${}^g X = U(g)XU(g)^{-1}$. $U$ is said to be a representation up to a factor. It can be chosen to be a representation if the cocycle $Z$ is a coboundary, that is, if there is a unitary $Y(g)$ in $\mathfrak{A}'$ such that

$$Y(g)^g Y(h) = Y(gh)Z(g, h)$$

In general, little is known about solving problems of this kind, but there are a number of results when $\pi$ is irreducible and the unitary group of its commutant reduces to the circle.

We turn now to consider the modular theory of von Neumann algebras. A vector $\Omega$ is said to be separating for a von Neumann algebra $\mathcal{R}$ if $A\Omega = 0$ and $A \in \mathcal{R}$ implies $A = 0$. If $\Omega$ is both cyclic and separating, there is a uniquely determined closed antilinear involution $S$ with $SA\Omega = A^*\Omega$ for $A \in \mathcal{R}$. If $S = J\Delta^{1/2}$ is the polar decomposition of $S$, then the unitary operators $\Delta^{it}$ induce automorphisms $\delta^{it}$ of $\mathcal{R}$ and $J\mathcal{R}J = \mathcal{R}'$. $J$ is called the modular conjugation, $\Delta$ the modular operator, and $\delta^{it}$ the modular automorphisms. The closure of $\{\Delta^{1/4}A\Omega : A \in \mathcal{R}, A \geq 0\}$ is a cone, called the natural cone. Every normal state of $\mathcal{R}$ is implemented by a unique vector in the natural cone. If $\alpha$ is an automorphism of $\mathcal{R}$, there is therefore a unique vector $\Omega_\alpha$ in the natural cone such that, for every $A \in \mathcal{R}$,

$$(\Omega, \alpha^{-1}(A)\Omega) = (\Omega_\alpha, A\Omega_\alpha)$$

There is now a canonical unitary operator $V_\alpha$ defined by

$$V_\alpha A\Omega = \alpha(A)\Omega_\alpha$$

$V_\alpha$ maps the natural cone into itself and $\alpha \mapsto V_\alpha$ is an implementing representation of the group of automorphisms of $\mathcal{R}$. Under these circumstances, we do not have to deal with representations up to a factor.

*See also:* Algebraic Approach to Quantum Field Theory; Axiomatic Quantum Field Theory; Boundary Conformal Field Theory; Current Algebra; Quantum Fields with Topological Defects; Supergravity; Symmetries in Quantum Field Theory of Lower Spacetime Dimensions; Two-Dimensional Models.

## Further Reading


Bisognano JJ and Wichmann EH (1976) On the duality condition for quantum fields. *Journal of Mathematical Physics* 17: 303–321.

Borchers HJ and Buchholz D (1985) The energy–momentum spectrum in local field theories with broken Lorentz symmetry. *Communications in Mathematical Physics* 97: 169–185.

Buchholz D (1982) The physical state space of quantum electrodynamics. *Communications in Mathematical Physics* 85: 49–71.

Buchholz D, Doplicher S, and Longo R (1986) On Noether's theorem in quantum field theory. *Annals of Physics* 170: 1–17.

Buchholz D, Doplicher S, Longo R, and Roberts JE (1992) A new look at Goldstone's theorem. *Reviews in Mathematical Physics* (Special Issue): 49–83.

Buchholz D, Dreyer O, Florig M, and Summers SJ (2000) Geometric modular action and spacetime symmetry groups. *Reviews in Mathematical Physics* 12: 475–560.

Coleman S and Mandula J (1967) All possible symmetries of the *S*-matrix. *Physical Review* 159: 1251–1256.

Doplicher S and Roberts JE (1990) Why there is a field algebra with a compact gauge group describing the superselection structure in particle physics. *Communications in Mathematical Physics* 131: 51–107.

Guido D and Longo R (1996) The conformal spin and statistics theorem. *Communications in Mathematical Physics* 181: 11–35.

Lopuszański J (1991) *An Introduction to Symmetry and Supersymmetry in Quantum Field Theory*. Singapore: World Scientific.

O'Raifeartaigh L (1965) Lorentz invariance and internal symmetry. *Physical Review* 139: B1052–B1062.

# Symmetry and Symmetry Breaking in Dynamical Systems

**I Melbourne**, University of Surrey, Guildford, UK



## Introduction

The same symmetries may underlie diverse contexts such as phase transitions of crystals (Landau theory), fluid dynamics, and problems in biology and chemical engineering. Hence, seemingly unrelated systems may exhibit similar phenomena in regard to symmetries of patterns and transitions between patterns (spontaneous symmetry breaking). It is natural to focus attention on aspects of pattern formation that are universal or model independent – aspects depending on underlying symmetries rather than model-specific details.

The general framework is that the underlying system is governed by an evolution equation

$$\dot{x} = f(x) \qquad [1]$$

with symmetry group $\Gamma$. To avoid technicalities, we assume that [1] is an ordinary differential equation (ODE), the vector field $f:\mathbb{R}^n \to \mathbb{R}^n$ is as smooth as desired, and $\Gamma$ is a compact Lie group acting linearly on $\mathbb{R}^n$. An inner product may be chosen so that $\Gamma$ acts orthogonally. The vector field in [1] is $\Gamma$-equivariant if

$$f(\gamma x) = \gamma f(x) \quad \text{for all } x \in \mathbb{R}^n,\ \gamma \in \Gamma \qquad [2]$$

Equivalently, if $x(t)$ is a solution and $\gamma \in \Gamma$, then $\gamma x(t)$ is a solution.

In this article, we are interested in the dynamics to be expected for equivariant vector fields, and transitions that arise as parameters are varied. The symmetry group $\Gamma$ is taken as given, whereas $f$ is a general $\Gamma$-equivariant vector field. (Other features such as energy conservation or time reversibility must be built into the general setup, but are excluded in this article.)

## Isotropy Subgroups and Commuting Linear Maps

Let $\Gamma$ be a compact Lie group acting linearly on $\mathbb{R}^n$. The isotropy subgroup of $x \in \mathbb{R}^n$ is defined to be

$$\Sigma_x = \{\gamma \in \Gamma:\ \gamma x = x\}$$

Note that $\Sigma_{\gamma x} = \gamma \Sigma_x \gamma^{-1}$ for all $x \in \mathbb{R}^n$, $\gamma \in \Gamma$.

Given an isotropy subgroup $\Sigma \subset \Gamma$, define the fixed-point subspace

$$\mathrm{Fix}\,\Sigma = \{y \in \mathbb{R}^n\colon \sigma y = y \text{ for all } \sigma \in \Sigma\}$$

If $f:\mathbb{R}^n \to \mathbb{R}^n$ is a $\Gamma$-equivariant vector field, then $f(\mathrm{Fix}\,\Sigma) \subset \mathrm{Fix}\,\Sigma$ for each isotropy subgroup $\Sigma$. Hence $\mathrm{Fix}\,\Sigma$ is flow invariant.

The normalizer $N(\Sigma) = \{\gamma \in \Gamma : \gamma\Sigma\gamma^{-1} = \Sigma\}$ is the largest subgroup of $\Gamma$ that acts on $\mathrm{Fix}\,\Sigma$, and $f_\Sigma = f|_{\mathrm{Fix}\Sigma}$ is $(N(\Sigma)/\Sigma)$-equivariant.

An isotropy subgroup $\Sigma$ is axial if $\dim \mathrm{Fix}\,\Sigma = 1$, and then $N(\Sigma)/\Sigma \cong \mathbb{Z}_2$ or $\mathbf{1}$. More generally, $\Sigma$ is maximal if there are no isotropy subgroups $T$ with $\Sigma \subset T \subset \Gamma$ other than $T = \Sigma$ and $T = \Gamma$. Then $N(\Sigma)/\Sigma$ acts fixed-point freely on $\mathrm{Fix}\,\Sigma$ and the connected component of the identity $(N(\Sigma/\Sigma)^0 \cong \mathbf{1}$, SO(2) or SU(2). Correspondingly $\Sigma$ is called real, complex, or quaternionic. In the complex case $\dim \mathrm{Fix}\,\Sigma$ is even; in the quaternionic case $\dim \mathrm{Fix}\,\Sigma \equiv 0 \bmod 4$.

The dihedral group $\Gamma = \mathbb{D}_m$ of order $m$ is the symmetry group of the regular $m$-gon, $m \geq 3$. Its standard action on $\mathbb{R}^2$ is generated by

$$\rho = \begin{pmatrix} \cos 2\pi/m & -\sin 2\pi/m \\ \sin 2\pi/m & \cos 2\pi/m \end{pmatrix}$$

$$\kappa = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

For $m$ even, the isotropy subgroups up to conjugacy are

$$\mathbb{D}_m, \quad \mathbb{Z}_2(\kappa), \quad \mathbb{Z}_2(\rho\kappa), \quad \mathbf{1}$$

where $\mathbb{Z}_j(g)$ denotes the cyclic group of order $j$ generated by $g$. The maximal isotropy subgroups $\Sigma = \mathbb{Z}_2(\kappa), \mathbb{Z}_2(\kappa\rho)$ are axial with $N(\Sigma)/\Sigma \cong \mathbb{Z}_2$. For $m$ odd, $\mathbb{Z}_2(\rho\kappa)$ is conjugate to $\mathbb{Z}_2(\kappa)$ leaving three conjugacy classes of isotropy subgroups, and $\Sigma = \mathbb{Z}_2(\kappa)$ is axial with $N(\Sigma)/\Sigma = \mathbf{1}$.

The space of commuting linear maps

$$\mathrm{Hom}_\Gamma(\mathbb{R}^n) = \{L : \mathbb{R}^n \to \mathbb{R}^n \text{ linear: } L\gamma = \gamma L \text{ for all } \gamma \in \Gamma\}$$

is completely described representation-theoretically. Recall that $\Gamma$ acts irreducibly on $\mathbb{R}^n$ if the only $\Gamma$-invariant subspaces of $\mathbb{R}^n$ are $\mathbb{R}^n$ and $\{0\}$. Then $\mathrm{Hom}_\Gamma(\mathbb{R}^n)$ is a real division ring (skew field) $\mathcal{D} \cong \mathbb{R}$, $\mathbb{C}$ or $\mathbb{H}$. The representation is called absolutely irreducible when $\mathcal{D} = \mathbb{R}$ and nonabsolutely irreducible when $\mathcal{D} = \mathbb{C}$ or $\mathbb{H}$.

If the action of $\Gamma$ is not irreducible, write $\mathbb{R}^n = V_1 \oplus \cdots \oplus V_k$ (nonuniquely) as a sum of irreducible subspaces. Summing together irreducible subspaces that are isomorphic to form isotypic components $W$ gives the (unique) isotypic decomposition $\mathbb{R}^n = W_1 \oplus \cdots \oplus W_\ell$. If $L \in \mathrm{Hom}_\Gamma(\mathbb{R}^n)$, then $L(W_j) \subset W_j$ for each $j$, hence $\mathrm{Hom}_\Gamma(\mathbb{R}^n) = \mathrm{Hom}_\Gamma(W_1) \oplus \cdots \oplus \mathrm{Hom}_\Gamma(W_\ell)$. Each $W_j$ consists of $k_j$ isomorphic copies of an irreducible representation with division ring $\mathcal{D}_j$. Let $M_k(\mathcal{D})$ denote the space of $k \times k$ matrices with entries in $\mathcal{D}$. Then

$$\mathrm{Hom}_\Gamma(\mathbb{R}^n) \cong M_{k_1}(\mathcal{D}_1) \oplus \cdots \oplus M_{k_\ell}(\mathcal{D}_\ell) \qquad [3]$$

Spectral properties of commuting linear maps can be recovered from the decomposition [3], paying due attention to multiplicity and complex conjugates of eigenvalues.

## Equivariant Dynamics

The dynamics of equivariant systems includes (relative) equilibria and periodic solutions, robust heteroclinic cycles/networks, and symmetric chaotic attractors.

### Equilibria

Consider the ODE [1] with $\Gamma$-equivariant vector field $f$ satisfying [2]. If $x(t) \equiv x_0$ is an equilibrium, $f(x_0) = 0$, then there is a group orbit $\Gamma x_0$ of equilibria.

Let $\Sigma = \Sigma_{x_0}$ be the isotropy subgroup of $x_0$. If $\dim \Sigma = \dim \Gamma$, then generically (for an open dense set of $\Gamma$-equivariant vector fields), the eigenvalues of $(df)_{x_0}$ have nonzero real part, hence $x_0$ is hyperbolic. If the eigenvalues all have negative real part, then $x_0$ is asymptotically stable. If at least one eigenvalue has positive real part, then $x_0$ is unstable. Hyperbolic equilibria are isolated and persist under perturbations of $f$; the perturbed equilibria continue to have isotropy $\Sigma$. Since $(df)_{x_0} \in \mathrm{Hom}_\Sigma(\mathbb{R}^n)$, decomposition [3] for the action of $\Sigma$ on $\mathbb{R}^n$ facilitates stability computations for $x_0$.

If $\dim \Sigma < \dim \Gamma$, then $\Gamma x_0$ is a continuous group orbit of equilibria. Generically, $\dim \ker (df)_{x_0} = \dim \Gamma - \dim \Sigma$ and $\ker (df)_0 = \{\xi x_0 : \xi \in L\Gamma\}$, where $L\Gamma$ is the Lie algebra of $\Gamma$. The remaining $k = n - \dim \Gamma + \dim \Sigma$ eigenvalues generically have nonzero real part so $\Gamma x_0$ is normally hyperbolic. If all $k$ eigenvalues have nonzero real part, then $\Gamma x_0$ is asymptotically stable. If at least one has positive real part, then $\Gamma x_0$ is unstable. When $N(\Sigma)/\Sigma$ is finite, generically $x_0$ is an isolated equilibrium in $\mathrm{Fix}\, \Sigma$ and persists as an equilibrium with isotropy $\Sigma$ under perturbation.

### Relative Equilibria and Skew Products

A point $x_0 \in \mathbb{R}^n$ (or the corresponding group orbit $\Gamma x_0$) is a relative equilibrium if $f(x_0) \in T_{x_0}\Gamma x_0 = L\Gamma x_0$. If $x_0$ has isotropy $\Sigma$, then $x_0$ is a relative equilibrium if $f(x_0) \in LD_\Sigma x_0$, where $D_\Sigma = (N(\Sigma)/\Sigma)^0$.

Write $f(x_0) = \xi x_0$, where $\xi \in LD_\Sigma$. The closure of the one-parameter subgroup $\exp(t\xi)$ is a maximal torus in $D_\Sigma$ for almost every $\xi$. All maximal tori are conjugate with common dimension $d = \mathrm{rank}\, D_\Sigma$. The solution $x(t) = \exp(t\xi)x_0$ is typically a $d$-dimensional quasiperiodic motion. "Typically" holds in both the topological and probabilistic sense and there is no phase-locking. When $d = 1$, $x(t)$ is periodic, often called a rotating wave.

Choose a $\Sigma$-invariant local cross section $X$ to the group orbit $\Gamma x_0$ at $x_0$. There is a $\Gamma$-invariant neighborhood of $\Gamma x_0$ that is $\Gamma$-equivariantly diffeomorphic to $(\Gamma \times X)/\Sigma$, where $\Sigma$ acts freely on $\Gamma \times X$ by

$$\sigma \cdot (\gamma, x) = (\gamma\sigma^{-1}, \sigma x)$$

and $\Gamma$ acts by left multiplication on the first factor. The $\Gamma$-equivariant ODE on $(\Gamma \times X)/\Sigma$ lifts to a $(\Gamma \times \Sigma)$-equivariant skew product on $\Gamma \times X$

$$\dot{\gamma} = \gamma\xi(x), \qquad \dot{x} = h(x) \qquad [4]$$

where $\xi : X \to L\Gamma$, $h : X \to X$ satisfy the $\Sigma$-equivariance conditions

$$\xi(\sigma x) = \mathrm{Ad}_\sigma \xi(x) = \sigma\xi(x)\sigma^{-1}$$
$$h(\sigma x) = \sigma h(x)$$

and $h(x_0) = 0$.

Thus, dynamics near the relative equilibrium $\Gamma x_0 \subset \mathbb{R}^n$ reduces to dynamics near the ordinary equilibrium $x_0 \in X$ for the $\Sigma$-equivariant vector $h : X \to X$, coupled with $\Gamma$ drifts. In particular, the stability of $\Gamma x_0$ is determined by $(dh)_{x_0}$.

### Periodic Solutions

A nonequilibrium solution $x(t)$ is periodic if $x(t + T) = x(t)$ for some $T > 0$. The least such $T$ is the (absolute) period. The spatial symmetry group $\Delta$ is the isotropy subgroup of $x(t)$ for some, and hence all, $t \in \mathbb{R}$. The periodic solution $P = \{x(t): 0 \le t < T\}$ lies inside $\mathrm{Fix}\, \Delta$. Define the spatiotemporal symmetry group $\Sigma = \{\gamma \in \Gamma : \gamma P = P\}$. Note that $\Delta$ is a normal subgroup of $\Sigma$ and either $\Sigma/\Delta \cong S^1$ ($P$ is a rotating wave) or $\Sigma/\Delta \cong \mathbb{Z}_q$ and $P$ is called a standing wave or a discrete rotating wave. For each $\sigma \in \Sigma$, there exists $T_\sigma \in [0, T)$ such that $\sigma x(t) = x(t + T_\sigma)$. The relative period of $x(t)$ is the least $T > 0$ such that $x(T) \in \Sigma x_0$.

If $\dim \Sigma = \dim \Gamma$, then generically $P$ is hyperbolic, hence isolated, the stability of $P$ is determined by its

Floquet exponents, and $P$ persists under perturbation as a periodic solution with spatial symmetry $\Delta$ and spatiotemporal symmetry $\Sigma$. For $\Gamma$ infinite and $N(\Delta)/\Delta$ finite, generically $P$ is isolated in $\mathrm{Fix}\,\Delta$ and the neutral Floquet exponent has multiplicity $\dim\Gamma - \dim\Sigma + 1$.

### Relative Periodic Solutions

A solution $x(t)$ is a relative periodic solution if it is not a relative equilibrium and $x(T) \in \Gamma x(0)$ for some $T > 0$. The least such $T$ is the relative period. The spatial symmetry group $\Delta = \Sigma_{x(t)}$ for some, hence all, $t$. The spatiotemporal symmetry group $\Sigma$ is the closed subgroup of $\Gamma$ generated by $\Delta$ and $\sigma$, where $x(T) = \sigma x(0)$, and generically $\Sigma/\Delta \cong \mathbb{T}^d \times \mathbb{Z}_q$ is a maximal topologically cyclic (Cartan) subgroup of $N(\Delta)/\Delta$ containing $\sigma\Delta$. Then $x(t)$ is a $(d+1)$-dimensional quasiperiodic motion.

The dynamics near the relative periodic solution is again governed by a skew product. There exists $n \geq 1$ such that $\sigma^n = \exp(n\xi)$, where $\xi \in LZ(\Sigma)$ and $Z(\Sigma) \subset \Gamma$ is the centralizer of $\Sigma$. Define $\alpha = \exp(-\xi)\sigma$. Form a semidirect product $\Delta \rtimes \mathbb{Z}_{2n}$ by adjoining to $\Delta$ an element $Q$ of order $2n$ such that $Q\delta Q^{-1} = \sigma\delta\sigma^{-1}$ for $\delta \in \Delta$.

In a comoving frame with velocity $\xi$, a neighborhood of the relative periodic orbit is $\Gamma$-equivariantly diffeomorphic to $(\Gamma \times X \times S^1)/\Delta \rtimes \mathbb{Z}_{2n}$, where $X$ is a $\Delta \rtimes \mathbb{Z}_{2n}$-invariant cross section, $S^1 = \mathbb{R}/2n\mathbb{Z}$ and $\Delta \rtimes \mathbb{Z}_{2n}$ acts on $\Gamma \times X \times S^1$ as

$$\delta \cdot (\gamma, x, \theta) = (\gamma\delta^{-1}, \delta x, \theta)$$
$$Q \cdot (\gamma, x, \theta) = (\gamma\alpha^{-1}, Qx, \theta + 1)$$

The $\Gamma$-equivariant ODE on $(\Gamma \times X \times S^1)/\Delta \rtimes \mathbb{Z}_{2n}$ lifts to a $\Gamma \times (\Delta \rtimes \mathbb{Z}_{2n})$-equivariant skew product

$$\dot{\gamma} = \gamma\xi(x, \theta), \quad \dot{x} = h(x, \theta), \quad \dot{\theta} = 1 \qquad [5]$$

where $\xi : X \times S^1 \to L\Gamma$, $h : X \times S^1 \to X$ satisfy appropriate $\Delta \rtimes \mathbb{Z}_{2n}$-equivariance conditions.

### Robust Heteroclinic Cycles

Heteroclinic cycles, degenerate in systems without symmetry, arise robustly in equivariant systems. Let $x_1, \ldots, x_m \in \mathbb{R}^n$ be saddles with $W^{\mathrm{u}}(x_i) - \{x_i\} \subset \Gamma W^{\mathrm{s}}(x_{i+1})$ (where $m + 1 = 1$). If $\Sigma_1, \ldots, \Sigma_m \subset \Gamma$ are isotropy subgroups, $W^{\mathrm{u}}(x_i) \subset \mathrm{Fix}\,\Sigma_i$, and $x_{i+1}$ is a sink in $\mathrm{Fix}\,\Sigma_i$, then saddle–sink connections from $x_i$ to $x_{i+1}$ persist for nearby $\Gamma$-equivariant flows. The union $\bigcup_{i=1}^{m} \Gamma W^{\mathrm{u}}(x_i)$ forms a robust heteroclinic cycle (see the subsection "Dynamics" for an example). Such cycles, when asymptotically stable, are a mechanism for intermittency or bursting, notably in rotating Rayleigh–Bénard convection (where rolls disappear and reorient themselves at approximately 60°), and provide a possible intrinsic explanation for irregular reversals of the Earth's magnetic field.

Asymmetric perturbations (deterministic or noisy) destroy the cycles, but the perturbed attractors inherit the bursting behavior.

Establishing the existence of heteroclinic connections is often straightforward when $\dim \mathrm{Fix}\,\Sigma_i = 2$ and nontrivial with $\dim \mathrm{Fix}\,\Sigma_i \geq 3$. Criteria for asymptotic stability of heteroclinic cycles are given in terms of real parts of eigenvalues of $(df)_{x_i}$, and depend on the geometry of the representation of $\Gamma$.

Robust cycles exist also between more complicated dynamical states such as periodic solutions or chaotic sets (cycling chaos). When $W^u(x_i)$ connects to two or more distinct states, the collection of unstable manifolds forms a heteroclinic network leading to competition between various subnetworks.

### Symmetric Attractors

Suppose that $\Gamma$ is a finite group acting linearly on $\mathbb{R}^n$. A closed subset $A \subset \mathbb{R}^n$ has symmetry groups $\Delta = \{\gamma \in \Gamma\colon \gamma x = x \text{ for all } x \in A\}, \Sigma = \{\gamma \in \Gamma\colon \gamma A = A\}$. Here, $\Delta$ is an isotropy subgroup and $\Delta \subset \Sigma \subset N(\Delta)$. In applications, $\Delta$ corresponds to instantaneous symmetry and $\Sigma$ to symmetry on average.

If $A$ is an attractor (a Lyapunov stable $\omega$-limit set) for a $\Gamma$-equivariant vector field $f : \mathbb{R}^n \to \mathbb{R}^n$, then $\Sigma$ fixes a connected component of $\mathrm{Fix}\,\Delta - L$, where $L$ is the union of proper fixed-point spaces in $\mathrm{Fix}\,\Delta$.

Provided $\dim \mathrm{Fix}\,\Delta \geq 3$, all pairs $\Delta, \Sigma$ satisfying the above restrictions arise as symmetry groups of a nonperiodic attractor $A$. If $\dim \mathrm{Fix}\,\Delta \geq 5$, then $A$ is realized by a uniformly hyperbolic (Axiom A) attractor.

If $\dim \mathrm{Fix}\,\Delta \geq 3$ and $\Sigma$ fixes a connected component of $\mathrm{Fix}\,\Delta - L$, then $A$ is realized by a periodic sink provided $\Sigma/\Delta$ is cyclic. If $\dim \mathrm{Fix}\,\Delta = 2$, then in addition either $\Sigma = \Delta$ or $\Sigma = N(\Delta)$.

Suppose $A$ is an attractor and $\gamma \in \Gamma - \Sigma$. Then $\gamma A \cap A = \emptyset$. Varying a parameter, $A$ may undergo a symmetry-increasing bifurcation: $A$ grows until it collides with $\gamma A$ producing a larger attractor with symmetry on average generated by $\Sigma$ and $\gamma$.

Determining symmetries of an attractor by inspection is often infeasible. A detective is a $\Gamma$-equivariant polynomial $\phi : \mathbb{R}^n \to V$ where every subgroup of $\Gamma$ is an isotropy subgroup for the action on $V$, and each component of $\phi$ is nonzero. Suppose that $A \subset \mathbb{R}^n$ is an attractor with physical (Sinai–Ruelle–Bowen) measure $\mu$. By ergodicity, the time average

$$\psi_A = \lim_{T \to \infty} \frac{1}{T} \int_0^T \phi(x(t))\mathrm{d}t \in V$$

is well defined for almost every trajectory $x(t)$ in $\operatorname{supp}\mu$. Generically, $\Sigma_{\psi_A} = \Sigma_A$ so computing the symmetry of $A$ reduces to computing the symmetry of a point.

If $\Gamma$ is an infinite compact Lie group, and $A$ is an $\omega$-limit set containing points of trivial isotropy, then $A$ cannot be uniformly hyperbolic. Hence partially hyperbolic flows arise naturally in systems with continuous symmetry. Consider the skew product [4] where $\Sigma = \mathbf{1}$ and $h: X \to X$ possesses a hyperbolic basic set $\Lambda \subset X$ with equilibrium measure $\mu$ (for a Hölder potential). Let $\nu$ denote Haar measure on $\Gamma$. Then $\Lambda \times \Gamma$ is partially hyperbolic, and $\mu \times \nu$ is ergodic (even Bernoulli) for an open dense set of equivariant flows. Such stably ergodic flows possess strong statistical properties (rapid decay of correlations, central-limit theorem); a possible explanation for hypermeander (Brownian-like motion) of spiral waves in planar excitable media.

### Forced Symmetry Breaking

In applications, symmetry is not perfect and account should be taken of $\Gamma'$-equivariant perturbations of [1] for $\Gamma'$ a subgroup of $\Gamma$ (including $\Gamma' = \mathbf{1}$). This topic is not discussed in this article, except in the subsections "Robust heteroclinic cycles" and "Branching patterns and finite determinacy."

## Equivariant Bifurcation Theory

Consider families of ODEs $\dot{x} = f(x, \lambda)$, with bifurcation parameter $\lambda \in \mathbb{R}$ and vector field $f: \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ satisfying $f(0, 0) = 0$ and the $\Gamma$-equivariance condition

$$f(\gamma x, \lambda) = \gamma f(x, \lambda) \quad \text{for all } x \in \mathbb{R}^n, \ \lambda \in \mathbb{R}, \ \gamma \in \Gamma$$

A local bifurcation from the equilibrium $x = 0$ occurs if $(df)_{0,0}$ is nonhyperbolic. The center subspace $E^c$ is the sum of generalized eigenspaces corresponding to eigenvalues on the imaginary axis, and is $\Gamma$-invariant. By center manifold theory, local dynamics ($(x, \lambda)$ near $(0, 0)$) are captured by the center manifold $W^c$. After center manifold reduction (or Lyapunov–Schmidt reduction if the focus is on equilibria), it may be assumed that $\mathbb{R}^n = E^c$.

If $(df)_{0,0}$ possesses zero eigenvalues, then there is a steady-state bifurcation. Generically, $(df)_{0,0} = 0$ and $E^c$ is absolutely irreducible. There are two subcases.

If $\Gamma$ acts trivially on $\mathbb{R}^n$, then $n = 1$ and generically there is a saddle–node (or limit point) bifurcation where the zero sets of $f(x, \lambda)$ and $\pm x^2 \pm \lambda$ are diffeomorphic for $(x, \lambda)$ near $(0, 0)$. Higher-order degeneracies can be treated using singularity theory. The equilibria and their stability determines the local dynamics. All bifurcating equilibria have isotropy $\Gamma$, so there is no symmetry breaking.

From now on, consider the remaining subcase where $\Gamma$ acts absolutely irreducibly and nontrivially on $\mathbb{R}^n$. Then $\operatorname{Fix}\Gamma = \{0\}$, $f(0, \lambda) \equiv 0$, and $(df)_{0,\lambda} = c(\lambda) I_n$ where generically $c'(0) \neq 0$. Assume that $c'(0) > 0$, so the "trivial solution" $x = 0$ is asymptotically stable subcritically ($\lambda < 0$) and unstable supercritically ($\lambda > 0$). Bifurcating solutions lie outside $\operatorname{Fix}\Gamma$ and hence there is spontaneous symmetry breaking.

### Axial Isotropy Subgroups

The "equivariant branching lemma" guarantees branches of equilibria with isotropy $\Sigma$ for each axial isotropy subgroup. There are three associated branching patterns, see Figure 1.

If $N(\Sigma)/\Sigma = \mathbb{Z}_2$, then $f_\Sigma$ is odd. Generically, $\partial_x^3 f_\Sigma(0, 0) \neq 0$, since $(x_1^2 + \cdots + x_n^2)x$ is $\Gamma$-equivariant, and there are two branches of equilibria bifurcating supercritically or subcritically together, and lying on the same group orbit. The branches form a symmetric pitchfork whose direction of branching is determined by $\operatorname{sgn} \partial_x^3 f_\Sigma(0, 0)$.

If $N(\Sigma)/\Sigma \cong \mathbf{1}$, then generically $f_\Sigma$ is even. If all quadratic $\Gamma$-equivariant maps vanish on $\operatorname{Fix}\Sigma$, then the bifurcation is sub/supercritical depending on $\operatorname{sgn} \partial_x^3 f_\Sigma(0, 0)$ but the branches lie on distinct group orbits. This is an asymmetric pitchfork.

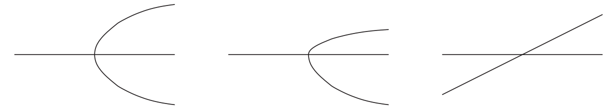
If $\partial_x^2 f_\Sigma(0, 0) \neq 0$, then the equilibria exist transcritically: for $\lambda < 0$ and $\lambda > 0$.

The natural actions of $\mathbb{D}_m$ on $\mathbb{R}^2$ are absolutely irreducible. The axial branches are symmetric pitchforks for $m \geq 4$ even, asymmetric pitchforks for $m \geq 5$ odd, and transcritical for $m = 3$.

The actions of $\mathbb{D}_m, m \geq 5$ odd, provide the simplest instances of hidden symmetries, where certain $N(\Sigma)/\Sigma$-equivariant mappings on $\operatorname{Fix}\Sigma$ do not extend to smooth $\Gamma$-equivariant mappings on $\mathbb{R}^n$.

### Nonaxial Maximal Isotropy Subgroups

For $\Sigma$ a real maximal isotropy subgroup, $\dim \operatorname{Fix}\Sigma$ odd, there exist branches of equilibria with isotropy



(a) (b) (c)

**Figure 1** Axial branches: (a) supercritical symmetric pitchfork, (b) supercritical asymmetric pitchfork, and (c) transcritical branches.

$\Sigma$. When $\dim \mathrm{Fix}\,\Sigma$ is even, there are examples where equilibria exist and examples where no equilibria exist. For $\Sigma$ complex or quaternionic, there exist branches of rotating waves with isotropy $\Sigma$. In the quaternionic case, the rotating waves foliate the SU(2) group orbits according to the Hopf fibration.

### Submaximal Isotropy Subgroups

It has been conjectured falsely that steady-state bifurcation leads generically to equilibria only with maximal isotropy. The simplest counterexample is the 24-element group $\Gamma = \mathbb{Z}_3\mathbb{Z}_2^3$ generated by

$$\rho = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad \kappa = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

(Alternatively, $\Gamma = \mathrm{T} \oplus \mathbb{Z}_2(-I_3)$, where $\mathrm{T} \subset \mathrm{SO}(3)$ is the tetrahedral group.)

The isotropy subgroup $\Sigma = \mathbb{Z}_2(\kappa)$ has two-dimensional fixed-point subspace $\mathrm{Fix}\,\Sigma = \{(x, y, 0)\}$. The only one-dimensional fixed-point spaces contained in $\mathrm{Fix}\,\Sigma$ are the $x$- and $y$-axes. The general $\Gamma$-equivariant vector field is

$$\begin{aligned} \dot{x} &= g(x^2, y^2, z^2, \lambda)x \\ \dot{y} &= g(y^2, z^2, x^2, \lambda)y \\ \dot{z} &= g(z^2, x^2, y^2, \lambda)z \end{aligned}$$

After scaling,

$$\begin{aligned} &g(x^2, y^2, z^2, \lambda) \\ &\quad = \lambda - x^2 - ay^2 - bz^2 + o(x^2, y^2, z^2, \lambda) \end{aligned} \qquad [6]$$

Restricting to $\mathrm{Fix}\,\Sigma$ and dividing out the axial solutions $x = 0$ and $y = 0$ yields at lowest order the equations $\lambda = x^2 + ay^2 = y^2 + bx^2$. Submaximal solutions exist provided $\mathrm{sgn}(a-1) = \mathrm{sgn}(b-1)$.

In general, the existence of equilibria with submaximal isotropy must be treated on a case-by-case basis (for each absolutely irreducible representation of $\Gamma$ and isotropy subgroup $\Sigma$).

### Asymptotic Stability

Subcritical and axial transcritical branches are automatically unstable. Moreover, the existence of a quadratic $\Gamma$-equivariant mapping $q : \mathbb{R}^n \to \mathbb{R}^n$ and $x \in \mathrm{Fix}\,\Sigma$ such that $(dq)_x$ has eigenvalues with nonzero real part guarantees that branches of equilibria with axial isotropy $\Sigma$ are generically unstable (even when $q|_{\mathrm{Fix}\Sigma} \equiv 0$).

There are no general results for asymptotic stability, and calculations must be done on a case-by-case basis. (The remarks in the subsection “Equilibria” are useful here.)

### Branching Patterns and Finite Determinacy

The following notion of finite determinacy is based on equivariant transversality theory. Assume $\Gamma$ acts absolutely irreducibly. Consider the set $\mathcal{F}$ of $\Gamma$-equivariant vector fields $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ satisfying $(df)_{0,0} = 0$. For an open dense subset of $\mathcal{F}$, branches of relative equilibria near $(0, 0)$ are normally hyperbolic. The collection of branches of relative equilibria, together with their isotropy type, direction of branching, and stability properties, is called a branching pattern. These persist under small perturbations and are finitely determined: there exist $q = q_\Gamma \geq 2$ and an open dense subset $\mathcal{U}(q) \subset \mathcal{F}$ such that the branching patterns of $f$ and $f + g$ are identical for $f \in \mathcal{U}(q)$, $g \in \mathcal{F}$, provided $g(x, \lambda) = o(\|x\|^q)$.

Furthermore, branching patterns are strongly finitely determined: there exist $d \geq 2$ and an open dense subset $\mathcal{S}(d) \subset \mathcal{F}$ such that the branching patterns of $f$ and $f + g$ are identical for $f \in \mathcal{S}(d)$ and all (not necessarily equivariant) $g$ satisfying $g(x, \lambda) = o(\|x\|^d)$.

For example, consider the hyperoctahedral group $S_n\mathbb{Z}_2^n$, $n \geq 1$. Here $S_n$ acts by permutations of the coordinates $(x_1, \ldots, x_n)$ and $\mathbb{Z}_2^n$ consists of diagonal matrices with entries $\pm 1$. Let $\Gamma = T\,\mathbb{Z}_2^n$, where $T \subset S_n$ is a transitive subgroup. Then $\Gamma$ acts absolutely irreducibly on $\mathbb{R}^n$ and is strongly 3-determined. Submaximal branches of equilibria exist except when $T = S_n$, $T = A_n$ and, if $n = 6$, $T = \mathrm{PGL}_2(\mathbb{F}_5)$.

### Dynamics

Absolutely irreducible representations have arbitrarily high dimension, so steady-state bifurcation leads to rich dynamics. The group $\Gamma = \mathbb{Z}_3\mathbb{Z}_2^3$ with $\mathrm{sgn}(a-1) \neq \mathrm{sgn}(b-1)$ and $a + b > 2$ in [6] yields asymptotically stable heteroclinic cycles with planar connections connecting equilibria in the $x$-, $y$- and $z$-axes (see Figure 2). In $\mathbb{R}^4$, there is the possibility of instant chaos where chaotic dynamics bifurcates directly from the equilibrium 0.

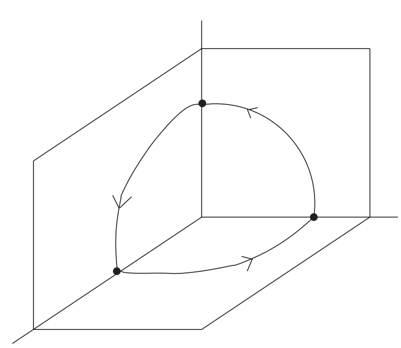


**Figure 2** Robust heteroclinic cycle for the group $\Gamma = \mathbb{Z}_3 \ltimes \mathbb{Z}_2^3$.

In the absence of quadratic equivariants, the invariant-sphere theorem gives an open set of equivariant vector fields for which an attracting normally hyperbolic flow-invariant $(n-1)$-dimensional sphere bifurcates supercritically. This simplifies computations of nontrivial dynamics.

## Hopf Bifurcation and Mode Interactions

### Equivariant Hopf Bifurcation

The setting is the same as in the last section, except that $L=(df)_{0,0}$ has imaginary eigenvalues $\pm i\omega$ of algebraic and geometric multiplicity $n/2$. Generically, $\mathbb{R}^n=E^c$ is Γ-simple: either the direct sum of two isomorphic absolutely irreducible subspaces, or nonabsolutely irreducible.

By Birkhoff normal-form theory (see below), for any $k\geq 1$ there is a Γ-equivariant change of coordinates after which $f(x,\lambda)=f_k(x,\lambda)+o(\|x\|^k)$, where $f_k$ is $(\Gamma\times S^1)$-equivariant. Here $S^1=\overline{\{\exp(tL)\colon t\in\mathbb{R}\}}$ acts freely on $\mathbb{R}^n$ and $\Gamma\times S^1$ acts complex irreducibly ($\mathcal{D}=\mathbb{C}$). Hence, $\dim \mathrm{Fix}\, J$ is even for each isotropy subgroup $J\subset\Gamma\times S^1$, and $N(J)/J\cong S^1$ when $J$ is maximal. The equivariant Hopf theorem guarantees, generically, branches of rotating waves with absolute period approximately $2\pi/\omega$ for each maximal isotropy subgroup $J$.

The notions of finite and strong finite determinacy extend to complex irreducible representations and the rotating waves persist as periodic solutions for the original Γ-equivariant vector field $f$. Define the spatial and spatiotemporal symmetry groups $\Delta\subset\Sigma\subset\Gamma$ as in the subsection "Periodic solutions." Then $J=\{(\sigma,\theta(\sigma))\colon \sigma\in\Sigma\}$ is a twisted subgroup, with $\theta:\Sigma\to S^1$ a homomorphism and $\Delta=J\cap\Gamma=\ker\theta$.

In the non-symmetry-breaking case, where Γ acts trivially on $\mathbb{R}^2$, phase–amplitude reduction leads to $\mathbb{Z}_2$-equivariant amplitude equations on $\mathbb{R}$ and higher-order degeneracies are amenable to $\mathbb{Z}_2$-equivariant singularity theory. Similar comments apply to O(2)-equivariant Hopf bifurcation where the amplitude equations are $\mathbb{D}_4$-equivariant. The technique fails for general groups Γ.

### Mode Interactions and Birkhoff Normal Form

Steady-state and Hopf bifurcations are codimension 1 and occur generically in one-parameter families of Γ-equivariant vector fields. Multiparameter families may undergo higher-codimension bifurcations called mode interactions. Suppressing parameters, steady-state/steady-state bifurcation occurs when $\mathbb{R}^n=E^c=V_1\oplus V_2$, where $V_1$ and $V_2$ are absolutely irreducible and $L=(df)_0$ has zero eigenvalues. If $V_1$ and $V_2$ are nonisomorphic then $L=0$, otherwise $L$ is nilpotent and there is an equivariant Takens–Bogdanov bifurcation. Similarly, there are codimension-2 steady-state/Hopf and Hopf/Hopf bifurcations.

Write $L=S+N$ (uniquely), where $S$ is semisimple, $N$ is nilpotent, and $SN=NS$. Then $\overline{\{\exp tS\colon t\in\mathbb{R}\}}$ is a torus $\mathbb{T}^p$, where $p\geq 0$ is the number of rationally independent eigenvalues for $L$.

For each $k\geq 1$, there is a Γ-equivariant degree-$k$ polynomial change of coordinates $P:\mathbb{R}^n\to\mathbb{R}^n$ satisfying $P(0)=0, (dP)_0=I$ transforming $f$ to Birkhoff normal form $f_k+o(\|x\|^k)$, where $f_k$ is $(\Gamma\times\mathbb{T}^p)$-equivariant.

If $N\neq 0$, then $\overline{\{\exp tN^T\colon t\in\mathbb{R}\}}\cong\mathbb{R}$ and $f_k$ can be chosen so that the nonlinear terms are $(\Gamma\times\mathbb{T}^p\times\mathbb{R})$-equivariant. The linear terms are not $\mathbb{R}$-equivariant.

The study of mode interactions proceeds by first analyzing $(\Gamma\times\mathbb{T}^p)$-equivariant normal forms, then considering exponentially small effects of the Γ-equivariant tail. Versions of the equivariant branching lemma and equivariant Hopf theorem establish existence of certain solutions. There are numerous examples of robust heteroclinic cycles connecting (relative) equilibria and periodic solutions, symmetric chaos, and symmetry-increasing bifurcations.

## Bifurcations from Relative Equilibria and Periodic Solutions

Using the skew product [4], bifurcations from a relative equilibrium with isotropy Σ for a Γ-equivariant vector field reduce to bifurcations from a fully symmetric equilibrium for a Σ-equivariant vector field $h$ coupled with Γ drifts. If $h$ possesses (relative) equilibria or periodic solutions, then the drift is determined generically as in the subsections "Relative equilibria and skew products" and "Relative periodic solutions." Nevertheless, solving the drift equation can be useful for understanding behavior in physical space. This is facilitated by making equivariant polynomial changes of coordinates $(\gamma Q(x),P(x))$ putting $h$ into Birkhoff normal form and simplifying $\xi$.

Bifurcations from (relative) periodic solutions also reduce, mainly, to bifurcations from equilibria (with enlarged symmetry group). Based on the discussion in the subsection "Relative periodic solutions," it suffices to consider bifurcations from isolated periodic solutions $P=\{x(t)\}$ with spatial symmetry Δ and spatiotemporal symmetry Σ. Write $x(T)=\sigma x(0)$, where $T$ is the relative period and $\sigma$ is chosen so that the automorphism $\delta\mapsto\sigma^{-1}\delta\sigma, \delta\in\Delta$, has finite order $k$. Form the semidirect product $\Delta\rtimes\mathbb{Z}_{2k}$ by adjoining to Δ an element τ of order

$2k$ such that $\tau^{-1}\delta\tau = \sigma^{-1}\delta\sigma$, for $\delta \in \Delta$. Codimension-1 bifurcations from $P$ are in one-to-one correspondence (modulo tail terms) with bifurcations from fully symmetric equilibria for a $(\Delta \rtimes \mathbb{Z}_{2k})$-equivariant vector field. In particular, period-preserving and period-doubling bifurcations from $P$ reduce to steady-state bifurcations, and Naimark–Sacker bifurcations reduce to Hopf bifurcations. This framework incorporates issues such as suppression of period doubling. Similar results hold for higher-codimension bifurcations.

The skew products [4] and [5] are valid for proper actions of certain noncompact Lie groups $\Gamma$ provided the spatial symmetries are compact, leading to explanations of spiral and scroll wave phenomena in excitable media.

When the spatial symmetry group is noncompact, $E^c$ may be infinite-dimensional and center manifold reduction may break down due to continuous-spectrum issues. For Euclidean symmetry, there is a theory of modulation or Ginzburg–Landau equations.

*See also:* Bifurcation Theory; Bifurcations in Fluid Dynamics; Bifurcations of Periodic Orbits; Central Manifolds, Normal Forms; Chaos and Attractors; Electroweak Theory; Finite Group Symmetry Breaking; Hyperbolic Dynamical Systems; Quantum Spin Systems; Quasiperiodic Systems; Singularity and Bifurcation Theory.

## Further Reading


Chossat P and Lauterbach R (2000) Methods in Equivariant Bifurcations and Dynamical Systems. *Advanced Series in Nonlinear Dynamics*, vol. 15. Singapore: World Scientific.

Crawford JD and Knobloch E (1991) Symmetry and symmetry-breaking in fluid dynamics. In: Lumley JL, Van Dyke M, and Reed HL (eds.) *Annual Review of Fluid Mechanics*, vol. 23, pp. 341–387. Palo Alto, CA: Annual Reviews.

Fiedler B and Scheel A (2003) Spatio-temporal dynamics of reaction-diffusion patterns. In: Kirkilionis M, Krömker S, Rannacher R, and Tomi F (eds.) *Trends in Nonlinear Analysis*, pp. 23–152. Berlin: Springer.

Field M (1996a) *Lectures on Bifurcations, Dynamics and Symmetry*. Pitman Research Notes in Mathematics Series, vol. 356. Harlow: Addison Wesley Longman.

Field M (1996b) *Symmetry Breaking for Compact Lie Groups*. Memoirs of the American Mathematical Society, vol. 574. Providence, RI: American Mathematical Society.

Golubitsky M and Stewart IN (2002) *The Symmetry Perspective*. Progress in Mathematics, vol. 200. Basel: Birkhäuser.

Golubitsky M, Stewart IN, and Schaeffer D (1988) *Singularities and Groups in Bifurcation Theory, Vol. II*, Applied Mathematical Sciences, vol. 69. New York: Springer.

Lamb JSW and Melbourne I (1999) Bifurcation from periodic solutions with spatiotemporal symmetry. In: Golubitsky M, Luss D, and Strogatz SH (eds.) *Pattern Formation in Continuous and Coupled Systems*, IMA Volumes in Mathematics and its Applications, vol. 115, pp. 175–191. New York: Springer.

Lamb JSW, Melbourne I, and Wulff C (2003) Bifurcation from periodic solutions with spatiotemporal symmetry, including resonances and mode interactions. *Journal Differential Equations* 191: 377–407.

Melbourne I (2000) Ginzburg–Landau theory and symmetry. In: Debnath L and Riahi DN (eds.) *Nonlinear Instability, Chaos and Turbulence, Vol 2*, Advances in Fluid Mechanics, vol. 25, pp. 79–109. Southampton: WIT Press.

Michel L (1980) Symmetry defects and broken symmetry. Configurations. Hidden symmetry. *Reviews of Modern Physics* 52: 617–651.

Sandstede B, Scheel A, and Wulff C (1999) Dynamical behavior of patterns with Euclidean symmetry. In: Golubitsky M, Luss D, and Strogatz SH (eds.) *Pattern Formation in Continuous and Coupled Systems*, IMA Volumes in Mathematics and its Applications, vol. 115, pp. 249–264. New York: Springer.


# Symmetry and Symplectic Reduction


**J-P Ortega**, Université de Franche-Comté, Besançon, France

**T S Ratiu**, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland




## Introduction

The use of symmetries in the quantitative and qualitative study of dynamical systems has a long history that goes back to the founders of mechanics. In most cases, the symmetries of a system are used to implement a procedure generically known under the name of "reduction" that restricts the study of its dynamics to a system of smaller dimension. This procedure is also used in a purely geometric context to construct new nontrivial manifolds having various additional structures.

Most of the reduction methods can be seen as constructions that systematize the techniques of elimination of variables found in classical mechanics. These procedures consist basically of two steps. First, one restricts the dynamics to flow-invariant submanifolds of the system in question and, second, one projects the restricted dynamics onto the symmetry orbit quotients of the spaces constructed in the first step. Sometimes, the

flow-invariant manifolds appear as the level sets of a momentum map induced by the symmetry of the system.

## Symmetry Reduction

### The Symmetries of a System

The standard mathematical fashion to describe the symmetries of a dynamical system (*see* Dynamical Systems in Mathematical Physics: An Illustration from Water Waves) $X \in \mathfrak{X}(M)$ defined on a manifold $M(\mathfrak{X}(M)$ denotes the Lie algebra of smooth vector fields on $M$ endowed with the Jacobi–Lie bracket $[\cdot,\cdot])$ consists in studying its invariance properties with respect to a smooth Lie group $\Phi : G \times M \to M$ (continuous symmetries) or Lie algebra $\phi : \mathfrak{g} \to \mathfrak{X}(M)$ (infinitesimal symmetry) action. Recall that $\Phi$ is a (left) action if the map $g \in G \mapsto \Phi(g,\cdot) \in \mathrm{Diff}(M)$ is a group homomorphism, where $\mathrm{Diff}(M)$ denotes the group of smooth diffeomorphisms of the manifold $M$. The map $\phi$ is a (left) Lie algebra action if the map $\xi \in \mathfrak{g} \mapsto \phi(\xi) \in \mathfrak{X}(M)$ is a Lie algebra antihomomorphism and the map $(m,\xi) \in M \times \mathfrak{g} \mapsto \phi(\xi)(m) \in TM$ is smooth. The vector field $X$ is said to be $G$-symmetric whenever it is equivariant with respect to the $G$-action $\Phi$, that is, $X \circ \Phi_g = T\Phi_g \circ X$, for any $g \in G$. The space of $G$-symmetric vector fields on $M$ is denoted by $\mathfrak{X}(M)^G$. The flow $F_t$ of a $G$-symmetric vector field $X \in \mathfrak{X}(M)^G$ is $G$-equivariant, that is, $F_t \circ \Phi_g = \Phi_g \circ F_t$, for any $g \in G$. The vector field $X$ is said to be $\mathfrak{g}$-symmetric if $[\phi(\xi), X] = 0$, for any $\xi \in \mathfrak{g}$.

If $\mathfrak{g}$ is the Lie algebra of the Lie group $G$ (*see* Lie Groups: General Theory) then the infinitesimal generators $\xi_M \in \mathfrak{X}(M)$ of a smooth $G$-group action defined by

$$\xi_M(m) := \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \Phi(\exp t\xi, m), \quad \xi \in \mathfrak{g},\ m \in M$$

constitute a smooth Lie algebra $\mathfrak{g}$-action and we denote in this case $\phi(\xi) = \xi_M$.

If $m \in M$, the closed Lie subgroup $G_m := \{g \in G \mid \Phi(g,m) = m\}$ is called the isotropy or symmetry subgroup of $m$. Similarly, the Lie subalgebra $\mathfrak{g}_m := \{\xi \in \mathfrak{g} \mid \phi(\xi)(m) = 0\}$ is called the isotropy or symmetry subalgebra of $m$. If $\mathfrak{g}$ is the Lie algebra of $G$ and the Lie algebra action is given by the infinitesimal generators, then $\mathfrak{g}_m$ is the Lie algebra of $G_m$. The action is called free if $G_m = \{e\}$ for every $m \in M$ and locally free if $\mathfrak{g}_m = \{0\}$ for every $m \in M$. We will write interchangeably $\Phi(g,m) = \Phi_g(m) = \Phi^m(g) = g \cdot m$, for $m \in M$ and $g \in G$.

In this article we will focus mainly on continuous symmetries induced by proper Lie group actions. The action $\Phi$ is called proper whenever for any two convergent sequences $\{m_n\}_{n\in\mathbb{N}}$ and $\{g_n \cdot m_n := \Phi(g_n, m_n)\}_{n\in\mathbb{N}}$ in $M$, there exists a convergent subsequence $\{g_{n_k}\}_{k\in\mathbb{N}}$ in $G$. Compact group actions are obviously proper.

### Symmetry Reduction of Vector Fields

Let $M$ be a smooth manifold and $G$ a Lie group acting properly on $M$. Let $X \in \mathfrak{X}(M)^G$ and $F_t$ be its (necessarily equivariant) flow. For any isotropy subgroup $H$ of the $G$-action on $M$, the $H$-isotropy type submanifold $M_H := \{m \in M \mid G_m = H\}$ is preserved by the flow $F_t$. This property is known as the law of conservation of isotropy. The properness of the action guarantees that $G_m$ is compact and that the (connected components of) $M_H$ are embedded submanifolds of $M$ for any closed subgroup $H$ of $G$. The manifolds $M_H$ are, in general, not closed in $M$. Moreover, the quotient group $N(H)/H$ (where $N(H)$ denotes the normalizer of $H$ in $G$) acts freely and properly on $M_H$. Hence, if $\pi_H : M_H \to M_H/(N(H)/H)$ denotes the projection onto orbit space and $i_H : M_H \hookrightarrow M$ is the injection, the vector field $X$ induces a unique vector field $X^H$ on the quotient $M_H/(N(H)/H)$ defined by $X^H \circ \pi_H = T\pi_H \circ X \circ i_H$, whose flow $F_t^H$ is given by $F_t^H \circ \pi_H = \pi_H \circ F_t \circ i_H$. We will refer to $X^H \in \mathfrak{X}(M_H/(N(H)/H))$ as the $H$-isotropy type reduced vector field induced by $X$.

This reduction technique has been widely exploited in handling specific dynamical systems. When the symmetry group $G$ is compact and we are dealing with a linear action, the construction of the quotient $M_H/(N(H)/H)$ can be implemented in a very explicit and convenient manner by using the invariant polynomials of the action and the theorems of Hilbert and Schwarz–Mather.

## Symplectic Reduction

Symplectic or Marsden–Weinstein reduction is a procedure that implements symmetry reduction for the symmetric Hamiltonian systems defined on a symplectic manifold $(M,\omega)$. The particular case in which the symplectic manifold is a cotangent bundle is dealt with separately (*see* Cotangent Bundle Reduction). We recall that the Hamiltonian vector field $X_h \in \mathfrak{X}(M)$ associated to the Hamiltonian function $h \in C^\infty(M)$ is uniquely determined by the equality $\omega(X_h,\cdot) = \mathrm{d}h$. In this context, the symmetries $\Phi : G \times M \to M$ of interest are given by symplectic or canonical transformations, that is, $\Phi_g^*\omega = \omega$, for any $g \in G$. For canonical actions each $G$-invariant function $h \in C^\infty(M)^G$ has an associated $G$-symmetric Hamiltonian vector field $X_h$. A Lie

algebra action $\varphi$ is called symplectic or canonical if $\pounds_{\phi(\xi)}\omega = 0$ for all $\xi \in \mathfrak{g}$, where $\pounds$ denotes the Lie derivative operator. If the Lie algebra action is induced from a canonical Lie group action by taking its infinitesimal generators, then it is also canonical.

#### Momentum Maps

The symmetry reduction described in the previous section for general vector fields does not produce a well-adapted answer for symplectic manifolds $(M, \omega)$ in the sense that the reduced spaces $M_H/(N(H)/H)$ are, in general, not symplectic. To solve this problem one has to use the conservation laws associated to the canonical action, which often appear as momentum maps.

Let $G$ be a Lie group acting canonically on the symplectic manifold $(M, \omega)$. Suppose that for any $\xi \in \mathfrak{g}$, the vector field $\xi_M$ is Hamiltonian, with Hamiltonian function $\boldsymbol{J}^\xi \in C^\infty(M)$ and that $\xi \in \mathfrak{g} \mapsto \boldsymbol{J}^\xi \in C^\infty(M)$ is linear. The map $\boldsymbol{J} : M \to \mathfrak{g}^*$ defined by the relation $\langle \boldsymbol{J}(z), \xi \rangle = \boldsymbol{J}^\xi(z)$, for all $\xi \in \mathfrak{g}$ and $z \in M$, is called a momentum map of the $G$-action (*see* Hamiltonian Group Actions). Momentum maps, if they exist, are determined up to a constant in $\mathfrak{g}^*$ for any connected component of $M$.

**Examples 1**

(i) (*Linear momentum*) The phase space of an $N$-particle system is the cotangent space $T^*\mathbb{R}^{3N}$ endowed with its canonical symplectic structure. The additive group $\mathbb{R}^3$, whose Lie algebra is abelian and is also equal to $\mathbb{R}^3$, acts canonically on it by spatial translation on each factor: $\boldsymbol{v} \cdot (\boldsymbol{q}_i, \boldsymbol{p}^i) = (\boldsymbol{q}_i + \boldsymbol{v}, \boldsymbol{p}^i)$, with $i = 1, \ldots, N$. This action has an associated momentum map $\boldsymbol{J} : T^*\mathbb{R}^{3N} \to \mathbb{R}^3$, where we identified the dual of $\mathbb{R}^3$ with itself using the Euclidean inner product, which coincides with the classical linear momentum $\boldsymbol{J}(\boldsymbol{q}_i, \boldsymbol{p}^i) = \sum_{i=1}^N \boldsymbol{p}_i$.

(ii) (*Angular momentum*) Let SO(3) act on $\mathbb{R}^3$ and then, by lift, on $T^*\mathbb{R}^3$, that is, $A \cdot (\boldsymbol{q}, \boldsymbol{p}) = (A\boldsymbol{q}, A\boldsymbol{p})$. This action is canonical and has as associated momentum map $\boldsymbol{J} : T^*\mathbb{R}^3 \to \mathfrak{so}(3)^* \cong \mathbb{R}^3$, the classical angular momentum $\boldsymbol{J}(\boldsymbol{q}, \boldsymbol{p}) = \boldsymbol{q} \times \boldsymbol{p}$.

(iii) (*Lifted actions on cotangent bundles*) The previous two examples are particular cases of the following situation. Let $\Phi : G \times M \to M$ be a smooth Lie group action. The (left) cotangent lifted action of $G$ on $T^*Q$ is given by $g \cdot \alpha_q := T^*_{g \cdot q}\Phi_{g^{-1}}(\alpha_q)$ for $g \in G$ and $\alpha_q \in T^*Q$. Cotangent lifted actions preserve the canonical 1-form on $T^*Q$ and hence are canonical. They admit an associated momentum map $\boldsymbol{J} : T^*Q \to \mathfrak{g}^*$ given by $\langle \boldsymbol{J}(\alpha_q), \xi \rangle = \alpha_q(\xi_Q(q))$, for any $\alpha_q \in T^*Q$ and any $\xi \in \mathfrak{g}$.

(iv) (*Symplectic linear actions*) Let $(V, \omega)$ be a symplectic linear space and let $G$ be a subgroup of the linear symplectic group, acting naturally on $V$. By the choice of $G$ this action is canonical and has a momentum map given by $\langle \boldsymbol{J}(v), \xi \rangle = (1/2)\omega(\xi_V(v), v)$, for $\xi \in \mathfrak{g}$ and $v \in V$ arbitrary.

#### Properties of the Momentum Map

The main feature of the momentum map that makes it of interest for use in reduction is that it encodes conservation laws for $G$-symmetric Hamiltonian systems. Noether's theorem states that the momentum map is a constant of the motion for the Hamiltonian vector field $X_h$ associated to any $G$-invariant function $h \in C^\infty(M)^G$ (*see* Symmetries and Conservation Laws).

The derivative $T\boldsymbol{J}$ of the momentum map satisfies the following two properties: range $(T_m\boldsymbol{J}) = (\mathfrak{g}_m)^\circ$ and $\ker T_m\boldsymbol{J} = (\mathfrak{g} \cdot m)^\omega$, for any $m \in M$, where $(\mathfrak{g}_m)^\circ$ denotes the annihilator in $\mathfrak{g}^*$ of the isotropy subalgebra $\mathfrak{g}_m$ of $m$, $\mathfrak{g} \cdot m := T_m(G \cdot m) = \{\xi_M(m) | \xi \in \mathfrak{g}\}$ is the tangent space at $m$ to the $G$-orbit that contains this point, and $(\mathfrak{g} \cdot m)^\omega$ is the symplectic orthogonal space to $\mathfrak{g} \cdot m$ in the symplectic vector space $(T_mM, \omega(m))$. The first relation is sometimes called the bifurcation lemma since it establishes a link between the symmetry of a point and the rank of the momentum map at that point.

The existence of the momentum map for a given canonical action is not guaranteed. A momentum map exists if and only if the linear map $\rho : [\xi] \in \mathfrak{g}/[\mathfrak{g}, \mathfrak{g}] \mapsto [\omega(\xi_M, \cdot)] \in H^1(M, \mathbb{R})$ is identically zero. Thus, if $H^1(M, \mathbb{R}) = 0$ or $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}] = H^1(\mathfrak{g}, \mathbb{R}) = 0$ then $\rho \equiv 0$. In particular, if $\mathfrak{g}$ is semisimple, the "first Whitehead lemma" states that $H^1(\mathfrak{g}, \mathbb{R}) = 0$ and therefore a momentum map always exists for canonical semisimple Lie algebra actions.

A natural question to ask is when the map $(\mathfrak{g}, [\cdot, \cdot]) \to (C^\infty(M), \{\cdot, \cdot\})$ defined by $\xi \mapsto \boldsymbol{J}^\xi, \xi \in \mathfrak{g}$, is a Lie algebra homomorphism, that is, $\boldsymbol{J}^{[\xi,\eta]} = \{\boldsymbol{J}^\xi, \boldsymbol{J}^\eta\}$, $\xi, \eta \in \mathfrak{g}$. Here $\{\cdot, \cdot\} : C^\infty(M) \times C^\infty(M) \to C^\infty(M)$ denotes the Poisson bracket associated to the symplectic form $\omega$ of $M$ defined by $\{f, h\} := \omega(X_f, X_h), f, h \in C^\infty(M)$. This is the case if and only if $T_z\boldsymbol{J}(\xi_M(z)) = -\mathrm{ad}^*_\xi \boldsymbol{J}(z)$, for any $\xi \in \mathfrak{g}$, $z \in M$, where $\mathrm{ad}^*$ is the dual of the adjoint representation $\mathrm{ad} : (\xi, \eta) \in \mathfrak{g} \times \mathfrak{g} \mapsto [\xi, \eta] \in \mathfrak{g}$ of $\mathfrak{g}$ on itself. A momentum map that satisfies this relation in called infinitesimally equivariant. The reason behind this terminology is that this is the infinitesimal version of global or coadjoint equivariance: $\boldsymbol{J}$ is $G$-equivariant if $\mathrm{Ad}^*_{g^{-1}} \circ \boldsymbol{J} = \boldsymbol{J} \circ \Phi_g$ or, equivalently,

$J_g^{\mathrm{Ad}}\xi(g \cdot z) = J^{\xi}(z)$, for all $g \in G$, $\xi \in \mathfrak{g}$, and $z \in M$; $\mathrm{Ad}^*$ denotes the dual of the adjoint representation Ad of $G$ on $\mathfrak{g}$. Actions admitting infinitesimally equivariant momentum maps are called Hamiltonian actions and Lie group actions with coadjoint equivariant momentum maps are called globally Hamiltonian actions. If the symmetry group $G$ is connected then global and infinitesimal equivariance of the momentum map are equivalent concepts. If $\mathfrak{g}$ acts canonically on $(M, \omega)$ and $H^1(\mathfrak{g}, \mathbb{R}) = \{0\}$ then this action admits at most one infinitesimally equivariant momentum map.

Since momentum maps are not uniquely defined, one may ask whether one can choose them to be equivariant. It turns out that if the momentum map is associated to the action of a compact Lie group, this can always be done. Momentum maps of cotangent lifted actions are also equivariant as are momentum maps defined by symplectic linear actions. Canonical actions of semisimple Lie algebras on symplectic manifolds admit infinitesimally equivariant momentum maps, since the "second Whitehead lemma" states that $H^2(\mathfrak{g}, \mathbb{R}) = 0$ if $\mathfrak{g}$ is semisimple. We shall identify below a specific element of $H^2(\mathfrak{g}, \mathbb{R})$ which is the obstruction to the equivariance of a momentum map (assuming it exists).

Even though, in general, it is not possible to choose a coadjoint equivariant momentum map, it turns out that when the symplectic manifold is connected there is an affine action on the dual of the Lie algebra with respect to which the momentum map is equivariant. Define the nonequivariance 1-cocycle associated to $J$ as the map $\sigma : G \longrightarrow \mathfrak{g}^*$ given by $g \mapsto J(\Phi_g(z)) - \mathrm{Ad}^*_{g^{-1}}(J(z))$. The connectivity of $M$ implies that the right-hand side of this equality is independent of the point $z \in M$. In addition, $\sigma$ is a (left) $\mathfrak{g}^*$-valued 1-cocycle on $G$ with respect to the coadjoint representation of $G$ on $\mathfrak{g}^*$, that is, $\sigma(gh) = \sigma(g) + \mathrm{Ad}^*_{g^{-1}}\sigma(h)$ for all $g, h \in G$. Relative to the affine action $\Theta : G \times \mathfrak{g}^* \longrightarrow \mathfrak{g}^*$ given by $(g, \mu) \longmapsto \mathrm{Ad}^*_{g^{-1}}\mu + \sigma(g)$, the momentum map $J$ is equivariant. The "reduction lemma," the main technical ingredient in the proof of the reduction theorem, states that for any $m \in M$ we have

$$\mathfrak{g}_{J(m)} \cdot m = \mathfrak{g} \cdot m \cap \ker T_m J = \mathfrak{g} \cdot m \cap (\mathfrak{g} \cdot m)^{\omega}$$

where $\mathfrak{g}_{J(m)}$ is the Lie algebra of the isotropy group $G_{J(m)}$ of $J(m) \in \mathfrak{g}^*$ with respect to the affine action of $G$ on $\mathfrak{g}^*$ induced by the nonequivariance 1-cocycle of $J$.

## The Symplectic Reduction Theorem

The symplectic reduction procedure that we now present consists of constructing a new symplectic manifold out of a given symmetric one in which the conservation laws encoded in the form of a momentum map and the degeneracies associated to the symmetry have been eliminated. This strategy allows the reduction of a symmetric Hamiltonian dynamical system to a dimensionally smaller one. This reduction procedure preserves the symplectic category, that is, if we start with a Hamiltonian system on a symplectic manifold, the reduced system is also a Hamiltonian system on a symplectic manifold. The reduced symplectic manifold is usually referred to as the symplectic or Marsden–Weinstein reduced space.

**Theorem 2** *Let $\Phi : G \times M \to M$ be a free proper canonical action of the Lie group $G$ on the connected symplectic manifold $(M, \omega)$. Suppose that this action has an associated momentum map $J : M \to \mathfrak{g}^*$, with nonequivariance 1-cocycle $\sigma : G \to \mathfrak{g}^*$. Let $\mu \in \mathfrak{g}^*$ be a value of $J$ and denote by $G_\mu$ the isotropy of $\mu$ under the affine action of $G$ on $\mathfrak{g}^*$. Then:*

(i) *The space $M_\mu := J^{-1}(\mu)/G_\mu$ is a regular quotient manifold and, moreover, it is a symplectic manifold with symplectic form $\omega_\mu$ uniquely characterized by the relation*

$$\pi_\mu^* \omega_\mu = i_\mu^* \omega$$

*The maps $i_\mu : J^{-1}(\mu) \hookrightarrow M$ and $\pi_\mu : J^{-1}(\mu) \to J^{-1}(\mu)/G_\mu$ denote the inclusion and the projection, respectively. The pair $(M_\mu, \omega_\mu)$ is called the symplectic point reduced space.*

(ii) *Let $h \in C^\infty(M)^G$ be a $G$-invariant Hamiltonian. The flow $F_t$ of the Hamiltonian vector field $X_h$ leaves the connected components of $J^{-1}(\mu)$ invariant and commutes with the $G$-action, so it induces a flow $F_t^\mu$ on $M_\mu$ defined by $\pi_\mu \circ F_t \circ i_\mu = F_t^\mu \circ \pi_\mu$.*

(iii) *The vector field generated by the flow $F_t^\mu$ on $(M_\mu, \omega_\mu)$ is Hamiltonian with associated reduced Hamiltonian function $h_\mu \in C^\infty(M_\mu)$ defined by $h_\mu \circ \pi_\mu = h \circ i_\mu$. The vector fields $X_h$ and $X_{h_\mu}$ are $\pi_\mu$-related. The triple $(M_\mu, \omega_\mu, h_\mu)$ is called the reduced Hamiltonian system.*

(iv) *Let $k \in C^\infty(M)^G$ be another $G$-invariant function. Then $\{h, k\}$ is also $G$-invariant and $\{h, k\}_\mu = \{h_\mu, k_\mu\}_{M_\mu}$, where $\{\cdot, \cdot\}_{M_\mu}$ denotes the Poisson bracket associated to the symplectic form $\omega_\mu$ on $M_\mu$.*

## Reconstruction of Dynamics

We pose now the question converse to the reduction of a Hamiltonian system. Assume that an integral curve $c_\mu(t)$ of the reduced Hamiltonian system $X_{h_\mu}$

on $(M_\mu, \omega_\mu)$ is known. Let $m_0 \in \mathbf{J}^{-1}(\mu)$ be given. One can determine from this data the integral curve of the Hamiltonian system $X_h$ with initial condition $m_0$. In other words, one can reconstruct the solution of the given system knowing the corresponding reduced solution. The general method of reconstruction is the following. Pick a smooth curve $d(t)$ in $\mathbf{J}^{-1}(\mu)$ such that $d(0)=m_0$ and $\pi_\mu(d(t))=c_\mu(t)$. Then, if $c(t)$ denotes the integral curve of $X_h$ with $c(0)=m_0$, we can write $c(t)=g(t)\cdot d(t)$ for some smooth curve $g(t)$ in $G_\mu$ that is obtained in two steps. First, one finds a smooth curve $\xi(t)$ in $\mathfrak{g}_\mu$ such that $\xi(t)_M(d(t))=X_h(d(t))-\dot{d}(t)$. With the $\xi(t)\in\mathfrak{g}_\mu$ just obtained, one solves the nonautonomous differential equation $\dot{g}(t)=T_eL_{g(t)}\xi(t)$ on $G_\mu$ with $g(0)=e$.

### The Orbit Formulation of the Symplectic Reduction Theorem

There is an alternative approach to the reduction theorem which consists of choosing as numerator of the symplectic reduced space the group invariant saturation of the level sets of the momentum map. This option produces as a result a space that is symplectomorphic to the Marsden–Weinstein quotient but presents the advantage of being more appropriate in the context of quantization problems. Additionally, this approach makes easier the comparison of the symplectic reduced spaces corresponding to different values of the momentum map which is important in the context of Poisson reduction (*see* Poisson Reduction). In carrying out this construction, one needs to use the natural symplectic structures that one can define on the orbits of the affine action of a group on the dual of its Lie algebra and that we now quickly review.

Let $G$ be a Lie group, $\sigma: G\to\mathfrak{g}^*$ a coadjoint 1-cocycle, and $\mu\in\mathfrak{g}^*$. Let $\mathcal{O}_\mu$ be the orbit through $\mu$ of the affine $G$-action on $\mathfrak{g}^*$ associated to $\sigma$. If $\Sigma:\mathfrak{g}\times\mathfrak{g}\to\mathbb{R}$ defined by

$$\sum(\xi,\eta):=\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0}\langle\sigma(\exp(t\xi),\eta\rangle$$

is a real-valued Lie algebra 2-cocycle (which is always the case if $\sigma$ is the derivative of a smooth real-valued group 2-cocycle or if $\sigma$ is the nonequivariance 1-cocycle of a momentum map), that is, $\Sigma:\mathfrak{g}\times\mathfrak{g}\to\mathbb{R}$ is skew-symmetric and $\Sigma([\xi,\eta],\zeta)+\Sigma([\eta,\zeta],\xi)+\Sigma([\zeta,\xi],\eta)=0$ for all $\xi,\eta,\zeta\in\mathfrak{g}$, then the affine orbit $\mathcal{O}_\mu$ is a symplectic manifold with $G$-invariant symplectic structure $\omega^\pm_{\mathcal{O}_\mu}$ given by

$$\omega^\pm_{\mathcal{O}_\mu}(\nu)(\xi_{\mathfrak{g}^*}(\nu),\eta_{\mathfrak{g}^*}(\nu))=\pm\langle\nu,[\xi,\eta]\rangle\mp\Sigma(\xi,\eta) \qquad [1]$$

for arbitrary $\nu\in\mathcal{O}_\mu$, and $\xi,\eta\in\mathfrak{g}$. The symbol $\xi_{\mathfrak{g}^*}(\nu):=-\mathrm{ad}^*_\xi\nu+\Sigma(\xi,\cdot)$ denotes the infinitesimal generator of the affine action on $\mathfrak{g}^*$ associated to $\xi\in\mathfrak{g}$. The symplectic structures $\omega_{\mathcal{O}}\mu^\pm$ on $\mathcal{O}_\mu$ are called the $(\pm)$-orbit or Kostant–Kirillov–Souriau (KKS) symplectic forms.

This symplectic form can be obtained from Theorem 2 by considering the symplectic reduction of the cotangent bundle $T^*G$ endowed with the magnetic symplectic structure $\overline{\omega_\Sigma}:=\omega_{\mathrm{can}}-\pi^*B_\Sigma$, where $\omega_{\mathrm{can}}$ is the canonical symplectic form on $T^*G,\pi:T^*G\to G$ is the projection onto the base, and $B_\Sigma\in\Omega^2(G)^G$ is a left-invariant 2-form on $G$ whose value at the identity is the Lie algebra 2-cocycle $\Sigma:\mathfrak{g}\times\mathfrak{g}\to\mathbb{R}$. Since $\Sigma$ is a cocycle, it follows that $B_\Sigma$ is closed and hence $\overline{\omega_\Sigma}$ is a symplectic form. Moreover, the lifting of the left translations on $G$ provides a canonical $G$-action on $T^*G$ that has a momentum map given by $\mathbf{J}(g,\mu)=\Theta(g,\mu),(g,\mu)\in G\times\mathfrak{g}^*\simeq T^*G$, where the trivialization $G\times\mathfrak{g}^*\simeq T^*G$ is obtained via left translations. Symplectic reduction using these ingredients yields symplectic reduced spaces that are naturally symplectically diffeomorphic to the affine orbits $\mathcal{O}_\mu$ with the symplectic form [1].

**Theorem 3** (Symplectic orbit reduction). *Let $\Phi:G\times M\to M$ be a free proper canonical action of the Lie group $G$ on the connected symplectic manifold $(M,\omega)$. Suppose that this action has an associated momentum map $\mathbf{J}:M\to\mathfrak{g}^*$, with nonequivariance 1-cocycle $\sigma:G\to\mathfrak{g}^*$. Let $\mathcal{O}_\mu:=G\cdot\mu\subset\mathfrak{g}^*$ be the G-orbit of the point $\mu\in\mathfrak{g}^*$ with respect to the affine action of G on $\mathfrak{g}^*$ associated to $\sigma$. Then the set $M_{\mathcal{O}_\mu}:=\mathbf{J}^{-1}(\mathcal{O}_\mu)/G$ is a regular quotient symplectic manifold with the symplectic form $\omega_{\mathcal{O}_\mu}$ uniquely characterized by the relation $i^*_{\mathcal{O}_\mu}\omega=\pi^*_{\mathcal{O}_\mu}\omega_{\mathcal{O}_\mu}+\mathbf{J}^*_{\mathcal{O}_\mu}\omega^+_{\mathcal{O}_\mu}$, where $\mathbf{J}_{\mathcal{O}_\mu}$ is the restriction of $\mathbf{J}$ to $\mathbf{J}^{-1}(\mathcal{O}_\mu)$ and $\omega^+_{\mathcal{O}_\mu}$ is the $(+)$–symplectic structure on the affine orbit $\mathcal{O}_\mu$. The maps $i_{\mathcal{O}_\mu}:\mathbf{J}^{-1}(\mathcal{O}_\mu)\hookrightarrow M$ and $\pi_{\mathcal{O}_\mu}:\mathbf{J}^{-1}(\mathcal{O}_\mu)\to M_{\mathcal{O}_\mu}$ are natural injection and the projection, respectively. The pair $(M_{\mathcal{O}_\mu},\omega_{\mathcal{O}_\mu})$ is called the symplectic orbit reduced space. Statements similar to (ii)–(iv) in Theorem 2 can be formulated for the orbit reduced spaces $(M_{\mathcal{O}_\mu},\omega_{\mathcal{O}_\mu})$.*

We emphasize that given a momentum value $\mu\in\mathfrak{g}^*$, the reduced spaces $M_\mu$ and $M_{\mathcal{O}_\mu}$ are symplectically diffeomorphic via the projection to the quotients of the inclusion $\mathbf{J}^{-1}(\mu)\hookrightarrow\mathbf{J}^{-1}(\mathcal{O}_\mu)$.

Reduction at a general point can be replaced by reduction at zero at the expense of enlarging the manifold by the affine orbit. Consider the canonical diagonal action of $G$ on the symplectic difference $M\ominus\mathcal{O}^+_\mu$, which is the manifold $M\times\mathcal{O}_\mu$ with the symplectic form $\pi_1^*\omega-\pi_2^*\omega_{\mathcal{O}^+_\mu}$, where $\pi_1:M\times\mathcal{O}_\mu\to M$ and $\pi_2:M\times\mathcal{O}_\mu\to\mathcal{O}_\mu$ are the projections.

A momentum map for this action is given by $\boldsymbol{J} \circ \pi_1 - \pi_2 : M \ominus \mathcal{O}_\mu^+ \to \mathfrak{g}^*$. Let $(M \ominus \mathcal{O}_\mu^+)_0 := ((\boldsymbol{J} \circ \pi_1 - \pi_2)^{-1}(0)/G, (\omega \ominus \omega_{\mathcal{O}_\mu}^+)_0)$ be the symplectic point reduced space at zero.

**Theorem 4** (Shifting theorem). *Under the hypotheses of the symplectic orbit reduction theorem (Theorem 3), the symplectic orbit reduced space $M_{\mathcal{O}_\mu}$, the point reduced spaces $M_\mu$, and $(M \ominus \mathcal{O}_\mu^+)_0$ are symplectically diffeomorphic.*

## Singular Reduction

In the previous section we carried out symplectic reduction for free and proper actions. The freeness guarantees via the bifurcation lemma that the momentum map $\boldsymbol{J}$ is a submersion and hence the level sets $\boldsymbol{J}^{-1}(\mu)$ are smooth manifolds. Freeness and properness ensure that the orbit spaces $M_\mu := \boldsymbol{J}^{-1}(\mu)/G_\mu$ are regular quotient manifolds. The theory of singular reduction studies the properties of the orbit space $M_\mu$ when the hypothesis on the freeness of the action is dropped. The main result in this situation shows that these quotients are symplectic Whitney stratified spaces, in the sense that the strata are symplectic manifolds in a very natural way; moreover, the local properties of this Whitney stratification make it into what is called a cone space. This statement is referred to as the "symplectic stratification theorem" and adapts to the symplectic symmetric context the stratification theorem of the orbit space of a proper Lie group action by using its orbit type manifolds. In order to present this result, we review the necessary definitions and results on stratified spaces (*see* Singularity and Bifurcation Theory for more information on singularity theory).

### Stratified Spaces

Let $\mathcal{Z}$ be a locally finite partition of the topological space $P$ into smooth manifolds $S_i \subset P, i \in I$. We assume that the manifolds $S_i \subset P, i \in I$, with their manifold topology are locally closed topological subspaces of $P$. The pair $(P, \mathcal{Z})$ is a decomposition of $P$ with pieces in $\mathcal{Z}$ when the following condition is satisfied:

**Condition (DS)** If $R, S \in \mathcal{Z}$ are such that $R \cap \bar{S} \neq \emptyset$, then $R \subset \bar{S}$. In this case we write $R \preceq S$. If, in addition, $R \neq S$ we say that $R$ is incident to $S$ or that it is a boundary piece of $S$ and write $R \prec S$.

The above condition is called the frontier condition and the pair $(P, \mathcal{Z})$ is called a decomposed space. The dimension of $P$ is defined as $\dim P = \sup\{\dim S_i \mid S_i \in \mathcal{Z}\}$. If $k \in \mathbb{N}$, the $k$-skeleton $P^k$ of $P$ is the union of all the pieces of dimension smaller than or equal to $k$; its topology is the relative topology induced by $P$. The depth $\mathrm{dp}(z)$ of any $z \in (P, \mathcal{Z})$ is defined as

$$\mathrm{dp}(z) := \sup\{k \in \mathbb{N} \mid \exists S_0, S_1, \ldots, S_k \in \mathcal{Z} \text{ with } z \in S_0 \prec S_1 \prec \cdots \prec S_k\}$$

Since for any two elements $x, y \in \mathcal{S}$ in the same piece $\mathcal{S} \in P$ we have $\mathrm{dp}(x) = \mathrm{dp}(y)$, the depth $\mathrm{dp}(S)$ of the piece $S$ is well defined by $\mathrm{dp}(S) := \mathrm{dp}(x), x \in \mathcal{S}$. Finally, the depth $\mathrm{dp}(P)$ of $(P, \mathcal{Z})$ is defined by $\mathrm{dp}(P) := \sup\{\mathrm{dp}(S) \mid S \in \mathcal{Z}\}$.

A continuous mapping $f : P \to Q$ between the decomposed spaces $(P, \mathcal{Z})$ and $(Q, \mathcal{Y})$ is a morphism of decomposed spaces if, for every piece $S \in \mathcal{Z}$, there is a piece $T \in \mathcal{Y}$ such that $f(S) \subset T$ and the restriction $f|_S : S \to T$ is smooth. If $(P, \mathcal{Z})$ and $(P, \mathcal{T})$ are two decompositions of the same topological space we say that $\mathcal{Z}$ is coarser than $\mathcal{T}$ or that $\mathcal{T}$ is finer than $\mathcal{Z}$ if the identity mapping $(P, \mathcal{T}) \to (P, \mathcal{Z})$ is a morphism of decomposed spaces. A topological subspace $Q \subset P$ is a decomposed subspace of $(P, \mathcal{Z})$ if, for all pieces $S \in \mathcal{Z}$, the intersection $S \cap Q$ is a submanifold of $S$ and the corresponding partition $\mathcal{Z} \cap Q$ forms a decomposition of $Q$.

Let $P$ be a topological space and $z \in P$. Two subsets $A$ and $B$ of $P$ are said to be equivalent at $z$ if there is an open neighborhood $U$ of $z$ such that $A \cap U = B \cap U$. This relation constitutes an equivalence relation on the power set of $P$. The class of all sets equivalent to a given subset $A$ at $z$ will be denoted by $[A]_z$ and called the set germ of $A$ at $z$. If $A \subset B \subset P$, we say that $[A]_z$ is a subgerm of $[B]_z$, and denote $[A]_z \subset [B]_z$.

A stratification of the topological space $P$ is a map $\mathcal{S}$ that associates to any $z \in P$ the set germ $\mathcal{S}(z)$ of a closed subset of $P$ such that the following condition is satisfied:

**Condition (ST)** For every $z \in P$ there is a neighborhood $U$ of $z$ and a decomposition $\mathcal{Z}$ of $U$ such that for all $y \in U$ the germ $\mathcal{S}(y)$ coincides with the set germ of the piece of $\mathcal{Z}$ that contains $y$.

The pair $(P, \mathcal{S})$ is called a stratified space. Any decomposition of $P$ defines a stratification of $P$ by associating to each of its points the set germ of the piece in which it is contained. The converse is, by definition, locally true.

### The Strata

Two decompositions $\mathcal{Z}_1$ and $\mathcal{Z}_2$ of $P$ are said to be equivalent if they induce the same stratification of $P$. If $\mathcal{Z}_1$ and $\mathcal{Z}_2$ are equivalent decompositions of $P$ then, for all $z \in P$, we have that $\mathrm{dp}_{\mathcal{Z}_1}(z) = \mathrm{dp}_{\mathcal{Z}_2}(z)$. Any stratified space $(P, \mathcal{S})$ has a unique decomposition $\mathcal{Z}_\mathcal{S}$ associated with the following maximality property: for any open subset $U \subset P$ and any

decomposition $\mathcal{Z}$ of $P$ inducing $\mathcal{S}$ over $U$, the restriction of $\mathcal{Z}_{\mathcal{S}}$ to $U$ is coarser than the restriction of $\mathcal{Z}$ to $U$. The decomposition $\mathcal{Z}_{\mathcal{S}}$ is called the canonical decomposition associated to the stratification $(P, \mathcal{S})$. It is often denoted by $\mathcal{S}$ and its pieces are called the strata of $P$. The local finiteness of the decomposition $\mathcal{Z}_{\mathcal{S}}$ implies that for any stratum $S$ of $(P, \mathcal{S})$ there are only finitely many strata $R$ with $S \prec R$. Henceforth, the symbol $\mathcal{S}$ in the stratification $(P, \mathcal{S})$ will denote both the map that associates to each point a set germ and the set of pieces associated to the canonical decomposition induced by the stratification of $P$.

### Stratified Spaces with Smooth Structure

Let $(P, \mathcal{S})$ be a stratified space. A singular or stratified chart of $P$ is a homeomorphism $\phi : U \to \phi(U) \subset \mathbb{R}^n$ from an open set $U \subset P$ to a subset of $\mathbb{R}^n$ such that for every stratum $S \in \mathcal{S}$ the image $\phi(U \cap S)$ is a submanifold of $\mathbb{R}^n$ and the restriction $\phi|_{U\cap S} : U \cap S \to \phi(U \cap S)$ is a diffeomorphism. Two singular charts $\phi : U \to \phi(U) \subset \mathbb{R}^n$ and $\varphi : V \to \varphi(V) \subset \mathbb{R}^m$ are compatible if for any $z \in U \cap V$ there exist an open neighborhood $W \subset U \cap V$ of $z$, a natural number $N \geq \max\{n, m\}$, open neighborhoods $O, O' \subset \mathbb{R}^N$ of $\phi(U) \times \{0\}$ and $\varphi(V) \times \{0\}$, respectively, and a diffeomorphism $\psi : O \to O'$ such that $i_m \circ \varphi|_W = \psi \circ i_n \circ \phi|_W$, where $i_n$ and $i_m$ denote the natural embeddings of $\mathbb{R}^n$ and $\mathbb{R}^m$ into $\mathbb{R}^N$ by using the first $n$ and $m$ coordinates, respectively. The notion of singular or stratified atlas is the natural generalization for stratifications of the concept of atlas existing for smooth manifolds. Analogously, we can talk of compatible and maximal stratified atlases. If the stratified space $(P, \mathcal{S})$ has a well-defined maximal atlas, then we say that this atlas determines a smooth or differentiable structure on $P$. We will refer to $(P, \mathcal{S})$ as a smooth stratified space.

### The Whitney Conditions

Let $M$ be a manifold and $R, S \subset M$ two submanifolds. We say that the pair $(R, S)$ satisfies the Whitney condition (A) at the point $z \in R$ if the following condition is satisfied:

**Condition (A)** For any sequence of points $\{z_n\}_{n\in\mathbb{N}}$ in $S$ converging to $z \in R$ for which the sequence of tangent spaces $\{T_{z_n}S\}_{n\in\mathbb{N}}$ converges in the Grassmann bundle of $\dim S$–dimensional subspaces of $TM$ to $\tau \subset T_zM$, we have that $T_zR \subset \tau$.

Let $\phi : U \to \mathbb{R}^n$ be a smooth chart of $M$ around the point $z$. The Whitney condition (B) at the point $z \in R$ with respect to the chart $(U, \phi)$ is given by the following statement:

**Condition (B)** Let $\{x_n\}_{n\in\mathbb{N}} \subset R \cap U$ and $\{y_n\}_{n\in\mathbb{N}} \subset S \cap U$ be two sequences with the same limit

$$z = \lim_{n\to\infty} x_n = \lim_{n\to\infty} y_n$$

and such that $x_n \neq y_n$, for all $n \in \mathbb{N}$. Suppose that the set of connecting lines $\overline{\phi(x_n)\phi(y_n)} \subset \mathbb{R}^n$ converges in projective space to a line $L$ and that the sequence of tangent spaces $\{T_{y_n}S\}_{n\in\mathbb{N}}$ converges in the Grassmann bundle of $(\dim S)$-dimensional subspaces of $TM$ to $\tau \subset T_zM$. Then, $(T_z\phi)^{-1}(L) \subset \tau$.

If the condition (A) (respectively (B)) is verified for every point $z \in R$, the pair $(R, S)$ is said to satisfy the Whitney condition (A) (respectively (B)). It can be verified that Whitney's condition (B) does not depend on the chart used to formulate it. A stratified space with smooth structure such that, for every pair of strata, Whitney's condition (B) is satisfied is called a Whitney space.

### Cone Spaces and Local Triviality

Let $P$ be a topological space. Consider the equivalence relation $\sim$ in the product $P \times [0, \infty)$ given by $(z, a) \sim (z', a')$ if and only if $a = a' = 0$. We define the cone $CP$ on $P$ as the quotient topological space $P \times [0, \infty)/\sim$. If $P$ is a smooth manifold then the cone $CP$ is a decomposed space with two pieces, namely, $P \times (0, \infty)$ and the vertex which is the class corresponding to any element of the form $(z, 0)$, $z \in P$, that is, $P \times \{0\}$. Analogously, if $(P, \mathcal{Z})$ is a decomposed (stratified) space then the associated cone $CP$ is also a decomposed (stratified) space whose pieces (strata) are the vertex and the sets of the form $S \times (0, \infty)$, with $S \in \mathcal{Z}$. This implies, in particular, that $\dim CP = \dim P + 1$ and $\mathrm{dp}(CP) = \mathrm{dp}(P) + 1$.

A stratified space $(P, \mathcal{S})$ is said to be locally trivial if for any $z \in P$ there exist a neighborhood $U$ of $z$, a stratified space $(F, \mathcal{S}^F)$, a distinguished point $\mathbf{0} \in F$, and an isomorphism of stratified spaces

$$\psi : U \to (S \cap U) \times F$$

where $S$ is the stratum that contains $z$ and $\psi$ satisfies $\psi^{-1}(y, 0) = y$, for all $y \in S \cap U$. When $F$ is given by a cone $CL$ over a compact stratified space $L$ then $L$ is called the link of $z$.

An important corollary of "Thom's first isotopy lemma" guarantees that every Whitney stratified space is locally trivial. A converse to this implication needs the introduction of cone spaces. Their definition is given by recursion on the depth of the space.

**Definition 5** Let $m \in \mathbb{N} \cup \{\infty, \omega\}$. A cone space of class $C^m$ and depth 0 is the union of countably many $C^m$ manifolds together with the stratification whose strata are the unions of the connected components of equal dimension. A cone space of class $C^m$ and depth $d+1, d \in \mathbb{N}$, is a stratified space $(P, \mathcal{S})$ with a $C^m$ differentiable structure such that for any $z \in P$ there exists a connected neighborhood $U$ of $z$, a compact cone space $L$ of class $C^m$ and depth $d$ called the link, and a stratified isomorphism

$$\psi : U \to (S \cap U) \times CL$$

where $S$ is the stratum that contains the point $z$, the map $\psi$ satisfies $\psi^{-1}(y, 0) = y$, for all $y \in S \cap U$, and $0$ is the vertex of the cone $CL$.

If $m \neq 0$ then $L$ is required to be embedded into a sphere via a fixed smooth global singular chart $\varphi : L \to S^l$ that determines the smooth structure of $CL$. More specifically, the smooth structure of $CL$ is generated by the global chart $\tau : [z, t] \in CL \longmapsto t\varphi(z) \in \mathbb{R}^{l+1}$. The maps $\psi : U \to (S \cap U) \times CL$ and $\varphi : L \to S^l$ are referred to as a cone chart and a link chart, respectively. Moreover, if $m \neq 0$ then $\psi$ and $\psi^{-1}$ are required to be differentiable of class $C^m$ as maps between stratified spaces with a smooth structure.

### The Symplectic Stratification Theorem

Let $(M, \omega)$ be a connected symplectic manifold acted canonically and properly upon by a Lie group $G$. Suppose that this action has an associated momentum map $\mathbf{J} : M \to \mathfrak{g}^*$ with nonequivariance 1–cocycle $\sigma : G \to \mathfrak{g}^*$. Let $\mu \in \mathfrak{g}^*$ be a value of $\mathbf{J}, G_\mu$ the isotropy subgroup of $\mu$ with respect to the affine action $\Theta : G \times \mathfrak{g}^* \to \mathfrak{g}^*$ determined by $\sigma$, and let $H \subset G$ be an isotropy subgroup of the $G$-action on $M$. Let $M_H^z$ be the connected component of the $H$-isotropy type manifold that contains a given element $z \in M$ such that $\mathbf{J}(z) = \mu$ and let $G_\mu M_H^z$ be its $G_\mu$-saturation. Then the following hold:

1. The set $\mathbf{J}^{-1}(\mu) \cap G_\mu M_H^z$ is a submanifold of $M$.
2. The set $M_\mu^{(H)} := [\mathbf{J}^{-1}(\mu) \cap G_\mu M_H^z]/G_\mu$ has a unique quotient differentiable structure such that the canonical projection $\pi_\mu^{(H)} : \mathbf{J}^{-1}(\mu) \cap G_\mu M_H^z \longrightarrow M_\mu^{(H)}$ is a surjective submersion.
3. There is a unique symplectic structure $\omega_\mu^{(H)}$ on $M_\mu^{(H)}$ characterized by
$$i_\mu^{(H)*}\omega = \pi_\mu^{(H)*}\omega_\mu^{(H)}$$
where $i_\mu^{(H)} : \mathbf{J}^{-1}(\mu) \cap G_\mu M_H^z \hookrightarrow M$ is the natural inclusion. The pairs $(M_\mu^{(H)}, \omega_\mu^{(H)})$ will be called singular symplectic point strata.
4. Let $h \in C^\infty(M)^G$ be a $G$-invariant Hamiltonian. Then the flow $F_t$ of $X_h$ leaves the connected components of $\mathbf{J}^{-1}(\mu) \cap G_\mu M_H^z$ invariant and commutes with the $G_\mu$-action, so it induces a flow $F_t^\mu$ on $M_\mu^{(H)}$ that is characterized by $\pi_\mu^{(H)} \circ F_t \circ i_\mu^{(H)} = F_t^\mu \circ \pi_\mu^{(H)}$.
5. The flow $F_t^\mu$ is Hamiltonian on $M_\mu^{(H)}$, with reduced Hamiltonian function $h_\mu^{(H)} : M_\mu^{(H)} \to \mathbb{R}$ defined by $h_\mu^{(H)} \circ \pi_\mu^{(H)} = h \circ i_\mu^{(H)}$. The vector fields $X_h$ and $X_{h_\mu^{(H)}}$ are $\pi_\mu^{(H)}$-related.
6. Let $k : M \to \mathbb{R}$ be another $G$-invariant function. Then $\{h, k\}$ is also $G$-invariant and $\{h, k\}_\mu^{(H)} = \{h_\mu^{(H)}, k_\mu^{(H)}\}_{M_\mu^{(H)}}$, where $\{\,,\}_{M_\mu^{(H)}}$ denotes the Poisson bracket induced by the symplectic structure on $M_\mu^{(H)}$.

**Theorem 6** (Symplectic stratification theorem). *The quotient $M_\mu := \mathbf{J}^{-1}(\mu)/G_\mu$ is a cone space when considered as a stratified space with strata $M_\mu^{(H)}$.*

As was the case for regular reduction, this theorem can also be formulated from the orbit reduction point of view. Using that approach one can conclude that the orbit reduced spaces $M_{\mathcal{O}_\mu}$ are cone spaces symplectically stratified by the manifolds $M_{\mathcal{O}_\mu}^{(H)} := G \cdot (\mathbf{J}^{-1}(\mu) \cap M_H^z)/G$ that have symplectic structure uniquely determined by the expression

$$i_{\mathcal{O}_\mu}^{(H)*}\omega = \pi_{\mathcal{O}_\mu}^{(H)*}\omega_{\mathcal{O}_\mu}^{(H)} + \mathbf{J}_{\mathcal{O}_\mu}^{(H)*}\omega_{\mathcal{O}_\mu}^{+}$$

where $i_{\mathcal{O}_\mu}^{(H)} : G \cdot (\mathbf{J}^{-1}(\mu) \cap M_H^z) \hookrightarrow M$ is the inclusion, $\mathbf{J}_{\mathcal{O}_\mu}^{(H)} : G \cdot (\mathbf{J}^{-1}(\mu) \cap M_H^z) \to \mathcal{O}_\mu$ is obtained by restriction of the momentum map $\mathbf{J}$, and $\omega_{\mathcal{O}_\mu}^{+}$ is the $(+)$–symplectic form on $\mathcal{O}_\mu$. Analogous statements to (7)–(6) above with obvious modifications are valid.

*See also:* Cotangent Bundle Reduction; Dynamical Systems in Mathematical Physics: An Illustration from Water Waves; Graded Poisson Algebras; Hamiltonian Group Actions; Lie Groups: General Theory; Poisson Reduction; Singularity and Bifurcation Theory; Symmetries and Conservation Laws.

## Further Reading


Abraham R and Marsden JE (1978) *Foundations of Mechanics*, 2nd edn. Reading, MA: Addison-Wesley.

Arms JM, Cushman R, and Gotay MJ (1991) A universal reduction procedure for Hamiltonian group actions. In: Ratiu TS (ed.) *The Geometry of Hamiltonian Systems*, pp. 33–51. New York: Springer.

Cendra H, Marsden JE, and Ratiu TS (2001) Lagrangian reduction by stages. *Memoirs of the American Mathematical Society*, Volume 152, No. 722.

Huebschmann J (2001) Singularities and Poisson geometry of certain representation spaces. In: Landsman NP, Pflaum M, and Schlichenmaier M (eds.) *Quantization of Singular Symplectic Quotients*, Progr. Math., vol. 198, pp. 119–135. Boston, MA: Birkhäuser.

Kazhdan D, Kostant B, and Sternberg S (1978) Hamiltonian group actions and dynamical systems of Calogero type. *Communications in Pure and Applied Mathematics* 31: 481–508.

Kirillov AA (1976) *Elements of the Theory of Representations*, Grundlehren Math. Wiss. Berlin–New York: Springer.

Kostant B (1970) *Quantization and Unitary Representations*, Lecture Notes in Math., vol. 570, pp. 177–306. Berlin: Springer.

Lerman E, Montgomery R, and Sjamaar R (1993) Examples of singular reduction. In: *Symplectic Geometry*, London Math. Soc. Lecture Note Ser., vol. 192, pp. 127–155. Cambridge: Cambridge University Press.

Marsden J (1981) *Lectures on Geometric Methods in Mathematical Physics*. Philadelphia: SIAM.

Marsden JE, Misiołek G, Ortega J-P, Perlmutter M, and Ratiu TS (2005) *Hamiltonian Reduction by Stages*, Lecture Notes in Mathematics. Berlin: Springer.

Marsden JE and Ratiu TS (2003) *Introduction to Mechanics and Symmetry*, 2nd edn., second printing (1st edn (1994)), Texts in Applied Mathematics, vol. 17. New York: Springer.

Marsden JE and Weinstein A (1974) Reduction of symplectic manifolds with symmetry. *Reports on Mathematical Physics* 5(1): 121–130.

Marsden JE and Weinstein A (2001) Comments on the history, theory, and applications of symplectic reduction. In: Landsman N, Pflaum M, and Schlichenmaier M (eds.) *Quantization of Singular Sympectic Quotients*, Progress in Mathematics, vol. 198, pp. 1–20. Boston: Birkhäuser.

Mayer KR (1973) Symmetries and integrals in mechanics. In: Peixoto M (ed.) *Dynamical Systems*, pp. 259–273. New York: Academic Press.

Ortega J-P and Ratiu TS (2003) *Momentum Maps and Hamiltonian Reduction*, Progress in Math., vol. 222. Boston: Birkhäuser.

Pflaum MJ (2001) *Analytic and Geometric Study of Stratified Spaces*, Lecture Notes in Mathematics, vol. 1768. Berlin: Springer–Verlag.

Sjamaar R and Lerman E (1991) Stratified symplectic spaces and reduction. *Annals of Mathematics* 134: 375–422.

Souriau J-M (1970) *Structure des Systèmes Dynamiques*. Paris: Dunod. (English translation by Cushman RH and Tuynman GM (1997)) Structure of Dynamical Systems, Progress in Math., vol. 149. Boston: Birkhäuser.

# Symmetry Breaking in Field Theory

**T W B Kibble**, Imperial College, London, UK



## Introduction

Spontaneous symmetry breaking in its simplest form occurs when there is a symmetry of a dynamical system that is not manifest in its ground state or equilibrium state. It is a common feature of many classical and quantum systems. In quantum field theories, in the infinite-volume limit, there are new features, the appearance of unitarily inequivalent representations of the canonical commutation relations, and the possibility of a true phase transition – a point in the phase space where the thermodynamic free energy is nonanalytic. The spontaneous breaking of a continuous global symmetry implies the existence of massless particles, the Goldstone bosons, while in the local-symmetry case some or all of these may be eliminated by the Higgs mechanism. Spontaneous symmetry breaking in gauge theories is however a more elusive concept.

## Breaking of Global Symmetries

In a quantum-mechanical system a (time-independent) symmetry is represented by a unitary operator $\hat{U}$ acting on the Hilbert space of quantum states which commutes with the Hamiltonian $\hat{H}$. If the ground state $|0\rangle$ of the system in not invariant under $\hat{U}$, then $|0'\rangle = \hat{U}|0\rangle \neq c|0\rangle$ is also a ground state. In other words, the ground state is degenerate.

For a system with a finite number of degrees of freedom, whose states are represented by vectors in a separable Hilbert space $\mathcal{H}$, symmetry breaking of an abelian symmetry group $G$ is impossible, unless there are additional accidental symmetries. Consider, for example, a particle in a double-well potential

$$V = \frac{m\omega^2}{4a^2}(x^2 - a^2)^2 \qquad [1]$$

which has the discrete symmetry group $G = \mathbb{Z}_2$; the inversion symmetry operator $\hat{U}$ satisfies $\hat{U}^2 = \hat{1}$. There are then two approximate ground states $|0\rangle$ and $|0'\rangle = \hat{U}|0\rangle$, with wave functions proportional to $\exp[-(1/2)\, m\omega(x \mp a)^2]$. However, there is an overlap between these, and the off-diagonal matrix element $\langle 0|\hat{H}|0'\rangle$ is nonzero, although exponentially small, so the true energy eigenstates are, approximately, $|0_\pm\rangle = (1/\sqrt{2})(|0\rangle \pm |0'\rangle)$. (More accurate energy eigenfunctions and eigenvalues may be found by using the WKB approximation.)

Of course, if the symmetry group is nonabelian, and the ground state belongs to a nontrivial representation, then degeneracy is unavoidable. For example, if $G$ is the rotation group SO(3) (or SU(2))

and the ground state has angular momentum $j \neq 0$, then it is $(2j+1)$-fold degenerate.

The situation is different, however, in a quantum field theory. In the infinite-volume limit, even abelian symmetries can be spontaneously broken. Take, for example, a real scalar field with Lagrangian

$$\mathcal{L} = \tfrac{1}{2}\partial_\mu\phi\partial^\mu\phi - V = \tfrac{1}{2}\dot{\phi}^2 - \tfrac{1}{2}(\nabla\phi)^2 - V \qquad [2]$$

(where we set $c = \hbar = 1$), again with a double-well potential

$$V = \tfrac{1}{8}\lambda(\phi^2 - \eta^2)^2 \qquad [3]$$

exhibiting a $\mathbb{Z}_2$ symmetry under which $\phi(x) \mapsto -\phi(x)$.

At least in the semiclassical or tree approximation, there are two degenerate vacuum states $|0\rangle$ and $|0'\rangle$, with

$$\langle 0|\hat{\phi}(x)|0\rangle \approx \eta \quad \text{and} \quad \langle 0'|\hat{\phi}(x)|0'\rangle \approx -\eta \qquad [4]$$

If we quantize the system in a box of finite volume $\mathcal{V}$, then, as earlier, there is an off-diagonal matrix element of the Hamiltonian connecting the two states, so the true ground state is (approximately) $(1/\sqrt{2})(|0\rangle + |0'\rangle)$. However, this matrix element goes to zero exponentially as $\mathcal{V} \to 0$. Even for large but finite volume, the rate of transitions from $|0\rangle$ to $|0'\rangle$ is exponentially slow.

Similarly, we can consider a complex scalar field theory with a sombrero potential:

$$\begin{aligned} \mathcal{L} &= |\dot{\phi}|^2 - |\nabla\phi|^2 - V \\ V &= \tfrac{1}{2}\lambda\left(|\phi|^2 - \tfrac{1}{2}\eta^2\right)^2 \end{aligned} \qquad [5]$$

This model is invariant under the U(1) group of phase transformations, $\phi(x) \mapsto \phi(x)e^{i\alpha}$, so we now have a continuously infinite set of degenerate vacuum states $|0_\alpha\rangle$ labeled by an angle $\alpha$, and satisfying

$$\langle 0_\alpha|\hat{\phi}(x)|0_\alpha\rangle \approx \frac{1}{\sqrt{2}}\eta e^{i\alpha} \qquad [6]$$

Once again, one finds that in the infinite-volume limit there are no matrix elements connecting the different vacuum states. Moreover, in this limit no polynomial formed from the field operators $\hat{\phi}(x)$ in a finite volume can have nonzero matrix elements between $|0_\alpha\rangle$ and $|0_\beta\rangle$ for $\alpha \neq \beta$. Applying the operators $\hat{\phi}(x)$ to any one of these vacuum states $|0_\alpha\rangle$, we can construct a Fock space $\mathcal{H}_\alpha$, and the representations of the canonical commutation relations on these separate Hilbert spaces are unitarily inequivalent. Formally, we can introduce operators $\hat{U}_\alpha$ that perform the symmetry transformations:

$$\hat{U}_\alpha\hat{\phi}(x)\hat{U}_\alpha^{-1} = \hat{\phi}(x)e^{i\alpha} \qquad [7]$$

However, these are not unitary operators on the spaces $\mathcal{H}_\beta$, but rather maps from one space to another: $\hat{U}_\alpha : \mathcal{H}_\beta \to \mathcal{H}_{\alpha+\beta}$ – or, alternatively, operators on the nonseparable Hilbert space $\mathcal{H} = \bigoplus_\alpha \mathcal{H}_\alpha$.

So far, our discussion has been restricted to the tree approximation. For a full quantum treatment, $V(\phi)$ must be replaced by the effective potential $V_{\text{eff}}(\phi)$, which may be defined as the minimum value of the mean energy density in all states in which the field $\hat{\phi}$ has the uniform expectation value $\langle\hat{\phi}(x)\rangle = \phi$. $V_{\text{eff}}$ may be computed by summing vacuum loop diagrams.

A point to note is that although the degenerate vacua $|0_\alpha\rangle$ are mathematically distinct, in the absence of any external definition of phase, they are physically identical. There is no internal observational test that will distinguish them.

## Symmetry-Breaking Phase Transitions

Spontaneous symmetry breaking often occurs in the context of a phase transition. At high temperature, $T \gg \eta$, there are large fluctuations in $\phi$ and the central hump of the potential is unimportant. Then the equilibrium state is symmetric, with $\langle\hat{\phi}\rangle = 0$. However, as the temperature falls, it becomes less probable that the field will fluctuate over the top of the hump. It will tend to fall into the trough, and acquire a nonzero average value $\langle\hat{\phi}\rangle$ – the order parameter for the phase transition – thus breaking the symmetry. The direction of symmetry breaking (e.g., the phase of $\phi$ in the U(1) model) is random, determined in practice by small preexisting fluctuations or interactions with the environment.

One way of studying this process is to compute the temperature-dependent effective potential $V_{\text{eff}}(\phi, T)$. In the one-loop approximation, at high temperature, the leading corrections to the zero-temperature effective potential $V_{\text{eff}}(\phi, T)$ are of the form

$$\begin{aligned} V_{\text{eff}}(\phi, T) = V_{\text{eff}}(\phi, 0) &- \frac{\pi^2}{90}N_* T^4 \\ &+ \frac{1}{24}M_*^2(\phi)T^2 + \mathcal{O}(T) \end{aligned} \qquad [8]$$

where $N_*$ is the total number of helicity states of light particles (those with masses $\ll T$), and $M_*^2$, which depends on $\phi$, is the sum of their squared masses. (Fermions if present contribute to $N_*$ with a factor of 7/8 and to $M_*^2$ with a factor of 1/2.) In the simplest case, where we have only a multiplet $\boldsymbol{\phi} = (\phi_a)_{a=1,\ldots,N}$ of real scalar fields, $N_* = N$ and $M_*^2 = M_{aa}^2$

(summation over $a$ implied), where the mass-squared matrix is

$$M_{ab}^2 = \frac{\partial^2 V}{\partial\phi_a \partial\phi_b} \quad [9]$$

For example, in an O($N$) theory, with $V = (1/8)\lambda(\boldsymbol{\phi}^2 - \eta^2)^2$, where $\boldsymbol{\phi}^2 = \phi_a\phi_a$, one has

$$M_{ab}^2 = \tfrac{1}{2}\lambda(\boldsymbol{\phi}^2 - \eta^2)\delta_{ab} + \lambda\phi_a\phi_b \quad [10]$$

whence

$$V_{\rm eff}(\boldsymbol{\phi}, T) \approx \frac{1}{8}\lambda(\boldsymbol{\phi}^2 - \eta^2)^2 - \frac{\pi^2}{90}NT^4 + \frac{1}{48}\lambda T^2[(N+2)\boldsymbol{\phi}^2 - N\eta^2] \quad [11]$$

It is then easy to see that the minimum occurs at $\boldsymbol{\phi} = 0$ for $T > T_{\rm c}$, where in this approximation ${T_{\rm c}}^2 = 12\eta^2/(N+2)$, while below the critical temperature the minimum is at

$$\boldsymbol{\phi}^2 = \phi_{\rm eq}^2(T) \approx \eta^2 - \frac{N+2}{12}T^2 \quad [12]$$

As $T \to 0$, the equilibrium state approaches one of the vacuum states $|0_{\boldsymbol{n}}\rangle$, labeled by an $N$-dimensional unit vector $\boldsymbol{n}$, such that $\langle 0_{\boldsymbol{n}}|\hat{\boldsymbol{\phi}}|0_{\boldsymbol{n}}\rangle = \eta\boldsymbol{n}$.

It is often convenient to introduce a classical symmetry-breaking potential. For example, in the O($N$) model, we may take $V_{\rm sb} = -\boldsymbol{j}\cdot\boldsymbol{\phi}(x)$, where $\boldsymbol{j}$ is a constant $N$-vector. This has the effect of tilting the potential, thus removing the degeneracy. A characteristic of spontaneous symmetry breaking is that the limits $\boldsymbol{j} \to \boldsymbol{0}$ and $\mathcal{V} \to \infty$ do not commute. If (for $T < T_{\rm c}$) we take the infinite-volume limit first, and then let $\boldsymbol{j} \to 0$, we get different equilibrium states, depending on the direction from which $\boldsymbol{j}$ approaches zero; if we fix $\boldsymbol{n}$ and let $\boldsymbol{j} = j\boldsymbol{n}$, $j \to 0$, then we find

$$\lim_{j\to 0}\lim_{\mathcal{V}\to\infty}\langle\hat{\boldsymbol{\phi}}(x)\rangle_{j\boldsymbol{n}} = \phi_{\rm eq}(T)\boldsymbol{n} \quad [13]$$

We may also regard $\boldsymbol{j}$ as representing an interaction with the external environment (e.g., other fields). If such a term is present during the cooling of the system through the phase transition, it will constrain the direction of the spontaneous symmetry breaking. Note that one always arrives in this way at one of the degenerate vacua $|0_{\boldsymbol{n}}\rangle$, not a linear combination of them.

## Goldstone Bosons

The Goldstone theorem states that spontaneous breaking of any continuous global symmetry leads inevitably (except, as we discuss later, in the presence of long-range forces) to the appearance of massless modes – the Goldstone bosons.

The proof is straightforward. Associated with any continuous symmetry there is a Noether current satisfying the continuity equation $\partial_\mu\hat{j}^\mu = 0$ and such that infinitesimal symmetry transformations are generated by the spatial integral of $\hat{j}^0$. The fact that the symmetry is broken means that there is some scalar field $\hat{\phi}(x)$ whose vacuum expectation value $\langle 0|\hat{\phi}(0)|0\rangle$ is not invariant under the symmetry transformation. Hence,

$$\lim_{\mathcal{V}\to 0} {\rm i}\int_{\mathcal{V}} {\rm d}^3x\langle 0|[\hat{j}^0(x), \hat{\phi}(0)]|0\rangle|_{x^0=0} \neq 0 \quad [14]$$

Moreover, the time derivative of this integral is

$$\lim_{\mathcal{V}\to 0} {\rm i}\int_{\mathcal{V}} {\rm d}^3x\langle 0|[\partial_0\hat{j}^0(x), \hat{\phi}(0)]|0\rangle|_{x^0=0} = -\lim_{\mathcal{V}\to 0} {\rm i}\int_{\partial\mathcal{V}} {\rm d}S_k\langle 0|[\hat{j}^k(x), \hat{\phi}(0)]|0\rangle|_{x^0=0} = 0 \quad [15]$$

where $\partial\mathcal{V}$ is the bounding surface of $\mathcal{V}$. This vanishes because the surface integral is zero – in a relativistic theory, because the commutator vanishes at space-like separation, and more generally in the absence of long-range interactions because it tends rapidly to zero at large spatial separation.

Now, inserting a complete set of momentum eigenstates $|n, \boldsymbol{p}\rangle$ in [14], we can see that there must exist states such that $\langle n, \boldsymbol{p}|\hat{\phi}(0)|0\rangle \neq 0$, with $p^0 \to 0$ in the limit $|\boldsymbol{p}| \to 0$, that is, massless modes.

One can see this more directly in the U(1) model above. Consider a vacuum state $|0\rangle$ such that $\langle 0|\hat{\phi}|0\rangle = \eta/\sqrt{2}$ is real. Then it is useful to shift the origin of $\phi$ by writing

$$\phi(x) = \frac{1}{\sqrt{2}}[\eta + \varphi_1(x) + {\rm i}\varphi_2(x)] \quad [16]$$

where $\varphi_1$ and $\varphi_2$ are real. Then the Lagrangian becomes

$$\mathcal{L} = \tfrac{1}{2}\Big[\dot{\varphi}_1^2 - (\nabla\varphi_1)^2 + \dot{\varphi}_2^2 - (\nabla\varphi_2)^2 - \lambda\eta^2\varphi_1^2 - \lambda\eta\varphi_1(\varphi_1^2 + \varphi_2^2) - \tfrac{1}{4}\lambda(\varphi_1^2 + \varphi_2^2)^2\Big] \quad [17]$$

Evidently, the field $\varphi_1$, corresponding to radial oscillations in $\phi$, is massive, with mass $\sqrt{\lambda}\eta$. But there is no term in $\varphi_2^2$, so $\varphi_2$ is massless.

In the case of spontaneous symmetry breaking of nonabelian symmetries, there may be several Goldstone bosons, one for each broken component of the continuous symmetry. In our theory with symmetry group $G = {\rm O}(N)$, the possible values of the vacuum expectation value at $T = 0$ are $\langle 0_{\boldsymbol{n}}|\hat{\phi}(0)|0_{\boldsymbol{n}}\rangle \approx \eta\boldsymbol{n}$,

where $\boldsymbol{n}$ is an arbitrary unit vector. In this case, for given $\boldsymbol{n}$, there is an unbroken symmetry subgroup

$$H = \{R \in \mathrm{O}(N) : R\boldsymbol{n} = \boldsymbol{n}\} = \mathrm{O}(N-1) \qquad [18]$$

and the number of broken symmetries is

$$\dim G - \dim H = N - 1 \qquad [19]$$

Thus, the radial component of $\phi$ is massive, and there are $N-1$ Goldstone bosons, the $N-1$ transverse components.

## Spontaneously Broken Gauge Theories

As we shall see, symmetry breaking in gauge theories is a more problematic concept but, for the moment, these complications are ignored and the present discussion will continue with an approach similar to that used above.

The simplest local gauge symmetry theory is a U(1) Higgs model, a model of a complex scalar field $\phi(x)$ interacting with a gauge potential $A_\mu(x)$, described by the Lagrangian

$$\mathcal{L} = D_\mu\phi^* D^\mu\phi - \tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} - V(|\phi|) \qquad [20]$$

where $V$ is a sombrero potential as in [5], while the covariant derivative $D_\mu\phi$ and gauge field $F_{\mu\nu}$ are given by

$$D_\mu\phi = \partial_\mu\phi + \mathrm{i}eA_\mu\phi, \qquad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \qquad [21]$$

The model is invariant under the local U(1) gauge transformations

$$\begin{aligned}\phi(x) &\mapsto \phi(x)\mathrm{e}^{\mathrm{i}\alpha(x)} \\ A_\mu(x) &\mapsto A_\mu(x) - \frac{1}{e}\partial_\mu\alpha(x)\end{aligned} \qquad [22]$$

The Goldstone theorem does not apply to local-symmetry theories. The problem is that to have a Hilbert space containing only physical states one must eliminate the gauge freedom by choosing a gauge condition (e.g., in the U(1) case the Coulomb gauge $\partial_k A^k(x) = 0$, which has the effect of restricting the number of polarization states of photons to two). This necessarily breaks manifest Lorentz invariance, although the theory is, of course, still fully Lorentz invariant. The proof of the theorem fails because the current is no longer local; the long-range Coulomb interaction makes the commutator fall off only like $1/r^2$, so the surface integral no longer vanishes in the infinite-volume limit. (The theorem also fails for nonrelativistic models with long-range forces.)

Again, consider a vacuum state $|0\rangle$ in which $\langle 0|\hat{\phi}|0\rangle = \eta/\sqrt{2}$, and make the same decomposition, [16]. Then, if we set

$$A'_\mu = A_\mu + \frac{1}{e\eta}\partial_\mu\varphi_2 \qquad [23]$$

we find that the kinetic term for $\varphi_2$ has been absorbed into a mass term $(1/2)e^2\eta^2 A'_\mu A'^\mu$ for the vector field. We have a model with only massive fields: the "Higgs field" $\varphi_1$ with mass $\sqrt{\lambda}\eta$ and the gauge field $A'_\mu$ with mass $e\eta$. The Goldstone bosons have been "eaten up" by the vector field to provide its longitudinal mode. This is the Higgs mechanism, first noted by Anderson in the context of the photon in a plasma becoming a massive plasmon.

A more elegant way of seeing this is to note that we can always make a gauge transformation to ensure that $\phi$ is real (at least so long as $\phi \neq 0$; where it is zero, there may be problems). This means that $\phi(x) = (1/\sqrt{2})(\eta + \varphi_1)$; $\varphi_2$ disappears altogether, and its kinetic term reduces to $(1/2)e^2 A_\mu A^\mu(\eta + \varphi_1)^2$, which includes the mass term for $A_\mu$ as well as cubic and quartic interaction terms.

As before, the discussion can be generalized to nonabelian theories, although there are additional problems to be discussed later. If we have a local symmetry group $G$ that breaks spontaneously to leave an unbroken subgroup $H$, then the gauge fields associated with $H$ remain massless. Each of the (dim $G$ – dim $H$) complementary fields "eats up" one of the Goldstone bosons, becoming massive in the process. We are left only with other, "radial" components of $\boldsymbol{\phi}$, the massive Higgs fields.

Consider, for example, a local SO(3) model, with scalar fields $\boldsymbol{\phi} = (\phi_a)_{a=1,2,3}$ and gauge potentials $\boldsymbol{A}_\mu = (A_{a\mu})$. The infinitesimal gauge transformations are

$$\delta\boldsymbol{\phi} = \boldsymbol{\delta\omega} \times \boldsymbol{\phi}, \qquad \delta\boldsymbol{A}_\mu = \boldsymbol{\delta\omega} \times \boldsymbol{A} - \frac{1}{e}\partial_\mu\boldsymbol{\delta\omega} \qquad [24]$$

where $\boldsymbol{\delta\omega}$ is the gauge parameter. The Lagrangian is

$$\mathcal{L} = \tfrac{1}{2}D_\mu\boldsymbol{\phi}\cdot D^\mu\boldsymbol{\phi} - \tfrac{1}{4}\boldsymbol{F}_{\mu\nu}\cdot\boldsymbol{F}^{\mu\nu} - \tfrac{1}{8}\lambda(\boldsymbol{\phi}^2 - \eta^2)^2 \qquad [25]$$

where the covariant derivative and gauge field are

$$\begin{aligned}D_\mu\boldsymbol{\phi} &= \partial_\mu\boldsymbol{\phi} + e\boldsymbol{A}_\mu \times \boldsymbol{\phi} \\ \boldsymbol{F}_{\mu\nu} &= \partial_\mu\boldsymbol{A}_\nu - \partial_\nu\boldsymbol{A}_\mu + e\boldsymbol{A}_\mu \times \boldsymbol{A}_\nu\end{aligned} \qquad [26]$$

If we take $\langle\hat{\boldsymbol{\phi}}\rangle$ in the 3-direction, the fields $A_{1\mu}$ and $A_{2\mu}$ absorb the Goldstone fields $\phi_1$, $\phi_2$ to become massive. As in the abelian case, we can use the local SO(3) invariance to rotate $\boldsymbol{\phi}$ everywhere to the 3-direction, and write $\boldsymbol{\phi} = (0, 0, \eta + \varphi_3)$. In this gauge the kinetic term $(1/2)(e\boldsymbol{A}_\mu \times \boldsymbol{\phi})^2$ gives a mass $e\eta$ to the fields $A_{1\mu}, A_{2\mu}$ while $A_{3\mu}$ remains massless, and the Higgs field $\varphi_3$ again has mass $\sqrt{\lambda}\eta$.

## Elitzur's Theorem; the Role of Gauge Fixing

The concept of spontaneous symmetry breaking in the context of a local symmetry requires further discussion, in particular because of Elitzur's theorem, proved in 1975, which states in essence that "spontaneous breaking of a local symmetry is impossible." In the light of this theorem, it may seem that a "spontaneously broken gauge theory" is an oxymoron. In fact, it means something rather different, although even that is not unproblematic.

The theorem was proved in the context of lattice gauge theory, where the spatial continuum is replaced by a discrete lattice. The scalar field is then represented by values $\boldsymbol{\phi}_x$ at each lattice site, and the gauge potential by values $\boldsymbol{A}_{x,\mu}$ on the links of the lattice. This is significant because on the lattice one can use a manifestly gauge-invariant formalism. Expectation values of gauge-invariant physical variables can be found, for example, by a Monte Carlo algorithm that effectively averages over all possible gauges. In this context, it is possible to show that the expectation value of any gauge-noninvariant operator (such as $\hat{\boldsymbol{\phi}}_x$) necessarily vanishes identically.

To be more specific, suppose we incorporate a symmetry-breaking term of the form $-\boldsymbol{j}\cdot\sum_x \boldsymbol{\phi}_x$, and consider the limits $\mathcal{V}\to\infty$ followed by $\boldsymbol{j}\to\boldsymbol{0}$. In the global-symmetry case, as we noted earlier, this yields the nonzero result [13]. However, in the case of a local gauge symmetry, one can show rigorously that

$$\lim_{j\to 0}\lim_{\mathcal{V}\to\infty}\langle\hat{\boldsymbol{\phi}}_x\rangle_{jn} = \boldsymbol{0} \qquad [27]$$

The essential reason for this is that we can make a gauge transformation in the neighborhood of the point $x$ to make $\boldsymbol{\phi}_x$ have any value we like without changing the energy by more than a very small amount that goes to zero as $j\to 0$. Within this manifestly gauge-invariant formalism, it is clear that the expectation value of a gauge-noninvariant operator such as $\hat{\boldsymbol{\phi}}$ is not an appropriate order parameter. One must instead look for a gauge-invariant order parameter.

It is important to note, however, that this result applies only in the context of a manifestly gauge-invariant formalism. But, in general, gauge theories cannot be quantized in a manifestly gauge-invariant way. In a path-integral formalism, the action functional, which appears in the exponent, is constant along the orbits of the gauge-group action. Consequently, the integral contains an infinite factor, the volume of the (infinite-dimensional) gauge group. There are corresponding divergences in the perturbation series. As is well known, this problem can be dealt with by introducing a gauge-fixing term, which explicitly breaks the gauge symmetry, and renders Elitzur's theorem inapplicable. But this procedure leaves a global symmetry unbroken, and it is in fact that global symmetry that is broken spontaneously.

One example is the Landau–Ginzburg model of a superconductor, which is essentially just the non-relativistic limit of the abelian Higgs model, although there is one significant difference: here the field $\hat{\phi}$ annihilates a Cooper pair, a bound pair of electrons with equal and opposite momenta and spins, so $e$ above is replaced by the charge $2e$ of a Cooper pair. The appearance of a condensate of Cooper pairs in the low-temperature superconducting phase corresponds to a state in which $\langle\hat{\phi}\rangle$ is nonzero. This would not be possible without fixing a gauge. In the nonrelativistic context, the obvious gauge to choose is the Coulomb gauge, defined by the condition $\partial_k A^k = 0$. This gauge-fixing condition breaks the local symmetry explicitly, but it leaves unbroken the global symmetry $\phi(x)\to\phi(x)\mathrm{e}^{\mathrm{i}\alpha}$ with constant $\alpha$. It is that global symmetry that is spontaneously broken when $\langle\hat{\phi}\rangle\neq 0$.

For a model with nonabelian local symmetry the standard procedure used to derive a perturbation expansion is that of Faddeev and Popov. Consider, for example, the SO(3) gauge theory discussed in the preceding section. To fix the gauge, we can choose a set of functions $\boldsymbol{F}=(F_a)$ of the fields, and introduce into the path integral a gauge-fixing term of the form

$$\mathcal{L}_{\mathrm{gf}} = -\frac{1}{2\xi}\boldsymbol{F}^2 \qquad [28]$$

where $\xi$ is an arbitrary real constant. However, to ensure that this does not bias the integral, so that the gauge-fixed theory is at least formally equivalent to the original gauge-invariant theory, one must also include the determinant of the Jacobian matrix

$$J_{ab}(x,y) = \frac{\delta F_a(x)}{\delta\omega_b(y)} \qquad [29]$$

The easiest way to do this is to introduce Faddeev–Popov ghost fields $\bar{\mathbf{C}},\mathbf{C}$, which are scalar Grassmann variables, and an appropriate term in the Lagrangian

$$\mathcal{L}_{\mathrm{FP}} = \bar{\boldsymbol{C}}\cdot\boldsymbol{J}\cdot\boldsymbol{C} \qquad [30]$$

For the SO(3) model, a convenient choice of gauge is the $R_\xi$ gauge defined by

$$\boldsymbol{F} = \partial_\mu\boldsymbol{A}^\mu - \xi e\eta\,\boldsymbol{n}\times\boldsymbol{\phi} \qquad [31]$$

where $\boldsymbol{n}$ is an arbitrarily chosen unit vector. It is clear that the full Lagrangian $\mathcal{L} + \mathcal{L}_{\text{gf}} + \mathcal{L}_{\text{FP}}$ is no longer invariant under the full SO(3) gauge group, although there is a residual U(1) gauge invariance corresponding to rotations about $\boldsymbol{n}$. In this gauge, the arbitrary choice of $\boldsymbol{n}$ means that the global SO(3) symmetry is also broken. However, for other choices, such as the Lorentz gauge $\boldsymbol{F} = \partial_\mu \boldsymbol{A}^\mu$ or axial gauge $\boldsymbol{F} = \boldsymbol{A}_3$, the Lagrangian *is* invariant under global SO(3) rotations of all the fields. This global symmetry is then spontaneously broken, with $\hat{\boldsymbol{\phi}}$ acquiring as before a nonzero expectation value of the form $\langle \hat{\boldsymbol{\phi}}(x) \rangle = \eta \boldsymbol{n}$.

It is interesting to look again at the particle content of this model. By setting $\boldsymbol{\phi}(x) = \eta \boldsymbol{n} + \boldsymbol{\varphi}(x)$ with $\boldsymbol{n} = (0, 0, 1)$, one finds that in the quadratic part of the Lagrangian, the cross-terms between $\boldsymbol{A}_\mu$ and $\boldsymbol{\varphi}$ combine to form a total divergence which can be dropped. As before, $\varphi_3$ is the Higgs field, with $m^2 = \lambda\eta^2$, $A_{3\mu}$ is the massless gauge field corresponding to the unbroken gauge symmetry, and the three transverse components of $A_{1\mu}$ and $A_{2\mu}$ represent the massive vector fields, with $m^2 = e^2\eta^2$. There are, however, also unphysical fields with $\xi$-dependent masses: $\varphi_{1,2}, C_{1,2}, \bar{C}_{1,2}$, and the longitudinal components $\partial_\mu A^\mu_{1,2}$ all have $m^2 = \xi e^2 \eta^2$. We can now compute the effective potential $V_{\text{eff}}(T, \boldsymbol{\varphi})$. One point that should be noted in performing this calculation is that the ghost fields $\boldsymbol{C}, \bar{\boldsymbol{C}}$ contribute negatively. Obviously, $V_{\text{eff}}$, being $\xi$-dependent, is not itself physically meaningful. Nevertheless, it can be shown that the stationary points of $V_{\text{eff}}$ are physical, and correspond to the possible equilibrium states of the theory. Moreover, the extremal values of $V_{\text{eff}}$ are independent of $\xi$ and give correctly the thermodynamic potential in the corresponding equilibrium states. The negative contributions from the ghost fields to $N_*$ and $M_*^2$ ensure that the $\xi$ dependence cancels out, and we find as expected $N_* = 9$ and $M_*^2 = (\lambda + 6e^2)\eta^2$.

## Phase Transitions and Crossovers

Our discussion so far has for the most part been restricted to a semiclassical or mean-field approximation. It is important to bear in mind, however, that this approximation does not suffice to determine whether a phase transition (where the thermodynamic free energy is nonanalytic) exists, or what its nature is. Determining the detailed characteristics of phase transitions requires other methods, such as the renormalization group or lattice simulations. In many cases, it is far from trivial to establish the order of the transition, or even whether a true phase transition actually exists.

Gauge theories pose particular problems because of the infrared divergences in the thermal field theory at high temperature, and because in asymptotically free nonabelian theories the coupling becomes large at very low energy. Even when they appear to exhibit spontaneous symmetry breaking, they do not necessarily undergo a true phase transition. Lattice gauge theory calculations have led to the conclusion that in nonabelian gauge theories with the Higgs field in the fundamental representation, there are values of the coupling constants for which there is no phase transition, only a rapid but smooth crossover from one type of behavior to another, so that the high- and low-temperature phases are analytically connected. If the coupling constant is small, there is a first-order phase transition, and for moderate values the theory exhibits a very rapid crossover that looks quite similar to a symmetry-breaking phase transition. Nevertheless, the analytic connection between the two phases implies that there cannot exist an order parameter that is strictly zero above the transition and nonzero below it.

In particular, it appears that for physical values of the Higgs mass, the electroweak theory does not undergo in fact undergo a true phase transition. It is somewhat ironic that the most famous example of a spontaneously broken gauge theory probably does not, strictly speaking, exhibit a symmetry-breaking phase transition!

## Conclusions

We have discussed the main features of spontaneous symmetry breaking in both the global- and local-symmetry cases, especially the appearance of Goldstone bosons when a continuous global symmetry breaks, and their elimination in the local-symmetry case by the Higgs mechanism, as well as the problems attaching to the concept of spontaneous symmetry breaking in gauge theories.

*See also:* Abelian Higgs Vortices; Effective Field Theories; Electroweak Theory; Finite Group Symmetry Breaking; Lattice Gauge Theory; Noncommutative Geometry and the Standard Model; Phase Transitions in Continuous Systems; Quantum Central Limit Theorems; Quantum Spin Systems; Symmetries in Quantum Field Theory of Lower Spacetime Dimensions; Topological Defects and their Homotopy Classification.

## Further Reading

Anderson PW (1963) Plasmons, gauge invariance and mass. *Physical Review* 130: 439–442.

Coleman S (1985) *Aspects of Symmetry*, ch. 5. Cambridge: Cambridge University Press.
Elitzur S (1975) Impossibility of spontaneously breaking local symmetries. *Physical Review D* 12: 3978–3982.
Englert F and Brout R (1964) Broken symmetry and the mass of gauge vector bosons. *Physical Review Letters* 13: 321–323.
Fradkin E and Shenker SH (1979) Phase diagrams of lattice gauge theories with Higgs fields. *Physical Review D* 19: 3682–3697.
Goldstone J, Salam A, and Weinberg S (1962) Broken symmetries. *Physical Review* 127: 965–970.
Guralnik GS, Hagen CR, and Kibble TWB (1964) Global conservation laws and massless particles. *Physical Review Letters* 13: 585–587.
Higgs PW (1964) Broken symmetries, massless particles and gauge fields. *Physics Letters* 12: 132–133.
Kibble TWB (1967) Symmetry-breaking in non-abelian gauge theories. *Physical Review* 155: 1554–1561.
Weinberg S (1996) *The Quantum Theory of Fields, Vol. II. Modern Applications*, chs. 19 and 21. Cambridge: Cambridge University Press.

# Symmetry Classes in Random Matrix Theory

**M R Zirnbauer**, Universität Köln, Köln, Germany



## Introduction

A classification of random matrix ensembles by symmetries was first established by Dyson, in an influential 1962 paper with the title "the threefold way: algebraic structure of symmetry groups and ensembles in quantum mechanics." Dyson's threefold way has since become fundamental to various areas of theoretical physics, including the statistical theory of complex many-body systems, mesoscopic physics, disordered electron systems, and the field of quantum chaos.

Over the last decade, a number of random matrix ensembles beyond Dyson's classification have come to the fore in physics and mathematics. On the physics side, these emerged from work on the low-energy Dirac spectrum of quantum chromodynamics (QCD) and from the mesoscopic physics of low-energy quasiparticles in disordered superconductors. In the mathematical research area of number theory, the study of statistical correlations in the values of Riemann zeta and similar functions has prompted some of the same generalizations.

In this article, Dyson's fundamental result will be reviewed from a modern perspective, and the recent extension of Dyson's threefold way will be motivated and described. In particular, it will be explained why symmetry classes are associated with large families of symmetric spaces.

## The Framework

Random matrices have their physical origin in the quantum world, more precisely in the statistical theory of strongly interacting many-body systems such as atomic nuclei. Although random matrix theory is nowadays understood to be of relevance to numerous areas of physics – *see* Random Matrix Theory in Physics – quantum mechanics is still where many of its applications lie. Quantum mechanics also provides a natural framework in which to classify random matrix ensembles.

Following Dyson, the mathematical setting for classification consists of two pieces of data:

- A finite-dimensional complex vector space $V$ with a Hermitian scalar product $\langle \cdot, \cdot \rangle$, called a "unitary structure" for short. (In physics applications, $V$ will usually be the truncated Hilbert space of a family of quantum Hamiltonian systems.)
- On $V$ there acts a group $G$ of unitary and antiunitary operators (the joint symmetry group of the multiparameter family of quantum systems).

Given this setup, one is interested in the linear space of self-adjoint operators on $V$ – the Hamiltonians $H$ – with the property that they commute with the $G$-action. Such a space is reducible in general, that is, the matrix of $H$ decomposes into blocks. The goal of classification is to list all of the irreducible blocks that occur.

### Symmetry Groups

Basic to classification is the notion of a symmetry group in quantum Hamiltonian systems, a notion that will now be explained.

In classical mechanics, the symmetry group $G_0$ of a Hamiltonian system is understood to be the group of canonical transformations that commute with the phase flow of the system. An important example is the rotation group for systems in a central field.

In passing from classical to quantum mechanics, one replaces the classical phase space by a quantum-mechanical Hilbert space $V$ and assigns to the symmetry group $G_0$ a (projective) representation by unitary $\mathbb{C}$-linear operators on $V$. Besides the one-parameter continuous subgroups, whose significance is highlighted by Noether's theorem, the components of $G_0$ not connected with the identity play an

important role. A prominent example is provided by the operator for space reflection. Its eigenspaces are the subspaces of states with positive and negative parity; these reduce the matrix of any reflection-invariant Hamiltonian to two blocks.

Not all symmetries of a quantum-mechanical system are of the canonical, unitary kind: the prime counterexample is the operation of inverting the time direction, called time reversal for short. In classical mechanics, this operation reverses the sign of the symplectic structure of phase space; in quantum mechanics, its algebraic properties reflect the fact that inverting the time direction, $t \mapsto -t$, amounts to sending $\mathrm{i} = \sqrt{-1}$ to $-\mathrm{i}$. Indeed, time $t$ enters in the Dirac, Pauli, or Schrödinger equation as $\mathrm{i}\hbar\mathrm{d}/\mathrm{d}t$. Therefore, time reversal is represented in the quantum theory by an antiunitary operator $T$, which is to say that $T$ is complex antilinear:

$$T(zv) = \bar{z}Tv \quad (z \in \mathbb{C}, v \in V)$$

and preserves the Hermitian scalar product or unitary structure up to complex conjugation:

$$\langle Tv_1 \, , Tv_2 \rangle = \overline{\langle v_1 \, , v_2 \rangle} = \langle v_2 \, , v_1 \rangle$$

Another operation of this kind is charge conjugation in relativistic theories such as the Dirac equation.

By the symmetry group $G$ of a quantum-mechanical system with Hamiltonian $H$, one then means the group of all unitary and antiunitary transformations $g$ of $V$ that leave the Hamiltonian invariant: $gHg^{-1} = H$. We denote the unitary subgroup of $G$ by $G_0$, and the set of antiunitary operators in $G$ by $G_1$ (not a group). If $V$ carries extra structure, as will be the case for some extensions of Dyson's basic scheme, the action of $G$ on $V$ has to be compatible with that structure.

The set $G_1$ may be empty. When it is not, the composition of any two elements of $G_1$ is unitary, so every $g \in G_1$ can be obtained from a fixed element of $G_1$, say $T$, by right multiplication with some $U \in G_0$: $g = TU$. In other words, when $G_1$ is nonempty the coset space $G/G_0$ consists of exactly two elements, $G_0$ and $T \cdot G_0 = G_1$. We shall assume that $T$ represents some inversion symmetry such as time reversal or charge conjugation. $T$ must then be a (projective) involution, that is, $T^2 = z \times \mathrm{Id}$ with $z$ a complex number of unit modulus, so that conjugation by $T^2$ is the identity operation. Since $T$ is complex antilinear, the associative law $T^2 \cdot T = T \cdot T^2$ forces $z$ to be real, and hence $T^2 = \pm\mathrm{Id}$.

Finding the total symmetry group of a Hamiltonian system need not always be straightforward, but this complication will not be an issue here: we take the symmetry group $G$ and its action on the Hilbert space $V$ as fundamental and given, and then ask what are the corresponding symmetry classes, meaning the irreducible spaces of Hamiltonians on $V$ that commute with $G$.

For technical reasons, we assume the group $G_0$ to be compact; this is an assumption that covers most (if not all) of the cases of interest in physics. The noncompact group of space translations can be incorporated, if necessary, by wrapping the system around a torus, whereby translations are turned into compact torus rotations.

While the primary objects to classify are the spaces of Hamiltonians $H$, we shall focus for convenience on the spaces of time evolutions $U_t = \mathrm{e}^{-\mathrm{i}tH/\hbar}$ instead. This change of focus results in no loss, as the Hamiltonians can always be retrieved by linearizing in $t$ at $t = 0$.

### Symmetric Spaces

We appropriate a few basic facts from the theory of symmetric spaces.

Let $M$ be a connected $m$-dimensional Riemannian manifold and $p$ a point of $M$. In some open subset $N_p$ of a neighborhood of $p$ there exists a map $s_p : N_p \to N_p$, the geodesic inversion with respect to $p$, which sends a point $x \in N_p$ with normal coordinates $(x_1, \ldots, x_m)$ to the point with normal coordinates $(-x_1, \ldots, -x_m)$. The Riemannian manifold $M$ is called locally symmetric if the geodesic inversion is an isometry, and is called globally symmetric if $s_p$ extends to an isometry $s_p : M \to M$, for all $p \in M$. A globally symmetric Riemannian manifold is called a symmetric space for short.

The Riemann curvature tensor of a symmetric space is covariantly constant, which leads one to distinguish between three cases: the scalar curvature can be positive, zero, or negative, and the symmetric space is said to be of compact type, Euclidean type, or noncompact type, respectively. (In mesoscopic physics, each type plays a role: the first provides us with the scattering matrices and time evolutions, the second with the Hamiltonians, and the third with the transfer matrices.) The focus in the current article will be on compact type, as it is this type that houses the unitary time evolution operators of quantum mechanics. The compact symmetric spaces are subdivided into two major subtypes, both of which occur naturally in the present context, as follows.

### Type II

Consider first the case where the antiunitary component $G_1$ of the symmetry group is empty, so the data are $(V, G)$ with $G = G_0$. Let $\mathscr{U}(V)$ denote the group of all complex linear transformations that

leave the structure of the vector space $V$ invariant. Thus, $\mathscr{U}(V)$ is a group of unitary transformations if $V$ carries no more than the usual Hermitian scalar product; and is some subgroup of the unitary group if $V$ does have extra structure (as is the case for the Nambu space of quasiparticle excitations in a superconductor). The symmetry group $G_0$, by acting on $V$ and preserving its structure, is contained as a subgroup in $\mathscr{U}(V)$.

Let now $H$ be any Hamiltonian with the prescribed symmetries. Then the time evolution $t \mapsto U_t = e^{-itH/\hbar}$ generated by $H$ is a one-parameter subgroup of $\mathscr{U}(V)$ which commutes with the $G_0$-action. The total set of transformations $U_t$ that arise in this way is called the (connected part of the) "centralizer" of $G_0$ in $\mathscr{U}(V)$, and is denoted by $Z$. This is the "good" set of unitary time evolutions – the set compatible with the given symmetries of an ensemble of quantum systems.

The centralizer $Z$ is obviously a group: if $U$ and $U'$ belong to $Z$, then so do their inverses and their product. What can one say about the structure of the group $Z$?

Since $G_0$ is compact by assumption, its group action on $V$ is completely reducible and $V$ is guaranteed to have an orthogonal decomposition

$$V = \bigoplus_{\lambda} V_\lambda$$

where the sum runs over isomorphism classes of irreducible $G_0$-representations $\lambda$, and the vector spaces $V_\lambda$ are called the $G_0$-*isotypic* components of $V$. For example, if $G_0$ is the rotation group $SO_3$, the $G_0$-isotypic component $V_\lambda$ of $V$ is the subspace spanned by all the states with total angular momentum $\lambda$.

Consider now any $U \in Z$. Since $U$ commutes with the $G_0$-action, it does not connect different $G_0$-isotypic components. (Indeed, in the example of $SO_3$-invariant dynamics, angular momentum is conserved and transitions between different angular momentum sectors are forbidden.) Thus, every $G_0$-isotypic component $V_\lambda$ is an invariant subspace for the action of $Z$ on $V$, and $Z$ decomposes as $Z = \prod_\lambda Z_\lambda$ with blocks $Z_\lambda = Z|_{V_\lambda}$.

To say more, fix a standard irreducible $G_0$-module $R_\lambda$ of isomorphism class $\lambda$ and consider

$$L_\lambda := \mathrm{Hom}_{G_0}(R_\lambda, V_\lambda)$$

the linear space of $\mathbb{C}$-linear maps $l : R_\lambda \to V_\lambda$ that intertwine the $G_0$-actions on $R_\lambda$ and $V_\lambda$. An element of $L_\lambda$ is called a $G_0$-*equivariant homomorphism*. By Schur's lemma, $L_\lambda \cong \mathbb{C}$ if $V_\lambda$ is $G_0$-irreducible. More generally, $\dim L_\lambda =: m_\lambda$ counts the multiplicity of occurrence of $R_\lambda$ in $V_\lambda$; for example, in the case of $G_0 = SO_3$ we take $R_\lambda$ to be the standard irreducible module of dimension $2\lambda + 1$; and $m_\lambda$ then is the number of times a multiplet of states with total angular momentum $\lambda$ occurs in $V_\lambda$.

The natural mapping $L_\lambda \otimes R_\lambda \to V_\lambda$ by $l \otimes r \mapsto l(r)$ is an isomorphism,

$$V_\lambda \cong L_\lambda \otimes R_\lambda$$

and using it we can transfer the entire discussion from $V_\lambda$ to $L_\lambda \otimes R_\lambda$. The group $G_0$ acts trivially on $L_\lambda \cong \mathbb{C}^{m_\lambda}$ and irreducibly on $R_\lambda$. Therefore, the component $Z_\lambda$ of the centralizer $Z$ is the unitary group

$$Z_\lambda \cong \mathrm{U}(L_\lambda) \cong \mathrm{U}_{m_\lambda}$$

if $V$ is a unitary vector space with no extra structure. In the presence of extra structure (which, by compatibility with the $G_0$-action, restricts to every subspace $V_\lambda$), the factor $Z_\lambda$ is some subgroup of $\mathrm{U}_{m_\lambda}$. In all cases, $Z$ is a direct product of connected compact Lie groups $Z_\lambda$.

To make the connection with symmetric spaces, write $M := Z_\lambda$. Since $M$ is a group, the operation of taking the inverse, $U \mapsto U^{-1}$, makes sense for all $U \in M$. Moreover, being a compact Lie group, the manifold $M$ admits a left- and right-invariant Riemannian structure in which the inversion $U \mapsto U^{-1}$ is an isometry. By translation, one gets an isometry $s_{U_1} : U \mapsto U_1 U^{-1} U_1$ for every $U_1 \in M$. All of these maps $s_{U_1}$ are globally defined, and the restriction of $s_{U_1}$ to some neighborhood of $U_1$ coincides with the geodesic inversion with respect to $U_1$. Thus, $M$ is a symmetric space by the definition given above. Symmetric spaces of this kind are called type II.

### Type I

Consider next the case of $G_1 \neq \emptyset$, where some antiunitary symmetry $T$ is present. As before, let $Z$ be the connected component of the centralizer of $G_0$ in $\mathscr{U}(V)$. Conjugation by $T$,

$$U \mapsto \tau(U) := TUT^{-1}$$

is an automorphism of $\mathscr{U}(V)$ and, owing to $T^2 = \pm \mathrm{Id}$, $\tau$ is involutive. Because $G_0 \subset G$ is a normal subgroup, $\tau$ restricts to an involutive automorphism (still denoted by $\tau$) of $Z$. Now recall that $T$ is complex antilinear and the good Hamiltonians are subject to $THT^{-1} = H$. The good time evolutions $U_t = e^{-itH/\hbar}$ clearly satisfy $\tau(U_t) = U_{-t} = U_t^{-1}$. Thus, the good set to consider is $\mathscr{M} := \{U \in Z \mid U = \tau(U)^{-1}\}$. The set $\mathscr{M}$ is a manifold, but in general is not a Lie group.

Further details depend on what $\tau$ does with the factorization $Z = \prod_\lambda Z_\lambda$. If $V_\lambda$ is a $G_0$-isotypic

component of $V$, then so is $TV_\lambda$, since $T$ normalizes $G_0$. Thus, either $V_\lambda \cap TV_\lambda = 0$, or $TV_\lambda = V_\lambda$. In the former case, the involutive automorphism $\tau$ just relates $U \in Z_\lambda$ with $\tau(U) \in Z_{TV_\lambda}$, whence no intrinsic constraint on $Z_\lambda$ results, and the time evolutions $(U, \tau(U)^{-1}) \in Z_\lambda \times Z_{TV_\lambda}$ constitute a type-II symmetric space, as before.

A novel situation occurs when $TV_\lambda = V_\lambda$, in which case $\tau$ restricts to an automorphism of $Z_\lambda$. Let therefore $TV_\lambda = V_\lambda$, put $K \equiv Z_\lambda$ for short, and consider

$$M := \{U \in K | U = \tau(U)^{-1}\}$$

Note that if two elements $p, p_0$ of $K$ are in $M$, then so is the product $p_0 p^{-1} p_0$. The group $K$ acts on $M \subset K$ by

$$k \cdot U = kU\tau(k)^{-1} \quad (k \in K)$$

and this group action is transitive, that is, every $U \in M$ can be written as $U = k\tau(k)^{-1}$ with some $k \in K$. (Finding $k$ for a given $U$ is like taking a square root, which is possible since $\exp : \mathrm{Lie}\, K \to K$ is surjective.) There exists such a $K$-invariant Riemannian structure for $M$ that for all $p_0 \in M$ the mapping $s_{p_0} : M \to M$ defined by

$$s_{p_0}(p) = p_0 p^{-1} p_0$$

is the geodesic inversion with respect to $p_0 \in M$. Thus, in this natural geometry $M$ is a globally symmetric Riemannian manifold and hence a symmetric space. The present kind of symmetric space is called type I. If $K_\tau$ is the set of fixed points of $\tau$ in $K$, the symmetric space $M$ is analytically diffeomorphic to the coset space $K/K_\tau$ by

$$K/K_\tau \to M \subset K, \qquad UK_\tau \mapsto U\tau(U)^{-1}$$

which we call the "Cartan embedding" of $K/K_\tau$ into $K$.

In summary, the solution to the problem of finding the set of unitary time evolution operators that are compatible with a given symmetry group $G$ and structure of Hilbert space $V$ is always a symmetric space. This is a valuable insight, as symmetric spaces are rigid objects and have been completely classified by Cartan.

If the dimension of $V$ is kept variable, the irreducible symmetric spaces that occur belong to one of the large families listed in Table 1.

**Table 1** The large families of symmetric spaces. The form of $H$ in the header applies to the last seven families

| *Family* | *Symmetric space* | *Form of* $H = \begin{pmatrix} W & Z \\ Z^\dagger & -\bar{W} \end{pmatrix}$ |
|---|---|---|
| $A$ | $\mathrm{U}_N$ | Complex Hermitian |
| $A$I | $\mathrm{U}_N/\mathrm{O}_N$ | Real symmetric |
| $A$II | $\mathrm{U}_{2N}/\mathrm{USp}_{2N}$ | Quaternion self-adjoint |
| $C$ | $\mathrm{USp}_{2N}$ | $Z$ complex symmetric, $W = W^\dagger$ |
| $C$I | $\mathrm{USp}_{2N}/\mathrm{U}_N$ | $Z$ complex symmetric, $W = 0$ |
| $D$ | $\mathrm{SO}_{2N}$ | $Z$ complex skew, $W = W^\dagger$ |
| $D$III | $\mathrm{SO}_{2N}/\mathrm{U}_N$ | $Z$ complex skew, $W = 0$ |
| $A$III | $\mathrm{U}_{p+q}/\mathrm{U}_p \times \mathrm{U}_q$ | $Z$ complex $p \times q$, $W = 0$ |
| $BD$I | $\mathrm{SO}_{p+q}/\mathrm{SO}_p \times \mathrm{SO}_q$ | $Z$ real $p \times q$, $W = 0$ |
| $C$II | $\mathrm{USp}_{2p+2q}/\mathrm{USp}_{2p} \times \mathrm{USp}_{2q}$ | $Z$ quaternion $2p \times 2q$, $W = 0$ |

## Dyson's Threefold Way

Recall the goal: given a Hilbert space $V$ and a symmetry group $G$ acting on it, one wants to classify the (irreducible) spaces of time evolution operators $U$ that are "compatible" with $G$, meaning

$$U = g_0 U g_0^{-1} = g_1 U^{-1} g_1^{-1} \quad (\text{for all } g_\sigma \in G_\sigma)$$

As we have seen, the spaces that arise in this way are symmetric spaces of type I or II depending on the nature of the time reversal (or other antiunitary symmetry) $T$.

An even stronger statement can be made when more information about the Hilbert space $V$ is specified. In Dyson's classification, the Hermitian scalar product of $V$ is assumed to be the only invariant structure that exists on $V$. With that assumption, only three large families of symmetric spaces arise; these correspond to what we call the "Wigner–Dyson symmetry classes."

### Class *A*

Recall that in Dyson's case, the connected part of the centralizer of $G_0$ in $\mathscr{U}(V)$ is a direct product of unitary groups, each factor being associated with one $G_0$-isotypic component $V_\lambda$ of $V$. The type-II situation occurs when the set $G_1$ of antiunitary symmetries is either empty or else exchanges different $V_\lambda$. In both cases, the set of good time evolution operators restricted to one $G_0$-isotypic component $V_\lambda$ is a unitary group $\mathrm{U}_{m_\lambda}$, with $m_\lambda$ being the multiplicity of the irreducible $G_0$-representation $\lambda$ in $V_\lambda$.

The unitary groups $\mathrm{U}_{N=m_\lambda}$ or, to be precise, their simple parts $\mathrm{SU}_N$, are called type-II symmetric spaces of the $A$ family or $A$ series – hence the name class $A$. The Hamiltonians $H$, the generators of time evolutions $U_t = \mathrm{e}^{-itH/\hbar}$, in this class are represented by

complex Hermitian $N \times N$ matrices. By putting a $U_N$-invariant Gaussian probability measure

$$\exp\left(-\mathrm{tr}H^2/2\sigma^2\right) \mathrm{d}H \quad (\sigma \in \mathbb{R})$$

on that space, one gets what is called the GUE – the Gaussian unitary ensemble – which defines the Wigner–Dyson universality class of unitary symmetry.

### Classes *A*I and *A*II

Consider next the case $G_1 \neq \emptyset$, with antiunitary generator $T$. Let $V_\lambda = TV_\lambda$ be any $G_0$-isotypic component of $V$ invariant under $T$ (the type-I situation). The mapping $U \mapsto TUT^{-1} = \tau(U)$ then is an automorphism of the groups $U(V_\lambda), G_0$ and $K = Z_\lambda \cong U_{m_\lambda}$. If $K_\tau$ is the subgroup of fixed points of $\tau$ in $K$, the space of good time evolutions can be identified with the symmetric space $K/K_\tau$ by the Cartan embedding. Our task is to determine $K_\tau$.

To simplify the notation let us write $V_\lambda \equiv V, R_\lambda \equiv R$, and $L_\lambda \equiv L$. We now ask what happens with $T: V \to V$ in the process of transfer to $L \otimes R \cong V$. The answer, so we claim, is that $T$ transfers to a *pure* tensor made from antiunitary maps $\alpha: L \to L$ and $\beta: R \to R$,

$$T = \alpha \otimes \beta$$

To prove this claim, let $C$ be the antilinear map from $V$ to the *dual* vector space $V^*$ by $v \mapsto \langle v, \cdot \rangle$. Because the elements of $G_0$ are represented by unitaries, the $\mathbb{C}$-linear operator $CT: V \to V^*$ intertwines $G_0$-actions:

$$CT\, a(g) = g^{-1^*} CT \quad (g \in G_0)$$

where $a$ is the automorphism $a(g) = T^{-1}gT$. From the irreducibility of $R$ it follows that the space of intertwiners $R \to R^*$ is one dimensional here (Schur's lemma). Therefore, $CT: L \otimes R \to L^* \otimes R^*$ must be a pure tensor (as opposed to a sum of such tensors), and since $C$ is clearly a pure tensor, so is $T$. This completes the proof.

By the involutive property $T^2 = \varepsilon_T \mathrm{Id}_V\, (\varepsilon_T = \pm 1)$, the two antiunitary factors of $T = \alpha \otimes \beta$ cannot but square to $\alpha^2 = \varepsilon_\alpha \mathrm{Id}_L$ and $\beta^2 = \varepsilon_\beta \mathrm{Id}_R$ where $\varepsilon_\alpha, \varepsilon_\beta = \pm 1$ are related by $\varepsilon_\alpha \varepsilon_\beta = \varepsilon_T$. The factor $\alpha$ determines a nondegenerate complex bilinear form $Q: L \times L \to \mathbb{C}$ by

$$Q(l_1, l_2) = \langle \alpha l_1, l_2 \rangle_L \quad (l_1, l_2 \in L)$$

Since $\alpha$ is antiunitary one has the exchange symmetry

$$Q(l_1, l_2) = \overline{\langle \alpha^2 l_1, \alpha l_2 \rangle}_L = \varepsilon_\alpha\, Q(l_2, l_1)$$

Thus, the complex bilinear form (or pairing) $Q$ is symmetric for $\varepsilon_\alpha = +1$ and alternating for $\varepsilon_\alpha = -1$.

Knowing the sign of $\varepsilon_\alpha = \pm 1$ we know the group $K_\tau$. Indeed, an element $k \in K_\tau$ commutes with $T$ and after transfer from $V$ to $L$ still commutes with $\alpha$. But since $K_\tau$ is a subgroup of $K = U_{m_\lambda}$, this means that $k \in K_\tau$ preserves $Q$. In the case of $\varepsilon_\alpha = +1$, what is preserved is a symmetric pairing, and therefore $K_\tau \cong O_{m_\lambda}$. For $\varepsilon_\alpha = -1$, the multiplicity $m_\lambda$ must be even and $K_\tau$ preserves an alternating pairing (or symplectic structure); in that case $K_\tau \cong \mathrm{USp}_{m_\lambda}$, the unitary symplectic group.

Thus, there is a dichotomy for the sets of good time evolutions $M \cong K/K_\tau$:

$$\text{Class } A\text{I}: \quad K/K_\tau \cong U_N/O_N \qquad (N = m_\lambda)$$
$$\text{Class } A\text{II}: \quad K/K_\tau \cong U_{2N}/\mathrm{USp}_{2N} \quad (2N = m_\lambda)$$

Again we are referring to symmetric spaces by the names they – or rather their simple parts $SU_N/SO_N$ and $SU_{2N}/\mathrm{USp}_{2N}$ – have in the Cartan classification.

In general, there is no immediate means of predicting the parity $\varepsilon_\alpha$, and one has no choice but to go through the steps of constructing $\alpha$. If $\beta: R \to R$ happens to be $G_0$-invariant, however, the situation simplifies. In that case $\beta$ determines a $G_0$-invariant pairing $R \times R \to \mathbb{C}$ (in the same way as $\alpha$ determines $Q: L \times L \to \mathbb{C}$ above). On general grounds, an irreducible $G_0$-representation space admits at most one such pairing. If that pairing is symmetric, then, as we have seen, $\varepsilon_\beta = 1$; if it is alternating, then $\varepsilon_\beta = -1$. The parity $\varepsilon_\alpha$ is given by $\varepsilon_\alpha \varepsilon_\beta = \varepsilon_T$.

**Example** Consider any physical system with spin-rotation symmetry ($G_0 = SU_2$) and time-reversal symmetry. The physical operation of time reversal, $T$, commutes with spin rotations and, hence, here is a case where the factor $\beta$ in $T = \alpha \otimes \beta$ is $G_0$-invariant. On fundamental physics grounds one has $T^2 = (-1)^{2S}$ on states with spin $S$. The spin-$S$ representation of $SU_2$ is known to carry an invariant pairing which is symmetric or skew depending on whether the integer $2S$ is even or odd. Therefore, $\varepsilon_T = \varepsilon_\beta$ and $\varepsilon_\alpha = +1$ in all cases.

Thus, $T$-invariant systems with no symmetries other than energy and spin invariably are class $A$I. By breaking spin-rotation symmetry ($G_0 = \{\mathrm{Id}\}$, $\varepsilon_\beta = +1$) while maintaining $T$-symmetry for states with half-integer spin (say single electrons, which carry spin $S = 1/2$), one gets $\varepsilon_\alpha = -1$, thereby realizing class $A$II.

**The Hamiltonians** By passing to the tangent space of $K/K_\tau$ at unity one obtains Hermitian matrices with entries that are real numbers (class $A$I) or real quaternions (class $A$II). When $K_\tau$-invariant Gaussian

probability measures (called GOE resp., GSE) are put on these spaces, one gets the Wigner–Dyson universality classes of orthogonal resp., symplectic symmetry. In mesoscopic physics, these are realized in disordered metals with time-reversal invariance (absence of magnetic fields and magnetic impurities). Spin-rotation symmetry is broken by strong spin–orbit scatterers such as gold impurities.

### Warning

The word "symmetry class" is not synonymous with "universality class." Indeed, inside a symmetry class many different types of physical behavior are possible. For example, random matrix models for disordered metallic grains with time-reversal symmetry belong to the symmetry class of the example above (class *A*I), and so do Anderson tight-binding models with real hopping. The former are known to exhibit energy level statistics of universal GOE type, whereas the latter have localized eigenfunctions and hence level statistics which is expected to approach the Poisson limit when the system size goes to infinity.

## Disordered Superconductors

When Dirac first wrote down his famous equation in 1928, he assumed that he was writing an equation for the wave function of the electron. Later, because of the instability caused by negative-energy solutions, the Dirac equation was reinterpreted (via second quantization) as an equation for the fermionic field operators of a quantum field theory. A similar change of viewpoint is carried out in reverse in the Hartree–Fock–Bogoliubov mean-field description of quasiparticle excitations in superconductors. There, one starts from the equations of motion for linear superpositions of the electron creation and annihilation operators, and reinterprets them as a unitary quantum dynamics for what might be called the quasiparticle "wave function."

In both cases – the Dirac equation and the quasiparticle dynamics of a superconductor – there enters a structure not present in the standard quantum mechanics underlying Dyson's classification: the field operators for fermionic particles are subject to a set of relations called the "canonical anticommutation relations," and these are preserved by the quantum dynamics. Therefore, whenever second quantization is undone (assuming it can be undone) to return from field operators to wave functions, the wave-function dynamics is required to preserve some extra structure. This puts a linear constraint on the good Hamiltonians $H$. For our purposes, the best viewpoint to take is to attribute the extra invariant structure to the Hilbert space $V$, thereby turning it into a Nambu space.

### Nambu Space

Adopting the standard physics conventions of second quantization, consider some set of single-particle creation and annihilation operators $c_i^\dagger$ and $c_i$, where $i = 1, 2, \ldots$ labels an orthonormal system of single-particle states. Such operators are subject to the canonical anticommutation relations (CARs)

$$\begin{aligned} c_i^\dagger c_j + c_j c_i^\dagger &= \delta_{ij} \\ c_i^\dagger c_j^\dagger + c_j^\dagger c_i^\dagger &= 0 = c_i c_j + c_j c_i \end{aligned} \qquad [1]$$

When written in terms of $c_j + c_j^\dagger$ and $\mathrm{i}(c_j - c_j^\dagger)$, these become the standard defining relations of a Clifford algebra over $\mathbb{R}$. Field operators are linear combinations $\boldsymbol{\psi} = \sum_i (u_i c_i + f_i c_i^\dagger)$ with complex coefficients $u_i$ and $f_i$.

Now take $H$ to be some Hamiltonian which is quadratic in the creation and annihilation operators:

$$H = \sum_{i,j} W_{ij} c_i^\dagger c_j + \frac{1}{2} \sum_{i,j} \left( Z_{ij} c_i^\dagger c_j^\dagger + \bar{Z}_{ij} c_j c_i \right)$$

and let $H$ act on field operators $\boldsymbol{\psi}$ by the commutator: $H \cdot \boldsymbol{\psi} \equiv [H, \boldsymbol{\psi}]$. The time evolution of $\boldsymbol{\psi}$ is then determined by the Heisenberg equation of motion

$$-\mathrm{i}\hbar \frac{\mathrm{d}\boldsymbol{\psi}}{\mathrm{d}t} = H \cdot \boldsymbol{\psi} \qquad [2]$$

which integrates to $\boldsymbol{\psi}(t) = \mathrm{e}^{\mathrm{i}tH/\hbar} \cdot \boldsymbol{\psi}(0)$, and is easily verified to preserve the CARs [1].

The dynamical equation [2] is equivalent to a system of linear differential equations for the amplitudes $u_i$ and $f_i$. If these are assembled into vectors, and the $W_{ij}$ and $Z_{ij}$ into matrices, eqn [2] becomes

$$-\mathrm{i}\hbar \frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{u} \end{pmatrix} = \begin{pmatrix} W & Z \\ Z^\dagger & -\bar{W} \end{pmatrix} \begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{u} \end{pmatrix}$$

The Hamiltonian matrix on the right-hand side has some special properties due to $Z_{ij} = -Z_{ji}$ (from $c_i c_j = -c_j c_i$) and $W_{ij} = \bar{W}_{ji}$ (from $H$ being self-adjoint as an operator in Fock space). To keep track of these properties while imposing some unitary and antiunitary symmetries, it is best to put everything in invariant form.

So, let $U$ be the unitary vector space of annihilation operators $u = \sum_i u_i c_i$, and view the creation operators $f = \sum_i f_i c_i^\dagger$ as lying in the dual vector space $U^*$. The field operators $\boldsymbol{\psi} = u + f$ then are elements of the direct sum $U \oplus U^* =: V$, called "Nambu

space." On $V$ there exists a canonical unitary structure expressed by

$$\langle \boldsymbol{\psi}, \tilde{\boldsymbol{\psi}} \rangle = \sum_i (\bar{u}_i \tilde{u}_i + \bar{f}_i \tilde{f}_i)$$

A second canonical structure on $V = U \oplus U^*$ is given by the symmetric complex bilinear form

$$\{\boldsymbol{\psi}, \tilde{\boldsymbol{\psi}}\} = \sum_i (\tilde{f}_i u_i + f_i \tilde{u}_i) = \tilde{f}(u) + f(\tilde{u})$$

where the last expression uses the meaning of $f$ as a linear function $f : U \to \mathbb{C}$. Note that $\{\boldsymbol{\psi}, \tilde{\boldsymbol{\psi}}\}$ agrees with the anticommutator of the field operators, $\boldsymbol{\psi}\tilde{\boldsymbol{\psi}} + \tilde{\boldsymbol{\psi}}\boldsymbol{\psi}$.

Now recall that the quantum dynamics is determined by a Hamiltonian $H$ that acts on $\boldsymbol{\psi}$ by the commutator $H \cdot \boldsymbol{\psi} = [H, \boldsymbol{\psi}]$. The one-parameter groups $t \mapsto e^{itH/\hbar}$ generated by this action (the time evolutions) preserve the symmetric pairing:

$$\{\boldsymbol{\psi}, \tilde{\boldsymbol{\psi}}\} = \{e^{itH/\hbar}\boldsymbol{\psi}, e^{itH/\hbar}\tilde{\boldsymbol{\psi}}\}$$

since the anticommutation relations [1] do not change with time. They also preserve the unitary structure,

$$\langle \boldsymbol{\psi}, \tilde{\boldsymbol{\psi}} \rangle = \left\langle e^{itH/\hbar}\boldsymbol{\psi}, e^{itH/\hbar}\tilde{\boldsymbol{\psi}} \right\rangle$$

because probability in Nambu space is conserved. (Physically speaking, this holds true as long as $H$ is quadratic, i.e., many-body interactions are negligible.)

One can now pose Dyson's question again: given Nambu space $V$ and a symmetry group $G$ acting on it, what is the set of time evolution operators that preserve the structure of $V$ and are compatible with $G$? From the section "The framework," we know the answer to be some symmetric space, but which are the symmetric spaces that occur?

### Class *D*

Consider a superconductor with no symmetries in its quasiparticle dynamics, so $G = \{\mathrm{Id}\}$. (A concrete example would be a disordered spin-triplet superconductor in the vortex phase.) The time evolutions $U_t = e^{itH/\hbar}$ are then constrained only by invariance of the unitary structure and the symmetric pairing $\{,\}$ of Nambu space. These two structures are consistent; they are related by particle–hole conjugation $C$:

$$\{\boldsymbol{\psi}, \tilde{\boldsymbol{\psi}}\} = \langle C\boldsymbol{\psi}, \tilde{\boldsymbol{\psi}} \rangle$$

which is an antiunitary operator with square $C^2 = +\mathrm{Id}$. Let $V_\mathbb{R} \subset V$ denote the real vector space of fixed points of $C$. (The field operators in $V_\mathbb{R}$ are called of "Majorana" type in physics.) The condition $\{\boldsymbol{\psi}, \tilde{\boldsymbol{\psi}}\} = \{U_t\boldsymbol{\psi}, U_t\tilde{\boldsymbol{\psi}}\}$ selects a complex orthogonal group $\mathrm{SO}(V)$, and imposing unitarity yields a real orthogonal subgroup $\mathrm{SO}(V_\mathbb{R})$ with $\dim V_\mathbb{R} \in 2\mathbb{N}$ – a symmetric space of the $D$ family.

When expressed in some basis of Majorana fermions (meaning a basis of $V_\mathbb{R}$), the matrix of the time evolution generator $iH \in \mathfrak{so}(V_\mathbb{R})$ is real skew, and that of $H$ imaginary skew. The simplest random matrix model for class $D$, the SO-invariant Gaussian ensemble of imaginary skew matrices, is analyzed in the second edition of Mehta's (1991) book. From the expressions given by Mehta it is seen that the level correlation functions at high energy coincide with those of the Wigner–Dyson universality class of unitary symmetry. The level correlations at low energy, however, show different behavior defining a separate universality class. This universal behavior at low energies has immediate physical relevance, as it is precisely the low-energy quasiparticles that determine the thermal transport properties of the superconductor at low temperatures.

### Class *D*III

Let now magnetic fields and magnetic impurities be absent, so that time reversal $T$ is a symmetry of the quasiparticle system: $G = \{\mathrm{Id}, T\}$. Following the section "The framework," the set of good time evolutions is $M \cong K/K_\tau$ with $K = \mathrm{SO}(V_\mathbb{R})$ and $K_\tau$ the set of fixed points of $U \mapsto \tau(U) = TUT^{-1}$ in $K$. What is $K_\tau$?

The square of the time-reversal operator is $T^2 = -\mathrm{Id}$ (for particles with spin 1/2), and commutes with particle–hole conjugation $C$, which makes $P := iCT$ a useful operator to consider. Since $C$ by definition commutes with the action of $K$, and hence also with that of $K_\tau$, the subgroup $K_\tau$ has an equivalent description as

$$K_\tau = \{k \in \mathrm{U}(V) \mid k = PkP^{-1} = \tau(k)\}$$

The operator $P$ is easily seen to have the following properties: (1) $P$ is unitary, (2) $P^2 = \mathrm{Id}$, and (3) $\mathrm{tr}_V P = 0$. Consequently, $P$ possesses two eigenspaces $V_\pm$ of equal dimension, and the condition $k = PkP^{-1}$ fixes a subgroup $\mathrm{U}(V_+) \times \mathrm{U}(V_-)$ of $\mathrm{U}(V)$. Since $P$ contains a factor $i = \sqrt{-1}$ in its definition, it anticommutes with the antilinear operator $T$. Therefore, the automorphism $\tau$ exchanges $\mathrm{U}(V_+)$ with $\mathrm{U}(V_-)$, and the fixed-point set $K_\tau$ is the same as $\mathrm{U}(V_+) \cong \mathrm{U}_{2N}$. Thus,

$$M = K/K_\tau \cong \mathrm{SO}_{4N}/\mathrm{U}_{2N} \quad (\dim V_+ = 2N)$$

a symmetric space in the $D$III family. Note that for particles with spin 1/2 the dimension of $V_+$ has to be even.

By realizing the algebra of involutions $C, T$ as $C\boldsymbol{\psi} = (1_{2N} \otimes i\sigma_x)\bar{\boldsymbol{\psi}}$ and $T\boldsymbol{\psi} = (1_{2N} \otimes i\sigma_y)\bar{\boldsymbol{\psi}}$, the Hamiltonians $H$ in class $D$III are brought into the standard form

$$H = \begin{pmatrix} 0 & Z \\ -\bar{Z} & 0 \end{pmatrix}$$

where the $2N \times 2N$ matrix $Z$ is complex and skew.

### Class *C*

Next let the spin of the quasiparticles be conserved, as is the case for a spin-singlet superconductor with no spin–orbit scatterers present, and let time-reversal invariance be broken by a magnetic field. The symmetry group of the quasiparticle system then is the spin-rotation group: $G = G_0 = \mathrm{Spin}_3 = \mathrm{SU}_2$.

Nambu space $V$ can be arranged to be a tensor product $V = L \otimes R$ so that $G_0$ acts trivially on $L$ and by the spinor representation on the spinor space $R \equiv \mathbb{C}^2$. Since two spinors combine to give a scalar, the latter comes with an alternating bilinear form $a : R \times R \to \mathbb{C}$. In a suitable basis, the anticommutation relations [1] factor on particle–hole and spin indices. The symmetric bilinear form {,} of $V$ correspondingly factors under the tensor product decomposition $V = L \otimes R$ as

$$\{l_1 \otimes r_1, l_2 \otimes r_2\} = [l_1, l_2] \times a(r_1, r_2)$$

where [ , ] is an alternating form on $L$, giving $L$ the structure of a complex symplectic vector space.

The good set $M$ now consists of the time evolutions that, in addition to preserving the structure of Nambu space, commute with the spin-rotation group $\mathrm{SU}_2$:

$$M = \{U \in \mathrm{U}(V) | UC = CU, \forall g \in \mathrm{SU}_2 : gU = Ug\}$$

By the last condition, all time evolutions act trivially on the factor $R$. The condition $UC = CU$, which expresses invariance of the symmetric form of $V$, then implies that time evolutions preserve the alternating form of $L$. Time evolutions therefore are unitary symplectic transformations of $L$, hence $M = \mathrm{USp}(L) \cong \mathrm{USp}_{2N}$ – a symmetric space of the $C$ family. The Hamiltonian matrices in class $C$ have the standard form

$$H = \begin{pmatrix} W & Z \\ Z^\dagger & -\bar{W} \end{pmatrix}$$

with $W$ being Hermitian and $Z$ complex and symmetric.

### Class *C*I

The next class is obtained by taking the time reversal $T$ as well as the spin rotations $g \in \mathrm{SU}_2$ to be symmetries of the quasiparticle system.

By arguments that should be familiar by now, the set of good time evolutions is a symmetric space $M \cong K/K_\tau$ with $K = \mathrm{USp}(L)$ and $K_\tau$ the set of fixed points of $\tau$ in $K$. Once again, the question to be answered is: what is $K_\tau$? The situation here is very similar to the one for class $D$III, with $L$ and $\mathrm{USp}(L)$ taking the roles of $V$ and $\mathrm{SO}(V_\mathbb{R})$. By adapting the previous argument to the present case, one shows that $K_\tau$ is the same as $\mathrm{U}(L_+) \cong \mathrm{U}_N$, where $L_+$ is the positive eigenspace of $P = iCT$ viewed as a unitary operator on $L$. Thus,

$$M = K/K_\tau \cong \mathrm{USp}_{2N}/\mathrm{U}_N$$

## Dirac Fermions: The Chiral Classes

Three large families of symmetric spaces remain to be implemented. Although these, too, occur in mesoscopic physics, their most natural realization is by 4D Dirac fermions in a random gauge field background.

Consider the Lagrangian $\mathscr{L}$ for the Euclidean spacetime version of QCD with $N_c \geq 3$ colors of quarks coupled to an $\mathrm{SU}_{N_c}$ gauge field $A_\mu$:

$$\mathscr{L} = i\bar{\boldsymbol{\psi}}\gamma^\mu(\partial_\mu - A_\mu)\boldsymbol{\psi} + im\bar{\boldsymbol{\psi}}\boldsymbol{\psi}$$

The massless Dirac operator $D = i\gamma^\mu(\partial_\mu - A_\mu)$ anticommutes with $\gamma_5 = \gamma^0\gamma^1\gamma^2\gamma^3$. Therefore, in a basis of eigenstates of $\gamma_5$ the matrix of $D$ takes the form

$$D = \begin{pmatrix} 0 & Z \\ Z^\dagger & 0 \end{pmatrix} \quad [3]$$

If the gauge field carries topological charge $\nu \in \mathbb{Z}$, the Dirac operator $D$ has at least $|\nu|$ zero modes by the index theorem. To make a simple model of the challenging situation where $A_\mu$ is distributed according to Yang–Mills measure, one takes the matrices $Z$ to be complex rectangular, of size $p \times q$ with $p - q = \nu$, and puts a Gaussian probability measure on that space. This random matrix model for $D$ captures the universal features of the QCD Dirac spectrum in the massless limit.

The exponential of the truncated Dirac operator, $e^{itD}$ (where $t$ is not the time), lies in a space equivalent to $\mathrm{U}_{p+q}/\mathrm{U}_p \times \mathrm{U}_q$ – a symmetric space of the $A$III family. We therefore say that the universal behavior of the QCD Dirac spectrum is that of symmetry class $A$III.

But hold on! Why are we entitled to speak of a symmetry class here? By definition, symmetries

always *commute* with the Hamiltonian, never do they anticommute! (The relation $D = -\gamma_5 D \gamma_5$ is not a symmetry in the sense of Dyson, nor is it a symmetry in our sense.)

### Class *A*III

To incorporate the massless QCD Dirac operator into the present classification scheme, we adapt it to the Nambu space setting. This is done by reorganizing the four-component Dirac spinor $\boldsymbol{\psi}, \bar{\boldsymbol{\psi}}$ as an eight-component Majorana spinor $\Psi$, to write

$$\mathscr{L}_{m=0} = \frac{\mathrm{i}}{2}\Psi\Gamma^{\mu}(\partial_{\mu} - \mathscr{A}_{\mu})\Psi$$

The $8 \times 8$ matrices $\Gamma^{\mu}$ are real symmetric besides satisfying the Clifford relations $\Gamma^{\mu}\Gamma^{\nu} + \Gamma^{\nu}\Gamma^{\mu} = 2\delta^{\mu\nu}$. A possible tensor product realization is

$$\Gamma^0 = 1 \otimes \sigma_z \otimes 1, \quad \Gamma^1 = \sigma_x \otimes \sigma_y \otimes \sigma_y$$
$$\Gamma^2 = \sigma_y \otimes \sigma_y \otimes 1, \quad \Gamma^3 = \sigma_z \otimes \sigma_y \otimes \sigma_y$$

The gauge field in this Majorana representation is $\mathcal{A}_{\mu} = 1 \otimes 1 \otimes (A_{\mu}^{(-)} - A_{\mu}^{(+)}\sigma_y)$ where $A_{\mu}^{(\pm)} = (1/2)(A_{\mu} \pm A_{\mu}^{\mathrm{t}})$ are the symmetric and skew parts of $A_{\mu} \in \mathfrak{su}(N_c)$.

The operator $H = \mathrm{i}\Gamma^{\mu}(\partial_{\mu} - \mathscr{A}_{\mu})$ is imaginary skew, therefore $\mathrm{e}^{\mathrm{i}tH}$ is real orthogonal. This means that there exists a Nambu space $V$ with unitary structure $\langle , \rangle$ and symmetric pairing $\{ , \}$, both of which are preserved by the action of $\mathrm{e}^{\mathrm{i}tH}$. No change of physical meaning or interpretation is implied by the identical rewriting from Dirac $D$ to Majorana $H$. The fact that Dirac fermions are not truly Majorana is encoded in a $U_1$-symmetry $H\mathrm{e}^{\mathrm{i}\theta Q} = \mathrm{e}^{\mathrm{i}\theta Q}H$ generated by $Q = 1 \otimes 1 \otimes \sigma_y$.

Now comes the essential point: since $H$ obeys $\bar{H} = -H$, the chiral "symmetry" $H = -\Gamma_5 H \Gamma_5$ with $\Gamma_5 = 1 \otimes \sigma_x \otimes 1$ can be recast as a true symmetry:

$$H = +\Gamma_5 \bar{H} \Gamma_5 = THT^{-1}$$

with antilinear $T : \Psi \mapsto \Gamma_5 \bar{\Psi}$. Thus, the massless QCD Dirac operator is indeed associated with a symmetry class in the present, post-Dyson sense: that is class *A*III, realized by self-adjoint operators on Nambu space with Dirac $U_1$-symmetry and an antiunitary symmetry $T$.

### Classes *BD*I and CII

Consider Hamiltonians $D$ still of the form [3] but now with matrix entries taken from either the real numbers or the real quaternions. Their one-parameter groups $\mathrm{e}^{\mathrm{i}tD}$ belong to two further families of symmetric spaces, namely the classes *BD*I and CII of Table 1. These large families are known to be realized as symmetry classes by the massless Dirac operator with gauge group $SU_2$ (for *BD*I), or with fermions in the adjoint representation (for CII). For the details we must refer to Verbaarschot's (1994) paper and the recent article by Heinzner *et al.* (2005).

*See also:* Classical Groups and Homogeneous Spaces; Compact Groups and Their Representations; Determinantal Random Fields; Dirac Fields in Gravitation and Nonabelian Gauge Theory; Dirac Operator and Dirac Field; High $T_c$ Superconductor Theory; Integrable Systems in Random Matrix Theory; Lie Groups: General Theory; Random Matrix Theory in Physics; Random Partitions; Supersymmetry Methods in Random Matrix Theory; Symmetries and Conservation Laws.

## Further Reading


Altland A, Simons BD, and Zirnbauer MR (2002) Theories of low-energy quasi-particle states in disordered *d*-wave superconductors. *Physics Reports* 359: 283–354.

Altland A and Zirnbauer MR (1997) Non-standard symmetry classes in mesoscopic normal-/superconducting hybrid systems. *Physical Review B* 55: 1142–1161.

Dyson FJ (1962) The threefold way: algebraic structure of symmetry groups and ensembles in quantum mechanics. *Journal of Mathematical Physics* 3: 1199–1215.

Heinzner P, Huckleberry A, and Zirnbauer MR (2005) Symmetry classes of disordered fermions. *Communications in Mathematical Physics* 257: 725–771.

Helgason S (1978) *Differential Geometry, Lie Groups and Symmetric Spaces*. New York: Academic Press.

Mehta ML (1991) *Random Matrices*. New York: Academic Press.

Weyl H (1939) *The Classical Groups: Their Invariants and Representations*. Princeton: Princeton University Press.

Verbaarschot JJM (1994) The spectrum of the QCD Dirac operator and chiral random matrix theory: the threefold way. *Physical Review Letters* 72: 2531–2533.

# Synchronization of Chaos

**M A Aziz-Alaoui**, Université du Havre, Le Havre, France



## Introduction: Chaotic Systems Can Synchronize

Synchronization is a ubiquitous phenomenon characteristic of many processes in natural systems and (nonlinear) science. It has permanently remained an objective of intensive research and is today considered as one of the basic nonlinear phenomena studied in mathematics, physics, engineering, or life science. This word has a Greek root, *syn* = common and *chronos* = time, which means to share the common time or to occur at the same time, that is, correlation or agreement in time of different processes (Boccaletti *et al.* 2002). Thus, synchronization of two dynamical systems generally means that one system somehow traces the motion of another. Indeed, it is well known that many coupled oscillators have the ability to adjust some common relation that they have between them due to weak interaction, which yields to a situation in which a synchronization-like phenomenon takes place.

The original work on synchronization involved periodic oscillators. Indeed, observations of (periodic) synchronization phenomena in physics go back at least as far as C Huygens (1673), who, during his experiments on the development of improved pendulum clocks, discovered that two very weakly coupled pendulum clocks become synchronized in phase: two clocks hanging from a common support (on the same beam of his room) were found to oscillate with exactly the same frequency and opposite phase due to the (weak) coupling in terms of the almost imperceptible oscillations of the beam generated by the clocks.

Since this discovery, periodic synchronization has found numerous applications in various domains, for instance, in biological systems and living nature where synchronization is encountered on different levels. Examples range from the modeling of the heart to the investigation of the circadian rhythm, phase locking of respiration with a mechanical ventilator, synchronization of oscillations of human insulin secretion and glucose infusion, neuronal information processing within a brain area and communication between different brain areas. Also, synchronization plays an important role in several neurological diseases such as epilepsies and pathological tremors, or in different forms of cooperative behavior of insects, animals, or humans (Pikovsky *et al.* 2001).

This process may also be encountered in celestial mechanics, where it explains the locking of revolution period of planets and satellites.

Its view was strongly broadened with the developments in radio engineering and acoustics, due to the work of Eccles and Vincent, 1920, who found synchronization of a triode generator. Appleton, Van der Pol, and Van der Mark, 1922–27, have, experimentally and theoretically, extended it and worked on radio tube oscillators, where they observed entrainment when driving such oscillators sinusoidally, that is, the frequency of a generator can be synchronized by a weak external signal of a slightly different frequency.

But, even though original notion and theory of synchronization implies periodicity of oscillators, during the last decades, the notion of synchronization has been generalized to the case of interacting chaotic oscillators. Indeed, the discovery of deterministic chaos introduced new types of oscillating systems, namely the chaotic generators.

Chaotic oscillators are found in many dynamical systems of various origins; the behavior of such systems is characterized by instability and, as a result, limited predictability in time.

Roughly speaking, a system is chaotic if it is deterministic, has a long-term aperiodic behavior, and exhibits sensitive dependence on initial conditions on a closed invariant set (the chaos theory is discussed in more detail elsewhere in this encyclopedia) (*see* Chaos and Attractors).

Consequently, for a chaotic system, trajectories starting arbitrarily close to each other diverge exponentially with time, and quickly become uncorrelated. It follows that two identical chaotic systems cannot synchronize. This means that they cannot produce identical chaotic signals, unless they are initialized at exactly the same point, which is in general physically impossible. Thus, at first sight, synchronization of chaotic systems seems to be rather surprising because one may intuitively (and naively) expect that the sensitive dependence on initial conditions would lead to an immediate breakdown of any synchronization of coupled chaotic systems. This scenario in fact led to the belief that chaos is uncontrollable and thus unusable. Despite this, in the last decades, the search for synchronization has moved to chaotic systems. Significant research has been done and, as a result, Yamada and Fujisaka (1983), Afraimovich *et al.* (1986), and Pecora and Carroll (1990) showed that

two chaotic systems could be synchronized by coupling them: synchronization of chaos is actual and chaos could then be exploitable. Ever since, many researchers have discussed the theory and the design or applications of synchronized motion in coupled chaotic systems. A broad variety of applications has emerged, for example, to increase the power of lasers, to synchronize the output of electronic circuits, to control oscillations in chemical reactions, or to encode electronic messages for secure communications.

The publication of the seminal paper of Pecora and Caroll (1990) had a very strong impact in the domain of chaos theory and chaos synchronization, and their applications. It had stimulated very intense research activities and the related studies continue to attract great attention. Many authors have contributed to developing this domain, theoretically or experimentally (Boccaletti *et al.* 2002, Pecorra *et al.* 1997, references therein).

However, the special features of chaotic systems make it impossible to directly apply the methods developed for synchronization of periodic oscillators. Moreover, in the topics of coupled chaotic systems, many different phenomena, which are usually referred to as synchronization, exist and have been studied now for over a decade. Thus, more precise descriptions of such systems are indeed desirable.

Several different regimes of synchronization have been investigated. In the following, the focus will be on explaining the essentials on this large topic, subdivided into four basic types of synchronization of coupled or forced chaotic systems which have been found and have received much attention, while emphasizing on the first three:

- identical (or complete) synchronization (IS), which is defined as the coincidence of states of interacting systems;
- generalized synchronization (GS), which extends the IS phenomenon and implies the presence of some functional relation between two coupled systems; if this relationship is the identity, we recover the IS;
- phase synchronization (PS), which means entrainment of phases of chaotic oscillators, whereas their amplitudes remain uncorrelated; and
- lag synchronization (LS), which appears as a coincidence of time-shifted states of two systems.

Other regimes exist, some of them will be briefly pointed out at the end of this article; we also will briefly discuss the very relevant issue of the stability of synchronous motions.

Our discussion and examples given here are based on unidirectionally continuous systems, most of the exposed ideas can be easily extended to discrete systems.

Let us also emphasize that the same year, 1990, saw the publication of another seminal paper, by Ott, Grebogi, and Yorke (OGY) on the control of chaos (Ott *et al.* 1990). Recently, it has been realized that synchronization and control of chaos share a common root in nonlinear control theory. Both topics were presented by many authors in a unified framework. However, synchronization of chaos has evolved in its own right, even if it is nowadays known as a part of the nonlinear control theory.

## Synchronization and Stability

For the basic master–slave configuration, where an autonomous chaotic system (the master)

$$\frac{dX}{dt} = F(X), \quad X \in \mathbb{R}^n \qquad [1]$$

drives another system (the slave),

$$\frac{dY}{dt} = G(X, Y), \quad Y \in \mathbb{R}^m \qquad [2]$$

synchronization takes place when $Y$ asymptotically copies, in a certain manner, a subset $X_p$ of $X$. That is, there exists a relation between the two coupled systems, which could be a smooth invertible function $\psi$, which transforms the trajectories on the attractor of a first system into those on the attractor of a second system. In other words, if we know, after a transient regime, the state of the first system, it allows us to predict the state of the second: $Y(t) = \psi(X(t))$. Generally, it is assumed that $n \geq m$; however, for the sake of easy readability (even if this is not a necessary restriction) the case $n = m$ will only be considered; thus, $X_p = X$. Henceforth, if we denote the difference $Y - \psi(X)$ by $X_\perp$, in order to arrive at a synchronized motion, it is expected that

$$||X_\perp|| \longrightarrow 0, \quad \text{as } t \longrightarrow +\infty \qquad [3]$$

If $\psi$ is the identity function, the process is called IS.

**Definition of IS** System [2] synchronizes with system [1], if the set $M = \{(X, Y) \in \mathbb{R}^n \times \mathbb{R}^n, Y = X\}$ is an attracting set with a basin of attraction $B(M \subset B)$ such that $\lim_{t \to \infty} \|X(t) - Y(t)\| = 0$, for all $(X(0), Y(0)) \in B$.

Thus, this regime corresponds to the situation where all the variables of two (or more) coupled chaotic systems converge.

If $\psi$ is not the identity function, the phenomenon is more general and is referred to as GS.

**Definition of GS** System [2] synchronizes with system [1], in the generalized sense, if there exists a transformation $\psi : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, a manifold $M = \{(X, Y) \in \mathbb{R}^{n+m}, Y = \psi(X)\}$ and a subset $B$ $(M \subset B)$, such that for all $(X_0, Y_0) \in B$, the trajectory based on the initial conditions $(X_0, Y_0)$ approaches $M$ as time goes to infinity. This is explained further in the following.

Henceforth, in the case of IS, eqn [3] above means that a certain hyperplane $M$, called synchronization manifold, within $\mathbb{R}^{2n}$, is asymptotically stable. Consequently, for the sake of synchrony motion, we have to prove that the origin of the transverse system $X_\perp = Y - X$ is asymptotically stable. That is, to prove that the motion transversal to the synchronization manifold dies out.

However, significant progress has been made by mathematicians and physicists in studying the stability of synchronous motions. Two main tools are used in the literature for this aim: conditional Lyapunov exponents and asymptotic stability. In the examples given below, we will essentially formulate conditions for synchronization in terms of Lyapunov exponents, which play a central role in chaos theory. These quantities measure the sensitive dependence on initial conditions for a dynamical system and also quantify synchronization of chaos.

The Lyapunov exponents associated with the variational equation corresponding to the transverse system $X_\perp$:

$$\frac{\mathrm{d}X_\perp}{\mathrm{d}t} = DF(X)X_\perp \qquad [4]$$

where $DF(X)$ is the Jacobian of the vector field evaluated onto the driving trajectory $X$, are referred to as transverse or conditional Lyapunov exponents (CLEs).

In the case of IS, it appears that the condition $L^\perp_{\max} < 0$ is sufficient to insure synchronization, where $L^\perp_{\max}$ is the largest CLE. Indeed, eqn [4] gives the dynamics of the motion transverse to the synchronization manifold; therefore, CLEs indicate if this motion dies out or not, and hence, whether the synchronization state is stable or not. Consequently, if $L^\perp_{\max}$ is negative, it insures the stability of the synchronized state. This will be best explained using two examples below.

Even if there exist other approaches for studying synchronization, one may ask if this condition on $L^\perp_{\max}$ is true in general. To answer this question, mathematicians have recently formulated it in terms of properties of manifolds (or synchronization hyperplanes). Some rigorous results on (generalized) synchronization, when the system is smooth, are given by Josic (2000). This approach relies on the Fenichel theory of normally hyperbolic invariant manifolds and quantities that resemble Lyapunov exponents, and is referred to as differentiable GS. However, many situations correspond to the case where, in some region of values of parameters coupling, the function $\psi$ is only continuous but not smooth, that is, the graph of $\psi$ is a complicated geometrical object. This kind of synchronization is called nonsmooth GS (Afraimovich *et al.* 2001).

Furthermore, the mathematical theory of IS often assumes the coupled oscillators to be identical, even if, in practice, no two oscillators are exact copies of each other. This leads to small differences in system parameters and then to synchronization errors. These errors have been studied by many authors (see, e.g., Illing (2002), and references therein).

## Identical Synchronization

Perhaps the best way to explain synchronization of chaos is through IS, also referred to as conventional or complete synchronization (Boccaletti *et al.* 2002). It is the simplest form of chaos synchronization and generalizes to the complete replacement which is explained below. It is also the most typical form of chaotic synchronization often observable in two identical systems.

There are various processes leading to synchronization; depending on the particular coupling configuration used these processes could be very different. So, one has to distinguish between the following two main situations, even if they are, in some sense, similar: the unidirectional and the bidirectional coupling. Indeed, synchronization of chaotic systems is often studied for schemes of the form

$$\begin{aligned} \frac{\mathrm{d}X}{\mathrm{d}t} &= F(X) + kN(X - Y) \\ \frac{\mathrm{d}Y}{\mathrm{d}t} &= G(Y) + kM(X - Y) \end{aligned} \qquad [5]$$

where $F$ and $G$ act in $\mathbb{R}^n, (X, Y) \in (\mathbb{R}^n)^2$, is a scalar, and $M$ and $N$ are coupling matrices belonging to $\mathbb{R}^{n\times n}$. If $F = G$ the two subsystems $X$ and $Y$ are identical. Moreover, when both matrices are nonzero then the coupling is called bidirectional, while it is referred to as unidirectional if one is the zero matrix, and the other nonzero.

### Constructing Pairs of Synchronized Systems: Complete Replacement

Pecora and Carroll (1990) proposed the use of stable subsystems of given chaotic systems to

construct pairs of unidirectionally coupled synchronizing systems. Since then generalizations of this approach have been developed and various methods now exist to synchronize systems (Wu 2002, Hasler 1998).

One way to build a couple of synchronized systems is then to use the basic construction method introduced by Pecora and Carroll, who made an important observation. They found that, when they make a replica of part of a chaotic system and send a system variable from the original system (transmitter) to drive this replica (receiver), sometimes the replica subsystem and the original chaotic one lock in their steps and evolve together chaotically in synchrony. This method can be described as follows. Consider the autonomous $n$-dimensional dynamical system,

$$\frac{\mathrm{d}u}{\mathrm{d}t} = F(u) \qquad [6]$$

divide this system into two subsystems ($u = (v, w)$),

$$\begin{aligned} \frac{\mathrm{d}v}{\mathrm{d}t} &= G(v, w) \\ \frac{\mathrm{d}w}{\mathrm{d}t} &= H(v, w) \end{aligned} \qquad [7]$$

where $v = (u_1, \ldots, u_m)$, $w = (u_{m+1}, \ldots, u_n)$, $G = (F_1, \ldots, F_m)$, and $H = (F_{m+1}, \ldots, F_n)$. Next, create a new subsystem $w'$ identical to the $w$-subsystem. This yields a $(2n - m)$-dimensional system:

$$\begin{aligned} \frac{\mathrm{d}v}{\mathrm{d}t} &= G(v, w) \\ \frac{\mathrm{d}w}{\mathrm{d}t} &= H(v, w) \\ \frac{\mathrm{d}w'}{\mathrm{d}t} &= H(v, w') \end{aligned} \qquad [8]$$

The first state-variable component $v(t)$ of the $(v, w)$ system is then used as the input to the $w'$-system. The coupling is unidirectional and the $(v, w)$ subsystem is referred to as the driving (or master) system, the $w'$-subsystem as the response (or slave) system. In this context, the following notions and results are useful.

**Definition** If $\lim_{t \to +\infty} \|w'(t) - w(t)\| = 0$ and $w'(t)$ continues to remain in step with $w(t)$ in the course of the time, the two subsystems are said to be synchronized.

**Definition** The Lyapunov exponents of the response subsystem $(w')$ for a particular driven trajectory $v(t)$ are called CLEs.

Let $w(t)$ be a chaotic trajectory with initial condition $w(0)$, and $w'(t)$ be a trajectory started at a nearly point $w'(0)$. The basic idea of the Pecora–Carroll approach is to establish the asymptotic stability of the solutions of $w'$-subsystem by means of CLEs. They have shown the following result (Pecora and Carroll 1990):

**Theorem** *A necessary and sufficient condition for the two subsystems, $w$ and$w'$, to be synchronized is that all of the CLEs be negative.*

Note that only a finite number of possible decompositions (or couplings) $v$–$w$ exist; this is bounded by the number of different possible subsystems, namely $N(N-1)/2$. (For a description and mathematical analysis of various coupling schemes see Wu (2002).) Furthermore, by splitting the main system [6] in a different way, (complete) synchronization could not exist. Indeed, in general, only a few of the possible response subsystems possess negative CLEs, and may thus be used to implement synchronizing systems using the Pecora–Caroll method. In fact, it has been pointed out in the literature that in some cases, the CLE criterion is not as practical as some other criteria.

For simplicity, the idea will now be developed on the following three-dimensional simple autonomous system, which belongs to the class of dynamical systems called generalized Lorenz systems (see Derivière and Aziz-Alaoui (2003), and references therein):

$$\begin{aligned} \dot{x} &= -9x - 9y \\ \dot{y} &= -17x - y - xz \\ \dot{z} &= -z + xy \end{aligned} \qquad [9]$$

(This should be compared with the well-known Lorenz system:

$$\begin{aligned} \dot{x} &= -10x + 10y \\ \dot{y} &= 28x - y - xz \\ \dot{z} &= -\tfrac{8}{3}z + xy \end{aligned}$$

which differs in the signs of various terms and the values of coefficients.) From previous observations, it was shown that system [9] oscillates chaotically; its Lyapunov exponents are +0.601, 0.000, and –16.470; it exhibits the chaotic attractor of Figure 1, with a three-dimensional feature very similar to that of Lorenz attractor (in fact, it satisfies the condition $z < 0$, but in our context it does not matter).

Let us divide system [9] into two subsystems $v = x_1$ and $w = (y_1, z_1)$. By creating a copy

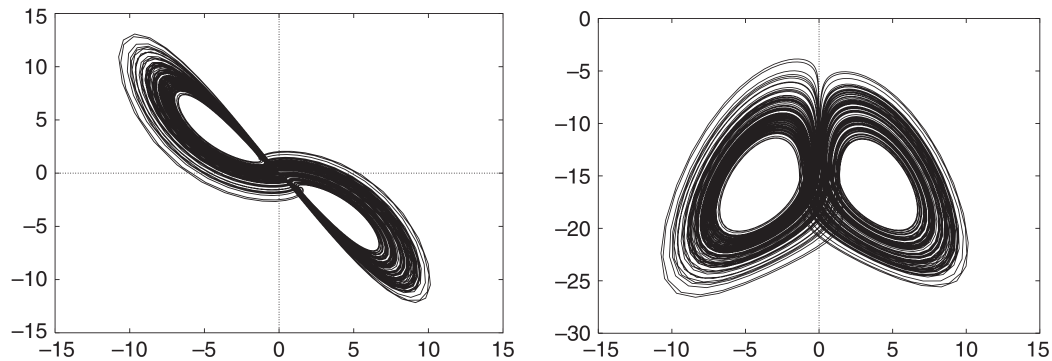


**Figure 1** The chaotic attractor of system [9]: $x$–$y$ and $x$–$z$ plane projections.

$w' = (y_2, z_2)$ of the $w$-subsystem, we obtain the following five-dimensional dynamical system:

$$\begin{aligned} \dot{x}_1 &= -9x_1 - 9y_1 \\ \dot{y}_1 &= -17x_1 - y_1 - x_1 z_1 \\ \dot{z}_1 &= -z_1 + x_1 y_1 \\ \dot{y}_2 &= -17x_1 - y_2 - x_1 z_2 \\ \dot{z}_2 &= -z_2 + x_1 y_2 \end{aligned} \qquad [10]$$

In numerical experiments, it was observed that the motion quickly results in the two equalities, $\lim_{t\to+\infty} |y_2 - y_1| = 0$ and $\lim_{t\to+\infty} |z_2 - z_1| = 0$, to be satisfied, that is, $\lim_{t\to+\infty} \|w' - w\| = 0$. These equalities persist as the system evolves. Hence, the two subsystems $w$ and $w'$ are synchronized. Figure 2 illustrates this phenomenon.

It is also easy to verify that the synchronization persists even if a slight change in the parameters of the system is made. The CLEs of the linearization of the system around the synchronous state, the negativity of which determines the stability of the synchronized solution, are also computed easily.

Pecora–Carroll similarly built the system [10] by using the following steps. Starting with two copies of system [9], a signal $x(t)$ is transmitted from the first to the second: in the second system all $x$-components are replaced with the signal from the first system, that is, $x_2$ is replaced by $x_1$ in the second system. Finally, the $\mathrm{d}x_2/\mathrm{d}t$ equation is eliminated, since it is exactly the same as $\mathrm{d}x_1/\mathrm{d}t$ equation, and is superfluous. This then results in system [10]. For this reason, Pecora–Carroll called this construction a complete replacement. Thus, it is natural to think of the $x_1$ variable as driving the second system, but also to label the first system the drive and the second system the response. In fact, this method is a particular case of the unidirectional coupling method explained below. Note also that this method could be modified by using a partial substitution approach, in which a response variable is replaced with the drive counterpart only in certain locations (Pecora *et al.* 1997).

### Unidirectional IS

The IS synchronization has also been called as one-way diffusive coupling, drive–response coupling, master–slave coupling, or negative feedback control.

System [5], $F = G$ and $N = 0$, becomes unidirectionally coupled, and reads

$$\begin{aligned} \frac{\mathrm{d}X}{\mathrm{d}t} &= F(X) \\ \frac{\mathrm{d}Y}{\mathrm{d}t} &= F(Y) + kM(X - Y) \end{aligned} \qquad [11]$$

$M$ is then a matrix that determines the linear combination of $X$ components that will be used in the difference, and $k$ determines the strength of the coupling (see, for an interesting review on this subject, Pecora *et al.* (1997)). In unidirectional synchronization, the evolution of the first system (the drive) is unaltered by the coupling, the second system (the response) is then constrained to copy the dynamics of the first. Let us consider an example with two copies of system [9], and for

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad [12]$$

that is, by adding a damping term to the first equation of the response system, we get a following unidirectionally coupled system, coupled through a linear term $k > 0$ according to variables $x_{1,2}$:

$$\begin{aligned} \dot{x}_1 &= -9x_1 - 9y_1 \\ \dot{y}_1 &= -17x_1 - y_1 - x_1 z_1 \\ \dot{z}_1 &= -z_1 + x_1 y_1 \\ \dot{x}_2 &= -9x_2 - 9y_2 - k(x_2 - x_1) \\ \dot{y}_2 &= -17x_2 - y_2 - x_2 z_2 \\ \dot{z}_2 &= -z_2 + x_2 y_2 \end{aligned} \qquad [13]$$

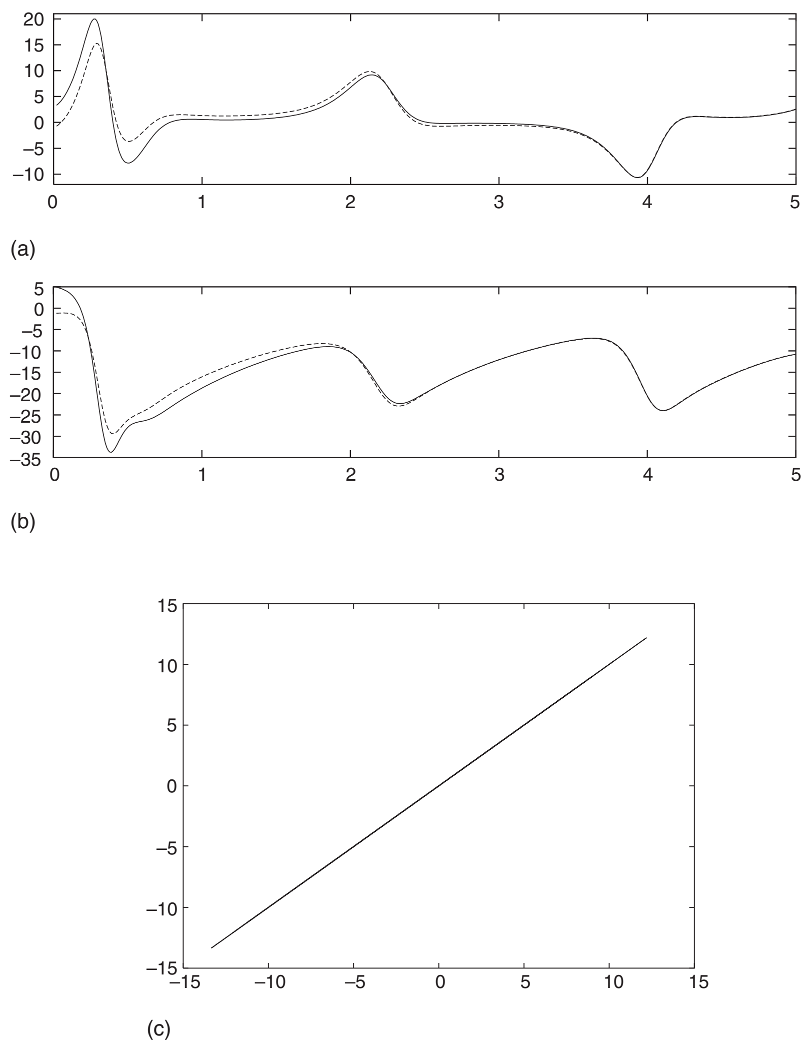



**Figure 2** Complete replacement synchronization. Time series for (a) $y_i(t)$ and (b) $z_i(t), i = 1, 2$, in system [10]. The difference between the variable of the transmitter and the variable of the receiver asymptotes tends to zero as time progresses, that is, synchronization occurs after transients die down. (c) The plot of amplitudes $y_1$ against $y_2$, after transients die down, shows a diagonal line, which also indicates that the receiver and the transmitter are maintaining synchronization. The plot of $z_1$ against $z_2$ shows a similar behavior.

For $k = 0$, the two subsystems are uncoupled; for $k > 0$ both subsystems are unidirectionally coupled; and for $k \longrightarrow +\infty$, we recover the complete replacement coupling scheme explained above. Our numerical computations yield the optimal value $\tilde{k}$ for the synchronization; we found that for $k \geq \tilde{k} = 4.999$, both subsystems of [13] synchronize. That is, starting from random initial conditions, and after some transient time, system [13] generates the same attractor as for system [9] (see **Figure 1**). Consequently, all the variables of the coupled chaotic subsystems converge: $x_2$ converges to $x_1$, $y_2$ to $y_1$, and $z_2$ to $z_1$ (see **Figure 3**). Thus, the second system (the response) is locked to the first one (the drive).

Alternatively, observation of diagonal lines in correlation diagrams, which plot the amplitudes $x_1$ against $x_2$, $y_1$ against $y_2$, and $z_1$ against $z_2$, can also indicate the occurrence of system synchronization.

IS was the first for which examples of unidirectionally coupled chaotic systems were presented. It is important for potential applications of chaos synchronization in communication systems, or for time-series analysis, where the information flow is also unidirectional.

## Bidirectional IS

A second brief example uses a bidirectional (also called mutual or two-way) coupling. In this situation, in contrast to the unidirectional coupling, both drive and response systems are connected in such a way that they influence each other's behavior. Many

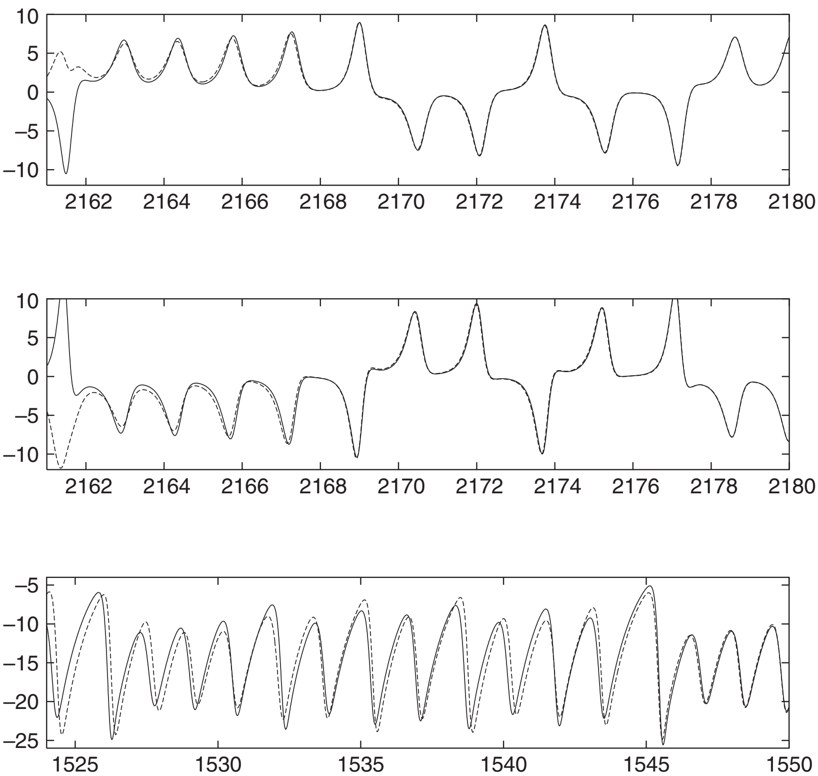



**Figure 3** Time series for $x_i(t), y_i(t)$, and $z_i(t) (i = 1, 2)$ in system [13] for the coupling constant $k = 5.0$, that is, beyond the threshold necessary for synchronization. After transients die down, the two subsystems synchronize perfectly.

biological or physical systems consist of bidirectionally interacting elements or components; examples range from cardiac and respiratory systems to coupled lasers with feedback. Let us then take two copies of the same system [9] as given above, but two-way coupled through a linear constant term $k > 0$ according to variables $x_{1,2}$:

$$\begin{aligned} \dot{x_1} &= -9x_1 - 9y_1 - k(x_1 - x_2) \\ \dot{y_1} &= -17x_1 - y_1 - x_1 z_1 \\ \dot{z_1} &= -z_1 + x_1 y_1 \\ \dot{x_2} &= -9x_2 - 9y_2 - k(x_2 - x_1) \\ \dot{y_2} &= -17x_2 - y_2 - x_2 z_2 \\ \dot{z_2} &= -z_2 + x_2 y_2 \end{aligned} \qquad [14]$$

We can get an idea of the onset of synchronization by plotting, for example, $x_1$ against $x_2$ for various values of the coupling-strength parameter $k$. Our numerical computations yield the optimal value $\tilde{k}$ for the synchronization: $\tilde{k} \simeq 2.50$ (**Figure 4**), both $(x_i, y_i, z_i)$ subsystems synchronize and system [14] also generates the attractor of **Figure 1**.

**Synchronization manifold and stability** Geometri-Geometrically, the fact that systems [13] and [14], beyond synchronization, generate the same attractor as system [9], implies that the attractors of these combined drive–response six-dimensional systems are confined to a three-dimensional hyperplane (the synchronization manifold) defined by $Y = X$. After the synchronization is reached, this manifold is a stable submanifold in the full phase space $\mathbb{R}^6$. **Figure 5** gives an idea of what the geometry of the synchronous attractor of system [13] or [14] looks like, by exhibiting the projection of the phase space $\mathbb{R}^6$ onto $(x_1, y_1, y_2)$ subspace. But, one can similarly plot any combination of variable $x_i$, $y_i$, and $z_i$ $(i = 1, 2)$, and get the same result, since the motion, in case of synchronization, is confined to the hyperplane defined in $\mathbb{R}^6$ by the equalities $x_1 = x_2$, $y_1 = y_2$, and $z_1 = z_2$.

This hyperplane is stable since small perturbations which take the trajectory off the synchronization manifold decay in time. Indeed, as stated earlier, CLEs of the linearization of the system around the synchronous state could determine the stability of the synchronized solution. This leads to requiring that the origin of the transverse system, $X_\perp$, is asymptotically stable. To see this, for both systems [13] and [14], we then switch to the new set of coordinates, $X_\perp = Y - X$, that is, $x_\perp = x_2 - x_1$, $y_\perp = y_2 - y_1$, and $z_\perp = z_2 - z_1$. The origin $(0, 0, 0)$ is obviously a fixed point for this transverse system,

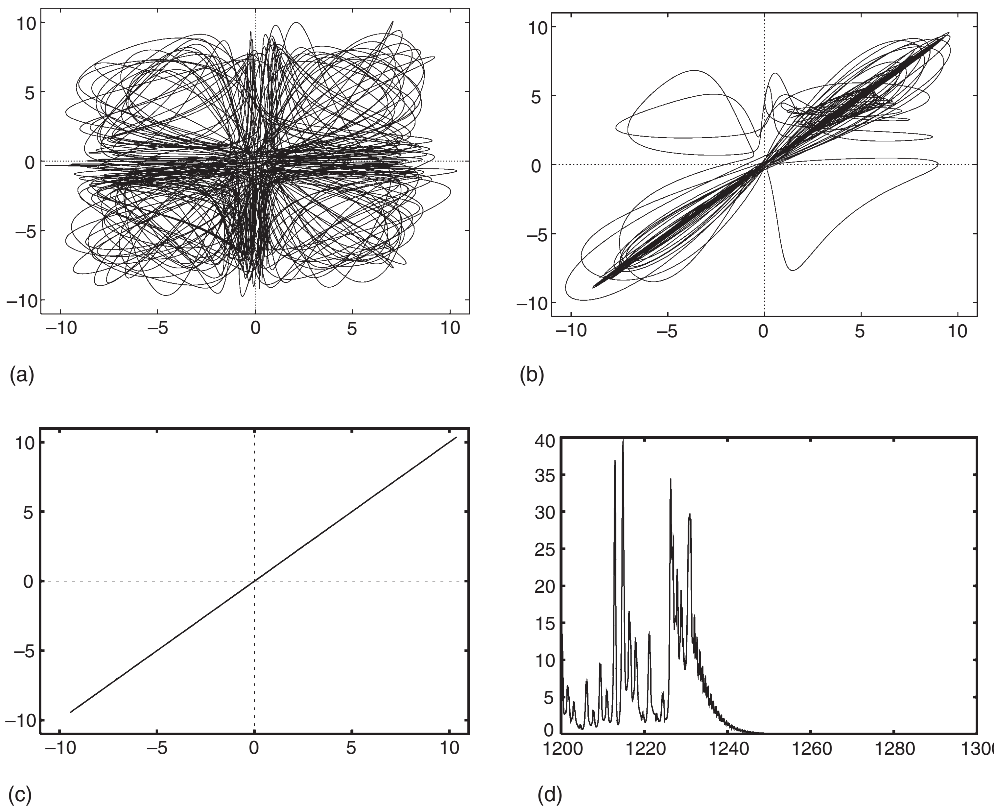



**Figure 4** Illustration of the onset of synchronization of system [14]. (a)–(c) Plots of amplitudes $x_1$ against $x_2$ for values of the coupling parameter $k = 0.5, 1.5, 2.8$, respectively. The system synchronizes for $k \geq 2.5$. (d) Plot, for $k = 2.8$, of the norm $N(X) = \|x_1 - x_2\| + \|y_1 - y_2\| + \|z_1 - z_2\|$ versus $t$, which shows that the system synchronizes very quickly.

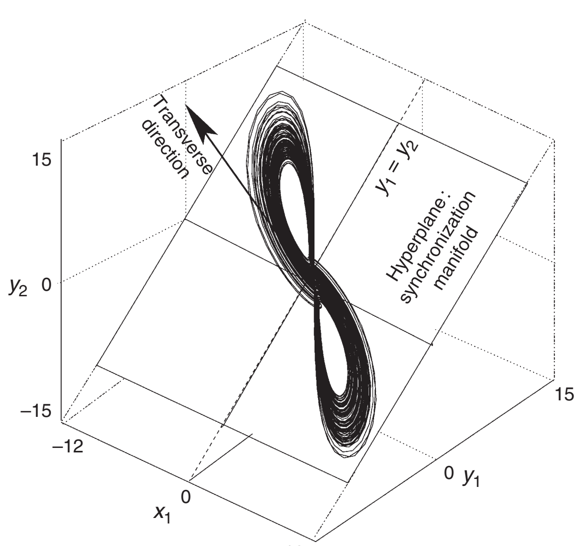



**Figure 5** The motion of synchronized system [13] or [14] takes place on a chaotic attractor which is embedded in the synchronization manifold, that is, the hyperplane defined by $x_1 = x_2$, $y_1 = y_2$, and $z_1 = z_2$.

within the synchronization manifold. Therefore, for small deviations from the synchronization manifold, this system reduces to a typical variational equation:

$$\frac{\mathrm{d}X_\perp}{\mathrm{d}t} = DF(X)X_\perp \qquad [15]$$

where $DF(X)$ is the Jacobian of the vector field evaluated onto the driving trajectory $X$, that is,

$$\begin{pmatrix} \frac{\mathrm{d}x_\perp}{\mathrm{d}t} \\ \frac{\mathrm{d}y_\perp}{\mathrm{d}t} \\ \frac{\mathrm{d}z_\perp}{\mathrm{d}t} \end{pmatrix} = V \begin{pmatrix} x_\perp \\ y_\perp \\ z_\perp \end{pmatrix} \qquad [16]$$

For systems [13] and [14], we obtain

$$V = V_i = \begin{pmatrix} -9 - k_i & -9 & 0 \\ -17 - z & -1 & -x \\ y & x & -1 \end{pmatrix} \qquad [17]$$

with $k_i = k$ for system [13] and $k_i = 2k$ for system [14]. Let us remark that the only difference between both matrices $V_i$ is the coupling $k$ which has a factor

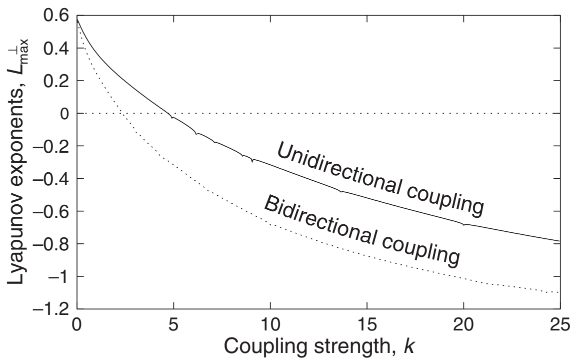



**Figure 6** The largest transverse Lyapunov exponents $L^{\perp}_{max}$ as a function of coupling strength $k$, in the unidirectional system [13] (solid) and the bidirectional system [14] (dotted).

2 in the bidirectional case. **Figure 6** shows the dependence of $L^{\perp}_{max}$ on $k$, for both examples of unidirectionally and bidirectionally coupling systems. $L^{\perp}_{max}$ becomes negative as $k$ increases, which insures the stability of the synchronized state for systems [13] and [14].

Let us note that this can also be proved analytically as done by Derivière and Aziz-Alaoui (2003) by using a suitable Lyapunov function, and using some new extended version of LaSalle invariance principle.

*Desynchronization motion* Synchronization depends not only on the coupling strength, but also on the vector field and the coupling function. For some choice of these quantities, synchronization may occur only within a finite range $[k_1, k_2]$ of coupling strength; in such a case a desynchronization phenomenon occurs. Thus, increasing $k$ beyond the critical value $k_2$ yields loss of the synchronized motion ($L^{\perp}_{max}$ becomes positive).

## Generalized Synchronization

Identical chaotic systems synchronize by following the same chaotic trajectory. However, real systems are in general not identical. For instance, when the parameters of two coupled identical systems do not match, or when these coupled systems belong to different classes, complete IS may not be expected, because there does not exist such an invariant manifold $Y = X$, as for IS. For non-identical systems, the possibility of some type of synchronization has been investigated (Afraimovich *et al.* 1986). It was shown that when two different systems are coupled with sufficiently strong coupling strength, a general synchronous relation between their states could exist and it could be expressed by a smooth invertible function, $Y(t) = \psi(X(t))$. This phenomenon, called GS, is thus a relaxed and extended form of IS in non-identical systems.

However, it may also occur for pairs of identical systems, for example, for systems having reflection symmetry, $F(-X) = -F(X)$. Besides these examples of GS, others also exist that exploit symmetries of the underlying systems (Parlitz and Kocarev 1999).

GS was introduced for unidirectionally coupled systems by Rulkov *et al.* (1995). For simplicity, we also focus on unidirectionally coupled continuous time systems:

$$\frac{dX}{dt} = F(X)$$
$$\frac{dY}{dt} = G(Y, u(t)) \quad [18]$$

where $X \in \mathbb{R}^n$, $Y \in \mathbb{R}^m$, $F : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, $G : \mathbb{R}^m \times \mathbb{R}^k \longrightarrow \mathbb{R}^m$, and $u(t) = (u_1(t), \ldots, u_k(t))$ with $u_i(t) = h_i(X(t, X_o))$. Two (non-identical) dynamical systems are said to be synchronized in a generalized sense if there is a continuous function $\psi$ from the phase space of the first to the phase space of the second, taking orbits of the first system to orbits of the second.

The main problem is to know when and under what conditions system [18] undergoes GS. Many authors have addressed this question, and it has been shown that asymptotic stability is equally significant for this more universal concept (for some theoretical results, see Rulkov *et al.* (1995) and Parlitz and Kocarev (1999)). For unidirectionally coupled continuous time systems, the following results hold:

**Theorem** *A necessary and sufficient condition for system [18] to be synchronized in the generalized sense is that for each $u(t) = u(X(t, X_o))$ the system-is asymptotically stable.*

When it is not possible to find a Lyapunov function in order to use this theorem, one can numerically compute the CLEs of the response system, and use the following result:

**Theorem** *The drive and response subsystems of system [18] synchronize in the generalized sense iff all of the CLEs of the response subsystem are negative.*

The definition of $\psi$ has the advantage that it allows the discussion of synchronization of non-identical systems and, at the same time, to consider synchronization in terms of the property of synchronization manifold. Therefore, it is important to study the existence of the transformation $\psi$ and its nature

(continuity, smoothness, ...). Unfortunately, except in special cases (Afraimovich *et al.* 1986), rarely will one be able to produce formulas exhibiting the mapping $\psi$.

An example of two unidirectionally coupled chaotic systems which synchronize in the generalized sense is given below. Consider the following Rössler system driven by system [9]:

$$\begin{aligned} \dot{x}_1 &= -9x_1 - 9y_1 \\ \dot{y}_1 &= -17x_1 - y_1 - x_1z_1 \\ \dot{z}_1 &= -z_1 + x_1y_1 \\ \dot{x}_2 &= -y_2 - z_2 - k(x_2 - (x_1^2 + y_1^2)) \\ \dot{y}_2 &= x_2 + 0.2y_2 - k(y_2 - (y_1^2 + z_1^2)) \\ \dot{z}_2 &= 0.2 + z_2(x_2 - 9.0) - k(z_2 - (x_1^2 + z_1^2)) \end{aligned} \qquad [19]$$

As shown in **Figure 7**, it appears impossible to tell what the relation is between the transmitter subsystem $(x_1, y_1, z_1)$ in eqn [19] and the two Rössler response subsystems $(x_2, y_2, z_2)$ at $k = 1$ and $k = 100$.

However, GS occurs for large values of the coupling-strength parameter $k$. Therefore, for such values we expect that orbits of [19] will lie in the vicinity of a certain synchronization manifold. Indeed, let us define the set

$$S = \{(x_1, y_1, z_1, x_2, y_2, z_2) \in \mathbb{R}^6 : x_2 = x_1^2 + y_1^2, \\ y_2 = y_1^2 + z_1^2, z_2 = x_1^2 + z_1^2\}$$

Since the projections of $S$ onto the coordinates $(x_1, y_1, x_2), (y_1, z_1, y_2)$, and $(x_1, z_1, z_2)$ are paraboloids, we can see how the synchronization manifold is approached. This is illustrated in **Figure 8**, where the $(x_1, y_1, x_2)$ projections of typical trajectories are shown at four different coupling values. (See Josic (2000) for other examples and further developments; see also Pecora *et al.* (1997), where the authors summarize a method in order to get an idea on the functional relation occurring in case of GS, between two coupled systems.)

## Phase Synchronization

For coupled non-identical chaotic systems, other types of synchronizations exist. Recently, a rather weak degree of synchronization, the PS, of chaotic systems has been described (Pikovsky *et al.* 2001). The Greek meaning of the word synchronization, mentioned in the introduction, is closely related to this type of processes. The synchronous motion is actually not visible. Indeed, in PS the phases of chaotic systems with PS are locked, that is, there exists a certain relation between them, whereas the amplitudes vary chaotically and are practically uncorrelated. Thus, it is mostly close to synchronization of periodic oscillators.

**Definition** PS of two coupled chaotic oscillators occurs if, for arbitrary integers $n$ and $m$, the phase locking condition between the corresponding phases, $|n\phi_1(t) - m\phi_2(t)| \leq$ constant, holds and the amplitudes of both systems remain uncorrelated.

Let us note that such a phenomenon occurs when a zero Lyapunov exponent of the response system becomes negative, while, as explained above, identical chaotic systems synchronize by following the same chaotic trajectory, when their largest transverse Lyapunov exponent of the synchronized manifold decreases from positive to negative values.

Moreover, following the definition above, this phenomenon is best observed when a well-defined phase variable can be identified in both coupled systems. This can be done for strange attractors that spiral around a "hole," or a particular (fixed) point in a two-dimensional projection of the attractor. The typical example is given by the Rössler system, which, for some range of parameters, exhibits a Möbius-

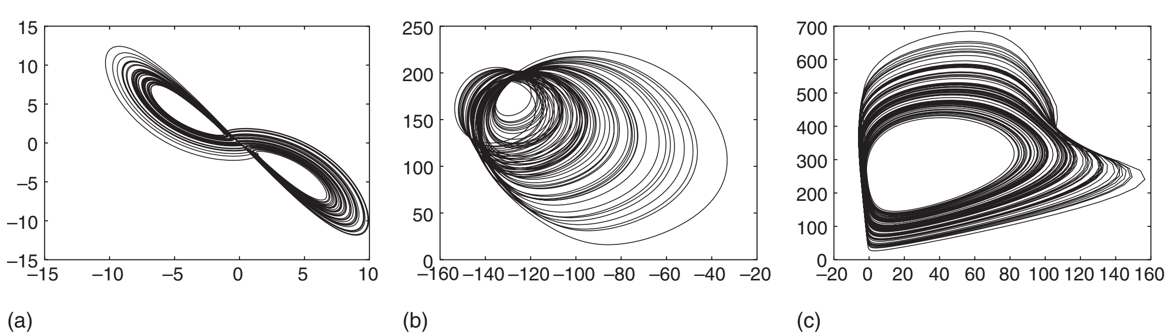



**Figure 7** Projections onto the ($x$–$y$) plane of typical trajectories of system [19]. (a) $(x_1, y_1)$ projection, that is, a typical trajectory of system [9]; (b) and (c) $(x_2, y_2)$ projections at, respectively, $k = 1$ and $k = 100$.

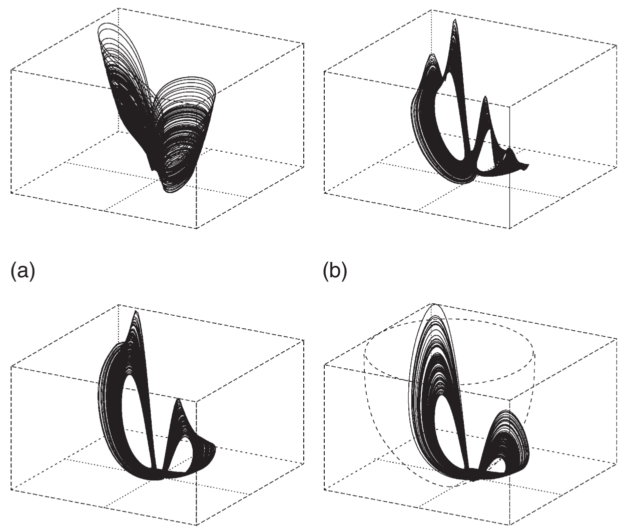



**Figure 8** Generalized synchronization. $(x_1, y_1, x_2)$ projections of typical trajectories of system [19] after transients die out, with (a) $k=1$, (b) $k=20$, (c) $k=100$, and (d) $k=200$. For the last value, the attractor lies in the set $S$, three-dimensional projections of which are paraboloïds.

strip-like chaotic attractor with a central hole. In such a case, a phase angle $\phi(t)$ can be defined that decreases or increases monotonically. For an illustration, we take the following two coupled Rössler oscillators:

$$\begin{aligned} \dot{x}_1 &= -\alpha_1 y_1 - z_1 + k(x_2 - x_1) \\ \dot{y}_1 &= \alpha_1 x_1 + 0.17 y_1 \\ \dot{z}_1 &= 0.2 + z_1(x_1 - 9.0) \\ \dot{x}_2 &= -\alpha_2 y_2 - z_2 + k(x_1 - x_2) \\ \dot{y}_2 &= \alpha_2 x_2 + 0.17 y_2 \\ \dot{z}_2 &= 0.2 + z_2(x_2 - 9.0) \end{aligned} \qquad [20]$$

with a small parameter mismatch $\alpha_{1,2} = 0.95 \pm 0.04, k$ governs the strength of coupling. If we can define a Poincaré section surface for the system, then, for each piece of a trajectory between two cross sections with this surface, we define the phase, as done in Pikovsky *et al.* (2001), as a piecewise linear function of time, so that the phase increment is $2\pi$ at each rotation:

$$\phi(t) = 2\pi \frac{t - t_n}{t_{n+1} - t_n} + 2\pi n, \quad t_n \le t \le t_{n+1}$$

where $t_n$ is the time of the $n$th crossing of the secant surface.

In our example, the last has been chosen as the negative $x$-axis and represented by the wide segment in Figure 9a. This definition of phases is clearly ambiguous since it depends on the choice of the Poincaré section; nevertheless, defined in this way, the phase has a physically important property, it does correspond to the direction with the zero Lyapunov exponent in the phase space, its perturbations neither grow nor decay in time. Figure 9c shows that there is a transition from the nonsynchronous phase regime, where the phase difference increases almost linearly with time ($k=0.01$ and $k=0.05$), to a synchronous state, where the relation $|\phi_1(t) - \phi_2(t)| < \text{constant}$ holds ($k=0.1$), that is, the phase difference does not grow with time. However, the amplitudes are obviously uncorrelated as seen in Figure 9b. This example shows that PS could takes place for weaker degree of synchronization in chaotic systems. Readers can find more rigorous mathematical discussion on this subject, and on the definition of phases of chaotic oscillators, in Pikovsky *et al.* (2001), see also Boccaletti *et al.* (2002) and references therein.

## Other Treatments and Types of Synchronization

### Lag Synchronization

PS synchronization occurs when non-identical chaotic oscillators are weakly coupled: the phases are locked, while the amplitudes remain uncorrelated. When the coupling strength becomes larger, some relationships between amplitudes may be established. Indeed, it has been shown (Rosenblum *et al.* 1997), in symmetrically coupled non-identical oscillators and in time-delayed systems, that there exists

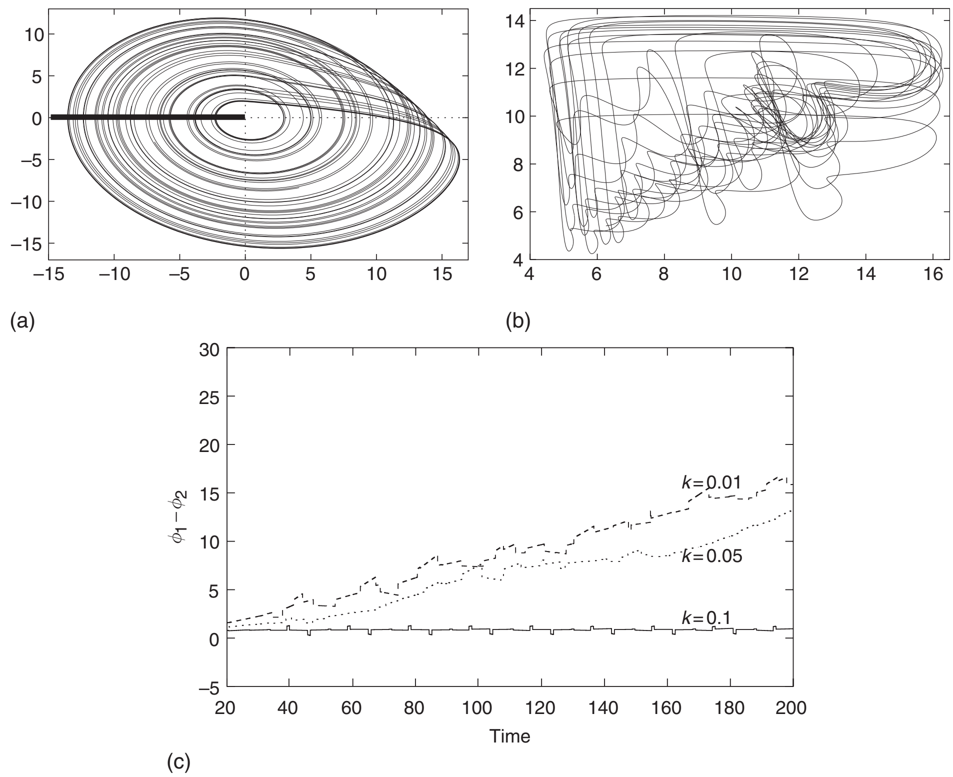



**Figure 9** (a) Rössler chaotic attractor projection onto *x*–*y* plane. (b) Amplitudes $A_1$ versus $A_2$ for the phase synchronized case at $k = 0.1$. (c) Time serie of phase difference for different coupling strengths $k$; for $k = 0.01$ PS is not achieved, while for $k = 0.1$ PS takes place. Although the phases are locked, for $k = 0.1$, the amplitudes remain chaotic and uncorrelated.

a regime of LS. This process appears as a coincidence of time-shifted states of two systems:

$$\lim_{t\to+\infty} ||Y(t) - X(t-\tau)|| = 0$$

where $\tau$ is a positive delay.

### Projective Synchronization

In coupled partially linear systems, it was reported by Mainieri and Rehacek (1999) that two identical systems could be synchronized up to a scaling factor. This type of chaotic synchronization is referred to as projective synchronization. Consider, for example, a three-dimensional chaotic system $\dot{X} = F(X)$, where $X = (x, y, z)$. Decompose $X$ into a vector $v = (x, y)$ and a scalar $z$; the system can then be rewritten as

$$\frac{\mathrm{d}u}{\mathrm{d}t} = g(v, z), \qquad \frac{\mathrm{d}z}{\mathrm{d}t} = h(v, z)$$

In projective synchronization, two identical systems $X_1 = (x_1, y_1, z_1)$ (drive) and $X_2 = (x_2, y_2, z_2)$ (response) are coupled through the scalar variable $z$. It occurs if the state vectors $v_1$ and $v_2$ synchronize up to a constant ratio, that is, $\lim_{t\to+\infty} ||\alpha v_1(t) - v_2(t)|| = 0$, where $\alpha$ is called a scaling factor. For partially linear systems, it may automatically occur provided that the systems satisfy some stability conditions.

However, this process could not be classified as GS, even if there exists a linear relation between the coupled systems, because the response system of projective synchronization is not asymptotically stable. For more information about this subject, the reader is referred to Mainieri and Rehacek (1999).

### Anticipating Synchronization

It is interesting to mention that a new form of synchronization has recently appeared, the so-called anticipating synchronization (Boccaletti *et al.* 2002). It shows that some coupled chaotic systems might synchronize such that their response anticipates the drivers by synchronizing with their future states.

It is also interesting to mention the nonlinear $H_\infty$ synchronization method for nonautonomous schemes introduced by Suykens *et al.* (1997).

### Spatio-Temporal Synchronization

Low-dimensional systems have rather limited usefulness in modeling real-world applications. This is why the synchronization of chaos has been carried

out in high dimensions (see Kocarev *et al.* (1997) for a review). See also Chen and Dong (2001) for a discussion of special high-dimensional systems, namely large arrays of coupled chaotic systems.

## Application to Transmission Systems and Secure Communication

Synchronization principles are useful in practical applications. Use of chaotic signals to transmit information has been a very active research topic in the last decade. Thus, it has been established that chaotic circuits may be used to transmit information by synchronization. As a result, several proposals for secure-communication schemes have been advanced (see, e.g., Cuomo *et al.* (1993), Hasler (1998), and Parlitz *et al.* (1999)). The first laboratory demonstration of a secure-communication system, which uses a chaotic signal for masking purposes, and which exploits the chaotic synchronization techniques to recover the signal, was reported by Kocarev *et al.* (1992).

It is difficult, within the scope of this article, to give a complete or detailed discussion, and it should be noted that there exist many competing and tested methods that are well established.

The main idea of the communication schemes is to encode a message by means of a chaotic dynamical system (the transmitter), and to decode it using a second dynamical system (the receiver) that synchronizes with the first. In general, secure-communication applications assume additionally that the coupled systems used are identical.

Different methods can be used to hide the useful information, for example, chaotic masking, chaotic switching, or direct chaotic modulation (Hasler 1998). For instance, in the chaotic masking method, an analog information carrying the signal $s(t)$ is added to the output $y(t)$ of the chaotic system in the transmitter. The receiver tries to synchronize with component $y(t)$ of the transmitted signal $s(t) + y(t)$. If synchronization takes place, the information signal can be retrieved by subtraction (**Figure 10**).

It is interesting to note that, in all proposed schemes for secure communications using the idea of synchronization (experimental realization or computer simulation), there is an inevitable noise degrading the fidelity of the original message.

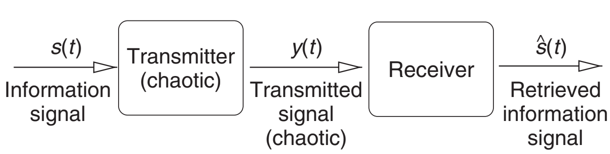



**Figure 10** A typical communication setup.

Robustness to parameter mismatch was addressed by many authors (Illing *et al.* 2002). Lozi *et al.* (1993) showed that, by connecting two identical receivers in cascade, a significant amount of the noise can be reduced, thereby allowing the recovery of a much higher quality signal.

Furthermore, different implementations of chaotic secure communication have been proposed during the last decades, as well as methods for cracking this encoding. The methods used to crack such a chaotic encoding make use of the low dimensionality of the chaotic attractors. Indeed, since the properties of low-dimensional chaotic systems with one positive Lyapunov exponent can be reconstructed by analyzing the signal, such as through the delay-time reconstruction methods, it seems unlikely that these systems might provide a secure encryption method. The hidden message can often be retrieved easily by an eavesdropper without using the receiver. But, chaotic masking and encoding are difficult to break, using the state-of-the-art analysis tools, if sufficiently high dimensional chaos generators with multiple positive Lyapunov exponents (i.e., hyperchaotic systems) are used (see Pecora *et al.* (1997), and references therein).

## Conclusion

In spite of the essential progress in theoretical and experimental studies, synchronization of chaotic systems continues to be a topic of active investigations and will certainly continue to have a broad impact in the future. Theory of synchronization remains a challenging problem of nonlinear science.

*See also:* Bifurcations of Periodic Orbits; Chaos and Attractors; Fractal Dimensions in Dynamics; Generic Properties of Dynamical Systems; Isochronous Systems; Lyapunov Exponents and Strange Attractors; Singularity and Bifurcation Theory; Stability Theory and KAM; Weakly Coupled Oscillators.

## Further Reading

Afraimovich V, Chazottes JR, and Cordonet A (2001) Synchronization in directionally coupled systems some rigourous results. *Discrete and Continuous Dynamical Systems B* 1(4): 421–442.

Afraimovich V, Verichev N, and Rabinovich MI (1986) Stochastic synchronization of oscillations in dissipative systems. *Radiophysics and Quantum Electron* 29: 795–803.

Boccaletti S, Kurths J, Osipov G, Valladares D, and Zhou C (2002) The synchronization of chaotic systems. *Physics Reports* 366: 1–101.

Chen G and Dong X (1998) *From Chaos to Order*. Singapore: World Scientific.

Cuomo K, Oppenheim A, and Strogatz S (1993) Synchronization of Lorenz-based chaotic circuits with applications to communications. *IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing* 40(10): 626–633.

Derivière S and Aziz-Alaoui MA (2003) Estimation of attractors and synchronization of generalized Lorenz systems. *Dynamics of Continuous, Discrete and Impulsive Systems Series B: Applications and Algorithms* 10(6): 833–852.

Hasler M (1998) Synchronization of chaotic systems and transmission of information. *International Journal of Bifurcation and Chaos* 8(4): 647–659.

Huygens Ch (Hugenii) (1673) Horologium Oscillatorium (English translation: 1986 The Pendulum Clock. Ames: Iowa State University Press). Parisiis, France: Apud F. Muguet.

Illing L (2002) *Chaos Synchronization and Communications in Semiconductor Lasers*. Ph.D. dissertation,. San Diego: University of California.

Illing L, Brocker J, Kocarev L, Parlitz U, and Abarbanel H (2002) When are synchronization errors small? *Physical Review E* 66: 036229.

Josic K (2000) Synchronization of chaotic systems and invariants manifolds. *Nonlinearity* 13: 1321–1336.

Kocarev Lj, Halle K, Eckert K, and Chua LO (1992) Experimental demonstration of secure communication via chaotic synchronization. *International Journal of Bifurcation and Chaos* 2(3): 709–713.

Kocarev Lj, Tasev Z, Stojanovski T, and CParlitz U (1997) Synchronizing spatiotemporel chaos. *Chaos* 7(4): 635–643.

Lozi R and Chua LO (1993) Secure communications via chaotic synchronization II: noise reduction by cascading two identical receivers. *International Journal of Bifurcation and Chaos* 3(5): 1319–1325.

Mainieri R and Rehacek J (1999) Projective synchronization in three-dimensional chaotic systems. *Physical Review Letters* 82: 3042–3045.

Ott E, Grebogi C, and Yorke JA (1990) Controlling chaos. *Physical Review Letters* 64: 1196–1199.

Parlitz U and Kocarev L (1999) Synchronization of chaotic systems. In: Schuster HG (ed.) *Handbooks of Chaos Control*, pp. 271–303. Germany: Wiley-VCH.

Pecora L and Carroll T (1990) Synchronization in chaotic systems. *Physical Review Letters* 64: 821–824.

Pecora L and Carroll T (1991) Driving systems with chaotic signals. *Physical Review A* 44: 2374–2383.

Pecora L, Carroll T, Johnson G, and Mar D (1997) Fundamentals of synchronization in chaotic systems, concepts and applications. *Chaos* 7(4): 520–543.

Pikovsky A, Rosenblum M, and Kurths J (2001) *Synchronization, A Universal Concept in Nonlinear Science*. Cambridge: Cambridge University Press.

Rosenblum M, Pikovsky A, and Kurths J (1997) From phase to lag synchronization in coupled chaotic oscillators. *Physical Review Letters* 78: 4193–4196.

Rulkov N, Sushchik M, Tsimring L, and Abarbanel H (1995) Generalized synchronization of chaos in directionally coupled chaotic systems. *Physical Review E* 51(2): 980–994.

Suykens J, Vandewalle J, and Chua LO (1997) Nonlinear $H_\infty$ synchronization of chaotic Lure systems. *International Journal of Bifurcation and Chaos* 7(6): 1323–1335.

Wu CW (2002) Synchronization in Coupled Chaotic Circuits and Systems. *Series on Nonlinear Science, Series A*, vol. 41. Singapore: World Scientific.

Yamada T and Fujisaka H (1983) Stability theory of synchronized motion in coupled-oscillator systems. *Progress of Theoretical Physics* 70: 1240–1248.

# T

*t* **Hooft–Polyakov Monopoles** *see* Solitons and Other Extended Field Configurations

## Thermal Quantum Field Theory

**C D Jäkel**, Ludwig-Maximilians-Universität München, München, Germany



### Introduction

Quantum field theory was initially invented in order to describe high-energy elementary particles, thereby unifying quantum mechanics and special relativity. In other words, quantum field theory was addressed to the so-called vacuum sector, that is, roughly speaking physics at zero temperature and zero particle density.

The same applies to the various mathematically rigorous versions of quantum field theory that have been developed since the mid-1950s. Indeed, in Wightman's axiomatic setting, quantum field theory is describes in terms of a set of the so-called vacuum expectation values. The "algebraic approach" to quantum field theory developed by Araki, Haag, Kastler, and their collaborators is more flexible in nature. In fact, right from the beginning, the new algebraic tools were successfully applied to lattice models and other nonrelativistic systems with infinitely many degrees of freedom (see *Operator algebras and quantum statistical mechanics* by O Bratteli and D W Robinson). But the need to treat large systems of relativistic particles was apparently not felt. Even in Haag's recent monograph, *Local Quantum Physics*, the subjects of algebraic quantum field theory and algebraic quantum statistical mechanics are treated separately.

It is remarkable that constructive field theory was ahead of its time in this respect. The famous $P(\phi)_2$ model (first constructed by Glimm and Jaffe) was adapted to thermal states by Høegh-Krohn as early as 1974 (see Høegh-Krohn (1974)). His paper was properly named "Relativistic quantum statistical mechanics in two-dimensional space-time," but only recently has it received proper attention.

At the same time, around 1974, cosmology and heavy-ion collisions drew the interest of phyiscists towards the quantum statistical mechanics of hot relativistic quantum systems. Well-known papers from this early stage include those by Weinberg, Bernard, and Dolan and Jackiw. While most of the papers used Euclidean path integrals, Umezawa and his school developed a real-time framework called "thermo-field dynamics," which involved a doubling of the degrees of freedom. The excellent review by Landsman and van Weert (1987) covers these early attempts; it also explains the basic connection to the algebraic approach.

In the following years, it became evident that statistical mechanics (in its standard formulation) is barely sufficient to derive the properties of bulk matter from the underlying microscopic description provided by quantum field theory. Thus, various people began to establish mathematically rigorous foundations for the description of thermal field theory. The most successful approach was launched by D Buchholz (with various collaborators), who, from about 1985 onwards, started applying the KMS condition (which describes a thermal equilibrium state in the operator-algebraic framework of local quantum physics) to relativistic quantum field theory. In 1994, Buchholz and Bros managed to integrate the holomorphic structure of Wightman field theory into Haag's operator-algebraic framework, which led them to the notion of a relativistic KMS condition.

The advanced mathematical concepts involved in the formulation of entropy densities for thermal quantum fields (see Narnhofer (1994)) do not allow us to present this topic. The reader is referred to the excellent book *Quantum Entropy and Its Use* by M Ohya and D Petz for an introduction to the subject. A discussion of the so-called thermalization

effects that occur as a result of a curved spacetime is provided in Quantum Field Theory in Curved Spacetime. Another subject, which is missing almost completely, is perturbation theory. This subject has been covered extensively in three well-known textbooks by Kapusta, Le Bellac, and Umezawa.

### Observables and States

Following Heisenberg, we start from the basic assumption that quantum theory can be formulated in terms of observables which form an algebra $\mathcal{A}$, that is, a vector space with a (noncommutative) multiplication law. Although our emphasis on the abstract algebraic structure may look strange, there is a profound reason for starting out with an abstract algebra of observables: as soon as one considers systems with infinitely many degrees of freedom, one encounters a possibility to realize the abstract elements of the algebra $\mathcal{A}$ as operators on a Hilbert space in various inequivalent ways. The famous equivalence between the Heisenberg and the Schrödinger picture simply breaks down. States which are macroscopically different (e.g., thermal equilibrium states for different temperatures) give rise – in a natural way, which will be discussed in the sequel – to unitarily inequivalent representations of the abstract algebra of observables $\mathcal{A}$, while states which only differ microscopically can be accommodated by density matrices within the same Hilbert space. In other words, a physical state is described macroscopically by specifying a representation, and microscopically by a density matrix in this representation.

In a Lagrangian approach, the algebra of observables $\mathcal{A}$ may be thought of as being generated by the underlying fields, currents, etc. This leads to the so-called polynomial algebras. It is mathematically convenient to assume that $\mathcal{A}$ is an algebra of bounded operators, generated by the bounded functions of the underlying quantum fields. If $\phi(x)$ is any such field and if $f \in \mathcal{S}(\mathbb{R}^{d+1})$ is any real test function with support in a bounded region of spacetime, then the corresponding operator

$$W(f) = \exp\left(\mathrm{i}\int \mathrm{d}x\, f(x)\phi(x)\right)$$

would be a typical element of $\mathcal{A}$. The set of operators $\{W(f) \mid \operatorname{supp} f \subset \mathcal{O}\}$ will generate a subalgebra $\mathcal{A}(\mathcal{O})$ of $\mathcal{A}$. The underlying fields can be recovered by taking (functional) derivatives, once a representation of $\mathcal{A}$ on a Hilbert space is specified.

The spacetime symmetry of Minkowski space manifests itself in the existence of a representation

$$\alpha : (\Lambda, x) \mapsto \alpha_{\Lambda,x} \in \mathrm{Aut}(\mathcal{A}), \quad (\Lambda, x) \in \mathcal{P}_+^{\uparrow}$$

of the (orthochronous) Poincaré group $\mathcal{P}_+^{\uparrow}$. Here $\alpha_{\Lambda,x}$ is an automorphism of $\mathcal{A}$, that is, a mapping from $\mathcal{A}$ to $\mathcal{A}$ which preserves the algebraic structure. Once a Lorentz frame is fixed by choosing a timelike vector $e \in V_+$, the time evolution $t \mapsto \alpha_{1,te}$ will be denoted by $t \mapsto \tau_t$.

For the free field, the group of automorphisms $(\Lambda, x) \mapsto \alpha_{\Lambda,x}$ is defined by

$$\alpha_{\Lambda,x}(W(f)) := W\big(f(\Lambda^{-1}(. - x))\big)$$

As before, $f \in \mathcal{S}(\mathbb{R}^{d+1})$ is a Schwarz function over the Minkowski space $\mathbb{R}^{d+1}$.

While the invariance of the equations of motion is reflected in the existence of a representation of the Poincaré group in terms of automorphisms in the Heisenberg picture, at least the invariance with respect to Lorentz boosts is spontaneously broken in the Schrödinger picture for a thermal equilibrium state.

The usual notions of vector states and density matrices associated with a given Hilbert space (usually Fock space) are *a priori* not general enough to cover all cases of interest in thermal field theory. The following algebraic definition of a state substantially generalizes the notion of a state: A state $\omega$ is a positive, linear, and normalized functional, that is, a linear map $\omega : \mathcal{A} \to \mathbb{C}$ such that

$$\omega(a^*a) \geq 0 \quad \text{and} \quad \omega(\mathbb{1}) = 1$$

Once a state $\omega$ is distinguished on physical grounds, the GNS reconstruction theorem provides a Hilbert space $\mathcal{H}_\omega$ and a representation $\pi_\omega$ of $\mathcal{A}$, that is, a map from $\mathcal{A}$ to the set of bounded operators $\mathcal{B}(\mathcal{H}_\omega)$, which preserves the algebraic relations.

It is instructive to consider the GNS representation of the Pauli matrices $\{\sigma_0 = \mathbb{1}, \sigma_1, \sigma_2, \sigma_3\}$. Given a state (a diagonal $2 \times 2$ matrix $\rho$ with positive entries and $\operatorname{tr} \rho = 1$), the left regular representation (a construction well known from group theory)

$$\pi(\sigma_i)|\sqrt{\rho}> = |\sigma_i\sqrt{\rho}>, \quad i = 0, 1, 2, 3$$

defines a reducible representation on $\mathbb{C}^4$, unless one of the entries in the diagonal of $\rho$ is zero (which corresponds to a pure state). In the latter case, the GNS Hilbert space is $\mathbb{C}^2$. By construction, $<\sqrt{\rho}|\pi(\sigma_i)|\sqrt{\rho}> = \operatorname{tr} \rho\sigma_i$, $i = 1, 2, 3$.

## Thermal Equilibrium

The variety of nonequilibrium states ranges from mild perturbations of equilibrium states through steady states, whose properties are governed by external heat baths, or hydrodynamic flows up to totally chaotic states which no longer

admit a description in terms of thermodynamic notions. Buchholz *et al.* (2002) have initiated an investigation of nonequilibrium states that are locally (but not globally) close to thermal equilibrium. Unfortunately, we will not be able to cover this topic. Instead, we will concentrate on states which deviate from a true equilibrium state only microscopically.

### Characterization of Thermal Equilibrium States

When the time evolution $t \mapsto \tau_t \in \mathrm{Aut}(\mathcal{A})$ is changed by a local perturbation, which is slowly switched on and slowly switched off again, then an equilibrium state $\omega$ returns to its original form at the end of this procedure. This heuristic condition of adiabatic invariance can be expressed by the stability requirement

$$\lim_{t\to\infty} \int_{-t}^{t} \mathrm{d}t \; \omega([a, \tau_t(b)]) = 0 \quad \forall a, b \in \mathcal{A} \qquad [1]$$

In a pioniering work Haag, Kastler, and Trych-Pohlmeyer showed that the characterization [1] of an equilibrium state leads to a sharp mathematical criterion, first encountered by Haag, Hugenholtz, and Winnink and more implicitly by Kubo, Martin, and Schwinger:

**Definition 1** A state $\omega_\beta$ over $\mathcal{A}$ is called a KMS state for some $\beta > 0$, if for all $a, b \in \mathcal{A}$, there exists a function $F_{a,b}$ which is continuous in the strip $0 \leq \Im z \leq \beta$ and analytic and bounded in the open strip $0 < \Im z < \beta$, with boundary values given by

$$F_{a,b}(t) = \omega_\beta(a\tau_t(b)) \quad \text{and} \quad F_{a,b}(t + \mathrm{i}\beta) = \omega_\beta(\tau_t(b)a) \quad \forall t \in \mathbb{R} \qquad [2]$$

Before we start analyzing the properties of KMS states, we should mention an alternative characterization of thermal equilibrium states: passivity. The amount of work a cycle can perform when applied to a moving thermodynamic equilibrium state is bounded by the amount of work an ideal windmill or turbine could perform; this property is called semipassivity (Kuckert 2002): a state $\omega$ is called semipassive (passive) if there is an "efficiency bound" $E \geq 0$ ($E = 0$) such that

$$-(W\Omega_\omega, H_\omega W\Omega_\omega) \leq E \cdot (W\Omega_\omega, |P_\omega| W\Omega_\omega) \quad \forall W \in \pi_\omega(\mathcal{A})''$$

with $W^{-1} = W^*, [H_\omega, W] \in \pi_\omega(\mathcal{A})''$, and $[P_\omega, W] \in \pi_\omega(\mathcal{A})''$. Here $(H_\omega, P_\omega)$ denote the generators implementing the spacetime translations in the GNS representation $(\mathcal{H}_\omega, \Omega_\omega, \pi_\omega)$. Generalizing the notion of complete passivity, the state $\omega$ is called completely semipassive if all its finite tensorial powers are semipassive with respect to one fixed efficiency bound $E$. It has been shown by Kuckert (2002) that a state is completely semipassive in all inertial frames if and only if it is completely passive in some inertial frame. The latter implies that $\omega$ is a KMS state or a ground state (a result due to Pusz and Woronowicz).

Let us now turn to properties of thermal equilibrium states which are specific for relativistic models. It was first recognized by Bros and Buchholz (1994) that KMS states of a relativistic theory have stronger analyticity properties in configuration space than those imposed by the traditional KMS condition:

**Definition 2** A KMS state $\omega_\beta$ satisfies the relativistic KMS condition (Bros and Buchholz 1994) if there exists a unit vector $e$ in the forward light cone $V_+$ such that for every pair of local elements $a$, $b$ of $\mathcal{A}$ the function $F_{a,b}$

$$F_{a,b}(x_1, x_2) = \omega_\beta(\alpha_{x_1}(a)\alpha_{x_2}(b))$$

extends to an analytic function in the tube domain $-\mathcal{T}_{\beta e/2} \times \mathcal{T}_{\beta e/2}$, where $\mathcal{T}_{\beta e/2} = \{z \in \mathbb{C} \mid \Im z \in V_+ \cap (\beta e/2 - V_+)\}$.

The relativistic KMS condition can be understood as a remnant of the relativistic spectrum condition in the vacuum sector. It has been rigorously established (Bros and Bruchholz 1994) for the KMS states constructed by Buchholz and Junglas (1989) and by C Gérard and the author for the $P(\phi)_2$ model. In the thermal Wightman framework (Bros and Buchholz 1996) it has been shown that the relativistic KMS condition implies existence of model-independent analyticity properties of thermal $n$-point functions. These properties also appear in perturbative computations of the thermal Wightman functions (Steinmann 1995).

We now turn to the properties of the set of KMS states. For given $\beta$, the convex set $\mathcal{S}_\beta$ of all KMS states is known to form a simplex; the extreme points in the set $\mathcal{S}_\beta$ are called extremal KMS states. As a consequence, the extremal states in $\mathcal{S}_\beta$ can be distinguished with the help of "classical" (central) observables, that is, by observables which commute with all other observables.

If $\omega$ is an extremal KMS state and $\gamma$ is an automorphism which commutes with the time evolution $t \mapsto \tau_t$, then the state $\omega'$ defined by

$$\omega'(a) := \omega(\gamma(a)), \quad a \in \mathcal{A}$$

is again an extremal KMS state to the same parameter values. If $\omega' \neq \omega$, one says that the symmetry is spontaneously broken.

Lorentz invariance with respect to boosts is always broken by a KMS state, since the KMS condition distinguishes a rest frame. A KMS state might also break spatial translation or rotation invariance. However, by averaging over the different configurations one can usually construct a translation- and rotation-invariant state. The situation is drastically different with respect to supersymmetry. Buchholz and Ojima (1997) have shown that supersymmetry is broken in any thermal state and it is impossible to proceed from it by "symmetrization" to states on which an action of supercharges can be defined.

### Existence of Thermal Equilibrium States

Buchholz and Junglas (1989) demonstrated that the existence of KMS states can be guaranteed for a large class of quantum field-theoretic models. The basic assumption to be met concerns the phase-space properties of the model. A generalized trace norm (the so-called "nuclear norm") is used to estimate the "number" of degrees of freedom in phase space.

The first step is to construction a subspace $\mathcal{H}(\Lambda)$ of the vacuum Hilbert space $\mathcal{H}_{\text{vac.}}$, which represents excitations of the vacuum strictly localized inside of a bounded spacetime region $\hat{\mathcal{O}}$. Due to the strong correlations present in the vacuum state of any relativistic model, as a consequence of the Reeh–Schlieder property (see the section "Analyticity of n-point functions") this is a delicate procedure, which involves the so-called "split property." This property ensures that there exists a product vector $\eta$ in vacuum Hilbert space $\mathcal{H}_{\text{vac.}}$ such that

$$(\eta, \pi_{\text{vac.}}(ab)\eta) = \omega_{\text{vac.}}(a)\cdot\omega_{\text{vac.}}(b) \quad \forall a \in \mathcal{A}(\mathcal{O}),\ b \in \mathcal{A}(\hat{\mathcal{O}})^{\text{c}} \qquad [3]$$

Here $\mathcal{O} \subset \hat{\mathcal{O}}$ denotes a slightly smaller open spacetime region (such that the closure $\overline{\mathcal{O}}$ is inside the interior of $\hat{\mathcal{O}}$) and $\mathcal{A}(\hat{\mathcal{O}})^{\text{c}} := \{A \in \mathcal{A} \mid [A, B] = 0\ \forall B \in \mathcal{A}(\hat{\mathcal{O}})\}$. The existence of a product vector can be ensured if the nuclear norm satisfies some mild bounds which are expected to hold in all models of physical interest. Given a product vector $\eta$ which satisfies [3], the sought after subspace is

$$\mathcal{H}(\Lambda) := \overline{\pi_{\text{vac.}}(\mathcal{A}(\mathcal{O}))''\Omega_{\text{vac.}}}$$

The crucial step in the proof of existence of KMS states is to show that

$$\operatorname{tr} E(\Lambda)\mathrm{e}^{-\beta H}E(\Lambda) < \infty \quad \text{for } \beta > 0$$

if the nuclearity condition holds. Here $E(\Lambda)$ denotes the projection onto the subspace $\mathcal{H}(\Lambda)$ representing localized excitations and $H$ denotes the Hamiltonian in the vacuum representation $\pi_{\text{vac.}}$. Next it is shown that the function

$$t \mapsto \omega_{\beta,\Lambda}(a\tau_t(b)) = \frac{1}{Z}\operatorname{tr} E(\Lambda)\mathrm{e}^{-\beta H}E(\Lambda)\pi_{\text{vac.}}(a\tau_t(b))$$

allows an analytic extension to a strip of width $\beta$ which satisfies the KMS boundary condition [2] for $|t| < \delta$ if $a, b \in \mathcal{A}(\mathcal{O}_\circ)$ and $\mathcal{O}_\circ + te \subset \mathcal{O}$ for $|t| < \delta$. In the final step, Buchholz and Junglas were able to demonstrate that bounds on the nuclear norm are even sufficient to control the thermodynamic limit.

Given a thermal field theory, a slight variation of the method used by Buchholz and Junglas allows one to construct a KMS state for a new temperature (Jäkel 2004), that is, to change the temperature of a thermal state.

### Thermal Representations

Given a KMS state $\omega_\beta$, the GNS construction gives rise to a Hilbert space $\mathcal{H}_\beta$ and a representation $\pi_\beta$, called a thermal representation, of $\mathcal{A}$. The algebra $\mathcal{R}_\beta := \pi_\beta(\mathcal{A})''$ possesses a cyclic (due to the GNS construction) and separating (due to the KMS condition) vector $\Omega_\beta$ such that

$$\omega_\beta(a) = \left(\Omega_\beta, \pi_\beta(a)\Omega_\beta\right) \quad \forall a \in \mathcal{A}$$

The KMS condition implies that $\omega_\beta$ is invariant under time translations, that is, $\omega_\beta \circ \tau_t = \omega_\beta$ for all $t \in \mathbb{R}$. Thus,

$$U(t)\pi_\beta(a)\Omega_\beta = \pi_\beta(\tau_t(a))\Omega_\beta, \quad a \in \mathcal{A}$$

defines a strongly continuous unitary group $\{U(t)\}_{t\in\mathbb{R}}$ implementing the time evolution in the representation $\pi_\beta$. By Stone's theorem there exists a self-adjoint generator $L$ such that

$$U(t) = \mathrm{e}^{\mathrm{i}Lt}, \quad t \in \mathbb{R} \qquad [4]$$

For $0 \le \beta < \infty$, the Liouville operator $L$ is not bounded from below; its spectrum is symmetric and consists typically of the whole real line. However, the negative part of $L$ is "suppressed" with respect to the algebra of observables $\mathcal{R}_\beta := \pi_\beta(\mathcal{A})''$ in the following sense (Haag 1992): let $\mathbb{1}_{]-\infty,-\kappa]}$ be the spectral projection of $L$ for the interval $]-\infty, -\kappa] \subset \mathrm{Sp}(L)$, then

$$\|\mathbb{1}_{]-\infty,-\kappa]}A\Omega_\beta\| \le \mathrm{e}^{-\beta\kappa}\|A\| \quad \forall A \in \mathcal{R}_\beta$$

We now turn to structural aspects which are characteristic for a relativistic model, namely the existence of strong spatial correlations and the connection between the decay of these correlations and the spectral properties of the Liouville operator.

Let $\omega_\beta$ be a state, which satisfies the relativistic KMS condition. It follows (using a theorem of Glaser) that for $a \in \mathcal{A}$ the function $\Phi_a : \mathbb{R}^4 \to \mathcal{H}_\beta$,

$$x \mapsto \pi_\beta(\alpha_x(a))\Omega_\beta$$

can be analytically continued from the real axis into the domain $\mathcal{T}_{\beta e/2}$ such that it is weakly continuous for $\Im z \searrow 0$. If the usual additivity assumption $\cup_i \mathcal{O}_i = \mathcal{O} \Rightarrow \vee_i \mathcal{R}_\beta(\mathcal{O}_i) = \mathcal{R}_\beta(\mathcal{O})$ for the local von Neumann algebras holds, then

$$\mathcal{H}_\beta = \overline{\pi_\beta(\mathcal{A}(\mathcal{O}))\Omega_\beta} \qquad [5]$$

for any open spacetime region $\mathcal{O} \subset \mathbb{R}^{d+1}$. Junglas has shown that the thermal Reeh–Schlieder property [5] follows as well from the standard KMS condition, if $\omega_\beta$ is locally normal with respect to the vacuum representation.

The decay of spatial correlations depends on infrared properties of the model, and the essential ingredients for the following cluster theorem are the continuity properties of the spectrum of $L$ near zero.

**Theorem 3** *Let $\Omega_\beta$ denote the unique (up to a phase) normalized eigenvector with eigenvalue $\{0\}$ of the Liouvillean $L$ and let $P^+$ denote the projection onto the strictly positive part of the spectrum of $L$. Assume that there exist positive constants $m > 0$ and $C_1(\mathcal{O}) > 0$ such that*

$$\|e^{-\lambda L} P^+ \pi_\beta(a)\Omega_\beta\| \le C_1(\mathcal{O}) \cdot \lambda^{-m}\|a\| \quad \forall a \in \mathcal{A}(\mathcal{O})$$

*Here $\mathcal{O} \subset \mathbb{R}^{d+1}$ is an open and bounded spacetime region. Now consider two spacelike separated spacetime regions $\mathcal{O}_1, \mathcal{O}_2$, which can be embedded into $\mathcal{O}$ by translation and such that $\mathcal{O}_1 + \delta e \subset \mathcal{O}_2', \delta >> \beta$. then, for $a \in \mathcal{A}(\mathcal{O}_1)$ and $b \in \mathcal{A}(\mathcal{O}_2)$,*

$$|\omega_\beta(ba) - \omega_\beta(b)\omega_\beta(a)| \le C_2 \cdot \delta^{-2m}\|a\|\,\|b\|$$

*The constant $C_2(\beta, \mathcal{O}) \in \mathbb{R}^+$ may depend on the temperature $\beta^{-1}$ and the size of the region $\mathcal{O}$ but is independent of $\delta, a$, and $b$.*

From explicit calculations one expects that $m = 1/2$ for free massless bosons in $3+1$ spacetime dimensions. Consequently, the exponent given on the right-hand side is optimal since it is well known that in this case the correlations decay only like $\delta^{-1}$.

A description of thermal representations would be inadequate without pointing out one of the deepest connections between pure mathematics and physics that emerged in the last century: consider a von Neumann algebra $\mathcal{R}$ which possesses a cyclic and separating vector $\Omega$. Then polar decomposition of the closeable operator $S : A\Omega \mapsto A^*\Omega$, $A \in \mathcal{R}$, provides an antiunitary operator $J$ (the modular conjugation) and a self-adjoint operator $\Delta^{1/2}$. The connection to physics was established independently by Takesaki and Winnink, showing that the pair $(\mathcal{R}, \sigma)$ satisfies the KMS condition for $\beta = -1$, if one sets $\sigma_t(A) = \Delta^{it} A \Delta^{-it}$ for $A \in \mathcal{R}$.

Taking advantage of the Reeh–Schlieder property [5], one can associate modular objects to certain spacetime regions $\mathcal{O}$. In general, a physical interpretation of these modular objects is missing. But for two-dimensional thermal models, which factorize in light-cone coordinates, the modular group corresponding to the algebra of a spacelike wedge admits a simple description: at large distances (compared to $\beta$) from the boundary, the flow pattern is essentially the same as time translations. These are results due to Borchers and Yngvason (1999).

### Analyticity Properties of *n*-Point Functions

The correlation functions describe the full physical content of the theory: all observable quantities can in principle be derived from them. This is so because according to the Wightman reconstruction theorem (which is closely related to the GNS construction) knowledge of the correlation functions allows the reconstruction of the full representation of the field algebra. The Wightman distributions $\{\mathcal{W}_\beta^{(n)}\}_{n\in\mathbb{N}}$,

$$\mathcal{W}_\beta^{(n)}(t_2 - t_1, x_2 - x_1, \ldots, t_n - t_{n-1}, x_n - x_{n-1}) := (\Omega_\beta, \phi_\beta(t_1, x_1) \cdots \phi_\beta(t_n, x_n)\Omega_\beta) \qquad [6]$$

where $\pi_\beta(W(f)) =: \exp(i \int dt\, dx\, f(t,x)\phi_\beta(t,x))$, satisfy a number of key properties: locality, positivity, Poincaré covariance, and temperedness. These properties have been formulated for thermal field by Bros and Buchholz (1996), and this section is entirely based on their work.

The relativistic KMS condition implies that the Wightman distributions $\{\mathcal{W}_\beta^{(n)}\}_{n\in\mathbb{N}}$ of a translation-invariant equilibrium state admit in the corresponding set of spacetime variables $(t_2 - t_1, x_2 - x_1), \ldots, (t_n - t_{n-1}, x_n - x_{n-1})$ an analytic continuation into the union of domains

$$(\alpha_1 \mathcal{T}_{\beta e}) \times \cdots \times (\alpha_{n-1} \mathcal{T}_{\beta e})$$

for $\alpha_i > 0$, $i = 1, \ldots, n-1$ and $\sum_{i=1}^{n-1} \alpha_i = 1$. The tube domains $\mathcal{T}_{\beta e}$ were specified in Definition 2. For $\beta \to \infty$, the tube $\mathcal{T}_{\beta e}$ tends to the vacuum tube $\mathcal{T}_{\text{vac.}} = \mathbb{R}^{d+1} + iV_+$; thus, one recovers the spectrum condition for the vacuum expectation values.

Let us now turn to the Fourier transformed Wightman correlation functions. Translation invariance implies

$$\widetilde{\mathcal{W}}_\beta^{(n)}(\nu_1, p_1, \ldots, \nu_n, p_n)\delta(\nu_1 + \cdots + \nu_n)\delta(p_1 + \cdots + p_n)$$

The Wightman distribution $\widetilde{\mathcal{W}}_\beta^{(n)}$ satisfies on the linear manifold $(\nu_1, p_1) + \cdots + (\nu_n, p_n) = 0$ the KMS relation in the energy variables: for any pair of multi-indices $(I, J)$ the identity

$$\widetilde{\mathcal{W}}_\beta^{(n)}(J, I) = \mathrm{e}^{-\beta\nu_I}\widetilde{\mathcal{W}}_\beta^{(n)}(I, J)$$

holds, where $\widetilde{\mathcal{W}}_\beta^{(n)}(J, I)$ is an abbreviation for $\widetilde{\mathcal{W}}_\beta^{(n)}(\{p_i\}_{i\in I}, \{p_j\}_{j\in J})$ and $\nu_I = \sum_{i\in I}\nu_i$.

We now specialize to the two-point function $\mathcal{W}_\beta^{(2)}$. The corresponding commutation function $\mathcal{C}(x)$ is given by

$$\mathcal{C}(x_1 - x_2) = \mathcal{W}_\beta^{(n)}(x_1, x_2) - \mathcal{W}_\beta^{(2)}(x_2, x_1)$$

Locality implies that $\operatorname{supp}\mathcal{C} \subset \overline{V_+} \cup \overline{V_-}$. The retarded and the advanced propagator $r$ and $a$, formally given by

$$r(x) = \mathrm{i}\theta(x_\circ)\mathcal{C}(x), \quad a(x) = -\mathrm{i}\theta(-x_\circ)\mathcal{C}(x)$$

satisfy the relation

$$r - a = -\mathrm{i}\mathcal{C}$$

which corresponds to a partition of the support of $\mathcal{C}$ in its convex components: $\operatorname{supp} r \subset \overline{V_+}$ and $\operatorname{supp} a \subset \overline{V_-}$. For the free scalar field of mass $m$ the commutator function is

$$\mathcal{C}^{(m)}(x) = \frac{1}{(2\pi)^2}\int_{\mathbb{R}^4} \mathrm{d}p\, \mathrm{e}^{-\mathrm{i}px}\tilde{\mathcal{C}}^{(m)}(p)$$

with

$$\tilde{\mathcal{C}}^{(m)}(p) = \frac{1}{2\pi}\operatorname{sgn}(\nu)\delta(\nu^2 - \boldsymbol{p}^2 - m^2)$$

and subsequently the retarded and advanced propagators $r^{(m)}$ and $a^{(m)}$ are structural functions of the field algebra, which are determined by the c-number commutation relations of the fields. Thus, they are independent of the temperature, in contrast to the two-point function:

$$\tilde{\mathcal{W}}_\beta^{(2)}(p) = \frac{\tilde{\mathcal{C}}^{(m)}(p)}{1 - \mathrm{e}^{-\beta\nu}} \qquad [7]$$

Let now $\tilde{\tau}(p)$ be the Fourier transform of the time-ordered function $\tau(x)$. The relation

$$\tilde{\tau}(p) = \frac{-\mathrm{i}\tilde{r}(p) + -\mathrm{i}\tilde{a}(p)\mathrm{e}^{-\beta\nu}}{1 - \mathrm{e}^{-\beta\nu}}$$

shows that $\tilde{\tau}(p)$ and $-\mathrm{i}\tilde{r}(p)$ only "coincide up to an exponential tail" at very high energies (Bros and Buchholz 1996).

### Particle Aspects

The condition of locality (together with the relativistic KMS condition) leads to strong constraints on the general form of the thermal two-point functions that allow one to apply the techniques of the Jost–Lehmann–Dyson representation. As has been shown by Bros and Buchholz (1996), the interacting two-point function $\mathcal{W}_\beta$ can be represented in the form

$$\mathcal{W}_\beta(t, \boldsymbol{x}) = \int_0^\infty \mathrm{d}m\, \mathcal{D}_\beta(\boldsymbol{x}, m)\mathcal{W}_\beta^{(2)}(t, \boldsymbol{x}, m)$$

Here $\mathcal{D}_\beta(\boldsymbol{x}, m)$ is a distribution in $\boldsymbol{x}, m$ which is symmetric in $\boldsymbol{x}$, and

$$\mathcal{W}_\beta^{(2)}(t, \boldsymbol{x}, m) = (2\pi)^{-1}\int \mathrm{d}\nu\mathrm{d}\boldsymbol{p}\ \mathrm{e}^{\mathrm{i}(\nu t - \boldsymbol{p}\boldsymbol{x})}\tilde{\mathcal{W}}_\beta^{(2)}(\nu, \boldsymbol{p})$$

is the two-point correlation function of the free thermal field of mass $m$. In contrast to the vacuum case, the damping factors $\mathcal{D}_\beta(\boldsymbol{x}, m)$ depend in a nontrivial way on the spatial variables $\boldsymbol{x}$. The damping factors describe the dissipative effects of the thermal system on the propagation of sharply localized excitations. Bros and Buchholz suggested that the damping factor $\mathcal{D}_\beta(\boldsymbol{x}, m)$ can be decomposed into a discrete and an absolute continuous part

$$\mathcal{D}_\beta(\boldsymbol{x}, m) = \delta(m - m_0)\mathcal{D}_{\beta,\mathrm{d}}(\boldsymbol{x}) + \mathcal{D}_{\beta,\mathrm{c}}(\boldsymbol{x}, m)$$

and that the $\delta$-contribution in the damping factors is due to stable constituent particles of mass $m_0$ out of which the thermal states are formed, whereas the collective quasiparticle-like excitations only contribute to the continuous part of the damping factors (Bros and Buchholz 1996).

In the case of spontaneously broken internal symmetries Bros and Buchholz (1998) have shown that the damping factors $\mathcal{D}_\beta^{\pm}(\boldsymbol{x}, m)$ which appear in the representation of current-field correlations functions

$$(\Omega_\beta, j_0(t, \boldsymbol{x})\phi_\beta(0, 0)\Omega_\beta) = \int_0^\infty \mathrm{d}m\Big(\mathcal{D}_\beta^+(\boldsymbol{x}, m)\partial_t\mathcal{W}_\beta^{(2)}(t, \boldsymbol{x}, m) + \mathcal{D}_\beta^-(\boldsymbol{x}, m)\mathcal{W}_\beta^{(2)}(t, \boldsymbol{x}, m)\Big)$$

indeed contain a discrete (in the sense of measures) zero-mass contribution and are slowly decreasing in $|\boldsymbol{x}|$ for small values of $m$. Thus, these damping factors coincide locally with the Källén–Lehmann weights appearing in the case of spontaneous symmetry breaking in the vacuum sector (Bros and Buchholz 1998). It is easily seen in examples that there is no sharp energy–momentum dispersion law for the Goldstone particles. Thus, the Källén–Lehmann representation is better suited than Fourier transformation to uncover the particle aspects of thermal equilibrium states.

## Models of Thermal Field Theory

In the simplest case, the classical Lagrangian density of the so-called $P(\phi)_2$ models is given by

$$\mathcal{L} = (\partial_\nu \phi)(\partial^\nu \phi) - m^2\phi^2 - \frac{\lambda}{4}\phi^4 \qquad [8]$$

Here $\phi(t,x)$ denotes a real scalar field over spacetime. The construction of the corresponding quantized thermal field presented in this section (Gérard and Jäkel 2005) is based on the original ideas of Høegh-Krohn (1974).

### Free Fields

Let $\mathfrak{h}_m$ denote the $L^2$-closure of $C_0^\infty(\mathbb{R})$ with respect to the norm $\|f\| = (f, (1/2\epsilon)f)$, where $\epsilon(k) = \sqrt{k^2+m^2}$ denotes the one-particle energy for a single neutral scalar boson and the scalar product is the usual $L^2$-scalar product. The subspaces associated to a double cone $\mathcal{O}$ are given by

$$\mathfrak{h}_m(\mathcal{O}) := \{h \in \mathfrak{h}_m | \operatorname{supp} \Re h, \operatorname{supp} \nu^{-1}\Im h \subset O\}$$

where $O$ denotes the basis of the double cone $\mathcal{O}$. The corresponding free quantum field is described by the Weyl algebra $\mathcal{W}(\mathfrak{h}_m) := \{W(f) \,|\, f \in \mathfrak{h}_m\}$, together with the time evolution $\{\tau_t^\circ\}_{t\in\mathbb{R}}$,

$$\tau_t^\circ(W(f)) = W(\mathrm{e}^{\mathrm{i}t\epsilon}f), \quad f \in \mathfrak{h}_m$$

If $m > 0$, the KMS condition allows just one unique (quasifree) $(\tau^\circ, \beta)$-KMS state:

$$\omega_\beta^\circ(W(f)) := \mathrm{e}^{-(1/4)(f,(1+2\rho)f)_m}, \quad \rho := (\mathrm{e}^{\beta\epsilon} - 1)^{-1}$$

The GNS representation associated to the pair $(\mathcal{W}(\mathfrak{h}_m), \omega_\beta^\circ)$ is the well-known Araki–Woods representation, given by

$$\mathcal{H}_{\mathrm{AW}} := \Gamma\big(\mathfrak{h}_m \oplus \overline{\mathfrak{h}_m}\big), \quad \Omega_{\mathrm{AW}} := \Omega_{\mathcal{F}},$$
$$\pi_{\mathrm{AW}}(W(h)) := W_{\mathcal{F}}\Big((1+\rho)^{1/2}h \oplus \overline{\rho}^{1/2}\overline{h}\Big), \quad h \in \mathfrak{h}_m$$

Here $\overline{\mathfrak{h}_m}$ is the Hilbert space conjugate to $\mathfrak{h}_m$, $W_{\mathcal{F}}(.)$ denotes the usual Weyl operator on the Fock space $\Gamma(\mathfrak{h}_m \oplus \overline{\mathfrak{h}_m})$ and $\Omega_{\mathcal{F}} \in \Gamma(\mathfrak{h}_m \oplus \overline{\mathfrak{h}_m})$ is the Fock vacuum. The Liouvillean $L_{\mathrm{AW}}$ (see [4]) can be identified with $d\Gamma(\epsilon \oplus -\epsilon)$.

The local von Neumann algebra generated by $\{\pi_{\mathrm{AW}}(W(h)) \,|\, h \in \mathfrak{h}_m(\mathcal{O})\}$ is denoted by $\mathcal{R}_{\mathrm{AW}}(\mathcal{O})$. The algebra of observables for the free quantum field (and, as we will see, the $P(\phi)_2$ model) is the norm closure

$$\mathcal{A} := \overline{\bigcup_{\mathcal{O}\subset\mathbb{R}^2} \mathcal{R}_{\mathrm{AW}}(\mathcal{O})}^{C^*}$$

of the local von Neumann algebras.

### The Thermal $P(\phi)_2$ Model

In $1+1$ spacetime dimensions Wick ordering is sufficient to eliminate the UV divergences of polynomial interactions. As it turns out, the leading order in the UV divergences is independent of the temperature (in agreement with the results found in Kopper *et al.* (2001)). Thus, it is a matter of convenience whether one uses the thermal covariance function $C_\beta$,

$$C_\beta(h_1,h_2) := \left(h_1, \frac{(1+\mathrm{e}^{-\beta\epsilon})}{2\epsilon(1-\mathrm{e}^{-\beta\epsilon})}h_2\right)_{L^2(\mathbb{R})}$$
$$h_1, h_2 \in \mathcal{S}(\mathbb{R})$$

or the vacuum covariance function $C_{\mathrm{vac.}}$ to define the Wick ordering:

$$:\phi(f)^n:_C = \sum_{m=0}^{[n/2]} \frac{n!}{m!(n-2m!)}\phi(f)^{n-2m}\left(-\frac{1}{2}C(f,f)\right)^m$$

Now let $P(\lambda)$ be a real-valued polynomial, which is bounded from below. Then Euclidean techniques can be used to define the operator sum

$$H_l := L_{\mathrm{AW}} + \int_{-l}^{l} :P(\phi_\beta(x)):_{C_\beta} \,\mathrm{d}x$$

in the Araki–Woods representation and to show that $H_l$ is essentially self-adjoint Gérard and Jäkel (2005). Thus, (the closure of) $H_l$ can be used to define a perturbed time evolution $t \mapsto \tau_t^l$ on $\mathcal{A}$ and the vector

$$\Omega_l := \frac{\mathrm{e}^{-(\beta/2)H_l}\Omega_{\mathrm{AW}}}{\|\mathrm{e}^{-(\beta/2)H_l}\Omega_{\mathrm{AW}}\|}$$

induces a KMS state $\omega_l$ for the dynamical system $(\pi_{\mathrm{AW}}(\mathcal{A})'', \tau^l)$.

A finite propagation speed argument (using Trotter's product formula) shows that

$$\tau_t^l(A) := \mathrm{e}^{\mathrm{i}tH_l}A\mathrm{e}^{-\mathrm{i}tH_l}, \quad t \in \mathbb{R} \qquad [9]$$

is independent of $l$ for $A \in \mathcal{R}_{\mathrm{AW}}(\mathcal{O}), t \in \mathbb{R}$ fixed and $l$ sufficiently large. Thus, there exists a limiting dynamics $\tau$ such that

$$\lim_{l\to\infty} \|\tau_t^l(A) - \tau_t(A)\| = 0 \qquad [10]$$

for all $A \in \mathcal{R}_{\mathrm{AW}}(\mathcal{O}), \mathcal{O}$ bounded. This norm convergence extends to the norm closure $\mathcal{A}$ of the local von Neumann algebras.

The existence of weak* limit points (which are states) of the (generalized) sequence $\{\omega_l\}_{l>0}$ is a consequence of the Banach–Alaoglu theorem. The fact that all limit states satisfy the KMS condition with respect to the pair $(\mathcal{A}, \tau)$ follows from [10]. To

prove that the sequence $\{\omega_l\}_{l>0}$ has only one accumulation point,

$$\omega_\beta = \lim_{l\to\infty} \omega_l \qquad [11]$$

is more delicate. Following Høegh-Krohn, Nelson symmetry is used in Gérard and Jäkel (2005) to relate the interacting thermal theory on the real line to the $P(\phi)_2$ model on the circle $S^1$ of length at temperature 0. The existence of the limit [11] then follows from the uniqueness of the vacuum state on the circle. The relativistic KMS condition can be derived by Nelson symmetry as well, using the fact that the discrete spectrum of the model on the circle satisfies the spectrum condition. Since the limit [11] exists on the norm closure $\mathcal{A}$ of the weakly closed local algebras, it follows from a result of Takesaki and Winnink that $\omega_\beta$ is locally normal with respect to the Araki–Woods representation (which itself is locally normal with respect to the Fock representation). Consequently,

$$\mathcal{R}_\beta(\mathcal{O}) := \pi_\beta(\mathcal{A}(\mathcal{O}))'' \cong \mathcal{R}_{\mathrm{AW}}(\mathcal{O}), \quad \mathcal{O} \text{ bounded}$$

that is, $\mathcal{R}_\beta(\mathcal{O})$ is (isomorphic to) the unique hyperfinite factor of type $\mathrm{III}_1$. Moreover, the local Fock property implies that the split property holds.

### Perturbation Theory

Steinmann (1995) has shown that perturbative expansions for the Wightman distributions of the $:\phi^4:_4$ model can be derived directly in the thermodynamic limit, using as only inputs the equations of motion and the (thermal) Wightman axioms. The result can be represented as a sum over generalized Feynman graphs.

The method consists in solving the differential equations for the correlation functions which follow from the field equation, by a power series expansion in the coupling constant, using the axiomatic properties of the Wightman functions as subsidiary conditions. The Wightman axioms are expected to hold separately in each order of perturbation theory, with the exception of the cluster property.

As expected, the UV renormalization can be chosen to be temperature independent, that is, one can use the same counterterms as in the vacuum case. But the infrared divergencies are more severe, they cannot be removed by minor adjustments of the renormalization procedure. Various elaborate resummation techniques have been proposed to (at least partially) remove the infrared singularities.

Another approach has been pursued by Kopper *et al.* (2001). They have investigated the perturbation expansion of the $:\phi^4:_4$ model in the imaginary-time formalism, using Wilson's flow equations. The result is once again that all correlation functions become ultraviolet-finite in all orders of the perturbation expansion, once the theory has been renormalized at zero temperature by usual renormalization prescriptions.

### Asymptotic Dynamics of Thermal Fields

Timelike asymptotic properties of thermal correlation functions cannot be interpreted in terms of free fields due to persistent dissipative effects of a thermal system. This well-known fact manifests itself in a softened pole structure of the Green's functions in momentum space and is at the root of the failure of the conventional approach to thermal perturbation theory (Bros and Bruchholz 2002). In fact, assuming a sharp dispersion law, one would be forced to conclude that the scattering matrix is trivial (a famous no-go theorem by Narnhofer *et al.* (1983)).

However, there seems to be a possibility to find an effective theory, which is much simpler and still reproduces the correct asymptotic behavior of the full theory. Disregarding low-energy excitations, Bros and Buchholz (2002) have shown that the $\delta$-contributions in the damping factors give rise to asymptotically leading terms which have a rather simple form: they are products of the thermal correlation function of a free field and a damping factor describing the dissipative effects of the model-dependent thermal background. This result is based on the assumption that the truncated $n$-point functions satisfy

$$\lim_{\mathrm{T}\to\infty} \mathrm{T}^{3(n-1)/2-\kappa} \mathcal{W}_\beta^{(n)}(t_1, x_1, \ldots, t_n, x_n)^{\text{trunc.}} = 0$$
$$\kappa > 0$$

while the $\delta$-contribution in the damping factors exhibit, for large timelike separations $T$, a $T^{-3/2}$ type behavior (in $3+1$ spacetime dimensions).

Bros and Buchholz (2002) have shown that the asymptotically dominating parts of the correlation functions can be interpreted in terms of quasifree states acting on the algebra generated by a Hermitian field $\phi_0$ satisfying the commutation relations

$$[\phi_0(t_1, x_1), \phi_0(t_2, x_2)] = \Delta_{m_0}(t_1 - t_2, x_1, x_2) Z(x_1 - x_2)$$

Here $\Delta_{m_0}$ is the usual commutator function of a free scalar field of mass $m_0$ and $Z$ is an operator-valued distribution commuting with $\phi_0$ such that $\hat{\omega}_\beta(Z(x_1 - x_2)) = \mathcal{D}_{\beta,\mathrm{d}}(x_1 - x_2)$. (Here $\hat{\omega}_\beta$ denotes a KMS state for the algebra generated by $\phi_0$.) Intuitively speaking, the field $\phi_0$ carries an additional stochastic degree of freedom, which manifests itself in a central element that appears in the commutation relations and couples to the thermal background.

As $\phi_0$ describes the interacting field asymptotically, one may expect that $\phi_0$ satisfies the field

equation of the interacting field in an asymptotic sense. Buchholz and Bros (2002) have demonstrated that this assumption allows one to derive an explicit expression for the discrete part of the damping factors $\mathcal{D}_{\beta,\mathrm{d}}(x)$ in simple models.

*See also:* Axiomatic Quantum Field Theory; Quantum Field Theory in Curved Spacetime; Scattering in Relativistic Quantum Field Theory: The Analytic Program; Tomita–Takesaki Modular Theory.

## Further Reading


Borchers H-J and Yngvason J (1999) Modular groups of quantum fields in thermal states. *Journal of Mathematical Physics* 40: 601–624.

Birke L and Fröhlich J (2002) KMS, etc., *Reviews in Mathematical Physics* 14: 829–871.

Bros J and Buchholz D (1994) Towards a relativistic KMS condition. *Nuclear Physics B* 429: 291–318.

Bros J and Buchholz D (1996) Axiomatic analyticity properties and representations of particles in thermal quantum field theory. *Annales de l'Institut Henri Poincaré* 64: 495–521.

Bros J and Buchholz D (1998) The unmasking of thermal Goldstone bosons. *Physical Review D* 58: 125012-1.

Bros J and Buchholz D (2002) Asymptotic dynamics of thermal quantum fields. *Nuclear Physics B* 627: 289–310.

Buchholz D and Junglas P (1989) On the existence of equilibrium states in local quantum field theory. *Communications in Mathematical Physics* 121: 255–270.

Buchholz D and Ojima I (1997) Spontaneous collapse of supersymmetry. *Nuclear Physics B* 498: 228–242.

Buchholz D, Ojima I, and Roos H (2002) Thermodynamic properties of non-equilibrium states in quantum field theory. *Annals of Physics* 297: 219–242.

Cuniberti G, De Micheli E, and Viano GA (2001) Reconstructing the thermal green functions at real times from those at imaginary times. *Communications in Mathematical Physics* 216: 59–83.

Derezinski J, Jaksic V, and Pillet CA (2003) Perturbations of $W^*$-dynamics, Liouvilleans and KMS states. *Reviews in Mathematical Physics* 15: 447–489.

Fröhlich J (1975) The reconstruction of quantum fields from Euclidean Green's functions at arbitrary temperatures. *Helvetica Physica Acta* 48: 355–363.

Gérard C and Jäkel C (2005) Thermal quantum fields with spatially cutoff interactions in $1+1$ space-time dimensions. *Journal of Functional Analysis* 220: 157–213.

Gérard C and Jäkel C (2005) Thermal quantum fields without cutoffs in $1+1$ space-time dimensions. *Reviews in Mathematical Physics* 17: 113–173.

Haag R (1992) *Local Quantum Physics*. Berlin: Springer.

Høegh-Krohn R (1974) Relativistic quantum statistical mechanics in two-dimensional space-time. *Communications in Mathematical Physics* 38: 195–224.

Jäkel CD (2004) The relation between KMS states for different temperatures. *Annales de l'Institut Henri Poincaré* 5: 1–30.

Kopper C, Müller VF, and Reisz T (2001) Temperature independant renormalization of finite temperature field theory. *Annales de l'Institut Henri Poincaré* 2: 387–402.

Kuckert B (2002) Covariant thermodynamics of quantum systems: passivity, semipassivity, and the Unruh effect. *Annals of Physics* 295: 216–229.

Landsman NP and van Weert Ch G (1987) Real- and imaginary-time field theory at finite temperature and density. *Physics Reports* 145: 141–249.

Narnhofer H (1994) Entropy density for relativistic quantum field theory. *Reviews in Mathematical Physics* 6: 1127–1145.

Narnhofer H, Requardt M, and Thirring W (1983) Quasiparticles at finite temperature. *Communications in Mathematical Physics* 92: 247–268.

Steinmann O (1995) Perturbative quantum field theory at positive temperatures: an axiomatic approach. *Communications in Mathematical Physics* 170: 405–415.


**Thermohydraulics** *see* Newtonian Fluids and Thermohydraulics

# Toda Lattices

**Y B Suris**, Technische Universität München, München, Germany



Lattices, or differential–difference equations, are a special class of ordinary differential equations, with the dependent variable $t$ playing the role of time and an infinite number of dependent variables $q_n = q_n(t)$ numbered by integer indices $n$, characterized by a translational invariance with respect to the shift $n \to n+1$. Due to this property, such equations are well suited for description of processes in translationally symmetric systems like crystals. On his search for lattice models admitting interesting explicit solutions, M Toda discovered in 1967 the lattice which nowadays carries his name:

$$\ddot{q}_n = \mathrm{e}^{q_{n+1}-q_n} - \mathrm{e}^{q_n-q_{n-1}} \qquad [1]$$

Toda lattice is one of the most celebrated systems of mathematical physics, and a large amount of literature is devoted to it and to its various generalizations. Its most prominent property is "integrability," so that it is amenable to a rather complete exact treatment; moreover, it can be regarded as one of the basic models, illustrating all the relevant

paradigms, notions, methods, and results of the theory of integrable systems (sometimes called the theory of solitons). One has a rare possibility to read the first-hand presentation of a large body of relevant results, including the authentic story of the original discovery, in Toda (1989).

## The Infinite Toda Lattice

### Model

The classical infinite Toda lattice [1] describes a one-dimensional chain of unit mass particles, each one interacting with the nearest neighbors only, $q_n$ being the displacement of the $n$th particle from equilibrium. It can be treated within the Hamiltonian formalism of the classical mechanics (with some care, because of the infinite number of degrees of freedom). In this framework, the second-order Newtonian equations of motion [1] are replaced by the first-order Hamiltonian ones, for the coordinates $q_n$ and canonically conjugate momenta $p_n$:

$$\dot{q}_n = p_n, \qquad \dot{p}_n = \mathrm{e}^{q_{n+1}-q_n} - \mathrm{e}^{q_n - q_{n-1}} \qquad [2]$$

The corresponding Hamilton function is

$$H_2(p,q) = \frac{1}{2}\sum_{n\in\mathbb{Z}} p_n^2 + \sum_{n\in\mathbb{Z}} (\mathrm{e}^{q_{n+1}-q_n} - 1) \qquad [3]$$

One can understand infinite sums here formally, or, alternatively, one can impose suitable boundary conditions, like $q_{n+1} - q_n \to 0, p_n \to 0$ as $|n| \to \infty$ (usually one requires decay faster than any degree of $1/|n|$).

### Multisoliton Solutions

M Toda found in 1967 a number of exact traveling wave solutions of this system, including the 1-soliton solution:

$$q_n(t) = \log \frac{1 + \mathrm{e}^{-2(\gamma_1 n - \beta_1 t + \delta_1)}}{1 + \mathrm{e}^{-2(\gamma_1 (n-1) - \beta_1 t + \delta_1)}} \qquad [4]$$

or, equivalently,

$$\mathrm{e}^{q_{n+1}(t) - q_n(t)} = 1 + \frac{\beta_1^2}{\cosh^2(\gamma_1 n - \beta_1 t + \delta_1)} \qquad [5]$$

where $\gamma_1 > 0$, $\beta_1 = \pm\sinh\gamma_1$, and $\delta_1$ is an arbitrary phase. Such a soliton moves with the velocity $v_1 = \beta_1/\gamma_1$ (to the right, if $v_1 > 0$, and to the left, if $v_1 < 0$). Note that the faster the soliton is, the larger its amplitude. Multisoliton solutions were constructed in 1973 by R Hirota with the help of his ingenious "direct" (or bilinear) method. They can be written as

$$\mathrm{e}^{q_{n+1}(t) - q_n(t)} = \frac{\tau_{n+1}(t)\tau_{n-1}(t)}{\tau_n^2(t)} \qquad [6]$$

where, for an $M$-soliton solution, $\tau_n(t)$ can be represented through the $M \times M$ determinant depending on $2M$ parameters $z_j \in (-1,1)$ and $c_j \in \mathbb{R}$:

$$\tau_n(t) = \det\left(\delta_{ij} + \frac{c_i(t)c_j(t)(z_i z_j)^{n+1}}{1 - z_i z_j}\right)_{1\le i,j\le M} \qquad [7]$$

where $c_j(t) = c_j \mathrm{e}^{\beta_j t}$, $\beta_j = (1/2)(z_j^{-1} - z_j)$. If one sets $z_j = \pm\mathrm{e}^{-\gamma_j}$ with $\gamma_j > 0$, then $\beta_j = \pm\sinh\gamma_j$, and one can show that asymptotically both for $t \to -\infty$ and for $t \to +\infty$ the solution [6] looks like the sum of well-separated solitons [4] with the velocities $v_j = \beta_j/\gamma_j$ and the respective phases $\gamma_j n - \beta_j t + \delta_j^{(\pm)}$. This is usually interpreted as a particle-like behavior of solitons. One can show that the scattering of solitons is factorized:

$$\delta_j^{(+)} - \delta_j^{(-)} = \sum_{v_k < v_j} \log\left|\frac{1 - z_j z_k}{z_j - z_k}\right| - \sum_{v_k > v_j} \log\left|\frac{1 - z_j z_k}{z_j - z_k}\right| \qquad [8]$$

which means that the phase shifts of individual solitons can be interpreted as coming from the pairwise interactions only.

### Integrability

The infinite Toda lattice is completely integrable in the sense of the classical Hamiltonian mechanics: it admits an infinite number of functionally independent integrals of motion in involution. This was demonstrated in 1974 by M Hénon. An instance of these higher integrals of motion is given by

$$H_3(p,q) = \frac{1}{3}\sum_{n\in\mathbb{Z}} p_n^3 + \sum_{n\in\mathbb{Z}} (p_n + p_{n+1})\mathrm{e}^{q_{n+1}-q_n} \qquad [9]$$

Hamiltonian flows corresponding to the higher integrals of motion (usually referred to as higher Toda flows) form the "Toda lattice hierarchy." A beautiful approach to this hierarchy is based on the Lax representation of the Toda lattice, discovered in 1974 independently by H Flaschka and S Manakov. In the variables $a_n, b_n$, related to $q_n, p_n$ by

$$a_n = \mathrm{e}^{q_{n+1}-q_n}, \qquad b_n = p_n \qquad [10]$$

equations of motion of the Toda lattice [2] are rewritten as

$$\dot{a}_n = a_n(b_{n+1} - b_n), \qquad \dot{b}_n = a_n - a_{n-1} \qquad [11]$$

It turns out that eqns [11] are equivalent to the operator equation

$$\dot{L} = [L, A_+] = [A_-, L] \quad [12]$$

where $L$ and $A_\pm$ are linear difference operators with coefficients depending on $a_n, b_n$:

$$L = \sum_{n\in\mathbb{Z}} b_n E_{n,n} + \sum_{n\in\mathbb{Z}} a_n E_{n,n+1} + \sum_{n\in\mathbb{Z}} E_{n+1,n} \quad [13]$$

$$A_+ = \sum_{n\in\mathbb{Z}} b_n E_{n,n} + \sum_{n\in\mathbb{Z}} E_{n+1,n}$$
$$A_- = \sum_{n\in\mathbb{Z}} a_n E_{n,n+1} \quad [14]$$

Here difference operators are represented as infinite matrices, $E_{m,n}$ being the matrix with the only nonvanishing element equal to 1 in the position $(m, n)$. A diagonal similarity (gauge) transformation of the matrix $L$ leads to an equivalent Lax representation of the Toda lattice:

$$\dot{L_0} = [L_0, A_0] \quad [15]$$

with

$$L_0 = \sum_{n\in\mathbb{Z}} b_n E_{n,n} + \sum_{n\in\mathbb{Z}} a_n^{1/2}\left(E_{n+1,n} + E_{n,n+1}\right) \quad [16]$$

$$A_0 = \frac{1}{2}\sum_{n\in\mathbb{Z}} a_n^{1/2}\left(E_{n+1,n} - E_{n,n+1}\right) \quad [17]$$

Being equivalent for the Toda lattice, these two Lax representations admit nonequivalent generalizations (see below). Note that the matrices $A_\pm$ in [14] may be interpreted as $A_\pm = \pi_\pm(L)$, where $\pi_\pm$ stands for the lower-triangular, resp., strictly upper-triangular part. The commuting higher members of the Toda lattice hierarchy (enumerated by $s \in \mathbb{N}$) are characterized by the Lax equations of the form [12] with the same Lax matrix $L$ as in [13] and with $A_\pm = \pi_\pm(L^s)$. In the Lax representation [15], the higher Toda flows are obtained by choosing $A_0 = \mathrm{skew}(L_0^s)$, where "skew" denotes the skew-symmetric part (strictly lower-triangular part minus strictly upper-triangular part) of the symmetric matrix. The Hamilton functions of the higher flows are obtained as $H_s \sim \mathrm{tr}(L^s) = \mathrm{tr}(L_0^s)$.

### Inverse Scattering

H Flaschka and S Manakov laid the Lax representation into the base of the application of the inverse-scattering, or inverse-spectral, transformation method (IST) to the infinite Toda lattice. It was the first application of IST in the lattice context. The matrix $L_0$ in [16] is symmetric tridiagonal, which yields that the operator $L_0$ is second order and self-adjoint. The direct and inverse-spectral problem for $L_0\psi = \mu\psi$ with such operators $L_0$ is well studied and parallel, to a large extent, to the corresponding theory for second-order differential operators. In the rapidly decaying case, the set of spectral data of the operator $L_0$, allowing for a solution of the inverse problem, consists of:

1. eigenvalues $\mu_j = z_j + z_j^{-1}$ of the discrete spectrum, with $z_j \in (-1, 1)$;
2. normalizing coefficients $\gamma_j$ of the corresponding eigenfunctions; and
3. reflection coefficient $r(z)$ for $|z| = 1$, characterizing the continuous spectrum $\mu = z + z^{-1} \in [-2, 2]$.

The solution of the inverse-spectral problem is given in terms of the Riemann–Hilbert problem or its variants, like the Gelfand–Levitan equation. Equation [12] means that the evolution of the operator $L$, induced by the evolution of $q_n(t), p_n(t)$ in virtue of the Toda lattice equations [2], is "isospectral." More precisely, the discrete eigenvalues are integrals of motion, while the evolution of other spectral data is governed by simple linear equations:

$$z_j = \text{const.}, \qquad \gamma_j(t) = \gamma_j(0)\mathrm{e}^{\left(z_j^{-1} - z_j\right)t}$$
$$r(z, t) = r(z, 0)\mathrm{e}^{(z^{-1} - z)t} \quad [18]$$

In particular, the multisoliton solutions correspond to the reflectionless case $r(z, t) \equiv 0$. The IST solution of the initial-value problem for the infinite Toda lattice can be schematically depicted as in Figure 1.

### Bi-Hamiltonian Structure

The canonical Poisson bracket for the variables $q_n, p_n$ turns in the Flaschka–Manakov variables [10] into

$$\{b_n, a_n\}_1 = -a_n, \qquad \{a_n, b_{n+1}\}_1 = -a_n \quad [19]$$

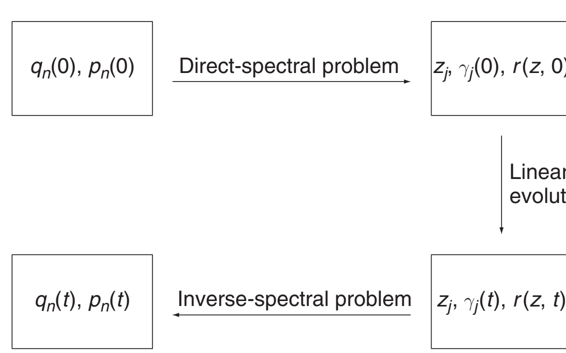



**Figure 1** General scheme of the IST.

(all other brackets of the coordinate functions vanish), and the system [11] is Hamiltonian with respect to this bracket, with the Hamilton function

$$H_2 = \frac{1}{2}\sum b_n^2 + \sum a_n$$

However, one can define also a different Poisson bracket for the variables $a_n, b_n$:

$$\begin{aligned} \{b_n, a_n\}_2 &= -b_n a_n \\ \{a_n, b_{n+1}\}_2 &= -a_n b_{n+1} \\ \{b_n, b_{n+1}\}_2 &= -a_n \\ \{a_n, a_{n+1}\}_2 &= -a_n a_{n+1} \end{aligned} \qquad [20]$$

with the following properties: it is compatible with the first one (i.e., their linear combinations are again Poisson brackets), and the system [11] is Hamiltonian with respect to this bracket, with the Hamilton function $H_1 = \sum b_n$. So, the Toda lattice in the form [11] is a bi-Hamiltonian system. This result is due to M Adler (1979). The bi-Hamiltonian property, introduced by F Magri in 1978 on the example of the Korteweg–de Vries equation, has been established since then as an alternative (and highly effective and informative) definition of integrability. Actually, the Toda lattice [11] is even tri-Hamiltonian, since there exists one more local Poisson bracket for the variables $a_n, b_n$ with similar properties, discovered by B Kupershmidt in 1985.

### Darboux–Bäcklund Transformations and Discretization

A further indispensable attribute of integrable systems are the so-called Darboux–Bäcklund transformations. For the Toda lattice they were first found by M Toda and M Wadati in 1975. A Bäcklund transformation $(q_n, p_n) \mapsto (\widetilde{q}_n, \widetilde{p}_n)$ with the parameter $h$ can be written as

$$\begin{aligned} 1 + hp_n &= e^{\tilde{q}_n - q_n} + h^2 e^{q_n - \tilde{q}_{n-1}} \\ 1 + h\widetilde{p}_n &= e^{\tilde{q}_n - q_n} + h^2 e^{q_{n+1} - \tilde{q}_n} \end{aligned} \qquad [21]$$

This is a canonical transformation, possessing a classical generating function. These formulas can be given a fundamentally important interpretation in terms of the matrices

$$U_+ = \sum_{n \in \mathbb{Z}} e^{\tilde{q}_n - q_n} E_{n,n} + h \sum_{n \in \mathbb{Z}} E_{n+1,n} \qquad [22]$$

$$U_- = I + h \sum_{n \in \mathbb{Z}} e^{q_{n+1} - \tilde{q}_n} E_{n,n+1} \qquad [23]$$

The first formula in [21] is equivalent to the factorization $I + hL = U_+ U_-$, while the second one is equivalent to the factorization $I + h\widetilde{L} = U_- U_+$ with the flipped factors. The Bäcklund transformation [21] serves also as an integrable discretization of the Toda flow [2] with the time step $h$.

## Finite Open-End Toda Lattice

### Model

The infinite Toda lattice [1] can be reduced to finite-dimensional systems by imposing suitable boundary conditions, different from the rapidly decaying ones. Particularly important are "open-end boundary conditions," which correspond to placing the particles 0 and $N+1$ at $q_0 = +\infty$ and $q_{N+1} = -\infty$, respectively. In terms of the Flaschka–Manakov variables, this means that $a_0 = a_N = 0$ and $b_0 = b_{N+1} = 0$. The Hamilton function of the resulting system with $N$ degrees of freedom is

$$H_2(p, q) = \frac{1}{2}\sum_{n=1}^{N} p_n^2 + \sum_{n=1}^{N-1} e^{q_{n+1} - q_n} \qquad [24]$$

This system consists of $N$ particles subject to repulsive forces between nearest neighbors, and exhibits a scattering behavior both as $t \to -\infty$ and $t \to +\infty$. It admits a Lax representation of the same form [12] or [15] as in the infinite case, but with all the matrices being now of finite size $N \times N$, so that [13]–[14] and [16]–[17] are replaced by

$$L = \sum_{n=1}^{N} b_n E_{n,n} + \sum_{n=1}^{N-1} a_n E_{n,n+1} + \sum_{n=1}^{N-1} E_{n+1,n} \qquad [25]$$

$$\begin{aligned} A_+ &= \sum_{n=1}^{N} b_n E_{n,n} + \sum_{n=1}^{N-1} E_{n+1,n} \\ A_- &= \sum_{n=1}^{N-1} a_n E_{n,n+1} \end{aligned} \qquad [26]$$

and

$$L_0 = \sum_{n=1}^{N} b_n E_{n,n} + \sum_{n=1}^{N-1} a_n^{1/2} \left(E_{n+1,n} + E_{n,n+1}\right) \qquad [27]$$

$$A_0 = \frac{1}{2}\sum_{n=1}^{N-1} a_n^{1/2} \left(E_{n+1,n} - E_{n,n+1}\right) \qquad [28]$$

The qualitative behavior of the solutions is easily understood: as a consequence of repulsive interactions, the pairwise distances between particles grow infinitely, $a_n(t) = e^{q_{n+1}(t) - q_n(t)} \to 0$ as $t \to \pm\infty$, so that the matrix $L_0$ becomes asymptotically diagonal, with the limit velocities $b_n(\pm\infty) = \dot{q}_n(\pm\infty)$ as the diagonal entries. Due to the isospectral evolution of $L_0$, these limit velocities have to coincide with the eigenvalues $\mu_j$ of $L_0$, which are integrals of motion.

As $t \to -\infty$, they appear on the diagonal in the increasing order (the rightmost particle $q_1$ being the slowest, and the leftmost $q_N$ being the fastest), while as $t \to +\infty$, their order on the diagonal changes to the decreasing one (the particle $q_1$ becoming the fastest and $q_N$ becoming the slowest).

### Moser's Solution

Integration of this system has been first performed by J Moser in 1975. His solution can be interpreted within the general scheme of the IST (see **Figure 1**). The spectral data in this case consist, for example, of the eigenvalues $\mu_j (j = 1, \ldots, N)$ of the matrix $L_0$ and the first components $r_j$ of the corresponding orthonormal eigenvectors. The evolution of these data induced by the Toda flow [2] turns out to be simple:

$$\mu_j = \text{const.}, \qquad r_j^2(t) = \frac{r_j^2(0)\mathrm{e}^{\mu_j t}}{\sum_{i=1}^{N} r_i^2(0)\mathrm{e}^{\mu_i t}} \qquad [29]$$

The IST is expressed by the identity

$$\sum_{j=1}^{N} \frac{r_j^2}{\mu - \mu_j} = \cfrac{1}{\mu - b_1 - \cfrac{a_1}{\mu - b_2 - \ddots - \cfrac{a_{N-1}}{\mu - b_N}}} \qquad [30]$$

both parts of which represent the entry $(1, 1)$ of the matrix $(\mu I - L_0)^{-1}$. It implies that all variables $a_n(t), b_n(t)$ are rational functions of $\mu_j$ and $\mathrm{e}^{\mu_j t}$; in particular, one finds:

$$a_n(t) = \mathrm{e}^{q_{n+1}(t) - q_n(t)} = \frac{\tau_{n-1}(t)\tau_{n+1}(t)}{\tau_n^2(t)} \qquad [31]$$

where $\tau_n(t)$ can be represented as an $n \times n$ Hankel determinant

$$\tau_n(t) = \det\big(c_{j+k}(t)\big)_{0 \le j,k \le n-1}$$
$$c_j(t) = \sum_{i=1}^{N} \mu_i^j r_i^2(t) \qquad [32]$$

### Factorization Solution

The Lax representation [12] is a particular instance of a general construction, known under the name of Adler–Kostant–Symes (AKS) method and found around 1980. The ingredients of this construction are:

- a Lie algebra $\mathfrak{g}$, equipped with a nondegenerate scalar product which is used to identify $\mathfrak{g}$ with its dual space $\mathfrak{g}^*$;
- a splitting of $\mathfrak{g}$ into a direct sum of its two subspaces $\mathfrak{g}_\pm$ which are also Lie subalgebras, with $\pi_\pm : \mathfrak{g} \to \mathfrak{g}_\pm$ being the corresponding projections;
- the Lie group $G$ of the Lie algebra $\mathfrak{g}$, and its Lie subgroups $G_\pm$ with the Lie algebras $\mathfrak{g}_\pm$; and
- a function $\phi : \mathfrak{g} \to \mathfrak{g}$ covariant with respect to the adjoint action of $G$ (in the case of matrix Lie algebras and groups, one can take, e.g., $\phi(L) = L^s$).

The AKS method provides a formula for the solution of the initial-value problem for Lax equations of the form [12] with the Lax matrix $L \in \mathfrak{g}$ and $A_\pm = \pi_\pm(\phi(L))$. The solution is given by

$$L(t) = U_+^{-1}(t)L(0)U_+(t) = U_-(t)L(0)U_-^{-1}(t) \qquad [33]$$

where the elements $U_\pm(t) \in G_\pm$ solve the factorization problem

$$\exp\big(t\phi(L(0))\big) = U_+(t)U_-(t) \qquad [34]$$

For the open-end Toda lattice $\mathfrak{g} = \mathrm{gl}(N)$, the Lie algebra of all $N \times N$ matrices, $\mathfrak{g}_\pm$ consist of all lower-triangular, resp., strictly upper-triangular, matrices. Accordingly, $G = \mathrm{GL}(N)$, the Lie group of all nondegenerate $N \times N$ matrices, and $G_\pm$ consist of all nondegenerate lower-triangular matrices, resp., of upper-triangular matrices with units on the diagonal. The corresponding factorization problem in $G$ is well known in the linear algebra under the name of LR factorization, and is related to the Gaussian elimination. From [33] and the well-known expression of the diagonal elements of the lower-triangular factor in the LR factorization through the minors of the factorized matrix, we find:

$$a_n(t) = \frac{\tau_{n+1}(t)\tau_{n-1}(t)}{\tau_n^2(t)} a_n(0) \qquad [35]$$

where $\tau_n(t)$ is the upper-left $n \times n$ minor of the matrix $\exp(tL(0))$. If $L(t)$ is the Lax matrix along the solution of the Toda flow ($\phi(L) = L$), then the sampling of the matrix $\exp(L(t))$ at the integer times $t \in \mathbb{Z}$ coincides with the result of application of the Rutishauser's *LR* algorithm to the matrix $\exp(L(0))$. The *LR* algorithm applied to the matrix $I + hL(0)$ is nothing other but the Bäcklund transformation [21] in the open-end situation.

## Finite Periodic Toda Lattice

### Model

A different reduction of the infinite Toda lattice to a finite-dimensional system appears by imposing periodic boundary conditions, $q_{n+N}(t) \equiv q_n(t)$ for all $n \in \mathbb{Z}$ (of course, such relations hold also for the Flaschka–Manakov variables $a_n, b_n$). The Hamilton

function of the resulting system with $N$ degrees of freedom is

$$H_2(p,q) = \frac{1}{2}\sum_{n\in\mathbb{Z}/N\mathbb{Z}} p_n^2 + \sum_{n\in\mathbb{Z}/N\mathbb{Z}} e^{q_{n+1}-q_n} \qquad [36]$$

This system consists of $N$ particles $q_n (n = 1, \ldots, N)$, and it is always assumed that $q_{N+1} \equiv q_1$ and $q_0 \equiv q_N$. Thus, the potential energy in [36] differs from the potential energy in [24] by one additional term $e^{q_1-q_N}$. However, this modest difference leads to much more complicated dynamics of the system (quasiperiodic instead of scattering). It is convenient to replace infinite matrices in the Lax representation [12] by finite ones, of size $N \times N$, but depending on an additional parameter $\lambda$ (called the spectral parameter):

$$L = \sum_{n\in\mathbb{Z}/N\mathbb{Z}} b_n E_{n,n} + \lambda^{-1} \sum_{n\in\mathbb{Z}/N\mathbb{Z}} a_n E_{n,n+1} + \lambda \sum_{n\in\mathbb{Z}/N\mathbb{Z}} E_{n+1,n} \qquad [37]$$

$$A_+ = \sum_{n\in\mathbb{Z}/N\mathbb{Z}} b_n E_{n,n} + \lambda \sum_{n\in\mathbb{Z}/N\mathbb{Z}} E_{n+1,n} \qquad [38]$$

$$A_- = \lambda^{-1} \sum_{n\in\mathbb{Z}/N\mathbb{Z}} a_n E_{n,n+1} \qquad [39]$$

The Lax representation [12] holds identically in $\lambda$, so that the spectral parameter drops out of the equations of motion. Note that, unlike the open-end case, $L$ is no more a tridiagonal matrix, because of the nonvanishing entries in the positions $(N, 1)$ and $(1, N)$.

### Inverse-Spectral Transformation

Solution of the periodic lattice in terms of multidimensional theta functions has been given independently by E Date and S Tanaka, and by I Krichever in 1976. In this case, the set of the spectral data is more complicated; it includes:

- a hyperelliptic Riemann surface $\mathcal{R}$ of genus $N-1$ determined by the eigenvalues of the periodic boundary-value problem for the operator $L$, or, in other words, by the equation $R(\lambda,\mu) = \det(L(\lambda) - \mu I) = 0$; and
- $N-1$ points $P_k$ on $\mathcal{R}$, which correspond to the eigenvalues of $L$ with vanishing boundary conditions.

Due to [12], the Riemann surface $\mathcal{R}$ itself is an integral of motion, and the evolution of points $P_k$ is such that the image of the divisor $P_1 + \cdots + P_{N-1}$ under the Abel map moves along a straight line in the Jacobi variety of $\mathcal{R}$. Solution of the inverse-spectral problem is given in terms of multidimensional theta-functions by formula [35] with $\tau_n(t) = \theta(nU - tV + D)$, where $U, V, D$ are certain vectors on the Jacobian of $\mathcal{R}$ (the first two of them depending on the spectrum $\mathcal{R}$ only).

### Loop Algebras

The periodic Toda lattice can be included into the general AKS scheme, if one interprets the Lax matrix $L$ as an element of the loop algebra $\boldsymbol{g}$ which consists of Laurent polynomials (in $\lambda$) with coefficients from $\mathrm{gl}(N)$, singled out by the additional condition

$$\boldsymbol{g} = \left\{ L(\lambda) \in \mathrm{gl}(N)[\lambda, \lambda^{-1}] : \Omega L(\lambda)\Omega^{-1} = L(\omega\lambda) \right\}$$

where $\Omega = \mathrm{diag}(1, \omega, \ldots, \omega^{N-1})$, $\omega = \exp(2\pi i/N)$. Subalgebras $\boldsymbol{g}_\pm$ consist of Laurent polynomials with respect to non-negative, resp., strictly negative powers of $\lambda$. The Lie group $\boldsymbol{G}$ corresponding to the Lie algebra $\boldsymbol{g}$ consists of $\mathrm{GL}(N)$-valued functions $U(\lambda)$ of the complex parameter $\lambda$, regular in $\mathbb{C}P^1 \backslash \{0, \infty\}$ and satisfying $\Omega U(\lambda)\Omega^{-1} = U(\omega\lambda)$. Its subgroups $\boldsymbol{G}_\pm$ corresponding to the Lie algebras $\boldsymbol{g}_\pm$ are singled out by the following conditions: elements of $\boldsymbol{G}_+$ are regular in the neighborhood of $\lambda = 0$, while elements of $\boldsymbol{G}_-$ are regular in the neighborhood of $\lambda = \infty$ and take at $\lambda = \infty$ the value $I$. The corresponding factorization is called the generalized LR factorization. As opposed to the open-end case, finding such a factorization is a problem of the Riemann–Hilbert type which is solved in terms of algebraic geometry and theta-functions rather than in terms of linear algebra and exponential functions. This approach to the periodic Toda lattice is due to Reyman and Semenov-Tian-Shansky (1979) and, independently, to M Adler and P van Moerbeke (1980).

## Generalizations: Lie-Algebraic Systems

The AKS interpretation of the finite Toda lattices leads directly to their generalizations by replacing the algebra $\mathrm{gl}(N)$, resp., the loop algebra over $\mathrm{gl}(N)$, by simple Lie algebras, resp. affine Lie algebras. These generalized Toda systems were introduced in 1976 by O Bogoyavlensky and solved in 1979 independently by M Olshanetsky, A Perelomov, and by B Kostant.

### Simple Lie Algebras

Let $\mathfrak{g}$ be a simple Lie algebra (complex or real split), and $\mathfrak{h}$ its Cartan subalgebra. Let further $\Delta = \Delta_+ \cup \Delta_-$ be the root system of $\mathfrak{g}$, decomposed into the sets of positive roots $\Delta_+$ and the set of negative roots $\Delta_-$. One has a direct vector space $\mathfrak{g} = \mathfrak{g}_+ \oplus \mathfrak{g}_-$, where $\mathfrak{g}_+$ is spanned by the root spaces for positive roots and by $\mathfrak{h}$,

while $\mathfrak{g}_-$ is spanned by the root spaces for negative roots (Borel decomposition). For $\alpha \in \Delta$ let $E_\alpha$ be a corresponding root vector. So, $[H, E_\alpha] = \alpha(H)E_\alpha$ for all $H \in \mathfrak{h}$. The root $\alpha \in \mathfrak{h}^*$ may be identified with $H_\alpha \in \mathfrak{h}$ defined by $\langle H_\alpha, H\rangle = \alpha(H)$ for all $H \in \mathfrak{h}$. It is easy to deduce that $[E_\alpha, E_{-\alpha}] = c_\alpha H_\alpha$, where $c_\alpha = \langle E_\alpha, E_{-\alpha}\rangle$. The system of simple roots will be denoted by $\Phi \subset \Delta_+$.

The generalized Toda lattice for the Lie algebra $\mathfrak{g}$ is the following system of differential equations on $\mathfrak{h} \times \mathfrak{h}$:

$$\begin{aligned} \dot{Q} &= P \\ \dot{P} &= -\sum_{\alpha\in\Phi} \mathrm{e}^{\alpha(Q)}[E_\alpha, E_{-\alpha}] = -\sum_{\alpha\in\Phi} c_\alpha \mathrm{e}^{\alpha(Q)} H_\alpha \end{aligned} \qquad [40]$$

This system can be given a Hamiltonian formulation, with the Hamilton function

$$H_{\mathfrak{g}} = \frac{1}{2}\langle P, P\rangle + \sum_{\alpha\in\Phi} c_\alpha \mathrm{e}^{\alpha(Q)} \qquad [41]$$

It is completely integrable, and has a Lax representation [12] with

$$L = P + \sum_{\alpha\in\Phi} E_\alpha + \sum_{\alpha\in\Phi} \mathrm{e}^{\alpha(Q)} E_{-\alpha} \qquad [42]$$

$$A_+ = P + \sum_{\alpha\in\Phi} E_\alpha, \quad A_- = \sum_{\alpha\in\Phi} \mathrm{e}^{\alpha(Q)} E_{-\alpha} \qquad [43]$$

The usual open-end Toda lattice corresponds to the algebra sl($N$) (series $A_{N-1}$), so that the Hamilton function [24] can be denoted by $H_{A_{N-1}}$. The Hamilton functions of the generalized lattices corresponding to other classical algebras so($2N+1$) (series $B_N$), sp($N$) (series $C_N$), and so($2N$) (series $D_N$) can be written in the canonically conjugate variables $q_n, p_n (n = 1, \ldots, N)$ as

$$H_{\mathfrak{g}}(p,q) = H_{A_{N-1}}(p,q) + \begin{cases} \mathrm{e}^{-q_N}, & \mathfrak{g} = B_N \\ \mathrm{e}^{-2q_N}, & \mathfrak{g} = C_N \\ \mathrm{e}^{-q_N - q_{N-1}}, & \mathfrak{g} = D_N \end{cases} \qquad [44]$$

### Affine Lie Algebras

Turning to the generalizations of the periodic Toda lattice, let $\theta$ be a Coxeter automorphism of a simple complex algebra $\mathfrak{g}$, the order of $\theta$ being $m$. Introduce the loop algebra $\boldsymbol{g}$ as the Lie algebra of Laurent polynomials

$$\boldsymbol{g} = \left\{ L(\lambda) \in \mathfrak{g}[\lambda, \lambda^{-1}] : \theta(L(\lambda)) = L(\omega\lambda) \right\}$$

where $\omega = \exp(2\pi\mathrm{i}/m)$. Denote by $\mathfrak{g}_j$ the eigenspaces of $\theta$ corresponding to the eigenvalues $\omega^j (j \in \mathbb{Z}/m\mathbb{Z})$. Set $\mathfrak{a} = \mathfrak{g}_0$, and let $s$ denote the dimension of $\mathfrak{a}$. By definition of the Coxeter automorphism, $\mathfrak{a}$ is an abelian subalgebra of $\mathfrak{g}$. Denote by $\Psi$ the set of $\alpha \in \mathfrak{a}^*$ for which there exist nonzero elements $E_\alpha \in \mathfrak{g}_1$ with $[H, E_\alpha] = \alpha(H)E_\alpha$ for all $H \in \mathfrak{a}$. The elements $E_{-\alpha} \in \mathfrak{g}_{-1}$ are defined similarly. It can be shown that $\Psi$ contains $s+1$ elements, so that between them there exists exactly one linear relation. The elements of $\Psi$ are called simple weights of the loop algebra $\boldsymbol{g}$. The Lie algebra $\boldsymbol{g}$ is a direct sum of its two subspaces $\boldsymbol{g}_\pm$ consisting of Laurent polynomials with non-negative, resp., with strictly negative powers of $\lambda$; these subspaces are also Lie subalgebras.

Now the generalized Toda lattice related to the loop algebra $\boldsymbol{g}$ can be introduced as the system of differential equations on $\mathfrak{a} \times \mathfrak{a}$, which looks formally exactly as [40], and has the Hamilton function which looks exactly as [41], but with the set of simple roots $\Phi$ of $\mathfrak{g}$ being replaced by the set of simple weights $\Psi$ of $\boldsymbol{g}$. The matrices participating in the Lax representation [12] belong now to the loop algebra $\boldsymbol{g}$:

$$L(\lambda) = P + \lambda \sum_{\alpha\in\Psi} E_\alpha + \lambda^{-1} \sum_{\alpha\in\Psi} \mathrm{e}^{\alpha(Q)} E_{-\alpha} \qquad [45]$$

$$\begin{aligned} A_+(\lambda) &= P + \lambda \sum_{\alpha\in\Psi} E_\alpha \\ A_-(\lambda) &= \lambda^{-1} \sum_{\alpha\in\Psi} \mathrm{e}^{\alpha(Q)} E_{-\alpha} \end{aligned} \qquad [46]$$

For the classical series of loop algebras, the Hamilton functions $H_{\boldsymbol{g}}$ in the canonically conjugate variables $q_n, p_n (n = 1, \ldots, N)$ can be presented as

$$H_{\boldsymbol{g}}(p,q) = H_{A_{N-1}}(p,q) + \begin{cases} \mathrm{e}^{-q_N} + \mathrm{e}^{q_1+q_2}, & \boldsymbol{g} = B_N^{(1)} \\ \mathrm{e}^{-2q_N} + \mathrm{e}^{2q_1}, & \boldsymbol{g} = C_N^{(1)} \\ \mathrm{e}^{-q_N - q_{N-1}} + \mathrm{e}^{q_1+q_2}, & \boldsymbol{g} = D_N^{(1)} \\ \mathrm{e}^{-2q_N} + \mathrm{e}^{q_1+q_2}, & \boldsymbol{g} = A_{2N-1}^{(2)} \\ \mathrm{e}^{-q_N} + \mathrm{e}^{2q_1}, & \boldsymbol{g} = A_{2N}^{(2)} \\ \mathrm{e}^{-q_N} + \mathrm{e}^{q_1}, & \boldsymbol{g} = D_{N+1}^{(2)} \end{cases} \qquad [47]$$

Actually, one can find even more general integrable systems of the Toda type: one can add to $H_{A_{N-1}}(p,q)$ any of the two potentials $\mathrm{e}^{-q_N - q_{N-1}}$ or $\alpha\mathrm{e}^{-q_N} + \beta\mathrm{e}^{-2q_N}$ on one end combined with any of the two potentials $\mathrm{e}^{q_1+q_2}$ or $\gamma\mathrm{e}^{q_1} + \delta\mathrm{e}^{2q_1}$ on the other end, where $\alpha, \beta, \gamma, \delta$ are arbitrary constants. This result is due to E Sklyanin (1987).

## Generalizations: Lattices with Nearest-Neighbor Interactions

There exist further integrable lattice systems with the nearest-neighbor interaction apart from the classical exponential Toda lattice [1]. Those of the

type $\ddot{q}_n = r(\dot{q}_n)(g(q_{n+1} - q_n) - g(q_n - q_{n-1}))$ have been classified by R Yamilov in 1982, and the list contains, apart from the usual Toda lattice [1], the following ones:

$$\ddot{q}_n = \dot{q}_n(e^{q_{n+1}-q_n} - e^{q_n-q_{n-1}}) \quad [48]$$

$$\ddot{q}_n = \dot{q}_n(q_{n+1} - 2q_n + q_{n-1}) \quad [49]$$

$$\ddot{q}_n = -(\dot{q}_n^2 - \nu^2)\left(\frac{1}{q_{n+1} - q_n} - \frac{1}{q_n - q_{n-1}}\right) \quad [50]$$

$$\ddot{q}_n = -(\dot{q}_n^2 - \nu^2)(\coth(q_{n+1} - q_n) - \coth(q_n - q_{n-1})) \quad [51]$$

Equations [48] are known as the "modified Toda lattice." Equations [49] describe the "dual Toda lattice" which was instrumental in the original discovery by Toda (see Toda (1989)). All systems [49]–[51] can be obtained from [11] via suitable parametrizations of the variables $a_n, b_n$ by canonically conjugate ones $q_n, p_n$, similar to [10] for [1], see Suris (2003).

A remarkable discovery of the integrable relativistic Toda lattice is due to S Ruijsenaars (1990). This lattice with the equations of motion

$$\ddot{q}_n = (1 + \alpha\dot{q}_n)\left((1 + \alpha\dot{q}_{n+1})\frac{e^{q_{n+1}-q_n}}{1 + \alpha^2 e^{q_{n+1}-q_k}} - (1 + \alpha\dot{q}_{n-1})\frac{e^{q_n-q_{n-1}}}{1 + \alpha^2 e^{q_n-q_{n-1}}}\right) \quad [52]$$

can be considered as the perturbation of the usual Toda lattice with the small parameter $\alpha$ (the inverse speed of light).

A class of integrable lattice systems of the relativistic Toda type $\ddot{q}_n = r(\dot{q}_n)(\dot{q}_{n+1}f(q_{n+1} - q_n) - \dot{q}_{n-1}f(q_n - q_{n-1}) + g(q_{n+1} - q_n) - g(q_n - q_{n-1}))$ is richer than that of the Toda type, and has been isolated by Yu B Suris and by V Adler and A Shabat in 1997. The list contains, apart from the relativistic Toda lattice [52], two more $\alpha$-perturbations of the usual Toda lattice [1]:

$$\ddot{q}_n = (1 + \alpha\dot{q}_{n+1})e^{q_{n+1}-q_n} - (1 + \alpha\dot{q}_{n-1})e^{q_n-q_{n-1}} - \alpha^2\left(e^{2(q_{n+1}-q_n)} - e^{2(q_n-q_{n-1})}\right) \quad [53]$$

$$\ddot{q}_n = (1 - \alpha\dot{q}_n)^2\left((1 - \alpha\dot{q}_{n+1})e^{q_{n+1}-q_n} - (1 - \alpha\dot{q}_{n-1})e^{q_n-q_{n-1}}\right) \quad [54]$$

two $\alpha$-perturbations of the modified Toda lattice [48]:

$$\ddot{q}_k = \dot{q}_n\left(e^{q_{n+1}-q_n} - e^{q_n-q_{n-1}} + \alpha\dot{q}_{n+1}\frac{e^{q_{n+1}-q_n}}{1 + \alpha e^{q_{n+1}-q_n}} - \alpha\dot{q}_{n-1}\frac{e^{q_n-q_{n-1}}}{1 + \alpha e^{q_n-q_{n-1}}}\right) \quad [55]$$

$$\ddot{q}_n = \dot{q}_n(1 - \alpha\dot{q}_n)\left((1 - \alpha\dot{q}_{n+1})\frac{e^{q_{n+1}-q_n}}{1 + \alpha e^{q_{n+1}-q_n}} - (1 - \alpha\dot{q}_{n-1})\frac{e^{q_n-q_{n-1}}}{1 + \alpha e^{q_n-q_{n-1}}}\right) \quad [56]$$

two $\alpha$-perturbations of the dual Toda lattice [49]:

$$\ddot{q}_n = \dot{q}_n(q_{n+1} - 2q_n + q_{n-1}) + \frac{\alpha\dot{q}_{n+1}\dot{q}_n}{1 + \alpha(q_{n+1} - q_n)} - \frac{\alpha\dot{q}_n\dot{q}_{n-1}}{1 + \alpha(q_n - q_{n-1})} \quad [57]$$

$$\ddot{q}_n = \dot{q}_n(1 + \alpha^2\dot{q}_n)\left(\frac{q_{n+1} - q_n - \alpha\dot{q}_{n+1}}{1 + \alpha(q_{n+1} - q_n)} - \frac{q_n - q_{n-1} - \alpha\dot{q}_{n-1}}{1 + \alpha(q_n - q_{n-1})}\right) \quad [58]$$

and one $\alpha$-perturbation of each of the systems [50] and [51]:

$$\ddot{q}_n = -(\dot{q}_n^2 - \nu^2)\left(\frac{q_{n+1} - q_n - \alpha\dot{q}_{n+1}}{(q_{n+1} - q_n)^2 - (\nu\alpha)^2} - \frac{q_n - q_{n-1} - \alpha\dot{q}_{n-1}}{(q_n - q_{n-1})^2 - (\nu\alpha)^2}\right) \quad [59]$$

$$\ddot{q}_n = -\frac{1}{2}(\dot{q}_n^2 - \nu^2) \times \left(\frac{\sinh 2(q_{n+1} - q_n) - \nu^{-1}\sinh(2\nu\alpha)\dot{q}_{n+1}}{\sinh^2(q_{n+1} - q_n) - \sinh^2(\nu\alpha)} - \frac{\sinh 2(q_n - q_{n-1}) - \nu^{-1}\sinh(2\nu\alpha)\dot{q}_{n-1}}{\sinh^2(q_n - q_{n-1}) - \sinh^2(\nu\alpha)}\right) \quad [60]$$

A detailed study of all these systems, their interrelations, and time discretizations can be found in Suris (2003).

There exist also lattices with more complicated nearest-neighbor interactions, involving elliptic functions. They were discovered by A Shabat and R Yamilov (1990), and by I Krichever (2000). For example, the nonrelativistic elliptic Toda lattice is governed by the equations

$$\ddot{q}_n = (\dot{q}_n^2 - 1)(V(q_n, q_{n+1}) + V(q_n, q_{n-1})) \quad [61]$$

where $V(q, q') = \zeta(q + q') + \zeta(q - q') - \zeta(2q)$ is an elliptic function in both arguments $q, q'$ (here $\zeta(q)$ is the Weierstrass $\zeta$-function).

## Further Developments and Generalizations

### Sato's Theory

Formulas [6], [31], and [35] have the same structure, with the case-dependent functions $\tau_n(t)$ given by the determinants [7] for the multisoliton solution in the

infinite case, by the Hankel determinants [32] or by the minors of the matrix $\exp(L(0))$ in the open case, and by the multidimensional theta functions in the periodic case. All these seemingly different objects are actually particular cases of a beautiful construction due to M Sato (1981), developed by E Date, M Jimbo, M Kashiwara, T Miwa (1981–83), and by G Segal and G Wilson (1985), which provides one of the major unifying schemes for the theory of integrable systems. In this construction, integrable systems are interpreted as simple dynamical systems on an infinite-dimensional Grassmannian. The $\tau$-function (first invented by R Hirota in 1971) receives in this theory a representation-theoretical interpretation in terms of the determinant bundle over the Grassmannian.

#### Band Matrices

The Lax matrices [13] and [16] in the Manakov–Flaschka variables can be easily generalized: in the symmetric matrix $L_0$ one can admit nonvanishing elements in the band of the width $2s+1>3$ around the main diagonal, in the Heisenberg matrix $L$ one can admit more nonvanishing diagonals in the upper-triangle part. A systematic presentation of a large body of relevant results is given in Kupershmidt (1985). In the setting of finite lattices, the integrability of such systems becomes a non-trivial problem (as opposed to the tridiagonal situation), because the number of independent conjugation-invariant functions $\mathrm{tr}(L^s)$ becomes less than the number of degrees of freedom. An effective approach to this problem based on the semi-invariant functions has been found by P Deift, L-Ch Li, T Nanda, and C Tomei in 1986.

#### Two-Dimensional Toda Lattices

Up to now, we considered integrable lattices with one continuous and one discrete independent variables. This allows for a further generalization. Integrable systems with two continuous and one discrete independent variables are well known and widely used as models of the field theory. For instance, the Toda field theory deals with the system

$$(q_n)_{xy} = \mathrm{e}^{q_{n+1}-q_n} - \mathrm{e}^{q_n-q_{n-1}} \qquad [62]$$

introduced in the soliton theory by A Mikhailov in 1979. This two-dimensional system admits all possible kinds of reductions and generalizations mentioned above for the usual Toda lattice. In particular, the periodic two-dimensional Toda lattice is referred to as the affine Toda field theory (with the prominent example of the sine-Gordon field which corresponds to the period 2). Later, it was realized that the equivalent equation $(\log v_n)_{xy} = v_{n+1} - 2v_n + v_{n-1}$, which is obtained from [62] by setting $v_n = \exp(q_{n+1}-q_n)$, already appeared in studies by G Darboux in the 1880s, as the equation satisfied by the Laplace invariants of the chain of Laplace transformations of a given conjugate net. This relation to the classical differential geometry was extensively studied by G Darboux, G Tzitzéica, and others long before the advent of the theory of integrable systems. Another link to the differential geometry is a more recent observation, and relates the two-dimensional Toda lattice, with the d'Alembert operator $(\cdot)_{xy}$ on the left-hand side of [62] replaced by the Laplace operator $(\cdot)_{z\bar{z}}$, to harmonic maps. For instance, the sinh-Gordon equation $u_{z\bar{z}} = \sinh u$ governs harmonic maps from $\mathbb{C}$ into the unit sphere $\mathbb{S}^2$, which can be interpreted also as Gauss maps of the constant mean curvature surfaces in $\mathbb{R}^3$. A review of this topic can be found in Guest (1997).

Discretization of Toda lattices, nonabelian Toda Lattices, quantization of Toda lattices, dispersionless limit of Toda lattices, etc., are only some of the further relevant topics, which cannot be discussed in any detail in the restricted frame of this article, and the same holds, unfortunately, for such fascinating applications of the Toda lattice as the Frobenius manifolds, Laplacian growth problem, quantum cohomology, random matrix theory, two-dimensional gravity, etc.

*See also:* Bäcklund Transformations; Bi-Hamiltonian Methods in Soliton Theory; Classical *r*-Matrices, Lie Bialgebras, and Poisson Lie Groups; Current Algebra; Dynamical Systems and Thermodynamics; Functional Equations and Integrable Systems; Integrable Discrete Systems; Integrable Systems and Discrete Geometry; Integrable Systems and the Inverse Scattering Method; Integrable Systems: Overview; Lie Groups: General Theory; Multi-Hamiltonian Systems; Quantum Calogero–Moser Systems; Separation of Variables for Differential Equations; Solitons and Kac–Moody Lie Algebras; WDVV Equations and Frobenius Manifolds.

## Further Reading

Adler M (1979) On a trace functional for formal pseudo–differential operators and the symplectic structure for Korteweg–de Vries type equations. *Invent. Mathematics* 50: 219–248.

Adler VE and Shabat AB (1997) On a class of Toda chains. *Teor. Mat. Phys.* 111: 323–334 (in Russian; English translation: *Theor. Math. Phys.* 111: 647–657); Generalized Legendre transformations. *Teor. Mat. Phys.* 112: 179–194 (in Russian; English translation: *Theor. Math. Phys.* 112: 935–948).

Adler M and van Moerbeke P (1980) Completely integrable systems, Kac–Moody algebras and curves; *Advances in Mathematics* 38: 267–317. Linearization of Hamiltonian systems, Jacobi varieties and representation theory. *Advances in Mathematics* 38: 318–379.

Bogoyavlensky OI (1976) On perturbations of the periodic Toda lattice. *Communications in Mathematical Physics* 51: 201–209.

Date F, Jimbo M, and Miwa T (1982–83) Method for generating discrete soliton equations. I–V. *Journal of the Physical Society of Japan* 51: 4116–4124, 4125–4131; 52: 388–393, 761–765, 766–771.

Date E and Tanaka S (1976) Analogue of inverse scattering theory for the discrete Hill's equation and exact solutions for the periodic Toda lattice. *Progress in Theoretical Physics* 55: 457–465.

Deift P, Li L-C, Nanda T, and Tomei C (1986) The Toda flow on a generic orbit is integrable. *Communications in Pure and Applied Mathematics* 39: 183–232.

Flaschka H (1974) On the Toda lattice II. Inverse scattering solution. *Progress in Theoretical Physics* 51: 703–716.

Guest MA (1997) *Harmonic Maps, Loop Groups, and Integrable Systems*. Cambridge: Cambridge University Press.

Hénon M (1974) Integrals of the Toda lattice. *Physical Review B* 9: 1921–1923.

Krichever IM (1978) Algebraic curves and nonlinear difference equations. *Uspekhi Mat. Nauk* 33: 215–216 (in Russian).

Krichever I (2000) Elliptic analog of the Toda Lattice. *Int. Math. Res. Notes* 8: 383–412.

Kupershmidt BA (1985) Discrete Lax equations and differential-difference calculus. *Asterisque* 123: 212.

Magri F (1978) A simple model of the integrable Hamiltonian equation. *Journal of Mathematical Physics* 19: 1156–1162.

Manakov SV (1974) On the complete integrability and stochastization in discrete dynamical systems. *Zh. Exp. Theoretical Physics* 67: 543–555 (in Russian; English translation: *Soviet Physics JETP* 40: 269–274).

Mikhailov AV (1979) Integrability of a two–dimensional generalization of the Toda chain. *Pis'ma Zh. Eksp. Teor. Fiz.* 30: 443–448 (in Russian; English translation: *JETP Letters* 30: 414–418).

Moser J (1975) Finitely many mass points on the line under the influence of an exponential potential – an integrable system. *Lecture Notes Physics* 38: 467–497.

Olshanetsky MA and Perelomov AM (1979) Explicit solutions of classical generalized Toda models. *Invent. Math.* 54: 261–269.

Reyman AG and Semenov-Tian-Shansky MA (1979–81) Reduction of Hamiltonian systems, affine Lie algebras and Lax equations. I, II. *Invent. Math.* 54: 81–100, 63: 423–432.

Reyman AG and Semenov-Tian-Shansky MA (1994) Group theoretical methods in the theory of finite dimensional integrable systems. In: Arnold VI and Novikov SP (eds.) *Encyclopaedia of Mathematical Science*, vol. 16. *Dynamical Systems VII*, pp. 116–225. Berlin: Springer.

Ruijsenaars SNM (1990) Relativistic Toda systems. *Communications in Mathematical Physics* 133: 217–247.

Sato M and Sato Y (1983) Soliton equations as dynamical systems on infinite-dimensional Grassmann manifold. In: *Nonlinear Partial Differential Equations in Applied Science (Tokyo, 1982)*, pp. 259–271. North-Holland: Amsterdam.

Segal G and Wilson G (1985) Loop groups and equations of KdV type. *Inst. Hautes Études Sci. Publ. Math.* 61: 5–65.

Shabat AB and Yamilov RI (1990) Symmetries on nonlinear chains. *Algebra i Analiz* 2: 183–208 (in Russian; English translation: *Leningrad Mathematical Journal* 2: 377–400).

Sklyanin EK (1987) Boundary conditions for integrable equations. *Funkts. Anal. Prilozh.* 21: 86–87 (in Russian; English translation: *Funct. Anal. Appl.* 21: 164–166).

Suris YB (1997) New integrable systems related to the relativistic Toda lattice. *Journal of Physics A: Math. and Gen.* 30: 1745–1761.

Suris YuB (2003) *The Problem of Integrable Discretization: Hamiltonian Approach*. Basel: Birkhäuser.

Toda M (1967) Vibration of a chain with nonlinear interaction. *Journal of Physical Society Japan* 22: 431–436.

Toda M (1989) *Theory of Nonlinear Lattices*. Berlin: Springer.

Toda M and Wadati M (1975) A canonical transformation for the exponential lattice. *Journal of the Physical Society of Japan* 39: 1204–1211.

Yamilov RI (1982) On the classification of discrete equations. In: *Integrable Systems*, Ufa, 95–114 (in Russian).


# Toeplitz Determinants and Statistical Mechanics

**E L Basor**, California Polytechnic State University, San Luis Obispo, CA, USA



## Introduction

A finite Toeplitz matrix is an $n \times n$ matrix with the following structure:

$$\begin{pmatrix} a_0 & a_{-1} & a_{-2} & \cdots & a_{-n+1} \\ a_1 & a_0 & a_{-1} & \cdots & a_{-n+2} \\ a_2 & a_1 & a_0 & \cdots & a_{-n+3} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n-1} & a_{n-2} & a_{n-3} & \cdots & a_0 \end{pmatrix} \qquad [1]$$

The entries depend on the difference $i - j$ and hence they are constant down all the diagonals. There are two cases when the determinant is easy to compute. One is when the matrix is upper- or lower-triangular and the determinant is $a_0^n$. The other case is when the matrix is of the form

$$\begin{pmatrix} a_0 & a_{n-1} & a_{n-2} & \cdots & a_1 \\ a_1 & a_0 & a_{n-1} & \cdots & a_2 \\ a_2 & a_1 & a_0 & \cdots & a_3 \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n-1} & a_{n-2} & a_{n-3} & \cdots & a_0 \end{pmatrix} \qquad [2]$$

In this latter case, the matrix is called a circulant matrix and the eigenvalues are given by the formula

$$\psi_n(\mathrm{e}^{\mathrm{i}2\pi k/n}), \quad 0 \le k \le n-1$$

where

$$\psi_n(\mathrm{e}^{\mathrm{i}\theta}) = \sum_{j=0}^{n-1} a_i \mathrm{e}^{\mathrm{i}j\theta}$$

The corresponding eigenvector for eigenvalue $\psi_n(e^{i2\pi k/n})$ is

$$[1, e^{i2\pi k/n}, \ldots, e^{i2\pi k(n-1)/n}]$$

This can be verified by direct computation. The role of circulant matrices will not be emphasized in this article, although they are used in the computation of the generating function for certain dimer configurations and also in applications using the discrete Fourier transform.

The most common way to generate a finite Toeplitz matrix is with the Fourier coefficients of an integrable function. Let $\phi : T \to \mathbb{C}$ be a function defined on the unit circle with Fourier coefficients

$$\phi_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi(e^{i\theta})e^{-ik\theta}d\theta \qquad [3]$$

We define $T_n(\phi)$ to be the Toeplitz matrix:

$$T_n(\phi) = (\phi_{i-j})_{i,j=0}^{n-1}$$

A basic problem that in large part has been motivated by statistical mechanics is to determine the behavior of the asymptotics of the determinant of $T_n(\phi)$ as $n \to \infty$. The determinant will be referred to as $D_n(\phi)$, where $\phi$ is called the generating function of the determinant. If the generating function has the property that its Fourier coefficients vanish for negative index (positive index) then the corresponding matrix is lower-triangular (upper-triangular) and hence the determinant is $\phi_0^n$. For other cases, the determinant is not easy to determine and requires additional mathematical machinery.

Some of the primary motivation to study the determinant of these matrices comes from the two-dimensional Ising model. We consider the Onsager lattice in the absence of a magnetic field with sites labeled by

$$(i,j), \quad 0 \le i \le M,\ 0, \le j \le N$$

and with a value $\sigma_{i,j} = \pm 1$ assigned to each site. In the Ising model, $\sigma_{i,j}$ signifies the state of the spin at the site $(i,j)$. To each possible configuration of spins, we define an energy

$$\mathcal{E}(\sigma) = -E_1\sum_{i,j}\sigma_{i,j}\sigma_{i+1,j} - E_2\sum_{i,j}\sigma_{i,j}\sigma_{i,j+1}$$

Let

$$Z = \sum_{\sigma} e^{-\beta\mathcal{E}(\sigma)}$$

be the partition function. Then the probability of a given configuration is

$$\frac{1}{Z}e^{-\beta\mathcal{E}(\sigma)}$$

Here $E_1$, $E_2$, and $\beta = 1/kT$ are, without loss of generality, assumed to be positive constants, $T$ is the temperature, and $k$ is the Boltzmann constant. If $X$ is a random variable defined on the space of configurations, the expectation is given by

$$E(X) = \frac{1}{Z}\sum_{\sigma=\pm 1} X(\sigma)e^{-\beta\mathcal{E}(\sigma)}$$

Let $n$ be fixed for the moment and assume toroidal boundary conditions for the lattice and then let $N, M \to \infty$. It is known that the random variable

$$X(\sigma) = \sigma_{0,0}\sigma_{0,n}$$

has expectation $\langle\sigma_{0,0}\sigma_{0,n}\rangle$ given by $D_n(\phi)$, where

$$\phi(e^{i\theta}) = \left(\frac{(1-\alpha_1 e^{i\theta})(1-\alpha_2 e^{-i\theta})}{(1-\alpha_1 e^{-i\theta})(1-\alpha_2 e^{i\theta})}\right)^{1/2}$$

$$\alpha_1 = z_1\left(\frac{1-z_2}{1+z_2}\right), \quad \alpha_2 = z_1^{-1}\left(\frac{1-z_2}{1+z_2}\right)$$

and

$$z_1 = \tanh\beta E_1, \qquad z_2 = \tanh\beta E_2$$

The square root is taken so that $\phi(e^{i\pi}) = 1$. This formula was first stated by Onsager and later verified in a difficult computation by Montroll, Potts, and Ward.

The spontaneous magnetization $M$ for the Ising model is defined by

$$M^2 = \lim_{n\to\infty}\langle\sigma_{0,0}\sigma_{0,n}\rangle = \lim_{n\to\infty} D_n(\phi)$$

Note that it is the square root of the correlation between two distant sites. Hence, the asymptotics of the Toeplitz determinants will determine whether the magnetization is positive or tends to zero as $n \to \infty$.

## Strong Szegö Limit Theorem

To determine the behavior of the determinants, we need to analyze the generating function $\phi$. Let us first consider the case where $\alpha_2 < 1$. (It is always the case that $0 < \alpha_1 < 1$.) This generating function is differentiable, nonzero and has winding number zero, and it is for functions of this type that a second-order expansion of the Toeplitz determinants can be described. The expansion first formulated by Szegö, in response to the question concerning the

spontaneous magnetization, is called the "strong Szegö limit theorem."

Before proving the Szegö theorem, it should be remarked that we can view the finite Toeplitz matrix as a truncation of an infinite array,

$$\begin{pmatrix} \phi_0 & \phi_{-1} & \phi_{-2} & \\ \phi_1 & \phi_0 & \phi_{-1} & \ddots \\ \phi_2 & \phi_1 & \phi_0 & \ddots \\ & \ddots & \ddots & \ddots \end{pmatrix} \qquad [4]$$

The above infinite array is the matrix representation for the Toeplitz operator

$$T(\phi) : H^2 \to H^2$$

defined by

$$T(\phi)f = P(\phi f)$$

where $H^2$ is the Hardy space

$$\{f \in L^2(T) \mid f_k = 0,\ k < 0\}$$

the function $\phi \in L^\infty(T)$, and $P$ is the orthogonal projection of $L^2(T)$ onto $H^2$. The matrix representation given in [4] is with respect to the Hilbert space basis of $H^2$,

$$\{e^{ik\theta} \mid 0 \le k < \infty\}$$

and $\phi$ is called the symbol of the operator. Now define $P_n : H^2 \to H^2$ by

$$P_n(f_0, f_1, f_2, \ldots) = (f_0, f_1, f_2, \ldots, f_{n-1}, 0, 0, \ldots)$$

The finite Toeplitz matrix can be thought of as the upper-left corner of the array given in [4] or as $P_nT(\phi)P_n$.

To prove the Strong Szegö limit theorem, we introduce the Banach algebra $\mathcal{B}$ of bounded functions $f$ satisfying $\sum_{k=-\infty}^{\infty} |k||f_k|^2 < \infty$.

**Theorem 1** (Strong Szegö limit theorem). *Assume $\phi = \phi_-\phi_+$, where $\phi_\pm$ have logarithms in $\mathcal{B}$. Suppose $\log\phi_-,\ \log\phi_+ \in H^2$. Then*

$$\lim_{n\to\infty} D_n(\phi)/G(\phi)^n = E(\phi) = \exp\left(\sum_{k=1}^{\infty} k s_k s_{-k}\right)$$

*where $G(\phi) = \exp((\log\phi)_0)$ and $s_k = \log\phi_k$.*

Since $\mathcal{B}$ is a Banach algebra, it follows that if $\log\phi_\pm$ belong to $\mathcal{B}$ so do

$$\phi_-,\ \phi_+,\ \phi_+^{-1},\ \phi_-^{-1},\ \phi,\ \phi^{-1}$$

and hence they are bounded. Since $\phi_+$ is in $H^2$ as well, its Fourier coefficients vanish for negative index and the Toeplitz operator has a corresponding infinite array that is lower-triangular. The Fourier coefficients vanish for positive index for $\phi_-$ and hence the infinite array is upper triangular. From this, it follows that

$$T(\phi_+)T(\phi_+^{-1}) = T(\phi_-^{-1})T(\phi_-) = I \qquad [5]$$

$$T(\phi_-)T(\phi_+) = T(\phi) \qquad [6]$$

and

$$\begin{aligned} P_nT(\phi_+) &= P_nT(\phi_+)P_n \\ P_nT(\phi_-)P_n &= T(\phi_-)P_n \end{aligned} \qquad [7]$$

This yields

$$D_n(\phi) = \det T_n(\phi) = \det P_nT(\phi)P_n \qquad [8]$$

$$= \det P_nT(\phi_+)T(\phi_+^{-1})T(\phi)T(\phi_-^{-1})T(\phi_-)P_n \qquad [9]$$

$$= \det P_nT(\phi_+)P_nT(\phi_+^{-1})T(\phi)T(\phi_-^{-1})P_nT(\phi_-)P_n \qquad [10]$$

$$z = \det P_nT(\phi_+)P_n \det(P_nT(\phi_+^{-1})T(\phi)T(\phi_-^{-1})P_n) \times \det P_nT(\phi_-)P_n \qquad [11]$$

The determinants of the right-hand side and the left-hand side of the above expression are $((\phi_\mp)_0)^n$, respectively. Now given the Banach algebra conditions imposed on the symbol $\phi$, it follows that the operator

$$T(\phi_+^{-1})T(\phi)T(\phi_-^{-1})$$

is of the form $I + K$, where $K$ is trace class. Hence, the eigenvalues $\lambda_i$ of $K$ satisfy

$$\sum |\lambda_i| < \infty$$

and the infinite (Fredholm) determinant of $I + K$ is defined. To verify the claim that the operator

$$T(\phi_+^{-1})T(\phi)T(\phi_-^{-1}) = T(\phi_+^{-1})T(\phi_-)T(\phi_+)T(\phi_-^{-1})$$

is $I$ plus a trace class operator, we use the identity

$$T(fg) - T(f)T(g) = H(f)H(\tilde{g}) \qquad [12]$$

where $H(f)$ has matrix form $(f_{i+j+1})_{i,j=0}^{\infty}$, and $\tilde{g}(e^{i\theta}) = g(e^{-i\theta})$. Our Banach algebra conditions show that if $f$ is in $\mathcal{B}$ then the operator $H(f)$ satisfies $\sum_{i,j} |a_{ij}|^2 < \infty$, where the $a_{ij}$ are the matrix entries of the operator. Any operator satisfying this is called a Hilbert–Schmidt operator, and it is known that the product of two Hilbert–Schmidt is trace class. Applying the identity to

$$T(\phi_+^{-1})T(\phi_-)$$

shows that this operator is $T(\phi_+^{-1}\phi_-)$ plus trace class. The operator

$$T(\phi_+)T(\phi_-^{-1})$$

is thus $T(\phi_+\phi_-^{-1})$ plus trace class and one more application of the identity combined with the fact that trace class operators form an ideal yield the desired result.

From the theory of infinite determinants, as $n \to \infty$,

$$\det P_n T(\phi_+^{-1}) T(\phi) T(\phi_-^{-1}) P_n \qquad [13]$$

converges to

$$\det\left(T(\phi_+^{-1}) T(\phi) T(\phi_-^{-1})\right) \qquad [14]$$

At this point, we have proved that

$$\begin{aligned} &\lim_{n\to\infty} D_n(\phi)/((\phi_-)_0)^n((\phi_+)_0)^n \\ &\quad = \lim_{n\to\infty} \det P_n T(\phi_+^{-1}) T(\phi) T(\phi_-^{-1}) P_n \\ &\quad = \det\left(T(\phi_+^{-1}) T(\phi) T(\phi_-^{-1})\right) \qquad [15] \end{aligned}$$

It only remains to identify the constants. To see that

$$G(\phi) = ((\phi_-)_0)^n((\phi_+)_0)^n$$

we note that

$$\begin{aligned} G(\phi) &= \exp((\log\phi)_0) = \exp\left(\frac{1}{2\pi}\int_0^{2\pi} \log\phi(\mathrm{e}^{\mathrm{i}\theta})\mathrm{d}\theta\right) \\ &= \exp\left(\frac{1}{2\pi}\int_0^{2\pi} (\log\phi_-(\mathrm{e}^{\mathrm{i}\theta}) + \log\phi(\mathrm{e}^{\mathrm{i}\theta}))\mathrm{d}\theta\right) \\ &= \exp(\log\phi_-)_0 \exp(\log\phi_+)_0 = (\phi_-)_0(\phi_+)_0 \end{aligned}$$

To compute the determinant of

$$T(\phi_+^{-1}) T(\phi) T(\phi_-^{-1})$$

we write

$$\begin{aligned} &\det T(\phi_+^{-1}) T(\phi) T(\phi_-^{-1}) \\ &\quad = \det T(\phi_+^{-1}) T(\phi_-\phi_+) T(\phi_-^{-1}) \\ &\quad = \det T(\phi_+^{-1}) T(\phi_-) T(\phi_+) T(\phi_-^{-1}) \end{aligned}$$

This last expression is the form

$$\mathrm{e}^A \mathrm{e}^B \mathrm{e}^{-A} \mathrm{e}^{-B}$$

where

$$A = -T(\log\phi_+) \quad \text{and} \quad B = T(\log\phi_-)$$

If $AB - BA$ is trace class then

$$\det \mathrm{e}^A \mathrm{e}^B \mathrm{e}^{-A} \mathrm{e}^{-B} = \mathrm{e}^{\mathrm{tr}(AB-BA)}$$

The operator $AB - BA$ is

$$-T(\log\phi_+)T(\log\phi_-) + T(\log\phi_-)T(\log\phi_+)$$

which equals

$$-T(\log\phi_+)T(\log\phi_-) + T((\log\phi_-)(\log\phi_+))$$

and, by the identity from eqn [12], becomes

$$H(\log\phi_-)H(\log\phi_+)$$

It can be directly computed that

$$\mathrm{tr}(H(\log\phi_-)H(\log\phi_+))$$

equals

$$\sum_{k=1}^{\infty} k s_k s_{-k}$$

and the theorem is proved.

Returning to the Ising model, one needs to compute the asymptotics of the determinants for the generating function

$$\phi(\mathrm{e}^{\mathrm{i}\theta}) = \left(\frac{(1-\alpha_1\mathrm{e}^{\mathrm{i}\theta})(1-\alpha_2\mathrm{e}^{-\mathrm{i}\theta})}{(1-\alpha_1\mathrm{e}^{-\mathrm{i}\theta})(1-\alpha_2\mathrm{e}^{\mathrm{i}\theta})}\right)^{1/2}$$

The term $G(\phi) = 1$ and for $k > 0$

$$k s_k s_{-k} = \frac{1}{4}\left(\frac{-\alpha_1^{2k}}{k} + \frac{-\alpha_2^{2k}}{k} + 2\frac{-\alpha_1^k\alpha_2^k}{k}\right)$$

from which it follows that

$$\lim_{n\to\infty} D_n(\phi) = \left[\frac{(1-\alpha_1^2)(1-\alpha_2^2)}{(1-\alpha_1\alpha_2)^2}\right]^{1/4}$$

Recalling the definition of $\alpha_1$ and $\alpha_2$ yields

$$\lim_{n\to\infty}\langle\sigma_{0,0}\sigma_{0,n}\rangle = \left[1 - \frac{1}{(\sinh 2\beta E_1 \ \sinh 2\beta E_2)^2}\right]^{1/4}$$

or the spontaneous magnetization $M$ as

$$M = \left(1 - \frac{1}{(\sinh 2\beta E_1 \ \sinh 2\beta E_2)^2}\right)^{1/8}$$

In order for this computation to be valid, it was necessary for $0 < \alpha_2 < 1$, and by elementary computations one can show that this is equivalent to the inequality

$$\sinh 2\beta E_1 \sinh 2\beta E_2 > 1$$

## Nonsmooth Symbols or $T = T_c$

A problem occurs in the analysis just outlined when the inequality $0 < \alpha_2 < 1$ does not hold. There are two separate possibilities, $\alpha_2 > 1$ or $\alpha_2 = 1$. First, we consider the latter case. For fixed $E_1$ and $E_2$, this happens for exactly one fixed value of the constant $\beta_c = 1/kT_c$ and the corresponding temperature $T_c$ is called the critical temperature. The "strong Szegö

limit theorem" does not apply since our generating function is of the form

$$\phi(e^{i\theta}) = \left(\frac{(1-\alpha_1 e^{i\theta})(1-e^{-i\theta})}{(1-\alpha_1 e^{-i\theta})(1-e^{i\theta})}\right)^{1/2} \qquad [16]$$

In 1968, Fisher and Hartwig raised a conjecture about $D_n(\phi)$ for nonsmooth $\phi$ which included the above example. They considered generating functions of the form

$$\phi(e^{i\theta}) = \psi(e^{i\theta}) \prod_{j=1}^{R} \phi_{\alpha_j,\beta_j}(e^{i(\theta-\theta_j)}) \qquad [17]$$

where

$$\phi_{\alpha,\beta}(e^{i\theta}) = (2-2\cos\theta)^{\alpha} e^{i\beta(\theta-\pi)}, \quad 0<\theta<2\pi$$

$\Re\alpha > -1/2$, and $\beta$ is not an integer. The function $\psi$ is assumed to be a smooth function. Using the Fisher–Hartwig notation, the symbol of interest in the Ising model from eqn [16] can be written as

$$\psi(e^{i\theta})\phi_{0,-1/2}(e^{i\theta})$$

where

$$\psi(e^{i\theta}) = \left(\frac{1-\alpha_1 e^{i\theta}}{1-\alpha_1 e^{-i\theta}}\right)^{1/2}$$

The conjecture of Fisher and Hartwig for general symbols of this type stated that

$$D_n(\phi) \sim G(\psi)^n n^p E$$

where

$$p = \sum_{r=1}^{R} (\alpha_r^2 - \beta_r^2)$$

and $E^*$ is a constant whose value they did not identify. The constant was later computed to be

$$\begin{aligned} E^*(\phi) = E(\psi) &\prod_{j=1}^{R} \psi_+(e^{i\theta_j})^{-\alpha_j+\beta_j}\psi_-(e^{i\theta_j})^{-\alpha_j-\beta_j} \\ &\times \prod_{1\le s\ne r\le R} (1-e^{i(\theta_s-\theta_r)})^{-(\alpha_s+\beta_s)(\alpha_r-\beta_r)} \\ &\times \prod_{j=1}^{R} \frac{G(1+\alpha_j+\beta_j)G(1+\alpha_j-\beta_j)}{G(1+2\alpha_j)} \end{aligned}$$

where $G(z)$ is the Barnes $G$-function satisfying

$$G(1+z) = \Gamma(z)G(z)$$

and is defined by

$$\begin{aligned} G(1+z) = {} & (2\pi)^{z/2} e^{-(z+1)z/2-\gamma z^2/2} \\ & \times \prod_{k=1}^{\infty}\left(1+\frac{z}{k}\right)^k e^{-z+z^2/2k} \end{aligned}$$

For the above factors, we normalize $\psi$ so that the geometric mean is 1. Then we may assume that the factors $\psi_+, \psi_-(\psi_-\psi_+ = \psi)$ are 1 at zero and infinity, respectively, and this defines the logarithms for the first product. The $E(\psi)$ term is the constant in Szegö's theorem, and the argument of a term of the form $(1-e^{i(\theta_s-\theta_r)})$ is taken between $-\pi/2$ and $\pi/2$.

In the case where $R=1$, the conjecture is known to hold if $\Re\alpha > -1/2$ and the function $b$ satisfies the conditions of Szegö's theorem and is infinitely differentiable. The theorem also has an extension to the case where $\Re\alpha < -1/2$, with $2\alpha$ not an integer, as long as the Fourier coefficients are defined as the coefficients of a distribution.

If we apply the theorem to the generating function from [16]

$$\psi(e^{i\theta})\phi_{0,-1/2}(e^{i\theta}) = \left(\frac{1-\alpha_1 e^{i\theta}}{1-\alpha_1 e^{-i\theta}}\right)^{1/2} \phi_{0,-1/2}(e^{i\theta})$$

we see that the asymptotic expansion is given by

$$n^{-1/4}\left(\frac{1+\alpha_1}{1-\alpha_1}\right)^{1/4} G(1/2)G(3/2)$$

This last formula shows that, at the critical temperature,

$$\lim_{n\to\infty}\langle\sigma_{0,0}\sigma_{0,n}\rangle = \lim_{n\to\infty} D_n(\phi) = 0$$

thus, $M=0$, and hence there is no correlation between distant lattice points.

It should be remarked here that the diagonal correlation at the critical temperature is also given by a singular Toeplitz determinant,

$$\langle\sigma_{0,0}\sigma_{n,n}\rangle = D_n(\phi_{0,-1/2}) \sim n^{-1/4}G(1/2)G(3/2)$$

and thus this limit is also zero.

The proof of the Fisher–Hartwig conjecture is much more complicated than the proof of the "strong Szegö limit theorem." For an indication of how it is proved, note that if we consider the generating function $\phi_{0,\beta}$, the Fourier coefficients are $(\sin\pi\beta)/[\pi(n-\beta)]$ and hence the matrix is Cauchy and the determinant can be computed exactly. From this the asymptotics can be derived and they yield a special case of the Fisher–Hartwig conjecture. The main idea in extending the result to a symbol of the form

$$\psi(e^{i\theta})\phi_{0,\beta}(e^{i\theta})$$

is to prove that the limit of

$$\frac{D_n(\psi\phi_{0,\beta})}{D_n(\psi)D_n(\phi_{0,\beta})}$$

exists. The proof uses much of the same trace-class approach used in proving the "strong Szegö limit theorem," although the results are more complicated. These ideas are then extended for $R > 1$ and also more general $\beta$ and $\alpha$.

It should be noted that in this article the Fisher–Hartwig conjecture does not always hold. If we consider the function

$$\phi(e^{i\theta}) = \begin{cases} -1, & -\pi < \theta < 0 \\ 1, & 0 < \theta < \pi \end{cases}$$

then

$$\phi_n = \begin{cases} 0, & \text{if } k \text{ is even} \\ -2i/(\pi k), & \text{if } k \text{ is odd} \end{cases}$$

The matrix $T_n(\phi)$ is antisymmetric and, if $n$ is odd, $D_n(\phi) = 0$. If $n$ is even, using elementary row and column operations, the determinant can be put in block form with each block of Cauchy type. The determinant can then be evaluated to find

$$D_n(\phi) \sim (i)^n n^{-1/2} K$$

where $K$ is a certain constant.

It is instructive to note that

$$\begin{aligned}\phi(e^{i\theta}) &= \phi_{0,1/2}(e^{i\theta})\phi_{0,-1/2}(e^{i(\theta-\pi)}) \\ &= \phi_{0,-1/2}(e^{i\theta})\phi_{0,1/2}(e^{i(\theta-\pi)})\end{aligned}$$

and thus that this particular symbol has two representations of the type given in [17] and each would give a different asymptotic expansion of the determinant if the conjecture were true for this set of parameters. Hence, it is clear that the conjecture must fail to hold in this case.

However, this example indicates that there might be a generalization of the original conjecture of Fisher and Hartwig. If

$$\phi_{\alpha,\beta}(e^{i(\theta-\gamma)}) = \phi_{\alpha,\beta,\gamma}$$

then

$$\phi = \psi \prod_{j=1}^{R} \phi_{\alpha_j,\beta_j,\theta_j}$$

it is also the case that

$$\phi = \psi^* \prod_{j=1}^{R} \phi_{\alpha_j,\beta_j+n_j,\theta_j}$$

where

$$\sum_{j=1}^{R} n_j = 0 \quad \text{and} \quad \psi^* = \psi \prod_{j=1}^{R} (-e^{i\theta_j})^{n_j}$$

In the example above, $\beta_1 = 1/2$, $\beta_2 = -1/2$, $\theta_1 = 0$, $\theta_2 = \pi$, $n_1 = -1$, and $n_2 = 1$. The result for the counterexample, combined with what is known for the case of integer values of $\alpha$ and $\beta$, leads to the following generalized conjecture. Suppose

$$\psi(e^{i\theta}) = \phi^k \prod_{j=1}^{R} \phi_{\alpha_j^k,\beta_j^k,\theta_j}$$

for some set of indices $k$. Define $Q(k) = \sum_{j=1}^{R} (\alpha_j^k)^2 - (\beta_j^k)^2$. Let $Q = \max_k \Re(Q(k))$ and

$$\mathcal{K} = \{k \mid \Re(Q(k)) = Q\}$$

The generalized asymptotic formula is conjectured to be

$$D_n(\psi) = \sum_{k \in \mathcal{K}} G(\psi^k) n^{Q(k)} E_k^* + o(|G(\phi)|^n n^Q)$$

It may turn out that there is only one element in $\mathcal{K}$ and for these symbols there is a unique representation that yields the highest power in the exponent of the asymptotic expansion. These are the symbols for which the original Fisher–Hartwig conjecture should be true and it is now confirmed in these cases. For example, the conjecture is known to hold for $R > 1$ when $|\Re\alpha_r| < 1/2$ and $|\Re\beta_r| < 1/2$.

## Symbols with Nonzero Index or $T > T_c$

The last possibility in computing the correlation asymptotics is the case where $\alpha_2 > 1$. Note that, for fixed $E_1$ and $E_2$, there is exactly one value of $\beta = 1/kT$ where

$$\alpha_2 = z_1^{-1}\left(\frac{1 - z_2}{1 + z_2}\right) = 1$$

For values of $T > T_c$, we have that the symbol

$$\left(\frac{(1 - \alpha_1 e^{i\theta})(1 - \alpha_2 e^{-i\theta})}{(1 - \alpha_1 e^{-i\theta})(1 - \alpha_2 e^{i\theta})}\right)^{1/2}$$

is the same as

$$e^{-i\theta}\left(\frac{(1 - \alpha_1 e^{i\theta})(1 - (1/\alpha_2)e^{i\theta})}{(1 - \alpha_1 e^{-i\theta})(1 - (1/\alpha_2)e^{-i\theta})}\right)^{1/2}$$

with the argument chosen so that the symbol is positive at $\pi$. Except for the extra factor of $e^{-i\theta}$, this is the same type of smooth symbol that was considered earlier (see the section "Strong Szegö limit theorem"). However, a factor of $e^{i\theta}$ can change the asymptotics considerably as can be seen by considering the simple example of the $\phi \equiv 1$. Fortunately, a variation of the Szegö theorem, first

considered by Fisher and Hartwig, holds for this case of smooth, nonvanishing index.

**Theorem 2** *Suppose that $\phi = \phi_-\phi_+$ satisfies the condition of the "strong Szegö limit theorem" and in addition is at least once continuously differentiable. Then, if $b = \phi_-\phi_+^{-1}$ and $c = \phi_-^{-1}\phi_+$,*

$$D_n(e^{im\theta}\phi) \sim (-1)^{(n-1+m)m} G(\phi)^n E(\phi) G(c)^m \qquad [18]$$

$$\times \left( \det \begin{pmatrix} b_n & \cdots & b_{n-m+1} \\ \vdots & \ddots & \vdots \\ b_{n-1+m} & \cdots & b_n \end{pmatrix} + O(n^{-3}) \right) \times (1 + O(n^{-1})) \qquad [19]$$

Applying this to the symbol

$$e^{-i\theta}\phi(e^{i\theta}) = e^{-i\theta} \left( \frac{(1-\alpha_1 e^{i\theta})(1-(1/\alpha_2)e^{i\theta})}{(1-\alpha_1 e^{-i\theta})(1-(1/\alpha_2)e^{-i\theta})} \right)^{1/2}$$

we have that $m = 1, G(\phi) = -1, G(c) = -1$, and

$$E(\phi) = \left[ (1-\alpha_1^2)\left(1 - \frac{1}{\alpha_2^2}\right)\left(1 - \frac{\alpha_1}{\alpha_2}\right)^2 \right]^{1/4} \qquad [20]$$

The determinant in the above formula is the constant

$$b_n = \frac{1}{2\pi}\int_0^{2\pi} \left( (1-\alpha_1 e^{i\theta})\left(1 - \frac{1}{\alpha_2}e^{i\theta}\right) \times (1-\alpha_1 e^{-i\theta})\left(1 - \frac{1}{\alpha_2}e^{-i\theta}\right) \right)^{-1/2} e^{-in\theta} d\theta$$

The last integral can be deformed to a segment of the real line and evaluated asymptotically to find that the leading term is

$$-\frac{1}{\sqrt{\pi}\alpha_2^n} \left( \left(1 - \frac{\alpha_1}{\alpha_2}\right)(1-\alpha_1\alpha_2)\left(1 - \frac{1}{\alpha_2^2}\right) \right)^{-1/2} \times \frac{\Gamma(n+1/2)}{\Gamma(n+1)}$$

Putting this together with the above constants, we have, for $T > T_c$,

$$\langle \sigma_{0,0}\sigma_{0,n} \rangle \sim \frac{1}{\sqrt{\pi n}\alpha_2^n} \left[ (1-\alpha_1^2)^{1/4} \left(1 - \frac{1}{\alpha_2^2}\right)^{-1/4} (1-\alpha_1\alpha_2)^{-1/2} \right]$$

This implies that the correlation tends to zero very rapidly as $n \to \infty$.

## Further Remarks

The interaction between statistical mechanics and the theory of Toeplitz determinants has a long history, and much of the motivation to describe the asymptotics of the determinants was spurred by the question of spontaneous magnetization in the two-dimensional Ising model. The previous three sections attempt to show how the very different physical situations – $T < T_c, T = T_c$, and $T > T_c$ – all correspond to very different behavior in the symbols of the generating functions. Critical systems predict qualitatively different Szegö type theorems. For example, the phase transition at $T_c$ predicts that the asymptotics for singular symbols cannot be predicted by the smooth symbols, that is, one cannot use continuous functions to approximate the results for singular symbols.

Onsager (1971) was the first to understand that the correlation function could be expressed as a Toeplitz determinant. This was made explicit by Montroll *et al.* (1963). For more information about the Ising model, the reader is referred to McCoy and Wu (1973), where a clear and complete description of the Ising model (and most of the notation used here in reference to this model) can be found.

Szegö (1915, 1952) had originally proved a weak form of the "limit" theorem and he understood that it was desirable to extend to a second-order term. Szegö first proved the "strong Szegö limit theorem" for positive generating functions and this was later extended to the nonpositive case.

The first to understand that a different asymptotic behavior was expected at the critical temperature was Fisher and this resulted in the conjecture for the class of determinants generated by what is now known as Fisher–Hartwig symbols (Fisher and Hartwig 1968). Progress on the conjecture was made by many authors. Böttcher and Silbermann (1998) have provided general results concerning Toeplitz operators and determinants. Additional information about the conjectures of Fisher and Hartwig can be found in Böttcher and Silbermann (1990, 1998), Ehrhardt (2001), and Ehrhardt and Silbermann (1997).

Toeplitz determinants are also important in many other applications. One more recent area of interest is the connection between random-matrix theory and Toeplitz determinants. Many statistical quantities for the circular unitary ensemble can be described as a Toeplitz determinant. For example, the probability of finding no eigenvalues in an interval can be expressed as a Toeplitz determinant. It is also the case that many of the most interesting

statistics correspond to singular symbols. For basic random-matrix theory information see Mehta (1991), and for connections between the circular unitary ensemble and Toeplitz determinants, see Hughes (2001), Tracy and Widom (1993), and Widom (1994).

*See also:* Integrable Systems in Random Matrix Theory; Two-Dimensional Ising Model.

## Further Reading


Böttcher A and Silbermann B (1990) *Analysis of Toeplitz Operators*. Berlin: Akademie-Verlag.

Böttcher A and Silbermann B (1998) *Introduction to Large Truncated Toeplitz Matrices*. Berlin: Springer.

Ehrhardt T (2001) A status report on the asymptotic behavior of Toeplitz determinants with Fisher–Hartwig singularities. *Operator Theory: Advances and Applications* 124: 217–241.

Ehrhardt T and Silbermann B (1997) Toeplitz determinants with one Fisher–Hartwig singularity. *Journal of Functional Analysis* 148: 229–256.

Fisher ME and Hartwig RE (1968) Toeplitz determinants: some applications, theorems, and conjectures. *Advances in Chemical Physics* 15: 333–353.

Hughes C, Keating JP, and O'Connell N (2001) On the characteristic polynomial of a random unitary matrix. *Communications in Mathematical Physics* 220: 429–451.

McCoy BM and Wu TT (1973) *The Two-Dimensional Ising Model*. Cambridge, MA: Harvard University Press.

Metha ML (1991) *Random Matrices*. San Diego: Academic Press.

Montroll EW, Potts RB, and Ward JC (1963) Correlations and spontaneous magnetization of the two-dimensional Ising model. *Journal of Mathematical Physics* 4(2): 308–322.

Onsager L (1971) The Ising model in two dimensions. In: Mills RE, Ascher E, and Jaffee RI (eds.) *Critical Phenomena in Alloys, Magnets, and Superconductors*, pp. 3–12. New York: McGraw-Hill.

Szegö G (1915) Ein Grenzwertsatz über die Toeplitzschen Determinanten einer reellen positiven Funktion. *Mathematische Annalen* 76: 490–503.

Szegö G (1952) *On Certain Hermitian Forms Associated with the Fourier Series of a Positive Function*, pp. 222–238. Lund: Festschrift Marcel Riesz.

Tracy CA and Widom H (1993) Introduction to random matrices. In: Helminck GF (ed.) *Geometric and Quantum Aspects of Integrable Systems*, (Scheveningen, 1992), Lecture Notes in Physics, vol. 424, pp. 103–130. Berlin: Springer.

Widom H (1994) Random Hermitian matrices and (nonrandom) Toeplitz matrices. In: Basor E and Gohberg I (eds.) *Toeplitz Operators and Related Topics* (Santa Cruz, CA, 1992), Oper. Theory Adv. Appl., vol. 71, pp. 9–15. Basel: Birkhuser.


# Tomita–Takesaki Modular Theory

**S J Summers**, University of Florida, Gainesville, FL, USA



## Basic Structure

The origins of Tomita–Takesaki modular theory lie in two unpublished papers of M Tomita in 1967 and a slim volume by Takesaki (1970). It has developed into one of the most important tools in the theory of operator algebras and has found many applications in mathematical physics.

Although the modular theory has been formulated in a more general setting, it will be presented in the form in which it most often finds application in mathematical physics (for generalizations, details, and further references concerning the material covered in this article, the reader is referred to the Further Reading section). Let $\mathcal{M}$ be a von Neumann algebra on a Hilbert space $\mathcal{H}$ containing a vector $\Omega$ which is cyclic and separating for $\mathcal{M}$. Define the operator $S_0$ on $\mathcal{H}$ as follows:

$$S_0 A\Omega = A^*\Omega, \quad \text{for all } A \in \mathcal{M}$$

This operator extends to a closed antilinear operator $S$ defined on a dense subset of $\mathcal{H}$. Let $\Delta$ be the unique positive, self-adjoint operator and $J$ the unique antiunitary operator occurring in the polar decomposition

$$S = J\Delta^{1/2} = \Delta^{-1/2}J$$

$\Delta$ is called the modular operator and $J$ the modular conjugation (or modular involution) associated with the pair $(\mathcal{M}, \Omega)$. Note that $J^2$ is the identity operator and $J = J^*$. Moreover, the spectral calculus may be applied to $\Delta$ so that $\Delta^{it}$ is a unitary operator for each $t \in \mathbb{R}$ and $\{\Delta^{it} \mid t \in \mathbb{R}\}$ forms a strongly continuous unitary group. Let $\mathcal{M}'$ denote the set of all bounded linear operators on $\mathcal{H}$ which commute with all elements of $\mathcal{M}$. The modular theory begins with the following remarkable theorem.

**Theorem 1** *Let $\mathcal{M}$ be a von Neumann algebra with a cyclic and separating vector $\Omega$. Then $J\Omega = \Omega = \Delta\Omega$, and the following equalities hold:*

$$J\mathcal{M}J = \mathcal{M}'$$

*and*

$$\Delta^{it}\mathcal{M}\Delta^{-it} = \mathcal{M}, \quad \text{for all } t \in \mathbb{R}$$

Note that if one defines $F_0 A'\Omega = A'^*\Omega$, for all $A' \in \mathcal{M}'$, and takes its closure $F$, then one has the relations

$$\Delta = FS, \quad \Delta^{-1} = SF, \quad F = J\Delta^{-1/2}$$

## Modular Automorphism Group

By Theorem 1, the unitaries $\Delta^{it}, t \in \mathbb{R}$, induce a one-parameter automorphism group $\{\sigma_t\}$ of $\mathcal{M}$ by

$$\sigma_t(A) = \Delta^{it} A \Delta^{-it}, \quad A \in \mathcal{M}, \quad t \in \mathbb{R}$$

This group is called the modular automorphism group of $\mathcal{M}$ (relative to $\Omega$). Let $\omega$ denote the faithful normal state on $\mathcal{M}$ induced by $\Omega$:

$$\omega(A) = \frac{1}{\|\Omega\|^2} \langle \Omega, A\Omega \rangle, \quad A \in \mathcal{M}$$

From Theorem 1 it follows that $\omega$ is invariant under $\{\sigma_t\}$, that is, $\omega(\sigma_t(A)) = \omega(A)$ for all $A \in \mathcal{M}$ and $t \in \mathbb{R}$.

The modular automorphism group contains information about both $\mathcal{M}$ and $\omega$. For example, the modular automorphism group is an inner automorphism on $\mathcal{M}$ if and only if $\mathcal{M}$ is semifinite. It is trivial if and only if $\omega$ is a tracial state on $\mathcal{M}$. Indeed, for any $B \in \mathcal{M}$, one has $\sigma_t(B) = B$ for all $t \in \mathbb{R}$ if and only if $\omega(AB) = \omega(BA)$ for all $A \in \mathcal{M}$. Let $\mathcal{M}^\sigma$ denote the set of all such $B$ in $\mathcal{M}$.

### The KMS Condition

The modular automorphism group satisfies a condition which had already been used in mathematical physics to characterize equilibrium temperature states of quantum systems in statistical mechanics and field theory – the Kubo–Martin–Schwinger (KMS) condition. If $\mathcal{M}$ is a von Neumann algebra and $\{\alpha_t \mid t \in \mathbb{R}\}$ is a $\sigma$-weakly continuous one-parameter group of automorphisms of $\mathcal{M}$, then the state $\phi$ on $\mathcal{M}$ satisfies the KMS condition at (inverse temperature) $\beta$ $(0 < \beta < \infty)$ with respect to $\{\alpha_t\}$ if for any $A,B \in \mathcal{M}$ there exists a complex function $F_{AB}(z)$ which is analytic on the strip $\{z \in \mathbb{C} \mid 0 < \operatorname{Im} z < \beta\}$ and continuous on the closure of this strip such that

$$F_{A,B}(t) = \phi(\alpha_t(A)B)$$
$$F_{A,B}(t + i\beta) = \phi(B\alpha_t(A))$$

for all $t \in \mathbb{R}$. In this case, $\phi(\alpha_{i\beta}(A)B) = \phi(BA)$, for all $A, B$ in a $\sigma$-weakly dense, $\alpha$-invariant $*$-subalgebra of $\mathcal{M}$. Such KMS states are $\alpha$-invariant, that is, $\phi(\alpha_t(A)) = \phi(A)$, for all $A \in \mathcal{M}, t \in \mathbb{R}$, and are stable and passive (cf. Bratteli and Robinson (1981) and Haag (1992)).

Every faithful normal state satisfies the KMS condition at $\beta = 1$ (henceforth called the modular condition) with respect to the corresponding modular automorphism group.

**Theorem 2** *Let $\mathcal{M}$ be a von Neumann algebra with a cyclic and separating vector $\Omega$. Then the induced state $\omega$ on $\mathcal{M}$ satisfies the modular condition with respect to the modular automorphism group $\{\sigma_t \mid t \in \mathbb{R}\}$ associated to the pair $(\mathcal{M}, \Omega)$.*

The modular automorphism group is, therefore, endowed with the analyticity associated with the KMS condition, and this is a powerful tool in many applications of the modular theory to mathematical physics. In addition, the physical properties and interpretations of KMS states are often invoked when applying modular theory to quantum physics.

Note that while the nontriviality of the modular automorphism group gives a measure of the nontracial nature of the state, the KMS condition for the modular automorphism group provides the missing link between the values $\omega(AB)$ and $\omega(BA)$, for all $A,B \in \mathcal{M}$ (hence the use of the term "modular," as in the theory of integration on locally compact groups).

The modular condition is quite restrictive. Only the modular group can satisfy the modular condition for $(\mathcal{M}, \Omega)$, and the modular group for one state can satisfy the modular condition only in states differing from the original state by the action of an element in the center of $\mathcal{M}$.

**Theorem 3** *Let $\mathcal{M}$ be a von Neumann algebra with a cyclic and separating vector $\Omega$, and let $\{\sigma_t\}$ be the corresponding modular automorphism group. If the induced state $\omega$ satisfies the modular condition with respect to a group $\{\alpha_t\}$ of automorphisms of $\mathcal{M}$, then $\{\alpha_t\}$ must coincide with $\{\sigma_t\}$. Moreover, a normal state $\psi$ on $\mathcal{M}$ satisfies the modular condition with respect to $\{\sigma_t\}$ if and only if $\psi(\,\cdot\,) = \omega(h\,\cdot\,) = \omega(h^{1/2} \cdot h^{1/2})$ for some unique positive injective operator $h$ affiliated with the center of $\mathcal{M}$.*

Hence, if $\mathcal{M}$ is a factor, two distinct states cannot share the same modular automorphism group. The relation between the modular automorphism groups for two different states will be described in more detail.

### One Algebra and Two States

Consider a von Neumann algebra $\mathcal{M}$ with two cyclic and separating vectors $\Omega$ and $\Phi$, and denote by $\omega$ and $\phi$, respectively, the induced states on $\mathcal{M}$. Let $\{\sigma_t^\omega\}$ and $\{\sigma_t^\phi\}$ denote the corresponding modular groups. There is a general relation between the modular automorphism groups of these states.

**Theorem 4** *There exists a σ-strongly continuous map* $\mathbb{R} \ni t \mapsto U_t \in \mathcal{M}$ *such that*

(i) $U_t$ *is unitary for all* $t \in \mathbb{R}$;
(ii) $U_{t+s} = U_t \sigma_t^{\omega}(U_s)$ *for all* $s,t \in \mathbb{R}$; *and*
(iii) $\sigma_t^{\phi}(A) = U_t \sigma_t^{\omega}(A) U_t^*$ *for all* $A \in \mathcal{M}$ *and* $t \in \mathbb{R}$.

The 1-cocycle $\{U_t\}$ is commonly called the cocycle derivative of $\phi$ with respect to $\omega$ and one writes $U_t = (D\phi : D\omega)_t$. There is a chain rule for this derivative, as well: If $\phi,\psi$, and $\rho$ are faithful normal states on $\mathcal{M}$, then $(D\psi : D\phi)_t = (D\psi : D\rho)_t (D\rho : D\phi)_t$, for all $t \in \mathbb{R}$. More can be said about the cocycle derivative if the states satisfy any of the conditions in the following theorem.

**Theorem 5** *The following conditions are equivalent*:

(i) $\phi$ *is* $\{\sigma_t^{\omega}\}$*-invariant*;
(ii) $\omega$ *is* $\{\sigma_t^{\phi}\}$*-invariant*;
(iii) *there exists a unique positive injective operator* $h$ *affiliated with* $\mathcal{M}^{\sigma^{\omega}} \cap \mathcal{M}^{\sigma^{\phi}}$ *such that* $\omega(\,\cdot\,) = \phi(h\,\cdot\,) = \phi(h^{1/2}\,\cdot\,h^{1/2})$;
(iv) *there exists a unique positive injective operator* $h'$ *affiliated with* $\mathcal{M}^{\sigma^{\omega}} \cap \mathcal{M}^{\sigma^{\phi}}$ *such that* $\phi(\,\cdot\,) = \omega(h'\,\cdot\,) = \omega(h'^{1/2}\,\cdot\,h'^{1/2})$;
(v) *the norms of the linear functionals* $\omega + \mathrm{i}\phi$ *and* $\omega - \mathrm{i}\phi$ *are equal*; *and*
(vi) $\sigma_t^{\omega}\sigma_s^{\phi} = \sigma_s^{\phi}\sigma_t^{\omega}$, *for all* $s,t \in \mathbb{R}$.

The conditions in Theorem 5 turn out to be equivalent to the cocycle derivative being a representation.

**Theorem 6** *The cocycle* $\{U_t\}$ *intertwining* $\{\sigma_t^{\omega}\}$ *with* $\{\sigma_t^{\phi}\}$ *is a group representation of the additive group of reals if and only if* $\phi$ *and* $\omega$ *satisfy the conditions in* Theorem 5. *In that case*, $U(t) = h^{-\mathrm{i}t}$.

The operator $h' = h^{-1}$ in Theorem 5 is called the Radon–Nikodym derivative of $\phi$ with respect to $\omega$ (often denoted by $\mathrm{d}\phi/\mathrm{d}\omega$), due to the following result, which, if the algebra $\mathcal{M}$ is abelian, *is* the well-known Radon–Nikodym theorem from measure theory.

**Theorem 7** *If* $\phi$ *and* $\omega$ *are normal positive linear functionals on* $\mathcal{M}$ such that $\phi(A) \le \omega(A)$, *for all positive elements* $A \in \mathcal{M}$, *then there exists a unique element* $h^{1/2} \in \mathcal{M}$ *such that* $\phi(\,\cdot\,) = \omega(h^{1/2}\,\cdot\,h^{1/2})$ *and* $0 \le h^{1/2} \le 1$.

The analogies with measure theory are not accidental, although these are not discussed in detail here. Indeed, any normal trace on a (finite) von Neumann algebra $\mathcal{M}$ gives rise to a noncommutative integration theory in a natural manner. Modular theory affords an extension of this theory to the setting of faithful normal functionals $\eta$ on von Neumann algebras $\mathcal{M}$ of any type, enabling the definition of noncommutative $L^p$ spaces, $L^p(\mathcal{M},\eta)$.

## Modular Invariants and the Classification of von Neumann Algebras

As already mentioned, the modular structure carries information about the algebra. This is best evidenced in the structure of type III factors. As this theory is rather involved, only a sketch of some of the results can be given.

If $\mathcal{M}$ is a type III algebra, then its crossed product $\mathcal{N} = \mathcal{M} \rtimes_{\sigma^{\omega}} \mathbb{R}$ relative to the modular automorphism group of any faithful normal state $\omega$ on $\mathcal{M}$ is a type $\mathrm{II}_{\infty}$ algebra with a faithful semifinite normal trace $\tau$ such that $\tau \circ \theta_t = \mathrm{e}^{-t}\tau$, $t \in \mathbb{R}$, where $\theta$ is the dual of $\sigma^{\omega}$ on $\mathcal{N}$. Moreover, the algebra $\mathcal{M}$ is isomorphic to the cross product $\mathcal{N} \rtimes_{\theta} \mathbb{R}$, and this decomposition is unique in a very strong sense. This structure theorem entails the existence of important algebraic invariants for $\mathcal{M}$, which has many consequences, one of which is made explicit here.

If $\omega$ is a faithful normal state of a von Neumann algebra $\mathcal{M}$ induced by $\Omega$, let $\Delta_{\omega}$ denote the modular operator associated to $(\mathcal{M},\Omega)$ and $\mathrm{sp}\,\Delta_{\omega}$ denote the spectrum of $\Delta_{\omega}$. The intersection

$$S'(\mathcal{M}) = \cap\ \mathrm{sp}\,\Delta_{\omega}$$

over all faithful normal states $\omega$ of $\mathcal{M}$ is an algebraic invariant of $\mathcal{M}$.

**Theorem 8** *Let* $\mathcal{M}$ *be a factor acting on a separable Hilbert space. If* $\mathcal{M}$ *is of type* III, *then* $0 \in S'(\mathcal{M})$; *otherwise*, $S'(\mathcal{M}) = \{0,1\}$ *if* $\mathcal{M}$ *is of type* $\mathrm{I}_{\infty}$ *or* $\mathrm{II}_{\infty}$ *and* $S'(\mathcal{M}) = \{1\}$ *if not. Let* $\mathcal{M}$ *now be a factor of type* III.

(i) $\mathcal{M}$ *is of type* $\mathrm{III}_{\lambda}, 0 < \lambda < 1$, *if and only if* $S'(\mathcal{M}) = \{0\} \cup \{\lambda^n \mid n \in \mathbb{Z}\}$.
(ii) $\mathcal{M}$ *is of type* $\mathrm{III}_0$ *if and only if* $S'(\mathcal{M}) = \{0,1\}$.
(iii) $\mathcal{M}$ *is of type* $\mathrm{III}_1$ *if and only if* $S'(\mathcal{M}) = [0,\infty)$.

In certain physically relevant situations, the spectra of the modular operators of all faithful normal states coincide, so that Theorem 8 entails that it suffices to compute the spectrum of any conveniently chosen modular operator in order to determine the type of $\mathcal{M}$. In other such situations, there are distinguished states $\omega$ such that $S'(\mathcal{M}) = \mathrm{sp}\,\Delta_{\omega}$. One such example is provided by asymptotically abelian systems. A von Neumann algebra $\mathcal{M}$ is said to be "asymptotically abelian" if there exists a sequence $\{\alpha_n\}_{n\in\mathbb{N}}$ of automorphisms of $\mathcal{M}$ such that the limit of $\{A\alpha_n(B) - \alpha_n(B)A\}_{n\in\mathbb{N}}$ in

the strong operator topology is zero, for all $A, B \in \mathcal{M}$. If the state $\omega$ is $\alpha_n$-invariant, for all $n \in \mathbb{N}$, then $\mathrm{sp}\,\Delta_\omega$ is contained in $\mathrm{sp}\,\Delta_\phi$, for all faithful normal states $\phi$ on $\mathcal{M}$, so that $S'(\mathcal{M}) = \mathrm{sp}\,\Delta_\omega$. If, moreover, $\mathrm{sp}\,\Delta_\omega = [0, \infty)$, then $\mathrm{sp}\,\Delta_\omega = \mathrm{sp}\,\Delta_\phi$, for all $\phi$ as described.

## Self-Dual Cones

Let $j : \mathcal{M} \to \mathcal{M}'$ denote the antilinear $*$-isomorphism defined by $j(A) = JAJ, A \in \mathcal{M}$. The natural positive cone $\mathcal{P}^\natural$ associated with the pair $(\mathcal{M}, \Omega)$ is defined as the closure, in $\mathcal{H}$, of the set of vectors

$$\{Aj(A)\Omega \mid A \in \mathcal{M}\}$$

Let $\mathcal{M}_+$ denote the set of all positive elements of $\mathcal{M}$. The following theorem collects the main attributes of the natural cone.

**Theorem 9**

(i) *$\mathcal{P}^\natural$ coincides with the closure in $\mathcal{H}$ of the set $\{\Delta^{1/4} A\Omega \mid A \in \mathcal{M}_+\}$.*
(ii) *$\Delta^{it}\mathcal{P}^\natural = \mathcal{P}^\natural$ for all $t \in \mathbb{R}$.*
(iii) *$J\Phi = \Phi$ for all $\Phi \in \mathcal{P}^\natural$.*
(iv) *$Aj(A)\mathcal{P}^\natural \subset \mathcal{P}^\natural$ for all $A \in \mathcal{M}$.*
(v) *$\mathcal{P}^\natural$ is a pointed, self-dual cone whose linear span coincides with $\mathcal{H}$.*
(vi) *If $\Phi \in \mathcal{P}^\natural$, then $\Phi$ is cyclic for $\mathcal{M}$ if and only if $\Phi$ is separating for $\mathcal{M}$.*
(vii) *If $\Phi \in \mathcal{P}^\natural$ is cyclic, and hence separating, for $\mathcal{M}$, then the modular conjugation and the natural cone associated with the pair $(\mathcal{M}, \Phi)$ coincide with $J$ and $\mathcal{P}^\natural$, respectively.*
(viii) *For every normal positive linear functional $\phi$ on $\mathcal{M}$, there exists a unique vector $\Phi_\phi \in \mathcal{P}^\natural$ such that $\phi(A) = \langle \Phi_\phi, A\Phi_\phi \rangle$ for all $A \in \mathcal{M}$.*

In fact, the algebras $\mathcal{M}$ and $\mathcal{M}'$ are uniquely characterized by the natural cone $\mathcal{P}^\natural$ [4]. In light of (viii), if $\alpha$ is an automorphism of $\mathcal{M}$, then

$$V(\alpha)\Phi_\phi = \Phi_{\phi \circ \alpha^{-1}}$$

defines an isometric operator on $\mathcal{P}^\natural$, which by (v) extends to a unitary operator on $\mathcal{H}$. The map $\alpha \mapsto V(\alpha)$ defines a unitary representation of the group of automorphisms $\mathrm{Aut}(\mathcal{M})$ on $\mathcal{M}$ in such a manner that $V(\alpha)AV(\alpha)^{-1} = \alpha(A)$ for all $A \in \mathcal{M}$ and $\alpha \in \mathrm{Aut}(\mathcal{M})$. Indeed, one has the following:

**Theorem 10** *Let $\mathcal{M}$ be a von Neumann algebra with a cyclic and separating vector $\Omega$. The group $\mathcal{V}$ of all unitaries $V$ satisfying*

$$V\mathcal{M}V^* = \mathcal{M}, \quad VJV^* = J, \quad V\mathcal{P}^\natural = \mathcal{P}^\natural$$

*is isomorphic to* $\mathrm{Aut}(\mathcal{M})$ *under the above map $\alpha \mapsto V(\alpha)$, which is called the "standard implementation" of* $\mathrm{Aut}(\mathcal{M})$.

Often of particular physical interest are (anti-)automorphisms of $\mathcal{M}$ leaving $\omega$ invariant. They can only be implemented by (anti)unitaries which leave the pair $(\mathcal{M}, \Omega)$ invariant. In fact, if $U$ is a unitary or antiunitary operator satisfying $U\Omega = \Omega$ and $U\mathcal{M}U^* = \mathcal{M}$, then $U$ commutes with both $J$ and $\Delta$.

## Two Algebras and One State

Motivated by applications to quantum field theory, the study of the modular structures associated with one state and more than one von Neumann algebra has begun (see Borchers (2000) for references and details). Let $\mathcal{N} \subset \mathcal{M}$ be von Neumann algebras with a common cyclic and separating vector $\Omega$, and $\Delta_\mathcal{N}$, $J_\mathcal{N}$ and $\Delta_\mathcal{M}, J_\mathcal{M}$ denote the corresponding modular objects. The structure $(\mathcal{M}, \mathcal{N}, \Omega)$ is called a $\pm$-half-sided modular inclusion if $\Delta_\mathcal{M}^{it}\mathcal{N}\Delta_\mathcal{M}^{-it} \subset \mathcal{N}$, for all $\pm t \geq 0$.

**Theorem 11** *Let $\mathcal{M}$ be a von Neumann algebra with cyclic and separating vector $\Omega$. The following are equivalent:*

(i) *There exists a proper subalgebra $\mathcal{N} \subset \mathcal{M}$ such that $(\mathcal{M}, \mathcal{N}, \Omega)$ is a $\mp$-half-sided modular inclusion.*
(ii) *There exists a unitary group $\{U(t)\}$ with positive generator such that*

$$U(t)\mathcal{M}U(t)^{-1} \subset \mathcal{M}, \quad \text{for all} \pm t \geq 0,$$
$$U(t)\Omega = \Omega, \quad \text{for all } t \in \mathbb{R}$$

*Moreover, if these conditions are satisfied, then the following relations must hold:*

$$\Delta_\mathcal{M}^{it}U(s)\Delta_\mathcal{M}^{-it} = \Delta_\mathcal{N}^{it}U(s)\Delta_\mathcal{N}^{-it} = U(\mathrm{e}^{\mp 2\pi t}s)$$

*and*

$$J_\mathcal{M}U(s)J_\mathcal{M} = J_\mathcal{N}U(s)J_\mathcal{N} = U(-s)$$

*for all $s,t \in \mathbb{R}$. In addition, $\mathcal{N} = U(\pm 1)\mathcal{M}U(\pm 1)^{-1}$, and if $\mathcal{M}$ is a factor, it must be type* $\mathrm{III}_1$.

The richness of this structure is further suggested by the next theorem.

**Theorem 12**

(i) *Let $(\mathcal{M}, \mathcal{N}_1, \Omega)$ and $(\mathcal{M}, \mathcal{N}_2, \Omega)$ be $-$-half-sided, resp. $+$-half-sided, modular inclusions satisfying the condition $J_{\mathcal{N}_1}J_{\mathcal{N}_2} = J_\mathcal{M}J_{\mathcal{N}_2}J_{\mathcal{N}_1}J_\mathcal{M}$. Then the modular unitaries $\Delta_\mathcal{M}^{it}, \Delta_{\mathcal{N}_1}^{is}, \Delta_{\mathcal{N}_2}^{iu}, s, t, u \in \mathbb{R}$, generate a faithful continuous unitary representation of the identity component of the*

*group of isometries of two-dimensional Minkowski space.*

(ii) *Let $\mathcal{M}, \mathcal{N}, \mathcal{N} \cap \mathcal{M}$ be von Neumann algebras with a common cyclic and separating vector $\Omega$. If $(\mathcal{M}, \mathcal{M} \cap \mathcal{N}, \Omega)$ and $(\mathcal{N}, \mathcal{M} \cap \mathcal{N}, \Omega)$ are $-$-half-sided, resp. $+$-half-sided, modular inclusions such that $J_{\mathcal{N}} \mathcal{M} J_{\mathcal{N}} = \mathcal{M}$, then the modular unitaries $\Delta^{it}_{\mathcal{M}}, \Delta^{is}_{\mathcal{N}}, \Delta^{iu}_{\mathcal{N} \cap \mathcal{M}}, s, t, u \in \mathbb{R}$, generate a faithful continuous unitary representation of* SL $(2, \mathbb{R})/\mathbb{Z}_2$.

This has led to a further useful notion. If $\mathcal{N} \subset \mathcal{M}$ and $\Omega$ is cyclic for $\mathcal{N} \cap \mathcal{M}$, then $(\mathcal{M}, \mathcal{N}, \Omega)$ is said to be a "$\pm$-modular intersection" if both $(\mathcal{M}, \mathcal{M} \cap \mathcal{N}, \Omega)$ and $(\mathcal{N}, \mathcal{M} \cap \mathcal{N}, \Omega)$ are $\pm$-half-sided modular inclusions and

$$J_{\mathcal{N}} \left[ \lim_{t \to \mp\infty} \Delta^{it}_{\mathcal{N}} \Delta^{-it}_{\mathcal{M}} \right] J_{\mathcal{N}} = \lim_{t \to \mp\infty} \Delta^{it}_{\mathcal{M}} \Delta^{-it}_{\mathcal{N}}$$

where the existence of the strong operator limits is assured by the preceding assumptions. An example of the utility of this structure is the following theorem.

**Theorem 13** *Let $\mathcal{N}, \mathcal{M}, \mathcal{L}$ be von Neumann algebras with a common cyclic and separating vector $\Omega$. If $(\mathcal{M}, \mathcal{N}, \Omega)$ and $(\mathcal{N}', \mathcal{L}, \Omega)$ are $-$-modular intersections and $(\mathcal{M}, \mathcal{L}, \Omega)$ is a $+$-modular intersection, then the unitaries $\Delta^{it}_{\mathcal{M}}, \Delta^{is}_{\mathcal{N}}, \Delta^{iu}_{\mathcal{L}}, s, t, u \in \mathbb{R}$, generate a faithful continuous unitary representation of* $\mathrm{SO}^{\uparrow}(1, 2)$.

These results and their extensions to larger numbers of algebras were developed for application in algebraic quantum field theory, but one may anticipate that half-sided modular inclusions will find wider use. Modular theory has also been applied fruitfully in the theory of inclusions $\mathcal{N} \subset \mathcal{M}$ of properly infinite algebras with finite or infinite index.

## Applications in Quantum Theory

The Tomita–Takesaki theory has found many applications in quantum field theory and quantum statistical mechanics. As mentioned earlier, the modular automorphism group satisfies the KMS condition, a property of physical significance in the quantum theory of many-particle systems, which includes quantum statistical mechanics and quantum field theory. In such settings, for a suitable algebra of observables $\mathcal{M}$ and state $\omega$, an automorphism group $\{\sigma_{\beta t}\}$ representing the time evolution of the system satisfies the modular condition. Hence, on the one hand, $\{\sigma_{\beta t}\}$ is the modular automorphism group of the pair $(\mathcal{M}, \Omega)$, and, on the other, $\omega$ is an equilibrium state at inverse temperature $\beta$, with all the consequences which both of these facts have.

But it has become increasingly clear that the modular objects $\Delta^{it}$, $J$, of certain algebras of observables and states encode additional physical information. In 1975, it was discovered that if one considers the algebras of observables associated with a finite-component quantum field theory satisfying the Wightman axioms, then the modular objects associated with the vacuum state and algebras of observables localized in certain wedge-shaped regions in Minkowski space have geometric content. In fact, the unitary group $\{\Delta^{it}\}$ implements the group of Lorentz boosts leaving the wedge region invariant (this property is now called modular covariance), and the modular involution $J$ implements the space-time reflection about the edge of the wedge, along with a charge conjugation. This discovery caused some intense research activity (see Baumgartel and Wollenberg 1992, Borchers 2000, Haag 1992).

### Positive Energy

In quantum physics the time development of the system is often represented by a strongly continuous group $\{U(t) = e^{itH} \mid t \in \mathbb{R}\}$ of unitary operators, and the generator $H$ is interpreted as the total energy of the system. There is a link between modular structure and positive energy, which has found many applications in quantum field theory. This result was crucial in the development of Theorem 11 and was motivated by the 1975 discovery mentioned above, now commonly called the Bisognano–Wichmann theorem.

**Theorem 14** *Let $\mathcal{M}$ be a von Neumann algebra with a cyclic and separating vector $\Omega$, and let $\{U(t)\}$ be a continuous unitary group satisfying $U(t)\mathcal{M}U(-t) \subset \mathcal{M}$, for all $t \geq 0$. Then any two of the following conditions imply the third*:

(i) $U(t) = e^{itH}$, *with* $H \geq 0$;
(ii) $U(t)\Omega = \Omega$, *for all* $t \in \mathbb{R}$; *and*
(iii) $\Delta^{it} U(s) \Delta^{-it} = U(e^{-2\pi t} s)$ *and* $JU(s)J = U(-s)$, *for all* $s, t \in \mathbb{R}$.

### Modular Nuclearity and Phase Space Properties

Modular theory can be used to express physically meaningful properties of quantum "phase spaces" by a condition of compactness or nuclearity of certain maps. In its initial form, the condition was formulated in terms of the Hamiltonian, the global energy operator of theories in Minkowski space. The above indications that the modular operators carry information about the energy of the system were reinforced when it was shown that a

formulation in terms of modular operators was essentially equivalent.

Let $\mathcal{O}_1 \subset \mathcal{O}_2$ be nonempty bounded open subregions of Minkowski space with corresponding algebras of observables $\mathcal{A}(\mathcal{O}_1) \subset \mathcal{A}(\mathcal{O}_2)$ in a vacuum representation with vacuum vector $\Omega$, and let $\Delta$ be the modular operator associated with $(\mathcal{A}(\mathcal{O}_2), \Omega)$ (by the Reeh–Schlieder theorem, $\Omega$ is cyclic and separating for $\mathcal{A}(\mathcal{O}_2)$). For each $\lambda \in (0, 1/2)$ define the mapping $\Xi_\lambda : \mathcal{A}(\mathcal{O}_1) \to \mathcal{H}$ by $\Xi_\lambda(A) = \Delta^\lambda A\Omega$. The compactness of any one of these mappings implies the compactness of all of the others. Moreover, the $l^p$ (nuclear) norms of these mappings are interrelated and provide a measure of the number of local degrees of freedom of the system. Suitable conditions on the maps in terms of these norms entail the strong statistical independence condition called the split property. Conversely, the split property implies the compactness of all of these maps. Moreover, the existence of equilibrium temperature states on the global algebra of observables can be derived from suitable conditions on these norms in the vacuum sector.

The conceptual advantage of the modular compactness and nuclearity conditions compared to their original Hamiltonian form lies in the fact that they are meaningful also for quantum systems in curved spacetimes, where global energy operators (i.e., generators corresponding to global timelike Killing vector fields) need not exist.

## Modular Position and Quantum Field Theory

The characterization of the relative "geometric" position of algebras based on the notions of modular inclusion and modular intersection was directly motivated by the Bisognano–Wichmann theorem. Observable algebras associated with suitably chosen wedge regions in Minkowski space provided examples whose essential structure could be abstracted for more general application, resulting in the notions presented in the preceding sections.

Theorem 12(ii) has been used to construct, from two algebras and the indicated half-sided modular inclusions, a conformal quantum field theory on the circle (compactified light ray) with positive energy. Since the chiral part of a conformal quantum field model in two spacetime dimensions naturally yields such half-sided modular inclusions, studying the inclusions in Theorem 12(ii) is equivalent to studying such field theories. Theorems 12(i) and 13 and their generalizations to inclusions involving up to six algebras have been employed to construct Poincaré-covariant nets of observable algebras (the algebraic form of quantum field theories) satisfying the spectrum condition on $(d+1)$-dimensional Minkowski space for $d = 1, 2, 3$. Conversely, such quantum field theories naturally yield such systems of algebras.

This intimate relation would seem to open up the possibility of constructing interacting quantum field theories from a limited number of modular inclusions/intersections.

## Geometric Modular Action

The fact that the modular objects in quantum field theory associated with wedge-shaped regions and the vacuum state in Minkowski space have geometric significance ("geometric modular action") was originally discovered in the framework of the Wightman axioms. As an algebraic quantum field theory (AQFT) does not rely on the concept of Wightman fields, it was natural to ask (i) when does geometric modular action hold in AQFT and (ii) which physically relevant consequences follow from this feature?

There are two approaches to the study of geometric modular action. In the first, attention is focused on modular covariance, expressed in terms of the modular groups associated with wedge algebras and the vacuum state in Minkowski space. Modular covariance has been proven to obtain in conformally invariant AQFT, in any massive theory satisfying asymptotic completeness, and also in the presence of other, physically natural assumptions. To mention only three of its consequences, both the spin–statistics theorem and the PCT theorem, as well as the existence of a continuous unitary representation of the Poincaré group acting covariantly upon the observable algebras and satisfying the spectrum condition follow from modular covariance.

In a second approach to geometric modular action, the modular involutions are the primary focus. Here, no *a priori* connection between the modular objects and isometries of the spacetime is assumed. The central assumption, given the state vector $\Omega$ and the von Neumann algebras of localized observables $\{\mathcal{A}(\mathcal{O})\}$ on the spacetime, is that there exists a family $\mathcal{W}$ of subsets of the spacetime such that $J_{W_1}\mathcal{R}(W_2)J_{W_1} \in \{\mathcal{R}(W) \mid W \in \mathcal{W}\}$, for every $W_1, W_2 \in \mathcal{W}$. This condition makes no explicit appeal to isometries or other special attributes and is thus applicable, in principle, to quantum field theories on general curved spacetimes.

It has been shown for certain spacetimes, including Minkowski space, that under certain additional technical assumptions, the modular involutions encode enough information to determine the dynamics of the theory, the isometry group of the spacetime, and a continuous unitary representation of the isometry group which acts covariantly upon the observables and leaves the state invariant. In certain

cases including Minkowski space, it is even possible to derive the spacetime itself from the group $\mathcal{J}$ generated by the modular involutions $\{J_W \mid W \in \mathcal{W}\}$.

The modular unitaries $\Delta_W^{it}$ enter in this approach through a condition which is designed to assure the stability of the theory, namely that $\Delta_W^{it} \in \mathcal{J}$, for all $t \in \mathbb{R}$ and $W \in \mathcal{W}$. In Minkowski space, this additional condition entails that the derived representation of the Poincaré group satisfies the spectrum condition.

### Further Applications

As previously observed, through the close connection to the KMS condition, modular theory enters naturally into the equilibrium thermodynamics of many-body systems. But in recent work on the theory of nonequilibrium thermodynamics it also plays a role in making mathematical sense of the notion of quantum systems in local thermodynamic equilibrium. Modular theory has also proved to be of utility in recent developments in the theory of superselection rules and their attendant sectors, charges and charge-carrying fields.

*See also:* Algebraic Approach to Quantum Field Theory; Axiomatic Quantum Field Theory; Quantum Central-Limit Theorems; Symmetries in Quantum Field Theory of Lower Spacetime Dimensions; Thermal Quantum Field Theory; Positive Maps on $C^*$-Algebras; Two-Dimensional Models; von Neumann Algebras: Introduction, Modular Theory, and Classification Theory.

## Further Reading

Baumgärtel H and Wollenberg M (1992) *Causal Nets of Operator Algebras*. Berlin: Akademie-Verlag.

Borchers HJ (2000) On revolutionizing quantum field theory with Tomita's modular theory. *Journal of Mathematical Physics* 41: 3604–3673.

Bratteli O and Robinson DW (1981) *Operator Algebras and Quantum Statistical Mechanics II*. Berlin: Springer.

Connes A (1974) Caractérisation des algèbres de von Neumann comme espaces vectoriels ordonnés. *Annales de l'Institut Fourier* 24: 121–155.

Haag R (1992) *Local Quantum Physics*. Berlin: Springer.

Kadison RV and Ringrose JR (1986) *Fundamentals of the Theory of Operator Algebras*, vol. II. Orlando: Academic Press.

Pedersen GK (1979) *$C^*$-Algebras and Their Automorphism Groups*. New York: Academic Press.

Stratila S (1981) *Modular Theory in Operator Algebras*. Tunbridge Wells: Abacus Press.

Takesaki M (1970) *Tomita's Theory of Modular Hilbert Algebras and Its Applications*, Lecture Notes in Mathematics, vol. 128. Berlin: Springer.

Takesaki M (2003) *Theory of Operator Algebras II*. Berlin: Springer.

# Topological Defects and Their Homotopy Classification

**T W B Kibble**, Imperial College, London, UK



## Introduction

Symmetry-breaking phase transitions occur in a wide variety of systems – from condensed matter to the early universe. One of the common features of such transitions is the appearance, in the broken-symmetry phase, of topological defects, trapped regions in which the symmetry is restored, or at least changed. Examples are vortices in superfluids, domain walls in ferromagnets, and disclination lines in liquid crystals. Often these defects are stable for topological reasons, and play an important role in the dynamics of the system. An astonishingly rich variety of defects can be found in various systems. They can usefully be classified using the tools of homotopy theory.

## Spontaneous Symmetry Breaking

Let us consider a quantum-mechanical system with a symmetry group $G$. This means that each $g \in G$ is represented on the Hilbert space of quantum states by a unitary operator $\hat{U}(g)$, which commutes with the Hamiltonian. Spontaneous symmetry breaking occurs if this symmetry is not shared by the ground state or vacuum state $|0\rangle$ of the system. In other words, for some $g \in G, \hat{U}(g)|0\rangle \neq |0\rangle$. Then the ground state is necessarily degenerate: $\hat{U}(g)|0\rangle$ must have the same energy as $|0\rangle$.

Spontaneous symmetry breaking is usually describable in terms of an order-parameter field, which vanishes above the transition and is nonzero below it. We can find a scalar field $\hat{\phi}(\boldsymbol{r})$, or multiplet of fields $\hat{\phi} = (\hat{\phi}_i, i = 1, \ldots, n)$ transforming according to some representation $D$ of $G$ (assumed not to contain the trivial representation), whose expectation value in the ground state is nonzero:

$$\langle 0|\hat{\phi}(\boldsymbol{r})|0\rangle = \phi_0 \neq 0 \qquad [1]$$

This is the order parameter. Since

$$\langle 0|\hat{U}^\dagger(g)\hat{\phi}(\boldsymbol{r})\hat{U}(g)|0\rangle = D(g)\phi_0 \qquad [2]$$

it follows that the only elements of $G$ that can be symmetries of the ground state are those in the

stability subgroup $H$ of $\phi_0$ (the group of unbroken symmetries in this ground state):

$$H = \{g \in G:\ D(g)\phi_0 = \phi_0\} \qquad [3]$$

In terms of this subgroup, we can find a useful characterization of the manifold $\mathcal{M}$ of degenerate ground states. As noted above, for each $g \in G$, $\hat{U}(g)|0\rangle$ is also a ground state. However, these are not all distinct, because clearly $\hat{U}(gh)|0\rangle = \hat{U}(g)|0\rangle$ for all $h \in H$. Hence, the distinct ground states are in one-to-one correspondence with the left cosets $gH$ of $H$ in $G$, and $\mathcal{M}$ may be identified with the quotient space $G/H$, the space of left cosets.

For example, suppose $G$ is the rotation group SO(3), and $\hat{\boldsymbol{\phi}}$ belongs to the three-dimensional vector representation. If $\boldsymbol{\phi} \neq \mathbf{0}$ in the ground state, we may choose $\boldsymbol{\phi}_0 = (0, 0, v)$. Then, clearly, $H = \mathrm{SO}(2)$, the group of rotations about the $z$-axis, and $\mathcal{M} = \mathrm{SO}(3)/\mathrm{SO}(2) = S^2$, the 2-sphere. It is useful to think of $\mathcal{M}$ as the subset of the order-parameter space comprising the possible expectation values $\boldsymbol{\phi} = \langle\hat{\phi}\rangle$ for the various degenerate ground states. For example, in this case, $\mathcal{M} = \{\boldsymbol{\phi}\text{: } \boldsymbol{\phi}^2 = v^2\}$.

## Defect Formation

It is often possible to characterize the dynamics at finite temperature in terms of a function of the order parameter, the effective potential $V(\phi)$, which is necessarily invariant under $G$, and whose minima define the equilibrium states. At low temperatures, it has a form like $V = \lambda(\boldsymbol{\phi}^2 - v^2)^2$, whose minima occur at nonzero values of $\boldsymbol{\phi}$. But above the critical temperature $T_c$, the only minimum is at $\boldsymbol{\phi} = \mathbf{0}$, so the equilibrium state is symmetric under $G$. In the high-temperature phase, there may be large fluctuations in $\hat{\boldsymbol{\phi}}$, but its mean value will be zero.

Now, when the system is cooled through the phase tränsition, $\hat{\phi}$ will acquire a nonzero expectation value, gradually approaching one of the degenerate ground states characterized by a point of $\mathcal{M}$. But the choice of which one is unpredictable; the symmetry breaking is spontaneous. Moreover, in a large system, there is no reason why the same choice should be made everywhere. For example, a ferromagnet cooling through its Curie point may acquire a spontaneous magnetization in different directions in different parts of the sample.

Of course, there is an energetic penalty to having a spatially varying order parameter, so it will tend to become more uniform as the temperature is lowered. But the question arises whether there may be any topological obstruction to this process. It can happen that if we choose points on $\mathcal{M}$ in a continuous manner everywhere around the periphery of some region, it is topologically impossible to complete the process throughout its interior. Continuity may require that there are points where $\phi$ leaves the surface $\mathcal{M}$. For example, if our ferromagnet has two opposite possible directions of easy magnetization, described by $\boldsymbol{\phi}_0$ and $-\boldsymbol{\phi}_0$, then $\mathcal{M}$ consists essentially of these two points. Regions where $\boldsymbol{\phi} \approx \boldsymbol{\phi}_0$ and where $\boldsymbol{\phi} \approx -\boldsymbol{\phi}_0$ must be separated by domain walls across which $\boldsymbol{\phi}$ varies smoothly from one to the other.

## Homotopy Groups

To classify the various possible types of defect, we need to consider the homotopy groups of the manifold $\mathcal{M}$ of degenerate ground states. In this section, we briefly review the necessary definitions.

A path in $\mathcal{M}$ is a map $\phi: I \rightarrow \mathcal{M}$ from the unit interval $I = [0, 1] \subset \mathbb{R}$. We choose a base point $m_0 \in \mathcal{M}$ (which may be identified with $\phi_0$), and consider loops in $\mathcal{M}$, paths such that $\phi(0) = \phi(1) = m_0$. We say that two loops are homotopic, and write $\phi \sim \psi$, if one can be continuously deformed into the other within $\mathcal{M}$, that is, if there exists a map $\chi: I^2 \rightarrow \mathcal{M}$ such that

$$\chi(0, t) = \phi(t) \quad \text{and} \quad \chi(1, t) = \psi(t) \qquad [4]$$

for all $t$, and

$$\chi(s, 0) = \chi(s, 1) = m_0 \qquad [5]$$

for all $s$. This is an equivalence relation. The set $\pi_1(\mathcal{M})$ is the set of equivalence classes $[\phi]$ of loops under this relation.

On the set of loops, we may define a product $\phi\psi$, comprising the loop $\phi$ followed by $\psi$ (see Figure 1). Explicitly,

$$(\phi\psi)(t) = \begin{cases} \phi(2t), & 0 \leq t \leq \frac{1}{2} \\ \psi(2t-1), & \frac{1}{2} < t \leq 1 \end{cases} \qquad [6]$$

It is easy to show that if $\phi \sim \phi'$ and $\psi \sim \psi'$, then $\phi\psi \sim \phi'\psi'$. Hence, this defines a product on $\pi_1(\mathcal{M})$, by $[\phi][\psi] = [\phi\psi]$. So equipped, $\pi_1(\mathcal{M})$ becomes the

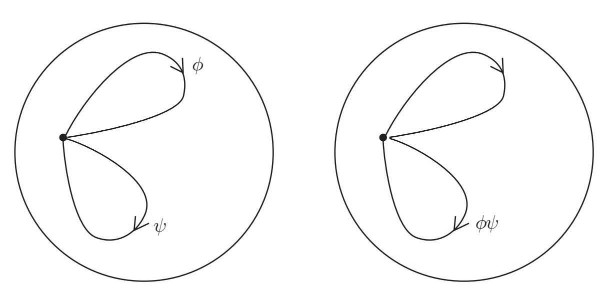



**Figure 1** The product of loops.

fundamental group or first homotopy group of $\mathcal{M}$. Note that the identity is the equivalence class $[\phi_0]$ of the trivial loop with $\phi_0(t) \equiv m_0$, while the inverse is $[\phi]^{-1} = [\tilde{\phi}]$, where the map $\tilde{\phi}$ is the reverse of $\phi$: $\tilde{\phi}(t) = \phi(1-t)$.

Strictly speaking, we should write $\pi_1(\mathcal{M}, m_0)$ in place of $\pi_1(\mathcal{M})$. However, for any path-connected space, the groups $\pi_1(\mathcal{M}, m_0)$ and $\pi_1(\mathcal{M}, m_0')$ are always isomorphic, and, more importantly, the same is true for any coset space $\mathcal{M} = G/H$, where $G$ is a Lie group and $H$ a closed subgroup. For a general manifold $\mathcal{M}, \pi_1(\mathcal{M})$ is not necessarily abelian, but it is so if $\mathcal{M}$ is a Lie group, or more generally a Riemannian symmetric space. The space $\mathcal{M}$ is said to be simply connected if $\pi_1(\mathcal{M}) = 0$, the group comprising only the identity element, $0 = \{[\phi_0]\}$. (Although $\pi_1(\mathcal{M})$ is not always abelian, it is conventional for homotopy groups to use an additive notation and represent the trivial group by 0 rather than 1.)

The $n$th homotopy group $\pi_n(\mathcal{M})$ may be defined similarly, as a set of equivalence classes of maps $\phi: I^n \to \mathcal{M}$ such that $\phi$ maps the entire boundary $\partial I^n$ to the base point $m_0$. Two such maps are homotopic ($\phi \sim \psi$) if there exists a map $\chi: I^{n+1} \to \mathcal{M}$ such that

$$\chi(0, t) = \phi(t) \quad \text{and} \quad \chi(1, t) = \psi(t) \qquad [7]$$

for all $t = (t_1, \ldots, t_n)$, and, for each $s \in I, \chi(s, t) = m_0$ for all $t \in \partial I^n$. The product $\phi\psi$ is defined by

$$(\phi\psi)(t_1, \ldots, t_n) = \begin{cases} \phi(2t_1, t_2, \ldots, t_n), & 0 \le t_1 \le \frac{1}{2} \\ \psi(2t_1 - 1, t_2, \ldots, t_n), & \frac{1}{2} < t_1 \le 1 \end{cases} \qquad [8]$$

The choice of $t_1$ rather than any other $t_j$ is arbitrary; all choices yield homotopic product maps. The product again defines a product on $\pi_n(\mathcal{M})$, which thereby becomes a group, the $n$th homotopy group. One new feature is that, for all $n > 1, \pi_n(\mathcal{M})$ is always abelian.

Note that since the entire boundary of $I_n$ is mapped to a single point, it is possible to collapse it, and talk instead about maps from the $n$-sphere $S^n$ to $\mathcal{M}$, taking one designated point to $m_0$. The fact that $\pi_n(\mathcal{M})$ is nontrivial indicates the existence in $\mathcal{M}$ of closed $n$-surfaces that cannot be smoothly shrunk to a point. In particular, it is worth noting that, for any $n$, $\pi_n(S^n) = \mathbb{Z}$, the additive group of integers, while $\pi_m(S^n) = 0$ for all $m < n$.

A special case is $n = 0$. Here, $S^0$ comprises two points only, and since one of them is always mapped to $m_0$, we really have to consider maps from a single point to $\mathcal{M}$, that is, points in $\mathcal{M}$. Two points are homotopic if they can be joined by a path in $\mathcal{M}$. Thus, $\pi_0(\mathcal{M})$ may be identified with the set of path-connected components of $\mathcal{M}$. Note, however, that in general no product can be defined on $\pi_0(\mathcal{M})$, so $\pi_0(\mathcal{M})$ should be called the zeroth homotopy set (not group). There is an important exception, however: if $G$ is a Lie group, and $G_0$ its connected subgroup (the subset of elements joined by paths to the identity $e$), then $\pi_0(\mathcal{M})$ may be identified with the quotient group $G/G_0$. Note, however, that this group $\pi_0(\mathcal{M}) = G/G_0$ is not necessarily abelian.

## Classification of Defects

We now turn to the classification of defects by means of homotopy groups. It will be useful to start with simple specific examples in three-dimensional space, $\mathbb{R}^3$.

First, suppose again that $\boldsymbol{\phi}$ belongs to the vector representation of $G = \mathrm{SO}(3)$. Then $\mathcal{M} = \mathrm{SO}(3)/\mathrm{SO}(2) = S^2$ may be identified with the sphere $\mathcal{M} = \{\boldsymbol{\phi}: \boldsymbol{\phi}^2 = v^2\}$ in $\phi$ space. Consider a closed surface $\mathcal{S}$, an embedding of a 2-sphere $S^2$ in $\mathbb{R}^3$. Assume that everywhere on $\mathcal{S}$ the field $\boldsymbol{\phi}(\boldsymbol{r})$ has one of the ground-state values. In other words, we have a map $\boldsymbol{\phi}: \mathcal{S} \to \mathcal{M}$, from one 2-sphere to another. The map $\boldsymbol{\phi}$ can be extended to a map from the interior of $\mathcal{S}$ to $\mathcal{M}$ only if it belongs to the trivial homotopy class $[\boldsymbol{\phi}_0] \in \pi_2(\mathcal{M})$, where $\phi_0: I^2 \to \mathcal{M}: (t_1, t_2) \mapsto m_0 = eH$. In all other cases, there must be at least one point where $\boldsymbol{\phi}(\boldsymbol{r}) = \boldsymbol{0}$; this is a point defect. The second homotopy group in this case is $\pi_2(S^2) = \mathbb{Z}$, so the possible point defects, or monopoles, are labeled by an integer $n \in \mathbb{Z}$, the winding number. (An example of a map with winding number $n$ is (in spherical polars) $(r, \theta, \varphi) \mapsto (v, \theta, n\varphi)$.)

More generally, point defects in $\mathbb{R}^d$ are classified by $\pi_{d-1}(\mathcal{M})$. A map $\phi$ from a closed $(d-1)$-dimensional surface $\mathcal{S} \subset \mathbb{R}^d$ to $\mathcal{M}$ can be extended to the interior of $\mathcal{S}$ if and only if it belongs to the trivial homotopy class $[\phi_0] \in \pi_{d-1}(\mathcal{M})$. If this is not the case, there must be at least one point around which $\phi(\boldsymbol{r})$ leaves the surface $\mathcal{M}$, although in general it is not required to vanish anywhere.

Second, take the case where $\phi$ is a single complex field, and $G$ is the phase symmetry group U(1). In this case, $H$ is the subgroup $\mathbf{1} = \{1\} \subset G$. Thus, $\mathcal{M} = \mathrm{U}(1)/\mathbf{1} = S^1$; this manifold may be identified with the circle $\{\phi: |\phi| = v\}$ in the order-parameter space. Now consider a closed loop $\mathcal{C}$ in space, an embedding of $S^1$ in $\mathbb{R}^3$ (see Figure 2). Suppose that on $\mathcal{C}, \phi(\boldsymbol{r})$ takes one of the ground-state values, say $\phi(\boldsymbol{r}) = v \exp[i\alpha(\boldsymbol{r})]$. If $\mathcal{S}$ is some surface with boundary $\mathcal{C}$, then the map $\phi: \mathcal{C} \to \mathcal{M}$ can be extended to a map $\phi: \mathcal{S} \to \mathcal{M}$ if and only if it belongs to the trivial homotopy class $[\phi_0] \in \pi_1(\mathcal{M})$. If it does not, then there must be at least one point on $\mathcal{S}$ within $\mathcal{C}$ where $\phi = 0$. Moreover, this

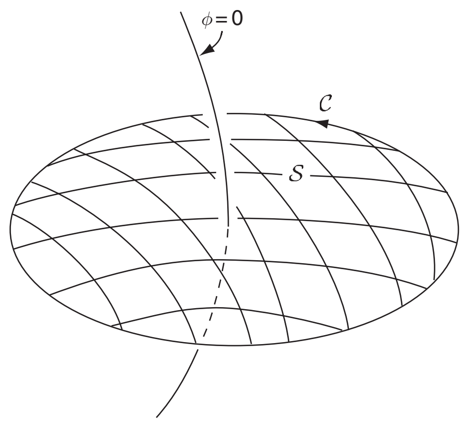



**Figure 2** A linear defect.

must be true of every surface $\mathcal{S}$ spanning $\mathcal{C}$, so there must be a curve passing through $\mathcal{C}$ along which $\phi = 0$. This is a linear defect, a string or vortex line. In this case, the first homotopy group is $\pi_1(S^1) = \mathbb{Z}$, so we see that the possible linear defects are classified by an integer, the winding number $n$. An example of a map with winding number $n$ is $\varphi \mapsto v e^{ni\varphi}$.

Again, this result can easily be generalized. Linear defects in $\mathbb{R}^d$ are classified by $\pi_{d-2}(\mathcal{M})$. If, on a $(d-2)$-dimensional surface $\mathcal{C}, \phi(\boldsymbol{r})$ takes values in $\mathcal{M}$, and if it does not belong to the trivial homotopy class, there must be a linear defect threading through $\mathcal{C}$, around which $\phi$ leaves the surface $\mathcal{M}$ – although again it need not necessarily vanish.

More generally yet, in the $d$-dimensional space $\mathbb{R}^d$, defects of dimension $p$ are classified by the homotopy group $\pi_{d-p-1}(\mathcal{M})$. For example, in three dimensions, planar defects – domain walls – are classified by $\pi_0(\mathcal{M})$.

## The Exact Sequence

There are mathematical theorems that greatly facilitate the computation of the homotopy group of homogeneous spaces, of the form $\mathcal{M} = G/H$.

We begin with the maps relating these spaces to each other. There is a canonical injective homomorphism $i: H \to G: h \mapsto h$, and a canonical projection associating each element of $G$ with its coset: $p: G \to \mathcal{M}: g \mapsto gH$. Moreover, it is clear that the image of $i$, namely the subgroup $H$, is also the kernel of $p$, the inverse image $p^{-1}m_0$ of the distinguished element $m_0 = eH$ of $\mathcal{M}$. These statements can be summarized by saying that

$$\mathbf{1} \to H \xrightarrow{i} G \xrightarrow{p} \mathcal{M} \to \mathbf{1}$$

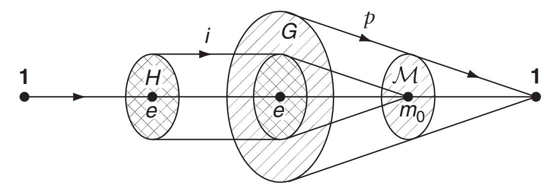



**Figure 3** An exact sequence.

is an exact sequence: the image of each map is the kernel of the following one (see Figure 3).

Next, we note that since any closed loops (or $n$-surfaces) in $H$ belonging to the same homotopy class are also homotopic as loops (or $n$-surfaces) in $G$, there is an induced homomorphism $i_*: \pi_n(H) \to \pi_n(G)$. Similarly, homotopic loops or $n$-surfaces in $G$ project to homotopic loops or $n$-surfaces in $\mathcal{M}$, so there is an induced homomorphism $p_*: \pi_n(G) \to \pi_n(\mathcal{M})$. Moreover, it is easy to see that although $i_*$ is not necessarily injective and $p_*$ not necessarily projective, it is true that the image of $i_*$ is the kernel of $p_*$. For example, any loop in $G$ will be mapped to a homotopically trivial loop in $\mathcal{M}$ if and only if it is homotopic to the image of a loop in $H$.

In addition, there is a boundary map that relates homotopy groups of different dimension: $\partial: \pi_{n+1}(\mathcal{M}) \to \pi_n(H)$. To see this, it is useful to think of $G$ as a fiber bundle with base space $\mathcal{M}$ and fiber $H$. Now consider a map $\phi: (I^{n+1}, \partial I^{n+1}) \to (\mathcal{M}, m_0)$. Since $p$ is a projection, $\phi$ can always be lifted to a map $\hat{\phi}: (I^{n+1}, \partial I^{n+1}) \to (G, H)$, that is, we can find a (nonunique) map $\hat{\phi}$ such that $\phi = p \circ \hat{\phi}$ (see Figure 4). However, $\hat{\phi}$ does not necessarily map the boundary to a single point; what is true is that $\hat{\phi}$ must map the boundary to a subset of $H$, and since topologically $\partial I^{n+1} \simeq S^n$, this defines a map $\tilde{\phi}: S^n \to H$. If we allow $\phi$ to vary over some homotopy class of maps, and $\hat{\phi}$ to vary continuously, then $\tilde{\phi}$ will

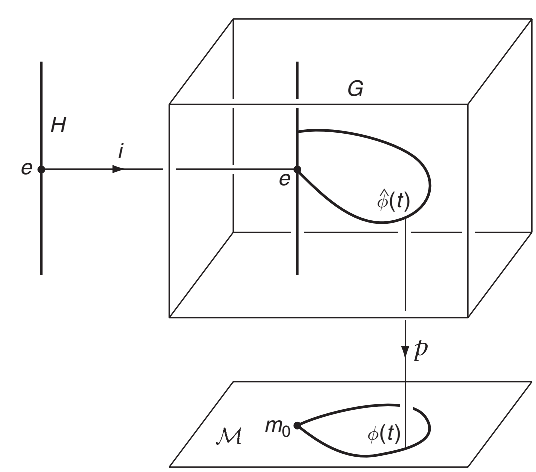



**Figure 4** Lift of a loop.

also remain in one homotopy class. Thus, we have defined a map $\partial : \pi_{n+1}(\mathcal{M}) \to \pi_n(H) : [\phi] \mapsto [\tilde{\phi}]$.

It is also easy to see that the image of $\partial : \pi_{n+1}(\mathcal{M}) \to \pi_n(H)$ is the kernel of $i_* : \pi_n(H) \to \pi_n(G)$, because the $n$-surface in $H$ defined by $\tilde{\phi}$ is necessarily homotopically trivial in $G$. Similarly, one can see that the image of $p_* : \pi_{n+1}(G) \to \pi_{n+1}(\mathcal{M})$ is the kernel of $\partial : \pi_{n+1}(\mathcal{M}) \to \pi_n(H)$.

Putting all these results together, we see that there is a (semi-infinite) exact sequence connecting all the homotopy groups:

$$\cdots \xrightarrow{p_*} \pi_{n+1}(\mathcal{M}) \xrightarrow{\partial} \pi_n(H) \xrightarrow{i_*} \pi_n(G) \xrightarrow{p_*} \pi_n(\mathcal{M})$$
$$\xrightarrow{\partial} \pi_{n-1}(H) \xrightarrow{i_*} \cdots \pi_1(G) \xrightarrow{p_*} \pi_1(\mathcal{M}) \xrightarrow{\partial} \pi_0(H) \quad [9]$$
$$\xrightarrow{i_*} \pi_0(G) \xrightarrow{p_*} \pi_0(\mathcal{M})$$

This sequence makes it easy to compute most of the low-dimensional homotopy groups of $\mathcal{M}$. Let us begin with $\pi_0(\mathcal{M})$, which merely labels its disconnected components. As noted earlier, for the Lie group $G, \pi_0(G)$ is the quotient group $\pi_0(G) = G/G_0$, where $G_0$ is the connected subgroup of $G$. Now the image of $\pi_0(H)$ under $i_*$ is clearly the set of connected components of $G$ that contain elements of $H$, so if $G$ has $m$ connected components, and $n$ of them contain elements of $H$, then $\pi_0(\mathcal{M})$ has $m/n$ elements (see Figure 5).

Next, we note that, for all the higher homotopy groups, disconnected pieces are irrelevant. Since a loop, for example, starting at $m_0$ must remain within its connected component $\mathcal{M}_0 \subset \mathcal{M}$, it follows that $\pi_1(\mathcal{M}) = \pi_1(\mathcal{M}_0)$, and similarly $\pi_n(\mathcal{M}) = \pi_n(\mathcal{M}_0)$ for all $n > 1$. So one can ignore any disconnected parts of the symmetry group $G$, and assume from now on that $\pi_0(G) = 0$. Moreover, it is always possible to replace $G$ by its simply connected covering group, replacing SO(3), for

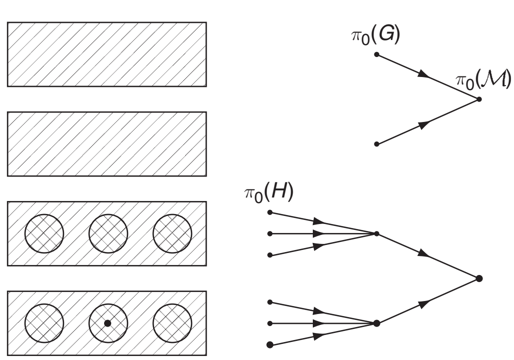



**Figure 5** The disconnected components of $G$ are shaded, those of $H$ are cross-hatched. Here $\pi_0(\mathcal{M})$ has two elements.

example, by SU(2). Thus, we may also assume that $\pi_1(G) = 0$. Then the section of the exact sequence in the second line of [9] becomes

$$0 \xrightarrow{p_*} \pi_1(\mathcal{M}) \xrightarrow{\partial} \pi_0(H) \xrightarrow{i_*} 0$$

which implies that the two groups in the center are isomorphic:

$$\pi_1(\mathcal{M}) = \pi_0(H) \quad [10]$$

For example, if the symmetry group $G = \mathrm{SO}(3)$ is completely broken, so that $H = \mathbf{1}$, then replacing $G$ by $\tilde{G} = \mathrm{SU}(2)$ requires replacing $H$ by $\tilde{H} = \{+1, -1\} \simeq \mathbb{Z}_2$, hence also $\pi_1(\mathcal{M}) = \pi_0(\tilde{H}) = \mathbb{Z}_2$; there is only one nontrivial class of linear defects in this model.

To find $\pi_2(\mathcal{M})$, we need a standard theorem about Lie groups, namely that the second homotopy group of any Lie group is trivial: for any $G, \pi_2(G) = 0$. (No details of the proof are given here. It derives from the fact that a generic element $g \in G$ belongs to a unique one-parameter subgroup $\{\exp(tX), t \in \mathbb{R}\} \subset G$, where $X$ is an element of the Lie algebra of $G$. Thus, all the points on a surface in $G$ may be joined by these paths to the identity, and the surface may then be shrunk along the resulting cone. There are exceptional elements for which this is not true, but it can be shown that in a $d$-dimensional group they lie on $(d-3)$-dimensional surfaces, so any 2-surface can be smoothly deformed to avoid them.)

It follows from this theorem that another section of the exact sequence is

$$0 \xrightarrow{p_*} \pi_2(\mathcal{M}) \xrightarrow{\partial} \pi_1(H) \xrightarrow{i_*} 0$$

which again implies an isomorphism:

$$\pi_2(\mathcal{M}) = \pi_1(H) \quad [11]$$

For example, if $G = \mathrm{SO}(3)$ and $H = \mathrm{SO}(2)$, or equivalently $\tilde{G} = \mathrm{SU}(2)$ and $\tilde{H} = \mathrm{U}(1)$ (a double cover of the SO(2)), then $\pi_2(\mathcal{M}) = \pi_1(\tilde{H}) = \mathbb{Z}$, so point defects in this theory are labeled by an integer winding number.

## Examples

The simplest continuous symmetry is the U(1) phase symmetry $\hat{\phi} \mapsto \hat{\phi}e^{i\alpha}$ of a complex field. In a weakly interacting Bose gas, below the Bose–Einstein condensation temperature, or in superfluid helium-4, a macroscopic fraction of the atoms occupies a single quantum state, and $\hat{\phi}$ acquires a nonzero expectation value, $\langle\hat{\phi}\rangle = \phi$, whose phase is arbitrary, so the symmetry is completely broken to $H = \mathbf{1}$. Thus, $\mathcal{M} = S^1$; we have a circle of equivalent degenerate ground states. (This corresponds to

spontaneous breaking of the particle-number symmetry. It is possible to describe the system in a U(1)-invariant way, by projecting out a state of definite particle number, a uniform superposition of all the states in $\mathcal{M}$, but it is generally less convenient to do so.) In this case, the only nontrivial homotopy group is $\pi_1(\mathcal{M}) = \mathbb{Z}$, so the only defects are linear defects classified by a winding number $n \in \mathbb{Z}$. The defects with $n = \pm 1$ are stable vortices. Those with $|n| > 1$ are in general unstable and tend to break up into $|n|$ single-quantum vortices.

Low-temperature superconductors also have a U(1) symmetry, although there are important differences. This is not a global symmetry but a local, gauge symmetry, with coupling to the electromagnetic field. Moreover, it is not single atoms that condense but Cooper pairs, pairs of electrons of equal and opposite momentum and spin. These systems too exhibit linear defects, magnetic flux tubes carrying a magnetic flux $4\pi n\hbar/e$.

A less trivial example is a nematic liquid crystal. These materials are composed of rod-shaped molecules that tend, at low temperatures, to line up parallel to one another. The nematic state is characterized by a preferred orientation, described by a unit vector $\boldsymbol{n}$, the director. (Note that $\boldsymbol{n}$ and $-\boldsymbol{n}$ are physically equivalent.) There is long-range orientational order, with molecules preferentially lining up parallel to $\boldsymbol{n}$, but unlike a solid crystal there is no long-range translational order – the molecules move freely past each other as in a normal liquid.

A convenient order parameter here is the mean mass quadrupole tensor $\Phi$ of a molecule. In the nematic state, $\Phi$ is proportional to $(3\boldsymbol{nn} - \mathbf{1})$; for example, if $\boldsymbol{n} = (0, 0, 1)$, then $\Phi$ is diagonal with diagonal elements proportional to $(-1, -1, 2)$. In this case, the symmetry group is SO(3) (or, more precisely, O(3); but the inversion symmetry is not broken, so we can restrict our attention to the connected part of the group). The subgroup $H$ that leaves this $\Phi$ invariant is a semidirect product, $H = \mathrm{SO}(2) \ltimes \mathbb{Z}_2$ (isomorphic to O(2)), composed of rotations about the $z$-axis and rotations through $\pi$ about axes in the $x$–$y$ plane. (If we enlarge $G$ to its simply connected covering group $\tilde{G} = \mathrm{SU}(2)$, then $H$ becomes $\tilde{H} = [\mathrm{U}(1) \ltimes \mathbb{Z}_4]/\mathbb{Z}_2$, where U(1) is generated as before by $J_z$. The essential difference is that the square of any of the elements in the disconnected piece of $\tilde{H}$ is not now the identity but the element $\mathrm{e}^{2\pi \mathrm{i} J_z} = -1 \in \mathrm{U}(1)$.) The manifold $\mathcal{M}$ of degenerate ground states in this case is the projective space $\mathbb{R}P^2$ (obtained by identifying opposite points of $S^2$).

Since $\tilde{H}$ has disconnected pieces, we have $\pi_1(\mathcal{M}) = \pi_0(\tilde{H}) = \mathbb{Z}_2$. Thus, there can be topologically stable linear defects, here called disclination lines, around which the director $\boldsymbol{n}$ rotates by $\pi$ (see Figure 6). The fact that these defects are classified by $\mathbb{Z}_2$ rather than $\mathbb{Z}$ means that a line around which $\boldsymbol{n}$ rotates by $2\pi$ is topologically trivial; indeed, $\boldsymbol{n}$ can be smoothly rotated near the line to run parallel to it, leaving a configuration with no defect.

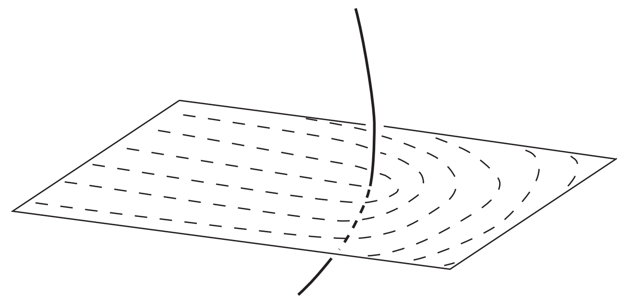


**Figure 6** Orientation of molecules around a disclination line.

There are also point defects; since $\pi_2(\mathcal{M}) = \pi_1(\tilde{H}) = \mathbb{Z}$, they are labeled by an integer winding number $n$. In a defect with $n = 1$, the vector $\boldsymbol{n}$ points radially outwards all round the defect position.

## Helium-3

Finally, let us turn to helium-3, one of the most fascinating and complex examples of spontaneous symmetry breaking, which becomes a superfluid at a temperature of a few millikelvin. Unlike helium-4, this is, of course, a Fermi liquid, so it is not the atoms that condense, but bound pairs of atoms, analogous to Cooper pairs. In this case, however, the most attractive channel is not the $^1$S, but the $^3$P, so the pairs have both orbital and spin angular momentum, $L = S = 1$. Therefore, the order parameter is not a single complex scalar field but a $3 \times 3$ complex matrix $\Phi_{jk}$, where the two indices label the orbital and spin angular momentum states.

To a good approximation, the system is invariant under separate rotations of $\boldsymbol{L}$ and $\boldsymbol{S}$ (the effects of the small spin–orbit coupling will be discussed later), so the symmetry group is

$$G = \mathrm{U}(1)_Y \times \mathrm{SO}(3)_L \times \mathrm{SO}(3)_S \qquad [12]$$

where the subscripts denote the generators and $\mathrm{U}(1)_Y$ represents multiplication by an overall phase factor, $\mathrm{e}^{\mathrm{i}\alpha Y} : \Phi_{jk} \mapsto \Phi_{jk}\mathrm{e}^{\mathrm{i}\alpha}$. This complicated symmetry allows much scope for a large variety of defects. There are, in fact, two distinct superfluid phases, $A$ and $B$, with different symmetries (and indeed in the presence of a magnetic field there is a third, $A1$).

In the $^3$He-$A$ phase, the order parameter has the form $\Phi_{jk} \propto (m_j + \mathrm{i}n_j)d_k$, where $\boldsymbol{m}$, $\boldsymbol{n}$, $\boldsymbol{d}$ are unit vectors, with $\boldsymbol{m} \perp \boldsymbol{n}$; if we set $\boldsymbol{l} = \boldsymbol{m} \wedge \boldsymbol{n}$, then

$\boldsymbol{l}$ defines the orbital angular momentum state by $\boldsymbol{l} \cdot \boldsymbol{L} = 1$, while $\boldsymbol{d}$ defines the spin quantization axis, such that $\boldsymbol{d} \cdot \boldsymbol{S} = 0$. The manifold $\mathcal{M}_A$ for this phase is

$$\mathcal{M}_A = [\mathrm{SO}(3) \times S^2]/\mathbb{Z}_2 \qquad [13]$$

where the $\mathbb{Z}_2$ is present because ($\boldsymbol{m}$, $\boldsymbol{n}$, $\boldsymbol{d}$) and ($-\boldsymbol{m}$, $-\boldsymbol{n}$, $-\boldsymbol{d}$) represent the same state. If, for example, we take $\boldsymbol{l}$ and $\boldsymbol{d}$ in the $z$-direction, the unbroken symmetry subgroup is

$$H_A = \mathrm{SO}(2)_{L_z+Y} \times [\mathrm{SO}(2)_{S_z} \ltimes \mathbb{Z}_2] \qquad [14]$$

where the nontrivial element of $\mathbb{Z}_2$ may be taken to be $e^{i\pi(S_x+L_z)}$. The covering group of $G$ is, of course,

$$\tilde{G} = \mathbb{R}_Y \times \mathrm{SU}(2)_L \times \mathrm{SU}(2)_S \qquad [15]$$

Correspondingly,

$$\tilde{H}_A = \mathbb{R}_{L_z+Y} \times [\mathrm{U}(1)_{S_z} \ltimes \mathbb{Z}_4] \qquad [16]$$

It follows that the homotopy groups are

$$\pi_0(\mathcal{M}_A) = 0, \quad \pi_1(\mathcal{M}_A) = \mathbb{Z}_4, \quad \pi_2(\mathcal{M}_A) = \mathbb{Z} \qquad [17]$$

There are linear defects labeled by a mod-4 quantum number and point defects labeled by an integer.

For the $^3$He-$B$ phase, by contrast, the order parameter is of the form

$$\Phi_{jk} \propto R_{jk} e^{i\theta} \qquad [18]$$

where $R$ is a rotation matrix, $R \in \mathrm{SO}(3)$. Here then,

$$\mathcal{M}_B = \mathrm{SO}(3) \times S^1 \qquad [19]$$

with homotopy groups

$$\pi_0(\mathcal{M}_B) = 0, \quad \pi_1(\mathcal{M}_B) = \mathbb{Z}_2 \times \mathbb{Z}, \quad \pi_2(\mathcal{M}_B) = 0 \qquad [20]$$

In this phase, there are two distinct types of linear defect, the mass vortices with an integer label, and the spin vortices with a mod-2 label. (One can also have a "spin–mass vortex" carrying both quantum numbers.)

## Composite Defects

There are several cases, including in particular helium-3, that exhibit symmetry breaking with multiple length or energy scales. For example, there may be two order parameters, say $\phi, \psi$, with $|\phi| \gg |\psi|$. If $|\psi|$ is negligible, the symmetry $G$ is broken by $\phi$ to $H$, and the manifold of degenerate ground states is $\mathcal{M} = G/H$. However, these states are not all exactly degenerate: $\psi$ breaks the symmetry further to $K \subset H$, so the precisely degenerate ground states form a submanifold $\mathcal{M}' = G/K$.

The case of helium-3 is slightly different. Here it is the small spin–orbit coupling, arising from long-range dipole–dipole interactions, that introduces the second scale. Its effect is only significant over large distances.

In the $^3$He-$A$ phase, at short range the $\boldsymbol{l}$ and $\boldsymbol{d}$ vectors are uncorrelated but, over large distances, they tend to be aligned parallel or antiparallel. We can use the $\mathbb{Z}_2$ symmetry mentioned earlier to choose $\boldsymbol{l} = \boldsymbol{d}$. Hence, the manifold $\mathcal{M}'_A$ of true ground states is only a submanifold of $\mathcal{M}_A$, namely $\mathcal{M}'_A = \mathrm{SO}(3)$, whose homotopy groups are

$$\pi_0(\mathcal{M}'_A) = 0, \quad \pi_1(\mathcal{M}'_A) = \mathbb{Z}_2, \quad \pi_2(\mathcal{M}'_A) = 0 \qquad [21]$$

Because of different behavior on different scales, "composite" defects can arise. For example, because $\pi_2(\mathcal{M}_A) = \mathbb{Z}$, there are short-range monopole configurations. For the $n = 1$ monopole, we have a configuration with uniform $\boldsymbol{l}$, and with $\boldsymbol{d}$ pointing outwards from the center. But, eventually the misalignment of $\boldsymbol{d}$ with $\boldsymbol{l}$ is energetically disfavored, and at large distances $\boldsymbol{d}$ tends to rotate to align with $\boldsymbol{l}$ except around one particular direction where it is oppositely aligned (see Figure 7). We have a composite defect: a small monopole coupled to a relatively fat string.

To see how the small- and large-scale structures fit together, one has to look also at the relative homotopy groups $\pi_n(\mathcal{M}, \mathcal{M}')$, whose elements are homotopy classes of maps from $I^n$ to $\mathcal{M}$ such that one face of the boundary is mapped into $\mathcal{M}'$, and the remainder to the chosen base point $m_0$. For example, $\pi_1(\mathcal{M}, \mathcal{M}')$ classifies paths that terminate at $m_0$ while beginning at any point of $\mathcal{M}'$. There is, in fact, a long exact sequence, similar to [9], relating these homotopy groups, of which a typical segment is

$$\cdots \xrightarrow{\partial} \pi_n(\mathcal{M}') \xrightarrow{i_*} \pi_n(\mathcal{M}) \xrightarrow{p_*} \pi_n(\mathcal{M}, \mathcal{M}') \xrightarrow{\partial} \pi_{n-1}(\mathcal{M}') \xrightarrow{i_*} \cdots \qquad [22]$$

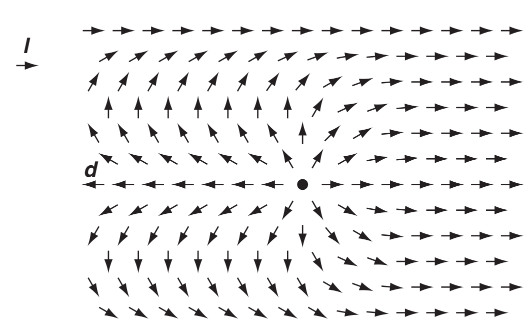



**Figure 7** Cross-section of a short-range monopole attached to a fat string.

The relevant groups in the present case are

$$\pi_1(\mathcal{M}_A, \mathcal{M}'_A) = \mathbb{Z}_2, \quad \pi_2(\mathcal{M}_A, \mathcal{M}'_A) = \mathbb{Z} \qquad [23]$$

Because $\pi_1(\mathcal{M}_A) = \mathbb{Z}_4$, there are three distinct classes of linear defects at small scales, but only those with quantum number $n = 2 \,(\mathrm{mod}\, 4)$ survive unchanged to large scales; they correspond to the nontrivial element of $\pi_1(\mathcal{M}'_A) = \mathbb{Z}_2$. On the other hand, the homotopy classes $n = \pm 1 \,(\mathrm{mod}\, 4)$ are mapped to nontrivial elements of $\pi_1(\mathcal{M}_A, \mathcal{M}'_A) = \mathbb{Z}_2$, which indicates that the corresponding linear defects are coupled at long range to fat domain walls, across which $\boldsymbol{d}$ rotates through $\pi$ with a compensating rotation through $\pi$ about $\boldsymbol{l}$. Similarly, the nontrivial elements of $\pi_2(\mathcal{M}_A) = \mathbb{Z}$ are mapped to nontrivial elements of $\pi_2(\mathcal{M}_A, \mathcal{M}'_A)$, confirming that these short-range monopoles are coupled to fat strings, as in Figure 7.

For $^3$He-$B$, the effect of the spin–orbit coupling is to make the most energetically favorable configurations those in which the rotation matrix $R$ in [18] represents a rotation about an arbitrary axis $\boldsymbol{n}$ through the Leggett angle $\theta_L = \arccos(-1/4) = 104°$: $R = \exp(-\mathrm{i}\theta_L \boldsymbol{n} \cdot \boldsymbol{J})$.

Consequently,

$$\mathcal{M}'_B = S^2 \times S^1 \qquad [24]$$

and so

$$\pi_0(\mathcal{M}'_B) = 0, \quad \pi_1(\mathcal{M}'_B) = \mathbb{Z}, \quad \pi_2(\mathcal{M}'_B) = \mathbb{Z} \qquad [25]$$

The relative homotopy groups are

$$\pi_1(\mathcal{M}_B, \mathcal{M}'_B) = \mathbb{Z}_2, \quad \pi_2(\mathcal{M}_B, \mathcal{M}'_B) = 0 \qquad [26]$$

Here the mass vortex persists at long range, but the configuration around the spin vortex deforms so that they become attached to fat domain walls. The "monopole" configurations corresponding to nontrivial elements of $\pi_2(\mathcal{M}'_B)$ have no short-range singularity at all.

*See also:* Abelian Higgs Vortices; Leray–Schauder Theory and Mapping Degree; Liquid Crystals; Phase Transition Dynamics; Quantum Field Theory: A Brief Introduction; Quantum Fields with Topological Defects; Solitons and Other Extended Field Configurations; String Topology: Homotopy and Geometric Perspectives; Symmetries and Conservation Laws; Symmetry Breaking in Field Theory; Variational Techniques for Ginzburg–Landau Energies.

## Further Reading


Helgason S (2001) *Differential Geometry, Lie Groups and Symmetric Spaces*. Providence, RI: American Mathematical Society.

Hu S-T (1959) *Homotopy Theory*. New York: Academic Press.

Kibble TWB (1976) Topology of cosmic domains and strings. *Journal of Physics A: Mathematical and General* 9: 1387–1398.

Kibble TWB (2000) Classification of topological defects and their relevance to cosmology and elsewhere. In: Bunkov YM and Godfrin H (eds.) *Topological Defects and the Non-Equilibrium-Dynamics of Symmetry Breaking Phase Transitions*, NATO Science Series C: Mathematical and Physical Sciences. vol. 249, pp. 7–31. Dordrecht: Kluwer Academic Publishers.

Shellard EPS and Vilenkin A (1994) *Cosmic Strings and Other Topological Defects*. Cambridge: Cambridge University Press.

Tilley DR and Tilley J (1990) *Superfluidity and Superconductivity*, 3rd edn. Bristol: IoP Publishing.

Toulouse G and Kléman M (1976) Principles of a classification of defects in ordered media. *Journal de Physique Lettres* 37: 149.

Vollhardt D and Wölfle P (1990) *The Superfluid Phases of Helium 3*. London: Taylor and Francis.

Volovik GE (1992) *Exotic Properties of Superfluid $^3$He*. Singapore: World Scientific.

Volovik GE and Mineev VP (1977) Investigation of singularities in superfluid $^3$He and in liquid crystals by the homotopic topology methods. *Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki* 72: 2256–2274 *(Soviet Physics–JETP 24:* 1186–1196).


# Topological Gravity, Two-Dimensional

**T Eguchi**, University of Tokyo, Tokyo, Japan



## Introduction

It is well known that large-$N$ Hermitian matrix models generate Feynman diagrams which represent the triangulation of Riemann surfaces. For instance, if we consider the integral of an $N \times N$ Hermitian matrix $H$

$$Z = \int \mathrm{d}H \exp\left(-N\left[\frac{1}{2}\mathrm{tr}\, H^2 + \frac{\lambda}{4}\mathrm{tr}\, H^4\right]\right) \qquad [1]$$

we find that the free energy $F = \log Z$ has the $1/N$ expansion

$$F = \sum_{g=0}^{\infty} N^{2-2g} F_g(\lambda) \qquad [2]$$

Inspection of the Feynman diagrams shows that $F_g$ reproduces the sum over the triangulations of genus $g$ Riemann surfaces. The theory [1] is obviously well defined for $\lambda \geq 0$. In the large-$N$ expansion, the theory continues to exist also at negative values of $\lambda$ down to the critical point $\lambda_c = -1/12$.

The double scaling limit of large-$N$ matrix models (Brézin and Kazakov 1990, Douglas and

Shenker 1990, Gross and Migdal 1990) is given by adjusting the coupling $\lambda$ to $\lambda_c$ and at the same time taking the limit $N \to \infty$. In this limit, contributions of all genera survive, and the theory describes the dynamics of fluctuating surfaces of arbitrary topologies. Results obtained in this way do not, in fact, depend on the detailed choice of the potential ($\phi^4$ type in [1]) and have a high degree of universality. Thus, it provides an interesting model of two-dimensional (2D) quantum gravity.

Soon after the discovery of double scaling limit of matrix models, Witten observed that the correlation functions of the 2D gravity theory may be given a geometrical interpretation as topological invariants of the moduli space of Riemann surfaces $\mathcal{M}$, and that the 2D gravity theory may be reformulated as a topological field theory (Witten 1990). This reformulation of the results of the 2D gravity theory is called "2D topological gravity."

In fact, 2D gravity theories come in a family parametrized by a pair of integers $(p, q)$. The double scaling limit of [1] gives the simplest example $(p=2, q=1)$. Models with a chain of $p-1$ Hermitian matrices give the $(p, q)$ 2D gravity theories. The label $q$ stands for the order of criticality of the model, and higher values of $q$ are achieved by fine-tuning the parameters of the potential. At $q=1$, 2D gravity theories possess a topological interpretation. The most basic case $(p=2, q=1)$ is called pure topological gravity, and in theories at higher values of $p$, topological gravity is coupled to a matter system, that is, topological minimal models. Topological minimal models are obtained by twisting $\mathcal{N}=2$ superconformal field theories.

Let us first consider the case of pure gravity $(p=2, 1)$. Let $\mathcal{O}_n$ denote the observables in the theory and $t_n$ the coupling constants to these operators. The correlation functions of topological gravity are given by

$$\langle \mathcal{O}_{n_1}\mathcal{O}_{n_2}\ldots\mathcal{O}_{n_s}\rangle_g, \quad n_i = 1, 2, \ldots \qquad [3]$$

where $\langle\cdots\rangle_g$ denotes the expectation value on a surface with $g$ handles. The precise significance of eqn [3] as the intersection number on the moduli space is discussed below. The string partition function $\tau(t)$ is defined as the generating function of all possible correlation functions

$$\tau(t) = \exp\sum_{g=0}^{\infty}\left\langle \exp\sum t_n\mathcal{O}_n\right\rangle_g \qquad [4]$$

The most striking aspect of topological gravity is the connection of the intersection theory on $\mathcal{M}$ to the theory of completely integrable systems, that is, Korteveg–de Vries (KdV) and KP hierarchies. Witten conjectured that the generating function of intersection numbers on moduli space $\tau(t)$ is the $\tau$-function of KdV hierarchy. KdV hierarchy is obtained by generalizing the well-known KdV equation

$$\frac{\partial u}{\partial t} = \frac{3}{2}u\frac{\partial u}{\partial x} + \frac{1}{4}\frac{\partial^3 u}{\partial x^3} \qquad [5]$$

Identification of the KdV equation with topological gravity is given by $u=2\langle\mathcal{O}_1\mathcal{O}_1\rangle, x=t_1, t=t_3$. Witten's conjecture was verified by Kontsevich (1991) by an explicit construction of a new type of matrix model which generates the triangulation of the moduli space of Riemann surfaces.

In the general case of $(p, 1)$ topological gravity, the partition function of the theory obeys the equations of $p$th generalized KdV hierarchy ($p$ reduction of KP hierarchy).

## Intersection Theory

We now present some basic features of intersection theory on the moduli space of Riemann surfaces. It is known that 2D oriented surfaces $\Sigma$ with $g$ handles and $s$ marked points $x_i$ $(i=1,\ldots,s)$ possess a finite number of inequivalent complex structures (complex structures are identified when they differ only by diffeomorphism). The space of inequivalent complex structures is called the moduli space $\mathcal{M}_{g,s}$ of the Riemann surface $\Sigma$. Its dimension is given by

$$\dim\mathcal{M}_{g,s} = 3g-3+s \qquad [6]$$

For a mathematically rigorous treatment, we have to consider a compactification $\bar{\mathcal{M}}_{g,s}$ of moduli space $\mathcal{M}_{g,s}$ by adding suitable boundary components which arise due to various types of degenerations of Riemann surfaces. In the Deligne–Mumford or stable compactification, one considers the following three classes of singular Riemann surfaces $\Sigma$:

1. Two points, $x_i$ and $x_j$, on $\Sigma$ come close together. In this case, an extra 2-sphere is pinched off from the surface by forming a thin neck. The sphere contains points $x_i$ and $x_j$ and also the point $x_l$ at the end of the neck (see Figure 1a). Since the original surface now has one point less and the 2-sphere with three points has no moduli, the degenerate surface has $3g-4+s$ parameters and forms a boundary divisor of the moduli space $\bar{\mathcal{M}}_{g,s}$.
2. If a cycle of nontrivial homology class shrinks to a point, we have a surface with one less genus and two extra marked points. Singular surface has $3(g-1)-3+s+2$ number of moduli and this is again a complex codimension-1 component (see Figure 1b).

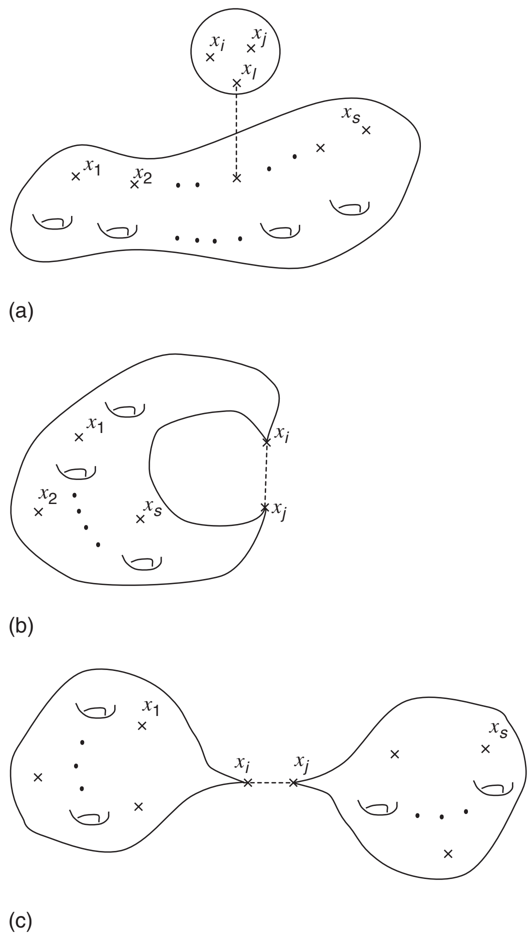



**Figure 1** Degenerate Riemann surface obtained when (a) the points $x_i$ and $x_j$ coincide; (b) a nontrivial cycle collapses, two new points $x_i$ and $x_j$ are created; (c) a pinching cycle collapses, two new points $x_i$ and $x_j$ are created.

3. Similarly, if a dividing cycle pinches, one obtains two disconnected surfaces of genus $g_i$ with $s_i+1$ marked points $(i=1,2; g_1+g_2=g, s_1+s_2=s)$. This type of degeneration also has the same number of parameters $\sum(3g_i-3)+\sum(s_1+1)=3g-4+s$ (see Figure 1c).

It is known that $\bar{\mathcal{M}}_{g,s}$ is a compact and smooth orbifold space, and observables of topological gravity are given by the cohomology classes on $\bar{\mathcal{M}}_{g,s}$. There exist special cohomology classes introduced by Mumford and Morita, which are defined as follows: There are natural line bundles $\mathcal{L}_1,\ldots,\mathcal{L}_s$ on the moduli space $\bar{\mathcal{M}}_{g,s}$. The fiber of the bundle $\mathcal{L}_i$ at a point $\Sigma\in\bar{\mathcal{M}}_{g,s}$ is the cotangent space $T^*_{x_i}\Sigma$ to the point $x_i$ on the surface $\Sigma$. These line bundles have the first Chern classes $c_1(\mathcal{L}_i)$ and by taking their exterior power we can define $2n$-dimensional classes

$$\sigma_n(i)=c_1(\mathcal{L}_i)^n\in H^{2n}(\bar{\mathcal{M}}_{g,s}) \qquad [7]$$

Correlation functions are defined by integrating these classes over the moduli space:

$$\langle\sigma_{n_1}\cdots\sigma_{n_s}\rangle_g\equiv\int_{\bar{\mathcal{M}}_{g,s}}c_1(\mathcal{L}_1)^{n_1}\wedge\cdots\wedge c_1(\mathcal{L}_s)^{n_s} \qquad [8]$$

These integrals are topological invariants of $\bar{\mathcal{M}}_{g,s}$ and are nonzero only when the degree of the cohomology classes adds up to the dimension of the moduli space

$$\sum_{i=1}^{s}n_i=3g-3+s \qquad [9]$$

$\sigma_n(i)$ is known as the $n$th descendant of the puncture operator $\sigma_0(i)$, since it is associated with the marked point $x_i$.

The above correlation functions are evaluated using various recursion relations. First, one has the puncture equation

$$\langle\sigma_0\sigma_{n_1}\cdots\sigma_{n_s}\rangle_g=\sum_{i=1,n_i\neq 0}^{s}\langle\sigma_{n_1}\cdots\sigma_{n_i-1}\cdots\sigma_{n_s}\rangle_g \qquad [10]$$

which can be derived by considering a map $\pi:\bar{\mathcal{M}}_{g,s+1}\to\bar{\mathcal{M}}_{g,s}$ where one forgets the position of an extra point. Contributions arise when the forgotten point coincides with the other points. This relation can be used to eliminate $\sigma_0$'s from correlation functions when they are well defined. At $g=0$, less than three insertions are ill-defined and one has

$$\langle\sigma_0\sigma_0\sigma_0\rangle_0=1 \qquad [11]$$

Another basic relation is the dilaton equation for the operator $\sigma_1$:

$$\langle\sigma_1\sigma_{n_1}\cdots\sigma_{n_s}\rangle_g=(2g-2+s)\langle\sigma_{n_1}\cdots\sigma_{n_s}\rangle_g \qquad [12]$$

The dilaton equation follows from the fact that since $\sigma_1$ is the first Chern class $c_1(\mathcal{L})$, it calculates the degree of the canonical line bundle of genus $g$ surface with $s$ punctures. At $g=1$, one insertion is required and one has

$$\langle\sigma_1\rangle_1=\tfrac{1}{24} \qquad [13]$$

By combining these recursion relations, one can evaluate the correlation functions. For instance, at $g=0$ one finds

$$\langle\sigma_{n_1}\cdots\sigma_{n_s}\rangle_0=\frac{(n_1+\cdots+n_s)!}{n_1!\cdots n_s!} \qquad [14]$$

A powerful way of computing correlation functions is given by the KdV hierarchies and Virasoro conditions as discussed below. In the context of integrable systems, it is convenient to redefine the observables as

$$\mathcal{O}_{2n+1}=(2n+1)!!\cdot\sigma_n,\quad n\geq 0 \qquad [15]$$

## Topological Minimal Models

Standard intersection theory applies to the case of pure topological gravity, $p=2$. At higher values of $p$, the theory is generalized as follows: one introduces the coupling of topological gravity to the topological matter sector which is obtained by twisting the $\mathcal{N}=2$ superconformal theories.

We recall that $\mathcal{N}=2$ superconformal symmetry is generated by the operators, stress tensor $T(z)$, U(1) current $J(z)$, and two types of supersymmetry generators $G(z)^{\pm}$. (In the holomorphic sector of the theory these operators depend on the holomorphic coordinate $z$ of the Riemann surface. In the antiholomorphic sector they depend on the antiholomorphic variable $\bar{z}$.) Mode expansion of the stress tensor and U(1) current is given by

$$T(z) = \sum_n L_n z^{-n-2}, \qquad J(z) = \sum_n J_n z^{-n-1} \qquad [16]$$

$L_n$ generates the Virasoro algebra

$$[L_m, L_n] = (m-n)L_{m+n} + \frac{c}{12} m(m^2-1)\delta_{m+n,0} \quad [17]$$

where $c$ denotes the central charge of the theory. Commutators of $J_n$ and $L_n$ are given by

$$[J_m, J_n] = \frac{c}{3} m\delta_{m+n,0}, \qquad [L_m, J_n] = -nJ_{m+n} \qquad [18]$$

It is known that there is a continuum of unitary $\mathcal{N}=2$ conformal theories in the range $c \geq 3$; however, only discrete values of the central charge $c = 3k/(k+2), k = 1, 2, \ldots$ are allowed in the region $3 \geq c \geq 1$. These are the $\mathcal{N}=2$ minimal models labeled by the level $k$. Only a finite number of primary fields exist in these theories.

In $\mathcal{N}=2$ theory, primary fields $\phi_\alpha$ are characterized by their conformal dimension and U(1) charge:

$$L_0|\phi_\alpha\rangle = h|\phi_\alpha\rangle, \qquad J_0|\phi_\alpha\rangle = q|\phi_\alpha\rangle \qquad [19]$$

There exists a special set of primary operators, chiral primary fields $\phi_\ell$ ($\ell = 0, \ldots, k$), which are annihilated by the supercharge operator $G^+$:

$$\oint dz G^+(z)|\phi_\ell\rangle = 0 \qquad [20]$$

$\phi_\ell$ has the dimension and U(1) charge

$$q(\phi_\ell) = \frac{\ell}{k+2}, \quad h(\phi_\ell) = \tfrac{1}{2}q(\phi_\ell), \quad \ell = 0, 1, \ldots, k \quad [21]$$

By considering primary fields annihilated by $G^-$, we can also define antichiral fields. Antichiral fields have U(1) charge opposite to those of chiral fields.

If one defines the twisted stress tensor by

$$T'(z) = T(z) + \tfrac{1}{2}\partial J(z) \qquad [22]$$

then $T'(z)$ has a vanishing central charge. Furthermore, the conformal dimensions of the supersymmetry operators $G^{\pm}$ become shifted from 3/2 to $h(G^+) = 1$ and $h(G^-) = 2$. It is then possible to integrate $G^+$ on the Riemann surface and define a fermionic scalar operator $G_0^+ = \oint dz G^+(z)$. From the $\mathcal{N}=2$ algebra, one has

$$(G_0^+)^2 = 0, \qquad \{G_0^+, G^-(z)\} = 2\,T'(z) \qquad [23]$$

If we identify $G_0^+$ as the Becchi–Rouet–Stora–Tyupin (BRST) operator of the theory, then the twisted stress tensor becomes BRST trivial, which is the characteristic feature of topological field theory. Thus, we obtain a topological field theory by twisting $\mathcal{N}=2$ conformal theory (Eguchi and Yang 1990). These are topological minimal models. BRST-invariant observables are given by the chiral primary fields [20]. (To be precise, when we take account of the antiholomorphic sector, we may define either $Q = G_0^+ + \bar{G}_0^+$ or $Q = G_0^+ + \bar{G}_0^-$ as the BRST operator. Thus, in general, we obtain two different topological field theories. This is the origin of the mirror symmetry. In the context of topological gravity, one takes the convention $Q = G_0^+ + \bar{G}_0^+$.)

Now, we consider the coupling of topological gravity to topological minimal models. We identify $k = p-2$. Making use of chiral fields $\phi_\ell$ ($\ell = 0, \ldots, p-2$), observables are constructed:

$$\sigma_{n,\ell} = \sigma_n \otimes \phi_\ell \qquad [24]$$

$\mathcal{N}=2$U(1) charge is identified as the degree of differential form of the moduli space. Thus, the degree of $\sigma_{n,\ell}$ is $n + \ell/p$. Correlation functions $\langle \prod_{i=1}^s \sigma_{n_i,\ell_i} \rangle_g$ are nonzero if the selection rule

$$\sum_{i=1}^s \left(n_i + \frac{\ell_i}{p}\right) = \left(3 - \frac{p-2}{p}\right)(g-1) + s \qquad [25]$$

is obeyed.

We may assemble $\sigma_{n,\ell}$ into operators with one index $\mathcal{O}_m$ as

$$\mathcal{O}_{np+\ell+1} = \prod_{r=0}^{n}(rp + \ell + 1) \cdot \sigma_{n,\ell} \qquad [26]$$

where one introduces a convenient normalization factor. Note that the operators $\mathcal{O}_m$ do not exist when $m \equiv 0 \bmod p$ and the corresponding paramters $t_m$ are absent. This is a characteristic feature of $p$ reduced KP hierarchy.

The puncture and dilaton equations for $(p, 1)$ theories read

$$\langle\sigma_{0,0}\sigma_{n_1,k_1}\cdots\sigma_{n_s,k_s}\rangle_g = \sum_{i=1,n_i\neq 0}^{s}\langle\sigma_{n_1,k_1}\cdots\sigma_{n_i-1,k_i}\cdots\sigma_{n_s,k_s}\rangle_g \quad [27]$$

$$\langle\sigma_{1,0}\sigma_{n_1,k_1}\cdots\sigma_{n_s,k_s}\rangle_g = (2g-2+s)\langle\sigma_{n_1,k_1}\cdots\sigma_{n_s,k_s}\rangle_g \quad [28]$$

The special terms at $g=0$ and $g=1$ are given by

$$\langle\sigma_{0,0}\sigma_{0,i}\sigma_{0,p-i-2}\rangle_0 = 1, \qquad \langle\sigma_{1,0}\rangle_1 = \frac{p-1}{24} \quad [29]$$

## Integrable Hierarchy

We now summarize some basic facts about the integrable hierarchy (see for instance eqn [5]). We introduce a $p$th order differential operator:

$$L = D^p + \sum_{i=0}^{p-2} u_i(x)D^i, \quad D \equiv \frac{\partial}{\partial x} \quad [30]$$

where the coefficient functions $u_i$ are arbitrary functions of $x$. This Lax operator describes the $p$th generalized KdV hierarchy. We consider the time evolution of the operator $L$ by an infinite set of commuting Hamiltonians:

$$\frac{\partial L}{\partial t_n} = [H_n, L], \quad n = 1, 2, \ldots \quad [31]$$

where $H_n$ is given by

$$H_n = \left(L^{n/p}\right)_+ \quad [32]$$

Here "+" denotes the non-negative part of a pseudodifferential operator and is defined as

$$A = \sum_{i=-\infty}^{n} f_i(x)D^i, \qquad A_+ = \sum_{i=0}^{n} f_i(x)D^i \quad [33]$$

We also use the notation

$$\operatorname{res} A = f_{-1}(x), \qquad A_- = \sum_{i=-\infty}^{-1} f_i(x)D^i \quad [34]$$

Note that $x$ is identified as the first time variable $t_1$, that is, $x = t_1$.

It is a basic result of the calculus of pseudodifferential operators that the above Hamiltonians satisfy the zero-curvature condition

$$\frac{\partial H_m}{\partial t_n} - \frac{\partial H_n}{\partial t_m} + [H_m, H_n] = 0 \quad [35]$$

Note that when $m$ is a multiple of $p$, $H_m$ becomes a power of $L$ and trivially commutes with $L$. Thus, the time variables $t_m$ are absent for $n \equiv 0 \bmod p$. In the simple case of $p = 2$, one has

$$L = D^2 + u(x) \quad [36]$$

and $H_3 = D^3 + (3/2)uD + (3/4)u'$. One finds

$$\frac{\partial L}{\partial t_3} = \frac{\partial u}{\partial t_3} = [H_3, L] = \frac{3}{2}u\frac{\partial u}{\partial x} + \frac{1}{4}\frac{\partial^3 u}{\partial x^3} \quad [37]$$

which is the standard KdV equation.

In the case of KP hierarchy, one starts with a pseudodifferential operator

$$Q = D + \sum_{i=1}^{\infty} a_i D^{-i} \quad [38]$$

and considers the time evolution equations

$$\frac{\partial Q}{\partial t_n} = [H_n, Q], \quad H_n = (Q^n)_+ \quad [39]$$

$p$-reduced KP hierarchy is obtained if one has

$$Q^p{}_- = 0 \quad [40]$$

By introducing a pseudodifferential operator $K$, one may bring $Q$ to the simple derivative operator $D$ as

$$Q = KDK^{-1} \quad [41]$$

$K$ has an expansion of the form

$$K = 1 + \sum_{i=1}^{\infty} a_i D^{-i} \quad [42]$$

After time evolution, the coefficient functions $u_i(x)$ of the Lax operator depend also on the variables $t_2, t_3, \ldots$ and become functions of $t \equiv \{t_1, t_2, \ldots\}$. These functions are expressed by the $\tau$-function $\tau(t)$ of the hierarchy in the following manner:

$$\operatorname{res} K = -\frac{\partial}{\partial x}\log\tau(t) \quad [43]$$

$$\operatorname{res} L^{i/p} = \frac{\partial^2}{\partial x \partial t_i}\log\tau(t) \quad [44]$$

These residues are expressed in terms of $\{u_i\}$ and their derivatives in $x$, and one can determine them in terms of the $\tau$-function.

In the case $p=2$, one has

$$[H_k, L] = 2D\,\mathrm{res}(L^{k/2}) = DR_k, \quad k = \text{odd} \qquad [45]$$

Here $\{R_k\}$ are the Gelfand–Dikii potentials

$$\begin{aligned} R_1 &= u, \qquad R_3 = \tfrac{1}{4}(3u^2 + u'') \\ R_5 &= \tfrac{1}{16}(10u^3 + 5u'^2 + 10uu'' + u'''') \\ &\vdots \end{aligned} \qquad [46]$$

and obey the recursion relation

$$DR_{k+2} = \tfrac{1}{4}\left(D^3 + 2(Du + uD)\right)R_k \qquad [47]$$

If one uses the relation [44], Gelfand–Dikii potentials are identified as

$$R_k = 2\langle \mathcal{O}_1 \mathcal{O}_k \rangle \qquad [48]$$

By setting $k=1$, we note $u=2\langle \mathcal{O}_1\mathcal{O}_1\rangle$ and find that the evolution equations [31] are all satisfied as

$$\begin{aligned} \frac{\partial L}{\partial t_k} &= \frac{\partial u}{\partial t_k} = 2\frac{\partial}{\partial t_k}\langle \mathcal{O}_1\mathcal{O}_1\rangle = 2D\langle \mathcal{O}_1\mathcal{O}_k\rangle \\ &= DR_k = [H_k, L] \end{aligned} \qquad [49]$$

Now it is possible to identify the initial condition for the Lax operator in the case of topological $(p, 1)$ gravity. By using the definition

$$\log \tau(t) = \sum_{g=0}^{\infty} \left\langle \exp \sum_n t_n \mathcal{O}_n \right\rangle_g \qquad [50]$$

one has

$$\mathrm{res}\, L^{i/p}(0) = \langle \mathcal{O}_1 \mathcal{O}_i \rangle, \quad i = 1, \ldots, p-1 \qquad [51]$$

From [29] one finds

$$\mathrm{res} L^{i/p}(0) = \mathrm{i}x \cdot \delta_{i,p-1} \qquad [52]$$

This gives the initial value of the Lax operator:

$$L(0) = D^p + px \qquad [53]$$

Thus, only the lowest term $u_0(x)=px$ is nonzero and higher coefficients all vanish at $t=0$. This is the special simplification which takes place in the topological gravity theory.

We note a relation

$$\left[\frac{1}{p}D, L(0)\right] = 1 \qquad [54]$$

This is the so-called "string equation" (at $t=0$). At nonzero values of $t$, the string equation takes the form

$$\begin{aligned} &[P, L] = 1 \\ &P = \frac{1}{p}\left( (L^{1/p})_+ - \sum_{k=p+1} kt_k (L^{(k-p)/p})_+ \right) \end{aligned} \qquad [55]$$

From [55], we see that $(p, 1)$ theory corresponds to the background value of the coupling $t_{p+1} = -1/(p+1)$. In the case of $(p, q)$ theory, background value is given by $t_{pq+1} = -1/(pq+1)$.

## Virasoro Conditions

A powerful algebraic machinery controlling the structure of 2D gravity is the so-called "Virasoro conditions." One introduces differential operators

$$L_{-1} = -\frac{\partial}{\partial t_1} + \sum_{k=p+1}^{\infty} kt_k \frac{\partial}{\partial t_{k-p}} + \frac{1}{2}\sum_{i+j=p} ijt_it_j \qquad [56]$$

$$L_0 = -\frac{\partial}{\partial t_{p+1}} + \sum_{k=1}^{\infty} kt_k \frac{\partial}{\partial t_k} + \frac{p^2-1}{24} \qquad [57]$$

By using the fact that derivative in $t_n$ brings down the operator $\mathcal{O}_n$ when acting on the $\tau$-function, it is easy to show that

$$L_{-1} \cdot \tau = 0 \qquad [58]$$

$$L_0 \cdot \tau = 0 \qquad [59]$$

reproduce the puncture [27] and dilaton equation [28], respectively. It is possible to show that the $L_{-1}$-condition, $L_{-1} \cdot \tau = 0$, is equivalent to the string equation [55].

Together with the operators $(n \geq 1)$

$$L_n = -\frac{\partial}{\partial t_{1+(n+1)p}} + \sum_{k=1}^{\infty} kt_k \frac{\partial}{\partial t_{k+np}} + \frac{1}{2}\sum_{i+j=np} \frac{\partial^2}{\partial t_i \partial t_j}$$

they generate Virasoro algebra $(L'_n \equiv (1/p)L_n)$

$$[L'_m, L'_n] = (m-n)L'_{m+n}, \quad n, m \geq -1 \qquad [60]$$

It is possible to show that the $(p, 1)$ model obeys the Virasoro conditions [6]

$$L_n \cdot \tau = 0, \quad n \geq -1 \qquad [61]$$

It is known that $(p, 1)$ models with $p > 2$ also obey constraints of $W$-algebra.

The relationship of the Virasoro conditions to KdV hierarchy is summarized as

string equation + KdV hierarchy
$\Longleftrightarrow$ Virasoro and $W$-algebra constraints

## Topological $\sigma$-Model

It is known that when the target space of a supersymmetric nonlinear $\sigma$-model is a Kahler manifold $K$, the theory acquires an enhanced $\mathcal{N}=2$

supersymmetry. Then we can twist the theory and converted into a topological field theory. This is the topological $\sigma$-model [7]. The partition function of the theory consists of a sum over world-sheet instantons, that is, holomorphic maps from the Riemann surface to the target space $K$. Due to supersymmetry, functional determinants around instantons cancel and the theory simply counts the number of holomorphic curves inside the Kahler manifold $K$. Thus, the topological $\sigma$-model has a close relationship with enumerative problems in algebraic geometry, that is, Gromov–Witten invariants and quantum cohomology theory.

When the topological $\sigma$-model is coupled to topological gravity, the BRST-invariant observables are given by $\sigma_n(\Phi_i) \equiv \sigma_n \otimes \Phi_i$, where $\Phi_i$ are cohomology classes of $K$. Correlation functions are defined as

$$\left\langle \prod_{i=1}^{s} \sigma_{n_i}(\Phi_i) \right\rangle_{g,d} = \int_{\bar{\mathcal{M}}(K;d)_{g,s}} \prod_{i=1}^{s} c_i(\mathcal{L}_i)^{n_i} \wedge e_i^*(\Phi_i) \quad [62]$$

Here $\bar{\mathcal{M}}_{g,s}(K;d)$ denotes the (stable compactification of) moduli space of degree $d$ holomorphic maps to $K$ from genus $g$ Riemann surfaces $\Sigma$. $e_i^*$ is the pullback of the evaluation map $e_i : (f; x_1, \ldots, x_s) \in \bar{\mathcal{M}}_{g,s}(K;d) \to f(x_i) \in K$ by $f$ where $f$ is a holomorphic map. Correlation functions [62] give topological (symplectic) invariants of the manifold $K$. In the cases $n_i = 0\,(i = 1, \ldots, s)$, they are known as Gromov–Witten invariants.

Equation [62] is nonvanishing if the selection rule

$$\begin{aligned} \sum_{i=1}^{s}(n_i + q_i) &= \dim \mathcal{M}_{g,s}(K;d) \\ &= c_1(K)d + (3 - \dim K)(g-1) + s \end{aligned} \quad [63]$$

is obeyed, where $q_i$ is the degree of cohomology class $\Phi_i$ and $c_1(K)$ is the first Chern class of the tangent bundle of $K$.

We see that there is a close parallel between the topological $\sigma$-model and ($p$, 1) topological gravity. If we formally set $q_i = \ell_i/p, c_1(K) = 0$, and $\dim K = (p-2)/p$, eqn [63] agrees with eqn [25]. Based on this analogy, Eguchi, Hori, and Xiong proposed the Virasoro conjecture [8], that is, generating functions of the number of holomorphic maps to arbitrary Kahler manifolds are annihilated by the Virasoro operators which are constructed by taking an analogy with those of $(p, 1)$ gravity. The Virasoro conjecture is a natural generalization of Witten's conjecture, and has recently been rigorously proved in the case of curves and projective spaces.

Excellent reviews on the theory of 2D topological gravity are given in Witten (1991) and Dijkgraaf (1991).

*See also:* Axiomatic Approach to Topological Quantum Field Theory; Large-*N* and Topological Strings; Mirror Symmetry: A Geometric Survey; Moduli Spaces: An Introduction; Riemann Surfaces; Topological Sigma Models; WDVV Equations and Frobenius Manifolds.

## Further Reading


Brézin E and Kazakov V (1990) Exactly soluble field theory of closed strings. *Physics Letters B* 236: 144.

Dijkgraaf R (1991) *Lectures Presented at Cargése Summer School on New Symmetry Principles in Quantum Field Theory*, hep-th/9201003.

Dijkgraaf R, Verlinde E, and Verlinde H (1991) Loop equations and Virasoro constraints in non-perturbative 2d quantum gravity. *Nulcear Physics B* 348: 435.

Douglas M and Shenker S (1990) Strings in less than one dimension. *Nuclear Physics B* 335: 635.

Eguchi T and Yang S-K (1990) $\mathcal{N} = 2$ superconformal models as topological field theories. *Modern Physics Letters A* 5: 1693.

Eguchi T, Hori K, and Xiong C-S (1997) Quantum cohomology and Virasoro algebra. *Physics Letters B* 402: 71.

Fukuma M, Kawai H, and Nakayama R (1991) Continuum Schwinger–Dyson equations and universal structures in two-dimensional quantum gravity. *International Journal of Modern Physics A* 6: 1385.

Gelfand IM and Dikii LA (1975) Asymptotic behavior of the resolvent of Sturm–Liouville equations and the algebra of the Korteweg–de Vries equations. *Russian Mathematical Surveys* 30(5): 77–113.

Gross D and Migdal A (1990) Non-perturbative two dimensional quantum gravity. *Physical Review Letters* 64: 17.

Kontsevich M (1991) Intersection theory on the moduli space of curves and the matrix Airy function. *Communications in Mathematical Physics*.

Witten E (1988) Topological sigma models. *Communications in Mathematical Physics* 118: 411.

Witten E (1990) On the topological phase of two dimensional gravity. *Nuclear Physics B* 340: 281.

Witten E (1991) Two-dimensional gravity and intersection theory of moduli space. *Surveys in Differential Geometry* 1: 243–310.

# Topological Knot Theory and Macroscopic Physics

**L Boi**, EHESS and LUTH, Paris, France



## Introduction to the Physical and Mathematical Contexts and Issues

One of the most exciting developments of mathematical physics in the last three decades has been the discovery of numerous intimate relationships between the topology and the geometry of knot theory and the dynamics of many domains of "classical" and "new" macroscopic physics. Indeed, complex systems of knotted and entangled filamentary structures are ubiquitous in nature and arise in such disparate contexts as electrodynamics, magnetohydrodynamics, fluid dynamics (vortex structures), superfluidity, dynamical systems, plasma physics, cosmic string theory, chaos of magnetic flows and nonlinear phenomena, turbulence, polymer physics, and molecular biology. In the recent years, mathematical tools have been developed to identify and analyze the geometrical and topological complex structures and behaviors of such systems and relate this information to energy levels and stable states.

The influence of geometry and topology on macroscopic physics has been especially fruitful in the study and comprehension of the following topics.

1. *Knots and braids in dynamical systems.* It is now clear that the chaotic behavior of the Hénon–Heiles system and other nonlinear systems is driven and controlled by topological properties. For example, it has been found that trajectories in the phase space form hyperbolic knots. The finding of knots in the Lorenz equations is another important theme closely related to the previous. By varying the Rayleigh number $r$, a parameter in the Lorenz equations, both chaotic and periodic behavior is observed. In the recent years, the knots (notably several torus knots) corresponding to the different periodic solutions of the system have been found and classified. By finding hyperbolic knots and in particular hyperbolic figure-8 knot as a solution to the Lorenz equations the suspicion that there exists a new route to chaos would be strengthened.

2. *Topological structures of electromagnetic fields.* Progress in the field of space physics, astronomy, and astrophysics over the last decade, increasingly reveals the significance of topological magnetic fields in these areas. In particular, the interaction of plasma and magnetic field can create an astonishing variety of structures, which often exhibit linked and knotted forms of magnetic flux. In these complex structures of the fields, huge amounts of magnetic energy can be stored. It is, however, a typical property of astrophysical plasmas, that the dynamics of magnetic fields is alternating between an ideal motion, where all forms of knottedness and linkage of the field are conserved (topology conservation), and a kind of disruption of the magnetic structure, the so-called magnetic reconnection. In the latter, the magnetic structure breaks up and reconnects, a process often accompanied by explosive eruptions, where enormous amounts of energy are set free. Magnetic reconnection is in close analogy to splitting of knots, which makes us confident that the global dynamics of magnetic and electromagnetic fields can be characterized with the help of such topological quantities as well.

3. *Knotting and unknotting of phase singularities.* It has long been known that dislocation lines can be closed, and recently it was shown that they can be knotted and linked. Moreover, Berry and Dennis (2001) constructed exact solutions of the Helmhotz equation representing torus knots and links; in fact, a straightforward application of this idea led to knotted and linked dislocation lines in stationary states of electrons in hydrogen. As a parameter, called $\alpha$, is varied, the topology of dislocation lines can change, leading to the creation of knots and links from initially simple dislocation loops, and the reverse process of unknotting and unlinking. The main purpose here is to elucidate the mechanism of these changes of topology. All waves are solutions of monochromatic wave equations, that is, stationary waves, and $\alpha$ is an external parameter that could be manipulated experimentally. However, $\alpha$ could represent time, and then the analogous solutions of time-dependent wave equations would describe knotting and linking events in the history of waves. The methods of Berry and Dennis are based on exact stationary solutions of wave equations, and lead to knots and links threaded by multistranded helices.

## The Origins of Topological Vortex Dynamics Ideas

The intimate relationship between three-dimensional vortex dynamics and topology was recognized as early as 1869 by W Thomson (Lord Kelvin) who tried to elaborate a theory of matter in which atoms were thought to be tiny vortex filaments embedded in an elastic-like fluid medium, called ether. Accordingly, the infinite variety of possible chemical compounds was given by the endless family of topological

combinations of linked and knotted vortices. Kelvin was inspired by the work of Gauss, who in an attempt to describe topologically the behavior of two inseparably closed linked circuits carrying electric current, found a relationship between the magnetic action induced by the currents and a pure number that depends only on the type of link, and not on the geometry: this number is the first topological invariant now known as the linking number.

In modern mathematical terms, Gauss introduced an invariant of a link consisting of two simple closed curves $\gamma_1, \gamma_2$ in $R^3$, namely the signed number of turns of one of the curves around the other, the linking coefficient $\{\gamma_1, \gamma_2\}$ of the link. His formula for this is

$$N = \{\gamma_1, \gamma_2\} = \frac{1}{4\pi}\int_{\gamma_1}\int_{\gamma_2} ([\mathrm{d}\gamma_1(t), \mathrm{d}\gamma_2(t)], \gamma_1 - \gamma_2)/ |\gamma_1(t) - \gamma_2(t)|^3 \qquad [1]$$

where [ , ] denotes the vector (or cross) product of vectors in $R^3$ and ( , ) the Euclidean scalar product. Thus, this integral always has an integer value $N$. If we take one of the curves to be the $z$-axis in $R^3$ and the other to lie in the $(x, y)$-plane, then the formula [1] gives the net number of turns of the plane curve around the $z$-axis. It is interesting to note that the linking coefficient [1] may be zero even though the curves are nontrivially linked. Thus, its having nonzero value represents only a sufficient condition for nontrivial linkage of the loops. This last consideration leads naturally to the mathematical concepts of knots and links whose most striking properties have been investigated in our introductory article (*see* Mathematical Knot Theory).

The other source of inspiration of Kelvin's theory of matter was the Helmholtz's laws of vortex motion, which state that in an ideal fluid (where there is no viscosity) vortex lives forever: two closed vortex rings, once linked, will always be linked. The classical results obtained by Helmholtz are basic to understanding the dynamics of Euler motions. The vorticity of a velocity field is its curl and is denoted $\omega_t(z) := \mathrm{curl}(X(z,t))$. In two dimensions, the vorticity is a real-valued function and $\omega_t = -\Delta\Psi$, where $\Psi$ is the stream function of $X(z, t)$. Recall that the push-forward of a scalar field (0-form) $s$ under a diffeomorphism $f$ is $f_* s = s \circ f^{-1}$. These results, in modern terms, can be stated as follows:

**Theorem** (Helmholtz–Kelvin). *An incompressible fluid motion* $(M_t, \phi_t)$ *with velocity field* $X$ *and vorticity* $\omega_t$ *is Euler if and only if its vorticity is passively transported,*

$$\phi_{t^*}\omega_0 = \omega_t$$

*and circulation around all smooth simple closed curves C are preserved under the flow,*

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\phi_t(C)} X \cdot \mathrm{d}\boldsymbol{r} = 0$$

One knows that in three dimensions, the Helmholtz–Kelvin theorem says that the vorticity (now a vector field) is transported. Thus, with generic initial vorticity a 3D time-periodic Euler fluid motion preserves a nontrivial vector field. One very interesting question that remains to be elucidated is the following: are there any chaotic, time-periodic Euler flows with stationary boundaries?

## The Connection between Topological and Numerical Invariants of Knots and the Physical Helicity of Vector Fields

The writhing number of a curve in Euclidean three-dimensional space is the standard measure of the extent to which the curve wraps and coils around itself; it has proved its importance for electrodynamics and fluid mechanics in the study of the knotted structures of magnetic vortices and dynamics flows, and for molecular biologists in the study of knotted duplex DNA and the enzymes which affect it. The helicity of a divergenceless vector field defined on a domain in Euclidean 3-space, introduced by Woltjer in 1958 in an astrophysical context and coined by Moffat in 1969 in the study of its topological meaning, is the standard measure of the extent to which the field lines wrap and coil around one another; it plays important roles in fluid mechanics, magnetohydrodynamics, and plasma physics. The "Biot–Savart operator" associates with each current distribution on a given domain the restriction of its magnetic field to the domain. When the domain is simply connected, the divergence-free fields which are tangent to the boundary and which minimize energy for given helicity provide models for stable force-free magnetic fields in space and laboratory plasmas; these fields appear mathematically as the extreme eigenfields for an appropriate modification of the Biot–Savart operator. Information about these fields can be converted into bounds on the writhing number of a given piece of DNA.

Recent researches (Cantarella *et al.* 2001) obtained rough upper bounds for the writhing number of a knot or link in terms of its length and thickness, and rough upper bounds for the helicity of a vector field in terms of its energy and the geometry of its domain. It was also showed that in the case of classical electrodynamics in vacuum, the

natural helicity invariant, called the electromagnetic helicity, has an important particle meaning: the difference between the numbers of right- and left-handed photons. Recently, a topological model of classical electrodynamics has been proposed in which the helicity is topologically quantized, in a relation that connects the wave and particle aspects of the fields (Trueba and Rañada 2000).

Consider two disjoint closed space curves, $C$ and $C'$, and the Gauss' integral formula for their linking number

$$\mathrm{Lk}(C, C') = \frac{1}{4\pi}\int_{C\times C'} (\mathrm{d}x/\mathrm{d}s \times \mathrm{d}y/\mathrm{d}t) \times x - y/|x-y|\,\mathrm{d}s\,\mathrm{d}t \qquad [2]$$

The curves $C$ and $C'$ are assumed to be smooth and to be parametrized by arclength. Now the question is to know what happens to this integral when the two space curves $C$ and $C'$ come together and coalesce as one curve $C$. At first glance, the integrand looks like it might blow up along the diagonal of $C \times C'$, but a careful calculation shows that in fact the integrand approaches zero on the diagonal, and so the integral converges. Its value is the writhing number $\mathrm{Wr}(C)$ of $C$ defined above:

$$\mathrm{Wr}(C) = \frac{1}{4\pi}\int_{C\times C} (\mathrm{d}x/\mathrm{d}s \times \mathrm{d}y/\mathrm{d}t) \times x - y/|x-y|\,\mathrm{d}s\,\mathrm{d}t \qquad [3]$$

Here is the very useful result, due to Fuller (1978). The writhing number of a knot $K$ is the average linking number of $K$ with its slight perturbations in every possible direction:

$$\mathrm{Wr}(K) = \frac{1}{4\pi}\int_{W\in S^2} \mathrm{Lk}(K, K+\varepsilon W)\,\mathrm{d(area)} \qquad [4]$$

This is helpful for getting a quick approximation to the writhing number of a knot which almost lies in a plane; in the example of a trefoil knot, $\mathrm{Wr}(K) \approx 3$.

Here, a very important result must be recalled, a "bridge theorem," proved by Berger and Field (1984), see also Ricca and Moffatt (1992), which connects helicity of vector fields to writhing of knots and links, and which can be used to convert upper bounds on helicity into upper bounds on writhing.

**Proposition** (Berger and Field). *Let $K$ be a smooth knot or link in 3-space and $\Omega = N(K, R)$ a tubular neighborhood of radius $R$ about $K$. Let $V$ be a vector field defined in $\Omega$, orthogonal to the cross-sectional disks, with length depending only on distance from $K$. This makes $V$ divergence-free and tangent to the boundary of $\Omega$. Then the writhing number $\mathrm{Wr}(K)$ of $K$ and the helicity $H(V)$ of the vector field $V$ are related by the formula*

$$H(V) = \mathrm{Flux}(V)^2\,\mathrm{Wr}(K)$$

In the formula, $\mathrm{Flux}(V)$ denotes the flux of $V$ through any of the cross-sectional disks $D$,

$$\mathrm{Flux}(V) = \int_D V\cdot n\,\mathrm{d(area)}$$

where $n$ is a unit normal vector field to $D$.

A key feature of this formula is that the helicity of $V$ depends on the writhing number of $K$, but not any further on its geometry; in particular, such quantities as the curvature and torsion of $K$ do not enter into the formula. Berger and Field actually showed that the helicity $H(V)$ is a sum of two terms: a "kink helicity," which is given by the right-hand side of the above formula, and a "twist helicity," which is easily shown in our case to be zero. Their proof assumes $K$ is a knot, but it is straightforward to extend it to cover links.

Let $\Omega$ be a compact domain in 3-space with smooth boundary $\partial\Omega$; we allow both $\Omega$ and $\partial\Omega$ to be disconnected. Let $V$ be a smooth vector field (where "smooth" means of class $C^\infty$), defined on the domain $\Omega$. The helicity $H(V)$ of the vector field $V$ is defined by the formula

$$H(V) = \frac{1}{4\pi}\int_{\Omega\times\Omega} V(x)\times V(y)\cdot x - y/|x-y|^3 \times \mathrm{d(vol)}_x\,\mathrm{d(vol)}_y \qquad [5]$$

Clearly, helicity for vector fields is the analog of writhing number for knots. Both formulas are variants of Gauss' integral formula for the linking number of two disjoint closed space curves.

In order to understand this formula for helicity, think of $V$ as a distribution of electric current, and use the Biot–Savart law of electrodynamics to compute its magnetic field:

$$\mathrm{BS}(V)(y) = \frac{1}{4\pi}\int_\Omega V(x)\times y - x/|y-x|^3\,\mathrm{d(vol)}_x \qquad [6]$$

Then the helicity of $V$ can be expressed as an integrated dot product of $V$ with its magnetic field $\mathrm{BS}(V)$:

$$\begin{aligned}H(V) &= \frac{1}{4\pi}\int_{\Omega\times\Omega} V(x)\times V(y)\cdot x - y/|y-x|^3 \times \mathrm{d(vol)}_x\,\mathrm{d(vol)}_y\\ &= \int_\Omega V(g)\cdot\left[\frac{1}{4\pi}\int_\Omega V(x)\times y - x/|x-y|^3\,\mathrm{d(vol)}_x\right] \times \mathrm{d(vol)}_y\\ &= \int_\Omega V(y)\cdot \mathrm{BS}(V)(y)\,\mathrm{d(vol)}_y\\ &= \int_\Omega V\cdot\mathrm{BS}(V)\,\mathrm{d(vol)}\end{aligned}$$

Cantarella *et al.* (2001) found two very interesting results.

**Theorem 1** *Let K be a smooth knot or link in 3-space, with length L and with an embedded tubular neighborhood of radius R. Then the writhing number* Wr(*K*) *of K is bounded by*

$$|\mathrm{Wr}(K)| < 1/4(L/R)^{4/3}$$

For the proof, see Cantarella *et al.* (2001).

**Theorem 2** *The helicity of a unit vector field V defined on the compact domain Ω is bounded by*

$$|H(V)| < 1/2\,\mathrm{vol}(\Omega)^{4/3}$$

Let us now give a brief overview of the methods used to find sharp upper bounds for the helicity of vector fields defined on a given domain Ω in 3-space. As usual, Ω will denote a compact domain with smooth boundary in 3-space. Let $K(\Omega)$ denotes the set of all smooth divergence-free vector fields defined on Ω and tangent to its boundary. These vector fields, sometimes called "fluid knots," are prominent for several reasons: (1) They are natural vector fields to study in a "fluid dynamics approach" to geometric knot theory. (2) They correspond to incompressible fluid flows inside a fixed container. (3) They are vector fields most often studied in plasma physics. (4) For given energy (equivalently minimize energy for given helicity), they provide models for stable force-free magnetic fields in gaseous nebulaes and laboratory plasmas. (5) The search for these helicity-maximizing fields can be converted to the task of solving a system of partial differential equations. (6) The fluid knots can reveal some fundamental and still unknown mechanisms, which characterize the phenomenon of phase transition, and in particular the transition from chaotic (unstable) phases and behaviors of matter to ordered (stable) ones.

## Knots and Fluid Mechanics (Vortex Lines, Magnetic Helicity, and Turbulence)

The Kelvin's theory of explaining atoms as knotted vortices in fluid ether was seminal in the development of topological fluid mechanics. The recent revival (starting in the 1970s) is mainly due to the work of Moffat, on topological interpretation of helicity, and Arnol'd, on asymptotic linking number of space-filling curves. Modern developments have been influenced by recent progress in the theory of knots and links.

### Influence of Geometry and Topology on Fluid Flows

Ideal topological fluid mechanics deals essentially with the study of fluid structures that are continuously deformed from one configuration to another by ambient isotopies. Since the fluid flow map $\varphi$ is both continuous and invertible, then $\varphi_{t_1}(K)$ and $\varphi_{t_2}(K)$ generate isotopies of a fluid structure $K$ (e.g., a vortex filament) for any $\{t_1, t_2\} \in I$. Isotopic flows generate equivalence classes of (linked and knotted) fluid structures. In the case of (vortex or magnetic) fluid flux tubes, fluid actions induce continuous deformations in $D$. One of the simplest deformations is local stretching of the tube. From a mathematical viewpoint, this deformation corresponds to a time-dependent, continuous reparametrization of the tube centerline. This reparametrization (via homotopy classes) generates ambient isotopies of the flux tube, with a continuous deformation of the integral curves.

Moreover, in the context of the Euler equations, the Reidemeister moves (or isotopic plane deformations), whose changes conserves the knot topology, are performed quite naturally by the action of local flows on flux tube strands. If the fluid in $(D - K)$ is irrotational, then these fluid flows (with velocity $\boldsymbol{u}$) must satisfy the Dirichlet problem for the Laplacian of the stream function $\varphi$, that is,

$$\begin{aligned} \boldsymbol{u} &= \nabla\varphi \quad \text{in } (D-K) \\ \nabla^2\varphi &= 0 \end{aligned} \qquad [7]$$

with normal component of the velocity to the tube boundary $\boldsymbol{u}_\perp$ given. Equations [7] admit a unique solution in terms of local flows, and these flows are interpretable in terms of Reidemeister's moves performed on the tube strands. Note that boundary conditions prescribe only $\boldsymbol{u}_\perp$, whereas no condition is imposed on the tangential component of the velocity. This is consistent with the fact that tangential effects do not alter the topology of the physical knot (or link). The three type of Reidemeister's moves are therefore performed by local fluid flows, which are solutions to [7], up to arbitrary tangential actions.

### Knotted and Linked Tubes of Magnetic Flux

Let $\boldsymbol{T}$ be the standard solid torus in $\boldsymbol{R}^3$ given by

$$((2 + \varepsilon\cos\theta)\cos\varphi,\ (2 + \varepsilon\cos\theta)\sin\varphi,\ \varepsilon\sin\theta)) \qquad [8]$$

where $0 \le \theta, \varphi < 2\pi$, and $0 \le \varepsilon < 1$. For relatively prime integers $p$ and $q$, let $F_{p,q}$ denote the foliation

of $\boldsymbol{T}$ by the curves $\gamma_{\varepsilon,\theta}$ (where $0 \leq \varepsilon \leq 1$ and $0 \leq \theta < 2\pi$) given by

$$\gamma_{\varepsilon,\theta}(s) = (2 + \varepsilon\cos(\theta + qs))\cos(ps), \\ (2 + \varepsilon\cos(\theta + qs))\sin(ps), \varepsilon\sin(\theta + qs) \qquad [9]$$

where $0 \leq s < 2\pi$.

**Definition** A magnetic tubular link (or magnetic link) is a smooth immersion into $\boldsymbol{R}^3$ of finitely many disjoint standard solid tori $\cup_{i=1}^{n}\boldsymbol{T}_i$

$$L : \cup_{i=1}^{n}\boldsymbol{T}_i \to \boldsymbol{R}^3$$

and a smooth magnetic field $\boldsymbol{B}$ on $\boldsymbol{R}^3$ such that

(i) $L$ is an imbedding when restricted to the interior of $\cup_{i=1}^{n}\boldsymbol{T}_i$,
(ii) the bounding surface of $\cup_i L(\boldsymbol{T}_i)$, that is, $\cup_i L(\partial\boldsymbol{T}_i)$ is a magnetic surface, and
(iii) for each component $L\boldsymbol{T}_i$, there exist relatively prime nonzero integers $p_i$ and $q_i$ such that $L$ maps the foliation $F_{p_i,q_i}$ of $\boldsymbol{T}_i$ onto the integral curves of $\boldsymbol{B}$ in $L\boldsymbol{T}_i$.

**Remark** Thus, for every fixed $i$ and $j$, the linking number between an arbitrary field line in $L\boldsymbol{T}_i$ and an arbitrary field line in $L\boldsymbol{T}_j$ is the same regardless of which integral curves are chosen from $L\boldsymbol{T}_i$ and $L\boldsymbol{T}_j$, respectively. This is true even when $i = j$.

It follows that a magnetic link $\cup_i L\boldsymbol{T}_i$ remains a magnetic link under the action of the fluid flow, that is, $\cup_i g_t L\boldsymbol{T}_i$ is a magnetic link for $t \geq 0$.

Keeping that the magnetic field $\boldsymbol{B}$ is frozen in the fluid, we can now find and study those properties of magnetic links that are invariant under the action of fluid flow. One obvious invariant is the volume $V_i$ of each flux tube $g_t L\boldsymbol{T}_i$, that is,

$$V_i = \mathrm{Vol}(L\boldsymbol{T}_i) = \mathrm{Vol}(g_t L\boldsymbol{T}_i) = \iiint_{g_t L T_i} \mathrm{d(vol)} \qquad [10]$$

which remains unchanged because of incompressibility.

Another invariant of fluid flow is defined as follows:

**Definition** Let $L$ be a magnetic link. For each solid torus $\boldsymbol{T}_i$, choose a meridional disk $D_i$. The magnetic flux $\Phi_i = \Phi(L\boldsymbol{T}_i)$ in the $i$th component is the surface integral defined as

$$\Phi_i = \Phi(L\boldsymbol{T}_i) : \iint_{LD_i} \boldsymbol{B}\cdot\boldsymbol{U}\,\mathrm{d(area)}$$

where $\boldsymbol{U}$ denotes the normal to the surface $LD_i$ pointing in the positive direction induced by the $\boldsymbol{B}$ field.

It can be shown that $\Phi_i$ is independent of the chosen meridional disk. It also can be shown that each $\Phi_i$ is a fluid flow invariant, that is,

$$\Phi_i(g_t L\boldsymbol{T}_i) = \iint_{g_t LD_i} \boldsymbol{B}\cdot\boldsymbol{U}\,\mathrm{d(area)} \qquad [11]$$

is independent of $t$.

One more fluid invariant that will play a central role in the energy minimization of magnetic links is given by the following definition.

**Definition** The helicity of a magnetic link $L$ is defined as

$$H(L) = \iiint_{\cup_i LT_i} \boldsymbol{A}\cdot\boldsymbol{B}\,\mathrm{d(vol)}$$

The term helicity was first introduced in a fluid context by Moffat, and it was previously used in particle physics for the scalar product of the momentum and spin of a particle. In another connection, note that the helicity $H(L)$ is the same as the Chern–Simons action:

$$H(L) = \int \boldsymbol{A}\wedge \mathrm{d}\boldsymbol{A} \\ = \int \mathrm{tr}(\boldsymbol{A}\wedge \mathrm{d}\boldsymbol{A} + \tfrac{2}{3}\boldsymbol{A}\wedge\boldsymbol{A}\wedge\boldsymbol{A}) \qquad [12]$$

where $\boldsymbol{A}$ now denotes the magnetic vector potential as a 1-form.

It can be shown that $H(L)$ is gauge invariant, and hence well defined.

**Theorem** (Moffat). *The helicity is invariant under fluid flow, that is,*

$$\frac{\mathrm{d}}{\mathrm{d}t}H(g_t L) = 0$$

Arnol'd (1998) defines the helicity in a more abstract setting and shows that it is invariant under the group $S(\mathrm{Diff})$ of volume-preserving diffeomorphisms.

The following theorem summarizes the many results due to Moffat, Ricca, Berger, Lomonaco, Hornig, Kauffman, and others, relating the helicity of magnetic links to linking and to magnetic flux.

**Theorem** *Let $L$ be a magnetic link. Then*

$$H(L) = \sum_{1\leq i\leq n} \Phi^2{}_i \mathrm{SL}_{F_i} + 2\sum_{1\leq i<j\leq n} \Phi_i\Phi_j \mathrm{LK}_{ij}$$

*where $\mathrm{SL}_{F_i}$ denotes the self-linking number of the axis curve of the tube $L\boldsymbol{T}_i$ with respect to the framing $F_i$ induced by the integral curves of the magnetic field $\boldsymbol{B}$ within $L\boldsymbol{T}_i$, and $\mathrm{LK}_{ij}$ denotes the linking number between any integral curve of*

*the magnetic field* $\boldsymbol{B}$ *in* $LT_i$ *with any integral curve of the magnetic field* $\boldsymbol{B}$ *in* $LT_j$.

**Remark** In fact, $\mathrm{SL}_{F_i}$ is the same as the linking number between any two integral curves of the magnetic field $\boldsymbol{B}$ within the tube $LT_i$.

Thus, as many authors have showed, the helicity does reflect the topology and the geometry of the magnetic lines of force within a magnetic link. If, for example, $L$ has only one component, that is, $L$ is a magnetic knot, then

$$H(L) = \Phi^2 \mathrm{SL}_F(C) \qquad [13]$$

where $\mathrm{SL}_F(C)$ is the self-linking number of the axis curve $C$ of the knotted tube with respect to the framing $F$ induced by the integral curves of the magnetic field $\boldsymbol{B}$ within the magnetic knot. If, for example, the tube is knotted in the form of a trefoil and if the magnetic lines of force appear to be parallel to the axis curve when the trefoil is placed on a plane flat surface, then $\mathrm{SL} = \pm 3$ and

$$H = \pm 3\Phi^2 \qquad [14]$$

On the other hand, if for example, the magnetic lines of force induce the trivial framing in each component, then

$$H(L) = 2 \sum_{1 \le i < j \le n} \Phi_i \Phi_j LK_{ij} \qquad [15]$$

Thus, if $L$ is a magnetic two-component Hopf link with no twisting of the integral curves of the magnetic field within the components of $L$, then

$$H(L) = \pm 2\Phi_1 \Phi_2 \qquad [16]$$

because the self-linking number based on the $\boldsymbol{B}$-field framing is zero for each component, and the linking number between the two components is ±1.

### Energy of Magnetic Knots and Links

Let us conclude this section with the definition of the energy of a magnetic link.

**Definition** The magnetic energy $E_{\mathrm{M}}(L)$ of a magnetic link $L$ is defined by the classical formula

$$E_{\mathrm{M}}(L) = \frac{1}{8\pi} \iiint_{\cup_i LT_i} |\mathbf{B}|^2 \, \mathrm{d(vol)} \ \text{(Gaussian units)}$$

Although the energy $E_{\mathrm{M}}$ is not flow invariant, it will play a central role in magnetic relaxation of knots and minimum energy magnetic links.

Consider a magnetic link $L$ in a perfectly conducting, incompressible, viscous fluid. As a result of dissipative frictional fluid forces, the magnetic energy $E_{\mathrm{M}}(g_t L)$ of $g_t L$ will decrease with time $t$. In losing energy, the magnetic lines of force will contract. On the other hand, since this is a volume-preserving process, the cross sections of the flux tubes of $g_t L$ will at the same time expand. These changes of topology occur while the flux $\Phi$, volume $V$, and helicity of $g_t L$ will remain the same. In other words, knotted magnetic flux tubes left free to evolve in such a fluid will do so by conserving their magnetic flux $\Phi$ and volume $V$, but converting their magnetic energy into kinetic energy, which in turn dissipates by internal friction. Magnetic links and knots evolve from high to low magnetic energy levels, conserving topology; and because of the induced shortening of field lines under conservation of volume, they become fatter and fatter, with an increase of the average tube cross section.

This process cannot continue indefinitely. Eventually, the magnetic flux tubes of $g_t L$ must make contact with each other. In other words, the topology of the magnetic link $g_t L$, as expressed in knotting and linking, creates a barrier to the full dissipation of the magnetic link's energy, that is, $E_{\mathrm{M}}(g_t L)$ has a positive lower bound that results from the topology of $g_t L$. That means, in other words, that relaxation is obstructed by the knottedness and entanglement of the field lines, and a minimum magnetic energy is reached. Thus, the magnetic link will reach a nontrivial stable and invariant energy state, much as Kelvin conjectured his atomic vortices would.

Various estimates of magnetomechanical energy in terms of topological quantities have been put forward in recent years (see Freedmann and He (1994)). These relations give lower bounds for the energy levels attainable by knot or link types by taking into account the effects that linking numbers and number of crossings have on the energy of the relaxed state. These bounds are expressed by relationship of the kind

$$E_{\min} \ge \phi(C_{\min}, \Phi, V, N) \qquad [17]$$

where $E_{\min}$ is the equilibrium energy and $\phi$ gives the relationship between physical quantities – such as total flux $\Phi$, number of tubes $N$, magnetic volume $V$ – and topology, given here by the minimal possible number of crossing $C_{\min}$. These relations offer numerous advantages due to the explicit dependence on qualitative properties of the flow field. A simple example is provided by the analysis of three braids, which shows that magnetic energy grows quadratically in time due to random braiding. This means that the least possible amount of magnetic energy that can be attained by the physical knot or link is determined purely by its topology. If topological information sets the levels of minimum energy accessible to the knot or link, geometric properties

may also influence the relaxation process. Considerations of helicity and linking numbers, for example, demonstrate that internal rearrangement of magnetic field geometry leads to a spectrum of different asymptotic endstates with the same topology. Moreover, magnetic knots have a natural tendency to get rid of excessive torsion of field lines and *S*-shaped tube geometries, and this may influence the relaxation process.

Since the helicity $H(g_t L)$ is both an invariant of fluid flow and an expression of the magnetic link $g_t L$'s topology, the following theorem, first stated by Moffat, is a mathematical expression of this topological bounds.

**Theorem** *Let L be a magnetic link. Then*

$$E_{\mathrm{M}}(L) \geq q_0 |H(L)|$$

*where $q_0$ is a nonzero constant that is independent of the magnetic link.*

Freedman and He (1991) obtain more subtle and tighter topological bounds on the minimum energy of magnetic links. For example, for a magnetic knot $K$, they prove that

$$E_{\mathrm{M}}(K) \geq \frac{1}{4\pi^{5/4}} \frac{\Phi(K)^{3/2}\ \mathrm{ac}(K)^{3/4}}{V(K)^{1/3}} \geq \frac{1}{4\pi^{5/4}} \frac{\Phi(K)^{3/2}(2g(K)-1)}{V(K)^{1/3}} \quad \text{(Gaussian units)} \qquad [18]$$

where $V(K)$ denotes the volume of the magnetic knot $K$, $\Phi(K)$ denotes the flux in $K$, $\mathrm{ac}(K)$ is the asymptotic crossing number, and $g(K)$ is the genus of the knot $K$. Freedman and He conjecture that $\mathrm{ac}(K) = \mathrm{c}(K)$, where $\mathrm{c}(K)$ is the crossing number, that is, the minimum number of crossings among all plane diagrams representing the knot $K$. Besides, Moffat (1990) suggests that the minimum energy spectrum of a magnetic knot can be used to construct new knot invariants.

## Topological Changes, Dissipation, and Reconnection in Fluid Patterns

As we saw above, topological changes do occur when dissipative effects become predominant over the coherency of structures. When this happens, there is a dramatic change of fluid patterns, often on small timescales compared to evolution. The change occurs through the formation and disappearance of physical reconnections in the fluid pattern. In real fluids, for example, vortex and magnetic tubes do interact and reconnect freely. From a dynamical system viewpoint, reconnections take place when the vector field lines (streamlines, vortex lines, or magnetic lines) cross each other. If two field lines meet, the point of crossing is a true nodal point, like a bifurcation in a path. Dissipative effects allow the reconnection to proceed through such points.

In dissipative fluids, mathematical and physical properties are no longer conserved, and during the process we lose part of the original information. However, some of the invariants are rather robust and may only degrade slowly. One of them is magnetic helicity, the magnetic analog of the kinetic helicity. Its dissipation during reconnection can be modest; in particular, if the reconnection timescale is small compared to classical dissipation times, then helicity loss will be negligible. The robustness of magnetic helicity plays a central role in fusion plasma physics and in many astrophysical contexts. On the other hand, large changes in kinetic helicity are intimately related to qualitative changes in the topology of vortex flows.

Under Euler's equations, the helicity of a vortex tube of vorticity $\omega$ and velocity $\boldsymbol{u}$ is defined by $H = \int \boldsymbol{u} \cdot \omega \, \mathrm{d}V$. The integral is taken over the tube volume $V$ occupied by $\omega$. Now, for $n$ knotted and linked vortex tubes, each of (constant) strength (total vorticity) $\Phi_i (1 \leq i \leq N)$, the helicity of the whole system can be expressed in terms of linking numbers $\mathrm{Lk}_{ij}$ as

$$H = \sum_{ij} \mathrm{Lk}_{ij} \Phi_i \Phi_j \cdot \mathrm{Lk}_{ij}$$

which is equal to $\mathrm{Lk}_{ji}$; this is a topological invariant whose value does not change under continuous deformation of the fluid structure. Since helicity and flux-tube strength are measurable conserved quantities, the above equation provides useful information about the topology of the flow field and flow structures. In addition, by direct measurements of helicity and application of conservation of topology, one can estimate average geometric quantities, such as the mean twist of field lines, and their contribution to the total energy.

## Brief Conclusion

In this article, we have made an attempt to indicate how "classical" field theories, which have been successfully used to describe physics of fundamental structures and forces of nature, can also be used to study geometry and topology of low-dimensional manifolds. These developments not only provide new insights into old problems of topology of these manifolds but also have been responsible for profoundly interesting new mathematics (fluid mechanics, dynamical flows, and polymer biophysics

are maybe the most significant examples in the last years). In particular fluid dynamics, a topological macroscopic field theory, provides a powerful framework for modern theory of knots and links in 3-manifolds. Moreover, as we saw here, it provides a physical interpretation of the link, self-linking, and writhing number of knots and links. The present article was essentially aimed to illustrate such a relationship. Thus, the most fundamental result we reported here is the relation (formula) connecting the helicity of vector (magnetic) fields to the writhing number of knots: $H(V) = \text{Flux}(V)^2\ \text{Wr}(K)$. So, writhing number for knots is the analog of helicity for vector fields. Both expressions of these invariants are variants of the (Gaussian) integral formula for the linking number of two disjoint closed space curves. Further investigations of these invariants and their mathematical properties might throw new light on the interfaces between many different areas of macroscopic and quantum physics.

*See also:* The Jones Polynomial; Knot Theory and Physics; Magnetohydrodynamics; Mathematical Knot Theory; Stability of Flows; Superfluids; Topological Quantum Field Theory: Overview; Vortex Dynamics; Yang–Baxter Equations.

## Further Reading


Arnol'd V and Khesin B (1998) *Topological Methods in Hydrodynamics*. Heidelberg: Springer.

Berger MA and Field GB (1984) The topological properties of magnetic helicity. *Journal of Fluid Mechanics* 147: 133–148.

Berry MV and Dennis MR (2001) Knotted and linked phase singularities in nonchromatic waves. *Proceedings of the Royal Society A* 457: 2251–2263.

Boi L (2005) Topological knots' models in physics and biology. In: Boi L (ed.) *Geometries of Nature, Living Systems and Human Cognition. New Interactions of Mathematics with Natural Sciences and Humanities*, pp. 211–294. Singapore: World Scientific.

Boyland P (2001) Fluid mechanics and mathematical structures. In: Ricca RL (ed.) *An Introduction to the Geometry and Topology of Fluid Flows*, pp. 105–134. NATO-ASI Series: Mathematics. Dordrecht: Kluwer.

Cantarella J, De Turk D, and Gkuck H (2001) The Biot–Savart operator for application to knot theory fluid dynamics, and plasma physics. *Journal of Mathematical Physics* 42: 876–905.

Freedman MH and Zheng-Xu He (1991) Divergence free fields: energy and asymptotic crossing number. *Annals of Mathematics* 134: 189–229.

Fuller FB (1978) Decomposition of the linking number of a closed ribbon: a problem from molecular biology. *Proceedings of the National Academy of Sciences, USA* 75: 3557–3561.

Ghrist RW, Holmes PhJ, and Sullivan MC (1997) *Knots and Links in Three-Dimensional Flows*. Heidelberg: Springer.

Hornig G (2002) *Topological Methods in Fluid Dynamics. Preprint*. Ruhr-Universität-Bochum.

Kauffman LH (1995) Knots and Applications. *Series on Knots and Everything*, vol. 6, Singapore: World Scientific.

Lomonaco SJ (1995) The modern legacies of Thomson's atomic vortex theory in classical electrodynamics. In: Kauffman LH (ed.) *The Interface of Knots and Physics*, Proc. Symp. Appl. Math., vol. 51, pp. 145–166. American Mathematical Society.

Moffatt HK (1969) The degree of knottedness of tangled vortex lines. *Journal of Fluid Mechanics* 35: 117–129.

Moffat HK (1990) The energy spectrum of knots and links. *Nature* 347: 367–369.

Moffatt HK, Zaslavsky GM, Comte P, and Tabor M (1992) *Topological Aspects of the Dynamics of Fluids and Plasmas*, NATO ASI Series, Series E: Applied Sciences, vol. 218. Dordrecht: Kluwer Academic.

Ricca RL (1998) New developments in topological fluid mechanics: from Kelvin's vortex knots to magnetic knots. In: Stasiak A, Katritch V, and Kauffman LH (eds.) *Ideal Knots*. Singapore: World Scientific.

Ricca RL and Moffat HK (1992) The helicity of a knotted vortex filament. In: Moffat HK (ed.) *Topological Aspects of Dynamics of Fluids and Plasmas*, pp. 225–236. Dordrecht: Kluwer.

Tait PG (1900) On Knots I, II, III. In: *Scientific Papers*. Cambridge: Cambridge University Press.

Thomson JJ (1883) *A Treatise on the Motion of Vortex Rings*. London: Macmillan.

Trueba JL and Rañada AF (2000) Helicity in classical electrodynamics and its topological quantization. *Apeiron* 7: 83–88.

Woltjer L (1958) A theorem on force-free magnetic fields. *Proceedings of the National Academy Sciences, USA* 44: 489–491.


# Topological Quantum Field Theory: Overview

**J M F Labastida**, CSIC, Madrid, Spain
**C Lozano**, INTA, Torrejón de Ardoz, Spain



## Introduction

Topological quantum field theory (TQFT) constitutes one of the most successful fields of mathematical physics since it originated in the 1980s. It possesses an inherent property which makes it unique: TQFT provides predictions in mathematics which open new fields of research. A well-known example is the prediction of Seiberg–Witten invariants as building blocks of Donaldson invariants. However, there are others such as the recent proposal for the coefficients of the HOMFLY polynomial invariants for knots as quantities related to enumerative geometry. These developments have drawn the attention of mathematicians and physicists into TQFT since the 1980s, a very fruitful period in which both communities have benefited from each other.

Topology has always been present in mathematical physics, in particular when dealing with aspects of

quantum physics. Global effects play an important role in quantum-mechanical models and topology becomes an essential ingredient in their description. TQFT itself appeared in the winter of 1987 after Witten's work (Witten 1988a) on Donaldson theory (Donaldson 1990), but a series of papers during the 1980s which dealt with topological aspects of field and string theory anticipated its existence. Two of these correspond to Witten's works on supersymmetric quantum mechanics and supersymmetric sigma models (Witten 1982) that led to a generalization of Morse theory. This generalization was considered by Floer (1987) in a new context that constituted the key element in Witten's construction of TQFT. These developments were certainly influenced by Atiyah (1988). TQFT was born as a result of the interplay between physics and mathematics. This has been a constant feature all along its development.

Soon after the formulation of the TQFT addressing Donaldson theory, now known as Donaldson–Witten theory, Witten formulated a new TQFT which focuses on knot invariants such as the Jones polynomial and its generalizations (Jones 1985). Witten (1989) constructed Chern–Simons gauge theory and proved its relation to the theory of knot and link invariants. This theory possesses different features than Donaldson–Witten theory, and in fact it turns out that these two theories fall into two different general types of TQFTs as will be explained in the following section. Anyhow, despite their formal differences, both Donaldson–Witten and Chern–Simons gauge theory emerged as a novel way to express topological invariants in terms of quantum field theory quantities as well as to generalize their previous formulation. But there was much more to them than it seemed in their beginnings. Once these topological invariants were formulated in field theory language, one had a huge machinery to study them from different points of view. Theoretical physicists have developed many useful tools to study quantum field theory. The use of these tools led to new frameworks for these topological invariants.

In this overview we are going to provide the basics of TQFT and briefly describe two examples – Donaldson–Witten theory and Chern–Simons gauge theory – to explain how the general features are implemented. Some excellent reviews on the subject (Birmingham *et al.* 1991, Cordes *et al.* 1996, Labastida and Mariño 2004) are available. The organization of this work is as follows. In the following section we present a general introduction to TQFT from a functional integral point of view. Next, we touch upon the twisting of extended supersymmetry as a general constructive approach to TQFT. This is followed by a section on Donaldson–Witten theory where we discuss the computation of its observables from a perturbative approach, showing their relation to the Donaldson invariants. Next, we introduce Chern–Simons gauge theory as a theory of knot and link invariants. The penultimate section deals with advanced developments in TQFT. Finally, we end up with some concluding remarks.

## Topological Quantum Field Theory

We will start our overview by presenting the most general structure of a TQFT from a functional integral point of view which, though not rigorously defined, is the approach that has led to the most important developments. As in conventional quantum field theory, axiomatic approaches to TQFT do exist, but we will not follow that route here.

Let us consider an $n$-dimensional Riemannian manifold $X$ endowed with a metric $g_{\mu\nu}$ and a quantum field theory on it. We will say that this theory is "topological" if there exist operators in the theory such that their correlation functions do not depend on the metric. If we denote these operators by $\mathcal{O}_i$ (where $i$ is a generic label), then

$$\frac{\delta}{\delta g_{\mu\nu}}\langle \mathcal{O}_{i_1}\cdots\mathcal{O}_{i_n}\rangle = 0 \qquad [1]$$

where $\langle\cdots\rangle$ denotes a vacuum expectation value. The operators that satisfy this equation are called "topological observables."

The simplest way to achieve metric independence is to consider a theory whose action and operators do not depend on the metric. In this situation, if no anomalous metric dependence is generated upon quantization, the correlation functions of these operators satisfy [1] and lead to topological invariants on $X$. Theories of this sort are collectively referred to as Schwarz-type TQFTs, and well-known examples are Chern–Simons gauge theory and BF theories. However, Schwarz-type theories are too restrictive. One would like to have a theory satisfying property [1] with a weaker condition on the action. This can be achieved with the help of a symmetry. The resulting TQFTs are called of Witten or cohomological type, the main examples being Donaldson–Witten theory and topological sigma models (Witten 1988b).

For TQFTs of Witten type, the action may depend on the metric. However, the theory has an underlying scalar symmetry $\delta$ acting on the fields $\phi_i$. Since $\delta$ is a symmetry, the action of the theory satisfies $\delta S(\phi_i) = 0$. In these theories, metric independence of the correlation functions is achieved as follows. Let $T_{\mu\nu} = (\delta/\delta g^{\mu\nu})S(\phi_i)$ be the energy–momentum tensor of

the theory. It turns out that the energy–momentum tensor is $\delta$-exact:

$$T_{\mu\nu} = -\mathrm{i}\delta G_{\mu\nu} \qquad [2]$$

$G_{\mu\nu}$ being some tensor. Indeed, if [2] is satisfied, it follows that for any set of operators $\mathcal{O}_i$ which are $\delta$-invariant,

$$\begin{aligned}\frac{\delta}{\delta g^{\mu\nu}}\langle \mathcal{O}_{i_1}\mathcal{O}_{i_2}\cdots\mathcal{O}_{i_n}\rangle &= \langle \mathcal{O}_{i_1}\mathcal{O}_{i_2}\cdots\mathcal{O}_{i_n}T_{\mu\nu}\rangle \\ &= -\mathrm{i}\langle \mathcal{O}_{i_1}\mathcal{O}_{i_2}\cdots\mathcal{O}_{i_n}\delta G_{\mu\nu}\rangle \\ &= \pm\mathrm{i}\langle \delta(\mathcal{O}_{i_1}\mathcal{O}_{i_2}\cdots\mathcal{O}_{i_n}G_{\mu\nu})\rangle \\ &= 0 \end{aligned} \qquad [3]$$

In this computation we have assumed that the symmetry $\delta$ is not anomalous and that there are no contributions coming from boundary terms since we have integrated by parts in field space. This is not always the case and in fact the situations in which one of these two properties fails lead to rich phenomena. In those cases, for example, in Donaldson–Witten theory on manifolds with $b_2^+ = 1$, the correlation functions fail to be topological invariants in a controlled manner which unveils many interesting properties.

We will now describe Witten-type theories in a general context. The general structure of Schwarz-type theories is much simpler and will be illustrated in the example presented below. In Witten-type theories the observables are the $\delta$-invariant operators. It is simple to prove that $\delta$-exact operators decouple from the theory. Indeed, if $\mathcal{O}_a$ is $\delta$-exact, $\mathcal{O}_a = \delta\hat{\mathcal{O}}_a$, then

$$\begin{aligned}\langle \mathcal{O}_a\mathcal{O}_{i_1}\mathcal{O}_{i_2}\cdots\mathcal{O}_{i_n}\rangle &= \langle \delta\hat{\mathcal{O}}_a\mathcal{O}_{i_1}\mathcal{O}_{i_2}\cdots\mathcal{O}_{i_n}\rangle \\ &= \langle \delta(\hat{\mathcal{O}}_a\mathcal{O}_{i_1}\mathcal{O}_{i_2}\cdots\mathcal{O}_{i_n})\rangle = 0\end{aligned} \qquad [4]$$

Thus, one can restrict the set of observables to the cohomology of $\delta$:

$$\mathcal{O} \in \frac{\operatorname{Ker}\delta}{\operatorname{Im}\delta} \qquad [5]$$

There is no reason *a priori* why the $\delta$-symmetry should be a scalar Grassmannian symmetry, but in all known models of Witten-type TQFTs this turns out to be the case. Thus, these theories violate the spin-statistics theorem. In all these models the algebra of the $\delta$ symmetry has the form

$$\delta^2 = Z \qquad [6]$$

where $Z$ is a symmetry transformation (typically a gauge symmetry of some sort). This property forces to consider $Z$-invariant observables and to work in the context of "equivariant cohomology."

The observables of Witten-type theories fit into a general pattern that we describe now. The key ingredient is a map between the homology of $X$ and the equivariant cohomology of $\delta$. Given an operator $\phi^{(0)}$ in the equivariant cohomology of $\delta$, let us consider the following set of equations:

$$\mathrm{d}\phi^{(n)} = \delta\phi^{(n+1)}, \quad n \geq 0 \qquad [7]$$

where the operators $\phi^{(n)}(n = 1, \ldots, \dim X)$ are differential forms of degree $n$ on $X$ and d is the de Rham differential. These differential equations are called "descent equations" and their solutions $\phi^{(n)}(n \geq 0)$ "topological descendants" of $\phi^{(0)}$. We will show how to construct a solution to these equations on general grounds.

The topological descendants lead to the construction of a set of elements of the equivariant cohomology of $\delta$. Let $\gamma_n$ be an $n$-cycle on $X$, $\gamma_n \in H_n(X)$, and let us consider the following operator:

$$W_{\phi^{(0)}}^{(\gamma_n)} = \int_{\gamma_n} \phi^{(n)} \qquad [8]$$

This operator is $\delta$-invariant,

$$\delta W_{\phi^{(0)}}^{(\gamma_n)} = \int_{\gamma_n} \delta\phi^{(n)} = \int_{\gamma_n} \mathrm{d}\phi^{(n-1)} = \int_{\partial\gamma_n} \phi^{(n-1)} = 0 \qquad [9]$$

since $\partial\gamma_n = 0$. On the other hand, if $\gamma_n$ were trivial in homology, that is, if $\gamma_n = \partial\gamma_{n+1}$, we would have that $W_{\phi^{(0)}}^{(\gamma_n)}$ is $\delta$-exact:

$$W_{\phi^{(0)}}^{(\gamma_n)} = \int_{\partial\gamma_{n+1}} \phi^{(n)} = \int_{\gamma_{n+1}} \mathrm{d}\phi^{(n)} = \delta\int_{\gamma_{n+1}} \phi^{(n+1)} \qquad [10]$$

Thus, given the operator $\phi^{(0)}$, we have constructed a map between the homology of $X$ and the equivariant cohomology of $\delta$. There are as many maps as basic operators $\phi^{(0)}$ one finds in the theory.

To actually construct these maps, we need to find a solution of the descent equations [7]. As announced before, there is a general solution to those equations in Witten-type theories. Since in this type of theories [2] holds, there exists an operator

$$G_\mu \equiv G_{0\mu} \qquad [11]$$

that satisfies

$$P_\mu = T_{0\mu} = -\mathrm{i}\delta G_\mu \qquad [12]$$

Notice that $G_\mu$ is an anticommuting operator and a 1-form in spacetime. With the aid of this operator, one constructs the following solution to the descent equations [7]:

$$\phi^{(n)} = \frac{1}{n!}\phi^{(n)}_{\mu_1\mu_2\ldots\mu_n}\,\mathrm{d}x^{\mu_1} \wedge \cdots \wedge \mathrm{d}x^{\mu_n} \qquad [13]$$

where

$$\phi^{(n)}_{\mu_1\mu_2\ldots\mu_n}(x) = G_{\mu_1}G_{\mu_2}\cdots G_{\mu_n}\phi^{(0)}(x), \quad n = 1, \ldots, \dim X \qquad [14]$$

One can easily check using [12] and the $\delta$-invariance of $\phi^{(0)}$ that the operators [13] do satisfy the descent equations [7].

We have seen that Witten-type TQFTs are characterized by property [2]. It would be desirable to have at hand a systematic procedure to build theories satisfying that property. It has been found that extended supersymmetry provides a very helpful starting point to build those theories. Although supersymmetry guarantees from first principles only the weaker condition [12] instead of [2], all TQFTs that have been constructed from extended supersymmetry actually satisfy [2]. To build a TQFT from a theory with extended supersymmetry, one needs to go through the twisting procedure that we now describe.

## Twisting of Extended Supersymmetry

All known Witten-type theories are related to an underlying extended supersymmetric quantum field theory. The topological theory is a modified version of the supersymmetric theory in which the Lorentz transformation properties (spins) of some of the fields have been modified. This modification of spin assignments is known as twisting, and it can be carried out on any theory with extended supersymmetry in any spacetime dimension. We will not consider the procedure in such a general setting but instead we will illustrate it by considering the case of $\mathcal{N}=2$ supersymmetry in four dimensions. We will begin with a general description and then we will apply it to a specific example: Donaldson–Witten theory.

Let us consider the Euclidean version of the $\mathcal{N}=2$ supersymmetry algebra with no central charges. Central charges can be included without much ado but we will not consider them for simplicity. The total symmetry group of the theory is $\mathcal{H}=\mathrm{SU}(2)_+\times \mathrm{SU}(2)_-\times \mathrm{SU}(2)_R \times \mathrm{U}(1)_R$, $\mathcal{K}=\mathrm{SU}(2)_+\times \mathrm{SU}(2)_-$ being the rotation group, and $\mathrm{SU}(2)_R\times \mathrm{U}(1)_R$ the internal symmetry group of the $\mathcal{N}=2$ supersymmetry algebra. The generator algebra takes the following form:

$$
\begin{aligned}
&\{Q_{\alpha v},\overline{Q}_{\dot\beta w}\} = 2\epsilon_{vw}\sigma^{\mu}_{\alpha\dot\beta}P_\mu, && \{Q_{\alpha v},Q_{\beta w}\}=0\\
&[P_\mu,Q_{\alpha v}]=0, && [P_\mu,\overline{Q}_{\dot\alpha v}]=0\\
&[M_{\alpha\beta},Q_{\delta v}]=\epsilon_{\delta(\alpha}Q_{\beta)v}, && [M_{\alpha\beta},\overline{Q}_{\dot\alpha v}]=0\\
&[\overline{M}_{\dot\alpha\dot\beta},Q_{\delta v}]=0, && [\overline{M}_{\dot\alpha\dot\beta},\overline{Q}_{\dot\delta v}]=\epsilon_{\dot\delta(\dot\alpha}\overline{Q}_{\dot\beta)v}\\
&[B^{vw},Q^{u}_{\alpha}]=\epsilon^{u(v}Q^{w)}_{\alpha}, && [B^{vw},\overline{Q}^{u}_{\dot\alpha}]=-\epsilon^{u(v}\overline{Q}^{w)}_{\dot\alpha}\\
&[Q_{\alpha v},R]=Q_{\alpha v}, && [\overline{Q}_{\dot\alpha v},R]=-\overline{Q}_{\dot\alpha v}
\end{aligned}
\qquad [15]
$$

In these relations $v,w\in\{1,2\}$ are $\mathrm{SU}(2)_R$ indices and $\alpha$ and $\dot\alpha$ denote spinorial indices of $\mathrm{SU}(2)_-$ and $\mathrm{SU}(2)_+$, respectively. The supersymmetry generators $Q_{\alpha v}$ and $\overline{Q}_{\dot\alpha v}$ transform under $\mathcal{H}$ as $(\mathbf{0},\mathbf{2},\mathbf{2})^1$ and $(\mathbf{2},\mathbf{0},\mathbf{2})^{-1}$, respectively. $\overline{M}_{\dot\alpha\dot\beta}$ and $M_{\alpha\beta}$ are the generators of $\mathrm{SU}(2)_+$ and $\mathrm{SU}(2)_-$, respectively, while $B^{vw}$ and $R$ generate $\mathrm{SU}(2)_R$ and $\mathrm{U}(1)_R$, respectively.

The twisting of a supersymmetric theory involves a modification of the couplings of the theory to a background metric on the space where the theory is defined. This modification is carried out redefining the Lorentz transformation properties of the different fields making use of the internal symmetry $\mathrm{SU}(2)_R$. In particular, we will redefine the couplings of the fields to the $\mathrm{SU}(2)_+$ spin connection according to the way they transform under $\mathrm{SU}(2)_R$. This is easily done by identifying the $\mathrm{SU}(2)_R$ indices $v$ with the $\mathrm{SU}(2)_+$ indices $\dot\alpha$. The procedure involves a redefinition of the rotation group into $\mathcal{K}'=\mathrm{SU}'(2)_+\otimes \mathrm{SU}(2)_-$, where $\mathrm{SU}'(2)_+$ is generated by

$$M'_{\dot\alpha\dot\beta}=M_{\dot\alpha\dot\beta}-B_{\dot\alpha\dot\beta} \qquad [16]$$

The supersymmetry generators $Q_{\alpha v}$ and $\overline{Q}_{\dot\alpha v}$ get transformed in the following way:

$$
\begin{aligned}
\overline{Q}_{\dot\alpha v} &\to \overline{Q}_{\dot\alpha\dot\beta}\\
Q_{\alpha v} &\to Q_{\alpha\dot\beta}
\end{aligned}
\qquad [17]
$$

which allows us to define the "topological supercharge":

$$\overline{\mathcal{Q}}\equiv \epsilon^{\dot\alpha\dot\beta}\overline{Q}_{\dot\alpha\dot\beta} \qquad [18]$$

It is simple to prove using [15] and [16] that this quantity is a scalar under the new rotation group $\mathcal{K}'$: $[M_{\alpha\beta},\overline{\mathcal{Q}}]=0$ and $[M'_{\dot\alpha\dot\beta},\overline{\mathcal{Q}}]=0$. In addition, from [15], it follows that $\overline{\mathcal{Q}}$ is nilpotent (in the absence of central charges):

$$\overline{\mathcal{Q}}^2=0 \qquad [19]$$

The scalar generator $\overline{\mathcal{Q}}$ leads to the topological symmetry $\delta$ of the previous section. Actually, the twisting procedure provides also the operator $G_\mu$ in [12]. Defining

$$G_\mu=\frac{\mathrm{i}}{4}(\overline{\sigma}_\mu)^{\dot\alpha\gamma}Q_{\gamma\dot\alpha} \qquad [20]$$

one easily finds, after using [15] and [18],

$$\{\overline{\mathcal{Q}},G_\mu\}=\partial_\mu \qquad [21]$$

which is indeed equivalent to [12]. On general grounds we cannot prove that twisted supersymmetric theories lead to theories which satisfy [12]. However relation [12], which is weaker, is guaranteed. It turns out that in all the models originated from extended supersymmetry which have been studied, [2] is satisfied and thus the resulting theories are TQFTs of Witten type.

## Donaldson–Witten Theory

One of the greatest successes of TQFT has been the discovery of Seiberg–Witten invariants as building blocks of Donaldson invariants. This was achieved in two main steps. First, Donaldson theory was reformulated in field-theoretical terms, using perturbative methods. Second, the resulting TQFT was solved using nonperturbative methods. In this section we are going to describe in some detail the first step. The second one will be briefly addressed later and is the main object of a separate article in the encyclopedia (*see* Seiberg–Witten Theory).

Let us consider $\mathcal{N}=2$ supersymmetric Yang–Mills theory in four dimensions. The field content of the theory is the following: a gauge field $A_\mu$, two spinors $\lambda_{v\alpha}$, and a complex scalar $\phi$, all of them in the adjoint representation of a gauge group $\mathcal{G}$. In addition, the theory possesses the auxiliary fields $D_{vw}$ in the **3** of the internal $\mathrm{SU}(2)_R$. The theory has the following action:

$$\int \mathrm{d}^4x\,\mathrm{tr}\Big(\nabla_\mu\phi^\dagger\nabla^\mu\phi - \mathrm{i}\lambda_v\sigma^\mu\nabla_\mu\overline{\lambda}^v - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{4}D_{vw}D^{vw} - \frac{1}{2}[\phi,\phi^\dagger]^2 - \frac{\mathrm{i}}{\sqrt{2}}\epsilon^{vw}\lambda_v{}^\alpha[\phi^\dagger,\lambda_{w\alpha}] - \frac{\mathrm{i}}{\sqrt{2}}\epsilon_{vw}\overline{\lambda}^v{}_{\dot\alpha}[\overline{\lambda}^{w\dot\alpha},\phi]\Big) \qquad [22]$$

This action is invariant under the following $\mathcal{N}=2$ supersymmetric transformations:

$$\begin{aligned}
\delta\phi &= \sqrt{2}\epsilon^{vw}\xi_v\lambda_w\\
\delta A_\mu &= \mathrm{i}\xi_v\sigma_\mu\overline{\lambda}^v - \mathrm{i}\lambda_v\sigma_\mu\overline{\xi}^v\\
\delta\lambda_{v\alpha} &= D_v{}^w\xi_{w\alpha} - \mathrm{i}\xi_{v\alpha}[\phi,\phi^\dagger] - \mathrm{i}\sigma^{\mu\nu}{}_\alpha{}^\beta\xi_{v\beta}F_{\mu\nu}\\
&\quad + \mathrm{i}\sqrt{2}\epsilon_{vw}\sigma^\mu_{\alpha\dot\alpha}\overline{\xi}^{w\dot\alpha}\nabla_\mu\phi\\
\delta D^{vw} &= 2\mathrm{i}\overline{\xi}^{(v}\overline{\sigma}^\mu\nabla_\mu\lambda^{w)} + 2\mathrm{i}\nabla_\mu\overline{\lambda}^{(v}\overline{\sigma}^\mu\xi^{w)}\\
&\quad + 2\mathrm{i}\sqrt{2}\xi^{(v}[\lambda^{w)},\phi^\dagger] + 2\mathrm{i}\sqrt{2}\overline{\xi}^{(v}[\overline{\lambda}^{w)},\phi]
\end{aligned} \qquad [23]$$

$\xi_v$ being spinorial $\mathcal{N}=2$ supersymmetric parameters.

We can now twist the above theory following the procedure explained in the previous section. Upon twisting, the fields of the theory change their spin content as follows:

$$\begin{aligned}
A_\mu(\mathbf{2},\mathbf{2},\mathbf{0})^0 &\to A_\mu(\mathbf{2},\mathbf{2})^0\\
\lambda_{\alpha v}(\mathbf{2},\mathbf{0},\mathbf{2})^1 &\to \psi_{\alpha\dot\beta}(\mathbf{2},\mathbf{2})^1\\
\overline{\lambda}_{\dot\alpha v}(\mathbf{0},\mathbf{2},\mathbf{2})^{-1} &\to \eta(\mathbf{0},\mathbf{0})^{-1},\chi_{\dot\alpha\dot\beta}(\mathbf{3},\mathbf{0})^{-1}\\
\phi(\mathbf{0},\mathbf{0},\mathbf{0})^2 &\to \phi(\mathbf{0},\mathbf{0})^2\\
\phi^\dagger(\mathbf{0},\mathbf{0},\mathbf{0})^{-2} &\to \phi^\dagger(\mathbf{0},\mathbf{0})^{-2}\\
D_{vw}(\mathbf{0},\mathbf{0},\mathbf{3})^0 &\to D_{\dot\alpha\dot\beta}(\mathbf{0},\mathbf{0})^0
\end{aligned} \qquad [24]$$

In this table the representations of the respective rotation groups carried by the fields have been indicated. The superindices refer to the $\mathrm{U}(1)_R$ charge which is also called "ghost number" in the context of TQFT. The fields $\eta$ and $\chi$ are given by the antisymmetric and symmetric pieces of $\bar\lambda_{\dot\alpha\dot\beta}: \chi_{\dot\alpha\dot\beta} = \bar\lambda_{(\dot\alpha\dot\beta)}$ and $\eta = (1/2)\epsilon^{\dot\alpha\dot\beta}\bar\lambda_{\dot\alpha\dot\beta}$.

Notice that the twisted fields in [24] are differential forms on $X$; therefore, the twisted theory makes sense globally on any arbitrary Riemannian 4-manifold. This is not the case with the original $\mathcal{N}=2$ supersymmetric Yang–Mills, which contains fermionic fields. Making global sense of those on arbitrary Riemannian 4-manifolds requires the manifold to be Spin.

The dynamics of the twisted theory is governed by an action which can be obtained by twisting the action [22]. On an arbitrary Riemannian 4-manifold endowed with a metric $g_{\mu\nu}$, the twisted action becomes

$$\begin{aligned}
S = \int \mathrm{d}^4x\sqrt{g}\,\mathrm{tr}\Big(&\nabla_\mu\phi\nabla^\mu\phi^\dagger - \mathrm{i}\psi^{\dot\beta}{}_\alpha\overline{\sigma}^{\mu\dot\alpha\alpha}\nabla_\mu\chi_{\dot\alpha\dot\beta}\\
&- \mathrm{i}\psi_{\alpha\dot\alpha}\nabla^{\dot\alpha\alpha}\eta - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{4}D_{\dot\alpha\dot\beta}D^{\dot\alpha\dot\beta}\\
&- \frac{1}{2}[\phi,\phi^\dagger]^2 - \frac{\mathrm{i}}{\sqrt{2}}\chi^{\dot\alpha\dot\beta}[\phi,\chi_{\dot\alpha\dot\beta}]\\
&+ \mathrm{i}\sqrt{2}\eta[\phi,\eta] - \frac{\mathrm{i}}{\sqrt{2}}\psi_{\alpha\dot\alpha}[\psi^{\alpha\dot\alpha},\phi^\dagger]\Big)
\end{aligned} \qquad [25]$$

where $\sqrt{g} = (\det(g_{\mu\nu}))^{1/2}$.

To obtain the transformations of the fields under the topological symmetry, we need to compute the $\overline{\mathcal{Q}}$-transformations. These are easily obtained using [18] and [23]. They turn out to be

$$\begin{aligned}
[\overline{\mathcal{Q}},\phi] &= 0\\
[\overline{\mathcal{Q}},A_\mu] &= \psi_\mu\\
\{\overline{\mathcal{Q}},\eta\} &= [\phi,\phi^\dagger]\\
\{\overline{\mathcal{Q}},\psi_\mu\} &= 2\sqrt{2}\nabla_\mu\phi\\
[\overline{\mathcal{Q}},\phi^\dagger] &= 2\sqrt{2}\mathrm{i}\eta\\
\{\overline{\mathcal{Q}},\chi_{\dot\alpha\dot\beta}\} &= \mathrm{i}(F^+_{\dot\alpha\dot\beta} - D_{\dot\alpha\dot\beta})\\
[\overline{\mathcal{Q}},D] &= (2\nabla\psi)^+ + 2\sqrt{2}[\phi,\chi]
\end{aligned} \qquad [26]$$

where $\psi_\mu = \sigma_{\mu\alpha\dot\beta}\psi^{\alpha\dot\beta}$ and $F^+_{\dot\alpha\dot\beta} = \sigma^{\mu\nu}_{\dot\alpha\dot\beta}F_{\mu\nu}$ is the self-dual part of $F_{\mu\nu}$. Using these transformations, one easily finds that $\overline{\mathcal{Q}}^2$ is a gauge transformation. This is not unexpected since the $\mathcal{N}=2$ supersymmetric transformations [23] are in the Wess–Zumino gauge and they close only up to gauge transformations. This property implies that one must consider the equivariant cohomology of $\overline{\mathcal{Q}}$ defined on the set of gauge-invariant operators.

The action [25] is $\overline{\mathcal{Q}}$-exact up to a topological term:

$$S = \{\overline{\mathcal{Q}}, V\} - \frac{1}{2}\int F \wedge F \qquad [27]$$

where

$$V = \int d^4x\sqrt{g}\,\mathrm{tr}\left(\frac{i}{4}\chi_{\dot{\alpha}\dot{\beta}}(F^{\dot{\alpha}\dot{\beta}} + D^{\dot{\alpha}\dot{\beta}}) - \frac{1}{2}\eta[\phi, \phi^\dagger] + \frac{1}{2\sqrt{2}}\psi_{\alpha\dot{\alpha}}\nabla^{\dot{\alpha}\alpha}\phi^\dagger\right) \qquad [28]$$

Actually, it turns out that in all the theories obtained after twisting extended supersymmetry, the resulting actions are $\overline{\mathcal{Q}}$-exact up to topological terms. In the case of $\mathcal{N} = 2$ theories, topological (theta) terms $\int F \wedge F$ are generically not observable (due to a chiral anomaly), so it is customary to pick

$$S_{\mathrm{DW}} = \{\overline{\mathcal{Q}}, V\} \qquad [29]$$

as the action of the theory, which immediately implies [2] and therefore the topological character of the theory. Notice, however, that [29] is stronger than [2].

As we described in the previous section, the observables of the theory can be constructed using the operator $G_\mu$ in [20]. Its action on the twisted fields is easily obtained using [23]:

$$\begin{aligned}
&[G_\mu, \phi] = \frac{1}{2\sqrt{2}}\psi_\mu \\
&[G_\nu, A_\mu] = \frac{i}{2}g_{\mu\nu}\eta - i\chi_{\mu\nu} \\
&[G, \eta] = -\frac{i\sqrt{2}}{4}\nabla\overline{\phi} \\
&\{G_\mu, \psi_\nu\} = -(F^-_{\mu\nu} + D^+_{\mu\nu}) \\
&[G, \overline{\phi}] = 0 \\
&[G, F^+] = i\nabla\chi + \frac{3i}{2} * \nabla\eta \\
&\{G, \chi\} = -\frac{3i\sqrt{2}}{8} * \nabla\overline{\phi} \\
&[G, D] = -\frac{3i}{4} * \nabla\eta + \frac{3i}{2}\nabla\chi
\end{aligned} \qquad [30]$$

We now need to fix the basic operator $\phi^{(0)}$ in [14]. The starting point must be a set of gauge-invariant, $\overline{\mathcal{Q}}$-closed operators which are not $\overline{\mathcal{Q}}$-trivial. Since $[\overline{\mathcal{Q}}, \phi] = 0$, these operators are the gauge-invariant polynomials in the field $\phi$. For a simple gauge group of rank $r$ the algebra of these polynomials is generated by $r$ elements, and we shall denote this basis by $\mathcal{O}_n, n = 1, \ldots, r$. A simple choice for SU($N$) consists of the following Casimirs:

$$\mathcal{O}_n = \mathrm{tr}(\phi^{n+1}), \quad n = 1, \ldots, N \qquad [31]$$

Using $G_\mu$ we can now construct the map between the homology of $X$ and the equivariant cohomology of $\overline{\mathcal{Q}}$. Let us consider the simple case SU(2). There exists only one independent Casimir and, correspondingly, only one basic operator:

$$\mathcal{O} = \mathrm{tr}(\phi^2) \qquad [32]$$

for which one finds the following set of descendants:

$$\begin{aligned}
\mathcal{O}^{(1)} &= \mathrm{tr}\left(\frac{1}{\sqrt{2}}\phi\psi_\mu\right)dx^\mu \\
\mathcal{O}^{(2)} &= -\frac{1}{2}\mathrm{tr}\left(\frac{1}{\sqrt{2}}\phi(F^-_{\mu\nu} + D_{\mu\nu}) - \frac{1}{4}\psi_\mu\psi_\nu\right)dx^\mu \wedge dx^\nu \\
&\ \ \vdots
\end{aligned} \qquad [33]$$

The map from the homology of $X$ to the equivariant cohomology of $\overline{\mathcal{Q}}$ can now be constructed very easily. Let $\gamma_i$ be an element of the homology group $H_i(X)$. We associate to it the following observable:

$$\gamma_i \to I_i(\gamma_i) = \int_{\gamma_i}\mathcal{O}^{(i)} \qquad [34]$$

where $\mathcal{O}^{(i)}$ is given in [33]. The construction assures that $I_i(\gamma_i)$ is invariant under $\overline{\mathcal{Q}}$ and gauge transformations. Furthermore, it is also assured that $I_i(\gamma_i)$ is not $\overline{\mathcal{Q}}$-exact.

Let us consider the computation of correlation functions. The discussion will be presented for a generic gauge group. We will consider the topological theory defined by the Donaldson–Witten action

$$S_{\mathrm{DW}} = \{\overline{\mathcal{Q}}, V\} \qquad [35]$$

where $V$ is defined in [28]. The property [35] has a very important consequence. The action $S_{\mathrm{DW}}$ shows up in the correlation functions as $\exp(-S_{\mathrm{DW}}/e^2)$, where $e$ is a free parameter which corresponds to the coupling constant of the $\mathcal{N} = 2$ theory. Since the term involving the coupling constant is $\overline{\mathcal{Q}}$-exact, the correlation functions of $\overline{\mathcal{Q}}$-invariant operators are independent of $e$. Let us explain this in some detail. The (unnormalized) correlation functions of the theory are defined by

$$\langle\phi_1\cdots\phi_n\rangle = \int \mathcal{D}\phi\,\phi_1\cdots\phi_n\,e^{-(1/e^2)S_{\mathrm{DW}}} \qquad [36]$$

where $\phi_1, \ldots, \phi_n$ are invariant under $\overline{\mathcal{Q}}$ transformations. Using the fact that $S_{\mathrm{DW}}$ is $\overline{\mathcal{Q}}$-exact, one obtains

$$\begin{aligned}
\frac{\partial}{\partial e}\langle\phi_1\cdots\phi_n\rangle &= \frac{2}{e^3}\langle\phi_1\cdots\phi_n S_{\mathrm{DW}}\rangle \\
&= \frac{2}{e^3}\langle\{\overline{\mathcal{Q}}, \phi_1\cdots\phi_n V\}\rangle = 0
\end{aligned} \qquad [37]$$

where we have used the fact that $\overline{\mathcal{Q}}$ is a symmetry of the theory, and therefore as in [3] the last functional integral gives zero. This result implies that one can compute these correlation functions in different limits of $e$. In the weak-coupling limit (semiclassical or saddle point approximation), one establishes the connection with Donaldson theory. In the strong-coupling limit, Seiberg–Witten invariants appear and one finds the connection between these two types of invariants. We will briefly explore the weak-coupling limit $e \to 0$. The functional integral [36] can be evaluated exactly in two steps: first one analyzes the zero modes or classical configurations that minimize the action, then one expands around them considering only quadratic fluctuations. The integration over these quadratic fluctuations involves ratios of determinants of kinetic operators that because of the $\overline{\mathcal{Q}}$-symmetry of the theory (which in fact is a Bose–Fermi symmetry) are $\pm 1$. One is then left with an integral over the bosonic zero modes which leads to a finite-dimensional integral over the space of bosonic collective coordinates, and a finite Grassmannian integral over the zero modes of the fermionic fields. A careful analysis of the zero modes, first carried out by Witten, reveals that the infinite-dimensional functional integral is replaced by a finite-dimensional integral over the moduli space of anti-self-dual (ASD) connections $\mathcal{M}_{\text{ASD}}$, that is, the space of connections satisfying $F^{+}_{\mu\nu} = 0$.

Therefore, the correlation functions [36] have the form

$$\langle \phi_1 \cdots \phi_n \rangle = \int_{\mathcal{M}_{\text{ASD}}} \hat{\phi}_1 \wedge \cdots \wedge \hat{\phi}_n \qquad [38]$$

where the fields in $\phi_1 \cdots \phi_n$ are mapped to differential forms $\hat{\phi}_1 \cdots \hat{\phi}_n$ on $\mathcal{M}_{\text{ASD}}$ – the degree of each form being given by the ghost number of its partner. Notice that the integral on the right-hand side vanishes unless the form has top degree. From the field-theoretical point of view, this is the requirement that the overall ghost number of the correlation function must be equal to $\dim \mathcal{M}_{\text{ASD}}$.

The quantities on the right-hand side of [38] are – for gauge group SU(2) – precisely the Donaldson invariants. Thus, Witten's work provided a new point of view on these invariants by reformulating them in a quantum field theory language. This is a very important contribution since quantum field theory is a very rich framework and a wide variety of methods can be used to analyze the correlation functions. This opened an entirely new strategy to investigate the Donaldson invariants. The emergence of Seiberg–Witten invariants is perhaps the greatest achievement of the implementation of this strategy.

We finish this section by pointing out that many features of the evaluation of the functional integral of the Donaldson–Witten theory developed here are common to most topological field theories of the Witten type. These features can be studied in the context of the Mathai–Quillen formalism which is the object of a separate article in the encyclopedia (*see* Mathai–Quillen Formalism).

## Chern–Simons Gauge Theory for Knots and Links

Chern–Simons gauge theory is the most important example of Schwarz-type TQFTs. Let us begin by introducing its basic elements. Chern–Simons gauge theory is a quantum field theory whose action is based on the Chern–Simons form associated to a nonabelian gauge group. The theory is defined by the following data: a smooth 3-manifold $M$ which will be taken to be compact, a gauge group $G$ which will be taken semisimple and compact, and an integer parameter $k$. The action of the theory is

$$S_{\text{CS}}(A) = \frac{k}{4\pi} \int_M \operatorname{tr}\left( A \wedge \mathrm{d}A + \frac{2}{3} A \wedge A \wedge A \right) \qquad [39]$$

where $A$ is a gauge connection and the trace is taken in the fundamental representation. The exponential of i times this action is invariant under gauge transformations,

$$A \to A + g^{-1}\,\mathrm{d}g \qquad [40]$$

where $g$ is a map $g: M \to G$.

Notice that the action [39] is independent of the metric on the 3-manifold $M$. In this theory, appropriate observables lead to correlation functions which correspond to topological invariants. Candidates to be observables of this type must be metric independent and gauge invariant. Wilson loops satisfy these properties. They correspond to the holonomy of the gauge connection $A$ along a loop. Given a representation $R$ of the gauge group $G$ and a 1-cycle $\gamma$ on $M$, it is defined as

$$W^R_\gamma(A) = \operatorname{tr}_R\big(\operatorname{Hol}_\gamma(A)\big) = \operatorname{tr}_R P \exp \int_\gamma A \qquad [41]$$

Products of these operators are the natural candidates to obtain topological invariants after computing their correlation functions. These correlation functions are formally written as

$$\langle W^{R_1}_{\gamma_1} W^{R_2}_{\gamma_2} \cdots W^{R_n}_{\gamma_n} \rangle = \int [DA]\, W^{R_1}_{\gamma_1}(A)\, W^{R_2}_{\gamma_2}(A) \cdots W^{R_n}_{\gamma_n}(A)\, \mathrm{e}^{\mathrm{i}S_{\text{CS}}(A)} \qquad [42]$$

where $\gamma_1, \gamma_2, \ldots, \gamma_n$ are 1-cycles on $M$ and $R_1, R_2$, and $R_n$ are representations of $G$. In [42], the quantity $[DA]$ denotes the functional integral measure and it is assumed that an integration over connections modulo gauge transformations is carried out. As usual in quantum field theory, this integration is not well defined. Field theorists have developed methods to assign a meaning to the right-hand side of [42]. These methods mainly fall into two categories – perturbative and nonperturbative – and their degree of success mostly depends on the quantum field theory under consideration. For gauge theories, it is also possible to take an alternative approach, the large-$N$ expansion, which in general provides further insights into the theory. In Chern–Simons gauge theory all these three methods have proved of great value.

Witten (1989) showed, using nonperturbative methods, that when one considers nonintersecting cycles $\gamma_1, \gamma_2, \ldots, \gamma_n$ without self-intersections, the correlation functions [42] lead to the polynomial invariants of knot theory discovered a few years earlier starting with the work of Jones (1985).

Knot theory studies embeddings $\gamma: S^1 \to M$. Any two of such embeddings are considered equivalent if the image of one of them can be deformed into the image of the other by a homeomorphism on $M$. The main goal of knot theory is to classify the resulting equivalence classes. Each of these classes is a knot. Most of the work on knot theory has been carried out for the simple case $M = S^3$. Chern–Simons gauge theory, however, being a formulation intrinsically three dimensional, provides a framework to study the case of more general 3-manifolds $M$.

A powerful approach to classify knots is based on the construction of knot invariants. These are quantities which can be computed for a representative of a class and are invariant within the class, that is, under continuous deformations of the chosen representative. At present, it is not known if there exist enough knot invariants to classify knots. Vassiliev invariants (Vassiliev 1990) are the most promising candidates, but it is already known that if they do provide such a classification, infinitely many of them are needed.

The problem of the classification of knots in $S^3$ can be reformulated in a two-dimensional framework using regular knot projections. Given a representative of a knot in $S^3$, deform it continuously in such a way that the projection on a plane has simple crossings. Draw the projection on the plane, and at each crossing use the convention that the line that goes under the crossing is erased in a neighborhood of the crossing. The resulting diagram is a set of segments on the plane, containing the relevant information at the crossings. The problem of classifying knots is equivalent to the problem of classifying knot projections modulo a series of relations among them. These relations are known as Reidemeister moves. Invariance of a quantity under the three Reidemeister moves is called invariance under ambient isotopy. If a quantity is invariant under all but the first move, it is said to possess invariance under regular isotopy.

The formalism described for knots generalizes to the case of links. For a link of $n$ components, one considers $n$ embeddings, $\gamma_i: S^1 \to M$ $(i = 1, \ldots, n)$, with no intersections among them. Again, the main problem that link theory faces is the problem of their classification modulo homeomorphisms on $M$. In this case one can also define regular projections and reformulate the problem in terms of their classification modulo the Reidemeister moves.

The study of knot and link invariants experimented important progress in the 1980s. Jones (1985) discovered a new invariant which carries his name. The Jones polynomial can be defined very simply in terms of skein relations. These are a set of rules that can be applied to the diagram of a regular knot projection to construct the polynomial invariant. They establish a relation between the invariants associated to three links which only differ in a region as shown in Figure 1 where arrows have been introduced to take into account that the Jones polynomial is defined for oriented links.

If one denotes by $V_L(t)$ the Jones polynomial corresponding to a link $L$, $t$ being the argument of the polynomial, it must satisfy the skein relation:

$$\frac{1}{t} V_{L_+} - t V_{L_-} = \left( \sqrt{t} - \frac{1}{\sqrt{t}} \right) V_{L_0} \qquad [43]$$

where $L_+, L_-$, and $L_0$ are the links shown in Figure 1. This relation plus a choice of normalization for the unknot ($U$) are enough to compute the Jones polynomial for any link. The standard choice for the unknot is

$$V_U = 1 \qquad [44]$$

though it is not the most natural one from the point of view of Chern–Simons gauge theory. After Jones

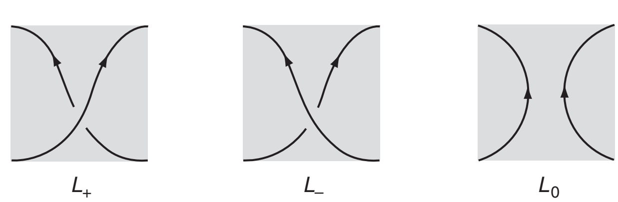



**Figure 1** Skein relations.

work in 1984, many other polynomial invariants were discovered, as the HOMFLY and the Kauffman polynomial invariants.

The pioneering work of Witten in 1988 showed that the correlation functions of products of Wilson loops [42] correspond to the Jones polynomial when one considers SU(2) as gauge group and all the Wilson loops entering in the correlation function are taken in the fundamental representation $F$. For example, if one considers a knot $K$, Witten showed that

$$V_K(t) = \langle W_K^F \rangle \qquad [45]$$

provided that one performs the identification

$$t = \exp\left(\frac{2\pi \mathrm{i}}{k+h}\right) \qquad [46]$$

where $h = 2$ is the dual Coxeter number of the gauge group SU(2). Witten also showed that if instead of SU(2) one considers SU($N$) and the Wilson loop carries the fundamental representation, the resulting invariant is the HOMFLY polynomial. The second variable of this polynomial originates in this context from the $N$ dependence. However, these cases are just a sample of the general framework intrinsic to Chern–Simons gauge theory. Taking other groups and other representations, one possesses an enormous set of knot and link invariants. These invariants can also be obtained in the context of quantum groups.

Many nonperturbative studies of Chern–Simons gauge theory have been carried out. The quantization of the theory has been studied from the point of view of the operator formalism as well as other more geometrical methods. Also, its connection to two-dimensional conformal field theory has been further elucidated, and a powerful method for the general computation of knot and link invariants has been developed by Kaul and collaborators.

Chern–Simons theory is also amenable to perturbative analysis, which has provided important representations of the Vassiliev invariants. These invariants, proposed by Vassiliev in 1990, turned out to be the coefficients of the perturbative series expansion of the correlators of Chern–Simons gauge theory. Perturbative studies can be carried out in different gauges, originating a variety of new representations of Vassiliev invariants. Among the most relevant results related to these topics are the integral expressions for Vassiliev invariants by Kontsevich and by Bott and Taubes, as well as the recent combinatorial ones. These developments are not described here but the interested reader is referred to the recent review (Labastida 1999).

## Advanced Developments

Topological sigma models are another important type of (Witten-type) TQFTs. These theories are obtained after twisting 2D $\mathcal{N} = 2$ supersymmetric sigma models. The twisting can be done in two different ways leading to two types of models, A and B. Their existence is related to mirror symmetry. Only type-A models will be described in what follows. These models can be defined on an arbitrary almost-complex manifold, though typically they are considered on Kähler manifolds. The theory involves maps from two-dimensional Riemann surfaces $\Sigma$ to target spaces $X$, together with fermionic degrees of freedom on $\Sigma$ which are mapped to tangent vectors on $X$. The functional integral of the resulting theory is localized on holomorphic maps, defining the corresponding moduli space. The corresponding $\overline{\mathcal{Q}}$-cohomology provides the set of physical observables, which can be mapped to cohomology classes on the moduli space and integrated to produce topological invariants.

Topological sigma models keep fixed the complex structure of the Riemann surface $\Sigma$. Motivated by string theory, one also considers the situation in which one integrates over complex structures. In this case, one ends up working with holomorphic maps in the entire moduli space of curves. The resulting theories are called topological strings.

We will review now a particular example of topological string theory which, besides being very interesting from the point of view of physics and mathematics, will be very useful in establishing a relation with Chern–Simons gauge theory. Let us consider topological strings with target manifold $X$ a Calabi–Yau 3-fold. In this case, the virtual dimension of the moduli space of holomorphic maps turns out to be zero. Two situations can occur: either the space is given by a number of points (the real dimension is zero) or the moduli space is finite dimensional and possesses a bundle of the same dimension as the tangent bundle. In the first case, topological strings count the number of points weighted by the exponential of the area of the holomorphic map (the pullback of the Kähler form integrated over the surface) times $x^{2g-2}$, where $x$ is the string-coupling constant and $g$ is the genus of $\Sigma$. In the second case, one computes the top Chern class of the appropriate bundles (properly defined), again weighted by the same factor. In both cases one can classify the contributions according to the cohomology class $\beta$ on $X$ in which the image of the holomorphic map is contained. The sum of the numbers obtained for each $\beta$ and fixed $g$ are known as Gromov–Witten

invariants, $N_g^\beta$. The topological string contribution takes the form

$$\sum_{g\geq 0} x^{2g-2}\left(\sum_{\beta\in H_2(X,Z)} N_g^\beta\, e^{\int_\beta \omega}\right) \qquad [47]$$

where $\omega$ is the Kähler class of the Calabi–Yau manifold. In general, the quantities $N_g^\beta$ are rational numbers.

The precedent discussion has shown how Gromov–Witten invariants can be interpreted in terms of string theory. One could think that this is just a fancy observation and that no further insight on these invariants can be gained from this formulation. The situation turns out to be quite the opposite. Once a string formulation has been obtained, the whole machinery of string theory is at our disposal. One should look to new ways to compute the quantity [47], where Gromov–Witten invariants are packed. The hope is that, if this is possible, the new emerging picture will provide new insights on these invariants. This is indeed what occurred recently. It turns out that the quantity [47] can be obtained from an alternative point of view in which the embedded Riemann surfaces are regarded as D-branes. The outcome of this approach is that the Gromov–Witten invariants can be written in terms of other invariants which are integers and that possess a geometrical interpretation. To be more specific, the quantity [47] takes the form

$$\sum_{\substack{g\geq 0\\ \beta\in H_2(X,Z)}} \sum_{d>0} n_g^\beta \frac{1}{d}\left(2\sin\left(\frac{dx}{2}\right)\right)^{2g-2} e^{d\int_\beta \omega} \qquad [48]$$

where $n_g^\beta$ are the new "integer" invariants. This prediction has been verified in all the cases in which it has been tested. A similar structure will be found in the next section in the context of knot theory in the large-$N$ limit.

Let us now consider also Donaldson–Witten theory from a new perspective. To be more specific, let us consider the case in which the gauge group is SU(2), and the 4-manifold $X$ is simply connected and has $b_2^+ > 1$ (the case $b_2^+ = 1$ is anomalous). In this situation there are $1 + b_2$ physical observables [34], $\mathcal{O} = I_1$ and $I(\Sigma_a) = I_2(\Sigma_a)$ $(a = 1, \ldots, b_2)$, where $\Sigma_a$ is a basis of $H_2(X)$. These can be packed in a generating functional:

$$\left\langle \exp\left(\sum_a \alpha_a I(\Sigma_a) + \lambda\mathcal{O}\right)\right\rangle \qquad [49]$$

where $\lambda$ and $\alpha_a\,(a = 1, \ldots, b_2)$ are parameters. In computing this quantity one can argue that the contribution is localized on the moduli space of instantons configurations and one ends up, after taking into account the selection rule dictated by the dimensionality of the moduli space, with integrations over the moduli space of the selected forms. The resulting quantities are Donaldson invariants.

As in the case of topological sigma models one could be tempted to argue that the observation leading to a field-theoretical interpretation of Donaldson invariants does not provide any new insight. Quite on the contrary, once a field theory formulation is available, one has at his disposal a huge machinery which could lead, on the one hand, to further generalizations of the theory and, on the other hand, to new ways to compute quantities such as [49], obtaining new insights on these invariants. This is indeed what happened in the 1990s, leading to an important breakthrough in 1994 when Seiberg and Witten calculated [49] in a different way and pointed out the relation of Donaldson invariants to new integer invariants that nowadays bear their names.

The localization argument that led to the interpretation of [49] as Donaldson invariants is valid because the theory under consideration is exact in the weak-coupling limit. Actually, the topological theory under consideration is independent of the coupling constant and thus calculations in the strong-coupling limit are also exact. These types of calculations were out of reach before 1994. The situation changed dramatically after the work of Seiberg and Witten in which $\mathcal{N} = 2$ super Yang–Mills theory was solved in the strong-coupling limit. Its application to the corresponding twisted version was immediate and it turned out that Donaldson invariants can be written in terms of new integer invariants now known as Seiberg–Witten invariants (Witten 1994). The development has a strong resemblance with the one described above for topological strings: certain noninteger invariants can be expressed in terms of new integer invariants.

The Seiberg–Witten invariants are actually simpler to compute than Donaldson invariants. They correspond to partition functions of topological Yang–Mills theories where the gauge group is abelian. These contributions can be grouped into classes labeled by $x = -2c_1(L)$, where $c_1(L)$ is the first Chern class of the corresponding line bundle. The sum of contributions, each being $\pm 1$, for a given class $x$ is the integer Seiberg–Witten invariant $n_x$. The strong-coupling analysis of topological Yang–Mills theory leads to the following expression for [49]:

$$2^{1+(1/4)(7\chi+11\sigma)}\Big(e^{((v^2/2)+2\lambda)}\sum_x n_x e^{v\cdot x} + i^{\chi+\sigma/4}\, e^{(-(v^2/2)-2\lambda)}\sum_x n_x\, e^{-iv\cdot x}\Big) \qquad [50]$$

where $v = \sum_a \alpha_a \Sigma_a$, and $\chi$ and $\sigma$ are the Euler number and the signature of the manifold $X$. This result matches the known structure of [49] (structure theorem of Kronheimer and Mrowka) and provides

a meaning to its unknown quantities in terms of the new Seiberg–Witten invariants. Equation [50] is a rather remarkable prediction that has been tested in many cases, and for which a general proof has been recently proposed. For a review of the subject, see Labastida and Lozano (1998).

The situation for manifolds with $b_2^+=1$ involves a metric dependence and has been worked out in detail (Moore and Witten 1998). The formulation of Donaldson invariants in field-theoretical terms has also provided a generalization of these invariants. This generalization has been carried out in several directions: (1) the consideration of higher-rank groups, (2) the coupling to matter fields after twisting $\mathcal{N}=2$ hypermultiplets, and (3) the twist of theories involving $\mathcal{N}=4$ supersymmetry.

We will now look at Chern–Simons gauge theory from the perspective that emerges from its treatment in the context of the large-$N$ expansion. We will restrict the discussion to the case of knots on $S^3$ with gauge group SU($N$). Gauge theories with gauge group SU($N$) admit, besides the perturbative expansion, a large-$N$ expansion. In this expansion correlators are expanded in powers of $1/N$ while keeping the 't Hooft coupling $t=Nx$ fixed, $x$ being the coupling constant of the gauge theory. For example, for the free energy of the theory, one has the general form

$$F=\sum_{\substack{g\geq 0\\ h\geq 1}}^{\infty} C_{g,h}N^{2-2g}t^{2g-2+h} \qquad [51]$$

In the case of Chern–Simons gauge theory, the coupling constant is $x=2\pi i/(k+N)$ after taking into account the shift in $k$. The large-$N$ expansion [51] resembles a string-theory expansion and indeed the quantities $C_{g,h}$ can be identified with the partition function of a topological open string with $g$ handles and $h$ boundaries, with $N$ D-branes on $S^3$ in an ambient six-dimensional target space $T^*S^3$. This was pointed out by Witten in 1992. The result makes a connection between a topological three-dimensional field theory and the topological strings described in the previous section.

In 1998 an important breakthrough took place which provided a new approach to compute quantities such as [51]. Using arguments inspired by the AdS/CFT correspondence, Gopakumar and Vafa (1999) provided a closed-string-theory interpretation of the partition function [51]. They conjectured that the free energy $F$ can be expressed as

$$F=\sum_{g\geq 0}^{\infty} N^{2-2g}F_g(t) \qquad [52]$$

where $F_g(t)$ corresponds to the partition function of a topological closed-string theory on the noncompact Calabi–Yau manifold $X$ called the resolved conifold, $\mathcal{O}(-1)\oplus\mathcal{O}(-1)\to \boldsymbol{P}^1$, $t$ being the flux of the $B$-field through $\boldsymbol{P}^1$. The quantities $F_g(t)$ have been computed using both physical and mathematical arguments, thus proving the conjecture.

Once a new picture for the partition function of Chern–Simons gauge theory is available, one should ask about the form that the expectation values of Wilson loops could take in the new context. The question was faced by Ooguri and Vafa and they provided the answer, later refined by Labastida, Mariño, and Vafa. The outcome is an entirely new point of view in the theory of knot and link invariants. The new picture provides a geometrical interpretation of the integer coefficients of the quantum group invariants, an issue that has been investigated during many years. To present an account of these developments, one needs to review first some basic facts of large-$N$ expansions.

To consider the presence of Wilson loops, it is convenient to introduce a particular generating functional. First, one performs a change of basis from representations $R$ to conjugacy classes $C(\boldsymbol{k})$ of the symmetric group, labeled by vectors $\boldsymbol{k}=(k_1,k_2,\ldots)$ with $k_i\geq 0$, and $|\boldsymbol{k}|=\sum_j k_j>0$. The change of basis is $W_{\boldsymbol{k}}=\sum_R \chi_R(C(\boldsymbol{k}))W_R$, where $\chi_R$ are characters of the permutation group $S_\ell$ of $\ell=\sum_j jk_j$ elements ($\ell$ is also the number of boxes of the Young tableau associated to $R$). Second, one introduces the generating functional:

$$F(V)=\log Z(V)=\sum_{\boldsymbol{k}}\frac{|C(\boldsymbol{k})|}{\ell!}W_{\boldsymbol{k}}^{(c)}\Upsilon_{\boldsymbol{k}}(V) \qquad [53]$$

where

$$Z(V)=\sum_{\boldsymbol{k}}\frac{|C(\boldsymbol{k})|}{\ell!}W_{\boldsymbol{k}}\Upsilon_{\boldsymbol{k}}(V)$$

$$\Upsilon_{\boldsymbol{k}}(V)=\prod_j(\mathrm{tr}\,V^j)^{k_j}$$

In these expressions $|C(\boldsymbol{k})|$ denotes the number of elements of the class $C(\boldsymbol{k})$ in $S_\ell$. The reason behind the introduction of this generating functional is that the large-$N$ structure of the connected Wilson loops, $W_{\boldsymbol{k}}^{(c)}$, turns out to be very simple:

$$\frac{|C(\boldsymbol{k})|}{\ell!}W_{\boldsymbol{k}}^{(c)}=\sum_{g=0}^{\infty}x^{2g-2+|\boldsymbol{k}|}F_{g,\boldsymbol{k}}(\lambda) \qquad [54]$$

where $\lambda=e^t$ and $t=Nx$ is the 't Hooft coupling. Writing $x=t/N$, it corresponds to a power series expansion in $1/N$. As before, the expansion looks like a perturbative series in string theory where $g$ is the genus and $|\boldsymbol{k}|$ is the number of holes. Ooguri and Vafa conjectured in 1999 the appropriate string-theory description of [54]. It corresponds to an open topological string theory (notice that the ones

described in the previous section were closed), whose target space is the resolved conifold $X$. The contribution from this theory will lead to open-string analogs of Gromov–Witten invariants.

In order to describe in more detail the fact that one is dealing with open strings, some new data need to be introduced. Here is where the knot description intrinsic to the Wilson loop enters. Given a knot $K$ on $S^3$, let us associate to it a Lagrangian submanifold $C_K$ with $b_1 = 1$ in the resolved conifold $X$ and consider a topological open string on it. The contributions in this open topological string are localized on holomorphic maps $f : \Sigma_{g,h} \to X$ with $h = |\boldsymbol{k}|$ which satisfy $f_*[\Sigma_{g,h}] = \mathcal{Q}$, and $f_*[C] = j[\gamma]$ for $k_j$ oriented circles $C$. In these expressions $\gamma \in H_1(C_K, \boldsymbol{Z})$, and $\mathcal{Q} \in H_2(X, C_K, \boldsymbol{Z})$, that is, the map is such that $k_j$ boundaries of $\Sigma_{g,h}$ wrap the knot $j$ times, and $\Sigma_{g,h}$ itself gets mapped to a relative two-homology class characterized by the Lagrangian submanifold $C_K$. The number of such maps (in the sense described in the previous section) is the open-string analog of Gromov–Witten invariants. They will be denoted by $N^{\mathcal{Q}}_{g,\boldsymbol{k}}$. Comparing to the situation that led to [47] in the closed-string case, one concludes that in this case the quantities $F_{g,\boldsymbol{k}}(\lambda)$ in [54] must take the form

$$F_{g,\boldsymbol{k}}(\lambda) = \sum_{\mathcal{Q}} N^{\mathcal{Q}}_{g,\boldsymbol{k}}\, \mathrm{e}^{\int_{\mathcal{Q}} \omega}, \qquad t = \int_{\boldsymbol{P}^1} \omega \qquad [55]$$

where $\omega$ is the Kähler class of the Calabi–Yau manifold $X$ and $\lambda = \mathrm{e}^t$. For any $\mathcal{Q}$, one can always write $\int_{\mathcal{Q}} \omega = Qt$, where $Q$ is in general a half-integer number. Therefore, $F_{g,\boldsymbol{k}}(\lambda)$ is a polynomial in $\lambda^{\pm 1/2}$ with rational coefficients.

The result [55] is very impressive but still does not provide a representation where one can assign a geometrical interpretation to the integer coefficients of the quantum-group invariants. Notice that to match a polynomial invariant to [55], after obtaining its connected part, one must expand it in $x$ after setting $q = \mathrm{e}^x$ keeping $\lambda$ fixed. One would like to have a refined version of [55], in the spirit of what was described in the previous section leading from the Gromov–Witten invariants $N^{\beta}_g$ of [47] to the new integer invariants $n^{\beta}_g$ of [48]. It turns out that, indeed, $F(V)$ can be expressed in terms of integer invariants in complete analogy with the description presented in the previous section for topological strings. A good review on the subject can be found in Mariño (2005).

## Concluding Remarks

In this overview we have introduced key features of TQFTs and we have described some of the most relevant results emerged from them. We have described how the many faces of TQFT provide a variety of important insights in a selected set of problems in topology. Among these outstand the reformulation of Donaldson theory and the discovery of the Seiberg–Witten invariants, and the string-theory description of the large-$N$ expansion of Chern–Simons gauge theory, which provides an entirely new point of view in the study of knot and link invariants and points to an underlying fascinating interplay between string theory, knot theory, and enumerative geometry which opens new fields of study.

In addition to their intrinsic mathematical interest, TQFTs have been found relevant to important questions in physics as well. This is so because, in a sense, TQFTs are easier to solve than conventional quantum field theories. For example, topological sigma models are relevant to the computation of certain couplings in string theory. Also, Witten-type gauge TQFTs such as Donaldson–Witten theories and its generalizations play a role in string theory as effective world-volume theories of extended string states (branes) wrapping curved spaces, and TQFTs arising from $\mathcal{N} = 4$ gauge theories in four dimensions have shed light on field- (and string-) theory dualities.

Most of these developments, and others that we have not touched upon or only mentioned in passing have their own entries in the encyclopedia, to which we refer the interested reader for further details.

*See also:* Axiomatic Approach to Topological Quantum Field Theory; BF Theories; Chern–Simons Models: Rigorous Results; Donaldson–Witten Theory; Gauge Theoretic Invariants of 4-Manifolds; Gauge Theory: Mathematical Applications; Hamiltonian Fluid Dynamics; The Jones Polynomial; Knot Theory and Physics; Mathai–Quillen Formalism; Mathematical Knot Theory; Schwarz-Type Topological Quantum Field Theory; Seiberg–Witten Theory; Stationary Phase Approximation; Topological Sigma Models.

## Further Reading


Atiyah MF (1988) New invariants of three and four dimensional manifolds. In: *The Mathematical Heritage of Herman Weyl*, Proc. Symp. Pure Math., vol. 48. American Math. Soc. pp. 285–299.

Birmingham D, Blau M, Rakowski M, and Thompson G (1991) Topological field theory. *Physics Reports* 209: 129–340.

Cordes S, Moore G, and Rangoolam S (1996) Lectures on 2D Yang–Mills theory, equivariant cohomology and topological field theories. In: David F, Ginsparg P, and Zinn-Justin J (eds.) *Fluctuating Geometries in Statistical Mechanics and Field Theory*, Les Houches Sesion LXII, p. 505 (hep-th/9411210). Elsevier.

Donaldson SK (1990) Polynomial invariants for smooth four-manifolds. *Topology* 29: 257–315.

Floer A (1987) Morse theory for fixed points of symplectic diffeomorphisms. *Bulletin of the American Mathematical Society* 16: 279.

Freyd P, Yetter D, Hoste J, Lickorish WBR, Millet K, and Ocneanu A (1985) A new polynomial invariant of knots and links. *Bulletin of the American Mathematical Society* 12(2): 239–246.
Gopakumar R and Vafa C (1999) On the gauge theory/geometry correspondence. *Advances in Theoretical and Mathematical Physics* 3: 1415 (hep-th/9811131).
Jones VFR (1985) A polynomial invariant for knots via von Neumann algebras. *Bulletin of the American Mathematical Society* 12: 103–112.
Jones VFR (1987) Hecke algebra representations of braid groups and link polynomials. *Annals of Mathematics* 126(2): 335–388.
Kauffman LH (1990) An invariant of regular isotopy. *Transactions of American Mathematical Society* 318(2): 417–471.
Labastida JMF (1999) Chern-Simons gauge theory: ten years after. In: Falomir H, Gamboa R, and Schaposnik F (eds.) *Trends in Theoretical Physics II*, ch. 484, 1. New York: AIP (hep-th/9905057).
Labastida JMF and Lozano C (1998) Lectures in topological quantum field theory. In: Falomir H, Gamboa R, and Schaposnik F (eds.) *Trends in Theoretical Physics*, ch. 419, 54. New York: AIP (hep-th/9709192).
Labastida JMF and Mariño M (2005) *Topological Quantum Field Theory and Four Manifolds*. Dordrecht: Elsevier; Norwell, MA: Springer.
Mariño M (2005) Chern–Simons theory and topological strings. *Reviews of Modern Physics* 77: 675–720.
Moore G and Witten E (1998) Integrating over the *u*-plane in Donaldson theory. *Advances in Theoretical and Mathematic Physics* 1: 298–387.
Vassiliev VA (1990) Cohomology of knot spaces. In: *Theory of Singularities and Its Applications*, Advances in Soviet Mathematics, vol. 1, pp. 23–69. American Mathematical Society.
Witten E (1982) Supersymmetry and Morse theory. *Journal of Differential Geometry* 17: 661–692.
Witten E (1988a) Topological quantum field theory. *Communications in Mathematical Physics* 117: 353.
Witten E (1988b) Topological sigma models. *Communications in Mathematical Physics* 118: 411.
Witten E (1989) Quantum field theory and the Jones polynomial. *Communications in Mathematical Physics* 121: 351.
Witten E (1994) Monopoles and four-manifolds. *Mathematical Research Letters* 1: 769–796.

# Topological Sigma Models

**D Birmingham**, University of the Pacific, Stockton, CA, USA



## Introduction

Topological sigma models govern the quantum mechanics of maps from a Riemann surface $\Sigma$ to a target space $M$. In contrast to the standard supersymmetric sigma model, the topological version has a special local shift symmetry. This symmetry takes the form $\delta u^i = \epsilon^i$, where $\epsilon^i$ is an arbitrary local function of the coordinates on the base manifold $\Sigma$. In essence, this topological shift symmetry ensures that all local degrees of freedom of the model can be gauged away. As a result, the dynamics of such a model resides in a finite number of global topological degrees of freedom. This feature is generic to all topological field theories of Witten type, also known as cohomological field theories (*see* Topological Quantum Field Theory: Overview). The topological shift symmetry is responsible for the special topological nature of the model, which is seen most readily by BRST quantizing the local shift symmetry. This gives rise to a nilpotent BRST operator $Q$. The properties of this BRST operator are crucial for establishing the topological nature of the model. The key point in the construction of any cohomological field theory is the fact that the full quantum action $S_{\mathrm{q}}$ can be written as a BRST commutator $S_{\mathrm{q}} = \{Q, V\}$, where $V$ is a function of the fields needed to define the path integral. In particular, one can show that the partition function and all correlation functions are independent of the metric on both the base manifold $\Sigma$ and the target space $M$. For example, let us define the path integral by

$$Z = \int \mathrm{d}\Phi \, \mathrm{e}^{-\{Q,V\}} \qquad [1]$$

where $\Phi$ denotes the full set of fields required at the quantum level. In general, the function $V$ depends on geometric data of both $\Sigma$ and $M$. Nevertheless, one can easily establish that the partition function is independent of this data by noting the following. Variation of $Z$ with respect to the metric of the target space $g$ (for example) gives

$$\delta_g Z = -\int \mathrm{d}\Phi \, \mathrm{e}^{-\{Q,V\}} \, \{Q, \delta_g V\} \qquad [2]$$

The right-hand side of this equation is nothing but the vacuum expectation value of a BRST commutator, and this vanishes by BRST invariance of the vacuum. It is important to note here that the BRST operator $Q$ can be constructed to be independent of $g$. Apart from the necessity of introducing the metric tensor, these models also require additional geometric data for their construction. The complex structure of $\Sigma$, and at least an almost-complex structure on $M$, is required. By a similar argument, one can show that the partition function and correlation functions are independent of this extra

geometric data. As mentioned above, these models possess no local degrees of freedom. One can then show that the path-integral expression for the correlation functions can be localized to a finite-dimensional moduli space of instanton configurations which minimize the classical action.

We will first show how the full quantum action of the theory can be obtained as a BRST quantization of a classical action with a local gauge symmetry. However, we shall then highlight the fact that the gauge algebra for this topological shift symmetry only closes on-shell. In order to proceed with a BRST quantization of the model, and obtain the complete quantum action, one must take recourse to the Batalin–Vilkovisky quantization scheme. This machinery is ideally tailored for such a problem, with the end result that quartic ghost terms are present in the action. However, the presence of such terms does not affect the arguments presented above, since the quantum is still obtained as a BRST commutator. Following this, we construct all observables of the theory and demonstrate their connection to the de Rham cohomology of the target space. The special topological properties of the observables are then discussed, and it is shown how their computation is localized to the moduli space $\mathcal{M}$ of holomorphic maps from $\Sigma$ to $M$. As a particular example, we show how the computation of a certain class of observables determines the intersection numbers of the moduli space $\mathcal{M}$. We present a brief discussion of the connection between topological sigma models with Calabi–Yau target space $M$, and the mirror symmetry of $M$.

## Construction of the Model

We begin with the following classical action:

$$S_{\mathrm{c}} = \int_{\Sigma} \mathrm{d}^2\sigma \sqrt{h}\, h_{\alpha\beta}\, g_{ij} K^{\alpha i} K^{\beta j} \qquad [3]$$

where

$$K^{\alpha i} = G^{\alpha i} - \tfrac{1}{2}\left(\partial^\alpha u^i + \epsilon^\alpha{}_\beta J^i{}_j \partial^\beta u^j\right) \qquad [4]$$

The fields $G^{\alpha i}$ and $K^{\alpha i}$ both satisfy the self-duality constraint

$$\begin{aligned} G^{\alpha i} &= P^{\alpha i}_{+\ \beta j} G^{\beta j} \\ K^{\alpha i} &= P^{\alpha i}_{+\ \beta j} K^{\beta j} \end{aligned} \qquad [5]$$

where the self-dual and anti-self-dual projection operators are defined as

$$P^{\alpha i}_{\pm\ \beta_j} = \tfrac{1}{2}\left(\delta^\alpha{}_\beta \delta^i{}_j \pm \epsilon^\alpha{}_\beta J^i{}_j\right) \qquad [6]$$

The above action describes a theory of maps $u^i(\sigma)$ from a Riemann surface $\Sigma$ to an almost complex manifold $M$. The coordinates on $\Sigma$ are denoted by $\sigma^\alpha (\alpha = 1, 2)$, while those on the target manifold $M$ are denoted by $u^i (i = 1, \ldots, \dim M)$. The metric and complex structure of $\Sigma$ are denoted by $h_{\alpha\beta}$ and $\epsilon^\alpha{}_\beta$, respectively; they obey the relations $\epsilon^\alpha{}_\beta \epsilon^\beta{}_\gamma = -\delta^\alpha{}_\gamma$ and $\epsilon_{\alpha\beta} = h_{\alpha\gamma}\epsilon^\gamma{}_\beta$. The metric tensor $g_{ij}$ and almost-complex structure $J^i{}_j$ of $M$ obey analogous relations to the above. In the general model, the target space need only be an almost-complex manifold. This requires the existence of a globally defined tensor field $J^i{}_j$ such that $J^i{}_j J^j{}_k = -\delta^i{}_k$.

The action [3] is invariant under the topological shift symmetry

$$\delta u^i = \epsilon^i \qquad [7]$$

where $\epsilon^i$ is an arbitrary local function of the coordinates on the base manifold $\Sigma$. Already, at this level, we see the distinction with the standard sigma model. The presence of this shift symmetry means that all local degrees of freedom can be gauged away, leaving only a finite number of global topological degrees of freedom. It requires some work to determine the corresponding transformation for $G^{\alpha i}$, the key point being the preservation of the self-duality constraint. We find

$$\begin{aligned} \delta G^{\alpha i} = {} & P^{\alpha i}_{+\ \beta j}\left(D^\beta \epsilon^j + \tfrac{1}{2}\epsilon^\beta{}_\gamma \epsilon^l \left(D_l J^j{}_k\right)\partial^\gamma u^k\right) \\ & + \tfrac{1}{2}\epsilon^\alpha{}_\beta \epsilon^k \left(D_k J^i{}_j\right) G^{\beta j} - \Gamma^i_{lk}\epsilon^k G^{\alpha l} \end{aligned} \qquad [8]$$

where the covariant derivative is defined by $D_\alpha \epsilon^i = \partial_\alpha \epsilon^i + \Gamma^i_{jk}(\partial_\alpha u^j)\epsilon^k$.

Having determined the classical symmetries of the model, we can now proceed with the BRST quantized form of the quantum action. As a topological field theory of Witten type, one can show that the quantum action can be written as a BRST commutator, that is, $S_{\mathrm{q}} = \{Q, V\}$, where the gauge fermion $V$ is defined by

$$V = \int \mathrm{d}^2\sigma \sqrt{h}\, \bar{C}_{\alpha i}\left(\partial^\alpha u^i - \frac{\alpha}{4} B^{\alpha i}\right) \qquad [9]$$

where $\alpha$ is an arbitrary gauge-fixing parameter. The BRST operator $Q$ is nilpotent $Q^2 = 0$, off-shell. It is defined by $\delta\Phi = \epsilon\{Q, \Phi\}$, and takes the form

$$\begin{aligned} \delta u^i &= -\epsilon C^i \\ \delta C^i &= 0 \\ \delta \bar{C}_{\alpha i} &= \epsilon\left(B_{\alpha i} + \frac{1}{2}\epsilon_\alpha{}^\beta \left(D_k J_i{}^j\right)\bar{C}_{\beta j} C^k + \Gamma^k_{ij}\bar{C}_{\alpha k} C^j\right) \\ \delta B^{\alpha i} &= \frac{\epsilon}{4} C^k C^l \left(R^{i_t}_{kl} + R_{klrs} J^{ri} J^s{}_t\right)\bar{C}^{\alpha t} \\ & \quad - \frac{\epsilon}{2}\epsilon^\alpha{}_\beta \left(D_k J^i{}_j\right) C^k B^{\beta j} \\ & \quad + \frac{\epsilon}{4}\left(C^k D_k J^i{}_s\right)\left(C^l D_l J_t{}^s\right)\bar{C}^{\alpha t} + \epsilon \Gamma^i_{jk} C^j B^{\alpha k} \end{aligned} \qquad [10]$$

In the above, the ghost field is denoted by $C^i$, while the anti-ghost field $\bar{C}_{\alpha i}$ and the multiplier field $B_{\alpha i}$ obey the self-duality constraint [5]. The key point to note in the above transformations is the fact that the ghost field $C^i$ is BRST invariant. Again, this is a feature which is generic to all cohomological field theories. The existence of such a field allows the construction of an entire set of topological correlation functions, as we shall see in the following section.

While the gauge-fixing parameter $\alpha$ is arbitrary, a conventional choice is to take $\alpha = 1$, and then integrate out the multiplier field $B$. This yields the action in the form

$$\begin{aligned} S_q = \int d^2\sigma\sqrt{h}\Big[&\frac{1}{2}h^{\alpha\beta}g_{ij}\partial_\alpha u^i\partial_\beta u^j + \frac{1}{2}\epsilon^{\alpha\beta}J_{ij}\partial_\alpha u^i\partial_\beta u^j \\ &+ \bar{C}_{\alpha i}\left(D^\alpha C^i + \frac{1}{2}\epsilon^\alpha{}_\beta(D_j J^i{}_k)\partial^\beta u^k C^j\right) \\ &+ \frac{1}{8}\bar{C}_\alpha{}^m\bar{C}^{\alpha k}R_{mkjr}C^jC^r \\ &+ \frac{1}{16}\bar{C}_{\alpha i}\bar{C}^{\alpha k}(D_j J^{li})(D_r J_{lk})C^jC^r\Big] \end{aligned} \qquad [11]$$

It should be stressed that the classical gauge algebra [7] and [8] only closes on-shell. Quantization of the model is therefore more subtle, and requires use of the Batalin–Vilkovisky formalism. The on-shell closure problem automatically results in the presence of quartic ghost coupling terms in the action and consequently cubic terms in the BRST transformations. Despite this, we have established that the full quantum action can be written as a BRST commutator.

The form of the action simplifies when the complex structure of the target manifold is covariantly constant, $D_k J^i{}_j = 0$. In this case, the target manifold $M$ is Kähler and we denote the complex coordinates as $u^I$, with their complex conjugates denoted by $u^{\bar{I}}$. The nonzero components of the metric tensor are then $g_{I\bar{J}}$. Similarly, the coordinates of $\Sigma$ are denoted $\sigma^\pm$, with nonzero metric components $h_{+-}$. The nonzero components of the ghost and anti-ghost are then given by $C^I, C^{\bar{I}}, \bar{C}_{+\bar{I}}, \bar{C}_{-I}$. The action can be written in the form

$$\begin{aligned} S_q = \int d^2\sigma\sqrt{h}\Big[&h^{+-}g_{I\bar{J}}\partial_+ u^I\partial_- u^{\bar{J}} + \frac{1}{2}\bar{C}_+{}^I(D_- C^{\bar{J}})h^{+-}g_{I\bar{J}} \\ &+ \frac{1}{2}\bar{C}_-{}^{\bar{I}}(D_+ C^J)h^{+-}g_{I\bar{J}} \\ &+ \frac{1}{4}h^{+-}\bar{C}_+{}^I\bar{C}_-{}^{\bar{I}}R_{I\bar{I}J\bar{J}}C^JC^{\bar{J}}\Big] \end{aligned} \qquad [12]$$

## Construction of Observables

Having defined the quantum action, it is now of interest to consider the correlation functions of the model. In the functional integral, we integrate over all maps $\Sigma \to M$ in a fixed homotopy class. Let us consider a correlation function

$$\langle\mathcal{O}\rangle = \int du^i\, d\bar{C}_{\alpha i}\, dC^i\, e^{-tS_q}\mathcal{O} \qquad [13]$$

where $t > 0$ is a parameter, and the observable $\mathcal{O}$ is BRST invariant $\{Q, \mathcal{O}\} = 0$. From the BRST invariance of the vacuum, it follows immediately that the vacuum expectation value of a BRST commutator is zero, $\langle\{Q, \mathcal{O}\}\rangle = 0$. An operator which is a BRST commutator is said to be $Q$-exact. Hence, our interest is in the $Q$-cohomology classes of operators, that is, BRST invariant operators modulo BRST exact operators. It is for this reason that such a model is called a cohomological field theory.

One can now show that the variation of [13] with respect to $t$ is a BRST commutator, namely

$$\delta_t\langle\mathcal{O}\rangle = -\delta t\int du^i\, d\bar{C}_{\alpha i}\, dC^i\, e^{-tS_q}\{Q, V\mathcal{O}\} = 0 \qquad [14]$$

As a result, one can evaluate the correlation function in the large-$t$ (weak-coupling) limit. In this limit, the path integral is dominated by fluctuations around the classical minima. For the sigma model under study, the classical action is minimized by the instanton configurations

$$\partial_\alpha u^i + \epsilon_\alpha{}^\beta J^i{}_j\partial_\beta u^i = 0 \qquad [15]$$

Indeed, this localization of the path integral to the moduli space of instantons can also be seen by choosing the $\alpha = 0$ gauge in [9]. Integration over the multiplier field then imposes a delta function constraint to the instanton configurations. The key point in the above derivation is the fact that the quantum action is a BRST commutator, $S_q = \{Q, V\}$. By a similar argument, one can show that variations of $\langle\mathcal{O}\rangle$ with respect to the metric and complex structure of $\Sigma$ and $M$ are also zero.

Our aim now is to construct the $Q$-cohomology classes of operators in the theory. Let us first associate an operator $\mathcal{O}_A^{(0)}$ to each $p$-form $A = A_{i_1\cdots i_p}\, du^{i_1}\wedge\cdots\wedge du^{i_p}$ on the target space $M$, given by

$$\mathcal{O}_A^{(0)} = A_{i_1\cdots i_p}C^{i_1}\cdots C^{i_p} \qquad [16]$$

where $C^i$ is the ghost field. Under a BRST transformation, we see that

$$\begin{aligned} \{Q, \mathcal{O}_A^{(0)}\} &= -\partial_{i_0}A_{i_1\cdots i_p}C^{i_0}\cdots C^{i_p} \\ &= -\mathcal{O}_{dA}^{(0)} \end{aligned} \qquad [17]$$

since the ghost fields are BRST invariant by [10]. Hence, $\mathcal{O}_A^{(0)}$ is BRST invariant if and only if $A$ is a closed $p$-form. Similarly, if $A$ is an exact $p$-form, then the corresponding operator is $Q$-exact. Hence, the BRST cohomology classes of these operators are in one to one correspondence with the de Rham cohomology classes on $M$. The reason for assigning the peculiar superscript to the operator $\mathcal{O}^{(0)}$ will become clear at the end of this construction. Notice also that operators of the form $\mathcal{O}_A^{(0)}$ can be used as building blocks for constructing new observables. If we consider a set of closed forms $A_1, \ldots, A_k$, then the product of the associated operators $\mathcal{O}_{A_1}^{(0)} \cdots \mathcal{O}_{A_k}^{(0)}$ is clearly $Q$-invariant as well.

When considering the vacuum expectation values of operators which are polynomials in the fields, there is an implicit dependence on the points where the operators are located. In the case at hand however, the operator $\mathcal{O}_A^{(0)}(\sigma)$ at the point $\sigma$ has a vacuum expectation value which is a topological invariant, and thus cannot depend on the chosen point. To see this explicitly, we consider all fields defined over $\Sigma$, and differentiate the operator with respect to some local coordinates $\sigma^\alpha$:

$$\frac{\partial}{\partial\sigma^\alpha} A_{i_1\cdots i_p} C^{i_1}\cdots C^{i_p} = (\partial_{i_0} A_{i_1\cdots i_p}) \frac{\partial u^{i_0}}{\partial\sigma^\alpha} C^{i_1}\cdots C^{i_p} + p A_{i_1\cdots i_p} (\partial_{i_0} C^{i_1}) \frac{\partial u^{i_0}}{\partial\sigma^\alpha} C^{i_2}\cdots C^{i_p} \qquad [18]$$

In terms of exterior derivatives, this takes the form,

$$\begin{aligned} d\mathcal{O}_A^{(0)} &= \partial_{i_0} A_{i_1\cdots i_p} du^{i_0} C^{i_1}\cdots C^{i_p} + p A_{i_1\cdots i_p} dC^{i_1} C^{i_2}\cdots C^{i_p} \\ &= \{Q, \mathcal{O}_A^{(1)}\} \end{aligned} \qquad [19]$$

where $\mathcal{O}_A^{(1)} = -p A_{i_1\cdots i_p} du^{i_1} C^{i_2}\cdots C^{i_p}$, and we have used the fact that $A$ is a closed $p$-form. If we let $\gamma$ represent any path between two arbitrary points $P$ and $P'$, then this expression has the integral form,

$$\mathcal{O}_A^{(0)}(P) - \mathcal{O}_A^{(0)}(P') = \left\{Q, \int_\gamma \mathcal{O}_A^{(1)}\right\} \qquad [20]$$

and we see that the vacuum expectation value of $\mathcal{O}_A^{(0)}$ is point independent by the BRST invariance of the vacuum. The same remark applies to any product of operators of the form we are considering.

To continue our construction, consider a one-dimensional homology cycle $\gamma(\partial\gamma = 0)$, and define

$$W_A^{(1)}(\gamma) = \int_\gamma \mathcal{O}_A^{(1)} \qquad [21]$$

This new operator $W_A^{(1)}(\gamma)$ is BRST invariant by inspection,

$$\left\{Q, W_A^{(1)}(\gamma)\right\} = \int_\gamma \left\{Q, \mathcal{O}_A^{(1)}\right\} = \int_\gamma d\mathcal{O}_A^{(0)} = 0 \qquad [22]$$

Moreover, if $\gamma$ happens to be the boundary of a two-dimensional surface $(\gamma = \partial\beta)$, so that $\gamma$ is trivial in homology, then this new operator is likewise trivial in $Q$ cohomology:

$$W_A^{(1)}(\gamma) = \int_\gamma \mathcal{O}_A^{(1)} = \int_\beta d\mathcal{O}_A^{(1)} = \left\{Q, \int_\beta \mathcal{O}_A^{(2)}\right\} \qquad [23]$$

where

$$\mathcal{O}_A^{(2)} = -\frac{p(p-1)}{2} A_{i_1\cdots i_p} du^{i_1} \wedge du^{i_2} C^{i_3}\cdots C^{i_p}$$

As before, let us now associate to each homology 2-cycle $\beta(\partial\beta = 0)$, another BRST invariant operator $W_A^{(2)}$ defined by

$$W_A^{(2)}(\beta) = \int_\beta \mathcal{O}_A^{(2)} \qquad [24]$$

The BRST invariance follows trivially as in [23].

In summary, we have produced three operators $\mathcal{O}_A^{(0)}, \mathcal{O}_A^{(1)}$, and $\mathcal{O}_A^{(2)}$ from any given closed form $A$, which satisfy the relations:

$$\begin{aligned} 0 &= \left\{Q, \mathcal{O}_A^{(0)}\right\}, & d\mathcal{O}_A^{(0)} &= \left\{Q, \mathcal{O}_A^{(1)}\right\} \\ d\mathcal{O}_A^{(1)} &= \left\{Q, \mathcal{O}_A^{(2)}\right\}, & d\mathcal{O}_A^{(2)} &= 0 \end{aligned} \qquad [25]$$

The BRST observables are then given by arbitrary products of the integrated operators $W_A^{(i)}(\gamma) = \int_\gamma \mathcal{O}_A^{(i)}$, where $\gamma$ is any $i$-cycle in homology.

## Observables and Intersection Theory

Let us consider the computation of the correlation function $\langle\mathcal{O}\rangle$ in the background field method. We first pick a background instanton configuration [15], and then integrate over the quantum fluctuations around that instanton. The relevant part of the quantum action is quadratic in the quantum fields, and localization of the model then ensures that such a computation is exact. The quantum fields are expanded into eigenfunctions of the operators that appear in the quadratic part of the action, and the functional integral is replaced by an integral over the eigenmodes. However, if there are fermionic zero modes, then those modes do not enter in the action. As a result, the fermionic integrals $(\int d\chi = 0)$ over those modes will cause $\langle\mathcal{O}\rangle$ to vanish unless it has the correct fermion content; the zero modes must be absorbed. In our case, a glance at the quantum action indicates that we should concern ourselves with the zero modes of the ghost $C^i$ and anti-ghost

$\bar{C}_{\alpha i}$. A $C^i$ zero mode is clearly in the kernel of the operator

$$\bar{D}^i_{\alpha j} = D_\alpha \delta^i{}_j + \epsilon_{\alpha\beta} J^i{}_j D^\beta + \epsilon_{\alpha\beta}(D_j J^i{}_k)\partial^\beta u^k \qquad [26]$$

and a $\bar{C}_{\alpha i}$ zero mode is a zero eigenfunction of its adjoint $\bar{D}^*$. In the BRST quantization of the model, the ghost fields $C^i$ are assigned ghost number +1, while the anti-ghost fields $\bar{C}_{\alpha i}$ have ghost number –1. It is therefore apparent that the vacuum expectation value of any observable will vanish unless that observable has a ghost number equal to the number of $\bar{D}$ zero modes, $a$, minus the number of $\bar{D}^*$ zero modes, $b$. This difference, $\omega = a - b$, is called the index of the operator $\bar{D}$.

There is a direct link between this index and the dimension of the moduli space of instantons. Recall that we are considering the space of maps $\Sigma \to M$ in a specified homotopy class, which satisfy equation [15]. It is then of interest to determine the dimension of the space of such solutions. To this aim, we examine the constraint that arises by considering an instanton $u^i$, and another neighboring solution $u^i + \hat{u}^i$, where $\hat{u}^i$ is an infinitesimal deformation. To first order in $\hat{u}^i$, we see that $\hat{u}^i$ must be a zero mode of the operator $\bar{D}$. This is no coincidence, and we can thus interpret the ghost fields $C^i$ as cotangent vectors to instanton moduli space $\mathcal{M}$. In particular, if $\mathcal{M}$ is a smooth manifold, then $\dim \mathcal{M} = a$. The index of the operator $\bar{D}$ is called the virtual dimension of the moduli space. In generic situations, the virtual dimension is equal to the actual dimension $\dim \mathcal{M}$.

It is possible to interpret some of the observables that we have described in terms of intersection theory applied to the moduli space of instantons. In particular, one can show that all correlation functions of the form

$$\left\langle \mathcal{O}^{(0)}_{A_1} \cdots \mathcal{O}^{(0)}_{A_s} \right\rangle \qquad [27]$$

are intersection numbers of certain submanifolds of moduli space. In order to see this in a simple example, we first recall the notion of Poincaré duality and the relationship between cohomology and homology.

Poincaré duality can be formulated as a relationship between de Rham cohomology (defined in terms of closed differential forms) and homology (defined in terms of subspaces of $M$). For our purposes here, it is sufficient to state that we can associate to each boundaryless submanifold $N$ of codimension $k$, a cohomology class $[\phi] \in H^k(M)$, such that

$$\int_M \phi \wedge \psi = \int_N \psi \qquad [28]$$

for all $[\psi] \in H^{n-k}(M)$. By $\psi$ on the right-hand side of this equation, we mean the pullback $i^*\psi$ under the inclusion $i : N \to M$. Conversely, to each closed $k$-form $\phi$ on $M$, we can associate an $(n-k)$-cycle $N$ (it is in general a chain of subspaces), unique up to homology, such that the previous relation is satisfied. Furthermore, one can show that the Poincaré dual to $N$ can be chosen in such a way that its support is localized within any given open neighborhood of $N$ in $M$ (essentially delta function support on $N$).

Let us now define the notion of transversal intersection. For simplicity, we will first consider the intersection of two submanifolds $M_1$ and $M_2$ contained in $M$. We will say that these two submanifolds have transversal intersection if the tangent spaces satisfy

$$T_x(M_1) + T_x(M_2) = T_x(M) \qquad [29]$$

for all $x \in M_1 \cap M_2$. It is a theorem that a submanifold of codimension $k$ can be locally "cut-out" by $k$ smooth functions, that is, the submanifold is locally specified by the zeros of this set of functions. It is a worthwhile exercise to convince oneself that the definition of transversal intersection is equivalent to the statement that the functions which cut-out $M_1$ are independent from those which cut-out $M_2$. Thus, we can write

$$\operatorname{codim}(M_1 \cap M_2) = \operatorname{codim}(M_1) + \operatorname{codim}(M_2) \qquad [30]$$

More generally, we say that the intersection $M_1 \cap \cdots \cap M_s$ of $s$ submanifolds is transversal if the intersection of every pair of them is transversal. It then follows trivially by the previous argument that the codimensions must satisfy

$$\operatorname{codim}(M_1 \cap \cdots \cap M_s) = \sum_{i=1}^{s} \operatorname{codim}(M_i) \qquad [31]$$

The special case which will be important for us occurs when the intersection of submanifolds is a collection of points, that is, when the codimension of the intersection is equal to the dimension of $M$. Since these points are isolated, the compactness of $M$ guarantees that they are finite in number.

We are now in a position to describe in what sense correlation functions of the form $\left\langle \mathcal{O}^{(0)}_{A_1} \cdots \mathcal{O}^{(0)}_{A_s} \right\rangle$ determine intersection numbers in the moduli space $\mathcal{M}$ of instantons. By definition, this moduli space is the set of maps from $\Sigma$ to $M$ which satisfy [15]. Let us consider the generic situation, where the virtual dimension of $\mathcal{M}$ (i.e., the index of $\bar{D}$) is equal to $\dim \mathcal{M}$. For convenience, let us begin by choosing the forms $A_i$ which represent de Rham cohomology classes on $M$, together with their Poincaré duals $M_i$, such that the forms have essentially delta function

support on their respective submanifolds. Since each of the operators in the correlation function depends on some fixed point $\sigma_i$, it is meaningful to define the submanifolds $L_i \equiv \{u \in \mathcal{M} \mid u(\sigma_i) \in M_i\} \subset \mathcal{M}$. Now, the correlation function represents a functional integral over the space of maps $\mathrm{Map}(\Sigma, M)$, and we have argued that this integral only receives contributions from the instanton configurations. Since the operators $A_i(u(\sigma_i))$ vanish unless $u \in L_i$ by our choice of the Poincaré duals, we see that the only contribution to the functional integral can be from those maps which lie in the intersection $L_1 \cap \cdots \cap L_s$. By ghost number considerations, this correlation function must vanish unless the codimension of the intersection equals the virtual dimension of $\mathcal{M}$. In the generic case where the virtual dimension is equal to $\dim \mathcal{M}$, this means that the intersection is simply a finite number of points. Intersection numbers $\pm 1$ can then be assigned to each point in the intersection $L_1 \cap \cdots \cap L_s$, by considering the relative orientation of the submanifolds $L_i$ at the intersection points. From the functional integral point of view, the computation reduces to an evaluation of the ratio of the bosonic determinant (integration over $u^i$) to the fermionic determinant (integration over $C^i$ and $\bar{C}_{\alpha i}$). In the Kähler case, for example, the intersection number assigned to each point in the intersection is always +1. This is due to the fact that the $C^I, \bar{C}_-{}^{\bar{I}}$ determinant is the complex conjugate of the $C^{\bar{I}}, \bar{C}_+{}^{I}$ determinant.

## *A* and *B* Models and Mirror Symmetry

The topological sigma model for a Kähler target space [12] is also known as the topological *A* model. In this case, the action can be recovered by twisting the standard $N=2$ supersymmetric sigma model. This twisting procedure amounts to a reassignment of the spins of the fields in the theory. However, there is an alternative twisting which can be done, and this leads to another model known as the topological *B* model. The usefulness of this observation lies in the fact that the topological *A* model on a Calabi–Yau target space *M* is related to the topological *B* model on the mirror of *M*. This relationship and the computation of correlation functions in the *A* and *B* models thus sheds light on the nature of mirror symmetry.

*See also:* Batalin–Vilkovisky Quantization; BRST Quantization; Functional Integration in Quantum Physics; Graded Poisson Algebras; Mathai–Quillen Formalism; Mirror Symmetry: A Geometric Survey; Several Complex Variables: Compact Manifolds; Singularities of the Ricci Flow; Topological Gravity, Two-Dimensional; Topological Quantum Field Theory: Overview; WDVV Equations and Frobenius Manifolds.

## Further Reading


Baulieu L and Singer I (1989) The topological sigma model. *Communications in Mathematical Physics* 125: 227.

Birmingham D, Blau M, Rakowski M, and Thompson G (1991) Topological field theory. *Physics Reports* 209: 129.

Birmingham D, Rakowski M, and Thompson G (1989) BRST quantization of topological field theory. *Nuclear Physics B* 315: 577.

Eguchi T and Yang SK (1990) $N=2$ superconformal models as topological field theories. *Modern Physics Letters A* 5: 1693.

Floer A (1987) Morse theory for fixed points of symplectic diffeomorphisms. *Bulletin of the American Mathematical Society* 16: 279.

Floer A (1988) An instanton invariant for three manifolds. *Communications in Mathematical Physics* 118: 215.

Gromov M (1985) Pseudo holomorphic curves in symplectic manifolds. *Inventiones Mathematicae* 82: 307.

Gromov M (1986) In: *Proceedings of the International Congress of Mathematicians*, p. 81. Berkeley.

Vafa C (1991) Topological Landau–Ginzburg models. *Modern Physics Letters A* 6: 337.

Witten E (1988) Topological sigma models. *Communications in Mathematical Physics* 118: 411.

Witten E (1991) Mirror manifolds and topological field theory. In: Yau ST (ed.) *Mirror Symmetry I*, p. 121. American Mathematical Society.


# Turbulence Theories

**R M S Rosa**, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil



## Introduction

Turbulence has initially been defined as an irregular motion in fluids. The cloud formations in the atmosphere and the motion of water in rivers make this point clear. These are but a few readily available examples of a multitude of flows which display turbulent regimes: from the blood that flows in our veins and arteries to the motion of air within our lungs and around us; from the flow of water in creeks to the atmospheric and oceanic currents; from the flows past submarines, ships, automobiles, and aircraft to the combustion processes propelling them; and in the flow of gas, oil, and water, from the prospecting end to the entrails of the cities. The great majority of flows in nature and in engineering applications are somehow turbulent.

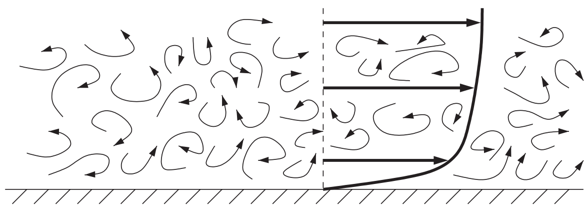


**Figure 1** Illustration of the irregular motion of a turbulent flow over a flat plate (thin lines), and of the well-defined velocity profile of the mean flow (thick lines).

But turbulent flows are much more than simply irregular. More refined definitions were desirable and were later coined. A definitive and precise one, however, may only come when the phenomenon is fully understood. Nevertheless, several characteristic properties of a turbulent flow can be listed:

*Irregularity and unpredictability* A turbulent flow is irregular both in space and time, displaying unpredictable, random patterns.

*Statistical order* From the irregularity of a turbulent motion there emerges a certain statistical order. Mean quantities and correlation are regular and predictable (Figure 1).

*Wide range of active scales* A wide range of scales of motion are active and display an irregular motion, yielding a large number of degrees of freedom.

*Mixing and enhanced diffusivity* The fluid particles undergo complicated and convoluted paths, causing a large mixing of different parts of fluid. This mixing significantly enhances diffusion, increasing the transport of momentum, energy, heat, and other advected quantities.

*Vortex stretching* When a moving portion of fluid also rotates transversally to its motion an increase in speed causes it to rotate faster, a phenomenon called vortex stretching. This causes that portion of fluid to become thinner and elongated, and fold and intertwine with other such portions. This is an intrinsically three-dimensional mechanism which plays a fundamental role in turbulence and is associated with large fluctuations in the vorticity field.

## Turbulent Regimes

Turbulence is studied from many perspectives. The subject of "transition to turbulence" attempts to describe the initial mechanisms responsible for the generation of turbulence starting from a laminar motion in particular geometries. This transition can be followed with respect to position in space (e.g., the flow becomes more complicated as we look further downstream on a flow past an obstacle or over a flat plate) or to parameters (e.g., as we increase the angle of attack of a wing or the pressure gradient in a pipe). This subject is divided into two cases: wall-bounded and free-shear flows. In the former, the viscosity, which causes the fluid to adhere to the surface of the wall, is the primary cause of the instability in the transition process. In the latter, inviscid mechanisms such as mixing layers and jets are the main factors. The tools for studying the transition to turbulence include linearization of the equations of motion around the laminar solution, nonlinear amplitude equations, and bifurcation theory.

"Fully developed turbulence," on the other hand, concerns turbulence which evolves without imposed constraints, such as boundaries and external forces. This can be thought of turbulence in its "pure" form, and it is somewhat a theoretical framework for research due to its idealized nature. Hypotheses of homogeneity (when the mean quantities associated with the statistical order characterizing a turbulent flow are independent in space), stationarity (*idem* in time), and isotropy (*idem* with respect to rotations in space) concern fully developed turbulent flows. The Kolmogorov theory was developed in this context and it is the most fundamental theory of turbulence. Current research is dedicated in great part to unveil the mechanisms behind a phenomenon called intermittency and how it affects the laws obtained from the conventional theory. Research is also dedicated to derive such laws as much from first principles as possible, minimizing the use of phenomenological and dimensional analysis.

Real turbulent flows involve various regimes at once. A typical flow past a blunt object, for instance, displays laminar motion at its upstream edge, a turbulent boundary layer further downstream, and the formation of a turbulent wake (Figure 2). The subject of turbulent boundary layer is a world in itself with current research aiming to determine mean properties of flows over rough surfaces and varied topography. Convective turbulence involves coupling with active scalars such as

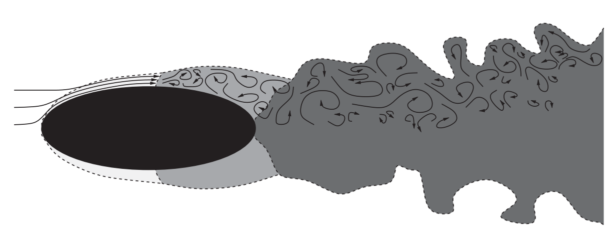


**Figure 2** Illustration of a flow past an object, with a laminar boundary layer (light gray), a turbulent boundary layer (medium gray), and a turbulent wake (dark gray).

large heat gradients, occurring in the atmosphere, and large salinity gradients, in the ocean. Geophysical turbulence involves also stratification and the anisotropy generated by Earth's rotation. Anisotropic turbulence is also crucial in astrophysics and plasma theory. Multiphase and multicomponent turbulence appear in flows with suspended particles or bubbles and in mixtures such as gas, water, and oil. Transonic and supersonic flows are also of great importance and fall into the category of compressible turbulence, much less explored than the incompressible case.

In all those real situations one would like, from the engineering point of view, to compute mean properties of the flow, such as drag and lift for more efficient designs of aircraft, ships, and other vehicles. Knowledge of the drag coefficient is also of fundamental importance in the design of pipes and pumps, from pipelines to artificial human organs. Mean turbulent diffusion coefficients of heat and other passive scalars – quantities advected by the flow without interfering on it, such as chemical products, nutrients, moisture, and pollutants – are also of major importance in industry, ecology, meteorology, and climatology, for instance. And in most of those cases a large amount of research is dedicated to the "control of turbulence," either to increase mixing or reduce drag, for instance. From a theoretical point of view, one would like to fully understand and characterize the mechanisms involved in turbulent flows, clarifying this fascinating phenomenon. This could also improve practical applications and lead to a better control of turbulence.

The concept of "two-dimensional turbulence" is controversial. A two-dimensional flow may be irregular and display mixing, statistical order, and a wide range of active scales but definitely it does not involve vortex stretching since the velocity field is always perpendicular to the vorticity field. For this reason many researchers discard two-dimensional turbulence altogether. It is also argued that real two-dimensional flows are unstable at complicated regimes and soon develop into a three-dimensional flow. Nevertheless, many believe that two-dimensional turbulence, even lacking vortex stretching, is of fundamental theoretical importance. It may shed some light into the three-dimensional theory and modeling, and it can serve as an approximation to some situations such as the motion of the atmosphere and oceans in the large and meso scales and some magnetohydrodynamic flows. The relative shallowness of the atmosphere and oceans or the imposition of a strong uniform magnetic field may force the flow into two-dimensionality, at least for a certain range of scales.

"Chaos" serves as a paradigm for turbulence, in the sense that it is now accepted that turbulence is a dynamic processes in a sensitive deterministic system. But not all chaotic motions in fluids are termed turbulent for they may not display mixing and vortex stretching or involve a wide range of scales. An important such example appears in the dispersive, nonlinear interactions of waves.

## The Equations of Motion

It is usually stressed that turbulence is a continuum phenomenon, in the sense that the active scales are much larger than the collision mean free path between molecules. For this reason, turbulence is believed to be fully accounted for by the Navier–Stokes equations.

In the case of incompressible homogeneous flows, the Navier–Stokes equations in the Eulerian form and in vector notation read

$$\frac{\partial \boldsymbol{u}}{\partial t} - \nu\Delta\boldsymbol{u} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} + \nabla p = \boldsymbol{f} \qquad [1a]$$

$$\nabla\cdot\boldsymbol{u} = 0. \qquad [1b]$$

Here, $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x}, t) = (u_1, u_2, u_3)$ denotes the velocity vector of an idealized fluid particle located at position $\boldsymbol{x} = (x_1, x_2, x_3)$, at time $t$. The mass density in a homogeneous flow is constant, denoted $\rho$. The constant $\nu$ denotes the kinematic viscosity of the fluid, which is the molecular viscosity $\mu$ divided by $\rho$. The variable $p = p(\boldsymbol{x}, t)$ is the kinematic pressure, and $\boldsymbol{f} = \boldsymbol{f}(\boldsymbol{x}, t) = (f_1, f_2, f_3)$ denotes the mass density of volume forces.

Equation [1a] expresses the conservation of linear momentum. The term $\nu\Delta\boldsymbol{u}$ accounts for the dissipation of energy due to molecular viscosity, and the nonlinear term $(\boldsymbol{u}\cdot\nabla)\boldsymbol{u}$, also called the inertial term, accounts for the redistribution of energy among different structures and scales of motion. Equation [1b] represents the incompressibility condition. In Einstein's summation convention, these equations can be written as

$$\frac{\partial u_i}{\partial t} + \nu\frac{\partial^2 u_i}{\partial x_j^2} + u_j\frac{\partial u_i}{\partial x_j} + \frac{\partial p}{\partial x_i} = f_i, \quad \frac{\partial u_j}{\partial x_j} = 0$$

## The Reynolds Number

The transition to turbulence was carefully studied by Reynolds in the late nineteenth century in a series of experiments in which water at rest in a tank was allowed to flow through a glass pipe. Starting with dimensional analysis, Reynolds argued that a critical

value of a certain nondimensional quantity was likely to exist beyond which a laminar flow gives rise to a "sinuous" motion. This was followed by observations of the flow for tubes with different diameter $L$, different mean velocities $U$ across the tube section, and with the kinematic viscosity $\nu = \rho/\mu$ being altered through changes in temperature. The experiments confirmed the existence of such a critical value for what is now called the Reynolds number:

$$Re = \frac{LU}{\nu}$$

The dimensional analysis argument can be reproduced in the following form: the physical dimension for the inertial term in [1a] is $U^2/L$, while that for the viscous term is $\nu U/L^2$. The ratio between them is precisely $Re = LU/\nu$. For small values of $Re$ viscosity dominates and the flow is laminar, whereas for large values of $Re$ the inertial term dominates, and the flow becomes more complicated and eventually turbulent. In applications, different types of Reynolds number can be used depending on the choice of the characteristic velocity and length, but in any case, the larger the Reynolds number, the more complicated the flow.

## The Reynolds Equations

Another advance put forward by Reynolds in a subsequent article was to decompose the flow into a mean component and the remaining fluctuations. In terms of the velocity and pressure fields this can be written as

$$\boldsymbol{u} = \bar{\boldsymbol{u}} + \boldsymbol{u}', \qquad p = \bar{p} + p' \qquad [2]$$

with $\bar{\boldsymbol{u}}$ and $\bar{p}$ representing the mean components and $\boldsymbol{u}'$ and $p'$, the fluctuations. By substituting [2] into [1], one finds the Reynolds-averaged Navier–Stokes (RANS) equations for the mean flow:

$$\frac{\partial \bar{\boldsymbol{u}}}{\partial t} - \nu \Delta \bar{\boldsymbol{u}} + (\bar{\boldsymbol{u}} \cdot \nabla)\bar{\boldsymbol{u}} + \nabla \bar{p} = \boldsymbol{f} + \nabla \cdot \tau$$

$$\nabla \cdot \bar{\boldsymbol{u}} = 0$$

It differs from [1] only by the addition of the Reynolds stress tensor:

$$\tau = -\overline{\boldsymbol{u}' \otimes \boldsymbol{u}'} = -\left(\overline{u_i' u_j'}\right)_{i,j=1}^{3}$$

In a laminar flow, the fluctuations are negligible, otherwise this decomposition shows how they influence the mean flow through this additional turbulent stresses.

## The Closure Problem and Turbulence Models

The RANS equations cannot be solved directly for the mean flow since the Reynolds stresses are unknown. Equations for these stress terms can be derived but they involve further unknown moments. This continues with equations for moments of a given order depending on new moments up to a higher order, leading to an infinite system of equations known as the Friedman–Keller system. For practical applications, approximations closing the system at some finite order are needed, in what is called the closure problem. Several *ad hoc* approximations exist, the most famous being the Boussinesq eddy-viscosity approximation, in which the turbulent fluctuations are regarded as increasing the viscosity of the flow. Prandtl's mixing-length hypothesis yields a prescription for the computation of this eddy viscosity, and together they form the basis of the algebraic models of turbulence. Other models involve additional equations, such as the $k$-$\epsilon$ and $k$-$\omega$ models. Most of the practical computations of industrial flows are based on such lower-order models, and a large amount of research is done to determine appropriate values for the various *ad hoc* parameters which appear in these models and which are highly dependent on the geometry of the flow. This dependency can be explained by the fact that the RANS is supposed to model the mean flow even at the large scales of motion, which are highly affected by the geometry.

Computational fluid dynamics (CFD) is indeed a fundamental tool in turbulence, both for research and engineering applications. From the theoretical side, direct numerical simulations (DNS), which attempt to resolve all the active scales of the flow, reveal some fundamental mechanisms involved in the transition to turbulence and in vortex stretching. As for applications, DNS applies to flows up to low-Reynolds turbulence, with the current computational power not allowing for a full resolution of all the scales involved in high-Reynolds flows. And the current rate of evolution of computational power predicts that this will continue so for several decades.

An intermediate CFD method between RANS and DNS is the large-eddy simulation (LES), which attempts to fully resolve the large scales while modeling the turbulent motion at the smaller scales. Several models have been proposed which have their own advantages and limitations as compared to RANS and DNS. It is currently a subject of intense research, particularly for the development of suitable models for the structure functions near the boundary. Theoretical results on fully developed turbulence play a fundamental role in the modeling process.

LESs are a promising tool and they have been successfully applied to a number of situations. The choice of the best method for a given application, however, depends very much on the Reynolds number of the flow and the prior knowledge of similar situations for adjusting the parameters.

## Elements of the Statistical Theory

Several types of averages can be used. The ensemble average is taken with respect to a number of experiments at nearly identical conditions. Despite the irregular motion of, say, the velocity vector $\boldsymbol{u}^{(n)}(\boldsymbol{x}, t)$ of each experiment $n = 1, \ldots, N$, the average value

$$\bar{u}(\boldsymbol{x}, t) = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{u}^{(n)}(\boldsymbol{x}, t)$$

is expected to behave in a more regular way. This type of averaging is usually denoted with the symbol $\langle \cdot \rangle$. This notion can be cast into the context of a probability space $(\mathcal{M}, \Sigma, \mathcal{P})$, where $\mathcal{M}$ is a set, $\Sigma$ is a $\sigma$-algebra of subsets of $\mathcal{M}$, and $\mathcal{P}$ is a probability measure on $\Sigma$. The velocity field is a random variable in the sense that it is a density function $\omega \mapsto \boldsymbol{u}(\boldsymbol{x}, t, \omega)$ from $\mathcal{M}$ into the space of time-dependent divergence-free velocity fields. The mean velocity field in this context is regarded as

$$\langle \boldsymbol{u}(\boldsymbol{x}, t)\rangle = \int_{\mathcal{M}} \boldsymbol{u}(\boldsymbol{x}, t, \omega)\mathrm{d}\mathcal{P}(\omega)$$

Other flow quantities such as energy and correlations in space and time can be expressed by means of a function $\varphi = \varphi(\boldsymbol{u}(\cdot, \cdot))$ of the velocity field, with their mean value given by

$$\langle \varphi(\boldsymbol{u}(\cdot, \cdot))\rangle = \int_{\mathcal{M}} \varphi(\boldsymbol{u}(\cdot, \cdot, \omega)\mathrm{d}\mathcal{P}(\omega)$$

In general, the statistics of the flow are allowed to change with time. A particular situation is when statistical equilibrium is reached, so that $\langle \boldsymbol{u}(\boldsymbol{x}, t)\rangle$ and, more generally, $\langle \varphi(\boldsymbol{u}(\cdot, \cdot + t))\rangle$ are independent of $t$. In this case, an ergodic assumption is usually invoked, which means that for "most" individual flows $\boldsymbol{u}(\cdot, \cdot, \omega_0)$ (i.e, for almost all $\omega_0$ with respect to the probability measure $\mathcal{P}$), the time averages along this flow converge to the mean ensemble value as the period of the average increases to the mean value obtained by the ensemble average:

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T \varphi(\boldsymbol{u}(\cdot, \cdot + s, \omega_0))\mathrm{d}s = \int_{\mathcal{M}} \varphi(\boldsymbol{u}(\cdot, \cdot, \omega))\mathrm{d}\mathcal{P}(\omega)$$

Based on this assumption, the averages may in practice be calculated as time averages over a sufficiently large period $T$. There is a related argument for substituting space averages by time averages and based on the mechanics of turbulence which is called the "Taylor hypothesis."

Another fundamental concept in the statistical theory is that of homogeneity, which is the spatial analog of the statistical equilibrium in time. In homogeneous turbulence, the statistical quantities of a flow are independent of translations in space, that is,

$$\langle \varphi(\boldsymbol{u}(\cdot + \ell, \cdot))\rangle = \langle \varphi(\boldsymbol{u}(\cdot, \cdot))\rangle$$

for all $\ell \in \mathbb{R}^3$. The concept of isotropic turbulence assumes further independence with respect to rotations and reflections in the frame of reference, that is,

$$\langle \varphi(Q^{\mathrm{t}}\boldsymbol{u}(Q\cdot, \cdot))\rangle = \langle \varphi(\boldsymbol{u}(\cdot, \cdot))\rangle$$

for all orthogonal transformations $Q$ in $\mathbb{R}^3$, with adjoint $Q^{\mathrm{t}}$.

Under the homogeneity assumption, mean quantities can be defined independently of position in space, such as the mean kinetic energy per unit mass

$$e = \frac{1}{2}\langle |\boldsymbol{u}(\boldsymbol{x})|^2\rangle = \frac{1}{2}\sum_{j=1}^{3}\langle |u_i(\boldsymbol{x})|^2\rangle$$

and the mean rate of viscous energy dissipation per unit mass and unit time

$$\epsilon = \nu \sum_{i=1}^{3}\langle |\nabla u_i(\boldsymbol{x})|^2\rangle = \nu \sum_{i,j=1}^{3}\left\langle \left|\frac{\partial u_i(\boldsymbol{x})}{\partial x_j}\right|^2\right\rangle$$

The mean kinetic energy can be written as $e = \mathrm{tr}R(0)/2$, where

$$\mathrm{tr}R(\ell) = R_{11}(\ell) + R_{22}(\ell) + R_{33}(\ell), \quad \ell \in \mathbb{R}^3,$$

is the trace of the correlation tensor

$$R(\ell) = \langle \boldsymbol{u}(\boldsymbol{x}) \otimes \boldsymbol{u}(\boldsymbol{x} + \ell)\rangle = (R_{ij}(\ell))_{i,j=1}^3 = (\langle u_i(\boldsymbol{x})u_j(\boldsymbol{x} + \ell)\rangle)_{i,j=1}^3$$

which measures the correlation between the velocity components at different positions in space. From the homogeneity assumption, this tensor is a function only of the relative position $\ell$. Then, assuming that the Fourier transform of $\mathrm{tr}R(\ell)$ exists, and denoting it by $Q(\kappa)$, for $\kappa \in \mathbb{R}^3$, we have

$$\mathrm{tr}R(\ell) = \frac{1}{(2\pi)^{3/2}}\int_{\mathbb{R}^3} Q(\kappa)\mathrm{e}^{\mathrm{i}\ell\cdot\kappa}\mathrm{d}\kappa = 2\int_0^\infty S(\kappa)\mathrm{e}^{\mathrm{i}\ell\cdot\kappa}\mathrm{d}\kappa$$

where $\mathcal{S}(\kappa)$ is the energy spectrum defined by

$$\mathcal{S}(\kappa) = \frac{1}{2(2\pi)^{3/2}} \int_{|\kappa|=\kappa} Q(\kappa) \mathrm{d}\Sigma(\kappa)$$
$$\forall \kappa > 0$$

with $\mathrm{d}\Sigma(\kappa)$ denoting the area element of the 2-sphere of radius $|\kappa|$. Then we can write

$$e = \frac{1}{2}\langle |\boldsymbol{u}(\boldsymbol{x})|^2 \rangle = \frac{1}{2}\mathrm{tr}R(0) = \int_0^\infty \mathcal{S}(\kappa)\mathrm{d}\kappa$$

By expanding the velocity coordinates into Fourier modes $\exp(\ell \cdot \kappa)$, with $\kappa \le |\kappa| \le \kappa + \mathrm{d}\kappa$ and interpreting them as "eddies" with characteristic wave number $|\kappa|$, the quantity $\mathcal{S}(\kappa)\mathrm{d}\kappa$ can be interpreted as the energy of the component of the flow formed by the "eddies" with characteristic wave number between $\kappa$ and $\kappa + \mathrm{d}\kappa$.
Similarly,

$$\epsilon = 2\nu \int_0^\infty \kappa^2 \mathcal{S}(\kappa)\mathrm{d}\kappa$$

and we obtain the dissipation spectrum $2\nu\kappa^2\mathcal{S}(\kappa)$, which can be interpreted as the density of energy dissipation occurring at wave number $\kappa$.

In the previous arguments it is assumed that the flow extends to all the space $\mathbb{R}^3$. This avoids the presence of boundaries, addressing the idealized case of fully developed turbulence. It is sometimes customary to assume as well that the flow is periodic in space to avoid problems with unbounded domains such as infinite kinetic energy.

The random nature of turbulent flows was greatly explored by Taylor in the early twentieth century, who introduced most of the concepts described above. Another important concept he introduced was the Taylor microlength $\ell_\mathrm{T}$, which is a characteristic length for the small scales based on the correlation tensor. A microscale Reynolds number based on the Taylor microlength is very often used in applications.

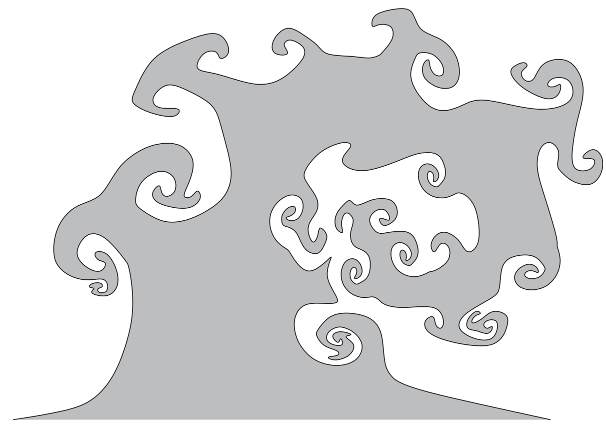


**Figure 3** Illustration of the eddy breakdown process in which energy is transferred to smaller eddies and so on until the smallest scales are reached and the energy is dissipated by viscosity.

## Kolmogorov Theory

An inspiring concept in the theory of turbulence is Richardson's "energy cascade" process. For large Reynolds numbers the nonlinear term dominates the viscosity according to the dimensional analysis, but this is valid only for the large-scale structures. The small scales have their own characteristic length and velocity. In the cascade process, the inertial term is responsible for the transfer of energy to smaller and smaller scales until small enough scales are reached for which viscosity becomes important (Figure 3). At those smallest scales kinetic energy is finally dissipated into heat. It should be emphasized that turbulence is a dissipative process; no matter how large the Reynolds number is, viscosity plays a role in the smallest scales.

The Kolmogorov theory of locally isotropic turbulence allows for inhomogeneity and anisotropy in the large scales, which contain most of the energy, assuming that with the cascade transfer of energy to smaller scales, the orienting effects generated in the large scales become weaker and weaker so that for sufficiently small eddies the motion becomes statistically homogeneous, isotropic, and independent of the particular energy-productive mechanisms. He proposed that the statistical regime of the small-scale eddies is then universal and depends only on $\nu$ and $\epsilon$. The equilibrium range is defined as the range of scales in which this universality holds.

Simple dimensional analysis shows that the only algebraic combination of $\nu$ and $\epsilon$ with dimension of length is $\ell_\epsilon = (\nu^3/\epsilon)^{1/4}$, which is then interpreted as that near which the viscous effect becomes important and hence most of the energy dissipation takes place. The scale $\ell_\epsilon$ is known as Kolmogorov dissipation length.

Kolmogorov theory gives particular attention to moments involving differences of velocities, such as the $p$th-order structure function

$$S_p(\ell) \stackrel{\mathrm{def}}{=} \langle (\boldsymbol{u}(\boldsymbol{x} + \ell\boldsymbol{e}) \cdot \boldsymbol{e} - \boldsymbol{u}(\boldsymbol{x})\boldsymbol{e})^p \rangle$$

where $\boldsymbol{e}$ may be taken as an arbitrary unit vector, thanks to the isotropy assumption. By restricting the search for universal laws for the structure functions only for small values of $\ell$ anisotropy and inhomogeneity are allowed in the large scales.

The theory assumes a wide separation between the energy-containing scales, of order say $\ell_0$, and the energy-dissipative scales, of order $\ell_\epsilon$, so that the cascade process occurs within a wide range of scales $\ell$ such that $\ell_0 \gg \ell \gg \ell_\epsilon$. In this range, termed the inertial range, the viscous effects are still negligible and the statistical regime should depend only on $\epsilon$. Then, the Kolmogorov "two-thirds law" asserts that within the inertial range the second-order correlations must be proportional to $(\epsilon\ell)^{2/3}$, that is,

$$S_2(\ell) = C_K(\epsilon\ell)^{2/3}$$

for some constant $C_K$ known as the Kolmogorov constant in physical space (there is a related constant in spectral space). The argument extends to higher-order structure functions, yielding

$$S_p(\ell) = C_p(\epsilon\ell)^{p/3}$$

Kolmogorov's derivation of these results was not by dimensional analysis, it was in fact a more convincing self-similarity argument based on the universality assumed for the equilibrium range. A different argument without resorting to universality assumptions, however, was applied to the third-order structure function, yielding the more precise "four-fifths law":

$$S_3(\ell) = -\tfrac{4}{5}\epsilon\ell$$

The "Kolmogorov five-thirds law" concerns the energy spectrum $\mathcal{S}(\kappa)$ and is the spectral version of the two-thirds law, given by Obukhoff:

$$\mathcal{S}(\kappa) = C'_K\epsilon^{2/3}\kappa^{-5/3}$$

The constant $C'_K$ is the Kolmogorov constant in spectral space. The spectral version of the dissipation length is the Kolmogorov wave number $\kappa_\epsilon = (\epsilon/\nu^3)^{1/4}$.

A typical distribution of energy in a turbulent flow is depicted in Figure 4. The energy is concentrated on the large scales, while the dissipation is concentrated near the Kolmogorov scale $\ell_\epsilon$. The four-fifths law becomes visible as a straight line in the logarithmic scale.

A more precise mechanism for the energy cascade assumes that in the inertial range, eddies with length scale $\ell$ transfer kinetic energy to smaller eddies during their characteristic timescale, also known as circulation time. If $u_\ell$ is their characteristic velocity, then $\tau_\ell = \ell/u_\ell$ is their circulation time, so that the kinetic energy transferred from these eddies during this time is

$$\epsilon_\ell \sim \frac{u_\ell^2}{\tau_\ell} = \frac{u_\ell^3}{\ell}$$

In statistical equilibrium, the energy lost to the smaller scales equals the energy gained from the larger scales, and that should also equal the total kinetic energy dissipated by viscous effects. Hence, $\epsilon_\ell \equiv \epsilon$, and we find

$$\epsilon \sim \frac{u_\ell^3}{\ell}$$

It also follows that $\tau_\ell = \ell/u_\ell = \ell(\epsilon\ell)^{-1/3} = \epsilon^{-1/3}\ell^{2/3}$ so that the circulation time decreases with the length scale and becomes of the order of the viscous dissipation time $(\nu/\epsilon)^{1/2}$ precisely when $\ell \sim \ell_\epsilon$.

A similar relation between $\epsilon$ and the large scales can also be obtained with heuristic arguments: let $e$ be the mean kinetic energy and $\ell_0$, a characteristic length for the large scales. Then $u_0$ given by $e = u_0^2/2$ is a characteristic velocity for the large scales, and $\tau_0 = \ell_0/u_0$ is the large-scale circulation time. In statistical equilibrium, the rate $\epsilon$ of kinetic energy dissipated per unit time and unit mass is expected to be of the order of $e/\tau_0$, hence

$$\epsilon \sim \frac{u_0^3}{\ell_0}$$

which is called the "energy dissipation law."

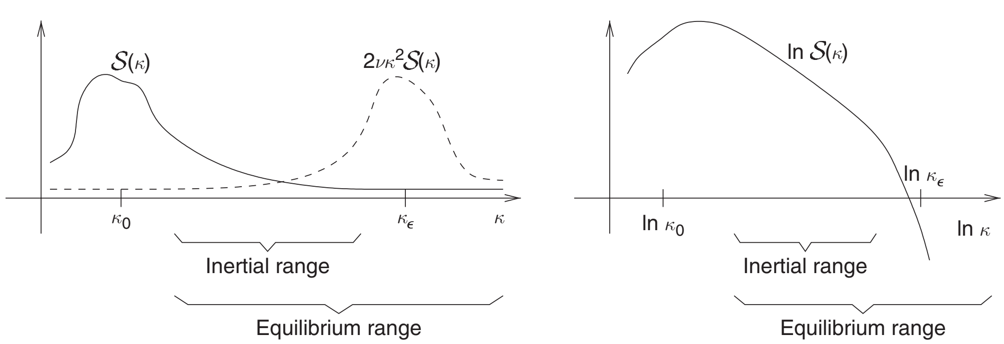



**Figure 4** A typical distribution for the energy spectrum $\mathcal{S}(\kappa)$ and the dissipation spectrum $2\nu\kappa^2\mathcal{S}(\kappa)$ in spectral space in nonlogarithmic and logarithmic scales. The energy is mostly concentrated on the large scales while the dissipation is concentrated near the dissipation scale. In the logarithmic scale, the four-fifths law for the energy spectrum stands out as a straight line with slope $-4/5$ over the inertial range.

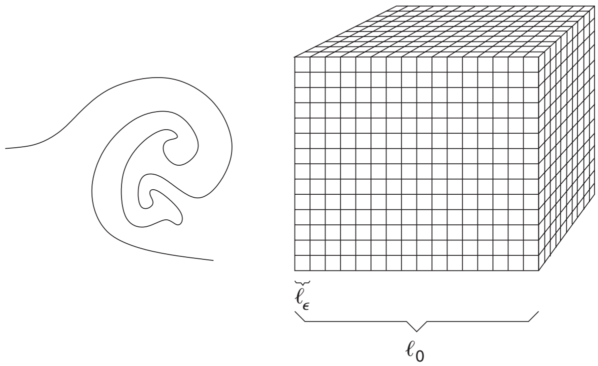



**Figure 5** A schematic representation of a flow structure displaying a range of active scales and a three-dimensional grid with linear dimension $\ell_0$ and mesh length $\ell_\epsilon$, sufficient to represent all the active scales in a turbulent flow. The number of degrees of freedom is the number of blocks: $(\ell_0/\ell_\epsilon)^3$.

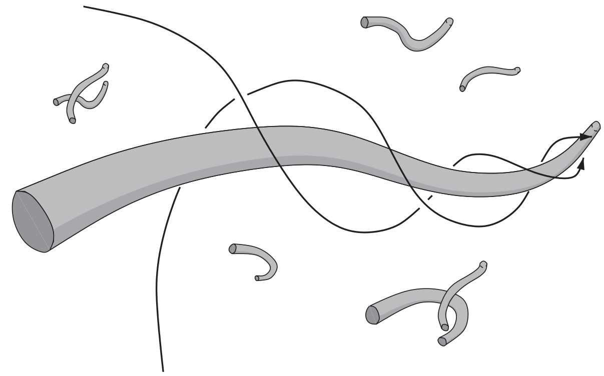


**Figure 6** A portion of rotating fluid gets stretched and thinned as the flow speeds up, generating one of many coherent structures of high vorticity and low dissipation.

From the energy dissipation law, several relations between characteristic quantities of turbulent flows can be obtained, such as $\ell_0/\ell_\epsilon \sim Re^{3/4}$, for $Re = \ell_0 u_0/\nu$.

Now, assuming the active scales in a turbulent flow exist down to the Kolmogorov scale $\ell_\epsilon$, one needs a three-dimensional grid with mesh spacing $\ell_\epsilon$ to resolve all the scales, which means that the number $N$ of degrees of freedom of the system is of the order of $N \sim (\ell_0/\ell_\epsilon)^3$ (see Figure 5). This number can be estimated in terms of the Reynolds number by $N \sim Re^{9/4}$. This relation is important in predicting the computational power needed to simulate all the active scales in turbulent flows.

Several such universal laws can be deduced and extended to other situations such as turbulent boundary layers, with the famous logarithmic law of the wall. They play a fundamental role in turbulence modeling and closure, for the calculation of the mean flow and other quantities.

## Intermittency

The universality hypothesis based on a constant mean energy dissipation rate throughout the flow received some criticisms and was later modified by Kolmogorov in an attempt to account for observed large deviations on the mean rate of energy dissipation. Such phenomenon of intermittency is related to the vortex stretching and thinning mechanism, which leads to the formation of coherent structures of vortex filaments of high vorticity and low dissipation (Figure 6). These filaments have diameter as small as the Kolmogorov scale and longitudinal length extending from the Taylor scale up to the large scales and with a lifetime of the order of the large-scale circulation time.

It has been argued based on experimental evidence that intermittency leads to modified power laws $S_p(\ell) \propto \ell^{\zeta(p)}, \zeta(p) < p/3$, for high-order ($p > 3$) structure functions. The issues of intermittency and coherent structures and whether and how they could affect the deductions of the universality theory such as the power laws for the structure functions are far from settled and are currently one of the major and most fascinating issues being addressed in turbulence theory. Several phenomenological theories attempt to adjust the universality theory to the existence of such coherent structures. Multifractal models, for instance, suppose that the eddies generated in the cascade process do not fill up the space and form multifractal structures. Field-theoretic renormalization group develops techniques based on quantum field renormalization theory. Intermediate asymptotics also exploits self-similar analysis and renormalization theory but with a somewhat different flavor. Detailed mathematical analysis of the vorticity equations is also playing a major role in the understanding of the dynamics of the vorticity field.

## Mathematical Aspects of Turbulence Theory

From a mathematical perspective, it is fundamental to develop a rigorous background upon which to study the physical quantities of a turbulent flow. The first problem in the mathematical theory is related to the deterministic nature of chaotic systems assumed in dynamical system theory and believed to hold in turbulence. This has actually not been proved for the Navier–Stokes equations. It is in fact one of the most outstanding open problems in mathematics to determine whether given an initial condition for the velocity field there exists, in some sense, a unique solution of the Navier–Stokes equations starting with this initial condition and valid for all later times. It has been proved that a global solution (i.e., valid for all later

times) exists but which may not be unique, and it has been proved that unique solutions exist which may not be global (i.e., they are guaranteed to exist as unique solutions only for a finite time).

The difficulty here is the possible existence of singularities in the vorticity field (vorticity becoming infinite at some points in space and time). Depending on how large the singularity set is, uniqueness may fail in strictly mathematical terms. The existence of singularities may not be a purely mathematical curiosity, it may in fact be related with the intermittency phenomenon. Rigorous studies of the vorticity equation may continue to reveal more fundamental aspects on vortex dynamics and coherent structures.

The statistical theory has also been put into a firm foundation with the notion of statistical solution of the Navier–Stokes equations. It addresses the existence and regularity of the probability distribution assumed for turbulent flows and of the fundamental elements of the statistical theory such as correlation functions and spectra. Based on that, a number of relations between physical quantities of turbulent flows may be derived in a mathematically sound and definitive way. This does not replace other theories, it is mostly a mathematical framework upon which other techniques can be applied to yield rigorous results.

Despite the difficulties in the mathematical theory of the NSE some successes have been collected such as estimates for the number of degrees of freedom in terms of fractal dimensions of suitable sets associated with the solutions of the Navier–Stokes equations, and partial estimates of a number of relations derived in the statistical theory of fully developed turbulence.

*See also:* Bifurcations in Fluid Dynamics; Geophysical Dynamics; Incompressible Euler Equations: Mathematical Theory; Intermittency in Turbulence; Inviscid Flows; Lagrangian Dispersion (Passive Scalar); Stochastic Hydrodynamics; Variational Methods in Turbulence; Viscous Incompressible Fluids: Mathematical Theory; Vortex Dynamics; Wavelets: Application to Turbulence.

## Further Reading


Adzhemyan LT, Antonov NV, and Vasiliev AN (1999) *The Field Theoretic Renormalization Group in Fully Developed Turbulence*. Amsterdam: Gordon and Breach.

Barenblatt GI and Chorin AJ (1998) New perspectives in turbulence: scaling laws, asymptotics, and intermittency. *SIAM Review* 40(2): 265–291.

Batchelor GK (1953) The Theory of Homogeneous Turbulence. *Cambridge Monographs on Mechanics and Applied Mathematics*. New York: Cambridge University Press.

Chorin AJ (1994) Vorticity and Turbulence. *Applied Mathematical Sciences* vol. 103, New York: Springer.

Constantin P (1994) Geometric statistics in turbulence. *SIAM Review* 36: 73–98.

Foias C, Manley OP, Rosa R, and Temam R (2001) Navier–Stokes Equations and Turbulence. *Encyclopedia of Mathematics and its Applications*, vol. 83. Cambridge: Cambridge University Press.

Friedlander S and Topper L (1961) *Turbulence. Classic Papers on Statistical Theory*. New York: Interscience Publisher.

Frisch U (1995) *Turbulence. The Legacy of A. N. Kolmogorov*. Cambridge: Cambridge University Press.

Hinze JO (1975) Turbulence. *McGraw-Hill Series in Mechanical Engineering*. New York: McGraw-Hill.

Holmes P, Lumley JL, and Berkooz G (1996) *Turbulence, Coherent Structures, Dynamical Systems, and Symmetry*. Cambridge: Cambridge University Press.

Lesieur M (1997) Turbulence in Fluids. *Fluid Mechanics and its Applications*, 3rd edn. vol. 40. Dordrecht: Kluwer Academic.

Monin AS and Yaglom AM (1975) *Statistical Fluid Mechanics: Mechanics of Turbulence 2*. Cambridge: MIT Press.

Sagaut P (2001) *Large Eddy Simulation for Incompressible Flows*. Berlin: Springer.

Schlichting H and Gersten K (2000) *Boundary Layer Theory*, 8th edn. Berlin: Springer.

Tennekes H and Lumley JL (1972) *A First Course in Turbulence*. Cambridge, MA: MIT Press.

Vishik MI and Fursikov AV (1988) *Mathematical Problems of Statistical Hydrodynamics*. Dordrecht: Kluwer.

Wilcox DC (2000) *Turbulence Modeling for CFD*, 2nd edn. Anaheim, CA: DCW Industries Inc.


# Twistor Theory: Some Applications

**L Mason**, University of Oxford, Oxford, UK



## Introduction

Roger Penrose introduced twistor theory as a geometrical framework for basic physics in order to unify quantum theory and gravity. This program has had many successes along the way, but the long-term goals of reformulating and superceding the established theories of basic physics are still a long way from being fulfilled. Nevertheless, the successes have had many important applications across mathematics and mathematical physics. This article will concentrate on three areas of application: integrable systems, geometry, and perturbative gauge theory (via twistor-string theory). It is intended to be self-contained as far as possible, but the reader may well find it easier to first read the article Twistors.

## Twistor Theory

A basic motivation of twistor theory is to bring out the complex (holomorphic) geometry that underlies real spacetime. In general relativity, a spacetime is a 4-manifold with metric $g$ of signature $(1,3)$, and when it is flat, that is, $g = \mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2$, where $(t,x,y,z)$ are coordinates on $\mathbb{R}^4$, it is called Minkowski space. The first appearance of a complex structure arises from the fact that, at a given event, the celestial sphere of light rays (directions of zero length with respect to $g$) naturally has the structure of the Riemann sphere, $\mathbb{CP}^1$, in such a way that Lorentz transformations (linear transformations of the tangent space preserving the metric) act on this sphere by Möbius transformations. These are the maximal group of complex analytic transformations of $\mathbb{CP}^1$.

Twistor space extends this idea to the whole of Minkowski space. Denoted $\mathbb{PT}$, the twistor space for Minkowski space is complex projective 3-space, $\mathbb{CP}^3$, the space of one-dimensional subspaces of $\mathbb{C}^4$; it is a three-dimensional complex manifold obtained by adding a "plane at infinity" to $\mathbb{C}^3$. Explicitly, we can introduce homogeneous coordinates $Z^\alpha \in \mathbb{C}^4 - \{0\}$ with $\alpha = 0,1,2,3$ but where $Z^\alpha \sim \lambda Z^\alpha$ for $\lambda \in \mathbb{C} - \{0\}$. Affine coordinates on a $\mathbb{C}^3$ chart $Z^3 \neq 0$ can be obtained by setting $(z_1, z_2, \lambda) = (Z^0/Z^3, Z^1/Z^3, Z^2/Z^3)$. Physically, points of twistor space correspond to spinning massless particles in Minkowski space. Mathematically, the correspondence can be understood as the Klein correspondence.

## The Klein Correspondence

The correspondence between $\mathbb{PT}$ and Minkowski space can be extended first to complexified Minkowski space so that the coordinates are allowed to take on values in $\mathbb{C}$, and then to its conformal compactification by including the "light cone at infinity." It then coincides with the classical complex Klein correspondence. The Klein correspondence is the one-to-one correspondence between lines in $\mathbb{CP}^3$ and points of a four complex-dimensional quadric, $\mathbb{CM}$, in $\mathbb{CP}^5$. The 4-quadric $\mathbb{CM}$ can be understood as conformally compactified complexified Minkowski space. Introducing affine coordinates $(z_1, z_2, \lambda)$ on $\mathbb{PT}$ and $(t,x,y,z)$ on $\mathbb{CM}$, we find that a point $(t,x,y,z)$ in $\mathbb{CM}$ corresponds to a line in $\mathbb{PT}$ according to

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} t - z & x + \mathrm{i}y \\ x - \mathrm{i}y & t + z \end{pmatrix} \begin{pmatrix} 1 \\ \lambda \end{pmatrix}$$

Alternatively, fixing $(\lambda, z_1, z_2)$ in these equations gives a 2-plane in complex Minkowski space corresponding to all the lines in $\mathbb{PT}$ through $(\lambda, z_1, z_2)$. Such 2-planes are called "$\alpha$-planes." They are totally null (i.e., the tangent vectors not only have zero length but are also mutually orthogonal) and also self-dual (under the differential geometer's notion of Hodge duality).

This complex correspondence can also be restricted to give correspondences for $\mathbb{R}^4$ with metrics of positive-definite, Euclidean, signature or ultrahyperbolic, $(2,2)$, signature. A particular simplification in Euclidean signature is that the complex $\alpha$-planes intersect the real slice in a point. The conformal compactification of Euclidean $\mathbb{R}^4$ is the 4-sphere $S^4$ given by adding a single point at infinity, and so we have a projection $p : \mathbb{PT} \to S^4$ whose fibers are holomorphically embedded $\mathbb{CP}^1$s. These fibers can be characterized as the lines in $\mathbb{PT}$ that are invariant under a quaternionic complex conjugation which is an antiholomorpic map $\hat{}: \mathbb{PT} \to \mathbb{PT}$ with no fixed points. (Here quaternionic means that on the nonprojective twistor space, $\mathbb{T} = \mathbb{C}^4$, the conjugation has the property $\hat{\hat{Z}} = -Z$ so that it defines a second complex structure anticommuting with the standard one; this is sufficient to express $\mathbb{T} = \mathbb{Q}^2$, where $\mathbb{Q}$ denotes the quaternions. The complex structures $i$, $j$, and $k$ of the quaternions are given by identifying $i$ with $\sqrt{-1}$ on $\mathbb{C}^4$ and $j$ with $\hat{}$ and $k = ij$.)

## The Penrose Transform

A basic task of twistor theory is to transform solutions to the field equations of mathematical physics into objects on twistor space. This works well for linear massless fields such as the Weyl neutrino equation, Maxwell's equations for electromagnetism and linearized gravity. In its general form, this transform has become known as the Penrose transform. Such fields correspond to freely prescribable holomorphic functions $f(\lambda, z_1, z_2)$ (or, more precisely, analytic cohomology classes) on regions of twistor space. The field can be obtained from this function by means of a contour integral. The simplest of these integral formulas is

$$\phi(x^a) = \oint f(\lambda, t - z + \lambda(x + \mathrm{i}y), x - \mathrm{i}y + \lambda(t + z))\mathrm{d}\lambda$$

and differentiation under the integral sign leads easily to the fact that $\phi$ satisfies the wave equation

$$\frac{\partial^2\phi}{\partial t^2} - \frac{\partial^2\phi}{\partial x^2} - \frac{\partial^2\phi}{\partial y^2} - \frac{\partial^2\phi}{\partial z^2} = 0$$

This formula was originally discovered by Bateman. Note that $f$ must have singularities on twistor space to yield a nontrivial $\phi$ and even then, there are many choices of $f$ that yield zero. For a solution $\phi$ defined

over a region $U$ in spacetime, the function $f$ is correctly understood as a representative of a Cech cohomology class defined on the region $U'$ in twistor space swept out by the lines corresponding to points of $U$. Furthermore, the function $f$ should be taken globally to be a function of homogeneity –2, $f(\lambda Z^\alpha) = \lambda^{-2} f(Z^\alpha)$. This formula has generalizations to massless fields of all helicities in which a field of helicity $s$ corresponds to a function (Cech cocycle) of homogeneity degree $2s - 2$.

The Penrose transform has found important applications in representation theory and integral geometry. For a review, the reader is referred to Baston and Eastwood (1989), the relevant survey articles in Bailey and Baston (1990), or Mason and Hughston (1990, chapter 1).

## Twistor Theory and Nonlinear Equations

The Penrose transform for the Maxwell equations and linearized gravity turns out to be linearizations of correspondences for the nonlinear analogs of these equations: the Einstein vacuum equations and the Yang–Mills equations. However, the constructions only work when these fields are anti-self-dual. This is the condition that the curvature 2-forms satisfy $F^* = -\mathrm{i}F$, where $*$ denotes the Hodge dual (which, up to certain factors of $\pm\mathrm{i}$, has the effect of interchanging electric and magnetic fields); it is a nonlinear generalization of the right-handed circular polarization condition. Explicitly, in terms of spacetime indices $a, b, \ldots = 0, 1, 2, 3$, $F^*_{ab} = (1/2)\varepsilon_{abcd}F^{cd}$, where $\varepsilon_{0123} = 1$ and $\varepsilon_{abcd} = \varepsilon_{[abcd]}$. In Minkowski signature, the i factor in the anti-self-duality condition implies that real fields cannot be anti-self-dual. Thus, these extensions are not sufficient to fulfill the ambitions of twistor theory to incorporate real classical nonlinear physics in Minkowski space. However, the factor of i is not present in Euclidean and ultrahyperbolic signature, so the anti-self-duality condition is consistent with real fields in these signatures and this is where the main applications of these constructions have been.

### The Nonlinear Graviton Construction and Its Generalizations

The first nonlinear twistor construction was due to Penrose (1976), and was inspired by Newman's (1976) construction of "heavens" from the infinities of asymptotically flat spacetimes in general relativity.

The nonlinear graviton construction proceeds from the definition of twistors in flat spacetime as $\alpha$-planes in complexified Minkowski space. It is natural to ask which complexified metrics admit a full family of $\alpha$-surfaces, that is, 2-surfaces that are totally null and self-dual. The answer is that a full family of $\alpha$-surfaces exists iff the conformally invariant part of the curvature tensor, the Weyl tensor, is anti-self-dual. If this is the case, twistor space can be defined to be the (necessarily three-dimensional) space of such $\alpha$-surfaces.

The remarkable fact is that the twistor space, together with its complex structure, is sufficient to determine the original spacetime. Twistor space is again a three-dimensional complex manifold, and contains holomorphically embedded rational curves, $\mathbb{CP}^1$s, at least one for each point of the spacetime. However, holomorphic rigidity implies that the family of rational curves is precisely four-dimensional over the complex numbers. Furthermore, incidence of a pair of curves can be taken to imply that the corresponding points in spacetime lie on a null geodesic and this yields a conformal structure on spacetime. Further structures on twistor space can be imposed to give the complex spacetime a metric that is vacuum, perhaps with a cosmological constant. The correspondence is stable under small deformations and so the data defining the twistor space is effectively freely prescribable, see Penrose (1976).

In Euclidean signature, again the complex $\alpha$-planes intersect the real spacetime in a point, so the twistor space again fibers over spacetime. The twistor fibration can be constructed as the projectivized bundle of self-dual spinors or more commonly as the unit sphere bundle in the space of self-dual 2-forms (Atiyah *et al.* 1978). In the latter formulation, the complex structure on the twistor space arises from the direct sum of the naturally defined complex structures on the horizontal and vertical tangent spaces to the bundle; that on the vertical subspace is the standard one on the sphere, and that on the horizontal subspace is a multiple of the self-dual 2-form at the given point of the fiber.

There are now large families of extensions, generalizations, and reductions of this construction. They are all based on the idea of realizing a space with a given complexified geometric structure as the parameter space of a family of holomorphically embedded submanifolds inside a twistor space. In general, the most useful of these constructions are those in which the "spacetime" is obtained as the space of rational curves in a twistor space. This is because the equations that are solved on the corresponding spacetime can be thought of as a completely integrable system in which the integrability condition for the generalized $\alpha$-surfaces is interpreted as the consistency condition of a Lax

pair or more general linear system. For a more detailed discussion from this point of view, see Mason and Woodhouse (1996, chapter 13).

### The Anti-Self-Dual Yang–Mills Equation and Its Twistor Correspondence

The anti-self-dual Yang–Mills equations extend Maxwell's equations for electromagnetism in the right-circularly polarized case. They are a family of equations that depend on a choice of Lie group $G$, usually taken to be a group of complex matrices; Maxwell's equations arise when $G = \mathrm{U}(1)$.

Introduce coordinates $x^a, a = 0, 1, 2, 3$, on $\mathbb{R}^4$ with metric $\mathrm{d}s^2 = \mathrm{d}x^0 \cdot \mathrm{d}x^3 - \mathrm{d}x^1 \cdot \mathrm{d}x^2$ (this is a metric of ultrahyperbolic signature – Euclidean signature can be obtained by choosing the coordinates to be complex, but with $(x^3, -x^2)$ the complex conjugates of $(x^0, x^1)$). The dependent variables are the components $A_a$ of a connection $D_a = \partial_a - A_a$, where $\partial_a = \partial/\partial x^a$ and $A_a = A_a(x^b) \in \mathrm{Lie}\, G$, the Lie algebra of $G$. This connection defines a method of differentiating vector-valued functions $s$ in some representation of $G$. The freedom in changing bases for the vector bundle induce the gauge transformations $A_a \to g^{-1}A_a g - g^{-1}\partial_a g, g(x) \in G$ on $A_a$; two connections that are related by a gauge transformation are deemed to be the same. The self-dual Yang–Mills equations are the condition

$$[D_0, D_2] = [D_1, D_3] = [D_0, D_3] - [D_1, D_2] = 0$$

They are the compatibility conditions

$$[D_0 + \lambda D_1,\, D_2 + \lambda D_3] = 0$$

for the linear system of equations

$$(D_0 - \lambda D_1)s = (D_2 - \lambda D_3)s = 0 \qquad [1]$$

where $\lambda \in C$ and $s$ is an $n$-component column vector. These latter equations form a "Lax pair" for the system.

The Ward (1977) construction provides a one–one correspondence between gauge equivalence classes of solutions of the self-dual Yang–Mills equations and holomorphic vector bundles on regions in twistor space. The key point here is that eqn [1] defines parallel propagation along $\alpha$-planes. To each point $Z$ in twistor space, we can associate the vector space $E_Z$ of solutions to eqn [1] along the corresponding $\alpha$-plane. These vector spaces vary holomorphically with $Z$ and that is what one means by a holomorphic vector bundle $E \to \mathbb{PT}$. The remarkable fact is that the anti-self-dual Yang–Mills field can be reconstructed up to gauge from $E$, and, in effect, for local analytic solutions, $E$ can be represented by freely prescribable "patching" data consisting of local holomorphic matrix-valued functions on twistor space. To construct the solution on spacetime, one must first find a Birkhoff factorization of the patching data on each Riemann sphere in twistor space corresponding to points of the appropriate region in spacetime. On each Riemann sphere, the Birkhoff factorization starts with the given patching function with values in $\mathrm{GL}(n, \mathbb{C})$ on the real axis in the complex plane, and expresses it as a product of functions with values in $\mathrm{GL}(n, \mathbb{C})$ one of which extends over the upper-half plane, and the other over the lower-half complex plane. The anti-self-dual connection can be obtained by differentiating the resulting matrices. See Penrose (1984, 1986), Ward and Wells (1990), or Mason and Woodhouse (1996) for a full discussion, and Atiyah (1979) for the formulation appropriate to Euclidean signature.

## Completely Integrable Systems

In effect, the twistor constructions amount to providing a geometric general local solution to the anti-self-duality equations; the twistor data is, for a local solution, freely prescribable. In this sense, they demonstrate complete integrability of the anti-self-duality equations. The reconstruction of a solution on spacetime from twistor data is not a quadrature – it involves, in the anti-self-dual Yang–Mills case, a Birkhoff factorization (also sometimes referred to as the solution to a Riemann–Hilbert problem), and in the case of the anti-self-dual Einstein equations, the construction of a family of rational curves inside a complex manifold. Nevertheless, such constructions are a familiar part of the apparatus of the theory of integrable systems.

In Ward (1985), this connection with integrable systems was developed further, and the anti-self-dual Yang–Mills equations were shown to yield many important integrable systems under symmetry reduction. Ward's list has been extended and now includes many of the most famous examples of integrable systems such as the Painlevé equations, the Korteweg–de Vries (KdV) equation, the nonlinear Schrödinger equation, the $n$-wave equations, and so on, see Mason and Woodhouse (1996) for a review. There are some notable omissions from the list such as the Kadomtsev–Petviashvili (KP) and Davey–Stewartson equations (at least if one restricts oneself to finite-dimensional gauge groups; reductions using infinite dimensional gauge groups have been obtained).

The list of integrable systems obtainable by symmetry reduction nevertheless remains impressive and provides a route to the classification of at least those integrable systems that can be obtained in this

way. Such systems can be classified by the choice of ingredients required in the symmetry reduction: the gauge group, the group of spacetime symmetries to be reduced by, the choice of Euclidean or ultra-hyperbolic signature, and the choice of certain constants of integration that arise in the reduction.

Another implication is that if an integrable system can be obtained from one of the self-duality equations by symmetry reduction, then it inherits a reduced twistor correspondence because the twistor correspondences share the symmetry groups of the spacetime field equations. These twistor correspondences can be seen to underlie much of the theory of these equations; for example, Backlund transformations of solutions correspond to elementary algebraic operations on the twistor data, similarly the Kac–Moody Lie algebras of hidden symmetries act locally on the twistor data by matrix multiplication of the appropriate loop algebras. Similarly, the inverse-scattering transform for the KdV and non-linear Schrodinger equations can be seen to arise as particular presentations of the twistor construction.

By and large, although twistor methods have yielded new insight into the geometry and structure of systems in dimensions 1 and 2, they have not necessarily superceded pre-existing techniques for constructing solutions and analyzing the solution space. The systems for which twistor methods have been particularly effective for constructing solutions and characterizing their properties are in $2+1$ or higher dimension. Key examples here are of course the anti-self-dual Yang–Mills and Einstein equations themselves, and their single translation reductions. In the anti-self-dual Yang–Mills case, these reductions lead either to Ward's or Manakov and Zakharov's chiral model in Lorentzian signature, $2+1$, or the Bogomolny equations for monopoles, the reduction from Euclidean signature. In both cases, the twistor construction has played a major role in constructing and studying the solitonic solutions.

See Ward and Wells (1990), Mason and Woodhouse (1996), Ward's article in Huggett *et al.* (1998) and the first few chapters of Mason *et al.* (1995), and Mason *et al.* (2001) for more examples of aspects of the theory of integrable systems arising from twistor correspondences.

## Applications to Geometry

These applications are, to a large extent, higher-dimensional analogs of those discussed above; most of the problems in geometry to which twistor theory has been applied are those for which the underlying differential equations are integrable. These start with the Euclidean signature versions of the original Ward construction for anti-self-dual Yang–Mills fields and Penrose's nonlinear graviton construction for Ricci-flat anti-self-dual metrics but, as we will discuss, these constructions have a number of extensions and generalizations.

The first dramatic application of these constructions was the ADHM construction of Yang–Mills instantons. These are absolute minima of the Yang–Mills action, $S[A] = \int \mathrm{tr}(F \wedge F^*)$ on the 4-sphere, $S^4$, with its round metric. A simple argument shows that the action is bounded below by the second Chern class of the bundle and that this bound is achieved only for anti-self-dual fields. Thus, the problem was to characterize all the anti-self-dual Yang–Mills fields on $S^4$. In this Euclidean context, twistor space, $\mathbb{CP}^3$, fibers over $S^4$ and the corresponding Ward vector bundle is a bundle over all of $\mathbb{CP}^3$. It turns out that all such bundles satisfying a certain stability condition had been constructed reasonably explicitly by algebraic geometers. Since the stability condition was implied by the context, this could be turned into an algebraic construction of the general instanton explicit enough to give some insight into both the local and global structure of the solution space. See Atiyah (1979) for a review.

Hitchin used the Euclidean version of the non-linear graviton to develop the theory of gravitational instantons that are asymptotically locally Euclidean (i.e., asymptotically $\mathbb{R}^4/\Gamma$, where $\Gamma$ is a finite subgroup of the rotation group). These were finally constructed by Kronheimer who again used twistor theory to identify the appropriate parameter space, see his article in Mason *et al.* (2001) and Dancer's review of hyper-Kähler manifolds in LeBrun and Wang (1999).

Even in four dimensions, there are a number of variants of the nonlinear graviton construction. The basic twistor correspondence produces a twistor space that is a complex 3-manifold $\mathcal{PT}$ for 4-manifolds with conformal structures whose Weyl tensor is anti-self-dual. There are four natural specializations that have attracted study: (1) the Ricci-flat case, (2) the Einstein case (with nonzero cosmological constant), (3) the scalar-flat Kähler case, and (4) the hypercomplex case.

The twistor space in the Ricci-flat case admits the additional structure of a fibration over $\mathbb{CP}^1$ together with a holomorphic Poisson structure on the fibers with values in the pullback of the 1-forms on $\mathbb{CP}^1$ (alternatively, the bundle of holomorphic 3-forms should be the pullback of the square of the bundle of holomorphic 1-forms on $\mathbb{CP}^1$). The Einstein case with nonzero cosmological constant is a variant of this in which the twistor space admits a

nondegenerate holomorphic contact structure, that is, a distribution of 2-plane elements, which are only integrable when the cosmological constant vanishes. It also admits a Kähler form when the scalar curvature is positive (in the negative case the corresponding Kähler form is indefinite). For the case of Kähler metrics with vanishing scalar curvature, the twistor space admits a holomorphic volume form with a double pole. The Ricci-flat case is equivalent to the case of hyper-Kähler metrics, those that are Kähler with respect to three different complex structures $I, J$, and $K$ satisfying the standard quaternionic relations $IJ = K$, etc. A hypercomplex structure is obtained when one only has the three integrable complex structures satisfying the quaternion relations. Such manifolds admit an underlying conformal structure that is anti-self-dual, and the corresponding twistor space admits a fibration to $\mathbb{CP}^1$.

These constructions have all played a significant role in the general analysis of these geometric structures, and the construction of examples. A striking example of an application of the nonlinear graviton construction to general properties is due to Donaldson and Friedman who show that if two 4-manifolds admit anti-self-dual conformal structures, then their direct sum does also.

In higher dimensions, most generalizations rely on quaternionic geometry and its reductions. The Euclidean signature formulation of the nonlinear graviton construction has natural extensions to quaternionic manifolds in $4k$ dimensions. These are manifolds with metric whose holonomies are contained in $\mathrm{Sp}(k) \times \mathrm{Sp}(1)$. The latter $\mathrm{SP}(1) = \mathrm{SU}(2)$ factor leads to an associated $S^2$ bundle whose total space is the twistor space $\mathcal{PT}$ and it naturally has the structure of a $(2k+1)$-dimensional complex manifold.

For a series of review articles, the reader is referred to Bailey and Baston (1990, chapters 3 and 4) and also LeBrun and Wang (1999, chapters 2, 5, 6, 10, and 14) which, despite being a book on the distinct subject of Einstein manifolds, is strongly influenced by twistor theory. Other applications along these lines are summarized in Mason *et al.* (2001, chapter 1).

There are a number of applications that go beyond complete integrability. A striking application is the twistor framework of Merkulov for studying arbitrary geometric structures. This has led to a classification of all possible irreducible holonomies of torsion-free affine connections, see Merkulov's article in Huggett *et al.* (1998). Another important area is in the field of conformal invariants in which the local twistor connection plays a prominent role. This is a connection that is naturally defined on any conformal manifold being the spinor representation of the Cartan conformal connection. An impressive application here is the construction of conformally invariant differential operators and other conformal invariants. See the article by Baston and Eastwood in Bailey and Baston (1990).

## Beyond Classical Integrability: Twistor-String Theory

Until Witten (2004), there was little indication that twistor theory would have much useful to say about Yang–Mills or gravitational fields that are not anti-self-dual. Furthermore, it was problematic to incorporate quantum field theory into twistor ideas. However, twistor-string theory has transformed the situation and has furthermore had impressive applications to the field of perturbative gauge theory.

The story starts with a formulation by Nair of the remarkable Park–Taylor formulas for the so-called maximal helicity violating (MHV) amplitudes in gauge theory. These are scattering amplitudes at tree level in which helicity conservation is maximally violated; using crossing symmetry to take all the particles to be outgoing, these are amplitudes in which $n-2$ of the particles have helicity –1 and two have helicity +1. These amplitudes can be expressed simply as follows. Let the $n$ particles have color $t_i$ in the Lie algebra of the gauge group and null momenta $p_i$ with spinor decompositions $p_i^a = \tilde{\pi}_i^A \pi_i^{A'}$, $i = 1, \ldots, n$ where the $\pi_i^{A'}$ are self-dual spinors and $\tilde{\pi}^A$ are anti-self-dual spinors using the index notation of Spinors and Spin Coefficients, and Twistors. Let $i = r$ and $i = s$ be the two gluons of helicity +1. Then the coefficient of the colour term $\mathrm{tr}(t_1 t_2 \cdots t_n)$ is

$$\delta^4\left(\sum_{i=1}^n p_i^a\right) \frac{\pi_r \cdot \pi_s}{\Pi_{i=1}^n \pi_i \cdot \pi_{i+1}}$$

where $\pi_i \cdot \pi_j = \pi_i^{A'} \pi_{jA'}$ denotes the standard skew-symmetric inner product on chiral spinors and $\pi_{n+1} = \pi_1$. A striking feature is that, except for the delta function, it is holomorphic in the $\pi_i$s except at the simple poles $\pi_i \cdot \pi_{i+1} = 0$. Nair interprets these poles as those associated to fermion correlators in a current algebra on a $\mathbb{CP}^1$ parametrized by $\pi$. Using a supersymmetric formulation adapted to $N = 4$ super Yang–Mills, he formulated the amplitude as arising from an integral over lines in supertwistor space $\mathbb{CP}^{3|4}$.

Witten extends these ideas to give, at least conjecturally, a complete theory. He proposes that full perturbative $N = 4$ super Yang–Mills theory on spacetime is equivalent to a string theory, a topological

$B$ model, on a supersymmetric version of twistor space, $\mathbb{PT}_s = \mathbb{CP}^{3|4}$. This is the space obtained by taking $\mathbb{C}^{4|4}$ with bosonic coordinates $Z^\alpha, \alpha = 0, \ldots, 3$ and fermionic coordinates $\eta^i, i = 1, \ldots, 4$ moduli the equivalence relation $(Z^\alpha, \eta^i) \sim \lambda(Z^\alpha, \eta^i)$ where $\lambda \in \mathbb{C}, \lambda \neq 0$.

The number 4 here plays two crucial but different roles. It is the maximum number of supersymmetries that Yang–Mills can have; it has the effect of incorporating both the positive and negative helicity parts of the gauge field in the same supermultiplet. It is also the only value of $N$ for which $\mathbb{CP}^{3|N}$ is a Calabi–Yau manifold and this is a necessary condition for the topological twisted $B$ model to be anomaly-free. The Calabi–Yau condition is the condition that the manifold admit a global holomorphic volume form which here is

$$\Omega_s = \varepsilon_{\alpha\beta\gamma\delta} Z^\alpha \mathrm{d}Z^\beta \wedge \mathrm{d}Z^\gamma \wedge \mathrm{d}Z^\delta \wedge \mathrm{d}\eta^1 \wedge \mathrm{d}\eta^2 \wedge \mathrm{d}\eta^3 \wedge \mathrm{d}\eta^4$$

This is invariant under $(Z^\alpha, \eta^i) \to (\lambda Z^\alpha, \lambda\eta^i)$ because $\mathrm{d}(\lambda\eta^i) = \lambda^{-1}\mathrm{d}\eta^i, \lambda \in \mathbb{C}$ follows from the Berezinian rule of integration $\int \theta \mathrm{d}\theta = 1$ for anticommuting variables.

Open-string topological twisted $B$ models are known to correspond to holomorphic Chern–Simons theories on their target space. A holomorphic Chern–Simons theory is a theory whose basic variable is a d-bar operator $\bar\partial_{\mathcal{A}} = \bar\partial + \mathcal{A}$ on a complex vector bundle $E \to \mathbb{PT}^{3|4}$, where $\mathcal{A}$ is a Lie algebra valued $(0,1)$-form on the target space and whose action is

$$S[\mathcal{A}] = \int \frac{1}{2}\left(\mathcal{A}\bar\partial\mathcal{A} + \frac{1}{3}\mathcal{A}^3\right) \wedge \Omega_s$$

The field equations are $\bar\partial_{\mathcal{A}}^2 = 0$. The classical solutions therefore consist of holomorphic vector bundles on the target space, here $\mathbb{CP}^{3|4}$. The twistor-space representation of the fields are obtained by expanding $\mathcal{A}$ in the anticommuting variables $\eta^i$ to obtain

$$\mathcal{A} = a + \eta^i b_i + \eta^i\eta^j c_{ij} + \eta^i\eta^j\eta^k d_{ijk} + \eta^1\eta^2\eta^3\eta^4 g$$

and $a$ has homogeneity zero, but because the homogeneity of $\eta^i$ is of degree 1, $b_i$ has homogeneity degree –1, and so on down to homogeneity degree $-4$ for $g$. Via the Ward construction, the $a$ component corresponds to an anti-self-dual Yang–Mills field on spacetime. The other components of $\mathcal{A}$ can be seen to correspond to spacetime fields with helicities –1/2 to +1 that are background coupled to the anti-self-dual Yang–Mills field.

As it stands, although this holomorphic Chern–Simons theory gives the correct field content of $N = 4$ super Yang–Mills, the couplings are only those of an anti-self-dual sector and more couplings are needed to obtain full $N = 4$ super Yang–Mills. The remarkable fact is that these can be naturally introduced by coupling in certain D1 instantons. The D1 instantons are algebraic curves $C$ in twistor space and the coupling is via a pair of spinor fields $\alpha$ and $\beta$ on $C$ with values in $E$ and $E^*$, respectively with action

$$S[\alpha, \beta, \mathcal{A}] = \int_C \beta\bar\partial_{\mathcal{A}}\alpha$$

This leads to explicit expressions for Yang–Mills scattering amplitudes in terms of integrals of fermion correlators over the moduli spaces of such algebraic curves in supertwistor space. In principle, the integral is over all algebraic curves. However, algebraic curves have two topological invariants, their degree denoted $d$ and genus $g$. An argument based on a classical scaling symmetry gives that integration over just those of curves of degree $d$ gives the subset of processes for which

$$d = q - 1 + l$$

where $q$ is the number of outgoing particles of helicity +1 in the process and $l$ is the number of loops. It is also the case that $g \leq l$.

An elegant formula for the amplitudes is that for the on-shell generating functional for tree-level scattering amplitudes $\mathscr{A}[\mathcal{A}]$, where $\mathcal{A}$ is the on-shell twistor field, being the above-mentioned $(0,1)$-form. The generating functional for processes with $q = d + 1$ external fields of helicity +1 is then

$$\mathscr{A}^d[\mathcal{A}] = \int_{C \in \mathscr{M}^d} \det(\bar\partial + \mathcal{A})|_C \mathrm{d}\mu$$

where $\mathrm{d}\mu$ is a natural measure on the moduli space $\mathscr{M}^d$ of connected rational (genus 0) curves in $\mathbb{CP}^{3|4}$ of degree $d$. This approach has been successfully exploited to obtain implicit algebraic formulas for all tree-level scattering amplitudes.

In an alternative version, the curves of degree $d$ can be taken to be maximally disconnected, being the union of $d$ lines. However, in this approach, we need to also incorporate Chern–Simons propagators which, for tree diagrams, join the lines into a tree. This gives a very flexible calculus for perturbative gauge theory in which scattering processes are obtained by gluing together MHV diagrams. It has been argued that the two formulations are equivalent. On the one hand, the Chern–Simons propagator has a simple pole when the lines meet and the

contour integral over the moduli space can be performed using residues in such a way as to eliminate the Chern–Simons propagators leaving an integral over *d* intersecting lines. On the other hand, the measure on the space of connected curves has a simple pole where the curve acquires double points and again the contour integral can be performed in such a way as to yield the same integral over *d* intersecting lines.

It should be mentioned that Berkovits has given an alternative version of twistor-string theory which is a heterotic open-string theory with target supertwistor space in which the strings are taken to have boundary on the real slice $\mathbb{RP}^3$ in $\mathbb{CP}^3$ (this is appropriate to a spacetime with split signature) and the D1-instanton expansions are replaced by expansions in the fundamental modes of the string (this is not a topological theory). This gives rise to the same formulas for scattering amplitudes as Witten's original model.

There have been many applications now of these ideas, perhaps the most striking being the recursion relations of Britto, Cachazo, Feng, and Witten which give, at tree level, on-shell recurrence relations for Yang–Mills scattering amplitudes that suggests a hitherto unsuspected underlying structure for Yang-Mills theory.

Despite all these successes, twistor-string theory is not thought by string theorists to be a good vehicle for basic physics. The most serious problem is that the closed-string sector gives rise to conformal supergravity which is an unphysical theory. This is particularly pernicious from the point of view of analyzing loop diagrams as from the point of view of string theory, loop diagrams will carry supergravity modes. From this point of view, twistor-string theory is another duality, like AdS-CFT etc., that gives insight into some standard physics but is fundamentally limited.

From the point of view of a twistor theorist, however, twistor-string theory has overcome major obstacles to the twistor programme. Hodges has used the BCFW recursion relations to provide all twistor diagrams for gauge theory. In Mason (2005) it is shown how to derive the main generating function formulas from Yang–Mills and conformal gravity spacetime action principles via a twistor space actions for these theories. These twistor actions can in the first instance be expressed purely bosonically and distinctly and the twistor-string generating function formulas are obtained by expanding and re-summing the classical limit of the path integral in a parameter that expands about the anti-self-dual sector. This allows one to decouple the Yang–Mills and conformal gravity modes, and indeed to work purely bosonically – one is not tied to super Yang–Mills. Although there is much work to be done to extend these ideas to provide a consistent approach to the main equations of basic physics, obstacles that seemed insurmountable a few years ago have been overcome.

*See also:* Chern–Simons Models: Rigorous Results; Einstein Equations: Exact Solutions; General Relativity: Overview; Instantons: Topological Aspects; Integrable Systems and the Inverse Scattering Method; Riemann–Hilbert Methods in Integrable Systems; Spinors and Spin Coefficients; Twistors; Classical Groups and Homogeneous Spaces; Quantum Mechanics: Foundations; Several Complex Variables: Compact Manifolds; Several Complex Variables: Basic Geometric Theory.

## Further Reading

Atiyah MF (1979) *Geometry of Yang–Mills Fields: Lezioni Fermiane*. Pisa: Accademia Nazionale dei Lincei Scuola Normale Superiore.

Atiyah MF, Hitchin NJ, and Singer IM (1978) Self-duality in four-dimensional Riemannian geometry. *Proceedings of the Royal Society A* 362: 425.

Bailey TN and Baston R (eds.) (1990) *Twistors in Mathematics and Physics*, LMS Lecture Notes Series, vol. 156. Cambridge: Cambridge University Press.

Baston RJ and Eastwood MG (1989) *The Penrose Transform: Its Interaction with Representation Theory*. Oxford: Oxford University Press.

Cachazo F, and Svrcek P (2005) Lectures on twistor strings and perturbative Yang–Mills theory, arXiv:hep-th/0504194.

Hitchin N (1987) *Monopoles, Minimal Surfaces and Algebraic Curves, Seminaire de Mathematiques supérieures*, vol. 105. NATO Advanced Study Institute. Les Presses de l'Universite De Montreal.

Huggett S, Mason LJ, Tod KP, Tsou TS, and Woodhouse NMJ (eds.) (1998) *The Geometric Universe*. Oxford: Oxford University Press.

LeBrun C and Wang M (eds.) (1999) *Essays on Einstein manifolds*, Surveys in Differential Geometry, vol. VI. Boston, MA: International Press.

Mason LJ (2005) Twistor actions for non-self-dual fields, a derivation of twistor-string theory, hep-th/0507269.

Mason LJ and Hughston LP (eds.) (1990) *Further Advances in Twistor Theory, Volume I: The Penrose Transform and Its Applications*. Pitman Research Notes in Maths, vol. 231. Harlow: Longman.

Mason LJ, Hughston LP, and Kobak PZ (1995) *Further Advances in Twistor Theory, Volume II: Integrable Systems, Conformal Geometry and Gravitation*. Pitman Research Notes in Maths, vol. 232. Harlow: Longman.

Mason LJ, Hughston LP, Kobak PZ, and Pulverer K (eds.) (2001) *Further Advances in Twistor Theory, Volume III: Curved Twistor Spaces*. Pitman Research Notes in Maths, vol. 424. Boca Raton, FL: Chapman and Hall/CRC Press.

Mason LJ and Woodhouse NMJ (1996) *Twistor Theory, Self-Duality and Twistor Theory*. Oxford: Oxford University Press.

Newman ET (1976) Heaven and its properties. *General Relativity and Gravitation* 7(1): 107–111.

Penrose R (1976) Nonlinear gravitons and curved twistor theory. *General Relativity and Gravitation* 7: 31–52.

Penrose R (1984, 1986) *Spinors and Space-Time*, vols. I and II. Cambridge: Cambridge University Press.
Ward RS (1977) On self-dual gauge fields. *Physics Letters A* 61: 81–82.
Ward RS (1985) Integrable and solvable systems and relations among them. *Philosophical Transactions of the Royal Society A* 315: 451–457.
Ward RS and Wells RO (1990) *Twistor Geometry and Field Theory*. Cambridge: Cambridge University Press.
Witten E (2004) Perturbative gauge theory as a string theory in twistor space. *Communications in Mathematical Physics* 252: 189 (arXiv:hep-th/0312171).

# Twistors

**K P Tod**, University of Oxford, Oxford, UK



## Introduction

Twistor theory initially arose from two principal motivations: a desire for a conformally invariant calculus for spacetime geometry and fields on spacetime, and a desire to unify and account for the various occurrences of complex numbers and holomorphic functions in mathematical physics, especially in general relativity (Penrose and MacCallum 1973). The theory leads to a nonlocal relation between spacetime and twistor space, whereby a point in one is an extended object in the other. Part of the present-day motivation of the subject is that this nonlocal relation will be a fruitful way to approach the quantization of spacetime. A comparison is often invoked with Hamiltonian mechanics, which is a formal rephrasing of classical mechanics that nonetheless provides a bridge from that theory to quantum mechanics. The hope is that the twistor theory has the right character to provide a bridge from general relativity to quantum theory, specifically to quantum gravity.

The principal successes of twistor theory in mathematical physics can be characterized as the linear Penrose transform, which provides a solution of the zero-rest-mass free-field equations in Minkowski space in terms of sheaf cohomology in twistor space, and the nonlinear Penrose transform, which provides solutions of certain nonlinear field equations in terms of holomorphic geometry. These are treated below, together with other applications of twistor theory, following a brief introduction to twistor geometry.

Very recently, there has been a resurgence of interest in twistor theory following Witten's introduction of twistor string theory (Witten 2003) as a string theory in twistor space. This is not treated here, but this article does provide the necessary background.

## Twistor Geometry

General references for this section are the books by Penrose and Rindler (1986) and Hugget and Tod (1994). It will be convenient to use Penrose's abstract index convention (Penrose and Rindler 1984, 1986), which is also used in Spinors and Spin Coefficients. This can be used wherever vector or tensor indices occur. Suppose that $V$ is a (real or complex) finite-dimensional vector space with dual $V'$. Elements of $V$ are written $v^a, u^b, w^c, \ldots$, where an index $a, b, c, \ldots$ is regarded not as an integer in the range 1 to $\dim V$ but simply as an abstract label indicating that the object to which it is attached is a vector. Elements of $V'$ are similarly written $u_a, v_b, w_c, \ldots$ and elements of the tensor algebra as $t^{a\cdots b}{}_{c\cdots d}$ according to valence, and so on. The usual operations of tensor algebra are written in the way that component calculations would suggest, but without necessitating a choice of basis. The jump to tensor fields on a manifold $M$ is immediate. A metric is a particular field $g_{ab}$ and determines a Levi-Civita connection $\nabla_a$ which defines maps $\nabla_a : v^b \to \nabla_a v^b$ and similar for other valences. The virtue of the formalism is that, while remaining invariant, it can harness the strength and flexibility of calculations in components.

With this understanding, twistors may first be defined as the fundamental representation of SU(2, 2), so that they are elements $Z^\alpha$ of a four-dimensional complex vector space $\mathbb{T}$. $\mathbb{T}$ carries a Hermitian form $\Sigma$ of signature $(++--)$ which is made explicit below and which provides an isomorphism from the complex conjugate of $\mathbb{T}$ to its dual. This isomorphism is used to eliminate all appearances of complex-conjugate twistors from the formalism and is therefore regarded as an antilinear map to the dual.

SU(2, 2) is the double cover of $O(2, 4)$, the rotation group of $E_{2,4}$, the six-dimensional space with flat metric $\eta_{2,4}$ of signature $(+---+-)$, which in turn is the double cover of $C(1, 3)$, the conformal group of Minkowski space $M$. This last group homomorphism may be made explicit as follows (suspending the abstract-index convention for the duration of this

aside): introduce pseudo-Cartesian coordinates $x^a = (x^0, x^1, x^2, x^3)$ on $M$ and $y^\alpha = (y^0, y^1, y^2, y^3, y^4, y^5)$ on $E_{2,4}$. The corresponding metrics are

$$\eta_{1,3} = \eta_{ab}\mathrm{d}x^a\mathrm{d}x^b = (\mathrm{d}x^0)^2 - (\mathrm{d}x^1)^2 - (\mathrm{d}x^2)^2 - (\mathrm{d}x^3)^2 \quad [1]$$

$$\eta_{2,4} = \eta_{\alpha\beta}\mathrm{d}y^\alpha\mathrm{d}y^\beta = (\mathrm{d}y^0)^2 - (\mathrm{d}y^1)^2 - (\mathrm{d}y^2)^2 - (\mathrm{d}y^3)^2 + (\mathrm{d}y^4)^2 - (\mathrm{d}y^5)^2 \quad [2]$$

We map $M$ into $E_{2,4}$ by

$$\phi(x^a) = (x^0, x^1, x^2, x^3, (1-\eta)/2, (1+\eta)/2) \quad [3]$$

where $\eta = \eta_{ab}x^a x^b$ with $\eta_{ab}$ as in [1], and it can be checked that $\phi(M)$ is the intersection of the null cone $N$ of the origin in $E_{2,4}$ with the plane $P$ defined by $y^4 + y^5 = 1$. $P$ is in fact a null hyperplane in $E_{2,4}$ and any point of $N$ not on the null hyperplane defined by

$$y^4 + y^5 = 0 \quad [4]$$

can be mapped along the generators of $N$ to a unique point of $P$ (recall that any point on a cone lies on a line through the vertex: these lines are the generators). Thus, the image of $M$ under $\phi$ gives a point on every generator of $N$ except those satisfying [4]. It can also be seen from [2] that the intrinsic metric in $E_{2,4}$ on the intersection of $N$ and $P$ is just $\eta_{1,3}$.

Now let $\mathbb{P}N$ be the projective null cone, or, equivalently, the space of generators of $N$. This is a compact manifold with topology $S^1 \times S^3$, as one can see by intersecting $N$ with the sphere

$$(y^0)^2 + (y^1)^2 + (y^2)^2 + (y^3)^2 + (y^4)^2 + (y^5)^2 = 2$$

Each generator meets this sphere twice at, say, $y^\alpha$ and $-y^\alpha$, and $\mathbb{P}N$ is the quotient by this identification of the two surfaces

$$(y^0)^2 + (y^4)^2 = 1 = (y^1)^2 + (y^2)^2 + (y^3)^2 + (y^5)^2$$

which define the intersection. The metric $\eta_{2,4}$ defines a degenerate metric on $N$, which, however, is nondegenerate on any smooth cross section of $N$ which meets each generator once. Furthermore, the map along the generators between any two such cross sections is conformal. Thus, there is a conformal metric on $\mathbb{P}N$ and it is conformal to $\eta_{1,3}$. We call $\mathbb{P}N$ compactified Minkowski space $M_c$ as it is compact and has the same conformal metric as Minkowski space. It can be thought of as $M$ compactified by the addition of some points, namely the points of $\mathbb{P}N$ corresponding to the generators satisfying [4]. To interpret these, we consider the points satisfying the similar equation $y^4 - y^5 = 0$. By inspection of $\phi$, [3], we see that these points correspond to the light cone of the origin in $M$. Thus, $M_c$ is obtained from $M$ by adding a single light cone, the light cone at infinity known as $\mathcal{I}$ and read as "scri," short for "script-I."

Now the rotation group $O(2,4)$ of $E_{2,4}$ maps $N$ to itself preserving the metric and consequently maps $\mathbb{P}N$ to itself, preserving the conformal metric. Thus, $O(2,4)$ defines conformal transformations of $M_c$ and a count of dimension shows that it is locally isomorphic to the conformal group $C(1,3)$. The map is two-to-one with $\pm I$ in $O(2,4)$ maping to $I$ in $C(1,3)$. The fact that $\mathrm{SU}(2,2)$ is four-to-one homomorphic to $C(1,3)$ follows from calculations below. It is because of this homomorphism of $\mathrm{SU}(2,2)$ and $C(1,3)$ that the geometry and analysis of twistors (i.e., twistor theory) provides a formalism adapted to conformally invariant or conformally covariant notions in $M$ or $M_c$.

A twistor may be expressed in terms of two-component spinors of $\mathrm{SL}(2,\mathbb{C})$, the double cover of the Lorentz group, as follows:

$$Z^\alpha = \left(\omega^A, \pi_{A'}\right) \quad [5]$$

where again indices are abstract, so that

$$\mathbb{T} = \mathbb{S} \oplus \bar{\mathbb{S}}'$$

in terms of the spin space $\mathbb{S}$ and complex-conjugate dual spin space $\bar{\mathbb{S}}'$ of $M$. Now we can write the action of infinitesimal elements of $C(1,3)$ explicitly as

$$\begin{aligned}\widehat{\omega}^A &= \phi^A{}_B\omega^B - \mathrm{i}T^{AA'}\pi_{A'} + \Lambda\omega^A \\ \widehat{\pi}_{A'} &= \bar{\phi}^{B'}_{A'}\pi_{B'} + \mathrm{i}B_{AA'}\omega^A + \Lambda\pi_{A'}\end{aligned} \quad [6]$$

where $T^{AA'}$ (a real vector) defines an infinitesimal translation, $B_{AA'}$ (another real vector) defines an infinitesimal special conformal transformation, $\Lambda$ (a real constant) defines a dilatation and the (real) bivector $M_{ab} = \phi_{AB}\epsilon_{A'B'} + \bar{\phi}_{A'B'}\epsilon_{AB}$ defines an infinitesimal rotation. This gives a total of 15 parameters for the transformation, which is the correct dimension for $C(1,3)$.

The Hermitian form $\Sigma(\,,\,)$ can be written as

$$\Sigma(Z, Z) = Z^\alpha\bar{Z}_\alpha = \omega^A\bar{\pi}_A + \bar{\omega}^{A'}\pi_{A'} \quad [7]$$

when it can be checked that the transformations [6] leave it invariant (and that its signature is $(++--)$; this establishes that $\mathrm{SU}(2,2)$ is locally isomorphic to $C(1,3)$). Equation [7] will be referred to as the norm of a twistor.

From [6], a twistor $Z^\alpha = (\omega^A, \pi_{A'})$ gives rise, under translation by a variable $x^{AA'}$, to a spinor field $\Omega^A$ given by

$$\Omega^A = \omega^A - \mathrm{i}x^{AA'}\pi_{A'} \qquad [8]$$

Differentiating [8] and symmetrizing, we see that $\Omega^A$ satisfies the differential equation

$$\nabla_{A'(A}\Omega_{B)} = 0 \qquad [9]$$

which is known as the twistor equation. In fact, the general solution of [9] takes the form of [8] for constant spinors $\omega^A$ and $\pi_{A'}$. Furthermore, the conformal group can be shown directly to act on solutions of [9], so that twistor theory can begin with the study of [9] and its solutions. In this approach, a twistor is precisely a solution of [9].

Given a spinor field $\Omega^A$ of the form of [8], we may seek the points of $M$ where it vanishes. In general, there are none, but if we consider complexified Minkowski space $\mathbb{C}M$, then $\Omega^A$ vanishes on a two-dimensional complex plane with the property that every tangent vector is of the form $\lambda^A\pi^{A'}$ for varying $\lambda^A$ and fixed $\pi^{A'}$. The 2-plane is flat and totally null, in that the (analytically extended) Minkowski metric vanishes identically on it, and it has a self-dual (SD) tangent bivector determined by $\pi^{A'}$. Such a 2-plane is known as an $\alpha$-plane (reserving the term $\beta$-plane for a totally null 2-plane with anti-self-dual (ASD) tangent bivector). At a given point $p$ in $\mathbb{C}M$, there is an $\alpha$-plane for each choice of $\pi_{A'}$ up to scale (in other words, for each element of the projective (primed) spin space at $p$) which is a copy of the complex projective line, $\mathbb{CP}^1$.

The $\alpha$-plane is determined by the twistor up to scale (in that a constant complex multiple of the field $\Omega^A$ determines the same $\alpha$-plane). Thus, we consider the projective twistor space $\mathbb{PT}$ which, since $\mathbb{T}$ is $\mathbb{C}^4$, is a copy of complex projective 3-space, $\mathbb{CP}^3$. This is now the space of $\alpha$-planes, but is also compact. We define complexified, compactified Minkowski space $\mathbb{C}M_c$ as the space of all (complex projective) lines in $\mathbb{PT}$; then it is easy to see that this includes $\mathbb{C}M$ as an open dense subset. $\mathbb{PT}$ is the space of $\alpha$-planes in $\mathbb{C}M_c$ and two lines meet in $\mathbb{PT}$ iff the corresponding points in $\mathbb{C}M_c$ lie on an $\alpha$-plane, or, equivalently, iff they are null separated. Thus, the conformal structure in $\mathbb{C}M_c$ is determined by incidence of lines in $\mathbb{PT}$.

To find $M$ and $M_c$ in this picture, we seek $\alpha$-planes containing real points. If $\Omega^A$ from [8] vanishes at a real $x^{AA'}$, then the contraction $\omega^A\bar{\pi}_A$ must be purely imaginary, so that, by [7], the norm of the twistor is zero. Conversely, one calculates that $\Omega^A$ can indeed vanish at real points if the norm is zero, and that it will then in fact vanish along a null geodesic with tangent vector (proportional to) $\bar{\pi}^A\pi^{A'}$. Twistors with norm zero are called null and the (five-dimensional, real) submanifold of them in $\mathbb{PT}$ is $\mathbb{PN}$. This is a compactification of the space of (unscaled) null geodesics in $M$ by the inclusion of the 2-sphere of null geodesics in $M_c$ which lie on the light cone at $\mathcal{I}$. For use in the next section, we note the definition of $\mathbb{PT}^+$ and $\mathbb{PT}^-$ as the projective twistors with positive and negative norm, respectively.

To summarize, we have found $M$ and $M_c$: (complex projective) lines in $\mathbb{PT}$ define points of $\mathbb{C}M_c$; lines in $\mathbb{PN}$ define points of $M_c$ with one such, call it $I$, picked out as the vertex of the null cone $\mathcal{I}$; lines in $\mathbb{PN}$ which meet $I$ correspond to points of $\mathcal{I}$; lines in $\mathbb{PN}$ which do not meet $I$ correspond to points in $M$. As for $\mathbb{C}M_c$, the conformal structure of $M$ and $M_c$ is determined by incidence in $\mathbb{PN}$. We may now note the nonlocal correspondence mentioned in the introduction: points in $\mathbb{C}M_c$ are lines in $\mathbb{PT}$ and points in $\mathbb{PT}$ are $\alpha$-planes in $\mathbb{C}M_c$.

It will be convenient to refer to the line in $\mathbb{PT}$ associated with a point $x$ in $\mathbb{C}M_c$ as $L_x$. With this notation, it is possible to characterize the forward or future tube in terms of twistor space: a point $x$ of $\mathbb{C}M$ is in the forward tube iff its imaginary part is timelike and past-pointing, and this is equivalent to $L_x$ lying in $\mathbb{PT}^-$.

The starting point for Riemannian twistor theory is the fact that $\mathbb{CP}^3$ is a fibration with fiber $\mathbb{CP}^1$ over $S^4$, where the fiber above a point $p$ can be interpreted as the almost-complex structures at $p$ (since this is the same as the projective primed spin space at $p$). In the picture developed above, this means that there is an $S^4$'s worth of lines filling out $\mathbb{CP}^3$, no two of which intersect (so that there are no null vectors and the metric is definite). The complexification of $S^4$ with its conformal structure is again $\mathbb{C}M_c$.

If a twistor has nonzero norm, say $Z^\alpha\bar{Z}_\alpha = s \neq 0$, then it can be interpreted as a massless particle with spin $s$: the momentum is $p_a = \bar{\pi}_A\pi_{A'}$ and the angular momentum bivector is $M^{ab} = \mathrm{i}\omega^{(A}\bar{\pi}^{B)}\epsilon^{A'B'} - \mathrm{i}\bar{\omega}^{(A'}\pi^{B')}\epsilon^{AB}$. The angular momentum transforms appropriately under translation by virtue of [6] and the (Pauli–Lubanski) spin vector is $sp_a$, as it should be for a massless spinning particle.

## The Linear Penrose Transform: Zero-Rest-Mass Free Fields

A zero-rest-mass free field of spin $s$ is a symmetric spinor field $\phi_{AB\ldots C}$ with $2s$ indices which satisfies the field equation

$$\nabla^{A'A}\phi_{AB\ldots C} = 0 \qquad [10]$$

The Weyl neutrino equation, source-free Maxwell equation, and linearized Einstein vacuum equation are examples of zero-rest-mass free-field equations, with spins 1/2, 1, and 2, respectively, so that these are equations of physical interest. Conventionally, one takes the $s=0$ case to be the wave equation, and the complex-conjugate fields $\psi_{A'B'\dots C'}$ to have the same spin but opposite helicity.

The conformal group acts on solutions of [10], so that the equations are conformally invariant. The equations can be solved by contour integral expressions involving homogeneous functions of a twistor variable. To be explicit, we define an operation $\rho_x$ of restriction to the line $L_x$ for a function of a twistor variable by the following:

$$\rho_x f(Z^\alpha) = f(\mathrm{i}x^{AA'}\pi_{A'}, \pi_{A'}) \qquad [11]$$

Now suppose that $f(Z^\alpha)$ is holomorphic and homogeneous of degree $-2s-2$ in the twistor variable for positive integer $2s$, but otherwise arbitrary, and consider the integral

$$\psi_{A'B'\dots C'}(x) = \int \pi_{A'}\pi_{B'}\dots\pi_{C'}\rho_x f(Z^\alpha)\epsilon^{E'F'}\pi_{E'}\,\mathrm{d}\pi_{F'} \qquad [12]$$

where there are $2s$ indices on $\psi$ and the integration is around a contour in the line $L_x$ in $\mathbb{PT}$. The choice of homogeneity ensures that the integral is well defined but, to obtain a nonzero answer, $\rho_x f$ must have some singularities as a function of $\pi_{A'}$ on $L_x$. The answer then automatically gives a helicity-$(-s)$ solution of [10], as may be checked by differentiation under the integral sign.

For a helicity-$s$ solution, we take an arbitrary function $f(Z^\alpha)$, holomorphic and of homogeneity $(2s-2)$, and consider the integral

$$\phi_{AB\dots C}(x) = \int \rho_x\left(\frac{\partial}{\partial\omega^A}\frac{\partial}{\partial\omega^B}\cdots\frac{\partial}{\partial\omega^C}f(Z^\alpha)\right)\epsilon^{E'F'}\pi_{E'}\mathrm{d}\pi_{F'} \qquad [13]$$

where there are $2s$ indices on $\phi$ and the integration is again around a contour in the line $L_x$. As before, one needs singularities to make the contour integral nonzero, but again the result satisfies [10].

The correct framework in which to understand these integrals is sheaf cohomology theory. For [12], the functions with singularities are actually elements of $H^1(\hat{\mathcal{U}}, \mathcal{O}(-2s-2))$, the first cohomology group of a region $\hat{\mathcal{U}}$ in $\mathbb{PT}$ with coefficients in the sheaf of germs of holomorphic functions of homogeneity $-2s-2$, while the fields are elements of $H^0(\mathcal{U}, \mathcal{Z}_s)$, the zeroth cohomology group of the corresponding region $\mathcal{U}$ of $M$ with coefficients in helicity-$s$ zero-rest-mass fields (thus, $\hat{\mathcal{U}}$ must contain the neighborhood of lines $L_x$ for points $x$ in $\mathcal{U}$). Similarly, [13] is interpreted cohomologically in terms of potentials modulo a gauge. With appropriate conditions on $\hat{\mathcal{U}}$ and $\mathcal{U}$ (for brevity, $\mathcal{U}$ is said to be elementary), these groups can be shown to be isomorphic and this isomorphism is known as the Penrose transform (Ward and Wells (1991)). A particular instance of an elementary $\mathcal{U}$ is the forward tube, when $\hat{\mathcal{U}}$ is $\mathbb{PT}^-$. Since the definition of positive frequency is holomorphicity on the forward tube, this observation geometrizes the notion of positive frequency in terms of twistor space.

For free fields with mass, there are generalizations of [12] and [13] to solve the Dirac equation for different spins. However, the integrands now involve functions of more than one twistor variable, subject to an equation. This equation is a counterpart of the Klein–Gordon equation and breaks the conformal invariance (as it must, since mass does). It can be imposed by a projection which can in turn be written as a contour integral over arbitrary holomorphic functions. It has been argued that the appropriate description of leptons and hadrons in twistor theory is with functions of two and three twistor variables, respectively. Such a function has two or three integer quantum numbers determined by the homogeneities in different variables, and this leads to a twistor particle classification scheme (see, e.g., Hughston and Sheppard (1980) and Sparling (1981)), similar in many respects to, but not identical with, the standard classifications.

Given that free fields, massive or massless, are determined from arbitrary twistor functions through contour integrals, one may translate the Feynman diagrams of a quantum field theory into contour integrals over twistor functions. In the massless case, the contours are compact, so that the integrals are finite without need for renormalization. The massive case is more complicated but essentially parallel. This is twistor diagam theory and there is a substantial literature on it (see, e.g., the article by Hodges in the volume edited by Huggett *et al.* (1998)). There is currently no new physical theory, distinct from a known quantum field theory, to generate the relevant diagrams.

## The Nonlinear Penrose Transform: Curved Twistor Spaces

The electromagnetic field, in Minkowski space say, can be regarded as a spinor field subject to field equations, in which case these equations can be

solved via the Penrose transform by contour integrals. Alternatively, it can be seen as the curvature of a connection on a U(1) bundle over $M$, which is a more active role for the field in curving a bundle. For SD or ASD electromagnetic fields, there are analogous active twistor constructions. From an ASD electromagnetic field, one may define a connection on the primed spin space of $\mathbb{C}M$ which is flat on $\alpha$-planes: if the tangents to the $\alpha$-plane are of the form $\lambda^A \pi^{A'}$ for varying $\lambda^A$ and with $\pi^{A'}$ fixed up to scale, then consider the propagation of $\pi_{A'}$ around the $\alpha$-plane given by

$$\pi^{A'}(\nabla_{A'A} - \mathrm{i}A_{A'A})\pi_{B'} = 0 \qquad [14]$$

where $A_{A'A}$ is a potential for the electromagnetic field. This connection is flat provided

$$\pi^{A'}\pi^{B'}\nabla_{AA'}A^A_{B'} = 0 \qquad [15]$$

and if this is to hold for all $\pi_{A'}$ then $\nabla_{A(A'}A^A_{B')}$ vanishes and the electromagnetic field, defined as usual as the exterior derivative of the potential, is necessarily ASD. Now the space of $\alpha$-planes in $\mathbb{C}M$ is projective twistor space $\mathbb{PT}$, so we define a holomorphic $\mathbb{C}^*$ bundle $\mathcal{T}$ over $\mathbb{PT}$ by taking the fiber above an $\alpha$-plane to be choices of $\pi_{A'}$ scaled as in [14]. If we restrict attention to the $\alpha$-planes through a given point $p$ of $\mathbb{C}M$, then by comparing the scalings at $p$ we can trivialize the bundle; thus, $\mathcal{T}$ is trivial on lines in $\mathbb{PT}$. There is a converse to this construction and we have: *there is a one-to-one correspondence between holomorphic $\mathbb{C}^*$ bundles on a region $\hat{\mathcal{U}}$ in $\mathbb{PT}$ which are trivial on lines and ASD electromagnetic fields on the corresponding region $\mathcal{U}$ of $\mathbb{C}M$ (for elementary $\mathcal{U}$).*

This construction can be extended to solve the ASD Yang–Mills equations with holomorphic vector bundles replacing holomorphic line bundles: *with $\hat{\mathcal{U}}$ and elementary $\mathcal{U}$ as above, there is a natural one-to-one correspondence between ASD $GL(n,\mathbb{C})$ gauge fields on $\mathcal{U}$ and holomorphic rank-n vector bundles $\mathcal{E}$ over $\hat{\mathcal{U}}$ which are trivial on $L_x$ for every $x$ in $\mathcal{U}$.*

ASD Yang–Mills fields cannot be real on $M$, but using Riemannian twistor theory, one can impose appropriate reality and globality conditions to ensure that these ASD Yang–Mills fields are both real and globally defined on $S^4$. These are then instantons. The Atiyah–Drinfeld–Hitchin–Manin (ADHM) construction of instantons (Atiyah *et al.* 1978) proceeds via construction of the corresponding holomorphic vector bundles over twistor space.

The construction of ASD Yang–Mills fields is also the starting point for the twistor theory of integrable systems (Mason and Woodhouse 1996), following the observation that many of the known completely integrable partial differential equations (PDEs) (including the sine-Gordon, Korteweg–de Vries (KdV) and nonlinear Schrödinger equations) are reductions of the ASD Yang–Mills equations. Solutions of these other integrable systems can be given in terms of a geometrical construction, usually of some structure in holomorphic geometry.

The other major active twistor construction, which historically preceded the Yang–Mills one, is Penrose's nonlinear graviton (Penrose 1976), which solves the ASD Einstein vacuum equations. For this, one starts from a complex, four-dimensional manifold $\mathcal{M}$ with holomorphic metric, vanishing Ricci curvature and ASD Weyl tensor. These conditions on the curvature are necessary and sufficient to allow the existence of $\alpha$-surfaces, which generalize $\alpha$-planes. They are two-dimensional totally null (complex) surfaces with SD tangent bivector, one for each choice of (null) SD bivector, or, equivalently, for each choice of primed spinor, at each point.

The space of $\alpha$-surfaces is a three-dimensional complex manifold, the curved twistor space $\mathbb{P}\mathcal{T}$. This is curved inasmuch as it is not now (part of) $\mathbb{CP}^3$, but it still contains complex projective lines: given a point $p$ in $\mathcal{M}$ there is an $\alpha$-surface through $p$ for every primed spinor at $p$ up to scale; these $\alpha$-surfaces make up a projective line $L_p$ in $\mathbb{P}\mathcal{T}$. The conditions on the curvature are equivalent to the statement that the Levi-Civita connection is flat on primed spinors, so that there exist constant primed spinors in $\mathcal{M}$, and the tangent bivector to an $\alpha$-surface can be taken to be constant, without loss of generality. The map associating a constant primed spinor with each $\alpha$-surface defines a projection $\pi$ from $\mathbb{P}\mathcal{T}$ to $\mathbb{CP}^1$, so that $\mathbb{P}\mathcal{T}$ is a fibration over $\mathbb{CP}^1$. The lines $L_p$ define a four-parameter family of sections of this fibration.

To define the metric of $\mathcal{M}$ from $\mathbb{P}\mathcal{T}$, one needs the notion of normal bundle: the normal bundle of a submanifold $Y$ in a manifold $X$ is $N = TX|_Y/TY$ in terms of the tangent bundles $TX$ and $TY$. The normal bundle $\mathcal{N}_p$ of a particular section $L_p$ is the same in $\mathbb{P}\mathcal{T}$ as it was in $\mathbb{PT}$, namely $H \oplus H$, where $H$ is the hyperplane-section line bundle over $\mathbb{CP}^1$ (Ward and Wells 1991). A section $S_V$ of $\mathcal{N}_p$ corresponds to a vector $V$ in $T_p\mathcal{M}$ (think of it as an infinitesimally neighboring point in $\mathcal{M}$) and $V$ is defined to be null iff $S_V$ has a zero. Because of the nature of $\mathcal{N}$, this defines a quadratic conformal metric, which, furthermore, agrees with the conformal metric on $\mathcal{M}$ and generalizes the definition of conformal metric for $\mathbb{C}M_c$ in terms of incidence in $\mathbb{PT}$. To define the actual metric, as opposed to just the conformal metric, one has a covariant-constant

choice of $\epsilon^{A'B'}$ in $\mathcal{M}$ which defines an $\epsilon$ on the base of the fibration, and a Poisson structure on the fibers $\mu$ of the projection. The definition of $\mu$ is more intricate, but the two structures enable the metric of $\mathcal{M}$ to be recovered from $\mathbb{P}\mathcal{T}$. Penrose (1976) and Huggett and Tod (1994) provide more details.

Now the metric and curvature properties of $\mathcal{M}$ are coded into holomorphic properties of $\mathbb{P}\mathcal{T}$ together with $\epsilon$ and $\mu$. These properties characterize $\mathcal{M}$: *subject to topological conditions on* $\mathcal{M}$, *there is a one-to-one correspondence between holomorphic solutions* $\mathcal{M}$ *of the Einstein vacuum equations with ASD Weyl tensor and three-dimensional complex manifolds* $\mathbb{P}\mathcal{T}$ *fibered over* $\mathbb{CP}^1$, *with a four-parameter of sections, each with normal bundle* $H \oplus H$, *and the forms* $\epsilon$ *and* $\mu$ *as above*.

In fact, one only needs to assume the existence of one section with the correct normal bundle and the full four-parameter family will automatically exist, at least near to the initial one. Penrose (1976) showed how curved twistor spaces with the necessary structures could be obtained by deforming the neighborhood of a line in the "flat" twistor space $\mathbb{PT}$. The Kodaira–Spencer theory of complex deformations ensures that the necessary lines continue to exist under this deformation.

The original nonlinear graviton construction has been extended in various ways including the following: to allow the possibility of a cosmological constant (Ward and Wells 1991); to produce real, Riemannian solutions (Hitchin 1995); to solve other but related field equations (e.g., those for hypercomplex metrics, scalar-flat Kahler metrics or Einstein–Weyl structures).

The search for a twistor construction of the SD Einstein equations (distinct from a construction in terms of dual twistors, which is, of course, provided by deforming dual twistor space) is an active area of research. This and other applications of twistor theory, including a quasilocal definition of mass in general relativity, the classification of affine holonomies and the construction of four-dimensional conformal field theories, may be found in the literature cited in the "Further reading" section.

*See also:* Classical Groups and Homogeneous Spaces; Clifford Algebras and Their Representations; Integrable Systems: Overview; Quantum Field Theory: A Brief Introduction; Quantum Mechanics: Foundations; Relativistic Wave Equations Including Higher Spin Fields; Riemann–Hilbert Problem; Spinors and Spin Coefficients; Twistor Theory: Some Applications.

## Further Reading


Atiyah MF, Hitchin NJ, Drinfeld VG, and Manin YuI (1978) Construction of instantons. *Physics Letters* A 65: 185–187.

Bailey TN and Baston RJ (eds.) (1990) *Twistors in Mathematics and Physics*, LMS Lecture Note Series, vol. 156. Cambridge: Cambridge University Press.

Frauendiener J and Penrose R (2001) Twistors and general relativity. In: Engquist B and Schmid W (eds.) *Mathematics Unlimited – 2001 and Beyond*. Berlin: Springer.

Hitchin NJ (1995) Twistor spaces, Einstein metrics and isomonodromic deformations. *Journal of Differential Geometry*. 42: 30–112.

Hitchin NJ, Segal GB, and Ward RS (1999) *Integrable Systems. Twistors, Loop Groups, and Riemann Surfaces*, Oxford Graduate Texts in Mathematics, vol. 4. Oxford: Oxford University Press.

Huggett SA and Tod KP (1994) *An Introduction to Twistor Theory*, LMS Student Text, vol. 4. Cambridge: Cambridge University Press.

Huggett SA, Mason LJ, Tod KP, Tsou TS, and Woodhouse NMJ (eds.) (1998) *The Geometric Universe – Science, Geometry and the Work of Roger Penrose*. Oxford: Oxford University Press.

Hughston LP and Sheppard M (1980) On the magnetic moments of hadrons. *Reports on Mathematical Physics* 18: 53–66.

Mason LJ, Hughston LP, Kobak PZ, and Pulverer K (eds.) (1995, 1998, 2001) *Further Advances in Twistor Theory,* vols. 1–3. Boca Raton, FL: Pitman Advanced Publishing Programme and Chapman and Hall, CRC.

Mason LJ and Woodhouse NMJ (1996) *Integrability, Self-Duality, and Twistor Theory*, LMS Monographs Series. Oxford: Clarendon.

Penrose R (1976) Nonlinear gravitons and curved twistor theory. *General Relativity and Gravitation* 7: 31–52.

Penrose R (1999) The central programme of twistor theory. *Chaos Solitons Fractals* 10: 581–611.

Penrose R and MacCallum MAH (1973) Twistor theory: an approach to the quantisation of fields and space–time. *Physics Reports* 6C: 241–316.

Penrose R and Rindler W (1984, 1986) *Spinors and Space–Time,* vols. 1 and 2. Cambridge: Cambridge University Press.

Sparling GAJ (1981) Theory of massive particles. I. Algebraic structure. *Philosophical Transactions of the Royal Society of London* A 301: 27–74.

Ward RS and Wells RO (1991) *Twistor Geometry and Field Theory*. Cambridge: Cambridge University Press.

Witten E (2003) Perturbative gauge theory as a string theory in twistor space, hep-th/0312171.

# Two-Dimensional Conformal Field Theory and Vertex Operator Algebras

**M R Gaberdiel**, ETH Zürich, Zürich, Switzerland



## Introduction

For the last twenty years or so, two-dimensional (2D) conformal field theories have played an important role in different areas of modern theoretical physics. One of the main applications of conformal field theory has been in string theory (*see* Compactification of Superstring Theory), where the excitations of the string are described, from the point of view of the world sheet, by a 2D conformal field theory. Conformal field theories have also been studied in the context of statistical physics, since the critical points of second-order phase transition are typically described by a conformal field theory. Finally, conformal field theories are interesting solvable toy models of genuinely interacting quantum field theories.

From an abstract point of view, conformal field theories are (Euclidean) quantum field theories that are characterized by the property that their symmetry group contains, in addition to the Euclidean symmetries, local conformal transformations, that is, transformations that preserve angles but not necessarily lengths. The local conformal symmetry is of special importance in two dimensions since the corresponding symmetry algebra is infinite dimensional in this case. As a consequence, 2D conformal field theories have an infinite number of conserved quantities, and are essentially solvable by symmetry considerations alone. The mathematical formulation of these symmetries has led to the concept of a vertex operator algebra, which has become a new branch of mathematics in its own right. In particular, it has played a major role in the explanation of "monstrous moonshine" for which Richard Borcherds received the Fields medal in 1998.

In the following, we want to explain the main features of conformal field theory using an algebraic approach that will naturally lead to the concept of a vertex operator algebra. There are other approaches to the subject, most notably the formulation, pioneered by Segal, of conformal field theory as a functor from the category of Riemann surfaces to the category of vector spaces. Due to limitations of space, however, we will not be able to discuss any of these other approaches here.

## The Conformal Symmetry Group

The conformal symmetry group of the $n$-dimensional Euclidean space $\mathbb{R}^n$ consists of the (locally defined) transformations that preserve angles but not necessarily lengths. The transformations that preserve angles as well as lengths are the well-known translations and rotations. The conformal group contains (in any dimension) in addition the dilatations or scale transformations

$$x^\mu \mapsto \tilde{x}^\mu = \lambda x^\mu \qquad [1]$$

where $\lambda \in \mathbb{R}$ and $x^\mu \in \mathbb{R}^n$, as well as the so-called special conformal transformations,

$$x^\mu \mapsto \tilde{x}^\mu = \frac{x^\mu + \boldsymbol{x}^2 a^\mu}{1 + 2(\boldsymbol{x}\cdot\boldsymbol{a}) + \boldsymbol{x}^2\boldsymbol{a}^2} \qquad [2]$$

where $a^\mu \in \mathbb{R}^n$ and $\boldsymbol{x}^2 = x^\mu x_\mu$. (Note that this last transformation is only defined for $x^\mu \neq -a^\mu/\boldsymbol{a}^2$.)

If the dimension $n$ of the space $\mathbb{R}^n$ is larger than 2, one can show that the full conformal group is generated by these transformations. For $n=2$, however, the group of (locally defined) conformal transformations is much larger. To see this, it is convenient to introduce complex coordinates for $(x,y)\in\mathbb{R}^2$ by defining $z = x + iy$ and $\bar{z} = x - iy$. Then any (locally) analytic function $f(z)$ defines a conformal transformation by $z \mapsto f(z)$, since analytic maps preserve angles. (Incidentally, the same also applies to $z \mapsto \overline{f(z)}$, but this would reverse the orientation.) Clearly, the group of such transformations is infinite dimensional; this is a special feature of two dimensions.

In this complex notation, the transformations that are generated by translations, rotations, dilatations, and special conformal transformations simply generate the Möbius group of automorphisms of the Riemann sphere

$$z \mapsto f(z) = \frac{az+b}{cz+d} \qquad [3]$$

where $a,b,c,d$ are complex constants with $ad - bc \neq 0$; since rescaling $a,b,c,d$ by a common complex number does not modify [3], the Möbius group is isomorphic to $\mathrm{SL}(2,\mathbb{C})/\mathbb{Z}_2$. In addition to these transformations (that are globally defined on the Riemann sphere), we have an infinite set of infinitesimal transformations generated by $L_n : z \mapsto z + \epsilon z^{n+1}$ for all $n\in\mathbb{Z}$. The generators $L_{\pm1}$ and $L_0$ generate the

subgroup of Möbius transformations, and their commutation relations are simply

$$[L_m, L_n] = (m - n)L_{m+n} \quad [4]$$

In fact, [4] describes also the commutation relations of all generators $L_n$ with $n \in \mathbb{Z}$: this is the Lie algebra of (locally defined) 2D conformal transformations – it is called the Witt algebra.

## The General Structure of Conformal Field Theory

A 2D conformal field theory is determined (like any other field theory) by its space of states and the collection of its correlation functions (vacuum expectation values). The space of states is a vector space $\mathcal{H}$ (which, in many interesting examples, is a Hilbert space), and the correlation functions are defined for collections of vectors in some dense subspace of $\mathcal{H}$. These correlation functions are defined on a 2D (Euclidean) space. We shall mainly be interested in the case where the underlying 2D space is a closed compact surface; the other important case concerning surfaces with boundaries (whose analysis was pioneered by Cardy) will be reviewed elsewhere (*see* the article Boundary Conformal Field Theory). The closed surfaces are classified (topologically) by their genus $g$, which counts the number of handles; the simplest such surface which we shall mainly consider is the sphere with $g = 0$, the surface with $g = 1$ is the torus, etc.

One of the special features of conformal field theory is the fact that the theory is naturally defined on a Riemann surface (or complex curve), that is, on a surface that possesses suitable complex coordinates. In the case of the sphere, the complex coordinates can be taken to be those of the complex plane that cover the sphere except for the point at infinity; complex coordinates around infinity are defined by means of the coordinate function $\gamma(z) = 1/z$ that maps a neighborhood of infinity to a neighborhood of zero. With this choice of complex coordinates, the sphere is usually referred to as the Riemann sphere, and this choice of complex coordinates is, up to Möbius transformations, unique. The correlation functions of a conformal field theory that is defined on the sphere are thus of the form

$$\langle 0|V(\psi_1; z_1, \bar{z}_1) \cdots V(\psi_n; z_n, \bar{z}_n)|0\rangle \quad [5]$$

where $V(\psi, z, \bar{z})$ is the field that is associated to the state $\psi$, and $z_i$ and $\bar{z}_i$ are complex conjugates of one another. Here $|0\rangle$ denotes the $\mathrm{SL}(2, \mathbb{C})/\mathbb{Z}_2$-invariant vacuum. The usual locality assumption of a 2D (bosonic) Euclidean quantum field theory implies that these correlation functions are independent of the order in which the fields appear in [5].

It is conventional to think of $z = 0$ as describing "past infinity," and $z = \infty$ as "future infinity"; this defines a time direction in the Euclidean field theory and thus a quantization scheme (radial quantization). Furthermore, we identify the space of states with the space of "incoming" states; thus, the state $\psi$ is simply

$$\psi = V(\psi; 0, 0)|0\rangle \quad [6]$$

We can think of $z_i$ and $\bar{z}_i$ in [5] as independent variables, that is, we may relax the constraint that $\bar{z}_i$ is the complex conjugate of $z_i$. Then we have two commuting actions of the conformal group on these correlations functions: the infinitesimal action on the $z_i$ variables is described (as before) by the $L_n$ generators, while the generators for the action on the $\bar{z}_i$ variables are $\bar{L}_n$. In a conformal field theory, the space of states $\mathcal{H}$ thus carries two commuting actions of the Witt algebra. The generator $L_0 + \bar{L}_0$ can be identified with the time-translation operator, and thus describes the energy operator. The space of states of the physical theory should have a bounded energy spectrum, and it is thus natural to assume that the spectrum of both $L_0$ and $\bar{L}_0$ is bounded from below; representations with this property are usually called positive-energy representations. It is relatively easy to see that the Witt algebra does not have any unitary positive-energy representations except for the trivial representation. However, as is common in many instances in quantum theory, it possesses many interesting projective representations. These projective representations are conventional representations of the central extension of the Witt algebra

$$[L_m, L_n] = (m - n)L_{m+n} + \frac{c}{12} m(m^2 - 1)\delta_{m,-n} \quad [7]$$

which is the famous Virasoro algebra. Here $c$ is a central element that commutes with all $L_m$; it is called the central charge (or conformal anomaly).

Given the actions of the two Virasoro algebras (that are generated by $L_n$ and $\bar{L}_n$), one can decompose the space of states $\mathcal{H}$ into irreducible representations as

$$\mathcal{H} = \bigoplus_{ij} M_{ij} \mathcal{H}_i \otimes \bar{\mathcal{H}}_j \quad [8]$$

where $\mathcal{H}_i(\bar{\mathcal{H}}_j)$ denotes the irreducible representations of the algebra of $L_n(\bar{L}_n)$, and $M_{ij} \in \mathbb{N}_0$ describe the multiplicities with which these combinations of representations occur. (We are assuming here that the space of states is completely reducible with

respect to the action of the two Virasoro algebras; examples where this is not the case are the so-called logarithmic conformal field theories.) The positive-energy representations of the Virasoro algebra are characterized by the value of the central charge, as well as the lowest eigenvalue of $L_0$; the state $\psi$ whose $L_0$ eigenvalue is smallest is called the highest-weight state, and its eigenvalue $L_0\psi = h\psi$ is the conformal weight. The conformal weight determines the conformal transformation properties of $\psi$: under the conformal transformation $z \mapsto f(z), \bar{z} \mapsto \bar{f}(\bar{z})$, we have

$$V(\psi; z, \bar{z}) \mapsto (f'(z))^h \left(\bar{f}'(\bar{z})\right)^{\bar{h}} V(\psi; f(z), \bar{f}(\bar{z})) \qquad [9]$$

where $L_0\psi = h\psi$ and $\bar{L}_0\psi = \bar{h}\psi$. The corresponding field $V(\psi; z, \bar{z})$ is then called a primary field; if [9] only holds for the Möbius transformations [3], the field is called quasiprimary.

Since $L_m$ with $m > 0$ lowers the conformal weight of a state (see [7]), the highest-weight state $\psi$ is necessarily annihilated by all $L_m$ (and $\bar{L}_m$) with $m > 0$. However, in general the $L_m$ (and $\bar{L}_m$) with $m < 0$ do not annihilate $\psi$; they generate the descendants of $\psi$ that lie in the same representation. Their conformal transformation property is more complicated, but can be deduced from that of the primary state [9], as well as the commutation relations of the Virasoro algebra.

The Möbius symmetry (whose generators annihilate the vacuum) determines the 1-, 2- and 3-point functions of quasiprimary fields up to numerical constants: the 1-point function vanishes, unless $h = \bar{h} = 0$, in which case $\langle 0|V(\psi; z, \bar{z})|0\rangle = C$, independent of $z$ and $\bar{z}$. The 2-point function of $\psi_1$ and $\psi_2$ vanishes unless $h_1 = h_2$ and $\bar{h}_1 = \bar{h}_2$; if the conformal weights agree, it takes the form

$$\langle 0|V(\psi_1; z_1, \bar{z}_1)V(\psi_2; z_2, \bar{z}_2)|0\rangle = C(z_1 - z_2)^{-2h}(\bar{z}_1 - \bar{z}_2)^{-2\bar{h}} \qquad [10]$$

Finally, the structure of the 3-point function of three quasiprimary fields $\psi_1, \psi_2$, and $\psi_3$ is

$$\langle 0|V(\psi_1; z_1, \bar{z}_1)V(\psi_2; z_2, \bar{z}_2)V(\psi_3; z_3, \bar{z}_3)|0\rangle = C\prod_{i<j}(z_i - z_j)^{(h_k - h_i - h_j)}(\bar{z}_i - \bar{z}_j)^{(\bar{h}_k - \bar{h}_i - \bar{h}_j)} \qquad [11]$$

where for each pair $i < j, k$ labels the third field, that is, $k \neq i$ and $k \neq j$. The Möbius symmetry also restricts the higher correlation function of quasiprimary fields: the 4-point function is determined up to an (undetermined) function of the Möbius invariant cross-ratio, and similar statements also hold for $n$-point functions with $n \geq 5$. The full Virasoro symmetry must then be used to restrict these functions further; however, since the generators $L_n$ with $n \leq -2$ do not annihilate the vacuum $|0\rangle$, the Virasoro symmetry leads to Ward identities that cannot be easily evaluated in general. (In typical examples, the Ward identities give rise to differential equations that must be obeyed by the correlation functions.)

## Chiral Fields and Vertex Operator Algebras

The decomposition [8] usually contains a special class of states that transform as the vacuum state with respect to $\bar{L}_m$; these states are the so-called chiral states. (Similarly, the states that transform as the vacuum state with respect to $L_m$ are the antichiral states.) Given the transformation properties described above, it is not difficult to see that the corresponding chiral fields $V(\psi; z, \bar{z})$ only depend on $z$ in any correlation function, that is $V(\psi; z, \bar{z}) \equiv V(\psi, z)$. (Similarly, the antichiral fields only depend on $\bar{z}$.) The chiral fields always contain the field corresponding to the state $L_{-2}|0\rangle$, that describes a specific component of the stress–energy tensor.

In conformal field theory, the product of two fields can be expressed again in terms of the fields of the theory. The conformal symmetry restricts the structure of this operator product expansion:

$$V(\psi_1; z_1, \bar{z}_1)V(\psi_2; z_2, \bar{z}_2) = \sum_i (z_1 - z_2)^{\Delta_i}(\bar{z}_1 - \bar{z}_2)^{\bar{\Delta}_i} \sum_{r,s\geq 0} V(\phi^i_{r,s}; z_2, \bar{z}_2)(z_1 - z_2)^r(\bar{z}_1 - \bar{z}_2)^s \qquad [12]$$

where $\Delta_i$ and $\Delta_{\bar{i}}$ are real numbers, and $r, s \in \mathbb{N}_0$. (Here $i$ labels the conformal representations that appear in the operator-product expansion, while $r$ and $s$ label the different descendants.) The actual form of this expansion (in particular, representations that appear) can be read off from the correlation functions of the theory since the identity [12] has to hold in all correlation functions.

Given that the chiral fields only depend on $z$ in all correlation functions, it is then clear that the operator-product expansion of two chiral fields again only contains chiral fields. Thus, the subspace of chiral fields closes under the operator-product expansion, and therefore defines a consistent (sub)-theory by itself. This subtheory is sometimes referred to as a meromorphic conformal field theory (Goddard 1989). (Obviously, the same also applies to the subtheory of antichiral fields.) The operator-product

expansion defines a product on the space of meromorphic fields. This product involves the complex parameters $z_i$ in a nontrivial way, and therefore does not directly define an algebra structure; it is, however, very similar to an algebra, and is therefore usually called a vertex operator algebra in the mathematical literature. The formal definition involves formal power series calculus and is quite complicated; details can be found in (Frenkel–Lepowski–Meurman 1988).

By virtue of its definition as an identity that holds in arbitrary correlation functions, the operator-product expansion is associative, that is,

$$\begin{aligned}(V(\psi_1;z_1,\bar{z}_1)V(\psi_2;z_2,\bar{z}_2))V(\psi_3;z_3,\bar{z}_3)\\ = V(\psi_1;z_1,\bar{z}_1)(V(\psi_2;z_2,\bar{z}_2)V(\psi_3;z_3,\bar{z}_3))\end{aligned} \quad [13]$$

where the brackets indicate which operator-product expansion is evaluated first. If we consider the case where both $\psi_1$ and $\psi_2$ are meromorphic fields, then the associativity of the operator-product expansion implies that the states in $\mathcal{H}$ form a representation of the vertex operator algebra. The same also holds for the vertex operator algebra associated to the antichiral fields. Thus the meromorphic fields encode in a sense the symmetries of the underlying theory: this symmetry always contains the conformal symmetry (since $L_{-2}\,|\,0\rangle$ is always a chiral field, and $\bar{L}_{-2}\,|\,0\rangle$ always an antichiral field). In general, however, the symmetry may be larger. In order to take full advantage of this symmetry, it is then useful to decompose the full space of states $\mathcal{H}$ not just with respect to the two Virasoro algebras, but rather with respect to the two vertex operator algebras; the structure is again the same as in [8], where, however, each $\mathcal{H}_i$ and $\bar{\mathcal{H}}_j$ is now an irreducible representation of the chiral and antichiral vertex operator algebra, respectively.

## Rational Theories and Zhu's Algebra

Of particular interest are the rational conformal field theories that are characterized by the property that the corresponding vertex operator algebras only possess finitely many irreducible representations. (The name "rational" stems from the fact that the conformal weights and the central charge of these theories are rational numbers.) The simplest example of such rational theories are the so-called minimal models, for which the vertex operator algebra describes just the conformal symmetry: these models exist for a certain discrete set of central charges $c<1$ and were first studied by Belavin, Polyakov, and Zamolodchikov in 1984. (Their paper is contained in the reprint volume of Goddard and Olive (1988).) It was this seminal paper that started many of the modern developments in conformal field theory. Another important class of examples are the Wess–Zumino–Witten (WZW) models that describe the world-sheet theory of strings moving on a compact Lie group. The relevant vertex operator algebra is then generated by the loop group symmetries. There is some evidence that all rational conformal field theories can be obtained from the WZW models by means of two standard constructions, namely by considering cosets and taking orbifolds; thus rational conformal field theory seems to have something of the flavor of (reductive) Lie theory.

Rational theories may be characterized in terms of Zhu's algebra that can be defined as follows. The chiral fields $V(\psi,z)$ that only depend on $z$ must by themselves define local operators; they can therefore be expanded in a Laurent expansion as

$$V(\psi,z)=\sum_{n\in\mathbb{Z}}V_n(\psi)\,z^{-n-h} \quad [14]$$

where $h$ is the conformal weight of the state $\psi$. For example, for the case of the holomorphic component of the stress–energy tensor one finds

$$T(z)=\sum_{n\in\mathbb{Z}}L_n z^{-n-2} \quad [15]$$

where the $L_n$ are the Virasoro generators. By the state/field correspondence [6], it then follows that

$$V_n(\psi)|0\rangle=0 \quad \text{for } n>-h \quad [16]$$

and that

$$V_{-h}(\psi)|0\rangle=\psi \quad [17]$$

(For an example of the above component of the stress–energy tensor, [16] implies that $L_{-1}|0\rangle = L_0\,|0\rangle = L_n|0\rangle = 0$ for $n\geq 0$ – thus the vacuum is in particular $\mathrm{SL}(2,\mathbb{C})/\mathbb{Z}_2$ invariant. Furthermore, [17] shows that $L_{-2}\,|\,0\rangle$ is the state corresponding to this component of the stress–energy tensor.) We denote by $\mathcal{H}_0$ the space of states that can be generated by the action of the modes $V_n(\psi)$ from the vacuum $|0\rangle$. On $\mathcal{H}_0$ we consider the subspace $\mathcal{O}(\mathcal{H}_0)$ that is spanned by the states of the form

$$V^{(N)}(\psi)\chi, \quad N>0 \quad [18]$$

where $V^{(N)}(\psi)$ is defined by

$$V^{(N)}(\psi)=\sum_{n=0}^{h}\binom{h}{n}V_{-n-N}(\psi) \quad [19]$$

and $h$ is the conformal weight of $\psi$. Zhu's algebra is then the quotient space

$$A=\mathcal{H}_0/\mathcal{O}(\mathcal{H}_0) \quad [20]$$

It actually forms an associative algebra, where the algebra structure is defined by

$$\psi \star \chi = V^{(0)}(\psi)\chi \qquad [21]$$

This algebra structure can be identified with the action of the "zero-mode algebra" on an arbitrary highest-weight state.

Zhu's algebra captures much of the structure of the (chiral) conformal field theory: in particular, it was shown by Zhu in 1996 that the irreducible representations of $A$ are in one-to-one correspondence with the representations of the full vertex operator algebra. A conformal field theory is thus rational (in the above, physicists', sense) if Zhu's algebra is finite dimensional. (In the mathematics literature, a vertex operator algebra is usually called rational if in addition every positive-energy representation is completely reducible. It has been conjectured that this is equivalent to the condition that Zhu's algebra is semisimple.)

In practice, the determination of Zhu's algebra is quite complicated, and it is therefore useful to obtain more easily testable conditions for rationality. One of these is the so-called $C_2$ condition of Zhu: a vertex operator algebra is $C_2$-cofinite if the quotient space $\mathcal{H}_0/\mathcal{O}_2(\mathcal{H}_0)$ is finite dimensional, where $\mathcal{O}_2(\mathcal{H}_0)$ is spanned by the vectors of the form

$$V_{-n-h}(\psi)\chi, \quad n \geq 1 \qquad [22]$$

It is easy to show that the $C_2$-cofiniteness condition implies that Zhu's algebra is finite dimensional. Gaberdiel and Neitzke have shown that every $C_2$-cofinite vertex operator algebra has a simple spanning set; this observation can, for example, be used to prove that all the fusion rules (see below) of such a theory are finite.

## Fusion Rules and Verlinde's Formula

As explained above, the correlation function of three primary fields is determined up to an overall constant. One important question is whether or not this constant actually vanishes since this determines the possible "couplings" of the theory. This information is encoded in the so-called fusion rules of the theory. More precisely, the fusion rules $N_{ij}{}^k \in \mathbb{N}_0$ determine the multiplicity with which the representation of the vertex operator algebra labeled by $k$ appears in the operator-product expansion of the two representations labeled by $i$ and $j$.

In 1988, Verlinde found a remarkable relation between the fusion rules of a vertex operator algebra and the modular transformation properties of its characters. To each irreducible representation $\mathcal{H}_i$ of a vertex operator algebra, one can define the character

$$\chi_i(\tau) = \mathrm{tr}_{\mathcal{H}_i}\left(q^{L_0-(c/24)}\right), \quad q = \mathrm{e}^{2\pi \mathrm{i}\tau} \qquad [23]$$

For rational vertex operator algebras (in the mathematical sense) these characters transform under the modular transformation $\tau \mapsto -1/\tau$ as

$$\chi(-1/\tau) = \sum_j S_{ij}\chi_j(\tau) \qquad [24]$$

where $S_{ij}$ are constant matrices. Verlinde's formula then states that, at least for unitary theories,

$$N_{ij}{}^k = \sum_l \frac{S_{il}\, S_{jl}\, S^*_{kl}}{S_{0l}} \qquad [25]$$

where the "0" label denotes the vacuum representation. A general argument for this formula has been given by Moore and Seiberg in 1989; very recently, this has been made more precise by Huang.

## Modular Invariance and the Conformal Bootstrap

Up to now, we have only considered conformal field theories on the sphere. In order for the theory to be well defined also on higher-genus surfaces, it is believed that the only additional requirement comes from the consistency of the torus amplitudes. In particular, the vacuum torus amplitude must only depend on the equivalence class of tori that is described by the modular parameter $\tau \in \mathbb{H}$, up to the discrete identifications that are generated by the usual action of the modular group $\mathrm{SL}(2,\mathbb{Z})$ on the upper half-plane $\mathbb{H}$. For the theory with decomposition [8] this requires that the function

$$Z(\tau,\bar{\tau}) = \sum_{ij} M_{ij}\chi_i(\tau)\chi_j(\bar{\tau}) \qquad [26]$$

is invariant under the action of $\mathrm{SL}(2,\mathbb{Z})$. This is a very powerful constraint on the multiplicity matrices $M_{ij}$ that has been analyzed for various vertex operator algebras. For example, Cappelli, Itzykson, and Zuber have shown that the modular invariant WZW models corresponding to the group SU(2) have an A–D–E classification. The case of SU(3) was solved by Gannon, using the Galois symmetries of these rational conformal field theories.

The condition of modular invariance is relatively easily testable, but it does not, by itself, guarantee that a given space of states $\mathcal{H}$ comes from a consistent conformal field theory. In order to construct a consistent conformal field theory, one needs to solve the conformal bootstrap, that is, one has to determine

all the normalization constants of the correlators so that the resulting set of correlators is local and factorizes appropriately into 3-point correlators (crossing symmetry). This is typically a difficult problem which has only been solved explicitly for rather few theories, for example, the minimal models. Recently, it has been noticed that the conformal bootstrap can be more easily solved for the corresponding boundary conformal field theory. Furthermore, Fuchs, Runkel, and Schweigert have shown that any solution of the boundary problem induces an associated solution for conformal field theory on surfaces without boundary. This construction relies heavily on the relation between 2D conformal field theory and 3D topological field theory (Turaev 1994).

*See also:* Boundary Conformal Field Theory; Compactification of Superstring Theory; Current Algebra; Knot Theory and Physics; String Field Theory; Superstring Theories; Symmetries in Quantum Field Theory of Lower Spacetime Dimensions.

## Further Reading


Di Francesco P, Mathieu P, and Sénéchal D (1997) *Conformal Field Theory*. New York: Springer.

Frenkel I, Lepowski J, and Meurman A (1988) *Vertex Operator Algebras and the Monster*. Boston, MA: Academic Press.

Gaberdiel MR (2000) An introduction to conformal field theory. *Reports on Progress in Physics* 63: 607 (arXiv:hep-th/9910156).

Gannon T (1999) Monstrous moonshine and the classification of CFT, arXiv:math.QA/9906167.

Gannon T (2006) *Moonshine beyond the Monster: The Bridge Connecting Algebra, Modular Forms and Physics* (to appear). Cambridge: Cambridge University Press.

Gawedzki K (1999) Lectures on conformal field theory. In: *Quantum Fields and Strings: A Course for Mathematicians*, vol. 2. Providence, RI: American Mathematical Society.

Ginsparg P (1988) *Applied Conformal Field Theory*. Lectures Given at the Les Houches Summer School in Theoretical Physics. Elsevier.

Goddard P (1989) Meromorphic conformal field theory. *Infinite Dimensional Lie Algebras and Lie Groups: Proceedings of the CIRM Luminy Conference, 1988, 556*. Singapore: World Scientific.

Goddard P and Olive DI (1988) *Kac–Moody and Virasoro Algebras, A Reprint Volume for Physicists*. Singapore: World Scientific.

Kac VG (1998) *Vertex Algebras for Beginners*. Providence, RI: American Mathematical Society.

Pressley A and Segal GB (1986) *Loop Groups*. Oxford: Clarendon.

Schweigert C, Fuchs J, and Walcher J (2000) Conformal field theory, boundary conditions and applications to string theory, arXiv:hep-th/0011109.

Turaev VG (1994) *Quantum Invariants of Knots and 3-Manifolds*. Berlin: de Gruyter.


# Two-Dimensional Ising Model

**B M McCoy**, State University of New York at Stony Brook, Stony Brook, NY, USA



## Introduction

The Ising model is a model of a classical ferromagnet on a lattice first introduced in 1925 in the one-dimensional case by E Ising. At each lattice site there is a "spin" variable $\sigma$, which takes on the values +1 (spin up) and –1 (spin down). The mutual interaction energy of the pair of spins $\sigma_\alpha$ and $\sigma_{\alpha'}$, where $\alpha$ and $\alpha'$ are nearest neighbors, is $-E(\alpha, \alpha')$ if $\sigma_\alpha = \sigma_{\alpha'}$ and is $E(\alpha, \alpha')$ if $\sigma_\alpha = -\sigma_{\alpha'}$. In addition, the spins can interact with an external magnetic field as $-H\sigma_\alpha$. On a square lattice, where $j$ specifies the row and $k$ specifies the column, the interaction energy for the homogeneous case where $E_{\rm v}(\alpha, \alpha')$ and $E_{\rm h}(\alpha, \alpha')$ are independent of the position $\alpha, \alpha'$ may be explicitly written as

$$E(H) = -\sum_{j,k}[E_{\rm h}\sigma_{j,k}\sigma_{j,k+1} + E_{\rm v}\sigma_{j,k}\sigma_{j+1,k} + H\sigma_{j,k}] \quad [1]$$

This very simple model [1] has the remarkable property that in two dimensions at $H = 0$ many properties of physical interest can be computed exactly. Furthermore, the model has a ferromagnetic phase transition at a critical temperature $T_{\rm c}$, at which the specific heat diverges and the magnetic susceptibility diverges to infinity and below which there is a nonzero spontaneous magnetization. In addition, the microscopic correlations between spins can also be exactly computed. These exact calculations are the basis of the modern theory of second-order phase transitions used to analyze real ferromagnets and real fluids near their critical points in both two and three dimensions. The model may also be interpreted as a lattice gauge theory.

## Solvability

The solvability of the Ising model at $H = 0$ was discovered by Onsager in 1944 in one of the most profound and inventive papers ever written in mathematical physics. Onsager discovered that the model possesses an infinite-dimensional symmetry, which allowed him to exactly compute the free

energy per site. This symmetry is generated by the relations

$$
\begin{aligned}
[A_l, A_m] &= 4G_{l-m} \\
[G_l, A_m] &= 2A_{m+l} - 2A_{m-l} \\
[G_l, G_m] &= 0
\end{aligned}
\qquad [2]
$$

This algebra of Onsager is a subalgebra of what is now called the loop algebra of the Lie algebra $Sl_2$ and it is the first infinite-dimensional algebra to be used in physics.

In the 60 years since Onsager first computed the free energy, several other methods of exact solution have been found. In 1949, Kaufman reduced the computation of the free energy to a problem of free fermions. A closely related combinatorial method was invented by Kac and Ward, Hurst and Green, and by Kastelyn. Baxter (1982) has computed the free energy by means of star triangle equations and functional equations in his book.

The fermionic and the combinatorial methods are powerful enough to compute the correlation functions but are not generalizable to other models. The functional equation methods of Baxter generalize to many other important models but they do not give correlation functions. There are still aspects of Onsager's method that remain unexplored.

The free energy per site in the thermodynamic limit is defined as

$$F = -k_{\mathrm{B}}T \lim_{\mathcal{N}\to\infty} \mathcal{N}^{-1} \ln Z_I(H) \qquad [3]$$

where $\mathcal{N}$ is the total number of sites of the lattice and the partition function $Z_I(H)$ is defined as

$$Z_I(H) = \sum_{\text{all } \sigma=\pm 1} \mathrm{e}^{E(H)/k_{\mathrm{B}}T} \qquad [4]$$

with the sum being over all values $\sigma_{j,k} = \pm 1$ and $k_{\mathrm{B}}$ is Boltzmann's constant. The result of Onsager is that, at $H=0$,

$$
\begin{aligned}
F/k_{\mathrm{B}}T = \ln 2 + \frac{1}{8\pi^2}\int_0^{2\pi} \mathrm{d}\theta_1 \int_0^{2\pi} \mathrm{d}\theta_2 \ln \Big[ &\cosh 2E_{\mathrm{h}}/k_{\mathrm{B}}T \\
&\times \cosh 2E_{\mathrm{v}}/k_{\mathrm{B}}T - \sinh 2E_{\mathrm{h}}/k_{\mathrm{B}}T \cos\theta_1 \\
&- \sinh 2E_{\mathrm{v}}/k_{\mathrm{B}}T \cos\theta_2 \Big]
\end{aligned}
\qquad [5]
$$

This free energy has a singularity at a temperature $T_{\mathrm{c}}$ defined from

$$\sinh(2E_{\mathrm{v}}/k_{\mathrm{B}}T_{\mathrm{c}}) \sinh(2E_{\mathrm{h}}/k_{\mathrm{B}}T_{\mathrm{c}}) = 1 \qquad [6]$$

and near $T_{\mathrm{c}}$ the specific heat diverges as

$$
\begin{aligned}
c \sim -\frac{2}{k_{\mathrm{B}}T^2\pi} \Big( &E_{\mathrm{h}}^2 \sinh^2 2E_{\mathrm{v}}/k_{\mathrm{B}}T_{\mathrm{c}} + 2E_{\mathrm{v}}E_{\mathrm{h}} \\
&+ E_{\mathrm{v}}^2 \sinh^2 2E_{\mathrm{h}}/k_{\mathrm{B}}T_{\mathrm{c}} \Big) \ln|1 - T/T_{\mathrm{c}}|
\end{aligned}
\qquad [7]
$$

The next property to be computed was the spontaneous magnetization, which is usually defined as

$$M_- = \lim_{H\to 0^+} M(H) \qquad [8]$$

However, because solution is only available at $H=0$, this definition cannot be used and instead $M_-$ is computed from an alternative definition in terms of the spin–spin correlation function

$$\langle\sigma_{0,0}\sigma_{M,N}\rangle = \frac{1}{Z_I(0)} \sum_{\sigma=\pm 1} \sigma_{0,0}\sigma_{M,N} \mathrm{e}^{-E(0)/k_{\mathrm{B}}T} \qquad [9]$$

as

$$M_-^2 = \lim_{M^2+N^2\to\infty} \langle\sigma_{0,0}\sigma_{M.N}\rangle \qquad [10]$$

The result for $M_-$, first announced by Onsager in 1949, is

$$M_- = \begin{cases} (1-k^2)^{1/8} & \text{for } T \le T_{\mathrm{c}} \\ 0 & \text{for } T > T_{\mathrm{c}} \end{cases} \qquad [11]$$

where

$$k = (\sinh 2E_{\mathrm{v}}/k_{\mathrm{B}}T \sinh 2E_{\mathrm{h}}/k_{\mathrm{B}}T)^{-1} \qquad [12]$$

A key point in the computation of the magnetization [11] from [9] is that the spin–spin correlation function can be written as a determinant. In fact, there are many such different, but equal, determinental representations and the size of the smallest one in general is $2(|M| + |N|)$. The simplest case is the diagonal correlation

$$\langle\sigma_{0,0}\sigma_{N,N}\rangle = \begin{vmatrix} a_0 & a_{-1} & a_{-2} & \cdots & a_{1-N} \\ a_1 & a_0 & a_{-1} & \cdots & a_{2-N} \\ a_2 & a_1 & a_0 & \cdots & a_{3-N} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_{N-1} & a_{N-2} & a_{N-3} & \cdots & a_0 \end{vmatrix} \qquad [13]$$

where

$$a_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{d}\theta \mathrm{e}^{-\mathrm{i}n\theta} \left[\frac{1-k\mathrm{e}^{-\mathrm{i}\theta}}{1-k\mathrm{e}^{\mathrm{i}\theta}}\right]^{1/2} \qquad [14]$$

Determinants of the form [13], where the elements on each diagonal are equal, are called Toeplitz.

The study of the spin–spin correlations of the Ising model provides a microscopic picture of the behavior of the ferromagnet near the phase transition temperature $T_{\mathrm{c}}$, and an entire branch of mathematics has developed from the study of the behavior of Toeplitz determinants when the size is large. The first such mathematical advance was the discovery by Szegö of a general formula for the limit as $N \to \infty$, from which the magnetization [11] is computed.

The simplest result for the approach to this $N \to \infty$ limit is the behavior of the diagonal correlation at $T = T_c(k = 1)$, where [13] exactly reduces to

$$<\sigma_{0,0}\sigma_{N,N}> = \left(\frac{2}{\pi}\right)^N \prod_{l=1}^{N-1}\left[1 - \frac{1}{4l^2}\right]^{l-N} \qquad [15]$$

which behaves as $N \to \infty$ as

$$<\sigma_{0,0}\sigma_{N,N}> \sim AN^{-1/4} \qquad [16]$$

where $A \sim 0.6450\cdots$ is a transcendental constant. Further results for large $N$ and $T$ fixed are for $T < T_c(k < 1)$,

$$<\sigma_{0,0}\sigma_{N,N}> \sim M_-{}^2\left\{1 + \frac{2k^{2(N+1)}}{\pi N^2 (k^{-2} - 1)^2} + \cdots\right\} \qquad [17]$$

and for $T > T_c(k > 1)$,

$$<\sigma_{0,0}\sigma_{N,N}> \sim \frac{k^{-N}}{(\pi N)^{1/2}(1 - k^{-2})^{1/4}} + \cdots \qquad [18]$$

By comparing [16] with [17] and [18], we see that at $T = T_c$ the correlations decay algebraically but for $T \neq T_c$ the decay is exponential. It is useful to write the exponential in [17] for $T < T_c$ as

$$k^{-N} = \mathrm{e}^{-N/\xi_-} \quad \text{with } \xi^{-1} = -\ln k \qquad [19]$$

and in [18] for $T > T_c$ as

$$k^N = \mathrm{e}^{-N/\xi_+} \quad \text{with } \xi^{-1} = \ln k \qquad [20]$$

The quantity $\xi$ is called the correlation length and as $T \to T_c$ the correlation length diverges as

$$\xi_\pm \sim |1 - k|^{-1} = \text{const.}\ |T - T_c|^{-1} \qquad [21]$$

A more profound property of the correlations is that they satisfy differential and difference equations. It was found by Jimbo and Miwa (1980) that the diagonal correlation function satisfies the nonlinear differential equation related to the sixth Painlevé function

$$\left(t(t-1)\frac{\mathrm{d}^2\sigma}{\mathrm{d}t^2}\right)^2 = N^2\left((t-1)\frac{\mathrm{d}\sigma}{\mathrm{d}t} - \sigma\right)^2 - 4\frac{\mathrm{d}\sigma}{\mathrm{d}t}\left((t-1)\frac{\mathrm{d}\sigma}{\mathrm{d}t} - \sigma - \frac{1}{4}\right)\left(t\frac{\mathrm{d}\sigma}{\mathrm{d}t} - \sigma\right) \qquad [22]$$

where for $T < T_c$ we set $t = k^{-2}$ and

$$\sigma_N(t) = t(t-1)\frac{\mathrm{d}}{\mathrm{d}t}\ln <\sigma_{0,0}\sigma_{N,N}> - \frac{1}{4} \qquad [23]$$

and for $T > T_c$ we set $t = k^2$ and

$$\sigma_N(t) = t(t-1)\frac{\mathrm{d}}{\mathrm{d}t}\ln <\sigma_{0,0}\sigma_{N,N}> - \frac{t}{4} \qquad [24]$$

Furthermore it was found by McCoy *et al.* (1981) that for a given temperature the general two spin correlation function and all multipoint correlations satisfy quadratic nonlinear partial difference equations in the locations of the spins.

## Scaling Theory

It is evident that the results [17] and [18] do not reduce to [16] when $k \to 1$. Therefore, in order to uniformly characterize the behavior of the correlation function in the critical region near $T_c$, it is necessary to introduce what is called the scaling function. This uniform expansion is obtained by introducing a scaled length defined as

$$r = N/\xi \qquad [25]$$

and considering the joint (scaling limit) where

$$N \to \infty \quad \text{and} \quad T \to \infty \quad \text{with } r \text{ fixed} \qquad [26]$$

We define the scaled correlation function as

$$G_\pm(r) = \lim_{\text{scaling}} M_\pm^{-2} <\sigma_{0,0}\sigma_{N,N}> \qquad [27]$$

where the subscript $\pm$ means that the limit is taken from $T > T_c$ or $T < T_c$, respectively, $M_-$ is the spontaneous magnetization [11] and

$$M_+ = (k^{-2} - 1)^{1/8} \qquad [28]$$

This concept of the scaling limit and scaling function is very general and can be defined for any system with a critical point that has an order parameter like $M_-$ that vanishes at $T_c$ and a correlation length that diverges at $T_c$. However, the Ising model has the further remarkable property discovered by Wu *et al.* (1976) that the scaled correlation function may be explicitly expressed in terms of a function which satisfies an ordinary nonlinear differential equation. Specifically,

$$G_\pm(r) = \frac{1}{2}[1 \mp \eta(r/2)]\eta(r/2)^{-1/2} \times \exp\int_{r/2}^{\infty} r'\frac{r'}{4}\eta^{-2}[(1-\eta^2)^2 - (\eta')^2] \qquad [29]$$

where the function $\eta(r)$ satisfies the Painlevé III equation

$$\eta'' = \frac{1}{\eta}(\eta')^2 - \frac{\eta'}{r} + \eta^3 - \eta^{-1} \qquad [30]$$

with the boundary condition that

$$\eta(r) \sim 1 - 2\lambda K_0(2r) \quad \text{as } r \to \infty \qquad [31]$$

where $K_0(r)$ is the modified Bessel function of the third kind and

$$\lambda = 1/\pi \qquad [32]$$

The leading behavior of $G_{\pm}(r)$ for $r \to \infty$ is

$$G_{+}(r) \sim \lambda K_0(r) \qquad [33]$$

$$G_{-}(r) \sim 1 + \lambda^2 \left\{ r^2[K_1^2(r) - K_0^2(r)] - rK_0(r)K_1(r) + \tfrac{1}{2}K_0^2(r) \right\} \qquad [34]$$

where $K_n(z)$ is the modified Bessel function of the third kind. When $\lambda$ is given by [32] these $r \to \infty$ limits of $G_{\pm}(r)$ agree with the behavior of $<\sigma_{0,0}\sigma_{N,N}>$ for $N \gg 1$ and $|T - T_c|$ small with $N|T - T_c| \gg 1$ which is obtained from [18] and [17]. The behavior of $G_{\pm}(r)$ for $r \to 0$ with the value of $\lambda$ given by [32] is

$$G_{\pm}(r) = \text{const.}\, r^{-1/4} \qquad [35]$$

where the constant agrees with that computed from the result [16] for $<\sigma_{0,0}\sigma_{N,N}>$ at $T = T_c$ for $N \gg 1$. For other values of the boundary condition constant $\lambda$, the scaling function $G_{\pm}(r)$ diverges with a power which differs from 1/4. The computation of the constant in [35] requires the evaluation of a nontrivial integral involving the Painlevé III function.

The agreement of the limits $r \to \infty$ and $r \to 0$ of the function $G_{\pm}(r)$ with the lattice results near $T_c$ means that this scaling function uniformly interpolates between $T \neq T_c$ and $T = T_c$ and that the lattice size (defined here as unity) and the self-generated correlation length $\xi$ are the only two length scales in the theory. This feature that the system generates only one new length scale near $T_c$ is referred to as one length scale scaling.

## Susceptibility

The final quantity of macroscopic thermodynamic interest is the magnetic susceptibility

$$\chi(T) = \left.\frac{\partial M(H)}{\partial H}\right|_{H=0} \qquad [36]$$

which is expressed in terms of the spin–spin correlation function as

$$\chi(T) = \frac{1}{k_B T}\sum_{M,N}\{<\sigma_{0,0}\sigma_{M,N}> - M_{-}^2\} \qquad [37]$$

The susceptibility may be studied by using the determinental expression for the correlation function. The simplest result is obtained (for the isotropic case, $E_v = E_h$) by using the scaling form [27] to find for $T \sim T_c$ that

$$k_B T \chi_{\pm}(T) \sim M_{\pm}^2 \xi^2 2\pi \int_0^\infty \mathrm{d}r\, r\{G_{\pm} - \Theta_{\pm}\} \qquad [38]$$

where $\Theta_{+} = 0$ and $\Theta_{-} = 1$. and thus $\chi_{\pm}(T)$ diverges at $T \to T_c$ as

$$\chi_{\pm}(T) \sim C_{\pm}|T - T_c|^{-7/4} \qquad [39]$$

where $C_{\pm}$ are transcendental constants given as integrals over the scaling function $G_{\pm}(r)$, which were first evaluated by Barouch *et al.* in 1973 as

$$C_{-} = 0.0255369719\ldots, \quad C_{+} = 0.9625817322\ldots \qquad [40]$$

## Critical-Exponent Phenomenology

From the behavior for the Ising model of the specific heat, magnetization, susceptibility, correlation length, and the correlation at $T_c$ given above we abstract for general systems the phenomenological critical-exponent parametrization for $T \to T_c\pm$ of

$$c \sim A_C^{\pm}|T - T_c|^{-\alpha_{\pm}} \qquad [41]$$

$$M \sim A_M|T_c - T|^{\beta} \qquad [42]$$

$$\chi \sim A_{\chi}^{\pm}|T - T_c|^{-\gamma_{\pm}} \qquad [43]$$

$$\xi \sim A_{\xi}^{\pm}|T - T_c|^{-\nu_{\pm}} \qquad [44]$$

and at $T = T_c$ for $R \to \infty$

$$<\sigma_0\sigma_R> \sim A_{\sigma}/R^{d-2+\eta} \quad \text{where } d \text{ is the dimension} \qquad [45]$$

The exponents $\alpha_{\pm}, \gamma_{\pm}, \nu_{\pm}$ above and below $T_c$ are usually found to be equal, and the exponent $\eta$ is usually called the anomalous dimension. If it is assumed that the scaling function [27] exists and that one length scale scaling holds then the exponents are related by what are called scaling laws, such as

$$2\beta = \nu_{-}(d - 2 + \eta) \qquad [46]$$

$$\alpha_{-} + 2\beta - \gamma_{-} = 2 \qquad [47]$$

$$d\nu_{-} = 2 - \alpha_{-} \qquad [48]$$

Thus, from the properties of the Ising model near $T_c$, we have obtained a phenomenology for use on all systems near the critical point.

## Fuchsian Equations and Natural Boundaries for Susceptibility

This critical phenomenology, however, has not taken into account the fact that the susceptibility is a much more complicated function than either the spontaneous magnetization [11] or the free energy [5], which have only isolated singularities at $k^2=1$, and that there is more structure to the susceptibility than the singularity of [39].

For arbitrary $T$, the susceptibility was shown by Wu, McCoy, Tracy, and Barouch to be expressible in the form

$$\chi_{\pm}(T) = M_{\pm}^2 \sum_j \tilde{\chi}^{(j)}(T) \qquad [49]$$

where in the sum $j$ is odd (even) for $T$ above (below) $T_c$. The quantities $\tilde{\chi}^{(j)}(T)$ are explicitly given as $j$-fold integrals of algebraic functions and thus will satisfy linear differential equations with polynomial coefficients. Such functions can have only isolated singularities. The function $\tilde{\chi}^{(1)}(T)$ is elementary and has a double pole at $T_c$ and $\tilde{\chi}^{(2)}(T)$ is given in terms of complete elliptic integrals. Quite recently, remarkable Fuchsian linear differential equations for $\tilde{\chi}^{(3)}(T)$ and $\tilde{\chi}^{(4)}(T)$ of seventh and tenth orders, respectively, have been obtained by Zenine, Boukraa, Hassani, and Maillard for the isotropic lattice.

Furthermore, it was shown by Orrick *et al.* (2001) that $\tilde{\chi}^{(j)}$ has singularities in the complex $T$ plane at

$$\begin{aligned} &\cosh(2E_v/kT)\cosh(2E_h/kT) \\ &\quad - \sinh(2E_v/kT)\cos(2\pi/j) \\ &\quad - \sinh(2E_h/kT)\cos(2\pi m'/j) = 0 \end{aligned} \qquad [50]$$

with $m, m' = 1, 2, \ldots, j$. The form of the singularity in $\tilde{\chi}^{(j)}(T)$ for $T > T_c$ is as

$$\epsilon^{(j^2-3)/2} \ln \epsilon \qquad [51]$$

and, for $T < T_c$, it is as

$$\epsilon^{(j^2-3)/2} \qquad [52]$$

where $\epsilon$ measures the deviation from the singular point [50]. These singularities become dense as $j \to \infty$ and, therefore, the singularity at $T = T_c$ is not isolated and instead the critical point is embedded in a natural boundary. Such a function cannot satisfy a linear differential equation of finite order with polynomial coefficients.

The existence of the natural boundary in the susceptibility is a new phenomenon which is not seen in either the free energy or magnetization and leads to the speculation that in the presence of a magnetic field the one length scale scaling property of the model at $H = 0$ may fail. If this proves to be correct, there will be physical effects which are not incorporated in the phenomenological scaling theory of critical phenomena.

## Impure Ising Models

The Ising model may also be studied when the interaction energies at sites $j, k$ are not chosen to be independent of position but are allowed to vary from site to site. When these interactions are chosen randomly out of some probability distribution, this is a model of a ferromagnet with frozen (quenched) impurities. All real systems will be impure to some extent, so the study of such dirty systems is of great practical importance.

The special case where the interactions are translationally invariant in the horizontal direction but are allowed to vary in a layered fashion from row to row was introduced by McCoy and Wu in 1968 and found to be dramatically different from the pure Ising model described above. In particular, what is a critical temperature $T_c$ in the pure case is now spread out into a region bounded by the temperatures the pure model would be critical if all the bonds took on the minimum or maximum value allowed by the probability distribution. In this new region, the correlations (in the direction of translational invariance) are found to decay as a power law which depends on the temperature; the specific heat is never infinite but the susceptibility is infinite in an entire temperature region that includes the temperature at which the spontaneous magnetization first appears as $T$ is lowered. The existence of this new region for Ising models with a general randomness in two and three dimensions has been demonstrated by Griffiths. More recently, this effect has been reinterpreted in terms of impurities in quantum spin chains.

## Quantum Field Theory

The Ising model of [1] may be reinterpreted as a two-dimensional lattice gauge theory of the gauge field

$s_{j+1/2,k} = \pm 1$

on the vertical link between $(j, k)$ and $(j+1, k)$

$s_{j,k+1/2} = \pm 1$

on the horizontal link between $(j, k)$

and $(j, k+1)$ [53]

and a “Higgs” field

$$\phi_{j,k} = \pm 1 \quad \text{on the site } (j, k) \qquad [54]$$

with the action

$$S_g = -E_g \sum_{j,k} s_{j+1/2,k}s_{j+1,k+1/2}s_{j+1/2,k+1}s_{j,k+1/2} - E_h \sum_{j,k} (\phi_{j,k}s_{j+1/2,k}\phi_{j+1,k} + \phi_{j,k}s_{j,k+1/2}\phi_{j,k+1}) \qquad [55]$$

If we define

$$z_{g,h} = \tanh E_{g,h}/k_B T \qquad [56]$$

the partition function of the gauge theory is expressed in terms of the Ising model partition function as

$$Z_g = [8\cosh(E_g/k_B T)\cosh^2 \times (E_h/k_B T) z_g^{1/2} z_h]^{\mathcal{N}} Z_I(H) \qquad [57]$$

where we make the identification

$$H/k_B T = \tfrac{1}{2}\ln z_g \quad \text{and} \quad E/k_B T = \tfrac{1}{2}\ln z_h \qquad [58]$$

This identification may be extended to correlation functions. Of particular interest for the gauge theory is the plaquette–plaquette correlation $< P_{0,0}P_{j,l} >$, where

$$P_{j,k} = s_{j+1/2,k}s_{j+1,k+1/2}s_{j+1/2,k+1}s_{j,k+1/2} \qquad [59]$$

which is expressed in terms of the Ising correlations at $H \neq 0$ as

$$< P_{0,0}P_{j,k} > - < P_{0,0} >^2 = \sinh^2(2H/k_B T)(< \sigma_{0,0}\sigma_{j,k} > - < \sigma_{0,0} >^2) \qquad [60]$$

To study this correlation further, we need to study the correlations of the Ising model in nonzero magnetic field. This has been done by McCoy and Wu in the scaling limit $H \to 0, T \to T_c$ with

$$h = \frac{H}{|T - T_c|^{15/8}} \text{ fixed} \qquad [61]$$

for $T < T_c$, where it is found that the scaling function $G(r,h)$ for small $h$ and large $r$ if

$$G(r,h) \sim \sum_l ahK_0\left[(2 + h^{2/3}\lambda_l)r\right] \sim \pi^{1/2} r^{-1/2} e^{-2r} \sum_l hae^{-rh^{2/3}\lambda_l} \qquad [62]$$

where $\lambda_l$ are the solutions of

$$J_{1/3}\left(\tfrac{1}{3}\lambda^{3/2}\right) + J_{-1/3}\left(\tfrac{1}{3}\lambda^{3/2}\right) = 0 \qquad [63]$$

with $J_n(z)$ the Bessel function of order $n$ and $K_0(z)$ the modified Bessel function of the third kind.

A field theory is said to possess a particle spectrum if the Fourier transform of the two-point function

$$G(k,h) = \int d^2 r e^{ik\cdot r} G(r,h) \qquad [64]$$

has poles of the form $A_l/(k^2 + m_l^2)$, where $m_l$ is the mass of the $l$th particle. If we note that the Fourier transform of $K_0(r)$ is

$$\int d^2 r e^{ik\cdot r} K_0(r) = \frac{2\pi}{k^2 + 1} \qquad [65]$$

we see that the Fourier transform of [62] is the sum of an infinite number of poles. This is to be compared with the Fourier transform of the scaled correlation function $G_-(r)$ at $H = 0$ and $T < T_c$ [34], which does not contain any poles at all and may instead be interpreted as having a two-particle cut. This phenomenon of a cut at $h = 0$ breaking up into an infinite number of poles for $h > 0$ is a signal that at $h = 0$ the theory has free unconfined two-particle states which become weakly confined by a linear confining potential for $h > 0$. This confinement is thought to be a characteristic of most gauge theories.

*See also:* Eight Vertex and Hard Hexagon Models; Holonomic Quantum Fields; Painlevé Equations; Percolation Theory; Phase Transitions in Continuous Systems; Statistical Mechanics and Combinatorial Problems; Toeplitz Determinants and Statistical Mechanics; Yang–Baxter Equations.

## Further Reading


Barouch E, McCoy BM, and Wu TT (1973) Zero-field susceptibility of the two dimensional Ising model near $T_c$. *Physical Review Letters* 31: 1409–1411.

Baxter RJ (1982) *Exactly Solved Models in Statistical Mechanics*. London: Academic Press.

Griffiths RB (1969) Nonanalytic behavior above the critical point in a random Ising ferromagnet. *Physical Review Letters* 23: 17–19.

Jimbo M and Miwa T (1980) Studies on holonomic quantum fields. XVII. *Proceedings of the Japanese Academy* 56A: 405; (1981) 57A: 347.

Kasteleyn PW (1963) Dimer statistics and phase transitions. *Journal of Mathematical Physics* 4: 287–293.

McCoy BM (1969) Incompleteness of the critical exponent description for ferromagnetic systems containing random impurities. *Physical Review Letters* 23: 383–388.

McCoy BM (1995) The connection between statistical mechanics and quantum field theory. In: Bazhanov VV and Burden CJ (eds.) *Statistical Mechanics and Field Theory*, pp. 26–128. Singapore: World Scientific.

McCoy BM and Wu TT (1968) Theory of a two dimensional Ising model with random impurities. *Physical Review* 176: 631–643.

McCoy BM and Wu TT (1973) *The Two Dimensional Ising Model*. Cambridge: Harvard University Press.

McCoy BM and Wu TT (1978) Two dimensional Ising field theory in a magnetic field: Breakup of the cut in the 2-point function. *Physical Review D* 18: 1259–1267.

McCoy BM, Perk JHH, and Wu TT (1981) Ising field theory: quadratic difference equations for the *n*-point Green's functions on the lattice. *Physical Review Letters* 46: 757.

Montroll EW, Potts RB, and Ward JC (1963) Correlations and spontaneous magnetization of the two dimensional Ising model. *Journal of Mathematical Physics* 4: 308–322.

Onsager L (1944) Crystal statistics I. A two dimensional model with an order disorder transition. *Physical Review* 65: 117–149.
Orrick WP, Nickel BG, Guttmann AJ, and Perk JHH (2001) Critical behavior of the two dimensional Ising susceptibility. *Physical Review Letters* 86: 4120–4123.
Wu TT, McCoy BM, Tracy CA, and Barouch E (1976) Spin–spin correlation functions for the two dimensional Ising model, exact theory in the scaling region. *Physical Review B* 13: 315–374.
Zenine N, Boukraa S, Hassani S, and Maillard JM (2004) The Fuchsian differential equation of the square lattice Ising model $\chi^{(3)}$ susceptibility. *Journal of Physics A* 37: 9651–9668.
Zenine N, Boukraa S, Hassani S, and Maillard JM (2005) Ising model susceptibility: Fuchsian differential equation for $\chi^{(4)}$ and its factorization properties. *Journal of Physics A* 38: 4149–4173.

# Two-Dimensional Models

**B Schroer**, Freie Universität Berlin, Berlin, Germany



## History and Motivation

Local quantum physics of systems with infinitely many interacting degrees of freedom leads to situations whose understanding often requires new physical intuition and mathematical concepts beyond that acquired in quantum mechanics and perturbative constructions in quantum field theory. In this situation, two-dimensional soluble models turned out to play an important role. On the one hand, they illustrate new concepts and sometimes remove misconceptions in an area where new physical intuition is still in the process of being formed. On the other hand, rigorously soluble models confirm that the underlying physical postulates are mathematically consistent, a task which for interacting systems with infinite degrees of freedom is mostly beyond the capability of pedestrian methods or brute force application of hard analysis on models whose natural invariances have been mutilated by a cutoff.

In order to underline these points and motivate the interest in two-dimensional QFT, let us briefly look at the history, in particular at the physical significance of the three oldest two-dimensional models of relevance for statistical mechanics and relativistic particle physics, in chronological order: the Lenz–Ising (L–I) model, Jordan's model of bosonization/fermionization, and the Schwinger model ($QED_2$). (A more detailed account of the changeful history concerning their correct physical interpretation and generalizations to higher dimensions of these models and the increasing conceptual role of low-dimensional models in QFT can be found in Schroer (2005).)

The L–I model was proposed in 1920 by Wihelm Lenz (see Lenz (1925)) as the simplest discrete statistical mechanics model with a chance to go beyond the P Weiss phenomenological ansatz involving long-range forces and instead explain ferromagnetism in terms of nonmagnetic short-range interactions. Its one-dimensional version was solved four years later by his student Ernst Ising. Its changeful history reached a temporary conceptual climax when Onsager succeeded to rigorously establish a second-order phase transition in two dimensions.

Another conceptually rich model which lay dormant for almost two decades as a result of a misleading speculative higher-dimensional generalization by its protagonist is the bosonization/fermionization model first proposed by Jordan (1937). This model establishes a certain equivalence between massless two-dimensional fermions and bosons and is related to Thirring's massless 4-fermion coupling model and also to Luttinger's one-dimensional model of an electron gas (Schroer). One reason why even nowadays hardly anybody knows Jordan's contribution is certainly the ambitious but unfortunate title "the neutrino theory of light" under which he published a series of papers.

Both discoveries demonstrate the usefulness of having controllable low-dimensional models; at the same time, their complicated history also illustrates the danger of rushing to premature "intuitive" conclusions about extensions to higher dimensions.

A review of the early historical benchmarks of conceptual progress through the study of solvable two-dimensional models would be incomplete without mentioning Schwinger's (1962) proposed solution of two-dimensional quantum electrodynamics, afterwards referred to as the Schwinger model. He used this model in order to argue that gauge theories are not necessarily tied to zero-mass vector particles. Some work was necessary (Schroer) to unravel its physical content with the result that the would-be charge of that $QED_2$ model was "screened" and its apparent chiral symmetry broken; in other words, the model exists only in the so-called Schwinger–Higgs phase with massive free scalar particles accounting for its physical content. Another closely related aspect of

this model which also arose in the Lagrangian setting of four-dimensional gauge theories was that of the $\theta$-angle parametrizing, an ambiguity in the quantization.

A coherent and systematic attempt at a mathematical control of two-dimensional models came in the wake of Wightman's first rigorous programmatic formulation of QFT (Schroer 2005). This formulation stayed close to the physical ideas underlying the impressive success of renormalized QED perturbation theory, although it avoided the direct use of Lagrangian quantization. The early attempts towards a "constructive QFT" found their successful realization in two-dimensional QFT (the $P\varphi_2$ models (Glimm and Jaffe 1987)); the restriction to low dimensions is related to the mild short-distance singularity behavior (super-renormalizability) which these methods require. We will focus our main attention on alternative constructive methods which, even though not suffering from such short-distance restrictions, also suffer from a lack of mathematical control in higher spacetime dimensions; the illustration of the constructive power of these new methods comes presently from massless $d=1+1$ conformal and chiral QFT as well as from massive factorizing models.

There are several books and review articles (Furlan *et al.* 1989, Ginsparg 1990, Di Francesco *et al.* 1996) on $d=1+1$ conformal as well as on massive factorizing models (Abdalla *et al.* 1991). To the extent that concepts and mathematical structures are used which permit no extension to higher dimensions (Kac–Moody algebras, loop groups, integrability, presence of an infinite number of conservation laws), this line of approach will not be followed in this article since our primary interest will be the use of two-dimensional models of QFT as "theoretical laboratories" of general QFT. Our aim is twofold; on the one hand, we intend to illustrate known principles of general QFT in a mathematically controllable context and on the other hand, we want to identify new concepts whose adaptation to QFT in $d=1+1$ lead to their solvability (Schroer).

## General Concepts and Their Two-Dimensional Manifestation

The general framework of QFT, to which the rich world of controllable two-dimensional models contributes as an important testing ground, exists in two quite different but nevertheless closely related formulations: the 1956 approach in terms of pointlike covariant fields due to Wightman (see Streater and Wightman (1964)) (*see* Axiomatic Quantum Field Theory), and the more algebraic setting which can be traced back to ideas which Haag (1992) developed shortly after and which are based on spacetime-indexed operator algebras and related concepts which developed over a long period of time, with contributions of many other authors to what is now referred to as algebraic QFT (AQFT) or simply local quantum physics (LQP). Whereas the Wightman approach aims directly at the (not necessarily observable) quantum fields, the operator-algebraic setting (*see* Algebraic Approach to Quantum Field Theory) is more ambitious. It starts from physically well-motivated assumptions about the algebraic structure of local observables and aims at the reconstruction of the full field theory (including the operators carrying the superselected charges) in the spirit of a local representation theory of (the assumed structure of the) local observables. This has the advantage that the somewhat mysterious concept of an inner symmetry (as opposed to outer (spacetime) symmetry) can be traced back to its physical roots which is the representation-theoretical structure of the local observable algebra (*see* Symmetries in Quantum Field Theory of Lower Spacetime Dimensions). In the standard Lagrangian quantization approach, the inner symmetry is part of the input (multiplicity indices of field components on which subgroups of U($n$) or O($n$) act linearly) and hence it is not possible to problematize this fundamental question. When in low-dimensional spacetime dimensions the sharp separation (the Coleman–Mandula theorems) of inner versus outer symmetry becomes blurred as a result of the appearance of braid group statistics, the standard Lagrangian quantization setting of most of the textbooks is inappropriate and even the Wightman framework has to be extended. In that case, the algebraic approach is the most appropriate.

The important physical principles which are shared between the Wightman approach (see Streater and Wightman (1964)) and the operator algebra (AQFT) setting (Haag 1992) are the spacelike locality or Einstein causality (in terms of pointlike fields or algebras localized in causally disjoint regions) and the existence of positive-energy representations of the Poincaré group implementing covariance and the stability of matter. In the algebraic approach, the observable content of the theory is encoded into a family of (weakly closed) operator algebras $\{\mathcal{A}(\mathcal{O})\}_{\mathcal{O}\in\mathcal{K}}$ indexed by a family of convex causally closed spacetime regions $\mathcal{O}$ (with $\mathcal{O}'$ denoting the spacelike complement and $\mathcal{A}'$ the von Neumann commutant) which act in one common Hilbert space. Covariant local fields lose their distinguished role

which they have in the classical setting and which (via Lagrangian quantization) was at least partially inherited by the Wightman approach and, apart in their role as local generators of symmetries (conserved currents), became mere "field coordinatizations" of local algebras. (There is a denumerable set of such pointlike field generators which form a local equivalence (Borchers) class of fields and in the absence of interactions permits a neat description in terms of Wick-ordered free-field polynomials (Haag 1992). Certain properties cannot be naturally formulated in the pointlike field setting (e.g., Haag duality for convex regions $\mathcal{A}(\mathcal{O}') = \mathcal{A}(\mathcal{O})'$), but apart from those properties the two formulations are quite close; in particular for two-dimensional theories there are convincing arguments that one can pass between the two without imposing additional technical requirements. (Haag duality holds for observable algebras in the vacuum sector in the sense that any violation can be explained in terms of a spontaneously broken symmetry; in local theories, it can always be enforced by dualization and the resulting Haag dual algebra has a charge superselection structure associated with the unbroken subgroup.) Haag duality is the statement that the commutant of observables not only contains the algebra of the causal complement that is, $\mathcal{A}(\mathcal{O}') \subset \mathcal{A}(\mathcal{O})'$ (Einstein causality) but is even exhausted by it; it is deeply connected to the measurement process and its violation in the vacuum sector for convex causally complete regions signals spontaneous symmetry breaking in the associated charge-carrying field algebra (Haag 1992). It can always be enforced (assuming that the wedge-localized algebras fulfill [1] below) by symmetry-reducing extension called Haag dualization. Its violation for multilocal region reveals the charge content of the model via charge–anticharge splitting in the neutral observable algebra (Schroer).

Another physically important property which has a natural algebraic formulation is the split property: for regions $\mathcal{O}_i$ separated by a finite spacelike distance, one finds $\mathcal{A}(\mathcal{O}_1 \cup \mathcal{O}_2) \simeq \mathcal{A}(\mathcal{O}_1) \otimes \mathcal{A}(\mathcal{O}_2)$ which can be derived from the Buchholz–Wichmann "nuclearity property" (Haag 1992) (an appropriate adaptation of the "finiteness of phase-space cell" property of QM to QFT). Related to the Haag duality is the local version of the "time slice property" (the QFT counterpart of the classical causal dependency property) sometimes referred to as "strong Einstein causality" $\mathcal{A}(\mathcal{O}'') = \mathcal{A}(\mathcal{O})''$.

One of the most astonishing achievements of the algebraic approach (which justifies its emphasis on properties of "local observables") is the DHR theory of superselection sectors (Doplicher *et al.* 1971), that is, the realization that the structure of charged (nonvacuum) representations (with the superposition principle being valid only within one representation) and the spacetime properties of the generating fields which are the carriers of these generalized charges (including their spacelike commutation relations which lead to the particle statistics and also to their internal symmetry properties) are already encoded in the structure of the Einstein causal observable algebra (Symmetries in Quantum Field Theory: Algebraic Aspects). The intuitive basis of this remarkable result (whose prerequisite is locality) is that one can generate charged sectors by spatially separating charges in the vacuum (neutral) sector and disposing of the unwanted charges at spatial infinity (Haag 1992).

An important concept which especially in $d = 1 + 1$ has considerable constructive clout is "modular localization." It is a consequence of the above algebraic setting if either the net of algebras have pointlike field generators, or if the one-particle masses are separated by spectral gaps so that the formalism of time-dependent scattering can be applied (Schroer 2005); in conformal theories, this property holds automatically in all spacetime dimensions. It rests on the basic observation (Tomita–Takesaki Modular Theory) that a standard pair $(\mathcal{A}, \Omega)$ of a von Neumann operator algebra and a standard vector (standardness means that the operator algebra of the pair $(\mathcal{A}, \Omega)$ acts cyclic and separating on the vector $\Omega$) gives rise to a Tomita operator $S$ through its star-operation whose polar decomposition yield two modular objects, a one-parametric subgroup $\Delta^{it}$ of the unitary group of operators in Hilbert space whose Ad-action defines the modular automorphism of $(\mathcal{A}, \Omega)$ whereas the angular part $J$ is the modular conjugation which maps $\mathcal{A}$ into its commutant $\mathcal{A}'$

$$\begin{aligned} &SA\Omega = A^*\Omega,\ S = J\Delta^{1/2} \\ &J_W = U(j_W) = S_{\mathrm{scat}}J_0, \Delta_W^{it} = U(\Lambda_W(2\pi t)) \\ &\sigma_W(t) := \mathrm{Ad}\Delta_W^{it} \end{aligned} \qquad [1]$$

The standardness assumption is always satisfied for any field-theoretic pair $(A(\mathcal{O}), \Omega)$ of a $O$-localized algebra and the vacuum state (as long as $\mathcal{O}$ has a nontrivial causal disjoint $\mathcal{O}'$), but it is only for the wedge region $W$ that the modular objects have a physical interpretation in terms of the global symmetry group of the vacuum as specified in the second line of [1]; the modular unitary $\Delta_W^{it}$ represents the $W$-associated boost $\Lambda_W(\chi)$ and the modular conjugation $J_W$ implements the TCP-like reflection along the edge of the wedge (Bisognano

and Wichmann 1975). The third line is the definition of the modular group. The importance of this theory for local quantum physics results from the fact that it leads to the concept of modular localization, an intrinsic new scenario for field-theoretic constructions which is different from the Lagrangian quantization schemes (Schroer 2005).

A special feature of $d = 1 + 1$ Minkowski spacetime is the disconnectedness of the right/left spacelike region leading to a right–left ordering structure. So in addition to the Lorentz-invariant timelike ordering $x \prec y$ ($x$ earlier than $y$, which is independent of spacetime dimensions), there is an invariant spacelike ordering $x < y$ ($x$ to the left of $y$) in $d = 1 + 1$ which opens the possibility of more general Lorentz-invariant spacelike commutation relations than those implemented by Bose/Fermi fields (Rehren and Schroer 1987) of fields with a spacelike braid group commutation structure. The appearance of such exotic statistics fields is not compatible with their Fourier transforms being creation/annihilation operators for Wigner particles; rather, the state vectors which they generate from the vacuum contain in addition to the one-particle contribution a vacuum polarization cloud (Schroer 2005). This close connection between new kinematic possibilities and interactions is one of the reasons why, (different from higher dimensions where interactions are prescribed by the recipe of local couplings of free fields) low-dimensional QFT offers a more intrinsic access to the central issue of interactions.

## Boson/Fermion Equivalence and Superselection Theory in a Special Model

The simplest and oldest but conceptually still rich model is obtained, as first proposed by Jordan (1937), by using a two-dimensional massless Dirac current and showing that it may be expressed in terms of scalar canonical Bose creation/annihilation operators

$$j_\mu =: \overline{\psi}\gamma_\mu\psi := \partial_\mu\phi,\ \phi := \int_{-\infty}^{+\infty} \{e^{ipx}a^*(p) + h.c.\}\frac{dp}{2|p|} \qquad [2]$$

Although the potential $\phi(x)$ of the current as a result of its infrared divergence is not a field in the standard sense of an operator-valued distribution in the Fock space of the $a(p)^{\#}$ (It becomes an operator after smearing with test functions whose Fourier transform vanishes at $p = 0$), the formal exponential defined as the zero-mass limit of a well-defined exponential free massive field

$$: e^{i\alpha\phi(x)} := \lim_{m\to 0} m^{\alpha^2/2} : e^{i}\alpha\phi_m(x) : \qquad [3]$$

turns out to be a *bona fide* quantum field in a larger Hilbert space (which extends the Fock space generated from applying currents to the vacuum). The power in front is determined by the requirement that all Wightman functions (computed with the help of free-field Wick combinatorics) stay finite in this massless limit; the necessary and sufficient condition for this is the charge conservation rule

$$\left\langle \prod_i : e^{i\alpha_i\phi(x)} : \right\rangle = \begin{cases} \prod_{i<j}\left(\dfrac{-1}{(\xi_{+ij})_\varepsilon(\xi_{-ij})_\varepsilon}\right)^{(1/2)\alpha_i\alpha_j}, & \Sigma\alpha_i = 0 \\ 0, & \text{otherwise} \end{cases} \qquad [4]$$

where the resulting correlation function has been factored in terms of light-ray coordinates $\xi_{\pm ij} = x_{\pm i} - x_{\pm j}, x_\pm = t \pm x$, and the $\varepsilon$-prescription stands for taking the standard Wightman boundary value $t \to t + i\varepsilon$, $\lim_{\varepsilon\to 0}$ which insures the positive-energy condition. The finiteness of the limit insures that the resulting zero-mass limiting theory is a *bona fide* quantum field theory that is, its system of Wightman functions permits the construction of an operator theory in a Hilbert space with a distinguished vacuum vector.

The factorization into light-ray components [4] shows that the exponential charge-carrying operators inherit this factorization into two independent chiral components $:\exp i\alpha\,\phi(x) : = \exp i\alpha\,\phi_+(x_+):$ $:\exp i\alpha\,\phi_-(x_-):$, each one being covariant under scaling $\xi \to \lambda\xi$ if one assigns the scaling dimension $d = \alpha^2/2$ to the chiral exponential field and $d = 1$ to the current. As any Wightman field, this is a singular object which only after smearing with Schwartz test functions yields an (unbounded) operator. But the above form of the correlation function belongs to a class of distributions which admits a much larger test-function space consisting of smooth functions which instead of decreasing rapidly only need to be bounded so that they stay finite on the compactified light-ray line $\dot{R} = S^1$. To make this visible, one uses the Cayley transform (now $x$ denotes either $x_+$ or $x_-$)

$$z = \frac{1 + ix}{1 - ix} \in S^1 \qquad [5]$$

This transforms the Schwartz test function into a space of test functions on $S^1$ which have an infinite order zero at $z = -1$ (corresponding to $x = \pm\infty$) but the rotational transformed fields $j(z)$, $:\exp i\alpha\,\phi(z):$ permit the smearing with all smooth functions on $S^1$, a characteristic feature of all conformal invariant

theories as the present one turns out to be. There is an additional advantage in the use of this compactification. Fourier transforming the circular current actually allows for a quantum-mechanical zero mode whose possible nonzero eigenvalues indicate the presence of additional charge sectors beyond the charge-zero vacuum sector. For the exponential field, this leads to a quantum-mechanical pre-exponential factor which automatically insures the charge selection rules so that unrestricted (by charge conservation) Wick contraction rules can be applied. In this approach, the original chiral Dirac fermion $\psi(x)$ (from which the current was formed as the $:\psi\bar{\psi}:$ composite) reappears as a charge-carrying exponential field for $\alpha = 1$ and thus illustrates the meaning of bosonization/fermionization. (It is interesting to note that Jordan's (1937) original treatment of fermionization had such a pre-exponential quantum-mechanical factor.) Naturally, this terminology has to be taken with a grain of salt in view of the fact that the bosonic current algebra only generates a superselected subspace into which the charge-carrying exponential field does not fit. Only in the case of massive two-dimensional QFT fermions can be incorporated into a Fock space of bosons (see last section). At this point, it should however be clear to the reader that the physical content of Jordan's paper had nothing to do with its misleading title "neutrino theory of light" but rather was an early illustration about charge superselection rules in two-dimensional QFT.

A systematic and rigorous approach consists in solving the problem of positive-energy representation theory for the Weyl algebra on the circle (which is the rigorous operator-algebraic formulation of the abelian current algebra). (The Weyl algebra originated in quantum mechanics around 1927; its use in QFT only appeared after the cited Jordan paper. By representation we mean here a regular representation in which the exponentials can be differentiated in order to obtain (unbounded) smeared current operators.) It is the operator algebra generated by the exponential of a smeared chiral current (always with real test functions) with the following relation between the generators

$$\begin{aligned} W(f) &= \mathrm{e}^{\mathrm{i}j(f)} \\ j(f) &= \int \frac{\mathrm{d}z}{2\pi\mathrm{i}} j(z) f(z), [j(z), j(z')] \\ &= -\delta'(z - z') \end{aligned} \quad [6a]$$

$$\begin{aligned} W(f)W(g) &= \mathrm{e}^{-(1/2)s(f,g)} W(f+g) \\ W^*(f) &= W(-f) \end{aligned} \quad [6b]$$

$$\begin{aligned} \mathcal{A}(S^1) &= \mathrm{alg}\{W(f), f \in C_\infty(S^1)\} \\ \mathcal{A}(I) &= \mathrm{alg}\{W(f), \mathrm{supp} f \subset I\} \end{aligned} \quad [6c]$$

where

$$s(.\,,.) = \int \frac{\mathrm{d}z}{2\pi\mathrm{i}} f'(z) g(z)$$

is the symplectic form which characterizes the Weyl algebra structure and [6c] denotes the unique $C^*$ algebra generated by the unitary objects $W(f)$. A particular representation of this algebra is given by assigning the vacuum state to the generators $\langle W(f)\rangle_0 = e^{-(1/2)\|f\|_0^2}$, $\|f\|_0^2 = \sum_{n\geq 1} n|f_n|^2$. Starting with the vacuum Hilbert space representation $\mathcal{A}(S^1)_0 = \pi_0(\mathcal{A}(S^1))$, one easily checks that the formula

$$\langle W(f)\rangle_\alpha := \mathrm{e}^{\mathrm{i}\alpha f_0} \langle W(f)\rangle_0 \quad [7a]$$

$$\pi_\alpha(W(f)) = \mathrm{e}^{\mathrm{i}\alpha f_0} \pi_0(W(f)) \quad [7b]$$

defines a state with positive energy, that is, one whose GNS representation for $\alpha \neq 0$ is unitarily inequivalent to the vacuum representation. Its incorporation into the vacuum Hilbert space [7b] is part of the DHR formalism. It is convenient to view this change as the result of an application of an automorphism $\gamma_\alpha$ on the $C^*$-Weyl algebra $\mathcal{A}(S^1)$ which is implemented by a unitary charge-generating operator $\Gamma_\alpha$ in a larger (nonseparable) Hilbert space which contains all charge sectors $H_\alpha = \Gamma_\alpha H_0$, $H_0 \equiv H_{\mathrm{vac}} = \overline{\mathcal{A}(S^1)\Omega}$:

$$\begin{aligned} \langle W(f)\rangle_\alpha &= \langle \gamma_\alpha(W(f))\rangle_0 \\ \gamma_\alpha(W(f)) &= \Gamma_\alpha W(f) \Gamma_\alpha^* \end{aligned} \quad [8]$$

$\Gamma_\alpha \Omega = \Omega_\alpha$ describes a state with a rotational homogeneous charge distribution; arbitrary charge distributions $\rho_\alpha$ of total charge $\alpha$ that is, $\int (\mathrm{d}z/\ 2\pi\mathrm{i})\rho_\alpha = \alpha$ are obtained in the form

$$\psi^\zeta_{\rho_\alpha} = \eta(\rho_\alpha) W(\hat{\rho}^\zeta_\alpha) \Gamma_\alpha \quad [9]$$

where $\eta(\rho_\alpha)$ is a numerical phase factor and the net effect of the Weyl operator is to change the rotational homogeneous charge distribution into $\rho_\alpha$. The necessary charge-neutral compensating function $\rho^\zeta_\alpha$ in the Weyl cocycle $W(\rho^\zeta_\alpha)$ is uniquely determined in terms of $\rho_\alpha$ up to the choice of one point $\zeta \in S^1$ (the determining equation involves the $\ln z$ function which needs the specification of a branch cut (Schroer 2005)). From this formula, one derives the commutation relations $\psi^\zeta_{\rho_\alpha} \psi^\zeta_{\rho_\beta} = \mathrm{e}^{\pm\mathrm{i}\pi\alpha\beta} \psi^\zeta_{\rho_\beta} \psi^\zeta_{\rho_\alpha}$ for spacelike separations of the $\rho$ supports; hence, these fields are relatively local (bosonic) for $\alpha\beta = 2\mathbb{Z}$. In particular, if only one

type of charge is present, the generating charge is $\alpha_{\mathrm{gen}} = \sqrt{2N}$ and the composite charges are multiples, that is, $\alpha_{\mathrm{gen}}\mathbb{Z}$. This locality condition providing bosonic commutation relations does not yet ensure the $\zeta$-independence. Since the equation which controls the $\zeta$-change turns out to be

$$\psi_{\rho_\alpha}^{\zeta_1}\left(\psi_{\rho_\alpha}^{\zeta_2}\right)^* = \mathrm{e}^{\pm \mathrm{i}\pi\alpha\beta}\mathrm{e}^{2\pi \mathrm{i} Q\alpha} \qquad [10]$$

one achieves $\zeta$-independence by restricting the Hilbert space charges to be "dual" to that of the operators, that is,

$$Q = \left\{\frac{1}{\sqrt{2N}}\mathbb{Z}\right\}$$

The localized $\psi_{\rho_\alpha}^{\zeta_1}$ operators acting on the restricted separable Hilbert space $H_{\mathrm{res}}$ generate a $\zeta$-independent extended observable algebra $\mathcal{A}_N(S^1)$ (Schroer) and it is not difficult to see that its representation in $H_{\mathrm{res}}$ is reducible and that it decomposes into $2N$ charge sectors

$$\left\{\frac{1}{\sqrt{2N}}n, n = 0, 1, \ldots, N-1\right\}$$

Hence, the process of extension has led to a charge quantization with a finite ("rational") number of charges relative to the new observable algebra which is neutral in the new charge counting

$$\frac{1}{\alpha_{\mathrm{gen}}}\mathbb{Z}/\alpha_{\mathrm{gen}}\mathbb{Z} = \mathbb{Z}/\alpha_{\mathrm{gen}}^2 = \mathbb{Z}_{2N}$$

The charge-carrying fields in the new setting are also of the above form [9], but now the generating field carries the charge

$$\int \frac{\mathrm{d}z}{2\pi \mathrm{i}}\rho_{\mathrm{gen}} = Q_{\mathrm{gen}}$$

which is a $(1/2N)$ fraction of the old $\alpha_{\mathrm{gen}}$. Their commutation relations for disjoint charge supports are "braidal" (or better "plektonic" which is more on par with being bosonic/fermionic). (In the abelian case like the present, the terminology "anyonic" enjoys widespread popularity, but in the present context the "any" does not go well with charge quantization.) These objects considered as operators localized on $S^1$ do depend on the cut $\zeta$, but using an appropriate finite covering of $S^1$ this dependence is removed (Schroer 2005). So the field algebra $\mathcal{F}_{\mathbb{Z}2N}$ generated by the charge-carrying fields (as opposed to the bosonic observable algebra $\mathcal{A}_N$) has its unique localization structure on a finite covering of $S^1$. An equivalent description which gets rid of $\zeta$ consists in dealing with operator-valued sections on $S^1$. The extension $\mathcal{A} \to \mathcal{A}_N$, which renders the Hilbert space separable and quantizes the charges, seems to be characteristic for abelian current algebra; in all other models which have been constructed up to now the number of sectors is at least denumerable and in the more interesting ones even finite (rational models). An extension is called maximal if there exists no further extension which maintains the bosonic commutation relation. For the case at hand, this would require the presence of another generating field of the same kind as above, which belongs to an integer $N'$ is relatively local to the first one. This is only possible if $N$ is divisible by a square.

In passing, it is interesting to mention a somewhat unexpected relation between the Schwinger model, whose charges are screened, and the Jordan model. Since the Lagrangian formulation of the Schwinger model is a gauge theory, the analog of the four-dimensional "asymptotic freedom" wisdom would suggest the possibility of "charge liberation" in the short-distance limit of this model. This seems to contradict the statement that the intrinsic content of the Schwinger model ($QED_2$ with massless Fermions) (after removing a classical degree of freedom) is the QFT of a free massive Bose field and such a simple free field is at first sight not expected to contain subtle information about asymptotic charge liberation. (In its original gauge-theoretical form, the Schwinger model has an infinite vacuum degeneracy. The removal of this degeneracy (restoration of the cluster property) with the help of the "$\theta$-angle formalism" leaves a massive free Bose field (the Schwinger–Higgs mechanism). As expected in $d = 1 + 1$ the model only possesses this phase.) Well, as we have seen above, the massless limit really does have liberated charges and the short-distance limit of the massive free field is the massless model (Schroer).

As a result of the peculiar bosonization/fermionization aspect of the zero-mass limit of the derivative of the massive free field, Jordan's model is also closely related to the massless Thirring model (and the related Luttinger model for an interacting one-dimensional electron gas) whose massive version is in the class of factorizing models (see later section). (Another structural consequence of this aspect leads to Coleman's theorem (Schroer 2005) which connects the Mermin–Wagner no-go theorem for two-dimensional spontaneous continuous symmetry breaking with these zero-mass peculiarities.) The Thirring model is a special case in a vast class of "generalized" multi-coupling multicomponent Thirring models, that is, models with 4-fermion interactions. Under this name they were studied in the early 1970s (Schroer) with the aim to identify massless subtheories for which the currents form chiral current algebras.

The counterpart of the potential of the conserved Dirac current in the massive Thirring model is the sine-Gordon field, that is, a composite field which in the attractive regime of the Thirring coupling again obeys the so-called sine-Gordon equation of motion. Coleman gave a supportive argument (Schroer 2005) but some fine points about the range of its validity in terms of the coupling strength remained open. (It was noticed that the current potential of the free massive Dirac Fermion ($g=0$) does not obey the sine-Gordon equation (Schroer 2005).) A rigorous confirmation of these facts was recently given in the bootstrap form-factor setting (Schroer 2005). Massive models which have a continuous or discrete internal symmetry have "disorder" fields which implement a "half-space" symmetry on the charge-carrying field (acting as the identity in the other half-axis) and together with the basic pointlike field form composites which have exotic commutation relations (see the last section).

## The Conformal Setting, Structural Results

Chiral theories play a special role within the setting of conformal quantum fields. General conformal theories have observable algebras which live on compactified Minkowski space ($S^1$ in the case of chiral models) and fulfill the Huygens principle, which in an even number of spacetime dimension means that the commutator is only nonvanishing for lightlike separation of the fields. The fact that this classically expected behavior breaks down for nonobservable conformal fields (e.g., the massless Thirring field) was noticed at the beginning of the 1970s and considered paradoxical at that time ("reverberation" in the timelike (Huygens) region). Its resolution around 1974–75 confirmed that such fields are genuine conformal covariant objects but that some fine points about their causality needed to be addressed. The upshot was the proposal of two different but basically equivalent concepts about globally causal fields. They are connected by the following global decomposition formula:

$$A(x_{\rm cov}) = \sum A_{\alpha,\beta}(x),\ A_{\alpha,\beta}(x) = P_\alpha A(x) P_\beta,\ Z = \sum e^{id_\alpha} P_\alpha \qquad [11]$$

On the left-hand side, the spacetime point of the field is a point on the universal covering of the conformal compactified Minkowski space. These are fields (Lüescher and Mack 1975) (Schroer 2005) which "live" in the sense of quantum (modular) localization on the universal covering spacetime (or on a finite covering, depending on the "rationality" of the model) and fulfill the global causality condition previously discovered by I Segal (Schroer 2005). They are generally highly reducible with respect to the center of the covering group. The family of fields on the right-hand side, on the other hand, are fields which were introduced (Schroer and Swieca 1974; Schroer *et al.* 1975) with the aim to have objects which live on the projection $x(x_{\rm cov})$, that is, on the spacetime of the physics laboratory instead of the "hells and heavens" of the covering (Schroer 2005). They are operator-distributional valued sections in the compactification of ordinary Minkowski spacetime. The connection is given by the above decomposition formula into irreducible conformal blocks with respect to the center $\mathbf{Z}$ of the noncompact covering group $\widetilde{\mathrm{SO}(2,n)}$ where $\alpha, \beta$ are labels for the eigenspaces of the generating unitary $Z$ of the abelian center $\mathbf{Z}$. The decomposition [11] is minimal in the sense that in general there generally will be a refinement due to the presence of additional charge superselection rules (and internal group symmetries). The component fields are not Wightman fields since they annihilate the vacuum if the right-hand projection differs from $P_0 = P_{\rm vac}$.

Note that the Huygens (timelike) region in Minkowski spacetime has a timelike ordering structure $x \prec y$ or $x \succ y$ (earlier or later). In $d=1+1$, the topology allows in addition a spacelike left–right ordering $x \lessgtr y$. In fact, it is precisely the presence of these two orderings in conjunction with the factorization of the vacuum symmetry group $\widetilde{\mathrm{SO}(2,2)} \simeq \widetilde{\mathrm{PSL}(2R)}_l \otimes \widetilde{\mathrm{PSL}(2,R)}_r$, in particular $\mathbf{Z} = \mathbf{Z}_l \otimes \mathbf{Z}_r$, which is at the root of a significant simplification. This situation suggested a tensor factorization into chiral components and led to an extremely rich and successful construction program of two-dimensional conformal QFT as a two-step process: the classification of chiral observable algebras on the light ray and the amalgamation of left–right chiral theories to two-dimensional local conformal QFT. The action on the circular coordinates $z$ is through fractional $\mathrm{SU}(1,1)$ transformations

$$g(z) = \frac{\alpha z + \beta}{\bar\beta z + \bar\alpha}$$

whereas the covering group acts on the Mack–Luescher covering coordinates.

The presence of an ordering structure permits the appearance of more general commutation relations for the above $A_{\alpha\beta}$ component fields namely

$$A_{\alpha,\beta}(x) B_{\beta,\gamma}(y) = \sum_{\beta'} R^{\alpha,\gamma}_{\beta,\beta'} B_{\alpha,\beta'}(y) A_{\beta',\gamma}(x), \quad x > y \qquad [12]$$

with numerical $R$-coefficients which, as a result of associativity and relative commutativity with respect to observable fields, have to obey certain structure relations; in this way, Artin braid relations emerge as a new manifestation of the Einstein causality principle for observables in low-dimensional QFT (Rehren and Schroer 1989) (see Schroer 2005). Indeed, the DHR method to interpret charged fields as charge superselection carriers (tied by local representation theory to the bosonic local structure of observable algebras) leads precisely to such a plektonic statistics structure (Fredenhagen *et al.* 1992, Gabbiani and Froehlich 1993) for systems in low spacetime dimension (*see* Symmetries in Quantum Field Theory of Lower Spacetime Dimensions). With an appropriately formulated adjustment to observables fulfilling the Huygens commutativity, this plektonic structure (but now disconnected from particle/field statistics) is also a possible manifestation of causality for the higher-dimensional timelike structure (Schroer 2005).

The only examples known up to the appearance of the seminal BPZ work (Belavin *et al.* 1984) were the abelian current models of the previous section which furnish a rather poor man's illustration of the richness of the decomposition theory. The floodgates of conformal QFT were only opened after the BPZ discovery of "minimal models," which was preceded by the observation (Friedan *et al.* 1984) that the algebra of the stress–energy tensor came with a new representation structure which was not compatible with an underlying internal group symmetry (*see* Symmetries in Quantum Field Theory: Algebraic Aspects).

An important step in the structural study of chiral models was the recognition that the energy–momentum tensor has the commutation structure of a Lie field (Schroer 2005); in the next section, its algebraic structure and its representation theory will be presented.

## Chiral Fields and Two-Dimensional Conformal Models

Let us start with a family which generalizes the abelian model of the previous section. Instead of a one-component abelian current we now take $n$ independent copies. The resulting multicomponent Weyl algebra has the previous form except that the current is $n$-component and the real function space underlying the Weyl algebra consists of functions with values in an $n$-component real vector space $f \in LV$ with the standard Euclidean inner product denoted by $(\,,)$. The local extension now leads to $(\alpha, \beta) \in 2\mathbb{Z}$, that is, an even-integer lattice $\mathcal{L}$ in $V$, whereas the restricted Hilbert subspace $H_{\mathcal{L}^*}$ which ensures $\zeta$-independence is associated with the dual lattice $L^*$: $(\lambda_i, \alpha_k) = \delta_{ik}$ which contains $\mathcal{L}$. The resulting superselection structure (i.e., the $Q$-spectrum) corresponds to the finite factor group $\mathcal{L}^*/\mathcal{L}$. For self-dual lattices $\mathcal{L}^* = \mathcal{L}$ (which only can occur if $\dim \mathrm{V}$ is a multiple of 8), the resulting observable algebra has only the vacuum sector; the most famous case is the Leech lattice $\Lambda_{24}$ in $\dim V = 24$, also called the "moonshine" model. The observation that the root lattices of the Lie algebras of types $A$, $B$, or $E$ (e.g., su($n$) corresponding to $A_{n-1}$) also appear among the even-integral lattices suggests that the nonabelian current algebras associated to those Lie algebras can also be implemented. This turns out to be indeed true as far as the level-1 representations are concerned which brings us to the second family: the nonabelian current algebras of level $k$ associated to those Lie algebras; they are characterized by the commutation relation

$$[J_\alpha(z), J_\beta(z')] = \mathrm{i} f^\gamma_{\alpha\beta} j_\gamma(z)\delta(z - z') - \tfrac{1}{2} k g_{\alpha\beta}\delta'(z - z') \qquad [13]$$

where $f^\gamma_{\alpha\beta}$ are the structure constants of the underlying Lie algebra, $g$ their Cartan–Killing form, and $k$, the level of the algebra, must be an integer in order that the current algebra can be globalized to a loop group algebra. The Fourier decomposition of the current leads to the so-called affine Lie algebras, a special family of Kac–Moody algebras. For $k = 1$, these currents can be constructed as bilinears in terms of the multicomponent chiral Dirac field; there exists also the mentioned possibility to obtain them by constructing their maximal Cartan currents within the above abelian setting and representing the remaining nondiagonal currents as certain charge-carrying ("vertex" algebra) operators. Level-$k$ algebras can be constructed from reducing tensor products of $k$ level-1 currents or directly via the representation theory of infinite-dimensional affine Lie algebras. (The global exponentiated algebras (the analogs to the Weyl algebra) are called loop group algebras.) Either way one finds that, for example, the SU(2) current algebra of level $k$ has (together with the vacuum sector) $k + 1$ sectors (inequivalent representations). The different sectors are already distinguished by the structure of their ground states of the conformal Hamiltonian $L_0$. Although the computation of higher point correlation functions for $k > 1$, there is no problem in securing the existence of the algebraic nets which define these chiral models as well as their $k + 1$

representation sectors and to identify their generating charge-carrying fields (primary fields) including their $R$-matrices appearing in their plektonic commutation relations. It is customary to use the notation $SU(2)_k$ for the abstract operator algebras associated with the current generators [13] and we will denote their $k+1$ equivalence classes of representations by $\mathcal{A}_{SU(2)_k,n}, n=0,\ldots,k$, whereas representations of current algebras for higher rank groups require a more complicated labeling (in terms of Weyl chambers).

The third family of models are the so-called minimal models which are associated with the Lie-field commutation structure of the chiral stress–energy tensor which results from the chiral decomposition of a conformally covariant two-dimensional stress–energy tensor

$$[T(z), T(z')] = \mathrm{i}(T(z) + T(z'))\delta'(z - z') + \frac{\mathrm{i}c}{24\pi}\delta'''(z - z') \qquad [14]$$

whose Fourier decomposition yields the Witt–Virasoro algebra, that is, a central extension of the Lie algebra of the $\mathrm{Diff}(S^1)$. (The presence of the central term in the context of QFT (the analog of the Schwinger term) was noticed later; however, the terminology Witt–Virasoro algebra in the physics literature came to mean the Lie algebra of diffeomorphisms of the circle including the central extension.) The first two coefficients are determined by the physical role of $T(z)$ as the generating field density for the Lie algebra of the Poincaré group whereas the central extension parameter $c > 0$ (positivity of the two-point function) for the connection with the generation of the Moebius transformations and the undetermined parameter $c > 0$ (the central extension parameter) is easily identified with the strength of the two-point function. Although the structure of the $T$-correlation functions resembles that of free fields (in the sense that is an algebraically computable unique set of correlation functions once one has specified the two-point function), the realization that $c$ is subject to a discrete quantization if $c < 1$ came as a surprise. As already mentioned, the observation that the superselection sectors (the positive-energy representation structure) of this algebra did not at all follow the logic of a representation theory of an inner symmetry group generated a lot of attention and stimulated a flurry of publications on symmetry concepts beyond groups (quantum groups). A concept of fundamental importance is the DHR theory of localized endomorphisms of operator algebras and the concept of operator-algebraic inclusions (in particular, inclusions with conditional expectations – V Jones inclusions).

The $SU(2)_k$ current coset construction (Goddard *et al.* 1985) revealed that the proof of existence and the actual construction of the minimal models is related to that of the $SU(2)_k$ current algebras. Constructing a chiral model does not necessarily mean the explicit determination of the $n$-point Wightman functions of their generating fields (which for most chiral models remains a prohibitively complicated task) but rather a proof of their existence by demonstrating that these models are obtained from free fields by a series of computational complicated but mathematically controlled operator-algebraic steps as reduction of tensor products, formation of orbifolds under group actions, coset constructions, and a special kind of extensions. The generating fields of the models are nontrivial in the sense of not obeying free-field equations (i.e., not being “on-shell”). The cases where one can write down explicit $n$-point functions of generating fields are very rare; in the case of the minimal family this is limited to the field theory of the Ising model (Schroer 2005).

To show the power of inclusion theory for the determination of the charge content of theory, let us look at a simple illustration in the context of the above multicomponent abelian current algebra. The vacuum representation of the corresponding Weyl algebra is generated from smooth $V$-valued functions on the circle modulo constant functions (i.e., functions with vanishing total integral) $f \in LV_0$. These functions equipped with the aforementioned complex structure and scalar product yield a Hilbert space. The $I$-localized subalgebra is generated by the Weyl image of $I$-supported functions (class functions whose representing functions are constant in the complement $I'$)

$$\begin{aligned} \mathcal{A}(I) &:= \mathrm{alg}\{W(f) | f \in K(I)\} \\ K(I) &= \{f \in LV_0 | f = \mathrm{const.in}\, I'\} \end{aligned} \qquad [15]$$

The one-interval Haag duality $\mathcal{A}(I)' = \mathcal{A}(I')$ (the commutant algebra equals the algebra localized in the complement) is simply a consequence of the fact that the symplectic complement $K(I)'$ in terms of $\mathrm{Im}(f, g)$ consists of real functions in that space which are localized in the complement, that is, $K(I)' = K(I')$. The answer to the same question for a double interval $I = I_1 \cup I_3$ (think of the first and third quadrant on the circle) does not lead to duality but rather to a genuine inclusion

$$\begin{aligned} K((I_1 \cup I_3)') = K(I_2 \cup I_4) &\subset K(I_1 \cup I_3)' \\ K(I_1 \cup I_3) &\subset K((I_1 \cup I_3)')' \end{aligned} \qquad [16]$$

The meaning of the left-hand side is clear; these are functions which are constant in $I_1 \cup I_3$ with the same constant in the two intervals whereas the functions on the right-hand side are less restrictive in that the constants can be different. The conversion of real subspaces into von Neumann algebras by the Weyl functor leads to the algebraic inclusion $\mathcal{A}(I_1 \cup I_3) \subset \mathcal{A}((I_1 \cup I_3)')'$. In physical terms, the enlargement results from the fact that within the charge neutral vacuum algebra a charge split with one charge in $I_1$ and the compensating charge in $I_2$ for all values of the (unquantized) charge occurs. A more realistic picture is obtained if one allows a charge split to be subjected to a charge quantization implemented by a lattice condition $f(I_2) - f(I_4) \in 2\pi L$ which relates the two multicomponent constant functions (where $f(I)$ denotes the constant value $f$ takes in $I$). As in the previous one-component case, the choice of even lattices corresponds to the local (bosonic) extensions. Although imposing such a lattice structure destroys the linearity of the $K$, the functions still define Weyl operators which generated operator algebras $\mathcal{A}_L(I_1 \cup I_2)$. (The linearity structure is recovered on the level of the operator algebra.) But now the inclusion involves the dual lattice $L^*$ (which of course contains the original lattice),

$$\begin{aligned} &\mathcal{A}_L(I_1 \cup I_2) \subset \mathcal{A}_{L^*}(I_1 \cup I_2) \\ &\operatorname{ind}\{\mathcal{A}_L(I_1 \cup I_2) \subset \mathcal{A}_L((I_1 \cup I_2)')'\} = |G| \\ &\mathcal{A}_L(I_1 \cup I_2) = \operatorname{inv}_G \mathcal{A}_{L^*}(I_1 \cup I_2) \end{aligned}$$

This time the possible charge splits correspond to the factor group $G = L^*/L$, that is, the number of possibilities is $|G|$ which measures the relative size of the bigger algebra in terms of the smaller. This is a special case of the general concept of the so-called Jones index of an inclusion which is a numerical measure of its depth. A prerequisite is that the inclusion permits a conditional expectation which is a generalization of the averaging under the "gauge group" $G$ on $\mathcal{A}_{L^*}(I_1 \cup I_2)$ in the third equation above, which identifies the invariant smaller algebra with the fix-point algebra (the invariant part) under the action of $G$. In fact, using the conceptual framework of Jones, one can show that the two-interval inclusion is independent of the position of the disjoint intervals characterized by the group $G$.

There exists another form of this inclusion which is more suitable for generalizations. One starts from the charge quantized extended local algebra $\mathcal{A}_L^{\text{ext}} \supset \mathcal{A}$ described earlier in terms of an even-integer lattice $L$ (which lives in the separable Hilbert space $H_{L^*}$) as our observable algebra. Again the Haag duality is violated and converted into an inclusion $\mathcal{A}_L^{\text{ext}}(I_1 \cup I_2) \subset \mathcal{A}_L^{\text{ext}}((I_1 \cup I_2)')'$ which turns out to have the same $G = L^*/L$ charge structure (it is in fact isomorphic to the previous inclusion). In the general setting (current algebras, minimal model algebras, ...), this double interval inclusion is particularly interesting if the associated Jones index is finite. One finds Kawahigashi *et al.* (2001) (Schroer 2005).

**Theorem 1** *A chiral theory with finite Jones index* $\mu = \operatorname{ind}\{\mathcal{A}((I_1 \cup I_2)')' : \mathcal{A}(I_1 \cup I_2)\}$ *for the double interval inclusion (always assuming that* $A(S^1)$ *is strongly additive and split) is a rational theory and the statistical dimensions* $d_\rho$ *of its charge sectors are related to* $\mu$ *through the formula*

$$\mu = \sum_\rho d_\rho^2 \qquad [17]$$

Instead of presenting more constructed chiral models, it may be more informative to mention some of the algebraic methods by which they are constructed and explored. The already mentioned DHR theory provides the conceptual basis for converting the notion of positive-energy representation sectors of the chiral model observable algebras $\mathcal{A}$ (equivalence classes of unitary representations) into localized endomorphisms $\rho$ of this algebra. This is an important step because contrary to group representations which have a natural tensor product composition structure, representations of operator algebras generally do not come with a natural composition structure. The DHR endomorphisms theory of $\mathcal{A}$ leads to fusion laws and an intrinsic notion of generalized statistics (for chiral theories: plektonic in addition to bosonic/fermionic). The chiral statistics parameters are complex numbers (Haag 1992) whose phase is related to a generalized concept of spin via a spin-statistics theorem and whose absolute value (the statistics dimension) generalized the notion of multiplicities of fields known from the description of inner symmetries in higher-dimensional standard QFTs. The different sectors may be united into one bigger algebra called the exchange algebra $\mathcal{F}_{\text{red}}$ in the chiral context (the "reduced field bundle" of DHR) in which every sector occurs by definition with multiplicity 1 and the statistics data are encoded into exchange (commutation) relations of charge-carrying operators or generating fields ("exchange algebra fields") (Schroer 2005). Even though this algebra is useful in that all properties concerning fusion and statistics are nicely encoded, it lacks some cherished properties of standard field theory

namely there is no unique state–field relation, that is, no Reeh–Schlieder property (a field $A_{\alpha\beta}$ whose source projection $P_\beta$ does not coalesce with the vacuum projection annihilates the vacuum); in operator-algebraic terms, the local algebras are not factors. This poses the question of how to manufacture from the set of all sectors natural (not necessarily local) extensions with these desired properties. It was found that this problem can be characterized in operator-algebraic terms by the existence of the so-called DHR triples (Schroer). In case of rational theories, the number of such extensions is finite and in the aforementioned "classical" current algebra and minimal models they all have been constructed by this method (thus confirming existing results completing the minimal family by adding some missing models). The same method adapted to the chiral tensor product structure of $d = 1 + 1$ conformal observables classifies and constructs all two-dimensional local (bosonic/ fermionic) conformal QFT $\mathcal{B}_2$ which can be associated with the observable chiral input. It turns out that this approach leads to another of those pivotal numerical matrices which encode structural properties of QFT: the coupling matrix $Z$,

$$\mathcal{A} \otimes \mathcal{A} \subset \mathcal{B}_2$$
$$\sum_{\rho\sigma} Z_{\rho,\sigma}\rho(\mathcal{A}) \otimes \sigma(\mathcal{A}) \subset \mathcal{A} \otimes \mathcal{A} \qquad [18]$$

where the second line is an inclusion solely expressed in terms of observable algebras from which the desired (isomorphic) inclusion in the first line follows by a canonical construction, the so-called Jones basic construction. The numerical matrix $Z$ is an invariant closely related to the so-called "statistics character matrix" (Schroer 2005) and in case of rational models it is even a modular invariant with respect to the modular $\mathrm{SL}(2, Z)$ group transformations (which are closely related to the matrix $S$ in the final section).

## Integrability, the Bootstrap Form-Factor Program

Integrability in QFT and the closely associated bootstrap form-factor construction of a very rich class of massive two-dimensional QFTs can be traced back to two observations made during the 1960s and 1970s ideas. On the one hand, there was the time-honored idea to bypass the "off-shell" field-theoretic approach to particle physics in favor of a pure on-shell $S$-matrix setting which (in particular recommended for strong interactions), as a result of the elimination of short distances via the mass-shell restriction, would be free of ultraviolet divergencies. This idea was enriched in the 1960s by the crossing property which in turn led to the bootstrap idea, a highly nonlinear seemingly self-consistent proposal for the determination of the $S$-matrix. However, the protagonists of this $S$-matrix bootstrap program placed themselves into a totally antagonistic fruitless position with respect to QFT so that the strong return of QFT in the form of gauge theory undermined their credibility. On the other hand, there were rather convincing quasiclassical calculations in certain two-dimensional massive QFTs as, for example, the sine-Gordon model which indicated that the obtained quasiclassical mass spectrum is exact and hence suggested that the associated QFTs are integrable (Dashen *et al.* 1975) and have no real particle creation. These provocative observations asked for a structural explanation beyond quasiclassical approximations, and it soon became clear that the natural setting for obtaining such mass formulas was that of the "fusion" of boundstate poles of unitary crossing-symmetric purely elastic $S$-matrices; first in the special context of the sine-Gordon model (Schroer *et al.* 1976) and later as a classification program from which factorizing $S$-matrices can be determined by solving well-defined equations for the elastic two-particle $S$-matrix (Karowski *et al.* 1977). (It was incorrectly believed that the "nontrivial elastic scattering implies particle creation" statement of Aks (Aks, 1963) is also valid for low-dimensional QFTs.) Some equations in this bootstrap approach resembled mathematical structures which appeared in C N Yang's work on nonrelativistic $\delta$-function particle interactions as well as relations for Boltzmann weights in Baxter's work on solvable lattice models; hence, they were referred to as Yang–Baxter relations. These results suggested that the old bootstrap idea, once liberated from its ideological dead freight (in particular from the claim that the bootstrap leads to a unique "theory of everything" (minus gravity)), generates a useful setting for the classification and construction of factorizing two-dimensional relativistic $S$-matrices. Adapting certain known relations between two-particle form factors of field operators and the $S$-matrix to the case at hand (Karowski and Weisz 1978), and extending this with hindsight to generalized (multiparticle) form factors, one arrived at the axiomatized recipes of the bootstrap form-factor program of $d = 1 + 1$ factorizable models (Smirnov 1992). Although this approach can be formulated within the

setting of the LSZ scattering formalism, the use of a certain algebraic structure (Zamolodchikov and Zamolodchikov 1979) which in the simplest version reads

$$Z(\theta)Z^*(\theta') = S^{(2)}(\theta - \theta')Z^*(\theta')Z(\theta) + \delta(\theta - \theta')$$
$$Z(\theta)Z(\theta') = S^{(2)}(\theta' - \theta)Z(\theta')Z(\theta) \quad [19]$$

(the $\delta$-term Faddeev is due to Faddeev) brought significant simplifications. In the general case, the $Z's$ are vector valued and the $S^{(2)}$-structure function is matrix valued. (The identification of the Z–F structure coefficients with the elastic two-particle $S$-matrix $S^{(2)}$ (which is prenempted by our notation) can be shown to follow from the physical interpretation of the Z-F structure in terms of localization.) In that case the associativity of the Z–F algebra is equivalent to the Yang–Baxter equations. Recently, it became clear that this algebraic relation has a deep physical interpretation; it is the simplest algebraic structure which can be associated with generators of nontrivial wedge-localized operator algebras (see the next section).

Conceptually as well as computationally it is much simpler to identify the intrinsic meaning of integrability in QFT with the factorization of its $S$-matrix or a certain property of wedge-localized algebras (see next section) than to establish integrability (*see* Integrability and Quantum Field Theory).

The first step of the bootstrap form-factor program namely the classification and construction of model $S$-matrices follows a combination of two patterns: prescribing particle multiplets transforming according to group symmetries and/or specifying structural properties of the particle spectrum. The simplest illustration for the latter strategy is supplied by the $\mathbb{Z}_N$ model. In terms of particle content, $\mathbb{Z}_N$ demands the identification of the $N$th bound state with the antiparticle. Since the fusion condition for the bound mass $m_b^2 = (p_1 + p_2)^2 = m_1^2 + m_2^2 + 2m_1 m_2 ch(\theta_1 - \theta_2)$ is only possible for a pure imaginary rapidity difference $\theta_{12} = \theta_1 - \theta_2 = \mathrm{i}\alpha$ ("binding angle"). Hence, the binding of two "elementary" particles of mass $m$ gives

$$m_2 = m\frac{\sin 2\alpha}{\sin \alpha}$$

and more generally of $k$ particles with

$$m_k = m\frac{\sin k\alpha}{\sin \alpha}$$

so that the antiparticle mass condition $m_N = \bar{m} = m$ fixes the binding angle to $\alpha = 2\pi/N$. (The quotation mark is meant to indicate that in contrast to the Schrödinger QM there is "nuclear democracy" on the level of particles. The inexorable presence of interaction-caused vacuum polarization limits a fundamental/fused hierarchy to the fusion of charges.) The minimal (no additional physical poles) two-particle $S$-matrix in terms of which the $n$-particle $S$-matrix factorizes is therefore

$$S^{(2)}_{\min} = \frac{\sin(1/2)(\theta + (2\pi \mathrm{i})/N)}{\sin(1/2)(\theta - (2\pi \mathrm{i})/N)} \quad [20]$$

(minimal = without so-called CDD poles) The SU($N$) model as compared with the U($N$) model requires a similar identification of bound states of $N - 1$ particles with an antiparticle. This $S$-matrix enters as in the equation for the vacuum to $n$-particle meromorphic form factor of local operators; together with the crossing and the so-called "kinematical pole equation," one obtains a recursive infinite system linking a certain residue with a form factor involving a lower number of particles. The solutions of this infinite system form a linear space from which the form factors of specific tensor fields can be selected by a process which is analogous but more involved than the specification of a Wick basis of composite free fields. Although the statistics property of two-dimensional massive fields is not intrinsic but a matter of choice, it would be natural to realize, for example, the $\mathbb{Z}_N$ fields as $\mathbb{Z}_N$-anyons.

Another rich class of factorizing models are the Toda theories of which the sine-Gordon and sinh–Gordon are the simplest cases. For their descriptions, the quasiclassical use of Lagrangians (supported by integrability) turns out to be of some help in setting up their more involved bootstrap form-factor construction.

The unexpected appearance of objects with new fundamental (solitonic) charges (e.g., the Thirring field as the carrier of a solitonic sine-Gordon charge) and the unexpected confinement of charges (e.g., the CP(1) model as a confined SU(2) model) turn out to be opposite sides of the same coin and both cases have realizations in the setting of factorizing models (Schroer 2005).

## Recent Developments

There are two ongoing developments which place the two-dimensional bootstrap form-factor program into a more general setting which permits to understand its position in the general context of local quantum physics.

One of these starts from the observation that the smallest spacetime localization region in which it is possible to find vacuum-polarization-free generators (PFG) in the presence of interactions is the wedge

region. If one demands in addition that these generators (necessarily unbounded operators) have the standard domain properties of QFT (which include stability of the domain under translations), then one finds that this leads precisely to the two-dimensional Z–F algebraic structure which in turn in this way a spacetime interpretation for the first time acquires. In these investigations (Schroer 2005), modular localization theory plays a prominent role and there are strong indications that with these methods one can show the nontriviality of intersections of wedge algebras which is the algebraic criterion for the existence of a model within local quantum physics.

There is a second constructive idea based on light-front holography which uses the radical reorganization of spacetime properties of the algebraic structure while maintaining the physical content including the Hilbert space. Since spacetime localization aspects (apart from the remark about wedge algebras and their PFG generators made before) are traditionally related to the concept of fields, holographic methods tend to de-emphasize the particle structure in favor of "field properties." Indeed, the transversely extended chiral theories which arise as the holographic image lead to simplification of many interesting properties with very similar aims to the old "light-cone quantization" except that light-front holography is another way of looking at the original local ambient theory without subjecting it to another quantization. (The price for this simplification is that as a result of the nonuniqueness of the holographic inversion certain problems cannot be formulated.)

Actually, as a result of the absence of a transverse direction in the two-dimensional setting, the family of factorizing models provides an excellent theoretical laboratory to study their rigorous "chiral encoding" which is conceptually very different from Zamolodchikov's perturbative relation (which is based on identifying a factorizing model in terms of a perturbation on a chiral theory).

It turns out that the issue of statistics of particles loses its physical relevance for two-dimensional massive models since they can be changed without affecting the physical content. Instead such notions as order/disorder fields and soliton take their place (Schroer 2005).

In accordance with its historical origin, the theory of two-dimensional factorizing models may also be viewed as an outgrowth of the quantization of classical integrable systems (Integrability and Quantum Field Theory). But in comparison with the rather involved structure of integrabilty (verifying the existence of sufficiently many commuting conservation laws), the conceptual setting of factorizing models within the scattering framework (factorization follows from existence of wedge-localized tempered PFGs) is rather simple and intrinsic (Schroer 2005).

Among the additional ongoing investigations in which the conceptual relation with higher-dimensional QFT is achieved via modular localization theory, we will select three which have caught our, active attention. One is motivated by the recent discovery of the adaptation of Einsteins classical principle of local covariance to QFT in curved spacetime. The central question raised by this work (*see* Algebraic Approach to Quantum Field Theory) is if all models of Minkowski spacetime QFTs permit a local covariant extension to curved spacetime and if not which models do? In the realm of chiral QFT, this would amount to ask if all Moebius-invariant models are also Diff($S^1$)-covariant. It has been known for sometime that a QFT with all its rich physical content can be uniquely defined in terms of a carefully chosen relative position of a finite number of copies of one unique von Neumann operator algebra within one common Hilbert space. This is a perfect quantum field-theoretical illustration for Leibnitz's philosophical proposal that reality results from the relative position of "monades" (As opposed to the more common (Newtonian) view that the material reality originates from a material content being placed into a spacetime vessel) if one takes the step of identifying the hyperfinite typ $III_1$ Murray von Neumann factor algebra with an abstract monade from which the different copies result from different ways of positioning in a shared Hilbert space (Schroer 2005). In particular, Moebius-covariant chiral QFTs arise from two monades with a joint intersection defining a third monade in such a way that the relative positions are specified in terms of natural modular concepts (without reference to geometry). This begs the question whether one can extend these modular-based algebraic ideas to pass from the global vacuum preserving Moebius invariance to local Diff(S) covariance Moeb $\rightarrow$ Diff($S^1$). This would be precisely the two-dimensional adaptation of the crucial problem raised by the recent successful generalization of the local covariance principle underlying Einstein's classical theory of gravity to QFT in curved spacetime: does every Poincaré covariant Minkowski spacetime QFT allow a unique correspondence with one curved spacetime (having the same abstract algebraic substrate but with a totally different spacetime encoding)? In the chiral context, one is led to the notion of "partially geometric modular groups" which only act geometrically if restricted to specific subalgebras (Schroer

2005). It is hard to imagine how one can combine quantum theory and gravity without understanding first the still mysterious links between spacetime geometry, thermal properties, and relative positioning of monades in a joint Hilbert space.

A second important umbilical cord with higher-dimensional theories is the issue of "Euclideanization" in particular the chiral counterpart of Osterwalder–Schrader localization and the closely related Nelson–Symanzik duality. In concrete chiral models (e.g., the models in the section "Chiral fields and two-dimensional conformal models"), it has been noted as a result of explicit calculations that the analytic continuation in the angular parametrization for thermal correlation functions leads to a duality relation in

$$\langle A(\varphi_1, \ldots, \varphi_n) \rangle_{\alpha, 2\pi\beta_t} = \left(\frac{\mathrm{i}}{\beta_t}\right)^a \sum_\gamma S_{\alpha\gamma} \left\langle A\left(\frac{\mathrm{i}}{\beta_t}\varphi_1, \ldots, \frac{\mathrm{i}}{\beta_t}\varphi_n\right)\right\rangle_{\gamma,(2\pi/\beta_t)} \qquad [21]$$

where the thermal correlation function is defined as

$$\begin{aligned} \langle A(\varphi_1, \ldots, \varphi_n) \rangle_{\rho_\alpha, 2\pi\beta_t} &:= \mathrm{tr}_{H_{\rho_\alpha}} \mathrm{e}^{-2\pi\beta_t \left(L_0^{\rho_\alpha} - (c/24)\right)} \pi_{\rho_\alpha}(A(\varphi_1, \ldots, \varphi_n)) \\ A(\varphi_1, \ldots, \varphi_n) &= \prod_{i=1}^n A_i(\varphi_i) \end{aligned} \qquad [22]$$

Compared with the thermally extended Nelson–Symanzik relation for two-dimensional QFT one notices that in addition to the expected behavior of real coordinates becoming imaginary and the $2\pi$-periodicity changing role with the (suitably normalized) KMS inverse temperature, there is a rotation in the space of superselected charges in terms of a unitary matrix $S$ whose origin lies in the braid group statistics (the statistics character matrix). The deeper structural explanation which shows that this relation is not just a property of special models, but rather a generic property of chiral QFT, comes from a very deep angular Euclideanization which is based on modular theory (Schroer). Specializing $A =$ identity, one obtains a relation for the partition function, the famous Verlinde identity which is part of the transformation law of the thermal angular correlation functions under the $\mathrm{SL}(2, \mathrm{R})$ modular group.

There are many additional important observations on factorizing models whose relation to the physical principles of QFT, unlike the bootstrap form-factor program, is not yet settled. The meaning of the $c$-parameter outside the chiral setting and ideas on its renormalization group flow as well as the various formulations of the thermodynamic Bethe ansatz belong to a series of interesting observations whose final relation to the principles of QFT still needs clarification.

*See also:* Algebraic Approach to Quantum Field Theory; Axiomatic Quantum Field Theory; Bosons and Fermions in External Fields; Euclidean Field Theory; Integrablility and Quantum Field Theory; Operator Product Expansion in Quantum Field Theory; Sine-Gordon Equation; Symmetries in Quantum Field Theory: Algebraic Aspects; Symmetries in Quantum Field Theory of Lower Spacetime Dimensions; Tomita–Takesaki Modular Theory.

## Further Reading


Abdalla E, Abdalla MCB, and Rothe K (1991) *Non-Perturbative Methods in 2-Dimensional Quantum Field Theory*. Singapore: World Scientific.

Belavin AA, Polyakov AM, and Zamolodchikov AB (1984) Infinite conformal symmetry in two-dimensional quantum field theory. *Nuclear Physics B* 241: 333.

Bisognano JJ and Wichmann EH (1975) *Journal of Mathematical Physics* 16: 985.

Dashen F, Hasslacher B, and Neveu A (1975) *Physics Reviews D* 11: 3424.

Di Francesco P, Mathieu P, and Sénéchal D (1996) *Conformal Field Theory*. Berlin: Springer.

Doplicher S, Haag R, and Roberts JE (1971/1974) *Communications in Mathematical Physics* 23: 199; 35: 49.

Fredenhagen K, Rehren KH, and Schroer B (1992) Superselection sector with braid group statistics and exchange algebras II: Geometric aspects and conformal invariance. *Reviews of Mathematical Physics* 1 (special issue): 113.

Furlan P, Sotkov G, and Todorov I (1989) Two-dimensional conformal quantum field theory. *Rivista del Nuovo Cimento* 12: 1–203.

Gabbiani F and Froehlich J (1993) Operator algebras and conformal field theory. *Communications in Mathematical Physics* 155: 569.

Glimm J and Jaffe A (1987) *Quantum Physics. A Functional Integral Point of View*. Berlin: Springer.

Ginsparg P (1990) Applied conformal field theory. In: Brezin E and Zinn-Justin J (eds.) *Fields, Strings and Critical Phenomena*, Les Houches 1988. Amsterdam: North-Holland.

Goddard P, Kent A, and Olive D (1985) Virasoro algebras and coset space models. *Physics Letters B* 152: 88.

Haag R (1992) *Local Quantum Physics*. Berlin: Springer.

Ising E (1925) *Zeitschrift für physik* 31: 253.

Jordan P (1937) Beiträge zur Neutrinotheorie des Lichts. *Zeitschrift für Physik* 114: 229 and earlier papers quoted therein.

Karowski M, Thun H-J, Truoung TT, and Weisz P (1977) *Physics Letters B* 67: 321.

Karowski M and Weisz P (1978) *Physics Reviews B* 139: 445.

Kawahigashi Y, Longo R, and Mueger M (2001) Multi-interval subfactors and modularity in representations of conformal field theory. *Communications in Mathematical Physics* 219: 631.

Lenz W (1920) *Physikalische Zeitschrift* 21: 613.

Lüescher M and Mack G (1975) Global conformal invariance in quantum field theory. *Communications in Mathematical Physics* 41: 203.

Rehren K-H and Schroer B (1987) Exchange algebra and Ising $n$-point functions. *Physics Letters B* 198: 84.

Rehren KH and Schroer B (1989) Einstein causality and Artin braids. *Nuclear Physics B* 312: 715.
Schroer B (2005) Two-dimensional models, a testing ground for principles and concepts of QFT, *Annals of Physics* (in print) (hep-th/0504206).
Schroer B and Swieca JA (1974) Conformational transformations for quantized fields. *Physics Reviews D* 10: 480.
Schroer B, Swieca JA, and Voelkel AH (1975) Global operator expansions in conformally invariant relativistic quantum field theory. *Physics Reviews D* 11: 11.
Schroer B, Truong TT, and Weisz P (1976) Towards an explicit construction of the Sine-Gordon field theory. *Annals of Physics (New York)* 102: 156.
Schwinger J (1962) *Physical Review* 128: 2425.
Schwinger J (1963) Gauge theory of vector particles. In: *Theoretical Physics Trieste Lectures 1962*. Wien: IAEA.
Smirnov FA (1992) *Advanced Series in Mathematical Physics 14*. Singapore: World Scientific.
Streater RF and Wightman AS (1964) *PCT, Spin and Statistics and All That*. New York: Benjamin.
Zamolodchikov AB and Zamolodchikov AB (1979) *Annals of Physics (New York)* 120: 253.

# U

## Universality and Renormalization

**M Lyubich**, University of Toronto, Toronto, ON, Canada and Stony Brook University, Stony Brook, NY, USA



### Introduction

Discovery of the universality phenomenon and the underlying renormalization mechanism by Feigenbaum and independently by Coullet and Tresser in late 1970s was one of the most influential events in the dynamical systems theory in the last quarter of the twentieth century. It was numerically observed that the cascades of doubling bifurcations leading to chaotic regimes in one-parameter families of interval maps, as well as the dynamical attractors that appear in the limits, exhibit the universal small-scale geometry. To explain this surprising observation, a "Renormalization Conjecture" was formulated which asserted that a natural renormalization operator acting in the space of dynamical systems has a unique hyperbolic fixed point.

It took about two decades to prove this conjecture rigorously (and without the help of computers). The proof revealed rich mathematical structures behind the universality phenomenon that linked it tightly to holomorphic dynamics and conformal and hyperbolic geometry.

Besides the universality *per se*, the renormalization theory led to many other important results. It includes the proof of the regular or stochastic dichotomy that gives us a complete understanding of the real quadratic family (and more general families of one-dimensional maps) from measure-theoretic point of view, as well as deep advances in several key problems of holomorphic dynamics.

Since the original discovery, many other manifestations of the universality have been observed, experimentally, numerically, and theoretically, in various classes of dynamical systems. However, in this article we will focus on mathematical aspects of the original phenomenon.

#### General Terminology and Notations

We will use general notations and terminology from Holomorphic Dynamics.

### Unimodal Maps

#### Definitions and Conventions

Let us consider a smooth interval map $f: I \to I$. It is called *unimodal* if it has a single critical point $c$ and this point is an extremum. We assume that the critical point is nondegenerate, unless otherwise it is explicitly stated. A unimodal map is called $S$-unimodal if it has a negative Schwarzian derivative:

$$Sf = \frac{f'''}{f'} - \frac{3}{2}\left(\frac{f''}{f'}\right)^2 < 0$$

For simplicity, we also assume that the map $f$ is even, and normalize it so that $c = 0$ and one of the endpoints of $I$ is a fixed point.

#### Topological Dynamics

Let $J \ni 0$ be a 0-symmetric periodic interval, that is, $f^p(J) \subset J$ for some $p \in \mathbb{N}$, such that the intervals $J_k = f^k(J), k = 0, 1, \ldots, p-1$, have disjoint interiors. Then we refer to $\cup J_k$ as a *cycle of intervals* of period $p$.

According to their topological dynamics, $S$-unimodal maps can be divided into three possible types (Sharkovskii, Singer, Guckenheimer, Misiurewicz, van Strien, Blokh, etc.):

- *Regular maps*. Such a map has an attracting or parabolic cycle $\boldsymbol{\alpha}$. In this case, almost all trajectories of $f$ converge to $\boldsymbol{\alpha}$. In case $\boldsymbol{\alpha}$ is attracting, the map $f$ is also called *hyperbolic* (*see* Holomorphic Dynamics).
- *Topologically chaotic maps*. For such a map, there is a cycle of intervals $\cup J_k$ such that the restriction $f | \cup J_k$ is *topologically transitive* (i.e., it has a dense orbit). Moreover, for almost all $z \in I$, $\operatorname{orb} z$ eventually lands in this cycle.
- *Infinitely renormalizable maps*. For such a map, there is a nested sequence of periodic intervals $J^1 \supset J^2 \supset \cdots \ni 0$ of periods $p_n \to \infty$. Then the

intersection of the corresponding cycles of intervals,

$$A = A_f = \bigcap_{n=0}^{\infty} \bigcup_{k=0}^{p_n - 1} f^k(J^n) \qquad [1]$$

is a Cantor set endowed with a natural group structure (inverse limit of cyclic groups $\mathbb{Z}/p_n\mathbb{Z}$) such that $f|A$ becomes a group translation. Moreover, $f^n z \to A$ for a.e. $z \in I$. This Cantor set is also called the *Feigenbaum attractor* of $f$.

### Kneading Theory

Kneading theory (Milnor and Thurston, mid-1970s) gives a complete topological classification of $S$-unimodal maps (and more general one-dimensional maps). Let $I_+$ and $I_-$ stand for the components of $I \backslash \{0\}$, where $I_+ \ni f(0)$. To any point $x \in I$, let us associate its *itinerary* $(\varepsilon_n)_{n=0}^{N}$, where $\varepsilon_n \in \{+, -, 0\}$, $N \in \mathbb{Z}_+ \cup \infty$, in the following way. If $x$ is precritical then $N \in \mathbb{Z}_+$ is the smallest number such that $f^N x = 0$, and we let $\varepsilon_N = 0$. Otherwise, $N = \infty$. For $n < N, \varepsilon_n = +$ if $f^n x \in I_+$, and $\varepsilon_n = -$ if $f^n x \in I_-$.

The *kneading sequence* of $f$ is the itinerary of the critical value $f(0)$. It essentially classifies $S$-unimodal maps: *two nonregular S-unimodal maps are topologically conjugate if and only if they have the same kneading sequence*. (In the regular case, one should state if the map is hyperbolic or parabolic and specify the sign of the multiplier of the corresponding cycle.)

The kneading theory completely describes *admissible* kneading sequences (realizable by some unimodal maps), and order them linearly in such a way that a bigger sequence corresponds to a more "complicated" map. The minimal admissible kneading sequence, $+ + +$, is realized by the parabolic map $x \mapsto x^2 + 1/4$, while the maximal one, $+ - - - -$, is realized by the Chebyshev map $x \mapsto x^2 - 2$.

A central result of the kneading theory is the *Intermediate Value Theorem* asserting that a smooth one-parameter family of $S$-unimodal maps $f_t$ containing two kneading sequences also contains all intermediate kneading sequences. In particular, a family that contains the above maximal and the minimal kneading sequences, contains *all* admissible kneading sequences. Such a family is called *full*. We see that the real quadratic family $P_c, c \in [-2, 1/4]$, is full: *any S-unimodal map is topologically equivalent to some quadratic polynomial*. This indicates dynamical significance of the quadratic family.

We say that a one-parameter family of unimodal maps $f_t$ is *almost full* if it contains all admissible kneading sequences except possibly the minimal one.

## Universality Phenomenon

### Universal Geometry of Doubling Bifurcations and the Feigenbaum Attractor

Let us consider the real quadratic family $P_c : x \mapsto x^2 + c$, $c \in [-2, 1/4]$. As the parameter $c$ moves down from 1/4, we observe a sequence of doubling bifurcations $c_n$ where the attracting cycle of period $2^n$ gives birth to an attracting cycle of period $2^{n+1}, n = 0, 1, \ldots$ (*see* Holomorphic Dynamics and Figure 1). This sequence converges to the *Feigenbaum parameter* $c_\infty$ at exponential rate: $c_n - c_\infty \sim \lambda^{-n}$, where $\lambda \approx 4.6$. It turns out that if we consider a similar one-parameter family of unimodal maps, say $x \mapsto a \sin x$, we observe a similar sequence of doubling bifurcations converging to the limit exponentially at the *same* rate $\lambda^{-n}$, independently of the family under consideration.

In the dynamical space, let us consider the Feigenbaum attractor $A_f$ [1] of an infinitely renormalizable $S$-unimodal map $f$ that appear in the limit of doubling bifurcations (so that the periods of periodic intervals $J^n$ are equal to $2^n$). Let us consider the *scaling factors* $\sigma_n = |J^n|/|J^{n-1}|$. Then $\sigma_n \to \sigma_\infty$, where the limiting scaling factor $\sigma_\infty \approx 2.6$ is

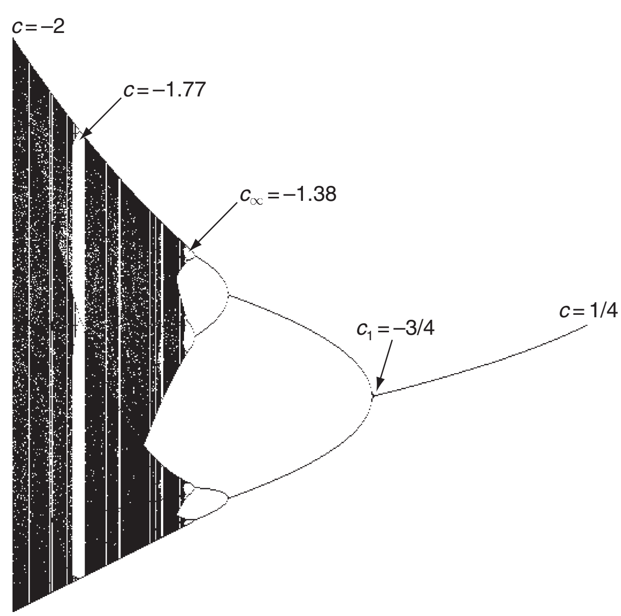



**Figure 1** Real quadratic family $P_c : x \mapsto x^2 + c$. This picture presents how the limit set of the orbit $\{P_c^n(0)\}_{n=0}^{\infty}$ bifurcates as the parameter $c$ changes from 1/4 on the right to −2 on the left. Three topological types of regimes are intertwined in an intricate way. The gaps correspond to the regular regimes. The black regions correspond to the chaotic regimes (though, of course, there are many narrow invisible gaps therein). In the beginning (on the right) one can see the cascade of doubling bifurcations. This picture became symbolic for one-dimensional dynamics.

*independent* of the particular map $f$ under consideration. Thus, the small-scale geometry of $A_f$ is universal.

This was historically the first observed manifestation of the *quantitative universality* of dynamical and parameter structures.

#### Feigenbaum–Coullet–Tresser Renormalization Conjecture

To explain the above universality phenomenon, Feigenbaum and independently Coullet and Tresser, formulated the following Renormalization Conjecture. Let us consider the space $\mathcal{U}$ of $S$-unimodal maps $f:[-1,1]\to[-1,1]$. A map $f\in\mathcal{U}$ is called *(doubling) renormalizable* if it has a cycle of intervals $J\to J_1\to J$ of period 2. Then, for any $n\in\mathbb{Z}_+\cup\{\infty\}$, we can naturally define $n$-times renormalizable maps, where $n=0$ corresponds to the *non-renormalizable* case, while $n=\infty$ corresponds to the *infinitely renormalizable* case.

Let $\mathcal{U}'\subset\mathcal{U}$ be the space of doubling renormalizable maps. If $f\in\mathcal{U}'$ then $f^2:J\to J$ is an $S$-unimodal map as well, and we define the *(doubling) renormalization operator* $R:\mathcal{U}'\to\mathcal{U}$ as the rescaling of this map:

$$Rf(x)=\sigma^{-1}f^2(\sigma x)$$

where $\sigma=|J|/2$.

The Renormalization Conjecture asserted that:

- The renormalization operator $R$ has a unique fixed point $f_*$, and this point is hyperbolic;
- the stable manifold $W^s(f_*)$ consists of infinitely renormalizable unimodal maps;
- the unstable manifold $W^u(f_*)$ is one dimensional and represents an almost full family of unimodal maps (see the section "Kneading theory"); and
- the quadratic family $\{P_c\}$ transversally intersects $W^s(f_*)$ (see Figure 2).

Assuming this conjecture, one can see that for any curve $t\mapsto g_t$ in $\mathcal{U}$ that transversally intersects the stable manifold $W^s(f_*)$ at some moment $t_*$, the doubling bifurcations parameters $t_n$ converge to $t_*$ at exponential rate $\lambda^{-n}$, where $\lambda$ is the unstable eigenvalue of the differential $DR(f_*)$. This explains the universal geometry of doubling bifurcations.

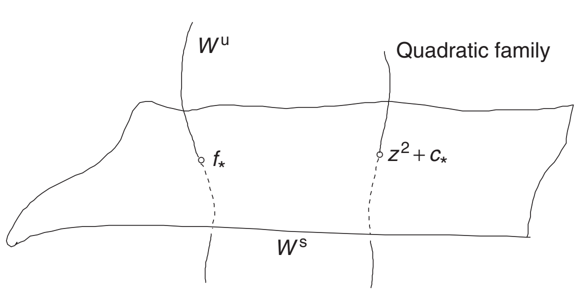


**Figure 2** Renormalization fixed point.

One can also show that the Feigenbaum attractor $A_f$ of any map $f\in W^s(f_*)$ is *smoothly equivalent* to $A_{f_*}$, which explains the universal small-scale geometry of these attractors.

#### Full Renormalization Horseshoe

Along with period doublings, one can consider period triplings, quadruplings, etc. A unimodal map $f\in\mathcal{U}$ is said to be *renormalizable with period* $p$ if it has a cycle of intervals $J\to J_1\to\cdots\to J_{p-1}\to J$ of period $p$. The corresponding renormalization operator is defined as $Rf(x)=\sigma^{-1}f^p(\sigma x)$, where $\sigma=|J|/2$.

The *combinatorics* or *type* $\tau$ of the renormalization operator is the order of the intervals $J_k, k=0,1,\ldots,p-1$, on the real line (up to reversal). (For instance, there are three admissible combinatorics $\tau$ of period 5.) If we want to specify combinatorics of the renormalization operator under consideration, we use notation $R_\tau$. This operator is defined on the "renormalization strip" $\mathcal{U}^\tau$ of unimodal maps $f\in\mathcal{U}$ that are renormalizable with combinatorics $\tau$.

The Renormalization Conjecture admits a straightforward generalization to any renormalization operator $R_\tau$. More interestingly, one can formulate a stronger version of it by putting all the admissible renormalization types together. Let $\mathcal{T}$ stand for the set of all *minimal* renormalization types, that is, the types that cannot be factored through other types. Then the renormalization strips $\mathcal{U}^\tau, \tau\in\mathcal{T}$, are pairwise disjoint, and we can define the *full renormalization operator*

$$R:\bigcup_{\tau\in\mathcal{T}}\mathcal{U}^\tau\to\mathcal{U} \qquad [2]$$

by letting $R|\mathcal{U}^\tau=R_\tau$. Then the strong version of the renormalization conjecture asserted that:

- there is an $R$-invariant hyperbolic subset $\mathcal{A}\subset\mathcal{U}$ called the *full renormalization horseshoe* such that the restriction $R|\mathcal{A}$ is topologically conjugate to the full shift $\sigma$ on the space $\Sigma$ of bi-infinite sequences $(\ldots,\tau_{-1},\tau_0,\tau_1,\ldots)$ of symbols $\tau_n\in\mathcal{T}$;
- for any $f_*\in\mathcal{A}$, the stable manifold $W^s(f_*)$ consists of infinitely renormalizable maps $f\in\mathcal{U}$ with the same combinatorics as $f_*$;
- for any $f_*\in\mathcal{A}$, the unstable manifold $W^u(f_*)$ is one-dimensional and represents an almost full family of unimodal maps; and
- the real quadratic family $\{P_c\}$ transversally intersects all stable manifolds $W^s(f_*)$.

## Complex Renormalization

### Polynomial-Like Maps

A *polynomial-like map* is a holomorphic branched covering of finite degree $f: U \to U'$, where $U \Subset U' \subset \mathbb{C}$ are topological disks (In other words, the maps $f$ is *proper*, that is, full preimages $f^{-1}(K)$ of compact sets $K \subset U'$ are compact). For instance, if $f$ is a polynomial of degree $d$ then for a sufficiently large radius $R > 0$, the map $f: f^{-1}(\mathbb{D}_R) \to \mathbb{D}_R$ is a polynomial-like map of the same degree $d$. We refer to such polynomial-like maps as "polynomials."

The *filled Julia set* of $f$ is the set of nonescaping points:

$$K(f) = \{z\colon f^n z \in U, n = 0, 1, \ldots\}$$

The Julia set of $f$ is the boundary of its filled Julia set: $J(f) = \partial K(f)$.

A polynomial-like map of degree $d$ has $d-1$ critical points counted with multiplicities. The Julia set (and the filled Julia set) is connected if and only if all the critical points $c_i$ are nonescaping, that is, $c_i \in K(f)$.

A polynomial-like map of degree 2 is called *quadratic-like*. The Julia set of a quadratic-like map is either connected or a Cantor set, depending on whether its critical point is nonescaping or otherwise.

The domain of a polynomial-like map is allowed to be slightly adjusted by taking $V'$ to be a topological disk such that $U \subset V' \subset U'$ and letting $V = f^{-1}(V')$. We say that two polynomial-like maps represent the same germ if one can be obtained from the other by a sequence of such adjustments.

We will be mostly interested in the quadratic case; so let $\mathcal{Q}$ be the space of quadratic-like germs considered up to affine conjugacy, and let $\mathcal{C}$ be the *connectedness locus* in $\mathcal{Q}$, that is, the subset of $f \in \mathcal{Q}$ with connected Julia set. The space $\mathcal{Q}$ has a natural complex analytic structure such that holomorphic curves in $\mathcal{Q}$ are represented by holomorphic families $f_\lambda(z)$ of quadratic-like maps.

Two polynomial-like maps are called *hybrid* equivalent if they are conjugate by a quasiconformal map $h$ such that $\bar{\partial} h = 0$ a.e. on $K(f)$ (in particular, $h$ is conformal on int $K(f)$). By the *Straightening Theorem*, any polynomial-like map is hybrid equivalent (after an adjustment of its domain) to a polynomial of the same degree (called the "straightening" of $f$). The straightening depends only on the germ of $f$.

For a quadratic-like map $f$ with connected Julia set, the straightening $P_c: z \mapsto z^2 + c$ is *unique*, $c = \chi(f)$. Thus, we obtain the straightening map $\chi: \mathcal{C} \to M$, where $M$ is the Mandelbrot set (*see* Holomorphic Dynamics). We let $\mathcal{H}_c = \chi^{-1}(c)$ be the hybrid class passing through a point $c \in M$. One can show that $\mathcal{H}_c$ is a *codimension-one submanifold* in $\mathcal{Q}$.

Any quadratic-like map has two fixed points counted with multiplicity. In the case of connected Julia set, these fixed points have a different dynamical meaning: one of them, called $\alpha$, is either attracting, or neutral, or repelling *separating*, that is, $J(f) \setminus \{\alpha\}$ is disconnected. Another one, called $\beta$, is either parabolic with multiplier 1 (and then it coincides with $\alpha$) of repelling *nonseparating*.

In what follows, we normalize quadratic-like maps so that 0 is their critical point.

### Complex Renormalization and Little Mandelbrot Sets

A quadratic-like map $f: U \to U'$ with connected Julia set is called *renormalizable* if there is a topological disk $V \ni 0$ and a natural number $p \geq 2$ called the *renormalization period* such that:

- letting $g = f^p | V$ and $V' = g(V)$, the map $g: V \to V'$ is quadratic-like;
- the *little Julia set* $K(g)$ is connected; and
- the sets $g^n(K(g)), n = 1, \ldots, p-1$, can intersect $K(g)$ only at the $\beta$-fixed point of $g$.

Under these circumstances, the quadratic-like germ $g$ considered up to affine conjugacy is called the *renormalization* of the quadratic-like germ $f$; $g = Rf$. Moreover, one says that $f$ is *primitively* renormalizable if the little Julia sets $g^n(K(g)), n = 1, \ldots, p-1$, are pairwise disjoint. Otherwise, $f$ is *satellite* renormalizable.

As in the unimodal case, one can define *combinatorics* or *type* $\tau$ of the complex renormalization. Roughly speaking, renormalizable maps with the same combinatorics have the same renormalization period and the "same position" of the little Julia sets $f^k(K(g))$ in $\hat{\mathbb{C}}$ (the rigorous definition is based on the notion of Thurston's equivalence from Holomorphic Dynamics).

**Theorem 1** (Douady and Hubbard 1986). *The set of parameters $c$ for which a quadratic map $P_c: z \mapsto z^2 + c$ is renormalizable with a given combinatorics $\tau$ assemble a homeomorphic copy $M^\tau$ of the Mandelbrot set $M$.*

This theorem explains the presence of many little Mandelbrot sets that are observable on the computer pictures of $M$ (see Figures 3 and 4). Moreover, the copies corresponding to the primitive renormalization originate at primitive hyperbolic components (*see* Holomorphic Dynamics), while the copies obtained by a satellite renormalization originate at satellite hyperbolic components attached to some

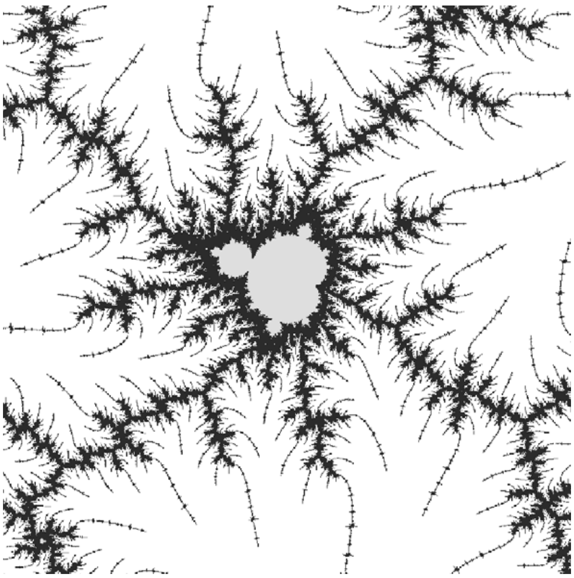


**Figure 3** A primitive copy of the Mandelbrot set.

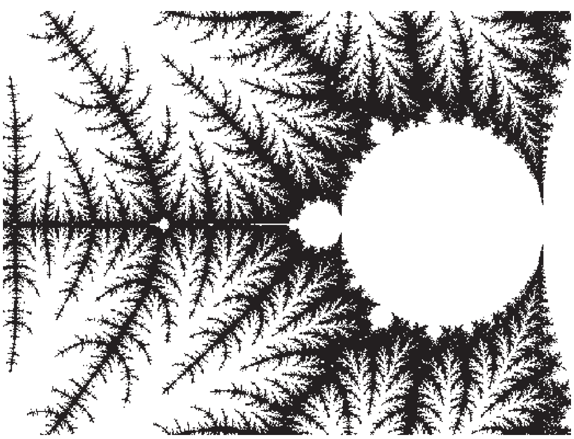


**Figure 4** The satellite copy of the Mandelbrot set attached to the main cardioid at the point of doubling bifurcation.

"mother" hyperbolic component. (Satellite copies attached to the main cardioid are particularly prominent on the pictures of $M$.)

Given a combinatorial type $\tau$, the set $\mathcal{Q}^\tau$ of quadratic-like germs $f \in \mathcal{Q}$ that are renormalizable with combinatorics $\tau$ (the complex renormalization strip) is the union of hybrid classes passing through the little copy $M^\tau$. As in the real case, let us consider the set $\mathcal{T}_{\mathbb{C}}$ of all minimal combinatorial types. Then the corresponding renormalization strips $\mathcal{Q}^\tau$ are pairwise disjoint, and we can define the *full* complex renormalization operator $R : \bigcup_{\tau \in \mathcal{T}_{\mathbb{C}}} \mathcal{Q}^\tau \to \mathcal{Q}$.

## Renormalization Theorem

The first proof of the Renormalization Conjecture in the period-doubling case was based on rigorous computer estimates (Lanford 1982). It followed, in the 1980s, by works of Epstein, Eckmann, Khanin, Sinai, among others, which gave a better conceptual understanding and provided proofs of many ingredients of the picture (without computer assistance).

The turning point in this development occurred when methods of holomorphic dynamics and conformal geometry were introduced into the subject (Douady and Hubbard 1985, Sullivan 1986). This led to the proof of the renormalization conjecture in the space of quadratic-like germs:

**Theorem 2** (Sullivan–McMullen–Lyubich, the 1990s). *For any real combinatorics $\tau \in \mathcal{T}$, the operator $R_\tau$ has a unique fixed point $f_\tau$ in the space $\mathcal{Q}$. Moreover, $f_\tau$ is hyperbolic, its stable manifold $W^s(f_\tau)$ coincides with the hybrid class $\mathcal{H}_c, c = \chi(f_\tau)$, while the real slice of the unstable manifold represents an almost full family of unimodal maps.*

This result was further extended to the smooth category by de Faria, de Melo, and Pinto.

## MLC, Density of Hyperbolicity, and Geometry of Feigenbaum Julia Sets

The "Mandelbrot set is locally connected" (MLC) conjecture (*see* Holomorphic Dynamics) is intimately related to the renormalization phenomenon. This connection was first revealed by the following result:

**Theorem 3** (Yoccoz 1990, unpublished). *Let us consider a nonrenormalizable quadratic polynomial $P_c : z \mapsto z^2 + c$ with connected Julia set and both fixed points repelling. Then the Julia set $J(P_c)$ is locally connected and the Mandelbrot set is locally connected at $c$.*

This result was recently extended to higher-degree unicritical polynomials $z \mapsto z^d + c$ (Kahn–Lyubich, preprint 2005).

The MLC Conjecture is still open for general infinitely renormalizable parameters. However, the similar problem for the real quadratic family has been resolved. It implies the real version of the Fatou conjecture in the quadratic case (*see* Holomorphic Dynamics):

**Theorem 4** (Lyubich 1997). *Hyperbolic maps are dense in the real quadratic family.*

This result was recently extended to higher-degree polynomials by Kozlovskii, Shen, and van Strien (preprint 2003).

Infinitely renormalizable quadratic maps of bounded combinatorial type (i.e., with bounded relative periods $p_{n+1}/p_n$) supply us with a rich class of fractals with very interesting geometry. These

Julia sets are "hairy" at the origin, that is, their blow-ups fill in densely the whole plane (this phenomenon is related to the universal geometry of the Feigenbaum attractors; McMullen (1996)). However, some of them have zero Lebesgue measure (Yarrington, thesis 1995) and Hausdorff dimension smaller than 2 (Avila–Lyubich, preprint 2004). It is unknown whether this happens for all of them or not (in particular, the answer is unknown for the Feigenbaum map born in the cascade of doubling bifurcations).

## Regular or Stochastic Dichotomy

### Stochastic Maps

An $S$-unimodal map $f$ is called *stochastic* if it has an absolutely continuous invariant measure $\mu$. In this case, $f$ is topologically chaotic (see the section "Topological dynamics") and $\mu$ is supported on the transitive cycle of intervals $\cup J_k$. Moreover, $\mu$ has a positive characteristic exponent,

$$\chi = \int \log|Df| \mathrm{d}\mu > 0$$

and Lebesgue almost all orbits are equidistributed with respect to $\mu$, that is, for Lebesgue a.e. $x \in I$,

$$\frac{1}{n}\sum \phi(f^n x) \to \int \phi \, \mathrm{d}\mu$$

for any continuous function $\phi$. The map $f^p \mid J$ is mixing with respect to $\mu$, and in fact, is *weakly Bernoulli*.

Here are two important criteria for stochasticity:

- *Collet–Eckmann condition* (*see* Holomorphic Dynamics). These maps have extra strong stochastic properties, notably, the exponential decay of correlations.
- *Martens–Nowicki condition*. To state it, we need to define the principal nest of intervals, $I^0 \supset I^1 \supset \cdots \ni 0$. Here $I^0 = [-\alpha, \alpha]$, where $\alpha$ is the fixed point with negative multiplier, and $I^{n+1}$ is inductively defined as the component of $f^{-l_n}(I^n)$ containing 0, where $l_n$ is the moment of first return of the orbit of 0 to $I^n$. Let us consider the scaling factors $\sigma_n = |I^n|/|I^{n-1}|$. If $\sum \sqrt{\sigma_n} < \infty$ *then* $f$ *is stochastic*.

Let $\mathcal{N} \subset [-2, 1/4]$ be the set of parameters $c$ for which the quadratic map $P_c$ is topologically chaotic. Not every such map is stochastic. However, *the set of stochastic parameters has positive Lebesgue measure* (Jakobson 1981), and in fact,

**Theorem 5** (Lyubich 2000). *For a.e.* $c \in \mathcal{N}$, *the map* $P_c$ *satisfies the Martens–Nowicki condition, and thus, is stochastic.*

Avila and Moreira (2005) went on to prove that *for a.e.* $c \in \mathcal{N}$, *the map* $P_c$ *is Collet–Eckmann.*

### Renormalization Horseshoe

Let us consider the complexification of the renormalization operator [2],

$$R : \bigcup_{\tau \in \mathcal{T}} \mathcal{Q}^\tau \to \mathcal{Q} \qquad [3]$$

acting in the space of quadratic-like maps.

**Theorem 6** (Lyubich 2002). *The "Strong Renormalization Conjecture" is valid for the operator* [3].

Let $\mathcal{I} \subset [-2, 1/4]$ be the set of parameters for which the quadratic map $P_c$ is infinitely renormalizable. The above theorem implies that this set has *zero Lebesgue measure*. (Avila and Moreira went on to prove that $\mathrm{HD}(\mathcal{I}) < 1$.)

### Regular or Stochastic Dichotomy

Putting together Theorems 5 and 6, we obtain:

**Theorem 7** *For a.e.* $c \in [-2, 1/4]$, *the quadratic map* $P_c$ *is either regular or stochastic.*

This result gives a complete probabilistic picture of dynamics in the real quadratic family. It has been later transferred to any nondegenerate real analytic family of $S$-unimodal maps (Avila–Lyubich–de Melo), and further to a generic smooth family of $S$-unimodal maps (Avila–Moreira).

Palis has formulated a strong general conjecture (in all dimensions) asserting that a typical (from the probabilistic point of view) smooth dynamical system $f$ has finitely many attractors supporting SRB measures (*see* Lyapunov Exponents and Strange Attractors) that govern the behavior of Lebesgue a.e. trajectories of $f$. The above results confirm the *Palis Conjecture* in the setting of $S$-unimodal maps.

## Other Universality Classes

From a more general point of view, renormalization is an appropriately rescaled return map to a relevant piece of the phase space, viewed as an operator in some class of dynamical systems. From this point of view, most dynamical systems are "renormalizable," and the renormalization approach often provides a deep insight into the nature of the systems in question.

Here is a partial list of classes of nonlinear systems that exhibit universality with an underlying renormalization mechanism (we provide a few

relevant names, but there are many more people who contributed to the corresponding theories):

- Holomorphic germs near indifferent equilibria (Yoccoz, Shishikura, McMullen);
- critical circle maps (Kadanoff, Feigenbaum, Rand, Lanford, Swiatek, de Faria, Yampolsky);
- non-renormalizable quadratic-like maps of Fibonacci type (Lyubich–Milnor);
- conservative two-dimensional diffeomorphisms near the point of breaking of KAM tori (MacKay, Koch); and
- dissipative Hénon-like maps (Collet–Eckmann–Koch, de Carvalho–Lyubich–Martens).

*See also:* Fractal Dimensions in Dynamics; Holomorphic Dynamics; Lyapunov Exponents and Strange Attractors; Multiscale Approaches.

## Further Reading

Collet P and Eckmann J-P (1980) *Iterated Maps of the Interval as Dynamical Systems*. Boston: Birkhäuser.

Cvitanović P (1984) *Universality in Chaos*. Bristol: Adam Hilger.

Douady A and Hubbard JH (1985) On the dynamics of polynomial-like maps. *Annales Scientifiques de l'École Normale Supérieure* 18: 287–343.

Lyubich M (1999) Feigenbaum–Coullet–Tresser universality and Milnor's hairiness conjecture. *Annals of Mathematics* 149: 319–420.

Lyubich M (2000) Quadratic family as a qualitatively solvable model of dynamics. *Notices of the American Mathematical Society* 47(9): 1042–1052.

McMullen C (1996) *Renormalization and 3-Manifolds Which Fiber Over the Circle*, Annals of Math. Studies, vol. 135. Princeton: Princeton University Press.

de Melo W and van Strien S (1993) *One-Dimensional Dynamics.*, Berlin: Springer.

Sullivan D (1993) Linking the universalities of Milnor–Thurston, Feigenbaum and Ahlfors–Bers. In: *Topological Methods in Modern Mathematics*, The Proceedings of Symposium held in honor of John Milnor's 60th Birthday, SUNY at Stony Brook, 1991, pp. 543–564. Houston, TX: Publish or Perish.

Vul EB, Sinai YaG, and Khanin KM (1984) Feigenbaum universality and the thermodynamical formalism. *Russian Mathematical Surveys* 39: 1–40.

# V

## Variational Methods in Turbulence

**F H Busse**, Universität Bayreuth, Bayreuth, Germany



### Introduction

The problem of fluid turbulence is commonly regarded as one of the most challenging problems of theoretical physics and mathematics. There is general agreement that the Navier–Stokes equations (NSEs) provide a satisfactory basis for the description of turbulent motions of homogeneous Newtonian fluids such as gases and most liquids. But the difficulty of generating solutions of these equations for high-Reynolds-number flows has prevented accurate answers to simple questions such as the question of the discharge of turbulent pipe flow as a function of the pressure head or the question of the heat transport by turbulent convection in a fluid layer heated from below. In view of this difficulty, it has become an attractive idea to obtain rigorous bounds on turbulent transports. Variational methods have played an important role in the derivation of such bounds.

There is another motivation for the use of variational methods for the understanding of turbulent fluid systems. Experimenters have sometimes noted the tendency of turbulent flows to maximize transports under given external conditions. In his pioneering paper, Howard (1963) mentions that the Malkus hypothesis of a maximum heat transport by thermal convection had motivated him to derive upper bounds through the use of variational methods. The techniques developed by Howard have later been applied to other kinds of turbulent transports by Busse. While relatively simple ordinary differential equations are obtained when the equation of continuity is not imposed as a constraint, the Euler–Lagrange equations for a stationary value of the variational functional lead to nonlinear partial differential equations when solenoidal extremalizing vector fields are required. Nevertheless, using boundary layer methods one can derive approximate analytical solutions even in the limit of asymptotically large Rayleigh and Reynolds numbers (Busse 1969, 1978).

In the following, we shall first discuss the energy method which provides necessary conditions for the existence of turbulent solutions of the underlying equations and then turn to the problem of upper bounds for the turbulent momentum transport in the plane Couette flow configuration as a particular example. The properties and physical relevance of the extremalizing vector fields will be discussed in a final section.

### Energy Method

For simplicity, we consider the NSEs for a homogeneous incompressible fluid with a constant kinematic viscosity $\nu$ in an arbitrary fixed domain $\mathcal{D}$. Using the diameter $d$ of the domain as length scale and $d^2/\nu$ as timescale, we can write the NSEs of motion in dimensionless form,

$$\frac{\partial}{\partial t}\boldsymbol{v} + \boldsymbol{v}\cdot\nabla\boldsymbol{v} = -\nabla p + \boldsymbol{f} + \nabla^2\boldsymbol{v} \qquad [1a]$$

$$\nabla\cdot\boldsymbol{v} = 0 \qquad [1b]$$

where $\boldsymbol{f}$ denotes some given steady distribution of a force density. On the boundary $\partial\mathcal{D}$ of the domain $\mathcal{D}$, steady velocities parallel to the boundary may be specified. We assume that the basic steady solution of the problem is given by $\boldsymbol{v}_s = Re\,\hat{\boldsymbol{v}}$ where the average of $(\hat{\boldsymbol{v}})^2/2$ over the domain $\mathcal{D}$ (indicated by angular brackets) is unity, $\langle|\hat{\boldsymbol{v}}|^2\rangle = 2$. Any velocity field $\boldsymbol{v}_t$ different from $\boldsymbol{v}_s$, that is, with $\boldsymbol{u} \equiv \boldsymbol{v}_t - \boldsymbol{v}_s \not\equiv 0$, must obey the equations

$$\frac{\partial}{\partial t}\boldsymbol{u} + \boldsymbol{v}_s\cdot\nabla\boldsymbol{u} + \boldsymbol{u}\cdot\nabla\boldsymbol{v}_s + \boldsymbol{u}\cdot\nabla\boldsymbol{u} = -\nabla\tilde{p} + \nabla^2\boldsymbol{u} \qquad [2a]$$

$$\nabla\cdot\boldsymbol{u} = 0 \qquad [2b]$$

together with the homogeneous boundary conditions for $\boldsymbol{u}$ on $\partial\mathcal{D}$. By multiplying eqn [2a] by $\boldsymbol{u}$ and averaging the result over the domain $\mathcal{D}$ we obtain the relationship

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\langle\boldsymbol{u}\cdot\boldsymbol{u}\rangle = -\langle|\nabla\boldsymbol{u}|^2\rangle - Re\langle\boldsymbol{u}\cdot(\boldsymbol{u}\cdot\nabla)\hat{\boldsymbol{v}}\rangle \qquad [3]$$

where the vanishing of $\boldsymbol{u}$ on $\partial\mathcal{D}$ and equations such as

$$\langle \boldsymbol{u}\cdot(\boldsymbol{v}_\mathrm{s}\cdot\nabla)\boldsymbol{u}\rangle = \tfrac{1}{2}\langle \boldsymbol{v}_\mathrm{s}\cdot\nabla\boldsymbol{u}\cdot\boldsymbol{u}\rangle = \tfrac{1}{2}\langle\nabla\cdot(\boldsymbol{v}_\mathrm{s}\boldsymbol{u}\cdot\boldsymbol{u})\rangle = 0$$

have been used to prove that the terms $\boldsymbol{v}_\mathrm{s}\cdot\nabla\boldsymbol{u}$, $\boldsymbol{u}\cdot\nabla\boldsymbol{u}$ and $\nabla\tilde{p}$ do not enter the balance [3]. This balance is called the Reynolds–Orr energy equation and is the basis for the application of the energy method. The lowest value $Re$ for which the right-hand side of [3] is non-negative is called the energy Reynolds number $Re_\mathrm{E}$. For $Re < Re_\mathrm{E}$ the steady solution $\boldsymbol{v}_\mathrm{s}$ is absolutely stable and the energy of any disturbance $\boldsymbol{u}$ must decay exponentially in time. $Re > Re_\mathrm{E}$ is a necessary condition for the existence of a persistent turbulent state of fluid flow. $Re_\mathrm{E}$ is determined as the solution of the variational problem:

*For a given flow $\hat{\boldsymbol{v}}$ in $\mathcal{D}$ find the minimum $Re_\mathrm{E}$ of the functional*

$$\mathcal{R}_\mathrm{E} \equiv \frac{\langle|\nabla\breve{\boldsymbol{u}}|^2\rangle}{\langle-\breve{\boldsymbol{u}}\cdot(\breve{\boldsymbol{u}}\cdot\nabla)\hat{\boldsymbol{v}}\rangle} \qquad [4]$$

*among all vector fields $\breve{\boldsymbol{u}}$ which satisfy the conditions $\nabla\cdot\breve{\boldsymbol{u}}=0$ in $\mathcal{D}$, $\breve{\boldsymbol{u}}=0$ on $\partial\mathcal{D}$, and $\langle\breve{\boldsymbol{u}}\cdot(\breve{\boldsymbol{u}}\cdot\nabla)\hat{\boldsymbol{v}}\rangle < 0$.*

For $Re \geq Re_\mathrm{E}$ there will exist at least one vector field $\boldsymbol{u}$, namely the minimizing solution $\breve{\boldsymbol{u}}$ of the variational problem [4], the energy of which does not decay, at least not initially. In the derivation of the Euler–Lagrange equations as necessary conditions for stationary values of the variational functional [4],

$$\tfrac{1}{2}G(\breve{u}_\kappa\partial_\kappa\hat{v}_i + \breve{u}_\kappa\partial_i\hat{v}_\kappa) = -\partial_i\breve{\pi} + \partial_\kappa\partial_\kappa\breve{u}_i \qquad [5a]$$

$$\partial_\kappa\breve{u}_\kappa = 0 \qquad [5b]$$

the constraint $\nabla\cdot\breve{\boldsymbol{u}}=0$ has been taken into account through the Lagrange multiplying function $\breve{\pi}$. $G$ is a stationary value of the functional [4] and in general there exist many of those which are determined as eigenvalues of the linear boundary value problem [5] together with its boundary condition $\breve{u}_i=0$ on $\partial\mathcal{D}$. Only the infinum of all $G$ provides the energy Reynolds number $Re_\mathrm{E}$. Many details on the energy method can be found in Joseph's book (1976). Here we just wish to remark that the Reynolds–Orr balance [3] remains valid when the problem is considered in a system rotating with a constant angular velocity $\Omega_D$ since the Coriolis force does not contribute to the energy balance [3]. The values of $Re_\mathrm{E}$ are usually much smaller than the critical values $Re_\mathrm{c}$ for the onset of infinitesimal disturbances as can be seen from Table 1. Here the experimentally determined values

**Table 1** Reynolds numbers for shear flows

| | $Re_\mathrm{E}$ | $Re_\mathrm{G}$ (from exp.) | $Re_\mathrm{c}$ |
|---|---|---|---|
| Plane Couette flow | 82.6 | ≈1300 | ∞ |
| Poiseuille flow (channel flow) | 99.2[a] | ≈2000[a] | 5772[a] |
| Hagen–Poiseuille flow (pipe flow) | 81.5[a] | ≈2100[a] | ∞ |
| Circular Couette flow with $\Omega_D = Re_\mathrm{E}/2$ | 82.6 | ≈82.6 | 82.6 |

[a] The maximum velocity and the channel width $d$ (radius $d$ in the case of pipe flow) have been used in definition of $Re$.

$Re_\mathrm{G}$ for the instability of the basic flow state have also been listed. A unique situation occurs in the small gap limit of the Taylor–Couette system where $Re_\mathrm{E}$ and $Re_\mathrm{c}$ coincide for a special value of the dimensionless mean rotation rate $\Omega_D$ (Busse 2002).

## Variational Problem for Turbulent Momentum Transport

In order to introduce the variational method for bounds on turbulent transports we consider the simplest configuration for which a nontrivial solution of the NSEs of motion exists: the configuration of plane Couette flow (Figure 1). The Reynolds number is defined in this case in terms of the constant relative motion $U_0\boldsymbol{i}$ between the plates, $Re = U_0 d/\nu$, where $\boldsymbol{i}$ is the unit vector parallel to the plates and $\nu$ is the kinematic viscosity of the fluid. Using the distance $d$ between the plates as length scale and $d^2/\nu$ as timescale, the basic equations can be written in the form

$$\frac{\partial}{\partial t}\boldsymbol{v} + \boldsymbol{v}\cdot\nabla\boldsymbol{v} = -\nabla p + \nabla^2\boldsymbol{v} \qquad [6]$$

$$\nabla\cdot\boldsymbol{v} = 0 \qquad [7]$$

We use a Cartesian system of coordinates with the $x$, $z$-coordinates in the directions of $\boldsymbol{i}$ and $\boldsymbol{k}$,

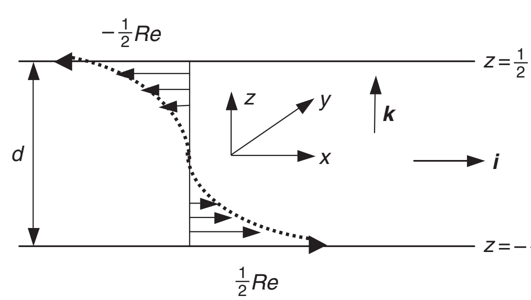



**Figure 1** Geometrical configuration of the plane Couette flow problem.

respectively, where $\boldsymbol{k}$ is the unit vector normal to the plates such that the boundary conditions are given by

$$\boldsymbol{v} = \mp \frac{1}{2} Re\, \boldsymbol{i} \quad \text{at } z = \pm \frac{1}{2} \qquad [8]$$

After separating the velocity field $\boldsymbol{v}$ into its mean and fluctuating parts, $\boldsymbol{v} = U + \check{\boldsymbol{v}}$ with $\bar{\boldsymbol{v}} = U$, $\bar{\check{\boldsymbol{v}}} = 0$, where the bar denotes the average over planes $z = \text{const.}$, we obtain by multiplying eqn [6] by $\check{\boldsymbol{v}}$ and averaging it over the entire fluid layer (indicated by angular brackets)

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\langle|\check{\boldsymbol{v}}|^2\rangle = -\left\langle \overline{\boldsymbol{u}w} \cdot \frac{\partial}{\partial z} U \right\rangle - \langle|\nabla\check{\boldsymbol{v}}|^2\rangle \qquad [9]$$

Here $\boldsymbol{u}$ denotes the component of $\check{\boldsymbol{v}}$ perpendicular to $\boldsymbol{k}$ and $w$ is its $z$-component. We define fluid turbulence under stationary conditions by the property that quantities averaged over planes $z = \text{const.}$ are time independent. Accordingly, the equation for the mean flow $U$ can be integrated to yield

$$\frac{\mathrm{d}}{\mathrm{d}z} U = \overline{w\boldsymbol{u}} - \langle w\boldsymbol{u}\rangle - Re\, \boldsymbol{i} \qquad [10]$$

where the boundary condition [8] has been employed. With this relationship, $U$ can be eliminated from the problem and the energy balance

$$\langle|\nabla\boldsymbol{u}|^2\rangle + \langle|\overline{\boldsymbol{u}w} - \langle\boldsymbol{u}w\rangle|^2\rangle = Re\langle u_x w\rangle \qquad [11]$$

is obtained where the identity $\langle\overline{\boldsymbol{u}w}^2\rangle - \langle\boldsymbol{u}w\rangle^2 = \langle|\overline{\boldsymbol{u}w} - \langle\boldsymbol{u}w\rangle|^2\rangle$ has been used.

Since the momentum transport in the $x$-direction between the moving rigid plates is described by $M = -\mathrm{d}U_x/\mathrm{d}z\,|_{z=\pm 1/2} = \langle u_x w\rangle + Re$, we can conclude immediately that the momentum transport by turbulent flow always exceeds the corresponding laminar value because $\langle u_x w\rangle$ is positive according to the relationship [11]. Since a lower bound on $M$ thus exists, an upper bound $\mu$ on $\langle u_x w\rangle$ as a function of $Re$ is of primary interest. Following Howard (1963), it can be shown that $\mu(Re)$ is a monotonous function and it is therefor equivalent to ask for a lower bound $R$ of $Re$ at a given value $\mu$ of $\langle u_x w\rangle$. We are thus led to the following formulation of the variational problem:

*Find the minimum $R(\mu)$ of the functional*

$$\mathcal{R}(\tilde{\boldsymbol{v}}, \mu) \equiv \frac{\langle|\nabla\tilde{\boldsymbol{v}}|^2\rangle}{\langle\tilde{u}_x\check{w}\rangle} + \mu\frac{\langle|\overline{\tilde{\boldsymbol{u}}\tilde{w}} - \langle\tilde{\boldsymbol{u}}\tilde{w}\rangle|^2\rangle}{\langle\tilde{u}_x\tilde{w}\rangle^2} \qquad [12]$$

*among all solenoidal vector fields $\tilde{\boldsymbol{v}} \equiv \tilde{\boldsymbol{u}} + \boldsymbol{k}\tilde{w}$ (with $\tilde{\boldsymbol{u}} \cdot \boldsymbol{k} = 0$) that satisfy the boundary condition $\tilde{\boldsymbol{v}} = 0$ at $z = \pm 1/2$ and the condition $\langle\tilde{u}_x\tilde{w}\rangle > 0$.*

The Euler–Lagrange equations as necessary conditions for an extremal value of the functional are given by

$$\tilde{w}\frac{\mathrm{d}}{\mathrm{d}z}U^* + \boldsymbol{k}\tilde{\boldsymbol{u}} \cdot \frac{\mathrm{d}}{\mathrm{d}z}U^* = -\nabla\pi + \nabla^2\tilde{\boldsymbol{v}} \qquad [13]$$

$$\nabla \cdot \tilde{\boldsymbol{v}} = 0 \qquad [14]$$

where $\mathrm{d}U^*/\mathrm{d}z$ is defined by

$$\frac{\mathrm{d}}{\mathrm{d}z}U^* = \overline{\tilde{\boldsymbol{u}}\tilde{w}} - \langle\tilde{\boldsymbol{u}}\tilde{w}\rangle - \boldsymbol{i}\left(R - \frac{\langle|\nabla\tilde{\boldsymbol{v}}|^2\rangle}{2\langle\tilde{\boldsymbol{v}}_x\tilde{w}\rangle}\right) \qquad [15]$$

and where $\mu = \langle\tilde{\boldsymbol{u}}_x\tilde{w}\rangle$ has been set. When eqns [13]–[15] are compared with the equations for $\check{\boldsymbol{v}}$ and for $\boldsymbol{U}$, a strong similarity can be noticed. The variational problem does not exhibit any time dependence, but the Euler–Lagrange equations may still be regarded as the symmetric analogue of the NSEs for steady flow.

## Upper Bounds on the Turbulent Momentum Transport

A simple analytical solution of the variational problem can be obtained when the constraint $\nabla \cdot \tilde{\boldsymbol{v}} = 0$ is dropped. In that case it is evident that the minimum of the functional [12] is reached when $\tilde{\boldsymbol{v}}$ is independent of $x$, $y$, and when $\tilde{u}_x = \tilde{w} = f(z)$ holds. The Euler–Lagrange equations then assume the form of an ordinary differential equation,

$$f'' = [\mu(f^2/\langle f^2\rangle - 1) - R + \langle f'^2\rangle/\langle f^2\rangle]f \qquad [16]$$

Since the variational functional [12] is homogeneous in $\tilde{\boldsymbol{v}}$, we are free to use a normalization condition for which we choose $\max[f(z)] = 1$. Multiplication of eqn [16] by $f'$ and integration yield

$$f'^2 = \frac{\mu}{2k^2\langle f^2\rangle}(1 - k^2f^2)(1 - f^2)$$
$$\text{with } k^2 = \mu/[2(R+\mu)\langle f^2\rangle - 2\langle f'^2\rangle - \mu] \qquad [17]$$

This equation can be solved in terms of elliptical integrals. The minimum $R(\mu)$ is determined by the relationships

$$R = \frac{8}{3}[K^2(1+k^2) + K^3/D - 3k^2KD)]$$
$$\mu = 8k^2\,KD \qquad [18]$$

where $D(k)$ and $K(k)$ are the complete elliptical integrals usually labeled by these letters. For details, see the analysis by Howard (1963) of an analogous problem. In the asymptotic case of large

Reynolds numbers, relationships [18] yield the upper bound

$$\mu(Re) = \frac{8}{128} Re^2 \qquad [19]$$

In solving the full eqns [13]–[15], it is convenient to eliminate eqn [14] through the general representation of the solenoidal vector field $\tilde{\boldsymbol{v}}$,

$$\tilde{\boldsymbol{v}} = \nabla \times (\nabla\phi \times \boldsymbol{k}) + \nabla\psi \times \boldsymbol{k} \qquad [20]$$

We assume that the minimizing vector field $\tilde{\boldsymbol{v}}$ does not depend on $x$, although a rigorous proof for this property can be given only for small values of $\mu$. Introducing the notations $\theta \equiv \partial\psi/\partial y$ and $w \equiv -\partial^2\phi/\partial y^2$ we are thus led to the general ansatz

$$w = w^{(N)} \equiv \sum_{n=1}^{N} \alpha_n^{-2} w_n(z)\phi_n(y) \qquad [21a]$$

$$\theta = \theta^{(N)} \equiv \sum_{n=1}^{N} \theta_n(z)\phi_n(y) \qquad [21b]$$

where $N$ may tend to infinity and the functions $\phi_n(y)$ satisfy the equation

$$\frac{\partial^2}{\partial y^2}\phi_n = -\alpha_n^2\phi_n \qquad [22]$$

In the following, it will be assumed that the positive wavenumbers $\alpha_n$ are ordered according to their size, $\alpha_{n-1} < \alpha_n < \alpha_{n+1}$. The solutions of the form [21] of the Euler–Lagrange equations exhibit a boundary layer structure for large $\mu$ as sketched in Figure 2. Accordingly, the $N$–$\alpha$ solutions are characterized by a hierarchy of $N$ boundary layers at each plate and provide the upper bound sequentially with increasing $\mu$ starting with $N = 1$. The extremalizing vector fields thus exhibit a bifurcation structure similar to that found in many cases of the transition to turbulence. The thicknesses of the boundary layers decrease with increasing $\mu$ and their ratio from one layer to the next approaches the factor 4 as indicated in Figure 3. The typical scale of motion increases linearly with

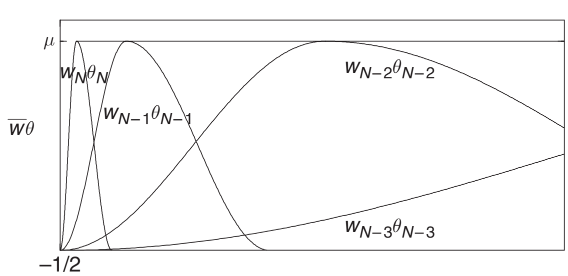



**Figure 2** Qualitative sketch of the boundary layer structure of the extremalizing $N-\alpha$ solution.

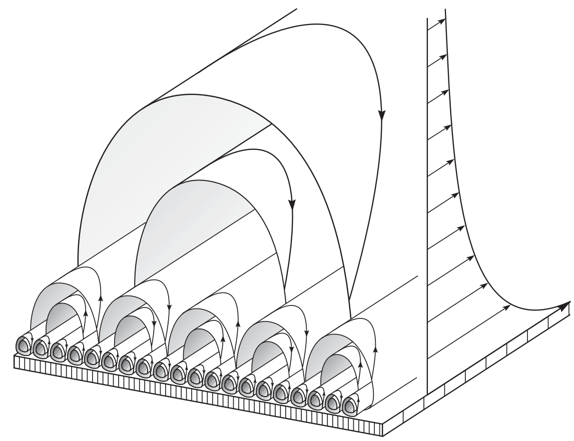


**Figure 3** Qualitative sketch of the nested boundary layers that characterize the vector field of maximum transport. The profile of the mean shear is shown on the right side.

distance from the wall as assumed in Prandtl's mixing-length theory. But the discreteness of the scales reflects the fact that effective transports require preferred scales. Asymptotically, the upper bound for the momentum transport approaches

$$\mu(Re) = 0.010\, Re^2 \qquad [23]$$

which represents a significant improvement over the relationship [19]. Nevertheless, the upper bound still exceeds the measured values of the momentum transport by more than a factor 10.

## Discussion

Bounds like those for the momentum transport have been obtained for many other kinds of turbulent transports. For details we refer to the review articles listed below. Usually, the formulation of the upper bound problem requires that the external conditions are homogeneous in two spatial dimensions such that a separation of the turbulent velocity, temperature, or magnetic fields into mean and fluctuating parts is possible. In this respect, the variational methods for upper bounds are more restricted than those used for determination of the energy Reynolds number $Re_E$. The latter problem, incidentally, corresponds to the limit $\mu \to 0$ of variational problems of the type [12] as can be seen from a comparison with expression [4].

In recent years, the background field method has been introduced by Doering and Constantin (1994) as an alternative way for obtaining bounds on properties of turbulent flows. When optimized, it becomes equivalent to the variational method discussed in this article as has been demonstrated by Kerswell (1998). The fact that not optimized bounds can be obtained

relatively easily emphasizes the point that the extremalizing vector fields are the most interesting aspect of the variational problems. They often exhibit similarities with the observed turbulent velocity fields, in particular as far as the mean flows are concerned. In the case of convection in a layer heated from below, the transition of the bound from the $1-\alpha$ solution to the $2-\alpha$ solution corresponds closely to the experimentally observed transition from convection rolls to bimodal convection (Busse 1969).

The close similarities between variational functionals for rather different physical systems suggest corresponding similarities between the respective turbulent fields. For example, the analogy between the fluctuating component of the temperature in turbulent convection and the streamwise component of the fluctuating velocity field in shear flow turbulence has been demonstrated and employed in a theory of the atmospheric boundary layer (Busse 1978). Better bounds and more physically realistic properties of the extremalizing vector fields can be expected when additional constraints are imposed. For example, the energy balances for poloidal and toroidal components of the velocity field can be applied separately. But these developments are still in their initial stages.

*See also:* Bifurcations in Fluid Dynamics; Fluid Mechanics: Numerical Methods; Turbulence Theories.

## Further Reading

Busse FH (1969) On Howard's upper bound for heat transport by turbulent convection. *Journal of Fluid Mechanics* 37: 457–477.

Busse FH (1978) The optimum theory of turbulence. *Advances in Applied Mechanics* 18: 77–121.

Busse FH (2002) The problem of turbulence and the manifold of asymptotic solutions of the Navier–Stokes equations. In: Oberlack M and Busse FH (eds.) *Theories of Turbulence*, pp. 77–121. Wien: Springer.

Doering CR and Constantin P (1994) Variational bounds on energy dissipation in incompressible flows: shear flow. *Physical Review E* 49: 4087–4099.

Howard LN (1963) Heat transport by turbulent convection. *Journal of Fluid Mechanics* 17: 405–432.

Howard LN (1972) Bounds on flow quantities. *Annual Review of Fluid Mechanics* 4: 473–494.

Joseph DD (1976) *Stability of fluid motions*. vol. 1. Berlin: Springer.

Kerswell RR (1998) Unification of variational principles for turbulent shear flows: the background method of Doering–Constantin and Howard–Busse's mean-fluctuation formulation. *Physica D* 121: 175–192.

# Variational Techniques for Ginzburg–Landau Energies

**S Serfaty**, New York University, New York, NY, USA



Ginzburg–Landau-type problems are variational problems which consider a Dirichlet-type energy posed on complex-valued functions, penalized by a potential term which has a well in the unit circle of the complex plane. The denomination comes from the physical model of superconductivity of Ginzburg and Landau. They are phase-transition-type models in the sense that they describe the state of the material according to different "phases" which can coexist in a sample and be separated by various types of interfaces. We start by presenting the physical model (readers familiar with it may wish to skip the next two sections and go straight to the section "The simplified model").

## Introduction to the Ginzburg–Landau Model

The Ginzburg–Landau model was introduced by Ginzburg and Landau in the 1950s as a phenomenological model to describe superconductivity, and was later justified as a limit of the quantum BCS theory of Bardeen–Cooper–Schrieffer. It is a model of great importance and recognition in physics (with several Nobel prizes awarded for it: Landau, Ginzburg, Abrikosov). In addition to its importance in the modeling of superconductivity, the Ginzburg–Landau model turns out to be mathematically extremely close to the Gross–Pitaevskii model for superfluidity, and models for rotating Bose–Einstein condensates, which all have in common the appearance of topological defects called "vortices."

Superconductivity, which was discovered in 1911 by Kammerling Ohnes, consists in the complete loss of resistivity of certain metals and alloys at very low temperatures: the two most striking consequences of it being the possibility of permanent superconducting currents and the particular behavior that an external magnetic field applied to the sample gets expelled from the material and can generate vortices, through which it penetrates the sample.

### The Energy Functional

After a series of dimension reductions, the Ginzburg–Landau model describes the state of the superconducting sample occupying a region $\Omega$ and submitted to the external magnetic field $h_{\text{ex}}$,

below the critical temperature, through its Gibbs energy:

$$G_\varepsilon(\psi, A) = \frac{1}{2}\int_\Omega |\nabla_A\psi|^2 + \frac{(1-|\psi|^2)^2}{2\varepsilon^2} + \frac{1}{2}\int_{\mathbb{R}^3} |\mathrm{curl}\, A - h_{\mathrm{ex}}|^2 \qquad [1]$$

In this expression, the first unknown $\psi$ is the "order parameter" in physics. It is a complex-valued condensed wave function, indicating the local state of the material, or the phase (in the Landau theory approach of phase transitions): $|\psi|^2$ is the density of the "Cooper pairs" of superconducting electrons explaining superconductivity in the BCS approach. With our normalization $|\psi| \le 1$ and where $|\psi| \sim 1$ the material is in the superconducting phase, while where $|\psi| \sim 0$, it is in the normal phase (i.e., behaves like a normal conductor), the two phases being able to coexist in the sample.

The second unknown $A$ is the electromagnetic vector potential of the magnetic field, a function from $\Omega$ to $\mathbb{R}^3$. The induced magnetic field in the sample is deduced by $h = \mathrm{curl}\, A$. The notation $\nabla_A$ denotes the covariant derivative $\nabla - \mathrm{i}A$. The superconducting current is the vector $j$ of components

$$j_k = \langle \mathrm{i}\psi, (\nabla_A)_k \psi\rangle \qquad [2]$$

where $\langle .,.\rangle$ denotes the scalar product in $\mathbb{C}$ identified with $\mathbb{R}^2$.

Finally, the parameter $\varepsilon$ is the inverse of the "Ginzburg–Landau parameter" $\kappa$, a dimensionless parameter (ratio of the penetration depth and the coherence length) depending on the material only.

Most variational studies of Ginzburg–Landau focus on the regime of large $\kappa$ or small $\varepsilon$, corresponding to "extreme type-II" superconductors, also called the London limit. In this limit, the potential term acts as a singular perturbation, and the characteristic size of the vortices is $\varepsilon \to 0$; vortices become line-like topological singularities, which makes it easier to extract and describe them.

This model is a $\mathbb{U}(1)$-gauge theory, that is, it is invariant under the gauge transformations:

$$\begin{aligned} \psi &\mapsto \psi e^{\mathrm{i}\Phi} \\ A &\mapsto A + \nabla\Phi \end{aligned} \qquad [3]$$

where $\Phi$ is a smooth real-valued function. The physically relevant quantities are those that are gauge invariant, such as the energy $G_\varepsilon, |\psi|, h$, and the superconducting current $j$.

For more on the model, we refer to the physics literature (e.g., DeGennes (1966) and Tinkham (1996)).

### Reductions of the Model

The goal of variational studies of the Ginzburg–Landau model is to relate the energy to the vortices and the applied field. In three dimensions (3D), vortices are filaments, or lines of zeros of the order parameter $\psi$, around which $\psi$ has a nonzero winding number. These are quite delicate to describe in 3D (we will mention some results below), so a simplification that is commonly made consists in reducing to a two-dimensional model.

When reducing to 2D, one assumes that everything is independent of the vertical direction, and that the applied magnetic field is also vertical. The domain $\Omega$ is then a two-dimensional, bounded and (for simplicity) simply connected open set, which is the horizontal section of an infinite vertical cylinder. One can also imagine it represents a thin film.

In 2D, the energy is written the same way:

$$G_\varepsilon(\psi, A) = \frac{1}{2}\int_\Omega |\nabla_A\psi|^2 + \frac{(1-|\psi|^2)^2}{2\varepsilon^2} + |\mathrm{curl}\, A - h_{\mathrm{ex}}|^2 \qquad [4]$$

where this time $A$ is $\mathbb{R}^2$-valued, and the induced magnetic field $h = \mathrm{curl}\, A = \partial_1 A_2 - \partial_2 A_1$ is now a real-valued function, which can be taken to be equal to $h_{\mathrm{ex}}$ (now a real positive number) in $\mathbb{R}^2 \backslash \Omega$.

The stationary states of the system are the critical points of $G_\varepsilon$, or the solutions of the Ginzburg–Landau equations:

$$\begin{aligned} -(\nabla_A)^2\psi &= \frac{1}{\varepsilon^2}\psi(1-|\psi|^2) && \text{in } \Omega \\ -\nabla^\perp h &= \langle \mathrm{i}\psi, \nabla_A\psi\rangle && \text{in } \Omega \\ h &= h_{\mathrm{ex}} && \text{on } \partial\Omega \\ \nabla_A\psi\cdot\nu &= 0 && \text{on } \partial\Omega \end{aligned} \qquad [5]$$

where $\nabla^\perp$ denotes $(-\partial_{x_2}, \partial_{x_1})$.

A common simplification consists in suppressing the magnetic field, and thus in studying the simplified energy

$$E_\varepsilon(u) = \frac{1}{2}\int_\Omega |\nabla u|^2 + \frac{(1-|u|^2)^2}{2\varepsilon^2} \qquad [6]$$

where the order parameter is commonly denoted by $u$, and is still complex valued. This energy, which can be seen as a complex analog of the real-valued Allen–Cahn model of phase transitions, has been extensively studied, especially since the work of Bethuel–Brezis–Hélein, where the domain $\Omega$ is assumed to be two dimensional and simply connected. The higher-dimensional case has also been considered.

### Vortices and Critical Fields

We now need to explain more precisely what a vortex is. In two dimensions, a vortex is an object centered at an isolated zero of $u$ (or $\psi$), around which the phase of $u$ has a nonzero winding number called the "degree of the vortex." It is the simplest example of a topological defect. If the zero is located at $x_0$, the winding number or degree is the integer that can be computed by

$$\frac{1}{2\pi}\int_{\partial B(x_0,r)}\frac{\partial\varphi}{\partial\tau}=d\in\mathbb{Z} \qquad [7]$$

where $r$ is small enough, and $\varphi$ is the phase of $u$, that is, $u$ can be written $u=|u|e^{i\varphi}$. For example, the phase $\varphi=d\theta$, where $\theta$ is the polar angle centered at $x_0$, yields a vortex of degree $d$. Observe that the phase $\varphi$ is not a well-defined function, it is multivalued (and defined up to $2\pi$); however, we have the important relation

$$\operatorname{curl}\nabla\varphi=2\pi\sum_i d_i\delta_{a_i} \qquad [8]$$

where the $a_i$'s are the zeros of $u$, $d_i$'s the associated degrees, and $\delta_x$ denotes the Dirac mass at $x$.

When $\varepsilon$ is small, it is clear from [4] or [6] that $|u|$ prefers to be close to 1, and a scaling argument hints that $|u|$ is different from 1 in regions of characteristic size $\varepsilon$. Of course this is an intuitive picture and several mathematical notions are used to describe the vortices.

Vortices appear due to the applied field $h_{ex}$. For type-II superconductors there are essentially three critical fields, $H_{c_1}$, $H_{c_2}$, $H_{c_3}$, critical values of $h_{ex}$ for which phase transitions occur. For $h_{ex}\le H_{c_1}=O(|\log\varepsilon|)$, there are no vortices and the superconductor is in the superconducting phase $|\psi|\simeq 1$ everywhere. At $H_{c_1}$ the first vortices appear, and their number increases as $h_{ex}$ is raised. When they become numerous they tend to arrange in triangular lattices called Abrikosov lattices, as observed in experiments and predicted by Abrikosov from the Ginzburg–Landau model, in a very influential work. At the second critical field $H_{c_1}=O(1/\varepsilon^2)$ bulk superconductivity is destroyed, and surface superconductivity remains until $H_{c_3}=O(1/\varepsilon^2)$, the third critical field, above which $\psi\equiv 0$ and the material is normal.

## Issues and Methods

The variational approach to Ginzburg–Landau consists in expressing the energy in terms of reduced quantities or objects, in particular in terms of the vortices. This requires to develop mathematical tools to describe and characterize the vortices (in particular give some suitable definitions of a "vortex structure" for a given $u$ or $\psi$), and estimate precisely the energetic cost of each vortex and of their interaction. This allows us to obtain results of variational convergence of the energy $G_\varepsilon, E_\varepsilon$ (or their variants), that is, to derive $\Gamma$-limits, or "reduced problems" posed in terms of the vortices, which are easier to minimize than the original ones. These limits depend on the regime of applied field, and allow to characterization of, in turn, the critical fields, and the optimal repartition and number of the vortices, if any.

Variational methods also serve to solve some inverse problems, that is, to prove the existence of solutions of the equation which have some given properties, such as a given repartition of vortices, through local minimization procedures, or the use of topological methods based on investigating the topology of the energy levels.

Nonvariational approaches of Ginzburg–Landau are also very useful, in particular to identify the profiles of the solutions, to describe vortices of nonminimizing critical points, or to perform a bifurcation analysis around the normal solution at $H_{c_3}$.

## The Simplified Model

We first present the variational study of $E_\varepsilon$ [6] in dimension 2, together with the mathematical tools used for both [6] and [4]. We will restrict to the asymptotics $\varepsilon\to 0$, since this is the situation where the most results are known.

Let us present informally the essential ingredients of the analysis.

### Tracing the Vortices

The easiest way to trace the vortices is to use the current $\langle iu, \nabla u\rangle$ (or the "superconducting current" $j=\langle i\psi, \nabla_A\psi\rangle$ for the case with magnetic field). Here we recall $\langle .,. \rangle$ denotes the scalar product in $\mathbb{C}$ as identified with $\mathbb{R}^2$, that is, $\langle iu, \nabla u\rangle=(u\times\partial_1 u, u\times\partial_2 u)$ with $\times$ the vector product in $\mathbb{R}^2$.

The curl of the current is the vorticity of the map $u$, exactly like in fluid mechanics. Writing $u=\rho e^{i\varphi}$ we have (at least formally) $\langle iu, \nabla u\rangle=\rho^2\nabla\varphi$ and since $\rho=|u|$ is close to 1 (other than in the small vortex regions), we have the approximation

$$\begin{aligned}\operatorname{curl}\langle iu,\nabla u\rangle &=\operatorname{curl}(\rho^2\nabla\varphi)\simeq\operatorname{curl}\nabla\varphi\\ &=2\pi\sum_i d_i\delta_{a_i}\end{aligned} \qquad [9]$$

where the $a_i$'s are the zeros of $u$ (or its vortices) and the $d_i$'s their degrees, or

$$\operatorname{curl}\langle i\psi,\nabla_A\psi\rangle+\operatorname{curl}A\simeq\operatorname{curl}\nabla\varphi$$

in the case with magnetic field. This can be made rigorous (see Jerrard and Soner (2002) and Sandier and Serfaty (to appear)), that is, one can express that

$$\operatorname{curl}\langle \mathrm{i}u, \nabla u\rangle - 2\pi \sum_i d_i \delta_{a_i} \to 0 \quad \text{as } \varepsilon \to 0 \qquad [10]$$

(or respectively $\operatorname{curl}\langle \mathrm{i}\psi, \nabla_A \psi\rangle + \operatorname{curl} A - 2\pi \sum_i d_i \delta_{a_i} \to 0$) in some weak functional norm, thus giving a rigorous use of [8]. The quantity

$$\mu(u) = \operatorname{curl}\langle \mathrm{i}u, \nabla u\rangle \qquad [11]$$

or

$$\mu(\psi, A) = \operatorname{curl}\langle \mathrm{i}\psi, \nabla_A \psi\rangle + \operatorname{curl} A = \operatorname{curl} j + h \qquad [12]$$

in the case with magnetic field, will thus be called the vorticity and be used to trace the vortices, in this limit $\varepsilon \to 0$. The relation

$$\mu - 2\pi \sum_i d_i \delta_{a_i} \to 0 \quad \text{as } \varepsilon \to 0 \qquad [13]$$

states that it is close to being a measure.

This is also called the Jacobian determinant if written (with differential forms) $Ju = \mathrm{d}\langle \mathrm{i}u, \mathrm{d}u\rangle = \langle \mathrm{i}\mathrm{d}u, \mathrm{d}u\rangle = 2(u_{x_1} \times u_{x_2})\mathrm{d}x_1 \wedge \mathrm{d}x_2$, and under this form it can be used in higher dimensions.

### The Cost of Each Vortex

Here we investigate informally the cost of a vortex of degree $d$. We know already that the characteristic length scale of variation of $u$ is $\varepsilon$, and that $(1 - |u|^2)^2$ is strongly penalized. Thus, we may expect that $|u|$ is close to 1 at a distance $\gg \varepsilon$ of the zeros. Assuming that $x_0$ is a zero of $u$, and taking formally $|u| = 1$ for $|x - x_0| \geq \varepsilon$, we may write $u = \mathrm{e}^{\mathrm{i}\varphi}$ and $|\nabla u| = |\nabla \varphi|$ for $|x - x_0| \geq \varepsilon$.

Then, we have

$$\begin{aligned} &\frac{1}{2}\int_{R \geq |x - x_0| \geq \varepsilon} |\nabla u|^2 \\ &\geq \frac{1}{2}\int_\varepsilon^R \left(\int_{\partial B(x_0, r)} \left|\frac{\partial \varphi}{\partial \tau}\right|^2\right) \mathrm{d}r \\ &\geq \frac{1}{2}\int_\varepsilon^R \left(\left(\int_{\partial B(x_0, r)} \frac{\partial \varphi}{\partial \tau}\right)^2 \frac{1}{2\pi r}\right) \mathrm{d}r \qquad [14] \\ &\geq \frac{1}{2}\,\frac{4\pi^2 d^2}{2\pi}\int_\varepsilon^R \frac{\mathrm{d}r}{r} = \pi d^2 \log\frac{R}{\varepsilon} \qquad [15] \end{aligned}$$

where we have used the Cauchy–Schwarz inequality for [14], and the characterization of the degree [7]. We may also observe that this lower bound is sharp if $\partial\varphi/\partial\tau$ is constant, that is, if the phase is $d\theta$ (and the vortex radial). The cost associated to $|u|$ in the energy imposes the length scale $\varepsilon$ and is generally of order 1 ($|\nabla u| \leq C/\varepsilon$), thus negligible compared to the cost associated to the phase, which blows up as $\log 1/\varepsilon$ as $\varepsilon \to 0$.

The above estimate is only valid as long as $B(x_0, R)$ does not contain any other zero of $u$. If vortices get close to each other or become numerous, one needs refined techniques to estimate their cost. This can be done through a "ball-construction method" introduced independently by Jerrard and Sandier.

### Evaluating the Total Interaction Cost of Vortices

In a first approach, one studies configurations which satisfy the upper bound $E_\varepsilon(u) \leq C|\log \varepsilon|$. Then, lower bounds of the type [15] show that the total sum of the degrees (hence the total number of vortices of nonzero degree) remains bounded as $\varepsilon \to 0$. Up to extraction, we may assume these zeros $a_i$ converge as $\varepsilon \to 0$ to a finite set of points $p_i$, with a total degree still denoted $d_i$. This can also be expressed as $\mu(u_\varepsilon) \to 2\pi \sum_i d_i \delta_{p_i}$ as $\varepsilon \to 0$.

This is not the only case of interest, since unbounded numbers of vortices do arise, especially in the physical situation of the energy with magnetic field, as we will see in the next section. However, this hypothesis, which was made in the work of Bethuel–Brezis–Hélein, makes the analysis easier and already allows us to exhibit the main phenomena.

Vortices in superconductors are generated by the presence of the external magnetic field $h_{\mathrm{ex}}$. For the energy without magnetic field, this has to be replaced by some boundary condition which forces some degree. Bethuel–Brezis–Hélein considered the fixed Dirichlet boundary condition $u_\varepsilon = g$ on $\partial\Omega$, where $g$ is a fixed unit-valued map on $\partial\Omega$, of degree $d > 0$. This forces $u$ to have a total degree $d$ in $\Omega$. However, the Neumann boundary condition, for instance, can also be considered (the minimizers of $E_\varepsilon$ are then simply constants, they are trivial, but one can still look for other critical points).

Let us return to lower bounds in order to look for the next order term in the energy (still with formal arguments). Cutting out holes $\cup_i B(p_i, \rho)$ of fixed size $\rho$ around the limiting vortices $p_i$, we may assume that $u = \mathrm{e}^{\mathrm{i}\varphi}$ in $\Omega \setminus \cup_i B(p_i, \rho) = \Omega_\rho$, with $\varphi$ a real-valued function, defined modulo $2\pi$. Minimizing the energy outside of the holes amounts to solving

$$\min_{\substack{u:\, \Omega_\rho \to \mathbb{S}^1 \\ u = g \text{ on } \partial\Omega \\ \deg(u, \partial B(p_i, \rho)) = d_i}} \frac{1}{2}\int_{\Omega_\rho} |\nabla u|^2$$

This is a harmonic map problem, whose solution is given in terms of $\varphi$ by

$$\begin{aligned} \Delta\varphi &= 0 \qquad \text{in } \Omega_\rho \\ \frac{\partial\varphi}{\partial\tau} &= \left\langle \mathrm{i}g, \frac{\partial g}{\partial\tau}\right\rangle \quad \text{on } \partial\Omega \\ \int_{\partial B(p_i,\rho)} \frac{\partial\varphi}{\partial\tau} &= 2\pi d_i \end{aligned}$$

and in terms of the harmonic conjugate $\Phi$ which is the function (up to a constant) such that $\nabla\varphi = \nabla^\perp\Phi$,

$$\begin{aligned} \Delta\Phi &= 0 \qquad \text{in } \Omega_\rho \\ \frac{\partial\Phi}{\partial\nu} &= \left\langle \mathrm{i}g, \frac{\partial g}{\partial\tau}\right\rangle \quad \text{on } \partial\Omega \\ \int_{\partial B(p_i,\rho)} \frac{\partial\Phi}{\partial\nu} &= 2\pi d_i \end{aligned} \qquad [16]$$

As $\rho \to 0$, $\Phi$ behaves like the solution of

$$\begin{aligned} \Delta\Phi_0 &= 2\pi\sum_i d_i\delta_{p_i} \quad \text{in } \Omega \\ \frac{\partial\Phi_0}{\partial\nu} &= \left\langle \mathrm{i}g, \frac{\partial g}{\partial\tau}\right\rangle \qquad \text{on } \partial\Omega \end{aligned} \qquad [17]$$

Hence, we have

$$\begin{aligned} \frac{1}{2}\int_{\Omega_\rho} |\nabla\varphi|^2 &= \frac{1}{2}\int_{\Omega_\rho} |\nabla\Phi|^2 \\ &\simeq \frac{1}{2}\int_{\Omega_\rho} |\nabla\Phi_0|^2 \\ &= \pi\sum_i d_i^2\log\frac{1}{\rho} + W_d(p_1,\ldots,p_n) \\ &\quad + o(1) \quad \text{as } \rho\to 0 \end{aligned} \qquad [18]$$

where

$$\begin{aligned} W_d(a_1,\ldots,a_n) = &-\pi\sum_{i\neq j} d_i d_j \log|p_i - p_j| \\ &-\pi\sum_i d_i R(a_i) \\ &+\frac{1}{2}\int_{\partial\Omega}\Phi_0\left(\mathrm{i}g, \frac{\partial g}{\partial\tau}\right) \end{aligned} \qquad [19]$$

and $R(x) = \Phi_0(x) - \sum_i d_i\log|x - p_i|$. The function $W$ was introduced by Bethuel–Brezis–Hélein and called the renormalized energy, since it consists in the part of the energy that is left after subtracting the "infinite part" in $|\log\varepsilon|$ from $E_\varepsilon$. It contains the (logarithmic) interaction energy between the vortices: we see that vortices with degrees of same sign repel one another while vortices with degrees of opposite signs attract one another. The $\pi d_i^2 \log 1/\rho$ term corresponds to the self-interaction, or cost of the vortex of core of size $\rho$; it is what replaces the infinite term in the formal calculation.

Now [18] is a good estimate for the optimal energy outside of the holes, while the energy in holes of size $\rho$ can be bounded below by [15]. Given the degree $d_i$ on the boundary $\partial B(p_i,\rho)$ of the small hole, $B(p_i,\rho)$ contains one or several zeros of $u$ of degrees $\delta_k$ with total degree $\sum_k \delta_k = d_i$. In view of [15], since the cost of a vortex of degree $d$ grows like $\pi d^2|\log\varepsilon|$, and since the infimum of $\sum_k \delta_k^2$ under the constraint $\sum_k \delta_k = d_i$ is $\delta_k = \mathrm{sign}(d_i)$, the least costly way to achieve this is to have $|d_i|$ vortices of degree $\mathrm{sign}(d_i)$. The smallest lower bound possible is thus

$$\frac{1}{2}\int_{B(p_i,\rho)} |\nabla u|^2 + \frac{(1-|u|^2)^2}{2\varepsilon^2} \geq \pi|d_i|\log\frac{\rho}{\varepsilon} + C \qquad [20]$$

where the constant $C$ can be described explicitly. Adding up the results of [20] and [18], we find

$$\begin{aligned} E_\varepsilon(u) \geq\ &\pi\sum_i d_i^2\log\frac{1}{\rho} \\ &+\pi\sum_i |d_i|\log\frac{\rho}{\varepsilon} + W_d(p_1,\ldots,p_n) \\ &+ nC + o_\rho(1) + o_\varepsilon(1) \\ \geq\ &\pi\sum_i |d_i|\log\frac{1}{\varepsilon} + W_d(p_1,\ldots,p_n) \\ &+ nC + o_\varepsilon(1) \end{aligned} \qquad [21]$$

with equality only if $u$ has $|d_i|$ zeros of degree $\mathrm{sign}(d_i)$ in each $B(p_i,\rho)$.

This provides a lower bound of the energy in terms of the vortices. Moreover, this bound is sharp: one can construct test configurations which have the given limiting vortices $(p_i, d_i)$, and an energy equal to the right-hand side of [21].

One can thus deduce the behavior of global minimizers of the energy. Given the total degree $d = \deg(g) > 0$ on $\partial\Omega$, we need $\sum_i d_i = d$, and the lowest value achievable under this constraint in the right-hand side of [21] is to have $d_i = 1$ for every $i$, and thus to have exactly $d$ vortices of degree 1. Moreover, the limiting points $p_i$'s should minimize $W$. We thus are led to the first main result.

**Theorem 1** (Bethuel–Brezis–Hélein). *Minimizers of $E_\varepsilon$ under the boundary condition $u = g$, $\deg(g) = d > 0$, have $d$ zeros of degree 1, which converge as $\varepsilon \to 0$ to a minimizer of $W$.*

This result can be rephrased as a result of $\Gamma$-convergence of $E_\varepsilon - \pi d|\log\varepsilon|$. It reduces the minimization of $E_\varepsilon$ to one of $W$, which is a finite-dimensional problem (interaction of point charges).

Thus, we see again the interest of studying this asymptotic limit $\varepsilon \to 0$ because the vortices become pointlike and the problem reduces to a finite-dimensional one, or one of minimizing the vortex interaction.

#### Further Results

A nonvariational approach also allowed Bethuel–Brezis–Hélein to prove a further correspondence between $E_\varepsilon$ and $W$: they obtained that critical points of $E_\varepsilon$, under the upper bound $E_\varepsilon \le C|\log \varepsilon|$, have vortices which converge to a critical point of $W$. Other important results are the study of the blow-up profiles or solutions in the whole plane, by Brezis–Merle–Rivière and Mironescu.

In two dimensions, the variational approach is also used to solve inverse problems (construct solutions) and study variants of the energy with pinning (or weighted) terms.

The variational approach is also fruitful in higher dimensions. In dimension 3, for example, vortices are not points but vortex lines, and the Jacobian $Ju = d(\mathrm{i}u, \mathrm{d}u)$ can be seen as a current carried by the vortex line, with $\|Ju\|$ total mass of the current equal to $\pi$ times the length of the line, and it was established by Jerrard and Soner that $Ju_\varepsilon$ is compact in some weak sense, and converges, up to extraction, to some $\pi$ times integer-multiplicity rectifiable current $J$, with

$$\liminf_{\varepsilon\to 0} \frac{E_\varepsilon(u_\varepsilon)}{|\log \varepsilon|} \ge \|J\|$$

In fact, a complete $\Gamma$-convergence result of $E_\varepsilon/|\log \varepsilon|$ can be proved, see the work of Alberti–Baldo–Orlandi, and thus minimizing $E_\varepsilon$ reduces at the limit to minimizing the length of the line, leading to straight lines, or in higher dimensions, to codimension-2 minimal currents. This is a nontrivial problem, contrarily to dimension 2, where the $\Gamma$-limit of $E_\varepsilon/|\log \varepsilon|$ is trivial, which required to go to the lower-order term to find the nontrivial renormalized energy limit $W$.

### The Functional with Magnetic Field

The aim here is to achieve the same objective: express or bound from below the energy by terms which depend only on the vortices and their degrees. The method consists in transposing the type of analysis above taking into account the magnetic field contribution to see how the external field triggers the sudden appearance of vortices, and for what values they appear (thus retrieving the critical fields, etc.). One of the main difficulties consists in the fact that the number of vortices becomes divergent, which requires more delicate estimates. Also, it is then no longer possible to study the convergence of the individual zeros of $\psi$, so one studies instead the limit of rescalings of the vorticity measures $\mu(\psi, A)$.

#### Splitting of the Energy and Main Results

Let us recall that in the case with magnetic field, the vorticity is given by [12]. In addition, we may assume that the second set of equations in [5]

$$-\nabla^\perp h = j \ \text{ in } \Omega, \quad h = h_{\mathrm{ex}} \text{ on } \partial\Omega \qquad [22]$$

is satisfied (if not, keeping $\psi$ fixed and choosing $A$ which satisfies this equation always decreases the energy). Taking the curl of this equation, we find exactly

$$\begin{aligned} -\Delta h + h &= \mu(\psi, A) && \text{in } \Omega \\ h &= h_{\mathrm{ex}} && \text{on } \partial\Omega \end{aligned} \qquad [23]$$

Thus, the vorticity and the induced magnetic field are in one-to-one correspondence with each other. Combining it to the relation [13], we are led to the approximate relation

$$\begin{aligned} -\Delta h + h &\simeq 2\pi \sum_i d_i \delta_{a_i} && \text{in } \Omega \\ h &= h_{\mathrm{ex}} && \text{on } \partial\Omega \end{aligned} \qquad [24]$$

where again the $a_i$'s are the vortex centers and $d_i$'s their degrees, well known in physics as the "London equation." It shows how the magnetic field is induced by the vortices which act like "charges," and how the magnetic field "penetrates the sample" around the positive vortex locations. Of course this equation is only an approximation, because the singularities at the $a_i$'s, where $h$ would become infinite, are really smoothed out in $\mu(\psi, A)$; however, the approximation is good far from the vortex cores, just as [17] is an approximation for [16].

It is then natural to introduce the field corresponding to the vortex-free situation, which is $h_{\mathrm{ex}} h_0$ where $h_0$ solves

$$\begin{aligned} -\Delta h_0 + h_0 &= 0 && \text{in } \Omega \\ h_0 &= 1 && \text{on } \partial\Omega \end{aligned} \qquad [25]$$

$h_0$ is thus a fixed smooth function, depending only on $\Omega$, and when there are no vortices, we expect $h$ to be approximately $h_{\mathrm{ex}} h_0$. Moreover, $h' := h - h_{\mathrm{ex}} h_0$ then solves

$$\begin{aligned} -\Delta h' + h' &= \mu(\psi, A) \simeq 2\pi \sum_i d_i \delta_{a_i} && \text{in } \Omega \\ h' &= 0 && \text{on } \partial\Omega \end{aligned} \qquad [26]$$

Defining the Green kernel $G(.,y)$ by

$$\begin{aligned} -\Delta G + G &= \delta_y \quad \text{in } \Omega \\ G &= 0 \quad \text{on } \partial\Omega \end{aligned} \qquad [27]$$

and $S$ by $S(x,y) = 2\pi G(x,y) + \log|x-y|$, for $x$ far enough from the $a_i$'s, we may approximate $h'$ by

$$h'(x) = 2\pi \sum_i G(x, a_i) \qquad [28]$$

Using the second Ginzburg–Landau equation [22] and the fact that $|\psi| \leq 1$, we have $|\nabla_A \psi| \geq |j| = |\nabla h|$, thus $G_\varepsilon(\psi, A) \geq (1/2)\int_\Omega |\nabla h|^2 + |h - h_{\text{ex}}|^2$. Plugging in the decomposition $h = h_{\text{ex}} h_0 + h'$ and using an integration by parts and [26], one finds

$$\begin{aligned} G_\varepsilon(\psi, A) &= \frac{1}{2} h_{\text{ex}}^2 \int_\Omega |\nabla h_0|^2 + |h_0 - 1|^2 \\ &\quad + h_{\text{ex}} \int_\Omega \nabla h_0 \cdot \nabla h' + (h_0 - 1) h' \\ &\quad + \frac{1}{2} \int_\Omega |\nabla h'|^2 + |h'|^2 \\ &= h_{\text{ex}}^2 J_0 + h_{\text{ex}} \int_\Omega (h_0 - 1) \mu(\psi, A) \\ &\quad + \frac{1}{2} \int_\Omega |\nabla h'|^2 + |h'|^2 \end{aligned} \qquad [29]$$

where $J_0$ is the constant $(1/2)\int_\Omega |\nabla h_0|^2 + |h_0 - 1|^2$. The right-hand side of eqn [29] can be expressed in terms of the vortices. First, using [26], we have $\int_\Omega (h_0 - 1)\mu(\psi, A) \simeq 2\pi \sum_i d_i (h_0 - 1)(a_i)$. Second, the expression $\int_\Omega |\nabla h'|^2 + |h'|^2$ can be treated exactly like $E_\varepsilon(u)$ in the previous section, using lower bounds for the cost of vortices provided by the Jerrard–Sandier method, we are led to the (approximate) relation

$$\begin{aligned} \frac{1}{2} \int_\Omega |\nabla h'|^2 + |h'|^2 \geq \pi \sum_i |d_i| \log \frac{1}{\varepsilon} \\ - \pi \sum_{i \neq j} d_i d_j \log |a_i - a_j| \\ + \pi \sum_{i,j} d_i d_j S(a_i, a_j) \end{aligned} \qquad [30]$$

Combining this to [29] we find the decomposition

$$\begin{aligned} G_\varepsilon(\psi, A) \geq h_{\text{ex}}^2 J_0 + \pi \sum_i |d_i| |\log \varepsilon| \\ + 2\pi h_{\text{ex}} \sum_i d_i (h_0 - 1)(a_i) \\ - \pi \sum_{i \neq j} d_i d_j \log |a_i - a_j| \\ + \pi \sum_{i,j} d_i d_j S(a_i, a_j) \end{aligned} \qquad [31]$$

On the other hand, this inequality is sharp: as before, given vortices $a_i$, one can construct a configuration $(\psi, A)$ for which this is an equality, at leading order.

In that relation, $h_{\text{ex}}^2 J_0$ is a fixed energy, the energy of the vortex-free configuration. To it are added the intrinsic cost of each vortex $\pi |d_i| |\log \varepsilon|$, the interaction cost between vortices, and the interaction between the vortices and the external field $2\pi h_{\text{ex}} \sum_i d_i (h_0 - 1)(a_i)$.

It is then simple, by minimizing the right-hand side with respect to the vortices for a given $h_{\text{ex}}$, and observing that $h_0 - 1 \leq 0$, to deduce a few basic facts about vortices: vortices of positive degree (and of degree +1) are preferred, each vortex costs $\pi|\log \varepsilon|$, and allows to gain at best an energy $2\pi h_{\text{ex}} \max |h_0 - 1|$ when placed at the minimum of $h_0 - 1$. Therefore, vortices become favorable when their cost becomes smaller than the gain, that is, when $h_{\text{ex}}$ becomes larger than the "first critical field"

$$H_{c_1} \sim \frac{|\log \varepsilon|}{2|\min(h_0 - 1)|} \qquad [32]$$

We have the first main result.

**Theorem 2** (Sandier–Serfaty). *When $\varepsilon$ is small enough and $h_{\text{ex}} \leq H_{c_1}$, then minimizers of $G_\varepsilon$ have no vortices.*

On the other hand, if $h_{\text{ex}} \geq H_{c_1}$, the vortices cannot all be located at the same minimum point of $h_0 - 1$, because their repulsion $-\pi \sum_{i \neq j} \log |a_i - a_j|$ would be infinite. There is thus a trade-off between their repulsion and the cost for being far from the minimum of $h_0 - 1$. Only if $n$, the number of vortices, is small compared to $h_{\text{ex}}$ do the vortices tend to concentrate near the minimum of $h_0 - 1$. If so, then, assuming for simplicity that the minimum of $h_0 - 1$ is achieved at a unique point $p$, and denoting by $Q$ the Hessian of $h_0 - 1$ at $p$, in the relation above $(h_0 - 1)(a_i)$ can be approximated by $\min(h_0 - 1) + (1/2) Q(a_i - p)$ and thus $G_\varepsilon(\psi, A)$ by

$$\begin{aligned} G_\varepsilon(\psi, A) \sim h_{\text{ex}}^2 J_0 + \pi n |\log \varepsilon| + 2\pi n h_{\text{ex}} \min(h_0 - 1) \\ + \pi h_{\text{ex}} \sum_i Q(a_i - p) \\ - \pi \sum_{i \neq j} d_i d_j \log |a_i - a_j| + \pi n^2 S(p, p) \end{aligned} \qquad [33]$$

From this relation, optimizing on $\ell$, the characteristic distance to $p$ and characteristic distance between the vortices, we find that $\ell = \sqrt{n / h_{\text{ex}}}$ is optimal.

Moreover, optimizing with respect to $n$, we find that $n$ should remain bounded (as $\varepsilon \to 0$) when $h_{\text{ex}} \leq H_{c_1} + O(\log|\log \varepsilon|)$. In that regime, rescaling by setting $x_i = ((a_i - p)/\ell)$, we have the following result:

**Theorem 3** (Sandier–Serfaty). *There exist fields $H_n \sim H_{c_1} + C(n-1)\log|\log\varepsilon|$ such that when $H_n \le h_{ex} < H_{n+1}$, minimizers of $G_\varepsilon$ have n vortices of degree 1, and the rescaled vortices $x_i$'s tend to minimize*:

$$w_n(x_1, \ldots, x_n) = -\pi \sum_{i\neq j} \log|x_i - x_j| + \pi n \sum_{i=1}^{n} Q(x_i) \qquad [34]$$

If $h_{ex} - H_{c_1} \gg \log|\log\varepsilon|$, then the optimal number of vortices $n$ becomes unbounded as $\varepsilon \to 0$. The analysis above still holds, but in order to get a convergence of the vortices, one needs to rescale the vorticity measure by $n$. There is an intermediate regime, for $\log|\log\varepsilon| \ll h_{ex} - H_{c_1} \ll |\log\varepsilon|$ for which $n$ should be $\gg 1$ but still $n \ll h_{ex}$, so $\ell \ll 1$: vortices are numerous, but still concentrate around $p$. Rescaling by the scale $\ell$ as above, we prove that the density of vortices (after dividing it by $n$) converges to a probability measure, minimizer of the energy

$$I(\mu) = -\pi \int_{\mathbb{R}^2\times\mathbb{R}^2} \log|x-y|\, d\mu(x)\, d\mu(y) + \pi \int_{\mathbb{R}^2} Q(x)\, d\mu(x) \qquad [35]$$

This is an averaged/continuous form of [34].

If $h_{ex} - H_{c_1}$ is of order $|\log\varepsilon|$, then the optimal number $n$ becomes of order $h_{ex}$ and the vortices no longer concentrate around a single point.

The simplest approach is then to simply consider the vorticity measure $\mu(\psi, A)$ and to rescale it by the order $n$, hence by $h_{ex}$. Then $(1/h_{ex})\mu(\psi, A)$ converges, after extraction, to some measure $\mu_*$. A continuous version of [31] can thus be written, using [12], as

$$G_\varepsilon(\psi, A) \ge \frac{1}{2} h_{ex}|\log\varepsilon| \int_\Omega |\mu_*| + \frac{1}{2} h_{ex}^2 \int_\Omega |\nabla h_{\mu_*}|^2 + |h_{\mu_*}|^2 \qquad [36]$$

where $h_{\mu_*}$ solves

$$\begin{aligned} -\Delta h_{\mu_*} + h_{\mu_*} &= \mu_* \quad \text{in } \Omega \\ h_{\mu_*} &= 1 \quad \text{on } \partial\Omega \end{aligned}$$

Again, this inequality can be proved to be sharp (by a construction) and allows to show that minimizers of $G_\varepsilon$ have a vorticity $\mu(\psi, A)$ such that $\mu(\psi, A)/h_{ex}$ converges to a minimizer of

$$\mathcal{G}(\mu_*) = \frac{1}{2}\left(\lim_{\varepsilon\to 0} \frac{|\log\varepsilon|}{h_{ex}}\right) \int_\Omega |\mu_*| + \frac{1}{2}\int_\Omega |\nabla h_{\mu_*}|^2 + |h_{\mu_*}|^2$$

In fact the stronger result holds, in that sense:

**Theorem 4** (Sandier–Serfaty). *$G_\varepsilon/h_{ex}^2$ $\Gamma$-converges to $\mathcal{G}$.*

The limit problem of minimizing $\mathcal{G}$ turns out to have a simple solution in terms of an obstacle problem: the optimal $\mu_*$ is a uniform density of vortices on a subdomain of $\Omega$ determined through a free boundary problem (and depending on $h_{ex}$), which is nonzero.

In all these regimes, we have thus been able to identify the optimal number and repartition of vortices through a $\Gamma$-convergence-type approach, that is, by reducing the minimization of the energy to the minimization of a limiting problem: $w_n$ or $I$ or $\mathcal{G}$, according to the regime.

### Further Results

Concerning vortices, in the same spirit as what was done for $E_\varepsilon$, we can obtain necessary conditions characterizing limiting vorticities obtained from sequences of (nonminimizing) critical points of the energy $G_\varepsilon$. They consist in passing to the limit in the conservative form of the Ginzburg–Landau equations [5].

Most of the results concerning the phase transitions at the next critical fields $H_{c_2}$ and $H_{c_3}$ are also obtained by nonvariational methods, and often by linear analysis.

The study of the Ginzburg–Landau energy in non-simply-connected domains is also very interesting because it leads to nontrivial topological effects, since in such domains there exist unit-valued maps with nonzero degree (corresponding to permanent currents).

*See also:* Abelian Higgs Vortices; Aharonov–Bohm Effect; Bose–Einstein Condensates; Gamma-Convergence and Homogenization; Gauge Theory: Mathematical Applications; Ginzburg–Landau Equation; High $T_c$ Superconductor Theory; Image Processing: Mathematics; Superfluids; Topological Defects and Their Homotopy Classification; Variational Techniques for Microstructures.

## Further Reading

Bethuel F, Brezis H, and Hélein F (1994) *Ginzburg–Landau Vortices*. Boston: Birkhäuser.

DeGennes PG (1966) *Superconductivity of Metal and Alloys*. New York: Benjamin.

Jerrard RL and Soner HM (2002) The Jacobian and the Ginzburg–Landau energy. *Calculus of Variations and Partial Differential Equations* 14(2): 151–191.

Sandier E and Serfaty S *Vortices in the Magnetic Ginzburg–Landau Model*. Birkhäuser (monograph to appear).

Tinkham M (1996) *Introduction to Superconductivity*, 2nd edn. McGraw-Hill.

# Variational Techniques for Microstructures

**G Dolzmann**, University of Maryland, College Park, MD, USA



## Austenite–Martensite Transformations and the Shape Memory Effect

Microstructures in materials that typically form in response to phase transformations in the solid state, and their impact on the elastic properties of these materials have been known for centuries. The discovery of the complex phase diagram of iron revolutionized the production of steels at the end of the nineteenth century. Starting in the 1980s, the mathematical description of microstructures in the framework of nonlinear elasticity has led to deep analytical questions and surprising developments in the calculus of variations and in nonlinear partial differential equations.

The mathematical approach outlined here is based on the following fundamental assumptions:

1. The observed configurations correspond to minimizers of or elements of minimizing sequences for an energy functional.
2. The qualitative properties of low energy states are determined from the set of minima of the free energy density.

Under these assumptions one aims at explaining experimental observations and to predict material properties based on minimizing an energy functional of the form

$$I(u) = \int_\Omega W(Du)\,\mathrm{d}x$$

Here $\Omega$ is an ideal, unstressed reference configuration in $\mathbb{R}^n, u:\Omega\to\mathbb{R}^m$ is an elastic deformation, and $W:\mathbb{M}^{m\times n}\to\mathbb{R}$ is the stored energy density. In the case of physical interest, $m=n=2$ or $m=n=3$. For applications in elasticity we assume that $m=n$, but this assumption is not needed in the general theory. The energy density $W$ and its structure depend critically on the temperature. However, since we are interested in the analysis of the material at a given temperature, we do not include this dependence explicitly.

The key ingredient of this model is the stored energy density $W$ which has to reflect the properties of the specific material one wants to model. Frequently these are alloys, in particular shape memory alloys that undergo an austenite–martensite transformation. For most materials a closed analytic expression for $W$ is not available. In the spirit of the fundamental assumption (2) one therefore focuses on the structure of the set of minima of $W$ which is determined from general invariance and symmetry principles. We may assume that $W\geq 0$ and that $K=\{X: W(X)=0\}\neq\emptyset$. The principle of material frame indifference then asserts that

$$W(RF) = W(F) \quad \text{for all } R\in \mathrm{SO}(n)$$

Here $\mathrm{SO}(n)$ is the group of proper rotations, that is, the set of all matrices $R\in\mathbb{M}^{n\times n}$ with $R^TR=\mathrm{Id}$ and $\det R=1$.

The symmetry of the austenitic (high-temperature) phase implies that the energy density in the martensitic (low-temperature) phase is invariant under all changes of basis that leave the underlying lattice in the austenitic phase invariant. Therefore,

$$W(R^TFR) = W(F) \quad \text{for all } R\in\mathcal{P}_a$$

where $\mathcal{P}_a$ is the point group of the austenite. In the case of a cubic to tetragonal phase transformation, this leads to $K=\mathrm{SO}(3)$ in the austenitic phase and to

$$K = \mathrm{SO}(3)U_1\cup\mathrm{SO}(3)U_2\cup\mathrm{SO}(3)U_3 \qquad [1]$$

with

$$U_i = \eta^2 e_i\otimes e_i + \frac{1}{\eta}(I - e_i\otimes e_i) \qquad [2]$$

in the martensitic phase (see Figure 1). A set of the form $\mathrm{SO}(n)U_i$ is often referred to as an energy well.

The origin of the shape memory effect is the availability of a rich class of geometric patterns in which the martensitic phases can be arranged, thus leading to a great flexibility of the material to accommodate macroscopic deformations. Upon heating of the material above the transformation temperature, the martensitic phases lose their stability and the material returns to its unique shape in the

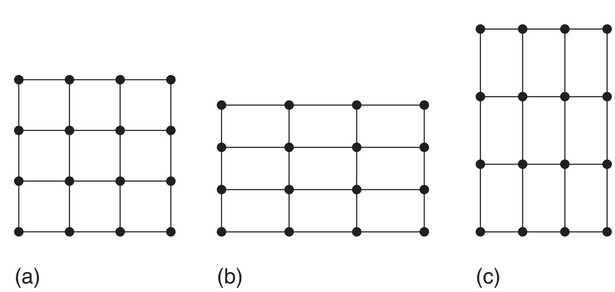
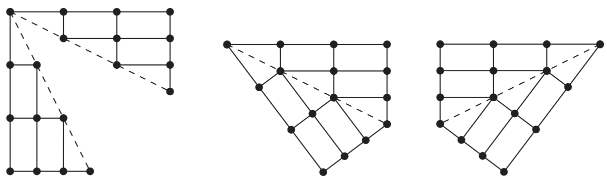



**Figure 1** Two-dimensional cartoon of a cubic to tetragonal phase transformation in a single crystal: (a) a cubic lattice, (b) and (c) tetragonal variants which are stretched in directions $e_1$ and $e_2$, respectively. (Sketch not to scale.)



(a) (b) (c)

**Figure 2** Formation of phase boundaries in a single crystal. (a) The upper right half of the lattice deforms into phase I with the constant deformation gradient $U_1$, the lower left half of the lattice deforms into phase II with constant deformation gradient $U_2$. (b) An additional rotation is needed to accomplish a continuous deformation, see formula [3]. (c) A different configuration with a different orientation of the interface. (Sketch not to scale.)

austenitic phase. The two solutions of Hadamard's compatibility condition

$$QU_2 - U_1 = a \otimes b, \quad Q \in SO(3)$$

are given by

$$Q_1 = \frac{1}{\eta^4+1}\begin{pmatrix} 2\eta^2 & \eta^4-1 & 0 \\ 1-\eta^4 & 2\eta^2 & 0 \\ 0 & 0 & \eta^4+1 \end{pmatrix} \qquad [3]$$

and $Q_2 = Q_1^T$ (see Figure 2). The normals (in the reference configuration) are given by $(1, \pm 1, 0)/\sqrt{2}$. It is one of the successes of the theory that it provides an analytical derivation of the normals to the twinning planes.

## The Direct Method in the Calculus of Variations

The mathematical interest in the variational problems described in the previous section lies in the fact that existence of minimizers cannot in general be obtained by a straightforward application of the direct methods in the calculus of variations. This approach is based on the idea to (1) choose a minimizing sequence for the functional $I$, (2) show that this sequence is bounded and precompact, and (3) prove that the functional is lower semicontinuous with respect to the notion of convergence,

$$I(u) \le \liminf_{j\to\infty} I(u_j) \quad \text{if } u_j \to u$$

The typical choice is to seek $u_j$ in a suitable Sobolev space $W^{1,p}(\Omega;\mathbb{R}^m)$ with $1 < p \le \infty$ which is related to growth and coercivity conditions for the energy density $W$,

$$c_1|F|^p - c_2 \le W(F) \le c_3\left(|F|^p + 1\right) \quad \text{for all } F \in \mathbb{M}^{m\times n} \qquad [4]$$

This leads to weak compactness in $W^{1,p}(\Omega;\mathbb{R}^m)$ (weak-$*$ compactness in $W^{1,\infty}(\Omega;\mathbb{R}^m)$) and to the requirement of sequential weak lower semicontinuity of the functional,

$$I(u) \le \liminf_{j\to\infty} I(u_j) \quad \text{if } u_j \rightharpoonup u \text{ in } W^{1,p}(\Omega;\mathbb{R}^m)$$

(sequential weak-$*$ lower semicontinuity for $p=\infty$). Morrey's fundamental work establishes a link between convexity conditions for the energy density and lower semicontinuity of the variational integral: under suitable growth and coercivity conditions, sequential weak-$*$ lower semicontinuity is equivalent to quasiconvexity of the integrand.

**Definition 1** A function $W : \mathbb{M}^{m\times n} \to \mathbb{R}$ is said to be quasiconvex at $F$ if

$$\int_\Omega W(F)\mathrm{d}x \le \int_\Omega W(F + D\phi)\mathrm{d}x \quad \text{for all } \phi \in W_0^{1,\infty}(\Omega;\mathbb{R}^m)$$

and for all open and bounded domains $\Omega \subset \mathbb{R}^n$ with $\mathcal{L}^n(\partial\Omega) = 0$. It is said to be quasiconvex if it is quasiconvex at all $F$.

In the language of nonlinear elasticity, $W$ is quasiconvex if affine functions are minimizers of the energy functional subject to their own boundary conditions. The direct method implies the following classical existence theorem.

**Theorem 1** *Suppose that* $W : \mathbb{M}^{m\times n} \to \mathbb{R}$ *is quasiconvex and satisfies the growth and coercivity condition* [4]. *Let* $u_0 \in W^{1,p}(\Omega;\mathbb{R}^m)$. *Then the variational problem: minimize* $I(u)$ *in*

$$\mathcal{A} = \left\{u \in W^{1,p}(\Omega;\mathbb{R}^m) : \ u - u_0 \in W_0^{1,p}(\Omega;\mathbb{R}^m)\right\}$$

*has a minimizer.*

The remarkable fact is that the structure of the zero set of a typical energy $W$ modeling a phase-transforming material in its low-temperature phase prevents $W$ from being quasiconvex. In order to see this, let $Q \subset \mathbb{R}^3$ be a cube with two of its sides perpendicular to $b = (1,1,0)/\sqrt{(2)}$ and let $h$ be the 1-periodic function with $h' = 0$ on $(0,\lambda)$ and $h' = 1$ on $(\lambda, 1)$ with $\lambda \in (0,1)$. Define $v_j(x) = U_1x + ah(jx\cdot b)/j$ and

$$\begin{aligned} u_j(x) &= \min\{v_j(x), \operatorname{dist}(x,\partial Q)\} \\ &= \min\{U_1x + ah(jx\cdot b)/j, \operatorname{dist}(x,\partial Q)\} \end{aligned}$$

where $\operatorname{dist}(x,\partial Q) = \inf\{\|x-y\|_\infty, y \in \partial Q\}$. Then $u_j \to u$, $u(x) = Cx$ strongly in $L^\infty(Q;\mathbb{R}^3)$ and weakly-$*$ in $W^{1,\infty}(Q;\mathbb{R}^3)$ with $C = \lambda U_1 + (1-\lambda)Q_1U_2 \notin K$ where $K$ is the zero set of $W$, see the previous section.

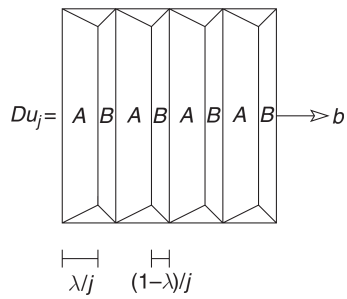



**Figure 3** Construction of a minimizing sequence $u_j$ with $Du_j \to \{A, B\}$ in measure and affine boundary conditions $u(x) = \lambda A + (1-\lambda)B$ Hadamard's compatibility condition requires that $A - B = a \otimes b$ is a rank-1 matrix and that the planar interfaces are perpendicular to $b$.

Moreover, $Du_j \in \{U_1, Q_1 U_2\}$ except in a small transition layer of volume $\mathcal{O}(1/j)$ close to $\partial Q$ and

$$I(u) = \int_\Omega W(C)\mathrm{d}x > \liminf_{j\to\infty} I(u_j) = 0$$

This inequality shows that the functional is not weakly-$*$ lower semicontinuous and therefore $W$ fails to be quasiconvex. The oscillations of $u_j$ on a scale $1/j$ are part of the mathematical model for the microstructures frequently observed in shape memory alloys. More generally, whenever $u$ is a Sobolev function on a domain $\Omega$ such that $Du$ takes only two values, say $Du \in \{A, B\}$, on open sets which are not empty and whose union is $\Omega$ (up to a set of measure zero), then the tangential continuity of the derivatives implies that the difference $A - B$ is a matrix of rank 1, $A - B = a \otimes b$, and that the interfaces between the regions with $Du = A$ and $Du = B$ are hyperplanes with normal parallel to $b$. This statement is usually referred to as "Hadamard's compatibility condition." Moreover, the pattern in Figure 3 is known as a "simple laminate" and the matrices $A$ and $B$ are said to be rank-1 connected.

## Relaxation

The discussion in the previous section shows that the variational problems related to models in materials science typically fail to be weakly lower semicontinuous. One approach which allows us to recover the macroscopic energy of the system and the macroscopic stress–strain relation is to pass to the relaxed variational problem which involves the quasiconvex envelope of the energy density $W$.

**Definition 2** Let $W : \mathbb{M}^{m\times n} \to \mathbb{R}$ be given. The function

$$W^{\mathrm{qc}} = \sup\{f : f \le W, f \text{ quasiconvex}\}$$

is called the quasiconvex envelope of $W$. Equivalently,

$$W^{\mathrm{qc}}(F) = \inf_{\phi \in W_0^{1,\infty}(\Omega;\mathbb{R}^m)} \frac{1}{|\Omega|}\int_\Omega W(F + D\phi)\mathrm{d}x$$

This formula implies that $W^{\mathrm{qc}}$ is the macroscopic energy of the system in the sense that it characterizes the smallest energy per unit volume that is required to subject a volume element to a deformation with affine boundary conditions. Here the system is allowed to minimize its energy with microstructures at any scale, a mechanism which was already explored in the previous section. The arguments in this section prove that $W^{\mathrm{qc}}(C) = 0$ and this shows that the zero set of $W^{\mathrm{qc}}$ can be strictly larger than the zero set of $K$, see Definition 4. The relaxed functional is given by

$$I^{\mathrm{qc}}(u) = \int_\Omega W^{\mathrm{qc}}(Du)\mathrm{d}x$$

Since $W^{\mathrm{qc}}$ satisfies the growth and coercivity conditions [4] if they are satisfied by $W$, the functional $I^{\mathrm{qc}}$ attains its minimum subject to given boundary conditions. The functional $I^{\mathrm{qc}}$ is the weakly lower semicontinuous envelope of $I$ in the sense that minimizing sequences for $I$ contain subsequences that converge to minimizers of $I^{\mathrm{qc}}$ and for all $u$ there exists a sequence $u_j$ which converges in $W^{1,p}(\Omega;\mathbb{R}^m)$ to $u$ such that the energies converge, $I(u_j) \to I(u)$. However, a lot of information in particular about oscillation patterns might be lost in the passage from $I$ to $I^{\mathrm{qc}}$ since the knowledge of a minimizer $u$ for $I^{\mathrm{qc}}$ does not provide any immediate information about the behavior of any minimizing sequence for $I$ that converges to $u$. Moreover, the minimization problem required in the definition of the relaxed energy has been solved explicitly only for very special energy densities.

In this context, one often relies on two related notions of convexity, one sufficient and the other necessary for quasiconvexity. For $F \in \mathbb{M}^{m\times n}$ let $M(F) \in \mathbb{R}^{d(m,n)}$ be the vector of all minors (subdeterminants) of $F$. In the special case $m = n = 2$ we have $M(F) = (F, \det F) \in \mathbb{R}^5$ and for $m = n = 3$ we find $M(F) = (F, \operatorname{cof} F, \det F) \in \mathbb{R}^{19}$ where cof $F$ is the $3 \times 3$ matrix of all $2 \times 2$ subdeterminants of $F$.

**Definition 3** Let $W : \mathbb{M}^{m\times n} \to \mathbb{R}$ be given. The function $W$ is said to be polyconvex if there exists a convex function $g : \mathbb{R}^{d(m,n)} \to \mathbb{R}$ such that $W(F) = g(M(F))$. The function $W$ is rank-1 convex if it is convex along all rank-1 lines in $\mathbb{M}^{m\times n}$, that is, the function $t \mapsto W(F + tR)$ is convex for all $F \in \mathbb{M}^{m\times n}$ and all $R \in \mathbb{M}^{m\times n}$ with $\mathrm{rank}(R) = 1$.

All notions of convexity reduce to classical convexity if $m=1$ or $n=1$. In the vector-valued case $m,n>1$ the following implications are true:

$$f \text{ convex} \Rightarrow f \text{ polyconvex} \Rightarrow f \text{ quasiconvex} \Rightarrow f \text{ rank-1 convex}$$

The reverse statements for the first two implications are not true. Rank-1 convexity does not imply quasiconvexity for $m \geq 3$ and it is a fundamental open problem with deep connections to harmonic analysis to decide whether rank-1 convexity and quasiconvexity are equivalent for $m=n=2$.

The polyconvex and the rank-1 convex envelope of an energy density $W$ are defined analogously to Definition 2. In view of the implications between the different notions of convexity, one has $W^{\mathrm{pc}} \leq W^{\mathrm{qc}} \leq W^{\mathrm{rc}}$ and essentially all explicitly known relaxation formulas are based on the approach to construct a candidate $W^*$ for $W^{\mathrm{rc}}$ and to verify that $W^*$ is polyconvex. Then the inequalities become equalities and one obtains a characterization for the relaxed energy. This approach does not work for extended-valued functions which are used in models for incompressible materials since quasiconvexity does not imply rank-1 convexity in this case. However, for a model system of particular interest, nematic elastomers, a complete characterization of the relaxed energy, the macroscopic stress–strain relation, and the macroscopic phase diagram have been obtained.

## Classical and Generalized Minimizers

The discussion of observed configurations as elements of minimizing sequences $\{u_j\}$ in the section "The direct method in the calculus of variations" leaves the question of the existence of minimizers open. The answer cannot be obtained via the direct methods since minimizing sequences do not need to converge strongly to minimizers. One approach to obtain the existence of solutions $u$ with $I(u)=0$ is to solve the differential relation $Du \in K, u(x)=Fx$ on $\partial\Omega$ by constructing special minimizing sequences that converge strongly so that one can pass to the limit in the energy integral. This idea has led to surprising solutions $u$ with affine boundary conditions for the two-well problem where $K=\mathrm{SO}(2)\mathrm{diag}(\eta,1/\eta) \cup \mathrm{SO}(2)\mathrm{diag}(1/\eta,\eta)$. However, the structure of the solutions is intrinsically complicated in the sense that the phase boundary has infinite length unless the boundary conditions are given by $u(x)=Fx$ with $F \in K$.

More generally, the right tool to pass to the limit in nonlinear functions of $z_j=Du_j$ like the energy is the "Young measure" generated by a subsequence. It is given by a family of probability measures $\nu_x$ that provide statistical information about the distribution of the values of $z_j$ close to a given point $x$. The existence and the fundamental properties of Young measures are described in the following theorem. For simplicity we assume that the sequence $z_j$ is uniformly bounded.

**Theorem 2** (Fundamental theorem on Young measures). *Let $E \subset \mathbb{R}^n$ be measurable, $\mathcal{L}^n(E) < \infty$, and let $z_j : E \to \mathbb{R}^d$ be a measurable and bounded sequence. Then there exists a subsequence $z_k$ and a weakly-$*$ measurable map $\nu : E \to \mathcal{M}(\mathbb{R}^d)$ such that the following assertions are true:*

(i) *The measures $\nu_x$ are non-negative probability measures.*
(ii) *If there exists a compact set $K$ such that $u_k \to K$ in measure, then $\operatorname{supp}\nu_x \subset K$ for a.e. $x \in E$.*
(iii) *If $f \in C(\mathbb{R}^d)$ and if $f(z_k)$ is relatively weakly compact in $L^1(E)$, then $f(z_k) \rightharpoonup \overline{f}$ in $L^1(E)$ where $\overline{f}(x) = \langle \nu_x, f \rangle$.*

Here $\langle \nu_x, f \rangle$ denotes the integration of the function $f$ with respect to the measure $\nu_x$. For example, the Young measure generated by the sequence $Du_j$ constructed in the section "The direct method in the calculus of variations" generates the Young measure $\nu_x = (1/2)\delta_A + (1/2)\delta_B$ (see Figure 3) and

$$I(u_j) = \int_\Omega W(Du_j)\,\mathrm{d}x \to \int_\Omega \int_{\mathbb{M}^{m\times n}} W(Y)\,\mathrm{d}\nu_x(Y)\,\mathrm{d}x = 0$$

A Young measure generated by a sequence of gradients is called a gradient Young measure (GYM). It is said to be homogeneous if $\nu_x = \nu$ is independent of $x$. We restrict our attention in the following to homogeneous GYMs generated by sequences that are bounded in $L^\infty$. The importance of quasiconvexity is also reflected in the following characterization of homogeneous GYMs.

**Theorem 3** *A non-negative probability measure $\nu$ is a GYM if and only if there exists a compact set $K \subset \mathbb{M}^{m\times n}$ with $\operatorname{supp}\nu \subset K$ and Jensen's inequality $\langle \nu, f \rangle \geq f(\langle \nu, \mathrm{id} \rangle)$ holds for all quasiconvex functions $f : \mathbb{M}^{m\times n} \to \mathbb{R}$.*

This motivates to characterize the generalized limits of minimizing sequences as

$$\mathcal{M}^{\mathrm{qc}}(K) = \{\nu \in \mathcal{M}(K) :\ f(\langle \nu, \mathrm{id}\rangle) \leq \langle \nu, f \rangle \text{ for all } f : \mathbb{M}^{m\times n} \to \mathbb{R} \text{ quasiconvex}\}$$

where $\mathcal{M}(K)$ is the set of all probability measures supported on $K$. If $\nu$ is generated by a sequence of

functions with affine boundary conditions $u_j(x) = Fx$, then $\langle \nu, \mathrm{id} \rangle = F$. The set of all affine deformations of the material that can be recovered by heating (shape memory effect) is therefore given as the set of all centers of mass of homogeneous GYMs supported on $K$, the so-called "quasiconvex hull" $K^{\mathrm{qc}}$ of $K$.

**Definition 4** Suppose that $K \subset \mathbb{M}^{m\times n}$ is compact. We define the quasiconvex hull of $K$ by

$$K^{\mathrm{qc}} = \{F = \langle \nu, \mathrm{id} \rangle : \ \nu \in \mathcal{M}^{\mathrm{qc}}(K)\}$$

There are several equivalent definitions of $K^{\mathrm{qc}}$. The foregoing definition corresponds to the definition of the convex hull of a set as the set of all centers of mass of probability measures supported on $K$ (which satisfy Jensen's inequality for all convex $f$). The set $K^{\mathrm{qc}}$ can also be defined as the set of all points that cannot be separated by quasiconvex functions from $K$ or as the zero set of the quasiconvex envelope of the distance function to $K$. The "polyconvex hull" $K^{\mathrm{pc}}$ and the "rank-1 convex hull" $K^{\mathrm{rc}}$ are defined analogously by replacing quasiconvexity with polyconvexity and rank-1 convexity in the foregoing definitions. It follows that $K^{\mathrm{rc}} \subset K^{\mathrm{qc}} \subset K^{\mathrm{pc}}$ and all of these inclusions can be strict.

A particularly useful set of conditions are the minors conditions

$$\langle \nu, M \rangle = M(\langle \nu, \mathrm{id} \rangle)$$

for all minors $M$ which follow from the weak continuity of the minors. For example, if $K = \{A, B\} \subset \mathbb{M}^{2\times 2}$, then any probability measure supported on $K$ is given by $\nu = \lambda\delta_A + (1-\lambda)\delta_B$. The minors condition with $M = \det$ implies that

$$\begin{aligned}\det(\lambda A + (1-\lambda)B) &= \det\langle \nu, \mathrm{id} \rangle = \langle \nu, \det \rangle \\ &= \lambda \det A + (1-\lambda)\det B\end{aligned}$$

This identity is equivalent to

$$\lambda(1-\lambda)\det(A-B) = 0$$

and therefore the quasiconvex hull is equal to $K$ if and only if $\det(A-B) \neq 0$. A very instructive example is the set $K = \{(1,3), (-1,-3), (-3,1), (3,-1)\}$ viewed as a subset of the space of all diagonal matrices in $\mathbb{M}^{2\times 2}$. It is frequently referred to as a $T_4$ configuration. The rank-1 convex hull is equal to the quasiconvex hull and given by the four points, the line segments, and the square in the center, the polyconvex hull is bounded by four hyperbolic arcs, and the convex hull is the square with the points as corners, see Figure 4. It is remarkable that the rank-1 convex hull is strictly larger than the set $K$ itself despite the fact that the set $K$ does not contain any rank-1 connections.

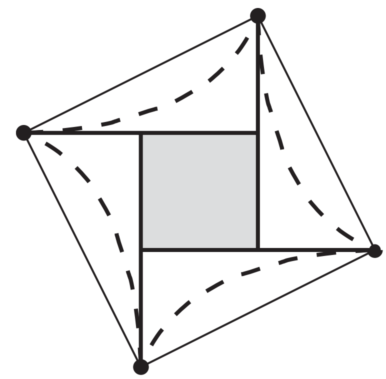


**Figure 4** The four-point subset $K$ in the space of all diagonal matrices and its convex hulls: $K^{\mathrm{rc}} = K^{\mathrm{qc}}$ are given by $K$, the line segments and the shade square, $K^{\mathrm{pc}}$ is bounded by the dashed hyperbolic arcs, and the convex hull is the outer square.

There are only a few examples in which explicit characterizations of the convex hulls for sets invariant under $SO(n)$ have been obtained. For $K = SO(3)U_1 \cup SO(3)U_2$ (see [2]), one finds

$$K^{\mathrm{qc}} = \left\{ F \in \mathbb{M}^{3\times 3} : F^T F = \begin{pmatrix} a & c & 0 \\ c & b & 0 \\ 0 & 0 & 1/\eta^2 \end{pmatrix}, \right.$$

$$\left. ab - c^2 = \eta^2, \ a + b + 2|c| \leq \eta^4 + \frac{1}{\eta^2} \right\}$$

The quasiconvex hull of the three-well problem [1] is not known. In two dimensions one finds for

$$\begin{aligned}&K = SO(2)U_1 \cup \cdots \cup SO(2)U_n, \\ &\det U_i = 1, \ i = 1, \ldots, n\end{aligned}$$

that

$$K^{\mathrm{qc}} = \left\{ F \in \mathbb{M}^{2\times 2} : \ \det F = 1, |Fe|^2 \leq \max_{i=1,\ldots,n} |U_i e|^2 \right\}$$

All examples in which envelopes of functions or hulls of sets have been obtained explicitly are based on the exceptional property that the polyconvex envelope coincides with the rank-1 convex envelope. The $T_4$ configuration in Figure 4 is one of the few cases where the quasiconvex hull is known to be different from the polyconvex hull. The construction of quasiconvex functions and the understanding of their properties is one of the challenges left for the future.

## Bibliographical Remarks

This article can only review some of the highlights of mathematical developments related to models in

nonlinear elasticity for solid–solid phase transformations based on a huge body of work in the original literature. The precise references can be found in the extensive bibliographies of the books and review articles that are cited in the subsequent section, in particular in Ball (2004), Bhattacharya (2003), Dolzmann (2003), James and Hane (2000), and Müller (1999). This article focuses on models for single crystals; the behavior of polycrystals (which strongly depends on the amount of symmetry breaking in the transformation) was studied by Bhattacharya and Kohn.

The formulation of solid–solid phase transformations via nonlinear continuum theory goes back to Ericksen and the analysis via tools in the calculus of variations was initiated by Ball and James, Chipot and Kinderlehrer, and Fonseca. The Russian school developed the theory in linear elasticity in the 1960s, see Khachaturyan (1983) for a review. A detailed discussion of the crystallographic and group-theoretical aspects is contained in Pitteri and Zanzotto (2002).

Quasiconvexity was introduced by Morrey (1966) and his results were extended to Carathéodory integrands by Acerbi and Fusco and Marcellini. A modern treatment including Dacorogna's relaxation theorem and a summary of the various notions of convexity and their properties can be found in Dacorogna (1989). Šverák proved that rank-1 convexity does not imply quasiconvexity for $m \geq 3$ and Milton modified his example to show that the rank-1 convex hull of a set can be strictly smaller than its quasiconvex hull. The explicit characterizations for nematic elastomers were obtained by DeSimone and Dolzmann.

Lipschitz solutions to differential inclusions were constructed by Müller and Šverák based on Gromov's concept of convex integration, by Dacorogna and Marcellini using Baire's category argument, and by Kirchheim in the framework of Banach Mazur games. The structure of solutions of the two-well problem with finite surface energy was analyzed by Dolzmann and Müller. Young measures (also called parametrized measures or chattering controls) were originally introduced as generalized solutions for optimal control problems which do not admit classical solutions (Young 1969). Tartar (1979) introduced Young measures as a fundamental tool for the analysis of oscillation effects in partial differential equations and for the passage from microscopic to macroscopic models. Gradient Young measures were characterized by Kinderlehrer and Pedregal. The four-point configuration was discovered independently in various contexts by several authors including Scheffer, Aumann and Hart, Casadio Tarabusi, Tartar, and Milton and Nesi. The characterization of the quasiconvex hull uses a quasiconvex function constructed by Šverák. The quasiconvex hull of the two-well problem in 3D was found by Ball and James, and the generalization to $n$ wells in 2D by Bhattacharya and Dolzmann.

## Acknowledgments

The work of G Dolzmann was supported by the NSF through grants DMS0405853 and DMS0104118.

*See also:* Gamma-Convergence and Homogenization; Variational Techniques for Ginzburg–Landau Energies.

## Further Reading

Ball JM (2004) Mathematical models of martensitic microstructure. *Materials Science and Engineering A* 378(1–2): 61–69.

Bhattacharya K (2003) *Microstructure of Martensite*. Oxford: Oxford University Press.

Dacorogna B (1989) *Direct Methods in the Calculus of Variations*. Berlin: Springer.

Dolzmann G (2003) *Variational Methods for Crystalline Microstructure: Analysis and Computation*, Lecture Notes in Mathematics, vol. 1803. Berlin: Springer.

James RD and Hane KF (2000) Martensitic transformations and shape-memory materials. *Acta Materiali* 48: 197–222.

Khachaturyan A (1983) *Theory of Structural Transformations in Solids*. New York: Wiley.

Morrey CB (1966) *Multiple Integrals in the Calculus of Variations*. Berlin: Springer.

Müller S (1999) Variational methods for microstructure and phase transitions. *Proc. C.I.M.E. Summer School "Calculus of Variations and Geometric Evolution Problems," Cetraro, 1996*, Lecture Notes in Mathematics, vol. 1713. Berlin–Heidelberg: Springer.

Pitteri M and Zanzotto G (2002) *Continuum Models for Phase Transitions and Twinning in Crystals*. London: Chapman and Hall.

Tartar L (1979) Compensated compactess and partial differential equations. In: Knops R (ed.) *Nonlinear Analysis and Mechanics: Heriot–Watt Symposion*, vol. IV. London: Pitman.

Young LC (1969) *Lectures on the Calculus of Variations and Optimal Control Theory*. Philadelphia–London–Toronto: Saunders.

**Vertex Operator Algebras** *see* Two-Dimensional Conformal Field Theory and Vertex Operator Algebras

# Viscous Incompressible Fluids: Mathematical Theory

**J G Heywood**, University of British Columbia, Vancouver, BC, Canada



## Introduction

The Navier–Stokes equations

$$\rho(u_t + u \cdot \nabla u) = -\nabla p + \mu \Delta u + f \qquad [1]$$

$$\nabla \cdot u = 0 \qquad [2]$$

provide the simplest model for the motion of a viscous incompressible fluid that is consistent with the principles of mass and momentum conservation, and with Stokes' hypothesis that the internal forces due to viscosity must be invariant with respect to any superimposed rigid motion of the reference frame. Despite their simplicity, they seem to govern the motion of air, water, and many other fluids very accurately over a wide range of conditions. Thus, their mathematical theory is central to the rigorous analysis of many experimental observations, from the asymptotics of steady wakes and jets, to the dynamics of convection cells, vortex shedding, and turbulence. During the last 80 years, a great deal of progress has been made on both the basic mathematical theory of the equations and on its application to the understanding of such phenomena. But one of the most important matters, that of estimating the regularity of solutions over long periods of time, remains a vexing and fascinating challenge. Such an estimate will almost certainly be needed to prove the "global" existence of smooth solutions. By that we mean the existence of smooth solutions of the initial-value problem over indefinitely long periods of time without any restriction on the "size" of the data. To date we can prove the "local" existence of smooth solutions, but there remains a concern that if the data are large, solutions may develop singularities within a finite period of time. In fact, there is a great deal more at issue than this question of existence. A regularity estimate is required to prove the reliability of the equations as a predictive model. That is because any estimate for the continuous dependence of solutions on the prescribed data for a problem depends upon a regularity estimate, as do error estimates for numerical approximations. A global estimate for the regularity of solutions is also required for a mathematically rigorous theory of turbulence. In fact, it may be hoped that the insight which ultimately yields a global regularity estimate will also be pivotal to our understanding of turbulence, perhaps justifying Kolmogorff theory; see Heywood (2003). In this article we aim to present a relatively simple approach to the local existence, uniqueness, and regularity theory for the initial boundary value problem for the Navier–Stokes equations, and to discuss some observations that bear on the question of global regularity. A wider-ranging review of open problems is given in Heywood (1990), and further observations concerning the problem of global regularity are given in Heywood (1994).

## Setting the Problem

To focus on core issues, we shall make some simplifying assumptions. The fluid under consideration will be assumed to completely fill (without free boundaries or vacuums) a bounded, connected, time-independent domain $\Omega \subset R^n, n = 2$ or 3, with smooth boundary $\partial\Omega$. We are mainly interested in the three-dimensional case, but comparisons with the two-dimensional case are illuminating. The $R^n$-valued velocity $u(x,t) = (u_1(x,t), \ldots, u_n(x,t))$ and $R$-valued pressure $p(x,t)$ are functions of the position $x = (x_1, \ldots, x_n) \in \Omega$ and time $t \geq 0$. Equation [1] is an expression of Newton's second law of motion, equating mass density times acceleration on the left with several force densities on the right, due to pressure and viscosity, and sometimes a prescribed external force $f$. Written in full, using the summation convention over repeated indices, its $i$th component is

$$\rho\left(\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j}\right) = -\frac{\partial p}{\partial x_i} + \mu \frac{\partial^2 u_i}{\partial x_j^2} + f_i$$

We will assume the density $\rho$ and the coefficient of viscosity $\mu$ are positive constants.

In this article, we consider the initial boundary value problem consisting of the equations [1], [2] together with the initial and boundary conditions

$$u|_{t=0} = u_0, \qquad u|_{\partial\Omega} = 0 \qquad [3]$$

The initial velocity $u_0(x)$ is prescribed. It will be assumed to possess whatever smoothness is convenient, and to satisfy $\nabla \cdot u_0 = 0$ and $u_0|_{\partial\Omega} = 0$. The boundary condition is a reasonable one, since fluids adhere to rigid surfaces.

Notice that a further condition would be needed to uniquely determine the pressure, since only its derivatives appear in the problem as posed. We prefer to do without auxiliary conditions for the pressure, and to refer to $u$ by itself as a solution of

the problem provided there exists a scalar function $p$ which together with $u$ satisfies [1]–[3]. The problem is said to be uniquely solvable if there is a unique solution $u$, in which case the gradient of the pressure is also uniquely determined, along with the pressure up to a constant. Notice also that under our assumptions a potential force like gravity has no effect on $u$. If $u$ solves the problem in the absence of such a force, then the inclusion of the force affects only the pressure, from which the potential must be subtracted. It turns out that the inclusion of a prescribed nonpotential force, while complicating many of the estimates below, does not affect in any essential way those parts of the theory to be presented here. Thus, for simplicity, we shall henceforth assume that $f \equiv 0$.

## Reynolds Number

We can make a slight further simplification of eqn [1] by rescaling, with the objective of setting $\rho = 1$, or even $\rho = 1$ and $\mu = 1$. This scaling is not required for the existence theory we are presenting, but provides an important insight for the study of stability, bifurcation, and turbulence. The Reynolds number

$$R = \frac{\max|u| \cdot |\Omega| \cdot \rho}{\mu}$$

plays an important role in rescaling. It expresses the ratio of the inertial to viscous effects. The notation $|\Omega|$ represents a characteristic length, such as the minimum diameter of a bounded domain. Generally speaking, a high Reynolds number corresponds to what is meant by "large" data, and the higher the Reynolds number the more inclined a flow is to instability and turbulence, and perhaps to the development of singularities. However, the size of the Reynolds number has precise implications only in comparing "dynamically similar" flows. We say that two vector fields $v(x,t)$ and $u(x,t)$ are dynamically similar if and only if $v(x,t) = \alpha u(x/\beta, t/\gamma)$ for some $\alpha, \beta, \gamma > 0$. In such a case, if $u$ is defined in $\Omega \times [0, T)$, then $v$ will be defined in $\beta\Omega \times [0, \gamma T)$, where $\beta\Omega = \{\beta x \colon x \in \Omega\}$. Furthermore, if $u$ satisfies the Navier–Stokes equations, then $v$ will satisfy

$$\begin{aligned} &\rho\alpha^{-1}\gamma v_t + \rho\alpha^{-2}\beta v \cdot \nabla v \\ &\quad = -\beta\nabla p(x/\beta, t/\gamma) + \alpha^{-1}\beta^2\mu\Delta v \end{aligned} \quad [4]$$

which has the form of the Navier–Stokes equations if and only if the coefficients of the two inertial terms on the left-hand side are equal. That is, if and only if

$$\alpha\gamma = \beta \quad [5]$$

in which case

$$v_t + v \cdot \nabla v = -\nabla q + \eta\Delta v \quad [6]$$

with

$$\eta = \alpha\beta\mu/\rho \quad [7]$$

and $q(x,t) = \alpha^2\rho^{-1}p(x/\beta, t/\gamma)$. We refer to such $u$ and $v$ as dynamically similar flows. The relation [7], that follows from [5], is equivalent to the equality of the Reynolds numbers for the two flows,

$$\begin{aligned} R(u) &= \frac{\max|u| \cdot |\Omega| \cdot \rho}{\mu} \\ &= \frac{\max|\alpha u| \cdot |\beta\Omega| \cdot 1}{\eta} = R(v) \end{aligned}$$

The condition [5] can be satisfied simultaneously with the condition $\eta = 1$. For example, one may choose $\beta = 1, \alpha = \rho/\mu$, and $\gamma = \mu/\rho$. This achieves a rescaling of the equation to

$$v_t + v \cdot \nabla v = -\nabla q + \Delta v \quad [8]$$

without changing the domain. Different Reynolds numbers result from varying the magnitude of the velocity. In what follows, we will work with the Navier–Stokes equation in this simplest possible form.

## Continuous Dependence on the Data

We begin our investigation of the initial boundary value problem

$$\begin{aligned} &u_t + u \cdot \nabla u = -\nabla p + \Delta u, \quad \nabla \cdot u = 0 \\ &\text{for } (x,t) \in \Omega \times (0, \infty), \\ &u|_{t=0} = u_0, \quad u|_{\partial\Omega} = 0 \end{aligned} \quad [9]$$

by considering two smooth solutions, say $u$ and $v$, taking possibly different initial values $u_0$ and $v_0$. Let their difference be $w = v - u$, with initial value $w_0$, and let $q$ be the difference of the corresponding pressures. Then, subtracting one equation from the other, one obtains

$$w_t + w \cdot \nabla w + u \cdot \nabla w + w \cdot \nabla u = -\nabla q + \Delta w \quad [10]$$

Multiplying this by $w$, integrating over $\Omega$, and integrating by parts, one then obtains

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|w\|^2 + \|\nabla w\|^2 = -(w \cdot \nabla u, w) \quad [11]$$

where

$$\|w\|^2 = \int_\Omega w^2\,\mathrm{d}x$$
$$\|\nabla w\|^2 = \int_\Omega \frac{\partial w_i}{\partial x_j}\frac{\partial w_i}{\partial x_j}\,\mathrm{d}x$$
$$(w\cdot\nabla u, w) = \int_\Omega w_j \frac{\partial u_i}{\partial x_j} w_i\,\mathrm{d}x$$

since (and this should further explain our notation)

$$\begin{aligned}(w_t, w) &= \int_\Omega w_t\cdot w\,\mathrm{d}x\\ &= \frac{1}{2}\frac{d}{dt}\int_\Omega w^2\,\mathrm{d}x = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|w\|^2\\ (\Delta w, w) &= \int_\Omega \frac{\partial^2 w_i}{\partial x_j^2} w_i\,\mathrm{d}x\\ &= -\int_\Omega \frac{\partial w_i}{\partial x_j}\frac{\partial w_i}{\partial x_j}\,\mathrm{d}x = -\|\nabla w\|^2\\ (\nabla q, w) &= \int_\Omega \frac{\partial q}{\partial x_i} w_i\,\mathrm{d}x = -\int_\Omega q\frac{\partial w_i}{\partial x_i}\,\mathrm{d}x = 0\\ (u\cdot\nabla w, w) &= \int_\Omega u_j\frac{\partial w_i}{\partial x_j} w_i\,\mathrm{d}x\\ &= -\frac{1}{2}\int_\Omega \frac{\partial u_j}{\partial x_j} w_i w_i\,\mathrm{d}x = 0\end{aligned}$$

and similarly $(w\cdot\nabla w, w)=0$. In deriving these we have used the fact that the vector fields are divergence free and vanish on the boundary. In the following, we will use such identities without further mention.

We can estimate the nonlinear term on the right-hand side of [11] by using the "Sobolev inequalities"

$$\begin{aligned}\|\phi\|_4^2 &\le \|\phi\|\|\nabla\phi\|, &&\text{if } n=2\\ \|\phi\|_4^2 &\le \|\phi\|^{1/2}\|\nabla\phi\|^{3/2}, &&\text{if } n=3\end{aligned} \qquad [12]$$

proved by Ladyzhenskaya (1969), though with larger constants. These are valid for any smooth function $\phi$ which vanishes on the boundary of $\Omega$. It may be either scalar or vector valued. The norms on the left are $L^4$-norms; we use the notation $\|\phi\|_p = (\int_\Omega|\phi|^p\,\mathrm{d}x)^{1/p}$ for any $p>1$, but usually drop the subscript when $p=2$. Using first Hölder's inequality and then [12], one obtains

$$|(w\cdot\nabla u, w)| \le \|w\|_4^2\|\nabla u\| \le \begin{cases}\|w\|\|\nabla w\|\|\nabla u\| & \text{if } n=2\\ \|w\|^{1/2}\|\nabla w\|^{3/2}\|\nabla u\| & \text{if } n=3\end{cases}$$

Young's inequality

$$ab \le \frac{1}{p}a^p + \frac{1}{q}b^q$$

holds if $a,b>0, p,q>1$ and $1/p+1/q=1$. Taking $a=\sqrt{2}\|\nabla w\|$, along with $p=q=2$ in the two-dimensional case, and $a=(4/3)^{3/4}\|\nabla w\|$, along with $p=4/3$, $q=4$ in the three-dimensional case, one obtains

$$|(w\cdot\nabla u, w)| \le \begin{cases}\|\nabla w\|^2+\frac{1}{4}\|\nabla u\|^2\|w\|^2, & \text{if } n=2\\ \|\nabla w\|^2+\frac{27}{256}\|\nabla u\|^4\|w\|^2, & \text{if } n=3\end{cases} \qquad [13]$$

Using these estimates for the right-hand side of [11], we obtain linear differential inequalities for $\|w\|^2$ that are easily integrated to give

$$\|w(t)\|^2 \le \begin{cases}\|w_0\|^2\exp\int_0^t \frac{1}{2}\|\nabla u\|^2\,\mathrm{d}\tau, & \text{if } n=2\\ \|w_0\|^2\exp\int_0^t \frac{27}{128}\|\nabla u\|^4\,\mathrm{d}\tau, & \text{if } n=3\end{cases} \qquad [14]$$

It follows that if we can estimate the integrals on the right, which concern only the solution $u$, and if $v$ is a second solution, perhaps differing only slightly from $u$ when $t=0$, then we can estimate the difference $\|v(t)-u(t)\|$ at later times. Moreover, at any particular time this difference will be bounded proportionally to $\|v(0)-u(0)\|$. The integral on the right-hand side of the two-dimensional version of [14] is easily estimated using the energy estimate [16] below. The estimation of the corresponding integral in the three-dimensional case, without a restriction on the size of the data, remains an open problem. It can be regarded as the most important open problem in the Navier–Stokes theory. It would never be enough to somehow prove that solutions are smooth without estimating this integral, or something equivalent to it. Of course, if solutions were known to be smooth one could infer their uniqueness from [14], since smoothness would imply that the integrals are finite, which is enough to conclude that $\|w(t)\|$ is zero if $\|w_0\|$ is zero.

## Energy Estimate

If one multiplies the Navier–Stokes equation for $u$ by $u$, and proceeds as in deriving [11], one obtains

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u\|^2+\|\nabla u\|^2=0 \qquad [15]$$

and hence

$$\frac{1}{2}\|u(t)\|^2+\int_0^t\|\nabla u\|^2\,\mathrm{d}\tau = \frac{1}{2}\|u_0\|^2 \qquad [16]$$

This settles the matter of continuous dependence in the two-dimensional case. Together with [16], the two-dimensional version of [14] implies

$$\|w(t)\|^2 \leq \|w_0\|^2 \exp\tfrac{1}{4}\|u_0\|^2, \quad \text{if } n = 2 \qquad [17]$$

We remark that the local rate of energy dissipation is $2|Du|^2$ rather than $|\nabla u|^2$, where $Du$ is the stress tensor $Du = (1/2)(\nabla u + (\nabla u)^{\mathrm{T}})$. However, integrating over the domain, and integrating by parts using the boundary condition $u|_{\partial\Omega} = 0$, one may verify that the rate of total energy dissipation $2\|Du\|^2$ equals $\|\nabla u\|^2$. For the purpose of this article, it is convenient to write the energy identity as [15].

## Estimates for $\|\nabla u(t)\|$ Pointwise in Time

Of course, an estimate for $\|\nabla u(t)\|$ pointwise in time would imply an estimate for the integral of $\|\nabla u(t)\|^4$ on the right-hand side of [14]. We can prove such an estimate for at least a finite interval of time by an argument due to Prodi (1962). It requires, in preparation, some deep results concerning the regularity of solutions of the steady Stokes equations. These cannot be proved here, but we can briefly summarize what will be needed. Let

$\boldsymbol{L}^2(\Omega)$ = space of vector fields $\phi$, with finite $\boldsymbol{L}^2$-norms $\|\phi\|$,
$\boldsymbol{C}_0^\infty(\Omega)$ = space of smooth vector fields with compact support in $\Omega$,
$\boldsymbol{D}(\Omega) = \{\phi \in \boldsymbol{C}_0^\infty(\Omega): \nabla\cdot\phi = 0\}$,
$\boldsymbol{J}(\Omega)$ = completion of $\boldsymbol{D}(\Omega)$ in the $\boldsymbol{L}^2$-norm $\|\phi\|$,
$\boldsymbol{J}_1(\Omega)$ = completion of $\boldsymbol{D}(\Omega)$ in the norm $\|\nabla\phi\|$,
$\boldsymbol{G}(\Omega) = \{\nabla p: p \in L^2(\Omega) \text{ with } \nabla p \in \boldsymbol{L}^2(\Omega)\}$, and
$P: \boldsymbol{L}^2(\Omega) \to \boldsymbol{J}(\Omega)$ be the $\boldsymbol{L}^2$-projection of $\boldsymbol{L}^2(\Omega)$ onto $\boldsymbol{J}(\Omega)$,

and define the Sobolev $W_2^2(\Omega)$ norm by

$$\|u\|^2_{W_2^2(\Omega)} = \|u\|^2 + \|\nabla u\|^2 + \int_\Omega \left|\partial^2 u_i/\partial x_j \partial x_k\right|^2 \mathrm{d}x$$

Furthermore, observe that $(\nabla p, \phi) = 0$ for $\nabla p \in \boldsymbol{G}(\Omega)$ and $\phi \in \boldsymbol{J}(\Omega)$, since it holds if $p$ is smooth and $\phi \in \boldsymbol{D}(\Omega)$. Therefore, $P\nabla p = 0$, since $(P\nabla p, \phi) = (\nabla p, \phi) = 0$, for all $\phi \in \boldsymbol{J}(\Omega)$. Later, when we need it, we will also argue that $\boldsymbol{L}^2(\Omega) = \boldsymbol{J}(\Omega) \oplus \boldsymbol{G}(\Omega)$.

With these preparations, it is evident that every smooth vector field $u$ satisfying $\nabla\cdot u = 0$ and $u|_{\partial\Omega} = 0$ can be regarded as a solution of the steady Stokes problem

$$-\Delta u + \nabla p = f \text{ and } \nabla\cdot u = 0 \text{ in } \Omega \;\; u|_{\partial\Omega} = 0 \qquad [18]$$

with $f = -P\Delta u$. For such solutions, and hence for all such $u$, we have the estimates

$$\|u\|_{W_2^2(\Omega)} \leq c\|P\Delta u\| \qquad [19]$$

and

$$\sup_\Omega |u|^2 \leq \begin{cases} c\|u\|\|P\Delta u\|, & \text{if } n = 2 \\ c\|\nabla u\|\|P\Delta u\|, & \text{if } n = 3 \end{cases} \qquad [20]$$

with constants independent of $u$. It can also be shown that every such vector field $u$ belongs to $\boldsymbol{J}_1(\Omega)$ and hence to $\boldsymbol{J}(\Omega)$; see Heywood (1973).

Some history and remarks are in order. The inequality [19] was proved independently by Solonnikov (1964, 1966), and by Prodi's student Cattabriga (1961). In fact, they gave $L^p$ versions of it for all orders of the derivatives. Several proofs specific to the $L^2$ case needed here have been given by Solonnikov and Sčadilov (1973) and by Beirão da Veiga (1997). The inequalities [20] can be proved by combining [19] with appropriate Sobolev inequalities, or better, by combining [19] with recent inequalities of Xie (1991) which are of precisely the form [20], but with $\triangle u$ instead of $P\triangle u$ on the right-hand side, and without the requirement that $\nabla\cdot u = 0$. The constant $c$ in [19] depends upon the regularity of the boundary, and tends to infinity along with a bound for the boundary curvature. Through the work of Xie (1992, 1997), there is reason to believe that the inequalities [20] are probably valid for arbitrary domains, with the constant $c = (2\pi)^{-1}$ if $n = 2$, and $c = (3\pi)^{-1}$ if $n = 3$. Xie's efforts to prove this have been continued by the author (Heywood 2001). If the inequalities [20] can be proved for arbitrary domains (i.e., arbitrary open sets), with these fixed constants, then the approach to Navier–Stokes theory presented in this article will extend immediately to arbitrary domains, as explained in Heywood and Xie (1997), with estimates independent of the domain.

We go on now with an estimation of $\|\nabla u(t)\|$ based on [20]. Multiplying the Navier–Stokes equation for $u$ by $-P\triangle u$, and integrating over $\Omega$, one obtains

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\nabla u\|^2 + \|P\Delta u\|^2 = (u\cdot\nabla u, P\Delta u) \leq \sup_\Omega |u|\|\nabla u\|\|P\Delta u\| \qquad [21]$$

since $(u_t, -P\Delta u) = (Pu_t, -\Delta u) = (u_t, -\Delta u) = (\nabla u_t, \nabla u)$ and $(\nabla p, P\Delta u) = 0$.

The right-hand side of [21] can be estimated using [20] and Young's inequality:

$$\sup_{\Omega} |u|\|\nabla u\|\|P\Delta u\|$$
$$\leq \begin{cases} c\|u\|^{1/2}\|\nabla u\|\|P\Delta u\|^{3/2}, & \text{if } n=2 \\ c\|\nabla u\|^{3/2}\|P\Delta u\|^{3/2}, & \text{if } n=3 \end{cases}$$
$$\leq \begin{cases} \frac{1}{2}\|P\Delta u\|^2 + c\|u\|^2\|\nabla u\|^4, & \text{if } n=2 \\ \frac{1}{2}\|P\Delta u\|^2 + c\|\nabla u\|^6, & \text{if } n=3 \end{cases}$$

Thus,

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\nabla u\|^2 + \|P\Delta u\|^2 \leq \begin{cases} c\|u\|^2\|\nabla u\|^4, & \text{if } n=2 \\ c\|\nabla u\|^6, & \text{if } n=3 \end{cases} \qquad [22]$$

These differential inequalities are at the core of present theory. Consider first the two-dimensional case. It can be viewed as a linear differential inequality

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\nabla u\|^2 \leq \left(c\|u\|^2\|\nabla u\|^2\right)\|\nabla u\|^2 \qquad [23]$$

with a coefficient $c\|u\|^2\|\nabla u\|^2$ that is integrable, in view of the energy estimate [16]. Integrating it yields a "global" estimate; for all $t \geq 0$,

$$\|\nabla u(t)\|^2 \leq \|\nabla u_0\|^2 \exp\int_0^t \|u\|^2\|\nabla u\|^2\,\mathrm{d}\tau$$
$$\leq \|\nabla u_0\|^2 \exp\frac{1}{2}\|u_0\|^4 \qquad [24]$$

However, if the three-dimensional version of [22] is viewed as a linear differential inequality, the coefficient to be integrated is $\|\nabla u(t)\|^4$. Thus, the same integral which is crucial to proving continuous dependence on the data is also crucial to proving regularity. What we can do in the three-dimensional case, is view [22] as a nonlinear differential inequality of the form

$$\varphi' \leq c\varphi^3 \quad \text{or} \quad \varphi' \leq c\|\nabla u\|^2\varphi^2 \qquad [25]$$

for $\varphi(t) = \|\nabla u(t)\|^2$. Integrating the first of these, one obtains a local estimate

$$\|\nabla u(t)\|^2 \leq \frac{\|\nabla u_0\|^2}{\sqrt{1-2c\|\nabla u_0\|^4 t}} \qquad [26]$$

for

$$0 \leq t < \frac{1}{2c\|\nabla u_0\|^4}$$

without any restriction on the size of the data. Integrating the second, one obtains a global estimate

$$\|\nabla u(t)\|^2 \leq \frac{\|\nabla u_0\|^2}{1 - c\|\nabla u_0\|^2 \int_0^t \|\nabla u\|^2\,\mathrm{d}\tau}$$
$$\leq \frac{\|\nabla u_0\|^2}{1-(c/2)\|u_0\|^2\|\nabla u_0\|^2} \qquad [27]$$

valid for all $t \geq 0$, provided

$$\|u_0\|^2\|\nabla u_0\|^2 < \frac{2}{c} \qquad [28]$$

This is a good interpretation of what we mean by "small data." If Xie's conjecture is correct, that the constant in the three-dimensional version of [20] is $c=(3\pi)^{-1}$, then we obtain [25]–[28] with the constant $c = 3/(128\pi^2)$. Thus, $2/c \simeq 842$.

## Further Regularity, Smoothing Estimates

Once one has an estimate of the form

$$\|\nabla u(t)\| \leq M(t), \quad \text{for } 0 \leq t < T \qquad [29]$$

as provided by [24], [26], or [27], one can estimate the solution's derivatives of all orders over the open time interval $(0, T)$. The initial time $t=0$ must be excluded from the interval, because the "imperfection" of prescribed data generally causes an impulsive acceleration along the boundary at time zero, resulting in a thin boundary layer in which the derivatives are so large that $\|\nabla u_t(t)\|$ and $\|u(t)\|_{W_2^3(\Omega)}$ tend to infinity as $t \to 0^+$. But the solution quickly smooths and remains smooth as long as [29] remains in force. Thus, our working assumption up to this point, that solutions are $C^\infty$ smooth in $\overline{\Omega} \times [0,\infty)$ is not valid at $t=0$. However, we will see that they are smooth in $\overline{\Omega} \times (0, T)$ and continuous in $\overline{\Omega} \times [0, T)$. They are also continuous on $[0, T)$ in the $W_2^2(\Omega)$ norm. This is sufficient regularity to justify everything that we have done to this point.

In this section, we give estimates for the derivatives of all orders with respect to time, of $u$ and its first- and second-order derivatives with respect to space. In the next section, we will prove an existence theorem by Galerkin approximation. It will be easily seen that all of the estimates proved in this and previous sections, for solutions that are assumed to be smooth, also hold for the approximations, without any unproven assumptions. Therefore, they will be inherited by the solution that is obtained upon passing to the limit of the approximations. At first, this solution will be something of a generalized solution, not fully classical, but one which is $C^\infty$ with respect to time over the interval $0 < t < T$, in the $W_2^2(\Omega)$ norm. In a final step,

viewing $u$ at any fixed time as a solution of the steady Stokes equations, we can apply regularity estimates for the Stokes equations to infer that it is $C^\infty$ in all variables throughout $\overline{\Omega} \times (0, T)$, with specific estimates for each derivative.

The estimates of this section are obtained by integrating an infinite sequence of differential inequalities, for $\|u\|, \|\nabla u\|, \|u_t\|, \|\nabla u_t\|, \|u_{tt}\|, \|\nabla u_{tt}\|, \ldots$. The first two are [15] and [21], which have already been dealt with. It turns out that after these first two, each succeeding differential inequality is linearized by the estimates obtained from its predecessor, which explains why the time intervals for these additional estimates do not become successively shorter. In fact, in the two-dimensional case, the energy estimate resulting from [15], which is valid for all time, already gives the linearization [23] of [21], which then provides an estimate valid for all time. Except for noting such differences between the two- and three-dimensional cases, we will henceforth deal with only the three-dimensional case.

The differential inequalities just mentioned are obtained by estimating the right-hand sides of two sequences of differential identities, and ordering them by an iteration between the two sequences. The first sequence begins with and is patterned after the energy identity,

$$\begin{aligned}
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u\|^2 + \|\nabla u\|^2 = 0\\
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u_t\|^2 + \|\nabla u_t\|^2 = -(u_t\cdot\nabla u, u_t)\\
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u_{tt}\|^2 + \|\nabla u_{tt}\|^2 = -(u_{tt}\cdot\nabla u, u_{tt})\\
&\qquad\qquad - 2(u_t\cdot\nabla u_t, u_{tt})
\end{aligned} \qquad [30]$$

etc.

while the second begins with and is patterned after Prodi's identity,

$$\begin{aligned}
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\nabla u\|^2 + \|P\Delta u\|^2 = (u\cdot\nabla u, P\Delta u)\\
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\nabla u_t\|^2 + \|P\Delta u_t\|^2 = (u_t\cdot\nabla u, P\Delta u_t)\\
&\qquad\qquad + (u\cdot\nabla u_t, P\Delta u_t)\\
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\nabla u_{tt}\|^2 + \|P\Delta u_{tt}\|^2 = (u_{tt}\cdot\nabla u, P\Delta u_{tt}) + \cdots
\end{aligned} \qquad [31]$$

etc.

Before going on, notice that we can return to [22] and use [29] to infer a more complete estimate of the form

$$\|\nabla u(t)\|^2 + \int_0^t \|P\Delta u\|^2\,\mathrm{d}\tau \le B(M,t), \quad \text{for } 0 \le t < T \qquad [32]$$

containing an integral of $\|P\Delta u\|^2$ on the left-hand side. We will use the notation $B(M,t)$ generically, for any bound that depends only on the function $M(t)$ and $t$. We remark, that a term $\|u_t\|^2$ can also be included under the integral sign on the left-hand side of [32], because $\|u_t\|$ and $\|P\Delta u\|$ are of essentially the same order, being the leading terms in the projection $u_t + P(u\cdot\nabla u) = P\Delta u$ of the Navier–Stokes equation. Finally, one can also include $\|u\|_{W_2^2(\Omega)}$ under the integral sign, in view of [19].

Going on, we obtain a third differential inequality from the second identity of the sequence [30]. Its right-hand side admits the estimate

$$\begin{aligned}
-(u_t\cdot\nabla u, u_t) &\le \|u_t\|_4^2\|\nabla u\|\\
&\le c\|u_t\|^{1/2}\|\nabla u_t\|^{3/2}\|\nabla u\|\\
&\le \frac{1}{2}\|\nabla u_t\|^2 + c\|\nabla u\|^4\|u_t\|^2
\end{aligned} \qquad [33]$$

which, in view of [29] or [32], produces a linear differential inequality with integrable coefficients. Its integration yields an estimate of the form

$$\|u_t(t)\|^2 + \int_0^t \|\nabla u_t\|^2\,\mathrm{d}\tau \le B(M, t, \|u_t(0)\|), \quad \text{for } 0 \le t < T \qquad [34]$$

provided $\|u_t(0)\|$ is bounded. Since $u_t = P(\Delta u - u\cdot\nabla u)$, we have the estimate

$$\|u_t(0)\| = \|P(\Delta u_0 - u_0\cdot\nabla u_0)\| \le \|\Delta u_0 - u_0\cdot\nabla u_0\| \le B\Big(\|u_0\|_{W_2^2(\Omega)}\Big) \qquad [35]$$

provided that $u$ is smooth in $\overline{\Omega}\times[0,T)$. This is a delicate point, having been forewarned of a regularity breakdown at $t=0$. But, we will be able to replicate the estimate [35] for the Galerkin approximations, ultimately validating [34] for the approximations and the solution.

The integration of the next differential inequality, which arises from the second of the identities [31], requires that $\|\nabla u_t(0)\| < \infty$. Similarly to [35], we have

$$\|\nabla u_t(0)\| = \|\nabla P(\Delta u_0 - u_0\cdot\nabla u_0)\| \le B\Big(\|u_0\|_{W_2^3(\Omega)}\Big) \qquad [36]$$

provided that $u$ is smooth in $\overline{\Omega}\times[0,T)$. However, there is a big difference between [35] and [36]. In the next section, we will not be able to obtain an analog of [36] for the Galerkin approximations. Consequently, the solution that is obtained will not be fully regular at time $t=0$. It will satisfy $u \in C(\overline{\Omega}\times[0,T)) \cap C^\infty(\overline{\Omega}\times(0,T))$, but not $u \in C^\infty(\overline{\Omega}\times[0,T))$. It will satisfy

$\|u(t) - u_0\|_{W_2^2(\Omega)} \to 0$ but not $\|u(t) - u_0\|_{W_2^3(\Omega)} \to 0$, as $t \to 0^+$.

One may wonder whether this is a fault or deficiency in the Galerkin method. It is not, remembering what was said at the beginning of this section. For most prescribed values of $u_0$, no matter how smooth, there is a breakdown in the regularity of the solution as $t \to 0^+$. In fact, it was proved in Heywood and Rannacher (1982) that if $\|\nabla u_t(t)\|$ or any one of several other quantities, including $\|u(t)\|_{W_2^3(\Omega)}$, remains bounded as $t \to 0^+$, then there exists a solution $p_0$ of the overdetermined Neumann problem

$$\begin{aligned} -\Delta p_0 &= \nabla \cdot (u_0 \cdot \nabla u_0) \quad \text{in } \Omega \\ \nabla p_0|_{\partial\Omega} &= \Delta u_0|_{\partial\Omega} \end{aligned} \qquad [37]$$

Generically speaking, this problem is not solvable, and therefore

$$\limsup_{t\to 0^+} \|\nabla u_t(t)\| = \infty$$

We mention that under our assumption that $u_0$ is smooth, the correctly posed Neumann problem, with boundary condition $\partial p_0/\partial n|_{\partial\Omega} = \Delta u_0 \cdot n|_{\partial\Omega}$, is uniquely solvable for a solution $p_0 \in W_2^1(\Omega)/R$, and $\|\nabla p(t) - \nabla p_0\| \to 0$, as $t \to 0^+$; see Heywood and Rannacher (1982).

Since solutions are smooth for $0 < t < T$, the pressure in the Navier–Stokes equations satisfies the overdetermined Neumann problem for all $t \in (0, T)$. So it may seem appropriate to require that the prescribed initial value $u_0$ be a function for which problem [37] is solvable. We do not agree with that. It is too difficult, if not impossible, to find such functions, except by solving the Navier–Stokes equations. For example, one might think that the condition that [37] should be solvable might be satisfied if $u_0 \in \boldsymbol{D}(\Omega)$, since such functions are zero in a neighborhood of the boundary. In fact, K Masuda has shown that if $\Omega$ is a three-dimensional sphere, then the overdetermined Neumann problem [37] is never solvable for nonzero $u_0 \in \boldsymbol{D}(\Omega)$. Hence, the gradient of the initial pressure will have a nonzero tangential component, causing an impulsive tangential acceleration along the boundary.

If we are to use the Navier–Stokes equations to make predictions of the future, we must solve the initial boundary value problem for "man-made" initial values, and accept the fact that there is a momentary breakdown in regularity along the boundary, immediately following the initial time. Thereafter, the solution smooths as "nature" takes over. To prove the reliability of our predictions, we need continuous dependence estimates and error estimates for numerical methods that take into account this initial breakdown in the regularity. The continuous dependence estimate [14] meets this requirement. So also do the error estimates given in a series of four papers by Rannacher and the author, beginning with Heywood and Rannacher (1982). They were based on the "smoothing" regularity estimates for solutions that are being presented here. We go on with these now, as models for similar estimates for the Galerkin approximations.

Estimating the right-hand side of the second of the identities [31] using [20] and Young's inequality, and then multiplying through by $t$, we get the linear differential inequality

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\left(t\|\nabla u_t\|^2\right) + t\|P\Delta u_t\|^2 \\ &\quad \le \|\nabla u_t\|^2 + c\left(\|\nabla u\|^4 + \|\nabla u\|^2 \right. \\ &\qquad \left. + \|P\Delta u\|^2\right)\left(t\|\nabla u_t\|^2\right) \end{aligned} \qquad [38]$$

for $t\|\nabla u_t\|^2$, with coefficients that are integrable in view of the previous estimates [32], [34], and [35]. Therefore, its integration yields an estimate analogous to [32] of the form

$$\begin{aligned} &t\|\nabla u_t(t)\|^2 + \int_0^t t\|P\Delta u_t\|^2 \,\mathrm{d}\tau \\ &\quad \le B\left(M, t, \|u_0\|_{W_2^2(\Omega)}\right), \quad \text{for } 0 < t < T \end{aligned} \qquad [39]$$

provided its "initial value" is finite. It is, due to the time weight, in the sense that

$$\limsup_{t\to 0^+} \left(t\|\nabla u_t(t)\|^2\right) = 0 \qquad [40]$$

This is proved by noting that if the lim sup were positive, then the integral on the left-hand side of [34] would be infinite. Finally, a term $t\|u_{tt}\|^2$ can be included under the integral sign on the left-hand side of [39], because $\|u_{tt}\|$ and $\|P\Delta u_t\|$ are of essentially the same order, being the leading terms in the projection $u_{tt} + P(u_t \cdot \nabla u + u \cdot \nabla u_t) = P\Delta u_t$ of the time differentiated Navier–Stokes equation.

We continue inductively. Estimating the right-hand side of the third of the identities [30] using [12], [20], and Young's inequality, and then multiplying through by $t^2$, we get the linear differential inequality

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}\left(t^2\|u_{tt}\|^2\right) + t^2\|\nabla u_{tt}\|^2 \\ &\quad \le 2t\|u_{tt}\|^2 + t^2\|\nabla u_t\|^2 + t^2\|P\Delta u_t\|^2 \\ &\qquad + c\left(\|\nabla u\|^4 + \|\nabla u_t\|^4\right)\left(t^2\|u_{tt}\|^2\right) \end{aligned} \qquad [41]$$

with coefficients that are integrable in view of preceding estimates. In particular, the integrability

of the first term on the right-hand side follows from the boundedness of the integral

$$\int_0^t t\|u_{tt}\|^2 \, d\tau \qquad [42]$$

which, we have pointed out, can be included on the left-hand side of [39]. Finally, notice that the boundedness of the integral [42] implies

$$\limsup_{t\to 0^+}\left(t^2\|u_{tt}(t)\|^2\right) = 0 \qquad [43]$$

Therefore, we can integrate [41] to get the estimate

$$t^2\|u_{tt}(t)\|^2 + \int_0^t t^2\|\nabla u_{tt}\|^2 \, d\tau \le B\left(M, t, \|u_0\|_{W_2^2(\Omega)}\right), \quad \text{for } 0 \le t < T \qquad [44]$$

analogous to [34].

At this point, we have introduced every device needed to proceed by induction to an infinite sequence of time-weighted estimates, similar to [39] and [44], but with successively higher orders of time derivatives and weights. The dependence of these estimates on $\|u_0\|_{W_2^2(\Omega)}$ was introduced through [34] and [35]. It can be eliminated by beginning the introduction of powers of $t$ as weight functions one step earlier, with the added advantage that the initial velocity $u_0$ needs only belong to $\boldsymbol{J}_1(\Omega)$. In the two-dimensional case, the weight functions can be introduced even another step earlier, with the advantage that the initial velocity $u_0$ needs only belong to $\boldsymbol{J}(\Omega)$. Each of these cases leads to an existence theorem for solutions $u \in C^\infty(\overline{\Omega} \times (0, T))$, with the initial values assumed in the norms of $\boldsymbol{J}_1(\Omega)$ and $\boldsymbol{J}(\Omega)$, respectively.

## Existence by Galerkin Approximation

Let $\{a^1, a^2, \ldots\}$ and $\{\lambda_1, \lambda_2, \ldots\}$ denote the eigenfunctions and eigenvalues of the Stokes equations,

$$-\Delta a^k + \nabla p = \lambda_k a^k, \quad \nabla \cdot a^k = 0 \quad \text{in } \Omega$$
$$a^k\big|_{\partial\Omega} = 0 \qquad [45]$$

chosen to be orthonormal in $L^2(\Omega)$. Clearly, $-P\Delta a_k = \lambda_k a^k$, so they are also the eigenfunctions and eigenvalues of the Stokes operator, $-P\Delta$. Using regularity estimates for the Stokes equations, each eigenfunction is known to be $C^\infty$ smooth in $\overline{\Omega}$.

The $n$th Galerkin approximation for problem [9] is the solution

$$u^n(x,t) = \sum_{k=1}^n c_{kn}(t) a^k(x)$$

of the system of ordinary differential equations

$$\left(u_t^n, a^l\right) + \left(u^n \cdot \nabla u^n, a^l\right) = \left(\Delta u^n, a^l\right)$$
$$\text{for } l = 1, 2, \ldots, n \qquad [46]$$

satisfying the initial conditions $(u^n(0), a^l) = (u_0, a^l)$, for $l = 1, 2, \ldots, n$. Of course, since $(u_t^n, a^l) = \partial c_{ln}/\partial t$ and $(\Delta u^n, a^l) = (P\Delta u^n, a^l) = -\lambda_l c_{ln}$, the differential equations can be written as

$$\frac{d}{dt} c_{ln} = -\sum_{i,j=1}^n c_{in} c_{jn}\left(a^i \cdot \nabla a^j, a^l\right) - \lambda_l c_{ln}$$

and the initial conditions as $c_{ln}(0) = (u^n(0), a^l)$, for $l = 1, 2, \ldots, n$.

The system [46] is at least locally solvable, on some interval $[0, T_n)$, with each coefficient satisfying $c_{ln} \in C^\infty[0, T_n)$. Therefore, since the eigenfunctions are also smooth, $u^n$ is $C^\infty$ smooth in $\overline{\Omega} \times [0, T_n)$. It also satisfies all of the identities [30] and [31] on the interval $[0, T_n)$. Indeed, multiplying [46] by $c_{ln}$ and summing over $l$ from 1 to $n$ has the effect of converting $a^l$ into $u^n$. The resulting identity for $u^n$ leads immediately to the energy identity

$$\frac{1}{2}\frac{d}{dt}\|u^n\|^2 + \|\nabla u^n\|^2 = 0 \qquad [47]$$

The remaining identities in the sequence [30] are obtained similarly. For example, the second is obtained by taking the time derivative of [46], multiplying through by $dc_{ln}/dt$ and summing over $l$.

Prodi's identity is obtained by multiplying [46] by $\lambda_l c_{ln}$ and summing, which has the effect of converting $a^l$ into $-P\Delta u^n$. To obtain the second of the identities [31] for $u^n$, one differentiates [46], multiplies by $\lambda_l dc_{ln}/dt$ and sums. The remaining identities in the sequence [31] are obtained similarly.

The initial conditions easily imply that $\|u^n(0)\| \le \|u_0\|$, because $u_0 \in \boldsymbol{J}(\Omega)$ and the eigenfunctions are orthogonal and complete in $\boldsymbol{J}(\Omega)$. Therefore, integration of [47] yields the energy estimate

$$\frac{1}{2}\|u^n(t)\|^2 + \int_0^t \|\nabla u^n\|^2 \, d\tau \le \frac{1}{2}\|u_0\|^2 \qquad [48]$$

which is uniform in $n$. Since $\|u^n(t)\|$ remains bounded, the solution $u^n(t)$ can be continued for all time. Thus, $T_n = \infty$, for all $n$. Hence, our early working assumption about solutions, that they are smooth in $\overline{\Omega} \times [0, \infty)$, is actually valid for the Galerkin approximations. The issue becomes one of obtaining estimates for their derivatives that are uniform in $n$. All of the estimates we have proved for solutions are proved in exactly the same way for the approximations. The only possible source of nonuniformity would arise from the initial values of $\|\nabla u^n\|$ and $\|u_t^n\|$.

The estimates [24], [26], and [27] are uniform in $n$, since $u_0 \in \boldsymbol{J}_1(\Omega)$ and hence $\|\nabla u^n(0)\| \le \|\nabla u_0\|$, due to the orthogonality of the eigenfunctions in the inner-product $(\nabla u, \nabla v)$, and their completeness with respect to functions in $\boldsymbol{J}_1(\Omega)$. We also obtain a uniform bound for $\|u_t^n(0)\|$ of the form [35], by multiplying [46] by $\partial c_{ln}/\partial t$ and summing over $l$. In the last step, we also need the inequality $\|u^n(0)\|_{W_2^2(\Omega)} \le \|u_0^n\|_{W_2^2(\Omega)}$, which follows from the orthogonality of the eigenfunctions in the inner product $(P\Delta u, P\Delta v)$, and their completeness with respect to functions in $\boldsymbol{J}_1(\Omega) \cap W_2^2(\Omega)$; see Ladyzhenskaya (1969, p. 46). Any attempt to find a bound for $\|\nabla u_t^n(0)\|$ analogous to [36] is certain to fail, as it would lead to a contradiction with aforementioned results from Heywood and Rannacher (1982).

### Passage to the Limit

We now have $L^2$-bounds for $u^n, \nabla u^n, u_t^n, \partial^2 u^n/\partial x_i \partial x_j$, and $\nabla u_t^n$ over any space-time region $\Omega \times (0, T')$, with $0 < T' < T$. We also have $L^2$-bounds for all orders of the time derivatives of these quantities over any subregion $\Omega \times (\varepsilon, T')$, with $0 < \varepsilon < T' < T$. From these $L^2$-bounds, we may infer the existence of a subsequence of the Galerkin approximations, again denoted by $\{u^n\}$, which converges, along with those of its derivatives for which we have bounds, to a limit $u$ and its derivatives. The convergence $u^n \to u$ and $\nabla u^n \to \nabla u$ is strong in $\boldsymbol{L}^2(\Omega \times (0, T'))$; the convergence of $u_t^n$ is strong in $\boldsymbol{L}^2(\Omega \times (\varepsilon, T'))$ and weak in $\boldsymbol{L}^2(\Omega \times (0, T'))$; the convergence of $P\Delta u^n$ is weak in $\boldsymbol{L}^2(\Omega \times (0, T'))$; all time derivatives of $u^n, \nabla u^n$ converge strongly in $\boldsymbol{L}^2(\Omega \times (\varepsilon, T'))$.

Because of estimates for the time derivatives, trace arguments give the strong convergence $u^n \to u$, $\nabla u^n \to \nabla u, u_t^n \to u_t$, and the weak convergence $P\Delta u^n \to P\Delta u$, in $\boldsymbol{L}^2(\Omega)$, for every $t > 0$.

For any fixed time, $u \in W_2^2(\Omega)$, and therefore $u$ is continuous in $\overline{\Omega}$ by a well known Sobolev inequality. Since $u \in \boldsymbol{J}_1(\Omega)$, it must equal zero along the boundary. The estimates for the time derivatives of $u^n, \nabla u^n, \partial^2 u^n/\partial x_i \partial x_j$ imply that $u$ and its time derivatives are time continuous in $W_2^2(\Omega)$. Therefore, $u, u_t, u_{tt}, \ldots$ are classically continuous in $\overline{\Omega} \times (0, T)$.

### Introduction of the Pressure

Because of the strong convergence $u^n \to u$, $\nabla u^n \to \nabla u, u_t^n \to u_t$ and the weak convergence $P\Delta u^n \to P\Delta u$, in $\boldsymbol{L}^2(\Omega)$, for any $t > 0$, it is an easy matter to let $n \to \infty$ in [46], obtaining, for all $t > 0$,

$$\left(u_t, a^l\right) + (u \cdot \nabla u, a^l) = \left(\Delta u, a^l\right) \quad \text{for } l = 1, 2, \ldots \qquad [49]$$

Since the eigenfunctions are complete in $\boldsymbol{J}(\Omega)$, and $\boldsymbol{D}(\Omega) \subset \boldsymbol{J}(\Omega)$, this implies

$$(u_t + u \cdot \nabla u - \Delta u, \phi) = 0, \quad \text{for all } \phi \in \boldsymbol{D}(\Omega) \qquad [50]$$

Therefore, there exists a vector field $\nabla p \in \boldsymbol{G}(\Omega)$ such that

$$u_t + u \cdot \nabla u - \Delta u = -\nabla p \qquad [51]$$

Indeed, the usual test to determine whether a smooth vector field $w$ is conservative in some domain $\Omega$, and therefore representable as a gradient, is to check whether the curve integrals

$$\oint_C w \cdot \tau \, \mathrm{d}s \qquad [52]$$

vanish for every smooth closed curve $C \subset \Omega$. Here, $\tau$ is the unit tangent to the curve and $\mathrm{d}s$ is its arc length. With a little reflection, one will realize that these curve integrals can be approximated by volume integrals of the form $(w, \phi)$ with $\phi \in \boldsymbol{D}(\Omega)$. For this, one should choose $\phi$ to have its support in a small tubular neighborhood of the curve, and its streamlines parallel to the curve, with unit net flux through any section of the tube. If $w$ is not smooth, but only known to belong to $\boldsymbol{L}^2(\Omega)$, one can approximate it with its smooth mollifications. This argument can be made rigorous. We previously showed that $\boldsymbol{J}(\Omega)$ and $\boldsymbol{G}(\Omega)$ are orthogonal subspaces of $\boldsymbol{L}^2(\Omega)$. Now we have argued that $\boldsymbol{L}^2(\Omega) = \boldsymbol{J}(\Omega) \oplus \boldsymbol{G}(\Omega)$.

### Classical $C^\infty$ Regularity

At any fixed time, we may regard $u$ as a solution of the steady Stokes problem [18] with $f = -u_t - u \cdot \nabla u$. Included in Cattabriga (1961) and Solonnikov (1964, 1966) are regularity estimates for all orders of derivatives of the form

$$\|u\|_{W_2^{k+2}(\Omega)} \le c\|f\|_{W_2^k(\Omega)}$$

From our estimates above, we easily conclude that $f \equiv -u_t - u \cdot \nabla u \in W_2^1(\Omega)$. Hence, $u \in W_2^3(\Omega)$. In fact, in view of the regularity we have proven with respect to time, $f \in C^\infty(0, T; W_2^1(\Omega))$ and $u \in C^\infty(0, T; W_2^3(\Omega))$. Thus begins a bootstrapping argument. In the next step, we observe that $f \in C^\infty(0, T; W_2^2(\Omega))$ and conclude that $u \in C^\infty(0, T; W_2^4(\Omega))$. By induction, one obtains $u \in C^\infty(0, T; W_2^k(\Omega))$ for every positive integer $k$. Then well-known Sobolev inequalities imply that $u \in C^\infty(\overline{\Omega} \times (0, T))$.

### Assumption of the Initial Values

We begin by showing that $u(t) \to u_0$, weakly in $L^2(\Omega)$, as $t \to 0^+$. Of course, $\|u(t)\|$ remains bounded as $t \to 0^+$, in virtue of [48], and the eigenfunctions $\{a^l\}$ are complete in $J(\Omega)$. Writing

$$\left(u(t) - u_0, a^l\right) = \left(u(t) - u^n(t), a^l\right) + \left(u^n(t) - u^n(0), a^l\right) + \left(u^n(0) - u_0, a^l\right)$$

note that the first and third terms on the right-hand side can be made small by choosing $n$ large. The second can be written as

$$\left(u^n(t) - u^n(0), a^l\right) = \int_0^t \left(u^n_t, a^l\right) \mathrm{d}\tau$$

which will be small if $t$ is small, in view of [34]. Thus, $(u(t) - u_0, a^l) \to 0$, as $t \to 0^+$, which implies the desired weak convergence.

The strong convergence $u(t) \to u_0$ in $L^2(\Omega)$ follows from the weak convergence if $\limsup_{t\to 0^+} \|u(t)\| \leq \|u_0\|$. The energy estimate [48] for the approximations implies this also.

To conclude that $u(t) \to u_0$ strongly in $J_1(\Omega)$, it only remains to be shown that $\limsup_{t\to 0^+} \|\nabla u(t)\| \leq \|\nabla u_0\|$. This readily follows from [29], provided the bounding function $M(t)$ satisfies $M(t) \to \|\nabla u_0\|$, as $t \to 0^+$. The bounding functions provided by our basic estimates [24], [26], and [27] all have this property.

We may conclude that $u(t) \to u_0$ weakly in $W_2^2(\Omega)$, provided $\|u(t)\|_{W_2^2(\Omega)}$ remains bounded as $t \to 0^+$. To see this, remember that $\|P\Delta u\|$ and $\|u_t\|$ are of essentially the same order. Thus the term $\|u_t(t)\|^2$ on the left-hand side of [34] can be accompanied by a term $\|u(t)\|^2_{W_2^2(\Omega)}$.

Finally, to prove that $u(t) \to u_0$ strongly in $W_2^2(\Omega)$, we need only show that $\limsup_{t\to 0^+} \|P\Delta u(t)\| \leq \|P\Delta u_0\|$, since $\|P\Delta\cdot\|$ and $\|\cdot\|_{W_2^2(\Omega)}$ are equivalent norms on $J_1(\Omega) \cap W_2^2(\Omega)$. To this end, multiply [46] by $\lambda_l \mathrm{d}c_{ln}/\mathrm{d}t$ and sum to get

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|P\Delta u^n\|^2 + \left\|\nabla u^n_t\right\|^2 = \left(u^n \cdot \nabla u^n, P\Delta u^n_t\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(u^n \cdot \nabla u^n, P\Delta u^n\right) - \left(u^n_t \cdot \nabla u^n + u^n \cdot \nabla u^n_t, P\Delta u^n\right)$$

Integrating this gives

$$\begin{aligned}\|P\Delta u^n(t)\|^2 - \|P\Delta u^n(0)\|^2 &= \int_0^t \frac{\mathrm{d}}{\mathrm{d}t}\|P\Delta u^n\|^2 \mathrm{d}s \\ &= 2(u^n \cdot \nabla u^n, P\Delta u^n)|_t - 2(u^n \cdot \nabla u^n, P\Delta u^n)|_0 \\ &\quad - 2\int_0^t \left\|\nabla u^n_t\right\|^2 \mathrm{d}s - 2\int_0^t \left(u^n_t \cdot \nabla u^n + u^n \cdot \nabla u^n_t, P\Delta u^n\right) \mathrm{d}s\end{aligned} \quad [53]$$

For the terms under the last integral we have

$$\left|\left(u^n_t \cdot \nabla u^n, P\Delta u^n\right) + \left(u^n \cdot \nabla u^n_t, P\Delta u^n\right)\right| \leq \left\|\nabla u^n_t\right\|^2 + c\|\nabla u^n\|^{1/2}\|P\Delta u^n\|^{3/2}$$

Therefore, [53] implies

$$\|P\Delta u^n(t)\|^2 \leq \|P\Delta u^n(0)\|^2 + 2(u^n \cdot \nabla u^n, P\Delta u^n)|_t - 2(u^n \cdot \nabla u^n, P\Delta u^n)|_0 + Kt$$

uniformly in $n$, as $t \to 0^+$, where $K$ is a constant depending on the estimates [32] and [34]. Letting $n \to \infty$, gives

$$\|P\Delta u(t)\|^2 \leq \|P\Delta u(0)\|^2 + 2(u \cdot \nabla u, P\Delta u)|_t - 2(u \cdot \nabla u, P\Delta u)|_0 + Kt$$

Since $u \cdot \nabla u \to u_0 \cdot \nabla u_0$ strongly in $L^2(\Omega)$, and $P\Delta u \to P\Delta u_0$ weakly in $L^2(\Omega)$, we get the desired result. The continuous assumption of the initial values in $W_2^2(\Omega)$ also implies their continuous assumption in the classical sense, and hence that $u \in C(\overline{\Omega} \times [0, T))$.

### Conclusion

Years ago, mathematical questions concerning the Navier–Stokes equations were usually considered in the context of generalized or weak solutions, which was a technical barrier to many in the scientific community. Nowadays, realizing that solutions are at least locally classical, fundamental questions such as that of global regularity can be studied within the classical context. If the estimate [29] is proved for classical solutions, with $T = \infty$, and without a restriction on the size of the data, this particular matter will be settled.

This work has been supported by the Natural Sciences and Engineering Research Council of Canada.

*See also:* Compressible Flows: Mathematical Theory; Elliptic Differential Equations: Linear Theory; Incompressible Euler Equations: Mathematical Theory; Interfaces and Multicomponent Fluids; Leray–Schauder Theory and Mapping Degree; Non-Newtonian Fluids; Partial Differential Equations: Some Examples; Stochastic Hydrodynamics; Turbulence Theories; Wavelets: Application to Turbulence.

## Further Reading

Beirão da Veiga H (1997) A new approach to the $L^2$-regularity theorem for linear stationary nonhomogeneous Stokes systems. *Portugaliae Mathematica* 54(Fasc. 3): 271–286.

Cattabriga L (1961) Su un probleme al contorno relativo al sistema di equazioni di Stokes. *Rendi conti del Seminario Matematico della Università di Padova* 31: 308–340.

Heywood JG (1976) On uniqueness questions in the theory of viscous flow. *Acta Mathematica* 136: 61–102.
Heywood JG (1980) The Navier–Stokes equations: on the existence, regularity and decay of solutions. *Indiana University Mathematics Journal* 29: 639–681.
Heywood JG (1990) Open problems in the theory of the Navier–Stokes equations for viscous incompressible flow. In: Heywood JG, Masuda K, Rautmann R, and Solonnikov VA (eds.) *The Navier–Stokes Equations: Theory and Numerical Methods*, Lecture Notes in Mathematics, vol. 1431, pp. 1–22. Berlin–Heidelberg: Springer Verlag.
Heywood JG (1994) Remarks concerning the possible global regularity of solutions of the three-dimensional incompressible Navier–Stokes equations. In: Galdi GP, Malek J, and Necas J (eds.) *Progress in Theoretical and Computational Fluid Dynamics*, Pitman Research Notes in Mathematics Series, vol. 308, pp. 1–32. Essex: Longman Scientific and Technical.
Heywood JG (2001) On a conjecture concerning the Stokes problem in nonsmooth domains. In: Neustupa J and Penel P (eds.) *Mathematical Fluid Mechanics: Recent Results and Open Problems*, Advances in Mathematical Fluid Mechanics vol. 2, pp. 195–206. Basel: Birkhauser Verlag.
Heywood JG (2003) A curious phenomenon in a model problem, suggestive of the hydrodynamic inertial range and the smallest scale of motion. *Journal of Mathematical Fluid Mechanics* 5: 403–423.
Heywood JG and Rannacher R (1982) Finite element approximation of the nonstationary Navier–Stokes problem, Part I: Regularity of solutions and second-order error estimates for the spatial discretization. *SIAM Journal on Numerical Analysis* 19: 275–311.
Heywood JG and Xie W (1997) Smooth solutions of the vector Burgers equation in nonsmooth domains. *Differential and Integral Equations* 10: 961–974.
Ladyzhenskaya OA (1969) *The Mathematical Theory of Viscous Incompressible Flow*, 2nd edn. New York: Gordon and Breach.
Prodi G (1962) Teoremi di tipo lacale per il sistema de Navier–Stokes e stabilità delle soluzione stazionarie. *Rendi conti del Seminario Matematico della Università di Padova* 32: 374–397.
Solonnikov VA (1964) On general boundary-value problems for elliptic systems in the sense of Douglis–Nirenberg. I. *Izvestiya Akademii Nauk SSSR, Sariya Matematicheskaya* 28: 665–706.
Solonnikov VA (1966) On general boundary-value problems for elliptic systems in the sense of Douglis–Nirenberg. II. *Trudy Matematiceskogo Instituta Imeni V.A. Steklova* 92: 233–297.
Solonnikov VA and Sčadilov VE (1973) On a boundary value problem for a stationary system of Navier–Stokes equations. *Proceedings. Steklov Institute of Matematics* 125: 186–199.
Xie W (1991) A sharp pointwise bound for functions with $L^2$ Laplacians and zero boundary values on arbitrary three-dimensional domains. *Indiana University Mathematics Journal* 40: 1185–1192.
Xie W (1992) On a three-norm inequality for the Stokes operator in nonsmooth domains. In: Heywood JG, Masuda K, Rautmann R, and Solonnikov VA (eds.) *The Navier–Stokes Equations II: Theory and Numerical Methods*, Lecture Notes in Mathematics, vol. 1530, pp. 310–315. Berlin–Heidelberg: Springer Verlag.
Xie W (1997) Sharp Sobolev interpolation inequalities for the Stokes operator. *Differential and Integral Equations* 10: 393–399.

# von Neumann Algebras: Introduction, Modular Theory, and Classification Theory

**V S Sunder**, The Institute of Mathematical Sciences, Chennai, India



## Introduction

von Neumann algebras, as they are called now, first made their appearance under the name "rings of operators" in a series of seminal papers – see Murray and von Neumann (1936, 1937, 1943) and von Neumann (1936) – by F J Murray and J von Neumann starting in 1936. Murray and von Neumann (1936) specifically cite "attempts to generalize the theory of unitary group representations" and "demands by various aspects of the quantum-mechanical formalism" among the reasons for the elucidation of this subject.

In fact, the simplest definition of a von Neumann algebra is via unitary group representations: a collection $M$ of continuous linear operators on a Hilbert space $\mathcal{H}$ (in order to avoid some potential technical problems, we shall restrict ourselves to separable Hilbert spaces throughout this article) is a von Neumann algebra precisely when there is a representation $\rho$ of a group $G$ as unitary operators on $\mathcal{H}$ such that

$$M = \{x \in \mathcal{L}(\mathcal{H}) : x\rho(t) = \rho(t)x \ \forall t \in G\}$$

As above, we shall write $\mathcal{L}(\mathcal{H})$ for the collection of all continuous linear operators on the Hilbert space $\mathcal{H}$; recall that a linear mapping $x : \mathcal{H} \to \mathcal{H}$ is continuous precisely when there exists a positive constant $K$ such that $\|x\xi\| \leq K\|\xi\| \ \forall \xi \in \mathcal{H}$. If the norm $\|x\|$ of the operator $x$ is defined as the smallest constant $K$ with the above property, then the set $\mathcal{L}(\mathcal{H})$ acquires the structure of a Banach space. In fact $\mathcal{L}(\mathcal{H})$ is a Banach $^*$-algebra with respect to the composition product, and involution $x \mapsto x^*$ given by

$$\langle x\xi, \eta\rangle = \langle \xi, x^*\eta\rangle \ \forall \xi, \eta \in \mathcal{H}$$

The first major result in the subject is the remarkable "double commutant theorem," which establishes the equivalence of a purely algebraic requirement to purely topological ones. We need

two bits of terminology to be able to state the theorem.

First, define the commutant $S'$ of a subset $S \subset \mathcal{L}(\mathcal{H})$ by

$$S' = \{x' \in \mathcal{L}(\mathcal{H}) : x'x = xx' \ \forall x \in S\}$$

Second, the strong (resp., weak) operator topology is the topology on $\mathcal{L}(\mathcal{H})$ of "pointwise strong (resp., weak) convergence": that is, $x_n \to x$ precisely when $\|x_n\xi - x\xi\| \to 0 \ \forall \xi \in \mathcal{H}$ (resp., $\langle x_n\xi - x\xi, \eta\rangle \to 0 \ \forall \xi, \eta \in \mathcal{H}$).

**Theorem 1** *The following conditions on a unital* $^*$*-subalgebra M of* $\mathcal{L}(\mathcal{H})$ *are equivalent*:

(i) $M = M''(= (M')')$.
(ii) *M is closed in the strong operator topology.*
(iii) *M is closed in the weak operator topology.*

The conventional definition of a von Neumann algebra is that it is a unital $^*$-subalgebra of $\mathcal{L}(\mathcal{H})$ which satisfies the equivalent conditions above. The equivalence with our earlier "simplest definition" is a consequence of the double commutant theorem and the fact that any element of a von Neumann algebra is a linear combination of four unitary elements of the algebra: simply take $G$ to be the group of unitary operators in $M'$.

Another consequence of the double commutant theorem is that von Neumann algebras are closed under any "canonical construction." For instance, the uniqueness of the spectral measure $E \mapsto P_x(E)$ associated to a normal operator $x$ shows that if $u$ is unitary, then $P_{uxu^*}(E) = uP_x(E)u^*$ for all Borel sets $E$. In particular, if $x \in M$ and $u' \in \mathcal{U}(M')$, then $u'P_x(E)u'^* = P_{u'xu'^*}(E) = P_x(E)$, and hence, we may conclude that $P_x(E) \in \mathcal{U}(M')' = (M')' = M$ (we will write $\mathcal{U}(N)$ (resp., $\mathcal{P}(N)$) to denote the collection of unitary (resp., projection) operators in any von Neumann algebra $N$); that is, if a von Neumann algebra contains a normal operator, it also contains all the associated spectral projections. This fact, together with the spectral theorem, has the consequence that any von Neumann algebra $M$ is the closed linear span of $\mathcal{P}(M)$.

The analogy with unitary group representations is fruitful. Suppose then that $M = \rho(G)'$, for a unitary representation of $G$. Then the last sentence of the previous paragraph implies that $\rho(G)' = \mathbb{C}$ precisely when there exist no nontrivial $\rho$-stable subspaces (here and in the sequel, we identify $\mathbb{C}$ with its image under the unique unital homomorphism of $\mathbb{C}$ into $\mathcal{L}(\mathcal{H})$; and we reserve the symbol $Z(M)$ to denote the center of $M$), that is, when $\rho$ is irreducible. In general, the $\rho$-stable subspaces are precisely the ranges of projection operators in $M$. The notion of unitary equivalence of subrepresentations of $\rho$ is seen to translate to the equivalence defined on the set $\mathcal{P}(M)$ of projections in $M$, whereby $p \sim q$ if and only if there exists an operator $u \in M$ such that $u^*u = p$ and $uu^* = q$. (Such a $u$ is called a partial isometry, with "initial space" = range $p$, and "final space" = range $q$.) This is the definition of what is known as the "Murray–von Neumann equivalence rel $M$" and is denoted by $\sim_M$ . The following accompanying definition is natural: if $p, q \in \mathcal{P}(M)$, say $p \preceq_M q$ if there exists $p_0 \in \mathcal{P}(M)$ such that $p \sim_M p_0 \leq q$ – where of course $e \leq f \Leftrightarrow \text{range}(e) \subset \text{range}(f)$.

## The Murray–von Neumann Classification of Factors

We start with a fact (whose proof is quite easy) and a consequent fundamental definition.

**Proposition 2** *The following conditions on a von Neumann algebra M are equivalent*:

(i) *for any* $p, q \in \mathcal{P}(M)$, *it is true that either* $p \preceq_M q$ *or* $q \preceq_M p$.
(ii) $Z(M) = M \cap M' = \mathbb{C}$.

*The von Neumann algebra M is called a* "factor" *if it satisfies the equivalent conditions above.*

The alert reader would have noticed that if $G$ is a finite group, then $\rho(G)'$ is a factor precisely when the representation $\rho$ is "isotypical." Thus, the "representation-theoretic fact," that any unitary representation is expressible as a direct sum of isotypical subrepresentations, translates into the "von Neumann algebraic fact" that any $*$-subalgebra of $\mathcal{L}(\mathcal{H})$ is isomorphic, when $\mathcal{H}$ is finite dimensional, to a direct sum of factors. In complete generality, von Neumann (1949) showed that any von Neumann algebra is expressible as a "direct integral of factors." We shall interpret this fact from "reduction theory" as the statement that all the magic/mystery of von Neumann algebras is contained in factors and hence restrict ourselves, for a while, to the consideration of factors.

Murray and von Neumann initiated the study of a general factor $M$ via a qualitative as well as a quantitative analysis of the relation $\preceq_M$ on $\mathcal{P}(M)$. First, call a $p \in \mathcal{P}(M)$ infinite if there exists a $p_0 \leq p$ such that $p \sim_M p_0$ and $p_0 \neq p$; otherwise, say $p$ is finite. They obtained an analog, called the "dimension function," of the Haar measure, as follows.

**Theorem 3**

(i) *With M as above, there exists a function* $D_M : \mathcal{P}(M) \to [0, \infty]$ *which satisfies the following*

*properties, and is determined up to a multiplicative constant, by them*:

- $p \preceq_M q \Leftrightarrow D_M(p) \leq D_M(q)$
- *p is finite if and only if* $D_M(p) < \infty$
- *If* $\{p_n : n = 1, 2, \ldots\}$ *is any sequence of pairwise orthogonal projections in* $\mathcal{P}(M)$ *and* $p = \sum_n p_n$, *then* $D_M(p) = \sum_n D_M(p_n)$

(ii) *M falls into exactly one of five possible cases, depending on which of the following sets is the range of some scaling of* $D_M$:

- $(I_n)$ $\{0, 1, 2, \ldots, n\}$
- $(I_\infty)$ $\{0, 1, 2, \ldots, \infty\}$
- $(II_1)$ $[0, 1]$
- $(II_\infty)$ $[0, \infty]$
- $(III)$ $\{0, \infty\}$

In words, we may say that a factor $M$ is of:

1. type $I$ (i.e., of type $I_n$ for some $1 \leq n \leq \infty$) precisely when $M$ contains a minimal projection,
2. type $II$ (i.e., of type $II_1$ or $II_\infty$) precisely when $M$ contains nonzero finite projections but no minimal projections, and
3. type $III$ precisely when $M$ contains no nonzero finite projections.

**Examples** $L^\infty(\Omega, \mu)$ may be regarded as a von Neumann algebra acting on $L^2(\Omega, \mu)$ as multiplication operators; thus, if we set $m_f(\xi) = f\xi$, then $m : f \mapsto m_f$ defines an isomorphism of $L^\infty(\Omega, \mu)$ onto a commutative von Neumann subalgebra of $\mathcal{L}(L^2(\Omega, \mu))$. In fact, "up to multiplicity," this is how any commutative von Neumann algebra looks.

It is a simple exercise to prove that $M \subset \mathcal{L}(\mathcal{H})$ is a factor of type $I_n, 1 \leq n \leq \infty$, if and only if there exist Hilbert spaces $\mathcal{H}_n$ and $\mathcal{K}$ and identifications $\mathcal{H} = \mathcal{H}_n \otimes \mathcal{K}, M = \{x \otimes \mathrm{id}_\mathcal{K} : x \in \mathcal{L}(\mathcal{H}_n)\}$ where $\dim \mathcal{H}_n = n$; and so $M \cong \mathcal{L}(\mathcal{H}_n)$.

To discuss examples of the other types, it will be convenient to use "crossed products" of von Neumann algebras by ergodically acting groups of automorphisms. We shall now digress with a discussion of this generalization of the notion of a semidirect product of groups.

If $\alpha : G \to \mathrm{Aut}(M)$ is an action of a countable group $G$ on $M$, where $M \subset \mathcal{L}(\mathcal{H})$ is a von Neumann algebra, and $\widetilde{\mathcal{H}} = \ell^2(G, \mathcal{H})$, there are representations $\pi : M \to \mathcal{L}(\widetilde{\mathcal{H}})$ and $\lambda : G \to \mathcal{U}(\mathcal{L}(\widetilde{\mathcal{H}}))$ defined by

$$(\pi(x)\xi)(s) = \alpha_{s^{-1}}(x)\xi(s), \quad (\lambda(t)\xi)(s) = \xi(t^{-1}s)$$

These representations satisfy the commutation relation $\lambda(t)\pi(x)\lambda(t^{-1}) = \pi(\alpha_t(x))$, and the crossed product $M \rtimes_\alpha G$ is the von Neumann subalgebra of $\mathcal{L}(\widetilde{\mathcal{H}})$ defined by $\widetilde{M} = (\pi(M) \cup \lambda(G))''$.

Let us restrict ourselves to the case of $M = L^\infty(\Omega, \mu)$ acting on $L^2(\Omega, \mu)$. In this case, it is true that any automorphism of $M$ is of the form $f \mapsto f \circ T^{-1}$, where $T$ is a "nonsingular transformation of the measure space $(\Omega, \mu)$" (= a bijection which preserves the class of sets of $\mu$-measure 0). So, an action of $G$ on $M$ is of the form $\alpha_t(f) = f \circ T_t^{-1}$, for some homomorphism $t \mapsto T_t$ from $G$ to the group of nonsingular transformations of $(\Omega, \mu)$. We have the following elegantly complete result from Murray and von Neumann (1936).

**Theorem 4** *Let* $M, G, \alpha$ *be as in the last section, and let* $\widetilde{M} = M \rtimes_\alpha G$. *Assume the G-action is "free," meaning that if* $t \neq 1 \in G$, *then* $\mu(\{\omega \in \Omega : T_t(\omega) = \omega\}) = 0$. *Then*:

(i) $\widetilde{M}$ *is a factor if and only if G acts ergodically on* $(\Omega, \mu)$ *– meaning that the only G-fixed functions in M are the constants.*

(ii) *Assume that G acts ergodically. Then the type of the factor* $\widetilde{M}$ *is determined as follows*:

- $\widetilde{M}$ *is of type I or II if and only if there exists a G-invariant measure* $\nu$ *which is mutually absolutely continuous with respect to* $\mu$, *meaning* $\nu(E) = 0 \Leftrightarrow \mu(E) = 0$; (*the ergodicity assumption implies that such a* $\nu$ *is necessarily unique up to scaling by a positive constant*;)
- $\widetilde{M}$ *is of type* $I_n$ *precisely when the* $\nu$ *as above is totally atomic, and* $\Omega$ *is the disjoint union of n atoms for* $\nu$;
- $\widetilde{M}$ *is of type II precisely when the* $\nu$ *as above is nonatomic*;
- $\widetilde{M}$ *is of finite type – meaning that* 1 *is a finite projection in* $\widetilde{M}$ *– precisely when the* $\nu$ *as above is a finite measure*;
- $\widetilde{M}$ *is of type III if and only if there exists no* $\nu$ *as above.*

Thus, we get all the types of factors by this construction; for instance, we may take:

$(I_n)$ $G = \mathbb{Z}_n$ acting on $\Omega = \mathbb{Z}_n$ by translation, and $\mu = \nu =$ counting measure
$(I_\infty)$ $G = \mathbb{Z}$ acting on $\Omega = \mathbb{Z}$ by translation, and $\mu = \nu =$ counting measure
$(II_1)$ $G = \mathbb{Z}$ acting on $\Omega = \mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$ by powers of an aperiodic rotation, and $\mu = \nu =$ arclength measure
$(II_\infty)$ $G = \mathbb{Q}$ acting on $\Omega = \mathbb{R}$ by translations, and $\mu = \nu =$ Lebesgue measure
$(III)$ $G = ax + b$ group acting in the obvious manner on $\Omega = \mathbb{R}, \mu = \nu =$ Lebesgue measure.

Such crossed products of a commutative von Neumann algebra by an ergodically acting countable

group were intensively studied by Krieger (1970, 1976). We shall simply refer to such factors as "Krieger factors." The term "Krieger factor" is actually used for factors obtained from a slightly more general construction, with ergodic group actions replaced by more general ergodic equivalence relations. Since there is no difference in the two notions at least in good (amenable) cases, we will say no more about this.

## Abstract von Neumann Algebras

So far, we have described matters as they were in von Neumann's time. To come to the modern era, it is desirable to "free a von Neumann algebra from the ambient Hilbert space" and to regard it as an abstract object in its own right which can act on different Hilbert spaces – for example, $L^\infty(\Omega,\mu)$ is an object worthy of study in its own right, without reference to $L^2(\Omega,\mu)$.

The abstract viewpoint is furnished by a theorem of Sakai (1983); let us define an abstract von Neumann algebra to be an abstract $C^*$-algebra (this is a Banach algebra with an involution related to the norm by the so-called $C^*$-identity $\|x\|^2 = \|x^*x\|$) $M$ which admits a pre-dual $M_*$ – i.e., $M$ is isometrically isomorphic to the Banach dual space $(M_*)^*$. It turns out that a predual of such an abstract von Neumann algebra is unique up to isometric isomorphism. Consequently, an abstract von Neumann algebra comes equipped with a canonical "weak$^*$-topology," usually called the "$\sigma$-weak topology" on $M$. The natural morphisms in the category of abstract von Neumann algebras are $*$-homomorphisms which are continuous with respect to $\sigma$-weak topologies on domain and range. It is customary to call a linear map between abstract von Neumann algebras "normal" if it is continuous with respect to $\sigma$-weak topologies on domain and range.

The equivalence of the "abstract" definition of this section, with the "concrete" one given earlier (which depends on an ambient Hilbert space), relies on the following four facts:

1. $\mathcal{L}(\mathcal{H})$ is an abstract von Neumann algebra, with the predual $\mathcal{L}(\mathcal{H})_*$ being the so-called "trace class" of operators, equipped with the "trace norm."
2. A self-adjoint subalgebra of $\mathcal{L}(\mathcal{H})$ is closed in the strong operator topology, and is hence a "concrete von Neumann algebra" precisely when it is closed in the $\sigma$-weak topology on $\mathcal{L}(\mathcal{H})$.
3. If $M$ is an abstract von Neumann algebra, and $N$ is a $^*$-subalgebra of $M$ which is closed in the $\sigma$-weak topology of $M$, then $N$ is also an abstract von Neumann algebra, with one candidate for $N_*$ being $M_*/N_\perp$ (where $N_\perp = \{\rho \in M_* : n(\rho) = 0\ \forall n \in N\}$).
4. Any abstract von Neumann algebra (with separable predual) is isomorphic (in the category of abstract von Neumann algebras) to a (concrete) von Neumann subalgebra of $\mathcal{L}(\mathcal{H})$ (for a separable $\mathcal{H}$).

With the abstract viewpoint available, we shall look for modules over a von Neumann algebra $M$, meaning pairs $(\mathcal{H},\pi)$ where $\pi : M \to \mathcal{L}(\mathcal{H})$ is a normal $^*$-homomorphism.

A brief digression into the proof of fact (4) above – which asserts the existence of faithful $M$-modules – will be instructive and useful. Suppose $M$ is an abstract von Neumann algebra. A linear functional $\phi$ on $M$ is called a normal state if:

- (positivity) $\phi(x^*x) \geq 0 \forall x \in M$;
- (normality) $\phi : M \to \mathbb{C}$ is normal; and
- (normalization) $\phi(1) = 1$.

(Normal states on $L^\infty(\Omega,\mu)$ correspond to non-negative probability measures on $\Omega$ which are absolutely continuous with respect to $\mu$.) It is true that there exist plenty of normal states on $M$. In fact, they linearly span $M_*$. This implies that if $M_*$ is separable, then there exist normal states on $M$ which are even "faithful" – meaning $\phi(x^*x) = 0 \Leftrightarrow x = 0$.

Fix a faithful normal state $\phi$ on $M$. (Consistent with our convention about separable $\mathcal{H}$'s, we shall only consider $M$'s with separable preduals.) The well-known "Gelfand–Naimark–Segal" construction then yields a faithful $M$-module which is usually denoted by $L^2(M,\phi)$ – motivated by the fact that if $M = L^\infty(\Omega,\mu)$, and $\phi(f) = \int f \,\mathrm{d}\nu$, with $\nu$ a probability measure mutually absolutely continuous with respect to $\mu$, then $L^2(M,\phi) = L^2(\Omega,\nu)$ with $L^\infty(\Omega,\mu)$ acting as multiplication operators. The construction mimics this case: the assumptions on $\phi$ ensure that the equation

$$\langle x, y \rangle = \phi(y^*x)$$

defines a positive-definite inner product on $M$; let $L^2(M,\phi)$ be the Hilbert space completion of $M$. It turns out that the operator of left-multiplication by an element of $M$ extends as a bounded operator to $L^2(M,\phi)$, and it then follows easily that $L^2(M,\phi)$ is indeed a faithful $M$-module, thereby establishing fact (4) above.

Since we wish to distinguish between elements of the dense subspace $M$ of $L^2(M,\phi)$ and the operators of left-multiplication by members of $M$, let us write $\hat{x}$ for an element of $M$ when thought of as an

element of $L^2(M,\phi)$, and $x$ for the operator of left-multiplication by $x$; thus, for instance, $\hat{x} = x\hat{1}$, and $x\hat{y} = \widehat{xy}$, $\langle x\hat{1}, \hat{1}\rangle = \phi(x)$, etc.

## Modular Theory

While type *III* factors were more or less an enigma at the time of von Neumann, all that changed with the advent of Connes. The first major result of this "type *III* era" is the celebrated "Tomita–Takesaki theorem" (cf. Takesaki (1970)), which views the adjoint mapping on $M$ as an appropriate operator on $L^2(M,\phi)$, and analyzes its polar decomposition. Specifically, we have:

**Theorem 5** *If $\phi$ is any faithful normal state on $M$, consider the densely defined conjugate-linear operator given, with domain $\{\hat{x}: x \in M\}$, by $S_\phi^{(0)}(\hat{x}) = \widehat{x^*}$. Then,*

(i) *there is a unique conjugate-linear operator $S_\phi$ (the "closure of $S_\phi^{(0)}$") whose graph is the closure of the graph of $S_\phi^{(0)}$; if we write $S_\phi = J_\phi \Delta_\phi^{1/2}$ for the polar decomposition of the conjugate-linear closed operator $S_\phi$, then*
(ii) *$J_\phi$ is an antiunitary involution on $L^2(M,\phi)$ (i.e., it is a conjugate-linear norm-preserving bijection of $L^2(M,\phi)$ onto itself which is its own inverse);*
(iii) *$\Delta_\phi$ is an injective positive self-adjoint operator on $L^2(M,\phi)$ such that $J_\phi f(\Delta_\phi) J_\phi = \bar{f}(\Delta_\phi^{-1})$ for all Borel functions $f: \mathbb{R} \to \mathbb{R}$, and most crucially*
(iv)

$$J_\phi M J_\phi = M' \quad \text{and} \quad \Delta_\phi^{it} M \Delta_\phi^{-it} = M \ \forall t \in \mathbb{R}$$

*(Here and elsewhere, we shall identify $x \in M$ with the operator of "left-multiplication by $x$" on $L^2(M,\phi)$.)*

Thus, each faithful normal state $\phi$ on $M$ yields a one-parameter group $\{\sigma_t^\phi : t \in \mathbb{R}\}$ of automorphisms of $M$ – referred to as the group of "modular automorphisms" – given by

$$\sigma_t^\phi(x) = \Delta_\phi^{it} x \Delta_\phi^{-it}$$

The extent of dependence of the modular group on the state is captured precisely by Connes' Radon Nikodym theorem (Connes 1973), which shows that the modular groups associated to two different faithful normal states are related by a "unitary cocycle in $M$." This has the consequence that if $\epsilon : \mathrm{Aut}(M) \to \mathrm{Out}(M) = \mathrm{Aut}(M)/\mathrm{Int}(M)$ is the quotient mapping – where $\mathrm{Int}(M)$ denotes the normal subgroup of inner automorphisms given by unitary elements of $M$ – then the one-parameter subgroup $\{\epsilon(\sigma_t^\phi) : t \in \mathbb{R}\}$ of $\mathrm{Out}(M)$ is independent of $\phi$.

## Connes' Classification and Injective Factors

Given a factor $M$, Connes defined

$$S(M) = \bigcap \{\mathrm{spec}(\Delta_\phi) : \phi \text{ a faithful normal state on } M\}$$

which is obviously an isomorphism invariant. He then classified (Connes 1973) type *III* factors into a continuum of factors:

**Theorem 6** *Let $M$ be a factor. Then,*

(i) *$0 \in S(M) \Leftrightarrow M$ is of type III; and*
(ii) *if $M$ is a type III factor, there are three mutually exclusive and exhaustive possibilities:*

- *$(III_0) S(M) = \{0, 1\}$*
- *$(III_\lambda) S(M) = \{0\} \cup \lambda^{\mathbb{Z}}$, for some $0 < \lambda < 1$*
- *$(III_1) S(M) = [0, \infty)$*

**Example 7** Consider the compact group $\Omega = \prod_{n=1}^\infty G_n$ where $G_n$ is a finite cyclic group of order $\nu_n$ for each $n$. Let $\mu = \prod_{n=1}^\infty \mu_n$, where $\mu_n$ is a probability measure defined on the subsets of $G_n$ which assigns positive mass to each singleton. Let $G = \oplus_{n=1}^\infty G_n$ be the dense subgroup of $\Omega$ consisting of finitely nonzero sequences. It is not hard to see that each translation $T_g, g \in G$ (given by $T_g(\omega) = g + \omega$) is a nonsingular transformation of the measure space $(\Omega, \mu)$. The density of $G$ in $\Omega$ shows that this action of $G$ on $L^\infty(\Omega,\mu)$ is fixed-point-free and ergodic, with the result that the crossed product $L^\infty(\Omega,\mu) \rtimes G$ is a factor.

Krieger showed that in the case of a Krieger factor $M = L^\infty(\Omega,\mu) \rtimes G$, the invariant $S(M)$ agrees with the so-called "asymptotic ratio set" of the group $G$ of nonsingular transformations, which is computable purely in terms of the Radon–Nikodym derivatives $\mathrm{d}(\mu \circ T_t)/\mathrm{d}\mu$. Using this ratio set description, it is not hard to see that the Krieger factor $M$ given by the infinite product $\Omega$

- is a factor of type $III_\lambda$ if $\nu_n = 2$ and $\mu_n\{0\} = \lambda/(1+\lambda)$ for all $n$;
- is a factor of type $III_1$ if $\nu_n = 2$ and $\mu_{2n}\{0\} = \lambda/(1+\lambda)$, $\mu_{2n+1}\{0\} = \kappa/(1+\kappa)$, for all $n$, provided that $\{\lambda, \kappa\}$ generates a dense multiplicative subgroup of $\mathbb{R}_+^\times$;
- can be of type $III_0$.

Among all factors, Connes identified one tractable class – the so-called injective factors – which are ubiquitous and amenable to classification. To start

with, he established the equivalence of several (seemingly quite disparate) requirements on a von Neumann algebra $M \subset \mathcal{L}(\mathcal{H})$ – ranging from injectivity (meaning the existence of a projection of norm 1 from $\mathcal{L}(\mathcal{H})$ onto $M$) to "approximate finite dimensionality" (meaning $M = (\cup_n A_n)''$ for some increasing sequence $A_1 \subset A_2 \subset \cdots \subset A_n \subset \cdots$ of finite-dimensional $*$-subalgebras). In the same paper, Connes (1976) essentially finished the complete classification of injective factors. Only the injective $III_1$ factor withstood his onslaught; but eventually even it had to surrender to the technical virtuosity of Haagerup (1987) a few years later!

In the language we have developed thus far, the classification of injective factors may be summarized as follows:

- Every injective factor is isomorphic to a Krieger factor.
- Up to isomorphism, there is a unique injective factor of each type with the solitary exception of $III_0$.
- Injective factors of type $III_0$ are classified (up to isomorphism) by an invariant of an ergodic-theoretic nature called the "flow of weights"; unfortunately, coming up with a crisp description of this invariant, which is simultaneously accessible to the nonexpert and is consistent with the stipulated size of this survey, is beyond the scope of this author.

The interested reader is invited to browse through one of the books (Connes 1994, Sunder 1986, Dixmier 1981) for further details; the third book is the oldest (a classic but the language has changed a bit since it was written), the second is more recent (but quite sketchy in many places), and the first is clearly the best choice (if one has the time to read it carefully and digest it). Alternatively, the interested reader might want to browse through the encyclopediac treatments (Kadison and Ringrose) or (Takesaki).

*See also:* Algebraic Approach to Quantum Field Theory; Bicrossproduct Hopf Algebras and Noncommutative Spacetime; Braided and Modular Tensor Categories; $C^*$-Algebras and Their Classification; Ergodic Theory; Finite-Type Invariants; Hopf Algebra Structure of Renormalizable Quantum Field Theory; Hopf Algebras and $q$-Deformation Quantum Groups; The Jones Polynomial; Knot Theory and Physics; Noncommutative Geometry and the Standard Model; Noncommutative Tori, Yang–Mills and String Theory; Positive Maps on $C^*$-Algebras; Quantum 3-Manifold Invariants; Quantum Entropy; Tomita–Takesaki Modular Theory; von Neumann Algebras: Subfactor Theory.

## Further Reading


Connes A (1973) Une classification des facteurs de type III. *Annales Scientifiques de l'Ecole Normale Superieure* 6: 133–252.

Connes A (1976) Classification of injective factors. *Annals of Mathematics* 104: 73–115.

Connes A (1994) *Non-commutative Geometry*. San Diego: Academic Press.

Dixmier J (1981) *von Neumann Algebras*. Amsterdam: North Holland.

Haagerup U (1987) Connes' bicentralizer problem and uniqueness of the injective factor of type $\mathrm{III}_1$. *Acta Mathematica* 158: 95–148.

Kadison RV and Ringrose JR (1983/1986) *Fundamentals of the Theory of Operator Algebras*, vol. I–IV. New York: Academic Press.

Krieger W (1970) On the Araki–Woods asymptotic ratio set and non-singular transformations of a measure space. In: *Contributions to Ergodic Theory and Probability*, pp. 158–177. Lecture Notes in Mathematics, vol. 160, Springer-Verlag.

Krieger W (1976) On ergodic flows and the isomorphism of factors. *Mathematische Annalen* 223: 19–70.

Murray FJ and von Neumann J (1936) Rings of operators. *Annals of Mathematics* 37: 116–229.

Murray FJ and von Neumann J (1937) On rings of operators, II. *Transactions of the American Mathematical Society* 41: 208–248.

Murray FJ and von Neumann J (1943) On rings of operators, IV. *Annals of Mathematics* 44: 716–808.

Sakai S (1983) *$C^*$-algebras and $W^*$-algebras*. Berlin–New York: Springer-Verlag.

Sunder VS (1986) *An Invitation to von Neumann Algebras*. New York: Springer-Verlag.

Takesaki M (1970) *Tomita's Theory of Modular Hilbert Algebras and its Applications*, Lecture Notes in Mathematics 128, Berlin–Heidelberg–New York: Springer-Verlag.

Takesaki M (1979/2003) *Theory of Operator Algebras*, vols. I–III. Heidelberg: Springer Verlag.

von Neumann J (1936) On rings of operators, III. *Annals of Mathematics* 37: 111–115.

von Neumann J (1949) On rings of operators. Reduction theory. *Annals of Mathematics*, 50: 401–485.

# von Neumann Algebras: Subfactor Theory

**Y Kawahigashi**, University of Tokyo, Tokyo, Japan



## Introduction

Subfactor theory was initiated by Jones (1983) and has experienced rapid progress beyond the framework of operator algebras. Here we start with a basic introduction in this section.

A factor is a von Neumann algebra with a trivial center. A von Neumann algebra $M$ is an algebra of bounded linear operators on a Hilbert space $H$, which contains the identitiy operator and is closed under the $*$-operation and weak operator topology, and its center is the intersection of $M$ and its commutant

$$M' = \{x \in B(H) | xy = yx \text{ for all } y \in M\}$$

where $B(H)$ denotes the set of all the bounded linear operators on $H$. (We are mostly interested in separable, infinite-dimensional Hilbert spaces. A von Neumann algebra is automatically closed also in the norm topology and thus it is also a $C^*$-algebra.) By definition, a factor $M$ acts on a certain Hilbert space $H$, but we also consider its action on another Hilbert space $K$, that is, a $\sigma$-weakly continuous homomorphism preserving the $*$-operation from $M$ into $B(K)$. A subfactor is a factor $N$ which is contained in another factor $M$ and has the same identity. A factor is classified into types $\mathrm{I}_n$ $(n = 1, 2, 3, \ldots)$, $\mathrm{I}_\infty$, $\mathrm{II}_1$, $\mathrm{II}_\infty$, and III. In most of the interesting studies of subfactors, the two factors are of both type $\mathrm{II}_1$ or both type III. A factor $M$ is said to be of type $\mathrm{II}_1$ if it is infinite dimensional and has a finite trace $\mathrm{tr}: M \to \mathbf{C}$. By definition, a finite trace tr is a linear functional on $M$ satisfying $\mathrm{tr}(1) = 1$, $\mathrm{tr}(xy) = \mathrm{tr}(yx)$ for all $x, y \in M$, and $\mathrm{tr}(x^*x) \geq 0$ for all $x \in M$. When a factor $M$, not isomorphic to $\mathbf{C}$, acts on a separable Hilbert space, it is of type III if and only if for any two nonzero projections $p, q \in M$, we have an operator $v \in M$ with $vv^* = p$ and $v^*v = q$. One obviously cannot have a trace on such a factor. (See Takesaki (2002, 2003) for a general theory on factors.)

Let $M$ be a type $\mathrm{II}_1$ factor acting on a Hilbert space $H$. We then have the coupling constant of Murray and von Neumann, which is denoted by $\dim_M H$ and belongs to $(0, \infty]$. This measures the relative dimension of $H$ with respect to $M$. Note that the factor $M$ acts on $M$ itself by the left multiplication. We introduce an inner product on $M$ by $(x, y) = \mathrm{tr}(y^*x)$ and denote the completion by $L^2(M)$. Then $M$ acts on this Hilbert space and we have $\dim_M L^2(M) = 1$.

Let $N \subset M$ be a subfactor and suppose that both $N$ and $M$ are of type $\mathrm{II}_1$. (We then simply say that $N \subset M$ is a type $\mathrm{II}_1$ subfactor.) Suppose that $M$ acts on a Hilbert space $H$ with $\dim_M H < \infty$. Then we define the Jones index of $N$ in $M$ by

$$[M : N] = \frac{\dim_N H}{\dim_M H} \in [1, \infty]$$

This number is independent of the choice of $H$, as long as $\dim_M H < \infty$, so we can take $H = L^2(M)$, then we have $[M : N] = \dim_N L^2(M)$. The equality $[M : N] = 1$ means $M = N$. The first major discovery of Jones (1983) is that the value of the Jones index is in the set

$$\{4\cos^2(\pi/m) | m = 3, 4, 5, \ldots\} \cup [4, \infty] \qquad [1]$$

and all the values in this set are indeed realized.

Suppose we have a $\mathrm{II}_1$ factor $M$ and an action of an at most countable, discrete group $G$ on $M$, that is, a homomorphism $\alpha: G \to \mathrm{Aut}(M)$, where $\mathrm{Aut}(M)$ is the automorphism group of $M$. Then we have a construction $M \rtimes_\alpha G$, called the crossed product. If $\alpha_g$ is not an inner automorphism of $M$ for any $g \in G$ other than the identity element of $G$, then $M \rtimes_\alpha G$ is also a type $\mathrm{II}_1$ factor. (An automorphism $\pi$ of $M$ is said to be inner if it is of the form $\pi(x) = uxu^*$ for some unitary operator $u \in M$.) The index of a subfactor $M \subset M \rtimes_\alpha G$ is the order of $G$, which can be infinite. If we have a subgroup $H$ of $G$, then we obtain a subfactor $M \rtimes_\alpha H \subset M \rtimes_\alpha G$ and its index is given by the index $[G : H]$ of the subgroup $H$. This analogy to the index of a subgroup is the origin of the terminology of the Jones index for a subfactor. The Jones index is also analogous to the degree of an extension of a field. From the viewpoint of this analogy, subfactor theory can be regarded as a certain generalized analogue (or the "quantum" version) of the classical Galois theory for field extensions. (The direct analog of the classical Galois correspondence for subfactors was studied by Nakamura–Takeda in the early days, and Izumi–Longo–Popa gave the most general form.)

The tools Jones (1983) has introduced to study subfactors are as follows. Let $N \subset M$ be a subfactor of type $\mathrm{II}_1$ with finite Jones index. We consider the actions of $N$, $M$ on $L^2(M)$. The completion of $N$ with respect to the inner product given by the trace gives $L^2(N)$, which is naturally regarded as a closed subspace of $L^2(M)$. Let $e_N$ be the projection on $L^2(M)$ onto $L^2(N)$, which is called the Jones

projection. We define $M_1$ to be the von Neumann algebra generated by $M$ and $e_N$ on $L^2(M)$. This is again a type $\mathrm{II}_1$ factor and denoted by $M_1$. This construction is called the basic construction. We obtain $[M_1 : M] = [M : N]$. Repeat the same procedure for $M \subset M_1$ acting on $L^2(M_1)$ this time. In this way, we have an increasing sequence of type $\mathrm{II}_1$ factors,

$$N \subset M \subset M_1 \subset M_2 \subset M_3 \subset \cdots$$

which is called the Jones tower. We label the corresponding Jones projections as $e_1 = e_N$, $e_2 = e_M$, $e_3 = e_{M_1}, \ldots$. We then have the following celebrated Jones relations:

$$e_j e_k = e_k e_j, \quad \text{if } |j-k| > 1$$
$$e_j e_{j\pm 1} e_j = \frac{1}{[M:N]} e_j \qquad [2]$$

Jones proved the above-mentioned restriction on the possible values of the Jones index using these relations. The realization of the index values below 4 in the set [1] by Jones also relies on these relations of the Jones projection. The basic construction is also possible for the other direction. That is, we can construct a subfactor $N_1 \subset N$ so that $N \subset M$ is the basic construction of $N_1 \subset N$. This is called the downward basic construction. This $N_1$ is not unique, but is unique up to an inner automorphism of $N$.

A subfactor $N \subset M$ is said to be irreducible if the relative commutant $N' \cap M$ is equal to $\mathbf{C}$. If a subfactor has Jones index less than 4, then it is automatically irreducible. The original realization of the Jones index values above 4 by Jones was through reducible subfactors. Popa proved that all the values above 4 are realized with irreducible subfactors. A factor is said to be hyperfinite if it has a dense subalgebra given as the union of increasing sequence of finite-dimensional $*$-algebras. If $M$ is a hyperfinite type $\mathrm{II}_1$ factor, then its subfactor is automatically also hyperfinite by a deep theorem of Connes. For hyperfinite, irreducible type $\mathrm{II}_1$ subfactors, it is still an open problem to determine all the possible values of the Jones index.

For type $\mathrm{II}_1$ factors $N \subset M \subset P$, the Jones index $[P:N]$ is equal to the product $[P:M][M:N]$. Thus for the Jones tower, we have $[M_k:N] = [M:N]^{k+1}$. In general, if a subfactor $N \subset M$ has a finite Jones index, then the relative commutant $N' \cap M$ is automatically finite dimensional. So, if we start with a type $\mathrm{II}_1$ subfactor $N \subset M$ with finite Jones index, we have an increasing sequence of finite-dimensional algebras as follows:

$$N' \cap M \subset N' \cap M_1 \subset N' \cap M_2 \subset N' \cap M_3 \subset \cdots \qquad [3]$$

These finite-dimensional algebras are called higher relative commutants of $N \subset M$. We draw the Bratteli diagram for the higher relative commutants as follows. Consider $N' \cap M_k$ (with convention $M_{-1} = N$, $M_0 = M$), then it is a finite-dimensional $*$-algebra; thus, it is of the form $\bigoplus_j M_{n_j}(\mathbf{C})$, where we have only finitely many direct summands. We draw a dot for each summand. We similarly draw a dot for each summand in $\bigoplus_l M_{m_l}(\mathbf{C})$ for $N' \cap M_{k+1}$. Let $\iota$ be the inclusion map from $N' \cap M_k = \bigoplus_j M_{n_j}(\mathbf{C})$ to $N' \cap M_{k+1} \bigoplus_l M_{m_l}(\mathbf{C})$ and $p_l$ the identity of $M_{m_l}(\mathbf{C})$, which is a projection in $N' \cap M_{k+1}$. We denote by $\mu_{jl}$ the multiplicity of the embedding map $x \mapsto \iota(x)p_l$ from $M_{n_j}(\mathbf{C})$ to $M_{m_l}(\mathbf{C})$. Then we draw $\mu_{jl}$ edges from the $j$th dot for $M_{n_j}(\mathbf{C})$ to the $l$th dot for $M_{m_l}(\mathbf{C})$. We repeat this procedure for all $k$, and get a picture as in Figure 1, which is called the Bratteli diagram of the higher relative commutants of $N \subset M$.

It turns out that the edges connecting the $k$th and $(k+1)$th steps of the Bratteli diagram consist of the reflection of those connecting the $(k-1)$th and $k$th steps, and a (possibly empty) new part. The "new" parts taken altogether in the above Bratteli diagram constitute the principal graph of a subfactor $N \subset M$. In the example of Figure 1, the principal graph is the Dynkin diagram $A_5$. In general, a principal graph can be finite or infinite. If it is finite, we say that a subfactor is of finite depth. If a subfactor has the Jones index less than 4, it is automatically of finite depth and the principal graph must be one of the $A$–$D$–$E$ Dynkin diagrams.

Pimsner and Popa (1986) obtained the characterization of the Jones index value in terms of the Pimsner–Popa inequality for a conditional expectation. This can be used as a definition of the index for a subfactor of arbitrary type (and even for $C^*$-subalgebras). Kosaki obtained a definition of the index for type III subfactors based on works of Connes and Haagerup.

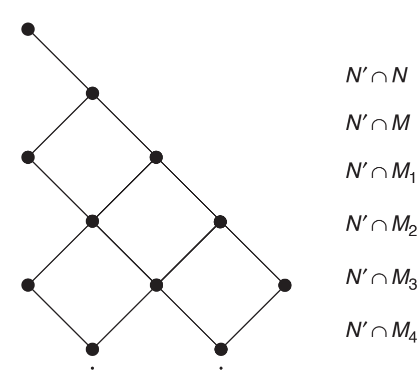



**Figure 1** The Bratteli diagram of the higher relative commutants.

## Analytic Classification Theory

If $M$ is a hyperfinite type $II_1$ factor, then it is unique up to isomorphism. So any subfactor of such $M$ is isomorphic to $M$ itself. We next consider the classification problem of hyperfinite type $II_1$ subfactors. We say that a subfactor $N \subset M$ is isomorphic to $P \subset Q$ if we have an isomorphism of $M$ onto $Q$ which maps $N$ onto $P$. The following tower of finite-dimensional algebras is a natural invariant for a type $II_1$ subfactor $N \subset M$ with finite Jones index and it is called the standard invariant for $N \subset M$:

$$\begin{array}{ccccccc} M' \cap M & \subset & M' \cap M_1 & \subset & M' \cap M_2 & \subset & \cdots \\ \cap & & \cap & & \cap & & \\ N' \cap M & \subset & N' \cap M_1 & \subset & N' \cap M_2 & \subset & \cdots \end{array} \qquad [4]$$

Each square

$$\begin{array}{ccc} M' \cap M_k & \subset & M' \cap M_{k+1} \\ \cap & & \cap \\ N' \cap M_k & \subset & N' \cap M_{k+1} \end{array}$$

is a special combination of inclusions called a commuting square. Under a fairly general condition (called extremality of a subfactor, which automatically holds for an irreducible subfactor), the above sequence [4] is anti-isomorphic to the following sequence of finite-dimensional algebras, including the trace values:

$$\begin{array}{ccccccc} M' \cap M & \subset & N' \cap M & \subset & N_1' \cap M & \subset & \cdots \\ \cap & & \cap & & \cap & & \\ M' \cap M_1 & \subset & N' \cap M_1 & \subset & N_1' \cap M_1 & \subset & \cdots \end{array} \qquad [5]$$

where $\cdots \subset N_3 \subset N_2 \subset N_1 \subset N \subset M$ is given by repeated downward basic constructions. So, if the closure of $\bigcup_j (N_j' \cap M_1)$ in the weak operator topology is equal to $M_1$ for an appropriate choice of $N_j$'s, then the closure of $\bigcup_j (N_j' \cap M)$ is also $M$, and the isomorphism class of the subfactor $N \subset M$ is recovered from the standard invariant. In such a case, we say that a subfactor has a generating property, and then we have a complete classification of subfactors in terms of the standard invariant. Popa (1994) introduced a notion called strong amenability and proved that a subfactor of type $II_1$ is strongly amenable if and only if it has the generating property. This is the fundamental result in the classification of subfactors. A hyperfinite type $II_1$ subfactor with finite Jones index and finite depth is automatically strongly amenable, so such a subfactor is covered by this classification theorem of Popa. Popa also has a similar result for subfactors of type III.

## Constructions and Combinatorial Classification

As mentioned in the above section, Jones constructed hyperfinite type $II_1$ subfactors for all possible index values below 4. They have the Dynkin diagrams $A_n$ as the principal graphs. It has been an important problem to construct new subfactors since then. Using the Hecke algebras, Wenzl constructed a series of subfactors with index values $\sin^2(N\pi/k)/\sin^2(\pi/k)$ with $N = 2, 3, 4, \ldots$, where the series for $N = 2$ coincide with the ones constructed by Jones. Wenzl's dimension estimate in this work for the relative commutant has been an important tool to study subfactors. It was soon noticed that the subfactors of Jones and Wenzl are related to the quantum groups $U_q(\mathrm{sl}_N)$ of Drinfel'd–Jimbo, at the value of the deformation parameter $q$ at $\exp(\pi i/k)$. Constructions of subfactors from other quantum groups have been given by Wenzl.

Ocneanu (1988) has introduced a notion of a paragroup and characterized the higher relative commutants arising from a type $II_1$ subfactor with finite Jones index and finite depth as a paragroup. If we start with a subfactor $N \subset N \rtimes_\alpha G$ for a finite group $G$, the corresponding paragroup contains complete information on the group $G$ and its representations. In this sense, a paragroup is a generalization of a (finite) group. The basic idea is to regard the bimodule ${}_N L^2(M)_M$ as an analog of the fundamental representation of a compact Lie group and make finite relative tensor products

$$\cdots \otimes_N L^2(M) \otimes_M L^2(M) \otimes_N L^2(M) \otimes_M \cdots$$

Then one makes an irreducible decomposition and studies various intertwiners arising from these irreducible bimodules. In this way, we obtain a certain combinatorial object and it is called a paragroup. The vertices of the principal graphs correspond to irreducible bimodules and the edges correspond to basis vectors in the intertwiner spaces. Note that by Popa's theorem explained in the previous section, a classification of subfactors of a hyperfinite type $II_1$ factor with finite Jones index and finite depth is reduced to one of paragroups.

Using this theory of paragroups, Ocneanu has found that the Dynkin diagrams $A_n$, $D_{2n}$, $E_6$, and $E_8$ are realized as principal graphs of subfactors, but $D_{2n+1}$ and $E_7$ are not. Furthermore, each of the graphs $A_n$ and $D_{2n}$ has unique realization and each of $E_6, E_8$ has two realizations. At the index value 4, the principal graph must be one of the extended Dynkin diagrams, $A_{2n-1}^{(1)}$, $D_n^{(1)}$, $E_6^{(1)}$, $E_7^{(1)}$, $E_8^{(1)}$, $A_\infty$, $A_{\infty,\infty}$, and $D_\infty$, and all are realized. (The last three correspond to subfactors of infinite depth.)

See Evans and Kawahigashi (1998) and Goodman *et al.* (1989) for these constructions and classifications. Evans-Kawahigashi and Xu studied the orbifold construction of subfactors applied to the Hecke algebra subfactors of Wenzl.

In a theory of integrable lattice models, we have squares with labeled edges, and we assign complex numbers to them. A paragroup has much formal similarity to such a lattice model, and the paragroups of subfactors of Jones and Wenzl correspond to the lattice models of Andrews–Baxter–Forrester.

Goodman–de la Harpe–Jones have another construction of subfactors from the Dynkin diagrams, and for $E_6$ this gives a hyperfinite type $\mathrm{II}_1$ subfactor with Jones index $3+\sqrt{3}$ and finite depth. Haagerup has made a combinatorial study on type $\mathrm{II}_1$ subfactors with Jones index values between 4 and $3+\sqrt{3}$ and obtained a list of candidates of possible higher relative commutants. Haagerup himself showed one in the list with Jones index $(5+\sqrt{13})/2$ is indeed realized. Asaeda–Haagerup showed that another in the list having the Jones index $(5+\sqrt{17})/2$ is also realized. These two examples are still among the most mysterious examples of subfactors today and do not seem to arise from other constructions using quantum groups or conformal field theory. Izumi has another construction of a subfactor with the Jones index $(7+\sqrt{29})/2$ using an endomorphism of the Cuntz algebra.

Popa has obtained a complete characterization of higher relative commutants including the case of infinite depth, and axiomatized the higher relative commutant as the standard $\lambda$-lattices. Xu has constructed standard $\lambda$-lattices, hence subfactors, from quantum groups. This realization of Popa of a given standard $\lambda$-lattice produces a nonhyperfinite type $\mathrm{II}_1$ subfactor. Popa–Shlyakhtenko later showed that any standard $\lambda$-lattice is realized for a subfactor of a single type $\mathrm{II}_1$ factor, a group $\mathrm{II}_1$ factor arising from the free group $F_\infty$ having countably many generators, which is not hyperfinite.

Jones (1999) has introduced a combinatorial characterization of standard $\lambda$-lattices as planar algebras. This approach uses planar operads based on tangles and provides a new viewpoint on the structure of higher relative commutants. More studies on planar algebras have been done by Bisch–Jones.

## Topological Invariants in Three Dimensions and Tensor Categories

Through the relations of the Jones projections, Jones (1985) discovered the Jones polynomial as an invariant for links. This was the beginning of series of entirely new theories in three-dimensional topology. The Jones polynomial was quickly generalized to the two-variable HOMFLY polynomial by Hoste, Ocneanu, Millet, Freyd, Lickorish, and Yetter.

A three-dimensional topological quantum field theory ($\mathrm{TQFT}_3$) assigns a complex number to each closed oriented 3-manifold and a finite dimensional vector space to each closed oriented surface. Furthermore, to each compact oriented 3-manifold with boundary, it assigns a vector in the vector space corresponding to its boundary. Turaev–Viro have constructed $\mathrm{TQFT}_3$ from combinatorial data called quantum $6j$-symbols arising from quantum groups. Ocneanu has found that a subfactor of finite index and finite depth also produces quantum $6j$-symbols, which give rise to a $\mathrm{TQFT}_3$ generalizing the Turaev–Viro construction. See Evans and Kawahigashi (1998) for this construction. Reshetikhin–Turaev have another construction of $\mathrm{TQFT}_3$ from a modular tensor category, which is a braided tensor category with nondegenerate braiding. Ocneanu has found a subfactor version of the quantum double construction which produces a modular tensor category from a type $\mathrm{II}_1$ subfactor of finite index and finite depth. From a type $\mathrm{II}_1$ subfactor of finite index and finite depth, we can apply Ocneanu's generalization of the Turaev–Viro construction on one hand, and also the Reshetikhin–Turaev construction to the modular tensor category arising from the quantum double construction of Ocneanu. The resulting two $\mathrm{TQFT}_3$s are shown to be equal by Kawahigashi–Sato–Wakui. Concrete computations of these topological invariants have been made by Sato–Wakui based on Izumi's work. Turaev and Wenzl have other constructions of $\mathrm{TQFT}_3$ and modular tensor categories.

## Algebraic Quantum Field Theory

An operator algebraic approach to quantum field theory is called algebraic quantum field theory and the standard reference is Haag (1996). In this approach, instead of quantum fields which are operator-valued distributions, we consider a family $\{A(O)\}$ of von Neumann algebras parametrized by spacetime regions $O$ in a Minkowski space. Each $A(O)$ is meant to be generated by self-adjoint operators which are observables in $O$. We axiomatize such a family of von Neumann algebras and call one a local net of von Neumann algebras. It is enough to take $O$ of a special form, called a double cone. The name "local" comes from the locality axiom which is a mathematical expression of the Einstein causality on a Minkowski space. The

Poincaré group is used as the spacetime symmetry of the Minkowski space. Doplicher *et al.* (1971, 1974) have introduced a representation theory of a local net $A$ of von Neumann algebras and found that a "physically nice" representation is realized as an endomorphism of a one von Neumann algebra $A(O)$ for some fixed $O$. They have a notion of a statistical dimension for such a representation and it is an integer (or infinite) if the spacetime dimension is larger than 2. Longo (1989, 1990) has shown that this statistical dimension of a representation is equal to the square root of the index $[A(O) : \lambda(A(O))]$, where $\lambda$ is the corresponding endomorphism of $A(O)$ to the representation. The relation between algebraic quantum field theory and subfactor theory has been found in this way. Longo (1989, 1990) has also started a theory of canonical endomorphisms for a subfactor and Izumi has further studied it. Longo has later obtained a characterization when an endomorphism of a factor becomes a canonical endomorphism by introducing a $Q$-system.

Recently, conformal field theory has attracted much attention. An approach based on algebraic quantum field theory describes a conformal field theory with a local net of von Neumann algebras on a two-dimensional Minkowski space with diffeomorphism group as the spacetime symmetry. We can restrict such a theory into a tensor product of two theories on the circle, the compactified one-dimensional Euclidean space. Each theory on the circle is called a chiral conformal field theory and described by a local conformal net of von Neumann algebras, which is a family of von Neumann algebras parametrized by intervals on the circle. The name "conformal" comes from the fact that we use the orientation preserving diffeomorphism group on the circle as the symmetry group of the space. For a local conformal net $A$ of von Neumann algebras on the circle with natural irreducibility assumption, each von Neumann algebra $A(I)$ is automatically a type III factor. The Doplicher–Haag–Roberts theory works in this setting after an appropriate adaptation as in Fredenhagen *et al.* (1989) and each representation of a local conformal net of von Neumann algebras is realized by an endomorphism of $A(I)$, where $I$ is an arbitrarily fixed interval on the circle. (Here we do not need an assumption that a representation is "physically nice" since it now automatically holds.) Now the representations give a braided tensor category.

Buchholz–Mack–Todorov constructed examples of local conformal nets of von Neumann algebras on the circle using the U(1)-current algebra. Wassermann (1998) has constructed more examples using positive energy representations of the loop groups LSU($N$) and computed their representation theory, and his construction has been extended to other Lie groups by Toledano Laredo and others. For the local conformal net $A$ of von Neumann algebras on the circle arising from LSU($N$), we take an endomorphism $\lambda$ of $A(I)$ arising from a representation of the local conformal net, then we have a subfactor $\lambda(A(I)) \subset A(I)$. This is isomorphic to the type $\mathrm{II}_1$ subfactor constructed by Jones and Wenzl tensored with a common type III factor.

Longo–Rehren (1995) started the study of a local net of subfactors, $A(I) \subset B(I)$. They have defined a certain induction procedure which gives a representation of the larger local conformal net $B$ from that of $A$. This procedure is today called $\alpha$-induction. Xu has studied this procedure and found several basic properties. In the cases of local conformal nets of subfactors arising from conformal embeddings, he has found a simple construction of subfactors with principal graphs $E_6$ and $E_8$ using $\alpha$-induction. In the context of subfactor theory, $\alpha$-induction has been further studied by Böckenhauer–Evans–Kawahigashi, together with graphical methods of Ocneanu on the Dynkin diagrams. More detailed studies on local conformal nets of factors on the circle have been pursued partly using various techniques of subfactor theory, including classification of local conformal nets of von Neumann algebras on the circle with central charge less than 1 by Kawahigashi–Longo.

*See also:* Algebraic Approach to Quantum Field Theory; Braided and Modular Tensor Categories; $C^*$-Algebras and Their Classification; Hopf Algebras and $q$-Deformation Quantum Groups; The Jones Polynomial; Quantum 3-Manifold Invariants; Quantum Entropy; von Neumann Algebras: Introduction, Modular Theory, and Classification Theory; Yang–Baxter Equations.

## Further Reading


Doplicher S, Haag R, and Roberts JE (1971) Local observables and particle statistics, I. *Communications in Mathematical Physics* 23: 199–230.

Doplicher S, Haag R, and Robert JE (1974) Local observables an particle statistics, II. *Communications in Mathematical Physics* 35: 49–85.

Evans DE and Kawahigashi Y (1998) *Quantum Symmetries on Operator Algebras.* Oxford: Oxford University Press.

Fredenhagen K, Rehren K-H, and Schroer B (1989) Superselection sectors with braid group statistics and exchange algebras. *Communications in Mathematical Physics* 125: 201–226.

Goodman F, de la Harpe P, and Jones VFR (1989) *Coxeter Graphs and Towers of Algebras*, vol. 14. Berlin: MSRI Publications, Springer.

Haag R (1996) *Local Quantum Physics.* Berlin: Springer.

Jones VFR (1983) Index for subfactors. *Inventiones Mathematical* 72: 1–25.

Jones VFR (1985) A polynomial invariant for knots via von Neumann algebras. *Bulletin of the American Mathematical Society* 12: 103–112.
Jones VFR (1999) Planar algebras, I, math.QA/9909027.
Longo R (1989) Index of subfactors and statistics of quantum fields I. *Communications in Mathematical Physics* 126: 217–247.
Longo R (1990) Index of subfactors and statistics of quantum fields II. *Communications in Mathematical Physics* 130: 285–309.
Longo R and Rehren K-H (1995) Nets of subfactors. *Reviews in Mathematical Physics* 7: 567–597.
Ocneanu A (1988) Quantized group, string algebras and Galois theory for algebras. In: Evans DE and Takesaki M (eds.) *Operator Algebras and Applications, vol. 2 (Warwick 1987)*, London Mathematical Society Lecture Note Series, vol. 36, pp. 119–172. Cambridge: Cambridge University Press.
Pimsner M and Popa S (1986) Entropy and index for subfactors. *Annales Scientifques de l'Ecole Normale* 19: 57–106.
Popa S (1994) Classification of amenable subfactors of type II. *Acta Mathematica* 172: 163–255.
Popa S (1995) An axiomatization of the lattice of higher relative commutants of a subfactor. *Inventiones Mathematicae* 120: 427–445.
Takesaki M (2002, 2003) *Theory of Operator Algebras I, II, III*. Berlin: Springer.
Wassermann A (1998) Operator algebras and conformal field theory III: Fusion of positive energy representations of SU(*N*) using bounded operators. *Inventiones Mathematicae* 133: 467–538.

# Vortex Dynamics

**M Nitsche**, University of New Mexico, Albuquerque, NM, USA



## Introduction

A vortex is commonly associated with the rotating motion of fluid around a common centerline. It is defined by the vorticity in the fluid, which measures the rate of local fluid rotation. Typically, the fluid circulates around the vortex, the speed increases as the vortex is approached and the pressure decreases. Vortices arise in nature and technology applications in a large range of sizes, as illustrated by the examples given in Table 1. The next section presents some of the mathematical background necessary to understand vortex formation and evolution. Next, some sample flows are described, including important instabilities and reconnection processes. Finally, some of the numerical methods used to simulate these flows are presented.

## Background

Let $D$ be a region in three-dimensional (3D) space containing a fluid, and let $\boldsymbol{x}=(x,y,z)^{\mathrm{T}}$ be a point in $D$. The fluid motion is described by its velocity $\boldsymbol{u}(\boldsymbol{x},t)=u(\boldsymbol{x},t)\boldsymbol{i}+v(\boldsymbol{x},t)\boldsymbol{j}+w(\boldsymbol{x},t)\boldsymbol{k}$, and depends on the fluid density $\rho(\boldsymbol{x},t)$, temperature $T(\boldsymbol{x},t)$, gravitational field $\boldsymbol{g}$, and other external forces possibly acting on it. The fluid vorticity is defined by $\boldsymbol{\omega}=\nabla\times\boldsymbol{u}$. The vorticity measures the local fluid rotation about an axis, as can be seen by expanding the velocity near $\boldsymbol{x}=\boldsymbol{x}_0$,

$$\boldsymbol{u}(\boldsymbol{x}) = \boldsymbol{u}(\boldsymbol{x}_0)+D(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0)+\tfrac{1}{2}\boldsymbol{\omega}(\boldsymbol{x}_0)\times(\boldsymbol{x}-\boldsymbol{x}_0) + O(|\boldsymbol{x}-\boldsymbol{x}_0|^2) \quad [1]$$

where

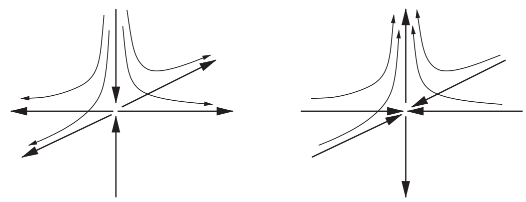
$$D(\boldsymbol{x}_0)=\tfrac{1}{2}(\nabla\boldsymbol{u}+\nabla\boldsymbol{u}^{\mathrm{T}}),\quad \nabla\boldsymbol{u}=\begin{bmatrix} u_x & u_y & u_z\\ v_x & v_y & v_z\\ w_x & w_y & w_z\end{bmatrix} \quad [2]$$

The first term $\boldsymbol{u}(\boldsymbol{x}_0)$ corresponds to translation: all fluid particles move with constant velocity $\boldsymbol{u}(\boldsymbol{x}_0)$. The second term $D(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0)$ corresponds to a strain field in the three directions of the eigenvectors of the symmetric matrix $D$. If the eigenvalue corresponding to a given eigenvector is positive, the fluid is stretched in that direction, if it is negative, the fluid is compressed. Note that, in incompressible flow, $\nabla\cdot\boldsymbol{u}=0$, so the sum of the eigenvalues of $D$ equals zero. Thus, at least one eigenvalue is positive and one negative. If the third eigenvalue is positive, fluid particles move towards sheets (Figure 1a). If the third eigenvalue is negative, fluid particles move towards tubes (Figure 1b). The last term in eqn [1], $(1/2)\boldsymbol{\omega}(\boldsymbol{x}_0)\times(\boldsymbol{x}-\boldsymbol{x}_0)$, corresponds to a rotation: near a point with $\boldsymbol{\omega}(\boldsymbol{x}_0)\neq 0$, the fluid rotates with angular velocity $|\boldsymbol{\omega}|/2$ in a plane normal to the vorticity vector $\boldsymbol{\omega}$. Fluid for which $\boldsymbol{\omega}=0$ is said to be irrotational.

A vortex line is an integral curve of the vorticity. For incompressible flow, $\nabla\cdot\boldsymbol{\omega}=\nabla\cdot(\nabla\times\boldsymbol{u})=0$, which implies that vortex lines cannot end in the interior of the flow, but must either form a closed loop

**Table 1** Sample vortices and typical sizes

| *Vortex* | *Diameter* |
|---|---|
| Superfluid vortices | $10^{-8}$ cm (= 1 Å) |
| Trailing vortex of Boeing 727 | 1–2 m |
| Dust devils | 1–10 m |
| Tornadoes | 10–500 m |
| Hurricanes | 100–2000 km |
| Jupiter's Red Spot | 25 000 km |
| Spiral galaxies | Thousands of light years |



(a) (b)

**Figure 1** Strain field: (a) two positive eigenvalues, sheet formation and (b) one positive eigenvalue, tube formation.

interior of the flow, but must either form a closed loop or start and end at a bounding surface. In 2D flow, $\boldsymbol{u} = u\boldsymbol{i} + v\boldsymbol{j}$ and the vorticity is $\boldsymbol{\omega} = \omega\boldsymbol{k}$, where $\omega$ is the scalar vorticity. Thus, in 2D, the vorticity points in the $z$-direction and the vortex lines are straight lines normal to the $x$–$y$ plane. A vortex tube is a bundle of vortex lines. The strength of a vortex tube is defined as the circulation $\int_C \boldsymbol{u} \cdot \mathrm{d}\boldsymbol{s}$ about a curve $C$ enclosing the tube. By Stokes' theorem,

$$\int_C \boldsymbol{u} \cdot \mathrm{d}\boldsymbol{s} = \iint_A \boldsymbol{\omega} \cdot \boldsymbol{n}\, \mathrm{d}S \qquad [3]$$

and thus the circulation can also be interpreted as the flux of vorticity through a cross section of the tube. In inviscid incompressible flow of constant density, Helmholtz' theorem states that the tube strength is independent of the curve $C$, and is therefore a well-defined quantity, and Kelvin's theorem states that a tube's strength remains constant in time. A vortex filament is an idealization in which a tube is represented by a single vortex line of nonzero strength.

The evolution equation for the fluid vorticity, as derived from the Navier–Stokes equations, is

$$\frac{\mathrm{d}\boldsymbol{\omega}}{\mathrm{d}t} = \boldsymbol{\omega} \cdot \nabla \boldsymbol{u} + \nu \Delta \boldsymbol{\omega} \qquad [4]$$

where $\mathrm{d}/\mathrm{d}t = \partial/\partial t + \boldsymbol{u} \cdot \nabla$ is the total time derivative. Equation [4] states that the vorticity is transported by the fluid velocity (first term), stretched by the fluid velocity gradient (second term), and diffused by viscosity $\nu$ (last term). These equations are usually nondimensionalized and written in terms of the Reynolds number, a dimensionless quantity inversely proportional to viscosity.

To understand high Reynolds number flow it is of interest to study the inviscid Euler equations. The corresponding vorticity evolution equation in 2D is

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = 0 \qquad [5]$$

which states that 2D vortex filaments in inviscid flow move with the fluid velocity. Furthermore, in incompressible flow, the fluid velocity is determined by the vorticity, up to an irrotational far-field component $\boldsymbol{u}_\infty$, through the Biot–Savart law,

$$\boldsymbol{u}(\boldsymbol{x}) = -\frac{1}{4\pi} \int \frac{(\boldsymbol{x} - \boldsymbol{x}') \times \boldsymbol{\omega}(\boldsymbol{x}')}{|\boldsymbol{x} - \boldsymbol{x}'|^3} \mathrm{d}\boldsymbol{x}' + \boldsymbol{u}_\infty \qquad [6]$$

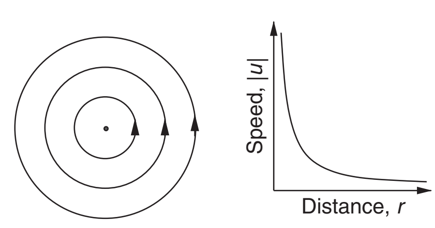
In planar 2D flow, eqn [6] reduces to

$$\boldsymbol{u}(\boldsymbol{x}) = K_{2\mathrm{D}}\omega, \quad K^*{}_{2\mathrm{D}}(\boldsymbol{x}) = \frac{1}{2\pi} \frac{-y\boldsymbol{i} + x\boldsymbol{j}}{|\boldsymbol{x}|^2} \qquad [7]$$

where $\omega(\boldsymbol{x})$ is the scalar vorticity. Equations [4], [5] and [6], [7] are the basis of the numerical methods discussed later in this article.

A vortex is typically defined by a region in the fluid of concentrated vorticity. A simple model is a point vortex in 2D flow, which corresponds to a straight vortex filament of unit circulation. The associated scalar vorticity is a $\delta$-function in the plane, and the induced velocity is obtained from the Biot–Savart law. For a point vortex at the origin, this reduces to the radial velocity field $\boldsymbol{u}(\boldsymbol{x}) = K^*{}_{2\mathrm{D}}\delta = K_{2\mathrm{D}}(\boldsymbol{x})$. Corresponding particle trajectories are shown in Figure 2a. The particle speed $|\boldsymbol{u}| = 1/(2\pi r)$ increases unboundedly as the vortex center is approached, and vanishes as $r \to \infty$ (Figure 2b). In general, the far-field velocity of a concentrated vortex behaves similarly to the one of a point vortex, with speeds decaying as $1/r$. Near the vortex center, the velocity typically increases in magnitude and, as a result, the fluid pressure decreases (Bernoulli's theorem). A vortex of arbitrary shape can be approximated by a sum of point vortices (in 2D) or vortex filaments (in 3D), as is often done for simulation purposes.

Vorticity can be generated by a variety of mechanisms. For example, vorticity can be generated by density gradients, which in turn are induced by spatial temperature variations. This mechanism explains the formation of warm-air vortices when a layer of hot air is trapped underneath cooler air. Vorticity is also generated near solid walls in the form of boundary layers caused by viscosity. To illustrate, imagine




(a) (b)

**Figure 2** Flow induced by a point vortex: (a) streamlines and (b) speed $|\boldsymbol{u}|$ vs. distance $r$.

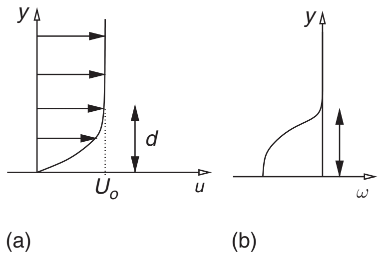



**Figure 3** Velocity and vorticity in boundary layer near a flat wall.

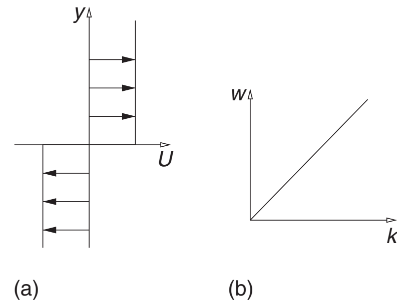



**Figure 5** Vortex sheet: (a) velocity profile and (b) dispersion relation.

horizontal flow with speed $U_o$ moving past a solid wall at rest (Figure 3a). Since in viscous flow the fluid sticks to the wall (the no-slip boundary condition), the fluid velocity at the wall is zero. As a result, there is a thin layer near the wall in which the horizontal velocity varies greatly while the vertical velocity gradients are small, yielding large negative vorticity values $\omega = v_x - u_y$ (Figure 3b). Similarity solutions to the approximating Prandtl boundary-layer equations show that the boundary-layer thickness $d$ grows proportional to $\sqrt{t}$, where $t$ measures the time from the beginning of the motion. Boundary layers can separate from the wall at corners or regions of high curvature and move into the fluid interior, as illustrated in several of the following examples.

## Sample Vortex Flows

### Shear Layers

A shear layer is a thin region of concentrated vorticity across which the tangential velocity component varies greatly. An example is the constant-vorticity layer given by parallel 2D flow $u(x, y) = U(y)$, $v(x, y) = 0$, where $U$ is as shown in Figure 4a. In this case, the velocity is constant outside the layer and linear inside. The vorticity $\omega = -U'(y)$ is zero outside the layer and constant inside. Shear layers occur naturally in the ocean or atmosphere when regions of distinct temperature or density meet. To illustrate this scenario, consider a tank containing two horizontal layers of fluids of different densities, one on top of the other. If the tank is tilted, the heavier bottom fluid moves downstream, and the lighter one moves upstream, creating a shear layer.

Flat shear layers are unstable to perturbations: they do not remain flat but roll up into a sequence of vortices. This is the Kelvin–Helmholtz instability, which can be deduced analytically using linear stability analysis. One shows that in a periodically perturbed flat shear layer, the amplitude of a perturbation with wave number $k$ will initially grow exponentially in time as $e^{wt}$, where $w = w(k)$ is the dispersion relation, leading to instability. The wave number of largest growth depends on the layer thickness. This is illustrated in Figure 4b, which plots $w(k)$ for a constant-vorticity layer of thickness $2d$. The wave number of maximal growth is proportional to $1/d$.

A vortex sheet is a model for a shear layer. The layer is approximated by a surface of zero thickness across which the tangential velocity is discontinuous, as illustrated in Figure 5a. In this case, the dispersion relation reduces to $w(k) = \pm k$. That is, for each wave number $k$ there is a growing and a decaying mode, and the growing mode grows faster the higher the wave number is, as shown in Figure 5b. The vortex sheet arises from a constant vorticity shear layer as the thickness $d \to 0$ and the vorticity $\omega \to \infty$ in such a way that the product $\omega d$ remains constant. Figure 6 shows the roll-up of a periodically perturbed vortex sheet due to the

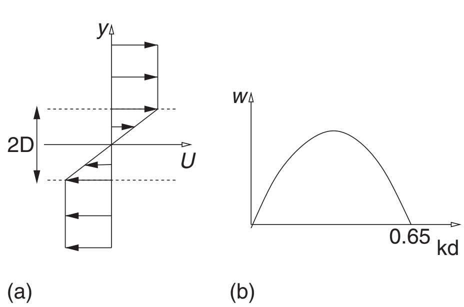



**Figure 4** Shear layer: (a) velocity profile and (b) dispersion relation.

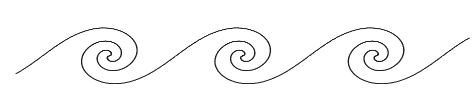


**Figure 6** Computation of vortex sheet roll-up.

Kelvin–Helmholtz instability, computed using one of the methods described in the next section.

### Aircraft Trailing Vortices

One can often observe trailing vortices that shed from the wings of a flying aircraft (also called contrails). These vortices are formed because the wing develops lift. The pressure on the top of the wing is lower than on bottom, causing air to move around the edge of the wing from the bottom surface to the top. The boundary layer on the wing separates as a shear layer that rolls up into a vortex attached to the tip of the wing (Figure 7). Since the velocity inside the vortex is high, the pressure is correspondingly low and causes water vapor in the air to condense, forming water droplets that visualize the vortices. The vortex strength increases with increasing lift, and is particularly strong in high-lift conditions such as take-off and landing. Since lift is proportional to weight, it also increases with the size of the airplane. Vortices of large planes are strong enough to flip a small one if it gets too close. Trailing vortices are the principal reason for the time delay between take-off and landing and are still a serious issue for crowded urban airports.

The trailing vortices can be modeled by a pair of counter-rotating vortex lines (Figure 8a). Two parallel vortex lines of opposite strength induce a downward motion on each other, similar to two point vortices, the zero-core limit. Two point vortices of strength $\pm\Gamma$ at a distance $2d$ from each other translate with self-induced velocity (Figure 9):

$$U = \frac{\Gamma}{4\pi d} \qquad [8]$$

As a result trailing vortices near takeoff hit the ground as a strong downwash air current.

Vortex decay results generally from the development of instabilities. Two parallel vortex tubes are subject to the long-wavelength Crow instability. Triggered by turbulence in the surrounding air, or by local variations in air temperature or density, the vortices develop symmetric sinusoidal perturbations with long wavelength, of the same order as the vortex separation (Figure 8b). As the perturbations grow to finite amplitude, the tubes reconnect and produce a sequence of vortex rings. Note that the two-dimensional schematic in Figure 8c does not convey the three-dimensional structure of the rings. The reconnection process destroys the initial wake structure more rapidly than viscous decay of the individual filaments.

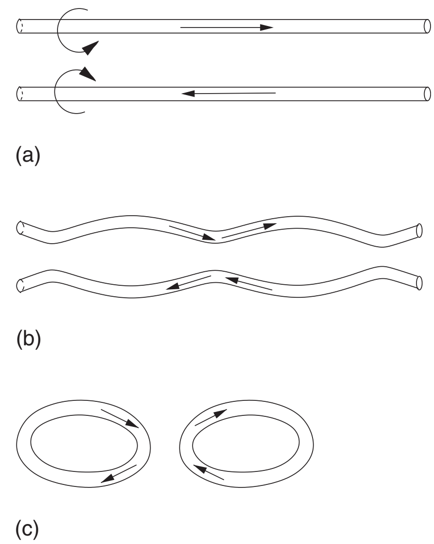



**Figure 8** Sketch. Onset of Crow instability in a pair of vortex lines and ensuing reconnection.

Of much interest is the study of how to accelerate the vortex decay. High-aspect-ratio vortices are subject to a shorter-wavelength elliptic instability, which leads to earlier destruction. However, such vortices are not realistic in current aircraft wakes. Wing designs have been proposed in which more than two trailing vortices form which interact strongly and lead to faster decay. Other interesting aspects are the effect of ambient turbulence and vortex breakdown. Breakdown refers to a disturbance in the vortex core in which it quickly, within an axial distance of few core diameters, develops a region of reversed flow and loses its laminar behavior.

Unlike the counter-rotating vortices discussed so far, two equally signed vortices rotate under their self-induced velocity about a common axis. If the separation distance between them is too small, two equally signed patches merge into one. Vortex merging occurs in two- or three-dimensional flows, as opposed to vortex reconnection, which is a strictly three-dimensional phenomenon.

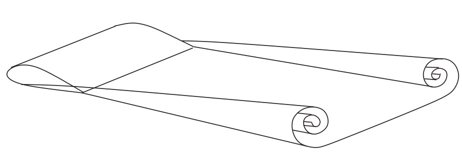


**Figure 7** Sketch. Shear layer separation and roll-up into trailing vortices behind an airfoil.

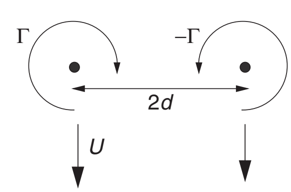



**Figure 9** Self-induced downward motion of a vortex pair.

### Vortex Rings

A vortex tube that forms a closed loop is called a vortex ring. Vortex rings can be formed by ejecting fluid from a circular opening, such as when a smoke ring is formed. The boundary layer wall vorticity separates at the opening as a cylindrical shear layer that rolls up at its edge into a ring (Figure 10). The vorticity is concentrated in a core, which may be thin or thick relative to the ring diameter. The limiting cases are an infinitely thin circular filament of nonzero circulation and the Hill's vortex, in which the vorticity occupies all the interior of a sphere.

Just as a counter-rotating vortex pair, a ring translates under its self-induced velocity $U$ in direction normal to the plane of the ring (Figure 11). However, unlike the vortex pair, the ring velocity depends significantly on its core thickness. For a ring with radius, circulation and core size, respectively, $R$, $\Gamma$, $a$, the self-induced velocity is

$$U \sim \frac{\Gamma}{4\pi R}\left(\log\frac{8R}{a} - \frac{1}{4}\right) \qquad [9]$$

asymptotically as $a \to 0$. Thus, the translation velocity becomes unbounded for rings with decreasing core size. In reality, at some point viscosity takes over and spreads the core vorticity, slowing the ring down.

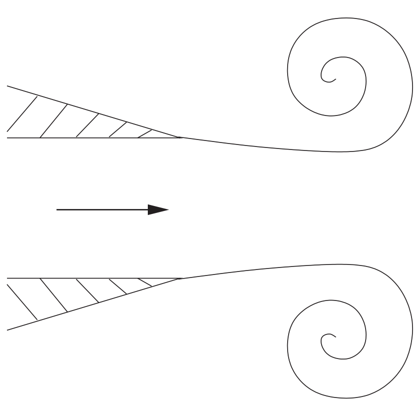


**Figure 10** Vortex ring, formed by ejecting fluid from a circular tube.

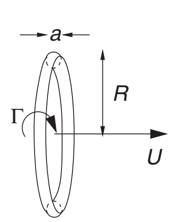



**Figure 11** Self-induced motion of a vortex ring.

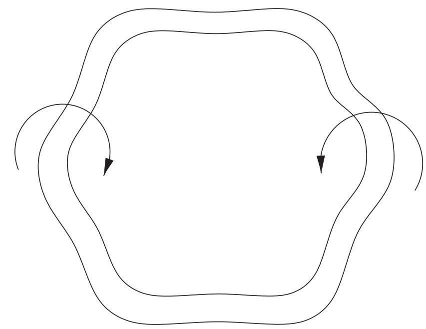


**Figure 12** Sketch. Onset of azimuthal vortex ring instability.

Vortex rings of small cross section are subject to an azimuthal instability. Theory, experiment, and simulations show that if a ring is perturbed in the azimuthal direction, there exists a dominant wave number which is unstable and grows (Figure 12). The unstable wave number increases as the core size decreases, while its spatial amplification rate is almost independent of the core size.

Interesting dynamics are obtained when two or more rings interact. Two coaxial vortex rings of equally signed circulation move in the same direction and exhibit leap-frogging: the rear ring causes the front ring to grow in radius and the front ring causes the rear one to decrease. From eqn [9] it can be seen that the ring velocity is inversely proportional to its radius. Consequently, the front ring slows down and the rear ring speeds up, until the rear ring travels through the front ring. This process repeats itself and is known as leap-frogging. On the other hand, two coaxial vortex rings of oppositely signed circulation approach each other and grow in radius. Their cores contract in order to preserve volume, and their vorticity increases in order to preserve circulation. Under certain experimental conditions, the azimuthal instability develops, the resulting waves on opposite rings reconnect and a sequence of smaller rings form.

### Vortices, Mixing, and Chaos

Mixing is important in many natural processes and technological applications. For example, mixing in shear flows and wakes is relevant to aeronautics and combustion, mixing and diffusion determine chemical reaction rates, and mixing of contaminants pollutes oceans and atmosphere. It is therefore important to understand and control mixing processes.

Efficient mixing of two fluids is obtained by efficient stretching and folding of material lines.

Stretching and folding in turn are the fingerprint of chaos; thus, mixing and chaos are intimately related. Mixing and associated chaotic fluid motion can be obtained by simple vortical motion. For example, two counter-rotating vortices subject to a periodic strain field oscillate in a regular fashion but induce chaos in a region of fluid moving with them. Similarly, two corotating vortices of equal strength that are turned on and off periodically so that one is on when the other is off, known as the blinking vortices, rotate around a common axis in a stepwise manner but induce chaos in nearby regions. On the other hand, if there are four or more vortices present, the vortex motion itself is generally chaotic. It should be noted that there are also nonchaotic equilibrium solutions of four or more vortices forming what is called a vortex crystal.

Information about chaotic particle motion is obtained by studying Poincaré sections, examining the associated stable and unstable manifolds, and investigating the existence of chaotic maps such as the horseshoe map.

### Atmospheric Vortices

Atmospheric vortices are driven by temperature gradients, Earth's rotation (Coriolis force), spatial landscape variations, and instabilities. For example, temperature differences between the equator and the poles and Earth's rotation lead to large-scale vortices such as the trade winds (Hadley cell), the jet streams, and the polar vortex (Figure 13). Semi-annual temperature oscillations are responsible for the Indian monsoons. Daily oscillations cause land- and sea-breezes. Landscape variations can cause urban–rural wind flows and mountain–valley circulations.

Instabilities are often responsible for large cyclonic vortices. Barotropic instability results from large horizontal velocity gradients, and has been deemed responsible for disturbances over the Sahara region that occasionally intensify into tropical cyclones. Baroclinic instability, which occurs when temperature advection is superposed on a velocity field, can lead to cyclonic vortices at the front between air of polar origin and that of tropical origin. The inertial or centrifugal instability occurs when air flows around high-pressure systems and the pressure gradient force is not large enough to balance the centripetal acceleration and the Coriolis effect.

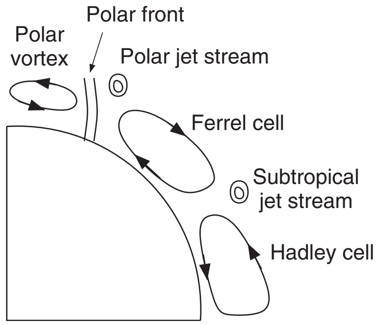



**Figure 13** Vortices in the atmosphere.

Vortices also form on other planets with an atmosphere. On Mars, dust devils are quite common. They are ~10–50 times larger than the ones on Earth and can carry high-voltage electric fields caused by the rubbing of dust grains against each other. Jupiter's characteristic spots are extremely large storm vortices. The Great Red Spot is a vortex spanning twice the diameter of the Earth. Unlike the low-pressure terrestrial storms and hurricanes, the Great Red Spot is a high-pressure system that has been stable for more than 300 years. Other vortices on Jupiter decay and vanish, such as the White Ovals, three large anticyclones which merged into one within two years. Recent computer simulations predict that many of Jupiter's vortices will merge and disappear in the next decade. As a result, mixing of heat across zones will decay and the planet's temperature is predicted to increase.

Numerical simulations of the atmosphere are expensive due to the large number of parameters and the relatively small scales that need to be resolved. For climate models and medium-range forecast models, the governing 3D compressible Euler equations are simplified using the hydrostatic approximation (in which only the pressure gradient and the gravitational forces are retained in the vertical-momentum equation) and the anelastic approximation (in which $d\rho/dt$ is neglected), to obtain the primitive equations. Additional vertical averaging yields the shallow-water equations. One big hurdle is to accurately incorporate the effect of clouds, which is significant and is usually treated using subgrid models.

### Vortices in Superfluids and Superconductors

At temperatures below 2.2 K, liquid helium is a superfluid, meaning that it acts essentially like a fluid with zero viscosity governed by the Euler equations. The fluid is irrotational, except for extremely thin vortex filaments, which are formed by quantum-mechanical processes. Since the vortices cannot end in the interior of the flow, they can be generated only at the surface or they nucleate as

vortex rings inside the fluid. As an example, if a cylindrical container with helium is rotated sufficiently fast, vortex lines attached to both ends of the container appear. These quantum vortices have discrete values of circulation ($= nh/m$, where $h =$ Planck's constant, $m =$ mass of helium atom, $n =$ integer), core sizes of about 1 Å (roughly the diameter of a single hydrogen atom) and move without viscosity.

Similarly, certain types of materials lose their electric resistance at low temperatures and become superconductors. One distinguishes type-I superconductors (most pure metals) from type-II superconductors (alloys). Using the Ginzburg–Landau theory it has been predicted that in type-II superconductors a lattice of vortex filaments forms, each carrying a quantized amount of magnetic flux. This was subsequently confirmed by experimental observation. More precisely, for temperatures $T$ below a critical value $T_c$, there are three regions corresponding to increasing values of the magnetic field (Figure 14). At low magnetic fields ($H < H_{c1}$), no vortices exist (superconducting phase). At intermediate values ($H_{c1} < H < H_{c2}$), the magnetic field penetrates the superconductor in the form of quantized vortices, also called flux lines (mixed phase). The values $H_{c1,c2}$ are determined by the London penetration depth $\lambda$, which measures the electromagnetic response of the superconductor. With increasing magnetic field, the density of flux lines increases until the vortex cores overlap when the upper critical field $H_{c2}$ is reached, beyond which one recovers the normal metallic state (normal conductor).

When an external current density $\boldsymbol{j}$ is applied to the vortex system, the flux lines start to move under the action of the Lorentz force. As a result, a dissipating electric field $\boldsymbol{E}$ appears that is parallel to $\boldsymbol{j}$, and the superconducting property of dissipation-free current flow is lost. In order to recover the desired property of dissipation-free flow, flux lines have to be pinned, for example, by introducing inhomogeneities and structural defects. For a given pinning force, flux lines remain pinned as long as the current density stays below a critical value. A major research objective is to optimize the pinning force in order to preserve superconductivity at larger current densities.

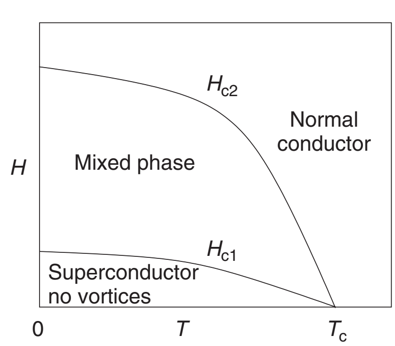



**Figure 14** Superconductor phase dependence on magnetic field $H$ and temperature $T$.

## Numerical Vortex Methods

Many numerical methods used to compute fluid flow are Eulerian schemes based on a fixed mesh, such as finite difference, finite element, and spectral methods, commonly used for example in atmosphere and ocean modeling. This section briefly describes alternative vorticity-tracking methods used to simulate incompressible inviscid vortex flows, and concludes with some extensions to viscous flows. The premise of these methods is that since the fluid velocity is determined by the vorticity through the Biot–Savart law (eqn [6]), it suffices to track only that portion of the fluid carrying nonzero vorticity. This region is often much smaller than the total fluid volume, and computational efficiency is gained. Numerical vortex methods are typically Lagrangian, that is, the computational elements move with the fluid velocity.

### Point-Vortex Approximation in 2D

To compute the evolution of a vorticity distribution $\omega(\boldsymbol{x}, t)$ in 2D, the simplest approach is to approximate the vorticity by a set of point vortices at $\boldsymbol{x}_j(t)$ with circulation $\Gamma_j$ and evolve them under their self-induced motion. The values $\Gamma_j$ are an estimate of the initial circulation around $\boldsymbol{x}_j(0)$. The vortex positions $\boldsymbol{x}_j(t)$ evolve in the induced velocity field

$$\frac{\mathrm{d}\boldsymbol{x}_j}{\mathrm{d}t} = \sum_{\substack{k=1\\k\neq j}}^{N} \Gamma_k \boldsymbol{K}_{2\mathrm{D}}(\boldsymbol{x}_j - \boldsymbol{x}_k) \qquad [10]$$

where the exclusion $k \neq j$ accounts for the fact that a point vortex induces zero velocity on itself. The solution to the system of ordinary differential equations [10] can be obtained using any method, such as Runge–Kutta or Adams–Bashforth.

The point-vortex approximation can be written in Hamiltonian form as

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = -\frac{1}{\Gamma_j}\frac{\partial H}{\partial y_j}, \qquad \frac{\mathrm{d}y_j}{\mathrm{d}t} = \frac{1}{\Gamma_j}\frac{\partial H}{\partial x_j} \qquad [11]$$

where the Hamiltonian

$$H(x,y) = \frac{1}{4\pi}\sum_{j=1}^{N}\sum_{\substack{k=1\\k>j}}^{N}\Gamma_j\Gamma_k \log\left[(x_j - x_k)^2 + (y_j - y_k)^2\right] \qquad [12]$$

is conserved along fluid particles, $\mathrm{d}H/\mathrm{d}t = 0$. The method also conserves the fluid circulation and the linear and angular momenta.

Ideally, the solution to [10] should converge as $N \to \infty$ to the solution of the Euler equations. This is true for smooth vorticity distributions, but for singular distributions such as a vortex sheet, the situation is more complicated. The vortex sheet, a curve in the plane, develops a singularity in finite time at which the curvature becomes unbounded at a point. The point-vortex approximation converges before the singularity formation time, provided the growth of spurious roundoff error due to Kelvin–Helmholtz instability is suppressed using a filter. However, past the singularity formation time, the point-vortex approximation no longer converges.

The general approach is to replace the singular kernel $\boldsymbol{K}_{2\mathrm{D}}$ by a regularization $\boldsymbol{K}_{2\mathrm{D}}^{\delta}$, such as

$$K_{2\mathrm{D}}^{\delta} = \frac{1}{2\pi}\frac{-y\boldsymbol{i} + x\boldsymbol{j}}{|\boldsymbol{x}|^2 + \delta^2} \qquad [13a]$$

$$K_{2\mathrm{D}}^{\delta} = \frac{1}{2\pi}\frac{-y\boldsymbol{i} + x\boldsymbol{j}}{|\boldsymbol{x}|^2}\left(1 - \mathrm{e}^{-|\boldsymbol{x}|^2/\delta^2}\right) \qquad [13b]$$

where $\delta$ is a numerical parameter. The regularization amounts to replacing the $\delta$-function vorticity of a point vortex by an approximate $\delta$-function. In order to recover the solution to the Euler equations, it is necessary to study the limit $N \to \infty, \delta \to 0$. For smooth vorticity distributions, this process converges. For vortex sheet initial data, there is evidence of convergence, but details of the limiting behavior remain under investigation. Regularized solutions with fixed value $\delta$ and vortex sheet initial data are shown in **Figures 6** and **15**. **Figure 6** shows the onset of the Kelvin–Helmholtz instability in a periodically perturbed flat vortex sheet. **Figure 15** shows the rollup of an elliptically loaded flat vortex sheet that models the evolution of an aircraft wake (see **Figure 7**). The correspondence between the two-dimensional simulation and the three-dimensional wake is made by replacing the spatial coordinate in the aircraft's line of flight by a time coordinate.

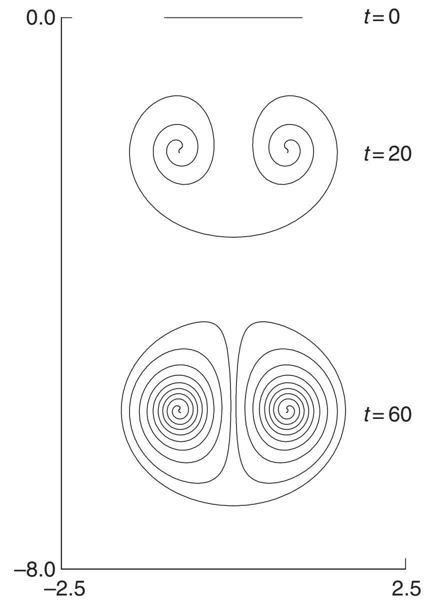



**Figure 15** Computed evolution of an elliptically loaded flat vortex sheet.

### Contour Dynamics in 2D

Consider a planar patch of constant vorticity $\omega_o$ bounded by a curve $\boldsymbol{x}(s,t), 0 \le s \le L$, moving in inviscid, incompressible flow. In view of Kelvin's theorem and eqn [5], the vorticity in the patch remains constant and equal to $\omega_o$ for all time, and the patch area remains constant. Only the patch boundary moves. The velocity at a point $\boldsymbol{x}(\alpha, t)$ on the boundary can be written as a line integral over the boundary:

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = -\frac{\omega_o}{2\pi}\int_C \log|\boldsymbol{x} - \boldsymbol{x}(s,t)|\frac{\partial \boldsymbol{x}}{\partial s}\mathrm{d}s \qquad [14]$$

The contour dynamics method consists of approximating a given vorticity distribution by a superposition of vortex patches, and moving their boundaries according to eqn [14]. This method has been applied to compute the evolution of single-vortex patches and shear layers, and to geophysical flows. Typically, filamentation occurs: the patch develops thin filaments which increase the boundary length significantly and thereby the computational expense. The approach generally taken is to remove the thin filaments at several times throughout the computation, which is referred to as contour surgery. The contour dynamics approach as well as the point-vortex approximation have also been generalized to treat quasigeostrophic flows.

### Vortex Filament Methods in 3D

Vortex simulations in 3D differ from those in 2D in that the stretching term in eqn [4] needs to be incorporated. The vortex filament method approximates the fluid vorticity by a finite number of filaments whose circulation remains constant in time. Each filament is marked by computational mesh points which move with the regularized induced velocity. The regularization is necessary to prevent the infinite self-induced velocities of curved vortex filaments. As in 2D, this method automatically conserves circulation. Vorticity stretching is accounted for by the stretching between computational mesh points. As the filament length increases, more meshpoints are typically introduced to keep it resolved. Also, the number of filaments can be increased throughout the simulation to maintain resolution.

### Viscous Vortex Methods

While inviscid models are expected to approximate small viscosity fluids well far from boundaries, near boundaries, where vortex shedding is an inherently viscous mechanism, it is important to incorporate the effects of viscosity. The first methods to do so used operator splitting in which inviscid and viscous terms of the Navier–Stokes equations were solved in a sequential manner. In each time step, the computational elements would first be convected, and then they would be diffused by a random-walk scheme. The particle strength exchange method, introduced more recently, does not rely on operator splitting and has better accuracy. The particle position and vorticity evolve simultaneously, and viscous diffusion is accounted for in a consistent manner.

Vortex dynamics continues to be a source of interesting problems of theoretical and practical importance. In particular, much remains to be learned to better understand turbulence and the transition to turbulence, a process dominated by deterministic vortex dynamics.

## Further Remarks

Finally, some remarks on relevant literature on this subject are in order. Lugt (1983) and Tritton (1988) are recommended as elementary introduction to vortex flows. van Dyke (1982) presents beautiful and instructive flow visualizations. Comprehensive treatments of incompressible fluid dynamics are given in Batchelor (1967), Chorin and Marsden (1992), Lamb (1932), and Saffman (1992), and compressible flow is treated in Anderson (1990). Cottet and Koumoutsakos (2000) give an overview of numerical vortex methods.

Special topics have also been addressed; atmosphere (Andrews *et al.* 1987), point vortex motion and chaos (Aref 1983, Newton 2001, Ottino 1989), superfluids and superconductors (Blatter *et al.* 1994, Donnelly 1991), turbulence theory using statistical mechanics ( Chorin 1994), vortex reconnection (Kida and Takaoka 1994), theory for Euler and Navier–Stokes equations (Majda and Bertozzi 2002), contour dynamics (Pullin 1992), vortex rings (Shariff and Leonard 1992), and aircraft trailing vortices (Spalart 1998). Green (1995) includes survey articles on various topics.

## Nomenclature

| | |
|---|---|
| $a$ | vortex ring core size |
| $\boldsymbol{g}$ | gravitational field |
| $H$ | Hamiltonian |
| $K_{2D}$ | singular velocity kernel |
| $K_{2D,\delta}$ | regularized velocity kernel |
| $\rho(\boldsymbol{x}, t)$ | fluid density |
| $R$ | vortex ring radius |
| $T(\boldsymbol{x}, t)$ | temperature |
| $U$ | translation velocity |
| $\boldsymbol{u}(\boldsymbol{x}, t) = u(\boldsymbol{x}, t)\boldsymbol{i} + v(\boldsymbol{x}, t)\boldsymbol{j} + w(\boldsymbol{x}, t)\boldsymbol{k}$ | fluid velocity |
| $w(k)$ | dispersion relation |
| $\boldsymbol{\omega} = \nabla \times \boldsymbol{u}$ | vorticity |
| $\omega = v_x - u_y$ | scalar vorticity |
| $\Gamma$ | ring circulation |

*See also:* Abelian Higgs Vortices; Incompressible Euler Equations: Mathematical Theory; Integrable Systems: Overview; Interfaces and Multicomponent Fluids; Intermittency in Turbulence; Newtonian Fluids and Thermohydraulics; Point-Vortex Dynamics; Stochastic Hydrodynamics; Superfluids; Topological Knot Theory and Macroscopic Physics; Turbulence Theories.

## Further Reading


Anderson JD (1990) *Modern Compressible Flow with Historical Perspective*, 2nd edn. New York: McGraw-Hill.

Andrews DG, Holton JR, and Leovy CB (1987) *Middle Atmosphere Dynamics*. Orlando: Academic Press.

Aref H (1983) Integrable, chaotic, and turbulent vortex motion in two-dimensional flows. *Annual Review of Fluid Mechanics* 15: 345–389.

Batchelor GK (1967) *An Introduction to Fluid Dynamics*. Cambridge: Cambridge University Press.

Blatter G, Feigel'man MV, Geshkenbein VB, Larkin AI, and Vinokur VM (1994) Vortices in high-temperature superconductors. *Reviews of Modern Physics* 66(4): 1125–1388.

Chorin AJ (1994) *Vorticity and Turbulence*. New York: Springer.

Chorin AJ and Marsden JE (1992) *A Mathematical Introduction to Fluid Mechanics*, 3rd edn. New York: Springer.

Cottet G-H and Koumoutsakos PD (2000) *Vortex Methods: Theory and Practice*. Cambridge: Cambridge University Press.

Donnelly RJ (1991) *Quantized Vortices in Helium II*. Cambridge: Cambridge University Press.
Green SI (ed.) (1995) *Fluid Vortices*. Dordrecht: Kluwer Academic.
Kida S and Takaoka M (1994) Vortex reconnection *Annual Review of Fluid Mechanics* 26: 169–189.
Lamb H (1932) *Hydrodynamics*, 6th edn. New York: Dover.
Lugt HJ (1983) *Vortex Flow in Nature and Technology*. New York: Wiley.
Majda AJ and Bertozzi AL (2002) *Vorticity and Incompressible Flow*. Cambridge: Cambridge University Press.
Newton PK (2001) *The N-Vortex Problem: Analytical Techniques*. New York: Springer.
Pullin DI (1992) Contour dynamics methods. *Annual Review of Fluid Mechanics* 24: 89–115.
Ottino JM (1989) *The Kinematics of Mixing: Stretching, Chaos, and Transport*. Cambridge: Cambridge University Press.
Saffman PG (1992) *Vortex Dynamics*. Cambridge: Cambridge University Press.
Shariff K and Leonard A (1992) Vortex rings. *Annual Review of Fluid Mechanics* 24: 235–279.
Spalart PR (1998) Airplane trailing vortices. *Annual Review of Fluid Mechanics* 30: 107–138.
Tritton DJ (1988) *Physical Fluid Dynamics*, 2nd edn. Oxford: Clarendon Press.
van Dyke M (1982) *Album of Fluid Motion*. Stanford: The Parabolic Press.

**Vortices** *see* Abelian Higgs Vortices; Point-Vortex Dynamics

# W

## Wave Equations and Diffraction

**M E Taylor**, University of North Carolina, Chapel Hill, NC, USA



### Introduction

The most basic wave equation is

$$\frac{\partial^2 u}{\partial t^2} - \Delta u = 0 \qquad [1]$$

for $u = u(t, x)$, where $\Delta$ is the Laplace operator, given by $\Delta u = \partial^2 u/\partial x_1^2 + \cdots + \partial^2 u/\partial x_n^2$ on $n$-dimensional Euclidean space $\mathbb{R}^n$. More generally, $u$ might be defined on $\mathbb{R} \times M$, where $\mathbb{R}$ is the $t$-axis and $M$ is a Riemannian manifold, with a metric tensor given in local coordinates by $(g_{jk})$. Then the Laplace–Beltrami operator is given, in local coordinates, by

$$\Delta u = g^{-1/2} \sum_{j,k} \frac{\partial}{\partial x_j}\left(g^{1/2} g^{jk} \frac{\partial u}{\partial x_k}\right) \qquad [2]$$

where $(g^{jk})$ is the matrix inverse to $(g_{jk})$ and $g = \det(g_{jk})$. Even if one concentrates on wave propagation in Euclidean space, one frequently wants to use curvilinear coordinates, and [2] is useful. Equation [1] is supplemented by initial conditions of the form

$$u(0, x) = f(x), \qquad \partial_t u(0, x) = g(x) \qquad [3]$$

called Cauchy data. If the spatial domain $M$ has a boundary $\partial M$ (e.g., if $M$ is a bounded region in $\mathbb{R}^n$), then boundary conditions are imposed. The most common are the Dirichlet boundary condition

$$u(t, x) = 0 \quad \text{for } x \in \partial M \qquad [4]$$

and the Neumann boundary condition

$$\partial_\nu u(t, x) = 0 \quad \text{for } x \in \partial M \qquad [5]$$

where $\partial_\nu u$ denotes the normal derivative of $u$ at the boundary. More generally, one might have a driving force, and replace 0 on the right-hand side of [1] by a function $F(t, x)$. Similarly, one can consider nonzero boundary data in [4] and [5].

The wave equation [1] models a number of physical phenomena, at least in the linear approximation. The vibration of a drum head is modeled by [1], with $M$ a planar domain, and with the Dirichlet boundary condition [4]. The motion of sound waves in a room with hard walls is modeled by [1], with $M$ a region in $\mathbb{R}^3$, and with the Neumann boundary condition [5]. The propagation of electromagnetic waves is given by Maxwell's equations:

$$\begin{aligned} \frac{\partial E}{\partial t} - \operatorname{curl} B &= -J \\ \frac{\partial B}{\partial t} + \operatorname{curl} E &= 0 \\ \operatorname{div} E &= \rho \\ \operatorname{div} B &= 0 \end{aligned} \qquad [6]$$

where $\rho$ is the electric charge density and $J$ the current. These equations yield [1] (with the right-hand side replaced by some function $F(t, x)$ if $J$ and $\rho$ are not zero) for the components of the electric field $E$ and the magnetic field $B$. If the propagation is in a region $M$ in $\mathbb{R}^3$ bounded by a perfect conductor, then the boundary conditions are that $E$ is normal to $\partial M$ and $B$ is tangential to $\partial M$. If $\partial M$ is flat, these equations can be decomposed into Dirichlet problems for some components and Neumann problems for the rest, but if $\partial M$ is curved such a decomposition is not possible.

Other models of vibrating objects produce variants of [1]. Examples include vibrating elastic solids, yielding an equation like [1] with $\Delta u$ replaced by $\mu \Delta u + (\lambda + \mu) \operatorname{grad} \operatorname{div} u$, for linear elasticity. Here $\Delta$ acts componentwise on $u$, and $\mu$ and $\lambda$ are constants, called Lamé constants. Other examples model vibrations of crystals and propagation of electromagnetic waves in crystals. Further interesting phenomena arise in these various cases, such as Rayleigh waves in linear elasticity and conical refraction in crystal optics.

Here we discuss the propagation of waves and their reflection and diffraction at boundaries. In the interest of providing reasonable coverage in a brief space, we restrict attention to the wave equation [1].

## Basic Propagation Phenomena

The simplest examples of waves propagating according to [1] are plane waves, of the form

$$u(t,x) = \varphi(x \cdot \omega - t) \qquad [7]$$

with $(t,x) \in \mathbb{R} \times \mathbb{R}^n, \omega$ a unit vector in $\mathbb{R}^n$, and $\varphi$ a function on $\mathbb{R}$. If $\varphi$ has two continuous derivatives, [7] defines a classical solution of [1]. More generally, one can allow $\varphi$ to be less regular. For example, it could be piecewise smooth with a jump discontinuity at some point $a \in \mathbb{R}$. In such a case, $u$ will be piecewise smooth with a jump across the $n$-dimensional surface $x \cdot \omega - t = a$ in $\mathbb{R} \times \mathbb{R}^n$, which will solve [1] in a weak, or distributional, sense. For fixed $t, u(t,\cdot)$ has a jump across the $(n-1)$-dimensional surface $\Sigma_t = \{x \in \mathbb{R}^n : x \cdot \omega = t + a\}$. As $t$ varies, $\Sigma_t$ moves in the direction $\omega$ with unit speed.

There are also spherical wave solutions to [1] on $\mathbb{R} \times \mathbb{R}^n$, such as

$$u(t,x) = \frac{\operatorname{sgn} t}{2\pi}(t^2 - |x|^2)_+^{-1/2} \qquad [8]$$

for $n = 2$, and

$$u(t,x) = \frac{1}{4\pi t}\delta(|x| - |t|) \qquad [9]$$

for $n = 3$. Here $s_+ = s$ for $s > 0, s_+ = 0$ for $s \le 0$, and $\delta(s)$ is the Dirac delta function. In fact, [8] and [9] are "fundamental solutions" (more on which in the section on harmonic analysis) to the wave equation on $\mathbb{R} \times \mathbb{R}^n$, for $n = 2$ and 3, respectively. In such cases, the singularity in $u(t,\cdot)$ for each fixed $t$ lies in $\Sigma_t = \{x \in \mathbb{R}^n : |x| = |t|\}$, a family of surfaces in $\mathbb{R}^n$ that moves, in the direction of the normal to $\Sigma_t$, at unit speed.

The examples mentioned above illustrate two general phenomena about the behavior of solutions to [1]. The first is finite propagation speed. Its general formulation is that, given a closed set $K \subset M$,

$$\begin{aligned} \operatorname{supp} f, g \subset K \ \Rightarrow\ & \operatorname{supp} u(t,\cdot) \\ & \subset \{x \in M : \operatorname{dist}(x,K) \le |t|\} \end{aligned} \qquad [10]$$

In fact, given that [8]–[9] are fundamental solutions, [10] is a consequence of these formulas when $M = \mathbb{R}^2$ or $\mathbb{R}^3$. The result [10] is true in great generality, with well-known demonstrations involving energy estimates.

The second phenomenon involves propagation of singularities. Typically, if the Cauchy data $f$ and $g$ in [3] are smooth on the complement of an $(n-1)$-dimensional surface $\Sigma_0$, perhaps with a jump across $\Sigma_0$, or such a singularity as in [8] or [9], the solution $u(t,x)$ will be a sum of two terms, with singularities of a similar nature on the surfaces $\Sigma_t^{\pm}$, moving at unit speed in the direction of their normals, $\Sigma_t^+$ flowing from $\Sigma_0$ in one direction and $\Sigma_t^-$ in the other. This also holds for the manifold case [2]. That happens at least until such surfaces develop singularities, when matters become more elaborate.

An alternative way to describe how the set of singularities evolves is the following. Let $S_1M$ denote the space of unit vectors tangent to $M$; this is a submanifold of the tangent bundle of $M, TM$. There is a natural projection $\pi : S_1M \to M$. Associated to a smooth surface $\Sigma_0$ of dimension $n-1$ in $M$ (of dimension $n$) are two preimages $\Lambda_0^+$ and $\Lambda_0^-$ in $S_1M$, consisting of unit vectors lying over points of $\Sigma_0$ and normal to $\Sigma_0$. The geodesic flow is a flow on $S_1M$, and it takes $\Lambda_0^{\pm}$ to smooth $(n-1)$-dimensional surfaces $\Lambda_t^{\pm}$ in $S_1M$. The sets $\Sigma_t^{\pm}$ are the images of $\Lambda_t^{\pm}$ under $\pi$. The geodesics starting out at points in $\Lambda_0^{\pm}$ and sweeping out $\Lambda_t^{\pm}$ are the rays along which the singularities of the solution $u$ propagate.

This latter description works for all $t$ if $M$ has no boundary and is complete, that is, all geodesics are defined for all $t$, although singularities develop in the images $\pi(\Lambda_t^{\pm}) = \Sigma_t^{\pm}$, at points $p \in \Sigma_t^{\pm}$, where $\Lambda_t^{\pm}$ meets $T_pM$ nontransversally. The behavior of $u$ near such singular points of $\Sigma_t^{\pm}$, known as caustics, is more complicated than that near regular points, but it can be captured in terms of integrals. Methods of establishing this propagation of singularities are discussed in the section on geometrical optics.

Such a description needs further elaboration if $M$ has a boundary. One of the principal problems of diffraction theory is to explain how singularities of solutions to [1], with a boundary condition such as [4] or [5], propagate and reflect off the boundary.

Considering the case where $M$ is a half-space in $\mathbb{R}^n$,

$$M = \mathbb{R}_+^n = \{x \in \mathbb{R}^n : x_n \ge 0\} \qquad [11]$$

provides a guide to the simplest reflection phenomena. In such a case, one can solve the Dirichlet or Neumann boundary problem for the wave equation [1] by the method of images. One extends $f$ and $g$ from $\mathbb{R}_+^n$ to $\mathbb{R}^n$. For the Dirichlet problem [4], one takes odd extensions, $f(x', -x_n) = -f(x', x_n)$, and similarly for $g$. For the Neumann problem [5], one takes even extensions, $f(x', -x_n) = f(x', x_n)$, etc. One then solves the wave equation [1] on $\mathbb{R} \times \mathbb{R}^n$ with the extended Cauchy data, and the restriction to $\mathbb{R} \times \mathbb{R}_+^n$ solves the respective boundary problem. Suppose $\Sigma_0$ is a smooth $(n-1)$-dimensional surface that does not meet $\partial\mathbb{R}_+^n$, and that $f$ and $g$ have singularities on $\Sigma_0$, as above. (Suppose for simplicity that $f$ and $g$ vanish near $\partial\mathbb{R}_+^n$.) Those rays issuing

from normals to $\Sigma_0$ have mirror images, which are rays in $\mathbb{R}^n_-$. If such a ray hits $\partial\mathbb{R}^n_+$, its mirror image does so also, and continues into $\mathbb{R}^n_+$, as the reflected ray. The singularities of $u$ propagate along such reflected rays.

Such a description extends to a general complete Riemannian manifold with boundary $M$, in the case of rays that hit the boundary transversally. Such a ray is reflected by retaining the tangential component of its velocity vector at the point of intersection $\partial M$ and reversing the sign of the normal component. One says that the ray is reflected according to the laws of geometrical optics. Singularities of $u$ carried by such rays that hit $\partial M$ are correspondingly reflected. Methods to establish such transversal reflection of singularities are natural extensions of those developed to treat the propagation away from $\partial M$, mentioned above.

Matters become more delicate when there are rays that are tangent to $\partial M$. A model example is given by

$$M = \mathbb{R}^n \setminus B, \quad B = \{x \in \mathbb{R}^n : |x| < 1\} \qquad [12]$$

which one takes when studying the scattering of waves in $\mathbb{R}^n$ by the obstacle $B$. Consider a solution to [1] with boundary condition given by [4] or [5] that has a simple singularity on $\Sigma_t = \{x \in \mathbb{R}^n : x_n = t\}$ for $t < -1$. The associated rays are of the form $\gamma_{x'}(t) = (x', t)$, for $t < -1$, with $x' \in \mathbb{R}^{n-1}$. If $|x'| > 1$, these rays continue on in $\mathbb{R}^n \setminus B$, for all $t \geq -1$. If $|x'| < 1$, these rays hit $\partial M = \partial B$ transversally, and their reflection is as described above. If $|x'| = 1$, these rays hit $\partial B$ tangentially, at $t = 0$; they are sometimes called grazing rays. One also continues them past $t = 0$. One defines in this fashion $\Sigma_t$ for $t \geq -1$. The region

$$\mathcal{S} = \{x = (x', x_n) \in \mathbb{R}^n \setminus B : |x'| < 1, x_n > 0\} \qquad [13]$$

is called the "shadow region." It is disjoint from $\Sigma_t$ for all $t$. The solution $u$ is smooth in $\mathcal{S}$ for all $t$, although it is not identically zero. The set

$$\mathcal{S}^b = \{x = (x', x_n) \in \mathbb{R}^n \setminus B : |x'| = 1, \, x_n \geq 0\} \qquad [14]$$

is the "shadow boundary."

One can replace $B$ in [12] by a more general smooth, convex obstacle $K$, with positive Gauss curvature everywhere, and the same considerations of transversal and grazing rays and shadow regions apply. These notions also extend to a more general class of Riemannian manifolds with boundary, called manifolds with diffractive boundary. In the case $K = B$, one can use separation of variables to reduce the problem of analyzing solutions to [1] and showing that singularities propagate along such rays to a problem in harmonic analysis on the sphere $S^{n-1}$. For more general convex obstacles $K$ or manifolds with diffractive boundary, other techniques are required, to show that waves reflect off the boundary in a fashion similar to the case [12].

Another situation arises if instead of [12] one takes $M = B$, or more generally $M = K$, a convex region as described above. A ray starting off from a point in $\partial M$, almost tangent to $\partial M$ but with a small component in the direction of the normal pointing into $M$, will undergo many reflections in a short time. Upon shrinking the normal component of the initial velocity to zero, one obtains in the limit a geodesic in $\partial M$, known as a gliding ray. In such a case, singularities of solutions to [1], with such a boundary condition as [4] or [5], propagate along both transversally reflected and gliding rays.

For the generic smooth obstacle $K$ in $\mathbb{R}^n$, the second fundamental form can have a variety of signatures at various boundary points. Various types of "generalized rays" occur – generally speaking limits of sequences of transversally reflected rays. This situation also holds for general complete Riemannian manifolds with smooth boundary. The main result about propagation of singularities in such a case is that it is always along such generalized rays. This was established by Melrose and Sjöstrand (1978).

Further diffraction effects arise when $\partial M$ has singularities, such as edges and corners. The simplest example is

$$M = \{x \in \mathbb{R}^2 : a \leq \theta \leq b, r \geq 0\} \qquad [15]$$

where $(r, \theta)$ are the polar coordinates of $x \in \mathbb{R}^2$, and we assume $0 \leq a < b \leq 2\pi$. Here one is studying the diffraction of waves by a wedge. In the limiting case $a = 0$, $b = 2\pi$, the wedge becomes a half-line, that is,

$$M = \mathbb{R}^2 \setminus \{(x_1, 0) : x_1 > 0\} \qquad [16]$$

Singularities of solutions to [1] on $\mathbb{R} \times M$ with such a boundary condition as [4] or [5] propagate in the interior of $M$ and reflect off the regular points of $\partial M$ as before. If a family of continuous, piecewise smooth curves $\Sigma_t$ carrying the singularity of $u$ hit the corner $x = 0$ at $t = a$, this reflection creates a tear in $\Sigma_t$ for $t > a$. In addition, a diffracted wave spreads out from the corner at unit speed. This diffracted wave carries a singularity that is weaker than that of the incident wave. For example, if one has a solution like [8], but shifted to have support in a disk of radius $|t|$ about a point $p \neq 0$ in $\mathbb{R}^2$, for small $|t|$, then the diffracted wave will have a jump discontinuity.

The space $M$ in [15] is a special case of a cone. More generally, if $N$ is a complete Riemannian

manifold (possibly with boundary), then the cone $C(N)$ with base $N$ is the set

$$C(N) = [0, \infty) \times N \qquad [17]$$

with all points $(0, x)$, $x \in N$, identified, with the metric tensor

$$ds^2 = dr^2 + r^2 g \qquad [18]$$

where $g$ is the metric tensor on $N$, and points on $C(N)$ are denoted $(r, x), r \in [0, \infty), x \in N$. The space in [14] has the form $M = C(N)$ with $N = [a, b]$, an interval. A cone in Euclidean space $\mathbb{R}^n$ is of the form $C(N)$ with $N$ a domain in the unit sphere $S^{n-1}$.

The propagation of singularities for solutions to [1] on $C(N)$, when $N$ has smooth boundary, has a description similar to that above for the case [15]. Again, there is a diffracted wave set off from the conic point $\{r = 0\}$ when a singularity of a wave hits it. The diffracted wave is typically $(n - 1)/2$ units smoother than the singular wave producing it, where $n = \dim C(N)$. For example, the fundamental solution to the wave equation on $C(N)$ produces a diffracted wave which is the sum of a jump discontinuity and (in general) a logarithmic singularity.

In fact, precise understanding of the behavior of the fundamental solution to the wave equation on $C(N)$ is encoded in terms of the behavior of the solution operator to the wave equation on the base $N$. This is discussed in further detail in the section on harmonic analysis. In the case where $C(N)$ is given by [15], we are dealing with the wave equation on an interval $[a, b]$, whose behavior is elementary.

One can use analysis of [15] together with finite propagation speed to get a good qualitative picture of diffraction of waves in $\mathbb{R}^2$ by a polygonal obstacle. A variation of this argument allows one to understand the behavior of the wave equation on a "polygonal" domain $N$ in $S^2$, that is, one whose boundary consists of a finite number of geodesic segments in $S^2$. Going from there to $C(N)$, one can then analyze diffraction of waves in $\mathbb{R}^3$ by a polyhedron.

It is worth remarking how the "shadow region" for such an obstacle as a wedge in $\mathbb{R}^2$ differs from that in [12]–[14]. For example, if one considers $M$ given by [16] and $u(t, x) = \delta(x_2 - t)$, for $t < 0$, then the region

$$\mathcal{S} = \{x = (x_1, x_2) : x_1, x_2 > 0\} \qquad [19]$$

is the "shadow region," in the sense that rays either missing or reflecting off the obstacle $\{(x_1, 0) : x_1 > 0\}$ do not enter the region [19]. However, unlike the case [13], the solution $u(t, x)$ is not smooth in the region [19] for $t > 0$. There is a singularity there, although it is weaker than the singularity of the main wave.

Taking Cartesian products of spaces of the form [15] with $\mathbb{R}^k$ yields spaces with $k$-dimensional edges. There are also spaces with curvy edges. Rather than continuing with further general description, one more particular case is discussed next, which has had a historical significance. Namely, we consider the reflection of waves in $\mathbb{R}^3$ off a disk, that is, take

$$M = \mathbb{R}^3 \setminus D, \quad D = \left\{(x_1, x_2, 0) : x_1^2 + x_2^2 \le 1\right\} \qquad [20]$$

Consider a wave given for $t < 0$ by $u(t, x) = \delta(x_3 - t)$. This wave hits $D = \partial M$ at $t = 0$, giving off a diffracted wave, traveling away from $\sigma = \{(x_1, x_2, 0) : x_1^2 + x_2^2 = 1\}$ at speed 1 for $t > 0$. This diffracted wave carries a singularity that blows up like the $-1/2$ power of the distance to the torus of points of distance $t$ from $\sigma$, for $t \in (0, 1)$. For $t > 1$, there is a focusing effect along the $x_3$-axis, producing a stronger singularity for $u(t, x)$ there.

This sort of phenomenon was understood, at least from a heuristic point of view, in the nineteenth century, and it played a role in an important argument of Poisson. At the time, there was a debate about whether the propagation of light was a wave phenomenon. Poisson did not think it was, and he noted that if it were, the light waves propagated past such an obstacle should produce a bright spot along the axis normal to the disk and through its center. The experiment was performed and the bright spot was observed. This is now called the Poisson spot, and its occurrence convinced many physicists, including Poisson, that the propagation of light is a wave phenomenon.

## Harmonic Analysis and the Wave Equation

The wave equation [1] with Cauchy data [3] can be regarded as an operator differential equation, with solution

$$u(t, x) = \cos t\sqrt{-\Delta}\, f(x) + \frac{\sin t\sqrt{-\Delta}}{\sqrt{-\Delta}}\, g(x) \qquad [21]$$

This brings one to investigate functions of the self-adjoint operator $\Delta$. If $M = \mathbb{R}^n$, one can do this using the Fourier transform, which is given by

$$\mathcal{F}f(\xi) = \hat{f}(\xi) = (2\pi)^{-n/2} \int f(x) e^{-ix\cdot\xi}\, dx \qquad [22]$$

One defines $\mathcal{F}^*$ by changing $e^{-ix\cdot\xi}$ to $e^{ix\cdot\xi}$ in [22], and the Fourier inversion formula says $\mathcal{F}$ and $\mathcal{F}^*$ are

inverses of each other on various function spaces, including $L^2(\mathbb{R}^n)$. Then one has

$$\varphi(\sqrt{-\Delta})f(x) = (2\pi)^{-n/2} \int \varphi(|\xi|)\hat{f}(\xi)\mathrm{e}^{\mathrm{i}x\cdot\xi}\,\mathrm{d}\xi \qquad [23]$$

Note that [23] is equal to

$$\int \Phi(x-y)f(y)\,\mathrm{d}y = \Phi * f(x) \qquad [24]$$

where

$$\Phi(x) = (2\pi)^{-n} \int \varphi(|\xi|)\mathrm{e}^{\mathrm{i}x\cdot\xi}\,\mathrm{d}\xi \qquad [25]$$

In particular, [21] becomes

$$u(t,x) = \frac{\partial}{\partial t}R_t * f(x) + R_t * g(x) \qquad [26]$$

where

$$R_t(x) = (2\pi)^{-n} \int \frac{\sin t|\xi|}{|\xi|}\mathrm{e}^{\mathrm{i}x\cdot\xi}\,\mathrm{d}\xi \qquad [27]$$

is the fundamental solution to the wave equation.

The integral [27] is not an easy integral when $n>1$, but the answer can be derived by analytic continuation from the Poisson kernel, that is,

$$\begin{aligned}\mathrm{e}^{-y\sqrt{-\Delta}}f(x) &= P_y * f(x)\\ P_y(x) &= C_n y(|x|^2+y^2)^{-(n+1)/2}\end{aligned} \qquad [28]$$

where $C_n = \pi^{-(n+1)/2}\Gamma((n+1)/2)$. One gets

$$R_t(x) = \lim_{\varepsilon\searrow 0}\frac{C_n}{n-1}\operatorname{Im}(|x|^2-(t-\mathrm{i}\varepsilon)^2)^{-(n-1)/2} \qquad [29]$$

Taking this limit for $n=2,3$ yields the formulas [8]–[9]. There are several ways to derive [28]. One, which is flexible and useful for other situations, derives it from the formula for the heat kernel,

$$\mathrm{e}^{t\Delta}f(x) = H_t * f(x), \quad H_t(x) = (4\pi t)^{-n/2}\,\mathrm{e}^{-|x|^2/4t} \qquad [30]$$

via the subordination identity:

$$\begin{aligned}\mathrm{e}^{-yA} &= \frac{y}{2\pi^{1/2}}\int_0^\infty \mathrm{e}^{-y^2/4t}\,\mathrm{e}^{-tA^2}t^{-3/2}\,\mathrm{d}t\\ &A>0, \quad y>0\end{aligned} \qquad [31]$$

with $A=\sqrt{-\Delta}$. The heat kernel can be computed via [23], which becomes a well-known Gaussian integral. The identity [31] can be proved using the fact that the Fourier integral formula for $P_y(x)$ is elementary to compute when $n=1$.

To understand functions of the Laplace operator on a cone $C(N)$, one uses

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{n-1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\Delta_N \qquad [32]$$

where $\Delta_N$ is the Laplace operator on $N$, which follows from [2] and [18]. Here $n=\dim C(N)$. This is a modified Bessel operator. We define the operator

$$\nu = (-\Delta_N+\alpha^2)^{1/2}, \quad \alpha = -\frac{n-2}{2} \qquad [33]$$

For each $\nu_j$ in the spectrum of $\nu$, we consider the Hankel transform

$$H_{\nu_j}g(\lambda) = \int_0^\infty g(r)J_{\nu_j}(\lambda r)r\,\mathrm{d}r \qquad [34]$$

where $J_{\nu_j}$ is the Bessel function of order $\nu_j$. The Hankel inversion formula says $H_{\nu_j}$ is unitary on $L^2(\mathbb{R}^+, r\,\mathrm{d}r)$, and is its own inverse. Consequently, we can write the action of $\varphi(\sqrt{-\Delta})$ on $L^2(C(N))$ as

$$\varphi(\sqrt{-\Delta})g(r,x) = \int_0^\infty K_\varphi(r,s,\nu)g(s,x)s^{n-1}\,\mathrm{d}s \qquad [35]$$

where $K_\varphi(r,s,\nu)$ is a family of operators on $L^2(N)$, given by

$$K_\varphi(r,s,\nu) = (rs)^\alpha \int_0^\infty \varphi(\lambda)J_\nu(\lambda r)J_\nu(\lambda s)\lambda\,\mathrm{d}\lambda \qquad [36]$$

To obtain the wave kernel on $C(N)$, one can analytically continue formulas for the Poisson kernel, for $\mathrm{e}^{-y\sqrt{-\Delta}}$. Such formulas arise from the Lipschitz–Hankel identity:

$$\begin{aligned}&\int_0^\infty \mathrm{e}^{-y\lambda}J_\nu(r\lambda)J_\nu(s\lambda)\,\mathrm{d}\lambda\\ &\quad = \frac{1}{\pi}(rs)^{-1/2}Q_{\nu-1/2}\left(\frac{r^2+s^2+y^2}{2rs}\right)\end{aligned} \qquad [37]$$

Here $Q_{\nu-1/2}(\tau)$ is a Legendre function. The identity [37] is one of the more difficult identities in the theory of Bessel functions. It is useful to know that it can be derived by applying a slight variant of the subordination identity [31] to the more elementary identity

$$\int_0^\infty \mathrm{e}^{-t\lambda^2}J_\nu(r\lambda)J_\nu(s\lambda)\lambda\,\mathrm{d}\lambda = \frac{1}{2t}\mathrm{e}^{-(r^2+s^2)/4t}I_\nu\left(\frac{rs}{2t}\right) \qquad [38]$$

(where $I_\nu(y)=\mathrm{e}^{-\pi\mathrm{i}\nu/2}J_\nu(\mathrm{i}y)$ for $y>0$), which describes the behavior of the heat kernel on $C(N)$.

Carrying out the analytic continuation of [37] to imaginary $y$ yields results stated in the section on basic propagation phenomena, once one

understands the behavior of families of functions of the operator $\nu$ so produced. An approach taken by Cheeger and Taylor (1982) to this was to synthesize these operators from $e^{is\nu}$, $s \in \mathbb{R}$, and deduce their behavior from the behavior of the solution operator to the wave equation on the base $N$.

One can apply similar considerations to $M = \mathbb{R}^n \setminus B$, which is the truncated cone $[1, \infty) \times S^{n-1}$, with metric tensor [18], where $g$ is the metric tensor on $S^{n-1}$, and Laplace operator given by [32], with $\Delta_N$ the Laplace operator on $S^{n-1}$. The problem of diffraction of waves by the ball $B$ can be recast as solving

$$\frac{\partial^2 u}{\partial t^2} - \Delta u = 0 \quad \text{on } \mathbb{R} \times M \\ u|_{\mathbb{R}\times\partial M} = f, \quad u(t,x) = 0 \quad \text{for } t \ll 0 \tag{39}$$

with $f$ compactly supported on $\mathbb{R} \times \partial M$. Taking the partial Fourier transform with respect to $t$ yields the reduced wave equation

$$(\Delta + \lambda^2)v = 0 \quad \text{for } |x| > 1, \quad v|_{S^{n-1}} = g(x, \lambda) \tag{40}$$

and the condition $u(t,x) = 0$ for $t \ll 0$ yields for $v$ the outgoing radiation condition

$$r^{(n-1)/2}\left(\frac{\partial v}{\partial r} - i\lambda v\right) \to 0 \quad \text{as } r \to \infty \tag{41}$$

The solution is

$$v(x, \lambda) = r^{-(n-2)/2} \frac{H_\nu^{(1)}(\lambda r)}{H_\nu^{(1)}(\lambda)} g(x, \lambda) \tag{42}$$

with $\nu$ as in [33] and $H^{(1)}$ the Hankel function.

The behavior of $H_{\nu_j}^{(1)}(\lambda r)/H_{\nu_j}^{(1)}(\lambda)$ as $\nu_j, \lambda \to \infty$ with ratio in a small neighborhood of 1 can be shown to control the behavior of the solution $u$ to [39] near grazing rays. There is an asymptotic formula for this, which is one of the most delicate analytical results in the theory of Bessel functions. The result is that, uniformly for $z$ near 1, as $\mu \to \infty$,

$$H_\mu^{(1)}(\mu z) \sim 2e^{-\pi i/3}\left(\frac{4\zeta}{1 - z^2}\right)^{1/4} \times \left\{ A_+(\mu^{2/3}\zeta) \sum_{k \geq 0} a_k(\zeta)\mu^{1/3-2k} + A_+'(\mu^{2/3}\zeta) \sum_{k \geq 0} b_k(\zeta)\mu^{5/3-2k} \right\} \tag{43}$$

Here

$$A_+(\xi) = \mathrm{Ai}(e^{-2\pi i/3}\xi) \tag{44}$$

where Ai is the Airy function. The coefficients $a_k(\zeta)$ and $b_k(\zeta)$ are smooth functions of their argument, $\zeta = \zeta(z)$, which is defined by

$$\frac{2}{3}\zeta^{3/2} = \int_z^1 \sqrt{1 - t^2}\,\frac{dt}{t} \tag{45}$$

Making use of [43] in [44], one can obtain a parametrix for $u$ (i.e., a solution modulo a $C^\infty$ error) whose form is a special case of the formula [50], which we will present in the next section.

## Geometrical Optics and Extensions

By results of the last section, the solution to [1] when $M = \mathbb{R}^n$ has the form

$$u(t, x) = \sum_{\pm} \int e^{\pm it|\xi| + ix\cdot\xi}\hat{h}_\pm(\xi)\, d\xi \tag{46}$$

where the functions $h_\pm$ are produced from the initial data via simpler transformations. For a general metric tensor, one can produce a parametrix (i.e., an approximation to $u(t,x)$ with a $C^\infty$ error) in the following form:

$$u(t, x) = \sum_{\pm} \int a^\pm(t, x, \xi)\, e^{i\varphi^\pm(t,x,\xi)}\hat{h}_\pm(\xi)\, d\xi \tag{47}$$

Here the phase functions $\varphi^\pm(t,x,\xi)$ are smooth for $\xi \neq 0$ and homogeneous of degree 1 in $\xi$. The amplitudes $a^\pm(t,x,\xi)$ are smooth and have asymptotic expansions as $|\xi| \to \infty$:

$$a^\pm(t, x, \xi) \sim \sum_{k \geq 0} a_k^\pm(t, x, \xi) \tag{48}$$

with $a_k^\pm(t,x,\xi)$ homogeneous of degree $-k$ in $\xi$. One applies $\partial_t^2 - \Delta$ to both sides of [47], and obtains an operator of a similar form, with new amplitudes $b^\pm(t,x,\xi) \sim \sum b_k^\pm(t,x,\xi)$. Setting the terms in this asymptotic expansion equal to zero yields, first for $\varphi^\pm(t,x,\xi)$, a partial differential equation known as the eikonal equation:

$$\frac{\partial \varphi^\pm}{\partial t} = \pm|\nabla_x \varphi^\pm| \tag{49}$$

where $|v|$ is the norm of a vector $v \in T_x M$, determined by the metric tensor. Setting $b_k^\pm(t,x,\xi) = 0$ for $k \geq 1$ yields linear differential equations for the amplitude terms in [48], known as transport equations.

Operators of the form [47] are special cases of Fourier integral operators. Seminal works of Keller (1953) and Lax (1957) gave an important stimulus to work on these operators, and work of Hörmander (1971) turned this into a systematic and powerful theory. A particular advance regards

producing a parametrix valid for all $t$. Generally, one can solve [49] and the associated transport equations for $t$ in some interval, past which the eikonal equation might break down. Hörmander's theory treats products of Fourier integral operators, yielding global constructions. This facilitates the treatment of caustics mentioned earlier. Stationary-phase methods can be brought to bear to relate the singularities of $Th$ to those of $h$, when $T$ is a Fourier integral operator.

To construct parametrices for waves reflecting off a boundary, one can again reduce the problem to one of the form [39]. Waves that reflect transversally are given by parametrices of the form [47], although with the role of the variables changed, so that $t$ in [47]–[49] is replaced by a coordinate that vanishes on $\mathbb{R} \times \partial M$.

A parametrix that treats grazing rays can be written in the form of a Fourier–Airy integral operator:

$$u(y) = \int_{\mathbb{R}^n} \left[ a\, A_+(\zeta) + \mathrm{i}|\xi|^{-1/3} b\, A'_+(\zeta) \right] \times A_+(\zeta_0)^{-1} \, \mathrm{e}^{\mathrm{i}\theta} \widehat{F}(\xi) \, \mathrm{d}\xi \qquad [50]$$

Here $y = (y_1, \ldots, y_{n+1})$ denotes a coordinate system on a neighborhood of a boundary point of $\mathbb{R} \times M$, with $y_{n+1} = 0$ on $\mathbb{R} \times \partial M$. We have a pair of phase functions $\theta(y, \xi)$ and $\zeta(y, \xi)$, homogeneous in $\xi$ of degree 1 and 2/3, respectively, and a pair of amplitudes $a(y, \xi)$ and $b(y, \xi)$, each having asymptotic expansions of the form [48]. The function $A_+$ is the Airy function [44]. The phase functions satisfy a coupled pair of eikonal equations:

$$\begin{aligned} \langle \nabla_y \theta, \nabla_y \theta \rangle + \zeta \langle \nabla_y \zeta, \nabla_y \zeta \rangle &= 0 \\ \langle \nabla_y \theta, \nabla_y \zeta \rangle &= 0 \end{aligned} \qquad [51]$$

where $\langle \cdot, \cdot \rangle$ denotes the Lorentz inner product on $T_y(\mathbb{R} \times M)$ given by $\mathrm{d}t^2 - g$. More precisely, [51] is to hold in the region where $\zeta \le 0$, and also to infinite order at $y_{n+1} = 0$, for $\zeta \ge 0$. One requires $\partial\theta/\partial\xi_j$ to have linearly independent $y$-gradients, for $j = 1, \ldots, n$, and

$$\zeta(y, \xi) = \zeta_0(\xi) = \xi_1^{-1/3} \xi_n \quad \text{for } y_{n+1} = 0 \qquad [52]$$

The terms in the asymptotic expansions of $a(y, \xi)$ and $b(y, \xi)$ satisfy coupled systems of transport equations. One can arrange that $b(y, \xi) = 0$ for $y_{n+1} = 0$. Then $u|_{\mathbb{R}\times\partial M} = TF$, where $T$ is a Fourier integral operator, which can be inverted, modulo a smooth error, by Hörmander's theory, producing a parametrix for [39].

The construction of solutions to [51] satisfying [52] is due to Melrose. This followed earlier works of Ludwig (1967), Melrose (1975), and Taylor (1976), which produced solutions satisfying [52] to infinite order at $\xi_n = 0$. This earlier construction is adequate to produce a grazing ray parametrix, but the sharper result [52] is extremely valuable for constructing a gliding ray parametrix. This has the form

$$u(y) = \int_{\mathbb{R}^n} [a \ \mathrm{Ai}(\zeta) + \mathrm{i}|\xi|^{-1/3} b \, \mathrm{Ai}'(\zeta)] \times \mathrm{Ai}(\zeta_0)^{-1} \, \mathrm{e}^{\mathrm{i}\theta} \widehat{F}(\xi) \, \mathrm{d}\xi \qquad [53]$$

It differs from [50] in the use of Ai rather than $A_+$. Since Ai has real zeros, it is also convenient to pick $T > 0$ and evaluate $\theta, \zeta, a$, and $b$ at $(\xi_1, \ldots, \xi_{n-1}, \xi_n + \mathrm{i}T)$, and take $\zeta_0 = \xi^{-1/3}(\xi_n + \mathrm{i}T)$. The treatment of the eikonal and transport equations is as above, though the Fourier–Airy integral operator [50] has a different behavior from [53], reflecting the difference between how singularities in solutions to the wave equation are carried by grazing and by gliding rays.

## Further Reading


Bowman J, Senior T, and Uslengi P (1969) *Electromagnetic and Acoustic Scattering by Simple Shapes*. Amsterdam: North-Holland.

Cheeger J and Taylor ME (1982) On the diffraction of waves by conical singularities, I, II. *Communications in Pure and Applied Mathematics* 25: 275–331; 487–529.

Garnir HG (ed.) (1981) *Singularities in Boundary Value Problems*, NATO Advanced Study Institute Series, vol. 65. Boston: D. Reidel.

Hörmander L (1971) Fourier integral operators, I. *Acta Mathematica* 127: 79–183.

Hörmander L (1985) *The Analysis of Partial Differential Operators*. vols. 3–4. New York: Springer.

Keller JB (1953) The geometrical theory of diffraction. *Proc. Symp. Microwave Optics*, Eaton Electronics Lab, McGill University.

Lax P (1957) Asymptotic solutions of oscillatory initial value problems. *Duke Mathematical Journal* 24: 627–646.

Ludwig D (1967) Uniform asymptotic expansion of the field scattered by a convex object at high frequencies. *Communications in Pure and Applied Mathematics* 19: 103–138.

Melrose RB (1975) Microlocal parametrices for diffractive boundary problems. *Duke Mathematical Journal* 42: 605–635.

Melrose RB and Sjöstrand J (1978) Singularities of boundary value problems. *Communications in Pure and Applied Mathematics* 31: 593–617.

Melrose RB and Taylor ME (1985) Near peak scattering and the corrected Kirchhoff approximation for a convex obstacle. *Advances in Mathematics* 55: 242–315.

Melrose RB and Wunsch J (2004) Propagation of singularities for the wave equation on conic manifolds. *Inventiones Mathematicae* 156: 235–299.

Taylor ME (1976) Grazing rays and reflection of singularities of solutions to wave equations. *Communications in Pure and Applied Mathematics* 29: 1–38.

Taylor ME (1981) *Pseudodifferential Operators*. Princeton, NJ: Princeton University Press.

Taylor ME (1996) *Partial Differential Equations*, vols. 1–3. New York: Springer.

# Wavelets: Application to Turbulence

**M Farge**, Ecole Normale Supérieure, Paris, France
**K Schneider**, Université de Provence, Marseille, France



## Introduction about Turbulence and Wavelets

### What is Turbulence?

Turbulence is a highly nonlinear regime encountered in fluid flows. Such flows are described by continuous fields, for example, velocity or pressure, assuming that the characteristic scale of the fluid motions is much larger than the mean free path of the molecular motions. The prediction of the spacetime evolution of fluid flows from first principles is given by the solutions of the Navier–Stokes equations. The turbulent regime develops when the nonlinear term of Navier–Stokes equations strongly dominates the linear term; the ratio of the norms of both terms is the Reynolds number *Re*, which characterizes the level of turbulence. In this regime nonlinear instabilities dominate, which leads to the flow sensitivity to initial conditions and unpredictability.

The corresponding turbulent fields are highly fluctuating and their detailed motions cannot be predicted. However, if one assumes some statistical stability of the turbulence regime, averaged quantities, such as mean and variance, or other related quantities, for example, diffusion coefficients, lift or drag, may still be predicted.

When turbulent flows are statistically stationary (in time) or homogeneous (in space), as it is classically supposed, one studies their energy spectrum, given by the modulus of the Fourier transform of the velocity autocorrelation.

Unfortunately, since the Fourier representation spreads the information in physical space among the phases of all Fourier coefficients, the energy spectrum loses all structural information in time or space. This is a major limitation of the classical way of analyzing turbulent flows. This is why we have proposed to use the wavelet representation instead and define new analysis tools that are able to preserve time and space locality.

The same is true for computing turbulent flows. Indeed, the Fourier representation is well suited to study linear motions, for which the superposition principle holds and whose generic behavior is, either to persist at a given scale, or to spread to larger ones. In contrast, the superposition principle does not hold for nonlinear motions, their archetype being the turbulent regime, which therefore cannot be decomposed into a sum of independent motions that can be separately studied. Generically, their evolution involves a wide range of scales, exciting smaller and smaller ones, even leading to finite-time singularities, e.g., shocks. The "art" of predicting the evolution of such nonlinear phenomena consists of disentangling the active from the passive elements: the former should be deterministically computed, while the latter could either be discarded or their effect statistically modeled. The wavelet representation allows to analyze the dynamics in both space and scale, retaining only those degrees of freedom which are essential to predict the flow evolution. Our goal is to perform a kind of "distillation" and retain only the elements which are essential to compute the nonlinear dynamics.

### How One Studies Turbulence?

When studying turbulence one is uneasy about the fact that there are two different descriptions, depending on which side of the Fourier transform one looks from.

- On the one hand, looking from the Fourier space representation, one has a theory which assumes the existence of a nonlinear cascade in an intermediate range of wavenumbers sets, called the "inertial range" where energy is conserved and transferred towards high wavenumbers, but only on average (i.e., considering either ensemble or time or space averages). This implies that a turbulent flow is excited at wavenumbers lower than those of the inertial range and dissipated at wavenumbers higher. Under these hypotheses, the theory predicts that the slope of the energy spectrum in the inertial range scales as $k^{-5/3}$ in dimension 3 and as $k^{-3}$ in dimension 2, $k$ being the wavenumber, i.e., the modulus of the wave vector.
- On the other hand, if one studies turbulence from the physical space representation, there is not yet any universal theory. One relies instead on empirical observations, from both laboratory and numerical experiments, which exhibit the formation and persistence of coherent vortices, even at very high Reynolds numbers. They correspond to the condensation of the vorticity field into some organized structures that contain most of the energy ($L^2$-norm of velocity) and enstrophy ($L^2$-norm of vorticity).

Moreover, the classical method for modeling turbulent flows consists in neglecting high-wavenumber motions and replacing them by their average, supposing their dynamics to be either linear or slaved to the low wavenumber motions. Such a method would work if there exists a clear separation between low and high wavenumbers, that is, a spectral gap.

Actually, there is now strong evidence, from both laboratory and direct numerical simulation (DNS) experiments, that this is not the case. Conversely, one observes that turbulent flows are nonlinearly active all along the inertial range and that coherent vortices seem to play an essential dynamical role there, especially for transport and mixing. One may then ask the following questions: Are coherent vortices the elementary building blocks of turbulent flows? How can we extract them? Do their mutual interactions have a universal character? Can we compress turbulent flows and compute their evolution with a reduced number of degrees of freedom corresponding to the coherent vortices?

The DNS of turbulent flows, based on the integration of the Navier–Stokes equations using either grid points in physical space or Fourier modes in spectral space, requires a number of degrees of freedom per time step that varies as $Re^{9/4}$ in dimension 3 (and as $Re$ in dimension 2). Due to the inherent limitation of computer performances, one can presently only perform DNS of turbulent flows up to Reynolds numbers $Re = 10^6$. To compute higher Reynolds flows, one should then design *ad hoc* turbulence models, whose parameters are empirically adjusted to each type of flows, in particular to their geometry and boundary conditions, using data from either laboratory or numerical experiments.

### What are Wavelets?

The wavelet transform unfolds signals (or fields) into both time (or space) and scale, and possibly directions in dimensions higher than 1. The starting point is a function $\psi \in L^2(R)$, called the "mother wavelet", which is well localized in physical space $x \in \mathbb{R}$, is oscillating ($\psi$ has at least a vanishing integral, or better, its first $m$ moments vanish), and is smooth (its Fourier transform $\hat{\psi}(k)$ exhibits fast decay for wave numbers $|k|$ tending to infinity). The mother wavelet then generates a family of dilated and translated wavelets

$$\psi_{a,b}(x) = a^{-1/2}\psi\left(\frac{x-b}{a}\right)$$

with $a \in R^+$ the scale parameter and $b \in R$ the position parameter, all wavelets being normalized in $L^2$-norm.

The wavelet transform of a function $f \in L^2(R)$ is the inner product of $f$ with the analyzing wavelets $\psi_{a,b}$, which gives the wavelet coefficients: $\widetilde{f}(a,b) = \langle f, \psi_{a,b}\rangle = \int f(x)\psi_{a,b}(x)\,\mathrm{d}x$. They measure the fluctuations of $f$ around the scale $a$ and the position $b$. $f$ can then be reconstructed without any loss as the inner product of its wavelet coefficients $\widetilde{f}$ with the analyzing wavelets

$$\psi_{a,b} \ : f(x) = C_\psi^{-1} \int\!\!\int \widetilde{f}(a,\ b)\ \psi_{a,b}(x) a^{-2} \mathrm{d}a\,\mathrm{d}b$$

$C_\psi = \int |\hat{\psi}|^2 |k|^{-1}\,\mathrm{d}k$ being a constant which depends on the wavelet $\psi$.

Like the Fourier transform, the wavelet transform realizes a change of basis from physical space to wavelet space which is an isometry. It thus conserves the inner product (Plancherel theorem), and in particular energy (Parseval's identity). Let us mention that, due to the localization of wavelets in physical space, the behavior of the signal at infinity does not play any role. Therefore, the wavelet analysis and synthesis can be performed locally, in contrast to the Fourier transform where the nonlocal nature of the trigonometric functions does not allow to perform a local analysis.

Moreover, wavelets constitute building blocks of various function spaces out of which some can be used to contruct orthogonal bases. The main difference between the continuous and the orthogonal wavelet transforms is that the latter is non-redundant, but only preserves the invariance by translation and dilation only for a discrete subset of wavelet space which corresponds to the dyadic grid $\lambda = (j, i)$, for which scale is sampled by octaves $j$ and space by positions $2^{-j}i$. The advantage is that all orthogonal wavelet coefficients are decorrelated, which is not the case for the continuous wavelet transform whose coefficients are redundant and correlated in space and scale. Such a correlation can be visualized by plotting the continuous wavelet coefficients of a white noise and the patterns one thus observes are due to the reproducing kernel of the continuous wavelet transform, which corresponds to the correlation between the analyzing wavelets themselves.

In practice, to analyze turbulent signals or fields, one should use the continuous wavelet transform with complex-valued wavelets, since the modulus of the wavelet coefficients allows to read the evolution of the energy density in both space (or time) and scales. If one uses real-valued wavelets instead, the modulus of the wavelet coefficients will present the same oscillations as the analyzing wavelets and it will then become difficult to sort out features

belonging to the signal or to the wavelets. In the case of complex-valued wavelets, the quadrature between the real and the imaginary parts of the wavelet coefficients eliminates these spurious oscillations; this is why we recommend to use complex-valued wavelets, such as the Morlet wavelet. To compress turbulent flows, and *a fortiori* to compute their evolution at a reduced cost, compared to standard methods (finite difference, finite volume, or spectral methods), one should use orthogonal wavelets. This avoids redundancy, since one has the same number of grid points as wavelet coefficients. Moreover there exists a fast algorithm to compute the orthogonal wavelet coefficients which is even faster than the fast Fourier transform, having $O(N)$ operations instead of $O(N \log_2 N)$.

The first paper about the continuous wavelet transform has been published by Grossmann and Morlet (1984). Then, discrete wavelets were constructed, leading to frames (Daubechies *et al.* 1986) and orthogonal bases (Lemarié and Meyer, 1986). From there the formalism of multiresolution analysis (MRA) has been constructed which led to the fast wavelet algorithm (Mallat 1989). The first application of wavelets to analyze turbulent flows has been published by Farge and Rabreau (1988). Since then a long-term research program has been developed for analyzing, computing and modeling turbulent flows using either continuous wavelets, orthogonal wavelets, or wavelet packets.

## Wavelet Analysis

### Wavelet Spectra

**Wavelet space** To study turbulent signals one uses the continuous wavelet transform for analysis, and the orthogonal wavelet transform for compression and computation. To perform a continuous wavelet transform, one can choose:

- either a real-valued wavelet, such as the Marr wavelet, also called "Mexican hat," which is the second derivative of a Gaussian,

$$\psi(x) = (1 - x^2) \exp\left(\frac{-x^2}{2}\right) \qquad [1]$$

- or a complex-valued wavelet, such as the Morlet wavelet,

$$\begin{cases} \widehat{\psi}(k) = \dfrac{1}{2\pi} \exp\left(-\dfrac{(k-k_\psi)^2}{2}\right) & \text{for } k > 0 \\ \widehat{\psi}(k) = 0 \quad \text{for } k \leq 0 \end{cases} \qquad [2]$$

with the wavenumber $k_\psi$ denoting the barycenter of the wavelet support in Fourier space computed as

$$k_\psi = \frac{\int_0^\infty k|\widehat{\psi}(k)|\mathrm{d}k}{\int_0^\infty |\widehat{\psi}(k)|\mathrm{d}k} \qquad [3]$$

For the orthogonal wavelet transform, there is a large collection of possible wavelets and the choice depends on which properties are preferred, for instance: compact support, symmetry, smoothness, number of cancelations, computational efficiency.

From our own experience, we tend to prefer the Coifman wavelet 12, which is compactly supported, has four vanishing moments, is quasi-symmetric, and is defined with a filter of length 12, which leads to a computational cost for the fast wavelet transform in $24N$ operations, since two filters are used.

As stated above, we recommend the complex-valued continuous wavelet transform for analysis. In this case, one plots the modulus and the phase of the wavelet coefficients in wavelet space, with a linear horizontal axis for the position $b$, and a logarithmic vertical axis for the scale $a$, with the largest scale at the bottom and the smallest scale at the top.

In **Figure 1a** we show the wavelet analysis of a turbulent signal, corresponding to the time evolution of the velocity fluctuations of two successive vortex breakdowns, measured by hot-wire anemometry at $N = 32768 = 2^{15}$ instants (Cuypers *et al.* 2003). The modulus of the wavelet coefficients (**Figure 1b**) shows that during the vortex breakdown, which is due to strong nonlinear flow instability, energy is spread over a wide range of scales. The phase of the wavelet coefficients (**Figure 1c**) is plotted only where the modulus is non-negligible, otherwise the phase information would be meaningless. In **Figure 1c**, one observes that the lines of constant phase point towards the instants where the signal is less regular, that is, during vortex breakdowns.

**Local wavelet spectrum** Since the wavelet transform conserves energy and preserves locality in physical space, one can extend the concept of energy spectrum and define a local energy spectrum, such that

$$\widetilde{E}(k,x) = \frac{1}{C_\psi k_\psi}\left|\widetilde{f}\left(\frac{k_\psi}{k}, x\right)\right|^2 \quad \text{for } k \geq 0 \qquad [4]$$

where $k_\psi$ is the centroid wavenumber of the analyzing wavelet $\psi$ and $C_\psi$ is defined in the

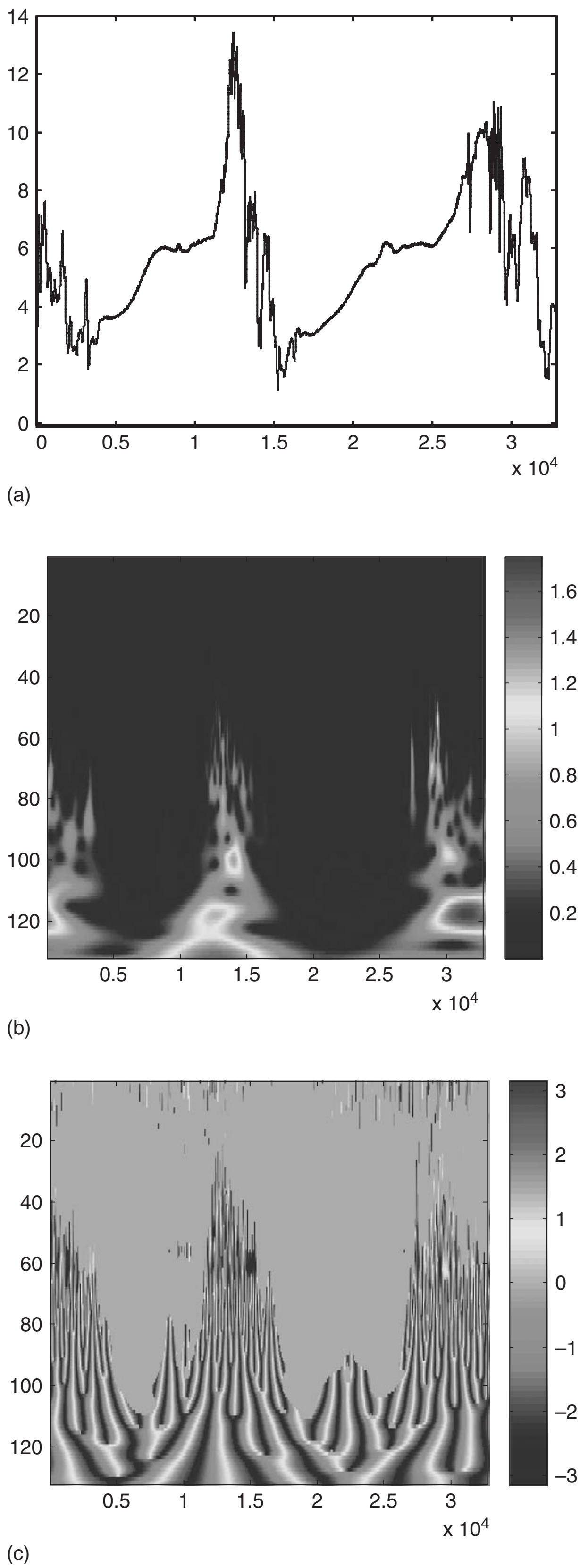



**Figure 1** Example of a one-dimensional continuous wavelet analysis. (a) the signal to be analyzed, (b) the modulus of its wavelet coefficients, (c) the phase of its wavelet coefficients.

admissibility condition (respectively, eqns [10] and [1] in the article Wavelets: Mathematical Theory).

By measuring $\widetilde{E}(k,x)$ at different instants or positions, one estimates which elements in the signal contribute most to the global Fourier energy spectrum, inorder to suggest a way to decompose the signal into different components. For example, if one considers turbulent flows, one can compare the energy spectrum of the coherent structures (such as isolated vortices in incompressible flows or shocks in compressible flows) and the energy spectrum of the incoherent background flow, since both elements exhibit different correlations and therefore different spectral slopes.

**Global wavelet spectrum** Although the wavelet transform analyzes the flow using localized functions rather than complex exponentials, one can show that the global wavelet energy spectrum converges towards the Fourier energy spectrum, provided the analyzing wavelet has enough vanishing moments. More precisely, the global wavelet spectrum, defined by integrating [4] over all positions,

$$\widetilde{E}(k) = \int_{-\infty}^{\infty} \widetilde{E}(k,x)\mathrm{d}x \qquad [5]$$

gives the correct exponent for a power-law Fourier energy spectrum $E(k) \propto k^{-\beta}$ if the analyzing wavelet has at least $M > (\beta - 1)/2$ vanishing moments. Thus, the steeper the energy spectrum one studies, the more vanishing moments the analyzing wavelet should have.

The inertial range which corresponds to the scales when turbulent flows are dominated by nonlinear interactions, exhibits a power-law behavior as predicted by the statistical theory of homogeneous and isotropic turbulence.

The ability to correctly evaluate the slope of the energy spectrum is an important property of the wavelet transform which is related to its ability to detect and characterize singularities. We will not discuss here how wavelet coefficients could be used to study singularities and fractal measures, since it is presented in detail elsewhere (*see* Wavelets: Applications).

### Relation to Classical Analysis

**Relation to Fourier spectrum** The global wavelet energy spectrum $\widetilde{E}(k)$ is actually a smoothed version of the Fourier energy spectrum $E(k)$. This can be

seen from the following relation between the two spectra:

$$\widetilde{E}(k) = \frac{1}{C_\psi k_\psi} \int_0^\infty E(k') \left| \widehat{\psi}\left(\frac{k_\psi k'}{k}\right) \right|^2 \mathrm{d}k' \qquad [6]$$

which shows that the global wavelet spectrum is an average of the Fourier spectrum weighted by the square of the Fourier transform of the analyzing wavelets at wavenumber $k$. Note that the larger $k$, the larger the averaging interval, because wavelets are bandpass filters with $\Delta k/k$ constant. This property of the global wavelet energy spectrum is particularly useful to study turbulent flows. Indeed, the Fourier energy spectrum of a single realization of a turbulent flow is too oscillating to be able to clearly detect a slope, while it is no more the case for the global wavelet energy spectrum, which is a better estimator of the spectral slope.

The real-valued Marr wavelet [1] has only two vanishing moments and thus can correctly measure the energy spectrum exponents up to $\beta < 5$. In the case of the complex-valued Morlet wavelet [2], only the zeroth-order moment is null, but the higher $m$th order moments are very small ($\propto k_\psi^m \, \mathrm{e}^{(-k_\psi^2/2)}$), provided that $k_\psi$ is larger than 5. For instance, the Morlet wavelet transform with $k_\psi = 6$ gives accurate estimates of the power–law exponent of the energy spectrum up to $\beta < 7$.

There is also a family of wavelets with an infinite number of cancelations

$$\widehat{\psi}_n(k) = \alpha_n \exp\left(-\frac{1}{2}\left(k^2 + \frac{1}{k^{2n}}\right)\right) \quad n \geq 1 \qquad [7]$$

where $\alpha_n$ is chosen for normalization.

These wavelets can therefore correctly measure any power–law energy spectrum, and thus detect the difference between a power–law energy spectrum and a Gaussian energy spectrum ($E(k) \propto \mathrm{e}^{(-(k/k_0)^2)}$). For instance, it is important in turbulence to determine the wavenumber after which the energy spectrum decays exponentially, since this wavenumber defines the end of the inertial range, dominated by nonlinear interactions, and the beginning of the dissipative range, dominated by linear dissipation.

**Relation to structure functions** In this subsection we will point out the limitations of classical measures of intermittency and present a set of wavelet-based alternatives.

The classical measures based on structure functions can be thought of as a special case of wavelet filtering using a nonsmooth wavelet defined as the difference of two Diracs (DOD). It is this lack of regularity of the underlying wavelet that limits the adequacy of classical measures to analyze smooth signals. Wavelet-based diagnostics can overcome these limitations, and produce accurate results, whatever the signal to be analyzed.

We will link the scale-dependent moments of the wavelet coefficients and the structure functions, which are classically used to study turbulence. In the case of second-order statistics, the global wavelet spectrum corresponds to the second-order structure function. Furthermore, a rigorous bound for the maximum exponent detected by the structure functions can be computed, but there is a way to overcome this limitation by using wavelets.

The increments of a signal, also called the modulus of continuity, can be seen as its wavelet coefficients using the DOD wavelet

$$\psi^\delta(x) = \delta(x+1) - \delta(x) \qquad [8]$$

We thus obtain

$$f(x+a) - f(x) = \widetilde{f}_{x,a} = \langle f, \psi^\delta_{x,a} \rangle \qquad [9]$$

with $\psi_{x,a}(y) = 1/a[\delta((y-x)/a+1) - \delta((y-x)/a)]$. Note that the wavelet is normalized with respect to the $L^1$-norm. The $p$th-order structure function $S_p(a)$ therefore corresponds to the $p$th-order moment of the wavelet coefficients at scale $a$

$$S_p(a) = \int (\widetilde{f}_{x,a})^p \mathrm{d}x \qquad [10]$$

As the DOD wavelet has only one vanishing moment (its mean), the exponent of the $p$th-order structure function in the case of a self-similar behavior is limited by $p$, that is, if $S_p(a) \propto a^{\zeta(p)}$, then $\zeta(p) < p$. To be able to detect larger exponents, one has to use increments with a larger stencil, or wavelets with more vanishing moments.

We now concentrate on the case $p = 2$, that is, the energy norm. Equation [6] gives the relation between the global wavelet spectrum $\widetilde{E}(k)$ and the Fourier spectrum $E(k)$ for an arbitrary wavelet $\psi$. For the DOD wavelet we find, since $\widehat{\psi^\delta}(k) = \mathrm{e}^{\mathrm{i}k} - 1 = \mathrm{e}^{\mathrm{i}k/2}(\mathrm{e}^{\mathrm{i}k/2} - \mathrm{e}^{\mathrm{i}\,k/2})$ and hence $|\widehat{\psi^\delta}(k)|^2 = 2(1 - \cos k)$, that

$$\widetilde{E}(k) = \frac{1}{C_\psi k} \int_0^\infty E(k') \left(2 - 2\cos\left(\frac{k_\psi k'}{k}\right)\right) \mathrm{d}k' \qquad [11]$$

Setting $a = k_\psi / k$, we see that the wavelet spectrum corresponds to the second-order structure function, such that

$$\widetilde{E}(k) = \frac{1}{C_\psi k} S_2(a) \qquad [12]$$

The above results show that, if the Fourier spectrum behaves like $k^{-\alpha}$ for $k \to \infty$, $\widetilde{E}(k) \propto k^{-\alpha}$ if $\alpha < 2M + 1$, where $M$ denotes the number of vanishing moments of the wavelets. Consequently, we find for $S_2(a)$ that $S_2(a) \propto a^{\zeta(p)} = (k_\psi/k)^{\zeta(p)}$ for $a \to 0$ if $\zeta(2) \leq 2M$. For the DOD wavelet, we have $M = 1$, therefore, the second-order structure function can only detect slopes smaller than 2, corresponding to an energy spectrum whose slope is shallower than $-3$. Thus, the usual structure functions give spurious results for sufficiently smooth signals. The relation between structure functions and wavelet coefficients can be generalized in the context of Besov spaces, which are classically used for nonlinear approximation theory (*see* Wavelets: Mathematical Theory).

#### Intermittency Measures

Intermittency is defined as localized bursts of high-frequency activity. This means that intermittent phenomena are localized in both physical and spectral spaces, and thus a suitable basis for representing intermittency should reflect this dual localization. The Fourier basis is well localized in spectral space, but delocalized in physical space. Therefore, when a turbulence signal is filtered using a high-pass Fourier transform and then reconstructed in physical space, for example, to calculate the flatness, some spatial information is lost. This leads to smoothing of strong gradients and spurious oscillations in the background, which come from the fact that the modulus and phase of the discarded high wavenumber Fourier modes have been lost. The spatial errors introduced by such a Fourier filtering lead to errors in estimating the flatness, and hence the signal's intermittency.

When a quantity (e.g., velocity derivative) is intermittent, it contains rare but strong events (i.e., bursts of intense activity), which correspond to large deviations reflected in the "heavy tails" of the PDF. Second-order statistics (e.g., energy spectrum, second-order structure function) are relatively insensitive to such rare events whose time or space supports are very small and thus do not dominate the integral. However, these events become increasingly important for higher-order statistics, where they finally dominate. High-order statistics therefore characterize intermittency. Of course, intermittency is not essential for all problems: second-order statistics are sufficient to measure dispersion (dominated by energy-containing scales), but not to calculate drag or mixing (dominated by vorticity production in thin boundary or shear layers).

To measure intermittency, one uses the space–scale information contained in the wavelet coefficients to define scale-dependent moments and moment ratios. Useful diagnostics to quantify the intermittency of a field $f$ are the moments of its wavelet coefficients at different scales $j$

$$M_{p,j}(f) = 2^{-j} \sum_{i=0}^{2^j-1} |\widetilde{f}_{j,i}|^p \qquad [13]$$

Note that the distribution of energy scale by scale, that is, the scalogram, can be computed from the second-order moment of the orthogonal wavelet coefficients: $E_j = 2^{j-1} M_{2,j}$. Due to orthogonality of the decomposition, the total energy is just the sum: $E = \sum_{j \geq 0} E_j$.

The sparsity of the wavelet coefficients at each scale is a measure of intermittency, and it can be quantified using ratios of moments at different scales

$$Q_{p,q,j}(f) = \frac{M_{p,j}(f)}{(M_{q,j}(f))^{p/q}} \qquad [14]$$

which may be interpreted as quotient norms computed in two different functional spaces, $L^p$-and $L^q$-spaces. Classically, one chooses $q = 2$ to define typical statistical quantities as a function of scale. Recall that for $p = 4$ we obtain the scale-dependent flatness $F_j = Q_{4,2,j}$. It is equal to 3 for a Gaussian white noise at all scales $j$, which proves that this signal is not intermittent. The scale-dependent skewness, hyperflatness, and hyperskewness are obtained for $p = 3, 5$, and 6, respectively. For intermittent signals $Q_{p,q,j}$ increases with $j$, whatever $p$ and $q$.

## Wavelet Compression

#### Principle

To study turbulent signals, we now propose to separate the rare and extreme events from the dense events, and then calculate their statistics independently. A major difficulty in turbulence research is that there is no clear scale separation between these two kinds of events. This lack of "spectral gap" excludes Fourier filtering for disentangling these two behaviors. Since the rare events are well

localized in physical space, one might try to use an on–off filter defined in physical space to extract them. However, this approach changes the spectral properties by introducing spurious discontinuities, adding an artificial scaling (e.g., $k^{-2}$ in one dimension) to the energy spectrum. To avoid these problems, we use the wavelet representation, which combines both physical and spectral space localizations (bounded from below by Heisenberg's uncertainty principle). In turbulence, the relevant rare events are the coherent vortices and the dense events correspond to the residual background flow. We have proposed a nonlinear wavelet filtering of the wavelet coefficients of vorticity to extract the coherent vortices out of turbulent flows. We now detail the different steps of this procedure.

### Extraction of Coherent Structures

**Principle** We propose a new method to extract coherent structures from turbulent flows, as encountered in fluids (e.g., vortices, shocklets) or plasmas (e.g., bursts), in order to study their role in transport and mixing.

We first replace the Fourier representation by the wavelet representation, which keeps track of both time and scale, instead of frequency only. The second improvement consists in changing our viewpoint about coherent structures. Since there is not yet a universal definition of coherent structures, we prefer starting from a minimal but more consensual statement about them, that everyone hopefully could agree with: "coherent structures are not noise." Using this apophatic method, we propose the following definition: "coherent structures are what remain after denoising."

For the noise we use the mathematical definition stating that a noise cannot be compressed in any functional basis. Another way to say this is to observe that the shortest description of a noise is the noise itself. Notice that often one calls "noise" what is actually "experimental noise," but not noise in the mathematical sense.

Considering our definition of coherent structures, turbulent signals can be split into two contributions: coherent bursts, corresponding to that part of the signal which can be compressed in a wavelet basis, and incoherent noise, corresponding to that part of the signal which cannot be compressed, neither in wavelets nor in any other basis. We will then check *a posteriori* that the incoherent contribution is spread, and therefore does not compress, in both Fourier and grid-point basis. Since we use the orthogonal wavelet representation, both coherent and incoherent components are orthogonal and therefore the $L^2$-norm, for example, energy or enstrophy, is a superposition of coherent and incoherent contributions (Mallat 1998).

Assuming that coherent structures are what remain after denoising, we need a model, not for the structures themselves, but for the noise. As a first guess, we choose the simplest model and suppose the noise to be additive, Gaussian and white, that is, uncorrelated. Having this model in mind, we use Donoho and Johnstone's theorem to compute the value to threshold the wavelet coefficients. Since the threshold value depends on the variance of the noise, which in the case of turbulence is not *a priori* known, we propose a recursive method to estimate it from the variance of the weakest wavelet coefficients, that is, those whose modulus is below the threshold value.

**Wavelet decomposition** We describe the wavelet algorithm to extract coherent vortices out of turbulent flows and apply it as example to a 3D turbulent flow. We consider the vorticity field $\boldsymbol{\omega} = \nabla \times \boldsymbol{v}$, computed at resolution $N = 2^{3J}, N$ being the number of grid points and $J$ the number of octaves in each spatial direction. Each vorticity component is developed into an orthogonal wavelet series from the largest scale $l_{\max} = 2^0$ to the smallest scale $l_{\min} = 2^{J-1}$ using a three-dimensional (3D) MRA:

$$\omega(\boldsymbol{x}) = \bar{\omega}_{0,0,0}\, \phi_{0,0,0}(\boldsymbol{x}) + \sum_{j=0}^{J-1} \sum_{i_x=0}^{2^j-1} \sum_{i_y=0}^{2^j-1} \sum_{i_z=0}^{2^j-1} \sum_{d=1}^{7} \tilde{\omega}^d_{j,i_x,i_y,i_z} \psi^d_{j,i_x,i_y,i_z}(\boldsymbol{x}) \qquad [15]$$

with $\phi_{j,i_x,i_y i,i_z}(\boldsymbol{x}) = \phi_{j,i_x}(x)\phi_{j,i_y}(y)\phi_{j,i_z}(z)$, and

$$\psi^d_{j,i_x,i_y,i_z}(\boldsymbol{x}) = \begin{cases} \psi_{j,i_x}(x)\phi_{j,i_y}(y)\phi_{j,i_z}(z) & d=1 \\ \phi_{j,i_x}(x)\psi_{j,i_y}(y)\phi_{j,i_z}(z) & d=2 \\ \phi_{j,i_x}(x)\phi_{j,i_y}(y)\psi_{j,i_z}(z) & d=3 \\ \psi_{j,i_x}(x)\phi_{j,i_y}(y)\psi_{j,i_z}(z) & d=4 \\ \psi_{j,i_x}(x)\psi_{j,i_y}(y)\phi_{j,i_z}(z) & d=5 \\ \phi_{j,i_x}(x)\psi_{j,i_y}(y)\psi_{j,i_z}(z) & d=6 \\ \psi_{j,i_x}(x)\psi_{j,i_y}(y)\psi_{j,i_z}(z) & d=7 \end{cases} \qquad [16]$$

where $\phi_{j,i}$ and $\psi_{j,i}$ are the one-dimensional scaling function and the corresponding wavelet, respectively. Due to orthogonality, the scaling coefficients are given by $\bar{\omega}_{0,0,0} = \langle \omega, \phi_{0,0,0} \rangle$ and the wavelet coefficients are given by $\tilde{\omega}^d_{j,i_x,i_y,i_z} = \langle \omega, \psi^d_{j,i_x,i_y,i_z} \rangle$, where $\langle \cdot,\cdot \rangle$ denotes the $L^2$-inner product.

**Nonlinear thresholding** The vorticity field is then split into $\boldsymbol{\omega}_C$ and $\boldsymbol{\omega}_I$ by applying a nonlinear thresholding to the wavelet coefficients. The threshold is defined as $\epsilon = (\frac{4}{3} Z \ln N)^{1/2}$. It only depends on the total enstrophy $Z = \frac{1}{2}\int |\boldsymbol{\omega}|^2 \mathrm{d}\boldsymbol{x}$ and on the number of grid points $N$ without any adjustable parameter. The choice of this threshold is based on theorems by Donoho and Johnstone proving optimality of the wavelet representation to denoise signals in the presence of Gaussian white noise, since this wavelet-based estimator minimizes the maximal $L^2$-error for functions with inhomogeneous regularity (Mallat 1998).

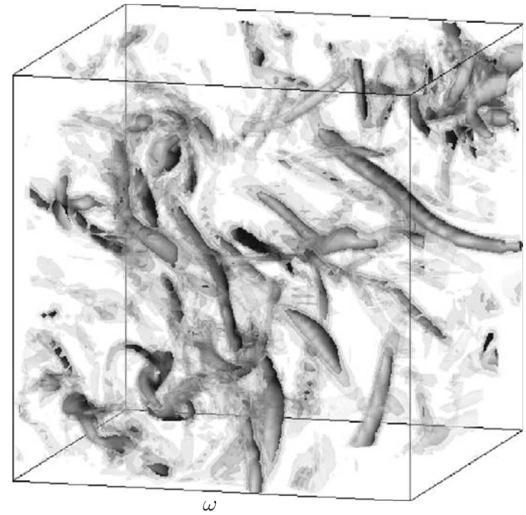
**Wavelet reconstruction** The coherent vorticity field $\boldsymbol{\omega}_C$ is reconstructed from the wavelet coefficients whose modulus is larger than $\epsilon$ and the incoherent vorticity field $\boldsymbol{\omega}_I$ from the wavelet coefficients whose modulus is smaller or equal to $\epsilon$. The two fields thus obtained, $\boldsymbol{\omega}_C$ and $\boldsymbol{\omega}_I$, are orthogonal, which ensures a separation of the total enstrophy into $Z = Z_C + Z_I$ because the interaction term $\langle \boldsymbol{\omega}_C, \boldsymbol{\omega}_I \rangle$ vanishes. We then use Biot–Savart's relation $\boldsymbol{v} = \nabla \times (\nabla^{-2}\boldsymbol{\omega})$ to reconstruct the coherent velocity $\boldsymbol{v}_C$ and the incoherent velocity $\boldsymbol{v}_I$ from the coherent and incoherent vorticities, respectively.

### Application to 3D Turbulence

We consider a 3D homogeneous isotropic turbulent flow, computed by DNS at resolution $N = 256^3$, which corresponds to a Reynolds number based on the Taylor microscale $R_\lambda = 168$ (Farge *et al.* 2003). The computation uses a pseudospectral code, with a Gaussian random vorticity field as initial condition, and the flow evolution is integrated until a statistically stationary state is reached. Figure 2 shows the modulus of the vorticity fluctuations of the total flow, zooming on a $64^3$ subcube to enhance structural details. The flow exhibits elongated, distorted, and folded vortex tubes, as observed in laboratory and numerical experiments.

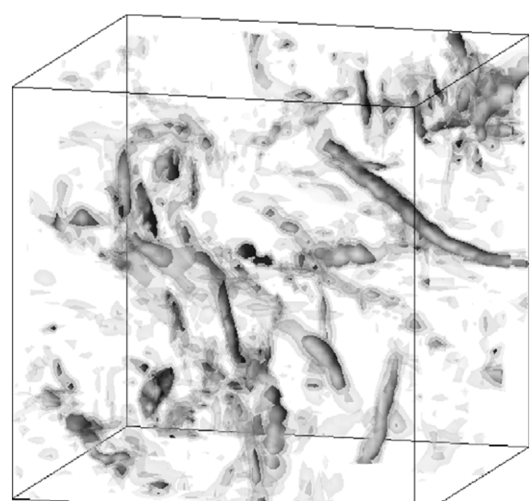
We apply to the total flow the wavelet compression algorithm described above. We find that only 2.9% wavelet modes correspond to the coherent flow, which retains 79% of the energy ($L^2$-norm of velocity) and 75% of the enstrophy ($L^2$-norm of vorticity), while the remaining 97.1% incoherent modes contain only 1% of the energy and 21% of the enstrophy. We display the modulus of the coherent (Figure 3) and incoherent (Figure 4) vorticity fluctuations resulting from the wavelet decomposition.

Note that the values of the three isosurfaces chosen for visualization ($|\omega| = 6Z^{1/2}$, $8Z^{1/2}$ and $10Z^{1/2}$, with $Z$ the total enstrophy) are the same for the total and coherent vorticities, but they have been reduced by a factor 2 for the incoherent vorticity whose fluctuations are much smaller. In the coherent vorticity (Figure 3) we recognize the same vortex tubes as those present in the total vorticity (Figure 2). In contrast, the remaining vorticity (Figure 4) is much more homogeneous and



$\omega$

**Figure 2** Isosurfaces of total vorticity field, for $|\boldsymbol{\omega}| = 3\sigma, 4\sigma, 5\sigma$ with opacity 1, 0.5, 0.1, respectively, and $\sigma^2$ the total enstrophy. Simulation with resolution $N = 256^3$ for $R_\lambda = 168$. Zoom on a subcube $64^3$. Reprinted with permission from Farge *et al.* Coherent vortex extraction in three-dimensional homogeneous turbulence: Comparison between CVS-wavelet and POD-Fourier decompositions. *Physics of Fluids* 15(10): 2886–2896. Copyright 2003, American Institute of Physics.



$\omega$>

**Figure 3** Isosurfaces of coherent vorticity field, for $|\boldsymbol{\omega}| = 3\sigma, 4\sigma, 5\sigma$ with opacity 1, 0.5, 0.1, respectively. Simulation with resolution $N = 256^3$. Zoom on a subcube $64^3$. Reprinted with permission from Farge *et al.* Coherent vortex extraction in three-dimensional homogeneous turbulence: Comparison between CVS-wavelet and POD-Fourier decompositions. *Physics of Fluids* 15(10): 2886–2896. Copyright 2003, American Institute of Physics.

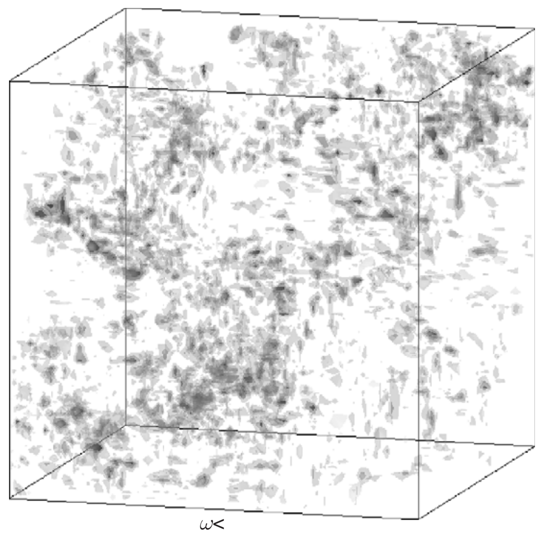



**Figure 4** Isosurfaces of incoherent vorticity field, for $|\boldsymbol{\omega}| = 3/2\sigma, 2\sigma, 5/2\sigma$ with opacity 1, 0.5, 0.1, respectively. Simulation with resolution $N = 256^3$. Zoom on a subcube $64^3$. Reprinted with permission from Farge *et al.* Coherent vortex extraction in three-dimensional homogeneous turbulence: Comparison between CVS-wavelet and POD-Fourier decompositions. *Physics of Fluids* 15(10): 2886–2896. Copyright 2003, American Institute of Physics.

does not exhibit coherent structures. Hence, the wavelet compression retains all the vortex tubes and preserves their structure at all scales. Consequently, the coherent flow is as intermittent as the total flow, while the incoherent flow is structureless and non intermittent. Modeling the effect of the incoherent flow onto the coherent flow should then be much simpler than with methods based on Fourier filtering.

Figure 5 shows the velocity PDF in semilogarithmic coordinates. We observe that the coherent velocity has

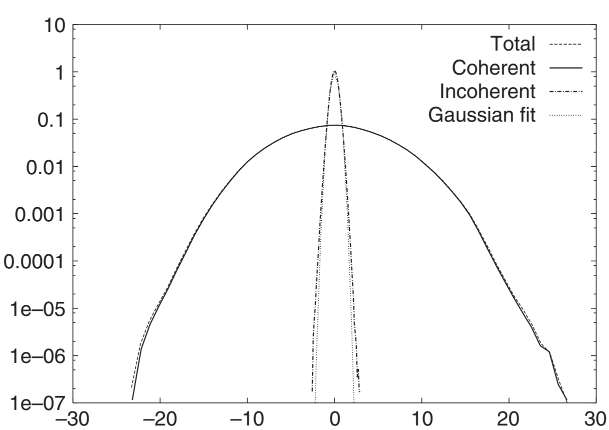



**Figure 5** Velocity PDF, resolution $N = 256^3$ with a zoom at $64^3$. Reprinted with permission from Farge *et al.* Coherent vortex extraction in three-dimensional homogeneous turbulence: Comparison between CVS-wavelet and POD-Fourier decompositions. *Physics of Fluids* 15(10): 2886–2896. Copyright 2003, American Institute of Physics.

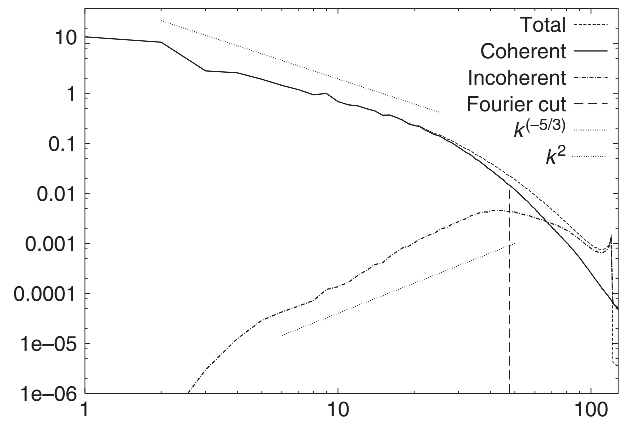



**Figure 6** Energy spectrum, resolution $N = 256^3$ with a zoom at $64^3$. Reprinted with permission from Farge *et al.* Coherent vortex extraction in three-dimensional homogeneous turbulence: Comparison between CVS-wavelet and POD-Fourier decompositions. *Physics of Fluids* 15(10): 2886–2896. Copyright 2003, American Institute of Physics.

the same Gaussian distribution as the total velocity, while the incoherent velocity remains Gaussian, but its variance is much smaller. The corresponding energy spectra are plotted on Figure 6. We observe that the spectrum of the coherent energy is identical to the spectrum of the total energy all along the inertial range. This implies that the vortex tubes are responsible for the $k^{-5/3}$ energy scaling, which corresponds to a long-range correlation, characteristic of 3D turbulence as predicted by Kolmogorov's theory. In contrast, the incoherent energy has a scaling close to $k^2$, which corresponds to an energy equipartition between all wave vectors $\boldsymbol{k}$, since the isotropic spectrum is obtained by integrating energy in 3D $\boldsymbol{k}$-space over 2D shells $k = |\boldsymbol{k}|$. The incoherent velocity field is therefore spatially uncorrelated, which is consistent with the observation that incoherent vorticity is structureless and homogeneous.

From these observations, we propose the following scenario to interpret the turbulent cascade: the coherent energy injected at large scales is transferred towards small scales by nonlinear interactions between vortex tubes. In the meantime, these nonlinear interactions also produce incoherent energy at all scales, which is dissipated at the smallest scales by molecular kinematic viscosity. Thus, the coherent flow causes direct transfer of the coherent energy into incoherent energy. Conversely, the incoherent flow does not trigger any energy transfer to the coherent flow, as it is structureless and uncorrelated. We conjecture that the coherent flow is dynamically active, while the incoherent flow is slaved to it, being only passively advected and mixed by the coherent vortex tubes. This is a different view from the classical interpretation since it does not suppose any scale separation. Both

coherent and incoherent flows are active all along the inertial range, but they are characterized by different probability distribution functions and correlations: non-Gaussian and long-range correlated for the former, while Gaussian and uncorrelated for the latter.

## Wavelet Computation

### Principle

The mathematical properties of wavelets (*see* Wavelets: Mathematical Theory) motivate their use for solving of partial differential equations (PDEs).

The localization of wavelets, both in scale and space, leads to effective sparse representations of functions and pseudodifferential operators (and their inverse) by performing nonlinear thresholding of the wavelet coefficients of the function and of the matrices representing the operators. Wavelet coefficients allow to estimate the local regularity of solutions of PDEs and thus can define autoadaptive discretizations with local mesh refinements. The characterization of function spaces in terms of wavelet coefficients and the corresponding norm equivalences lead to diagonal preconditioning of operators in wavelet space.

Moreover, the existence of the fast wavelet transform yields algorithms with optimal linear complexity. The currently existing algorithms can be classified in different ways. We can distinguish between Galerkin, collocation, and hybrid schemes. Hybrid schemes combine classical discretizations, for example, finite differences or finite volumes, and wavelets, which are only used to speed up the linear algebra and to define adaptive grids. On the other hand, Galerkin and collocation schemes employ wavelets directly for the discretization of the solution and the operators. Wavelet methods have been developed to solve Burger's, Stokes, Kuramoto–Sivashinsky, nonlinear Schrödinger, Euler, and Navier–Stokes equations. As an example, we present an adaptive wavelet algorithm, of Galerkin type, to solve the 2D Navier–Stokes equations.

### Adaptive Wavelet Scheme

We consider the 2D Navier–Stokes equations written in terms of vorticity $\omega$ and stream function $\Psi$, which are both scalars in two dimensions,

$$\partial_t \omega + \boldsymbol{v} \cdot \nabla \omega - \nu \nabla^2 \omega = \nabla \times \boldsymbol{F} \qquad [17]$$

$$\nabla^2 \Psi = \omega \quad \text{and} \quad \boldsymbol{v} = \nabla^{\perp} \Psi \qquad [18]$$

for $\boldsymbol{x} \in [0,1]^2, t > 0$. The velocity is denoted by $\boldsymbol{v}, \boldsymbol{F}$ is an external force, $\nu > 0$ is the molecular kinematic viscosity, and $\nabla^{\perp} = (-\partial_y, \partial_x)$.

The above equations are completed with boundary conditions and a suitable initial condition.

**Time discretization** Introducing a classical semi-implicit time discretization with a time step $\Delta t$ and setting $\omega^n(\boldsymbol{x}) \approx \omega(\boldsymbol{x}, n\Delta t)$, we obtain

$$(1 - \nu \Delta t \nabla^2)\omega^{n+1} = \omega^n + \Delta t(\nabla \times F^n - \boldsymbol{v}^n \cdot \nabla \omega^n) \qquad [19]$$

$$\nabla^2 \Psi^{n+1} = \omega^{n+1} \quad \text{and} \quad \boldsymbol{v}^{n+1} = \nabla^{\perp} \Psi^{n+1} \qquad [20]$$

Hence, in each time step two elliptic problems have to be solved and a differential operator has to be applied.

Formally the above equations can be written in the abstract form $Lu = f$, where $L$ is an elliptic operator with constant coefficients. This corresponds to a Helmholtz type equation for $\omega$ with $L = (1 - \nu \Delta t \nabla^2)$ and a Poisson equation for $\Psi$ with $L = \nabla^2$.

**Spatial discretization** For the spatial discretization, we use the method of weighted residuals, that is, a Petrov–Galerkin scheme. The trial functions are orthogonal wavelets $\phi$ and the test functions are operator adapted wavelets, called "vaguelettes," $\theta$. To solve the elliptic equation $Lu = f$ at time step $t^{n+1}$, we develop $u^{n+1}$ into an orthogonal wavelet series, that is, $u^{n+1} = \sum_\lambda \widetilde{u}_\lambda^{n+1} \psi_\lambda$, where $\lambda = (j, i_x, i_y, d)$ denotes the multi-index for scale $j$, space $i$, and direction $d$. Requiring that the residual vanishes with respect to all test functions $\theta_\lambda$, we obtain a linear system for the unknown wavelet coefficients $\widetilde{u}_\lambda^{n+1}$ of the solution $u$:

$$\sum_\lambda \widetilde{u}_\lambda^{n+1} \langle L\psi_\lambda, \theta_{\lambda'} \rangle = \langle f, \theta_{\lambda'} \rangle \qquad [21]$$

The test functions $\theta$ are defined such that the stiffness matrix turns out to be the identity. Therefore, the solution of $Lu = f$ reduces to a change of basis, that is, $u^{n+1} = \sum_\lambda \langle f, \theta_\lambda \rangle \psi_\lambda$. The right-hand side (RHS) $f$ can then be developed into a biorthogonal operator adapted wavelet basis $f = \sum_\lambda \langle f, \theta_\lambda \rangle \zeta_\lambda$, with $\theta_\lambda = L^{\star -1} \psi_\lambda$ and $\zeta_\lambda = L\psi_\lambda$, $^\star$ denoting the adjoint operator. By construction, $\theta$ and $\zeta$ are biorthogonal, that is, such that $\langle \theta_\lambda, \zeta_{\lambda'} \rangle = \delta_{\lambda, \lambda'}$. It can be shown that both have similar localization properties in physical and Fourier space as $\psi$, and that they form a Riesz basis.

**Adaptive discretization** To get an adaptive space discretization for the linear problem $Lu = f$, we

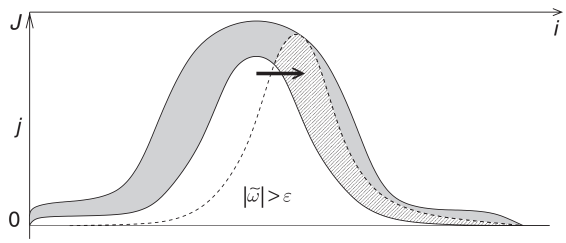



**Figure 7** Illustration of the dynamic adaption strategy in wavelet coefficient space.

consider only the significant wavelet coefficients of the solution. Hence, we only retain coefficients $\widetilde{u}^n_\lambda$ whose modulus is larger than a given threshold $\varepsilon$, that is, $|\widetilde{u}^n_\lambda| > \varepsilon$. The corresponding coefficients are shown in Figure 7 (white area under the solid line curve).

**Adaption strategy** To be able to integrate the equation in time we have to account for the evolution of the solution in wavelet coefficient space (indicated by the arrow in Figure 7). Therefore, we add at time step $t^n$ the neighbors to the retained coefficients, which constitute a security zone (gray area in Figure 7). The equation is then solved in this enlarged coefficient set (white and gray areas below the curves in Figure 7) to obtain $\widetilde{u}^{n+1}_\lambda$. Subsequently, we threshold the coefficients and retain only those whose modulus $|\widetilde{u}^{n+1}_\lambda| > \varepsilon$ (coefficients under the dashed curve in Figure 7). This strategy is applied in each time step and hence allows to automatically track the evolution of the solution in both scale and space.

**Evaluation of the nonlinear term** For the evaluation of the nonlinear term $f(u^n)$, where the wavelet coefficients $\widetilde{u}^n$ are given, there are two possibilities:

- *Evaluation in wavelet coefficient space*. As illustration, we consider a quadratic nonlinear term, $f(u) = u^2$. The wavelet coefficients of $f$ can be calculated using the connection coefficients, that is, one has to calculate the bilinear expression, $\sum_\lambda \sum_{\lambda'} \widetilde{u}_\lambda \mathcal{I}_{\lambda\lambda'\lambda''} \widetilde{u}_{\lambda'}$ with the interaction tensor $\mathcal{I}_{\lambda\lambda'\lambda''} = \langle \psi_\lambda \psi_{\lambda'}, \theta_{\lambda''} \rangle$. Although many coefficients of $\mathcal{I}$ are zero or very small, the size of $\mathcal{I}$ leads to a computation which is quite untractable in practice.
- *Evaluation in physical space*. This approach is similar to the pseudospectral evaluation of the nonlinear terms used in spectral methods, therefore it is called pseudowavelet technique. The advantage of this scheme is that general nonlinear terms, for example, $f(u) = (1-u)\,e^{-C/u}$, can be treated more easily. The method can be summarized as follows: starting from the significant wavelet coefficients, $|\widetilde{u}_\lambda| > \varepsilon$, one reconstructs $u$ on a locally refined grid and gets $u(x_\lambda)$. Then one can evaluate $f(u(x_\lambda))$ pointwise and the wavelet coefficients $\widetilde{f}_\lambda$ are calculated using the adaptive decomposition.

Finally, one computes the scalar products of the RHS of [21] with the test functions $\theta$ to advance the solution in time. We compute $\widetilde{u}_\lambda = \langle f, \theta_\lambda \rangle$ belonging to the enlarged coefficient set (white and gray regions in Figure 7).

The algorithm is of $O(N)$ complexity, where $N$ denotes the number of wavelet coefficients retained in the computation.

### Application to 2D Turbulence

To illustrate the above algorithm we present an adaptive wavelet computation of a vortex dipole in a square domain, impinging on a no-slip wall at Reynolds number $Re = 1000$. To take into account the solid wall, we use a volume penalization method, for which both the fluid flow and the solid container are modeled as a porous medium whose porosity tends towards zero in the fluid and towards infinity in the solid region.

The 2D Navier–Stokes equations are thus modified by adding the forcing term $\boldsymbol{F} = -(1/\eta)\chi \boldsymbol{v}$ in eqn [18], where $\eta$ is the penalization parameter and $\chi$ is the characteristic function whose value is 1 in the solid region and 0 elsewhere. The equations are solved using the adaptive wavelet method in a periodic square domain of size 1.1, in which the square container of size 1 is imbedded, taking $\eta = 10^{-3}$. The maximal resolution corresponds to a fine grid of $1024^2$ points. Figure 8a shows snapshots of the vorticity field at times $t = 0.2, 0.4, 0.6,$ and $0.8$ (in arbitrary units). We observe that the vortex dipole is moving towards the wall and that strong vorticity gradients are produced when the dipole hits the wall. The computational grid is dynamically adapted during the flow evolution, since the nonlinear wavelet filter automatically refines the grid in regions where strong gradients develop. Figure 8b shows the centers of the retained wavelet coefficients at corresponding times.

Note that during the computation only 5% out of $1024^2$ wavelet coefficients are used. The time evolution of total kinetic energy and the total enstrophy $\boldsymbol{F} = -(\frac{1}{y})\chi \boldsymbol{v}$, are plotted in Figure 9 to

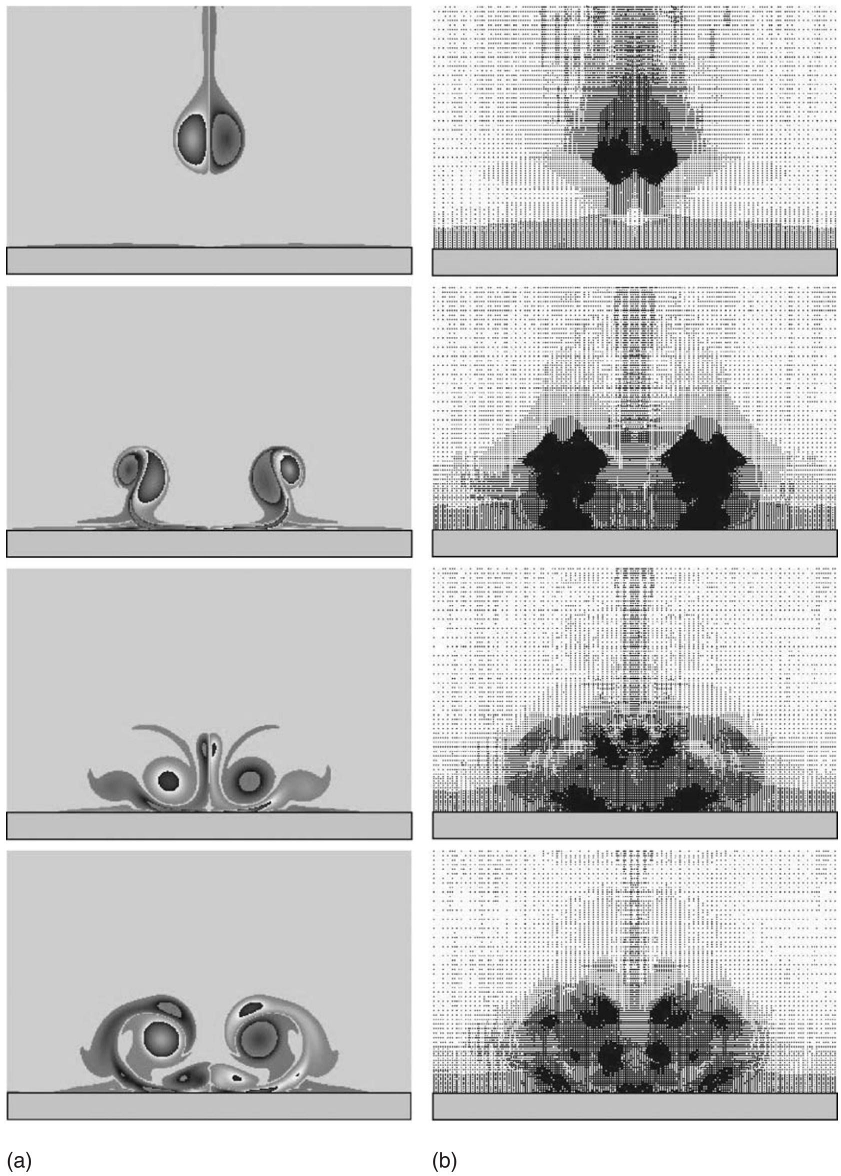


Figure 8 Dipole wall interaction at $Re = 1000$. (a) Vorticity field, (b) corresponding centers of the active wavelets, at $t = 0.2, 0.4, 0.6$, and 0.8 (from top to bottom).

show the production of enstrophy and the concomitant dissipation of energy when the vortex dipole hits the wall.

This computation illustrates the fact that the adaptive wavelet method allows an automatic grid refinement, both in the boundary layers at the wall and also in shear layers which develop during the flow evolution far from the wall. Therewith, the number of grid points necessary for the computation is significantly reduced, and we conjecture that the resulting compression rate will increase with the Reynolds number.

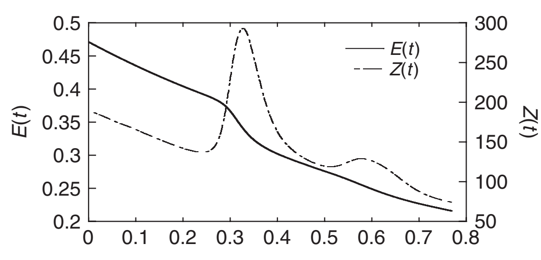



Figure 9 Time evolution of energy (solid line) and enstrophy (dashed line).

## Acknowledgments

Marie Farge thankfully acknowledges Trinity College, Cambridge, UK, and the Centre International de Rencontres Mathématiques (CIRM), Marseille, France, for hospitality while writing this paper.

*See also:* Turbulence Theories; Viscous Incompressible Fluids: Mathematical Theory; Wavelets: Applications; Wavelets: Mathematical Theory.

## Further Reading

Cohen A (2000) Wavelet methods in numerical analysis. In: Ciarlet PG and Lions JL (eds.) *Handbook of Numerical Analysis*, vol. 7, Amsterdam: Elsevier.

Cuypers Y, Maurel A, and Petitjeans P (2003) *Physics Review Letters* 91: 194502.

Dahmen W (1997) Wavelets and multiscale methods for operator equations. *Acta Numerica* 6: 55–228.

Daubechies I (1992) *Ten Lectures on Wavelets*. SIAM.

Daubechies I, Grossmann A, and Meyer Y (1986) *Journal of Mathematical Physics* 27: 1271.

Farge M (1992) Wavelet transforms and their applications to turbulence. *Annual Reviews of Fluid Mechanics* 24: 395–457.

Farge M and Rabreau G (1988) *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences, Paris* 2: 307.

Farge M, Kevlahan N, Perrier V, and Goirand E (1996) Wavelets and turbulence. *Proceedings of the IEEE* 84(4): 639–669.

Farge M, Kevlahan N, Perrier V, and Schneider K (1999) Turbulence analysis, modelling and computing using wavelets. In: van den Berg JC (ed.) *Wavelets in Physics*, pp. 117–200. Cambridge: Cambridge University Press.

Farge M and Schneider K (2002) Analysing and computing turbulent flows using wavelets. In: Lesieur M, Yaglom A, and David F (eds.) *New trends in turbulence*, Les Houches 2000, vol. 74, pp. 449–503. Springer.

Farge M, Schneider K, Pellegrino G, Wray AA, and Rogallo RS (2003) *Physics of Fluids* 15(10): 2886.

Grossmann A and Morlet J (1984) *SIAM J. Appl. Anal* 15: 723.

Lemarié P-G and Meyer Y (1986) *Revista Matematica Iberoamericana* 2: 1.

Mallat S (1989) *Transactions of the American Mathematical Society* 315: 69.

Mallat S (1998) *A Wavelet Tour of Signal Processing*. Academic Press.

Schneider K, Farge M, and Kevlahan N (2004) Spatial intermittency in two-dimensional turbulence. In: Tongring N and Penner RC (eds.) *Woods Hole Mathematics, Perspectives in Mathematics and Physics*, pp. 302–328. Word Scientific.

http://wavelets.ens.fr – other papers about wavelets and turbulence can be downloaded from this site.

# Wavelets: Applications

**M Yamada**, Kyoto University, Kyoto, Japan



## Introduction

Wavelet analysis was first developed in the early 1980s in the field of seismic signal analysis in the form of an integral transform with a localized kernel function with continuous parameters of dilation and translation. When a seismic wave or its derivative has a singular point, the integral transform has a scaling property with respect to the dilation parameter; thus, this scaling behavior can be available to locate the singular point. In the mid-1980s, the orthonormal smooth wavelet was first constructed, and later the construction method was generalized and reformulated as multiresolution analysis (MRA). Since then, several kinds of wavelets have been proposed for various purposes, and the concept of wavelet has been extended to new types of basis functions. In this sense, the most important effect of wavelets may be that they have awakened deep interest in bases employed in data analysis and data processing. Wavelets are now widely used in various fields of research; some of their applications are discussed in this article.

From the perspective of time–frequency analysis, the wavelet analysis may be regarded as a windowed Fourier analysis with a variable window width, narrower for higher frequency. The wavelets can therefore give information on the local frequency structure of an event; they have been applied to various kinds of one-dimensional (1D) or multidimensional signals, for example, to identify an event or to denoise or to sharpen the signal.

1D wavelets $\psi^{(a,b)}(x)$ are defined as

$$\psi^{(a,b)}(x) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{x-b}{a}\right)$$

where $a(\neq 0), b$ are real parameters and $\psi(x)$ is a spatially localized function called "analyzing wavelet" or "mother wavelet." Wavelet analysis gives a decomposition of a function into a linear combination of those wavelets, where a perfect reconstruction requires the analyzing wavelet to satisfy some mathematical conditions.

For the continuous wavelet transform (CWT), where the parameters $(a, b)$ are continuous, the analyzing wavelet $\psi(x)L^2(\boldsymbol{R})$ has to satisfy the admissibility condition

analyzing wavelet $\psi(x)L^2(\boldsymbol{R})$ has to satisfy the admissibility condition

$$C_\psi \equiv \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|}\, \mathrm{d}\omega < \infty$$

where $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(x)$:

$$\hat{\psi}(\omega) = \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\omega x}\psi(x)\,\mathrm{d}x$$

The admissibility condition is known to be equivalent to the condition that $\psi(x)$ has no zero-frequency component, that is, $\hat{\psi}(0)=0$, under some mild condition for the decay rate at infinity. Then the CWT and its inverse transform of a data function $f(x) \in L^2(\boldsymbol{R})$ is defined as

$$T_\psi(a,b) = \frac{1}{\sqrt{C_\psi}} \int_{-\infty}^{\infty} \overline{\psi^{(a,b)}(x)} f(x)\,\mathrm{d}x$$

$$f(x) = \frac{1}{\sqrt{C_\psi}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} T_\psi(a,b)\psi^{(a,b)}(x)\,\frac{\mathrm{d}a\,\mathrm{d}b}{a^2}$$

In the case of the discrete wavelet transform (DWT), the parameters $(a,b)$ are taken discrete; a typical choice is $a = 1/2^j, b = k/2^j$, where $j$ and $k$ are integers:

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$$

In order that the wavelets $\{\psi_{j,k}(x)\,|\,j,k \in \boldsymbol{Z}\}$ may constitute a complete orthonormal system in $L^2(\boldsymbol{R})$, the analyzing wavelet should satisfy more stringent conditions than the admissibility condition for the CWT, and is now constructed in the framework of MRA. A data function is then decomposed by the DWT as

$$f(x) = \sum_{j=-\infty}^{\infty} \alpha_{j,k}\psi_{j,k}(x), \quad \alpha_{j,k} = \int_{-\infty}^{\infty} \overline{\psi_{j,k}(x)} f(x)\,\mathrm{d}x$$

Even when the discrete wavelets do not constitute a complete orthnormal system, they often form a wavelet frame if linear combinations of the wavelets are dense in $L^2(\boldsymbol{R})$ and if there are two constants $A$, $B$ such that the inequality

$$A\|f\|^2 \le \sum_{j,k} |\langle \psi_{j,k}, f\rangle|^2 \le B\|f\|^2$$

holds for an arbitrary $f(x) \in L^2(\boldsymbol{R})$. For the wavelet frame $\{\psi_{j,k}\}$, there is a corresponding dual frame, $\{\tilde{\psi}_{j,k}\}$, which permits the following expansion of $f(x)$:

$$f(x) = \sum_{j,k} \langle \psi_{j,k}, f\rangle \tilde{\psi}_{j,k}(x) = \sum_{j,k} \langle \tilde{\psi}_{j,k}, f\rangle \psi_{j,k}(x)$$

The wavelet frame is also employed in several applications.

From the prospect of applications, the CWTs are better adapted for the analysis of data functions, including the detection of singularities and patterns, while the DWTs are adapted to the data processing, including signal compression or denoising.

## Singularity Detection and Multifractal Analysis of Functions

Since its birth, the wavelet analysis has been applied for the detection of singularity of a data function. Let us define the Hölder exponent $h(x_0)$ at $x_0$ of a function $f(x)$ is defined here as the largest value of the exponent $h$ such that there exists a polynomial $P_n(x)$ of degree $n$ that satisfies for $x$ in the neighborhood of $x_0$:

$$|f(x) - P_n(x - x_0)| = \mathrm{O}(|x - x_0|^h)$$

The data function is not differentiable if $h(x_0) < 1$, but if $h(x_0) > 1$ then it is differentiable and a singularity may arise in its higher derivatives. The wavelet transform is applied to find the Hölder exponent $h(x_0)$, because $T_\psi(a,b)$ has an asymptotic behavior $T_\psi(a,b) = \mathrm{O}(a^{h(x_0+1/2)})(a \to 0)$ if the analyzing wavelet has $N(>h(x_0))$ vanishing moments, that is,

$$\int_{-\infty}^{\infty} x^m \psi(x)\,\mathrm{d}x = 0, \quad m \in \boldsymbol{Z}, 0 \le m < N$$

A commonly used analyzing wavelet for this purpose may be the $N$-time derivative of the Gaussian function $\psi(x) = \mathrm{d}^N(\mathrm{e}^{-x^2/2})/\mathrm{d}x^N$. This method works well to examine a single or some finite number of singular points of the data function.

When the data function is a multifractal function with an infinite number of singular point of various strengths, the multifractal property of the data function is often characterized by the singularity spectrum $D(h)$ which denotes the Hausdorff dimension of the set of points where $h(x) = h$. The singularity spectrum is, however, difficult to obtain directly from the CWT, and the Legendre transformation is introduced to bypass the difficulty.

Fully developed 3D fluid turbulence may be a typical example of wavelet application to the singularity detection. The Kolmogorov similarity law of fluid turbulence for the longitudinal velocity increment $\Delta u(r) \equiv \boldsymbol{e}\cdot(\boldsymbol{u}(\boldsymbol{x}+r\boldsymbol{e}) - \boldsymbol{u}(\boldsymbol{x}))$, where $\boldsymbol{u}(\boldsymbol{x})$ is the velocity field and $\boldsymbol{e}$ is a constant unit vector,

predicts a scaling property of the structure function; for $r$ in the inertial subrange,

$$\langle (\Delta u(r))^p \rangle \sim r^{\zeta_p}, \quad \zeta_p = p/3$$

where $\langle \cdot \rangle$ denotes the statistical mean. In reality, however, the scaling exponent $\zeta_p$ measured in experiments shows a systematic deviation from $p/3$, which is considered to be a reflection of intermittency, namely the spatial nonuniformity or multifractal property of active vortical motions in turbulence. For simplicity, let us consider the velocity field on a linear section of the turbulence field. According to the multifractal formalism, the turbulence velocity field has singularities of various strengths described by the singularity spectrum $D(h)$, which is related to the scaling exponent $\zeta_p$ through the Legendre transform, $D(h) = \inf_p(ph - \zeta_p + 1)$. This relation is often used to determine $D(h)$ from the knowledge of $\zeta_p$ (structure function method). However, this method does not necessarily work well because, for example, it does not capture the singular points of the Hölder exponent larger than 1 and it is unstable for $h < 0$.

These difficulties are not restricted to the turbulence research, but arise commonly when the structure function is employed to determine the singularity spectrum. In these problems, the CWT $T_\psi(a, b)$ provides an alternative method. An ingenious technique is to take only the modulus maxima of $T_\psi(a, b)$ (for each of fixed $a$) to construct a partition function

$$Z(a, q) = \sum_{l \in L_{\max}} \left[ \sup_{(a, b') \in l} |T_\psi(a, b')| \right]^q$$

where $q \in \boldsymbol{R}$, and $L_{\max}$ denotes the set of all maxima lines, each of which is a continuous curve for small value of $a$, and there exists at least one maxima line toward a singular point of the Hölder exponent $h(x_0) < N$. In the limit of $a \to 0$, defining the exponent $\tau(q)$ as $Z(a, q) \sim a^{\tau(q)}$, one can obtain the singularity spectrum through the Legendre transform:

$$D(h) = \inf_q \left[ q\left(h + \tfrac{1}{2}\right) - \tau(q) \right]$$

This method (wavelet-transform modulus-maxima (WTMM) method) is advantageous in that it works also for singularities of $h > 1$ and $h < 0$. Several simple examples of multifractal functions have been successfully analyzed by this method. For fluid turbulence, this method gives a singularity spectrum $D(h)$ which has a peak value of $\sim 1$ at $h \sim 1/3$, consistently with Kolmogorov similarity law, but has a convex shape around $h = 1/3$ suggesting a multifractal property. For a fractal signal, we note that the WTMM method enlightens the hierarchical organization of the singularities, in the branching structure of the WT skeleton defined by the maxima lines arrangement in the $(a, b)$ half-plane.

Though the above discussion also applies to the DWT, the detection of the Hölder exponent $h$ in experimental situations is usually performed by the CWT, which has no restriction on possible values of $a$, while the DWT is often employed for theoretical discussions of singularity and multifractal structure of a function.

## Multiscale Analysis

Wavelet transform expands a data function in the time–frequency or the position–wavenumber space, which has twice the dimension of the original signal, and makes it easier to perform a multiscale analysis and to identify events involved in the signal. In the wavelet transform, as stated above, the time resolution is higher at higher frequency, in contrast with the windowed Fourier transform where the time and the frequency resolutions are independent of frequency. Another advantage of wavelet is a wide variety of analyzing wavelet, which enables us to optimize the wavelet according to the purpose of data analysis. Both the CWT and the DWT are available for these time–frequency or position–wavenumber analysis. However, the CWT has properties quite different from those of familiar orthonormal bases of discrete wavelets.

### Multidimensional CWT

The CWT can be formulated in an abstract way. We can regard $G = \{(a, b) \mid a(\neq 0), b \in \boldsymbol{R}\}$ as an affine group on $\boldsymbol{R}$ with the group operation of $(a, b)(a', b') = (aa', ab' + b)$ associated with the invariant measure $\mathrm{d}\mu = \mathrm{d}a\,\mathrm{d}b/a^2$. The group $G$ has its unitary representation in the Hilbert space $H = L^2(\boldsymbol{R})$:

$$(U(a, b)f)(x) = \frac{1}{\sqrt{|a|}} f\left(\frac{x - b}{a}\right)$$

and then we can consider the CWT can be constructed as a linear map $W$ from $L^2(\boldsymbol{R})$ to $L^2(G; \mathrm{d}a\,\mathrm{d}b/a^2)$:

$$W : f(x) \mapsto T_\psi(a, b) = \frac{1}{\sqrt{C_\psi}} \langle U(a, b)\psi, f \rangle$$

where $\langle \cdot, \cdot \rangle$ is the inner product of $L^2(\boldsymbol{R})$ with the complex conjugate taken at the first element, and

$\psi(x)$ is a unit vector (analyzing wavelet) satisfying the abstract admissibility condition

$$C_\psi = \int_G |\langle U(a,b)\psi, \psi\rangle|^2 \, d\mu < \infty$$

This formulation is applicable also to a locally compact group $G$ and its unitary and square integrable representation in a Hilbert space $H$. Note that even the canonical coherent states are included in this framework by taking the Weyl–Heisenberg group and $L^2(\boldsymbol{R})$ for $G$ and $H$, respectively. This abstract formulation allows us to extend the CWT to higher-dimensional Euclidean spaces and other manifolds: for example, 2D sphere $S^2$ for geophysical application and 4D manifold of spacetime taking the Poincaré group into consideration.

In $\boldsymbol{R}^n$, the CWT of $f(\boldsymbol{x}) \in L^2(\boldsymbol{R}^n)$ and its inverse transform are given by

$$T_\psi(a, r, \boldsymbol{b}) = \frac{1}{\sqrt{C_\psi}} \int_{\boldsymbol{R}^n} \overline{\psi^{(a,r,\boldsymbol{b})}(\boldsymbol{x})} f(\boldsymbol{x}) \, d\boldsymbol{x}$$

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{C_\psi}} \int_G T(a, r, \boldsymbol{b}) \psi^{(a,r,\boldsymbol{b})}(\boldsymbol{x}) \frac{da \, dr \, d\boldsymbol{b}}{a^{n+1}}$$

where $r \in \mathrm{SO}(n), \boldsymbol{b} \in \boldsymbol{R}^n, dr$ is the normalized invariant measure of $G = \mathrm{SO}(n)$, and the wavelets are defined as $\psi^{(a,\, r,\, b)}(\boldsymbol{x}) = (1/a^{n/2})\psi(r^{-1}(\boldsymbol{x} - \boldsymbol{b})/a)$, with the analyzing wavelet satisfying the admissibility condition

$$C_\psi = \int_{\boldsymbol{R}^n} \frac{|\hat{\psi}(\boldsymbol{\omega})|}{|\boldsymbol{\omega}|^n} \, d\boldsymbol{\omega} < \infty$$

Note that these wavelets are constructed not only by dilation and translation but also by rotation which therefore gives the possibility for directional pattern detection in a data function. In the case of 2D sphere $S^2$, on the other hand, the dilation operation should be reinterpreted in such a way that at the North Pole, for example, it is the normal dilation in the tangent plane followed by lifting it to $S^2$ by the stereographic projection from the South Pole.

Generally, the abstract map $W$ thus defined is injective and therefore reversal, but not surjective in contrast with the Fourier case. Actually in the case of 1D CWT, $T_\psi(a, b)$ is subject to an integral condition:

$$T_\psi(a, b) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{da \, db}{a^2} K(a, b; a', b') T_\psi(a', b')$$

$$K(a, b; a', b') = \int_{-\infty}^{\infty} \overline{\psi^{(a,b)}(x)} \psi^{(a',b')}(x) \, dx$$

which defines the range of the CWT, a subspace of $L^2(\boldsymbol{R})$. Therefore, if one wants to modify $T_\psi(a, b)$ by, for example, assigning its value as zero in some parameter region just as in a filter process, care should be taken for the resultant $T_\psi(a, b)$ to be in the image of the CWT. The reason may be understood intuitively by noticing that the wavelets $\psi^{(a,b)}(x)$ are linearly dependent on each other. The expression of a data function by a linear combination of the wavelets is therefore not unique, and thus is redundant. The CWT gives only $T_\psi(a, b)$ of the least norm in $L^2(\boldsymbol{R}^2; da \, db/a^2)$. In physical interpretations of the CWT, however, this nonuniqueness is often ignored.

## Pattern Detection

**Edge detection** The edges of an object are often the most important components for pattern detection. The edge may be considered to consist of points of sharp transition of image intensity. At the edge, the modulus of the gradient of the image $f(x, y)$ is expected to take a local maximum in the 1D direction perpendicular to the edge. Therefore, the local maxima of $|\nabla f(x, y)|$ may be the indicator of the edge. However, the image textures can also give similar sharp transitions of $f(x, y)$, and one should take into account the scale dependence which distinguishes between edges and textures. One of the practically possible ways for this purpose is to use dyadic wavelets $\psi_j^m(x, y) = 2^j \psi^m(2^j x, 2^j y)$ which are generated from the two wavelets $(\psi^1, \psi^2) = (-\partial\theta/\partial x, -\partial\theta/\partial y)$, where $\theta$ is a localized function (multiscale edge detection method). The dyadic wavelet transform of the image $f(x, y)$

$$T_j^m(b_1, b_2) = \langle f(x, y), \psi_j^m(x - b_1, y - b_2)\rangle, \quad m = 1, 2$$

defines the multiscale edges as a set of points $\boldsymbol{b} = (b_1, b_2)$ where the modulus of the wavelet transform, $|(T_j^1, T_j^2)|$, takes a locally maximum value (WTMM) in a 1D neighborhood of $\boldsymbol{b}$ in the direction of $(T_j^1(\boldsymbol{b}), T_j^2(\boldsymbol{b}))$. Scale dependence of the magnitude of the modulus maxima is related to the Hölder exponent of $f(x, y)$ similarly to 1D case, and thus gives information to distinguish between the edges and the textures.

Inversely, the information of WTMM $\boldsymbol{b}_{j,p} = \{(b_{1,j,p}, b_{2,j,p})\}$ of multiscale edges can be made use of for an approximate reconstruction of the original image, although the perfect reconstruction cannot be expected because of the noncompleteness of the modulus maxima wavelets. Assuming that $\{\psi_{j,p}^1, \psi_{j,p}^2\} = \{\psi_j^1(\boldsymbol{x} - \boldsymbol{b}_{j,p}), \psi_j^2(\boldsymbol{x} - \boldsymbol{b}_{j,p})\}$ constitutes a frame of the linear closed space generated by

$\{\psi^1_{j,p}, \psi^2_{j,p}\}$, an approximate image $\hat{f}$ is obtained by inverting the relation

$$L\hat{f} \equiv \sum_m \sum_{j,p} \langle \hat{f}, \psi^m_{j,p} \rangle \psi^m_{j,p} = \sum_m \sum_{j,p} T^m_j(\boldsymbol{b}_{j,p}) \psi^m_{j,p}$$

using, for example, a conjugate gradient algorithm, where a fast calculation is possible with a filter bank algorithm for the dyadic wavelet ("algorithm à trous"). This algorithm gives only the solution of minimum norm among all possible solutions, but it is often satisfactory for practical purposes and thus is applicable also to data compression.

**Directional detection** For oriented features such as segments or edges in images to be detected, a directionally selective wavelet for the CWT is desired. A useful wavelet for this purpose is one that has the effective support of its Fourier transform in a convex cone with apex at the origin in wave number space. A typical example of the directional wavelet may be the 2D Morlet wavelet:

$$\psi(\boldsymbol{x}) = \exp(\mathrm{i}\boldsymbol{k}_0 \cdot \boldsymbol{x}) \exp(-|A\boldsymbol{x}|^2)$$

where $\boldsymbol{k}_0$ is the center of the support in Fourier space, and $A$ is a $2 \times 2$ matrix diag$[\epsilon^{1/2}, 1](\epsilon \leq 1)$, where the admissibility condition for the CWT is approximately satisfied for $|k_0| \geq 5$. Another example is the Cauchy wavelet which has the support strictly in a convex cone in wave number space.

These wavelets have the directional selectivity with preference to a slender object in a specific direction. One of their applications is the analysis of the velocity field of fluid motion from an experimental data, where many tiny plastic balls distributed in fluid give a lot of line segments in a picture taken with a short exposure. The directional wavelet analysis of the picture classifies the line segments according to their directions, indicating the directions of fluid velocity. Another example may be a wave-field analysis where many waves in different directions are superimposed; the directional wavelets allow one to decompose the wave field into the component waves. Directional wavelets have also been applied successfully to detect symmetry of objects such as crystals or quasicrystals.

**Denoising and separation of signals** The wavelet frame as well as the CWT give a redundant representation of a data function. If, instead of the original data, the redundant expression is transmitted, the redundancy is used to reduce the noise included in the received data because the redundancy requires the data to belong to a subspace, and the projection of the received data to the subspace reduces the noise component orthogonal to it. More specifically, the wavelet frame gives a representation of a data function as $f(t) = \sum_{j,k} \alpha_{j,k}\psi_{j,k}$, where the expansion coefficients $\alpha_{j,k} = \langle \psi_{j,k}, f(x) \rangle$ satisfy the defining equation of the subspace

$$\alpha_{j',k'} = \sum \alpha_{j,k} \langle \psi_{j',k'}, \psi_{j,k} \rangle$$

If the frame coefficients are transmitted, the projection operator $P$, which is defined on the right-hand side of the above equation, reduces the noise in the received coefficients $\alpha_{j,k}$ contaminated during the transmission.

However, this method is not applicable if the transmitted signal is not redundant. Then some *a priori* criterion is necessary to discriminate between signal and noise. Various criteria have been proposed in different fields. If the signal and the noise, or plural signals have different power-law forms of spectra, then their discrimination may be possible by the DWT at higher-frequency region where the difference in the magnitude of the coefficients is significant. In this approach, the wavelets of Meyer type, that is, an orthogonal wavelet with a compact support in Fourier space, may be preferable because the wavelets of different scales are separated, at least to some extent, in Fourier space.

In fluid dynamics, the vorticity field of 2D turbulence is found to be decomposed into coherent and incoherent vorticity fields, according as the CWT is larger than a threshold value or not, respectively. These two fields give different Fourier spectra of the velocity field ($k^{-5}$ for coherent part while $k^{-3}$ for incoherent part), showing that the coherent structures are responsible for the deviation from $k^{-3}$ predicted by the classical enstrophy cascade theory. In an astronomical application, on the other hand, the data processing is performed by a more sophisticated method taking into account interscale relation in the wavelet transform, because an astronomical image contains various kinds of objects, including stars, double-stars, galaxies, nebulas, and clusters. In a medical image however contrast analysis is indispensable for diagnostic imaging to get a clear detailed picture of organic structure. A scale-dependent local contrast is defined as the ratio of the CWT to that given by an analyzing wavelet with a larger support. A multiplicative scheme to improve the contrast is constructed by using the local contrast.

### Signal Compression

Signal compression is quite an important technology in digital communication. Speech, audio, image, and digital video are all important fields of signal

compression, and plenty of compression methods have been put to practical use, but we mention here only a few.

The MRA for orthogonal wavelets gives a successive procedure to decompose a subspace of $L^2(\boldsymbol{R})$ into a direct sum of two subspaces corresponding to higher- and lower-frequency parts; only the latter of which is decomposed again into its higher- and lower-frequency parts. Algebraically, this procedure was already known before the discovery of MRA in filter theory in electrical engineering, where a discretely sampled signal is convoluted with a filter series to give, for example, a high-pass-filtered or low-pass-filtered series. An appropriate designed pair of a high-pass and a low-pass filters followed by the downsampling yields two new series corresponding to the higher- and lower-frequency parts, respectively, which are then reversible by another two reconstruction filters with the upsampling. These four filters which are often employed in a widely used technique of "subband coding" then constitute a perfect reconstruction filter bank. Under some conditions, successive applications of this decomposition process to the series of lower-frequency parts, which is equivalent to the nesting structure of MRA, have been used for data compression (quadrature mirror filter). A famous example is a data compression system of FBI for finger prints, consisting of wavelet coding with scalar quantization.

In MRA, however, it is only the lower-frequency parts that are successively decomposed. If both the lower- and the higher-frequency parts are repeatedly decomposed by the decomposition filters, then the successive convolution processes correspond to a decomposition of data function by a set of wavelet-like functions, called "wavelet packet," where there are choices whether to decompose the higher- and/or the lower-frequency parts. The best wavelet packet, in the sense of the entropy, for example, within a specified number of decompositions, often provides with a powerful tool for data compression in several areas, including speech analysis and image analysis. We also note that from the viewpoint of the best basis which minimizes the statistical mean square error of the thresholded coefficients, an orthonormal wavelet basis gives a good concentration of the energy if the original signal is a piecewise smooth function superimposed by a white noise, which is thus efficiently removed by thresholding the coefficients. The efficiency of a wavelet expansion of a signal is sometimes evaluated with the entropy of "probability" defined as $|\alpha_{j,k}|^2/||f||^2$. A better wavelet can be selected by reducing the entropy, practically from among some set of wavelets, and its restricted expansion coefficients give a compressed signal. One of the systematic methods to generate such a suitable basis is also to employ the wavelet packets.

## Numerical Calculation

Application of wavelet transform, especially of the DWT, to numerical solver for a differential equation (DE) has long been studied. At the first sight, the wavelets appear to give a good DE solver because the wavelet expansion is generally quite efficient compared to Fourier series due to its spatial localization. But its implementation to an efficient computer code is not so straightforward; research is still continuing for concrete problems. Application of the CWT to spectral method for partial differential equation (PDE) has been studied extensively. There is no wavelet which diagonalizes the differential operator $\partial/\partial x$; therefore, an efficient numerical method is necessary for derivatives of wavelets. Products of wavelets also yield another numerical problem. MRA brings about mesh points which are adaptive to some extent, but finite element method still gives more flexible mesh points.

For some scaling-invariant differential or integral operators, including $\partial^2/\partial x^2$, Abel transformations, and Reisz potential, adaptive biorthogonal wavelets can be provided with block-diagonal Galerkin representations, which has been applied to data processing. Generally, simultaneous localization of wavelets, both in space and in scale, leads to a sparse Galerkin representation for many pseudodifferential operators and their inverses. A thresholding technique with DWT has been introduced to coherent vortex simulation of the 2D Navier–Stokes equations, to reduce the relevant wavelet coefficients. Another promising application of wavelet occurs as a preprocessor for an iterative Poisson solver, where a wavelet-based preconditioning leads to a matrix with a bounded condition number.

## Other Wavelets and Generalizations

Several new types of wavelets have been proposed: "coiflet" whose scaling function has vanishing moments giving expansion coefficients approximately equal to values of the data functions, and "symlet" which is an orthonormal wavelet with a nearly symmetric profile. Multiwavelets are wavelets which give a complete orthonormal system in $L^2$ space. In 2D or multidimensional applications of the DWT, separable orthonormal wavelets consisting of tensor products of 1D orthonormal wavelets are frequently used, while nonseparable orthonormal wavelets are also available. Another generalization

of wavelets is the Malvar basis which is also a generalization of local Fourier basis, and gives a perfect reconstruction. A new direction of wavelet is the second-generation wavelets which are constructed by lifting scheme and free from the regular dyadic procedure, and thus applicable to compact regions as $S^2$ and a finite interval.

*See also:* Fractal Dimensions in Dynamics; Image Processing: Mathematics; Intermittency in Turbulence; Wavelets: Application to Turbulence; Wavelets: Mathematical Theory.

## Further Reading


Benedetto JJ and Frazier W (eds.) (1994) *Wavelets: Mathematics and Applications*. Boca Raton, FL: CRC Press.

van den Berg JC (ed.) (1999) *Wavelets in Physics*. Cambridge: Cambridge University Press.

Daubechies I (1992) *Ten Lectures on Wavelets*, SIAM, CBMS61, Philadelphia.

Mallat S (1998) *A Wavelet Tour of Signal Processing*. San Diego: Academic Press.

Strang G and Nguyen T (1997) *Wavelet and Filter Banks*. Wellesley: Wellesley-Cambridge Press.


# Wavelets: Mathematical Theory


**K Schneider**, Université de Provence, Marseille, France
**M Farge**, Ecole Normale Supérieure, Paris, France




## Introduction

The wavelet transform unfolds functions into time (or space) and scale, and possibly directions. The continuous wavelet transform has been discovered by Alex Grossmann and Jean Morlet who published the first paper on wavelets in 1984. This mathematical technique, based on group theory and square-integrable representations, allows us to decompose a signal, or a field, into both space and scale, and possibly directions. The orthogonal wavelet transform has been discovered by Lemarié and Meyer (1986). Then, Daubechies (1988) found orthogonal bases made of compactly supported wavelets, and Mallat (1989) designed the fast wavelet transform (FWT) algorithm. Further developments were done in 1991 by Raffy Coifman, Yves Meyer, and Victor Wickerhauser who introduced wavelet packets and applied them to data compression. The development of wavelets has been interdisciplinary, with contributions coming from very different fields such as engineering (sub-band coding, quadrature mirror filters, time–frequency analysis), theoretical physics (coherent states of affine groups in quantum mechanics), and mathematics (Calderon–Zygmund operators, characterization of function spaces, harmonic analysis). Many reference textbooks are available, some of them we recommend are listed in the "Further reading" section. Meanwhile, a large spectrum of applications has grown and is still developing, ranging from signal analysis and image processing via numerical analysis and turbulence modeling to data compression.

In this article, we will first define the continuous wavelet transform and then the orthogonal wavelet transform based on a multiresolution analysis. Properties of both transforms will be discussed and illustrated by examples. For a general introduction to wavelets, *see* Wavelets: Applications.

## Continuous Wavelet Transform

Let us consider the Hilbert space of square-integrable functions $L^2(\mathbb{R}) = \{f : \||f\|_2 < \infty\}$, equipped with the scalar product $\langle f, g \rangle = \int_{\mathbb{R}} f(x) g^\star(x)\, dx$ ($^\star$ denotes the complex conjugate in the case of complex-valued functions) and where the norm is defined by $\|f\|_2 = \langle f, f \rangle^{1/2}$.

### Analyzing Wavelet

The starting point for the wavelet transform is to choose a real- or complex-valued function $\psi \in L^2(\mathbb{R})$, called the "mother wavelet," which fulfills the admissibility condition,

$$C_\psi = \int_0^\infty \left|\widehat{\psi}(k)\right|^2 \frac{dk}{|k|} < \infty \qquad [1]$$

where

$$\widehat{\psi}(k) = \int_{-\infty}^{\infty} \psi(x)\, e^{-\imath 2\pi k x}\, dx \qquad [2]$$

denotes the Fourier transform, with $\imath = \sqrt{-1}$ and $k$ the wave number. If $\psi$ is integrable, that is, $\psi \in L^1(\mathbb{R})$, this implies that $\psi$ has zero mean,

$$\int_{-\infty}^{\infty} \psi(x)\, dx = 0 \qquad \text{or} \qquad \widehat{\psi}(0) = 0 \qquad [3]$$

In practice, however, one also requires the wavelet $\psi$ to be well localized in both physical and Fourier

$$\int_{-\infty}^{\infty} x^m \psi(x)\,\mathrm{d}x = 0 \qquad \text{for } m = 0,\, M-1 \qquad [4]$$

that is, monomials up to degree $M-1$ are exactly reproduced. In Fourier space, this property is equivalent to

$$\frac{\mathrm{d}^m}{\mathrm{d}k^m}\widehat{\psi}(k)\,|_{k=0} = 0 \quad \text{for } m = 0, M-1 \qquad [5]$$

therefore, the Fourier transform of $\psi$ decays smoothly at $k=0$.

## Analysis

From the mother wavelet $\psi$, we generate a family of continuously translated and dilated wavelets,

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right) \quad \text{for } a > 0 \text{ and } b \in \mathbb{R} \qquad [6]$$

where $a$ denotes the dilation parameter, corresponding to the width of the wavelet support, and $b$ the translation parameter, corresponding to the position of the wavelet. The wavelets are normalized in energy norm, that is, $\|\psi_{a,b}\|_2 = 1$.

In Fourier space, eqn [6] reads

$$\widehat{\psi}_{a,b(k)} = \sqrt{a}\widehat{\psi}(ak)\,\mathrm{e}^{-\imath 2\pi kb} \qquad [7]$$

where the contraction with $1/a$ in [6] is reflected in a dilation by $a$ [7] and the translation by $b$ implies a rotation in the complex plane.

The continuous wavelet transform of a function $f$ is then defined as the convolution of $f$ with the wavelet family $\psi_{a,b}$:

$$\widetilde{f}(a,b) = \int_{-\infty}^{\infty} f(x)\psi^*_{a,b}(x)\,\mathrm{d}x \qquad [8]$$

where $\psi^*_{a,b}$ denotes, in the case of complex-valued wavelets, the complex conjugate.

Using Parseval's identity, we get

$$\widetilde{f}(a,b) = \int_{-\infty}^{\infty} \widehat{f}(k)\widehat{\psi}^*_{a,b}(k)\,\mathrm{d}k \qquad [9]$$

and the wavelet transform could be interpreted as a frequency decomposition using bandpass filters $\widehat{\psi}_{a,b}$ centered at frequencies $k = k_\psi/a$. The wave number $k_\psi$ denotes the barycenter of the wavelet support in Fourier space

$$k_\psi = \frac{\int_0^\infty k|\widehat{\psi}(k)|\,\mathrm{d}k}{\int_0^\infty |\widehat{\psi}(k)|\,\mathrm{d}k} \qquad [10]$$

Note that these filters have a variable width $\Delta k/k$; therefore, when the wave number increases, the bandwidth becomes wider.

## Synthesis

The admissibility condition [1] implies the existence of a finite energy reproducing kernel, which is a necessary condition for being able to reconstruct the function $f$ from its wavelet coefficients $\widetilde{f}$. One then recovers

$$f(x) = \frac{1}{C_\psi}\int_0^\infty \int_{-\infty}^{\infty} \widetilde{f}(a,b)\psi_{a,b}(x)\frac{\mathrm{d}a\mathrm{d}b}{a^2} \qquad [11]$$

which is the inverse wavelet transform.

The wavelet transform is an isometry and one has Parseval's identity. Therefore, the wavelet transform conserves the inner product and we obtain

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(x)g^*(x)\,\mathrm{d}x = \frac{1}{C_\psi}\int_0^\infty \int_{-\infty}^{\infty} \widetilde{f}(a,b)\widetilde{g}^*(a,b)\frac{\mathrm{d}a\mathrm{d}b}{a^2} \qquad [12]$$

As a consequence, the total energy $E$ of a signal can be calculated either in physical space or in wavelet space, such as

$$E = \int_{-\infty}^{\infty} |f(x)|^2\,\mathrm{d}x = \frac{1}{C_\psi}\int_0^\infty \int_{-\infty}^{\infty} |\widetilde{f}(a,b)|^2\frac{\mathrm{d}a\mathrm{d}b}{a^2} \qquad [13]$$

This formula is also the starting point for the definition of wavelet spectra and scalogram (*see* Wavelets: Application to Turbulence).

## Examples

In the following, we apply the continuous wavelet transform to different academic signals using the Morlet wavelet. The Morlet wavelet is complex valued, and consists of a modulated Gaussian with width $k_0/\pi$:

$$\psi(x) = (\mathrm{e}^{2\imath\pi x} - \mathrm{e}^{-k_0^2/2})\,\mathrm{e}^{-2\pi^2x^2/k_0^2} \qquad [14]$$

The envelope factor $k_0$ controls the number of oscillations in the wave packet; typically, $k_0 = 5$ is used. The correction factor $\mathrm{e}^{-k_0^2/2}$, to ensure its vanishing mean, is very small and often neglected. The Fourier transform is

$$\widehat{\psi}(k) = \frac{k_0}{2\sqrt{\pi}}\,\mathrm{e}^{-(k_0^2/2)(1+k^2)}(\mathrm{e}^{-k_0^2k} - 1) \qquad [15]$$

Figure 1 shows wavelet analyses of a cosine, two sines, a Dirac, and a characteristic function. Below

the four signals we plot the modulus and the phase of the corresponding wavelet coefficients.

### Higher Dimensions

The continuous wavelet transform can be extended to higher dimensions in $L^2(\mathbb{R}^n)$ in different ways. Either we define spherically symmetric wavelets by setting $\psi(x) = \psi^{1d}(|x|)$ for $x \in \mathbb{R}^n$ or we introduce in addition to dilations $a \in \mathbb{R}^+$ and translations $b \in \mathbb{R}^n$ also rotations to define wavelets with a directional sensitivity. In the two-dimensional case, we obtain for example,

$$\psi_{a,b,\theta}(x) = \frac{1}{a}\psi\left(R_\theta^{-1}\left(\frac{x-b}{a}\right)\right) \qquad [16]$$

where $a \in \mathbb{R}^+, b \in \mathbb{R}^2$, and where $R_\theta$ is the rotation matrix

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \qquad [17]$$

The analysis formula [8] then becomes

$$\widetilde{f}(a,b,\theta) = \int_{\mathbb{R}^2} f(x)\psi^*_{a,b,\theta}(x)\,\mathrm{d}x \qquad [18]$$

and for the corresponding inverse wavelet transform [11] we obtain

$$f(x) = \frac{1}{C_\psi}\int_0^\infty \int_{\mathbb{R}^2}\int_0^{2\pi} \widetilde{f}(a,b,\theta)\psi_{a,b,\theta}(x)\frac{\mathrm{d}a\mathrm{d}b\mathrm{d}\theta}{a^3} \qquad [19]$$

Similar constructions can be made in dimensions larger than 2 using $n-1$ angles of rotation.

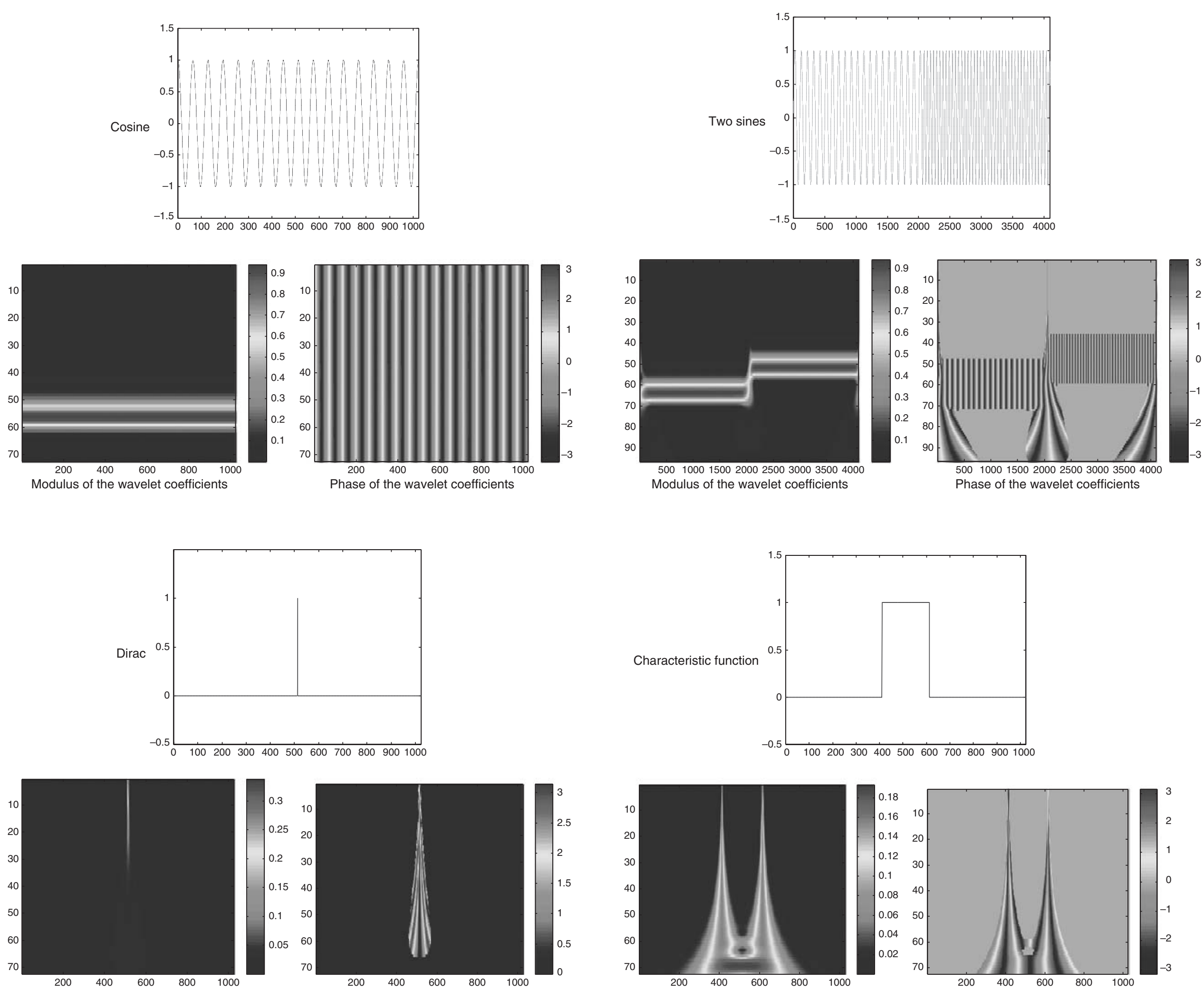



**Figure 1** Examples of a one-dimensional continuous wavelet analysis using the complex-valued Morlet wavelet. Each subfigure shows on the top the function to be analyzed and below (left) the modulus of its wavelet coefficients and below (right) the phase of its wavelet coefficients.

## Discrete Wavelets

### Frames

It is possible to obtain a discrete set of quasiorthogonal wavelets by sampling the scale and position axes $a, b$. For the scale $a$ we use a logarithmic discretization: $a$ is replaced by $a_j = a_0^{-j}$, where $a_0$ is the sampling rate of the log $a$ axis ($a_0 = \Delta(\log a)$) and where $j \in \mathbb{Z}$ is the scale index. The position $b$ is discretized linearly: $b$ is replaced by $x_{ji} = ib_0 a_0^{-j}$, where $b_0$ is the sampling rate of the position axis at the largest scale and where $i \in \mathbb{Z}$ is the position index. Note that the sampling rate of the position varies with scale, that is, for finer scales (increasing $j$ and hence decreasing $a_j$), the sampling rate increases. Accordingly, we obtain the discrete wavelets (cf. Figure 2)

$$\psi_{ji}(x') = a_j^{-1/2}\psi\left(\frac{x' - x_{ji}}{a_j}\right) \qquad [20]$$

and the corresponding discrete decomposition formula is

$$\widetilde{f}_{ji} = \langle \psi_{ji}, f \rangle = \int_{-\infty}^{\infty} f(x')\psi_{ji}^*(x')\,\mathrm{d}x' \qquad [21]$$

Furthermore, the wavelet coefficients satisfy the following estimate:

$$A\|f\|_2^2 \le \sum_{j,i} |\widetilde{f}_{ji}|^2 \le B\|f\|_2^2 \qquad [22]$$

with frame bounds $B \ge A > 0$. In the case $A = B$ we have a tight frame.

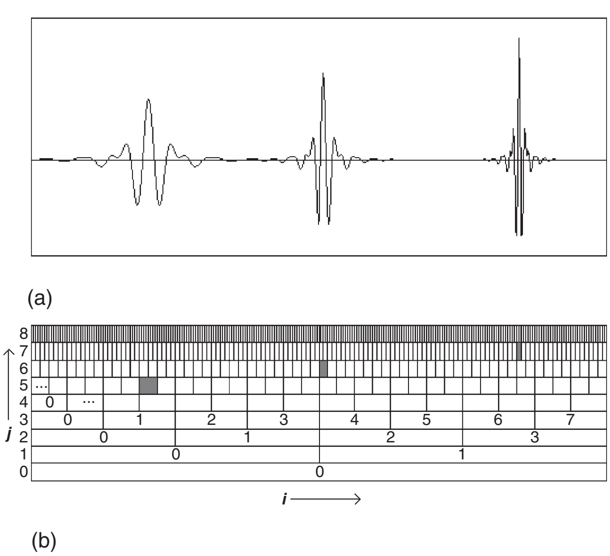


Figure 2 Orthogonal quintic spline wavelets $\psi_{j,i}(x) = 2^{j/2}\psi(2^j x - i)$ at different scales and positions: (a) $\psi_{5,6}(x)$, $\psi_{6,32}(x)$, $\psi_{7,108}(x)$, and (b) corresponding wavelet coefficients.

The discrete reconstruction formula is

$$f(x) = C \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} \widetilde{f}_{ji}\psi_{ji}(x) + R(x) \qquad [23]$$

where $C$ is a constant and $R(x)$ is a residual, both depending on the choice of the wavelet and the sampling of the scale and position axes. For the particular choice $a_0 = 2$ (which corresponds to a scale sampling by octaves) and $b_0 = 1$, we have the dyadic sampling, for which there exist special wavelets $\psi_{ji}$ that form an orthonormal basis of $L^2(\mathbb{R})$, that is, such that

$$\langle \psi_{ji}, \psi_{j'i'} \rangle = \delta_{jj'}\delta_{ii'} \qquad [24]$$

where $\delta$ denotes the Kronecker symbol. This means that the wavelets $\psi_{ji}$ are orthogonal with respect to their translates by discrete steps $2^{-j}i$ and their dilates by discrete steps $2^{-j}$ corresponding to octaves. In this case, the reconstruction formula is exact with $C = 1$ and $R = 0$. Note that the discrete wavelet transform has lost the invariance by translation and dilation of the continuous one.

### Orthogonal Wavelets and Multiresolution Analysis

The construction of orthogonal wavelet bases and the associated fast numerical algorithm is based on the mathematical concept of multiresolution analysis (MRA). The underlying idea is to consider approximations $f_j$ of the function $f$ at different scales $j$. The amount of information needed to go from a coarse approximation $f_j$ to a finer resolution approximation $f_{j+1}$ is then described using orthogonal wavelets. The orthogonal wavelet analysis can thus be interpreted as decomposing the function into approximations of the function at coarser and coarser scales (i.e., for decreasing $j$), where the differences between the approximations are encoded using wavelets.

The definition of the MRA was introduced by Stéphane Mallat in 1988 (Mallat 1989). This technique constitutes a mathematical framework of orthogonal wavelets and the related FWT.

A one-dimensional orthogonal MRA of $L^2(\mathbb{R})$ is defined as a sequence of successive approximation spaces $V_j, j \in \mathbb{Z}$, which are closed imbedded subspaces of $L^2(\mathbb{R})$. They verify the following conditions:

$$V_j \subset V_{j+1} \quad \forall j \in \mathbb{Z} \qquad [25]$$

$$\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L^2(\mathbb{R}) \qquad [26]$$

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\} \qquad [27]$$

$$f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1} \qquad [28]$$

A scaling function $\phi(x)$ is required to exist. Its translates generate a basis in each $V_j$, that is,

$$V_j V_j = \overline{\text{span}}\{\phi_{ji}\}_{i\in\mathbb{Z}} \quad [29]$$

where

$$\phi_{ji}(x) = 2^{j/2}\phi(2^j x - i), \quad j,i \in \mathbb{Z} \quad [30]$$

At a given scale $j$, this basis is orthonormal with respect to its translates by steps $i/2^j$ but not to its dilates,

$$\langle \phi_{ji}, \phi_{jk}\rangle = \delta_{ik} \quad [31]$$

The nestedness of the approximation spaces [28] generated by the scaling function $\phi$ implies that it satisfies a refinement equation:

$$\phi_{j-1,i}(x) = \sum_{n=-\infty}^{\infty} h_{n-2i}\phi_{jn}(x) \quad [32]$$

with the filter coefficients $h_n = \langle \phi_{jn}, \phi_{j-1,0}\rangle$, which determine the scaling function completely. In general, only the filter coefficients $h_n$ are known and no analytical expression of $\phi$ is given. Equation [32] implies that the approximation of a function at coarser scale can be described by linear combinations of the same function at finer scales.

The orthogonal projection of a function $f \in L^2(\mathbb{R})$ on $V_J$ is defined as

$$P_{V_J} : f \longrightarrow P_{V_J} f = f_J \quad [33]$$

with

$$f_J(x) = \sum_{k\in\mathbb{Z}} \langle f, \phi_{jk}\rangle \phi_{jk}(x) \quad [34]$$

This coarse graining at a given scale $J$ is done by filtering the function with the scaling function $\phi$. As a filter, the scaling function $\phi$ does not have vanishing mean but is normalized so that $\int_{-\infty}^{\infty} \phi(x)\,\mathrm{d}x = 1$.

As $V_{J-1}$ is included in $V_J$, we can define its orthogonal complement space in $V_J$:

$$V_J = V_{J-1} \oplus W_{J-1} \quad [35]$$

Correspondingly, the approximation of the function $f$ at scale $2^{-J}$, belonging to $V_J$, can be decomposed as a sum of orthogonal projections on $V_{J-1}$ and $W_{J-1}$, such that

$$P_{V_J} f = P_{V_{J-1}} f + P_{W_{J-1}} f \quad [36]$$

Based on the scaling function $\phi$, one can construct a function $\psi$, the so-called mother wavelet, given by the relation

$$\psi_{ji}(x) = \sum_{n\in\mathbb{Z}} g_{n-2i}\phi_{j,n}(x) \quad [37]$$

with $g_n = \langle \phi_{jn}, \psi_{j-1,0}\rangle$, and where $\psi_{ji}(x) = 2^{j/2}\psi(2^j x - i)$, $j,i \in \mathbb{Z}$ (cf. **Figure 2**). The filter coefficients $g_n$ can be computed from the filter coefficients $h_n$ using the relation

$$g_n = (-1)^{1-n} h_{1-n} \quad [38]$$

The translates and dilates of the wavelet $\psi$ constitute orthonormal bases of the spaces $W_j$,

$$W_j = \overline{\text{span}}\{\psi_{ji}\}_{i\in\mathbb{Z}} \quad [39]$$

As in the continuous case, the wavelets have vanishing mean, and also possibly vanishing higher-order moments; therefore,

$$\int_{-\infty}^{\infty} x^m \psi(x)\,\mathrm{d}x = 0 \quad \text{for } m = 0, \ldots, M-1 \quad [40]$$

Let us now consider approximations of a function $f \in L^2(\mathbb{R})$ at two different scales $j$:

- at scale $j$

$$f_j(x) = \sum_{i=-\infty}^{\infty} \overline{f}_{ji}\phi_{ji}(x) \quad [41]$$

- at scale $j-1$

$$f_{j-1}(x) = \sum_{i=-\infty}^{\infty} \overline{f}_{j-1,i}\phi_{j-1,i}(x) \quad [42]$$

with the scaling coefficients

$$\overline{f}_{ji} = \langle f, \phi_{ji}\rangle \quad [43]$$

which correspond to local averages of the function $f$ at position $i2^{-j}$ and at scale $2^{-j}$.

The difference between the two approximations is encoded by the wavelets

$$f_j(x) - f_{j-1}(x) = \sum_{i=-\infty}^{\infty} \widetilde{f}_{j-1,\,i}\psi_{j-1,i}(x) \quad [44]$$

with the wavelet coefficients

$$\widetilde{f}_{ji} = \langle f, \psi_{ji}\rangle \quad [45]$$

which correspond to local differences of the function at position $(2i+1)2^{-(j+1)}$ between approximations at scales $2^{-j}$ and $2^{-(j+1)}$.

Iterating the two-scale decomposition [44], any function $f \in L^2(\mathbb{R})$ can be expressed as a sum of a coarse-scale approximation at a reference scale $j_0$ that we set to 0 here, and their successive

differences. These details are needed to go from one scale $j$ to the next finer scale $j+1$ for $j=0,\dots,J-1$,

$$f(x) = \sum_{i=-\infty}^{\infty} \overline{f}_{0,i}\phi_{0,i}(x) + \sum_{j=0}^{\infty}\sum_{i=-\infty}^{\infty} \widetilde{f}_{ji}\psi_{ji}(x) \qquad [46]$$

For numerical applications, the sums in eqn [46] have to be truncated in both scale $j$ and position $i$. The truncation in scale corresponds to a limitation of $f$ to a given finest scale $J$, which is in practice imposed by the available sampling rate. Due to the finite length of the available data, the sum over $i$ also becomes finite. The decomposition [46] is orthogonal, as, by construction,

$$\langle \psi_{ji}, \psi_{j'i'} \rangle = \delta_{jj'}\delta_{ii'} \qquad [47]$$

$$\langle \psi_{ji}, \phi_{j'i'} \rangle = 0 \quad \text{for } j \geq j' \qquad [48]$$

in addition to [31].

### Fast Wavelet Transform

Starting with a function $f \in L^2(\mathbb{R})$ given at the finest resolution $2^{-J}$ (i.e., we know $f_J \in V_J$ and hence the coefficients $\overline{f}_{Ji}$ for $i \in \mathbb{Z}$), the FWT computes its wavelet coefficients $\widetilde{f}_{ji}$ by decomposing successively each approximation $f_J$ into a coarser scale approximation $f_{J-1}$, plus the corresponding details which are encoded by the wavelet coefficients. The algorithm uses a cascade of discrete convolutions with the low pass filter $h_n$ and the bandpass filter $g_n$, followed by downsampling, in which only one coefficient out of two is retained. The direct wavelet transform algorithm is

- initialization

$$\text{given } f \in L^2(\mathbb{R}) \text{ and } \overline{f}_{Ji} = f\left(\frac{i}{2^J}\right) \text{ for } i \in \mathbb{Z}$$

- decomposition
  for $j=J$ to 1, step –1, do

$$\overline{f}_{j-1,i} = \sum_{n\in\mathbb{Z}} h_{n-2i}\overline{f}_{jn} \qquad [49]$$

$$\widetilde{f}_{j-1,i} = \sum_{n\in\mathbb{Z}} g_{n-2i}\overline{f}_{jn} \qquad [50]$$

The inverse wavelet transform is based on successive reconstructions of fine-scale approximations $f_j$ from coarser scale approximations $f_{j-1}$, plus the differences between approximations at scale $j-1$ and the finer scale $j$ which are encoded by $\widetilde{f}_{j-1,i}$. The algorithm uses a cascade of discrete convolutions with the filters $h_n$ and $g_n$, preceded by upsampling which adds zeros in between two successive coefficients.

- reconstruction
  for $j=1$ to $J$, step 1, do

$$\overline{f}_{ji} = \sum_{n=-\infty}^{\infty} h_{i-2n}\overline{f}_{j-1,n} + \sum_{n=-\infty}^{\infty} g_{i-2n}\widetilde{f}_{j,n} \qquad [51]$$

The FWT has been introduced by Stéphane Mallat in 1989. If the scaling functions (and wavelets) are compactly supported, the filters $h_n$ and $g_n$ have only a finite number of nonvanishing coefficients. In this case, the numerical complexity of the FWT is $\mathcal{O}(N)$ where $N$ denotes the number of samples.

### Choice of Wavelets

Orthogonal wavelets are typically defined by their filter coefficients $h_n$, since in general no analytic expression for $\psi$ is available. In the following, we give the filter coefficients of $h_n$ for some typical orthogonal wavelets. The filter coefficients of $g_n$ can be obtained using the quadrature relation between the two filters [38].

- Haar D1 (one vanishing moment):

$$h_0 = 1/\sqrt{2}$$
$$h_1 = 1/\sqrt{2}$$

- Daubechies D2 (two vanishing moments):

$$h_0 = 0.482\,962\,913\,145$$
$$h_1 = 0.836\,516\,303\,736$$
$$h_2 = 0.224\,143\,868\,042$$
$$h_3 = -0.129\,409\,522\,551$$

- Daubechies D3 (three vanishing moments):

$$h_0 = 0.332\,670\,552\,950$$
$$h_1 = 0.806\,891\,509\,311$$
$$h_2 = 0.459\,877\,502\,118$$
$$h_3 = -0.135\,011\,020\,010$$
$$h_4 = -0.085\,441\,273\,882$$
$$h_5 = 0.035\,226\,291\,882$$

- Coiflets C12 (four vanishing moments): the wavelets and the corresponding scaling function are shown in Figure 3.

**Remarks** The construction of orthogonal wavelets in $L^2(\mathbb{R})$ can be modified to obtain wavelets on the interval, that is, in $L^2([0,1])$. Therewith, boundary wavelets are introduced, while in the interior of the interval the wavelets are not modified.

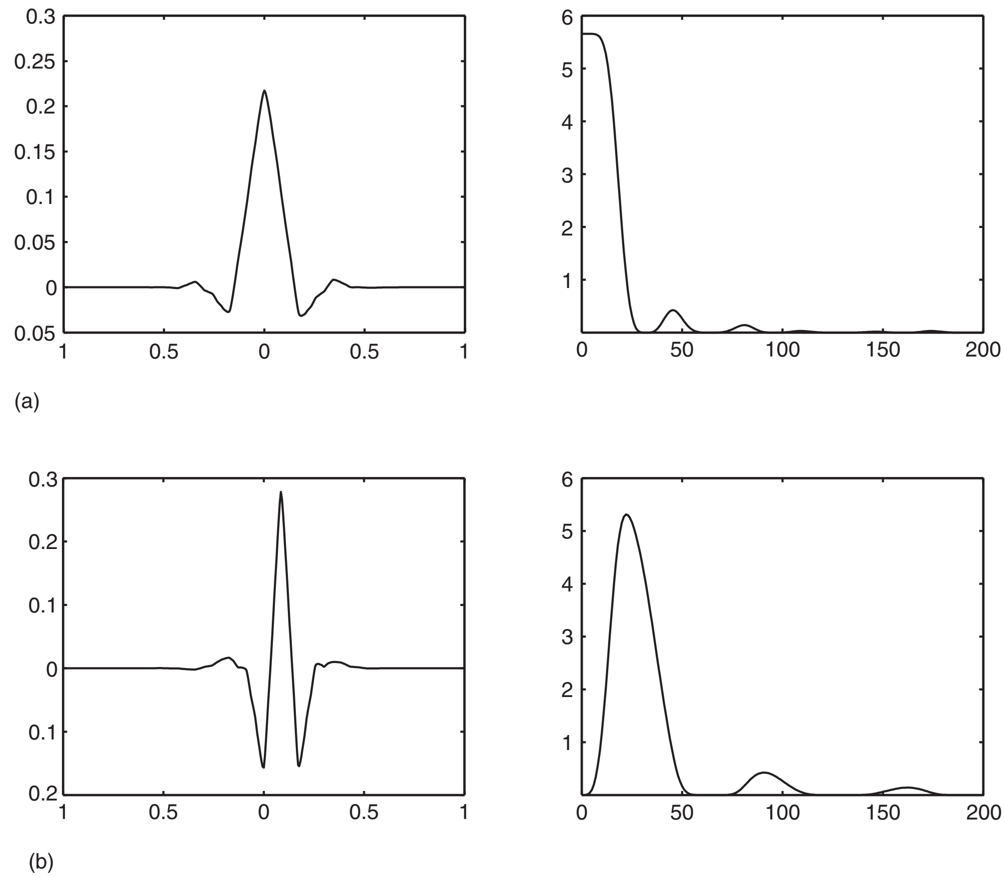



**Figure 3** Orthogonal wavelets Coiflet C12. (a) Scaling function $\phi(x)$ (left) and $|\hat{\phi}(\omega)|$. (b) Wavelet $\psi(x)$ (left) and $|\hat{\psi}(\omega)|$.

A periodic MRA of $L^2(\mathbb{T})$, where $\mathbb{T}=\mathbb{R}/\mathbb{Z}$ denotes the torus, can also be constructed by periodizing the wavelets in $L^2(\mathbb{R})$, using

$$\psi^{\mathrm{per}}(x) = \sum_{k\in\mathbb{Z}} \psi(x+k)$$

Relaxing the condition of orthogonality allows greater flexibility in the choice of the basis functions. For example, biorthogonal wavelets can be designed using different basis functions for analysis ($^{\mathrm{a}}$) and synthesis ($^{\mathrm{s}}$) which are related but no longer orthogonal. A couple of refinable scaling functions $(\phi^{\mathrm{a}}, \phi^{\mathrm{s}})$ with related wavelets $(\psi^{\mathrm{a}}, \psi^{\mathrm{s}})$ which are by construction biorthogonal generate a biorthogonal MRA $V_j^{\mathrm{a}}, V_j^{\mathrm{s}}$. From an algorithmic point of view, only two different filter couples $(g^{\mathrm{a}}, h^{\mathrm{a}})$ for the forward and $(g^{\mathrm{s}}, h^{\mathrm{s}})$ for the backward FWT are used, without changing the algorithm.

The multiresolution approach can be further generalized, for samplings on nonequidistant grids leading to the so-called second-generation wavelets.

### Higher Dimensions

The previously presented one-dimensional construction can be extended to higher dimensions. For simplicity, we will consider only the two-dimensional case, since higher dimensions can be treated analogously.

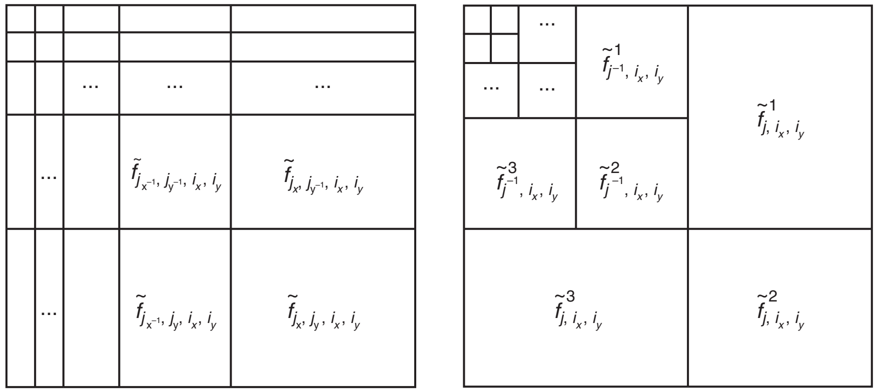
**Tensor product construction** Having developed a one-dimensional orthonormal basis $\psi_{ji}$ of $L^2(\mathbb{R})$, one could use these functions as building blocks in higher dimensions. One way of doing so is to take the tensor product of two one-dimensional bases and to define

$$\psi_{j_x,j_y,i_x,i_y}(x,y) = \psi_{j_x,i_x}(x)\psi_{j_y,i_y}(y) \qquad [52]$$

The resulting functions constitue an orthonormal wavelet basis for $L^2(\mathbb{R}^2)$. Each function $f \in L^2(\mathbb{R}^2)$ can then be developed into

$$f(x,y) = \sum_{j_x,i_x} \sum_{j_y,i_y} \widetilde{f}_{j_x,j_y,i_x,i_y} \psi_{j_x,j_y,i_x,i_y}(x,y) \qquad [53]$$

with $\widetilde{f}_{j_x,j_y,i_x,i_y} = \langle f, \psi_{j_x,j_y,i_x,i_y} \rangle$. However, in this basis the two variables $x$ and $y$ are dilatated separately




(a) (b)

**Figure 4a** Schematic representation of the 2D (b) wavelet transforms: (a) Tensor product construction and (b) 2D MRA.

and therefore no longer form an MRA. This means that the functions $\psi_{j_x,j_y}$ involve two scales, $2^{j_x}$ and $2^{j_y}$, and each of the functions is essentially supported on a rectangle with these side-lengths. Hence, the decomposition is often called rectangular wavelet decomposition (cf. Figure 4a). From the algorithmic viewpoint, this is equivalent to applying the one-dimensional wavelet transform to the rows and the columns of a matrix or a function. For some applications, such a basis is advantageous, for others not. Often the notion of a scale has a certain meaning. For an application, one would like to have a unique scale assigned to each basis function.

**Multiresolution construction** Another much more interesting construction is the construction of a truly two-dimensional MRA of $L^2(\mathbb{R}^2)$. It can be obtained through the tensor product of two one-dimensional MRAs of $L^2(\mathbb{R})$. More precisely, one defines the spaces $\boldsymbol{V}_j, j \in \mathbb{Z}$ by

$$\boldsymbol{V}_j = V_j \otimes V_j \qquad [54]$$

and $\boldsymbol{V}_j = \overline{\text{span}}\{\phi_{j,i_x,i_y}(x,y) = \phi_{j,i_x}(x)\phi_{j,i_y}(y), i_x, i_y \in \mathbb{Z}\}$ fulfilling analogous properties as in the one-dimensional case.

Likewise, we define the complement space $\boldsymbol{W}_j$ to be the orthogonal complement of $\boldsymbol{V}_j$ in $\boldsymbol{V}_{j+1}$, that is,

$$\begin{aligned}\boldsymbol{V}_{j+1} &= V_{j+1} \otimes V_{j+1} \\ &= (V_j \oplus W_j) \otimes (V_j \oplus W_j)\end{aligned} \qquad [55]$$

$$\begin{aligned}&= V_j \otimes V_j \oplus ((W_j \otimes V_j) \\ &\quad \oplus (V_j \otimes W_j) \oplus (W_j \otimes W_j))\end{aligned} \qquad [56]$$

$$= \boldsymbol{V}_j \oplus \boldsymbol{W}_j \qquad [57]$$

It follows that the orthogonal complement $\boldsymbol{W}_j = \boldsymbol{V}_{j+1} \ominus \boldsymbol{V}_j$ consists of three different types of functions and is generated by three different wavelets

$$\psi^{\varepsilon}_{j,i_x,i_y}(x,y) = \begin{cases} \psi_{j,i_x}(x)\phi_{j,i_y}(y), & \varepsilon = 1 \\ \phi_{j,i_x}(x)\psi_{j,i_y}(y), & \varepsilon = 2 \\ \psi_{j,i_x}(x)\psi_{j,i_y}(y), & \varepsilon = 3 \end{cases} \qquad [58]$$

Observe that here the scale parameter $j$ simultaneously controls the dilatation in $x$ and $y$. We recall that in $d$ dimensions this construction yields $2^d - 1$ types of wavelets spanning $\boldsymbol{W}_j$.

Using [58], each function $f \in L^2(\mathbb{R}^2)$ can be developed into a multiresolution basis as

$$f(x,y) = \sum_j \sum_{i_x,i_y} \sum_{\varepsilon=1,2,3} \tilde{f}^{\varepsilon}_{j,i_x,i_y} \psi^{\varepsilon}_{j,i_x,i_y}(x,y) \qquad [59]$$

with $\tilde{f}^{\varepsilon}_{j,i_x,i_y} = <f, \psi^{\varepsilon}_{j,i_x,i_y}>$. A schematic representation of the wavelet coefficients is shown in Figure 4b. The algorithmic structure of the one-dimensional transforms carries over to the two-dimensional case by simple tensorization, that is, applying the filters at each decomposition step to rows and columns.

**Remark** The described two-dimensional wavelets and scaling functions are separable. This advantage is the ease of generation starting from one-dimensional MRAs. However, the main drawback of this construction is that three wavelets are needed to span the orthogonal complement space $\boldsymbol{W}_j$. Another property should be mentioned. By construction, the wavelets are anisotropic, that is, horizontal, diagonal, and vertical directions are preferred.

## Approximation Properties

### Reproduction of Polynomials

A fundamental property of the MRA is the exact reproduction of polynomials. The vanishing moments of the wavelet $\psi$, that is, $\int_{\mathbb{R}} x^m \psi(x) \mathrm{d}x = 0$

for $m=0, M-1$, is equivalent to the fact that polynomials up to degree $M-1$, can be expressed exactly as a linear combination of scaling functions, $p_m(x)=\sum_{n\in\mathbb{Z}} n^m \phi(x-n)$ for $m=0,M-1$. This so-called Strang–Fix condition proves that $\psi$ has $M$ vanishing moments if and only if any polynomial of degree $M-1$ can be written as a linear combination of scaling functions $\phi$. Note that, as $p_m \notin L^2(\mathbb{R})$, the coefficients $n^m$ are not in $l^2(\mathbb{Z})$.

### Regularity and Local Decay of Wavelet Coefficients

The local or global regularity of a function is closely related to the decay of its wavelet coefficients. If a function is locally in $C^s(\mathbb{R})$ (the space of $s$-times continuously differentiable functions), it can be well approximated locally by a Taylor series of degree $s$. Consequently, its wavelet coefficients are small at fine scales, as long as the wavelet $\psi$ has enough vanishing moments. The decay of the coefficients hence determines directly the error being made when truncating a wavelet sum at some scale.

Depending on the type of norm used and whether global or local characterization is concerned, various relations of this kind have been developed. Let us take as example the case of an $\alpha$-Lipschitz function.

Suppose $f \in L^2(\mathbb{R})$, then for $[a,b] \subset \mathbb{R}$ the function $f$ is $\alpha$-Lipschitz with $0 < \alpha < 1$ for any $x_0 \in [a,b]$, that is, $|f(x_0+h) - f(x_0)| \le C|h|^\alpha$, if and only if there exists a constant $A$ such that $|\widetilde{f}_{ji}| \le A2^{-j\alpha-1/2}$ for any $(j,i)$ with $i/2^j \in [a,b]$.

This shows the relation between the local regularity of a function and the decay of its wavelet coefficients in scale.

**Example** To illustrate the local decay of the wavelet coefficients, we consider in Figure 5 the function $f(x) = \sin(2\pi x)$ for $x \le 1/4$ and $x \ge 3/4$ and $f(x) = -\sin(2\pi x)$ for $1/4 < x < 3/4$. The corresponding wavelet coefficients for quintic spline wavelets are plotted in logarithmic scale. The wavelet coefficients show that only in a local region around singularities the fine-scale coefficients are significant.

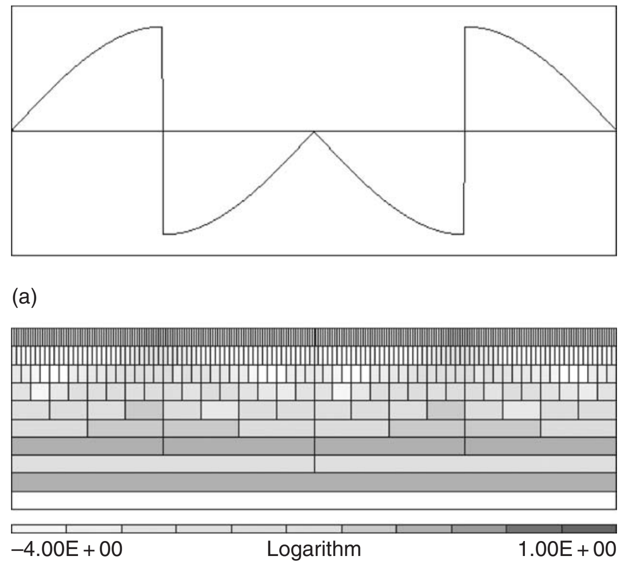



**Figure 5** Orthogonal wavelet decomposition using quintic spline wavelets: (a) function $f(x) = \sin(2\pi x)$ for $x \le 1/4$ and $x \ge 3/4$ and $f(x) = -\sin(2\pi x)$ for $1/4 < x < 3/4$ sampled on a grid $x_i = i/2^J, i=0,\ldots,2^J-1$ with $J=9$ and (b) corresponding wavelet coefficients $\log_{10}|\widetilde{f}_{j,i}|$ for $i=0,\ldots,2^j-1$ and $j=0,\ldots,J-1$.

### Linear Approximation

The exact reproduction of polynomials can be used to derive error estimates for the approximation of a function $f$ at a given scale, which corresponds to linear approximation. We consider $f$ belonging to the Sobolev space $W^{s,p}(\mathbb{R}^d)$, that is, the weak derivatives of $f$ up to order $s$ belong to $L^p(\mathbb{R}^d)$. The linear approximation of $f$ at scale $J$, corresponding to the projection of $f$ onto $V_J$, is then given by

$$f_J(x) = \sum_{j=0}^{J-1} \sum_{i\in\mathbb{Z}} \widetilde{f}_{j,i} \psi_{j,i}(x) \qquad [60]$$

The approximation error can be estimated by

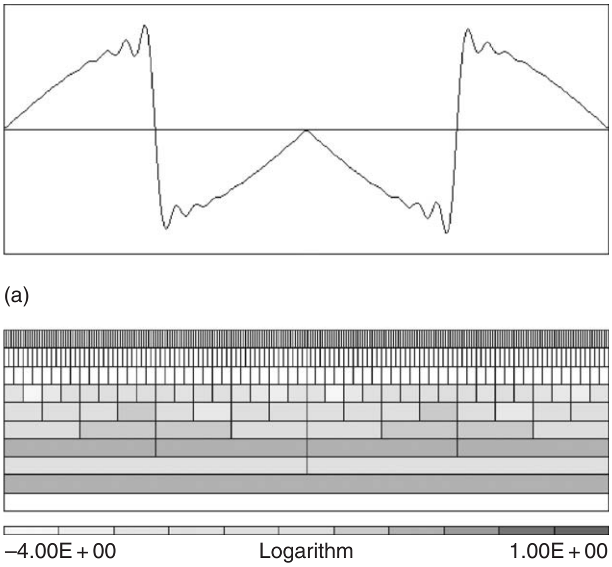
$$\|f - f_J\|_{L^p} < C2^{-J\min(s,m)/d} \qquad [61]$$

where $s$ denotes the smoothness of the function in $L^p$, $d$ the space dimension, and $m$ the number of vanishing moments of the wavelet $\psi$. In the case of poor global regularity of $f$, that is, for small $s$, a large number of scales $J$ is needed to get a good approximation of $f$.

In Figure 6, we plot the linear approximation of the function $f$ shown in Figure 5. The function $f_6$ is reconstructed using wavelet coefficients up to scale $J-1=5$, so that in total only 64 out of 512 coefficients are retained. We observe an oscillating behavior of $f_J$ near the discontinuities of $f$ which dominates the approximation error.

### Nonlinear Approximation

Retaining the $N$ largest wavelet coefficients in the wavelet expansion of $f$ in [46], without imposing any *a priori* cutoff scale, yields the best $N$-term approximation $f^N$. In contrast to the linear approximation [60], it is called nonlinear approximation, since the choice of the retained coefficients depends




(b)

**Figure 6** (a) Linear approximation $f_J$ of the function $f$ in Figure 5 for $J=6$, reconstructed from 64 wavelet coefficients using quintic splines wavelets and (b) corresponding wavelet coefficients $\log_{10}|\widetilde{f}_{j,i}|$ for $i=0,\ldots,2^j-1$ and $j=0,\ldots,J-1$. Note that the coefficients for $J>5$ have been set to zero.

on the function $f$. The mathematical theory has been formalized by Cohen, Dahmen, and De Vore.

The nonlinear approximation of the function $f$ can then be written as

$$f^N(x) = \sum_{(j,i)\in\Lambda_N} \widetilde{f}_{j,i}\,\psi_{j,i}(x) \qquad [62]$$

where $\Lambda_N$ denotes the ensemble of all multi-indices $\lambda=(j,i)$, indexing the $N$ largest coefficients (measured in the $l^p$ norm),

$$\Lambda_N = \{\lambda_k, k=1,N|\ \ \|\widetilde{f}_{\lambda_k}\|_{l^p} > \|\widetilde{f}_\mu\|_{l^p}\ \ \forall\mu\in\Lambda\} \qquad [63]$$

with $\Lambda=\{\mu=(j,i), j\geq 0, i\in\mathbb{Z}\}$. The nonlinear approximation leads to the following error estimate:

$$\|f-f^N\|_{L^p} < CN^{-s/d} \qquad [64]$$

where $s$ denotes the smoothness of $f$ in the larger space $L^q(\mathbb{R}^d)$ with

$$\frac{1}{q}=\frac{1}{p}+\frac{s}{d}$$

which corresponds to the Sobolev embedding line (Figure 7). This estimate shows that the nonlinear approximation converges faster than the linear one, if $f$ has a larger regularity in $L^q$, that is, $f\in W^{s,q}(\mathbb{R}^d)$, which is for example the case for functions with isolated singularities and for small $q$.

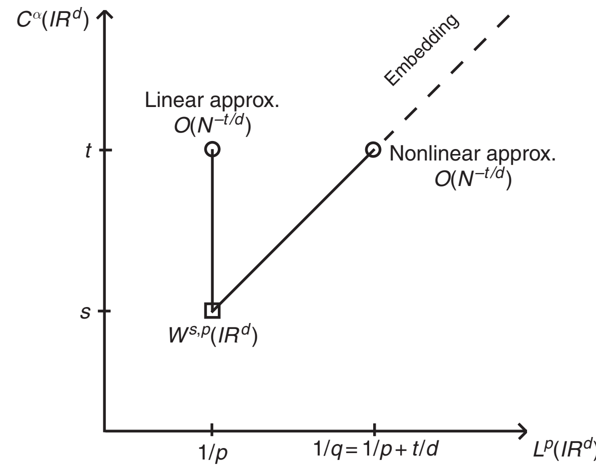
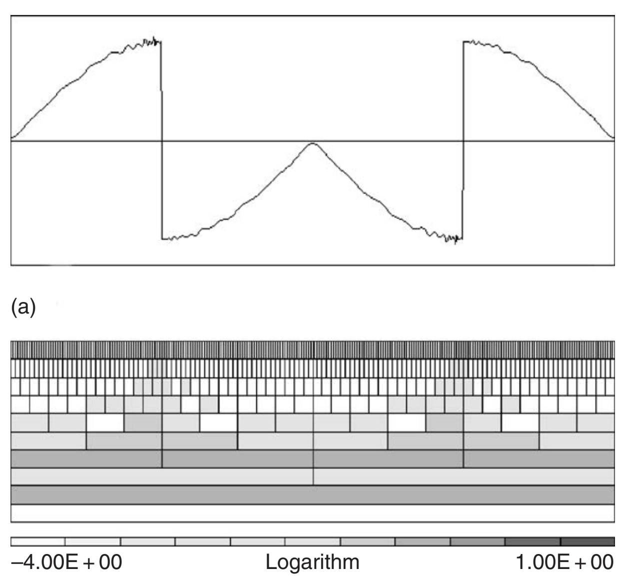



**Figure 7** Schematic representation of linear and nonlinear approximation.

In Figure 8, we plot the nonlinear approximation of the function $f$ shown in Figure 5. The function $f^N$ is reconstructed using the strongest 64 wavelet coefficients out of 512 coefficients. Compared to the linear approximation (cf. Figure 6), the oscillations around the discontinuities disappear and the approximation error is reduced while using the same number of coefficients.

## Compression and Preconditioning of Operators

The nonlinear approximation of functions can be extended to certain operators leading to an efficient




(b)

**Figure 8** (a) Nonlinear approximation $f^N$ of the function $f$ in Figure 5 reconstructed from the 64 largest wavelet coefficients using quintic splines wavelets, (b) retained wavelet coefficients $\log_{10}|\widetilde{f}_{j,i}|$ for $i=0,\ldots,2^j-1$ and $j=0,\ldots,J-1$.

representation in wavelet space, that is, to sparse matrices. For integral operators, for example, Calderon–Zygmund operators $T$ on $\mathbb{R}$ defined by

$$Tf(x) = \int_{\mathbb{R}} K(x,y)f(y)\,dy \qquad [65]$$

where the kernel $k$ satisfies

$$|k(x,y,)| \leq \frac{C}{|x-y|}$$

and

$$\left|\frac{\partial}{\partial x}k(x,y)\right| + \left|\frac{\partial}{\partial y}k(x,y,)\right| \leq \frac{C}{|x-y|^2}$$

their wavelet representation $\langle T\psi_{j,i}, \psi_{j',i'}\rangle$ is sparse and a large number of weak coefficients can be suppressed by simple thresholding of the matrix entries while controlling the precision. The resulting numerical scheme is called BCR algorithm and is due to Beylkin *et al.* (1991).

The characterization of function spaces by the decay of the wavelet coefficients and the corresponding norm equivalences can be used for diagonal preconditioning of integral or differential operators which leads to matrices with uniformly bounded condition numbers. For elliptic differential operators, for example, the Laplace operator $\nabla^2$ the norm equivalence $\|\nabla^2 f\| \simeq \|2^{2j}\widetilde{f}_{ji}\|$ can be used for preconditioning the matrix $\langle \nabla^2\psi_{j,i}, \psi_{j',i'}\rangle$ by a simple diagonal scaling with $2^{-2j}$ to obtain a uniformly bounded condition number. For further details, we refer to the book of Cohen (2000).

### Wavelet Denoising

We consider a function $f$ which is corrupted by a Gaussian white noise $n \in \mathcal{N}(0,\sigma^2)$. The noise is spread over all wavelet coefficients $\widetilde{s}_\lambda$, while, typically, the original function $f$ is determined by only few significant wavelet coefficients. The aim is then to reconstruct the function $f$ from the observed noisy signal $s = f + n$.

The principle of the wavelet denoising can be summarized in the following procedure:

- *Decomposition*. Compute the wavelet coefficients $\widetilde{s}_\lambda$ using the FWT.
- *Thresholding*. Apply the thresholding function $\rho_\varepsilon$ to the wavelet coefficients $\widetilde{s}_\lambda$, thus reducing the relative importance of the coefficients with small absolute value.
- *Reconstruction*. Reconstruct a denoised version $s_C$ from the thresholded wavelet coefficients using the fast inverse wavelet transform.

The thresholding parameter $\varepsilon$ depends on the variance of the noise and on the sample size $N$. The thresholding function $\rho$ we consider corresponds to hard thresholding:

$$\rho_\varepsilon(a) = \begin{cases} a & \text{if } |a| > \varepsilon \\ 0 & \text{if } |a| \leq \varepsilon \end{cases} \qquad [66]$$

Donoho and Johnstone (1994) have shown that there exists an optimal $\varepsilon$ for which the relative quadratic error between the signal $s$ and its estimator $s_C$ is close to the minimax error for all signals $s \in \mathcal{H}$, where $\mathcal{H}$ belongs to a wide class of function spaces, including Hölder and Besov spaces. They showed using the threshold

$$\varepsilon_D = \sigma_n\sqrt{2\ln N} \qquad [67]$$

yields an error which is close to the minimum error. The threshold $\varepsilon_D$ depends only on the sampling $N$ and on the variance of the noise $\sigma_n$; hence, it is called universal threshold. However, in many applications, $\sigma_n$ is unknown and has to be estimated from the available noisy data $s$. For this, the present authors have developed an iterative algorithm (see Azzolini *et al.* (2005)), which is sketched in the following:

1. *Initialization*
   (a) given $s_k, k = 0, \ldots, N-1$. Set $i = 0$ and compute the FWT of $s$ to obtain $\widetilde{s}_\lambda$;
   (b) compute the variance $\sigma_0^2$ of $s$ as a rough estimate of the variance of $n$ and compute the corresponding threshold $\varepsilon_0 = (2\ln N\sigma_0^2)^{1/2}$;
   (c) set the number of coefficients considered as noise $N_{\text{noise}} = N$.
2. *Main loop repeat*
   (a) set $N'_{\text{noise}} = N_{\text{noise}}$ and count the wavelet coefficients $N_{\text{noise}}$ with modulus smaller than $\varepsilon_i$;
   (b) compute the new variance $\sigma_{i+1}^2$ from the wavelet coefficients whose modulus is smaller than $\varepsilon_i$ and the new threshold $\varepsilon_{i+1} = (2(\ln N)\sigma_{i+1}^2)^{1/2}$;
   (c) set $\text{i} = \text{i} + 1$ until $(N'_{\text{noise}} == N_{\text{noise}})$.
3. *Final step*
   (a) compute $s_C$ from the coefficients with modulus larger than $\varepsilon_i$ using the inverse FWT.

**Example** To illustrate the properties of the denoising algorithm, we apply it to a one-dimensional test signal. We construct a noisy signal $s$ by superposing a Gaussian white noise, with zero mean and variance $\sigma_W^2 = 1$, to a function $f$, normalized such that $((1/N)\sum_k |f_k|^2)^{1/2} = 10$. The number of samples is

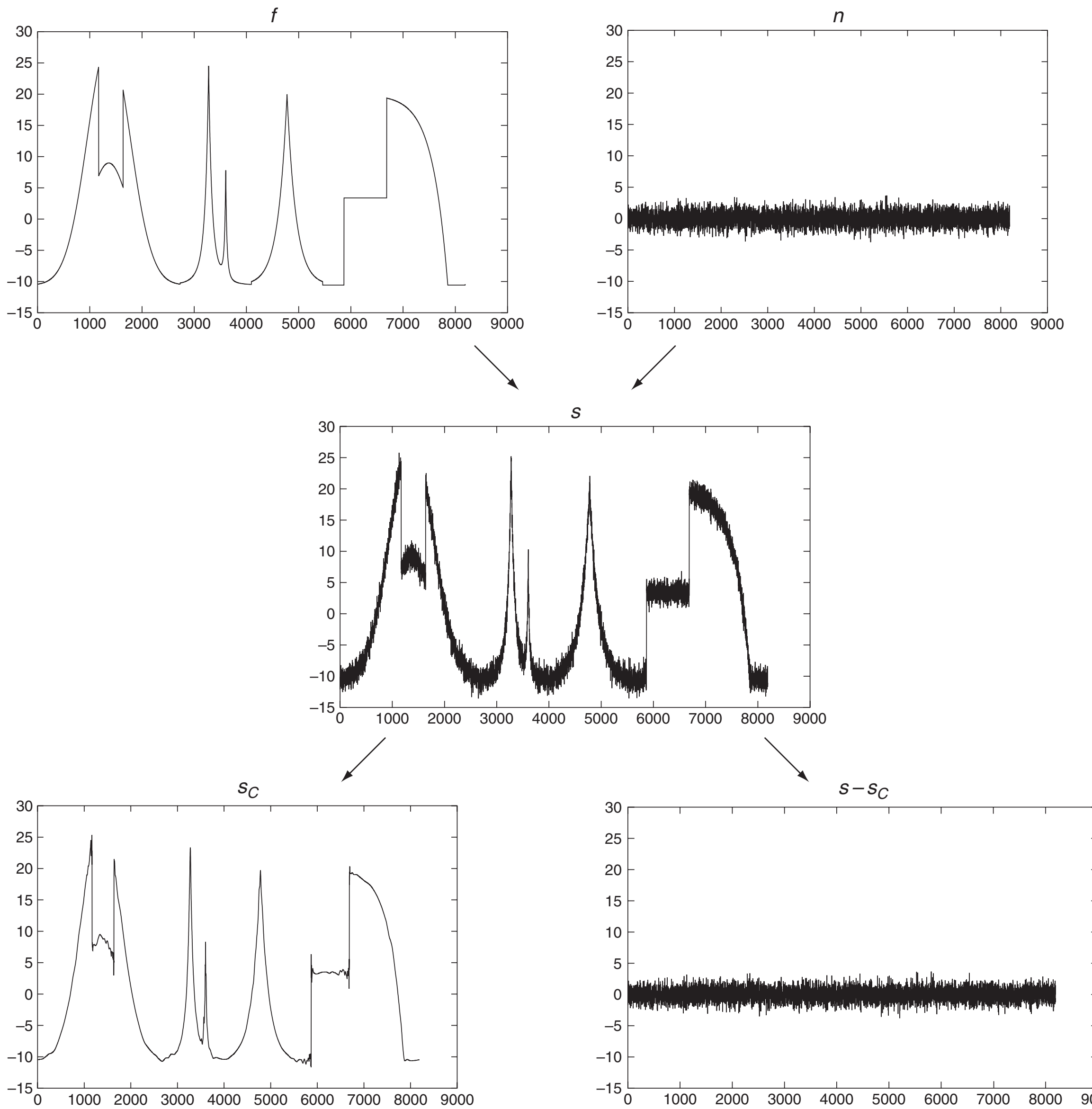



**Figure 9** Construction (top) of a 1D noisy signal $s = f + n$ (middle), and results obtained by the recursive denoising algorithm (bottom).

$N = 8192$. Figure 9a shows the function $f$ together with the noise $n$; Figure 9b shows the constructed noisy signal $s$ and Figure 9c shows the wavelet denoised signal $s_C$ together with the extracted noise.

## Acknowledgments

Marie Farge thankfully acknowledges Trinity College, Cambridge, UK, and CIRM, Marseille, France, for support while writing this paper. The authors also thank Barbara Burke for kindly revising their English.

*See also:* Coherent States; Fractal Dimensions in Dynamics; Homeomorphisms and Diffeomorphisms of the Circle; Image Processing: Mathematics; Wavelets: Application to Turbulence; Wavelets: Applications.

## Further Reading

Azzolini A, Farge M, and Schneider K (2005) Nonlinear wavelet thresholding: A recursive method to determine the optimal denoising threshold. *Applied and Computational Harmonic Analysis* 18(2): 177.

Beylkin, Coifman, and Rohklin (1991) Fast wavelet transforms and numerical algorithms. *Communications in Pure and Applied Mathematics* 44: 141.

Cohen A (2000) Wavelet methods in numerical analysis. In: Ciarlet PG and Lions JL (eds.) *Handbook of Numerical Analysis*, vol. 7. Amsterdam: Elsevier.

Dahmen W (1997) Wavelets and multiscale methods for operator equations. *Acta Numerica* 6: 55–228.
Daubechies I (1988) Orthonormal bases of compactly supported wavelets. *Communications in Pure and Applied Mathematics* 41: 909.
Daubechies I (1992) *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM.
Donoho and Johnstone (1994) Ideal spatial adaptation via wavelet shrinkage. *Biometrika* 81: 425.
Grossmann A and Morlet J (1984) Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal of Mathematical Analysis*. 15(4): 723–736.
Lemarié P-G and Meyer Yves (1986) Ondolettes et bases hilbertiennes. *Revista Matematica Iberoamericano* 2: 1.
Mallat S (1989) Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. *Transactions of the American Mathematical Society* 315: 69.
Mallat S (1998) *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic Press.

# WDVV Equations and Frobenius Manifolds

**B Dubrovin**, SISSA-ISAS, Trieste, Italy



## Main Definition

WDVV equations of associativity (after E Witten, R Dijkgraaf, E Verlinde, and H Verlinde) is tantamount to the following problem: find a function $F(v)$ of $n$ variables $v=(v^1, v^2, \ldots, v^n)$ satisfying the conditions [1], [3], and [4] given below. First,

$$\frac{\partial^3 F(v)}{\partial v^1 \partial v^\alpha \partial v^\beta} \equiv \eta_{\alpha\beta} \qquad [1]$$

must be a constant symmetric nondegenerate matrix. Denote $(\eta^{\alpha\beta}) = (\eta_{\alpha\beta})^{-1}$ the inverse matrix and introduce the functions

$$c^\gamma_{\alpha\beta}(v) = \eta^{\gamma\epsilon} \frac{\partial^3 F(v)}{\partial v^\epsilon \partial v^\alpha \partial v^\beta}, \quad \alpha, \beta, \gamma = 1, \ldots, n \qquad [2]$$

The main condition says that, for arbitrary $v^1, \ldots, v^n$ these functions must be structure constants of an associative algebra, that is, introducing a $v$-dependent multiplication law in the $n$-dimensional space by

$$a \cdot b := \left( c^1_{\alpha\beta}(v) a^\alpha b^\beta, \ldots, c^n_{\alpha\beta}(v) a^\alpha b^\beta \right)$$

one obtains an $n$-parameter family of $n$-dimensional associative algebras (these algebras will automatically be also commutative). Spelling out this condition one obtains an overdetermined system of nonlinear PDEs for the function $F(v)$ often also called WDVV associativity equations

$$\frac{\partial^3 F(v)}{\partial v^\alpha \partial v^\beta \partial v^\lambda} \eta^{\lambda\mu} \frac{\partial^3 F(v)}{\partial v^\mu \partial v^\gamma \partial v^\delta} = \frac{\partial^3 F(v)}{\partial v^\delta \partial v^\beta \partial v^\lambda} \eta^{\lambda\mu} \frac{\partial^3 F(v)}{\partial v^\mu \partial v^\gamma \partial v^\alpha} \qquad [3]$$

for arbitrary $1 \le \alpha, \beta, \gamma, \delta \le n$. (Summation over repeated indices will always be assumed.) The last one is the so-called quasihomogeneity condition

$$EF = (3-d)F + \tfrac{1}{2} A_{\alpha\beta} v^\alpha v^\beta + B_\alpha v^\alpha + C \qquad [4]$$

where

$$E = \left( a^\alpha_\beta v^\beta + b^\alpha \right) \frac{\partial}{\partial v^\alpha}$$

for some constants $a^\alpha_\beta, b^\alpha$ satisfying

$$a^\alpha_1 = \delta^\alpha_1, \qquad b^1 = 0$$

$A_{\alpha\beta}, B_\alpha, C, d$ are some constants. $E$ is called Euler vector field and $d$ is the charge of the Frobenius manifold.

For $n=1$ one has $F(v) = (1/6)v^3$. For $n=2$ one can choose

$$F(u, v) = \tfrac{1}{2} u v^2 + f(u)$$

only the quasihomogeneity [4] makes a constraint for $f(v)$. The first nontrivial case is for $n=3$. The solution to WDVV is expressed in terms of a function $f=f(x,y)$ in one of the two forms (in the examples all indices are written as lower):

$$\begin{aligned} d \neq 0: \quad & F = \tfrac{1}{2} v_1^2 v_3 + \tfrac{1}{2} v_1 v_2^2 + f(v_2, v_3) \\ & f_{xxy}^2 = f_{yyy} + f_{xxx} f_{xyy} \\ d = 0: \quad & F = \tfrac{1}{6} v_1^3 + v_1 v_2 v_3 + f(v_2, v_3) \\ & f_{xxx} f_{yyy} - f_{xxy} f_{xyy} = 1 \end{aligned} \qquad [5]$$

The function $f(x, y)$ satisfies additional constraint imposed by [4]. Because of this the above PDEs [5] can be reduced (Dubrovin 1992, 1996) to a particular case of the Painlevé-VI equation (*see* Painlevé Equations).

The problem [1], [3], [4] is invariant with respect to linear changes of coordinates preserving the direction of the vector $\partial/\partial v^1$:

$$v^\alpha \mapsto \tilde{v}^\alpha = P^\alpha_\beta v^\beta + Q^\alpha, \ \det(P^\alpha_\beta) \neq 0, \quad P^\alpha_1 = \delta^\alpha_1$$

It is also allowed to add to $F(v)$ a polynomial of the degree at most 2. To consider more general nonlinear changes of coordinates one has to give a coordinate-free form of the above equations [1], [3], [4]. This gives rise to the notion of Frobenius manifold introduced in Dubrovin (1992).

Recall that a Frobenius algebra is a pair $(A, <,>)$, where $A$ is a commutative associative algebra with a unity $e$ over a field $k$ (we will consider only the cases $k = \mathbb{R}, \mathbb{C}$) and $<,>$ is a $k$-bilinear symmetric nondegenerate invariant form on $A$, that is,

$$<x \cdot y, z> = <x, y \cdot z>$$

for arbitrary vectors $x, y, z$ in $A$.

**Definition** Frobenius structure $(\cdot, e, <,>, E, d)$ on the manifold $M$ is a structure of a Frobenius algebra on the tangent spaces $T_v M = (A_v, <,>_v)$ depending (smoothly, analytically, etc.) on the point $v \in M$. It must satisfy the following axioms.

**FM1.** The curvature of the metric $<,>_v$ on $M$ (not necessarily positive definite) vanishes. Denote $\nabla$ the Levi-Civita connection for the metric. The unity vector field $e$ must be flat, $\nabla e = 0$.

**FM2.** Let $c$ be the 3-tensor $c(x, y, z) := <x \cdot y, z>$, $x, y, z \in T_v M$. The 4-tensor $(\nabla_w c)(x, y, z)$ must be symmetric in $x, y, z, w \in T_v M$.

**FM3.** A linear vector field $E \in \mathrm{Vect}(M)$ (called Euler vector field) must be fixed on $M$, that is, $\nabla\nabla E = 0$, such that

$$\mathrm{Lie}_E(x \cdot y) - \mathrm{Lie}_E x \cdot y - x \cdot \mathrm{Lie}_E y = x \cdot y$$
$$\mathrm{Lie}_E <,> = (2 - d) <,>$$

for some number $d \in k$ called "charge."

The last condition (also called quasihomogeneity) means that the derivations $Q_{\mathrm{Func}(M)} := E$, $Q_{\mathrm{Vect}(M)} := \mathrm{id} + \mathrm{ad}_E$ define on the space $\mathrm{Vect}(M)$ of vector fields on $M$ a structure of graded Frobenius algebra over the graded ring of functions $\mathrm{Func}(M)$.

Flatness of the metric $<,>$ implies local existence of a system of flat coordinates $v^1, \ldots, v^n$ on $M$. Usually, they are chosen in such a way that

$$e = \frac{\partial}{\partial v^1}$$

is the unity vector field. In such coordinates, the problem of local classification of Frobenius manifolds reduces to the WDVV associativity equations [1], [3], [4]. Namely, $\eta_{\alpha\beta}$ is the constant Gram matrix of the metric in these coordinates

$$\eta_{\alpha\beta} := \left\langle \frac{\partial}{\partial v^\alpha}, \frac{\partial}{\partial v^\beta} \right\rangle$$

The structure constants of the Frobenius algebra $A_v = T_v M$

$$\frac{\partial}{\partial v^\alpha} \cdot \frac{\partial}{\partial v^\beta} = c^\gamma_{\alpha\beta}(v) \frac{\partial}{\partial v^\gamma} \qquad [6]$$

can be locally represented by third derivatives [2] of a function $F(v)$ satisfying [1], [3], [4]. The function $F(v)$ is called "potential" of the Frobenius manifold. It is defined up to adding of an at most quadratic polynomial in $v^1, \ldots, v^n$.

A generalization of the above definition to the case of Frobenius supermanifolds can be found in Manin (1999). For the more general class of the so-called $F$-manifolds, the requirement of the existence of a flat invariant metric has been relaxed.

## Deformed Flat Connection

One of the main geometrical structures of the theory of Frobenius manifolds is the deformed flat connection. This is a symmetric affine connection on $M \times \mathbb{C}^*$ defined by the following formulas:

$$\begin{aligned} \tilde\nabla_x y &= \nabla_x y + z x \cdot y, \quad x, y \in TM, z \in \mathbf{C}^* \\ \tilde\nabla_{\mathrm{d}/\mathrm{d}z} y &= \partial_z y + E \cdot y - \frac{1}{z}\mathcal{V} y \\ \tilde\nabla_x \frac{\mathrm{d}}{\mathrm{d}z} &= \tilde\nabla_{\mathrm{d}/\mathrm{d}z} \frac{\mathrm{d}}{\mathrm{d}z} = 0 \end{aligned} \qquad [7]$$

where, as above, $\nabla$ is the Levi-Civita connection for the metric $<,>$ and

$$\mathcal{V} := \frac{2-d}{2} - \nabla E \qquad [8]$$

is an operator on the tangent bundle $TM$ antisymmetric with respect to $<,>$,

$$<\mathcal{V}x, y> = -<x, \mathcal{V}y>$$

Observe that the unity vector field $e$ is an eigenvector of this operator with the eigenvalue

$$\mathcal{V}e = -\frac{d}{2} e$$

The connection $\tilde\nabla = \tilde\nabla(z)$ is not metric but it satisfies

$$\nabla <x, y> = <\tilde\nabla(-z)x, y> + <x, \tilde\nabla(z)y>$$
$$x, y \in TM$$

for any $z \in \mathbb{C}^*$. As it was discovered in Dubrovin (1992), vanishing of the curvature of the connection $\tilde\nabla$ is essentially equivalent to the axioms of Frobenius manifold.

**Definition** A "deformed flat function" $f(v;z)$ on a domain in $M \times \mathbb{C}^*$ is defined by the requirement of horizontality of the differential d$f$

$$\tilde{\nabla} df = 0 \qquad [9]$$

Due to vanishing of the curvature of $\tilde{\nabla}$ locally there exist $n$ independent deformed flat functions $f_1(v;z), \ldots, f_n(v;z)$ such that their differentials, together with the flat 1-form d$z$, span the cotangent plane $T^*_{(v;z)}(M \times \mathbb{C}^*)$. They will be called "deformed flat coordinates." The global analytic properties of deformed flat coordinates can be derived, for the case of semisimple Frobenius manifolds, from the results of the section "Moduli of semisimple Frobenius manifolds" discussed later.

One can relax the definition of Frobenius manifold dropping the last axiom FM3. The potential $F(v)$ in this case satisfies [1] and [3] but not [4]. In this case, the deformed flat connection $\tilde{\nabla}$ is just a family of affine flat connections on $M$ depending on the parameter $z \in \mathbb{C}$ given by the first line in [7]. The curvature and torsion of this family of connections vanishes identically in $z$. The deformed flat functions of $\tilde{\nabla}$ defined as in [9] can be chosen in the form of power series in $z$. The flatness equations written in the flat coordinates on $M$ yield a recursion equation for the coefficients of these power series

$$\tilde{\nabla} df = 0, \quad f = \sum_{p \geq 0} \theta_p(v) z^p$$
$$\partial_\lambda \partial_\mu f = z c^\nu_{\lambda\mu}(v) \partial_\nu f$$

$$\left.\begin{aligned} \partial_\lambda \partial_\mu \theta_0(v) &= 0 \\ \partial_\lambda \partial_\mu \theta_{p+1}(v) &= c^\nu_{\lambda\mu}(v) \partial_\nu \theta_p(v) \end{aligned}\right. \quad p \geq 0 \qquad [10]$$

Thus, $f(v;0)$ is just an affine linear function of the flat coordinates $v^1, \ldots, v^n$; the dependence on $z$ can be considered as a deformation of the affine structure. This motivates the name "deformed flat coordinates." The coefficients of the expansions of the deformed flat coordinates are the leading terms of the $\varepsilon$-expansion of the Hamiltonian densities of the integrable hierarchies associated with the Frobenius manifolds (see below).

## Intersection Form of a Frobenius Manifold

Another important geometric structure on $M$ is the intersection form of the Frobenius manifold. It is a symmetric bilinear form on the cotangent bundle $T^*M$ defined by the formula

$$(\omega_1, \omega_2) = \mathrm{i}_E \omega_1 \cdot \omega_2, \quad \omega_1, \omega_2 \in T^*M \qquad [11]$$

Here the multiplication law on the cotangent planes is defined by means of the isomorphism.

$$<\,,\,>: TM \to T^*M$$

The discriminant $\Sigma \subset M$ is a proper analytic (for an analytic $M$) subset where the intersection form degenerates. One can introduce a new metric on the open subset $M \backslash \Sigma$ taking the inverse of the intersection form. A remarkable result of the theory of Frobenius manifolds is vanishing of the curvature of this new metric. Moreover, the new flat metric together with the following new multiplication:

$$x * y := x \cdot y \cdot E^{-1}$$

defines on $M \backslash \Sigma$ a structure of an almost-dual Frobenius manifold (Dubrovin 2004). In the original flat coordinates $v^1, \ldots, v^n$ the coordinate expressions for the new metric and for the associated Levi-Civita connection $\nabla^*$, called the Gauss–Manin connection, read

$$\begin{gathered} g^{\alpha\beta}(v) := (\mathrm{d}v^\alpha, \mathrm{d}v^\beta) = E^\gamma(v) c^{\alpha\beta}_\gamma(v) \\ \nabla^{*\alpha}\, \mathrm{d}v^\beta = \Gamma^{\alpha\beta}_\gamma(v)\, \mathrm{d}v^\gamma \\ \Gamma^{\alpha\beta}_\gamma(v) := -g^{\alpha\nu}(v) \Gamma^\beta_{\nu\gamma}(v) = c^{\alpha\epsilon}_\gamma(v) \left(\frac{1}{2} - \mathcal{V}\right)^\beta_\epsilon \end{gathered} \qquad [12]$$

The pair $(\,,\,)$ and $<\,,\,>$ of bilinear forms on $T^*M$ possesses the following property crucial for understanding the relationships between Frobenius manifolds and integrable systems: they form a flat pencil. That means that on the complement to the subset

$$\Sigma_\lambda := \left\{ v \in M \,|\, \det\left(g^{\alpha\beta}(v) - \lambda \eta^{\alpha\beta}\right) = 0 \right\}$$

The inverse to the bilinear form

$$(\,,\,)_\lambda := (\,,\,) - \lambda <\,,\,> \qquad [13]$$

defines a metric with vanishing curvature. Flat functions $p = p(v;\lambda)$ for the flat metric are determined from the system

$$(\nabla^* - \lambda \nabla)\, \mathrm{d}p = 0 \qquad [14]$$

They are called "periods" of the Frobenius manifold. The periods $p(v;\lambda)$ are related to the deformed flat functions $f(v;z)$ by the suitably regularized Laplace-type integral transform

$$p(v;\lambda) = \int_0^\infty \mathrm{e}^{-\lambda z} f(v;z) \frac{\mathrm{d}z}{\sqrt{z}} \qquad [15]$$

Choosing a system of $n$ independent periods, one obtains a system of flat coordinates $p^1(v;\lambda), \ldots, p^n(v;\lambda)$ for the metric $(\,,\,)_\lambda$ on $M \backslash \Sigma_\lambda$,

$$\left(\mathrm{d}p^i(v;\lambda), \mathrm{d}p^j(v;\lambda)\right)_\lambda = G^{ij} \qquad [16]$$

for some constant nondegenerate matrix $G^{ij}$.

The structure of a flat pencil on the Frobenius manifold $M$ gives rise to a natural Poisson pencil (= bi-Hamiltonian structure) on the infinite-dimensional "manifold" $\mathcal{L}(M)$ consisting of smooth maps of a circle to $M$ (the so-called loop space). In the flat coordinates $v^1, \ldots, v^n$ for the metric $< , >$ the Poisson pencil has the form

$$\begin{aligned}\{v^\alpha(x), v^\beta(y)\}_1 &= \eta^{\alpha\beta}\delta'(x-y)\\ \{v^\alpha(x), v^\beta(y)\}_2 &= g^{\alpha\beta}(v(x))\delta'(x-y)\\ &\quad + \Gamma^{\alpha\beta}_\gamma(v(x))v^\gamma_x\delta(x-y)\end{aligned} \quad [17]$$

By definition of the Poisson pencil, the linear combination $a_1\{,\}_1 + a_2\{,\}_2$ of the Poisson brackets is again a Poisson bracket for arbitrary constants $a_1, a_2$. Choosing a system of $n$ independent periods $p^i(v;\lambda), i = 1, \ldots, n$, as a new system of dependent variables, one obtains a reduction of the Poisson bracket $\{,\}_\lambda := \{,\}_2 - \lambda\{,\}_1$ for a given $\lambda$ to the canonical form

$$\{p^i(v(x);\lambda), p^j(v(y);\lambda)\}_\lambda = G^{ij}\delta'(x-y) \quad [18]$$

Under an additional assumption of existence of tau function (Dubrovin 1996, Dubrovin and Zhang), one can prove that any Poisson pencil on $\mathcal{L}(M)$ of the form [17] with a nondegenerate matrix $(\eta^{\alpha\beta})$ comes from a Frobenius structure on $M$.

## Canonical Coordinates on Semisimple Frobenius Manifolds

**Definition** The Frobenius manifold $M$ is called semisimple if the algebras $T_vM$ are semisimple for $v$ belonging to an open dense subset in $M$.

Any $n$-dimensional semisimple Frobenius algebra over $\mathbb{C}$ is isomorphic to the orthogonal direct sum of $n$ copies of one-dimensional algebras. In this section, all the manifolds will be assumed to be complex analytic.

Near a semisimple point, the roots $u_i = u_i(v)$, $i = 1, \ldots, n$, of the characteristic equation

$$\det\left(g^{\alpha\beta}(v) - \lambda\eta^{\alpha\beta}\right) = 0 \quad [19]$$

can be used as local coordinates. The vectors $\partial/\partial u_i, i = 1, \ldots, n$, are basic idempotents of the algebras $T_vM$

$$\frac{\partial}{\partial u_i}\cdot\frac{\partial}{\partial u_j} = \delta_{ij}\frac{\partial}{\partial u_i}$$

We call $u_1, \ldots, u_n$ "canonical coordinates." Observe that we violate the indices convention labeling the canonical coordinates by subscripts. We will never use summation over repeated indices when working in the canonical coordinates. Actually, existence of canonical coordinates can be proved without using [4] (see details in Dubrovin (1992)).

Choosing locally branches of the square roots

$$\psi_{i1}(u) := \sqrt{<\partial/\partial u_i, \partial/\partial u_i>}, \quad i = 1, \ldots, n \quad [20]$$

we obtain a transition matrix $\Psi = (\psi_{i\alpha}(u))$,

$$\frac{\partial}{\partial v^\alpha} = \sum_{i=1}^n \frac{\psi_{i\alpha}(u)}{\psi_{i1}(u)}\frac{\partial}{\partial u_i} \quad [21]$$

from the basis $\partial/\partial v^\alpha$ to the orthonormal basis

$$\begin{aligned}\langle f_i, f_j\rangle &= \delta_{ij}\\ f_1 &= \psi_{11}^{-1}(u)\frac{\partial}{\partial u_1}\\ f_2 &= \psi_{21}^{-1}(u)\frac{\partial}{\partial u_2}, \ldots\\ f_n &= \psi_{n1}^{-1}(u)\frac{\partial}{\partial u_n}\end{aligned} \quad [22]$$

The matrix $\Psi(u)$ satisfies orthogonality condition

$$\Psi^*(u)\Psi(u) \equiv \eta, \quad \eta = (\eta_{\alpha\beta}), \quad \eta_{\alpha\beta} := \left\langle\frac{\partial}{\partial v^\alpha}, \frac{\partial}{\partial v^\beta}\right\rangle$$

In this formula $\Psi^*$ stands for the transposed matrix. The lengths [20] coincide with the first column of this matrix.

Denote $V(u) = (V_{ij}(u))$ the matrix of the antisymmetric operator $\mathcal{V}$ [8] with respect to the orthonormal frame

$$V(u) := \Psi(u)\mathcal{V}\Psi^{-1}(u) \quad [23]$$

The antisymmetric matrix $V(u) = (V_{ij}(u))$ satisfies the following system of commuting time-dependent Hamiltonian flows on the Lie algebra $\mathrm{so}(n)$ equipped with the standard Lie–Poisson brackets $\{V_{ij}, V_{kl}\} = V_{il}\delta_{jk} - V_{jl}\delta_{ik} + V_{jk}\delta_{il} - V_{ik}\delta_{jl}$:

$$\frac{\partial V}{\partial u_i} = \{V, H_i(V;u)\}, \quad i = 1, \ldots, n \quad [24]$$

with quadratic Hamiltonians

$$H_i(V;u) = \frac{1}{2}\sum_{j\neq i}\frac{V_{ij}^2}{u_i - u_j} \quad [25]$$

The matrix $\Psi(u)$ satisfies

$$\begin{aligned}\frac{\partial\Psi}{\partial u_i} &= V_i(u)\Psi,\\ V_i(u) &:= \mathrm{ad}_{E_i}\,\mathrm{ad}_U^{-1}(V(u)),\ i = 1, \ldots, n\end{aligned} \quad [26]$$

Here the matrix unity $E_i$ has the entries $(E_i)_{ab} = \delta_{ai}\delta_{ib}, U = \mathrm{diag}(u_1, \ldots, u_n)$. Conversely, given a solution to [24] and [26], one can reconstruct the

Frobenius manifold structure by quadratures (Dubrovin 1998). The reconstruction depends on a choice of an eigenvector of the constant matrix $\mathcal{V} = \Psi^{-1}(u)\ V(u)\Psi(u)$.

The system [24] coincides with the equations of isomonodromic deformations (*see* Isomonodromic Deformations) of the following linear differential operator with rational coefficients:

$$\frac{dY}{dz} = \left(U + \frac{V}{z}\right)Y \qquad [27]$$

The latter is nothing but the last component of the deformed flat connection [7] written in the orthonormal frame [22]. Other components of the horizontality equations yield

$$\partial_i Y = (zE_i + V_i(u))Y, \quad i = 1, \ldots, n \qquad [28]$$

The compatibility conditions of the system [27] and [28] coincide with [24].

The integration of [24], [26] and, more generally, the reconstruction of the Frobenius structure can be reduced to a solution of a certain Riemann–Hilbert problem (*see* Riemann–Hilbert Problem).

The isomonodromic tau function of the semisimple Frobenius manifold is defined by

$$d \log \tau_I(u) = \sum_{i=1}^{n} H_i(V(u); u) du_i \qquad [29]$$

It is an analytic function on a suitable unramified covering of the semisimple part of $M$.

Alternatively, eqns [24] can be represented as the isomonodromy deformations of the dual Fuchsian system

$$[U - \lambda]\frac{d\phi}{d\lambda} = \left(\frac{1}{2} + V\right) \qquad [30]$$

The latter comes from the Gauss–Manin system for the periods $p = p(v; \lambda)$ of the Frobenius manifold written in the canonical coordinates [22].

## Moduli of Semisimple Frobenius Manifolds

All $n$-dimensional semisimple Frobenius manifolds form a finite-dimensional space. They depend on $n(n-1)/2$ essential parameters. To parametrize the Frobenius manifolds one can choose, for example, the initial data for the isomonodromy deformation equations [24]. Alternatively, they can be parametrized by monodromy data of the deformed flat connection according to the following construction.

The first part of the monodromy data is the spectrum $(V, <\,,\,>, \hat{\mu}, R)$ of the Frobenius manifold associated with the Poisson pencil. Here $V$ is an $n$-dimensional linear space equipped with a symmetric nondegenerate bilinear form $<\,,\,>$. Two linear operators on $V$, a semisimple operator $\hat{\mu} : V \to V$, and a nilpotent operator $R : V \to V$ must satisfy the following properties. First, the operator $\hat{\mu}$ is antisymmetric:

$$\hat{\mu}^* = -\hat{\mu} \qquad [31]$$

and the operator $R$ satisfies

$$R^* = -e^{-\pi i \hat{\mu}} R\, e^{\pi i \hat{\mu}} \qquad [32]$$

Here the adjoint operators are defined with respect to the bilinear form $<\,,\,>$. The last condition to be imposed onto the operator $R$ can be formulated in a simple way by choosing a basis $e_1, \ldots, e_n$ of eigenvectors of the semisimple operator $\hat{\mu}$,

$$\hat{\mu} e_\alpha = \mu_\alpha e_\alpha, \quad \alpha = 1, \ldots, n$$

We require the existence of a decomposition

$$R = R_0 + R_1 + R_2 + \cdots \qquad [33]$$

where for any integer $k \geq 0$ the linear operator $R_k$ satisfies

$$R_k e_\alpha \in \text{span}\{e_\beta \,|\, \mu_\beta = \mu_\alpha + k\} \ \ \forall \alpha = 1, \ldots, n \qquad [34]$$

In the nonresonant case, such that none of the differences of the eigenvalues of $\hat{\mu}$ being equal to a positive integer, all the matrices $R_1, R_2, \ldots$, are equal to zero. Observe a useful identity

$$z^{\hat{\mu}} R\, z^{-\hat{\mu}} = R_0 + zR_1 + z^2 R_2 + \cdots \qquad [35]$$

More generally, for any operator $A : V \to V$ commuting with $e^{2\pi i \hat{\mu}}$ a decomposition is defined as

$$A = \bigoplus_{k \in \mathbb{Z}} [A]_k$$
$$z^{\hat{\mu}} A\, z^{-\hat{\mu}} = \sum_{k \in \mathbb{Z}} z^k [A]_k \qquad [36]$$

In particular, $[R]_k = R_k, k \geq 0$, $[R]_k = 0$, $k < 0$.

One has to also choose an eigenvector $e$ of the operator $\hat{\mu}$ such that $R_0 e = 0$; denote $-d/2$ the corresponding eigenvalue

$$e \in V, \qquad \hat{\mu} e = -\frac{d}{2} e, \qquad R_0 e = 0 \qquad [37]$$

The second part of the monodromy data is a pair of linear operators

$$C : V \to \mathbb{C}^n, \qquad S : \mathbb{C}^n \to \mathbb{C}^n$$

The space $\mathbb{C}^n$ is assumed to be equipped with the standard complex Euclidean structure given by the sum of squares. The properties of the operators $S$, $C$ depend on the choice of an unordered set

$u^0 = (u_1^0, \ldots, u_n^0)$ of $n$ pairwise distinct complex numbers and on a choice of a ray $\ell_+$ on an auxiliary complex $z$-plane starting at the origin such that

$$\mathrm{Re}\, z\left(u_i^0 - u_j^0\right) \neq 0, \quad i \neq j,\ z \in \ell_+ \qquad [38]$$

Let us order the complex numbers in such a way that

$$e^{z(u_i^0 - u_j^0)} \to 0, \quad i < j, |z| \to \infty,\ z \in \ell_+ \qquad [39]$$

The operator $S$ must be upper triangular

$$\begin{aligned} S = (S_{ij}), \quad S_{ij} = 0,\ i > j \\ S_{ii} = 1, \qquad i = 1, \ldots, n \end{aligned} \qquad [40]$$

The operator $C$ must satisfy

$$C^* S\, C = e^{\pi i \hat{\mu}}\, e^{\pi i R} \qquad [41]$$

Here the adjoint operator $C^*$ is understood as follows:

$$C^*\colon \mathbb{C}^n \xrightarrow{\simeq} \mathbb{C}^{n*} \to V^* \xrightarrow{<,>^{-1}} V$$

The group of diagonal $n \times n$ matrices

$$D = \mathrm{diag}(\pm 1, \ldots, \pm 1)$$

acts on the pairs $(S, C)$ by

$$S \mapsto DSD, \qquad C \mapsto DC$$

One is to factor out the action of this diagonal group. Besides, the operator $C$ is defined up to a left action of certain group of linear operators depending on the spectrum.

For the generic (i.e., nonresonant) case where $e^{2\pi i \hat{\mu}}$ has simple spectrum, the operator $C$ is defined up to left multiplication by any matrix commuting with $e^{2\pi i \hat{\mu}}$. In this situation, the monodromy data $(\hat{\mu}, R, S, C)$ are locally uniquely determined by the $n(n-1)/2$ entries of the matrix $S$. Therefore, near a generic point, the variety of the monodromy data is a smooth manifold of the dimension $n(n-1)/2$. At nongeneric points, the variety can get additional strata.

The monodromy data $S$, $C$ are determined at an arbitrary semisimple point of a Frobenius manifold in terms of the analytic properties of horizontal sections of the deformed flat connection $\tilde{\nabla}$ [7] in the complex $z$-plane (the so-called "Stokes matrix" and the "central connection matrix" of the operator [27]). Locally, they do not depend on the point of the semisimple Frobenius manifold (the isomonodromicity property).

We will now describe the reconstruction procedure giving a parametrization of semisimple Frobenius manifolds in terms of the monodromy data $(\hat{\mu}, R, S, C)$.

Conversely, to reconstruct the Frobenius manifold near a semisimple point with the canonical coordinates $u_1^0, \ldots, u_n^0$, one is to solve the following boundary-value problem. Let

$$\ell = (-\ell_-) \cup \ell_+$$

be the oriented line on the complex $z$-plane chosen as in [38]. Here the ray $\ell_-$ is the opposite to $\ell_+$. Denote $\Pi_R/\Pi_L$ the right/left half-planes with respect to $\ell$. To reconstruct the Frobenius manifold, one is to find three matrix-valued functions $\Phi_0(z; u)$, $\Phi_R(z; u)$, and $\Phi_L(z; u)$:

$$\begin{aligned} \Phi_0(z; u) &: V \to \mathbb{C}^n \\ \Phi_{R/L}(z; u) &: \mathbb{C}^n \to \mathbb{C}^n \end{aligned}$$

for $u$ close to $u^0$ such that $\Phi_0(z; u)$ is analytic and invertible for $z \in \mathbb{C}, \Phi_R(z; u)/\Phi_L(z; u)$ are analytic and invertible for $z \in \Pi_R/\Pi_L$ resp., and continuous up to the boundary $\ell \backslash 0$ and

$$\Phi_{R/L}(z; u) \sim 1 + O(1/z), \quad |z| \to \infty,\ z \in \Pi_R/\Pi_L$$

The boundary values of the functions $\Phi_0(z; u)$,$\Phi_R(z; u)$, and $\Phi_L(z; u)$ must satisfy the following boundary-value problem (as above $U = \mathrm{diag}(u_1, \ldots, u_n)$):

$$\Phi_R(z; u) = \Phi_L(z; u) e^{zU} S e^{-zU}, \quad z \in \ell_+ \qquad [42]$$

$$\Phi_R(z; u) = \Phi_L(z; u) e^{zU} S^* e^{-zU}, \quad z \in \ell_- \qquad [43]$$

$$\begin{aligned} \Phi_0(z; u) z^{\hat{\mu}} z^R &= \Phi_R(z; u) e^{zU} C, \quad z \in \Pi_R \\ \Phi_0(z; u) z^{\hat{\mu}} z^R &= \Phi_L(z; u) e^{zU} SC, \quad z \in \Pi_L \end{aligned} \qquad [44]$$

Here $z^{\hat{\mu}} := e^{\hat{\mu} \log z}$, $z^R := e^{R \log z}$ are considered as $\mathrm{Aut}(V)$-valued functions on the universal covering of $\mathbb{C} \backslash 0$; the branch cut in the definition of $\log z$ is chosen to be along $\ell_-$.

The solution of the above boundary-value problem [42]–[44], if exists, is unique. It can be reduced to a certain Riemann–Hilbert problem, that is, to a problem of factorization of an analytic $n \times n$ nondegenerate matrix-valued function on the annulus

$$G(z; u), \quad r < |z| < R, \quad \det G(z; u) \neq 0$$

depending on the parameter $u = (u_1, \ldots, u_n)$ in a product

$$G(z; u) = G_0(z; u)^{-1} G_\infty(z; u) \qquad [45]$$

of two matrix-valued functions $G_0(z;\, u)$ and $G_\infty(z;\, u)$ analytic for $|z| < R$ and $r < |z| \leq \infty$ resp., with nowhere-vanishing determinant.

Existence of a solution to the Riemann–Hilbert problem for a given $u = (u_1, \ldots, u_n), u_i \neq u_j$ for $i \neq j$, means triviality of certain $n$-dimensional vector bundle over the Riemann sphere with the transition functions given by $G(z; u)$. Existence of the solution for $u = u^0$ implies solvability of the Riemann–Hilbert problem for $u$ sufficiently close to $u^0$. From these arguments, it can be deduced that the matrices $\Phi_0(z; u), \Phi_{R/L}(z; u)$ are analytic in $(z; u)$ for $u$ sufficiently close to $u^0$. Moreover, they can be analytically continued in $u$ to the universal covering of the space of configurations of $n$ distinct points on the complex plane:

$$\left(\mathbb{C}^n \backslash \cup_{i \neq j} \{u_i = u_j\}\right) / S_n \qquad [46]$$

The resulting functions are meromorphic on the universal covering, according to the results of B Malgrange and T Miwa. The structure of the global analytic continuation is given (Dubrovin 1999) in terms of a certain action of the braid group

$$B_n = \pi_1\left(\left(\mathbb{C}^n \backslash \cup_{i \neq j} \{u_i = u_j\}\right) / S_n\right)$$

on the monodromy data.

## Examples of Frobenius Manifolds

**Example 0** Trivial Frobenius manifold, $M = A_0$ a graded Frobenius algebra, $F(v) = (1/6) < e, v \cdot v \cdot v >$ is a cubic polynomial.

First nontrivial examples appeared in the setting of 2D topological field theories (Dijkgraaf *et al.* 1991, Witten 1991) (*see* Topological Quantum Field Theory: Overview). Mathematical formalization of these ideas gives rise to the following two classes of examples.

**Example 1** Frobenius structure on the base of an isolated hypersurface singularity. The construction (Hertling 2002, Sabbah 2002) uses the K Saito theory of periods of primitive forms. For the example of $A_n$ singularity $f(x) = x^{n+1}$ the Frobenius structure on the base of universal unfolding

$$M_{A_n} = \{f_s(x) = x^{n+1} + s_1 x^{n-1} + \cdots + s_n \,|\, s_1, \ldots, s_n \in \mathbb{C}\}$$

is constructed as follows (Dijkgraaf *et al.* 1991):

$$e = \frac{\partial}{\partial s_n}$$
$$E = \frac{1}{n+1} \sum (k+1) s_k \frac{\partial}{\partial s_k}$$
$$d = \frac{n-1}{n+1}$$

The multiplication is introduced by identifying the tangent space $T_s M$ with the quotient algebra

$$T_s M_{A_n} = \mathbb{C}[x]/(f_s'(x))$$

The metric has the form

$$< \partial_{s_i}, \partial_{s_j} > = -(n+1)\, \mathrm{res}_{x=\infty} \frac{\partial f_s(x)/\partial s_i \partial f_s(x)/\partial s_j}{f_s'(x)}\, dx$$

The flat coordinates $v_\alpha = v_\alpha(s)$ can be found from the expansion of the solution to the equation $f_s(x) = k^{n+1}$,

$$x = k - \frac{1}{n+1}\left(\frac{v_n}{k} + \frac{v_{n-1}}{k^2} + \cdots + \frac{v_1}{k^n}\right) + O\left(\frac{1}{k^{n+2}}\right)$$

The potentials of the Frobenius manifolds $M_{A_n}$ for $n = 1, 2, 3$ read

$$\begin{aligned} F_{A_1} &= \tfrac{1}{6} v_1^3 \\ F_{A_2} &= \tfrac{1}{2} v_1^2 v_2 + \tfrac{1}{72} v_2^4 \\ F_{A_3} &= \tfrac{1}{2} v_1 v_2^2 + \tfrac{1}{2} v_1^2 v_3 + \tfrac{1}{16} v_2^2 v_3^2 + \tfrac{1}{960} v_3^5 \end{aligned} \qquad [47]$$

The space of polynomials $M_{A_n}$ can be identified with the orbit space of $\mathbb{C}/W(A_n)$ of the Weyl group of the type $A_n$. More generally (Dubrovin 1996), the orbit space $M_W := \mathbb{C}^n/W$ of an arbitrary irreducible finite Coxeter group $W \subset O(n)$ carries a natural structure of a polynomial semisimple Frobenius manifold. Conversely, all irreducible polynomial semisimple Frobenius manifolds with positive degrees of the flat coordinates can be obtained by this construction (Hertling 2002). Generalizations for the orbit spaces of certain infinite groups were obtained in Dubrovin and Zhang (1998b) and Bertola (2000).

**Example 2** Gromov–Witten (GW) invariants (*see* Topological Sigma Models). Let $X$ be a smooth projective variety. We will assume for simplicity that $H^{\mathrm{odd}}(X) = 0$. To every such variety, one can associate a bunch of rational numbers. They are expressed in terms of intersection theory of certain cycles on the moduli spaces $X_{g,m,\beta}$ of stable genus $g$ and degree $\beta$ curves on $X$ with $m$ marked points (see details in Kontsevich and Manin (1994)):

$$\begin{aligned} X_{g,m,\beta} := \{ & f : (C_g, x_1, \ldots, x_m) \to X, \\ & f_*[C_g] = \beta \in H_2(X; \mathbf{Z})\} \end{aligned} \qquad [48]$$

Denote $n := \dim H^*(X; \mathbb{C})$. Choosing a basis $\phi_1 = 1, \phi_2, \ldots, \phi_n$ we define the numbers

$$\begin{aligned} & < \tau_{p_1}(\phi_{\alpha_1}) \ldots \tau_{p_m}(\phi_{\alpha_m}) >_{g,\beta} \\ & := \int_{[X_{g,m,\beta}]^{\mathrm{virt}}} ev_1^*(\phi_{\alpha_1}) \wedge c_1^{p_1}(\mathcal{L}_1) \\ & \quad \wedge \cdots \wedge ev_m^*(\phi_{\alpha_m}) \wedge c_1^{p_m}(\mathcal{L}_m) \end{aligned} \qquad [49]$$

for arbitrary non-negative integers $p_1, \ldots, p_m$. Here the evaluation maps $ev_i, i = 1, \ldots, m$, are given by

$$ev_i : X_{g,m,\beta} \to X, \qquad f \mapsto f(x_i)$$

The so-called tautological line bundles $\mathcal{L}_i$ over $X_{g,m,\beta}$ by definition have the fiber $T^*_{x_i} C_g, i = 1, \ldots, m$ (see the article Moduli Spaces: An Introduction regarding the construction of the so-called virtual fundamental class $[X_{g,m,\beta}]^{\text{virt}}$). The numbers [49] can be defined for an arbitrary compact symplectic manifold $X$ where one is to deal with the intersection theory on the moduli spaces of pseudoholomorphic curves fixing a suitable almost-complex structure on $X$. They depend only on the symplectic structure on $X$. In particular, the numbers

$$<\tau_0(\phi_{\alpha_1}) \ldots \tau_0(\phi_{\alpha_m})>_{g,\beta} \qquad [50]$$

are called the genus $g$ and degree $\beta$ GW invariants of $X$. In certain cases, they admit an interpretation in terms of enumerative geometry of the variety $X$ (Kontsevich and Manin 1994). The numbers [49] with some of $p_i > 0$ are called "gravitational descendents."

One can form a generating functions of the numbers [49]

$$\mathcal{F}_g^X = \sum_m \sum_{\beta \in H_2(X;\mathbf{Z})} \frac{1}{m!} t^{\alpha_1,p_1} \ldots t^{\alpha_m,p_m} <\tau_{p_1}(\phi_{\alpha_1}) \ldots \tau_{p_m}(\phi_{\alpha_m})>_{g,\beta} \qquad [51]$$

(summation over repeated indices $1 \leq \alpha_1, \ldots, \alpha_m \leq n$ will always be assumed). Here $t^{\alpha,p}$ are indeterminates labeled by pairs $(\alpha, p)$ with $\alpha = 1, \ldots, n$, $p = 0, 1, 2, \ldots$. (Usually one is to insert in the definition of $\mathcal{F}_g^X$ elements $q^\beta$ of the Novikov ring $\mathbb{C}[H_2(X; \mathbf{Z})]$. However, due to the divisor axiom (Kontsevich and Manin 1994) and these insertions can be compensated by a suitable shift in the space of couplings $\boldsymbol{t} = (t^{\alpha,p})$.) We finally introduce the full generating function called total GW potential (it is also called the free energy of the topological sigma model with the target space $X$)

$$\mathcal{F}^X(\boldsymbol{t}; \epsilon) = \sum_{g \geq 0} \epsilon^{2g-2} \mathcal{F}_g^X \qquad [52]$$

Restricting the genus-zero generating function onto the so-called small phase space

$$F^X(v) := \mathcal{F}_0^X(t^{\alpha,0} = v^\alpha, t^{\alpha,p>0} = 0) \qquad v = (v^1, \ldots, v^n) \qquad [53]$$

one obtains a solution to the WDVV associativity equations. This solution defines a structure of (formal) Frobenius manifold on $H^*(X)$ with the bilinear form $\eta$ given by the Poincaré pairing

$$\eta_{\alpha\beta} = \int_X \phi_\alpha \wedge \phi_\beta$$

the unity

$$e = \frac{\partial}{\partial v^1}$$

and the Euler vector field

$$E = \sum_{\alpha=1}^n [(1 - q_\alpha) v^\alpha + r_\alpha] \frac{\partial}{\partial v^\alpha}$$

Here the numbers $q_\alpha, r_\alpha$ are defined by the conditions

$$\phi_\alpha \in H^{2q_\alpha}(X), \qquad c_1(X) = \sum_\alpha r_\alpha \phi_\alpha$$

The resulting Frobenius manifold will be denoted $M_X$. The corresponding $n$-parameter family of $n$-dimensional algebras on the tangent spaces $T_v M_X$ is also called "quantum cohomology" $\mathrm{QH}^*(X)$. At the point $v_{\text{cl}} \in M_X$ of classical limit, the algebra $T_{v_{\text{cl}}} M_X$ coincides with the cohomology ring $H^*(X)$. In all known examples, the series [53] actually converges in a neighborhood of the point $v_{\text{cl}}$. Therefore, one obtains a genuine Frobenius structure on a domain $M_X \subset H^*(X; \mathbb{C})/2\pi \mathrm{i} H_2(X; \mathbf{Z})$. However, a general proof of convergence is still missing.

In particular, for $d = 1$, the quantum cohomology of complex projective line $\boldsymbol{P}^1$ is a two-dimensional Frobenius manifold with the potential, unity, and the Euler vector field

$$F(u, v) = \tfrac{1}{2} u v^2 + e^u, \quad e = \frac{\partial}{\partial v}, \quad E = v \frac{\partial}{\partial v} + 2 \frac{\partial}{\partial u}$$

For $d = 2$ one has a three-dimensional Frobenius manifold $QH^*(\boldsymbol{P}^2)$ with

$$\begin{aligned} F(v_1, v_2, v_3) &= \tfrac{1}{2} v_1^2 v_3 + \tfrac{1}{2} v_1 v_2^2 + \sum_{k \geq 1} N_k \frac{v_3^{3k-1}}{(3k-1)!} e^{k v_2} \\ e &= \frac{\partial}{\partial v_1} \\ E &= v_1 \frac{\partial}{\partial v_1} + 3 \frac{\partial}{\partial v_2} - v_3 \frac{\partial}{\partial v_3} \end{aligned} \qquad [54]$$

where $N_k$ = number of rational curves on $\boldsymbol{P}^2$ passing through $3k - 1$ generic points. WDVV [5] yields (Kontsevich and Manin 1994) recursion relations for

the numbers $N_k$ starting from $N_1=1$. The closed analytic formula for the function [54] is still unknown.

Only for certain very exceptional $X$ the Frobenius manifold $M_X$ is semisimple (e.g., for $X=\boldsymbol{P}^d$). The general geometrical reasons of the semisimplicity of $M_X$ are still to have been understood.

For the case $X=$Calabi–Yau manifold, the Frobenius manifold $QH^*(X)$ is never semisimple. This Frobenius structure can be computed in terms of the mirror symmetry construction (*see* Mirror Symmetry: A Geometric Survey).

## Frobenius Manifold and Integrable Systems

The identities in the cohomology ring generated by the cocycles $ev_i^*(\phi_\alpha)$ and $\psi_j:=c_1(\mathcal{L}_j)$ can be recast into the form of differential equations for the generating function [52]. The variable $x:=t^{1,0}$ corresponding to $\phi_1=1$ plays a distinguished role in these differential equations. According to the idea of Witten (1991), the differential equations for the generating functions can be written as a hierarchy of systems of $n$ evolutionary PDEs ($n=\dim H^*(\mathrm{X})$) for the unknown functions

$$w_\alpha=\langle\langle\tau_0(\phi_\alpha)\tau_0(\phi_1)\rangle\rangle=\epsilon^2\frac{\partial^2\mathcal{F}^X(\boldsymbol{t},\epsilon)}{\partial t^{1,0}\partial t^{\alpha,0}} \qquad [55]$$

The variable $x=t^{1,0}$ is the spatial variable of the equations of the hierarchy. The remaining parameters (coupling constants) $t^{\alpha,p}$ of the generating function play the role of the time variables. Witten suggested to use the two-point correlators

$$h_{\alpha,p}=\langle\langle\tau_{p+1}(\phi_\alpha)\tau_0(\phi_1)\rangle\rangle=\epsilon^2\frac{\partial^2\mathcal{F}^X(\boldsymbol{t},\epsilon)}{\partial t^{1,0}\partial t^{\alpha,p}} \qquad [56]$$

as the densities of the Hamiltonians of the flows of the hierarchy.

Existence of such a hierarchy can be proved for the case of GW invariants (and their descendents) of complex projective spaces $\boldsymbol{P}^d$ (the results of Givental (2001) along with Dubrovin and Zhang (2005) can be used). For $d=0$ one obtains, according to the celebrated result by Kontsevich conjectured by Witten (*see* Topological Gravity, Two-Dimensional), the tau function of the solution to the KdV hierarchy (*see* Korteweg–de Vries Equation and Other Modulation Equations) specified by the initial condition,

$$u(x)\,|_{t=0}=x$$

For $d=1$ the hierarchy in question is the extended Toda lattice (see details in Dubrovin and Zhang (2004); *see also* Toda Lattices). For all other $d\geq 2$, the needed integrable hierarchy is a new one. It can be associated (Dubrovin and Zhang) with an arbitrary $n$-dimensional semisimple Frobenius manifold $M$. The equations of the hierarchy have the form

$$w_t^i=A_j^i(w)w_x^j+\epsilon^2\Big[B_j^i(w)w_{xxx}^i+C_{jk}^i(w)w_x^jw_{xx}^k + D_{jkl}^i(w)w_x^jw_x^kw_x^l\Big]+O(\epsilon^4),\quad i=1,\ldots,n \qquad [57]$$

The coefficients of $\epsilon^{2g}$ are graded homogeneous polynomials in $u_x, u_{xx}$, etc., of the degree $2g+1$,

$$\deg \mathrm{d}^m u/\mathrm{d}x^m=m$$

The construction of the hierarchy is done in two steps. First, we construct the leading approximation (Dubrovin 1992). The equation of the hierarchy specifying the dependence on $t=t^{\alpha,p}$ at $\epsilon=0$ reads

$$\frac{\partial v}{\partial t^{\alpha,p}}=\partial_x\left(\nabla\theta_{\alpha,p+1}(v)\right)\qquad \alpha=1,\ldots,n,\quad p\geq 0 \qquad [58]$$

The functions $\theta_{\alpha,p}(v), v\in M$, are the coefficients of expansion [10] of the deformed flat functions normalized by $\theta_{\alpha,0}=v_\alpha$. The solution $v=v(x,\boldsymbol{t})$ of interest is determined from the implicit function equations

$$v=xe+\sum_{\alpha,p}t^{\alpha,p}\nabla\theta_{\alpha,p}(v) \qquad [59]$$

Next, one has to find solution

$$\Delta\mathcal{F}=\sum_{g\geq 1}\epsilon^{2g-2}\mathcal{F}_g(v;v_x,\ldots,v^{(3g-2)}) \qquad [60]$$

of the following universal loop equation (closely related with the Virasoro conjecture of Eguchi and Xiong (1998)):

$$\begin{aligned}&\sum_{r\geq 0}\frac{\partial\Delta\mathcal{F}}{\partial v^{\gamma,r}}\partial_x^r\left(\frac{1}{E(v)-\lambda}\right)^\gamma\\&\quad+\sum_{r\geq 1}\frac{\partial\Delta\mathcal{F}}{\partial v^{\gamma,r}}\sum_{k=1}^r\binom{r}{k}\partial_x^{k-1}\partial_e p_\alpha G^{\alpha\beta}\partial_x^{r-k+1}\partial^\gamma p_\beta\\&=-\frac{1}{16}\mathrm{tr}(\mathcal{U}-\lambda)^{-2}+\frac{1}{4}\mathrm{tr}\Big[(\mathcal{U}-\lambda)^{-1}\mathcal{V}\Big]^2\\&\quad+\frac{\epsilon^2}{2}\sum\left(\frac{\partial^2\Delta\mathcal{F}}{\partial v^{\gamma,k}\partial v^{\rho,l}}+\frac{\partial\Delta\mathcal{F}}{\partial v^{\gamma,k}}\frac{\partial\Delta\mathcal{F}}{\partial v^{\rho,l}}\right)\\&\quad\times\partial_x^{k+1}\partial^\gamma p_\alpha G^{\alpha\beta}\partial_x^{l+1}\partial^\rho p_\beta\\&\quad+\frac{\epsilon^2}{2}\sum\frac{\partial\Delta\mathcal{F}}{\partial v^{\gamma,k}}\partial_x^{k+1}\\&\quad\times\left[\nabla\frac{\partial p_\alpha(v;\lambda)}{\partial\lambda}\cdot\nabla\frac{\partial p_\beta(v;\lambda)}{\partial\lambda}\cdot v_x\right]^\gamma G^{\alpha\beta}\end{aligned} \qquad [61]$$

Here $\mathcal{U}=\mathcal{U}(v)$ is the operator of multiplication by $E(v), p_\alpha = p_\alpha(v;\lambda), \alpha = 1,\ldots,n$, is a system of flat coordinates [16] of the bilinear form [13]. The substitution

$$v_\alpha \mapsto w_\alpha = v_\alpha + \epsilon^2 \partial_x \partial_{t^{\alpha,0}} \Delta\mathcal{F}(v; v_x, v_{xx}, \ldots; \epsilon^2) \quad \alpha = 1,\ldots,n \qquad [62]$$

transforms [58] to [57]. The terms of the expansion [60] are not polynomial in the derivatives. For example (Dubrovin and Zhang 1998a),

$$\mathcal{F}_1 = \frac{1}{24}\sum_{i=1}^{n} \log u_i' + \log \frac{\tau_I(u)}{J^{1/24}(u)}$$
$$J(u) = \det\left(\frac{\partial v^\alpha}{\partial u_i}\right) = \pm \prod_{i=1}^{n} \psi_{i1}(u) \qquad [63]$$

(the canonical coordinates have been used) where $\tau_I(u)$ is the isomonodromic tau function [29]. The transformation [62] applied to the solution [59] expresses higher-genus GW invariants of a variety $X$ with semisimple quantum cohomology $QH^*(X)$ via the genus-zero invariants. For the particular case of $X = \boldsymbol{P}^2$, the formula [63] yields (Dubrovin and Zhang 1998a)

$$\frac{\phi''' - 27}{8(27 + 2\phi' - 3\phi'')} = -\frac{1}{8} + \sum_{k\geq 1} k N_k^{(1)} \frac{e^{kz}}{(3k)!}$$

Here

$$\phi(z) = \sum_{k\geq 0} N_k \frac{e^{kz}}{(3k-1)!}$$

is the generating function of the genus-zero GW invariants of $\boldsymbol{P}^2$ (see [54]) and $N_k^{(1)}=$ the number of elliptic plane curves of the degree $k$ passing through $3k$ generic points.

*See also:* Bi-Hamiltonian Methods in Soliton Theory; Functional Equations and Integrable Systems; Integrable Systems: Overview; Isomonodromic Deformations; Korteweg–de Vries Equation and Other Modulation Equations; Mirror Symmetry: A Geometric Survey; Moduli Spaces: An Introduction; Painlevé Equations; Riemann–Hilbert Problem; Toda Lattices; Topological Gravity, Two-Dimensional; Topological Quantum Field Theory: Overview; Topological Sigma Models.

## Further Reading


Bertola M (2000) Frobenius manifold structure on orbit space of Jacobi groups. I, II. *Differential Geometry and Applications* 13: 19–41, 213–233.

Dijkgraaf R, Verlinde H, and Verlinde E (1991) Topological strings in $d < 1$. *Nuclear Physics* B 352: 59–86.

Dubrovin B (1992) Integrable systems in topological field theory. *Nuclear Physics* B 379: 627–689.

Dubrovin B (1996) Geometry of 2D topological field theories. In: Francaviglia M and Greco S (eds.) *Integrable Systems and Quantum Groups, Montecatini Terme, 1993*, Springer Lecture Notes in Math., vol. 1620, pp. 120–348.

Dubrovin B (1998) Geometry and analytic theory of Frobenius manifolds. *Proceedings of the International Congress of Mathematicians*, vol. II (Berlin, 1998). *Doc. Math.* 1998, Extra Vol. II, 315–326.

Dubrovin B (1999) Painlevé transcendents in two-dimensional topological field theory. In: Conte R (ed.) *The Painlevé Property: 100 Years Later*, CRM Ser. Math. Phys. pp. 287–412. New York: Springer.

Dubrovin B (2004) On almost duality for Frobenius manifolds. In: Buchstaber VM and Krichever IM (eds.) *Geometry, Topology, and Mathematical Physics*, Amer. Math. Soc. Transl. Ser. 2, vol. 212, pp. 75–132. Providence, RI: American Mathematical Society.

Dubrovin B and Zhang Y (1998a) Bi-Hamiltonian hierarchies in 2D topological field theory at one-loop approximation. *Communications in Mathematical Physics* 198: 311–361.

Dubrovin B and Zhang Y (1998b) Extended affine Weyl groups and Frobenius manifolds. *Compositio Math.* 111: 167–219.

Dubrovin B and Zhang Y (2004) Virasoro symmetries of the extended Toda hierarchy. *Communications in Mathematical Physics* 250: 161–193.

Dubrovin B and Zhang Y (2001) Normal forms of hierarchies of integrable PDEs, Frobenius manifolds and Gromov–Witten invariants, math/0108160.

Dubrovin B and Zhang Y (2005) Integrable hierarchies of the topological type (to appear).

Eguchi T and Xiong CS (1998) Quantum cohomology at higher genus: topological recursion relations and Virasoro conditions. *Advances in Theoretical and Mathematical Physics* 2: 219–229.

Givental A (2001) Gromov–Witten invariants and quantization of quadratic Hamiltonians. *Moscow Mathematical Journal* 1: 551–568, 645.

Hertling C (2002) Frobenius Manifolds and Moduli Spaces for Singularities, *Cambridge Tracts in Mathematics,* vol. 151. Cambridge: Cambridge University Press.

Hitchin N (1997) Frobenius manifolds. With Notes by David Calderbank. In: *Gauge Theory and Symplectic Geometry (Montreal, PQ, 1995).* , NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. vol. 488, pp. 69–112. Dordrecht: Kluwer Academic Publishers.

Kontsevich M and Manin Yu (1994) Gromov–Witten classes, quantum cohomology and enumerative geometry. *Communications in Mathematical Physics* 164: 525–562.

Manin Yu (1999) *Frobenius Manifolds, Quantum Cohomology, and Moduli Spaces,* American Mathematical Society Colloquium Publications, vol. 47. Providence, RI: American Mathematical Society.

Sabbah C (2002) Déformations isomonodromiques et variétés de Frobenius. *Savoirs Actuels (Les Ulis).* Paris: CNRS Éditions.

Witten E (1991) Two-dimensional gravity and intersection theory on moduli space. *Surv. of Differential Geometry* 1: 243–310.

# Weakly Coupled Oscillators

**E M Izhikevich**, The Neurosciences Institute, San Diego, CA, USA
**Y Kuramoto**, Hokkaido University, Sapporo, Japan



## Introduction

Practically any physical, chemical, or biological system can exhibit rhythmic oscillatory activity, at least when the conditions are right. Winfree (2001) reviews the ubiquity of oscillations in nature, ranging from autocatalytic chemical reactions to pacemaker cells in the heart, to animal gates, and to circadian rhythms. When coupled, even weakly, oscillators interact via adjustment of their phases, that is, their timing, often leading to synchronization. In this chapter, we review the most important concepts needed to study and understand the dynamics of coupled oscillators.

From a mathematical point of view, an oscillator is a dynamical system,

$$\dot{x} = f(x), \quad x \in \mathbb{R}^m \qquad [1]$$

having a limit-cycle attractor – periodic orbit $\gamma \subset \mathbb{R}^m$. Its period is the minimal $T > 0$ such that

$$\gamma(t) = \gamma(t + T) \quad \text{for any } t$$

and its frequency is $\Omega = 2\pi/T$. Let $x(0) = x_0 \in \gamma$ be an arbitrary point on the attractor, then the state of the system, $x(t)$, is uniquely defined by its phase $\vartheta \in \mathbb{S}^1$ relative to $x_0$, where $\mathbb{S}^1$ is the unit circle.

Throughout this article, we assume that the periodic orbit $\gamma$ is exponentially stable, which implies normal hyperbolicity. In this case, there is a continuous transformation $\Theta : U \to \mathbb{S}^1$ defined in a neighborhood $U \supset \gamma$ such that $\vartheta(t) = \Theta(x(t))$ for any trajectory in $U$, that is, $\Theta$ maps solutions of [1] to solutions of

$$\dot{\vartheta} = \Omega \qquad [2]$$

Such a transformation removes the amplitude but saves the phase of oscillation.

Accordingly, there is a continuous transformation that maps solutions of the weakly coupled network of $n$ oscillators,

$$\dot{x}_i = f_i(x_i) + \varepsilon g_i(x_1, \ldots, x_n, \varepsilon), \quad \varepsilon \ll 1 \qquad [3]$$

onto solutions of the phase system

$$\dot{\vartheta}_i = \Omega_i + \varepsilon h_i(\vartheta_1, \ldots, \vartheta_n, \varepsilon), \quad \vartheta_i \in \mathbb{S}^1 \qquad [4]$$

which is easier for studying the collective properties of [3].

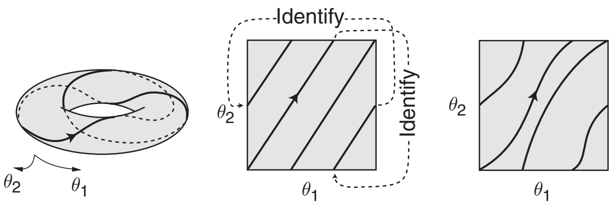



**Figure 1** A 2-torus and its representation on the square. (Modified from Hoppensteadt and Izhikevich 1997.)

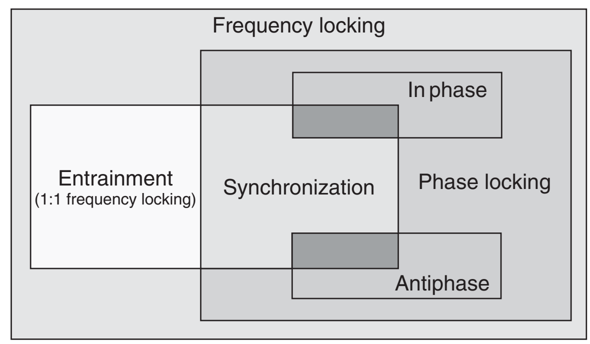



**Figure 2** Various degrees of locking of oscillators. (Modified from Izhikevich 2006.)

The oscillators are said to be frequency locked when [4] has a stable periodic orbit $\vartheta(t) = (\vartheta_1(t), \ldots, \vartheta_n(t))$ on the $n$-torus $\mathbb{T}^n$, as in Figure 1a. The "rotation vector" or "winding ratio" of the orbit is the set of integers $q_1 : q_2 : \cdots : q_n$ such that $\vartheta_1$ makes $q_1$ rotations while $\vartheta_2$ makes $q_2$ rotations, etc., as in the 2 : 3 frequency locking in Figure 1a. The oscillators are entrained when they are $1:1:\cdots:1$ frequency locked. The oscillators are phase locked when there is an $(n-1) \times n$ integer matrix $K$ having linearly independent rows such that $K\vartheta(t) = \text{const}$. For example, the two oscillators in Figure 1b are phase locked with $K = (2, 3)$, while those in Figure 1c are not. The oscillators are synchronized when they are entrained and phase locked. Synchronization is in-phase when $\vartheta_1(t) = \cdots = \vartheta_n(t)$ and out-of-phase otherwise. Two oscillators are said to be synchronized antiphase when $\vartheta_1(t) - \vartheta_2(t) = \pi$. Frequency locking without phase locking, as in Figure 1c, is called phase trapping. The relationship between all these definitions is depicted in Figure 2.

## Phase Resetting

An exponentially stable periodic orbit is a normally hyperbolic invariant manifold, hence its sufficiently small neighborhood, $U$, is invariantly foliated by

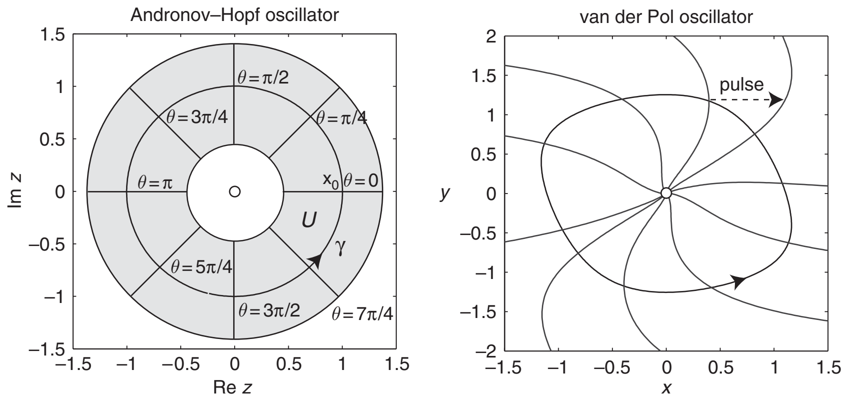



**Figure 3** Isochrons of Andronov–Hopf oscillator ($\dot{z} = (1 + i)z - z|z|^2, z \in C$) and van der Pol oscillator ($\dot{x} = x - x^3 - y, \dot{y} = x$).

stable submanifolds (Guckenheimer 1975) illustrated in **Figure 3**. The manifolds represent points having equal phases and, for this reason, they are called isochrons (from Greek "iso" meaning equal and "chronos" meaning time).

The geometry of isochrons determines how the oscillators react to perturbations. For example, the pulse in **Figure 3**, right, moves the trajectory from one isochron to another, thereby changing its phase. The magnitude of the phase shift depends on the amplitude and the exact timing of the stimulus relative to the phase of oscillation $\vartheta$. Stimulating the oscillator at different phases, one can measure the phase transition curve (Winfree 2001)

$$\vartheta_{\text{new}} = \text{PTC}(\vartheta_{\text{old}})$$

and the phase resetting curve

$$\text{PRC}(\vartheta) = \text{PTC}(\vartheta) - \vartheta$$
$$(\text{shift} = \text{new phase} - \text{old phase})$$

Positive (negative) values of the PRC correspond to phase advances (delays). PRCs are convenient when the phase shifts are small, so that they can be magnified and clearly seen, as in **Figure 4**. PTCs are convenient when the phase shifts are large and comparable with the period of oscillation.

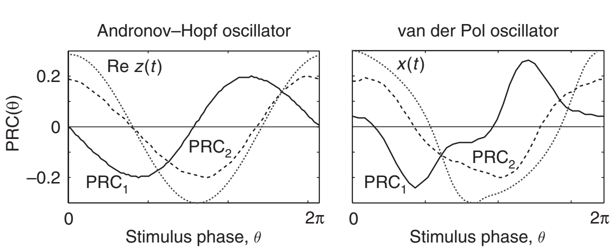



**Figure 4** Examples of phase response curves (PRCs) of the oscillators in **Figure 3**. $\text{PRC}_1(\vartheta)$ and $\text{PRC}_2(\vartheta)$ correspond to horizontal (along the first variable) and vertical (along the second variable) pulses with amplitudes 0.2. An example of oscillation is plotted as a dotted curve in each subplot (not to scale).

In **Figure 5** we depict phase portraits of the Andronov–Hopf oscillator receiving pulses of magnitude 0.5 (left) and 1.5 (right). Notice the drastic difference between the corresponding PRCs or PTCs. Winfree (2001) distinguishes two cases:

1. type 1 (weak) resetting results in continuous PRCs and PTCs with mean slope 1, and
2. type 0 (strong) resetting results in discontinuous PRCs and PTCs with mean slope 0.

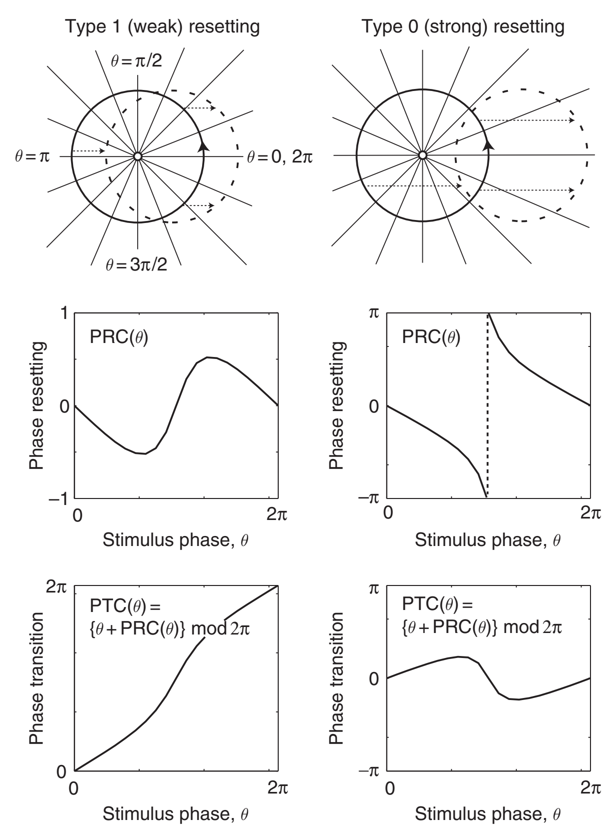



**Figure 5** Types of phase resetting of the Andronov–Hopf oscillator in **Figure 3**.

The discontinuity of type 0 PRC in Figure 5 is a topological property that cannot be removed by reallocating the initial point $x_0$ that corresponds to zero phase. The discontinuity stems from the fact that the shifted image of the limit cycle (dashed circle) goes beyond the central equilibrium at which the phase is not defined.

The stroboscopic mapping of $\mathbb{S}^1$ to itself, called Poincaré phase map,

$$\vartheta_{k+1} = \mathrm{PTC}(\vartheta_k) \qquad [5]$$

describes the response of an oscillator to a $T$-periodic pulse train. Here, $\vartheta_k$ denotes the phase of oscillation when the $k$th input pulse arrives. Its fixed points correspond to synchronized solutions, and its periodic orbits correspond to phase-locked states.

## Weak Coupling

Now consider dynamical systems of the form

$$\dot{x} = f(x) + \varepsilon s(t) \qquad [6]$$

describing periodic oscillators, $\dot{x} = f(x)$, forced by a weak time-depended input $\varepsilon s(t)$, for example, from other oscillators in a network. Let $\Theta(x)$ denote the phase of oscillation at point $x \in U$, so that the map $\Theta : U \to \mathbb{S}^1$ is constant along each isochron. This mapping transforms [6] into the phase model

$$\dot{\vartheta} = \Omega + \varepsilon Q(\vartheta) \cdot s(t)$$

with function $Q(\vartheta)$, illustrated in Figure 6, satisfying three equivalent conditions:

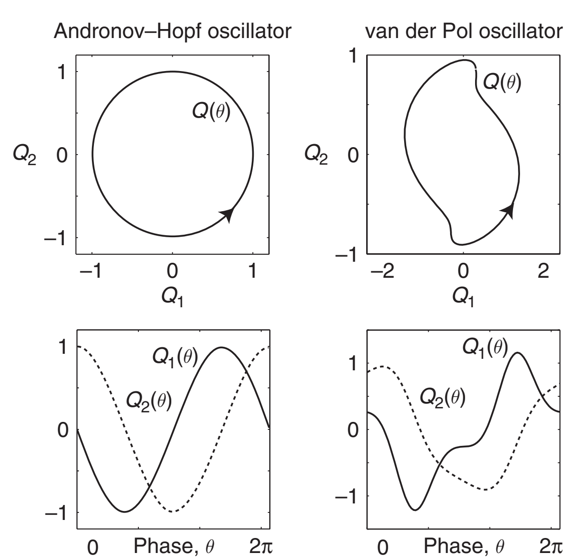



**Figure 6** Solutions $Q = (Q_1, Q_2)$ to the adjoint problem [7] for oscillators in **Figure 3**.

1. Winfree: $Q(\vartheta)$ is normalized PRC to infinitesimal pulsed perturbations;
2. Kuramoto: $Q(\vartheta) = \mathrm{grad}\ \Theta(x)$; and
3. Malkin: $Q$ is the solution to the adjoint problem

$$\dot{Q} = -\{Df(\gamma(t))\}^{\top} Q \qquad [7]$$

with the normalization $Q(t) \cdot f(\gamma(t)) = \Omega$ for any $t$.

The function $Q(\vartheta)$ can be found analytically in a few simple cases:

1. a nonlinear phase oscillator $\dot{x} = f(x)$ with $x \in \mathbb{S}^1$ and $f > 0$ has $Q(\vartheta) = \Omega / f(\gamma(\vartheta))$;
2. a system near saddle-node on invariant circle bifurcation has $Q(\vartheta)$ proportional to $1 - \cos\vartheta$; and
3. a system near supercritical Andronov–Hopf bifurcation has $Q(\vartheta)$ proportional to $\sin(\vartheta - \psi)$, where $\psi \in \mathbb{S}^1$ is a constant phase shift.

Other interesting cases, including homoclinic, relaxation, and bursting oscillators are considered by Izhikevich (2006).

Treating $s(t)$ in [6] as the input from the network, we can transform weakly coupled oscillators

$$\dot{x}_i = f_i(x_i) + \varepsilon \overbrace{\sum_{j=1}^{n} g_{ij}(x_i, x_j)}^{s_i(t)}, \quad x_i \in \mathbb{R}^m \qquad [8]$$

to the phase model

$$\dot{\vartheta}_i = \Omega_i + \varepsilon\, Q_i(\vartheta_i) \cdot \overbrace{\sum_{j=1}^{n} g_{ij}(x_i(\vartheta_i), x_j(\vartheta_j))}^{s_i(t)} \qquad [9]$$

having the form [4] with $h_i = Q_i \sum g_{ij}$, or the form

$$\dot{\vartheta}_i = \Omega_i + \varepsilon \sum_{j=1}^{n} h_{ij}(\vartheta_i, \vartheta_j)$$

where $h_{ij} = Q_i g_{ij}$. Introducing phase deviation variables $\vartheta_i = \Omega_i t + \varphi_i$, we transform this system into the form

$$\dot{\varphi}_i = \varepsilon \sum_{j=1}^{n} h_{ij}(\Omega_i t + \varphi_i,\ \Omega_j t + \varphi_j)$$

which can be averaged to

$$\dot{\varphi}_i = \varepsilon \sum_{j=1}^{n} H_{ij}(\varphi_i - \varphi_j) \qquad [10]$$

with the functions

$$H_{ij}(\chi) = \lim_{T\to\infty} \frac{1}{T} \int_0^T h_{ij}(\Omega_i t,\ \Omega_j t - \chi)\, dt \qquad [11]$$

describing the interaction between oscillators (Ermentrout and Kopell 1984). To summarize, we transformed weakly coupled system [8] into the phase model [10] with $H$ given by [11] and each $Q$ being the solution to the adjoint problem [7]. This constitutes the Malkin theorem for weakly coupled oscillators (Hoppensteadt and Izhikevich 1997, theorem 9.2).

Existence of one equilibrium of the phase model [10] implies the existence of the entire circular family of equilibria, since translation of all $\varphi_i$ by a constant phase shift does not change the phase differences $\varphi_i - \varphi_j$ and hence the form of [10]. This family corresponds to a limit cycle of [8], on which all oscillators have equal frequencies and constant phase shifts, that is, they are synchronized, possibly out of phase.

We say that two oscillators, $i$ and $j$, have resonant (or commensurable) frequencies when the ratio $\Omega_i/\Omega_j$ is a rational number, for example, it is $p/q$ for some integer $p$ and $q$. They are nonresonant when the ratio is an irrational number. In this case, the function $H_{ij}$ defined above is constant regardless of the details of the oscillatory dynamics or the details of the coupling, that is, dynamics of two coupled nonresonant oscillators is described by an uncoupled phase model. Apparently, such oscillators do not interact; that is, the phase of one of them cannot change the phase of the other one even on the long timescale of order $1/\varepsilon$.

## Synchronization

Consider [8] with $n=2$, describing two mutually coupled oscillators. Let us introduce "slow" time $\tau = \varepsilon t$ and rewrite the corresponding phase model [10] in the form

$$\varphi_1' = \omega_1 + H_{12}(\varphi_1 - \varphi_2)$$
$$\varphi_2' = \omega_2 + H_{21}(\varphi_2 - \varphi_1)$$

where $' = d/d\tau$ and $\omega_i = H_{ii}(0)$ is the frequency deviation from the natural oscillation, $i = 1, 2$. Let $\chi = \varphi_2 - \varphi_1$ denote the phase difference between the oscillators; then

$$\chi' = \omega + H(\chi) \qquad [12]$$

where

$$\omega = \omega_2 - \omega_1 \text{ and } H(\chi) = H_{21}(\chi) - H_{12}(-\chi)$$

is the frequency mismatch and the antisymmetric part of the coupling, respectively, illustrated in Figure 7, dashed curves. A stable equilibrium of [12] corresponds to a stable limit cycle of the phase model.

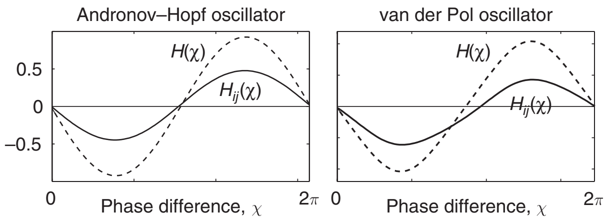



**Figure 7** Solid curves: functions $H_{ij}(\chi)$ defined by [11] corresponding to the gap–junction input $g(x_i, x_j) = (x_{j1} - x_{i1}, 0)$. Dashed curves: functions $H(\chi) = H_{ji}(\chi) - H_{ij}(-\chi)$.. Parameters are as in Figure 3.

All equilibria of [12] are solutions to $H(\chi) = -\omega$, and they are intersections of the horizontal line $-\omega$ with the graph of $H$. They are stable if the slope of the graph is negative at the intersection. If oscillators are identical, then $H(\chi)$ is an odd function (i.e., $H(-\chi) = -H(\chi)$), and $\chi = 0$ and $\chi = \pi$ are always equilibria, possibly unstable, corresponding to the in-phase and antiphase synchronized solutions. The in-phase synchronization of gap–junction coupled oscillators in Figure 7 is stable because the slope of $H$ (dashed curves) is negative at $\chi = 0$. The max and min values of the function $H$ determine the tolerance of the network to the frequency mismatch $\omega$, since there are no equilibria outside this range.

Now consider a network of $n > 2$ weakly coupled oscillators [8]. To determine the existence and stability of synchronized states in the network, we need to study equilibria of the corresponding phase model [10]. The vector $\phi = (\phi_1, \ldots, \phi_n)$ is an equilibrium of [10] when

$$0 = \omega_i + \sum_{j \neq 1}^{n} H_{ij}(\phi_i - \phi_j) \quad (\text{for all } i) \qquad [13]$$

It is stable when all eigenvalues of the linearization matrix (Jacobian) at $\phi$ have negative real parts, except one zero eigenvalue corresponding to the eigenvector along the circular family of equilibria ($\phi$ plus a phase shift is a solution of [13] too since the phase shifts $\phi_j - \phi_i$ are not affected).

In general, determining the stability of equilibria is a difficult problem. Ermentrout (1992) found a simple sufficient condition. If

1. $a_{ij} = H'_{ij}(\phi_i - \phi_j) \leq 0$, and
2. the directed graph defined by the matrix $a = (a_{ij})$ is connected, (i.e., each oscillator is influenced, possibly indirectly, by every other oscillator),

then the equilibrium $\phi$ is neutrally stable, and the corresponding limit cycle $x(t + \phi)$ of [8] is asymptotically stable.

Another sufficient condition was found by Hoppensteadt and Izhikevich (1997). If system [10] satisfies

1. $\omega_1 = \cdots = \omega_n = \omega$ (identical frequencies)
2. $H_{ij}(-\chi) = -H_{ji}(\chi)$ (pairwise odd coupling)

for all $i$ and $j$, then the network dynamics converge to a limit cycle. On the cycle, all oscillators have equal frequencies $1 + \varepsilon\omega$ and constant phase deviations.

The proof follows from the observation that [10] is a gradient system in the rotating coordinates $\varphi = \omega\tau + \phi$ with the energy function

$$E(\phi) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} R_{ij}(\phi_i - \phi_j)$$

where

$$R_{ij}(\chi) = \int_0^{\chi} H_{ij}(s)\,\mathrm{d}s$$

One can check that $\mathrm{d}E(\phi)/\mathrm{d}\tau = -\sum(\phi_i')^2 \le 0$ along the trajectories of [12] with equality only at equilibria.

## Mean-Field Approximations

Let us represent the phase model [10] in the form

$$\varphi_i' = \omega_i + \sum_{j \ne i}^{n} H_{ij}(\varphi_i - \varphi_j)$$

where $' = \mathrm{d}/\mathrm{d}\tau, \tau = \varepsilon t$ is the slow time, and $\omega_i = H_{ii}(0)$ are random frequency deviations. Collective dynamics of this system can be analyzed in the limit $n \to \infty$. We illustrate the theory using the special case, $H(\chi) = -\sin\chi$, known as the Kuramoto (1984) model:

$$\varphi_i' = \omega_i + \frac{K}{n}\sum_{j=1}^{n}\sin(\varphi_j - \varphi_i), \quad \varphi_i \in [0, 2\pi] \qquad [14]$$

where $K > 0$ is the coupling strength and the factor $1/n$ ensures that the model behaves well as $n \to \infty$. The complex-valued sum of all phases,

$$re^{\mathrm{i}\psi} = \frac{1}{n}\sum_{j=1}^{n} e^{\mathrm{i}\varphi_j}$$

(Kuramoto synchronization index) [15]

describes the degree of synchronization in the network. Apparently, the in-phase synchronized state $\varphi_1 = \cdots = \varphi_n$ corresponds to $r = 1$ with $\psi$ being the population phase. In contrast, the incoherent state with all $\varphi_i$ having different values randomly distributed on the unit circle, corresponds to $r \approx 0$. Intermediate values of $r$ correspond to a partially synchronized or coherent state, depicted in Figure 8. Some phases are synchronized forming a cluster, while others roam around the circle.

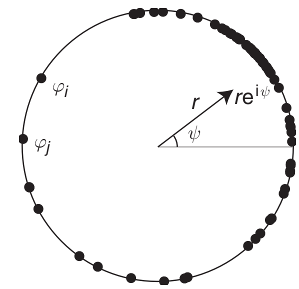



**Figure 8** Kuramoto synchronization index [15] describes the degree of coherence in the network [14].

Multiplying both sides of [15] by $e^{-\mathrm{i}\varphi_i}$ and considering only the imaginary parts, we can rewrite [14] in the equivalent form

$$\varphi_i' = \omega_i + Kr\sin(\psi - \varphi_i)$$

that emphasizes the mean-filed character of interactions between the oscillators: they all are pulled into the synchronized cluster ($\varphi_i \to \psi$) with the effective strength proportional to the cluster size $r$. This pull is offset by the random frequency deviations $\omega_i$ that pull away from the cluster.

Let us assume that $\omega_i$'s are distributed randomly around 0 with a symmetrical probability density function $g(\omega)$, for example, Gaussian. Kuramoto has shown that in the limit $n \to \infty$, the cluster size $r$ obeys the self-consistency equation

$$r = rK\int_{-\pi/2}^{+\pi/2} g(Kr\sin\varphi)\cos^2\varphi\,\mathrm{d}\varphi \qquad [16]$$

Notice that $r = 0$, corresponding to the incoherent state, is always a solution of this equation. When the coupling strength $K$ is greater than a certain critical value,

$$K_c = \frac{2}{\pi g(0)}$$

an additional, nontrivial solution $r > 0$ appears, which corresponds to a partially synchronized state. Expanding $g$ in a Taylor series, one gets the scaling $r = \sqrt{16(K - K_c)/(-g''(0)\pi K_c^4)}$. Thus, the stronger the coupling $K$ relative to the random distribution of frequencies, the more oscillators synchronize into a coherent cluster. The issue of stability of incoherent and partially synchronized states is discussed by Strogatz (2000). Other generalizations of the Kuramoto model are reviewed by Acebron *et al.* (2005). An extended version of this article with the

emphasis on computational neuroscience can be found in the recent book by Izhikevich (2006).

*See also:* Bifurcations of Periodic Orbits; Dynamical Systems and Thermodynamics; Hamiltonian Systems: Stability and Instability Theory; Singularity and Bifurcation Theory; Stability Theory and KAM; Synchronization of Chaos.

## Further Reading


Acebron JA, Bonilla LL, Vincente CJP, Ritort F, and Spigler R (2005) The Kuramoto model: a simple paradigm for synchronization phenomena. *Reviews of Modern Physics* 77: 137–186.

Ermentrout GB (1992) Stable periodic solutions to discrete and continuum arrays of weakly coupled nonlinear oscillators. *SIAM Journal on Applied Mathematics* 52: 1665–1687.

Ermentrout GB and Kopell N (1984) Frequency plateaus in a chain of weakly coupled oscillators, I. *SIAM Journal on Applied Mathematics* 15: 215–237.

Glass L and MacKey MC (1988) *From Clocks to Chaos.* Princeton: Princeton University Press

Guckenheimer J (1975) Isochrons and phaseless sets. *Journal of Mathematical Biology* 1: 259–273.

Hoppensteadt FC and Izhikevich EM (1997) *Weakly Connected Neural Networks.* New York: Springer.

Izhikevich EM (1999) Weakly connected quasiperiodic oscillators, FM interactions, and multiplexing in the brain. *SIAM Journal on Applied Mathematics* 59: 2193–2223.

Izhikevich EM (2006) *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting.* Cambridge, MA: The MIT Press.

Kuramoto Y (1984) *Chemical Oscillations, Waves, and Turbulence.* New York: Springer.

Pikovsky A, Rosenblum M, and Kurths J (2001) *Synchronization: A Universal Concept in Nonlinear Science.* Cambridge: Cambridge University Press.

Strogatz SH (2000) From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D* 143: 1–20.

Winfree A (2001) *The Geometry of Biological Time*, 2nd edn. New York: Springer.


# Wheeler–De Witt Theory

**J Maharana**, Institute of Physics, Bhubaneswar, India



## Introduction

It is recognized that one of the outstanding problems in modern physics is to formulate the quantum theory of gravity, synthesizing the principles of quantum mechanics and general theory of relativity. The fundamental units for measuring time, length, and energy, known as Planck time, Planck length, and Planck energy, respectively, are defined to be $t_{\mathrm{Pl}} = (\hbar G/c^5)^{1/2} = 5.39 \times 10^{-44}\,\mathrm{s}$, $l_{\mathrm{Pl}} = (\hbar G/c^3)^{1/2} = 1.61 \times 10^{-33}\,\mathrm{cm}$, and $m_{\mathrm{Pl}} = (\hbar c/G)^{1/2} = 2.17 \times 10^{-5}\,\mathrm{g}$, in terms of the Newton's constant, $G$, velocity of light, $c$, and $\hbar = h/2\pi, h$ being the Planck's constant. We may conclude, on dimensional arguments, that quantum gravity effects will play an important role when we consider physical phenomena in the vicinity of these scales. Therefore, when we probe very short distances, consider collisions at Planckian energies, and envisage evolution of the universe in the Planck era, the quantum gravity will come into play in a predominant manner. The purpose of this article is to present an overview of an approach to quantize Einstein's theory of gravity, pioneered by Wheeler and De Witt almost four decades ago. We proceed to recapitulate various prescriptions for quantizing gravitation and then discuss simple derivation of the Wheeler–De Witt (WDW) equation in general relativity and some of its applications in the study of quantum cosmology. There are, broadly speaking, three different approaches to quantize gravity.

The general theory of relativity has been tested to great degree of accuracy in the classical regime. The geometrical description of spacetime plays a cardinal role in Einstein's theory. Therefore, the general relativists emphasize the geometrical attributes of the theory and the central role played by the spacetime structure in their formulation of quantum theory. It is natural to adopt a background-independent approach. In contrast, the path followed by quantum field theorists, where the prescription is valid in the weak-field approximation, the theory is quantized in a given background, usually the Minkowskian space. It is argued by the proponents of the geometric approach, that the background metric should emerge from the theory in a self-consistent manner rather than being introduced by hand when we quantize the theory. One of the earliest attempts to quantize gravity was to follow the route of canonical method. The canonical quantization approach has many advantages. One of the important features is that it is quite similar to the prescriptions adopted in quantum field theory where one uses notion of operators, commutation relations, etc. Moreover, the subtleties encountered in quantizing gravity are transparent. Therefore, the canonical procedure is preferred over the path-integral formulation, although the latter has its own advantages too. Another positive aspect of the canonical approach is that the requirement of background-independent

formulation could be maintained to some extent. Thus, there is room for exploring some of the nonperturbative attributes of the theory. The relativists favor canonical formulation, since some of the geometrical features of general theory of relativity could be incorporated here and be explored to see how far the quantum theory captures such properties of the classical theory. As we shall discuss in sequel, some of the interesting issues of quantum cosmology are addressed in this approach. However, there are limitations and short comings in this formulation and we refer the reader to the text books and review articles for further reading and critical assessments of canonical approach to quantize gravity.

The second approach is primarily the endeavor of physicists who have devoted their research to quantum field theory. Feynman's seminal work on quantization of gravity from this perspective has profoundly influenced the subsequent developments. The quantization of gravity is carried out in the weak-field approximation such that the graviton is identified as the fluctuation over the Minkowski background metric. It is a massless spin-2 field as one concludes from the properties of low-energy gravitational interaction in the classical limit. Furthermore, the gauge invariance associated with a spin-2 massless field gets intimately related with invariance of Einstein's theory under general coordinate transformation. In this setup, the field-theoretic techniques could be employed to quantize theory and to consider perturbative expansions for the scattering amplitudes. It is realized that low-energy amplitudes computed from the massless spin-2 theory match with those derived from the Einstein–Hilbert action in the weak-field approximation. Furthermore, the theory is not perturbatively renormalizable since the coupling constant carries dimension. One of the most important outcomes of the investigations from this perspective is the discovery, due to Feynman, that the introduction of ghost fields is necessary in order to maintain unitarity of the $S$-matrix when one goes beyond the tree level. As is well known, this work has profoundly influenced frontiers of research in physics leading to quantization of Yang–Mills theory which, in turn, paved way for electroweak theory and the QCD. It is worthwhile to mention in passing that the quantum phenomena associated with gravity in the nonperturbative regime cannot be addressed in this framework.

In recent years, superstring theory has been at the center stage in order to provide a unified theory of fundamental interactions. It is postulated that all elementary constituents of matter and the carriers of the interactions such as gauge bosons and graviton are excitations of one-dimensional extended objects: the strings. The superstring theories are perturbatively consistent in critical ten dimensions. The closed-superstring spectrum contains a spin-2 massless state which is identified to be the graviton. It is well known that perturbative computation of processes involving graviton turn out to be finite. Moreover, the Einstein–Hilbert term appears naturally when one derives the string effective action. Therefore, it is expected that string theory will be able to provide answers to questions related to quantum gravity. Indeed, the theory has met with success in resolving some important issues. We note that cosmological scenario has been discussed in the string theory framework and the WDW equation has played an important role in study of quantum string cosmology. We shall comment on this aspect towards the end of this article.

## The Canonical Structure of Einstein Gravity

The Einstein–Hilbert action is

$$S = \frac{1}{16\pi G}\int_M \sqrt{-g}\,\mathrm{d}^4x(R - 2\Lambda) \qquad [1]$$

where $R$ is the Ricci scalar derived from the metric, $g_{\mu\nu}$, and $\Lambda$ is the cosmological constant. The field equations are derived from the action by the standard variational technique. Note that $R$ involves second derivative of the metric. If we have compact manifolds with boundary $\partial M$ such that variations of the metric vanish on the boundary and the normal derivatives do not, it is necessary to add a surface term to this action. The exact form of this term will be discussed later. The Einstein's theory of gravitation is manifestly covariant. The associated action [1] is invariant under general coordinate transformations: under $x^\mu \to x'^\mu(x)$,

$$g'^{\mu\nu}(x') = g^{\rho\lambda}(x)\frac{\partial x'^\mu}{\partial x^\rho}\frac{\partial x'^\nu}{\partial x^\lambda} \qquad [2]$$

Therefore, we expect that the theory will be endowed with constraints expressed in terms of the canonical variables. One can implement general coordinate transformations so that there are only two pairs of canonical phase-space variables on a spacelike hypersurface. In other words, from physical considerations, graviton has only two polarizations whereas the metric has ten components. Therefore, the two physical degrees of freedom can be obtained using the freedom of choosing the "gauge" transformations in this context. It is desirable to identify the constraints and analyze their structure, most appropriately in Dirac's

formalism, and to quantize the theory canonically as the next step. This is the path we intend to follow in order to arrive at the WDW equation.

### The Classical Constraints

The Hamiltonian approach is most appropriate to employ the constraint formalism due to Dirac. We recall that the Lagrangian formulation is manifestly covariant as is reflected in the field equations; whereas the spacetime covariance is lost in the passage to the Hamiltonian approach. Furthermore, the spatial components of the metric are the dynamical degrees of freedom. We adopt the formalism introduced by Arnowitt, Deser, and Misner (ADM) for the so-called $3+1$ split of the hyperbolic Riemannian spacetime metric, $g_{\mu\nu}$. One introduces the lapse function, $N^{\perp}$, and the shift function, $N^i$. We suppress the factors of $1/16\pi G$, etc., for the time being for the general discussions and shall reintroduce them later. The family of spacelike hypersurfaces, $\Sigma_t$, are constructed, with metric $h_{ij}$ induced on it. Here $t$ is a timelike parameter, parametrize $\Sigma_t$. The distance between points on two neighboring hypersurface, $\Sigma_t$ and $\Sigma_{t+dt}$, with coordinates $(t, x^i)$ and $(t+\mathrm{d}t, x^i+\mathrm{d}x^i)$, respectively, is given by

$$\begin{aligned} \mathrm{d}s^2 &= -(N^{\perp})^2\,\mathrm{d}t^2 + h_{ij}(N^i\,\mathrm{d}t+\mathrm{d}x^i)(N^j\,\mathrm{d}t+\mathrm{d}x^j) \\ &= g_{\mu\nu}\,\mathrm{d}x^{\mu}\,\mathrm{d}x^{\nu} \end{aligned} \qquad [3]$$

The indices of tensors defined on $\Sigma_t$ are raised and lowered by $h_{ij}$ and its inverse $h^{ij}$. The relations between the components of $g_{\mu\nu}$ and $N^{\perp}, N^i, h_{ij}$ can be obtained easily,

$$g_{00} = h_{ij}N^iN^j - (N^{\perp})^2, \qquad g_{0i} = h_{ij}N^j \qquad [4]$$

The above relations can be inverted to give

$$N^i = h^{ij}g_{0i} \qquad N^{\perp} = \frac{1}{\sqrt{-g^{00}}} \qquad [5]$$

The relations between spatial components, $g_{ij}$, of $g_{\mu\nu}$ and $h_{ij}$ and some other useful relations are listed below for later conveniences:

$$\begin{aligned} g^{ij} &= h^{ij} - \frac{N^iN^j}{(N^{\perp})^2} \\ \sqrt{-g} &= N^{\perp}\sqrt{h} \\ g^{0i} &= \frac{N^i}{(N^{\perp})^2} \end{aligned} \qquad [6]$$

Note that $(N^{\perp}, N^i)$ are introduced to specify the deformation of the hypersurface and therefore, the evolution equations through the Hamiltonian will not determine them; they are arbitrary functions. Consequently, [4] implies that $g_{00}$ and $g_{0i}$ will enter the Hamiltonian as arbitrary functions. As alluded to above, $h_{ij}$ and their conjugate momenta $\pi^{ij}$ are the dynamical degrees of freedom. We may choose $(N^{\perp}, N^i) = N^{\mu}$ and $h_{ij}$ as independent variables rather than $(g_{00}, g_{0i}) = g_{0\mu}$ and $h_{ij}$ for convenience and go back to the other set of variables through [4] and [5] if we desire. Let $\pi_{\mu}$ be canonically conjugate momenta to $N^{\mu}$, then it is obvious that a Lagrangian multiplier, $\chi^{\mu}$, is necessary so that $\pi.\chi$ term has to be supplemented to the Hamiltonian due to the arbitrariness of $N^{\mu}$. We remind the reader that in electrodynamics an analogous situation arises while analyzing its canonical structure – local gauge symmetry plays a crucial role there. It is obvious that the generic form of the Hamiltonian is (we shall introduce $1/16\pi G$, etc., later)

$$\mathbf{H} = \int \mathrm{d}^3x \left(N^{\perp}\mathcal{H}_{\perp}[h_{ij}, \pi^{ij}] + N^i\mathcal{H}_i[h_{ij}, \pi^{ij}] + \chi.\pi\right) \qquad [7]$$

From the perspective of constraint analysis, it is natural that $\pi^{\mu} \approx 0$ appears as a first-class constraint as they are multiplied by arbitrary functions. Moreover, this constraint must hold good under the deformation of the surface which implies $\{\pi^{\mu}, H\}_{\mathrm{PB}}$ must vanish weakly leading to $\mathcal{H}_{\mu} \approx 0$. As a consistency requirement, these must be first-class constraints if $N^{\mu}$ are to be arbitrary functions. We identify that $\pi^{\mu} \approx 0$ and $\mathcal{H}_{\mu} \approx 0$ are the primary and secondary constraints, respectively. Thus far, we have discussed the case for pure gravity; the presence of matter fields in the full action modifies the treatment appropriately.

Let us analyze the structure of the constraints for the Einstein–Hilbert action [1]. For a compact manifold with boundary $\partial M$, we have to add the surface term which takes the form:

$$\frac{1}{8\pi G}\int_{\partial M} \mathrm{d}^3x\sqrt{h}K$$

Here $K$ stands for the trace of the extrinsic curvature of the boundary 3-surface and $h = \det h_{ij}$; note that $h_{ij}$ is the induced metric on the 3-surface. If we include matter fields, the corresponding action is to be taken into account. Once we make the $3+1$ split of the metric, the action assumes the following form:

$$\begin{aligned} S = \frac{1}{16\pi G}&\int \mathrm{d}^3x\,\mathrm{d}tN^{\perp}\sqrt{h} \\ &\times \left(K_{ij}K^{ij} - K^2 + {}^3R - 2\Lambda\right) \end{aligned} \qquad [8]$$

where

$$K_{ij} = \frac{1}{N^{\perp}}\left(-\frac{\partial h_{ij}}{\partial t} + D_iN_j + D_jN_i\right) \qquad [9]$$

Here $D_i N_j$ represents covariant derivative of $N_j$ with the connections computed from $h_{ij}$ and ${}^3R$ is curvature of the 3-surface. The canonical momenta are

$$\pi^{ij} = \frac{\sqrt{h}}{16\pi G}\left(K^{ij} - h^{ij}K^l_l\right) \qquad [10]$$

and we can invert this relation to get

$$K^{ij} = -\frac{1}{16\pi G\sqrt{h}}\left(\pi^{ij} - \frac{1}{2}h^{ij}\pi^l_l\right)$$

The Hamiltonian form of action is given by

$$S_H = \int \mathrm{d}^3x\,\mathrm{d}t\left(\dot{h}_{ij}\pi^{ij} - N_\perp\mathcal{H}^\perp - N^i\mathcal{H}_i\right) \qquad [11]$$

Notice that [8] does not involve time derivatives of $N^\perp$ and $N^i$, their corresponding canonical momenta vanish.

$$\pi_\perp \approx 0, \qquad \pi_i \approx 0 \qquad [12]$$

as expected from our earlier discussions about the role of $N^\mu$. A straightforward constraint analysis leads to the pair of constraints

$$\mathcal{H}_i = -2D_j\pi^j_i \approx 0 \qquad [13]$$

$$\mathcal{H}_\perp = \frac{16\pi G}{\sqrt{h}}\left(h_{ij}h_{kl} - \frac{1}{2}h_{ik}h_{jl}\right)\pi^{ik}\pi^{jl} - \frac{\sqrt{h}}{16\pi G}{}^3R \approx 0 \qquad [14]$$

We mention in passing that the above constraint equations get modified in the presence of matter fields in the theory. This is relevant. The WDW equation plays an important role in quantum cosmology to describe the evolution of the universe in early epochs and the equation is studied in the presence of a generic matter content, that is, a scalar field with potential. The constraint equations [13] and [14] modify to

$$\mathcal{H}^{\mathrm{T}}_i = \mathcal{H}_i + \mathcal{H}^{\mathrm{matter}}_i \approx 0 \qquad [15]$$

$$\mathcal{H}^{\mathrm{T}}_\perp = \mathcal{H}_\perp + \mathcal{H}^{\mathrm{matter}} \approx 0 \qquad [16]$$

### The Algebra of Constraints

In order to compute the classical Poisson bracket algebra of the constraints [13] and [14], we use the canonical Poisson bracket relations for the phase-space variables on $\Sigma_t$:

$$\{h_{ij}(\boldsymbol{x}), h_{kl}(\boldsymbol{x}')\} = 0 \qquad [17]$$

$$\{\pi^{ij}(\boldsymbol{x}), \pi^{kl}(\boldsymbol{x}')\} = 0 \qquad [18]$$

$$\{h_{ij}(\boldsymbol{x}), \pi^{kl}(\boldsymbol{x}')\} = \delta^k_{(i}\delta^l_{j)}\delta(\boldsymbol{x}, \boldsymbol{x}') \qquad [19]$$

Thus, Poisson brackets among the constraints [13] and [14] are

$$\{\mathcal{H}_i(\boldsymbol{x}), \mathcal{H}_j(\boldsymbol{x}')\} = -\mathcal{H}_j(\boldsymbol{x})\partial^{x'}_i\delta(\boldsymbol{x}, \boldsymbol{x}') + \mathcal{H}_i\partial^x_j\delta(\boldsymbol{x}, \boldsymbol{x}') \qquad [20]$$

$$\{\mathcal{H}_i(\boldsymbol{x}), \mathcal{H}_\perp(\boldsymbol{x}')\} = \mathcal{H}_\perp(\boldsymbol{x})\partial^x_i\delta(\boldsymbol{x}, \boldsymbol{x}') \qquad [21]$$

$$\{\mathcal{H}_\perp(\boldsymbol{x}), \mathcal{H}_\perp(\boldsymbol{x}')\} = h^{ij}(\boldsymbol{x})\mathcal{H}_i(\boldsymbol{x})\partial^{x'}_j\delta(\boldsymbol{x}, \boldsymbol{x}') - h^{ij}(\boldsymbol{x}')\mathcal{H}_i(\boldsymbol{x}')\partial^x_j\delta(\boldsymbol{x}, \boldsymbol{x}') \qquad [22]$$

When we resort to canonical quantization, the starting point is the Hamiltonian action in the first-order formalism, where the canonical variables are subjected to the constraints [13] and [14] in terms of $\mathcal{H}_\perp$ and $\mathcal{H}_i$ satisfying the algebra given by [20]–[22]. One encounters a number of important issues while proceeding to canonically quantize the theory. We shall mention only a few of them in what follows. It is important to address issues related to the role of the constraints in the quantized theory and how to deal with the Lagrange multipliers $N^\perp$ and $N^i$. A simple proposal is to solve the constraints at the classical level and identify the physical degrees of freedom and quantize the theory subsequently. There are four constraints (first class), $\mathcal{H}_\perp, \mathcal{H}_i$, therefore, out of the 12 phase-space variables, $(h_{ij}, \pi^{ij})$, only eight are independent. We need to supply four gauge conditions in order to render the theory (classically) solvable. Thus, we are left with four physical degrees of freedom in the Hamiltonian phase space and we can quantize them. The implementation of this idea is easier said than done. One obstacle is that the constraints cannot be solved in a closed form in this formalism. If we fix a gauge and quantize the theory, we obviously break the gauge invariance. It is essential to show, subsequently, that all physically observable quantities are independent of the gauge choice. Another criticism of this formalism is that we already get rid of some of the components of the metric. Therefore, the spirit of the general theory of relativity, which is based on the geometrical structure of spacetime, is somewhat diluted. There are other suggestions where $h_{ij}$ and their conjugate momenta are elevated to quantum status before supplying the gauge conditions. The issues of gauge fixing and dealing with the constraints are addressed at the quantum level. We replace the canonical Poisson bracket

algebra by the canonical commutators and proceed further. The momentum operator assumes the form

$$\hat{\pi}^{ij} = -\mathrm{i}\hbar \frac{\delta}{\delta h_{ij}}$$

and the wave functional depends on $h_{ij}$ that is, $\Psi[h]$. There are many technical problems related to the properties of the states and we shall not deal with them due to limitations of space. It is essential to discuss the role of the constraints in the quantum theory. We demand that the quantum constraints annihilate the physical states (recall the Gauss law constraint in gauge theories). However, the issue of operator ordering is to be dealt with which in turn is connected with the Hermiticity properties of the quantum constraints. The Hamiltonian constraint $\mathcal{H}_\perp \approx 0$ (henceforth denoted as $\mathcal{H}$ and defined as the Hamiltonian) is a product of the metric $\hat{h}_{ij}$ and $\boldsymbol{\pi}^{ij}$. There is certain ambiguity in defining the constraint. Therefore, one has to choose a convention. The condition that the Hamiltonian, $\hat{\mathcal{H}}^{\mathrm{T}}$, consisting of gravitational and matter components, annihilates the state is expressed as

$$\hat{\mathcal{H}}^{\mathrm{T}} \Psi = 0 \qquad [23]$$

When we adopt coordinate representation for $\boldsymbol{\pi}^{ij}$, the above equation takes the form

$$\left[-16\pi G\, G_{ijkl} \frac{\delta}{\delta h_{ij}} \frac{\delta}{\delta h_{kl}} - \frac{\sqrt{h}}{16\pi G}\left({}^3R - 2\Lambda\right) + \mathcal{H}^{\mathrm{matter}}\right] \Psi[h, \phi] = 0 \qquad [24]$$

This is the celebrated WDW equation. Here we have considered a simple case where matter Hamiltonian density generically contains a single scalar field, $\phi$, and therefore $\Psi$ is functional of 3-metric on $\Sigma_t$ and $\phi$. $G_{ijkl}$ is the De Witt metric in the superspace:

$$G_{ijkl} = \frac{1}{\sqrt{h}}\left(h_{ik} hjl + h_{il} h_{jk} - h_{ij} h_{kl}\right) \qquad [25]$$

**Remarks** The space of all 3-metrics and the scalar field $(h_{ij}, \phi)$, on $\Sigma_t$, for the description of classical evolutions is called the superspace (no connection with the superspace of supersymmetry). Thus, $\Psi[h_{ij}, \phi]$ is a functional on superspace. Furthermore, $\Psi$ carries no explicit dependence on $t$. This is a consequence of the fact that "time" plays the role of a parameter in the general theory of relativity, thus the dynamical variables $h_{ij}$ and $\phi$ already provide the evolutionary processes although $t$ does not make its appearance. As mentioned earlier, we always discuss the case when $\Sigma_t$ is compact. Another point to note is that the quantum momentum constraint, $\hat{\mathcal{H}}_i$, as an operator annihilates the wave function which is a statement of the quantum-mechanical invariance of the theory under three-dimensional diffeomorphisms. However, the WDW equation conveys invariance of the theory under reparametrization, although careful analysis is necessary to prove this point. Now we proceed to discuss the solutions of the WDW equation.

## WDW Equation and the Solutions

It is recognized that the WDW equation [24] is a second-order hyperbolic functional differential equation and naturally it has enormous number of solutions. Therefore, if we want the WDW equation to have any predictive power, it is necessary to introduce boundary conditions. One of the possible choice is to specify the wave function on the boundary of the superspace. Indeed, the central issue of quantum cosmology is about the choice of various boundary conditions which has been an important topic of debates. This point will be briefly discussed later. Notice that the boundary condition has to be introduced keeping in mind how the universe is expected to behave as it evolves. There is a proposition that the boundary condition for the quantum evolution of the universe be given the status of a physical law. Therefore, the role of the wave functional, $\Psi[h_{ij}(\boldsymbol{x}), \phi(\boldsymbol{x}), B]$, its evolution, and interpretation are central to the development of quantum cosmology. Thus, $\Psi$ represents the amplitude for the universe to have $h_{ij}(\boldsymbol{x})$ on the 3-surface, $B$, and matter field $\phi(\boldsymbol{x})$. It is argued that path-integral formalism should be adopted as an alternative to the canonical prescription to solve for the wave function, rather the transition amplitude, satisfying the WDW equation. Here the first step is to define the Euclidean version of the gravitational action keeping in mind the subtleties. As is well known, we deal with propagator (or transition amplitude) in the path-integral approach where the functional integral is carried out over a set of 4-metrics and matter fields with Euclidean action inside the integral acting as the weight factor. We recall that while formulating quantum mechanics in the path-integral approach, we sum over all possible paths in the functional integral. However, in the semiclassical approximation, the amplitude is dominated by the action corresponding to the classical path and we approximate the wave function as $\psi \sim \mathrm{e}^{(\mathrm{i}/\hbar)S_{\mathrm{cl}}}$ and it gets modified appropriately in the Euclidean formulation. In this background, we briefly discuss how the wave function of the universe is obtained in the path-integral formalism.

According to the proposal of Hartle and Hawking, one adopts path-integral formalism for the Euclidean action where the functional integral is not only carried out over the 4-metric, $g_{\mu\nu}$, and the scalar field $\phi$, but also one takes sum over the class of manifolds, $M$. Note that $B$ is a part of the boundary of this set of manifold. If $\bar{h}_{ij}$ and $\bar{\phi}$ are the induced metric and the configuration of the scalar field, $\phi$, on the boundary, $B$, then the propagator (henceforth we just call it the wave function) $\Psi[\bar{h}_{ij}, \bar{\phi}, B]$ can be given a functional-integral representation. Indeed, obtaining the most general form of the path integral, summing over the 4-manifolds, is quite a formidable task. On the other hand, if one chooses a class of 4-manifolds which can be decomposed as a product (foliation) $\boldsymbol{R} \times B$, the wave function is expressed as

$$\Psi[\bar{h}, \bar{\phi}, B] = \int \mathcal{D}N^\mu \int \mathcal{D}h_{ij}\mathcal{D}\phi\, f(N^\mu)\Delta_{\mathrm{FP}}\mathrm{e}^{-S_{\mathrm{E}}[g_{\mu\nu},\phi]} \qquad [26]$$

We have introduced the gauge-fixing condition as $f(N^\mu)$, which is usually taken to be $\dot{N}^\mu = l^\mu$ and then the corresponding Faddeev–Popov determinant, $\Delta_{\mathrm{FP}}$, has to be inserted into the path-integral measure. We recall from our earlier discussions that $N^\mu$ has to be unrestricted on the boundary, $B$, since they have no dynamical role when we express the action in terms of the variables defined on the 3-surface. As noted in the previous discussion, explicit time dependence does not appear after the $3+1$ split and $(h_{ij}(\boldsymbol{x}), \phi(\boldsymbol{x}))$ have no dependence on $t$. Therefore, we introduce a parameter to designate the paths over which the functional integral is to be taken. Recall that in the quantum-mechanical case, the paths are parametrized as $q_i(t)$ for the coordinates. However, when we resort to a parametrization of the variables for the case at hand, certain conditions must be fulfilled. We are permitted to integrate over $h_{ij}$ and $\phi$ over only those paths, while parametrizing them as $(h_{ij}(\boldsymbol{x}, \tau), \phi(\boldsymbol{x}, \tau))$, so that they match the arguments of the wave function on the boundary $B$. Therefore, we may define the metric and the scalar field configuration so that at $\tau = 1$ they assume their functional values on the boundary: in other words, $\bar{h}_{ij}(\boldsymbol{x}) = h_{ij}(\boldsymbol{x}, \tau = 1)$ and $\bar{\phi}(\boldsymbol{x}) = \phi(\boldsymbol{x}, \tau = 1)$. It is worthwhile to go back to the quantum-mechanical analogy once more. When we compute amplitudes/propagators in quantum mechanics, the functional integral is defined for the amplitude of going from a configuration $q_i$ to $q_f$ while summing over all possible paths originating from one endpoint $q_i$ and ending at the final point $q_f$. On this occasion, we have imposed the constraint on the final endpoint belonging to the boundary $B$. Thus, in order to determine the wave function of the universe, we are required to specify the initial configurations of $h_{ij}$ and $\phi$ at $\tau = 0$. We shall not enter into important issues related with the properties of the Euclidean action, the problems associated with the choice of contours of the path integrals, and related topics. The reader will find detailed discussions in the lectures and monographs referred in the "Further reading" section.

It is important to re-emphasize that boundary conditions are to be introduced while solving the WDW equation. It was argued by De Witt that the wave function will be determined uniquely from the mathematical consistency of the theory and that hope has not been realized. Whether one attempts to solve the functional differential WDW equation or obtain the wave function in the path-integral formalism, the issue of boundary condition is unavoidable. There are mainly three different kinds of boundary conditions in quantum cosmology: Hartle–Hawking (HH) no-boundary proposal, Vilenkin's tunneling mechanism, and Linde's boundary condition. We shall briefly discuss the first two proposals. Instead of stating the boundary conditions in full generality, we shall envisage quantum cosmology in a minisuperspace and provide illustrative examples to compare the main features of HH and Vilenkin solutions to the WDW equation.

It is realized that the discussion and solutions of quantum cosmology in the superspace is rather difficult, since we deal with functional differential equations and the configuration space is infinite dimensional. Therefore, it is worthwhile to consider a system, as a simple model, which has finite degrees of freedom. Thus, we assume that the metric and matter fields depend only on cosmic time to begin with. There is a physical motivation behind this assumption, since the present classical state of the universe is described by the Friendmann–Robertson–Walker (FRW) metric corresponding to an isotropic and homogeneous universe. Notice that the classical evolution equation resembles that of the motion of a particle. The quantum evolution equations are now given by differential equations of quantum mechanics rather than functional differential equations. Similarly, the path-integral formulation becomes analogous to the quantum-mechanical frame work. Of course, adopting such a simplified approach deprives us from describing some of the important aspects of quantum gravity. However, within this framework, several essential features can be exhibited and deep insight might be gained into the physics of the very early universe. The first step in getting the minisuperspace metric is to assume that the lapse is homogeneous, that is, $N^\perp = N^\perp(t)$

and the shift is set to zero, $N^i = 0$. Thus, the metric takes the form

$$ds^2 = -(N^{\perp}(t))^2 dt^2 + h_{ij}(\boldsymbol{x}, t) dx^i dx^j \quad [27]$$

The relevant choice of 3-metric for FRW isotropic and homogeneous universe is

$$h_{ij}(\boldsymbol{x}, t) dx^i dx^j = a(t)^2 d\Omega_3^2 \quad [28]$$

Note that $d\Omega_3^2$ is the metric on a 3-sphere. It is straightforward to derive the Friedmann equations for such a geometry.

The HH no-boundary condition can be interpreted as a topological proposition about the set of path over which we have to sum. The 3-surface $B$ is to be taken as the only surface of compact 4-manifold $M$ which is endowed with the metric $g_{\mu\nu}$, and $\bar{h}_{ij}$ and $\bar{\phi}$ are the induced metric and the scalar field on the surface. The wave function is obtained by using the matching condition supplemented with initial condition. For the minisuperspace case, initial conditions impose constraints on the scale factor $a(\tau = 0)$ and $(da/d\tau)(\tau = o)$, and $N^{\perp}$ is to be gauge fixed. These conditions are to be implemented in the context of determining the wave function of the universe. In the case of the tunneling boundary condition of Vilenkin, the qualitative scenario is as follows. If we look at the solution to the WDW equation (in the path-integral approach, Vilenkin considers Lorentzian action), the solution, crudely speaking, has both ingoing and outgoing modes at the boundary. In his proposal, the outgoing mode at the boundary is to be accepted. The exact prescription is lot more subtle than the above statement, since one has to define the meaning of outgoing mode carefully in the absence of a timelike Killing vector when we write the WDW equation on the superspace. The qualitative picture for Vilenkin's boundary condition, in the minisuperspace, is like tunneling solutions in quantum mechanics when a particle penetrates through a potential barrier.

Let us consider a minisuperspace model with the scalar field and potential $V(\phi)$. The action is

$$S = \frac{1}{2}\int dt a^3 \left[ -\frac{1}{N^{\perp}}\left(\frac{\dot{a}}{a}\right)^2 + \frac{(\dot{\phi})^2}{N^{\perp}} - N^{\perp} V(\phi) + \frac{N^{\perp}}{a^2} \right] \quad [29]$$

A few comments are in order here. For the FRW metric, we have $\sqrt{g}R = 6(-a\dot{a} + ka) +$ a total derivative term; the total derivative term can be removed by adding a boundary term and $k$ is positive since we take the spatial part to be closed. We have redefined the scale factor, the scalar field, the potential term, and $k$ such that the Einstein–Hilbert action with matter field assumes the form of [29] and this action facilitates the definition of conjugate momenta without cumbersome numerical factors, and the Hamiltonian takes a simple form. The conjugate momenta and resulting Hamiltonian are

$$\pi_a = -\frac{a\dot{a}}{N^{\perp}}, \qquad \pi_{\phi} = \frac{a^3\dot{\phi}}{N^{\perp}} \quad [30]$$

$$H_c = \frac{N^{\perp}}{2}\left[ -\frac{\pi_a^2}{a} + \frac{\pi_{\phi}^2}{a^3} + a^3 V(\phi) - a \right] = N^{\perp} H \quad [31]$$

and the constraint is $H = 0$. In the quantum cosmology context, we solve the WDW equation: $H\Psi = 0$. Since the exact solution is not possible, one resorts to some approximation with simple assumptions. The differential equation is

$$\left[ \frac{\partial^2}{\partial a^2} - \frac{1}{a^2}\frac{\partial^2}{\partial \phi^2} + a^4 V(\phi) - a^2 \right]\Psi = 0 \quad [32]$$

Let us consider the case when $V(\phi)$ does not grow very fast, that is, $V(\phi)/V(\phi)' << 1$ and consider the solution to the WDW equation where $\Psi$ has weak dependence on $\phi$. Consequently, we may ignore the $\phi$ derivative term in [32]. The purpose of these assumptions is to reduce the problem to a one-dimensional quantum mechanics problem and then employ WKB method. It is hoped that at least some of the nonperturbative aspects can still be captured. When the effective potential appearing in [32] is negative (this is a classically inaccessible region), the wave function is

$$\Psi(a, \phi) \approx e^{\pm(1/3V(\phi))(1 - a^2 V(\phi))^{3/2}} \quad [33]$$

We expect the wave function have oscillatory behavior in the classically allowed domain and it does have that property,

$$\Psi(a, \phi) \approx e^{\pm(i/3V(\phi))(a^2 V(\phi) - 1)^{3/2}} \quad [34]$$

The choice of signs is decided from the boundary conditions imposed and the usual matching of the wave functions of the two regions is done as is the case with the WKB approximation. Note that we are considering the metric and the scalar field on

the boundary which were denoted by $\bar{h}_{ij}$ and $\bar{\phi}$; strictly speaking, we should denote the solutions as $\bar{a}$ and $\bar{\phi}$. But from now on, we drop this bar on $a$ and $\phi$.

Let us momentarily assume that $V$ is $\phi$-independent and therefore, we have an effective cosmological constant. The problem is identical to the motion of a particle in a potential well. There are two turning points. In one region, the particle starts from $a = 0$, reaches one turning point $r_1$ and returns back. In another case, it starts from $a = \infty$, travels up to $a = r_2$ and reflects back. In the quantum-mechanical case, the particle can tunnel through the barrier. The wave function has both decaying and growing modes under the barrier, and boundary conditions tell us which mode to choose. One possibility is that the particle starts from $a = 0$, tunnels through and proceeds towards $a = \infty$, that is, it has outgoing mode. The other possibility is that the wave function has both outgoing and ingoing modes. In this simple scenario, the former corresponds to Vilenkin's tunneling boundary condition, where the universe is created at $a = 0$ and it keeps growing. The latter is HH no-boundary proposal where the wave function has both modes and the universe contracts and expands.

Now we discuss the two boundary conditions in the presence of the potential, with the approximations mentioned above. The proposition of Vilenkin amounts to the following conditions on the wave function: the region of the boundary which is nonsingular is $\phi$ finite and $a = 0$. Other than this domain, either $a$ or $\phi$ diverge on any other region of the boundary; both can diverge in this singular boundary. Notice from the expression for [33] and [34] that the tunneling region corresponds to $a^2V(\phi) < 1$, whereas, the oscillatory domain is $a^2V(\phi) > 1$. If we use the saddle-point approximation, $\Psi \approx e^{\pm iS_{cl}}$. Vilenkin's boundary condition corresponds to $\Psi \approx e^{-iS_{cl}}$, with

$$S_{cl} = \frac{(\sqrt{a^2V(\phi) - 1})^3}{3V(\phi)}$$

So far, we considered the situation where differential operator for $\phi$ is dropped in [32]. In order to account for weak $\phi$-dependence, we could introduce it by multiplying a slowly varying function, say $F(\phi)$ and write $\Psi(a, \phi) \sim F(\phi)e^{-iS_{cl}}$. Similarly, the wave function can be obtained under the barrier and required to satisfy WKB matching conditions. Furthermore, the regularity condition on the wave function in small scale factor limit and behavior of its derivative with respect to $\phi$ in that limit determine the form of $F(\phi)$. In summary, the Vilenkin boundary conditions yield the following wave functions:

$$\Psi(a, \phi)_V \approx e^{-(1/3V(\phi))(1-[1-a^2V(\phi)]^{3/2})} \qquad [35]$$

$$\Psi(a, \phi)_V \approx e^{-1/3V(\phi)}\, e^{-(i/3V(\phi))[a^2V(\phi)-1]^{3/2}} \qquad [36]$$

Note that [35] is the wave function under the barrier, that is, $a^2V(\phi) < 1$ in this region, whereas [36] is in the classically accessible domain ($a^2V(\phi) > 1$) which is reflected by the oscillatory character. The slowly varying function $F(\phi) \sim e^{-1/3V(\phi)}$ appears as the common factor for the wave functions in the two domains.

The HH no-boundary proposal to derive the wave function of the universe was formulated in the Euclidean path-integral formalism. A considerable amount of attention has been focused in this area. We shall present the HH wave function providing only a sketchy argument. In the Euclidean description, 4-metric is $ds^2 = (N^\perp)^2 d\tau^2 + a^2(\tau)d\Omega_3^2$. The 4-geometry should close in a regular way. If we make the bounding 3-space smaller and smaller, it can be closed with flat space. We can infer about the behavior of the scale factor in the limit $\tau \to 0$ from this consideration. Furthermore, in the semiclassical approximation $\Psi(a, \phi) \sim e^{-S_E}$; we have replaced $(\bar{a}, \bar{\phi})$ by $(a, \phi)$ as remarked earlier. Thus, our aim is to evaluate $S_E$ at the saddle point. This is achieved by writing down the (Euclidean version) field equations for $a$ and $\phi$ and the Hamiltonian constraint, and then solve for $a(\tau), \phi(\tau)$, and $N^\perp(\tau)$. Eventually, we want to eliminate $N^\perp$ and then obtain $S_E$. After all, the path integral is dominated by the classical trajectory, $a(\tau)$, and one does not fix the gauge for $N^\perp$ while solving for $a$. In fact, the lapse gets eliminated by utilizing the Hamiltonian constraint which involve $\tau$-derivatives of both $a$ and $\phi$. We mention, without going into details, that the classical action is not unique. One of the ways to visualize it is to note that the solutions obtained for the lapse from the Hamiltonian constraint have sign ambiguities.

The classical action is

$$S_E^{\pm} = -\frac{1}{3V(\phi)}\left(1 \pm [1 - a^2V(\phi)]^{3/2}\right) \qquad [37]$$

Note that the two solutions correspond to 3-sphere boundary being closed off by sections of 4-sphere. Moreover, the Euclidean action is negative. Hartle and Hawking argue that the negative sign in [37] gives the correct answer since the wave function peaks for that choice. However, there is no unanimity

for HH argument and some authors have put forward a point of view that additional inputs are necessary to arrive at the HH conclusion about choosing the negative sign for $S_E$ in [37]. We refer the reader to the reviews of Hartle and Halliwell for detailed discussions on the choice of contours for path integrals, subtleties involved in getting various solutions for the lapse and their interpretations. We give below the wave function under the barrier (with choice of negative sign in [37]):

$$\Psi_{\mathrm{HH}}(a,\phi) \approx \mathrm{e}^{(1/3v(\phi))(1-[1-a^2V(\phi)]^{3/2})} \qquad [38]$$

$$\Psi_{\mathrm{HH}}(a,\phi) \approx \mathrm{e}^{1/3V(\phi)} \times \cos\left(\frac{1}{3V(\phi)}[a^2V(\phi)-1]^{3/2} - \frac{\pi}{4}\right) \qquad [39]$$

**Remarks** The wave function in [38] is obtained in the classically inaccessible region under the condition $a^2V(\phi) < 1$, and wave function [39] corresponds to the case $a^2V(\phi) > 1$, where the particle motion is permissible classically. Note the factor $\mathrm{e}^{1/3V(\phi)}$ in the wave functions in both the regions and compare that with the Vilenkin's wave function which has the opposite sign. We may conclude where the wave function will peak for each of the two boundary conditions. Whereas Vilenkin's proposal implies that $\Psi_V(a,\phi)$ peaks when $V(\phi)$ takes large values, HH no-boundary condition tells us that it peaks when $V(\phi) \to 0$. Furthermore, we note that $\Psi_V$ is complex and $\Psi_{\mathrm{HH}}$ is real in the oscillatory region. Although the debates on the merits and demerits of each of the boundary proposals are going on for more than two decades, the issue is far from being settled. In the absence of any experimental tests, there is no way to favor one boundary proposal over another. Then, boundary conditions do have predictions about the evolution of the universe after the quantum era and have predictions in that (classical) regime. Therefore, determination of the wave function with specific boundary conditions does have some connections with the laws that govern the evolution of our universe in the present epoch.

It is worthwhile to dwell on the WDW equation from the perspectives of string theories. Indeed, there have been important developments to understand the dynamics of the universe in the string-theoretic framework. It is important to note the key role played by dilaton in string theory: (1) it is one of the massless states of the theory, and (2) the vacuum expectation value (VEV) of this field determines the coupling constants we hope to use in describing fundamental interactions. Therefore, the graviton is always accompanied by the dilaton in any string-theoretic approach to study the universe. The duality symmetries are recognized to provide deep understanding of the string dynamics. Therefore, the investigations of quantum gravity phenomena from the string-theory viewpoint are necessarily influenced by above mentioned facts. Indeed, classical cosmological solutions, derived from string effective action, have several interesting characteristics. We mention is passing that the WDW equation has played an important role to study quantum evolution equations in string cosmology. The choice of operator-ordering prescription in defining the WDW Laplace–Beltrami operator can be resolved by appealing to the duality symmetries. Furthermore, the boundary conditions imposed on the wave function are dictated by string symmetries and therefore, the resulting wave function has very interesting properties. The string theory has addressed some of the most important problems in quantum gravity and it has provided resolutions to several key issues. It is expected that string theory will provide answers to challenging questions in quantum cosmology. In summary, we have conveyed some of the salient aspects of the WDW equation. The canonical quantization technique is adopted to study quantum gravity in this approach. We have illustrated the crucial role of the constraint formalism due to Dirac and argued that some of the nonperturbative aspects of quantum gravity could be retained. In a short article of this nature, it is not possible to provide detailed discussion about the general derivation of the WDW equation and discuss the role of boundary conditions more exhaustively. Instead, we presented some of the key steps in the derivation of the WDW equation adopting the canonical formalism and provided simple examples. The subject is still an active area of research. The interested reader may benefit from the bibliography.

*See also:* Canonical General Relativity; Loop Quantum Gravity; Quantum Cosmology; Quantum Dynamics in Loop Quantum Gravity; Quantum Geometry and its Applications; Superstring Theories.

## Further Reading


De Witt BS (1967) Quantum theory of gravity. *Physical Review* 160: 1113.

Feynman RP, Morinigo FB, and Wagner WG (1995) *Feynman Lectures on Gravity*. New York: Addison-Wesley.

Halliwell JJ (1990) Quantum cosmology. In: Randjbar-Daemi S, Sezgin E, and Shafi Q (eds.) *Summer School in High Energy Physics and Cosmology*, p. 513. Singapore: World Scientific.

Hartle JB (1989) Introductory lectures on quantum cosmology. In: Coleman S, Hartle JB, Piran T, and Weinberg S (eds.) *Quantum Cosmology and Baby Universes*, Proceedings of 7th Winter School in Theoretical Physics, p. 159. Jerusalem: World Scientific.

Hartle JB and Hawking SW (1983) Wave function of the universe. *Physical Review D* 28: 2960.

Hawking SW (1983) Quantum cosmology. Les Houches Lectures on Quantum Cosmology, p. 333. (Les Houches Publication) in Einstein Centenary Volume.

Vilenkin A (1984) Quantum creation of universe. *Physical Review D* 30: 509.

Wheeler JA (1963) Geometrodynamics. In: De Witt C and De Witt BS (eds.) *Relativity, Groups and Topology*. New York: Gordon and Breach.

**Wightman Axioms** *see* Axiomatic Quantum Field Theory

# Wulff Droplets

**S Shlosman**, Université de Marseille, Marseille, France



## Introduction

Historically, the first question where the Wulff shapes have appeared is the one of the formation of a droplet or a crystal of one substance inside another. The natural problem here is: what shape such a formation would take? The statement that such a shape should be defined by the minimum of the overall surface energy subject to the volume constraint is physically very natural. In the isotropic case, when the surface tension does not depend on the orientation of the surface, and so is just a positive number, the shape in question should be of course spherical (provided we neglect the gravitational effects). In a more general situation the shape in question is less symmetric. The corresponding variational problem is called the Wulff problem. Wulff (1901) formulated it in his paper, where he also presented a geometric solution to it, called the "Wulff construction."

The Wulff variational problem is formulated as follows. Let $\tau(\boldsymbol{n})$, $\boldsymbol{n} \in S^{d-1}$, be some continuous function on the unit sphere $S^{d-1} \subset \mathbb{R}^d$. We suppose that $\tau > 0$, and that $\tau$ is even: $\tau(\boldsymbol{n}) = \tau(-\boldsymbol{n})$. The value $\tau(\boldsymbol{n})$ plays the role of the surface tension between two phases separated by the hyperplane orthogonal to the vector $\boldsymbol{n}$. For every closed compact (hyper)surface $M^{d-1} \subset \mathbb{R}^d$, we define its surface energy as

$$\mathcal{W}_\tau(M) = \int_M \tau(\boldsymbol{n}_s)\,\mathrm{d}s$$

where $\boldsymbol{n}_s$ is the normal vector to $M$ at $s \in M$. The functional $\mathcal{W}_\tau(M)$ has the meaning of the surface energy of the $M$-shaped droplet made from one of these two phases. It is called the Wulff functional. Let $\mathfrak{W}_\tau$ be the surface which minimizes $\mathcal{W}_\tau(\cdot)$ over all the surfaces enclosing the unit volume. Such a minimizer does exist and is unique up to translation. It is called the Wulff shape.

The following is the geometric construction of $\mathfrak{W}_\tau$. Consider the set

$$K_\tau = \left\{ \boldsymbol{x} \in \mathbb{R}^d:\ \forall \boldsymbol{n} \in S^{d-1} (\boldsymbol{x}, \boldsymbol{n}) \le \tau(\boldsymbol{n}) \right\}$$

If we define the half-spaces

$$L_{\tau,\boldsymbol{n}} = \left\{ \boldsymbol{x} \in \mathbb{R}^d:\ (\boldsymbol{x}, \boldsymbol{n}) \le \tau(\boldsymbol{n}) \right\}$$

then

$$K_\tau = \cap_{\boldsymbol{n}} L_{\tau,\boldsymbol{n}} \qquad [1]$$

In particular, $K_\tau$ is convex. It turns out that

$$\mathfrak{W}_\tau = \lambda_\tau \partial(K_\tau)$$

where the dilatation factor $\lambda_\tau$ is defined by the normalization: $\mathrm{vol}(\lambda_\tau K_\tau) = 1$. The relation [1] is called the Wulff construction. For the future use, we introduce the notation $w_\tau$ for the value of the surface energy of the Wulff shape:

$$w_\tau = \mathcal{W}_\tau(\mathfrak{W}_\tau)$$

The Wulff construction was considered by the rigorous statistical mechanics as just a phenomenological statement, though the notion of the surface tension was among its central notions. The situation changed after the appearance of the book by Dobrushin *et al.* (1992). There it was shown that in the setting of the canonical ensemble formalism, in the regime of the first-order phase transition, the (random) shape occupied by one of the phases has asymptotically (in the thermodynamic limit) a nonrandom shape, given precisely by the Wulff construction! In other words, a typical macroscopic random droplet looks very close to the Wulff shape.

In what follows we will explain the above result. Another important application of the concepts introduced above – the role played by the Wulff

shapes in the theory of metastability – is also described (*see* Metastable States).

## Crystals in the Ising Model

Ising spins $\sigma_x$ take values $\pm 1$, with $x \in \mathbb{Z}^d$. We will wrap $\mathbb{Z}^d$ into a torus $T_N^d$ by taking a factor lattice: $T_N^d = \mathbb{Z}^d / N\mathbb{Z}^d$. Ising-model grand canonical Gibbs state in $T_N^d$ is the probability measure $\mu_N^\beta$:

$$\mu_N^\beta(\sigma) = Z_{N,\beta}^{-1} \exp(-\beta H_N(\sigma))$$

Here $H_N(\sigma) = -\sum_{x,y\,\text{n.n.},\,x,y \in T_N^d} \sigma_x \sigma_y, \beta > 0$ is the inverse temperature, and $Z_{N,\beta}$ is the normalization factor. Ising-model canonical Gibbs state in $T_N^d$ is the probability measure $\mu_N^{\beta,\rho}$, obtained from $\mu_N^\beta$ by taking its conditional distribution:

$$\mu_N^{\beta,\rho}(\sigma) = \mu_N^\beta \left( \sigma \mid \sum_{x \in T_N^d} \sigma_x = \rho N^d \right), \quad |\rho| < 1$$

(Here we make a slight abuse of notation. More precisely, since $\sigma_x = \pm 1$, one has to consider the conditioning $\sum \sigma_x = \rho_N N^d$, where $\rho_N \to \rho$ as $N \to \infty$, while the numbers $(1 - \rho_N)N^d$ are even integers; otherwise the condition is empty.) We will characterize the canonical state $\mu_N^{\beta,\rho}$ by describing the properties of contours, $\{\gamma_i(\sigma)\}$, of configuration $\sigma$. Contours $\gamma_i$ of configuration $\sigma$ are hypersurfaces made of elementary $(d-1)$-dimensional unit cubes of the dual lattice, which separate the nearest-neighbor (n.n.) points $x, y \in T_N^d$ where $\sigma_x \neq \sigma_y$.

Suppose that the temperature $\beta^{-1}$ is low enough, while the density parameter $\rho$ satisfies the constraints:

$$m_d^*(\beta) > \rho > g_d$$

Here $m_d^*(\beta)$ is the spontaneous magnetization of the $d$-dimensional Ising model, while $g_d$ is some geometric factor, the role of which will be explained later. The above constraint forces some amount of the $(-)$-phase into the $(+)$-phase. It turns out that this amount gathers into one big droplet, which has approximately the Wulff shape.

We first formulate the known rigorous results for the case $d = 2$, and then indicate some extensions.

### Two-Dimensional Case

The following holds with $\mu_N^{\beta,\rho}$-probability approaching 1 as $N \to \infty$:

- The set $\{\gamma_i(\sigma)\}$ of contours of $\sigma$ has precisely one "big" contour, $\Gamma(\sigma)$; the diameters of other contours do not exceed $K \ln N$, $K = K(\beta)$.
- The area $|\mathrm{Int}\,\Gamma(\sigma)|$ inside $\Gamma(\sigma)$ satisfies

$$\big| |\mathrm{Int}\,\Gamma(\sigma)| - \nu N^2 \big| \leq K N^{6/5} (\ln N)^{\varkappa}$$

where

$$\nu = \frac{m_2^*(\beta) - |\rho|}{2 m_2^*(\beta)}, \quad \varkappa = \varkappa(\rho)$$

- There is a point $x = x(\sigma)$ – the "center" of $\Gamma(\sigma)$ – such that the shift of $\Gamma(\sigma)$ by $-x(\sigma)$ brings the contour $\Gamma(\sigma)$ very close to the scaled Wulff curve, defined by the Ising-model surface tension $\tau$:

$$r_{\mathrm{H}}\left( \Gamma(\sigma) - x(\sigma), \sqrt{\frac{2\nu}{w_\tau}} N W_\tau \right) \leq K N^{2/3} (\ln N)^{\varkappa} \quad [2]$$

(Here $r_{\mathrm{H}}$ is the Hausdorff distance: for every two sets $A, C \in \mathbb{R}^d$, $r_{\mathrm{H}}(A, C)$ is defined as $\max\{\inf[r : A \subset C + B_r], \inf[r : C \subset A + B_r]\}$, where $B_r$ is the ball of radius $r$.)

The proof of the above result is the content of the book by Dobrushin *et al.* (1992). In the two-dimensional case, it remains true for all temperatures $\beta^{-1}$ below the critical one (Ioffe and Schonmann 1998). The value 2/3 of the exponent is an improvement of the original 3/4 result (Alexander 1992). Probably, it can be further improved down to 1/2. Though Dobrushin *et al.* (1992) treat only the Ising model, their results are valid for a wide range of other models.

The restriction $\rho > g_d$ in the theorem is needed because without it the droplet may prefer to assume the shape of a strip between two meridians rather than to take the Wulff shape.

### Three-Dimensional Case

In the case $d = 3$ or $d \geq 3$, the statement that a typical configuration $\sigma$ has only one big contour $\Gamma(\sigma)$ is still true. But the analog of [2] is not known. It is natural to conjecture that it holds at low temperatures, even in a stronger version, with only a logarithmic term $K(\ln N)^{\varkappa}$ in the RHS. It probably fails at higher subcritical temperatures.

What is known to hold is a weaker version of this theorem, where the distance between random droplet and the Wulff shape is measured not in Hausdorf distance, but in $L^1$ sense. To state the corresponding theorem, we will associate with every configuration $\sigma$ on a lattice torus $T_N^d$ a real-valued function $M_\sigma(t)$ on the unit torus $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$, and we then compare this function with the indicator function $\mathbb{I}_{sK_\tau}$, where $sK_\tau \subset \mathbb{T}^d$ is the Wulff body, properly scaled.

The function $M_\sigma(t)$ is defined as follows. We denote by $i_N$ the natural embedding of the discrete torus $T_N^d$ into $\mathbb{T}^d$, the image of $i_N$ being the grid with spacing $1/N$. For $t \in \mathbb{T}^d$ we define $b_N(t) \subset \mathbb{T}^d$ to be

the ball centered at $t$ with radius $\sqrt[d]{1/N}$, and let $B_N(t) \subset \Lambda(N)$ be its preimage under $i_N$. Then

$$M_\sigma(r) = \frac{1}{|B_N(t)|} \sum_{x \in B_N(t)} \sigma(x)$$

We have to expect to see a droplet $sK_\tau$ with

$$s = \sqrt[d]{\frac{d}{w_\tau} \frac{m_d^*(\beta) - \rho}{2m_d^*(\beta)}}$$

Let us introduce for every subset $A \subset \mathbb{T}^d$ the indicator

$$\mathbb{I}_A(t) = \begin{cases} 1, & t \in A \\ -1, & t \in A^c \end{cases}$$

For every function $v$ in $L^1(\mathbb{T}^d)$ we denote by $U(v, \delta)$ its $\delta$-neighborhood in $L^1(\mathbb{T}^d)$.

The result can now be formulated. Suppose the temperature $\beta^{-1}$ is below the critical one. Then the function $M_\sigma(t)$ is close to the characteristic function of the Wulff shape: For every $\delta > 0$

$$\lim_{N \to \infty} \mu_N^{\beta,\rho} \left\{ \frac{1}{m_d^*(\beta)} M_\sigma(\cdot) \in \bigcup_{t \in \mathbb{T}^d} U(\mathbb{I}_{sK_\tau + t}, \delta) \right\} = 1$$

The shifts by all $t$–s of the Wulff shape $sK_\tau$ appear in the statement since the location of the droplet can be arbitrary. Note that if a point $t$ is such that the ball $B_N(t)$ stays away from the boundary of the droplet $\Gamma(\sigma)$ present in the configuration $\sigma$, then the value $M_\sigma(t)$ should be expected to be $\pm m_d^*(\beta)$, depending on whether $t$ is outside or inside the droplet, which explains the factor $1/m_d^*(\beta)$.

For a proof, see Bodineau (1999) and Cerf and Pizstora (1999).

*See also:* Cluster Expansion; Large Deviations in Equilibrium Statistical Mechanics; Metastable States; Percolation Theory; Statistical Mechanics of Interfaces.

## Further Reading

Alexander K (1992) Stability of the Wulff minimum and fluctuations in shape for large clusters in two-dimensional percolation. *Probability Theory and Related Fields* 91: 507–532.

Bodineau T (1999) The Wulff construction in three and more dimensions. *Communications in Mathematical Physics* 207(1): 197–229.

Cerf R and Pizstora A (2000) On the Wulff crystal in the Ising model. *Annals of Probability* 28(3): 947–1017.

Dobrushin RL, Kotecky R, and Shlosman SB (1992) *Wulff Construction: A Global Shape from Local Interaction*. AMS Translations Series. Providence, RI: American Mathematical Society.

Ioffe D and Schonmann R (1998) Dobrushin–Kotecky–Shlosman theory up to the critical temperature. *Communications in Mathematical Physics* 199: 117–167.

Wulff G (1901) Zur frage der geschwindigkeit des wachsturms under auflosung der kristallchen. *Z. Kristallogr.* 34: 449–530.

# Y

## Yang–Baxter Equations

**J H H Perk and H Au-Yang**, Oklahoma State University, Stillwater, OK, USA



### Introduction

The term Yang–Baxter equations (YBEs) was coined by Faddeev in the late 1970s to denote a principle of integrability, that is, exact solvability, in a wide variety of fields in physics and mathematics. Since then it has become a common name for several classes of local equivalence transformations in statistical mechanics, quantum field theory, differential equations, knot theory, quantum groups, and other disciplines. We shall cover the various versions and their relationships, paying attention also to the early historical development.

#### Electric Networks

The first such transformation came up as early as 1899 when the Brooklyn engineer Kennelly published a short paper, entitled "The equivalence of triangles and three-pointed stars in conducting networks." This work gave the definite answer to such questions as whether it is better to have the three coils in a dynamo – or three resistors in a network – arranged as a star or as a triangle, see Figure 1. Using Kirchhoff's laws, the two situations in Figure 1 can be shown to be equivalent provided

$$Z_1\overline{Z}_1 = Z_2\overline{Z}_2 = Z_3\overline{Z}_3$$

$$= Z_1Z_2 + Z_2Z_3 + Z_3Z_1 \qquad [1]$$

$$= \overline{Z}_1\overline{Z}_2\overline{Z}_3/(\overline{Z}_1 + \overline{Z}_2 + \overline{Z}_3) \qquad [2]$$

Here one has to take either [1] or [2] as second line of the equation, depending on which direction the transformation is to go. The star–triangle transformation thus defined is also known under other names within the electric network theory literature as wye–delta $(\mathrm{Y}-\Delta)$, upsilon–delta $(\Upsilon-\Delta)$, or tau–pi $(\mathrm{T}-\Pi)$ transformation.

#### Spin Models

When Onsager wrote his monumental paper on the Ising model published in 1944, he made a brief remark on an obvious star–triangle transformation relating the model on the honeycomb lattice with the one on the triangular lattice. His details on this were first presented in Wannier's review article of 1945. However, the star–triangle transformation played a much more crucial role in Onsager's reasoning, as it is also intimately connected with his elliptic function uniformizing parametrization.

Furthermore, it implies the commutation of transfer matrices and spin-chain Hamiltonians. Only in his Battelle lecture of 1970 did Onsager explain how he used this remarkable observation in his derivation of the formula for the spontaneous magnetization which he had announced as a conference remark in 1948 and of which the first complete derivation had been published by Yang in 1952 using a completely different method.

Many other applications and generalizations have since appeared. Most generally, we can consider a system whose state variables – also called spins – take values from some suitable discrete or continuous sets. The interactions between spins $a$ and $b$ are given in terms of weight factors $W_{ab}$ and $\overline{W}_{ab}$, which are complex numbers in general, see Figure 2. One quantity of special interest is the partition function – sum of the product of all weight factors over all allowed spin values. The integrability of the model is expressed by the existence of spectral variables – rapidities $p, q, r, \ldots$ – that live on oriented lines, two of which cross between $a$ and $b$ as indicated by the dotted lines in Figure 2. Arrows from $a$ to $b$ are added to keep track of the ordering of $a$ and $b$ in case the weights are chiral (not symmetric).

In Onsager's special choice of the Ising model the spins take values $a, b, c, \ldots = \pm 1$ and the weight factors are the usual real positive Boltzmann weights depending on the product $ab = \pm 1$, uniformizing variable $p - q$, and elliptic modulus $k$. In the integrable chiral Potts model the weights depend on $a - b \bmod N$, with $a, b = 1, \ldots, N$, whereas the rapidities $p$ and $q$ are living in general on a higher-genus curve.

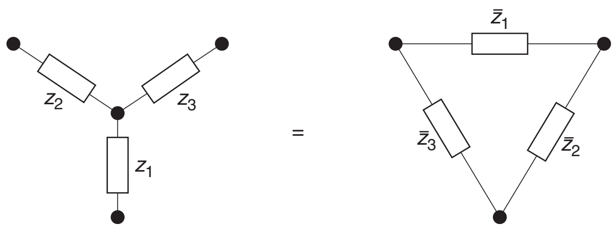



**Figure 1** Star–triangle equation for impedances.

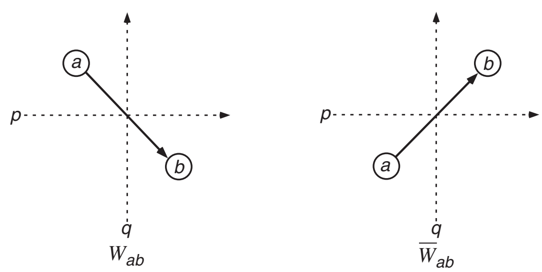



**Figure 2** Spin model weights $W_{ab}(p,q)$ and $\overline{W}_{ab}(p,q)$.

When the weights are asymmetric in the spins, there are two sets of star–triangle equations which can be expressed both pictorially (Figure 3) and algebraically:

$$\sum_d \overline{W}_{cd}(p,q)\overline{W}_{db}(q,r)W_{da}(p,r) = R(p,q,r)W_{ba}(p,q)W_{ca}(q,r)\overline{W}_{cb}(p,r) \quad [3]$$

$$\overline{R}(p,q,r)W_{ab}(p,q)W_{ac}(q,r)\overline{W}_{bc}(p,r) = \sum_d \overline{W}_{dc}(p,q)\overline{W}_{bd}(q,r)W_{ad}(p,r) \quad [4]$$

Note that eqns [3] and [4] differ from each other by the transposition of both spin variables in all six weight factors. In general, there may also appear scalar factors $R(p,q,r)$ and $\overline{R}(p,q,r)$, which can often be eliminated by a suitable renormalization of the weights. If $a$, $b$, and $c$ take values in the same set, we can sum over $a=b=c$, showing that $R=\overline{R}$ in that case.

The Kennelly star–triangle equation [1], [2] can be recovered as a special limit of a spin model where the states are continuous variables.

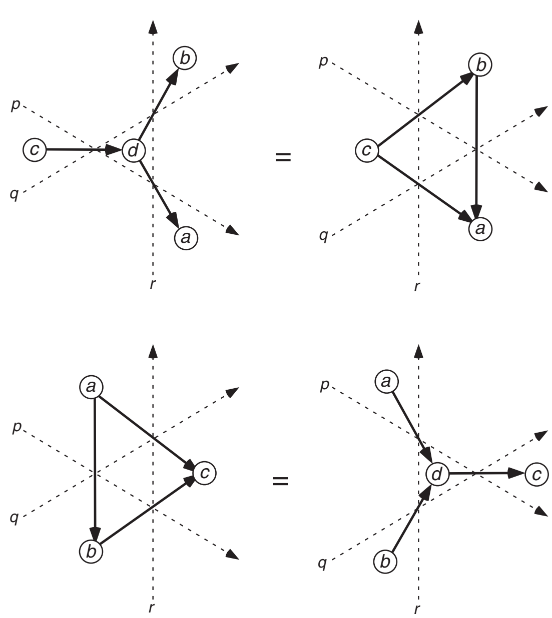



**Figure 3** Star–triangle equation.

### Knot Theory and Braid Group

A seemingly totally different situation occurs in the theory of knots, links, tangles, and braids. In 1926, Reidemeister showed that only three types of moves suffice to show the equivalence between two different configurations, see Figure 4. Moves of type I – removing simple loops – do not apply to braids. Moves of type II, for which one strand crosses twice over another strand, can be reformulated for braids, namely that an overcrossing is the inverse of an undercrossing. The Reidemeister move of type III is a precursor of the more general Yang–Baxter moves and can be represented also by the defining relations of Artin's braid group. Let $\boldsymbol{R}_{i,i+1}$ be the operator representing the situation in which the strand in position $i$ crosses over the one in position $i+1$. Then a braid can be represented by a product of $\boldsymbol{R}_{j,j+1}$'s and their inverses, provided

$$\boldsymbol{R}_{i,i+1}\boldsymbol{R}_{i+1,i+2}\boldsymbol{R}_{i,i+1} = \boldsymbol{R}_{i+1,i+2}\boldsymbol{R}_{i,i+1}\boldsymbol{R}_{i+1,i+2} \quad [5]$$

and

$$[\boldsymbol{R}_{i,i+1}, \boldsymbol{R}_{j,j+1}] = 0, \quad \text{if } |i-j| \geq 2 \quad [6]$$

and similar relations in which $\boldsymbol{R}_{i,i+1}$ and/or $\boldsymbol{R}_{i+1,i+2}$ are replaced by their inverses.

### Factorizable *S*-Matrices and Bethe Ansatz

In the early 1960s, Lieb and Liniger solved the one-dimensional Bose gas with delta-function interaction using the Bethe ansatz. Yang and McGuire then tried to generalize this result to systems with internal degrees of freedom and to fermions. This led to the

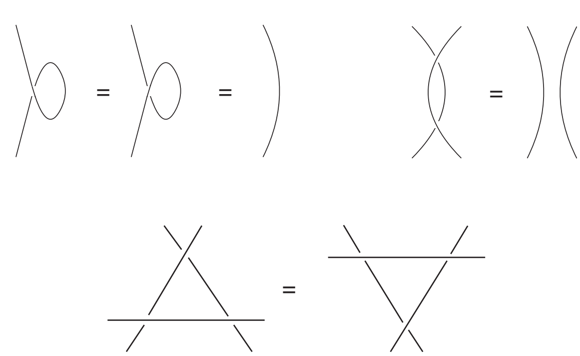


**Figure 4** Reidemeister moves of types I, II, and III.

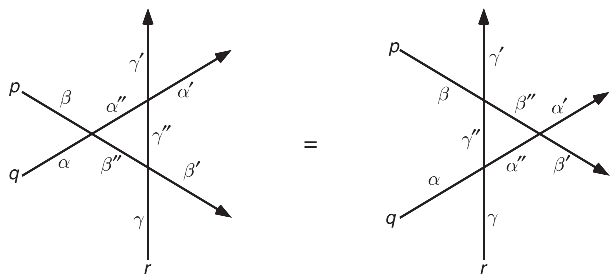



**Figure 5** Vertex model YBE.

discovery of the condition for factorizable *S*-matrices by McGuire in 1964, represented pictorially by Figure 5, where the world lines of the particles are given. Upon collisions the particles can only exchange their rapidities $p, q, r$, so that there is no dispersion. Also indicated are the internal degrees of freedom in Greek letters. In other words, the three-body *S*-matrix can be factorized in terms of two-body contributions and the order of the collisions does not affect the final outcome. McGuire also realized that this condition is all one needs for the consistency of factoring the *n*-body *S*-matrix in terms of two-body *S*-matrices. The consistency condition is obviously related to the Reidemeister move of type III in Figure 4.

Yang succeeded in solving the spin-1/2 fermionic model using a nested Bethe ansatz, utilizing a generalization of Artin's braid relations [5] and [6],

$$\check{R}_{i,i+1}(p-q)\check{R}_{i+1,i+2}(p-r)\check{R}_{i,i+1}(q-r)$$
$$= \check{R}_{i+1,i+2}(q-r)\check{R}_{i,i+1}(p-r)\check{R}_{i+1,i+2}(p-q) \qquad [7]$$

He submitted his findings in two short papers in 1967. The $\check{R}$ operators in eqn [7] – a notation introduced later by the Leningrad school – depend on differences of two momenta or two relativistic rapidities. Sutherland solved the general spin case using repeated nested Bethe ansätze, while Lieb and Wu used Yang's work to solve the one-dimensional Hubbard model.

### Vertex Models

Since Lieb's solution of the ice model by a Bethe ansatz, there have been many developments on vertex models, in which the state variables live on line segments and weight factors $\omega_{\alpha\mu}^{\lambda\beta}$ are assigned to a vertex where four line segments with the four states $\alpha, \mu, \lambda, \beta$ on them meet, see Figure 6.

Baxter solved the eight-vertex model in 1971, using a method based on commuting transfer matrices, starting from a solution of what he then called the generalized star–triangle equation, but what is now commonly called the Yang–Baxter equation (YBE):

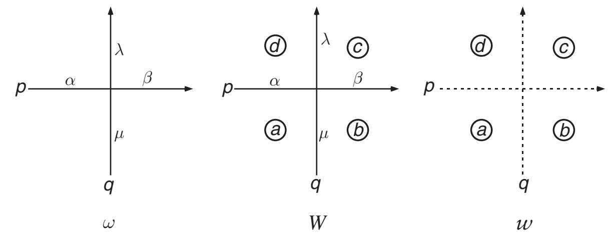



**Figure 6** Vertex model weight $\omega_{\alpha\mu}^{\lambda\beta}(p,q)$, mixed model weight $W_{\alpha\mu}^{\lambda\beta}|_{ab}^{dc}(p,q)$ and IRF model weight $w_{ab}^{dc}(p,q)$.

$$\sum_{\alpha''}\sum_{\beta''}\sum_{\gamma''}\omega_{\beta\alpha}^{\alpha''\beta''}(p,q)\omega_{\alpha''\gamma''}^{\gamma'\alpha'}(q,r)\omega_{\beta''\gamma}^{\gamma''\beta'}(p,r)$$
$$= \sum_{\alpha''}\sum_{\beta''}\sum_{\gamma''}\omega_{\beta''\alpha''}^{\alpha'\beta'}(p,q)\omega_{\alpha\gamma}^{\gamma''\alpha''}(q,r)\omega_{\beta\gamma''}^{\gamma'\beta''}(p,r) \qquad [8]$$

This equation is represented graphically in Figure 5. From it one can also derive a sufficient condition for the commutation of transfer matrices and spin-chain Hamiltonians, generalizing the work of McCoy and Wu, who had earlier initiated the search by showing that the general six-vertex model transfer matrix commutes with a Heisenberg spin-chain Hamiltonian. To be more precise, Baxter found that if $\omega_{\alpha\mu}^{\lambda\beta} = \delta_\alpha^\lambda\delta_\mu^\beta$ for some choice of $p$ and $q$, some spin-chain Hamiltonians could be derived as logarithmic derivatives of the transfer matrix.

### Interaction-Round-a-Face Model

Baxter introduced another language, namely that of the IRF or "interaction-round-a-face" model, which he introduced in connection with his solution of the hard-hexagon model. This formulation is convenient when studying one-point functions using the corner-transfer-matrix method. Now the integrability condition can be represented graphically as in Figure 7 or algebraically as

$$\sum_d w_{cb'}^{a'd}(p,q)w_{dc'}^{a'b}(q,r)w_{b'a}^{dc'}(p,r)$$
$$= \sum_{d'} w_{d'a}^{bc'}(p,q)w_{b'a}^{cd'}(q,r)w_{cd'}^{a'b}(p,r) \qquad [9]$$

The spins live on faces enclosed by rapidity lines and the weights $w_{ab}^{dc}(p,q)$ are assigned as in Figure 6.

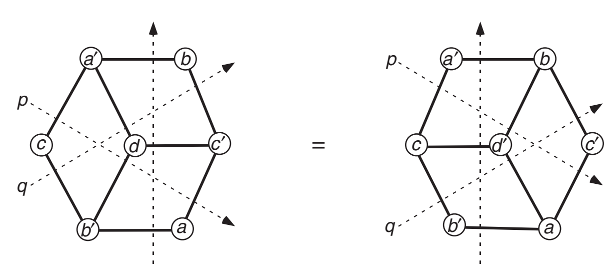



**Figure 7** IRF model YBE.

Baxter discovered a new principle based on eqns [8] and [9], which he called $Z$-invariance, as it expresses an invariance of the partition function $Z$ under moves of rapidity lines. This also implies that typical one-point functions are independent of the values of the rapidities, while two-point functions can only depend on the values of the rapidities of rapidity lines crossing between the two spins considered. Many recent results on correlation functions in integrable models depend on this observation of Baxter.

#### IRF-Vertex Model

In Figure 6, we have also defined mixed IRF-vertex model weights $W^{\lambda\beta}_{\alpha\mu}|^{dc}_{ab}(p,q)$. (We could put further state variables on the vertices, but then the natural thing to do is to introduce new effective weights summing over the states at each vertex.) With the choice made a more general YBE can be represented as in Figure 8, or by

$$\sum_{\alpha''}\sum_{\beta''}\sum_{\gamma''}\sum_{d} W^{\alpha''\beta''}_{\beta\alpha}|^{a'd}_{cb'}(p,q) \times W^{\gamma'\alpha'}_{\alpha''\gamma''}|^{a'b}_{dc'}(q,r)\,W^{\gamma''\beta'}_{\beta''\gamma}|^{dc'}_{b'a}(p,r) = \sum_{\alpha''}\sum_{\beta''}\sum_{\gamma''}\sum_{d'} W^{\alpha'\beta'}_{\beta''\alpha''}|^{bc'}_{d'a}(p,q) \times W^{\gamma''\alpha''}_{\alpha\gamma}|^{cd'}_{b'a}(q,r)\,W^{\gamma'\beta''}_{\beta\gamma''}|^{a'b}_{cd'}(p,r) \qquad [10]$$

#### Quantum Inverse-Scattering Method

The Leningrad school of Faddeev incorporated the methods of Baxter and Yang in their so-called

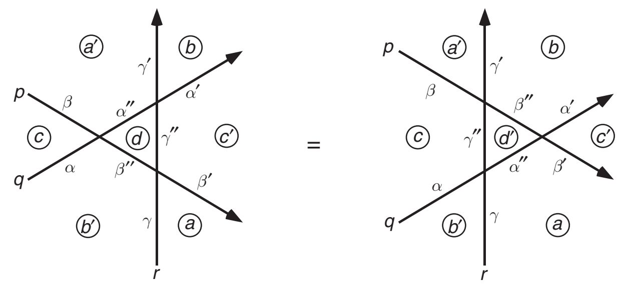


**Figure 8** General YBE.

quantum inverse-scattering method (QISM), coining the term quantum YBEs (QYBEs) for eqns [8]. If special limiting values of $p$ and $q$ can be found, say as $\hbar \to 0$, such that $\omega^{\lambda\beta}_{\alpha\mu} = \delta^{\lambda}_{\mu}\delta^{\beta}_{\alpha} + O(\hbar)$, one can reduce [8] to the classical Yang–Baxter equations (CYBEs) by expanding up to the first nontrivial order in expansion variable $\hbar$. These determine the integrability of certain models of classical mechanics by the inverse-scattering method and the existence of Lax pairs.

## Checkerboard generalizations

Star–triangle equations [3] and [4] imply that there are further generalizations of the YBEs, namely those for which the faces enclosed by the rapidity lines are alternatingly colored black and white in a checkerboard pattern. We can then introduce either vertex model weights $\omega^{\lambda\beta}_{\alpha\mu}(p,q)$ and $\overline{\omega}^{\lambda\beta}_{\alpha\mu}(p,q)$, or IRF-vertex model weights $W^{\lambda\beta}_{\alpha\mu}|^{dc}_{ab}(p,q)$ and $\overline{W}^{\lambda\beta}_{\alpha\mu}|^{dc}_{ab}(p,q)$, or IRF model weights $w^{dc}_{ab}(p,q)$ and $\overline{w}^{dc}_{ab}(p,q)$, see Figure 9.

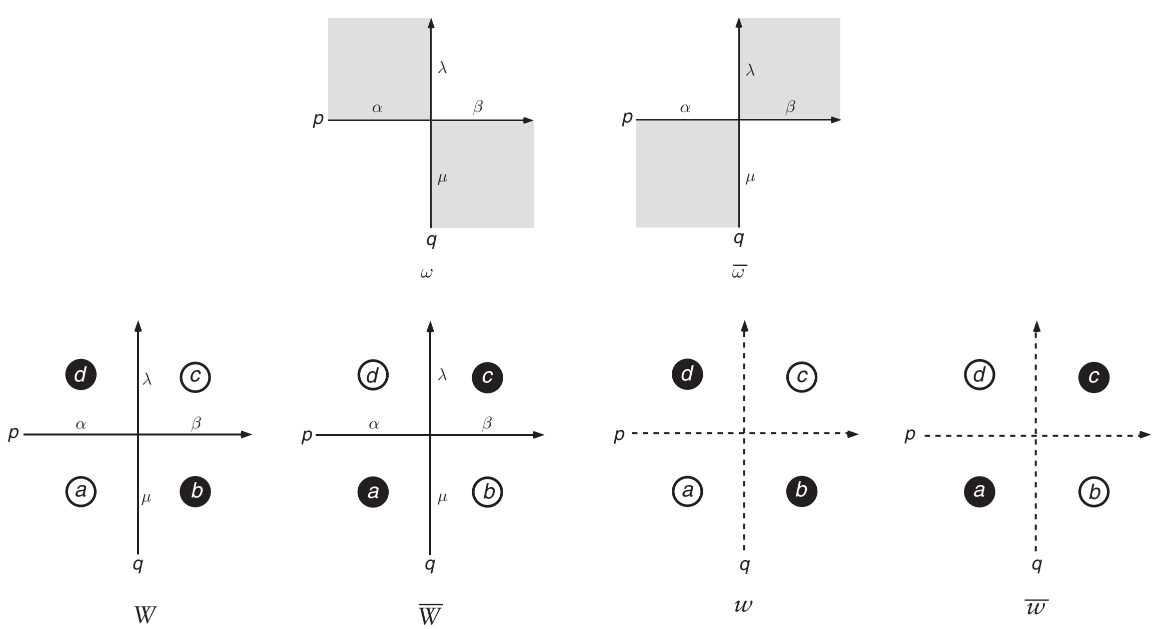


**Figure 9** Checkerboard versions of the weights.

The black faces are those where the spins of the spin model with weights defined in Figure 2 live; the white faces are to be considered empty in Figures 2 and 3 (or, equivalently, they can be assumed to host trivial spins that take on only a single value). Clearly, the IRF-vertex model description contains all the other versions.

### Checkerboard Vertex Model

First we consider the checkerboard vertex model with weights $\omega_{\alpha\mu}^{\lambda\beta}(p,q)$ and $\varpi_{\alpha\mu}^{\lambda\beta}(p,q)$ as assigned in Figure 9. The YBE [8] then generalizes to two sets of equations:

$$\begin{aligned}\sum_{\alpha''}\sum_{\beta''}\sum_{\gamma''}\omega_{\beta\alpha}^{\alpha''\beta''}(p,q)\,\omega_{\alpha''\gamma''}^{\gamma'\alpha'}(q,r)\,\varpi_{\beta''\gamma}^{\gamma''\beta'}(p,r) &= R(p,q,r)\sum_{\alpha''}\sum_{\beta''}\sum_{\gamma''}\varpi_{\beta''\alpha''}^{\alpha'\beta'}(p,q)\\ &\quad\times\varpi_{\alpha\gamma}^{\gamma''\alpha''}(q,r)\,\omega_{\beta\gamma''}^{\gamma'\beta''}(p,r)\end{aligned} \qquad [11]$$

$$\begin{aligned}\overline{R}(p,q,r)\sum_{\alpha''}\sum_{\beta''}\sum_{\gamma''}\varpi_{\beta\alpha}^{\alpha''\beta''}(p,q)\,\varpi_{\alpha''\gamma''}^{\gamma'\alpha'}(q,r)\,\omega_{\beta''\gamma}^{\gamma''\beta'}(p,r) &= \sum_{\alpha''}\sum_{\beta''}\sum_{\gamma''}\omega_{\beta''\alpha''}^{\alpha'\beta'}(p,q)\,\omega_{\alpha\gamma}^{\gamma''\alpha''}(q,r)\,\varpi_{\beta\gamma''}^{\gamma'\beta''}(p,r)\end{aligned} \qquad [12]$$

where scalar factors $R$ and $\overline{R}$ have been added as in [3] and [4]. These equations are represented graphically by Figure 10.

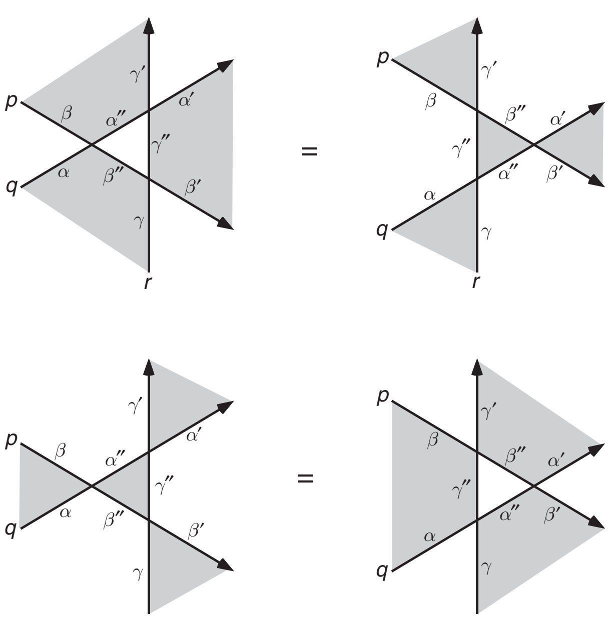


**Figure 10** Checkerboard vertex model YBE.

### Checkerboard IRF Model

The checkerboard IRF version of the YBE [8] becomes

$$\sum_{d} w_{cb'}^{a'd}(p,q)\,w_{dc'}^{a'b}(q,r)\,\overline{w}_{b'a}^{dc'}(p,r) = R(p,q,r)\sum_{d'}\overline{w}_{d'a}^{bc'}(p,q)\,\overline{w}_{b'a}^{cd'}(q,r)\,w_{cd'}^{a'b}(p,r) \qquad [13]$$

$$\overline{R}(p,q,r)\sum_{d}\overline{w}_{cb'}^{a'd}(p,q)\,\overline{w}_{dc'}^{a'b}(q,r)\,w_{b'a}^{dc'}(p,r) = \sum_{d'}w_{d'a}^{bc'}(p,q)\,w_{b'a}^{cd'}(q,r)\,\overline{w}_{cd'}^{a'b}(p,r) \qquad [14]$$

again with scalar factors $R$ and $\overline{R}$ added as in [3] and [4]. These equations can now be represented graphically as in Figure 11. Note that these equations reduce to eqns [3] and [4] if the spins on the white faces are allowed to take only one value, which means that they can be ignored.

### Checkerboard IRF-Vertex Model

Finally, the most general case is represented by the checkerboard IRF-vertex model, with weights defined in Figure 9. For this case the YBEs are given by

$$\begin{aligned}&\sum_{\alpha''}\sum_{\beta''}\sum_{\gamma''}\sum_{d} W_{\beta\alpha}^{\alpha''\beta''}\Big|_{cb'}^{a'd}(p,q)\\ &\quad\times W_{\alpha''\gamma''}^{\gamma'\alpha'}\Big|_{dc'}^{a'b}(q,r)\,\overline{W}_{\beta''\gamma}^{\gamma''\beta'}\Big|_{b'a}^{dc'}(p,r)\\ &= R(p,q,r)\sum_{\alpha''}\sum_{\beta''}\sum_{\gamma''}\sum_{d'}\overline{W}_{\beta''\alpha''}^{\alpha'\beta'}\Big|_{d'a}^{bc'}(p,q)\\ &\quad\times\overline{W}_{\alpha\gamma}^{\gamma''\alpha''}\Big|_{b'a}^{cd'}(q,r)\,W_{\beta\gamma''}^{\gamma'\beta''}\Big|_{cd'}^{a'b}(p,r)\end{aligned} \qquad [15]$$

$$\begin{aligned}&\overline{R}(p,q,r)\sum_{\alpha''}\sum_{\beta''}\sum_{\gamma''}\sum_{d}\overline{W}_{\beta\alpha}^{\alpha''\beta''}\Big|_{cb'}^{a'd}(p,q)\\ &\quad\times\overline{W}_{\alpha''\gamma''}^{\gamma'\alpha'}\Big|_{dc'}^{a'b}(q,r)\,W_{\beta''\gamma}^{\gamma''\beta'}\Big|_{b'a}^{dc'}(p,r)\\ &= \sum_{\alpha''}\sum_{\beta''}\sum_{\gamma''}\sum_{d'}W_{\beta''\alpha''}^{\alpha'\beta'}\Big|_{d'a}^{bc'}(p,q)\\ &\quad\times W_{\alpha\gamma}^{\gamma''\alpha''}\Big|_{b'a}^{cd'}(q,r)\,\overline{W}_{\beta\gamma''}^{\gamma'\beta''}\Big|_{cd'}^{a'b}(p,r)\end{aligned} \qquad [16]$$

with its graphical representation in Figure 12.

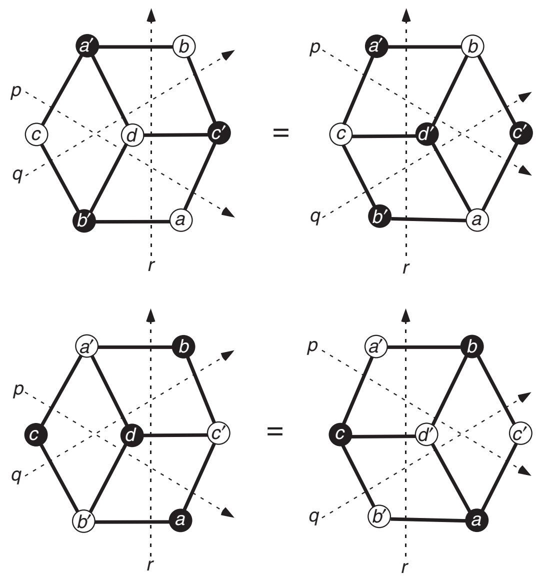



**Figure 11** Checkerboard IRF model YBE.

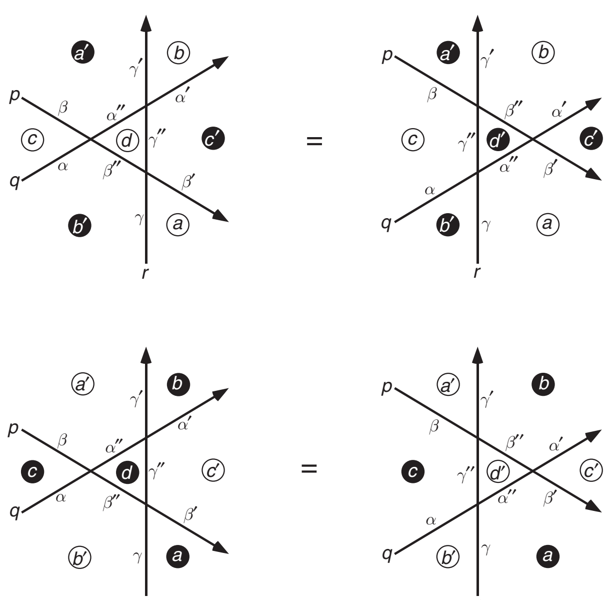



**Figure 12** Checkerboard YBE.

## Formal Equivalence of Languages

### The Square Weight

Combining four weights of a checkerboard model in a square, as is done with four spin model weights in Figure 13, we find a regular vertex model weight with rapidities that are now pairs of the original ones. This gives

$$W_{\alpha\mu}(p_1,q_1)\overline{W}_{\mu\beta}(p_1,q_2)\overline{W}_{\alpha\lambda}(p_2,q_1)W_{\lambda\beta}(p_2,q_2) = \omega_{\alpha\mu}^{\lambda\beta}(p_1,p_2;q_1,q_2) \qquad [17]$$

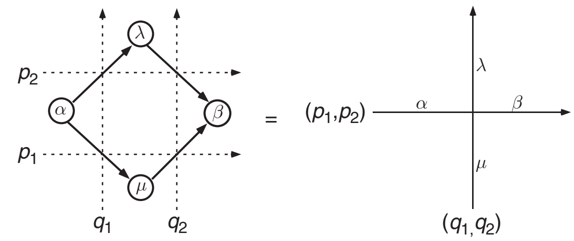



**Figure 13** Square weight as vertex weight.

From any solution of [3] and [4] we can thus construct a solution of YBE [8]. This has been used by Bazhanov and Stroganov to relate the integrable chiral Potts model with a cyclic representation of the six-vertex model.

### Map to Checkerboard Vertex Model

The checkerboard IRF-vertex model formulation contains all other versions mentioned above as special cases. However, collecting the state variables in triples, we can immediately translate it to a vertex model version, writing

$$\omega_{\hat\alpha\hat\mu}^{\hat\lambda\hat\beta}(p,q) = W_{\alpha\mu}^{\lambda\beta}\big|_{ab}^{dc}(p,q), \quad \overline{\omega}_{\hat\alpha\hat\mu}^{\hat\lambda\hat\beta}(p,q) = \overline{W}_{\alpha\mu}^{\lambda\beta}\big|_{ab}^{dc}(p,q)$$
$$\text{if } \begin{cases} \hat\lambda = (d,\lambda,c), & \hat\beta = (b,\beta,c) \\ \hat\alpha = (a,\alpha,d), & \hat\mu = (a,\mu,b) \end{cases} \qquad [18]$$

$$\omega_{\hat\alpha\hat\mu}^{\hat\lambda\hat\beta}(p,q) = \overline{\omega}_{\hat\alpha\hat\mu}^{\hat\lambda\hat\beta}(p,q) = 0 \quad \text{otherwise} \qquad [19]$$

In eqn [19], we have set all vertex model weights zero that are inconsistent with IRF-vertex configurations. Clearly, the translation of IRF models and spin models to vertex models can be done similarly.

### Map to Spin Model

We can, furthermore, translate each vertex model with weights assigned as in Figures 6 or 9 into a spin model with weights as in Figure 2 by defining suitable spins in the black faces, after checkerboard coloring. Each spin is then defined to be the ordered set of states on the line segments of the vertex model, $\underline{a} = (\alpha_1, \alpha_2, \ldots)$, ordering the line segments counterclockwise starting at, say, 12 o'clock. We can then identify $\omega_{\alpha\mu}^{\lambda\beta}(p,q) = W_{\underline{a},\underline{b}}(p,q)$, $\overline{\omega}_{\alpha\mu}^{\lambda\beta}(p,q) = \overline{W}_{\underline{a},\underline{b}}(p,q)$. This is surely not very economical, as many of the weights will be equal, but it helps show that all different versions of the checkerboard YBE are formally equivalent.

Hence, we shall only use the vertex model language in the following. It is fairly straightforward to convert to the other formulations.

#### An sl($m|n$) Example

One fundamental example is a $Q$-state model for which the rapidities have $2Q+1$ components, $\boldsymbol{p}=(p_{-Q},\ldots,p_Q)$, $\boldsymbol{q}=(q_{-Q},\ldots,q_Q)$, etc., and the states on the line segments are arranged in strings of continuing conserved color. The vertex weights, for $\alpha$, $\beta$, $\lambda$, $\mu=1,\ldots,Q$, are given by

$$\omega_{\alpha\mu}^{\lambda\beta}(\boldsymbol{p},\boldsymbol{q})=\omega_{0\alpha\mu}^{\lambda\beta}(p_0,q_0)\frac{p_{+\lambda}q_{-\beta}}{q_{+\alpha}p_{-\mu}} \qquad [20]$$

with ($\rho\neq\sigma$)

$$\begin{aligned}\omega_{0\rho\rho}^{\rho\rho}(p_0,q_0)&=\mathcal{N}\sinh[\eta+\varepsilon_\rho(p_0-q_0)]\\ \omega_{0\sigma\rho}^{\rho\sigma}(p_0,q_0)&=\mathcal{N}G_{\rho\sigma}\sinh(p_0-q_0)\\ \omega_{0\sigma\rho}^{\sigma\rho}(p_0,q_0)&=\mathcal{N}\,\mathrm{e}^{(p_0-q_0)\mathrm{sign}(\rho-\sigma)}\sinh\eta\\ \omega_{0\alpha\mu}^{\lambda\beta}(p_0,q_0)&=0,\quad \text{otherwise}\end{aligned} \qquad [21]$$

where $\mathcal{N}$ is an arbitrary overall normalization factor and $\eta$ is a constant. Furthermore, $\varepsilon_\rho=\pm1$ for $\rho=1,\ldots,Q$, where $m$ of them equal +1 and $n$ of them equal –1. The $G_{\rho\sigma}$'s are constants satisfying $G_{\rho\sigma}=1/G_{\sigma\rho}$, which freedom is allowed because the number of $\rho$-$\sigma$ crossings minus the number of $\sigma$-$\rho$ crossings is fixed by the states on the boundary only, that is, the choice of $\alpha,\alpha',\beta,\beta',\gamma,\gamma'$ in YBE [8] and Figure 5.

The solution [20], [21] has many applications. The case $m=0$, $n=2$ leads to the general six-vertex model; the $m=0, n=n$ case produces the fundamental intertwiner of affine quantum group $U_q\widehat{\mathfrak{sl}}(n)$, whereas the case $m=2, n=1$ corresponds to the supersymmetric one-dimensional $t$–$J$ model.

## Operator Formulations

#### The $R$-Matrix

For a problem with $N$ rapidity lines, carrying rapidities $p_1,\ldots,p_N$, we can introduce a set of matrices $\boldsymbol{R}_{ij}(p_i,p_j)$, for $1\leqslant i<j\leqslant N$, with elements

$$R_{ij}(p_i,p_j)_{\alpha_1\ldots\alpha_N}^{\beta_1\ldots\beta_N}=\omega_{\alpha_i\alpha_j}^{\beta_j\beta_i}(p_i,p_j)\prod_{k\neq i,j}\delta_{\alpha_k}^{\beta_k} \qquad [22]$$

In terms of these, the YBE [8] can be rewritten in matrix form as

$$\begin{aligned}&\boldsymbol{R}_{jk}(p_j,p_k)\boldsymbol{R}_{ik}(p_i,p_k)\boldsymbol{R}_{ij}(p_i,p_j)\\ &\quad=\boldsymbol{R}_{ij}(p_i,p_j)\boldsymbol{R}_{ik}(p_i,p_k)\boldsymbol{R}_{jk}(p_j,p_k)\end{aligned} \qquad [23]$$

where $1\leqslant i<j<k\leqslant N$.

#### The $\check{R}$-Matrix

If we transpose the $\beta$ indices $\beta_i$ and $\beta_j$ in eqn [22], we can define a set of matrices $\check{\boldsymbol{R}}_{i,i+1}(p,q)$ with elements

$$\check{\boldsymbol{R}}_{i,i+1}(p,q)_{\alpha_1\ldots\alpha_N}^{\beta_1\ldots\beta_N}=\omega_{\alpha_i,\alpha_{i+1}}^{\beta_i,\beta_{i+1}}(p,q)\prod_{k\neq i,i+1}\delta_{\alpha_k}^{\beta_k} \qquad [24]$$

Using these, the YBE [8] can be rewritten in matrix form as

$$\begin{aligned}&\check{\boldsymbol{R}}_{i,i+1}(q,r)\check{\boldsymbol{R}}_{i+1,i+2}(p,r)\check{\boldsymbol{R}}_{i,i+1}(p,q)\\ &\quad=\check{\boldsymbol{R}}_{i+1,i+2}(p,q)\check{\boldsymbol{R}}_{i,i+1}(p,r)\check{\boldsymbol{R}}_{i+1,i+2}(q,r)\end{aligned} \qquad [25]$$

and

$$[\check{\boldsymbol{R}}_{i,\,i+1}(p,q),\check{\boldsymbol{R}}_{j,\,j+1}(r,s)]=0,\quad \text{if } |i-j|\geqslant2 \qquad [26]$$

In this formulation, it is clear that many solutions can be found "Baxterizing" Temperley–Lieb and Iwahori–Hecke algebras.

## Classical YBEs

If we expand

$$\boldsymbol{R}_{ij}(p_i,p_j)=\boldsymbol{1}+\hbar\boldsymbol{X}_{ij}(p_i,p_j)+\mathrm{O}(\hbar^2) \qquad [27]$$

in [23], we get in second order in $\hbar$ the classical YBE (CYBE) as the vanishing of a sum of three commutators, that is,

$$\begin{aligned}&[\boldsymbol{X}_{ij}(p_i,p_j),\boldsymbol{X}_{ik}(p_i,p_k)]+[\boldsymbol{X}_{ij}(p_i,p_j),\boldsymbol{X}_{jk}(p_j,p_k)]\\ &\quad+[\boldsymbol{X}_{ik}(p_i,p_k),\boldsymbol{X}_{jk}(p_j,p_k)]=0\end{aligned} \qquad [28]$$

introduced by Belavin and Drinfel'd, where $\boldsymbol{X}_{ij}$ is called the classical $r$-matrix.

## Reflection YBEs

Cherednik and Sklyanin found a condition determining the solvability of systems with boundaries, the reflection YBEs (RYBEs), see Figure 14. Upon

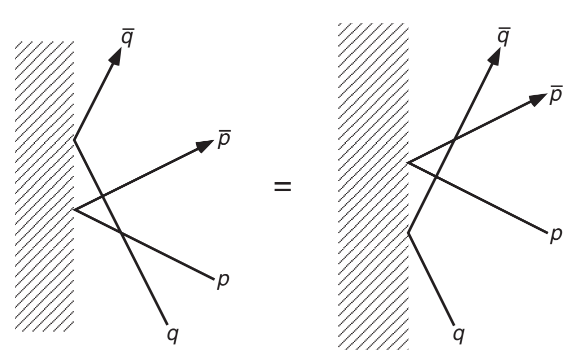



**Figure 14** Reflection YBE.

collisions with a left or right wall the rapidity variable changes from $p$ to $\overline{p}$ and back. In most examples, in which the rapidities are difference variables such that $\boldsymbol{R}(p,q)=\boldsymbol{R}(p-q)$, one also has $\overline{p}=\mu-p$, with $\mu$ some constant. The corresponding left boundary weights are $K_{\alpha}^{\beta}(p,\overline{p})$ satisfying

$$\check{\boldsymbol{K}}_1(q,\overline{q})\check{\boldsymbol{R}}_{12}(\overline{p},q)\check{\boldsymbol{K}}_1(p,\overline{p})\check{\boldsymbol{R}}_{12}(q,p) = \check{\boldsymbol{R}}_{12}(\overline{p},\overline{q})\check{\boldsymbol{K}}_1(p,\overline{p})\check{\boldsymbol{R}}_{12}(\overline{q},p)\check{\boldsymbol{K}}_1(q,\overline{q}) \quad [29]$$

with $\check{\boldsymbol{K}}_1(p,\overline{p})$ defined by a direct product as in [24] appending unit matrices for positions $i\geqslant 2$, and a similar equation must hold for the right boundary. Most work has been done for vertex models, while Pearce and co-workers wrote several papers on the IRF-model version.

## Higher-Dimensional Generalizations

In 1980 Zamolodchikov introduced a three-dimensional generalization of the YBE, the so-called tetrahedron equations (TEs), and he found a special solution. Baxter then succeeded in proving that this solution satisfies all TEs. Baxter and Bazhanov showed in 1992 that this solution can be seen as a special case of the sl($\infty$) chiral Potts model. Several authors found further generalizations more recently.

## Inversion Relations

When $\omega_{\alpha\mu}^{\lambda\beta}(p,\,p) \propto \delta_{\alpha}^{\lambda}\delta_{\mu}^{\beta}$, that is, the weight decouples when the two rapidities are equal, one can derive the local inverse relation depicted in Figure 15, which is a generalization of the Reidemeister move of type II in Figure 4. It is easily shown that $C(q,p)=C(p,q)$.

This local relation implies also a global inversion relation which can be found in many ways. The following heuristic way is the easiest: consider the situation in Figure 16, with $N$ closed $p$-rapidity lines and $M$ closed $q$-rapidity lines. For $M$ and $N$ large, we may expect the partition function of Figure 16 to factor asymptotically in top- and bottom-half contributions. If each line segment carries a state

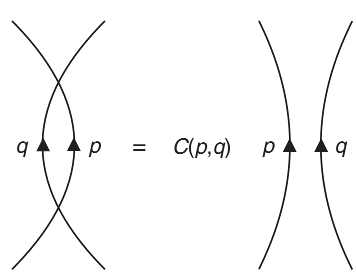



**Figure 15** Local inversion relation.

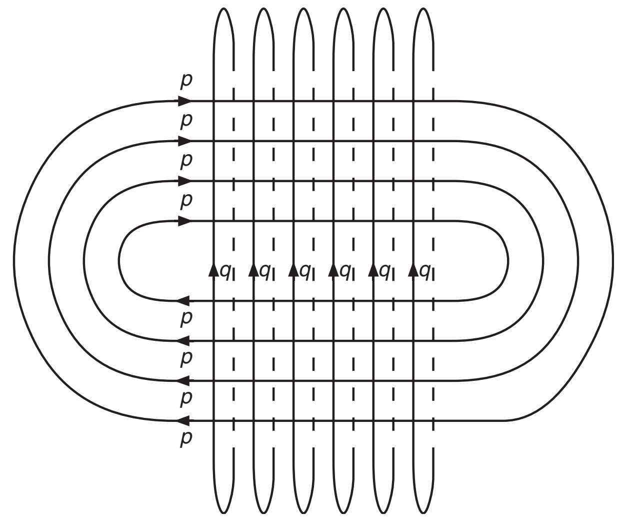



**Figure 16** Heuristic derivation of inversion relation.

variable that can assume $Q$ values, then the total partition function factors by repeated application of the relation in Figure 15 into the contribution of $M+N$ circles. Therefore,

$$Z = Q^{M+N}C(p,q)^{MN} \approx Z_{M,N}(p,q)Z_{N,M}(q,p) \quad [30]$$

Taking the thermodynamic limit,

$$z(p,q) \equiv \lim_{M,N\to\infty} Z_{M,N}(p,q)^{1/MN} \quad [31]$$

one finds

$$z(p,q)z(q,p) = \mathrm{C}(q,p) \quad [32]$$

In many models, eqn [32], supplemented with some suitable symmetry and analyticity conditions, can be used to calculate the free energy per site.

*See also:* Affine Quantum Groups; Bethe Ansatz; Classical *r*-matrices, Lie Bialgebras, and Poisson Lie Groups; Eight Vertex and Hard Hexagon Models; Hopf Algebras and *q*-Deformation Quantum Groups; Integrability and Quantum Field Theory; Integrable Discrete Systems; Integrable Systems: Overview; The Jones Polynomial; Knot Invariants and Quantum Gravity; Knot Theory and Physics; Sine-Gordon Equation; Topological Knot Theory and Macroscopic Physics; Two-Dimensional Ising Model; von Neumann Algebras: Subfactor Theory.

## Further Reading

Au-Yang H and Perk JHH (1989) Onsager's star-triangle equation: master key to integrability. *Advanced Studies in Pure Mathematics* 19: 57–94.

Baxter RJ (1982) *Exactly Solved Models in Statistical Mechanics*. London: Academic Press.

Behrend RE, Pearce PA, and O'Brien DL (1996) Interaction-round-a-face models with fixed boundary conditions: the ABF fusion hierarchy. *Journal of Statistical Physics* 84: 1–48.

Gaudin M (1983) *La Fonction d'Onde de Bethe*. Paris: Masson.

Jimbo M (ed.) (1987) *Yang–Baxter Equation in Integrable Systems*. Singapore: World Scientific.

Kennelly AE (1899) The equivalence of triangles and three-pointed stars in conducting networks. *Electrical World and Engineer* 34: 413–414.

Korepin VE, Bogoliubov NM, and Izergin AG (1993) *Quantum Inverse Scattering Method and Correlation Functions*. Cambridge: Cambridge University Press.

Kulish PP and Sklyanin EK (1981) Quantum spectral transform method. Recent developments. In: Hietarinta J and Montonen C (eds.) *Integrable Quantum Field Theories*, Lecture Notes in Physics, vol. 151, pp. 61–119. Berlin: Springer.

Lieb EH and Wu FY (1972) Two-dimensional ferroelectric models. In: Domb C and Green MS (eds.) *Phase Transitions and Critical Phenomena*, vol. 1, pp. 331–490. London: Academic Press.

Onsager L (1971) The Ising model in two dimensions. In: Mills RE, Ascher E, and Jaffee RI (eds.) *Critical Phenomena in Alloys, Magnets and Superconductors*, pp. xix–xxiv, 3–12. New York: McGraw-Hill.

Perk JHH (1989) Star-triangle equations, quantum Lax pairs, and higher genus curves. *Proceedings of Symposia in Pure Mathematics* 49(1): 341–354.

Perk JHH and Schultz CL (1981) New families of commuting transfer matrices in $q$-state vertex models. *Physics Letters A* 84: 407–410.

Perk JHH and Wu FY (1986) Graphical approach to the nonintersecting string model: star-triangle equation, inversion relation, and exact solution. *Physica A* 138: 100–124.

Reidemeister K (1926a) Knoten und Gruppen. *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität* 5: 7–23.

Reidemeister K (1926b) Elementare Begründung der Knotentheorie. *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität* 5: 24–32.

Yang CN (1967) Some exact results for the many-body problem in one dimension with repulsive delta-function interaction. *Physical Review Letters* 19: 1312–1314.

Yang CN (1968) $S$-matrix for the one-dimensional $N$-body problem with repulsive or attractive $\delta$-function interaction. *Physical Review* 167: 1920–1923.